Islet Long Noncoding RNAs: A Playbook for Discovery and Characterization

Ruth A. Singer[1,2] and Lori Sussel[1,3]

Diabetes 2018;67:1461–1470 | https://doi.org/10.2337/dbi18-0001

Diabetes is a complex group of metabolic disorders that can be accompanied by several comorbidities, including increased risk of early death. Decades of diabetes research have elucidated many genetic drivers of normal islet function and dysfunction; however, a lack of suitable treatment options suggests our knowledge about the disease remains incomplete. The establishment of long noncoding RNAs (lncRNAs), once dismissed as "junk" DNA, as essential gene regulators in many biological processes has redefined the central role for RNA in cells. Studies showing that misregulation of lncRNAs can lead to disease have contributed to the emergence of lncRNAs as attractive candidates for drug targeting. These findings underscore the need to reexamine islet biology in the context of a regulatory role for RNA. This review will *1*) highlight what is known about lncRNAs in the context of diabetes, *2*) summarize the strategies used in lncRNA discovery pipelines, and *3*) discuss future directions and the potential impact of studying the role of lncRNAs in diabetes.

Diabetes represents a complex set of diseases with genetic, immunological, and environmental etiologies. Although the exact pathophysiology differs for each diabetes subtype, the disease generally results from the failure of pancreatic β-cells to meet the insulin demands required for glucose homeostasis. In recent years, there has been an unprecedented increase in diabetes prevalence worldwide. This alarming trend, coupled with limited treatment options and severe comorbidities, highlights the urgent need to address gaps in our understanding of the genetic and molecular mechanisms of diabetes.

Decades of mouse research and advances in genome-wide association studies have identified several genetic drivers of monogenic syndromes of β-cell dysfunction, as well as 113 distinct type 2 diabetes (T2D) susceptibility loci (1) and ∼60 loci associated with an increased risk of developing type 1 diabetes (T1D) (2). Interestingly, these studies discovered that most T1D and T2D susceptibility loci fall outside of coding regions, which suggests a role for noncoding elements in the development of disease (3,4). Several studies have demonstrated that many causal variants of diabetes are significantly enriched in regions containing islet enhancers, promoters, and transcription factor binding sites (5,6); however, not all diabetes susceptibility loci can be explained by associations with these regulatory regions. This highlights our incomplete understanding of the islet regulome and the need for detailed functional analyses of noncoding genes to precisely determine their contribution to diabetes susceptibility and disease progression.

NONCODING RNAS

Advances in RNA sequencing (RNA-seq) technologies have revealed that mammalian genomes encode tens of thousands of RNA transcripts that have similar features to mRNAs, yet are not translated into proteins (7). Although the percentage of these nonprotein-coding RNAs that are functional is still unknown, detailed characterization of many of these transcripts has challenged the idea that the central role for RNA in a cell is to give rise to proteins. Instead, these RNA transcripts make up a class of molecules called noncoding RNAs (ncRNAs) that function either as "housekeeping" ncRNAs, such as transfer RNAs (tRNAs) and ribosomal RNAs (rRNAs), that are expressed ubiquitously and are required for protein synthesis or as "regulatory" ncRNAs that control gene expression. While the functional mechanisms of short regulatory ncRNAs, such as microRNAs (miRNAs), small interfering RNAs (siRNAs), and Piwi-interacting RNAs

[1]Columbia University Medical Center, New York, NY
[2]The Integrated Graduate Program in Cellular, Molecular and Biomedical Studies, Graduate School of Arts and Sciences, Columbia University Medical Center, New York, NY
[3]Barbara Davis Center for Diabetes, University of Colorado Anschutz Medical Campus, Aurora, CO
Corresponding author: Lori Sussel, lori.sussel@ucdenver.edu.

Received 5 February 2018 and accepted 20 March 2018.

See accompanying article, p. 1471.

(piRNAs), have been described in detail (8–10), the most abundant and functionally enigmatic regulatory ncRNAs are called long noncoding RNAs (lncRNAs) that are loosely defined as RNAs larger than 200 nucleotides (nt) that do not encode for protein (11–13). Although using a definition based strictly on size is somewhat arbitrary, this definition is useful both bioinformatically (the need for an effective size gap to distinguish lncRNAs from the ~20-nt short ncRNAs) and technically (200 nt is approximately the retention size of most silica-based columns used in standard RNA extraction kits) (14). While the 200-nt size cutoff has simplified identification of lncRNAs, this rather broad classification means several features of lncRNAs, including abundance, cellular localization, stability, conservation, and function, are inherently heterogeneous (15–17). Although this represents one of the major challenges of lncRNA biology, it also highlights the untapped potential of lncRNAs to provide a novel layer of gene regulation that influences islet physiology and pathophysiology.

Major efforts to unify the lncRNA field by clarifying nomenclature and identifying common attributes of lncRNAs have yielded several important conclusions about their shared characteristics (reviewed in ref. 17). In general, despite a lack of translatable open reading frames, most lncRNAs are biochemically indistinguishable from their mRNA counterparts: they are transcribed by RNA polymerase II (Pol II) from genetic loci that contain classic active chromatin marks at their promoters (H3K4me3) and gene bodies (H3K36me3), and they are 5′-capped, spliced, and polyadenylated (18). Beyond these common characteristics, lncRNAs represent a heterogeneous population of functional RNAs with diverse functions that have yet to be fully explored. As the lncRNA field has grown, several databases, such as NONCODE (19), lncRNAdb (20), and LNCipedia (21), have emerged to provide researchers with catalogs of empirically identified lncRNAs (Fig. 1). Additional attempts to categorize the myriad of newly identified lncRNAs have yielded several classifications based on genomic proximity to the nearest protein-coding gene (22). Although it can be useful to have classifications that do not rely on functional annotation, the genomic contexts of lncRNAs do not necessarily provide insights into function. For example, two well-characterized lncRNAs, *HOTTIP* (HOXA transcript at the distal tip) and *HOTAIR* (HOX transcript antisense RNA), are both transcribed from the HOXA gene locus; however, *HOTTIP* regulates a nearby HOXA gene in *cis,* whereas *HOTAIR* recruits chromatin-modifying complexes to other chromosomes in *trans* (23,24). Functional classification based on sequence alone has also been difficult because lncRNAs are not under the same evolutionary pressure as proteins to convey a genetic code (25). Instead, information from the small number of well-characterized lncRNAs has been distilled to describe four nonmutually exclusive archetypes of lncRNA molecular mechanisms: signals, decoys, guides, and scaffolds (26). These archetypes emphasize the unique ability of lncRNAs to bind to DNA, RNA, and proteins to regulate all layers of gene expression. As more lncRNAs undergo functional characterization, improved classifications will enable predictive modeling of lncRNA function.

lncRNAs AND DIABETES

Although the role of miRNAs in diabetes has been well established (9), analyses of lncRNAs in islets have lagged behind their short ncRNA counterparts. However, several recent studies provide evidence that lncRNAs are crucial components of the islet regulome and may have a role in diabetes (27). For example, a genome-wide association study examined T2D susceptibility loci within unknown genomic associations and found that a significant percentage (>16%) of T2D loci contained islet lncRNAs within 150 kb of the reported lead single nucleotide polymorphism (SNP) (28), suggesting lncRNAs are essential for normal pancreatic function. Furthermore, misexpression of several lncRNAs has been correlated with diabetes complications, such as diabetic nephropathy and retinopathy (29–31). There are also preliminary studies suggesting that circulating lncRNAs, such as Gas5, MIAT1, and SENCR, may represent effective molecular biomarkers of diabetes and diabetes-related complications (32,33). Finally, several recent studies have explored the role of lncRNAs in the peripheral metabolic tissues that contribute to energy homeostasis (reviewed in ref. 34).

In addition to their potential as genetic drivers and/or biomarkers of diabetes and diabetes complications, lncRNAs can be exploited for the treatment of diabetes. For example, although tremendous efforts have been dedicated to generating replacement β-cells for individuals with diabetes (35,36), human pluripotent stem cell–based β-cell differentiation protocols remain inefficient, and the end product is still functionally and transcriptionally immature compared with primary human β-cells (reviewed in ref. 37). This is largely due to our incomplete knowledge of in vivo differentiation regulatory pathways, which likely include a role for lncRNAs. Once we gain additional understanding of lncRNA function during the pancreatic endocrine differentiation process, we can incorporate lncRNAs into the in vitro differentiation protocols to optimize the generation of β-cells derived from human pluripotent stem cells.

Inherent characteristics of lncRNAs have also made them attractive candidates for drug targeting, which could be exploited for developing new diabetes therapies. lncRNAs can be targeted by antisense oligonucleotides (ASOs), which is technically simpler than small molecule screening or inhibitory antibody development. Also, as strategies to therapeutically target mRNA molecules date back to the 1970s, developing next-generation technologies to target lncRNAs is quite feasible and easily adapted from traditional mRNA targeting approaches. For example, several currently available biochemical modifications to ASOs, such as a locked nucleic acid (LNA) backbone, increase their stability and reduce toxicity (38). Unique features of lncRNAs also make them more amenable to targeting, including their highly tissue-specific expression and their relatively lower expression levels as compared with coding mRNAs, which may allow the use of

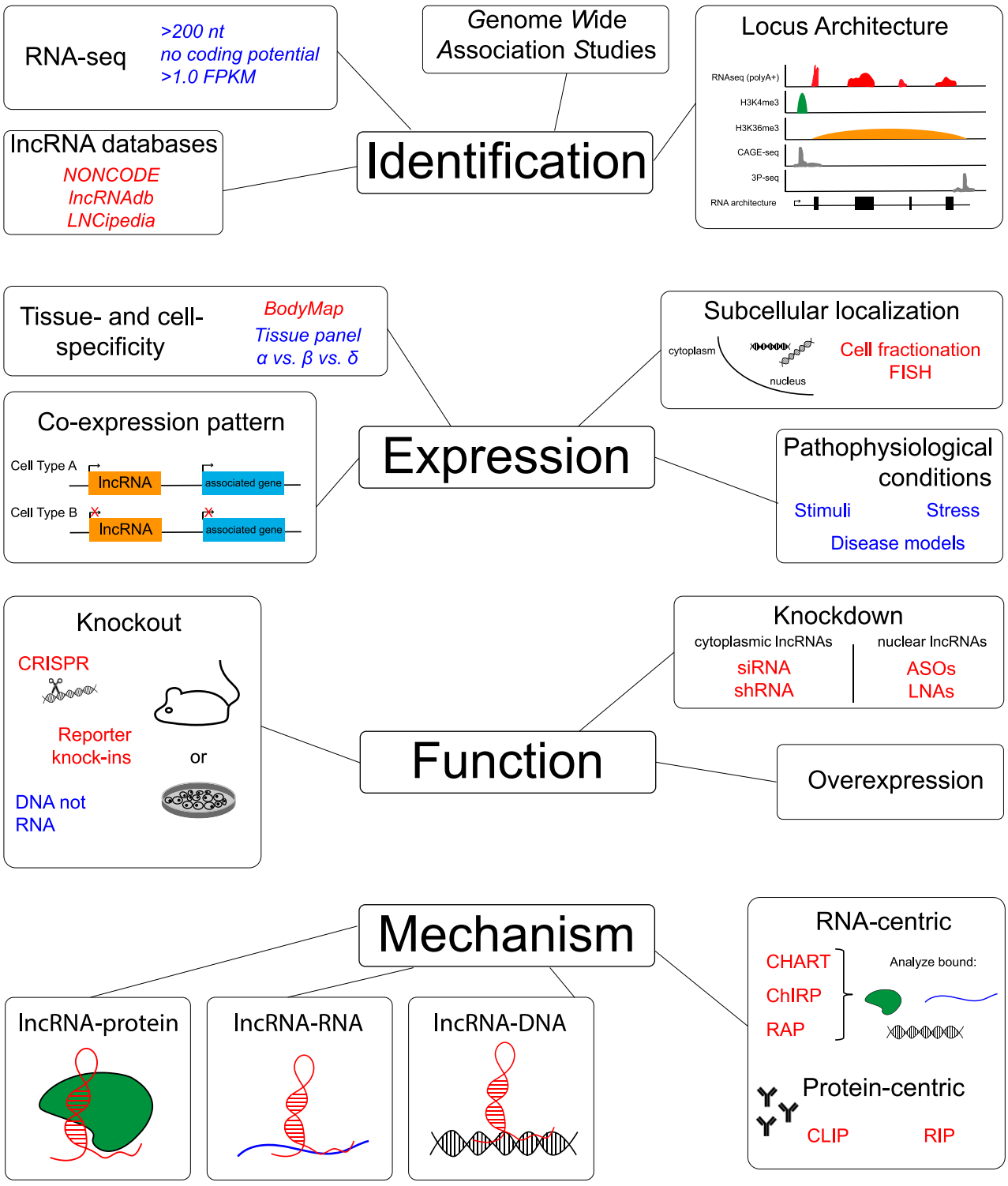

Figure 1—Overview of the lncRNA discovery and characterization pipeline. The lncRNA pipeline outlined in this review flows through four main phases: *1*) identification of lncRNAs, *2*) expression analyses, *3*) functional interrogation, and *4*) determination of a regulatory mechanism. Highlighted are the tools and techniques (red), experimental parameters to consider (blue), and general strategies to characterize functional lncRNAs. 3P-seq, poly(A)-position profiling by sequencing; CAGE-seq; cap analysis gene expression; FPKM, fragments per kilobase million.

lower doses of targeting molecules, thus alleviating drug toxicity (11,14). A study in human pancreatic islets found that although 9.4% of Ref-seq–annotated protein-coding genes were islet specific, 55% of intergenic lncRNAs were expressed only in islets (28). Taken together, these findings suggest that lncRNAs could be easily targeted and their restricted expression profile would result in fewer unwanted pleiotropic phenotypes.

FROM DISCOVERY TO FUNCTION

Identifying lncRNAs

Although the rapid evolution of the lncRNA field has provided remarkable insight into lncRNA functions in many different organ systems and disease states (11,14,18,22), these studies have also revealed important aspects of lncRNA biology that need be considered to optimize a discovery pipeline (Fig. 1). Unlike the discovery of novel coding genes, whose functions can be inferred from the functional protein domains they encode, a different set of strategies must be used for lncRNAs. To optimally characterize and exploit lncRNAs for treating diabetes, it will be useful to develop a common set of guidelines to most effectively identify those lncRNAs that are essential for pancreatic function. With this goal in mind, we will discuss studies that have identified mouse and human pancreatic lncRNAs, highlight the major steps of lncRNA discovery pipelines for diabetes-related lncRNAs, and identify the advantages and disadvantages of each approach (Table 1). The information gained from these pioneering studies will inform the identification of functionally relevant islet-specific lncRNAs that can be exploited to promote the efficient generation or regeneration of β-cells or can be targeted for novel therapies to treat diabetes.

The first systematic mapping of lncRNAs in the pancreas was described in 2012; Morán et al. (28) performed deep RNA-seq on human islets in combination with chromatin immunoprecipitation sequencing for three epigenetic markers of active transcription: H3K4me3, H3K36me3, and RNA Pol II. This strategy made it possible to leverage the histone marks to identify 1,128 human islet lncRNAs that showed classic marks of active transcription (28). More recently, this same group reported a similar, updated set of parameters including the presence of H3K4me3 enrichment in a region +1 kb to −0.5 kb from a transcript's 5′ end to identify 2,226 human islet lncRNAs (39). Subsequently, Fadista et al. (40) used associations with disease loci to identify potential diabetes-related lncRNAs; genetic variants regulating gene expression, referred to as expression quantitative trait loci (eQTLs), that were implicated in altered glucose metabolism and insulin secretion determined that 33 out of 616 diabetic eQTLs influenced the expression of human islet lncRNAs. As the human diabetes field moves beyond exomic sequencing to whole-genome sequence analysis, it is likely that additional lncRNAs will be identified as contributing factors to the etiology of diabetes.

Although using human pancreatic islets as a source of RNA has the advantage of identifying lncRNAs most relevant for human disease (28,39–43), cadaveric islets are not only scarce but also often highly heterogeneous and have reduced RNA integrity, which can result in RNA artifacts and confound the identification of bona fide lncRNA molecules. Furthermore, the options for downstream functional analyses of human lncRNAs are limited. As an alternative, several lncRNA screens have been performed in mouse islets (44–46). This approach benefits from higher-quality starting material, increased sample homogeneity, and the ability to use cell-specific reporters to isolate purified islet cell populations. Rodent models also afford the option to interrogate lncRNA function in vivo. Furthermore, although lncRNAs are often poorly conserved at the nucleotide level between mice and humans, Morán et al. (28) reported that 70% of human islet lncRNAs had an orthologous mouse genomic region and that 47% of those orthologous mouse loci produced a corresponding lncRNA transcript. Currently, there is a paucity of information regarding the minimal amount of nucleotide conservation that would predict conserved function, especially as the conservation may be at the level of RNA secondary structure (47). However, the characterization of increased numbers of conserved orthologous lncRNAs will pave the way to understanding the contribution of lncRNA function to β-cell biology and inform whether lncRNAs that are identified and characterized in mice contribute to human β-cell function and disease pathologies.

Linking Expression to Function

As it remains difficult to use lncRNA nucleotide sequence as a predictor of functional activity, cell-specific and/or regulated expression can be used to predict lncRNA function. Although most transcriptional regulatory proteins are expressed in many different tissues and cell types, expression of functionally important lncRNAs tends to be much more restricted (14). This can present a challenge for the identification of lncRNAs as expression must be assessed in the correct cell type, but it also provides the exciting possibility that lncRNAs confer highly specialized cell-specific regulatory functions. To identify islet cell–type specific lncRNA transcripts, a study used β-cell–specific reporter mice (MIP:GFP) to compare the transcriptomes of MIP:GFP-positive mouse islet cells (β-cells) to MIP:GFP-negative (non–β-cell) islet cells and discovered ∼12% of islet lncRNAs were β-cell specific (44). Similarly, studies have identified differentially expressed lncRNAs in human α- versus β-cells (41) and at different stages of in vitro human pancreas differentiation (48). Given the highly cell-type specific expression of lncRNAs, it is tempting to speculate that these molecules provide the regulatory specificity to gene networks that define cell-specific identities and functions. For example, the lncRNA *βlinc1* is expressed exclusively in pancreatic β-cells, where it appears to regulate a β-cell–specific regulatory program (49). Coexpression and/or cross-regulation of lncRNAs with a nearby protein-coding gene can also provide important functional and regulatory information; the human islet lncRNA *PLUTO* had a highly correlated tissue-specific expression pattern with its neighboring gene, the essential pancreatic transcription factor PDX1, and was further shown to directly regulate PDX1 transcription (39).

Altered expression in pathophysiological conditions could also be a defining characteristic in the identification of functionally important lncRNAs. The lncRNAs *Meg3* and *Tug1* were both shown, in separate studies, to be downregulated in the NOD T1D mouse model (50,51). Similarly,

Table 1—Identification of islet lncRNAs

Tissue/cell type	lncRNAs identified	lncRNA annotation parameters	Reference
Human and mouse islets; human purified β-cells	1,128 lncRNAs	>200-nt long; overlap with a H3K4me3 peak; >0.5 RPKM; lack of evidence for splicing to any annotated coding gene	Morán et al., 2012 (28)
Mouse β- and non–β-cell islet endocrine cells	2,790 transcripts, 2,425 loci, 1,342 novel genes	No overlap with annotated gene or 5′ or 3′ UTR of a Ref-seq protein-coding gene; >200-nt long; PhyloCSF score <0	Ku et al., 2012 (44)
FACS purified human β- and α-cells	12 β-cell–specific and 5 α-cell–specific noncoding transcript	>200-nt long; >1 RPKM; no overlap with the UCSC Repeat Masker track or other ncRNAs; within 5 kb of a H3K4me3 peak	Bramswig et al., 2013 (41)
Human islets; β-cells; non–β-cell islet endocrine cells	280 lncRNAs enriched in β-cells compared with whole islets or non–β-endocrine cells	Limited analyses to annotated lncRNAs in Morán et al. (28)	Nica et al., 2013 (43)
Human islets exposed to inflammatory cytokines	262 lncRNAs, 177 novel, none differentially expressed	>200-nt long; >1 exon; negative CPAT score; novel lncRNAs did not match with known human lncRNA sequences in the NONCODE database	Li et al., 2014 (42)
89 human islet samples	493 Ref-seq lncRNAs expressed in human islets, 54 lncRNAs with HbA_{1c} eQTLs	No overlap with any annotated genes; >1 exon; expressed (nonnull read coverage) in at least 5% of the samples; negative CPAT score	Fadista et al., 2014 (40)
FACS purified mouse β- and α-cells	145 novel lncRNAs	No overlap with any known Ref-seq or UCSC gene; transcript length >3 kb; >1 RPKM; no overlap with known rRNA loci; PhastCons score >0.4	Benner et al., 2014 (45)
41 human islet samples	2,226 de novo lncRNAs	H3K4me3 enrichment +1 kb to −0.5 kb from the transcript 5′ end; CPAT score <0.364; no sense overlap with a coding exon; >0.05 FPKM in β-cells; acinar to β-cell expression ratio <3	Akerman et al., 2017 (39)
Islets from mice fed a regular or high-fat diet	1,761 annotated lncRNAs, 1,558 novel lncRNAs	>200-nt long; at least 2 exons; FPKM >5; GeneID v1.4.4 coding potential score <4	Motterle et al., 2017 (46)

CPAT, coding potential assessment tool (68); FACS, fluorescent automated cell sorting; FPKM, fragments per kilobase million; PhastCons, phylogenetic analysis with space/time models (Phast) conservation (Cons) (69); PhyloCSF, phylo codon substitution frequencies (70); RPKM, reads per kilobase million; UCSC, University of California Santa Cruz; UTR, untranslated region.

a study in human islets found that two lncRNAs *KCNQ1OT1* and *HI-LNC45* were significantly upregulated or downregulated in T2D islets, respectively (28). Several groups also identified lncRNAs misregulated in islets exposed to altered stimuli or physiological challenges, such as high glucose (45), high-fat diet (46), cytokines (52), and pregnancy (53). Interestingly, although exposure of mouse islets to inflammatory cytokines caused upregulation of four lncRNAs (42), a study on human islets cultured in the presence or absence of cytokines found no differential expression of lncRNAs between the two conditions (52), suggesting these analyses may be influenced by low lncRNA expression levels and detection limits of the assay.

Subcellular localization can yield perhaps the most telling clues about a lncRNA's mechanism of action. Nuclear lncRNAs are more likely to regulate transcription, whereas cytoplasmic lncRNAs more often influence translation or mRNA stability (26). An individual lncRNA can also be expressed in both the nucleus and cytoplasm and have a different function in each compartment (54). Although subcellular fractionation has traditionally been used to assess cellular localization, RNA fluorescent in situ hybridization (FISH) is a highly sensitive assay that can also provide crucial spatial information about lncRNA localization within the cell. For example, *DEANR1* was shown to directly regulate the transcription factor FOXA2 in differentiated endoderm by using RNA-DNA-FISH experiments showing that *DEANR1* localized to two punctae corresponding to the FOXA2 DNA locus (48).

Although expression analyses can be performed using a combination of quantitative PCR and RNA in situ hybridization, in silico discovery of islet lncRNAs is also possible as a result of the comprehensive mapping of mouse α-, β-, and δ-cell transcriptome (55). Additionally, global expression information can often be obtained in silico through publically available data sets, such as BodyMap (56); however, these resources rarely include pancreas and/or islet RNA in their arrays due to challenges associated with pancreatic RNA integrity. Furthermore, standard RNA-seq parameters often lack the coverage needed to identify low abundant lncRNAs, many of which remain unannotated. Another obstacle to lncRNA discovery is the paucity of relevant tissue, given that islets make up only about 5–10% of the entire pancreas. As new molecular technologies with increased detection sensitivities are developed, these challenges will be overcome; however, this will also increase the importance of functional validation for newly identified pancreatic lncRNAs.

Functional Characterization of Islet lncRNAs

With the advancement of high-throughput sequencing techniques, the list of islet-specific lncRNAs is growing exponentially; however, functional characterization is missing for the majority of these lncRNAs. Studies that have experimentally determined the function of an islet lncRNA have used several different approaches (Table 2). The most straightforward strategy to test the regulatory function of islet lncRNAs is by loss-of-function assays using RNA interference (RNAi). Knockdown (KD) of the lncRNA *βlinc1* in MIN6 cells demonstrated that *βlinc1* is a novel *cis* regulator of the islet transcription factor Nkx2–2 (49), whereas short hairpin RNA (shRNA) KD of the lncRNA *DEANR1* led to a drastic downregulation of the nearby gene *FOXA2* (48). In vitro RNAi is also the optimal technique to identify lncRNAs that function in *trans* to regulate essential islet genes: HI-LNC25 was shown to regulate the distant islet transcription factor GLIS3 in EndoC-βH1 cells (28), the lncRNA *Meg3* positively regulated *Pdx1* and *MafA* in MIN6 cells (51), and the *lncRNA-ROR* regulated *Insulin*, *Pdx1*, and *Glut2* in human amniotic epithelial cells differentiated into β-like cells (57). Furthermore, KD of lncRNA *Tug1* in NIT-1 cells was correlated with changes in β-cell function, including decreased glucose-stimulated insulin secretion (50). In a more comprehensive screen, Akerman et al. (39) functionally interrogated 12 human lncRNAs with lentiviral vectors containing Pol II–transcribed artificial miRNAs (coined amiRNAs) with perfect homology to the target lncRNA to elicit degradation via the RNAi pathway. Remarkably, KD of 9 out of 12 lncRNAs in EndoC-βH1 cells elicited significant gene expression changes, and KD of three of those lncRNAs led to impaired insulin secretion (39).

Although siRNAs and shRNAs are well suited to downregulating the expression and/or translation of protein-coding genes, there is concern in the field about the use of cytoplasmic RNAi machinery to KD nuclear lncRNAs (58). Modified ASOs, such as LNA GapmeRs, are a valuable alternative for lncRNAs enriched in the nucleus. They are stable high-affinity RNA analogs that readily permeate the nucleus and function by RNase H–dependent degradation of complementary RNA targets (59). Gain-of-function experiments, mainly lncRNA overexpression, should also be considered when overexpression more closely mimics an endogenous or diseased state, as was the case for four lncRNAs shown to be upregulated in MIN6 cells exposed to inflammatory cytokines (52).

Although in vitro functional analysis has been informative, there are caveats associated with using immortalized cell lines. For example, several lncRNAs with significant functions in vitro have had no phenotype when knocked out (KO) in mice (58). This discrepancy may be due to improper use of KD tools or because lncRNA KO phenotypes can be subtle and may only appear after physiological stress. Genetic manipulation is the optimal approach for assessing lncRNA function in an endogenous in vivo context, although these studies face their own challenges. There are several commonly used gene-targeting strategies that vary in their efficiency and disruption of genomic contexts (i.e., enhancers), including deletions (whole gene or promoter), inversions (whole gene or promoter), and insertions (premature termination sequence or reporter) (58). To our knowledge, *βlinc1* is the first islet lncRNA that has been genetically disrupted at the DNA level to generate a KO mouse with no detectable *βlinc1* RNA (49). *βlinc1* KO mice exhibited impaired glucose tolerance due to defects in insulin

Table 2—Characterization of functional islet lncRNAs

lncRNAs	Expression	Function	Reference
PVT1	Human mesangial cells	PVT1 mediates the development and progression of diabetic nephropathy through mechanisms involving extracellular matrix accumulation.	Alvarez et al., 2011 (29)
HI-LNC25	Enriched in human islets and purified β-cells compared with other tissues	Positively regulates GLIS3 (which contains both T1D and T2D risk variants) in EndoC-βH1 human β-cell line.	Morán et al., 2012 (28)
LINC01611	Human islets from patients with T2D	SNP (rs9362054) associated with diabetic retinopathy mapped to this lncRNA.	Awata et al., 2014 (31)
MALAT1	Upregulated in retinas of diabetic (streptozotocin or *db*/*db*) mice	Upregulation of MALAT1 is associated with microvascular dysfunction. MALAT1 KD alleviates diabetic retinopathy.	Liu et al., 2014 (66)
DEANR1	Definitive endoderm differentiated from human embryonic stem cells	Positively regulates expression of the endoderm factor FOXA2 by facilitating SMAD2/3 recruitment to the FOXA2 promoter.	Jiang et al., 2015 (48)
lncRNAs 1, 2, 3, 4	Increased expression in MIN6 cells exposed to cytokines and NOD mice	Overexpression in MIN6 cells increased apoptosis (lncRNAs 1–4) and caused nuclear translocation of p65 (lncRNA 1).	Motterle et al., 2015 (52)
Tug1	Enriched in mouse islets compared with other tissues, downregulated in NOD islets	Downregulation of lncRNA TUG1 expression increased apoptosis and reduced insulin secretion in mouse β-cells.	Yin et al., 2015 (50)
βlinc1 (HI-LNC15)	Restricted expression in mouse β-cells	Regulates β-cell identity and function in vivo, partially through the regulation of its neighboring gene Nkx2–2.	Arnes et al., 2016 (49)
NONHSAG011351	Human islets with T1D-associated SNPs	rs705708 had a *cis*-eQTL effect on both ERBB3 and NONHSAG011351, causing reduced expression of both genes.	Kaur et al., 2016 (67)
Meg3	Enriched in mouse islets compared with other tissues, downregulated in *db*/*db* islets	KD in MIN6 cells causes impaired β-cell function and increased apoptosis.	You et al., 2016 (51)
lncRNA-ROR	Reduced expression during differentiation of human amniotic epithelial cells into β-like cells	Loss of lncRNA-ROR during differentiation impairs β-cell function and reduced expression of Insulin, Pdx1, and Glut2.	Zou et al., 2016 (57)
PLUTO (HI-LNC71)	Enriched in human islets and purified β-cells relative to the exocrine pancreas and nonpancreatic tissues	Regulates transcription of the nearby gene PDX1 and influences 3D chromatin structure surrounding PDX1.	Akerman et al., 2017 (39)
TUNAR (HI-LNC78)	Enriched in human islets and purified β-cells relative to the exocrine pancreas and nonpancreatic tissues	KD of TUNAR caused impaired glucose-stimulated insulin secretion.	Akerman et al., 2017 (39)
βlinc2, βlinc3	Restricted expression in mouse β-cells, increased (βlinc2) or decreased (βlinc3) expression in mice given high-fat diet and *db*/*db* islets	Upregulation of βlinc2 caused increased apoptosis in MIN6 cells and isolated islets.	Motterle et al., 2017 (46)

secretion (49). A major caveat of gene deletion, however, is the possibility that disruption of the genomic DNA, not the RNA, is responsible for any observed phenotype. Additional KD experiments were therefore required to show that the in vivo phenotype was not due to deletion of an enhancer for the nearby gene Nkx2–2 (49). As technologies for in vivo gene editing improve and become more widely implemented, it is likely that we will see a significant increase in lncRNA

in vivo functional studies in complex animal models to further solidify their important functions in islet biology and diabetes.

Molecular Characterization of lncRNAs

Tens of thousands of lncRNAs have been identified in different cell types and model organisms; however, their functions largely remain unknown. Although the tools for determining lncRNA function are technically restrictive, uncovering novel regulatory mechanisms will have the greatest impact on understanding islet function and identifying novel therapeutics for diabetes. To date, no biochemical assay has been used to directly determine the molecular mechanisms by which islet lncRNAs function, which highlights both the infancy of the field and the difficulty in implementing these techniques. The different lncRNA regulatory subtypes represent the mechanisms by which lncRNAs can act on DNA in *cis* or in *trans*, bind to complementary mRNA molecules to influence their translation, and recruit proteins to either enable or prevent their function (26). Based on lncRNA studies in other tissues, the most straightforward way to characterize the molecular activity of a lncRNA is to identify its interacting partners, using either protein-centric or RNA-centric approaches. Protein-centric methods, such as RNA immunoprecipitation (RIP) (60) or cross-linking immunoprecipitation (CLIP) (61), use antibodies to immunoprecipitate RNA binding protein complexes from cellular homogenate in vivo. lncRNAs stably associated with these proteins, either directly (native RIP) or indirectly (CLIP), can be extracted and measured by quantitative PCR or nonbiasedly identified by RNA-seq. If empirical evidence suggests lncRNA interaction with a specific protein, then these techniques are feasible. However, when functional data do not indicate a role for specific protein interactions, an RNA-centric approach is the ideal a priori strategy to probe lncRNA regulatory mechanisms. As the number of identified lncRNAs climbed exponentially, so did the need for RNA-centric biochemical purification methods. To address this technology gap, three techniques emerged almost concurrently: Capture Hybridization Analysis of RNA Targets (CHART) (62), Chromatin Isolation by RNA Purification (ChIRP) (63), and RNA Antisense Purification (RAP) (64) (Table 3). Differences between the protocols largely pertain to the cross-linking method and probe design: CHART uses probes designed based on empirical evidence of RNase H accessibility, whereas ChIRP and RAP both tile the whole RNA molecule, albeit with different sized oligos, 20 mer and 120 nt, respectively. The most powerful aspect of these protocols is that once RNA pull-down is successful, the readout can be tailored for tandem analyses of the DNA, RNA, or protein bound to a lncRNA in either a systematic or candidate-driven approach. For example, a comprehensive study used CHART to pull down lncRNAs NEAT1 and MALAT1, followed by both DNA sequencing to identify genome-wide DNA binding sites and mass spectrometry to identify all interacting proteins (65).

Table 3—Biochemical tools for identifying lncRNA targets

Technique	Cross-linking method	Pull-down	Oligo design	Oligo modifications	Elution	Reference
CHART	Whole-cell fixation with 1% formaldehyde for 10′ and nuclei fixation with 3% formaldehyde for 30′	25-mer DNA antisense capture oligos	C-oligos designed to target RNase H sensitivity sites	Biotinylated at 3′ end with 18-carbon spacer arm	RNase H	Simon et al., 2011 (62)
ChIRP	Whole-cell fixation with 3% formaldehyde for 30′	20-mer DNA antisense tiling oligos	Tiling entire RNA sequence, split into "even" and "odd" sets	Biotinylated at 3′ end with 18-carbon spacer arm	RNase H or RNase A	Chu et al., 2011 (63)
RAP	Whole-cell fixation with 2 mmol/L disuccinimidyl glutarate for 45′, followed by 3% formaldehyde for 10′	120-nt RNA tiling oligos	Tiling every 15 nt across the entire RNA sequence, excluding repetitive regions	Probes transcribed in vitro in the presence of biotin–uridine triphosphate	Proteinase K	Engreitz et al., 2013 (64)

Although identification of a lncRNA's molecular function is theoretically straightforward, the paucity of studies using either protein-centric or RNA-centric techniques on islet lncRNAs exemplifies the challenges associated with these tools. A major source of difficulty associated with these techniques is that they were all developed using ubiquitous lncRNAs with relatively high endogenous expression levels, including *roX2* (CHART), *Xist* (RAP), and *roX2*, *TERC*, and *HOTAIR* (ChIRP) (62–64). Furthermore, these lncRNAs had previously characterized regulatory mechanisms, which meant positive controls were already available to troubleshoot RNA pull-downs. Conversely, the application of these techniques to characterize the molecular mechanism of novel cell-specific lncRNAs faces challenges associated with insufficient tissue (starting material), low abundant transcripts, and unknown binding partners. lncRNA overexpression or in vitro transcription may be an option when endogenous expression is too low; however, these approaches can also introduce experimental artifacts.

Due to the infancy of the lncRNA field, most of the biochemical and genetic tools used to interrogate lncRNA function have only recently been developed or are adapted from techniques used to study protein-coding genes and we are only beginning to appreciate the limits and challenges of borrowing strategies from the protein-coding world. Given the growing appreciation for lncRNAs in biology, it is likely that increased efforts will be made to adapt and optimize these technologies to enable mechanistic characterization of all functional lncRNAs, regardless of their abundance.

CONCLUSIONS

The discovery of lncRNAs as a novel class of tissue-specific regulatory molecules has spawned an exciting new field of biology that will significantly impact our understanding of pancreas physiology and pathophysiology. As the field continues to grow, there is growing appreciation that lncRNAs will provide many of the missing components to existing molecular pathways that regulate islet biology and contribute to diabetes when they become dysfunctional. However, to date, most of the experimental emphasis on lncRNAs has focused on large-scale discovery using genome-wide approaches, and there remains a paucity of functional analysis. With improved RNA-centric imaging and molecular technologies, combined with the advent of novel gene-editing tools, it is likely that our knowledge of lncRNA functions in the islet will expand exponentially to rival what is currently known about canonical transcriptional regulatory programs. These advances will pave the way to a greater understanding of pancreas biology and enable the development of better prediction tools and therapies for the treatment of diabetes.

Acknowledgments. The authors thank David Lorberbaum and other members of the Sussel laboratory for helpful discussions and critical reading of the manuscript.

Funding. Funding was provided by National Institute of Diabetes and Digestive and Kidney Diseases grants R01 DK082590 (L.S.), R01 DK111405 (L.S.), and F31 DK107028 (R.A.S.).

Duality of Interest. No potential conflicts of interest relevant to this article were reported.

Author Contributions. R.A.S. and L.S. wrote, reviewed, and edited the manuscript.

Prior Presentation. Parts of this article were presented at the 78th Scientific Sessions of the American Diabetes Association, Orlando, FL, 22–26 June 2018.

References

1. Scott RA, Scott LJ, Mägi R, et al.; DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) Consortium. An expanded genome-wide association study of type 2 diabetes in Europeans. Diabetes 2017;66:2888–2902
2. Mirza AH, Kaur S, Pociot F. Long non-coding RNAs as novel players in β cell function and type 1 diabetes. Hum Genomics 2017;11:17
3. McCarthy MI. Genomics, type 2 diabetes, and obesity. N Engl J Med 2010;363:2339–2350
4. Ricaño-Ponce I, Wijmenga C. Mapping of immune-mediated disease genes. Annu Rev Genomics Hum Genet 2013;14:325–353
5. Fuchsberger C, Flannick J, Teslovich TM, et al. The genetic architecture of type 2 diabetes. Nature 2016;536:41–47
6. Pasquali L, Gaulton KJ, Rodríguez-Seguí SA, et al. Pancreatic islet enhancer clusters enriched in type 2 diabetes risk-associated variants. Nat Genet 2014;46:136–143
7. Carninci P, Kasukawa T, Katayama S, et al.; FANTOM Consortium; RIKEN Genome Exploration Research Group and Genome Science Group (Genome Network Project Core Group). The transcriptional landscape of the mammalian genome. Science 2005;309:1559–1563
8. Mendell JT, Olson EN. MicroRNAs in stress signaling and human disease. Cell 2012;148:1172–1187
9. LaPierre MP, Stoffel M. MicroRNAs as stress regulators in pancreatic beta cells and diabetes. Mol Metab 2017;6:1010–1023
10. Weick EM, Miska EA. piRNAs: from biogenesis to function. Development 2014;141:3458–3471
11. Derrien T, Johnson R, Bussotti G, et al. The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. Genome Res 2012;22:1775–1789
12. Batista PJ, Chang HY. Long noncoding RNAs: cellular address codes in development and disease. Cell 2013;152:1298–1307
13. Rinn JL, Chang HY. Genome regulation by long noncoding RNAs. Annu Rev Biochem 2012;81:145–166
14. Cabili MN, Trapnell C, Goff L, et al. Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. Genes Dev 2011;25:1915–1927
15. Ponting CP, Oliver PL, Reik W. Evolution and functions of long noncoding RNAs. Cell 2009;136:629–641
16. Wilusz JE, Sunwoo H, Spector DL. Long noncoding RNAs: functional surprises from the RNA world. Genes Dev 2009;23:1494–1504
17. Guttman M, Rinn JL. Modular regulatory principles of large non-coding RNAs. Nature 2012;482:339–346
18. Guttman M, Amit I, Garber M, et al. Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. Nature 2009;458:223–227
19. Luo H, Bu D, Sun L, Fang S, Liu Z, Zhao Y. Identification and function annotation of long intervening noncoding RNAs. Brief Bioinform 2017;18:789–797
20. Amaral PP, Clark MB, Gascoigne DK, Dinger ME, Mattick JS. lncRNAdb: a reference database for long noncoding RNAs. Nucleic Acids Res 2011;39:D146–D151
21. Volders PJ, Helsens K, Wang X, et al. LNCipedia: a database for annotated human lncRNA transcript sequences and structures. Nucleic Acids Res 2013;41:D246–D251
22. Mercer TR, Dinger ME, Mattick JS. Long non-coding RNAs: insights into functions. Nat Rev Genet 2009;10:155–159

23. Wang KC, Yang YW, Liu B, et al. A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. Nature 2011;472:120–124
24. Rinn JL, Kertesz M, Wang JK, et al. Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. Cell 2007;129:1311–1323
25. Smith MA, Mattick JS. Structural and functional annotation of long noncoding RNAs. Methods Mol Biol 2017;1526:65–85
26. Wang KC, Chang HY. Molecular mechanisms of long noncoding RNAs. Mol Cell 2011;43:904–914
27. Motterle A, Sanchez-Parra C, Regazzi R. Role of long non-coding RNAs in the determination of β-cell identity. Diabetes Obes Metab 2016;18(Suppl. 1):41–50
28. Morán I, Akerman I, van de Bunt M, et al. Human β cell transcriptome analysis uncovers lncRNAs that are tissue-specific, dynamically regulated, and abnormally expressed in type 2 diabetes. Cell Metab 2012;16:435–448
29. Alvarez ML, DiStefano JK. Functional characterization of the plasmacytoma variant translocation 1 gene (PVT1) in diabetic nephropathy. PLoS One 2011;6:e18671
30. Hanson RL, Craig DW, Millis MP, et al. Identification of PVT1 as a candidate gene for end-stage renal disease in type 2 diabetes using a pooling-based genome-wide single nucleotide polymorphism association study. Diabetes 2007;56:975–983
31. Awata T, Yamashita H, Kurihara S, et al. A genome-wide association study for diabetic retinopathy in a Japanese population: potential association with a long intergenic non-coding RNA. PLoS One 2014;9:e111715
32. Carter G, Miladinovic B, Patel AA, Deland L, Mastorides S, Patel NA. Circulating long noncoding RNA GAS5 levels are correlated to prevalence of type 2 diabetes mellitus. BBA Clin 2015;4:102–107
33. de Gonzalo-Calvo D, Kenneweg F, Bang C, et al. Circulating long-non coding RNAs as biomarkers of left ventricular diastolic function and remodelling in patients with well-controlled type 2 diabetes. Sci Rep 2016;6:37354
34. Giroud M, Scheideler M. Long non-coding RNAs in metabolic organs and energy homeostasis. Int J Mol Sci 2017;18:E2578
35. Pagliuca FW, Millman JR, Gürtler M, et al. Generation of functional human pancreatic β cells in vitro. Cell 2014;159:428–439
36. Russ HA, Parent AV, Ringler JJ, et al. Controlled induction of human pancreatic progenitors produces functional beta-like cells in vitro. EMBO J 2015;34:1759–1772
37. Millman JR, Pagliuca FW. Autologous pluripotent stem cell-derived β-like cells for diabetes cellular therapy. Diabetes 2017;66:1111–1120
38. Adams BD, Parsons C, Walker L, Zhang WC, Slack FJ. Targeting noncoding RNAs in disease. J Clin Invest 2017;127:761–771
39. Akerman I, Tu Z, Beucher A, et al. Human pancreatic β cell lncRNAs control cell-specific regulatory networks. Cell Metab 2017;25:400–411
40. Fadista J, Vikman P, Laakso EO, et al. Global genomic and transcriptomic analysis of human pancreatic islets reveals novel genes influencing glucose metabolism. Proc Natl Acad Sci U S A 2014;111:13924–13929
41. Bramswig NC, Everett LJ, Schug J, et al. Epigenomic plasticity enables human pancreatic α to β cell reprogramming. J Clin Invest 2013;123:1275–1284
42. Li B, Bi CL, Lang N, et al. RNA-seq methods for identifying differentially expressed gene in human pancreatic islet cells treated with pro-inflammatory cytokines. Mol Biol Rep 2014;41:1917–1925
43. Nica AC, Ongen H, Irminger JC, et al. Cell-type, allelic, and genetic signatures in the human pancreatic beta cell transcriptome. Genome Res 2013;23:1554–1562
44. Ku GM, Kim H, Vaughn IW, et al. Research resource: RNA-Seq reveals unique features of the pancreatic β-cell transcriptome. Mol Endocrinol 2012;26:1783–1792
45. Benner C, van der Meulen T, Cacéres E, Tigyi K, Donaldson CJ, Huising MO. The transcriptional landscape of mouse beta cells compared to human beta cells reveals notable species differences in long non-coding RNA and protein-coding gene expression. BMC Genomics 2014;15:620
46. Motterle A, Gattesco S, Peyot ML, et al. Identification of islet-enriched long non-coding RNAs contributing to β-cell failure in type 2 diabetes. Mol Metab 2017;6:1407–1418
47. Ulitsky I, Shkumatava A, Jan CH, Sive H, Bartel DP. Conserved function of lincRNAs in vertebrate embryonic development despite rapid sequence evolution. Cell 2011;147:1537–1550
48. Jiang W, Liu Y, Liu R, Zhang K, Zhang Y. The lncRNA DEANR1 facilitates human endoderm differentiation by activating FOXA2 expression. Cell Reports 2015;11:137–148
49. Arnes L, Akerman I, Balderes DA, Ferrer J, Sussel L. βlinc1 encodes a long noncoding RNA that regulates islet β-cell formation and function. Genes Dev 2016;30:502–507
50. Yin DD, Zhang EB, You LH, et al. Downregulation of lncRNA TUG1 affects apoptosis and insulin secretion in mouse pancreatic β cells. Cell Physiol Biochem 2015;35:1892–1904
51. You L, Wang N, Yin D, et al. Downregulation of long noncoding RNA Meg3 affects insulin synthesis and secretion in mouse pancreatic beta cells. J Cell Physiol 2016;231:852–862
52. Motterle A, Gattesco S, Caille D, Meda P, Regazzi R. Involvement of long non-coding RNAs in beta cell failure at the onset of type 1 diabetes in NOD mice. Diabetologia 2015;58:1827–1835
53. Sisino G, Zhou AX, Dahr N, et al. Long noncoding RNAs are dynamically regulated during β-cell mass expansion in mouse pregnancy and control β-cell proliferation in vitro. PLoS One 2017;12:e0182371
54. Kino T, Hurt DE, Ichijo T, Nader N, Chrousos GP. Noncoding RNA Gas5 is a growth arrest– and starvation-associated repressor of the glucocorticoid receptor. Sci Signal 2010;3:r8
55. DiGruccio MR, Mawla AM, Donaldson CJ, et al. Comprehensive alpha, beta and delta cell transcriptomes reveal that ghrelin selectively activates delta cells and promotes somatostatin release from pancreatic islets. Mol Metab 2016;5:449–458
56. Li B, Qing T, Zhu J, et al. A comprehensive mouse transcriptomic BodyMap across 17 tissues by RNA-seq. Sci Rep 2017;7:4200
57. Zou G, Liu T, Guo L, et al. miR-145 modulates lncRNA-ROR and Sox2 expression to maintain human amniotic epithelial stem cell pluripotency and β islet-like cell differentiation efficiency. Gene 2016;591:48–57
58. Bassett AR, Akhtar A, Barlow DP, et al. Considerations when investigating lncRNA function in vivo. eLife 2014;3:e03058
59. Xing Z, Lin A, Li C, et al. lncRNA directs cooperative epigenetic regulation downstream of chemokine signals. Cell 2014;159:1110–1125
60. Zhao J, Ohsumi TK, Kung JT, et al. Genome-wide identification of polycomb-associated RNAs by RIP-seq. Mol Cell 2010;40:939–953
61. Darnell R. CLIP (cross-linking and immunoprecipitation) identification of RNAs bound by a specific protein. Cold Spring Harb Protoc 2012;2012:1146–1160
62. Simon MD, Wang CI, Kharchenko PV, et al. The genomic binding sites of a noncoding RNA. Proc Natl Acad Sci U S A 2011;108:20497–20502
63. Chu C, Qu K, Zhong FL, Artandi SE, Chang HY. Genomic maps of long noncoding RNA occupancy reveal principles of RNA-chromatin interactions. Mol Cell 2011;44:667–678
64. Engreitz JM, Pandya-Jones A, McDonel P, et al. The Xist lncRNA exploits three-dimensional genome architecture to spread across the X chromosome. Science 2013;341:1237973
65. West JA, Davis CP, Sunwoo H, et al. The long noncoding RNAs NEAT1 and MALAT1 bind active chromatin sites. Mol Cell 2014;55:791–802
66. Liu JY, Yao J, Li XM, et al. Pathogenic role of lncRNA-MALAT1 in endothelial cell dysfunction in diabetes mellitus. Cell Death Dis 2014;5:e1506
67. Kaur S, Mirza AH, Brorsson CA, et al.; Hvidoere International Study Group. The genetic and regulatory architecture of ERBB3-type 1 diabetes susceptibility locus. Mol Cell Endocrinol 2016;419:83–91
68. Wang L, Park HJ, Dasari S, Wang S, Kocher JP, Li W. CPAT: Coding-Potential Assessment Tool using an alignment-free logistic regression model. Nucleic Acids Res 2013;41:e74
69. Siepel A, Bejerano G, Pedersen JS, et al. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. Genome Res 2005;15:1034–1050
70. Lin MF, Jungreis I, Kellis M. PhyloCSF: a comparative genomics method to distinguish protein coding and non-coding regions. Bioinformatics 2011;27:i275–i282