Cell Metabolism
Article

Human β Cell Transcriptome Analysis Uncovers lncRNAs That Are Tissue-Specific, Dynamically Regulated, and Abnormally Expressed in Type 2 Diabetes

Ignasi Morán,[1,19] İldem Akerman,[1,19] Martijn van de Bunt,[3,4] Ruiyu Xie,[5] Marion Benazra,[6] Takao Nammo,[1,2,7] Luis Arnes,[8] Nikolina Nakić,[1,2] Javier García-Hurtado,[1,2] Santiago Rodríguez-Seguí,[1,2] Lorenzo Pasquali,[1,2] Claire Sauty-Colace,[6] Anthony Beucher,[1,2] Raphael Scharfmann,[9] Joris van Arensbergen,[1,2] Paul R. Johnson,[4,10,16] Andrew Berry,[11] Clarence Lee,[12] Timothy Harkins,[12] Valery Gmyr,[13] François Pattou,[13] Julie Kerr-Conte,[13] Lorenzo Piemonti,[14] Thierry Berney,[15] Neil Hanley,[11] Anna L. Gloyn,[4,16] Lori Sussel,[8] Linda Langman,[17] Kenneth L. Brayman,[17] Maike Sander,[5] Mark I. McCarthy,[3,4,14] Philippe Ravassard,[6] and Jorge Ferrer[1,2,18,*]

[1]Genomic Programming of Beta Cells Laboratory, Institut d'Investigacions Biomèdiques August Pi I Sunyer (IDIBAPS), 08036 Barcelona, Spain
[2]CIBER de Diabetes y Enfermedades Metabólicas Asociadas (CIBERDEM), 08036 Barcelona, Spain
[3]The Wellcome Trust Centre for Human Genetics, University of Oxford, OX1 2JD Oxford, UK
[4]Oxford Centre for Diabetes, Endocrinology and Metabolism, Churchill Hospital, OX3 7LJ Oxford, UK
[5]Department of Cellular and Molecular Medicine, University of California at San Diego, La Jolla, CA 92093, USA
[6]Centre de recherche de l'institut du cerveau et de la moelle, Biotechnology & Biotherapy team, CNRS UMR7225; INSERM U975; University Pierre et Marie Curie, 75005 Paris, France
[7]Department of Metabolic Disorders, Diabetes Research Center, Research Institute, National Center for Global Health and Medicine, Shinjuku-ku, 162-8655 Tokyo, Japan
[8]Department of Genetics and Development, Russ Berrie Medical Pavilion, Columbia University, New York, NY 10032, USA
[9]Institut National de la Santé et de la Recherche Médicale (INSERM) U845, Research Center Growth and Signalling, Paris Descartes University, Sorbonne Paris Cité, Necker Hospital, 75001 Paris, France
[10]Oxford Islet Transplant Programme, Nuffield Department of Surgical Sciences, John Radcliffe Hospital, OX3 9DU Oxford, UK
[11]Developmental Biomedicine Research Group, School of Biomedicine, Manchester Academic Health Sciences Centre, University of Manchester and Central Manchester University Hospitals NHS Foundation Trust, M13 9WL Manchester, UK
[12]Genome Sequencing Collaborations Group, Life Technologies, Beverly, MA 01915, USA
[13]University of Lille Nord de France, INSERM U859 Biotherapies of Diabete, Lille, France
[14]Diabetes research institute (HSR-DRI), San Raffaele Scientific Institute, 20132 Milano, Italy
[15]Cell Isolation and Transplantation Center, 1121 Geneva 4, Switzerland
[16]Oxford NIHR Biomedical Research Centre, Churchill Hospital, OX3 7LJ Oxford, UK
[17]Division of Transplantation, Department of Surgery, Center for Cellular Therapy and Biotherapeutics, University of Virginia, Charlottesville, VA 22904, USA
[18]Department of Endocrinology and Nutrition, Hospital Clínic de Barcelona, 170 08036 Barcelona, Spain
[19]These authors contributed equally to this work
*Correspondence: jferrer@clinic.ub.es
http://dx.doi.org/10.1016/j.cmet.2012.08.010

SUMMARY

A significant portion of the genome is transcribed as long noncoding RNAs (lncRNAs), several of which are known to control gene expression. The repertoire and regulation of lncRNAs in disease-relevant tissues, however, has not been systematically explored. We report a comprehensive strand-specific transcriptome map of human pancreatic islets and β cells, and uncover >1100 intergenic and antisense islet-cell lncRNA genes. We find islet lncRNAs that are dynamically regulated and show that they are an integral component of the β cell differentiation and maturation program. We sequenced the mouse islet transcriptome and identify lncRNA orthologs that are regulated like their human counterparts. Depletion of *HI-LNC25,* a β cell-specific lncRNA, downregulated *GLIS3* mRNA, thus exemplifying a gene regulatory function of islet lncRNAs. Finally, selected islet lncRNAs were dysregulated in type 2 diabetes or mapped to genetic loci underlying diabetes susceptibility. These findings reveal a new class of islet-cell genes relevant to β cell programming and diabetes pathophysiology.

INTRODUCTION

During recent years, it has become apparent that the genomes of species as diverse as zebrafish, mice, and humans transcribe thousands of RNAs that do not encode for proteins (Bertone et al., 2004; Birney et al., 2007; Carninci et al., 2005; Guttman et al., 2009; Ulitsky et al., 2011). A subset of noncoding transcripts are larger than 200 nucleotides and are known as long noncoding RNAs (lncRNAs) (Mattick and Makunin, 2006). The function of most lncRNAs remains unknown. However, several dozen lncRNAs are known to exert nonredundant roles in processes such as X inactivation, imprinting, splicing, transcriptional regulation, pluripotency, cancer, cell cycle, or survival

(Gupta et al., 2010; Guttman et al., 2011; Hu et al., 2011; Penny et al., 1996; Rinn et al., 2007; Sleutels et al., 2002). In one example, a lncRNA has been shown to promote reprogramming of pluripotent cells from somatic cells (Loewer et al., 2010). Available evidence thus indicates that lncRNAs represent a still poorly understood layer of gene regulation.

Numerous mammalian lncRNAs are expressed in a cell-type specific manner (Cabili et al., 2011; Mercer et al., 2008). Together with knowledge that several such transcripts are functional, this raises the intriguing possibility that lncRNAs could be previously unsuspected mediators of lineage-specific differentiation or specialized cellular functions. Defects in lncRNAs could thus underlie human disease, and cell-specific regulatory lncRNAs might provide therapeutic targets. This warrants the need to explore the repertoires of lncRNAs of disease-relevant cell types and tissues.

Pancreatic islets of Langerhans are an excellent model of a specialized tissue that is closely linked to human disease. Islets comprise insulin-secreting β cells and other polypeptide hormone-producing cells, including glucagon-secreting α-cells. Islet-cell dysfunction is central to the pathophysiology of type 2 diabetes (T2D), the most prevalent form of diabetes (Bell and Polonsky, 2001). Recent genome-wide association studies for T2D and related traits have revealed >50 susceptibility loci, most of which are not known to carry variants that alter protein-coding sequences (McCarthy, 2010). A common hypothesis is that such variants impact regulatory elements of protein-coding genes, although they could equally affect other non-protein-coding elements such as lncRNAs.

In type 1 diabetes, β cells are destroyed by autoimmune mechanisms, and consequently several experimental approaches are being developed to replace destroyed cells (Halban et al., 2001). One approach is based on the recent discovery of β cell transcription factors, some of which have been misexpressed in somatic cells to create insulin-expressing cells (Collombat et al., 2009; Zhou et al., 2008). Another approach is to derive β cells from pluripotent cells (Kroon et al., 2008). However, existing strategies have not yet succeeded in generating fully functional therapeutic β cells in vitro. Clearly, the identification of novel regulators of β cell differentiation and maturation remains a major challenge, and islet-specific regulatory transcripts are logical targets for this effort. Despite the potential implications for human diabetes, information on islet-cell lncRNAs is lacking.

In this study, we integrated sequence-based transcriptome and chromatin maps of human islets and β cells to define 1128 islet lncRNA genes. We show that these lncRNAs are an integral component of the dynamic β cell-specific differentiation program, suggesting a role as biomarkers and potential regulators for programming efforts. We extend existing knowledge on lncRNAs by disclosing orthologous mouse transcripts that are regulated in an evolutionarily conserved manner. We focused on one islet-specific lncRNA and show that it impacts the expression of a regulatory target. Finally, we show dysregulation of islet lncRNAs in T2D and map selected lncRNAs to human diabetes genetic susceptibility loci. Collectively, these studies describe a new class of islet genes and provide diverse lines of evidence to suggest that islet lncRNAs may impact diabetes pathophysiology and efforts to program therapeutic β cells.

RESULTS

High-Resolution Maps of Active Human Pancreatic Islet Cell Genes

To identify human β cell lncRNAs we integrated transcriptional and chromatin maps of purified human islet cells (Figure 1A). We generated directional and nondirectional cDNA libraries from PolyA+ and Ribo− RNA fractions of six human islets and two FACS-purified β cell samples, and we collected >460 million uniquely mapped paired-end sequence reads (Table S1). To increase our ability to define the structure of novel genes, we mapped epigenetic gene landmarks in three human islet samples, including H3K4me3, which is enriched in active promoter regions (Table S1). These data sets were then integrated to build strand-specific models of active human islet-cell genes with their predominant splicing patterns (Figure 1A).

As a measure of quality, we first examined annotated genes and recovered ~70% as expressed above a threshold level of 0.5 reads per million mapped reads (RPKM) in islet cells. Despite the fact that the samples originated from diverse human islet donors (Table S2), both mRNA and H3K4me3 levels showed a high correlation between the different human islet samples ($r^2 > 0.94$; Figures S1A and S1B). Consistent with the high islet purity, the most abundant H3K4me3-positive transcripts included known islet-cell and βcell genes (Figure 1B, Table S3). Remarkably, ~20% of exonic reads in islets and ~45% in β cells originated from the Proinsulin mRNA (Figure 1B). Moreover, known islet-specific genes displayed coherent chromatin enrichment patterns (see, for example, Figures 1C and S1C).

Genomic inventories of human islets are unavoidably confounded by signals from contaminant exocrine cells. Our approach, however, allowed us to define lncRNAs that are truly expressed in β cells and other islet cell types. Transcripts from few contaminating acinar cells were efficiently excluded because they lacked H3K4me3 enrichment in islet preparations (Figures S1D and S1H, Table S3). Accordingly, <1% of genes that we defined as active in islet cells were contaminant acinar transcripts, on the basis of a relative enrichment in acinar versus islet-cell RNA. This initial analysis thus validated our strategy to map bona-fide active genes in human islets and β cells.

Human Islet Cells Transcribe Intergenic and Antisense lncRNAs

Over 19% of the transcribed genome in human islets mapped outside of annotated protein-coding genes (Figure S1E). To define discrete islet-cell lncRNA genes, we selected 1128 transcripts with the following properties: (1) length >200 bp; (2) H3K4me3 enrichment in a coherent location relative to the transcribed strand; (3) expression >0.5 RPKM in all five islet and β cell PolyA+ samples; (4) no splicing or overlap with any coding gene present in RefSeq, UCSC, or Ensembl annotations; and (5) low protein-coding potential. Of these, 761 were "antisense" (AS) lncRNAs, located <1 kb from an annotated gene but in a divergent orientation (Figure 1D). Thirty-two were <1 kb from an annotated gene but in a convergent orientation. Another set of 335 were "intergenic" (IG) lncRNAs, located >1 kb from any coding gene. Finally, we identified 55 annotated lncRNAs that were located within the boundaries of coding genes and termed these "overlapping antisense" lncRNAs (Figure 1D).

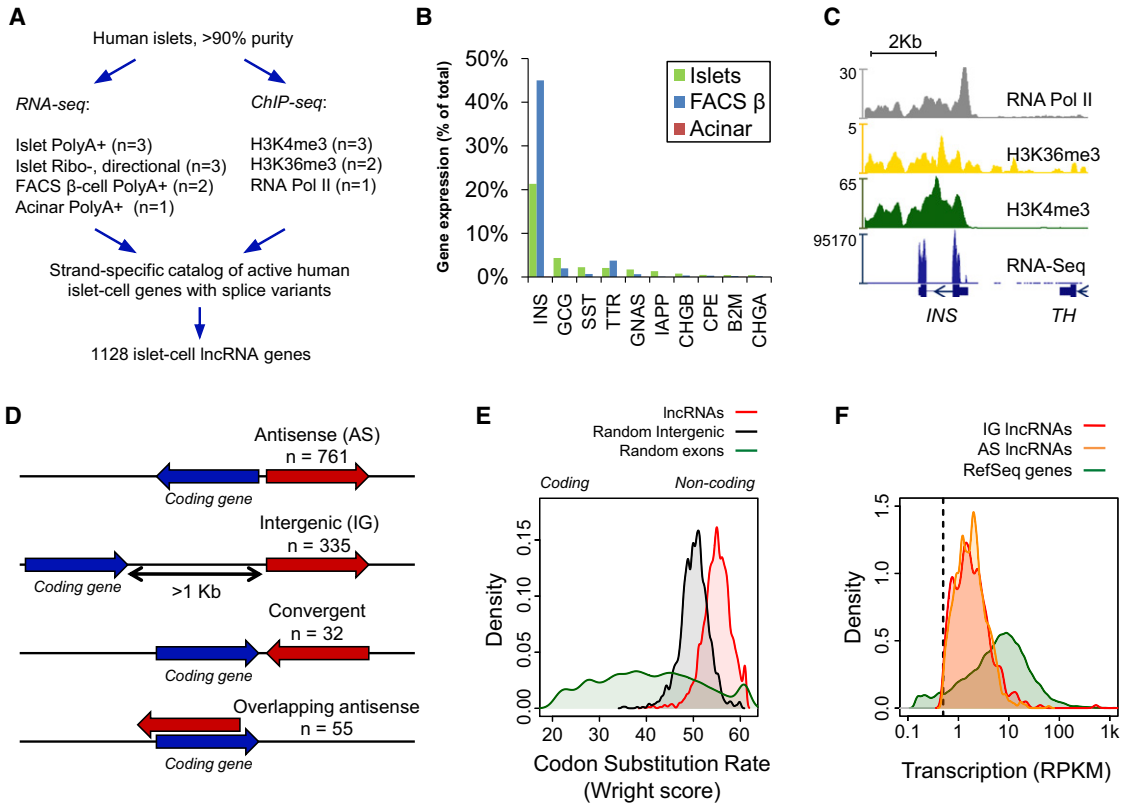


Figure 1. Systematic Identification of Human Islet and β cell lncRNAs

(A) Outline of the analysis pipeline.

(B) RNA-seq expression levels in the three pancreatic cell fractions for the ten most-expressed H3K4me3-positive genes in human islets.

(C) Chromatin and RNA-seq landscape of the *INS* locus.

(D) Definitions and counts of different lncRNA classes.

(E) Kernel density plot of the codon substitution rate of islet lncRNAs (red), intergenic control regions (black) and randomly selected exons of protein-coding genes (green).

(F) Kernel density plot of the transcriptional abundance of all RefSeq genes (green), intergenic (red), and antisense (orange) lncRNAs. The dashed line marks the 0.5 RPKM expression threshold used for defining lncRNAs. See also Figure S1 and Tables S1–S5.

For subsequent analysis we merged convergent and intergenic lncRNAs (a total of 367) as neither was closely associated with a promoter of a protein-coding gene. Comparison with acinar RNA-seq showed that only one lncRNA showed >3-fold acinar enrichment. Furthermore, qPCR analysis confirmed expression of 31/31 lncRNAs in human islets, and 26/31 in the human β cell-line EndoC-β H1 (Ravassard et al., 2011) (data not shown). These transcripts are therefore bona-fide islet-cell and in most cases β cell lncRNAs. Their genomic location and exon annotations are provided in Table S4.

Further analysis of the genomic properties of the 1128 islet lncRNAs showed that their overall protein-coding parameters resembled those of randomly chosen intergenic regions, confirming that most are likely to be truly noncoding transcripts (Figures 1E and S1F). Antisense and intergenic lncRNAs had a similar average length as RefSeq genes, yet showed ~10-fold lower expression (Figure 1F). Despite the low abundance of lncRNAs, the top quartile were expressed at levels comparable to mRNAs encoding transcriptional regulators linked to human diabetes (*HNF1A*, *HNF4A*, and *TCF7L2*; 6.4, 7.1 and 5.7 RPKM in human islets, respectively). Compared with RefSeq transcripts, islet lncRNA genes displayed similar RNA Pol II enrichment but were less often H3K36me3 enriched, in keeping with less frequent splicing (36% for intergenic lncRNAs, versus 93% in RefSeq genes) (Table S5). Finally, lncRNAs overlapped interspersed repeat regions more often than coding exons, yet less than random intergenic control regions (Figure S1G). In conclusion, these studies uncover a high-confidence set of 1128 human islet-cell genes that have several expected properties for lncRNAs.

Islet lncRNAs Are Highly Tissue Specific

Because many islet lncRNAs have escaped annotation, we reasoned that this could be in part because this class of transcripts is often cell-type specific. We realigned 16 human non-pancreatic RNA-seq data sets (Table S1), and found that 9.4% of RefSeq annotated genes were islet specific (Figure 2A, Table S6; see Experimental Procedures for a definition of the islet-specificity score). By contrast, 55% of intergenic lncRNAs, and 40% of antisense islet lncRNAs were islet specific, a significant enrichment over protein-coding genes ($p < 3 \cdot 10^{-16}$ for both comparisons) (Figure 2B). This set of lncRNAs was similarly

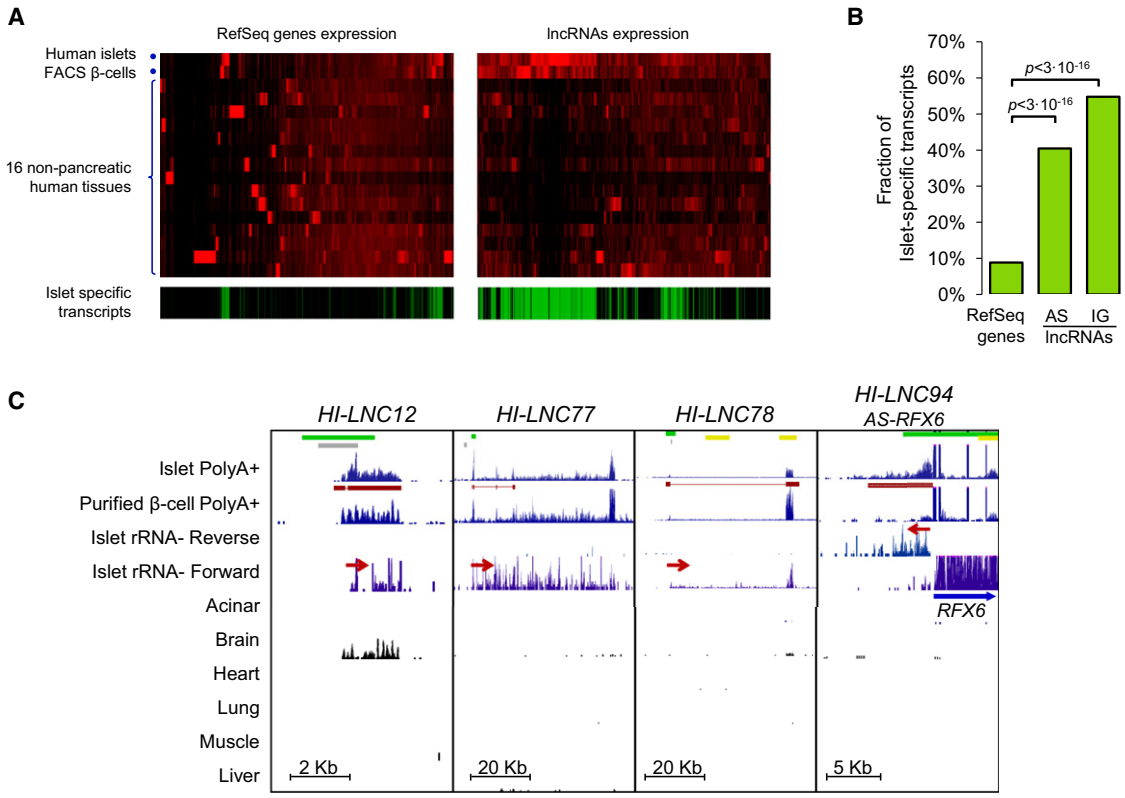


Figure 2. Islet lncRNAs Are Highly Tissue Specific

(A) Hierarchical clustering of expression levels of all RefSeq genes (left) and lncRNAs (right) across human islets, FACS-purified β cells and 16 nonpancreatic human tissues. Bottom panels highlight islet-specific transcripts in green.

(B) Islet specificity of antisense and intergenic lncRNAs compared to all RefSeq genes. *P* values were calculated by Chi-Square test.

(C) Representative examples of islet-enriched lncRNAs. Gene models are depicted in red and transcript orientation with a red arrow. All rows depict RNA-seq except the top row, which shows consistent peaks for H3K4me3 (green), H3K36 (yellow), and RNA Polymerase II (gray). The vertical axis is scaled at the same expression level for all tissues. Only 5/16 representative nonpancreatic tissues are shown for simplicity. See also Figure S2 and Table S6.

enriched in purified β cells (Figure 2A and not shown). Many examples of lncRNAs were thus found to be unique or highly specific to β cells within the entire panel of 18 tissues (Figure 2C). The islet selectivity was confirmed by qPCR in 12/12 lncRNAs using an independent panel of nine tissues (Figure S2). Thus, human islet-cell lncRNAs are frequently transcribed in a highly cell-type specific manner.

Islet lncRNAs Often Map near Islet-Specific Chromatin Domains and Coding Genes

To explore the biological significance of human islet lncRNAs, we next examined their relationship with nearby protein-coding genes. The abundance of each lncRNA transcript in human islets was not related to that of the neighboring protein-coding mRNA (Figure S3A). However, the expression of lncRNAs and their most proximal protein-coding gene significantly correlated across different tissues ($p < 3 \cdot 10^{-14}$) (Figure 3A). Accordingly, the closest protein-coding genes to islet intergenic and antisense lncRNAs were more often islet-specific than the entire collection of annotated genes ($p < 10^{-10}$), up to >3-fold when considering only islet-specific lncRNAs ($p < 2 \cdot 10^{-11}$) (Figures 3B and 3C). Notably, islet lncRNAs were preferentially located in genomic regions where the closest protein-coding gene has been linked to β cell function, development, and transcription (Figure 3C). Examples of known islet protein-coding genes paired with intergenic lncRNAs include *MAFB*, *FOXA2, PCSK1*, and *ISL1* (Figure S3B and data not shown). We compiled a list of 20 genes encoding known islet-enriched transcription factors, and found that remarkably 13 (65%), including *HNF1A*, *PDX1, PAX6*, *ISL1*, *INSM1*, *NEUROD1, GATA6, NKX2-2,* and *RFX6,* were associated with antisense lncRNAs (Figure S3B and data not shown). Thus, islet lncRNAs often map to the same genomic region as islet-enriched protein-coding genes.

It was also apparent that lncRNAs were often located in gene-poor areas. We examined the genomic intergenic spaces where lncRNAs reside and found that they are on average nearly 3-fold larger than the entire set of intergenic intervals in the human genome ($p < 3 \cdot 10^{-12}$) (Figure 3D). The fact that islet lncRNAs were often located in large gene-poor spaces, yet associated to genes with similar cell-type specific expression, suggested the existence of broad cell-specific regulatory domains. Human islets harbor tissue-specific clusters of open chromatin regulatory elements (COREs), often linked to a single islet-enriched protein-coding gene (Gaulton et al., 2010). We thus examined the relationship of islet lncRNAs to islet-specific COREs (see Experimental Procedures). Both intergenic and antisense

lncRNAs, as well as their nearby annotated genes, were more often located near islet-selective COREs than random intergenic regions ($p < 3 \cdot 10^{-9}$) or expressed annotated genes ($p < 0.013$) (Figure 3E). Thus, islet lncRNAs are often associated to broad cell-specific regulatory domains, many of which appear to be shared by protein-coding genes.

HI-LNC25 illustrates many of the above-mentioned features. It is a multiexonic transcript located in a broad ~1.6 Mb space that lacks any protein-coding gene but contains clusters of islet-specific active chromatin (Figure 3F). The most proximal protein-coding gene is *MAFB*, an essential regulator of islet-cell maturation that is abundantly expressed in human β cells (Artner et al., 2007; Dorrell et al., 2011). Interestingly, whereas *MAFB* is expressed in several tissues, *HI-LNC25* shows a much more restricted tissue distribution (Figure 3F). Taken together, these results indicate that islet-selective lncRNAs are frequently associated with cell-specific higher-order chromatin domains that potentially underlie coregulation of noncoding and coding transcript pairs.

Stage-Specific Activation of lncRNAs during β Cell Maturation

To understand the developmental regulation of islet lncRNAs, we first used qPCR to test the expression of β cell selective lncRNAs (ten intergenic, three antisense) in dissected Carnegie stage 17–19 human embryonic pancreas, a progenitor stage that shows scarce signs of cytodifferentiation (Piper et al., 2004). Of the 13 lncRNAs that were examined, all except one were silent or expressed at low levels in pancreatic progenitors and were subsequently active in adult islets (Figure 4A). Thus, islet lncRNAs are not only often islet specific, but they are also linked to the pancreatic endocrine differentiation program.

We next examined the dynamics of islet lncRNAs in a human embryonic stem-cell (hES) differentiation model. Current methods to differentiate β cells from stem cells are limited by difficulties in completing the maturation of β cells in vitro. We thus used a hES cell protocol that involves multiple in vitro differentiation steps, followed by encapsulation of differentiated pancreatic endoderm, which after implantation into mice produces mature endocrine cells over a 140 day in vivo incubation period (Kroon et al., 2008). We profiled the same lncRNAs at each stage of the protocol and discovered that all islet lncRNAs were markedly induced during the in vivo maturation step (Figure 4B). Notably, six lncRNAs were expressed at very low or undetectable levels throughout all in vitro differentiation steps and were only activated during the in vivo maturation step (Figure 4B). Thus, islet-specific lncRNA gene activation is linked to pancreatic endocrine differentiation during embryogenesis and during in vivo differentiation of hES cell-derived pancreatic progenitors. Islet lncRNAs are therefore candidate markers and/or regulators of β cell differentiation and maturation.

Glucose-Dependent Regulation of Islet lncRNAs

Mature β cells exhibit a transcriptional response to increased demand (Bensellam et al., 2009; Schuit et al., 2002). Selected islet protein-coding transcripts are thus moderately induced in response to high glucose under appropriate experimental conditions. To assess if islet lncRNAs are also dynamically regulated, we exposed human islets to 4 mM or 11 mM glucose for 72 hrs. This revealed glucose-dependent changes in protein-coding mRNAs related to insulin secretion (*INS*, *IAPP*, *PCSK1*), whereas the stress-responsive marker *DDIT3* (CHOP) was not induced (Figure 4C). We then tested the set of 13 islet lncRNAs and discovered that *HI-LNC78* and *HI-LNC80,* two intergenic multiexonic lncRNAs that are not located near islet-enriched protein coding genes, were consistently upregulated in a glucose-dependent manner in five individual donors (Figure 4C). Thus, selected islet lncRNAs are dynamically regulated in settings that are relevant for mature islet-cell physiology.

Identification of Conserved β Cell-Specific lncRNAs in Mice

We next assessed the evolutionary conservation of islet lncRNAs. Mammalian sequence conservation scores of islet lncRNAs were markedly lower than protein-coding exons, yet higher than random intergenic regions, in keeping with previous results (Pauli et al., 2012; Ulitsky et al., 2011) (Figure 5A). For nearly 70% of human lncRNAs, we identified orthologous mouse genomic regions of >200 bp, compared to 58% of random human intergenic regions ($p < 2 \cdot 10^{-9}$) (Figure 5B).

The degree of sequence identity that is required for functional conservation of noncoding transcripts is uncertain. To improve our understanding of the evolutionary conservation of islet lncRNAs, we examined if orthologous lncRNAs were transcribed in mouse islets. To this end, we generated 78 million paired-end reads from purified mouse islet PolyA+ RNA directional libraries (Table S7). We found that 47% of the mouse orthologous regions were transcribed in mouse islets, as opposed to <3% random intergenic regions of the same size ($p < 3 \cdot 10^{-16}$) (Figure 5C). This is likely an underestimate of orthologous mouse lncRNAs since qPCR analysis confirmed the expression of 7/7 orthologous mouse lncRNAs, and it also showed that 5/7 orthologous mouse genomic regions that did not surpass our detection threshold in the RNA-seq analysis were nevertheless detected at low levels in mouse islet and/or β cell-line RNA, whereas control intergenic regions were undetectable (Figures 5D and 5E).

Next, we assessed if mouse orthologous lncRNA transcripts are regulated in a similar manner as their human counterparts. Like human lncRNAs, mouse orthologous transcripts were >4-fold more frequently islet-cell-specific than protein-coding genes ($p < 3 \cdot 10^{-16}$) (Figures 5F, 5G, and S4A). Moreover, the analysis of E13.5 embryonic pancreas showed that five out eight lncRNAs were inactive prior to endocrine differentiation, indicating that they are typically regulated in a similar stage-specific fashion as their human orthologs (Figures 5H and S4B). Furthermore, a comparison of islets from neonate and adult mice showed that most lncRNAs were further upregulated during postnatal islet maturation (Figure S4C). Likewise, five out of eight tested mouse islet lncRNAs were induced upon exposure to high-glucose concentrations, including *Mi-Linc80*, whose ortholog is glucose-responsive in human islets (Figure 5I). Several islet lncRNAs were also regulated in vivo in glucose intolerant Leptin-deficient mice (Figure S4D). Thus, genomic mouse regions that are orthologous to human islet lncRNA genes are frequently also transcribed in islets, and like their human counterparts they exhibit a highly dynamic and cell-specific regulation.

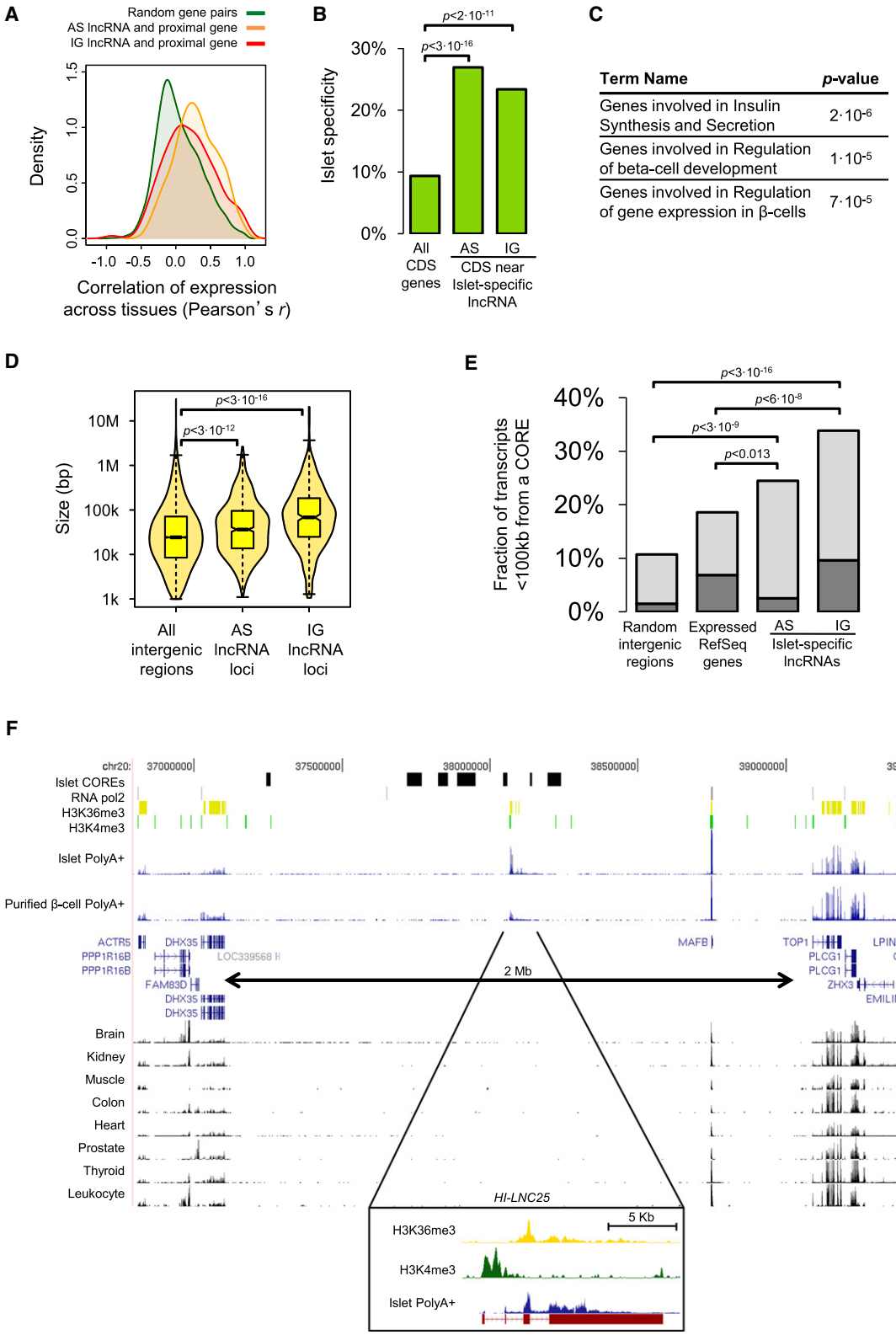


Figure 3. Many Islet lncRNAs Map Near Islet-Specific Chromatin Domains and Regulatory Genes

(A) LncRNA transcript levels correlate with those of their nearest coding gene across tissues. Kernel density plots of correlation of RNA levels for random gene pairs (green), intergenic (IG) lncRNAs paired with their nearest coding genes (red) ($p < 3 \cdot 10^{-14}$ compared to random genes pairs, Mann-Whitney *U* test), and antisense (AS) lncRNAs with their proximal coding genes (orange) ($p < 3 \cdot 10^{-16}$).

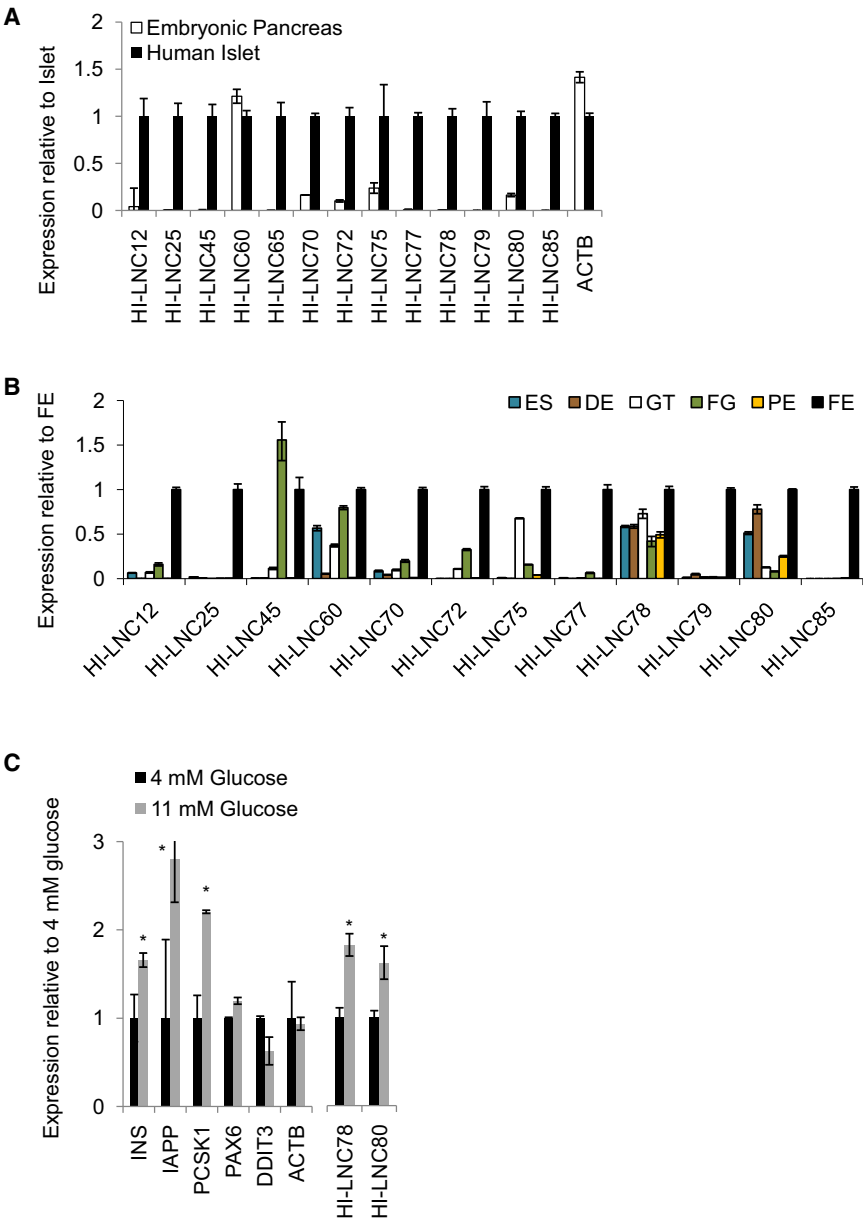


Figure 4. Human lncRNAs Are Dynamically Regulated

(A) Expression of human islet-specific lncRNAs (10 intergenic, 3 antisense) was assayed in mature islets and pooled embryonic pancreas (Carnegie stage 17–19) by qPCR. The results were normalized to *TBP* mRNA and expressed as a fraction of expression in islets. All but one lncRNAs were activated after the pancreatic progenitor stage.

(B) Human islet lncRNAs are induced upon in vivo differentiation of embryonic stem cells. Human embryonic stem cells (ES) were subjected to a differentiation protocol that undergoes stage-specific differentiation to definitive endoderm (DE), primitive gut tube (GT), foregut endoderm (FG), and pancreatic endoderm (PE), respectively. The cells are then transplanted into mice to allow for in vivo maturation of functional endocrine cells (FE). ES, DE, GT, FG, PE, and FE represent days 0, 2, 5, 7, 10, and 150 of the differentiation protocol. *HI-LNC65* was not detected in any of the maturation steps and is not shown.

(C) Expression of a subset of human lncRNAs is regulated by glucose. Human islets were either cultured under low (4 mM) or high (11 mM) glucose concentrations for 72 hr. Panels from left to right illustrate expected glucose responsiveness of *INS*, *IAPP*, and *PCSK1*, and the glucose responsiveness of *HI-LNC78* and *HI-LNC80* in five independent human islet samples. Values are normalized to *TBP* mRNA and expressed as a fraction of the condition denoted in the vertical axis label. Bars represent means ± SEM and asterisks denote $p < 0.001$ (Student's t test).

HI-LNC25 Regulates *GLIS3* mRNA

The evolutionary conserved regulation of islet lncRNAs points to a functional role. LncRNAs have been linked to diverse types of functions, frequently involving direct or indirect regulation of expression of protein-coding genes (Wang and Chang, 2011). To begin to understand islet lncRNA function we focused on *HI-LNC25,* which fulfills several typical features of human islet lncRNAs, as shown in Figure 3F. As an experimental model we used a human β cell line (EndoC-β H1) that exhibits glucose-induced insulin secretion (Ravassard et al., 2011). We transduced human β cells with lentiviral vectors expressing two independent shRNA hairpins that suppressed *HI-LNC25* RNA and examined the expression of a panel of 24 islet mRNAs. This screen identified *GLIS3* mRNA as a potential regulatory target of *HI-LNC25*. *GLIS3* encodes an islet transcription factor, it is mutated in a form of monogenic diabetes, and contains T2D risk variants (Cho et al., 2012; Senée et al., 2006). We therefore

(B) Islet specificity of all coding RefSeq genes (CDS) compared with the nearest coding RefSeq gene to islet-specific antisense and intergenic lncRNAs. *P* values were calculated by Chi-square test.

(C) Top-enriched gene ontology terms for genes closest to islet-specific lncRNAs.

(D) LncRNAs are preferentially located in gene-poor regions. Violin plots of the length distribution of all intergenic spaces in the genome >1 kb, compared with the intergenic spaces where antisense and intergenic lncRNAs reside.

(E) Percentage of islet-specific lncRNAs located near islet-specific clusters of regulatory elements (COREs), compared with random intergenic regions and RefSeq genes. Dark gray areas of the bars represent lncRNAs directly overlapping a CORE element, while light gray areas represent non-overlapping lncRNAs located <100 kb away from COREs.

(F) Birds-eye view of a ~2 Mb region surrounding the islet regulator *MAFB*, centered on *HI-LNC25*. All rows depict RNA-seq (vertical axis fixed at 50 RPKMs), except for the first four rows, which show the islet CORE regions (black), enriched regions for RNA polymerase II (gray), H3K36 (yellow), and H3K4me3 (green). The bottom inset shows a zoomed-in view of the *HI-LNC25* gene. Only 8/16 representative human tissues are shown for simplicity. See also Figure S3.

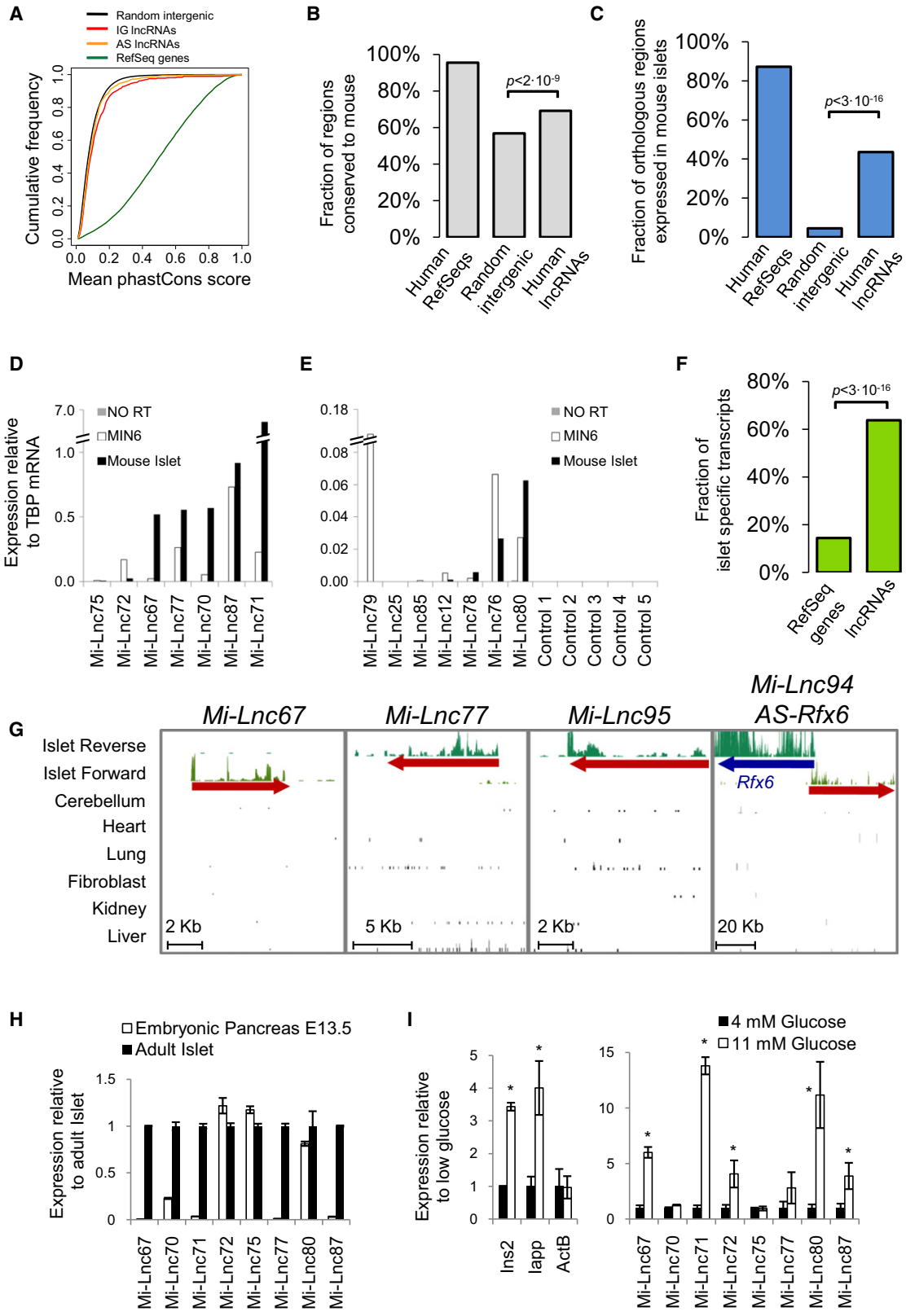


Figure 5. Mouse Islet lncRNA Orthologs Exhibit Conserved Regulation

(A) Mean phastCons scores for intergenic lncRNAs (red), antisense lncRNAs (orange), random intergenic fragments (black), and coding genes (green).

(B) Fraction of human RefSeq exons, random intergenic regions, and lncRNAs with an orthologous mouse genomic region of at least 200 bp.

performed four independent *HI-LNC25* knockdown experiments with two separate shRNA hairpins. Although this depletion of *HI-LNC25* did not cause significant changes in glucose-stimulated insulin secretion (not shown), we observed a consistent reduction of *GLIS3* mRNA in comparison to five control shRNA hairpins, whereas other control genes remained unaltered (Figure 6A). This effect was stably maintained over several days in culture (Figure 6B). Thus, *HI-LNC25* positively regulates *GLIS3* mRNA, supporting the notion that our collection of islet lncRNAs contains regulatory transcripts.

Dysregulation of Islet lncRNAs in Human T2D

To determine the potential role of islet lncRNAs in the pathogenesis of diabetes, we first examined whether lncRNAs are abnormally expressed in human islets from donors with T2D.

We examined the panel of 13 lncRNAs and added *KCNQ1OT1,* a lncRNA that was detected as an overlapping antisense lncRNA in our analysis, and has been previously genetically associated with T2D (Voight et al., 2010). We compared these lncRNAs in islets from 19 nondiabetic and 16 T2D donors. We found that two lncRNAs, namely *KCNQ1OT1* and *HI-LNC45*, were significantly increased or decreased in T2D islets, respectively ($p < 0.02$) (Figure 7A). To ensure that this result was not due to differences in islet purity between groups of samples, we normalized the expression levels of HI-LNCs relative to the islet transcription factor *PAX6*, and found similar results (data not shown). Thus, selected lncRNAs are dysregulated in T2D islets.

A Subset of lncRNAs Map to Genomic Loci Underlying Human Diabetes

To further examine the potential role of islet lncRNAs in the genetic susceptibility for diabetes, we first examined loci implicated in monogenic syndromes of β cell dysfunction disorders. Islet antisense lncRNAs were found within loci underlying neonatal diabetes (*ABCC8/KCNJ11*), pancreas agenesis (*GATA6*), and monogenic diabetes (*HNF1A*) (Allen et al., 2012; Bell and Polonsky, 2001; Gloyn et al., 2004). Because several antisense lncRNAs have been implicated in *cis* regulatory effects (Wilusz et al., 2009; Xu et al., 2011), these islet lncRNAs provide additional candidate regulatory elements that may contain pathogenic variants.

Next, we examined genetic association data from multifactorial T2D and related continuous glycemic measures, including fasting glucose in nondiabetic individuals (Dupuis et al., 2010; Voight et al., 2010). Currently, >50 loci are known to show association to T2D and related traits, and at most loci noncoding variants are likely to be causal (McCarthy, 2010). Earlier studies showed that two annotated lncRNAs, *ANRIL/CDKN2BAS* and *KCNQ1OT1*, map within established T2D susceptibility loci (Voight et al., 2010; Zeggini et al., 2007). To evaluate the potential overlap between newly identified islet lncRNAs with T2D loci, we used MAGENTA (Segrè et al., 2010). This tool tests whether a prespecified set of genes is enriched for trait associations in genome-wide data, and was applied here to available data sets for T2D and related continuous glycemic measures, as well as five nonislet related control phenotypes.

We found that islet lncRNA transcripts were, as a group, enriched for association with both T2D ($p = 0.02$) and fasting glucose ($p = 0.01$), with no enrichment for the control phenotypes apart from height ($p = 0.02$) (Lango Allen et al., 2010) and waist-hip ratio ($p = 0.001$) (Heid et al., 2010). When lncRNAs were examined by type (intergenic versus antisense), enrichment of association signal for intergenic lncRNAs was only observed for T2D and fasting glucose (T2D $p = 0.03$; fasting glucose $p = 0.003$) (Table S8). Of 55 T2D susceptibility loci, 9 contained islet lncRNAs within 150 kb of the reported lead SNP, 6 of which have been linked directly to β cell dysfunction (Figure 7A) (Cho et al., 2012; Dupuis et al., 2010; Kooner et al., 2011; Strawbridge et al., 2011; van de Bunt and Gloyn, 2010; Voight et al., 2010). Examples include a lncRNA in the vicinity of *PROX1*, which overlaps the region of strongest association, very likely to contain the casual SNP (Figure 7B), and the most significant lncRNA in the MAGENTA analysis, near *WFS1* (Figure 7C). These studies therefore offer a new class of genomic elements that can be interrogated to dissect the functional etiology of T2D susceptibility.

DISCUSSION

Increasing evidence points to a regulatory function of many lncRNAs, which suggests a pivotal role in physiology and disease (Guttman and Rinn, 2012; Wang and Chang, 2011). We have integrated transcriptional and chromatin maps to systematically annotate lncRNA genes in pancreatic islet cells, a key tissue for human diabetes. Strand-specific analysis uncovered hundreds of intergenic islet lncRNAs and many others that were antisense to protein-coding genes and hence not discernable using conventional cDNA libraries. Human islet lncRNAs were found to be highly cell-type specific, developmentally regulated, and tightly linked to the differentiation of ES cell-derived islet cells. We identified mouse orthologous transcripts, and discovered that they are dynamically regulated in a similar manner as human islet lncRNAs. We focused on a prototypical islet lncRNA and demonstrate that it acts as a positive regulator of an islet mRNA. Finally, we show examples of lncRNAs

(C) Fraction of orthologous regions to human RefSeq exons, random intergenic regions, and lncRNAs that are transcribed higher than 0.5 RPKM in mouse islet RNA-seq.
(D) qPCR confirmation in mouse islets and MIN6 mouse β cell line for 7/7 orthologous mouse islet lncRNAs.
(E) qPCR reveals islet expression of 4/7 orthologous regions where islet RNA-seq did not reveal transcription, but not in random intergenic regions. No RT denotes control reactions lacking reverse transcriptase.
(F) Orthologous mouse lncRNAs are more frequently islet specific than RefSeq genes.
(G) Examples of conserved mouse islet-specific lncRNAs. Red arrows depict lncRNA orientation. The y axis for all tissues is adjusted to the same value as islets (RPKM).
(H) qPCR shows that 4/7 lncRNAs were silent in E13.5 mouse embryonic pancreas and were induced in adult islets.
(I) Expression of several orthologous lncRNAs is regulated by glucose. Mouse islets were cultured either under low (4 mM) or high (11 mM) glucose concentrations for 72 hr. Expression is shown as a fraction of expression under low-glucose conditions (n = 4). All qPCR values are normalized to *Tbp* mRNA. Bars represent average ± SEM and asterisks denote $p < 0.001$ (Student's t test). See also Figure S4 and Table S7.

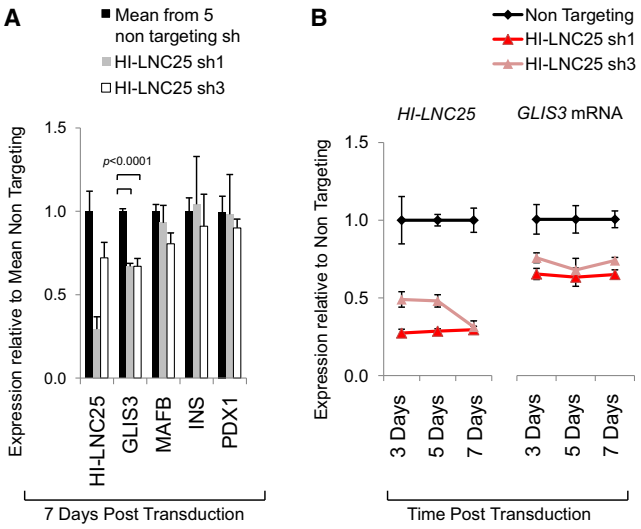


Figure 6. Knockdown of *HI-LNC25* in β Cells Causes Downregulation of *GLIS3*

(A) Human β cells (EndoC-β H1) were transduced with lentiviral vectors expressing two independent RNA hairpins that target *HI-LNC25* RNA or five independently transduced negative control nontargeting shRNA sequences. Cells were harvested at 7 days and assayed for *GLIS3* mRNA and several control genes.

(B) Effects on *GLIS3* mRNA were stable 3–7 days post transduction. Expression levels were normalized to *TBP* mRNA and shown as a fraction of nontargeting vector cells. The results reflect four independent experiments. Bars are average ± SEM, *p* values were obtained with Student's t test.

that are dysregulated in T2D islets and others that map to T2D susceptibility loci. These studies open new avenues to study the role of functional lncRNAs in islet-cell disease and therapeutic programming.

LncRNAs Are Dynamically Regulated during Islet Cell Maturation

Islet lncRNAs show a striking cell-type specific expression pattern. This finding is in line with a recent analysis of intergenic lncRNAs across multiple tissues (Pauli et al., 2012), and with earlier descriptions of lncRNAs that have distinct cell-type specific expression patterns in brain (Mercer et al., 2008). Our studies additionally demonstrate that islet-specific lncRNA genes are inactive in embryonic pancreatic progenitors and subsequently become activated in islet cells. Likewise, numerous lncRNAs were activated during the final endocrine differentiation step of a pluripotent cell differentiation protocol. This indicates that islet lncRNAs are an integral component of the endocrine differentiation program.

Islet-specific lncRNA transcription was linked to clusters of open chromatin. Such clusters were previously shown to be associated with islet-specific transcription (Gaulton et al., 2010). Interestingly, we also observed concordant cell-specific transcription of lncRNAs with their nearest protein-coding genes, in accordance with other genomic analysis of lncRNAs (Ponjavic et al., 2009). These data suggest that concordant tissue-specific expression of lncRNA and protein-coding pairs is linked to shared chromosomal regulatory domains.

The islet specificity of lncRNAs is reminiscent of the recent discovery of transcription factors that show a similar stage- and cell-specific activation pattern (reviewed in Servitja and Ferrer, 2004). Several such islet-enriched factors have now been shown to control islet development and have consequently been exploited to program insulin-expressing cells from somatic cells (Collombat et al., 2009; Zhou et al., 2008). By analogy, islet lncRNAs that exert gene-regulatory functions could also be employed in efforts to program functional β cells. This prospect is supported by recent studies showing that a lncRNA can promote reprogramming pluripotent cells from somatic cells (Loewer et al., 2010).

Several islet lncRNAs were specifically activated during the in vivo maturation step of the ES cell differentiation protocol. This result is significant because the major roadblock for efforts to derive fully functional β cells from pluripotent cells lies in the inability to complete the differentiation process in vitro (Van Hoof et al., 2009). Our findings suggest that lncRNAs that are induced at this late stage can be used as biomarkers of mature β cells and potentially be exploited as effectors to promote β cell programming.

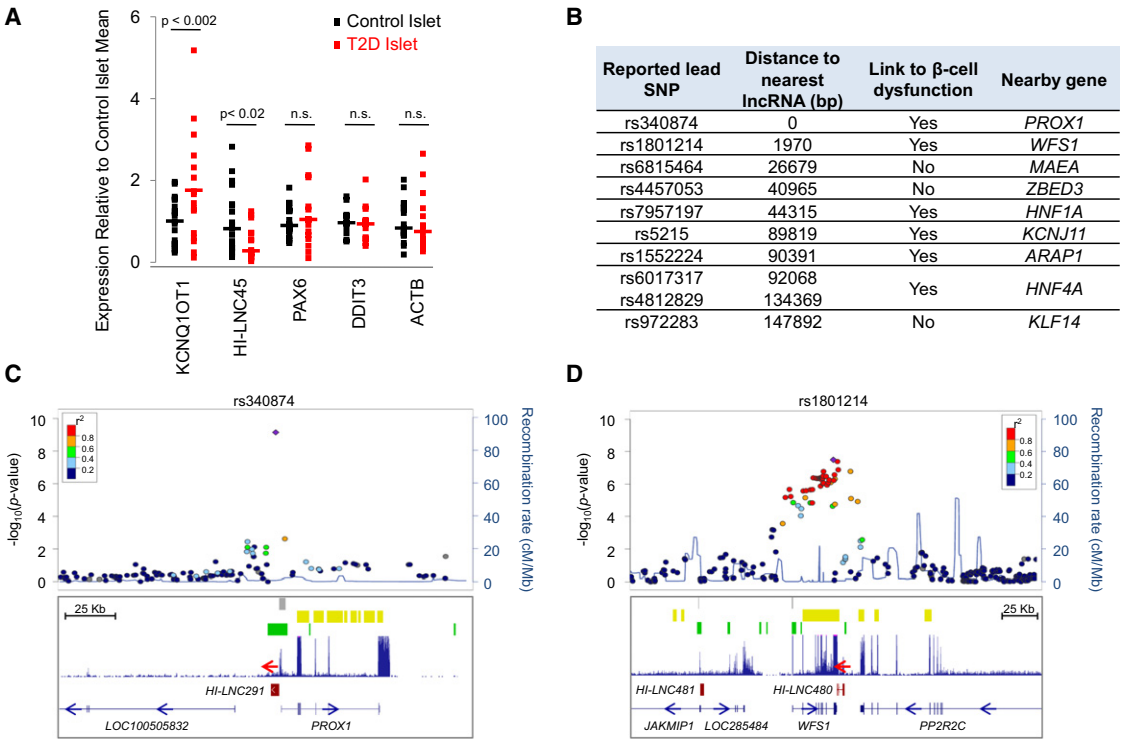
Evolutionary Conserved Regulation of Islet lncRNAs

For a major fraction of lncRNAs we identified orthologous mouse genomic regions, although the overall sequence identity was clearly not subject to a similar evolutionary constraint as protein-coding genes, in keeping with earlier studies of lncRNAs (Pauli et al., 2012; Ulitsky et al., 2011). Several mechanisms have been previously proposed to explain this comparatively low sequence conservation, including a hypothetically less stringent requirement for primary sequence conservation to maintain functional secondary structures, or the existence of very short stretches of functional sequences (Ulitsky et al., 2011). It is also conceivable that evolutionary changes in lncRNA sequences are functionally coupled to other parallel genomic changes, such as the acquisition of species-specific repetitive elements.

Importantly, in the current study we discovered that a significant fraction of orthologous mouse sequences were transcribed in mouse islet cells and β cell lines. Orthologous lncRNA pairs display similar cell-type specific expression and stage-specific regulation during embryonic development. This finding points to an evolutionary conserved functional property of lncRNAs that extends beyond primary sequence and further supports that lncRNAs are an integral component of the mammalian islet-cell differentiated phenotype.

Islet lncRNAs and Diabetes

The identification of lncRNAs opens a new framework to study human pathophysiology. Our study revealed lncRNA genes at six neonatal diabetes loci and in *HNF1A*, the most common monogenic diabetes locus (Bell and Polonsky, 2001). Most such lncRNAs run antisense from the coding gene. Antisense transcripts have been shown to control the transcription of protein-coding genes in *cis* (Wilusz et al., 2009; Xu et al., 2011). Antisense lncRNAs identified in this study are thus candidate regulators of protein-coding genes implicated in human β cell monogenic disorders and can be potentially affected by pathogenic mutations.



Reported lead SNP	Distance to nearest lncRNA (bp)	Link to β-cell dysfunction	Nearby gene
rs340874	0	Yes	*PROX1*
rs1801214	1970	Yes	*WFS1*
rs6815464	26679	No	*MAEA*
rs4457053	40965	No	*ZBED3*
rs7957197	44315	Yes	*HNF1A*
rs5215	89819	Yes	*KCNJ11*
rs1552224	90391	Yes	*ARAP1*
rs6017317	92068	Yes	*HNF4A*
rs4812829	134369		
rs972283	147892	No	*KLF14*

Figure 7. Several Human lncRNAs Map to T2D-Associated Loci

(A) *KCNQ1OT1* and *HI-LNC45* are dysregulated in islets from T2D patients. Islet LncRNAs and control mRNAs were assessed by qPCR in islets isolated from 19 control and 16 T2D donors. Values are normalized to TBP mRNA and expressed as a fraction of controls. *P* values represent Mann-Whitney significance test.

(B) Nine T2D-associated loci contain islet lncRNAs mapping within 150 kb of the reported lead SNPs. Six of these loci have been linked to β cell dysfunction (Dupuis et al., 2010; van de Bunt and Gloyn, 2010; Voight et al., 2010).

(C and D) Islet lncRNAs map to established T2D association signals located near *PROX1* and *WFS1*. The top panel shows T2D association *p* values for all analyzed SNPs in the published DIAGRAM genome-wide association meta analysis (Voight et al., 2010) and the combined *p* value for the "lead" SNP after further follow up (rs340874, $p = 7.2 \cdot 10^{-10}$) (rs1801214, $p = 3.16 \cdot 10^{-8}$), which are denoted as purple diamonds. SNP colors indicate LD relationships. RNA-seq and ChIP-seq is presented in the bottom panel as described in the legend for Figure 2. See also Table S8.

These results are also relevant to the molecular etiology of T2D, a disease caused by abnormal β cell function or growth (Bell and Polonsky, 2001; McCarthy, 2010). We identified examples of islet lncRNAs that are dysregulated in T2D. Our findings also revealed islet-cell lncRNAs as candidates to dissect the underpinnings of noncoding variants underlying T2D risk. More generally, these results set the stage for future studies to dissect the potential role of inherited and acquired defects in lncRNA genes in human β cell physiology and disease. Several experimental model systems, including mouse genetics coupled with perturbation studies in human β cells, are now available to address this challenge.

EXPERIMENTAL PROCEDURES

Pancreatic Islets and Tissues

Human islets were isolated at the University of Virginia, University of Geneva, University of Lille, and San Raffaele Scientific Institute islet centers (Bucher et al., 2005). Human islets used for RNA-seq and ChIP-seq were cultured with CMRL 1066 medium containing 10% fetal calf serum (FCS) before shipment, after which they were cultured for 3 days with RPMI 1640 medium containing 11 mM glucose, supplemented with 10% FCS. For glucose regulation experiments, human or mouse islets were cultured for 72 hr in RPMI 1640 medium containing 4 or 11 mM glucose. Donor information for nondiabetic and diabetic donors is provided in Table S2. Samples were selected based on islet cell purity, as assessed by dithizone staining, immunofluorescence analysis, and qPCR analysis of cell-specific mRNAs.

RNA Analysis

RNA was isolated with Trizol (Invitrogen) or RNeasy (QIAGEN). DNase I treatment and a control lacking reverse transcriptase were performed on all RNA samples. Quantitative PCR was performed as described based on SYBR green detection (van Arensbergen et al., 2010). See Table S9 for oligonucleotide sequences.

miR-Based Knockdown Assays

Lentiviral vectors carrying two miRNA-based shRNAs targeting *HI-LNC25* and five nontargeting control sequences were transduced into the EndoC-β H1 human β cell line (Castaing et al., 2005; Ravassard et al., 2011; Scharfmann et al., 2008). Oligonucleotide sequences are shown in Table S9. Nontransduced cells were assayed in parallel. Cells were assayed 3 or 7 days post transduction.

ES Cell Differentiation

Human ES cells were differentiated to pancreatic insulin-producing cells as described (Kroon et al., 2008).

RNA-Seq and ChIP-Seq

Chromatin immunoprecipitations (ChIP) were performed as described (Boj et al., 2009), with modifications described in Supplemental Experimental

Procedures. ChIP libraries were prepared according to Illumina protocols and sequencing of single-end reads was performed on a GAIIx system. RNA-seq was performed from unidirectional and nondirectional cDNA libraries prepared from PolyA+ or rRNA-depleted pancreatic islet RNA, and sequenced using GAIIx, HiSeq2000, or SOLiD 4 systems. Nonpancreatic RNA-seq reads were retrieved from the Illumina human BodyMap2 data set (http://www.ebi.ac.uk/arrayexpress/browse.html?keywords=E-MTAB-513) and the ENCODE/LICR Project (Birney et al., 2007). Reads were aligned to the NCBI36/hg18 and NCBI37/mm9 genomes using Bowtie v0.11.3 for ChIP-seq, allowing one mismatch per read, no multimapping, and no clonal reads. Nondirectional RNA-seq libraries were mapped with TopHat v1.2.0, and unidirectional libraries with Bowtie, using parameters described in the Supplemental Experimental Procedures. The number of uniquely aligned reads, read length, and sample information are shown in Tables S1 and S7.

LncRNA Characterization and Analysis

Transcriptional units were defined from the RNA-seq alignments and joined by splice junctions and paired-end mapping to create putative gene models. Transcriptional units that were expressed at >0.5 RPKMs, with H3K4me3 enrichment in the 5′ region in at least two samples, and either did not overlap with annotated protein-coding genes or did overlap but were transcribed in the opposite strand and were assessed for protein-coding potential and processed through our lncRNA discovery pipeline. To assess evolutionary sequence conservation human islet lncRNA exon locations or the same fragments randomized 1000 times in the alignable portion of intergenic space were analyzed. Orthologous mouse genomic sequences were identified with the LiftOver tool from the UCSC browser. Mouse islet RNA-seq was used to detect transcripts expressed at >0.5 RPKM in orthologus or randomized control regions. To evaluate islet specificity of lncRNAs, Cluster 3.0 v1.5a was used for hierarchical clustering of transcript expression in islet cells and 16 nonpancreatic tissues. Furthermore, a score was created that measures the difference between the average expression in 3 human islet PolyA+ samples and the average +2 standard deviations in the remaining 16 tissues. MAGENTA (Segrè et al., 2010) was used to evaluate whether lncRNAs were enriched for association to T2D and related glycemic traits, with adaptions described in the Supplemental Experimental Procedures.

Statistical Analysis

Chi-square, Pearson's *r* coefficient, and independent 2-group Mann-Whitney *U* tests were implemented with the R statistical package (http://www.r-project.org). Student's t test was performed with Microsoft Excel for quantitative PCR experiments. Error bars in Figures 4–6 represent SEM.

ACCESSION NUMBERS

Data sets have been deposited in the ArrayExpress Archive under accession number E-MTAB-1294. Islet lncRNAs coordinates are available at http://www.betacellregulation.net.

SUPPLEMENTAL INFORMATION

Supplemental Information includes nine figures, four tables, Supplemental Experimental Procedures, and Supplemental References and can be found with this article online at http://dx.doi.org/10.1016/j.cmet.2012.08.010.

ACKNOWLEDGMENTS

This work was funded by grants from the NIH-BCBC (2U01 DK072473-06 to J.F., P.R., and L.S.; U01-DK089567 to M.S. and J.F.), the Juvenile Diabetes Research Foundation (26-2008-633 to J.F. and 31-2008-416 to T.B., F.P., and L.P. to support human islet isolation), Ministerio de Economía y Competitividad (SAF2008-03116 to J.F. and a Ph.D. fellowship to Ignasi Morán). A.L.G. is a Wellcome Trust Senior Fellow in Basic Biomedical Research (WT 09510/Z/10/Z). We thank Rudolph Leibel (University of Columbia) for providing B6.V-Lepob/J mice, Bing Ren (Ludwig Institute for Cancer Research) for generating the ENCODE mouse tissue RNA-seq data, Kelli Bramlett (Genome Sequencing Collaborations Group) for her help in generation of SOLiD data and Viacyte, Inc. for performing implantations of hES cell derivatives. We are also grateful to Ayellet Segrè (Broad Institute) for helpful discussion on MAGENTA, and to Chris Stoeckert and Elisabetta Manduchi (University of Pennsylvania) for help with the online submission.

Received: March 21, 2012
Revised: July 30, 2012
Accepted: August 31, 2012
Published online: October 2, 2012

REFERENCES

Allen, H.L., Flanagan, S.E., Shaw-Smith, C., De Franco, E., Akerman, I., Caswell, R., Ferrer, J., Hattersley, A.T., and Ellard, S. (2012). GATA6 haploinsufficiency causes pancreatic agenesis in humans. Nat. Genet. *44*, 20–22.

Artner, I., Blanchi, B., Raum, J.C., Guo, M., Kaneko, T., Cordes, S., Sieweke, M., and Stein, R. (2007). MafB is required for islet beta cell maturation. Proc. Natl. Acad. Sci. USA *104*, 3853–3858.

Bell, G.I., and Polonsky, K.S. (2001). Diabetes mellitus and genetically programmed defects in beta-cell function. Nature *414*, 788–791.

Bensellam, M., Van Lommel, L., Overbergh, L., Schuit, F.C., and Jonas, J.C. (2009). Cluster analysis of rat pancreatic islet gene mRNA levels after culture in low-, intermediate- and high-glucose concentrations. Diabetologia *52*, 463–476.

Bertone, P., Stolc, V., Royce, T.E., Rozowsky, J.S., Urban, A.E., Zhu, X., Rinn, J.L., Tongprasit, W., Samanta, M., Weissman, S., et al. (2004). Global identification of human transcribed sequences with genome tiling arrays. Science *306*, 2242–2246.

Birney, E., Stamatoyannopoulos, J.A., Dutta, A., Guigó, R., Gingeras, T.R., Margulies, E.H., Weng, Z., Snyder, M., Dermitzakis, E.T., Thurman, R.E., et al; ENCODE Project Consortium; NISC Comparative Sequencing Program; Baylor College of Medicine Human Genome Sequencing Center; Washington University Genome Sequencing Center; Broad Institute; Children's Hospital Oakland Research Institute. (2007). Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature *447*, 799–816.

Boj, S.F., Servitja, J.M., Martin, D., Rios, M., Talianidis, I., Guigo, R., and Ferrer, J. (2009). Functional targets of the monogenic diabetes transcription factors HNF-1alpha and HNF-4alpha are highly conserved between mice and humans. Diabetes *58*, 1245–1253.

Bucher, P., Mathe, Z., Morel, P., Bosco, D., Andres, A., Kurfuest, M., Friedrich, O., Raemsch-Guenther, N., Buhler, L.H., and Berney, T. (2005). Assessment of a novel two-component enzyme preparation for human islet isolation and transplantation. Transplantation *79*, 91–97.

Cabili, M.N., Trapnell, C., Goff, L., Koziol, M., Tazon-Vega, B., Regev, A., and Rinn, J.L. (2011). Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. Genes Dev. *25*, 1915–1927.

Carninci, P., Kasukawa, T., Katayama, S., Gough, J., Frith, M.C., Maeda, N., Oyama, R., Ravasi, T., Lenhard, B., Wells, C., et al; FANTOM Consortium; RIKEN Genome Exploration Research Group and Genome Science Group (Genome Network Project Core Group). (2005). The transcriptional landscape of the mammalian genome. Science *309*, 1559–1563.

Castaing, M., Duvillié, B., Quemeneur, E., Basmaciogullari, A., and Scharfmann, R. (2005). Ex vivo analysis of acinar and endocrine cell development in the human embryonic pancreas. Dev. Dyn. *234*, 339–345.

Cho, Y.S., Chen, C.H., Hu, C., Long, J., Ong, R.T., Sim, X., Takeuchi, F., Wu, Y., Go, M.J., Yamauchi, T., et al; DIAGRAM Consortium; MuTHER Consortium. (2012). Meta-analysis of genome-wide association studies identifies eight new loci for type 2 diabetes in east Asians. Nat. Genet. *44*, 67–72.

Collombat, P., Xu, X., Ravassard, P., Sosa-Pineda, B., Dussaud, S., Billestrup, N., Madsen, O.D., Serup, P., Heimberg, H., and Mansouri, A. (2009). The ectopic expression of Pax4 in the mouse pancreas converts progenitor cells into alpha and subsequently beta cells. Cell *138*, 449–462.

Dorrell, C., Schug, J., Lin, C.F., Canaday, P.S., Fox, A.J., Smirnova, O., Bonnah, R., Streeter, P.R., Stoeckert, C.J., Jr., Kaestner, K.H., and

Grompe, M. (2011). Transcriptomes of the major human pancreatic cell types. Diabetologia *54*, 2832–2844.

Dupuis, J., Langenberg, C., Prokopenko, I., Saxena, R., Soranzo, N., Jackson, A.U., Wheeler, E., Glazer, N.L., Bouatia-Naji, N., Gloyn, A.L., et al; DIAGRAM Consortium; GIANT Consortium; Global BPgen Consortium; Anders Hamsten on behalf of Procardis Consortium; MAGIC investigators. (2010). New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. Nat. Genet. *42*, 105–116.

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., et al. (2010). A map of open chromatin in human pancreatic islets. Nat. Genet. *42*, 255–259.

Gloyn, A.L., Pearson, E.R., Antcliff, J.F., Proks, P., Bruining, G.J., Slingerland, A.S., Howard, N., Srinivasan, S., Silva, J.M., Molnes, J., et al. (2004). Activating mutations in the gene encoding the ATP-sensitive potassium-channel subunit Kir6.2 and permanent neonatal diabetes. N. Engl. J. Med. *350*, 1838–1849.

Gupta, R.A., Shah, N., Wang, K.C., Kim, J., Horlings, H.M., Wong, D.J., Tsai, M.C., Hung, T., Argani, P., Rinn, J.L., et al. (2010). Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. Nature *464*, 1071–1076.

Guttman, M., Amit, I., Garber, M., French, C., Lin, M.F., Feldser, D., Huarte, M., Zuk, O., Carey, B.W., Cassady, J.P., et al. (2009). Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. Nature *458*, 223–227.

Guttman, M., Donaghey, J., Carey, B.W., Garber, M., Grenier, J.K., Munson, G., Young, G., Lucas, A.B., Ach, R., Bruhn, L., et al. (2011). lincRNAs act in the circuitry controlling pluripotency and differentiation. Nature *477*, 295–300.

Guttman, M., and Rinn, J.L. (2012). Modular regulatory principles of large non-coding RNAs. Nature *482*, 339–346.

Halban, P.A., Kahn, S.E., Lernmark, A., and Rhodes, C.J. (2001). Gene and cell-replacement therapy in the treatment of type 1 diabetes: how high must the standards be set? Diabetes *50*, 2181–2191.

Heid, I.M., Jackson, A.U., Randall, J.C., Winkler, T.W., Qi, L., Steinthorsdottir, V., Thorleifsson, G., Zillikens, M.C., Speliotes, E.K., Mägi, R., et al; MAGIC. (2010). Meta-analysis identifies 13 new loci associated with waist-hip ratio and reveals sexual dimorphism in the genetic basis of fat distribution. Nat. Genet. *42*, 949–960.

Hu, W., Yuan, B., Flygare, J., and Lodish, H.F. (2011). Long noncoding RNA-mediated anti-apoptotic activity in murine erythroid terminal differentiation. Genes Dev. *25*, 2573–2578.

Kooner, J.S., Saleheen, D., Sim, X., Sehmi, J., Zhang, W., Frossard, P., Been, L.F., Chia, K.S., Dimas, A.S., Hassanali, N., et al; DIAGRAM; MuTHER. (2011). Genome-wide association study in individuals of South Asian ancestry identifies six new type 2 diabetes susceptibility loci. Nat. Genet. *43*, 984–989.

Kroon, E., Martinson, L.A., Kadoya, K., Bang, A.G., Kelly, O.G., Eliazer, S., Young, H., Richardson, M., Smart, N.G., Cunningham, J., et al. (2008). Pancreatic endoderm derived from human embryonic stem cells generates glucose-responsive insulin-secreting cells in vivo. Nat. Biotechnol. *26*, 443–452.

Lango Allen, H., Estrada, K., Lettre, G., Berndt, S.I., Weedon, M.N., Rivadeneira, F., Willer, C.J., Jackson, A.U., Vedantam, S., Raychaudhuri, S., et al. (2010). Hundreds of variants clustered in genomic loci and biological pathways affect human height. Nature *467*, 832–838.

Loewer, S., Cabili, M.N., Guttman, M., Loh, Y.-H., Thomas, K., Park, I.H., Garber, M., Curran, M., Onder, T., Agarwal, S., et al. (2010). Large intergenic non-coding RNA-RoR modulates reprogramming of human induced pluripotent stem cells. Nat. Genet. *42*, 1113–1117.

Mattick, J.S., and Makunin, I.V. (2006). Non-coding RNA. Hum. Mol. Genet. *15* (Spec No 1), R17–R29.

McCarthy, M.I. (2010). Genomics, type 2 diabetes, and obesity. N. Engl. J. Med. *363*, 2339–2350.

Mercer, T.R., Dinger, M.E., Sunkin, S.M., Mehler, M.F., and Mattick, J.S. (2008). Specific expression of long noncoding RNAs in the mouse brain. Proc. Natl. Acad. Sci. USA *105*, 716–721.

Pauli, A., Valen, E., Lin, M.F., Garber, M., Vastenhouw, N.L., Levin, J.Z., Fan, L., Sandelin, A., Rinn, J.L., Regev, A., and Schier, A.F. (2012). Systematic identification of long noncoding RNAs expressed during zebrafish embryogenesis. Genome Res. *22*, 577–591.

Penny, G.D., Kay, G.F., Sheardown, S.A., Rastan, S., and Brockdorff, N. (1996). Requirement for Xist in X chromosome inactivation. Nature *379*, 131–137.

Piper, K., Brickwood, S., Turnpenny, L.W., Cameron, I.T., Ball, S.G., Wilson, D.I., and Hanley, N.A. (2004). Beta cell differentiation during early human pancreas development. J. Endocrinol. *181*, 11–23.

Ponjavic, J., Oliver, P.L., Lunter, G., and Ponting, C.P. (2009). Genomic and transcriptional co-localization of protein-coding and long non-coding RNA pairs in the developing brain. PLoS Genet. *5*, e1000617.

Ravassard, P., Hazhouz, Y., Pechberty, S., Bricout-Neveu, E., Armanet, M., Czernichow, P., and Scharfmann, R. (2011). A genetically engineered human pancreatic β cell line exhibiting glucose-inducible insulin secretion. J. Clin. Invest. *121*, 3589–3597.

Rinn, J.L., Kertesz, M., Wang, J.K., Squazzo, S.L., Xu, X., Brugmann, S.A., Goodnough, L.H., Helms, J.A., Farnham, P.J., Segal, E., and Chang, H.Y. (2007). Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. Cell *129*, 1311–1323.

Scharfmann, R., Xiao, X., Heimberg, H., Mallet, J., and Ravassard, P. (2008). Beta cells within single human islets originate from multiple progenitors. PLoS ONE *3*, e3559.

Schuit, F., Flamez, D., De Vos, A., and Pipeleers, D. (2002). Glucose-regulated gene expression maintaining the glucose-responsive state of beta-cells. Diabetes *51* (*Suppl 3*), S326–S332.

Segrè, A.V., Groop, L., Mootha, V.K., Daly, M.J., and Altshuler, D.; DIAGRAM Consortium; MAGIC investigators. (2010). Common inherited variation in mitochondrial genes is not enriched for associations with type 2 diabetes or related glycemic traits. PLoS Genet. *6*, 6.

Senée, V., Chelala, C., Duchatelet, S., Feng, D., Blanc, H., Cossec, J.C., Charon, C., Nicolino, M., Boileau, P., Cavener, D.R., et al. (2006). Mutations in GLIS3 are responsible for a rare syndrome with neonatal diabetes mellitus and congenital hypothyroidism. Nat. Genet. *38*, 682–687.

Servitja, J.M., and Ferrer, J. (2004). Transcriptional networks controlling pancreatic development and beta cell function. Diabetologia *47*, 597–613.

Sleutels, F., Zwart, R., and Barlow, D.P. (2002). The non-coding Air RNA is required for silencing autosomal imprinted genes. Nature *415*, 810–813.

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., et al; DIAGRAM Consortium; GIANT Consortium; MuTHER Consortium; CARDIoGRAM Consortium; C4D Consortium. (2011). 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.

Ulitsky, I., Shkumatava, A., Jan, C.H., Sive, H., and Bartel, D.P. (2011). Conserved function of lincRNAs in vertebrate embryonic development despite rapid sequence evolution. Cell *147*, 1537–1550.

van Arensbergen, J., García-Hurtado, J., Moran, I., Maestro, M.A., Xu, X., Van de Casteele, M., Skoudy, A.L., Palassini, M., Heimberg, H., and Ferrer, J. (2010). Derepression of Polycomb targets during pancreatic organogenesis allows insulin-producing beta-cells to adopt a neural gene activity program. Genome Res. *20*, 722–732.

van de Bunt, M., and Gloyn, A.L. (2010). From genetic association to molecular mechanism. Curr. Diab. Rep. *10*, 452–466.

Van Hoof, D., D'Amour, K.A., and German, M.S. (2009). Derivation of insulin-producing cells from human embryonic stem cells. Stem Cell Res. (Amst.) *3*, 73–87.

Voight, B.F., Scott, L.J., Steinthorsdottir, V., Morris, A.P., Dina, C., Welch, R.P., Zeggini, E., Huth, C., Aulchenko, Y.S., Thorleifsson, G., et al; MAGIC investigators; GIANT Consortium. (2010). Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. Nat. Genet. *42*, 579–589.

Wang, K.C., and Chang, H.Y. (2011). Molecular mechanisms of long noncoding RNAs. Mol. Cell *43*, 904–914.

Wilusz, J.E., Sunwoo, H., and Spector, D.L. (2009). Long noncoding RNAs: functional surprises from the RNA world. Genes Dev. *23*, 1494–1504.

Xu, Z., Wei, W., Gagneur, J., Clauder-Münster, S., Smolik, M., Huber, W., and Steinmetz, L.M. (2011). Antisense expression increases gene expression variability and locus interdependency. Mol. Syst. Biol. *7*, 468.

Zeggini, E., Weedon, M.N., Lindgren, C.M., Frayling, T.M., Elliott, K.S., Lango, H., Timpson, N.J., Perry, J.R., Rayner, N.W., Freathy, R.M., et al; Wellcome Trust Case Control Consortium (WTCCC). (2007). Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. Science *316*, 1336–1341.

Zhou, Q., Brown, J., Kanarek, A., Rajagopal, J., and Melton, D.A. (2008). In vivo reprogramming of adult pancreatic exocrine cells to beta-cells. Nature *455*, 627–632.