UNIVERSITY OF CALIFORNIA SAN DIEGO

Human pluripotent stem cell systems as a model for beta cell development, disease risk, and disease pathogenesis

A dissertation submitted in partial satisfaction of the requirements for the degree Doctor of Philosophy

in

Biomedical Sciences

by

Ryan Joseph Geusz

Committee in charge:

Professor Maike Sander, Chair
Professor Nunzio Bottini
Professor Kyle Gaulton
Professor Ananda Goldrath
Professor Bing Ren

2022

The Dissertation of Ryan Joseph Geusz is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2022

DEDICATION

I would like to thank everyone who has supported me over the course of my PhD and throughout the path that has led me here. I thank my family for nurturing my love of science, and my friends in San Diego for their unending encouragement. In particular, I thank Emily, Aaron, and Mike—I couldn't have done this without you.

TABLE OF CONTENTS

LIST OF SUPPLEMENTAL FILES

Chapter 1 Supplemental Figures: Chap1Supplemental.txt

Chapter 2 Supplemental Figures: Chap2Supplemental.txt

Chapter 3 Supplemental Figures: Chap3Supplemental.txt

Chapter 4 Supplemental Figures: Chap4Supplemental.txt

LIST OF FIGURES

ACKNOWLEDGEMENTS

I would like to acknowledge Drs. Maike Sander and Allen Wang for their mentorship and guidance, as well as all members of the Sander lab for their support throughout my dissertation. I would also like to acknowledge all committee members for their helpful insights and suggestions.

Chapter 1, in full, has been submitted as it appears in Nature Communications 2021. This material was co-authored with Geusz, R.J., Wang, A., Lam, D., Vinckier, N.K., Alysandratos, K., Roberts, D., Wang, J., Kefalopoulou, S., Ramirez, A., Qui, Y., Chiou, J., Gaulton, K.J., Ren, B., Kotton, D.N., and Sander, M. The dissertation author was the principal author of this chapter.

Chapter 2, in full, is a reprint of the material as it appears in Nature Communications 2020. This material was co-authored with Vinckier, N. K., Patel, N. A., Geusz, R. J., Wang, A., Wang, J., Matta, I., Harrington, A., Wortham, M., Wetton, N., Wang, J., Jhala, U., Rosenfeld, M., Benner, C. W., Shih, H. P., and Sander, M. The dissertation author was a co-author of this chapter.

Chapter 3, in full, is a reprint of the material as it appears in eLife 2021. This material was co-authored with Wang, A., Chiou, J., Lancman, J.J., Wetton, N., Kelalopoulou, S., Wang, J., Qiu, Y., Yan, J., Aylward, A., Ren, B., Dong, P., Gaulton, K.J., and Sander, M. The dissertation author was the principal author of this chapter.

Chapter 4 was co-authored with Clarke, D., Zhou, H., Sharma, S., Teyton, L., and Sander, M. The dissertation author was the principal author of this chapter.

VITA

2016 Bachelor of Science, University of Miami

2022 Doctor of Philosophy, University of California San Diego

PUBLICATIONS

Geusz, R.J.*, Wang, A.*, Lam, D., Vinckier, N.K., Alysandratos, K., Roberts, D., Wang, J., Kefalopoulou, S., Ramirez, A., Qui, Y., Chiou, J., Gaulton, K.J., Ren, B., Kotton, D.N., Sander, M. Sequence logic at enhancers governs a dual mechanism of endodermal organ fate induction by FOXA pioneer factors. *Nature Communications* **12**, 6636, https://doi.org/10.1038/s41467-021-26950-0 (2021).

Heller, S., Li, Z., **Geusz, R.,** Breunig, M., Hohwieler, M., Zhang, X., Nair, G., Seufferlein, T., Hebrok, M., Sander, M., Julier, C., Kleger, A., Costa, I. Transcriptional and open chromatin landscape during pancreatic differentiation reveals a role of ONECUT1 in inducing the endocrine transcriptional program. *Communications Biology* **4**, 1298 https://doi.org/10.1038/s42003-021-02818-3 (2021).

Philippi, A., Heller, S., Costa, I.G., Senee, V., Breunig, M., Zhijian, L., Kwon, G., Illing, A., Lin, Q., Hohwieler, M., Degavre, A., Zalloua, P., Liebau, S., Schuster, M., Krumm, J., Zhang, X., **Geusz, R.J.**, Benthuysen, J., Wang, A., Gaulton, K., Neubauer, H., Simon, E., Klein, T., Wagner, M., Russell, R., Nair, G., Besse, C., Olaso, R., Deleuze, J., Kuster, B., Hebrok, M., Seufferlein, T., Sander, M., Boehm, B., Oswald, F., Nicolino, M., Julier, C., Kleger, A. ONECUT1 mutations and variants in diabetes. *Nature Medicine,* doi:10.1038/s41591-021-01502-7 (2021).

Chiou, J., **Geusz, R.J.**, Okino, M., Han, J.Y., Miller, M., Benaglio, P., Huang, S., Korgaonkar, K., Heller, S., Kleger, A., Preissle, S., Gorkin, D., Sander, M., Gaulton, K.J. Interpreting type 1 diabetes risk with genetics and single-cell epigenomics. *Nature* **594**, 398-402, doi:10.1038/s41586-021-03552-w (2021).

Geusz, R.J.*, Wang, A.*, Chiou, J.*, Lancman, J.J., Wetton, N., Kelalopoulou, S., Wang, J., Qiu, Y., Yan, J., Aylward, A., Ren, B., Dong, P., Gaulton, K.J., Sander, M. Pancreatic progenitor epigenome maps prioritize type 2 diabetes risk genes with roles in development. *Elife* **10**, doi:10.7554/eLife.59067 (2021).

Vinckier, N. K.*, Patel, N. A.*, **Geusz, R. J.***, Wang, A., Wang, J., Matta, I., Harrington, A., Wortham, M., Wetton, N., Wang, J., Jhala, U., Rosenfeld, M., Benner, C. W., Shih, H. P., Sander, M. LSD1-mediated enhancer silencing attenuates retinoic acid signalling during pancreatic endocrine cell development. *Nature Communications* **11**, 2082, doi:10.1038/s41467-020-16017-x (2020).

Coronel, M., **Geusz, R.**, Stabler, C. Mitigating hypoxic stress on pancreatic islets via in situ oxygen generating biomaterial. *Biomaterials* **129**, 139-151, doi:10.1016/j.biomaterials.2017.03.018 (2017).

FIELDS OF STUDY

Major Field: Biomedical Sciences

Studies in genetics and genomics
Dr. Maike Sander

ABSTRACT OF THE DISSERTATION

Human pluripotent stem cell systems as a model for beta cell development, disease risk, and disease pathogenesis

by

Ryan Joseph Geusz

Doctor of Philosophy in Biomedical Sciences

University of California San Diego, 2022

Professor Maike Sander, Chair

In response to external signaling factors, human pluripotent stem cells (hPSCs) can be induced to differentiate into diverse cell types. This unique property establishes hPSCs as a valuable model system to study human developmental processes. Furthermore, mature cell types

generated through *in vitro* differentiation protocols can be studied in place of primary cells in circumstances where primary cells are difficult to obtain or exist in small numbers.

Here we used established protocols to differentiate hPSCs towards the pancreatic beta cell fate. Using genomic assays, we characterized changes in the epigenome that accompany the commitment of differentiating cells to the pancreatic organ lineage. Focusing on distal regulatory elements, we investigated the function of FOXA pioneer transcription factors in the activation of pancreatic enhancers. We find that these factors are necessary for pancreatic lineage commitment and show that differential DNA sequence specifies distinct temporal patterns of FOXA recruitment to pancreatic enhancers, with profound effects on gene expression. Using models of liver and lung development, we show relevance of our findings across endodermal lineages.

We next identified a group of pancreatic enhancers that are activated in response to retinoic acid signaling and whose subsequent deactivation is dependent on the enzyme LSD1. We demonstrate a critical role for LSD1 in limiting the duration of signal-dependent enhancer activation during pancreatic lineage commitment.

We then integrated epigenomic analyses with data from genome-wide association studies to identify type 2 diabetes (T2D) risk variants within pancreatic enhancers. Using gene editing in hPSCs, we assigned target genes to variant-containing pancreatic enhancers and performed knockdown experiments in zebrafish to determine the developmental roles of these target genes. These experiments identify a developmental contribution to risk of T2D pathogenesis.

Finally, we established a co-culture system between mature beta-like cells derived from induced pluripotent stem cells of donors with type 1 diabetes (T1D) and autoreactive T cell clones. This system is currently being used to identify a potential contribution of direct antigen presentation by beta cells to T1D pathogenesis.

Altogether, our work demonstrates the utility of *in vitro* hPSC differentiation systems in modeling pancreatic development, disease risk, and disease pathogenesis.

INTRODUCTION

Human pluripotent stem cell (hPSC) systems allow for the generation of mature human cell types using *in vitro* directed differentiation protocols. Gene regulatory elements activated and gene expression profiles induced through these protocols closely resemble those of *in vivo* human development. Here we employ directed differentiation of hPSCs towards human beta-like cells to study gene regulatory elements that orchestrate pancreatic differentiation and to functionally annotate type 2 diabetes (T2D)-associated variants present within these elements. We then use mature hPSC-derived beta-like cells to model cell interactions that may contribute to pathogenesis of type 1 diabetes (T1D).

Human pluripotent stem cells

Human pluripotent stem cells (hPSCs) encompass embryonic stem cells (ESCs) derived from embryos and induced pluripotent stem cells (iPSCs) derived from mature somatic cells. In response to proper environmental signals, these cells can differentiate towards any given cell type, following trajectories that mimic *in vivo* human development[1]. The scalability afforded by stem cell systems allows for generation of large numbers of developmental intermediates and mature cell types. Furthermore, stem cell systems are highly amenable to gene editing or other genomic perturbation techniques. Thus, hPSCs offer a valuable alternative to animal models and primary cell sources for the study of human developmental pathways and generation of mature cell types which may be rare or difficult to obtain.

Here we employ various directed differentiation protocols to guide hPSCs towards the mature beta cell fate. While remaining differences between hPSC-derived and primary beta cells necessitate further protocol optimization[2], cells generated through existing protocols closely resemble primary beta cells in functionality[3], gene expression profiles[4], and chromatin landscape

(unpublished data, Sander laboratory). Thus, directed differentiation of hPSCs provides an ideal model for the study of human pancreatic development and beta cell function.

Gene regulatory elements and the epigenome

Regulation of gene expression is generally conferred through distal noncoding elements known as enhancers. These enhancers can act locally or at long distances to exert transcriptional control over one or several target genes, often in a highly cell type- and/or context-specific manner[5]. The activity of enhancers is dependent upon the local chromatin state, which itself is dynamic throughout development. In general, chromatin at developmental enhancers is thought to transition between defined states that correlate with gene expression patterns.

At inactive enhancers, chromatin is tightly compacted by histone proteins, rendering it inaccessible to most transcription factors and leaving the enhancer unlikely to induce transcription of target genes. Contrastingly, at primed enhancers, displacement of histones allows transcription factors to access chromatin. This primed state correlates with a deposition of the H3K4me1 and H3K3me2 modifications at histones flanking the enhancer[6]. Although enhancers in this primed state do not promote transcription of target genes, they can be rapidly activated upon exposure to environmental stimuli. This induction yields an active enhancer which promotes expression of target genes. An active enhancer can be distinguished from a primed enhancer by the additional presence of the H3K27ac modification at flanking histones[6].

The chromatin state across the genome of a given cell constitutes the cellular epigenome[7]. Several epigenomic assays exist which allow for measurement of chromatin state and enhancer activity on a genome-wide basis. For example, mRNA sequencing (RNA-seq) allows for quantification of all genes expressed within a given cell. Additionally, the Assay for Transposase Accessible Chromatin with high-throughput sequencing (ATAC-seq)[8] provides locations in the genome in which histones have been displaced, leaving DNA accessible to binding by transcription factors. Chromatin immunoprecipitation with high-throughput sequencing[9] identifies

genomic regions bound by a given transcription factor or at which histones exhibit the presence of a given modification. Finally, Hi-C[10] identifies the 3D conformation of chromatin within a cell, providing evidence for potential target genes of enhancers. Although single-cell technologies have recently become available[11-14], by and large these assays remain relatively inefficient and require large numbers of input cells. Thus, the scalability of stem cell systems makes them ideal candidates for use in assays to profile the epigenome.

Diabetes

Diabetes is a chronic disease defined by dysregulation of blood glucose levels. The disease currently affects over 300 million people worldwide, with 90-95% of these cases consisting of T2D[15]. T2D is characterized by insulin resistance and dysfunction of insulin-secreting pancreatic beta cells. Genetic as well as lifestyle factors contribute to risk for T2D pathogenesis, although the exact causes remain largely unknown[16]. T2D can often be treated or reversed with a combination of drugs and lifestyle changes. Contrastingly, T1D results from autoimmune attack and destruction of a patient's beta cells. The disease is incurable, and patients are treated with a lifelong regimen of exogenous insulin delivery. Both genetic and environmental factors contribute to T1D pathogenesis[17]. However, like T2D, the exact mechanisms of disease pathogenesis are not well understood.

Even when properly treated, diabetes generally results in suboptimal control of blood glucose levels. This leaves patients at risk of developing severe complications such as cardiovascular issues, retinopathy, and neuropathy. Complications associated with diabetes cost an estimated $827 billion globally each year[15] and were the 8th leading cause of death in the U.S. during 2020[18]. A more complete understanding of pathogenic mechanisms for T2D and T1D can inform future treatment options or identify opportunities for therapeutic interventions.

hPSCs as a model for pancreatic development

Chromatin state within the epigenome is dynamic during development as cells integrate external signals and respond with appropriate changes in gene expression. However, the mechanisms of these dynamics are not well understood. For example, transition from an inactive chromatin state to a primed or activated chromatin state generally requires displacing histones at a given enhancer. The process by which this displacement occurs during development is the subject of debate. Certain classes of transcription factors known as pioneer factors have the demonstrated ability to displace histones and create regions of accessible chromatin *in vitro*[19]; however, the *in vivo* relevance of these factors is unclear. Conversely, upon withdrawal of a developmental signal, cells must attenuate gene expression and limit signal-responsiveness. While equally important in development as the acquisition of signal-responsiveness, the mechanisms of this phenomenon are less frequently studied.

We address these questions in chapters 1 and 2, respectively, of our work. In chapter 1 we examine the function of the FOXA family of pioneer transcription factors in priming and activating pancreatic enhancers during pancreatic lineage induction. In chapter 2, we identify a role for the histone demethylase LSD1 in acting at pancreatic enhancers to dampen responsiveness to retinoic acid signaling upon withdrawal of retinoic acid during pancreatic lineage induction. In both chapters, we utilize epigenomic assays to characterize regulatory regions in developing cells on a genome-wide scale. We also utilize gene editing-based perturbations to more thoroughly evaluate the function of specific transcription factors. Thus, these chapters of our work leverage the advantages of stem cell systems to investigate transitions of the epigenome during commitment to the pancreatic organ lineage in development.

hPSCs as a model to interrogate type 2 diabetes risk

Genetic components strongly influence disease pathogenesis of both T1D and T2D. Recent genome-wide association studies (GWAS) have identified multiple genetic variants that are associated with risk for both forms of diabetes[20,21]. However, most of these variants lie within non-coding regulatory elements with unknown target genes. Furthermore, the specific cell types in which these regulatory elements are active cannot be discerned through sequence alone.

In chapter 3 of this work, we integrate chromatin maps of active regulatory elements within pancreatic progenitors with T2D-associated variants identified through GWAS. In doing so, we identify progenitor-specific regulatory elements that harbor variants associated with disease occurrence. We then utilize gene editing within our stem cell system to link variant-containing enhancers to their respective target genes, and we validate a role for each of these target genes in pancreatic beta cell development using a zebrafish system. Altogether our results identify a developmental contribution to T2D risk and demonstrate the utility of stem cell systems for linking target genes to variant-containing regulatory elements, elucidating mechanisms of disease association.

hPSCs as a model of type 1 diabetes pathogenesis

Genetic risk for developing T1D correlates most closely with variants found at the major histocompatibility complex (MHC) gene locus[22]. Genes at this locus are involved in presentation of antigens to CD8+ T cells (MHC class I genes) or CD4+ T cells (MHC class II genes). Variants at the locus tend to be inherited in groups known as haplotypes. Among these haplotypes, the MHC class II HLA-DQ8 haplotype is most strongly associated with risk for developing T1D[22].

Structural studies of the MHC class II molecule encoded by the HLA-DQ8 haplotype have implicated a potentially pathogenic role of flexible peptide interactions that take place between this molecule and insulin produced by beta cells[23-26]. However, the context in which these interactions take place and the cell types involved remain unclear. Lack of access to human tissues during disease pathogenesis has precluded in-depth study of this process.

In chapter 4 of this work, we generate mature islet-like cells from iPCSs derived from donors harboring the HLA-DQ8 haplotype and coculture these cells with CD4+ T cell clones recognizing various beta cell-derived autoantigens. Specifically focusing on a role for beta cells in the direct stimulation of CD4+ T cells, our work demonstrates the power of stem cell systems in modeling patient-specific cell type interactions which may lead to disease pathogenesis. Future studies using the system established here will incorporate patient-derived immune cells for a fully autologous patient-specific co-culture system.

CHAPTER 1: SEQUENCE LOGIC AT ENHANCERS GOVERNS A DUAL MECHANISM OF ENDODERMAL ORGAN FATE INDUCTION BY FOXA PIONEER FACTORS

Abstract

FOXA pioneer transcription factors (TFs) associate with primed enhancers in endodermal organ precursors. Using a human stem cell model of pancreas differentiation, we here discover that only a subset of pancreatic enhancers is FOXA-primed, whereas the majority is unprimed and engages FOXA upon lineage induction. Primed enhancers are enriched for signal-dependent TF motifs and harbor abundant and strong FOXA motifs. Unprimed enhancers harbor fewer, more degenerate FOXA motifs, and FOXA recruitment to unprimed but not primed enhancers requires pancreatic TFs. Strengthening FOXA motifs at an unprimed enhancer near *NKX6.1* renders FOXA recruitment pancreatic TF-independent, induces priming, and broadens the *NKX6.1* expression domain. We make analogous observations about FOXA binding during hepatic and lung development. Our findings suggest a dual role for FOXA in endodermal organ development: First, FOXA facilitate signal-dependent lineage initiation via enhancer priming, and second, FOXA enforce organ cell type-specific gene expression via indirect recruitment by lineage-specific TFs.

Introduction

The pancreas, liver, and lung develop from the foregut endoderm in response to local signaling cues that specify lineage identity by inducing organ-specific gene expression. The competence of organ lineage precursors to activate lineage-specific genes in response to inductive signals is acquired during endoderm development[27,28]. Coincident with the acquisition of competence, the transcription factors (TFs) FOXA1 and FOXA2 (henceforth abbreviated FOXA1/2) are recruited to enhancers of foregut-derived organ lineages, leading to a gain in chromatin accessibility and H3K4me1 deposition[27,29,30], a phenomenon referred to as enhancer priming. Thus, current evidence suggests that FOXA1/2's role in endodermal organ development

is to render foregut endoderm competent to activate organ-specific genes by broadly priming pancreas-, liver-, and lung-specific enhancers before organ-inductive signals trigger enhancer activation. Consistent with this model, studies in model organisms and human pluripotent stem cell (hPSC)-based differentiation systems have shown a requirement for FOXA1/2 in pancreas, liver, and lung development, with the two FOXA TFs functioning in a partially or fully redundant manner[29-32]. However, whether chromatin priming is the only mechanism by which FOXA TFs control endodermal organ development is unknown.

The mechanisms by which FOXA TFs engage with and open chromatin have been the subject of debate. *In vitro* experiments have shown that FOXA TFs possess pioneering activity, which refers to the specific ability of a TF to engage target sites on nucleosomal DNA and to remodel such regions to increase chromatin accessibility[33-35]. Through their chromatin remodeling activity, FOXA TFs facilitate subsequent binding of other TFs and co-factors that further modify chromatin state and initiate gene expression[35-40]. However, despite their ability to access target sites in closed chromatin *in vitro*, binding site selection of FOXA and other pioneer TFs in cellular contexts has been shown to depend on additional features, such as the local chromatin landscape[41], presence of cooperative binding partners[42,43], and strength of the binding motif[43-45]. For example, steroid receptor activation in breast cancer cell lines induces FOXA1 recruitment to sites with degenerate FOXA1 binding motifs[44,46], exemplifying heterogeneity in FOXA target site engagement. The determinants that underlie FOXA binding site selection and FOXA-mediated enhancer priming during cellular transitions of development remain to be explored.

Here, we sought to determine the specific mechanisms that underlie the regulation of endodermal organ development by FOXA TFs. To this end, we mapped FOXA1/2 genomic association with pancreas-specific enhancers throughout a time course of hPSC differentiation into pancreas. Surprisingly, only a minority of pancreas-specific enhancers are FOXA1/2-bound prior to lineage induction and exhibit priming, whereas the majority engage FOXA1/2 concomitant with pancreas induction. Compared to unprimed enhancers, primed enhancers contain DNA

sequences more closely matching FOXA consensus motifs and harbor additional sequence motifs for signal-dependent TFs. By contrast, unprimed enhancers contain degenerate and fewer FOXA motifs, are enriched for motifs of lineage-specific TFs, and depend on the pancreas-specific TF PDX1 for FOXA1/2 recruitment. We further show that CRISPR/Cas9-mediated optimization of FOXA motifs in an unprimed enhancer near the pancreatic TF *NKX6.1* is sufficient to redefine patterns of FOXA binding and to broaden *NKX6.1* expression within the pancreatic progenitor domain, suggesting that FOXA motif strength is relevant for fine-tuning developmental gene expression. In-depth analysis of FOXA binding during hPSC differentiation toward hepatocytes and lung alveolospheres revealed similar patterns of FOXA binding and sequence logic at FOXA-bound enhancers. Our findings show that FOXA1/2 regulate foregut organ development through two distinct and complementary mechanisms: priming of a small subset of organ-specific enhancers before lineage induction and activation of a larger cohort of enhancers through cooperative binding with organ lineage-specific TFs. We propose that priming of a small enhancer subset permits precise spatial and temporal regulation of organ induction by lineage-inductive signals, whereas cooperative FOXA binding with lineage-specific TFs ensures cell type specificity of gene expression, providing a safeguard against broad activation of alternative lineage programs during developmental transitions.

Results

FOXA1 and FOXA2 are necessary for pancreatic lineage induction

To investigate the role of FOXA1/2 in pancreas development, we employed a hPSC differentiation protocol in which cells transition stepwise to the pancreatic fate through sequential exposure to developmental signaling cues (**Fig. 1.1a**). The pancreatic lineage is induced by retinoic acid from gut tube (GT) intermediates, resulting in expression of the pancreatic markers PDX1 in early pancreatic progenitors (PP1) and NKX6.1 in late pancreatic progenitors (PP2). FOXA1 and FOXA2 were expressed from the definitive endoderm (DE) stage onwards

(**Supplementary Figure 1.1a,b**), and levels of *FOXA1* and *FOXA2* were similar in GT, PP1, and PP2 (**Supplementary Figure 1.1a**).

To determine a possible requirement for FOXA1 and FOXA2 in pancreas development, we deleted *FOXA1* or *FOXA2* in human embryonic stem cells (hESCs) (**Fig. 1.1a** and **Supplementary Figure 1.1c,d**) and differentiated control, *FOXA1*$^{-/-}$, and *FOXA2*$^{-/-}$ hESC lines into pancreatic progenitors. Analysis of PDX1 and NKX6.1 expression revealed a requirement for FOXA2 but not FOXA1 for pancreatic lineage induction (**Fig. 1.1b, Supplementary Figure 1.2**), consistent with recent findings[29]. The presence of residual PDX1^{+} and NKX6.1^{+} cells and increased *FOXA1* levels in *FOXA2*$^{-/-}$ pancreatic progenitors (**Fig. 1.1b,c**) suggests FOXA1 partially compensates for FOXA2 deficiency. Therefore, we generated *FOXA1*$^{-/-}$;*FOXA2*$^{-/-}$ (*FOXA1/2*$^{-/-}$) hESC lines (**Supplementary Figure 1.1e**) and analyzed phenotypes at the DE, GT, and PP2 stages. At the DE and GT stages, similar numbers of *FOXA1/2*$^{-/-}$ and control cells expressed the DE marker SOX17 and GT marker HNF1B, respectively (**Supplementary Figure 1.1f,g)**. In contrast, pancreas induction was blocked in *FOXA1/2*$^{-/-}$ cells, as evidenced by an almost complete absence of PDX1^{+} and NKX6.1^{+} cells, reduced expression of early pancreatic TFs, and down-regulation (≥ 2 fold change, FDR ≤ 0.05) of genes associated with pancreas-specific biological processes (**Fig. 1.1d-f**). Principal component analysis (PCA) of transcriptome data further confirmed that *FOXA1/2*$^{-/-}$ and control cells were more similar at the GT stage than at the PP2 stage (**Fig. 1.1g**). Together, these findings show that FOXA1 and FOXA2 control pancreatic lineage induction from gut tube lineage intermediates in a partially redundant manner.

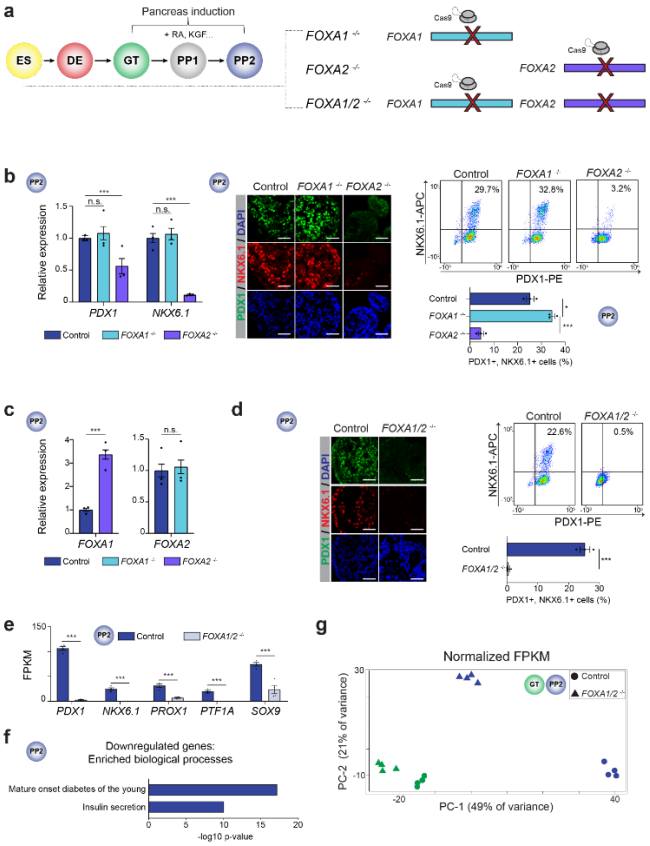

Figure 1.1. Partially redundant requirement for FOXA1 and FOXA2 in pancreatic lineage induction. (**a**) Schematic of stepwise pancreatic differentiation protocol from hESCs (ES): definitive endoderm (DE), primitive gut tube (GT), early pancreatic progenitor cells (PP1), and late pancreatic progenitor cells (PP2), with indicated genetic modifications in ES. Select growth factors for pancreatic lineage induction are indicated. RA, retinoic acid; KGF, keratinocyte growth factor. (**b**) qPCR analysis of *PDX1* and *NKX6.1* (left), representative immunofluorescent staining (middle), and flow cytometry analysis and quantification of PDX1$^+$ and NKX6.1$^+$ cells (right) in control, *FOXA1*$^{-/-}$ and *FOXA2*$^{-/-}$ PP2 cells (n = 3 independent differentiations; qPCR: P = 0.493, 0.590, 3.12 x 10^{-3}, and < 1.00 x 10^{-6} for *PDX1* and *NKX6.1* in control compared to *FOXA1*$^{-/-}$ and *FOXA2*$^{-/-}$ PP2 cells, respectively; flow cytometry: P = 1.15 x 10^{-2} and 7.00 x 10^{-4} in control compared to *FOXA1*$^{-/-}$ and *FOXA2*$^{-/-}$ PP2 cells, respectively; student's t-test, 2-sided; n.s., not significant). (**c**) qPCR analysis of *FOXA1* and *FOXA2* in control, *FOXA1*$^{-/-}$ and *FOXA2*$^{-/-}$ PP2 cells (n = 3 independent differentiations; P = < 1.00 x 10^{-6} and 0.700 for *FOXA1* and *FOXA2* in control compared to *FOXA1*$^{-/-}$ and *FOXA2*$^{-/-}$ PP2 cells, respectively; student's t-test, 2-sided). (**d**) Representative immunofluorescent staining (left) and flow cytometry analysis and quantification (right) of PDX1$^+$ and NKX6.1$^+$ cells in control and *FOXA1/2*$^{-/-}$ PP2 cells. (n = 3 independent differentiations; P = 2.6 x 10^{-3} in control compared to *FOXA1/2*$^{-/-}$ PP2 cells; student's t-test, 2-sided). (**e**) mRNA expression levels of pancreatic transcription factors determined by RNA-seq in control and *FOXA1/2*$^{-/-}$ PP2 cells (n = 4 independent differentiations; P adj. = 1.08 x 10^{-42}, 2.56 x 10^{-12}, 4.93 x 10^{-20}, 1.00 x 10^{-49}, and 2.82 x 10^{-4} for *PDX1*, *NKX6.1*, *PROX1*, *PTF1A*, and *SOX9*, respectively; DESeq2; FPKM, fragments per kilobase per million fragments mapped). (**f**) Enriched gene ontology terms of 2833 downregulated genes (≥ 2-fold decrease, P adj. < 0.05) in *FOXA1/2*$^{-/-}$ compared to control PP2 cells. (**g**) Principal component analysis showing variance in total normalized transcriptome between control and *FOXA1/2*$^{-/-}$ cells in GT and PP2. Each plotted point represents one biological replicate.

For all qPCR, each plotted point represents the average of three technical replicates. For all immunofluorescence, representative images are shown from n ≥ 2 independent differentiations. Scale bars, 50 μm. For all flow cytometry analyses, representative plots are shown from n = 3 independent differentiations.

FOXA transcription factors exhibit two temporal patterns of recruitment to pancreatic enhancers

To identify transcriptional targets of FOXA1/2 during pancreatic lineage induction, we mapped FOXA1/2 binding sites at the GT and PP2 stages. Consistent with the partial functional redundancy between FOXA1 and FOXA2 (**Fig. 1.1b-d**), FOXA1 and FOXA2 binding sites were highly correlated at both stages (**Supplementary Figure 1.3a**). FOXA1/2 mostly bound to distal sites (> 2.5 kb from TSS; **Supplementary Figure 1.3b**), suggesting regulation of enhancers by FOXA1/2. To test this, we defined GT and PP2 enhancers as distal H3K27ac peaks (> 2.5 kb from TSS) and compared enhancer activity based on H3K27ac signal in control and *FOXA1/2*$^{-/-}$ cells at the GT and the PP2 stages. Like gene expression (**Fig. 1.1g**), H3K27ac profiles in *FOXA1/2*$^{-/-}$ and control cells differed more substantially at the PP2 than at the GT stage (**Supplementary Figure 1.3c**), showing that *FOXA1/2* deletion has broad impact on regulation of enhancer activity during the GT to PP2 transition.

To investigate specific mechanisms by which FOXA1/2 mediates pancreatic lineage induction, we identified all FOXA1/2-bound pancreatic enhancers that are activated upon pancreatic lineage induction. To this end, we first identified enhancers that exhibited a ≥ 2-fold increase in H3K27ac signal from the GT to the PP2 stage (2574 enhancers, hereafter referred to as pancreatic enhancers; **Supplementary Figure 1.3d,e**). As expected, genes near these enhancers were predicted to regulate biological processes associated with pancreas development. Second, we analyzed FOXA1/2 binding at these pancreatic enhancers, revealing that 72% were FOXA1/2-bound at the PP2 stage (**Supplementary Figure 1.3f**). Consistent with prior reports[27,29], we observed FOXA1/2 occupancy at the GT stage preceding pancreatic lineage induction. Surprisingly, however, the percentage of pancreatic enhancers bound by FOXA1/2 was significantly lower at the GT compared to the PP2 stage, implying that not all pancreatic enhancers engage FOXA1/2 before lineage induction. To comprehensively characterize temporal patterns of FOXA1/2 recruitment, we identified all pancreatic enhancers with FOXA1 or FOXA2

binding at the GT and/or PP2 stages and quantified FOXA1/2 ChIP-seq signal at these sites (**Fig. 1.2a**). We observed three distinct patterns of FOXA1/2 occupancy: class I enhancers (561) were bound by FOXA1/2 at both the GT and PP2 stages, class II enhancers (1422) were FOXA1/2-bound only at the PP2 stage, and the overall small group of class III enhancers (118) was FOXA1/2-bound only at the GT stage (**Fig. 1.2a**). Analysis of H3K27ac signal intensity at the GT and PP2 stages showed similar patterns of H3K27ac signal at all enhancers (**Fig. 1.2b**), suggesting that enhancers of all classes are mostly inactive at the GT stage and become activated during pancreatic lineage induction. Activation of enhancers of all classes during the GT to PP2 transition was dependent on FOXA1/2 (**Fig. 1.2c** and **Supplementary Figure 1.3g**). Since the predominant patterns were either maintenance of FOXA1/2 binding (class I) or *de novo* FOXA1/2 occupancy (class II) after pancreas induction, we excluded class III enhancers from further analyses. We identified examples of both class I and class II enhancers in proximity to gene bodies of pancreatic lineage-determining TFs, such as *PDX1, HNF1B*, *NKX6.1*, and *MNX1* (**Fig. 1.2d**). Consistent with the H3K27ac pattern, the *PDX1* class I enhancer and the *NKX6.1* class II enhancer are both inactive in GT and active in PP2 in enhancer reporter assays[27]. Together, this analysis shows that FOXA1/2 recruitment to pancreatic enhancers precedes lineage induction at only a small subset of enhancers, while FOXA1/2 recruitment to most pancreatic enhancers coincides with lineage induction (**Fig. 1.2e**).

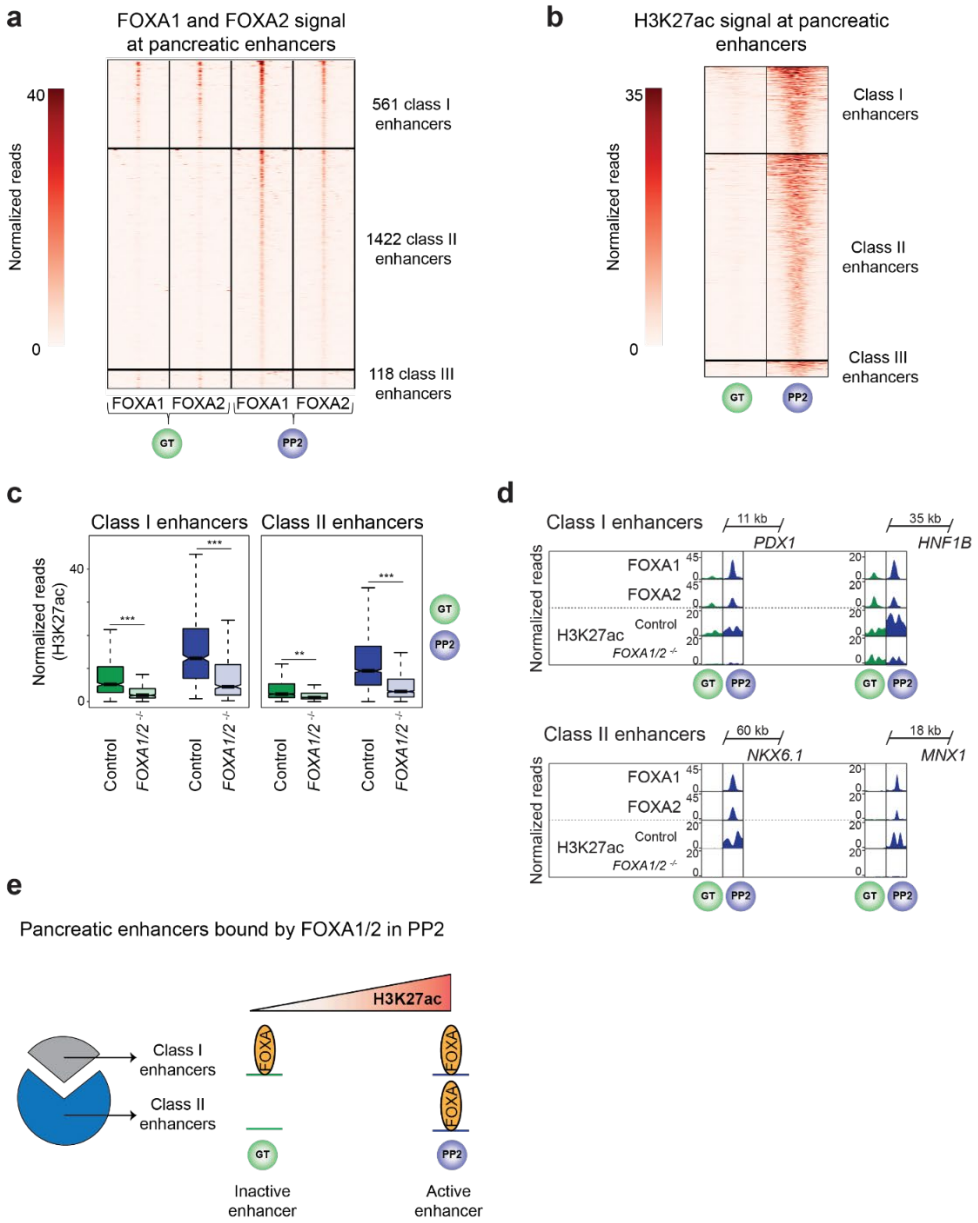

Figure 1.2. Two distinct temporal patterns of FOXA1 and FOXA2 binding to pancreatic enhancers. (**a** and **b**) Heatmaps showing density of FOXA1 and FOXA2 ChIP-seq reads (**a**) and H3K27ac ChIP-seq reads (**b**) at pancreatic enhancers in GT and PP2. Heatmaps are centered on FOXA1, FOXA2, and H3K27ac peaks, respectively, and span 5 kb. Pancreatic enhancers are classified based on temporal pattern of FOXA1 and FOXA2 occupancy. (**c**) Box plots of H3K27ac ChIP-seq counts at class I and class II pancreatic enhancers in control and *FOXA1/2*$^{-/-}$ GT and PP2 cells (P = < 2.2 x 10^{-16}, < 2.2 x 10^{-16}, 0.009, and < 2.2 x 10^{-16} for control versus *FOXA1/2*$^{-/-}$ at class I enhancers in GT, class I enhancers in PP2, class II enhancers in GT, and class II enhancers in PP2, respectively; Wilcoxon rank sum test, 2-sided). (**d**) Genome browser snapshots showing FOXA1, FOXA2, and H3K27ac ChIP-seq signal at class I pancreatic enhancers near PDX1 and HNF1B and class II pancreatic enhancers near NKX6.1 and MNX1 in GT and PP2. Approximate distance between enhancer and gene body is indicated. (**e**) Schematic illustrating the identified pattern of FOXA1/2 occupancy at pancreatic enhancers. All ChIP-seq experiments, n = 2 replicates from independent differentiations.

Primed and unprimed pancreatic enhancers reside in distinct regulatory domains

Given early recruitment of FOXA1/2 to class I but not class II enhancers, we hypothesized that the two classes could differ in their temporal pattern of gain in chromatin accessibility and H3K4me1 deposition, predicting that early FOXA1/2 occupancy at class I enhancers would lead to chromatin priming. As predicted, class I enhancers exhibited open chromatin and H3K4me1 deposition at the GT stage (**Fig. 1.3a** and **Supplementary Figure 1.4a,b**). By contrast, class II enhancers acquired these features largely with pancreatic lineage induction (**Fig. 1.3a** and **Supplementary Figure 1.4a,b**), identifying primed chromatin as a feature of class I enhancers. Although a subset of class II enhancers was marked by H3K4me1 at the GT stage, this population comprised the minority of class II enhancers (**Supplementary Figure 1.4c**). At both class I and class II enhancers, H3K4me1 deposition and gain in chromatin accessibility during lineage induction was FOXA1/2-dependent (**Fig. 1.3b** and **Supplementary Figure 1.4b**), demonstrating that FOXA1/2 are necessary for chromatin remodeling at both classes of enhancers.

We next sought to determine whether class I and class II enhancers function together within larger regions of active chromatin such as super-enhancers[47], or whether they reside in distinct regulatory domains. To distinguish between these possibilities, we defined 167 super-enhancers among the 2574 pancreatic enhancers identified in Supplementary Figure 3d (**Supplementary Figure 1.4d**) and found that 160 (96%) were FOXA1/2-bound at the PP2 stage (**Supplementary Figure 1.4e**). Analysis of overlap between class I or class II enhancers and FOXA-bound super-enhancers revealed that most FOXA-bound super-enhancers (76%) contained either class I or class II enhancers but not both (**Fig. 1.3c**). Furthermore, we analyzed Hi-C datasets produced from PP2 stage cells and found that class I and class II enhancers were mostly located in non-overlapping 3D chromatin loops (**Fig. 1.3d**). This evidence indicates that class I and class II enhancers reside largely within distinct gene regulatory domains and therefore likely function independently.

To identify target genes of class I and class II enhancers, we assigned enhancers to their nearest expressed gene at the PP2 stage and validated predictions by showing regulation of these genes by FOXA1/2 (**Supplementary Figure 1.4f**). Consistent with their location in distinct regulatory domains (**Fig. 1.3c,d**), class I and class II enhancers mostly associated with distinct genes, including pancreatic lineage-determining TFs (**Fig. 1.3e**). Of note, gene ontology analysis of genes regulated by class I compared to class II enhancers revealed roles for class I enhancer-associated genes in cellular signal transduction pathways (**Supplementary Figure 1.4g**), whereas no comparative enrichment of specific gene ontology terms was observed for class II enhancer-associated genes. Together, these results suggest that two distinct mechanisms establish the pancreatic gene expression program: a subset of pancreatic genes is regulated by enhancers that undergo FOXA1/2-mediated chromatin priming at the gut tube stage, whereas most pancreatic genes are regulated by enhancers that are unprimed prior to pancreatic lineage induction, and to which FOXA1/2 are recruited upon lineage induction (**Fig. 1.3f**).

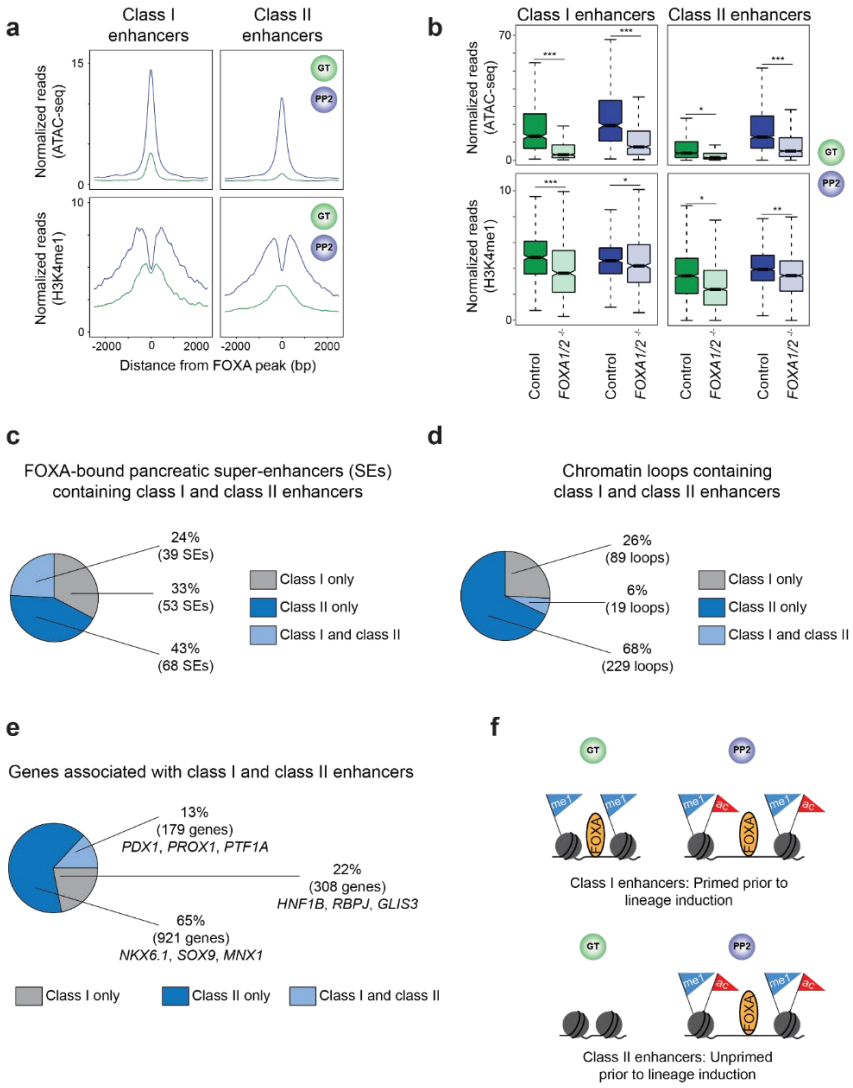

Figure 1.3. Class I and class II pancreatic enhancers largely map to distinct gene regulatory elements. (**a**) Tag density plots for class I and class II pancreatic enhancers displaying ATAC-seq (top) and H3K4me1 ChIP-seq (bottom) read density in GT and PP2. Plots are centered on FOXA1/2 peaks and span 5 kb. (**b**) Box plots of ATAC-seq (top) and H3K4me1 ChIP-seq (bottom) counts at class I and class II pancreatic enhancers in GT and PP2 for control and *FOXA1/2*$^{-/-}$ cells (P = < 2.2 x 10^{-16}, < 2.2 x 10^{-16}, 0.01, and < 2.2 x 10^{-16} for control versus *FOXA1/2*$^{-/-}$ of ATAC-seq signal at class I in GT, class I in PP2, class II in GT, and class II in PP2, respectively. P = < 2.2 x 10^{-16}, 0.01, 0.02, and 0.01 for control versus *FOXA1/2*$^{-/-}$ of H3K4me1 signal at class I in GT, class I in PP2, class II in GT, and class II in PP2, respectively; Wilcoxon rank sum test, 2-sided). (**c**) Percentage of FOXA1- and/or FOXA2-bound pancreatic super-enhancers (SEs) in PP2 containing only class I, only class II, or both class I and class II enhancers. (**d**) Percentage of chromatin loop anchors in PP2 containing only class I, only class II, or both class I and class II enhancers. (**e**) Percentage of genes associated with only class I, only class II, or both class I and class II enhancers. Target genes were assigned to enhancers based on nearest TSS of expressed genes (fragments per kilobase per million fragments mapped (FPKM) ≥ 1) in PP2. (**f**) Schematic illustrating FOXA1/2 occupancy, chromatin accessibility, and presence of H3K4me1 and H3K27ac at class I and class II enhancers in GT and PP2. All ChIP-seq and ATAC-seq experiments, n = 2 replicates from independent differentiations.

Distinct DNA sequence motifs at primed and unprimed pancreatic enhancers

We next investigated mechanisms that could explain the observed temporal differences in FOXA1/2 binding to class I (primed) and class II (unprimed) pancreatic enhancers. To test whether differences in DNA sequence could provide an explanation, we conducted *de novo* motif analysis to identify motifs enriched at class I enhancers against a background of class II enhancers. Class I enhancers were enriched for FOXA motifs and motifs for several signal-dependent TFs, including the ETS family TFs GABPA and SPDEF, the downstream effector of Hippo signaling TEAD, and the retinoic acid receptor RXRA (**Fig. 1.4a**). Work in model organisms has identified critical roles for ETS TFs as well as Hippo and retinoic acid signaling in early pancreatic development[48-51], suggesting that pancreatic lineage-inductive signals are read at class I enhancers by partnering of FOXA1/2 with signal-dependent TFs. ChIP-seq analysis for RXR confirmed preferential RXR binding to class I compared to class II enhancers at the PP1 stage (**Fig. 1.4b**). Class I enhancers were also enriched for GATA TF motifs (**Fig. 1.4a**) and a higher percentage of class I than class II enhancers bound GATA4 and GATA6 at the GT stage (**Fig. 1.4b**). Given that GATA TFs cooperatively bind with FOXA1/2 to DNA[43], GATA4/6 could facilitate FOXA1/2 recruitment to a subset of class I enhancers prior to pancreas induction.

Since FOXA1/2 binding to class I enhancers precedes binding to class II enhancers (**Fig. 1.2a**) and FOXA motifs are enriched at class I compared to class II enhancers (**Fig. 1.4a**), we postulated that different mechanisms could underlie FOXA1/2 recruitment to the two classes of enhancers. Binding site selection of pioneer TFs such as FOXA1/2 has been shown to depend on motif abundance, strength, and position[43-45,52]. Therefore, we analyzed FOXA motifs at class I and class II enhancers for these features. To determine abundance and strength of FOXA motifs, we selected position-weighted matrices (PWMs) corresponding to three FOXA1 and three FOXA2 motifs from JASPAR[53] (**Supplementary Figure 1.5a**), identified occurrences of each motif at class I and class II enhancers, and generated a log-odds score to measure how closely the DNA sequence at each identified motif occurrence matched the PWM. Class I enhancers were

significantly enriched for occurrences of all six FOXA motifs compared to class II enhancers (**Fig. 1.4c**). Furthermore, three of the FOXA motifs had significantly higher log-odds scores at class I than class II enhancer occurrences (MA0047.2, MA0148.1, and MA0148.3; $P = 1.54 \times 10^{-2}$, 1.10×10^{-3}, and 1.03×10^{-2}, respectively; Wilcoxon rank sum test). Thus, class II enhancers contain more degenerate and fewer FOXA motifs compared to class I enhancers. We additionally examined the positioning of FOXA motifs relative to open chromatin by identifying regions of greatest chromatin accessibility at class I and class II enhancers in PP2 stage cells (n = 531 and n = 1257 ATAC-seq summits in class I and class II enhancers, respectively) and determining enrichment of each FOXA motif at these regions. Occurrence of all FOXA motifs was enriched at ATAC-seq summits at class I compared to class II enhancers (**Fig. 1.4d** and **Supplementary Figure 1.5b**), indicating that regions of greatest chromatin accessibility at class I enhancers are more likely to harbor FOXA motifs. ATAC-seq footprinting analysis further revealed a higher occurrence of FOXA footprints at class I than at class II enhancers (**Fig. 1.4e**), indicative of either longer FOXA1/2 DNA residence times or more direct interaction of FOXA1/2 with DNA at class I enhancers[54]. Together, this analysis reveals features of FOXA motifs at class I pancreatic enhancers previously associated with canonical FOXA1/2 pioneer TF activity[43,44].

To further elucidate differences in mechanisms of FOXA recruitment to class I and class II enhancers, we identified *de novo* motifs enriched at class II enhancers against a background of class I enhancers. Here, we observed enrichment of motifs for pancreatic lineage-determining TFs, such as ONECUT (HNF6), SOX (SOX9), HNF1B, and PDX1 (**Fig. 1.4a**), which sharply increased in expression during pancreatic lineage induction (**Supplementary Figure 1.5c**). To determine whether these TFs exhibit preferential binding to class II enhancers, we mapped HNF6, PDX1, and SOX9 binding sites genome-wide at the PP2 stage (**Fig. 1.4b** and **Supplementary Figure 1.5d**). Overall, we found that similar percentages of class I and class II enhancers were bound by HNF6, PDX1, and SOX9 at the PP2 stage (**Fig. 1.4b**). To determine whether the difference in sequence motif enrichment between class I and class II enhancers is also observed

when focusing on enhancers bound by a specific TF, we analyzed motifs at HNF6-, PDX1-, or SOX9-bound enhancers. Still, class I enhancers were enriched for FOXA and class II enhancers for ONECUT (HNF6), PDX1, and SOX motifs (**Supplementary Figure 1.5e**). Thus, despite differences in DNA sequence motifs between primed (class I) and unprimed (class II) enhancers, both classes of enhancers are occupied by FOXA1/2 as well as pancreatic lineage-determining TFs after pancreatic lineage induction.

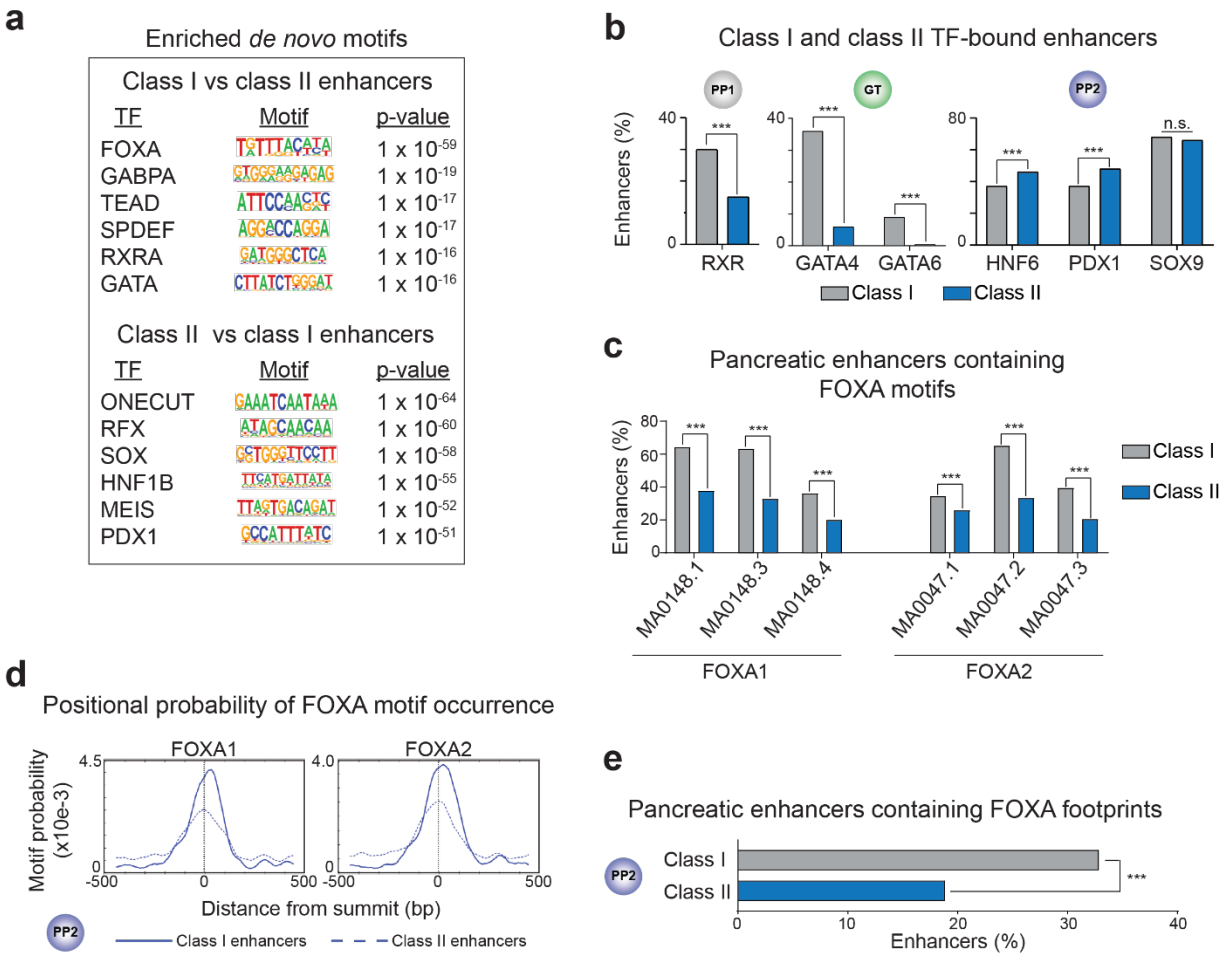

Figure 1.4. FOXA1/2 binding sites at class I and class II pancreatic enhancers differ in DNA sequence. (**a**) Enriched *de novo* transcription factor (TF) binding motifs at class I against a background of class II pancreatic enhancers and vice versa. Fisher's exact test, 1-sided, corrected for multiple comparisons. (**b**) Percentage of class I and class II enhancers overlapping RXR ChIP-seq peaks in PP1; GATA4 and GATA6 ChIP-seq peaks in GT; and HNF6, PDX1, and SOX9 ChIP-seq peaks (within 100 bp from peak) in PP2 (P = 8.27 x 10^{-14}, < 2.2 x 10^{-16}, < 2.2 x 10^{-16}, 3.52 x 10^{-4}, 1.01 x 10^{-5}, and 0.40 for comparisons of overlap with binding sites for RXR, GATA4, GATA6, HNF6, PDX1, and SOX9, respectively; Fisher's exact test, 2-sided). (**c**) Percentage of class I and class II enhancers with at least one occurrence of selected FOXA1 and FOXA2 motifs (P = < 2.2 x 10^{-16}, < 2.2 x 10^{-16}, 1.76 x 10^{-13}, 1.61 x 10^{-4}, < 2.2 x 10^{-16}, and < 2.2 x 10^{-16} for comparisons of occurrences of MA0148.1, MA0148.3, MA0148.4, MA0047.1, MA0047.2, and MA0047.3, respectively. Fisher's exact test, 2-sided). (**d**) Probability (motif occurrence per base pair) of FOXA1 (MA0148.3) and FOXA2 (MA0047.2) motifs relative to ATAC-seq peak summits at class I (solid line) and class II (dashed line) enhancers. ATAC-seq peak summits at class I enhancers are enriched for occurrences of MA0148.3 (P = 2.1 x 10^{-14}; Fisher's exact test, 1-sided) and MA0047.2 (P = 6.8 x 10^{-14}) compared to summits at class II enhancers. (**e**) Percentage of class I and class II enhancers containing FOXA TF ATAC-seq footprints in PP2 (P = 1.01 x 10^{-10} for comparison of class I and class II enhancers; Fisher's exact test, 2-sided). All ChIP-seq experiments, n = 2 replicates from independent differentiations.

FOXA1/2 binding to a subset of unprimed enhancers depends on PDX1

Since motifs for pancreatic lineage-determining TFs, such as PDX1, were enriched at class II compared to class I enhancers (**Fig. 1.4a**), we hypothesized that FOXA1/2 recruitment to class II enhancers could require cooperativity with lineage-determining TFs. To test this, we analyzed FOXA1/2 binding, chromatin accessibility, and H3K27ac signal in *PDX1*-deficient pancreatic progenitors (**Fig. 1.5a** and **Supplementary Figure 1.6a**). Focusing on PDX1-bound enhancers (n = 205 class I enhancers and 682 class II enhancers), we found that loss of *PDX1* reduced FOXA1/2 binding to a greater extent at class II than class I enhancers (**Fig. 1.5b** and **Supplementary Figure 1.6b,c**), exemplified by class I enhancers near *PDX1* and *HNF1B*, and class II enhancers near *NKX6.1* and *MNX1* (**Fig. 1.5c**). In total, 23% of PDX1-bound class II enhancers exhibited a significant loss (≥ 2-fold decrease, *P*. adj. < 0.05) in FOXA1/2 ChIP-seq signal after *PDX1* knock-down compared to only 3% of PDX1-bound class I enhancers (**Supplementary Figure 1.6b**). Furthermore, PDX1-bound class II enhancers showed greater loss of FOXA1/2 signal than PDX1-bound class I enhancers (**Supplementary Figure 1.6c**). Given substantial overlap between binding sites for pancreatic lineage-determining TFs (**Supplementary Figure 1.5d**), it is possible that other TFs recruit FOXA1/2 to PDX1-bound class II enhancers where FOXA1/2 occupancy is not significantly affected. Loss of *PDX1* led to a significant reduction in ATAC-seq and H3K27ac signal at both class I and class II enhancers (**Supplementary Figure 1.6d**), showing that full acquisition of chromatin accessibility and enhancer activation during pancreas induction require PDX1 at primed and unprimed enhancers.

Collectively, our findings show that despite similar mechanisms for their activation, primed and unprimed pancreatic enhancers differ in sequence logic and mechanism of FOXA1/2 recruitment (**Fig. 1.5d**). Primed enhancers have abundant and strong FOXA motifs, and FOXA1/2 are recruited to primed enhancers prior to pancreatic lineage induction largely independent of the pancreatic TF PDX1. By contrast, unprimed enhancers have fewer and weaker FOXA motifs, and a proportion of unprimed enhancers requires PDX1 for FOXA1/2 recruitment.

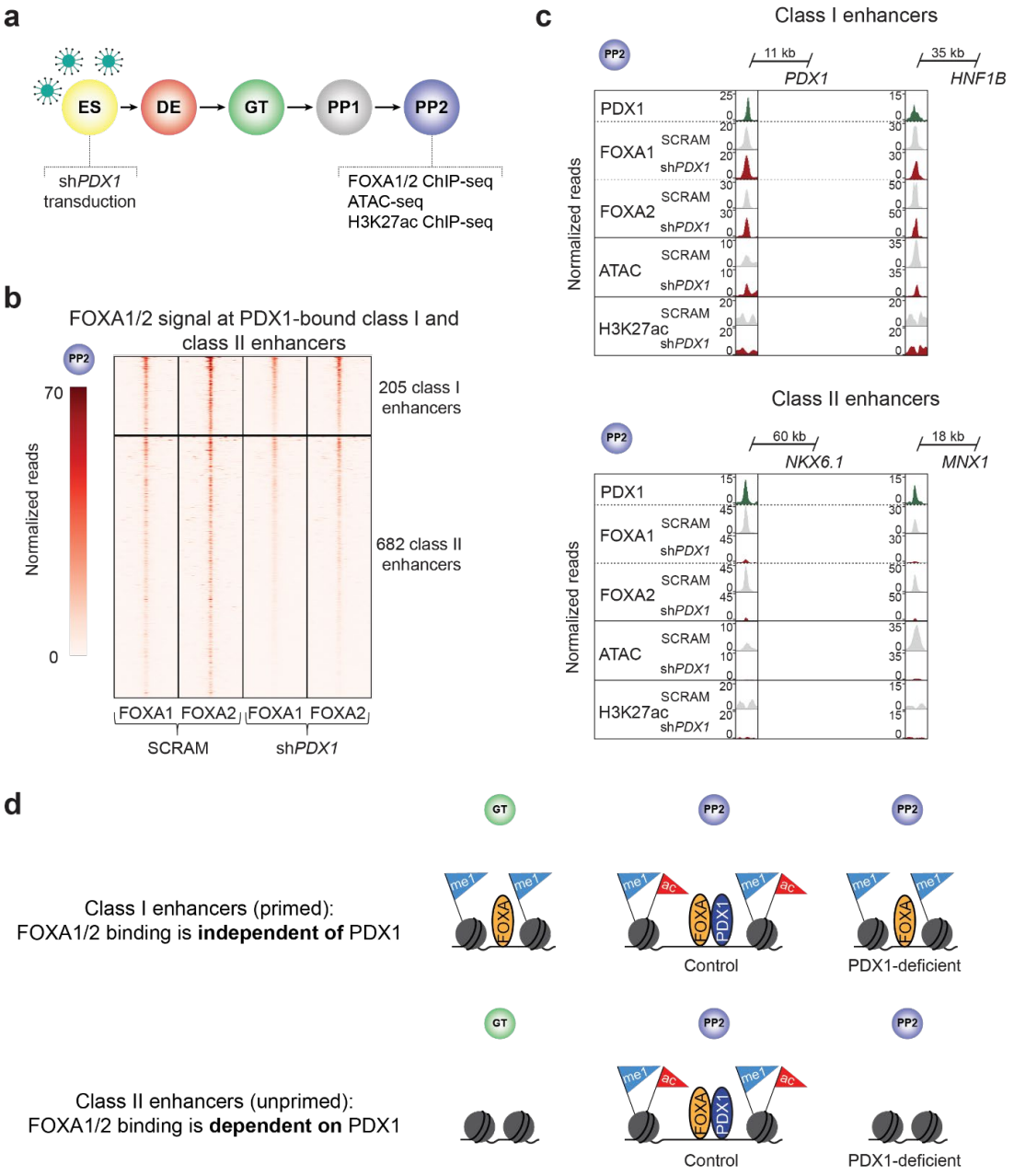

Figure 1.5. FOXA1/2 binding at class II enhancers is dependent on PDX1. (**a**) Schematic of experimental design for *PDX1* knock-down in hESCs and subsequent differentiation into PP2 stage pancreatic progenitors. (**b**) Heatmap showing density of FOXA1 and FOXA2 ChIP-seq reads at PDX1-bound class I and class II pancreatic enhancers in hESCs transduced with scrambled control (SCRAM) or *PDX1* shRNA (sh*PDX1*) in PP2. Heatmap is centered on FOXA1 and FOXA2 peaks, respectively, and spans 5 kb. (**c**) Genome browser snapshots showing PDX1, FOXA1, and FOXA2 ChIP-seq, ATAC-seq, and H3K27ac ChIP-seq signal at class I enhancers near *PDX1* and *HNF1B* and class II enhancers near *NKX6.1* and *MNX1* in PP2. Approximate distance between enhancer and gene body is indicated. (**d**) Schematic illustrating distinct modes of FOXA TF recruitment at class I and class II pancreatic enhancers. FOXA1/2 recruitment depends on the lineage-determining TF PDX1 at class II enhancers. Both enhancer classes require PDX1 or activation. All ChIP-seq and ATAC-seq experiments, n = 2 replicates from independent differentiations.

Altering FOXA motif strength redefines temporal FOXA1/2 binding patterns

We next sought to determine the extent to which the timing and mechanism of FOXA1/2 recruitment are solely dependent on DNA sequence. Since stronger FOXA motifs are a characteristic of class I enhancers, we tested this by optimizing FOXA motifs at a class II enhancer via CRISPR-Cas9 genome editing and mapping FOXA1/2 binding. For this we selected an unprimed class II enhancer near *NKX6.1* for editing in hESCs. This enhancer lacks FOXA1/2 binding (**Fig. 1.2d**), accessible chromatin **(Supplementary Figure 1.4b**), and H3K4me1 signal (**Supplementary Figure 1.4b**) prior to pancreas induction. Furthermore, in the absence of PDX1, FOXA1/2 do not bind to this enhancer (**Fig. 1.5c**). Examination of the *NKX6.1* enhancer revealed four degenerate FOXA motifs surrounding the ATAC-seq summit (**Fig. 1.6a**). We altered six base pairs within the enhancer to strengthen the FOXA motifs (referred to as motif optimized) (**Fig. 1.6a**). Optimizing FOXA motifs resulted in FOXA1/2 recruitment to the *NKX6.1* enhancer at the GT stage prior to pancreas induction (**Fig. 1.6b**). Early FOXA1/2 recruitment was accompanied by H3K4me1 but not H3K27ac deposition at the GT stage (**Fig. 1.6b**), supporting that FOXA1/2 prime enhancers prior to activation. Thus, optimization of FOXA binding motifs is sufficient to convert an unprimed class II enhancer into a primed class I enhancer.

Optimizing FOXA motifs broadens the domain of target gene expression

To define the relationship between FOXA motif strength and *NKX6.1* target gene expression, we conducted single cell RNA-sequencing of PP2 cells from control and motif optimized cell lines. Consistent with prior studies[55], we observed a population of multipotent pancreatic progenitor cells expressing high levels of pancreatic lineage-determining TFs (e.g., *PDX1*, *HNF6*, *SOX9*, and *PTF1A*) as well as a population of early endocrine progenitor cells expressing endocrine TFs and genes (e.g., *NEUROG3*, *NEUROD1, FEV,* and *CHGA*) but lower levels of *PDX1* (**Fig. 1.6c** and **Supplementary Figure 1.7a**). In control PP2 cultures, *NKX6.1* expression was restricted to multipotent pancreatic progenitors with high *PDX1* expression (**Fig.**

1.6c,d and **Supplementary Figure 1.7b,c**). By contrast, *NKX6.1* was broadly expressed in motif optimized cultures, including in cells expressing lower levels of *PDX1*. Consistent with the lack of enhancer activation in motif optimized GT stage cells (**Fig. 1.6b**), there was no premature expression of *NKX6.1* at the GT stage (**Supplementary Figure 1.7d**). These findings indicate that optimizing FOXA motif strength renders *NKX6.1* expression independent of high levels of PDX1. Corroborating this conclusion, we found NKX6.1 protein restricted to progenitors with high levels of PDX1 in control cultures, whereas motif optimized cultures contained a population of NKX6.1^{+}/PDX1^{low} cells (**Fig. 1.6e** and **Supplementary Figure 1.7e**). In sum, these findings show that increasing FOXA motif strength is sufficient to allow for FOXA recruitment independent of cooperative interactions with pancreatic lineage-determining TFs and that converting an unprimed into a primed enhancer lowers the target gene expression threshold (**Fig. 1.6f**).

Given that alpha cells are derived from NKX6.1^{-} endocrine progenitors, whereas beta cells arise from NKX6.1^{+} endocrine progenitors[56], we examined effects of broader *NKX6.1* expression among progenitors on cell fate allocation. To this end, we differentiated motif optimized and control cells to the early endocrine cell stage, when pre-alpha and pre-beta cells can be distinguished[55] (**Supplementary Figure 1.7f**). We observed a two-fold increase in NKX6.1^{+}/insulin^{+} cells accompanied by a decrease in glucagon expression (**Supplementary Figure 1.7g**), suggesting a pre-alpha to a pre-beta cell fate shift. These results suggest that barriers to enhancer activation and target gene expression imposed by DNA sequence at class II enhancers are biologically relevant for cell lineage allocation during development.

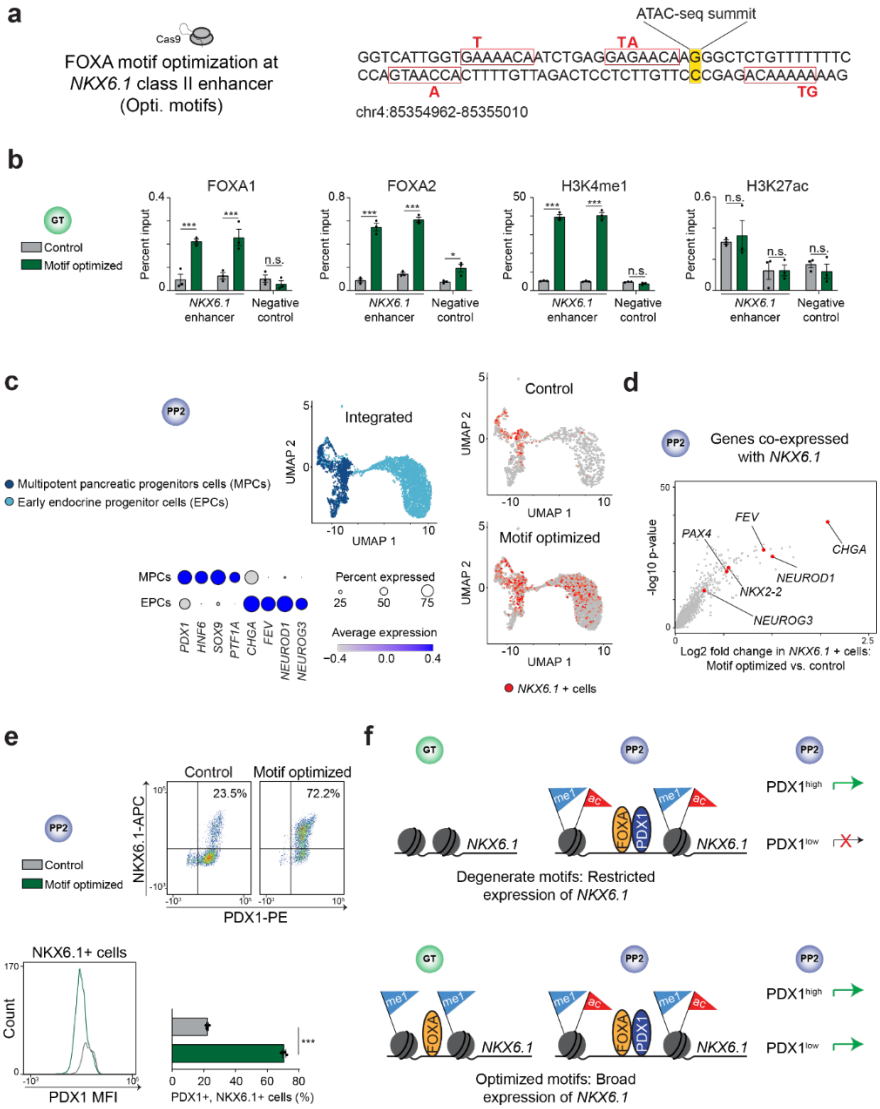

Figure 1.6. Optimization of FOXA binding motifs at an *NKX6.1* enhancer redefines patterns of FOXA association and gene expression. (**a**) Schematic illustrating base editing strategy at *NKX6.1* enhancer via CRISPR-Cas9. Degenerate FOXA binding motifs and base edits are indicated in red. (**b**) ChIP-qPCR comparing FOXA1, FOXA2, H3K4me1, and H3K27ac ChIP-seq signal at the *NKX6.1* enhancer in control and motif optimized hESC lines at GT stage. Plots show two independent primer pairs for *NKX6.1* enhancer and one primer pair for a negative control region (n = 3 technical replicates; P = 3.04 x 10^{-3}, 1.22 x 10^{-2}, and 0.35 for FOXA1; P = 2.22 x 10^{-4}, 4.3 x 10^{-5}, and 2.57 x 10^{-2} FOXA2; P = 1.0 x 10^{-5}, 2.3 x 10^{-5}, and 0.07 for H3K4me1; and P = 0.70, 0.99, and 0.44 for H3K27ac; student's t-test, 2-sided; n.s., not significant). (**c**) UMAP representation of single cell RNA-seq data from both control and motif optimized PP2 cells (integrated) and dot plot showing expression of marker genes in each population (bottom). *NKX6.1* expression across populations in control and motif optimized cell lines (right). (**d**) Volcano plot comparing genes co-expressed with *NKX6.1* in motif optimized compared to control PP2 cells. (**e**) Representative flow cytometry analysis for PDX1 and NKX6.1, mean fluorescence intensity (MFI) of PDX1 signal in NKX6.1^{+} cells, and quantification of PDX1^{+} and NKX6.1^{+} cells in control and motif optimized PP2 cells (n = 3 independent differentiations; P < 1.0 x 10^{-4}; student's t-test, 2-sided). (**f**) Schematic illustrating temporal patterns of FOXA recruitment and *NKX6.1* expression at the PP2 stage in cells with degenerate and optimized FOXA motifs at the *NKX6.1* enhancer.

Distinct temporal patterns of FOXA1/2 occupancy distinguish hepatic and alveolar enhancers

To determine whether the identified mechanisms of enhancer activation during organ development are universal across endodermal lineages, we also analyzed liver and lung enhancers, which like pancreatic enhancers undergo chromatin priming in gut endoderm[27]. Like pancreas development, both early liver and lung development depend on FOXA TFs[30-32]. Furthermore, previous studies have demonstrated FOXA binding to primed liver enhancers in gut endoderm prior to organ lineage induction[27,37]. To test whether class I and class II enhancers can be distinguished during liver and lung development, we induced the hepatic fate from hESC-GT stage intermediates (**Fig. 1.7a**), and generated distal lung alveolar epithelial type 2-like cells (iAT2s) grown at 95% purity as 3D alveolospheres (ALV) from iPSCs (**Fig. 1.7b**)[27,57]. For liver, we analyzed H3K27ac signal and FOXA1/2 binding before liver induction at the GT stage and in hepatic progenitors (HP). For lung, we analyzed H3K27ac signal and FOXA1 binding before lung induction in anteriorized foregut (AFG) and at the ALV stage.

Analogous to the strategy used for identifying pancreatic enhancers (**Supplementary Figure 1.3d**), we identified hepatic and alveolar enhancers based on gain in H3K27ac signal during the GT to HP and AFG to ALV transitions, respectively (≥ 2 fold change in H3K27ac, FDR ≤ 0.05; **Supplementary Figure 1.8a-d**). Subsequently, we quantified FOXA1/2 binding at the identified enhancers. As in pancreas, we observed two distinct patterns of FOXA1/2 occupancy (**Fig. 1.7c,d**) despite similar dynamics in H3K27ac signal (**Supplementary Figure 1.8e,f**): a subset of class I enhancers exhibited FOXA1/2 occupancy prior to lineage induction (488 class I hepatic enhancers and 368 class I alveolar enhancers), whereas class II enhancers constituted the majority and exhibited *de novo* FOXA1/2 binding with lineage induction (965 class II hepatic enhancers and 2924 class II alveolar enhancers). These patterns were exemplified by enhancers near hepatic genes Alpha1-Antitrypsin (*AAT*) and *CEBPA* (**Fig. 1.7e**), as well as lung developmental TF genes *SOX2* and *NKX2.1* (**Fig. 1.7f**).

De novo motif analysis at class I against a background of class II hepatic enhancers revealed enrichment for FOXA motifs, GATA motifs, and the motif for the signal-dependent nuclear receptor NR2E1[58]. Class II enhancers showed comparative enrichment for motifs of the hepatic lineage-determining TFs CEBPA, HNF4A, and TBX[59,60] (**Fig. 1.7g**), which increased in expression upon liver induction from hESC-GT intermediates (**Supplementary Figure 1.9a**). FOXA2, HNF4A, and CEBP have been shown to co-bind liver-specific enhancers after liver induction[33], supporting a potential role for cooperative recruitment of FOXA TFs by these factors. Analogous to the motif enrichment patterns observed in pancreas and liver, alveolar class I enhancers were comparatively enriched for FOXA motifs, GATA motifs, and motifs for signal-dependent TFs NR5A1 (SF1) and TEAD with roles in lung development[61,62], whereas alveolar class II enhancers showed comparative motif enrichment for SOX family TFs and the lung master TF NKX2.1[63] (**Fig. 1.7h**). Thus, as in pancreas, a subset of hepatic and alveolar enhancers with canonical FOXA motifs and enrichment for motifs of signal-dependent TFs are FOXA1/2-bound prior to lineage induction, while *de novo* FOXA1/2 recruitment occurs at most hepatic and alveolar enhancers upon lineage induction.

To gain further insight into the architecture of hepatic and alveolar enhancers, we examined abundance, strength, and positioning of FOXA motifs. Using the same six FOXA PWMs as for pancreatic enhancers (**Supplementary Figure 1.5a**), we observed significant enrichment for occurrence of FOXA motifs at both class I hepatic and class I alveolar enhancers (**Supplementary Figure 1.9b,c**). We also found significantly higher log-odds scores for three FOXA PWMs (MA0047.2, MA0148.1, and MA0148.3; P = 1.40 x 10^{-3}, 2.00 x 10^{-3}, and 1.60 x 10^{-2}, respectively; Wilcoxon rank sum test) at class I compared to class II hepatic enhancers, and two FOXA PWMs (MA0047.3 and MA0148.1; P = 3.1 x 10^{-2} and 4.1 x 10^{-2}, respectively; Wilcoxon rank sum test) at class I compared to class II alveolar enhancers. Furthermore, FOXA motif occurrence at ATAC-seq summits (444 and 701 ATAC-seq summits in class I and class II enhancers, respectively, at HP stage; **Supplementary Figure 1.9d**) and occurrence of FOXA

footprints (**Supplementary Figure 1.9e**) were enriched at class I compared to class II hepatic enhancers. Thus, like pancreatic class I enhancers, hepatic and alveolar class I enhancers exhibit sequence features that have been associated with FOXA1/2 pioneering in other contexts[43,44]. Moreover, analogous to pancreatic enhancers, we observed preferential binding of GATA4 and GATA6 to class I compared to class II hepatic enhancers at the GT stage (**Supplementary Figure 1.9f**), but no binding preference of the hepatic lineage-determining TF HNF4A at class II compared to class I hepatic enhancers despite HNF4A motif enrichment at HNF4A-bound class II enhancers (**Supplementary Figure 1.9f,g**). These results show that similar characteristics of sequence architecture distinguish pancreatic, hepatic, and alveolar class I and class II enhancers.

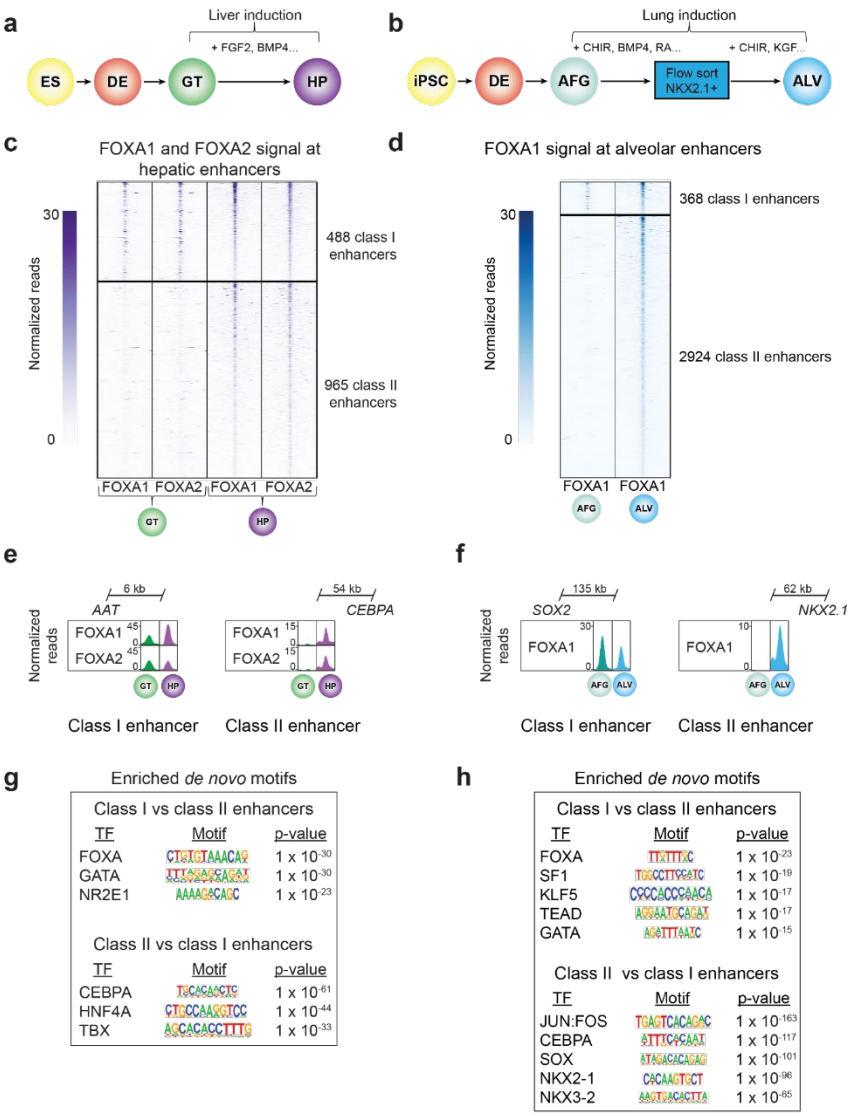

Figure 1.7. Class I and class II enhancers can be distinguished in liver and lung development. (**a** and **b**) Schematic of stepwise differentiation of hESCs to hepatic progenitors (HP) (**a**) and induced human pluripotent stem cells (iPSC) into alveolosphere organoids (ALV) (**b**). AFG, anteriorized foregut. Select growth factors for hepatic (**a**) and alveolar (**b**) lineage induction are indicated. FGF2, fibroblast growth factor 2; BMP4, bone morphogenic protein 4; CHIR, CHIR99021; RA, retinoic acid. (**c**) Heatmap showing density of FOXA1 and FOXA2 ChIP-seq reads at hepatic enhancers in GT and HP. Heatmap is centered on FOXA1 and FOXA2 peaks, respectively, and spans 5 kb. Hepatic enhancers are classified based on temporal pattern of FOXA1 and FOXA2 occupancy. (**d**) Heatmap showing density of FOXA1 ChIP-seq reads at hepatic enhancers in AFG and ALV. Heatmap is centered on FOXA1 peaks and spans 5 kb. Alveolar enhancers are classified based on temporal pattern of FOXA1 occupancy. (**e**) Genome browser snapshots showing FOXA1 and FOXA2 ChIP-seq signal at a class I hepatic enhancer near *AAT* and a class II hepatic enhancers near *CEBPA* in GT and HP. (**f**) Genome browser snapshots showing FOXA1 ChIP-seq signal at a class I alveolar enhancer near *SOX2* and a class II alveolar enhancer near *NKX2.1* in AFG and ALV. (**g** and **h**) Enriched *de novo* transcription factor (TF) binding motifs at class I against a background of class II enhancers and vice versa for hepatic (**g**) and alveolar enhancers (**h**). Fisher's exact test, 1-sided, corrected for multiple comparisons. All ChIP-seq experiments, n = 2 replicates from independent differentiations.

Lineage-specific recruitment of FOXA1/2 to unprimed enhancers

Our results suggest a model whereby the full enhancer complement for each endodermal organ lineage is established through (i) FOXA1/2-mediated priming of a small subset of enhancers for each lineage in endodermal precursors prior to lineage induction, and (ii) activation of a larger subset of unprimed enhancers by organ lineage-determining TFs that cooperatively recruit FOXA1/2 upon lineage induction. To determine the relationship between class I and class II enhancers across different endodermal lineages, we performed differential motif enrichment analysis, comparing class I or class II enhancers of each lineage against a background of class I or class II enhancers, respectively, of the alternate lineages. As expected, motifs for lineage-determining TFs for each lineage were enriched at both classes of enhancers. However, motif enrichment was stronger at class II than at class I enhancers (**Fig. 1.8a**), lending further support to the model that cooperativity with lineage-determining TFs facilitates lineage-specific FOXA1/2 association with class II enhancers of each organ. Consistent with the binding of FOXA1/2 to class I enhancers in shared developmental precursors prior to lineage induction, we found that class I enhancers of one organ lineage were more frequently bound by FOXA1/2 in alternate lineages than class II enhancers (**Fig. 1.8b,c**). Altogether, these findings support establishment of organ-specific gene expression programs through two distinct mechanisms of FOXA1/2-mediated enhancer activation (**Fig. 1.8d**).

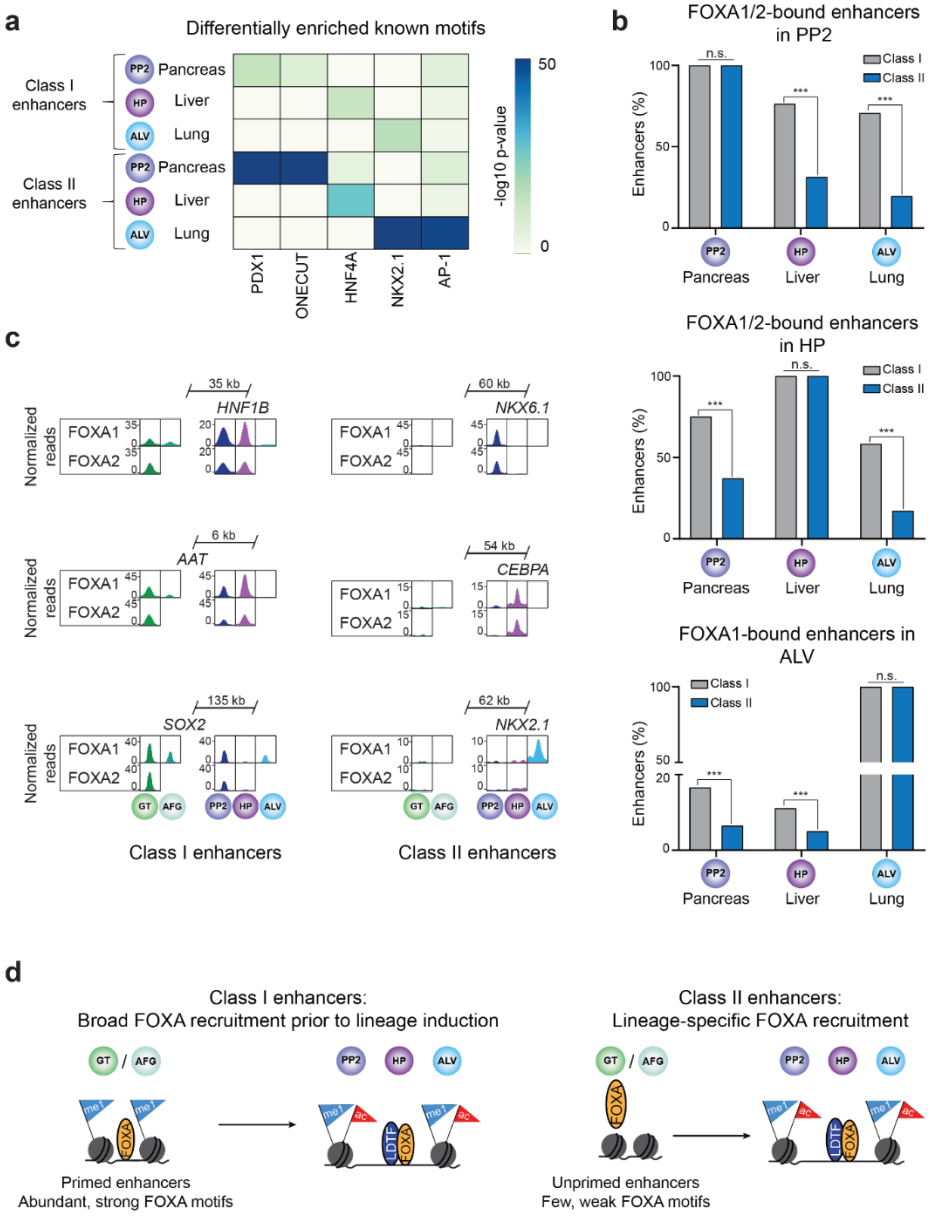

Figure 1.8. Recruitment of FOXA1/2 to class II enhancers is lineage-specific. (**a**) Heatmap showing enrichment of known binding motifs for lineage-determining transcription factors at pancreatic, hepatic, and alveolar class I and class II enhancers. Class I and class II enhancers of each lineage were compared against a background of class I and class II enhancers, respectively, of all other lineages. Fisher's exact test, 1-sided, corrected for multiple comparisons. (**b**) Percentage of pancreatic, hepatic, and alveolar class I and class II enhancers overlapping FOXA1/2 ChIP-seq peaks (within 100 bp from peak) in PP2 (pancreas), HP (liver) and ALV (lung). For ALV only FOXA1 peaks were considered. P = 1, < 2.2 x 10^{-16}, < 2.2 x 10^{-16}, < 2.2 x 10^{-16}, 1, < 2.2 x 10^{-16}, 5.86 x 10^{-11}, 3.08 x 10^{-5}, and 1 for comparisons of FOXA occupancy at class I and class II pancreatic, hepatic, and alveolar enhancers at PP2, HP, and ALV stage cells, respectively; Fisher's exact test, 2-sided). (**c**) Genome browser snapshots showing FOXA1/2 ChIP-seq signal across endodermal lineages at example pancreatic, hepatic, and alveolar class I and class II enhancers. Approximate distance between enhancer and gene body is indicated. (**d**) Schematic showing differential recruitment of FOXA TFs to endodermal organ class I and class II enhancers during endoderm development. LDTF, lineage determining transcription factors. All ChIP-seq experiments, n = 2 replicates from independent differentiations.

Discussion

FOXA TFs are generally thought to control developmental transitions by mediating chromatin priming owing to FOXA's pioneer TF activity[19,27,37]. We have previously reported that chromatin priming and FOXA1/2 recruitment precede organ lineage induction at pancreas, liver, and lung enhancers[27]. Here, we show that chromatin priming and early FOXA1/2 recruitment are limited to a small subset of organ lineage enhancers, whereas the majority transitions from unprimed to active and engages FOXA1/2 upon lineage induction. We demonstrate that DNA sequence logic is the primary determinant of whether an enhancer is primed and recruits FOXA1/2 independent of lineage-specific TFs or whether it is unprimed and requires lineage-specific TFs for FOXA1/2 binding. The results presented here provide a molecular framework for understanding gene regulatory principals that underlie lineage induction and cell type diversification during organogenesis. Our findings support a model whereby FOXA-mediated priming of a subset of organ-specific enhancers enables the initiation of organ-specific gene expression programs by lineage inductive cues, whereas secondary recruitment of FOXA by lineage-specific TFs to most organ-specific enhancers helps establish cell type-specific gene expression by safeguarding against broad target gene expression within the organ progenitor domain.

We observed stronger and more abundant FOXA motifs at primed compared to unprimed enhancers and found that FOXA1/2 recruitment to a proportion of unprimed enhancers depends on the pancreatic TF PDX1. Furthermore, we show that strengthening FOXA motifs at an unprimed enhancer obviates dependency of FOXA1/2 binding on PDX1, resulting in FOXA recruitment and enhancer priming prior to lineage induction. Our findings are consistent with prior observations in tumor cell line models, which have suggested that the ability of FOXA TFs to stably bind and remodel chromatin is DNA sequence dependent[43,44,46]. Our results extend these observations in immortalized cell lines to demonstrate relevance of distinct mechanisms of FOXA recruitment for developmental gene regulation.

Our observation that FOXA1/2 bind primed enhancers without cooperative recruitment by pancreatic TFs raises the question of how FOXA TFs engage their target sites at primed enhancers. We find that a subset of primed enhancers is bound by both FOXA and GATA TFs prior to lineage induction. Given previously demonstrated cooperativity between FOXA and GATA TFs[43], it is possible that GATA TFs help recruit FOXA to a subset of primed enhancers. However, we show that strengthening FOXA motifs is sufficient to enable FOXA1/2 binding to an enhancer not bound by GATA TFs. Therefore, our data support the conclusion that strong FOXA motifs are sufficient to facilitate FOXA TF engagement and chromatin priming during development, consistent with observations that FOXA1/2 can engage target sites on nucleosomal DNA *in vitro*[33-35].

Our findings provide insight into the gene regulatory mechanisms that underlie endodermal organ lineage induction and cell fate specification. We observed enrichment of binding motifs for signal-dependent TFs and binding of the retinoic acid receptor subunit RXR at primed pancreatic enhancers. These findings suggest that organ lineage-inductive cues are read by primed enhancers to initiate expression of lineage-determining TFs. In support of this, primed enhancers are found near *PDX1*, *HNF1B,* and *MEIS1*, which are among the first TFs expressed upon pancreas induction. By contrast, unprimed enhancers are enriched for binding motifs of organ-specific TFs which recruit FOXA1/2 secondarily. Given that FOXA TFs are broadly expressed across endodermal organ lineages, indirect FOXA recruitment by organ-specific TFs provides a safeguard against lineage-aberrant enhancer activation and gene expression. This agrees with studies in *Drosophila* and *Ciona* which suggest that suboptimization of TF binding motifs could be a general principal by which to confer cell specificity to enhancers[52,64].

Replacing low affinity FOXA binding sites at an unprimed enhancer for *NKX6.1* with higher affinity sites broadened the domain of *NKX6.1* expression among pancreatic progenitors. As we show, *NKX6.1* was not prematurely expressed, demonstrating that motif optimization does not eliminate the dependency of target gene expression on lineage-specific cues. This suggests that

early FOXA recruitment through high affinity binding sites lowers the threshold for target gene expression, which could reflect an increased sensitivity of the enhancer to activation by lineage-specific TFs. Thus, higher thresholds to target gene expression conferred by unprimed enhancers will restrict target gene expression to specific cell populations, as enhancer activation will only occur when a specific complement of lineage-specific TFs is present in sufficient concentrations. We propose that gene regulation by unprimed enhancers provides a mechanism for specifying different cell types early in organ development. Small differences in TF expression among early organ progenitors would be sufficient to activate different repertoires of unprimed enhancers, thereby creating divergent gene expression patterns and cell populations. Consistent with this concept, it has been shown that PDX1^{high} and PDX1^{low} cells in the early pancreatic epithelium acquire different cell identities[65].

We demonstrate that conversion of a single enhancer near *NKX6.1* from an unprimed to a primed state is sufficient to alter cell fate due to broadened expression of *NKX6.1* within the progenitor cell domain. These findings show that in a developmental context, differences in FOXA binding affinity at enhancers can affect cell fate allocation. It is therefore possible that polymorphisms at FOXA binding sites determine interindividual differences in endodermal organ cell type composition. Consistent with this possibility, islet cell type composition is known to vary greatly in humans[66] and the *NKX6.1* enhancer contains twelve known polymorphisms predicted to alter the strength and spacing of FOXA motifs. While the importance of polymorphisms for organ cell type composition remains to be demonstrated, our findings support the concept that FOXA TF motif strength at developmental enhancers provides a tunable threshold for target gene expression.

Methods

Human cell culture experiments

hESC research was approved by the University of California, San Diego (UCSD), Institutional Review Board and Embryonic Stem Cell Research Oversight Committee (protocol 090165ZX). Human iPSC research was approved by the Boston University Institutional Review Board (protocol H-33122).

Maintenance of HEK293T cells

HEK293T cells (female) were cultured in a humidified incubator at 37 °C with 5% CO_2 using Dulbecco's Modified Eagle Medium (Cat# 45000-312; 4.5 g/L glucose, [+] L-glutamine, [-] sodium pyruvate) supplemented with 10% fetal bovine serum (FBS).

Maintenance and differentiation of CyT49 hESCs

CyT49 hESCs (male) were maintained and differentiated as described[27,67,68]. Propagation of CyT49 hESCs was carried out by passing cells every 3 to 4 days using Accutase™ (eBioscience) for enzymatic cell dissociation, and with 10% (v/v) human AB serum (Valley Biomedical) included in the hESC media the day of passage. hESCs were seeded into tissue culture flasks at a density of 50,000 cells/cm^2.

Pancreatic differentiation was performed as previously described[27,67,68]. Briefly, a suspension-based culture format was used to differentiate cells in aggregate form. Undifferentiated aggregates of hESCs were formed by re-suspending dissociated cells in hESC maintenance medium at a concentration of 1 x 10^6 cells/mL and plating 5.5 mL per well of the cell suspension in 6-well ultra-low attachment plates (Costar). The cells were cultured overnight on an orbital rotator (Innova2000, New Brunswick Scientific) at 95 rpm. After 24 hours the undifferentiated aggregates were washed once with RPMI medium and supplied with 5.5 mL of day 0 differentiation medium. Thereafter, cells were supplied with the fresh medium for the

appropriate day of differentiation (see below). Cells were continually rotated at 95 rpm, or 105 rpm on days 4 through 8, and no media change was performed on day 10. Both RPMI (Mediatech) and DMEM High Glucose (HyClone) medium were supplemented with 1X GlutaMAX™ and 1% penicillin/streptomycin. Human activin A, mouse Wnt3a, human KGF, human noggin, and human EGF were purchased from R&D systems. Other added components included FBS (HyClone), B-27® supplement (Life Technologies), Insulin-Transferrin-Selenium (ITS; Life Technologies), TGFβ R1 kinase inhibitor IV (EMD Bioscience), KAAD-Cyclopamine (KC; Toronto Research Chemicals), and the retinoic receptor agonist TTNPB (RA; Sigma Aldrich). Day-specific differentiation media formulations were as follows:

Days 0 and 1: RPMI + 0.2% (v/v) FBS, 100 ng/mL Activin, 50 ng/mL mouse Wnt3a, 1:5000 ITS.

Days 1 and 2: RPMI + 0.2% (v/v) FBS, 100 ng/mL Activin, 1:5000 ITS

Days 2 and 3: RPMI + 0.2% (v/v) FBS, 2.5 mM TGFβ R1 kinase inhibitor IV, 25ng/mL KGF, 1:1000 ITS

Days 3 - 5: RPMI + 0.2% (v/v) FBS, 25 ng/mL KGF, 1:1000 ITS

Days 5 - 8: DMEM + 0.5X B-27® Supplement, 3 nM TTNPB, 0.25 mM KAAD-Cyclopamine, 50 ng/mL Noggin

Days 8 - 10: DMEM/B-27, 50 ng/mL KGF, 50 ng/mL EGF

Cells at D0 correspond to the embryonic stem cell (ES) stage, cells at D2 correspond to the definitive endoderm (DE) stage, cells at D5 correspond to the gut tube (GT) stage, cells at D7 correspond to the early pancreatic progenitor (PP1) stage, and cells at D10 correspond to the late pancreatic progenitor (PP2) stage.

Hepatic differentiation was performed as previously described[27]. Briefly, cells were treated identically as in pancreatic differentiation until the GT stage at D5. At this point cells were treated with 50 ng/ml BMP4 (Millipore) and 10 ng/ml FGF2 (Millipore) in RPMI media (Mediatech) supplemented with 0.2% (vol/vol) FBS (HyClone) for 3 days with daily media changes. Cells at D8 correspond to the hepatic progenitor (HP) cell stage.

Maintenance and differentiation of H1 hESCs

H1 hESCs (male) were maintained and differentiated as described with some modifications[69,70]. In brief, hESCs were cultured in mTeSR1 media (Stem Cell Technologies) and propagated by passaging cells every 3 to 4 days using Accutase (eBioscience) for enzymatic cell dissociation.

For differentiation, cells were dissociated using Accutase for 10 min, then reaggregated by plating the cells at a concentration of ~5.5 x 10^6 cells/well in a low attachment 6-well plate on an orbital shaker (100 rpm) in a 37 °C incubator. The following day, undifferentiated cells were washed in base media (see below) and then differentiated using a multi-step protocol with stage-specific media and daily media changes.

All stage-specific base media were comprised of MCDB 131 medium (Thermo Fisher Scientific) supplemented with NaHCO3, GlutaMAX, D-Glucose, and BSA using the following concentrations:

Stage 1/2 base medium: MCDB 131 medium, 1.5 g/L NaHCO3, 1X GlutaMAX, 10 mM D-Glucose, 0.5% BSA

Stage 3/4 base medium: MCDB 131 medium, 2.5 g/L NaHCO3, 1X GlutaMAX, 10 mM D-glucose, 2% BSA

Stage 5 medium: MCDB 131 medium, 1.5 g/L NaHCO3, 1X GlutaMAX, 20 mM D-glucose, 2% BSA

Media compositions for each stage were as follows:

Stage 1 (days 0 - 2): base medium, 100 ng/ml Activin A, 25 ng/ml Wnt3a (day 0). Day 1-2: base medium, 100 ng/ml Activin A

Stage 2 (days 3 - 5): base medium, 0.25 mM L-Ascorbic Acid (Vitamin C), 50 ng/mL FGF7

Stage 3 (days 6 - 7): base medium, 0.25 mM L-Ascorbic Acid, 50 ng/mL FGF7, 0.25 μM SANT-1, 1 μM Retinoic Acid, 100 nM LDN193189, 1:200 ITS-X, 200 nM TPB

Stage 4 (days 8 - 10): base medium, 0.25 mM L-Ascorbic Acid, 2 ng/mL FGF7, 0.25 μM SANT-1, 0.1 μM Retinoic Acid, 200 nM LDN193189, 1:200 ITS-X, 100nM TPB

Stage 5 (days 11 - 13): base medium, 0.25 μM SANT-1, 0.05 μM RA, 100 nM LDN-193189, 1 μM T3, 10 μM ALK5i II, 10 μM ZnSO4, 10 μg/mL heparin, 1:200 ITS-X

Cells at D0, D3, D6, D8, D11, and D14 correspond to the ES DE, GT, PP1, PP2, and EN stages, respectively.

Maintenance and differentiation of iPSCs

SPC2 iPSCs (male; clone SPC2-ST-B2[71]) were maintained in feeder-free culture conditions in 6-well tissue culture dishes (Corning) coated with growth factor reduced Matrigel (Corning), in mTeSR1 medium (Stem Cell Technologies) and passaged using gentle cell dissociation reagent (GCDR). Details of iPSC derivation, characterization, and differentiation into anterior foregut endoderm and alveolar epithelial type 2 cells (iAT2s; also known as iAEC2s) have been previously published[57,71,72] and are available for free download at http://www.bu.edu/dbin/stemcells/protocols.php. Briefly, the SPC2-ST-B2 iPSC clone, engineered to carry a tdTomato reporter knocked into one allele of the endogenous *SFTPC* locus[71], underwent directed differentiation to generate iAT2s in 3D Matrigel cultures as follows. Cells were first differentiated into definitive endoderm using the STEMdiff Definitive Endoderm Kit (Stem Cell Technologies) for 72 hours and subsequently dissociated with GCDR and passaged as small clumps into growth factor reduced Matrigel-coated (Corning) 6-well culture plates (Corning) in ‘‘DS/SB’’ foregut endoderm anteriorization media, consisting of complete serum-free differentiation medium (cSFDM) base as previously described[57], supplemented with 10 μm SB431542 (‘‘SB’’; Tocris) and 2 μm Dorsomorphin (‘‘DS’’; Stemgent), to pattern cells towards anterior foregut endoderm (AFE; day 6 of differentiation). For the first 24 hours after passaging, media was supplemented with 10 μM Y-27632. After anteriorization in DS/SB media for 72 hours, beginning on day 6 of differentiation cells were cultured in ‘‘CBRa’’ lung progenitor-induction

medium for 9 additional days. ‘‘CBRa’’ medium consists of cSFDM base supplemented with 3 μM CHIR99021 (Tocris), 10 ng/mL recombinant human BMP4 (rhBMP4, R&D Systems), and 100 nM retinoic acid (RA, Sigma), as described[57]. On differentiation day 15, NKX2-1^{+} lung progenitors were isolated based on CD47^{hi}/CD26^{neg} gating[73] using a high-speed cell sorter (MoFlo Legacy or MoFlo Astrios EQ). Purified day 15 lung progenitors were resuspended in undiluted growth factor-reduced 3D Matrigel (Corning) at a concentration of 400 cells/μl and distal/alveolar differentiation was performed in ‘‘CK+DCI’’ medium, consisting of cSFDM base supplemented with 3 μm CHIR99021 (Tocris), 10 ng/mL rhKGF (R&D Systems), and 50 nM dexamethasone (Sigma), 0.1 mM 8-Bromoadenosine 30, 50-cyclic monophosphate sodium salt (Sigma) and 0.1 mM 3-Isobutyl-1-methylxanthine (IBMX; Sigma) (DCI) with a brief period of CHIR99021 withdrawal between days 34-39 to achieve iAT2 maturation. To establish pure cultures of iAT2s, cells were sorted by flow cytometry on day 45 to purify SFTPC$^{tdTomato+}$ cells. iAT2s were maintained as self-renewing monolayered epithelial spheres (“alveolospheres”) through serial passaging every 10-14 days and replating in undiluted growth factor-reduced 3D Matrigel (Corning) droplets at a density of 400 cells/μl in CK+DCI medium, as described[72]. iAT2 culture quality and purity was monitored at each passage by flow cytometry, with 95.2 ± 4.2% (mean ± S.D.) of cells expressing SFTPC^{tdTomato} over time, as we have previously detailed[57,71].

Cells at day 6 correspond to the AFG stage and day 261 iAT2s were used for the alveolar stage.

Generation of FOXA1-/-, FOXA2-/-, and FOXA1/2-/- H1 hESC lines

To generate homozygous *FOXA1*, *FOXA2*, and FOXA1/2 deletion hESC lines, sgRNAs targeting coding exons within each gene were cloned into Px333-GFP, a modified version of Px333[74], which was a gift from Andrea Ventura (Addgene, #64073). The plasmid was transfected into H1 hESCs with XtremeGene 9 (Roche), and 24 hours later 8000 GFP^{+} cells were sorted into a well of six-well plate. Individual colonies that emerged within 5-7 days were subsequently

transferred manually into 48-well plates for expansion, genomic DNA extraction, PCR genotyping, and Sanger sequencing. For control clones, the Px333-GFP plasmid was transfected into H1 hESCs, and cells were subjected to the same workflow as H1 hESCs transfected with sgRNAs.

sgRNA oligo used to generate *FOXA1*$^{-/-}$ hESCs: CGCCATGAACAGCATGACTG

sgRNA oligo used to generate *FOXA2*$^{-/-}$ hESCs: CATGAACATGTCGTCGTACG

sgRNA oligos used to generate *FOXA1/2*$^{-/-}$ frameshift hESCs:

FOXA1: CGCCATGAACAGCATGACTG

FOXA2: CATGAACATGTCGTCGTACG

sgRNA oligos used to generate *FOXA1/2*$^{-/-}$ exon deletion hESCs:

FOXA1 upstream: GCGACTGGAACAGCTACTAC

FOXA1 downstream: GCACTGCAATACTCGCCTTA

FOXA2 upstream: TCCGACTGGAGCAGCTACTA

FOXA2 downstream: CGGCTACGGTTCCCCCATGC

Generation of NKX6.1 enhancer motif optimized H1 hESC line

To generate base substitutions in the *NKX6.1* enhancer, a sgRNA targeting the enhancer was cloned into the Px458 plasmid[75], which was a gift from Feng Zhang (Addgene, #48138). The plasmid and an asymmetric single-stranded oligodeoxynucleotide donor template (ssODN) were transfected into H1 hESCs with XtremeGene 9 (Roche), and cells were treated with 1μM SCR7 DNA ligase IV inhibitor to promote homology-directed repair. 24 hours later 8000 GFP^{+} cells were sorted into a well of six-well plate. Individual colonies that emerged within 5-7 days were subsequently transferred manually into 48-well plates for expansion, genomic DNA extraction, PCR genotyping, and Sanger sequencing.

sgRNA oligo used to target *NKX6.1* enhancer: GAAGCTCTCTACCTAGTGTG

ssODN sequence:

TGCCTATGATTTATGTATTTGTTTAGTCAATAGTCTAATGTAAATGATGTAATTAATTATAGAT

GGTGGTGTCAGGTCATTTGTGTAAACAATCTGAGGTAAACAAGGGCTCTGTTTACTTCATG
ACAGATGCAGGGGGGTGGGGGGCTGAGTTGAGGGAATTCCAGGGGAACTTTTTCACGTG
TGAATGGCGGCTGGGA

Transduction of CyT49 hESCs with SCRAM and shPDX1

To generate shRNA expression vectors, shRNA guide sequences were placed under the control of the human U6 pol III promoter in the pLL3.7 backbone[76], which was a gift from Luk Parijs (Addgene, plasmid #11795). Guide sequences are listed in **Supplementary Figure 1.10**.

High-titer lentiviral supernatants were generated by co-transfection of the shRNA expression vector and the lentiviral packaging construct into HEK293T cells as described [67]. Briefly, shRNA expression vectors were co-transfected with the pCMV-R8.74 and pMD2.G expression plasmids (Addgene #22036 and #12259, respectively, gifts from Didier Trono) into HEK293T cells using a 1 mg/ml PEI solution (Polysciences). Lentiviral supernatants were collected at 48 hours and 72 hours after transfection. Lentiviruses were concentrated by ultracentrifugation for 120 min at 19,500 rpm using a Beckman SW28 ultracentrifuge rotor at 4°C.

CyT49 hESCs were plated onto a six-well plate at a density of 1 million cells per well. The following morning, concentrated lentivirus was added at 5 µL/mL media, as well as 8 µg/mL polybrene. After 30 minutes of incubation, the 6 well plate was spun in a centrifuge (Sorvall Legend RT) for 1 hour at 30°C at 950 G. 6 hours later, viral media was replaced with fresh base culture media. After 72 hours, cells were sorted for GFP expression and re-cultured.

Immunofluorescence analysis

Cell aggregates derived from hESCs were allowed to settle in microcentrifuge tubes and washed twice with PBS before fixation with 4% paraformaldehyde (PFA) for 30 min at room temperature. Fixed samples were washed twice with PBS and incubated overnight at 4°C in 30% (w/v) sucrose in PBS. Samples were then loaded into disposable embedding molds (VWR),

covered in Tissue-Tek® O.C.T. Sakura® Finetek compound (VWR) and flash frozen on dry ice to prepare frozen blocks. The blocks were sectioned at 10 µm and sections were placed on Superfrost Plus® (Thermo Fisher) microscope slides and washed with PBS for 10 min. Slide-mounted cell sections were permeabilized and blocked with blocking buffer, consisting of 0.15% (v/v) Triton X-100 (Sigma) and 1% (v/v) normal donkey serum (Jackson Immuno Research Laboratories) in PBS, for 1 hour at room temperature. Slides were then incubated overnight at 4°C with primary antibody solutions. The following day slides were washed five times with PBS and incubated for 1 hour at room temperature with secondary antibody solutions. Cells were washed five times with PBS before coverslips were applied.

All antibodies were diluted in blocking buffer at the ratios indicated below. Primary antibodies used were mouse anti-FOXA1 (1:100 or 1:1000 dilution, Abcam); goat anti-FOXA2 (1:300 dilution, R&D systems); goat anti-SOX17 (1:300 dilution, R&D systems); goat anti-HNF4A (1:1000 dilution, Santa Cruz Biotechnology); rabbit anti-PDX1 (1:500 dilution, Abcam); and mouse anti-NKX6.1 (1:300 dilution, Developmental Studies Hybridoma Bank). Secondary antibodies against mouse, rabbit, and goat were Alexa488- and Cy3-conjugated donkey antibodies (Jackson Immuno Research Laboratories), and were used at dilutions of 1:500 (anti-rabbit Alexa488) or 1:1000 (all other secondary antibodies). Cell nuclei were stained with Hoechst 33342 (1:3000, Invitrogen). Representative images were obtained with a Zeiss Axio-Observer-Z1 microscope equipped with a Zeiss ApoTome and AxioCam digital camera. Figures were prepared in Adobe Creative Suite 5.

Flow cytometry analysis

Cell aggregates derived from hESCs were allowed to settle in microcentrifuge tubes and washed with PBS. Cell aggregates were incubated with Accutase® at 37°C until a single-cell suspension was obtained. Cells were washed with 1 mL ice-cold flow buffer comprised of 0.2% BSA in PBS and centrifuged at 200 g for 5 min. BD Cytofix/Cytoperm™ Plus

Fixation/Permeabilization Solution Kit was used to fix and stain cells for flow cytometry according to the manufacturer's instructions. Briefly, cell pellets were re-suspended in ice-cold BD Fixation/Permeabilization solution (300 µL per microcentrifuge tube). Cells were incubated for 20 min at 4°C. Cells were washed twice with 1 mL ice-cold 1X BD Perm/Wash™ Buffer and centrifuged at 10°C and 200 x g for 5 min. Cells were re-suspended in 50 µL ice-cold 1X BD Perm/Wash™ Buffer containing diluted antibodies, for each staining performed. Cells were incubated at 4°C in the dark for 1-3 hours. Cells were washed with 1.25 mL ice-cold 1X BD Wash Buffer and centrifuged at 200 g for 5 min. Cell pellets were re-suspended in 300 µL ice-cold flow buffer and analysed in a FACSCanto™ II (BD Biosciences). Antibodies used were PE-conjugated anti-SOX17 antibody (1:20 dilution, BD Biosciences); mouse anti-HNF1B antibody (1:100 dilution, Santa Cruz Biotechnology); PE-conjugated anti-mouse IgG (1:50 dilution, BD Biosciences); PE-conjugated anti-PDX1 (1:10 dilution, BD Biosciences); AlexaFluor® 647-conjugated anti-NKX6.1 (1:5 dilution, BD Biosciences); and PE-conjugated anti-Insulin (1:50 dilution, Cell Signaling). Data were processed using FlowJo software v10.

Chromatin Immunoprecipitation Sequencing (ChIP-seq)

ChIP-seq was performed using the ChIP-IT High-Sensitivity kit (Active Motif) according to the manufacturer's instructions. Briefly, for each cell stage and condition analyzed, 5-10 x 10^6 cells were harvested and fixed for 15 min in an 11.1% formaldehyde solution. Cells were lysed and homogenized using a Dounce homogenizer and the lysate was sonicated in a Bioruptor® Plus (Diagenode), on high for 3 x 5 min (30 sec on, 30 sec off). Between 10 and 30 µg of the resulting sheared chromatin was used for each immunoprecipitation. Equal quantities of sheared chromatin from each sample were used for immunoprecipitations carried out at the same time. 4 µg of antibody were used for each ChIP-seq assay. Chromatin was incubated with primary antibodies overnight at 4°C on a rotator followed by incubation with Protein G agarose beads for 3 hours at 4°C on a rotator. Antibodies used were rabbit anti-H3K27ac (Active Motif 39133); rabbit

anti-H3K4me1 (Abcam ab8895); goat anti-FOXA1 (Abcam Ab5089); goat-anti-FOXA2 (Santa Cruz SC-6554); goat anti-GATA4 (Santa Cruz SC-1237); mouse anti-GATA6 (Santa Cruz SC-9055); and mouse anti-HNF4A (Novus PP-H1415). Reversal of crosslinks and DNA purification were performed according to the ChIP-IT High-Sensitivity instructions, with the modification of incubation at 65°C for 2-3 hours, rather than at 80°C for 2 hours. Sequencing libraries were constructed using KAPA DNA Library Preparation Kits for Illumina® (Kapa Biosystems) and library sequencing was performed on either a HiSeq 4000 System (Illumina®) or NovaSeq 6000 System (Illumina®) with single-end reads of either 50 or 75 base pairs (bp). Sequencing was performed by the UCSD Institute for Genomic Medicine (IGM) core research facility. For ChIP-seq experiments at the DE, AFG, and ALV stages in iAEC2 cells, two technical replicates from a single differentiation were generated. For all other ChIP-seq experiments, replicates from two independent hESC differentiations were generated.

ChIP-qPCR

For ChIP-qPCR, immunoprecipitation, reversal of crosslinks, and DNA purification were performed as for ChIP-seq. Antibodies used were rabbit anti-H3K27ac (Active Motif 39133); rabbit anti-H3K4me1 (Abcam ab8895); goat anti-FOXA1 (Abcam Ab5089); and goat anti-FOXA2 (R&D AF2400). After DNA purification, each sample and a 1% dilution of input DNA used for immunoprecipitation were amplified using 2 independent primers targeting either the histones flanking the *NKX6.1* enhancer (for measurements of H3K4me1 and H3K27ac) or the FOXA binding site (for measurements of FOXA1 and FOXA2), as well as a negative control region. qPCR reactions were performed in technical triplicates using a CFX96™ Real-Time PCR Detection System and the iQ™ SYBR® Green Supermix (Bio-Rad). Primer sequences are listed in **Supplementary Figure 1.11**.

ChIP-seq data analysis

ChIP-seq reads were mapped to the human genome consensus build (hg19/GRCh37) and visualized using the UCSC Genome Browser[77]. Burrows-Wheeler Aligner (BWA)[78] version 0.7.13 was used to map data to the genome. Unmapped and low-quality (q<15) reads were discarded. SAMtools[79] version 1.5 was used to remove duplicate sequences and HOMER[80] version 4.10.4 was used to call peaks using the findPeaks command with default parameters. The command "-style factor" was used for TFs and the command "-style histone" was used for histone modifications. Stage- and condition-matched input DNA controls were used as background when calling peaks. The BEDtools[81] version 2.26.0 suite of programs was used to perform genomic algebra operations. Tag directories were created for each replicate using HOMER. Directories from each replicate were then combined, and peaks were called from the combined replicates using HOMER. These peaks were then intersected with pancreatic enhancers, hepatic enhancers, or alveolar enhancers, respectively. Pearson correlations for the intersecting peaks were calculated between each pair of replicates using the command multiBamSummary from the deepTools2 package[82] version 3.1.3 and are listed in **Supplementary Figure 1.12**.

RNA isolation and sequencing (RNA-seq) and qRT-PCR

RNA was isolated from cell samples using the RNeasy® Micro Kit (Qiagen) according to the manufacturer instructions. For each cell stage and condition analyzed between 0.1 and 1 x 10^6 cells were collected for RNA extraction. For qRT-PCR, cDNA synthesis was first performed using the iScript™ cDNA Synthesis Kit (Bio-Rad) and 500 ng of isolated RNA per reaction. qRT-PCR reactions were performed in triplicate with 10 ng of template cDNA per reaction using a CFX96™ Real-Time PCR Detection System and the iQ™ SYBR® Green Supermix (Bio-Rad). PCR of the TATA binding protein (TBP) coding sequence was used as an internal control and relative expression was quantified via double delta CT analysis. For RNA-seq, stranded, single-end sequencing libraries were constructed from isolated RNA using the TruSeq® Stranded mRNA Library Prep Kit (Illumina®) and library sequencing was performed on either a HiSeq 4000 System

(Illumina®) or NovaSeq 6000 System (Illumina®) with single-end reads of either 50 or 75 base pairs (bp). Sequencing was performed by the UCSD IGM core research facility. RT-qPCR primer sequences are listed in **Supplementary Figure 1.13**.

RNA-seq data analysis

Reads were mapped to the human genome consensus build (hg19/GRCh37) using the Spliced Transcripts Alignment to a Reference (STAR) aligner version 2.4[83]. Normalized gene expression (fragments per kilobase per million mapped reads; FPKM) for each sequence file was determined using Cufflinks[84] version 2.2.1 with the parameters: --library-type fr-firststrand --max-bundle-frags 10000000. Differential gene expression was determined using DESeq2[85]. Adjusted *P*-values < 0.05 and fold change ≥ 2 were considered significant. For RNA-seq corresponding to cells at the HP stage, one replicate was generated. For all other RNA-seq experiments, replicates from two independent hESC differentiations were generated. Pearson correlations between bam files corresponding to each pair of replicates were calculated and are listed in **Supplementary Figure 1.14**.

Assay for Transposase Accessible Chromatin Sequencing (ATAC-seq)

ATAC-seq[86] was performed on approximately 50,000 nuclei. The samples were permeabilized in cold permabilization buffer (0.2% IGEPAL-CA630 (I8896, Sigma), 1 mM DTT (D9779, Sigma), Protease inhibitor (05056489001, Roche), 5% BSA (A7906, Sigma) in PBS (10010-23, Thermo Fisher Scientific) for 10 minutes on the rotator in the cold room and centrifuged for 5 min at 500 xg at 4°C. The pellet was resuspended in cold tagmentation buffer (33 mM Tris-acetate (pH = 7.8) (BP-152, Thermo Fisher Scientific), 66 mM K-acetate (P5708, Sigma), 11 mM Mg-acetate (M2545, Sigma), 16% DMF (DX1730, EMD Millipore) in Molecular biology water (46000-CM, Corning)) and incubated with tagmentation enzyme (FC-121-1030; Illumina) at 37 °C for 30 min with shaking at 500 rpm. The tagmented DNA was purified using

MinElute PCR purification kit (28004, QIAGEN). Libraries were amplified using NEBNext High-Fidelity 2X PCR Master Mix (M0541, NEB) with primer extension at 72°C for 5 minutes, denaturation at 98°C for 30 s, followed by 8 cycles of denaturation at 98°C for 10 s, annealing at 63°C for 30 s and extension at 72°C for 60 s. After the purification of amplified libraries using MinElute PCR purification kit (28004, QIAGEN), double size selection was performed using SPRIselect bead (B23317, Beckman Coulter) with 0.55X beads and 1.5X to sample volume. Finally, libraries were sequenced on HiSeq4000 (Paired-end 50 cycles, Illumina).

ATAC-seq data analysis

ATAC-seq reads were mapped to the human genome (hg19/GRCh37) using Burrows-Wheeler Aligner[78] (BWA) version 0.7.13, and visualized using the UCSC Genome Browser[77]. SAMtools[79] was used to remove unmapped, low-quality (q<15), and duplicate reads. MACS2[87] version 2.1.4 was used to call peaks, with parameters “shift set to 100 bps, smoothing window of 200 bps” and with “nolambda” and “nomodel” flags on. MACS2 was also used to call ATAC-Seq summits, using the same parameters combined with the “call-summits” flag.

For all ATAC-seq experiments, replicates from two independent hESC differentiations were generated. Bam files for each pair of replicates were merged for downstream analysis using SAMtools, and Pearson correlations between bam files for each individual replicate were calculated over a set of peaks called from the merged bam file. Correlations were performed using the command multiBamSummary from the deepTools2 package[82] with the “--removeOutliers” flag and are listed in **Supplementary Figure 1.15**.

Hi-C data analysis

Hi-C data were processed as previously described[88]. Read pairs were aligned to the hg19 reference genome separately using BWA-MEM with default parameters[78]. Specifically, chimeric reads were processed to keep only the 5’ position and reads with low mapping quality (<10) were

filtered out. Read pairs were then paired using custom scripts. Picard tools were then used to remove PCR duplicates. Bam files with alignments were further processed into text format as required by Juicebox tools[89]. Juicebox tools were then applied to generate Hi-C files containing normalized contact matrices. All downstream analysis was based on 10 Kb resolution KR normalized matrices.

Chromatin loops were identified by comparing each pixel with its local background, as described previously[90] with some modifications. Specifically, only the donut region around the pixel was compared to model the expected count. Briefly, the KR-normalized contact matrices at 10 Kb resolution were used as input for loop calling. For each pixel, distance-corrected contact frequencies were calculated for each surrounding bin and the average of all surrounding bins. The expected counts were then transformed to raw counts by multiplying the counts with the raw-to-KR normalization factor. The probability of observing raw expected counts was calculated using Poisson distribution. All pixels with *P*-value < 0.01 and distance less than 10 Kb were selected as candidate pixels. Candidate pixels were then filtered to remove pixels without any neighboring candidate pixels since they were likely false positives. Finally, pixels within 20 Kb of each other were collapsed and only the most significant pixel was selected. The collapsed pixels with *P*-value < 1×10^{-5} were used as the final list of chromatin loops.

Single cell RNA sequencing library preparation

Pancreatic progenitor cells at day 11 of differentiation were allowed to settle in microcentrifuge tubes and washed with PBS. Cell aggregates were incubated with Accutase® at 37°C until a single-cell suspension was obtained. Cells were then resuspended in 1 mL ice-cold flow buffer comprised of 0.2% BSA in PBS and stained with propidium iodide (Sigma) to distinguish live cells. 500,000 live cells were collected using a FACSAria™ Fusion Flow Sorter, and 10,000 cells per sample were then loaded onto a 10X Chromium Controller and run using Next GEM Single Cell 3' v3.1 reagents. Library preparation was performed according to

manufacturer's instructions, and libraries were sequenced using a NovaSeq S4 (Paired-end 100 bp reads, Illumina).

Single cell RNA-sequencing data analysis

Sequencing reads were processed using CellRanger[91] version 6.0.0, and matrices generated by CellRanger were imported into Seurat[92] version 3 for further processing. Doublet cells (>8000 total features for control cells and >6000 total features for motif optimized cells), low-coverage cells (<3000 total features for control cells and <2500 total features for motif optimized cells), and poor-quality cells (>10% mitochondrial reads for both conditions) were removed from further analysis. Each dataset was Log Normalized with a scale factor of 10000 using the command "NormalizeData." Percentage of mitochondrial genes were regressed out of each dataset using the command "ScaleData." Integration anchors for each dataset were identified using "FindIntegrationAnchors," and datasets were integrated using the command "IntegrateData." Principal component analysis was performed for the integrated dataset using the command "RunPCA," and UMAP plots were generated through "RunUMAP." Clusters were defined running the commands "FindNeighbors" and "FindClusters" at a resolution of 0.03, and marker genes were identified using "FindMarkers." Feature plots and dot plots were generated using the commands "Featureplot" and "Dotplot," and differential expression of genes co-expressed with *NKX6.1* was calculated by subsetting for cells expressing *NKX6.1* and using "FindMarkers" to determine differential genes between control and motif optimized cells. Wilcoxon rank sum tests were used to calculate differential expression.

Gene Ontology analysis

Gene ontology analysis for enhancer groups was performed using GREAT[93] version 4.0.4 with the default parameters. Gene ontology for differentially expressed genes and genes

associated with class I and class II enhancers was performed using Metascape[94] using default parameters.

Identification of super-enhancers

To define pancreatic super-enhancers, we first identified pancreatic enhancers as distal genomic regions exhibiting a ≥ 2-fold increase in H3K27ac ChIP-seq signal during pancreas induction. We then used Rank Ordering of Super-enhancers (ROSE) software[47,95] to join identified pancreatic enhancers within a 12.5 kb span and rank these joined enhancers based on intensity of H3K27ac ChIP-seq signal. These joined enhancers were plotted based on H3K27ac signal, and pancreatic super-enhancers were defined as joined enhancers ranking above the inflection point of the resulting graph.

Principal component analysis

For RNA-seq data, transcriptomes were first filtered for genes expressed (FPKM ≥ 1) in at least one condition, then log10 transformed. For distal H3K27ac signals, H3K27ac peaks were filtered for distal enhancers (≥ 2.5 kb from any annotated TSS). Based on filtered values, PCA plots were generated using the PRComp package in R.

Quantification of changes in H3K27ac signal

HOMER[80] was used to annotate raw H3K27ac ChIP-seq reads over distal enhancers at developmental stages both before and after lineage induction. HOMER was then used to invoke the R package DESeq2[85] version 3.10 for differential analysis, using default parameters.

Quantification of changes in TF ChIP-seq and ATAC-seq signal

HOMER[80] was used to annotate raw FOXA1 and FOXA2 ChIP-seq reads, as well as ATAC-seq reads over PDX1-bound class I and class II enhancers in cells transfected with

SCRAM and sh*PDX1* lentivirus. HOMER was then used to invoke the R package DESeq2[85] for differential analysis, using the flag “norm2total.”

Assignment of enhancer target genes

RNA-seq data were filtered for expressed genes (FPKM ≥ 1) at the PP2 stage, and BEDTools[81] “closest” command was used to assign each enhancer to the nearest annotated TSS.

Motif enrichment analysis

HOMER[80] was used for comparative motif enrichment analyses, using the command findMotifsGenome.pl. *de novo* motifs were assigned to TFs based on suggestions generated by HOMER.

Identification of FOXA motifs and generation of log-odds scores

FOXA1 and FOXA2 PWMs were selected to encompass the most divergent PWMs for each TF. PWMs were downloaded from the JASPAR database[53], and occurrences with associated log-odds scores were quantified using the FIMO feature within the MEMEsuit package[96] version 5.1.1.

Calculation of positional motif enrichment

Identified ATAC-seq summits on class I and class II enhancers were flanked by 500 bp in each direction, and the CENTRIMO feature within the MEMEsuit package[97] version 5.1.1 was used to determine enrichment at summits for selected PWMs associated with FOXA1 and FOXA2, as well as to graph the positional probability of motif occurrence with respect to ATAC-seq summits.

ATAC-seq footprinting analysis

ATAC-seq footprinting was performed as previously described[98]. In brief, diploid genomes for CyT49 were created using vcf2diploid (version 0.2.6a)[99] and genotypes called from whole genome sequencing and scanned for a compiled database of TF sequence motifs from JASPAR[100] and ENCODE[101] with FIMO (version 4.12.0)[96] using default parameters for p-value threshold and a 40.9% GC content based on the hg19 human reference genome. Footprints within ATAC-seq peaks were discovered with CENTIPEDE (version 1.2)[102] using cut-site matrices containing Tn5 integration counts within a ±100 bp window around each motif occurrence. Footprints were defined as those with a posterior probability ≥ 0.99.

Permutation-based significance

A random sampling approach (10,000 iterations) was used to obtain null distributions for enrichment analyses, in order to obtain *P*-values. Null distributions for enrichments were obtained by randomly shuffling enhancer regions using BEDTools[81] and overlapping with FOXA1/2 binding sites. *P*-values < 0.05 were considered significant.

Quantification and statistical analysis

Statistical analyses were performed using GraphPad Prism (v8.1.2), and R (v3.6.1). Statistical parameters such as the value of n, mean, standard deviation (SD), standard error of the mean (SEM), significance level (n.s., not significant; *p < 0.05; **p < 0.01; and ***p < 0.001), and the statistical tests used are reported in the figures and figure legends. The ‘‘n’’ refers to the number of independent hESC differentiation experiments analyzed (biological replicates). All bar graphs and line graphs are displayed as mean ± S.E.M, and all box plots are centered on median, with box encompassing 25th-75th percentile and whiskers extending up to 1.5 interquartile range. Statistically significant gene expression changes were determined with DESeq2[85].

Data sources

The following datasets used in this study were obtained from the GEO and ArrayExpress repositories:

RNA-seq: Pancreatic differentiation of CyT49 hESC line (E-MTAB-1086)

ChIP-seq: H3K27ac in CyT49 hESC, DE, GT, PP1, PP2 (GSE54471 and GSE149148); H3K27ac in CyT49 PP2 SCRAM and PP2 shPDX1 (GSE54471); H3K4me1 in CyT49 GT and PP2 (GSE54471 and GSE149148); RXR in CyT49 PP1 (GSE104840); PDX1 in CyT49 PP2 (GSE54471 and GSE149148); HNF6 in CyT49 PP2 (GSE149148); SOX9 in CyT49 PP2 (GSE149148); FOXA1 in CyT49 PP2 (GSE149148); FOXA2 in CyT49 PP2 (GSE149148).

ATAC-seq: CyT49 GT and PP2 (GSE149148)

Hi-C datasets were generated as a component of the 4D Nucleome Project78. Datasets corresponding to the PP2 stages of differentiation can be found under accession number 4DNES0LVRKBM.

Data availability

All mRNA-seq, ChIP-seq, and ATAC-seq datasets generated for this study have been deposited at GEO under the accession number GSE148368.

Author contributions

A.W. and M.S. conceived the project. R.J.G, A.W., and M.S. designed experiments. A.W., R.J.G., D.K.L, N.K.V, D.A.R, K.D.A, J.W., A.R., and S.K. performed experiments. A.W., R.J.G, N.K.V., Y.Q., and J.C. analyzed sequencing data. R.J.G, A.W, and M.S. interpreted data. R.J.G. and M.S. wrote the manuscript. B.R., K.J.G., D.N.K., and M.S. supervised all research.

Declaration of interests

K.J.G. does consulting for Genentech. The authors declare no other competing interests.

Acknowledgements

We thank Ileana Matta for assistance with ATAC-seq assays and library preparations, and members of the Sander laboratory, Dr. Christopher W. Benner, Dr. Emma Farley, and Dr. Xin Sun for helpful discussions and critical reading of the manuscript. We acknowledge support of the UCSD Human Embryonic Stem Cell Core for cell sorting and K. Jepsen and the UCSD IGM Genomic Center (supported by P30 DK063491) for library preparation and sequencing. This work was supported by grant T32 GM008666 (R.J.G.), R01 DK068471 and R01 DK078803 (M.S.), U54 DK107977 (B.R.), the I.M. Rosenzweig Junior Investigator Award from the Pulmonary Fibrosis Foundation (K.D.A.), R01 HL095993, R01HL128172, U01TR001810, and N01 75N92020C00005 (D.N.K).

Chapter 1, in full, is a reprint of the material as it appears in Nature Communications 2021. This material was co-authored with Geusz, R.J., Wang, A., Lam, D., Vinckier, N.K., Alysandratos, K., Roberts, D., Wang, J., Kefalopoulou, S., Ramirez, A., Qui, Y., Chiou, J., Gaulton, K.J., Ren, B., Kotton, D.N., and Sander, M. The dissertation author was the principal author of this chapter.

CHAPTER 2: LSD1-MEDIATED ENHANCER SILENCING ATTENUATES RETINOIC ACID SIGNALLING DURING PANCREATIC ENDOCRINE CELL DEVELOPMENT

Abstract

Developmental progression depends on temporally defined changes in gene expression mediated by transient exposure of lineage intermediates to signals in the progenitor niche. To determine whether cell-intrinsic epigenetic mechanisms contribute to signal-induced transcriptional responses, we here manipulated the signalling environment and activity of the histone demethylase LSD1 during differentiation of hESC-gut tube intermediates into pancreatic endocrine cells. We identify a transient requirement for LSD1 in endocrine cell differentiation spanning a short time-window early in pancreas development, a phenotype we reproduced in mice. Examination of enhancer and transcriptome landscapes revealed that LSD1 silences transiently active retinoic acid (RA)-induced enhancers and their target genes. Furthermore, prolonged RA exposure phenocopies LSD1 inhibition, suggesting that LSD1 regulates endocrine cell differentiation by limiting the duration of RA signalling. Our findings identify LSD1-mediated enhancer silencing as a cell-intrinsic epigenetic feedback mechanism by which the duration of the transcriptional response to a developmental signal is limited.

Introduction

During development, intermediate progenitors progress toward a distinct cell fate as a result of sequential instructions by signalling cues in the progenitor niche. The duration of a developmental signal has to be limited in order for developmental intermediates to appropriately respond to the next inductive cue. For example, pancreas induction from the foregut endoderm requires retinoic acid (RA) signalling, but thereafter RA signalling activity needs to be dampened for pancreatic progenitors to correctly interpret pro-endocrine differentiation cues[50]. Thus,

signalling cues are interpreted in a highly context-dependent manner and signals need to be temporally limited to delineate critical competence windows for developmental transitions. An open question is whether removal of the signal is sufficient to terminate a response to a signal or whether cell-intrinsic mechanisms at the level of the responder tissue enable developmental transitions by limiting the duration of signal-induced transcriptional responses.

Spatiotemporal gene expression during development is regulated by transcriptional enhancers[103]. Chromatin state at enhancers is a significant determinant of transcriptional responsiveness to environmental signals, and enhancers respond to signalling cues by modifying their chromatin state. Enhancers can exhibit an inactive, poised, or active chromatin state. Inactive enhancers are characterised by compact chromatin and absence of active histone modifications, whereas poised enhancers are nucleosome-free and marked by mono- and di-methylation of histone H3 at lysine 4 (H3K4me1 and H3K4me2)[6,27,104]. The transition from an inactive to a poised enhancer state during development coincides with a gain in cellular competence of lineage intermediates to respond to inductive signalling cues[27]. Thus, developmental competence can be defined as a temporal state during which the epigenetic landscape is permissive for responding to environmental signals. Signal-dependent transcription factors (TFs) activate poised enhancers by recruiting co-activator complexes containing histone acetyltransferases (HATs) that deposit H3K27 acetylation (H3K27ac) marks, thereby transforming the poised enhancer into one that actively supports transcription[105]. It is unknown whether or not the erasure of these epigenetic marks is a prerequisite for termination of one competence window and transition to the next.

Lysine-specific demethylase 1 (LSD1), also known as KDM1A, regulates chromatin by catalysing the removal of mono- and di-methyl marks from K4 at histone H3[106], thus rendering poised enhancer chromatin inactive[107]. This process has been called enhancer decommissioning and is coupled to complete silencing of associated genes[107]. Despite its role in enhancer silencing, LSD1 frequently resides in complexes of active enhancers[108-110]. In the context of acetylated

histones, LSD1 activity and demethylation of H3K4 is inhibited[108]. Therefore, current evidence suggests that histones need to be deacetylated before LSD1 can decommission active enhancers. Consistent with this mechanism, LSD1 occupies enhancers of pluripotency genes in pluripotent stem cells and decommissions these enhancers only when pluripotent stem cells undergo differentiation[107]. Whether LSD1-mediated regulation of enhancer chromatin plays a role in defining developmental competence windows and enabling sequential cell state transitions remains unknown.

Here, we asked whether epigenetic mechanisms can limit the duration of an inductive signal throughout a developmental time course, thereby defining distinct competence windows and preventing inappropriate responses to developmental signals. To investigate this, we manipulated LSD1 activity and RA signalling in a human embryonic stem cell (hESC)-based differentiation system, where cells progress stepwise in defined conditions toward the pancreatic endocrine cell fate. We show that LSD1-mediated enhancer decommissioning limits the time window, during which cells express RA-induced genes. When LSD1 activity is inhibited immediately after pancreas induction, RA-induced genes fail to be silenced despite removal of RA as an inductive signal, which is associated with an inability of the cells to undergo endocrine cell differentiation. Thus, our results show that loss of LSD1 function critically alters the epigenetic landscape that terminates the competence window for RA signalling. These findings identify modification of the epigenome as an important cell-intrinsic mechanism for sharpening transcriptional responses to developmental signals.

Results

Human Pancreatic Endocrine Cell Development Requires LSD1

To investigate possible roles for LSD1 during defined windows of transition to a differentiated cell type, we employed a hESC differentiation system, in which cells progress stepwise toward the pancreatic endocrine cell lineage through sequential exposure to signalling cues that guide

corresponding cell state transitions in the developing embryo (**Fig. 2.1a**)[27,67,68,111]. In this differentiation system, LSD1 was broadly expressed throughout progression to the endocrine cell stage (EN) (**Supplementary Figure 2.1a**,**b**). Likewise, levels of flavin adenine dinucleotide (FAD), which is a metabolic cofactor of LSD1[112], did not change substantially throughout the differentiation time course (**Supplementary Figure 2.1c**). We verified LSD1 expression in pancreatic progenitor cells and differentiated endocrine cells in human foetal and adult tissue (**Supplementary Figure 2.1d**).

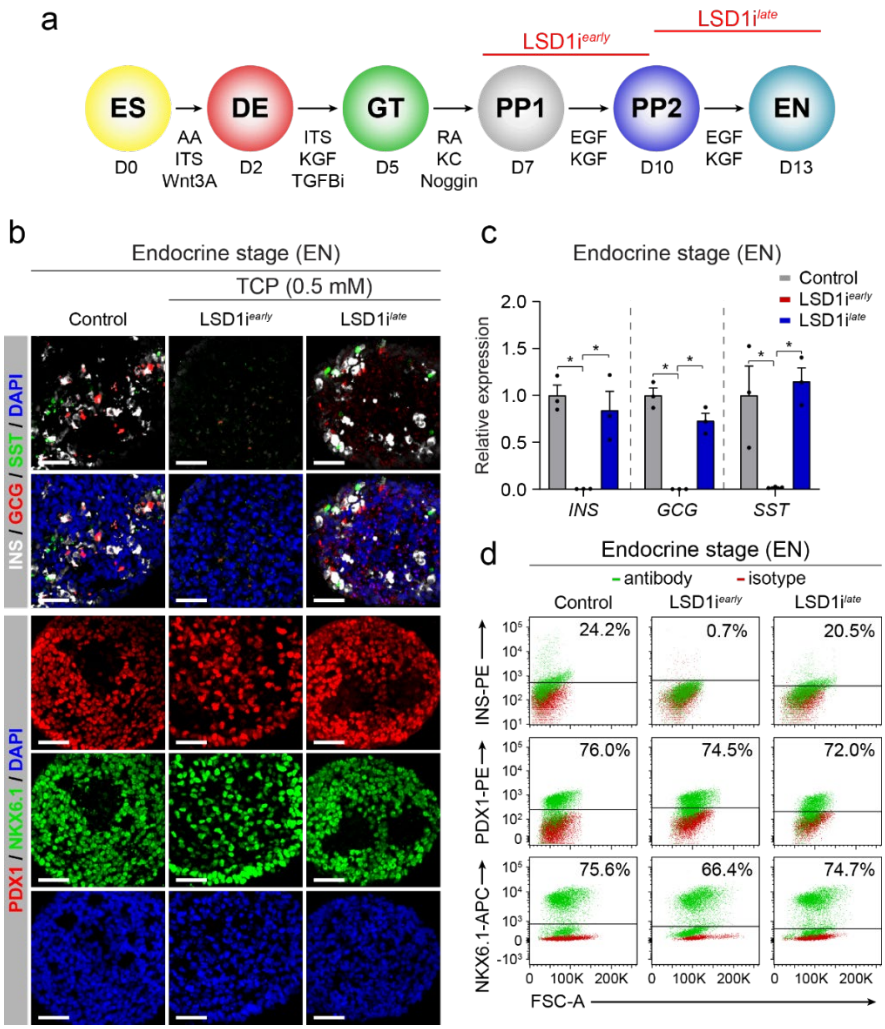

Figure 2.1. Endocrine cell formation requires LSD1 activity during a short window in early pancreatic development. a, Schematic of the human embryonic stem cell (hESC) differentiation protocol to the endocrine cell stage (EN) and experimental plan for LSD1 inhibition. **b**, Immunofluorescent staining for pancreatic hormones insulin (INS), glucagon (GCG) and somatostatin (SST) or PDX1 and NKX6.1 in control EN cells compared to EN cells with early (LSD1i^{early}) and late (LSD1i^{late}) LSD1 inhibition (representative images, n = 10 independent differentiations). Scale bar, 50 µm. **c**, qRT-PCR analysis for *INS*, *GCG* and *SST* in control, LSD1i^{early} and LSD1i^{late} EN cells. Data are shown as mean ± S.E.M. (n = 3 replicates from independent differentiations with n = 3 technical replicates per sample; source data are provided as a Source Data file). p = 7.93 e-4, 1.42 e-2, 2.32 e-4, 8.71 e-4, 3.5 e-2, and 1.52 e-3, respectively, Student's t-test, 2 sided. **d**, Flow cytometry analysis at EN stage for NKX6.1, PDX1 and INS comparing control, LSD1i^{early} and LSD1i^{late} cells. Isotype control for each antibody is shown in red and target protein staining in green. Percentage of cells expressing each protein is indicated (representative experiment, n = 2 independent differentiations). D, day; AA, activin A; ITS, insulin-transferrin-selenium; TGFBi, TGFβ R1 kinase inhibitor; KC, KAAD-cyclopamine; KGF, keratinocyte growth factor; RA, retinoic acid; EGF, epidermal growth factor; ES, human embryonic stem cells; DE, definitive endoderm; GT, primitive gut tube; PP1, early pancreatic progenitors; PP2, late pancreatic progenitors; EN, endocrine cell stage; FSC-A, forward scatter area.

To assess whether LSD1 is required for pancreatic development, we started by blocking LSD1 activity immediately after the initiation of pancreas induction during the transition from the early (PP1) to the late (PP2) pancreatic progenitor cell stage (LSD1i^{early}), using the irreversible LSD1 inhibitor tranylcypromine (TCP) (**Fig. 2.1a**). PP1 and PP2 progenitors are distinguished by increasing expression of pancreatic TFs that commit progenitors to the endocrine cell fate, including NKX6.1 and NGN3[67]. Thus, PP1 cells represent a less committed pancreatic progenitor cell stage, whereas PP2 cells exhibit features of endocrine cell commitment. LSD1 inhibition during the PP1 to PP2 transition did not negatively affect expression of PDX1, NKX6.1, or NGN3 (**Supplementary Figure 2.1e-h**), indicating that endocrine-committed pancreatic progenitors can form in the absence of LSD1. However, when LSD1i^{early} cells were further differentiated to the EN stage, we observed a striking absence of endocrine cells at the EN stage, while progenitor cell markers remained largely unaffected (**Fig. 2.1b-d** and **Supplementary Figure 2.2**). The same phenotype was observed when culturing in the presence of several other irreversible and reversible LSD1 inhibitors during the PP1 to PP2 transition or by transducing cells with a lentivirus expressing shRNAs for *LSD1* a day prior to the PP1 stage (**Supplementary Figure 2.3a-d** and **Supplementary Figure 2.4a-c**). The normal progression through endocrine commitment but absence of endocrine cells after LSD1 inhibition indicated a specific requirement for LSD1 activity during endocrine cell differentiation. To directly test whether the endocrine cell differentiation step requires LSD1 activity, we added TCP or the LSD1 inhibitor GSK2879552 during the PP2 to EN transition (LSD1i^{late}). Surprisingly, this later inhibition of LSD1 did not perturb endocrine cell formation (**Fig. 2.1b-d** and **Supplementary Figure 2.3b-d**). Thus, endocrine cell development requires LSD1 activity during a narrow time window after pancreas induction, but not during endocrine cell differentiation. This indicates that LSD1-mediated changes during the PP1 to PP2 transition affect the ability of developmental precursors to undergo endocrine differentiation later in development.

LSD1 Inhibition Prevents Enhancer Silencing

Given LSD1's role as a chromatin modifier[106], we investigated whether loss of LSD1 activity during the PP1 to PP2 transition could block endocrine cell development due to aberrant regulation of the epigenome. To this end, we performed chromatin immunoprecipitation sequencing (ChIP-seq) for LSD1 at the PP1 stage and mapped chromatin state changes at LSD1-bound sites during the PP1 to PP2 transition without and with LSD1 inhibition. We identified a total of 15,084 LSD1 peaks at the PP1 stage throughout the genome (**Supplementary Figure 2.5a**). Of these, the vast majority were promoter-distal (11,799; > 3kb from TSS; **Supplementary Figure 2.5a**), which is consistent with prior observations in hESCs[107]. Distal LSD1 peaks at PP1 overlapped with binding sites for the early pancreatic TFs FOXA1, FOXA2, GATA4, GATA6, and HNF6, suggesting that these TFs reside in a complex with LSD1 (**Supplementary Figure 2.5b,c)**.

Distal enhancers are highly dynamic during pancreatic development[27], leading us to postulate that LSD1 controls endocrine cell differentiation by regulating changes in enhancer chromatin state during the PP1 to PP2 transition. To test this, we performed ChIP-seq for the active enhancer mark H3K27ac[104,113,114] at the PP1 and PP2 stage. Reasoning that effects of LSD1 on the active enhancer landscape would be most likely to affect gene expression and therefore have high propensity to be causal for the phenotype, we isolated enhancers that are active at PP1 and/or PP2 and also bound by LSD1 at the PP1 stage. This analysis revealed three groups of LSD1-bound enhancers: Group 1 (G1) enhancers (n = 1345) underwent deactivation during the PP1 to PP2 transition (≥ 2-fold decrease in H3K27ac); Group 2 (G2) enhancers (n = 765) were active at both PP1 and PP2 (< 2-fold change in H3K27ac); and Group 3 (G3) (n = 511) enhancers underwent activation (≥ 2-fold increase in H3K27ac) during the PP1 to PP2 transition (**Fig. 2.2a**). We next examined the "poised" chromatin modifications H3K4me1 and H3K4me2 at these three enhancer groups during the PP1 to PP2 transition. We observed that LSD1-bound G1 enhancers exhibited a marked decrease in H3K4me1 and H3K4me2 (**Fig. 2.2b**), consistent with known roles of LSD1 as a H3K4me2 and H3K4me1 demethylase[106,115]. Thus, G1 enhancers

are decommissioned during the PP1 to PP2 transition. To investigate whether LSD1 recruitment is regulated at these decommissioned enhancers, we examined LSD1 occupancy also in gut tube (GT) and PP2 cells. At G1 enhancers, we observed an increase in LSD1 ChIP-seq signal from GT to PP1 and a decrease from PP1 to PP2, whereas LSD1 signal was similar at all stages in the G2 and G3 enhancer groups (**Fig. 2.2c**). Since endocrine cell development requires LSD1 activity during the PP1 to PP2, but not the PP2 to EN transition, the transient recruitment of LSD1 to G1 enhancers at the PP1 stage could signify a specific importance of this enhancer group for the endocrine differentiation phenotype.

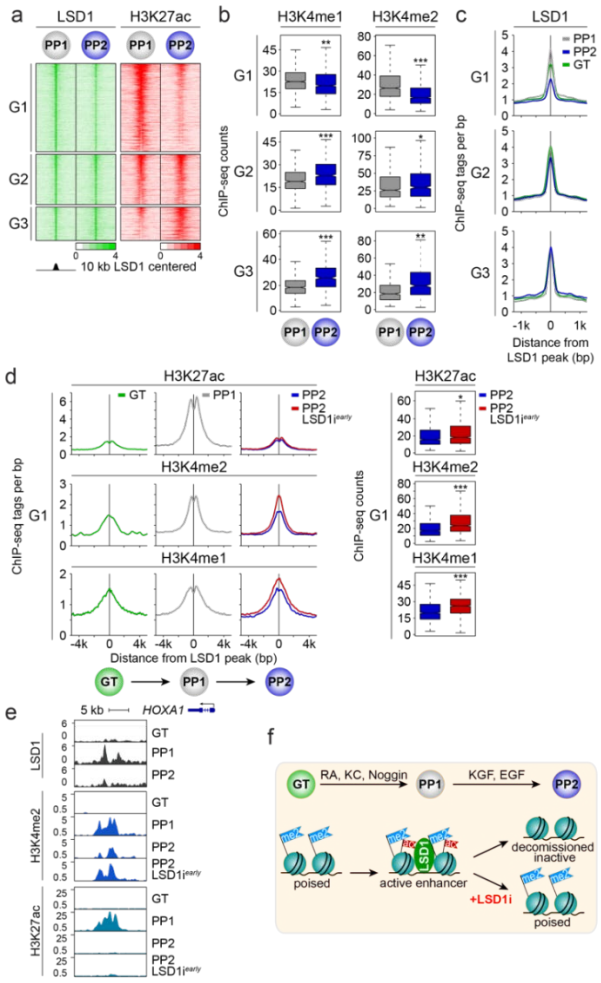

Figure 2.2. LSD1 inhibition prevents decommissioning of transiently active early pancreatic enhancers. a, Heatmap showing density of ChIP-seq reads for LSD1 and H3K27ac centred on LSD1 peaks, spanning 10 kb. G1, G2 and G3 groups of LSD1-bound enhancers are deactivated (G1; n=1345), remain active (G2; n=765), or are activated (G3; n=511) from PP1 to PP2. **b**, Box plots of H3K4me1 and H3K4me2 ChIP-seq counts at G1, G2 and G3 enhancers at PP1 and PP2 stages. Plots are centred on median, with box encompassing 25th-75th percentile and whiskers extending up to 1.5 interquartile range (Tukey style). P = 5.0 e-12, < 2.2 e-16, < 2.2 e-16, 1.73 e-2, < 2.2 e-16, and 8.23 e-14, respectively, Wilcoxon rank-sum test, 2 sided. **c**, Tag density plots displaying LSD1 tag distribution at G1, G2 and G3 enhancers at GT, PP1 and PP2 stages, centred on PP1 LSD1 peaks. **d**, Tag density plots (left) for G1 enhancers displaying H3K27ac, H3K4me2 and H3K4me1 tag distribution at GT and PP1 stages, and at PP2 stage with and without early LSD1 inhibition (TCP, LSD1i^{early}). Plots are centred on PP1 LSD1 peaks. Box plots (right) of H3K27ac, H3K4me2 and H3K4me1 ChIP-seq counts at G1 enhancers at PP2 stage with and without early LSD1 inhibition (LSD1i^{early}). Plots are centred on median, with box encompassing 25th-75th percentile and whiskers extending up to 1.5 interquartile range (Tukey style). P = 4.59 e-5, < 2.2 e-16, and < 2.2 e-16, respectively, Wilcoxon rank-sum test, 2 sided. **e**, LSD1, H3K4me2, and H3K27ac ChIP-seq profiles at an enhancer near *HOXA1*. **f**, Model for LSD1-dependent enhancer decommissioning. Enhancer deactivation by removal of acetylation from H3K27 occurs independent of LSD1 activity. LSD1 subsequently mediates enhancer decommissioning by removal of H3K4me2 marks. KC, KAAD-cyclopamine; KGF, keratinocyte growth factor; RA, retinoic acid; EGF, epidermal growth factor. GT, primitive gut tube; PP1, early pancreatic progenitors; PP2, late pancreatic progenitors. All ChIP-seq experiments, n = 2 replicates from independent differentiations.

To determine whether LSD1 activity is required for remodelling enhancer chromatin during the PP1 to PP2 transition, we analysed H3K27ac, H3K4me2 and H3K4me1 modifications in PP2 cells after LSD1 inhibition (LSD1i^{early}). In all three enhancer clusters, we observed little effect of LSD1 inhibition on H3K27ac dynamics during the PP1 to PP2 transition (**Fig. 2.2d,e** and **Supplementary Figure 2.5d**). The activation (i.e. acetylation) of G1 enhancers coincided with the pancreas induction step from GT to PP1 (**Fig. 2.2d**,**e** and **Supplementary Figure 2.5e**). Confirming our prior observation that pancreas-specific enhancers are poised prior to activation[27], G1 enhancers exhibited significant deposition of H3K4me1 at the GT stage (**Fig. 2.2d**). Thus, G1 enhancers become activated during pancreas induction and are quickly fully inactivated (i.e. decommissioned) as pancreatic endocrine development proceeds. Consistent with LSD1's enzymatic activity, LSD1 inhibition during the PP1 to PP2 transition led to significant accumulation of H3K4me1 and H3K4me2, particularly at G1 enhancers (**Fig. 2.2d**,**e** and **Supplementary Figure 2.5e**; $p < 2.2e\text{-}16$, Wilcoxon rank-sum test). H3K4me1 and H3K4me2 levels at G1 enhancers in LSD1i^{early} PP2 cells were similar to levels at PP1, showing a requirement for LSD1 in decommissioning these enhancers during the PP1 to PP2 transition. Although H3K4me1 and H3K4me2 levels were also increased at G2 and G3 enhancers after LSD1 inhibition, the effect was less pronounced compared to G1 enhancers (**Supplementary Figure 2.5d**). Importantly, H3K4me1 and H3K4me2 deposition was not increased at enhancers not bound by LSD1 (**Supplementary Figure 2.5f**), demonstrating specificity of the effect to LSD1-bound enhancers. Combined, this analysis identified a LSD1-regulated set of enhancers that is activated upon addition of pancreas-inductive factors during the GT to PP1 transition and deacetylated and decommissioned (i.e. demethylated) when these factors are withdrawn from PP1 to PP2 (**Fig. 2.2f**). We find that deacetylation of these enhancers occurs largely independent of LSD1, but that LSD1 is required for enhancer decommissioning and thus complete enhancer silencing. Given prior findings that LSD1 activity is inhibited in context of acetylated histones[108], these results suggest that histone acetylation from GT to PP1 prevents LSD1-mediated enhancer silencing and

that LSD1-independent H3K27ac removal allows LSD1 to silence these enhancers during the PP1 to PP2 transition.

LSD1 Represses Transiently Expressed, RA-Dependent Genes

We next sought to investigate possible effects of the observed chromatin changes on gene expression and compared RNA-seq profiles of control PP2 cells and PP2 cells after LSD1 inhibition (LSD1i^{early}). This analysis identified 445 genes that decreased and 955 genes that increased in expression due to LSD1 inhibition (**Fig. 2.3a**; $p < 0.05$, ≥ 1.5-fold change). To identify those genes most likely directly regulated by LSD1, we performed enrichment analysis for G1, G2, and G3 enhancers as well as other distal LSD1 binding sites near genes up- and down-regulated after LSD1 inhibition (TSS ± 100 kb from LSD1 peak). G1, G2, and G3 enhancers, but not other distal LSD1 binding sites, showed significant enrichment close to genes up-regulated due to LSD1 inhibition (**Fig. 2.3b**). The majority of the enhancer-associated up-regulated genes were near G1 enhancers (**Fig. 2.3c**). By contrast, we observed significant depletion or lack of enrichment of distal LSD1-bound sites near genes down-regulated in LSD1i^{early} cells (**Supplementary Figure 2.6a**). Together, this analysis suggests that direct LSD1 target genes are overrepresented among genes up-regulated after LSD1 inhibition, whereas down-regulated genes are not directly LSD1-regulated.

We next determined how candidate G1, G2 and G3 enhancer target genes are regulated over the developmental time course (**Fig. 2.3d** and **Supplementary Figure 2.6b**,**c**). G1 enhancer-associated genes that were up-regulated by LSD1 inhibition were induced during the GT to PP1 transition and then down-regulated during the transition to PP2 (**Fig. 2.3d**). Thus, the expression pattern of G1 enhancer-associated, LSD1-regulated genes mirrors the acetylation pattern of G1 enhancers, which are not acetylated at the GT stage, acetylated at the PP1 stage, and LSD1-dependently decommissioned during the PP1 to PP2 transition (**Fig. 2.2d**). Unbiased analysis of over-represented pathways among genes up-regulated by LSD1 inhibition revealed

enrichment for genes linked to RA signalling in the G1, but not G2 or G3, enhancer-associated group of genes, suggesting an important role for RA signalling in the regulation of G1 enhancer-associated genes (**Fig. 2.3e** and **Supplementary Figure 2.6b**,**c**). To simulate the requirement for RA signalling in pancreatic lineage induction in vivo[116,117], RA is one of three growth factors added to the culture medium during the GT to PP1 transition to induce pancreatic genes (**Fig. 2.1a**)[67]. RA is subsequently withdrawn during the transition from PP1 to PP2. During the PP1 to PP2 transition, the only possible source for stimulation of the RA receptor (RAR) are traces of retinol in the B27 supplement. Thus, the activity of G1 enhancers and expression of associated genes precisely coincides with the addition and removal of exogenous RA.

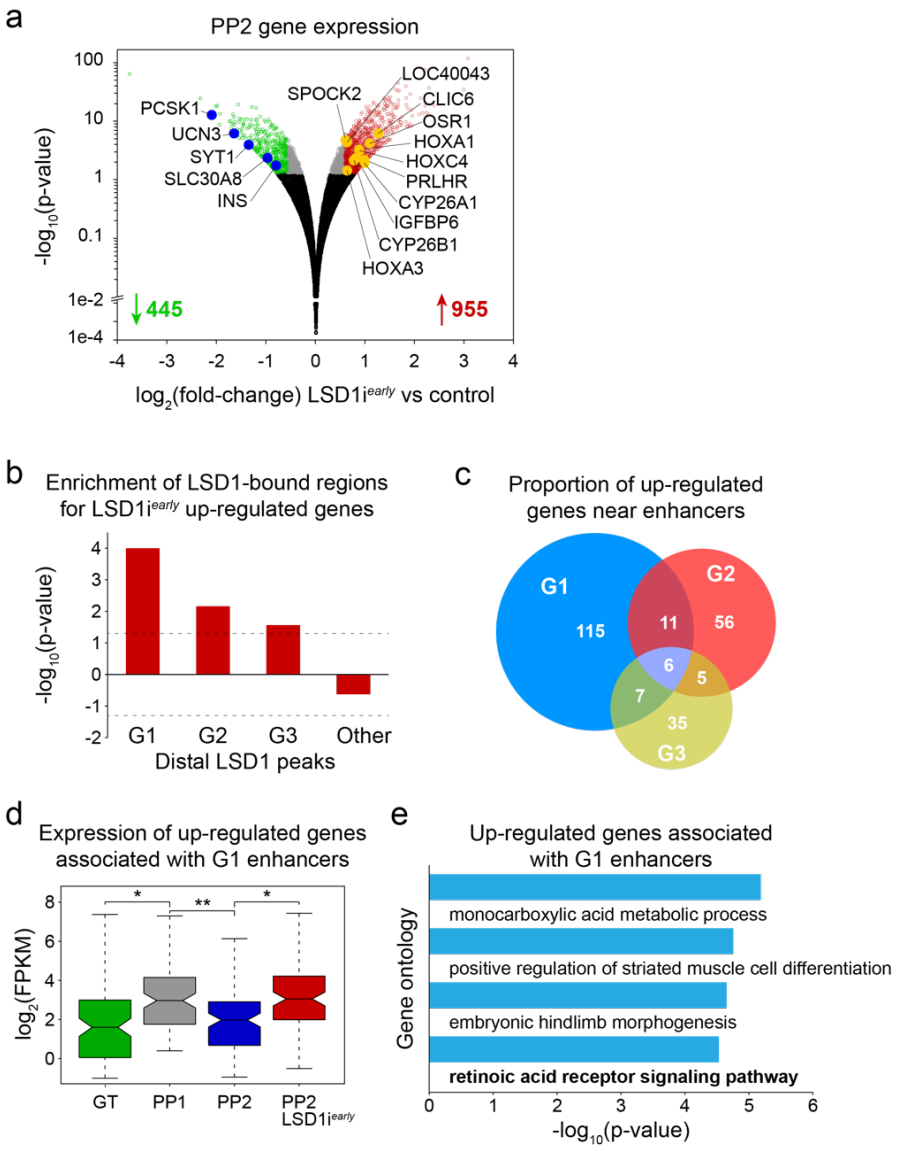

Figure 2.3. LSD1 activity is necessary for repressing transiently expressed retinoic acid-dependent genes. a, Volcano plot of differentially expressed genes at PP2 after LSD1 inhibition from PP1 to PP2 (TCP, LSD1i^{early}). Differential expression calculated with DESeq2 and genes with ≥ 1.5-fold change up or down. Adjusted p-values of < 0.05 were considered differentially expressed. 445 genes were down-regulated and 955 were up-regulated in LSD1i^{early} PP2 cells. Black dots indicate genes not significantly changed (p-value > 0.05), grey dots genes significantly changed (p-value < 0.05) but less than 1.5-fold compared to control, red and green dots genes significantly up- and down-regulated (p-value < 0.05 and ≥ 1.5-fold change), respectively (n = 2 replicates from independent differentiations). **b**, Enrichment analysis of genes up-regulated by LSD1i^{early} within 100kb of G1, G2 and G3 or other distal LSD1 peaks. Dashed lines indicate p-value = 0.05 for enrichment (positive value) or depletion (negative value), permutation test. **c**, Percentage of LSD1i^{early} up-regulated genes near G1 (n=139), G2 (n=78) and G3 (n=53) enhancers (within 100kb). **d**, Box plot of mRNA levels for 139 LSD1i^{early} up-regulated genes near G1 enhancers. Plots are centred on median, with box encompassing 25th-75th percentile and whiskers extending up to 1.5 interquartile range (Tukey style). P = 2.30 e-3, 4.38 e-6, and 2.25 e-7, respectively, Wilcoxon rank-sum test, 2 sided. **e**, Gene ontology analysis for 139 LSD1i^{early} up-regulated genes near G1 enhancers. GT, primitive gut tube; PP1, early pancreatic progenitors; PP2, late pancreatic progenitors.

Prolonged Exposure to RA Phenocopies LSD1 Inhibition

RA regulates gene expression by binding to its heterodimeric receptor composed of RAR and retinoid X receptor (RXR)[118]. In the absence of RA, the RAR/RXR heterodimer recruits co-repressors leading to histone deacetylation and gene silencing, while RA binding to RAR/RXR induces recruitment of HATs, mediating histone acetylation and activation of RA-dependent genes. Hence, the observed pattern of H3K27 acetylation at G1 enhancers during progression from GT to PP2 (**Fig. 2.2d**) is consistent with RA-dependent regulation of these enhancers. To determine whether G1 enhancers are indeed regulated by RA, we performed TF binding motif enrichment analysis. This analysis revealed significant enrichment of the motif for the RAR/RXR heterodimer at G1 compared to G2 and G3 enhancers (**Fig. 2.4a**). When motifs at G2 and G3 enhancers were compared against the entire genome, excluding G1, G2, and G3 regions, no RAR/RXR motif enrichment was observed, further supporting specific enrichment of the RAR/RXR motif at G1 enhancers. ChIP-seq analysis for RXR, which is the obligatory binding partner for all RAR isoforms, confirmed RXR binding to G1 enhancers at the PP1 stage (**Fig. 2.4b** and **Supplementary Figure 2.5c**). RXR binding was enriched at G1 enhancers when compared to either G2 or G3 enhancers (Fisher's exact test, p=1.728 e-8 and p=9.427 e-15, respectively), indicating RA-dependent regulation particularly of G1 enhancers.

To determine whether failure to silence RA-induced genes could be the mechanism by which LSD1 inhibition blocks endocrine cell differentiation, we tested whether extended RA exposure of pancreatic progenitors abrogates endocrine cell differentiation in a similar manner as LSD1 inhibition. To extend the time period of RA signalling, we added the RA analogue TTNPB not only from the GT to PP1 stage, but also during the PP1 to PP2 transition (**Fig. 2.4c**, $RA^{extended}$) and then differentiated $RA^{extended}$ cultures to the EN stage. Mimicking the $LSD1i^{early}$ phenotype, EN stage $RA^{extended}$ cultures exhibited a striking absence of endocrine cells, while the progenitor cell markers PDX1, NKX6.1, and NGN3 were unaffected (**Fig. 2.4d**,**e** and **Supplementary Figure**

2.7a-d). Thus, RA exposure of pancreatic progenitors has to be transient for endocrine cell differentiation to occur.

We next employed RNA-seq analysis to identify genes dysregulated as a result of prolonged RA exposure. Only 96 genes were down-regulated and 69 up-regulated in RA^{extended} PP2 cultures (**Fig. 2.4f**; $p < 0.05$, ≥ 1.5-fold change), suggesting that the endocrine differentiation block is mediated by dysregulation of a modest number of genes. Consistent with the small number of dysregulated genes in RA^{extended} cells compared to LSD1i^{early} cells, RA^{extended} cells more closely resembled untreated PP2 cells than LSD1i^{early} cells (**Supplementary Figure 2.6e**). Strikingly, genes up-regulated after LSD1 inhibition were significantly enriched among the genes also increased in expression after prolonged RA exposure (**Fig. 2.4f**), and genes near G1 enhancers largely accounted for this enrichment (**Fig. 2.4g**). Genes down-regulated after LSD1 inhibition were likewise enriched among genes decreased in expression after prolonged RA, but LSD1-occupied enhancers were not enriched in the vicinity of these genes (**Supplementary Figure 2.7f**), suggesting indirect effects. Among the genes up-regulated by both LSD1 inhibition and prolonged RA exposure were numerous genes known to be regulated by RA, including genes encoding HOX TFs and RA-inactivating enzymes of the CYP26 family (**Fig. 2.4h**), supporting the notion that RA-induced genes need to be silenced for cells to acquire competence for endocrine cell differentiation. Together, our findings support a model whereby LSD1 silences RA-regulated genes by decommissioning their enhancers upon RA withdrawal, thereby ensuring transient, ligand-dependent expression of RA-induced genes. Under conditions of prolonged RA exposure, RA-mediated recruitment of HATs maintains histone acetylation[118], which inhibits LSD1 activity[108] and prevents enhancer decommissioning during the PP1 to PP2 transition.

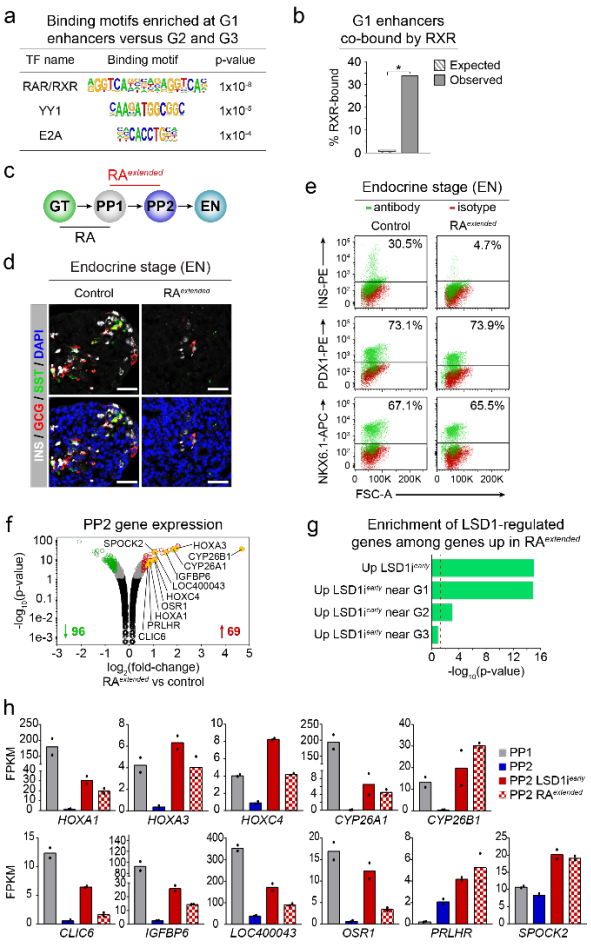

Figure 2.4. Prolonged retinoic acid exposure of early pancreatic progenitor cells phenocopies LSD1 inhibition. a, Enriched transcription factor (TF) binding motifs with associated p-values for G1 enhancers compared to G2 and G3 enhancers. Fisher's exact test, 2 sided, corrected for multiple comparisons. **b**, Enrichment for RXR peaks (± 1kb) among G1 enhancers versus random genomic regions. *p = 0, permutation test. **c**, Experimental plan to extend retinoic acid (RA) exposure through PP1 to PP2 ($RA^{extended}$) during hESC differentiation to the endocrine cell stage (EN). **d**, Immunofluorescent staining for insulin (INS), glucagon (GCG) and somatostatin (SST) in control EN cells compared to EN cells with extended RA treatment ($RA^{extended}$) (representative images, n = 3 independent differentiations). Scale bar, 50 µm. **e**, Flow cytometry analysis at EN stage for NKX6.1, PDX1 and INS comparing control and $RA^{extended}$ cultures. Isotype control for each antibody is shown in red and target protein staining in green. Percentage of cells expressing each protein is indicated (representative experiment, n = 2 independent differentiations). **f**, Volcano plot of differentially expressed genes at PP2 in $RA^{extended}$ cultures. Differential expression calculated with DESeq2 and genes with ≥ 1.5-fold change up or down. Adjusted p-values < 0.05 were considered differentially expressed. 96 genes were down-regulated and 69 were up-regulated in $RA^{extended}$ cultures. Black dots indicate genes not significantly changed (p-value > 0.05), grey dots genes significantly changed (p-value < 0.05) but less than 1.5-fold compared to control, red and green dots genes significantly up- and down-regulated (p-value < 0.05 and ≥ 1.5-fold change), respectively. Yellow dots highlight genes also up-regulated after LSD1 inhibition from PP1 to PP2 (TCP, $LSD1i^{early}$) (n = 2 replicates from independent differentiations). **g**, Enrichment analysis of genes associated with LSD1-bound enhancers and up-regulated by $LSD1i^{early}$ among those up-regulated by $RA^{extended}$. Dashed line indicates p-value = 0.05, Fisher's exact test, 2 sided. **h**, mRNA levels of select genes significantly up-regulated in both $LSD1i^{early}$ and $RA^{extended}$ PP2 cells. Levels at PP1 stage are also displayed. Data shown as mean FPKM ± S.E.M. (n = 2 replicates from independent differentiations; source data are provided as a Source Data file). GT, primitive gut tube; PP1, early pancreatic progenitors; PP2, late pancreatic progenitors.

LSD1 Dampens Future Responses to RA

We sought to further substantiate that LSD1 regulates RA responsiveness and that the block in endocrine cell differentiation after LSD1 inhibition is linked to aberrant expression of RA-dependent genes. We predicted that endocrine cell differentiation should not be perturbed when cells are re-exposed to RA during the PP2 to EN transition, because enhancers of early RA-responsive genes are already decommissioned at the PP2 stage (**Fig. 2.2d**,**e** and **Supplementary Figure 2.5d**). To test this, we re-introduced RA into the culture medium during the PP2 to EN transition (RA^{late}; **Fig. 2.5a**). As hypothesised, and in stark contrast to $RA^{extended}$ cultures (**Fig. 2.4d**,**e**), endocrine cells were present in RA^{late} EN stage cultures in numbers almost identical to control cultures (**Fig. 2.5b**,**c** and **Supplementary Figure 2.8a**,**b**). Thus, similar to addition of the LSD1 inhibitor (**Fig. 2.1a-d**), addition of RA prevents endocrine cell formation only during the PP1 to PP2 but not the PP2 to EN transition.

To further test whether enhancer decommissioning is a mechanism by which to regulate RA responsiveness, we re-exposed cells to RA from PP2 to EN (RA^{late}) with or without prior LSD1 inhibition during the PP1 to PP2 transition (**Fig. 2.5a**). As expected, $LSD1i^{early}$ + RA^{late} treatment completely blocked endocrine cell differentiation, phenocopying $LSD1i^{early}$ cultures (**Supplementary Figure 2.8c**,**d**). To determine whether prior LSD1 inhibition alters the extent to which RA-regulated genes can be induced by RA, we measured gene expression changes in response to RA. To this end, we compared gene expression at the EN stage in $LSD1i^{early}$ vs. $LSD1i^{early}$ + RA^{late} and control vs. RA^{late} conditions. We used $LSD1i^{early}$ EN cells, rather than control EN stage cultures, as a reference for the cells re-exposed to RA after $LSD1i^{early}$ treatment to control for the population bias caused by the lack of endocrine cells after LSD1 inhibition. Most genes near G1 enhancers that are up-regulated by both LSD1 inhibition (**Fig. 2.3a**) and extended RA exposure (**Fig. 2.4f**), including *HOXA1*, *HOXA3*, *HOXC4*, and *CYP26B1*, exhibited a higher degree of inducibility by RA with prior LSD1 inhibition (**Fig. 2.5d, Supplementary Figure 2.8e**). Thus, LSD1 appears to dampen, although not obliterate, future RA responsiveness in cells that

have been previously exposed to RA. We observed no difference in the expression of *LSD1*, *RARs*, or *RXRs* between the different conditions (**Supplementary Figure 2.8f**), indicating that alteration of the epigenetic state rather than differences in TF and co-factor expression explain the heightened RA responsiveness after LSD1 inhibition. We note that other factors must also control RA responsiveness since *HOXA1*, *HOXA3*, *HOXC4*, *CYP26A1,* and *CYP26B1* are still induced by late RA treatment without prior LSD1 inhibition.

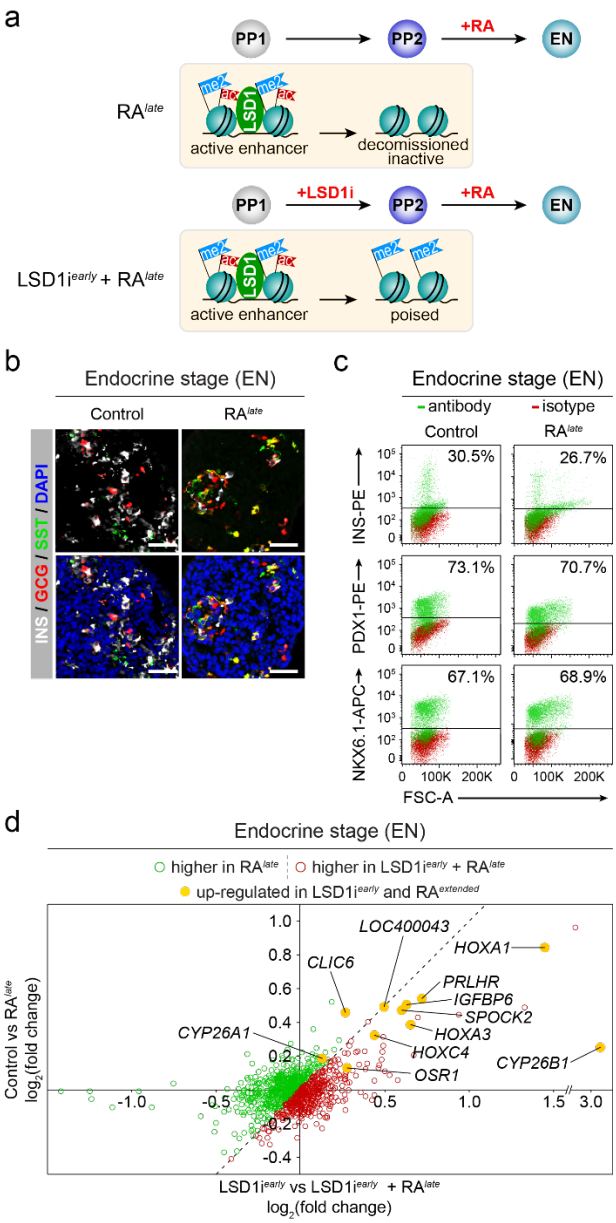

Figure 2.5. LSD1 decreases future inducibility of RA-dependent genes by retinoic acid. a, Experimental plan to re-introduce retinoic acid (RA) during the PP2 to endocrine (EN) transition of hESC differentiation without (RA^{late}) and with prior inhibition of LSD1 ($LSD1i^{early}$ + RA^{late}). The cartoon depicts the chromatin state as determined in Figure 2. **b**, Immunofluorescent staining for insulin (INS), glucagon (GCG) and somatostatin (SST) in control EN cells compared to EN cells with late RA treatment (RA^{late}) (representative images, n = 3 independent differentiations). Scale bar, 50 µm. **c**, Flow cytometry analysis at EN stage for NKX6.1, PDX1 and INS comparing control and RA^{late} cells. Isotype control for each antibody is shown in red and target protein staining in green. Percentage of cells expressing each protein is indicated (representative experiment, n = 2 independent differentiations). **d**, Gene expression changes in RA^{late} vs control EN cells compared to gene expression changes in $LSD1i^{early}$ + RA^{late} vs $LSD1i^{early}$ EN cells (calculated with DESeq2). Green dots indicate genes more increased in RA^{late} and red dots indicate genes more increased in $LSD1i^{early}$ + RA^{late} compared to respective controls. Yellow dots highlight genes up-regulated after both $LSD1i^{early}$ and extended RA treatment ($RA^{extended}$) from PP1 to PP2. (n = 2 replicates from independent differentiations per condition). PP1, early pancreatic progenitors; PP2, late pancreatic progenitors; FSC-A, forward scatter area.

Lsd1 Is Required for Endocrine Cell Development In Vivo

To verify our in vitro findings in an in vivo model, we deleted *Lsd1* conditionally in mice to determine whether endocrine cell differentiation requires Lsd1 activity transiently in early pancreas development, as observed in the hESC differentiation system. As in the human pancreas (**Supplementary Figure 2.1d**), Lsd1 was highly expressed in pancreatic progenitors and endocrine cells (**Supplementary Figure 2.9a**,**b**). To selectively inactivate *Lsd1* in early pancreatic progenitors similar to LSD1 inhibition at PP1, we generated *Pdx1Cre*;*Lsd1*$^{flox/flox}$ (*Lsd1*$^{\Delta pan}$) mice (**Fig. 2.6a**). In *Lsd1*$^{\Delta pan}$ embryos, key aspects of early pancreatic development, such as the induction of early pancreatic markers and outgrowth of the tissue buds, were unperturbed (**Fig. 2.6b**,**c** and **Supplementary Figure 2.9c**). However, by embryonic day (e) 15.5, when widespread endocrine cell differentiation was evident in control mice, *Lsd1*$^{\Delta pan}$ embryos exhibited an almost complete lack of endocrine cells (**Fig. 2.6b**), a phenotype that remained apparent at postnatal day (P) 0 (**Fig. 2.6b**,**d**). In vivo inactivation of *Lsd1* further revealed that Lsd1 activity is selectively required for development of the endocrine lineage, while being dispensable for exocrine cell formation and key aspects of early pancreatic development, such as maintenance of pancreatic progenitors and growth of the developing organ (**Fig. 2.6b,c** and **Supplementary Figure 2.9c-e**). Analysis of the endocrine progenitor marker Ngn3 further revealed that endocrine lineage commitment was unaffected in *Lsd1*$^{\Delta pan}$ embryos (**Fig. 2.6b**,**c**). Thus, as in the hESC differentiation system, *Lsd1* inactivation in early pancreatic progenitors of mice prevents endocrine cell differentiation after endocrine fate commitment.

To determine when precisely Lsd1 is required for endocrine cell development, we crossed *Lsd1*$^{flox/flox}$ and *Pdx1CreER*^{TM} mice, allowing for temporally controlled *Lsd1* inactivation in pancreatic progenitors by tamoxifen administration (**Fig. 2.6e**). Consistent with the phenotype of *Lsd1*$^{\Delta pan}$ mice, tamoxifen administration at e10.5 (*Lsd1*$^{\Delta early}$ mice) resulted in almost complete absence of endocrine cells (**Fig. 2.6f** and **Supplementary Figure 2.9f**). Remaining endocrine cells in *Lsd1*$^{\Delta early}$ mice were mostly Lsd1^{+} due to mosaic deletion (**Supplementary Figure 2.9f,g**).

By contrast, tamoxifen injection at e12.5 (*Lsd1*$^{\Delta late}$ mice), which targets late pancreatic progenitors shortly before endocrine cell differentiation similar to PP2 cells, did not affect endocrine cell formation, as evidenced by the presence of Lsd1-deficient hormone^{+} cell clusters (**Fig. 2.6f** and **Supplementary Figure 2.9g**). Therefore, as in the hESC differentiation system, endocrine cell differentiation in mice requires Lsd1 activity during a narrow time window in early pancreas development. Furthermore, early pancreatic inactivation of *Lsd1* resulted in up-regulation of *HoxA1* transcripts (**Fig. 2.6g**) as observed in hESC-PP2 cells (**Fig. 2.3a**). Combined our in vitro and in vivo findings support a model whereby LSD1 controls progression to the endocrine cell stage by limiting the duration of early RA signalling through chromatin modification at RA-responsive enhancers.

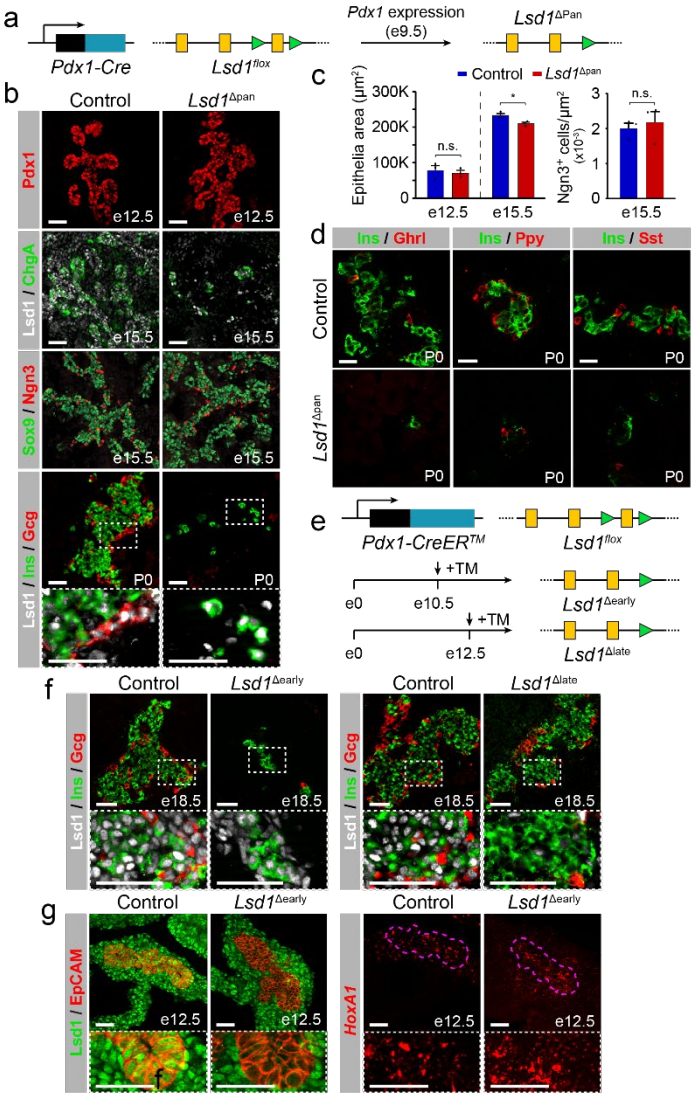

Figure 2.6. Lsd1 is required for endocrine cell formation in mice during a short window in early pancreatic development. a, Strategy for conditional *Lsd1* deletion in embryonic pancreatic progenitors of mice (*Lsd1*$^{\Delta pan}$ mice). Yellow boxes: exons; green triangles: *loxP* sites. **b**, Immunofluorescent staining for Pdx1 at embryonic day (e) 12.5, Lsd1 and chromogranin A (ChgA) or Sox9 and Ngn3 at e15.5, and Lsd1, insulin (Ins) and glucagon (Gcg) at postnatal day (P) 0 in control and *Lsd1*$^{\Delta pan}$ mice (representative images, n=6 embryos per genotype). Boxed areas are shown in higher magnification. Scale bar, 50 μm. **c**, Quantification of pancreatic epithelial area at e12.5 and e15.5 and Ngn3^{+} cells relative to epithelial area. Data are shown as means ± S.E.M. (n=3 embryos per genotype; source data are provided as a Source Data file). P = 0.65, 0.02, and 0.39, respectively, Student's t-test, 2 sided. **d**, Immunofluorescent staining for Ins with somatostatin (Sst), pancreatic polypeptide (Ppy) and ghrelin (Ghrl) at P0 in control and *Lsd1*$^{\Delta pan}$ mice. Scale bar, 25 μm. **e**, Strategy for tamoxifen-inducible *Lsd1* deletion in embryonic pancreatic progenitors of mice at e10.5 (*Lsd1*$^{\Delta early}$) and e12.5 (*Lsd1*$^{\Delta late}$). Yellow boxes: exons; green triangles: *loxP* sites. **f**, Immunofluorescent staining for Lsd1, Ins and Gcg at e18.5 in control, *Lsd1*$^{\Delta early}$ and *Lsd1*$^{\Delta late}$ mice (representative images, n=3 embryos per genotype). Boxed areas are shown in higher magnification. Scale bar, 50 μm. **g**, Immunofluorescent staining for Lsd1 and EpCAM (left panels) and RNAscope in situ hybridization for *HoxA1* (right panels, adjacent sections to left panels) in control and *Lsd1*$^{\Delta early}$ embryos at e12.5 (representative images, n=3 *Lsd1*$^{\Delta early}$ embryos and n=6 control embryos). Bottom images show higher magnification. Scale bar, 50 μm.

Discussion

Proper formation of terminally differentiated cell types requires precise timing, amplitude, and, as suggested by this study, duration of developmental signals. Our findings show that decommissioning (i.e. complete inactivation) of RA-dependent early pancreatic enhancers temporally limits the expression of RA-induced genes after RA is removed, thereby creating a sharp and transient gene expression response to RA. This mechanism can explain how the duration of a transcriptional response to a transient inductive signal is limited during a developmental time course. Furthermore, we observed that enhancer decommissioning dampens future gene inducibility by the same signal. Our findings suggest that developmental competence windows are terminated through erasure of epigenetic marks, providing a mechanistic understanding of why developmental signals evoke context-dependent cellular responses. Underscoring the importance of enhancer silencing for developmental progression, we find that the inability to decommission RA-dependent enhancers and down-regulate RA-induced genes coincides with subsequent failure to initiate endocrine cell differentiation. We propose that LSD1-mediated enhancer decommissioning is a responder tissue-intrinsic mechanism by which perduring transcriptional effects of transient developmental signals are prevented. By helping close competence windows during rapid developmental transitions, this mechanism could help create a conducive state for correct interpretation of the next inductive signal.

Our findings indicate that LSD1-mediated silencing of a subset of RA-dependent early pancreatic genes is necessary for pancreatic progenitors to acquire competence for endocrine cell differentiation. We identified a set of LSD1-regulated genes, including many *HOX* genes and other genes known to be directly regulated by RA, that are transiently expressed after pancreas induction and quickly down-regulated as cells transition to an endocrine-committed pancreatic progenitor cell state, marked by NKX6.1 and NGN3. Several observations suggest that the mechanism by which LSD1 controls endocrine cell development is to silence these RA-regulated genes prior to endocrine cell differentiation. First, we observed that endocrine cell formation in

both the hESC-based system and in mice proceeds normally when LSD1 is inhibited or deleted during endocrine cell differentiation, implying that LSD1 has no immediate role in the activation of endocrine genes. Second, like LSD1 inhibition, extended exposure of uncommitted hESC-pancreatic progenitors to RA blocks endocrine cell differentiation, whereas RA exposure during endocrine cell differentiation does not. Similarly, exposure of mouse embryonic pancreatic explants to high RA concentrations prior to endocrine cell differentiation has been shown to impair endocrine cell formation[119]. Furthermore, we showed that after both early LSD1 inhibition and extended RA exposure of hESC-pancreatic progenitors, RA-induced genes are up-regulated, which indicates that dysregulation of these genes is relevant for the phenotype. Third, *Lsd1* deletion in pancreatic progenitors of mice has no overt effect on exocrine cell differentiation or other aspects of pancreatic development, suggesting that Lsd1-mediated silencing of these RA-regulated genes occurs specifically during endocrine lineage progression. Consistent with this notion, the expression of HOXA1 is maintained in pancreatic exocrine cells[120]. Similar to pancreas development, LSD1 is expressed in the developing nervous system and plays an important role in neural differentiation[121,122]. Given that RA has time-dependent roles during different phases of neural development[123], it is possible that a similar connection between LSD1 and the regulation of RA signalling windows exists during neurogenesis.

Our findings suggest that the presence of LSD1 in enhancer complexes enables silencing of RA-induced genes after withdrawal of exogenous RA. In the context of LSD1 inhibition, silencing of RA-dependent genes does not occur despite RA removal, showing that absence of the RA signal is not sufficient for gene inactivation. RA is known to maintain histone acetylation by mediating recruitment of HATs, while removal of RA results in a co-factor switch leading to histone deacetylation[118]. Because presence of acetylated histones inhibits LSD1 activity[108], LSD1-mediated enhancer decommissioning is prevented as long as RA is present, linking enhancer decommissioning and target gene silencing to RA withdrawal. Thus, presence of LSD1 in enhancer complexes provides a cell-intrinsic epigenetic mechanism that couples the duration of

gene expression to the presence of the extrinsic signal. It is possible that LSD1 at promoters also contributes to gene regulation after LSD1 inhibition. However, TSSs of the majority of RA-induced genes that were up-regulated after LSD1 inhibition (e.g. *HOXA1 HOXA3, HOXC4, CYP26A1*, *CYP26B1*) were not LSD1-bound, suggesting that effects on gene expression are mediated by distal enhancers.

We show that inhibition of LSD1 activity during developmental progression coincides with increased H3K4me2/me1 deposition and increased expression of RA-dependent genes after RA removal. This raises the question of whether the catalytic activity of LSD1 and increased H3K4me2/me1 levels or demethylase-independent functions of LSD1 are responsible for the increased expression of RA-dependent genes. Our data suggest that the catalytic activity of LSD1 is indeed important, because TCP and other LSD1 inhibitors inhibit LSD1's catalytic activity. Furthermore, control and LSD1i^{early} PP2 cells exhibited no difference in LSD1 recruitment to G1 enhancers near genes up-regulated after LSD1 inhibition (data not shown). Of note, a recent study in ESCs showed that H3K4me1 is not required for transcription of nearby genes in the context of acetylated enhancers[124]. Our findings suggest that this is not true in the context of deacetylated enhancers, where H3K4me2/me1 appears to provide an enhancer activation memory that has impact on gene expression. Our observations are consistent with evidence that H3K4me2 deposition is dependent on transcription at enhancers, and that enhancer RNA transcripts correlate with the expression of nearby genes[125]. The exact contexts in which H3K4me2/me1 can drive gene expression and the mechanisms employed will require further studies.

We find that in addition to RAR motifs, LSD1-regulated enhancers are also enriched for FOXA and GATA motifs and that FOXA1, FOXA2, GATA4, and GATA6 binding is enriched at LSD1-occupied enhancers (**Supplementary Figure 2.5b**). FOXA1/2 and GATA4/6 have known functions in early pancreas development[126-128]. This indicates that regulation of early pancreatic enhancers requires collaborative interactions of signal-dependent RARs with other TFs that

regulate cell identity. We have previously shown that recruitment of FOXAs to pancreas enhancers precedes pancreas induction, and occurs prior to the addition of RA during hESC differentiation[27]. Before RA is added, FOXA-occupied pancreatic enhancers are poised. Combined with our present findings, this suggests that RA induces the pancreatic lineage by binding to RARs that act upon a set of primed enhancers established by FOXAs and other lineage-determining TFs. FOXAs have previously been shown to broadly prime enhancers of multiple gut tube-derived organs, including liver and lung[27]. While pancreas induction requires RA, lung and liver induction depend on BMP or BMP and WNT signalling, respectively. Therefore, collaboration between signal-dependent TFs and lineage-determining TFs, such as FOXAs, could be a broadly used mechanism to specify different organ lineages from a field of multipotent progenitors. Such mechanism is consistent with the importance of niche signals for specifying progenitor subdomains during development.

One open question is whether the here-described mechanism for LSD1-mediated enhancer silencing is limited to RA-dependent enhancers or could operate also in the context of other signal-dependent enhancers. LSD1 has been shown to reside in transcriptional complexes with signal-dependent TFs of numerous signalling pathways, including the Notch, Wnt, and multiple nuclear hormone receptor signalling pathways[129-131]. Similar to RARs, which associate with HATs upon RA binding[118], the Notch intracellular domain facilitates recruitment of HATs to Notch-responsive enhancers in the presence of ligand, and β-catenin recruits HATs when the Wnt signalling pathway is activated[132,133]. It is therefore conceivable that LSD1 limits the duration of a transcriptional response at these enhancers in a similar manner as shown in this study for RA-responsive enhancers. Such unified mechanism for LSD1 function would explain why LSD1 activity is required in numerous development contexts throughout phylogeny[134-138].

Methods

Cell Lines

CyT49 embryonic stem cells were maintained in DMEM F12 (without L-glutamine; VWR) + 10% knockout serum replacement (KSR; Thermo Fisher Scientific), 1% non-essential amino acids (NEAA; Thermo Fisher Scientific), 1% GlutaMAX (Thermo Fisher Scientific), 0.2% β-mercaptoethanol (Thermo Fisher Scientific) and 1% penicillin-streptomycin (Thermo Fisher Scientific). HEK293T were maintained in DMEM F12 containing 100 units/mL penicillin and 100 mg/mL streptomycin sulfate supplemented with 10% foetal bovine serum (FBS).

Animals

Pdx1-Cre, *Pdx1-CreER*^{TM} [139], *Lsd1*^{flox} [134], and *Rosa26-eYFP* [140] mouse strains have been described previously. *Lsd1*$^{\Delta pan}$ mice were generated by crossing *Pdx1-Cre* and *Lsd1*^{flox} mice. Conditional *Lsd1* knockouts were generated by crossing *Pdx1-CreER*^{TM} and *Lsd1*^{flox} mice. Tamoxifen (Sigma) was dissolved in corn oil (Sigma) at 10 mg/mL, and a single dose of 3.5 mg/40 g or 4.5 mg/40 g body weight was administered by intraperitoneal injection at embryonic day (e) 10.5 or e12.5, respectively. Control mice were LSD1$^{+/+}$ littermates carrying the *Pdx1-Cre* or the *Pdx1-CreER*^{TM} transgene. Midday on the day of vaginal plug appearance was considered embryonic day (e) 0.5. All animal experiments were approved by the Institutional Animal Care and Use Committees of the University of California, San Diego. The numbers of animals studied per genotype are indicated within each experiment.

Human Tissue

Human foetal pancreas donor tissue was obtained from the Birth Defects Research Laboratory of the University of Washington. Cadaveric adult pancreata used in this study were from non-diabetic donors and were acquired through the Network for Pancreatic Organ Donors with Diabetes (nPOD)[141]. Protein expression was analysed in nPOD donors: LSD1 and GCG in

#6140 (38-year-old male); LSD1 and CHGA in #6160 (22-year-old male); LSD1 and SST in 6178 (25-year-old female); and LSD1, INS and GCG in 6179 (21-year-old female). All studies for use of foetal and adult human tissue were approved by the Institutional Review Board of the University of California, San Diego.

Pancreatic Endocrine Differentiation of Human Embryonic Stem Cells (hESCs)

Pancreatic differentiation was performed as previously described[27,67,68]. Briefly, a suspension-based culture format was used to differentiate cells in aggregate form. Undifferentiated aggregates of hESCs were formed by re-suspending dissociated cells in hESC maintenance medium at a concentration of 1 x 10^6 cells/mL and plating 5.5 mL per well of the cell suspension in 6-well ultra-low attachment plates (Costar). The cells were cultured overnight on an orbital rotator (Innova2000, New Brunswick Scientific) at 95 rpm. After 24 hours the undifferentiated aggregates were washed once with RPMI medium and supplied with 5.5 mL of day 0 differentiation medium. Thereafter, cells were supplied with the fresh medium for the appropriate day of differentiation (see below). Cells were continually rotated at 95 rpm, or 105 rpm on days 4 through 8, and no media change was performed on day 10. Both RPMI (Mediatech) and DMEM High Glucose (HyClone) medium were supplemented with 1X GlutaMAX™ and 1% penicillin/streptomycin. Human activin A, mouse Wnt3a, human KGF, human noggin, and human EGF were purchased from R&D systems. Other added components included FBS (HyClone), B-27® supplement (Life Technologies), Insulin-Transferrin-Selenium (ITS; Life Technologies), TGFβ R1 kinase inhibitor IV (EMD Bioscience), KAAD-Cyclopamine (KC; Toronto Research Chemicals), and the retinoic receptor agonist TTNPB (RA; Sigma Aldrich). Day-specific differentiation media formulations were as follows: Days 0 and 1: RPMI + 0.2% (v/v) FBS, 100 ng/mL Activin, 50 ng/mL mouse Wnt3a, 1:5000 ITS. Days 1 and 2: RPMI + 0.2% (v/v) FBS, 100ng/mL Activin, 1:5000 ITS. Days 2 and 3: RPMI + 0.2% (v/v) FBS, 2.5 mM TGFβ R1 kinase inhibitor IV, 25ng/mL KGF, 1:1000 ITS. Days 3 – 5: RPMI + 0.2% (v/v) FBS, 25ng/mL KGF, 1:1000

ITS. Days 5 – 8: DMEM + 0.5X B-27® Supplement (contains ~0.1mg/L of retinol, all trans), 3 nM TTNPB, 0.25 mM KAAD-Cyclopamine, 50ng/mL Noggin. Days 8 – 12: DMEM/B-27, 50ng/mL KGF, 50ng/mL EGF.

LSD1 Inhibition during Pancreatic Differentiation

Early inhibition of LSD1 (LSD1i^{early}) was performed by addition of the irreversible LSD1 inhibitor tranylcypromine (TCP) to cell culture wells on days 7, 8 and 9 at a final concentration of 0.5 μM. Late inhibition of LSD1 (LSD1i^{late}) was performed by addition of TCP to cell culture wells on days 10, 11 and 12 at a final concentration of 0.5 μM. Additional LSD1 inhibitors used to perform LSD1i^{early} and LSD1i^{late} experiments were SP2509 (Selleck Chemicals), GSK-LSD1 (Sigma) and GSK2879552 (Chemietek); and experiments were performed in the same way as with TCP, at a final concentrations of 1 μM.

Alteration of RA Treatment during Pancreatic Differentiation

Extended RA treatment (RA^{extended}) was performed by addition of the RA analogue TTNPB to cell culture wells on days 8 and 9 at a final concentration of 3 nM. Late RA treatment (RA^{late}) was performed by addition of TTNPB to cell culture wells on days 10 and 11 at a final concentration of 3 nM.

Chromatin Immunoprecipitation Sequencing (ChIP-seq)

ChIP-seq experiments for histone modifications were performed on day 10 with no TCP treatment or after TCP treatment (treatment on days 7, 8, and 9). ChIP-seq experiments for LSD1 were conducted on day 5 (GT stage) day 7 (PP1 stage) and day 10 (PP2 stage) without addition of TCP. ChIP-seq was performed using the ChIP-IT High-Sensitivity kit (Active Motif) according to the manufacturer's instructions. Briefly, for each cell stage and condition analyzed, 5-10 x 10^6 cells were harvested and fixed for 15 min in an 11.1% formaldehyde solution. Cells were lysed

and homogenised using a Dounce homogeniser and the lysate was sonicated in a Bioruptor® Plus (Diagenode), on high for 3 x 5 min (30 sec on, 30 sec off). Between 10 and 30 µg of the resulting sheared chromatin was used for each immunoprecipitation. Equal quantities of sheared chromatin from each sample were used for immunoprecipitations carried out at the same time. 4 µg of antibody were used for each ChIP-seq assay. Antibodies used were: rabbit anti-H3K27ac (Active Motif); rabbit anti-H3K4me1 (Abcam); rabbit anti-H3K4me2 (Millipore); rabbit anti-LSD1 (Abcam); goat anti-FOXA1 (Abcam); goat anti-FOXA2 (Santa Cruz); goat anti-GATA4 (Santa Cruz); mouse anti-GATA6 (Santa Cruz); rabbit anti-HNF6 (Santa Cruz); and rabbit anti-RXRA (Santa Cruz). Chromatin was incubated with primary antibodies overnight at 4 °C on a rotator followed by incubation with Protein G agarose beads for 3 hours at 4 °C on a rotator. Reversal of crosslinks and DNA purification were performed according to the ChIP-IT High-Sensitivity instructions, with the modification of incubation at 65 °C for 2-3 hours, rather than at 80 °C for 2 hours. Sequencing libraries were constructed using KAPA DNA Library Preparation Kits for Illumina® (Kapa Biosystems) and library sequencing was performed on a HiSeq 4000 System (Illumina®) with single-end reads of 50 base pairs (bp). Both library construction and sequencing were performed by the Institute for Genomic Medicine (IGM) core research facility at the University of California at San Diego (UCSD). Two replicates from independent hESC differentiations were generated for each ChIP-seq experiment, except for RXR where only one data set was generated.

ChIP-seq Data Analysis

Single-end 50-bp ChIP-seq reads were mapped to the human genome consensus build (hg19/GRCh37) and visualised using the UCSC Genome Browser[77]. Bowtie 2, version 2.2.7[142] was used to map data to the genome and unmapped reads were discarded. SAMtools[79] was used to remove duplicate sequences and HOMER[80] was used to call peaks using default parameters and to generate tag density plots. Stage- and condition-matched input DNA controls were used as background when calling peaks. The BEDtools[81] suite of programs was used to analyze

whether certain peaks overlapped with other peaks or modified histone regions. Differential peak analysis using HOMER, with default parameters, identified enhancer dynamics between the PP1 to PP2 stage and classify LSD1-bound enhancers into the different enhancer groups (G1, G2 and G3). Each ChIP-seq analysis was performed with two biological replicates, except H3K4me2 at GT, for which pseudo-replicates were generated. The first replicate was analysed and correlated with the appropriate second replicate. Pearson correlations between replicates are listed in **Supplementary Figure 2.10**.

RNA Isolation and Sequencing (RNA-seq) and qRT-PCR

RNA was isolated from cell samples using the RNeasy® Micro Kit (Qiagen) according to the manufacturer instructions. For each cell stage and condition analyzed between 0.1 and 1 x 10^6 cells were collected for RNA extraction. For qRT-PCR, cDNA synthesis was first performed using the iScript™ cDNA Synthesis Kit (Bio-Rad) and 500 ng of isolated RNA per reaction. qRT-PCR reactions were performed in triplicate with 10 ng of template cDNA per reaction using a CFX96™ Real-Time PCR Detection System and the iQ™ SYBR® Green Supermix (Bio-Rad). PCR of the TATA binding protein (TBP) coding sequence was used as an internal control and relative expression was quantified via double delta C_T analysis. For RNA-seq, stranded, single-end sequencing libraries were constructed from isolated RNA using the TruSeq® Stranded mRNA Library Prep Kit (Illumina®) and library sequencing was performed on a HiSeq 4000 System (Illumina®) with single-end reads of 50-bp. Both library construction and sequencing were performed by the IGM core research facility at UCSD. A complete list of RT-qPCR primer sequences can be found in **Supplementary Figure 2.11.**

RNA-seq Data Analysis

Single-end 50-bp reads were mapped to the human genome consensus build (hg19/GRCh37) using the Spliced Transcripts Alignment to a Reference (STAR) aligner[83]. Tag

directories were constructed from STAR outputs and normalised gene expression (fragments per kilobase per million mapped reads; FPKM) for each sequence file was determined using HOMER[80]. HOMER was used to annotate all RefSeq genes with FPKM values and to invoke the R package DESeq2[85] for differential expression analysis. Each RNA-seq analysis was performed on at least two biological replicates with DESeq2, using the built-in option to account for batch effects.

Assignment of Enhancer Target Genes and Motif Enrichment Analysis

Enhancer target genes were assigned using BEDtools to identify transcription start sites (TSSs) located ± 100 kb from LSD1-bound enhancers (groups G1, G2 and G3). HOMER[80] was used to identify transcription factor (TF) binding motifs enriched in the G1 enhancer group versus the G2 and G3 groups. G2 and G3 enhancer peak files were merged and set as the background. G1 enhancers associated with one or more genes with FPKM ≥ 1 at the PP1 stage were used for motif analysis. Analogous motif enrichment analysis was conducted at G2 and G3 enhancers versus the entire genome, excluding G1, G2, and G3 regions.

Flavin Adenine Dinucleotide (FAD) Measurements

FAD on hESC-derived cells was isolated with the FAD Assay Kit (Abcam, ab204710). Cell lysates were deproteinated using perchloric acid and FAD was quantified using the colorimetric assay according to the manufacturer's instructions. The isolation procedure was performed in triplicate with one to two million cells on differentiation days 0, 2, 5, 10, and 13. FAD measurements were performed simultaneously and normalized to protein concentration determined by the Micro BCA Protein Assay Kit (Thermo Scientific, 23235).

Immunofluorescence Analysis

Cell aggregates derived from hESCs were allowed to settle in microcentrifuge tubes and washed twice with PBS before fixation with 4% paraformaldehyde (PFA) for 30 min at room temperature. Dissected e10.5 - e13.5 mouse embryos, e15.5 - postnatal day (P) 0 pancreata, and pancreata from 3-month-old mice were fixed in 4% PFA in PBS for 30 minutes, 3 hours, and overnight, respectively. Fixed samples were washed twice with PBS and incubated overnight at 4 °C in 30% (w/v) sucrose in PBS. Samples were then loaded into disposable embedding molds (VWR), covered in Tissue-Tek® O.C.T. Sakura® Finetek compound (VWR) and flash frozen on dry ice to prepare frozen blocks. The blocks were sectioned at 10 µm and sections were placed on Superfrost Plus® (Thermo Fisher) microscope slides and washed with PBS for 10 min. Slide-mounted cell sections were permeabilised and blocked with blocking buffer, consisting of 0.15% (v/v) Triton X-100 (Sigma) and 1% (v/v) normal donkey serum (Jackson Immuno Research Laboratories) in PBS, for 1 hour at room temperature. Slides were then incubated overnight at 4 °C with primary antibody solutions. The following day slides were washed five times with PBS and incubated for 1 hour at room temperature with secondary antibody solutions. Cells were washed five times with PBS before coverslips were applied.

All antibodies were diluted in blocking buffer at the ratios indicated below. Primary antibodies used were: goat anti-carboxypeptidase A1 (Cpa1) (1:1000 dilution, R&D systems); goat anti-ghrelin (Ghrl) (1:300 dilution, Santa Cruz); goat anti-glucagon (Gcg) (1:1000 dilution, Santa Cruz); goat anti-osteopontin (Opn) (1:300 dilution, R&D systems); guinea pig anti-Ngn3 (1:1000, [143]); goat anti-PDX1 (1:500 dilution, Abcam); guinea pig anti-insulin (INS) (1:500 dilution, Dako); mouse anti-glucagon (GCG) (1:500 dilution, Sigma); rabbit anti-somatostatin (SST) (1:500 dilution, Dako); mouse anti-NKX6.1 (1:300 dilution, Developmental Studies Hybridoma Bank); rabbit anti-amylase (Amy) (1:500 dilution, Sigma); rabbit anti-chromogranin A (ChgA) (1:1000 dilution, Dako); rabbit anti-LSD1 (1:500 dilution, Abcam); rabbit anti-Phospo-Histone3 (Ser10) (pHH3) (1:1000 dilution, Cell Signaling); rabbit anti-polypeptide Y (Ppy) (1:1000 dilution, Dako); rabbit anti-Ptf1a (1:500 dilution, BCBC); rabbit anti-SOX9 (1:1000 dilution, Millipore); rat anti-E-

cadherin (Cdh1) (1:300 dilution, Sigma); rat anti-EpCAM (1:100, DSHB # G8.8); sheep anti-NGN3 (1:300, R&D Systems). Secondary antibodies against sheep, rabbit, goat, mouse, rat, and guinea pig were Alexa488-, Cy3- and Cy5-conjugated donkey antibodies and were used at dilutions of 1:1000, 1:2000, and 1:250, respectively (Jackson Immuno Research Laboratories). Cell nuclei were stained with Hoechst 33342 (1:3000, Invitrogen). TUNEL staining was performed using the ApopTag® Plus Peroxidase In Situ Apoptosis Kit (Millipore). Representative images were obtained with a Zeiss Axio-Observer-Z1 microscope equipped with a Zeiss ApoTome and AxioCam digital camera. Figures were prepared in Adobe Creative Suite 5.

Flow Cytometry Analysis

Cell aggregates derived from hESCs were allowed to settle in microcentrifuge tubes and washed with PBS. Cell aggregates were incubated with Accutase® at room temperature until a single-cell suspension was obtained. Cells were washed with 1 mL ice-cold flow buffer comprised of 0.2% BSA in PBS and centrifuged at 200 g for 5 min. BD Cytofix/Cytoperm™ Plus Fixation/Permeabilization Solution Kit was used to fix and stain cells for flow cytometry according to the manufacturer's instructions. Briefly, cell pellets were re-suspended in ice-cold BD Fixation/Permeabilization solution (300 µL per microcentrifuge tube). Cells were incubated for 20 min at 4 °C. Cells were washed twice with 1 mL ice-cold 1X BD Perm/Wash™ Buffer and centrifuged at 10 °C and 200 x g for 5 min. Cells were re-suspended in 50 µL ice-cold 1X BD Perm/Wash™ Buffer containing diluted antibodies, for each staining performed. Cells were incubated at 4 °C in the dark for 1 hr. Cells were washed with 1.25 mL ice-cold 1X BD Wash Buffer and centrifuged at 200 g for 5 min. Cell pellets were re-suspended in 300 µL ice-cold flow buffer and analysed in a FACSCanto™ (BD Biosciences). Antibodies used for flow cytometry: PE-conjugated anti-PDX1 (1:20 dilution, BD Biosciences); AlexaFluor® 647-conjugated anti-NKX6.1 (1:20 dilution, BD Biosciences); PE-conjugated anti-INS (1:50 dilution, BD Biosciences). Flow cytometry data was processed using FlowJo v10 software.

Generation of LSD1 shRNA Lentiviruses

To generate shRNA expression vectors, shRNA guide sequences were placed under the control of the human U6 pol III promoter in the pLKO.1 backbone (Addgene, plasmid #10878). Small hairpin sequences are listed in **Supplementary Figure 2.12**.

High-titer lentiviral supernatants were generated by co-transfection of the shRNA expression vector and the lentiviral packaging construct into HEK293T cells as described[67]. Briefly, shRNA expression vectors were co-transfected with the pCMV-R8.74 (Addgene, #22036) and pMD2.G (Addgene, #12259) expression plasmids into HEK293T cells using a 1mg/ml PEI solution (Polysciences). Lentiviral supernatants were collected at 48 hr and 72 hr after transfection. Lentiviruses were concentrated by ultracentrifugation for 120 min at 19,500 rpm using a Beckman SW28 ultracentrifuge rotor at 4°C.

Transduction of hESC Endodermal Lineage Intermediates

Differentiation toward the pancreatic progenitor cell stage was initiated as described. At day 6 of differentiation, cells were dissociated with Accutase, washed in PBS + 0.02% BSA and counted. Cells were then distributed onto a 6 well plate at a density of 5 million cells per well. Concentrated lentivirus was then added at 1 µL/mL media, as well as 8 µg/mL polybrene, 10 µM Rock Inhibitor, and 5 units/mL heparin. Cells were then re-aggregated at 100 rpm. After 6 hours, viral media was replaced with fresh day 6 differentiation media. Differentiation was then continued as described, and cells were collected for analysis at day 13.

RNAscope

Mouse embryos at e12.5 were fixed in 4% PFA at 4 °C overnight, embedded in OCT (Sakura Finetek), frozen, and sectioned at 10 µm. Serial sections were prepared as described in the “Immunofluorescence Analysis” section. The expression of mouse *HoxA1* transcripts was

detected using the RNAscope Probe-Mm-Hoxa1 #542391, Multiplex Fluorescent Reagent Kit v2 (#323100, Advanced Cell Diagnostics) and Opal 570 Reagent Pack (Akoya Biosciences #FP1488001KT) according to the manufacturers' recommendations, with the following specifics: target retrieval was omitted, slides received a 30 min treatment with Protease IV and Opal 570 fluorophore and DAPI was diluted at 1:1500. Images were captured with a Zeiss Apotome microscope.

Gene Ontology

Gene ontology analysis was performed using Metascape (http://metascape.ncibi.org) with the default parameters.

Principal Component Analysis

For all samples, FPMKs for total transcriptome were calculated as described above. Genes were then filtered for FPKMs greater than or equal to one, and genes showing the top 25% Median Absolute Deviation (MAD) were selected. Based on these values, PCA plots were generated using the PRComp package in R.

Morphometric Analysis and Cell Counting

At e12.5, every pancreas section was analysed from a minimum of three embryos per genotype, while every fifth pancreas section was analysed at e15.5. For determination of total pancreatic epithelial area, E-cadherin$^+$ area per section was measured using ZEN Digital Imaging for Light Microscopy software (ZEISS), which was calibrated to calculate values in μm^2. The number of marker$^+$ cells was determined by counting every marker$^+$/DAPI$^+$ cell in each section. The number of marker$^+$ cells per section was subsequently divided by the total epithelial area of the section and expressed as marker$^+$ cells/μm^2. All values are shown as mean ± standard error

of the mean (SEM); p-values calculated using unpaired Student's t-test; $p < 0.05$ was considered significant.

Experimental Comparisons

All experiments were independently repeated at least twice. Results are shown as mean ± SEM. Statistical analyses were conducted using GraphPad Prism 6 and R.

Differential Gene Expression Analysis

The DESeq2 Bioconductor package for R was used to calculate gene expression changes. Adjusted p-values < 0.05 and log2(fold-change) ≥ 1.5 were considered significantly different.

Permutation-Based Significance

A random sampling approach (10,000 iterations) was used to obtain null distributions for enrichment analyses, in order to obtain p-values. Null distributions for enrichment of enhancers for gene sets were obtained by randomly shuffling enhancer regions using BEDtools and overlapping with nearby genes. P-values < 0.05 were considered significant.

Data Availability

The accession number for the ChIP-seq and RNA-seq data reported in this manuscript is GSE104840. The accession number for previously reported H3K4me1 and H3K27ac ChIP-seq data is GSE54471. The accession number for previously reported RNA-seq data is E-MTAB-1086.

Author Contributions

N.K.V., N.A.P., A.W., and M.S. conceived the project and designed the experiments. N.K.V., N.A.P., R.J.G., I.M., A.R.H., H.P.S., N.W., and Jinzhao W. performed experiments. N.K.V., R.J.G. and C.W.B. analysed sequencing data. Jianxun W. and M.G.R. provided mice. H.P.S., M.W. and U.S.J. provided technical expertise and critical suggestions for experimental design. N.K.V., N.A.P., R.J.G., A.W., H.P.S., and M.S. interpreted data. N.K.V., N.A.P., R.J.G., and M.S. wrote the manuscript.

Competing Interests

The authors declare no competing interests.

Acknowledgements

We thank FenFen Liu, Nancy Rosenblatt, Karen V. Navarro, and Thomas Harper for experimental assistance, Yinghui Sui for help with data analysis, and the Sander laboratory for helpful discussions and comments on the manuscript. We are grateful to Tung Nguyen, Rizi Ai and Wei Wang for help with data analysis at early stages of this project and Sven Heinz for advice with ChIP-seq experiments. We acknowledge support of the UCSD Human Embryonic Stem Cell Core for cell sorting and the UCSD IGM Genomic Center (supported by P30 DK064391) for library preparation and sequencing. This work was supported by a California Institute for Regenerative Medicine (CIRM) training grant (N.A.P.), Juvenile Diabetes Research Foundation postdoctoral fellowships (H.P.S., A.W.), National Institutes of Health (NIH) grant T32 GM008666 (R.G), and by NIH grants DK089567 and DK068471, and CIRM grant RB5-07236 (M.S.).

Chapter 2, in full, is a reprint of the material as it appears in Nature Communications 2020. This material was co-authored with Vinckier, N. K., Patel, N. A., Geusz, R. J., Wang, A., Wang, J., Matta, I., Harrington, A., Wortham, M., Wetton, N., Wang, J., Jhala, U., Rosenfeld, M., Benner, C. W., Shih, H. P., and Sander, M. The dissertation author was a co-author of this chapter.

CHAPTER 3: PANCREATIC PROGENITOR EPIGENOME MAPS PRIORITIZE TYPE 2 DIABETES RISK GENES WITH ROLES IN DEVELOPMENT

Abstract

Genetic variants associated with type 2 diabetes (T2D) risk affect gene regulation in metabolically relevant tissues, such as pancreatic islets. Here, we investigated contributions of regulatory programs active during pancreatic development to T2D risk. Generation of chromatin maps from developmental precursors throughout pancreatic differentiation of human embryonic stem cells (hESCs) identifies enrichment of T2D variants in pancreatic progenitor-specific stretch enhancers that are not active in islets. Genes associated with progenitor-specific stretch enhancers are predicted to regulate developmental processes, most notably tissue morphogenesis. Through gene editing in hESCs, we demonstrate that progenitor-specific enhancers harboring T2D-associated variants regulate cell polarity genes LAMA1 and CRB2. Knockdown of lama1 or crb2 in zebrafish embryos causes a defect in pancreas morphogenesis and impairs islet cell development. Together, our findings reveal that a subset of T2D risk variants specifically affects pancreatic developmental programs, suggesting that dysregulation of developmental processes can predispose to T2D.

Introduction

Type 2 diabetes (T2D) is a multifactorial metabolic disorder characterized by insulin insensitivity and insufficient insulin secretion by pancreatic beta cells[144]. Genetic association studies have identified hundreds of loci influencing risk of T2D[20]. However, disease-relevant target genes of T2D risk variants, the mechanisms by which these genes cause disease, and the tissues in which the genes mediate their effects remain poorly understood.

The majority of T2D risk variants map to non-coding sequence, suggesting that genetic risk of T2D is largely mediated through variants affecting transcriptional regulatory activity.

Intersection of T2D risk variants with epigenomic data has uncovered enrichment of T2D risk variants in regulatory sites active in specific cell types, predominantly in pancreatic beta cells, including risk variants that affect regulatory activity directly[20,145-153]. T2D risk-associated variants are further enriched within large, contiguous regions of islet active chromatin, referred to as stretch or super-enhancers[146]. These regions of active chromatin preferentially bind islet cell-restricted transcription factors and drive islet-specific gene expression[145,146].

Many genes associated with T2D risk in islets are not uniquely expressed in differentiated islet endocrine cells, but also in pancreatic progenitor cells during embryonic development. For example, T2D risk variants map to *HNF1A, HNF1B, HNF4A, MNX1, NEUROG3, PAX4,* and *PDX1*[20,154,155], which are all transcription factors also expressed in pancreatic developmental precursors. Studies in model organisms and hESC-based models of pancreatic endocrine cell differentiation have shown that inactivation of these transcription factors causes defects in endocrine cell development, resulting in reduced beta cell numbers[156]. Furthermore, heterozygous mutations for *HNF1A, HNF1B, HNF4A, PAX4,* and *PDX1* are associated with maturity onset diabetes of the young (MODY), which is an autosomal dominant form of diabetes with features similar to T2D[157]. Thus, there is evidence that reduced activity of developmentally expressed transcription factors can cause diabetes later in life.

The role of these transcription factors in T2D and MODY could be explained by their functions in regulating gene expression in mature islet cells. However, it is also possible that their function during endocrine cell development could predispose to diabetes instead of, or in addition to, endocrine cell gene regulation. One conceivable mechanism is that individuals with reduced activity of these transcription factors are born with either fewer beta cells or beta cells more prone to fail under conditions of increased insulin demand. Observations showing that disturbed intrauterine metabolic conditions, such as maternal malnutrition, can lead to reduced beta cell mass and T2D predisposition in the offspring[158-160] support the concept that compromised beta cell development could predispose to T2D. However, whether there is T2D genetic risk relevant

to the regulation of endocrine cell development independent of gene regulation in mature islet cells has not been explored.

In this study, we investigated the contribution of gene regulatory programs specifically active during pancreatic development to T2D risk. First, we employed a hESC-based differentiation system to generate chromatin maps of hESCs during their stepwise differentiation into pancreatic progenitor cells. We then identified T2D-associated variants localized in active enhancers in developmental precursors but not in mature islets, used genome editing in hESCs to define target genes of pancreatic progenitor-specific enhancers harboring T2D variants, and employed zebrafish genetic models to study the role of two target genes in pancreatic and endocrine cell development.

Results

Pancreatic progenitor stretch enhancers are enriched for T2D risk variants

To determine whether there is a development-specific genetic contribution to T2D risk, we generated genome-wide chromatin maps of hESCs during their stepwise differentiation into pancreatic progenitors through four distinct developmental stages: definitive endoderm (DE), gut tube (GT), early pancreatic progenitors (PP1), and late pancreatic progenitors (PP2) (**Fig. 3.1a**). We then used ChromHMM[161] to annotate chromatin states, such as active promoters and enhancers, at all stages of hESC differentiation as well as in primary islets (**Supplementary Figure 3.1a,b**).

Large, contiguous regions of active enhancer chromatin, which have been termed stretch- or super-enhancers[47,146], are highly enriched for T2D risk variants in islets[145,146]. We therefore partitioned active enhancers from each hESC developmental stage and islets into stretch enhancers (SE) and traditional (non-stretch) enhancers (TE) (**Fig. 3.1b**). Consistent with prior observations of SE features[47,146], SE comprised a small subset of all active enhancers (7.7%, 7.8%, 8.8%, 8.1%, 8.1%, and 10.4% of active enhancers in ES, DE, GT, PP1, PP2, and islets,

respectively; **Fig. 3.1b** and **Supplementary Figure 3.1c**) and genes proximal to SE were more highly expressed than genes proximal to TE ($p = 4.68 \times 10^{-7}$, 4.64×10^{-11}, 1.31×10^{-5}, 8.85×10^{-9}, 5.34×10^{-6}, and $< 2.2 \times 10^{-16}$ for expression of genes near TE vs SE in ES, DE, GT, PP1, PP2, and islets, respectively; **Supplementary Figure 3.1d**). Genes near SE in pancreatic progenitors included transcription factors involved in the regulation of pancreatic cell identity, such as *NKX6.1* and *PDX1* (**Fig. 3.1c**). Since disease-associated variants are preferentially enriched in narrow peaks of accessible chromatin within broader regions of active chromatin[150,151,153], we next used ATAC-seq to generate genome-wide maps of chromatin accessibility across all time points of differentiation. Nearly all identified SE contained at least one ATAC-seq peak (**Fig. 3.1d** and **Supplementary Figure 3.1e,f**). At the PP2 stage, 62.3% of SE harbored one, 32.2% two or three, and 0.7% four or more ATAC-seq peaks (**Supplementary Figure 3.1f**). Similar percentages were observed in earlier developmental precursors and islets.

Having annotated accessible chromatin sites within SE, we next tested for enrichment of T2D-associated variants in SE active in mature islets and in pancreatic developmental stages. We observed strongest enrichment of T2D-associated variants in islet SE (log enrichment = 2.18, 95% CI = 1.80, 2.54) and late pancreatic progenitor SE (log enrichment = 2.17, 95% CI = 1.40, 2.74), which was more pronounced when only considering variants in accessible chromatin sites within these elements (islet log enrichment = 3.20, 95% CI = 2.74, 3.60; PP2 log enrichment = 3.18, 95% CI = 2.35, 3.79; **Fig. 3.1e**). Given that a subset of pancreatic progenitor SE is also active in islets, we next determined whether pancreatic progenitor SE contribute to T2D risk independently of islet SE. Variants in accessible chromatin sites of late pancreatic progenitor SE were enriched for T2D association in a joint model including islet SE (islet log enrichment = 2.94, 95% CI = 2.47, 3.35; PP2 log enrichment = 1.27, 95% CI = 0.24, 2.00; **Fig. 3.1f**). We also observed enrichment of variants in accessible chromatin sites of pancreatic progenitor SE after conditioning on islet SE (log enrichment = 0.60, 95% CI = -0.87, 1.48), as well as when excluding pancreatic progenitor SE active in islets (log enrichment = 1.62, 95% CI = <-20, 3.14). Examples

of known T2D loci with T2D-associated variants in SE active in pancreatic progenitors but not in islets included *LAMA1* and *PROX1*. These results suggest that a subset of T2D variants may affect disease risk by altering regulatory programs specifically active in pancreatic progenitors.

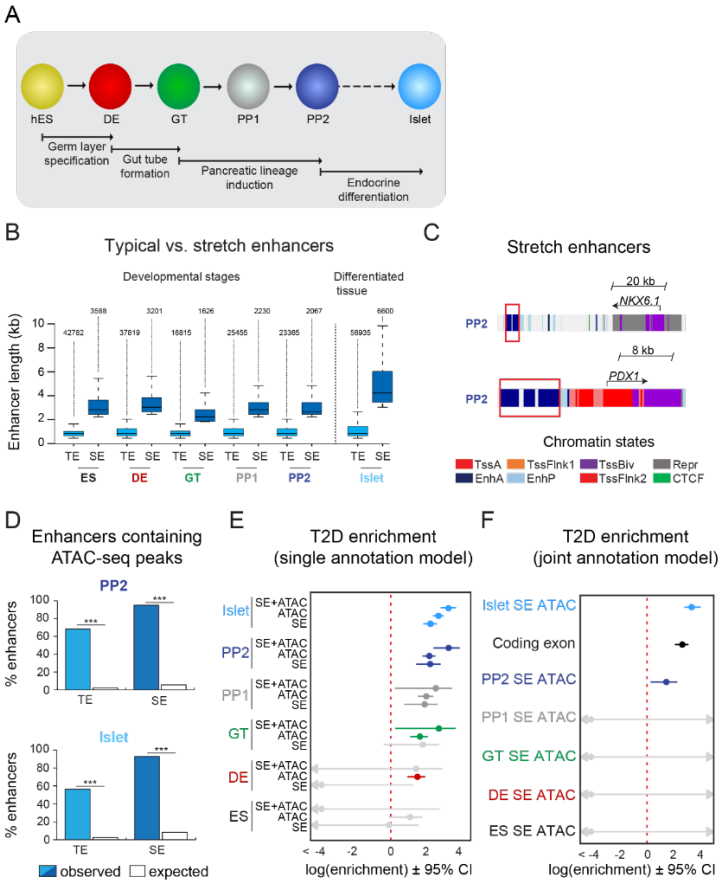

Figure 3.1: T2D-associated risk variants are enriched in stretch enhancers of pancreatic progenitors independent of islet stretch enhancers. (**A**) Schematic illustrating the stepwise differentiation of human embryonic stem cells (hES) into pancreatic progenitors (solid arrows) and lineage relationship to islets (dotted arrow). Developmental intermediates include definitive endoderm (DE), gut tube (GT), early pancreatic progenitor (PP1), and late pancreatic progenitor (PP2) cells. (**B**) Box plots depicting length of typical enhancers (TE) and stretch enhancers (SE) at each developmental stage and in primary human islets. Plots are centered on median, with box encompassing 25th-75th percentile and whiskers extending up to 1.5 interquartile range. Total numbers of enhancers are shown above each box plot. (**C**) Examples of stretch enhancers (denoted with red boxes) near the genes encoding the pancreatic lineage-determining transcription factors NKX6.1 and PDX1, respectively. Chromatin states are based on ChromHMM classifications: TssA, active promoter; TssFlnk, flanking transcription start site; TssBiv, bivalent promoter; Repr, repressed; EnhA, active enhancer; EnhP, poised enhancer. (**D**) Percentage of TE and SE overlapping with at least one ATAC-seq peak at PP2 or in islets. Enrichment analysis comparing observed and expected overlap based on random genomic regions of the same size and located on the same chromosome averaged over 10,000 iterations (*** $p < 1 \times 10^{-4}$; permutation test). ATAC-seq peaks were merged from 2 independent differentiations for PP2 stage cells and 4 donors for primary islets. (**E**) Genome-wide enrichment of T2D-associated variants (minor allele frequency > 0.0025) in stretch enhancers, ATAC-seq peaks, and ATAC-seq peaks within stretch enhancers for all developmental stages when modelling each annotation separately. Points and lines represent log-scaled enrichment estimates and 95% confidence intervals from functional genome wide association analysis (fgwas), respectively. ATAC-seq peaks were merged from 2 independent differentiations for ES, DE, GT, PP1, and PP2 stage cells and from 4 donors for primary islets. (**F**) Genome-wide enrichment of T2D-associated variants (minor allele frequency > 0.0025) in ATAC-seq peaks within stretch enhancers for all developmental stages and coding exons when considering all annotations in a joint model. Points and lines represent log-scaled enrichment estimates and 95% confidence intervals from fgwas, respectively. ATAC-seq peaks were merged from 2 independent differentiations for ES, DE, GT, PP1, and PP2 stage cells and from 4 donors for primary islets.

Pancreatic progenitor-specific stretch enhancers are near genes that regulate tissue morphogenesis

Having observed enrichment of T2D risk variants in pancreatic progenitor SE independent of islet SE, we next sought to further characterize the regulatory programs of SE with specific function in pancreatic progenitors. We therefore defined a set of pancreatic progenitor-specific stretch enhancers (PSSE) based on the following criteria: (i) annotation as a SE at the PP2 stage, (ii) no classification as a SE at the ES, DE, and GT stages, and (iii) no classification as a TE or SE in islets. Applying these criteria, we identified a total of 492 PSSE genome-wide (**Fig. 3.2a**).

As expected based on their chromatin state classification, PSSE acquired broad deposition of the active enhancer mark H3K27ac at the PP1 and PP2 stages (**Fig. 3.2b,c**). Coincident with an increase in H3K27ac signal, chromatin accessibility at PSSE also increased (**Fig. 3.2b**), and 93.5% of PSSE contained at least one accessible chromatin site at the PP2 stage (**Supplementary Figure 3.2a,b**). Further investigation of PSSE chromatin state dynamics at earlier stages of pancreatic differentiation revealed that PSSE were often poised (defined by H3K4me1 in the absence of H3K27ac) prior to activation (42%, 48%, 63%, and 17% of PSSE in ES, DE, GT, and PP1, respectively; **Fig. 3.2c**), consistent with earlier observations that a poised enhancer state frequently precedes enhancer activation during development[6,27]. Intriguingly, a subset of PSSE was classified as TE earlier in development (13%, 23%, 29%, and 46% of PSSE in ES, DE, GT, and PP1, respectively; **Fig. 3.2c**), suggesting that SE emerge from smaller regions of active chromatin seeded at prior stages of development. During differentiation into mature islet cells, PSSE lost H3K27ac but largely retained H3K4me1 signal (62% of PSSE) (**Fig. 3.2c**), persisting in a poised state in terminally differentiated islet cells.

To gain insight into the transcription factors that regulate PSSE, we conducted motif enrichment analysis of accessible chromatin sites within PSSE (**Supplementary Figure 3.2c**). Consistent with the activation of PSSE upon pancreas induction, motifs associated with transcription factors known to regulate pancreatic development[162,163] were enriched, including

FOXA ($p = 1 \times 10^{-34}$), PDX1 ($p = 1 \times 10^{-30}$), GATA ($p = 1 \times 10^{-25}$), ONECUT ($p = 1 \times 10^{-17}$), and RBPJ ($p = 1 \times 10^{-14}$), suggesting that pancreatic lineage-determining transcription factors activate PSSE. Analysis of the extent of PSSE overlap with ChIP-seq binding sites for FOXA1, FOXA2, GATA4, GATA6, PDX1, HNF6, and SOX9 at the PP2 stage substantiated this prediction ($p < 1 \times 10^{-4}$ for all transcription factors; permutation test; **Fig. 3.2d**).

Annotation of biological functions of predicted target genes for PSSE (nearest gene with FPKM ≥ 1 at PP2 stage) revealed gene ontology terms related to developmental processes, such as tissue morphogenesis ($p = 1 \times 10^{-7}$) and vascular development ($p = 1 \times 10^{-8}$), as well as developmental signaling pathways, including BMP ($p = 1 \times 10^{-5}$), NOTCH ($p = 1 \times 10^{-4}$), and canonical Wnt signaling ($p = 1 \times 10^{-4}$; **Figure 3.2e**), which have demonstrated roles in pancreas morphogenesis and cell lineage allocation[164-168]. Consistent with the temporal pattern of H3K27ac deposition at PSSE, transcript levels of PSSE-associated genes increased upon pancreatic lineage induction and peaked at the PP2 stage ($p = 1.8 \times 10^{-8}$; **Supplementary Figure 3.2d**). Notably, expression of these genes sharply decreased in islets ($p < 2.2 \times 10^{-16}$), underscoring the likely role of these genes in regulating pancreatic development but not mature islet function.

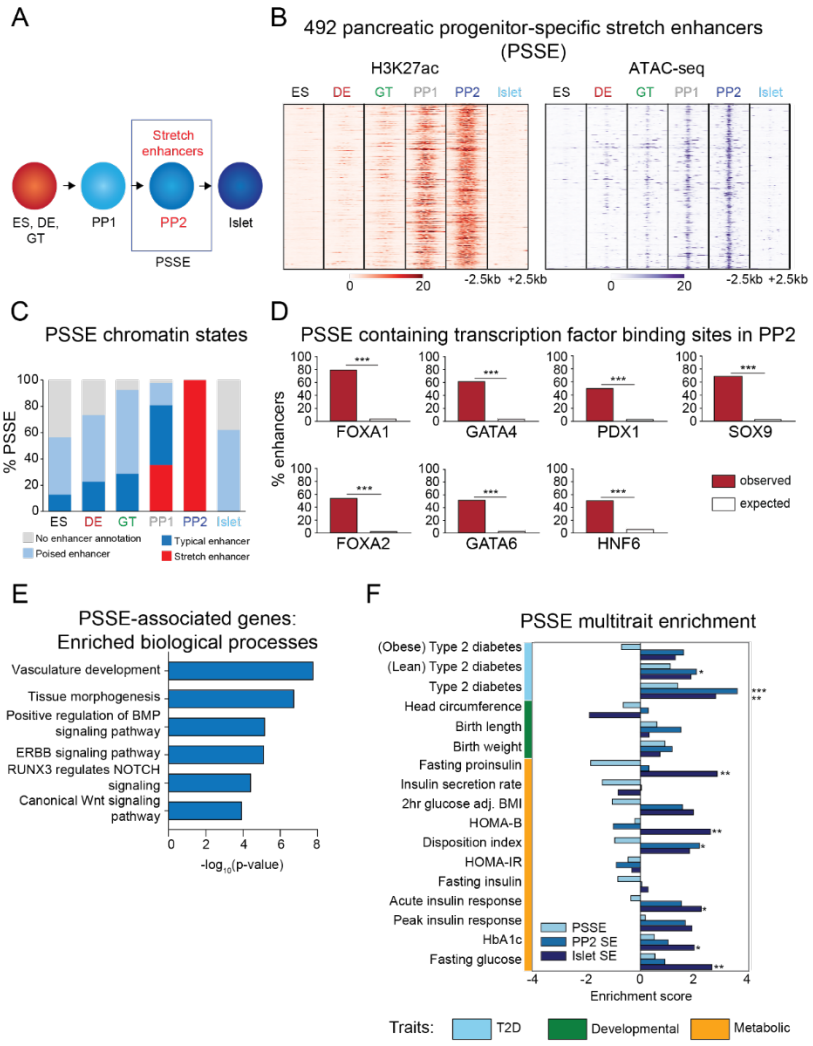

Figure 3.2: Candidate target genes of pancreatic progenitor-specific stretch enhancers regulate developmental processes. (**A**) Schematic illustrating identification of pancreatic progenitor-specific stretch enhancers (PSSE). (**B**) Heatmap showing density of H3K27ac ChIP-seq and ATAC-seq reads at PSSE, centered on overlapping H3K27ac and ATAC-seq peaks, respectively, and spanning 5 kb in ES, DE, GT, PP1, PP2, and islets. PSSE coordinates in Figure 2 – source data 1. (**C**) Percentage of PSSE exhibiting indicated chromatin states at defined developmental stages and in islets. (**D**) Percentage of PSSE overlapping with at least one ChIP-seq peak at PP2 for the indicated transcription factors. Enrichment analysis comparing observed and expected overlap based on random genomic regions of the same size and located on the same chromosome averaged over 10,000 iterations (*** $p < 1 \times 10^{-4}$; permutation test). (**E**) Gene ontology analysis for nearest expressed genes (fragments per kilobase per million fragments mapped (FPKM) ≥ 1 at PP2) to the 492 PSSE. See also Figure 2 – source data 2. (**F**) Enrichment (LD score regression coefficient z-scores) of T2D, developmental, and metabolic GWAS trait-associated variants at accessible chromatin sites in PSSE as compared with PP2 and islet stretch enhancers. Significant enrichment was identified within accessible chromatin at PP2 stretch enhancers for lean type 2 diabetes ($Z = 2.06$, $*p = 3.94 \times 10^{-2}$), at PP2 stretch enhancers for type 2 diabetes ($Z = 3.57$, $***p = 3.52 \times 10^{-4}$), at islet stretch enhancers for type 2 diabetes ($Z = 2.78$, $**p = 5.46 \times 10^{-3}$), at islet stretch enhancers for fasting proinsulin levels ($Z = 2.83$, $**p = 4.61 \times 10^{-3}$), at islet stretch enhancers for HOMA-B ($Z = 2.58$, $**p = 9.85 \times 10^{-3}$), at PP2 stretch enhancers for disposition index ($Z = 2.18$, $*p = 2.94 \times 10^{-2}$), at islet stretch enhancers for acute insulin response ($Z = 2.24$, $*p = 2.51 \times 10^{-2}$), at islet stretch enhancers for HbA1c ($Z = 1.98$, $*p = 4.72 \times 10^{-2}$), and at islet stretch enhancers for fasting glucose levels ($Z = 2.64$, $**p = 8.31 \times 10^{-3}$).

Pancreatic progenitor-specific stretch enhancers are highly specific across T2D-relevant tissues and cell types

We next sought to understand the phenotypic consequences of PSSE activity in the context of T2D pathophysiology. Variants in accessible chromatin sites of PSSE genome-wide were enriched for T2D association (log enrichment = 2.85, 95% CI = <-20, 4.09). We determined enrichment of genetic variants for T2D-related quantitative endophenotypes within accessible chromatin sites of PSSE, as well as all pancreatic progenitor SE (not just progenitor-specific) and islet SE, using LD score regression[169,170]. As expected based on prior observations[145,146], we observed enrichment ($Z > 1.96$) of variants associated with quantitative traits related to insulin secretion and beta cell function within islet SE, exemplified by fasting proinsulin levels, HOMA-B, and acute insulin response ($Z = 2.8$, $Z = 2.6$, and $Z = 2.2$, respectively; **Fig. 3.2f**). Conversely, PSSE showed a trend toward depletion for these traits although the estimates were not significant. We further tested for enrichment in the proportion of variants in PSSE and islet SE nominally associated ($p < 0.05$) with beta cell function traits compared to background variants. There was significant enrichment of beta cell trait association among islet SE variants (χ^2 test; $p < 0.05$ for all beta cell functional traits except for insulin secretion rate), but no corresponding enrichment for PSSE.

A prior study found that variants at the *LAMA1* locus had stronger effects on T2D risk among lean relative to obese cases[171]. Since we identified a PSSE at the *LAMA1* locus, we postulated that variants in PSSE collectively might have differing impact on T2D risk in cases segregated by BMI. We therefore tested PSSE, as well as pancreatic progenitor SE and islet SE, for enrichment of T2D association using GWAS of lean and obese T2D[171], using LD score regression[169,170]. We observed nominally significant enrichment of variants in pancreatic progenitor SE for T2D among lean cases ($Z = 2.1$). Variants in PSSE were mildly enriched for T2D among lean ($Z = 1.1$) and depleted among obese ($Z = -0.70$) cases, although neither estimate was significant. By comparison, islet SE showed positive enrichment for T2D among both lean (Z

= 1.9) and obese cases (Z = 1.3; **Figure 3.2f**). Together, these results suggest that PSSE may affect T2D risk in a manner distinct from islet SE function.

Having observed little evidence for enrichment of PSSE variants for traits related to beta cell function, we asked whether the enrichment of PSSE for T2D-associated variants could be explained by PSSE activity in T2D-relevant tissues and cell types outside the pancreas. We assessed PSSE activity by measuring H3K27ac signal in 95 representative tissues and cell lines from the ENCODE and Epigenome Roadmap projects[172]. Interestingly, there was group-wide specificity of PSSE to pancreatic progenitors relative to other cells and tissues including those relevant to T2D, such as adipose tissue, skeletal muscle, and liver (**Supplementary Figure 3.2e**). Since gene regulation in adipocyte precursors also contributes to T2D risk[173], we further examined PSSE specificity with respect to chromatin states during adipogenesis, using data from human adipose stromal cell differentiation stages (hASC1-4)[151,174]. PSSE exhibited virtually no active chromatin during adipogenesis (9, 8, 6, and 8 out of the 492 PSSE were active enhancers in hACS-1, hASC-2, hASC-3, and hASC-4, respectively; **Supplementary Figure 3.2f**). These findings identify PSSE as highly pancreatic progenitor-specific across T2D-relevant tissues and cell types.

Identification of pancreatic progenitor-specific stretch enhancers harboring T2D-associated variants

Given the relative specificity of PSSE to pancreatic progenitors, we next sought to identify T2D-associated variants in PSSE at specific loci which may affect pancreatic development. We therefore identified variants in PSSE with evidence of T2D association (at $p = 4.7 \times 10^{-6}$) after correcting for the total number of variants in PSSE genome-wide (n = 10,738). In total there were 49 variants in PSSE with T2D association exceeding this threshold mapping to 11 loci (**Fig. 3.3a**). This included variants at 9 loci with known genome-wide significant T2D association (*PROX1*, *ST6GAL1*, *SMARCAD1*, *XKR6*, *INS-IGF2*, *HMGA2*, *SMEK1*, *HMG20A*, and *LAMA1*), as well as

at two previously unreported loci with sub-genome-wide significant association, *CRB2* and *PGM1*. To identify candidate target genes of the T2D-associated PSSE in pancreatic progenitors, we analyzed the expression of all genes within the same topologically associated domain (TAD) as the PSSE in PP2 cells and in primary human embryonic pancreas tissue (**Fig. 3.3b** and **Supplementary Figure 3.3a**). These expressed genes are candidate effector transcripts of T2D-associated variants in pancreatic progenitors.

As many pancreatic progenitor SE remain poised in mature islets (**Fig. 3.2c**), we considered whether T2D-associated variants in PSSE could have gene regulatory function in islets that is re-activated in the disease state. We therefore assessed overlap of PSSE variants with accessible chromatin of islets from T2D donors[175]. None of the strongly T2D-associated variants in PSSE ($p = 4.7 \times 10^{-6}$) overlapped an islet accessible chromatin site in T2D islets, arguing against the relevance of PSSE in broadly regulating islet gene activity during T2D.

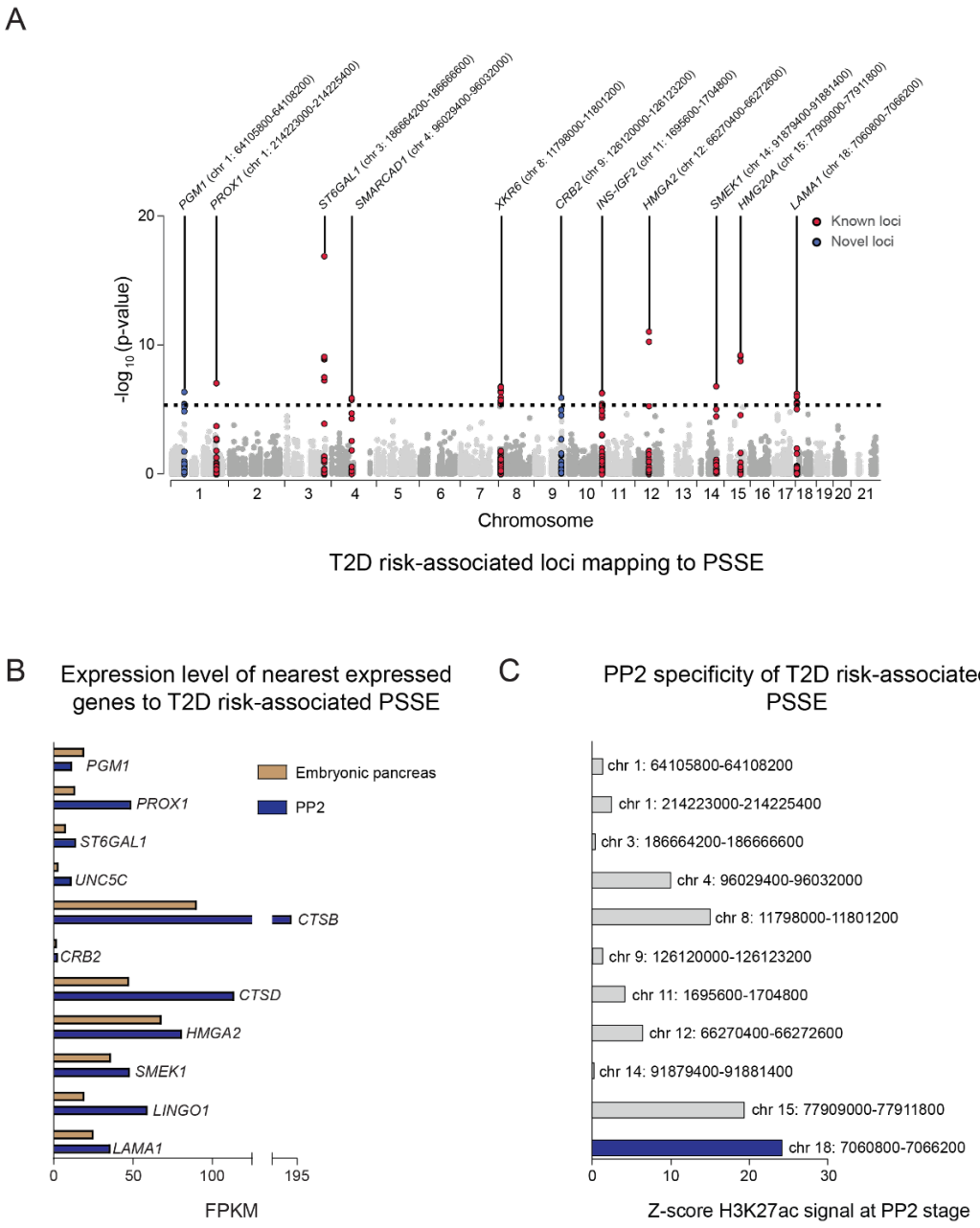

Figure 3.3: Identification of T2D risk variants associated with pancreatic progenitor-specific stretch enhancers. (**A**) Manhattan plot showing T2D association p-values (from Mahajan et al., 2018) for 10,738 variants mapping within PSSE. The dotted line shows the threshold for Bonferroni correction ($p = 4.66 \times 10^{-6}$). Novel loci identified with this threshold and mapping at least 500 kb away from a known locus are highlighted in blue. Chromosomal coordinates of T2D-associated PSSE are indicated. (**B**) mRNA levels (measured in fragments per kilobase per million fragments mapped (FPKM)) at PP2 (blue) and in human embryonic pancreas (54 and 58 days gestation, gold) of nearest expressed (FPMK ≥ 1) gene at PP2 for PSSE harboring T2D variants identified in Fig. 3A. (**C**) PP2 specificity of H3K27ac signal at PSSE harboring T2D variants identified in Fig. 3A. Z-score comparing H3K27ac signal at PP2 to H3K27ac signal in tissues and cell lines from the ENCODE and Epigenome Roadmap projects.

A pancreatic progenitor-specific stretch enhancer at *LAMA1* harbors T2D risk variants and regulates *LAMA1* expression selectively in pancreatic progenitors

Variants in a PSSE at the *LAMA1* locus were associated with T2D at genome-wide significance (**Fig. 3.3a**), and *LAMA1* was highly expressed in the human embryonic pancreas (**Fig. 3.3b**). Furthermore, the activity of the PSSE at the *LAMA1* locus was almost exclusively restricted to pancreatic progenitors (**Supplementary Figure 3.3b,c**), and was further among the most progenitor-specific across all PSSE harboring T2D risk variants (**Fig. 3.3c**). In addition, reporter gene assays in zebrafish embryos have shown that this enhancer drives gene expression specific to pancreatic progenitors *in vivo*[51]. We therefore postulated that the activity of T2D-associated variants within the *LAMA1* PSSE is relevant for gene regulation in pancreatic progenitors, and we sought to characterize the *LAMA1* PSSE in greater depth.

Multiple T2D-associated variants mapped within the *LAMA1* PSSE, and these variants were further in the 99% credible set in fine-mapping data from the DIAMANTE consortium[20] (**Fig. 3.4a**). No other variants in the 99% credible set mapped in an accessible chromatin site active in islets from either non-diabetic or T2D samples. The PSSE is intronic to the *LAMA1* gene and contains regions of poised chromatin and TE at prior developmental stages (**Fig. 3.4a**). Consistent with its stepwise genesis as a SE throughout development, regions of open chromatin within the *LAMA1* PSSE were already present at the DE and GT stages. Furthermore, pancreatic lineage-determining transcription factors, such as FOXA1, FOXA2, GATA4, GATA6, HNF6, SOX9, and PDX1, were all bound to the PSSE at the PP2 stage (**Fig. 3.4b**). Among credible set variants in the *LAMA1* PSSE, rs10502347 overlapped an ATAC-seq peak as well as ChIP-seq sites for multiple pancreatic lineage-determining transcription factors. Additionally, rs10502347 directly coincided with a SOX9 footprint identified in ATAC-seq data from PP2 cells, and the T2D risk allele C is predicted to disrupt SOX9 binding (**Fig. 3.4b**). Consistent with the collective endophenotype association patterns of PSSE (**Fig. 3.2f**), rs10502347 showed no association with beta cell function ($p = 0.81, 0.23, 0.46$ for fasting proinsulin levels, HOMA-B, and acute insulin

response, respectively; **Supplementary Figure 3.4a**). Thus, T2D variant rs10502347 is predicted to affect the binding of pancreatic transcription factors and does not appear to affect beta cell function.

Enhancers can control gene expression over large genomic distances, and therefore their target genes cannot be predicted based on proximity alone. To directly assess the function of the *LAMA1* PSSE in regulating gene activity, we utilized CRIPSR-Cas9-mediated genome editing to generate two independent clonal human hESC lines harboring homozygous deletions of the *LAMA1* PSSE (hereafter referred to as Δ*LAMA1*Enh; **Supplementary Figure 3.4b**). We examined *LAMA1* expression in Δ*LAMA1*Enh compared to control cells throughout stages of pancreatic differentiation. Consistent with the broad expression of *LAMA1* across developmental and mature tissues, control cells expressed *LAMA1* at all stages (**Fig. 3.4c**). *LAMA1* was expressed at similar levels in Δ*LAMA1*Enh and control cells at early developmental stages, but was significantly reduced in PP2 cells derived from Δ*LAMA1*Enh clones ($p = 0.319$, 0.594, 0.945, 0.290, and $< 1 \times 10^{-6}$ for comparisons in ES, DE, GT, PP1, and PP2, respectively; **Fig. 3.4d**). To next investigate whether the *LAMA1* PSSE regulates other genes at this locus, we utilized Hi-C datasets from PP2 cells to identify topologically associated domains (TADs). We then examined expression of genes mapping in the same TAD as the *LAMA1* PSSE. *ARHGAP28* was the only other expressed gene within the TAD, and albeit not significantly different from controls ($p.adj > 0.05$), showed a trend toward lower expression in Δ*LAMA1*Enh PP2 cells (**Fig. 3.4e**), raising the possibility that *ARHGAP28* is an additional target gene of the *LAMA1* PSSE. Together, these results demonstrate that while *LAMA1* itself is broadly expressed across developmental stages, the T2D-associated PSSE regulates *LAMA1* expression specifically in pancreatic progenitors.

To determine whether deletion of the *LAMA1* PSSE affects pancreatic development, we generated PP2 stage cells from Δ*LAMA1*Enh and control hESC lines and analyzed pancreatic cell fate commitment by flow cytometry and immunofluorescence staining for PDX1 and NKX6.1

(**Supplementary Figure 3.4c,d**). At the PP2 stage, Δ*LAMA1*Enh and control cultures contained similar percentages of PDX1- and NKX6.1-positive cells. Furthermore, mRNA expression of *PDX1*, *NKX6.1*, *PROX1*, *PTF1A*, and *SOX9* was either unaffected or only minimally reduced (p adj. = 3.56×10^{-2}, 0.224, 0.829, 8.14×10^{-2}, and 0.142, for comparisons of *PDX1*, *NKX6.1*, *PROX1*, *PTF1A*, and *SOX9* expression, respectively; **Supplementary Figure 3.4e**), and the overall gene expression profiles as determined by RNA-seq were similar in Δ*LAMA1*Enh and control PP2 cells (**Supplementary Figure 3.4f**). To examine effects of complete *LAMA1* loss-of-function, we additionally generated a hESC line harboring a deletion of the *LAMA1* coding sequences (hereafter referred to as Δ*LAMA1;* **Supplementary Figure 3.5a,b)**, and produced PP2 stage cells. Similar to Δ*LAMA1*Enh cultures, Δ*LAMA1* and control PP2 stage cultures contained similar numbers of PDX1- and NKX6.1-positive cells (**Supplementary Figure 3.5c,d**). Likewise, mRNA expression of *PDX1, NKX6.1, PROX1, PTF1A,* and *SOX9* was similar in Δ*LAMA1* and control PP2 cells ($p = 4.3 \times 10^{-2}$, 0.19, 0.16, 0.17, and 8.7×10^{-2}, respectively; **Supplementary Figure 3.5e)**. These findings indicate that *in vitro* pancreatic lineage induction is unperturbed in both Δ*LAMA1*Enh cells exhibiting reduced *LAMA1* expression, as well as Δ*LAMA1* cells where *LAMA1* coding sequences are disrupted.

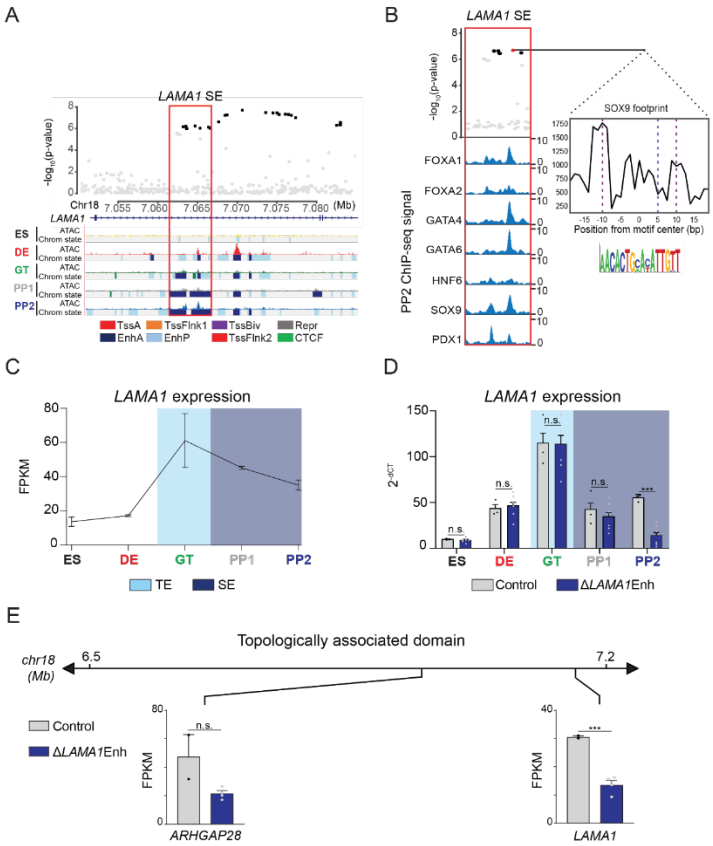

Figure 3.4: A T2D risk-associated *LAMA1* pancreatic progenitor-specific stretch enhancer regulates *LAMA1* expression specifically in pancreatic progenitors. (**A**) (Top) Locus plots showing T2D association p-values for variants in a 35 kb window (hg19 chr18:7,050,000-7,085,000) at the *LAMA1* locus and *LAMA1* PSSE (red box). Fine mapped variants within the 99% credible set for the *LAMA1* locus are colored black. All other variants are colored light gray. (Bottom) Chromatin states and ATAC-seq signal in ES, DE, GT, PP1, and PP2. TssA, active promoter; TssFlnk, flanking transcription start site; TssBiv, bivalent promoter; Repr, repressed; EnhA, active enhancer; EnhP, poised enhancer. (**B**) FOXA1, FOXA2, GATA4, GATA6, HNF6, SOX9, and PDX1 ChIP-seq profiles at the *LAMA1* PSSE in PP2. The variant rs10502347 (red) overlaps transcription factor binding sites and a predicted ATAC-seq footprint for the SOX9 sequence motif. Purple dotted lines indicate the core binding profile of the average SOX9 footprint genome-wide and the blue dotted line indicates the position of rs10502347 within the SOX9 motif. (**C**) *LAMA1* mRNA expression at each developmental stage determined by RNA-seq, measured in fragments per kilobase per million fragments mapped (FPKM). Data shown as mean ± S.E.M. (n = 3 replicates from independent differentiations). Light blue and purple indicate classification of the *LAMA1* PSSE as typical enhancer (TE) and stretch enhancer (SE), respectively. (**D**) *LAMA1* mRNA expression at each developmental stage determined by qPCR in control and Δ*LAMA1*Enh cells. Data are shown as mean ± S.E.M. (n = 3 replicates from independent differentiations for control cells. Δ*LAMA1*Enh cells represent combined data from 2 clonal lines with 3 replicates for each line from independent differentiations. n = 3 technical replicates for each sample; p = 0.319, 0.594, 0.945, 0.290, and < 1×10^{-6} for comparisons in ES, DE, GT, PP1, and PP2, respectively; student's t-test, 2 sided; *** p < 0.001, n.s., not significant). Light blue and purple indicate classification of the *LAMA1* PSSE as TE and SE, respectively. Each plotted point represents the average of technical replicates for each differentiation. (**E**) mRNA expression determined by RNA-seq at PP2 of genes expressed in either control or Δ*LAMA1*Enh cells (FPKM ≥ 1 at PP2) and located within the same topologically associated domain as *LAMA1*. Data are shown as mean FPKM ± S.E.M. (n = 2 replicates from independent differentiations for control cells. Δ*LAMA1*Enh cells represent combined data from 2 clonal lines with 2 replicates for each line from independent differentiations. p adj. = 0.389 and 8.11×10^{-3} for *ARHGAP28* and *LAMA1*, respectively; DESeq2).

Pancreatic progenitor-specific stretch enhancers at the CRB2 and PGM1 loci harbor T2D-associated variants

Multiple variants with evidence for T2D association in PSSE mapped outside of known risk loci, such as those mapping to *CRB2* and *PGM1* (**Fig. 3.3a**). As with the *LAMA1* PSSE, PSSE harboring variants at *CRB2* and *PGM1* were intronic to their respective genes, contained ATAC-seq peaks, and bound pancreatic lineage-determining transcription factors FOXA1, FOXA2, GATA4, GATA6, HNF6, SOX9, and PDX1 (**Fig. 3.5a,b** and **Supplementary Figure 3.6a,b**). Compared to the *LAMA1* PSSE, *CRB2* and *PGM1* PSSE were less specific to pancreatic progenitors and exhibited significant H3K27ac signal in several other tissues and cell types, most notably brain, liver, and the digestive tract (**Supplementary Figure 3.6c,d**).

CRB2 is a component of the Crumbs protein complex involved in the regulation of cell polarity and neuronal, heart, retinal, and kidney development[176-180]. However, its role in pancreatic development is unknown. To determine whether the *CRB2* PSSE regulates *CRB2* expression in pancreatic progenitors, we generated two independent hESC clones with homozygous deletions of the *CRB2* PSSE (hereafter referred to as Δ*CRB2*Enh; **Supplementary Figure 3.7a**) and performed pancreatic differentiation of Δ*CRB2*Enh and control hESC lines. In control cells, *CRB2* was first expressed at the GT stage and increased markedly at the PP1 stage (**Fig. 3.5c**). This pattern of *CRB2* expression is consistent with H3K27ac deposition at the *CRB2* PSSE in GT stage cells and classification as a SE at the PP1 and PP2 stages (**Fig. 3.5a** and **Supplementary Figure 3.6c**). In Δ*CRB2*Enh cells, we observed upregulation of *CRB2* expression at earlier developmental stages, in particular at the DE and GT stages ($p < 1 \times 10^{-6}$ at both stages; **Fig. 3.5d**), suggesting that the *CRB2* PSSE may be associated with repressive transcriptional complexes prior to pancreas induction. At the PP2 stage, *CRB2* expression was significantly reduced in Δ*CRB2*Enh cells (p adj. = 3.51×10^{-3}; **Fig. 3.5d**), whereas the expression of other genes in the same TAD was not affected (p adj. ≥ 0.05; **Fig. 3.5e**). Thus, the *CRB2* PSSE specifically regulates *CRB2* and is required for *CRB2* expression in pancreatic progenitors.

Phenotypic characterization of PP2 stage Δ*CRB2*Enh cultures revealed similar percentages of PDX1- and NKX6.1-positive cells as in control cultures (**Supplementary Figure 3.7b,c**). The expression of pancreatic transcription factors and global gene expression profiles were also similar in Δ*CRB2*Enh and control PP2 cells (**Supplementary Figure 3.7d,e**). Likewise, *CRB2* deletion hESCs (Δ*CRB2*) differentiated to the PP2 stage (**Supplementary Figure 3.8a,b**) produced similar numbers of PDX1- and NKX6.1-positive cells and expressed pancreatic transcription factors at levels similar to control cells (**Supplementary Figure 3.8c-e**). Thus, neither deletion of the *CRB2* PSSE nor the *CRB2* gene overtly impairs pancreatic lineage induction in the *in vitro* hESC differentiation system.

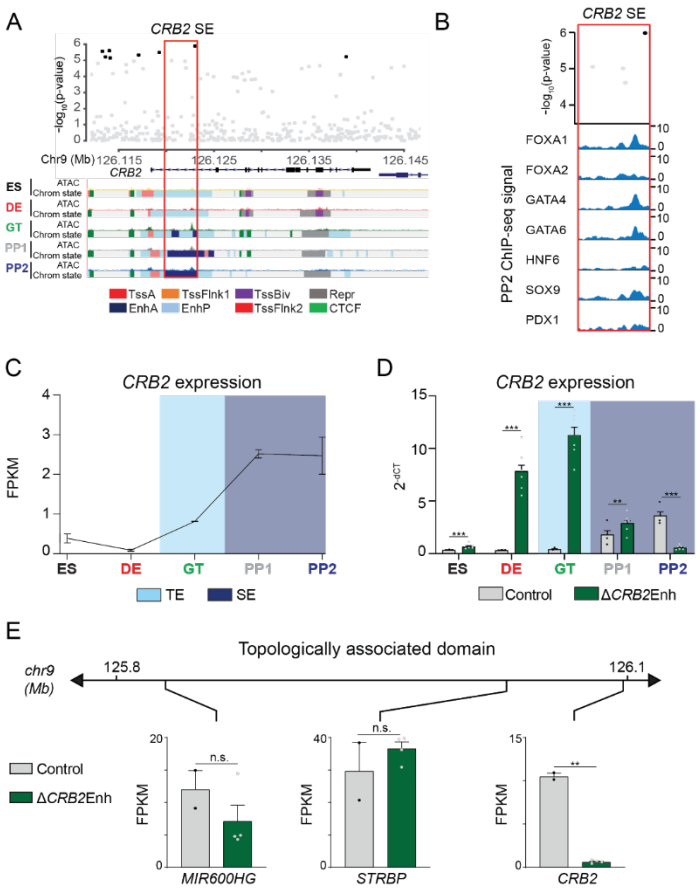

Figure 3.5: A T2D risk-associated *CRB2* pancreatic progenitor-specific stretch enhancer regulates *CRB2* expression specifically in pancreatic progenitors. (**A**) (Top) Locus plots showing T2D association p-values for variants in a 35 kb window (hg19 chr9:126,112,000-126,147,000) at the *CRB2* locus and *CRB2* PSSE (red box). Fine mapped variants within the 99% credible set for the novel *CRB2* locus are colored black. All other variants are colored light gray. (Bottom) Chromatin states and ATAC-seq signal in ES, DE, GT, PP1, and PP2. TssA, active promoter; TssFlnk, flanking transcription start site; TssBiv, bivalent promoter; Repr, repressed; EnhA, active enhancer; EnhP, poised enhancer. (**B**) FOXA1, FOXA2, GATA4, GATA6, HNF6, SOX9, and PDX1 ChIP-seq profiles at the *CRB2* PSSE in PP2. The variant rs2491353 (black) overlaps with transcription factor binding sites. (**C**) *CRB2* mRNA expression at each developmental stage determined by RNA-seq, measured in fragments per kilobase per million fragments mapped (FPKM). Data shown as mean ± S.E.M. (n = 3 replicates from independent differentiations). Light blue and purple indicate classification of the *CRB2* PSSE as typical enhancer (TE) and stretch enhancer (SE), respectively. Plotted points represent average of technical replicates for each differentiation. (**D**) *CRB2* mRNA expression at each developmental stage determined by qPCR in control and Δ*CRB2*Enh cells. Data are shown as mean ± S.E.M. (n = 3 replicates from independent differentiations for control cells. Δ*CRB2*Enh cells represent combined data from 2 clonal lines with 3 replicates for each line from independent differentiations. n = 3 technical replicates for each sample; $p = 7.03 \times 10^{-4}$, $< 1 \times 10^{-6}$, $< 1 \times 10^{-6}$, 1.46×10^{-2}, and $< 1 \times 10^{-6}$ for comparisons in ES, DE, GT, PP1, and PP2, respectively; student's t-test, 2 sided; *** $p < 0.001$ ** $p < 0.01$). Light blue and purple indicate classification of the *CRB2* PSSE as TE and SE, respectively. Each plotted point represents the average of technical replicates for each differentiation. (**E**) mRNA expression determined by RNA-seq at PP2 of genes expressed in either control or Δ*CRB2*Enh cells (FPKM ≥ 1 at PP2) and located within the same topologically associated domain as *CRB2*. Data are shown as mean FPKM ± S.E.M. (n = 2 replicates from independent differentiations for control cells. Δ*CRB2*Enh cells represent combined data from 2 clonal lines with 2 replicates for each line from independent differentiations. p adj. = 0.158, 1.00, and 3.51×10^{-3}, for *MIR600HG*, *STRBP*, and *CRB2*, respectively; DESeq2; **$p < 0.01$, n.s., not significant).

lama1 and crb2 zebrafish morphants display annular pancreas and decreased beta cell mass

Based on their classification as extracellular matrix and cell polarity proteins, respectively, Laminin (encoded by *LAMA1*) and CRB2 are predicted to regulate processes related to tissue morphogenesis, such as cell migration, tissue growth, and cell allocation within the developing organ. Furthermore, PSSE in general were enriched for proximity to genes involved in tissue morphogenesis (**Fig. 3.2e**), suggesting that T2D risk variants acting within PSSE could have roles in pancreas morphogenesis. Since cell migratory processes and niche-specific signaling events are not fully modeled during hESC differentiation, we reasoned that the *in vitro* pancreatic differentiation system might not be suitable for studying Laminin and CRB2 function in pancreatic development.

To circumvent these limitations, we employed zebrafish as an *in vivo* vertebrate model to study the effects of reduced *lama1* and *crb2* levels on pancreatic development. The basic organization and cell types in the pancreas as well as the genes regulating endocrine and exocrine pancreas development are highly conserved between zebrafish and mammals[181-183]. To analyze pancreatic expression of Laminin and Crb proteins, we used *Tg(ptf1a:eGFP)*^{jh1} embryos to visualize pancreatic progenitor cells and the acinar pancreas by eGFP expression. At 48 hours post-fertilization (hpf), both Laminin and Crb proteins were detected in the eGFP and Nkx6.1 co-positive pancreatic progenitor cell domain (**Supplementary Figure 3.9a,b**).

To determine the respective functions of *lama1* and *crb2* in pancreatic development, we performed knockdown experiments using anti-sense morpholinos directed against *lama1* and the two zebrafish *crb2* genes, *crb2a* and *crb2b*[184,185]. Knockdown efficiency of each morpholino was validated using whole-mount immunohistochemistry. We observed significant reduction of Laminin staining throughout the pancreatic progenitor cell domain in embryos treated with morpholinos targeting *lama1* (**Supplementary Figure 3.10a-d**). In embryos treated with morpholinos targeting *crb2a* or *crb2a* and *crb2b*, we observed loss of staining in the pancreatic progenitor cell domain using antibodies specific to Crb2a or antibodies detecting all Crb proteins,

respectively (**Supplementary Figure 3.11a-h**) Residual panCrb protein signal was observed in the dorsal pancreas, which may be the result of expression of Crb proteins other than Crb2a and Crb2b in this region.

Consistent with prior studies[184], *lama1* morphants exhibited reduced body size and other gross anatomical defects at 78 hpf, whereas *crb2a/b* morphants appeared grossly normal. Both *lama1* and *crb2a/b* morphants displayed an annular pancreas (15 out of 34 *lama1* and 27 out of 69 *crb2a/b* morphants) characterized by pancreatic tissue partially or completely encircling the duodenum (**Fig. 3.6a-d**), a phenotype indicative of impaired migration of pancreatic progenitors during pancreas formation. These findings suggest that both *lama1* and *crb2a/b* control cell migratory processes during early pancreatic development and that reduced levels of *lama1* or *crb2a/b* impair pancreas morphogenesis.

To gain insight into the effects of *lama1* and *crb2a/b* knockdown on pancreatic endocrine cell development, we examined beta cell numbers (insulin$^+$ cells) at 78 hpf. We also evaluated potential synergistic effects of combined *lama1* and *crb2a/b* knockdown. To account for the reduction in body and pancreas size in *lama1* morphants, we compared cell numbers in 78 hpf *lama1* morphants with 50 hpf control embryos, which have a similarly sized acinar compartment as 78 hpf *lama1* morphants. Beta cell numbers were significantly reduced in both *lama1* and *crb2a/b* morphants ($p = 8.0 \times 10^{-3}$ and 4.0×10^{-3} for comparisons of *lama1* and *crb2a/b* morphants, respectively; **Fig. 3.6e,f**), as well as in morphants with a combined knockdown of *lama1* and *crb2a/b* ($p = 2.0 \times 10^{-4}$; **Fig. 3.6f**), showing that reduced *lama1* and *crb2a/b* levels, both individually and in combination, impair beta cell development. Furthermore, we found that nearly all *lama1*, *crb2a/b,* and combined *lama1* and *crb2a/b* morphants without an annular pancreas had reduced beta cell numbers, indicating independent roles of *lama1* and *crb2* in pancreas morphogenesis and beta cell differentiation. Finally, to investigate the contributions of individual *crb2* genes to the observed phenotype, we performed knockdown experiments using morpholinos against *crb2a* and *crb2b* alone. Only *crb2b* morphants showed a significant reduction in beta cell numbers (p =

4.4×10^{-2}; **Supplementary Figure 3.12**), suggesting that *crb2b* is the predominant *crb2* gene required for beta cell development. Combined, these findings demonstrate that *lama1* and *crb2* are regulators of pancreas morphogenesis and beta cell development *in vivo*.

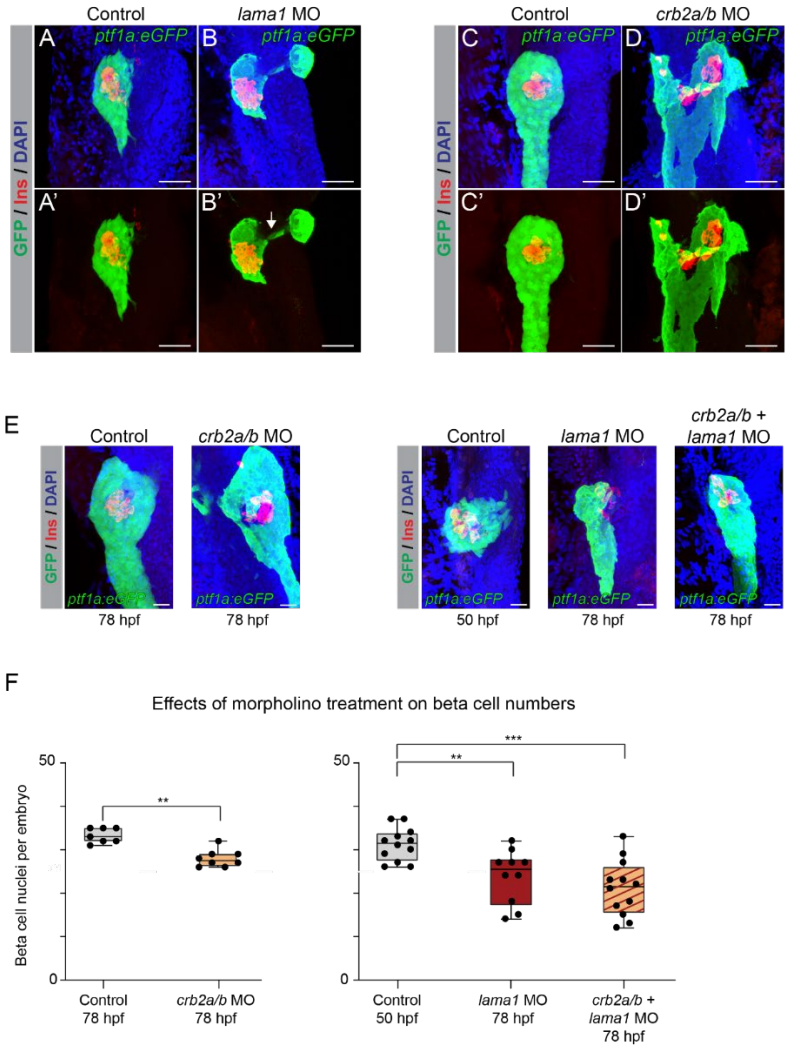

Figure 3.6: *lama1* and *crb2* regulate pancreas morphogenesis and beta cell differentiation. (**A,B**) Representative 3D renderings of *Tg(ptf1a:eGFP)* control zebrafish embryos (**A,A'**) and *lama1* morphants (**B,B'**) stained with DAPI (nuclei, blue) and antibody against insulin (red); n ≥ 15 embryos per condition. To account for reduced acinar pancreas size in *lama1* morphants, control embryos were imaged at 50 hours post fertilization (hpf) and *lama1* morphants at 78 hpf. 15 out of 34 *lama1* morphants displayed an annular pancreas with two acinar pancreas domains (green) connected behind the presumptive intestine (**B'**, white arrow). Scale bar, 40 μM. (**C,D**) Representative 3D renderings of 78 hpf *Tg(ptf1a:eGFP)* control zebrafish embryos (**C,C'**) and *crb2a/b* morphants (**D,D'**) stained with DAPI (nuclei, blue) and antibodies against insulin (red); n ≥ 15 embryos per condition. 27 out of 69 *crb2a/b* morphants displayed an annular pancreas with the acinar pancreas (green) completely surrounding the presumptive intestine. Scale bar, 40 μM. (**E**) Representative 3D renderings of *Tg(ptf1a:eGFP)* control zebrafish embryos and *crb2a/b, lama1*, or *crb2a/b* + *lama1* morphants stained with DAPI (nuclei, blue) and antibody against insulin (red). All embryos were imaged at 78 hpf except for controls to *lama1* and *crb2a/b* + *lama1* morphants, which were imaged at 50 hpf to account for reduced acinar pancreas size of *lama1* morphants. Scale bar, 20 μM. (**F**) Quantification of beta (insulin$^+$) cell nuclei per embryo from experiment in (E). p adj. = 4.0×10^{-3}, 8.0×10^{-3}, and 2.0×10^{-4} for comparison of hfp 78 control (n = 7 embryos) to hfp 78 *crb2a/b* (n = 8), hpf 50 control (n = 12) to hpf 78 *lama1* (n = 10), or *crb2a/b* + *lama1* (n = 12) morphants, respectively; ANOVA-Dunnett's multiple comparison test; *** $p < 0.001$ ** $p < 0.01$. 5 out of 8 *crb2a/b*, 3 out of 10 *lama1*, and 9 out of 12 *crb2a/b* + *lama1* morphants displayed an annular pancreas. MO, morpholino; Control, standard control morpholino.

Discussion

In this study, we identify T2D-associated variants localized within chromatin active in pancreatic progenitors but not islets or other T2D-relevant tissues, suggesting a novel mechanism whereby a subset of T2D risk variants specifically alters pancreatic developmental processes. We link T2D-associated enhancers active in pancreatic progenitors to the regulation of *LAMA1* and *CRB2* and demonstrate a functional requirement in zebrafish for *lama1* and *crb2* in pancreas morphogenesis and endocrine cell formation. Furthermore, we provide a curated list of T2D risk-associated enhancers and candidate effector genes for further exploration of how the regulation of developmental processes in the pancreas can predispose to T2D.

Our analysis identified eleven loci where T2D-associated variants mapped in SE specifically active in pancreatic progenitors. Among these loci was *LAMA1*, which has stronger effects on T2D risk in lean compared to obese individuals[171]. We also found evidence that variants in PSSE collectively have stronger enrichment for T2D in lean individuals, although the small number of PSSE and limited sample size of the BMI-stratified T2D genetic data prohibits a more robust comparison. There was also a notable lack of enrichment among PSSE variants for association with traits related to insulin secretion and beta cell function. If T2D-associated variants in PSSE indeed confer diabetes susceptibility by affecting beta cell development, the question arises as to why variants associated with traits related to beta cell function are not enriched within PSSE. As genetic association studies of endophenotypes are based on data from non-diabetic subjects, a possible explanation is that variants affecting beta cell developmental processes have no overt phenotypic effect under physiological conditions and contribute to T2D pathogenesis only during the disease process.

Since the genomic position of enhancers and transcription factor binding sites is not well conserved between species[186], a human cell model is necessary to identify target genes of enhancers associated with disease risk. By employing enhancer deletion in hESCs, we demonstrate that T2D-associated PSSE at the *LAMA1* and *CRB2* loci regulate *LAMA1* and *CRB2*,

respectively, and establish *LAMA1* and *CRB2* as the predominant target gene of their corresponding PSSE within TAD boundaries. By analyzing *LAMA1* and *CRB2* expression throughout the pancreatic differentiation time course, we show that the identified PSSE control *LAMA1* and *CRB2* expression in a temporal manner consistent with the activation pattern of their associated PSSE. While the specific T2D-relevant target genes of the majority of T2D-associated PSSE remain to be identified, it is notable that several are localized within TADs containing genes encoding transcriptional regulators. These include *PROX1* and *GATA4*, which are known to regulate pancreatic development[127,128,187], as well as *HMGA2* and *BCL6* with unknown functions in the pancreas. Our catalogue of T2D-associated PSSE provides a resource to fully characterize the gene regulatory program associated with developmentally mediated T2D risk in the pancreas. Our finding that predicted target genes of PSSE are similarly expressed in hESC-derived pancreatic progenitors and primary human embryonic pancreas (**Fig. 3.3b** and **Supplementary Figure 3.3a**) further underscores the utility of the hESC-based system for these studies.

In the embryo, endocrine cells differentiate by delaminating from a polarized epithelium of progenitors governed by local cell-cell and cell-matrix signaling events[49]. These processes are not well-recapitulated in the hESC-based pancreatic differentiation system, highlighting a limitation of this system for studying the function of Laminin and CRB2, which are mediators of mechanical signals within an epithelium. Therefore, we analyzed their function in zebrafish as an *in vivo* model. We show that *lama1* or *crb2* knockdown leads to an annular pancreas and reduced beta cell numbers. The beta cell differentiation defect was also evident in embryos not displaying an annular pancreas, suggesting independent mechanisms.

Consistent with our findings in *lama1* morphants, culture of pancreatic progenitors on Laminin-based substrates promotes endocrine cell differentiation[49]. During *in vivo* pancreatic development, endothelial cells are an important albeit not the only source of Laminin in the pancreas[49,188,189]. While we do not know the respective contributions of endothelial cell- and pancreatic progenitor cell-derived Laminin to the phenotype of *lama1* morphants, the T2D-

associated *LAMA1* PSSE is not active in endothelial cells (**Supplementary Figure 3.3c**). Furthermore, we found no other T2D-associated variants at the *LAMA1* locus mapping in endothelial cell enhancers or accessible chromatin sites in islets, suggesting that T2D risk is linked to *LAMA1* regulation in pancreatic progenitors.

Similar to Laminin, CRB2 has been shown to regulate mechanosignaling[190]. Our observation that pancreatic progenitor cells express Crb proteins is consistent with the phenotype of *crb2* morphants reflecting a progenitor-autonomous role of Crb2. Furthermore, the similarity in pancreatic phenotype between *lama1* or *crb2* morphants raises the possibility that signals from Laminin and Crb2 could converge on the same intracellular pathways in pancreatic progenitors.

Our findings suggest that variation in gene regulation during pancreatic development can predispose to T2D later in life. Several lines of evidence support the concept of a developmental impact on T2D risk. First, human genetic studies have shown a strong correlation between birth weight and adult cardiometabolic traits and disease[191]. Second, epidemiological studies provide evidence that offspring of mothers who were pregnant during a famine have a higher prevalence of T2D[158]. This phenomenon has been experimentally reproduced in rodents, where maternal malnutrition has been shown to cause reduced beta cell mass at birth and to render beta cells more prone to failure under stress[160]. Together, our results provide a strong rationale for further exploration of how genetic variants affecting developmental gene regulation in the pancreas contribute to T2D risk.

Methods

Maintenance and differentiation of CyT49 hESCs

Genomic and gene expression analyses (ChIP-seq, ATAC-seq, RNA-seq) for generation of chromatin maps and target gene identification were performed in CyT49 hESCs (male). Propagation of CyT49 hESCs was carried out by passing cells every 3 to 4 days using Accutase™ (eBioscience) for enzymatic cell dissociation, and with 10% (v/v) human AB serum (Valley Biomedical) included in the hESC media the day of passage. hESCs were seeded into tissue culture flasks at a density of 50,000 cells/cm^2. hESC research was approved by the University of California, San Diego, Institutional Review Board and Embryonic Stem Cell Research oversight committee.

Pancreatic differentiation was performed as previously described[27,67,68]. Briefly, a suspension-based culture format was used to differentiate cells in aggregate form. Undifferentiated aggregates of hESCs were formed by re-suspending dissociated cells in hESC maintenance medium at a concentration of 1×10^6 cells/mL and plating 5.5 mL per well of the cell suspension in 6-well ultra-low attachment plates (Costar). The cells were cultured overnight on an orbital rotator (Innova2000, New Brunswick Scientific) at 95 rpm. After 24 hours the undifferentiated aggregates were washed once with RPMI medium and supplied with 5.5 mL of day 0 differentiation medium. Thereafter, cells were supplied with the fresh medium for the appropriate day of differentiation (see below). Cells were continually rotated at 95 rpm, or 105 rpm on days 4 through 8, and no media change was performed on day 10. Both RPMI (Mediatech) and DMEM High Glucose (HyClone) medium were supplemented with 1X GlutaMAX™ and 1% penicillin/streptomycin. Human activin A, mouse Wnt3a, human KGF, human noggin, and human EGF were purchased from R&D systems. Other added components included FBS (HyClone), B-27® supplement (Life Technologies), Insulin-Transferrin-Selenium (ITS; Life Technologies), TGFβ R1 kinase inhibitor IV (EMD Bioscience), KAAD-Cyclopamine (KC; Toronto Research

Chemicals), and the retinoic receptor agonist TTNPB (RA; Sigma Aldrich). Day-specific differentiation media formulations were as follows:

Days 0 and 1: RPMI + 0.2% (v/v) FBS, 100 ng/mL Activin, 50 ng/mL mouse Wnt3a, 1:5000 ITS.

Days 1 and 2: RPMI + 0.2% (v/v) FBS, 100ng/mL Activin, 1:5000 ITS

Days 2 and 3: RPMI + 0.2% (v/v) FBS, 2.5 mM TGFβ R1 kinase inhibitor IV, 25ng/mL KGF, 1:1000 ITS

Days 3 – 5: RPMI + 0.2% (v/v) FBS, 25ng/mL KGF, 1:1000 ITS

Days 5 – 8: DMEM + 0.5X B-27® Supplement, 3 nM TTNPB, 0.25 mM KAAD-Cyclopamine, 50ng/mL Noggin

Days 8 – 10: DMEM/B-27, 50ng/mL KGF, 50ng/mL EGF

Cells at D0 correspond to the embryonic stem cell (ES) stage, cells at D2 correspond to the definitive endoderm (DE) stage, cells at D5 correspond to the gut tube (GT) stage, cells at D7 correspond to the early pancreatic progenitor (PP1) stage, and cells at D10 correspond to the late pancreatic progenitor (PP2) stage.

Maintenance and differentiation of H1 hESCs

Δ*LAMA1Enh* and Δ*CRB2Enh* clonal lines were derived by targeting H1 hESCs (male). Cells were maintained and differentiated as described with some modifications[69,70]. In brief, hESCs were cultured in mTeSR1 media (Stem Cell Technologies) and propagated by passaging cells every 3 to 4 days using Accutase (eBioscience) for enzymatic cell dissociation. hESC research was approved by the University of California, San Diego, Institutional Review Board and Embryonic Stem Cell Research Oversight Committee.

For differentiation, cells were dissociated using Accutase for 10 min, then reaggregated by plating the cells at a concentration of ~5.5 e6 cells/well in a low attachment 6-well plate on an orbital shaker (100 rpm) in a 37 °C incubator. The following day, undifferentiated cells were

washed in base media (see below) and then differentiated using a multi-step protocol with stage-specific media and daily media changes.

All stage-specific base media were comprised of MCDB 131 medium (Thermo Fisher Scientific) supplemented with NaHCO3, GlutaMAX, D-Glucose, and BSA using the following concentrations:

Stage 1/2 medium: MCDB 131 medium, 1.5 g/L NaHCO3, 1X GlutaMAX, 10 mM D-Glucose, 0.5% BSA

Stage 3/4 medium: MCDB 131 medium, 2.5 g/L NaHCO3, 1X GlutaMAX, 10 mM D-glucose, 2% BSA

Media compositions for each stage were as follows:

Stage 1 (day 0-2): base medium, 100 ng/ml Activin A, 25 ng/ml Wnt3a (day 0). Day 1-2: base medium, 100 ng/ml Activin A

Stage 2 (day 3-5): base medium, 0.25 mM L-Ascorbic Acid (Vitamin C), 50 ng/mL FGF7

Stage 3 (day 6-7): base medium, 0.25 mM L-Ascorbic Acid, 50 ng/mL FGF7, 0.25 μM SANT-1, 1 μM Retinoic Acid, 100 nM LDN193189, 1:200 ITS-X, 200 nM TPB

Stage 4 (day 8-10): base medium, 0.25 mM L-Ascorbic Acid, 2 ng/mL FGF7, 0.25 μM SANT-1, 0.1 μM Retinoic Acid, 200 nM LDN193189, 1:200 ITS-X, 100nM TPB

Cells at D0 correspond to the embryonic stem cell (ES) stage, cells at D3 correspond to the definitive endoderm (DE) stage, cells at D6 correspond to the gut tube (GT) stage, cells at D8 correspond to the early pancreatic progenitor (PP1) stage, and cells at D11 correspond to the late pancreatic progenitor (PP2) stage.

Generation of ΔLAMA1Enh, ΔCRB2Enh, ΔLAMA1, and ΔCRB2 hESC lines

To generate clonal homozygous LAMA1Enh and CRB2Enh deletion hESC lines, sgRNAs targeting each relevant enhancer were designed and cloned into Px333-GFP, a modified version of Px333 (Addgene, #64073). To generate clonal homozygous LAMA1 and CRB2 deletion hESC

lines, sgRNAs targeting the second exon of each gene were designed and cloned into Px458 (Addgene, #48138). Plasmids expressing the sgRNAs were transfected into H1 hESCs with XtremeGene 9 (Roche). 24 hours later, 8000 GFP+ cells were sorted into a well of six-well plate. Individual colonies that emerged within 5-7 days after transfection were subsequently transferred manually into 48-well plates for expansion, genomic DNA extraction, PCR genotyping, and Sanger sequencing. sgRNA oligos are listed below.

LAMA1Enh Upstream Guide: GTCAAATTGCTATAACACGG

LAMA1Enh Downstream Guide: CCACTTTAAGTATCTCAGCA

CRB2Enh Upstream Guide: ATACAAAGCACGTGAGA

CRB2Enh Downstream Guide: GAATGCGGATGACGCCTGAG

LAMA1 Exon 2 Guide: ATCAGCACCAATGCCACCTG

CRB2 Exon 2 Guide: TCGATGTCCAGCTCGCAGCG

Human tissue

Human embryonic pancreas tissue was obtained from the Birth Defects Research Laboratory of the University of Washington. Studies for use of embryonic human tissue were approved by the Institutional Review Board of the University of California, San Diego. A pancreas from a 54 and 58 day gestation embryo each were pooled for RNA-seq analysis.

Zebrafish husbandry

Adult zebrafish and embryos were cared for and maintained under standard conditions. All research activity involving zebrafish was reviewed and approved by SBP Medical Discovery Institute Institutional Animal Care and Use Committee. The following transgenic lines were used: *Tg(ptf1a:eGFP)*^{jh1192}.

Morpholino injections in zebrafish

The following previously validated morpholinos were injected into the yolk at the 1-cell stage in a final volume of either 0.5 or 1 nl: 0.75 ng lama1-ATG (5'- TCATCCT CATCTCCATCATCGCTCA -3'); 3 ng crb2a-SP, (5'-ACGTTGCCAGTACCTGTGTATCCTG-3')[185,193]; 3 ng crb2b-SP, (5'-TAAAGATGTCCTACCCAGCTTGAAC-3')[185]; 6.75 ng standard control MO (5'- CCTCTTACCTCAGTTACAATTTATA -3'). All morpholinos were obtained from GeneTools, LLC.

Chromatin Immunoprecipitation Sequencing (ChIP-seq)

ChIP-seq was performed using the ChIP-IT High-Sensitivity kit (Active Motif) according to the manufacturer's instructions. Briefly, for each cell stage and condition analyzed, 5-10 × 106 cells were harvested and fixed for 15 min in an 11.1% formaldehyde solution. Cells were lysed and homogenized using a Dounce homogenizer and the lysate was sonicated in a Bioruptor® Plus (Diagenode), on high for 3 × 5 min (30 sec on, 30 sec off). Between 10 and 30 µg of the resulting sheared chromatin was used for each immunoprecipitation. Equal quantities of sheared chromatin from each sample were used for immunoprecipitations carried out at the same time. 4 µg of antibody were used for each ChIP-seq assay. Chromatin was incubated with primary antibodies overnight at 4 °C on a rotator followed by incubation with Protein G agarose beads for 3 hours at 4 °C on a rotator. Antibodies used were rabbit anti-H3K27ac (Active Motif 39133), rabbit anti-H3K4me1 (Abcam ab8895), rabbit anti-H3K4me3 (Millipore 04-745), rabbit anti-H3K27me3 (Millipore 07-499), goat anti-CTCF (Santa Cruz Biotechnology SC-15914X), goat anti-GATA4 (Santa Cruz SC-1237), rabbit anti-GATA6 (Santa Cruz SC-9055), goat anti-FOXA1 (Abcam Ab5089), goat-anti-FOXA2 (Santa Cruz SC-6554), rabbit anti-PDX1 (BCBC AB1068), rabbit anti-HNF6 (Santa Cruz SC-13050), and rabbit anti-SOX9 (Chemicon AB5535). Reversal of crosslinks and DNA purification were performed according to the ChIP-IT High-Sensitivity instructions, with the modification of incubation at 65 °C for 2-3 hours, rather than at 80 °C for 2 hours. Sequencing libraries were constructed using KAPA DNA Library Preparation Kits for

Illumina® (Kapa Biosystems) and library sequencing was performed on either a HiSeq 4000 System (Illumina®) or NovaSeq 6000 System (Illumina®) with single-end reads of either 50 or 75 base pairs (bp). Sequencing was performed by the Institute for Genomic Medicine (IGM) core research facility at the University of California at San Diego (UCSD). Two replicates from independent hESC differentiations were generated for each ChIP-seq experiment.

ChIP-seq data analysis

ChIP-seq reads were mapped to the human genome consensus build (hg19/GRCh37) and visualized using the UCSC Genome Browser[77]. Burrows-Wheeler Aligner (BWA)[78] version 0.7.13 was used to map data to the genome. Unmapped and low-quality ($q<15$) reads were discarded. SAMtools[79] was used to remove duplicate sequences and HOMER[80] was used to call peaks using default parameters and to generate tag density plots. Stage- and condition-matched input DNA controls were used as background when calling peaks. The BEDtools suite of programs[81] was used to perform genomic algebra operations. For all ChIP-seq experiments, replicates from two independent hESC differentiations were generated. Tag directories were created for each replicate using HOMER. Directories from each replicate were then combined, and peaks were called from the combined replicates. For histone modifications and CTCF peaks, pearson correlations between each pair of replicates were calculated over the called peaks using the command multiBamSummary from the deepTools2 package[82]. For pancreatic lineage-determining transcription factors (GATA4, GATA6, FOXA1, FOXA2, HNF6, PDX1, SOX9), correlations were calculated for peaks overlapping PSSE. Calculated Pearson correlations are listed in **Supplementary Figure 3.13**.

RNA isolation and sequencing (RNA-seq) and qRT-PCR

RNA was isolated from cell samples using the RNeasy® Micro Kit (Qiagen) according to the manufacturer instructions. For each cell stage and condition analyzed between 0.1 and 1 ×

10^6 cells were collected for RNA extraction. For qRT-PCR, cDNA synthesis was first performed using the iScript™ cDNA Synthesis Kit (Bio-Rad) and 500 ng of isolated RNA per reaction. qRT-PCR reactions were performed in triplicate with 10 ng of template cDNA per reaction using a CFX96™ Real-Time PCR Detection System and the iQ™ SYBR® Green Supermix (Bio-Rad). PCR of the TATA binding protein (TBP) coding sequence was used as an internal control and relative expression was quantified via double delta CT analysis. For RNA-seq, stranded, single-end sequencing libraries were constructed from isolated RNA using the TruSeq® Stranded mRNA Library Prep Kit (Illumina®) and library sequencing was performed on either a HiSeq 4000 System (Illumina®) or NovaSeq 6000 System (Illumina®) with single-end reads of either 50 or 75 base pairs (bp). Sequencing was performed by the Institute for Genomic Medicine (IGM) core research facility at the University of California at San Diego. RT-qPCR primer sequences are listed in **Supplementary Figure 3.14**.

RNA-seq data analysis

Reads were mapped to the human genome consensus build (hg19/GRCh37) using the Spliced Transcripts Alignment to a Reference (STAR) aligner v2.4[83]. Normalized gene expression (fragments per kilobase per million mapped reads; FPKM) for each sequence file was determined using Cufflinks v2.2.1[84] with the parameters: --library-type fr-firststrand --max-bundle-frags 10000000. For all RNA-Seq experiments, replicates from two independent hESC differentiations were generated. Pearson correlations between bam files corresponding to each pair of replicates are listed in **Supplementary Figure 3.15**.

Assay for Transposase Accessible Chromatin Sequencing (ATAC-seq)

ATAC-seq[86] was performed on approximately 50,000 nuclei. The samples were permeabilized in cold permeabilization buffer (0.2% IGEPAL-CA630 (I8896, Sigma), 1 mM DTT (D9779, Sigma), Protease inhibitor (05056489001, Roche), 5% BSA (A7906, Sigma) in PBS

(10010-23, Thermo Fisher Scientific) for 10 minutes on the rotator in the cold room and centrifuged for 5 min at 500 × g at 4°C. The pellet was resuspended in cold tagmentation buffer (33 mM Tris-acetate (pH = 7.8) (BP-152, Thermo Fisher Scientific), 66 mM K-acetate (P5708, Sigma), 11 mM Mg-acetate (M2545, Sigma), 16% DMF (DX1730, EMD Millipore) in Molecular biology water (46000-CM, Corning)) and incubated with tagmentation enzyme (FC-121-1030; Illumina) at 37 °C for 30 min with shaking at 500 rpm. The tagmented DNA was purified using MinElute PCR purification kit (28004, QIAGEN). Libraries were amplified using NEBNext High-Fidelity 2X PCR Master Mix (M0541, NEB) with primer extension at 72°C for 5 minutes, denaturation at 98°C for 30 s, followed by 8 cycles of denaturation at 98°C for 10 s, annealing at 63°C for 30 s and extension at 72°C for 60 s. After the purification of amplified libraries using MinElute PCR purification kit (28004, QIAGEN), double size selection was performed using SPRIselect bead (B23317, Beckman Coulter) with 0.55X beads and 1.5X to sample volume. Finally, libraries were sequenced on HiSeq4000 (Paired-end 50 cycles, Illumina).

ATAC-seq data analysis

ATAC-seq reads were mapped to the human genome (hg19/GRCh37) using Burrows-Wheeler Aligner (BWA) version 0.7.13[78], and visualized using the UCSC Genome Browser[77]. SAMtools[79] was used to remove unmapped, low-quality (q<15), and duplicate reads. MACS2[87] was used to call peaks, with parameters “shift set to 100 bps, smoothing window of 200 bps” and with “nolambda” and “nomodel” flags on. MACS2 was also used to call ATAC-Seq summits, using the same parameters combined with the “call-summits” flag.

For all ATAC-Seq experiments, replicates from two independent hESC differentiations were generated. Bam files for each pair of replicates were merged for downstream analysis using SAMtools, and Pearson correlations between bam files for each individual replicate were calculated over a set of peaks called from the merged bam file. Correlations were performed using

the command multiBamSummary from the deepTools2 package[82] with the "—removeOutliers" flag and are listed in **Supplementary Figure 3.16**.

For downstream analysis, ATAC-seq peaks were merged from 2 independent differentiations for ES, DE, GT, PP1, and PP2 stage cells and from 4 donors for primary islets. Primary islet ATAC-seq data was obtained from previously published datasets[153].

Hi-C data analysis

Hi-C data were processed as previously described with some modifications[194]. Read pairs were aligned to the hg19 reference genome separately using BWA-MEM with default parameters[78]. Specifically, chimeric reads were processed to keep only the 5' position and reads with low mapping quality (<10) were filtered out. Read pairs were then paired using custom scripts. Picard tools were then used to remove PCR duplicates. Bam files with alignments were further processed into text format as required by Juicebox tools[89]. Juicebox tools were then applied to generate hic files containing normalized contact matrices. All downstream analysis was based on 10 Kb resolution KR normalized matrices.

Chromatin loops were identified by comparing each pixel with its local background, as described previously[90] with some modifications. Specifically, only the donut region around the pixel was compared to model the expected count. Briefly, the KR-normalized contact matrices at 10 Kb resolution were used as input for loop calling. For each pixel, distance-corrected contact frequencies were calculated for each surrounding bin and the average of all surrounding bins. The expected counts were then transformed to raw counts by multiplying the counts with the raw-to-KR normalization factor. The probability of observing raw expected counts was calculated using Poisson distribution. All pixels with p-value < 0.01 and distance less than 10 Kb were selected as candidate pixels. Candidate pixels were then filtered to remove pixels without any neighboring candidate pixels since they were likely false positives. Finally, pixels within 20 Kb of each other

were collapsed and only the most significant pixel was selected. The collapsed pixels with p-value < 1e-5 were used as the final list of chromatin loops.

A full set of scripts used for processing Hi-C data is available at https://github.com/MSanderlab/Pancreatic-progenitor-epigenome-maps-prioritize-type-2-diabetes-risk-genes-with-roles-in-development/tree/master.

Definition of chromatin states

We collected or generated H3K4me1, H3K27ac, H3K4me1, H3K4me3, H3K27me3 and CTCF ChIP-seq data at each developmental stage and in mature islets. Data corresponding to mature islets was downloaded from previously published studies[145,146,195]. Sequence reads were mapped to the human genome hg19 using bwa (version 0.7.12)[78], and low quality and duplicate reads were filtered using samtools (version 1.3)[79]. Using these reads, we then called chromatin states jointly across all data using chromHMM (version 1.12)[161] and used a 10-state model and 200bp bin size, as models with larger state numbers did not empirically resolve any additional informative states. We then assigned state names based on patterns defined by the NIH Epigenome Roadmap[172], which included active promoter/TssA (high H3K4me3, high H3K27ac), flanking TSS/TssFlnk1 (high H3K4me3), flanking TSS/TssFlnk 2 (high H3K4me3, high H3K27ac, high H3K4me1), bivalent Tss/TssBiv (high H3K27me3, high H3K4me3), poised enhancer/EnhP (high H3K4me1), insulator/CTCF (high CTCF), active enhancer/EnhA (high H3K27ac, high H3K4me1), repressor (high H3K27me3), and two quiescent (low signal for all assays) states. The state map with assigned names is shown in **Supplementary Figure 3.1a**.

We next defined stretch enhancer elements at each developmental stage and in mature islets. For each active enhancer (EnhA) element, we determined the number of consecutive 200bp bins covered by the enhancer. We then modeled the resulting bin counts for enhancers in each cell type using a Poisson distribution. Enhancers with a p-value less than .001 were labelled as stretch enhancers and otherwise labelled as traditional enhancers.

Permutation-based significance

A random sampling approach (10,000 iterations) was used to obtain null distributions for enrichment analyses, in order to obtain p-values. Null distributions for enrichments were obtained by randomly shuffling enhancer regions using BEDTools[81] and overlapping with ATAC-seq peaks. P-values < 0.05 were considered significant.

Assignment of enhancer target genes

Transcriptomes were filtered for genes expressed (FPKM ≥ 1) at each relevant stage, and BEDTools[81] was used to assign each enhancer to the nearest annotated TSS.

Gene ontology

All gene ontology analyses were performed using Metascape[94] with default parameters.

Motif enrichment analysis

The findMotifsGenome.pl. command in HOMER[80] was used to identify enriched transcription factor binding motifs. *de novo* motifs were assigned to transcription factors based on suggestions generated by HOMER.

T2D-relevant trait enrichment analysis

GWAS summary statistics for T2D[20,171], metabolic traits (HOMA-B, HOMA-IR[196], fasting glucose, fasting insulin[197], fasting proinsulin[198], 2 hour glucose adjusted for BMI[199], HbA1c, insulin secretion rate, disposition index, acute insulin response, peak insulin response)[200], and developmental traits (head circumference[201], birth length[202], birth weight[191]) conducted with individuals of European ancestry were obtained from various sources including the MAGIC consortium, EGG consortium, and authors of the studies. Custom LD score annotation files were

created for PSSE, PP2 stretch enhancers, and islet stretch enhancers using LD score regression version 1.0.1[169]. Enrichments for GWAS trait-associated variants within PSSE, PP2 stretch enhancers, and islet stretch enhancers were estimated with stratified LD score regression[170]. We next determined enrichment in the proportion of variants in accessible chromatin sites within islet SE and PSSE with nominal association to beta cell-related glycemic traits. For each trait, we calculated a 2x2 table of variants mapping in and outside of islet SE or PSSE and with or without nominal association and then determined significance using a chi-square test.

Adipocyte differentiation analysis

Chromatin states for human adipose stromal cell (hASC) differentiation stages (1-4) were obtained from a published study[151]. PSSE were intersected with hASC chromatin states using BEDTools intersect (version 2.26.0)[81] with default parameters.

Identification of T2D risk loci intersecting PSSE

T2D GWAS summary statistics were obtained from the DIAMANTE consortium[20]. Intersection of variants and PSSE was performed using BEDTools intersect (version 2.26.0)[81] with default parameters. The adjusted significance threshold was set at $P < 4.66 \times 10^{-6}$ (Bonferroni correction for 10,738 variants mapping in PSSE). Putative novel loci were defined as those with 1) at least one variant in a PSSE reaching the adjusted significance threshold and 2) mapping at least 500 kb away from a known T2D locus.

ATAC-seq footprinting analysis

ATAC-seq footprinting was performed as previously described[98]. In brief, diploid genomes for CyT49 were created using vcf2diploid (version 0.2.6a)[99] and genotypes called from whole genome sequencing and scanned for a compiled database of TF sequence motifs from JASPAR[100] and ENCODE[101] with FIMO (version 4.12.0)[96] using default parameters for p-value

threshold and a 40.9% GC content based on the hg19 human reference genome. Footprints within ATAC-seq peaks were discovered with CENTIPEDE (version 1.2)[102] using cut-site matrices containing Tn5 integration counts within a ±100 bp window around each motif occurrence. Footprints were defined as those with a posterior probability ≥ 0.99.

Generation of similarity matrices for total transcriptomes

For each replicate, FPKM values corresponding to total transcriptome were filtered for genes expressed (FPKM ≥ 1) in ≥ 1 replicate. For expressed genes, log(FPKM+1) values were used to calculate Pearson correlations.

Immunofluorescence analysis

Cell aggregates derived from hESCs were allowed to settle in microcentrifuge tubes and washed twice with PBS before fixation with 4% paraformaldehyde (PFA) for 30 min at room temperature. Fixed samples were washed twice with PBS and incubated overnight at 4 °C in 30% (w/v) sucrose in PBS. Samples were then loaded into disposable embedding molds (VWR), covered in Tissue-Tek® O.C.T. Sakura® Finetek compound (VWR) and flash frozen on dry ice to prepare frozen blocks. The blocks were sectioned at 10 µm and sections were placed on Superfrost Plus® (Thermo Fisher) microscope slides and washed with PBS for 10 min. Slide-mounted cell sections were permeabilized and blocked with blocking buffer, consisting of 0.15% (v/v) Triton X-100 (Sigma) and 1% (v/v) normal donkey serum (Jackson Immuno Research Laboratories) in PBS, for 1 hour at room temperature. Slides were then incubated overnight at 4 °C with primary antibody solutions. The following day slides were washed five times with PBS and incubated for 1 hour at room temperature with secondary antibody solutions. Cells were washed five times with PBS before coverslips were applied.

All antibodies were diluted in blocking buffer at the ratios indicated below. Primary antibodies used were goat anti-PDX1 (1:500 dilution, Abcam ab47383), mouse anti-NKX6.1

(1:300 dilution, Developmental Studies Hybridoma Bank F64A6B4), rabbit anti-Laminin (1:30, Sigma L-9393), and rabbit anti-CRB2 (1:500, Sigma SAB1301340). Secondary antibodies against goat and mouse were Alexa488- and Cy3-conjugated donkey antibodies, respectively (Jackson Immuno Research Laboratories 705-545-003 and 715-165-150, respectively), and were used at dilutions of 1:500 (anti-goat Alexa488) or 1:1000 (anti-mouse Cy3). Cell nuclei were stained with Hoechst 33342 (1:3000, Invitrogen). Representative images were obtained with a Zeiss Axio-Observer-Z1 microscope equipped with a Zeiss ApoTome and AxioCam digital camera. Figures were prepared in Adobe Creative Suite 5.

Flow cytometry analysis

Cell aggregates derived from hESCs were allowed to settle in microcentrifuge tubes and washed with PBS. Cell aggregates were incubated with Accutase® at room temperature until a single-cell suspension was obtained. Cells were washed with 1 mL ice-cold flow buffer comprised of 0.2% BSA in PBS and centrifuged at 200 g for 5 min. BD Cytofix/Cytoperm™ Plus Fixation/Permeabilization Solution Kit was used to fix and stain cells for flow cytometry according to the manufacturer's instructions. Briefly, cell pellets were re-suspended in ice-cold BD Fixation/Permeabilization solution (300 µL per microcentrifuge tube). Cells were incubated for 20 min at 4 °C. Cells were washed twice with 1 mL ice-cold 1 × BD Perm/Wash™ Buffer and centrifuged at 10 °C and 200 × g for 5 min. Cells were re-suspended in 50 µL ice-cold 1 × BD Perm/Wash™ Buffer containing diluted antibodies, for each staining performed. Cells were incubated at 4 °C in the dark for 1-3 hrs. Cells were washed with 1.25 mL ice-cold 1X BD Wash Buffer and centrifuged at 200 × g for 5 min. Cell pellets were re-suspended in 300 µL ice-cold flow buffer and analyzed in a FACSCanto™ II (BD Biosciences). Antibodies used were PE-conjugated anti-PDX1 (1:10 dilution, BD Biosciences); and AlexaFluor® 647-conjugated anti-NKX6.1 (1:5 dilution, BD Biosciences). Data were processed using FlowJo software v10.

Whole mount immunohistochemistry

Zebrafish larvae were fixed and stained according to published protocols[203] using the following antibodies: chicken anti-GFP (1:200; Aves Labs; GFP-1020), guinea pig anti-insulin (1:200; Biomeda; v2024), mouse anti-Crb2a (1:100; ZIRC; zs-4), rabbit anti-panCrb (1:100; provided by Dr. Abbie M. Jensen at University of Massachusetts, Amherst[204]), rabbit anti-Laminin (1:100; Sigma;L9393), mouse anti-Nkx6.1 (1:10; DSHB; F55A10), and DAPI (1:200; 500 mg/ml, Invitrogen; D1306).

Imaging and quantification of beta cell numbers in zebrafish

To quantify beta cell numbers, 50 and 78 hpf zebrafish larvae were stained for confocal imaging using DAPI and guinea pig anti-insulin antibody (1:200; Biomeda; v2024). Whole mount fluorescent confocal Z-stacks (0.9 µm steps) images were collected for the entire islet with optical slices captured at a focal depth of 1.8 µm. Samples were imaged using a Zeiss 710 confocal microscope running Zen 2010 (Black) software. Final images were generated using Adobe Photoshop CS6 and/or ImageJ64 (vs.1.48b).

Data sources

The following datasets used in this study were downloaded from the GEO and ArrayExpress repositories:

RNA-seq: Pancreatic differentiation of CyT49 hESC line (E-MTAB-1086); primary islet data (GSE115327)

ChIP-seq: H3K27ac data in primary islets (E-MTAB-1919 and GSE51311); H3K27ac data in pancreatic differentiation of CyT49 hESC line (GSE54471); H3K4me1 data in pancreatic differentiation of CyT49 hESC line (GSE54471); H3K4me1 data in primary islets (E-MTAB-1919 and E-MTAB 189); H3K27me3 and H3K4me3 in pancreatic differentiation of CyT49 hESC line (E-MTAB-1086); H3K4me3 and H3K27me3 in primary islets (E-MTAB-189); CTCF in primary

islets (E-MTAB-1919); PDX1 in CyT49 PP2 (GSE54471); samples from ROADMAP consortium: http://ncbi.nlm.nih.gov/geo/roadmap/epigenomics

ATAC-seq: primary islet data (PRJN527099); CyT49 PP2 (GSE115327).

Hi-C datasets were generated in collaboration with the Ren laboratory at University of California, San Diego as a component of the 4D Nucleome Project (Dekker et al., 2017) under accession number 4DNES0LVRKBM.

Quantification and statistical analyses

Statistical analyses were performed using GraphPad Prism (v8.1.2), and R (v3.6.1). Statistical parameters, such as the value of n, mean, standard deviation (SD), standard error of the mean (SEM), significance level (*$p < 0.05$, **$p < 0.01$, and ***$p < 0.001$), and the statistical tests used, are reported in the figures and figure legends. The ‘‘n’’ refers to the number of independent pancreatic differentiation experiments analyzed (biological replicates).

Statistically significant gene expression changes were determined with DESeq2[85].

Data availability

All mRNA-seq, ChIP-seq, and ATAC-seq datasets generated for this study have been deposited at GEO under the accession number GSE149148.

Declaration of interests

KJG does consulting for Genentech. The authors declare no other competing interests.

Acknowledgements

We thank Ileana Matta for assistance with ATAC-seq assays and library preparations, as well as the Sander and Gaulton laboratories for helpful discussions. We also thank Dr. Abbie

Jensen at University of Massachusetts, Amherst for the anti-panCrb antibody. We acknowledge support of the UCSD Human Embryonic Stem Cell Core for cell sorting and the K. Jepsen and UCSD IGM Genomic Center for library preparation and sequencing. This work was supported by NIH grants T32 GM008666 (R.J.G.), P30 DK064391 (K.J.), R01 DK068471 (M.S.), 1DP2DK098092 (P.D.S.D.), and U01 DK105541 (M.S., B.R., P.D.S.D.); as well as the W.M. Keck Foundation 2017-01 (P.D.S.D.), and Diabetes Research Connection Project #08 (J.J.L.)

Chapter 3, in full, is a reprint of the material as it appears in eLife 2021. This material was co-authored with Wang, A., Chiou, J., Lancman, J.J., Wetton, N., Kelalopoulou, S., Wang, J., Qiu, Y., Yan, J., Aylward, A., Ren, B., Dong, P., Gaulton, K.J., and Sander, M. The dissertation author was the principal author of this chapter.

CHAPTER 4: INVESTIGATING A ROLE FOR BETA CELLS IN DIRECT ANTIGEN PRESENTATION VIA HLA-DQ8

Abstract

The majority of genetic risk for type 1 diabetes (T1D) is conferred by genes at the major histocompatibility complex (MHC) locus. Specifically, the HLA-DQ8 MHC class II haplotype is most strongly associated with T1D occurrence. This association is thought to be driven by increased flexibility in peptide interactions with HLA-DQ8, which allows for presentation of insulin and other autoantigens on HLA-DQ8 and subsequent stimulation of autoreactive CD4+ T cells. Given that beta cells have been shown to be able to express MHC class II genes, we examined whether beta cells are capable of directly presenting insulin peptides to CD4+ T cells via HLA-DQ8. To this end, we first used a retrovirus to express HLA-DQ8 in the human beta cell line EndoC-βH1 and co-cultured transduced cells with CD4+ T cell clones recognizing various insulin peptides. In this context, we observed stimulation of the CD4+ T cell clones. Next, we tested whether induced human pluripotent stem cell-derived islets (SC-islets) from donors harboring the HLA-DQ8 haplotype are similarly capable of directly stimulating CD4+ T cell clones. In response to cytokine treatment, we observed significant upregulation of MHC class II genes in SC-islet cells, with a subpopulation expressing HLA-DQ8 at the cell surface. However, SC-islets failed to stimulate CD4+ T cell clones upon co-culture, indicating that insulin peptides are not presented on HLA-DQ8 in this context. Current efforts are focused on examining contributions of cell state and environmental effects to the potential for presentation of insulin peptides on HLA-DQ8 by human beta cells.

Introduction

Type 1 diabetes (T1D) occurs when the body develops an autoimmune response against beta cells, insulin-secreting cells located in the islet of Langerhans which are responsible for maintaining blood glucose homeostasis. Currently, T1D affects over 400 million people worldwide, with rates of diagnosis increasing[205]. Development of T1D is the result of complex genetic and environmental factors, many of which remain unknown. However, much of the genetic risk associated with disease development can be attributed to variants within the major histocompatibility complex (MHC) locus[22].

Variants within the MHC locus are generally inherited in groups known as haplotypes. Among MHC haplotypes associated with T1D, risk is most highly correlated with the MHC class II HLA-DQ8 haplotype[206]. The MHC class II molecule facilitates antigen presentation to CD4+ T cells via expression on the surface of an antigen presenting cell. Structural changes conferred by the HLA-DQ8 haplotype have been demonstrated to allow for promiscuous peptide binding on the surface of MHC class II, which may lead to the presentation of autoantigens to autoreactive T cells, thus triggering autoimmunity[23,207]. Specifically, alternate binding registers for the B chain of insulin have been implicated in T1D pathogenesis in humans as well as in mouse models of T1D[24-26,208].

The cell types participating in the presentation of autoantigens and the environmental conditions that drive antigen presentation are not well understood. Although “professional antigen presenting cells” such as macrophages and dendritic cells are likely candidates, beta cells also express MHC class II under various conditions[209-214]. The relevance of this phenomenon to T1D pathogenesis remains largely unexplored.

Here we examine the role of beta cells in presenting antigens to autoreactive CD4+ T cells via the HLA-DQ8 MHC class II molecule. We first demonstrate the potential for direct presentation of insulin peptides by beta cells through retroviral expression of HLA-DQ8 in the EndoC-βH1 beta cell line. We then derive induced pluripotent stem cells from diabetic and healthy control patients harboring the HLA-DQ8 haplotype and use these cells to generate islet-like cell clusters. Upon

cytokine treatment, a fraction of these islet-like cells expresses HLA-DQ8. However, upon co-culture these cells fail to stimulate CD4+ T cell clones recognizing insulin peptides. We additionally demonstrate that co-culture of thapsigargin-treated islet-like cells with CD4+ T cells results in CD4+ T cell stimulation via a T cell receptor-independent mechanism, arguing against the relevance of this experimental approach for modeling immune cell-beta cell interactions. The co-culture system established in this work will help develop future studies to identify a potential role for beta cells in contributing to stimulation of CD4+ T cells via presentation of insulin and other endogenous antigens on HLA-DQ8.

Results

HLA-DQ8-expressing beta cells present endogenous antigens to CD4+ T cells

To determine whether beta cells are capable of properly processing, loading, and presenting antigens on HLA-DQ8, we utilized the immortalized human beta cell line EndoC-βH1[215]. Because these cells do not harbor the HLA-DQ8 haplotype, we used retroviral transduction to induce expression of the molecule. Upon transduction we observed extensive co-localization of intracellular insulin with HLA-DQ8 (**Fig. 4.1a**). When we co-cultured transduced cells with CD4+ T cell clones recognizing either the C-peptide component of proinsulin[216] or residues 13-21 of the insulin B chain (Ins_{13-21})[217], we observed robust stimulation of each clone regardless of whether or not exogenous peptide was added to the co-cultures (**Fig. 4.1b**). This indicates that stimulation takes place through presentation of endogenous rather than exogenous antigens on HLA-DQ8. Altogether, these experiments demonstrate that beta cells expressing HLA-DQ8 are capable of processing, loading, and presenting endogenous antigens such as insulin to CD4+ T cells.

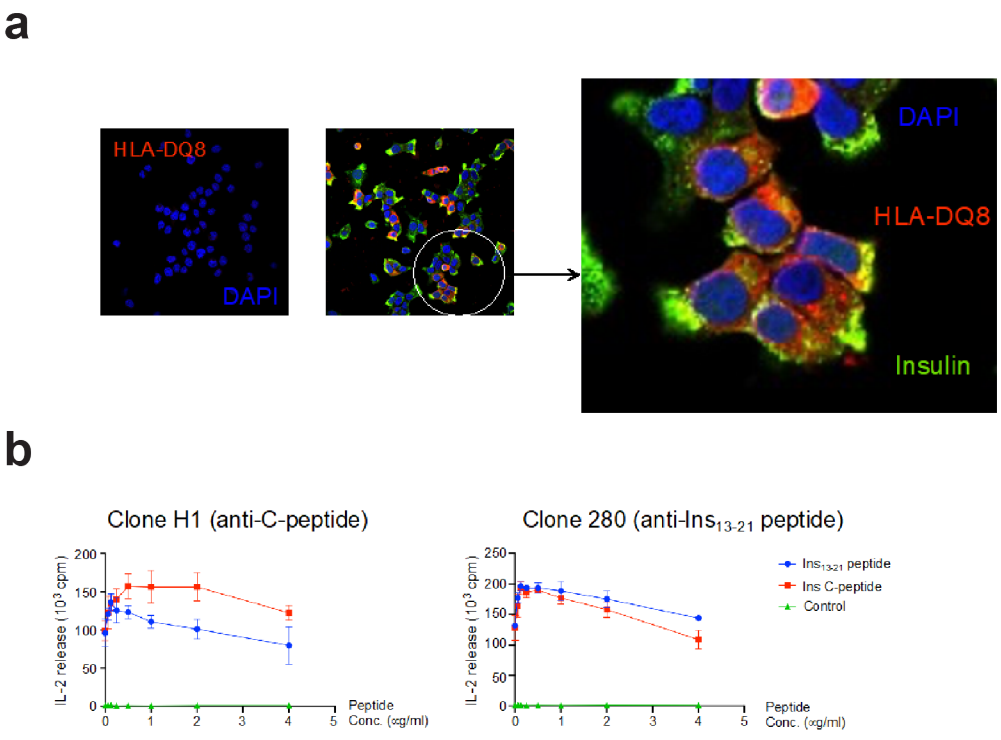

Figure 4.1. HLA-DQ8-expressing β-cells present endogenous antigens to CD4+ T cells. (**a**) EndoC-βH1 cells do not express HLA-DQ8 under normal conditions (left panel) but do after transduction with an HLA-DQ8 retroviral vector (right panels). HLA-DQ8 overlaps with insulin-containing vesicles. (**b**) Endogenous antigen presentation by HLA-DQ8+ EndoC-βH1 cells and not non-transduced control cells. H1 is an anti-C-peptide clone, whereas 280 is an anti-Ins_{13-21} clone. The presence of peptide did not modify the response, indicating endogenous presentation. Activation was measured by IL-2 release in biological triplicates.

HLA-DQ8+ iPSC-islet cells are functional

We next sought to determine whether endogenous expression of HLA-DQ8 in beta cells is sufficient to allow for antigen presentation. To this end, we recruited patients harboring the HLA-DQ8 haplotype. Five such patients were recruited, three of which had established T1D and two of which were healthy controls. A final non-diabetic patient was recruited harboring the HLA-DQ6 haplotype, which has been shown to be protective against T1D occurrence[218]. From each donor, we collected peripheral blood samples and derived clonal induced pluripotent stem cell (iPSC) lines (**Fig. 4.2a**). We then adapted an established protocol[3] to differentiate iPSCs into mature islet-like cells (iPSC-islet cells). For iPSCs derived from both diabetic and non-diabetic donors, we observed similar yields of beta cells (**Fig. 4.2b**) and similar functionality as measured by insulin release in response to external glucose concentrations (**Fig. 4.2c**).

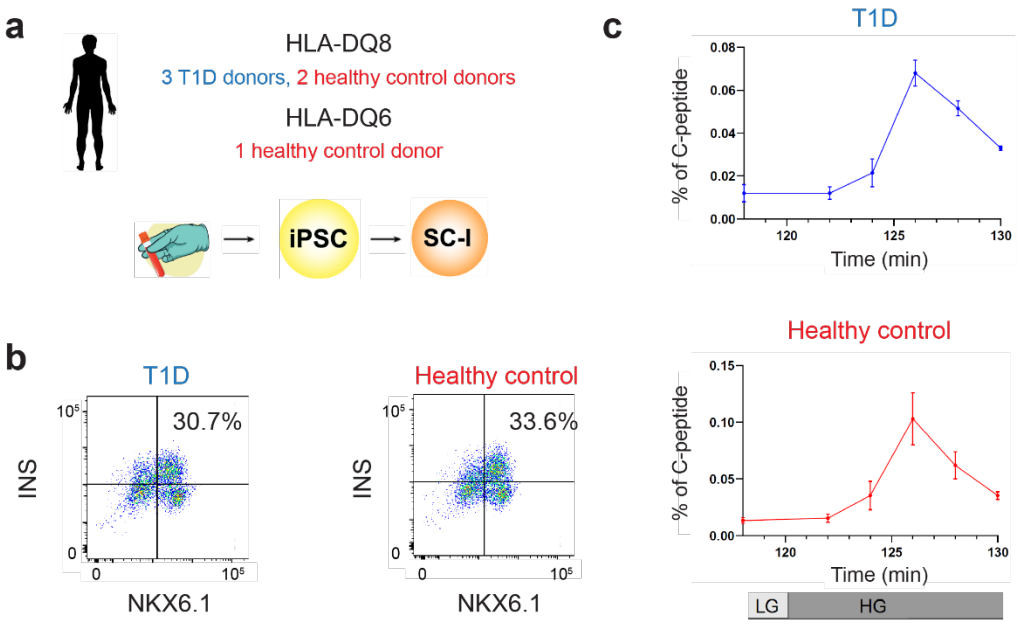

Figure 4.2. Derivation of iPSCs from HLA-DQ8+ donors and differentiation into SC-islets. (**a**) Schematic illustrating haplotypes of donors, collection of peripheral blood, and derivation and differentiation of iPSCs. (**b**) Flow cytometry plots showing percentage of beta-like cells as defined by NKX6.1+/INS+ cells. Plots are shown for iPSCs derived from T1D and healthy control donors. (**c**) Insulin secretion from SC-islet cells derived from T1D and healthy control donors in response to low glucose (LG, 2.8 mM) and high glucose (HG, 16.8 mM) concentrations.

iPSC-islet cells express HLA-DQ8 but do not stimulate CD4+ T cell clones

We next tested environmental conditions that may lead to expression of HLA-DQ8 in iPSC-islet cells. Based on previous observations in primary human islets[209,212,214], we first tested the effects of cytokine treatment. We found that treatment of iPSC-islet cells with 30 ng/mL IFNg and 10ng/mL TNFa for a duration of 96 hours led to robust upregulation of various genes involved in the MHC class II presentation pathway (**Fig. 4.3a**). Flow cytometry revealed that HLA-DQ8 was presented on the cell surface in approximately 5% of iPSC-islet cells in culture (**Fig. 4.3b**). This non-uniform expression is consistent with previous findings of sporadic MHC class II expression in cytokine-treated primary human beta cells[214]. However, additional flow cytometry analysis showed that HLA-DQ8 was expressed mainly within a population of iPSC-islet cells that did not express the beta cell markers insulin and NKX6-1 (**Fig. 4.3c**). While beta cells have previously been demonstrated to downregulate these markers in response to cytokines[219], these data could also indicate that cytokines induce expression of HLA-DQ8 primarily within a non-beta cell population in the iPSC-islet cultures. Consistent with the latter possibility, when cytokine-treated iPSC-islets were co-cultured with CD4+ T cell clones recognizing C-peptide, we observed no stimulation (**Fig. 4.3d**). These data demonstrate that iPSC-islet cells do not present insulin peptides on HLA-DQ8 in response to cytokine treatment.

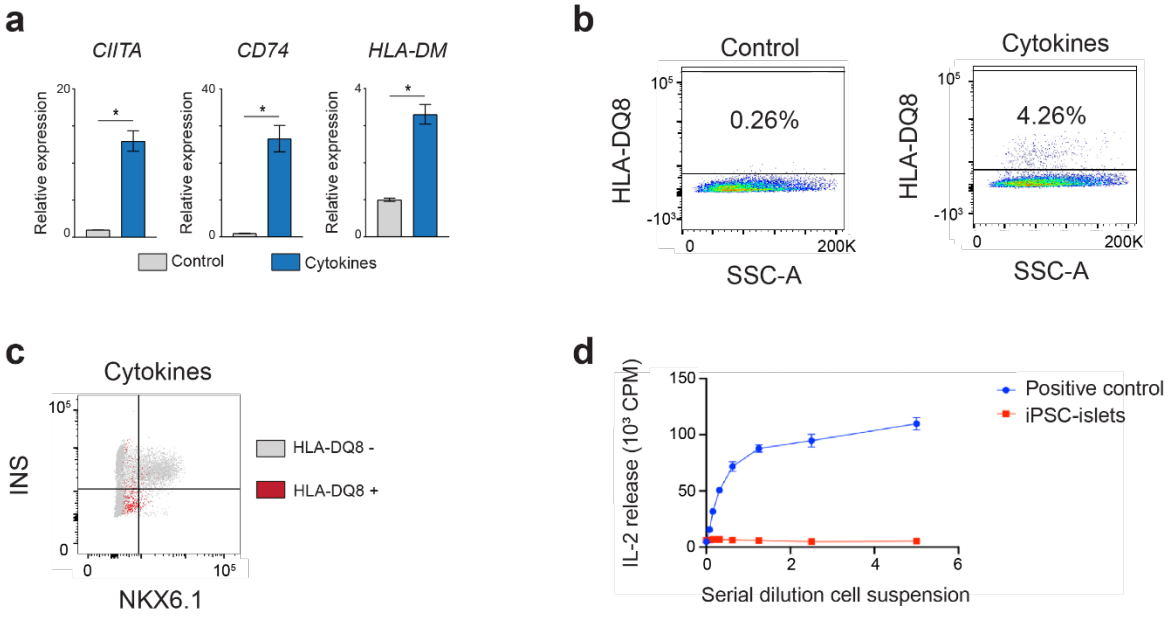

Figure 4.3. Expression of HLA-DQ8 and stimulation of CD4+ T cells upon cytokine treatment. (**a**) qPCR analysis of MHC class II pathway genes *CIITA*, *CD74*, and *HLA-DM* in untreated compared to cytokine-treated iPSC-islet cells (P = 1.31 x 10^{-2}, 1.90 x 10^{-2}, and 1.33 x 10^{-2} for comparisons of *CIITA*, *CD74*, and *HLA-DM*, respectively, in cytokine-treated compared to untreated SC-islet cells; student's t-test, 2-sided). (**b**) Flow cytometry plots showing percentage of untreated and cytokine-treated iPSC-islet cells expressing HLA-DQ8 on the cell surface. (**c**) Flow cytometry plot showing presence of NKX6.1 and INS in cytokine-treated cells expressing HLA-DQ8. (**d**) Stimulation of anti-C-peptide CD4+ T cell clones as measured by IL-2 release in biological triplicates upon co-culture with cytokine-treated iPSC-islets. Treatment with exogenous IL-2 was used as a positive control.

Thapsigargin-treated iPSC-islet cells stimulate CD4+ T cells through off-target mechanisms

Because cytokine treatment does not result in the stimulation of CD4+ T cell clones, we investigated other environmental factors which may lead to presentation of insulin peptides on HLA-DQ8. A previous study of iPSC-islets co-cultured with autologous peripheral blood mononuclear cells found that treatment of iPSC-islets with cytokines was insufficient to induce stimulation of CD8+ T cells via MHC class I[220]. However, treatment of iPSC-islets with the ER stress inducer thapsigargin prior to co-culture led to robust T cell stimulation. Given our analogous findings that cytokine treatment is insufficient for stimulation of CD4+ T cells via MHC class II, we hypothesized that thapsigargin-induced ER stress may be necessary for presentation of insulin peptides on HLA-DQ8. Therefore, as performed in the previous study[220], we pre-treated iPSC-islet cells with 5 μM thapsigargin for 5 hours. We then rinsed cells twice with PBS and co-cultured with CD4+ T cell clones recognizing C-peptide. Upon doing so, we observed robust stimulation of CD4+ T cells (**Fig. 4.4a**). However, blocking HLA-DQ8 through addition of an antibody did not diminish activation of CD4+ T cells, suggesting that activation of CD4+ T cells in this context does not take place through interaction of a T cell receptor with MHC class II-bound peptide. To further examine this possibility, we generated iPSC-islet cells using the H1 hESC line, which does not harbor the HLA-DQ8 haplotype. We then treated these cells with thapsigargin as described and co-cultured with CD4+ T cell hybridomas lacking a functional T cell receptor. Upon co-culture, these T cells were likewise stimulated, confirming that pre-treatment of iPSC-islets with thapsigargin leads to stimulation of CD4+ T cells through mechanisms that are independent of T cell receptor signaling.

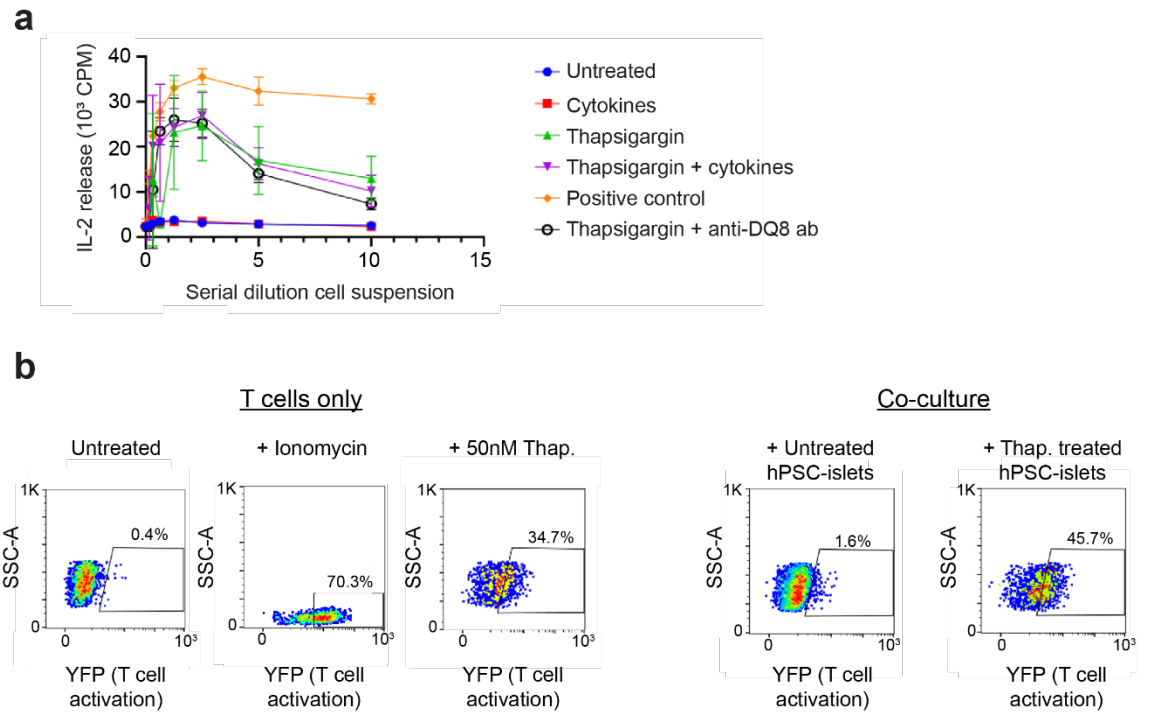

Figure 4.4. Thapsigargin-treated iPSC-islets stimulate CD4+ T cells in a T cell receptor-independent manner. (**a**) Stimulation of anti-C-peptide CD4+ T cell clones as measured by IL-2 release in biological triplicates upon co-culture with cytokine-treated and/or thapsigargin-treated iPSC-islets. Treatment with exogenous IL-2 was used as a positive control. (**b**) Flow cytometry plots showing activation of CD4+ T cell hybridomas lacking a T cell receptor when cultured alone or with SC-islet cells derived from the HLA-DQ8 – H1 hESC line. Activation was measured in biological triplicates.

Discussion

The role of MHC class II expression in pancreatic beta cells and its relevance to T1D remain unclear. It has been hypothesized that presentation of islet autoantigens in the absence of costimulatory factors could be a means of promoting anergy in autoreactive T cells[214]. Such a mechanism would inhibit activation of naïve CD4+ T cells, but may contribute to further stimulation of CD4+ T cells that have already been activated against islet autoantigens. Thus, in the context of T1D, beta cell presentation of autoantigens on MHC class II may "add fuel to the fire" of autoimmunity. Supporting this possibility, we demonstrate that beta cells expressing HLA-DQ8 are capable of processing, loading, and presenting endogenous antigens to CD4+ T cells (**Fig. 4.1a,b**).

Using iPSC-islet cells derived from HLA-DQ8+ donors, we find that cytokine treatment induces presentation of HLA-DQ8 on the surface of a small population of cells (**Fig. 4.3b**). However, it is unclear whether this population constitutes beta cells (**Fig. 4.3c**), and co-culture experiments fail to provide stimulation of CD4+ T cells (**Fig. 4.3d**). Thus, we find discrepancies between immortalized beta cell lines and iPSC-islets in this regard. It is known that hPSC-islets are functionally immature in terms of insulin secretion and metabolic profile[2,4,221], and the inability to present endogenous peptides via HLA-DQ8 may be another aspect in which hPSC-islets fail to recapitulate primary beta cell biology. Future studies are focused on examining this possibility by engrafting iPSC-islets into immunodeficient mice as a means of maturing the cells prior to co-culture experiments. Conversely, it is possible that our findings in immortalized cell lines are an artifact of non-physiological expression of HLA-DQ8; under physiological conditions, beta cells may not present endogenous antigens to CD4+ T cells. *In vivo* maturation of iPSC-islets will be useful in discriminating between these possibilities. Notably, lack of endogenous antigen presentation does not necessarily preclude activation of CD4+ T cells. Given high concentrations of insulin and other autoantigens within the islet, binding of exogenous antigens to HLA-DQ8 expressed on beta cells may contribute to stimulation of CD4+ T cells.

Unexpectedly, we find that pre-treatment of iPSC-islet cells with thapsigargin leads to stimulation of CD4+ T cells through T cell receptor-independent mechanisms (**Fig. 4.4a,b**). Thapsigargin induces ER stress by releasing intracellular Ca^{2+} stores[222], and T cell receptor signal transduction takes place through a similar mechanism[223]. Nanomolar concentrations of thapsigargin have been demonstrated to generate IL-2 release in CD4+ T cells[224]. Although iPSC-islets were washed thoroughly between pre-treatment and co-culture experiments, the high concentration of thapsigargin used may have led to residual compound present within the extracellular matrix formed between iPSC-islet cells. Additionally, thapsigargin may have been taken up into secretory vesicles of iPSC-islet cells and been released along with insulin or other hormones upon co-culture. Regardless of the mechanism, our findings demonstrate that pre-treatment with thapsigargin does not effectively model the effects of ER stress on stimulation of CD4+ T cells via HLA-DQ8. We posit that future experiments seeking to study the effects of ER stress on this phenomenon are better served to use physiological stressors such as lipotoxicity or compounds with fewer possibilities of off-target effects.

Altogether, our experiments suggest the possibility of direct presentation of islet autoantigens to CD4+ T cells via HLA-DQ8. The here-established iPSC-islet co-culture platform provides a foundation for future studies examining the effects of functional maturation on beta cell antigen presentation, as well as studies which may incorporate patient-derived immune cells into a fully autologous co-culture system to model autoimmunity.

Methods

Cell Lines

HEK293T were maintained in DMEM F12 containing 100 units/mL penicillin and 100 mg/mL streptomycin sulfate supplemented with 10% foetal bovine serum (FBS). EndoC-βH1 cells were maintained in low glucose DMEM containing 0.02 g/mL BSA, .0035 uL/mL beta-mercaptoethanol, .0012 g/mL nicotinamide, .0055 mg/mL transferrin, 6.7 µg/mL sodium selenite, 100 units/mL penicillin, and 100 mg/mL streptomycin sulfate. CD4+ T cell hybridomas were cultured in DMEM supplemented with 10% FBS.

Generation of HLA-DQ8 Retrovirus

To generate the HLA-DQ8 expression vector, the coding sequences for α and β chain of HLA-DQ8 were cloned into the pMIG backbone (Addgene, plasmid #9044).

High-titer lentiviral supernatants were generated by co-transfection of the expression vector and the lentiviral packaging construct into HEK293T cells as described[67]. Briefly, the expression vector was co-transfected with the pCMV-R8.74 (Addgene, #22036) and pMD2.G (Addgene, #12259) expression plasmids into HEK293T cells using a 1mg/ml PEI solution (Polysciences). Retroviral supernatants were collected at 48 hr and 72 hr after transfection. Retroviruses were concentrated by ultracentrifugation for 120 min at 19,500 rpm using a Beckman SW28 ultracentrifuge rotor at 4°C.

Transduction of EndoC-βH1 cells

EndoC-βH1 cells were distributed onto a 6 well plate at a density of approximately 1 million cells per well. Concentrated retrovirus and 8 µg/mL polybrene were then added to media in each well. After 30 minutes of incubation, the 6 well plate was spun in a centrifuge (Sorvall Legend RT) for 1 hour at 30°C at 950 G. 6 hours later, viral media was replaced with fresh base culture media.

Immunofluorescence analysis

EndoC-βH1 cells were grown and retrovirally transduced on coverslips and fixed with 4% PFA. Slide-mounted cells were permeabilized and blocked with blocking buffer, consisting of 0.15% (v/v) Triton X-100 (Sigma) and 1% (v/v) normal donkey serum (Jackson Immuno Research Laboratories) in PBS, for 1 hour at room temperature. Slides were then incubated overnight at 4°C with primary antibody solutions. The following day slides were washed five times with PBS and incubated for 1 hour at room temperature with secondary antibody solutions. Cells were washed five times with PBS before coverslips were applied.

All antibodies were diluted in blocking buffer at the ratios indicated below. Primary antibodies used were guinea pig anti-insulin (INS) (1:500 dilution, Dako) and mouse anti-HLA-DQ8 (1:200, cs.200.1). Secondary antibodies were Alexa488- and Cy3-conjugated donkey antibodies (Jackson Immuno Research Laboratories), and were used at dilutions of 1:500 (anti-guinea pig Alexa488) or 1:1000 (anti-mouse Cy3). Cell nuclei were stained with Hoechst 33342 (1:3000, Invitrogen). Representative images were obtained with a Zeiss Axio-Observer-Z1 microscope equipped with a Zeiss ApoTome and AxioCam digital camera. Figures were prepared in Adobe Creative Suite 5.

Flow cytometry analysis

Cell aggregates derived from hPSCs were allowed to settle in microcentrifuge tubes and washed with PBS. Cell aggregates or EndoC-βH1 cells were incubated with Accutase® at 37°C until a single-cell suspension was obtained. Cells were washed with 1 mL ice-cold flow buffer comprised of 0.2% BSA in PBS and centrifuged at 200 g for 5 min. BD Cytofix/Cytoperm™ Plus Fixation/Permeabilization Solution Kit was used to fix and stain cells for flow cytometry according to the manufacturer's instructions. Briefly, cell pellets were re-suspended in ice-cold BD Fixation/Permeabilization solution (300 µL per microcentrifuge tube). Cells were incubated for 20 min at 4°C. Cells were washed twice with 1 mL ice-cold 1X BD Perm/Wash™ Buffer and

centrifuged at 10°C and 200 x g for 5 min. Cells were re-suspended in 50 µL ice-cold 1X BD Perm/Wash™ Buffer containing diluted antibodies, for each staining performed. Cells were incubated at 4°C in the dark for 1-3 hours. Cells were washed with 1.25 mL ice-cold 1X BD Wash Buffer and centrifuged at 200 g for 5 min. Cell pellets were re-suspended in 300 µL ice-cold flow buffer and analysed in a FACSCanto™ II (BD Biosciences). Antibodies used were mouse anti-HLA-DQ8 (1:200, cs.200.1), PE-conjugated anti-Insulin (1:50 dilution, Cell Signaling), AlexaFluor® 647-conjugated anti-NKX6.1 (1:5 dilution, BD Biosciences), and anti-mouse Alexa488 (1:50, Jackson Immuno Research Laboratories). Data were processed using FlowJo software v10.

RNA Isolation and Sequencing (RNA-seq) and qRT-PCR

RNA was isolated from cell samples using the RNeasy® Micro Kit (Qiagen) according to the manufacturer instructions. For each cell stage and condition analyzed between 0.1 and 1 x 10^6 cells were collected for RNA extraction. For qRT-PCR, cDNA synthesis was first performed using the iScript™ cDNA Synthesis Kit (Bio-Rad) and 500 ng of isolated RNA per reaction. qRT-PCR reactions were performed in triplicate with 10 ng of template cDNA per reaction using a CFX96™ Real-Time PCR Detection System and the iQ™ SYBR® Green Supermix (Bio-Rad). PCR of the TATA binding protein (TBP) coding sequence was used as an internal control and relative expression was quantified via double delta C_T analysis. A complete list of RT-qPCR primer sequences can be found in **Supplementary Figure 4.1.**

CD4+ T cell co-culture assays

For co-culture assays involving anti-$Ins_{13\text{-}21}$ and anti-C-peptide T cell clones, serial dilutions of suspended iPSC-islet aggregates or EndoC-βH1 cells were co-cultured with T cell hybridomas. After 24 hours of incubation at 37°C, supernatants were harvested and assayed for IL-2 content using a radioactive IL-2–dependent natural killer cell line bioassay. For assays

involving addition of exogenous peptides, peptides were added to media in increasing concentration. All assays were performed in biological triplicates. For each T cell hybridoma, receptors were derived from autoreactive CD4+ T cells found in type 1 diabetic patients[216,217] and cloned onto a hybridoma expressing the human CD4 molecule with substitutions at positions 40 and 45 (Gln40Tyr and Thr45Trp)[216].

For assays involving activation of hybridomas lacking a T cell receptor, a reporter line was used in which T cell activation triggers expression of an NFAT response element linked to mAmetrine. Activation was then measured via flow cytometry.

Derivation of iPSC lines

50 mL blood draws were taken from consented patients, and peripheral blood mononuclear cells were isolated via flow cytometry assisted cell sorting. Clonal iPSC lines were then derived at the Salk Institute Stem Cell Core. Each clonal line was karyotyped, donor-matched to patient cells via STR analysis, and tested for pluripotency prior to differentiation.

Maintenance and pancreatic differentiation of iPSCs

iPSCs were cultured in mTeSR1 media (Stem Cell Technologies) and propagated by passaging cells every 3 to 4 days using Accutase (eBioscience) for enzymatic cell dissociation.

For differentiation, cells were dissociated using Accutase for 10 min, then reaggregated by plating the cells at a concentration of ~5 x 10^6 cells/well in a 6-well plate coated with Matrigel (Corning). The following day, undifferentiated cells were washed with PBS and then differentiated using a multi-step protocol with stage-specific media and daily media changes. At day 29, cells were dissociated and resuspended in stage 7 medium at a concentration of ~3 x 10^6 cells per well of a 6-well ultra-low attachment plates (Costar). The cells were cultured overnight on an orbital rotator (Innova2000, New Brunswick Scientific) at 100 rpm, then maintained at 118 rpm until day 40. At day 40, cells were collected for analysis or used for further experiments.

All stage-specific base media were comprised of MCDB 131 medium (Thermo Fisher Scientific) supplemented with NaHCO3, GlutaMAX, D-Glucose, and BSA using the following concentrations:

Stage 1/2 base medium: MCDB 131 medium, 1.5 g/L NaHCO3, 1X GlutaMAX, 10 mM D-Glucose, 0.5% BSA

Stage 3/4 base medium: MCDB 131 medium, 2.5 g/L NaHCO3, 1X GlutaMAX, 10 mM D-glucose, 2% BSA

Stage 5/6 medium: MCDB 131 medium, 1.5 g/L NaHCO3, 1X GlutaMAX, 15 mM D-glucose, 2% BSA

Stage 7 medium: MCDB 131 medium, 1.5 g/L NaHCO3, 2% BSA

Media compositions for each stage were as follows:

Stage 1 (days 0 - 3): base medium, 100 ng/ml Activin A, 3 μM CHIR (day 0). Day 1-2: base medium, 100 ng/ml Activin A

Stage 2 (days 4 - 6): base medium, 0.25 mM L-Ascorbic Acid (Vitamin C), 50 ng/mL KGF

Stage 3 (days 7 - 8): base medium, 0.25 mM L-Ascorbic Acid, 50 ng/mL KGF, 0.25 μM SANT-1, 1 μM Retinoic Acid, 100 nM LDN193189, 1:200 ITS-X, 200 nM TPB

Stage 4 (days 9 - 11): base medium, 0.25 mM L-Ascorbic Acid, 2 ng/mL KGF, 0.25 μM SANT-1, 0.1 μM Retinoic Acid, 200 nM LDN193189, 1:200 ITS-X, 100nM TPB

Stage 5 (days 12 - 14): base medium, 0.25 μM SANT-1, 0.05 μM RA, 100 nM LDN-193189, 1 μM T3, 10 μM ALK5i II, 10 μM ZnSO4, 10 μg/mL heparin, 1:200 ITS-X

Stage 6 (days 15 - 21): base medium, 0.05 μM RA, 100 nM LDN-193189, 1 μM T3, 10 μM ALK5i II, 10 μM ZnSO4, 10 μg/mL heparin, 100nM gamma-secretase inhibitor, 1:200 ITS-X

Stage 7 (days 22 - 40): base medium, 10 μg/mL heparin, 10 μM ZnSO4, 1X trace element A, 1X trace element B

Measurement of glucose-stimulated insulin secretion (GSIS)

Perifusion was carried out using the Biorep perifusion system. hPSC-islet aggregates were first starved for 30 min in Krebs-RingersBicarbonate-HEPES (KRBH) buffer (130 mM NaCl, 5 mM KCl, 1.2 mM CaCl2, 1.2 mM MgCl2, 1.2 mM KH2PO4, 20 mM HEPES pH 7.4, 25 mM NaHCO3, and 0.1% bovine serum albumin) containing 2.8 mM glucose at 37°C with 5% CO2 and were then loaded into perifusion chambers (70 cell aggregates/chamber).

hPSC-islet aggregates were perifused with KRBH containing 2.8 mM glucose for an additional 2 hours, at which point aggregates were stimulated with KRBH containing 16.8 mM glucose. At each timepoint, perifusate was collected for analysis. Perifusion was carried out at 37°C with a flow rate of 80 µL/min. At the end of each experiment, hPSC-islet aggregates were transferred to Eppendorf tubes and lysed by sonication for insulin content determination. Insulin content was measured using the ALPCO C-peptide ELISA kit according to manufacturer's instructions.

Cytokine treatment of iPSC-islet cells

10 ng/mL TNFa (Peprotech) and 30ng/mL IFNg (Peprotech) were added daily to fresh Stage 7 differentiation media for 4 days. Cells were then rinsed with PBS prior to co-culture experiments or downstream analysis.

Thapsigargin treatment of iPSC-islet cells

5 µM thapsigargin (Sigma) was added to Stage 7 differentiation media 5 hours. Cells were then rinsed twice with PBS prior to coculture experiments and downstream analysis.

Quantification and statistical analyses

Statistical analyses were performed using GraphPad Prism (v8.1.2). Statistical parameters, such as the value of n, mean, standard deviation (SD), standard error of the mean

(SEM), significance level (*p < 0.05, **p < 0.01, and ***p < 0.001), and the statistical tests used, are reported in the figures and figure legends. The ‘‘n’’ refers to the number of biological replicates.

Declaration of interests

The authors declare no competing interests.

Acknowledgements

This work was supported by the NIH grants 5UG3DK122639-02 (M.S and L.T.) and T32 GM008666 (R.J.G.).

This material was co-authored with Clarke, D., Zhou, H., Sharma, S., Teyton, L., and Sander, M. The dissertation author was the principal author of this chapter.

REFERENCES

1 Cohen, D. E. & Melton, D. Turning straw into gold: directing cell fate for regenerative medicine. *Nat Rev Genet* **12**, 243-252, doi:10.1038/nrg2938 (2011).

2 Davis, J. C., Alves, T. C., Helman, A., Chen, J. C., Kenty, J. H., Cardone, R. L., Liu, D. R., Kibbey, R. G. & Melton, D. A. Glucose Response by Stem Cell-Derived beta Cells In Vitro Is Inhibited by a Bottleneck in Glycolysis. *Cell Rep* **31**, 107623, doi:10.1016/j.celrep.2020.107623 (2020).

3 Hogrebe, N. J., Augsornworawat, P., Maxwell, K. G., Velazco-Cruz, L. & Millman, J. R. Targeting the cytoskeleton to direct pancreatic differentiation of human pluripotent stem cells. *Nat Biotechnol* **38**, 460-470, doi:10.1038/s41587-020-0430-6 (2020).

4 Augsornworawat, P., Maxwell, K. G., Velazco-Cruz, L. & Millman, J. R. Single-Cell Transcriptome Profiling Reveals beta Cell Maturation in Stem Cell-Derived Islets after Transplantation. *Cell Rep* **32**, 108067, doi:10.1016/j.celrep.2020.108067 (2020).

5 Long, H. K., Prescott, S. L. & Wysocka, J. Ever-Changing Landscapes: Transcriptional Enhancers in Development and Evolution. *Cell* **167**, 1170-1187, doi:10.1016/j.cell.2016.09.018 (2016).

6 Rada-Iglesias, A., Bajpai, R., Swigut, T., Brugmann, S. A., Flynn, R. A. & Wysocka, J. A unique chromatin signature uncovers early developmental enhancers in humans. *Nature* **470**, 279-283, doi:10.1038/nature09692 (2011).

7 Jaenisch, R. & Bird, A. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nat Genet* **33 Suppl**, 245-254, doi:10.1038/ng1089 (2003).

8 Buenrostro, J. D., Wu, B., Chang, H. Y. & Greenleaf, W. J. ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. *Curr Protoc Mol Biol* **109**, 21 29 21-21 29 29, doi:10.1002/0471142727.mb2129s109 (2015).

9 Orlando, V. Mapping chromosomal proteins in vivo by formaldehyde-crosslinked-chromatin immunoprecipitation. *Trends Biochem Sci* **25**, 99-104, doi:10.1016/s0968-0004(99)01535-2 (2000).

10 Belton, J. M., McCord, R. P., Gibcus, J. H., Naumova, N., Zhan, Y. & Dekker, J. Hi-C: a comprehensive technique to capture the conformation of genomes. *Methods* **58**, 268-276, doi:10.1016/j.ymeth.2012.05.001 (2012).

11 Tang, F., Barbacioru, C., Wang, Y., Nordman, E., Lee, C., Xu, N., Wang, X., Bodeau, J., Tuch, B. B., Siddiqui, A., Lao, K. & Surani, M. A. mRNA-Seq whole-transcriptome analysis of a single cell. *Nat Methods* **6**, 377-382, doi:10.1038/nmeth.1315 (2009).

12 Mezger, A., Klemm, S., Mann, I., Brower, K., Mir, A., Bostick, M., Farmer, A., Fordyce, P., Linnarsson, S. & Greenleaf, W. High-throughput chromatin accessibility profiling at single-cell resolution. *Nat Commun* **9**, 3647, doi:10.1038/s41467-018-05887-x (2018).

13 Hainer, S. J., Boskovic, A., McCannell, K. N., Rando, O. J. & Fazzio, T. G. Profiling of Pluripotency Factors in Single Cells and Early Embryos. *Cell* **177**, 1319-1329 e1311, doi:10.1016/j.cell.2019.03.014 (2019).

14 Nagano, T., Lubling, Y., Stevens, T. J., Schoenfelder, S., Yaffe, E., Dean, W., Laue, E. D., Tanay, A. & Fraser, P. Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. *Nature* **502**, 59-64, doi:10.1038/nature12593 (2013).

15 Organization, W. H. Global report on diabetes. (2016).

16 Halban, P. A., Polonsky, K. S., Bowden, D. W., Hawkins, M. A., Ling, C., Mather, K. J., Powers, A. C., Rhodes, C. J., Sussel, L. & Weir, G. C. beta-cell failure in type 2 diabetes: postulated mechanisms and prospects for prevention and treatment. *J Clin Endocrinol Metab* **99**, 1983-1992, doi:10.1210/jc.2014-1425 (2014).

17 Atkinson, M. A., Eisenbarth, G. S. & Michels, A. W. Type 1 diabetes. *Lancet* **383**, 69-82, doi:10.1016/S0140-6736(13)60591-7 (2014).

18 Ahmad, F. B. & Anderson, R. N. The Leading Causes of Death in the US for 2020. *JAMA* **325**, 1829-1830, doi:10.1001/jama.2021.5469 (2021).

19 Zaret, K. S. & Carroll, J. S. Pioneer transcription factors: establishing competence for gene expression. *Genes Dev* **25**, 2227-2241, doi:10.1101/gad.176826.111 (2011).

20 Mahajan, A., Taliun, D., Thurner, M., Robertson, N. R., Torres, J. M., Rayner, N. W., Payne, A. J., Steinthorsdottir, V., Scott, R. A., Grarup, N., Cook, J. P., Schmidt, E. M., Wuttke, M., Sarnowski, C., Magi, R., Nano, J., Gieger, C., Trompet, S., Lecoeur, C., Preuss, M. H., Prins, B. P., Guo, X., Bielak, L. F., Below, J. E., Bowden, D. W., Chambers, J. C., Kim, Y. J., Ng, M. C. Y., Petty, L. E., Sim, X., Zhang, W., Bennett, A. J., Bork-Jensen, J., Brummett, C. M., Canouil, M., Ec Kardt, K. U., Fischer, K., Kardia, S. L. R., Kronenberg, F., Lall, K., Liu, C. T., Locke, A. E., Luan, J., Ntalla, I., Nylander, V., Schonherr, S., Schurmann, C., Yengo, L., Bottinger, E. P., Brandslund, I., Christensen, C., Dedoussis, G., Florez, J. C., Ford, I., Franco, O. H., Frayling, T. M., Giedraitis, V., Hackinger, S., Hattersley, A. T., Herder, C., Ikram, M. A., Ingelsson, M., Jorgensen, M. E., Jorgensen, T., Kriebel, J., Kuusisto, J., Ligthart, S., Lindgren, C. M., Linneberg, A., Lyssenko, V., Mamakou, V., Meitinger, T., Mohlke, K. L., Morris, A. D., Nadkarni, G., Pankow, J. S., Peters, A., Sattar, N., Stancakova, A., Strauch, K., Taylor, K. D., Thorand, B., Thorleifsson, G., Thorsteinsdottir, U., Tuomilehto, J., Witte, D. R., Dupuis, J., Peyser, P. A., Zeggini, E., Loos, R. J. F., Froguel, P., Ingelsson, E., Lind, L., Groop, L., Laakso, M., Collins, F. S., Jukema, J. W., Palmer, C. N. A., Grallert, H., Metspalu, A., Dehghan, A., Kottgen, A., Abecasis, G. R., Meigs, J. B., Rotter, J. I., Marchini, J., Pedersen, O., Hansen, T., Langenberg, C., Wareham, N. J., Stefansson, K., Gloyn, A. L., Morris, A. P., Boehnke, M. & McCarthy, M. I. Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. *Nat Genet* **50**, 1505-1513, doi:10.1038/s41588-018-0241-6 (2018).

21 Chiou, J., Geusz, R. J., Okino, M. L., Han, J. Y., Miller, M., Melton, R., Beebe, E., Benaglio, P., Huang, S., Korgaonkar, K., Heller, S., Kleger, A., Preissl, S., Gorkin, D. U., Sander, M. & Gaulton, K. J. Interpreting type 1 diabetes risk with genetics and single-cell epigenomics. *Nature* **594**, 398-402, doi:10.1038/s41586-021-03552-w (2021).

22 Noble, J. A. & Valdes, A. M. Genetics of the HLA region in the prediction of type 1 diabetes. *Curr Diab Rep* **11**, 533-542, doi:10.1007/s11892-011-0223-x (2011).

23 Corper, A. L., Stratmann, T., Apostolopoulos, V., Scott, C. A., Garcia, K. C., Kang, A. S., Wilson, I. A. & Teyton, L. A Structural Framework for Deciphering the Link Between I-A^{g7} and Autoimmune Diabetes. *Science* **288**, 505-511, doi:10.1126/science.288.5465.505 (2000).

24 Yoshida, K., Corper, A. L., Herro, R., Jabri, B., Wilson, I. A. & Teyton, L. The diabetogenic mouse MHC class II molecule I-Ag7 is endowed with a switch that modulates TCR affinity. *J Clin Invest* **120**, 1578-1590, doi:10.1172/JCI41502 (2010).

25 Gioia, L., Holt, M., Costanzo, A., Sharma, S., Abe, B., Kain, L., Nakayama, M., Wan, X., Su, A., Mathews, C., Chen, Y. G., Unanue, E. & Teyton, L. Position beta57 of I-A(g7) controls early anti-insulin responses in NOD mice, linking an MHC susceptibility allele to type 1 diabetes onset. *Sci Immunol* **4**, doi:10.1126/sciimmunol.aaw6329 (2019).

26 Levisetti, M. G., Suri, A., Petzold, S. J. & Unanue, E. R. The insulin-specific T cells of nonobese diabetic mice recognize a weak MHC-binding segment in more than one form. *J Immunol* **178**, 6051-6057, doi:DOI 10.4049/jimmunol.178.10.6051 (2007).

27 Wang, A., Yue, F., Li, Y., Xie, R., Harper, T., Patel, N. A., Muth, K., Palmer, J., Qiu, Y., Wang, J., Lam, D. K., Raum, J. C., Stoffers, D. A., Ren, B. & Sander, M. Epigenetic priming of enhancers predicts developmental competence of hESC-derived endodermal lineage intermediates. *Cell Stem Cell* **16**, 386-399, doi:10.1016/j.stem.2015.02.013 (2015).

28 Zorn, A. M. & Wells, J. M. Vertebrate endoderm development and organ formation. *Annu Rev Cell Dev Biol* **25**, 221-251, doi:10.1146/annurev.cellbio.042308.113344 (2009).

29 Lee, K., Cho, H., Rickert, R. W., Li, Q. V., Pulecio, J., Leslie, C. S. & Huangfu, D. FOXA2 Is Required for Enhancer Priming during Pancreatic Differentiation. *Cell Rep* **28**, 382-393 e387, doi:10.1016/j.celrep.2019.06.034 (2019).

30 Genga, R. M. J., Kernfeld, E. M., Parsi, K. M., Parsons, T. J., Ziller, M. J. & Maehr, R. Single-Cell RNA-Sequencing-Based CRISPRi Screening Resolves Molecular Drivers of Early Human Endoderm Development. *Cell Rep* **27**, 708-718 e710, doi:10.1016/j.celrep.2019.03.076 (2019).

31 Wan, H., Dingle, S., Xu, Y., Besnard, V., Kaestner, K. H., Ang, S. L., Wert, S., Stahlman, M. T. & Whitsett, J. A. Compensatory roles of Foxa1 and Foxa2 during lung morphogenesis. *J Biol Chem* **280**, 13809-13816, doi:10.1074/jbc.M414122200 (2005).

32 Lee, C. S., Friedman, J. R., Fulmer, J. T. & Kaestner, K. H. The initiation of liver development is dependent on Foxa transcription factors. *Nature* **435**, 944-947, doi:10.1038/nature03649 (2005).

33 Iwafuchi-Doi, M., Donahue, G., Kakumanu, A., Watts, J. A., Mahony, S., Pugh, B. F., Lee, D., Kaestner, K. H. & Zaret, K. S. The Pioneer Transcription Factor FoxA Maintains an Accessible Nucleosome Configuration at Enhancers for Tissue-Specific Gene Activation. *Mol Cell* **62**, 79-91, doi:10.1016/j.molcel.2016.03.001 (2016).

34 Soufi, A., Garcia, M. F., Jaroszewicz, A., Osman, N., Pellegrini, M. & Zaret, K. S. Pioneer transcription factors target partial DNA motifs on nucleosomes to initiate reprogramming. *Cell* **161**, 555-568, doi:10.1016/j.cell.2015.03.017 (2015).

35 Cirillo, L. A., Lin, F. R., Cuesta, I., Friedman, D., Jarnik, M. & Zaret, K. S. Opening of compacted chromatin by early developmental transcription factors HNF3 (FoxA) and GATA-4. *Mol Cell* **9**, 279-289, doi:10.1016/s1097-2765(02)00459-8 (2002).

36 Carroll, J. S., Liu, X. S., Brodsky, A. S., Li, W., Meyer, C. A., Szary, A. J., Eeckhoute, J., Shao, W., Hestermann, E. V., Geistlinger, T. R., Fox, E. A., Silver, P. A. & Brown, M. Chromosome-wide mapping of estrogen receptor binding reveals long-range regulation requiring the forkhead protein FoxA1. *Cell* **122**, 33-43, doi:10.1016/j.cell.2005.05.008 (2005).

37 Gualdi, R., Bossard, P., Zheng, M., Hamada, Y., Coleman, J. R. & Zaret, K. S. Hepatic specification of the gut endoderm in vitro: cell signaling and transcriptional control. *Genes Dev* **10**, 1670-1682, doi:10.1101/gad.10.13.1670 (1996).

38 Hurtado, A., Holmes, K. A., Ross-Innes, C. S., Schmidt, D. & Carroll, J. S. FOXA1 is a key determinant of estrogen receptor function and endocrine response. *Nat Genet* **43**, 27-33, doi:10.1038/ng.730 (2011).

39 Magnani, L. & Lupien, M. Chromatin and epigenetic determinants of estrogen receptor alpha (ESR1) signaling. *Mol Cell Endocrinol* **382**, 633-641, doi:10.1016/j.mce.2013.04.026 (2014).

40 Wang, Q., Li, W., Zhang, Y., Yuan, X., Xu, K., Yu, J., Chen, Z., Beroukhim, R., Wang, H., Lupien, M., Wu, T., Regan, M. M., Meyer, C. A., Carroll, J. S., Manrai, A. K., Janne, O. A., Balk, S. P., Mehra, R., Han, B., Chinnaiyan, A. M., Rubin, M. A., True, L., Fiorentino, M., Fiore, C., Loda, M., Kantoff, P. W., Liu, X. S. & Brown, M. Androgen receptor regulates a distinct transcription program in androgen-independent prostate cancer. *Cell* **138**, 245-256, doi:10.1016/j.cell.2009.04.056 (2009).

41 Lupien, M., Eeckhoute, J., Meyer, C. A., Wang, Q., Zhang, Y., Li, W., Carroll, J. S., Liu, X. S. & Brown, M. FoxA1 translates epigenetic signatures into enhancer-driven lineage-specific transcription. *Cell* **132**, 958-970, doi:10.1016/j.cell.2008.01.018 (2008).

42 Caizzi, L., Ferrero, G., Cutrupi, S., Cordero, F., Ballare, C., Miano, V., Reineri, S., Ricci, L., Friard, O., Testori, A., Cora, D., Caselle, M., Di Croce, L. & De Bortoli, M. Genome-wide activity of unliganded estrogen receptor-alpha in breast cancer cells. *Proc Natl Acad Sci U S A* **111**, 4892-4897, doi:10.1073/pnas.1315445111 (2014).

43 Donaghey, J., Thakurela, S., Charlton, J., Chen, J. S., Smith, Z. D., Gu, H., Pop, R., Clement, K., Stamenova, E. K., Karnik, R., Kelley, D. R., Gifford, C. A., Cacchiarelli, D., Rinn, J. L., Gnirke, A., Ziller, M. J. & Meissner, A. Genetic determinants and epigenetic effects of pioneer-factor occupancy. *Nat Genet* **50**, 250-258, doi:10.1038/s41588-017-0034-3 (2018).

44 Swinstead, E. E., Miranda, T. B., Paakinaho, V., Baek, S., Goldstein, I., Hawkins, M., Karpova, T. S., Ball, D., Mazza, D., Lavis, L. D., Grimm, J. B., Morisaki, T., Grontved, L., Presman, D. M. & Hager, G. L. Steroid Receptors Reprogram FoxA1 Occupancy

through Dynamic Chromatin Transitions. *Cell* **165**, 593-605, doi:10.1016/j.cell.2016.02.067 (2016).

45 Zhang, Y., Fang, B., Emmett, M. J., Damle, M., Sun, Z., Feng, D., Armour, S. M., Remsberg, J. R., Jager, J., Soccio, R. E., Steger, D. J. & Lazar, M. A. GENE REGULATION. Discrete functions of nuclear receptor Rev-erbalpha couple metabolism to the clock. *Science* **348**, 1488-1492, doi:10.1126/science.aab3021 (2015).

46 Paakinaho, V., Swinstead, E. E., Presman, D. M., Grontved, L. & Hager, G. L. Meta-analysis of Chromatin Programming by Steroid Receptors. *Cell Rep* **28**, 3523-3534 e3522, doi:10.1016/j.celrep.2019.08.039 (2019).

47 Whyte, W. A., Orlando, D. A., Hnisz, D., Abraham, B. J., Lin, C. Y., Kagey, M. H., Rahl, P. B., Lee, T. I. & Young, R. A. Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* **153**, 307-319, doi:10.1016/j.cell.2013.03.035 (2013).

48 Kobberup, S., Nyeng, P., Juhl, K., Hutton, J. & Jensen, J. ETS-family genes in pancreatic development. *Dev Dyn* **236**, 3100-3110, doi:10.1002/dvdy.21292 (2007).

49 Mamidi, A., Prawiro, C., Seymour, P. A., de Lichtenberg, K. H., Jackson, A., Serup, P. & Semb, H. Mechanosignalling via integrins directs fate decisions of pancreatic progenitors. *Nature* **564**, 114-118, doi:10.1038/s41586-018-0762-2 (2018).

50 Huang, W., Wang, G., Delaspre, F., Vitery Mdel, C., Beer, R. L. & Parsons, M. J. Retinoic acid plays an evolutionarily conserved and biphasic role in pancreas development. *Dev Biol* **394**, 83-93, doi:10.1016/j.ydbio.2014.07.021 (2014).

51 Cebola, I., Rodriguez-Segui, S. A., Cho, C. H., Bessa, J., Rovira, M., Luengo, M., Chhatriwala, M., Berry, A., Ponsa-Cobas, J., Maestro, M. A., Jennings, R. E., Pasquali, L., Moran, I., Castro, N., Hanley, N. A., Gomez-Skarmeta, J. L., Vallier, L. & Ferrer, J. TEAD and YAP regulate the enhancer network of human embryonic pancreatic progenitors. *Nat Cell Biol* **17**, 615-626, doi:10.1038/ncb3160 (2015).

52 Farley, E. K., Olson, K. M., Zhang, W., Brandt, A. J., Rokhsar, D. S. & Levine, M. S. Suboptimization of developmental enhancers. *Science* **350**, 325-328, doi:10.1126/science.aac6948 (2015).

53 Fornes, O., Castro-Mondragon, J. A., Khan, A., van der Lee, R., Zhang, X., Richmond, P. A., Modi, B. P., Correard, S., Gheorghe, M., Baranasic, D., Santana-Garcia, W., Tan, G., Cheneby, J., Ballester, B., Parcy, F., Sandelin, A., Lenhard, B., Wasserman, W. W. & Mathelier, A. JASPAR 2020: update of the open-access database of transcription factor binding profiles. *Nucleic Acids Res* **48**, D87-D92, doi:10.1093/nar/gkz1001 (2020).

54 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, doi:10.1016/j.molcel.2014.08.016 (2014).

55 Veres, A., Faust, A. L., Bushnell, H. L., Engquist, E. N., Kenty, J. H., Harb, G., Poh, Y. C., Sintov, E., Gurtler, M., Pagliuca, F. W., Peterson, Q. P. & Melton, D. A. Charting

cellular identity during human in vitro beta-cell differentiation. *Nature* **569**, 368-373, doi:10.1038/s41586-019-1168-5 (2019).

56 Schaffer, A. E., Taylor, B. L., Benthuysen, J. R., Liu, J., Thorel, F., Yuan, W., Jiao, Y., Kaestner, K. H., Herrera, P. L., Magnuson, M. A., May, C. L. & Sander, M. Nkx6.1 controls a gene regulatory network required for establishing and maintaining pancreatic Beta cell identity. *PLoS Genet* **9**, e1003274, doi:10.1371/journal.pgen.1003274 (2013).

57 Jacob, A., Morley, M., Hawkins, F., McCauley, K. B., Jean, J. C., Heins, H., Na, C. L., Weaver, T. E., Vedaie, M., Hurley, K., Hinds, A., Russo, S. J., Kook, S., Zacharias, W., Ochs, M., Traber, K., Quinton, L. J., Crane, A., Davis, B. R., White, F. V., Wambach, J., Whitsett, J. A., Cole, F. S., Morrisey, E. E., Guttentag, S. H., Beers, M. F. & Kotton, D. N. Differentiation of Human Pluripotent Stem Cells into Functional Lung Alveolar Epithelial Cells. *Cell Stem Cell* **21**, 472-488 e410, doi:10.1016/j.stem.2017.08.014 (2017).

58 Corso-Diaz, X., de Leeuw, C. N., Alonso, V., Melchers, D., Wong, B. K., Houtman, R. & Simpson, E. M. Co-activator candidate interactions for orphan nuclear receptor NR2E1. *BMC Genomics* **17**, 832, doi:10.1186/s12864-016-3173-5 (2016).

59 Kheolamai, P. & Dickson, A. J. Liver-enriched transcription factors are critical for the expression of hepatocyte marker genes in mES-derived hepatocyte-lineage cells. *BMC Mol Biol* **10**, 35, doi:10.1186/1471-2199-10-35 (2009).

60 Papaioannou, V. E. The T-box gene family: emerging roles in development, stem cells and cancer. *Development* **141**, 3819-3833, doi:10.1242/dev.104471 (2014).

61 Ramana, C. V. Insights into the Signal Transduction Pathways of Mouse Lung Type II Cells Revealed by Transcription Factor Profiling in the Transcriptome. *Genomics Inform* **17**, e8, doi:10.5808/GI.2019.17.1.e8 (2019).

62 Ikonomou, L., Herriges, M. J., Lewandowski, S. L., Marsland, R., 3rd, Villacorta-Martin, C., Caballero, I. S., Frank, D. B., Sanghrajka, R. M., Dame, K., Kandula, M. M., Hicks-Berthet, J., Lawton, M. L., Christodoulou, C., Fabian, A. J., Kolaczyk, E., Varelas, X., Morrisey, E. E., Shannon, J. M., Mehta, P. & Kotton, D. N. The in vivo genetic program of murine primordial lung epithelial progenitors. *Nat Commun* **11**, 635, doi:10.1038/s41467-020-14348-3 (2020).

63 Herriges, M. & Morrisey, E. E. Lung development: orchestrating the generation and regeneration of a complex organ. *Development* **141**, 502-513, doi:10.1242/dev.098186 (2014).

64 Crocker, J., Abe, N., Rinaldi, L., McGregor, A. P., Frankel, N., Wang, S., Alsawadi, A., Valenti, P., Plaza, S., Payre, F., Mann, R. S. & Stern, D. L. Low affinity binding site clusters confer hox specificity and regulatory robustness. *Cell* **160**, 191-203, doi:10.1016/j.cell.2014.11.041 (2015).

65 Li, L. C., Qiu, W. L., Zhang, Y. W., Xu, Z. R., Xiao, Y. N., Hou, C., Lamaoqiezhong, Yu, P., Cheng, X. & Xu, C. R. Single-cell transcriptomic analyses reveal distinct dorsal/ventral pancreatic programs. *EMBO Rep* **19**, doi:10.15252/embr.201846148 (2018).

66 Brissova, M., Fowler, M. J., Nicholson, W. E., Chu, A., Hirshberg, B., Harlan, D. M. & Powers, A. C. Assessment of human pancreatic islet architecture and composition by laser scanning confocal microscopy. *J Histochem Cytochem* **53**, 1087-1097, doi:10.1369/jhc.5C6684.2005 (2005).

67 Xie, R., Everett, L. J., Lim, H. W., Patel, N. A., Schug, J., Kroon, E., Kelly, O. G., Wang, A., D'Amour, K. A., Robins, A. J., Won, K. J., Kaestner, K. H. & Sander, M. Dynamic chromatin remodeling mediated by polycomb proteins orchestrates pancreatic differentiation of human embryonic stem cells. *Cell Stem Cell* **12**, 224-237, doi:10.1016/j.stem.2012.11.023 (2013).

68 Schulz, T. C., Young, H. Y., Agulnick, A. D., Babin, M. J., Baetge, E. E., Bang, A. G., Bhoumik, A., Cepa, I., Cesario, R. M., Haakmeester, C., Kadoya, K., Kelly, J. R., Kerr, J., Martinson, L. A., McLean, A. B., Moorman, M. A., Payne, J. K., Richardson, M., Ross, K. G., Sherrer, E. S., Song, X., Wilson, A. Z., Brandon, E. P., Green, C. E., Kroon, E. J., Kelly, O. G., D'Amour, K. A. & Robins, A. J. A scalable system for production of functional pancreatic progenitors from human embryonic stem cells. *PLoS One* **7**, e37004, doi:10.1371/journal.pone.0037004 (2012).

69 Rezania, A., Bruin, J. E., Arora, P., Rubin, A., Batushansky, I., Asadi, A., O'Dwyer, S., Quiskamp, N., Mojibian, M., Albrecht, T., Yang, Y. H., Johnson, J. D. & Kieffer, T. J. Reversal of diabetes with insulin-producing cells derived in vitro from human pluripotent stem cells. *Nat Biotechnol* **32**, 1121-1133, doi:10.1038/nbt.3033 (2014).

70 Jin, W., Mulas, F., Gaertner, B., Sui, Y., Wang, J., Matta, I., Zeng, C., Vinckier, N., Wang, A., Nguyen-Ngoc, K. V., Chiou, J., Kaestner, K. H., Frazer, K. A., Carrano, A. C., Shih, H. P. & Sander, M. A Network of microRNAs Acts to Promote Cell Cycle Exit and Differentiation of Human Pancreatic Endocrine Cells. *Iscience* **21**, 681-694, doi:10.1016/j.isci.2019.10.063 (2019).

71 Hurley, K., Ding, J., Villacorta-Martin, C., Herriges, M. J., Jacob, A., Vedaie, M., Alysandratos, K. D., Sun, Y. L., Lin, C., Werder, R. B., Huang, J., Wilson, A. A., Mithal, A., Mostoslavsky, G., Oglesby, I., Caballero, I. S., Guttentag, S. H., Ahangari, F., Kaminski, N., Rodriguez-Fraticelli, A., Camargo, F., Bar-Joseph, Z. & Kotton, D. N. Reconstructed Single-Cell Fate Trajectories Define Lineage Plasticity Windows during Differentiation of Human PSC-Derived Distal Lung Progenitors. *Cell Stem Cell* **26**, 593-608 e598, doi:10.1016/j.stem.2019.12.009 (2020).

72 Jacob, A., Vedaie, M., Roberts, D. A., Thomas, D. C., Villacorta-Martin, C., Alysandratos, K. D., Hawkins, F. & Kotton, D. N. Derivation of self-renewing lung alveolar epithelial type II cells from human pluripotent stem cells. *Nat Protoc* **14**, 3303-3332, doi:10.1038/s41596-019-0220-0 (2019).

73 Hawkins, F., Kramer, P., Jacob, A., Driver, I., Thomas, D. C., McCauley, K. B., Skvir, N., Crane, A. M., Kurmann, A. A., Hollenberg, A. N., Nguyen, S., Wong, B. G., Khalil, A. S., Huang, S. X., Guttentag, S., Rock, J. R., Shannon, J. M., Davis, B. R. & Kotton, D. N. Prospective isolation of NKX2-1-expressing human lung progenitors derived from pluripotent stem cells. *J Clin Invest* **127**, 2277-2294, doi:10.1172/JCI89950 (2017).

74 Maddalo, D., Manchado, E., Concepcion, C. P., Bonetti, C., Vidigal, J. A., Han, Y. C., Ogrodowski, P., Crippa, A., Rekhtman, N., de Stanchina, E., Lowe, S. W. & Ventura, A.

In vivo engineering of oncogenic chromosomal rearrangements with the CRISPR/Cas9 system. *Nature* **516**, 423-427, doi:10.1038/nature13902 (2014).

75 Ran, F. A., Hsu, P. D., Wright, J., Agarwala, V., Scott, D. A. & Zhang, F. Genome engineering using the CRISPR-Cas9 system. *Nat Protoc* **8**, 2281-2308, doi:10.1038/nprot.2013.143 (2013).

76 Rubinson, D. A., Dillon, C. P., Kwiatkowski, A. V., Sievers, C., Yang, L., Kopinja, J., Rooney, D. L., Zhang, M., Ihrig, M. M., McManus, M. T., Gertler, F. B., Scott, M. L. & Van Parijs, L. A lentivirus-based system to functionally silence genes in primary mammalian cells, stem cells and transgenic mice by RNA interference. *Nat Genet* **33**, 401-406, doi:10.1038/ng1117 (2003).

77 Kent, W. J., Sugnet, C. W., Furey, T. S., Roskin, K. M., Pringle, T. H., Zahler, A. M. & Haussler, D. The human genome browser at UCSC. *Genome Res* **12**, 996-1006, doi:10.1101/gr.229102 (2002).

78 Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-1760, doi:10.1093/bioinformatics/btp324 (2009).

79 Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R. & Genome Project Data Processing, S. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079, doi:10.1093/bioinformatics/btp352 (2009).

80 Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y. C., Laslo, P., Cheng, J. X., Murre, C., Singh, H. & Glass, C. K. 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).

81 Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-842, doi:10.1093/bioinformatics/btq033 (2010).

82 Ramirez, F., Ryan, D. P., Gruning, B., Bhardwaj, V., Kilpert, F., Richter, A. S., Heyne, S., Dundar, F. & Manke, T. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res* **44**, W160-165, doi:10.1093/nar/gkw257 (2016).

83 Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M. & Gingeras, T. R. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21, doi:10.1093/bioinformatics/bts635 (2013).

84 Trapnell, C., Williams, B. A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M. J., Salzberg, S. L., Wold, B. J. & Pachter, L. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol* **28**, 511-515, doi:10.1038/nbt.1621 (2010).

85 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).

86 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, doi:10.1038/nmeth.2688 (2013).

87 Zhang, Y., Liu, T., Meyer, C. A., Eeckhoute, J., Johnson, D. S., Bernstein, B. E., Nusbaum, C., Myers, R. M., Brown, M., Li, W. & Liu, X. S. Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**, R137, doi:10.1186/gb-2008-9-9-r137 (2008).

88 Geusz, R. J., Wang, A., Chiou, J., Lancman, J. J., Wetton, N., Kefalopoulou, S., Wang, J., Qiu, Y., Yan, J., Aylward, A., Ren, B., Dong, P. D. S., Gaulton, K. J. & Sander, M. Pancreatic progenitor epigenome maps prioritize type 2 diabetes risk genes with roles in development. *Elife* **10**, doi:10.7554/eLife.59067 (2021).

89 Durand, N. C., Robinson, J. T., Shamim, M. S., Machol, I., Mesirov, J. P., Lander, E. S. & Aiden, E. L. Juicebox Provides a Visualization System for Hi-C Contact Maps with Unlimited Zoom. *Cell Syst* **3**, 99-101, doi:10.1016/j.cels.2015.07.012 (2016).

90 Rao, S. S., Huntley, M. H., Durand, N. C., Stamenova, E. K., Bochkov, I. D., Robinson, J. T., Sanborn, A. L., Machol, I., Omer, A. D., Lander, E. S. & Aiden, E. L. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665-1680, doi:10.1016/j.cell.2014.11.021 (2014).

91 Zheng, G. X., Terry, J. M., Belgrader, P., Ryvkin, P., Bent, Z. W., Wilson, R., Ziraldo, S. B., Wheeler, T. D., McDermott, G. P., Zhu, J., Gregory, M. T., Shuga, J., Montesclaros, L., Underwood, J. G., Masquelier, D. A., Nishimura, S. Y., Schnall-Levin, M., Wyatt, P. W., Hindson, C. M., Bharadwaj, R., Wong, A., Ness, K. D., Beppu, L. W., Deeg, H. J., McFarland, C., Loeb, K. R., Valente, W. J., Ericson, N. G., Stevens, E. A., Radich, J. P., Mikkelsen, T. S., Hindson, B. J. & Bielas, J. H. Massively parallel digital transcriptional profiling of single cells. *Nat Commun* **8**, 14049, doi:10.1038/ncomms14049 (2017).

92 Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W. M., 3rd, Hao, Y., Stoeckius, M., Smibert, P. & Satija, R. Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888-1902 e1821, doi:10.1016/j.cell.2019.05.031 (2019).

93 McLean, C. Y., Bristor, D., Hiller, M., Clarke, S. L., Schaar, B. T., Lowe, C. B., Wenger, A. M. & Bejerano, G. GREAT improves functional interpretation of cis-regulatory regions. *Nat Biotechnol* **28**, 495-501, doi:10.1038/nbt.1630 (2010).

94 Zhou, Y., Zhou, B., Pache, L., Chang, M., Khodabakhshi, A. H., Tanaseichuk, O., Benner, C. & Chanda, S. K. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat Commun* **10**, 1523, doi:10.1038/s41467-019-09234-6 (2019).

95 Loven, J., Hoke, H. A., Lin, C. Y., Lau, A., Orlando, D. A., Vakoc, C. R., Bradner, J. E., Lee, T. I. & Young, R. A. Selective inhibition of tumor oncogenes by disruption of super-enhancers. *Cell* **153**, 320-334, doi:10.1016/j.cell.2013.03.036 (2013).

96 Grant, C. E., Bailey, T. L. & Noble, W. S. FIMO: scanning for occurrences of a given motif. *Bioinformatics* **27**, 1017-1018, doi:10.1093/bioinformatics/btr064 (2011).

97 Bailey, T. L. & Machanick, P. Inferring direct DNA binding from ChIP-seq. *Nucleic Acids Res* **40**, e128, doi:10.1093/nar/gks433 (2012).

98 Aylward, A., Chiou, J., Okino, M. L., Kadakia, N. & Gaulton, K. J. Shared genetic risk contributes to type 1 and type 2 diabetes etiology. *Hum Mol Genet*, doi:10.1093/hmg/ddy314 (2018).

99 Rozowsky, J., Abyzov, A., Wang, J., Alves, P., Raha, D., Harmanci, A., Leng, J., Bjornson, R., Kong, Y., Kitabayashi, N., Bhardwaj, N., Rubin, M., Snyder, M. & Gerstein, M. AlleleSeq: analysis of allele-specific expression and binding in a network framework. *Mol Syst Biol* **7**, 522, doi:10.1038/msb.2011.54 (2011).

100 Mathelier, A., Fornes, O., Arenillas, D. J., Chen, C. Y., Denay, G., Lee, J., Shi, W., Shyr, C., Tan, G., Worsley-Hunt, R., Zhang, A. W., Parcy, F., Lenhard, B., Sandelin, A. & Wasserman, W. W. JASPAR 2016: a major expansion and update of the open-access database of transcription factor binding profiles. *Nucleic Acids Res* **44**, D110-115, doi:10.1093/nar/gkv1176 (2016).

101 Consortium, E. P. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57-74, doi:10.1038/nature11247 (2012).

102 Pique-Regi, R., Degner, J. F., Pai, A. A., Gaffney, D. J., Gilad, Y. & Pritchard, J. K. Accurate inference of transcription factor binding from DNA sequence and chromatin accessibility data. *Genome Res* **21**, 447-455, doi:10.1101/gr.112623.110 (2011).

103 Buecker, C. & Wysocka, J. Enhancers as information integration hubs in development: lessons from genomics. *Trends in genetics : TIG* **28**, 276-284, doi:10.1016/j.tig.2012.02.008 (2012).

104 Creyghton, M. P., Cheng, A. W., Welstead, G. G., Kooistra, T., Carey, B. W., Steine, E. J., Hanna, J., Lodato, M. A., Frampton, G. M., Sharp, P. A., Boyer, L. A., Young, R. A. & Jaenisch, R. Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 21931-21936, doi:10.1073/pnas.1016071107 (2010).

105 Heinz, S., Romanoski, C. E., Benner, C. & Glass, C. K. The selection and function of cell type-specific enhancers. *Nature reviews. Molecular cell biology* **16**, 144-154, doi:10.1038/nrm3949 (2015).

106 Shi, Y., Lan, F., Matson, C., Mulligan, P., Whetstine, J. R., Cole, P. A., Casero, R. A. & Shi, Y. Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. *Cell* **119**, 941-953, doi:10.1016/j.cell.2004.12.012 (2004).

107 Whyte, W. A., Bilodeau, S., Orlando, D. A., Hoke, H. A., Frampton, G. M., Foster, C. T., Cowley, S. M. & Young, R. A. Enhancer decommissioning by LSD1 during embryonic stem cell differentiation. *Nature* **482**, 221-225, doi:10.1038/nature10805 (2012).

108 Shi, Y. J., Matson, C., Lan, F., Iwase, S., Baba, T. & Shi, Y. Regulation of LSD1 histone demethylase activity by its associated factors. *Molecular cell* **19**, 857-864, doi:10.1016/j.molcel.2005.08.027 (2005).

109 Metzger, E., Wissmann, M., Yin, N., Muller, J. M., Schneider, R., Peters, A. H., Gunther, T., Buettner, R. & Schule, R. LSD1 demethylates repressive histone marks to promote androgen-receptor-dependent transcription. *Nature* **437**, 436-439, doi:10.1038/nature04020 (2005).

110 Lee, M. G., Wynder, C., Cooch, N. & Shiekhattar, R. An essential role for CoREST in nucleosomal histone 3 lysine 4 demethylation. *Nature* **437**, 432-435, doi:10.1038/nature04021 (2005).

111 D'Amour, K. A., Bang, A. G., Eliazer, S., Kelly, O. G., Agulnick, A. D., Smart, N. G., Moorman, M. A., Kroon, E., Carpenter, M. K. & Baetge, E. E. Production of pancreatic hormone-expressing endocrine cells from human embryonic stem cells. *Nature biotechnology* **24**, 1392-1401, doi:10.1038/nbt1259 (2006).

112 Yang, S. J., Park, Y. S., Cho, J. H., Moon, B., An, H. J., Lee, J. Y., Xie, Z., Wang, Y., Pocalyko, D., Lee, D. C., Sohn, H. A., Kang, M., Kim, J. Y., Kim, E., Park, K. C., Kim, J. A. & Yeom, Y. I. Regulation of hypoxia responses by flavin adenine dinucleotide-dependent modulation of HIF-1alpha protein stability. *EMBO J* **36**, 1011-1028, doi:10.15252/embj.201694408 (2017).

113 Zentner, G. E. & Scacheri, P. C. The chromatin fingerprint of gene enhancer elements. *The Journal of biological chemistry* **287**, 30888-30896, doi:10.1074/jbc.R111.296491 (2012).

114 Heintzman, N. D., Hon, G. C., Hawkins, R. D., Kheradpour, P., Stark, A., Harp, L. F., Ye, Z., Lee, L. K., Stuart, R. K., Ching, C. W., Ching, K. A., Antosiewicz-Bourget, J. E., Liu, H., Zhang, X., Green, R. D., Lobanenkov, V. V., Stewart, R., Thomson, J. A., Crawford, G. E., Kellis, M. & Ren, B. Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature* **459**, 108-112, doi:10.1038/nature07829 (2009).

115 Wang, J., Lu, F., Ren, Q., Sun, H., Xu, Z., Lan, R., Liu, Y., Ward, D., Quan, J., Ye, T. & Zhang, H. Novel histone demethylase LSD1 inhibitors selectively target cancer cells with pluripotent stem cell properties. *Cancer research* **71**, 7238-7249, doi:10.1158/0008-5472.CAN-11-0896 (2011).

116 Martin, M., Gallego-Llamas, J., Ribes, V., Kedinger, M., Niederreither, K., Chambon, P., Dolle, P. & Gradwohl, G. Dorsal pancreas agenesis in retinoic acid-deficient Raldh2 mutant mice. *Dev Biol* **284**, 399-411, doi:10.1016/j.ydbio.2005.05.035 (2005).

117 Stafford, D. & Prince, V. E. Retinoic acid signaling is required for a critical early step in zebrafish pancreatic development. *Curr Biol* **12**, 1215-1220 (2002).

118 Cunningham, T. J. & Duester, G. Mechanisms of retinoic acid signalling and its roles in organ and limb development. *Nat Rev Mol Cell Biol* **16**, 110-123, doi:10.1038/nrm3932 (2015).

119 Ostrom, M., Loffler, K. A., Edfalk, S., Selander, L., Dahl, U., Ricordi, C., Jeon, J., Correa-Medina, M., Diez, J. & Edlund, H. Retinoic acid promotes the generation of pancreatic endocrine progenitor cells and their further differentiation into beta-cells. *PloS one* **3**, e2841, doi:10.1371/journal.pone.0002841 (2008).

120 Lomberk, G. A., Imoto, I., Gebelein, B., Urrutia, R. & Cook, T. A. Conservation of the TGFbeta/Labial homeobox signaling loop in endoderm-derived cells between Drosophila and mammals. *Pancreatology* **10**, 74-84, doi:10.1159/000276895 (2010).

121 Han, X., Gui, B., Xiong, C., Zhao, L., Liang, J., Sun, L., Yang, X., Yu, W., Si, W., Yan, R., Yi, X., Zhang, D., Li, W., Li, L., Yang, J., Wang, Y., Sun, Y. E., Zhang, D., Meng, A. & Shang, Y. Destabilizing LSD1 by Jade-2 promotes neurogenesis: an antibraking system in neural development. *Molecular cell* **55**, 482-494, doi:10.1016/j.molcel.2014.06.006 (2014).

122 Sun, G., Alzayady, K., Stewart, R., Ye, P., Yang, S., Li, W. & Shi, Y. Histone demethylase LSD1 regulates neural stem cell proliferation. *Mol Cell Biol* **30**, 1997-2005, doi:10.1128/MCB.01116-09 (2010).

123 Maden, M. Retinoic acid in the development, regeneration and maintenance of the nervous system. *Nat Rev Neurosci* **8**, 755-765, doi:10.1038/nrn2212 (2007).

124 Dorighi, K. M., Swigut, T., Henriques, T., Bhanu, N. V., Scruggs, B. S., Nady, N., Still, C. D., 2nd, Garcia, B. A., Adelman, K. & Wysocka, J. Mll3 and Mll4 Facilitate Enhancer RNA Synthesis and Transcription from Promoters Independently of H3K4 Monomethylation. *Molecular cell* **66**, 568-576 e564, doi:10.1016/j.molcel.2017.04.018 (2017).

125 Kaikkonen, M. U., Spann, N. J., Heinz, S., Romanoski, C. E., Allison, K. A., Stender, J. D., Chun, H. B., Tough, D. F., Prinjha, R. K., Benner, C. & Glass, C. K. Remodeling of the enhancer landscape during macrophage activation is coupled to enhancer transcription. *Molecular cell* **51**, 310-325, doi:10.1016/j.molcel.2013.07.010 (2013).

126 Gao, N., LeLay, J., Vatamaniuk, M. Z., Rieck, S., Friedman, J. R. & Kaestner, K. H. Dynamic regulation of Pdx1 enhancers by Foxa1 and Foxa2 is essential for pancreas development. *Genes Dev* **22**, 3435-3448, doi:10.1101/gad.1752608 (2008).

127 Shi, Z. D., Lee, K., Yang, D., Amin, S., Verma, N., Li, Q. V., Zhu, Z., Soh, C. L., Kumar, R., Evans, T., Chen, S. & Huangfu, D. Genome Editing in hPSCs Reveals GATA6 Haploinsufficiency and a Genetic Interaction with GATA4 in Human Pancreatic Development. *Cell Stem Cell* **20**, 675-688 e676, doi:10.1016/j.stem.2017.01.001 (2017).

128 Tiyaboonchai, A., Cardenas-Diaz, F. L., Ying, L., Maguire, J. A., Sim, X., Jobaliya, C., Gagne, A. L., Kishore, S., Stanescu, D. E., Hughes, N., De Leon, D. D., French, D. L. & Gadue, P. GATA6 Plays an Important Role in the Induction of Human Definitive Endoderm, Development of the Pancreas, and Functionality of Pancreatic beta Cells. *Stem Cell Reports* **8**, 589-604, doi:10.1016/j.stemcr.2016.12.026 (2017).

129 Cai, C., He, H. H., Gao, S., Chen, S., Yu, Z., Gao, Y., Chen, S., Chen, M. W., Zhang, J., Ahmed, M., Wang, Y., Metzger, E., Schule, R., Liu, X. S., Brown, M. & Balk, S. P. Lysine-specific demethylase 1 has dual functions as a major regulator of androgen receptor transcriptional activity. *Cell Rep* **9**, 1618-1627, doi:10.1016/j.celrep.2014.11.008 (2014).

130 Choi, S. H., Estaras, C., Moresco, J. J., Yates, J. R., 3rd & Jones, K. A. alpha-Catenin interacts with APC to regulate beta-catenin proteolysis and transcriptional repression of Wnt target genes. *Genes Dev* **27**, 2473-2488, doi:10.1101/gad.229062.113 (2013).

131 Mulligan, P., Yang, F., Di Stefano, L., Ji, J. Y., Ouyang, J., Nishikawa, J. L., Toiber, D., Kulkarni, M., Wang, Q., Najafi-Shoushtari, S. H., Mostoslavsky, R., Gygi, S. P., Gill, G., Dyson, N. J. & Naar, A. M. A SIRT1-LSD1 corepressor complex regulates Notch target gene expression and development. *Molecular cell* **42**, 689-699, doi:10.1016/j.molcel.2011.04.020 (2011).

132 Kitagawa, M. Notch signalling in the nucleus: roles of Mastermind-like (MAML) transcriptional coactivators. *J Biochem* **159**, 287-294, doi:10.1093/jb/mvv123 (2016).

133 Novellasdemunt, L., Antas, P. & Li, V. S. Targeting Wnt signaling in colorectal cancer. A Review in the Theme: Cell Signaling: Proteins, Pathways and Mechanisms. *Am J Physiol Cell Physiol* **309**, C511-521, doi:10.1152/ajpcell.00117.2015 (2015).

134 Wang, J., Scully, K., Zhu, X., Cai, L., Zhang, J., Prefontaine, G. G., Krones, A., Ohgi, K. A., Zhu, P., Garcia-Bassets, I., Liu, F., Taylor, H., Lozach, J., Jayes, F. L., Korach, K. S., Glass, C. K., Fu, X. D. & Rosenfeld, M. G. Opposing LSD1 complexes function in developmental gene activation and repression programmes. *Nature* **446**, 882-887, doi:10.1038/nature05671 (2007).

135 Popova, E. Y., Pinzon-Guzman, C., Salzberg, A. C., Zhang, S. S. & Barnstable, C. J. LSD1-Mediated Demethylation of H3K4me2 Is Required for the Transition from Late Progenitor to Differentiated Mouse Rod Photoreceptor. *Mol Neurobiol* **53**, 4563-4581, doi:10.1007/s12035-015-9395-8 (2016).

136 Kerenyi, M. A., Shao, Z., Hsu, Y. J., Guo, G., Luc, S., O'Brien, K., Fujiwara, Y., Peng, C., Nguyen, M. & Orkin, S. H. Histone demethylase Lsd1 represses hematopoietic stem and progenitor cell signatures during blood cell maturation. *eLife* **2**, e00633, doi:10.7554/eLife.00633 (2013).

137 Di Stefano, L., Ji, J. Y., Moon, N. S., Herr, A. & Dyson, N. Mutation of Drosophila Lsd1 disrupts H3-K4 methylation, resulting in tissue-specific defects during development. *Curr Biol* **17**, 808-812, doi:10.1016/j.cub.2007.03.068 (2007).

138 Katz, D. J., Edwards, T. M., Reinke, V. & Kelly, W. G. A C. elegans LSD1 demethylase contributes to germline immortality by reprogramming epigenetic memory. *Cell* **137**, 308-320, doi:10.1016/j.cell.2009.02.015 (2009).

139 Gu, G., Dubauskaite, J. & Melton, D. A. Direct evidence for the pancreatic lineage: NGN3+ cells are islet progenitors and are distinct from duct progenitors. *Development* **129**, 2447-2457 (2002).

140 Srinivas, S., Watanabe, T., Lin, C. S., William, C. M., Tanabe, Y., Jessell, T. M. & Costantini, F. Cre reporter strains produced by targeted insertion of EYFP and ECFP into the ROSA26 locus. *BMC Dev Biol* **1**, 4 (2001).

141 Campbell-Thompson, M., Wasserfall, C., Kaddis, J., Albanese-O'Neill, A., Staeva, T., Nierras, C., Moraski, J., Rowe, P., Gianani, R., Eisenbarth, G., Crawford, J., Schatz, D.,

Pugliese, A. & Atkinson, M. Network for Pancreatic Organ Donors with Diabetes (nPOD): developing a tissue biobank for type 1 diabetes. *Diabetes/metabolism research and reviews* **28**, 608-617, doi:10.1002/dmrr.2316 (2012).

142 Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357-359, doi:10.1038/nmeth.1923 (2012).

143 Henseleit, K. D., Nelson, S. B., Kuhlbrodt, K., Hennings, J. C., Ericson, J. & Sander, M. NKX6 transcription factor activity is required for alpha- and beta-cell development in the pancreas. *Development* **132**, 3139-3149, doi:10.1242/dev.01875 (2005).

144 Halban, P. A., Polonsky, K. S., Bowden, D. W., Hawkins, M. A., Ling, C., Mather, K. J., Powers, A. C., Rhodes, C. J., Sussel, L. & Weir, G. C. beta-cell failure in type 2 diabetes: postulated mechanisms and prospects for prevention and treatment. *Diabetes Care* **37**, 1751-1758, doi:10.2337/dc14-0396 (2014).

145 Pasquali, L., Gaulton, K. J., Rodriguez-Segui, S. A., Mularoni, L., Miguel-Escalada, I., Akerman, I., Tena, J. J., Moran, I., Gomez-Marin, C., van de Bunt, M., Ponsa-Cobas, J., Castro, N., Nammo, T., Cebola, I., Garcia-Hurtado, J., Maestro, M. A., Pattou, F., Piemonti, L., Berney, T., Gloyn, A. L., Ravassard, P., Skarmeta, J. L. G., Muller, F., McCarthy, M. I. & Ferrer, J. Pancreatic islet enhancer clusters enriched in type 2 diabetes risk-associated variants. *Nat Genet* **46**, 136-143, doi:10.1038/ng.2870 (2014).

146 Parker, S. C., Stitzel, M. L., Taylor, D. L., Orozco, J. M., Erdos, M. R., Akiyama, J. A., van Bueren, K. L., Chines, P. S., Narisu, N., Program, N. C. S., Black, B. L., Visel, A., Pennacchio, L. A., Collins, F. S., National Institutes of Health Intramural Sequencing Center Comparative Sequencing Program, A. & Authors, N. C. S. P. 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 (2013).

147 Gaulton, K. J., Nammo, T., Pasquali, L., Simon, J. M., Giresi, P. G., Fogarty, M. P., Panhuis, T. M., Mieczkowski, P., Secchi, A., Bosco, D., Berney, T., Montanya, E., Mohlke, K. L., Lieb, J. D. & Ferrer, J. A map of open chromatin in human pancreatic islets. *Nat Genet* **42**, 255-259, doi:10.1038/ng.530 (2010).

148 Gaulton, K. J., Ferreira, T., Lee, Y., Raimondo, A., Magi, R., Reschen, M. E., Mahajan, A., Locke, A., Rayner, N. W., Robertson, N., Scott, R. A., Prokopenko, I., Scott, L. J., Green, T., Sparso, T., Thuillier, D., Yengo, L., Grallert, H., Wahl, S., Franberg, M., Strawbridge, R. J., Kestler, H., Chheda, H., Eisele, L., Gustafsson, S., Steinthorsdottir, V., Thorleifsson, G., Qi, L., Karssen, L. C., van Leeuwen, E. M., Willems, S. M., Li, M., Chen, H., Fuchsberger, C., Kwan, P., Ma, C., Linderman, M., Lu, Y., Thomsen, S. K., Rundle, J. K., Beer, N. L., van de Bunt, M., Chalisey, A., Kang, H. M., Voight, B. F., Abecasis, G. R., Almgren, P., Baldassarre, D., Balkau, B., Benediktsson, R., Bluher, M., Boeing, H., Bonnycastle, L. L., Bottinger, E. P., Burtt, N. P., Carey, J., Charpentier, G., Chines, P. S., Cornelis, M. C., Couper, D. J., Crenshaw, A. T., van Dam, R. M., Doney, A. S., Dorkhan, M., Edkins, S., Eriksson, J. G., Esko, T., Eury, E., Fadista, J., Flannick, J., Fontanillas, P., Fox, C., Franks, P. W., Gertow, K., Gieger, C., Gigante, B., Gottesman, O., Grant, G. B., Grarup, N., Groves, C. J., Hassinen, M., Have, C. T., Herder, C., Holmen, O. L., Hreidarsson, A. B., Humphries, S. E., Hunter, D. J., Jackson, A. U., Jonsson, A., Jorgensen, M. E., Jorgensen, T., Kao, W. H., Kerrison, N. D.,

Kinnunen, L., Klopp, N., Kong, A., Kovacs, P., Kraft, P., Kravic, J., Langford, C., Leander, K., Liang, L., Lichtner, P., Lindgren, C. M., Lindholm, E., Linneberg, A., Liu, C. T., Lobbens, S., Luan, J., Lyssenko, V., Mannisto, S., McLeod, O., Meyer, J., Mihailov, E., Mirza, G., Muhleisen, T. W., Muller-Nurasyid, M., Navarro, C., Nothen, M. M., Oskolkov, N. N., Owen, K. R., Palli, D., Pechlivanis, S., Peltonen, L., Perry, J. R., Platou, C. G., Roden, M., Ruderfer, D., Rybin, D., van der Schouw, Y. T., Sennblad, B., Sigurethsson, G., Stancakova, A., Steinbach, G., Storm, P., Strauch, K., Stringham, H. M., Sun, Q., Thorand, B., Tikkanen, E., Tonjes, A., Trakalo, J., Tremoli, E., Tuomi, T., Wennauer, R., Wiltshire, S., Wood, A. R., Zeggini, E., Dunham, I., Birney, E., Pasquali, L., Ferrer, J., Loos, R. J., Dupuis, J., Florez, J. C., Boerwinkle, E., Pankow, J. S., van Duijn, C., Sijbrands, E., Meigs, J. B., Hu, F. B., Thorsteinsdottir, U., Stefansson, K., Lakka, T. A., Rauramaa, R., Stumvoll, M., Pedersen, N. L., Lind, L., Keinanen-Kiukaanniemi, S. M., Korpi-Hyovalti, E., Saaristo, T. E., Saltevo, J., Kuusisto, J., Laakso, M., Metspalu, A., Erbel, R., Jocke, K. H., Moebus, S., Ripatti, S., Salomaa, V., Ingelsson, E., Boehm, B. O., Bergman, R. N., Collins, F. S., Mohlke, K. L., Koistinen, H., Tuomilehto, J., Hveem, K., Njolstad, I., Deloukas, P., Donnelly, P. J., Frayling, T. M., Hattersley, A. T., de Faire, U., Hamsten, A., Illig, T., Peters, A., Cauchi, S., Sladek, R., Froguel, P., Hansen, T., Pedersen, O., Morris, A. D., Palmer, C. N., Kathiresan, S., Melander, O., Nilsson, P. M., Groop, L. C., Barroso, I., Langenberg, C., Wareham, N. J., O'Callaghan, C. A., Gloyn, A. L., Altshuler, D., Boehnke, M., Teslovich, T. M., McCarthy, M. I., Morris, A. P., Replication, D. I. G. & Meta-analysis, C. Genetic fine mapping and genomic annotation defines causal mechanisms at type 2 diabetes susceptibility loci. *Nat Genet* **47**, 1415-1425, doi:10.1038/ng.3437 (2015).

149 Fuchsberger, C., Flannick, J., Teslovich, T. M., Mahajan, A., Agarwala, V., Gaulton, K. J., Ma, C., Fontanillas, P., Moutsianas, L., McCarthy, D. J., Rivas, M. A., Perry, J. R. B., Sim, X., Blackwell, T. W., Robertson, N. R., Rayner, N. W., Cingolani, P., Locke, A. E., Tajes, J. F., Highland, H. M., Dupuis, J., Chines, P. S., Lindgren, C. M., Hartl, C., Jackson, A. U., Chen, H., Huyghe, J. R., van de Bunt, M., Pearson, R. D., Kumar, A., Muller-Nurasyid, M., Grarup, N., Stringham, H. M., Gamazon, E. R., Lee, J., Chen, Y., Scott, R. A., Below, J. E., Chen, P., Huang, J., Go, M. J., Stitzel, M. L., Pasko, D., Parker, S. C. J., Varga, T. V., Green, T., Beer, N. L., Day-Williams, A. G., Ferreira, T., Fingerlin, T., Horikoshi, M., Hu, C., Huh, I., Ikram, M. K., Kim, B. J., Kim, Y., Kim, Y. J., Kwon, M. S., Lee, J., Lee, S., Lin, K. H., Maxwell, T. J., Nagai, Y., Wang, X., Welch, R. P., Yoon, J., Zhang, W., Barzilai, N., Voight, B. F., Han, B. G., Jenkinson, C. P., Kuulasmaa, T., Kuusisto, J., Manning, A., Ng, M. C. Y., Palmer, N. D., Balkau, B., Stancakova, A., Abboud, H. E., Boeing, H., Giedraitis, V., Prabhakaran, D., Gottesman, O., Scott, J., Carey, J., Kwan, P., Grant, G., Smith, J. D., Neale, B. M., Purcell, S., Butterworth, A. S., Howson, J. M. M., Lee, H. M., Lu, Y., Kwak, S. H., Zhao, W., Danesh, J., Lam, V. K. L., Park, K. S., Saleheen, D., So, W. Y., Tam, C. H. T., Afzal, U., Aguilar, D., Arya, R., Aung, T., Chan, E., Navarro, C., Cheng, C. Y., Palli, D., Correa, A., Curran, J. E., Rybin, D., Farook, V. S., Fowler, S. P., Freedman, B. I., Griswold, M., Hale, D. E., Hicks, P. J., Khor, C. C., Kumar, S., Lehne, B., Thuillier, D., Lim, W. Y., Liu, J., van der Schouw, Y. T., Loh, M., Musani, S. K., Puppala, S., Scott, W. R., Yengo, L., Tan, S. T., Taylor, H. A., Jr., Thameem, F., Wilson, G., Sr., Wong, T. Y., Njolstad, P. R., Levy, J. C., Mangino, M., Bonnycastle, L. L., Schwarzmayr, T., Fadista, J., Surdulescu, G. L., Herder, C., Groves, C. J., Wieland, T., Bork-Jensen, J., Brandslund, I., Christensen, C., Koistinen, H. A., Doney, A. S. F., Kinnunen, L., Esko, T., Farmer, A. J., Hakaste, L., Hodgkiss, D., Kravic, J., Lyssenko, V., Hollensted, M., Jorgensen, M. E., Jorgensen, T., Ladenvall, C., Justesen, J. M., Karajamaki, A., Kriebel, J., Rathmann, W., Lannfelt, L., Lauritzen, T., Narisu, N., Linneberg, A., Melander, O., Milani, L., Neville, M., Orho-

Melander, M., Qi, L., Qi, Q., Roden, M., Rolandsson, O., Swift, A., Rosengren, A. H., Stirrups, K., Wood, A. R., Mihailov, E., Blancher, C., Carneiro, M. O., Maguire, J., Poplin, R., Shakir, K., Fennell, T., DePristo, M., de Angelis, M. H., Deloukas, P., Gjesing, A. P., Jun, G., Nilsson, P., Murphy, J., Onofrio, R., Thorand, B., Hansen, T., Meisinger, C., Hu, F. B., Isomaa, B., Karpe, F., Liang, L., Peters, A., Huth, C., O'Rahilly, S. P., Palmer, C. N. A., Pedersen, O., Rauramaa, R., Tuomilehto, J., Salomaa, V., Watanabe, R. M., Syvanen, A. C., Bergman, R. N., Bharadwaj, D., Bottinger, E. P., Cho, Y. S., Chandak, G. R., Chan, J. C. N., Chia, K. S., Daly, M. J., Ebrahim, S. B., Langenberg, C., Elliott, P., Jablonski, K. A., Lehman, D. M., Jia, W., Ma, R. C. W., Pollin, T. I., Sandhu, M., Tandon, N., Froguel, P., Barroso, I., Teo, Y. Y., Zeggini, E., Loos, R. J. F., Small, K. S., Ried, J. S., DeFronzo, R. A., Grallert, H., Glaser, B., Metspalu, A., Wareham, N. J., Walker, M., Banks, E., Gieger, C., Ingelsson, E., Im, H. K., Illig, T., Franks, P. W., Buck, G., Trakalo, J., Buck, D., Prokopenko, I., Magi, R., Lind, L., Farjoun, Y., Owen, K. R., Gloyn, A. L., Strauch, K., Tuomi, T., Kooner, J. S., Lee, J. Y., Park, T., Donnelly, P., Morris, A. D., Hattersley, A. T., Bowden, D. W., Collins, F. S., Atzmon, G., Chambers, J. C., Spector, T. D., Laakso, M., Strom, T. M., Bell, G. I., Blangero, J., Duggirala, R., Tai, E. S., McVean, G., Hanis, C. L., Wilson, J. G., Seielstad, M., Frayling, T. M., Meigs, J. B., Cox, N. J., Sladek, R., Lander, E. S., Gabriel, S., Burtt, N. P., Mohlke, K. L., Meitinger, T., Groop, L., Abecasis, G., Florez, J. C., Scott, L. J., Morris, A. P., Kang, H. M., Boehnke, M., Altshuler, D. & McCarthy, M. I. The genetic architecture of type 2 diabetes. *Nature* **536**, 41-47, doi:10.1038/nature18642 (2016).

150 Thurner, M., van de Bunt, M., Torres, J. M., Mahajan, A., Nylander, V., Bennett, A. J., Gaulton, K. J., Barrett, A., Burrows, C., Bell, C. G., Lowe, R., Beck, S., Rakyan, V. K., Gloyn, A. L. & McCarthy, M. I. Integration of human pancreatic islet genomic data refines regulatory mechanisms at Type 2 Diabetes susceptibility loci. *Elife* **7**, doi:10.7554/eLife.31977 (2018).

151 Varshney, A., Scott, L. J., Welch, R. P., Erdos, M. R., Chines, P. S., Narisu, N., Albanus, R. D., Orchard, P., Wolford, B. N., Kursawe, R., Vadlamudi, S., Cannon, M. E., Didion, J. P., Hensley, J., Kirilusha, A., Program, N. C. S., Bonnycastle, L. L., Taylor, D. L., Watanabe, R., Mohlke, K. L., Boehnke, M., Collins, F. S., Parker, S. C. & Stitzel, M. L. Genetic regulatory signatures underlying islet gene expression and type 2 diabetes. *Proc Natl Acad Sci U S A* **114**, 2301-2306, doi:10.1073/pnas.1621192114 (2017).

152 Chiou, J., Zeng, C., Cheng, Z., Han, J. Y., Schlichting, M., Huang, S., Wang, J., Sui, Y., Deogaygay, A., Okino, M.-L., Qiu, Y., Sun, Y., Kudtarkar, P., Fang, R., Preissl, S., Sander, M., Gorkin, D. & Gaulton, K. J. Single cell chromatin accessibility reveals pancreatic islet cell type- and state-specific regulatory programs of diabetes risk. *bioRxiv*, 693671, doi:10.1101/693671 (2019).

153 Greenwald, W. W., Chiou, J., Yan, J., Qiu, Y., Dai, N., Wang, A., Nariai, N., Aylward, A., Han, J. Y., Kadakia, N., Regue, L., Okino, M. L., Drees, F., Kramer, D., Vinckier, N., Minichiello, L., Gorkin, D., Avruch, J., Frazer, K. A., Sander, M., Ren, B. & Gaulton, K. J. Pancreatic islet chromatin accessibility and conformation reveals distal enhancer networks of type 2 diabetes risk. *Nat Commun* **10**, 2078, doi:10.1038/s41467-019-09975-4 (2019).

154 Flannick, J., Mercader, J. M., Fuchsberger, C., Udler, M. S., Mahajan, A., Wessel, J., Teslovich, T. M., Caulkins, L., Koesterer, R., Barajas-Olmos, F., Blackwell, T. W., Boerwinkle, E., Brody, J. A., Centeno-Cruz, F., Chen, L., Chen, S., Contreras-Cubas, C.,

Cordova, E., Correa, A., Cortes, M., DeFronzo, R. A., Dolan, L., Drews, K. L., Elliott, A., Floyd, J. S., Gabriel, S., Garay-Sevilla, M. E., Garcia-Ortiz, H., Gross, M., Han, S., Heard-Costa, N. L., Jackson, A. U., Jorgensen, M. E., Kang, H. M., Kelsey, M., Kim, B. J., Koistinen, H. A., Kuusisto, J., Leader, J. B., Linneberg, A., Liu, C. T., Liu, J., Lyssenko, V., Manning, A. K., Marcketta, A., Malacara-Hernandez, J. M., Martinez-Hernandez, A., Matsuo, K., Mayer-Davis, E., Mendoza-Caamal, E., Mohlke, K. L., Morrison, A. C., Ndungu, A., Ng, M. C. Y., O'Dushlaine, C., Payne, A. J., Pihoker, C., Broad Genomics, P., Post, W. S., Preuss, M., Psaty, B. M., Vasan, R. S., Rayner, N. W., Reiner, A. P., Revilla-Monsalve, C., Robertson, N. R., Santoro, N., Schurmann, C., So, W. Y., Soberon, X., Stringham, H. M., Strom, T. M., Tam, C. H. T., Thameem, F., Tomlinson, B., Torres, J. M., Tracy, R. P., van Dam, R. M., Vujkovic, M., Wang, S., Welch, R. P., Witte, D. R., Wong, T. Y., Atzmon, G., Barzilai, N., Blangero, J., Bonnycastle, L. L., Bowden, D. W., Chambers, J. C., Chan, E., Cheng, C. Y., Cho, Y. S., Collins, F. S., de Vries, P. S., Duggirala, R., Glaser, B., Gonzalez, C., Gonzalez, M. E., Groop, L., Kooner, J. S., Kwak, S. H., Laakso, M., Lehman, D. M., Nilsson, P., Spector, T. D., Tai, E. S., Tuomi, T., Tuomilehto, J., Wilson, J. G., Aguilar-Salinas, C. A., Bottinger, E., Burke, B., Carey, D. J., Chan, J. C. N., Dupuis, J., Frossard, P., Heckbert, S. R., Hwang, M. Y., Kim, Y. J., Kirchner, H. L., Lee, J. Y., Lee, J., Loos, R. J. F., Ma, R. C. W., Morris, A. D., O'Donnell, C. J., Palmer, C. N. A., Pankow, J., Park, K. S., Rasheed, A., Saleheen, D., Sim, X., Small, K. S., Teo, Y. Y., Haiman, C., Hanis, C. L., Henderson, B. E., Orozco, L., Tusie-Luna, T., Dewey, F. E., Baras, A., Gieger, C., Meitinger, T., Strauch, K., Lange, L., Grarup, N., Hansen, T., Pedersen, O., Zeitler, P., Dabelea, D., Abecasis, G., Bell, G. I., Cox, N. J., Seielstad, M., Sladek, R., Meigs, J. B., Rich, S. S., Rotter, J. I., Discov, E. H. R. C., Charge, LuCamp, ProDiGy, GoT2D, Esp, Sigma, T. D., T2D, G., Amp T2D, G., Altshuler, D., Burtt, N. P., Scott, L. J., Morris, A. P., Florez, J. C., McCarthy, M. I. & Boehnke, M. Exome sequencing of 20,791 cases of type 2 diabetes and 24,440 controls. *Nature* **570**, 71-76, doi:10.1038/s41586-019-1231-2 (2019).

155 Steinthorsdottir, V., Thorleifsson, G., Sulem, P., Helgason, H., Grarup, N., Sigurdsson, A., Helgadottir, H. T., Johannsdottir, H., Magnusson, O. T., Gudjonsson, S. A., Justesen, J. M., Harder, M. N., Jorgensen, M. E., Christensen, C., Brandslund, I., Sandbaek, A., Lauritzen, T., Vestergaard, H., Linneberg, A., Jorgensen, T., Hansen, T., Daneshpour, M. S., Fallah, M. S., Hreidarsson, A. B., Sigurdsson, G., Azizi, F., Benediktsson, R., Masson, G., Helgason, A., Kong, A., Gudbjartsson, D. F., Pedersen, O., Thorsteinsdottir, U. & Stefansson, K. Identification of low-frequency and rare sequence variants associated with elevated or reduced risk of type 2 diabetes. *Nat Genet* **46**, 294-298, doi:10.1038/ng.2882 (2014).

156 Gaertner, B., Carrano, A. C. & Sander, M. Human stem cell models: lessons for pancreatic development and disease. *Genes Dev* **33**, 1475-1490, doi:10.1101/gad.331397.119 (2019).

157 Urakami, T. Maturity-onset diabetes of the young (MODY): current perspectives on diagnosis and treatment. *Diabetes Metab Syndr Obes* **12**, 1047-1056, doi:10.2147/DMSO.S179793 (2019).

158 Lumey, L. H., Khalangot, M. D. & Vaiserman, A. M. Association between type 2 diabetes and prenatal exposure to the Ukraine famine of 1932-33: a retrospective cohort study. *Lancet Diabetes Endocrinol* **3**, 787-794, doi:10.1016/S2213-8587(15)00279-X (2015).

159 Portha, B., Chavey, A. & Movassat, J. Early-life origins of type 2 diabetes: fetal programming of the beta-cell mass. *Exp Diabetes Res* **2011**, 105076, doi:10.1155/2011/105076 (2011).

160 Nielsen, J. H., Haase, T. N., Jaksch, C., Nalla, A., Sostrup, B., Nalla, A. A., Larsen, L., Rasmussen, M., Dalgaard, L. T., Gaarn, L. W., Thams, P., Kofod, H. & Billestrup, N. Impact of fetal and neonatal environment on beta cell function and development of diabetes. *Acta Obstet Gynecol Scand* **93**, 1109-1122, doi:10.1111/aogs.12504 (2014).

161 Ernst, J. & Kellis, M. ChromHMM: automating chromatin-state discovery and characterization. *Nat Methods* **9**, 215-216, doi:10.1038/nmeth.1906 (2012).

162 Conrad, E., Stein, R. & Hunter, C. S. Revealing transcription factors during human pancreatic beta cell development. *Trends Endocrinol Metab* **25**, 407-414, doi:10.1016/j.tem.2014.03.013 (2014).

163 Masui, T., Long, Q., Beres, T. M., Magnuson, M. A. & MacDonald, R. J. Early pancreatic development requires the vertebrate Suppressor of Hairless (RBPJ) in the PTF1 bHLH complex. *Genes Dev* **21**, 2629-2643, doi:10.1101/gad.1575207 (2007).

164 Ahnfelt-Ronne, J., Ravassard, P., Pardanaud-Glavieux, C., Scharfmann, R. & Serup, P. Mesenchymal bone morphogenetic protein signaling is required for normal pancreas development. *Diabetes* **59**, 1948-1956, doi:10.2337/db09-1010 (2010).

165 Sui, L., Geens, M., Sermon, K., Bouwens, L. & Mfopou, J. K. Role of BMP signaling in pancreatic progenitor differentiation from human embryonic stem cells. *Stem Cell Rev Rep* **9**, 569-577, doi:10.1007/s12015-013-9435-6 (2013).

166 Li, X. Y., Zhai, W. J. & Teng, C. B. Notch Signaling in Pancreatic Development. *Int J Mol Sci* **17**, doi:10.3390/ijms17010048 (2015).

167 Sharon, N., Vanderhooft, J., Straubhaar, J., Mueller, J., Chawla, R., Zhou, Q., Engquist, E. N., Trapnell, C., Gifford, D. K. & Melton, D. A. Wnt Signaling Separates the Progenitor and Endocrine Compartments during Pancreas Development. *Cell Rep* **27**, 2281-2291 e2285, doi:10.1016/j.celrep.2019.04.083 (2019).

168 Murtaugh, L. C. The what, where, when and how of Wnt/beta-catenin signaling in pancreas development. *Organogenesis* **4**, 81-86, doi:10.4161/org.4.2.5853 (2008).

169 Bulik-Sullivan, B. K., Loh, P. R., Finucane, H. K., Ripke, S., Yang, J., Schizophrenia Working Group of the Psychiatric Genomics, C., Patterson, N., Daly, M. J., Price, A. L. & Neale, B. M. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet* **47**, 291-295, doi:10.1038/ng.3211 (2015).

170 Finucane, H. K., Bulik-Sullivan, B., Gusev, A., Trynka, G., Reshef, Y., Loh, P. R., Anttila, V., Xu, H., Zang, C., Farh, K., Ripke, S., Day, F. R., ReproGen, C., Schizophrenia Working Group of the Psychiatric Genomics, C., Consortium, R., Purcell, S., Stahl, E., Lindstrom, S., Perry, J. R., Okada, Y., Raychaudhuri, S., Daly, M. J., Patterson, N., Neale, B. M. & Price, A. L. Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat Genet* **47**, 1228-1235, doi:10.1038/ng.3404 (2015).

171 Perry, J. R., Voight, B. F., Yengo, L., Amin, N., Dupuis, J., Ganser, M., Grallert, H., Navarro, P., Li, M., Qi, L., Steinthorsdottir, V., Scott, R. A., Almgren, P., Arking, D. E., Aulchenko, Y., Balkau, B., Benediktsson, R., Bergman, R. N., Boerwinkle, E., Bonnycastle, L., Burtt, N. P., Campbell, H., Charpentier, G., Collins, F. S., Gieger, C., Green, T., Hadjadj, S., Hattersley, A. T., Herder, C., Hofman, A., Johnson, A. D., Kottgen, A., Kraft, P., Labrune, Y., Langenberg, C., Manning, A. K., Mohlke, K. L., Morris, A. P., Oostra, B., Pankow, J., Petersen, A. K., Pramstaller, P. P., Prokopenko, I., Rathmann, W., Rayner, W., Roden, M., Rudan, I., Rybin, D., Scott, L. J., Sigurdsson, G., Sladek, R., Thorleifsson, G., Thorsteinsdottir, U., Tuomilehto, J., Uitterlinden, A. G., Vivequin, S., Weedon, M. N., Wright, A. F., Magic, Consortium, D., Consortium, G., Hu, F. B., Illig, T., Kao, L., Meigs, J. B., Wilson, J. F., Stefansson, K., van Duijn, C., Altschuler, D., Morris, A. D., Boehnke, M., McCarthy, M. I., Froguel, P., Palmer, C. N., Wareham, N. J., Groop, L., Frayling, T. M. & Cauchi, S. Stratifying type 2 diabetes cases by BMI identifies genetic risk variants in LAMA1 and enrichment for risk variants in lean compared to obese cases. *PLoS Genet* **8**, e1002741, doi:10.1371/journal.pgen.1002741 (2012).

172 Roadmap Epigenomics, C., Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., Heravi-Moussavi, A., Kheradpour, P., Zhang, Z., Wang, J., Ziller, M. J., Amin, V., Whitaker, J. W., Schultz, M. D., Ward, L. D., Sarkar, A., Quon, G., Sandstrom, R. S., Eaton, M. L., Wu, Y. C., Pfenning, A. R., Wang, X., Claussnitzer, M., Liu, Y., Coarfa, C., Harris, R. A., Shoresh, N., Epstein, C. B., Gjoneska, E., Leung, D., Xie, W., Hawkins, R. D., Lister, R., Hong, C., Gascard, P., Mungall, A. J., Moore, R., Chuah, E., Tam, A., Canfield, T. K., Hansen, R. S., Kaul, R., Sabo, P. J., Bansal, M. S., Carles, A., Dixon, J. R., Farh, K. H., Feizi, S., Karlic, R., Kim, A. R., Kulkarni, A., Li, D., Lowdon, R., Elliott, G., Mercer, T. R., Neph, S. J., Onuchic, V., Polak, P., Rajagopal, N., Ray, P., Sallari, R. C., Siebenthall, K. T., Sinnott-Armstrong, N. A., Stevens, M., Thurman, R. E., Wu, J., Zhang, B., Zhou, X., Beaudet, A. E., Boyer, L. A., De Jager, P. L., Farnham, P. J., Fisher, S. J., Haussler, D., Jones, S. J., Li, W., Marra, M. A., McManus, M. T., Sunyaev, S., Thomson, J. A., Tlsty, T. D., Tsai, L. H., Wang, W., Waterland, R. A., Zhang, M. Q., Chadwick, L. H., Bernstein, B. E., Costello, J. F., Ecker, J. R., Hirst, M., Meissner, A., Milosavljevic, A., Ren, B., Stamatoyannopoulos, J. A., Wang, T. & Kellis, M. Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317-330, doi:10.1038/nature14248 (2015).

173 Claussnitzer, M., Dankel, S. N., Klocke, B., Grallert, H., Glunk, V., Berulava, T., Lee, H., Oskolkov, N., Fadista, J., Ehlers, K., Wahl, S., Hoffmann, C., Qian, K., Ronn, T., Riess, H., Muller-Nurasyid, M., Bretschneider, N., Schroeder, T., Skurk, T., Horsthemke, B., Diagram+Consortium, Spieler, D., Klingenspor, M., Seifert, M., Kern, M. J., Mejhert, N., Dahlman, I., Hansson, O., Hauck, S. M., Bluher, M., Arner, P., Groop, L., Illig, T., Suhre, K., Hsu, Y. H., Mellgren, G., Hauner, H. & Laumen, H. Leveraging cross-species transcription factor binding site patterns: from diabetes risk loci to disease mechanisms. *Cell* **156**, 343-358, doi:10.1016/j.cell.2013.10.058 (2014).

174 Mikkelsen, T. S., Xu, Z., Zhang, X., Wang, L., Gimble, J. M., Lander, E. S. & Rosen, E. D. Comparative epigenomic analysis of murine and human adipogenesis. *Cell* **143**, 156-169, doi:10.1016/j.cell.2010.09.006 (2010).

175 Khetan, S., Kursawe, R., Youn, A., Lawlor, N., Jillette, A., Marquez, E. J., Ucar, D. & Stitzel, M. L. Type 2 Diabetes-Associated Genetic Variants Regulate Chromatin Accessibility in Human Islets. *Diabetes* **67**, 2466-2477, doi:10.2337/db18-0393 (2018).

176 Bulgakova, N. A. & Knust, E. The Crumbs complex: from epithelial-cell polarity to retinal degeneration. *J Cell Sci* **122**, 2587-2596, doi:10.1242/jcs.023648 (2009).

177 Dudok, J. J., Murtaza, M., Henrique Alves, C., Rashbass, P. & Wijnholds, J. Crumbs 2 prevents cortical abnormalities in mouse dorsal telencephalon. *Neurosci Res* **108**, 12-23, doi:10.1016/j.neures.2016.01.001 (2016).

178 Alves, C. H., Bossers, K., Vos, R. M., Essing, A. H., Swagemakers, S., van der Spek, P. J., Verhaagen, J. & Wijnholds, J. Microarray and morphological analysis of early postnatal CRB2 mutant retinas on a pure C57BL/6J genetic background. *PLoS One* **8**, e82532, doi:10.1371/journal.pone.0082532 (2013).

179 Jimenez-Amilburu, V. & Stainier, D. Y. R. The transmembrane protein Crb2a regulates cardiomyocyte apicobasal polarity and adhesion in zebrafish. *Development* **146**, doi:10.1242/dev.171207 (2019).

180 Slavotinek, A., Kaylor, J., Pierce, H., Cahr, M., DeWard, S. J., Schneidman-Duhovny, D., Alsadah, A., Salem, F., Schmajuk, G. & Mehta, L. CRB2 mutations produce a phenotype resembling congenital nephrosis, Finnish type, with cerebral ventriculomegaly and raised alpha-fetoprotein. *Am J Hum Genet* **96**, 162-169, doi:10.1016/j.ajhg.2014.11.013 (2015).

181 Dong, P. D., Provost, E., Leach, S. D. & Stainier, D. Y. Graded levels of Ptf1a differentially regulate endocrine and exocrine fates in the developing pancreas. *Genes Dev* **22**, 1445-1450, doi:10.1101/gad.1663208 (2008).

182 Kimmel, R. A., Dobler, S., Schmitner, N., Walsen, T., Freudenblum, J. & Meyer, D. Diabetic pdx1-mutant zebrafish show conserved responses to nutrient overload and anti-glycemic treatment. *Sci Rep* **5**, 14241, doi:10.1038/srep14241 (2015).

183 Field, H. A., Dong, P. D., Beis, D. & Stainier, D. Y. Formation of the digestive system in zebrafish. II. Pancreas morphogenesis. *Dev Biol* **261**, 197-208, doi:10.1016/s0012-1606(03)00308-7 (2003).

184 Pollard, S. M., Parsons, M. J., Kamei, M., Kettleborough, R. N., Thomas, K. A., Pham, V. N., Bae, M. K., Scott, A., Weinstein, B. M. & Stemple, D. L. Essential and overlapping roles for laminin alpha chains in notochord and blood vessel formation. *Dev Biol* **289**, 64-76, doi:10.1016/j.ydbio.2005.10.006 (2006).

185 Omori, Y. & Malicki, J. oko meduzy and related crumbs genes are determinants of apical cell features in the vertebrate embryo. *Curr Biol* **16**, 945-957, doi:10.1016/j.cub.2006.03.058 (2006).

186 Villar, D., Berthelot, C., Aldridge, S., Rayner, T. F., Lukk, M., Pignatelli, M., Park, T. J., Deaville, R., Erichsen, J. T., Jasinska, A. J., Turner, J. M., Bertelsen, M. F., Murchison, E. P., Flicek, P. & Odom, D. T. Enhancer evolution across 20 mammalian species. *Cell* **160**, 554-566, doi:10.1016/j.cell.2015.01.006 (2015).

187 Westmoreland, J. J., Kilic, G., Sartain, C., Sirma, S., Blain, J., Rehg, J., Harvey, N. & Sosa-Pineda, B. Pancreas-specific deletion of Prox1 affects development and disrupts homeostasis of the exocrine pancreas. *Gastroenterology* **142**, 999-1009 e1006, doi:10.1053/j.gastro.2011.12.007 (2012).

188 Heymans, C., Degosserie, J., Spourquet, C. & Pierreux, C. E. Pancreatic acinar differentiation is guided by differential laminin deposition. *Sci Rep* **9**, 2711, doi:10.1038/s41598-019-39077-6 (2019).

189 Nikolova, G., Jabs, N., Konstantinova, I., Domogatskaya, A., Tryggvason, K., Sorokin, L., Fassler, R., Gu, G., Gerber, H. P., Ferrara, N., Melton, D. A. & Lammert, E. The vascular basement membrane: a niche for insulin gene expression and Beta cell proliferation. *Dev Cell* **10**, 397-405, doi:10.1016/j.devcel.2006.01.015 (2006).

190 Varelas, X., Samavarchi-Tehrani, P., Narimatsu, M., Weiss, A., Cockburn, K., Larsen, B. G., Rossant, J. & Wrana, J. L. The Crumbs complex couples cell density sensing to Hippo-dependent control of the TGF-beta-SMAD pathway. *Dev Cell* **19**, 831-844, doi:10.1016/j.devcel.2010.11.012 (2010).

191 Horikoshi, M., Beaumont, R. N., Day, F. R., Warrington, N. M., Kooijman, M. N., Fernandez-Tajes, J., Feenstra, B., van Zuydam, N. R., Gaulton, K. J., Grarup, N., Bradfield, J. P., Strachan, D. P., Li-Gao, R., Ahluwalia, T. S., Kreiner, E., Rueedi, R., Lyytikainen, L. P., Cousminer, D. L., Wu, Y., Thiering, E., Wang, C. A., Have, C. T., Hottenga, J. J., Vilor-Tejedor, N., Joshi, P. K., Boh, E. T. H., Ntalla, I., Pitkanen, N., Mahajan, A., van Leeuwen, E. M., Joro, R., Lagou, V., Nodzenski, M., Diver, L. A., Zondervan, K. T., Bustamante, M., Marques-Vidal, P., Mercader, J. M., Bennett, A. J., Rahmioglu, N., Nyholt, D. R., Ma, R. C. W., Tam, C. H. T., Tam, W. H., Group, C. C. H. W., Ganesh, S. K., van Rooij, F. J., Jones, S. E., Loh, P. R., Ruth, K. S., Tuke, M. A., Tyrrell, J., Wood, A. R., Yaghootkar, H., Scholtens, D. M., Paternoster, L., Prokopenko, I., Kovacs, P., Atalay, M., Willems, S. M., Panoutsopoulou, K., Wang, X., Carstensen, L., Geller, F., Schraut, K. E., Murcia, M., van Beijsterveldt, C. E., Willemsen, G., Appel, E. V. R., Fonvig, C. E., Trier, C., Tiesler, C. M., Standl, M., Kutalik, Z., Bonas-Guarch, S., Hougaard, D. M., Sanchez, F., Torrents, D., Waage, J., Hollegaard, M. V., de Haan, H. G., Rosendaal, F. R., Medina-Gomez, C., Ring, S. M., Hemani, G., McMahon, G., Robertson, N. R., Groves, C. J., Langenberg, C., Luan, J., Scott, R. A., Zhao, J. H., Mentch, F. D., MacKenzie, S. M., Reynolds, R. M., Early Growth Genetics, C., Lowe, W. L., Jr., Tonjes, A., Stumvoll, M., Lindi, V., Lakka, T. A., van Duijn, C. M., Kiess, W., Korner, A., Sorensen, T. I., Niinikoski, H., Pahkala, K., Raitakari, O. T., Zeggini, E., Dedoussis, G. V., Teo, Y. Y., Saw, S. M., Melbye, M., Campbell, H., Wilson, J. F., Vrijheid, M., de Geus, E. J., Boomsma, D. I., Kadarmideen, H. N., Holm, J. C., Hansen, T., Sebert, S., Hattersley, A. T., Beilin, L. J., Newnham, J. P., Pennell, C. E., Heinrich, J., Adair, L. S., Borja, J. B., Mohlke, K. L., Eriksson, J. G., Widen, E. E., Kahonen, M., Viikari, J. S., Lehtimaki, T., Vollenweider, P., Bonnelykke, K., Bisgaard, H., Mook-Kanamori, D. O., Hofman, A., Rivadeneira, F., Uitterlinden, A. G., Pisinger, C., Pedersen, O., Power, C., Hypponen, E., Wareham, N. J., Hakonarson, H., Davies, E., Walker, B. R., Jaddoe, V. W., Jarvelin, M. R., Grant, S. F., Vaag, A. A., Lawlor, D. A., Frayling, T. M., Davey Smith, G., Morris, A. P., Ong, K. K., Felix, J. F., Timpson, N. J., Perry, J. R., Evans, D. M., McCarthy, M. I. & Freathy, R. M. Genome-wide associations for birth weight and correlations with adult disease. *Nature* **538**, 248-252, doi:10.1038/nature19806 (2016).

192 Godinho, L., Mumm, J. S., Williams, P. R., Schroeter, E. H., Koerber, A., Park, S. W., Leach, S. D. & Wong, R. O. Targeting of amacrine cell neurites to appropriate synaptic laminae in the developing zebrafish retina. *Development* **132**, 5069-5079, doi:10.1242/dev.02075 (2005).

193 Watanabe, K., Nishimura, Y., Oka, T., Nomoto, T., Kon, T., Shintou, T., Hirano, M., Shimada, Y., Umemoto, N., Kuroyanagi, J., Wang, Z., Zhang, Z., Nishimura, N., Miyazaki, T., Imamura, T. & Tanaka, T. In vivo imaging of zebrafish retinal cells using fluorescent coumarin derivatives. *BMC Neurosci* **11**, 116, doi:10.1186/1471-2202-11-116 (2010).

194 Dixon, J. R., Jung, I., Selvaraj, S., Shen, Y., Antosiewicz-Bourget, J. E., Lee, A. Y., Ye, Z., Kim, A., Rajagopal, N., Xie, W., Diao, Y., Liang, J., Zhao, H., Lobanenkov, V. V., Ecker, J. R., Thomson, J. A. & Ren, B. Chromatin architecture reorganization during stem cell differentiation. *Nature* **518**, 331-336, doi:10.1038/nature14222 (2015).

195 Bhandare, R., Schug, J., Le Lay, J., Fox, A., Smirnova, O., Liu, C., Naji, A. & Kaestner, K. H. Genome-wide analysis of histone modifications in human pancreatic islets. *Genome Res* **20**, 428-433, doi:10.1101/gr.102038.109 (2010).

196 Dupuis, J., Langenberg, C., Prokopenko, I., Saxena, R., Soranzo, N., Jackson, A. U., Wheeler, E., Glazer, N. L., Bouatia-Naji, N., Gloyn, A. L., Lindgren, C. M., Magi, R., Morris, A. P., Randall, J., Johnson, T., Elliott, P., Rybin, D., Thorleifsson, G., Steinthorsdottir, V., Henneman, P., Grallert, H., Dehghan, A., Hottenga, J. J., Franklin, C. S., Navarro, P., Song, K., Goel, A., Perry, J. R., Egan, J. M., Lajunen, T., Grarup, N., Sparso, T., Doney, A., Voight, B. F., Stringham, H. M., Li, M., Kanoni, S., Shrader, P., Cavalcanti-Proenca, C., Kumari, M., Qi, L., Timpson, N. J., Gieger, C., Zabena, C., Rocheleau, G., Ingelsson, E., An, P., O'Connell, J., Luan, J., Elliott, A., McCarroll, S. A., Payne, F., Roccasecca, R. M., Pattou, F., Sethupathy, P., Ardlie, K., Ariyurek, Y., Balkau, B., Barter, P., Beilby, J. P., Ben-Shlomo, Y., Benediktsson, R., Bennett, A. J., Bergmann, S., Bochud, M., Boerwinkle, E., Bonnefond, A., Bonnycastle, L. L., Borch-Johnsen, K., Bottcher, Y., Brunner, E., Bumpstead, S. J., Charpentier, G., Chen, Y. D., Chines, P., Clarke, R., Coin, L. J., Cooper, M. N., Cornelis, M., Crawford, G., Crisponi, L., Day, I. N., de Geus, E. J., Delplanque, J., Dina, C., Erdos, M. R., Fedson, A. C., Fischer-Rosinsky, A., Forouhi, N. G., Fox, C. S., Frants, R., Franzosi, M. G., Galan, P., Goodarzi, M. O., Graessler, J., Groves, C. J., Grundy, S., Gwilliam, R., Gyllensten, U., Hadjadj, S., Hallmans, G., Hammond, N., Han, X., Hartikainen, A. L., Hassanali, N., Hayward, C., Heath, S. C., Hercberg, S., Herder, C., Hicks, A. A., Hillman, D. R., Hingorani, A. D., Hofman, A., Hui, J., Hung, J., Isomaa, B., Johnson, P. R., Jorgensen, T., Jula, A., Kaakinen, M., Kaprio, J., Kesaniemi, Y. A., Kivimaki, M., Knight, B., Koskinen, S., Kovacs, P., Kyvik, K. O., Lathrop, G. M., Lawlor, D. A., Le Bacquer, O., Lecoeur, C., Li, Y., Lyssenko, V., Mahley, R., Mangino, M., Manning, A. K., Martinez-Larrad, M. T., McAteer, J. B., McCulloch, L. J., McPherson, R., Meisinger, C., Melzer, D., Meyre, D., Mitchell, B. D., Morken, M. A., Mukherjee, S., Naitza, S., Narisu, N., Neville, M. J., Oostra, B. A., Orru, M., Pakyz, R., Palmer, C. N., Paolisso, G., Pattaro, C., Pearson, D., Peden, J. F., Pedersen, N. L., Perola, M., Pfeiffer, A. F., Pichler, I., Polasek, O., Posthuma, D., Potter, S. C., Pouta, A., Province, M. A., Psaty, B. M., Rathmann, W., Rayner, N. W., Rice, K., Ripatti, S., Rivadeneira, F., Roden, M., Rolandsson, O., Sandbaek, A., Sandhu, M., Sanna, S., Sayer, A. A., Scheet, P., Scott, L. J., Seedorf, U., Sharp, S. J., Shields, B., Sigurethsson, G., Sijbrands, E. J., Silveira, A., Simpson, L., Singleton, A., Smith, N. L., Sovio, U., Swift, A., Syddall, H., Syvanen, A. C., Tanaka, T., Thorand, B., Tichet, J., Tonjes, A., Tuomi, T., Uitterlinden, A. G., van Dijk, K. W., van Hoek, M., Varma, D., Visvikis-Siest, S., Vitart, V., Vogelzangs, N., Waeber, G., Wagner, P. J., Walley, A., Walters, G. B., Ward, K. L., Watkins, H., Weedon, M. N., Wild, S. H., Willemsen, G., Witteman, J. C., Yarnell, J. W., Zeggini, E., Zelenika, D., Zethelius, B., Zhai, G., Zhao, J. H., Zillikens, M. C., Consortium, D.,

Consortium, G., Global, B. C., Borecki, I. B., Loos, R. J., Meneton, P., Magnusson, P. K., Nathan, D. M., Williams, G. H., Hattersley, A. T., Silander, K., Salomaa, V., Smith, G. D., Bornstein, S. R., Schwarz, P., Spranger, J., Karpe, F., Shuldiner, A. R., Cooper, C., Dedoussis, G. V., Serrano-Rios, M., Morris, A. D., Lind, L., Palmer, L. J., Hu, F. B., Franks, P. W., Ebrahim, S., Marmot, M., Kao, W. H., Pankow, J. S., Sampson, M. J., Kuusisto, J., Laakso, M., Hansen, T., Pedersen, O., Pramstaller, P. P., Wichmann, H. E., Illig, T., Rudan, I., Wright, A. F., Stumvoll, M., Campbell, H., Wilson, J. F., Anders Hamsten on behalf of Procardis, C., investigators, M., Bergman, R. N., Buchanan, T. A., Collins, F. S., Mohlke, K. L., Tuomilehto, J., Valle, T. T., Altshuler, D., Rotter, J. I., Siscovick, D. S., Penninx, B. W., Boomsma, D. I., Deloukas, P., Spector, T. D., Frayling, T. M., Ferrucci, L., Kong, A., Thorsteinsdottir, U., Stefansson, K., van Duijn, C. M., Aulchenko, Y. S., Cao, A., Scuteri, A., Schlessinger, D., Uda, M., Ruokonen, A., Jarvelin, M. R., Waterworth, D. M., Vollenweider, P., Peltonen, L., Mooser, V., Abecasis, G. R., Wareham, N. J., Sladek, R., Froguel, P., Watanabe, R. M., Meigs, J. B., Groop, L., Boehnke, M., McCarthy, M. I., Florez, J. C. & Barroso, I. New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. *Nat Genet* **42**, 105-116, doi:10.1038/ng.520 (2010).

197 Manning, A. K., Hivert, M. F., Scott, R. A., Grimsby, J. L., Bouatia-Naji, N., Chen, H., Rybin, D., Liu, C. T., Bielak, L. F., Prokopenko, I., Amin, N., Barnes, D., Cadby, G., Hottenga, J. J., Ingelsson, E., Jackson, A. U., Johnson, T., Kanoni, S., Ladenvall, C., Lagou, V., Lahti, J., Lecoeur, C., Liu, Y., Martinez-Larrad, M. T., Montasser, M. E., Navarro, P., Perry, J. R., Rasmussen-Torvik, L. J., Salo, P., Sattar, N., Shungin, D., Strawbridge, R. J., Tanaka, T., van Duijn, C. M., An, P., de Andrade, M., Andrews, J. S., Aspelund, T., Atalay, M., Aulchenko, Y., Balkau, B., Bandinelli, S., Beckmann, J. S., Beilby, J. P., Bellis, C., Bergman, R. N., Blangero, J., Boban, M., Boehnke, M., Boerwinkle, E., Bonnycastle, L. L., Boomsma, D. I., Borecki, I. B., Bottcher, Y., Bouchard, C., Brunner, E., Budimir, D., Campbell, H., Carlson, O., Chines, P. S., Clarke, R., Collins, F. S., Corbaton-Anchuelo, A., Couper, D., de Faire, U., Dedoussis, G. V., Deloukas, P., Dimitriou, M., Egan, J. M., Eiriksdottir, G., Erdos, M. R., Eriksson, J. G., Eury, E., Ferrucci, L., Ford, I., Forouhi, N. G., Fox, C. S., Franzosi, M. G., Franks, P. W., Frayling, T. M., Froguel, P., Galan, P., de Geus, E., Gigante, B., Glazer, N. L., Goel, A., Groop, L., Gudnason, V., Hallmans, G., Hamsten, A., Hansson, O., Harris, T. B., Hayward, C., Heath, S., Hercberg, S., Hicks, A. A., Hingorani, A., Hofman, A., Hui, J., Hung, J., Jarvelin, M. R., Jhun, M. A., Johnson, P. C., Jukema, J. W., Jula, A., Kao, W. H., Kaprio, J., Kardia, S. L., Keinanen-Kiukaanniemi, S., Kivimaki, M., Kolcic, I., Kovacs, P., Kumari, M., Kuusisto, J., Kyvik, K. O., Laakso, M., Lakka, T., Lannfelt, L., Lathrop, G. M., Launer, L. J., Leander, K., Li, G., Lind, L., Lindstrom, J., Lobbens, S., Loos, R. J., Luan, J., Lyssenko, V., Magi, R., Magnusson, P. K., Marmot, M., Meneton, P., Mohlke, K. L., Mooser, V., Morken, M. A., Miljkovic, I., Narisu, N., O'Connell, J., Ong, K. K., Oostra, B. A., Palmer, L. J., Palotie, A., Pankow, J. S., Peden, J. F., Pedersen, N. L., Pehlic, M., Peltonen, L., Penninx, B., Pericic, M., Perola, M., Perusse, L., Peyser, P. A., Polasek, O., Pramstaller, P. P., Province, M. A., Raikkonen, K., Rauramaa, R., Rehnberg, E., Rice, K., Rotter, J. I., Rudan, I., Ruokonen, A., Saaristo, T., Sabater-Lleal, M., Salomaa, V., Savage, D. B., Saxena, R., Schwarz, P., Seedorf, U., Sennblad, B., Serrano-Rios, M., Shuldiner, A. R., Sijbrands, E. J., Siscovick, D. S., Smit, J. H., Small, K. S., Smith, N. L., Smith, A. V., Stancakova, A., Stirrups, K., Stumvoll, M., Sun, Y. V., Swift, A. J., Tonjes, A., Tuomilehto, J., Trompet, S., Uitterlinden, A. G., Uusitupa, M., Vikstrom, M., Vitart, V., Vohl, M. C., Voight, B. F., Vollenweider, P., Waeber, G., Waterworth, D. M., Watkins, H., Wheeler, E., Widen, E., Wild, S. H., Willems, S. M., Willemsen, G., Wilson, J. F., Witteman, J. C., Wright, A. F., Yaghootkar, H., Zelenika, D.,

Zemunik, T., Zgaga, L., Replication, D. I. G., Meta-analysis, C., Multiple Tissue Human Expression Resource, C., Wareham, N. J., McCarthy, M. I., Barroso, I., Watanabe, R. M., Florez, J. C., Dupuis, J., Meigs, J. B. & Langenberg, C. A genome-wide approach accounting for body mass index identifies genetic variants influencing fasting glycemic traits and insulin resistance. *Nat Genet* **44**, 659-669, doi:10.1038/ng.2274 (2012).

198 Strawbridge, R. J., Dupuis, J., Prokopenko, I., Barker, A., Ahlqvist, E., Rybin, D., Petrie, J. R., Travers, M. E., Bouatia-Naji, N., Dimas, A. S., Nica, A., Wheeler, E., Chen, H., Voight, B. F., Taneera, J., Kanoni, S., Peden, J. F., Turrini, F., Gustafsson, S., Zabena, C., Almgren, P., Barker, D. J., Barnes, D., Dennison, E. M., Eriksson, J. G., Eriksson, P., Eury, E., Folkersen, L., Fox, C. S., Frayling, T. M., Goel, A., Gu, H. F., Horikoshi, M., Isomaa, B., Jackson, A. U., Jameson, K. A., Kajantie, E., Kerr-Conte, J., Kuulasmaa, T., Kuusisto, J., Loos, R. J., Luan, J., Makrilakis, K., Manning, A. K., Martinez-Larrad, M. T., Narisu, N., Nastase Mannila, M., Ohrvik, J., Osmond, C., Pascoe, L., Payne, F., Sayer, A. A., Sennblad, B., Silveira, A., Stancakova, A., Stirrups, K., Swift, A. J., Syvanen, A. C., Tuomi, T., van 't Hooft, F. M., Walker, M., Weedon, M. N., Xie, W., Zethelius, B., Consortium, D., Consortium, G., Mu, T. C., Consortium, C. A., Consortium, C. D., Ongen, H., Malarstig, A., Hopewell, J. C., Saleheen, D., Chambers, J., Parish, S., Danesh, J., Kooner, J., Ostenson, C. G., Lind, L., Cooper, C. C., Serrano-Rios, M., Ferrannini, E., Forsen, T. J., Clarke, R., Franzosi, M. G., Seedorf, U., Watkins, H., Froguel, P., Johnson, P., Deloukas, P., Collins, F. S., Laakso, M., Dermitzakis, E. T., Boehnke, M., McCarthy, M. I., Wareham, N. J., Groop, L., Pattou, F., Gloyn, A. L., Dedoussis, G. V., Lyssenko, V., Meigs, J. B., Barroso, I., Watanabe, R. M., Ingelsson, E., Langenberg, C., Hamsten, A. & Florez, J. C. Genome-wide association identifies nine common variants associated with fasting proinsulin levels and provides new insights into the pathophysiology of type 2 diabetes. *Diabetes* **60**, 2624-2634, doi:10.2337/db11-0415 (2011).

199 Saxena, R., Hivert, M. F., Langenberg, C., Tanaka, T., Pankow, J. S., Vollenweider, P., Lyssenko, V., Bouatia-Naji, N., Dupuis, J., Jackson, A. U., Kao, W. H., Li, M., Glazer, N. L., Manning, A. K., Luan, J., Stringham, H. M., Prokopenko, I., Johnson, T., Grarup, N., Boesgaard, T. W., Lecoeur, C., Shrader, P., O'Connell, J., Ingelsson, E., Couper, D. J., Rice, K., Song, K., Andreasen, C. H., Dina, C., Kottgen, A., Le Bacquer, O., Pattou, F., Taneera, J., Steinthorsdottir, V., Rybin, D., Ardlie, K., Sampson, M., Qi, L., van Hoek, M., Weedon, M. N., Aulchenko, Y. S., Voight, B. F., Grallert, H., Balkau, B., Bergman, R. N., Bielinski, S. J., Bonnefond, A., Bonnycastle, L. L., Borch-Johnsen, K., Bottcher, Y., Brunner, E., Buchanan, T. A., Bumpstead, S. J., Cavalcanti-Proenca, C., Charpentier, G., Chen, Y. D., Chines, P. S., Collins, F. S., Cornelis, M., G, J. C., Delplanque, J., Doney, A., Egan, J. M., Erdos, M. R., Firmann, M., Forouhi, N. G., Fox, C. S., Goodarzi, M. O., Graessler, J., Hingorani, A., Isomaa, B., Jorgensen, T., Kivimaki, M., Kovacs, P., Krohn, K., Kumari, M., Lauritzen, T., Levy-Marchal, C., Mayor, V., McAteer, J. B., Meyre, D., Mitchell, B. D., Mohlke, K. L., Morken, M. A., Narisu, N., Palmer, C. N., Pakyz, R., Pascoe, L., Payne, F., Pearson, D., Rathmann, W., Sandbaek, A., Sayer, A. A., Scott, L. J., Sharp, S. J., Sijbrands, E., Singleton, A., Siscovick, D. S., Smith, N. L., Sparso, T., Swift, A. J., Syddall, H., Thorleifsson, G., Tonjes, A., Tuomi, T., Tuomilehto, J., Valle, T. T., Waeber, G., Walley, A., Waterworth, D. M., Zeggini, E., Zhao, J. H., consortium, G., investigators, M., Illig, T., Wichmann, H. E., Wilson, J. F., van Duijn, C., Hu, F. B., Morris, A. D., Frayling, T. M., Hattersley, A. T., Thorsteinsdottir, U., Stefansson, K., Nilsson, P., Syvanen, A. C., Shuldiner, A. R., Walker, M., Bornstein, S. R., Schwarz, P., Williams, G. H., Nathan, D. M., Kuusisto, J., Laakso, M., Cooper, C., Marmot, M., Ferrucci, L., Mooser, V., Stumvoll, M., Loos, R. J., Altshuler, D., Psaty, B. M., Rotter, J.

I., Boerwinkle, E., Hansen, T., Pedersen, O., Florez, J. C., McCarthy, M. I., Boehnke, M., Barroso, I., Sladek, R., Froguel, P., Meigs, J. B., Groop, L., Wareham, N. J. & Watanabe, R. M. Genetic variation in GIPR influences the glucose and insulin responses to an oral glucose challenge. *Nat Genet* **42**, 142-148, doi:10.1038/ng.521 (2010).

200 Wood, A. R., Jonsson, A., Jackson, A. U., Wang, N., van Leewen, N., Palmer, N. D., Kobes, S., Deelen, J., Boquete-Vilarino, L., Paananen, J., Stancakova, A., Boomsma, D. I., de Geus, E. J. C., Eekhoff, E. M. W., Fritsche, A., Kramer, M., Nijpels, G., Simonis-Bik, A., van Haeften, T. W., Mahajan, A., Boehnke, M., Bergman, R. N., Tuomilehto, J., Collins, F. S., Mohlke, K. L., Banasik, K., Groves, C. J., McCarthy, M. I., Diabetes Research on Patient, S., Pearson, E. R., Natali, A., Mari, A., Buchanan, T. A., Taylor, K. D., Xiang, A. H., Gjesing, A. P., Grarup, N., Eiberg, H., Pedersen, O., Chen, Y. D., Laakso, M., Norris, J. M., Smith, U., Wagenknecht, L. E., Baier, L., Bowden, D. W., Hansen, T., Walker, M., Watanabe, R. M., t Hart, L. M., Hanson, R. L. & Frayling, T. M. A Genome-Wide Association Study of IVGTT-Based Measures of First-Phase Insulin Secretion Refines the Underlying Physiology of Type 2 Diabetes Variants. *Diabetes* **66**, 2296-2309, doi:10.2337/db16-1452 (2017).

201 Taal, H. R., Pourcain, B. S., Thiering, E., Das, S., Mook-Kanamori, D. O., Warrington, N. M., Kaakinen, M., Kreiner-Moller, E., Bradfield, J. P., Freathy, R. M., Geller, F., Guxens, M., Cousminer, D. L., Kerkhof, M., Timpson, N. J., Ikram, M. A., Beilin, L. J., Bonnelykke, K., Buxton, J. L., Charoen, P., Chawes, B. L. K., Eriksson, J., Evans, D. M., Hofman, A., Kemp, J. P., Kim, C. E., Klopp, N., Lahti, J., Lye, S. J., McMahon, G., Mentch, F. D., Muller, M., O'Reilly, P. F., Prokopenko, I., Rivadeneira, F., Steegers, E. A. P., Sunyer, J., Tiesler, C., Yaghootkar, H., Cohorts for, H., Aging Research in Genetic Epidemiology, C., Breteler, M. M. B., Debette, S., Fornage, M., Gudnason, V., Launer, L. J., van der Lugt, A., Mosley, T. H., Seshadri, S., Smith, A. V., Vernooij, M. W., Early, G., Lifecourse Epidemiology, c., Blakemore, A. I., Chiavacci, R. M., Feenstra, B., Fernandez-Benet, J., Grant, S. F. A., Hartikainen, A. L., van der Heijden, A. J., Iniguez, C., Lathrop, M., McArdle, W. L., Molgaard, A., Newnham, J. P., Palmer, L. J., Palotie, A., Pouta, A., Ring, S. M., Sovio, U., Standl, M., Uitterlinden, A. G., Wichmann, H. E., Vissing, N. H., DeCarli, C., van Duijn, C. M., McCarthy, M. I., Koppelman, G. H., Estivill, X., Hattersley, A. T., Melbye, M., Bisgaard, H., Pennell, C. E., Widen, E., Hakonarson, H., Smith, G. D., Heinrich, J., Jarvelin, M. R., Early Growth Genetics, C. & Jaddoe, V. W. V. Common variants at 12q15 and 12q24 are associated with infant head circumference. *Nat Genet* **44**, 532-538, doi:10.1038/ng.2238 (2012).

202 van der Valk, R. J., Kreiner-Moller, E., Kooijman, M. N., Guxens, M., Stergiakouli, E., Saaf, A., Bradfield, J. P., Geller, F., Hayes, M. G., Cousminer, D. L., Korner, A., Thiering, E., Curtin, J. A., Myhre, R., Huikari, V., Joro, R., Kerkhof, M., Warrington, N. M., Pitkanen, N., Ntalla, I., Horikoshi, M., Veijola, R., Freathy, R. M., Teo, Y. Y., Barton, S. J., Evans, D. M., Kemp, J. P., St Pourcain, B., Ring, S. M., Davey Smith, G., Bergstrom, A., Kull, I., Hakonarson, H., Mentch, F. D., Bisgaard, H., Chawes, B., Stokholm, J., Waage, J., Eriksen, P., Sevelsted, A., Melbye, M., Early, G., Lifecourse Epidemiology, C., van Duijn, C. M., Medina-Gomez, C., Hofman, A., de Jongste, J. C., Taal, H. R., Uitterlinden, A. G., Genetic Investigation of, A. T. C., Armstrong, L. L., Eriksson, J., Palotie, A., Bustamante, M., Estivill, X., Gonzalez, J. R., Llop, S., Kiess, W., Mahajan, A., Flexeder, C., Tiesler, C. M., Murray, C. S., Simpson, A., Magnus, P., Sengpiel, V., Hartikainen, A. L., Keinanen-Kiukaanniemi, S., Lewin, A., Da Silva Couto Alves, A., Blakemore, A. I., Buxton, J. L., Kaakinen, M., Rodriguez, A., Sebert, S., Vaarasmaki, M., Lakka, T., Lindi, V., Gehring, U., Postma, D. S., Ang, W., Newnham, J. P., Lyytikainen,

L. P., Pahkala, K., Raitakari, O. T., Panoutsopoulou, K., Zeggini, E., Boomsma, D. I., Groen-Blokhuis, M., Ilonen, J., Franke, L., Hirschhorn, J. N., Pers, T. H., Liang, L., Huang, J., Hocher, B., Knip, M., Saw, S. M., Holloway, J. W., Melen, E., Grant, S. F., Feenstra, B., Lowe, W. L., Widen, E., Sergeyev, E., Grallert, H., Custovic, A., Jacobsson, B., Jarvelin, M. R., Atalay, M., Koppelman, G. H., Pennell, C. E., Niinikoski, H., Dedoussis, G. V., McCarthy, M. I., Frayling, T. M., Sunyer, J., Timpson, N. J., Rivadeneira, F., Bonnelykke, K., Jaddoe, V. W. & Early Growth Genetics, C. A novel common variant in DCST2 is associated with length in early life and height in adulthood. *Hum Mol Genet* **24**, 1155-1168, doi:10.1093/hmg/ddu510 (2015).

203 Lancman, J. J., Zvenigorodsky, N., Gates, K. P., Zhang, D., Solomon, K., Humphrey, R. K., Kuo, T., Setiawan, L., Verkade, H., Chi, Y. I., Jhala, U. S., Wright, C. V., Stainier, D. Y. & Dong, P. D. Specification of hepatopancreas progenitors in zebrafish by hnf1ba and wnt2bb. *Development* **140**, 2669-2679, doi:10.1242/dev.090993 (2013).

204 Hsu, Y. C. & Jensen, A. M. Multiple domains in the Crumbs Homolog 2a (Crb2a) protein are required for regulating rod photoreceptor size. *BMC Cell Biol* **11**, 60, doi:10.1186/1471-2121-11-60 (2010).

205 Mathers, C. D. & Loncar, D. Projections of global mortality and burden of disease from 2002 to 2030. *PLoS Med* **3**, e442, doi:10.1371/journal.pmed.0030442 (2006).

206 Todd, J. A., Bell, J. I. & McDevitt, H. O. HLA-DQ beta gene contributes to susceptibility and resistance to insulin-dependent diabetes mellitus. *Nature* **329**, 599-604, doi:10.1038/329599a0 (1987).

207 Stratmann, T., Apostolopoulos, V., Mallet-Designe, V., Corper, A. L., Scott, C. A., Wilson, I. A., Kang, A. S. & Teyton, L. The I-A^{g7} MHC Class II Molecule Linked to Murine Diabetes Is a Promiscuous Peptide Binder. *The Journal of Immunology* **165**, 3214-3225, doi:10.4049/jimmunol.165.6.3214 (2000).

208 Mohan, J. F., Levisetti, M. G., Calderon, B., Herzog, J. W., Petzold, S. J. & Unanue, E. R. Unique autoreactive T cells recognize insulin peptides generated within the islets of Langerhans in autoimmune diabetes. *Nat Immunol* **11**, 350-U332, doi:10.1038/ni.1850 (2010).

209 Pujol-Borrell, R., Todd, I., Doshi, M., Bottazzo, G. F., Sutton, R., Gray, D., Adolf, G. R. & Feldmann, M. HLA class II induction in human islet cells by interferon-gamma plus tumour necrosis factor or lymphotoxin. *Nature* **326**, 304-306, doi:10.1038/326304a0 (1987).

210 Bottazzo, G. F., Dean, B. M., McNally, J. M., MacKay, E. H., Swift, P. G. & Gamble, D. R. In situ characterization of autoimmune phenomena and expression of HLA molecules in the pancreas in diabetic insulitis. *N Engl J Med* **313**, 353-360, doi:10.1056/NEJM198508083130604 (1985).

211 Foulis, A. K., Farquharson, M. A. & Hardman, R. Aberrant expression of class II major histocompatibility complex molecules by B cells and hyperexpression of class I major histocompatibility complex molecules by insulin containing islets in type 1 (insulin-dependent) diabetes mellitus. *Diabetologia* **30**, 333-343, doi:10.1007/BF00299027 (1987).

212 Campbell, I. L., Oxbrow, L., West, J. & Harrison, L. C. Regulation of MHC protein expression in pancreatic beta-cells by interferon-gamma and tumor necrosis factor-alpha. *Mol Endocrinol* **2**, 101-107, doi:10.1210/mend-2-2-101 (1988).

213 Russell, M. A., Redick, S. D., Blodgett, D. M., Richardson, S. J., Leete, P., Krogvold, L., Dahl-Jorgensen, K., Bottino, R., Brissova, M., Spaeth, J. M., Babon, J. A. B., Haliyur, R., Powers, A. C., Yang, C., Kent, S. C., Derr, A. G., Kucukural, A., Garber, M. G., Morgan, N. G. & Harlan, D. M. 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).

214 Zhao, Y., Scott, N. A., Quah, H. S., Krishnamurthy, B., Bond, F., Loudovaris, T., Mannering, S. I., Kay, T. W. & Thomas, H. E. Mouse pancreatic beta cells express MHC class II and stimulate CD4(+) T cells to proliferate. *Eur J Immunol* **45**, 2494-2503, doi:10.1002/eji.201445378 (2015).

215 Tsonkova, V. G., Sand, F. W., Wolf, X. A., Grunnet, L. G., Kirstine Ringgaard, A., Ingvorsen, C., Winkel, L., Kalisz, M., Dalgaard, K., Bruun, C., Fels, J. J., Helgstrand, C., Hastrup, S., Oberg, F. K., Vernet, E., Sandrini, M. P. B., Shaw, A. C., Jessen, C., Gronborg, M., Hald, J., Willenbrock, H., Madsen, D., Wernersson, R., Hansson, L., Jensen, J. N., Plesner, A., Alanentalo, T., Petersen, M. B. K., Grapin-Botton, A., Honore, C., Ahnfelt-Ronne, J., Hecksher-Sorensen, J., Ravassard, P., Madsen, O. D., Rescan, C. & Frogne, T. The EndoC-betaH1 cell line is a valid model of human beta cells and applicable for screenings to identify novel drug target candidates. *Mol Metab* **8**, 144-157, doi:10.1016/j.molmet.2017.12.007 (2018).

216 Michels, A. W., Landry, L. G., McDaniel, K. A., Yu, L., Campbell-Thompson, M., Kwok, W. W., Jones, K. L., Gottlieb, P. A., Kappler, J. W., Tang, Q., Roep, B. O., Atkinson, M. A., Mathews, C. E. & Nakayama, M. Islet-Derived CD4 T Cells Targeting Proinsulin in Human Autoimmune Diabetes. *Diabetes* **66**, 722-734, doi:10.2337/db16-1025 (2017).

217 So, M., Elso, C. M., Tresoldi, E., Pakusch, M., Pathiraja, V., Wentworth, J. M., Harrison, L. C., Krishnamurthy, B., Thomas, H. E., Rodda, C., Cameron, F. J., McMahon, J., Kay, T. W. H. & Mannering, S. I. Proinsulin C-peptide is an autoantigen in people with type 1 diabetes. *Proc Natl Acad Sci U S A* **115**, 10732-10737, doi:10.1073/pnas.1809208115 (2018).

218 Sang, Y., Yan, C., Zhu, C. & Ni, G. Relationship between HLA-DRB1 and DQ alleles and the genetic susceptibility to type 1 diabetes. *Chin Med J (Engl)* **114**, 407-409 (2001).

219 Nordmann, T. M., Dror, E., Schulze, F., Traub, S., Berishvili, E., Barbieux, C., Boni-Schnetzler, M. & Donath, M. Y. The Role of Inflammation in beta-cell Dedifferentiation. *Sci Rep* **7**, 6285, doi:10.1038/s41598-017-06731-w (2017).

220 Leite, N. C., Sintov, E., Meissner, T. B., Brehm, M. A., Greiner, D. L., Harlan, D. M. & Melton, D. A. Modeling Type 1 Diabetes In Vitro Using Human Pluripotent Stem Cells. *Cell Rep* **32**, 107894, doi:10.1016/j.celrep.2020.107894 (2020).

221 Balboa, D., Barsby, T., Lithovius, V., Saarimäki-Vire, J., Omar-Hmeadi, M., Dyachok, O., Montaser, H., Lund, P.-E., Yang, M., Ibrahim, H., Näätänen, A., Chandra, V., Vihinen,

H., Jokitalo, E., Kvist, J., Ustinov, J., Nieminen, A. I., Kuuluvainen, E., Hietakangas, V., Katajisto, P., Lau, J., Carlsson, P.-O., Barg, S., Tengholm, A. & Otonkoski, T. Functional, metabolic and transcriptional maturation of stem cell derived beta cells. *bioRxiv*, 2021.2003.2031.437748, doi:10.1101/2021.03.31.437748 (2021).

222 Wong, W. L., Brostrom, M. A., Kuznetsov, G., Gmitter-Yellen, D. & Brostrom, C. O. Inhibition of protein synthesis and early protein processing by thapsigargin in cultured cells. *Biochem J* **289 (Pt 1)**, 71-79, doi:10.1042/bj2890071 (1993).

223 Cantrell, D. A. T-cell antigen receptor signal transduction. *Immunology* **105**, 369-374, doi:10.1046/j.1365-2567.2002.01391.x (2002).

224 Kim, K. H., Kim, S. H., Jung, H. H., Moon, J. H., Jeong, S. U., Yu, K. & Lee, C. K. Thapsigargin Increases IL-2 Production in T Cells at Nanomolar Concentrations. *Immune Netw* **18**, e26, doi:10.4110/in.2018.18.e26 (2018).