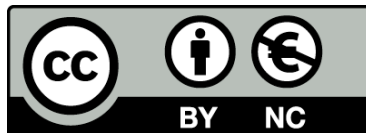




UNIVERSITAT_{DE}
BARCELONA

β -cells cis-regulatory networks and type 1 diabetes

Mireia Ramos Rodríguez



Aquesta tesi doctoral està subjecta a la llicència **Reconeixement- NoComercial 4.0. Espanya de Creative Commons**.

Esta tesis doctoral está sujeta a la licencia **Reconocimiento - NoComercial 4.0. España de Creative Commons**.

This doctoral thesis is licensed under the **Creative Commons Attribution-NonCommercial 4.0. Spain License**.



Programa de Doctorat en Biomedicina
Facultat de Biologia

β -cells cis-regulatory networks and type 1 diabetes

Dissertation submitted by
Mireia Ramos Rodríguez
to opt for a
Doctoral Degree from Universitat de Barcelona

Thesis developed under the supervision of Dr. Lorenzo Pasquali at the Endocrine Regulatory Genomics Lab in Fundació Institut d'Investigació en Ciències de la Salut Germans Trias i Pujol (IGTP).

Universitat de Barcelona
2016-2020

A handwritten signature in black ink, appearing to read 'Lorenzo Pasquali'.

Lorenzo Pasquali
Director

A handwritten signature in blue ink, appearing to read 'Modesto Orozco'.

Modesto Orozco
Tutor

A handwritten signature in black ink, appearing to read 'Mireia Ramos Rodríguez'.

Mireia Ramos Rodríguez
Candidate

Amb els suport de la Secretaria d'Universitats i Recerca de la Generalitat de Catalunya i del Fons Social Europeu.



Generalitat de Catalunya
Departament d'Empresa i Coneixement
Secretaria d'Universitats i Recerca



Unió Europea
Fons social europeu
L'FSE inverteix en el teu futur

*To Isra,
my family,
and my three mus(cat)eers*

Acknowledgements

This PhD thesis was written during the world-wide COVID19 pandemic. Yes, it sounds like the beginning of a science-fiction novel but it is not. A PhD is already an amazing, though challenging period but I guess no one would have imagined that some of us would be writing the thesis during a nation-wide lockdown. I guess life is really full of surprises.

First and foremost, I would like to thank my thesis director, Lorenzo. I remember when I first interviewed for this position, he told me that doing a PhD in a new small lab might be challenging, but also has many upsides. Now, after almost five years, I totally agree. Another key person during this PhD has been Helena, my better, wet-lab half. All the amazing findings summarized here could not be possible without her expertise and patience in the lab, but also her comfort and good advice in the hardest moments of the last few years. And of course, all the subsequent additions to our lab: Richard, Marc and Bea. Thank you for always being there to discuss science or just to hang out after a long hard day.

Our collaborators in the Décio Eizirik's lab in Brussels have also been key to the development of this thesis. Thank you Décio, Maikel, Jonàs and Jean-Valery, for helping with the experiments, bioinformatic analysis and with amazing exchanges on the results we've been obtaining throughout the years. I'd also like to thank Lorís for sharing his excitement for the Islet Regulome Browser with me and Iñaki and Lloyd for providing the technical support to develop both the Islet Regulome and all the bioinformatic analyses presented in this thesis.

All these years would have been much harder if I had not had the luck of meeting all the amazing past and present members of the *Endocrinologia i Nutrició* department at IGTP. Joan, Núria, Júlia, Mireia, Irene, David, Lorena, Esme, thank you for your advice and your invaluable company and friendship! I'd also like to extend that to all other groups in the IGTP and IJC that have attended the seminars and given me valuable input on my analyses.

Thank you to all past and present PhD committee members for organizing the most amazing activities in the Can Ruti Campus and for trying to improve the experience of all PhD students in the campus. Thanks especially to Marguerite-Marie, Laure, Bruna and Ana, for picking up the slack when things got crazy or complicated.

During these years, I have also had the honor of being a co-organizer of R-Ladies Barcelona. Thanks to all the R-Ladies community and especially to the co-organizers Ania, Sarah and Natàlia. We have organized some amazing meetings and workshops and created a great R-Ladies community!

To my college friends (aka BnD) Marina, Xavi, Manu, Tiffany, Silvia, Ferran,

Fatima and Roger, thank you for awakening my passion for science and for still being there after all this years!

Finally, to my friends and family, who have been there throughout all this exciting but difficult stage: thank you and I love you all. Thank you for always making the effort of trying to understand the topic of my research and keep listening to my day-to-day problems even when you have no idea what I'm talking about. Mama i Enrique, gràcies per tots els sopars i consells. Papa i Gina, gràcies pel suport i els dies de platja. Artistocuarzos, gràcies per les estones divertides i els sopars de posar-nos al dia. Isra, gracias por todo, absolutamente todo. Sin ti esto no habría sido posible.

Abstract

Type 1 Diabetes (T1D) is a β -cell-targeted autoimmune disease, leading to a reduction in pancreatic β -cell mass that renders patients insulin-dependent for life. In early stages of the disease, cells from the immune system infiltrate pancreatic islets in a process called insulinitis. During this stage, a cross-talk is established between cells in the pancreatic islets and the infiltrating immune cells, mediated by the release of cytokines and chemokines. Studying the gene regulatory networks driving β cell responses during insulinitis, will allow us to pinpoint key gene pathways leading to β -cell loss-of-function and apoptosis, and also to understand the role β cells have in their own demise. In the present thesis, we used two different cytokine cocktails, IFN- α and IFN- γ + IL-1 β , to model early and late insulinitis, respectively. After exposing β cells and pancreatic islets to such proinflammatory cytokines, we characterized the changes in their chromatin landscape, gene networks and protein profiles. Using both models, we observed dramatic chromatin remodeling in terms of accessibility and/or H3K27ac histone modification enrichment, coupled with up-regulation of the nearby genes and increased abundance of the corresponding protein. Mining gene regulatory networks of β -cells exposed to IFN- α revealed two potential therapeutic interventions which were able to reduce interferon signature in β cells: 1) Inhibition of bromodomain proteins, which resulted in a down-regulation of IFN- α -induced *HLA-I* and *CXCL10* expression; 2) Baricitnib, a JAK1/2 inhibitor, which was able to reduce both IFN- α -induced *HLA-I* and *CXCL10* expression levels and β cell apoptosis. In β cells exposed to IFN- γ + IL-1 β , we were able to identify a subset of novel regulatory elements uncovered upon the exposure, which we named Induced Regulatory Elements (IREs). Such regions were enriched for T1D-associated risk variants, suggesting that β cells might carry a portion of T1D genetic risk. Interestingly, we identified two T1D lead variants overlapping IREs, in which the risk allele modulated the IRE enhancer activity, exposing a potential T1D mechanism acting through β cells. To facilitate the access to these genomic data, together with other datasets relevant for the pancreatic islet community, we developed the Islet Regulome Browser (<http://www.isletregulome.org/>), a free web application that allows exploration and integration of pancreatic islet genomic data.

Contents

Abbreviations	3
I General introduction	7
1 Chromatin and gene expression regulation	9
1.1 DNA and chromatin structure	9
1.2 Chromatin three-dimensional organization	11
1.3 Regulation of gene expression	14
1.4 Characterization of regulatory elements using chromatin features	18
1.5 Chromatin dynamics in response to external stimuli	26
2 Pancreatic islets and glucose homeostasis	29
2.1 Pancreatic islet composition	29
2.2 Glucose homeostasis	30
2.3 Deciphering regulatory maps in pancreatic islets	31
3 Type 1 diabetes	33
3.1 Epidemiology	33
3.2 Aetiological factors	34
3.3 Pathogenesis	41
3.4 The role of islet inflammation	43
3.5 Current treatments and perspectives	46
II Hypothesis and Objectives	49
4 Hypothesis and Objectives	51
III Results	53
5 Preamble	55
Impact factor report from the thesis director	57
6 Interferon-α to model human β-cell responses to early insulinitis	59
6.1 Graphical abstract	59
6.2 Highlights	60
6.3 Article	61
7 Interferon-γ and Interleukin-1β to model human β cell responses to late insulinitis	79

7.1	Graphical abstract	79
7.2	Highlights	80
7.3	Article	81
8	The Pancreatic Islet Regulome Browser	101
8.1	Graphical abstract	101
8.2	Highlights	102
8.3	Article	103
8.4	State of the art of the Islet Regulome Browser	111
IV	Discussion	117
9	Regulatory networks driving responses to cytokines in β cells	119
9.1	Proinflammatory cytokines induce extensive β cell chromatin remodeling	119
9.2	Uncovering novel regulatory elements	121
9.3	Tissue-specific mechanisms in the response to proinflammatory cytokines	122
9.4	Changes in 3D chromatin structure	122
10	Exploring gene networks to find T1D therapeutic targets	125
11	Finding a mechanism for the β cell role in T1D using GWAS	127
12	Facilitating access to pancreatic islet genomic data	129
V	Conclusions	131
13	Conclusions	133
VI	References	135
14	References	137
VII	Appendix	155
A	Supplementary Information Colli et al. 2020	157
B	Supplementary Information Ramos-Rodríguez et al. 2019	193

Abbreviations

Abbreviation	Term
250HD	25-hydroxyvitamin D
4C	Circular Chromosome Conformation Capture
ATAC-seq	Assay for Transposase Accessible Chromatin sequencing
bp	Base pairs
ChIC	Chromatin ImmunoCleavage
ChIP-seq	Chromatin Immunoprecipitation sequencing
ChromHMM	Chromatin Hidden Markov Model
CORE	Cluster of Open Regulatory Elements
CpG	5'-Cytosine-phosphate-Guanine-3'
CTCF	CCCTC binding factor
CUT&RUN	Cleavage Under Targets and Release Using Nuclease
CUT&Tag	Cleavage Under Targets and Tagmentation
DiaMond	Diabetes Mondiale Project
DiViD	Diabetes Virus Detection Study
DNA	Deoxiribonucleic Acid
dsRNA	Dobule-Stranded RNA
ENCODE	Encyclopedia of DNA Elements
ER	Endoplasmic Reticulum
FACS	Fluorescence-Activated Cell Sorting
FAIRE-seq	Formaldehyde-Assisted Isolation of Regulatory Elements using sequencing
GAD65	65 kDa Glutamic Acid Decarboxylase
GTF	General Transcription Factor
GWAS	Genome-Wide Association Studies
HLA	Human Leukocyte Antigen
IAPP	Islet Amyloid Polypeptide

(continued)

Abbreviation	Term
IFN	Interferon
IL	Interleukin
INS	Insulin
IRB	Islet Regulome Browser
IRE	Induced Regulatory Element
kb	Kilobase
LPS	Lipopolysaccharide
Mb	Megabase
MHC	Major Histocompatibility Complex
mRNA	Messenger Ribonucleic Acid
OCR	Open Chromatin Region
pA-MNase	Protein A/Micrococcal Nuclease
PCR	Polymerase Chain Reaction
PIC	Pre-Initiation Complex
Pol II	RNA Polymerase II
PRR	Pattern Recognition Receptor
RE	Regulatory Element
SNP	Single-Nucleotide Polymorphism
T1D	Type 1 Diabetes
T2D	Type 2 Diabetes
TAD	Topologically Associating Domain
TBP	TATA-box-Binding Protein
TCGA	The Cancer Genome Atlas
TF	Transcription Factor
TFIID	Transcription Factor IID
TGF	Tumor Growth Factor
TLR	Toll-like Receptor
TNF	Tumor Necrosis Factor
TSS	Transcription Start Site

(continued)

Abbreviation	Term
UMI	Unique Molecular Identifier
UMI-4C	4C using Unique Molecular Identifiers
WAPL	Wings Apart-Like protein

Part I

General introduction

1

Chromatin and gene expression regulation

1.1 DNA and chromatin structure

Deoxyribonucleic acid, globally known with its acronym DNA, is a molecule formed by two complementary nucleotide chains that are structured as a double helix. The nucleotides are molecules that compose DNA and consist of one of four nitrogenous bases (cytosine (C), guanine (G), adenine (A) or thymine ((T)), deoxyribose and a phosphate group (**Figure 1.1 A**). As the nitrogenous bases are the elements that distinguish one nucleotide from another, they are usually used to identify the whole monomer. Therefore, a DNA molecule can be reduced to a sequence of nucleotides (A, C, T or G).

DNA can be subject to biologically relevant modifications that do not alter the nucleotide sequence. This is the case of **DNA methylation**, which consists on the addition of a methyl group to the cytosine residues of CpG dinucleotides (**Figure 1.1 B** and **Figure 1.4 A**). These dinucleotides can be scattered along the genome or concentrated in regions with high CpG density, named **CpG islands**.

In humans, each of the diploid cells contains 46 DNA molecules corresponding to 23 pairs of chromosomes, and a smaller DNA molecule found within mitochondria. The nucleotide sequences corresponding to all these DNA molecules is known as the **human genome** and contain all the genetic information necessary for proper development and function of a human individual.

Genes are units of genome sequences that are transcribed into mRNA – another nucleic acid – by the RNA polymerase II (Pol II). The mRNA molecules will in turn leave the cell's nucleus and be translated into proteins, which regulate cell metabolism and function. In a single gene, we can find several stretches of nucleotides that are translated into mRNA – exons – intertwined with non-coding regions – introns. Many times, gene exons are referred to as coding regions or sequences – as they code for proteins –, representing around 2% of the human genome (Carninci et al. 2005).

For a long time it was thought that only the coding sequences in our genome had biological relevance. The rest of the genome (98%), consisting of non-coding sequences, was disregarded and referred to as “junk DNA”. In 2012, the Encyclopedia of DNA Elements (ENCODE) project published

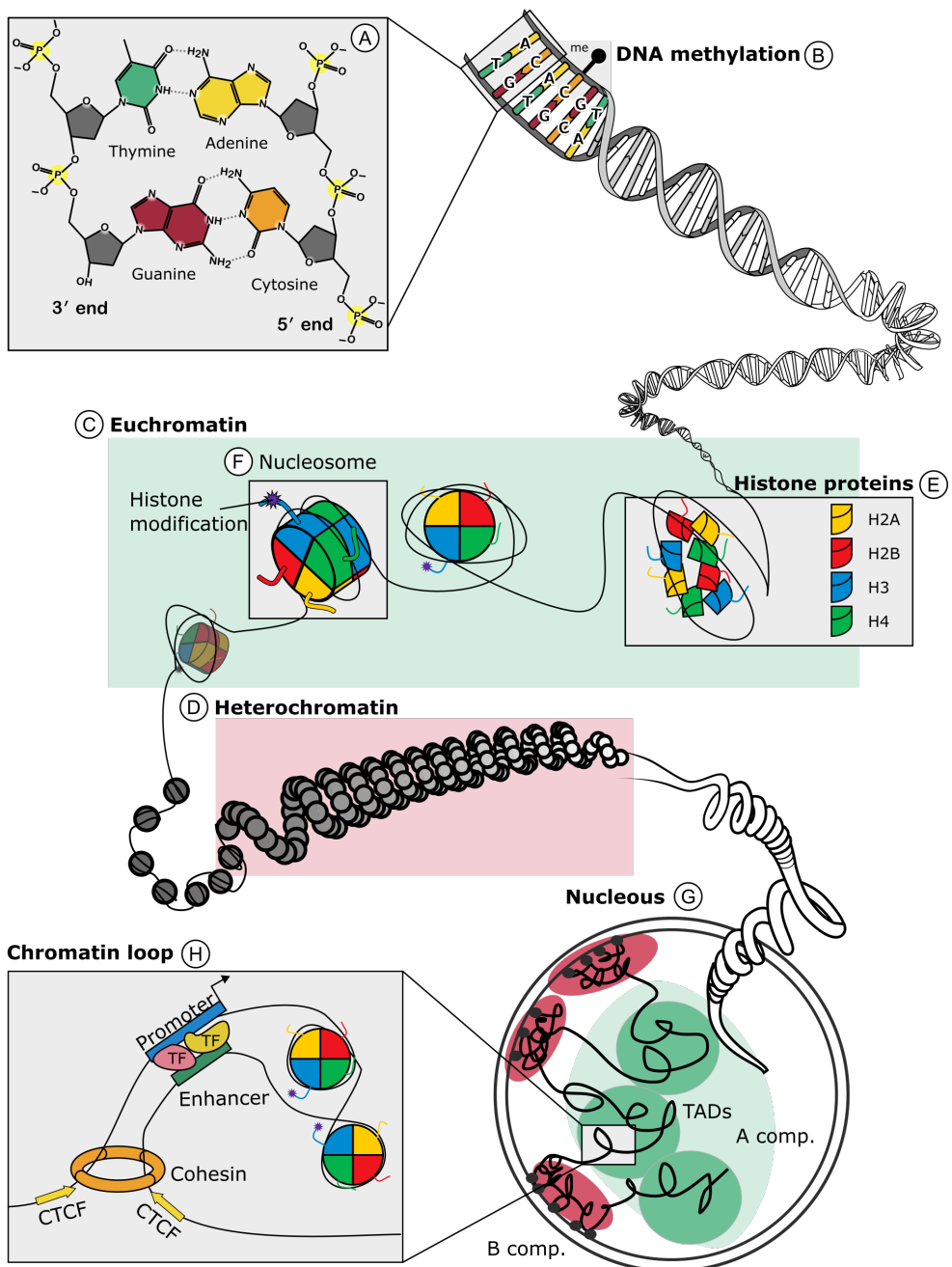


Figure 1.1: DNA and chromatin structure. **A**, The DNA molecule is composed of nucleotides, which can contain one of four nitrogenous bases (Adenine, Thymine, Cytosine and Guanine). **B**, DNA methylation involves the covalent addition of a methyl group to a cytosine followed by a guanine. **C** and **D**, Chromatin can be classified into euchromatin, more dispersed and accessible (**C**), and heterochromatin, more compacted (**D**). **E**, Histone complexes are formed by two copies of each of the four core histones. **F**, A nucleosome is a structure formed by the histone complex and the DNA portion wrapped around it. **G**, Chromatin inside the cell nucleus is organized in different domains of interaction of variable size, such as A/B compartments (comp.) and TADs. **H**, Chromatin loops are the smaller domains of interaction, where enhancer-promoter contacts take place.

several decade-long studies that suggested that 80% of the human genome has a biochemically relevant function (Pennisi 2012), thus challenging the conception of non-coding sequences as mere “junk”. During the last decade, many studies have unraveled functions of the non-coding genome, including non-coding genes and regulatory elements, which are indispensable for the regulation of gene transcription.

If stretched end-to-end, the DNA contained in one cell measures about 2 meters long, and is tightly folded to fit inside the cell nucleus, which has an average diameter of 6 micrometers (μm). This is achieved with the help of many different proteins that bind DNA. The structure formed by DNA and all the proteins bound to it form what we know as **chromatin**. Thus, these DNA-bound proteins affect packaging and condensation of chromatin. Taking this into consideration, chromatin can be classified into *euchromatin*, which is more disperse and accessible, and *heterochromatin*, which is more tightly packaged (**Figure 1.1 C-D**). Interestingly, this condensed or accessible structure also has a biological meaning: euchromatin is enriched in actively transcribed regions, as it allows the transcription machinery to access DNA; heterochromatin, on the other hand, is enriched for inactive and repressed regions.

Histones are key proteins that bind and help package DNA. Histone complexes are organized in the form of octamers, where each complex contains two copies of the four core histones (H2A, H2B, H3 and H4) (**Figure 1.1 E**). The different histone proteins have long tails that allow different modifications to be added covalently, which can affect the function and packaging of the neighboring DNA sequences (Bannister and Kouzarides 2011; Buschbeck and Hake 2017). Thus, a portion of the DNA molecule (~150 base pairs [bp]) is wrapped around the histone complex, resulting in a tight packaging favored by the negative charge of the DNA being attracted to the positive charge of the histones. This structure formed by the histone octamer and the 150 bp DNA portion wrapped around it is called a **nucleosome** (**Figure 1.1 F**).

1.2 Chromatin three-dimensional organization

Genome organization inside the cell's nucleus is highly complex. Chromosomes are folded in different hierarchical domains resulting in functional compartments that are biologically relevant, as they are key for maintenance of proper gene expression regulation.

Several architectural proteins are involved in maintaining the different layers of genome organization. Some of the key players in maintaining genome architecture are the CCCTC binding factor (CTCF), the cohesin complex and the mediator complex. The cohesin complex consists of three core subunits (SMC1, SMC3 and RAD21) that form a ring shaped structure. This complex

usually co-localizes with CTCF, a transcriptional repressor that is involved in gene expression regulation and insulator activity (Phillips and Corces 2009). However, cohesin binds to thousands of genomic sites independently of CTCF, many of which are co-bound by the mediator complex. The mediator complex is essential for Pol II transcription and is thus proposed to also have a role in mediating transcriptional regulation.

Taking into account the different probabilities and size of interactions, we can identify hierarchical layers of genome organization of decreasing sizes (**Figure 1.1 G**) (Kempfer and Pombo 2020):

- **Chromosome territories.** These domains are the highest order of nuclear organization, consisting on different chromosomes occupying specific territories within the cellular nucleus (Meaburn and Misteli 2007). This layer of organization favors intra-chromosomal chromatin interactions.
- **A/B compartments.** These multi-megabase-sized domains distinguish between active and gene-rich regions (“A” compartment) and silenced and compact regions (“B” compartment).
- **Topologically Associating Domains (TADs).** TADs are conserved and stable domains with increased interactions between elements within the same domain, usually ~1Mb in size.
- **Chromatin loops.** Chromatin loops or “insulated neighborhoods” function as structural units of gene expression control.

1.2.1 A/B compartments

A/B compartments are multi-megabase spatial domains with increased probability of genomic interactions in regions found within the same compartment. They were first identified in Hi-C experiments (Lieberman-Aiden et al. 2009; Fortin and Hansen 2015).

The “A” or active compartment contains gene-rich regions and is enriched for accessible chromatin regions and histone modifications associated with active transcription. On the other hand, the “B” or inactive compartment tends to be gene-poor and compact, contains histone marks for gene silencing, and overlaps with lamina-associating domains.

The composition of A/B compartments seems to be plastic and cell-type specific. For example, during human cell lineage specification extensive (36%) compartment switching was reported (Dixon et al. 2015). Additionally, Schmitt et al. (2016) analyzed 21 primary human tissues and observed over 60% of compartment switching between different cell types.

In contrast to TADs and chromatin loops, the formation of A/B compartments seems to be independent of cohesin, as shown by some recent cohesin-depletion studies (Schwarzer et al. 2017; Rao et al. 2017).

1.2.2 Topologically Associating Domains

TADs are submegabase-size domains that have increased contact frequency of intra-domain interactions (Nora et al. 2012; Laat and Duboule 2013; Dixon et al. 2012). They were first characterized genome-wide in mouse embryonic stem cells, where more than 90% of the genome was found to be organized into ~2,200 TADs with a median size of 880 kilobases (kb). TADs are considered to be the fundamental unit of genome organization, as TADs are mostly stable across different cell types and even conserved across species (Dixon et al. 2012).

It has been proposed that TAD boundaries act as insulators that facilitate intra-TAD interactions – which would otherwise be too infrequent – and thus ensure proper transcriptional activation. In support of this theory, Javierre et al. (2016) observed that one-third of significant genomic interactions crossed TAD boundaries, which represents a smaller proportion than what would be expected by chance. Therefore, even though TADs are effective in insulating and increasing intra-domain interactions, they are not absolute, as some inter-TAD interactions might be found.

However, recent studies have challenged the conception of TADs as biologically relevant domains of genome organization, as they may be artifacts arising from the use low-resolution assays [Rowley2018]. Thus, by increasing the resolution of 3D chromatin assays, smaller interaction domains can be found, which appeared as homogeneous TADs when using lower resolutions. This smaller domains of interaction may correspond to sub-TADs (Phillips-Cremins et al. 2013), of 185 kb median size (Rao et al. 2014). Sub-TADs also display increased intra-domain interactions, with smaller contact frequencies with genomic regions outside that specific sub-domain. The main difference between TADs and sub-TADs is related to their conservation in different tissues: sub-TADs appear to be much more cell-type-specific. This variant sub-TAD partitioning may be caused by cell-type-specific interactions mediated by Mediator/Cohesin and lineage-specific proteins (Phillips-Cremins et al. 2013).

1.2.3 Chromatin loops

Chromatin loops allow long-range interactions between different genomic regions. Loops may be formed between two convergent CTCF sites that are co-bound by cohesin, establishing the insulator anchor (**Figure 1.1 H**). This insulation increases the interaction probability of two distant regions in the linear genomic space, which would otherwise happen at very low frequency due to random collisions.

A proposed model for the chromatin loop formation is the “loop extrusion model” (Sanborn et al. 2015; Fudenberg et al. 2016), in which CTCF, the cohesin

complex and their cofactors are responsible for establishing these spatial regulatory domains. When the cohesin complex is loaded onto the chromatin, it entraps the DNA fiber in the lumen of its ring, forming a small chromatin loop. This loop is then extruded until cohesin encounters two CTCF-bound sites in convergent orientation. At this point, the cohesin-unloading factor WAPL (Wings apart-like protein) removes cohesin from the chromatin.

The loop extrusion model is supported mainly by two observations:

1. Chromatin loops and TADs are dramatically weakened upon depletion of chromatin-bound cohesin (Wutz et al. 2017).
2. The amount of residence time of chromatin-bound cohesin directly determines the chromatin size, showing an increased loop size with increased cohesin residence time (Haarhuis et al. 2017; Wutz et al. 2017).

1.3 Regulation of gene expression

The human genome contains about 3 billion bp, which carry all the information needed to develop from a single cell – a fertilized egg – to an adult functional organism. During development, different rounds of gene transcription drive the differentiation of all cell types that compose an adult organism. Interestingly, even after the adult organism has developed, it is necessary to activate or inhibit expression of genes in order maintain cell function and to respond to external stimuli.

The switching on and off of gene transcription is mediated by different mechanisms including the modulation of non-coding DNA sequences named **cis-regulatory elements** which include **promoters** and **enhancers**, among others. Promoters are non-coding regions located around the transcription start site (TSS) of genes, that allow binding of different proteins to initiate gene transcription. Conversely, enhancers can be located from bases to megabases away from their target genes. They are also bound by different proteins, such as transcription factors (TFs), and may physically interact with their target gene promoter to modulate gene expression.

Additional mechanisms of gene expression regulation are also in place, such as non-coding RNAs. However, as they are not the focus of this thesis, they will not be discussed in this introduction.

1.3.1 Promoters and transcription initiation

Promoters are the basic units of gene regulation. They are indispensable for gene transcription, as promoter sequences allow the assembly of the Pol II transcription machinery and auxiliary factors that are key to start gene

transcription. As mentioned above, they are located adjacently to their target genes, surrounding their TSS.

We can divide promoters into three distinct elements (Hernandez-Garcia and Finer 2014) (**Figure 1.2**):

- **Core promoter.** The core promoter is a short sequence located 50 bp upstream and downstream of the TSS and represents the minimal sequence needed to initiate transcription. It contains specific binding motifs for general transcription factors (GTFs), which are proteins that form the transcription pre-initiation complex (PIC) together with Pol II. One example of GTF is the TATA-box-binding protein (TBP), part of the transcription factor IID (TFIID) complex, that mediates Pol II recruitment and PIC assembly.
- **Proximal promoter.** The proximal promoter is formed by the 250 bp sequence upstream of the core promoter. It is characterized by containing binding motifs for sequence-specific TFs that activate or inhibit the expression of the target gene.
- **Distal promoter.** Distal promoters are not as clearly characterized as the two elements above, but are usually defined as sequences up to 2 kb upstream of the TSS. Similarly to proximal promoters, they are also rich in TF binding motifs.

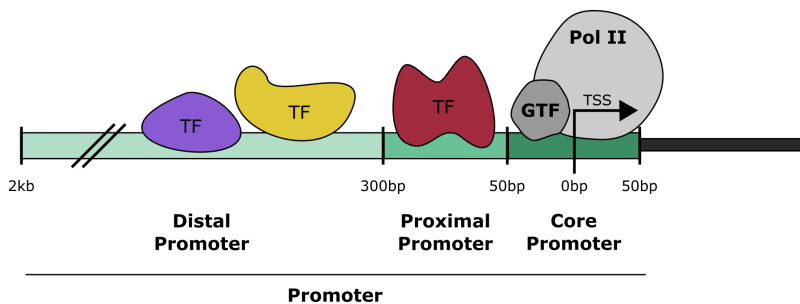


Figure 1.2: **Distinct sequence elements that compose promoters.**

The levels to which a promoter can activate gene transcription will depend on its sequence composition and its occupancy by sequence-specific TFs. For example the core promoter can induce basal transcription, but the presence of a proximal promoter can further increment the transcriptional activity. Additionally, enhancers can also interact with their target promoters to modulate gene transcription.

Promoter sequences also have a distinct nucleotide composition, presenting an elevated GC content compared to the rest of the genome. Moreover, 70% of human proximal promoters contain regions with high density of CpG dinucleotides, known as CpG islands (Saxonov, Berg, and Brutlag 2006).

These CpG islands can be methylated and thus prevent binding of the transcription machinery leading to the inhibition of gene transcription (see the *DNA methylation* section for more information).

In a simplified model, gene promoters can be classified based on their basal level of activity (Haberle and Lenhard 2016):

- **Constitutive promoters.** Constitutive promoters are those that present high expression levels in most cell types. They are used in many genetic engineering techniques to express foreign genes.
- **Inducible or regulated promoters.** Inducible or regulated promoters are activated in a cell-type specific molecular context, by external environmental stimuli to induce a specific transcriptional response.

1.3.2 Enhancers and tissue-specific regulation

Enhancers are non-coding regulatory sequences that are generally located distally relative to their target gene TSS. Their distribution is mostly intergenic, though some enhancers can also be found within genes, especially at introns. They are able to increase basal gene activity levels by interacting physically with their target gene promoter (**Figure 1.3**).

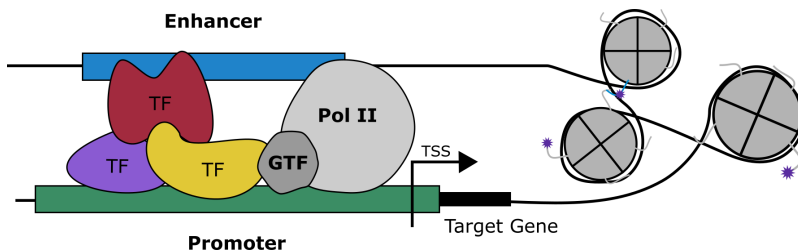


Figure 1.3: **Enhancers regulate gene expression by interacting with gene promoters.**

Enhancers are generally considered to be 200-500 bp in length. Their sequence composition facilitates the binding of TFs and the interaction with their target promoters resulting in the induction of gene transcription. Similarly to promoters, enhancers may contain CpG islands which can in turn repress enhancer activity if methylated (Bell and Vertino 2017).

The mechanism by which a single enhancer affects the transcription of a target gene is still poorly understood. Considering that enhancers are defined by their potential to increase gene transcription and this effect has been mostly studied in a bulk of cells, there are two models that could explain their individual action (García-González et al. 2016):

- The **binary model.** Enhancers increase the probability that a higher proportion of cells activate transcription at a given locus within a cell

population.

- The **progressive** or **rheostatic model**. Enhancers increase the number of RNA molecules transcribed from the target gene in a single cell, but not the number of cells that initiate transcription.

Until recently, it was not clear which of the above models was true. However, with the recent availability of single cell methodologies, Larsson et al. (2019) were able to show that, transcriptome-wide, core promoters affect mRNA burst size, that is, the number of RNA molecules transcribed, while enhancers regulate burst frequencies, which is proposed to be the primary driver of tissue-specific gene expression. This result goes in line with other observations suggesting that enhancers confer **tissue-specificity** to a linked gene. First, cell-type specific differences between histone modification patterns, TF binding and chromatin accessibility, which are key regulatory features, are mainly localized at enhancers and not promoters (Xi et al. 2007; Heintzman et al. 2009). Second, in functional assays performed with constructs containing an enhancer and a minimal promoter, the expression of the reporter gene is governed by the enhancer, even if enhancer and promoter come from genomic loci that have very different expression patterns *in vivo* (Bulger and Groudine 2010).

Additionally, we know that the relationship between enhancers and promoters is not linear, as an enhancer can regulate several promoters and a promoter can be regulated by many enhancers. Thus, enhancers are proposed to act as modular units of gene expression showing seemingly additive and/or redundant effects on their target genes. This redundancy ensures robustness of gene expression.

1.3.3 Transcription factors as modulators of gene transcription

TFs are proteins with domains that can bind short DNA sequences, usually enriched at enhancers and promoters, and participate in the modulation of gene transcription. The DNA sequences bound by a specific TF include a consensus sequence, known as **motif**. TF motifs are usually 6-10 bp long and contain degenerate positions in which different nucleotides facilitate the binding of the protein.

The presence of a TF motif, however, is not sufficient to direct TF binding. Instead, TFs bind to a small fraction of their binding motif occurrences in the genome. Additional factors that vary in different tissues and cell types facilitate TF binding; these include for example the chromatin context, the presence of partner TFs, enrichment of specific histone modification or the chromatin accessibility. As a result, broadly expressed TFs may exert cell-type specific binding and function.

As mentioned above, several TFs can be bound at the same time in close

physical proximity at an enhancer. There are currently two models that attempt to explain how this binding affects enhancer activity and transcription induction (García-González et al. 2016):

- The “**enhanceosome**” **model**. The enhancer DNA sequence acts as a scaffold for the ordered and cooperative binding of TFs to form a protein complex that can activate transcription. Enhancer activity emerges from a network of interactions and its action is lost just by the absence of one of the proteins involved.
- The “**billboard**” **model**. The binding of each TF is independent from each other and they do not act as a single unit.

1.4 Characterization of regulatory elements using chromatin features

Regulatory elements are key to understanding the cis-regulatory networks that guide gene expression. Because of their tissue- and state-specific nature, and despite the effort of the scientific community during the last decade in mapping regulatory maps (The ENCODE Project Consortium 2012; Martens and Stunnenberg 2013; Andersson et al. 2014), a comprehensive map and characterization of active regulatory elements in each human cell type developmental stage and disease relevant state is still far to be completed.

Proper characterization of cis-regulatory elements requires the depiction of three main features:

1. **Regulatory element genomic position.** Knowing the genomic coordinates of a regulatory element is key to study its sequence features, such as TF binding motifs, which can guide characterization of the networks involved in its activation.
2. **Regulatory element activity status.** Regulatory elements can be in various activity states, such as active, poised or repressed. Such states may be cell-type- and state-specific. Understanding the activity status can help clarify the role of the regulatory element in the cell’s gene regulatory programs.
3. **Regulatory element gene targets.** Identifying the putative gene targets of a regulatory element is challenging, as distal enhancers can be megabases away from their target genes.

Sequence features of enhancers alone, such as phylogenetic sequence conservation or TF motifs, have limited capacity to properly predict enhancer characteristics. Meanwhile, complementary information can be obtained by the study of chromatin structure that can inform on the regulatory element features, such as its position and/or activity.

In the following sections I will describe the main chromatin features that are used to identify regulatory elements, establish their activity level and help to pinpoint their putative target genes (**Table 1.1**), together with the available high-throughput technologies that allow retrieving genome-wide information of these chromatin elements.

Table 1.1: **Characterization of regulatory elements using chromatin features.** Summary of some of the main associations found between chromatin features and regulatory elements (RE) characteristics.

Chromatin feature	Main association
DNA methylation	Repressed or inactive RE.
Histone modification	Depending on the queried histone mark, can help identify active, poised or repressed RE. Can approximate the genomic coordinates of the RE.
Chromatin accessibility	Accessibility may be used as a proxy for active RE. Can approximate the genomic coordinates of the RE.
Chromatin contacts	Guides identification of RE putative target genes, which may be in physical contact.

1.4.1 DNA methylation

DNA methylation consists of the covalent addition of a methyl group to a cytosine (**Figure 1.4 A**) followed by a guanine, known as CpG sites. The methylated CpGs can be sparsely distributed along the genome or concentrated in what we call CpG islands, mainly located at repetitive sequences and promoter regions.

So far, DNA methylation has been associated with a large variety of biological processes such as development, tissue-specific gene expression, imprinting, inactivation of the X chromosome and silencing of repetitive DNA sequences (Law and Jacobsen 2010). Thus, dysregulation of DNA methylation is present in many different diseases, such as cancer (Klutstein et al. 2016) and psychiatric disorders, including autism (Tremblay and Jiang 2019) and schizophrenia (Pries, Gülöksüz, and Kenis 2017). Moreover, treatments that target DNA methylation have been approved for many cancer types (Pan et al. 2018).

DNA methylation – especially at CpG islands – is associated with transcriptional repression (Goll and Bestor 2005). Mechanistically, DNA methylation does not necessarily initiate gene repression, but rather maintains the repression stable (Curradi et al. 2002). Interestingly, DNA methylation has also been shown to physically interfere with the binding of some TFs to DNA (Yin et al. 2017).

DNA methylation does not guide the identification of the regulatory element genomic coordinates, as 70 to 80% of CpGs in mammal genomes are methylated (Jabbari and Bernardi 2004). Querying DNA methylation at specific regulatory elements, on the other hand, can provide insights on their activity levels: CpG methylation is associated with inactive regulatory elements. However, the absence of DNA methylation does not necessarily identify active regulatory elements.

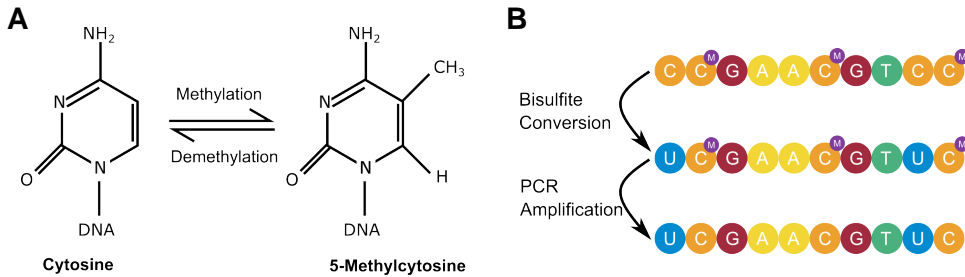


Figure 1.4: **DNA methylation.** **A**, Chemical structure of cytosine and 5-methylcytosine. **B**, Schematic of the bisulfite conversion process, followed by PCR amplification.

1.4.1.1 High-throughput assays for detecting DNA methylation

Most of the techniques developed to query genome-wide DNA methylation status take advantage of the **bisulfite conversion**, a process that turns unmethylated cytosines into uracils using sodium bisulfite (**Figure 1.4 B**). This process can be performed in the whole genome and coupled with massive sequencing (Bisulfite-seq) to obtain the methylation status of all the cytosines genome-wide. Alternatively, microarrays that query specific CpGs are broadly used to obtain high throughput information on the DNA methylation genome wide.

Within several commercially available platforms, Illumina provides two microarrays – HumanMethylation450 BeadChip (Infinium) methylation microarray (450K) (Sandoval et al. 2011) and MethylationEPIC BeadChip (Infinium) methylation microarray (850K) (Moran, Arribas, and Esteller 2016) – that query the methylation state of a comprehensive number of human CpGs, preferentially located in CpG islands, promoters or enhancers. Even though this technology fails to query DNA methylation status in all cytosines, it is the preferred method for large consortia such as The Cancer Genome Atlas (TCGA), due to their robustness and very high reproducibility (Moran, Arribas, and Esteller 2016).

1.4.2 Chromatin accessibility and nucleosome occupancy

Nucleosomes are key in the compaction of DNA, allowing it to fit inside the nucleus of a cell. However, the level of compaction of different genomic regions has a crucial functional role in determining the activity of regulatory elements. Thus, promoters of active genes and their putative enhancers are located in nucleosome-depleted regions. The “openness” of these regions allows the transcriptional machinery to bind and initiate gene transcription. Nucleosomes also act as “gatekeepers” of inactive regions, as the compacted chromatin is inaccessible to the binding of the transcriptional machinery. Chromatin accessible regions are also known as **DNase I hypersensitivity sites**, as the DNA is exposed and thus sensitive to cleavage by the DNase I enzyme.

Nucleosome occupancy is not fixed: different nucleosome-depleted regions – or open chromatin sites – have been identified in different cell types (Xi et al. 2007; The ENCODE Project Consortium 2012) and these accessible regions are dynamically regulated during development (Xu and Xie 2018) or upon exposure to different stimuli (Calderon et al. 2019).

Changes in chromatin accessibility are known to be regulated by different proteins, such as a specific class of TFs known as **pioneer transcription factors**. Pioneer factors, unlike other TFs, are able to bind compacted DNA sequences and recruit ATP-dependent chromatin modifiers, histone modifiers and other TFs, culminating in the activation of the underlying regulatory sequences (Zaret and Carroll 2011).

Thus, chromatin accessibility can be used as a proxy for the genomic location of functional regulatory elements (**Figure 1.5**). However, accessibility *per se* does not imply regulatory elements are in an active state, as some regions might become accessible by indirect mechanisms and not induce expression of their target genes.

1.4.2.1 High-throughput assays for detecting chromatin accessibility

Open chromatin regions can be interrogated using different methodologies that rely on the fact that open regions are accessible to DNA-cutting enzymes, as opposed to compacted regions. This process generates nucleosome-depleted fragments that can be sequenced and aligned to a reference genome.

Many different techniques have been developed throughout the years, including FAIRE-seq (Giresi et al. 2007), MNase-seq (Schones et al. 2008), DNase-seq (Boyle et al. 2008) and the Assay for Transposase-Accessible Chromatin using sequencing (ATAC-seq) (Buenrostro et al. 2013). ATAC-seq takes advantage of a hyperactive mutant Tn5 Transposase, that cuts DNA accessible regions and ligates sequencing adapters in a single step. Therefore,

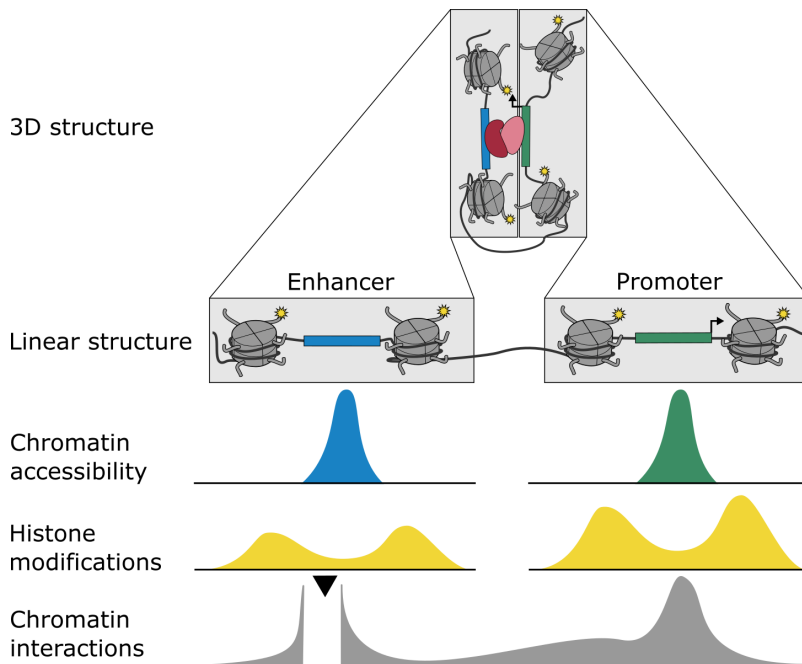


Figure 1.5: **Chromatin features allow identification of regulatory elements.** Summary of the profiles obtained when using specific high-throughput sequencing techniques to query different chromatin features.

it allows reducing the protocol time and the amount of starting material – even to a single cell level – to produce high-quality open chromatin profiles.

1.4.3 Histone modifications

Histones are proteins with a key role in the structure and function of chromatin. They form octamer complexes that are wrapped by DNA, forming structures called nucleosomes.

Interestingly, the role of histone complexes is not limited to DNA packaging: all four core histones have long N-terminal tails which are extruded from their central domains and can carry different covalent modifications. Such modifications at different N-terminal residues are associated with regulatory element activity and gene transcription.

Some of the most studied histone modifications affect the lysine 4 of histone H3 (H3K4). The monomethylation of H3K4 (H3K4me1) was the first modification found to be specifically enriched at distal enhancers, whereas the trimethylation of H3K4 (H3K4me3) was found to be associated with active promoters (Heintzman et al. 2007). Di- (H3K4me2) and tri-methylation (H3K4me3) of H3K4 can also be found at distal regulatory elements, though at lower enrichment compared to proximal regulatory elements (Trynka and

Raychaudhuri 2013). Interestingly, H3K4me1 is a marker for distal regulatory elements that can also be found at inactive enhancers, meaning that distal regulatory elements involved in the regulation of diverse processes, such as development, can be pre-labeled by this mark (Bernstein et al. 2006; Vastenhouw and Schier 2012). This suggests that H3K4 methylation might protect enhancers from being targeted by inhibitory complexes, such as DNA methylation enzymes, thus maintaining a poised state that may become active after the appropriate stimulus (Ooi et al. 2007).

The modifications in lysine 27 of histone H3 have become much more informative in terms of regulatory activity. Acetylation of this residue (H3K27ac) at both enhancers and promoters has been clearly associated with the induction of gene expression in the neighboring genes (Creighton et al. 2010; Rada-Iglesias et al. 2011). On the other hand, tri-methylation (H3K27me3) is a mark of active repression of regulatory elements that is deposited by the Polycomb repressive complex (Morey and Helin 2010; Margueron and Reinberg 2011). Usually, this mark is linked to facultative heterochromatin, a type of silent chromatin that is regulated during development. Defects in the deposition or removal of H3K27me3 have also been associated with the development of several types of cancer (Conway, Healy, and Bracken 2015).

In general, histone modifications can approximate the genomic location of regulatory elements, as they are present in the nucleosomes that delimit the accessible regulatory site (Figure 1.5).

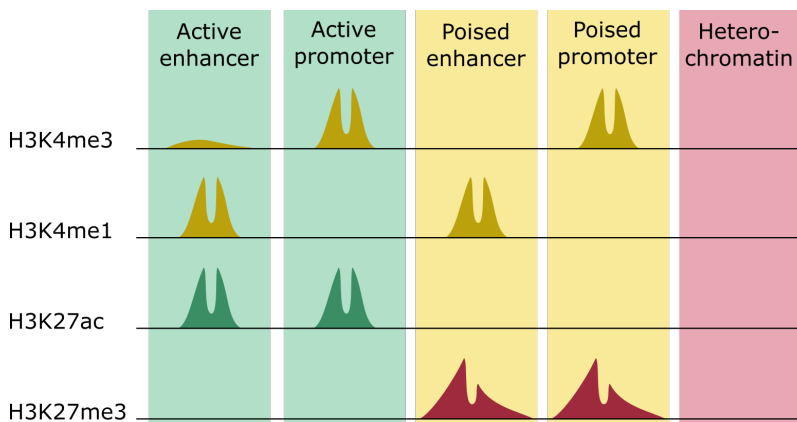


Figure 1.6: **Histone modifications as proxies of regulatory element activity.** Summary of some of the main histone modifications used to identify regulatory elements and characterize their activity status.

More importantly, querying histone marks can provide a clear picture of the regulatory element activity state. Enhancers marked with H3K4me1 and H3K27me3 and promoters marked with H3K4me3 and H3K27me3 can be classified as **poised**, which are actively repressed by the polycomb complex

but primed for activation. Conversely, enhancers enriched in H3K4me1 and H3K27ac and promoters marked with H3K4me3 and H3K27ac are considered **active**, as H3K27ac in both enhancers and promoters is correlated with the induction of gene expression (**Figure 1.6**).

Of note, correlation or association of histone modification enrichment with enhancer and/or promoter activity and gene transcription is not necessarily a causal clue, as it may reflect a bystander event or a secondary result of the transcriptional machinery recruitment and gene transcription activation.

1.4.3.1 High-throughput assays for detecting histone modifications

The most popular method for detecting histone modification enrichment is chromatin immunoprecipitation (ChIP), a technique that uses antibodies that recognize specific epitopes to selectively precipitate genomic regions enriched for different histone modifications (Collas 2010). The combination of this technique with high-throughput sequencing (ChIP-seq) allows querying genome-wide for the specific location of the histone modification. However, there are two main limitations of the classical ChIP-seq: it is strongly dependent on the quality of the antibody and requires a relatively large number of cells as starting material.

The number of cells required to perform the experiment represents a critical limitation of ChIP-seq, together with the fact that ChIP-seq protocols are quite tedious, time-consuming and expensive. ChIPmentation (Schmidl et al. 2015) emerged as an alternative to reduce both the number of cells and the protocol time. This method combines chromatin immunoprecipitation with sequencing library preparation by Tn5 transposase ('tagmentation'), which adds sequencing adapters in a single step, thus reducing both the time and cost of performing this protocol. Additionally, as few as 10,000 cells can be used to obtain reproducible histone modification profiles.

An alternative to using the ChIP technology is the Cleavage Under Targets and Release Using Nuclease (CUT&RUN) (Skene and Henikoff 2017), which is based on the Chromatin ImmunoCleavage (ChIC) strategy. This method allows protein mapping by successive binding of a specific antibody, followed by tethering a Protein A/Micrococcal Nuclease (pA-MNase) fusion protein in permeabilized cells without cross-linking. As with ChIP-seq, a modification of this technique has been developed in which the Tn5 transposase can add the sequencing adapters in situ, having the same benefits described above with ChIPmentation. This modification is called Cleavage Under Targets and Tagmentation (CUT&Tag) (Kaya-Okur et al. 2019).

1.4.4 Chromatin 3D interactions

The physical location of the regulatory elements in the cell nucleus is central to the regulation of gene expression. Thus, assessing chromatin three-dimensional contacts, allows identification – or at least prioritization – of the putative gene targets of distal enhancers and facilitates a deeper understanding of cell-specific gene regulatory programs.

Chromatin loops represent the structural units of gene expression control that allow gene promoters to physically interact with their putative enhancers. Many different techniques allow identifying chromatin interactions at various levels of resolution, influencing our ability to pinpoint the specific gene target of a distal regulatory element.

1.4.4.1 High-throughput assays for detecting chromatin interactions

Most assays for detecting chromatin interactions are based on C-technologies, in which the nucleus is cross-linked to preserve 3D chromatin conformation. This is followed by DNA fragmentation and re-ligation of loci that were spatially close at the time of cross-linking (Dekker et al. 2002). This process creates chimerical molecules that contain the two loci found to be in contact or in close proximity. Finally, the frequency of ligation determined by PCR or DNA sequencing can finally be used as a proxy of spatial proximity.

Many different techniques have been developed and coupled with DNA sequencing technologies to obtain high-throughput chromatin interactions with different levels of coverage and resolution. Here, I highlight some of the most used methods classified by the number of queried loci in the genome:

- **All loci vs all the genome.** Hi-C (Lieberman-Aiden et al. 2009) was the first technique that allowed retrieval of chromatin interactions genome-wide. While the major advantage of this technique is that of generating a comprehensive 3D interaction map of the genome, a limitation resides in the reduced resolution obtained from the individual contacts, which can be partially overcome by using frequent restriction enzyme cutters and high sequencing depth.
- **Many loci vs all the genome.** With these techniques, the query is narrowed down to a large number of genomic loci. In Promoter Capture Hi-C (Javierre et al. 2016), for example, the query is limited by sequence capture to a selection of gene promoters, resulting in a strong enrichment for promoter interactions, a reduction of the overall library complexity compared to Hi-C and an increased chromatin contact resolution. Another interesting example is Hi-ChIP (Mumbach et al. 2016), in which the loci containing the feature of interest, such as relevant histone modifications, are selected for sequencing by immunoprecipitation.

- **1 locus vs all the genome.** In Circular Chromosome Conformation Capture (4C) techniques (Simonis et al. 2009), a locus of interest – named viewpoint or bait – is queried for genome-wide interactions. For instance, the use of a specific enhancer as a viewpoint would result in the observation of a profile with 4C reads enrichment at genomic regions the enhancer is interacting with (**Figure 1.5**). These chromatin contact profiles can be obtained by using primers designed to capture the region of interest, which will select all the chimeric fragments containing the viewpoint sequence. The benefit of 4C techniques resides in the fact that it can achieve very high resolution (very few kb) with only a few million sequencing reads. Interestingly, a modification of this technique with unique molecular identifiers (UMIs), called UMI-4C (Schwartzman et al. 2016), was developed to reduce the PCR amplification bias, thus allowing a more accurate quantification of chromatin interactions.

1.5 Chromatin dynamics in response to external stimuli

1.5.1 Uncovering novel regulatory elements

Regulation of gene expression is key not only to maintain proper cell function, but also to allow terminally differentiated cells to respond to external stimuli. This permits cells to accommodate changes in their environment and to respond accordingly. The stimuli-induced remodeling of the regulatory landscape has been mostly studied in immune specialized cell types, as one of their key hallmarks is their ability to transition between resting and stimulated states upon the appropriate stimulus.

An initial view of gene regulatory mechanisms hypothesized that changes in gene expression were induced by means of a predetermined regulatory landscape enforced by TFs controlling cell identity. This would imply that an external stimulus would affect gene expression through a fixed repertoire of regulatory elements established after terminal differentiation, allowing the maintenance of cell identity in spite of a changing environment.

However, Ostuni et al. (2013) proposed that external stimuli modulating cell behavior might be associated with a partial reprogramming of the pre-established cis-regulatory elements. To test this hypothesis, they stimulated mouse bone-marrow-derived macrophages with lipopolysaccharide (LPS) – a canonical inflammatory agent – for 4 and 24 hours. Next, to characterize the repertoire of regulatory elements in unstimulated and stimulated cells, they performed ChIP-seq experiments of two histone marks – H3K4me1 and H3K27ac – and of the macrophage lineage determining TF Pu.1. These experiments revealed the activation of regulatory elements

that were **poised** in unstimulated conditions, and identified **latent** regulatory elements, which were activated by LPS, as they were unbound by TFs and lacked enrichment of H3K4me1 and H3K27ac in unstimulated cells. To explore whether these new regulatory elements were stimuli-specific, they next treated macrophages with different molecules, such as Toll-like receptor (TLR) agonists, tumor growth factor- β (TGF- β) or tumor necrosis factor- α (TNF α) and interleukin 1-beta (IL-1 β). Indeed, they observed diverse latent enhancers being activated with different stimuli, with an overlap reflecting the functional similarity of the used stimuli. This was the first observation that terminally differentiated cells were able to expand their predetermined regulatory element repertoire in response to external stimuli.

To deepen the understanding of the macrophage-specific regulatory landscape plasticity, Lavin et al. (2014) and Gosselin et al. (2014) explored how the cellular microenvironment could affect the repertoire of active regulatory elements. Their studies provide insights into the differences in the enhancer landscape of different tissue-residing macrophage populations, such as microglia or large peritoneal macrophages, which are more divergent than what can be explained by their developmental origin. Indeed, they observed that different populations showed different enhancer landscapes, although sharing a common developmental origin. Moreover, Lavin et al. (2014) showed that when transplanting bone marrow precursors or even differentiated macrophages to another tissue, they could be reprogrammed by their new microenvironment.

Recently, Calderon et al. (2019) explored the landscape of stimulation-responsive chromatin for up to 32 human immune cell populations by querying chromatin accessibility and mRNA expression under resting and stimulated conditions. They observed that the chromatin landscape features that appeared in response to stimulation were also dependent on the cell lineage.

These works demonstrate that regulatory maps of immune lineage cell types are stimuli-responsive. Nevertheless, less is known regarding the regulatory landscape plasticity of other disease-relevant terminally differentiated tissues, such as the pancreatic islets.

1.5.2 Dynamics of chromatin interactions

The uncovering of novel regulatory elements upon stimuli raises the question whether chromatin interactions may also be modified in response to the environment. A negative answer would suggest that enhancer-promoter contacts are already pre-formed, even when the regulatory elements are inactive, while a positive answer would indicate that differentiated cells retain the ability to form new enhancer-promoter interactions in response to their environment.

During development, different chromatin interactions are established to regulate gene expression patterns and give rise to the different cell types in the human body. Once cells are terminally differentiated, higher order structures like chromosome territories or TADs are usually shared among cell types, while smaller domains and chromatin loops are considered tissue-specific.

Chromatin organization has been shown to be disrupted in cancer, usually as a consequence of different processes, such as chromosome structural variation (Spielmann, Lupiáñez, and Mundlos 2018) or copy-number alterations (Taberlay et al. 2016). However, the observed changes are likely due to the high number of genetic and epigenetic alterations found in cancerous cell types.

To provide insights into the role of TADs during transient gene expression changes, such as those induced by hormonal exposure, Le Dily et al. (2014) analyzed changes in TAD structure after exposing breast cancer cells to progesterone. Interestingly, they observed that TAD structure was modified upon treatment, with smaller TADs enabling transient coordinated regulation of genes.

Nonetheless, the mentioned studies focus in large structures such as TADs, while changes in chromatin interactions are likely to happen in the chromatin loops, reflecting enhancer-promoter dynamics. Thus, more studies evaluating the dynamics of chromatin at a greater resolution may help understanding how novel regulatory elements driving the response to transient factors interact with their target genes.

2

Pancreatic islets and glucose homeostasis

The islets of Langerhans, or pancreatic islets, are endocrine micro-organs embedded in the exocrine pancreas. The pancreas of a healthy adult human hosts about 1 million islets, representing less than 2% of the pancreatic mass. Each micro-organ has a 50–500 μm diameter and is composed of 50 to 3,000 cells. The main function of pancreatic islets is the release of hormones, mainly involved in the regulation of glucose homeostasis.

2.1 Pancreatic islet composition

In humans, pancreatic islets are composed of different cell types exerting distinct endocrine functions (Röder et al. 2016):

- **Alpha (α) cells** (15-20%) which primarily secrete **glucagon**, a catabolic hormone that induces an increase in blood glucose levels.
- **Beta (β) cells** (65-80%) represent the most predominant pancreatic islet cell type. These cells release **insulin** and **amylin** at a 100:1 ratio. Insulin is an anabolic hormone that promotes absorption of glucose from the bloodstream into peripheral tissues. Amylin, or islet amyloid polypeptide (IAPP), is a hormone that plays a role in the regulation of glycemia by slowing gastric emptying and promoting satiety.
- **Delta (δ) cells** (3-10%) are in charge of the production and release of **somatostatin**, or growth hormone-inhibiting hormone (GHIH), which can inhibit insulin and glucagon secretion.
- **Epsilon (ϵ) cells** (<1%) produce **ghrelin**, a hormone that increases food intake, gastric motility and gastric acid secretion.
- **PP (Gamma, γ or F) cells** (3-5%) secrete **pancreatic polypeptide**, involved in self-regulating both endocrine and exocrine pancreatic secretion.

All these different cell types and others are intermingled in each pancreatic islet, forming trilaminar epithelial plates. Plates are surrounded by blood vessels that allow release of the secreted hormones into the bloodstream. In the epithelial plates, most β -cells are found in a central position and show cytoplasmic extensions between outlying non- β -cells (Bosco et al. 2010).

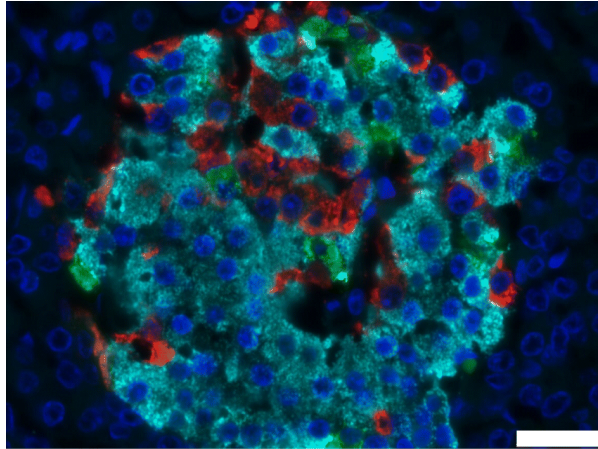


Figure 2.1: **Pancreatic islet immunostained cell types.** Immunofluorescent staining of insulin (light blue), glucagon (red), somatostatin (green) and cell nuclei (DAPI, dark blue). Scale bar, 25 μm . Extracted from Morgan and Richardson (2018).

This specific distribution favors both heterologous contacts between β and α cells and homologous contacts between β cells.

α and β cells together compose $\sim 90\%$ of the islet cell population. Because of their function in secreting blood-regulating hormones, pancreatic islets are key in maintaining glucose homeostasis.

2.2 Glucose homeostasis

Glucose is a key molecule central to cell metabolism and used in many complex and coordinated biochemical reactions. In humans, glucose is obtained from food ingestion and digestion, and is transported to all the cells and tissues through the bloodstream. To ensure proper function, glucose levels in blood – or glycemia – should be maintained within a very narrow range (3.3-7.8mM). The preservation of the physiological levels of glycemia is primarily obtained through the opposite functions of insulin and glucagon (**Figure 2.2**).

Pancreatic β cells respond to increased glycemia by releasing insulin, which will bind cell receptors in peripheral tissues, including adipose and skeletal muscle, to induce glucose uptake allowing to reduce the levels of glucose in the bloodstream. Decreases in the glucose levels will instead result in the release of glucagon from α cells, which will in turn target various organs, including liver and kidneys, to facilitate both glycogenolysis and gluconeogenesis and induce the release of glucose in the bloodstream.

Maintenance of euglycemia is vital, as hypoglycemia is a life-threatening event and chronic hyperglycemia can result in serious complications, including macrovascular disease and microangiopathy. Pancreatic islets are thus

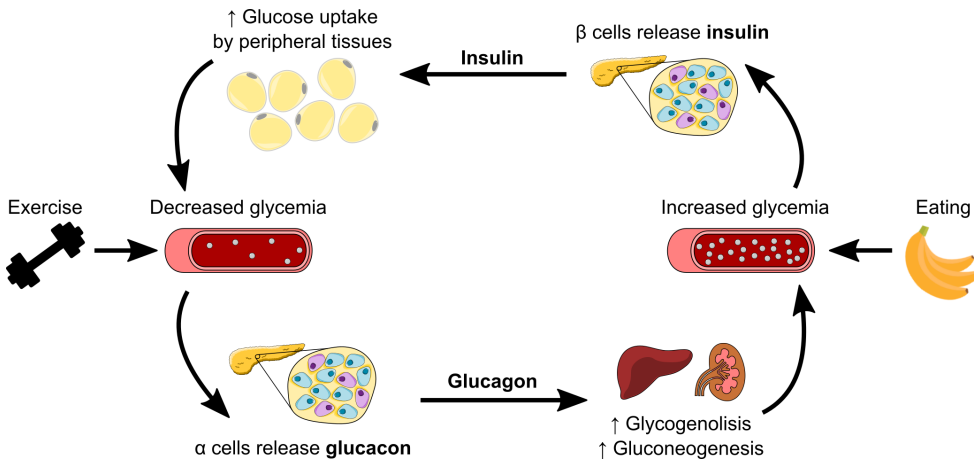


Figure 2.2: **Insulin and glucagon control glycemia.** After a meal, nutrients are absorbed throughout the digestive tract and converted into simple molecules, such as glucose, which are then released into the bloodstream, resulting in raise of blood glucose levels (hyperglycemia). When pancreatic β cells detect high levels of glucose in the bloodstream, they release insulin, which will bind its receptors in peripheral tissues, such as adipose tissue, and induce glucose uptake. After fasting or exercise, blood glucose levels will drop (hypoglycemia), thus inducing the release of glucagon by pancreatic α cells. Glucagon will target the liver and kidney, among other tissues, and induce the production and release of glucose by glycogenolysis and gluconeogenesis, thus restoring blood glucose levels to normal values.

fundamental in maintaining glucose homeostasis and failure of pancreatic cell function can result in diseases of the glucose metabolism such as diabetes.

2.3 Deciphering regulatory maps in pancreatic islets

With the advent of next-generation sequencing techniques, many different human tissues and cell populations have been analyzed to produce tissue-specific regulatory maps. Large consortia such as ENCODE and the Epigenome Roadmap allowed mapping human tissues including skeletal muscle, liver and adipose tissue, which are relevant to glucose metabolism diseases. Pancreatic islets and their cell types, however, are less accessible, and, for this reason, were not initially prioritized in these large consortia.

Human pancreatic endocrine tissue is difficult to access, being embedded inside the exocrine pancreas, and obtaining viable pancreatic islets cells requires a complex isolation procedure (Bucher et al. 2005; Melzi et al. 2010). Moreover, pancreas biopsies or resections, which can result in undesirable effects such as acute pancreatitis, are performed only under very special circumstances. For this reason, human pancreatic islets used for islet transplantation or research purposes are mainly obtained from deceased organ

donors. Heterogeneity of pancreatic islets represents an additional challenge to the study of the different islet cell population. Fluorescence-activated cell sorting (FACS) has been successful in separating major cell populations, such as β and α cells (Bramswig et al. 2013; Ackermann et al. 2016; Muraro et al. 2016; Arda et al. 2018), though specific markers for separating minority cell types are still needed.

Nevertheless, due to the central role of pancreatic islets in diseases such as diabetes, different laboratories embarked in deciphering the regulatory landscape of this tissue, in an ongoing effort to dissect the molecular mechanisms underlying glucose metabolism diseases. To this end, several groups focused on profiling the chromatin landscape of adult pancreatic islets (Bhandare et al. 2010; Gaulton et al. 2010; Stitzel et al. 2010; Parker et al. 2013; Pasquali et al. 2014) and pancreatic progenitors (Cebola et al. 2015).

One early study mapped the first atlas of regulatory elements in pancreatic islets (Gaulton et al. 2010). In this work, the authors used FAIRE-seq to identify accessible regions in human pancreatic islets and compared them to those of five non-islet cell lines. This allowed the characterization of islet-specific accessible sites. When studying the genomic distribution of these sites they identified islet-selective **clusters of open regulatory elements** (COREs), which were associated with islet-specific gene expression.

Subsequent studies used chromatin states identified by the ChromHMM algorithm (Parker et al. 2013) or the distribution of distal regulatory enhancers (Pasquali et al. 2014) to link **stretch enhancers** or **enhancer clusters** to tissue-specificity in pancreatic islets. By integrating this information with gene transcription data and islet-specific TF binding sites, Pasquali et al. (2014) were able to create integrative regulatory maps of human pancreatic islets and demonstrate that enhancer clusters are enriched of type 2 diabetes (T2D) risk variants, implicating their dysregulation to the susceptibility to develop T2D.

Recently, Miguel-Escalada et al. (2019) updated the linear conception of enhancer clusters by adding three-dimensional chromatin interactions, analyzed using promoter-capture Hi-C experiments. By characterizing enhancer-promoter interactions, they were able to define groups of tissue-specific regulatory elements that were not necessarily close in the linear space. They named these compact co-regulated regions **enhancer hubs**.

Nowadays, with the increasing availability of single cell methods to profile not only the transcriptome, but also the epigenome, we might be able to deepen our understanding of the regulatory circuitry acting through the different cell types that compose pancreatic islets.

3

Type 1 diabetes

Type 1 Diabetes (T1D), also known as autoimmune or juvenile diabetes, is an autoimmune disease in which insulin-producing pancreatic β cells are targeted and attacked by the immune system. This results in patients suffering a chronic insulin deficiency and a lifelong dependence on insulin injections.

T1D onset symptoms include increased urination (polyuria), thirst (polydipsia) and increased hunger (polyphagia), together with fatigue and weight loss. If untreated, T1D can lead to mild to severe complications, including ketoacidosis, which may present with rapid and deep breaths, drowsiness, thirst, abdominal pain and vomiting. T1D diagnosis often occurs in patients hospitalized due to life-threatening ketoacidosis episodes.

At the time of diagnosis, the autoimmune attack on β cells has likely been developing for many years. Prospective studies on pre-diabetic relatives of patients with T1D have shown that the process of β cell degeneration can take more than 3 years before the clinical manifestation of the disease (The DCCT Research Group 1998). Therefore, a clear asynchronicity between the trigger of the autoimmune attack and the clinical onset characterize the pathogenesis of T1D.

3.1 Epidemiology

T1D usually develops in children and young adults, although cases of adult-onset T1D have been reported (Buzzetti, Zampetti, and Maddaloni 2017). The age of onset is generally associated with the severity of the disease, where late cases show milder symptoms and a slower decrease of β -cell mass. This milder presentation in adults together with the increased incidence of other types of diabetes at that age makes the distinction between T1D and T2D more difficult.

Worldwide, T1D represents around 5-10% of all the cases of diabetes and shows variable incidence in different populations (**Figure 3.1**). In Spain, the annual incidence of T1D is 9.5-16/100,000 yearly cases in children under 14 years and 9.9/100,000 in young adults from 15 to 29 years of age (Lopez-Bastida et al. 2013). Over the past 60 years, the incidence of T1D has increased by 3–5 % per year (Bruno, Gruden, and Songini 2016). The Diabetes Mondiale (DiaMond) study, which monitored T1D incidence from 1990 to 1994, reported a 40/100,000 yearly children cases in Sardinia and

Finland, whose populations have the highest T1D rate (Karvonen et al. 2000). More recent studies (Harjutsalo, Sjöberg, and Tuomilehto 2008; Songini et al. 2017), have reported an incidence of 60 cases per 100,000 children in such regions, thus highlighting a considerable increase in the number of yearly diagnosed T1D cases. The increasing incidence of T1D is also associated with an onset peak occurring at a younger age; some studies, though, associate the onset peak shift to earlier diagnosis, primarily occurring in high-incidence countries (Patterson et al. 2009). Increase of T1D incidence has been associated with changes in lifestyle occurred during the industrial and economic revolution that took place during the latter half of the twentieth century (Ilonen, Lempainen, and Veijola 2019).

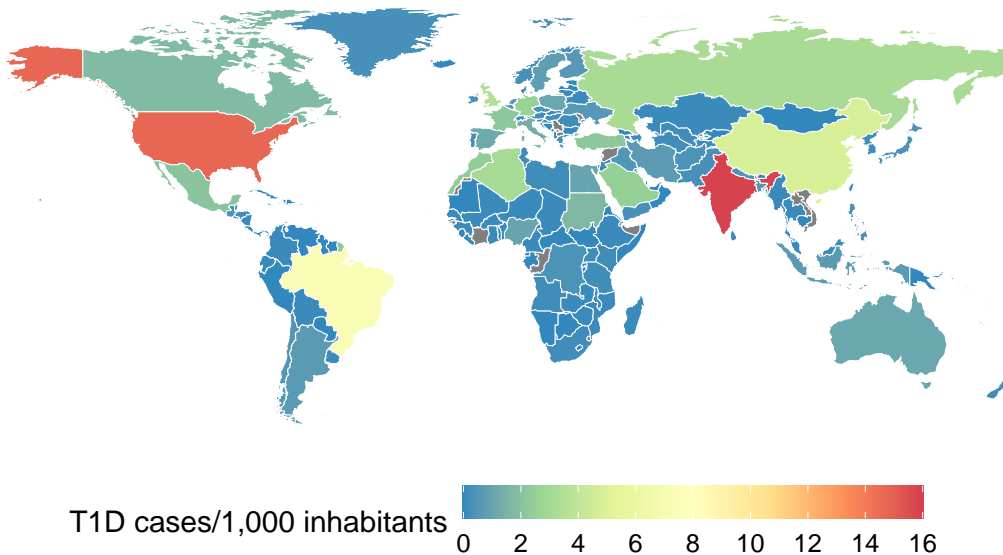


Figure 3.1: **Estimated worldwide incidence of Type 1 Diabetes (2019)**. Estimated new T1D cases per country in children (<15 years of age) per 1,000 individuals (2019). Data from the International Diabetes Federation (<http://www.diabetesatlas.org/data/en/>)

3.2 Aetiological factors

The aetiology of T1D is not completely understood, but it is likely to consist of a combination of genetic and environmental factors targeting immune and/or β cells, which end up triggering, permitting and aggravating a β -cell targeted autoimmune attack (**Figure 3.2**). Clarifying the precise aetiological mechanisms of T1D is challenging because of the heterogeneous nature of the

disease, exemplified by the presence of different islet-specific autoantibodies, the variability of age at disease onset and the complexity of the genetic factors.

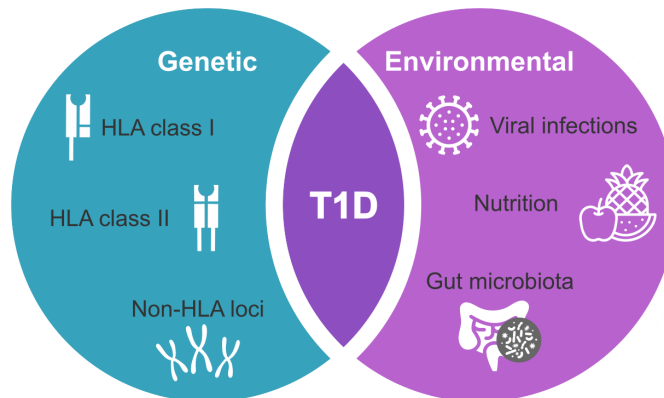


Figure 3.2: **Proposed Type 1 Diabetes aetiological factors.** T1D development may include an interaction between genetic susceptibility and environmental factors.

In the following section I will analyze the current knowledge on the implication of genetic susceptibility and environmental factors to the aetiology of T1D.

3.2.1 Genetic susceptibility

Genetics represents a primary risk factor for the development of β cell autoimmunity. Studies on T1D indicate that the disease risk for non-identical twins is similar to that observed for siblings, while this raises up to 50-70% in monozygotic twins, suggesting a key role for genetics in T1D aetiology (Redondo et al. 2008; Tuomilehto 2013).

Studying the genetic component in multifactorial and polygenic diseases is an extremely challenging task. Some insights may be obtained from studies exploring some specific candidate loci, such as the insulin locus. Additionally, Genome-Wide Association Studies (GWAS) take advantage of studying single-nucleotide polymorphisms (SNPs) present in the population using large cohorts of patients and healthy controls to associate different alleles to the risk of developing a disease.

Current insights into T1D genetics architecture allow distinguishing into HLA and non-HLA genetic susceptibility.

3.2.1.1 HLA locus

Human Leukocyte Antigen (HLA) is the name given to the **Major Histocompatibility Complex (MHC)** gene complex in humans. This largely polymorphic locus, located in a 3 Mb stretch in chromosome 6, codes for cell surface proteins

that are key to the adaptive immune system. MHC complexes are able to bind small peptides – the antigens – and are expressed in the cell surface, where they will be recognized by appropriate immune cell types.

MHC, and thus HLA, can be classified into three groups of proteins with distinct immune functions:

- **MHC class I** (HLA-A, HLA-B, HLA-C). HLA class I proteins are expressed by most cell types in the human organism. They are involved in presenting potentially dangerous internally digested peptides. These peptide-bound cell surface complexes are recognized by CD8⁺ cytotoxic T lymphocytes.
- **MHC class II** (HLA-DP, HLA-DM, HLA-DO, HLA-DQ and HLA-DR). HLA class II proteins, in turn, are expressed in the surface of professional antigen-presenting cells, such as thymic, dendritic and B cells. Thus, HLA-class II proteins bind processed peptides as well, but in this case they are obtained from exogenous phagocytosed proteins. This complex is involved in antigen presentation to CD4⁺ T cells, which will in turn launch an immune attack against foreign microorganisms.
- **MHC class III**. HLA class III locus encodes proteins with a role in the complement system, key in innate immunity. Other immunological important genes, such as tumor necrosis factors and lymphotoxin are also encoded in this region.

HLA genes currently account for 50% of the genetic risk of developing T1D (Pociot and Lernmark 2016). Thus, an individual's T1D risk is defined by the combination of the inherited parental HLA haplotypes. Nonetheless, different classes of HLA have different strengths of association with T1D development and are involved in different mechanisms of T1D progression (**Table 3.1**).

Among all HLA genes, HLA class II show the strongest association with T1D risk, especially **HLA-DR** and **HLA-DQ** genes. Different polymorphisms in the alleles of the above-mentioned genes affect the selection of epitopes that are presented to CD4⁺ T cells. **Molecular mimicry** is thought to be a key event to generate cross-reactive antigens, as two peptides – one from the organism and one from foreign microorganisms – could be recognized by a single antigen-specific receptor. Additionally, the involvement of HLA class II molecules in T1D has been associated with the development of islet autoimmunity, as the frequency of **seroconversion** – the development of islet-specific antibodies – is strongly correlated with the strength of the association between the presented HLA class II genotypes and T1D. Thus, different HLA class II haplotypes are mainly associated with the generation of β -cell targeting antibodies, specifically with the age of onset (Krischer et al. 2015; Ziegler et al. 2013) and the type of antibody that appears first (Törn et al. 2015). For instance, the appearance of the insulin autoantibody as the

first autoantibody is associated with the HLA-DR4-DQ8 haplotype, while the glutamic acid decarboxylase autoantibody is associated with the HLA-DR3-DQ2 haplotype (Ilonen, Lempainen, and Veijola 2019).

HLA class I genes are also associated with T1D risk, even though their contribution is considered to be less than HLA class II. In this case, the encoded proteins are key to the activation of CD8⁺ T cells, which represent the major cell population that infiltrates the pancreatic islets and mediates β cell death. Moreover, HLA class I protein expression is increased in islet cells during immune infiltration (Richardson et al. 2016). Thus, their role has been mostly associated with later stages of T1D, after the development of autoantibodies and autoimmunity.

3.2.1.2 Non-HLA loci

Even though the HLA locus has a major effect on T1D genetic risk, GWAS unmasked over 50 additional genomic loci that individually contribute with a modest effect to the total risk of developing T1D. However, the molecular mechanism of action of many of these individual loci is still unknown (Pociot and Lernmark 2016). In most cases, the prioritized target genes are associated with immune regulation or β cell function (Table 3.2). Interestingly, many genes are also associated with other autoimmune diseases, suggesting shared pathological mechanisms (Pociot and Lernmark 2016).

Among the non-HLA loci, the strongest association signals are observed at two loci: the *PTPN22* locus, which encodes the non-receptor protein tyrosine phosphatase type 22, a molecule involved in T cell and B cell responsiveness; and the *INS* locus, which codes for insulin.

INS polymorphisms are suggested to influence processes that are either involved in thymic immune tolerance and lead to the inadequate deletion of harmful β -cell antigen-reactive T cells or in the insufficient generation of T regulatory cells that are specific for β cell antigens. Indeed, some *INS* polymorphisms protect against T1D development by increasing insulin expression in thymic cells that present self-antigens to newly forming T cells (Pugliese et al. 1997).

Nonetheless, most target causal genes for the non-HLA loci are yet unknown. This is due to some inherent challenges from GWAS. First, variants are inherited in haplotype blocks and thus, may share strong p-values of association, limiting the identification of the causal variants solely based on p-values. Second, most (88%) of GWAS-associated variants fall in the non-coding part of the genome, suggesting that they may exert regulatory functions rather than affect the gene coding potential (Hindorff et al. 2009). Third, the regulatory landscape is dynamic, being cell-type- and state-specific and thus, identification of the causal regulatory variants requires knowledge

Table 3.1: **HLA gene polymorphisms associated with the risk of T1D.** Adapted from Ilonen, Lempainen, and Veijola (2019).

Effect	Proposed function	Haplotypes and alleles
HLA-DR and HLA-DQ		
Risk	Presentation of auto-antigens that induce autoimmunity	Far-eastern/Asian: DRB1*04:05-DQA1*03-DQB1*04:01; DRB1*09-DQA1*03-DQB1*03:03; DRB1*08-DQA1*03-DQB1*03:02. African: DRB1*07-DQA1*03-DQB1*02; DRB1*09-DQA1*03-DQB1*03:02. European: DRB1*04:01/2/4/5-DQA1*03-DQB1*0302; DRB1*03-DQA1*05-DQB1*02; DRB1*04:05-DQA1*03-DQB1*02
Protection	Incapacity to present auto-antigens to T helper cells and competitive binding with risk haplotypes	Far-eastern/Asian: DRB1*04:10-DQA1*03-DQB1*04:02. African: DRB1*03-DQA1*04:01-DQB1*04:02; DRB1*08-DQA1*04:01-DQB1*03:01. European: DRB1*15-DQA1*01-DQB1*06:02; DRB1*15-DQA1*01-DQB1*06:01; DRB1*14-DQA1*01-DQB1*05:03; DRB1*07-DQA1*02-DQB1*03:03; DRB1*04:03-DQA1*03-DQB1*03:02
HLA-DP		
Risk	Presentation of auto-antigens that induce autoimmunity	DPB1*03:01
Protection	Incapacity to present auto-antigens to T helper cells and competitive binding with risk haplotypes	DPB1*04:02
HLA class I		
Risk	Presentation of auto-antigens to cytotoxic CD8+ T cells	HL A- A*24; HL A- B*18; HL A-B*39:01; HL A- B*39:06

of the regulatory functions of the disease-implicated tissues and their state-specific cis-regulation. Moreover, linking non-coding variants to their gene target requires additional knowledge of tissue- and state-specific enhancer-promoter relationships. To overcome these challenges and improve our knowledge of T1D genetic mechanisms, different studies have performed

co-localization studies between T1D variants and regulatory elements from different cell populations, aiming to unravel the cell types through which the different variants may be acting, together with their potential target genes.

With the aim of understanding the role of T1D risk variants, Onengut-Gumuscu et al. (2015) fine-mapped T1D SNPs to active enhancers mapped in different cell populations, finding a significant enrichment in thymus, T and B cells, among others. However, no enrichment for T1D-associated variants was found in pancreatic islet enhancers, supporting the findings of other studies indicating that pancreatic islet enhancers are enriched for T2D variants, but not for T1D (Pasquali et al. 2014).

Similarly, Farh et al. (2015) fine-mapped disease-associated polymorphisms in order to find candidate causal variants for different autoimmune diseases. They observed an enrichment of T1D disease variants primarily in T-cell enhancers, but also reported a slight enrichment in pancreatic islet enhancers. Such observation opens the avenue for a possible mechanism of action of T1D risk variants through the pancreatic islet tissue, in addition to their action through immune cell types.

Table 3.2: **Selection of non-HLA T1D-associated risk variants.** Adapted from Ilonen, Lempainen, and Veijola (2019).

T1D SNPs	Candidate gene	Proposed function
rs689, VNTR	<i>INS</i>	Low expression in thymus, which affects central tolerance induction against proinsulin
rs2476601	<i>PTPN22</i>	Autoreactive T and B cells escaping negative selection in thymus
rs12722495	<i>IL2RA</i>	Response to IL-2 in regulatory T cells
rs3087243	<i>CTLA4</i>	Negative regulation of autoreactive T cells
rs1990760	<i>IFIH1</i>	Induction of type I interferon production
rs2292239	<i>ERBB3</i>	Regulation of β -cell apoptosis and antigen presenting function in dendritic cells
rs45450798	<i>PTPN2</i>	Regulation of β -cell apoptosis induction and regulation of T cell activation
rs3825932	<i>CTSH</i>	Regulation of β -cell function and protection from immune-mediated damage
rs3757247	<i>BACH2</i>	Regulation of β -cell apoptosis
rs11202303, rs80054410	<i>UBASH3A</i>	Regulation of CD4 ⁺ T cell activation

3.2.2 Environmental factors

The fact that T1D incidence has increased during the past decades points to a role of the environment in the pathology of the disease. This is also supported by the increasing number of T1D patients with low HLA-defined genetic risk. Currently, the main environmental factors that are associated with T1D aetiology are: viral infections, composition of gut microbiota and nutrition.

Viral infections are thought to accelerate T1D onset due to the generation of an increased insulin demand, known as the overload effect (Dahlquist 2006). Risk of T1D has been specifically associated with mumps and rubella viral outbreaks, although studies in the last years are focused on human enterovirus B infection. Enterovirus is associated with β -cell damage through different mechanisms. Some of them are direct, like cytolysis provoked by viral replication or functional deficiency of β cells. Other indirect mechanisms may involve an infection-associated inflammation which results in an increased β cell expression of HLA class I proteins and enhanced antigen presentation to CD8⁺ T cells. Viral infections may also increase molecular mimicry, as some antibodies produced as a response to an enterovirus infection were shown to recognize islet antigens (Härkönen et al. 2003).

Gut microbiota is important for the development of the immune system and has been suggested to have a role in autoimmunity. Different works have studied the proportion of different bacterial groups and have observed some differences between T1D patients and healthy controls. For instance, *Bacterioides* are more abundant in populations with high T1D incidence, while *Escherichia coli* is far more common in populations with lower T1D incidence (Vatanen et al. 2016).

Several **nutritional** factors have been associated with T1D. Obesity might be an important factor for developing T1D, especially in children with lower-risk HLA haplotypes (Corbin et al. 2018). Some studies have also associated increased consumption of cow's milk with T1D and unsaturated fish oils with some T1D protective effects (Niinistö et al. 2017). However, one key nutritional factor that has been the focus of most studies has been **vitamin D**. Vitamin D intake and subsequent 25-hydroxyvitamin D (25OHD) serum levels have been considered to play a role in the development of T1D, mainly because of their immune modulatory effects. However, results obtained from different population studies are contradictory, as some studies show decreased plasma levels of 25OHD in T1D patients compared to healthy controls (Littorin et al. 2006; Norris et al. 2018), while others report no differences (Simpson et al. 2011; Mäkinen et al. 2016). Thus, more studies are needed to clarify the role of vitamin D in the development of T1D.

3.3 Pathogenesis

T1D pathogenesis is highly complex and particularly challenging to study primarily due to two reasons: 1) Asynchronicity between onset and diagnosis makes it difficult to obtain samples and follow presymptomatic diabetes patients, and 2) Pancreatic islets are difficult to obtain and biopsies are at high risk of causing pancreatitis. Thus, most knowledge has been obtained from peripheral blood samples and cadaveric organ donors at variable disease stages. Animal models, such as NOD mice, have also provided useful insights, although limited considering that some immune processes associated with the development of diabetes differ between humans and the NOD mouse model

In global terms, T1D progression can be classified in three main stages (Pociot and Lernmark 2016; Katsarou et al. 2017):

- **Stage 1.** In this stage, β -cell autoimmunity, indicated by the presence of β -cell targeted autoantibodies, has already started and there is a decrease in β cell mass. However, glucose levels are maintained in the physiological range and there are no symptoms of the disease.
- **Stage 2.** The decrease in β cell mass is sufficient to impair glucose tolerance. However, whether the loss of glucose tolerance owing to impaired insulin secretion is entirely due to decreasing β -cell mass or also involves dysfunctional β -cells remains to be determined. Conversely, even though glucose levels may be altered, symptoms of the disease are not present yet.
- **Stage 3.** This stage corresponds to the clinical onset of the disease, as the β cell mass decrease aggravates hyperglycemia and the first T1D symptoms appear. At this stage, immunity is likely to have occurred for a prolonged period, as indicated by the presence of CD4⁺ and CD8⁺ T cells, dendritic cells, macrophages and B cells in and around the pancreatic islets in many newly diagnosed T1D patients (Gepts 1965; Krogvold et al. 2016).

According to this classification, stage 1 and 2 would correspond to pre-symptomatic T1D and stage 3 to symptomatic T1D (**Figure 3.3**).

The **trigger** able to start the autoimmune attack is commonly thought to be of environmental nature, such as a viral infection that would start an immune response, and lead to the production of the first autoantibody.

The **first autoantibodies** usually target insulin (INS) or 65 kDa glutamic acid decarboxylase (GAD65). The order of appearance of these two autoantibodies is strongly associated with the age of onset and some specific HLA haplotypes. Although not proven, autoantibody production might reflect continued presentation of β -cell autoantigens by dendritic cells, favoring the

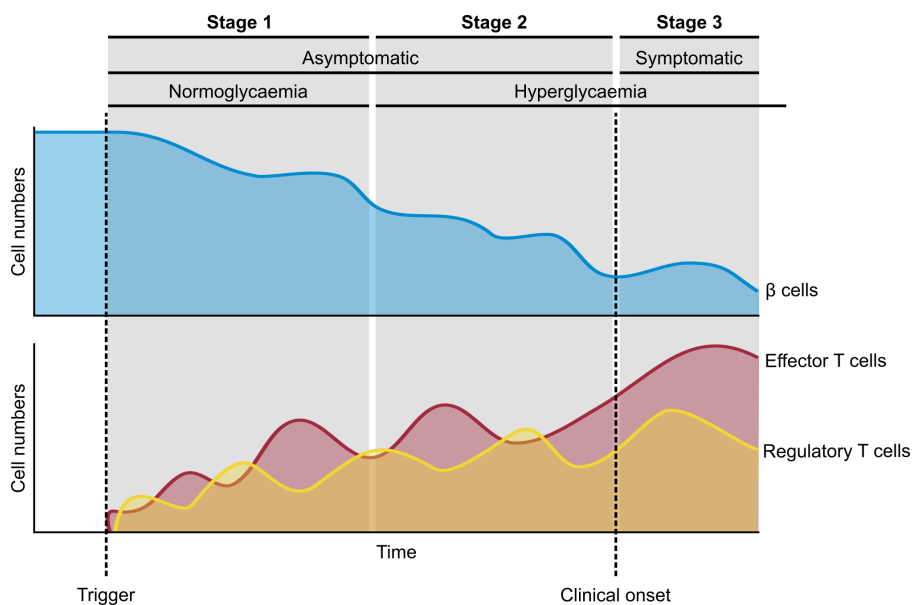


Figure 3.3: **Proposed staging for T1D and relative cell numbers.** Adapted from Herrath, Sanda, and Herold (2007), Pociot and Lernmark (2016) and Katsarou et al. (2017).

subsequent immunogenic response by $CD4^+$ and $CD8^+$ T cells.

Years after the first autoantibody appearance, tolerance to β cell autoantigens is finally breached. This loss of tolerance is marked by the appearance of **additional autoantibodies**, now targeting specific β cell proteins such as islet antigen-2 (IA-2) or the ZnT8 transporter. Increasing numbers of additional autoantibodies translates into an increased risk of rapid progression to T1D clinical onset (Pociot and Lernmark 2016). At this point, T cells recognize post-translationally modified β -cell peptides, likely induced by the presence of endoplasmic reticulum stress in β cells. Such stress-induced protein modifications may then cause the breach of β -cell tolerance (Van Lummel et al. 2014; Delong et al. 2016).

As with any other immune responses, regulatory mechanisms in the form of negative feedback circuits are in place to ensure stability of the system and minimize damage to the host. Due to the fact that these processes are cyclic in nature, Herrath, Sanda, and Herold (2007) proposed that T1D development and progression is also cyclical. According to this model, during insulinitis, the population of effector T cells in the islets increases, inducing β cell death. To compensate such process, the numbers of regulatory T cells increase to be able to modulate and repress the immune response, preventing further damage. In response to the immune attack and the consequent β cell mass reduction, β cells attempt to restore their functional mass by inducing

their proliferation and/or hypertrophy. However, this leads to the release of additional β cell antigens that in turn increase effector T cell population, which will renew the autoimmune attack to the β cell population (**Figure 3.3**). This cyclic mechanism of T1D progression would also explain the chronicity of the disease, as an enhanced exposure over time to an increasing number of β cell antigens leads to an exacerbation of the autoimmune attack, with the disease aggravating with increasing number of β -cell antigens.

As most information from T1D is obtained from blood samples, study of T1D pathogenesis has been focused on the presence of autoantibodies as indicators of β -cell directed autoimmunity (Campbell-Thompson 2015). However, even though autoantibodies represent precious biomarkers to study the progression of T1D, they are not necessarily pathogenic, as shown by the lack of correlation between autoantibody detection and immune infiltration in pancreatic islets from organ donors (Wiberg et al. 2015; Babon et al. 2016). Indeed, T1D is considered a cell-mediated disease, with the actual triggers of β cell death being immune CD8⁺ cytotoxic T cells. Thus, to obtain clearer insights into T1D pathology, the focus of future studies should be posed on specific events happening in pancreatic islets during immune infiltration.

Finally, the proinflammatory environment in the islet and other environmental factors that induce excessive β cell stress, might accelerate their deterioration and death (Dahlquist 2006). Indeed, stress signals such as proinflammatory cytokines, increase the accumulation of misfolded proteins with antigen capabilities in the endoplasmic reticulum. Such proteins and peptides will be released to the extracellular environment and may exacerbate the autoimmune attack.

3.4 The role of islet inflammation

Insulinitis represents a key hallmark of T1D and is the result of pancreatic islet immune infiltration. This immune infiltration is a heterogeneous process in terms of the number of islets affected and the pancreatic distribution of the preserved β cell mass.

Insulinitis can arbitrarily be classified into three different stages (Eizirik, Colli, and Ortis 2009). Of note, each stage can progress to the next stage or be resolved by endogenous factors and interrupt the progression to T1D.

1. **Induction.** Activation of both endogenous and exogenous ligands of pattern-recognition receptors (PRRs) can induce islet inflammation and death of pancreatic β cells.
2. **Amplification.** Cross-talk between immune and β cells may lead to amplification of the immune attack by transitioning from innate to adaptive

immune systems. This cross-talk is mediated by local production of chemokines, cytokines and danger signals from dying β cells.

- Maintenance or resolution.** After transitioning to the adaptive immune response, inflammatory mediators, such as cytokines, may contribute to prolonged functional suppression and death of β cells, modulation of β -cell regeneration and insulin resistance.

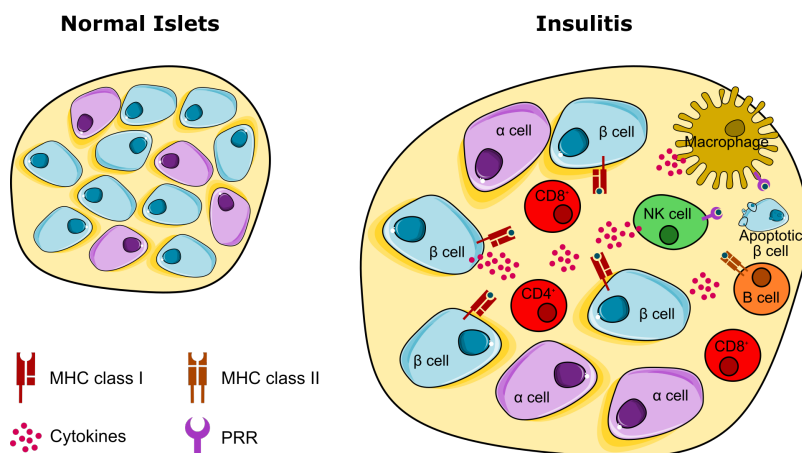


Figure 3.4: **Representation of insulinitis in pancreatic islets.**

The first immune response is obtained by the activation of components of the innate immunity. In the context of T1D, a specific type of PRRs named toll-like receptors (TLRs) are induced in β cells in the presence of double-stranded RNA (dsRNA) molecules, and can, in turn, induce transcription and translation of type I interferons (IFNs). If IFNs production is prolonged in time, it can lead to β cell apoptosis, which may be at least in part mediated by ER stress. Additionally, IFNs will increase the release of proinflammatory cytokines, which will induce overexpression of HLA class-I in β cells and will thus attract additional immune cells from both innate and adaptive immune systems.

The immune attack can be further amplified by transitioning to the adaptive immune system. This amplification involves recruiting of T cells to the pancreatic islets. Simultaneously, a cross-talk between β cells and infiltrating immune cells is established through the release of cytokines and chemokines. β cells will also show signs of ER stress, due to the proinflammatory environment. This stress can in turn be translated into an induction of antigen presentation through HLA-class I molecules. Finally, β cells undergoing apoptosis due to proinflammatory signals can also become 'danger signals' and induce further activation of the immune system.

Later stages of insulinitis involve the maintenance or resolution of the autoimmune attack. This stage is likely to be dominated by the adaptive immune system. Thus, individuals that develop mild insulinitis or that do not

have a high-risk genetic background, might resolve insulinitis and regain normal β cell function. However, patients with specific T1D-favoring conditions might see their β cell mass considerably reduced, thus leading to T1D progression and eventual disease clinical onset.

In summary, insulinitis is a very complex process that involves many different cell types, proteins and molecules. To deepen our understanding of this process, *in vitro* models were developed using different cytokine cocktails to reproduce some key insulinitis hallmarks in β cells. For the studies described in this thesis, we modeled different stages of insulinitis using two combinations of proinflammatory cytokines: IFN- α to mimic early insulinitis and IL-1 β + IFN- γ to model late insulinitis.

3.4.1 Interferon- α

Interferon- α (IFN- α), secreted mainly by macrophages and other cells from the innate immune system, belongs to a group of proteins called type I interferons, which are involved in stimulating an immune response after a viral infection.

The role of IFN- α in T1D is mostly associated with early insulinitis in T1D patients. This association was first suggested by the presence of IFN- α in pancreata of T1D patients with recent disease onset (Foulis, Farquharson, and Meager 1987). Moreover, INF- α serum levels were also elevated in children and adults around T1D onset (Chehadeh et al. 2000). In patients from the Diabetes Virus Detection (DiViD) cohort, composed of pancreatic biopsies from early diagnosed T1D patients, expression of IFN-induced genes was increased (Lundberg et al. 2016). Interestingly, even before disease onset or detection of autoantibodies, children with high genetic risk of T1D present a type I IFN transcriptional signature in blood (Kallionpaa et al. 2014).

IFN- α is used to treat infections caused by hepatitis C and some specific types of leukemia and melanoma. Of note, a break in β cell tolerance leading to T1D onset has been observed in the context of such therapeutic approach (Fabris et al. 1992; Schreuder et al. 2007; Sossau, Kofler, and Eigentler 2017).

Conversely, in patients with mutations in the thymus Autoimmune Regulator transcription factor (AIRE), which leads to the development of a rare autoimmune syndrome called autoimmune polyendocrinopathy syndrome type 1 (APS-1), the presence of self-reactive antibodies targeting IFN- α was associated with protection against T1D, which in some cases is an associated condition (Meyer et al. 2016). These observations support the hypothesis that IFN- α may play a central role in T1D pathogenesis.

IFN- α signals through the IFN- α receptor (IFNAR1) and the downstream tyrosine kinases JAK1 and TYK. Interestingly, *TYK2* gene overlaps a T1D GWAS risk locus. Human islets and β cells with *TYK2* knockdown show

decreased inflammatory gene expression, reduced HLA class I overexpression, and decreased apoptosis in response to dsRNA (Marroqui et al. 2015).

3.4.2 Interferon- γ and Interleukin-1 β

Interferon- γ (IFN- γ) is the sole member of the type II interferon group and is mainly secreted by CD8⁺ T lymphocytes. The main function of IFN- γ is the activation of macrophages and the induction of HLA class II molecule expression.

Interleukin-1 β (IL-1 β), on the other hand, is a proinflammatory cytokine that is produced by activated macrophages. IL-1 β is a key regulator of the inflammatory response and is involved in cell proliferation, differentiation and apoptosis, among other processes.

The combination of IFN- γ and IL-1 β , in contrast to IFN- α , is associated with later stages of insulinitis. These proinflammatory cytokines, in the context of late insulinitis stages, induce β cell release of additional chemokines, such as CXCL10 and CCL2. These molecules, in turn, attract more immune cells, thus increasing the local production of late proinflammatory cytokines (Eizirik, Colli, and Ortis 2009).

In patients with T1D, islets showed expression of the chemokine CXCL10, while the invading T-cells expressed the corresponding chemokine receptor CXCR3 (Roep et al. 2010). Moreover, the combination of IFN- γ + IL-1 β has already been shown to reproduce the cytokine-exposure signature observed in islets from T1D donors (Eizirik et al. 2012).

3.5 Current treatments and perspectives

Nowadays, T1D can be clinically managed through the exogenous administration of insulin and lifestyle modifications, such as low-carbohydrate diets and exercise. Nonetheless, an effective treatment that can cure T1D is not currently available.

Many different approaches are being studied to find a successful treatment for T1D. However, considering the large heterogeneity of T1D, which underlies different mechanisms and/or aetiological factors, finding an effective treatment is a challenging task.

The first course of action was developing drugs that aimed to suppress or modulate the autoimmune attack, in order to preserve β cell mass. Indeed, many clinical studies involving a large spectrum of immunosuppressing agents have been developed for many years, although with limited success (Skyler 2011).

Alternatively, some attempts at restoring glucose regulation through pancreas or islet transplantation have been successful, with 70% of transplanted T1D patients achieving insulin independence (Matsumoto 2010). However, donor shortage and the important side effects of immunosuppressant therapy, mandatory after the transplant, are important setbacks for using this approach.

After many years of isolated research, scientists worldwide started to join efforts and created large consortia to promote and develop clinical trials and mechanistic studies for T1D. One key example of this is the Type 1 Diabetes TrialNet, an international consortium of researchers aimed at developing studies focused on prevention and intervention clinical trials in T1D. Simultaneously, TrialNet is also gathering a large amount of longitudinal samples from T1D patients, accounting for more than 70,000 samples from over 7500 donors (Battaglia et al. 2017). Results from different studies developed through TrialNet are poised to provide key insights into the pathology of T1D, and show that collaborative research can enhance the discovery of new disease mechanisms.

In summary, it is key to deepen our understanding of T1D aetiology and pathology to understand the precise mechanisms of disease progression. This knowledge will enable identification of additional biomarkers of disease progression and of new potential therapeutic targets.

Part II

Hypothesis and Objectives

4

Hypothesis and Objectives

The regulatory landscape of adult cells changes to accommodate responses to their environment. We thus considered that studying the role of β cells in the development of T1D, by charting their regulatory landscape upon exposure to disease-relevant stimuli, could improve our understanding of the disease pathogenesis. The main **hypothesis** of this thesis is that β cells have an active role in the development of T1D, mediated by changes in their cis-regulatory networks in response to T1D-relevant stimuli such as proinflammatory cytokines.

By mapping cytokine-responsive regulatory circuitries, we will be able to identify key genes that could serve as potential T1D biomarkers or therapeutic targets. Additionally, by exploring the association between GWAS and β -cell cytokine-responsive regulatory elements, we will be able to determine the mechanism by which β cells are implicated in the risk of developing T1D.

More specifically, we define the following **objectives**:

1. Evaluate if exposure to proinflammatory cytokines induces changes in the chromatin regulatory landscape of pancreatic β cells.
2. Explore the cytokine-responsive β -cell regulatory circuitries and their gene targets to find potential biomarkers and/or therapeutic targets for T1D.
3. Assess if part of the T1D genetic risk can be explained by β -cell cis-regulatory networks that respond to the proinflammatory environment.
4. Make the generated regulatory data available to the scientific community in an easy and accessible way.

Part III

Results

5

Preamble

The results obtained by addressing the objectives of the present thesis have been organized and published in three different scientific papers (**Table 5.1**). The “Results” section contains the copies of said publications, together with a graphical abstract and the highlights of the most relevant results obtained in these studies, in order to guide the reader to the main contributions of each publication.

Results have been assembled into three different publications, the first two addressing the role of β cells in the development of T1D and the third one describing the development of an original tool to facilitate exploration and visualization of pancreatic islet genomic data.

1. *An integrated multi-omics approach identifies the landscape of IFN- α -mediated responses of human pancreatic β cells* (Colli et al. 2020). In this publication, we modeled **early insulinitis** by exposing human β cells to IFN- α at different time points (2, 8 and 24 hours). We assayed and integrated multi-omic data to pinpoint the key players in inducing the β -cell interferon signature. Finally, we selected some potential drug interventions to reverse this signature in human islets and β cells.
2. *The impact of proinflammatory cytokines on the β -cell regulatory landscape provides insights into the genetics of type 1 diabetes* (Ramos-Rodríguez et al. 2019). Similarly to the above, in this article we published multi-omics data obtained by exposing for 48 hours human pancreatic islets and human β cells to IFN- γ and IL-1 β . By integrating the generated data, we were able to dissect β -cell regulatory networks, possibly responding to a **late insulinitis** stage. Moreover, we observed an enrichment of T1D-associated SNPs in regulatory elements that were induced by the cytokine exposure. We validated the functional role for two of these variants, by showing that the T1D risk alleles alter the cytokine-responsive enhancer activity in β cells.
3. *The Pancreatic Islet Regulome Browser* (Mularoni, Ramos-Rodríguez, and Pasquali 2017). This last publication describes the implementation of the Islet Regulome Browser, an interactive online application that provides access and an integrated visualization to many genomic datasets relevant to the pancreatic islet community. The Islet Regulome

Browser was developed with the aim of facilitating access and exploration of complex genomic data to scientists with no bioinformatic background.

Table 5.1: Articles generated as a result of this thesis and included in the Results section.

Publication	Ref	Description	Obj.
An integrated multi-omics approach identifies the landscape of IFN- α -mediated responses of human pancreatic β cells	Colli et al. (2020)	IFN- α to model human β responses to early insulinitis	1, 2
The impact of proinflammatory cytokines on the β -cell regulatory landscape provides insights into the genetics of type 1 diabetes	Ramos-Rodríguez et al. (2019)	IFN- γ + IL-1 β to model human β cell responses to late insulinitis	1, 3
The Pancreatic Islet Regulome Browser	Mularoni, Ramos-Rodríguez, and Pasquali (2017)	Interactive exploration of pancreatic islet genomic data	4

Taking into account the Bioinformatic nature of the results obtained in this thesis, I considered it important to make all the code I produced for the analyses performed in Colli et al. (2020) and Ramos-Rodríguez et al. (2019) publicly available. Thus, all the code and resulting figures are gathered into a publicly accessible website at: mireia-bioinfo.github.io/phdthesis_code. A summary of the identifiers and accessors of the publicly available datasets released with the publications described above can also be found in the website.

Impact factor report from the thesis director

Dr. Lorenzo Pasquali, as director of the doctoral thesis titled “Beta-cells cis-regulatory networks and type 1 diabetes” presented by the PhD candidate Mireia Ramos Rodríguez, informs that results of the research project developed by the PhD candidate were accepted for publication in different relevant international journals. He also certifies that the works included in the present thesis will not be used by any of the co-authors in any other doctoral thesis.

The publication reference, together with the journal impact factor (according to *Journal Citation Report* database) and the specific contributions of the candidate are listed hereunder.

1. Colli, M.L. **Ramos-Rodríguez, M.**, Nakayasu, E.S. et al. An integrated multi-omics approach identifies the landscape of interferon- α -mediated responses of human pancreatic beta cells. *Nat Commun.* **11**, 2584 (2020). <https://doi.org/10.1038/s41467-020-16327-0>
 - **Journal Impact Factor:** Nature Communications, **11.878** (2018).
 - **Candidate Contributions:** Contributed to the project design. Performed analyses related to the chromatin structure as well as its integration with RNA-seq and proteomics. She actively participated in the discussion of the results obtained, including those obtained from the molecular biology experiments. She critically helped in addressing the comments and the discussion raised during the peer review process.
2. **Ramos-Rodríguez, M.**, Raurell-Vila, H., Colli, M.L. et al. The impact of proinflammatory cytokines on the β -cell regulatory landscape provides insights into the genetics of type 1 diabetes. *Nat Genet.* **51**, 1588–1595 (2019). <https://doi.org/10.1038/s41588-019-0524-6>
 - **Journal Impact Factor:** Nature Genetics, **25.455** (2018).
 - **Candidate Contributions:** Conception, coordination and overall lead of the project. She participated to the experimental design. She led and performed most of the bioinformatic analyses. She participated in writing the manuscript and in addressing the comments and the discussion raised during the peer review process.
3. Mularoni L., **Ramos-Rodríguez M.** and Pasquali L. The Pancreatic Islet Regulome Browser. *Front. Genet.* **8** (2017). <https://doi.org/10.3389/fgene.2017.00013>
 - **Journal Impact Factor:** Frontiers in Genetics, **3.517** (2018).
 - **Candidate Contributions:** Implementation of the Islet Regulome

Plot. She participated in writing the manuscript and in addressing the comments and the discussion raised during the peer review process.

Lorenzo Pasquali MD, PhD

Thesis Director

A handwritten signature in black ink that reads "Lorenzo Pasquali". The signature is written in a cursive style with a large initial 'L' and 'P'.

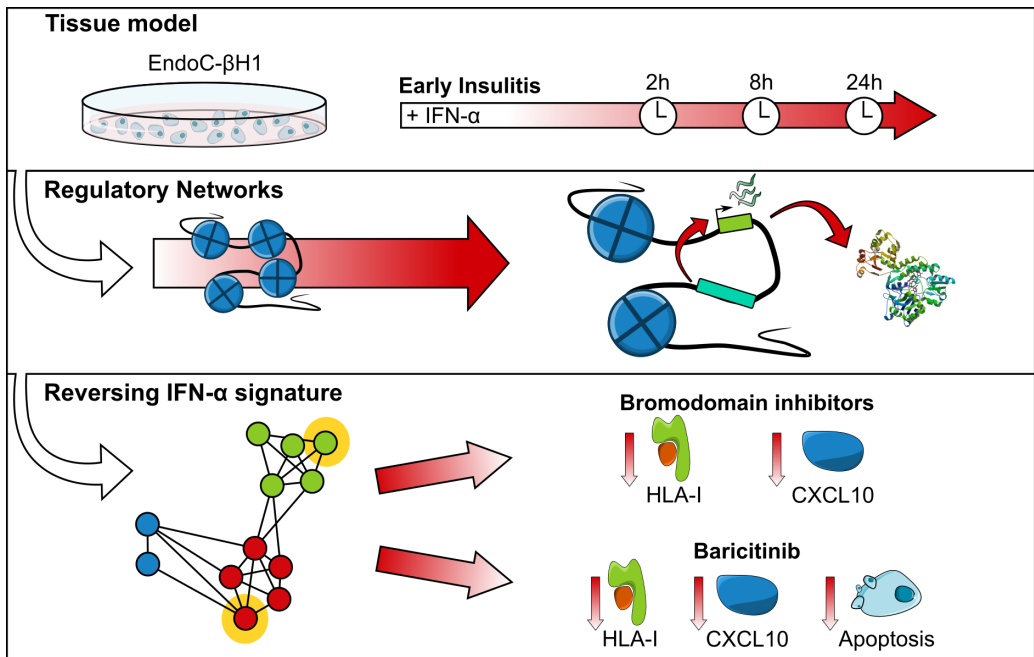
6

Interferon- α to model human β -cell responses to early insulinitis

In this study, we evaluated the effect of IFN- α on human β cells using a multi-omics approach, combining information obtained from chromatin accessibility, gene expression and protein abundance assays. IFN- α is a key cytokine that is known to play an important role in early insulinitis by inducing *HLA-I* expression, ER stress and β cell apoptosis (Marroqui et al. 2017).

Colli, M.L., **Ramos-Rodríguez, M.**, Nakayasu, E.S. et al. An integrated multi-omics approach identifies the landscape of interferon- α -mediated responses of human pancreatic beta cells. *Nat Commun.* **11**, 2584 (2020). <https://doi.org/10.1038/s41467-020-16327-0>










6.1 Graphical abstract



6.2 Highlights

- Exposure to IFN- α resulted in global changes in chromatin accessibility, gene expression and protein abundance. Importantly, all these changes were inter-correlated at the different time points.
- Reconstruction of the dynamic regulatory networks uncovered *IRF1*, *STAT1* and *STAT2* as the main TFs driving the response to IFN- α . When these genes were inhibited, we observed a down-regulation of checkpoint genes such as *PDL1* and *HLA-E*, suggesting that β cells play an active role during insulinitis by expressing these immune co-inhibitory molecules.
- An expansion of the β cell mRNA repertoire, mediated by the use of alternative splicing and alternative transcription start sites, was observed in β cells exposed to IFN- α . This mechanism might increase the β cell antigen repertoire in T1D.
- Mining gene networks induced by IFN- α uncovered two potentially interesting T1D therapeutic interventions:
 - **Bromodomain inhibitors**, which reverted the effect of IFN- α on β cells by decreasing in IFN- α -induced *HLA-I* and *CXCL10* expression, without altering IL-1 β + IFN- α induced apoptosis.
 - **Baricitinib**, an inhibitor of the kinase JAK1/2, showed a more clear reversion of the interferon signature by inducing a decrease in the expression of *HLA-I*, *CXCL10* and *CHOP*, together with a decrease in the cell surface expression of MHC class I and protection from IFN- α + IL-1 β induced apoptosis.

An integrated multi-omics approach identifies the landscape of interferon- α -mediated responses of human pancreatic beta cells

Maïkel L. Colli ^{1✉}, Mireia Ramos-Rodríguez^{2,3}, Ernesto S. Nakayasu ⁴, Maria I. Alvelos¹, Miguel Lopes¹, Jessica L. E. Hill⁵, Jean-Valery Turatsinze ¹, Alexandra Coomans de Brachène¹, Mark A. Russell⁵, Helena Raurell-Vila^{2,3}, Angela Castela¹, Jonàs Juan-Mateu¹, Bobbie-Jo M. Webb-Robertson⁴, Lars Krogvold⁶, Knut Dahl-Jorgensen⁶, Lorella Marselli ⁷, Piero Marchetti⁷, Sarah J. Richardson ⁵, Noel G. Morgan ⁵, Thomas O. Metz ⁴, Lorenzo Pasquali ^{2,3,8} & Décio L. Eizirik ^{1,9,10}

Interferon- α (IFN α), a type I interferon, is expressed in the islets of type 1 diabetic individuals, and its expression and signaling are regulated by T1D genetic risk variants and viral infections associated with T1D. We presently characterize human beta cell responses to IFN α by combining ATAC-seq, RNA-seq and proteomics assays. The initial response to IFN α is characterized by chromatin remodeling, followed by changes in transcriptional and translational regulation. IFN α induces changes in alternative splicing (AS) and first exon usage, increasing the diversity of transcripts expressed by the beta cells. This, combined with changes observed on protein modification/degradation, ER stress and MHC class I, may expand antigens presented by beta cells to the immune system. Beta cells also up-regulate the checkpoint proteins PDL1 and HLA-E that may exert a protective role against the auto-immune assault. Data mining of the present multi-omics analysis identifies two compound classes that antagonize IFN α effects on human beta cells.

¹ULB Center for Diabetes Research, Medical Faculty, Université Libre de Bruxelles, Brussels 1070, Belgium. ²Endocrine Regulatory Genomics, Department of Experimental & Health Sciences, University Pompeu Fabra, 08003 Barcelona, Spain. ³Endocrine Regulatory Genomics Laboratory, Germans Trias i Pujol University Hospital and Research Institute, Badalona, Spain. ⁴Biological Sciences Division, Pacific Northwest National Laboratory, Richland, WA 99352, USA. ⁵Institute of Biomedical & Clinical Science, University of Exeter Medical School, Exeter EX2 5DW, UK. ⁶Division of Pediatric and Adolescent Medicine, Faculty of Medicine, Oslo University Hospital, Oslo, Norway. ⁷Department of Clinical and Experimental Medicine, Islet Cell Laboratory, University of Pisa, 56126 Pisa, Italy. ⁸Josep Carreras Leukaemia Research Institute (IJC), Badalona, Barcelona, Catalonia, Spain. ⁹WELBIO, Université Libre de Bruxelles, Brussels, Belgium. ¹⁰Indiana Biosciences Research Institute, Indianapolis, IN, USA. ✉email: mcolli@ulb.ac.be

Type 1 diabetes (T1D) is a chronic autoimmune disease leading to pancreatic islet inflammation (insulinitis) and progressive beta cell loss¹. Type I interferons (IFN-I), a class of cytokines involved in antiviral immune responses², are involved in insulinitis. Viral infections are a risk factor associated with T1D development³ and individuals at risk of T1D show a type I interferon signature⁴. The type I interferon, interferon- α (IFN α), is expressed in islets of T1D patients⁵, and antibodies neutralizing different isoforms of IFN α prevent T1D development in individuals with polyglandular autoimmune syndrome type 1⁶. Exposure of human pancreatic beta cells to IFN α recapitulates three key findings observed in human insulinitis, namely HLA class I overexpression, endoplasmic reticulum (ER) stress and beta cell apoptosis⁷.

Combination of genome-wide association studies (GWAS)⁸ and studies using the ImmunoChip⁹ have identified around 60 loci associated with the risk of developing T1D. Transcriptomic studies revealed that >70% of the T1D risk genes are expressed in human pancreatic beta cells¹⁰, and many of these genes regulate innate immunity and type I IFN signaling¹¹.

Type I IFN signaling is often cell-specific, an effect mediated by differences in cell surface receptor expression, and activation of downstream kinases and transcription factors¹². Thus, and considering the potential relevance of this cytokine to the pathogenesis of T1D, it is crucial to characterize its effects on human beta cells. To define the global impact of IFN α on human beta cells, we presently performed an integrative multi-omics analysis (ATAC-seq, RNA-seq and proteomics) of IFN α -treated human beta cells to determine the early, intermediate and late responses to the cytokine. The findings obtained indicate that IFN α promotes early changes in chromatin accessibility, activating distant regulatory elements (RE) that control gene expression and protein abundance. IFN α activates key transcription factors (TFs), including IRF1, which act as a mediator of the crosstalk between beta cells and immune cells via the expression of the checkpoint proteins PDL1 and HLA-E. Furthermore, IFN α induces modules of co-expressed mRNA and proteins that physically interact and have relevance to T1D pathogenesis. The integration of high-coverage RNA-seq and ATAC-seq indicates regulatory gene networks and reveals that alternative splicing and different first exon usage are key mechanisms expanding the repertoire of mRNAs and proteins expressed by stressed beta cells. Finally, mining the modules of co-expressed genes and the IFN α beta cell signature against the most recent catalogs of experimental and clinical drugs identifies two potentially interesting therapeutic targets for future trials.

Results

IFN α modifies beta cell mRNA expression similarly to T1D.

We performed a time course multi-omics experiment combining ATAC-seq, RNA-seq and proteomics of the human beta cell line EndoC- β H1 exposed or not to IFN α . The data were integrated to determine the dynamics of chromatin accessibility, gene/transcript expression and protein translation, respectively (Fig. 1a). We also performed RNA-seq of 6 independent human pancreatic islet preparations exposed or not to the cytokine at similar time points (Supplementary Fig. 1a). To assess whether our in vitro model is relevant for the in vivo islet inflammation (insulinitis) in T1D, we took two approaches: (1) Examine whether candidate genes for T1D expressed in human islets are involved in IFN signaling (Supplementary Fig. 2a); and (2) Compare our in vitro data of IFN α -treated EndoC- β H1 cells and human islets with available RNA-seq data of human beta cells from T1D patients. In line with previous findings suggesting a role for IFNs on the pathogenesis of T1D¹³, we found that T1D risk genes expressed

in human islets^{10,14} are significantly enriched in immune-related pathways, including type I and II interferon regulation/signaling (Supplementary Fig. 2b). Next, we performed a Rank-Rank Hypergeometric Overlap (RRHO) analysis (which estimates the similarities between two ranked lists¹⁵) comparing the log₂ fold-change (FC) ranked list from RNA-seqs of EndoC- β H1 cells and human islets (IFN α -treated vs untreated) against an equally ranked list of genes obtained from RNA-seq of purified primary beta cells¹⁶ from T1D and healthy individuals (Supplementary Fig. 2c and Supplementary Data 1). There was a significant intersection between upregulated genes induced by IFN α in both, EndoC- β H1 cells (362 overlapping genes) and human islets (850 overlapping genes), and genes induced by the local pro-inflammatory environment affecting primary beta cells from T1D individuals (Supplementary Fig. 2d, f). We also compared these two IFN α -treated datasets against beta cells from T2D patients¹⁷, a condition mostly characterized by metabolic stress¹⁸. By contrast with the observations made in beta cells from T1D individuals, there was no statistically significant correlation between IFN α -regulated genes in EndoC- β H1 cells and human islets and the gene expression profile present in T2D beta cells (Supplementary Fig. 2e, g).

IFN α induces early changes in chromatin accessibility. The ATAC-seq experiments demonstrated that IFN α induces early changes in chromatin accessibility, with >4400 regions of gained open chromatin regions (OCRs) detected at 2 h, which decreased to 1000 regions by 24 h (Fig. 1b and Supplementary Data 2); only nine regions had loss of chromatin accessibility (Fig. 1b). Most of the OCRs at 24 h were already modified at 2 h (fast response), and only 10% of OCRs were specifically gained at 24 h (late response). The gained OCRs were mostly localized distally to gene transcription starting sites (TSS) (Supplementary Fig. 3a) acting, therefore, as potential regulatory elements. These regions are evolutionary conserved (Supplementary Fig. 3b), and enriched for transcription factors (TFs) binding motifs (Supplementary Fig. 3c), including islet-specific TFs binding sequences.

To assess whether changes in chromatin remodeling were associated with variations in gene expression, we first quantified the frequency of ATAC-seq regions gained or stable in the proximity (40 kb window centered on the TSS) of genes with differential mRNA expression (up/down/non-regulated or non-expressed) (Supplementary Data 2). There was a higher proportion of upregulated genes associated with gained OCRs in comparison to stable regions at each time point analyzed (Fig. 1c). Moreover, the number of gained OCRs was associated with changes in both the proportion (Fig. 1d) and the intensity (Supplementary Fig. 3e) of transcript induction (Supplementary Fig. 3d, see Methods for more information). There was also a minor association between the number of stable regions and upregulated mRNAs at 2 h (Supplementary Fig. 3e), likely due to the activation of already nucleosome-depleted regions ahead of cytokine exposure¹⁹. Consistently with these results, there was an increase in the frequency of upregulated proteins coded by genes proximal to gained OCRs (Fig. 1e). Likewise, there was a progressive increase in IFN α -induced protein abundance depending on the number of linked gained open chromatin regions (Fig. 1f).

There was a strong correlation between upregulated mRNAs and induced proteins (r^2 : 0.66 and 0.65 at, respectively, 8 and 24 h, $p < 2.2 \times 10^{-16}$) (Fig. 1g, first column), but a much lower similarity between downregulated mRNAs and proteins (Fig. 1g, second column). Gene ontology analysis of differentially abundant proteins upon IFN α treatment identified several biological processes involved in the pathogenesis of T1D, such

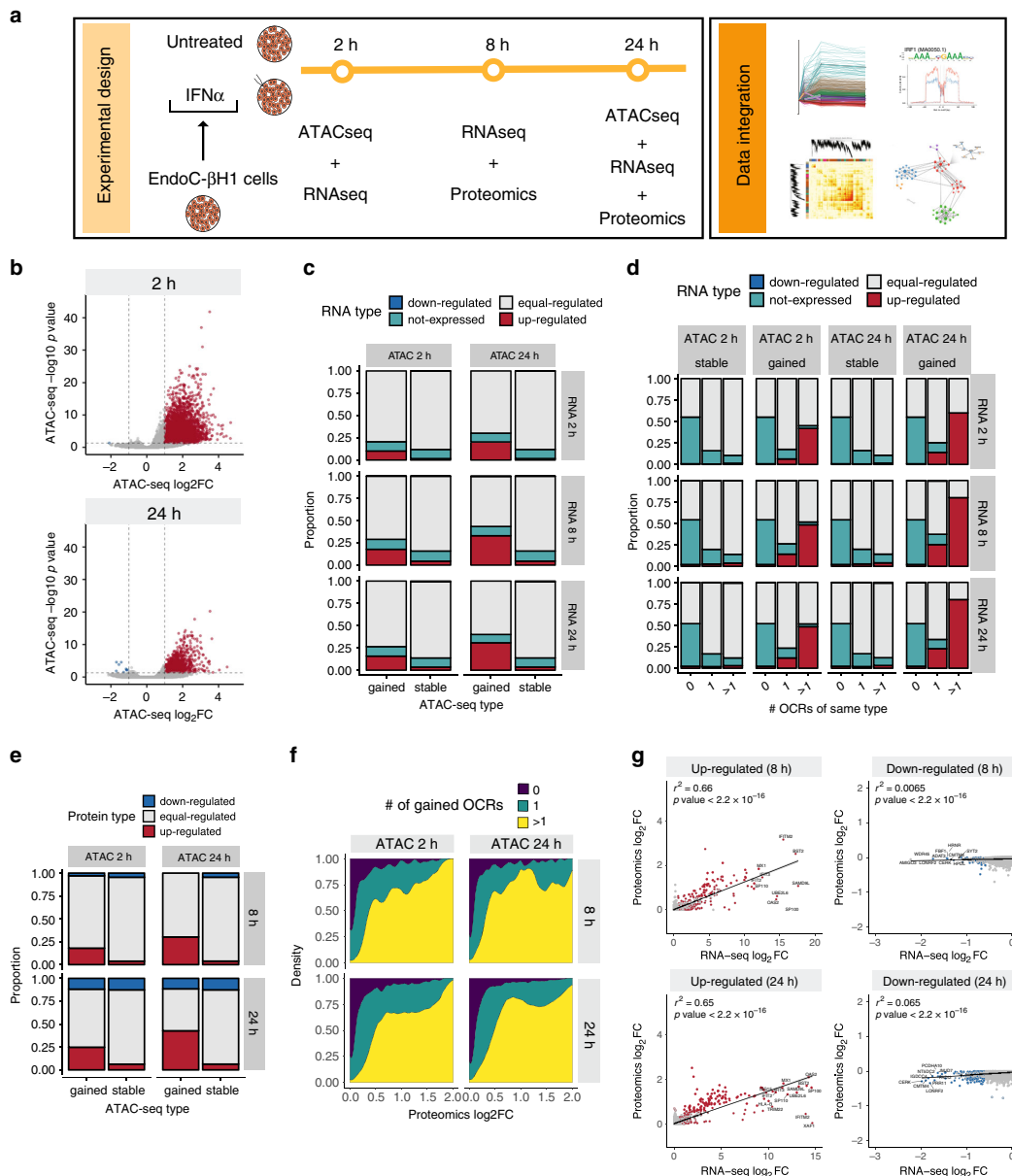


Fig. 1 Exposure of EndoC-βH1 cells to interferon-α promotes changes in chromatin accessibility, which are correlated with gene transcription and translation. **a** EndoC-βH1 cells were exposed or not to IFNα (2000 U/ml) for the indicated time points and different high-throughput techniques were performed to study chromatin accessibility (ATAC-seq, $n = 4$), transcription (RNA-seq, $n = 5$) and translation (Proteomics, $n = 4$). **b** Volcano plot showing changes in chromatin accessibility measured by ATAC-seq. Open chromatin regions indicated as gained (red) or lost (blue) had an absolute \log_2 fold-change (\log_2FC) > 1 , and a false discovery rate (FDR) < 0.05 . The regions that did not reach such threshold were considered “stable” (gray). **c, d** Frequency of upregulated, downregulated or stable transcripts in the vicinity (< 20 kb transcription start site (TSS) distance) of one or multiple open chromatin regions (OCRs) as classified in **b**. **e** Frequency of differentially abundant proteins in the vicinity (< 20 kb TSS distance) of gained or stable open chromatin regions. **f** Distribution of IFNα-induced changes in protein abundance among upregulated proteins based on the number of linked gained OCRs. **g** Correlation between RNA-seq and proteomics of EndoC-βH1 cells exposed to IFNα. The x axis represents the mRNA \log_2FC . The most upregulated ($\log_2FC > 0.58$, FDR < 0.05) and downregulated ($\log_2FC < -0.58$, FDR < 0.05) mRNAs are filled in red and blue, respectively. The y axis indicates the proteomics \log_2FC . The proteins most upregulated ($\log_2FC > 0.58$, FDR < 0.15) or downregulated ($\log_2FC < -0.58$, FDR < 0.15) are represented by red and blue borders, respectively. mRNAs and proteins not meeting these criteria were considered equal-regulated (gray fill and border, respectively).

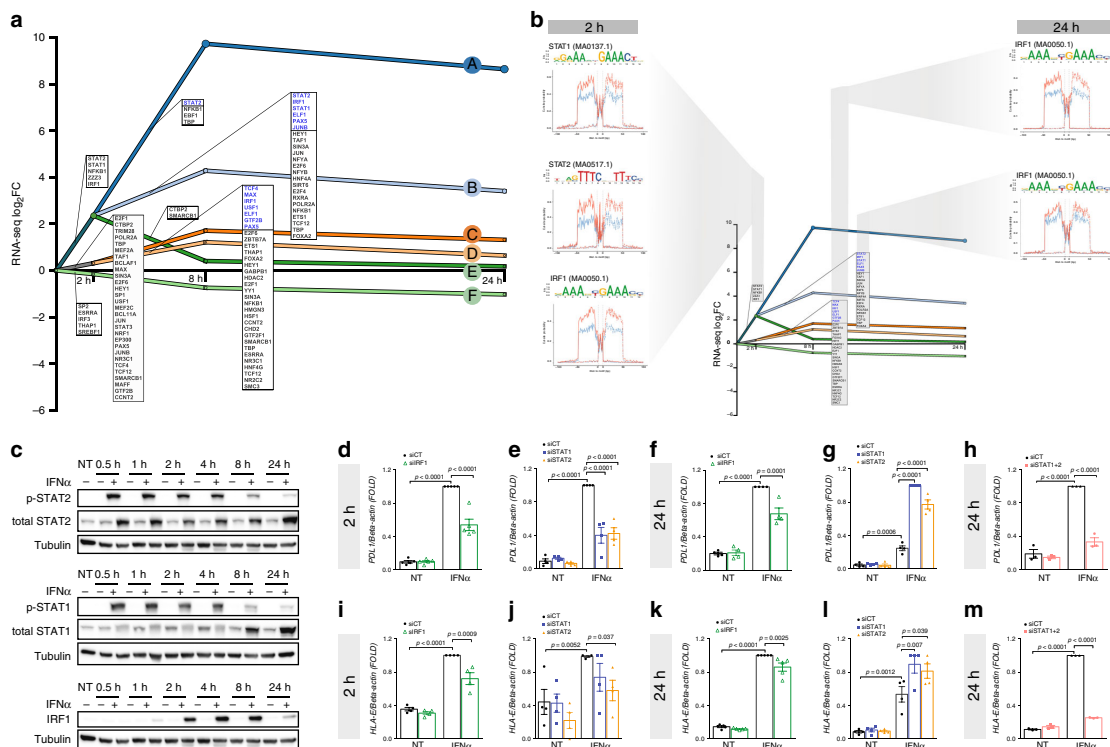


Fig. 2 IRF1, STAT1 and STAT2 regulate IFN α -induced transcription and the expression of checkpoint proteins. **a** The regulatory paths summarize the temporal patterns of the differentially expressed genes (DEG) detected by RNA-seq ($|\log_2FC| > 0.58$ and $FDR < 0.05$, $n = 5$) (evaluated by DREM²⁵). The x axis represents the time and the y axis the mRNA \log_2FC . Each path corresponds to a set of co-expressed genes. Split nodes (circles) represent a temporal event where co-expressed genes diverge in expression. In blue are the TFs upregulated at the respective time points of the RNA-seq that may regulate the pathways. **b** IFN α promoted TFs footprint deepening in open chromatin regions (OCR) associated to genes from the indicated DREM pathways. OCRs were associated to the nearest gene TSS with a maximum distance of 1 Mb. Previously annotated TF matrices⁷⁹ were used to identify differential DNA-footprints induced by IFN α (blue lines = untreated cells, red lines = IFN α (24 h), dashed lines = reverse strand, continuous line = forward strand, Methods, $n = 4$). **c** Time course profile of STAT1, STAT2 and IRF1 protein activation in EndoC- β H1 cells exposed to IFN α (representative of four independent experiments). **d-m** EndoC- β H1 cells were transfected with an inactive control siRNA (siCT) or previously validated^{7,24} siRNAs targeting IRF1 (siIRF1), STAT1 (siSTAT1), STAT2 (siSTAT2) or STAT1 plus STAT2 (siSTAT1+2). After 48 h the cells were exposed to IFN α (24 h). The values were normalized by the housekeeping gene β -actin (mRNA) and then by the highest value of each experiment considered as 1 (for **h** and **m** ($n = 3$); for **e-g**, **i**, **j** and **l** ($n = 4$); for **d**, **k** ($n = 5$)), ANOVA with Bonferroni correction for multiple comparisons (**d-m**). Values are mean \pm SEM (**d-m**). Source data are provided as Source Data file.

as antigen processing and presentation, responses to viruses, apoptosis and NK/T-cell responses (Supplementary Fig. 4a, b); groups of genes associated to protein modification and degradation were also present (Supplementary Fig. 4a, c). Furthermore, genes related to endoplasmic reticulum (ER) stress, another post-transcriptional mechanism that downregulates translation of many mRNAs²⁰, were also upregulated by IFN α at both the mRNA and protein levels (Supplementary Fig. 4d). These findings are in line with our previous observations⁷ and were confirmed here in independent samples for two key ER stress markers, namely the transcription factor *ATF3*²¹ and the ER chaperon *HSPA5* (also known as BiP/GRP78)²² (Supplementary Fig. 4e-h). ER stress often decreases translation, which may explain the weak association observed between mRNA and protein expression in downregulated mRNAs and proteins (Fig. 1g).

IRF1, STAT1 and STAT2 are key regulators of IFN α signaling.
To identify the key transcription factors involved, the expression

of differentially expressed genes (DEG) from all RNA-seq time points (Supplementary Data 3) was analyzed using the dynamic regulatory events miner (DREM) model²³. This approach identified six patterns of co-expressed genes (Fig. 2a); 5 out of 6 pathways had an early peak of induction (2 or 8 h), which then decreased or remained stable until 24 h (Fig. 2a). The model compared the frequency of TF binding sites in the gene promoters between divergent branches of co-expressed genes, assuming that these TFs are responsible for the observed differences in gene expression profiles (Fig. 2a). This was compared with the TF occupancy determined by assaying the protection of the bound sequence to ATAC-seq transposase cleavage (footprint) (Supplementary Fig. 5a and Methods). There were footprints for the transcription factors IRF1, STAT1 and STAT2, which were deepened upon IFN α exposure in pathway B (which had the highest transcriptional upregulation at 2 h) and for IRF1 in two independent pathways, namely B and D at 24 h (Fig. 2b). Western blot analysis confirmed the activation of these TFs (Fig. 2c). STAT1 and STAT2 phosphorylation peaked between 0.5 and 1 h and then returned to near-basal levels at 24 h, while IRF1

peaked later, at 4–8 h decreased by 24 h (Fig. 2c); these findings support the observed TF footprint profiles (Fig. 2b). There was also a close correlation between DEGs induced by IFN α in RNA-seq of EndoC- β H1 cells and in human pancreatic islets (Supplementary Fig. 1b; $p < 2.2 \times 10^{-22}$ at 2, 8 and 24 h), which resulted in a similar pattern of gene activation under the control of analogous TFs (Supplementary Fig. 1c and Supplementary Data 4).

Individual DREM pathways usually regulate specific biological processes (GO) (Supplementary Fig. 5b, 1d). Among them, was the term “Regulation of immune responses” (Supplementary Fig. 5b). This pathway comprises several genes involved in the crosstalk between beta cells and the immune system, such as *PDL1* (*CD274*), an immune checkpoint protein expressed in the islets of T1D individuals²⁴, and a second co-inhibitory molecule, *HLA-E*, recently identified as potential target for cancer immunotherapy²⁵ (Fig. 2d–m).

By using a previously validated siRNA targeting *IRF1*²⁴, we obtained around 60% knockdown (KD) of IFN α -induced IRF1 protein and mRNA expression at 2 and 24 h (Supplementary Fig. 5c–f). IRF1 silencing led to a significant decrease in IFN α -induced *PDL1* and *HLA-E* mRNA expression (Fig. 2d, f, i, k). Silencing of IRF1 also decreased IFN α -induced upregulation of the chemokines *CXCL1* and *CXCL10*, the HLA-I component beta-2-microglobulin (*B2M*) and the suppressor of cytokine signaling 3 (*Socs3*) (Supplementary Fig. 5d, f). Small interference RNAs targeting *STAT1* (siSTAT1) or *STAT2* (siSTAT2) promoted >70% KD of their respective proteins and mRNAs, (Supplementary Fig. 5g–j). Inhibiting *STAT1* or *STAT2* alone partially blocked the induction of *PDL1* and *HLA-E* at 2 h (Fig. 2e, j), but led to a paradoxical increase in *PDL1* and *HLA-E* expression at 24 h (Fig. 2g, i), which is probably due to a compensatory increase in expression of the non-targeted *STAT2*²⁴. In line with this, double KD of *STAT1* + *STAT2* led to downregulation of both *PDL1* and *HLA-E* (Fig. 2h, m). *STAT2* inhibition decreased the 2 h expression of IFN α -induced *CXCL1/10*, *Socs1* and *MX1*, whereas *STAT1* KD only prevented *CXCL10* induction (Supplementary Fig. 5h). At 24 h only 2 out of 4 genes remained partially inhibited by siSTAT2 (Supplementary Fig. 5j), whereas double KD of *STAT1* + *STAT2* prevented IFN α -induced gene upregulation at 24 h in most cases (Supplementary Fig. 5k).

Exposure of FACS-purified human beta cells (Supplementary Fig. 6a–c) to IFN α confirmed the upregulation of genes related to antigen presentation (*HLA-I*), antiviral responses (*MX1*, *MDA5*), ER stress (*CHOP*), immune cells recruitment (*CXCL10*) and checkpoint regulators (*PDL1*) (Supplementary Fig. 6d).

The checkpoint protein *PDL1* is overexpressed in beta cells from people with T1D²⁴, and we presently evaluated the expression of another checkpoint protein, i.e. *HLA-E*²⁵. IFN α upregulated *HLA-E* mRNA expression in EndoC- β H1 cells (Fig. 3a), dispersed human islets (Fig. 3b) and FACS-purified human beta cells (Fig. 3c) and augmented *HLA-E* protein expression in both EndoC- β H1 cells (Fig. 3d) and human islets (Fig. 3e), with peak at 24 h. The inhibitory effects of *HLA-E* on immune cells require its expression on the cell surface or its secretion²⁶. Flow cytometry confirmed that IFN α increases surface *HLA-E* expression (Fig. 3f, g, Supplementary Fig. 5l), but there was no *HLA-E* release to the supernatant (Supplementary Fig. 5m). *HLA-E* mRNA expression was upregulated by 8-fold in human islets of donors with recent-onset T1D in the DiViD study²⁷ and *HLA-E* protein expression was significantly increased in insulin-containing islets, but not in insulin-deficient islets, of T1D individuals in comparison to healthy individuals (Fig. 3h, i). *HLA-E* expression was present in both beta and alpha cells (but not delta cells; Supplementary Fig. 5n) in the islets of people with T1D, with a predominance of expression among

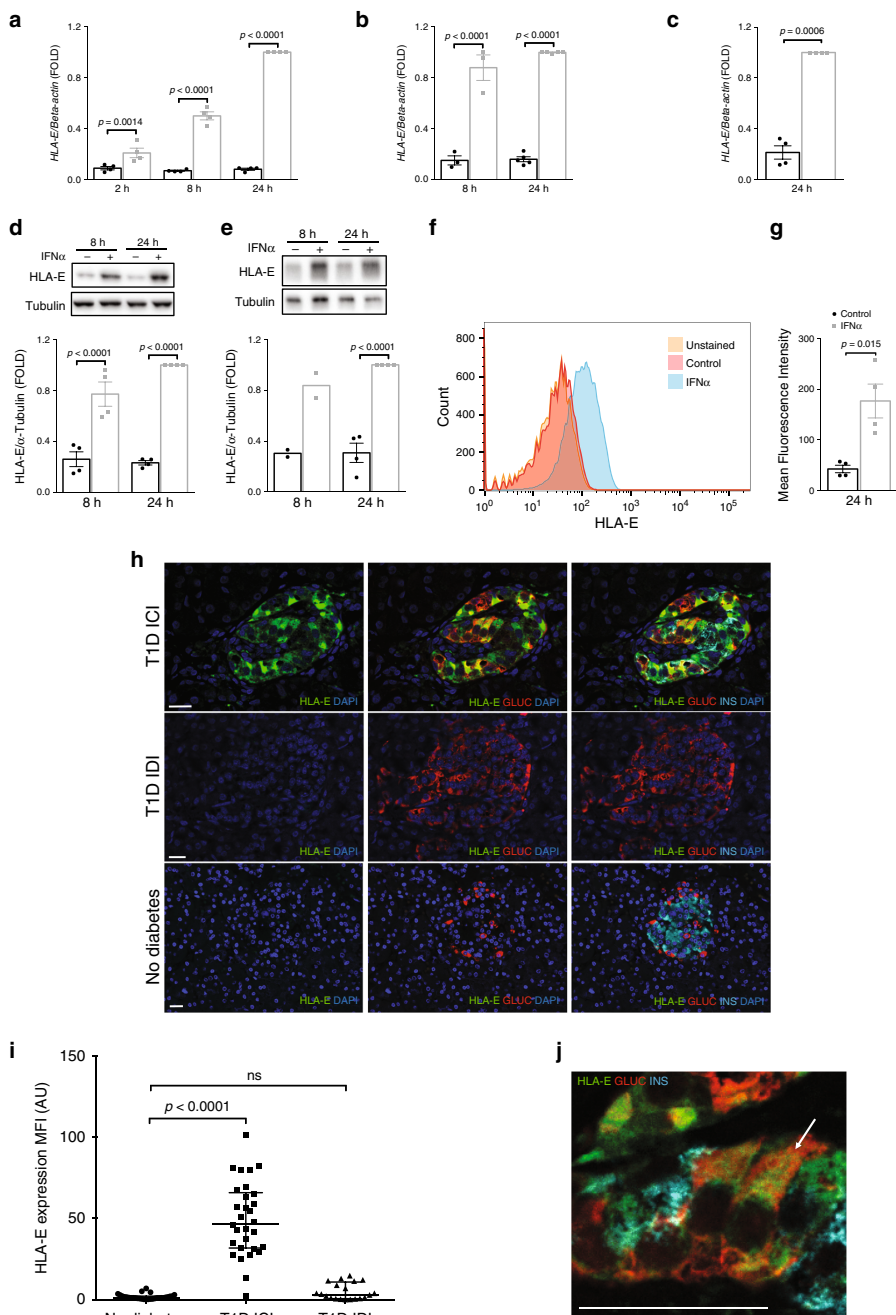
alpha cells as compared to beta cells (Fig. 3j). This may help to explain why alpha cells are more resistant to the immune assault in T1D.

mRNA and protein modules regulated by interferon- α . We integrated the RNA-seq and proteomics data (using all the samples from both 8 and 24 h) using the weighted correlation network analysis package (WGCNA)²⁸. The heatmaps of the topological overlap matrix from each dataset with module assignment are shown in Fig. 4a. There were initially 32 eigengene modules of mRNAs and 27 of proteins, which were merged (considering a dissimilarity threshold of 0.25) reducing the numbers of mRNA and protein modules to 8 and 7, respectively (Supplementary Fig. 7a–c). The quality of these modules was determined using a combined score of density and separability measures (Methods)²⁹, which indicated that they were well-defined ($Z_{\text{summary}} > 10$) (Supplementary Fig. 7d). WGCNA analysis of the RNA-seq of human islets exposed to IFN α identified well-defined modules of mRNAs (Supplementary Fig. 8a–d), similar to the ones identified in EndoC- β H1 cells exposed to the cytokine (Supplementary Fig. 8e). To focus on central modules induced by IFN α exposure, we selected only the differentially expressed genes (DEG) (Supplementary Data 3) and abundant proteins (DAP) (Supplementary Data 5) in each eigengene module, representing 49% of the protein-coding DEGs and 89% of the DAPs, and then examined the overlap between these datasets. There was a significant overlap between five modules of mRNAs and proteins (minimum of 10 elements in common, $FDR < 0.05$) (Fig. 4b). The two main new modules, called #1 and #2 (Fig. 4c), were composed of highly correlated mRNAs and proteins (Supplementary Fig. 7e, g) predominantly upregulated by IFN α at both 8 and 24 h (Supplementary Fig. 7f, h). Module #5 also had significantly correlated members (Supplementary Fig. 7i), but enriched in downregulated mRNAs/proteins at both 8 and 24 h (Supplementary Fig. 7j). Interestingly, there was significant enrichment of ATAC-seq gained OCRs in module #2 (Fig. 4d). They were enriched for TF binding motifs including both the pro-inflammatory motifs ISRE / IRF and the islet-specific transcription factor FOXA2 (Fig. 4e).

To identify the gene regulatory network (GRN) of module #2, we integrated information from two sources: (1) literature-based collection of TF-target interactions³⁰, and (2) the present de novo TF binding motifs and their predicted targets (Supplementary Fig. 9a). This allowed us to add information from *cis*-regulatory elements (in orange) acting on the IFN α -induced GRN in human beta cells (Supplementary Fig. 9b). A similar approach was used for modules #1 and #5, but considering only data from the literature (Supplementary Fig. 10a, c).

The PPI network InWeb InBio Map³¹ was used to assess the presence of protein–protein interaction (PPI) networks in the different modules. This generated networks of interacting proteins for modules #1, #2 and #5 (Fig. 4f and Supplementary Fig. 10b, d) and allowed the recognition of protein communities (grouped by colors) that regulate specific and common biological functions (Fig. 4f and Supplementary Fig. 10b, d). Module #2, which presents the higher number of connections, showed an enrichment for several key biological processes activated by IFN α and relevant for the pathogenesis of T1D, including cellular response to viruses, antigen processing and presentation via MHC class I, inflammatory and acute phase responses (Fig. 4g).

Interferon- α changes the alternative splicing landscape. The present high-coverage RNA-sequencing (>200 million reads) allowed the detection of ~47,000 splicing variants, with IFN α -induced 343 differentially expressed transcripts (DETs) at 2h,



and 1690 and 1669, respectively at 8 and 24 h, with predominance of upregulated transcripts (Fig. 5a and Supplementary Data 6 and 7). Considering all the DETs, 4%, 32% and 32% were exclusively modified at 2, 8 and 24 h, respectively, indicating a predominance of intermediary to late transcriptional changes induced by IFN α . Next, we evaluated the frequency of each individual splicing events (with an absolute difference in percent spliced-in ($|\Delta\text{PSI}| > 0.2$) regulated by IFN α at 8 and 24 h. There were 3140 events at

8 h and 2344 events at 24 h (FDR < 0.05) (Fig. 5b). The most frequent AS event modified by IFN α was cassette exons (CEX), with predominantly increased exon inclusion (represented by $\Delta\text{PSI} > 0.2$, FDR < 0.05) (Fig. 5c). An example of a cassette exon showing increased inclusion upon IFN α treatment is the gene *OASL* (Fig. 5d, e), an antiviral factor targeting single-stranded RNA viruses such as picornaviruses³². Exposure to IFN α for 24 h increased exon 4 inclusion in both EndoC- β H1 cells and human

Fig. 3 HLA-E is overexpressed in pancreatic islets of T1D individuals. EndoC- β H1 cells (**a, d**), human islets (**b, e**) or FACS-purified human beta cells (**c**) were exposed (gray bars) or not (black bars) to IFN α for the indicated time points and HLA-E mRNA (**a–c**) and protein (**d, e**) evaluated. The values were normalized by the housekeeping gene β -actin (mRNA) or α -tubulin (protein) and then by the highest value of each experiment considered as 1 (for **a** ($n = 4$); **b** ($n = 3$ (8 h), $n = 5$ (24 h)); **c** ($n = 4$); **d** ($n = 4$) and **e** ($n = 2$ (8 h), $n = 4$ (24 h))), ANOVA with Bonferroni correction for multiple comparisons (**a–e**). **f, g** HLA-E cell surface expression was quantified in EndoC- β H1 cells by flow cytometry. Histograms (**f**) represent changes in mean fluorescence intensity (MFI). The MFI values (**g**) were quantified at baseline and after 24 h exposure to IFN α ($n = 4$, two-sided paired *t*-test). Values are mean \pm SEM (**a–g**). **h** Immunostaining of HLA-E (green), glucagon (red) and insulin (light blue) in representative islets from individuals with or without diabetes. The top and middle panels represent an insulin-containing islet (ICI) and insulin-deficient islet (IDI) from T1D sample DiViD 3, and the lower panel represents an islet from a control donor (EADB sample 333/66). DAPI (dark blue). Scale bar 20 μ m. **i** The MFI analysis of HLA-E expression. 30 ICIs from 6 independent individuals with T1D (5 islets per individual), 20 IDIs from 4 independent individuals with T1D (5 islets per individual), and 30 ICIs from 6 independent individuals without diabetes (5 islets per individual) were analyzed. Values are median \pm interquartile range; ANOVA with Bonferroni correction for multiple comparisons, AU (arbitrary units), ns = (non-significant). **j** Higher magnification image demonstrating that HLA-E (green) localizes predominantly to alpha cells in a T1D donor islet (glucagon (red); insulin (light blue)) but is also expressed in beta cells, as indicated in **h** and **j**. Scale bar 30 μ m. Source data are provided as Source Data file.

islets (Fig. 5d). In line with this, the protein encoded by the isoform *OASL*–001 (which retains exon 4) displayed a higher IFN α -induced upregulation in comparison with the protein encoded by the isoform *OASL*–002, which has exon 4 exclusion (Fig. 5e). Interestingly, the isoform *OASL*–001 has antiviral activity, whereas the isoform 002 lacks the ubiquitin-like domain required for this response (Supplementary Fig. 11A)³³.

Intron retention is an important mechanism of gene expression regulation, promoting nuclear sequestration of transcripts or cytoplasmic degradation via nonsense-mediated decay³⁴. There was a predominance for intron removal after 24 h (represented by Δ PSI < -0.2 , FDR < 0.05), but not at 8 h (Fig. 5f). To understand how this impacts protein translation, we compared changes in protein abundance among three categories of Δ PSI. Genes presenting intron removal had a significant increase in protein expression after IFN α exposure for 24 h in comparison to those with intron retention (Δ PSI > 0.2 , FDR < 0.05) or with non-significant intron changes (Δ PSI -0.2 – 0.2 or FDR > 0.05) (Fig. 5g).

There were clear variations in the mRNA expression of several well-known RNA-binding proteins (RBPs)³⁵ upon IFN α exposure (Fig. 5h, left panel), but the impact on the respective proteins was less pronounced (Fig. 5h, right panel). We focused on a group of IFN α -modified RBPs at both mRNA and protein levels after 24 h, and mapped their RNA-binding motifs among upregulated and downregulated alternative exons. In support of a biological role for these RBPs on alternative exon splicing, there was an enrichment of their binding motifs in regions controlling alternative cassette exon inclusion/exclusion (Fig. 5i). To further study some of these findings, we first reproduced the IFN α -induced downregulation of two RBPs, ELAV-like protein 1 (*ELAVL1*) and heterogeneous nuclear ribonucleoprotein (*HNRNPA1*), by using specific siRNAs (Supplementary Fig. 12a, e). Next, we evaluated whether this inhibition reproduced the changes induced by IFN α in the exon usage of four-and-a-half LIM domain protein 1 (*FHL1*) and Caprin Family Member 2 (*CAPRN2*) (Supplementary Fig. 12b, f) two potential targets of, respectively, *ELAVL1*³⁶ and *HNRNPA1*³⁷. Silencing these RBPs promoted changes on exon usage (Supplementary Fig. 12c, g) that were similar to the ones observed after IFN α treatment (Supplementary Fig. 12b, f). This is especially relevant in the context of the IFN α -induced exon exclusion *FHL1*, which decreases the expression of transcripts coding for the protein *FHL1A* (Supplementary Fig. 12d), an isoform described as a key host factor for the replication of the RNA virus Chikungunya³⁸.

RBPs can also control gene expression by blocking RNA translation, as described for the Fragile X Mental Retardation 1 (*FMRI*) gene³⁹. Indeed, there was a significant downregulation of previously validated bona fide targets of *FMRI* (Supplementary

Table 1)⁴⁰ in IFN α -treated EndoC- β H1 cells as compared to the remaining proteins (Fig. 5j).

IFN α induces increased alternative transcription start sites.

The usage of alternative transcription start (TSS) sites is another mechanism that generates different transcripts from the same gene⁴¹. We used the SEASTAR pipeline⁴² for the computational identification and quantitative analysis of first exon usage. This approach recognized >250 events of alternative first exon (AFE) usage occurring in 166 different genes at 8 h, and >130 events of AFE usage in 88 genes at 24 h (Fig. 6a). In agreement with this, 118 and 64 alternative promoters (± 2 kb around FE TSS) detected by SEASTAR at 8 and 24 h, respectively, overlapped peaks of TSS identified by the FAMTOM5 Consortium⁴³. Among these genes was the 5'-nucleotidase cytosolic IIIA (*NT5C3A*), a negative regulator of IFN-I signaling⁴⁴. This gene had two AFEs identified by the SEASTAR modeling. In untreated condition (controls), there was a higher usage of the proximal first exon (FE), present in the isoforms *NT5C3A*–001 and 002 in beta cells (Fig. 6b, upper panel). After IFN α exposure, however, there was increased usage of the distal FE from the transcript *NT5C3A*–004 (Δ PSI: 0.71 (8 h)/0.65 (24 h), both FDR < 0.001), which is supported by the cap analysis of gene expression (CAGE) of TSSs⁴⁵ (Fig. 6b, upper panel). This was confirmed in independent samples of EndoC- β H1 cells and human islets using specific primers (Fig. 6b, lower panel). Exon Ontology analysis⁴⁶ indicated that this FE shift probably has functional impact, since the distal FE lacks both the endoplasmic reticulum (ER) retention signal and the transmembrane helix (Supplementary Fig. 11b), enabling its encoded protein to remain in the cytosol where *NT5C3A* acts⁴⁴.

Next, we compared the frequency of gained OCRs among alternative promoters. As the SEASTAR pipeline mainly recognizes non-redundant FEs, we evaluated alternative promoters identified by both the SEASTAR pipeline and the FAMTOM5 database of alternative TSSs⁴⁵ (Supplementary Methods). We thus identified 198 and 51 gained OCRs present in alternative promoter regions at 2 and 24 h, respectively. Characterization of the IFN α -induced alternative promoters presenting a major gain in chromatin accessibility pointed to the T1D risk gene *RM12*⁴⁷. At gene level, there was only a ~ 1.4 -fold upregulation of *RM12* expression, but at the transcript level there was a >60 -fold increase in two isoforms, *RM12*–002 and –004. Visualization of the *RM12* locus combined with ATAC-seq and RNA-seq peaks indicated that the isoform *RM12*–004 gained chromatin accessibility in its promoter leading then an increase in mRNA expression (Fig. 6c). Data from CAGE analysis⁴⁵ and RNA polymerase II ChIP-seq of another human cell type exposed to IFN α ⁴⁸ (Fig. 6c, lower part) confirms the presence of the *RM12*

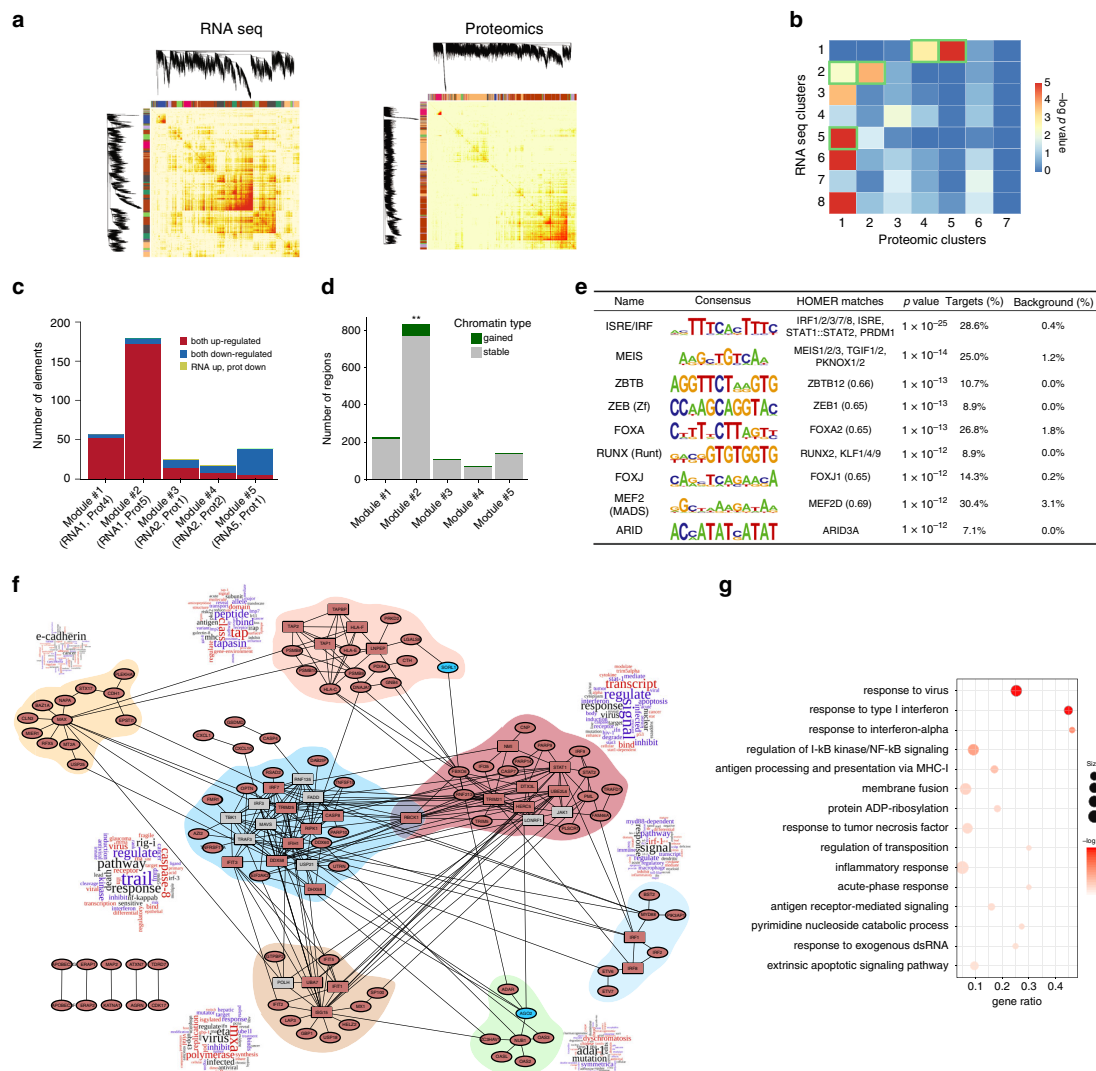
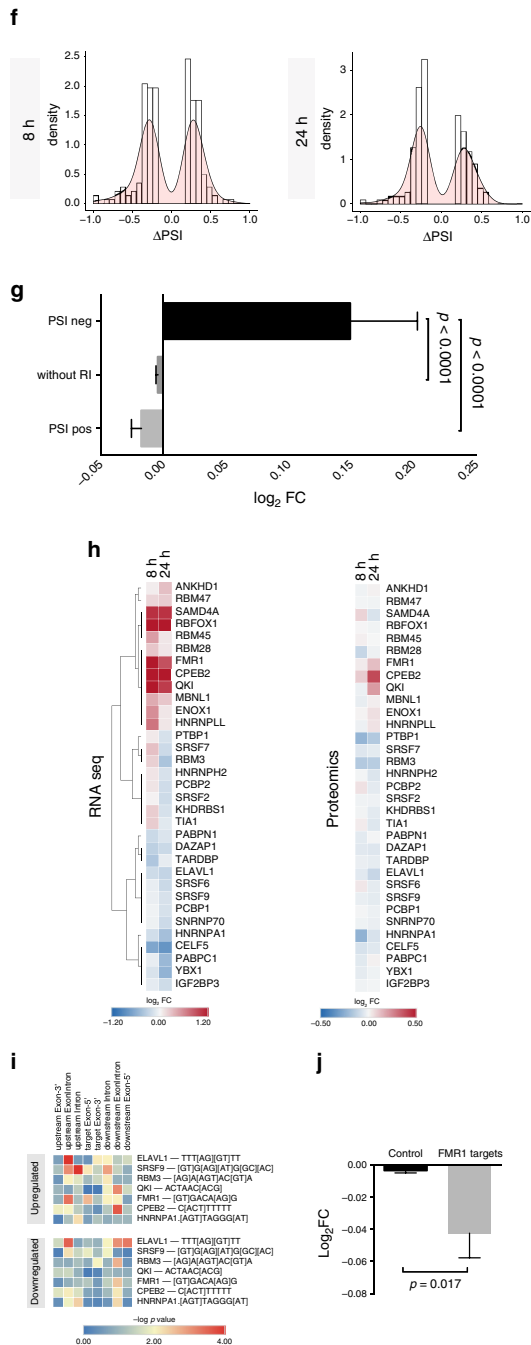
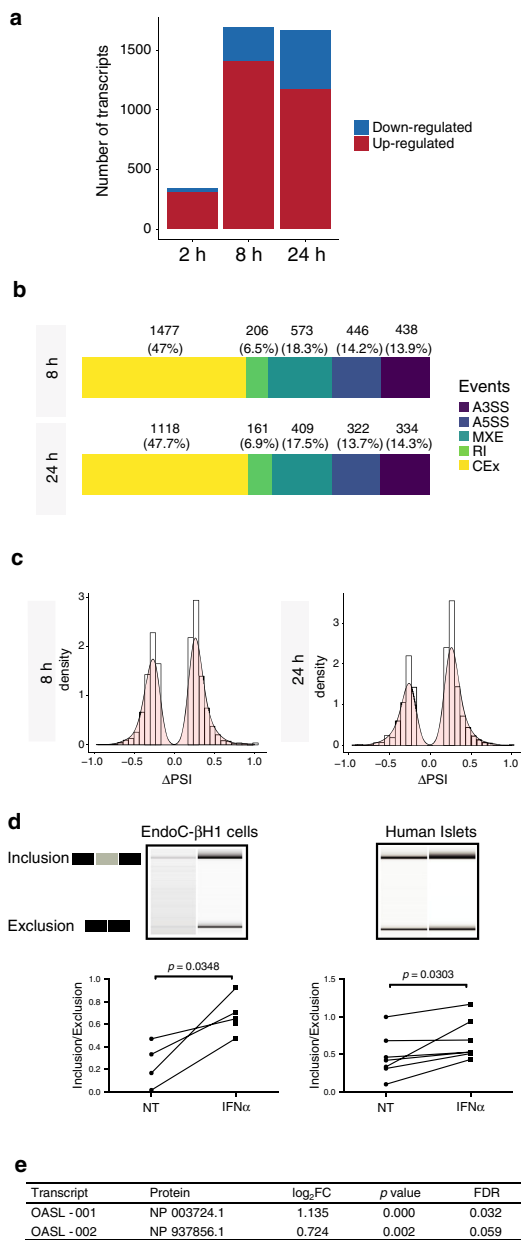


Fig. 4 Weighted correlation network analysis (WGCNA) identifies IFN α -regulated mRNA and protein modules. **a** Heatmap representation of the topological overlap matrix. Rows and columns correspond to single genes/proteins, light colors represent low topological overlap, and progressively darker colors represent higher topological overlap. The corresponding gene dendrograms and initial module assignment are also displayed. **b** Identification of modules presenting significant overlap (FDR < 0.05 and a minimum of 10 members in common) (green border) between differentially expressed genes (DEG) and their translated differentially abundant proteins (DAP). **c** Composition, number of elements and type of DEG and DAP present in each of the significantly overlapping modules. **d** ATAC-seq-identified open chromatin regions at 2 h were linked to gene transcription start sites (TSS) in a 40 kb window. These genes and their open chromatin regions were associated to the modules of DEG and DAP. The enrichment for gained open chromatin regions was then evaluated in each module. (** represents a p -value = 0.002343, one-sided χ^2 test). **e** De novo HOMER motifs present in the ATAC-seq regions overlapping module #2 as described in Methods. The unadjusted p -values were obtained using the hypergeometric test from the HOMER package⁷⁷. **f** The protein-protein interaction (PPI) network of module #2 was done using the InWeb InBio Map database³¹. Enriched proteins (FDR < 0.05 and minimum number of connections = 5, represented as squares) were identified and added to the network if they were not already present. Red fill identifies upregulated proteins, blue fills downregulated proteins and gray fill equal-regulated. Colored regions delimitate communities of proteins, as described in Methods. The wordcloud next to each community presents their enriched geneRIFs terms. **g** The biological processes (GO) overrepresented in module #2 summarize the main findings observed in IFN α -treated human beta cells. The present results were based on RNA-seq ($n = 5$) and proteomics ($n = 4$) data of EndoC- β H1 cells.



alternative promoter. The IFN α -induced *RMI2*-004 upregulation was confirmed using specific primers in both EndoC- β H1 cells and human islets (independent samples) (Fig. 6d). These findings support a double mechanism by which IFN α affects human beta cells, i.e. first a massive change in open chromatin regions followed by later changes in gene expression and AS (see above) and also AFE usage.

Mining IFN α signatures to identify T1D therapeutic targets. Considering the significant overlap observed between gene profiles of IFN α -exposed EndoC- β H1 cells and beta cells from T1D individuals (Supplementary Fig. 2d), mining these common signatures might identify relevant T1D therapeutic targets. First, the top 150 commonly upregulated genes detected by the RRHO analysis of both IFN α -exposed EndoC- β H1 cells and beta cells

Fig. 5 Interferon- α changes the alternative splicing landscape. **a** EndoC- β H1 cells were exposed to IFN α for the indicated time points. The significantly upregulated (red) and downregulated (blue) transcripts were identified using Flux Capacitor ($n = 5$, $|\log_2FC| > 0.58$ and $FDR < 0.05$). **b** Frequency of individual alternative splicing events regulated by IFN α ($n = 5$, $|\Delta PSI| > 0.2$, minimum 5 reads, $FDR < 0.05$). **c** Frequency distribution of alternative cassette exon (CEX) events altered by IFN α ($n = 5$, $|\Delta PSI| > |0.2|$ and $FDR < 0.05$). **d** Confirmation of the increased exon 4 inclusion in the antiviral gene OASL by IFN α (24 h). cDNA was amplified by RT-PCR using primers located in the upstream and downstream exons of the splicing event and the product evaluated using a Bioanalyzer 2100 ($n = 4$ (EndoC) and $n = 7$ (human islets), two-sided paired t -test). **e** The \log_2FC s of the proteins coding for OASL-001 and -002 isoforms from IFN α -treated EndoC- β H1 cells proteomics (24 h) ($n = 4$). **f** Frequency distribution of retained intron (RI) events altered by IFN α ($n = 5$, $|\Delta PSI| > 0.2$ and $FDR < 0.05$). **g** The protein \log_2FC values obtained by proteomics analysis of EndoC- β H1 cells exposed to IFN α for 24 h were classified in three categories according to the levels of retained intron ΔPSI ($n = 5$, mean \pm SEM, ANOVA with Bonferroni correction). **h** Expression of RNA-binding proteins (left) that are significantly modified at mRNA level ($FDR < 0.05$) after exposure to IFN α and their respective proteins (right) in the indicated time points ($n = 4-5$). **i** Positional enrichment of motifs from significantly modified RBPs among regions involved in the regulation of modified cassette exons (CEX) after exposure to IFN α for 24 h. ($n = 5$, $|\Delta PSI| > 0.2$, $FDR < 0.05$). **j** Comparison between the \log_2FC of a curated list (Supplementary Table 1) of known FMRI target proteins against the \log_2FC of the remaining proteins detected by the proteomics of EndoC- β H1 cells exposed to IFN α for 24 h ($n = 4$, mean \pm SEM; two-sided unpaired t -test). Source data are provided as a Source Data file.

from T1D individuals were selected (Supplementary Fig. 2d and Fig. 7a) to query the Connectivity Map database⁴⁹. We focused in opposite signatures of perturbagens that may reverse the effects of IFN α . To decrease off-target findings based on individual compounds, the analysis was performed considering only the classes of perturbagens. Four main classes, including bromodomain inhibitors, potentially reversed the signature from our query (tau score < -90) (Fig. 7b). Comparable results were obtained when analyzing the intersection of IFN α -exposed pancreatic human islets and beta cells from T1D individuals (Supplementary Fig. 13a). Bromodomain inhibitors have been shown to prevent autoimmune diabetes in animal models⁵⁰ and the KD of the bromodomain containing 2 gene (*BRD2*) induced an opposite signature to our model (Supplementary Fig. 13b). Pre-treatment of EndoC- β H1 cells with two bromodomain inhibitors decreased both IFN α -induced *HLA-I* and *CXCL10* induction, with no changes in *CHOP* (*DDIT3*) expression (Fig. 7c, e) or in apoptosis induced by IL1 β + IFN α (Fig. 7d, f). In human islets, these inhibitors induced a ~30% decrease in IFN α -induced *HLA-I* expression and a 90% reduction in *CXCL10* expression; at least in the context of I-BET-151, there was a 60% reduction of the ER stress marker *CHOP* (*DDIT3*) (Supplementary Fig. 13c, d).

Next, we searched for clinically approved drugs (DrugBank 5.1⁵¹) among the PPI network of the WGCNA module #2 (Fig. 4f), with a view to possible drug repurposing. Module #2 is particularly interesting in this context as it recapitulates many of the key biological processes induced by IFN α (Fig. 4g), and because ~50% of its members were also present among the most upregulated genes from the RRHO analysis (Supplementary Fig. 2d). An interesting target recognized as a hub for different drugs was the kinase JAK1 (Fig. 8a) and its inhibitor baricitinib, which has shown promising effects in the treatment of human rheumatoid arthritis⁵². Baricitinib prevented IFN α -induced mRNA expression of *HLA-I*, *CXCL10* and *CHOP* (*DDIT3*) in EndoC- β H1 cells (Fig. 8b) and human islets (Fig. 8c) and it completely protected EndoC- β H1 cells (Fig. 8d) and human islets (Fig. 8e) against the pro-apoptotic effects of IFN α + IL1 β . Furthermore, baricitinib decreased the cell surface protein expression of MHC class I by >90% in EndoC- β H1 cells (Fig. 9a) and human islets (Fig. 9b, c).

Discussion

We presently modeled the initial changes observed in the islets of Langerhans during T1D by performing an integrated multi-omics approach in EndoC- β H1 cells exposed to the early cytokine IFN α . The model was validated using human islets RNA-seq and independent experiments using the same human beta cell line, pancreatic human islets and FACS-purified human beta cells. Of relevance, taking into account the major differences between

human and rodent beta cell responses to stressful stimuli^{53,54}, all experiments were performed in clonal or primary human beta cells/islets. This approach identified very rapid and broad beta cell responses to IFN α including: (1) major early modifications in chromatin remodeling, which activates regulatory elements; (2) the key TFs regulating signaling, and the crosstalk between beta cells and immune cells; (3) the functional modules of genes and their regulatory networks; and (4) alternative splicing and first exon usage as important drivers of transcript diversity. Finally, an integrative analysis led to the identification of two compound classes that reverse all or part of these alterations in EndoC- β H1 cells and human islets and may be potential therapeutic targets for future trials in T1D prevention/treatment.

During viral infection a prompt innate immune response, mediated to a largest extent via type I interferons, is critical to control virus replication and spreading⁵⁵. In line with this, exposure of human beta cells to IFN α leads to changes in chromatin accessibility already at 2 h, which correlates with subsequent changes in mRNA and protein expression at 8 and 24 h. The majority of these regions are localized distally to TSSs, indicating that they may act primarily as distal regulatory elements. Interestingly, these regions were enriched in motifs of islets-specific TFs, suggesting that tissue-restricted characteristics regulate the local responses during insulinitis, as we have recently described for the cytokines IL1 β + IFN γ ¹⁹. This could explain the preferential expression of HLA class I (both the classical ABC members and the presently described inhibitory HLA-E) by pancreatic islets in comparison to the surrounding exocrine pancreas. Islet HLA class I overexpression is a key finding during T1D development⁵⁶, contributing for the recruitment of auto-reactive CD8⁺ T cells that selectively attack beta cells¹. IFN α also induces pathways involved in protein modification (ubiquitination, sumoylation, etc), degradation (proteasome, etc) and ER stress, which can generate neoantigens¹⁴.

The IRF and STAT family members are master TFs involved in IFN-I signaling². Viruses have developed several species-specific mechanisms to antagonize STAT1 and STAT2 activation⁵⁵. For instances, the NS5 protein of Zika virus degrades human but not mouse STAT2⁵⁷. In the present work, we confirmed the importance of both STAT1 and 2 for IFN α signaling in beta cells, and observed that their individual KD is compensated in most cases by the remaining member, as a possible backup mechanism to protect against pathogens⁵⁸. Interestingly, IRF1 seems to be a critical regulator of the IFN α -mediated “defense” responses in beta cells, including induction of checkpoint proteins such as PDL1 and HLA-E (present data), and the suppressors of cytokine signaling 1 and 3 (SOCS3) (ref. ⁵⁹ and present data). This stands in contrast to its pro-inflammatory effects in immune cells⁶⁰. In line with the possible role for IRF1 in dampening islet

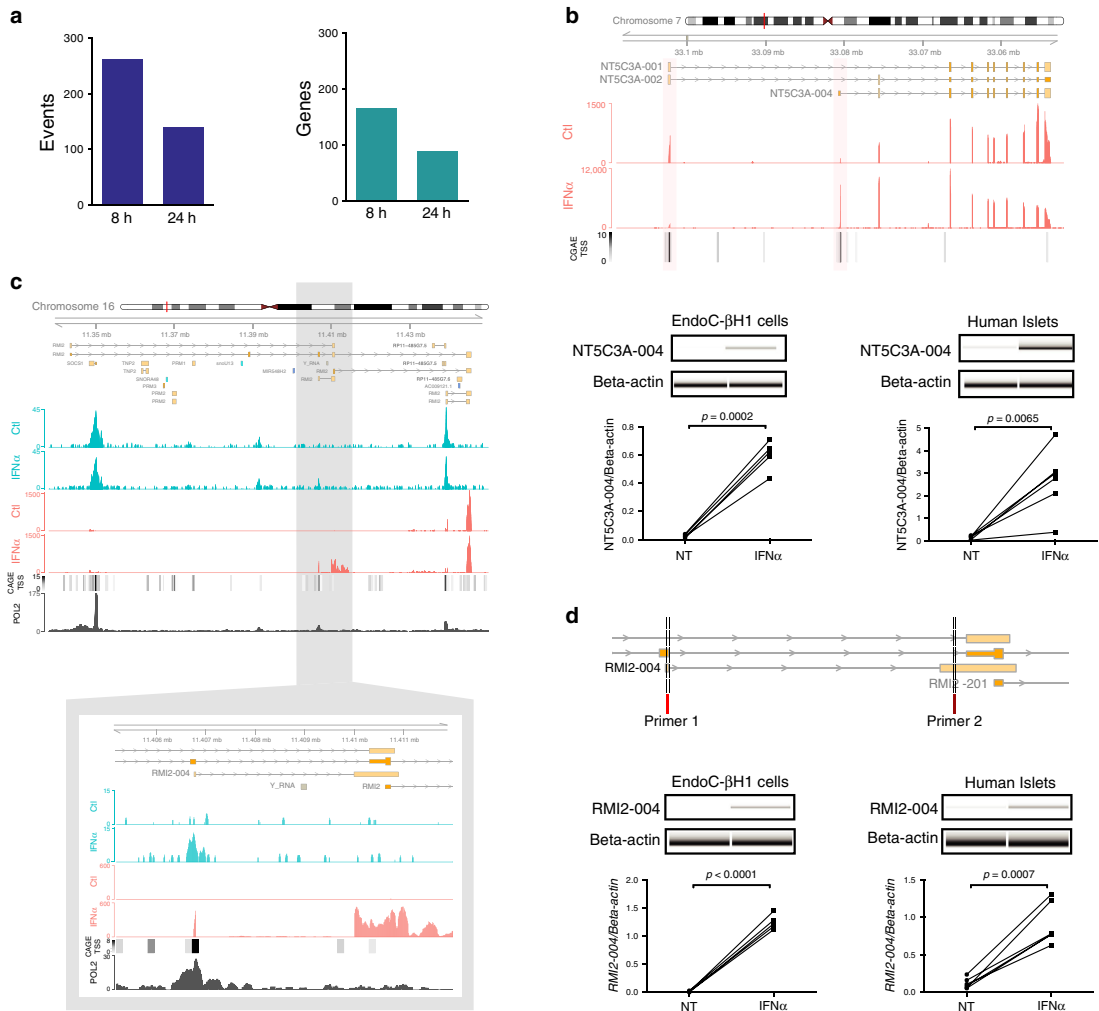


Fig. 6 Changes in the alternative transcription start site (TSS) initiation increase the repertoire of IFN α -regulated transcripts. **a** The tool SEASTAR⁴² was used to estimate the frequency of differential alternative first exon (AFE) usage induced by IFN α in human beta cells. The total number of IFN α -dependent AFEs events (left) and number of genes with AFEs (right) in the indicated time point are shown ($n = 5$, $\Delta\text{PSI} > [0.2]$, $\text{FDR} < 0.05$). **b** View of the NT5C3A locus showing the transcripts with AFE usage, the RNA-seq (red) signals of EndoC- β H1 cells exposed or not to IFN α and the CAGE TSSs information (black scale)⁴⁵ (upper panel). Confirmation of the AFE usage identified by SEASTAR in the gene NT5C3A (lower panel). cDNA was amplified by RT-PCR using primers located in the AFE and in its downstream exon. The PCR products were analyzed by automated electrophoresis using a Bioanalyzer 2100 and quantified by comparison with a loading control. The values were then corrected by the housekeeping gene β -actin. ($n = 4$ (EndoC) and $n = 6$ (human islets), two-sided paired t -test). **c** View of the RMI2 locus showing all the transcripts in this region, the ATAC-seq (blue) and the RNA-seq (red) signals of EndoC- β H1 cells exposed or not to IFN α for 24 h, the CAGE TSSs information (black scale)⁴⁵ and RNA polymerase II ChIP-seq signal of human K562 cells exposed to IFN α (black)⁴⁸. A higher magnification of the RMI2-004 locus is presented below (image representative of 4–5 independent experiments). **d** Confirmation of the AFE usage in the gene RMI2. Genome mapping (upper part) showing the genomic regions used to design-specific primers located in the AFE of the transcript RMI2-004 and in its downstream exon. The PCR product was analyzed by automated electrophoresis using a Bioanalyzer 2100 and quantified by comparison with a loading control. The values were then corrected by the housekeeping gene β -actin. ($n = 4$ (EndoC) and $n = 6$ (human islets), two-sided paired t -test). Source data are provided as a Source Data file.

inflammation, systemic knockout of IRF1 prevents autoimmune diabetes in NOD mice⁶¹, whereas IRF1 deletion in islets is associated with shorter mouse allograft graft function and survival⁶².

Alternative splicing (AS) is a species, tissue and context-specific post-transcriptional mechanism that expands the number of transcripts originated from the same gene thus increasing protein

diversity⁶³. Pancreatic beta cells share many characteristics with neuronal cells, including analogous signal transduction, developmental steps and splicing networks⁶⁴. Both T1D risk genes⁶⁵ and the cytokines IL1 β + IFN γ ¹⁰ modify AS in beta cells. We presently identified a preferential alternative exon inclusion after IFN α exposure and mapped the potentially involved RBPs, which included the upregulated protein Quaking (*QKI*). *QKI* activation

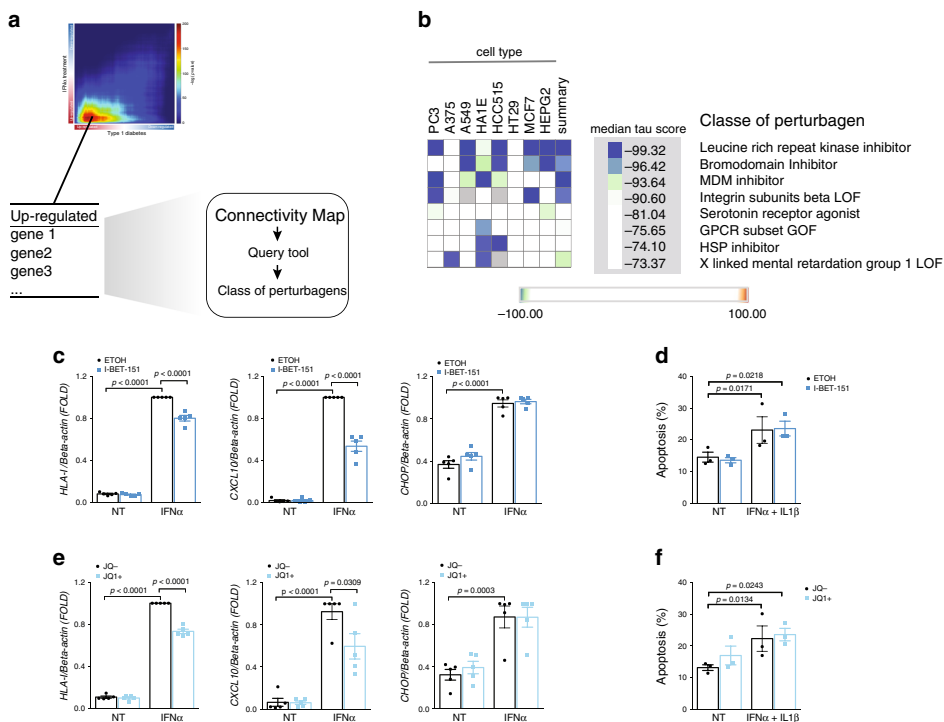


Fig. 7 Mining the type I interferon signature of pancreatic beta cells for identification of potentially T1D therapeutic targets. **a** The top 150 upregulated genes identified in Supplementary Fig. 2d were used to query the Connectivity MAP database of cellular signatures⁴⁹. **b** Connectivity map classes of perturbagens that promote an opposite signature to the one shared between beta cells of T1D individuals and EndoC-βH1 cells exposed to IFN α (Supplementary Fig. 2d). **c, e** EndoC-βH1 cells were pretreated for 2 h with the bromodomain inhibitors I-BET-151 (1 μ M) (**c**) or JQ1+ (0.4 μ M) (**e**) and then exposed to IFN α for 24 h. Cells were collected and the mRNA expression for HLA class I (ABC), the chemokine CXCL10 and the ER stress marker CHOP (DDIT3) evaluated. Ethanol (vehicle) and an inactive enantiomer (JQ1-) were used as respective controls for I-BET-151 and JQ1+. ($n = 5$, mean \pm SEM, ANOVA with Bonferroni correction for multiple comparisons). **d, f** Cell viability after exposure to the combination of cytokines IFN α (2000 U/ml) + IL1 β (50 U/ml) in the presence or not of the bromodomain inhibitors ($n = 3$, mean \pm SEM, ANOVA with Bonferroni correction for multiple comparisons).

in monocytes promotes extensive changes in AS, favoring their differentiation into pro-inflammatory macrophages⁶⁶. Furthermore, QKI binds to the genome of RNA viruses and inhibits their replication⁶⁷. A similar mechanism was recently described for *FMR1*⁶⁸, another RBP induced by IFN α , which controls protein translation in beta cell (present data). Several other RBPs were observed as downregulated by IFN α and identified as potential regulators of IFN α -induced AS events. Thus, inhibition of ELAVL1 and HNRNPA1 reproduced IFN α -mediated changes in exon usage. Different RNA viruses can use both ELAVL1⁶⁹ and HNRNPA1⁷⁰ to support their replication, indicating that the decreased expression of these proteins may provide an additional IFN-triggered antiviral mechanism. These findings suggest that during potentially diabetogenic viral infections, RBPs may have a dual role: first as splicing regulators and second as regulators of viral replication.

In order to identify novel approaches to protect beta cells in T1D, we analyzed the similarities between beta cell signatures from T1D donors and those following IFN α exposure, and compared the top identified genes/pathways with the Connectivity Map⁴⁹ and the DrugBank⁵¹ database. This identified two groups of potential therapeutic agents, namely bromodomain and JAK inhibitors. Bromodomain (BRD) proteins are components of chromatin-remodeling complexes that promote chromatin decompaction and transcriptional activation. BET inhibitors have

shown protective effects in different animal models of autoimmunity⁷¹, including the diabetes-prone NOD mice⁵⁰. We have now expanded these findings to human beta cells, showing that two distinctive BET inhibitors (JQ1+ and I-BET-151) decrease IFN α -induced responses, including HLA class I and chemokine overexpression.

After binding to its receptor, IFN α promotes phosphorylation of two tyrosine kinases, JAK1 and TYK2, which then trigger the downstream signaling cascade. Chemical inhibition of JAK1 + JAK2 prevents autoimmune diabetes in NOD mice⁷² and polymorphisms associated with decreased TYK2 function are protective against human T1D⁷³. We presently observed that baricitinib, a JAK1/2 inhibitor recently approved for use in rheumatoid arthritis by the FDA⁵², decreased all the three hallmarks previously identified in islets of T1D individuals and initiated by IFN α in human beta cells, namely HLA class I overexpression, ER stress and beta cell apoptosis, supporting its future testing in T1D.

In conclusion, we have applied a multi-omics approach to study the different levels of gene regulation induced by IFN α in EndoC-βH1 cells and pancreatic human islets. This in vitro modeling showed strong correlation with the mRNA profile from beta cells of T1D individuals. At the genomic level, early chromatin remodeling activated *cis*-regulatory elements, many of them presenting motifs for islets-specific TFs, providing a

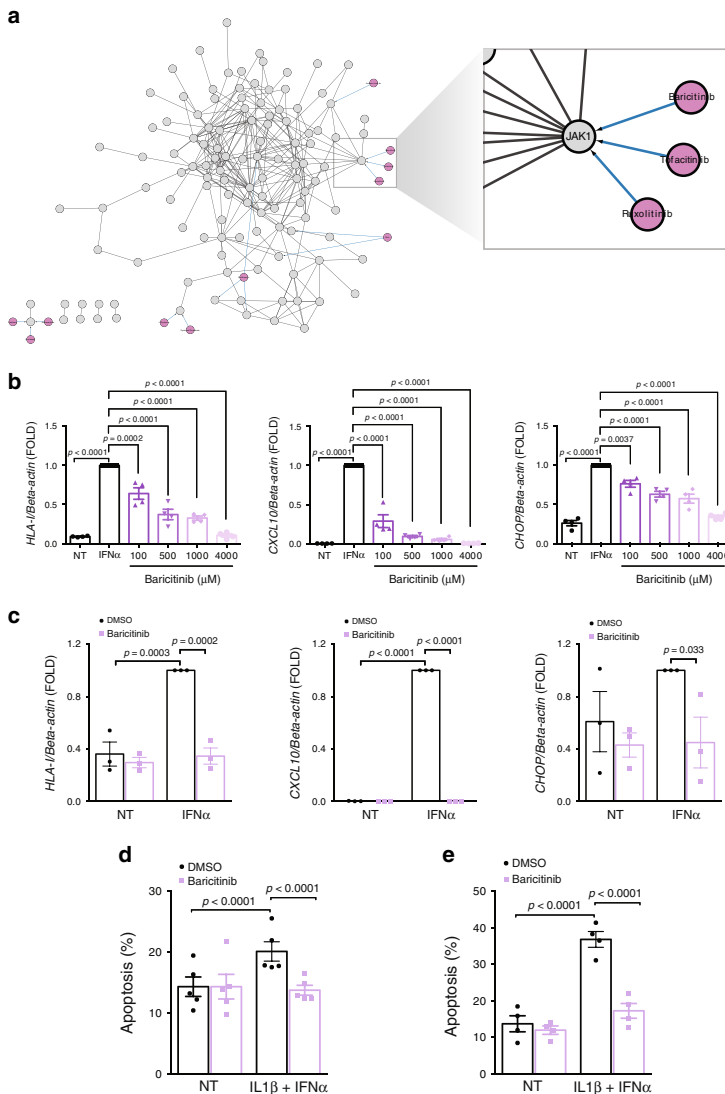


Fig. 8 Establishing JAK1 inhibition as protective mechanism against IFNα-mediated inflammation and apoptosis. **a** The PPI network of module #2 was integrated with the DrugBank repository⁵¹ using the CyTargetLinker app⁷⁸ in Cytoscape. A higher magnification on JAK1 is shown. **b** EndoC-βH1 cells were pretreated with DMSO (NT) or baricitinib at the indicated concentrations for 2 h. Cells were then left untreated (black bars), or treated with IFNα alone (white bars) without or with the presence of different concentrations of baricitinib (purple scale bars) for 24 h and mRNA expression of HLA class I (ABC), CXCL10 and CHOP (DDIT3) analyzed. The values were normalized by the housekeeping gene β-actin and then by the highest value of each experiment considered as 1 ($n = 4$, mean ± SEM, ANOVA with Bonferroni correction for multiple comparisons). **c** Human islets were pretreated with baricitinib (4 μM) or DMSO (vehicle) and then exposed or not to IFNα for 24 h in the presence or not of baricitinib. mRNA expression of HLA class I (ABC), CXCL10 and CHOP (DDIT3) was analyzed and values normalized by the housekeeping gene β-actin and then by the highest value of each experiment considered as 1. ($n = 3$, mean ± SEM, ANOVA with Bonferroni correction for multiple comparisons). **d**, **e** EndoC-βH1 cells (**d**) and human islets (**e**) were pretreated with DMSO or baricitinib (4 μM) for 2 h. Subsequently, cells were left untreated or treated with IFNα (2000 U/ml) + IL1β (50 U/ml) in the absence or presence of baricitinib for 24 h. Cell viability was evaluated using nuclear dyes by two independent observers. (**d** ($n = 5$), **e** ($n = 4$), mean ± SEM, ANOVA with Bonferroni correction for multiple comparisons). Source data are provided as a Source Data file.

possible mechanism by which tissue-restricted autoimmune diseases might arise. Post-translational modifications, alternative splicing and first exon usage were induced by IFNα, likely expanding the repertoire of proteins and transcripts generated by beta cells in response to this inflammatory stimuli. This can also

be a source of potential neoantigens. Interestingly, IFNα-exposed human beta cells upregulate co-inhibitory proteins such as PDL1 and HLA-E, which may attenuate or delay the autoimmune assault. Finally, the present results provide a useful resource for the discovery of compounds that may be used to reverse the

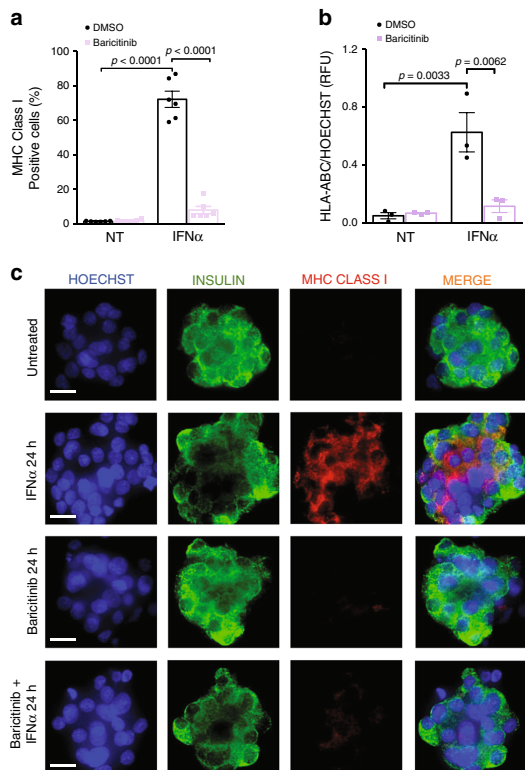


Fig. 9 Baricitinib decreases IFN α -mediated MHC class I protein expression in beta cells.

a EndoC- β H1 cells were pretreated with baricitinib (4 μ M) or DMSO and then exposed or not to IFN α for 24 h in the presence or not of baricitinib. MHC class I (ABC) protein expression was measured by flow cytometry. The percentage of positive cells was quantified. ($n = 6$, mean \pm SEM, ANOVA with Bonferroni correction for multiple comparisons). **b, c** Dispersed human islets were pretreated with baricitinib (4 μ M) or DMSO (vehicle). Next, cells were left untreated, treated with IFN α alone or with IFN α in the presence of baricitinib for 24 h. MHC class I intensity was quantified in each condition (**b**) using Fiji software⁸⁰ and normalized by the HOECHST intensity to correct for the number of cell per area ($n = 3$, ANOVA with Bonferroni correction for multiple comparisons, RFU (relative fluorescence units)). Immunocytochemistry (ICC) analysis (**c**) of MHC class I (ABC) (red), insulin (green) and HO (blue) was performed to confirm MHC class I expression in three independent human islet preparations. Scale bar 10 μ m.

effects of IFN α on human pancreatic beta cells, paving the way for potential T1D interventional trials.

Methods

Culture of EndoC- β H1 cells and human islets, cell treatment. The human pancreatic beta cell line EndoC- β H1 was kindly provided by Dr. R. Scharfmann, University of Paris, France⁷⁴. Human islet isolation from 20 non-diabetic organ donors (Supplementary Table 2) was performed in accordance with the local Ethical Committee in Pisa, Italy. The use of pancreatic human islets for this project was approved by the Comité d'Ethique hospitalo-facultaire Erasme-ULB. These cells were maintained in culture and treated as described in Supplementary Methods.

FACS-purified human beta cells isolation and treatment. Whole pancreatic human islets were exposed or not to IFN α for 24 h. After this period, the islets were dispersed into single cells and surface staining was carried out in FACS buffer (PBS with 0.5% BSA and EDTA 2 mM final concentration). Indirect antibody labeling

was performed with two sequential incubation at 4 $^{\circ}$ C and one wash in FACS buffer followed each step. Cells were resuspended in FACS buffer, viability dye was added (DAPI) and cells were sorted on a FACSARIA III cell sorter (BD Biosciences). Primary (mouse anti-human NTPDase3, hN3-B3S, www.econucleotidases-ab.com) and secondary (Alexa Fluor 546 conjugated donkey anti-mouse IgG (A10036, Thermo-Fisher Scientific)) antibodies were used with the dilutions described in Supplementary Table 5. Data analysis was carried out with FlowJo software (Version 10).

ATAC sequencing processing and analysis. ATAC sequencing was performed in four independent experiments for each time point (2 and 24 h)⁷⁵. For ATAC-seq 50,000 EndoC- β H1 cells were exposed or not to IFN α for 2 or 24 h. After that, the cells were harvested, and the nuclei isolated by using 300 μ l of cold lysis buffer (10 mM Tris-HCl pH 7.4, 10 mM NaCl, 3 mM MgCl₂, 0.1% Igepal CA-630). The nuclei pellet was resuspended in a 25 μ l transposase reaction mix containing 2 μ l of Tn5 transposase per reaction and incubated at 37 $^{\circ}$ C for 1 h. The tagmented DNA was isolated using SPRI cleanup beads (Agencourt AMPure XP, Beckman Coulter). For library amplification two sequential 9-cycle PCR were performed (72 $^{\circ}$ C for 5 min; 98 $^{\circ}$ C for 30 s; 9 cycles of 98 $^{\circ}$ C for 10 s, 63 $^{\circ}$ C for 30 s; and 72 $^{\circ}$ C for 1 min; and at 4 $^{\circ}$ C hold). Finally, the DNA library was purified using the MinElute PCR Purification Kit (Qiagen, Venlo, Netherlands). TapeStation and semi-quantitative PCR assays at target positive and negative controls were performed to ensure the quality and estimate the efficiency of the experiment before sequencing. Libraries were sequenced single-end on an Illumina HiSeq 2500. Data processing and analysis is described in Supplementary Methods.

RNA-sequencing processing and analysis. Total RNA of five independent experiments with EndoC- β H1 cells and six independent preparation of pancreatic human islets exposed or not to IFN α for different time points was obtained using the RNeasy Mini Kit (Qiagen, Venlo, Netherlands). RNA integrity number (RIN) values were evaluated using the 2100 Bioanalyzer System (Agilent Technologies, Wokingham, UK). All the samples analyzed had RIN values >9. mRNA was obtained from 500 ng of total RNA using oligo (dT)beads, before it was fragmented and randomly primed for reverse transcription followed by second-strand synthesis to generate double-stranded cDNA fragments. The cDNA undergone paired-end repair to convert overhangs into blunt ends. After 3'-monoadenylation and adaptor ligation, cDNAs were purified. Next, cDNA was amplified by PCR using primers specific for the ligated adaptors. (Illumina, Eindhoven, Netherlands). The generated libraries were submitted to quality control before being sequenced on an Illumina HiSeq 2500. RNA-seq data processing and analysis is described in Supplementary Methods.

Proteomics processing and analysis. EndoC- β H1 cells exposed or not to IFN α were extracted using the Metabolite, Protein and Lipid Extraction (MPLEX) approach. A detailed description of the method used for proteomics processing and analysis is provided in Supplementary Methods.

Rank-rank hypergeometric overlap (RRHO) analysis. To compare the signature induced by IFN α with the one present during insulinitis in T1D individuals, we performed the RRHO mapping¹⁵. For this goal, a full list of log₂FC ranked genes from our RNA-seq of EndoC- β H1 cells and human islets (IFN α vs Control, 24 h) were compared against similarly ranked lists of purified primary beta cells obtained from individuals with T1D¹⁶ and T2D¹⁷ (T1D/T2D vs non-diabetic).

In a RRHO map, the hypergeometric p -value for enrichment of k overlapping genes is calculated for all possible threshold pairs for each experiment, generating a matrix where the indices are the current rank in each experiment. The log-transformed hypergeometric p -values are then plotted in a heatmap indicating the degree of statistically significant overlap between the two ranked lists in that position of the map. Multiple correction was applied using the Benjamini-Yekutieli FDR correction.

Dynamic regulatory events miner (DREM) modeling. For reconstructing dynamic regulatory networks, we have used the DREM method²³, which integrates times series and static data using an Input-Output Hidden Markov Model (IOHMM), where the TF-DNA interaction information obtained from ChIP-seq experiments⁴⁸ was used as the input and our RNA-seq time series expression data as the output. A detailed description of DREM-based modeling is provided in Supplementary Methods.

Weighted gene co-expression network analysis (WGCNA). On each dataset (RNA-seq and proteomics), we obtained modules of genes/proteins of similar expression profiles using WGCNA²⁸. The soft threshold parameter for the RNA-seq dataset was chosen to be 10 (value to approximate a scale-free topology). Similar parameters were used for the analysis of RNA-seq of pancreatic human islets exposed or not to IFN α . Regarding the proteomics dataset, in order to achieve an approximated scale-free topology, we first normalized each protein expression in each temporal group (subtraction by mean and division by standard deviation), and then selected the soft threshold parameter as 14. After merging the modules

using a dissimilarity threshold of 0.25, we identified 8 modules in the RNA-seq dataset and 7 modules in the proteomics dataset.

To analyze module quality, we have used a set of statistics (density and separability metrics) from the *modulePreservation* function of the R package WGCNA²⁹. For this purpose, we resampled the dataset 1000-times to create reference and test sets from the original data and evaluate module preservation, represented as the Z_{summary} for each module across the resulting networks. $Z_{\text{summary}} > 2$ indicates moderate preservation and $Z > 10$ high quality/preservation for each module²⁹. To evaluate WGCNA module preservation in independent samples, we used the same R function, but in this case applying metrics based on module density and intramodular connectivity to give a composite statistic Z_{summary} .

To evaluate the overlap of RNA-seq and proteomics modules, we considered a mRNA to be differentially expressed at 8 or 24 h if its absolute fold-change was >1.5 and its FDR < 0.05 . Regarding the proteomics dataset, we considered a protein to be differentially abundant at 8 or 24 h if the *t*-test *p*-value was < 0.05 . We selected only the differentially expressed genes/abundant proteins in the identified WGCNA modules. We then searched for the overlap between the elements of the RNA-seq and proteomics modules and obtained an overlap *p*-value (hypergeometric probability). We retained overlapping modules with a FDR < 0.05 and a minimum of 10 common elements.

Protein-protein interaction network analysis. The inBio Map protein-protein interaction (PPI) network database³¹ was obtained from <https://www.intomics.com/inbio/>. We first restricted the network to contain only the elements expressed in human beta cells based on our RNA-seq database (mean RPKM > 0.5 in at least one condition). For each WGCNA overlapping module, we identified the proteins in the PPI network with a significantly high number of protein-to-protein connections to the set of elements in the module (FDR < 0.01 , and minimum number of connections equal to 5). We considered only networks obtained for the overlapping modules #1, #2 and #5, as the other overlapping modules returned empty PPI networks. We then obtained PPI networks for each WGCNA overlapping modules, involving the original set of module elements, plus the respective identified connecting proteins. Communities of interacting proteins were identified using the EAGLE algorithm⁷⁶ with the following parameters: CliqueSize threshold: 6 and ComplexSize threshold: 2. Wordclouds of each community were generated using information from geneRIFs terms.

Gene regulatory network analysis. A network of regulatory interactions was obtained from RegNetworks³⁰ (www.regnetworkweb.org). As in the PPI network, we first restricted the network to contain only the elements we found to be expressed in the RNA-seq dataset. Similarly to the PPI network analysis, for each WGCNA overlapping modules, we identified regulators with a significantly high number of regulatory connections to the set of elements in the module (FDR < 0.01 , and minimum number of connections equal to 4). We then obtained regulatory networks for each WGCNA overlapping modules, involving the original set of module elements, plus the respective identified regulators.

To create a non-redundant dataset of motifs from regions of gained open chromatin, we used the compareMotifs.pl script from the package HOMER⁷⁷ to merge motifs with a similarity score threshold of 0.7. The remaining motifs were mapped to the gain open chromatin regions using the annotatePeaks.pl script.

Transcription factor motif analysis. Sequence composition analysis of *de novo* motifs was performed using findMotifGenome.pl from the package HOMER⁷⁷ with parameters '-size given -bits -mask'. The motifs having a $p \leq 10^{-12}$ and observed in $>3.5\%$ of the targets were chosen for subsequent analysis. All *de novo* matches having a similarity score to known TF motifs higher than 0.7 are shown in the tables (Fig. 4e and Supplementary Fig. 3c), or when no match was present over this threshold, the first hit was elected and its score is presented.

Alternative splicing changes validation. Alternative splicing changes identified from RNA-seq were validated by RT-PCR using specifically designed primers (Supplementary Table 3). To confirm cassette exons, the primers were adjacent to the predicted splicing event. This approach allowed us to discriminate between variants based on their fragment sizes. For alternative first exon usage (AFE) validation, we have designed primers spanning regions that are unique to the isoform of interest (Fig. 6g), and then normalized the results by the housekeeping gene β -actin. cDNA was amplified using MyTaq Red DNA polymerase (Bioline, London, UK), and PCR products were analyzed using an Agilent 2100 Bioanalyzer system (Agilent Technologies, Wokingham, U.K.). The molarity of each PCR band corresponding to a specific splice variant was quantified using the 2100 Expert Software (Agilent Technologies, Diegem, Belgium), and used to calculate the ratio inclusion/exclusion (SE) or isoform- X/β -actin (AFE).

Small-RNA interference. Transfection was performed using Lipofectamine RNAiMAX (Invitrogen) as described in Supplementary Methods. After that, the cells were kept in culture for a 48 h recovery period and subsequently exposed or not to IFN α as indicated. Supplementary Table 3 describes the sequences of siRNAs used in the present study.

Real-time PCR analysis. After harvesting of the cells, Poly(A) + mRNA was obtained using the Dynabeads mRNA DIRECT kit (Invitrogen) and reverse transcribed. Detailed description is provided in Supplementary Methods.

Western blot, immunocytochemistry and flow cytometry. Detailed description together with additional information on western blot, immunocytochemistry and flow cytometry analysis is provided in Supplementary Methods.

Immunofluorescence. After dewaxing and rehydration, samples were subjected to heat-induced epitope retrieval (HIER) in 10 mM citrate buffer pH 6.0, then probed in a sequential manner with appropriate antibodies as indicated in Supplementary Table 4. The relevant antigen-antibody complexes were detected using secondary antibodies conjugated with fluorescent dyes (Invitrogen, Paisley, U.K.). Cell nuclei were stained with DAPI. After mounting, images were captured with a Leica AF6000 microscope (Leica, Milton Keynes, UK) and processed using the standard LASX Leica software platform (Version 1.9.013747). For quantification studies, randomly selected insulin-containing islets (ICIs) from individuals with or without diabetes were imaged, in addition to insulin-deficient islets (IDIs) from individuals with diabetes. Thirty ICIs were analyzed from 6 independent individuals (5 islets per individual), 20 IDIs were analyzed from 4 independent individuals (5 islets per individual) and 30 ICIs were analyzed from 6 independent control individuals (5 islets per individual). The mean fluorescence intensity (MFI) arising from detection of HLA-E was measured using LASX Leica quantification software.

Therapeutic targets identification. The top 150 upregulated genes shared among the RNA-seq of EndoC- β H1 cells and human islets exposed to IFN α for 24 h and the RNA-seq of beta cells¹⁶ from T1D individuals were identified by the RRHO analysis. This list of genes was used to query the Connectivity Map dataset of L1000 cellular signatures, which has transcriptional responses of human cells to different chemical and genetic perturbations, using the CLUE platform (<https://clue.io>)⁴⁹. To identify compounds potentially reverting the effects induced by interferons in beta cells, we have focused on perturbagens promoting signatures that were opposite (negative tau score) to our query list. Only perturbagens having a median tau score < -90 were considered for further evaluation.

Additionally, aiming at potential repurposing of drugs under clinical investigation for treatment of other pathologies, we have integrated the DrugBank database v5.1⁵¹ with the PPI network obtained from WGCNA module #2 using the CyTargetLinker v4.0.0⁷⁸ within Cytoscape v3.6 to build a biological network annotated with drugs.

The small molecules and drugs pointed out by these two approaches were then validated *in vitro* as described above to verify their impact on IFN α -induced upregulation of cytokines/chemokines, ER stress markers, HLA class I and beta cell apoptosis.

Cell viability assessment. The cell viability is described in details in Supplementary Methods.

Statistical analysis. Data of the confirmatory experiments are expressed as means \pm SEM. A significant difference between experimental conditions was assessed by paired *t*-test, unpaired *t*-test, one-way or two-ways ANOVA followed by Bonferroni correction for multiple comparisons as indicated using the GraphPad Prism program version 6.0 (www.graphpad.com). Results with $p \leq 0.05$ were considered statistically significant.

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

Data availability

All raw and processed ATAC and RNA-sequencing data that support the findings of this study have been deposited in NCBI Gene Expression Omnibus (GEO) with the primary accession code GSE133221 (subseries are GSE133218: RNA-seq of EndoC- β H1 cells, GSE148058: RNA-seq of human islets, GSE133219: ATAC-seq of EndoC- β H1 cells). The proteomics datasets have been submitted to Pride under identifier number PXD014244 (<http://www.ebi.ac.uk/pride/archive/projects/PXD014244>). The network of regulatory interactions can be obtained from RegNetworks (<http://www.regnetworkweb.org/download/RegulatoryDirections.zip>). The DrugBank database v5.1 can be downloaded from: <https://www.drugbank.ca/releases/5-1-0/downloads/all-full-database>. The inBio Map protein-protein interaction (PPI) network database can be obtained from: http://www.intomics.com/inbio/api/data/map_public/2016_09_12/inBio_Map_core_2016_09_12.zip. The CAGE peaks from FANTOM5 database can be obtained on: http://fantom.gsc.riken.jp/5/databases/phase2.5/extra/CAGE_peaks/. The Connectivity Map database can be accessed using the CLUE platform (<https://clue.io>). The RNA polymerase II (POLR2A) ChIP-seq of human K562 cells can be obtained from the ENCODE project (GSM935474, <https://www.encodeproject.org/experiments/ENCSCR000FAX/>). The Exon Ontology database can be accessed from: <http://fasterdb.ens-lyon.fr/ExonOntology/>. The information about T1D risk genes can be found on immunobase (www.immunobase.org) and GWAS catalog (<https://www.ebi.ac.uk/gwas/>).

The source data underlying Figs. 2c–m, 3a–e, g, i, 5d, g, j, 6b, d, 7c–f, 8b–e, 9a, b and Supplementary Figs. 4e–h, 5c–m, 6b, d, 12a–c, 12e–g, 13c–d are provided as a Source data file.

Received: 15 August 2019; Accepted: 23 April 2020;

Published online: 22 May 2020

References

- Eizirik, D. L., Colli, M. L. & Ortis, F. The role of inflammation in insulinitis and beta-cell loss in type 1 diabetes. *Nat. Rev. Endocrinol.* **5**, 219–226 (2009).
- Crow, M. K., Olleriev, M. & Kirou, K. A. Type 1 interferons in autoimmune disease. *Annu Rev. Pathol.* **14**, 369–393 (2019).
- Eizirik, D. L. & Op de Beeck, A. Coxsackievirus and type 1 diabetes mellitus: the Wolf's footprints. *Trends Endocrinol. Metab.* **29**, 137–139 (2018).
- Ferreira, R. C. et al. A type 1 interferon transcriptional signature precedes autoimmunity in children genetically at risk for type 1 diabetes. *Diabetes* **63**, 2538–2550 (2014).
- Foulis, A. K., Farquharson, M. A. & Meager, A. Immunoreactive alpha-interferon in insulin-secreting beta cells in type 1 diabetes mellitus. *Lancet* **2**, 1423–1427 (1987).
- Meyer, S. et al. AIRE-deficient patients harbor unique high-affinity disease-ameliorating autoantibodies. *Cell* **166**, 582–595 (2016).
- Marroqui, L. et al. Interferon-alpha mediates human beta cell HLA class I overexpression, endoplasmic reticulum stress and apoptosis, three hallmarks of early human type 1 diabetes. *Diabetologia* **60**, 656–667 (2017).
- Cooper, N. J. et al. Type 1 diabetes genome-wide association analysis with imputation identifies five new risk regions. Preprint at <https://doi.org/10.1101/120022> (2017).
- Onengut-Gumuscu, S. et al. Fine mapping of type 1 diabetes susceptibility loci and evidence for colocalization of causal variants with lymphoid gene enhancers. *Nat. Genet.* **47**, 381–386 (2015).
- Eizirik, D. L. et al. The human pancreatic islet transcriptome: expression of candidate genes for type 1 diabetes and the impact of pro-inflammatory cytokines. *PLoS Genet.* **8**, e1002552 (2012).
- Op de Beeck, A. & Eizirik, D. L. Viral infections in type 1 diabetes mellitus—why the beta cells? *Nat. Rev. Endocrinol.* **12**, 263–273 (2016).
- Schreiber, G. The molecular basis for differential type I interferon signaling. *J. Biol. Chem.* **292**, 7285–7294 (2017).
- Huang, X. et al. Interferon expression in the pancreases of patients with type 1 diabetes. *Diabetes* **44**, 658–664 (1995).
- Gonzalez-Duque, S. et al. Conventional and neo-antigenic peptides presented by beta cells are targeted by circulating naive CD8+ T cells in type 1 diabetic and healthy donors. *Cell Metab.* **28**, 946–960.e946 (2018).
- Plaisier, S. B., Taschereau, R., Wong, J. A. & Graeber, T. G. Rank-rank hypergeometric overlap: identification of statistically significant overlap between gene-expression signatures. *Nucleic Acids Res.* **38**, e169 (2010).
- Russell, M. A. et al. HLA class II antigen processing and presentation pathway components demonstrated by transcriptome and protein analyses of islet beta-cells from donors with type 1 diabetes. *Diabetes* **68**, 988–1001 (2019).
- Xin, Y. et al. RNA sequencing of single human islet cells reveals type 2 diabetes genes. *Cell Metab.* **24**, 608–615 (2016).
- Cnop, M. et al. Mechanisms of pancreatic beta-cell death in type 1 and type 2 diabetes: many differences, few similarities. *Diabetes* **54**, S97–S107 (2005).
- Ramos-Rodriguez, M. et al. The impact of proinflammatory cytokines on the beta-cell regulatory landscape provides insights into the genetics of type 1 diabetes. *Nat. Genet.* **51**, 1588–1595 (2019).
- Eizirik, D. L. & Cnop, M. ER stress in pancreatic beta cells: the thin red line between adaptation and failure. *Sci. Signal* **3**, pe7 (2010).
- Gurzov, E. N. et al. Pancreatic beta-cells activate a JunB/ATF3-dependent survival pathway during inflammation. *Oncogene* **31**, 1723–1732 (2012).
- Rondas, D. et al. Citrullinated glucose-regulated protein 78 is an autoantigen in type 1 diabetes. *Diabetes* **64**, 573–586 (2015).
- Ernst, J., Vainas, O., Harbison, C. T., Simon, I. & Bar-Joseph, Z. Reconstructing dynamic regulatory maps. *Mol. Syst. Biol.* **3**, 74 (2007).
- Colli, M. L. et al. PDL1 is expressed in the islets of people with type 1 diabetes and is up-regulated by interferons-alpha and-gamma via IRF1 induction. *EBioMedicine* **36**, 367–375 (2018).
- Andre, P. et al. Anti-NKG2A mAb is a checkpoint inhibitor that promotes anti-tumor immunity by unleashing both T and NK cells. *Cell* **175**, 1731–1743 e1713 (2018).
- Couplé, S. et al. Expression and release of soluble HLA-E is an immunoregulatory feature of endothelial cell activation. *Blood* **109**, 2806–2814 (2007).
- Lundberg, M., Krogvold, L., Kuric, E., Dahl-Jorgensen, K. & Skog, O. Expression of interferon-stimulated genes in insulinitic pancreatic islets of patients recently diagnosed with type 1 diabetes. *Diabetes* **65**, 3104–3110 (2016).
- Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559 (2008).
- Langfelder, P., Luo, R., Oldham, M. C. & Horvath, S. Is my network module preserved and reproducible? *PLoS Comput. Biol.* **7**, e1001057 (2011).
- Liu, Z. P., Wu, C., Miao, H. & Wu, H. RegNetwork: an integrated database of transcriptional and post-transcriptional regulatory networks in human and mouse. *Database (Oxford)* **2015**, <https://doi.org/10.1093/database/bav095> (2015).
- Li, T. et al. A scored human protein-protein interaction network to catalyze genomic interpretation. *Nat. Methods* **14**, 61–64 (2017).
- Choi, U. Y., Kang, J. S., Hwang, Y. S. & Kim, Y. J. Oligoadenylate synthase-like (OASL) proteins: dual functions and associations with diseases. *Exp. Mol. Med.* **47**, e144 (2015).
- Marques, J. et al. The p59 oligoadenylate synthetase-like protein possesses antiviral activity that requires the C-terminal ubiquitin-like domain. *J. Gen. Virol.* **89**, 2767–2772 (2008).
- Jacob, A. G. & Smith, C. W. J. Intron retention as a component of regulated gene expression programs. *Hum. Genet.* **136**, 1043–1057 (2017).
- Ray, D. et al. A compendium of RNA-binding motifs for decoding gene regulation. *Nature* **499**, 172–177 (2013).
- Mukherjee, N. et al. Integrative regulatory mapping indicates that the RNA-binding protein HuR couples pre-mRNA processing and mRNA stability. *Mol. Cell* **43**, 327–339 (2011).
- Bruun, G. H. et al. Global identification of hnRNP A1 binding sites for SSO-based splicing modulation. *BMC Biol.* **14**, 54 (2016).
- Meertens, L. et al. FHL1 is a major host factor for chikungunya virus infection. *Nature* **574**, 259–263 (2019).
- Maurin, T. & Bardoni, B. Fragile X mental retardation protein: to be or not to be a translational enhancer. *Front. Mol. Biosci.* **5**, 113 (2018).
- Pasciuto, E. & Bagni, C. SnapShot: FMRP mRNA targets and diseases. *Cell* **158**, 1446–1446.e1441 (2014).
- Alasoo, K. et al. Genetic effects on promoter usage are highly context-specific and contribute to complex traits. *Elife* **8**, e41673 (2019).
- Qin, Z., Stoilov, P., Zhang, X. & Xing, Y. SEASTAR: systematic evaluation of alternative transcription start sites in RNA. *Nucleic Acids Res.* **46**, e45 (2018).
- Gel, B. et al. regionR: an R/Bioconductor package for the association analysis of genomic regions based on permutation tests. *Bioinformatics* **32**, 289–291 (2016).
- Al-Haj, L. & Khabar, K. S. A. The intracellular pyrimidine 5'-nucleotidase NT5C3A is a negative epigenetic factor in interferon and cytokine signaling. *Sci. Signal* **11**, eaal2434 (2018).
- Consortium, F. et al. A promoter-level mammalian expression atlas. *Nature* **507**, 462–470 (2014).
- Tranchevent, L. C. et al. Identification of protein features encoded by alternative exons using exon ontology. *Genome Res* **27**, 1087–1097 (2017).
- Inshaw, J. R. J., Cutler, A. J., Burren, O. S., Stefana, M. I. & Todd, J. A. Approaches and advances in the genetic causes of autoimmune disease and their implications. *Nat. Immunol.* **19**, 674–684 (2018).
- Consortium, E. P. et al. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* **447**, 799–816 (2007).
- Subramanian, A. et al. A next generation connectivity map: L1000 platform and the first 1,000,000 profiles. *Cell* **171**, 1437–1452.e1417 (2017).
- Fu, W. et al. Epigenetic modulation of type-1 diabetes via a dual effect on pancreatic macrophages and beta cells. *Elife* **3**, e04631 (2014).
- Wishart, D. S. et al. DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res.* **46**, D1074–D1082 (2018).
- Genovese, M. C. et al. Baricitinib in patients with refractory rheumatoid arthritis. *N. Engl. J. Med.* **374**, 1243–1252 (2016).
- Eizirik, D. L. et al. Major species differences between humans and rodents in the susceptibility to pancreatic beta-cell injury. *Proc. Natl. Acad. Sci. USA* **91**, 9253–9256 (1994).
- Brozzi, F. et al. Cytokines induce endoplasmic reticulum stress in human, rat and mouse beta cells via different mechanisms. *Diabetologia* **58**, 2307–2316 (2015).
- Mesev, E. V., LeDesma, R. A. & Ploss, A. Decoding type I and III interferon signalling during viral infection. *Nat. Microbiol.* **4**, 914–924 (2019).
- Richardson, S. J. et al. Islet cell hyperexpression of HLA class I antigens: a defining feature in type 1 diabetes. *Diabetologia* **59**, 2448–2458 (2016).
- Grant, A. et al. Zika virus targets human STAT2 to inhibit type 1 interferon signaling. *Cell Host Microbe* **19**, 882–890 (2016).
- Blaszczyk, K. et al. The unique role of STAT2 in constitutive and IFN-induced transcription and antiviral responses. *Cytokine Growth Factor Rev.* **29**, 71–81 (2016).
- Moore, F. et al. STAT1 is a master regulator of pancreatic {beta}-cell apoptosis and islet inflammation. *J. Biol. Chem.* **286**, 929–941 (2011).

60. Langlais, D., Barreiro, L. B. & Gros, P. The macrophage IRF8/IRF1 regulome is required for protection against infections and is associated with chronic inflammation. *J. Exp. Med.* **213**, 585–603 (2016).
61. Nakazawa, T. et al. Complete suppression of insulinitis and diabetes in NOD mice lacking interferon regulatory factor-1. *J. Autoimmun.* **17**, 119–125 (2001).
62. Gysemans, C. et al. Interferon regulatory factor-1 is a key transcription factor in murine beta cells under immune attack. *Diabetologia* **52**, 2374–2384 (2009).
63. Alvelos, M. I., Juan-Mateu, J., Colli, M. L., Turatsinze, J. V. & Eizirik, D. L. When one becomes many—alternative splicing in beta-cell function and failure. *Diabetes Obes. Metab.* **20**, 77–87 (2018).
64. Juan-Mateu, J. et al. Neuron-enriched RNA-binding proteins regulate pancreatic beta cell function and survival. *J. Biol. Chem.* **292**, 3466–3480 (2017).
65. Nogueira, T. C. et al. GLIS3, a susceptibility gene for type 1 and type 2 diabetes, modulates pancreatic beta cell apoptosis via regulation of a splice variant of the BH3-only protein Bim. *PLoS Genet.* **9**, e1003532 (2013).
66. de Bruin, R. G. et al. Quaking promotes monocyte differentiation into pro-atherogenic macrophages by controlling pre-mRNA splicing and gene expression. *Nat. Commun.* **7**, 10846 (2016).
67. Liao, K. C. et al. Identification and characterization of host proteins bound to dengue virus 3′ UTR reveal an antiviral role for quaking proteins. *RNA* **24**, 803–814 (2018).
68. Soto-Acosta, R. et al. Fragile X mental retardation protein is a Zika virus restriction factor that is antagonized by subgenomic flaviviral RNA. *Elife* **7**, e39023 (2018).
69. Sokolowski, K. J. et al. Sindbis virus usurps the cellular HuR protein to stabilize its transcripts and promote productive infections in mammalian and mosquito cells. *Cell Host Microbe* **8**, 196–207 (2010).
70. Lin, J. Y. et al. hnRNP A1 interacts with the 5′ untranslated regions of enterovirus 71 and Sindbis virus RNA and is required for viral replication. *J. Virol.* **83**, 6106–6114 (2009).
71. Mele, D. A. et al. BET bromodomain inhibition suppresses TH17-mediated pathology. *J. Exp. Med.* **210**, 2181–2190 (2013).
72. Trivedi, P. M. et al. Repurposed JAK1/JAK2 inhibitor reverses established autoimmune insulinitis in NOD mice. *Diabetes* **66**, 1650–1660 (2017).
73. Marroqui, L. et al. TYK2, a candidate gene for type 1 diabetes, modulates apoptosis and the innate immune response in human pancreatic beta-cells. *Diabetes* **64**, 3808–3817 (2015).
74. Ravassard, P. et al. A genetically engineered human pancreatic beta cell line exhibiting glucose-inducible insulin secretion. *J. Clin. Invest.* **121**, 3589–3597 (2011).
75. 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).
76. Shen, H., Cheng, X., Cai, K. & Hu, M.-B. Detect overlapping and hierarchical community structure in networks. *Phys. A Stat. Mech. Appl.* **388**, 1706–1712 (2009).
77. Heinz, S. et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).
78. Kutmon, M., Kelder, T., Mandaviya, P., Evelo, C. T. & Coort, S. L. CyTargetLinker: a cytoscape app to integrate regulatory interactions in network analysis. *PLoS ONE* **8**, e82160 (2013).
79. Tan, G. JASPAR2016: data package for JASPAR 2016. R package version 1.12.0. <http://jaspar.genereg.net/> (2019).
80. Schindelin, J. et al. Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682 (2012).

Acknowledgements

We are grateful to Dr. Andrea A. Schiavo for help in the implementation of the protocol for FACS-purified human beta cells isolation and Isabelle Millard, Anyishai Musaya, Nathalie Pachera and Michael Pangerl of the ULB Center for Diabetes Research for excellent technical support. This work was supported by grants from the Fonds National de la Recherche Scientifique (FNRS), Welbio CR-2015A-06 and CR-2019C-04, Belgium; the Horizon 2020 Program, T2Systems (GA667191); the National Institutes of Health,

NIH-NIDDK-HIRN Consortium 1UC4DK104166-01 and Innovate2CureType1 - Dutch Diabetes Research Foundation (DDRF). D.L.E. and P.M. have received funding from the Innovative Medicines Initiative 2 Joint Undertaking under Grant Agreement No. 115797 (INNODIA). This Joint Undertaking receives support from the Union's Horizon 2020 research and innovation programme and "EPPIA", "JDRF" and "The Leona M. and Harry B. Helmsley Charitable Trust". The DiViD study is funded by The South-Eastern Norway Regional Health Authority (grant to K.D.-J.), the Novo Nordisk Foundation (grant to K.D.-J.), and through the PEVNET (Persistent Virus Infection in Diabetes Network) Study Group funded by the European Union's Seventh Framework Programme (FP7/2007-2013) under grant Agreement Number 261441 PEVNET. Additional support was from a JDRF Career Development Award (5-CDA-2014-221-A-N) to S.J.R., a JDRF research grant awarded to the network of Pancreatic Organ Donors – Virus (nPOD-V) consortium (JDRF 25-2012-516); an MRC Project Grant MR/P010695/1 awarded to S.J.R. and N.G.M., a studentship grant from the Norman Family Trust (to S.J.R. and N.G.M.), a Spanish Ministry of Economy and Competitiveness (SAF2017-86242-R to L.P.) and Marató TV3 (201624.10 to L.P.). L.P. is a recipient of a Ramon y Cajal contract from the Spanish Ministry of Economy and Competitiveness (RYC-2013-12864) and M.R. is supported by an FI Agencia de Gestió d'Ajuts Universitaris i de Recerca (AGAUR) PhD fellowship. Part of the work was performed in the Environmental Molecular Sciences Laboratory, a U.S. Department of Energy (DOE) national scientific user facility at Pacific Northwest National Laboratory (PNNL) in Richland, WA. Battelle operates PNNL for the DOE under contract DE-AC05-76RLO01830.

Author contributions

M.L.C., L.P., T.O.M. and D.L.E. conceived, designed and supervised the experiments. M.L.C., M.I.A., J.L.E.H., A.C.B., A.C., H.R.V., E.S.N., N.G.M. and S.J.R. performed and analyzed the experiments. L.K., K.D.-J., M.L., M.A.R., J.J.-M., L.M. and P.M. contributed with material and reagents. M.L.C., M.R.R., B.-J.M.W.-R. and J.-V.T. performed bioinformatic analyses. M.L.C. and D.L.E. wrote the manuscript, and all authors revised it. M.L.C. and D.L.E. are the guarantors of this work and, as such, have full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41467-020-16327-0>.

Correspondence and requests for materials should be addressed to M.L.C.

Peer review information *Nature Communications* thanks David Blodgett and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permission information is available at <http://www.nature.com/reprints>

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



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020

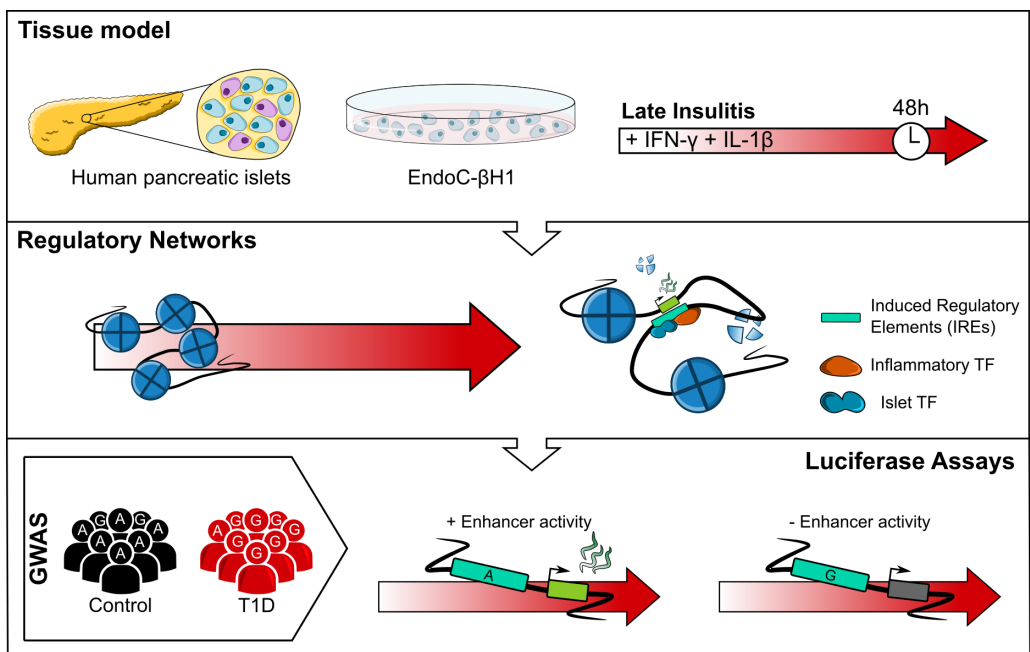
7

Interferon- γ and Interleukin-1 β to model human β cell responses to late insulinitis

In this study we sought to explore the role of β cells in late stages of the T1D-induced inflammation, specifically during the amplification of the immune response. To mimic these conditions, we exposed both a human β cell line (EndoC- β H1) and human pancreatic islets to proinflammatory cytokines IFN- γ and IL-1 β for 48 hours. We then performed several assays to reconstruct chromatin structure, gene expression and protein abundance in both conditions.

Ramos-Rodríguez, M., Raurell-Vila, H., Colli, M.L. et al. The impact of proinflammatory cytokines on the β -cell regulatory landscape provides insights into the genetics of type 1 diabetes. *Nat Genet.* **51**, 1588–1595 (2019). <https://doi.org/10.1038/s41588-019-0524-6>





7.1 Graphical abstract



7.2 Highlights

- Using the cytokine-induced gain of H3K27ac at ATAC-seq derived open chromatin regions, we identified 3,798 **Induced Regulatory Elements** (IREs). IREs present enhancer characteristics, such as distal location from gene promoters, phylogenetic sequence conservation and enrichment for TF binding motifs.
- Over 3,000 genes and 223 proteins were up-regulated after the treatment, showing a correlation between protein and the corresponding mRNA levels. Moreover, induction of gene expression and protein abundance was associated with the presence of IREs in proximity of the coding gene location.
- IREs were classified into two different groups according to the exhibited chromatin accessibility dynamics (see **Figure 2a** in Ramos-Rodríguez et al. (2019)):
 - **Opening IREs**, which showed increased accessibility coupled with enrichment in the H3K27ac histone modification. They were further classified into **neo IREs**, consisting of those regions that were completely closed prior to the cytokine exposure.
 - **Primed IREs**, which showed stable chromatin accessibility together with an increase in H3K27ac deposition.
- Both neo and primed enhancers were found to be occupied by inflammatory TFs, but only primed enhancers were also co-bound by islet-specific TFs. Additionally, neo enhancers were enriched in methylated CpG sites that underwent a demethylation process after the exposure.
- Chromatin capture experiments (UMI-4C, Schwartzman et al. (2016)) uncovered cytokine-induced changes in the 3D chromatin structure, including formation of enhancer-promoter interactions involving IREs and up-regulated genes.
- T1D but not T2D risk SNPs were enriched at both islet and β -cell IREs, suggesting a role for these regulatory regions in the development of T1D.
- *In vitro* testing of two T1D risk SNPs – rs78037977 and rs183778 – demonstrated functional properties of the allele variants in modulating IREs enhancer activity.

The impact of proinflammatory cytokines on the β -cell regulatory landscape provides insights into the genetics of type 1 diabetes

Mireia Ramos-Rodríguez¹ , Helena Raurell-Vila¹, Maikel L. Colli², Maria Inês Alvelos², Marc Subirana-Granés¹, Jonàs Juan-Mateu², Richard Norris¹, Jean-Valery Turatsinze², Ernesto S. Nakayasu³ , Bobbie-Jo M. Webb-Robertson³, Jamie R. J. Inshaw⁴, Piero Marchetti⁵, Lorenzo Piemonti⁶ , Manel Esteller^{7,8,9,10}, John A. Todd⁴, Thomas O. Metz³, Décio L. Eizirik² and Lorenzo Pasquali^{1,7,11*} 

The early stages of type 1 diabetes (T1D) are characterized by local autoimmune inflammation and progressive loss of insulin-producing pancreatic β cells. Here we show that exposure to proinflammatory cytokines reveals a marked plasticity of the β -cell regulatory landscape. We expand the repertoire of human islet regulatory elements by mapping stimulus-responsive enhancers linked to changes in the β -cell transcriptome, proteome and three-dimensional chromatin structure. Our data indicate that the β -cell response to cytokines is mediated by the induction of new regulatory regions as well as the activation of primed regulatory elements prebound by islet-specific transcription factors. We find that T1D-associated loci are enriched with newly mapped *cis*-regulatory regions and identify T1D-associated variants disrupting cytokine-responsive enhancer activity in human β cells. Our study illustrates how β cells respond to a proinflammatory environment and implicate a role for stimulus response islet enhancers in T1D.

In T1D, early inflammation of the pancreatic islets (insulinitis) by T and B cells contributes to both the primary induction and secondary amplification of the immune assault, with inflammatory mediators such as the cytokines interleukin-1 β (IL-1 β) and interferon- γ (IFN- γ) contributing to the functional suppression and apoptosis of β cells^{1–3}.

Genome-wide association studies (GWAS) have made a substantial contribution to the knowledge of T1D genetic architecture, uncovering >60 regions containing thousands of associated genetic variants. Nevertheless, translating variants to function is a main challenge for T1D and other complex diseases. Most of the associated variants do not reside in coding regions⁴, suggesting that they may influence transcript regulation rather than altering protein coding sequences. Recent studies showed a primary enrichment of T1D association signals in T- and B-cell enhancers^{4,5}. A secondary⁵, or a lack of enrichment, was instead observed in islet regulatory regions. While such observation points to a major role of the immune system, we hypothesize that a subset of T1D variants may also act at the β -cell level but only manifest on islet cell perturbation and are thus not captured by the current maps of islet regulatory elements.

We have now mapped inflammation-induced *cis*-regulatory networks, transcripts, proteins and three-dimensional (3D) chromatin structure changes in human β cells (Fig. 1a). We leverage these data

to reveal functional T1D genetic variants as well as key candidate genes and regulatory pathways contributing to β -cell autoimmune destruction. Such analyses permit elucidation of the role of gene regulation and its interaction with T1D genetics in the context of the autoimmune reaction that drives β -cell death.

Results

Proinflammatory cytokines impact the β -cell chromatin landscape. To characterize the effect of proinflammatory cytokines on the β -cell regulatory landscape, we first mapped all accessible or open chromatin sites in human pancreatic islets exposed or unexposed to IFN- γ and IL-1 β . We assayed chromatin accessibility by assay for transposase-accessible chromatin using sequencing (ATAC-seq) and, to focus on the β -cell fraction and decrease interindividual variability, in parallel with human pancreatic islet assays, we performed ATAC-seq in the clonal human β -cell line EndoC- β H1 (EC)⁶, exposed or unexposed to the proinflammatory cytokines (overall number of peaks identified in human islets: 92,610–229,588; in EC cells: 52,735–110,715; see Extended Data Fig. 1a). Such experiments revealed an important remodeling of β -cell chromatin resulting in approximately 12,500 highly confident chromatin sites that gained accessibility (false discovery rate (FDR)-adjusted $P < 0.05$; \log_2 fold change > 1; Extended Data Fig. 1b) on exposure to proinflammatory cytokines. Importantly, the changes

¹Endocrine Regulatory Genomics Laboratory, Germans Trias i Pujol University Hospital and Research Institute, Badalona, Spain. ²Center for Diabetes Research and Welbio, Medical Faculty, Université Libre de Bruxelles, Brussels, Belgium. ³Biological Sciences Division, Pacific Northwest National Laboratory, Richland, WA, USA. ⁴JDRF/Wellcome Diabetes and Inflammation Laboratory, Wellcome Centre for Human Genetics, Nuffield Department of Medicine, NIHR Oxford Biomedical Research Centre, University of Oxford, Oxford, UK. ⁵Department of Clinical and Experimental Medicine, University of Pisa, Pisa, Italy. ⁶Diabetes Research Institute, San Raffaele Scientific Institute, Milan, Italy. ⁷Josep Carreras Leukaemia Research Institute, Barcelona, Spain. ⁸Centro de Investigación Biomédica en Red Cáncer, Madrid, Spain. ⁹Institució Catalana de Recerca i Estudis Avançats, Barcelona, Spain. ¹⁰Physiological Sciences Department, School of Medicine and Health Sciences, University of Barcelona, Barcelona, Spain. ¹¹CIBER de Diabetes y Enfermedades Metabólicas Asociadas, Barcelona, Spain. *e-mail: lpasquali@igtp.cat

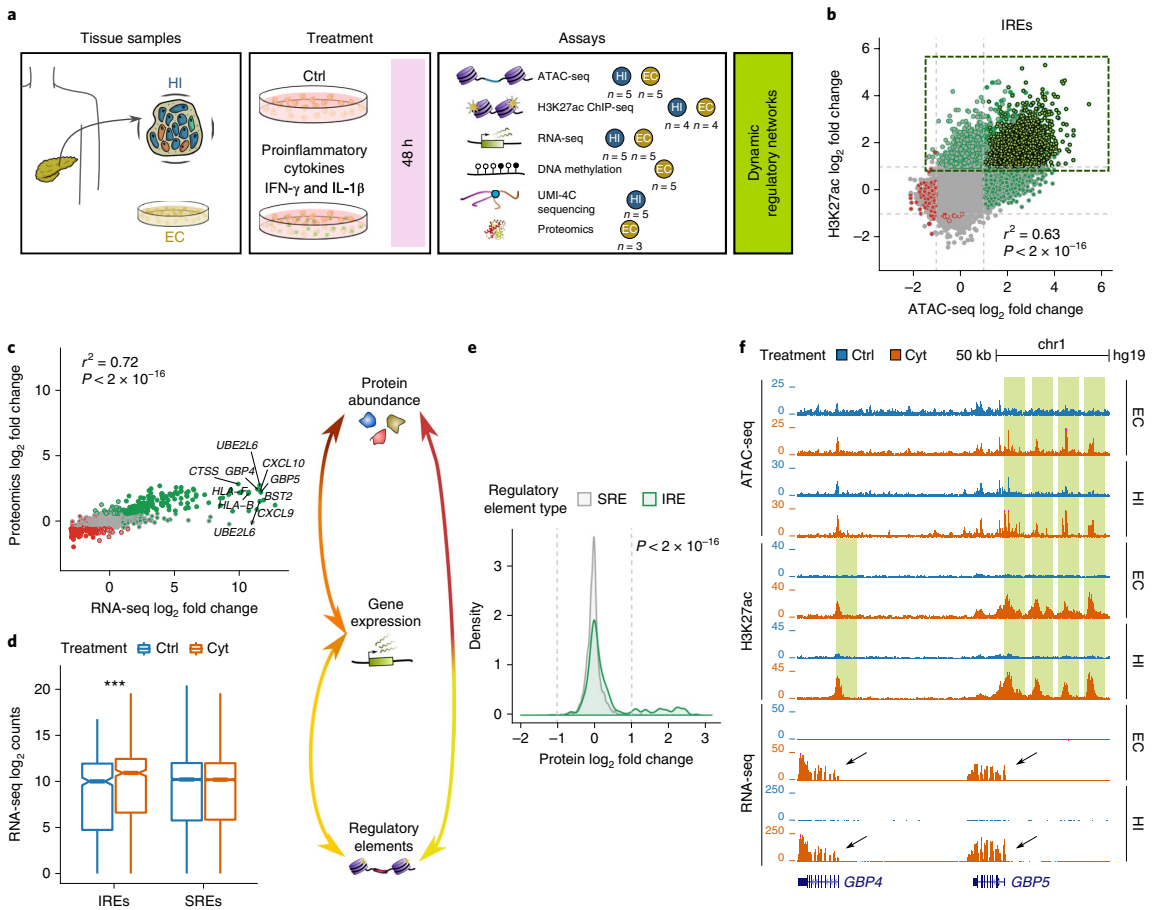


Fig. 1 | Proinflammatory cytokine exposure causes profound remodeling of the β -cell regulatory landscape. **a**, Summary of the experimental design. The number of EC and human pancreatic islet (HI) samples used in different assays is shown. **b**, Correlation between chromatin accessibility and H3K27ac deposition; each dot corresponds to a chromatin site. The point fill refers to the ATAC-seq; the border refers to the H3K27ac classification (green = gained; red = lost; gray = stable). The dashed box depicts the regulatory elements (IREs) and the lighter shade of green depicts a subtype named neo-IREs (see text). **c**, Correlation between changes in RNA expression and protein abundance in EC cells. The point fill and border indicate the classification of RNA-seq and proteins, respectively (upregulated = green; downregulated = red; equally regulated = gray). **d**, Genes proximal to the IREs (see Methods) show cytokine-induced expression in EC cells exposed or unexposed to proinflammatory treatment. CYT = cytokine-exposed. Two-sided Wilcoxon test $***P < 0.001$. The boxplot limits show the upper and lower quartiles; the whiskers extend to 1.5 \times the interquartile range. **e**, Translation of proteins encoded by IRE-associated genes is induced by cytokine exposure in EC cells. This is shown by the significantly different (two-sided Wilcoxon test $P < 2 \times 10^{-16}$) \log_2 fold change distribution of protein abundance obtained after cytokine exposure for proteins encoded by genes associated with IREs or SREs. **f**, Representative view of the *GBP4* and *GBP5* genes, encoding the IFN-inducible guanylate binding proteins, illustrating their upregulation on cytokine exposure and the nearby induction of IREs characterized by gains in chromatin accessibility and enrichment in H3K27ac (green boxes).

observed in the human β -cell line were concordant with those observed in the human islet preparations (Extended Data Fig. 1c).

We reasoned that changes in chromatin accessibility may reflect the activation of noncoding *cis*-regulatory elements. Thus, we used chromatin immunoprecipitation with sequencing (ChIP-seq) to map cytokine-induced changes in H3K27ac (Extended Data Fig. 1a), a key histone modification associated with active *cis*-regulatory elements shown to be dynamically regulated in response to acute stimulation⁷. We observed genome-wide deposition of the active histone modification mark on exposure to proinflammatory cytokines in both EC and human pancreatic islets (Extended Data Fig. 1b,c).

Integrative analysis of ATAC-seq and ChIP-seq indicates that changes in chromatin accessibility are strongly correlated with deposition of H3K27ac ($P < 2 \times 10^{-16}$, $r^2 = 0.63$) allowing the identification of approximately 3,800 open chromatin regions that gained H3K27ac (FDR-adjusted $P < 0.05$; \log_2 fold change > 1) on exposure to proinflammatory cytokines (Fig. 1b and Extended Data Fig. 1d). We found that this subset of open chromatin regions is preferentially located distally to gene transcription start sites (TSS) (Extended Data Fig. 1e), and their sequence is evolutionarily conserved (Extended Data Fig. 1f) and enriched for specific transcription factor binding sites (Extended Data Fig. 1g). We named

these newly mapped regions induced regulatory elements (IREs) (Supplementary Table 1 and Supplementary Table 2).

Chromatin changes link to variation in transcription and translation. We next explored whether the newly identified IREs were associated with changes in gene expression and protein translation. To identify β -cell transcripts and proteins induced by the proinflammatory cytokines, we assayed gene expression by RNA sequencing (RNA-seq; five replicates in EC and five replicates in human pancreatic islets⁹; see Extended Data Fig. 1a) and collected multiplex proteomics data for three EC replicates after exposure or lack of exposure to proinflammatory cytokines.

In line with the chromatin assays, which indicated extensive gene regulatory activation, we unraveled cytokine-induced transcriptional activation resulting in approximately 1,200 upregulated genes (FDR-adjusted $P < 0.05$; \log_2 fold change > 1 ; Extended Data Fig. 2a,b). By multiplex proteomics, after rigorous filtering, a subset of 10,166 proteins was confidently quantified and retained for significance testing. A total of 348 proteins displayed significant changes in abundance (FDR/ $Q < 0.15$ and absolute fold change > 1.5 ; absolute \log_2 fold change > 0.58). Of the overall detected proteins, 2.19% were upregulated (Extended Data Fig. 2c), 76% of which had induced messenger RNA levels at 48 h, confirming consistency between RNA-seq and protein changes ($r^2 = 0.72$, $P < 2 \times 10^{-16}$; Fig. 1c). Protein–protein interactions inferred from β -cell cytokine-induced proteins resulted in a network more connected than expected by chance ($P < 10^{-3}$) and significantly enriched for Molecular Signatures Database (<http://software.broadinstitute.org/gsea/msigdb/>) pathways including IFN- γ signaling, antigen processing and presentation, apoptosis and T1D (Kyoto Encyclopedia of Genes and Genomes T1D $P = 7.9 \times 10^{-8}$; Extended Data Fig. 2d).

As expected, we found that IREs were linked to the upregulation of the nearby gene(s) as well as to an induced abundance of the corresponding protein (Fig. 1d,e and Extended Data Fig. 2e). Moreover, gene induction was highly correlated with the number of associated IREs, suggesting a cumulative effect of IREs on cytokine-induced changes in gene expression (Extended Data Fig. 2f).

Taken together, these findings reveal that the pancreatic β -cell response to proinflammatory cytokines is dynamic, involving extensive chromatin remodeling and profound changes in the regulatory landscape (Fig. 1f and Extended Data Fig. 2g). Such changes are associated with induction of transcription and protein translation including pathways implicated in the pathogenesis of T1D. Newly defined regulatory maps can be visualized online along with other islet regulatory annotations at www.isletregulome.org.

Primed and neo-regulatory elements mediate cytokine response.

We next sought to gain an insight into the dynamic activation of IREs. The relationship between chromatin openness and H3K27ac deposition on exposure to proinflammatory cytokines allows the distinction of two classes of IREs (Fig. 1b and Fig. 2a–c): opening IREs ($n = 2,436$), which gain both chromatin accessibility (\log_2 fold change > 1) and H3K27ac (\log_2 fold change > 1); and primed IREs ($n = 1,362$), which are already accessible chromatin sites before treatment (ATAC-seq \log_2 fold change < 1) and gain H3K27ac (\log_2 fold change > 1) on exposure to the stimulus. Primed and opening IREs are both associated with gene expression induction (Extended Data Fig. 3a), are phylogenetically conserved (Extended Data Fig. 3b) and preferentially map distally relative to a gene's TSS (Extended Data Fig. 3c). We further revealed that 70% of opening IREs ($n = 1,716$), before cytokine exposure, are inactive and inaccessible (that is, undetectable by ATAC-seq under basal conditions; see Methods). We named the latter neo-IREs. Neo-IREs represent 45% of all IREs and may mirror a class of regulatory elements identified on stimulation of mouse macrophages and named 'latent enhancers'²⁷.

Because chromatin openness, the feature distinguishing the two classes of IREs, is believed to reflect transcription factor occupancy, we analyzed their sequence composition in search of recognition sequences of key transcription factors orchestrating the β -cell response to proinflammatory cytokines. Even though IREs are mostly distal to TSS (Extended Data Fig. 3c), to reduce sequence bias, we excluded all annotated promoters from this analysis. The two classes of distal IREs predominantly mapped to the enhancer chromatin state (Extended Data Fig. 3d) and showed clear differences in sequence composition. Newly induced enhancers were enriched for the binding motifs of inflammatory response transcription factors including IFN-sensitive response element (ISRE), signal transducer and activator of transcription (STAT) and nuclear factor kappa-light-chain-enhancer of activated B cells (NF- κ B) (Extended Data Fig. 3e). Instead, primed enhancers were enriched for binding motifs of inflammatory response transcription factors (ISRE, STAT) and, unexpectedly, islet-specific transcription factors (HNF1A/B, NEUROD1, PDX1, MAFB, NKX6.1; Extended Data Fig. 3f). Importantly, we found that in primed enhancers, inflammatory response and islet-specific transcription factor binding motifs mapped to the same genomic regions, suggesting cobinding and possibly cooperation of the two classes of transcription factors (Extended Data Fig. 3g,h).

Sequence composition bias per se does not imply transcription factor occupancy. Thus, we took advantage of published ChIP-seq datasets of islet-specific transcription factors (MAFB, PDX1, FOXA2, NKX6.1 and NKX2.2) mapped in unstimulated human pancreatic islets⁹ to measure transcription factor occupancy in primed and neo-enhancers before the proinflammatory stimulus. As expected from the sequence composition analysis, primed enhancers (unlike neo-enhancers) are bound by tissue-specific transcription factors even before their activation (Fig. 2d and Extended Data Fig. 3i). Transcription factor occupancy can also be indirectly assessed by ATAC-seq, which assays the protection of the bound sequence to transposase cleavage (footprint). Footprint analysis is effective for transcription factors with a long residence time¹⁰, such as the IFN regulatory factor (IRF) and STAT transcription factor families. Our analyses revealed the emergence of footprint marks on proinflammatory treatment in correspondence to ISRE motifs in both primed and neo-enhancers (Fig. 2e), indicating cytokine-induced transcription factor occupancy of IREs.

Gene regulation is orchestrated by different epigenetic mechanisms. DNA methylation is a relatively stable epigenetic mark contributing to maintenance of cellular identity^{11,12}. Moreover, high-resolution DNA methylation maps, obtained from multiple tissues, suggested that the vast majority of tissue-specific, differentially methylated regions are located at distal, mostly noncoding regulatory sites¹³. Consequently, characterization of the DNA methylome in the context of relevant stimuli is important for understanding the functional mechanisms of tissue-specific responses in human disease¹⁴. Thus, we explored if cytokine-induced chromatin remodeling is associated with changes in DNA methylation. We quantified DNA methylation changes by performing dense methylation arrays in EC cells exposed or unexposed to IFN- γ and IL-1 β . The Infinium MethylationEPIC array was designed to interrogate with high precision and coverage $> 850,000$ CpG sites (approximately 3% of all sites in the genome) selected primarily because of their location close to gene promoters and CpG island regions. By focusing on the 1,230 IRE enhancers harboring one or more CpG sites interrogated by the array, we observed that primed enhancers overlap lowly methylated CpGs (median $\beta = 0.12 \pm 0.08$), which did not vary significantly on cytokine exposure. Such observation is in sharp contrast with neo-enhancers, which were highly methylated under control conditions (median $\beta = 0.77 \pm 0.10$), but underwent a significant loss of DNA methylation (two-sided Wilcoxon test, $P = 4.13 \times 10^{-4}$) on treatment (Fig. 2f). While we did not observe cytokine-induced methylation,

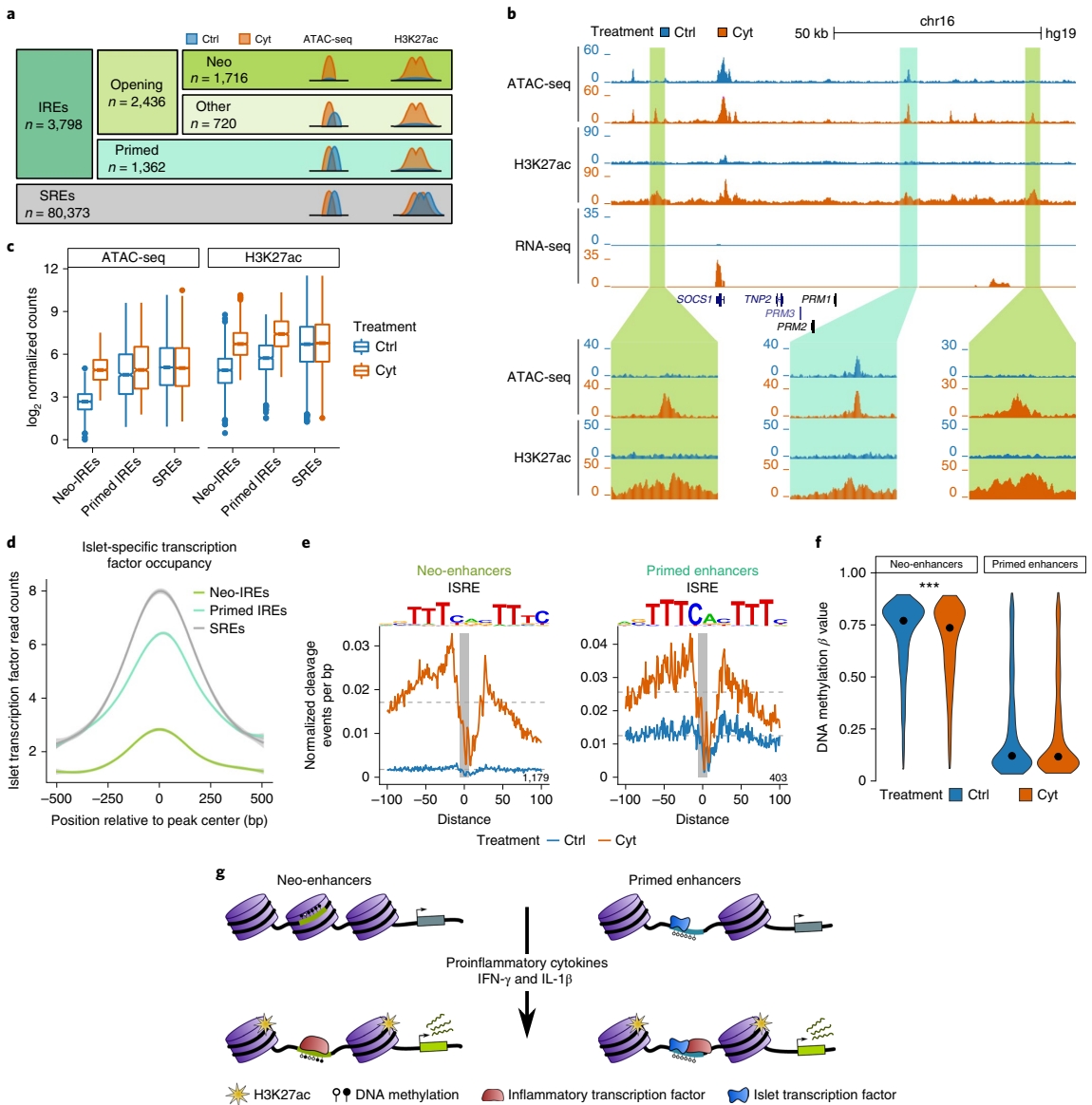


Fig. 2 | The β -cell response to proinflammatory cytokines unveils neo and primed IREs. a, Classification of ATAC-seq open chromatin sites on exposure of human β cells to IFN- γ and IL-1 β . **b**, View of the *SOCS1* locus, a gene strongly induced on proinflammatory cytokine exposure. Representative examples of primed (blue box) and neo-IREs (green boxes) are shown. **c**, Boxplot distribution of ATAC-seq and H3K27ac normalized tag counts at different classes of IREs. The boxplot limits show the upper and lower quartiles; the whiskers extend to 1.5 \times the interquartile range; the individual data points represent the outliers; the notch represents the confidence interval around the median. **d**, Islet-specific transcription factor occupancy at neo, primed and stable regulatory elements. Read density for PDX1, NKX2.2, FOXA2, NKX6.1 and MAFB was calculated in 10 bp bins in 1 kb windows centered on the regulatory element. The lines represent the mean, while the gray shading depicts the s.d. **e**, Footprint analysis of ISRE motifs in neo (left) and primed regulatory elements (right) in cells exposed or unexposed to IFN- γ and IL-1 β (blue = Ctrl; orange = cytokines). **f**, Violin plots showing the distribution of DNA methylation β values in neo and primed enhancers, exposed or unexposed to proinflammatory cytokines. Two-sided Wilcoxon test *** $P < 0.001$. **g**, Model showing two types of IREs driving the response to proinflammatory cytokines in human β cells.

we found that approximately 70% of significantly demethylated probes (FDR-adjusted $P \leq 0.05$; β -value differences between cytokine-exposed and untreated cells ($\beta_{\text{Cyt}} - \beta_{\text{Ctrl}} < -0.20$) mapping to IREs were located at neo-enhancers (Extended Data Fig. 3j,k).

These results suggest that neo-enhancers are enriched for methylated CpGs that undergo preferential demethylation on cytokine treatment, whereas primed enhancers are enriched for unmethylated CpGs that do not change their methylation status on exposure to cytokines.

Taken together these analyses lead to a model, where proinflammatory cytokines elicit a regulatory response in β cells characterized by: (1) new induction of distal regulatory elements coupled with reduction of DNA methylation and binding of inflammatory response transcription factors; and (2) activation of regulatory elements prebound by islet-specific transcription factors and induced by inflammatory response transcription factors (Fig. 2g).

Collectively, these results allow the reconstruction of *cis*-regulatory networks activated in human pancreatic β cells on exposure to the proinflammatory cytokines IFN- γ and IL-1 β (Extended Data Fig. 4a–c and Supplementary Table 1).

Changes in islet 3D chromatin structure. Regulatory regions can exert control over genes at megabase distances through the formation of DNA loops. These loops are often confined within structures known as topologically associating domains^{15–17}. Topologically associating domains are largely conserved on evolution, are invariant in different cell types and have their boundaries defined by the regulatory scope of tissue-specific enhancers^{18,19}. Our knowledge regarding the general characteristics and mechanisms of loops is improving^{20–23}, but much less is known regarding the mechanisms and functional significance of dynamic looping events during biological processes.

We took advantage of promoter capture Hi-C performed in human pancreatic islets²⁴ to explore the long-range interactions between gene promoters and cytokine-induced and invariant distant regulatory elements. Interestingly, we observed that the interaction confidence scores captured between IRE enhancers and gene promoters in untreated islets were significantly reduced compared with distal stable regulatory element (SRE) ($P=1.8 \times 10^{-11}$; Extended Data Fig. 5a). Since this finding pointed to potential dynamic properties of the interaction maps, we next sought to investigate if cytokine-induced regulatory changes are linked to modification of 3D chromatin structure and if induction of β -cell cytokine-responsive regulatory elements is coupled with the new formation of DNA looping interactions.

Hi-C profiles are limited in sequencing coverage and library complexity, resulting in maps of reduced resolution relative to regulatory maps of functional elements. On the other hand, circular chromosome conformation capture (4C) approaches are difficult to interpret quantitatively mainly due to potential amplification bias. Thus, we applied targeted chromosome capture with unique molecular identifiers (UMI-4C), a recently developed method²⁵, to quantitatively measure interaction intensities in human islets before and after exposure to proinflammatory cytokines. We centered the conformation capture viewpoint at the promoter of 13 genes (*TNFSF10*, *GBP1* and *CIITA*, among others) whose expression was strongly induced by cytokine exposure.

UMI-4C showed marked changes in the 3D chromatin structure at the analyzed loci. Promoters of the induced genes gained chromatin interactions, with distal genomic regions reflecting the formation of new DNA looping events (Fig. 3a,b and Extended Data Fig. 5b–d). Importantly, such new contacts were preferentially engaged with newly mapped human islet cytokine-responsive IREs (Fig. 3c).

These results demonstrate that cytokine exposure induces changes in human islet 3D chromatin conformation including the formation of new enhancer–promoter interactions. Such changes allow the newly activated distal IREs to contact their target gene promoters.

Islet cytokine enhancers are implicated in T1D genetic susceptibility. GWAS have identified approximately 60 chromosome regions associated with T1D²⁶, with many of the association signals having been assigned to candidate genes with immunological functions. Consistent with this notion, several studies reported a primary enrichment of T1D risk variants in T- and B-cell regulatory

elements^{4,5}. Furthermore, there is a substantial lack of statistically significant overlap of T1D-associated variants in islet enhancers, while such regulatory elements are instead enriched for GWAS signals for type 2 diabetes (T2D) and fasting glucose^{9,27}. Nonetheless, the molecular mechanisms linking T1D association signals to cellular functions are poorly described for most of the regions of association identified.

We hypothesized that a subset of T1D genetic signals may reflect an altered capacity of β cells to react to an inflammatory environment. Thus, we sought to explore to what extent genetic signals underlying T1D susceptibility act through pancreatic islet regulatory response to proinflammatory cytokines.

Causal *cis* variants are expected to be found in sequences that act as regulatory regions in state-specific and disease-relevant tissues. Thus, we examined nonshared loci with genome-wide significant association to T2D and T1D in European populations and considered all variants in high linkage disequilibrium (1000 Genomes Project, phase 3 European population (EUR), $R^2 > 0.8$) with a lead SNP reported in the National Human Genome Research Institute-European Bioinformatics Institute (NHGRI-EBI) GWAS catalog²⁶. In line with previous observations^{4,9}, we found that T2D but not T1D risk variants overlap human islet noncytokine-responsive regulatory elements (that is, SREs) more than expected by chance (T2D SNPs in SREs $P < 2 \times 10^{-16}$, $z = 5.47$). In contrast, we found that human islet IREs are enriched for T1D but not T2D risk variants (T1D SNPs in IREs $P = 3 \times 10^{-6}$, $z = 4.61$) (Fig. 4a). This result was reproduced when using regulatory elements detected in EC cells (Extended Data Fig. 6a). Such findings revealed 9 T1D-associated regions (13% of the total) containing at least 1 islet cytokine-induced regulatory element directly overlapping a T1D-associated variant (Supplementary Table 3 and Extended Data Fig. 6b–f).

We noted that the two T1D lead SNPs at the 1q24.3 and 16q13.13 loci (rs78037977 (refs. 28,29) and rs193778 (ref. 4), respectively) were directly overlapping IREs in the islets. We used GWAS genotyping data from a cohort of 14,575 individuals (5,909 T1D cases and 8,721 controls (Ctrls); see Methods) to confirm their association with T1D. Both variants were included in the 99% credible set of their respective locus and displayed strong association P values (rs78037977, $P = 6.94 \times 10^{-10}$; rs193778, $P = 1.33 \times 10^{-7}$; see Supplementary Table 4 for the posterior probability of association and variant ranking in the credible set), indicating that they could potentially be causal.

At the 1q24.3 locus, rs78037977 (NC_000001.10:g.172715702 A>G) overlaps an islet cytokine-induced chromatin site (Fig. 4b), which is prebound by islet-specific transcription factors and is a predicted enhancer in other cell types (Extended Data Fig. 6g). We created allele-specific luciferase reporter constructs and measured enhancer activity in the EC cell line before and after cytokine exposure. The sequence exerts enhancer activity exclusively after cytokine exposure, which is disrupted by the rs78037977 T1D-associated G allele (one-way analysis of variance (ANOVA), $F = 26$, $P = 4.34 \times 10^{-5}$; Fig. 4c and Extended Data Fig. 6h). This is consistent with a causal role of the variant at this locus. To identify the gene target of this T1D-susceptible enhancer, we reconstructed the 3D chromatin structure using chromatin capture experiments. UMI-4C in human islets identified a cytokine-induced interaction of the enhancer with *TNFSF18*, a gene activated in islets on cytokine exposure (Fig. 4d,e). *TNFSF18* encodes a cytokine, glucocorticoid-induced tumor necrosis factor receptor-related protein (GITR; also known as TNFRSF18), which modulates the inflammatory reaction and regulation of autoimmune responses³⁰. Interestingly, we noted that cytokine exposure results in upregulation of *TNFSF18* in human islets but not in the EC β -cell line, suggesting differences in gene regulatory dynamics in primary tissue or the activation of an islet cell subpopulation.

At the 16q13.13 locus, rs193778 (NC_000016.9:g.11351211 A>G) maps to a phylogenetically conserved, cytokine-responsive

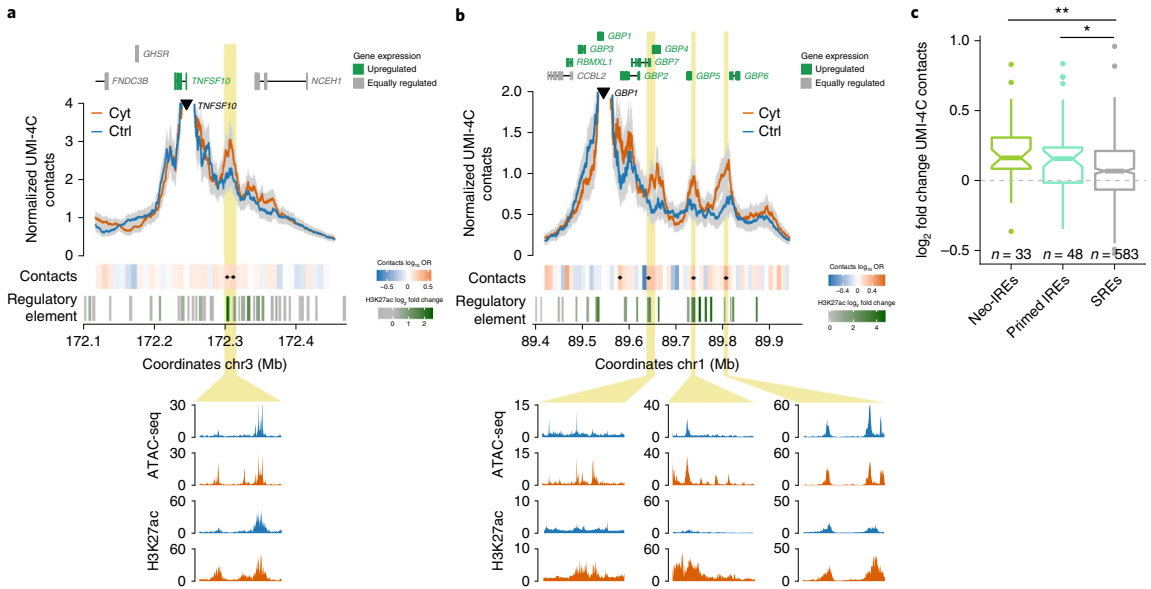


Fig. 3 | Cytokine exposure induces changes in human islet 3D chromatin structure. a, b, View of the UMI-4C chromatin contacts of the *TNFSF10* (**a**) and *GBP1* (**b**) promoters, before and after exposure to proinflammatory cytokines. The yellow boxes indicate IREs that gained contacts with the upregulated gene promoters. A heatmap under the 4C track represents the \log_{10} odds ratio of the UMI-4C contact difference in cytokine-exposed (Cyt) versus untreated cells (Ctrl). The small black diamonds over the contact heatmap indicate a significant difference between cytokine-treated and untreated 3D chromatin contacts (chi-squared test $P < 0.05$). ATAC-seq peaks are represented by rectangles shaded from gray to green proportionally to the cytokine-induced H2K27ac \log_2 fold change observed at that site. **c**, Distribution of the UMI-4C contact \log_2 fold changes (Cyt versus Ctrl) at the different types of islet open chromatin sites classified in Fig. 2a. The data, obtained by analyzing viewpoints centered at the promoter of cytokine-induced genes, show that the chromatin structural changes are preferentially happening at IREs. The boxplot limits show the upper and lower quartiles; the whiskers extend to 1.5x the interquartile range; the individual data points represent outliers; and the notch represents the confidence interval around the median. Two-sided Wilcoxon test * $P < 0.05$, ** $P < 0.01$.

regulatory element (Fig. 4f). This sequence displays enhancer activity in both treated and untreated β cells. However, exclusively in cytokine-exposed β cells, the T1D-associated G allele exerts significantly higher enhancer activity than the protective variant (one-way ANOVA, $F = 12.34$, $P = 1.23 \times 10^{-3}$; Fig. 4g and Extended Data Fig. 6). The locus includes several upregulated genes (*SOCS1*, *DEXI*, *CIITA*, *RMI2*) that could represent potential targets of this IRE. Recent research points to *DEXI* as a T1D candidate gene in immune cells and β cells^{31,32}. By performing UMI-4C experiments in human islets, we observed a strong chromatin contact between the promoter of *DEXI* and the regulatory element bearing the rs193778 T1D-associated variant (Fig. 4h). Such data points to *DEXI* as a potential causal gene in pancreatic islets.

Altogether, these results illustrate how unraveling cytokine-induced chromatin dynamics in human islets can guide the identification of *cis*-regulatory variants that are strong candidates in driving T1D-association signals.

Discussion

Our work illustrates the human pancreatic β -cell chromatin dynamics in response to an external stimulus that may be relevant in the context of T1D. We show that exposure to proinflammatory cytokines causes profound remodeling of the β -cell regulatory landscape coupled with changes in gene expression and protein production. The degree of remodeling of the regulatory network was comparable to that previously shown for macrophages or mouse dendritic cells exposed to similar stimuli⁷. We unveil the activation of approximately 3,600 cytokine-responsive distal *cis*-regulatory

elements and reveal a lack of homogeneity in their molecular mechanism of activation. We observe that the induction of a subset of regulatory regions (neo-IREs) require transcription factor binding and chromatin opening, while other chromatin sites are primed to their activation being prebound by islet-specific transcription factors. Our observations suggest a model where binding of tissue-specific transcription factors may facilitate chromatin accessibility at a subset of chromatin sites that can then be promptly activated by the induction of inflammatory response transcription factors. Such a model is supported by very recent findings³³ and it is consistent with observations in murine macrophages^{7,34} and dendritic cells³⁵; however, thus far, it has not been demonstrated in a highly differentiated and nonimmune-related tissue, such as pancreatic islets. Even though our model suggests that exposure to proinflammatory cytokines causes predominantly induction of gene transcription rather than transcript downregulation, we cannot exclude that a more prolonged stimulus could induce loss of critical β -cell processes resulting from the reduction of β -cell *cis*-regulatory network activity.

Importantly, we show that such regulatory changes are coupled with 3D chromatin remodeling, allowing the newly activated regulatory elements to contact their target genes. Several reports have described the properties of 3D chromatin dynamics in the cell developmental context^{36,37}, on loss of cell fate^{38,39}, senescence^{40,41} or in response to hormonal exposure⁴². Our observations indicate that the capacity of enhancer loop formation is maintained in a highly differentiated tissue such as the islets and it is coupled with transcriptional regulatory changes in response to an external stimulus.

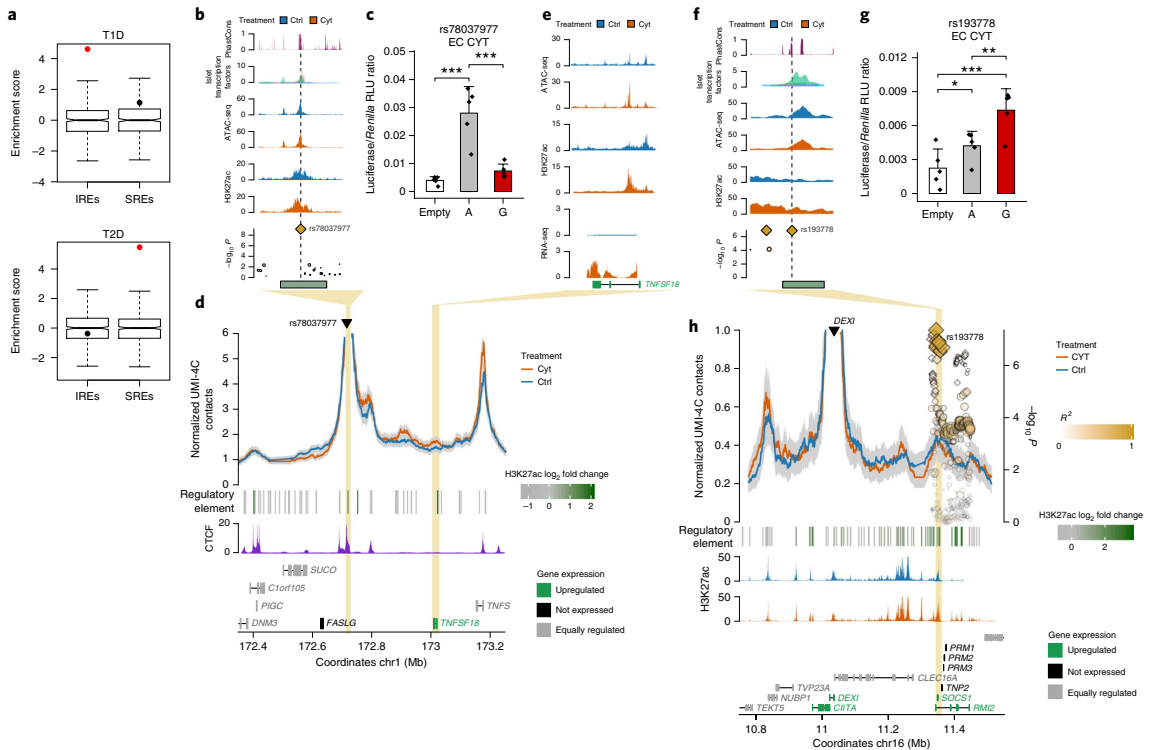


Fig. 4 | Cytokine-induced islet regulatory elements map to T1D-associated regions and guide the identification of functional risk variants. **a**, Islet IREs are enriched for T1D but not T2D risk variants while the opposite is true for islet SREs. Significant VSE scores are shown in red (Bonferroni-adjusted $P < 0.05$). The boxplot limits show the upper and lower quartiles; the whiskers extend to 1.5x the interquartile range; the notch represents the median confidence interval for distributions of matched null sets (500 permutations). **b**, rs78037977 overlaps an IRE bound by islet-specific transcription factors under basal conditions⁹. PhastCons, evolutionary conservation scores from phylogenetic analysis with space/time models. **c**, Luciferase assays in EC cells exposed to cytokines show that the sequence exerts enhancer activity that is reduced in the T1D-associated allele (G). RLU, relative light units. **d**, UMI-4C in human islets shows that the IRE containing rs78037977 engages multiple distal chromatin contacts. **e**, One induced chromatin contact mapping to the upregulated *TNFSF18* gene. **f**, The rs193778 variant maps to a phylogenetically conserved IRE. **g**, Luciferase assays in EC cells exposed to cytokines show significantly increased enhancer activity of the risk (G) allele compared to the nonrisk (A) allele. **h**, UMI-4C in islets, using the promoter of *DEX1* as the viewpoint, show a chromatin contact with the IRE bearing the T1D risk variant. **d,h**, ATAC-seq peaks are represented by rectangles shaded proportionally to the H2K27ac log₂ fold change. **c,g**, Statistical significance was determined by one-way ANOVA followed by Bonferroni correction. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$. The bars represent the mean \pm s.d.

The model used in our study to explore chromatin dynamics is of particular interest because it mimics the inflammatory environment that pancreatic islets may face in the early stages of T1D. While several T1D candidate genes regulating key steps related to danger signal recognition and innate immunity are expressed in human islets⁴³, T1D-associated variants are enriched for immune cell types but not in stable pancreatic islets regulatory elements⁴. Such apparent contradiction may be reconciled by our findings showing that human islet cytokine-responsive regulatory elements are enriched for T1D risk variants. Our data, supported by recent findings revealing regulatory variants that affect enhancer activation in the immune response^{33,44}, opens the avenue to identify T1D molecular mechanisms acting at the pancreatic islet cell level.

Although we cannot exclude that functional variants disrupting the β -cell regulatory mechanisms may at the same time affect the regulatory potential of immune-related cell types, the availability of stimulus-responsive *cis*-regulatory maps in pancreatic islets will facilitate hypothesis-driven experiments to uncover how common and lower-frequency genetic variants impact islet cells in T1D. In this study, we

researched the human islet responses to a specific proinflammatory stimulus. Future work studying additional immune-mediated stresses potentially affecting β cells at different stages of the disease may uncover other association signals acting at the islet cell level.

More generally, our findings could apply by extension to other diseases where primed enhancers may facilitate cell type-specific responses to ubiquitous signals resulting in tissue-specific genetic susceptibility in autoimmune diseases.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information, details of author contributions and competing interests, statements of code and data availability and associated accession codes are available at <https://doi.org/10.1038/s41588-019-0524-6>.

Received: 30 November 2018; Accepted: 27 September 2019; Published online: 01 November 2019

References

- Todd, J. A. Etiology of type 1 diabetes. *Immunity* **32**, 457–467 (2010).
- Ziegler, A.-G. & Nepom, G. T. Prediction and pathogenesis in type 1 diabetes. *Immunity* **32**, 468–478 (2010).
- Eizirik, D. L., Colli, M. L. & Ortis, F. The role of inflammation in insulinitis and β -cell loss in type 1 diabetes. *Nat. Rev. Endocrinol.* **5**, 219–226 (2009).
- Onengut-Gumuscu, S. et al. Fine mapping of type 1 diabetes susceptibility loci and evidence for colocalization of causal variants with lymphoid gene enhancers. *Nat. Genet.* **47**, 381–386 (2015).
- Farh, K. K. et al. Genetic and epigenetic fine mapping of causal autoimmune disease variants. *Nature* **518**, 337–343 (2015).
- Ravassard, P. et al. A genetically engineered human pancreatic β cell line exhibiting glucose-inducible insulin secretion. *J. Clin. Invest.* **121**, 3589–3597 (2011).
- Ostuni, R. et al. Latent enhancers activated by stimulation in differentiated cells. *Cell* **152**, 157–171 (2013).
- Gonzalez-Duque, S. et al. Conventional and neo-antigenic peptides presented by β cells are targeted by circulating naive CD8+ T cells in type 1 diabetic and healthy donor. *Cell Metab.* **28**, 946–960.e6 (2018).
- Pasquali, L. et al. Pancreatic islet enhancer clusters enriched in type 2 diabetes risk-associated variants. *Nat. Genet.* **46**, 136–143 (2014).
- Sung, M.-H., Guertin, M. J., Baek, S. & Hager, G. L. DNase footprint signatures are dictated by factor dynamics and DNA sequence. *Mol. Cell* **56**, 275–285 (2014).
- Bird, A. DNA methylation patterns and epigenetic memory. *Genes Dev.* **16**, 6–21 (2002).
- Hemmerger, M., Dean, W. & Reik, W. Epigenetic dynamics of stem cells and cell lineage commitment: digging Waddington's canal. *Nat. Rev. Mol. Cell Biol.* **10**, 526–537 (2009).
- Lister, R. et al. Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* **462**, 315–322 (2009).
- Feinberg, A. P. Phenotypic plasticity and the epigenetics of human disease. *Nature* **447**, 433–440 (2007).
- Dixon, J. R. et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376–380 (2012).
- Nora, E. P. et al. Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **485**, 381–385 (2012).
- Dekker, J., Marti-Renom, M. A. & Mirny, L. A. Exploring the three-dimensional organization of genomes: interpreting chromatin interaction data. *Nat. Rev. Genet.* **14**, 390–403 (2013).
- Symmons, O. et al. Functional and topological characteristics of mammalian regulatory domains. *Genome Res.* **24**, 390–400 (2014).
- Krivega, I., Dale, R. K. & Dean, A. Role of LDB1 in the transition from chromatin looping to transcription activation. *Genes Dev.* **28**, 1278–1290 (2014).
- Yu, M. & Ren, B. The three-dimensional organization of mammalian genomes. *Annu. Rev. Cell Dev. Biol.* **33**, 265–289 (2017).
- Rowley, M. J. & Corces, V. G. Organizational principles of 3D genome architecture. *Nat. Rev. Genet.* **19**, 789–800 (2018).
- Schoenfelder, S. & Fraser, P. Long-range enhancer-promoter contacts in gene expression control. *Nat. Rev. Genet.* **20**, 437–455 (2019).
- Stadhouders, R., Filion, G. J. & Graf, T. Transcription factors and 3D genome conformation in cell-fate decisions. *Nature* **569**, 345–354 (2019).
- Miguel-Escalada, I. et al. Human pancreatic islet three-dimensional chromatin architecture provides insights into the genetics of type 2 diabetes. *Nat. Genet.* **51**, 1137–1148 (2019).
- Schwartzman, O. et al. UMI-4C for quantitative and targeted chromosomal contact profiling. *Nat. Methods* **13**, 685–691 (2016).
- MacArthur, J. et al. The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucleic Acids Res.* **45**, D896–D901 (2017).
- Parker, S. C. J. 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).
- Cooper, N. J. et al. Type 1 diabetes genome-wide association analysis with imputation identifies five new risk regions. Preprint at *bioRxiv* <https://doi.org/10.1101/120022> (2017).
- Fortune, M. D. et al. Statistical colocalization of genetic risk variants for related autoimmune diseases in the context of common controls. *Nat. Genet.* **47**, 839–846 (2015).
- Ray, A., Basu, S., Williams, C. B., Salzman, N. H. & Dittel, B. N. A novel IL-10-independent regulatory role for B cells in suppressing autoimmunity by maintenance of regulatory T cells via GITR ligand. *J. Immunol.* **188**, 3188–3198 (2012).
- Davison, L. J. et al. Long-range DNA looping and gene expression analyses identify *DEXI* as an autoimmune disease candidate gene. *Hum. Mol. Genet.* **21**, 322–333 (2012).
- Dos Santos, R. S. et al. *DEXI*, a candidate gene for type 1 diabetes, modulates rat and human pancreatic beta cell inflammation via regulation of the type I IFN/STAT signalling pathway. *Diabetologia* **62**, 459–472 (2019).
- Alasoo, K. et al. Shared genetic effects on chromatin and gene expression indicate a role for enhancer priming in immune response. *Nat. Genet.* **50**, 424–431 (2018).
- Heinz, S. et al. Effect of natural genetic variation on enhancer selection and function. *Nature* **503**, 487–492 (2013).
- Vandenbon, A., Kumagai, Y., Lin, M., Suzuki, Y. & Nakai, K. Waves of chromatin modifications in mouse dendritic cells in response to LPS stimulation. *Genome Biol.* **19**, 138 (2018).
- Phanstiel, D. H. et al. Static and dynamic DNA loops form AP-1-bound activation hubs during macrophage development. *Mol. Cell* **67**, 1037–1048.e6 (2017).
- Mumbach, M. R. et al. Enhancer connectome in primary human cells identifies target genes of disease-associated DNA elements. *Nat. Genet.* **49**, 1602–1612 (2017).
- Taberlay, P. C. et al. Three-dimensional disorganization of the cancer genome occurs coincident with long-range genetic and epigenetic alterations. *Genome Res.* **26**, 719–731 (2016).
- Barutcu, A. R. et al. Chromatin interaction analysis reveals changes in small chromosome and telomere clustering between epithelial and breast cancer cells. *Genome Biol.* **16**, 214 (2015).
- Chandra, T. et al. Global reorganization of the nuclear landscape in senescent cells. *Cell Rep.* **10**, 471–483 (2015).
- Criscione, S. W. et al. Reorganization of chromosome architecture in replicative cellular senescence. *Sci. Adv.* **2**, e1500882 (2016).
- Le Dily, F. et al. Distinct structural transitions of chromatin topological domains correlate with coordinated hormone-induced gene regulation. *Genes Dev.* **28**, 2151–2162 (2014).
- Op de Beeck, A. & Eizirik, D. L. Viral infections in type 1 diabetes mellitus: why the β cells? *Nat. Rev. Endocrinol.* **12**, 263–273 (2016).
- Kim-Hellmuth, S. et al. Genetic regulatory effects modified by immune activation contribute to autoimmune disease associations. *Nat. Commun.* **8**, 266 (2017).

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

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2019

Methods

Human islets and EC cells. Human islets from 14 multiorgan donors without a history of glucose intolerance were isolated in compliance with ethical regulations (Supplementary Note 1) and according to established isolation procedures^{45,46} (Supplementary Note 2 and Supplementary Table 5). The human insulin-producing EC cells were kindly provided by R. Scharfmann⁶ and cultured in DMEM medium (Supplementary Note 3).

Human islets and EC cells were either exposed or unexposed to a cocktail of proinflammatory cytokines IFN- γ and IL-1 β for 48 h. The cytokine concentrations used were those described in previous dose–response experiments^{47–49} (Supplementary Note 2). The glucose stimulation index was tested on human islet preparations and EC cell samples to confirm functional competence of the samples (Supplementary Note 4 and Extended Data Fig. 7).

ChIP-seq and ATAC-seq. ATAC-seq library preparations were carried out as described previously⁵⁰ with minor modifications^{51,52} (Supplementary Note 5). ChIP-seq was carried out using tagmentation (ChIPmentation) as described previously⁵³ (Supplementary Note 6).

ATAC-seq and ChIP-seq libraries were sequenced on a HiSeq 2500 system (Illumina). Reads were aligned to the hg19 reference genome using Bowtie 2 v.2.3.4.1 (ref. ⁵⁴) using default parameters. After alignment, reads mapping to Encyclopedia of DNA Elements blacklist regions⁵⁵, noncanonical chromosomes or mitochondrial DNA were discarded. Duplicates were removed using samtools markdup v.1.8 (ref. ⁵⁶). See Supplementary Table 6 for the number of mapped reads per experiment and Extended Data Fig. 8 for measures of ATAC-seq quality.

Peaks were called with MACS2 callpeak v.2.1 (ref. ⁵⁷) with the parameters ‘-q 0.05 -nomodel -shift -100 -extsize 200’ for ATAC-seq and ‘-broad -broad-cutoff 0.1 --nomodel’ for H3K27ac ChIP-seq. A more detailed description of bioinformatics processing can be found in Supplementary Note 7.

RNA-seq. Total RNA was isolated from EC cells and human islets⁶ using the RNeasy Mini Kit (QIAGEN), which retrieves RNA molecules longer than 200 nucleotides, as described in detail previously⁵⁸. RNA integrity number values were evaluated using the Bioanalyzer 2100 (Agilent Technologies). All the samples had RNA integrity number values >8 (Supplementary Note 8).

RNA-seq libraries were sequenced on a HiSeq 2000 platform (Illumina) to produce 100 base pair (bp)-long paired-end reads with an average of 180 million reads per replicate (EC cells, $n=5$). Reads were aligned using TopHat v.2.0.13 (ref. ⁵⁹) to the GChr37 genome with default parameters. Then reads were assigned to GENCODE release 18 gene annotation⁶⁰ using htseq-count v.0.6.1p1 (ref. ⁶¹) with default parameters. The RNA-seq of five human islet preparations⁶ was used for comparison and processed in an identical way.

Differential analysis of ATAC-seq, ChIP-seq and RNA-seq. For both ATAC-seq and ChIP-seq, aligned reads from all replicates were merged into a single BAM file to identify a comprehensive set of peaks. Next, we used the comprehensive peak set to compute read counts separately for each replicate and condition. In the case of the RNA-seq data, the output of htseq-count was used as the input matrix for downstream analysis. The generated matrices were normalized and differential analysis was performed using DESeq2 v.1.24.0 (ref. ⁶²) using a paired sample design (Supplementary Note 9). Thresholds for significance were set at an FDR-adjusted $P < 0.05$ and an absolute log₂ fold change > 1. All regions/genes that did not reach significance or did not pass the log₂ fold change cutoff were classified as stable/equally regulated.

Proteomics. For the proteomics analysis, 1.5 million EC cells treated or not treated with cytokines (IL-1 β +IFN- γ) were processed using the metabolite, protein and lipid extraction approach⁶³ (Supplementary Note 10).

Collected data were processed using Decon2LS_V2 v.2.3.1.4 (ref. ⁶⁴) and DtaRefinery v.1.2 (ref. ⁶⁵), both using default parameters, to recalibrate the runs and generate peak lists. Peptide identification was done using MS-GF+ v.2017.08.23 (ref. ⁶⁶) by searching peak lists against islet protein sequences deduced from a transcriptomics experiment⁶⁷ and supplemented with keratin sequences (32,780 total protein sequences) (Supplementary Note 10).

Extracted reporter ion intensities (Supplementary Note 10) were then converted into log₂ and normalized by standard median centering. Proteins were quantified using a Bayesian proteoform discovery methodology (Bayesian proteoform quantification) in combination with standard reference-based median quantification⁶⁷ and were considered significant at a cutoff of $P \leq 0.05$ based on a paired t -test.

Protein–protein interaction network analysis was performed with GeNets⁶⁸ using Metanetworks v.1.0, which integrates protein–protein interactions from InWeb v.3 (ref. ⁶⁹) and ConsensusPathDB v.32 (ref. ⁷⁰). Default parameters were applied and Molecular Signatures Database-enriched v.6.1 (ref. ⁷¹) pathways were overlaid.

Defining classes of IREs. To characterize the dynamics of chromatin accessibility on exposure of human islets and EC cells to proinflammatory cytokines, we processed the results obtained from the DESeq2 differential analysis and computed

the overlap between ATAC-seq peaks and H3K27ac-enriched sites, allowing a 200 bp gap. Regions annotated as stable for both ATAC-seq and H3K27ac assays were classified as SREs. Regions classified as either stable or gained in ATAC-seq differential analysis and as gained in H3K27ac were classified as IREs.

IREs were classified in two groups: opening IREs ($n=2,436$), corresponding to regions annotated as gained for both ATAC-seq and H3K27ac; and primed IREs ($n=1,362$) for regions annotated as stable for ATAC-seq and gained for H3K27ac. Since opening IREs include a gradient of cytokine-induced chromatin accessibility changes, we next selected only those opening regions that were completely closed before cytokine exposure. For this purpose, we considered newly open chromatin regions as those opening ATAC-seq peaks that were not called in the control samples using a relaxed threshold ($P \leq 0.05$). Such analysis allowed us to identify a subset of 1,716 opening regions that we named neo-IREs. A similar approach was used to identify macrophage latent enhancers⁷.

See Supplementary Note 11 for the sequence conservation analysis performed at the different classes of IREs.

Assigning regulatory elements to target genes. To annotate regulatory elements as distal or proximal, we assigned each regulatory element to the nearest TSS of a coding gene (using GENCODE release 18 annotation⁶⁰). Those regions lying within 2 kilobases (kb) from the nearest TSS were annotated as promoters while the rest were considered as distal regulatory elements.

To test the association between different classes of open chromatin and changes in gene expression and protein abundance (Fig. 1d,e and Extended Data Figs. 2e and 3a) in an unbiased manner, we assigned ATAC-seq sites to genes closer than 15 kb from their TSS. To analyze the additive effect of IREs on gene expression changes, we associated to a gene all IREs within 40 kb of their TSS (Extended Data Fig. 2f).

Finally, to detect all possible IRE gene targets, we assigned to each IRE all upregulated genes whose TSS was closer than 40 kb. When an upregulated gene could not be found in <40 kb, the IRE was assigned to the closest, but <1 Mb far, induced gene (Extended Data Fig. 4a and Supplementary Tables 1 and 2; see Supplementary Note 12).

Sequence composition and transcription factor analysis. De novo motif analysis was performed using HOMER v.4.8.2 (ref. ⁷²) findMotifGenome.pl tool with the parameter ‘-size given -bits -mask’. Only enriched sequences present in >1.5% of targets were retained. Selection of best matches was performed as follows: all matches with scores >0.80 were included in the table; for those hits without any match >0.80, the top 3 hits were selected and their score was included in the table (Extended Data Figs. 1g and 3e,f).

To assay motif colocalization, we used all motif instances identified in the de novo analysis in primed enhancers. First we used the findMotifGenome.pl tool from HOMER to map all these motif instances in primed enhancers and SRE enhancers (that is, excluding all sites <2 kb from a TSS). Next, the motif colocalization was calculated by counting motif pairs found in each ATAC-seq peak. Significance was determined by Fisher’s exact test comparing the colocalization of motif pairs in distal IREs versus distal SREs. Only significant pairs (Fisher’s exact test, FDR-adjusted $P < 0.001$) were retained (Extended Data Fig. 3g,h).

To evaluate islet-specific transcription factor occupancy, we used ChIP-seq BAM files for PDX1, NKX2.2, FOXA2, NKX6.1 and MAFB⁶. We computed the read coverage in the regions of interest over 10 bp bins. Reads were quantile-normalized, the mean counts in each bin for each transcription factor were calculated and the mean for all transcription factors was plotted (Fig. 2d).

To identify footprints from the ATAC-seq data, we generated tag directories with all ATAC-seq replicates in each condition using HOMER makeTagDirectory. Neo and primed enhancers were centered on the ISRE motif matrix annotated with annotatePeaks.pl with the option ‘-center motif -size given’ and tag means for the 5’ and 3’ read ends were obtained using annotatePeaks.pl with the option ‘-size -100,100 -hist 1 -d tagsDir’. The resulting 5’ ends were plotted using ggplot2 v.3.2.0 (ref. ⁷³) (Fig. 2e).

To create a nonredundant dataset of motifs for the gene regulatory network analysis (Extended Data Fig. 4a), motifs from primed and opening enhancers were reduced to a nonredundant set with the compareMotifs.pl script from HOMER using a similarity score of 0.7 as the threshold for merging similar motifs. The motifs were then mapped to primed and opening enhancers using annotatePeaks.pl.

Infinium MethylationEPIC array. DNA from EC cells exposed or unexposed to IL-1 β and IFN- γ for 48 h, as described earlier (5 replicates per condition), was extracted using QIAamp DNA Mini Kit (QIAGEN); 1 μ g DNA aliquots ($n=10$) were processed for 850K Infinium MethylationEPIC array (Illumina) as described previously⁷⁴.

The resulting array signals were processed and analyzed with the RnBeads R package v.3.2.0 (ref. ⁷⁵). The method used by RnBeads to assess differences between groups consists of fitting a hierarchical linear model (the empirical Bayesian method from the limma package v.3.40.0 (ref. ⁷⁶)) using M values (log of β values) as metrics to measure methylation levels⁷⁷. All P values were corrected for multiple testing using the Benjamini–Hochberg method for controlling the FDR. CpGs

were considered as differentially methylated when FDR-adjusted $P < 0.05$ and the absolute difference in methylation β values between cytokine and control samples was >0.2 (20% change in methylation). Information on the differentially methylated CpGs can be found in Supplementary Table 7.

UMI-4C. UMI-4C was performed as described previously²⁵ with minor modifications (see Supplementary Note 13). To increase molecular complexity, each library was obtained by pooling 5–10 PCRs per viewpoint. The PCR primers used in UMI-4C are listed in Supplementary Table 8. Each library was sequenced to a depth >1 million 75-bp-long paired-end reads using either the NextSeq 550 or HiSeq 2500 platforms.

Paired-end reads were demultiplexed according to the viewpoint sequence using fastq-multx from ea-utils v.1.30 (ref.⁷⁸) and analyzed with the umi4cPackage v.0.0.0.9000 (ref.⁷⁹). 4C tracks were created by selecting viewpoint-specific reads, aligning them to the genome and extracting the number of UMIs using the p4cCreate4CseqTrack function (see quality control statistics in Supplementary Table 9). Cytokine-treated profiles were then scaled to the control profile using the umi4cPackage function p4cSmoothedTrendComp. Profiles were also smoothed based on the total number of UMIs present in a 2 Mb region centered on the viewpoint and excluding the 3 kb around it. The following formula was used to calculate the minimum UMIs needed for smoothing. If the fragment did not reach this minimum, it was merged with successive fragments until a minimum was reached:

$$\text{Minimum UMIs} = \frac{\sum \text{UMIs}_{\text{region}}}{2,000} \times 50$$

To detect differential chromatin contacts we focused on a 2 Mb region centered on the viewpoint, but excluding 1.5 kb on each side of the viewpoint. We then partitioned the region into windows of width proportional to the mean restriction fragment length in the region ($\text{Mean}_{\text{fragment}}$):

$$\text{Width}_{\text{window}} = \text{Mean}_{\text{fragment}} \times 20$$

Differential contact analysis was performed for each of these windows using a chi-squared test, comparing UMIs in such windows with the total number of UMIs in the 2 Mb region. Windows with a chi-squared $P < 0.05$ are highlighted in Fig. 3a,b and Extended Data Fig. 5b–d with small black diamonds. To quantify the chromatin contact changes, we counted the number of cytokine-treated and untreated UMIs for each window and computed their odds ratio (OR) based on the total UMI counts in the region, following the formula:

$$\text{OR}_{\text{window}} = \frac{\text{Ctrl}_{\text{region}} \times \text{Cyt}_{\text{window}}}{\text{Ctrl}_{\text{window}} \times \text{Cyt}_{\text{region}}}$$

where Ctrl and Cyt represent the number of UMIs in unexposed and cytokine-exposed conditions.

Variant set enrichment (VSE) analyses. We used the VSE R package v.0.99 (ref.⁷⁸) to assess the enrichment of T1D and T2D risk variant for IRE and SRE regulatory annotations. We first selected from the NHGRI-EBI GWAS catalog²⁶ all leading SNPs with disease trait matching either 'type 1 diabetes' or 'type 2 diabetes' (24 April 2019). Next, we extended our collection of associated variants to all those in strong linkage disequilibrium ($R^2 \geq 0.8$, EUR) with the lead SNP (source of linkage disequilibrium information, 1000 Genomes Project phase 3 (ref.⁸⁰)). These SNPs and their proxies were used to generate the associated variant set (AVS)⁷⁸, resulting in 83 disjointed regions for T1D and 389 for T2D, after removing shared loci between T1D and T2D. A null distribution or matched random variant set, matched in size and structure to the original AVS, was generated from the 1000 Genomes Project phase 3 by permutating the AVS 500 times. The number of independent SNPs from the AVS overlapping the regulatory annotations was computed and compared with the intersections obtained with the matched random variant set. The enrichment score was defined as the number of s.d. that the overlapping tally deviates from the null overlapping tally median. The exact P value was then calculated by fitting a density function to the null distribution derived from the matched random variant set. This P value was finally corrected for multiple testing using the Bonferroni method. Enrichments or depletions with a Bonferroni-adjusted $P < 0.05$ were considered statistically significant (Fig. 4a and Extended Data Fig. 6a).

T1D-associated regions were generated by selecting all SNPs in strong linkage disequilibrium ($R^2 \geq 0.8$, EUR) with the T1D leading SNPs. We defined the risk loci boundaries using the most upstream and downstream SNPs. Next, we merged the overlapping loci to obtain a total of 71 T1D risk regions. All T1D-associated regions containing IREs and T1D risk variants directly overlapping human islet cytokine-induced regulatory elements are shown in Supplementary Table 3. For this analysis, to extract all possible cytokine-induced regulatory elements located at the T1D risk loci, we used a less stringent set of human islet IREs by lowering the H3K27ac log₂ fold change threshold from 1 to 0.8.

For details regarding the GWAS association analysis, see Supplementary Note 14.

Luciferase reporter assays. For episomal reporter assays in the EC cell line, selected human cytokine-induced regulatory elements regions were first amplified from genomic DNA using primers (Supplementary Table 10) containing the XhoI/HindIII restriction sites. The amplicons were then cloned into the pGL4.23[luc2/ minP] luciferase reporter vector (Promega Corporation) as described previously³¹. Briefly, the amplicon and the vector were simultaneously digested. Next, the vector was dephosphorylated with FastAP (Thermo Fisher Scientific). The DNA was then purified and ligated with a T4 DNA Ligase (Promega Corporation). Next, the generated reporter vectors were transformed into *Escherichia coli* (DH5 α) and purified with the NucleoSpin Plasmid (catalog no. 740588.250; Macherey-Nagel). Site-directed mutagenesis was used to introduce single-nucleotide variants into the generated construct. The variants were generated by PCR using the primers shown in Supplementary Table 10. The parental supercoiled double-stranded DNA was digested with DpnI (catalog no. R0176S; New England Biolabs) 1 h at 37 °C and the constructs were transformed in competent *E. coli* cells (DH5 α) by thermal shock. Finally, the introduced variants were checked using Sanger sequencing.

EC cells were transfected in 24-well plates at a density of 300,000 cells per well, with 200 ng of reporter vectors or empty vectors plus 20 ng of pHL-CMV *Renilla* luciferase to control for transfection efficiency.

Transfections were performed with Lipofectamine 2000 (Thermo Fisher Scientific) for 8 h, according to the manufacturer's instructions. On transfection, the EC medium was supplemented with 2% FCS⁵² and exposed or unexposed to the cytokines for 48 h. After 48 h, cells were assayed using the Dual Luciferase Assay (Promega Corporation), according to the manufacturer's instructions. The luciferase units were measured with the VICTOR Multilabel Plate Reader (PerkinElmer). Firefly luciferase activity was normalized to *Renilla* luciferase activity and then divided by the values obtained for the empty pGL4.23. The assays were performed in at least three independent experiments.

Statistical differences were calculated using a one-way ANOVA. P values were then Bonferroni-corrected.

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

Data availability

Datasets for the IREs are available for download and visualization at the Islet Regulome Browser³¹ (www.isletregulome.com). Raw sequencing reads for the different high-throughput assays can be accessed at the Gene Expression Omnibus with the following identifiers: GSE123404 (ATAC-seq); GSE133135 (H3K27ac data); GSE137136 (RNA-seq); and GSE136865 (UMI-4C). Raw proteomics data can be accessed at the ProteomeXchange with the identifier PXD011902.

Code availability

The code and scripts used in this study are available from the corresponding author upon reasonable request.

References

- Bucher, P. et al. Assessment of a novel two-component enzyme preparation for human islet isolation and transplantation. *Transplantation* **79**, 91–97 (2005).
- Melzi, R. et al. Role of CCL2/MCP-1 in islet transplantation. *Cell Transplant.* **19**, 1031–1046 (2010).
- Eizirik, D. L. et al. The human pancreatic islet transcriptome: expression of candidate genes for type 1 diabetes and the impact of pro-inflammatory cytokines. *PLoS Genet.* **8**, e1002552 (2012).
- Colli, M. L., Moore, F., Gurzov, E. N., Ortis, F. & Eizirik, D. L. MDA5 and PTPN2, two candidate genes for type 1 diabetes, modify pancreatic β -cell responses to the viral by-product double-stranded RNA. *Hum. Mol. Genet.* **19**, 135–146 (2010).
- Ortis, F. et al. Cytokine-induced proapoptotic gene expression in insulin-producing cells is related to rapid, sustained, and nonoscillatory nuclear factor- κ B activation. *Mol. Endocrinol.* **20**, 1867–1879 (2006).
- 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).
- Lavin, Y. et al. Tissue-resident macrophage enhancer landscapes are shaped by the local microenvironment. *Cell* **159**, 1312–1326 (2014).
- Raurell-Vila H., Ramos-Rodríguez M., & Pasquali L. Assay for transposase accessible chromatin (ATAC-Seq) to chart the open chromatin landscape of human pancreatic islets. *Methods Mol. Biol.* **1766**, 197–208 (2018).
- Schmidl, C., Rendeiro, A. F., Sheffield, N. C. & Bock, C. ChIPmentation: fast, robust, low-input ChIP-seq for histones and transcription factors. *Nat. Methods* **12**, 963–965 (2015).
- Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).

55. Dunham, I. et al. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
56. Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
57. Zhang, Y. et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).
58. Juan-Mateu, J. et al. SRp55 regulates a splicing network that controls human pancreatic β -cell function and survival. *Diabetes* **67**, 423–436 (2018).
59. Trapnell, C., Pachter, L. & Salzberg, S. L. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**, 1105–1111 (2009).
60. Harrow, J. et al. GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res.* **22**, 1760–1774 (2012).
61. Anders, S., Pyl, P. T. & Huber, W. HTSeq: a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).
62. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
63. Nakayasu, E. S. et al. MPLEx: a robust and universal protocol for single-sample integrative proteomic, metabolomic, and lipidomic analyses. *mSystems* **1**, e00043-16 (2016).
64. Mayampurath, A. M. et al. DeconMSn: a software tool for accurate parent ion monoisotopic mass determination for tandem mass spectra. *Bioinformatics* **24**, 1021–1023 (2008).
65. Petyuk, V. A. et al. DtaRefinery, a software tool for elimination of systematic errors from parent ion mass measurements in tandem mass spectra data sets. *Mol. Cell. Proteomics* **9**, 486–496 (2010).
66. Kim, S. & Pevzner, P. A. MS-GF+ makes progress towards a universal database search tool for proteomics. *Nat. Commun.* **5**, 5277 (2014).
67. Webb-Robertson, B.-J. M. et al. Bayesian proteoform modeling improves protein quantification of global proteomic measurements. *Mol. Cell. Proteomics* **13**, 3639–3646 (2014).
68. Li, T. et al. GeNets: a unified web platform for network-based genomic analyses. *Nat. Methods* **15**, 543–546 (2018).
69. Li, T. et al. A scored human protein–protein interaction network to catalyze genomic interpretation. *Nat. Methods* **14**, 61–64 (2017).
70. Herwig, R., Hardt, C., Lienhard, M. & Kamburov, A. Analyzing and interpreting genome data at the network level with ConsensusPathDB. *Nat. Protoc.* **11**, 1889–1907 (2016).
71. Liberzon, A. et al. Molecular signatures database (MSigDB) 3.0. *Bioinformatics* **27**, 1739–1740 (2011).
72. Heinz, S. et al. Simple combinations of lineage-determining transcription factors prime *cis*-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).
73. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis* (Springer, 2009).
74. Moran, S., Arribas, C. & Esteller, M. Validation of a DNA methylation microarray for 850,000 CpG sites of the human genome enriched in enhancer sequences. *Epigenomics* **8**, 389–399 (2016).
75. Assenov, Y. et al. Comprehensive analysis of DNA methylation data with RnBeads. *Nat. Methods* **11**, 1138–1140 (2014).
76. Ritchie, M. E. et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).
77. Du, P. et al. Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. *BMC Bioinformatics* **11**, 587 (2010).
78. Aronesty, E. Comparison of sequencing utility programs. *Open Bioinforma. J.* **7**, 1–8 (2013).
79. Ahmed, M. et al. Variant Set Enrichment: an R package to identify disease-associated functional genomic regions. *BioData Min.* **10**, 9 (2017).
80. Auton, A. et al. A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
81. Jordà, M. et al. Upregulation of MMP-9 in MDCK epithelial cell line in response to expression of the Snail transcription factor. *J. Cell Sci.* **118**, 3371–3385 (2005).
82. Brozzi, F. et al. Cytokines induce endoplasmic reticulum stress in human, rat and mouse beta cells via different mechanisms. *Diabetologia* **58**, 2307–2316 (2015).
83. Mularoni, L., Ramos-Rodríguez, M. & Pasquali, L. The pancreatic Islet Regulome Browser. *Front. Genet.* **8**, 13 (2017).

Acknowledgements

L.Pasquali was supported by grants from the Spanish Ministry of Economy and Competitiveness (nos. BFU2014-58150-R and SAF2017-86242-R), Marató TV3 (no. 201624.10) and a young investigator award from the Spanish Society of Diabetes (Ayuda SED a Proyectos de Investigación, no. 2017-SED). L.Pasquali is a recipient of a Ramon y Cajal contract from the Spanish Ministry of Economy and Competitiveness (no. RYC-2013-12864). The Pasquali lab is further supported by Instituto de Salud Carlos III (no. PIE16/00011). M.R. is supported by an FI Agència de Gestió d'Ajuts Universitaris i de Recerca PhD fellowship (no. 2019FI_B100203). J.J. was supported by a Marie Skłodowska-Curie Actions fellowship grant from the Horizons 2020 European Union (EU) Programme (project no. 660449). M.I.A. was supported by a FRIA fellowship from the Fonds de la Recherche Scientifique (FNRS) (no. 26410496). Human islets were provided through the European Consortium for Islet Transplantation distribution program for basic research supported by JDRF (no. 31-2008-416). D.L.E. was supported by the Walloon Region through the FRFS-WELBIO Fund for Strategic Fundamental research (no. CR-2015A-06s and CR-2019C-04) and by grants from the Fonds National de la Recherche Scientifique (no. T003613F), the Horizon 2020 Programme (project T2Dsystems, no. GA667191), Brussels Capital Region Innoviris (project DiaType, no. 2017-PFS-24), Dutch Diabetes Research Foundation (project Innovate2CureType1, DDRF; no. 2018.10.002) and the Innovative Medicines Initiative 2 Joint Undertaking (project INNODIA, no. 115797). This Innovative Medicines Initiative 2 Joint Undertaking receives support from the EU's Horizon 2020 Research and Innovation Programme and European Federation of Pharmaceutical Industries and Associations, JDRF and the Leona M. and Harry B. Helmsley Charitable Trust (project INNODIA, no. 115797). T.O.M. and D.L.E. were supported by a grant from the National Institutes of Health–National Institute of Diabetes and Digestive and Kidney Diseases–Human Islet Research Network Consortium (no. 1UC4DK104166-01). Part of the work was performed at the Environmental Molecular Sciences Laboratory, a US Department of Energy national scientific user facility located at Pacific Northwest National Laboratory. Battelle operates the Pacific Northwest National Laboratory for the Department of Energy (contract no. DE-AC05-76RLO01830). We thank J. Mercader (Broad Institute) and M. Guindo Martínez (Barcelona Supercomputing Center) for helpful discussions regarding GWAS enrichment analyses and A. Castela (Université Libre de Bruxelles Center for Diabetes Research) for helping with the glucose-stimulated insulin secretion experiments.

Author contributions

L.Pasquali, M.R. and D.L.E. designed the experiments. H.R., M.L.C., M.A., J.J., R.N. and E.N. performed and analyzed the experiments. M.R. performed the bioinformatic analyses with contribution from M.S., J.-V.T., B.W. and J.I. L.Piemonti, P.M., M.E. and T.O.M. provided materials and resources. L.Pasquali, D.L.E., T.O.M. and J.A.T. supervised the study. L.Pasquali and M.R. coordinated and conceived the project, and wrote the manuscript with contribution from D.L.E. All authors reviewed the final manuscript.

Competing interests

The authors declare no competing interests.

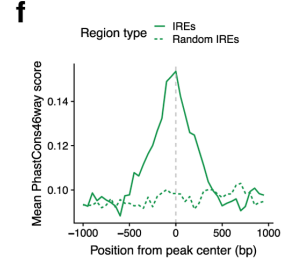
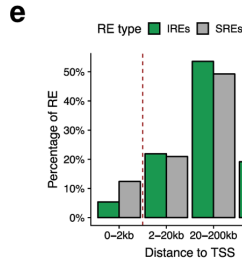
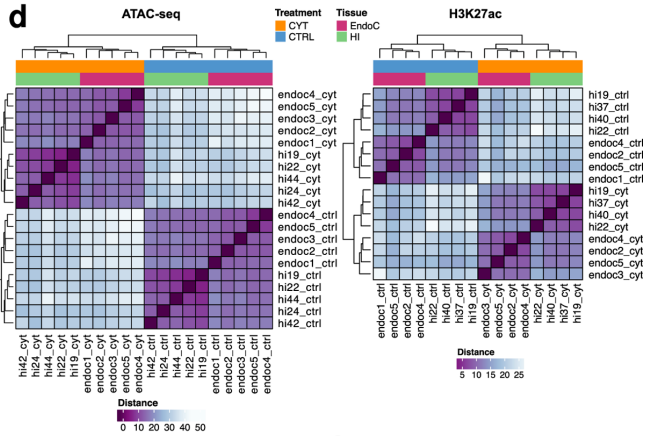
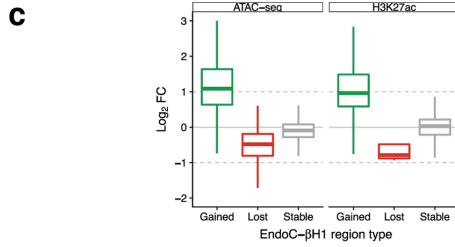
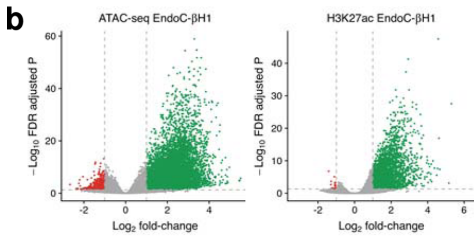
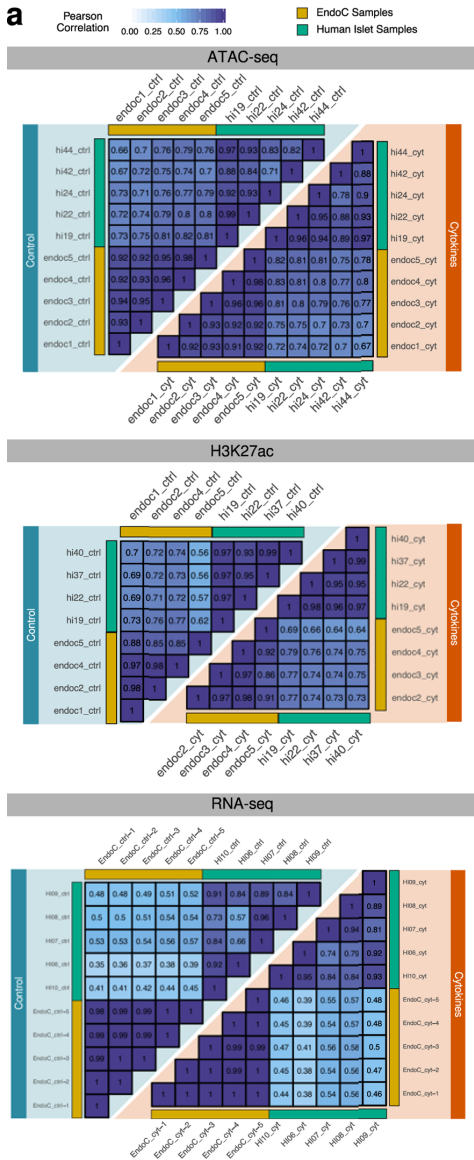
Additional information

Extended data is available for this paper at <https://doi.org/10.1038/s41588-019-0524-6>.

Supplementary information is available for this paper at <https://doi.org/10.1038/s41588-019-0524-6>.

Correspondence and requests for materials should be addressed to L.P.

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



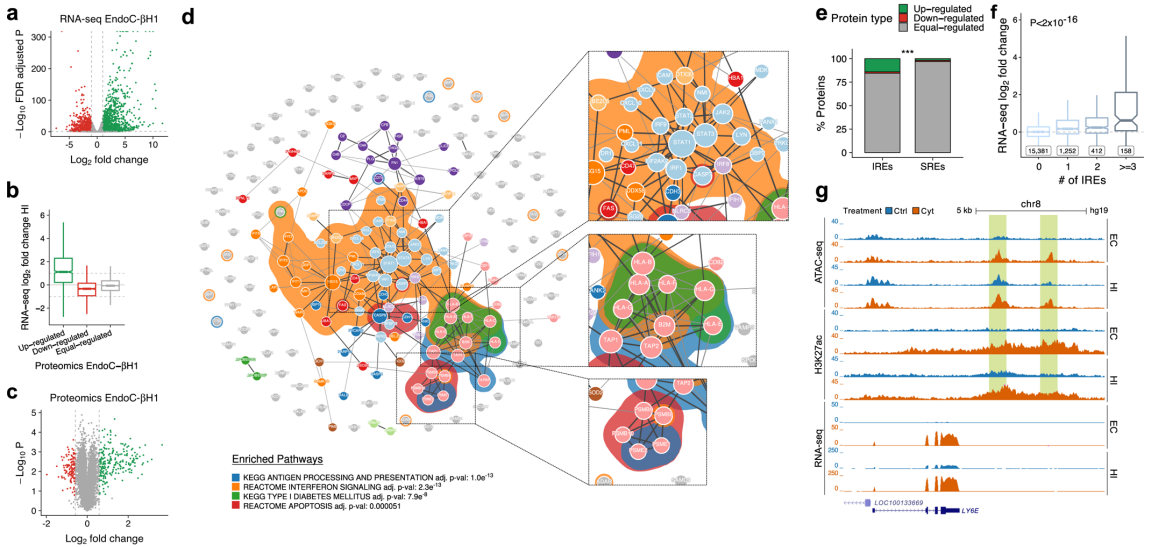
g

IREs (3,009 regions)

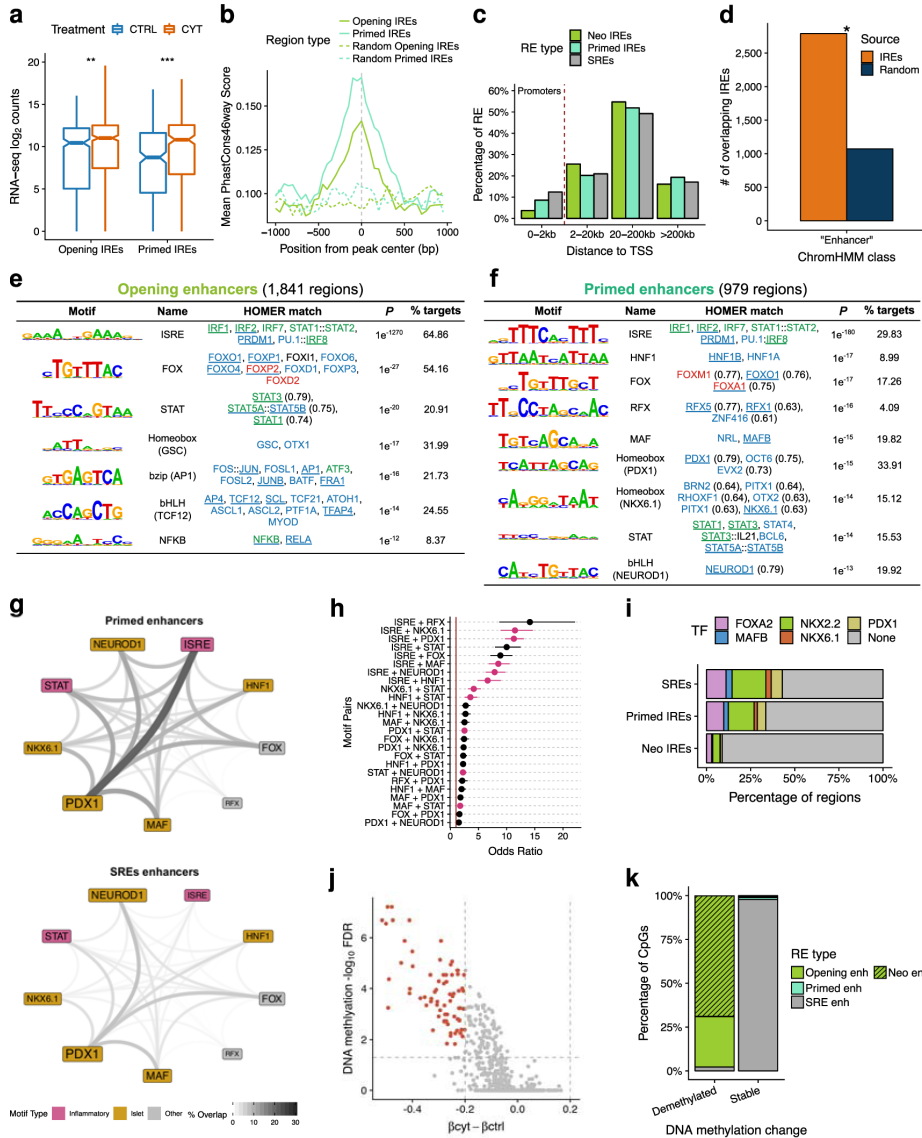
Motif	Name	HOMER match	P	% targets
	ISRE	IRF1, IRF2, STAT1::STAT2, IRF7, PRDM1, IRF8	$1e^{-1396}$	51.71
	FOXP	FOXP3, FOXJ1, FOXP1, FOXL1, FOXD2, FOXO6, FOXO4, FOXD1, FOXP2	$1e^{-39}$	44.13
	STAT	STAT4, STAT1, STAT3::IL21, BCL6, STAT5A::STAT5B	$1e^{-31}$	15.29
	HNF1	HNF1B, HNF1A	$1e^{-29}$	4.62
	bHLH (NEUROD1)	NEUROD1, OLIG2, TCF21, BHLHA15, TAL1::TGFB3, ATCH1, ASCL1, TCF12, MYOG	$1e^{-25}$	29.88
	NFKB	RELA, NFKB	$1e^{-21}$	5.85
	bZIP (AP1)	FOSL1, FOS::JUN, AP1, FOSL2, ATF3, JUNB, BATF	$1e^{-16}$	17.35
	MAF	NRL, TBX20, MAFB, MAFA, MAFK, MAFK	$1e^{-16}$	16.38
	IRF	IRF4 (0.76), RUNX1 (0.69)	$1e^{-13}$	11.67

Extended Data Fig. 1 | See next page for caption.

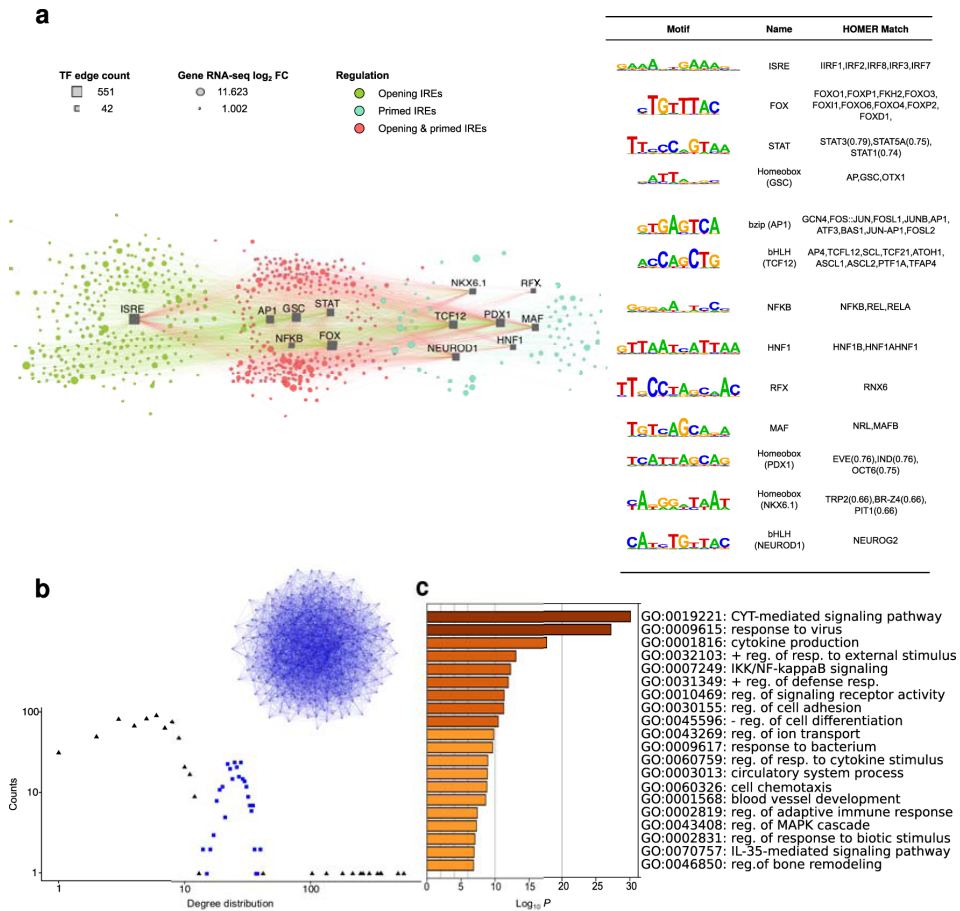
Extended Data Fig. 1 | Chromatin characterization of human pancreatic β cells exposed to pro-inflammatory cytokines. **a**, Pearson correlation values between replicates in different assays and conditions (see Supplementary Note 2). **b**, Volcano plots of ATAC-seq (left) and H3K27ac ChIP-seq (right) changes obtained after exposure of EndoC- β H1 to IFN- γ and IL-1 β ; green and red dots correspond to sites with absolute \log_2 fold change > 1 and FDR adjusted $P < 0.05$ as calculated by fitting a negative binomial model in DESeq2. Chromatin changes are classified as 'gained' and 'lost' chromatin sites whereas non-significant changes are defined as 'stable'. **c**, Chromatin accessibility and H3K27ac enrichment changes observed in EndoC- β H1 are largely replicated in human pancreatic islets as illustrated by the distribution of \log_2 fold change at regions as classified in **b** in EndoC- β H1. Dotted lines indicate \log_2 fold change thresholds (absolute \log_2 fold change > 1). Box plot limits show upper and lower quartiles, whiskers extend to 1.5 times the interquartile range and the notch represents the confidence interval around the median. **d**, Hierarchical clustering using normalized ATAC-seq and H3K27ac read counts at EndoC- β H1 IREs shows that samples cluster according to treatment, suggesting that the differences caused by the proinflammatory cytokines are greater than those derived by the sample heterogeneity. HI = Human pancreatic islets, EndoC = EndoC- β H1 **e**, Distribution of distances to nearest TSS for the different types of regulatory elements, showing that IREs, compared with stable regulatory elements (SREs), are preferentially located distally to TSS. **f**, Mean sequence conservation score of IREs and a randomized set of IREs in placental mammals. Peaks were extended from the center 1 kb to each direction and mean score was calculated in 50 bp windows. **g**, Sequence composition analysis of IREs ($n = 3,009$) illustrating the top identified *de novo* motifs. Colors for matched genes correspond to RNA-seq (name) or protein (underlined) status (red = down-regulated, blue = equal-regulated, green = up-regulated, black/no line = not expressed/detected).



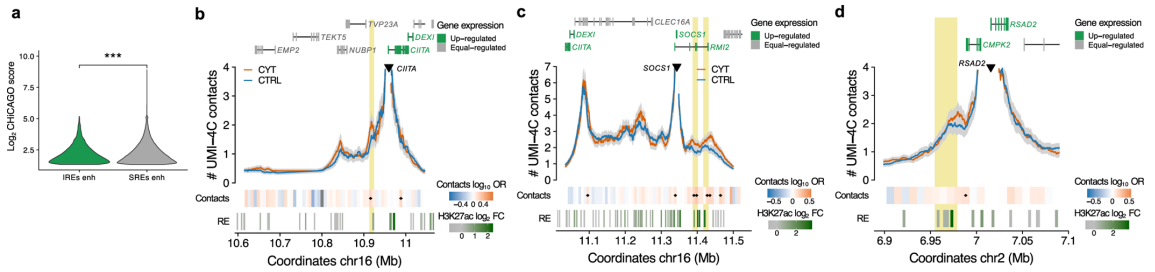
Extended Data Fig. 2 | Exposure to pro-inflammatory cytokines drives changes in the transcriptome and proteome of pancreatic β cells. **a**, Volcano plot of RNA-seq genes, showing up-regulated genes (green) and down-regulated genes (red) upon exposure of EndoC- β H1 to cytokines. Vertical lines indicate the log₂ fold change threshold (absolute log₂ fold change > 1) and horizontal line indicates the FDR adjusted *P* cutoff for significance (FDR adjusted *P* < 0.05) calculated by fitting a negative binomial model in DESeq2. **b**, Distribution of RNA-seq counts in human islet samples in the genes previously classified as up, down or equal-regulated in EndoC- β H1 cells. Boxplot limits show upper and lower quartiles, whiskers extend to 1.5 times the interquartile range and the notch represents the confidence interval around the median. **c**, Volcano plot for multiplex proteomics, showing in green the up-regulated proteins and in red the down-regulated, which have a *Q*-value < 0.1 and absolute log₂ fold change > 0.58. Vertical lines indicate the log₂ fold change thresholds. **d**, Protein-protein Interaction (PPI) network generated from up-regulated proteins after cytokine exposure. Node color indicates belonging to same interacting community and background corresponds to specific pathway enrichment. **e**, Proportion of up, equal or down-regulated proteins encoded by genes located <15 kb from IREs or SREs. *** Chi-squared test *P* < 0.001. **f**, An additive effect on gene up-regulation was observed for multiple IREs located at <40 kb of a gene. Box plot limits show upper and lower quartiles, whiskers extend to 1.5 times the interquartile range and the notch represents the confidence interval around the median. ANOVA *P* < 2.2 × 10⁻¹⁶. **g**, View of the *LY6E* locus, whose expression is induced after cytokine exposure and is coupled with chromatin changes in the vicinity.



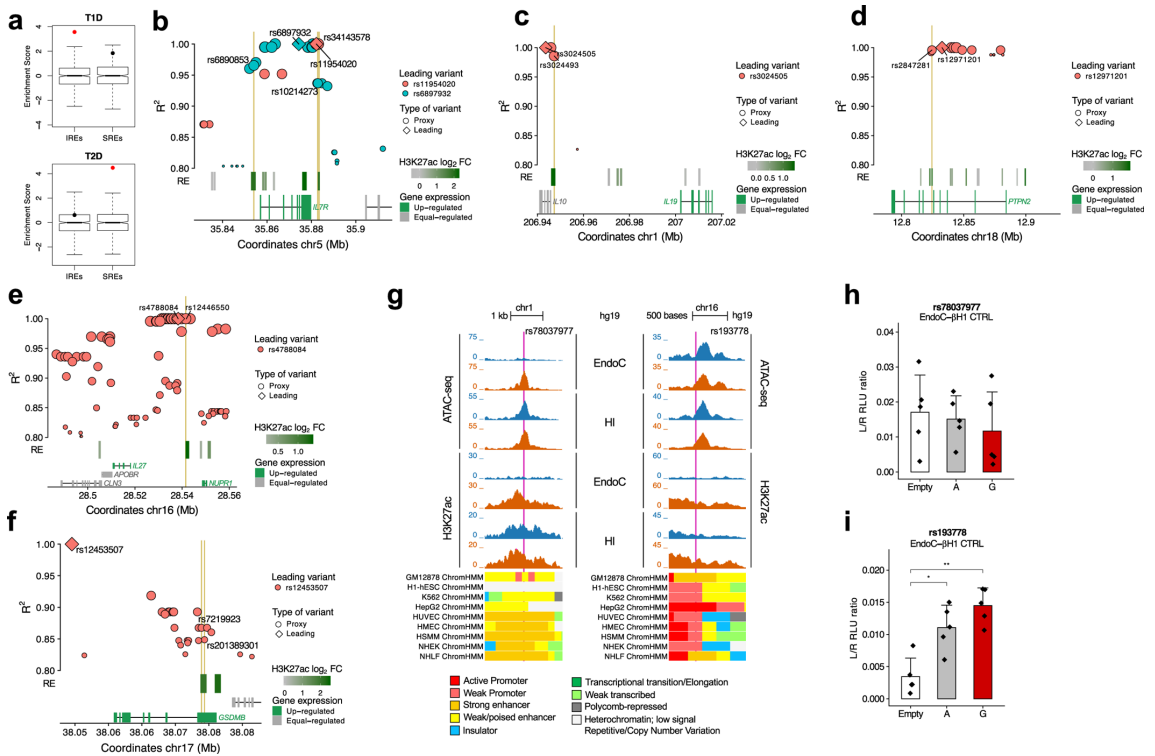
Extended Data Fig. 3 | Characterization of β -cell IREs. **a**, Genes associated to different classes of IREs (classified as in Fig. 2a) show cytokine-induced expression in EndoC- β H1. CYT = cytokine exposed, CTRL = control. Boxplot limits show upper and lower quartiles, whiskers extend to 1.5 times the interquartile range and notch represents the median confidence interval. ***Wilcoxon test $P < 0.001$. **b**, Sequence conservation score of IREs and a corresponding randomized set used as control. **c**, Distribution of distances to nearest TSS of the different classes of open chromatin sites. Line indicates the threshold used to classify them as 'promoters'. **d**, Number of IREs overlapping regions annotated as 'Strong' or 'Weak' enhancers by ENCODE ChromHMM. *Chi-squared $P < 2 \times 10^{-16}$. **e**, **f**, Top hits for *de novo* motif analysis in opening (**e**) and primed enhancers (**f**). Colors for matched genes correspond to RNA-seq (name) or protein (underlined) status (red = down-regulated, blue = equal-regulated, green = up-regulated, black/no-line = not-expressed/detected). **g**, Diagram showing the percentage of colocalization between the TF binding sites identified by *de novo* motif analysis in SRE and primed enhancers (that is excluding sites < 2Kb from a TSS). Label size indicates number of regions containing the TF binding sites and line width/intensity percentage of regions in which two motifs colocalize. **h**, Odds-ratio for finding a motif pair in the same enhancer in primed vs. SRE. Only significant pairs (FDR-adjusted Fisher's Exact test $P < 0.001$) are shown. Immune and islet-specific TF motifs colocalize more often in primed compared to SRE chromatin sites. **i**, Percentage of overlap between EndoC- β H1 different classes of open chromatin and islet-specific TFs obtained by ChIP-seq in untreated human islets. **j**, Volcano plot showing differentially methylated sites (depicted in red) in EndoC- β H1 exposed or not to cytokines. Dotted lines indicate the threshold for methylation differences or significance using limma moderated t-test. **k**, Distribution of demethylated and stable CpGs according to different classes of open chromatin.



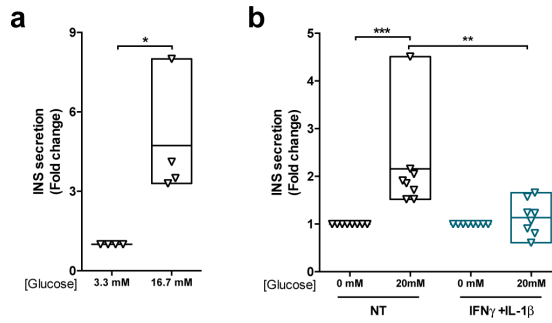
Extended Data Fig. 4 | Deconstructing cytokine induced cis-regulatory networks in β cells. a, Gene Regulatory Network (GRN) derived from IREs and their putative target genes. Squares represent the IREs inferred TF binding sites (motifs logos and TF matches are shown on the right side) and the ellipses represent their putative target genes (see Methods). The size of the squares reflects the number of connections (edge count) while the gene node size reflects the log₂ fold change of RNA expression after cytokine exposure. The resulting GRN is an interconnected scale-free network composed of 648 nodes and 3,589 edges. Genes regulated exclusively by primed IREs are represented in blue while green depicts opening IREs regulated genes. Red denotes genes regulated by both types of IREs. In each of these three groups the representation of the hierarchy is based on the principle of network centrality where authoritative nodes are located more proximal to the core. **b**, Comparison between the degree distribution of the observed GRN (black triangles) and a random generated network (blue squares) having the same number of nodes and edges. The bell-shaped degree distribution of random graph denotes a statistically homogeneity in the degree of small and large nodes. In contrast, the observed network showed a high frequency of small degree nodes and a low frequency of highly connected nodes as is typical of a scale-free network. **c**, Bar plot of gene ontology biological process enrichment analysis. Gene-ontology analysis was performed using all target genes in the GRN. Functional enrichment analysis was performed by Metascape (<http://metascape.org>). Only terms with $P < 0.001$ and with at least 3 enriched genes were considered as significant. Color is proportional to their P values.



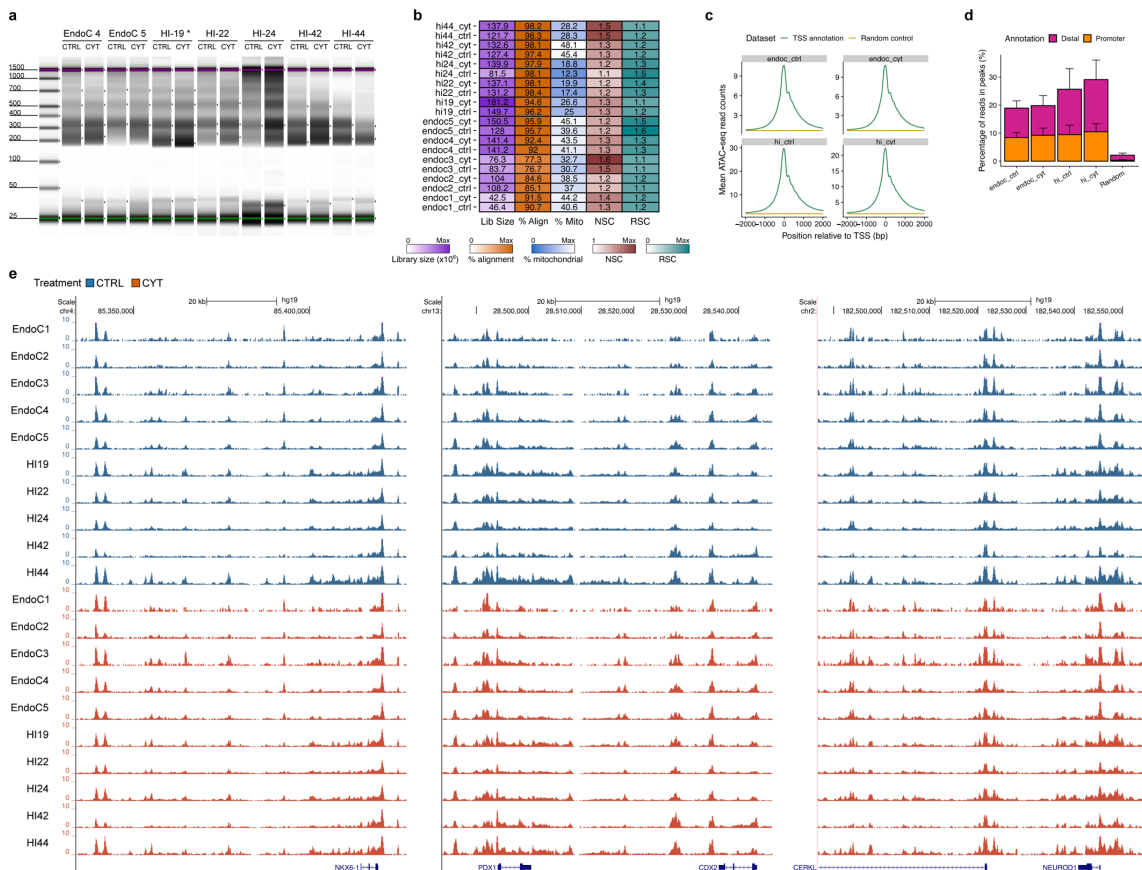
Extended Data Fig. 5 | 3D chromatin changes induced by exposure of human islets to pro-inflammatory cytokines. a, Violin plots showing the distribution of CHiCAGO scores of contacts, detected by promoter capture HiC experiments in untreated human islets²⁴, between stable and induced enhancers and their target genes. SREs engage chromatin contacts with higher interaction scores compared to those detected for IREs. **** Wilcoxon test $P < 0.001$. **b, c, d**, Views of the 3D chromatin contacts of *CIITA* (**b**), *SOCS1* (**c**) and *RSAD2* (**d**) promoters obtained by UMI-4C performed in islets exposed or not to pro-inflammatory cytokines. In yellow we highlight those IREs that gain contacts with the up-regulated gene promoter. A heatmap under the 4C track represents the \log_{10} odds ratio (OR) of the UMI-4C contacts difference in cytokine vs. control and a small black diamond on top of the contact heatmap indicates a significant difference in contacts between cytokine-treated and control samples (Chi-squared $P < 0.05$). ATAC-seq peaks are represented by rectangles, shaded from gray to green proportionally to the cytokine-induced H2K27ac \log_2 fold change observed at that site.



Extended Data Fig. 6 | Cytokine-induced islet regulatory elements are enriched in T1D associated variants. **a**, Endo-βH1 cytokine-induced regulatory elements (IREs) overlap more often than expected T1D associated variants while the opposite is true for T2D. Endo-βH1 cytokine-invariant regulatory elements (SREs) are instead enriched for T2D, but not T1D associated variants. Each dot denotes the Varien Set Enrichment (VSE) score in IREs or SREs regions. Boxplot shows the enrichment distribution of the matched null permuted data sets. Red dots indicate that the difference is statistically significant as determined by VSE (Bonferroni adjusted $P < 0.05$). Box plot limits show upper and lower quartiles, whiskers extend to 1.5 times the interquartile range and the notch represents the confidence interval around the median. **b-f**, Representative regional plots of different T1D risk loci containing T1D variants overlapping IREs and up-regulated genes. R^2 values are based on 1KG phase 3 EUR and the leading SNPs in the locus is represented by a diamond. If different leading variants are present in the same locus, their proxies are depicted in different colors. Yellow squares highlight those variants that overlap a human islet IRE. IREs are depicted as boxes, with the filling color corresponding to the H3K27ac log₂ fold change. **g**, The IRE bearing the T1D associated variant rs78037977 is marked by the ENCODE ChromHMM classification as a 'strong enhancer' (orange) in other non-β-cell lines (left). ENCODE ChromHMM classification in non-β-cell lines for the IRE bearing the T1D associated variant rs193778. **h**, **i**, Allele-specific luciferase experiments for rs78037977 (**h**) and rs193778 (**i**) in untreated EndoC-βH1. ANOVA followed by Bonferroni correction * $P < 0.05$; ** $P < 0.01$. Bars represent mean \pm sd.



Extended Data Fig. 7 | Human islets and EndoC- β H1 Glucose-Stimulated Insulin Secretion (GSIS). GSIS was assessed, in pancreatic human islets (**a**) and EndoC- β H1 cells (**b**). In the case of EndoC- β H1 cells, the experiments were performed upon exposure or not to IFN γ (1000 U/ml) +IL1 β (50U/ml) for 48 h. Data are mean plus range of four to eight independent experiments, and are expressed as the ratio between glucose stimulated and basal insulin secretion. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, for the indicated comparisons (paired t test (a) or ANOVA followed by Bonferroni correction (b)). NT = Non treated.



Extended Data Fig. 8 | ATAC-seq quality control. **a**, Agilent TapeStation profiles obtained by chromatin tagmentation of human islets and EndoC-β1 samples showing the laddering pattern of ATAC-seq libraries. The band sizes correspond to the expected nucleosomal pattern. *Notice that samples HI-19 CTRL and CYT were used as examples to illustrate the expected fragment distribution pattern in ATAC-seq experiments in Raurell-Vila et al.⁵² **b**, Summary of per-replicate sequencing metrics, showing total library sizes, percentage of aligned reads, percentage of mitochondrial aligned reads, normalized strand cross-correlation coefficient (NSC, values significantly lower than 1.1 (<0.05) tend to have low signal to noise or few peaks) and relative strand cross-correlation coefficient (RSC, values significantly lower than 1 (<0.8) tend to have low signal to noise). **c**, TSS enrichment over a 4 kb window centered on genes TSS compared to a set of genes randomized along the genome, showing the expected pattern of open chromatin centered on the TSS. **d**, Percentage of total reads found at called open chromatin peaks classified as distal (>2 kb from TSS) or promoters (≤2 kb from TSS) compared to a randomized set of peaks. **e**, UCSC views at islet-specific loci (*NFKX6.1*, *PDX1* and *NEUROD1*) showing the high reproducibility of ATAC-seq profiles among independent replicates.

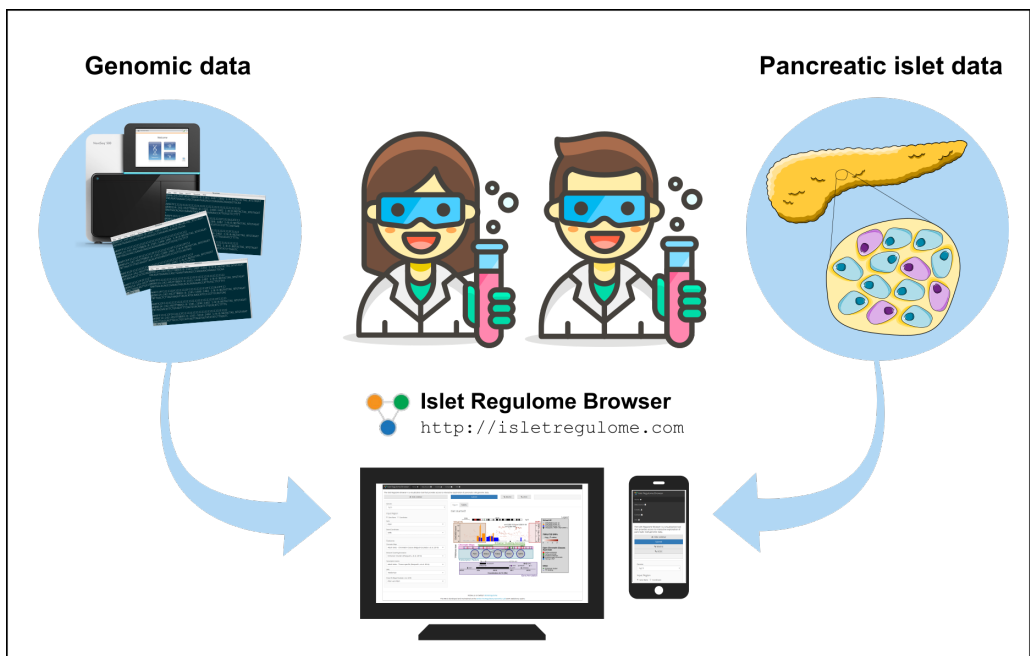
8

The Pancreatic Islet Regulome Browser

In the following article, we describe the **Islet Regulome Browser (IRB)**, a web application created with the aim of gathering relevant pancreatic islet genomic data and, at the same time, making it accessible to researchers with no bioinformatic expertise. As with other bioinformatic applications, since the article publication, the structure and overall implementation of the tool was changed and improved in order to keep up with new datasets and technologies. See section *State of the art of the Islet Regulome Browser* for an updated description of the IRB implementation.

Mularoni L., **Ramos-Rodríguez M.** and Pasquali L. The Pancreatic Islet Regulome Browser. *Front. Genet.* **8** (2017). <https://doi.org/10.3389/fgene.2017.00013>

8.1 Graphical abstract



8.2 Highlights

- The IRB is an online tool with a user-friendly interface that provides access to pancreatic islet genomic datasets.
- This tool allows interactive visualization of the genomic data. The user can select a genomic region or gene of interest, and several datasets to be visualized simultaneously.
- The plot generated by the IRB provides an integrated view of different types of genomic, epigenomic and GWAS data.
- The IRB is freely accessible at isletregulome.com.



The Pancreatic Islet Regulome Browser

Loris Mularoni^{1,2}, Mireia Ramos-Rodríguez³ and Lorenzo Pasquali^{3,4,5*}

¹ Research Program on Biomedical Informatics, IMIM Hospital del Mar Medical Research Institute and Universitat Pompeu Fabra, Barcelona, Spain, ² Biomedical Genomics, Institute for Research in Biomedicine, The Barcelona Institute of Science and Technology, Barcelona, Spain, ³ Program of Predictive and Personalized Medicine of Cancer, Germans Trias i Pujol University Hospital and Research Institute, Badalona, Spain, ⁴ Germans Trias i Pujol Campus, Josep Carreras Leukaemia Research Institute, Badalona, Spain, ⁵ CIBER de Diabetes y Enfermedades Metabólicas Asociadas, Barcelona, Spain

The pancreatic islet is a highly specialized tissue embedded in the exocrine pancreas whose primary function is that of controlling glucose homeostasis. Thus, understanding the transcriptional control of islet-cell may help to puzzle out the pathogenesis of glucose metabolism disorders. Integrative computational analyses of transcriptomic and epigenomic data allows predicting genomic coordinates of putative regulatory elements across the genome and, decipher tissue-specific functions of the non-coding genome. We herein present the Islet Regulome Browser, a tool that allows fast access and exploration of pancreatic islet epigenomic and transcriptomic data produced by different labs worldwide. The Islet Regulome Browser is now accessible on the internet or may be installed locally. It allows uploading custom tracks as well as providing interactive access to a wealth of information including Genome-Wide Association Studies (GWAS) variants, different classes of regulatory elements, together with enhancer clusters, stretch-enhancers and transcription factor binding sites in pancreatic progenitors and adult human pancreatic islets. Integration and visualization of such data may allow a deeper understanding of the regulatory networks driving tissue-specific transcription and guide the identification of regulatory variants. We believe that such tool will facilitate the access to pancreatic islet public genomic datasets providing a major boost to functional genomics studies in glucose metabolism related traits including diabetes.

Keywords: pancreatic islet, epigenome, non-coding DNA, gene regulation, β -cells

OPEN ACCESS

Edited by:

Guy A. Rutter,
Imperial College London, UK

Reviewed by:

Apiwat Mutirangura,
Chulalongkorn University, Thailand
Qiao Li,
University of Ottawa, Canada

*Correspondence:

Lorenzo Pasquali
lpasquali@igt.p.cat

Specialty section:

This article was submitted to
Epigenomics and Epigenetics,
a section of the journal
Frontiers in Genetics

Received: 31 October 2016

Accepted: 23 January 2017

Published: 14 February 2017

Citation:

Mularoni L, Ramos-Rodríguez M and
Pasquali L (2017) The Pancreatic Islet
Regulome Browser.
Front. Genet. 8:13.
doi: 10.3389/fgene.2017.00013

INTRODUCTION

During the last decade, the advent of high-throughput “-omics” technologies, has greatly promoted advances in the study of human diseases at the genomic, transcriptomic, and epigenomic levels. Sequence databases and software analysis tools are now crucial tools for molecular biologist to understand the molecular mechanisms underlying tissue-specific functions. Nevertheless, the systematic acquisition of large bioinformatic datasets has created a tremendous gap between available data and their biological interpretation. Frameworks to access processed and integrated genomic datasets may assist, computational and non-computational scientists, to bridge this gap and provide understanding and biological interpretations to the regulatory and transcriptional complexity of the genome.

In this context genome browsers are key tools in the accomplishment of this task. The UCSC Genome Browser (Speir et al., 2016), ENSEMBL (Yates et al., 2016) and NCBI's Sequence

Viewer (Wolfsberg, 2011), for example, provide to the research community a wealth of integrated information and represent nowadays essential instruments to assist the interpretation of genomic data.

The pancreatic islets of Langerhans constitute an endocrine tissue embedded in the exocrine pancreas and represent the sole source of insulin in the human body. Pancreatic islets play a crucial role in maintaining normal glucose homeostasis, and islet-cell dysfunction and/or reduction in islet-cell mass are key elements in the development of diabetes mellitus. For these reasons, understanding the regulatory networks controlling the tissue-specific expression of pancreatic islets, is key to shed light on the molecular mechanisms underlying diabetes.

Large consortia such as ENCODE (Dunham et al., 2012) and the Epigenome Roadmap (Bernstein et al., 2010) provided extensive epigenetics maps allowing annotation of the non-coding regions of the human genome for a large amount of cell lines and tissues including several relevant to diabetes such as adipose tissue and skeletal muscle, while other less accessible primary tissues such as the endocrine pancreas were not prioritized in these studies. For their central role in diabetes pathogenesis, different laboratories embarked in profiling the transcriptomic and epigenetic landscape of human pancreatic islet-cells (Bhandare et al., 2010; Gaulton et al., 2010; Stitzel et al., 2010; Parker et al., 2013; Dayeh et al., 2014; Pasquali et al., 2014) in an ongoing effort to shed light on the pancreatic islets tissue-specific gene regulation. Free access to such data represents an invaluable opportunity for the research community to dissect the molecular mechanisms of glucose metabolism diseases (Ashcroft and Rorsman, 2012). Nevertheless, these datasets are deposited in different repositories, often in bulky raw format files, thus of difficult immediate access especially to non-bioinformatic users.

Here we present the Islet Regulome Browser, an intuitive web tool providing access to interactive exploration of a wealth of pancreatic islet genomic data allowing the visualization of different classes of regulatory elements and transcription factor binding sites obtained from experiments performed by different labs worldwide. The Islet Regulome Browser is addressed to molecular biologists, human geneticist and clinicians with or without bioinformatics skills.

MATERIALS AND METHODS

The overall structure of the Islet Regulome Browser is illustrated in **Figure 1**.

The Islet Regulome Browser internal structure is composed of three main components: (a) the database, which is saved in binary format as RData objects and tabix indexed files, (b) the code for computing the graphic image, written in R (Rizzo¹), and (c) the interface and the framework for the web service, written in Python (<http://www.python.org>). The Islet Regulome Browser is compatible with all the most popular web browsers and operative systems. It can be explored via web at <http://www.isletregulome.com> or can be installed in a workstation or laptop through the Python package management system with the command *pip*

¹Statistical computing with R.

install regulome_web. The source code is available under the MIT license at https://bitbucket.org/batterio/regulome_web.

Web Interface and Plot Generation

The code for running the Islet Regulome Browser is composed of two main blocks that interact with each other: a Python framework that creates the web interface and retrieves the user input, and the R code which generates the plot and the tables.

On the server side the Islet Regulome Browser is managed by the Flask framework (<http://flask.pocoo.org/>). The web interface allows users to generate plots and tables by querying for a gene name or for a specific genomic region. In addition, users can customize their analyses by choosing which datasets to use. The interactivity of the web application is achieved by using Brython (<http://brython.info/>), a Python 3 implementation for client-side web programming. The options selected by the user are forwarded to the R script that generates both the plot and the result tables (**Figure 1**).

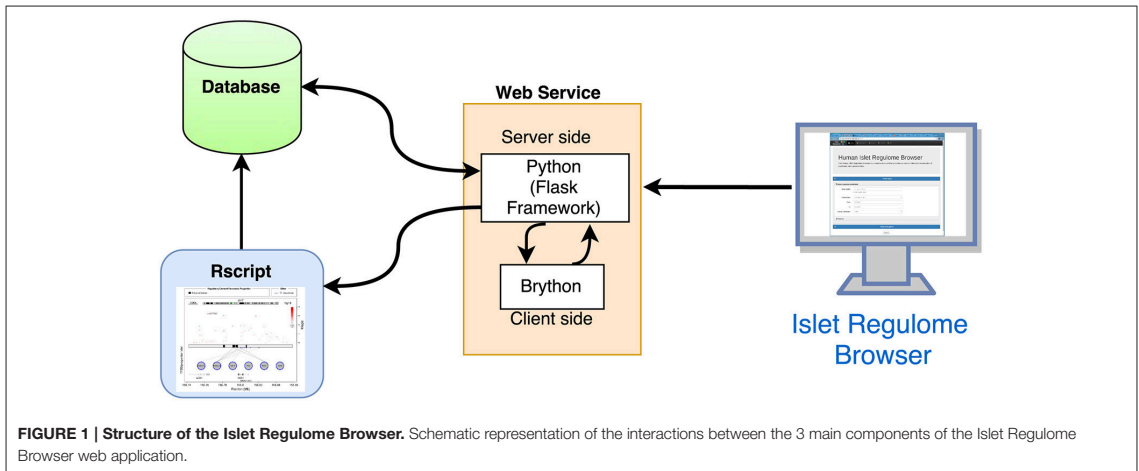
The plot is generated by an R script (R version 3.3.1) that takes as input the user specified features, such as the genomic location and the datasets to use. Several Bioconductor packages have been used to read the database and render the final plot: *Rsamtools* (Morgan et al., 2016), *rtracklayer* (Lawrence et al., 2009), and *Sushi* (Phanstiel et al., 2014). A plot may also be generated via command line, using the code as a stand-alone script.

The plots are converted from PDF to PNG format by the ImageMagick converter tool (<http://www.imagemagick.org/>) and cached, along with the produced text tables. This allows to rapidly load a plot, instead of generating a new one, in case the same query is repeated. The cache is not used when the users upload their own data.

Code Structure and Development

We deposited the Islet Regulome Browser code in a publicly accessible Bitbucket repository (https://bitbucket.org/batterio/regulome_web). Even though the web application can be explored at <http://www.isletregulome.com>, we created a Python package to easily install the Islet Regulome Browser on a personal computer. The recommended way to install the package is by using the Python package management system (*pip install regulome_web*). The main requirement for the web application is Python (version 3.5 or above), R (version 3.3.1 and above), and ImageMagick (<http://www.imagemagick.org>). Other Python related dependencies are listed in the “requirement.txt” file, however, by using the Python package management system all the libraries are automatically installed. Once installed, the Islet Regulome Browser can be executed with the command *regulome_web*. The program has two sub-commands: *init* and *start*. *regulome_web init* will create several folders following a structure required by the program, and a configuration file that needs to be modified by the user. The sub-command *regulome_web start* runs the Islet Regulome Browser web server, locally accessible at the url *localhost:5000*.

The R code to render the plot contains two main scripts: (1) *plot_IRB_main.R*, which is the script that needs to be executed to call all other scripts and to draw each part of the plot. (2), *plot_IRB_config.R* contains all configuration variables, including



the path to the database. The R script are integrated in the web application but they can also be used via command prompt as a stand-alone program.

Database

Central to the system is the database, which stores the genomic annotations, chromatin tracks, genome-wide association study (GWAS) variants and transcription factor binding sites that may be visualized by the browser. The publicly available data that can be currently visualized by the Islet Regulome Browser consists of transcription factor binding sites obtained from ChIP-seq experiments in adult human pancreatic islets (PDX1, FOXA2, NKX2.2, NKX6.1, and MAFB) (Pasquali et al., 2014) and pancreatic progenitors (PDX1, FOXA2, ONECUT1, HNF1B, and TEAD1) (Cebola et al., 2015); open chromatin classes and chromatin states in adult pancreatic islets (Parker et al., 2013; Pasquali et al., 2014), enhancer predictions in pancreatic progenitors (Cebola et al., 2015); enhancer clusters and stretched enhancers in adult pancreatic islets (Parker et al., 2013; Pasquali et al., 2014); open chromatin profiles of α - and β -cells FACS purified from adult human pancreatic islets (Ackermann et al., 2016); expression data obtained from RNA-seq experiments including coding (Morán et al., 2012) and non-coding RNA in adult pancreatic islets (Akerman et al., in press), and datasets for genome wide association studies for type 2 diabetes, DIAGRAM (Cho et al., 2012) and fasting glycemia, MAGIC (Scott et al., 2012).

While the above description summarizes the data currently available, the Islet Regulome Browser is a dynamic project. We periodically revise the database and the literature with the aim of providing the most updated and relevant datasets to the pancreatic islet community. We will ensure the future maintenance the Islet Regulome Browser and will interact with other members of the pancreatic islets community to collect their feedback and improve the user interaction with browser.

For each dataset visualized in the browser we provide, in the “Data Source” page, full reference of publication as well as links to the repositories where the raw data was deposited for bulk download.

RESULTS

The Islet Regulome Browser (<http://www.isletregulome.com>) provides interactive access to a wealth of information, allowing the visualization of GWAS variants, different classes of regulatory elements, together with enhancer clusters, stretch-enhancers and transcription factor binding sites in pancreatic progenitors and adult human pancreatic islets. Integration and visualization of such data may help in the interpretation of the regulatory networks driving tissue-specific transcription and guide the identification of regulatory variants.

From the initial page (Figure 2) a plot can be generated by selecting a valid gene name or an absolute chromosomal location by specifying the genomic coordinates (chromosome, start, and end). The available human builds are: hg18, hg19 (default), and hg38. The plot can be extended at both sides of the gene/location by selecting a range that by default is 50 Kb. To limit the computational load on the server, on the web applications, plots can span a maximum 5 Mb of genomic space and a minimum of 10 bp. These restrictions can be changed in a local installation of the Islet Regulome Browser. Four major track types can be loaded to obtain the desired plot. (1) Tracks named “chromatin maps” refer to genomic maps of regions that may be involved in gene transcription regulation. Such publicly available maps were inferred from experimental datasets such as open chromatin and histone modification profiles, performed in adult human pancreatic islets and pancreatic progenitors. (2) “transcription factors” tracks are maps of transcription factors binding sites obtained from Chip-seq experiments performed in human adult pancreatic islets and pancreatic progenitors. (3) “SNPs” tracks include GWAS variants datasets associated to type 2 diabetes and

The screenshot displays the front panel of the Islet Regulome Browser. At the top, a navigation bar includes the logo and links for Home, Data Source, Credits, Contact, and Info. The main heading is "The Islet Regulome Browser", followed by a descriptive paragraph: "The Islet Regulome Browser is a visualization tool that provides access to interactive exploration of pancreatic islet genomic data."

The interface is divided into three main sections:

- Create a plot:** This section contains two expandable panels.
 - Gene or genome coordinates:** Includes a text input for "Gene symbol" (with "PDX1" as an example), a "Chromosome" dropdown (set to "chromosome 13"), "From" and "To" coordinate inputs (set to "28494167" and "28500451" respectively), and an "Extend coordinates" dropdown (set to "50Kb").
 - Features:** Includes a "Genome assembly" dropdown (set to "GRCh37/hg19"), "Chromatin map" and "Transcription factors" dropdowns (both set to "Adult pancreatic islets - Open chromatin classes" and "Adult pancreatic islets - TFs" respectively), a "Chromatin profile" dropdown (set to "None"), and an "SNPs" dropdown (set to "Fasting glycaemia – MAGIC").
- Advanced options:** Contains two upload sections: "Upload chromatin map" and "Upload SNPs", each with a file input field and a "Load ..." button. Links for "Read about chromatin map file format here" and "Read about SNPs file format here" are provided below each input field.

A "Submit" button is located at the bottom center of the form area.

FIGURE 2 | Front panel of the Islet Regulome Browser. From the initial page the user can generate a plot by selecting the desired parameters and a valid gene name or an absolute chromosomal location.

fasting glycemia. (4) An optional “chromatin profile” track can be loaded to visualize open chromatin profiles obtained from ATAC-Seq experiments performed in FACS purified alpha and beta cells (Figure 2).

Variants or chromatin maps tracks can be uploaded by the user for temporary display from the home page, “Advanced options” section. The file size of the uploaded file should not exceed 50 Mb. If a file contains a header, this should start with the “#” symbol. A “variant file” should consist of three or four tab-delimited fields. Mandatory fields are those of chromosome, position, and *p*-value. The files can also contain an optional fourth field with the reference number of the variant, additional columns will be ignored. A “chromatin map file” has a typical BED file format and should be composed of 3 tab-delimited fields: chromosome, start, and end, additional columns will be ignored. The fields with positional information should only contain integer values while the *p*-values should be numerical values. Upon data upload, a “Share uploaded files” option may be selected. This will provide a link that can be copy-pasted to a browser address bar in order to reproduce the Islet Regulome Browser session in use, including the uploaded data. Such link may be shared with other users in order to share data on the Islet Regulome Browser. Data uploaded by the user will be available for 1 month.

Plot Description

For any given gene or genomic region selected by the user, a plot is generated (Figure 3).

The plot illustrates the regulatory regions, transcription factors binding sites and GWAS variants in which the sequence of the base genome is represented on the horizontal axis. In the upper part of the plot a red line on the chromosome ideogram reflects the portion of the chromosome displayed.

Each dot represents a genomic variant, being the color intensity of the dot proportional to $-\log p$ -value of association, as indicated on the side of the plot. A black box in the central part of the plot contains vertical colored bands depicting different chromatin states, open chromatin classes or regulatory elements as described in the legend above the plot. Black lines connecting the circles (each representing a different transcription factor) to the black box, point to the genomic location of each transcription factor binding site. The color intensity of such lines is proportional to the number of co-bound transcription factors. Annotated genes are depicted as horizontal gray lines at the bottom of the plot, with transcriptional orientation indicated by arrows. Boxes along the line correspond to positions of coding exons. Islet-specific genes are shown in dark gray.

Plotting Versatility

Graphical outputs are highly dynamic, being rendered on the fly. The user can zoom in and out at different resolutions as well as slide left or right 25, 50, and 75% of the length of the plot.

The “Data displayed” panel, selectable from top left corner of the plot page, allows reviewing all the settings used to make the plot including genomic coordinates, genome build and all the features selected.

Retrieve Results

Graphical representations and text tables are available for download (Figure 4A).

The plot can be downloaded as PNG (Portable Network Graphics) or as PDF (Adobe Portable Document) format by clicking on the download icon above the plot. The difference between the two formats is that the latter uses vector graphics that is more suitable for high resolution publication figures while PNG compresses the image to a bitmap.

A button above the plot provides a link to a UCSC browser (Speir et al., 2016) session containing all the data currently available in the Islet Regulome Browser for classic UCSC visualization. For this purpose bigwig files were generated from BAM files obtained by aligning the raw data using Bowtie2 (Langmead and Salzberg, 2012) (default parameters).

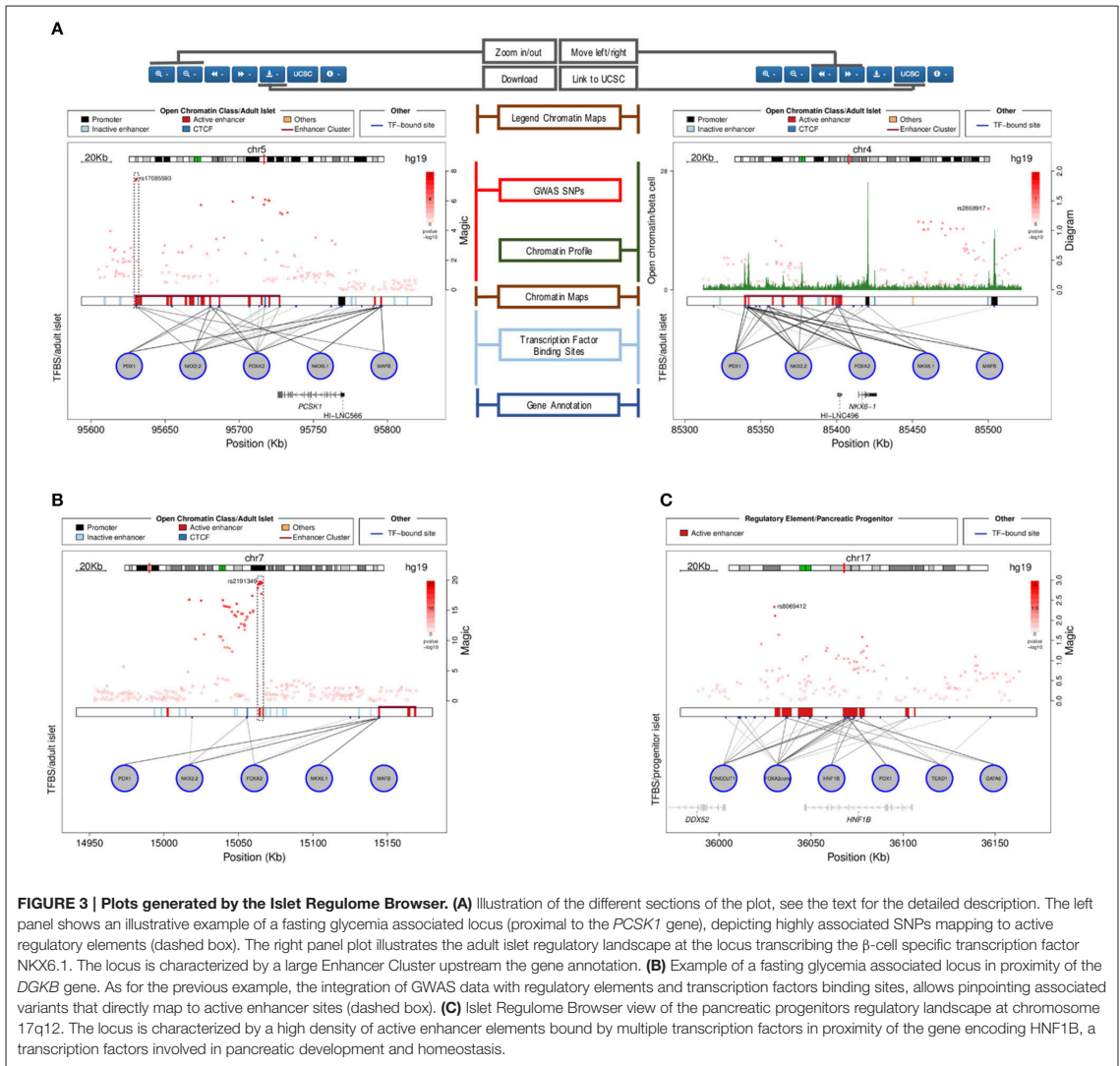
Three tables related to the selected locus can be downloaded from the “Table” panel, selectable from the top left corner of the plot page. One table contains the regulatory regions, open chromatin classes or chromatin states selected for display along with the transcription factors whose binding sites overlap them (Figure 4B). A second table lists the variants contained in the selected locus along with their *p*-value of association (Figure 4C). Finally a third table includes reference ID and expression level of the different transcript isoforms overlapping the selected locus (Figure 4D).

A link at the top left corner of the plot page named “Data displayed” redirects the user to the “Data Source” used to create the plot displayed, including reference, date of publication and links to the databases where the raw data is deposited.

DISCUSSION

With the advent of high-throughput sequencing technologies we are assisting to an exponential production of data relevant to different fields of research including pancreatic islet regulatory genomics. Scientists are now facing new challenges by shifting the research efforts from data acquisition to data processing, and knowledge extraction. The role of the Islet Regulome Browser is to provide to the pancreatic islet community fast accessibility to processed genomic data obtained from experiments performed on the endocrine pancreatic tissue. Such data is otherwise of difficult accessibility to non-bioinformatics laboratories being publicly available but usually deposited in bulky unprocessed formats.

Much of the scientific effort in the pancreatic islet field is nowadays dedicated to the understanding of the non-coding genome functions in diabetes, in an effort of translating the GWAS genetic signal of association to a molecular mechanism. Compared to preliminary meeting communications (Ramos et al., 2016) the Islet Regulome Browser now allows the visualization of different classes of regulatory elements and transcription factor binding sites obtained from experiments performed by different labs worldwide. The original view of the data provided by the Islet Regulome Browser allows to easily integrating GWAS raw files with epigenomic and transcriptomic

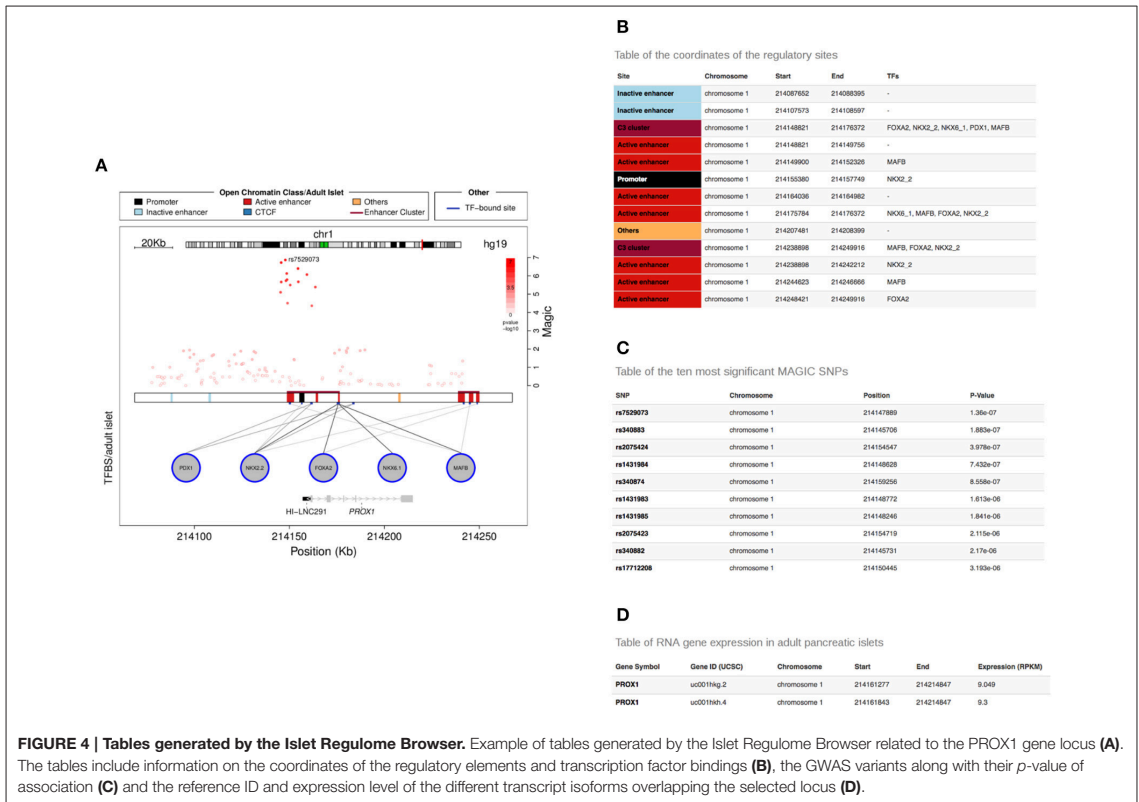


datasets. The user can thus visualize the whole spectrum of variants with different *p*-values of association and contrast them with non-coding regulatory elements and transcription factor binding sites in simple way. We believe that such level of data integration is novel compared to other available genome browsers and can assist researchers in prioritizing diabetes associated variants and to boost their functional validations.

The Islet Regulome Browser is not intended to compete with other genomic browser tools rather to integrate data of specific interest to a relative small scientific community with genomic annotation and epigenetic features obtained from other tissues. To this end we provide the data available at the Islet Regulome

Browser processed and organized in UCSC genomic browser sessions as well as direct links to the raw fastq files.

The Islet Regulome Browser is an intuitive interface to explore pancreatic islet genomic datasets. Publicly available experimental data sets such as open chromatin assays, transcription factor binding assays or GWAS variants are readily visualized at loci of interest and provided in the form of summary tables, facilitating the selection of candidate loci to be considered in experimental settings. We believe that such tool will facilitate the access to pancreatic islet public genomic datasets providing a major boost to functional genomics studies in glucose metabolism related traits including diabetes.



The Islet Regulome Browser is freely accessible at <http://www.isletregulome.com>.

AUTHOR CONTRIBUTIONS

LP and LM conceived the project. LM designed and implemented the interface, the Web page, and the R code with contribution from MR. LP wrote the paper with contributions from LM and MR. All the authors read and approved it.

FUNDING

This work was supported by a grant from the Spanish Ministry of Economy and Competitiveness (BFU2014-58150-R). LP is a

REFERENCES

Ackermann, A. M., Wang, Z., Schug, J., Naji, A., and Kaestner, K. H. (2016). Integration of ATAC-seq and RNA-seq identifies human alpha cell and beta cell signature genes. *Mol. Metabol.* 5, 233–244. doi: 10.1016/j.molmet.2016.01.002

recipient of a Ramon y Cajal contract from the Spanish Ministry of Economy and Competitiveness (RYC-2013-12864).

ACKNOWLEDGMENTS

The Islet Regulome Browser was originally developed at Jorge Ferrer's Laboratory at the Institut D'Investigacions Biomediques August Pi i Sunyer (IDIBAPS). We would like to thank Prof. Jorge Ferrer Imperial College London (ICL) as he provided initial supervision and support to The Islet Regulome Browser project. We would also like to thank Iñaki Martinez, System Administrator at the Program for Predictive and Personalized Medicine of Cancer at the Institute Germans Trias i Pujol (IGPT) Bioinformatics Core, for excellent technical support.

Akerman, I., Tu, Z., Beucher, A., Rolando, D. M. Y., Sauty-Colace, C., Benazra, M., et al. (in press). Human pancreatic β cell lncRNAs control cell-specific regulatory networks. *Cell Metab.* doi: 10.1016/j.cmet.2016.11.016

Ashcroft, F. M., and Rorsman, P. (2012). Diabetes mellitus and the beta cell: the last ten years. *Cell* 148, 1160–1171. doi: 10.1016/j.cell.2012.02.010

- Bernstein, B. E., Stamatoyannopoulos, J. A., Costello, J. F., Ren, B., Milosavljevic, A., Meissner, A., et al. (2010). The nih roadmap epigenomics mapping consortium. *Nat. Biotechnol.* 28, 1045–1048. doi: 10.1038/nbt1010-1045
- Bhandare, R., Schug, J., Le Lay, J., Fox, A., Smirnova, O., Liu, C., et al. (2010). Genome-wide analysis of histone modifications in human pancreatic islets. *Genome Res.* 20, 428–433. doi: 10.1101/gr.102038.109
- Cebola, I., Rodríguez-Seguí, S. A., Cho, C. H., Bessa, J., Rovira, M., Luengo, M., et al. (2015). TEAD and YAP regulate the enhancer network of human embryonic pancreatic progenitors. *Nat. Cell Biol.* 17, 615–626. doi: 10.1038/ncb3160
- Cho, Y. S., Chen, C. H., Hu, C., Long, J., Ong, R. T., Sim, X., et al. (2012). Meta-analysis of genome-wide association studies identifies eight new loci for type 2 diabetes in east Asians. *Nat. Genet.* 44, 67–72. doi: 10.1038/ng.1019
- Dayeh, T., Volkov, P., Saló, S., Hall, E., Nilsson, E., Olsson, A. H., et al. (2014). Genome-wide DNA methylation analysis of human pancreatic islets from type 2 diabetic and non-diabetic donors identifies candidate genes that influence insulin secretion. *PLoS Genet.* 10:e1004160. doi: 10.1371/journal.pgen.1004160
- Dunham, I., Kundaje, A., Aldred, S. F., Collins, P. J., Davis, C. A., Doyle, F., et al. (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature* 489, 57–74. doi: 10.1038/nature11247
- Gaulton, K. J., Nammo, T., Pasquali, L., Simon, J. M., Giresi, P. G., Fogarty, M. P., et al. (2010). A map of open chromatin in human pancreatic islets. *Nat. Genet.* 42, 255–259. doi: 10.1038/ng.530
- Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods.* 9, 357–359. doi: 10.1038/nmeth.1923
- Lawrence, M., Gentleman, R., and Carey, V. (2009). rtracklayer: an R package for interfacing with genome browsers. *Bioinformatics* 25, 1841–1842. doi: 10.1093/bioinformatics/btp328
- Morán, I., Akerman, I., van de Bunt, M., Xie, R., Benazra, M., Nammo, T., et al. (2012). Human beta cell transcriptome analysis uncovers lncRNAs that are tissue-specific, dynamically regulated, and abnormally expressed in type 2 diabetes. *Cell Metab.* 16, 435–448. doi: 10.1016/j.cmet.2012.08.010
- Morgan, M., Pagès, H., Obenchain, V., and Hayden, N. (2016). *Rsamtools: Binary Alignment (BAM), FASTA, Variant Call (BCF), and Tabix File Import.* R Package Version 1.26.1. Available online at: <http://bioconductor.org/packages/release/bioc/html/Rsamtools.html>
- Parker, S. C., Stitzel, M. L., Taylor, D. L., Orozco, J. M., Erdos, M. R., Akiyama, J. A., et al. (2013). Chromatin stretch enhancer states drive cell-specific gene regulation and harbor human disease risk variants. *Proc. Natl. Acad. Sci. U.S.A.* 110, 17921–17926. doi: 10.1073/pnas.1317023110
- Pasquali, L., Gaulton, K. J., Rodríguez-Seguí, S. A., Mularoni, L., Miguel-Escalada, I., Akerman, I., et al. (2014). Pancreatic islet enhancer clusters enriched in type 2 diabetes risk-associated variants. *Nat. Genet.* 46, 136–143. doi: 10.1038/ng.2870
- Phanstiel, D. H., Boyle, A. P., Araya, C. L., and Snyder, M. P. (2014). Sushi.R: flexible, quantitative and integrative genomic visualizations for publication-quality multi-panel figures. *Bioinformatics* 30, 2808–2810. doi: 10.1093/bioinformatics/btu379
- Ramos, M., Mularoni, L., and Pasquali, L. (2016). The Islet regulome browser. *Endocrinol. Nutr.* 63, 73.
- Scott, R. A., Lagou, V., Welch, R. P., Wheeler, E., Montasser, M. E., Luan, J., et al. (2012). Large-scale association analyses identify new loci influencing glycemic traits and provide insight into the underlying biological pathways. *Nat. Genet.* 44, 991–1005. doi: 10.1038/ng.2385
- Speir, M. L., Zweig, A. S., Rosenbloom, K. R., Raney, B. J., Paten, B., Nejad, P., et al. (2016). The UCSC Genome Browser database: 2016 update. *Nucleic Acids Res.* 44, D717–D725. doi: 10.1093/nar/gkv1275
- Stitzel, M. L., Sethupathy, P., Pearson, D. S., Chines, P. S., Song, L., Erdos, M. R., et al. (2010). Global epigenomic analysis of primary human pancreatic islets provides insights into type 2 diabetes susceptibility loci. *Cell Metab.* 12, 443–455. doi: 10.1016/j.cmet.2010.09.012
- Wolfsberg, T. G. (2011). Using the NCBI Map Viewer to browse genomic sequence data. *Curr. Protoc. Hum. Genet.* Chapter 18:Unit 18 5. doi: 10.1002/0471142905.hg1805s69
- Yates, A., Akanni, W., Amode, M. R., Barrell, D., Billis, K., Carvalho-Silva, D., et al. (2016). Ensembl 2016. *Nucleic Acids Res.* 44, D710–D716. doi: 10.1093/nar/gkv1157

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Mularoni, Ramos-Rodríguez and Pasquali. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

8.4 State of the art of the Islet Regulome Browser

The publication of the above article coincided with the first version of the IRB, during the first year of my PhD. Since then, I have grown the expertise and proficiency to take over the development and maintenance of the IRB. As part of the development and enhancement of the site, I have improved the whole implementation to make it more scalable and simple. For this reason, many of the methods and results described in the article are now obsolete. Thus, I provide this additional section to update the methods and results to match the new implementation of the IRB.

8.4.1 Methods

In **Figure 8.1** I show the current structure of the IRB. Similarly to the older version, it consists of three main components:

- **Database.** The database contains all the datasets available in the IRB, which are saved as RData objects.
- **plotRegulome R package.** Consists on a custom R package available on github¹ that uses the datasets present in the database to produce the typical Islet Regulome plot.
- **Shiny web application.** Shiny (Chang et al. 2019) is an R package that provides the necessary tools to build interactive web applications using the R programming language. The code for the IRB shiny application can be found in the corresponding github repository².

These three components interact with one another to render the IRB web application, process user queries and return the resulting plots and tables. All these components are encapsulated inside a Docker (Docker Inc n.d.) container, which ensures stability and reproducibility of the site.

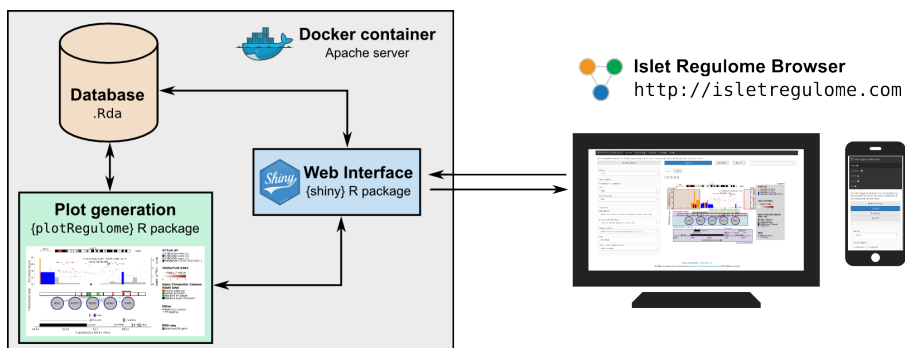


Figure 8.1: Updated structure of the Islet Regulome Browser.

¹<https://github.com/mireia-bioinfo/plotRegulome>

²https://github.com/mireia-bioinfo/isletregulome_shiny

8.4.1.1 Database

In order to generate the database, processed files were obtained from the original publications, coordinates were lifted over to the hg19 genome build and datasets were transformed to match the requirements of the `plotRegulome` package. These datasets were then stored as RData files, which are lightweight binary compressed files that R can write and read very efficiently.

According to the type of data they represent, they can be classified in different groups:

- **Enhancer clustering annotation.** Represents different grouping of enhancer elements such as enhancer clusters, stretch enhancers or enhancer hubs.
- **Chromatin maps.** Contains chromatin annotations in islet-relevant tissues and cell lines.
- **Transcription factors.** Depicts binding sites for TFs.
- **GWAS.** Association studies analyzing islet relevant diseases and phenotypes. Contains the studied SNPs together with their p-value of association with the studied phenotype.
- **Chromatin contact data.** Contains virtual 4C data derived from Miguel-Escalada et al. (2019) promoter-capture Hi-C experiments in human pancreatic islets.
- **Gene annotations.** Represents gene annotations, together with mRNA expression levels, and also lncRNAs found in human pancreatic islets.

All these datasets are public and can be easily downloaded with the function `downloadIRB()` from the `plotRegulome` R package. The different datasets included in the IRB are listed in **Table 8.1**.

8.4.1.2 `plotRegulome` package

The `plotRegulome` package was developed with the idea of providing consistent and scalable functions to produce the characteristic Islet Regulome plot, using the Islet Regulome database.

With this idea in mind, `plotRegulome` contains several functions for plotting all the different panels inside the regulome plot, allowing the user to easily produce the independent panels – for example, the GWAS SNPs – if needed. This new implementation in the form of an R package makes it easier to add new visualizations and implement other datasets to the Islet Regulome plot.

Internally, `plotRegulome` takes advantage of the `ggplot2` (Wickham et al. 2019) R package to easily produce the required plots from the different

Table 8.1: **Datasets available for visualization at the Islet Regulome Browser.**

Dataset type	Dataset name
Ramos-Rodríguez et al. (2019)	
Chromatin Maps	Adult Islets - Cytokine-Responsive REs
Chromatin Maps	EndoC- β H1 - Cytokine-Responsive REs
Gene Annotation	Adult Islets - Cytokine-Responsive expression
Gene Annotation	EndoC-bH1 - Cytokine-Responsive expression
Miguel-Escalada et al. (2019)	
Chromatin Maps	Adult Islets - Chromatin Classes
Enhancer Clustering Annotation	Super Enhancers
Enhancer Clustering Annotation	Enhancer Hubs
Transcription Factors	Adult Islets - Structural (MED1, CTCF)
Virtual 4C	Virtual 4C
Bonàs-Guarch et al. (2018)	
SNPs	70KforT2D
Akerman et al. (2017)	
Gene Annotation	lncRNAs in human islets
Cebola et al. (2015)	
Chromatin Maps	Pancreatic Progenitors
Transcription Factors	Pancreatic Progenitors (ONECUT1, FOXA2, PDX1, TEAD1, GATA6)
Pasquali et al. (2014)	
Chromatin Maps	Adult Islets - Chromatin Classes
Enhancer Clustering Annotation	Transcription Factors Enhancer Clusters
Transcription Factors	Adult Islets - Tissue-specific (PDX1, NKX2.2, FOXA2, NKX6.1, MAFB)
Parker et al. (2013)	
Chromatin Maps	Adult Islets - Chromatin States
Enhancer Clustering Annotation	Stretch Enhancers
Morán et al. (2012)	
Gene Annotation	RNA-seq in human islets
Morris et al. (2012)	
SNPs	Diagram
Scott et al. (2012)	
SNPs	Magic
Gaulton et al. (2010)	
Enhancer Clustering Annotation	COREs

datasets. After all the plots are generated, they are combined into one single figure using the `cowplot` (Wilke 2019) R package. This whole process is automated by the `plotRegulome` function, which takes as input the name of the different datasets the user wants to plot and automatically generates the output Islet Regulome Plot.

8.4.1.3 Shiny web application

Shiny (Chang et al. 2019) is a very powerful R package that allows easy creation of interactive and responsive web applications using the R programming language. It provides easy customization and full integration with other R packages and R code.

The IRB shiny app has three basic parts:

- `global.R`. Loads and defines elements that need to be used by all the other parts, such as R packages and datasets.
- `ui.R`. This script is used to define the user interface. It contains the overall structure of the site, together with the layout of the application and the different menus and widgets that need to be included.
- `server.R`. Deals with parsing the inputs from the user and producing the different outputs. Here, the different arguments are passed to the `plotRegulome` function to produce the Islet Regulome plot. The code for generating the different tables is also included in this script.

This app can be run locally or it can be launched at the IRB site (isletregulome.org).

8.4.2 Results

8.4.2.1 User interface

The use of shiny allows for a nice and responsive interface that automatically adjusts to the size of the user's screen. Compared to the older IRB version, the selection of coordinates and datasets and the output plot are located in the same window (Figure 8.2). This allows the user to keep track of the datasets that are being plotted and change them at will.

8.4.2.2 Islet Regulome plot

The overall IRB plot is very similar to the one described in the publication, with the addition of some new data visualizations, such as the virtual 4C profiles. Briefly, the IRB plot can enclose many different genomic and epigenomic information in one single figure (Figure 8.3):

- **Virtual 4C**. 3D contact data of the gene of interest (black triangle) with the rest of the genome. Fill color represents CHiCAGO confidence scores.
- **SNPs**. GWAS SNP data showing each individual SNP (points) and its p-value of association with the disease (T2D = Diagram, 70KforT2D; Fasting glycemia = Magic).
- **Chromatin Maps**. Open chromatin regions classified in different types (see original publications in **Table 8.1**), represented as colored

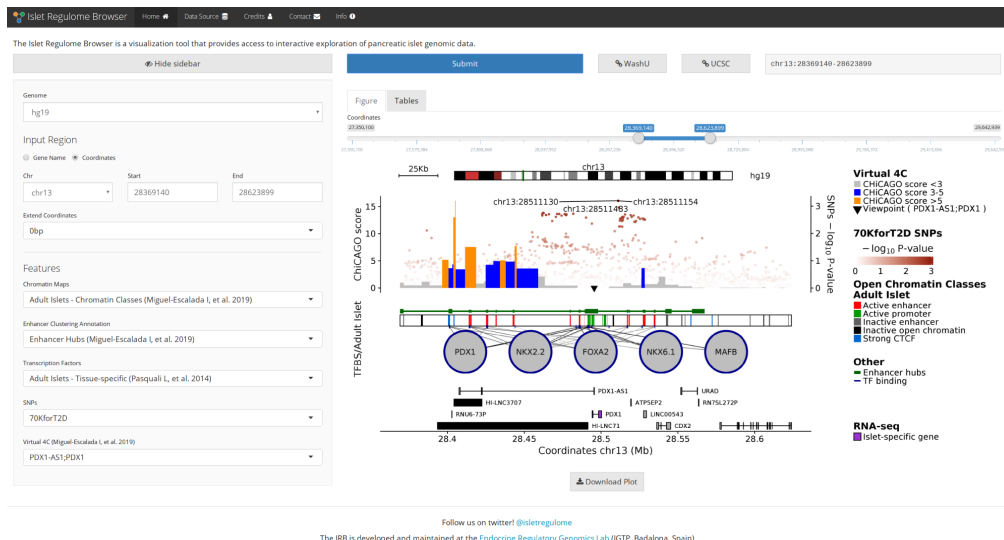


Figure 8.2: View of the Islet Regulome Browser.

rectangles.

- **Enhancer Clustering Annotation.** Grouping information of open chromatin regions obtained by different algorithms (see original publications in **Table 8.1**), represented as green lines covering the genomic regions included in the cluster.
- **Transcription Factors.** TF binding sites obtained from ChIP-seq assays in human islets or pancreatic progenitors. Lines indicate the location of the binding site and the darkest the line, the more TFs are found binding to the same region.
- **Gene Annotation.** This annotation includes protein coding genes (exons are represented as rectangles) and long non-coding RNAs from human islet samples. Islet-specific genes are colored in purple.

The zooming in and out of the plot is now provided as a slider at the top of the plot, through which the user can change both the start and end coordinates of the view and/or move it around by clicking in the blue bar that joins the two dots that represent the start and end coordinates.

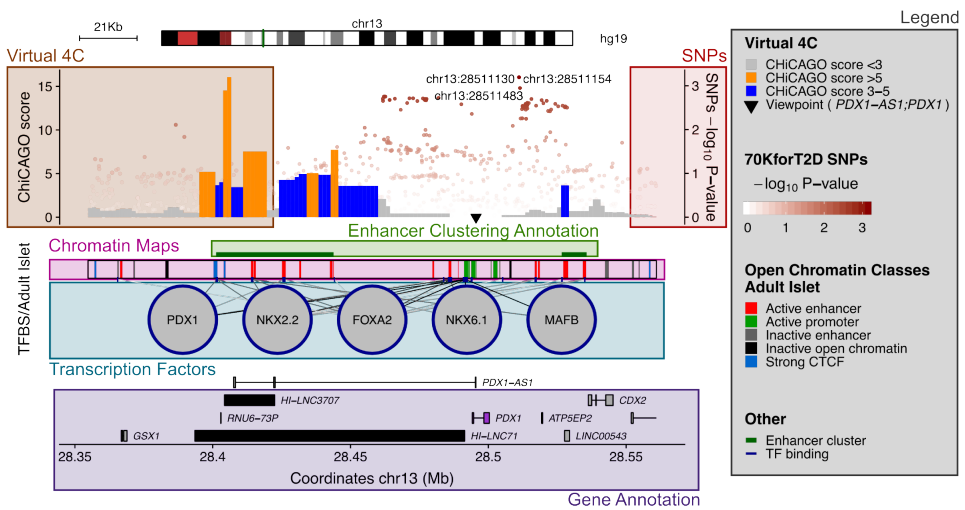


Figure 8.3: Plot generated by the Islet Regulome Browser via the `plotRegulome` R package. The different information integrated in the plot are marked by colored rectangles.

Part IV

Discussion

9

Regulatory networks driving responses to cytokines in β cells

Insulinitis is a key process in the development of T1D. During this immune islet infiltration, different molecules, such as cytokines and chemokines, are released by both immune and β cells. Such molecules act as messengers that will in turn induce changes in the surrounding cells by inducing the modulation of gene expression and protein profile. Thus, understanding the regulatory responses activated in different cell types during insulinitis is key to further dissect T1D pathological mechanisms and ultimately pinpoint new biomarkers of disease progression or new therapeutic intervention targets.

9.1 Proinflammatory cytokines induce extensive β cell chromatin remodeling

Exposure to both cocktails of proinflammatory cytokines tested in the studies described in the present thesis (IFN- α and IFN- γ + IL-1 β), induces considerable chromatin remodeling in pancreatic β cells, potentially uncovering new cis-regulatory networks driving the β -cell response to a proinflammatory environment.

In the case of IFN- α , which was used to model early insulinitis, >4,400 gained open chromatin regions were observed as early as 2 hours after IFN- α exposure. Of those accessible regions, only 1,000 were also identified as gained after 24 hours of exposure, suggesting that some regions act as fast-responders at 2 hours and then return to their basal, less accessible status. Even though chromatin accessibility can be used to identify regulatory elements, chromatin openness is not necessarily a measure of regulatory element activity. Thus, solely by measuring chromatin accessibility we cannot determine if the identified regions are in fact active regulatory elements. Therefore, these regions were simply defined as **Open Chromatin Regions** (OCRs).

Next, a cocktail of IFN- γ + IL-1 β was used to study β -cell chromatin remodeling in late insulinitis. Similarly, a great degree of chromatin remodeling was observed after a 48 hour exposure. We identified genomic regions gaining chromatin accessibility and/or increasing their enrichment in H3K27ac, a histone modification that marks active regulatory regions such as promoters

and enhancers (Creyghton et al. 2010; Rada-Iglesias et al. 2011). In this study, the use of both chromatin accessibility and the H3K27ac histone modification allowed for a better characterization of the potential regulatory elements driving the response to proinflammatory cytokines. Thus, changes in H3K27ac were used as proxies for regulatory element activation, with the coordinates of the accessible sites pinpointing the genomic coordinates of such elements. This allowed us to define a subset of 3,798 genomic regions becoming activated after cytokine exposure, which we called **Induced Regulatory Elements** (IREs).

Of note, regulatory element losses upon the stimuli, represented by chromatin regions reducing their accessibility or showing a reduction of H3K27ac, was residual in both studies. Nonetheless, it could be possible that longer exposure times could further modify the β -cell regulatory landscape by inducing loss of β -cell specific processes, as reported by other studies (Eizirik, Colli, and Ortis 2009).

In general, the magnitude of chromatin changes observed in both studies is comparable to other works using differentiated cells exposed to similar stimuli (Ostuni et al. 2013; Calderon et al. 2019). Nonetheless, the cited works use immune-related cell types, whose main function is to respond effectively to these type of stimuli. Thus, the results from the present thesis show that highly specialized cells, such as β cells, are also able to modify their regulatory landscape to respond to proinflammatory stimuli.

To what extent do the regulatory elements activated in models of early and late insulinitis overlap? We tried to address this question, but encountered some difficulties due to key differences in the experimental design of these studies that impede direct comparison of their results. First, characterization of regulatory elements in both studies is different: for the IFN- α study (solely based on ATAC-seq), all gained open chromatin regions were considered potential regulatory elements, whereas for the IFN- γ + IL-1 β study (based on ATAC-seq and H3K27ac ChIP-seq) only accessible regions – either gained or stable – with gains in H3K27ac enrichment were defined as induced regulatory elements. Secondly, the exposure times for both studies are different, that is, two different time points of 2 and 24 hours for the IFN- α study, and a single time point of 48 hours for the IFN- γ + IL-1 β study. Finally, even though bioinformatics methods for processing and performing differential analysis were similar, the number of replicates also differed between experiments: ATAC-seq data was obtained from 4 replicates in the IFN- α study, while 5 ATAC-seq replicates were used in the IFN- γ + IL-1 β studies. All these differences made the comparison of cytokine-responsive regulatory elements obtained from the two studies unreliable, based on the current data.

Both OCR and IREs show correlation with the induction of the nearby genes and their corresponding protein. In the case of OCRs, we observed that their

gains in chromatin accessibility precede up-regulation of their putative target genes, showing highest correlation between detected OCRs at 2 hours and up-regulated genes at 8 hours. Regarding IREs, their association with gene up-regulation was maintained when this group was sub-divided into opening and primed IREs. This observation suggests that both classes of IREs are able to up-regulate their target genes independently from their pre-existing accessibility status, that is, closed or already accessible. Further study on the gene networks regulated by such putative regulatory elements could help to clarify the cellular processes involved in responses to cytokines and could help to identify potential therapeutic interventions to reverse the detrimental effects of cytokine-induced β -cell stress and apoptosis.

9.2 Uncovering novel regulatory elements

Both proinflammatory cytokine cocktails studied in this thesis – IFN- α and IFN- γ + IL-1 β , modeling early and late insulinitis, respectively – trigger genome-wide chromatin remodeling, resulting in changes in gene expression and in the cell's protein profile. Such remodeling may uncover the activation of novel regulatory elements that drive the β -cell response to proinflammatory cytokines.

In Chapter 6, we hypothesized that gained OCRs were mostly composed of cytokine-induced distal regulatory elements, as suggested by their high sequence conservation scores, distal localization relative to gene TSS and enrichment for TF binding motifs. Moreover, the identified regions were correlated with the induction of gene expression of the nearby genes, and footprint analysis indicated likely binding of TFs.

Characterization of regulatory elements in Chapter 7, however, was more accurate, as we intersected changes in chromatin accessibility with changes in H3K27ac enrichment, allowing a more precise prediction of active enhancers. Similarly to OCRs, IREs were also phylogenetically conserved, distal to TSS and enriched for TF binding motifs. In this case, we were also able to further classify these accessible regions into different groups, depending on their chromatin accessibility changes: **Opening IREs** gained both H3K27ac enrichment and chromatin accessibility, while **primed IREs** were already accessible before the stimulus but were enriched in H3K27ac after the proinflammatory cytokine exposure. Taking into account that opening IREs showed different gradients of chromatin accessibility changes, we further divided them into **neo IREs**, representing those open chromatin regions that were totally closed before the stimulus but afterwards opened up and gained H3K27ac. Moreover, neo IREs are analogous to the latent enhancers described by Ostuni et al. (2013) in murine macrophages.

9.3 Tissue-specific mechanisms in the response to proinflammatory cytokines

Different classes of IREs also presented intrinsically different mechanisms of activation. **Primed IREs** were already accessible and pre-bound by islet-specific regulatory elements. After the cytokine exposure, our analyses suggest that new TFs, mostly inflammatory, bind these regulatory elements and cooperate with β -cell specific TFs to activate the response to proinflammatory cytokines. A similar cooperation between tissue- and signal-specific TFs has been described before in murine and human macrophages (Heinz et al. 2013; Alasoo et al. 2018).

Conversely, **neo IREs** showed no enrichment in islet-specific TFs, neither before nor after the exposure. On the other hand, they were likely bound by inflammatory TFs, as suggested by the *de novo* motif analysis and the ISRE footprint observed at those genomic sites. DNA methylation analysis also revealed the analyzed CpGs within these neo IREs were demethylated upon the proinflammatory signal.

These findings are in line with some studies revealing that external stimuli can change the enhancer repertoire of terminally differentiated cells from the immune system (Ostuni et al. 2013; Alasoo et al. 2018; Calderon et al. 2019). Moreover, these novel cis-regulatory elements elicited by a disease-relevant stimuli might help explain the immune attack tissue-specificity in autoimmune pathologies. Indeed, Alasoo et al. (2018) proposed that disease variants lying in such genomic regions can influence enhancer function, for instance by altering tissue-specific TF binding, which may in turn alter the cooperative binding of stimulus-specific TFs. The generated aberrant response would then impede the resolution of the immune attack and lead to a progression of the autoimmune process.

9.4 Changes in 3D chromatin structure

The uncovering of new putative enhancers driving β -cell responses to proinflammatory cytokines raises the question of whether such IREs are already in contact with their gene targets or new enhancer-promoter interactions are formed upon the cytokine exposure. Some studies already explored the dynamics of established chromatin interactions during cell development (Phanstiel et al. 2017; Mumbach et al. 2016), in carcinogenic processes (Taberlay et al. 2016; Barutcu et al. 2015) or after hormonal exposure in breast cancer cells (Le Dily et al. 2014), among others. However, at the time of publication of the work presented in Chapter 7, the chromatin dynamics of highly specialized and differentiated cells, such as pancreatic islets, upon an environmental stimulus was not known.

As TADs, mainly described by Hi-C chromatin capture technology (see *Chromatin three-dimensional organization*), were shown to be established during development and to be mostly stable among cell types, we aimed to obtain targeted promoter-centered chromatin interactions with higher resolution using UMI-4C assays. By using the promoters of 13 up-regulated genes as viewpoints, we observed that, upon proinflammatory exposure, chromatin contacts were preferentially gained with IREs. This observation suggests that new enhancer-promoter loops were formed, allowing IREs to get in physical proximity with their target genes. Moreover, these findings suggest that differentiated cells maintain the ability to form new chromatin interactions to respond to environmental changes.

In summary, the studies presented in this thesis provide evidence that the regulatory landscape of terminally differentiated cells, such as β cells, can be reshaped upon stimuli, revealing novel regulatory elements involved in the response to such stimuli. Such regulatory elements are, in turn, able to form new enhancer-promoter interactions to modulate the expression of their target genes. This is the first time, to our knowledge, that such level of remodeling is described in a human specialized, differentiated and non-immune related tissue, such as β cells and human pancreatic islets. Linking these newly uncovered IREs to T1D will provide more insights into the role of these putative enhancers in the development of the disease.

10

Exploring gene networks to find T1D therapeutic targets

Studying gene and protein networks driving β -cell responses to proinflammatory cytokines is vital to understand the cellular processes involved in such responses. Analyses of protein-protein interaction (PPI) networks in both IFN- α and IFN- γ + IL1- β studies revealed that up-regulated proteins were enriched for processes such as cellular response to viruses, antigen processing and presentation via MHC class I for **early insulinitis**, and IFN- γ signaling, antigen processing and presentation, apoptosis and T1D for **late insulinitis**. These results suggest that both cocktails are able to induce the key T1D signatures resulting from the different insulinitis stages.

Sequence analysis of open chromatin regions can provide clues as to which TFs are regulating the response to proinflammatory cytokines. *De novo* motif analysis in OCRs and IREs revealed both inflammatory and islet-specific TF motifs (see Chapter 9). Of note, IRF and STAT motifs, belonging to a family of TFs involved in IFN signaling, were found in both OCRs and IREs. The role of both families of TFs was further studied in Chapter 6, showing that individual protein knock-downs were usually compensated by other proteins from the same family. Moreover, IRF1 appeared to act as a key regulator of β -cell INF- α -mediated responses, by inducing inhibitory checkpoint proteins such as *PDL1* and *HLA-E*. Interestingly, the effect of IRF1 is different depending on the cell type in which it is studied: the systemic *IRF1* knockout in NOD mice prevents autoimmune diabetes (Nakazawa et al. 2001), while islet-specific deletion correlates with shorter islet mouse graft function and survival (Gysemans et al. 2009). Such differences may be explained by IRF1 acting through cell-type-specific regulatory elements, thus highlighting the importance of dissecting cis-regulatory networks in target tissues.

Study of gene and protein networks driving the response to IFN- α in β cells revealed several candidate genes and pathways, which are also present in the β -cell gene signatures of T1D patients. Mining such networks and comparing them to pathway and drug databases, revealed two potential therapeutic interventions inhibiting either bromodomain proteins or JAK tyrosine kinases.

Bromodomain proteins are part of chromatin remodeling complexes and are able to recognize acetylated lysine residues in histones, thus promoting chromatin decompaction. Previous studies have demonstrated

that inhibition of bromodomain proteins in NOD mice produces protective effects against T1D development (Fu et al. 2014). Here, we show that inhibition of bromodomain proteins by two different molecules (JQ1+ and I-BAT-151) is able to down-regulate HLA class I and chemokine expression in β cells. Nonetheless, such molecules are not able to protect β cells against cytokine-induced apoptosis.

On the other hand, the **JAK protein family** is implicated in the IFN- α signaling pathway by triggering the JAK-STAT signaling cascade. The inhibition of two of its members, JAK1 and JAK2, is able to prevent development of T1D in NOD mice (Trivedi et al. 2017). Here, we used baricitinib, a drug that has already been approved as treatment in rheumatoid arthritis (Genovese et al. 2016), to inhibit JAK1/2. Baricitinib treatment in β cells was able to inhibit HLA class I overexpression and reduce cytokine-induced β cell ER stress and apoptosis.

To summarize, the compounds presented in this thesis are good candidates for reversing the deleterious effects of IFN- α in β cells during early insulinitis, thus favoring survival of β cells.

11

Finding a mechanism for the β cell role in T1D using GWAS

GWAS provide great resources for studying and understanding the genetic risk of complex diseases, by comparing the polymorphic alleles present in affected population versus healthy controls. As 88% of GWAS-identified T1D variants are located in non-coding sequences, they are candidates for potentially disrupting gene regulatory activity. Thus, comparison of disease variants with enhancer maps is critical to shed light onto the potential disease mechanisms. However, much effort is still needed to understand the molecular mechanisms underlying candidate risk variants affecting enhancer function. Challenges involve the identification of the putative gene targets of the candidate enhancer, the understanding of the cell types in which risk variants operate and the lack of accurate *in vitro* and *in vivo* disease models.

In the case of T1D, several studies have observed enrichment of T1D candidate SNPs at immune cell enhancers (Onengut-Gumuscu et al. 2015; Farh et al. 2015) and, more subtly, in pancreatic islet enhancers (Farh et al. 2015). Nonetheless, studying static regulatory maps might be limiting our understanding of the genetic mechanisms underlying disease risk, as many variants may be context-specific. This means that it might be necessary to profile a large range of disease-specific cellular states in order to understand the individual contributions of each candidate risk variant.

The need for querying different cellular states to understand disease risk was already presented in previous works in immune cell types (Alasoo et al. 2018; Kim-Hellmuth et al. 2017). With the work presented in this thesis, we demonstrate that this concept is valid as well for pancreatic islets in the context of T1D. Thus, by intersecting IREs with T1D-associated variants, we were able to observe an enrichment of T1D risk SNPs in these cytokine-induced enhancers. Nonetheless, with the current data, we cannot rule out the possibility that the T1D risk variants overlapping islet IREs are also acting through immune cell types, as we observe that islet IREs may in some cases act as active enhancers in some immune cell types.

Luciferase assays provide further proof that a subset of T1D risk variants are functional by altering enhancer activity in β cells. We provide evidence that rs78037977 and rs193778, both T1D leading variants, are able to modify the enhancer activity of the IRE they overlap. More studies are needed to

fully understand the effect of the risk alleles at specific enhancers. One mechanism of action may be that of altering the binding affinity of cell-type specific TFs and thus indirectly decreasing binding of stimuli-specific TFs, mirroring a model proposed by Alasoo et al. (2018). These findings could be also exported to other autoimmune diseases, as disease variants acting through stimuli-specific enhancers in target tissues may explain the immune preference of the autoimmune attack.

To understand the biological consequences of the gene regulatory changes caused by functional risk variants, it is vital to identify their target genes. This is a challenging task, as enhancers can regulate genes located megabases away in the linear genomic space. As enhancers are believed to get in physical proximity with their target gene's promoter, analyzing chromatin interactions can guide their identification. By using UMI-4C, we provide some potential gene targets of the studied enhancers containing T1D risk variants. However, more studies are necessary to actually identify how changes in enhancer activity can affect gene networks in T1D. For example, Nasrallah et al. (2020) coupled analysis of chromatin interactions with subsequent CRISPR functional mapping to shed light onto the role of a specific enhancer in ulcerative colitis. Gaining insight onto the gene networks affected by T1D risk alleles may also allow to identify novel therapeutic targets for pre-clinical T1D studies.

12

Facilitating access to pancreatic islet genomic data

The growing availability of high-throughput assays to query the genome and epigenome, together with the drop in sequencing cost, has increased the production of genomic datasets in research laboratories worldwide. All these datasets are usually publicly available in data repositories, theoretically facilitating re-exploration and re-use of these load of the data by the scientific community. However, such data is usually deposited in unprocessed formats, which pose a challenge in terms of data processing, as it requires extensive bioinformatic knowledge, and in terms of uncovering of the underlying biological information. The Islet Regulome Browser (IRB) addresses both challenges by facilitating access and exploration of processed genomic and epigenomic data to the scientific community. Exploration of the IRB data does not require any bioinformatic knowledge, as the data is already processed and ready for visualization.

Different genomic and epigenomic factors can be queried using available high-throughput methodologies, allowing integration of such data with GWAS information, thus gaining insights into a certain disease (see *Finding a mechanism for the β cell role in T1D using GWAS* for an example). The IRB facilitates such integration by providing a unified view of GWAS data with islet epigenomic and transcriptomic information. Data integration can in turn help researchers to prioritize variants of interest by visualizing chromatin features, TF binding sites and chromatin interactions that might help uncover their underlying mechanism. Moreover, the IRB provides a user-friendly interface to easily explore genomic data and provides links for downloading the visualized datasets.

Many different browsers for genomic data visualization are currently available, such as the UCSC Genome Browser (Kent et al. 2002), the WashU Epigenome Browser (Zhou and Wang 2012) or the Integrative Genomics Viewer (Robinson et al. 2011). The IRB, however, does not aim to compete in terms of data visualization, but it rather aims to integrate and gather data on pancreatic islet regulation, making it accessible to a specialized community of researchers. For this reason, the IRB also provides links to sessions in the UCSC and WashU browsers, to allow further data exploration.

Considering that technology, both in terms of development of genomic assays

and of bioinformatic applications and tools, is advancing fast, the IRB is also evolving at the same pace. For this reason, since its official release, different types of genomic information have been added, such as virtual 4C tracks showing chromatin interactions in pancreatic islets. Additionally, the interface and the underlying code is improving to keep up with different available tools that facilitate integration and navigation at the IRB site.

In conclusion, the IRB is an important resource for the pancreatic islet community, as it gathers relevant genomic and epigenomic datasets, and allows for easy data exploration without requiring any bioinformatic background.

Part V

Conclusions

13

Conclusions

1. Exposure of β cells to both IFN- α (early insulinitis) and IFN- γ + IL-1 β (late insulinitis) induce **extensive chromatin remodeling** coupled with changes in gene expression and protein abundance. While with IFN- α exposure we only queried chromatin accessibility, thus identifying different sets of **Open Chromatin Regions** (OCRs) gained at the different time points, in the case of IFN- γ + IL-1 β we were able to uncover 3,798 **Induced Regulatory Elements** (IREs), which showed gains in chromatin accessibility and/or in enrichment of H3K27ac – a histone modification marking active regulatory elements – after the exposure.
2. Motif analysis of both OCRs and IREs revealed **binding sites for inflammatory TFs**, such as IRF and STAT. Interestingly, a subset of IREs called primed IREs, which were already accessible before exposure to the proinflammatory stimuli, were **pre-bound by islet-specific TFs**, which were likely cooperating with inflammatory TFs to drive changes in gene expression.
3. The study of changes in chromatin interactions by UMI-4C, revealed that after exposure to IFN- γ + IL-1 β , **new enhancer-promoter contacts were formed** between IREs and their up-regulated target genes. This suggests that the chromatin contacts in differentiated tissues are not fixed and can be modified to accommodate new enhancer-promoter interactions.
4. Proinflammatory cytokines induced regulation of genes related with T1D. By mining such induced gene regulatory networks, we identified two compounds able to reduce the interferon signatures in β cells:
 - **Bromodomain inhibitors** were able to reduce interferon-induced β cell stress.
 - **Baricitinib**, a JAK1/2 inhibitor, was revealed as a promising therapeutic drug for reducing both interferon-induced cellular stress and apoptosis in β cells.
5. IREs uncovered upon exposure to IFN- γ + IL-1 β were **enriched for T1D-associated risk variants**. Interestingly, the risk alleles for two T1D risk variants were tested and shown to have an **effect on enhancer activity** in response to proinflammatory cytokines, revealing a potential

β-cell-mediated mechanism for T1D progression.

6. The Islet Regulome Browser is a useful resource that allows **visualization and exploration of human pancreatic genomic datasets**, making them accessible to all researchers, including scientists with no bioinformatics background.

Part VI

References

References

- Ackermann, Amanda M., Zhiping Wang, Jonathan Schug, Ali Najji, and Klaus H. Kaestner. 2016. "Integration of ATAC-seq and RNA-seq identifies human alpha cell and beta cell signature genes." *Molecular Metabolism* 5 (3): 233–44. <https://doi.org/10.1016/j.molmet.2016.01.002>.
- Akerman, Ildem, Zhidong Tu, Anthony Beucher, Delphine M. Y. Rolando, Claire Sauty-Colace, Marion Benazra, Nikolina Nakic, et al. 2017. "Human Pancreatic β Cell lncRNAs Control Cell-Specific Regulatory Networks." *Cell Metabolism* 25 (2): 400–411. <https://doi.org/10.1016/j.cmet.2016.11.016>.
- Alasoo, Kaur, Julia Rodrigues, Subhankar Mukhopadhyay, Andrew J. Knights, Alice L. Mann, Kousik Kundu, Christine Hale, Gordon Dougan, and Daniel J. Gaffney. 2018. "Shared genetic effects on chromatin and gene expression indicate a role for enhancer priming in immune response." *Nature Genetics* 50 (3): 424–31. <https://doi.org/10.1038/s41588-018-0046-7>.
- Andersson, Robin, Claudia Gebhard, Irene Miguel-Escalada, Ilka Hoof, Jette Bornholdt, Mette Boyd, Yun Chen, et al. 2014. "An atlas of active enhancers across human cell types and tissues." *Nature* 507 (7493): 455–61. <https://doi.org/10.1038/nature12787>.
- Arda, H. Efsun, Jennifer Tsai, Yenny R. Rosli, Paul Giresi, Rita Bottino, William J. Greenleaf, Howard Y. Chang, and Seung K. Kim. 2018. "A Chromatin Basis for Cell Lineage and Disease Risk in the Human Pancreas." *Cell Systems* 7 (3): 310–322.e4. <https://doi.org/10.1016/j.cels.2018.07.007>.
- Babon, Jenny Aurielle B., Megan E. Denicola, David M. Blodgett, Inne Crèvecoeur, Thomas S. Buttrick, René Maehr, Rita Bottino, et al. 2016. "Analysis of self-antigen specificity of islet-infiltrating T cells from human donors with type 1 diabetes." *Nature Medicine* 22 (12): 1482–7. <https://doi.org/10.1038/nm.4203>.
- Bannister, Andrew J., and Tony Kouzarides. 2011. "Regulation of chromatin by histone modifications." *Cell Res.* <https://doi.org/10.1038/cr.2011.22>.
- Barutcu, A. Rasim, Bryan R. Lajoie, Rachel P. McCord, Coralee E. Tye, Deli Hong, Terri L. Messier, Gillian Browne, et al. 2015. "Chromatin interaction analysis reveals changes in small chromosome and telomere clustering between epithelial and breast cancer cells." *Genome Biology* 16 (1): 214. <https://doi.org/10.1186/s13059-015-0768-0>.

- Battaglia, Manuela, Mark S. Anderson, Jane H. Buckner, Susan M. Geyer, Peter A. Gottlieb, Thomas W. H. Kay, Åke Lernmark, et al. 2017. "Understanding and preventing type 1 diabetes through the unique working model of TrialNet." Springer Verlag. <https://doi.org/10.1007/s00125-017-4384-2>.
- Bell, Joshua S. K., and Paula M. Vertino. 2017. "Orphan CpG islands define a novel class of highly active enhancers." *Epigenetics* 12 (6): 449–64. <https://doi.org/10.1080/15592294.2017.1297910>.
- Bernstein, Bradley E., Tarjei S. Mikkelsen, Xiaohui Xie, Michael Kamal, Dana J. Huebert, James Cuff, Ben Fry, et al. 2006. "A Bivalent Chromatin Structure Marks Key Developmental Genes in Embryonic Stem Cells." *Cell* 125 (2): 315–26. <https://doi.org/10.1016/j.cell.2006.02.041>.
- Bhandare, Reena, Jonathan Schug, John Le Lay, Alan Fox, Olga Smirnova, Chengyang Liu, Ali Najj, and Klaus H. Kaestner. 2010. "Genome-wide analysis of histone modifications in human pancreatic islets." *Genome Research* 20 (4): 428–33. <https://doi.org/10.1101/gr.102038.109>.
- Bonàs-Guarch, Sílvia, Marta Guindo-Martínez, Irene Miguel-Escalada, Niels Grarup, David Sebastian, Elias Rodriguez-Fos, Friman Sánchez, et al. 2018. "Re-analysis of public genetic data reveals a rare X-chromosomal variant associated with type 2 diabetes." *Nature Communications* 9 (1): 321. <https://doi.org/10.1038/s41467-017-02380-9>.
- Bosco, Domenico, Mathieu Armanet, Philippe Morel, Nadja Niclauss, Antonino Sgroi, Yannick D. Muller, Laurianne Giovannoni, Géraldine Parnaud, and Thierry Berney. 2010. "Unique arrangement of α - and β -cells in human islets of Langerhans." *Diabetes* 59 (5): 1202–10. <https://doi.org/10.2337/db09-1177>.
- Boyle, Alan P., Sean Davis, Hennady P. Shulha, Paul Meltzer, Elliott H. Margulies, Zhiping Weng, Terrence S. Furey, and Gregory E. Crawford. 2008. "High-Resolution Mapping and Characterization of Open Chromatin across the Genome." *Cell* 132 (2): 311–22. <https://doi.org/10.1016/j.cell.2007.12.014>.
- Bramswig, Nuria C., Logan J. Everett, Jonathan Schug, Craig Dorrell, Chengyang Liu, Yanping Luo, Philip R. Streeter, Ali Najj, Markus Grompe, and Klaus H. Kaestner. 2013. "Epigenomic plasticity enables human pancreatic α to β cell reprogramming." *Journal of Clinical Investigation* 123 (3): 1275–84. <https://doi.org/10.1172/JCI66514>.
- Bruno, G., G. Gruden, and M. Songini. 2016. "Incidence of type 1 diabetes in age groups above 15 years: facts, hypothesis and prospects for future epidemiologic research." Springer-Verlag Italia s.r.l. <https://doi.org/10.1007/s00592-015-0835-8>.

Bucher, Pascal, Zoltan Mathe, Philippe Morel, Domenico Bosco, Axel Andres, Manfred Kurfuest, Olaf Friedrich, Nicole Raemisch-Guenther, Leo H Buhler, and Thierry Berney. 2005. "Assessment of a Novel Two-Component Enzyme Preparation for Human Islet Isolation and Transplantation." *Transplantation* 79 (1): 91–97. <https://doi.org/10.1097/01.TP.0000147344.73915.C8>.

Buenrostro, Jason D., Paul G. Giresi, Lisa C. Zaba, Howard Y. Chang, and William J. Greenleaf. 2013. "Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position." *Nature Methods* 10 (12): 1213–8. <https://doi.org/10.1038/nmeth.2688>.

Bulger, Michael, and Mark Groudine. 2010. "Enhancers: The abundance and function of regulatory sequences beyond promoters." *Developmental Biology* 339 (2): 250–57. <https://doi.org/10.1016/j.ydbio.2009.11.035>.

Buschbeck, Marcus, and Sandra B. Hake. 2017. "Variants of core histones and their roles in cell fate decisions, development and cancer." *Nature Reviews Molecular Cell Biology* 18 (5): 299–314. <https://doi.org/10.1038/nrm.2016.166>.

Buzzetti, Raffaella, Simona Zampetti, and Ernesto Maddaloni. 2017. "Adult-onset autoimmune diabetes: Current knowledge and implications for management." Nature Publishing Group. <https://doi.org/10.1038/nrendo.2017.99>.

Calderon, Diego, Michelle L. T. Nguyen, Anja Mezger, Arwa Kathiria, Fabian Müller, Vinh Nguyen, Ninnia Lescano, et al. 2019. "Landscape of stimulation-responsive chromatin across diverse human immune cells." *Nature Genetics* 51 (10): 1494–1505. <https://doi.org/10.1038/s41588-019-0505-9>.

Campbell-Thompson, Martha. 2015. "Organ donor specimens: What can they tell us about type 1 diabetes?" Blackwell Publishing Ltd. <https://doi.org/10.1111/pedi.12286>.

Carninci, P., T. Kasukawa, S. Katayama, J. Gough, M. C. Frith, N. Maeda, R. Oyama, et al. 2005. "Molecular biology: The transcriptional landscape of the mammalian genome." *Science* 309 (5740): 1559–63. <https://doi.org/10.1126/science.1112014>.

Cebola, Inês, Santiago A. Rodríguez-Seguí, Candy H. H. Cho, José Bessa, Meritxell Rovira, Mario Luengo, Mariya Chhatrivala, et al. 2015. "TEAD and YAP regulate the enhancer network of human embryonic pancreatic progenitors." *Nature Cell Biology* 17 (5): 615–26. <https://doi.org/10.1038/ncb3160>.

Chang, Winston, Joe Cheng, JJ Allaire, Yihui Xie, and Jonathan McPherson. 2019. *Shiny: Web Application Framework for R*. <https://CRAN.R-project.org/package=shiny>.

- Cehadeh, Wassim, Jacques Weill, Marie-Christine Vantghem, Gunar Alm, Jean Lefèbvre, Pierre Wattré, and Didier Hober. 2000. "Increased Level of Interferon- α in Blood of Patients with Insulin-Dependent Diabetes Mellitus: Relationship with Coxsackievirus B Infection." *The Journal of Infectious Diseases* 181 (6): 1929–39. <https://doi.org/10.1086/315516>.
- Collas, Philippe. 2010. "The current state of chromatin immunoprecipitation." Springer. <https://doi.org/10.1007/s12033-009-9239-8>.
- Colli, Maikel L., Mireia Ramos-Rodríguez, Ernesto S. Nakayasu, Maria I. Alvelos, Miguel Lopes, Jessica L. E. Hill, Jean-Valery Turatsinze, et al. 2020. "An integrated multi-omics approach identifies the landscape of interferon- α -mediated responses of human pancreatic beta cells." *Nature Communications* 2020 11:1 11 (1): 1–17. <https://doi.org/10.1038/s41467-020-16327-0>.
- Conway, Eric, Evan Healy, and Adrian P. Bracken. 2015. "PRC2 mediated H3K27 methylations in cellular identity and cancer." *Current Opinion in Cell Biology* 37: 42–48. <https://doi.org/10.1016/j.ceb.2015.10.003>.
- Corbin, Karen D., Kimberly A. Driscoll, Richard E. Pratley, Steven R. Smith, David M. Maahs, and Elizabeth J. Mayer-Davis. 2018. "Obesity in type 1 diabetes: Pathophysiology, clinical impact, and mechanisms." Oxford University Press. <https://doi.org/10.1210/er.2017-00191>.
- Creyghton, Menno P., Albert W. Cheng, G. Grant Welstead, Tristan Kooistra, Bryce W. Carey, Eveline J. Steine, Jacob Hanna, et al. 2010. "Histone H3K27ac separates active from poised enhancers and predicts developmental state." *Proceedings of the National Academy of Sciences* 107 (50): 21931–6. <https://doi.org/10.1073/pnas.1016071107>.
- Curradi, Michela, Annalisa Izzo, Gianfranco Badaracco, and Nicoletta Landsberger. 2002. "Molecular Mechanisms of Gene Silencing Mediated by DNA Methylation." *Molecular and Cellular Biology* 22 (9): 3157–73. <https://doi.org/10.1128/mcb.22.9.3157-3173.2002>.
- Dahlquist, G. 2006. "Can we slow the rising incidence of childhood-onset autoimmune diabetes? The overload hypothesis." Springer. <https://doi.org/10.1007/s00125-005-0076-4>.
- Dekker, Job, Karsten Rippe, Martijn Dekker, and Nancy Kleckner. 2002. "Capturing chromosome conformation." *Science* 295 (5558): 1306–11. <https://doi.org/10.1126/science.1067799>.
- DeLong, Thomas, Timothy A. Wiles, Rocky L. Baker, Brenda Bradley, Gene Barbour, Richard Reisdorph, Michael Armstrong, et al. 2016. "Pathogenic CD4 T cells in type 1 diabetes recognize epitopes formed by peptide fusion." *Science* 351 (6274): 711–14. <https://doi.org/10.1126/science.aad2791>.

Dixon, Jesse R., Inkyung Jung, Siddarth Selvaraj, Yin Shen, Jessica E. Antosiewicz-Bourget, Ah Young Lee, Zhen Ye, et al. 2015. "Chromatin architecture reorganization during stem cell differentiation." *Nature* 518 (7539): 331–36. <https://doi.org/10.1038/nature14222>.

Dixon, Jesse R., Siddarth Selvaraj, Feng Yue, Audrey Kim, Yan Li, Yin Shen, Ming Hu, Jun S. Liu, and Bing Ren. 2012. "Topological domains in mammalian genomes identified by analysis of chromatin interactions." *Nature* 485 (7398): 376–80. <https://doi.org/10.1038/nature11082>.

Docker Inc. n.d. "Empowering App Development for Developers | Docker." Accessed March 2, 2020. <https://www.docker.com/>.

Eizirik, Décio L., Maikel L. Colli, and Fernanda Ortis. 2009. "The role of inflammation in insulinitis and B-cell loss in type 1 diabetes." *Nature Reviews Endocrinology* 5 (4): 219–26. <https://doi.org/10.1038/nrendo.2009.21>.

Eizirik, Décio L., Michael Sammeth, Thomas Bouckennooghe, Guy Bottu, Giorgia Sisino, Mariana Igoillo-Esteve, Fernanda Ortis, et al. 2012. "The human pancreatic islet transcriptome: Expression of candidate genes for type 1 diabetes and the impact of pro-inflammatory cytokines." Edited by Greg Gibson. *PLoS Genetics* 8 (3): e1002552. <https://doi.org/10.1371/journal.pgen.1002552>.

Fabris, P, A Floreani, F.De Lazzari, C Betterle, N. A. Greggio, R Naccarato, and M Chiaramonte. 1992. "Development of type 1 diabetes mellitus during interferon alfa therapy for chronic HCV hepatitis." *The Lancet* 340 (8818): 548. [https://doi.org/10.1016/0140-6736\(92\)91744-S](https://doi.org/10.1016/0140-6736(92)91744-S).

Farh, Kyle Kai-How, Alexander Marson, Jiang Zhu, Markus Kleinewietfeld, William J Housley, Samantha Beik, Noam Shoresh, et al. 2015. "Genetic and epigenetic fine mapping of causal autoimmune disease variants." *Nature* 518 (7539): 337–43. <https://doi.org/10.1038/nature13835>.

Fortin, Jean Philippe, and Kasper D. Hansen. 2015. "Reconstructing A/B compartments as revealed by Hi-C using long-range correlations in epigenetic data." *Genome Biology* 16 (1). <https://doi.org/10.1186/s13059-015-0741-y>.

Foulis, Alan K., Maura A. Farquharson, and Anthony Meager. 1987. "Immunoreactive A-Interferon in Insulin-Secreting B Cells in Type 1 Diabetes Mellitus." *The Lancet* 330 (8573): 1423–7. [https://doi.org/10.1016/S0140-6736\(87\)91128-7](https://doi.org/10.1016/S0140-6736(87)91128-7).

Fu, Wenxian, Julia Farache, Susan M. Clardy, Kimie Hattori, Palwinder Mander, Kevin Lee, Inmaculada Rioja, et al. 2014. "Epigenetic modulation of type-1 diabetes via a dual effect on pancreatic macrophages and β cells." *eLife* 3 (November): 1–20. <https://doi.org/10.7554/eLife.04631>.

Fudenberg, Geoffrey, Maxim Imakaev, Carolyn Lu, Anton Goloborodko, Nezar

- Abdennur, and Leonid A. Mirny. 2016. "Formation of Chromosomal Domains by Loop Extrusion." *Cell Reports* 15 (9): 2038–49. <https://doi.org/10.1016/j.celrep.2016.04.085>.
- García-González, Estela, Martín Escamilla-Del-Arenal, Rodrigo Arzate-Mejía, and Félix Recillas-Targa. 2016. "Chromatin remodeling effects on enhancer activity." *Cellular and Molecular Life Sciences* 73 (15): 2897–2910. <https://doi.org/10.1007/s00018-016-2184-3>.
- Gaulton, Kyle J, Takao Nanno, Lorenzo Pasquali, Jeremy M Simon, Paul G Giresi, Marie P Fogarty, Tami M Panhuis, et al. 2010. "A map of open chromatin in human pancreatic islets." *Nature Genetics* 42 (3): 255–59. <https://doi.org/10.1038/ng.530>.
- Genovese, Mark C., Joel Kremer, Omid Zamani, Charles Ludivico, Marek Krogulec, Li Xie, Scott D. Beattie, et al. 2016. "Baricitinib in patients with refractory rheumatoid arthritis." *New England Journal of Medicine* 374 (13): 1243–52. <https://doi.org/10.1056/NEJMoa1507247>.
- Gepts, W. 1965. "Pathologic Anatomy of the Pancreas in Juvenile Diabetes Mellitus." *Diabetes* 14 (10): 619–33. <https://doi.org/10.2337/diab.14.10.619>.
- Giresi, Paul G., Jonghwan Kim, Ryan M. McDaniell, Vishwanath R. Iyer, and Jason D. Lieb. 2007. "FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) isolates active regulatory elements from human chromatin." *Genome Research* 17 (6): 877–85. <https://doi.org/10.1101/gr.5533506>.
- Goll, Mary Grace, and Timothy H. Bestor. 2005. "Eukaryotic cytosine methyltransferases." *Annual Review of Biochemistry* 74 (1): 481–514. <https://doi.org/10.1146/annurev.biochem.74.010904.153721>.
- Gosselin, David, Verena M. Link, Casey E. Romanoski, Gregory J. Fonseca, Dawn Z. Eichenfield, Nathanael J. Spann, Joshua D. Stender, et al. 2014. "Environment drives selection and function of enhancers controlling tissue-specific macrophage identities." *Cell* 159 (6): 1327–40. <https://doi.org/10.1016/j.cell.2014.11.023>.
- Gysemans, C., H. Callewaert, F. Moore, M. Nelson-Holte, L. Overbergh, D. L. Eizirik, and C. Mathieu. 2009. "Interferon regulatory factor-1 is a key transcription factor in murine beta cells under immune attack." *Diabetologia* 52 (11): 2374–84. <https://doi.org/10.1007/s00125-009-1514-5>.
- Haarhuis, Judith H. I., Robin H. van der Weide, Vincent A. Blomen, J. Omar Yáñez-Cuna, Mario Amendola, Marjon S. van Ruiten, Peter H. L. Krijger, et al. 2017. "The Cohesin Release Factor WAPL Restricts Chromatin Loop Extension." *Cell* 169 (4): 693–707.e14. <https://doi.org/10.1016/j.cell.2017.04.013>.
- Haberle, Vanja, and Boris Lenhard. 2016. "Promoter architectures and

developmental gene regulation.” Academic Press. <https://doi.org/10.1016/j.semcdb.2016.01.014>.

Harjutsalo, Valma, Lena Sjöberg, and Jaakko Tuomilehto. 2008. “Time trends in the incidence of type 1 diabetes in Finnish children: a cohort study.” *The Lancet* 371 (9626): 1777–82. [https://doi.org/10.1016/S0140-6736\(08\)60765-5](https://doi.org/10.1016/S0140-6736(08)60765-5).

Härkönen, Taina, Anja Paananen, Hilikka Lankinen, Tapani Hovi, Outi Vaarala, and Merja Roivainen. 2003. “Enterovirus infection may induce humoral immune response reacting with islet cell autoantigens in humans.” *Journal of Medical Virology* 69 (3): 426–40. <https://doi.org/10.1002/jmv.10306>.

Heintzman, Nathaniel D., Gary C. Hon, R. David Hawkins, Pouya Kheradpour, Alexander Stark, Lindsey F. Harp, Zhen Ye, et al. 2009. “Histone modifications at human enhancers reflect global cell-type-specific gene expression.” *Nature* 459 (7243): 108–12. <https://doi.org/10.1038/nature07829>.

Heintzman, Nathaniel D., Rhona K. Stuart, Gary Hon, Yutao Fu, Christina W Ching, R. David Hawkins, Leah O. Barrera, et al. 2007. “Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome.” *Nature Genetics* 39 (3): 311–18. <https://doi.org/10.1038/ng1966>.

Heinz, S., C. E. Romanoski, C. Benner, K. A. Allison, M. U. Kaikkonen, L. D. Orozco, and C. K. Glass. 2013. “Effect of natural genetic variation on enhancer selection and function.” *Nature* 503 (7477): 487–92. <https://doi.org/10.1038/nature12615>.

Hernandez-Garcia, Carlos M., and John J. Finer. 2014. “Identification and validation of promoters and cis-acting regulatory elements.” Elsevier. <https://doi.org/10.1016/j.plantsci.2013.12.007>.

Herrath, Matthias von, Srinath Sanda, and Kevan Herold. 2007. “Type 1 diabetes as a relapsing–remitting disease?” *Nature Reviews Immunology* 7 (12): 988–94. <https://doi.org/10.1038/nri2192>.

Hindorf, Lucia A., Praveen Sethupathy, Heather A. Junkins, Erin M. Ramos, Jayashri P. Mehta, Francis S. Collins, and Teri A. Manolio. 2009. “Potential etiologic and functional implications of genome-wide association loci for human diseases and traits.” *Proceedings of the National Academy of Sciences of the United States of America* 106 (23): 9362–7. <https://doi.org/10.1073/pnas.0903103106>.

Ilonen, Jorma, Johanna Lempainen, and Riitta Veijola. 2019. “The heterogeneous pathogenesis of type 1 diabetes mellitus.” *Nature Reviews Endocrinology* 15 (11): 635–50. <https://doi.org/10.1038/s41574-019-0254-y>.

Jabbari, Kamel, and Giorgio Bernardi. 2004. “Cytosine methylation and CpG, TpG (CpA) and TpA frequencies.” *Gene* 333 (SUPPL.): 143–49. <https://doi.org/10.1016/j.gene.2004.05.014>.

[org/10.1016/j.gene.2004.02.043](https://doi.org/10.1016/j.gene.2004.02.043).

Javierre, Biola M., Sven Sewitz, Jonathan Cairns, Steven W. Wingett, Csilla Várnai, Michiel J. Thiecke, Paula Freire-Pritchett, et al. 2016. “Lineage-Specific Genome Architecture Links Enhancers and Non-coding Disease Variants to Target Gene Promoters.” *Cell* 167 (5): 1369–1384.e19. <https://doi.org/10.1016/j.cell.2016.09.037>.

Kallionpaa, H., Laura L. Elo, E. Laajala, J. Mykkanen, I. Ricano-Ponce, Matti Vaarma, Teemu D. Laajala, et al. 2014. “Innate Immune Activity Is Detected Prior to Seroconversion in Children With HLA-Conferred Type 1 Diabetes Susceptibility.” *Diabetes* 63 (7): 2402–14. <https://doi.org/10.2337/db13-1775>.

Karvonen, Marjatta, Maarit Viik-Kajander, Elena Moltchanova, Ingrid Libman, Ronald LaPorte, and Jaakko Tuomilehto. 2000. “Incidence of childhood type 1 diabetes worldwide.” *Diabetes Care* 23 (10): 1516–26. <https://doi.org/10.2337/diacare.23.10.1516>.

Katsarou, Anastasia, Soffia Gudbjörnsdottir, Araz Rawshani, Dana Dabelea, Ezio Bonifacio, Barbara J. Anderson, Laura M. Jacobsen, Desmond A. Schatz, and Ake Lernmark. 2017. “Type 1 diabetes mellitus.” *Nature Reviews Disease Primers* 3: 1–18. <https://doi.org/10.1038/nrdp.2017.16>.

Kaya-Okur, Hatice S., Steven J. Wu, Christine A. Codomo, Erica S. Pledger, Terri D. Bryson, Jorja G. Henikoff, Kami Ahmad, and Steven Henikoff. 2019. “CUT&Tag for efficient epigenomic profiling of small samples and single cells.” *Nature Communications* 10 (1): 1930. <https://doi.org/10.1038/s41467-019-09982-5>.

Kempfer, Rieke, and Ana Pombo. 2020. “Methods for mapping 3D chromosome architecture.” *Nature Reviews Genetics* 21 (4): 207–26. <https://doi.org/10.1038/s41576-019-0195-2>.

Kent, W. J., C. W. Sugnet, T. S. Furey, K. M. Roskin, T. H. Pringle, A. M. Zahler, and a. D. Haussler. 2002. “The Human Genome Browser at UCSC.” *Genome Research* 12 (6): 996–1006. <https://doi.org/10.1101/gr.229102>.

Kim-Hellmuth, Sarah, Matthias Bechheim, Benno Pütz, Pejman Mohammadi, Johann Nédélec, Nicholas Giangreco, Jessica Becker, et al. 2017. “Genetic regulatory effects modified by immune activation contribute to autoimmune disease associations.” *Nature Communications* 8 (1): 266. <https://doi.org/10.1038/s41467-017-00366-1>.

Klutstein, Michael, Deborah Nejman, Razi Greenfield, and Howard Cedar. 2016. “DNA methylation in cancer and aging.” American Association for Cancer Research Inc. <https://doi.org/10.1158/0008-5472.CAN-15-3278>.

Krischer, Jeffrey P., Kristian F. Lynch, Desmond A. Schatz, Jorma Ilonen, Åke Lernmark, William A. Hagopian, Marian J. Rewers, et al.

2015. “The 6 year incidence of diabetes-associated autoantibodies in genetically at-risk children: the TEDDY study.” *Diabetologia* 58 (5): 980–87. <https://doi.org/10.1007/s00125-015-3514-y>.

Krogvold, Lars, Anna Wiberg, Bjørn Edwin, Trond Buanes, Frode Lars Jahnsen, Kristian F. Hanssen, Erik Larsson, Olle Korsgren, Oskar Skog, and Knut Dahl-Jørgensen. 2016. “Insulinitis and characterisation of infiltrating T cells in surgical pancreatic tail resections from patients at onset of type 1 diabetes.” *Diabetologia* 59 (3): 492–501. <https://doi.org/10.1007/s00125-015-3820-4>.

Laat, Wouter de, and Denis Duboule. 2013. “Topology of mammalian developmental enhancers and their regulatory landscapes.” *Nature* 502 (7472): 499–506. <https://doi.org/10.1038/nature12753>.

Larsson, Anton J. M., Per Johnsson, Michael Hagemann-Jensen, Leonard Hartmanis, Omid R. Faridani, Björn Reinius, Åsa Segerstolpe, Chloe M. Rivera, Bing Ren, and Rickard Sandberg. 2019. “Genomic encoding of transcriptional burst kinetics.” *Nature* 565 (7738): 251–54. <https://doi.org/10.1038/s41586-018-0836-1>.

Lavin, Yonit, Deborah Winter, Ronnie Blecher-Gonen, Eyal David, Hadas Keren-Shaul, Miriam Merad, Steffen Jung, and Ido Amit. 2014. “Tissue-resident macrophage enhancer landscapes are shaped by the local microenvironment.” *Cell* 159 (6): 1312–26. <https://doi.org/10.1016/j.cell.2014.11.018>.

Law, Julie A., and Steven E. Jacobsen. 2010. “Establishing, maintaining and modifying DNA methylation patterns in plants and animals.” Nature Publishing Group. <https://doi.org/10.1038/nrg2719>.

Le Dily, François, Davide Baù, Andy Pohl, Guillermo P. Vicent, François Serra, Daniel Soronellas, Giancarlo Castellano, et al. 2014. “Distinct structural transitions of chromatin topological domains correlate with coordinated hormone-induced gene regulation.” *Genes & Development* 28 (19): 2151–62. <https://doi.org/10.1101/gad.241422.114>.

Lieberman-Aiden, Erez, Nynke L. Van Berkum, Louise Williams, Maxim Imakaev, Tobias Ragozy, Agnes Telling, Ido Amit, et al. 2009. “Comprehensive mapping of long-range interactions reveals folding principles of the human genome.” *Science* 326 (5950): 289–93. <https://doi.org/10.1126/science.1181369>.

Littorin, B., P. Blom, A. Schölin, H. J. Arnqvist, G. Blohmé, J. Bolinder, A. Ekbom-Schnell, et al. 2006. “Lower levels of plasma 25-hydroxyvitamin D among young adults at diagnosis of autoimmune type 1 diabetes compared with control subjects: Results from the nationwide Diabetes Incidence Study in Sweden (DISS).” *Diabetologia* 49 (12): 2847–52. <https://doi.org/10.1007/s00125-006-0426-x>.

- Lopez-Bastida, Julio, Mauro Boronat, Juan Oliva Moreno, and Willemien Schurer. 2013. "Costs, outcomes and challenges for diabetes care in Spain." *Globalization and Health* 9 (1): 17. <https://doi.org/10.1186/1744-8603-9-17>.
- Lundberg, Marcus, Lars Krogvold, Enida Kuric, Knut Dahl-Jørgensen, and Oskar Skog. 2016. "Expression of Interferon-Stimulated Genes in Insulinitic Pancreatic Islets of Patients Recently Diagnosed With Type 1 Diabetes." *Diabetes* 65 (10): 3104–10. <https://doi.org/10.2337/db16-0616>.
- Margueron, Raphaël, and Danny Reinberg. 2011. "The Polycomb complex PRC2 and its mark in life." *Nature*. <https://doi.org/10.1038/nature09784>.
- Marroqui, Laura, Reinaldo S. Dos Santos, Anne Op de beeck, Alexandra Coomans de Brachène, Lorella Marselli, Piero Marchetti, and Decio L. Eizirik. 2017. "Interferon- α mediates human beta cell HLA class I overexpression, endoplasmic reticulum stress and apoptosis, three hallmarks of early human type 1 diabetes." *Diabetologia* 60 (4): 656–67. <https://doi.org/10.1007/s00125-016-4201-3>.
- Marroqui, Laura, Reinaldo Sousa Dos Santos, Tina Fløyel, Fabio A Grieco, Izortze Santin, Anne Op de beeck, Lorella Marselli, Piero Marchetti, Flemming Pociot, and Decio L Eizirik. 2015. "TYK2 , a Candidate Gene for Type 1 Diabetes, Modulates Apoptosis and the Innate Immune Response in Human Pancreatic β -Cells." *Diabetes* 64 (11): 3808–17. <https://doi.org/10.2337/db15-0362>.
- Martens, Joost H. A., and Hendrik G. Stunnenberg. 2013. "BLUEPRINT: Mapping human blood cell epigenomes." Ferrata Storti Foundation. <https://doi.org/10.3324/haematol.2013.094243>.
- Matsumoto, Shinichi. 2010. "Islet cell transplantation for Type 1 diabetes." *J Diabetes*. <https://doi.org/10.1111/j.1753-0407.2009.00048.x>.
- Mäkinen, Marjaana, Juha Mykkänen, Maarit Koskinen, Ville Simell, Riitta Veijola, Heikki Hyöty, Jorma Ilonen, Mikael Knip, Olli Simell, and Jorma Toppari. 2016. "Serum 25-hydroxyvitamin d concentrations in children progressing to autoimmunity and clinical type 1 diabetes." *Journal of Clinical Endocrinology and Metabolism* 101 (2): 723–29. <https://doi.org/10.1210/jc.2015-3504>.
- Meaburn, Karen J., and Tom Misteli. 2007. "Cell biology: Chromosome territories." Nature Publishing Group. <https://doi.org/10.1038/445379a>.
- Melzi, Raffaella, Alessia Mercalli, Valeria Sordi, Elisa Cantarelli, Rita Nano, Paola Maffi, Giovanni Sitia, et al. 2010. "Role of CCL2/MCP-1 in Islet Transplantation." *Cell Transplantation* 19 (8): 1031–46. <https://doi.org/10.3727/096368910X514639>.
- Meyer, Steffen, Martin Woodward, Christina Hertel, Philip Vlaicu, Yasmin Haque, Jaanika Kärner, Annalisa Macagno, et al. 2016. "AIRE-Deficient

Patients Harbor Unique High-Affinity Disease-Ameliorating Autoantibodies.” *Cell* 166 (3): 582–95. <https://doi.org/10.1016/j.cell.2016.06.024>.

Miguel-Escalada, Irene, Silvia Bonàs-Guarch, Inês Cebola, Joan Ponsa-Cobas, Julen Mendieta-Esteban, Goutham Atla, Biola M. Javierre, et al. 2019. “Human pancreatic islet three-dimensional chromatin architecture provides insights into the genetics of type 2 diabetes.” *Nature Genetics* 51 (7): 1137–48. <https://doi.org/10.1038/s41588-019-0457-0>.

Moran, Sebastian, Carles Arribas, and Manel Esteller. 2016. “Validation of a DNA methylation microarray for 850,000 CpG sites of the human genome enriched in enhancer sequences.” *Epigenomics* 8 (3): 389–99. <https://doi.org/10.2217/epi.15.114>.

Morán, Ignasi, İdem Akerman, Martijn van de Bunt, Ruiyu Xie, Marion Benazra, Takao Nammo, Luis Arnes, et al. 2012. “Human β Cell Transcriptome Analysis Uncovers lncRNAs That Are Tissue-Specific, Dynamically Regulated, and Abnormally Expressed in Type 2 Diabetes.” *Cell Metabolism* 16 (4): 435–48. <https://doi.org/10.1016/j.cmet.2012.08.010>.

Morey, Lluís, and Kristian Helin. 2010. “Polycomb group protein-mediated repression of transcription.” Elsevier Current Trends. <https://doi.org/10.1016/j.tibs.2010.02.009>.

Morgan, Noel G., and Sarah J. Richardson. 2018. “Fifty years of pancreatic islet pathology in human type 1 diabetes: insights gained and progress made.” Springer Verlag. <https://doi.org/10.1007/s00125-018-4731-y>.

Morris, Andrew P, Benjamin F Voight, Tanya M Teslovich, Teresa Ferreira, Ayellet V Segrè, Valgerdur Steinthorsdottir, Rona J Strawbridge, et al. 2012. “Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes.” *Nature Genetics* 44 (9): 981–90. <https://doi.org/10.1038/ng.2383>.

Mularoni, Loris, Mireia Ramos-Rodríguez, and Lorenzo Pasquali. 2017. “The Pancreatic Islet Regulome Browser.” *Frontiers in Genetics* 8 (FEB): 13. <https://doi.org/10.3389/fgene.2017.00013>.

Mumbach, Maxwell R., Adam J. Rubin, Ryan A. Flynn, Chao Dai, Paul A. Khavari, William J. Greenleaf, and Howard Y. Chang. 2016. “HiChIP: efficient and sensitive analysis of protein-directed genome architecture.” *Nature Methods* 13 (11): 919–22. <https://doi.org/10.1038/nmeth.3999>.

Muraro, Mauro J., Gitanjali Dharmadhikari, Dominic Grün, Nathalie Groen, Tim Dielen, Erik Jansen, Leon van Gurp, et al. 2016. “A Single-Cell Transcriptome Atlas of the Human Pancreas.” *Cell Systems* 3 (4): 385–394.e3. <https://doi.org/10.1016/j.cels.2016.09.002>.

Nakazawa, Tetsuya, Jo Satoh, Kazuma Takahashi, Yoshiyuki Sakata,

- Fumiko Ikehata, Yumiko Takizawa, Shin Ichiro Bando, et al. 2001. "Complete suppression of insulinitis and diabetes in NOD mice lacking interferon regulatory factor-1." *Journal of Autoimmunity* 17 (2): 119–25. <https://doi.org/10.1006/jaut.2001.0531>.
- Nasrallah, Rabab, Charlotte J. Imianowski, Lara Bossini-Castillo, Francis M. Grant, Mikail Dogan, Lindsey Placek, Lina Kozhaya, et al. 2020. "A distal enhancer at risk locus 11q13.5 promotes suppression of colitis by Treg cells." *Nature*, no. March 2019 (May). <https://doi.org/10.1038/s41586-020-2296-7>.
- Niinistö, Sari, Hanna-Mari Takkinen, Iris Erlund, Suvi Ahonen, Jorma Toppari, Jorma Ilonen, Riitta Veijola, Mikael Knip, Outi Vaarala, and Suvi M. Virtanen. 2017. "Fatty acid status in infancy is associated with the risk of type 1 diabetes-associated autoimmunity." *Diabetologia* 60 (7): 1223–33. <https://doi.org/10.1007/s00125-017-4280-9>.
- Nora, Elphège P., Bryan R. Lajoie, Edda G. Schulz, Luca Giorgetti, Ikuhiro Okamoto, Nicolas Servant, Tristan Piolot, et al. 2012. "Spatial partitioning of the regulatory landscape of the X-inactivation centre." *Nature* 485 (7398): 381–85. <https://doi.org/10.1038/nature11049>.
- Norris, Jill M., Hye Seung Lee, Brittni Frederiksen, Iris Erlund, Ulla Uusitalo, Jimin Yang, Åke Lernmark, et al. 2018. "Plasma 25-Hydroxyvitamin D concentration and risk of islet autoimmunity." *Diabetes* 67 (1): 146–54. <https://doi.org/10.2337/db17-0802>.
- Onengut-Gumuscu, Suna, Wei-Min Chen, Oliver Burren, Nick J Cooper, Aaron R. Quinlan, Josyf C. Mychaleckyj, Emily Farber, et al. 2015. "Fine mapping of type 1 diabetes susceptibility loci and evidence for colocalization of causal variants with lymphoid gene enhancers." *Nature Genetics* 47 (4): 381–86. <https://doi.org/10.1038/ng.3245>.
- Ooi, Steen K. T., Chen Qiu, Emily Bernstein, Keqin Li, Da Jia, Zhe Yang, Hediye Erdjument-Bromage, et al. 2007. "DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA." *Nature* 448 (7154): 714–17. <https://doi.org/10.1038/nature05987>.
- Ostuni, Renato, Viviana Piccolo, Iros Barozzi, Sara Polletti, Alberto Termanini, Silvia Bonifacio, Alessia Curina, Elena Prosperini, Serena Ghisletti, and Gioacchino Natoli. 2013. "Latent Enhancers Activated by Stimulation in Differentiated Cells." *Cell* 152 (1-2): 157–71. <https://doi.org/10.1016/j.cell.2012.12.018>.
- Pan, Yunbao, Guohong Liu, Fuling Zhou, Bojin Su, and Yirong Li. 2018. "DNA methylation profiles in cancer diagnosis and therapeutics." Springer-Verlag Italia s.r.l. <https://doi.org/10.1007/s10238-017-0467-0>.
- Parker, Stephen C J, Michael L Stitzel, D Leland Taylor, Jose Miguel Orozco,

Michael R Erdos, Jennifer A Akiyama, Kelly Lammerts van Bueren, et al. 2013. "Chromatin stretch enhancer states drive cell-specific gene regulation and harbor human disease risk variants." *Proceedings of the National Academy of Sciences* 110 (44): 17921–6. <https://doi.org/10.1073/pnas.1317023110>.

Pasquali, Lorenzo, Kyle J. Gaulton, Santiago A. Rodríguez-Seguí, Loris Mularoni, Irene Miguel-Escalada, İldem İldem Akerman, Juan J. Tena, et al. 2014. "Pancreatic islet enhancer clusters enriched in type 2 diabetes risk-associated variants." *Nature Genetics* 46 (2): 136–43. <https://doi.org/10.1038/ng.2870>.

Patterson, Christopher C., Gisela G. Dahlquist, Eva Gyürüs, Anders Green, Gyula Soltész, E. Schober, T. Waldhoer, et al. 2009. "Incidence trends for childhood type 1 diabetes in Europe during 1989-2003 and predicted new cases 2005-20: a multicentre prospective registration study." *The Lancet* 373 (9680): 2027–33. [https://doi.org/10.1016/S0140-6736\(09\)60568-7](https://doi.org/10.1016/S0140-6736(09)60568-7).

Pennisi, Elizabeth. 2012. "ENCODE Project Writes Eulogy for Junk DNA." *Science* 337 (6099): 1159–61. <https://doi.org/10.1126/science.337.6099.1159>.

Phanstiel, Douglas H., Kevin Van Bortle, Damek Spacek, Gaelen T. Hess, Muhammad Saad Shamim, Ido Machol, Michael I. Love, Erez Lieberman Aiden, Michael C. Bassik, and Michael P. Snyder. 2017. "Static and Dynamic DNA Loops form AP-1-Bound Activation Hubs during Macrophage Development." *Molecular Cell* 67 (6): 1037–1048.e6. <https://doi.org/10.1016/j.molcel.2017.08.006>.

Phillips, Jennifer E., and Victor G. Corces. 2009. "CTCF: Master Weaver of the Genome." NIH Public Access. <https://doi.org/10.1016/j.cell.2009.06.001>.

Phillips-Cremins, Jennifer E, Michael E G Sauria, Amartya Sanyal, Tatiana I Gerasimova, Bryan R Lajoie, Joshua S K Bell, Chin-Tong Ong, et al. 2013. "Architectural protein subclasses shape 3D organization of genomes during lineage commitment." *Cell* 153 (6): 1281–95. <https://doi.org/10.1016/j.cell.2013.04.053>.

Pociot, Flemming, and Åke Lernmark. 2016. "Genetic risk factors for type 1 diabetes." *The Lancet* 387 (10035): 2331–9. [https://doi.org/10.1016/S0140-6736\(16\)30582-7](https://doi.org/10.1016/S0140-6736(16)30582-7).

Pries, Lotta Katrin, Sinan Gülöksüz, and Gunter Kenis. 2017. "DNA methylation in schizophrenia." In *Advances in Experimental Medicine and Biology*, 978:211–36. Springer New York LLC. https://doi.org/10.1007/978-3-319-53889-1_12.

Pugliese, Alberto, Markus Zeller, Alarico Fernandez, Laura J. Zalcberg, Richard J. Bartlett, Camillo Ricordi, Massimo Pietropaolo, George S.

- Eisenbarth, Simon T. Bennett, and Dhavalkumar D. Patel. 1997. "The insulin gene is transcribed in the human thymus and transcription levels correlate with allelic variation at the INS VNTR-IDD2 susceptibility locus for type 1 diabetes." *Nature Genetics* 15 (3): 293–97. <https://doi.org/10.1038/ng0397-293>.
- Rada-Iglesias, Alvaro, Ruchi Bajpai, Tomek Swigut, Samantha A. Brugmann, Ryan A. Flynn, and Joanna Wysocka. 2011. "A unique chromatin signature uncovers early developmental enhancers in humans." *Nature* 470 (7333): 279–83. <https://doi.org/10.1038/nature09692>.
- Ramos-Rodríguez, Mireia, Helena Raurell-Vila, Maikel L Colli, Maria Inês Alvelos, Marc Subirana-Granés, Jonàs Juan-Mateu, Richard Norris, et al. 2019. "The impact of proinflammatory cytokines on the β -cell regulatory landscape provides insights into the genetics of type 1 diabetes." *Nature Genetics* 51 (11): 1588–95. <https://doi.org/10.1038/s41588-019-0524-6>.
- Rao, Suhas S. P., Su Chen Huang, Brian Glenn St Hilaire, Jesse M. Engreitz, Elizabeth M. Perez, Kyong Rim Kieffer-Kwon, Adrian L. Sanborn, et al. 2017. "Cohesin Loss Eliminates All Loop Domains." *Cell* 171 (2): 305–320.e24. <https://doi.org/10.1016/j.cell.2017.09.026>.
- Rao, Suhas S. P., Miriam H. Huntley, Neva C. Durand, Elena K. Stamenova, Ivan D. Bochkov, James T. Robinson, Adrian L. Sanborn, et al. 2014. "A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping." *Cell* 159 (7): 1665–80. <https://doi.org/10.1016/j.cell.2014.11.021>.
- Redondo, Maria J, Joy Jeffrey, Pamela R Fain, George S Eisenbarth, and Tihamer Orban. 2008. "Concordance for islet autoimmunity among monozygotic twins." <https://doi.org/10.1056/NEJMc0805398>.
- Richardson, Sarah J., Teresa Rodriguez-Calvo, Ivan C. Gerling, Clayton E. Mathews, John S. Kaddis, Mark A. Russell, Marie Zeissler, et al. 2016. "Islet cell hyperexpression of HLA class I antigens: a defining feature in type 1 diabetes." *Diabetologia* 59 (11): 2448–58. <https://doi.org/10.1007/s00125-016-4067-4>.
- Robinson, James T., Helga Thorvaldsdóttir, Wendy Winckler, Mitchell Guttman, Eric S. Lander, Gad Getz, and Jill P. Mesirov. 2011. "Integrative genomics viewer." Nature Publishing Group. <https://doi.org/10.1038/nbt.1754>.
- Roep, B. O., F. S. Kleijwegt, A. G. S. Van Halteren, V. Bonato, U. Boggi, F. Vendrame, P. Marchetti, and F. Dotta. 2010. "Islet inflammation and CXCL10 in recent-onset type 1 diabetes." *Clinical and Experimental Immunology* 159 (3): 338–43. <https://doi.org/10.1111/j.1365-2249.2009.04087.x>.
- Röder, Pia V, Bingbing Wu, Yixian Liu, and Weiping Han. 2016. "Pancreatic regulation of glucose homeostasis." *Experimental & Molecular Medicine*, 219.

<https://doi.org/10.1038/emm.2016.6>.

Sanborn, Adrian L, Suhas S P Rao, Su-Chen Huang, Neva C Durand, Miriam H Huntley, Andrew I Jewett, Ivan D Bochkov, et al. 2015. "Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes." *Proceedings of the National Academy of Sciences of the United States of America* 112 (47): E6456–65. <https://doi.org/10.1073/pnas.1518552112>.

Sandoval, Juan, Holger Heyn, Sebastian Moran, Jordi Serra-Musach, Miguel A Pujana, Marina Bibikova, and Manel Esteller. 2011. "Validation of a DNA methylation microarray for 450,000 CpG sites in the human genome." *Epigenetics* 6 (6): 692–702. <http://www.ncbi.nlm.nih.gov/pubmed/21593595>.

Saxonov, Serge, Paul Berg, and Douglas L. Brutlag. 2006. "A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters." *Proceedings of the National Academy of Sciences of the United States of America* 103 (5): 1412–7. <https://doi.org/10.1073/pnas.0510310103>.

Schmidl, Christian, André F Rendeiro, Nathan C Sheffield, and Christoph Bock. 2015. "ChIPmentation: fast, robust, low-input ChIP-seq for histones and transcription factors." *Nature Methods* 12 (10): 963–65. <https://doi.org/10.1038/nmeth.3542>.

Schmitt, Anthony D., Ming Hu, Inkyung Jung, Zheng Xu, Yunjiang Qiu, Catherine L. Tan, Yun Li, et al. 2016. "A Compendium of Chromatin Contact Maps Reveals Spatially Active Regions in the Human Genome." *Cell Reports* 17 (8): 2042–59. <https://doi.org/10.1016/j.celrep.2016.10.061>.

Schones, Dustin E., Kairong Cui, Suresh Cuddapah, Tae Young Roh, Artem Barski, Zhibin Wang, Gang Wei, and Keji Zhao. 2008. "Dynamic Regulation of Nucleosome Positioning in the Human Genome." *Cell* 132 (5): 887–98. <https://doi.org/10.1016/j.cell.2008.02.022>.

Schreuder, Tim C M A, Huub C Gelderblom, Christine J Weegink, Dörte Hamann, Henk W Reesink, J. Hans DeVries, Joost B L Hoekstra, and Peter L M Jansen. 2007. "High incidence of type 1 diabetes mellitus during or shortly after treatment with pegylated interferon α for chronic hepatitis C virus infection." *Liver International* 28 (1): 39–46. <https://doi.org/10.1111/j.1478-3231.2007.01610.x>.

Schwartzman, Omer, Zohar Mukamel, Noa Oded-Elkayam, Pedro Olivares-Chauvet, Yaniv Lubling, Gilad Landan, Shai Izraeli, and Amos Tanay. 2016. "UMI-4C for quantitative and targeted chromosomal contact profiling." *Nature Methods* 13 (8): 685–91. <https://doi.org/10.1038/nmeth.3922>.

Schwarzer, Wibke, Nezar Abdennur, Anton Goloborodko, Aleksandra

- Pekowska, Geoffrey Fudenberg, Yann Loe-Mie, Nuno A. Fonseca, et al. 2017. "Two independent modes of chromatin organization revealed by cohesin removal." *Nature* 551 (7678): 51–56. <https://doi.org/10.1038/nature24281>.
- Scott, Robert A, Vasiliki Lagou, Ryan P Welch, Eleanor Wheeler, May E Montasser, Jian'an Luan, Reedik Mägi, et al. 2012. "Large-scale association analyses identify new loci influencing glycemic traits and provide insight into the underlying biological pathways." *Nature Genetics* 44 (9): 991–1005. <https://doi.org/10.1038/ng.2385>.
- Simonis, Marieke, Petra Klous, Irene Homminga, Robert-Jan Galjaard, Erik-Jan Rijkers, Frank Grosveld, Jules P P Meijerink, and Wouter de Laat. 2009. "High-resolution identification of balanced and complex chromosomal rearrangements by 4C technology." *Nature Methods* 6 (11): 837–42. <https://doi.org/10.1038/nmeth.1391>.
- Simpson, M., H. Brady, X. Yin, J. Seifert, K. Barriga, M. Hoffman, T. Bugawan, et al. 2011. "No association of vitamin D intake or 25-hydroxyvitamin D levels in childhood with risk of islet autoimmunity and type 1 diabetes: the Diabetes Autoimmunity Study in the Young (DAISY)." *Diabetologia* 54 (11): 2779–88. <https://doi.org/10.1007/s00125-011-2278-2>.
- Skene, Peter J., and Steven Henikoff. 2017. "An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites." *eLife* 6 (January). <https://doi.org/10.7554/eLife.21856>.
- Skyler, J. S. 2011. "Immune intervention for type 1 diabetes mellitus." *Int J Clin Pract Suppl*. <https://doi.org/10.1111/j.1742-1241.2010.02580.x>.
- Songini, M., C. Mannu, C. Targhetta, and G. Bruno. 2017. "Type 1 diabetes in Sardinia: facts and hypotheses in the context of worldwide epidemiological data." Springer-Verlag Italia s.r.l. <https://doi.org/10.1007/s00592-016-0909-2>.
- Sossau, Daniel, Lukas Kofler, and Thomas Eigentler. 2017. "Type 1 diabetes mellitus caused by treatment with low-dose interferon- α in a melanoma patient." *Melanoma Research* 27 (5): 516–18. <https://doi.org/10.1097/CMR.000000000000381>.
- Spielmann, Malte, Darío G. Lupiáñez, and Stefan Mundlos. 2018. "Structural variation in the 3D genome." Nature Publishing Group. <https://doi.org/10.1038/s41576-018-0007-0>.
- Stitzel, Michael L., Praveen Sethupathy, Daniel S. Pearson, Peter S. Chines, Lingyun Song, Michael R. Erdos, Ryan Welch, et al. 2010. "Global epigenomic analysis of primary human pancreatic islets provides insights into type 2 diabetes susceptibility loci." *Cell Metabolism* 12 (5): 443–55. <https://doi.org/10.1016/j.cmet.2010.09.012>.
- Taberlay, Phillippa C., Joanna Achinger-Kawecka, Aaron T. L. Lun, Fabian

A Buske, Kenneth Sabir, Cathryn M Gould, Elena Zotenko, et al. 2016. "Three-dimensional disorganization of the cancer genome occurs coincident with long-range genetic and epigenetic alterations." *Genome Research* 26 (6): 719–31. <https://doi.org/10.1101/gr.201517.115>.

The DCCT Research Group. 1998. "Effect of Intensive Therapy on Residual β -Cell Function in Patients with Type 1 Diabetes in the Diabetes Control and Complications Trial." *Annals of Internal Medicine* 128 (7): 517. <https://doi.org/10.7326/0003-4819-128-7-199804010-00001>.

The ENCODE Project Consortium. 2012. "An integrated encyclopedia of DNA elements in the human genome." *Nature* 489 (7414): 57–74. <https://doi.org/10.1038/nature11247>.

Törn, Carina, David Hadley, Hye Seung Lee, William Hagopian, Åke Lernmark, Olli Simell, Marian Rewers, et al. 2015. "Role of type 1 diabetes- Associated snps on risk of autoantibody positivity in the TEDDY study." *Diabetes* 64 (5): 1818–29. <https://doi.org/10.2337/db14-1497>.

Tremblay, Martine W., and Yong-hui Jiang. 2019. "DNA Methylation and Susceptibility to Autism Spectrum Disorder." *Annual Review of Medicine* 70 (1): 151–66. <https://doi.org/10.1146/annurev-med-120417-091431>.

Trivedi, Prerak M., Kate L. Graham, Nicholas A. Scott, Misty R. Jenkins, Suktilang Majaw, Robyn M. Sutherland, Stacey Fynch, et al. 2017. "Repurposed JAK1/JAK2 inhibitor reverses established autoimmune insulinitis in NOD mice." *Diabetes* 66 (6): 1650–60. <https://doi.org/10.2337/db16-1250>.

Trynka, Gosia, and Soumya Raychaudhuri. 2013. "Using chromatin marks to interpret and localize genetic associations to complex human traits and diseases." NIH Public Access. <https://doi.org/10.1016/j.gde.2013.10.009>.

Tuomilehto, Jaakko. 2013. "The emerging global epidemic of type 1 diabetes." *Current Diabetes Reports* 13 (6): 795–804. <https://doi.org/10.1007/s11892-013-0433-5>.

Van Lummel, Menno, Gaby Duinkerken, Peter A. Van Veelen, Arnoud De Ru, Robert Cordfunke, Arnaud Zaldumbide, Iria Gomez-Touriño, et al. 2014. "Posttranslational modification of HLA-DQ binding islet autoantigens in type 1 diabetes." *Diabetes* 63 (1): 237–47. <https://doi.org/10.2337/db12-1214>.

Vastenhouw, Nadine L., and Alexander F. Schier. 2012. "Bivalent histone modifications in early embryogenesis." *Curr Opin Cell Biol*. <https://doi.org/10.1016/j.ceb.2012.03.009>.

Vatanen, Tommi, Aleksandar D. Kostic, Eva D'Hennezel, Heli Siljander, Eric A. Franzosa, Moran Yassour, Raivo Kolde, et al. 2016. "Variation in Microbiome LPS Immunogenicity Contributes to Autoimmunity in Humans." *Cell* 165 (4): 842–53. <https://doi.org/10.1016/j.cell.2016.04.007>.

- Wiberg, A., A. Granstam, S. Ingvast, T. Härkönen, M. Knip, O. Korsgren, and O. Skog. 2015. "Characterization of human organ donors testing positive for type 1 diabetes-associated autoantibodies." *Clinical and Experimental Immunology* 182 (3): 278–88. <https://doi.org/10.1111/cei.12698>.
- Wickham, Hadley, Winston Chang, Lionel Henry, Thomas Lin Pedersen, Kohske Takahashi, Claus Wilke, Kara Woo, and Hiroaki Yutani. 2019. *Ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics*. <https://CRAN.R-project.org/package=ggplot2>.
- Wilke, Claus O. 2019. *Cowplot: Streamlined Plot Theme and Plot Annotations for 'Ggplot2'*. <https://CRAN.R-project.org/package=cowplot>.
- Wutz, Gordana, Csilla Várnai, Kota Nagasaka, David A Cisneros, Roman R Stocsits, Wen Tang, Stefan Schoenfelder, et al. 2017. "Topologically associating domains and chromatin loops depend on cohesin and are regulated by CTCF, WAPL, and PDS5 proteins." *The EMBO Journal* 36 (24): 3573–99. <https://doi.org/10.15252/emj.201798004>.
- Xi, Hualin, Hennady P. Shulha, Jane M. Lin, Teresa R. Vales, Yutao Fu, David M. Bodine, Ronald D. G. McKay, et al. 2007. "Identification and characterization of cell type-specific and ubiquitous chromatin regulatory structures in the human genome." *PLoS Genetics* 3 (8): 1377–88. <https://doi.org/10.1371/journal.pgen.0030136>.
- Xu, Qianhua, and Wei Xie. 2018. "Epigenome in Early Mammalian Development: Inheritance, Reprogramming and Establishment." Elsevier Ltd. <https://doi.org/10.1016/j.tcb.2017.10.008>.
- Yin, Yimeng, Ekaterina Morgunova, Arttu Jolma, Eevi Kaasinen, Biswajyoti Sahu, Syed Khund-Sayeed, Pratyush K. Das, et al. 2017. "Impact of cytosine methylation on DNA binding specificities of human transcription factors." *Science* 356 (6337). <https://doi.org/10.1126/science.aaj2239>.
- Zaret, Kenneth S., and Jason S. Carroll. 2011. "Pioneer transcription factors: establishing competence for gene expression." *Genes & Development* 25 (21): 2227–41. <https://doi.org/10.1101/gad.176826.111>.
- Zhou, Xin, and Ting Wang. 2012. "Using the wash U Epigenome browser to examine genome-wide sequencing data." *Current Protocols in Bioinformatics* 0 10 (SUPPL.40): Unit10.10. <https://doi.org/10.1002/0471250953.bi1010s40>.
- Ziegler, Anette G., Marian Rewers, Olli Simell, Tuula Simell, Johanna Lempainen, Andrea Steck, Christiane Winkler, et al. 2013. "Seroconversion to multiple islet autoantibodies and risk of progression to diabetes in children." *JAMA - Journal of the American Medical Association* 309 (23): 2473–9. <https://doi.org/10.1001/jama.2013.6285>.

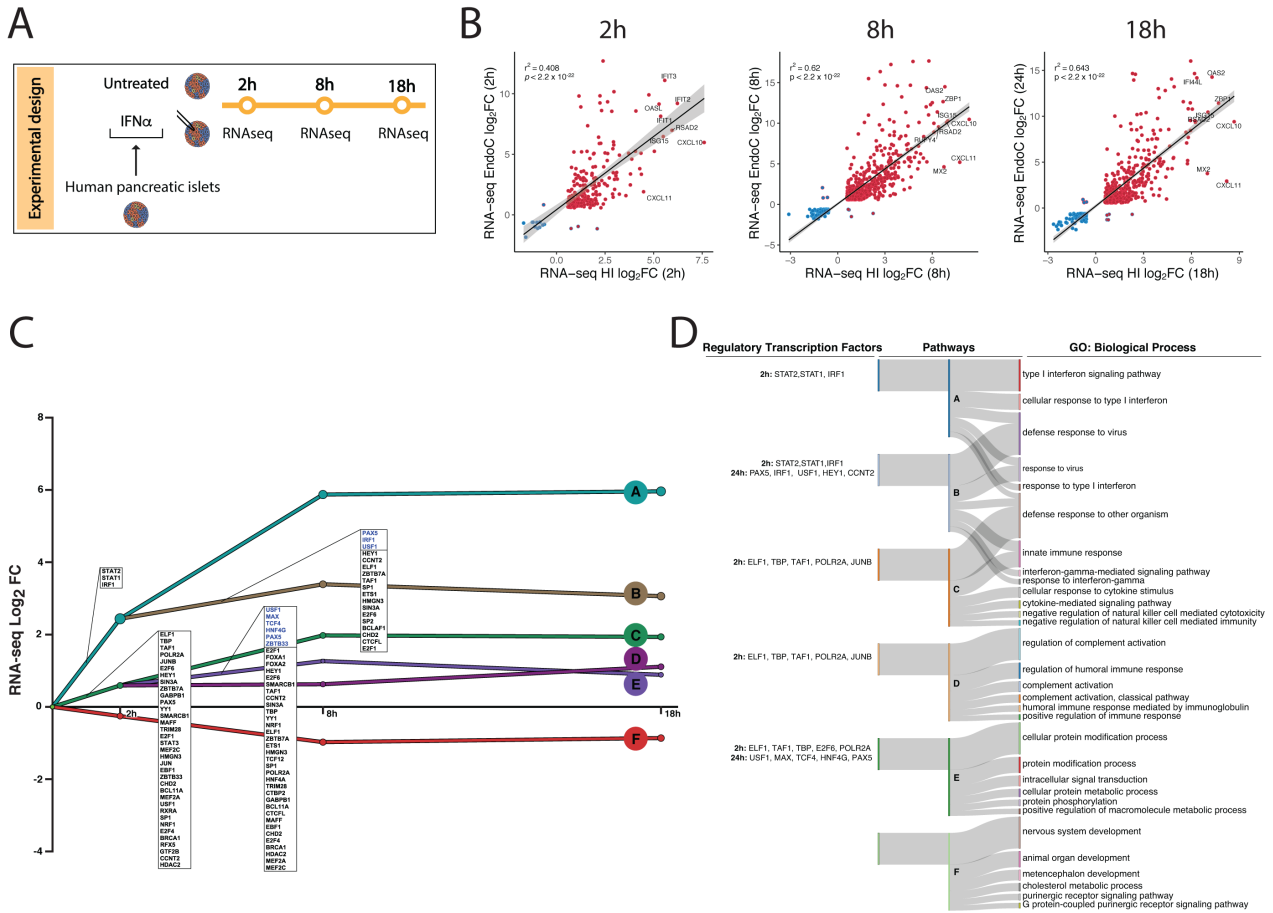
Part VII

Appendix

An integrated multi-omics approach identifies the landscape of interferon- α -mediated responses of human pancreatic beta cells

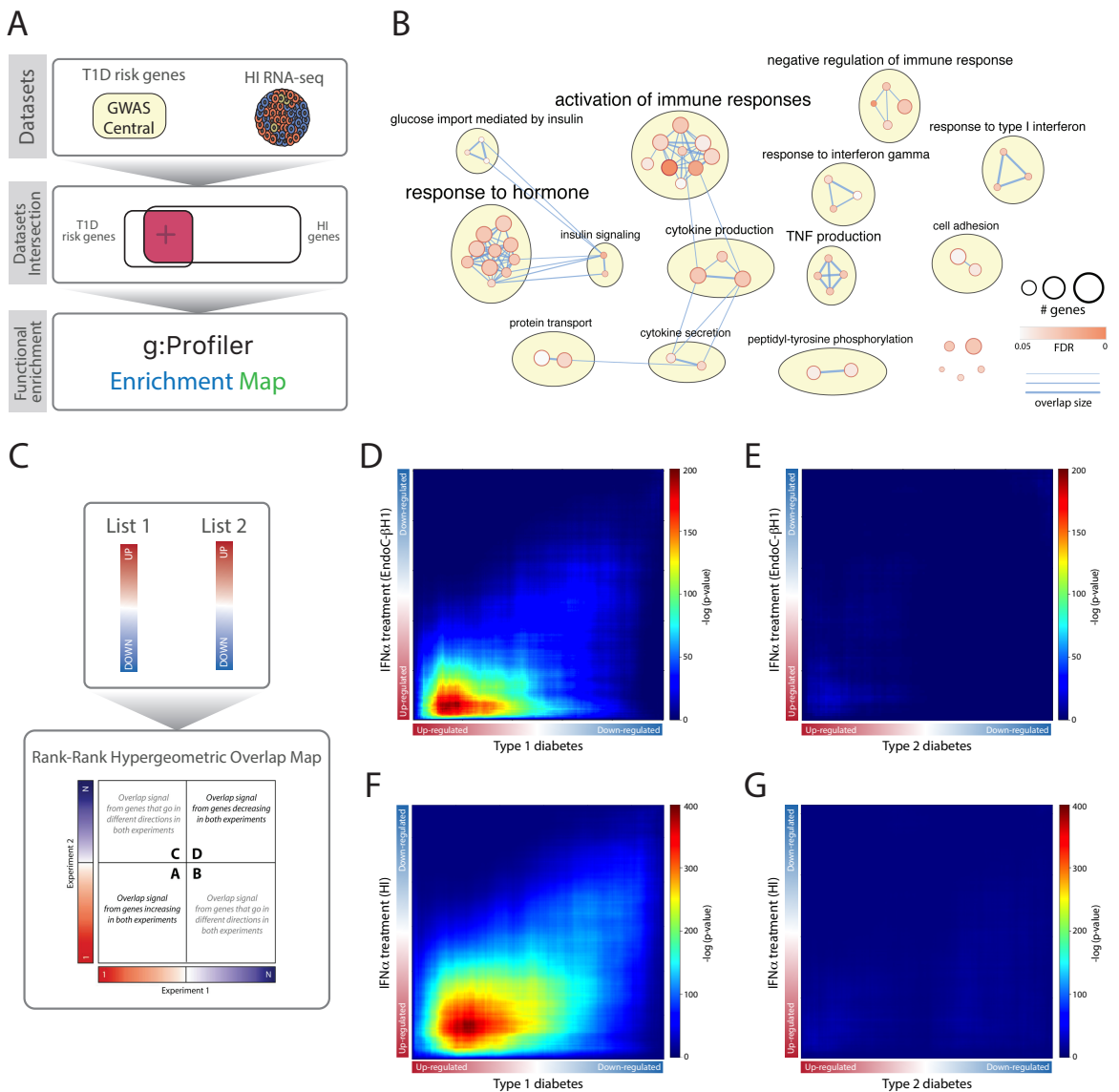
Colli et al.

Supplementary Figures



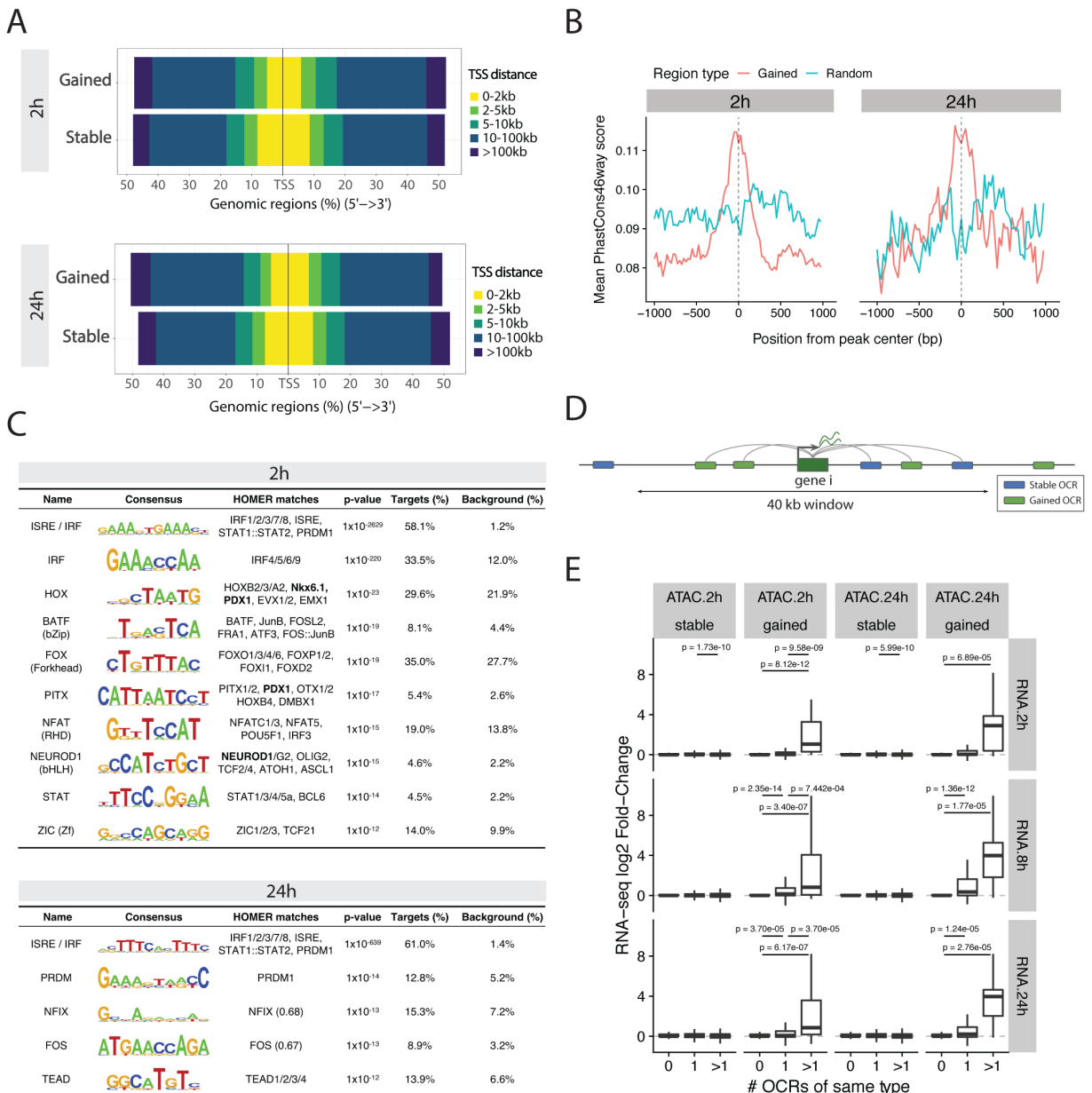
Supplementary Figure 1. Characterization of the global transcriptional responses of human islets to interferon- α . **A.** Schematic representation of the present experimental design. In brief, pancreatic human islets were exposed or not to IFN α (2000 U/ml) for the indicated time points and RNA-sequencing was performed (n = 6). **B.** Correlation between RNA-seq of EndoC- β H1 cells (n = 5) and RNA-seq of human islets (n = 6) exposed to IFN α . The x axis represents the log₂FC of DEGs from human islets RNA-seq and the y axis the DEG from EndoC- β H1 cells RNA-seq at different time points. The up-regulated (log₂FC > 0.58, FDR < 0.05) and down-regulated (log₂FC < -0.58, FDR < 0.05) mRNAs in human islets are filled in red and blue, respectively. The up-regulated (log₂FC > 0.58, FDR < 0.15) or down-regulated (log₂FC < -0.58, FDR < 0.15) mRNAs in EndoC- β H1 cells are represented by red and blue borders, respectively. **C.** The DREM¹ regulatory paths summarizing the temporal patterns of the differentially expressed genes (DEG) detected by RNA-seq (|log₂FC| > 0.58 and FDR < 0.05, n = 6). The x axis represents the time and the y axis the mRNA log₂FC. Each path corresponds to a set of co-expressed genes. Split nodes (circles) represent a temporal event where a group of genes co-expressed up to that point diverge in expression,

most probably due to regulatory events. These annotations are placed on the path immediately after the split to indicate whether the transcription factor (TF) controls the upper or lower paths following the split. In blue are the TFs upregulated at the respective time points of the RNA-seq. **D.** Transcription factors and biological processes (GO) enriched in the main pathways defined by the DREM model (FDR < 0.05).



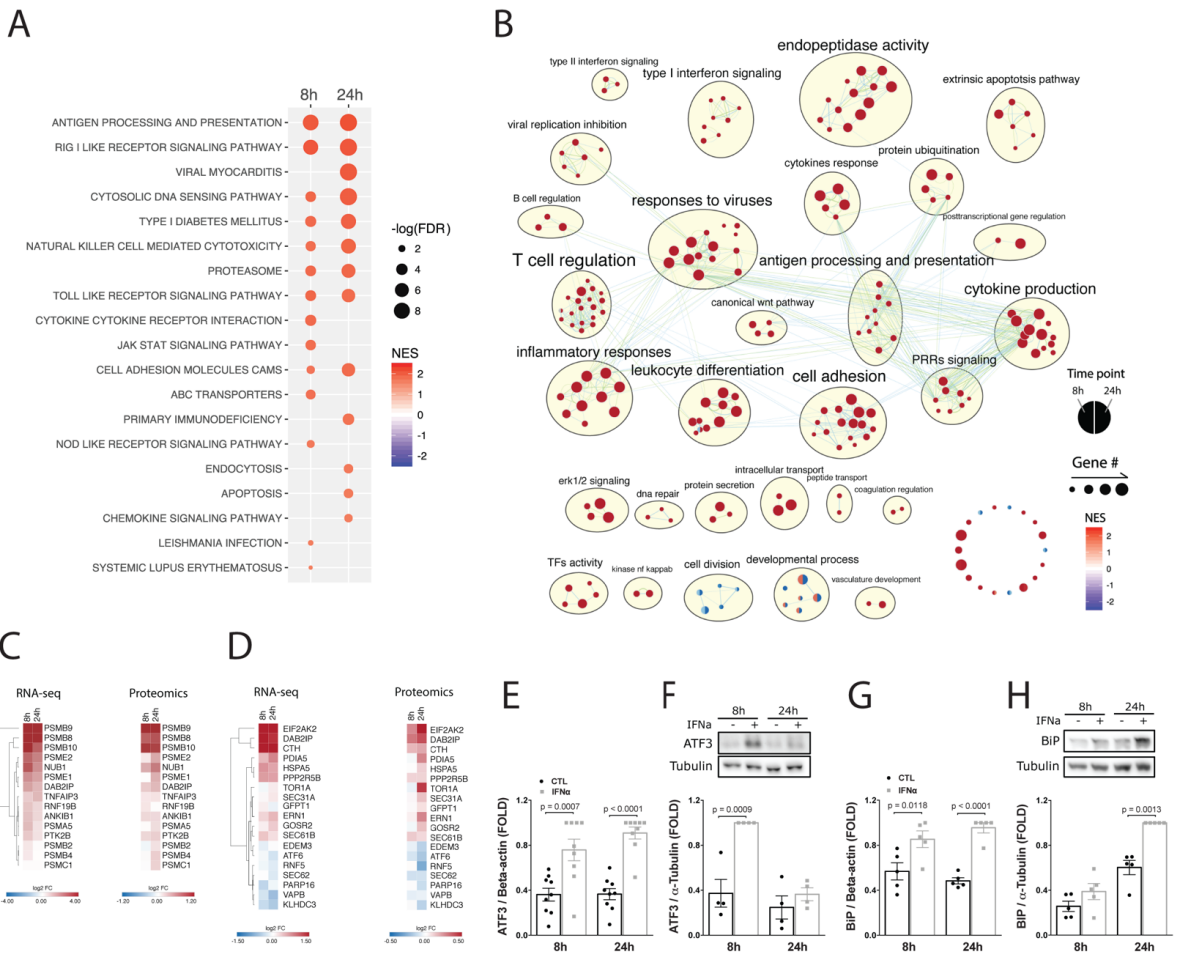
Supplementary Figure 2. Type I interferon-induced genes are enriched among T1D candidate genes and the *in vitro* modelling of this cytokine partially recapitulates the beta cell gene expression signature from T1D individuals. . A. Summary of the approach used to evaluate the biological processes regulated by T1D candidate genes expressed in human islets. Initially, two lists were generated, one containing the T1D risk genes (see Methods) and another the genes expressed in pancreatic human islets (RPKM > 0.5, (GSE108413)²). Next, these two lists were intersected and functional enrichment analysis was performed using the overlapping genes. **B.** Enrichment map of the biological processes (GO) significantly overrepresented in T1D candidate genes expressed in pancreatic human islets performed as described in Methods. **C.** Schematic illustration of the steps followed for the Rank-Rank Hypergeometric Overlap (RRHO) analysis. In brief, ranked-lists of genes based on the log₂FC (T1D or T2D

vs non-diabetic individuals) were generated (List 1) and compared to similarly ranked lists of EndoC- β H1 cells (**D and E**) and human islets (**F and G**) exposed to IFN α for 24 or 18h, respectively, vs untreated cells (List2). Next, the RRHO map was used to screen the sample permutation maps for the higher overlap in genes among the lists. **D, E, F and G**. RRHO map comparing the gene expression profile of EndoC- β H1 cells (**D**) or human islets (HI) (**F**) treated with IFN α to the one present in primary beta cells of individuals affected by T1D³ identified by RNA-seq. This IFN α -induced profile of EndoC- β H1 cells (**E**) and human islets (HI) (**G**) was also compared against the mRNA expression profile of beta cells isolated from individuals affected by T2D⁴, to exclude the impact of metabolically-induced changes.

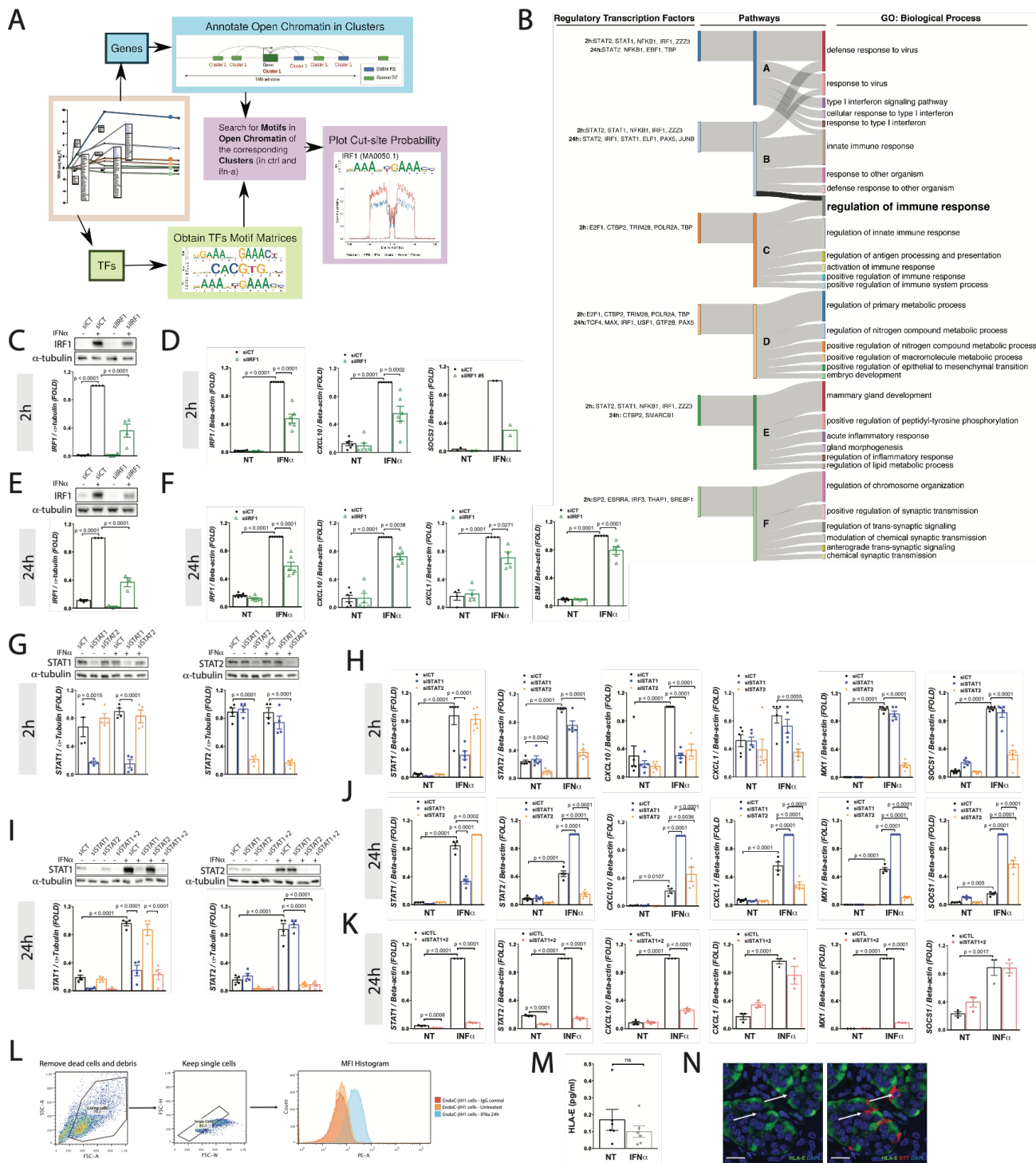


Supplementary Figure 3. Gained open chromatin regions are mainly localized distally to gene transcription starting sites (TSSs), evolutionary conserved and enriched in transcription factors (TFs) binding motifs. A. Distribution of the open chromatin regions (OCRs) distance to the closest gene TSS in different classes and time points. **B.** Mean sequence conservation score of gained open chromatin regions (2 and 24h) and randomized sets in placental mammals. Peaks are extended from the center to 1kb in each direction and mean score was calculated in 50bp windows. **C.** TFs binding motifs enriched in regions of IFN α -induced gained open chromatin. **D.** Scheme representing the strategy to analyze OCRs and their potential target genes. (n = 4, unadjusted p-values were obtained using the hypergeometric test from the HOMER package⁵). **E.** mRNA log₂ fold-changes detected by RNA-seq are related to the type

and the number of REs. Boxplot limits show upper and lower quartiles, centered on the median value. Whiskers extend to 1.5 times the interquartile range ($n = 5$, p-values were adjusted using the Benjamini-Hochberg method, two-sided Mann-Whitney test).

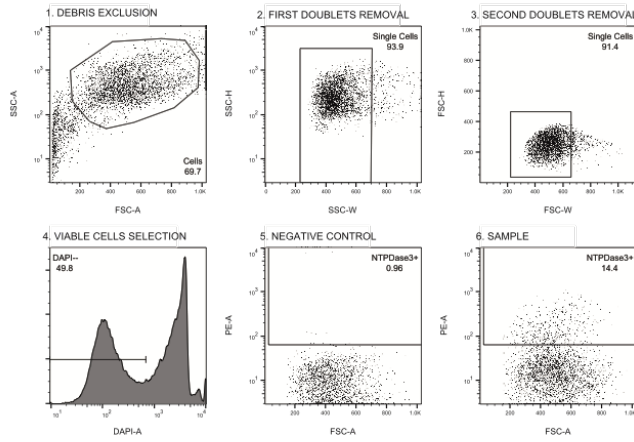
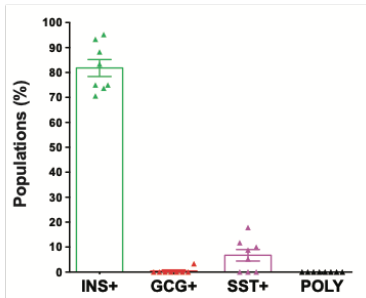
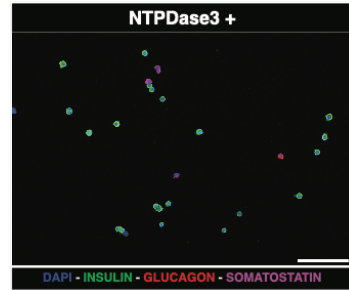
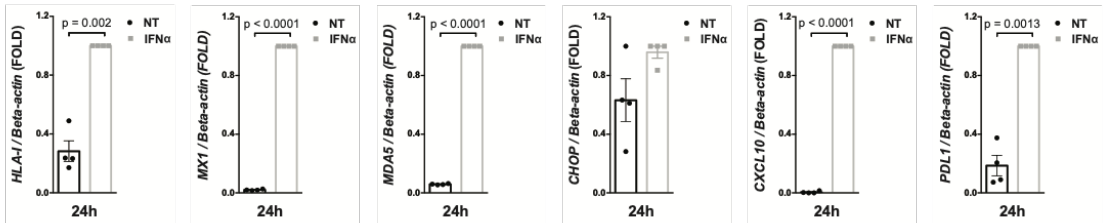


Supplementary Figure 4. IFN α decreases expression of proteins via proteasome activation and triggering of ER stress in EndoC- β H1 cells . A, B. Gene set enrichment analysis (GSEA) of KEGG pathways (A) and enrichment map of biological processes (Gene Ontology) (B) significantly enriched among proteins modified by IFN α in EndoC- β H1 cells. C, D. mRNAs and proteins involved in proteasomal degradation and triggering ER stress are significantly modified by IFN α in EndoC- β H1 cells (C: (n = 5), D: (n = 4), genes/proteins were selected if they present a FDR < 0.05 in at least one time point/condition). E to H. Validation of key mRNAs and proteins involved in protein folding and degradation. EndoC- β H1 cells were treated or not with IFN α for the indicated time points in independent experiments and expression of the stress-responsive transcription factor *ATF3* (E, F) and the chaperone BiP (*HSPA5*) (G, H) were evaluated at the mRNA (E, G) and protein levels (F, H). (for E (n = 9), F (n = 4), G and H (n = 5), mean \pm SEM, ANOVA with Bonferroni correction for multiple comparisons). Source data are provided as a Source Data file.

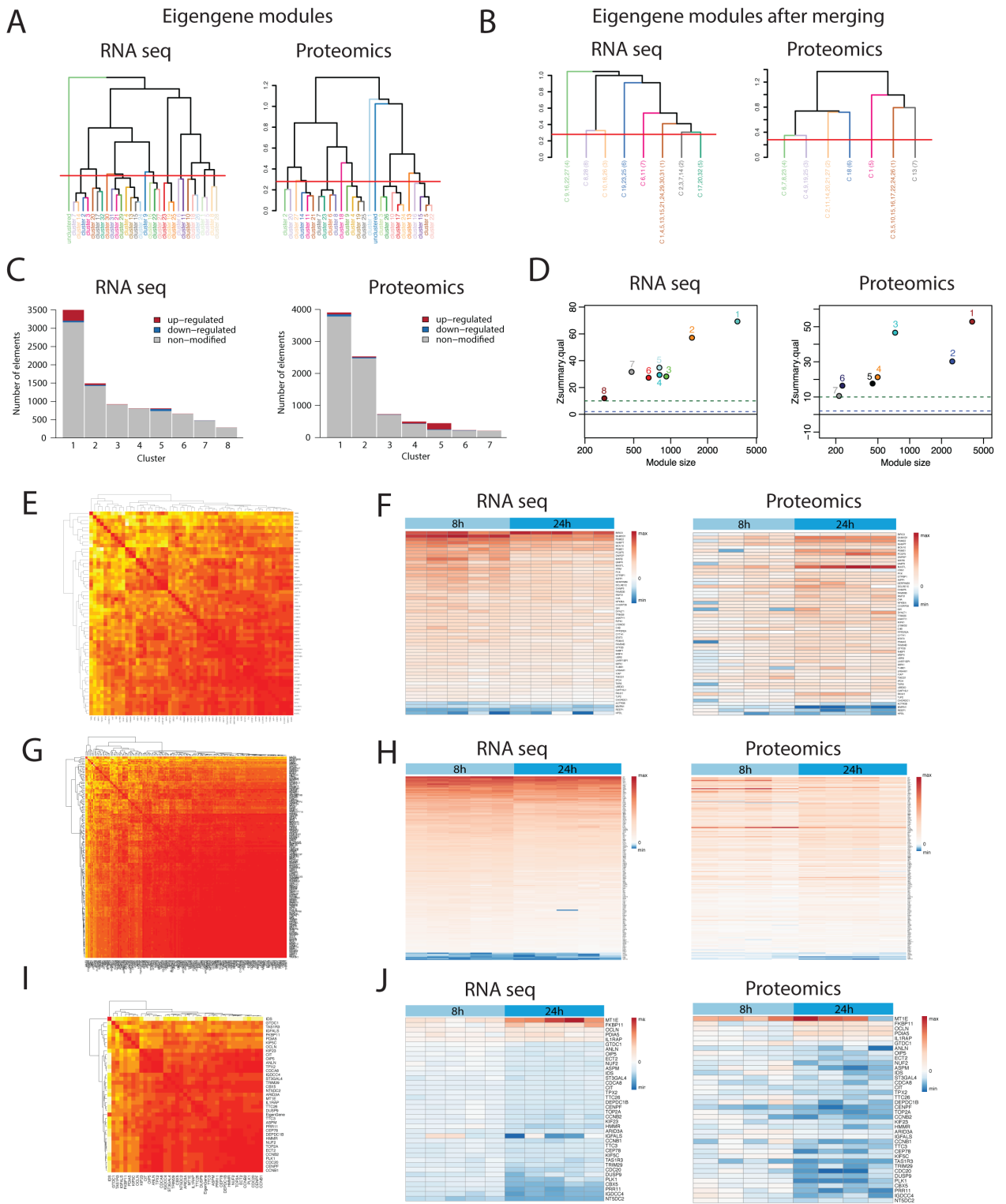


Supplementary Figure 5. Identification of key transcription factors regulating EndoC- β H1 cells responses to IFN α . **A.** Flowchart of the experimental approach: First, the DREM model divided the differentially expressed genes in pathways based on their expression pattern and pointed out TFs potentially controlling paths diverging over time. Second, the coordinates of the genes classified in each

of the pathways were obtained. Third, OCRs at 2 and 24h were associated to the nearest gene classified in one of the paths at 2 and 24h, respectively. Forth, motif matrices of the putative TFs regulating a specific path were identified. Fifth, matches for these matrices were searched for in the respective subset of OCRs annotated to the cluster regulated by the respective TFs. Sixth, the difference in cutting probability between control and IFN α samples was plotted. **B.** TFs and biological processes (GO) enriched in the main pathways defined by the DREM model (FDR < 0.05). **C to K.** EndoC- β H1 cells were transfected with a control non-specific siRNA (siCT) or previously validated siRNAs^{6,7} against *IRF1* (siIRF1) (**C to F**), *STAT1* (siSTAT1) (**G to J**), *STAT2* (siSTAT2) (**G to J**) or *STAT1* plus *STAT2* (siSTAT1+2) (**I and K**). After 48h of recovery, the knockdown efficiency was confirmed at the protein level by Western blot and quantified by densitometry (**C, E, G, I**). The values were normalized by the housekeeping protein α -tubulin and then by the highest value of each experiment considered as 1. The Western blot images are representative of three (**E**) and four (**C, G, I**) independent experiments (ANOVA with Bonferroni correction for multiple comparisons). The mRNA induction of key IFN α -target genes (**D, F, H, J and K**) was determined by real-time PCR at the indicated time points. The values were normalized by the housekeeping gene *β -actin* and then by the highest value of each experiment considered as 1 (for **D**: *IRF1* and *CXCL10* (n = 6), *SOCS1* (n = 2); for **F**: *IRF1* and *CXCL10* (n = 6), *CXCL1* (n = 4); *B2M* (n = 5); for **H**: all (n = 5); for **J**: all (n = 4); for **K**: all (n = 3), ANOVA with Bonferroni correction for multiple comparisons). **L.** Gating strategy used to evaluate HLA-E expression on EndoC- β H1 cells exposed or not to IFN α for 24h (see Supplementary Methods, the same gating strategy was used for MHC Class I analysis). **M.** HLA-E protein in the supernatant of EndoC- β H1 cells exposed or not to IFN α for 24h was measured by ELISA. (n = 6, ns: non-significant, two-sided paired *t* test) **N.** Higher magnification image of an islet from a T1D donor demonstrating that HLA-E (green) does not localize to delta cells (somatostatin; red – white arrows). DAPI – dark blue. Scale bar 20 μ m. Source data are provided as a Source Data file.

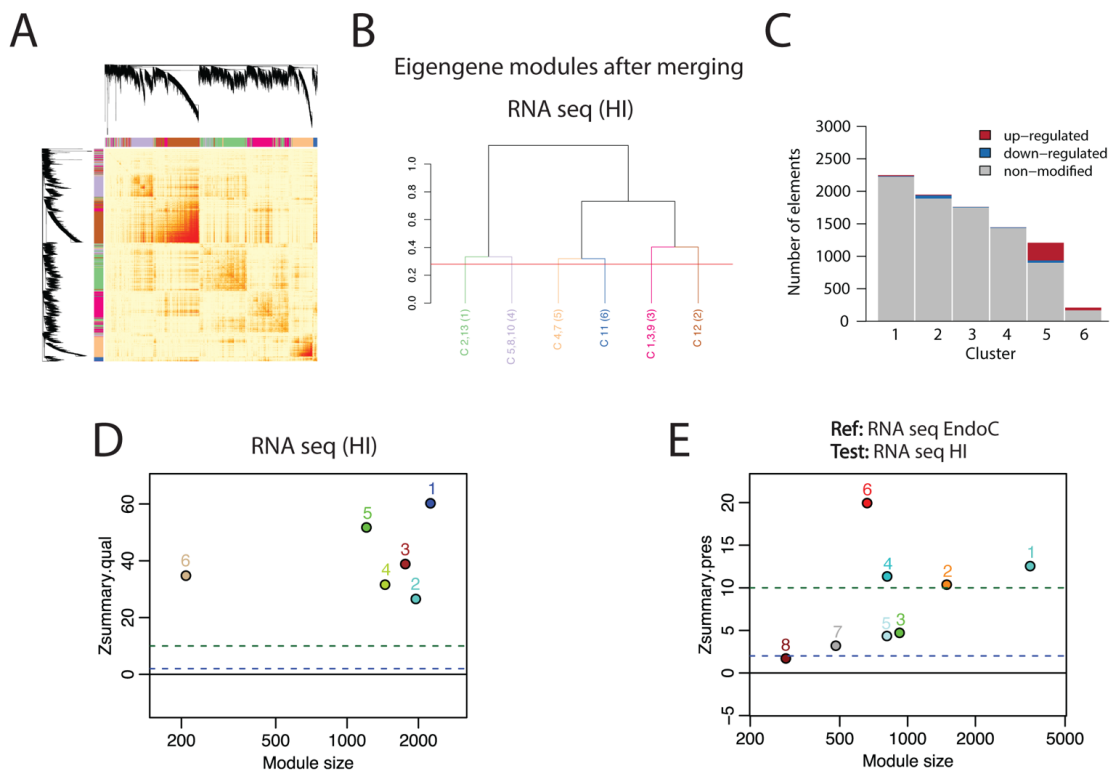
A**B****C****D**

Supplementary Figure 6. Evaluation of interferon- α -induced genes on FACS-purified primary human beta cells. Pancreatic human islets were FACS-sorted using a beta cell specific cell surface marker (*NTPDase3*) (see Methods). **A.** Gating strategy used to obtain purified human beta cells. **B.** The islet cell population frequency in the beta cell enriched fraction was determined by immunocytochemistry (ICC) with antibodies against insulin (INS), glucagon (GCG) or somatostatin (SST) (8 fields + were analysed from 2 independent experiments). **C.** Representative ICC image of the beta cell enriched fraction after FACS-sorting of pancreatic human islets. Image representative of 2 independent experiments. Scale bar 100 μ m. **D.** mRNA expression of *HLA class I* (ABC), *MX1*, *MDA5*, *CHOP*, *CXCL10* and *PDL1* was evaluated by real-time RT-PCR on FACS-purified human beta cells exposed or not to IFN α for 24h. The values were normalized by the housekeeping gene β -actin and then by the highest value of each experiment considered as 1 (n = 4, two-sided paired *t* test). Source data are provided as a Source Data file.

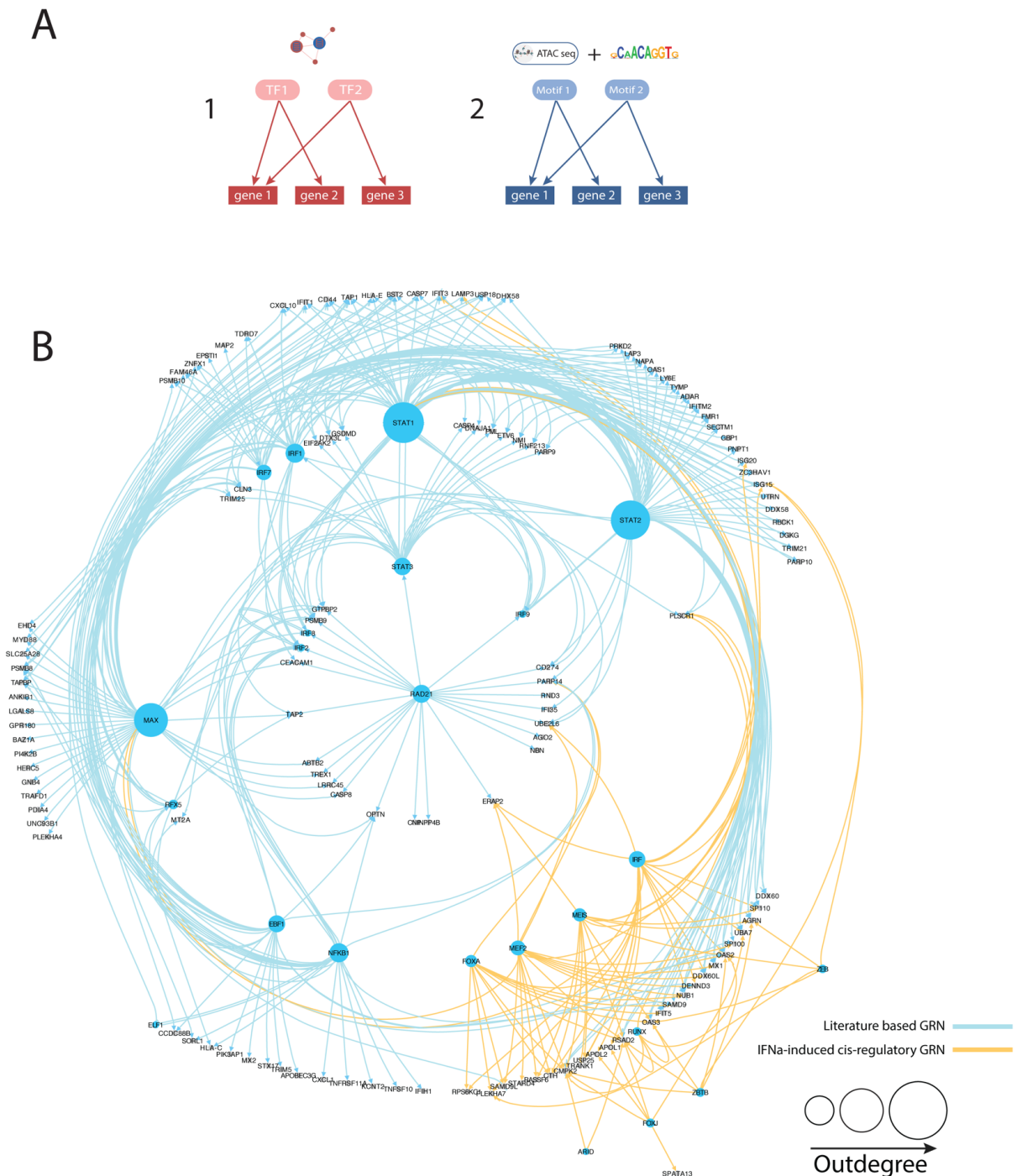


Supplementary Figure 7. Description of interferon- α -regulated modules of genes and proteins. A. Dendrogram of consensus module eigengenes identified by WGCNA mRNA and protein clustering. **B.**

Modules obtained after merging eigengenes groups whose expression profiles were similar. A threshold of eigengene correlation of 0.75 (corresponding to the dissimilarity threshold 0.25) was used for the merging. **C.** Number and status of the mRNAs and proteins in each module from RNA-seq and proteomics of EndoC- β H1 cells. **D.** Module quality is represented as the Zsummary for each module from the resulting networks from RNA-seq and proteomics of EndoC- β H1 cells. For this analysis, the original dataset was re-sampled 1000-times to create reference and test sets and then module quality was evaluated using density and separability metrics and represented as the Zsummary for each module. Zsummary > 2 indicates moderate and Z > 10 high quality/robustness for each module⁸. **E, G, I.** Heatmap plots depicting the relation between genes and proteins in each module. Progressively darker colors correspond to higher correlations. **F, H, J.** Log₂ fold-changes of mRNAs and proteins present in modules #1 (**F**), #2 (**H**) and #5 (**J**), which were generated as described before in Figure 4B and C. The present results show RNA-seq (n = 5) and proteomics (n = 4) data of EndoC- β H1 cells.



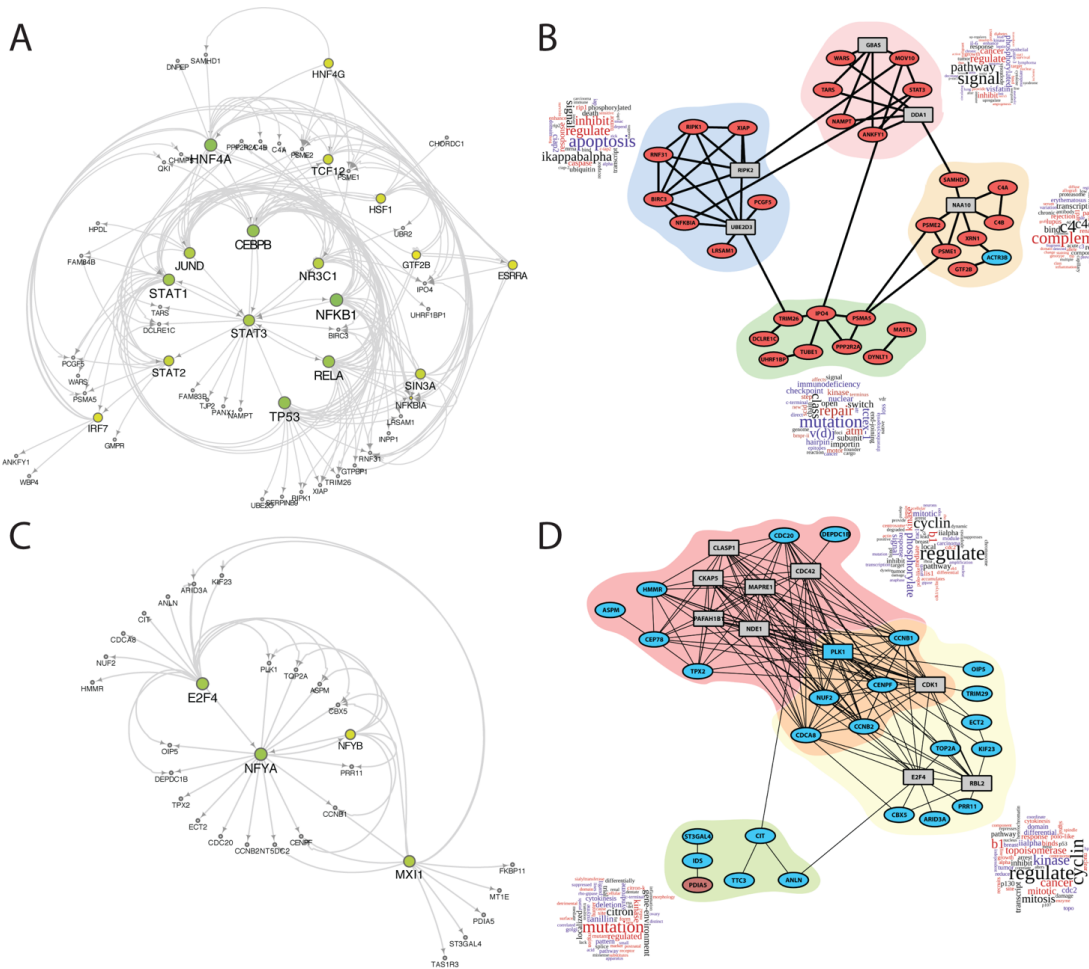
Supplementary Figure 8. Weighted correlation network analysis (WGCNA) of human pancreatic islets RNA-seq. **A.** Heatmap representation of the topological overlap matrix. Rows and columns correspond to single genes, light colors represent low topological overlap, and progressively darker colors represent higher topological overlap. The corresponding gene dendrograms and initial module assignment are also displayed. **B.** Modules obtained after merging eigengenes groups whose expression profiles were similar. A threshold of eigengene correlation of 0.75 (corresponding to the dissimilarity threshold 0.25) was used for the merging. **C.** Number and status of the mRNAs in each module based on the RNA-seq data from pancreatic human islets. **D.** Module quality is represented as the Zsummary for each module from the resulting networks. For this analysis, the original dataset was re-sampled 1000-times to create reference and test sets and then module quality was evaluated using density and separability metrics, and represented as the Zsummary for each module. Zsummary > 2 indicates moderate and Z > 10 high quality/robustness for each module⁸. **E.** Similarity between EndoC- β H1 cells mRNAs modules (reference set) and the human islets RNA-seq modules (test set) was evaluated as described in (C), but using density and intramodular connectivity metrics. Module preservation is represented as the Zsummary for each module. Zsummary > 2 indicates moderate and Z > 10 high preservation for each module⁸. The present results show RNA-seq data (n = 6) of pancreatic human islets.



Supplementary Figure 9. Reconstructing the gene regulatory networks induced by interferon- α in EndoC- β H1 cells. **A.** A gene regulatory network of module #2 was built combining data from two sources: 1) Transcription factor (TF)-target information obtained from publicly available data-repository of regulations (www.regnetworkweb.org). Enriched regulators (FDR < 0.01 and minimum number of

connections = 4) were identified and added to the network if they were not already present. 2) Previously recognized *de novo* HOMER motifs were linked to their target genes using *annotatePeaks.pl*.

B. Gene regulatory network from module #2. Light blue edges represent the literature-based network. Orange edges indicate the cis-regulatory network presently identified. The size of the nodes is proportional to the outdegree of the regulatory components. The present results were based on RNA-seq (n = 5) and proteomics (n = 4) data of EndoC- β H1 cells.



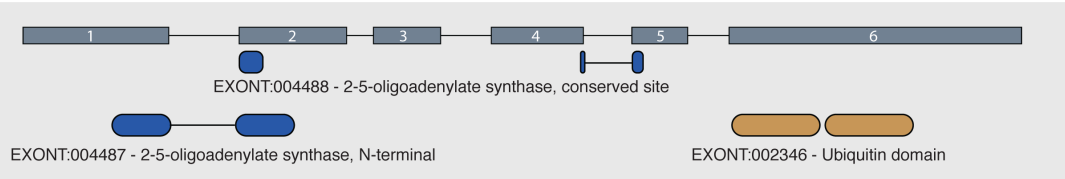
Supplementary Figure 10. Gene regulatory networks and protein-protein interaction (PPI) networks from modules #1 and #5. **A, C.** Gene regulatory networks of modules #1 (**A**) and #5 (**C**) were built using TF-target information obtained from publicly available data-repository of regulations (www.regnetworkweb.org). Enriched regulators (FDR < 0.05 and minimum number of connections = 5) were identified and added to the network if they were not already present. The size of the TFs nodes is proportional to the outdegree. **B, D.** The PPI network of modules #1 (**B**) and #5 (**D**) was done using the InWeb InBioMap database⁹. Enriched proteins (FDR < 0.05 and minimum number of connections = 5, represented as squares) were identified and added to the network if they were not already present. Red fill identifies upregulated proteins, blue fill indicates downregulated proteins and gray fill equal-regulated. Colored regions delimitate communities of proteins identified using the EAGLE algorithm, as described in Material and Methods. The wordcloud next to each community presents their enriched geneRIFs terms. The present results were based on RNA-seq (n = 5) and proteomics (n = 4) data of Endo-βH1 cells.

A

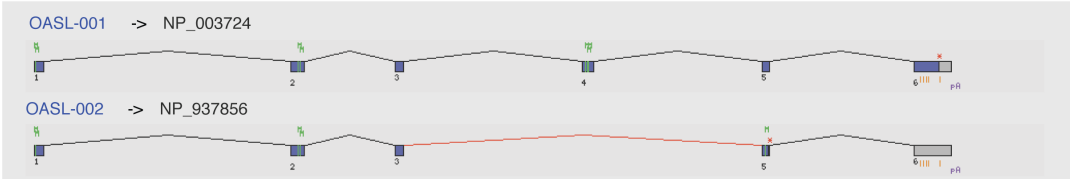
Gene structure (exons)



Protein features

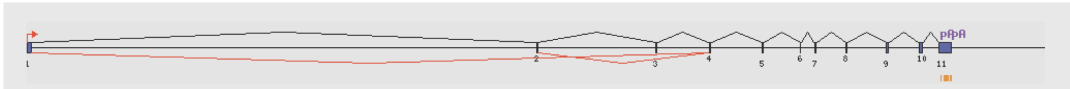


Transcripts

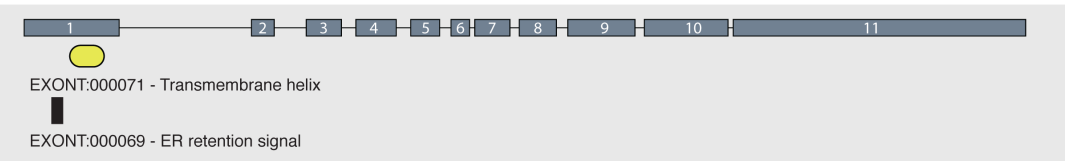


B

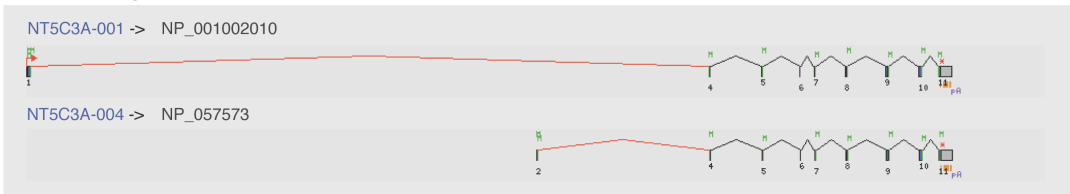
Gene structure (exons)



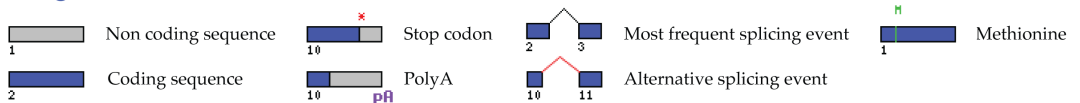
Protein features



Transcripts

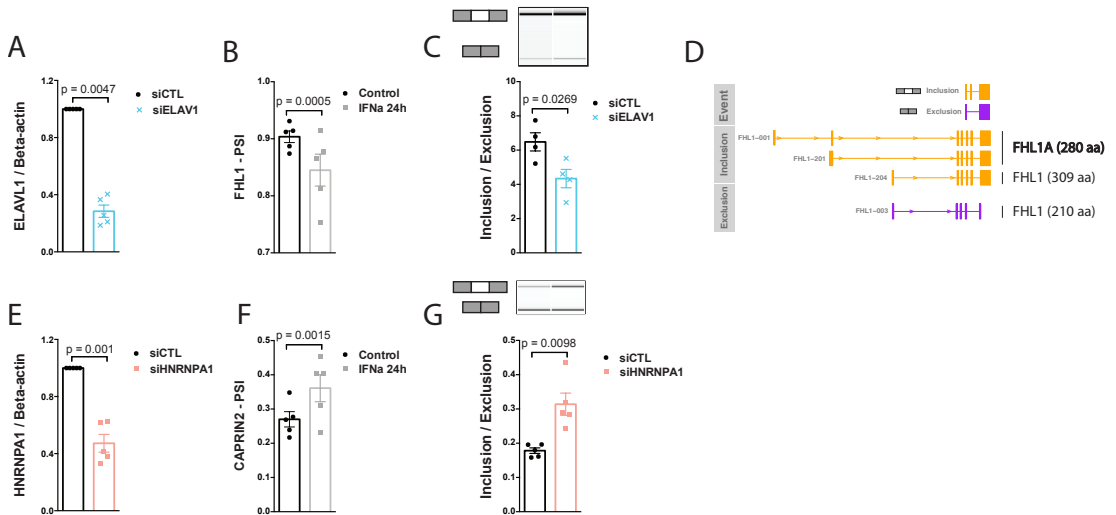


Legends

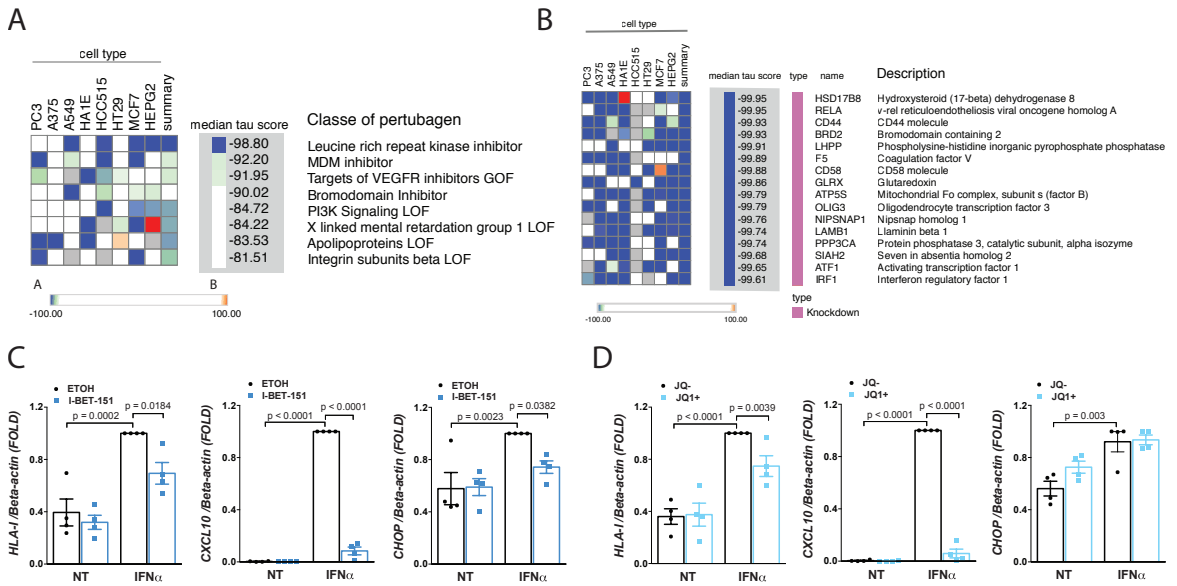


Supplementary Figure 11. Interferon- α leads to alternative exon usage with functional impact. Graphical representation of *OASL* (A) and *NT5C3A* (B) gene structure with all its exons and the most

frequent alternative exon skipping/retention events (red). In the middle it is shown the protein features associated to their respective exons and at the bottom the transcripts regulated by IFN α that are affected by the usage of alternative exons.



Supplementary Figure 12. Inhibition of ELAVL1 and HNRNPA1 reproduces IFN α -induced changes in alternative splicing. A and E. ELAVL1 (A) and HNRNPA1 (E) were silenced in EndoC- β H1 cells using specific siRNAs and their mRNA expression were evaluated by real-time RT-PCR. The values were normalized by the housekeeping gene β -actin and then by the highest value of each experiment considered as 1 (for A and E (n = 5), mean \pm SEM, two-sided paired *t* test). **B and F.** *FHL1* and *CAPRN2* percent splice-in (PSI) values in RNA-seq of EndoC- β H1 cells exposed to IFN α for 24h. PSI represents the ratio of normalized read counts, indicating inclusion of the transcript element over the total normalized reads for that event (inclusion and exclusion reads) (for B and F (n = 5), mean \pm SEM, adjusted p-values (FDR). **C and G.** Confirmation of exon 6 (*FHL1*) (C) and exon 5 (*CAPRN2*) (G) inclusion in EndoC- β H1 cells after silencing of *ELAVL1* and *HNRNPA1*, respectively. cDNA was amplified by RT-PCR using primers located in the up-stream and down-stream exons of the splicing event. The PCR products were analysed by automated electrophoresis using a Bioanalyzer 2100 machine and quantified by comparison with a loading control. (for C (n = 4) and G (n = 5), mean \pm SEM, two-sided paired *t* test). **D.** Schematic representation of the transcripts expressed in EndoC- β H1 cells with differential usage of exon 6 on *FHL1* and the proteins encoded by these transcripts (aa: amino acids). Source data are provided as a Source Data file.



Supplementary Figure 13. Bromodomain inhibitors decrease the type I interferon-induced signature in pancreatic human islets. **A.** Classes of drugs that have an opposite signature to the one shared among beta cells of individuals with T1D and human pancreatic islets exposed to IFN α for 18h (see Supplementary Fig. 2D). **B.** Knockdown of the gene *BRD2* (bromodomain containing 2) generates an inversely related signature in comparison to IFN α in the Connectivity MAP database¹⁰. **C, D.** Dispersed human islets were pretreated for 2h with the bromodomain inhibitors I-BET-151 (1 μ M) (**C**) or JQ1+ (0.4 μ M) (**D**) and then exposed to IFN α for 24h. After this period the cells were collected and the mRNA expression of *HLA class I* (*ABC*), the chemokine *CXCL10* and the ER stress marker *CHOP* were evaluated by real-time RT-PCR. The vehicle Ethanol and an inactive enantiomer (JQ1-) were used as respective controls for I-BET-151 and JQ1+. (for **C** and **D** (n = 4) mean \pm SEM, ANOVA with Bonferroni correction for multiple comparison). Source data are provided as a Source Data file.

Supplementary Tables

Supplementary Table 1. List of *FMR1* targets for which mRNA binding and protein regulation have been reported in more than one study¹¹.

Gene name	Gene symbol
Apoptosis associated tyrosine kinase	<i>AATK</i>
Aldolase, fructose-bisphosphate A	<i>ALDOA</i>
Adaptor related protein complex 2 subunit beta 1	<i>AP2B1</i>
APC, WNT signaling pathway regulator	<i>APC</i>
Amyloid beta precursor protein	<i>APP</i>
Activity regulated cytoskeleton associated protein	<i>ARC</i>
Rho guanine nucleotide exchange factor 12	<i>ARHGEF12</i>
Calcium/calmodulin dependent protein kinase II alpha	<i>CAMK2A</i>
Catenin beta 1	<i>CTNNB1</i>
Dystroglycan 1	<i>DAG1</i>
DSCS large MAGUK scaffold protein 4	<i>DLG4</i>
DLG associated protein 4	<i>DLGAP4</i>
Eukaryotic translation elongation factor 2	<i>EEF2</i>
Fragile X mental retardation 1	<i>FMR1</i>
FUS RNA binding protein	<i>FUS</i>
Gamma-aminobutyric acid type A receptor beta1 subunit	<i>GABRB1</i>
Gamma-aminobutyric acid type A receptor delta subunit	<i>GABRD</i>
Heterogeneous nuclear ribonucleoprotein A2/B1	<i>HNRNPA2B1</i>
Potassium voltage-gated channel subfamily C member 1	<i>KCNC1</i>
Potassium voltage-gated channel subfamily D member 2	<i>KCND2</i>
Microtubule associated protein 1B	<i>MAP1B</i>
Microtubule associated protein 2	<i>MAP2</i>
Myelin basic protein	<i>MBP</i>
Matrix metalloproteinase 9	<i>MMP9</i>
Neurologin 2	<i>NLGN2</i>
Nuclear receptor subfamily 3 group C member 1	<i>NR3C1</i>
Oligophrenin 1	<i>OPHN1</i>
Protocadherin 10	<i>PCDH10</i>
Piccolo presynaptic cytomatrix protein	<i>PCLO</i>
Phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit beta	<i>PIK3CB</i>
Plakophilin 4	<i>PKP4</i>
Proteolipid protein 1	<i>PLP1</i>
Protein phosphatase 2 catalytic subunit alpha	<i>PPP2CA</i>
Protein tyrosine phosphatase, non-receptor type 5	<i>PTPN5</i>
Rac family small GTPase 1	<i>RAC1</i>
Regulator of G protein signaling 5	<i>RGSS5</i>
Ras homolog family member A	<i>RHOA</i>
Superoxide dismutase 1	<i>SOD1</i>
Spen family transcriptional repressor	<i>SPEN</i>
Voltage dependent anion channel 1	<i>VDAC1</i>

Supplementary Table 3. List of siRNAs and primers used in the present study.

siRNAs		
	Distributor	Sequence
siCT	Qiagen, Venlo, Netherlands	Allstars Negative Control siRNA, sequence not provided
Human siIRF1	Qiagen, Venlo, Netherlands	5'- CAAGCATGGCTGGGACATCAA -3'
Human siSTAT1	Invitrogen, Pasley, UK	5'-GGAUUGAAAGCAUCCUAGAACUCAU-3'
Human siSTAT2	Invitrogen, Pasley, UK	5'-CAGCAGCAUGUCUUCUGCUUCCGAU-3'
Human siELAVL1	Ambion, Austin, USA	5'-UGUCAAACCCGUAUAAACGC-3'
Human siHNRNPA1	Ambion, Austin, USA	5'-AUCACUUUUUAACCAUUC-3'
Primers		
	Forward	Reverse
	Sequence (5'-3')	Sequence (5'-3')
Human ATF3	GCTGTCACCACGTGCAGTAT	TTTGTGTTAACGCTGGGAGA
Human β -actin	CTGTACGCCAACACAGTGCT	GCTCAGGAGGAGCAATGATC
Human B2M	TGCTGCTCCATGTTGATGTA	GACCAAGATGTTGATGTTGGATAAG
Human CHOP	Qiagen QuantiTect primer, cat# QT00082278	
Human CXCL1	AGAACATCCAAAGTGTGAAC	TTTCTTAACTATGGGGATG
Human CXCL10	GTGGCATTCAAGGAGTACCTC	GCCTTCGATTCTGGATTGAG
Human HLA-E	TGGTTGCTGCTGTGATATGGA	GCTCCACTCAGCCTTAGAGT
Human HLA-I (ABC)	GAGAACGGGAAGGAGACGC	CATCTCAGGGTGAGGGGCT
Human HSPA5	Qiagen QuantiTect primer, cat# QT00096404	
Human IRF1	CATTCACACAGGCCGATACA	TGGTCTTTCACCTCCTCGATAT
Human MX1	AGACAGGACCATCGGAATCT	GTAACCCCTTCAGGTGGAAC
Human NT5C3A -004	TGTAGCTCGTCCGGTTACCT	ACCAGCTTACAGTTGTCAATGAT
Human OASL (exon 4)	CCCCTGAGGTCTATGTGAGC	ATCTGTACCCTTCTGCCACG
Human PDL1	CCAGTCACCTCTGAACATGAA	ACTTGATGGTCACTGCTTGT
Human RMI2 -004	CGCCAGACATCTTATGCCCT	TCACAGCAAGGCAGTGTGAA
Human SOCS1	GACGCCTGCGGATTCTAC	GAGGCCATCTTACGCTAA
Human SOCS3	CCTCGCCACCTACTGAAGGC	CCCGGAGTAGATGAATA
Human STAT1	GACCCAATCCAGATGCTATGA	CCCGACTGAGCCTGATTA
Human STAT2	GTTGGCAGTTCTCCTCCTATG	GAAGTCAGCCCAGGACAATAA
Human ELAVL1	TTTGATCGTCAACTACCTCCCTC	CTGTGCTCTGCTACTTTATCCCG
Human HNRNPA1	ATTTGGACTTTCCTACCCACTC	CAGCTAGTTTTCTATTCCTGGCA
Human FHL1	TTTGCCAAGCATTGCGTGAA	GGCAGAGTCGGGACAATACA
Human CAPRN2	TCAAAACTGACCTGCCCTGA	TGGCATTTTTGGAACTGGG
Human MDA5	GAGGAATCAGCACGAGGAATAA	TCAGATGGTGGGCTTTGAC

Supplementary Table 4. List of antibodies and conditions used for the immunofluorescence.

Primary Antibody	Manufacturer and clone	Antigen Retrieval	Antibody Dilution	Incubation time with primary antibody	Secondary Detection System
Step 1: HLA-E	Abcam	10mM Citrate pH6	1:150	Overnight at 4°C	Goat anti-mouse IgG (H+L) HRP then tyramide Alexa 488 (as per manufacturers protocol; Thermofisher Tyramide SuperBoost kit Cat# B40922)
	MEM-E/02				
	Mouse Monoclonal				
	Cat#2216				
Step 2: Glucagon or Somatostatin	Abcam	10mM Citrate pH6	1:4000	1h at RT	Goat anti-rabbit IgG (H+L) Alexa Fluor™-conjugated secondary antibodies (1/400 for 1h)
	EP3070				
	Rabbit				
	Monoclonal				
	Cat# 92517				
	Abcam	10mM Citrate pH6	1:200	1h at RT	Goat anti-rat IgG (H+L) Alexa Fluor™-conjugated secondary antibodies (1/400 for 1h)
	Rat				
	M09204				
Monoclonal					
Cat# 30788					
Step 3: Insulin	Dako	10mM Citrate pH 6	1:700	1h at RT	Goat anti-guinea-pig IgG (H+L) Alexa Fluor™-conjugated secondary antibodies (1/400 for 1h)
	Guinea-Pig Polyclonal				
	Cat#A0546				

Supplementary Table 5. List of antibodies used for Western blot and flow cytometry.

Antibody	Supplier	Identifier	Dilution
HLA-E clone 3D12 (Flow cytometry (FC))	Biologend	Cat#342602; RRID: AB_1659247	1:250 (FC)
HLA-E clone MEM-E/02 (Western blot (WB) and IF)	Abcam	Cat#ab2216; RRID: AB_302895	1:500 (WB) 1:150 (IF)
IRF1	Cell signaling	Cat#8478; RRID: AB_10949108	1:1000 (WB)
phospho-STAT1	Cell signaling	Cat#9167; RRID: AB_561284	1:1000 (WB)
phospho-STAT2	Cell signaling	Cat#88410; RRID:AB_2800123	1:1000 (WB)
total STAT1	Cell signaling	Cat#14495; RRID: AB_2716280	1:1000 (WB)
total STAT2	Cell signaling	Cat#72604; RRID:AB_2799824	1:1000 (WB)
α -tubulin	Sigma	Cat#T9026; RRID:AB_477593	1:5000 (WB)
Peroxidase- conjugated donkey anti-rabbit IgG	Jackson ImmunoResearch	Cat#715-036-152; RRID:AB_2340590	1:1000 (WB)
Peroxidase- conjugated donkey anti-mouse IgG	Jackson ImmunoResearch	Cat#711-036-150; RRID:AB_2340773	1:1000 (WB)
Alexa Fluor 488 goat anti-Guinea-Pig IgG	Life technologies, USA	Cat#A11073; RRID: AB_2534117	1:500 (ICC)
Alexa Fluor 568 rabbit anti-mouse IgG	Life technologies, USA	Cat#A11061; RRID: AB_2534108	1:500 (ICC)
Polyclonal Goat Anti-Mouse Immunoglobulins/RPE Goat	Dako	Cat#R0480; RRID: AB_579538	1:500 (FC)
Mouse anti-human NTPDase3	www.ectonucleotidases-ab.com	Cat#hN3-B35	5 μ g/mL (FC)
ATF3	Santa Cruz	Cat#SC-188	1:1000 (WB)
BIP	Cell signaling	Cat##3177	1:1000 (WB)
MHC class I antibody (W6/32)	Enzo	Cat##ALX-805-711-C100	1:500 (FC) 1:1000 (ICC)
Insulin	DAKO	Cat#A0546	1:1000 (ICC)

Supplementary Methods

Culture of human EndoC- β H1 cells and human islets, and cell treatments

EndoC- β H1 cells were cultured in DMEM containing 5.6 mmol/l glucose (Gibco, Thermo-Fisher Scientific, Paisley, UK), 2% BSA fraction V, fatty acid free (Roche, Mannheim, Germany), 50 μ mol/l 2-mercaptoethanol (Sigma-Aldrich, Poole, UK), 10 mmol/l nicotinamide (Calbiochem, Darmstadt, Germany), 5.5 μ g/ml transferrin (Sigma-Aldrich), 6.7 ng/ml selenite (Sigma-Aldrich), 100 units/ml penicillin and 100 μ g/ml streptomycin (Lonza, Leusden, Netherlands) in Matrigel-fibronectin-coated plates¹².

After isolation, the human islets were cultured in M199 culture medium (5.5 mmol/l of glucose) before being sent to Brussels, Belgium, where they were dispersed and, in some experiments FACS-sorted for mRNA determination. Dispersed cells were cultured in Ham's F-10 medium containing 6.1 mM glucose (Gibco, Thermo-Fisher Scientific), 10% fetal bovine serum (FBS) (Gibco, Thermo-Fisher Scientific), 2 mM GlutaMAX (Sigma-Aldrich), 50 mM 3-isobutyl-1-methylxanthine (Sigma-Aldrich), 1% BSA fraction V, fatty acid free (Roche), 50 U/ml penicillin and 50 mg/ml streptomycin (Lonza)¹². All the results from experiments shown with EndoC- β H1 cells or human islet cells refer to independent biological samples. EndoC- β H1 cells or pancreatic islets were treated with human IFN α 2000 U/ml (equivalent to 11.1 pg/ml) (PeproTech Inc., Rocky Hill, NJ) alone, or in combination with human IL1 β , 50 U/ml (equivalent to 240 pg/ml) (R&D Systems, Abingdon, UK), for the indicated time points, based on previous dose-response experiments⁷. Of note, peripheral blood levels of IFN α in individuals affected by T1D are in the range of 0 - 30 pg/ml and infection of peripheral blood mononuclear cells (PBMCs) from T1D patients with Coxsackievirus B4, a strain previously identified in the pancreas of T1D patients¹³, produces IFN α concentration in the range of 0-65 pg/ml¹⁴. Patients with severe and complicated hepatitis C have serum IL1 β concentrations in the range of 0.7 – 187 pg/ml¹⁵. The different time points were selected to represent early, intermediary and late responses to IFN α . In some conditions the cells were pre-treated for 2h with chemical inhibitors of JAK signaling (Baricitinib), bromodomain and extra-terminal proteins (BET) (JQ1+ and I-BET-151) (Selleckchem, Munich, Germany) or their respective vehicles. These compounds were maintained in the medium during the subsequent exposure to cytokines.

Tissue

Formalin-fixed paraffin embedded pancreatic sections from the Exeter Archival Diabetes Biobank (EADB; <http://foulis.vub.ac.be/>) or from the DiViD biopsy study of living donors with recent-onset T1D¹⁶ were studied. These comprised 13 samples in total; 6 from non-diabetic subjects (age range: 4 - 58 years) and 7 from individuals with T1D (age range: 1.3 - 42 years; duration of disease range 3 days - 18 months; Supplementary Table 2). All samples were studied with full ethical approval and using adequate reporting standards for the study of human tissues (15/W/0258).

ATAC sequencing processing and analysis

Sequenced reads were mapped to the human reference genome version hg19 using bowtie2 (version 2.3.4.1)¹⁷ and unaligned reads were filtered out. Afterwards, duplicate reads were removed using Picard MarkDuplicates (version 2.5.0) (Picard Tools, retrieved November 29, 2018, from <https://broadinstitute.github.io/picard/index.html>) and reads aligned to the mitochondrial DNA or to ENCODE blacklisted regions¹⁸ were filtered out using samtools (version 1.8)¹⁹. Finally, ATAC-seq reads were shifted 4bp in the forward strand and -5bp on the reverse strand.

Bam files from different replicates were merged into a single one using samtools (version 1.8)¹⁹ and peaks were called from that merged file using MACS2 callpeak (version 2.1)²⁰ using the parameters “-q 0.05 --nomodel --shift -100 --extsize 200”.

Differential analysis of ATAC-seq chromatin accessibility was performed using the R package DESeq2²¹. A consensus peak set was generated for each timepoint by merging the called peaks. Then, reads were counted at each peak and the resulting matrix was used as input for DESeq2. Briefly, DESeq2 normalizes samples according to per-sample sequencing depth and accounting for intra-sample variability. Then, it fits data to a negative binomial generalized linear model (GLM) and calculates the Wald statistic. Finally, raw p-values are false discovery rate (FDR)-corrected for multiple testing using the Benjamini-Hochberg method. Peaks with an absolute \log_2 fold change higher than 1 ($|\log_2FC| > 1$) and an FDR < 0.05 were considered differentially accessible after the IFN α treatment.

RNA sequencing processing and analysis

Sequenced reads were mapped to the human genome (version GRCh37/hg19) using the software Tophat 2 (v2.0.13)²² with default parameters. Mapped reads were then annotated based on the Gencode version 18²³ using Flux Capacitor²⁴ (<http://confluence.sammeth.net>) with default parameters. The relative expression of genes and transcripts is represented in RPKM units (“reads per kilobase per million mapped reads”)²⁵.

The genes and transcripts differentially expressed after IFN α exposure were identified by using the R/Bioconductor package EdgeR²⁶. For the present analysis read counts were normalized using the Trimmed Mean of M-values (TMM) method, which calculates a set of normalization factors, one for each sample, and try to eliminate composition biases between libraries. The dispersion estimates was obtained and the negative binomial generalized linear model fitted. Different contrasts were used to allow the evaluation among the time points. The differential expression was determined using the quasi-likelihood (QL) F-test. Genes or transcripts were considered differentially expressed if presented a FDR < 0.05 and a $|\log_2FC| > 0.58$ (FC > 1.5).

Alternative splicing (AS) events were analyzed by rMATS²⁷ using paired settings. This generated percentage splicing index (PSI) values and the false discovery rates for five different splicing events:

cassette exons (CE_x), mutually exclusive exons (MXE), retained introns (RI), and 5' and 3' alternative splice sites (5/3ASS). To select the more biologically relevant AS events a threshold of $|\Delta\text{PSI}| > 20\%$ and $\text{FDR} < 0.05\%$ was applied. The computational motif enrichment analysis of RNA-binding proteins (RBPs) in the vicinity of alternatively spliced cassette exons was performed using rMAPS²⁸ by comparing the spatial occurrence of RBPs motifs among flanking regions of exons whose inclusion/exclusion was modified by IFN α -exposure to the one non-modified. Functional impact of the splicing changes was performed using the Exon Ontology database²⁹.

The quantification of alternative first exon (AFE) usage from RNA-seq data was performed using the software SEASTAR³⁰. In brief, all the putative non-redundant first exons (FEs) were identified by transcript assembly from our RNA-sequencing and then their usage was quantified by counting the reads aligned to the FEs and to their downstream splice junctions. Finally, the PSI values for AFEs detected in each sample were calculated and the rMATS statistical method²⁷ used to determine whether there was significant differential usage induced by IFN α . The standard parameters were used, except: '-c 0.05 -t P'. RNA polymerase II (POLR2A) ChIP-seq of human K562 cells exposed to IFN α for 6h was obtained from the ENCODE project (GSM935474); signal tracks represent the control-normalized tag density (bigWig format) of pooled replicates. The signal is expressed as p-value to reject the null hypothesis that the signal at that location is present in the control.

Sequence conservation analysis

The sequence conservation among stable and gained open chromatin regions detected by ATAC-seq was assessed by determining average phastCons 46 way score in placental mammals³¹. The scores average was calculated in 50bp bins over a 2kb window center on the open chromatin site. Such regions were then randomized along the mappable genome using regioner³² and used as control.

Assignment of open chromatin regions to target genes

To annotate open chromatin as distal or proximal, we assigned each ATAC-seq region to the nearest major TSS of a coding gene (the most upstream annotated TSS) using information from Gencode version 18²³. Regions closer than 2kb to the nearest TSS were annotated as promoters, all the other regions were considered as distal regulatory elements.

We evaluated the correlation between different classes of open chromatin and changes in mRNA expression and protein abundance (Figs. 1C-F, Supplementary Fig. 3E), and assigned ATAC-seq regions to a gene when closer than 20kb of its TSS.

Annotation of CAGE-defined TSSs

The human TSSs peaks detected by genome-wide 5'-RNA sequencing of capped RNAs (Cap Analysis of Gene Expression, CAGE) were obtained from the FANTOM5 database³³ (http://fantom.gsc.riken.jp/5/datafiles/phase2.5/extra/CAGE_peaks/). Canonical TSSs were defined as

the most upstream annotated TSS based on the Gencode version 18. All the other TSSs non-overlapping the principal TSS and Gencode-annotated were defined as alternative TSSs. Only TSSs having peaks with > 10 Tags Per Million (TPM) were considered in the analysis.

The promoter regions were considered as \pm 2kb around the TSS peak, and defined based on the type of TSS (canonical or alternative). The overlap between OCRs and promoters was performed using the function *overlapRegions* from the R package *regioneR*³², with standard parameters.

For visualization, CAGE tag start site (CTSS) human tracks from the FANTOM5 Phase1+2³³ were pooled and filtered with 3 or more tags per library. After that, RLE (Relative Log Expression)-factor normalization was applied and the results are represented using a log scale.

Proteomics processing

Protein extracts were dissolved in 50 mM NH_4HCO_3 containing 8 M urea and 10 mM dithiothreitol and vortexed at 800 rpm for 1h at 37°C. Sulfhydryl groups were alkylated by using 400 mM iodoacetamide (40 mM final concentration), and incubating for 1h in the dark at room temperature. Samples were then diluted 8-fold with 50 mM NH_4HCO_3 , and CaCl_2 was added to a final concentration of 1 mM. The digestion was performed with trypsin at 1:50 enzyme:protein ratio and incubation at 37°C for 5h. The final peptides were extracted using C18 cartridges (Discovery, 50 mg, Sulpelco) and concentrated in a vacuum centrifuge. Peptides quantification was done by BCA, normalized and labeled with tandem mass tags (TMT-10plex, ThermoFisher Scientific) according to the manufacturer's instructions. Labeled peptides were extracted using C18 cartridges and fractionated into 24 fractions using high-pH reversed phase chromatography. Peptide fractions were loaded into a C18 column (70 cm \times 75 μm i.d. packed with Phenomenex Jupiter, 3 μm particle size, 300 Å pore size) connected to Waters NanoAquity UPLC system. A gradient of water (solvent A) and acetonitrile (solvent B), both containing 0.1% formic acid, was used to elute the peptides, which were directly analyzed by nanoelectrospray ionization on a Q-Exactive mass spectrometer (Thermo Fisher Scientific). Scans were collected with a resolution of 35,000 at 400 m/z in a 400-2000 m/z range. High-energy collision induced dissociation (HCD) fragmentation were set for the 12 most intense parent ions using the following parameters: peptide charge \geq 2, 2.0 m/z isolation width, 30% normalized collision energy and 17,500 resolution at 400 m/z. Each parent ion was fragmented only once before being dynamically excluded for 30s.

Collected data were processed using Decon2LS_v2.0³⁴ and DTARefinery³⁵, both with default parameters, to generate peak lists. Peptide identification was done using MS-GF+³⁶ by searching peak lists against islet protein sequences deduced from transcriptomic experiments³⁷ and supplemented with keratin sequences (32,780 total protein sequences). For MS-GF+ searches, a parent ion mass tolerance of 10 ppm, partially tryptic digestion and 2 missed cleavages were allowed. The following modifications were also considered during searches: cysteine carbamidomethylation and N-terminal/lysine TMT

addition as static modifications, and methionine oxidation as a variable modification. Results were filtered in two steps to a final false-discovery rate <1%: spectral-peptide matches - MS-GF probability < 10^{-9} , and protein level < 10^{-10} . The intensity of TMT reporter ions was extracted using MASIC³⁸. Finally, the data was \log_2 converted and normalized by standard median centering. Proteins were quantified using a Bayesian proteoform discovery methodology (BP-Quant) in combination with standard reference-based median quantification³⁹. Proteins were considered significant with a cutoff of $p \leq 0.05$ based on a paired *t*-test.

Dynamic Regulatory Events Miner (DREM) modeling

The DREM divided genes in paths based on their expression profiles and identified bifurcation points, which are moments in time series where previously co-expressed groups of genes start to diverge. These points were then annotated with the TFs potentially regulating the split. The following parameters were used: Minimum Absolute Expression Change = 1.5 (difference from 0); Model Selection Options = Penalized Likelihood, Node Penalty: 40; Expression Scaling Options = Incorporate expression in regulator data for TF; default values were applied for other parameters. The key TFs were selected based on the split scores (score threshold < 0.05).

In order to validate the binding of these key TFs in promotor regions, but also in distant REs, we have integrated the DREM outputs with the changes in chromatin accessibility induced by IFN α . For this purpose, we first associated each open chromatin peak to the nearest TSS of a gene (if closer than 1Mb) and assigned the DREM pathway of the gene to the associated ATAC-seq peak. Next, binding motif matrices from TF appearing in the DREM pathways were selected from JASPAR2016 R package⁴⁰. Footprint was assayed for each relevant TF in ATAC-seq peaks annotated to its DREM pathway by using factorFootprints function from the R package ATACseqQC⁴¹ specifying a minimum score of 95% for finding a motif match. The genes selected for validation with specific siRNAs were identified based on the presence of predicted binding sites for the targeted TF using the i-cisTarget tool⁴².

Functional enrichment evaluation

The functional enrichment results were generated using the Gene Set Enrichment Analysis (GSEA) software v3.0⁴³ when comparing full list of genes (Supplementary Figs. 4A and B). For the functional analysis of WGCNA modules (Fig. 3G) and T1D risk genes (Supplementary Figs. 2A and B) the hypergeometric distribution was used to estimate the significance of enriched pathways and biological processes using g:Profiler⁴⁴. The whole list of genes identified by RNA-seq with a mean RPKM > 0.5 in at least one condition was used as background. The standard parameters were applied in both methods, except for minimum and maximum size of functional category values that were adjusted to 5 and 350, respectively. Enrichment maps of significantly modified biological process (Gene Ontology (GO)) were

generated using the plugin “Enrichment Map” v3.1⁴⁵ and visualized within Cytoscape v3.6, using standard parameters.

T1D risk genes were identified from immunobase (www.immunobase.org, accessed on 11/2018), GWAS catalog (<https://www.ebi.ac.uk/gwas/>, accessed on 11/2018) and⁴⁶. The risk genes were selected based on the following criteria: 1) T1D as the Disease/Trait evaluated by the study, 2) a p-value for the lead SNP < 0.5 x 10⁻⁸, 3) selecting the reported genes linked to the lead SNP by the original study, 4) filtering only the reported genes expressed in human islets (mean RPKM > 0.5 at basal condition or after exposure to IL1 β +IFN γ)^{2,37} for functional enrichment analysis as described above.

Real-time PCR

The real-time PCR quantification was performed using SYBR Green. Gene expression values were normalized by the housekeeping gene β -actin, as its expression is not affected by the present experimental conditions⁷, and then by the highest value of each experiment considered as 1. The sequences of primers used are shown in Supplementary Table 3.

Western blot, immunocytochemistry, ELISA and flow cytometry analysis

For Western blot analysis, the cells were lysed using Laemmli buffer. Total protein was resolved by 10% SDS-PAGE gel and then transferred to nitrocellulose membranes. Immunoblot analysis was performed with specific antibodies as indicated in Supplementary Table 5. Peroxidase-conjugated antibodies were used as secondary antibodies (Supplementary Table 5). The detection of immunoreactive bands was performed using chemiluminescent substrate (SuperSignal West Femto, Thermo Scientific, Chicago, USA) in a Bio-Rad chemi DocTM XRS+ system (Bio-Rad laboratories). The ImageLab software (Bio-Rad laboratories) was used for densitometric quantification of the bands. The values obtained were normalized by the housekeeping protein α -tubulin and then by the highest value of each experiment considered as 1.

Cell supernatants of EndoC- β H1 cells exposed or not to IFN α for 24h were retrieved and human HLA-E protein expression was measured by enzyme-linked immunosorbent assay (Biorbyt, Cambridge, UK).

For the flow cytometry analysis of HLA-E and MHC class I the EndoC- β H1 cells were seeded in 24-well plates (300,000 cells per condition) 72h before being exposed or not to IFN α for 24h. Next, the cells were incubated with primary antibody (mouse monoclonal anti-HLA-E antibody clone 3D12 (1:250) (Biolegend, San Diego, USA) or rabbit anti-MHC class I (W6/32) (Enzo Life Sciences, NY, USA) (1:1000)) for 2h at 4°C, without permeabilization, and subsequently with secondary antibody conjugated with fluorescent dyes for 1h at 4°C (1:500) (Alexa FluorTM anti-mouse or anti-rabbit) before performing flow cytometry analysis (BD LSRFortessaTM X-20, San Jose, CA, USA). Data analysis and graphical representation were performed using FlowJo software version v10 (Tree Star, Ashland, USA). The gating strategy is represented in Supplementary Fig. 5L. In summary, the cells were identified by FSC/SSC

morphological gates (to exclude debris and dead cells). Next, the FSC-W and FSC-H relation was used to eliminate doublets of cells. Finally, histogram distributions of the fluorophore were generated and the mean fluorescence intensity determined. The positive and negative thresholds were set using an isotype Ig control with the same fluorophore. For immunocytochemistry (ICC), 30,000 cells per condition were plated on polylysine-coated coverslips, then treated with IFN α for 24h and fixed with 4% paraformaldehyde. Cells were permeabilized with 0.1 % Triton X100 for 5 min and after incubated with rabbit anti-MHC class I (W6/32) (1:1000) and mouse monoclonal anti-insulin (1:1000). Alexa Fluor-conjugated secondary antibodies were used (Supplementary Table 5). After nuclear staining with Hoechst 33342 (HO), coverslips were mounted with fluorescent mounting medium (Dako, Carpinteria, CA, USA) and immunofluorescence was visualized on a Zeiss microscope (Zeiss-Vision, Munich, Germany). Images were obtained at $\times 20$ or $\times 40$ magnification and analysed using AxioVision software (version 4.7.2; Zeiss-Vision, Munich, Germany).

Small-RNA interference

For transfection, the siRNA and the Lipofectamine RNAiMAX were diluted separately in OptiMEM medium and incubated at room temperature for 5 min. The Lipid-siRNA complexes were then formed at room temperature for 20 min in a proportion of 1.3 μ l Lipofectamine RNAiMAX to 150 nmol/l of siRNA. The complexes were diluted five times in antibiotic-free medium and added to the cells at a final concentration of 30 nmol/l siRNA for overnight transfection.

The non-specific control siRNA (siCT) (Allstar Negative Control siRNA (Qiagen, Netherlands) does not affect gene expression or insulin secretion by human beta cells⁴⁷. The concentration of siRNAs presently used (30 nM) was selected based on dose-response studies⁴⁷.

Cell viability assessment

The percentage of viable, apoptotic and necrotic cells was determined using nuclear dyes (propidium iodide (10 μ g/ml, Sigma) and Hoechst 33342 (10 μ g/ml, Sigma)). This method has been validated for use in pancreatic beta cells by systematic comparison with electron microscopy, caspase 3 activation and DNA laddering⁴⁸⁻⁵⁰. A minimum of 500 cells was counted per condition. Viability was evaluated by two independent observers, one of them being unaware of sample identity. The agreement between the two observers was > 90%.

Supplementary References:





- 1 Ernst, J., Vainas, O., Harbison, C. T., Simon, I. & Bar-Joseph, Z. Reconstructing dynamic regulatory maps. *Mol Syst Biol* **3**, 74, doi:10.1038/msb4100115 (2007).
- 2 Gonzalez-Duque, S. *et al.* Conventional and Neo-antigenic Peptides Presented by beta Cells Are Targeted by Circulating Naive CD8+ T Cells in Type 1 Diabetic and Healthy Donors. *Cell Metab* **28**, 946-960 e946, doi:10.1016/j.cmet.2018.07.007 (2018).
- 3 Russell, M. A. *et al.* HLA Class II Antigen Processing and Presentation Pathway Components Demonstrated by Transcriptome and Protein Analyses of Islet beta-Cells From Donors With Type 1 Diabetes. *Diabetes* **68**, 988-1001, doi:10.2337/db18-0686 (2019).
- 4 Xin, Y. *et al.* RNA Sequencing of Single Human Islet Cells Reveals Type 2 Diabetes Genes. *Cell Metab* **24**, 608-615, doi:10.1016/j.cmet.2016.08.018 (2016).
- 5 Heinz, S. *et al.* Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* **38**, 576-589, doi:10.1016/j.molcel.2010.05.004 (2010).
- 6 Colli, M. L. *et al.* PDL1 is expressed in the islets of people with type 1 diabetes and is up-regulated by interferons-alpha and-gamma via IRF1 induction. *EBioMedicine* **36**, 367-375, doi:10.1016/j.ebiom.2018.09.040 (2018).
- 7 Marroqui, L. *et al.* Interferon-alpha mediates human beta cell HLA class I overexpression, endoplasmic reticulum stress and apoptosis, three hallmarks of early human type 1 diabetes. *Diabetologia* **60**, 656-667, doi:10.1007/s00125-016-4201-3 (2017).
- 8 Langfelder, P., Luo, R., Oldham, M. C. & Horvath, S. Is my network module preserved and reproducible? *PLoS Comput Biol* **7**, e1001057, doi:10.1371/journal.pcbi.1001057 (2011).
- 9 Li, T. *et al.* A scored human protein-protein interaction network to catalyze genomic interpretation. *Nat Methods* **14**, 61-64, doi:10.1038/nmeth.4083 (2017).
- 10 Subramanian, A. *et al.* A Next Generation Connectivity Map: L1000 Platform and the First 1,000,000 Profiles. *Cell* **171**, 1437-1452 e1417, doi:10.1016/j.cell.2017.10.049 (2017).
- 11 Pasciuto, E. & Bagni, C. SnapShot: FMRP mRNA targets and diseases. *Cell* **158**, 1446-1446 e1441, doi:10.1016/j.cell.2014.08.035 (2014).
- 12 Brozzi, F. *et al.* Cytokines induce endoplasmic reticulum stress in human, rat and mouse beta cells via different mechanisms. *Diabetologia* **58**, 2307-2316, doi:10.1007/s00125-015-3669-6 (2015).
- 13 Dotta, F. *et al.* Coxsackie B4 virus infection of beta cells and natural killer cell insulinitis in recent-onset type 1 diabetic patients. *Proc Natl Acad Sci U S A* **104**, 5115-5120, doi:10.1073/pnas.0700442104 (2007).
- 14 Xia, C. Q. *et al.* Increased IFN-alpha-producing plasmacytoid dendritic cells (pDCs) in human Th1-mediated type 1 diabetes: pDCs augment Th1 responses through IFN-alpha production. *J Immunol* **193**, 1024-1034, doi:10.4049/jimmunol.1303230 (2014).
- 15 Antonelli, A. *et al.* Serum levels of proinflammatory cytokines interleukin-1beta, interleukin-6, and tumor necrosis factor alpha in mixed cryoglobulinemia. *Arthritis Rheum* **60**, 3841-3847, doi:10.1002/art.25003 (2009).
- 16 Krogvold, L. *et al.* Pancreatic biopsy by minimal tail resection in live adult patients at the onset of type 1 diabetes: experiences from the DiViD study. *Diabetologia* **57**, 841-843, doi:10.1007/s00125-013-3155-y (2014).
- 17 Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357-359, doi:10.1038/nmeth.1923 (2012).
- 18 Consortium, E. P. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57-74, doi:10.1038/nature11247 (2012).
- 19 Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079, doi:10.1093/bioinformatics/btp352 (2009).

- 20 Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**, R137, doi:10.1186/gb-2008-9-9-r137 (2008).
- 21 Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550, doi:10.1186/s13059-014-0550-8 (2014).
- 22 Kim, D. *et al.* TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol* **14**, R36, doi:10.1186/gb-2013-14-4-r36 (2013).
- 23 Frankish, A. *et al.* GENCODE reference annotation for the human and mouse genomes. *Nucleic Acids Res* **47**, D766-D773, doi:10.1093/nar/gky955 (2019).
- 24 Montgomery, S. B. *et al.* Transcriptome genetics using second generation sequencing in a Caucasian population. *Nature* **464**, 773-777, doi:10.1038/nature08903 (2010).
- 25 Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L. & Wold, B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods* **5**, 621-628, doi:10.1038/nmeth.1226 (2008).
- 26 Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139-140, doi:10.1093/bioinformatics/btp616 (2010).
- 27 Shen, S. *et al.* rMATS: robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. *Proc Natl Acad Sci U S A* **111**, E5593-5601, doi:10.1073/pnas.1419161111 (2014).
- 28 Park, J. W., Jung, S., Rouchka, E. C., Tseng, Y. T. & Xing, Y. rMAPS: RNA map analysis and plotting server for alternative exon regulation. *Nucleic Acids Res* **44**, W333-338, doi:10.1093/nar/gkw410 (2016).
- 29 Tranchevent, L. C. *et al.* Identification of protein features encoded by alternative exons using Exon Ontology. *Genome Res* **27**, 1087-1097, doi:10.1101/gr.212696.116 (2017).
- 30 Qin, Z., Stoilov, P., Zhang, X. & Xing, Y. SEASTAR: systematic evaluation of alternative transcription start sites in RNA. *Nucleic Acids Res* **46**, e45, doi:10.1093/nar/gky053 (2018).
- 31 Siepel, A. *et al.* Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res* **15**, 1034-1050, doi:10.1101/gr.3715005 (2005).
- 32 Gel, B. *et al.* regioner: an R/Bioconductor package for the association analysis of genomic regions based on permutation tests. *Bioinformatics* **32**, 289-291, doi:10.1093/bioinformatics/btv562 (2016).
- 33 Consortium, F. *et al.* A promoter-level mammalian expression atlas. *Nature* **507**, 462-470, doi:10.1038/nature13182 (2014).
- 34 Mayampurath, A. M. *et al.* DeconMSn: a software tool for accurate parent ion monoisotopic mass determination for tandem mass spectra. *Bioinformatics* **24**, 1021-1023, doi:10.1093/bioinformatics/btn063 (2008).
- 35 Petyuk, V. A. *et al.* DtaRefinery, a software tool for elimination of systematic errors from parent ion mass measurements in tandem mass spectra data sets. *Mol Cell Proteomics* **9**, 486-496, doi:10.1074/mcp.M900217-MCP200 (2010).
- 36 Kim, S. & Pevzner, P. A. MS-GF+ makes progress towards a universal database search tool for proteomics. *Nat Commun* **5**, 5277, doi:10.1038/ncomms6277 (2014).
- 37 Eizirik, D. L. *et al.* The human pancreatic islet transcriptome: expression of candidate genes for type 1 diabetes and the impact of pro-inflammatory cytokines. *PLoS Genet* **8**, e1002552, doi:10.1371/journal.pgen.1002552 (2012).
- 38 Monroe, M. E., Shaw, J. L., Daly, D. S., Adkins, J. N. & Smith, R. D. MASIC: a software program for fast quantitation and flexible visualization of chromatographic profiles from detected LC-MS(/MS) features. *Comput Biol Chem* **32**, 215-217, doi:10.1016/j.compbiolchem.2008.02.006 (2008).
- 39 Webb-Robertson, B. J. *et al.* Bayesian proteoform modeling improves protein quantification of global proteomic measurements. *Mol Cell Proteomics* **13**, 3639-3646, doi:10.1074/mcp.M113.030932 (2014).

- 40 Tan, G. JASPAR2016: Data package for JASPAR 2016. R package version 1.12.0, <http://jaspar.genereg.net/>. (2019).
- 41 Ou, J. *et al.* ATACseqQC: a Bioconductor package for post-alignment quality assessment of ATAC-seq data. *BMC Genomics* **19**, 169, doi:10.1186/s12864-018-4559-3 (2018).
- 42 Imrichova, H., Hulselmans, G., Atak, Z. K., Potier, D. & Aerts, S. i-cisTarget 2015 update: generalized cis-regulatory enrichment analysis in human, mouse and fly. *Nucleic Acids Res* **43**, W57-64, doi:10.1093/nar/gkv395 (2015).
- 43 Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* **102**, 15545-15550, doi:10.1073/pnas.0506580102 (2005).
- 44 Reimand, J. *et al.* g:Profiler-a web server for functional interpretation of gene lists (2016 update). *Nucleic Acids Res* **44**, W83-89, doi:10.1093/nar/gkw199 (2016).
- 45 Merico, D., Isserlin, R., Stueker, O., Emili, A. & Bader, G. D. Enrichment map: a network-based method for gene-set enrichment visualization and interpretation. *PLoS One* **5**, e13984, doi:10.1371/journal.pone.0013984 (2010).
- 46 Cooper, N. J. *et al.* Type 1 diabetes genome-wide association analysis with imputation identifies five new risk regions. *bioRxiv*, doi:10.1101/120022 (2017).
- 47 Moore, F., Cunha, D. A., Mulder, H. & Eizirik, D. L. Use of RNA interference to investigate cytokine signal transduction in pancreatic beta cells. *Methods Mol Biol* **820**, 179-194, doi:10.1007/978-1-61779-439-1_11 (2012).
- 48 Moore, F. *et al.* PTPN2, a candidate gene for type 1 diabetes, modulates interferon-gamma-induced pancreatic beta-cell apoptosis. *Diabetes* **58**, 1283-1291, doi:10.2337/db08-1510 (2009).
- 49 Rasschaert, J. *et al.* Toll-like receptor 3 and STAT-1 contribute to double-stranded RNA+ interferon-gamma-induced apoptosis in primary pancreatic beta-cells. *J Biol Chem* **280**, 33984-33991, doi:10.1074/jbc.M502213200 (2005).
- 50 Cunha, D. A. *et al.* Initiation and execution of lipotoxic ER stress in pancreatic beta-cells. *J Cell Sci* **121**, 2308-2318, doi:10.1242/jcs.026062 (2008).

In the format provided by the authors and unedited.

The impact of proinflammatory cytokines on the β -cell regulatory landscape provides insights into the genetics of type 1 diabetes

Mireia Ramos-Rodríguez ¹, Helena Raurell-Vila¹, Maikel L. Colli², Maria Inês Alvelos², Marc Subirana-Granés¹, Jonàs Juan-Mateu², Richard Norris¹, Jean-Valery Turatsinze², Ernesto S. Nakayasu ³, Bobbie-Jo M. Webb-Robertson³, Jamie R. J. Inshaw⁴, Piero Marchetti⁵, Lorenzo Piemonti ⁶, Manel Esteller^{7,8,9,10}, John A. Todd⁴, Thomas O. Metz³, Décio L. Eizirik² and Lorenzo Pasquali ^{1,7,11*}

¹Endocrine Regulatory Genomics Laboratory, Germans Trias i Pujol University Hospital and Research Institute, Badalona, Spain. ²Center for Diabetes Research and Welbio, Medical Faculty, Université Libre de Bruxelles, Brussels, Belgium. ³Biological Sciences Division, Pacific Northwest National Laboratory, Richland, WA, USA. ⁴JDRF/Wellcome Diabetes and Inflammation Laboratory, Wellcome Centre for Human Genetics, Nuffield Department of Medicine, NIHR Oxford Biomedical Research Centre, University of Oxford, Oxford, UK. ⁵Department of Clinical and Experimental Medicine, University of Pisa, Pisa, Italy. ⁶Diabetes Research Institute, San Raffaele Scientific Institute, Milan, Italy. ⁷Josep Carreras Leukaemia Research Institute, Barcelona, Spain. ⁸Centro de Investigación Biomédica en Red Cáncer, Madrid, Spain. ⁹Institució Catalana de Recerca i Estudis Avançats, Barcelona, Spain. ¹⁰Physiological Sciences Department, School of Medicine and Health Sciences, University of Barcelona, Barcelona, Spain. ¹¹CIBER de Diabetes y Enfermedades Metabólicas Asociadas, Barcelona, Spain. *e-mail: lpasquali@igtp.cat

Table of Contents

Supplementary Notes:	2
1. Ethics.....	2
2. Human Islets.....	2
3. EndoC- β H1.....	3
4. Glucose-Stimulated Insulin Secretion.....	3
5. ATAC-seq.....	3
6. ChIP-seq.....	4
7. ATAC-seq, ChIP-seq read mapping and data processing.....	5
8. RNA-seq.....	6
9. Differential analysis of ATAC-seq, RNA-seq, ChIP-seq.....	6
10. Proteomics.....	7
11. Sequence conservation analysis.....	8
12. Gene regulatory network analysis.....	8
13. UMI-4C data generation.....	8
14. GWAS association analysis.....	9
Supplementary Tables:	11
Supplementary Table 3. T1D risk loci overlapping human islets induced regulatory elements*.....	11
Supplementary Table 5. Metadata from human islet samples.....	12
Supplementary Table 8. Oligonucleotides used for UMI-4C experiments.....	13
Supplementary Table 9. UMI-4C sequencing statistics.....	14
Supplementary Table 10. Oligonucleotides used for luciferase assays.....	15
References	16

Supplementary Notes:

1. Ethics

Human islets were isolated from brain-dead organ donors in accordance with national laws and institutional ethical requirements at the Istituto Scientifico San Raffaele, Milan, Italy and University of Pisa, Italy. All experiments were performed according to protocols approved by the institutional research committee of the Institute for Health Science Research Germans Trias i Pujol, Badalona, Spain.

2. Human Islets

A total of 14 human pancreatic islets were isolated from multiorgan donors without a history of glucose intolerance after informed consent from family members. Pancreatic islets were isolated according to established isolation procedures^{4,5}. After isolation, islets were incubated at 37°C in CMRL 1066 medium with 10% fetal calf serum for 24 h prior to shipment at room temperature in the same culture medium. Upon arrival, samples were recultured at 37°C in Ham's F-10 medium containing 10% fetal bovine serum (FBS), 2 mM GlutaMAX, 50 U/ml penicillin and 50 µg/ml streptomycin (GIBCO), 6.1 mM glucose, 50 µM 3-isobutyl-1-methylxanthine, 1% BSA (Sigma) for 1-3 days. The islets were then exposed or not to cytokines in the same medium without FBS for 48 hours. The following cytokine concentrations were used, as from previous dose-response experiments⁶⁻⁸: recombinant human IL-1β (specific activity 1.8×10⁷ U/mg; 201-LB-005, R&D Systems, Abingdon, UK) at 50 U/ml; recombinant human IFN-γ (specific activity 2×10⁷ U/mg; AF-300-02, Peprotech, London, UK) at 1000 U/ml as described⁶. Islets were next rinsed with phosphate buffered saline at room temperature, formaldehyde was added to a final concentration of 1%, and samples were incubated at room temperature with constant shaking for 10 minutes. Glycine was added to a final concentration of 125 mM for 5 minutes, and cells were rinsed two times with phosphate-buffered saline containing 1x protease inhibitor cocktail (Millipore) at 4°C. The cells were spun down at 1000 rpm for 4 minutes, snap frozen, stored at -80°C and then used for ChIP, and 4C experiments. Additional aliquots of the same sample were processed similarly but without exposing to formaldehyde or glycine and frozen for RNA extraction or directly processed for ATAC-seq experiments. Islet purity was initially assessed by dithizone staining using an aliquot of islets immediately prior to fixation. Only islet preparations with minimal exocrine contamination were selected for the experiments described here (**Supplementary Table 5**). Cell viability was assessed in 4 of the 5 human islets exposed or not to IL-1β + IFN-γ and used for RNAseq, as previously described⁹. The results confirmed, as previously observed^{9,10}, that the cytokine cocktail used in this study affects islet cell viability (control human islets viability 94 ± 1%; cytokine-exposed human islet viability 72 ± 4%. Paired T-test $P=0.02$).

High correlation levels are observed within replicates of the EndoC-βH1 β cell line and within biological replicates of human pancreatic islets primary tissues (**Extended Data Fig. 1a**). Importantly, we still observe high correlation when comparing human islets and EndoC-βH1. Lower correlation values in the latter comparison, compared to inter-replicates correlations are due to the heterogeneity of the human islet cell population composed primarily of β cells (~50-60%) but including other cell types such as α, γ, δ or ε cells.

3. EndoC-βH1

The human insulin-producing EndoC-βH1 cells, kindly provided by Dr R. Scharfmann, University of Paris, France¹¹, were cultured in DMEM medium containing 5.6 mmol/l glucose, 2% BSA fraction V, 50 μmol/l 2-mercaptoethanol (Sigma-Aldrich, Poole, UK), 10 mmol/l nicotinamide (Calbiochem, Darmstadt, Germany), 5.5 μg/ml transferrin, 6.7 ng/ml sodium selenite (Sigma-Aldrich), 100 units/ml penicillin and 100 μg/ml streptomycin (Lonza, Leiden, Netherlands). The same concentrations of cytokines as described for the human islets experiments were used in the treatment of EndoC-βH1 cells¹⁰.

4. Glucose-Stimulated Insulin Secretion

The glucose stimulation index tested on 4 human islet preparations ranged from 3.3 to 7.9, indicating that the islets were functionally competent. The glucose stimulation index was also assessed for EndoC-βH1 cells exposed or not to proinflammatory cytokines. In line with the cell viability experiments we observed that exposure to IL-1β + IFN-γ resulted in reduction of the glucose-induced insulin secretion (**Extended Data Fig. 7**).

Insulin secretion studies in human islets were performed as previously described¹². Briefly, islets were first kept at 37°C for 45 min in Krebs–Ringer bicarbonate (KRB), 0.5% (vol/vol) albumin, pH 7.4, containing 3.3 mmol/L glucose (wash-out phase). Then, the medium was replaced with KRB containing 3.3 mmol/L glucose to assess basal insulin secretion during 45 min, followed by further 45 min incubation at 16.7 mmol/L glucose to assess insulin response to acute challenge. Insulin was quantified by RIA (Pantec Forniture Biomediche, Turin, Italy). Results were corrected for the number of islets. EndoC-βH1 cells were incubated in Krebs-Ringer buffer without glucose for 1h and then in the same buffer with 0 mmol/L glucose or 20 mmol/L glucose for 40 min (modified from¹³) following Univercell-Biosolutions' protocol (<http://www.univercell-biosolutions.com>). Insulin release and insulin content were measured in cell-free supernatants and acid/ethanol-extracted cell lysates, respectively, using a human insulin ELISA kit (Merckodia, Uppsala, Sweden). Results were normalized by total protein content.

5. ATAC-seq

ATAC-seq library preparations were carried out as previously described¹⁴ with minor modifications^{3,15}. Briefly, we selected 50 healthy and acinar-free islets corresponding approximately to 50,000 cells and isolated their nuclei by incubating the islets in 300 μl cold lysis buffer for 25 min on ice and resuspending them after 5 and 15 min using a syringe with a 29G needle. The pellet was next washed in 100 μl of lysis buffer and centrifuged for 15 min at 500 g at 4°C. The transposition reaction was carried out in a 25 μl reaction mix containing 2 μl of Tn5 transposase, 12.5 μl of TD buffer (Nextera DNA Library Prep Kit, 15028212, Illumina, San Diego, USA) and 10.5 μl DEPC-treated water. The transposition reaction mix was incubated at 37°C for 1 h following inactivation by incubating for 30 min at 40°C after addition of 5 μl of clean up buffer (900 mM NaCl, 300 mM EDTA), 2 μl of 5% SDS and 2 μl of Proteinase K. Isolation of the tagmented DNA was performed with 2x SPRI beads cleanup (Agencourt AMPure XP, Beckman Coulter) and was eluted in 20 μl DEPC treated water.

Two sequential 9-cycle PCR were performed in order to enrich for small tagmented DNA fragments. The PCR mix consisted of 2 μ l of PCR Primer 1 (25 μ M working stock), 2 μ l of Barcoded PCR Primer 2 (25 μ M working stock; sequences provided in¹⁴), 25 μ l of NeBNext High-Fidelity 2x PCR Master Mix, 1 μ l of DEPC-treated water and 20 μ l of the eluted sample (DEPC water was added to compensate in case the volume of eluted DNA was less than 20 μ l). The library was amplified in a thermocycler using the following program (leave preheated to 72°C): 72 °C for 5 min; 98 °C for 30 s; 9 cycles of 98 °C for 10 s, 63 °C for 30 s; and 72 °C for 1 min; and at 4 °C hold. After the first PCR round, fragments smaller than 600bp were selected using SPRI cleanup beads. The DNA library was finally purified using the MinElute PCR Purification Kit (Qiagen, Hilden, Germany), following kit instructions and eluting 2 x 10 μ l with the elution buffer. TapeStation was performed to check library quality and nucleosomal pattern of the fragments distribution resulted from the tagmentation reaction (**Extended Data Fig. 8a**). Semi-quantitative PCR assays at target positive and negative control sites were performed to estimate the efficiency of the ATAC-seq experiment before sequencing (data not shown).

6. ChIP-seq

ChIP-seq was carried out using tagmentation (ChIPmentation) as previously described¹⁶. Briefly, 4,000 islet equivalents or ~4 million 1% formaldehyde fixed cells were lysed in sonication buffer (2% Triton X-100, 100mM NaCl, 10mM Tris-HCl pH 8.0, 1mM EDTA pH 8.0, 1% SDS, 1 x protease inhibitors cocktail (Millipore)) and sonicated with a Diagnode Bioruptor with the following conditions: 20 cycles of 1 minute (40 seconds on, 20 seconds off) at high power, to obtain a fragment size in the range of 100–500 bp. Lysates were centrifuged at 13,000 RPM for 5 min at 4°C, and the supernatant containing the sonicated chromatin was diluted with a ChIP dilution buffer (50mM Hepes pH 8.0, 1mM EDTA pH 8.0, 140mM NaCl, 0.75% Triton X-100, 0.1% sodium deoxycholate, 1 x protease inhibitors cocktail (Millipore)) to a final volume of 1ml per immunoprecipitation (IP). The antibody against H3K27ac epitope (1.5 μ g per IP, Abcam ab4729) was added along with 50 μ l 10% BSA and the IP was incubated over night at 4 °C. For each IP, 20 μ l magnetic Protein A+G (Millipore) were washed twice and resuspended in PBS. The beads were then added to the IP and incubated for 2h at 4°C on a rotator. Blocked antibody-conjugated beads were then placed on a magnet and the supernatant was removed. Beads were subsequently washed with a low salt wash buffer (20 mM Tris-HCl pH 8.0, 2mM EDTA pH 8.0, 150 mM NaCl, 1% Triton X-100, 0.1% SDS), high salt wash buffer (20 mM Tris-HCl pH 8.0, 2mM EDTA pH 8.0, 500 mM NaCl, 1% Triton X-100, 0.1% SDS) and LiCl wash buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA pH 8.0, 250 mM LiCl, 1% DOC and 1% NP40). The IP was incubated for 4 minutes at 4°C with rotation for each wash buffer. Beads were then washed with cold 10mM Tris-Cl pH 8.0, to remove detergent, salts and EDTA. The whole reaction, including beads, was then transferred to a new tube and placed on a magnet to remove supernatant to decrease background. Beads were then carefully resuspended in 30 μ l of the tagmentation reaction mix (10 mM Tris-HCl pH 8.0, 5 mM MgCl₂) containing 1 μ l Tn5 transposase from the Nextera DNA Library Prep Kit (15028212, Illumina, San Diego, USA) and incubated at 37°C for 10 min. The beads were then washed twice with RIPA (10 mM Tris-HCl, pH 8.0, 1 mM EDTA, pH 8.0, 140 mM NaCl, 1% Triton x-100, 0.1% SDS, 0.1% DOC), once with cold Tris-EDTA. Beads were next incubated with 150 μ l elution buffer (1% SDS, 0.1M NaHCO₃) for 15 min with rotation at room temperature. The chromatin immunoprecipitate was

separated from the beads using a magnet. The ChIP was then re-eluted by adding another 150 μ l elution buffer to the beads. Both ChIP elutions were combined in the same tube. Each ChIP elute was then incubated with 0.75 μ l RNaseA (20mg/ml) for 30 min at 37°C. 4.5 μ l Proteinase K (Thermo Scientific), and 12 μ l 5M NaCl was then added and the ChIP was incubated at 65°C overnight to revert formaldehyde cross-linking. Finally, ChIP DNA was purified using a Phenol-chloroform extraction protocol.

Two μ l of each library were amplified in a 10 μ l qPCR reaction containing 0.15 μ M primers, 1 \times SYBR Green and 5 μ l *NEBNext High-Fidelity 2X PCR Master Mix (NEB M0541S)*, to estimate the optimum number of enrichment cycles with the following program: 72°C for 5 min, 98°C for 30 s, 24 cycles of 98°C for 10 s, 63°C for 30 s and 72°C for 30 s, and a final elongation at 72°C for 1 min. Final enrichment of the libraries was performed in a 50 μ l reaction using 0.75 μ M primers and 25 μ l *NEBNext High-Fidelity 2X PCR Master Mix*. Libraries were amplified for N+1 cycles, where N is equal to the rounded-up Cq value determined in the qPCR reaction. Enriched libraries were purified using SPRI AMPure XP beads at a beads-to-sample ratio of 0.7:1, followed by a size selection using AMPure XP beads to recover libraries with a fragment length of 100–500 bp. Semi-quantitative PCR assays at target positive and negative control sites were performed to estimate the efficiency of the ChIP experiment before sequencing (data not shown).

7. ATAC-seq, ChIP-seq read mapping and data processing

ATAC-seq and ChIP-seq libraries were sequenced on Illumina HiSeq 2500 platform to generate 49-50 bp single or paired end reads. 42-180M ATAC-seq reads and 18-120M ChIP-seq reads were aligned to the hg19 reference genome using Bowtie 2 (version 2.3.4.1)¹⁷ with default parameters. After alignment, reads mapping to ENCODE blacklist regions¹⁸, to non-canonical chromosomes or to mitochondrial DNA were discarded. Duplicates were removed using samtools markdup (version 1.8)¹⁹ (see **Supplementary Table 6** for number of mapped reads per experiment). For ATAC-seq samples we performed an offset correction of 4 bp on the + strand and 5 bp on the - strand to adjust the read start sites to the center of the transposon's binding event as previously described¹⁴ using an in-house script.

As measures of quality control, normalized strand cross-correlation coefficients (NSC) and relative strand cross-correlation coefficient (RSC) were calculated for each sample using phantompeakqualtools²⁰ (**Extended Data Fig. 8b**). The NSC represents the normalized ratio between the fragment-length cross-correlation peak and the background cross-correlation and values lower than 1.05 represent low signal to noise ratios and few peaks, which might be due to biological reasons or due to poor quality of the experiment. The RSC represents the ratio between the fragment-length peak and the read-length peak and values lower than 0.8 are also indicative of low signal to noise ratios, probably to a failed or poor quality experiments, low read sequence quality or shallow sequencing depth. As additional measures of the overall quality of the ATAC-seq experiments, we show: 1) the enrichment of ATAC-seq reads 4kb around the TSS of coding genes, compared to a random control set of regions computed using regioneR²¹ (**Extended Data Fig. 8c**); 2) the signal to noise ratios both at distal and promoter regions (**Extended Data Fig. 8d**); 3) Some representative views of ATAC-seq profiles at islet-specific genes *NKX6.1*, *PDX1* and *NEUROD1*, showing the replicability of the experiments (**Extended Data Fig. 8e**).

Open chromatin peaks were called with MACS2 (version 2.1)²² callpeak using the following parameters “-q 0.05 --nomodel --shift -100 --extsize 200”. H3K27ac enriched regions were identified with the same software using the following parameters “--broad --broad-cutoff 0.1 --nomodel”. We compared the number of called peaks with other published ATAC-seq datasets in human islets and β cells^{23–25} and observed that ~80% of the stronger peaks found in such datasets are also present in our merged peak set.

For both assays, peaks were called separately for each replicate (ATAC-seq: EndoC- β H1 n=5, human islets n=5; ChIP-seq H3K27ac: EndoC- β H1 n=4, human islets n=4) as a measure of quality control. Merged BAM files for each condition and experiment were converted to bedgraph using bedtools (version 2.26)²⁶ genomeCoverageBed and transformed into bigWig (bedGraphToBigWig UCSC tool²⁷) to be uploaded to a public server for visualization.

8. RNA-seq

Total RNA was isolated from EndoC- β H1 cells and human islets²⁸ using the RNeasy Mini Kit (Qiagen) which retrieves RNA molecules longer than 200 nucleotides, as previously described in detail¹³. RNA integrity number (RIN) values were evaluated using the Agilent bioanalyzer 2100 (Agilent Technologies, Wokingham, UK). All the samples had RIN values of > 8. RNA-seq libraries generation was performed as described by the manufacturer (Illumina, Eindhoven, The Netherlands). In brief, mRNA was purified from 1–2 μ g of total RNA using oligo (dT) beads, before it was fragmented and randomly primed for reverse transcription followed by second-strand synthesis to create fragments of double-stranded cDNA. The obtained cDNA underwent paired-end repair to produce blunted ends. After 3'-monoadenylation and adaptor ligation, cDNAs were purified on an agarose gel and 200 bp products were excised from the gel. The purified cDNA was amplified by PCR using primers specific for the ligated adaptors. Finally, the libraries were submitted to quality control with the Agilent bioanalyzer 2100.

9. Differential analysis of ATAC-seq, RNA-seq, ChIP-seq

For both ATAC-seq and ChIP-seq, aligned reads from all replicates were merged into a single BAM file to identify a comprehensive set of peaks. We next used such peak set to compute read counts, separately for each replicate and condition. In the case of RNA-seq data, the output of htseq-count²⁹ was used as the input matrix for downstream analysis. The generated matrices were normalized and differential analysis was performed using DESeq2³⁰ using a paired sample design in which the model is defined as “~ replicate + treatment”. Briefly, DESeq2 algorithm performs size factors estimation, to normalize each sample by its library size; next, calculates dispersions for each feature to account for intra-sample variability; finally, it fits a negative binomial generalized linear model (GLM) and calculates the Wald statistic. To improve statistical power, DESeq2 performs an independent step to filter out low mean normalized counts that bear a limited chance of being significant. To control for type I errors, raw *P*-values are adjusted for the False Discovery Rate (FDR) using the Benjamini-Hochberg method. The DESeq2 method allows to detect, for the number of replicates used in this study, differential features with a statistical power of >80% (as calculated by PROPER³¹).

Thresholds for significance were set at an FDR adjusted $P < 0.05$ and $|\log_2 FC| > 1$. All regions/genes that did not reach significance or did not pass \log_2 fold change cutoff were classified as stable/equal-regulated.

10. Proteomics

For the proteomic analysis 1.5 million EndoC- β H1 cells treated or not with cytokines (IL-1 β + IFN- γ) were processed using the Metabolite, Protein and Lipid Extraction (MPLEx) approach³². Protein pellets were dissolved in 50 mM NH_4HCO_3 containing 8 M urea and 10 mM dithiothreitol and shaken at 800 rpm for 1 h at 37 °C. Sulfhydryl groups were alkylated by adding 400 mM iodoacetamide (40 mM final concentration), and incubated for another hour in the dark at room temperature. Samples were then 8-fold diluted with 50 mM NH_4HCO_3 , and CaCl_2 was added to a final concentration of 1 mM from a 1 M stock solution. The digestion was carried out by adding trypsin at 1:50 enzyme:protein ratio and incubation at 37 °C for 5 h. The resulting peptides were extracted using C18 cartridges (Discovery, 50 mg, Sulpelco) and concentrated in a vacuum centrifuge. Peptides were then quantified by BCA, normalized and labeled with tandem mass tags (TMT-10plex, ThermoFisher Scientific) according to the manufacturer's instructions. Labeled peptides were extracted using C18 cartridges and fractionated into 24 fractions using high-pH reversed phase chromatography³³. Peptide fractions were loaded into a C18 column (70 cm \times 75 μm i.d. packed with Phenomenex Jupiter, 3 μm particle size, 300 Å pore size) connected to a Waters NanoAquity UPLC system. A gradient of water (solvent A) and acetonitrile (solvent B) both containing 0.1% formic acid (1-8% B in 2 min, 8-12% B in 18 min, 12-30% B in 55 min, 30-45% B in 22 min, 45-95% B in 3 min, hold for 5 min in 95% B and 99-1% B in 10 min) was used to elute the peptides, which were directly analyzed by nanoelectrospray ionization on a Q-Exactive mass spectrometer (Thermo Fisher Scientific). Scans were collected with a resolution of 35,000 at 400 m/z in a 400–2000 m/z range. High-energy collision-induced dissociation (HCD) fragmentation were set for the 12 most intense parent ions using the following parameters: peptide charge ≥ 2 , 2.0 m/z isolation width, 30% normalized collision energy and 17,500 resolution at 400 m/z. Each parent ion was fragmented only once before being dynamically excluded for 30 s.

Collected data were processed using Decon2LS_V2³⁴ and DTARefinery³⁵, both using default parameters, to recalibrate the runs and generate peak lists. Peptide identifications were done using MSGF+³⁶ by searching peak lists against islet protein sequences deduced from a transcriptomics experiment⁶ and supplemented with keratin sequences (32,780 total protein sequences). For MSGF+ searches, a parent ion mass tolerance of 10 ppm, partially tryptic digestion and 2 missed cleavages were allowed. The following modifications were also considered during searches: cysteine carbamidomethylation and N-terminal/lysine TMT addition as static modifications, and methionine oxidation as a variable modification. Results were filtered in two steps to a final false-discovery rate $< 1\%$: spectral-peptide matches - MSGF probability $\leq 1.0 \times 10^{-9}$, and protein level $\leq 1.0 \times 10^{-10}$. The intensity of TMT reporter ions was extracted using MASIC³⁷.

Data quality of the multiple omics sets was assessed by evaluating of the distribution of the data for each sample via the robust Mahalanobis distance abundance vector (rMd-PAV) algorithm³⁸. Data were then converted into \log_2 and normalized by standard median centering. Proteins were quantified using a

Bayesian proteoform discovery methodology (BP-Quant) in combination with standard reference-based median quantification³⁹ and were considered significant with a cutoff of $p \leq 0.05$ based on a paired T-test.

11. Sequence conservation analysis

Sequence conservation at different classes of induced regulatory elements was assessed by determining average phastCons 46 way score in placental mammals⁴⁰. The scores average was calculated in 50bp bins over a 2kb window centered on the open chromatin site. A set of regions identical in number and size shuffled over the mappable genome using regioneR²¹ were used as control set.

12. Gene regulatory network analysis

Regulatory networks were constructed using 1.9 *networkx* python module⁴¹. IRE enhancers, harboring *de novo* enriched TF binding motifs, were linked to their inferred target gene as described above. A random network with the same number of nodes as the observed network and an edge creation probability $p=0.1$ was generated for comparison using the *erdos_renyi_graph* module. The inferred networks were visualized using Cytoscape (version 3.6.0)⁴² applying a force directed layout.

Functional enrichment analysis of the IREs gene targets (**Extended Data Fig. 4c**) was performed by Metascape¹. The enrichment analysis for Gene Ontology biological process included only terms with $P < 0.001$ and with at least 3 enriched genes.

13. UMI-4C data generation

UMI-4C was performed as described⁴³ with minor modifications. Briefly, formaldehyde crosslinked human islets (estimated ~4 million individual cells) were treated with ice-cold lysis buffer (50 mM Tris-HCl pH 8, 150 mM NaCl, 1% TX-100, 0.5% IGEPAL-CA-630, Sigma; 5mM EDTA, AppliChem; 1X protease inhibitor cocktail, Millipore) and permeabilized with SDS (Millipore) and TX-100. Nuclei were digested with DpnII (New England Biolabs) for 16h at 37°C. Afterwards, more DpnII was added for 4h and ligated with T4 DNA ligase (Promega) at 16°C overnight. Cross-link was reversed by overnight incubation with Proteinase K at 65°C. Next, after 30 min incubation with 30 μ l of 10 mg/ml RNase A at 37°C, DNA was purified by phenol/chloroform and ethanol precipitation. DNA was then resuspended in 10 mM Tris-HCl pH 8 and 5 μ g aliquots were sonicated by Covaris S2 to fragment sizes in the range of 450-550 bp. Sonicated DNA was next incubated for 30 min at 20°C with 20 μ l 10 \times end-repair buffer and 10 μ l end-repair mix (NEB E6050S). The reaction was next cleared by 2 \times AmpureXP beads (Agencourt AMPure XP, Beckman Coulter) and eluted in 75 μ l EB (10 mM Tris-HCl pH 8). A-tailing was performed on the elute by adding 10 μ l NEB buffer 2, 4 μ l Klenow fragment (NEB M0212L) and 10 μ l 10 nM dATP 20 min at 37 °C and 20 min at 75°C to inactivate the reaction. DNA was next dephosphorylated with 2 μ l calf intestinal alkaline phosphatase (NEB M0290S) and incubated at 50°C for 60 min and immediately purified by 2 \times AmpureXP beads. Illumina forked adaptors were added (final concentration 0.4 μ M) and incubated with 80 μ l quick ligase buffer and 5 μ l of Quick Ligase (NEB M2200S) for 20 min at 25°C. DNA was next denatured at 96°C for 5 min, moved to ice for 5 min and then purified by 0,8 \times AmpureXP beads to release the non-ligated strand of the adaptor. Denatured DNA was quantified by Qubit (ssDNA HS Assay, Q32851, Invitrogen).

200 ng of DNA were used for library preparation by nested PCR. The first PCR reaction was performed with 10 μ l GoTaq Flexi Buffer, 3 μ l MgCl₂ 25mM, 1 μ l dNTP 10 mM, 2 μ l US bait primer and 2 μ l Illumina universal primer 10 μ M (0,4 μ M final concentration each primer) and 1 μ l GoTaq (Promega, M5005) with the following program: 2 min 95°C, 20 cycles of 30 s 95°C, 30 s 56°C and 60 s 72°C and final extension of 5 min 72°C. After the first PCR reaction, the DNA was purified with 1 \times AmpureXP beads, and 31 μ l of the elute was used to perform a second PCR reaction following the same conditions as the first one but using the DS bait primer and 17 amplification cycles instead of 20. Finally, DNA was purified by 0,8x AmpureXP beads. To increase molecular complexity, each library was obtained by pooling 5-10 PCRs per viewpoint. The PCR primers used in UMI-4C are listed in **Supplementary Table 8**. Each library was sequenced to a depth of >1M 75bp long paired-end reads using either NextSeq or HiSeq 2500 platforms.

14. GWAS association analysis

Genome wide association statistics were generated from 5,909 type 1 diabetes cases (from the UK GRID⁴⁴ cohort) and 8,721 controls (1,397 from the UK blood service⁴⁵, 5,472 from the 1958 British birth cohort⁴⁶ and 1,852 from the bipolar disorder cases from the welcome trust case control consortium⁴⁵). 6,653 (1,926 cases and 4,727 controls) were genotyped using the Affymetrix GeneChip 500k array and 7,992 (3,983 cases and 3,994 controls) were genotyped using the Illumina Infinium 550k array. Prior to imputation, we filtered SNPs with a call rate of less than 95%. Shapeit2⁴⁷ was used to pre-phase individual haplotypes using the haplotype reference consortium (HRC) as the haplotype reference panel, then minimac3⁴⁸ was used to impute both batches genome-wide for chromosomes 1-22. As post imputation quality control, we excluded SNPs with a Minor Allele Frequency (MAF) <0.01 (or MAF<0.05 if the SNP was only present in one of the genotyping batches), an imputation information score <0.3 a difference in MAF between Illumina and Affymetrix controls of >5%, a difference between Illumina and Affymetrix cases of >5% and a difference in MAF of >5% between combined controls and European members of the HRC. As a further quality control (QC), we removed SNPs where the difference in log-odds ratio between cases and controls was >0.3, or the odds ratio difference was >0.3 and the effects were in opposite directions between batches. Finally, if a SNP was present in only one of the genotyping batches, we removed SNPs with an imputation r^2 of less than 0.8. This conservative QC procedure was implemented to minimize the probability of observing spurious associations due to imputation.

SNPTEST⁴⁹ was used to test the association of each SNP with type 1 diabetes status, adjusting for the top 3 principal components, for those genotyped with the Illumina chip and the Affymetrix chips separately, then results were combined in an inverse-variance weighted fixed effects meta-analysis.

Once association results were generated, we filtered-out rare variants with MAF < 0.01. For the definition of 99% credible sets we followed previously defined methodology⁵⁰. Shortly, we took all SNPs with $P < 5 \times 10^{-8}$ and created a 1Mb window around them, merging all overlapping windows. This allows us to generate the potential risk loci; we considered as the top SNP of the locus the one with the smallest P -value. Next, taking advantage from the R^2 scores from 1,000 Genomes phase 3 we removed from further analysis those variants whose association (R^2) with the top SNP in the locus was < 0.1. Finally, we calculated the

Approximate Bayes Factor (ABF) and Posterior Probability (PP) of each variant within the locus, and included in the credible set all those variants until reaching a cumulative PP over 0.99.

Supplementary Tables:

Supplementary Table 3. T1D risk loci overlapping human islets induced regulatory elements*.

Mapped genes	Lead variants**	locus coordinates	# Overlapping HI IREs***	Closest up-regulated genes****	SNPs directly overlapping IREs*****
<i>SLC25A38P1, FASLG</i>	rs78037977	chr1:172715702-172715702	1	<i>TNFSF18</i>	rs78037977
<i>CAMSAP2</i>	rs6691977	chr1:200718569-200832857	1	<i>GPR37L1</i>	None
<i>RF00019, IL10</i>	rs3024505	chr1:206939904-206955041	2	<i>IL19</i>	rs3024493
<i>IL2RA</i>	rs61839660, rs12722495	chr10:6078553-6097283	1	<i>IL2RA</i>	None
<i>IL2RA, RPL32P23, RBM17</i>	rs10795791, rs12251307, rs41295121	chr10:6098949-6217022	2	<i>IL2RA</i>	None
<i>AL353149.1, RNLS</i>	rs10509540, rs12416116	chr10:90015401-90051317	1	<i>LIPM</i>	None
<i>SH2B3, ATXN2</i>	rs3184504, rs653178	chr12:111826477-112059557	2	<i>SH2B3</i>	None
<i>NAA25</i>	rs17696736	chr12:112486818-112610714	4	<i>TRAFD1</i>	None
<i>CLEC16A</i>	rs12708716, rs12927355, rs2903692	chr16:11158885-11240854	3	<i>RM12</i>	None
<i>RM12</i>	rs193778, rs416603	chr16:11342785-11367477	4	<i>RM12, SOCS1</i>	rs149310, rs193779, rs193778, rs243330, rs243327
<i>CLN3, IL27, NUPR1</i>	rs151234, rs4788084	chr16:28483061-28653523	2	<i>IL27, NUPR1</i>	rs12446550
<i>ZPBP2, GSDMB</i>	rs2872507, rs12453507, rs2290400	chr17:37922259-38088417	2	<i>GSDMB</i>	rs7219923, rs201389301
<i>PTPN2</i>	rs12971201	chr18:12820900-12884343	2	<i>PTPN2</i>	rs2847281
<i>UBASH3A</i>	rs11203202, rs11203203, rs9976767, rs3788013	chr21:43820573-43841827	1	<i>UBASH3A</i>	None
<i>CCRL2, LINC02009</i>	rs113010081	chr3:46384204-46457412	2	<i>CCR5, CCRL2</i>	None
<i>KIAA1109, ADAD1, IL2, IL21</i>	rs4505848, rs6534347, rs75793288, rs17388568, rs2069762	chr4:123067808-123501086	7	<i>FGF2</i>	rs55969942, rs55904957, rs4505848, rs77516441
<i>LINC02357</i>	rs10517086	chr4:26085480-26128710	3	<i>SLIT2</i>	rs7441808, rs874040
<i>IL7R, AC112204.3, CAPSL</i>	rs6897932, rs11954020, rs1445898	chr5:35827351-35919378	9	<i>IL7R</i>	rs6890853, rs10213865, rs10214273, rs34143578

* T1D-T2D shared loci were excluded from this analysis (see text)

** Leading variants for the region as defined in the GWAS Catalog⁶² and Cooper et al⁶³.

*** IREs H3K27ac log₂ FC>0.8

**** Closest gene is defined as the overlapping or nearest up-regulated genes to the locus.

***** SNPs defined as those with strong LD (R²≥0.8) with the lead variant.

Supplementary Table 5. Metadata from human islet samples.

Sample	Gender	Age	BMI	% Purity	Cause of Death	Pre-mortem diabetes	Experiments
HI19	Male	34	23.1	80	Cerebral bleeding	No	ATAC-seq, ChIP-seq H3K27ac, UMI-4C
HI22	Male	52	25.1	85	Trauma	No	ATAC-seq, ChIP-seq H3K27ac, UMI-4C
HI24	Male	54	26.5	90	Anoxia	No	ATAC-seq, UMI-4C
HI28	Male	60	22.7	85	Cerebral bleeding	No	UMI-4C
HI32	Male	62	23.1	90	Cerebral bleeding	No	UMI-4C
HI37	Female	53	21.8	85	Cerebral bleeding	No	ChIP-seq H3K27ac
HI40	Female	62	29.3	90	Cerebral bleeding	No	ChIP-seq H3K27ac
HI42	Female	81	21.5	70	Cerebral bleeding	No	ATAC-seq
HI44	Male	58	27.8	90	Cerebral bleeding	No	ATAC-seq
HI06	Female	23	22.5	46	Cardiac arrest	No	RNA-seq*
HI07	Male	31	27.8	66	Cerebral bleeding	No	RNA-seq*
HI08	Male	77	24.5	59	Cerebral bleeding	No	RNA-seq*
HI09	Female	64	29.4	47	Cerebral bleeding	No	RNA-seq*
HI10	Female	58	21.3	67	Cerebral bleeding	No	RNA-seq*

* Previously reported in Gonzalez-Duque, S. et al. Conventional and Neo-Antigenic Peptides Presented by β Cells Are Targeted by Circulating Naïve CD8+ T Cells in Type 1 Diabetic and Healthy Donors. *Cell Metab.* (2018).

Supplementary Table 8. Oligonucleotides used for UMI-4C experiments.

Up-regulated genes

Bait Name	US Primer	DS Primer	PAD	Bait Position
<i>DOCK9</i>	GAGTTGGACAGGGGAGAAGAGAAG	AACTCCCAGACCCCTCCC	CGCCACCCAAGATC	chr13:99738257
<i>IFIH1</i>	GCAAACCTGTGAAGAACTGCCTG	CTCAAAGCTCCTACCCGAGTG	TGCAGCAGGATC	chr2:163175129
<i>SOCS1</i>	GCCGGAGAAAGGCTGTGC	CCCAAATCGCCAGACCAG	GCGCGGATC	chr16:11349720
<i>GBP1</i>	ATAGAACTGAGCACAGAAAGACT	CACTGTAACCTAATCTGAACCCAAGC	ACTCATGATC	chr1:89531571
<i>CIITA</i>	CTGAGTTGGAGAGAAACAGAGACC	GGACAAGCTCCCTGCAACTCA	GGACTTGCAGATC	chr16:10972544
<i>TNFSF10</i>	TCAAAGGAGAGCAAGAAAGAGAAG	TTGAGGTGAGTGCAGATAAGGG	GTGCATGGATC	chr3:172241434
<i>ATF3</i>	CTCGCCTTCATCCCTTCGC	TCTAAGTAACCAGTCCGCCG	ACCCCTGACCCACCCCGGATC	chr1:212782375
<i>CMPK2</i>	AAATCAGGTAGGGAAGGTGCCTG	CAGGCAACTCTTTTAAGATCTTGT	TTGCAAGGATC	chr2:7004967
<i>LAMP3</i>	GTGGTCTGACAGATTGTGCCAT	AGGTGTAGGAAGAGTGTCAAAGAC	TTTTTTTTTTTTAAATGATC	chr3:182879972
<i>RSAD2</i>	GCTGAGTTAGAGTTTCGATTTTTCC	AAGTTGAAACAGGGCCAAGAC	TGAAGACTAGAGATC	chr2:7017732
<i>CXCL11</i>	CTTCAGTAACTCCATGCCACG	GTTAGGGGTGAAGCCACACA	CCTCTGGGATC	chr4:76958088
<i>IFI1</i>	TGGCTGCTGTTTAGCTCCCTTAT	ACGTAAC TGAAAATCCACAAGACA	GAATAGCCAGATC	chr10:91152408
<i>IRF1</i>	ACGCGGCGTGGACCG	CGGACGAGGCTGCCG	GCGCCCGGCAGCTTTCGCAGATC	chr5:131825992

T1D risk loci

Bait Name	US Primer	DS Primer	PAD	Bait Position
<i>DEXI</i>	AACGCCCGCGCTGATT	TTGCTTTCCTGGTTCGGGTC	TGCCCTCAGATC	chr16:11036502
rs78037977	GAGTGTTTGCCAATGTTTGAAGTA	TTAAAGTGATGTTTGTGGCTAAGAG	CTCAAGATC	chr1:172715942

Supplementary Table 9. UMI-4C sequencing statistics.

Bait	Treatment	Raw reads	Bait-specific reads	% specificity	# UMIs
<i>ATF3</i>	IFN- γ + IL-1 β 48h	17,748,014	16,227,284	91	11,184
<i>ATF3</i>	Control	19,033,231	17,404,550	91	16,240
<i>CIITA</i>	IFN- γ + IL-1 β 48h	8,510,230	7,759,081	91	9,146
<i>CIITA</i>	Control	8,902,916	8,189,179	92	11,221
<i>CMPK2</i>	IFN- γ + IL-1 β 48h	4,130,031	3,778,254	91	10,290
<i>CMPK2</i>	Control	3,801,896	3,489,780	92	11,619
<i>CXCL11</i>	IFN- γ + IL-1 β 48h	8,441,058	7,653,148	91	10,276
<i>CXCL11</i>	Control	9,066,024	8,243,721	91	12,981
<i>DEXI</i>	IFN- γ + IL-1 β 48h	12,214,438	10,728,755	88	6,622
<i>DEXI</i>	Control	11,777,000	10,422,707	89	9,657
<i>DOCK9</i>	IFN- γ + IL-1 β 48h	1,172,909	577,184	49	9,955
<i>DOCK9</i>	Control	4,972,466	1,871,266	38	12,673
<i>GBP1</i>	IFN- γ + IL-1 β 48h	1,498,820	1,399,799	93	4,630
<i>GBP1</i>	Control	1,630,829	1,525,062	94	6,070
<i>IFIH1</i>	IFN- γ + IL-1 β 48h	1,561,369	1,351,149	87	26,263
<i>IFIH1</i>	Control	2,066,371	1,766,431	85	29,074
<i>IFIT1</i>	IFN- γ + IL-1 β 48h	7,997,529	7,361,083	92	9,657
<i>IFIT1</i>	Control	8,097,838	7,440,040	92	11,529
<i>IRF1</i>	IFN- γ + IL-1 β 48h	1,374,134	514,925	37	2,399
<i>IRF1</i>	Control	1,650,279	773,405	47	3,242
<i>LAMP3</i>	IFN- γ + IL-1 β 48h	2,495,385	847,831	34	9,437
<i>LAMP3</i>	Control	2,616,115	887,928	34	10,227
rs78037977	IFN- γ + IL-1 β 48h	1,505,820	1,353,025	90	24,163
rs78037977	Control	1,262,331	1,118,647	89	28,079
<i>RSAD2</i>	IFN- γ + IL-1 β 48h	14,440,385	13,144,612	91	13,247
<i>RSAD2</i>	Control	15,730,693	14,350,630	91	16,257
<i>SOCS1</i>	IFN- γ + IL-1 β 48h	3,336,826	2,118,570	63	19,843
<i>SOCS1</i>	Control	4,277,311	2,663,116	62	23,934
<i>TNFSF10</i>	IFN- γ + IL-1 β 48h	10,273,497	9,438,691	92	11,429
<i>TNFSF10</i>	Control	11,566,075	10,669,468	92	14,495

Supplementary Table 10. Oligonucleotides used for luciferase assays.

Amplification RE		
SNP	Fw Primer	Rv Primer
rs78037977	CCGCTCGAGAGTGGGCTCTTGTTTCAATGT	CCGAAGCTTTGGCAAACACTCAAAGATTCAA
rs193778	CCGCTCGAGGTGGTCGGGGCTGTAATTTG	CCGAAGCTTCACACACACGCACAACCTAG

Site Directed Mutagenesis		
SNP	Fw Primer	Rv Primer
rs78037977	CTTTCCTCTGAGGAGTAAGAGTGACCCCTTGCTTAAAAG	CTTTAAGCAAGGGTCACTCTTACTCCTCAGAGGAAAAG
rs193778	GCTCCGTGTAATTTTGGATGAAGGCATGTGAGG	CCTCACATGCCTTCATCCAAAATACACGGAGC

References

1. Tripathi, S. *et al.* Meta- and Orthogonal Integration of Influenza “OMICs” Data Defines a Role for UBR4 in Virus Budding. *Cell Host Microbe* **18**, 723–735 (2015).
2. Miguel-Escalada, I. *et al.* Human pancreatic islet three-dimensional chromatin architecture provides insights into the genetics of type 2 diabetes. *Nat. Genet.* **51**, 1137–1148 (2019).
3. Raurell-Vila, H., Ramos-Rodríguez, M. & Pasquali, L. Assay for Transposase Accessible Chromatin (ATAC-Seq) to Chart the Open Chromatin Landscape of Human Pancreatic Islets. in *Methods in Molecular Biology* (eds. Vavouri, T. & Peinado, M. A.) 197–208 (Human Press, 2018). doi:10.1007/978-1-4939-7768-0_11
4. Bucher, P. *et al.* Assessment of a Novel Two-Component Enzyme Preparation for Human Islet Isolation and Transplantation. *Transplantation* **79**, 91–97 (2005).
5. Melzi, R. *et al.* Role of CCL2/MCP-1 in Islet Transplantation. *Cell Transplant.* **19**, 1031–1046 (2010).
6. Eizirik, D. L. *et al.* The Human Pancreatic Islet Transcriptome: Expression of Candidate Genes for Type 1 Diabetes and the Impact of Pro-Inflammatory Cytokines. *PLoS Genet.* **8**, e1002552 (2012).
7. Colli, M. L., Moore, F., Gurzov, E. N., Ortis, F. & Eizirik, D. L. MDA5 and PTPN2, two candidate genes for type 1 diabetes, modify pancreatic β -cell responses to the viral by-product double-stranded RNA. *Hum. Mol. Genet.* **19**, 135–146 (2010).
8. Ortis, F. *et al.* Cytokine-Induced Proapoptotic Gene Expression in Insulin-Producing Cells Is Related to Rapid, Sustained, and Nonoscillatory Nuclear Factor- κ B Activation. *Mol. Endocrinol.* **20**, 1867–1879 (2006).
9. Delaney, C. A., Pavlovic, D., Hoorens, A., Pipeleers, D. G. & Eizirik, D. L. Cytokines Induce Deoxyribonucleic Acid Strand Breaks and Apoptosis in Human Pancreatic Islet Cells 1. *Endocrinology* **138**, 2610–2614 (1997).
10. Brozzi, F. *et al.* Cytokines induce endoplasmic reticulum stress in human, rat and mouse beta cells via different mechanisms. *Diabetologia* **58**, 2307–2316 (2015).
11. Ravassard, P. *et al.* A genetically engineered human pancreatic β cell line exhibiting glucose-inducible insulin secretion. *J. Clin. Invest.* **121**, 3589–3597 (2011).
12. Rondas, D. *et al.* Glucagon-Like Peptide-1 Protects Human Islets against Cytokine-Mediated β -Cell Dysfunction and Death: A Proteomic Study of the Pathways Involved. *J. Proteome Res.* **12**, 4193–4206 (2013).
13. Juan-Mateu, J. *et al.* SRp55 Regulates a Splicing Network That Controls Human Pancreatic β -Cell Function and Survival. *Diabetes* **67**, 423–436 (2018).
14. 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).
15. Lavin, Y. *et al.* Tissue-Resident Macrophage Enhancer Landscapes Are Shaped by the Local Microenvironment. *Cell* **159**, 1312–1326 (2014).
 16. Schmidl, C., Rendeiro, A. F., Sheffield, N. C. & Bock, C. CHIPmentation: fast, robust, low-input ChIP-seq for histones and transcription factors. *Nat. Methods* **12**, 963–965 (2015).
 17. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
 18. Consortium, T. E. P. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
 19. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
 20. Landt, S. G. *et al.* ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. *Genome Res.* **22**, 1813–1831 (2012).
 21. Gel, B. *et al.* regioneR: an R/Bioconductor package for the association analysis of genomic regions based on permutation tests. *Bioinformatics* **32**, btv562 (2015).
 22. Zhang, Y. *et al.* Model-based Analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).
 23. Ackermann, A. M., Wang, Z., Schug, J., Naji, A. & Kaestner, K. H. Integration of ATAC-seq and RNA-seq identifies human alpha cell and beta cell signature genes. *Mol. Metab.* **5**, 233–244 (2016).
 24. Arda, H. E. *et al.* A Chromatin Basis for Cell Lineage and Disease Risk in the Human Pancreas. *Cell Syst.* **7**, 310-322.e4 (2018).
 25. Thurner, M. *et al.* Integration of human pancreatic islet genomic data refines regulatory mechanisms at Type 2 Diabetes susceptibility loci. *Elife* **7**, 1–30 (2018).
 26. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
 27. Kent, W. J., Zweig, A. S., Barber, G., Hinrichs, A. S. & Karolchik, D. BigWig and BigBed: enabling browsing of large distributed datasets. *Bioinformatics* **26**, 2204–2207 (2010).
 28. Gonzalez-Duque, S. *et al.* Conventional and Neo-Antigenic Peptides Presented by β Cells Are Targeted by Circulating Naïve CD8⁺ T Cells in Type 1 Diabetic and Healthy Donors. *Cell Metab.* (2018). doi:10.1016/j.cmet.2018.07.007
 29. Anders, S., Pyl, P. T. & Huber, W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).
 30. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

31. Wu, H., Wang, C. & Wu, Z. PROPER: comprehensive power evaluation for differential expression using RNA-seq. *Bioinformatics* **31**, 233–241 (2015).
32. Nakayasu, E. S. *et al.* MPLEX: a Robust and Universal Protocol for Single-Sample Integrative Proteomic, Metabolomic, and Lipidomic Analyses. *mSystems* **1**, (2016).
33. Wang, Y. *et al.* Reversed-phase chromatography with multiple fraction concatenation strategy for proteome profiling of human MCF10A cells. *Proteomics* **11**, 2019–2026 (2011).
34. Mayampurath, A. M. *et al.* DeconMSn: a software tool for accurate parent ion monoisotopic mass determination for tandem mass spectra. *Bioinformatics* **24**, 1021–1023 (2008).
35. Petyuk, V. A. *et al.* DtaRefinery, a Software Tool for Elimination of Systematic Errors from Parent Ion Mass Measurements in Tandem Mass Spectra Data Sets. *Mol. Cell. Proteomics* **9**, 486–496 (2010).
36. Kim, S. & Pevzner, P. A. MS-GF+ makes progress towards a universal database search tool for proteomics. *Nat. Commun.* **5**, 5277 (2014).
37. Monroe, M. E., Shaw, J. L., Daly, D. S., Adkins, J. N. & Smith, R. D. MASIC: A software program for fast quantitation and flexible visualization of chromatographic profiles from detected LC–MS(/MS) features. *Comput. Biol. Chem.* **32**, 215–217 (2008).
38. Matzke, M. M. *et al.* Improved quality control processing of peptide-centric LC-MS proteomics data. *Bioinformatics* **27**, 2866–2872 (2011).
39. Webb-Robertson, B.-J. M. *et al.* Bayesian Proteoform Modeling Improves Protein Quantification of Global Proteomic Measurements. *Mol. Cell. Proteomics* **13**, 3639–3646 (2014).
40. Siepel, A. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res.* **15**, 1034–1050 (2005).
41. Hagberg, A. A., Schult, D. A. & Swart, P. J. Exploring Network Structure, Dynamics, and Function using NetworkX. in *Proceedings of the 7th Python in Science Conference* (eds. Varoquaux, G., Vaught, T. & Millman, J.) 11–15 (2008).
42. Shannon, P. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Res.* **13**, 2498–2504 (2003).
43. Schwartzman, O. *et al.* UMI-4C for quantitative and targeted chromosomal contact profiling. *Nat. Methods* **13**, 685–691 (2016).
44. Todd, J. A. *et al.* Robust associations of four new chromosome regions from genome-wide analyses of type 1 diabetes. *Nat. Genet.* **39**, 857–864 (2007).
45. Burton, P. R. *et al.* Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* **447**, 661–678 (2007).
46. Power, C. & Elliott, J. Cohort profile: 1958 British birth cohort (National Child Development Study).

Int. J. Epidemiol. **35**, 34–41 (2006).

47. Delaneau, O., Marchini, J. & Zagury, J.-F. A linear complexity phasing method for thousands of genomes. *Nat. Methods* **9**, 179–181 (2012).
48. Das, S. *et al.* Next-generation genotype imputation service and methods. *Nat. Genet.* **48**, 1284–1287 (2016).
49. Marchini, J., Howie, B., Myers, S., McVean, G. & Donnelly, P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat. Genet.* **39**, 906–913 (2007).
50. Bonàs-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).
51. Jorda, M. Upregulation of MMP-9 in MDCK epithelial cell line in response to expression of the Snail transcription factor. *J. Cell Sci.* **118**, 3371–3385 (2005).
52. MacArthur, J. *et al.* The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucleic Acids Res.* **45**, D896–D901 (2017).
53. Cooper, N. J. *et al.* Type 1 diabetes genome-wide association analysis with imputation identifies five new risk regions. *bioRxiv* 120022 (2017). doi:10.1101/120022