EFFECTS OF RS3842753 ON *INS* EXPRESSION IN SINGLE BETA CELLS

by

Su Wang

B.Sc., The University of British Columbia, 2018

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES

(Cell and Developmental Biology)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

October 2020

The following individuals certify that they have read, and recommend to the Faculty of Graduate and Postdoctoral Studies for acceptance, the thesis entitled:

Effects of rs3842753 on INS expression in single beta cells

submitted by Su Wang in partial fulfillment of the requirements for

the degree of Master of Science

in Cell and Developmental Biology

Examining Committee:

James D. Johnson, Professor, Department of Cellular and Physiological Sciences, UBC

Supervisor

Francis C. Lynn, Professor, Department of Cellular and Physiological Sciences, UBC

Supervisory Committee Member

Timothy J. Kieffer, Professor, Department of Cellular and Physiological Sciences, UBC

Supervisory Committee Member

William T. Gibson, Professor, Department of Medical Genetics, UBC

Additional Examiner

Abstract

Type 1 diabetes (T1D) is characterized by the autoimmune destruction of insulin-secreting beta-cells. Genetic variations upstream at the insulin (*INS*) locus contribute to ~10% of T1D heritable risk. Multiple studies showed an association between rs3842753 C/C genotype and T1D susceptibility. Three small studies reported an association between rs3842753 C allele and increased whole pancreas *INS* expression. To date, no large-scale studies have looked at the effect of those genetic variations on insulin expression at the single cell level. We aligned all available human pancreatic single cell RNA sequencing datasets using STAR and used Samtools mpileup to genotype rs3842753. Using Seurat, we integrated 2315 beta-cells from 13 A/A donors, 23 A/C heterozygous donors, and 35 C/C at-risk donors. The donors included persons with and without type 2 diabetes, but not T1D. We compared variance using Bartlett's test or Fligner-Killeen test and means using Wilcox Rank Sum, Student-t, ANOVA, or Kruskal-Wallis tests. Per β-cell *INS* expression mean and variance were significantly higher in females compared with males. In male cells, *INS* expression appeared to be significantly lower in T2D compared to non-diabetic cells. Comparing across all cells, we found that rs3842753 A/C genotype had the highest *INS* expression followed by C/C genotype, lastly by A/A genotype. Donor level comparisons between genotypes were not statistically significant. Conversely, within A/C heterozygous β-cells, A allele specific *INS* expression was higher. This association was consistent at the donor level. Lastly, we examined whole pancreatic islets from a small subset of donors and found no relationship between insulin protein abundance and rs3842753 genotype. Our analysis suggests that in single β-cells, rs3842753 may affect *INS* variance and expression. The contribution of these differences to T1D risk remains unclear.

Lay Summary

Type 1 diabetes (T1D) is caused by the body's own immune attack of pancreatic β-cells which produce insulin, a hormone that regulates blood sugar. The rs3842753 variant in the insulin gene changes the DNA sequence from A to C. Individuals with the C allele are more likely to develop T1D and have more pancreatic insulin expression. However, those studies included a small number of people and used whole pancreas tissue instead of pancreatic β-cells. In this study, we used gene expression data for pancreatic cells from 71 people. When comparing across all of the single cells in our data set, there was a relationship between rs3842753 genotype and pancreatic insulin expression. However, when looking within the individual cells that had one copy of each genotype, we found the opposing result. More research is therefore required to determine the molecular mechanisms of T1D disease development.

Preface

I designed all the bioinformatics experiments and performed all the alignment and analysis in the laboratory of Dr. James D. Johnson, who provided the direction for the project, unless otherwise specified below. Dr. Stephane Flibotte provided integral help and direction at all stages of the bioinformatics analysis.

All datasets are publicly available. As such, all cell sequencing preparations and sequencing runs are conducted according by the publications' authors as cited accordingly.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

ABCC8	ATP-binding cassette transporter sub-family C member 8
ADP	Adenosine diphosphate
AS160	Akt substrate of 160 kDa
ATP	Adenosine triphosphate
BMI	Body mass index
EMBL-EBI	European bioinformatics institute
ER	Endoplasmic reticulum
FACS	Fluorescence-activated cell sorting
FAD	Flavin adenine dinucleotide
FFAs	Free fatty acids
GDM	Gestational diabetes mellitus
GDP	Guanosine diphosphate
GEO	Gene expression omnibus
GTP	Guanosine-5'-triphosphate
GWAS	Genome-wide association studies
HLA	Human leukocyte antigen
HPAP	Human pancreas analysis program
IGF2	Insulin-like growth factor 2
INS	Insulin
IRS1	Insulin receptor substrate 1
K-ATP channel	Adenosine triphosphate sensitive potassium (channel)
KCNJ11	Potassium inwardly rectifying channel subfamily J member 11
MAFA	V-maf musculoaponeurotic fibrosarcoma oncogene homolog A
MODY	Maturity-onset diabetes of the young
mRNA	Messenger ribonucleic acid
NAD	Nicotinamide adenine dinucleotide
NDM	Neonatal diabetes mellitus
NEUROD1	Neurogenic differentiation 1
OR	Odds ratio
PCR	Polymerase chain reaction

PDK1	Pyruvate dehydrogenase kinase 1
PDX1	Pancreatic and duodenal homeobox factor-1
PI3K	Phosphoinositide 3-kinases
PIP2	Phosphatidylinositol 4,5-bisphosphate
PIP3	Phosphatidylinositol (3,4,5)-trisphosphate
PKB	Protein kinase B
PTEN	Phosphatase and tensin homolog deleted on chromosome 10
scRNAseq	Single cell RNA sequencing
SNP	Single nucleotide polymorphism
T1D	Type 1 diabetes
T2D	Type 2 diabetes
TIAR	T-cell restricted intracellular antigen 1-related protein
UMI	Unique molecular identifier
UTR	Untranslated region
VNTR	Variable number of tandem repeats

Acknowledgements

I would like to thank Dr. James D. Johnson for guiding me through this project and tirelessly discussing its direction with me as it morphed from a purely wet lab project to a purely bioinformatics focus. It is through these discussions that I dove head first into new fields and directions.

I am deeply grateful to Dr. Stephane Flibotte for answering my extensive questions about every aspect of transcriptomic bioinformatics I encountered. I entered into the bioinformatics field with nearly zero knowledge concerning anything programing related. His extensive help made this project possible.

I would also like to thank my committee members: Dr. Susanne Clee, Dr. Francis Lynn, and Dr. Tim Kieffer for their guidance and questions, which often opened my eyes to new directions the project could evolve into. Sadly Dr. Clee passed away in July 2020. During my time working on this project, I had numerous discussions with Dr. Clee. Her enthusiasm for genetics and kindness was truly infectious.

I would like to acknowledge the members of the Johnson lab for always providing laugher and scientific discussions during those long days where nothing seems to work.

I would like to thank the donors and their family for dedicating their organs to scientific research. I am grateful to be able to do this project with donor samples and hopefully move closer to understanding diabetes.

Lastly, I would like to thank my parents for supporting of all my endeavors and paths in life. They were always eager to listen as I described every turn of the project, and every new finding (and failure).

Chapter 1: Introduction

1.1: Global Burden of Diabetes

Diabetes mellitus is a disease characterized by the body's chronic inability to regulate blood glucose levels resulting in hyperglycemia (1; 2). This could be due to inadequate insulin synthesis, inadequate insulin secretion, and/or inadequate insulin action (insulin resistance) (1; 2). Chronic hyperglycemia, clinically defined by a fasting glucose of over 7.0mM or random glucose levels over 11.1mM, is the hallmark of all diabetes (1; 2). Insulin is a hormone primarily produced by pancreatic β-cells to promote the uptake of glucose from blood into skeletal muscle, adipose, and liver cells (1). The two main types of diabetes are type 1 diabetes (T1D) and type 2 diabetes (T2D), accounting for approximately 5 to 10% and 90 to 95% of total diabetes cases respectively (1). Other types of diabetes including monogenic diabetes and gestational diabetes account for 1 to 5% of total diabetes cases (1).

In 2014, diabetes affected 422 million people worldwide and directly caused 1.5 million deaths in 2012 (3). As well, complications due to diabetes involve cardiovascular diseases, chronic kidney disease, and neuropathy (3). This has led to an estimated global financial burden of US$1.3 trillion related to diabetes in 2015 with a projected increase to US$2.5 trillion based on past trends by 2020 (4; 5). The global estimate for the prevalence of diabetes mellitus is projected to increase to 642 million people by 2040 (6).

1.2: Type 1 Diabetes

T1D is an autoimmune disease characterized by the autoimmune destruction of pancreatic β-cells (1). A small percentage of T1D cases is caused by idiopathic destruction of pancreatic β-cells or β-cell failure (1). This loss of pancreatic β-cells results in the depletion of the body's main source of insulin leading to inadequate insulin production (1; 7). The lack of insulin results in the body's inability to regulate blood glucose levels which results in the rapid onset of diabetes (1; 7). T1D is the main type of diabetes that occurs in children accounting for more than 85% of all diabetes cases in youth under 20 years old (8-11). Most cases onset between 10 to 14 years old during puberty (12). The main risk factors for T1D are genetic though environmental components are also involved (13-15). Genetic components are evident by the concordance of the disease in monozygotic twins, which ranges from 30 to 70% over time (16).

Genome-wide association studies (GWAS), together with previous gene mapping efforts, have identified nearly 60 susceptibility loci for T1D (17; 18). The strongest genetic association for T1D are genes located in the human leukocyte antigen (*HLA*) loci (odds ratio [OR] = 0.02 to 11). The second strongest association is at the insulin (*INS*) loci (OR = 2.38) (19-21). Odds ratio indicates the likelihood of an individual developing T1D given mutations at those loci. Some environmental factors include early childhood virus infections, vitamin D deficiency, and decreased gut-microbiome diversity (14; 22-24). Since insulin producing pancreatic β-cells are destroyed in T1D, treatments commonly include lifelong insulin replacement therapy and/or pancreatic islet transplantations though research is emerging in stem cell and immune therapies (25). To date, there are no therapies that can cure T1D in humans.

1.3: Type 2 Diabetes

T2D is a metabolic disease caused by insufficient compensatory insulin secretion response in the context of insulin resistance (1; 26; 27). One of the early hallmarks of T2D, and its main risk factor obesity, is hyperinsulinemia (28; 29). Hyperinsulinemia itself can contribute to insulin resistance and manipulations that reduce insulin production have been shown to prevent adiposity and age-dependent insulin resistance in mouse models (30-32). Insulin resistance affects the muscle and adipose cells' ability to absorb glucose from the bloodstream. The resulting chronic hyperglycemia leads to further pancreatic β-cell stress, failure, and death (1; 26; 27).

T2D mainly occurs in older adults with the majority of patients between 45 to 64 years old (27; 33). However, there has been an increasing trend of children, adolescents, and younger adults developing T2D (34). This is especially concerning as these younger onset T2D patients have higher rates of microvascular and macrovascular complications, increased need for insulin treatment and increased mortality (8; 27; 35-37). Environmental and genetic factors are both important for understanding T2D development. Obesity is the most important risk for developing T2D though other environmental factors are also involved as such: sedentary lifestyle, physical inactivity and smoking (38-41). Through GWAS, more than 400 alleles that associate with increased risk for developing T2D have been observed (42). Life style changes to control blood glucose concentrations and lower body mass index (BMI) through diet and physical activity

continue to be the primary treatments for T2D. In the DiRECT study focused on dietary restriction and weight loss as a treatment for T2D, sustained diabetes remission was linked to sustained weight loss(43). Pharmacological agents, such as metformin, are also employed to inhibit gluconeogenesis in the liver (44). For advanced T2D stages, insulin replacement therapy can be used to maintain glucose homeostasis(44).

1.4: Monogenic Diabetes

Two main types of monogenic diabetes are neonatal diabetes mellitus (NDM) and maturity onset diabetes of the young (MODY). NDM is a rare form of diabetes that occurs in around 1 in 90,000-400,000 live births (45-47). It is often characterized by insulin dependent hyperglycemia in infants 6 less than months old, though some cases can present up to 1 year old (1; 45; 47). There are over 20 genetic mutations that cause NDM with mutations in the potassium inwardly rectifying channel subfamily J member 11 (*KCNJ11*) gene and ATP-binding cassette transporter sub-family C member 8 (*ABCC8*) genes accounting for nearly 50% of all NDM cases (48). Both *KCNJ11* and *ABCC8* encode proteins that make up the sub-units of the pancreatic β-cell adenosine triphosphate sensitive potassium (K-ATP) channels (49). These channels regulate the amount of insulin that is released into the bloodstream in response to glucose stimulation (50; 51). The mutations in *KCNJ11* and *ABCC8* directly prevent the closing of K-ATP channels, increasing the concentration of blood glucose needed to stimulate insulin release resulting in permanent NDM (49; 52).

MODY is also a form of monogenic diabetes characterized by early onset (before 25 years old), autosomal mode of inheritance, absence of pancreatic β-cell autoimmunity, sustained pancreatic β-cell function and few defects in insulin action (1). There are mutations in 14 genes associated with MODY which accounts for around 1 to 2% of total diabetes cases (53-58). The majority of MODY cases (45-70%) involve mutations in genes encoding pancreatic β-cell transcription factors that are essential for the development of pancreatic cells (five genes) and glucokinase which is the first rate-limiting step in glycolysis (58; 59). MODY10 is caused by heterozygous mutations on *INS* and affects less than 1% of all MODY cases. Edghill et al. (2008) and Molven et al. (2008) found the individuals that had MODY10 were mostly diagnosed with diabetes in their late teens and treatments were mostly non-insulin dependent (60; 61). Since

monogenic diabetes is inherited, genetic testing is encouraged for early detection and intervention to decrease hyperglycemia related complications. Sulfonylureas are often used to increase the release of insulin by pancreatic β-cells (47). Insulin replacement therapy is also used in cases where sulfonylureas use is insufficient to control hyperglycemia (47).

1.5: Gestational Diabetes Mellitus

Gestational diabetes mellitus (GDM) is classified as hyperglycemia that is first diagnosed during pregnancy with glucose levels below diabetes diagnosis outside of pregnancy (62). This is a sex specific form of diabetes that only affects women. The World Health Organization has estimated the prevalence of GDM to be highest in the Middle East and some North African countries (15.2% of pregnancies), followed by South-East Asia (15% of pregnancies), with the lowest in Europe (6.1% of pregnancies). However, even thought it affects with 6.1% of pregnancies, it still represents a sizable population and a significant burden especially as the complications associated with GDM include pre-eclampsia, preterm delivery, and increased risk of metabolic and cardiovascular disease in mothers (63).

Some of the main risk factors for developing GDM are obesity, advanced age, and ethnicity (63). Obesity (BMI > 25 kg m^2) prior to pregnancy carries the most significant risk for developing GDM (63). A meta-study using over 20 studies found that the OR for developing GDM was 2.14 in overweight, 3.56 in obese, and 8.56 in severely obese pregnant women compared to normal weight women (64). Age is also a major risk factor with women over 40 having a two-fold increase in risk for developing GDM compared to women between 25 to 29 years old (65). Another major risk factor is ethnicity. From world-wide estimations of GDM prevalence, it is clear that geography, and perhaps ethnicity, plays a major role in developing GDM. Other studies that examined GDM prevalence between ethnicities in the same geographical location also found and increase in Asian women compared to Caucasian women (66; 67). It is especially interesting that in Asian, there is an increased risk for developing GDM even at BMI < 25 kg m^2 compared to Caucasian women (66). This difference in GDM prevalence is likely due to a combination of genetic susceptibility and lifestyle.

Pregnancy induces a large amount of stress on the pancreatic β-cells. Mechanistically, the combined increase in insulin synthesis demand and increased insulin resistance lead to the

development of GDM. Insulin sensitivity decreases by around 50% during late gestation, which is associated with a 65% increased in faster insulin levels (68). This is associated with defects in the insulin signalling pathway in skeletal muscle cells. In GDM, there is less autophosphorylation of tyrosine in the insulin receptor β subunit and the amount of insulin receptor substrate 1 (IRS1) in the skeletal muscle cell membranes is reduced. These changes lower the amount of glucose uptake from the blood into the skeletal muscle cells. During pregnancy, there is also a large and rapid increase (30%) in basal endogenous glucose production (69). Whether sex specific differences in insulin target tissues or β cells contribute to the pathogenesis of GDM remains unclear.

1.6: Role of Insulin in Glucose Homeostasis

1.6.1: Insulin Biosynthesis

Insulin plays a crucial role in glucose homeostasis and is intrinsically linked to the development and progression of diabetes. Thus, it is essential to understand the mechanisms and factors that affect insulin biosynthesis, secretion, and signalling.

The human insulin protein (INS) is encoded by *INS* located on the reverse strand on chromosome 11p15.5 (70). It is 1431bp long with three exons and two introns (Figure 1) (70). In pancreatic β-cells, around 50% of total gene expression is due to *INS* expression. Humans have a single copy of the *INS* gene whereas in both mice and rats, there are two insulin genes (*Ins1* and *Ins2*) (71; 72). The proximal promoter region of human *INS* has been extensively studied and located in the 400bp region flanking the transcription start site (Figure 2) (73). Both ubiquitous and pancreatic β-cell specific transcription factors bind to this promotor region (74). Some key transcription factors include pancreatic and duodenal homeobox factor-1 (PDX1), V-maf musculoaponeurotic fibrosarcoma oncogene homolog A (MAFA), and neurogenic differentiation 1 (NEUROD1) (74-76). All three genes play an important role in pancreas development and glucose stimulated insulin secretion. Under high glucose conditions, PDX1 increasingly binds to the A3 binding site, MAFA binds to the C1 binding site and NEUROD1 binds to the E1 binding site (77-79). Under high glucose conditions, PDX1 is phosphorylated, NEUROD1 is glycosylated, and MAFA expression is increased. The three transcription factors are translocated to the nucleus where they aid in recruiting p300, a histone acetyltransferase, to the insulin

promoter (77-83). This opens the chromatin structure and allows for increased *INS* transcription. PDX1 and MAFA expression levels in T2D islets were also selectively decreased (84). Mutations in *INS*, *PDX1*, and *NEUROD1* directly correspond to MODY10, MODY4, and MODY6 respectively (58). As well, the *INS* promotor region has been implicated in T1D via GWAS (85). Increased DNA methylation has also been shown to decrease *INS* transcription, including in the context of T2D (86; 87). The genetic regulation of *INS* expression is therefore important in the context of diabetes and will be discussed further below.

Steady state insulin mRNA levels are a product of balance between mRNA transcription, mRNA stability, and mRNA degradation. Increased transcription and mRNA stability lead to increased levels of mRNA. Whereas, increased degradation and decreased mRNA stability lead to decreased levels of mRNA. Insulin transcription can be altered by transcription factors, chromatin remodeling and DNA methylation in response to changes in glucose concentration as previously discussed. Insulin mRNA has a long half-life in pancreatic β-cells (29 to 77 hours depending on glucose concentration) (88). Several proteins have been found to bind to insulin mRNA to either promote stability or degradation. In rat pancreatic islets, poly-pyrminidine-tract binding protein have been shown to bind to the 3' untranslated region (UTR) of insulin and increase insulin mRNA stability(89). Another protein, T-cell restricted intracellular antigen 1-related protein (TIAR) also compete for the same binding site (90). In an experiment using rat and human pancreatic β-cell line (INS-1 and EndoC-βH1 cells), inhibition of TIAR resulted in increased insulin mRNA levels. There was also an association between stress granule formation and the relocalization of TIAR from the cytosol to the stress granules (90). Stress granules contain mRNA and proteins during times when the cell is under stress. Translation cannot occur off the mRNA inside the granules leading to decreased protein amount. Together, there is evidence pointing to the role of TIAR in decreasing insulin mRNA stability.

INS is primarily produced by the pancreatic β-cells which comprise around 60% of the 2000 cells in a typical human pancreatic islet (91). After transcription from DNA to messenger RNA (mRNA), the INS mRNA is translated into the preproinsulin protein which is 110 amino acids long. Preproinsulin is composed of a hydrophobic N-terminal signal peptide, the B-chain, the connecting (C) peptide, and the A-chain (92). The signal peptide interacts with ribonucleoprotein signal recognition particles to translocate preproinsulin through the peptide-

conducting channel into the rough endoplasmic reticulum (ER) lumen where post-translational modifications occur (93). The signal peptide is cleaved after lumen translocation, becoming proinsulin (93; 94). In the ER, folding and three disulfide bonds are added to proinsulin (95). Two disulfide bonds are located in between the B-chain and A-chain (A7-B7, A20-B19) and one disulfide bond within the A-chain (A7-A11) (96). Proinsulin is finally translocated into the Golgi apparatus where it is engulfed by secretory vesicles and cleaved by prohormone convertases and carboxypeptidase E, resulting in the bioactive insulin (51 amino acids) and C-peptide (97). Insulin is stored in these insulin secretory granules prior to secretion, with a sub-set of granules that are primed to fuse with the cell membrane upon glucose stimulation (98).

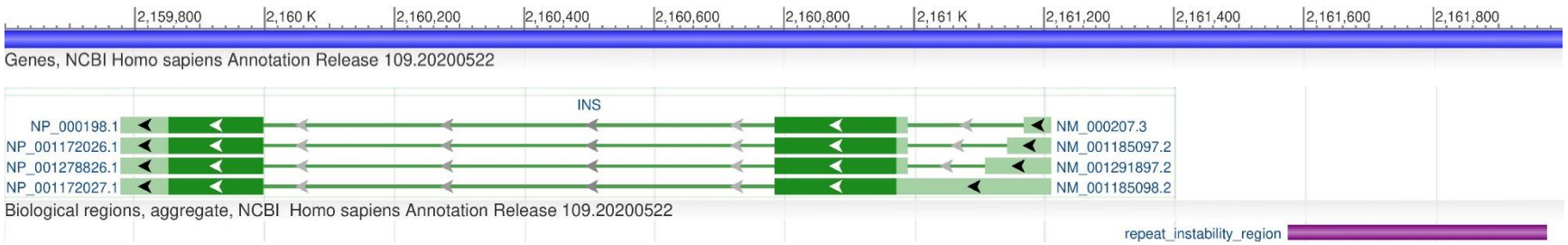


Figure 1: Human Insulin Gene and Surrounding Structure. Aligned to Homo sapiens genome assembly GRCh38.p13 from Genome Reference Consortium. Region shown is located on chromosome 11 from position 2,159,600 to 2,162,000. The light green boxes indicate untranslated exons, green lines indicate introns, dark green boxes indicate translated exons, purple box indicates variable number tandem repeat (VNTR) region, and arrows indicate transcription direction. From top to bottom, the four insulin mRNA variants are shown (variants 1, 2, 4, 3). Figure was obtained from NCBI https://www.ncbi.nlm.nih.gov/gene/3630 (70).

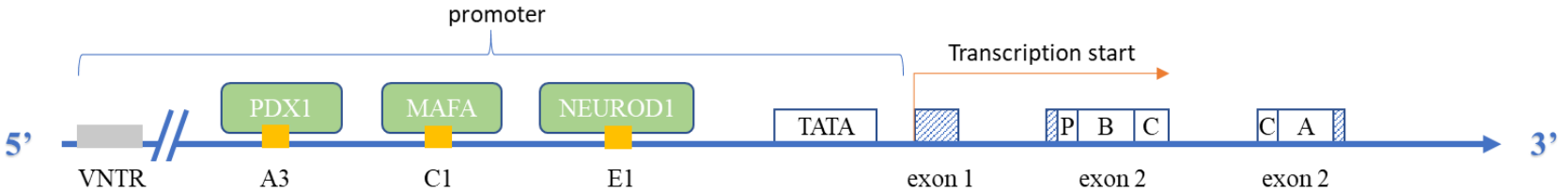


Figure 2: Detailed Human Insulin Gene Structure. Region shown located on chromosome 11p15.5 reverse strand. Arrow indicates transcription direction. Grey box represents the variable number of tandem repeat region. Orange boxes represent key transcription factor binding sites. Green ovals represent key transcription factors. Shaded blue boxes represent the untranslated exons. The peptides are represented as boxes inside the exons. The peptides depicted are the signal peptide (P), B-chain (B), C-peptide (C), and A-chain (A). Image not drawn to scale and adapted from German et al., 1995 and Hay and Docherty, 2006 (73; 99).

1.6.2: Mechanisms of Glucose Stimulated Insulin Secretion

After a meal, increased glucose concentration in the bloodstream stimulate pancreatic β-cells. In humans, glucose predominantly enters the pancreatic β-cell through glucose transporter 1 and is phosphorylated by glucokinase, generating pyruvate. Pyruvate is oxidized by pyruvate dehydrogenase into acetyl coenzyme A which enters the citric acid cycle (100-102). In the matrix of the mitochondria, acetyl coenzyme A in metabolized in the citric acid cycle leading to the production of the reduced form of nicotinamide adenine dinucleotide (NADH) from NAD, the reduced form of flavin adenine dinucleotide ($FADH_2$) from FAD, and guanosine-5'-triphosphate (GTP) from guanosine diphosphate (GDP) (102). NADH and $FADH_2$ are used in the electron transport chain to produce ATP thorough the transportation of protons by four inner mitochondria membrane bound complexes (I, II, III, IV) from the mitochondria matrix into the inter membrane space (102). As more protons are transported into the inter membrane space, an electro-chemical gradient is established, and the protons cannot move freely across the inner mitochondrial membrane. Instead, the protons move through ATP synthase into the mitochondrial matrix which catalyzes the formation of ATP from ADP and inorganic phosphate (102). Through this process, one molecule of glucose produces 32 molecules of ATP. As the ratio of ATP/ADP (Adenosine diphosphate) rises in the pancreatic β-cell, the K-ATP channels at the cell membrane are inhibited from opening, thus depolarizing the cell membrane (51; 101; 103). When the membrane potential depolarized to approximately -50 mV, the voltage gated L-type calcium channels open and Ca^{2+} rapidly enters the cell (101; 104). This influx of Ca^{2+} triggers the fusion of insulin granules bilayer with the cell membrane, via a process known as exocytosis, releasing insulin into the bloodstream (101; 105).

1.6.3: Molecular Mechanisms of Insulin Action

Insulin flows via the bloodstream to act on fat and muscle cells to promote the uptake of glucose and liver cells to synthesize glycogen to store glucose and suppress gluconeogenesis (101). In muscle and fat cells, insulin binds to and activates insulin receptors located on the surface of the cell membrane through autophosphorylation of the β subunit to initiate insulin signalling (106). The activated insulin receptor β subunit phosphorylates the insulin receptor substrate proteins especially the IRS1 isoform through Serine/Tyrosine kinases (107; 108).

Phosphoinositide 3-kinases (PI3K) bind to insulin receptor substrate proteins to recruit phosphatidylinositol 4,5-bisphosphate (PIP2) and phosphorylate to phosphatidylinositol (3,4,5)-trisphosphate (PIP3) (109). Phosphatase and tensin homolog deleted on chromosome 10 (PTEN) is a key inhibitor of IRS1 and PIP3 by dephosphorylation (110; 111). As PIP3 concentrations increase, pyruvate dehydrogenase kinase 1 (PDK1) is recruited towards the plasma membrane and phosphorylated triggering PDK1 activation (112). The phosphorylated PDK1 along with the mammalian target of rapamycin complex 2 phosphorylates and activates protein kinase B (PKB). The activation of PKB initiates a cascade of downstream phosphorylation events resulting in the phosphorylation and inhibition of Akt substrate of 160 kDa (AS160) (113). AS160 normally inhibits the translocation of glucose transporter type 4 to the cell membrane (114). Thus, the inhibition of AS160 allows glucose transporter type 4 to embed into the lipid bilayer allowing for the intake of glucose into the muscle or adipose cells.

Defects in the insulin signalling pathway in muscle and adipose cells result in insulin resistance and are especially important in T2D. Obesity is the cornerstone factor for developing insulin resistance through a combination of chronic inflammation and lipotoxicity (115). Chronic inflammation has been associated with obesity, insulin resistance and T2D (116; 117). The increase in cytokine secretion has been shown to contribute to insulin resistance through the activation of Serine/Tyrosine kinases thus inhibiting the phosphorylation of IRS1, PKB and downstream proteins (118; 119). This results in the lack of glucose uptake into muscle and adipose cells resulting in insulin resistance. Lipotoxicity results from the overflow of FFAs from adipose tissues into muscle, liver and pancreatic β-cells (120). The increase of circulating FFAs can induce cytokine production, thus insulin resistance, as mentioned above. As well, toxic lipid metabolites, especially ceramide, inhibit the activation of PBK through the phosphorylation of PBK at tyrosine-34 thus leading to insulin resistance (121).

1.6.4: Defects in the Insulin Production and Diabetes

Since pancreatic β-cells are the main producers of insulin in the body, it is important to consider ER homeostasis, which is the balance between the physiological demand for protein production and the ER folding capacity to produce mature proteins (122). It has been proposed that conditions of increased insulin demand (hyperglycemia, decreased pancreatic β-cells, and/or

insulin resistance) could generate large proportions of misfolded/unfolded proinsulin compared to properly folded proinsulin, leading to ER stress (123; 124). Additionally, a recent study has shown that insulin production even under basal conditions in mice is a significant source of ER stress (125). Upon ER stress, the unfolded protein response is activated to increase the production of proteins related to ER-associated protein degradation, autophagy, and apoptosis. This promotes the degradation and clearance of misfolded proteins or during prolonged stress, programmed cell death (122; 126).

In T1D patients, ER stress can trigger increased production of cytokines during apoptosis, alternative splicing, and the presentation of unknown (neo)autoantigens to immune cells which further exacerbates the ER stress response (127-129). Neoautoantigens are unique antigen expression that may present exclusively under specific conditions, such as ER stress, rather than ubiquitously. These neoautoantigens could be the result of errors in transcription, translation, post-translational processing, and disposal due to ER stress (130-132). Multiple neoautoantigens have already been identified including an alternative reading frame *INS* mRNA translated product named INS-DRiP (131; 133; 134). This experiment was conducted in transfected 293T cells treated with the ER stressor, thapsigargin, and showed INS-DRiP could stimulate T-cell proliferation and cytokine production (131). The importance of INS-DRiP in T1D autoimmunity is unclear.

In T2D patients and rodents, environmental factors include hyperglycemia, free fatty acids (FFAs), and inflammatory cytokines have been shown to triggers for ER stress in pancreatic β-cells (135-139). In rodents, this has been shown to be especially detrimental during chronic hyperglycemia since chronic ER stress triggers the apoptotic response leading to the destruction of pancreatic β-cells (135). In obese T2D patients, hyperglycemia and excess FFAs interact to induce chronic ER stress leading to decreased β-cell dysfunction and apoptosis (140). Since insulin overproduction induced ER stress plays a role in both T1D and T2D, it is important to study the genetic variations in *INS* that can result in increased insulin expression.

1.7: The Genetics of Insulin and Insulin Expression Regulation

1.7.1: Insulin Genetic Landscape

The *INS* gene is located on chromosome 11p15.5 (70). There are four variants of the *INS* mRNA with all variants encoding the same INS protein with only variance in the 5'UTR (Figure 1) (70). Exon 1 is untranslated, exon 2 encodes the signal peptide, B-chain, and part of the C peptide, and exon 3 encodes part of the C peptide and A-chain (Figure 2) (73). The proximal promotor region 0 to 300 bp upstream of *INS* contains three key transcription factors binding sites A3, C1, and E1 which bind to key transcription factors: PDX1, MAFA, and NEUROD1, respectively (73). These transcription factors have been shown to play a key role in pancreatic β-cell development and *INS* expression (73-77). Around 400bp upstream of *INS* is a region of variable number of tandem repeats (VNTR) that has been the main region implicated in increased T1D susceptibility (141). The two genes flanking *INS* are tyrosine hydroxylase and paternally imprinted insulin like growth factor 2 (*IGF2*) on the forward strand. There is a readthrough transcript starting from the *INS* start codon going into and through *IGF2* (*INS-IGF2*). Since the 4.1kb region surrounding the insulin gene is the second most associated genetic region with increased T1D susceptibility, the specific genetic variations warrant further investigation.

1.7.2: Clinical Phenotypes Associated With INS-VNTR Variants

The *INS* gene accounts for around 10% of T1D disease genetic risk (19-21). This association with increased T1D susceptibility has been found across numerous GWAS studies and population groups (142-146). The specific genetic variant associated with T1D susceptibility at the *INS* loci is the *INS*-VNTR region which is around 400bp upstream of the *INS* gene (15; 141-143; 146; 147). This genetic variant is in the human genome and not mice. The *INS*-VNTR are 14bp to 15bp long groups of tandem repeat sequences that are classified into the shorter Class I (26 to 63 repeats) and longer Class III (141 to 209 repeats) (141). Class II is rarely found in Caucasian populations which is the main demographic studied in previous literature. There are two single nucleotide polymorphisms (SNPs), rs689 and rs3842753, that are in strong linkage disequilibrium with the *INS*-VNTR thus also implicated in increased T1D susceptibility (142). This linkage disequilibrium is almost 100% in the Caucasian population (147). The global rs689

reference allele (A) frequency is around 0.28 with variance from 0.04 (East Asian population) to 0.49 (African populations) (148-152). The rs689 reference allele (T) is in linkage disequilibrium with the rs3842753 reference allele (A) and class III *INS*-VNTR (147). These are often referred to as the protective allele. By extension, the rs689 alternate allele (A) is in linkage disequilibrium with the rs3842753 alternate allele (C) and class I *INS*-VNTR (142). These are often referred to as the at-risk allele due to its association with increased T1D susceptibility.

INS-VNTR has also been linked to other conditions outside of T1D. Earlier studies have found an association between these genetic variations (Class I/I *INS*-VNTR, rs689 A/A genotype, and rs3842753 C/C genotype) and an increase in childhood obesity, but recent studies have not been able to replicate this finding (153-160). Another study found an association between paternally inherited Class III *INS*-VNTR (and by extension: rs689 T allele, and rs3842753 A allele) with an increase in T2D susceptibility (161). However, later studies have not been able find any associations between Class III/III *INS*-VNTR genotype and T2D in larger cohorts (162; 163). Currently, the most consistent genotype/phenotype association at the *INS*-VNTR locus is with T1D susceptibility.

1.7.3: INS-VNTR and INS Secretion

Numerous studies have investigated the relationship between *INS*-VNTR genotype and insulin secretion both under basal and glucose stimulations, and insulin sensitivity. The majority of studies have found no association between *INS*-VNTR genotype and basal insulin in children and adults (156; 157; 160; 163; 164). However, upon glucose stimulation, the associations between *INS*-VNTR genotype and insulin secretion become unclear depending on stimulation method, age of individuals, and diabetes status (156-158; 160; 163-165). One large study in non-diabetic Danish young adults (n = 358) found decreased acute serum insulin concentrations after intravenous glucose tolerance test in young adults with Class III/III *INS*-VNTR genotype compared to individuals with Class I *INS*-VNTR allele (heterozygous or homozygous) (163) . However, this association was no longer found in middle-aged adults (n = 4444) (163). Of note, studies using cohorts from United Kingdom, France, and Germany have failed to find an association between *INS*-VNTR genotype and insulin sensitivity (156; 157; 164). Thus, the role

of the *INS*-VNTR genotype on glucose stimulated insulin secretion and insulin resistance is unclear.

1.7.4: INS-VNTR and INS Expression

Despite multiple studies that have focused on the relationship between *INS*-VNTR genotype and T1D susceptibility, only a few studies have reported on the relationship between *INS*-VNTR genotype and gene expression. Twenty-five years ago, a key study examined the insulin expression from three adult heterozygous Class I/III *INS*-VNTR whole pancreases and found an association between the at-risk allele, Class I *INS*-VNTR allele, and increased pancreatic insulin expression (143). This association have been re-confirmed two additional small studies, both using fetal heterozygous Class I/III *INS*-VNTR pancreas tissues (n = 1 and n = 10) (166; 167). These three studies all used heterozygous donors to account for donor specific differences and found a Class I *INS*-VNTR allele specific increase in pancreatic *INS* expression.

Class I/I *INS*-VNTR correlated increase in pancreatic insulin expression could indicate increased pancreatic insulin production. If there were a reproducible link between increased T1D susceptibility and increased pancreatic *INS* expression due to the Class I *INS*-VNTR allele, it would support a possible increase in β-cell ER stress in at-risk individual. This may in turn lead to increased generation of neoautoantigens, a provoked autoimmune response, and ultimately in increased susceptibility to T1D.

Another working model posits that *INS*-VNTR affects central tolerance. Vafiadis et al. (1997) also found an association between Class I/I *INS*-VNTR and decreased thymus *INS* expression using human fetal thymus tissue (166). That study measured the insulin expression in twelve human fetal thymus tissues that are heterozygous for the *INS*-VNTR allele and found within the same fetus, Class I *INS*-VNTR correlates with lower *cis* thymic *INS* expression and Class III *INS*-VNTR correlates with higher *cis* thymic *INS* expression. The findings have been reconfirmed in another study using eight Class I *INS*-VNTR/Class III *INS*-VNTR heterozygous fetal thymuses (168). These studies again used heterozygous donors to account for donor specific differences. The researchers proposed that increased *INS* expression in the thymus (correlated with Class III *INS*-VNTR) leads to increased immune tolerance and therefore less immune activation targeting the pancreatic β-cells. Additional evidence for this hypothesis comes from

the significant decrease in proinsulin reactive T cells in either non-diabetic or T1D individuals with the Class III *INS*-VNTR allele (homozygous or heterozygous) compared to individuals with the Class I/I *INS*-VNTR genotype (169). The decrease in immune activation targeting the pancreatic β-cells is hypothesized to protect the pancreatic β-cells from autoimmune attacks which is the key characteristic of T1D.

All previous studies examining the effect of *INS*-VNTR genotype and tissue expression used whole organ tissue (whole pancreas or whole thymus). This presents a problem as recent discoveries have highlighted the heterogeneity of pancreatic β-cell gene expression using single cell RNA sequencing (scRNAseq) (170-181). Thus, previous studies could have not been able to detect the variance in insulin expression within a donor. As well, they would not have detected pancreatic β-cell specific insulin expression.

1.8: Single Cell RNA sequencing and Pancreatic β-cell Physiology

1.8.1: RNA sequencing

RNA sequencing measures transcriptome expression and consists of two main types: bulk sequencing and single cell sequencing. Bulk RNAseq is widely used and produces the average transcriptome information of all the cells provided (182). Thus, it is commonly used to identify differences in gene expression between different conditions at the tissue level. ScRNAseq has gained popularity recently due to its ability to extract gene expression information for each cell sequenced (182; 183). Therefore, it is able to identify differences in gene expression between cells, cell types, and conditions. As well, it is commonly used to identify novel cell types/sub-populations. However, due to the small amount of RNA material provided, scRNAseq produces data with more technical noise and less read depth, and is thus inadequate for very lowly expressed transcripts (183). Some technical noise can be eliminated through the use of unique molecular identifiers (UMIs) which are attached to unique transcripts. This reduces technical noise due to polymerase chain reaction (PCR) amplification through removing duplicate reads due to PCR while maintaining biological duplicates due to increased expression (184). This is very important in experiments with low input (single cell) projects as more PCR cycles are needed to achieve sufficient material for sequencing. Not all scRNAseq protocol includes this process (183).

1.8.2: scRNAseq and β-cell Heterogeneity

In recent years, there has been an explosion in the number of studies using scRNAseq in human pancreatic islet cells. Through the lowered cost and wider commercial availability, researchers have been able to sequence more cells to greater sequencing depths aiding in the discovery of rare cell types/subtypes and differential gene expression between disease conditions and cell subpopulations. Multiple studies using scRNAseq on human pancreatic β-cells have found the presence of distinct subpopulations of β-cells (171; 172; 177; 180; 181). Interestingly, the genes markers for the β-cells subpopulations have concentrated in genes in the ER stress and unfolded protein response pathways in non-diabetic donor cells (171; 172; 180). As well, Xin et al. (2018) found that different subpopulations of pancreatic β-cells have different *INS* expression (171). Though Segerstolpe et al. (2016) have not found a difference in *INS* expression in their pancreatic β-cell subpopulations (177). These results clearly indicate the presence of heterogeneity among pancreatic β-cells though the specific characteristics of these subpopulations are not clearly identified.

Several studies have examined whether there are transcriptomic changes between pancreatic β-cells from non-diabetic versus diabetic donors (170; 174; 177-179). In the review by Wang and Kaestner (2019), the differential expression of genes between T2D and non-diabetic pancreatic β-cells was largely non-overlapping (185). This could be due to a combination of the complexity of T2D etiology and limited number of donors. Often, there are few donors for either disease status with a wide range in age and BMI. This could be overcome by expanding the number of donors in future studies or integrating several studies to increase donor number.

Functional heterogeneity was observed in human pancreatic β-cells long before scRNAseq was available. The main areas were insulin secretory response to glucose stimulation and calcium oscillations. Sensitivity to glucose-stimulated insulin secretion was not uniform across all pancreatic β-cells and a small group of cells account for the majority of insulin secreted (186). This difference in glucose stimulated insulin secretion in the same individuals could indicate the presence of different subpopulations of pancreatic β-cells within one individual. As well, Ca^{2+} responses varied between β-cells in the pancreas though signal was synchronized in small cluster regions of pancreatic islets (187; 188). This type of

synchronization could indicate the communication between pancreatic β-cells as well as separation of different β-cells groups.

1.9: Thesis Investigation

The relationship between *INS*-VNTR genotype (and by extension SNPs rs689 and rs3842753) and pancreatic *INS* expression and protein abundance is currently unclear. The studies conducted previously had small sample sizes and were not pancreatic β-cell specific. However, they used heterozygous donors to account for donor specific differences. To address the shortfalls, we conducted a correlational study using all available pancreatic scRNAseq data to find the association between *INS*-VNTR genotype and pancreatic β-cell *INS* expression. We also examined allele specific *INS* expression in single cells of heterozygous donors to account for donor specific differences. We integrated the data from different studies to increase the donor number. As well, we conducted a pilot correlational study between *INS*-VNTR (and SNPs rs689 and rs3842753) genotype and insulin protein abundance to determine if the changes in *INS* expression translated to changes in protein abundance. Filling this knowledge gap is critical for testing multiple theories on the pathogenesis of T1D.

Chapter 2: Materials and Methods

All scRNAseq alignments and initial dataset integration analysis where conducted using the Cedar Compute Canada cloud compute resource (www.computecanada.ca). All subsequent analysis using Seurat was done in R version 3.6.0 (189).

2.1: Gene Expression Datasets

The majority of datasets containing scRNAseq data were found through searching in PubMed for keywords including: “single cell”, “scRNAseq”, “human”, and/or “pancreas”. PubMed searches resulted in nine datasets (170-174; 177-180). Xin et al. (2018) was excluded as the raw files did not contain sufficient information for genotype analysis (171). Datasets in Baron et al. (2016) and Muraro et al. (2016) were excluded from the final analysis because the scRNA sequencing reads were UMI counts as opposed to read counts (172; 180). There were more datasets with read counts than UMI counts. The five datasets found from PubMed searches were obtained through the Gene Expression Omnibus (GEO) or the European Bioinformatics Institute (EMBL-EBI). Both the raw sequencing files and metadata information were downloaded, except for the metadata from Camunas-Soler et al. (2020) which were obtained through the GitHub link https://github.com/jcamunas/patchseq/tree/master/data (179). The Human Pancreas Analysis Program (HPAP) datasets were found online (190). The datasets from HPAP were obtained directly from investigators (190). The datasets were split by donor as each donor was sequenced at separate times and on separate runs. All datasets metadata contains donors’ age, sex, and diabetes status. Some datasets metadata contain donors’ BMI, ethnicity, and cell type (as analysed by publication). The Wang et al. (2016) and HPAP datasets metadata do not include cell type classifications.

In total, we used data from 71 donors (22452 raw files [cells]), split into 48 non-diabetic donors (13433 raw files [cells]) and 23 T2D donors (9019 raw files [cells]). There were 27 male non-diabetic donors and 21 female non-diabetic donors. There were 10 male T2D donors and 13 females T2D donors (Table 1). As of this writing, we believe it is the largest compiled set of human islet cell scRNAseq data.

Table 1: Dataset Summary Statistics

Study	Date	Non-Diabetic Donor #		Non-Diabetic Cell #	T2D Donor #		T2D Cell #
		Male	Female		Male	Female	
Xin	2016	7	5	651	3	3	949
Wang	2016	1	1	686	1	1	412
Segerstolpe	2016	5	1	1980	2	2	1534
Camunas-Soler	2020	9	9	4346	3	4	3048
Enge	2017	3	2	1814	NA		NA
HPAP Data							
HPAP-1	2016	NA	NA	NA	1	NA	798
HPAP-3	2017	1	NA	799	NA	NA	NA
HPAP-4	2017	NA	1	760	NA	NA	NA
HPAP-6	2017	1	NA	799	NA	NA	NA
HPAP-7	2017	NA	NA	NA	NA	1	799
HPAP-8	2017	NA	1	799	NA	NA	NA
HPAP-10	2017	NA	NA	NA	NA	1	680
HPAP-12	2017	NA	1	799	NA	NA	NA
HPAP-13	2017	NA	NA	NA	NA	1	799
TOTAL		27	21	13433	10	13	9019

2.2: Insulin Protein Abundance

Only the Camunas-Solar et al. (2020) and HPAP datasets contain insulin protein abundance measurements (179). The insulin abundance data for the Camunas-Solar et al (2020) dataset was obtained from the Alberta Diabetes Institute IsletCore database. The insulin abundance data for the HPAP datasets were obtained from the HPAP consortium. Both datasets report insulin content per islet equivalent. Specific protocols for insulin abundance determination for the Camunas-Solar et al. (2020) and HPAP datasets can be found on the respective database (https://www.protocols.io/view/static-glucose-stimulated-insulin-secretion-gsis-p-wy4ffyw, https://hpap.pmacs.upenn.edu/explore/workflow/islet-physiology-studies?protocol=5).

2.3: Data Accession and Single Cell RNA Sequencing

For datasets that were stored in GEO, sra-tookit version 2.9.6 (http://ncbi.github.io/sra-tools/) was used to convert SRA files to fastq files using the fasterq-dump option, adjusting for single or paired end reads as appropriate. The raw sequencing files for Segerstolpe et al., (2016) and HPAP datasets were uploaded to EMBL-EBI and HPAP databases as fastq files thus no conversion was necessary (177; 190).

For all six datasets used in this study, the Smart-seq or Smart-seq2 methods were used for cDNA synthesis and PCR amplification (191; 192). The read length ranges from 43bp to 100bp and median read depth per cell ranges from 0.75 million reads per cell to 4.4 million. Specific sequencing protocols for individual datasets can be found in Table 2.

Table 2: Dataset Sequencing and Accession Number.

Study	Date	Protocol	Sequencing System	Read Type	Read Length	Read Depth per Cell	Data Accession Location	Accession Method
Xin	2016	SmartSeq	Illumina HiSeq2500	Single end	75bp	0.95 million	GSE81608	GEO
Wang	2016	SmartSeq	Illumina HiSeq2500	Single end	100bp	2.2 million	GSE83139	GEO
Segerstolpe	2016	SmartSeq2	Illumina HiSeq2000	Single end	43bp	0.75 million	E-MTAB-5061	EMBL-EBI
Camunas	2020	SmartSeq2	NextSeq 500 or NovaSeq platform	Paired end	75bp	1 million	GSE124742	GEO
Enge	2017	SmartSeq2	Illumina NextSeq	Paired end	75bp	1 million	GSE81547	GEO
HPAP								
HPAP-1	2016	SmartSeq	Illumina HiSeq2500	Single end	100bp	4.4 million	https://hpap.pmacs.upenn.edu	provided by investigator
HPAP-3	2017							
HPAP-4	2017							
HPAP-6	2017							
HPAP-7	2017							
HPAP-8	2017							
HPAP-10	2017							
HPAP-12	2017							
HPAP-13	2017							

All HPAP datasets protocol, read type, accession location, and accession methods are indicated by HPAP Data row. Specific HPAP dataset creation date indicated in each respective HPAP dataset row.

2.4: Read Alignment and Genotyping rs689 and rs3842753

Reads were aligned to the GRCh38.98 human reference genome using STAR version 2.7.1a (193). Both bam files and gene read counts were obtained using –outSAMtype BAM SortedByCoordinate and --quantMode GeneCounts, respectively. The read count files were aggregated into study specific read count matrixes using custom code.

Each cell was genotyped at SNPs rs689 and rs3842753 which are in tight linkage disequilibrium with the *INS*-VNTR (142). Since it is not possible to accurately align RNA sequencing reads to VNTRs longer than the length of the individual reads (43bp to 100bp) using PCR based sequencing, the read alignment at the SNPs are used as surrogate for *INS*-VNTR class determination. Samtools version 1.9 and Bcftools version 1.9 were used to genotype the SNPs(194; 195). Samtools index was used to index the sorted BAM files obtained from STAR. Samtools mpileup was used to obtain read counts for rs689 (chr 11: 2160994 - 2160994) and rs3842753 (chr 11:2159830-2159830), saved as binary variant call format (BCF) files. Bcftools call was used to convert BCF files to the variant call format (VCF) files for rs689 and rs3842753. The VCF files were aggregated into study specific genotype matrix using custom code.

For each cell, if the DP4 total (sum of read alignments to the reference and alternate allele, forward and reverse) was larger than 100 then the reference allele percent, alternate allele percent, and genotype were calculated. The cut off is set to eliminate the cells that are not β-cells thus do not express sufficient *INS* for reads to align to the SNP location. Reference allele percent was calculated by $\frac{(DP4ReferenceForward+DP4ReferenceReverse)}{DP4\ total}$. The alternate allele percent was calculated by $\frac{(DP4AlternateForward+DP4AlternateReverse)}{DP4\ total}$. The genotype was determined as reference (rs689 T allele or rs3842753 A allele) if the reference allele percent was over 80, alternate (rs689 A allele or rs3842753 C allele) if the alternate allele percent was over 80, or heterozygous otherwise. The genotype for an individual is determined as the vast majority of the individual's cells' genotype.

2.5: Dataset Filtering and Cell Type Analysis

The main analysis tool used for was Seurat version 3.0(196). Dataset integration was conducted in Cedar Compute Canada. The 5 published datasets and 9 HPAP datasets were made

into Seurat objects. For filtering and subsequent analysis, all datasets are split into non-diabetic donors and T2D donors, except HPAP datasets which were split into individual donors. Dataset from Camunas-Soler et al. (2020) was split into patch-clamp data and fluorescence-activated cell sorting (FACS) data since different sequencing protocols were used(179). Initial filtering was conducted to remove doublets and low viability cells. Doublets were removed by subsetting out cells with high number of total RNA counts, and genes. Low viability cells were removed by subsetting out cells with low number of genes and high mitochondrial gene percent. The upper and lower limits for total number of RNA read counts (nCount_RNA) and genes (nFeature_RNA) are determined specific to each dataset to account for library preparation and sequencing protocol differences. This was done upon visual inspection of the distribution graphs for each dataset. The highest mitochondrial gene percent is set to 25%.

All fourteen datasets were integrated into a single Seurat object using SCTransform following the default settings. UMAP clusters were determined using default Seurat settings. Top ten differentially expressed genes in each cluster (compared to all other cells) was used to determine cluster cell type identify. As well, the location of pancreatic hormone genes (*GCG, INS, PPY* and *SST*) expression was used to determine Alpha (α), β, PP, and delta cell clusters respectively. Enrichr (gene enrichment analysis) was also used when cluster identity was more difficult to determine (197).

2.6: Insulin Expression Analysis

Due to the overwhelming amount of *INS* expression in pancreatic β-cells (50% of all gene expression) and to prevent 'self-normalization', *INS* gene expression was normalized against select housekeeping genes rather than against all expressed genes in a cell which is the standard Seurat protocol. The inter-dataset variance and relative expression (to all genes in cell) of established human housekeeping genes are examined. Housekeeping genes are selected for *INS* normalization based on small inter-dataset expression variance and sufficient expression to accurately determine presence of gene expression. *INS* expression was normalized using custom code we wrote. β-cells *INS* read counts are normalized against the sum of select housekeeping genes read counts, scaled by factor of 10000, and natural log transformed. All cells with zero values for *INS* read counts or select house keeping genes read counts were not included.

2.7: Allele Specific Insulin Expression Analysis

To account for differences between donors, allele specific *INS* expression analysis was done on donors that were heterozygous for rs3842753. The reference allele read counts and alternate allele read counts were obtained for each cell. Then, the reference allele (A) percent was calculated for each cell and compared against a theoretical normal distribution with a mean of 50 and standard deviation of 10. The number of samples in the theoretical distribution was the same as in the heterozygous group (782 for cell level analysis and 23 for donor level analysis). Non-allele biased gene expression should have 50% of each allele expressed. The expression distribution was assumed to be normally distributed.

2.8: Statistical Analyses

All statistical analyses were done in R 3.6.0 (189). Shapiro-Wilk test was used to test for normality (198). Fligner-Killeen or Bartlett's test were used as a non-parametric or parametric test for difference in variance, respectively (199; 200). Wilcox Rank Sum or Student-t test was used as a non-parametric or parametric tests, for differences in normalized *INS* expression and insulin protein abundance between sexes and between disease statuses (201). Normalized *INS* expression within sex and between disease statuses was also examined using Wilcox Rank Sum test. Wilcox Rank Sum or Student-t test was also used to test for differences in allele expression percent between genotype, either sex, either disease status and theoretical normal distribution with a mean of 50 and standard deviation of 10. Between sex and between disease status allele expression percent differences were not tested. Pairwise Kruskal-Wallis tests were used as a non-parametric test for differences in cell specific normalized *INS* expression between genotypes, adjusting for multiple comparisons with Bonferroni correction (202; 203). One-way ANOVA test was used as a parametric test for differences in donor specific mean and median normalized *INS* expression between genotypes (204; 205).

Chapter 3: Results

3.1: Genotyping and Dataset Genotype Summary

Upon genotyping for rs689 and rs384275, reads could only be aligned to SNP rs3842753 (Figure 3). The genotype at this location is used in all subsequent analysis. Reads for rs689 were not included since it is in the 5' intronic region thus excised from the mature mRNA (Figure 1). Out of 71 total donors, there were 48 non-diabetic donors and 23 T2D donors. Within non-diabetic donors, there were 9 donors with the homozygous A allele, 12 with heterozygous (A/C allele), and 27 with the homozygous C allele. Within T2D donors, there were 4 donors with the homozygous A allele, 11 with heterozygous (A/C allele), and 8 with the homozygous C allele. There were no T2D male donors with the homozygous A allele. None of the donors from the HPAP database had the homozygous A allele (Table 3).

3.2: Dataset Quality Control

An important first step in the analysis was quality control of the data. Figure 4 to 6 shows the unfiltered and filtered distribution graphs for the nCount_RNA, nFeature_RNA, and mitochondrial gene percent (compared to total gene) for all datasets analysed. In the majority of datasets, clear upper and lower cutoffs for nCount_RNA and nFeature_RNA could be visually determined. There were very few cells from the Camunas-Soler et al. (2020) patch-clamp dataset as each cell was individually patch-clamped (Figure 5C, D)(179). As well, in the FACS dataset, many cells had higher than 25% of mitochondrial genes, indicating large numbers of cells undergoing apoptosis or necrosis (Figure 5E, G). As well, many cells in the HPAP-003 dataset had low number of aligned genes (less than 1000) (Figure 6A, B). This could indicate a problem with sequencing; thus, results from theses samples should be interpreted conservatively.

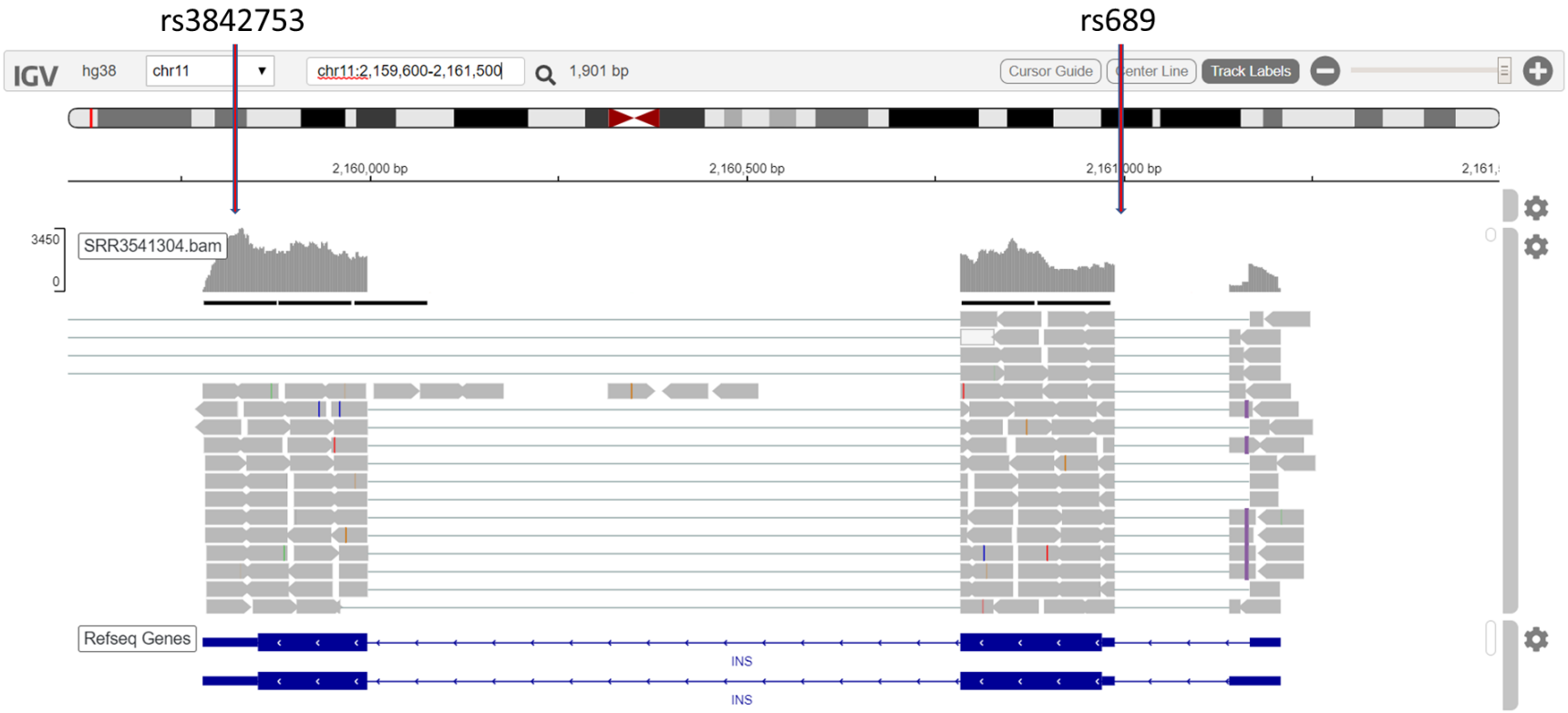


Figure 3: Representative reads aligned to the *INS* gene locus for non-diabetic donor cell. The reference gene *INS* is at the bottom with small blue boxes representing UTRs, large blue boxes representing exons, blue lines representing introns, and arrows representing transcription direction. The grey arrows represent read fragments with arrow direction representing alignment direction. The thin coloured vertical bars represent SNPs, the thicker purple coloured vertical bars represent insertions. Large red arrows represent locations of the SNPs rs689 and rs3842753.

Table 3: Donor number by sex and genotype for each dataset.

		Non-Diabetic				T2D			
Study		**A/A**	**A/C**	**C/C**	**Total**	**A/A**	**A/C**	**C/C**	**Total**
Xin et al., 2016	**Total Donor #**	3	4	5	12	1	3	2	6
	Male	3	1	3	7	0	2	1	3
	Female	0	3	2	5	1	1	1	3
Wang et al., 2016	**Total Donor #**	0	0	2	2	1	0	1	2
	Male	0	0	1	1	0	0	1	1
	Female	0	0	1	1	1	0	0	1
Segerstolpe et al., 2016	**Total Donor #**	3	0	3	6	1	2	1	4
	Male	3	0	2	5	0	1	1	2
	Female	0	0	1	1	1	1	0	2
Camunas-Soler et al., 2020	**Total Donor #**	1	4	13	18	1	4	2	7
	Male	1	2	6	9	0	2	1	3
	Female	0	2	7	9	1	2	1	4
Enge et al., 2017	**Total Donor #**	2	1	2	5	NA			
	Male	0	1	2	3				
	Female	2	0	0	2				
HPAP	**Total Donor #**	0	3	2	5	0	2	2	4
	Male	0	1	1	2	0	0	1	1
	Female	0	2	1	3	0	2	1	3
Total Donor #		9	12	27	48	4	11	8	23
Male		7	5	15	27	0	5	5	10
Female		2	7	12	21	4	6	3	13

Genotype is at SNP rs384275.

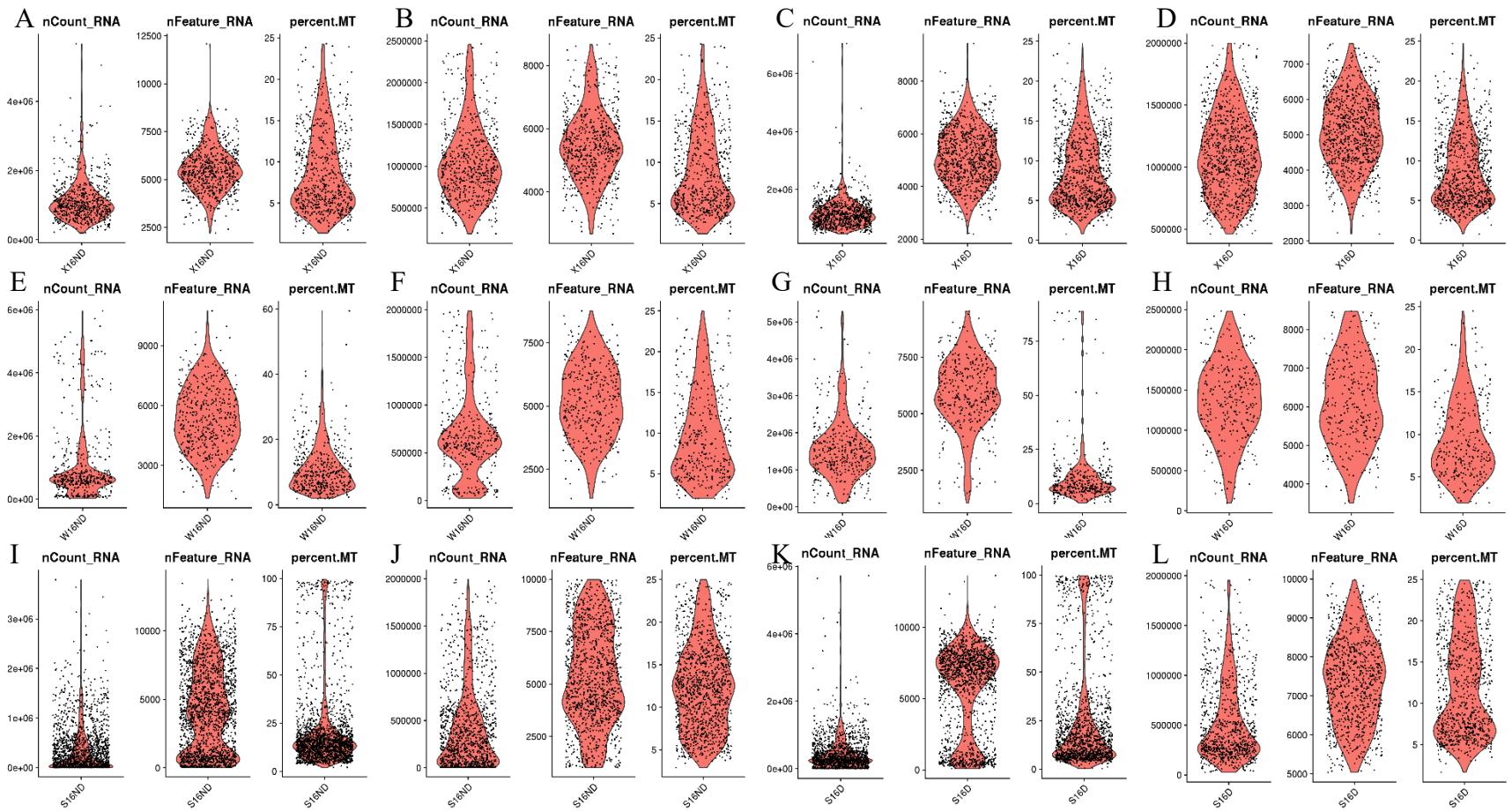


Figure 4: Quality control statistics for Xin (X16), Wang (W16), and Segerstolpe (S16) datasets. Unfiltered (A, C, E, G, I, and K) and filtered plots (B, D, F, H, J, and L) are shown. nCount_RNA represents number of RNA reads, nFeature_RNA represents number of genes, and percent.MT represents mitochondrial gene percent relative to total genes. Each point represents one cell. ND = non-diabetic, D = T2D.

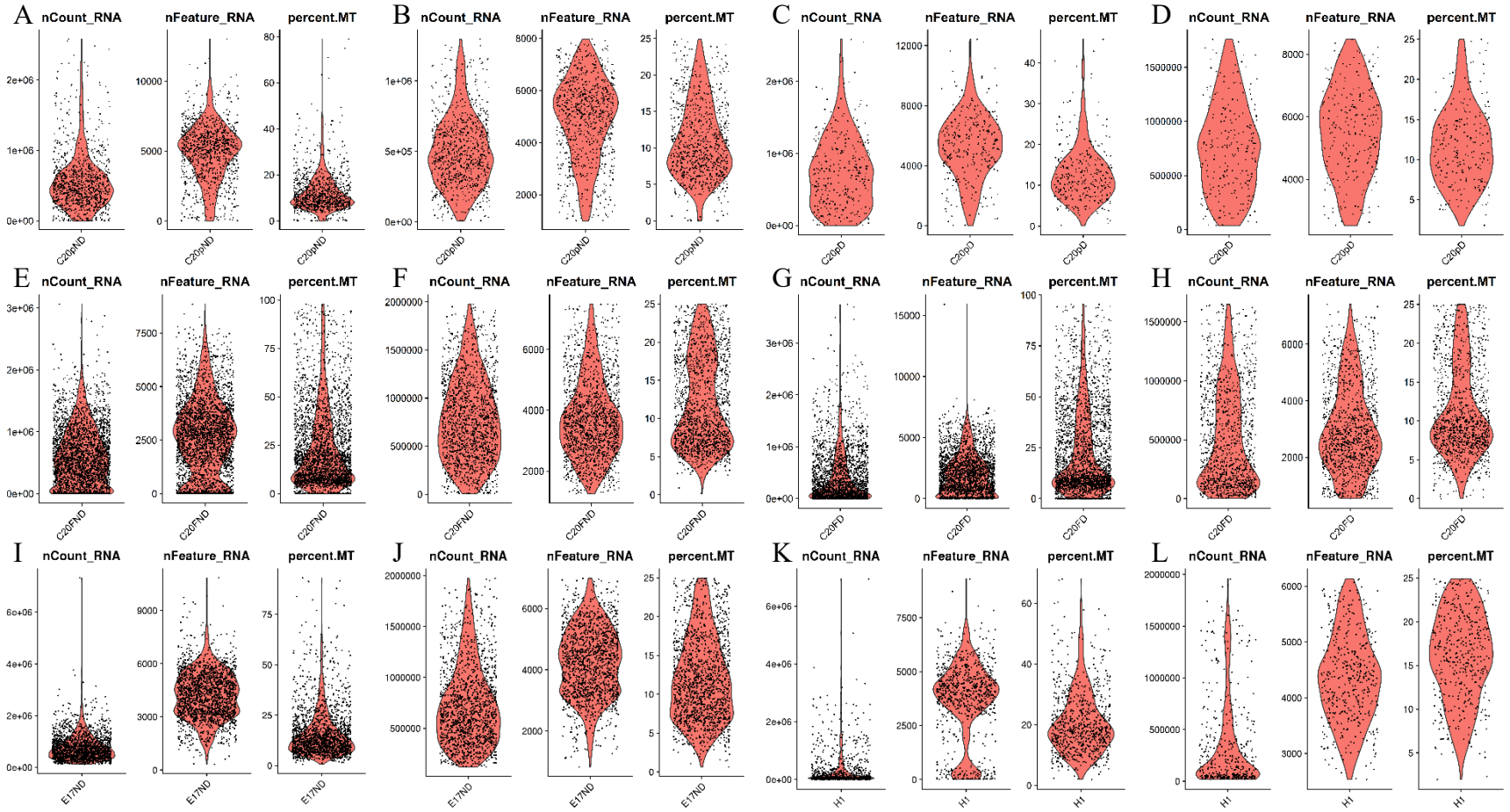


Figure 5: Quality control statistics for Camunas-Solar patch clamp (C20p) and FACS (C20F), Enge (E17), and HPAP-001 datasets. Unfiltered (A, C, E, G, I, and K) and filtered plots (B, D, F, H, J, and L) are shown. nCount_RNA represents number of RNA reads, nFeature_RNA represents number of genes, and percent.MT represents mitochondrial gene percent relative to total genes. Each point represents one cell. ND = non-diabetic, D = T2D.

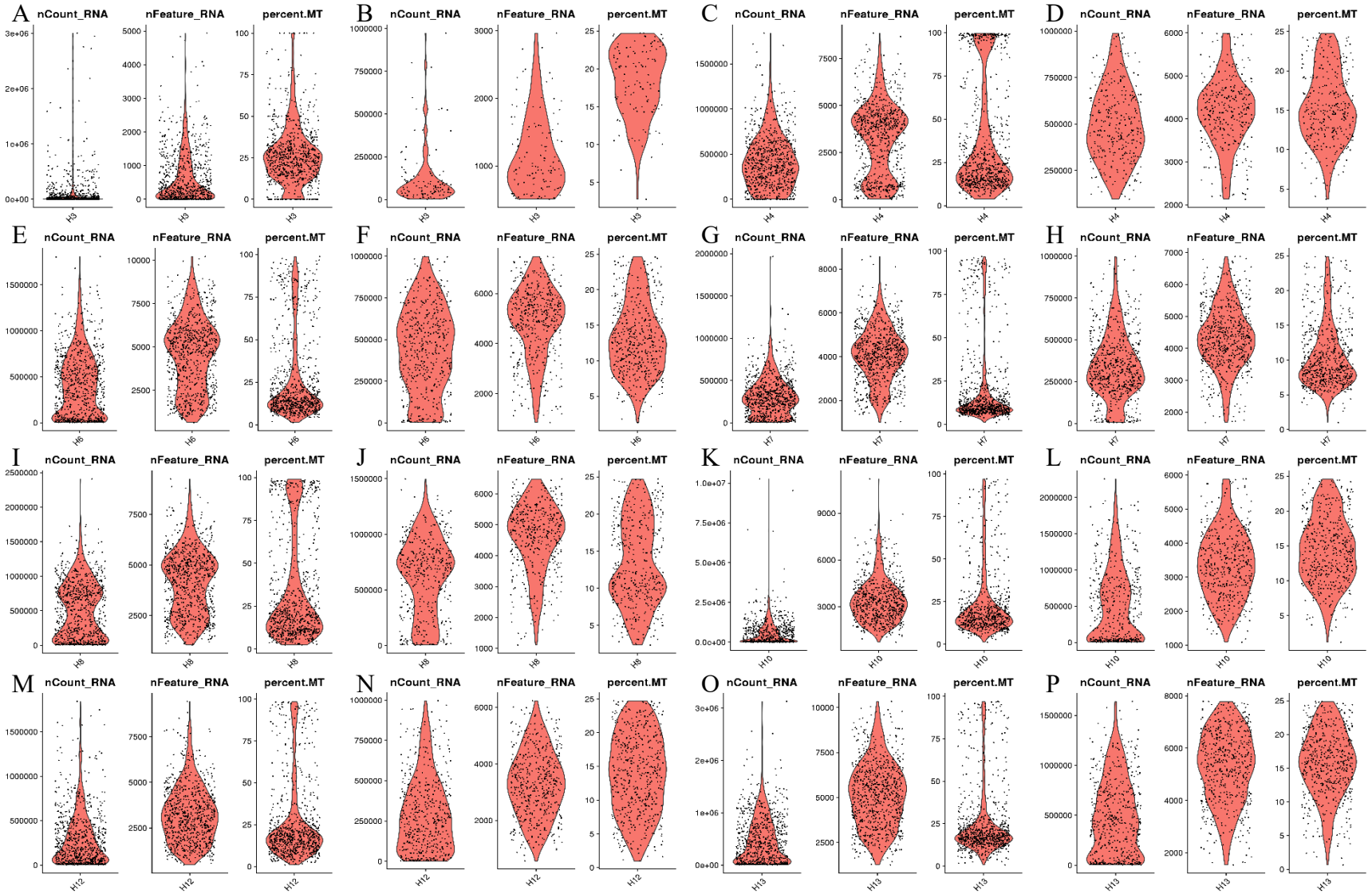


Figure 6: Quality control statistics for HPAP 003 to 013 datasets. Unfiltered (A, C, E, G, I, and K) and filtered plots (B, D, F, H, J, and L) are shown. nCount_RNA represents number of RNA reads, nFeature_RNA represents number of genes, and percent.MT represents mitochondrial gene percent relative to total genes. Each point represents one cell. ND = non-diabetic, D = T2D.

3.3: Cell Type Clustering and Identification

The scRNAseq data used included all the cells in pancreatic islets. Clustering was used to separate the different cell types/subtypes. Clusters were labeled based on default Seurat settings (Figure 7A). Clusters were formed based on cell type rather than datasets, disease status, or sex (Figure 7B to 7D). Based on pancreatic hormonal gene expression (*GCG, INS, PPY,* and *SST*) localization, clusters 0, 2, 4, and 12 were classified as α-cells (5519 cells), clusters 1, and 8 were classified as β-cells (2315 cells), cluster 7 was classified as PP cells (586 cells), and cluster 10 was classified as delta cells (451 cells) (Figure 8A, C, E). Based on the gene expression localization of *PECAM1*, *SPP1*, and *PRSS1*, cluster 11 was classified as endothelial cells (226 cells), cluster 3 was classified as duct cells (1502 cells), and cluster 6 was classified as acinar cells (1188 cells) (Figure 8B, C, E). Clusters 9 and 13 were classified as mesenchyme (531 cells), and CD 14+ monocytes (81 cells), respectively with Enrichr using the top ten conserved genes (specific cluster cells compared to all other cells) for each cluster (Figure 8C, E; Table 4 and 5). Cluster 5 was not clearly defined in UMAP space. We classified this cluster as Beta-like (1223 cells) because the seventh most conserved gene (specific cluster cells compared to all other cells) is *INS* (Figure 8C, E; Table 4). This cluster could indicate *INS* expression contamination during library preparation and/or sequencing. The identified clusters generally aligned with the clusters identified in the original publications. There were no clusters for epsilon cells. Similarly, Segerstolpe et al. (2016) labeled the epsilon cells based on the expression of *GHRL* rather than based on a distinct cluster. There is a discrepancy in the distribution in cell number for different cell types between *in vivo* pancreatic islets and our clustered cells. Specifically, we had more than 2 times more α-cells than β-cells compared to an *in vivo* human islet which comprises of around 30% α-cells, and 60% β-cells with the remaining 10% comprising of PP cells, delta cells, and epsilon cells (206). This could be because α-cells are more likely to survive high stress conditions (eg. islet isolation, islet dispersion, cell culture, and sample preparation) than β-cells. The β-cells were labelled with the genotype at rs384275. Across all datasets, there were 341 β-cells with the A/A homozygous allele, 819 with the A/C heterozygous, and 1015 with the homozygous C/C allele (Table 6).

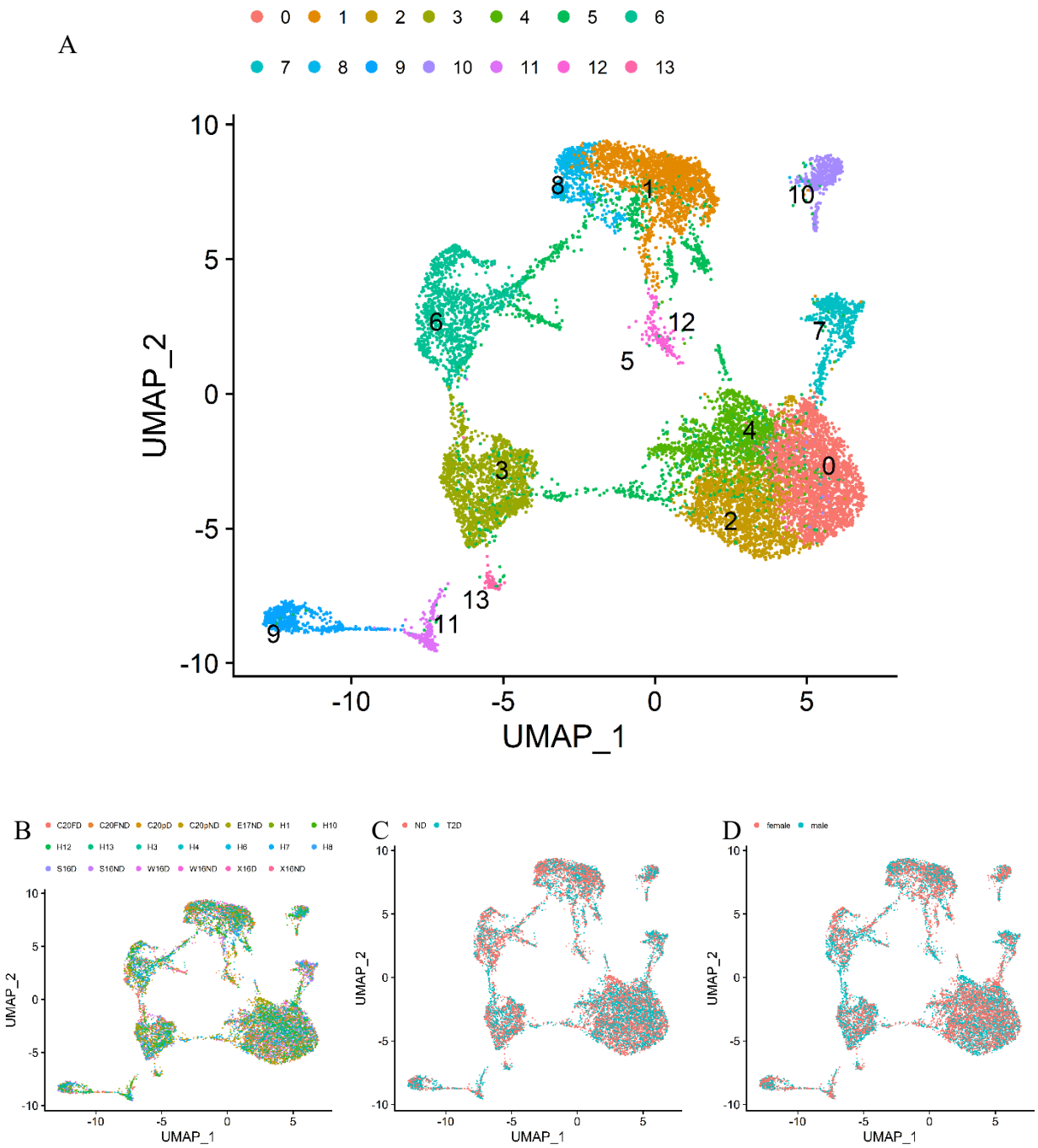
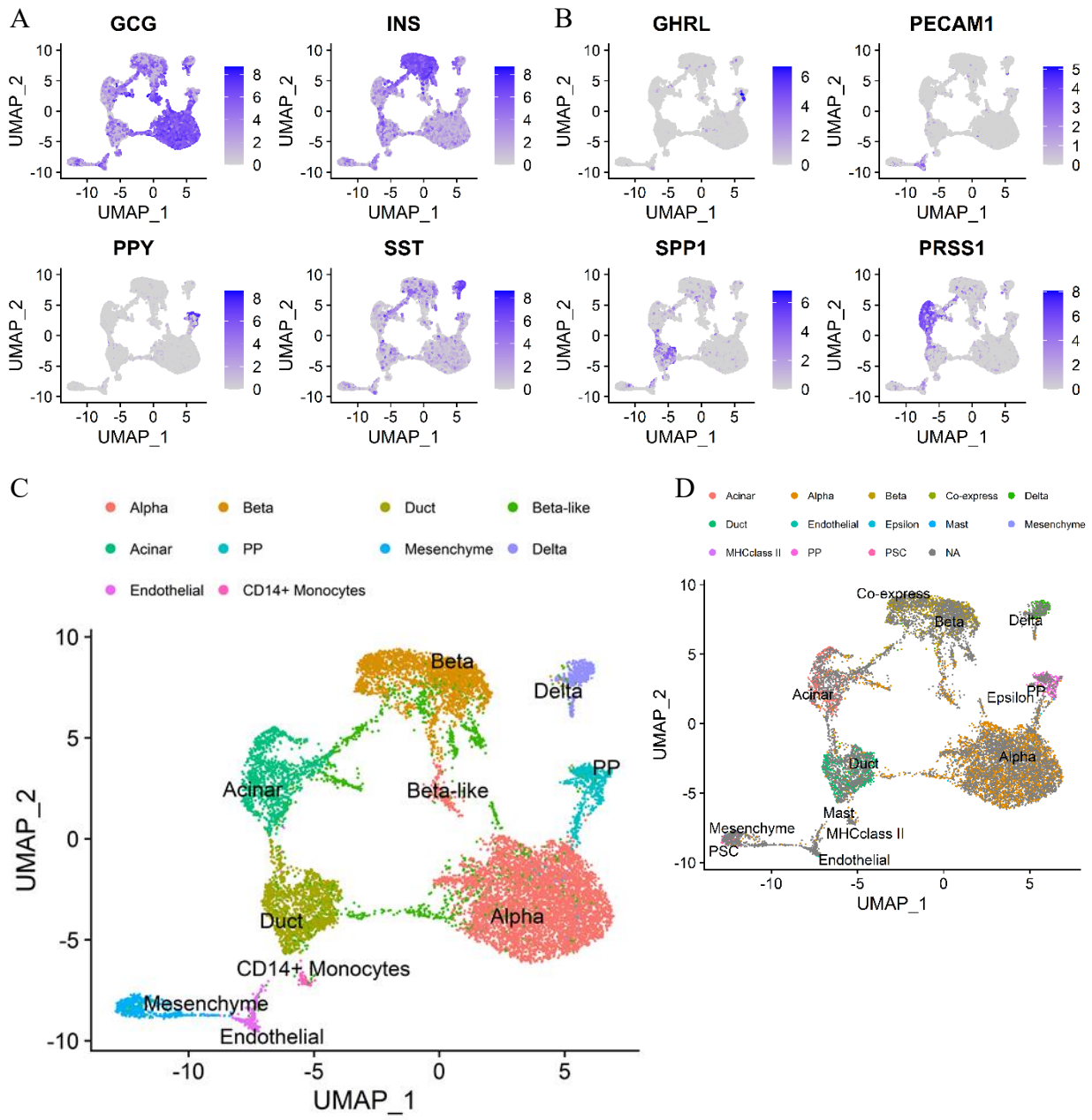


Figure 7: Clustering of integrated dataset in UMAP space. (A) UMAP projection of integrated dataset with clusters. UMAP projection of integrated dataset grouped by dataset (B), disease status (C), and sex (D). Each point represents one cell. ND = non-diabetic and D = diabetic.



E

Alpha	Beta	Duct	Beta-like	Acinar	PP	Mesen-chyme	Delta	Endo-thelial	CD14+ Monocyte
5519	2315	1502	1223	1188	586	531	451	226	81

Figure 8: Cell type identification. (A and B) UMAP projection of integrated dataset with gene expression localization for pancreatic hormonal genes and other cell type markers. Cells coloured by amount of gene expression. UMAP projection of integrated dataset with clusters labeled and coloured by cell type identity (C) and cell type identification from original publications (D). Each point represents one cell. (E) Table with each identified cluster cell number.

Table 4: Top ten genes differentially expressed in cluster 5.

Gene	P value	Average Log(fold change)	Pct.1	Pct.2	Adjusted P value	Cluster
UBB	1.44E-68	-0.32015	0.718	0.907	8.84E-64	5
PRRG3	4.93E-56	0.827248	0.551	0.358	3.02E-51	5
FTH1	1.74E-52	-0.2891	0.623	0.843	1.07E-47	5
HMGN2	5.72E-52	-0.39412	0.601	0.81	3.50E-47	5
TECR	2.57E-46	-0.25566	0.447	0.663	1.57E-41	5
INS	8.20E-43	0.387695	0.902	0.863	5.02E-38	5
MT-ND4L	2.16E-39	-0.28884	0.947	0.98	1.32E-34	5
PDIA6	7.31E-39	-0.27726	0.497	0.671	4.48E-34	5
GREM1	1.05E-35	0.637366	0.516	0.365	6.43E-31	5
PDIA3	2.75E-35	-0.32448	0.806	0.913	1.68E-30	5

Average Log(fold change) is the log fold-change of the average expression between the specified cluster and all other cells. Pct.1 is the percentage of cells where the gene is detected in the specified cluster. Pct.2 is the percentage of cells where the gene is detected all other cells. P values are adjusted using the Bonferroni correction using all genes in dataset.

Table 5: Top ten genes differentially expressed in cluster 13.

Gene	P value	Average Log(fold change)	Pct.1	Pct.2	Adjusted P value	Cluster
LAPTM5	0	2.440237	0.815	0.016	0	13
C1QA	0	2.240455	0.395	0.002	0	13
TYROBP	0	2.140109	0.58	0.006	0	13
C1QB	0	2.020045	0.395	0.003	0	13
C1QC	0	1.617708	0.358	0.003	0	13
CD53	0	1.141412	0.333	0.004	0	13
PIK3R5	0	1.036569	0.333	0.004	0	13
RUNX3	0	0.888774	0.457	0.007	0	13
LST1	0	0.527813	0.321	0.002	0	13
LILRB2	3.34E-307	0.571122	0.259	0.002	2.05E-302	13

Average Log(fold change) is the log fold-change of the average expression between the specified cluster and all other cells. Pct.1 is the percentage of cells where the gene is detected in the specified cluster. Pct.2 is the percentage of cells where the gene is detected all other cells. P values are adjusted using the Bonferroni correction using all genes in dataset.

Table 6: Total and beta cell specific number by sex and genotype for each dataset.

		Non-Diabetic				T2D			
Study		**A/A**	**A/C**	**C/C**	**Total**	**A/A**	**A/C**	**C/C**	**Total**
Xin et al., 2016	**Total Cell #**	NA			625	NA			905
	Total Beta Cell #	68	41	45	154	11	97	112	220
	Male	68	16	32	116	0	79	9	88
	Female	0	25	13	38	11	18	103	132
Wang et al., 2016	**Total Cell #**	NA			378	NA			297
	Total Beta Cell #	0	0	106	106	47	0	16	63
	Male	0	0	31	31	0	0	16	16
	Female	0	0	75	75	47	0	0	47
Segerstolpe et al., 2016	**Total Cell #**	NA			1225	NA			868
	Total Beta Cell #	48	0	101	149	17	4	17	38
	Male	48	0	69	117	0	3	17	20
	Female	0	0	32	32	17	1	0	18
Camunas-Soler et al., 2020 (patch clamp)	**Total Cell #**	NA			862	NA			233
	Total Beta Cell #	4	63	125	192	2	31	37	70
	Male	4	8	43	55	0	12	1	13
	Female	0	55	82	137	2	19	36	57
Camunas-Soler et al., 2020 (FACS)	**Total Cell #**	NA			1648	NA			1188
	Total Beta Cell #	60	36	274	370	45	86	4	135
	Male	60	23	231	314	0	53	4	57
	Female	0	13	43	56	45	33	0	78

Enge et al., 2017	**Total Cell #**	NA			1451	NA			
	Total Beta Cell #	56	51	39	146				
	Male	0	51	39	90				
	Female	56	0	0	56				
HPAP	**Total Cell #**	NA			1865	NA			2080
	Total Beta Cell #	0	209	96	305	0	202	75	277
	Male	0	51	6	57	0	0	31	31
	Female	0	158	90	248	0	202	44	246
Total Beta Cell #		236	400	754	1390	105	419	261	785
Male		180	149	451	780	0	147	78	225
Female		56	251	335	642	122	273	183	578

Genotype is at SNP rs384275.

3.4: Insulin Expression

3.4.1: Insulin Normalization

Gene expression normalization is extremely important in RNAseq analysis as it corrects for differences between different sequencing coverage while preserving biological variance. Using the default Seurat normalization, whereby *INS* expression was normalized against all genes in the cells, we found large inter-dataset and intra-dataset variance (Figure 9A). This could be due to the overwhelming amount of *INS* mRNA in β-cells. Therefore, we proceeded to internally normalize the *INS* expression to housekeeping genes. First, we had to determine empirically which housekeeping genes would provide the optimal normalization. The two housekeeping genes with the lowest inter-dataset variability are *B2M* and *RPLP0* (Figure 9G, H). These two housekeeping genes were therefore used for subsequent *INS* normalization. After normalization with housekeeping genes, there is visibly less inter-dataset variability in *INS* expression (Figure 10). We adopted this approach going forward.

3.4.2: Insulin Variance and Mean Expression by Sex and Disease Status

Sex and diabetes status have been repeatedly shown to affect gene expression profiles in many tissues, including pancreatic islets. It is important to test whether sex and/or diabetes status were affecting *INS* expression in order to study the effect of rs3842753 genotype on pancreatic *INS* expression. The variance and mean of normalized *INS* expression in β-cells were both significantly lower in males compared to females (male: var = 2.41, mean = 11.25, SD = 1.55; females: var = 2.95, mean = 11.56, SD = 1.72; variance p value = 1.938e-10, mean p value = 1.00e-05) (Figure 11A, B). The variance of normalized *INS* expression in β-cells was lower in non-diabetic donors compared to T2D donors with no significant different in mean (non-diabetic: var = 2.52; T2D: var = 3.11; variance p = 0.008) (Figure 11C, D). The variance within either sex across diabetes status was not significantly different. Interestingly in males, there was a decrease in mean of normalized *INS* expression in T2D compared to non-diabetic cells (male non-diabetic: mean = 11.27, SD = 1.57; male T2D: mean = 11.14, SD = 1.50; p value = 0.036) (Figure 11E, F). The difference in *INS* expression between males and females indicates that sex plays a role in insulin expression. The male specific lowering of *INS* expression in T2D compared to non-diabetic indicates diabetes status play a role in insulin expression. The

difference in variance of *INS* expression supports the heterogeneous β-cell population paradigm in a sex and diabetes status dependent manner.

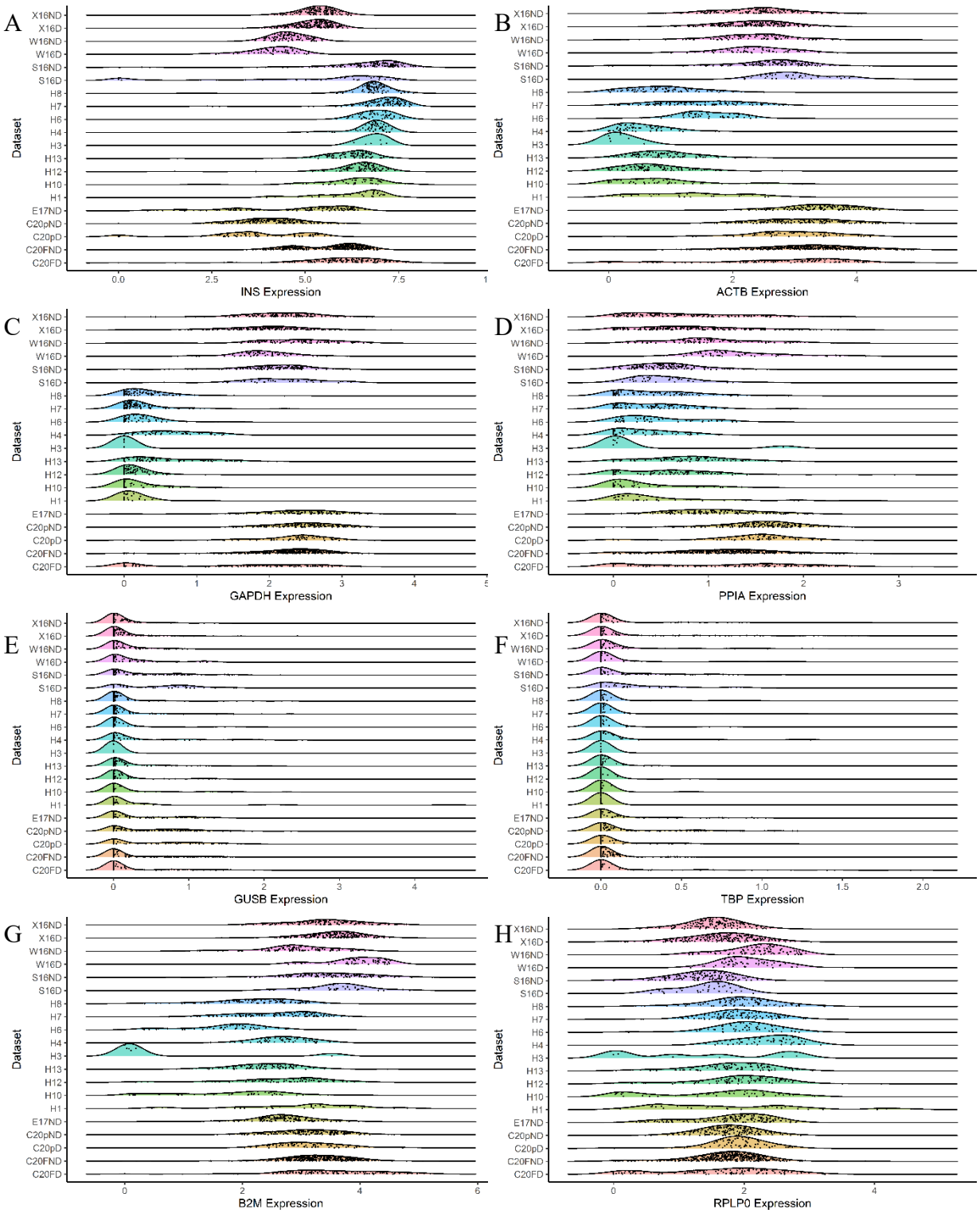


Figure 9: Default normalized INS and housekeeping genes expression by dataset. (A) *INS* expression normalized to all genes. (B) Housekeeping genes expression normalized to all genes. Gene expression is separated and coloured by datasets. Each point represents one cell. ND = non-diabetic, D = T2D.

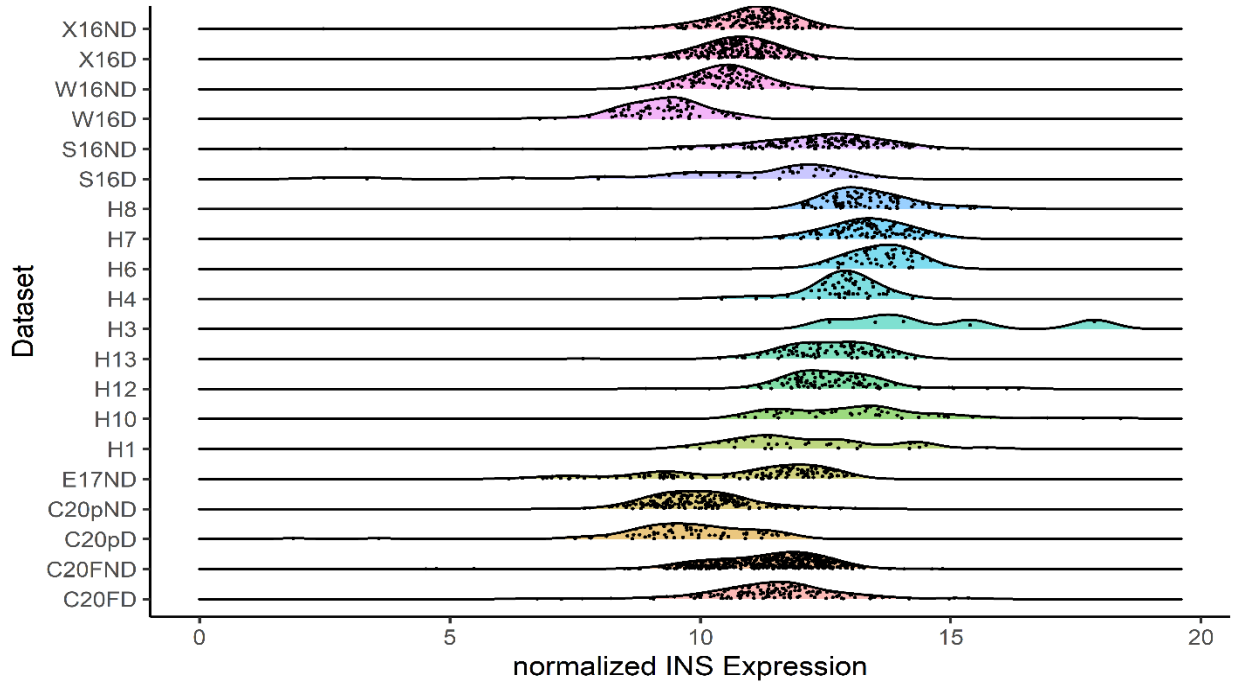


Figure 10: Normalized INS expression in beta cells by dataset. Normalized *INS* expression separated and coloured by dataset. Each point represents one cell.

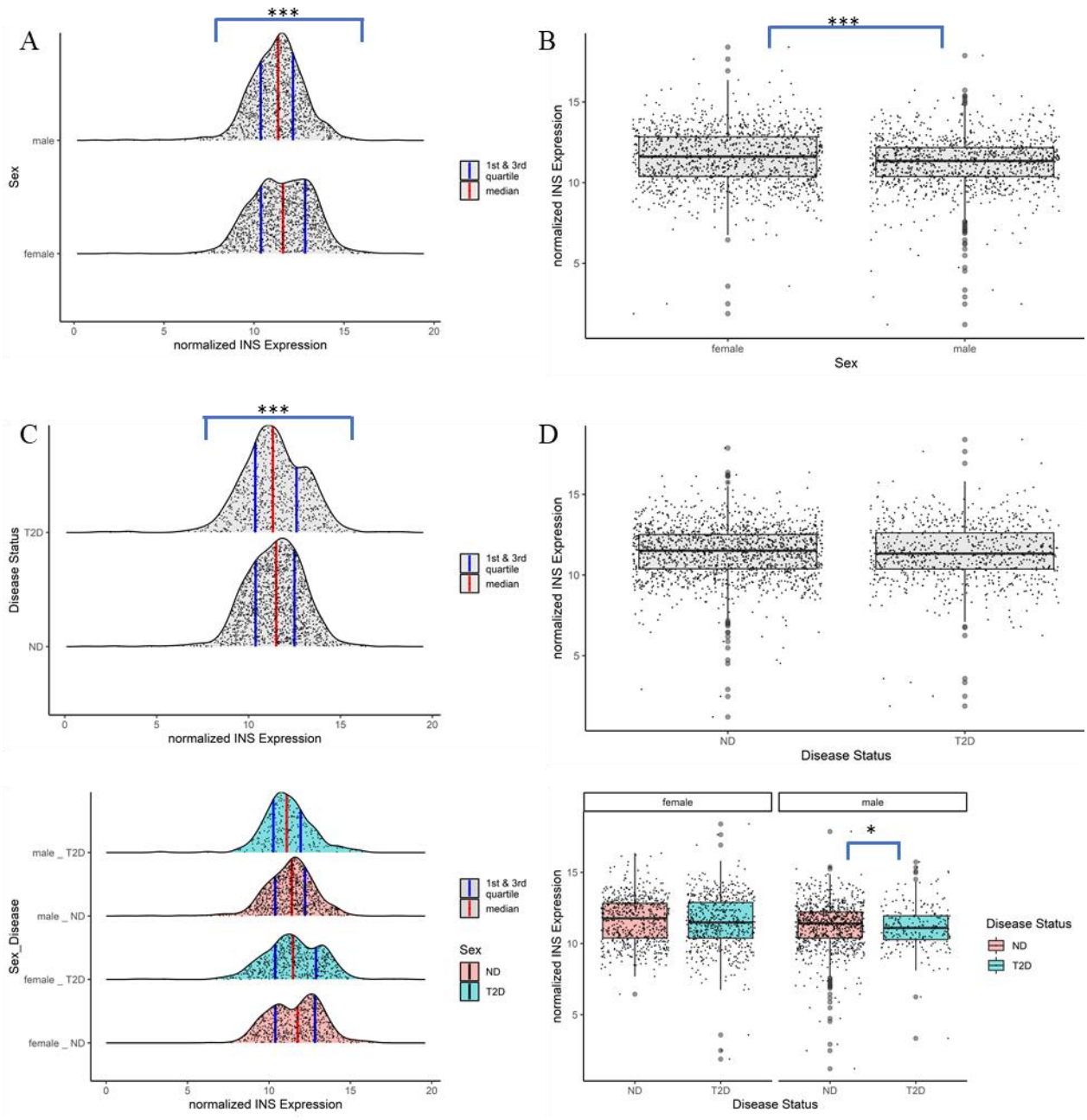


Figure 11: Normalized INS expression in beta cells by sex and disease status. Distribution and box plot of normalized *INS* expression separated by sex (A, B), disease status (C, D), and combined sex and disease status (E, F). Red lines represent median, blue lines represent 1st and 3rd quartiles. Each point represents one cell. P < 0.05 = *, P < 0.001 = ***. Graphs are coloured by disease status (E, F). Fligner-Killeen and Wilcox Rank Sum tests were used as non-parametric test for differences in variance and mean respectively.

3.4.3: Insulin Expression by Genotype

Previous studies have shown a correlation between the 'C' T1D at-risk rs3842753 allele and increased pancreatic *INS* expression in 10 heterozygous whole pancreas tissues, using allele specific expression. The sample size in these previous studies was small and *INS* expression was studied in whole islets, rather than single pancreatic β-cells. Here, we focused on the effects of rs3842753 genotype on pancreatic *INS* expression in single pancreatic β-cells. The variance of normalized *INS* expression in β-cells was significantly different between the three genotypes (A/A: var = 2.71; A/C: var = 3.03; C/C: var = 2.37; p value = 3.38e-11) (Figure 12A). Donors with the protective A/A homozygous genotype had the lowest normalized *INS* expression, followed by the at-risk C/C homozygous donors, with A/C heterozygous donors having the highest normalized *INS* expression (A/A: mean = 10.92, SD = 1.65; A/C: mean = 12.00, SD = 1.74; C/C mean = 11.36, SD = 1.54, adjusted p value < 3.27e-06) (Figure 12B). In addition to cell-level comparisons, we also tested whether the mean and median of normalized *INS* expression were different between donors (donor level comparisons). In this analysis, although some of the same tendencies were observed, we found that there was no significant difference in variance or mean between genotypes, perhaps because it was statistically underpowered (Figure 12C, D).

When examining the effect of rs3842753 genotype within sex, there was significant difference in both the variance and mean of normalized *INS* expression between genotypes. Males had significantly different variance between all three genotypes (A/A: var = 2.71; A/C: var = 3.44; C/C var = 1.82; p value = 7.51e-16) (Figure 13A). Male A/C heterozygous donors had the lowest normalized *INS* expression, followed by the protective A/A homozygous donors, with at-risk C/C homozygous donors having the highest normalized *INS* expression (A/A: mean = 11.31, SD = 1.65; A/C: mean = 11.03, SD = 1.86; mean = 11.33, SD = 1.35, adjusted p value < 1.19e-06) (Figure 13B). For females, there was only significant difference in variance between the A/C heterozygous donors and C/C homozygous donors (A/C: var = 2.4; C/C var = 3.02; p value = 0.03) (Figure 13A). Female A/A homozygous donors had the lowest normalized *INS* expression, followed by C/C homozygous donors, with A/C heterozygous donors having the highest normalized *INS* expression (A/A: mean = 10.51, SD = 1.55; A/C: mean = 12.08, SD = 1.55; mean = 11.39, SD = 1.74, adjusted p value < 0.01) (Figure 13B). When the donor specific mean and median of normalized *INS* expression were examined, there was no significant

difference in variance or mean between genotypes for either sex, although the same tendencies remained (Figure 13C, D).

Next, we examined the effect of genotype on β-cell *INS* expression in the context of disease status. Non-diabetic donors had significantly different variance between the three genotypes (A/A: var = 2.3; A/C: var = 3.65; C/C var = 2.07; p value < 2.2e-16) (Figure 14A). Non-diabetic donors with the protective A/A homozygous genotype had the lowest normalized *INS* expression, followed by at-risk C/C homozygous donors, with A/C heterozygous donors having the highest normalized *INS* expression (A/A: mean = 11.37, SD = 1.52; A/C: mean = 11.47, SD = 1.91; mean = 11.42, SD = 1.44, adjusted p value < 1.65e-05) (Figure 14B). For T2D donors, there was no significant difference in variance between the genotypes (Figure 14A). However, there was a significant difference in mean normalized *INS* expression with A/A homozygous donors have the lowest normalized *INS* expression, followed by C/C homozygous donors, with A/C heterozygous donors having the highest normalized *INS* expression (A/A: mean = 10.00, SD = 1.52; A/C: mean = 11.93, SD = 1.53; mean = 11.16, SD = 1.83, adjusted p value < 9.83e-09) (Figure 14B). Donor level comparisons showed no significant differences in either variance or mean for donor specific mean and median of normalized *INS* expression (Figure 14C, D). Together, these analyses show the rs3842753 C allele affects *INS* expression, with the A/C genotype associated with both increased pancreatic *INS* expression variance and level at the cell level.

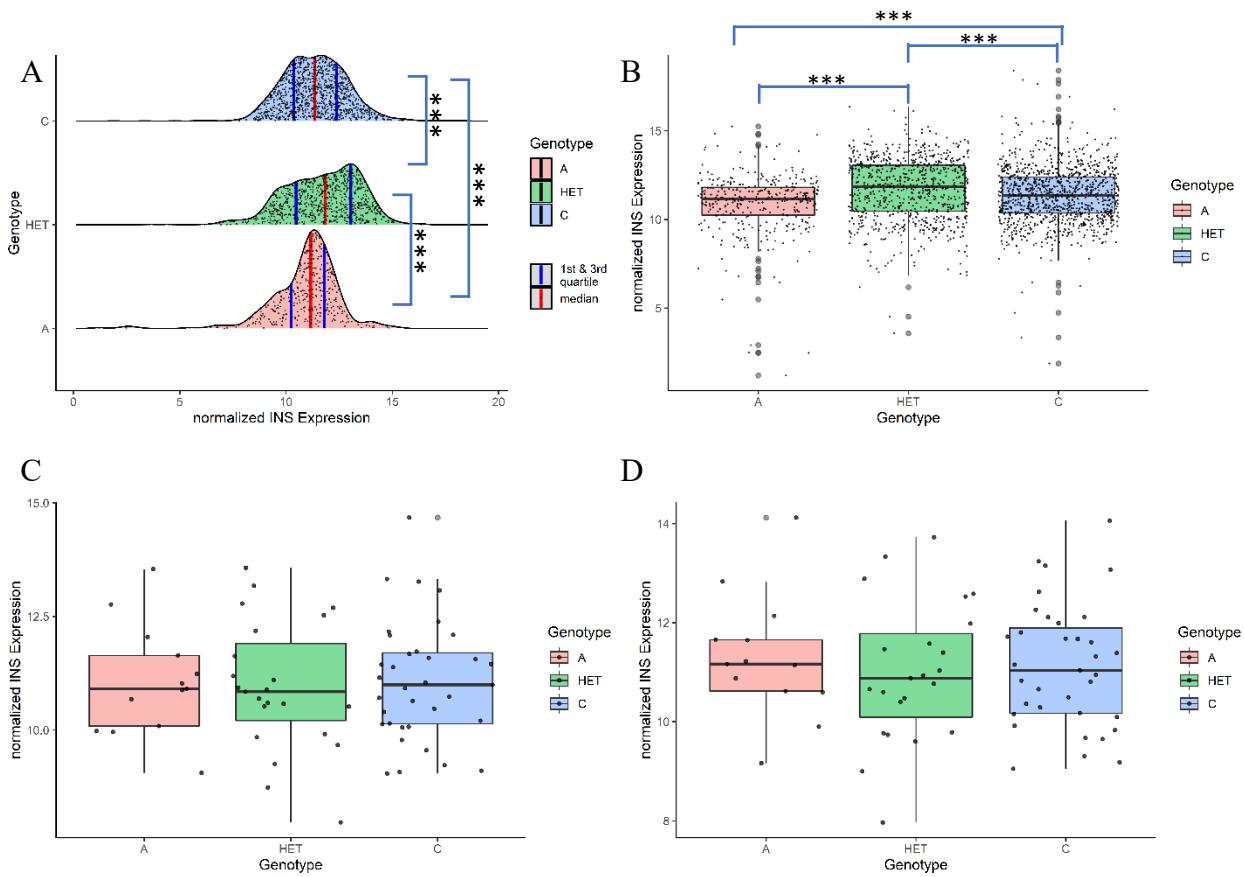


Figure 12: Raw, mean and median normalized INS expression in β-cells by genotype.
Cell specific distribution and box plot of normalized *INS* expression separated by genotype (A, B). Donor specific mean and median normalized *INS* expression box plot separated by genotype (C, D). Red lines represent median, blue lines represent 1st and 3rd quartiles. In A and B, each point represents one cell. In C and D, each point represents one donor. P < 0.001 = ***. Graphs are coloured by genotype. A represent A/A genotype, HET represent A/C genotype, C represent A/C genotype at rs384275. For cell specific analyses, Fligner-Killeen and pairwise Kruskal-Wallis tests were used as non-parametric test for differences in variance and mean respectively. For donor specific analyses, Bartlett and ANOVA tests were used as parametric test for differences in variance and mean respectively.

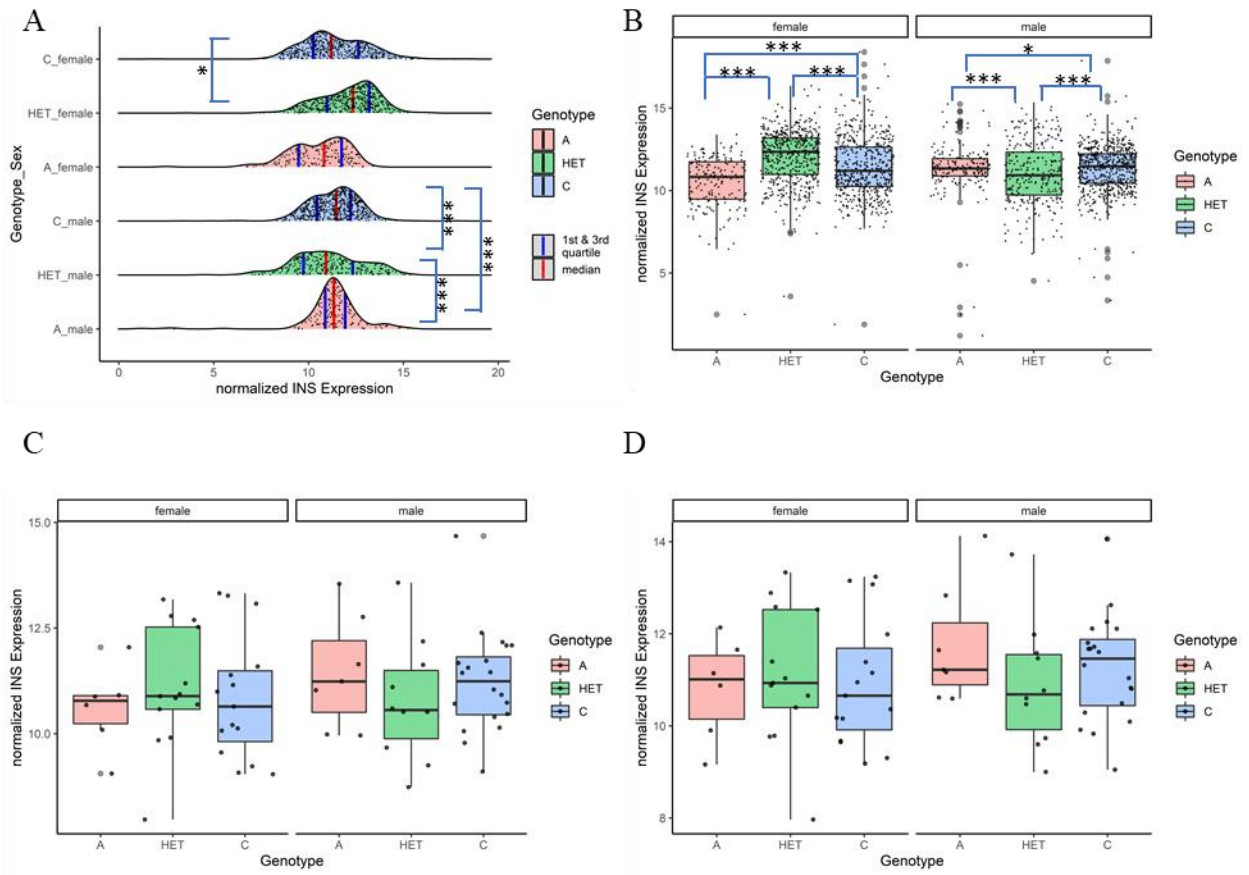


Figure 13: Raw, mean and median normalized INS expression in β-cells by genotype and sex. Distribution and box plot of normalized *INS* expression separated by genotype and sex (A, B). Donor specific mean and median normalized *INS* expression box plot separated by genotype and sex (C, D). Red lines represent median, blue lines represent 1st and 3rd quartiles. In A and B, each point represents one cell. In C and D, each point represents one donor. P < 0.05 = *, P < 0.001 = ***. Graphs are coloured by genotype. A represent A/A genotype, HET represent A/C genotype, C represent A/C genotype at rs384275. For cell specific analyses, Fligner-Killeen and pairwise Kruskal-Wallis tests within sex were used as non-parametric test for differences in variance and mean respectively. For donor specific analyses, Bartlett and ANOVA tests within sex were used as parametric test for differences in variance and mean respectively.

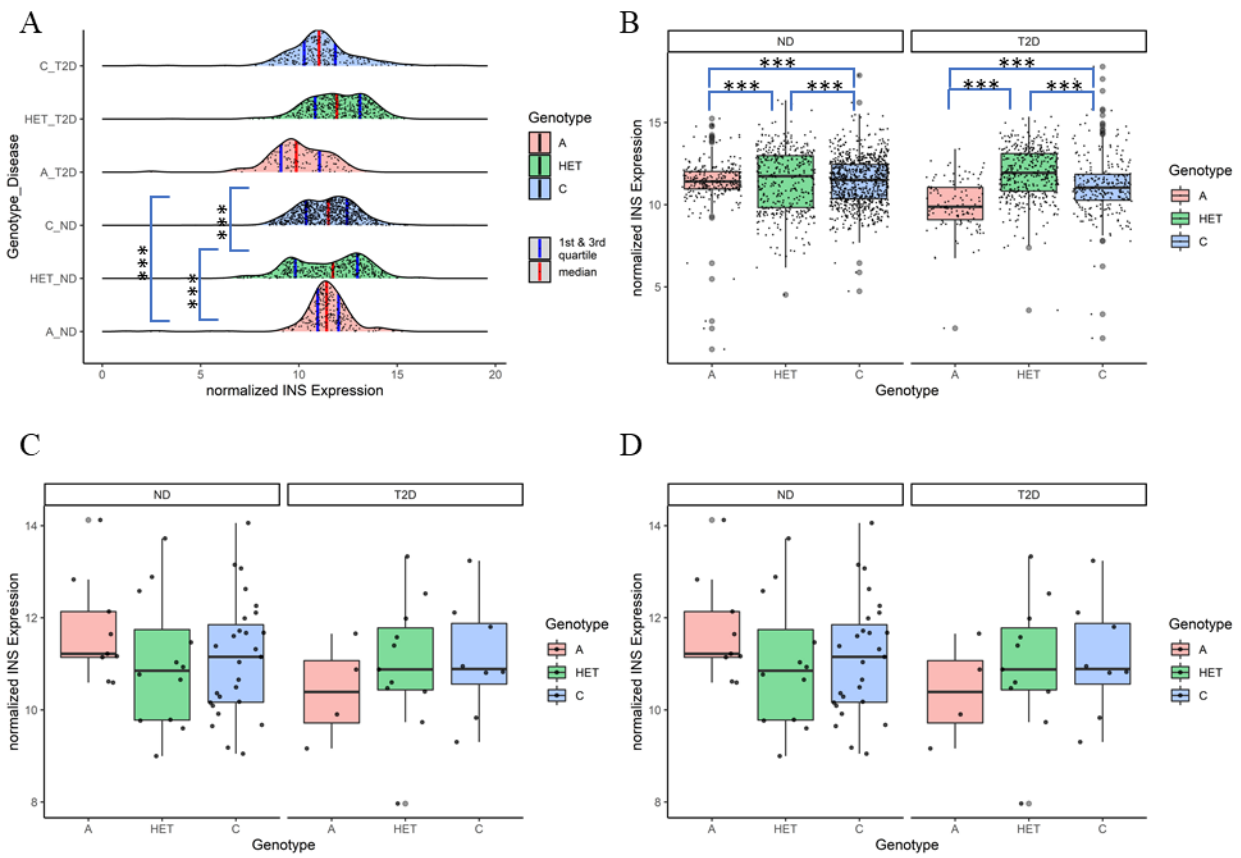


Figure 14: Raw, mean and median normalized INS expression in beta cells by genotype and disease status. Distribution and box plot of normalized *INS* expression separated by genotype and disease status (A, B). Donor specific mean and median normalized *INS* expression box plot separated by genotype and sex (C, D). Red lines represent median, blue lines represent 1st and 3rd quartiles. In A and B, each point represents one cell. In C and D, each point represents one donor. P < 0.001 = ***. Graphs are coloured by genotype. ND = non-diabetic. A represent A/A genotype, HET represent A/C genotype, C represent A/C genotype at rs384275. For cell specific analyses, Fligner-Killeen and pairwise Kruskal-Wallis tests within disease status were used as non-parametric test for differences in variance and mean respectively. For donor specific analyses, Bartlett and ANOVA tests within disease status were used as parametric test for differences in variance and mean respectively.

3.5: rs3842753 Allele Specific Expression for Heterozygous Donors

Another approach to test the hypothesis that the rs3842753 genotype might affect *INS* mRNA expression is to examine the allele specific transcript levels within cells from heterozygous donors. This approach potentially eliminates some technical sources of error between donors while preserving genotype specific effects on *INS* expression. There was no difference in the variance between heterozygous protective reference A allele percent and theoretical normal distribution at either the cell level or donor level (Figure 15A, C). When examining the mean reference allele (A) percent, the A allele specific expression in heterozygous was significantly higher than the theoretical 50% both at the cell level and donor level (HET: cell level mean = 61.39%, SD = 9.10; donor level mean = 59.06%, SD = 7.25; Theory: cell level mean = 49.70%, SD = 10.1; donor level mean = 48.03%, SD = 10.6; p value < 0.0002) (Figure 15B, D). This suggests that the protective reference A allele is preferentially transcribed compared to the at-risk C allele in one cell/donor. This surprising result is inconsistent with established literature which found the C allele to be overexpressed in heterozygous donors.

There was a difference in both the variance and mean protective reference A allele percent between either sex and the theoretical distribution. At the cell level, both sexes had significantly different variance for reference A allele percent expression from the normal distribution (theory var = 102.0; male: var = 52.8; female: var = 76.6; p value < 3.501e-05) (Figure 16A). As well, both sexes had significantly different mean reference A allele percent expression from the normal distribution (Theory: mean = 49.70%, SD = 10.1; male mean 56.32%, SD = 7.26; female: mean = 64.32%, SD = 8.75; p value < 2.2e-16) (Figure 16B). This difference in mean A allele percent remained when examining at the donor level with females consistently expressing higher A allele than 50% (female mean = 60.10%, SD = 6.83; median = 60.33%, SD = 7.13; Theory: mean = 48.03%, SD = 10.59; median = 52.27%, SD = 10.12; p value < 0.009) (Figure 16C, D). For males at the donor level, only the mean A allele expression percent was significantly higher than theoretical (male mean = 57.69%, SD = 7.91; Theory: mean = 48.03%, SD = 10.59; p value < 0.008) (Figure 16C, D).

Next, we examined the difference in the protective reference A allele percent variance and mean between either non-diabetic donors or T2D donors and the theoretical distribution. There was no difference in variance of reference A allele percent between either disease status

and the theoretical normal distribution (Figure 17A). Both disease statuses had significantly different mean reference A allele percent expression from the normal distribution (Theory: mean = 49.70%, SD = 10.1; non-diabetic mean 61.22%%, SD = 9.46; T2D: mean = 61.54%, SD = 8.72; p value < 2.2e-16) (Figure 17B). This difference in mean A allele percent remained when examining at the donor level with non-diabetic donors consistently expressing higher A allele than 50% (non-diabetic mean = 59.42%, SD = 7.23; median = 59.28%, SD = 7.13; Theory: mean = 48.03%, SD = 10.59; median = 52.27%, SD = 10.12; p value < 0.02) (Figure 17C, D). For T2D at the donor level, only the mean A allele expression percent was significantly higher than theoretical (T2D mean = 58.66%, SD = 7.60; Theory: mean = 48.03%, SD = 10.59; p value < 0.002) (Figure 17C, D). Taken together, these results indicate an apparently robust increase in transcription from the protective rs3842753 A allele within heterozygous A/C genotype pancreatic β-cells, in contrast to our comparisons between donor cells and in opposition with the previous literature that showed the rs3842753 C allele was preferentially expressed in heterozygous whole pancreas tissue. Some sequencing technical challenges are present given the high GC content in the region surrounding rs3842753 which could preferentially increase the sequencing efficiency of the reference A allele compared to the alternate C allele. This could contribute to the increase in A allele read counts compared to the C allele. Further studies will be required to resolve this issue.

3.6: Insulin Protein Abundance by Genotype

Due to the inconsistent data from *INS* expression and allele specific expression, we were interested in whether there was relationship between rs3842753 genotype and insulin protein abundance. We had the opportunity to conduct a small exploratory pilot study. We had access to whole islet insulin protein abundance for donors in Camunas et al. (2020) and the HPAP dataset. Only donors in Camunas et al. (2020) were used as there were no donors in the HPAP dataset with rs3842753 A/A homozygous allele (Table 3). Insulin protein abundance was not different between sexes, disease statuses or genotype (Figure 18A, B). As well, there was no difference in insulin protein abundance between genotypes after taking sex and disease status into account (Figure 18C to D). This is a pilot study using whole pancreatic islets, rather than pure or single β-cells, with a very sample size (n = 23 total, but with n = 2 A/A donors). Taken together, the

results point to a lack of evidence supporting the association between rs3842753 C allele and increased insulin protein production.

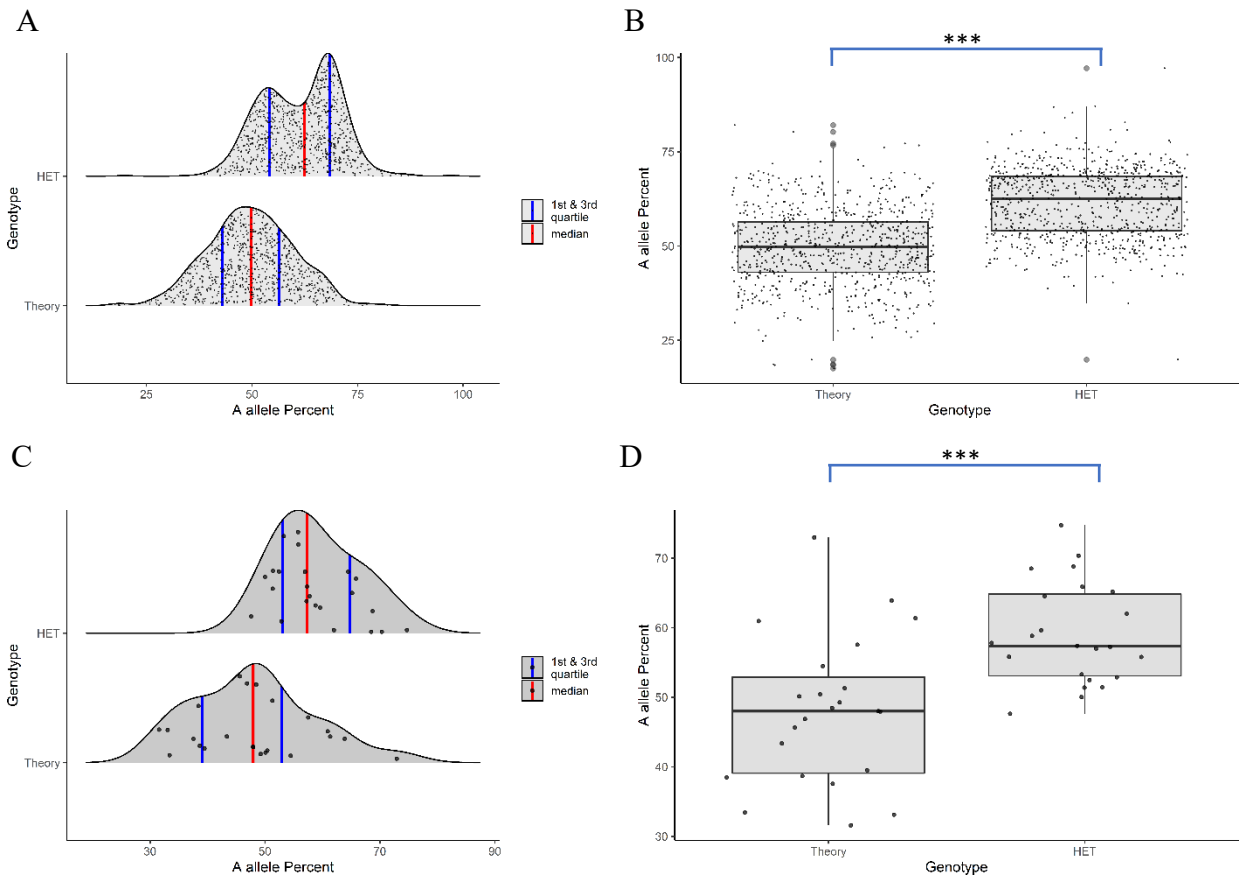


Figure 15: rs3842753 reference allele percent for heterozygous donors. Cell specific (A, B), and donor specific mean (C, D) of rs3842753 reference allele percent distribution and box plot separated by genotype against theoretical normal distribution with mean at 50. Red lines represent median, blue lines represent 1st and 3rd quartiles. In A and B, each point represents one cell. In C and D, each point represents one donor. P < 0.001 = ***. HET = heterozygous, theory = theoretical distribution. For cell specific analyses, Fligner-Killeen and Wilcox Rank Sum tests were used as non-parametric test for differences in variance and mean respectively. For donor specific analyses, Bartlett and student T tests were used as parametric test for differences in variance and mean respectively.

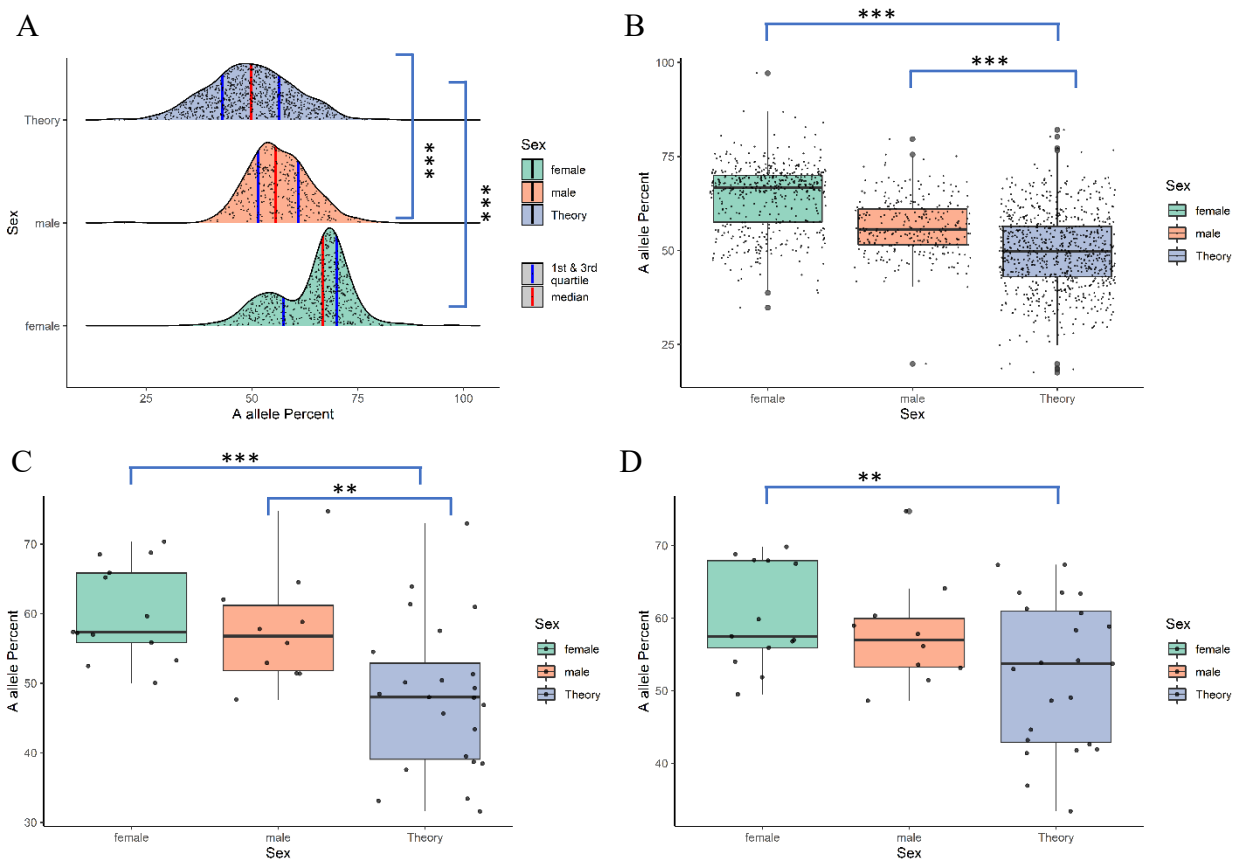


Figure 16: rs3842753 reference allele percent separated by sex for heterozygous donors. Cell specific (A, B), donor specific mean and median (C, D), of rs3842753 reference allele percent distribution and box plot separated by sex against theoretical normal distribution with mean at 50. Red lines represent median, blue lines represent 1st and 3rd quartiles. In A and B, each point represents one cell. In C and D, each point represents one donor. P < 0.01 = **. P < 0.001 = ***. Graphs are coloured by sex. Theory = theoretical distribution. For cell specific analyses, Fligner-Killeen and Wilcox Rank Sum tests between either sex and theoretical were used as non-parametric test for differences in variance and mean respectively. For donor specific analyses, Bartlett and Student T tests between either sex and theoretical were used as parametric test for differences in variance and mean respectively.

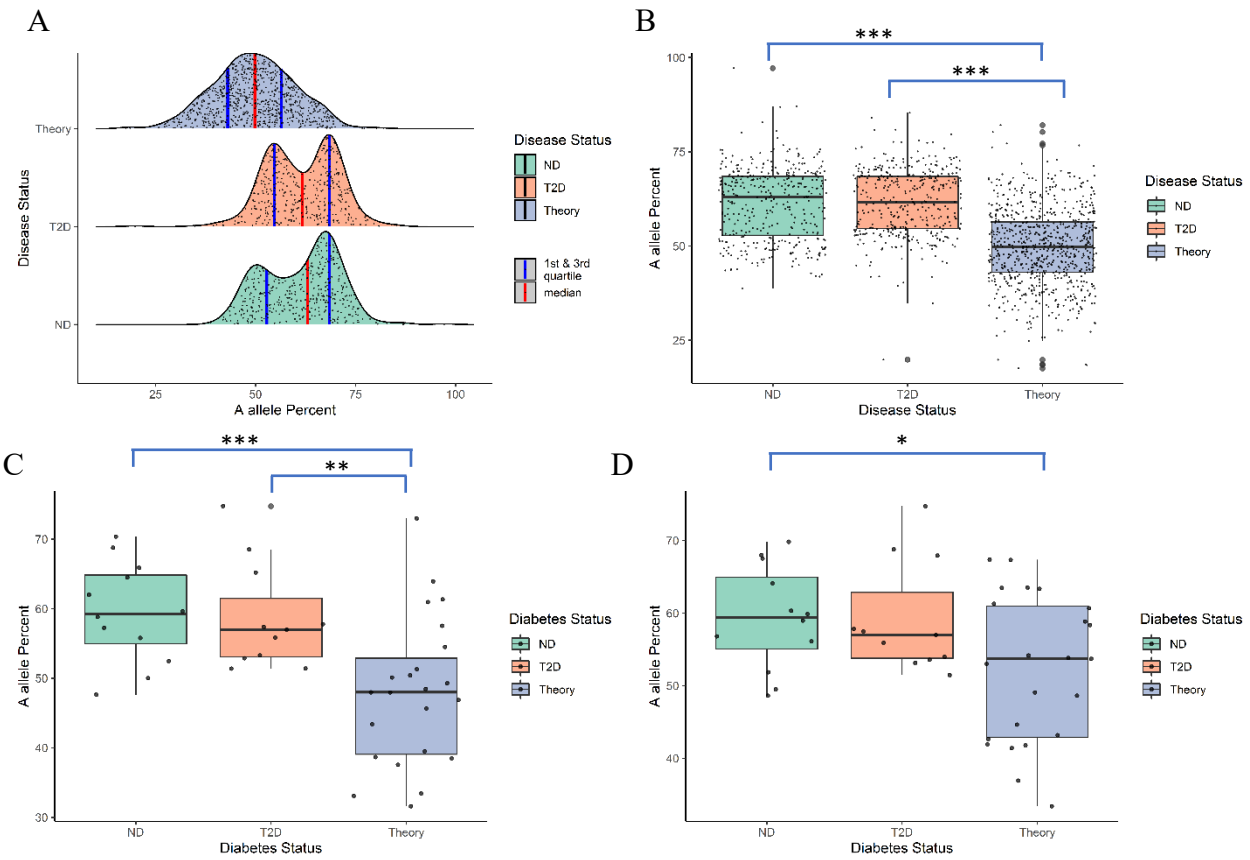


Figure 17: rs3842753 reference allele percent separated by donor and disease status for heterozygous donors. Cell specific (A, B) and donor specific mean and median (C, D) of rs3842753 reference allele percent distribution and box plot separated by disease status against theoretical normal distribution with mean at 50. Red lines represent median, blue lines represent 1st and 3rd quartiles. In A and B, each point represents one cell. In C and D, each point represents one donor. $P < 0.05$ = *. $P < 0.01$ = **. $P < 0.001$ = ***. Graphs are coloured by disease status. ND = non-diabetic. Theory = theoretical distribution. For cell specific analyses, Fligner-Killeen and Wilcox Rank Sum tests between either disease status and theoretical were used as non-parametric test for differences in variance and mean respectively. For donor specific analyses, Bartlett and Student T tests between either disease status and theoretical were used as parametric test for differences in variance and mean respectively.

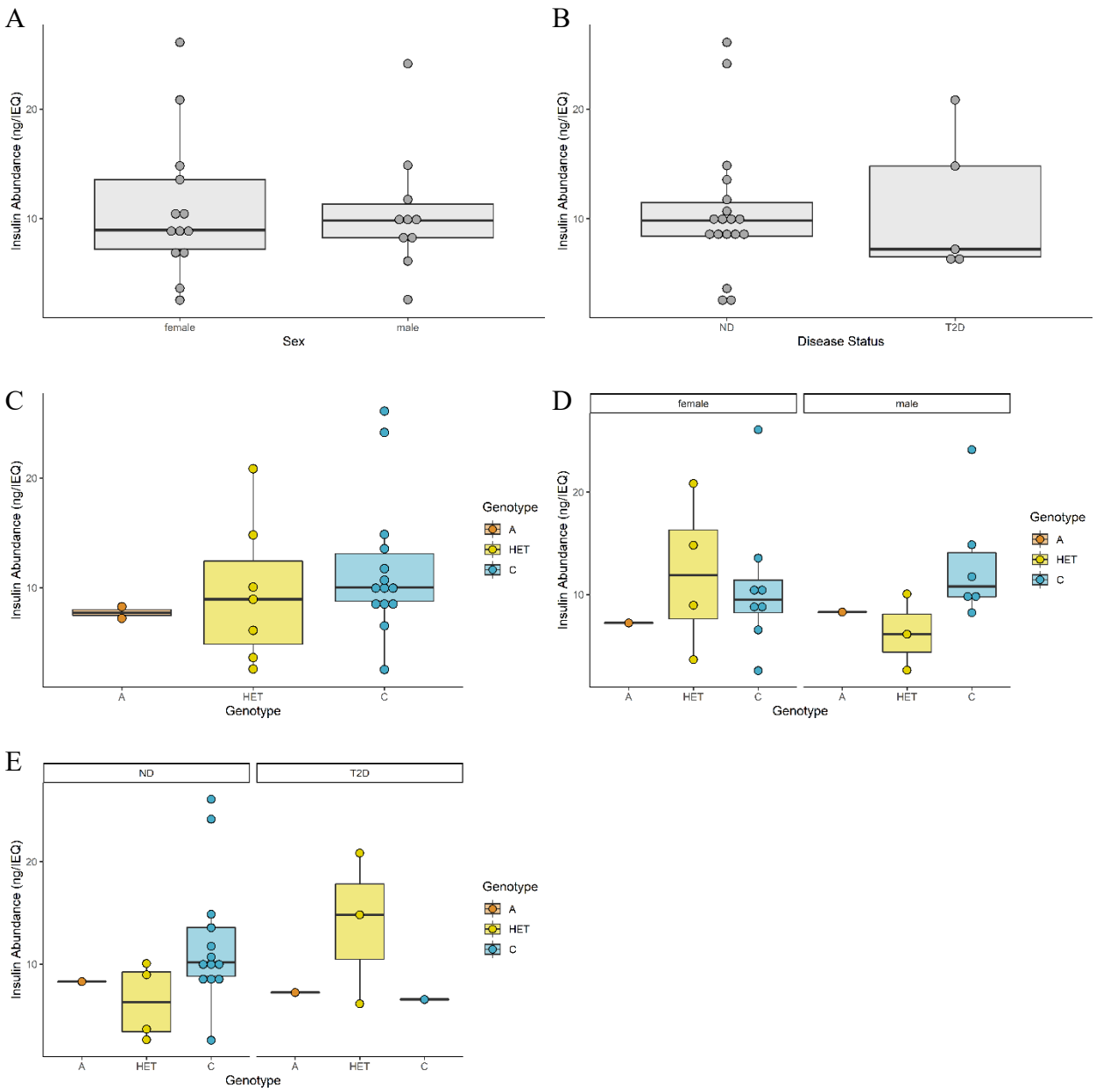


Figure 18: Insulin protein abundance of donors in Camunas et al. (2020) dataset. Insulin protein abundance for each donor separated by sex (A) and disease status (B) and genotype (C). Insulin protein abundance for each genotype separated by sex (D) and disease status (E). Each dot represents one donor. Graphs coloured by genotype. ND = non-diabetic. A represent A/A genotype, HET represent A/C genotype, C represent A/C genotype at rs384275.

Chapter 4: Discussion

4.1: rs3842753 Genotype and Pancreatic INS Expression

Several genetic variants surrounding *INS* have been implicated in T1D development (142; 143). Among those, the linked genetic variants: Class I *INS*-VNTR, rs689 A allele and rs3842753 C allele are the main variations associated with increased T1D susceptibility. Previously, there were only three studies examining the effect of genotype at the *INS* loci and pancreatic *INS* expression despite the numerous studies finding an association between genotype at the *INS* loci and T1D susceptibility (143; 166; 167). These studies included very few numbers of donors (n = 1, 3, and 10) and used whole pancreas tissues. Thus, the recent expansion of published pancreatic scRNAseq data has allowed previous questions to be re-examined with greater number of donors and with greater cell type specificity (185).

This thesis focused on the association between the genotype at SNP rs3842753 and pancreatic *INS* expression using published scRNAseq datasets. When examining normalized *INS* expression per pancreatic β-cell, we found an association between C allele and increased *INS* expression. Our study found that the heterozygous rs3842753 A/C genotype was associated with the highest *INS* expression followed by C/C genotype, then by the A/A genotype in total cells, female specific cell, and either diabetes status. In male cells, C/C genotype had the highest *INS* expression followed by A/A genotype, lastly by the A/C genotype. This somewhat consistent with previous literature which found the C allele to be associated with increased *INS* expression in heterozygous donors using whole pancreas tissues(143; 166; 167). Sex of the donors for previous studies were not reported. Thus, we could not be certain that our results' concordance to previous literature are not due to differences in distribution between sex.

However, the associations found in our study did not maintain statistical significance when assessed at the donor level, rather than cell level. Our finding of no association between rs3842753 genotype and pancreatic *INS* expression at the donor level is not consistent with previous smaller studies in adult and fetal whole pancreas tissues. Our study included 71 unique adult donors (48 non-diabetic and 23 T2D) cumulating in 2175 pancreatic β-cells. This is a substantial improvement from previous studies with a maximum of ten fetal pancreas tissues (167). With small sample numbers, variation between individuals could play a greater role than

genotypic effects. Perhaps at the donor level, our current study is likely to be still underpowered compared to the cell level *INS* expression analysis.

Also, we found an association between rs3842753 genotype and increased variance of *INS* expression. We found the heterozygous rs3842753 A/C genotype was associated with the highest variance in *INS* expression followed by A/A genotype, then by the C/C genotype in total cells, male specific cell, and either diabetes status. In female specific cells, the C/C genotype had the highest variance in *INS* expression followed by the A/C genotype. It is interesting that the A/C genotype mainly had the highest variance. This could be explained due to the cells' expression profile being influenced by both the rs3842753 A allele and C allele leading to increased variance. This difference in variance supports the established model that pancreatic β-cells are heterogeneous and perhaps the rs3842753 genotype could play a role in heterogeneity of *INS* expression.

4.2: rs3842753 Genotype and Allele Specific Expression

Previous studies that investigated the effect of rs3842753 on *INS* mRNA levels in whole pancreas tissue did so using heterozygous donors. Here we adopted the same approach in single cells. Allele specific expression for heterozygous rs3842753 alleles (A or C) can be quantified to remove inter-donor variance. Our finding of rs3842753 A allele (linked with Class III *INS*-VNTR) biased expression in rs3842753 A/C heterozygous pancreatic β-cells is not consistent with previous literature. In the three key studies examining *INS*-VNTR genotype and pancreatic *INS* expression, all individuals were heterozygous with one study involving three adults donors and two studies involving one and ten fetal donors (143; 166; 167). All studies found a significant increase in pancreatic *INS* expression from the Class I *INS*-VNTR haplotype compared with the Class III *INS*-VNTR haplotype, the opposite association as found in our study. The reasons for this are unclear, but deserve further investigation. For example, our study used data that had UMI counts rather than read counts which could be a source of error in allele specific expression quantification. Also, the region surrounding rs3842753 has a high GC content which could contribute to sequencing difficulty.

4.3: rs3842753 Genotype and Pancreatic INS Protein Abundance

We were also interested in whether changes in pancreatic *INS* expression could be translated into changes in *INS* protein abundance in whole islets. In our pilot study, we found no significant association between rs3842753 genotype and insulin protein abundance in whole pancreatic islets. Our *in vitro* finding of no association between rs3842753 genotype and insulin protein abundance is consistent with previous literature examining basal insulin secretion *in vivo* (156; 157; 160; 163; 164). This lack of association between rs3842753 genotype and insulin protein abundance is consistent with our finding that there is also no association with pancreatic *INS* expression level at the donor level. Limitation with our protein abundance study include the severe lack of power and type of INS protein quantification. We could only obtain INS protein abundance for a subset of donors (n = 23) with only 2 donors for the A/A genotype split across sex and diabetes status (1 male non-diabetic donor with rs3842753 A/A genotype, and 1 female T2D donor with rs3842753 A/A genotype). As well, we could only obtain the INS protein abundance for whole pancreatic islets as opposed to single pancreatic β-cells thus losing cell type specificity. This pilot study lacks statistical power and further investigation should be conducted possibly with single cell protein abundance.

4.4: An Alternative Model for Role of rs3842753 in T1D Susceptibility

We were unable to conclusively demonstrate a consistent relationship between rs3842753 genotype and pancreatic *INS* expression, allele specific expression and whole islet insulin protein abundance in our study of adult non-diabetic and T2D donors. Therefore, we are unable to lend strong support for the hypothesis that the Class III *INS*-VNTR allele is associated with sustained overproduction of insulin leading to ER stress thus increased T1D susceptibly. An alternative hypothesis for the mechanism of Class III *INS*-VNTR associated T1D susceptibly could be alternative protein products leading to autoimmune attack on pancreatic β-cells thus inducing T1D. Kracht et al. (2017) reported that neoautoantigen products due to alternative open reading frame (INS-DRiP - 2+ open reading frame shift at start codon AUG_{341}) could induce T-cell proliferation and targeted killing of pancreatic β-cell upon ER stress *in vitro* (131). Due to this frame shift, SNP rs3842753 becomes protein coding instead of being in the *INS* 3'UTR. In this situation, SNP rs3842753 would encode a missense mutation of histidine to proline (A allele to C

allele) thus creating a genotype specific neoautoantigen. However, Kracht et al. (2017) showed T cell proliferative response against both the histidine and proline INS-DRiP SNP variants individuals with T1D suggesting the SNP might not play a large role in immunogenicity (131). The specific role of rs384275, rs689, and *INS*-VNTR genotype in T1D development remains unclear.

4.5: Sex and Diabetes Status Affect Pancreatic *INS* Expression Variation But Not *INS* Expression Level

Sex and diabetes status play an important role in insulin resistance and secretion. Repeated studies have shown males to be more insulin resistant and secrete less insulin after glucose stimulation than females though basal insulin levels have been less extensively studied (207; 208). As well, during pregnancy, females undergo a sudden increase in inflammation, insulin resistance, and insulin demand which can result in developing GDM (209). We found that females had more pancreatic *INS* expression than males. Interestingly, we also found an increase in *INS* expression variance in females and T2D compared to males and non-diabetic donors, respectively. This is consistent with the hypothesis that female pancreatic β-cells need have more plasticity in gene expression to accommodate the various physiological demands and T2D β-cells could have more variance due to the range in disease state and its effects on cell health and function.

As for diabetes and insulin demand, the hallmark of T2D is increased insulin resistance and decreased insulin secretion (210). Some previous studies have suggested that *INS* mRNA levels are lower in whole islets and single cells from donors with T2D compared to non-diabetic (87; 174; 177). We found a male specific lowering of *INS* expression in T2D (Figure 10). This result indicates that insulin expression is different between diabetes status perhaps in a sex dependent manner.

4.6: Limitations

Our study has multiple limitations which mean that the results should be extrapolated with caution. One of the major limitations was that our study was not conducted in β- cells from

islets of children who either had or were at risk for T1D. Most cases of T1D onset between 10 to 14 years old during puberty; since all donors included this study are adults, this study could be studying the wrong age demographic for genotype associated pancreatic *INS* expression(12). Perhaps, the changes in pancreatic *INS* expression (and associations with the rs3842753 genotype) occurs in children when the genetics risk factors for developing T1D play a more prominent role. Over time, environmental risk factors could become more prominent and overshadow the genetics risk factors. Some evidence for this is comes from the Hansen et al. (2004) study focusing on glucose stimulated insulin release(163). They found an association between decreased insulin concentrations after glucose stimulation in young adults with Class III/III *INS*-VNTR genotype. However, this association could not be found in middle aged adults. As well, two of the early studies that found an association between Class I *INS*-VNTR allele and increased pancreatic *INS* expression studied fetal pancreas tissues. Our study could be focusing on the wrong age demographic thus unable to find an association between rs3842753 genotype and pancreatic *INS* expression. This limitation could only be solved through increased sample size in younger age groups. As well, there are many factors that affect insulin expression which were not accounted for in this study including by not limited to: age, BMI, dietary differences, and ethnicity. Although our study used all available datasets and is a major improvement in sample size compared to previous studies, many of the metadata mentioned above were not available and could not be accounted for. For this study, our datasets were all from human cadaver donors which has the benefit compared to mice samples as mice do not have the *INS*-VNTR variant.

Another limitation to our study, and the genetics of T1D in general, is the excessive focus on the Caucasian population. The linkage disequilibrium for the three genetic variants *INS*-VNTR, rs689, and rs3842753 has only been extensively studied in Caucasian populations with the r^2 unknown for other ethnicities(142). Our dataset only included ethnicity metadata for 31 out of 71 donors (29 were of Caucasian descent). Since we are unable to determine the ethnicity of 38 donors, our assumption that genotyping for SNP rs3842753 as a proxy for *INS*-VNTR could be incorrect. This limitation could be overcome through only examining donors of Caucasian descent though this would greatly decrease sample size thus lose statistical power.

There are also many technical limitations to our study and its underlying data, especially in the allele-specific *INS* expression analysis. For example, previous literature found in heterozygous whole pancreas tissues, the Class I *INS*-VNTR (linked to rs3842753 C allele) is associated with higher mRNA expression compared to the Class III *INS*-VNTR allele (linked to rs3842753 A allele). Our study found the opposite result in heterozygous single pancreatic β-cells. The difference between previous literature and our findings for the allele specific expression could be due to the technical limitations of the datasets used in this study. None of our datasets are UMI tagged. This means that PCR amplification plays a major role in final transcript counts. By studying the allele specific expression from single cell datasets, there is great reliance on accurate read counts from the two alleles. However, since each cDNA fragment is not tagged with UMI, there is an assumption that PCR amplification efficiency from both alleles is the same. This assumption is untested with no UMI tagging and could cause an inflated proportion of rs3842753 A allele read counts in individuals with heterozygous rs3842753 genotype. Some sequencing technical challenges are present given the high GC content (65%) in the 3'UTR containing rs3842753. The 20bp region immediately upstream and downstream of rs3842753 only contains 3 A nucleotides. Thus, a A>C mutation for rs3842753 would increase the difficulty of sequencing in that region. This could result in the preferential sequencing of the reference A allele compared to the alternate C allele. Further investigation is required to rule out artifacts due to sequencing and to identify a biological mechanism for these results.

4.7: Future Directions

This integrated dataset includes the most recent pancreatic islet scRNAseq data and will be available to the public. This tool will aid in any transcriptome investigations particularly between adult non-diabetic donors versus T2D. As well, sex and age are included in the metadata allowing for investigations focusing on sex specific transcriptome differences and age mediated transcriptome change. To date, no studies have specifically focused on sex specific differences in pancreatic islet single cells. Though Enge et al. (2017) have examined the transcriptional signatures of aging in human pancreas data with eight non-diabetic donors, it would be interesting to study the transcriptional signatures of aging between non-diabetic versus T2D pancreatic islet cells (173).

For the INS protein abundance analysis, more donors would need to be included to increase statistical power. Recently, studies have combined single-cell protein and RNA profiling to obtain mRNA expression and protein abundance on a per cell basis(211). This will allow for cell specific *INS* expression and protein abundance measurement. As well, direct correlation could be made between mRNA expression level and protein abundance within one cell. This will verify if any genotype associated changes in *INS* expression is translated into changes in protein production.

It would be useful to verify increased genetic variant associated thymic *INS* expression in single cell datasets using the same protocol as in this study as previous studies have also been small using whole thymic tissue (166; 167). Class III *INS*-VNTR is associated with increased thymic *INS* expression which has been proposed to increase immune tolerance and reduce immune activation targeting the pancreatic β-cells(166). Durinovic-Belló et al. (2005) have found a significant decrease in proinsulin reactive T cells in individuals with the Class III *INS*-VNTR allele (homozygous or heterozygous) compared to individuals with the Class I/I *INS*-VNTR genotype (169). This finding is consistent with the stated hypothesis. However, the Class III *INS*-VNTR associated increase in thymic *INS* expression should be verified with increased sample size and lower age range.

Taken together, our study has found that some, but not all, of the available evidence supports the association between Class I *INS*-VNTR, A allele rs689, and C allele rs3842753 and increased pancreatic *INS* expression and/or protein production in pancreatic β-cells. We found evidence for the association between the rs3842753 C allele and increased *INS* expression at the pancreatic β-cell level using 2315 cells from 71 donors. This increase in *INS* expression was highest in heterozygous rs3842753 A/C donors followed by rs3842753 C/C donors pointing to an overdominance of expression as opposed to simply an additive model. This association was not significant at the donor level though the same trends were observed. We also found significant increase in the variation of *INS* expression in rs3842753 A/C donors, followed by C/C donors, then A/A donors, meaning that some cells within the C allele containing population would have unusually high *INS* mRNA levels. This association was not significant at the donor level though similar trends persisted. This is somewhat consistent with previous literature which found a C allele associated increase in *INS* expression in heterozygous whole pancreas donors.

However, in our allele specific expression in 23 heterozygous rs3842753 A/C donors, we found preferential rs3842753 A allele expression compared to the C allele which is the opposite association direction compared to our *INS* expression findings and to previous literature. This deserves further study to rule out technical sequencing artifacts and to identify a biological mechanism. Taken together, our findings suggest that in single β-cells, rs3842753 may affect *INS* variance and expression, although the mechanisms by which this could dictate T1D risk remain unclear.

References

1. American Diabetes A: Diagnosis and classification of diabetes mellitus. Diabetes Care 2014;37 Suppl 1:S81-90

2. Alberti KG, Zimmet PZ: Definition, diagnosis and classification of diabetes mellitus and its complications. Part 1: diagnosis and classification of diabetes mellitus provisional report of a WHO consultation. Diabet Med 1998;15:539-553

3. Organization WH: Global report on diabetes. Geneva, Switzerland, 2016

4. Bommer C, Heesemann E, Sagalova V, Manne-Goehler J, Atun R, Barnighausen T, Vollmer S: The global economic burden of diabetes in adults aged 20-79 years: a cost-of-illness study. Lancet Diabetes Endocrinol 2017;5:423-430

5. Bommer C, Sagalova V, Heesemann E, Manne-Goehler J, Atun R, Barnighausen T, Davies J, Vollmer S: Global Economic Burden of Diabetes in Adults: Projections From 2015 to 2030. Diabetes Care 2018;41:963-970

6. Ogurtsova K, da Rocha Fernandes JD, Huang Y, Linnenkamp U, Guariguata L, Cho NH, Cavan D, Shaw JE, Makaroff LE: IDF Diabetes Atlas: Global estimates for the prevalence of diabetes for 2015 and 2040. Diabetes Res Clin Pract 2017;128:40-50

7. Daneman D: Type 1 diabetes. Lancet 2006;367:847-858

8. Liese AD, D'Agostino RB, Jr., Hamman RF, Kilgo PD, Lawrence JM, Liu LL, Loots B, Linder B, Marcovina S, Rodriguez B, Standiford D, Williams DE: The burden of diabetes mellitus among US youth: prevalence estimates from the SEARCH for Diabetes in Youth Study. Pediatrics 2006;118:1510-1518

9. Vandewalle CL, Coeckelberghs MI, De Leeuw IH, Du Caju MV, Schuit FC, Pipeleers DG, Gorus FK: Epidemiology, clinical aspects, and biology of IDDM patients under age 40 years. Comparison of data from Antwerp with complete ascertainment with data from Belgium with 40% ascertainment. The Belgian Diabetes Registry. Diabetes Care 1997;20:1556-1561

10. Thunander M, Petersson C, Jonzon K, Fornander J, Ossiansson B, Torn C, Edvardsson S, Landin-Olsson M: Incidence of type 1 and type 2 diabetes in adults and children in Kronoberg, Sweden. Diabetes Res Clin Pract 2008;82:247-255

11. Maahs DM, West NA, Lawrence JM, Mayer-Davis EJ: Epidemiology of type 1 diabetes. Endocrinol Metab Clin North Am 2010;39:481-497

12. Group DP: Incidence and trends of childhood Type 1 diabetes worldwide 1990-1999. Diabet Med 2006;23:857-866

13. Simmons KM, Michels AW: Type 1 diabetes: A predictable disease. World J Diabetes 2015;6:380-390

14. Knip M, Veijola R, Virtanen SM, Hyoty H, Vaarala O, Akerblom HK: Environmental triggers and determinants of type 1 diabetes. Diabetes 2005;54 Suppl 2:S125-136

15. Barrett JC, Clayton DG, Concannon P, Akolkar B, Cooper JD, Erlich HA, Julier C, Morahan G, Nerup J, Nierras C, Plagnol V, Pociot F, Schuilenburg H, Smyth DJ, Stevens H, Todd JA, Walker NM, Rich SS, Type 1 Diabetes Genetics C: Genome-wide association study and meta-analysis find that over 40 loci affect risk of type 1 diabetes. Nat Genet 2009;41:703-707

16. Redondo MJ, Jeffrey J, Fain PR, Eisenbarth GS, Orban T: Concordance for islet autoimmunity among monozygotic twins. N Engl J Med 2008;359:2849-2850

17. Ram R, Mehta M, Nguyen QT, Larma I, Boehm BO, Pociot F, Concannon P, Morahan G: Systematic Evaluation of Genes and Genetic Variants Associated with Type 1 Diabetes Susceptibility. J Immunol 2016;196:3043-3053

18. Onengut-Gumuscu S, Chen WM, Burren O, Cooper NJ, Quinlan AR, Mychaleckyj JC, Farber E, Bonnie JK, Szpak M, Schofield E, Achuthan P, Guo H, Fortune MD, Stevens H, Walker NM, Ward LD, Kundaje A, Kellis M, Daly MJ, Barrett JC, Cooper JD, Deloukas P, Type 1 Diabetes Genetics C, Todd JA, Wallace C, Concannon P, Rich SS: Fine mapping of type 1 diabetes susceptibility loci and evidence for colocalization of causal variants with lymphoid gene enhancers. Nat Genet 2015;47:381-386

19. Erlich H, Valdes AM, Noble J, Carlson JA, Varney M, Concannon P, Mychaleckyj JC, Todd JA, Bonella P, Fear AL, Lavant E, Louey A, Moonsamy P, Type 1 Diabetes Genetics C: HLA DR-DQ haplotypes and genotypes and type 1 diabetes risk: analysis of the type 1 diabetes genetics consortium families. Diabetes 2008;57:1084-1092

20. Bell GI, Horita S, Karam JH: A polymorphic locus near the human insulin gene is associated with insulin-dependent diabetes mellitus. Diabetes 1984;33:176-183

21. Bradfield JP, Qu HQ, Wang K, Zhang H, Sleiman PM, Kim CE, Mentch FD, Qiu H, Glessner JT, Thomas KA, Frackelton EC, Chiavacci RM, Imielinski M, Monos DS, Pandey R, Bakay M, Grant SF, Polychronakos C, Hakonarson H: A genome-wide meta-analysis of six type 1 diabetes cohorts identifies multiple associated loci. PLoS Genet 2011;7:e1002293

22. Rewers M, Ludvigsson J: Environmental risk factors for type 1 diabetes. Lancet 2016;387:2340-2348

23. Kimpimaki T, Kupila A, Hamalainen AM, Kukko M, Kulmala P, Savola K, Simell T, Keskinen P, Ilonen J, Simell O, Knip M: The first signs of beta-cell autoimmunity appear in infancy in genetically susceptible children from the general population: the Finnish Type 1 Diabetes Prediction and Prevention Study. J Clin Endocrinol Metab 2001;86:4782-4788

24. Zipitis CS, Akobeng AK: Vitamin D supplementation in early childhood and risk of type 1 diabetes: a systematic review and meta-analysis. Arch Dis Child 2008;93:512-517

25. Pathak V, Pathak NM, O'Neill CL, Guduric-Fuchs J, Medina RJ: Therapies for Type 1 Diabetes: Current Scenario and Future Perspectives. Clin Med Insights Endocrinol Diabetes 2019;12:1179551419844521

26. Robertson RP: Antagonist: diabetes and insulin resistance--philosophy, science, and the multiplier hypothesis. J Lab Clin Med 1995;125:560-564; discussion 565

27. Olokoba AB, Obateru OA, Olokoba LB: Type 2 diabetes mellitus: a review of current trends. Oman Med J 2012;27:269-273

28. Ferrannini E, Natali A, Bell P, Cavallo-Perin P, Lalic N, Mingrone G: Insulin resistance and hypersecretion in obesity. European Group for the Study of Insulin Resistance (EGIR). J Clin Invest 1997;100:1166-1173

29. Thomas DD, Corkey BE, Istfan NW, Apovian CM: Hyperinsulinemia: An Early Indicator of Metabolic Dysfunction. J Endocr Soc 2019;3:1727-1747

30. Mehran AE, Templeman NM, Brigidi GS, Lim GE, Chu KY, Hu X, Botezelli JD, Asadi A, Hoffman BG, Kieffer TJ, Bamji SX, Clee SM, Johnson JD: Hyperinsulinemia drives diet-induced obesity independently of brain insulin production. Cell Metab 2012;16:723-737

31. Morita I, Tanimoto K, Akiyama N, Naya N, Fujieda K, Iwasaki T, Yukioka H: Chronic hyperinsulinemia contributes to insulin resistance under dietary restriction in association with altered lipid metabolism in Zucker diabetic fatty rats. Am J Physiol Endocrinol Metab 2017;312:E264-E272

32. Templeman NM, Skovso S, Page MM, Lim GE, Johnson JD: A causal role for hyperinsulinemia in obesity. J Endocrinol 2017;232:R173-R183

33. Wild S, Roglic G, Green A, Sicree R, King H: Global prevalence of diabetes: estimates for the year 2000 and projections for 2030. Diabetes Care 2004;27:1047-1053

34. Fagot-Campagna A: Emergence of type 2 diabetes mellitus in children: epidemiological evidence. J Pediatr Endocrinol Metab 2000;13 Suppl 6:1395-1402

35. Hillier TA, Pedula KL: Complications in young adults with early-onset type 2 diabetes: losing the relative protection of youth. Diabetes Care 2003;26:2999-3005

36. Sattar N, Rawshani A, Franzen S, Rawshani A, Svensson AM, Rosengren A, McGuire DK, Eliasson B, Gudbjornsdottir S: Age at Diagnosis of Type 2 Diabetes Mellitus and Associations With Cardiovascular and Mortality Risks. Circulation 2019;139:2228-2237

37. Ang GY: Age of onset of diabetes and all-cause mortality. World J Diabetes 2020;11:95-99

38. Zimmet P, Alberti KG, Shaw J: Global and societal implications of the diabetes epidemic. Nature 2001;414:782-787

39. Al-Goblan AS, Al-Alfi MA, Khan MZ: Mechanism linking diabetes mellitus and obesity. Diabetes Metab Syndr Obes 2014;7:587-591

40. Manson JE, Ajani UA, Liu S, Nathan DM, Hennekens CH: A prospective study of cigarette smoking and the incidence of diabetes mellitus among US male physicians. Am J Med 2000;109:538-542

41. Wu Y, Ding Y, Tanaka Y, Zhang W: Risk factors contributing to type 2 diabetes and recent advances in the treatment and prevention. Int J Med Sci 2014;11:1185-1200

42. Mahajan A, Taliun D, Thurner M, Robertson NR, Torres JM, Rayner NW, Payne AJ, Steinthorsdottir V, Scott RA, Grarup N, Cook JP, Schmidt EM, Wuttke M, Sarnowski C, Magi R, Nano J, Gieger C, Trompet S, Lecoeur C, Preuss MH, Prins BP, Guo X, Bielak LF, Below JE, Bowden DW, Chambers JC, Kim YJ, Ng MCY, Petty LE, Sim X, Zhang W, Bennett AJ, Bork-Jensen J, Brummett CM, Canouil M, Ec Kardt KU, Fischer K, Kardia SLR, Kronenberg F, Lall K, Liu CT, Locke AE, Luan J, Ntalla I, Nylander V, Schonherr S, Schurmann C, Yengo L, Bottinger EP, Brandslund I, Christensen C, Dedoussis G, Florez JC, Ford I, Franco OH, Frayling TM, Giedraitis V, Hackinger S, Hattersley AT, Herder C, Ikram MA, Ingelsson M, Jorgensen ME, Jorgensen T, Kriebel J, Kuusisto J, Ligthart S, Lindgren CM, Linneberg A, Lyssenko V, Mamakou V, Meitinger T, Mohlke KL, Morris AD, Nadkarni G, Pankow JS, Peters A, Sattar N, Stancakova A, Strauch K, Taylor KD, Thorand B, Thorleifsson G, Thorsteinsdottir U, Tuomilehto J, Witte DR, Dupuis J, Peyser PA, Zeggini E, Loos RJF, Froguel P, Ingelsson E, Lind L, Groop L, Laakso M, Collins FS, Jukema JW, Palmer CNA, Grallert H, Metspalu A, Dehghan A, Kottgen A, Abecasis GR, Meigs JB, Rotter JI, Marchini J, Pedersen O, Hansen T, Langenberg C, Wareham NJ, Stefansson K, Gloyn AL, Morris AP, Boehnke M, McCarthy MI: Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. Nat Genet 2018;50:1505-1513

43. Lean MEJ, Leslie WS, Barnes AC, Brosnahan N, Thom G, McCombie L, Peters C, Zhyzhneuskaya S, Al-Mrabeh A, Hollingsworth KG, Rodrigues AM, Rehackova L, Adamson

AJ, Sniehotta FF, Mathers JC, Ross HM, McIlvenna Y, Welsh P, Kean S, Ford I, McConnachie A, Messow CM, Sattar N, Taylor R: Durability of a primary care-led weight-management intervention for remission of type 2 diabetes: 2-year results of the DiRECT open-label, cluster-randomised trial. Lancet Diabetes Endocrinol 2019;7:344-355

44. Blaslov K, Naranda FS, Kruljac I, Renar IP: Treatment approach to type 2 diabetes: Past, present and future. World J Diabetes 2018;9:209-219

45. Letourneau LR, Carmody D, Wroblewski K, Denson AM, Sanyoura M, Naylor RN, Philipson LH, Greeley SAW: Diabetes Presentation in Infancy: High Risk of Diabetic Ketoacidosis. Diabetes Care 2017;40:e147-e148

46. Kanakatti Shankar R, Pihoker C, Dolan LM, Standiford D, Badaru A, Dabelea D, Rodriguez B, Black MH, Imperatore G, Hattersley A, Ellard S, Gilliam LK, Group SfDiYS: Permanent neonatal diabetes mellitus: prevalence and genetic diagnosis in the SEARCH for Diabetes in Youth Study. Pediatr Diabetes 2013;14:174-180

47. Lemelman MB, Letourneau L, Greeley SAW: Neonatal Diabetes Mellitus: An Update on Diagnosis and Management. Clin Perinatol 2018;45:41-59

48. De Franco E, Flanagan SE, Houghton JA, Lango Allen H, Mackay DJ, Temple IK, Ellard S, Hattersley AT: The effect of early, comprehensive genomic testing on clinical care in neonatal diabetes: an international cohort study. Lancet 2015;386:957-963

49. Smith AJ, Taneja TK, Mankouri J, Sivaprasadarao A: Molecular cell biology of KATP channels: implications for neonatal diabetes. Expert Rev Mol Med 2007;9:1-17

50. Dean PM, Matthews EK: Glucose-induced electrical activity in pancreatic islet cells. J Physiol 1970;210:255-264

51. MacDonald PE, Joseph JW, Rorsman P: Glucose-sensing mechanisms in pancreatic beta-cells. Philos Trans R Soc Lond B Biol Sci 2005;360:2211-2225

52. Gloyn AL, Pearson ER, Antcliff JF, Proks P, Bruining GJ, Slingerland AS, Howard N, Srinivasan S, Silva JM, Molnes J, Edghill EL, Frayling TM, Temple IK, Mackay D, Shield JP,

Sumnik Z, van Rhijn A, Wales JK, Clark P, Gorman S, Aisenberg J, Ellard S, Njolstad PR, Ashcroft FM, Hattersley AT: Activating mutations in the gene encoding the ATP-sensitive potassium-channel subunit Kir6.2 and permanent neonatal diabetes. N Engl J Med 2004;350:1838-1849

53. Froguel P, Vaxillaire M, Sun F, Velho G, Zouali H, Butel MO, Lesage S, Vionnet N, Clement K, Fougerousse F, et al.: Close linkage of glucokinase locus on chromosome 7p to early-onset non-insulin-dependent diabetes mellitus. Nature 1992;356:162-164

54. Horikawa Y, Iwasaki N, Hara M, Furuta H, Hinokio Y, Cockburn BN, Lindner T, Yamagata K, Ogata M, Tomonaga O, Kuroki H, Kasahara T, Iwamoto Y, Bell GI: Mutation in hepatocyte nuclear factor-1 beta gene (TCF2) associated with MODY. Nat Genet 1997;17:384-385

55. Stoffers DA, Ferrer J, Clarke WL, Habener JF: Early-onset type-II diabetes mellitus (MODY4) linked to IPF1. Nat Genet 1997;17:138-139

56. Vionnet N, Stoffel M, Takeda J, Yasuda K, Bell GI, Zouali H, Lesage S, Velho G, Iris F, Passa P, et al.: Nonsense mutation in the glucokinase gene causes early-onset non-insulin-dependent diabetes mellitus. Nature 1992;356:721-722

57. Yamagata K, Furuta H, Oda N, Kaisaki PJ, Menzel S, Cox NJ, Fajans SS, Signorini S, Stoffel M, Bell GI: Mutations in the hepatocyte nuclear factor-4alpha gene in maturity-onset diabetes of the young (MODY1). Nature 1996;384:458-460

58. Hattersley AT, Greeley SAW, Polak M, Rubio-Cabezas O, Njolstad PR, Mlynarski W, Castano L, Carlsson A, Raile K, Chi DV, Ellard S, Craig ME: ISPAD Clinical Practice Consensus Guidelines 2018: The diagnosis and management of monogenic diabetes in children and adolescents. Pediatr Diabetes 2018;19 Suppl 27:47-63

59. Urakami T: Maturity-onset diabetes of the young (MODY): current perspectives on diagnosis and treatment. Diabetes Metab Syndr Obes 2019;12:1047-1056

60. Edghill EL, Flanagan SE, Patch AM, Boustred C, Parrish A, Shields B, Shepherd MH, Hussain K, Kapoor RR, Malecki M, MacDonald MJ, Stoy J, Steiner DF, Philipson LH, Bell GI, Neonatal Diabetes International Collaborative G, Hattersley AT, Ellard S: Insulin mutation

screening in 1,044 patients with diabetes: mutations in the INS gene are a common cause of neonatal diabetes but a rare cause of diabetes diagnosed in childhood or adulthood. Diabetes 2008;57:1034-1042

61. Molven A, Ringdal M, Nordbo AM, Raeder H, Stoy J, Lipkind GM, Steiner DF, Philipson LH, Bergmann I, Aarskog D, Undlien DE, Joner G, Sovik O, Norwegian Childhood Diabetes Study G, Bell GI, Njolstad PR: Mutations in the insulin gene can cause MODY and autoantibody-negative type 1 diabetes. Diabetes 2008;57:1131-1135

62. Diagnostic criteria and classification of hyperglycaemia first detected in pregnancy: a World Health Organization Guideline. Diabetes Res Clin Pract 2014;103:341-363

63. McIntyre HD, Catalano P, Zhang C, Desoye G, Mathiesen ER, Damm P: Gestational diabetes mellitus. Nat Rev Dis Primers 2019;5:47

64. Chu SY, Callaghan WM, Kim SY, Schmid CH, Lau J, England LJ, Dietz PM: Maternal obesity and risk of gestational diabetes mellitus. Diabetes Care 2007;30:2070-2076

65. Solomon CG, Willett WC, Carey VJ, Rich-Edwards J, Hunter DJ, Colditz GA, Stampfer MJ, Speizer FE, Spiegelman D, Manson JE: A prospective study of pregravid determinants of gestational diabetes mellitus. JAMA 1997;278:1078-1083

66. Hedderson M, Ehrlich S, Sridhar S, Darbinian J, Moore S, Ferrara A: Racial/ethnic disparities in the prevalence of gestational diabetes mellitus by BMI. Diabetes Care 2012;35:1492-1498

67. Anna V, van der Ploeg HP, Cheung NW, Huxley RR, Bauman AE: Sociodemographic correlates of the increasing trend in prevalence of gestational diabetes mellitus in a large population of women between 1995 and 2005. Diabetes Care 2008;31:2288-2293

68. Catalano PM, Tyzbir ED, Roman NM, Amini SB, Sims EA: Longitudinal changes in insulin release and insulin resistance in nonobese pregnant women. Am J Obstet Gynecol 1991;165:1667-1672

69. Catalano PM, Tyzbir ED, Wolfe RR, Roman NM, Amini SB, Sims EA: Longitudinal changes in basal hepatic glucose production and suppression during insulin infusion in normal pregnant women. Am J Obstet Gynecol 1992;167:913-919

70. Brown GR, Hem V, Katz KS, Ovetsky M, Wallin C, Ermolaeva O, Tolstoy I, Tatusova T, Pruitt KD, Maglott DR, Murphy TD: Gene: a gene-centered information resource at NCBI. Nucleic Acids Res 2015;43:D36-42

71. Soares MB, Schon E, Henderson A, Karathanasis SK, Cate R, Zeitlin S, Chirgwin J, Efstratiadis A: RNA-mediated gene duplication: the rat preproinsulin I gene is a functional retroposon. Mol Cell Biol 1985;5:2090-2103

72. Davies PO, Poirier C, Deltour L, Montagutelli X: Genetic reassignment of the insulin-1 (Ins1) gene to distal mouse chromosome 19. Genomics 1994;21:665-667

73. German M, Ashcroft S, Docherty K, Edlund H, Edlund T, Goodison S, Imura H, Kennedy G, Madsen O, Melloul D, et al.: The insulin gene promoter. A simplified nomenclature. Diabetes 1995;44:1002-1004

74. Melloul D, Marshak S, Cerasi E: Regulation of insulin gene transcription. Diabetologia 2002;45:309-326

75. Matsuoka TA, Artner I, Henderson E, Means A, Sander M, Stein R: The MafA transcription factor appears to be responsible for tissue-specific expression of insulin. Proc Natl Acad Sci U S A 2004;101:2930-2933

76. Naya FJ, Stellrecht CM, Tsai MJ: Tissue-specific regulation of the insulin gene by a novel basic helix-loop-helix transcription factor. Genes Dev 1995;9:1009-1019

77. Zhu Y, Liu Q, Zhou Z, Ikeda Y: PDX1, Neurogenin-3, and MAFA: critical transcription regulators for beta cell development and regeneration. Stem Cell Res Ther 2017;8:240

78. Sharma A, Moore M, Marcora E, Lee JE, Qiu Y, Samaras S, Stein R: The NeuroD1/BETA2 sequences essential for insulin gene transcription colocalize with those necessary for neurogenesis and p300/CREB binding protein binding. Mol Cell Biol 1999;19:704-713

79. Andrali SS, Qian Q, Ozcan S: Glucose mediates the translocation of NeuroD1 by O-linked glycosylation. J Biol Chem 2007;282:15589-15596

80. Fujimoto K, Polonsky KS: Pdx1 and other factors that regulate pancreatic beta-cell survival. Diabetes Obes Metab 2009;11 Suppl 4:30-37

81. Mosley AL, Corbett JA, Ozcan S: Glucose regulation of insulin gene expression requires the recruitment of p300 by the beta-cell-specific transcription factor Pdx-1. Mol Endocrinol 2004;18:2279-2290

82. Qiu Y, Sharma A, Stein R: p300 mediates transcriptional stimulation by the basic helix-loop-helix activators of the insulin gene. Mol Cell Biol 1998;18:2957-2964

83. Vanderford NL, Andrali SS, Ozcan S: Glucose induces MafA expression in pancreatic beta cell lines via the hexosamine biosynthetic pathway. J Biol Chem 2007;282:1577-1584

84. Guo S, Dai C, Guo M, Taylor B, Harmon JS, Sander M, Robertson RP, Powers AC, Stein R: Inactivation of specific beta cell transcription factors in type 2 diabetes. J Clin Invest 2013;123:3305-3316

85. Pociot F: Type 1 diabetes genome-wide association studies: not to be lost in translation. Clin Transl Immunology 2017;6:e162

86. Kuroda A, Rauch TA, Todorov I, Ku HT, Al-Abdullah IH, Kandeel F, Mullen Y, Pfeifer GP, Ferreri K: Insulin gene expression is regulated by DNA methylation. PLoS One 2009;4:e6953

87. Yang BT, Dayeh TA, Kirkpatrick CL, Taneera J, Kumar R, Groop L, Wollheim CB, Nitert MD, Ling C: Insulin promoter DNA methylation correlates negatively with insulin gene expression and positively with HbA(1c) levels in human pancreatic islets. Diabetologia 2011;54:360-367

88. Welsh M, Nielsen DA, MacKrell AJ, Steiner DF: Control of insulin gene expression in pancreatic beta-cells and in an insulin-producing cell line, RIN-5F cells. II. Regulation of insulin mRNA stability. J Biol Chem 1985;260:13590-13594

89. Tillmar L, Carlsson C, Welsh N: Control of insulin mRNA stability in rat pancreatic islets. Regulatory role of a 3'-untranslated region pyrimidine-rich sequence. J Biol Chem 2002;277:1099-1106

90. Fred RG, Mehrabi S, Adams CM, Welsh N: PTB and TIAR binding to insulin mRNA 3'- and 5'UTRs; implications for insulin biosynthesis and messenger stability. Heliyon 2016;2:e00159

91. Brissova M, Fowler MJ, Nicholson WE, Chu A, Hirshberg B, Harlan DM, Powers AC: Assessment of human pancreatic islet architecture and composition by laser scanning confocal microscopy. J Histochem Cytochem 2005;53:1087-1097

92. Egea PF, Stroud RM, Walter P: Targeting proteins to membranes: structure of the signal recognition particle. Curr Opin Struct Biol 2005;15:213-220

93. Patzelt C, Labrecque AD, Duguid JR, Carroll RJ, Keim PS, Heinrikson RL, Steiner DF: Detection and kinetic behavior of preproinsulin in pancreatic islets. Proc Natl Acad Sci U S A 1978;75:1260-1264

94. Liu M, Lara-Lemus R, Shan SO, Wright J, Haataja L, Barbetti F, Guo H, Larkin D, Arvan P: Impaired cleavage of preproinsulin signal peptide linked to autosomal-dominant diabetes. Diabetes 2012;61:828-837

95. Fu Z, Gilbert ER, Liu D: Regulation of insulin synthesis and secretion and pancreatic Beta-cell dysfunction in diabetes. Curr Diabetes Rev 2013;9:25-53

96. Abel JJ: Crystalline Insulin. Proc Natl Acad Sci U S A 1926;12:132-136

97. Steiner DF, Rouille Y, Gong Q, Martin S, Carroll R, Chan SJ: The role of prohormone convertases in insulin biosynthesis: evidence for inherited defects in their action in man and experimental animals. Diabetes Metab 1996;22:94-104

98. Hou JC, Min L, Pessin JE: Insulin granule biogenesis, trafficking and exocytosis. Vitam Horm 2009;80:473-506

99. Hay CW, Docherty K: Comparative analysis of insulin gene promoters: implications for diabetes research. Diabetes 2006;55:3201-3213

100. McCulloch LJ, van de Bunt M, Braun M, Frayn KN, Clark A, Gloyn AL: GLUT2 (SLC2A2) is not the principal glucose transporter in human pancreatic beta cells: implications for understanding genetic association signals at this locus. Mol Genet Metab 2011;104:648-653

101. Tokarz VL, MacDonald PE, Klip A: The cell biology of systemic insulin function. J Cell Biol 2018;217:2273-2289

102. Berg J, Tymoczko J, Stryer L: Biochemistry. 5th Edition. In *Chapter 17, The Citric Acid Cycle* New York, W H Freeman and Co, 2002

103. Sehlin J, Taljedal IB: Glucose-induced decrease in Rb+ permeability in pancreatic beta cells. Nature 1975;253:635-636

104. Rorsman P, Renstrom E: Insulin granule dynamics in pancreatic beta cells. Diabetologia 2003;46:1029-1045

105. Gandasi NR, Yin P, Riz M, Chibalina MV, Cortese G, Lund PE, Matveev V, Rorsman P, Sherman A, Pedersen MG, Barg S: Ca2+ channel clustering with insulin-containing granules is disturbed in type 2 diabetes. J Clin Invest 2017;127:2353-2364

106. White MF, Shoelson SE, Keutmann H, Kahn CR: A cascade of tyrosine autophosphorylation in the beta-subunit activates the phosphotransferase of the insulin receptor. J Biol Chem 1988;263:2969-2980

107. White MF: Regulating insulin signaling and beta-cell function through IRS proteins. Can J Physiol Pharmacol 2006;84:725-737

108. Sun XJ, Rothenberg P, Kahn CR, Backer JM, Araki E, Wilden PA, Cahill DA, Goldstein BJ, White MF: Structure of the insulin receptor substrate IRS-1 defines a unique signal transduction protein. Nature 1991;352:73-77

109. Ruderman NB, Kapeller R, White MF, Cantley LC: Activation of phosphatidylinositol 3-kinase by insulin. Proc Natl Acad Sci U S A 1990;87:1411-1415

110. Shi Y, Wang J, Chandarlapaty S, Cross J, Thompson C, Rosen N, Jiang X: PTEN is a protein tyrosine phosphatase for IRS1. Nat Struct Mol Biol 2014;21:522-527

111. Maehama T, Dixon JE: The tumor suppressor, PTEN/MMAC1, dephosphorylates the lipid second messenger, phosphatidylinositol 3,4,5-trisphosphate. J Biol Chem 1998;273:13375-13378

112. Alessi DR, Cohen P: Mechanism of activation and function of protein kinase B. Curr Opin Genet Dev 1998;8:55-62

113. Zheng X, Cartee GD: Insulin-induced Effects on the Subcellular Localization of AKT1, AKT2 and AS160 in Rat Skeletal Muscle. Sci Rep 2016;6:39230

114. Sano H, Kane S, Sano E, Miinea CP, Asara JM, Lane WS, Garner CW, Lienhard GE: Insulin-stimulated phosphorylation of a Rab GTPase-activating protein regulates GLUT4 translocation. J Biol Chem 2003;278:14599-14602

115. Chadt A, Scherneck S, Joost HG, Al-Hasani H: Molecular links between Obesity and Diabetes: "Diabesity". In *Endotext* Feingold KR, Anawalt B, Boyce A, Chrousos G, Dungan K, Grossman A, Hershman JM, Kaltsas G, Koch C, Kopp P, Korbonits M, McLachlan R, Morley JE, New M, Perreault L, Purnell J, Rebar R, Singer F, Trence DL, Vinik A, Wilson DP, Eds. South Dartmouth (MA), 2000

116. Rasouli N, Kern PA: Adipocytokines and the metabolic complications of obesity. J Clin Endocrinol Metab 2008;93:S64-73

117. Osborn O, Olefsky JM: The cellular and signaling networks linking the immune system and metabolism in disease. Nat Med 2012;18:363-374

118. Ozes ON, Akca H, Mayo LD, Gustin JA, Maehama T, Dixon JE, Donner DB: A phosphatidylinositol 3-kinase/Akt/mTOR pathway mediates and PTEN antagonizes tumor necrosis factor inhibition of insulin signaling through insulin receptor substrate-1. Proc Natl Acad Sci U S A 2001;98:4640-4645

119. Jager J, Gremeaux T, Cormont M, Le Marchand-Brustel Y, Tanti JF: Interleukin-1beta-induced insulin resistance in adipocytes through down-regulation of insulin receptor substrate-1 expression. Endocrinology 2007;148:241-251

120. Tan CY, Vidal-Puig A: Adipose tissue expandability: the metabolic problems of obesity may arise from the inability to become more obese. Biochem Soc Trans 2008;36:935-940

121. Teruel T, Hernandez R, Lorenzo M: Ceramide mediates insulin resistance by tumor necrosis factor-alpha in brown adipocytes by maintaining Akt in an inactive dephosphorylated state. Diabetes 2001;50:2563-2571

122. Fonseca SG, Gromada J, Urano F: Endoplasmic reticulum stress and pancreatic beta-cell death. Trends Endocrinol Metab 2011;22:266-274

123. Harding HP, Zeng H, Zhang Y, Jungries R, Chung P, Plesken H, Sabatini DD, Ron D: Diabetes mellitus and exocrine pancreatic dysfunction in perk-/- mice reveals a role for translational control in secretory cell survival. Mol Cell 2001;7:1153-1163

124. Scheuner D, Vander Mierde D, Song B, Flamez D, Creemers JW, Tsukamoto K, Ribick M, Schuit FC, Kaufman RJ: Control of mRNA translation preserves endoplasmic reticulum function in beta cells and maintains glucose homeostasis. Nat Med 2005;11:757-764

125. Szabat M, Page MM, Panzhinskiy E, Skovso S, Mojibian M, Fernandez-Tajes J, Bruin JE, Bround MJ, Lee JT, Xu EE, Taghizadeh F, O'Dwyer S, van de Bunt M, Moon KM, Sinha S, Han J, Fan Y, Lynn FC, Trucco M, Borchers CH, Foster LJ, Nislow C, Kieffer TJ, Johnson JD: Reduced Insulin Production Relieves Endoplasmic Reticulum Stress and Induces beta Cell Proliferation. Cell Metab 2016;23:179-193

126. Bravo R, Parra V, Gatica D, Rodriguez AE, Torrealba N, Paredes F, Wang ZV, Zorzano A, Hill JA, Jaimovich E, Quest AF, Lavandero S: Endoplasmic reticulum and the unfolded protein response: dynamics and metabolic integration. Int Rev Cell Mol Biol 2013;301:215-290

127. Eizirik DL, Colli ML, Ortis F: The role of inflammation in insulitis and beta-cell loss in type 1 diabetes. Nat Rev Endocrinol 2009;5:219-226

128. Gonzalez-Duque S, Azoury ME, Colli ML, Afonso G, Turatsinze JV, Nigi L, Lalanne AI, Sebastiani G, Carre A, Pinto S, Culina S, Corcos N, Bugliani M, Marchetti P, Armanet M, Diedisheim M, Kyewski B, Steinmetz LM, Buus S, You S, Dubois-Laforgue D, Larger E, Beressi JP, Bruno G, Dotta F, Scharfmann R, Eizirik DL, Verdier Y, Vinh J, Mallone R: Conventional and Neo-antigenic Peptides Presented by beta Cells Are Targeted by Circulating Naive CD8+ T Cells in Type 1 Diabetic and Healthy Donors. Cell Metab 2018;28:946-960 e946

129. Eizirik DL, Sammeth M, Bouckenooghe T, Bottu G, Sisino G, Igoillo-Esteve M, Ortis F, Santin I, Colli ML, Barthson J, Bouwens L, Hughes L, Gregory L, Lunter G, Marselli L, Marchetti P, McCarthy MI, Cnop M: The human pancreatic islet transcriptome: expression of candidate genes for type 1 diabetes and the impact of pro-inflammatory cytokines. PLoS Genet 2012;8:e1002552

130. Marre ML, James EA, Piganelli JD: beta cell ER stress and the implications for immunogenicity in type 1 diabetes. Front Cell Dev Biol 2015;3:67

131. Kracht MJ, van Lummel M, Nikolic T, Joosten AM, Laban S, van der Slik AR, van Veelen PA, Carlotti F, de Koning EJ, Hoeben RC, Zaldumbide A, Roep BO: Autoimmunity against a defective ribosomal insulin gene product in type 1 diabetes. Nat Med 2017;23:501-507

132. van Lummel M, Zaldumbide A, Roep BO: Changing faces, unmasking the beta-cell: post-translational modification of antigens in type 1 diabetes. Curr Opin Endocrinol Diabetes Obes 2013;20:299-306

133. Kanatsuna N, Taneera J, Vaziri-Sani F, Wierup N, Larsson HE, Delli A, Skarstrand H, Balhuizen A, Bennet H, Steiner DF, Torn C, Fex M, Lernmark A: Autoimmunity against INS-IGF2 protein expressed in human pancreatic islets. J Biol Chem 2013;288:29013-29023

134. Strollo R, Vinci C, Arshad MH, Perrett D, Tiberti C, Chiarelli F, Napoli N, Pozzilli P, Nissim A: Antibodies to post-translationally modified insulin in type 1 diabetes. Diabetologia 2015;58:2851-2860

135. Qian B, Wang H, Men X, Zhang W, Cai H, Xu S, Xu Y, Ye L, Wollheim CB, Lou J: TRIB3 [corrected] is implicated in glucotoxicity- and endoplasmic reticulum-stress-induced [corrected] beta-cell apoptosis. J Endocrinol 2008;199:407-416

136. Karaskov E, Scott C, Zhang L, Teodoro T, Ravazzola M, Volchuk A: Chronic palmitate but not oleate exposure induces endoplasmic reticulum stress, which may contribute to INS-1 pancreatic beta-cell apoptosis. Endocrinology 2006;147:3398-3407

137. Cnop M, Ladriere L, Igoillo-Esteve M, Moura RF, Cunha DA: Causes and cures for endoplasmic reticulum stress in lipotoxic beta-cell dysfunction. Diabetes Obes Metab 2010;12 Suppl 2:76-82

138. Cardozo AK, Ortis F, Storling J, Feng YM, Rasschaert J, Tonnesen M, Van Eylen F, Mandrup-Poulsen T, Herchuelz A, Eizirik DL: Cytokines downregulate the sarcoendoplasmic reticulum pump Ca2+ ATPase 2b and deplete endoplasmic reticulum Ca2+, leading to induction of endoplasmic reticulum stress in pancreatic beta-cells. Diabetes 2005;54:452-461

139. Lipson KL, Fonseca SG, Ishigaki S, Nguyen LX, Foss E, Bortell R, Rossini AA, Urano F: Regulation of insulin biosynthesis in pancreatic beta cells by an endoplasmic reticulum-resident protein kinase IRE1. Cell Metab 2006;4:245-254

140. Back SH, Kaufman RJ: Endoplasmic reticulum stress and type 2 diabetes. Annu Rev Biochem 2012;81:767-793

141. Owerbach D, Gabbay KH: Localization of a type I diabetes susceptibility locus to the variable tandem repeat region flanking the insulin gene. Diabetes 1993;42:1708-1714

142. Barratt BJ, Payne F, Lowe CE, Hermann R, Healy BC, Harold D, Concannon P, Gharani N, McCarthy MI, Olavesen MG, McCormack R, Guja C, Ionescu-Tirgoviste C, Undlien DE, Ronningen KS, Gillespie KM, Tuomilehto-Wolf E, Tuomilehto J, Bennett ST, Clayton DG, Cordell HJ, Todd JA: Remapping the insulin gene/IDDM2 locus in type 1 diabetes. Diabetes 2004;53:1884-1889

143. Bennett ST, Lucassen AM, Gough SC, Powell EE, Undlien DE, Pritchard LE, Merriman ME, Kawaguchi Y, Dronsfield MJ, Pociot F, et al.: Susceptibility to human type 1 diabetes at

IDDM2 is determined by tandem repeat variation at the insulin gene minisatellite locus. Nat Genet 1995;9:284-292

144. Guja C, Guja L, Nutland S, Rance H, Todd JA, Ionescu-Tirgoviste C: Strong association of insulin gene INS-VNTR polymorphisms with type 1 diabetes in the Romanian population. Rom J Intern Med 2004;42:313-323

145. Zhang N, Huang W, Dong F, Liu Y, Zhang B, Jing L, Wang M, Yang G, Jing C: Insulin gene VNTR polymorphisms -2221MspI and -23HphI are associated with type 1 diabetes and latent autoimmune diabetes in adults: a meta-analysis. Acta Diabetol 2015;52:1143-1155

146. Walter M, Albert E, Conrad M, Keller E, Hummel M, Ferber K, Barratt BJ, Todd JA, Ziegler AG, Bonifacio E: IDDM2/insulin VNTR modifies risk conferred by IDDM1/HLA for development of Type 1 diabetes and associated autoimmunity. Diabetologia 2003;46:712-720

147. Spielman RS, McGinnis RE, Ewens WJ: Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). Am J Hum Genet 1993;52:506-516

148. Genomes Project C, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, Marchini JL, McCarthy S, McVean GA, Abecasis GR: A global reference for human genetic variation. Nature 2015;526:68-74

149. Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, O'Donnell-Luria AH, Ware JS, Hill AJ, Cummings BB, Tukiainen T, Birnbaum DP, Kosmicki JA, Duncan LE, Estrada K, Zhao F, Zou J, Pierce-Hoffman E, Berghout J, Cooper DN, Deflaux N, DePristo M, Do R, Flannick J, Fromer M, Gauthier L, Goldstein J, Gupta N, Howrigan D, Kiezun A, Kurki MI, Moonshine AL, Natarajan P, Orozco L, Peloso GM, Poplin R, Rivas MA, Ruano-Rubio V, Rose SA, Ruderfer DM, Shakir K, Stenson PD, Stevens C, Thomas BP, Tiao G, Tusie-Luna MT, Weisburd B, Won HH, Yu D, Altshuler DM, Ardissino D, Boehnke M, Danesh J, Donnelly S, Elosua R, Florez JC, Gabriel SB, Getz G, Glatt SJ, Hultman CM, Kathiresan S, Laakso M, McCarroll S, McCarthy MI, McGovern D, McPherson R, Neale BM, Palotie A, Purcell SM, Saleheen D, Scharf JM, Sklar P, Sullivan PF, Tuomilehto J, Tsuang MT, Watkins HC, Wilson

JG, Daly MJ, MacArthur DG, Exome Aggregation C: Analysis of protein-coding genetic variation in 60,706 humans. Nature 2016;536:285-291

150. Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alfoldi J, Wang Q, Collins RL, Laricchia KM, Ganna A, Birnbaum DP, Gauthier LD, Brand H, Solomonson M, Watts NA, Rhodes D, Singer-Berk M, England EM, Seaby EG, Kosmicki JA, Walters RK, Tashman K, Farjoun Y, Banks E, Poterba T, Wang A, Seed C, Whiffin N, Chong JX, Samocha KE, Pierce-Hoffman E, Zappala Z, O'Donnell-Luria AH, Minikel EV, Weisburd B, Lek M, Ware JS, Vittal C, Armean IM, Bergelson L, Cibulskis K, Connolly KM, Covarrubias M, Donnelly S, Ferriera S, Gabriel S, Gentry J, Gupta N, Jeandet T, Kaplan D, Llanwarne C, Munshi R, Novod S, Petrillo N, Roazen D, Ruano-Rubio V, Saltzman A, Schleicher M, Soto J, Tibbetts K, Tolonen C, Wade G, Talkowski ME, Genome Aggregation Database C, Neale BM, Daly MJ, MacArthur DG: The mutational constraint spectrum quantified from variation in 141,456 humans. Nature 2020;581:434-443

151. Le VS, Tran KT, Bui HTP, Le HTT, Nguyen CD, Do DH, Ly HTT, Pham LTD, Dao LTM, Nguyen LT: A Vietnamese human genetic variation database. Hum Mutat 2019;40:1664-1675

152. Jeon S, Bhak Y, Choi Y, Jeon Y, Kim S, Jang J, Jang J, Blazyte A, Kim C, Kim Y, Shim J, Kim N, Kim YJ, Park SG, Kim J, Cho YS, Park Y, Kim HM, Kim BC, Park NH, Shin ES, Kim BC, Bolser D, Manica A, Edwards JS, Church G, Lee S, Bhak J: Korean Genome Project: 1094 Korean personal genomes with clinical information. Sci Adv 2020;6:eaaz7835

153. Le Stunff C, Fallin D, Schork NJ, Bougneres P: The insulin gene VNTR is associated with fasting insulin levels and development of juvenile obesity. Nat Genet 2000;26:444-446

154. Heude B, Dubois S, Charles MA, Deweirder M, Dina C, Borys JM, Ducimetiere P, Froguel P, Fleurbaix Laventie Ville Sante Study G: VNTR polymorphism of the insulin gene and childhood overweight in a general population. Obes Res 2004;12:499-504

155. Sandhu MS, Heude B, Young EH, Luben R, Luan J, Khaw KT, Todd J, Wareham NJ: INS VNTR class genotype and indexes of body size and obesity: population-based studies of 7,999 middle-aged men and women. Diabetes 2005;54:2812-2815

156. Heude B, Petry CJ, Avon Longitudinal Study of Parents Children study t, Pembrey M, Dunger DB, Ong KK: The insulin gene variable number of tandem repeat: associations and interactions with childhood body fat mass and insulin secretion in normal children. J Clin Endocrinol Metab 2006;91:2770-2775

157. Bouatia-Naji N, De Graeve F, Bronner G, Lecoeur C, Vatin V, Durand E, Lichtner P, Nguyen TT, Heude B, Weill J, Levy-Marchal C, Hebebrand J, Froguel P, Meyre D: INS VNTR is not associated with childhood obesity in 1,023 families: a family-based study. Obesity (Silver Spring) 2008;16:1471-1475

158. Stawerska R, Szalapska M, Borowiec M, Antosik K, Mlynarski W, Lewinski A: The influence of INS VNTR class III allele on auxological parameters, glucose, insulin, lipids, and adipocytokines secretion in prepubertal children born small for gestational age. Endokrynol Pol 2016;67:585-591

159. Maas JA, Mook-Kanamori DO, Ay L, Steegers EA, van Duijn CM, Hofman A, Hokken-Koelega AC, Jaddoe VW: Insulin VNTR and IGF-1 promoter region polymorphisms are not associated with body composition in early childhood: the generation R study. Horm Res Paediatr 2010;73:120-127

160. Vu-Hong TA, Durand E, Deghmoun S, Boutin P, Meyre D, Chevenne D, Czernichow P, Froguel P, Levy-Marchal C: The INS VNTR locus does not associate with smallness for gestational age (SGA) but interacts with SGA to increase insulin resistance in young adults. J Clin Endocrinol Metab 2006;91:2437-2440

161. Huxtable SJ, Saker PJ, Haddad L, Walker M, Frayling TM, Levy JC, Hitman GA, O'Rahilly S, Hattersley AT, McCarthy MI: Analysis of parent-offspring trios provides evidence for linkage and association between the insulin gene and type 2 diabetes mediated exclusively through paternally transmitted class III variable number tandem repeat alleles. Diabetes 2000;49:126-130

162. Bennett A, Sovio U, Ruokonen A, Martikainen H, Pouta A, Taponen S, Hartikainen AL, Franks S, Peltonen L, Elliott P, Jarvelin MR, McCarthy MI: No association between insulin gene variation and adult metabolic phenotypes in a large Finnish birth cohort. Diabetologia 2005;48:886-891

163. Hansen SK, Gjesing AP, Rasmussen SK, Glumer C, Urhammer SA, Andersen G, Rose CS, Drivsholm T, Torekov SK, Jensen DP, Ekstrom CT, Borch-Johnsen K, Jorgensen T, McCarthy MI, Hansen T, Pedersen O: Large-scale studies of the HphI insulin gene variable-number-of-tandem-repeats polymorphism in relation to Type 2 diabetes mellitus and insulin release. Diabetologia 2004;47:1079-1087

164. Mitchell SM, Hattersley AT, Knight B, Turner T, Metcalf BS, Voss LD, Davies D, McCarthy A, Wilkin TJ, Smith GD, Ben-Shlomo Y, Frayling TM: Lack of support for a role of the insulin gene variable number of tandem repeats minisatellite (INS-VNTR) locus in fetal growth or type 2 diabetes-related intermediate traits in United Kingdom populations. J Clin Endocrinol Metab 2004;89:310-317

165. Fendler W, Wyka K, Cieslik-Heinrich A, Polakowska E, Jastrzebska J, Mlynarski W: The 5'VNTR proinsulin gene (INS) polymorphism and the functional reserve of beta cells in the intravenous glucose tolerance test. Pediatr Endocrinol Diabetes Metab 2011;17:5-9

166. Vafiadis P, Bennett ST, Todd JA, Nadeau J, Grabs R, Goodyer CG, Wickramasinghe S, Colle E, Polychronakos C: Insulin expression in human thymus is modulated by INS VNTR alleles at the IDDM2 locus. Nat Genet 1997;15:289-292

167. Vafiadis P, Bennett ST, Colle E, Grabs R, Goodyer CG, Polychronakos C: Imprinted and genotype-specific expression of genes at the IDDM2 locus in pancreas and leucocytes. J Autoimmun 1996;9:397-403

168. Pugliese A, Zeller M, Fernandez A, Jr., Zalcberg LJ, Bartlett RJ, Ricordi C, Pietropaolo M, Eisenbarth GS, Bennett ST, Patel DD: The insulin gene is transcribed in the human thymus and transcription levels correlated with allelic variation at the INS VNTR-IDDM2 susceptibility locus for type 1 diabetes. Nat Genet 1997;15:293-297

169. Durinovic-Bello I, Wu RP, Gersuk VH, Sanda S, Shilling HG, Nepom GT: Insulin gene VNTR genotype associates with frequency and phenotype of the autoimmune response to proinsulin. Genes Immun 2010;11:188-193

170. Xin Y, Kim J, Okamoto H, Ni M, Wei Y, Adler C, Murphy AJ, Yancopoulos GD, Lin C, Gromada J: RNA Sequencing of Single Human Islet Cells Reveals Type 2 Diabetes Genes. Cell Metab 2016;24:608-615

171. Xin Y, Dominguez Gutierrez G, Okamoto H, Kim J, Lee AH, Adler C, Ni M, Yancopoulos GD, Murphy AJ, Gromada J: Pseudotime Ordering of Single Human beta-Cells Reveals States of Insulin Production and Unfolded Protein Response. Diabetes 2018;67:1783-1794

172. Baron M, Veres A, Wolock SL, Faust AL, Gaujoux R, Vetere A, Ryu JH, Wagner BK, Shen-Orr SS, Klein AM, Melton DA, Yanai I: A Single-Cell Transcriptomic Map of the Human and Mouse Pancreas Reveals Inter- and Intra-cell Population Structure. Cell Syst 2016;3:346-360 e344

173. Enge M, Arda HE, Mignardi M, Beausang J, Bottino R, Kim SK, Quake SR: Single-Cell Analysis of Human Pancreas Reveals Transcriptional Signatures of Aging and Somatic Mutation Patterns. Cell 2017;171:321-330 e314

174. Lawlor N, George J, Bolisetty M, Kursawe R, Sun L, Sivakamasundari V, Kycia I, Robson P, Stitzel ML: Single-cell transcriptomes identify human islet cell signatures and reveal cell-type-specific expression changes in type 2 diabetes. Genome Res 2017;27:208-222

175. Report of the Expert Committee on the Diagnosis and Classification of Diabetes Mellitus. Diabetes Care 1997;20:1183-1197

176. Ma L, Zheng J: Single-cell gene expression analysis reveals beta-cell dysfunction and deficit mechanisms in type 2 diabetes. BMC Bioinformatics 2018;19:515

177. Segerstolpe A, Palasantza A, Eliasson P, Andersson EM, Andreasson AC, Sun X, Picelli S, Sabirsh A, Clausen M, Bjursell MK, Smith DM, Kasper M, Ammala C, Sandberg R: Single-Cell Transcriptome Profiling of Human Pancreatic Islets in Health and Type 2 Diabetes. Cell Metab 2016;24:593-607

178. Wang YJ, Schug J, Won KJ, Liu C, Naji A, Avrahami D, Golson ML, Kaestner KH: Single-Cell Transcriptomics of the Human Endocrine Pancreas. Diabetes 2016;65:3028-3038

179. Camunas-Soler J, Dai XQ, Hang Y, Bautista A, Lyon J, Suzuki K, Kim SK, Quake SR, MacDonald PE: Patch-Seq Links Single-Cell Transcriptomes to Human Islet Dysfunction in Diabetes. Cell Metab 2020;31:1017-1031 e1014

180. Muraro MJ, Dharmadhikari G, Grun D, Groen N, Dielen T, Jansen E, van Gurp L, Engelse MA, Carlotti F, de Koning EJ, van Oudenaarden A: A Single-Cell Transcriptome Atlas of the Human Pancreas. Cell Syst 2016;3:385-394 e383

181. Fang Z, Weng C, Li H, Tao R, Mai W, Liu X, Lu L, Lai S, Duan Q, Alvarez C, Arvan P, Wynshaw-Boris A, Li Y, Pei Y, Jin F, Li Y: Single-Cell Heterogeneity Analysis and CRISPR Screen Identify Key beta-Cell-Specific Disease Genes. Cell Rep 2019;26:3132-3144 e3137

182. Stark R, Grzelak M, Hadfield J: RNA sequencing: the teenage years. Nat Rev Genet 2019;20:631-656

183. Chen G, Ning B, Shi T: Single-Cell RNA-Seq Technologies and Related Computational Data Analysis. Front Genet 2019;10:317

184. Islam S, Zeisel A, Joost S, La Manno G, Zajac P, Kasper M, Lonnerberg P, Linnarsson S: Quantitative single-cell RNA-seq with unique molecular identifiers. Nat Methods 2014;11:163-166

185. Wang YJ, Kaestner KH: Single-Cell RNA-Seq of the Pancreatic Islets--a Promise Not yet Fulfilled? Cell Metab 2019;29:539-544

186. Wojtusciszyn A, Armanet M, Morel P, Berney T, Bosco D: Insulin secretion from human beta cells is heterogeneous and dependent on cell-to-cell contacts. Diabetologia 2008;51:1843-1852

187. Quesada I, Todorova MG, Alonso-Magdalena P, Beltra M, Carneiro EM, Martin F, Nadal A, Soria B: Glucose induces opposite intracellular Ca2+ concentration oscillatory patterns in identified alpha- and beta-cells within intact human islets of Langerhans. Diabetes 2006;55:2463-2469

188. Cabrera O, Berman DM, Kenyon NS, Ricordi C, Berggren PO, Caicedo A: The unique cytoarchitecture of human pancreatic islets has implications for islet cell function. Proc Natl Acad Sci U S A 2006;103:2334-2339

189. Team RC: R: A language and environment for statistical computing. In *R Foundation for Statistical Computing* Vienna, Austria, 2019

190. Kaestner KH, Powers AC, Naji A, Consortium H, Atkinson MA: NIH Initiative to Improve Understanding of the Pancreas, Islet, and Autoimmunity in Type 1 Diabetes: The Human Pancreas Analysis Program (HPAP). Diabetes 2019;68:1394-1402

191. Ramskold D, Luo S, Wang YC, Li R, Deng Q, Faridani OR, Daniels GA, Khrebtukova I, Loring JF, Laurent LC, Schroth GP, Sandberg R: Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells. Nat Biotechnol 2012;30:777-782

192. Picelli S, Bjorklund AK, Faridani OR, Sagasser S, Winberg G, Sandberg R: Smart-seq2 for sensitive full-length transcriptome profiling in single cells. Nat Methods 2013;10:1096-1098

193. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR: STAR: ultrafast universal RNA-seq aligner. Bioinformatics 2013;29:15-21

194. Li H: A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics 2011;27:2987-2993

195. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Genome Project Data Processing S: The Sequence Alignment/Map format and SAMtools. Bioinformatics 2009;25:2078-2079

196. Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM, 3rd, Hao Y, Stoeckius M, Smibert P, Satija R: Comprehensive Integration of Single-Cell Data. Cell 2019;177:1888-1902 e1821

197. Kuleshov MV, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, Koplev S, Jenkins SL, Jagodnik KM, Lachmann A, McDermott MG, Monteiro CD, Gundersen GW,

Ma'ayan A: Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic Acids Res 2016;44:W90-97

198. Shapiro SS, Wilk MB: An Analysis of Variance Test for Normality (Complete Samples). Biometrika 1965;52:591-611

199. Conover WJ, Johnson ME, Johnson MM: A Comparative Study of Tests for Homogeneity of Variances, with Applications to the Outer Continental Shelf Bidding Data. Technometrics 1981;23:351-361

200. Levene H: Robust tests for the equality of variance. p. 278–292. I. Olkin (ed.) Contributions to probability and statistics. Stanford Univ. Press, Palo Alto, CA. Robust tests for the equality of variance p 278–292 In I Olkin (ed) Contributions to probability and statistics Stanford Univ Press, Palo Alto, CA 1960:-

201. Mann HB, Whitney DR: On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other. Ann Math Statist 1947;18:50-60

202. Kruskal WH, Wallis WA: Use of Ranks in One-Criterion Variance Analysis. Journal of the American Statistical Association 1952;47:583-621

203. Dunn OJ: Multiple Comparisons Using Rank Sums. Technometrics 1964;6:241-252

204. Quirk T: One-Way Analysis of Variance (ANOVA). 2016:165-182

205. Keppel G: *Design and analysis: A researcher's handbook*. Prentice-Hall, Inc, 1991

206. Da Silva Xavier G: The Cells of the Islets of Langerhans. J Clin Med 2018;7

207. Geer EB, Shen W: Gender differences in insulin resistance, body composition, and energy balance. Gend Med 2009;6 Suppl 1:60-75

208. Gannon M, Kulkarni RN, Tse HM, Mauvais-Jarvis F: Sex differences underlying pancreatic islet biology and its dysfunction. Mol Metab 2018;15:82-91

209. Mirghani Dirar A, Doupis J: Gestational diabetes from A to Z. World J Diabetes 2017;8:489-511

210. Mezza T, Cinti F, Cefalo CMA, Pontecorvi A, Kulkarni RN, Giaccari A: beta-Cell Fate in Human Insulin Resistance and Type 2 Diabetes: A Perspective on Islet Plasticity. Diabetes 2019;68:1121-1129

211. Reimegård J, Danielsson M, Tarbier M, Schuster J, Baskaran S, Panagiotou S, Dahl N, Friedländer M, Gallant CJ: Combined mRNA and protein single cell analysis in a dynamic cellular system using SPARC. bioRxiv 2019;