The Effects of Metabolic Conditions Relevant to Type 2 Diabetes on Alternative Splicing

Mahfuja Jannat Bulu

Director of studies: Dr Natasha J Hill

Supervisory Team: Dr Lucy Jones and Dr Jessica L Buxton

A thesis submitted in partial fulfilment of the requirements of Kingston University for the award of a MSc by Research Degree

School of Life Science, Pharmacy and Chemistry

Kingston University London

August 2018

Declaration

I, Mahfuja Jannat Bulu, declare that this thesis has been submitted for the degree of MSc by Research at Kingston University London. This thesis is my own research and contribution to this work by other individuals has been fully acknowledged. Previous published work has been quoted and the authors of those publications have been acknowledged by referencing.

I confirm that this thesis has not been previously submitted for the award of a degree by this or any other university.

I give consent to Kingston University's Learning Resource Centre to hold a copy of this thesis.

Mahfuja Jannat Bulu

Date: 26/10/2018

I dedicate this thesis to my beloved father,

who left this world on 30th October 2017.

Acknowledgements

I would like to thank my supervisor, Associate professor Dr Natasha Hill, for her kind guidance, patience and dedication throughout the period of my research. I came back to my studies after a ten-year break and it is her non-stop guidance and encouragement, that has brought me here today. Her support extended beyond academics as she looked after me in many ways, especially during the period when I lost my father. It is my pleasure to work have worked with her.

I would like to express my gratitude to my second supervisor Dr Jessica Buxton for her continued help and guidance. Thank you for being present at all my presentations, sharing your knowledge and giving valuable advice throughout my studentship.

I would also like to thank Dr Lucy Jones for her support especially at the beginning when I was not very confident in my own abilities.

I would like to thank Professor Dr Mehmet T Dorak for his inspirational talk which motivated me in a difficult time which he may not be even aware of. Thank you for providing us with computer and statistics lessons which have helped me develop many skills. You deliver lessons to students in an easy and understandable way. You may not even recognise me among your thousands of students.

I want to thank all my group members: Amtul, Jamila, Sharan, Cris and Katrina for their non-stop support and help. I would like to give a special thanks to Amanda for her help; she never said no when I asked for any kind of help.

I would like to express my gratitude to our lab manager Dr Lauren for her relentless assistance throughout my entire thesis period. Thanks for trusting me and my capabilities. It was a real pleasure working with you. I am always inspired by your honesty. I would like to thank all members of the Integrated Research and Teaching Laboratory (IRTL) - Dr John, Lauren, Natasha, Najma, Hiba, Avinish and Luke; all of you have helped me greatly. When I lost my father, you guys all directly and indirectly gave me support so thank you all for that.

A special thanks to Dr Michael, Jayne for your kind support.

I am overwhelmed by the support from the library staff and my language teacher, Gabi, during my period of study.

Finally, I want to show my gratitude to my daughter. You are my mental strength and I am truly blessed to have you in my life. Grateful to my Husband for his relentless support, providing a shoulder to lean on during all the ups and downs in my life. Because of you, I dared to come back to study. Thanks to my son for his inspiration and support. And finally, thank you to my mother for everything in my life.

Abstract

Type 2 diabetes is a metabolic disorder associated with chronic hyperglycaemia and hyperlipidaemia, with consequent insulin resistance and a progressive deterioration of pancreatic β cell mass and function. The combined deleterious effects of these metabolic changes on β cell are known as glucolipotoxicity (GLT). Alternative splicing is a post-transcriptional process that allows a single gene to encode multiple transcripts and is a crucial mechanism for generating proteomic diversity. However, little is known about the role of alternative splicing in diabetes. The aim of this study was to test whether metabolic conditions relevant to type 2 diabetes affect alternative splicing in β cells.

INS1 β cells were treated under GLT or control conditions and the expression of a panel of genes important for the regulation of alternative splicing was examined by quantitative reverse transcriptase PCR (RT-qPCR). These experiments showed that GLT conditions significantly increase expression of the stress-associated mRNA processing and stability gene *Elavl1* in β cells ($p=0.04$, $n=4$ independent experiments). The role of *Elavl1* in β cells is not known, but it has been shown previously that siRNA knockdown of the related family member, *Elavl4*, increases β cell apoptosis, suggesting that this family of genes regulates β cell survival. Using a combination of exon arrays and a candidate gene approach, we have further shown that the increased *Elavl1* expression in β cells by GLT conditions is associated with alternative splicing of apoptosis genes, including a variant of the *TCF7L2* gene thought to induce β cell apoptosis.

This suggests that the metabolic conditions existing in many people with type 2 diabetes may cause aberrant alternative splicing in β cells increasing susceptibility to apoptosis and subsequent β cell loss.

Table of Contents

List of Figures

Chapter 1- Introduction

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Chapter 6

List of Tables

Abbreviations

ADA	American Diabetes Association
AREs	AU-rich elements
AS	Alternative Splicing
ATF	Activating Transcription Factor
ATP	Adenosine Tri-Phosphate
B2M	Beta-2-microglobulin
BSA	Bovine Serum Albumin
CtBP	C-terminal binding protein
$Ca2^{+}$	Calcium ion
DE	Differentially Expressed
DM	Dimethyl sulfoxide
DNA	Deoxyribonucleic Acid
DP5	Death Protein 5
ELAV	Embryonic lethal abnormal vision
ER	Endoplasmic Reticulum
ESE	Exonic Splicing Enhancer
ESS	Exonic Splicing Silencer
FasL	Fas Ligand
FBS	Fetal Bovine Serum
FFAs	Free Fatty Acids
Gapdh	Glyceraldehyde-3-phosphate dehydrogenase
GLP-1	Glucagon Like Peptide 1
GLT	Glucolipotoxicity
GLUT2	Glucose transporter 2
GSIS	Glucose Stimulated Insulin Secretion
HnRNP	Heterogeneous Ribonucleoprotein
IR	Insulin Resistance
ISE	Intronic splicing Enhancer
ISS	Intronic Splicing Silencer
JNK	c-Jun N-terminal Kinase
KD	Knock Down

LC-CoA	Long Chain Acyl-CoA
mRNA	Messenger RNA
MAPK	Mitogen activated protein kinases
NCBI	National Centre for Biotechnology Information
NCDs	Noncommunicable diseases
NTC	No template control
ORF	Open Reading Frame
PBS	Phosphate Buffer Saline
Ppia	Peptidylprolyl isomerase A
PCR	Polymerase chain reaction
PL	phospholipids
PSR	Probe selection region
qPCR	Quantitative PCR
RBP	RNA Binding Protein
RIN	RNA Integrity Number
RNA	Ribonucleic Acid
RNA	Ribonucleic Acid
RNP	Ribonucleoprotein
rpm	Revolutions per minute
RT-PCR	Reverse Transcriptase PCR
SiRNA	Small interfering RNA
SNP	Single nucleotide polymorphism
SR protein	Serine/Arginine Rich Protein
T1D	Type 1 Diabetes
T2D	Type 2 Diabetes
TAE buffer	Tris-acetate buffer
TCF7L2	T-cell transcription factor 7 like 2
Tubb5	Tubulin, beta 5
UCP2	Mitochondrial Uncoupling Protein 2
UPR	Unfolded Protein Response
UTR	Untranslated region
VDCC	Voltage Dependent Calcium Channel

WHO	World Health Organisation
Ywhaz	Tyrosine 2/tryptophan 5-monooxygenase activation protein

1 Chapter 1- Introduction

1.1 A brief history and introduction to diabetes

Diabetes Mellitus (DM) is a complex heterogenous group of metabolic conditions characterized by hyperglycaemia resulting from impairment in insulin secretion, insulin action, or both (ADA, 2005; Das and Elbein, 2006). The pancreatic β cell and its secretory product insulin are central in the pathophysiology of diabetes (Atikson and Elisenbarth *et al*, 2001). The potential complications of T2D are heart disease and long-term damage, to different organs including the eyes, kidneys, the heart, nerves and blood vessels (WHO, 1999). Diabetes mellitus has become one of the main threats to human health in the 21st century (Zimmet *et al*, 2000) and is adding a huge cost to health care systems.

According to the American Diabetes Association (1997) and World Health Organisation (1999), there are two major types of diabetes:

Type 1 diabetes (T1D) results from the cell-mediated autoimmune destruction of pancreatic β-cells. Exogenous insulin is currently required for survival to prevent the development of ketoacidosis, coma and death (Piya and Michels, 2012). It is a complex, multifactorial disease caused by the interaction of genetic and poorly defined environmental factors. T1D occurs mainly in children and young adults and accounts for about 10 percent of diabetes cases (ADA Expert Committee, 1997). The incidence of T1D has doubled in last two decades (Harjutsalo *et al*, 2005; Harjutsalo *et al*, 2008, Chen *et al*, 2011).

Type 2 diabetes (T2D) is the most common form of diabetes and is characterized by insulin resistance and either the dysfunction or loss of pancreatic β cells. The development of these abnormalities is not fully understood (WHO, 1999). People with T2D can often control the disease with a combination of diet, exercise and non-insulin drugs for example, to enhance insulin sensitivity, though insulin is required in some cases. Insulin resistance and T2D are strongly associated with

age, obesity and physical inactivity and may remain undiagnosed for several years where hyperglycaemia is not severe enough to cause clinical symptoms. These individuals are, however, at increased risk of developing macrovascular and microvascular complications.

The prevalence of T2D varies in different racial/ethnic subgroups (Weyer *et al*, 2001). Candidate gene association and genome-wide association studies (GWAS) have identified many diabetes susceptibility loci associated with T2D corresponding in many cases to genes that impact glucose induced insulin secretion, and include: *TCF7L2*, *CAPN10*, *HNF4α*, *PTPN1, ADAMTS9, FOX01,* HNEX, APOB etc (Sladek *et al*, 2007; Imamura and Maeda, 2011; Horikawa *et al*, 2000; Das and Elbein, 2006).

According to the International Diabetes Federation (IDF), approximately 425 million adults were living with diabetes in 2017 which will rise to 629 million by 2045 and 352 million people were at risk of developing T2D. The proportion of people with T2D increasing in most countries (IDF Diabetes Atlas, 2017). This issue needs to be addressed urgently.

1.2 Molecular and metabolic mechanisms of insulin resistance and β cell failure in T2D

Insulin resistance can be defined as the inability of insulin to produce its usual biological actions at circulating concentrations that are effective in normal subjects. It can be caused by defects such as decreases in receptor concentration and kinase activity, and in the activity of intracellular enzymes (Pessin *et al*, 2000).

Studies of T2D have shown that insulin secretion and insulin action control glucose homeostasis in a dual regulatory cycle: a rise in glucose concentration stimulates insulin secretion which lowers plasma glucose in a time and concentration manner. Additionally, sustained hyperinsulinemia inhibits insulin

secretion and insulin action (Ferrannini,1998). When genetic predisposition and/or an unhealthy lifestyle results in decreased insulin sensitivity, this is initially compensated for by an increase in β cell function or mass, so that glucose tolerance remains normal. When β cell function and insulin sensitivity both deteriorate, postprandial glucose tolerance becomes abnormal. At this point, β cell function is typically sufficient to maintain normal fasting plasma glucose concentration. As a result of further deterioration in β cell function and insulin sensitivity fasting glucose concentrations increase due to an increase in basal endogenous glucose production. Finally, because of further deterioration in β cell function, both fasting and postprandial glucose levels reach diabetic levels (Kahn, 2001 and Weyer *et al*, 2001). The potential mechanisms by which the metabolic changes arise as a consequence of insulin resistance lead to β cell loss/dysfunction are discussed in section 1.4.

1.3 Normal function of β cells in response to glucose

Glucose stimulated insulin release is biphasic comprising a rapid first phase lasting 5-10 mins, followed by a prolonged second phase, which continues for the duration of the stimulus. In T2D both responses are diminished or blunted (Luzi *et al*,1989). The events which lead to the release of insulin from β cells are complex. The insulin released from the pancreas during first phase has been synthesized and stored in the secretory granules of β cells. In response to an increase in extracellular glucose, ATP levels increase causing closure of ATP-sensitive potassium channels and depolarization of the β cell membrane. This results in an influx of calcium that triggers the exocytosis of insulin-containing granules and release of granule contents (Figure1.1). The second phase insulin release involves synthesis of new insulin molecules as well as ATP-dependent mobilization of granules from a storage pool into the rapidly releasable pool. Impaired pulsatile insulin secretion is an early marker of β cell dysfunction in T2D (Mitrakou *et al*, 1990).

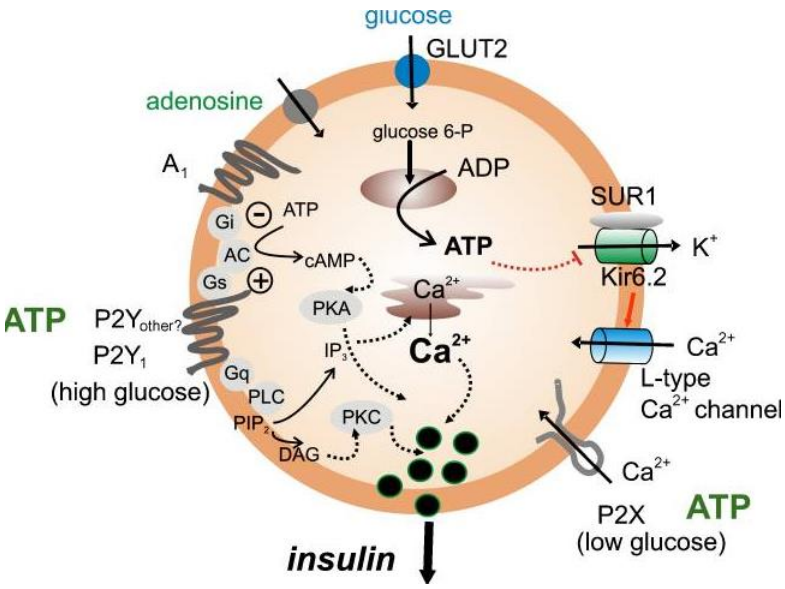


Figure1.1: Regulation of insulin secretion in the β-cell. Glucose enters the cell via the GLUT2 receptor (in rodents). The glucose transporter in human beta cells is thought to be GLUT1 and GLUT3, it is GLUT2 in rodents (McCulloch *et al*, 2011). ATP levels increase by glucose metabolism which closes the ATP-sensitive channel, K^+ATP, which depolarises the cell membrane and opens voltage-gated Ca^{2+} channel. As a result, Ca^{2+} action potentials are generated. An increase in the cellular Ca^{2+} is the triggering event that leads to exocytosis of secretory vesicles containing insulin. Image Reference, Novak, 2008.

1.4 Glucolipotoxicity of the pancreatic β cells

Hyperglycaemia and elevated circulating free fatty acids (FFAs) are thought to contribute to β cell failure by causing cellular stress and ultimately β cell apoptosis, resulting in insulin insufficiency (Deborah *et al*, 2008). However, the underlying molecular and cellular mechanisms of glucolipotoxicity in T2D are still debated.

Elevated FFAs in healthy individuals enhance insulin secretion but contribute to insulin resistance and progressive β cell failure in individuals with genetic predisposition to diabetes. Increased concentrations of saturated fatty acids are toxic to islet β cells (Montane *et al*, 2014). Palmitate is the most abundant saturated fatty acids in human plasma and its deleterious effects are mediated via ceramide mitochondrial apoptotic pathways (Maedler *et al*, 2001). Perturbations in mitochondrial permeability are observed early in the

development of fatty acid induced cell death in β cells (Koshkin in Poitout *et al*, 2010). In rats and humans, both *in vitro* and *in vivo*, prolonged exposure of β cells to FFA inhibits glucose stimulated insulin secretion (Mason and Paolisso in Poitout *et al*, 2010). Furthermore, uncoupling protein 2 (UCP2), a mitochondrial carrier that regulates ATP production, is upregulated by prolonged elevation of FFAs, and is thought to also play a role in glucolipotoxicity (Joseph *et al*, 2004).

However, further experiments have shown that exposure of rodent islets to high fatty acid conditions only impairs glucose-stimulated insulin secretion (GSIS) when the sugar concentration goes above a threshold concentration of approximately 8 mM (Muoio and Newgard, 2008). This suggests that β cell functional impairment is a consequence of the combination of elevated levels of both sugar and fatty acids, or glucolipotoxicity (GLT) rather than an exposure of either alone.

According to this model (outlined in Figure 1.2), in the presence of simultaneously elevated levels of glucose and fatty-acid (FA), there is an increase in cytosolic malonyl-CoA and the formation of cataplerotic signals e.g. citrate. Since fatty acid synthase activity is lower than that of acetyl CoA carboxylase in the β cell, the predominant activity of malonyl-CoA is to inhibit carnitine-palmitoyl transferase-1 (CPT-1) activity, which in turn blocks fatty acid oxidation and leads to accumulation of long-chain acyl CoA esters (LC-CoA) in the cytosol. Accumulation of cytosolic LC-CoA adversely affects β cell function. Transport of long-chain acyl-CoA (LC-CoA) in the mitochondria is reduced, and the esterification pathway is preferentially activated, leading to cytosolic accumulation of lipid-derived signalling molecules such as ceramide, diglycerides (DG), phosphatidic acid, phospholipids (PL) and triglycerides (TG). Therefore, fatty acid partitioning towards esterification and cellular lipid synthesis strengthens the cellular mechanism of glucolipotoxicity in β cells and the nature of the lipid derived metabolites are responsible for the deleterious effects (Prentki and Corkey, 1996; Poitout *et al*, 2010).

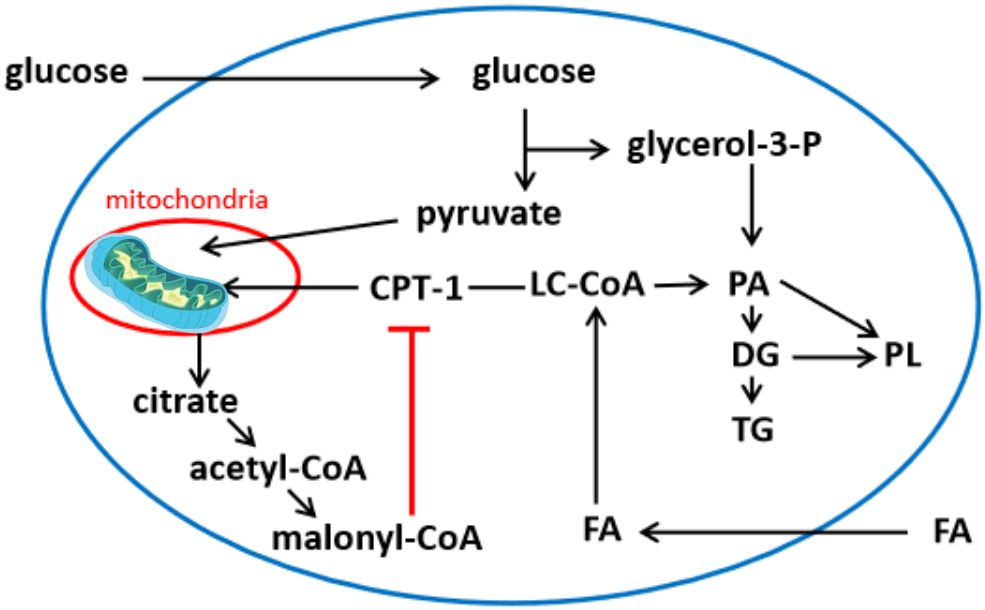


Figure 1.2: The permissive effects of glucose on the deleterious actions of fatty acids in pancreatic β cells. High sugar and fatty acids cause an increased cytosolic concentration of malonyl CoA which inhibits the enzyme carnitine palmitoyl transferase 1. As a result, long chain acyl CoA (LC-CoA) are reduced in mitochondria and an esterification pathway is preferentially activated. It is the lipid metabolites that is responsible for their deleterious effects.

Using the rat INS1 cell line, it has been shown that after exposure of β cells to high sugar and high fatty acid condition, there is a compensatory phase during which β cell proliferates which ultimately leads to increased insulin gene expression and secretion (Mason and Paolisso in Poitout *et al*, 2010, Sako in Poitout *et al*, 2010; Deborah *et al*, 2008). However, prolonged exposure of β cells to GLT conditions ultimately causes β cell failure and apoptosis (Jacqueminet in Poitout *et al*, 2010; Briaud in Poitout *et al*, 2010). A model showing the etiology of T2D in genetically susceptible individuals with excessive carbohydrate and fat intake is shown in Figure 1.3 (Prentki *et al*, 2002).

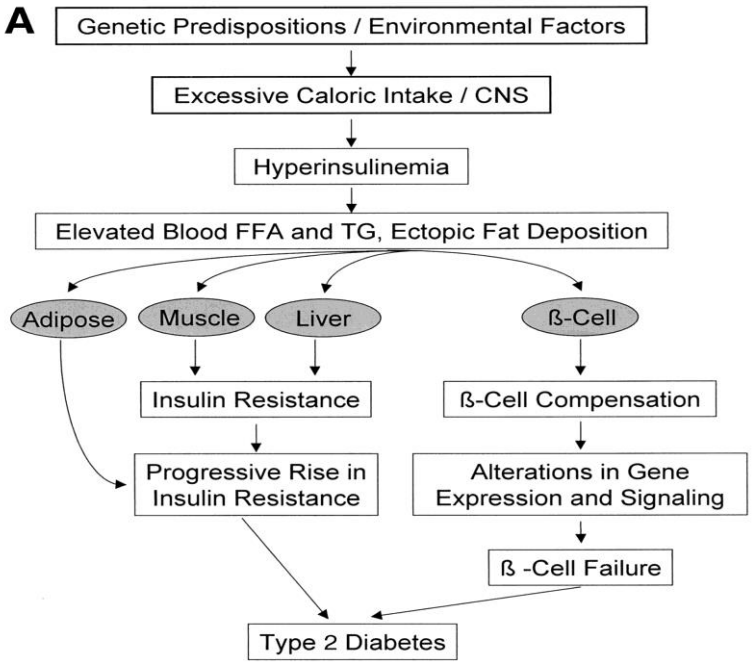


Figure1.3: Model illustrating the association of genetic predisposition and overnutrition in the etiology of T2D. In this model, genetically susceptible individuals, excessive carbohydrate and fat intake causes hyperinsulinemia along with increased hepatic lipoprotein secretion and elevated FFAs. This eventually causes insulin resistance in muscle and liver. On the other hand, excessive fuel allows β cell compensation by increased insulin secretion and β cell growth. The progressive rise in insulin resistance with altered β cell gene expression will lead to increase postprandial hyperglycaemia, which along with FFAs, will cause β cell failure and as a result of this changes in insulin secretion and β cell apoptosis diabetes emerges. Figure taken from Prentki *et al*, 2002.

1.4.1 ER Stress

It has been suggested that β cell failure may also be caused by activation of the unfolded protein response (UPR) and ER stress. When β cells are exposed to an increased secretory demand due to high sugar and overnutrition, the endoplasmic reticulum may struggle to cope with the increased rate of protein synthesis needed to maintain insulin secretion and blood glucose homeostasis, in ER stress (Cunha *et al*, 2008 and Bagnati *et al*, 2016).

The molecular mechanisms by which glucolipotoxicity induced ER stress causes β cell death are not fully understood but may involve a depletion of ER Ca^{2+} stores (Cunha *et al*, 2008 and Bagnati *et al*, 2016) and involve in activation of *ATF3*,

CHOP, Jun N-terminal kinase (*JNK*), caspases 4 and 12, and the mitochondrial apoptosis pathway (Eizirik *et al*, 2008 and Chang-Chen *et al*, 2008). On the other hand, other studies suggest that high glucose levels only mildly activate ER stress signalling (Elouil *et al*, 2007). The long chain FFAs, palmitate and oleate, can induce ER stress via the activation of mitochondrial pathway of cell death through up regulation of pro apoptotic BCL-2 family member death protein 5 (*DP5*), (Oliveira *et al*, 2015). But the combined effect of FFAs and glucose can trigger β cells to activate apoptosis via any of the mechanisms described above such as UPR and/or the mitochondrial and oxidative stress etc.

1.4.2 Limitation of glucolipotoxicity

To define the cellular and molecular mechanisms of glucolipotoxicity, *in vitro* models using insulin-secreting cells are extremely valuable. But proper care should be taken when interpreting the results. For example, it is known that there are species-related differences in the sensitivity of β cells to fatty acid-induced cell death (Poitout *et al*, 2010). Using insulin-secreting cells and isolated islets, elevated glucose and fatty acid-induced β cell apoptosis can be observed after 24h culture in human cells (El-Assad *et al*, 2003), whereas for rat cells 72h culture of β cells was not enough to induce apoptosis (Kelpe *et al*, 2003). However, due to this species related sensitivity, research results using model organisms should ideally be validated in human cells.

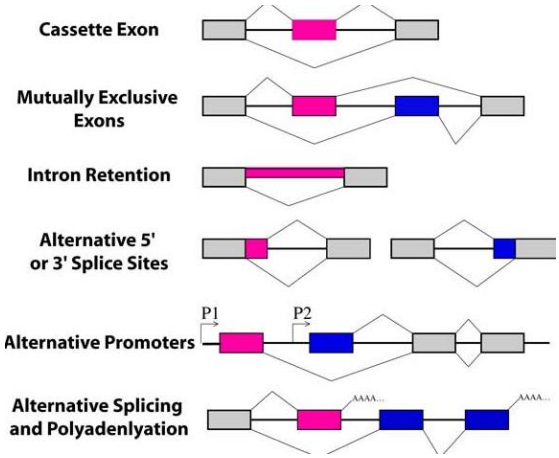
Another limitation of using INS1 cells are that INS1 are proliferative, whereas native beta cells are not usually (Skelin *et al*, 2010).

1.5 Alternative Splicing

1.5.1 Understanding of alternative splicing and its function

Alternative splicing (AS) is a regulated process during gene expression that results in a single gene coding for multiple proteins (Black, 2003). In other way, many genes in higher eukaryotes encode RNAs that can be spliced in alternative ways to generate two or more different mRNAs and thus different protein products (isoforms). AS is a crucial mechanism for gene regulation and for proteomic diversity and markedly affects human development. Its misregulation also underlies many human diseases (Zhou and Fou, 2013). Alternative splicing is achieved by exonic/intronic splicing enhancers/silencers (ESE, ISE, ESS, ISS) in combination with spatial and temporal expression of trans-acting splicing factors such as serine/arginine-rich (SR) proteins and heterogenous nuclear ribonucleoproteins (Zhou and Fou, 2013). It can be regulated at different stages of spliceosome assembly by different splicing factors and by many mechanisms that rely on *cis*-acting elements.

Transcriptome analysis shows that 90% mammalian genes have several transcript isoforms along with around 160,000 alternative spliced transcripts that code for proteins. Those isoforms can potentially play an important role in disease (Webster, 2017).



Taken from 'WormBook by Zahler, A. (2012).

Figure 1.4: Schematic presentation of different types of alternative splicing. Grey boxes represent constitutive exons and others are alternatively spliced exons or regions. Solid lines indicate introns and dashed lines are indicative of alternative pattern. Promoters are indicated by arrows and polyadenylation site by AAA. Types of alternative splicing are cassette exon, retained introns, mutually exclusive exons, use of alternative 5' splice sites and alternative 3' splice sites.

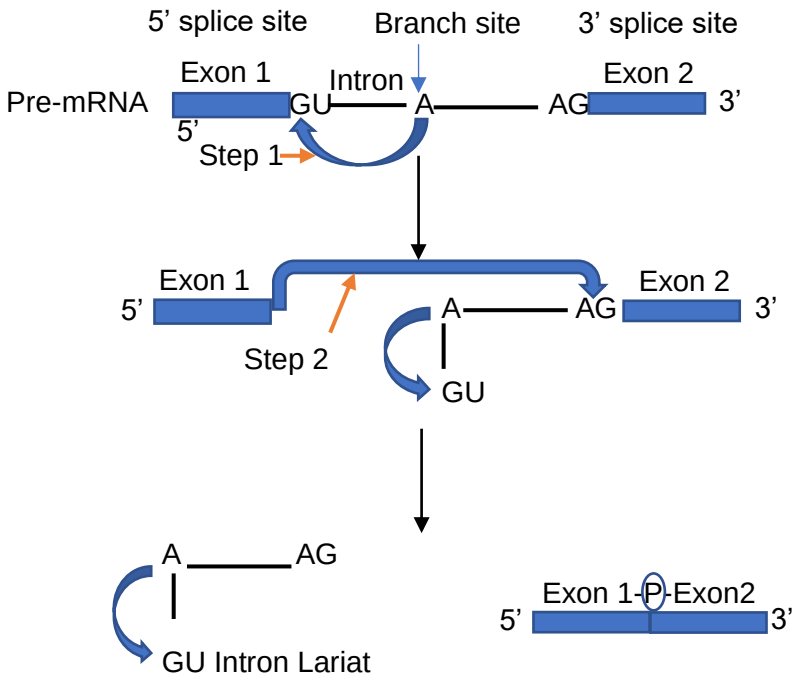


Figure 1.5: The splicing reaction. Two transesterifications of splicing reaction in which phosphodiester linkage within the pre-mRNA are broken and new ones are formed; introns are removed in a form called a lariat.

There are numerous types of alternative splicing that can be observed, depending on the way exons are combined. The most frequently observed events are exon skipping or cassette exons, mutually exclusive exons, alternative donor or acceptor sites and intron retention. A schematic presentation of these alternative splicing events is shown in Figure 1.4.

1.5.2 Mechanism of AS and its regulation

The core splicing reaction consists of two successful transesterification reactions which involves two nucleophilic attacks on the phosphodiester bonds (Figure 1.5). The first reaction is carried out by 2'-OH group of the branch point which is usually adenine and in the second reaction by 3'-OH group of the upstream exon.

Splicing is carried out by an RNA-dependent molecular machine called the spliceosome, which consists of five small nuclear ribonucleoprotein particles (snRNPs) and a group of auxiliary proteins that cooperate to accurately recognise the splice site and catalyse the splicing reaction. Initially the 5' splice site is recognised by small ribonucleoprotein U1 snRNP and the binding of splicing factor 1 (SF1) to the branch point in an ATP independent manner (E' complex). The E' complex converts in E complex by the binding of auxiliary factor U2AF to the polypyrimidine tract and the 3' splice site AG. The E complex is then converted into ATP-independent pre-spliceosome A complex by a replacement of SF1 by U2 snRNP at the branch point. In constitutive splicing, U2 binds the branch point via SF3B1 (Holly *et al*, 2014). Further recruitment of the U4/U6-U5 tri-SNP complex leads to the formation of the ATP dependent B complex, which is converted into catalytically competent splicing complex (Chen *et al*, 2010). The complex containing the pre-mRNA, and U1 and U2 snRNPs is called the pre-spliceosomal complex and defines the introns when the size of introns are less than 250 nucleotides (Hertel and Berget in Webster, 2017). However, most introns are kilobases in length and it is thought that splicing *in vivo* requires interaction of U1 and U2 snRNPs across the exons rather than introns, a process termed as 'exon definition'. This is followed by long range splice site pairing

across the intron to assemble the functional spliceosome (Figure 1.6); (Hertel and Berget in Webster, 2017).

The interaction between *cis*-acting elements and trans-acting factors determines whether such exons will be present in the mature mRNA. *Cis*-acting elements include exonic splicing enhancers (ESEs) and intronic splicing silencers (ISEs) those are bound by positive trans-acting factors e.g. serine/arginine-rich (SR) family proteins, depending on their location and on how they affect the selection of a splice site. Conversely, exonic splicing silencers (ESSs) and intronic splicing silencers (ISSs) are bound by negative trans-acting factors e.g. heterogenous nuclear ribonucleoproteins (hnRNPs). The hnRNPs are extremely conserved from nematodes to mammals (Chen *et al*, 2010; Webster, 2017; Wang *et al*, 2015). The interaction between the different components of the spliceosome specifically defines the alternative splicing of each exon (Pradas *et al*, 2014).

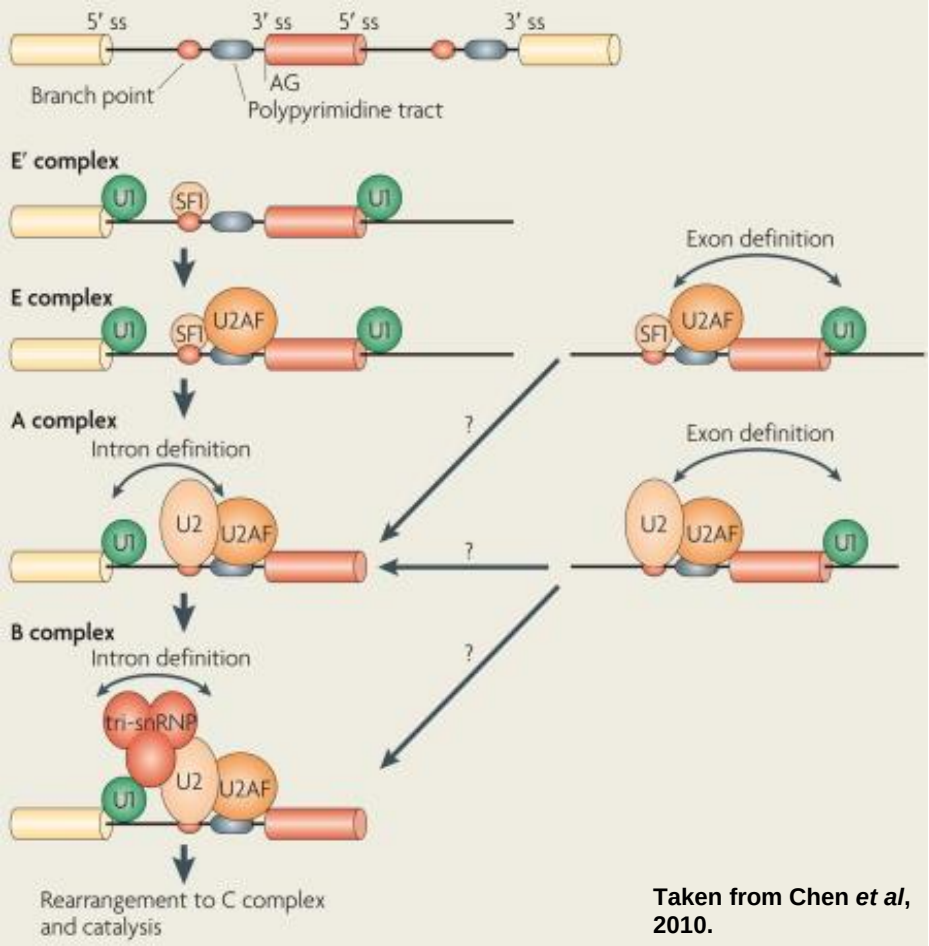


Figure 1.6: Molecular mechanisms of alternative splicing. The process begins with binding of U1 snRNA to the 5'splice site and the binding of splicing factor 1 (SF1) to the branch point by an ATP independent manner. The E' complex converts in E complex by binding of auxiliary factor U2AF to polypyrimidine tract. E complex is then converted into ATP-independent pre-spliceosome A complex by a replacement of SF1 by U2 snRNP at the branch point. Tri-snRNP U4/U6-U5 lead to the formation of B complex. This complex contained all spliceosome subunit necessary for pre-mRNA splicing. The loss of U1 and U4 snRNPs and conformational changes catalytically active spliceosome C complex is formed. Exons are defined by binding of U1 and U2 snRNP across the exon in the primary transcript followed (Webster, 2017).

1.5.3 Alternative splicing in ageing and disease

Alternative splicing is essential in cellular growth, differentiation and tissue development. Many studies have found links between the altered expression of splicing factors to be associated with disease. For example, cancer is known to be associated with changes in AS, and these changes have been shown to contribute to disease progression (Sveen *et al*, 2016). Data from mouse studies

have shown that overexpression of SR proteins, Srsf1, Srsf3 and Srsf10 transforms fibroblasts and increases tumour growth, and these proteins are also elevated in several human cancers. Besides the SR proteins, other RNA binding proteins e.g. hnRNPA1, hnRNPH, and PTB, which are hnRNPs, are overexpressed in cancer (Webster, 2017), suggesting that deregulated splicing of some splicing factors especially negative *trans*-acting factors (hnRNPs) or aberrant mRNA processing may associate with cancer in human and mice. By selecting a splice site more or less accessible to the spliceosome, hnRNPs regulate splicing reaction (Matlin *et al*, 2005).

The strict regulation of AS is essential for physiological and pathological conditions and deregulation of this process is responsible for different chronic diseases and ageing (Latorre and Harries, 2017). Lee *et al* (2016) have shown that changes in splicing factors' expression and regulation were associated with strain longevity in old and young mice in cellular model strain lifespan in spleen. They have also shown that splicing factors *HNRNPA1* and *HNRNPA2B1* were found to be associated in parental longevity in humans and changes isoform expression, specially hnRNP splicing inhibitor transcripts, helping in reducing cellular senescence and resuming the cellular proliferative capacity in long lived human model strains (Harries *et al*., 2011, Holly *et al*., 2013, Latorre *et al*., 2017, 2018), suggesting that splicing factor expression is strongly associated with age and cellular senescence, and correct regulation may increase lifespan in mice and humans.

There are several mechanisms, including AS of many cancer related genes, that caused the inactivation of oncogene suppression (Lamolle *et al*, 2006) or increase expression of some proteins resulting in cancer susceptibility or changed control over cell motility (Mazoyer *et al,*1998; Ghigna *et al*, 2005). A general loss of splicing fidelity changes the splicing profile which is the consequence of cancer-specific isoforms formation suggesting all these events may contribute to carcinogenesis (Brinkma, 2004). For example, alteration of SFs regulation changes the gene expression profile in non-malignant cells have been confirmed by next generation sequencing. Some SFs are directly related to

cancer progression and acted as cancer biomarkers e.g. SRSF1 is a prototype splicing factor because of its connection between AS and cancer.

As discussed above, normal regulation of SFs is hugely important for our normal physiology. Evidence suggests that dysregulated splicing is fundamental to cancer and neurodegeneration. Several spliced isoforms which are regulated by 29 SFs are found to be related to Alzheimer's disease (AD) where a reduction of splicing factors *PTBP2, RAVER1* and an increase of *ROD, PTBP1* are observed in brains from AD patients (Tollervev *et al*, 2011; Sato et al, 1999). SFs dysregulation also found to be reported in Parkinson's disease (Beyer and Ariza, 2013) suggesting changes in splicing regulation in the brain are responsible for some ageing related neurodegenerative diseases and these are associated with dysregulation of some negative *trans*-acting splicing factors (hnRNPs).

Great progress has been made by studying individual transcripts and, more recently, through genome-wide approaches, which together provide a picture of the mechanistic regulation of alternative pre-mRNA splicing. Changes in alternative splicing have also been observed in myotonic dystrophy where 25 novel aberrantly spliced exons were found (Yamashita *et al*, 2012), and in response to ionising radiation (Sprung *et al*, 2011). An exciting recent study in *C. elegans* showed that changes in splicing are associated with ageing, and that splicing factor1 (SFA1, Branch point binding protein) is essential for mediating the increased longevity associated with dietary restriction in this model organism (Heinz *et al*, 2017).

AS is controlled by the actions between *cis*-acting factors and *trans*-acting regulators which resume normal gene expression in our body and it will be interesting to see how splicing factors expression modifies e.g. the SFs expression changes or splicing regulators themselves in altered metabolic conditions in T2D.

1.6 The role of *TCF7L2* in diabetes

TCF7L2 variants are associated with type 2 diabetes in multiple ethnic groups and it is the genetic risk factor with largest effect size for T2D (Cauchi *et al*, 2006; Sale *et al*, 2007; Scott *et al*, 2006 and Grant *et al*, 2006). But the understanding of functional mechanisms is largely elusive. Evidence that attenuation of *TCF7L2* activity is associated with isoforms formation which caused impaired beta cell function and apoptosis and may be a mechanism of beta cell failure in T2D (Le Bacquer, *et al*, 2011). The association may be because of the effect on *TCF7L2* transcription, splicing or polyadenylation (Locke *et al*, 2011). Most of the alternative splicing events reported for this gene are involved in 3' splice site and research on the molecular diversity generated by AS may help to understand why this gene has emerged, since this gene has an association with T2D.

1.6.1 *TCF7L2* genomic variants are associated with the risk of T2D

TCF7L2 is well known as an important transcription factor in the canonical Wnt signalling pathway that affects a wide range of fundamental cellular process e.g. embryonic development, stem cell maintenance, cell fate, cell proliferation, cell migration, tumour suppression, and oncogenesis. In β cell, it also specifically plays a role in the regulation of insulin processing, secretion and β cell survival (Hanson *et al*, 2010).

The TCF*7L2* single nucleotide polymorphism (rs7903146) associates with T2D is located in intron 4 towards the 5' end of the gene (Grant *el al*, 2006). Since this is within a non-coding region, the molecular mechanism of this association is not yet clear. It has been shown that expression of the risk allele is associated with impaired conversion of proinsulin to insulin (Loss *et al*, 2007 and Hansson *et al*, 2010). However, other mechanisms have also been proposed. The risk allele is also associated with reduced insulin content and impaired glucose stimulated insulin secretion, as well as a reduction in glucagon-like peptide 1 (GLP-1; the incretin hormone main role in glucose homeostasis) (Hanson *et al*, 2010, Cropano *et al*, 2017, Pradas *et al*, 2014; Le Bacquer, *et al*, 2011). Locke *et al* (2011) also identified a novel polyadenylation signal within intron 4 of *TCF7L2*

that results in isoforms production which inhibits the full length *TCF7L2* isoforms activity, full form *TCF7L2* is involved in the pathogenesis of T2D and the relative use of this polyadenylation site in human tissues. The risk allele may therefore result in alternative splicing of *TCF7L2*, resulting in production of an inhibitory truncated isoforms inhibit *TCF7L2*-dependent target genes by sequestering β-catenin.

1.6.2 Expression of *TCF7L2* isoforms in T2D

A large-scale and complex splicing pattern of *TCF7L2* AS was shown by Duval *et al*, (2000) and Hansson *et al*, (2010). The splicing pattern is different in different tissues (Hansson *et al*, 2010). Expression of the high-risk allele at the intronic rs7903146 has been shown to be associated with higher *TCF7L2* mRNA expression (Lyssenko cited in Hansson *et al*, 2010), but lower protein expression, in donor (T2D) islets (Cauchi and Froguel cited in Hansson *et al*, 2010, Le Bacquer *et al*, 2011). As shown in Figure 1.8, there are two main regions that are alternatively spliced in β cells: (i) cassette exon 4, and (ii) cassette exons 13-16.

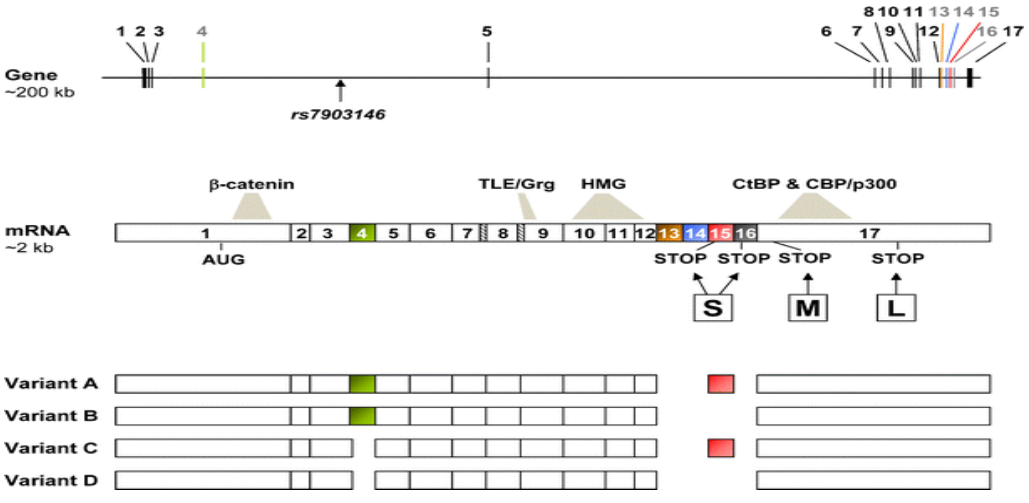


Image taken from Hansson et al, 2010

Figure 1.7: Alternative splicing of *TCF7L2* analysis in human islets. Four *TCF7L2* splice variants (A, B, C and D) are expressed in human pancreatic β cells.

As discussed above, AS increases the functionally different isoforms. A recent study by Kaminska *et al*, (2012) has shown that a complex pattern of *TCF7L2* is regulated by weight loss and associated with plasma glucose and serum FFAs along with obesity. The expression of a short mRNA variant, lacking exons 13, 14 and 16, is associated with hyperglycaemia and impaired insulin action in adipose tissue which was more common in subcutaneous fat in T2D patients. The association between *TCF7L2* splicing was independent of the *TCF7L2* genotype. The reason may be the impaired ability of this short mRNA variant to activate Wnt signalling in adipose tissue which, according to these researchers, induced beta cell apoptosis. They have shown that weight loss was associated with decreased expression of this short variants in adipose tissue in obese individuals. These findings may be important for in subjects with obesity which leading to increased diabetes risk.

AS of exons 13-17 results in altered reading frames and can lead to transcription termination due to the introduction of stop codons (Kaminska *et al*, 2012; Hansson *et al*, 2010). The differential splicing results in several isoforms competing for promoter binding and thus shows different activation or repression properties (Weise *et al*, in Hansson, 2010). C-terminus is necessary for efficient *TCF7L2*/β-catenin activation of the target promoter (Weise *et al* in Hansson, 2010). For example, the corepressor C-terminal binding protein (CtBP), an inhibitor of Wnt signalling pathway, and p300 are recognised by *TCF7L2* long isoform (Duval cited in Hansson *et al*, 2010), suggesting the role of long isoform in beta cell survival. The sites CtBP interacts with are encoded by exon 17 and thus it inhibits *TCF7L2* mediated expression, depending on the predicted stop codons are being used (Cuilliere-Dartigues cited in Hansson *et al*, 2010).

Another study by Weise *et al* (2010), had shown that, in human, short transcript variants lacking the exons 13, 14 and 16 decrease the activation of β-catenin target gene promoters in Wnt signalling pathway in opposite to full length transcripts, isoform having exons 13, 14 and 16. Therefore, expression of short

variants lead to switch off the a chain of signalling events and thus function as inhibitor of Wnt signalling pathway.

Therefore, AS in this 3' region is crucial to maintain normal function of *TCF7L2* and alteration in splicing may lead to formation of isoforms which can lead to termination of *TCF7L2* function and thus can impair β cell function, and lead to apoptosis.

1.7 The wider role of alternative splicing in β cells

Alternative splicing is a highly tissue-specific process, and only few studies have examined AS in β cells. Interestingly, one study has shown that a number of splicing regulators important in neurons are also expressed in β cells and regulate β cell survival and function (Mateu *et al*, 2017), specifically four RBPs (*Elavl4*, *Rbfox1*, *Nova2* and *Rbfox2)*.

Knockdown of *Rbfox1* using siRNA was shown to induce insulin secretion by a mechanism associated with alternative splicing of the actin remodelling protein gelsolin (*Gsn*) and the voltage-gated calcium channel 1C (*Cacna1c*). Splicing regulation of insulin receptor (*Insr*) was also modulated by *Rbfox1* knockdown (Mateu *et al*, 2017). Another splicing regulator *Nova1* has been reported to be a master regulator of AS in pancreatic β cells (Villate *et al*, 2014). It is expressed and regulates insulin secretion and apoptosis in β cells. Splicing factor *Rbfox1* controls the AS events of key genes (*Gsn*, *Cacna1c* and *Insr*) and which are involved in insulin secretion, gene expression, cell division and cell death (Mateu *et al*, Wong King Yuen *et al*, 2018). The AS events of these tissue specific splicing regulators are needing to explore in altered metabolic conditions to see their impact on β cells.

1.8 Aims

We hypothesized that the metabolic conditions associated with type 2 diabetes may affect the regulation of alternative splicing. Although there is progress in understanding the molecular mechanisms in the pathogenesis of diabetes, little is currently known about the effect of metabolic changes that exist in many people with undiagnosed or uncontrolled diabetes on splicing in β cells, and how this may be linked to the decline in β cell mass and function. Therefore, the aim of this study was to determine the effects of metabolic conditions associated with T2D, specifically high levels of glucose and free fatty acids, on the regulation of alternative splicing. The specific aims of this project were:

1. To determine the effect of high glucose and fatty acid conditions (glucolipotoxicity, GLT) on the expression of a panel of genes that regulate alternative splicing, using the rat INS1 cell line.

2. Analyse the effect of GLT on alternative splicing of specific candidate genes e.g. *Tcf7l2, Gsn and Cacna1c*.

3. To examine global changes in exon usage and splice junctions under GLT conditions in β cells using the Clariom D exon arrays.

Chapter 2- Materials and Methods

2.1 Cell line and cell culture

2.1.1 INS1 Cell line

To study the effect of metabolic conditions on pancreatic β cells, INS1 cells was used, which are derived from a rat insulinoma that arose following irradiation (Asfari *et al*, 1992). INS1 cells were kindly donated by Dr Charlotte Edling of the Blizard Institute of Cell and Molecular Biology at Queen Mary University of London.

After passage number 15, INS1 cells were not used for our experiments as they are very sensitive to environment and susceptible for mutation. There is a correlation between beta cell lines and phosphorylation of hexokinase activity which is key to glucose sensing and accurate insulin secretion. In cell culture, low Km value is observed for this glucose phosphorylating enzyme hexokinase, which is a deviation from normal phenotype that often β cell lines sustain during the first 20 passage in culture (Efrat, 1999). Therefore, cells were discarded after 3 passages when they were received (passage number 12 was received) for our experiments. Mostly Passage number 13 was used for our GLT treatment.

2.1.2 Cell culture media

INS1 cells were maintained in RPMI 1640 medium (Sigma-R8758 series) containing 11.1mmol/l D-glucose supplemented with L-glutamine (Fisher Scientific), 10 mM HEPES buffer (Sigma), 1 mM Sodium-pyruvate (Sigma), 100µg/µl penicillin-streptomycin (Fisher Scientific) and 0.05 mM 2-mercaptoethanol, and 10% Fetal Bovine Serum albumin (Gibco-10270-106; South American origin). Serum free medium was of the same formulation except 10% Fetal Bovine Serum (FBS) was added.

2.1.3 Cell storage and maintenance

For long term storage, INS1 cells were stored at the vapour phase of liquid nitrogen (higher racks) at temperature between -140°C and -180°C. To remove cells from liquid nitrogen, the vial was first sprayed with 70% ethanol and cells were thawed until a small piece of frozen medium remained (2-3 mins). The cell suspension was transferred very slowly or dropwise into a filter capped T75 flask (Thermo Fisher Scientific) with 25-28 ml of complete medium. The medium was replaced with fresh complete medium the following day. Cells were cultured in a T75 flask and incubated at 37°C in a 5% CO_2 atmosphere.

2.1.4 Cell sub-culturing

Cell lines were passaged when cells were 75-80% confluent. Growth medium was aspirated and discarded. Cells were briefly washed in five millilitre (ml) of sterile phosphate buffered saline solution (PBS), three ml of (1x) trypsin-EDTA solution (Sigma) was added and then incubated for approximately 4 minutes at 37°C in 5% CO_2 to detach cells, a process known as trypsinization. Cell detachment was observed using a bright field inverted microscope (Motic AE31). Seven ml complete medium was added to the flask and gently pipetted up and down a few times. An aliquot of three ml cell suspension was moved into a new T75 flask containing 24-28 ml medium.

2.1.5 Cryo-preservation of INS1 cells

When INS1 cells were 75-85% confluent they were used for preservation. Growth medium was aspirated, and cells were trypsinized as described above. Cells were resuspended in ten ml complete medium and centrifuged at 1000 rpm for 3 minutes. Cell pellets were resuspended in one ml of complete media. According to cell count, freezing media (3 ml of serum free RPMI 1640, 2 ml of 10% FBS, 100 µg/ml Penicillin-Streptomycin) was filter sterilised (0.2 µm) and added to cells. All solutions using for cell preservation were stored on ice. A 20% DMSO

solution (100% DMSO-Sigma Aldrich, UK), was added dropwise to cell suspension and the total volume transferred in cryovials. Again, serum free RPMI 1640 was used to make the DMSO solution. Cells were stored in freezer (-80°C) for at least 24h before being transferred in liquid nitrogen for long-term storage.

2.1.6 Cell counting and seeding density

As described in the above section 2.1.4, after trypsinisation ten ml of cell suspension was transferred to a 20 ml universal tube. After that the cell suspension was centrifuged for 3 minutes at 1000 rpm to pellet the cells. The medium as aspirated carefully without disrupting the cell pellet. Once removed, the pellet was disrupted by gently tapping the side of the tube and afterward cells were resuspended in 1 ml of fresh medium. Ten microliter (µl) of cell suspension was pipetted into an Eppendorf for cell counting. Trypan blue was used to distinguish viable cells from dead cells and served as a diluent. Approximately ten µl of the cell-trypan suspension were inserted into a chamber of a disposable haemocytometer (#BSV100H, Immune Systems) for cell counting and the concentration of cell per ml was calculated using the following formula:

Cells/ml = Average no of cells per large square x Dilution factor x 10^4

To achieve a growth curve, INS1 cells were seeded for 24 hours at density of 30,000 cells/well in a 96 (VWR) well plate, as described before (Viloria *et al*, 2016). The culture media was aspirated and replaced with freshly prepared GLT and control media to appropriate wells of the 96 well plate. The plate then placed into the IncuCyte (Essen Bioscience) live cell analysis system and allowed the plate to warm to 37 °C for 30 minutes prior to scanning. Cell growth was monitored using the Imaging system to determine cell area or cell confluency over the indicated time period, as describe in section 2.1.9.

2.1.7 GLT media

To mimic *in vivo* diabetic extracellular glucolipotoxic conditions, INS1 cells were treated with 27mM glucose, 200 µM oleic acid (C18:1) and 200 µM palmitic acid (C16:0) in RPMI media (which contains 11mM glucose before supplementation). This is referred to as GLT treatment. *In vitro* exposure of high FFAs caused cellular dysfunction and death (Cnop, 2008). Although oleate can abrogate the effects of palmitate, it is more physiological to deliver a high FFAs solution which consists of palmitic acid and stearic acid as those two are the most common fatty acids in human. The experimental condition was adopted from Bagnati *et al*, 2016. The GLT media was prepared on the day of use and filtered using 0.2µm filter prior to cell treatment.

2.1.8 Free Fatty acid (FFA) conjugation

Long chain fatty acids are important substrate for energy production, among them most common fatty acids in humans are the long-chain saturated fatty acid palmitate (C16:0) and the monounsaturated oleate (C18:1). The low solubility of long-chain fatty acids in aqueous solutions is one of the major limitations for *in vitro* and *in vivo* studies. To overcome this problem, FFAs were conjugated to albumin which allowed the preparation of solutions in the physiological concentration range (Oliveira *et al*, 2015).

Palmitic acid (#P0500-10G; Sigma) was dissolved (100 mM) in absolute ethanol and oleic acid (#07501-250MG; Sigma) was dissolved (100 mM) in 50% ethanol (absolute ethanol was dissolved in distilled water in 1:1 ratio). To achieve solubility, vortexing was required for both solutions. Both serum free and complete media were supplemented with two percent BSA (#A7030-10G; Sigma) to facilitate free fatty acids conjugation. Despite of 11 mM glucose in RPMI media, extra 16 mM glucose and fatty acids were added, so the total glucose was 27 mM. The media was filtered (0.2 µm) and the solution was incubated for 1 hour at 37°C which facilitated free fatty acid conjugation with albumin. Cells were

subsequently treated with GLT media in both complete and serum free conditions. Controls were made with the corresponding amounts of ethanol and BSA.

Glucolipotoxic media was prepared fresh for each day of use although the FFA-albumin complexes can be prepared in large amounts and preserved for several years at -20°C.

2.1.9 Image analysis using IncuCyte Zoom

IncuCyte Zoom 2015A and 2016A software was used for the growth analysis in the following procedure:

1) Open Schedule Scan from the Task list > click on the position- add your plate-choose vessel type.

2) New processing definition—Training image collection- change ‘segmentation adjustment’ bar to separate cells from background -select ‘area (um^2)’ under ‘filters’ and add a minimum value to exclude debris (250 um^2 was used in this study)-select ‘preview’- adjust further parameters if required to increase accuracy- select ‘preview’ again- ‘scratch wound mask’ and ‘confluence mask’ which will help visualize the experiment during this stage but will not change processing parameters. Once appropriate parameters have been chosen then save processing definition-File- ‘save as’- give a name. This can be used for further experiments and it is a requirement of the machine that one is selected to run the assay.

3) Scan pattern-Choose scan pattern from the list -job type- ‘Basic Analyser’

4) Select ‘Properties Tab’-put experiment name-cell type-plate map-passage no-then ‘APPLY’.

5) Launch Analysis Job-select ‘Scanned Vessel’ from Task List-choose experiment-job type- basic analyse-select processing definition-define time points you wish to use-select the wells- click ‘lock’.

6) Exporting Results-Choose your analysis result from Task List-select ‘Graph/Export’-choose ‘Metric’ you wish to export- select ‘Group’-select ‘None’- you can see the microplate graph and print that out-click ‘Export’ which will export a txt file that can open in excel.

7) To export images, select ‘image type’-Export- it will export images.

8) To export representative images- select’ utilities tab’- Export single image-select ‘phase contrast ‘select time, date and bar stamps-select ‘continue’ begin export.

2.2 RNA preparation and analysis

2.2.1 RNA extraction

For GLT treatment, INS1 cells were cultured at 30,000 cells/100µl in 6 well plate and allowed to adhere 24 hours, before treatment with GLT or control media for 46 hours. For GLT experiment, four biological replicates were there with three technical replicates.

RNA extraction was performed using RNeasy columns (Qiagen). Briefly, cells were lysed with the lysis buffer (RLT) and vortexed. An equal volume of 70% ethanol was added to the homogenized lysate and mixed well by pipetting and transferred to the RNeasy spin column. The solution was centrifuged at 10,000 rpm for 15 seconds. To remove genomic DNA from RNA preparations DNase digestion was carried out by using the DNase I in RDD buffer according to the protocol. A series of washes followed as an RNA purification process and each time the flow-through had been discarded. Finally, the purified RNA was eluted

from the spin column in 30 µl of RNase free water. RNA integrity and quantity were assessed using an Agilent Bioanalyzer or Nanovue spectrophotometer (discussed in detail in results chapter 4). For long term storage, RNA was stored in -80°C.

Spectrophotometry provides both individual absorbance values and absorbance ratios. RNA concentration is measured at absorbance of 260 nm where the light absorbed by a substance is directly proportional to the concentration of the substance (Beer-Lambert Law). RNA absorbs light at 260 nm whereas protein and ethanol absorb at 280 and 230 nm respectively. A ratio of 260/280 and 260/230 is an indicative of RNA purity, a range in between 1.9-2.1 demonstrates that the RNA isolated was of high quality and purity. All RNA samples used in this experiment fell into this range.

2.2.2 First strand cDNA synthesis

cDNA synthesis was performed using RevertAid First Strand cDNA synthesis kit (Thermo Scientific) with 500ng of total RNA. The reaction (20 µl reaction) set up with reverse transcriptase (RT), no reverse transcriptase control (NRT) and no template control (NTC) is shown in Table2.1. All reaction tubes were incubated on ice while performing the reaction. Table 2.2 shows the PCR cycling conditions for reverse transcription. The oligo d(T) method was used, with total RNA as a template.

Table 2.1: PCR reaction for reverse transcription.

Reagents	**RT (master mix)**	**NRT (master mix)**	**NTC (master mix)**
Total RNA	500 ng	500 ng	Not applicable
Oligo $(dt)_{18}$ primer	0.3 µM	0.3 µM	0.3 µM
Water, nuclease-free	To 12 µl	To 12 µl	To 12 µl
5x Reaction Buffer	4 µl	4 µl	4 µl
RiboLock RNase inhibitor (20 U/ µl)	1 µl	1 µl	1 µl
10 mM dNTP Mix	2 µl	2 µl	2 µl
RevertAid M-MuLV RT (200 U/ µl)	1 µl	Not applicable	1 µl

2.2.3 RT- PCR

Polymerase Chain Reaction is a technique that used to amplify a specific segment of DNA from a small amount of DNA template. Dream Taq™ (Thermo Fisher Scientific) was used and is a ready to use master mix buffer supplemented with $MgCl_2$, dNTPs, two tracking dyes and a density reagent that allows direct loading of the PCR product on a gel. A thermal cycler is to maintain a series of temperature for a set amount of time (Table 2.2).

Table 2.2: Standard PCR Cycling Condition.

Initial Denaturation	Stage 1 (1 Cycle)	95°C for 2 mins
Denaturation Annealing Extension	Stage 2 (35 Cycles)	95°C for 30 secs 55°C for 30 secs 72° for 30 secs
Final Extension	Stage 3 (1 Cycle)	72°C for 5 mins

For any RT-PCR reaction, 5 ng/µl concentration of INS1 cDNA was used (Table 2.3). The reaction volume was 20 µl and a thermocycler (Applied Biosystems) was used for cDNA amplification.

Table 2.3: RT- PCR master mix.

Reagent	X1 Reaction
Dream Taq 2x	1x
cDNA concentration	5 ng/µl
Forward Primer	0.3 µM
Reverse Primer	0.3 µM

2.3 Primers Selection, qPCR method validation and data analysis

2.3.1 Designing Primers

Exon-spanning primers were designed using National Centre for Biotechnology Information (NCBI) BLAST software (https://www.ncbi.nlm.nih.gov/tools/primer-blast/). Whilst designing the primers, extra care was taken to avoid primer-dimers, hairpin loop formation and they were all designed to anneal at 55°C.

When designing primers, amplicon size was restricted to be 100 to 170 bp, because shorter amplicons are amplified with higher efficiency, but also to keep amplicon size at least 100 bp which helped to distinguish it from primer-dimers that might form. Other factors that were considered where to maintain GC content of 45-60%, a melting temperature (Tm) between 58°C and 62°C, to keep the primers' length not more than 23 bp and to ensure to design primers in a way which will facilitate to reduce 3' complementary to avoid primer-dimer formation.

2.3.2 Gene expression analysis by quantitative RT- PCR

The use of fluorescent DNA-binding dyes to detect and quantitate a PCR product allows quantitative PCR to be performed in real time. Specifically, designed instruments perform both thermal cycling to amplify the target and fluorescence detection to monitor PCR product amplification. A SYBR green based reverse transcriptase qPCR reactions (qRT-PCR) were performed to validate primers and to assess gene expression analysis according to MIQE guidelines (Bustin, 2005). The qRT-PCR was performed in duplicate using PrimePro48 (TECHNE) and the thermal profile is shown in Table 2.4.

At the early stage of PCR cycle very little amount of florescence is generated which is considered as noise. As the fluorescence approaches the optical detection level, the exponential phase, which is the region where the Cq is determined. Cq is the phase where the fluorescence signal crosses the detection threshold level and is used for quantification. The fluorescent dye binds to the double stranded DNA and the fluorescence is directly proportional to the DNA concentration. Once the reaction components begin to be consumed, the reaction reaches the plateau stage. The higher the initial amount of sample DNA the lower the Cq values as the product is detected soon. The principal of this method is based on fluorescent chemistries, an optical system that can capture the emitted fluorescence at every PCR cycle, and the software which is able to quantify the PCR amplified product.

Table 2.4: Thermal Profile of qRT-PCR reaction.

Set up the thermal profile	Cycling Temperature (°C)	Cycling Time
Denaturation and Polymerase activation	95	10 mins
PCR cycling	95	15 seconds
	56	30 seconds
	72	30 seconds
Melt curve	95	15 seconds
	55	15 seconds
	95	15 seconds

A melt curve analysis was carried out at the end of the qPCR reaction which assess the primers' quality and specificity. PCR products were also analysed by agarose gel electrophoresis during validation to confirm the presence of a single product. Primer efficiencies were determined for all primer pairs.

To determine the primer efficiency, a standard curve approach was used where a serial dilution (10-fold dilution) of cDNA template of known concentration (5 ng/ul) was used to create a four/three-point standard curve. A number of rat housekeeping genes (TATA Biocentre) for example, Tyrosine 3//tryptophan 5-monooxygenase activation protein (*Ywhaz)* and peptitylprolyl isomerase A (*ppia)* were also tested.

Amplification efficiency, E, is calculated from the slope of the standard curve using the following formula:

E=10-1/slope

% efficiency was determined by using the formula below:

%Efficiency = (E-1) x 100%

Solutions/dilutions for calibration curves were made on the day of analysis and then discarded.

2.3.3 Data analysis by relative quantification method

Messenger RNA concentrations were measured of target genes relative to a reference gene. In the relative quantification method, one sample is usually chosen as the calibrator or control and the expression of the target gene in all other samples is expressed as an increase or decrease relative to the control or calibrator. In our experiment, GLT treated INS1 cells are test samples which calibrated against *Ppia* in comparison to untreated samples or controls. The results (RQ) obtained were expressed as the fold change and the relative amount is known as the 'fold difference'. Relative gene expression was determined using the ΔΔCq method (Livak *et al*, 2001).

The fold change relative to reference gene was calculated as follows:

ΔCq = Cq target gene – Cq reference gene

- ΔΔCq = ΔCq sample- ΔCq of control 1

Fold changes to control 1 $=2^{-\Delta\Delta Cq}$

2.4 Microarray analysis

Hybridisation using Affymetrix rat Clariom D arrays (see Figure 2.1 below for chip characteristics) was performed at the Genomics Centre, King's College London. Two control- and two GLT-treated samples were analysed (1 control- and 1 GLT-treated samples from two independent experiments). Analysis was performed using the Transcriptome Analysis Console (TAC) 4.0.1 software (ThermoFisher Scientific) to analyse and visualise global expression patterns of genes, exons, pathways and alternative splicing events. CEL files were provided by the Genomics facility, and SST-RMA normalisation was performed in order to overcome the historical problem that array data tends to underestimate fold changes compared to RT-PCR. Both gene level and exon level analysis were performed but focused primarily on the exon analysis.

To identify enriched pathways and functions of the differentially expressed genes PANTHER gene list analysis (version 13.1), a large-scale gene function analysis was used. We uploaded differentially expressed genes from our exon array analysis (Gene+ exon) and got functional classification with gene ontology (GO) which allowed us the identification of molecular functions, biological process and cellular components among that list of genes provided. The following link was used: http: //www.pantherdb.org/data/, for the identification of enriched pathways.

Content summary	Human	Mouse	Rat
Genes[1]	>134,700	>66,100	>68,900
Transcripts[1]	>542,500	>214,900	>495,200
Exons[1]	>948,300	>498,500	>320,400
Exon-exon splice junctions[1]	>484,900	>282,500	>293,700
Total probes[1]	>6,765,500	>6,022,300	>5,946,400
Probes targeting exons[1]	>4,781,200	>4,895,600	>4,780,700
Probes targeting exon-exon splice junctions[1]	>1,984,300	>1,126,700	>1,165,700
Probe length (bases)	25	25	25
Probe feature size	5 µm	5 µm	5 µm
Background probes	Antigenomic set	Antigenomic set	Antigenomic set

1. Numbers are representative annotation as of April, 2016. All numbers have been rounded down to the nearest hundred.

Figure 2.1: Characteristics of Affymetrix rat Clariom D arrays. Clariom D solution is a GeneChip™ for the next generation transcriptome level providing tools or instrument system, providing view of the transcriptome for a faster path to biomarker discovery, generate robust expression profiles from total RNA or cells. The chips confidently detect genes, exons and alternative splicing events that give rise to coding and long non-coding isoforms. In array specifications, numbers are representative of annotations and antigenomic background control contains the antigenomic background probes that can be used to estimate background expression level. Specification information is collected from the data sheet provided on Thermo Fisher Scientific website.

2.5 Gel purification of RT-PCR products for Sanger sequencing

To purify a specific RT-PCR product (e.g. where there were multiple variants giving different molecular weight PCR products), and to remove residual dNTP/primers, the QIAquick gel purification method was used.

Gel purification was carried out as the following procedure, using transilluminator and with proper UV protection, especially for eyes, DNA fragments were cut from the gel and collected in an empty pre-weighed Eppendorf tube. After weighing out the gel, three volumes of buffer QG added to one volume gel (100 mg gel =

100 µl). The maximum amount of gel per spin column is 400 mg. The gel was dissolved by vortex and with the help of temperature (50°C). A yellow colour was the indication of gel slice has dissolved completely. One gel volume of isopropanol was added to the sample and placed the sample in a QIAquick spin column on a two ml collection tube. To bind DNA, the sample applied to the column and centrifuged for 1 minute at 12,000 rpm. The flow-through was discarded. At this stage DNA was bound to silica-gel membrane which was then thoroughly washed with the ethanol containing PE buffer to remove any remaining contaminants and centrifuged for 1 minute at 12,000 rpm, the flow through was discarded. Fifty ml of Buffer EB (low concentration of Tris buffer) was added to the centre of the column and centrifuged at 12,000 rpm for 1 min and collected the yield in a sterile collection tube. DNA quantity and quality were then assessed by Nanovue spectrophotometer as described above (section 2.2.1).

2.5.1 Sanger sequencing and analysis

The purified DNA bands were sent to Beckman Coulter Genomics (GENEWIZ) for Sanger sequencing using both forward and reverse primers (as used for RT-PCR amplification). To analyse gene sequence, it was determined:

- The composite sequence from the forward and reverse reactions
- Nucleotide blast search to determine sequence identity
- The location of specific exons in the sequence

FinchTV 1.4 was used to examine the sequence chromatograms and perform a quality check. For reverse primer sequences, an online reverse complementation tool was used (‘reverse-complement.com’). Alignment of the forward and reverse sequences, or against database sequences, was done using the pairwise alignment software Clustalo online. In many cases clustalo was not able to account for gaps due to missing exons, and exon identification was performed in a semi-manual fashion by aligning one exon at a time.

2.5.2 Database

The database programmes used for sequence analysis were as followed:

- To open DNA chromatogram files, analyse and manipulate the data from Sanger DNA sequencing, the desktop application tool Softpedia was used by the following link: http://www.softpedia.com/get/Science-CAD/FinchTV.shtml
- To perform BLAST searches, the programme 'NCBI Blast Toolkit' was used which compares nucleotide sequence to sequence database and calculate the statistical significance of matches as well as to help identify members of gene families the weblink was used are as follows: https://blast.ncbi.nlm.nih.gov/Blast.cgi
- For multiple sequence alignment, the program 'Clustao' was used and the link is: https://www.ebi.ac.uk/Tools/msa/clustalo/

2.6 Statistical analysis

Microsoft Excel 2016 spreadsheet was used to calculate mean, fold change, standard deviation and standard error of the mean. Two sample t-tests or paired t-tests were used to calculate statistical significance. A p value ≤ 0.05 was considered significant.

Chapter 3 - The Effect of Glucolipotoxic Conditions on β Cell Viability

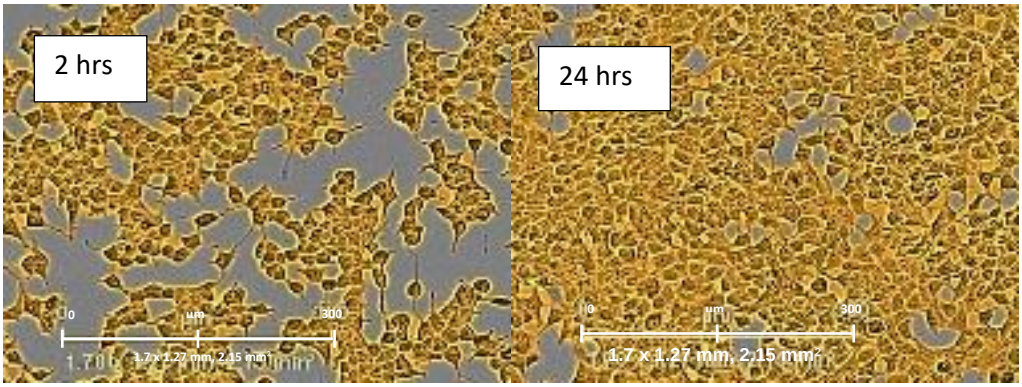
In order to mimic the metabolic conditions that exist in many people with type 2 diabetes and evaluate the effect on the regulation of splicing in pancreatic β cells, we first established and validated an *in vitro* model of β cell glucolipotoxicity, based on published protocols (Bagnati *et al*, 2016). To study glucolipotoxicity on the pancreatic β cell, we used rat insulinoma cells (INS1), which are pancreatic β cells immortalised by radiation (Asfari *et al*, 1992).

3.1 Validation of INS1 β cell growth characteristics

Understanding cell seeding variation and its effects on proliferation is important for controlling assay variability. We therefore performed initial experiments to validate growth and seeding densities using the IncuCyte live cell imaging system (Method section 2.1.9).

INS1 β cells were plated at a density of 30,000 cells/well in 96 well plates in complete (10% FBS) media, and 24 hours growth was monitored as described in Methods section 2.1.6. As shown in Figure 1.1A, the Incucyte parameters chosen correctly identify cell area at both low confluency (2 hrs) and high confluency (24 hrs). Cell growth was seen to increase over time (Figure 1.1B), demonstrating the expected pattern of β cell growth including lag phase and exponential cell growth. Due to the short timeframe, the plateau phase was not reached in this experiment. However, the cells appeared to be healthy and exhibited the expected growth pattern, validating the growth of INS1 cells in my hands.

A)



B)

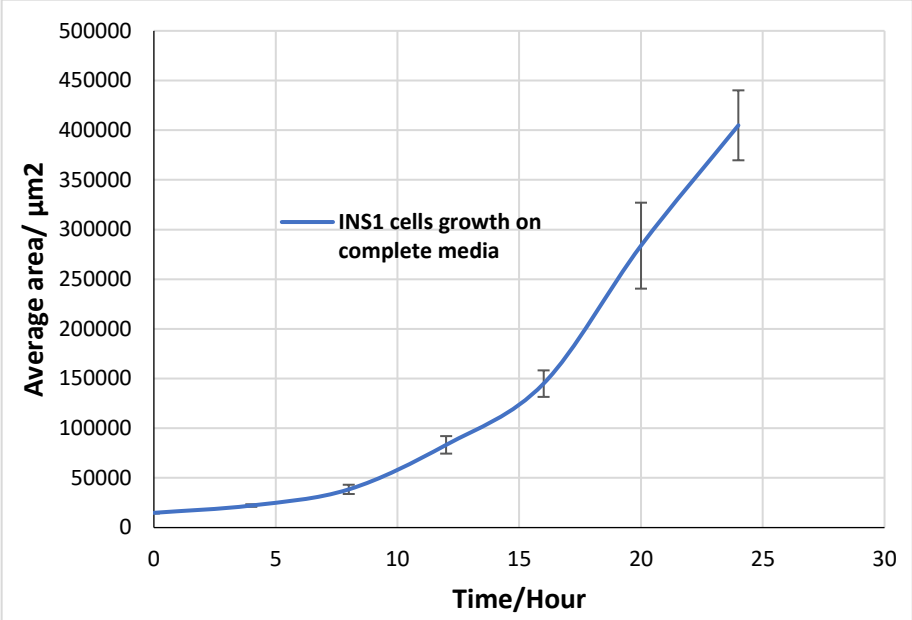


Figure 3.1: Validation of β cell growth characteristics under normal conditions. A) Representative phase IncuCyte images of INS1 cells showing confluence mask correctly quantifies cell area at both low and high cell density. **B)** Graph shows mean cell area +/- SEM, n=3 technical replicates from a single validation experiment.

3.2 Validation of INS1 cell growth in the presence of BSA supplementation

The *in vitro* glucolipotoxicity protocol used in later experiments is based on the method reported by Oliveira *et al*, 2015, which involves the use of free fatty acids bound to albumin to resolve the problem of low free fatty acid solubility. Since supplementation with two percent bovine serum albumin will be used for fatty acid

conjugation in later experiments, the effect of this supplementation on β cell growth was therefore first validated.

INS1 cells were plated as above then, after allowing the cells to adhere for 24 hours, the media was replaced with fresh media supplemented with or without two percent BSA. Both serum free and complete media was used for this experiment. INS1 cell growth was monitored by live cell imaging for 35 hours.

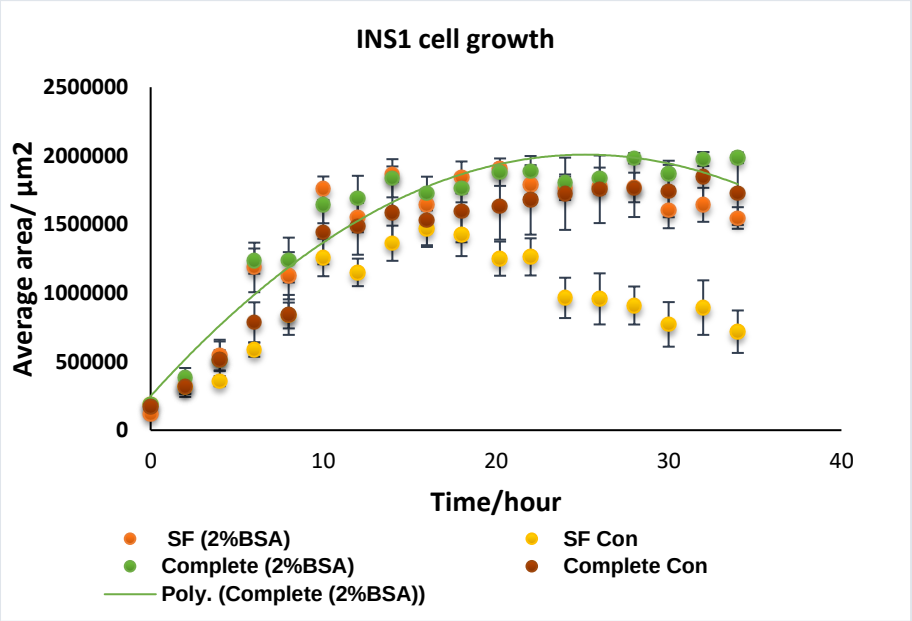


FIGURE 3.2: β cell growth under conditions of BSA supplementation. INS1 cells were cultured in either complete or serum free media (SF) and supplemented with with/without two percent BSA. Cells were incubated without 2% BSA in both complete and serum free media served as 'control' in this experiment in order to observe the effect of extra 2% BSA on INS1 cell growth. Graph shows mean +/- SEM (n=6 technical replicates from a single experiment). A polynomial trending curve has drawn for 2% BSA complete (poly. (Complete (2%BSA)) with a R^2 value is 0.925 which indicates a good fit of the line to data means less fluctuation of data with a smooth curve.

Both cell confluency and cell area were used to evaluate the growth pattern of INS1 cells in two percent BSA conditions and gave similar results. As shown in Figure 3.2, the addition of two percent BSA had no effect on cell growth in complete media, with both control and BSA treated cells demonstrating similar lag, growth and plateau stages. In serum free conditions, a decrease in cell number was observed at around 24 hours, presumably due to the lack of essential nutrients and growth factors provided by serum. The presence of two percent BSA prevented this decrease, and cell growth in the presence of BSA and absence of serum was similar to growth in complete media. Subsequent

experiments will use complete media, and the presence of BSA supplementation does not affect cell growth under these conditions.

3.3 The Effect of glucolipotoxic treatment on β cell growth

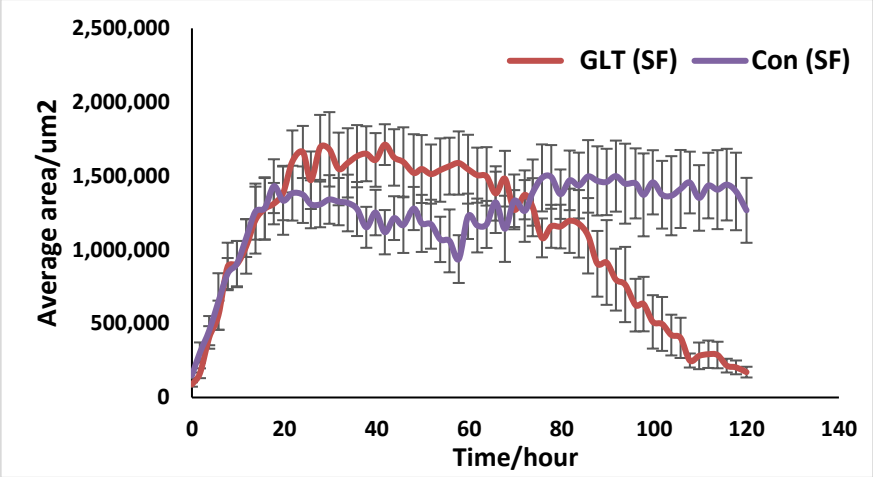
To validate GLT treatment, INS1 cells were treated with GLT media (Method section 2.1.7) to see the effect of a high concentration of sugar and fatty acids on β cell growth.

For GLT treatment, INS1 cells were incubated for five days in the presence of 200 µM palmitic acid, 200 µM oleic acid and 27mM glucose (Bagnati *et al*, 2016). Both serum free and complete media conditions were used initially to assess the effect of GLT treatment on β cell growth. The control media contained the same BSA/ethanol concentration but lacked free fatty acids (see methods for further details). Standard RPMI media contains 11 mM glucose which is euglycemic in human. The normal blood sugar level in human is between 4.0 to 5.5 mmol/L which represents a species difference between rodent and human. However, we supplemented the media with an additional 16 mM glucose (Bagnati *et al*, 2016) to create glucose toxicity in INS1 cells. INS1 cell growth was then monitored using the IncuCyte for five days following GLT or control treatment and images were taken each two-hour interval.

As shown in Figure 3.3, in serum-free conditions, GLT treatment initially offers some growth advantage, perhaps due to the high glucose levels, since glucose is known to induce β cell proliferation (Stamateris *et al*, 2016). However, after 72 hours, cell growth under GLT conditions was decreased dramatically in comparison to control and from this time point cells start to change their morphology, assuming a rounded shape and reducing adherence to the plate. A two-sample independent t-test was carried out as a statistical end point analysis (120 hours) and p=0.2 value in the Kolmogorov-Smirnov test (K-S) confirmed the normal data distribution which is prerequisite to perform a t-test. The end point,

analysis has shown that a significant decrease ($p=0.0006$) in cell number is observed in GLT treated cells due to the toxicity experienced by the cells.

A)



B)

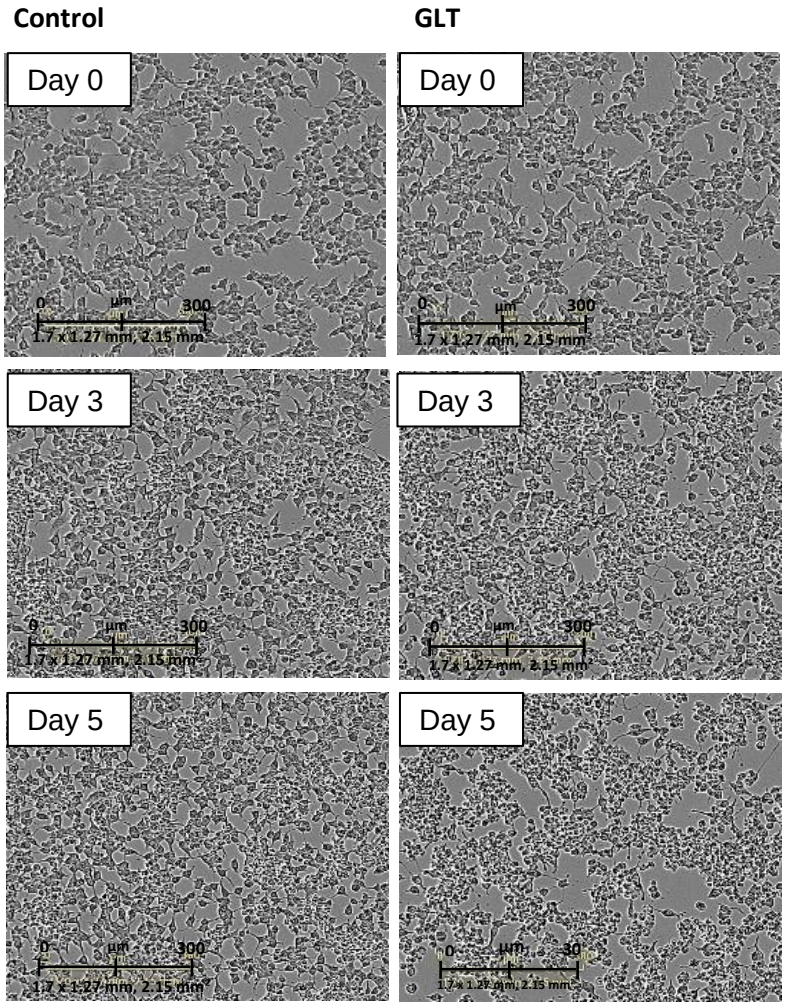


FIGURE 3.3: A) The effect of GLT treatment on cell growth (serum free conditions). The graph shows mean cell area per square meter +/- SEM, (n=6). Two independent experiments have been carried out where technical replicates were six. **B) Images (serum free) on control and GLT conditions (CON vs GLT)**.

Similar to serum free conditions, three growth stages (Figure 3.4) are observed in complete conditions, for example at the growth stage (up to 18 hours), initial plateau (from 18 to 72 hours) and toxicity stage (after 72 hours). As we discussed earlier, initial growth advantage is observed in the growth stage and the dramatic fall in cell area is observed in cells with high levels of sugar and fatty acids after 72 hours. An end point growth analysis at 120 hours shows that the average cell area significantly decreases ($p=0.002$; $n=6$; two sample t-test) in GLT treated cells in comparison to controls. Data was checked for normality and a $p=0.2$ in the Kolmogorov-Smirnov test confirmed normal data distribution. Similar results were obtained in 2/3 experiments (data not shown).

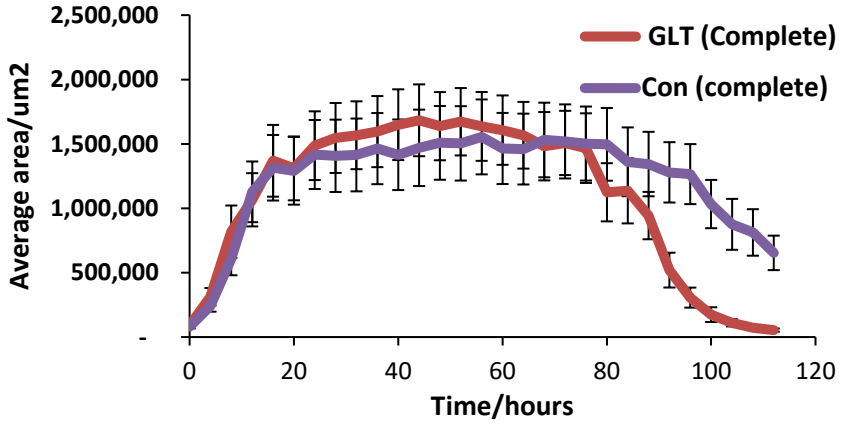
3.4 Discussion

In conclusion, GLT treatment caused a dramatic decrease in cell number both in serum free and complete media conditions. Although the concentration of free fatty acids (free rather than albumin conjugated) is easier to quantify in serum free conditions; it is more physiological to perform the growth curve in the presence of serum, and complete media conditions were used for all experiments going forwards. In future, the effect of high glucose and free fatty acids on INS1 cells can be assessed on caspase activation.

To demonstrate the toxicity effect of GLT on β cell growth, apoptosis should be monitored with progressive activation of caspase 3 and this can be a part of our future analysis.

To study glucolipotoxicity on the pancreatic β cell, we used the rat INS1 β cell line. However, we could have used at least another rat beta cell line e.g. RIN, MIN, HIT etc to see if similar changes occur in our study. It would be very useful to examine the results in human beta cell line EndoC-βH1 which can also be a part of future study.

A)



B)

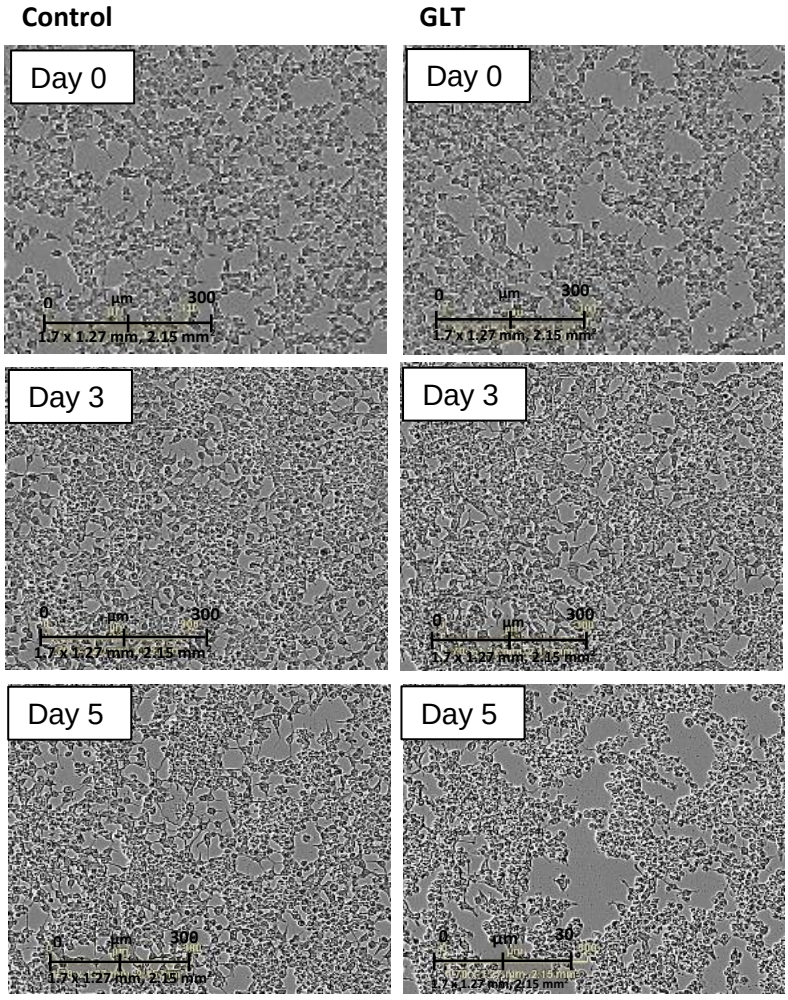


FIGURE 3.4: A) β cell growth characteristics on GLT conditions under physiological conditions. Average area per square meter was used for the calculation of cell growth. Data is presented as mean +/- SEM, of six replicates of each parameter. **B) Images (complete media) on control and GLT conditions (CON vs GLT)**.

Chapter 4 - The effect of GLT treatment on the expression of alternative splicing (AS) regulatory genes

The aim of this part of the study was to determine the effects of high glucose and fatty acid conditions on β cell expression of a panel of genes known to regulate alternative splicing.

4.1 Selection of a Panel of Genes that Regulate Alternative Splicing

Based on a review of the literature, we selected a panel of eleven trans-acting splicing factors, including serine/arginine rich (SR) proteins, heterogenous nuclear ribonucleoproteins (hnRNPs), and a set of neuron-specific splicing regulatory proteins (RBPs) that are known to be expressed in β cells e.g. Elav-like family member proteins, *Rbfox1 & 2* and *Nova1 & 2* family. The genes chosen are not the core components of the splicing machinery, rather they are involved in the regulation of alternative splicing. A full list of the genes selected and a brief description of their role in AS is given in Table 4.5. Primers were then designed (Method 2.3.1) to analyse the expression of these genes by qRT-PCR, and for another set of genes to be analysed for suitability as housekeeping genes. The selected AS genes and corresponding Gene bank accession numbers and the sequences of the primers designed, are listed in Table 4.6. Exon-spanning primers were designed using NCBI Blast software and validated for SYBR green qRT-PCR using PrimePro-48 (Techne) real time PCR system.

4.2 GLT treatment and validation of RNA integrity

The titration of the growth curve, as discussed in chapter 3.3, helped us to choose a suitable time-point during the plateau phase, since β cells are not highly replicative in the adult pancreas, but before the initiation of β cell loss since secondary changes are likely to occur during the process of cell death. The time-

point chosen was 46-hour. Four independent experiments were performed, each with n=3 technical replicates. Growth was monitored at regular 8-hour intervals under an inverted microscope. RNA was then extracted, and the integrity and quantity assessed using an Agilent Bioanalyzer or Nanovue spectrophotometer (Method section 2.2.1). For selective samples, RNA integrity and quality were assessed using Agilent Bioanalyzer (Figure 4.1) which was performed at the Genomics Centre, Kings College London.

Table 4.5: The panel of AS regulatory genes selected for analysis and function of the protein product is summarised based on information obtained from the NCBI database and different publications.

Name of Splicing Regulator/Factor	Gene symbol	Function
Serine and Arginine rich splicing factor 1	*Srsf1*	Required in splice site recognition, ATP-dependent interactions of both U1 and U2 snRNP with pre-mRNA plays a role in preventing exon skipping ensuring the accuracy of splicing and regulating AS (Claudia *et al*, 2010). Acts as an adaptor protein to recruit signalling molecules during tumorigenesis, overexpressed in some cancers and regulates alternative splicing of many cancers regulated genes (Claudia *et al*, 2010).
Serine and Arginine rich splicing factor 3	*Srsf3*	Shuttles between nucleus and cytoplasm (shuttling SR proteins), (Jeong *et al,* 2017). Involved in RNA processing in relation with cellular proliferation and/or maturation (Claudia *et al*, 2010).
Serine and Arginine rich splicing factor 5	*Srsf5*	Involved in constitutive splicing, can modulate the selection of AS (Claudia *et al*, 2010).
Serine and Arginine rich splicing factor 6	*Srsf6*	Promotes expression of isoforms of extracellular matrix protein tenascin C, Dvinge *et al*, 2016. Involved in constitutive splicing, can modulate the AS sites. Repress the splicing of MAPT/Tau exon 10. RNA binding (Claudia *et al*, 2010).
Serine and Arginine rich splicing factor 10	*Srsf10*	Involved in constitutive and regulated RNA splicing (Claudia *et al*, 2010).

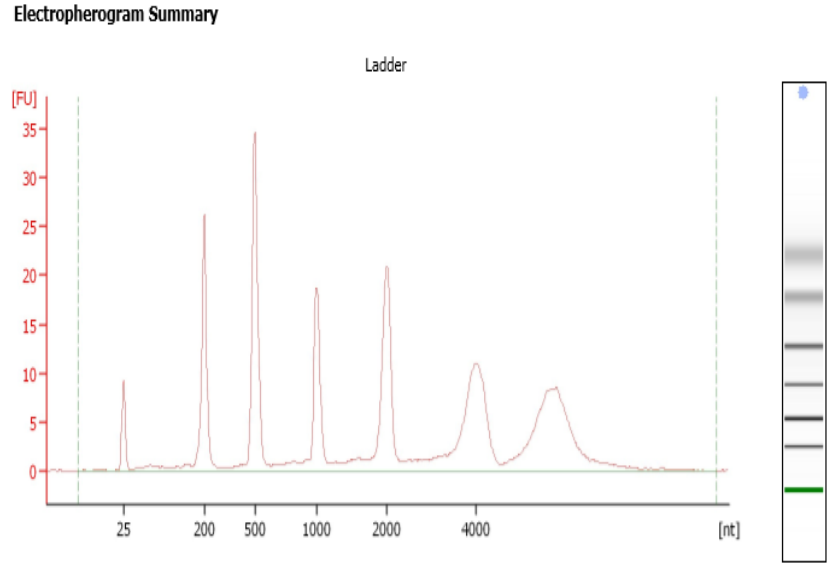
Heterogenous nuclear ribonucleoprotein F	*Hnrnpf*	Belongs to hnRNP family, highly conserved from nematode to mammals and play a role in pre-mRNA maturation. Their function is to bind the exonic splicing silencer (ESS) to exclude *trans*-acting factors such as SR proteins from *cis*-acting elements (Wang *et al,* 2014). A pre-mRNA splicing factor that may be involved in the coordination of transcription and splicing (RGD, Feb 2006).
RNA binding fox-1 homolog 1	*Rbfox1*	The Fox-1 family of RNA-binding proteins is evolutionarily conserved and regulates tissue-specific alternative splicing in metazoa. Fox-1 recognizes a (U)GCAUG stretch in regulated exons or in flanking introns. Several alternatively spliced transcript variants encoding different isoforms have been found for this gene. [provided by RefSeq, Nov 2011]. *Rbfox1* regulates β cell function of insulin secretory capacity without affecting the cells viability (Mateu *et al*, 2017).
RNA binding fox-1 homolog 2	*Rbfox2*	*Rbfox2* regulates β cell function affecting the insulin secretory capacity but without affecting the cells viability (Mateu *et al*, 2017). This gene encodes an RNA binding protein that is thought to be a key regulator of alternative exon splicing in the nervous system and other cell types (Jiang, *et al*, 2008). The protein binds to a conserved UGCAUG element found downstream of many alternatively spliced exons and promotes inclusion of the alternative exon in mature transcripts. Multiple transcript variants encoding different isoforms have been found for this gene. [provided by RefSeq, Jul 2008]
NOVA alternative splicing regulator 1	*Nova1*	Required for β cell survival, and their depletion leads to activation of the intrinsic pathway of apoptosis. (Mateu *et al*, 2017). Alternatively, spliced transcripts encoding distinct isoforms have been described. [provided by RefSeq, Jul 2008]
NOVA alternative splicing regulator 1	*Nova2*	Required for β cell survival, and their depletion leads to activation of the intrinsic pathway of apoptosis. (Mateu *et al*, 2017).
Polypyrimidine tract binding protein 1	*Ptbp1*	Belongs to heterogenous ribonucleoprotein family (hnRNPs). The hnRNPs are RNA-binding proteins and they complex with heterogeneous nuclear RNA (hnRNA) (Ghetti, *et al*, 1990). They are present in the nucleus, associated with pre-mRNA processing and some of them shuttle between nucleus and cytoplasm. The hnRNP proteins have distinct nucleic acid binding

		properties (Ghetti *et al*, 1990). This protein binds to the intronic polypyrimidine tracts that is required for pre-mRNA splicing and acts via the protein degradation ubiquitin-proteasome pathway. It may also promote the binding of U2 snRNP to pre-mRNAs [RefSeq, Jul 2008]. Involved in the negative regulation of alternative splicing [RGD, Feb 2006].
ELAV like RNA binding protein 1 (HuR)	*Elavl1*	This gene belongs to the ELAV family of RNA binding proteins predominantly localized within the nucleus. As a trans-acting factor, it can bind AU-rich element (AREs) of 3' untranslated region. AREs signal degradation of mRNAs to regulate gene expression, thus by binding AREs, the ELAVL family of proteins play a role in stabilizing ARE-containing mRNAs and/or protein translation. (RefSeq, Sep 2012 and Wang *et al*, 2013). The role of Elavl1 in β cells is not known, but it has been shown previously that siRNA knockdown of the related family member, Elavl4, increased β cells apoptosis (Mateu *et al*, 2017), suggesting that this family genes can regulate β cell survival.
ELAV like RNA binding protein 4	*Elavl4*	siRNA knockdown of the ELAVL family member, Elavl4, increases of β cells apoptosis, suggesting that this family of genes can regulate β cell survival. It functions in mRNA turnover and function including AS, stability and translation. (Mateu *et al*, 2017)

Table 4.6: List of genes with their nucleotide sequences of primers utilized in this study.

Gene name	Database accession number	Forward primers' sequence (5'-3')	Reverse primers' (5'-3')	Expected product size
Hnrnpf	NM_001037285.1	CCTTCCTCCCGTGGAAAACTTG	GCATCATGGACACTTGACAGG	133
Ptbp1	NM_001271057.1	AAGCTCACCAGTCTCAACGTC	TACGGAGAGGCTGACATTATGC	149
Srsf1	NM_001109552.2	CCGCGAGACGCGGAAGAT	CACTCTGTTTTCAGACCGCCT	180
Sfsf3	NM_001047907.3	CACTCCGAAGTGTGTGGGTT	GGCAGCCACATAGTGTTCTTC	120
Srsf5	NM_001195505.1	AGCGCAGTTGATTGAGGAAAC	CTGATCCGTCCGTAACCCTTG	127
Srsf6	NM_001014185.2	AAAGTCGCTCGCGATCTAGG	AATGTGAATGGGAACCCCGA	120
Srsf10	NM_001025738.1	CTACACTCGACGTCCAAGAGG	TGATTTGGTGTCTTCCGATCCC	150
Rbfox1	NM_001106974.2	GCGGTGTTGTTTACCAGGAT	GTCGGCAGCATAAACTCGTC	134
Rbfox2	NM_001079895.1	GTGCTGACCTCTATGGTGGA	TGTGTACACCCTGCCGTAAC	122
Nova2	XM_008759097.1	CAGGATTGAGTGACGGCTGAG	ACAGATATCACATAGACCTGGACC	132
Elavl4	NM_001077651.2	CTGTAAGAAGGGAATGTCGGCT	ATGTCGGTCCATTTGACACC	150
Gapdh	NM_017008.4	AGTGCCAGCCTCGTCTCATA	ATGAAGGGGTCGTTGATGGC	133
Dp5	NM_057130.4	GATTGTGCCAGAGCTTCACA	GATTGTGCCAGAGCTTCACA	125
Atf3	NM_012912.2	GCTGGAGTCAGTCACCATCA	ACACTTGGCAGCAGCAATTT	127
Elavl1	NM_001108848.1	TGACAAACGGTCTGAAGCAG	CCTGAATCTCTGTGCCTGGT	196
Nova1	NM_001100541.1	GGAGCAGTCAGGGGCTTGGG	TGAGACAGCTGCCACTCTGTGGA	161
Gsn	NM_001004080.1	CTTTGTTTGGAAAGGCAAGC	GGGGTACTGCATCTTGGAGA	221
Cacna1c	NM_012517.2	AGTGATCCCTGGAATGTTTTTG	AGGACTTGATGAAGGTCCACAG	210
Insr	NM_017071.2	TGCACAACGTGGTTTTTGTT	TCTCTTCTGGGGAGTCCTGA	291
Tubb5	NM_173102.2	GTGATAAGCGATGAACACGGC	GGGACATACTTGCCACCTGTA	116
B2m	NM_012512.2	TCGTGCTTGCCATTCAGAAAAC	GTGGAACTGAGACACGTAGCA	100
Ppia	NM_017101.1	CAGACGCCGCTGTCTCTTTTC	TGCAAACAGCTCGAAGCAGAC	115
Ywhaz	NM_013011.3	ACAACTTGACATTGTGGACATCG	TGGGACAGCATGGATGACAAA	143

A)



B)

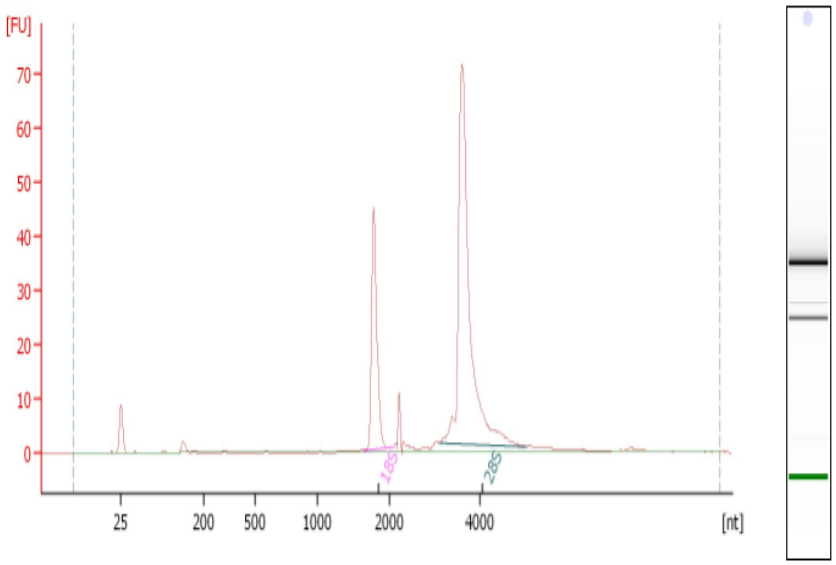


FIGURE 4.1: Analysis of RNA integrity and quality by Agilent Bioanalyzer. All the samples were assigned a RIN number 10, indicative of high quality RNA. **A)** The Electropherogram and gel images of RNA 6000 Nano Ladder and **B)** Total RNA sample had shown two distinct, 18s and 28s, ribosomal peaks and relatively flat baseline between 5s and 18s ribosomal peaks which confirms the RNA integrity and a successful run of the Bio analyser experiment.

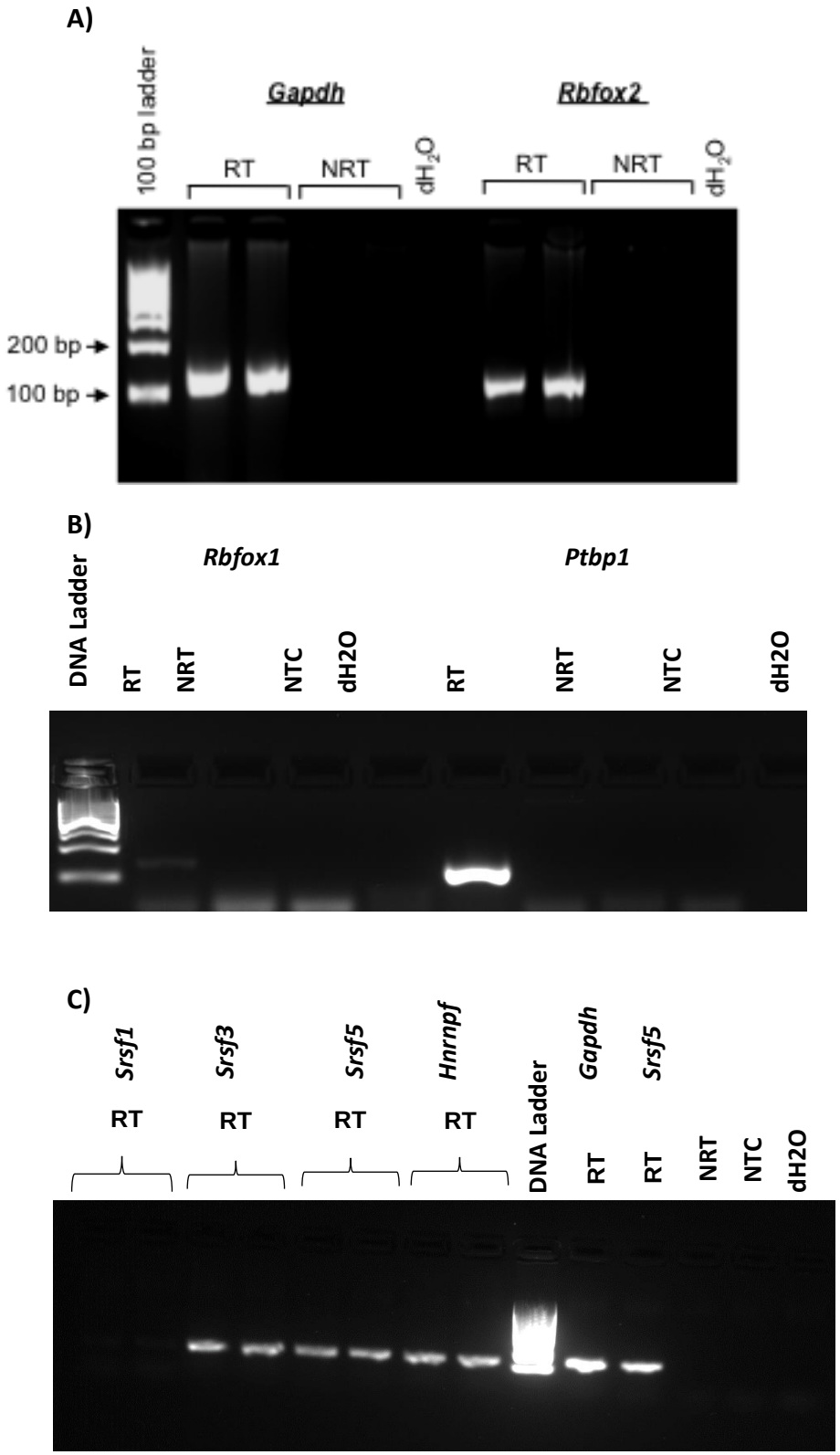
4.3 Validation of primer specificity by RT-PCR.

Expression of a single PCR product of the expected size in INS1 cells using the gene-specific primers was initially validated by RT-PCR. Amplification was performed as indicated in Method (section 2.2.3), and products were loaded onto 2% agarose. Negative controls were also performed to test for genomic and other contamination. A specific product of the expected size was seen for all selected genes in Table 6, except for *Rbfox1* and *Srsf1* as shown in Figure 4.2. *Rbfox1* and *Srsf1* genes were not detectable in INS1 cells using the current primers (Table 6) and were not analysed further. Agarose gel analysis (Figure 4.2) has shown that all genes were lacking in primer dimer formation which is important for subsequent qRT-PCR experiments using SYBR green.

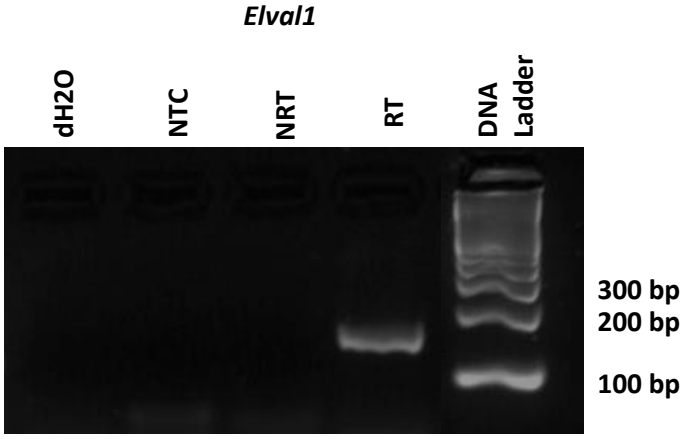
Table 6 shows all predicted amplicon sizes. No PCR amplifications in negative controls in our validation reaction confirmed the absence of contamination, in particular the NRT which tests for genomic contamination. NTCs detect PCR contamination and can also distinguish unintended amplification (Figure 4.2).

4.4 Evaluation using melt curve analysis

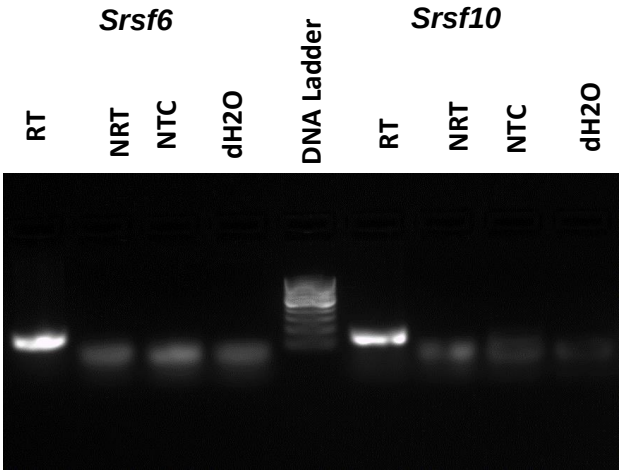
It must be ensured that PCR conditions yield only specific product which we verified by a melt curve analysis (Figure 4.3). Thermal melt curves are formed by placing all double-stranded DNA products at lower temperatures, which is around 60°C, and allowing a slow increase in temperature to denaturing level which is around 95°C. The length of product and sequence affects melting temperatures; therefore, the melt curve is used to characterise amplicon homogeneity. Non-specific amplification is identified by wide peaks in the melt curve or peaks with unexpected Tm values. In conclusion, via a melt curve analysis, we have shown the specificity of primers in a qRT-PCR (Figure 4.3).


A)
100 bp ladder
Gapdh
Rbfox2
RT
NRT
dH2O
RT
NRT
dH2O
200 bp
100 bp
B)
Rbfox1
Ptbp1
DNA Ladder
RT
NRT
NTC
dH2O
RT
NRT
NTC
dH2O
C)
Srsf1
Srsf3
Srsf5
Hnrnpf
RT
RT
RT
RT
DNA Ladder
Gapdh
Srsf5
RT
RT
NRT
NTC
dH2O

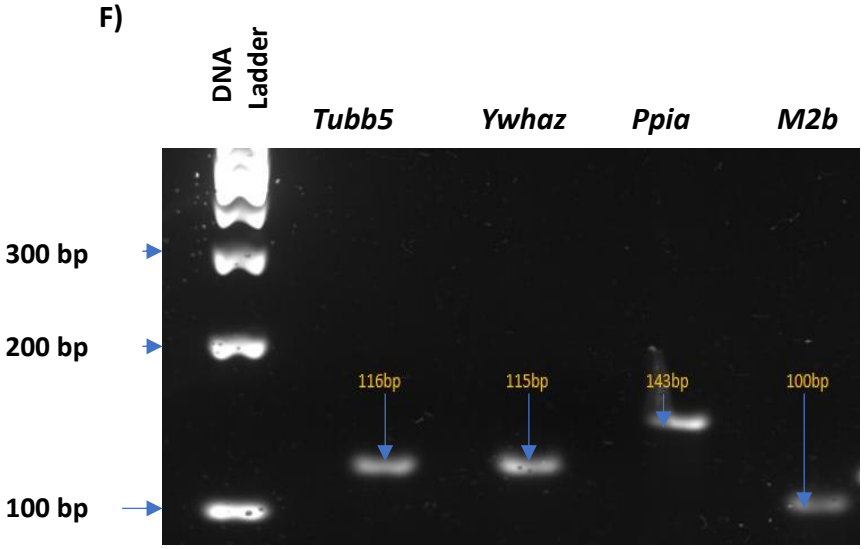
D)


Elval1
dH2O
NTC
NRT
RT
DNA Ladder
300 bp
200 bp
100 bp

E)


Srsf6
Srsf10
RT
NRT
NTC
dH2O
DNA Ladder
RT
NRT
NTC
dH2O

F)


DNA Ladder
Tubb5
Ywhaz
Ppia
M2b
300 bp
200 bp
100 bp
116bp
115bp
143bp
100bp

G)

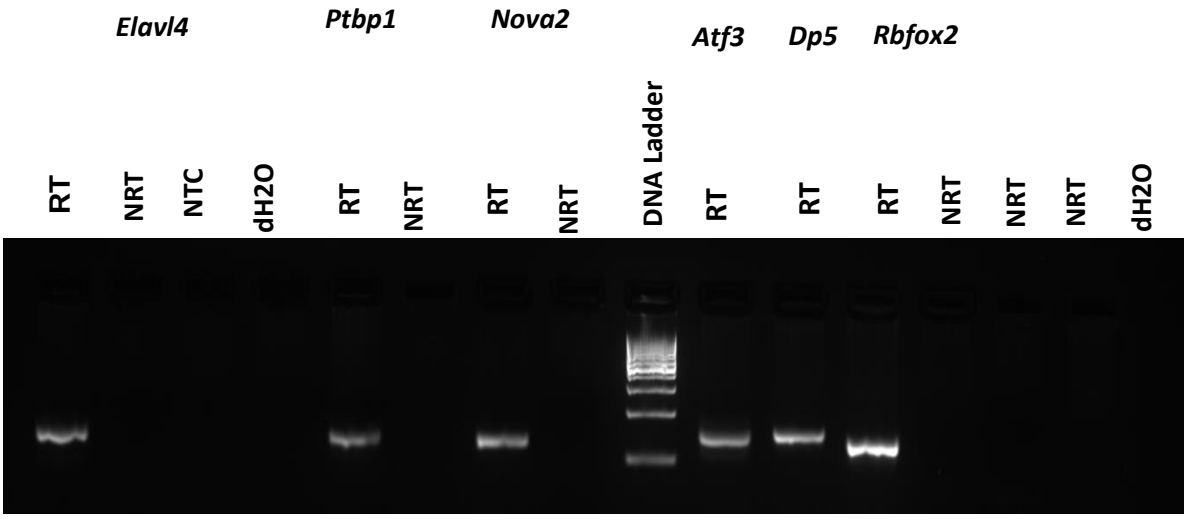
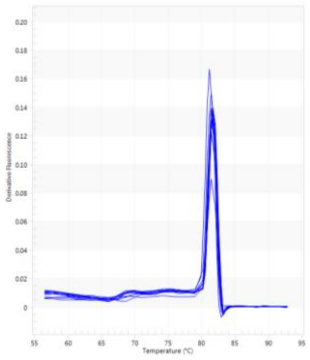
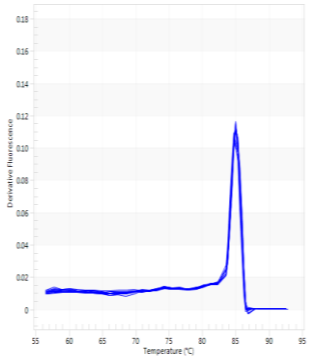


Figure 4.2: Validation of a panel of alternative splicing regulators and endogenous control genes by reverse transcriptase PCR. Images of agarose (2%) gels showing the PCR product specificity of **A)** positive control *Gapdh* and splicing factor *Rbfox2* along with reverse transcriptase minus (RT-) negative controls (NRT) and dH2O. No amplifications were observed in negative controls. Expected PCR products were measured by comparing the amplicons with 100bp DNA ladder. No primer dimers formation was observed on gel. **B)** Splicing factor *Rbfox1* shown very faint PCR amplicon (around 35-36 Cq value in qPCR), but Ptbp1 showing clear specific product. No amplification was observed in negative controls. **C)** Splicing factors *Srsf1* shows very low amplification (Cq value was around 36-37 in qPCR). However, SR primers: *Srsf3* and *Srsf5* have shown specific single PCR product with clean negative controls. Similarly, **D)** Splicing factor *Elavl1,* **E)** SR primers *Srsf6* and *Srsf10;* **F)** rat endogenous control genes *Tubb5*, *Ywhaz*, *Ppia* and *M2b* and **G)** *Elavl4*, *Ptbp1*, *Nova2*, stress primers *Atf3*, *Dp5* and *Rbfox2* primers amplification showing single specific product.

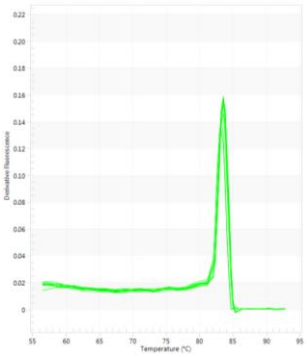
Gapdh



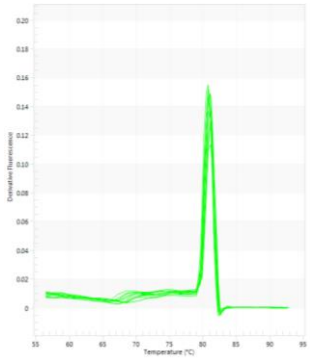
Rbfox2



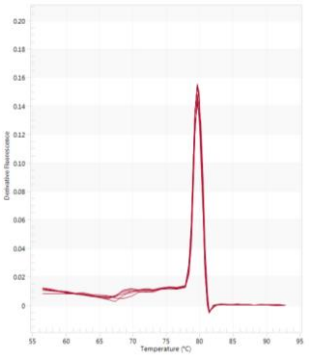
Ptbp1



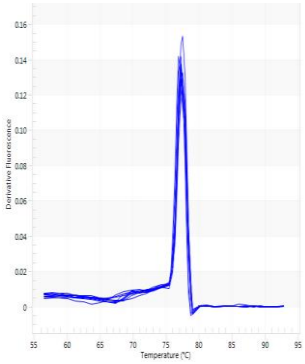
Srsf3



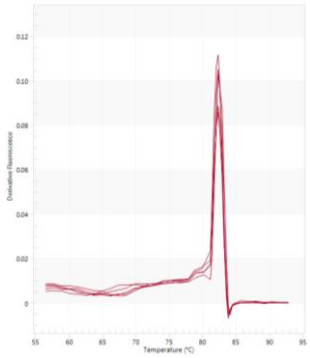
Srsf5



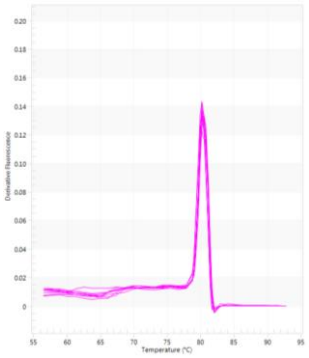
Hnrnpf



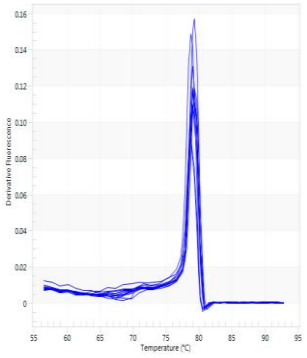
Elavl1



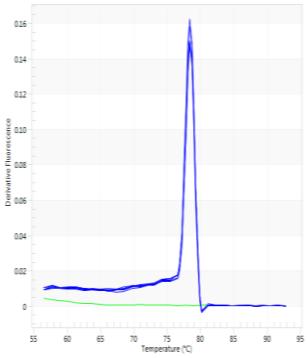
Srsf6



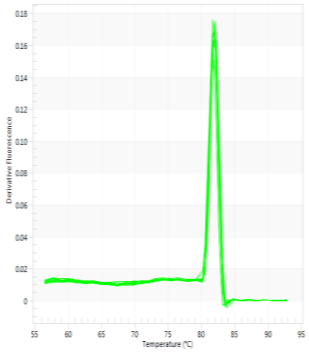
Ywhaz



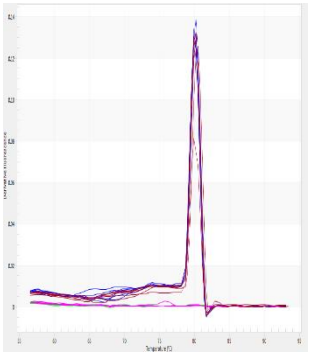
Elavl4



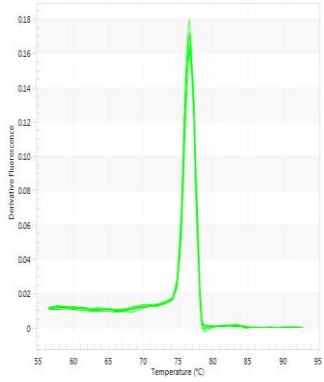
Nova2



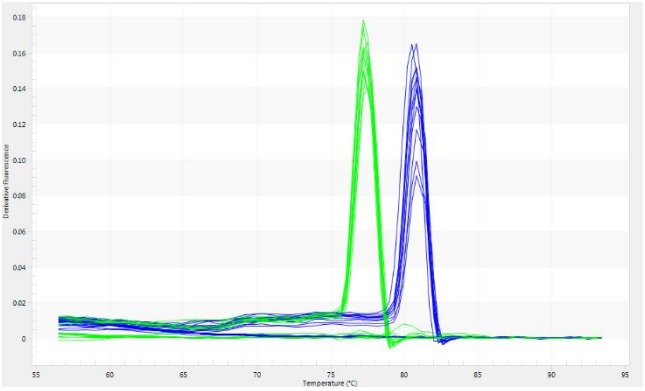
Atf3



Dp5



Green peak corresponds to *Gapdh* and blue to *Srsf10.*



Green peak corresponds to *Ppia* and blue to *Ywhaz.*

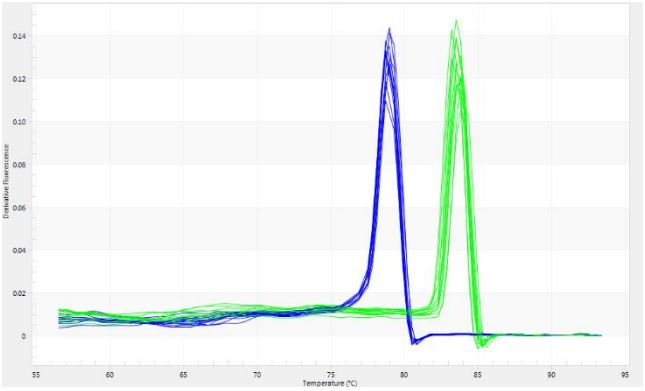


Figure 4.3: Melt curve analysis of all primers used in qPCR experiments. A single peak for each primer represents primers specificity; the presence of wider peaks or more than one peak is the indication of formation of primers dimers.

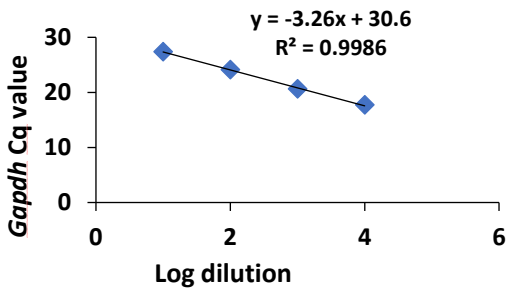
4.4 Validation of primer efficiency and specificity using qRT-PCR

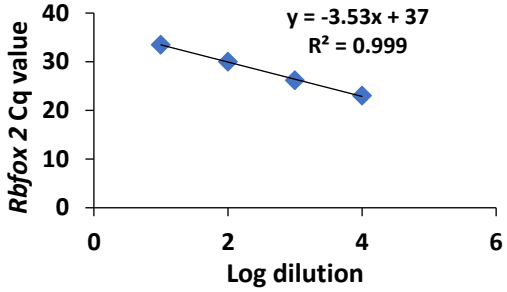
Further validation of primers' specificity and efficiency was performed using SYBR green based qRT-PCR before embarking on further gene expression analysis. To optimise the PCR amplification, primers efficiency is essential for accurate and reproducible quantification. Therefore, we validated primer efficiency for both target genes and a panel of reference genes. Primers *Atf3 and Dp5* have not been fully validated as those primers' sequences were obtained from Oliveira *et al*, 2015.

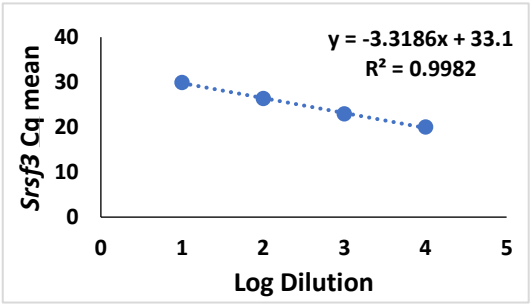
4.4.1 Primer efficiency

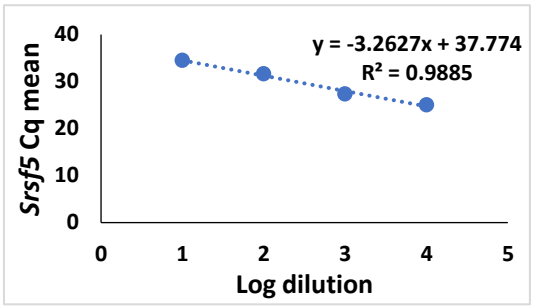
To determine the primer efficiency, serial dilution of the cDNA template was performed and Cq values were plotted against the log dilution. Amplification efficiency was calculated as indicated in the method section 2.3.2. Amplification of each sample was conducted in duplicate.

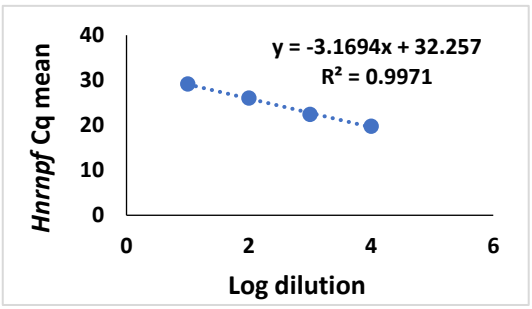
Figure 4.4 has shown that linear standard curves for all our selected panel of genes and also for endogenous control genes*, Ppia and Ywhaz*, with R^2, the coefficient of determination, values >0.99 (Table 4.7). The assay reflected primers high amplification efficiency with a range of 90-110% (Techne manual), A 100% efficiency indicates that the amount of product doubles with each cycle and a range outside that acceptable range is an indication of a problem with PCR reaction set up or with the template.

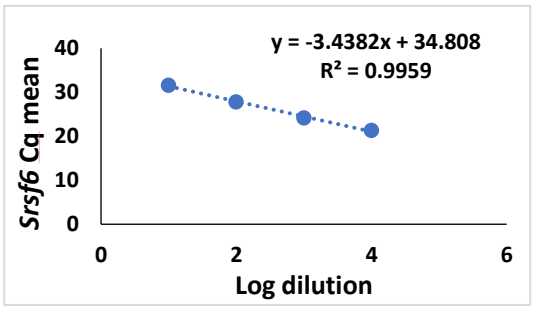

y = -3.26x + 30.6
R² = 0.9986
Gapdh Cq value
Log dilution

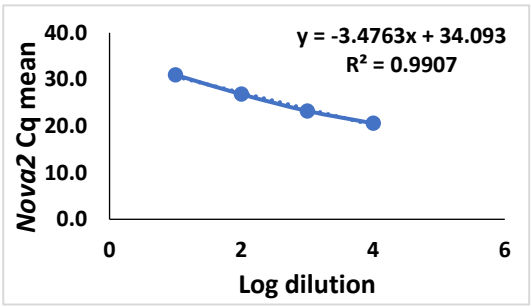

y = -3.53x + 37
R² = 0.999
Rbfox 2 Cq value
Log dilution

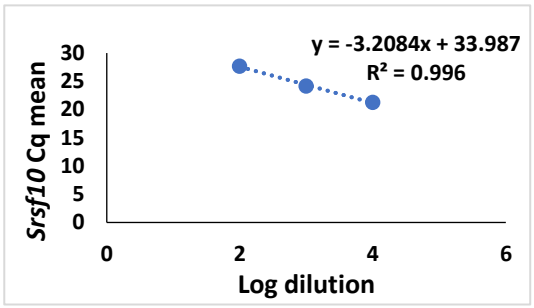

y = -3.3186x + 33.1
R² = 0.9982
Srsf3 Cq mean
Log Dilution

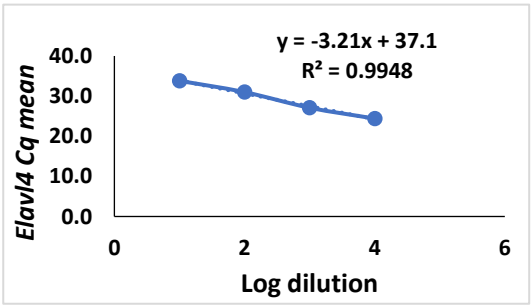

y = -3.2627x + 37.774
R² = 0.9885
Srsf5 Cq mean
Log dilution

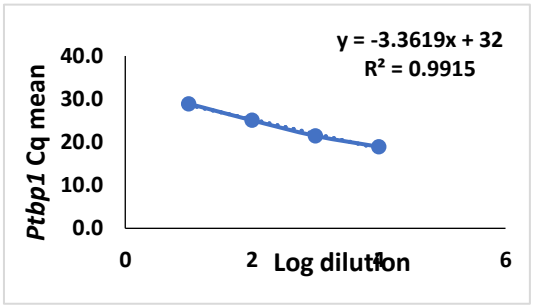

y = -3.1694x + 32.257
R² = 0.9971
Hnrnpf Cq mean
Log dilution


y = -3.4382x + 34.808
R² = 0.9959
Srsf6 Cq mean
Log dilution


y = -3.4763x + 34.093
R² = 0.9907
Nova2 Cq mean
Log dilution


y = -3.2084x + 33.987
R² = 0.996
Srsf10 Cq mean
Log dilution


y = -3.21x + 37.1
R² = 0.9948
Elavl4 Cq mean
Log dilution


y = -3.3619x + 32
R² = 0.9915
Ptbp1 Cq mean
Log dilution

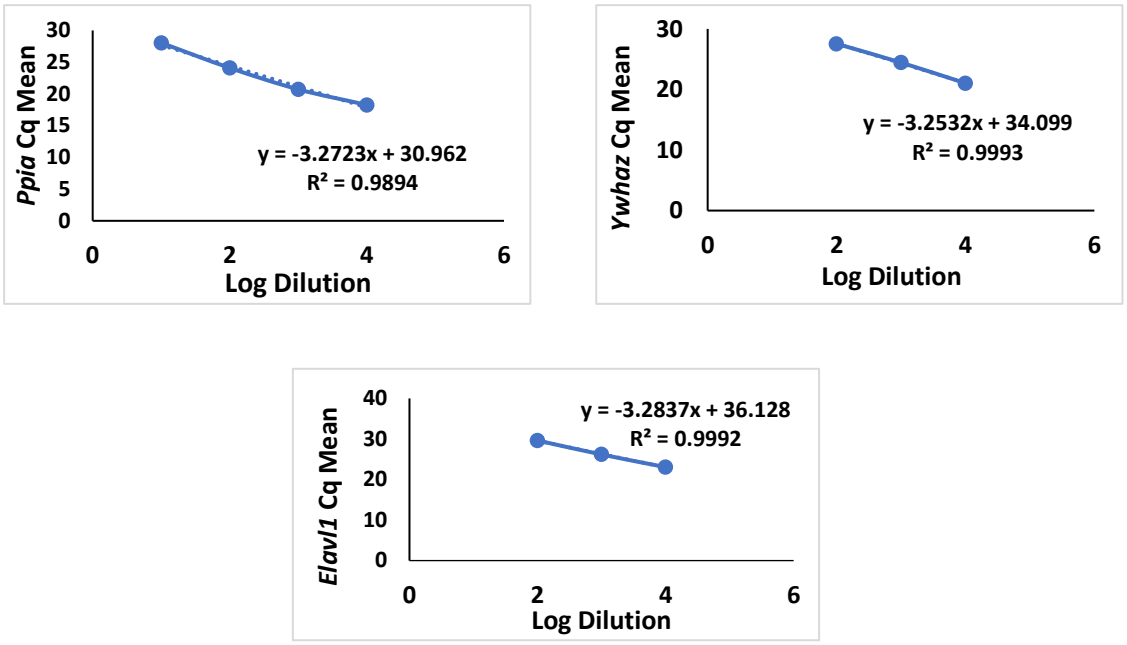


Figure 4.4: Generation of a standard curve from serial dilution (10-fold) of nucleic acid template of known concentration. Standard curves with the Cq mean plotted against the log of the starting quantity of template for each dilution. The equations for the regression line and R^2 values are shown on graphs.

A summary of the primers' efficiency performance is shown in Table 4.7.

Table 4.7: A summary of Primers' Efficiency Performance.

Gene name	Single PCR product with no primer dimer	Melt curve analysis	%Efficiency Accepted range (90-110; Techne Mannual)	R^2 value	Clean negative controls	Passed all validation
Gapdh	Yes	Single peak	103	0.99	Yes	Yes
Rbfox2	Yes	Single peak	92	0.99	Yes	Yes
Nova2	Yes	Single peak	94	0.99	Yes	Yes
Elavl4	Yes	Single peak	104	0.99	Yes	Yes
Hnrnpf	Yes	Single peak	106	0.99	Yes	Yes
Ptbp1	Yes	Single peak	98	0.99	Yes	Yes
Srsf3	Yes	Single peak	100	0.99	Yes	Yes
Srsf5	Yes	Single peak	102	0.99	Yes	Yes
Srsf6	Yes	Single peak	95	0.99	Yes	Yes
Srsf10	Yes	Single peak	104	0.99	Yes	Yes
Elavl1	Yes	Single peak	102	0.99	Yes	Yes
Ywhaz	Yes	Single peak	103	0.99	Yes	Yes
Ppia	Yes	Single peak	102	0.99	Yes	Yes
Rbfox1	No	Not applicable	Not applicable	Not applicable	Yes	No
Srsf1	No	Not applicable	Not applicable	Not applicable	Yes	No

4.5 Validation of housekeeping gene stability under GLT conditions

To compensate for slight variation in template amount qRT-PCR, we need to normalise expression to a housekeeping gene and that expression of the housekeeping gene must be constant across different treatment conditions for this to be valid. Therefore, we examined the stability of the initial reference gene, *Gapdh*. To do that, four independent experiments were performed in which INS1 cells were treated with high sugar and high fatty acids (GLT treatment) or control as described in method section 2.1.7. Total RNA was extracted (section 2.2.1) and cDNA was synthesised and amplified by qRT-PCR using *Gapdh* gene. A statistical data analysis by K-S (Kolmogorov-Smirnov analysis) test ensured a normal data distribution by a p value >0.05 which were analysed by SPSS (v.24). We collected our relative analysis data, $\Delta\Delta$Cq values expressed as fold change compare to control 1 (Livak *et al*, 2001), of *Gapdh* control vs. GLT for all four experiments and allowed them in statistical analysis of two samples independent t-test (Mean +/- SEM; Figure 4.5) each with three technical replicates.

As shown in Figure 4.5, our data showed decreased expression of *Gapdh* mRNA in the presence of high glucose and high fatty acids conditions decreased in (Exp A, $p<0.09$; Exp B, $p<0.06$ and in Exp D, $p<0.1$) and a significant increased expression was observed in Exp C, $p>0.004$. Despite *Gapdh* being a commonly used reference gene for GLT studies (Mateu *et al*, 2016; Bagnati *et al*, 2016), in our hands *Gapdh* expression is therefore unstable under GLT conditions and should not be used. In such a situation, we decided to select a further set of housekeeping genes (Table 6) from the list provided by 'TATAA Biocentre' on their web page, designed their exon-exon spanning primers and validated them as described above.

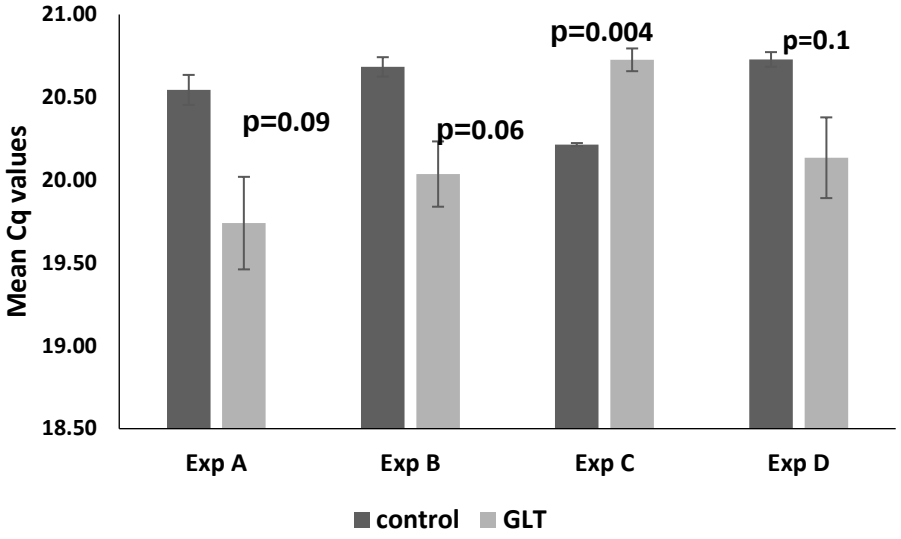


FIGURE 4.5: Stabilisation of *Gapdh* as a housekeeping gene on Glucolipotoxic Conditions. qRT-PCR analysis of *Gapdh* expression under GLT and control conditions. Data represents 4 independent experiments. Data shows Cq mean +/-SEM. Statistical analysis using two sample independent t-test. A *p* value <0.05 indicates statistical significance.

Among the four housekeeping genes (*Tubb5*, *M2b*, *Ywhaz* and *Ppia*; Table 6), *Ppia* and *Ywhaz* were randomly selected for further validation on GLT conditions. Using the same cDNA samples prepared from 4 independent experiments, as described above in section 4, a qRT-PCR was performed using primers specific for those genes (*Ywhaz* and *Ppia*) to determine Cq values in both control and treatment conditions in all four experiments. A two-tailed independent two sample t-test was carried out to analyse the data. Data analysis clearly guided us by concluding that among the three reference genes, *Ppia* (p=0.7) is highly stable under GLT conditions while *Gapdh* (p=0.2) and *Ywhaz* (p=0.2) were the least stable (Figure 4.6).

In conclusion, we have shown that *Ppia* expression is stable under GLT conditions, and conversely that *Gapdh* should not be used as a qRT-PCR housekeeping gene in this type of study.

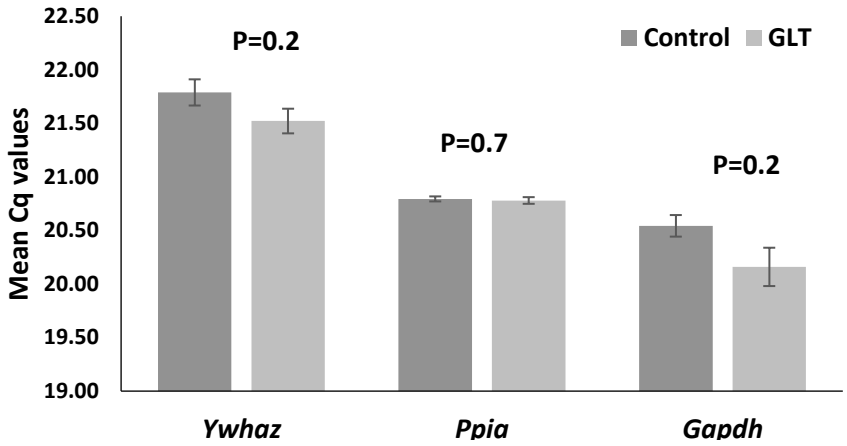


Figure 4.6: Endogenous housekeeping gene *Ppia* is the most conserved on GLT conditions. *Graph* shows Cq mean +/- SEM, n=4. Statistical analysis using two tailed paired t-test.

4.6 Validation of stress response gene upregulation following GLT treatment

In order to validate the response to GLT treatment at the level of gene expression, we used published primers to two genes involved in stress responses that are known to be upregulated in INS1 cells following GLT treatment (Cunha *et al,* 2008 and Oliveira *et al,* 2015). Glucolipotoxicity induced β cells apoptosis via induction of ER stress or activation of mitochondrial pathway of cell death which has already been discussed in section 1.4.1. Activation transcription factor 3 mediates the ER stress response along with spliced X-box binding protein 1 (XBP1s) and C/EBP homologous protein (CHOP), as well as the proapoptotic BCL-2 family member death protein 5. We therefore randomly selected *Atf3* and *Dp5* to investigate the stress response on INS1 β cells induced by GLT.

Using the same cDNA samples prepared from 4 independent experiments, as described above in section 4.5, expression of the stress response gene *Dp5* and *Atf3* was determined by a qRT-PCR, using *Ppia* as a reference gene. Normal distribution of data was confirmed before any further statistical analysis. As shown in Figure 4.7, although no change in *Atf3* expression was observed, a significant increase in *Dp5* expression (p=0.01, n=4) following GLT treatment was observed, consistent with previous publications (Oliveira *et al,* 2015).

In conclusion, increased expression of the stress-marker *Dp5* mRNA following GLT treatment, validated the effectiveness of our GLT treatment.

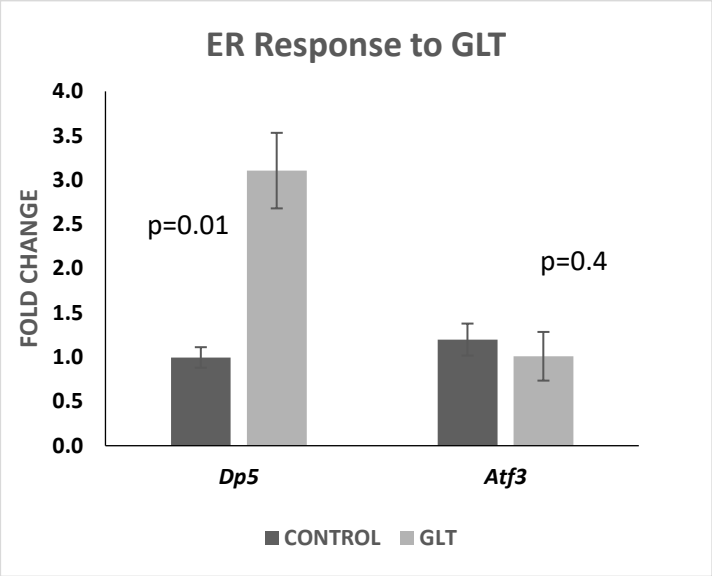


Figure 4.7: Validation of glucolipotoxicity by induction of stress markers. Data from four independent experiments, they were normalised for the expression level of the housekeeping gene *Ppia* and a significant increase (two sample paired t-test) in expression of *Dp5* mRNA validates the effectiveness of GLT treatment. Graph shows mean fold change +/-SEM.

4.7 Expression of alternative splicing regulatory genes following GLT treatment

To test the hypothesis that the expression of genes that regulate AS is affected by GLT treatment, we used the panel of AS gene primers validated the above sections, cDNA synthesized using RNA isolated from the four GLT treatment experiments (each with n=3 technical replicates) as described previously (section 4.5) was used in qRT-PCR reactions with each of the primer pairs, and *Ppia* was used as a reference gene to calculate fold changes using the ΔΔCq method.

As shown in Figure 4.8, the splicing regulator *Elavl1* is upregulated in INS1 β cells following GLT treatment (p=0.04, n=4, two-tailed paired sample t-test).

No significant changes in the regulation of serine/arginine rich (SR) family genes e.g. *Srsf3* (p=0.8, n=4)*, Srsf5* (p=0.5, n=4)*, Srsf6* (p=0.7, n=4) and *Srsf10* (p=0.9, n=4.) were observed, suggesting that these trans-activating splice factors are not affected by GLT treatment. A slight decrease in *Ptbp1* and *Hnrnpf* expression was observed but this did not reach statistical significance (p=0.3, n=4 for both genes). Further experiments would be required to determine whether these are real decreases or just experimental variation.

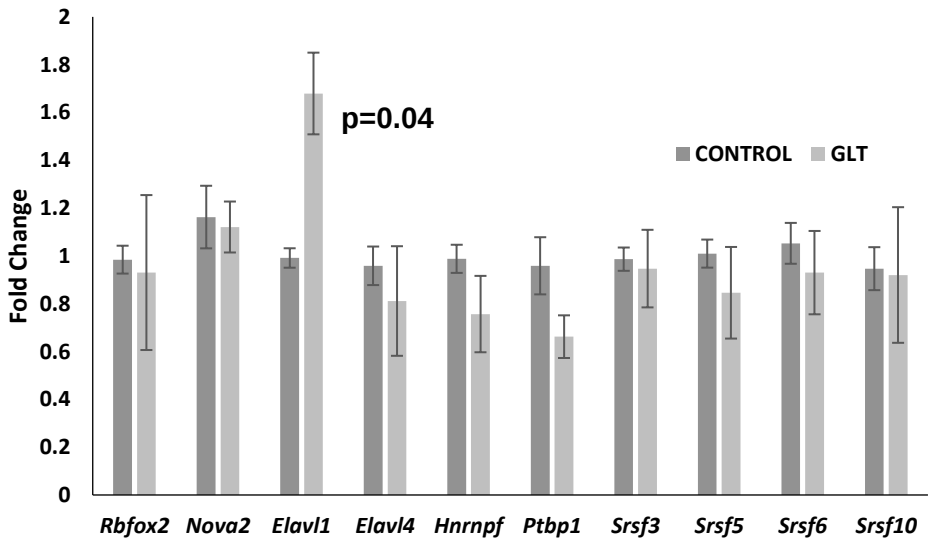


Figure 4.8: Regulation of RBPs, hnRNP and SR family genes following GLT treatment. Expression was normalised to *Ppia*, and relative to a single control value. Data represents 3-4 independent experiments, each with 3 technical replicates. Graph shows mean fold change +/- SEM. Graph indicates mean fold change relative to a single control sample. A p value <0.05 indicates statistical significance

4.8 Discussion

- In this study, we selected and tested 13 splicing regulatory genes under GLT conditions based on some published literatures. There are some notable omissions that are expressed in β cells for example, some negative *trans*-acting factors hnRNPs (*Hnrnpa1* and *Hnrnpa2b1)* which are found to be associated with cancer and ageing (Section 1.5.3), should be tested under GLT conditions. In future, more SFs expression should be tested and the Clariom data will be useful in selecting and testing SFs in altered metabolic conditions.

- We have shown that reference gene rat *Ppia* expression is stable under GLT conditions, and conversely that *Gapdh* should not be used as a RT-qPCR loading control in this type of study.

- These experiments showed that GLT conditions significantly increase ($p=0.04$, $n=4$) expression of the stress-associated mRNA processing and stability gene *Elavl1* in β cells.

- The role of *Elavl1* in β cells is not known, but it has been shown previously that siRNA knockdown of the related family member, *Elavl4*, increased β cell apoptosis (Mateu *et al*, 2017), suggesting that this family of genes can regulate β cell survival.

- *Elavl1*, also known as HuR, is an RNA binding protein that is known to be upregulated in multiple types of cancer (Jun *et al*, 2013). It has been shown to bind near 3' splice acceptor sites and to regulate alternative splicing, for example of the translation regulator *Eif4enif1* and the apoptosis receptor *Fas* (Chang *et al*, 2014). The protein functions of HuR are not clear as no target sequences of HuR have not been identified yet (Izquierdo, 2008).

In summary, these experiments demonstrate that exposure to high glucose and free fatty acid conditions increases expression of the splicing regulator *Elavl1* in β cells. This suggests that the metabolic conditions that exist in some people with type 2 diabetes may similarly upregulate *Elavl1*.

Chapter 5 - RT-PCR alternative splicing of candidate gene

Variation in *TCF7L2* is an important factor determining T2D risk and is known to be regulated by alternative splicing (as summarised in introduction section 1.6.1). We therefore wished to examine whether GLT treatment and the upregulation of *Elavl1* were associated with changes in *Tcf7l2* isoform expression in β cells.

5.1 Identification of rat *Tcf7l2* splice variants analysis and their regulation by GLT conditions

5.1.1 Initial analysis of rat *Tcf7l2*

The human genomic structure of *TCF7L2* consists of 17 exons and is abundantly expressed in pancreatic islets, adipose tissue, (Osmark *et al*, 2009). Previous research revealed by other groups has revealed tissue-specific splicing pattern of this gene and four splice variants expressed in pancreatic islets, (Hansson *et al*, 2010). In β cells, four primary splice variants (see introduction; Figure 1.7) were identified in one study (Hansson *et al*, 2010). However, other studies have identified additional variants based on alternative splicing of the same regions (Le Bacquer *et al*, 2011). So far, no splice variants have been reported for rat *Tcf7l2* in the database. We therefore performed an initial comparative analysis of the rat and human exon structures to identify in rats, the regions known to be alternatively spliced in human β cells.

There are 30 transcripts of the human *TCF7L2* gene identified in ENSEMBL, of which 21 are predicted to be protein coding. As shown in Figure 5.1 two regions that are alternatively spliced in human β cells are exon 4 at the 5' region, and exons 13-16 in the 3'end (Duval *et al* in Hansson, 2010). The first exon in the rat *Tcf7l2* gene is not annotated as an exon in the human *TCF7L2* sequence. Human exons1-3 are also not annotated as exons in the rat sequence, raising

the possibility that the 5' mRNA sequence is distinct in rats and humans. However, the sixth human exon is homologous with rat exon 2. The alignment of specific human/rat exons are shown in Appendix Figure 9.4.

Looking at the second alternatively spliced region in the human gene (exons 13-16), human exon 11 and 12 are homologous to rat exons 7 and 8, respectively, human exon 15 is homologous to rat exon 9, and human exon 17 is homologous to rat exon 10 (Figure 5.1). However, the alternatively spliced human exons 13, 14 and 16 are not annotated as exons in the rat sequence.

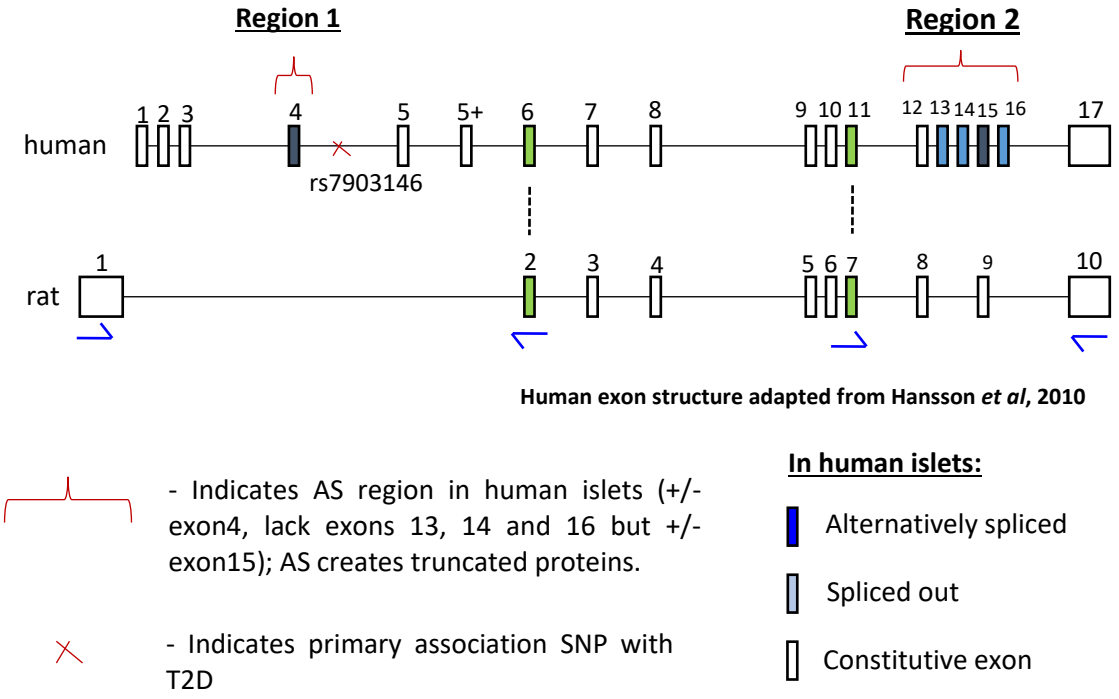


Figure 5.1: A comparison of *TCF7L2* exon structure between human and rat in database. Human exon structure and numbering (see Figure 1.7) has been adapted from Hansson *et al*, 2010.

Based on this analysis, we designed primers to amplify and sequence the rat exons 1 to 2 (Region 1) and exons 7 to 10 (Region 2) using NCBI database (Table 5.8), and look for multiple PCR products as evidence of *Tcf7l2* alternative splice variants in the rat.

Table 5.8: Nucleotide sequence of primers for *Tcf7l2* amplification in region 1 and region 2 by RT-PCR in rat pancreatic β cells.

Primer name	Database accession number	Forward primers' seq	Reverse primers' seq	Expected product size
Tcf7l2 1F and 2R	NM_001191052.1	ACTCCTTACAAAAGGTTGGGGA	TTGCTGTACGTGATGAGAGGTG	109
Tcf7l2 7F and 10R	NM_001191052.1	CTGGTCTGCACGGGATAACT	GCAGCTGCCTTCACCTTGTA	190

5.1.2 Analysis of the *Tcf7l2* splice pattern in INS1 cell

To analyse the splice pattern in rat INS1 cells, total RNA was extracted and reverse transcribed, and RT-PCR was conducted using two pairs of primers listed above. Exon 1-2 PCR showing the presence of four amplicons, named as A1-A4; low to high molecular weight. In region 2, another four PCR amplicons were identified (A5-A8; low to high molecular weight). After sanger sequencing of purified PCR products, no splice variants were identified in region 1, the reason for which may be that the 5' mRNA site is distinct in rats as we have discussed in above section 5.1.1. In region 2, we successfully identified three novel alternatively spliced isoforms, named as B (A5), C (A6) and D (A8), along with discovering three new novel exons, namely 8b, 8c and 9b, which are assumed to be the primary variants expressed in β cells (Figure 5.2). Predicted PCR amplicon was named as A. A schematic presentation summarizing the splice pattern in *Tcf7l2* has shown in Figure 5.3. A summary of this sequence analysis has shown in Table (Appendix 9.9).

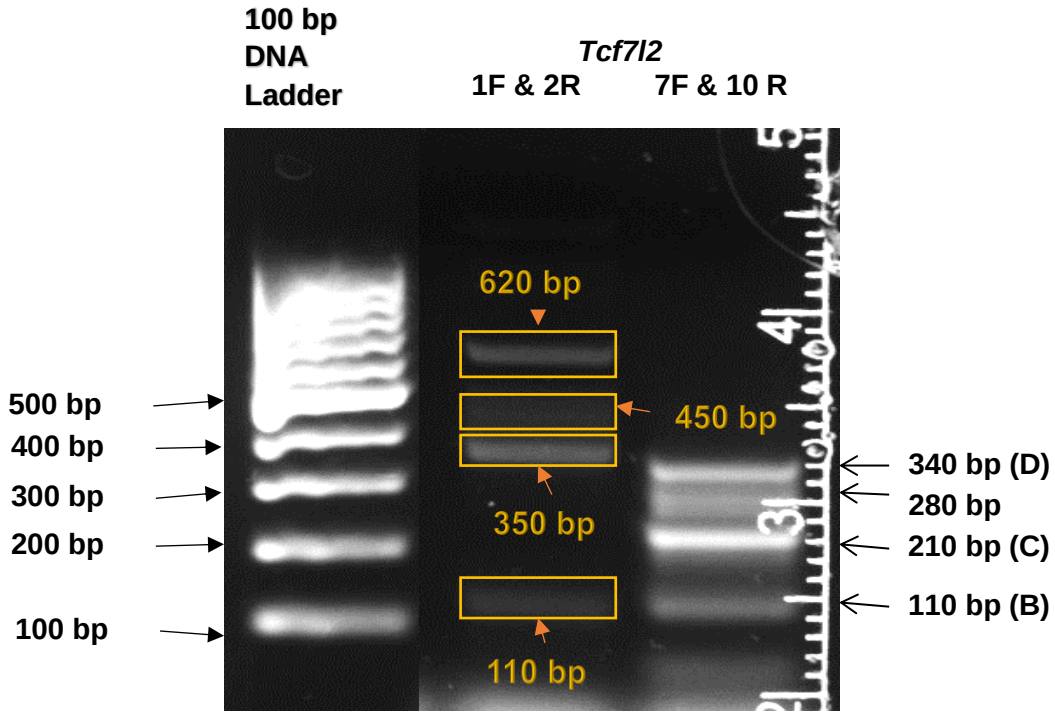


Figure 5.2: Splice variants analysis of rat *Tcf7l2* on pancreatic β cells by RT-PCR. Amplicons visualised on one percent low melting point agarose gel and purified for Sanger sequencing. A 100 bp DNA ladder was used to estimate the PCR products.

In region 2, three newly identified exons between exons 7 and 10 are shown in red colours, exons 8b,8c and 9b (Figure 5.3). These correspond to human exons 13, 14 and 16 respectively, the alignments are shown in Appendix 9.4. The difference between the isoforms involves the presence or absence of exons 8b, 8c, 9 and 9b: none of these exons were present in short isoform B (119bp), while isoform C contains the exons 9 and 9b, and isoform D contains all newly identified exons along with exon 9. Exon 7 and 10 are present in all isoforms (Figure 5.3; Appendix 9.5 to 9.8).

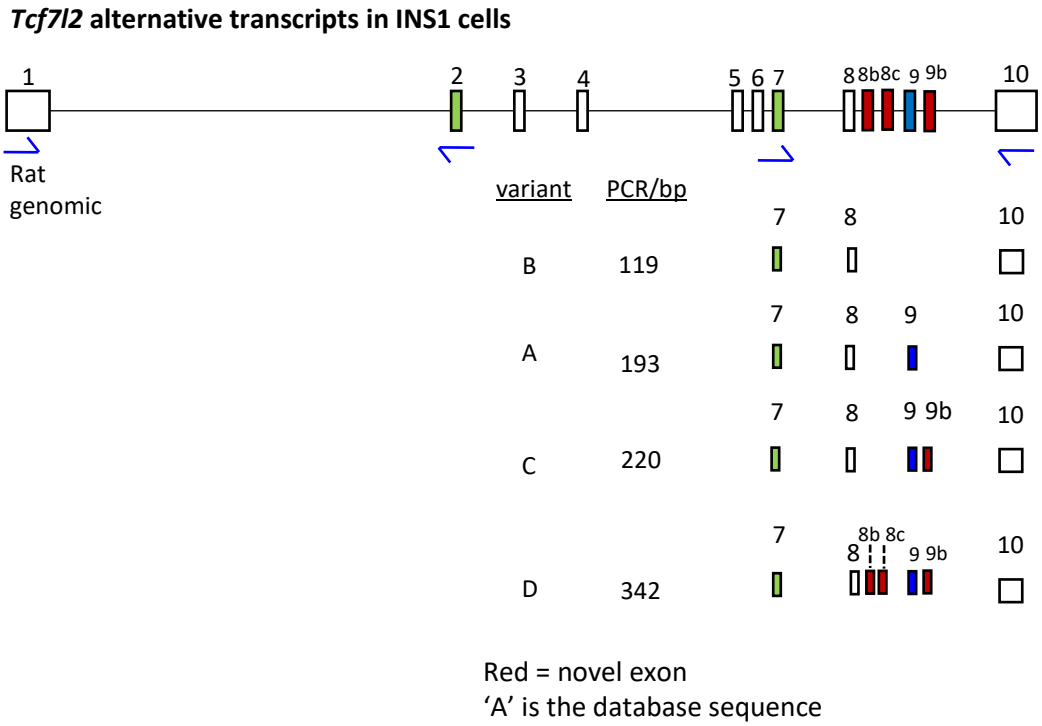


Figure 5.3: A schematic presentation of three novel primary *Tcf7l2* transcript variants in the pancreatic β cell line INS1 in region 2. Three new novel exons in this region are 8b,8c and 9b; are shown in red colour.

5.2 The splice pattern of *Tcf7l2* on glucolipotoxic conditions

We then examined the expression of *Tcf7l2* isoforms following GLT treatment, focussing on region 2 as we have not found any isoforms in region 1. The first replicate, both control and GLT, from each of the four independent experiments performed were analysed.

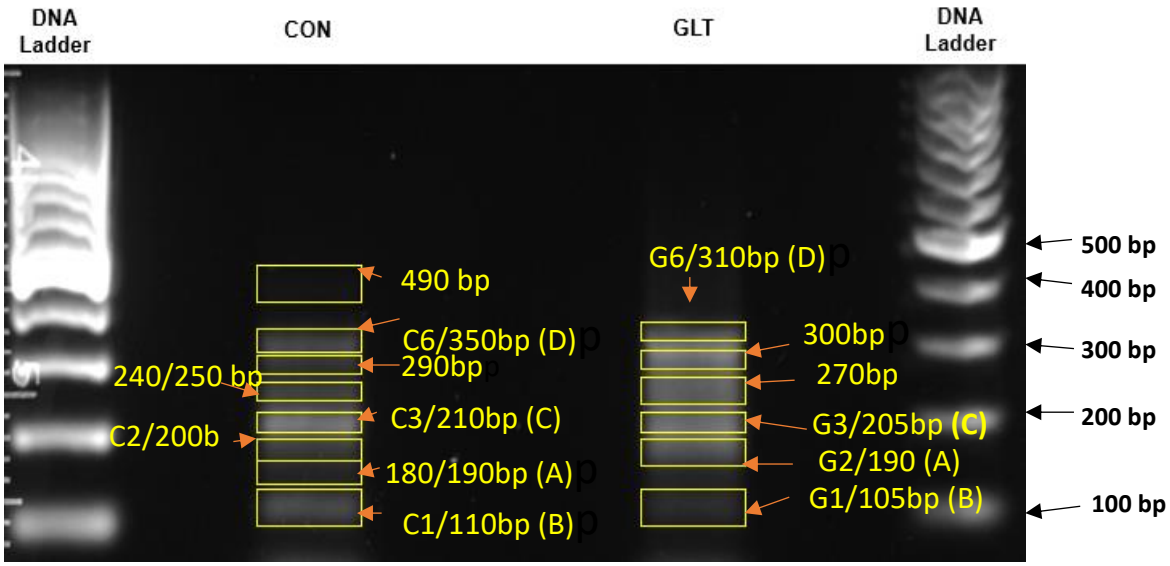


Figure 5.4: Amplification of *Tcf7l2* splice variants in region 2 both in control and altered metabolic conditions by RT-PCR. Amplicons visualised on one percent low melting point agarose gel and purified for Sanger sequencing. A 100 bp DNA ladder was used to estimate the PCR products. Estimatedd band sizes are shown in the figure.

After running the RT-PCR amplicons on 1% agarose gel (Figure 5.4), up to eight variants in GLT and six in control were observed. Due to time constraints, only four were successfully sequenced for both control (C1, C2 and C3 and C6) and GLT (G1, G2, G3 and G6). A summary of this analysis is shown in Table in Appendix 9.10.

Based on our findings above, we analysed the changes in expression of those alternatively spliced isoforms in altered metabolic conditions by comparing them with control on four percent agarose gel. As indicated in Figure 5.5, a shift to lower molecular weight transcript was observed after GLT treatment, where expression of short isoforms B and C went up and isoform D went down (Figure 5.5). Four independent experiments were carried out. In future mRNA (isoforms) quantification should be carried out to confirm this analysis.

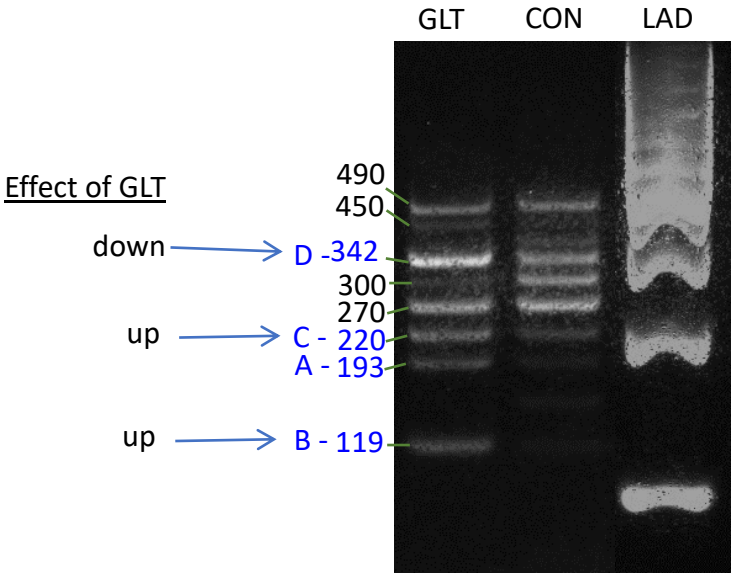


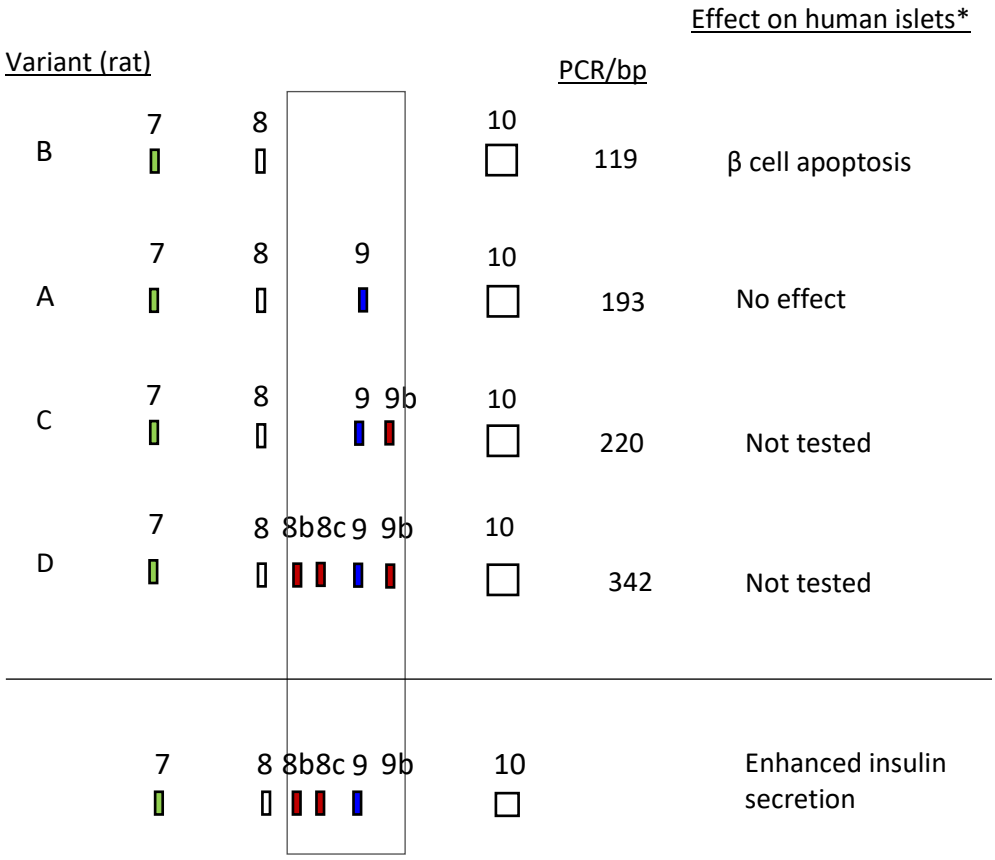
Figure 5.5: Shift to lower molecular weight transcript after GLT treatment. Representative image of four independent experiments, showing the first replicon of first experiment (Control and GLT treated). *Tcf7l2* splice variants on high sugar and free fatty acids and compare them with controls. Four independent experiments (Appendix 9.11 B).

5.3 Discussion

In this experiment, we have successfully identified three new different alternatively spliced products in rat Tcf7l2, named as isoform B, C, and D along with discovering three new novel exons, namely 8b, 8c and 9b.

In humans, protein isoforms formed by AS of *TCF7L2* exerts a distinct role depending on the exons present on that isoform and that may affect the β cell function and survival (Le Bacquer *et al*, 2011). They have demonstrated that isoforms containing the combination of exons 13, 14 and 15 produced different ORF and overexpression of this full length *TCF7L2* had the protective effect on β cell survival and function. On the other hand, overexpression of short variant (deficient of exons 13,14, 15 and 16) induced impaired insulin secretion and apoptosis in human islets (Weise *et al*, 2010; Le Bacquer *et al*, 2011). We have already shown that rat exons 8b, 8c, 9 and 9b are aligned with human 13, 14, 15 and 16 (Appendix 9.4). In a comparison *TCF7L2* transcript analysis published by Le Bacquer *et al* (2011), we found that short isoform B in rat pancreatic β cells are devoid of exons 8b, 8c, 9 and 9b and overexpressed on altered metabolic

conditions (Figure 5.5) which may induce β cell apoptosis under this altered conditions rather than low expression of isoform D (containing exons 8b, 8c, 9 and 9b) which in normal condition may enhance insulin secretion and has a protective effect on β cells survival but these are all suggestive roles, the biological effect of those isoforms in rat have not been tested yet (Figure 5.6).



*Based on data of comparative transcripts from study in human islets published by Le Bacquer *et al*, 2010.

Figure 5.6: *Tcf7l2* variants have distinct activity in islets, a comparison of isoforms between human islets by Le Bacquer *et al* (2010) and rat isoforms discovered in INS1 β cells.

In summary, we have identified three novel *Tcf7l2* isoforms (B, C, D) in rat with the discovery of three novel exons (exon 8b, 8c and 9b) and we have shown that GLT treatment is associated with upregulation of a variant (short isoform B) that may be associated with β cell death. This conclusion of the functional consequence of *Tcf7l2* results is based on comparative isoforms study in humans. In future, we need to study the isoforms expression at the protein level in rat β cells.

Chapter 6- Exon array analysis of alternative splicing events in β cells following GLT treatment.

The aim of the exon array performed in this section was to identify changes in gene splicing induced by GLT treatment of β cells at the global level, and to the main cellular pathways affected by alternative splicing.

RNA previously isolated as part of the qRT-PCR analysis (chapter 4) was used for hybridisation to Affymetrix Clariom D exon arrays (see method section 2.4). Single replicate control and GLT samples from each of the two experiments were submitted to the Genomic Centre at King's College London for hybridisation. Analysis was performed using 'Transcriptome Analysis Console software (T.A.C; Thermo Fisher Scientific), as described in the methods. Clariom D is a step up from a conventional array because it provides a highly detailed view of transcriptome and a faster path to actionable results, investigating altered pathways from well-annotated genes, enables rapid discovery of coding and non-coding genes, exons, splice variants and rare transcripts, able to deliver the accuracy in gene expression measurements throughout the transcriptome by sequencing on a High^{TM}2000 system (ThermoFisher Scientific).

CEL files (raw data) were processed using a modification of the normalisation method 'Robust Multi-array Average' (SST-RMA). Differences arising from inter-experiment variation were reduced by using the 'batch effects' setting during processing. The results of the data normalisation process can be illustrated using box plots as shown in Figure 6.1.

The signal box plot is based on either CHP or CEL files data before normalisation. The purpose for examining box plots is to look for whether individual array data are different from others and, most importantly, from other replicates in the same group. It is a useful graph to show the shape of the distribution of the data as well as its central value (median) and variability (inter-quartile range, and usually the maximum and minimum values). The box plot helps to find outlier samples. For

example, in Figure 6.1, the plots in blue colour are before normalisation and pink are experiments after background correction and within array normalisation. The plots clearly show that there is no outlier of the samples.

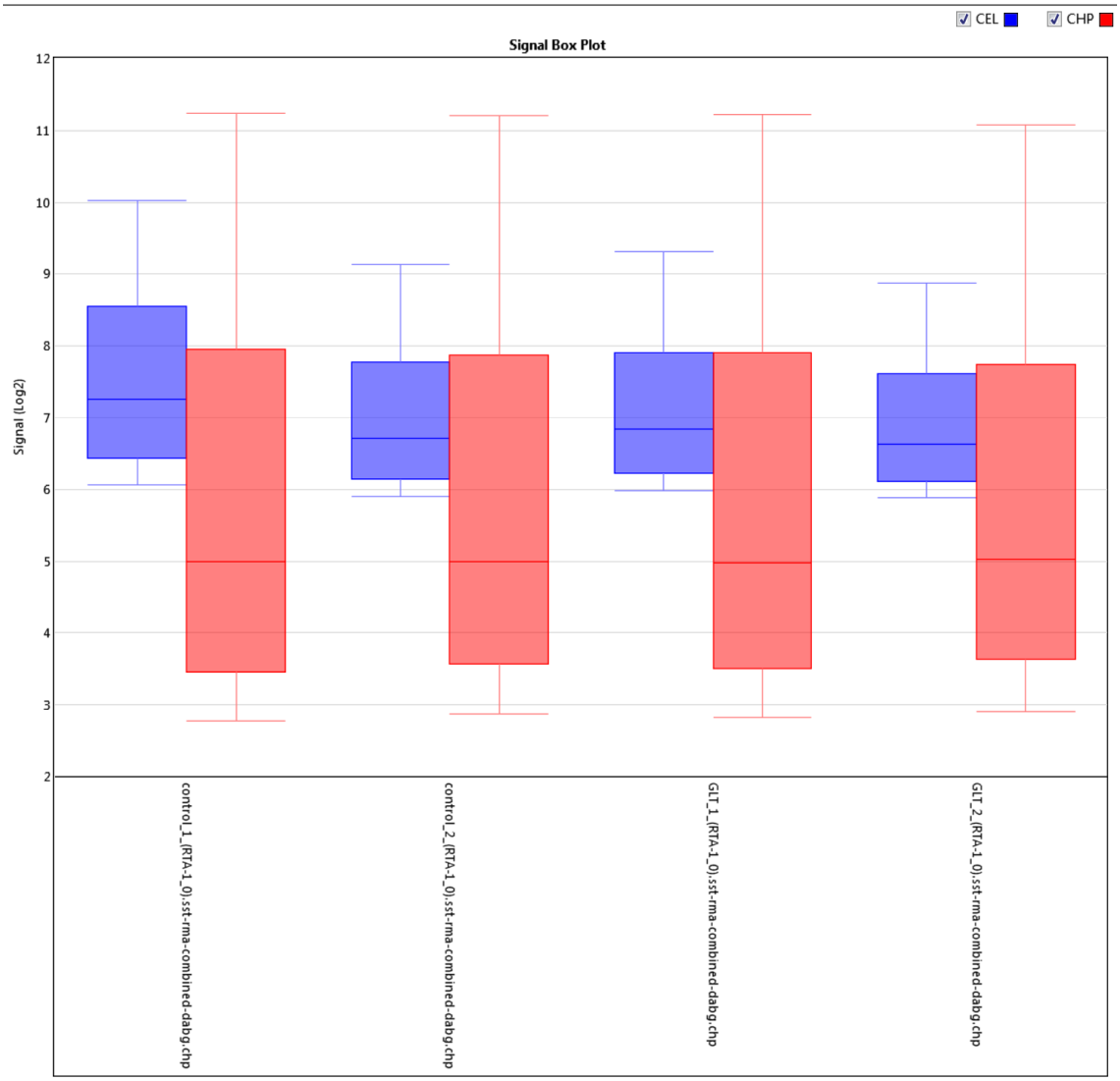


Figure 6.1: Data normalisation in Signal Box Plot. Normalisation compares between CEL and CHP files in Signal Box Plot. The plots (pink) clearly show that there is no outlier of the samples, the median values are similar across the array and the data distribution is normal after normalisation.

6.1 Gene level analysis

Default criteria was used for the gene filtering process and the criteria was: fold changes >2 or <-2 and p value <0.05. With this default filtration, the total number of genes passed through the filter was 68,842. A total of 911 genes were differentially expressed following GLT treatment, among them 402 genes were upregulated, and 509 genes were downregulated. The pie charts are showing the distribution of differentially expressed genes in altered metabolic conditions (Figure 6.2).

To further compare transcript expression between control and treated samples, scatter and volcano plots were generated. Figure 6.3 shows genes that are upregulated (red) and downregulated (green) in GLT compared to control conditions; the grey transcripts were filtered out by the default criteria. The volcano plot (Figure 6.3) combines statistically significant p values and fold changes, allowing quick visual identification of thousands of datapoints between two conditions. The highly significant values are found at the top of the plot. Due to quick visual identification of thousands of replicates results, it is very popular among genomics and metabolomics (Li, 2012).

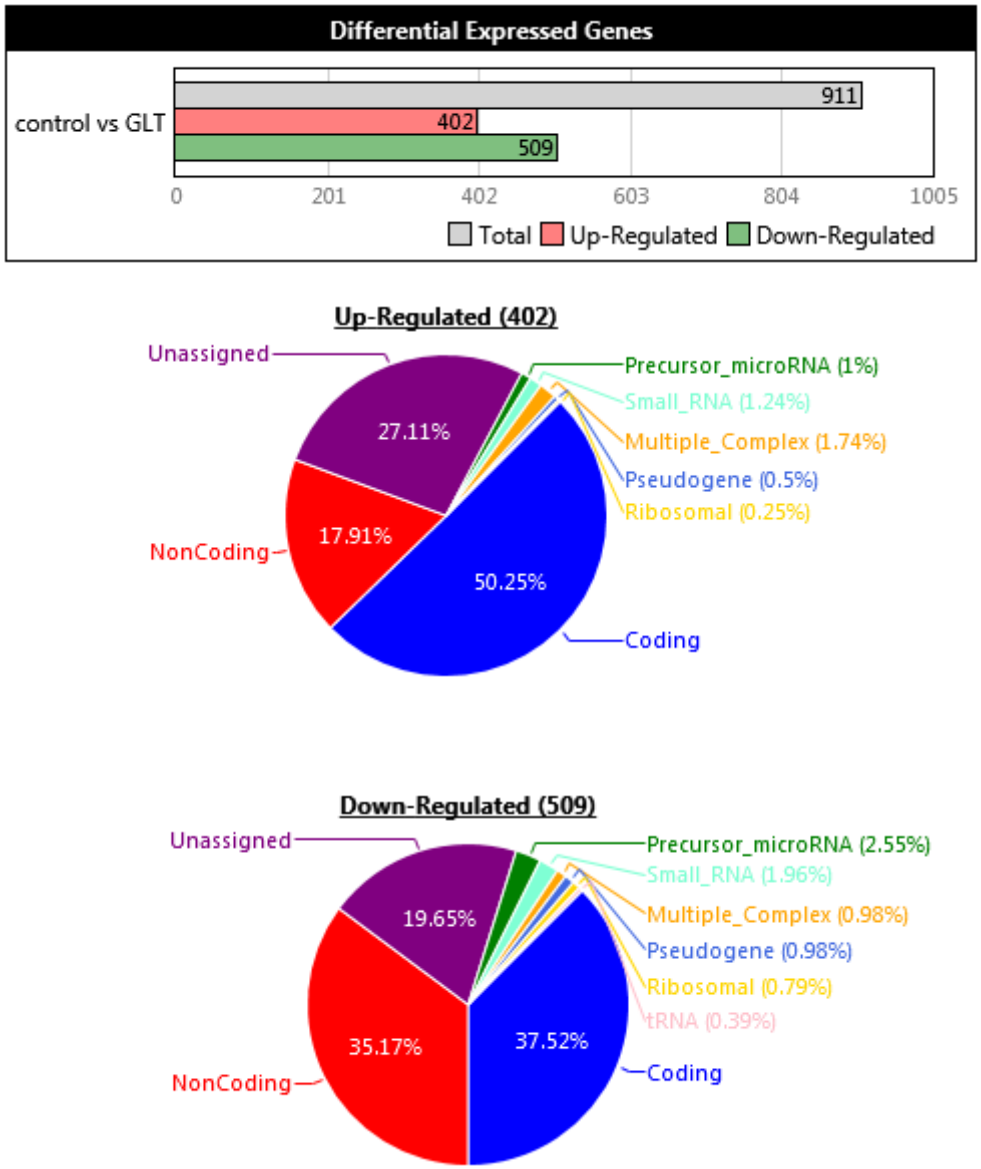


Figure 6.2: Pie charts showing a gene view summary of differentially expressed genes in metabolic conditions. Based on default filter criteria, the analysis is showing the breakdown of a total of 911 differentially expressed genes in glucolipotoxic conditions, of which 402 are up-regulated genes and 509 are down-regulated genes. The pie chart provides the percent of type of transcripts within the differentially expressed genes for example, coding (protein coding transcripts), non-coding, precursor-microRNA (gene contains microRNA precursor), pseudogenes, ribosomal (gene contains only rRNA), small RNA, tRNA and unassigned (locus type cannot determine based on source data).

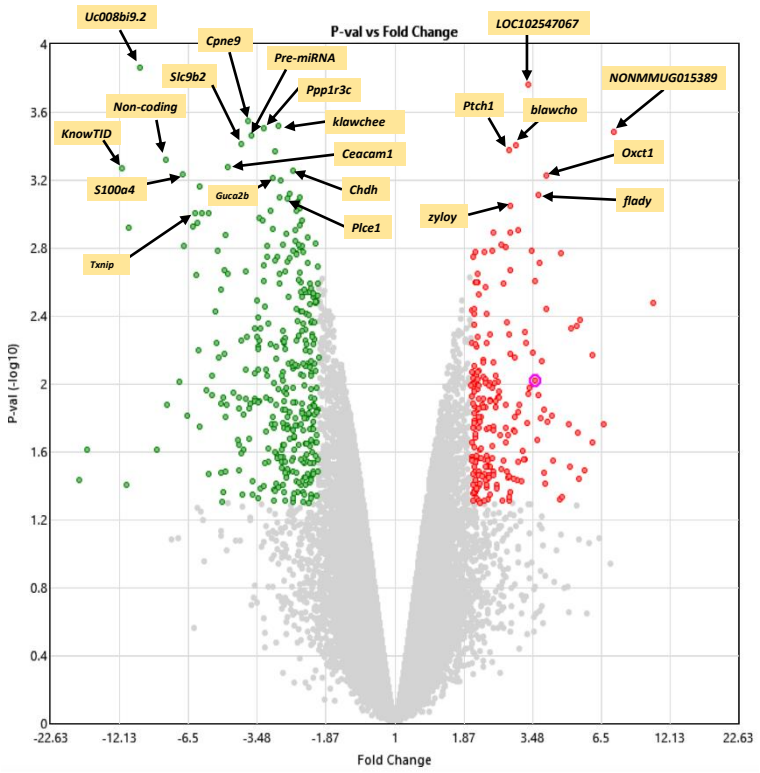


Figure 6.3: Volcano plot of differentially expressed genes under GLT conditions (Con vs GLT). For volcano plot, significance p value versus fold change on the y and x axis; n=2. Each dot represents a gene dot. The dot represents a gene; red dots are genes overexpressed and dots in green line are downregulated on glucolipotoxic conditions. Some top regulated genes are highlighted in figure.

This analysis shows that the number of genes that are highly differentially expressed is quite low, and the statistical analysis performed by the software indicated that there are no gene expression changes that exceed correction for genome-wide significance (FDR<0.05). This is not unexpected given the very small sample number, since this experiment was designed to gain initial insight into pathways affected by alternative splicing rather than as an end-point analysis for gene expression changes. Interestingly, *Elavl1* was upregulated 1.8-fold in the array analysis, consistent with the qRT-PCR analysis described in section 4.8. The rest of the SFs (Table 4.6) were also similar at the gene level to the Clariom data.

As shown in Table 6.9, a Wiki pathway, which is a website that is maintained by the biological community that shows biological pathways, analysis of the

differentially expressed genes meeting both fold change and significance criteria shows expected changes in lipid/cholesterol and cellular metabolism as well as changes in the MAPK signalling pathway. In Wiki pathways, genes are ranked based on tissue expression profile.

Table 6.9: Summary of some top hit metabolic and apoptotic enzymes in WikiPathways analysis.

Signalling pathway	**Genes differentially expressed**	**Description**	**Regulation on GLT**
Cholesterol Biosynthesis	*Lss*	Involved in cholesterol biosynthesis	up
Mitochondrial LC-Fatty Acid Beta Oxidation	*Acadl*	Fatty acid beta oxidation	Down
Lipid metabolism	*Gpd1*	Involved in carbohydrate metabolic process	Down
TCA cycle	*Pdhb*	Involved in acetyl CoA biosynthetic process from pyruvate	Down
Cell Cycle	*Abl1, Ccnd2* (also downregulated in Wnt signalling pathway)	Involved in cellular response to oxidative stress, found in endoplasmic reticulum/ Involved in cellular response to insulin stimulus	Down
Fatty Acid Biosynthesis	*Acaca*	Acetyl CoA metabolic process, FA biosynthesis process, insulin signalling pathway and associated with fatty liver and insulin resistance	Down
Fatty Acid Biosynthesis	*Acly*	FA metabolic process	Down
MAPK Signalling Pathway	*Rasgrp2*	Ca ion binding and positive regulation of GTPase	Down
MAPK Signalling Pathway	*Ntrk1*	Participates in apoptotic cell death	Up

MAPK Signalling Pathway	*Cacna1a*	Involved in calcium channel voltage dependent processes	Down

6.2 Analysis of enriched Pathways

The program we used for this analysis was PANTHER (see Method Section 2.4). We uploaded our 911 differentially expressed genes (Section 6.1) and identified the enriched pathways of molecular functions, cellular components and biological process as well as the results are displayed in a pie chart.

6.2.1 Recognition of molecular functions and cellular components in DE genes

Figure (6.4) shows the molecular functions of differentially expressed genes on altered metabolic conditions. Among those DE genes, the functions mostly affected by glucolipotoxicity were catalytic activity, binding, receptor and transporter activity.

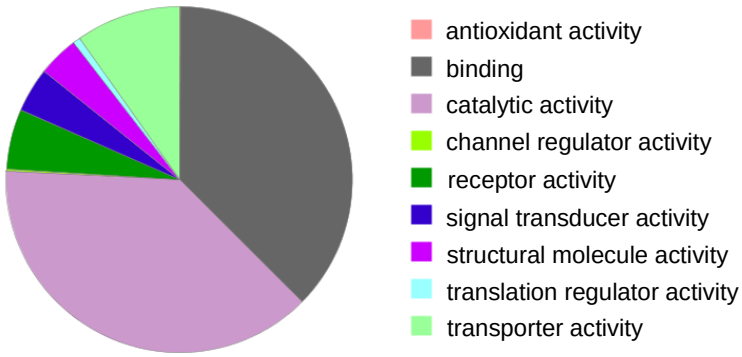


Figure 6.4: The pie chart represents the enriched molecular functions among the differentially expressed genes on GLT. The analysis indicates the association of differentially expressed genes with the molecular functions.

Catalytic activity such as enzyme regulator activity, transferase activity and oxidoreductase activity could be affected due to metabolic conditions. Another enriched molecular function for DE genes was binding activity, which is in line

with many molecules implicated in the impairment of calcium ion binding (MAPK signalling pathway; Table 6.9), calcium dependent phospholipid binding, carbohydrate, lipid, nucleic acid, nucleotide, oxygen and protein biding etc in insulin secretion and metabolic pathways, and consequently receptor activity and transporter activity, for example, glucose and lipid transporter activity along with transmembrane transporter activity are also impaired in β cells in T2D (Bagnati *et al*, 2016 and Saltiel and Khan, 2001).

Figure 6.5 shows the majority of differentially expressed genes are located on cell part (any constituent part of the cell) and on cell junction (multiprotein complexes that provides communication between a cell and extracellular matrix and between cell-cell communication) which is in line with the impairment of insulin secretion, glucose transport, and consequently receptor and glucose transporter activity are impaired in T2D (Saltier and Khan, 2011). Macromolecular complexes for example, ribonucleoprotein complexes (e.g. RBPs), protein complexes and protein-DNA complexes are affected by altered metabolic conditions (Chapter 4).

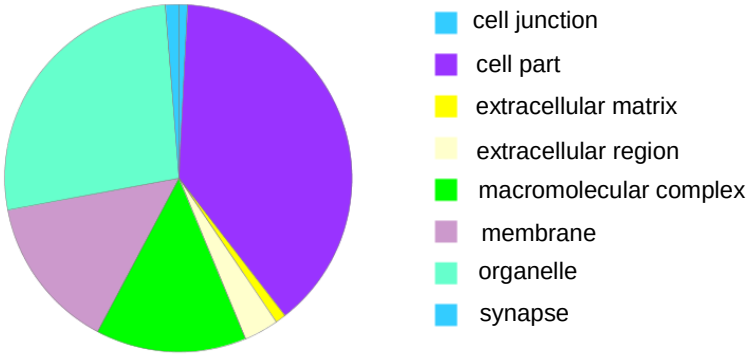


Figure 6.5: The pie chart presenting the enriched cellular component, where a gene acts in the cell, among the differentially expressed genes in high glucose and FFAs conditions.

6.2.2 Identification of enriched biological process

The enriched biological process (Figure 6.6), some biological phenomena that affects the state of an organism, is mostly affected in high sugar and fatty acids conditions were cellular and development process such as cell communication, cell cycle, cell growth, cell proliferation, cell recognition, cellular component movement, cell differentiation and death. These functional enrichments in differentially expressive genes could be explained by the impairment of β cell function and proliferation. Impairment in the cell cycle and cell growth could lead β cell death or trigger apoptosis. The over expression of *Tcf7l2* short isoform (Section 5.2) and *Elavl1* (Section 4.7) support the detrimental effects of glucolipotoxicity in β cell. The growth curve analysis also supports the adverse effects of high sugar and fatty acids on β cell growth and viability. DE genes in biological process, such as the homeostasis process, reflects impairment of glucose homeostasis in altered metabolic conditions.

Differentially expressed genes on GLT conditions in localisation such as RNA and protein localisation and transport, are in line with the finding of the overexpression of the stress-associated mRNA processing and stability gene *Elavl1* in β cells (Section 4.7). The Wiki pathways analysis (Table 6.9) supports the enriched functions of DE genes in metabolic processes affected by glucolipotoxicity.

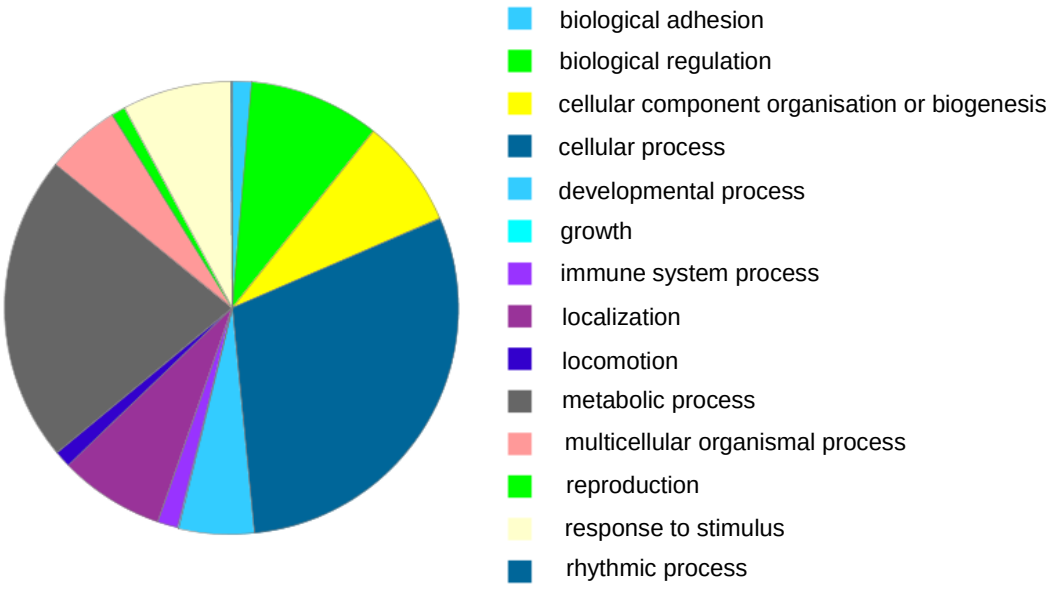


Figure 6.6: Enriched biological process involved in altered metabolic conditions. The pie chart is representing biological process affected by glucolipotoxicity and another dimension of biological function expressed by differentially expressed genes under that conditions.

6.3 Top hit results associated with diabetes, metabolism and apoptosis

A 'Comparison Wizard' is a tool to customise the analysis which can decrease the number of contrasts to simplify the results. The contrast based on the question queries the differential expression between two groups, for example, the 'control' and 'GLT' groups. By using this tool, we analysed our data to find the differential expression genes on GLT conditions. The top hit results have been summarised (Table 6.10) with corresponding fold changes and p values in conditions of high glucose and free fatty acids compared to controls. The results suggest the involvement of the specific genes in glucose metabolism (glycolysis), calcium and calcium signalling pathways, insulin signalling pathways, adipose tissue development, cell apoptosis regulation process and associate with T2D.

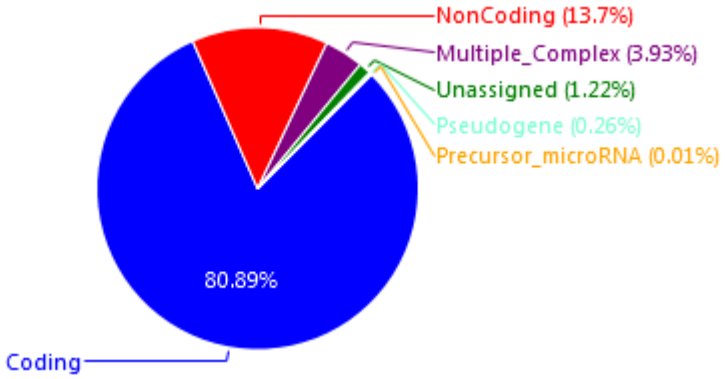
Table 6.10: List of top hits array results on GLT conditions in β cells; n=2 (con vs GLT).

Gene name	Description	Regulation	Fold Change	P value
Ppp1r3c	Involved in regulation of glycogen biosynthetic process and glycogen catabolic process; participates in insulin signalling pathway.	Down	3.28	0.0003
Slc9b2	Interacts with nuclear gene encoding mitochondrial protein; integral component of membrane	Down	4.04	0.0004
Ptch1	Encode a protein that exhibit cholesterol binding; involve in response to drug and mechanical stimulus	Up	2.82	0.0004
S100a4	Calcium binding protein; participates in calcium/ calcium mediated signalling pathway	Down	6.86	0.0006
Oxct1	Involved in adipose tissue development/ participates in ketone bodies metabolic pathways/ associates with T2D	up	3.93	0.0006
Guca2b	Participates in calcium and calcium signalling pathway	Down	3.02	0.0006
Lgals5	Encodes a protein that exhibits carbohydrate binding	Down	2.38	0.0008
Plce1	Involved in activation of MAPK activity; participates in calcium and calcium mediated signalling pathway	Down	2.64	0.0008
Txnip	Involved in apoptosis regulation process	Down	6.12	0.001

6.4 Exon level analysis of alternative splicing

For the exon level analysis, which allows identification of alternative splicing events in a specific gene, we again used the default criteria of an exon splicing index (The fold change of normalised expression levels for condition 1 versus condition 2 of the PSR or junction's expression level which is normalised to the gene's expression level) of +/- 2 fold, and an exon p value <0.05. Of the 68,599 genes, 8,958 (13.06%) met these criteria of exon level changes (Figure 6.4). Exon p value is the probability that there is no actual difference in the expression for the PSR and any observed difference is due to variance in the measurement (TAC 4.0.1 user guide; Thermo Fisher Scientific).

A)



B)

Events	Count
All	77
Alternative 5' Donor Site	24
Intron Retention	19
Cassette Exon	19
Alternative 3' Acceptor Site	15

C)

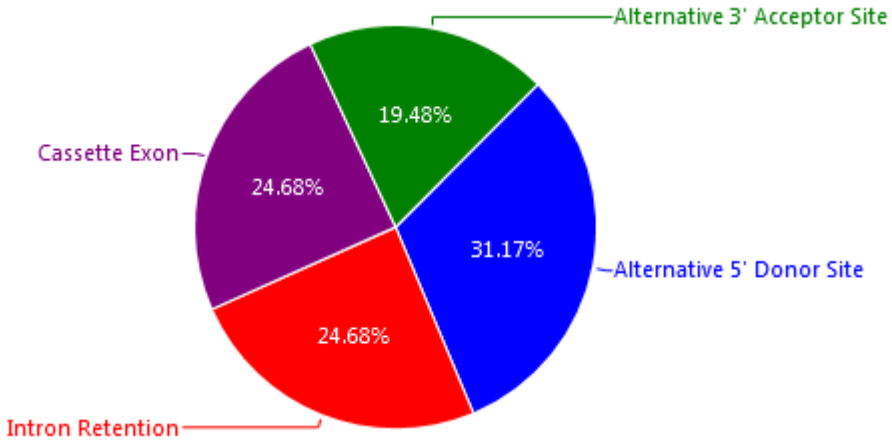


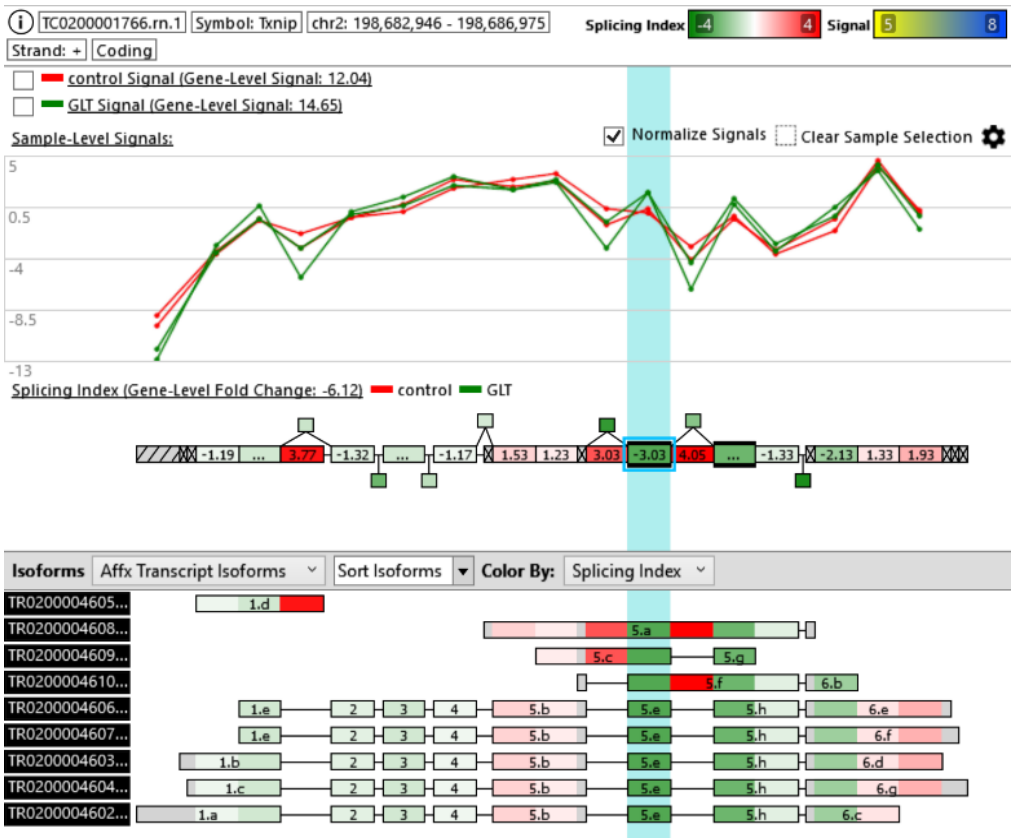
Figure 6.7: Exon expression analysis of alternative splicing on glucolipotoxic conditions (control vs. GLT). A) The pie chart provides a breakdown of genes by groups meeting the filter criteria **B)** Around 77 events were tested and genes above 13% passed the filter which means it can be considered to be alternatively spliced following GLT treatment in rat β cells C) the breakdown of identified standard alternative splicing events i.e. cassette exon, mutually exclusive exons, alternate 5' donor site, alternate 5' acceptor site and intron retention.

However, only 74 of these genes were assigned an exon score (consensus alternative splicing event score), representing the set of genes that are most likely to present a genuine alternative splicing event. The score is value from 0 to 1, where 1 represents a high probability for the event. Of these 74 genes, the top 5

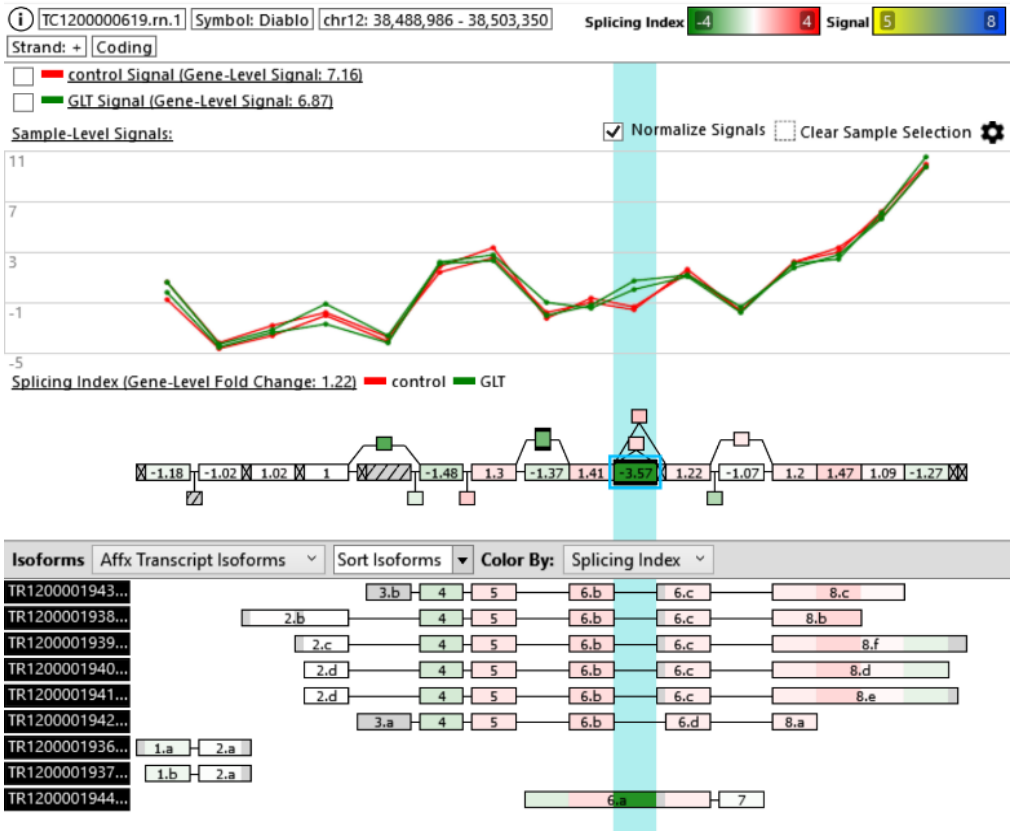
were: *Txnip, Diablo, Cyp4f1, Tsc22d1 and Cox15.* Three of these (*Txnip*, *Diablo*, *Tsc22d1*) are directly involved in the regulation of apoptosis, while *Cyp41* is involved in cytochrome c synthesis, potentially providing an additional link to the mitochondrial pathway of apoptosis (gene + exon; total 911 filtered genes).

The structure view of the TAC software was used to visualise the specific splicing events identified, as shown in Figures 6.5, where a cross-out box represents no data, a diagonally crossed-out box represents that is not expressed in either condition and PSR/junction indicates by box containing splicing index number on the top. Isoforms are displayed in table below the structural view; transcript isoforms are sorted as ‘up regulation’ on GLT conditions at the top and ‘down regulated’ are at the bottom. The colour coding is associated with exons; transcripts containing red exons are overexpressed and green exons are downregulated (Figure 6.5).

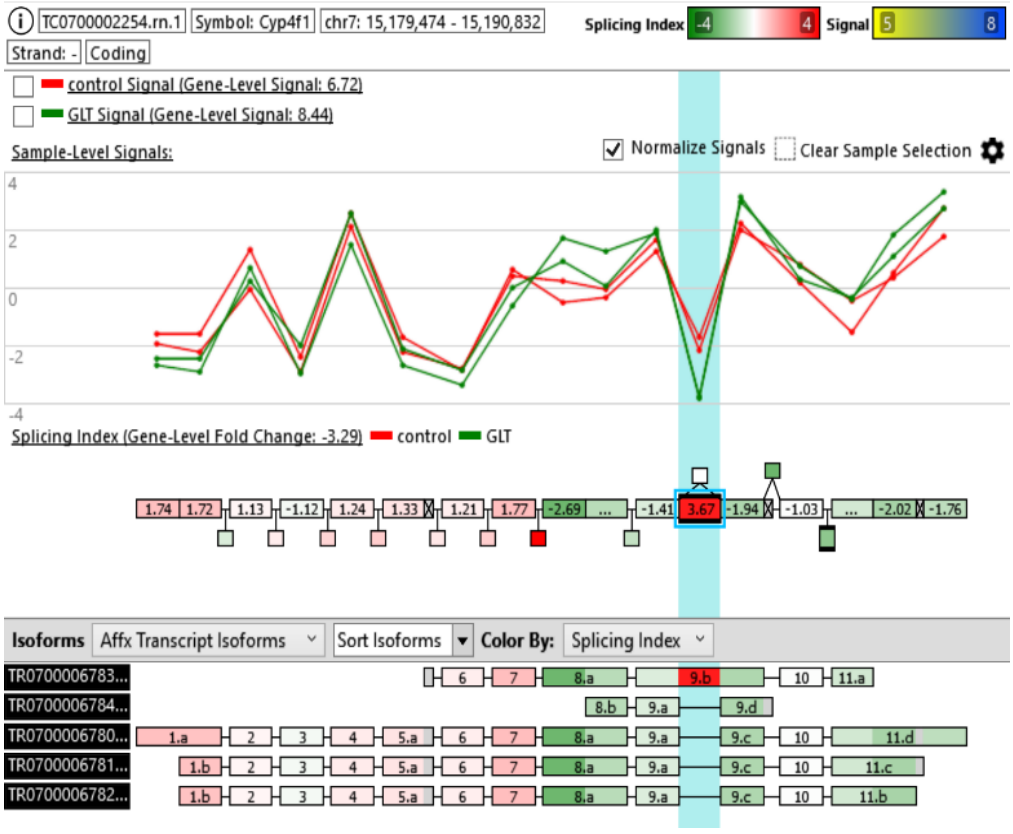
A)



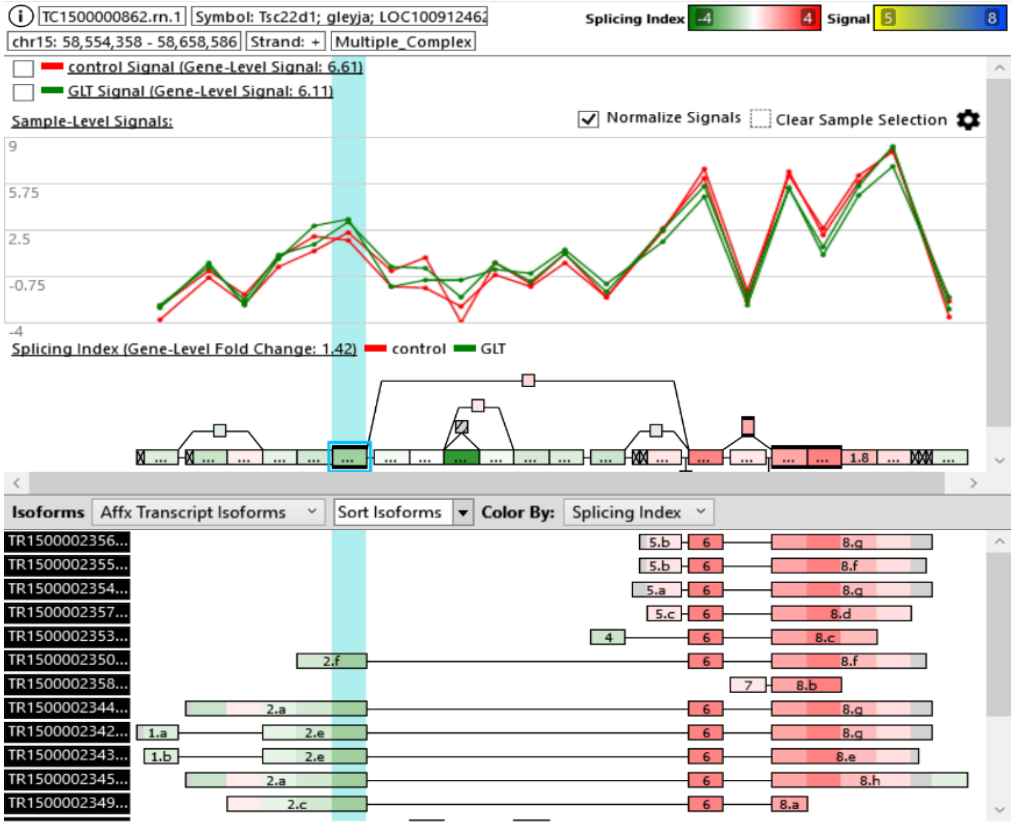
B)


TC1200000619.rn.1
Symbol: Diablo
chr12: 38,488,986 - 38,503,350
Splicing Index
Signal
Strand: +
Coding
control Signal (Gene-Level Signal: 7.16)
GLT Signal (Gene-Level Signal: 6.87)
Sample-Level Signals:
Normalize Signals
Clear Sample Selection
Splicing Index (Gene-Level Fold Change: 1.22)
control
GLT
Isoforms
Affx Transcript Isoforms
Sort Isoforms
Color By:
Splicing Index

C)


TC0700002254.rn.1
Symbol: Cyp4f1
chr7: 15,179,474 - 15,190,832
Splicing Index
Signal
Strand: -
Coding
control Signal (Gene-Level Signal: 6.72)
GLT Signal (Gene-Level Signal: 8.44)
Sample-Level Signals:
Normalize Signals
Clear Sample Selection
Splicing Index (Gene-Level Fold Change: -3.29)
control
GLT
Isoforms
Affx Transcript Isoforms
Sort Isoforms
Color By:
Splicing Index

D)


TC1500000862.rn.1
Symbol: Tsc22d1; gleyja; LOC10091246
chr15: 58,554,358 - 58,658,586
Strand: +
Multiple_Complex
Splicing Index
Signal
control Signal (Gene-Level Signal: 6.61)
GLT Signal (Gene-Level Signal: 6.11)
Sample-Level Signals:
Normalize Signals
Clear Sample Selection
Splicing Index (Gene-Level Fold Change: 1.42)
control
GLT
Isoforms
Affx Transcript Isoforms
Sort Isoforms
Color By:
Splicing Index

E)

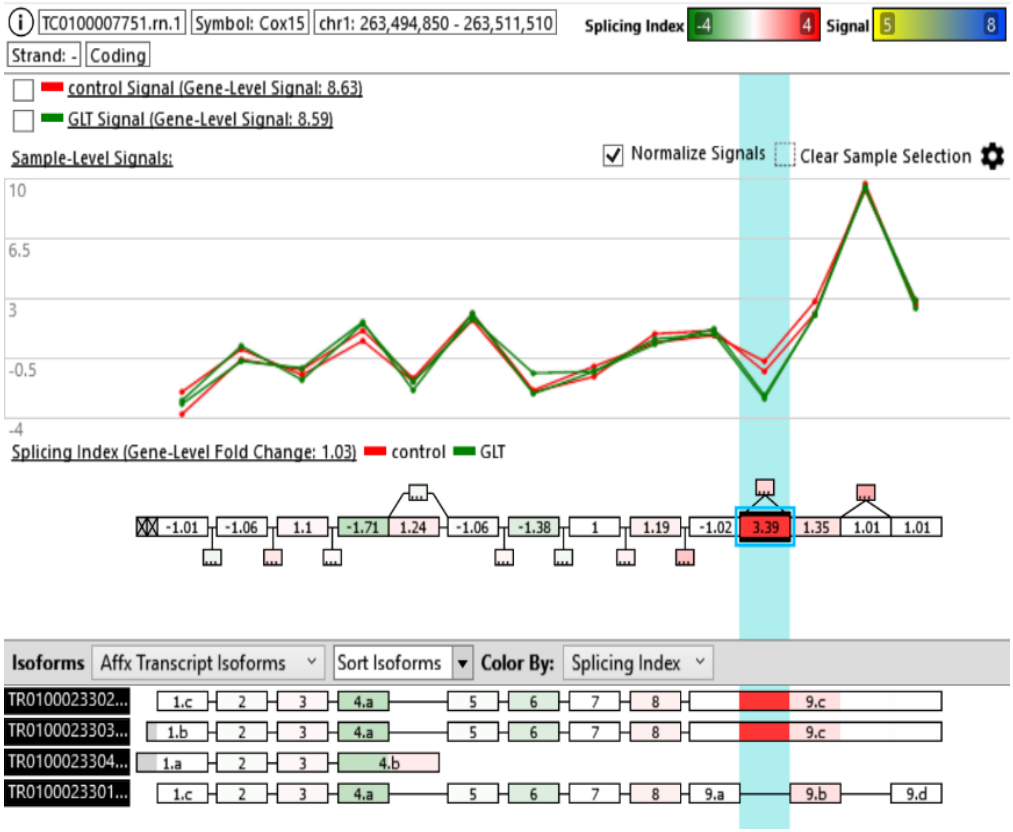


Figure 6.8: Exon expression analysis of AS of top 5 genes on GLT conditions. Image showing the normalised view of exon splicing. For each PSR/junction, fold change measurement is based on colour and the intensity of colour reflects the amount of fold change, with the pink line showing a selected sample, the red being control and the green GLT. Transcript isoforms can be viewed based on database selection, we have sorted isoforms based on Affymetrix database. **A)** AS events of *Txnip* which involved in positive regulation of apoptotic process, cell cycle proliferation and response to calcium ion. **B)** As events of *Diablo* involved in intrinsic apoptotic signalling pathway in response to oxidative stress. **C)** As events *Cyp4f1* response to toxic substance and participates in cytochrome P450 monooxygenase mediated pathway of arachidonic acid metabolism and arachidonic acid metabolic pathway. **D)** AS events of *Tsc22d1* involved in positive and negative regulation of apoptotic process and positive regulation of cell proliferation. **E)** AS events of *Cox15* involved in oxidative phosphorylation pathway and associated with Cytochrome-c Oxidase deficiency.

6.4 Discussion

The Clariom D array provided a complete view of all the differentially expressed and alternatively spliced genes in a single experiment. The main purpose was to discover novel alternatively spliced molecules that could be involved in the impairment of β cell function after exposure to high concentrations of glucose and fatty acids and use these results in future experimental designs. We initially focused on the general characterisations of differentially expressed genes and pathways analysis to identify the novel mechanism affected by GLT.

The pathway analysis has shown the most dysregulated genes which ultimately reflects the pathways that are differentially affected by glucolipotoxicity. For example, differentially expressed or dysregulated genes observed in cholesterol biosynthesis, mitochondrial long chain fatty acid beta oxidation, TCA cycle and fatty acid biosynthesis which reflects the consequence of altered metabolic conditions relevant to T2D in pancreatic β cells. On the other hand, differentially expressed genes in the cell cycle, Wnt signalling pathway and MAPK signalling pathway reflect the adverse effect of GLT on β cell proliferation, insulin sensitivity and insulin resistance; these pathways are also found to be affected in obesity, hyperlipidaemia, hypertension and T2D (Saltiel and Khan, 2001). Our Wiki pathways analysis results and the top hit array results suggest the involvement of the specific genes (DEs) in glucose metabolism, calcium and calcium signalling pathways, insulin signalling pathways, adipose tissue development, cell apoptosis regulation process and associate with T2D. The identification of enriched pathways analysis has also shown the enriched molecullar, cellular and biological functions of DE genes.

Our exon level analysis of alternative splicing results has shown the association of alternative splicing of top five genes: *Txnip*, *Diablo*, *Cyp4f1*, *Tsc22d1* and *Cox15*, among them *Txnip*, *Diablo, Cyp4f1* and *Tsc22d1* are directly involved in the regulation of apoptosis suggesting the association of β cell apoptosis with alterted metabolic conditions in T2D. Whereas altered isofoms formation of

Cyp4f1 and *Cox15* reflects the alternative splicing pattern changes to physiological and environmental stressors.

The biological replicates (n=2) for exon-array analysis are not enough for a powerful statistical test over fold changes threshold for selecting DE genes; the threshold incorporates variation between measurements, estimate for error rate, detection of minor changes and the ranking of DE genes. In future, more replicates should have been analysed for more powerful statistical results.

6.4 Limitation:

- Only two replicates in each category were analysed for genome-wide analysis that underpowers the Clariom genome wide analysis.
- Although multiple testing has not been applied, it needs to be recognised that over 68,000 transcripts tested from where 700 could appear positive by chance at p= 0.05 and that is why these findings should be viewed with caution. Same issue of multiple testing needs to be recognised for specific isoforms testing.
- Need to look in more detail at the type of splicing event being indicated for top 74 AS hits which is essential to assess what impact these treatments have had on the splicing patterns in rat β cells.
- No alternative splicing events were analysed for *Tcf7l2* in this array since only a single transcript is reported in the database.
- Needed to validate array hits by RT-qPCR but due to time constrains we could not validate the DE and alternatively spliced genes under metabolic conditions.

In conclusion, our transcriptome analysis has shown that metabolic conditions alter the alternative splicing regulation of some apoptotic genes which support our finding of *Elavl1* overexpression on these conditions which suggests that high sugar and fatty acids may contribute to β cell loss in some peole with type 2 diabetes.

Chapter 7- Discussion

According to International Diabetes Federation (IDF) approximately 425 million adults were suffering from diabetes in 2017 and predicts that the number will increase to 629 million by 2045. This issue needs to address urgently. Both genetics and epigenetics are interacting factors where unhealthy lifestyle influences. We hypothesized that the metabolic conditions are associated with T2D may affect the regulation of alternative splicing, a regulated process during gene expression that results in a single gene coding for multiple proteins and a crucial mechanism for gene regulation and for proteomic diversity. Little is currently known about the effect of metabolic changes that exist in many people with T2D on splicing in β cells. Therefore, we aimed to investigate how this may link to decline of β cell mass and function which can pave the way to future studies in order to develop therapeutic strategies that are able to restore β cell function in T2D.

Based on our research design - at first, we wanted to test the effect of glucolipotoxic conditions on β cell viability. We replicated *in vitro* diabetic extracellular glucolipotoxic conditions in INS1 rat pancreatic β cells (insulinoma) and treated them with GLT media (Bagnati *et al*, 2016) to analyse the toxicity effect on β cell growth and compare that with control by a statistical end-point growth analysis. We found that this condition decreased the viability of INS1 cells both in serum free and complete. A statistical end-point analysis has shown a significant decrease in cell numbers in both serum free ($p=0.0006$, $n=6$) and complete media ($p=0.002$, $n=6$) in 2/3 experiments, suggesting that metabolic conditions may affect the pancreatic β cell growth. A previous study by Zhou *et al* (2014) demonstrated that an increased apoptosis with progressive activation of caspase 3 happened in INS1 cells when treated with 30 mM glucose and 200 µM palmitic acid, supporting our outcome in terms of decrease viability of INS1 cells under our experimental toxic conditions (INS1 cells treated with 27mM glucose, 200 µM palmitic acid and 200 µM oleic acid), although they did not use oleic acid and they used an extra 3 mM glucose in their experiment. It is known

that there is an antagonism between saturated and unsaturated fatty acids (Gehrmann *et al*, 2015) but it is more physiological to use palmitic and oleic acid both in *in vitro* experiment. Our growth curve experiment was well designed because there was no missing data from replicates, timepoints were equally spaced, data was normally distributed, and experiment was repeated 2/3 times with six technical replicates. Hence, growth curve analysis suggests that INS1 average cell area decreased under altered metabolic conditions.

The growth curve titration helped us to choose a timepoint (46 hours) to analyse our selected panel of splicing regulators on GLT conditions by RT-qPCR. Among the selected splicing factors our data analysis has demonstrated the overexpression ($p=0.04$) of HuR family gene *Elavl1* under altered metabolic conditions.

Therefore, in this study, we have shown that metabolic conditions associated with T2D, specifically high levels of glucose and free fatty acids in presence of control, affect the global regulation of alternative splicing in β cells. Specifically, expression of the mRNA processing and stability regulator *Elavl1* is upregulated by GLT treatment of INS1 β cells *in vitro*. No study of *Elavl1* expression under GLT conditions in INS1 β cell has been published yet.

Despite HuR's predominant nuclear localisation and extensive characterisation, the functions of this family genes as splicing regulator still not identified yet. One of the reasons may be no target sequences for this regulatory gene has been discovered yet (Izquierdo, 2008).

Elavl1 has previously been shown in other cell types to regulate the alternative splicing of the apoptotic death receptor *Fas* and *Tra2β*, itself an important trans-acting regulator of alternative splicing. *Elavl1* is also known to regulate mRNA stability of target genes under conditions of stress. It will be important in the future to determine the role of *Elavl1* in β cells. This has not previously been examined,

although work by Mateu *et al* (2017) has shown that siRNA knockdown of the related family member *Elavl4* modulates β cell death. This supports the hypothesis that GLT stress in β cells activates a programme of alternative splicing events, mediated by *Elavl1* upregulation, that activate β cell apoptosis.

We have also demonstrated that reference gene *Ppia* expression is stable under altered metabolic conditions and conversely that *Gapdh* should not be used as a RT-qPCR loading control in this type of study.

Diabetes candidate gene *TCF7L2* is an important factor determining T2D risk and is known to be regulated by alternative splicing. We investigated whether altered metabolic conditions and overexpression of splicing regulatory gene *Elavl1* were associated with changes in *Tcf7l2* isoform expression. We worked with rat INS1 cell line and there were no splice variants have been reported for rat *Tcf7l2* in the database. However, in human genomic structure of *TCF7L2* consists of 17 exons (Database reported 10 exons for rat *Tcf7l2*) and only four splice variants were reported to be expressed in human β cells by Hanson *et al*, 2010. By an end-point PCR/ conventional PCR followed by Sanger sequencing we first discovered three novel *Tcf7l2* isoforms in rat INS1 β cells with the identification of three new novel exons (Exon 8b, exon 8c and exon 9b).

Therefore, we, have identified three novel rat isoforms (Isoform B, C and D) of *Tcf7l2* that are not present in the database, and we have further shown that GLT treatment appears to regulate the pattern of *Tcf7l2* isoform expression. By comparison with the functional analysis of human *TCF7L2* isoforms by Le Bacquer *et al* (2011), our data suggests that GLT treatment induces expression of a *Tcf7l2* isoform (short isoform B) that induces β cell apoptosis. This will require further experimental validation but provides further support for the link between GLT stress, alternative splicing and β cell apoptosis. Therefore, in future, we need to examine the functional consequence at the protein level for all three new isoforms in rat pancreatic β cells.

The top hit results at the exon level analysis (Microarray) support the involvement of the specific genes (DEs) in glucose metabolism, calcium and calcium signalling pathways, insulin signalling pathways, adipose tissue development and cell apoptosis regulation process associate with T2D. The identification of enriched pathways analysis has also shown the enriched molecular, cellular and biological functions of DE genes. On the other hand, Wiki pathway analysis results of DE genes have shown the differentially expressed genes in cell cycle, Wnt signalling pathway and MAPK signalling pathway suggesting the adverse effect of altered metabolic conditions in β cell.

Further evidence supporting our hypothesis was provided by the exon-level analysis of alternative splicing at the global level by GLT conditions. Four of the top five alternative splicing events (*Txnip*, *Diablo*, *Cyp4f1* and *Tsc22d1*) were genes linked to the regulation of apoptosis. However, it will be important to confirm the induction of these specific splicing events by GLT conditions using RT-qPCR, and in the long-term to determine the effect of the induced isoforms on β cell apoptosis.

Hence, the exon level analysis is suggesting that alternative splicing pattern changes continuously in respose to physiological conditions and environmental stressors which activating cells to react many challenges by differential expression of genes and by altering the balance between different mRNA isofomrs.

The results of this study have potential relevance for people with pre-diabetes and will be important to examine whether similar changes occur in human β cell lines, and in β cells isolated from people with diabetes. However, there are significant barriers to obtaining such samples, which can only be obtained post-mortem. However, if supported by future experimental results, this study suggests that the AS pathway in β cells, perhaps specifically *Elavl1*, could

potentially be used to promote β cell survival in people with pre-diabetes. This will require further experiments to evaluate this possibility, and to identify viable approaches to pharmacological regulation and delivery.

In conclusion, this study demonstrates that:

- High glucose and fatty acid (GLT) metabolic conditions induce expression of the RNA binding protein / splicing regulator *Elavl1* in β cells.
- GLT conditions increase expression of *Tcf7l2* transcript variants that may induce β cell apoptosis rather than enhancing insulin secretion.

Exon array data demonstrates that GLT treatment induces alternative transcript expression of a panel of apoptotic regulatory genes in β cells. This Suggests that *Elavl1* may be a novel master regulator of alternative splicing of apoptotic genes in β cells during altered metabolic conditions, and that alternative splicing of apoptotic genes in β cells may contribute to β cell loss in people with metabolic syndrome and pre-diabetes.

7.1 Limitation

- To study glucolipotoxicity on the pancreatic β cell, we used the rat INS1 β cell line. However, we could have used at least another rat beta cell line e.g. RIN, MIN, HIT etc to see if similar changes occur in our study. It would be very useful to examine the results in human beta cell line EndoC-βH1 which can be a part of future study.

- Our data showed the GLT effect on the expression of 13 splicing regulatory genes among about 300 SFs. There may be some notable omissions that are expressed in β cells for example, some negative *trans*-acting factors hnRNPs (*Hnrnpa1* and *Hnrnpa2b1)* which are found to be associated with cancer and ageing in human, should be tested under GLT conditions. In

future, it could be interesting to investigate more SFs and the Clariom data by selecting and testing SFs in altered metabolic conditions.

- The SFs expression changes were only investigated at 46-hour point, could have missed the detection of some important mechanisms at the early or at later stage. In future, the results we observed at 46-hour time point could investigate at other time points as well to investigate their molecular mechanisms at different time point of β cell growth under GLT conditions.

- More than one reference gene should be used in relative gene expression analysis by RT-qPCR. We have used only one reference gene *Ppia* in our study. In future study we can use more than one reference gene in RT-qPCR analysis.

- The biological replicates for the analysis of splicing regulators' expression by RT-qPCR under altered metabolic conditions were enough to carry on a two-tailed paired sample t-test. However, for a powerful statistical analysis biological replicates should be higher.

- The functional consequence of *Tcf7l2* isoform results is based on comparative isoforms study in humans (Le Bacquer *et al,* 2011). Despite rodent and human genome share high similarity there are some species related differences. To confirm this functional effect we need to see the protein level expression in rat β cells. It can be a part of future study.

- The Clariom analysis for genome wide analysis with only 2 replicates in each category is not enough for a powerful statistical analysis. Although the purpose of exon array analysis was to have a future experimental design, however for this kind of high throughput genome wide analysis the biological replicates should be higher.

- The top hit results of differentially expressed genes and alternatively spliced isoforms has to be confirmed by RT-qPCR. Due to time constrain we were not able to carry on that validation work but could be done in future.

7.3 Conclusion

This study demonstrates that the metabolic conditions existing in many people with type 2 diabetes may cause aberrant alternative splicing in β cells increasing susceptibility to apoptosis and subsequent β cell loss.

Chapter 8 – References

Abdelmohsen, K. and Gorospe, M. (2010). Posttranscriptional regulation of cancer traits by HuR. *Wiley Interdisciplinary Reviews: RNA*, 1(2), pp.214-229.

Al-Ahmadi, W., Al-Ghamdi, M., Al-Haj, L., Al-Saif, M. and Khabar, K. (2009). Alternative polyadenylation variants of the RNA binding protein, HuR: abundance, role of AU-rich elements and auto-Regulation. *Nucleic Acids Research*, 37(11), pp.3612-3624.

American Diabetes Association (ADA) (1997) Expert Committee on the Diagnosis and Classification of Diabetes Mellitus. *Diab Care*, 20: 1183-1197.

American Diabetes Association (ADA) (2005). Diagnosis and classification of diabetes mellitus (position statement). *Diab Care*, 28(Suppl.l): S37-S42.

Anuradha, R., Saraswati, M., Kumar, K. and Rani, S. (2014). Apoptosis of β cells in Diabetes Mellitus. *DNA and Cell Biology*, 33(11), pp.743-748.

Asfari,M, Janjic, D, Meda, P. *et al*. (1992). Establishment of 2-mercaptoethanol-dependent differentiated insulin-secreting cell lines. *Endocrinology*, 130, 167-178.

Atkinson, M. and Eisenbarth, G. (2001). Type 1 diabetes: new perspectives on disease pathogenesis and treatment. *The Lancet*, 358(9277), pp.1428-1435.

Back, S. and Kaufman, R. (2012). Endoplasmic Reticulum Stress and Type 2 Diabetes. *Annual Review of Biochemistry*, 81(1), pp.767-793.

Bagnati, M., Ogunkolade, B., Marshall, C., Tucci, C., Hanna, K., Jones, T., Bugliani, M., Nedjai, B., Caton, P., Kieswich, J., Yaqoob, M., Ball, G., Marchetti, P., Hitman, G. and Turner, M. (2016). Glucolipotoxicity initiates pancreatic β cell death through TNFR5/CD40-mediated STAT1 and NF-κB activation. *Cell Death and Disease*, 7(8), p. e2329.

Beyer, K. and Ariza, A. (2013). Alpha-Synuclein Posttranslational Modification and Alternative Splicing as a Trigger for Neurodegeneration. *Molecular Neurobiology*, 47(2), pp.509-524.

Bettger, W.J. and McKeehan, W. L. (1986). Mechanisms of cellular nutrition. *Physiol*. Rev. 66,1-35.

Black, Douglas L. (2003). "Mechanisms of alternative pre-messenger RNA splicing". *Annual Review of Biochemistry*. **72** (1): 291–336.

Bonner-Weir, S. *et al*. 1989. Compensatory growth of pancreatic β cells in adult rats after short-term glucose infusion. *Diabetes* 38: 49–53.

Bonner-Weir, S., Life and death of the pancreatic β cells. *Trends Endocrinol Metab*, 2000. 11(9): p. 375-8.

Brinkman, B. (2004). Splice variants as cancer biomarkers. *Clinical Biochemistry*, 37(7), pp.584-594.

Bronicki, L. and Jasmin, B. (2013). Emerging complexity of the HuD/ELAVl4 gene; implications for neuronal development, function, and dysfunction. *RNA*, 19(8), pp.1019-1037.

Bustin, S., Benes, V., Garson, J., Hellemans, J., Huggett, J., Kubista, M., Mueller, R., Nolan, T., Pfaffl, M., Shipley, G., Vandesompele, J. and Wittwer, C. (2009). The MIQE Guidelines: Minimum Information for Publication of Quantitative Real-Time PCR Experiments. *Clinical Chemistry*, 55(4), pp.611-622.

Cauchi, S., Meyre, D., Dina, C., Choquet, H., Samson, C., Gallina, S., Balkau, B., Charpentier, G., Pattou, F., Stetsyuk, V., Scharfmann, R., Staels, B., Fruhbeck, G. and Froguel, P. (2006). Transcription Factor TCF7L2 Genetic Study in the French Population: Expression in Human -Cells and Adipose Tissue and Strong Association With Type 2 Diabetes. *Diabetes*, 55(10), pp.2903-2908.

Chang, S., Elemento, O., Zhang, J., Zhuang, Z., Simons, M. and Hla, T. (2014). ELAVL1 regulates alternative splicing of eIF4E transporter to promote postnatal angiogenesis. *Proceedings of the National Academy of Sciences*, 111(51), pp.18309-18314.

Chang-Chen, K.J., R, Mullur, and E, Bernal-Mizrachi, (2008). β cell failure as a complication of diabetes. *Rev Endor Metab Discord*, 9(4), pp.329-43.

Chang, S.,Elemento,O.,Zhang,J.,Zhuang,Z.,Simons,M.andHla,T.(2014).ELAVL1 regulatesalternativesplicingofeIF4Etransporterto promote postnatal angiogenesis. Proceedings of *the National Academy of Sciences*, 111(51), pp.18309-18314

Chen L, Magliano DJ, Zimmet PZ. The worldwide epidemiology of type 2 diabetes mellitus–present and future perspectives. *Nat Rev Endocrinol*. 2011; 8:228–236.

Chen, L. *et al*. 1992. Factors regulating islet regeneration in the post-insulinoma NEDH rat. Adv. Exp. *Med. Biol*. 321: 71–80; discussion 81–74.

Chen, M. and Manley, J. (2009). Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches. *Nature Reviews Molecular Cell Biology*, 10(11), pp.741-754.

Cieply, B. and Carstens, R. (2015). Functional roles of alternative splicing factors in human disease. *Wiley Interdisciplinary Reviews: RNA*, 6(3), pp.311-326.

Cnop, M. (2008). Fatty acids and glucolipotoxicity in the pathogenesis of Type 2 diabetes. *Biochemical Society Transactions*, 36(3), pp.348-352.

Cockburn, B., Ostrega, D., Sturis, J., Kubstrup, C., Polonsky, K. and Bell, G. (1997). Changes in pancreatic islet glucokinase and hexokinase activities with increasing age, obesity, and the onset of diabetes. *Diabetes*, 46(9), pp.1434-1439.

Corbo, C., Orrù, S. and Salvatore, F. (2013). SRp20: An overview of its role in human diseases. *Biochemical and Biophysical Research Communications*, 436(1), pp.1-5.

Cropano, C., Santoro, N., Groop, L., Dalla Man, C., Cobelli, C., Galderisi, A., Kursawe, R., Pierpont, B., Goffredo, M. and Caprio, S. (2017). The rs7903146 Variant in the TCF7L2 Gene Increases the Risk of Pre-diabetes/Type 2 Diabetes in Obese Adolescents by Impairing β cell Function and Hepatic Insulin Sensitivity. *Diabetes Care*, 40(8), pp.1082-1089.

Cummings, B. (2008). *Interactive physiology 10-systme suite student version*. [Place of publication not identified]: Pearson Educ.

Cunha, D., Hekerman, P., Ladriere, L., Bazarra-Castro, A., Ortis, F., Wakeham, M., Moore, F., Rasschaert, J., Cardozo, A., Bellomo, E., Overbergh, L., Mathieu, C., Lupi, R., Hai, T., Herchuelz, A., Marchetti, P., Rutter, G., Eizirik, D. and Cnop, M. (2008). Initiation and execution of lipotoxic ER stress in pancreatic β cells. *Journal of Cell Science*, 121(14), pp.2308-2318.

Curran, P., Obeidat, K. and Losardo, D. (2010). Twelve Frequently Asked Questions About Growth Curve Modeling. *Journal of Cognition and Development*, 11(2), pp.121-136.

da Silva Xavier, G., Loder, M., McDonald, A., Tarasov, A., Carzaniga, R., Kronenberger, K., Barg, S. and Rutter, G. (2009). TCF7L2 Regulates Late Events in Insulin Secretion from Pancreatic Islet cells. *Diabetes*, 58(4), pp.894-905.

Dai, W., Zhang, G. and Makeyev, E. (2011). RNA-binding protein HuR autoregulates its expression by promoting alternative polyadenylation site usage. *Nucleic Acids Research*, 40(2), pp.787-800.

Das, S. and Elbein, S. (2006). The Genetic Basis of Type 2 Diabetes. *Cellscience*, 2(4), pp.100-131.

Diiorio P, Jurczyk A, Yang C, *et al*. (2011). Hyperglycemia-induced proliferation of adult human β cells engrafted into spontaneously diabetic immunodeficient NOD-Rag1null IL2rγnull Ins2Akita mice. *Pancreas*; 40:1147–1149.

Duval A, Rolland S, Tubacher E, Bui H, Thomas G, Hamelin R. (2000) The human T cell transcription factor-4 gene: structure, extensive characterization of alternative splicings, and mutational analysis in colorectal cancer cell lines. *Cancer Res*, 60:3872–3879.

Dvinge, H., Kim, E., Abdel-Wahab, O. and Bradley, R. (2016). RNA splicing factors as oncoproteins and tumour suppressors. *Nature Reviews Cancer*, 16(7), pp.413-430.

Efrat, S. (1999). Genetically engineered pancreatic beta-cell lines for cell therapy of diabetes. Ann. *N Y Acad. Sci*. 875, 286-293.

Eizirik, D., Cardozo, A. and Cnop, M. (2008). The Role for Endoplasmic Reticulum Stress in Diabetes Mellitus. *Endocrine Reviews*, 29(1), pp.42-61.

Eizirik, D.L., A.K. Cardozo, and M. Cnop. (2008) The role for endoplasmic reticulum stress in diabetes mellitus. *Endocr Rev*, 29(1): p. 42-61.

El-Assaad W, Buteau J, Peyot ML, Nolan C, Roduit R, Hardy S, Joly E, Dbaibo G, Rosenberg L, Prentki M. (2003) Saturated fatty acids synergize with elevated glucose to cause pancreatic beta-cell death. *Endocrinology*, 144:4154–4163.

Elouil, H., Bensellam, M., Guiot, Y., Vander Mierde, D., Pascal, S., Schuit, F. and Jonas, J. (2007). Acute nutrient regulation of the unfolded protein response and integrated stress response in cultured rat pancreatic islets. *Diabetologia*, 50(7), pp.1442-1452.

Ferrannini, E. (1998). Insulin Resistance versus Insulin Deficiency in Non-Insulin-Dependent Diabetes Mellitus: Problems and Prospects. *Endocrine Reviews*, 19(4), pp.477-490.

Garcia-Ocana, A. (2001). Using Cell Growth Factors to Enhance Human Pancreatic Islet Transplantation. *Journal of Clinical Endocrinology & Metabolism*, 86(3), pp.984-988.

Garneau, D., Revil, T., Fisette, J. and Chabot, B. (2005). Heterogeneous Nuclear Ribonucleoprotein F/H Proteins Modulate the Alternative Splicing of the Apoptotic Mediator Bcl-x*. *Journal of Biological Chemistry*, 280(24), pp.22641-22650.

Ge, Z., Quek, B., Beemon, K. and Hogg, J. (2016). Polypyrimidine tract binding protein 1 protects mRNAs from recognition by the nonsense-mediated mRNA decay pathway. *eLife*, 5.

Gehrmann, W., Würdemann, W., Plötz, T., Jörns, A., Lenzen, S. and Elsner, M. (2015). Antagonism Between Saturated and Unsaturated Fatty Acids in ROS Mediated Lipotoxicity in Rat Insulin-Producing Cells. *Cellular Physiology and Biochemistry*, 36(3), pp.852-865.

Ghigna, C., Giordano, S., Shen, H., Benvenuto, F., Castiglioni, F., Comoglio, P., Green, M., Riva, S. and Biamonti, G. (2005). Cell Motility Is Controlled by

SF2/ASF through Alternative Splicing of the Ron Protooncogene. *Molecular Cell*, 20(6), pp.881-890.

Ghetti A, Bolognesi M, Cobianchi F, Morandi C (1991). "Modeling by homology of RNA binding domain in A1 hnRNP protein". *FEBS Lett*. 277 (1-2): 272–6.

Grant SF, Thorleifsson G, Reynisdottir I, *et al*. (2006). Variant of transcription factor 7-like 2 (*TCF7L2*) gene confers risk of type 2 diabetes. *Nat Genet*. Mar. 38(3):320-3.

Hansson, O., Zhou, Y., Renstrom, E. and Osmark, P. (2010). Molecular Function of TCF7L2: Consequences of TCF7L2 Splicing for Molecular Function and Risk for Type 2 Diabetes. *Current Diabetes Report*, 10(6), pp.444-451.

Harjutsalo V, Podar T, Tuomilehto J, (2005). Cumulative incidence of type 1 diabetes in 10,168 siblings of Finnish young-onset type 1 diabetic patients. *Diabetes*, 54(2):563-569.

Harjutsalo V, Sjöberg L, Tuomilehto J. (2008). Time trends in the incidence of type 1 diabetes in Finnish children: a cohort study. *Lancet*, 371(9626):1777–82.

Harries, L., Hernandez, D., Henley, W., Wood, A., Holly, A., Bradley-Smith, R., Yaghootkar, H., Dutta, A., Murray, A., Frayling, T., Guralnik, J., Bandinelli, S., Singleton, A., Ferrucci, L. and Melzer, D. (2011). Human aging is characterized by focused changes in gene expression and deregulation of alternative splicing. *Aging Cell*, 10(5), pp.868-878.

Heintz, C., Doktor, T., Lanjuin, A., Escoubas, C., Zhang, Y., Weir, H., Dutta, S., Silva-García, C., Bruun, G., Morantte, I., Hoxhaj, G., Manning, B., Andresen, B. and Mair, W. (2016). Splicing factor 1 modulates dietary restriction and TORC1 pathway longevity in C. elegans. *Nature*, 541(7635), pp.102-106.

Heitjan, D., Manni, A. and Santen, R. (1993). Statistical Analysis of *in Vivo* Tumor Experiments. *American Association for Cancer Research*, 53, pp.6042-6050.

Holly, A., Melzer, D., Pilling, L., Fellows, A., Tanaka, T., Ferrucci, L. and Harries, L. (2013). Changes in splicing factor expression are associated with advancing age in man. *Mechanisms of Ageing and Development*, 134(9), pp.356-366.

Horikawa, Y., et al., (2000) Genetic variation in the gene encoding calpain-10 is associated with type 2 diabetes mellitus. *Nat Genet*, 26(2): p. 163-75.

Huggett, J., Dheda, K., Bustin, S. and Zumla, A. (2005). Real-time RT-PCR normalisation; strategies and considerations. *Genes & Immunity*, 6(4), pp.279-284.

Iannaccone, P. and Jacob, H. (2009). Rats! *Disease Models & Mechanisms*, 2(5-6), pp.206-210.

Imaizumi, K., Wanaka, A. and Tohyama, M. (1998). Death-promoting gene DP5, which interacts with BCL2 family, is induced during neuronal apoptosis following exposure to amyloid β-protein. *Neuroscience Research*, 31, p. S362.

Imamura, M. and S. Maeda. (2011). Genetics of type 2 diabetes: the GWAS era and future perspectives [Review]. *Endocr J*, 58(9): p. 723-39.

Izquierdo, J. (2008). Hu Antigen R (HuR) Functions as an Alternative Pre-mRNA Splicing Regulator of Fas Apoptosis-promoting Receptor on Exon Definition. *Journal of Biological Chemistry*, 283(27), pp.19077-19084.

Jacqueminet, S., Briaud, I., Rouault, C., Reach, G. and Poitout, V. (2000). Inhibition of insulin gene expression by long-term exposure of pancreatic β cells to palmitate is dependent on the presence of a stimulatory glucose concentration. *Metabolism*, 49(4), pp.532-536.

Jeong, S. (2017). SR Proteins: Binders, Regulators, and Connectors of RNA. *Molecules and Cells*, 40(1), pp.1-9.

Joseph, J., Koshkin, V., Saleh, M., Sivitz, W., Zhang, C., Lowell, B., Chan, C. and Wheeler, M. (2004). Free Fatty Acid-induced β cell Defects Are Dependent on Uncoupling Protein 2 Expression. *Journal of Biological Chemistry*, 279(49), pp.51049-51056.

Juan-Mateu, J., Rech, T., Villate, O., Lizarraga-Mollinedo, E., Wendt, A., Turatsinze, J., Brondani, L., Nardelli, T., Nogueira, T., Esguerra, J., Alvelos, M., Marchetti, P., Eliasson, L. and Eizirik, D. (2017). Neuron-enriched RNA-binding Proteins Regulate Pancreatic β cell Function and Survival. *Journal of Biological Chemistry*, 292(8), pp.3466-3480.

Jun W *et al*, 2013. Multiple Functions of the RNA Binding Protein HUR in cancer Progression, Treatment Responses and Prognosis. *Int. J. Mol. Sci.* 14,10015-10015.

Kahn, S. (2001). The Importance of cell Failure in the Development and Progression of Type 2 Diabetes. *Journal of Clinical Endocrinology & Metabolism*, 86(9), pp.4047-4058.

Kaminska, D., Kuulasmaa, T., Venesmaa, S., Kakela, P., Vaittinen, M., Pulkkinen, L., Paakkonen, M., Gylling, H., Laakso, M. and Pihlajamaki, J. (2012). Adipose Tissue *TCF7L2* Splicing Is Regulated by Weight Loss and Associates With Glucose and Fatty Acid Metabolism. *Diabetes*, 61(11), pp.2807-2813.

Kelpe CL, Moore PC, Parazzoli SD, Wicksteed B, Rhodes CJ, Poitout V, (2003) Palmitate inhibition of insulin gene expression is mediated at the transcriptional level via ceramide synthesis. *J Biol Chem*, 278:30015–30021.

Kii, I. and Ito, H. (2017). Periostin and its interacting proteins in the construction of extracellular architectures. *Cellular and Molecular Life Sciences*, 74(23), pp.4269-4277.

Kracht, M., van Lummel, M., Nikolic, T., Joosten, A., Laban, S., van der Slik, A., van Veelen, P., Carlotti, F., de Koning, E., Hoeben, R., Zaldumbide, A. and Roep, B. (2017). Autoimmunity against a defective ribosomal insulin gene product in type 1 diabetes. *Nature Medicine*, 23(4), pp.501-507.

Lamolle, G., Marin, M. and Alvarez-Valin, F. (2006). Silent mutations in the gene encoding the p53 protein are preferentially located in conserved amino acid positions and splicing enhancers. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, 600(1-2), pp.102-112.

Larionov, A., Krause, A. and Miller, W. (2005). A standard curve-based method for relative real time PCR data processing. *BioMed Central*, doi:10.1186/1471-2105-6-62.

Latorre, E. and Harries, L. (2017). Splicing regulatory factors, ageing and age-related disease. *Ageing Research Reviews*, 36, pp.165-170.

Le Bacquer, O., Shu, L., Marchand, M., Neve, B., Paroni, F., Kerr Conte, J., Pattou, F., Froguel, P. and Maedler, K. (2011). TCF7L2 splice variants have distinct effects on β cell turnover and function. *Human Molecular Genetics*, 20(10), pp.1906-1915.

Lee, B., Pilling, L., Emond, F., Flurkey, K., Harrison, D., Yuan, R., Peters, L., Kuchel, G., Ferrucci, L., Melzer, D. and Harries, L. (2016). Changes in the expression of splicing factor transcripts and variations in alternative splicing are associated with lifespan in mice and humans. *Aging Cell*, 15(5), pp.903-913.

Lee, E., Kim, W., Tominaga, K., Martindale, J., Yang, X., Subaran, S., Carlson, O., Mercken, E., Kulkarni, R., Akamatsu, W., Okano, H., Perrone-Bizzozero, N., de Cabo, R., Egan, J. and Gorospe, M. (2012). RNA-Binding Protein HuD Controls Insulin Translation. *Molecular Cell*, 45(6), pp.826-835.

Levitt, H., Cyphert, T., Pascoe, J., Hollern, D., Abraham, N., Lundell, R., Rosa, T., Romano, L., Zou, B., O'Donnell, C., Stewart, A., Garcia-Ocaña, A. and Alonso, L. (2010). Glucose stimulates human β cell replication in vivo in islets transplanted into NOD–severe combined immunodeficiency (SCID) mice. *Diabetologia*, 54(3), pp.572-582.

LI, W. (2012). Volcano plots in analysing differential expressions with mRNA microarrays. *Journal of Bioinformatics and Computational Biology*, 10(06), p.1231003.

Lin, L., Park, J., Ramachandran, S., Zhang, Y., Tseng, Y., Shen, S., Waldvogel, H., Curtis, M., Faull, R., Troncoso, J., Pletnikova, O., Ross, C., Davidson, B. and Xing, Y. (2016). Transcriptome sequencing reveals aberrant alternative splicing in Huntington's disease. *Human Molecular Genetics*, 25(16), pp.3454-3466.

Livak, K. and Schmittgen, T. (2001). Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the 2−ΔΔCT Method. *Methods*, 25(4), pp.402-408.

Llorian, M., Gooding, C., Bellora, N., Hallegger, M., Buckroyd, A., Wang, X., Rajgor, D., Kayikci, M., Feltham, J., Ule, J., Eyras, E. and Smith, C. (2016). The alternative splicing program of differentiated smooth muscle cells involves concerted non-productive splicing of post-transcriptional regulators. *Nucleic Acids Research*, 44(18), pp.8933-8950.

Locke, J., Da Silva Xavier, G., Rutter, G. and Harries, L. (2011). Erratum to: An alternative polyadenylation signal in TCF7L2 generates isoforms that inhibit T cell factor/lymphoid-enhancer factor (TCF/LEF)-dependent target genes. *Diabetologia*, 54(12), pp.3170-3170.

Long, J. and Caceres, J. (2009). The SR protein family of splicing factors: master regulators of gene expression. *Biochemical Journal*, 417(1), pp.15-27.

Loos, R., Franks, P., Francis, R., Barroso, I., Gribble, F., Savage, D., Ong, K., O'Rahilly, S. and Wareham, N. (2007). TCF7L2 Polymorphisms Modulate Proinsulin Levels and Beta-Cell Function in a British Europid Population. *Diabetes*, 56(7), pp.1943-1947.

Lu, J. and Schneider, R. (2004). Tissue Distribution of AU-rich mRNA-binding Proteins Involved in Regulation of mRNA Decay. *Journal of Biological Chemistry*, 279(13), pp.12974-12979.

Luzi, L. and DeFronzo, R. (1989). Effect of loss of first-phase insulin secretion on hepatic glucose production and tissue glucose disposal in humans. *American Journal of Physiology-Endocrinology and Metabolism*, 257(2), pp.E241-E246.

Ma, W., Cheng, S., Campbell, C., Wright, A. and Furneaux, H. (1996). Cloning and Characterization of HuR, a Ubiquitously Expressed Elav-like Protein. *Journal of Biological Chemistry*, 271(14), pp.8144-8151.

Matlin, A., Clark, F. and Smith, C. (2005). Understanding alternative splicing: towards a cellular code. *Nature Reviews Molecular Cell Biology*, 6(5), pp.386-398.

Matunis, M., Xing, J. and Dreyfuss, G. (1994). The hnRNP F protein: unique primary structure, nucleic acid-binding properties, and subcellular localization. *Nucleic Acids Research*, 22(6), pp.1059-1067.

Mazoyer, S., Puget, N., Perrin-Vidoz, L., Lynch, H., Serova-Sinilnikova, O. and Lenoir, G. (1998). A BRCA1 Nonsense Mutation Causes Exon Skipping. *The American Journal of Human Genetics*, 62(3), pp.713-715.

McCulloch, L., van de Bunt, M., Braun, M., Frayn, K., Clark, A. and Gloyn, A. (2011). GLUT2 (SLC2A2) is not the principal glucose transporter in human pancreatic beta cells: Implications for understanding genetic association signals at this locus. *Molecular Genetics and Metabolism*, 104(4), pp.648-653.

Mitrakou, A., Kelley, D., Veneman, T., Jenssen, T., Pangburn, T., Reilly, J. and Gerich, J. (1990). Contribution of Abnormal Muscle and Liver Glucose Metabolism to Postprandial Hyperglycemia in NIDDM. *Diabetes*, 39(11), pp.1381-1390.

Muoio, D. and Newgard, C. (2008). Molecular and metabolic mechanisms of insulin resistance and β-cell failure in type 2 diabetes. *Nature Reviews Molecular Cell Biology*, 9(3), pp.193-205.

Murphy-Ullrich, J. (1989). Thrombospondin modulates focal adhesions in endothelial cells. *The journal of Cell Biology*, 109 (3), pp. 1309-1319.

Novak, I., Purinergic receptors in the endocrine and exocrine pancreas. *Purinergic Signal*, 2008. 4(3): p. 237-53.

Oliveira, A., Cunha, D., Ladriere, L., Igoillo-Esteve, M., Bugliani, M., Marchetti, P. and Cnop, M. (2015). In vitro use of free fatty acids bound to albumin: A comparison of protocols. *BioTechniques*, 58(5): p. 228-233.

Osmark, P., Hansson, O., Jonsson, A., Rönn, T., Groop, L. and Renström, E. (2009). Unique splicing pattern of the TCF7L2 gene in human pancreatic islets. *Diabetologia*, 52(5), pp.850-854.

Perseghin, G., Ghosh, S., Gerow, K. and Shulman, G. (1997). Metabolic defects in lean nondiabetic offspring of NIDDM parents: a cross-sectional study. *Diabetes*, 46(6), pp.1001-1009.

Pessin, J. and Saltiel, A. (2000). Signaling pathways in insulin action: molecular targets of insulin resistance. *Journal of Clinical Investigation*, 106(2), pp.165-169.

Pineda, D., Rittenhouse, D., Valley, C., Cozzitorto, J., Burkhart, R., Leiby, B., Winter, J., Weber, M., Londin, E., Rigoutsos, I., Yeo, C., Gorospe, M., Witkiewicz, A., Sachs, J. and Brody, J. (2012). HuR's post-transcriptional regulation of death receptor 5 in pancreatic cancer cells. *Cancer Biology & Therapy*, 13(10), pp.946-955.

Piya, A. and Michels, A. (2010). Understanding the Immunology of Type 1 Diabetes – An Overview of Current Knowledge and Perspectives for the Future. *European Endocrinology*, 8(2), p.70.

Piya, A. and Michels, A. (2012). Understanding the Immunology of Type 1 Diabetes- An Overview of Current Knowledge and Perspectives for the Future. *US Endocrinology*, 08(01), p.12.

Poitout, V., Amyot, J., Semache, M., Zarrouki, B., Hagman, D. and Fontés, G. (2010). Glucolipotoxicity of the pancreatic β cell. *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids*, 1801(3), pp.289-298.

Porat, S., Weinberg-Corem, N., Tornovsky-Babaey, S., Schyr-Ben-Haroush, R., Hija, A., Stolovich-Rain, M., Dadon, D., Granot, Z., Ben-Hur, V., White, P., Girard, C., Karni, R., Kaestner, K., Ashcroft, F., Magnuson, M., Saada, A., Grimsby, J., Glaser, B. and Dor, Y. (2011). Control of Pancreatic β Cell Regeneration by Glucose Metabolism. *Cell Metabolism*, 13(4), pp.440-449.

Pradas-Juni, M., Nicod, N., Fernández-Rebollo, E. and Gomis, R. (2014). Differential Transcriptional and Posttranslational Transcription Factor 7-Like 2 Regulation Among Nondiabetic Individuals and Type 2 Diabetic Patients. *Molecular Endocrinology*, 28(9), pp.1558-1570.

Prentki, M. and Corkey, B. (1996). Are the β-cell signaling molecules malonyl-CoA and cystolic long-chain acyl-CoA implicated in multiple tissue defects of obesity and NIDDM?. *Diabetes*, 45(3), pp.273-283.

Prokunina-Olsson, L., Welch, C., Hansson, O., *et al*. (2009) Tissue-specific alternative splicing of TCF7L2, *Hum Mol Genet*, 18:3795–3804.

Reaven, G., Hollenbeck, C. and Chen, Y. (1989). Relationship between glucose tolerance, insulin secretion, and insulin action in non-obese individuals with varying degrees of glucose tolerance. *Diabetologia*, 32(1): 52-55.

Ruxton, G. and Neuhäuser, M. (2010). When should we use one-tailed hypothesis testing? *Methods in Ecology and Evolution*, 1(2), pp.114-117.

Sale, M., Smith, S., Mychaleckyj, J., Keene, K., Langefeld, C., Leak, T., Hicks, P., Bowden, D., Rich, S. and Freedman, B. (2007). Variants of the Transcription Factor 7-Like 2 (TCF7L2) Gene Are Associated with Type 2 Diabetes in an

African-American Population Enriched for Nephropathy. *Diabetes*, 56(10), pp.2638-2642.

Saltiel, A. and Kahn, C. (2001). Insulin signalling and the regulation of glucose and lipid metabolism. *Nature*, 414(6865), pp.799-806.

Sato, N., Hori, O., Tohyama, M. and Takagi, T. (1998). A novel presenilin-2 (PS-2) splice variant in human Alzheimer's disease brain tissue. *Neuroscience Research*, 31, p.S362.

Scott, L., Bonnycastle, L., Willer, C., Sprau, A., Jackson, A., Narisu, N., Duren, W., Chines, P., Stringham, H., Erdos, M., Valle, T., Tuomilehto, J., Bergman, R., Mohlke, K., Collins, F. and Boehnke, M. (2006). Association of Transcription Factor 7-Like 2 (TCF7L2) Variants With Type 2 Diabetes in a Finnish Sample. *Diabetes*, 55(9), pp.2649-2653.

Shepard, P. and Hertel, K. (2009). The SR protein family. *Genome Biology*, 10(10), p.242.

Sicree, R., Shaw, J. and Zimmet, P. (2003). The global burden of diabetes. Diabetes and impaired glucose tolerance: Prevalence and projections. *Diabetes*, 27, pp.15-71.

Skelin, M., Rupnik, M. and Cencic, A. (2010). Pancreatic beta cell lines and their applications in diabetes mellitus research. *ALTEX*, pp.105-113.

Sladek, R., et al., (2007) A genome-wide association study identifies novel risk loci for type 2 diabetes. *Nature*, 445(7130): p. 881-5

Spellman, R., Llorian, M. and Smith, C. (2007) Cross regulation and Functional Redundancy between the Splicing Regulator PTB and Its Paralogs nPTB and ROD1. *Molecular Cell*, 27(3), pp.420-434.

Srikantan, S., Tominaga, K. and Gorospe, M. (2012). Functional Interplay between RNA-Binding Protein HuR and microRNAs. *Current Protein & Peptide Science*, 13(4), pp.372-379.

Stamateris, R., Sharma, R., Kong, Y., Ebrahimpour, P., Panday, D., Ranganath, P., Zou, B., Levitt, H., Parambil, N., O'Donnell, C., García-Ocaña, A. and Alonso, L. (2016). Glucose Induces Mouse β cell Proliferation via IRS2, MTOR, and Cyclin D2 but Not the Insulin Receptor. *Diabetes*, 65(4), pp.981-995.

Stern, M. (2000). Strategies and prospects for finding insulin resistance genes. *Journal of Clinical Investigation*, 106(3), pp.323-327.

Stoffers, D.A., M.K. Thomas, and J.F. (1997) Habener, Homeodomain protein IDX-1: a master regulator of pancreas development and insulin gene expression. Trends *Endocrinol Metab*, 8(4): p. 145-51.

Stubbs, S. and Wickremesekera, K. (2001). Insulin resistance and type 2 diabetes: time for a new hypothesis. *The New Zealand Medical Journal*, 114(11), pp.239-240.

Suckale, J., Wendling, O., Masjkur, J., Jäger, M., Münster, C., Anastassiadis, K., Stewart, A. and Solimena, M. (2011). PTBP1 Is Required for Embryonic Development before Gastrulation. *PLoS ONE*, 6(2), p. e16992.

Tollervey, J., Wang, Z., Hortobagyi, T., Witten, J., Zarnack, K., Kayikci, M., Clark, T., Schweitzer, A., Rot, G., Curk, T., Zupan, B., Rogelj, B., Shaw, C. and Ule, J. (2011). Analysis of alternative splicing associated with aging and neurodegeneration in the human brain. *Genome Research*, 21(10), pp.1572-1582.

Villate, O., Turatsinze, J., Mascali, L., Grieco, F., Nogueira, T., Cunha, D., Nardelli, T., Sammeth, M., Salunkhe, V., Esguerra, J., Eliasson, L., Marselli, L., Marchetti, P. and Eizirik, D. (2014). Nova1 is a master regulator of alternative splicing in pancreatic β cells. *Nucleic Acids Research*, 42(18), pp.11818-11830.

Viloria, K., Munasinghe, A., Asher, S., Bogyere, R., Jones, L. and Hill, N. (2016). A holistic approach to dissecting SPARC family protein complexity reveals FSTL-1 as an inhibitor of pancreatic cancer cell growth. *Scientific Reports*, 6(1), p. 37839.

Vuong, J., Lin, C., Zhang, M., Chen, L., Black, D. and Zheng, S. (2016). PTBP1 and PTBP2 Serve Both Specific and Redundant Functions in Neuronal Pre-mRNA Splicing. *Cell Reports*, 17(10), pp.2766-2775.

Wagner, E. and Garcia-Blanco, M. (2001). Polypyrimidine Tract Binding Protein Antagonizes Exon Definition. *Molecular and Cellular Biology*, 21(10), pp.3281-3288.

Wahl, M., Will, C. and Lührmann, R. (2009). The Spliceosome: Design Principles of a Dynamic RNP Machine. *Cell*, 136(4), pp.701-718.

Wang, G. and Bushman, F. (2006). A statistical method for comparing viral growth curves. *Journal of Virological Methods*, 135(1), pp.118-123.

Wang, J., Guo, Y., Chu, H., Guan, Y., Bi, J. and Wang, B. (2013). Multiple Functions of the RNA-Binding Protein HuR in Cancer Progression, Treatment Responses and Prognosis. *International Journal of Molecular Sciences*, 14(5), pp.10015-10041.

Wang, Y., Liu, J., Huang, B., Li, J., Huang, L., Lin, J., Zhang, J., Min, Q., Yang, W. and Wang, X. (2014). Mechanism of alternative splicing and its regulation (Review). *Biomedical Reports*, 3, pp.152-158.

Watson, J., Baker, T., Bell, S., Gann, A., Levine, M. and Losick, R. (2008). *Molecular Biology of The Gene*. 6th ed. San Francisco: *Pearson Education Inc*, pp.415-455.

Webster, N. (2017). Alternative RNA Splicing in the Pathogenesis of Liver Disease. *Frontiers in Endocrinology*, 8.

Weir, G. and Bonner-Weir, S. (2013). Islet β cell mass in diabetes and how it relates to function, birth, and death. *Annals of the New York Academy of Sciences*, 1281(1), pp.92-105.

Weise, A., Bruser, K., Elfert, S., Wallmen, B., Wittel, Y., Wöhrle, S. and Hecht, A. (2010). Alternative splicing of Tcf7l2 transcripts generates protein variants with differential promoter-binding and transcriptional activation properties at Wnt/β-catenin targets. *Nucleic Acids Research*, 38(6), pp.1964-1981.

Weyer, C., Tataranni, P., Bogardus, C. and Pratley, R. (2001). Insulin Resistance and Insulin Secretory Dysfunction Are Independent Predictors of Worsening of Glucose Tolerance During Each Stage of Type 2 Diabetes Development. *Diabetes Care*, 24(1), pp.89-94.

Witten, J. and Ule, J. (2011). Understanding splicing regulation through RNA splicing maps. *Trends in Genetics*, 27(3), pp.89-97.

Wong King Yuen, S., Campiglio, M., Tung, C., Flucher, B. and Van Petegem, F. (2018). Structural Insights into the STAC Adaptor Protein and Voltage-Gated Calcium Channel Interaction. *Biophysical Journal*, 114(3), p.40a.

World Health Organization (1999). Report of a WHO Consultation-Definition, Diagnosis and Classification of Diabetes Mellitus. Report of a WHO Consultation, Part 1: Diagnosis and Classification of Diabetes Mellitus. Geneva: WHO; NCDs

.

Yang, S. and Berggren, P. (2006). The Role of Voltage-Gated Calcium Channels in Pancreatic β-Cell Physiology and Pathophysiology. *Endocrine Reviews*, 27(6), pp.621-676.

Youngren, J. and Goldfine, I. (1997). The molecular basis of insulin resistance. *Int J Obesity*, 26, pp.510-516.

Zahler, A (2012). Pre-mRNA splicing and its regulation in *Caenorhabditis elegans* (March 21, 2012), *WormBook*, ed. The C. elegans Research Community, Wormbook, doi/10. 1895/wormbook. 1.31.2, http://www.woormbook.org.

Zahler, A. (2012). Pre-mRNA splicing and its regulation in Caenorhabditis elegans. *WormBook*.

Zhou, L., Cai, X., Han, X. and Ji, L. (2014). P38 Plays an Important Role in Glucolipotoxicity-Induced Apoptosis in INS-1 Cells. *Journal of Diabetes Research*, pp.1-7.

Zhou, Z. and Fu, X. (2013). Regulation of splicing by SR proteins and SR protein-specific kinases. *Chromosoma*, 122(3), pp.191-207.

Zimmet, P. (2000). Globalization, coca-colonization and the chronic disease epidemic: can the Doomsday scenario be averted? *Journal of Internal Medicine*, 247(3), pp.301-310.

Chapter 9-Appendix

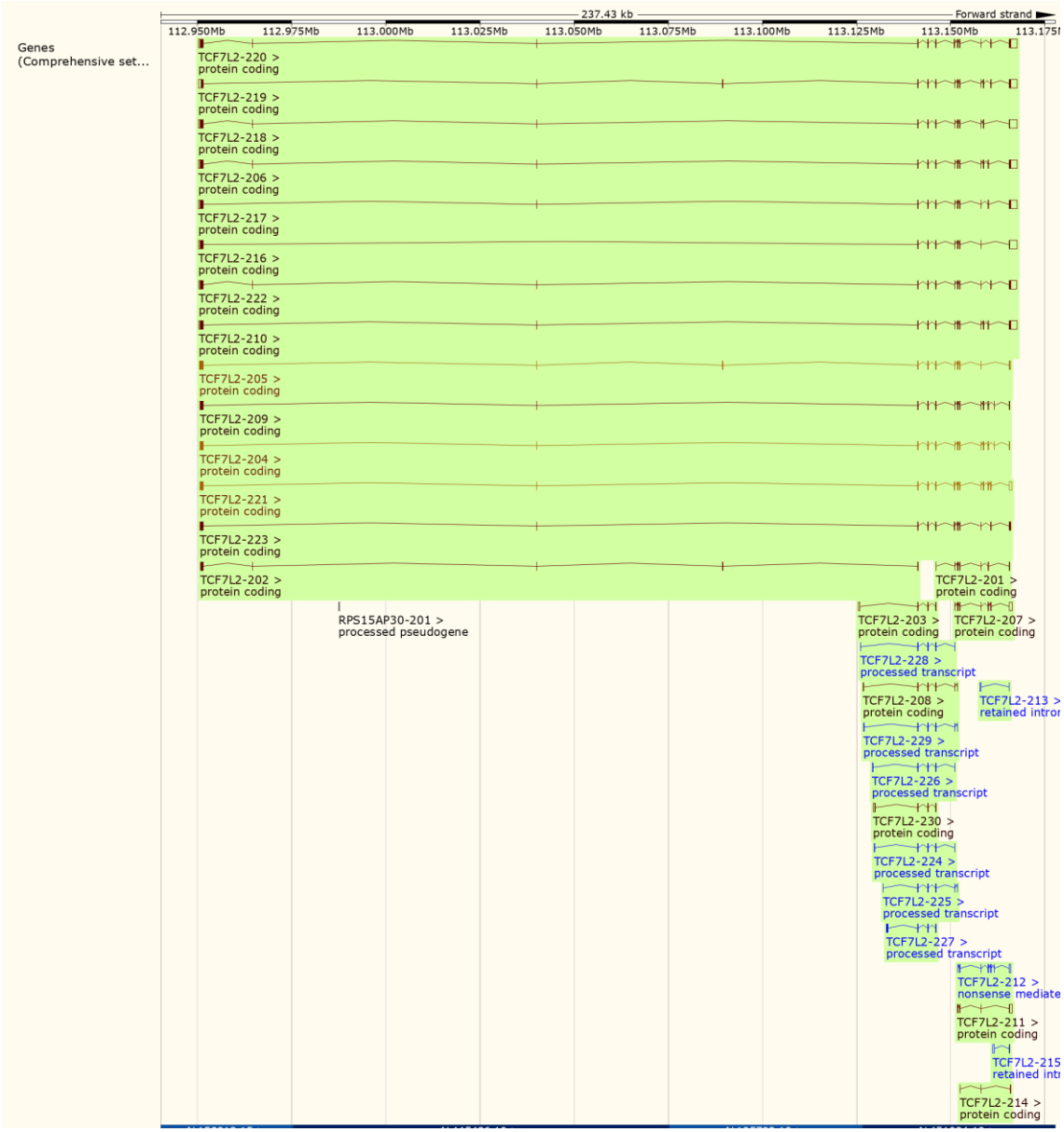
9.1 qPCR fold changes calculations for *Elavl1*, control vs. GLT (*Ppia* as a reference gene)

	FOLD CHANGES CALCULATIONS FOR *Elavl1* (Exp-1)										
			ΔCq	neg ΔΔCq	fold changes to control 1						
Sample Name	*Elavl1* Mean	*Ppia* mean	*Elavl1*	*Elavl1*	*Elavl1*	Mean	sd	t-test		*Elavl1* Mean	*Ppia* mean
CONTROL RT 1	23.2	20.6	2.6	0.0	1.00						
CONTROL RT 1											
CONTROL RT 2	23.4	21.0	2.4	0.3	1.20	1.1	0.09	0.052557	**CON**	23.3	20.9
CONTROL RT 2											
CONTROL RT 3	23.3	21.0	2.4	0.2	1.18						
CONTROL RT 3											
GLT RT 1	22.3	20.0	2.3	0.3	1.26						
GLT RT 1											
GLT RT 2	22.5	20.9	1.6	1.0	2.02	1.8	0.38		**GLT**	22.5	20.7
GLT RT 2											
GLT RT 3	22.7	21.1	1.6	1.1	2.09						
GLT RT 3											

9.2 qPCR analysis of splice factors on GLT conditions (*Ppia* as a reference gene)

Gene name	Paired t-test (*Ppia*, a reference gene)	N number
Rbfox2	0.9	3
Nova2	0.7	3
Elavl1	0.04	4
Elavl4	0.5	3
Hnrnpf	0.3	4
Ptbp1	0.2	4
Srsf3	0.8	4
Srsf5	0.5	4
Srsf6	0.7	4
Srsf10	0.9	4

9.2 Human TCF7L2 splice variants in database



9.3: Exon sequences for human variant 222 (11/9/17)

>Tcf7l2-201 cdna:protein_coding

TCATCCATAAGTATGTTTCTGTGAGGTGCAAGCATGGGCGATCCCATTAAGTTTATCCTA
ATAGATGCTTCAGTAAGGTAAGTTGAGAGAAGGCACAGCTAGGCAGCTGGCTGGTTGCTT
AGAGGGAGGAGAAGCCATATTCATGGGGGGATGGTGTTAGGGATGGGGACGGTGGGAAGG
GGCTTCGTAGGAACGCTGCATGAGGAAGCCGTGTTTGAGATGTCCCTCAGGGACCCCAGC
TGCATGGCAGGGAGGTGTGGAGAATTCTGGGTATGCCTTGGTGACTTGGCCCTCTCTCCT
CAAGCCTCACCCTGTCCCCCTTTCTTCCTGGCAGCAGAGCCCCCTCCCTTGCTGCACTCA
GGGCCATGCCTGTCCGCACTTCTACACCCCCTCAGACTTCACTGTCAGCACGCAAGTCTT
CAGGGACACGAAAAGCAGCCACTCCTTACAAAAGGTTGGGGAGCCCTGGTTTCTGGAGTC
GAACAAAGTGCCGGTGGTGCAGCACCCCCACCATGTCCACCCACTCACACCTCTCATCAC
GTACAGCAATGAACACTTCACCCCGGGAAATCCACCTCCGCACTTACCAGCTGACGTAGA
CCCCAAAACAGGAATTCCACGGCCTCCGCACCCTCCAGATATCTCTCCATATTACCCGCT
GTCGCCCGGCACCGTAGGACAAATCCCCCATCCGCTAGGATGGTTAGTACCACAGCAAGG
TCAGCCTGTGTACCCAATCACGACAGGAGGATTCAGACACCCCTACCCCACAGCGCTGAC
AGTCAACGCGTCTATGTCTAGGTTCCCTCCCCACATGGTCCCTCCCCATCACACTCTGCA
CACGACCGGAATCCCCCACCCGGCCATTGTCACACCGACAGTCAAGCAGGAGTCCTCACA
GAGTGACGTCGGCTCACTCCATAGCTCAAAGCATCAGGACTCCAAGAAGGAAGAAGAGAA
GAAAAAGCCCCACATAAAGAAGCCCCTTAATGCATTCATGTTGTATATGAAGGAGATGAG
AGCGAAGGTGGTGGCTGAGTGCACATTGAAAGAGAGTGCAGCCATCAACCAGATCCTCGG
GCGCAGGTGGCACGCCCTGTCCAGGGAAGAACAGGCAAAGTACTACGAGCTAGCCCGGAA
GGAACGACAGCTTCACATGCAGCTGTACCCTGGCTGGTCTGCACGGGATAACTATGGAAA
GAAGAAGAAGAGAAAAAGAGACAAGCAGCCGGGGGAGACCAATGATGCAAATACTCCAAA
GAAGTGTCGGGCACTGTTCGGGCTTGACCGACAGACTTTATGGTGCAAACCCTGCAGGAG
AAAAAAAAAGTGCGTTCGCTACATACAAGGTGAAGGCAGCTGCCTCAGCCCACCCTCTTC
AGATGGAAGCTTACTAGACTCACCTCCCCCCTCACCGCATCTGCTAGGCTCCCCTCCCCA
AGACGCCAAGTCACAGACTGAGCAGACACAGCCGCTCTCGCTGTCCCTGAAGCCTGATCC
TCTGGCCCACCTGTCCATGATGCCTCCACCACCCGCGCTCCTGCTGGCTGAAGCTGCCCA
CGGCAAGGCCTCTGCCCTCTGTCCCAACGGGGCCCTGGACCTGCCACCTGCCACTTTGCA
GCCGTCCATGGTCCCTTCCTCATCGCTCGCACAACCATCAACTTCTTCCTTACATTCCCA
CAACTCGCTGGCTGGGACGCAACCCCAGCCTCTGTCTCTGGTGACCAAGTCTTTAGAATTA
GCTTCGCCTCCTCGTCTTCCGTGCCCCTGCTTGATTGAAGTGTTTCCTTCTGGTTCTTGG
TTCGCCCTTCCCCCGCTTGGAAAGCCTTTTTTCTTGTTCTTTAATTTTGTGCCACTGTGG
CTACATGAGTTAATGTTTATGGAGTTCATTGGTCAATATTTGACCCATTCTTATTTCAAT
TTCTCCTTTTAAATATGTAGATGAGAGAAACCTCATGATTCTACCAAAATTTTTATCAAC
AGCTGTTTAAAGTCTTTGTAGCGTTTAAAAAAATATATAAATATATATATACATAACTGT
TATGTAGTTCGGATAGCTTAGTTTTAAAAGACTGATTTAAGGAACAAAAAGAAAAAGAAA
AGAGAGAGAAGCAACTTTGAAGCACCCTCCAGAAGGAGTTGGTTCTGTATTATTTGTATT
AAATACGAGCTTGCGAACCAATCATTTCCATCTTGGTATTTAAACCGCGAGGGCACGATG
AATGCAGTGCCGTTTCTTTCTTTCTTTTTTCCCTCCCATGTGAAACAACTTTTATTGTGAT
GTTACTTGTTATTGTTTAAATGTACAGAAACAAAGGGTTAAAATGTGTTAATATACCTTT
TTCCATGGTGTTGTTCTTTTTGGGGGGAGGGGACGCTACTCAACATGAATGGAATCAACA
CTGTTGGACCAGTAGTATTTATTACTTTAGAGGTTGCTTGTCCGTACCTGTCCGTCCGTC
CAATTTTAAATATTGTTTCCCTTTTTTTTTTGAAACTGTATAAAGTCAACCCCCCCAACA
TAAGCATCTTATATACAACAACTCATTTGTACAAGGTTTTTAAGTTTATATATAAATTGT
GTATATATTTTTTTGTTTCCTTTTCTTGACTTTTCGTTTGTTTGTTTGTTTCTGTATAAA
ACCCAGATGCCACCAAATGGACATTGATAGCTGCATTAAGGATCAGTAGCATTAACAAAA
GTTGCTTTAAAAGCCATTACGTAAAACAAGACTTGAACATTAGCGAGAGGGTCGAGCAGA
GCAGCCGTGGGAACCACCCCGTTTCCCCTTGGACCTCCTAGTCAATTACCCTGTGCTCTT
AGAACATGGACTGACCGTGTGCAAAACTTTTTATGTGCCAAAATTCTCAGTGACTTTAGC
TTTCTCCCTCCTTTTTGATGCCGTACTTTCTGTTCGCCATGTTTTGCTGTGATGTTACAT
AGATAGATTTGTATGTAGTTTTAATGTCACCTATAACAAAATGTGTTTGGTAGCAGACTG
TCCAGAAAGCATTTTAAATGAAGAGGTCTAAACCCTTAAGGGCCAAAAATTCTGTATATT
AGATTACTCTTAAAATGAAAAAGAAAAAAAGTAAAAAAAAAAAAAAACCCAGCTGCCGCTT
CTGTGTACAAGTATGACATACGAGTAGGTAGCGAACTTTAAAAACAAGAGAACCTAGGCA
TGCCGATGTTGTAAAATGCTGTATAAAGCCGAGACCTGTCTTTTCAGTGCCATCGTAGTT
GACATGAAGCGATTGTAAAACTGTCTCCGGTTTTCTCTGGTTTATTAAAATGCTAACTAT
AACAATTTTTTTTTGTGAATACTTTGAATGTTTCCTAACAGTTGTGGTGTTACCGTTCTG
TTTGATGCTTATTCCAAGTTCATTTTTGGATGGTTTGGAAGCCATTTTGTAATGAATAAA
TGTCCATGCTGTACAGTATCTGTAGCATGCCGTTCTGGATTAATAAAAGCAACTTAGTAT
GTGCAG

9.4: human and rat (TCF7L2) exons alignment results.

Alignment between human exon5 and rat exon 1: (Date of alignment 130818)

```
CLUSTAL O(1.2.4) multiple sequence alignment


rat1      TCATCCATAAGTATGTTTCTGTGAGGTGCAAGCATGGGCGATCCCATTAAGTTTATCCTA      60
hu5       ------------------------------------------------------------      0


rat1      ATAGATGCTTCAGTAAGGTAAGTTGAGAGAAGGCACAGCTAGGCAGCTGGCTGGTTGCTT      120
hu5       ------------------------------------------------------------      0


rat1      AGAGGGAGGAGAAGCCATATTCATGGGGGGATGGTGTTAGGGATGGGGACGGTGGGAAGG      180
hu5       ------------------------------------------------------------      0


rat1      GGCTTCGTAGGAACGCTGCATGAGGAAGCCGTGTTTGAGATGTCCCTCAGGGACCCCAGC      240
hu5       ------------------------------------------------------------      0


rat1      TGCATGGCAGGGAGGTGTGGAGAATTCTGGGTATGCCTTGGTGACTTGGCCCTCTCTCCT      300
hu5       ------------------------------------------------------------      0


rat1      CAAGCCTCACCCTGTCCCCCTTTCTTCCTGGCAGCAGAGCCCCCTCCCTTGCTGCACTCA      360
hu5       ----------------------------------CAGAGCCCCCTCCCTTGCTGCACTCA      26
                                             **************************


rat1      GGGCCATGCCTGTCCGCACTTCTACACCCCCTCAGACTTCACTGTCAGCACGCAAGTCTT      420
hu5       GGGACATGACTGTCAGCACTTCTACCCCCCCTCAGACTTCACTGTCAGCACTCAAGTCTT      86
          *** **** ***** ********** ************************ ********


rat1      CAGGGACACGAAAAGCAGCCACTCCTTACAAAAGGTTGGGGAGCCCTGGTTTCTGGAG       478
hu5       CAGGGACATGAAAAGGAGCCACTCCTTACAAAAAGTTGGGGAGCCCTGGTGTATTGAG       144
          ******** ****** ***************** *************** * * ***
```

Human exon 6 matched with rat exon 2: (Date of alignment 11.09.17)

```
CLUSTAL O(1.2.4) multiple sequence alignment


hu_exon6        TCTAACAAAGTGCCAGTGGTGCAGCACCCTCACCATGTCCACCCCCTCACGCCTCTTATC 60
rat_exonn2      TCGAACAAAGTGCCGGTGGTGCAGCACCCCCACCATGTCCACCCACTCACACCTCTCATC 60
                ** *********** ************** ************** ***** ***** ***

hu_exon6        ACGTACAGCAATGAACACTTCACGCCGGGAAACCCACCTCCACACTTACCAGCCGACGTA 120
rat_exonn2      ACGTACAGCAATGAACACTTCACCCCGGGAAATCCACCTCCGCACTTACCAGCTGACGTA 120
                ************************ ******** ******** *********** ******

hu_exon6        GACCCCAAAACAG         133
rat_exonn2      GACCCCAAAACAG         133
                *************
```

Human exon 7 matched with rat exon 3: (Date of alignment 11.09.17)

```
CLUSTAL O(1.2.4) multiple sequence alignment


hu_exon7        GAATCCCACGGCCTCCGCACCCTCCAGATATATCCCCGTATTACCCACTATCGCCTGGCA  60
rat_exon3       GAATTCCACGGCCTCCGCACCCTCCAGATATCTCTCCATATTACCCGCTGTCGCCCGGCA  60
                **** ************************** ** ** ******** ** ***** ****

hu_exon7        CCGTAGGACAAATCCCCCATCCGCTAGGATGGTTAGTACCACA         103
rat_exon3       CCGTAGGACAAATCCCCCATCCGCTAGGATGGTTAGTACCACA         103
                *****************************************
```

Human exon 8 matched with rat exon 4: (Date of alignment 11.09.17)

```
CLUSTAL O(1.2.4) multiple sequence alignment


hu_exon8        GCAAGGTCAACCAGTGTACCCAATCACGACAGGAGGATTCAGACACCCCTACCCCACAGC  60
rat_exon4       GCAAGGTCAGCCTGTGTACCCAATCACGACAGGAGGATTCAGACACCCCTACCCCACAGC  60
                ********* ** ************************************************

hu_exon8        TCTGACCGTCAATGCTTCCATGTCCAG      87
rat_exon4       GCTGACAGTCAACGCGTCTATGTCTAG      87
                 ***** ***** ** ** ***** **
```

Human exon 9 matched with rat exon 5: (Date of alignment 11.09.17)

```
CLUSTAL O(1.2.4) multiple sequence alignment


hu_exon9        GTTCCCTCCCCATATGGTCCCACCACATCATACGCTACACACGACGGGCATTCCGCATCC  60
rat_exon5       GTTCCCTCCCCACATGGTCCCTCCCCATCACACTCTGCACACGACCGGAATCCCCCACCC  60
                ************ ******** ** ***** ** ** ******** ** ** ** ** **

hu_exon9        GGCCATAGTCACACCAACAGTCAAACAGGAATCGTCCCAGAGTGATGTCGGCTCACTCCA  120
rat_exon5       GGCCATTGTCACACCGACAGTCAAGCAGGAGTCCTCACAGAGTGACGTCGGCTCACTCCA  120
                ****** ******** ******** ***** ** ** ******** **************

hu_exon9        TAGTTC        126
rat_exon5       TAGCTC        126
                *** **
```

Human exon 10 matched with rat exon 6: (Date of alignment 11.09.17)

```
LUSTAL O(1.2.4) multiple sequence alignment


hu_exon10       AAAGCATCAGGACTCCAAAAAGGAAGAAGAAAAGAAGAAGCCCCACATAAAGAAACCTCT  60
rat_exon6       AAAGCATCAGGACTCCAAGAAGGAAGAAGAGAAGAAAAAGCCCCACATAAAGAAGCCCCT  60
                ****************** *********** ***** ***************** ** **

hu_exon10       TAATGCATTCATGTTGTATATGAAGGAAATGAGAGCAAAGGTCGTAGCTGAGTGCACGTT  120
rat_exon6       TAATGCATTCATGTTGTATATGAAGGAGATGAGAGCGAAGGTGGTGGCTGAGTGCACATT  120
                ************************** ******** ***** ** ************ **

hu_exon10       GAAAGAAAGCGCGGCCATCAACCAGATCCTTGGGCGGAGG  160
rat_exon6       GAAAGAGAGTGCAGCCATCAACCAGATCCTCGGGCGCAGG  160
                ****** ** ** ***************** ***** ***
```

Human exon 11 matched with rat exon 7: (Date of alignment 11.09.17)

```
CLUSTAL O(1.2.4) multiple sequence alignment


hu_exon11       TGGCATGCACTGTCCAGAGAAGAGCAAGCGAAATACTACGAGCTGGCCCGGAAGGAGCGA  60
rat_exon7       TGGCACGCCCTGTCCAGGGAAGAACAGGCAAAGTACTACGAGCTAGCCCGGAAGGAACGA  60
                ***** ** ******** ***** ** ** ** *********** *********** ***

hu_exon11       CAGCTTCATATGCAACTGTACCCCGGCTGGTCCGCGCGGGATAACTAT    108
rat_exon7       CAGCTTCACATGCAGCTGTACCCTGGCTGGTCTGCACGGGATAACTAT    108
                ******** ***** ******** ******** ** ************
```

Human exon 12 matched with rat exon 8: (Date of alignment 11.09.17)

```
CLUSTAL O(1.2.4) multiple sequence alignment


Hu_exon12       GGAAAGAAGAAGAAGAGGAAAAGGGACAAGCAGCCGGGAGAGACCAATG   49
rat_exon8       GGAAAGAAGAAGAAGAGAAAAAGAGACAAGCAGCCGGGGGAGACCAATG   49
                ***************** ***** ************** **********
```

Alignment between human exon 13 and new novel rat exon '8b': (Date of alignment 01.07.18)

```
CLUSTAL O(1.2.4) multiple sequence alignment


Rat8b     AACACAGCGAATGTTTCCTAAATCCTTGCCTTTCGCTTCCTCCGATCACAG       51
HU13      AACACAGCGAATGTTTCCTAAATCCTTGCCTTTCACTTCCTCCGATTACAG       51
          ********************************** *********** ****
```

Alignment between human exon 14 and new novel rat exon '8c': (Date of alignment 01.07.18)

```
CLUSTAL O(1.2.4) multiple sequence alignment


Rat8c     ACCTGAGCGCTCCTAAGAAATGCCGAGCGCGCTTTGGCCTTGATCAACAGAATAACTGGT       60
HU14      ACCTGAGCGCTCCTAAGAAATGCCGAGCGCGCTTTGGCCTTGATCAACAGAATAACTGGT       60
          ************************************************************

C68c      GCGGCCCCTGCAG   73
HU14      GCGGCCCTTGCAG   73
          ******* *****
```

Human exon15 matched with rat exon 9: (Date of alignment 11.09.17)

```
CLUSTAL O(1.2.4) multiple sequence alignment


hu_exon13       ATGCAAATACTCCAAAGAAGTGTCGGGCACTGTTCGGGCTTGACCGACAGACTTTATGGT  60
rat_exon9       ATGCAAATACTCCAAAGAAGTGTCGGGCACTGTTCGGGCTTGACCGACAGACTTTATGGT  60
                ************************************************************

hu_exon13       GCAAACCGTGCAG      73
rat_exon9       GCAAACCCTGCAG      73
                ******* *****
```

Alignment between human exon 13 and new novel rat exon '9b': (Date of alignment 01.07.18)

```
CLUSTAL O(1.2.4) multiple sequence alignment


Rat9b       TCTTTGAATTTGGAATATTAAGATG       25
HU16        TCTTTGAATTTGGAATATTACAATG       25
            ********************  ***
```

Human exon17 matched with rat exon 10: (Date of alignment 11.09.17)

```
CLUSTAL O(1.2.4) multiple sequence alignment


hu_exon14       GAGAAAAAAAAAGTGCGTTCGCTACATACAAGGTGAAGGCAGCTGCCTCAGCCCACCCTC 60
rat_exon10      GAGAAAAAAAAAGTGCGTTCGCTACATACAAGGTGAAGGCAGCTGCCTCAGCCCACCCTC 60
                ************************************************************

hu_exon14       TTCAGATGGAAGCTTACTAGATTCGCCTCCCCCCTCCCCGAACCTGCTAGGCTCCCCTCC 120
rat_exon10      TTCAGATGGAAGCTTACTAGACTCACCTCCCCCCTCACCGCATCTGCTAGGCTCCCCTCC 120
                ********************* ** *********** *** * ****************

hu_exon14       CCGAGACGCCAAGTCACAGACTGAGCAGACCCAGCCTCTGTCGCTGTCCCTGAAGCCCGA 180
rat_exon10      CCAAGACGCCAAGTCACAGACTGAGCAGACACAGCCGCTCTCGCTGTCCCTGAAGCCTGA 180
                ** ************************** ***** ** ***************** **

hu exon14       CCCCCTGGCCCACCTGTCCATGATGCCTCCGCCACCCGCCCTCCTGCTCGCTGAGGCCAC 240
rat_exon10      TCCTCTGGCCCACCTGTCCATGATGCCTCCACCACCCGCGCTCCTGCTGGCTGAAGCTGC 240
                 ** ************************** ******** ******** ***** **  *

hu_exon14       ---CCACAAGGCCTCCGCCCTCTGTCCCAACGGGGCCCTGGACCTGCCCCCAGCCGCTTT 297
rat_exon10      CCACGGCAAGGCCTCTGCCCTCTGTCCCAACGGGGCCCTGGACCTGCCACCTGCCACTTT 300
                   *  ********* ******************************** ** *** ****

hu_exon14       GCAGCCTGCC---GCCCCCTCCTCATCAATTGCACAGCCGTCGACTTCTTCCTTACATTC 354
rat_exon10      GCAGCCGTCCATGGTCCCTTCCTCATCGCTCGCACAACCATCAACTTCTTCCTTACATTC 360
                ******  **   * *** ********  * ***** ** ** ****************

hu_exon14       CCACAGCTCCCTGGCCGGGACCCAGCCCCAGCCGCTGTCGCTCGTCACCAAGTCTTTAGA 414
rat_exon10      CCACAACTCGCTGGCTGGGACGCAACCCCAGCCTCTGTCTCTGGTGACCAAGTCTTTAGA 420
                ***** *** ***** ***** ** ******** ***** ** ** **************

hu_exon14       ATAGCTTTAGCGTCGTGAACCCCGCTGCTTTGTTTAT-GGTTTTGTTTCACTTTTCTTAA 473
rat_exon10      ATAGCTTCGCCTCCTCGTCTTCCGTGCCCCTGCTTGATTGAAGTGTTTCCTTCTGGTTCT 480
                *******   *  *  *    ***   *  ** **    *   ******  * *  **

hu_exon14       TTTGCCCCCCACCCCCACCTTGAAAGGTTTTGTTTTGTACTCTCTTAATTTTGTGCCATG 533
rat_exon10      TGGTTCGCCCTTCCCCCGCTTGGAAAGCCTTTTTTCTTGTTCTTTAATTTTGTGCCACTG 540
                *    * ***  ****  **** ** *  ** ***  *  *** * * ***    *  **

hu_exon14       TGGCTACATTAGTTGATGTTTATCGAGTTCATTGGTCAATATTTGACCCATTCTTATTTC 593
rat_exon10      TGGCTACATGAGTTAATGTTTATGGAGTTCATTGGTCAATATTTGACCCATTCTTATTTC 600
                ********* **** ******** ************************************

hu_exon14       AATTTCTCCTTTTAAATATGTAGATGAGAGAAGAACCTCATGATTCTACCAAAATTTTTA 653
rat_exon10      AATTTCTCCTTTTAAATATGTAGATGAGAG--AAACCTCATGATTCTACCAAAATTTTTA 658
                ****************************   *************************

hu_exon14       TCAACAGCTGTTTAAAGTCTTTGTAGCGTTTAAAAAA-------TATATATATATACATA 706
rat_exon10      TCAACAGCTGTTTAAAGTCTTTGTAGCGTTTAAAAAAATATATAAATATATATATACATA 718
                ************************************       ***************

hu_exon14       ACTGTTATGTAGTTCGGATAGCTTAGTTTTAAAAGACTGATTAAAAAACAAAAAGAAAAA 766
rat_exon10      ACTGTTATGTAGTTCGGATAGCTTAGTTTTAAAAGACTGATTTAAGGAACAAAAAGAAAA 778
```

```
                ****************************************** **  *  ****  ****

hu_exon14       AA----------AAGCAATTTTGAAGCAGCCCTCCAGAAGGAGTTGGTTCTGTATTATTT 816
rat_exon10      AGAAAAGAGAGAGAAGCAACTTTGAAGCACCCTCCAGAAGGAGTTGGTTCTGTATTATTT 838
                *            *   *  **  *    ********************************

hu_exon14       GTATTAAATACGAGCTTGCGAACCAATCATTTTACATCTGGTTTTTAAACCGTAAGGGCA 876
rat_exon10      GTATTAAATACGAGCTTGCGAACCAATCATTTCCATCTTGGTATTTAAACCGCGAGGGCA 898
                ********************************      **** *********  ******

hu_exon14       CCATGAATGCAGTGCCGTTACTTTTTTTTTTTT---TTTCTGTGTGAAACAACTCTTATT 933
rat_exon10      CGATGAATGCAGTGCCGTTTCTTTCTTTCTTTTTTCCTCCCATGTGAAACAACTTTTATT 958
                * ***************** **** *** ****    * *  ************ *****

hu exon14       GTGATGTTACTTGTTATTGTTTAAATGTACAGAAACAAAGGGTAAAAATGTGTTAATATA 993
rat_exon10      GTGATGTTACTTGTTATTGTTTAAATGTACAGAAACAAAGGGTTAAAATGTGTTAATATA 1018
                ****************************************** ****************

hu_exon14       CCTTGTTCCATGGTGTTGTTCTTTTGGGGGGAGGGGACGCTACTCAACACTTAATAGAAT 1053
rat_exon10      CCTTTTTCCATGGTGTTGTTCTTTTTGGGGGGAGGGGACGCTACTCAACATGAAT--GGA 1076
                **** ******************** *****  ***          *   * ***

hu_exon14       CACAACGCTGTTGGGCCAGTAGTATTTATTGCTTTAGAGATTGCTTGTCGTACCTGTAT- 1112
rat_exon10      ATCAACACTGTTGGACCAGTAGTATTTATTACTTTAGAGGTTGCTTGTCCGTACCTGTCC 1136
                  **** ******* *************** ******** *********    *

hu_exon14       G-TCGTCCCTTTTTAAATATGTTTTCCTTTTTCTTGAAACTGTATAAAGTTTTTTTCCCC 1171
rat_exon10      GTCCGTCCAATTTTAAATATTGTTTCCCTTTTTTTTTTGAAACTGTATAAAGTCAACCCC 1196
                *  *****  **********  ***** **** **          *     *   ****

hu_exon14       CTTAGCATAAGCATCTTATATATAACAACTCATTTGTACAAGGTTTTTAAGTTTATATAT 1231
rat_exon10      CCCAACATAAGCATCTTATATACAACAACTCATTTGTACAAGGTTTTTAAGTTTATATAT 1256
                *  * ***************** ************************************

hu_exon14       AAAATGTGTATATATATTTTTGTTTCCCCTTTTTGACTT-----------TTTTTTTTCT 1280
rat_exon10      AAATTGTGTATATATTTTTTTGTTTCCTTTTCTTGACTTTTCGTTTGTTTGTTTGTTTCT 1316
                *** *********** ***********  ** *******            *** *****

hu_exon14       GTATGAAACCCAGATGTCACCAAATGGACATTAATAGTTGCATTAAGGATCAGTAGCATT 1340
rat_exon10      GTATAAAACCCAGATGCCACCAAATGGACATTGATAGCTGCATTAAGGATCAGTAGCATT 1376
                **** *********** *************** **** **********************

hu_exon14       AACAAAAGTTGCTTTAAAAGCCATTATGTAAAACAAGACTTGAAAATGAGTGAGGGAATT 1400
rat_exon10      AACAAAAGTTGCTTTAAAAGCCATTACGTAAAACAAGACTTGAACATTAGCGAGAGGGTC 1436
                ************************** ***************** ** ** *** *  *

hu_exon14       TTAGCGACACTGTCTGAGCAGCAGTGGGAACCATCTTCGTTTCCCCTTTGAACTCCCAGT 1460
rat_exon10      GAGCAGAGCAGCCGTGGGAA----------CCA-CCCCGTTTCCCCTTGGACCTCCTAGT 1485
                     **       ** * *          *** *  *********** ** **** ***

hu_exon14       GGGATGCCCTACCCTGCGCCCTTAGGACCCGGACTGACCGTGTACAAA--ACTTTACGTG 1518
rat_exon10      C-----AATTACCCTGTGCTCTTAGAACATGGACTGACCGTGTGCAAAACTTTTTATGTG 1540
                      ******* ** ***** **  ************* ****    **** ***

hu_exon14       CCAAAATTCTCAGTGAATTTAGCTTTCTCCCT-CTTTTTGATGCTGTAATTTTTGTTCAT 1577
rat_exon10      CCAAAATTCTCAGTGACTTTAGCTTTCTCCCTCCTTTTTGATGCCGTACTTTCTGTTCGC 1600
                ***************** *************** *********** *** *** *****

hu_exon14       CATGTTTTGCTGTGATGTTACATAGGTAGATTTGTATGTAGTTTTAATGTCACCTATAAC 1637
rat_exon10      CATGTTTTGCTGTGATGTTACATAGATAGATTTGTATGTAGTTTTAATGTCACCTATAAC 1660
                ************************** **********************************

hu_exon14       AAAATGTGTTTGGTAGCAGATTGTCCAGAAAGCATTTTAAATGAAGAGGTATAAACCCTT 1697
rat_exon10      AAAATGTGTTTGGTAGCAGACTGTCCAGAAAGCATTTTAAATGAAGAGGTCTAAACCCTT 1720
                ******************** ***************************** *********

hu_exon14       AAGGGCCAAAAT-TCTGTATATTAGATTACTCTTAAAC----------------------- 1734
rat_exon10      AAGGGCCAAAAATTCTGTATATTAGATTACTCTTAAAATGAAAAAGAAAAAAAGTAAAAA 1780
                ***********  ************************

hu exon14       ----GAAAAACCAGCTGCCGCTTTTATGTACACATATTACATACGAGTAGGCAGCAGACT 1790
rat_exon10      AAAAAAAAACCCAGCTGCCGCTTCTGTGTACAAGTATGACATACGAGTAGGTAGCGAACT 1840
                     **** ************* * ******  *** ************* ***  ***

hu_exon14       TTAAAAATAAAAAAAACCTAGGCATGTTGATGTTGCAAAATGCTGTATAAAGCTGAAACC 1850
rat_exon10      TTAAA-AACAAGAGAACCTAGGCATGCCGATGTTGTAAAATGCTGTATAAAGCCGAGACC 1899
                ***** *  ** * ************  ******* ***************** ** ***
```

```
hu_exon14       TGTTCATTCAGTGCCATTGTAGTTGACATGAAGCGATTGTAAAACTGTCTCCGATTTTTC 1910
rat_exon10      TGTCTTTTCAGTGCCATCGTAGTTGACATGAAGCGATTGTAAAACTGTCTCCGGTTTTC- 1958
                ***   *********** ********************************** ****

hu_exon14       TCTGGTTTATTAAAATGCTAACTATAACA------------------------------ 1939
rat_exon10      TCTGGTTTATTAAAATGCTAACTATAACAATTTTTTTTTGTGAATACTTTGAATGTTTCC 2018
                ****************************

hu_exon14       ------------------------------------------------------------ 1939
rat_exon10      TAACAGTTGTGGTGTTACCGTTCTGTTTGATGCTTATTCCAAGTTCATTTTTGGATGGTT 2078


hu_exon14       ------------------------------------------------------------ 1939
rat exon10      TGGAAGCCATTTTGTAATGAATAAATGTCCATGCTGTACAGTATCTGTAGCATGCCGTTC 2138


hu_exon14       -------------------------------- 1939
rat_exon10      TGGATTAATAAAAGCAACTTAGTATGTGCAG 2169
```

9.5 Alignment results between *Tcf7l2* protein coding (Expected) and splice variants A (composite) in pancreatic β cells.

(Exon7_Exon8_Exon9_Exon10)

CLUSTAL O (1.2.4) multiple sequence alignment

```
Composite       ------------------------------------------------------------  0
expected        GCGCAGGTGGCACGCCCTGTCCAGGGAAGAACAGGCAAAGTACTACGAGCTAGCCCGGAA  60


Composite       ---------------------------------CTGGTTTGCACGGGATAACTATGGAAA  27
expected        GGAACGACAGCTTCACATGCAGCTGTACCCTGGCTGGTCTGCACGGGATAACTATGGAAA  120
                                                 ***** ********************

Composite       GAAGAAGAAGAGAAAAAGAGACAAGCAGCCGGGGGAGACCAATGATGCAAATACTCCAAA  87
expected        GAAGAAGAAGAGAAAAAGAGACAAGCAGCCGGGGGAGACCAATGATGCAAATACTCCAAA  180
                ************************************************************

Composite       GAAGTGTCGGGCACTGTTCGGGCTTGACCGACAGACTTTATGGTGCAAACCCTGCAGGAG  147
expected        GAAGTGTCGGGCACTGTTCGGGCTTGACCGACAGACTTTATGGTGCAAACCCTGCAGGAG  240
                ************************************************************


Composite       AAAAAAAAAGTGCGTTCGCTACATACAAGGTGAAGGCCCCTGCAAC-------------- 193
expected        AAAAAAAAAGTGCGTTCGCTACATACAAGGTGAAGGCAGCTGCCTCAGCCCACCCTCTTC 300
                ************************************  ****  *
```

9.6 Alignment results between *Tcf7l2* protein coding (Expected) and splice variants B (composite) in pancreatic β cells.

(Exon7_Exon8_Exon10)

CLUSTAL O (1.2.4) multiple sequence alignment

```
Composite:      ------------------------------------------------------------ 0
Expected        TGGCACGCCCTGTCCAGGGAAGAACAGGCAAAGTACTACGAGCTAGCCCGGAAGGAACGA 60

Composite:      -------------------------TTCTGGTCTGCACGGGATAACTATGGAAAGAAGAAG 36
Expected        CAGCTTCACATGCAGCTGTACCCTGGCTGGTCTGCACGGGATAACTATGGAAAGAAGAAG 120
                                         **********************************

Composite:      AAGAGAAAAAGAGACAAGCAGCCGGGGGGAGACCAATG------------------------ 88
Expected        AAGAGAAAAAGAGACAAGCAGCCGGGGGAGACCAATGATGCAAATACTCCAAAGAAGTGT 180
                ************************************

Composite:      ----------------------------------------------------GAGAAAAAAA 88
Expected        CGGGCACTGTTCGGGCTTGACCGACAGACTTTATGGTGCAAACCCTGCAGGAGAAAAAAA 240
                                                                   **********

Composite:      AAGTGCGTTCGCTACATACAAGGTGAAGGCAGCTGC------------------------119
Expected        AAGTGCGTTCGCTACATACAAGGTGAAGGCAGCTGCCTCAGCCCACCCTCTTCAGATGGA 300
                ************************************

Composite:      ------------------------------------------------------------ 119
Expected        AGCTTACTAGACTCACCTCCCCCCTCACCGCATCTGCTAGGCTCCCCTCCCCAAGACGCC 360
```

9.7 Alignment results between *Tcf7l2* protein coding (Expected) and splice variants C (composite) in pancreatic β cells.

(Exon7_Exon8_Exon9_Exon9b_Exon10)

CLUSTAL O (1.2.4) multiple sequence alignment

```
Composite       ------------------------------------------------------------  0
expected        GCGCAGGTGGCACGCCCTGTCCAGGGAAGAACAGGCAAAGTACTACGAGCTAGCCCGGAA  60

Composite       --------------------------------TCTGGTCTGCACGGGATAACTATGGAAA  28
expected        GGAACGACAGCTTCACATGCAGCTGTACCCTGGCTGGTCTGCACGGGATAACTATGGAAA  120
                                               ****************************

Composite       GAAGAAGAAGAGAAAAAGAGACAAGCAGCCGGGGGAGACCAATGATGCAAATACTCCAAA  88
expected        GAAGAAGAAGAGAAAAAGAGACAAGCAGCCGGGGGAGACCAATGATGCAAATACTCCAAA  180
                ************************************************************

Composite       GAAGTGTCGGGCACTGTTCGGGCTTGACCGACAGACTTTATGGTGCAAACCCTGCAGTCT  148
expected        GAAGTGTCGGGCACTGTTCGGGCTTGACCGACAGACTTTATGGTGCAAACCCTGCAG---  240
                ************************************************************

Composite       TTGAATTTGGAATATTATGATGGAGAAAAAAAAAGGGCGTTCGCTACATACAAGGTGAAG  208
expected        ----------------------GGAGAAAAAAAAAGTGCGTTCGCTACATACAAGGTGAAG  275
                                       ************************************

Composite       GCAGCTGCAACA------------------------------------------------  220
expected        GCAGCTGCCTCAGCCCACCCTCTTCAGATGGAAGCTTACTAGACTCACCTCCCCCCTCAC  335
                ********  **

Composite       ------------------------------------------------------------  220
expected        CGCATCTGCTAGGCTCCCCTCCCCAAGACGCCAAGTCACAGACTGAGCAGACACAGCCGC  395
```

9.8 Alignment results between *Tcf7l2* protein coding (Expected) and splice variants D (composite) in pancreatic β cells.

(Exon7_Exon8_Exon8b_Exon8c_Exon9_Exon9b_Exon10)

CLUSTAL O (1.2.4) multiple sequence alignment

```
Composite       ------------------------------------------------------------ 0
expected        GCGCAGGTGGCACGCCCTGTCCAGGGAAGAACAGGCAAAGTACTACGAGCTAGCCCGGAA 60

Composite       ------------------------------TTCTGGTCTGCACGGGATAACTATGGAAA 29
expected        GGAACGACAGCTTCACATGCAGCTGTACCCTGGCTGGTCTGCACGGGATAACTATGGAAA 120
                                              **************************

Composite       GAAGAAGAAGAGAAAAAGAGACAAGCAGCCGGGGGAGACCAATGAACACAGCGAATGTTT 89
expected        GAAGAAGAAGAGAAAAAGAGACAAGCAGCCGGGGGAGACCAATG---------------- 162
                ****************************************

Composite       CCTAAATCCTTGCCTTTCGCTTCCTCCGATCACAGACCTGAGCGCTCCTAAGAAATGCCG 149
expected        ------------------------------------------------------------ 162

Composite       AGCGCGCTTTGGCCTTGATCAACAGAATAACTGGTGCGGCCCCTGCAGATGCAAATACTC 209
expected        -------------------------------------------------ATGCAAATACTC 176
                                                                ************

Composite       CAAAGAACTGTCGGGCACTGTTCGGGCTTGACCGACAGACTTTATGGTGCAAACCCTGCA 269
expected        CAAAGAAGTGTCGGGCACTGTTCGGGCTTGACCGACAGACTTTATGGTGCAAACCCTGCA 236
                ******* ************************************************

Composite       GCCTTTGAATTTGGAATATTATGATGGAGAAAAAAAAAGTG------------------- 310
expected        -------------------------GGAGAAAAAAAAAGTGCGTTCGCTACATACAAGGT 296
                                         ***************

Composite       ------------------------------------------------------------ 310
expected        CTTCAGATGGAAGCTTACTAGACTCACCTCCCCCCTCACCGCATCTGCTAGGCTCCCCTC 356
```

9.9 A Summary of rat *Tcf7l2* Splice Variants Analysis in INS1 cell.

Sample code	Rat *Tcf7l2* primers	Estimated band size based on the gel (bp)	Sequence similarity with *Tcf7l2* (Blastn Result)	Exon Position
A1	1F and 2R	110 (Sequence length-114 bp)	similar	Exon1-exon2
A2	1F and 2R	350	Sequencing quality was good. Blastn results shown Rattus bacterial Prosc.; need to purify bands further to identify.	Not applicable
A3	1F and 2R	450	Sequence quality was poor. Blastn results shown no sequence similarities with *Tcf7l2.*	Not applicable
A4	1F and 2R	620	Sequence had similarities with Streptomyces strain.	Not applicable
A5	7F and 10R	110 (Sequence length 119 bp)	Sequence quality was good and exact matched with *Tcf7l2.*	Exon7_Exon8_Exon 10
A6	7F and 10R	210 (Sequence length 220 bp)	Sequence quality was good and exact matched with *Tcf7l2.*	Exon7_Exon 8_Exon 9_Exon9b_Exon 10

A7	7F and 10R	280 (Sequence length 221 bp)	Sequence quality was good and exact matched with *Tcf7l2.*	Exon7_Exon 8_Exon 9_Exon9b_Exon10
A8	7F and 10R	340 (Sequence length 342 bp)	Sequence quality was good and exact matched with *Tcf7l2.*	Exon7 (starts with primer seq)_Exon8_Exon 8b_Exon 8c_Exon9_Exon 9b_Exon10

9.10 Analysis of *Tcf7l2* splice variants on Con vs GLT.

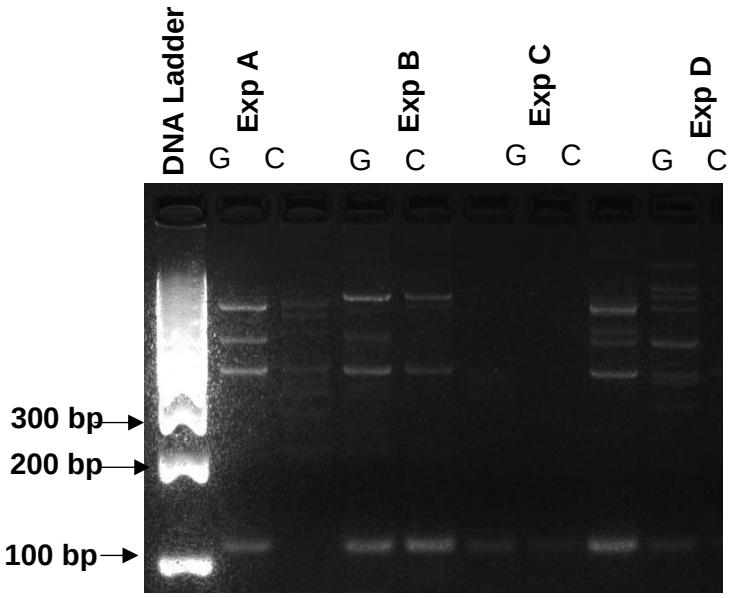
Sample (Tcf7l2); con-C, GLT- G	**Estimated band size based on the gel (bp)**	**Sequence similarity with Tcf7 (Blastn Result)**	**Exon Position**
C1	110 bp (Sequence length 119 bp)	similar	Exon 7 (though only 2bp) – exon 8 – exon 10
C2	180/190 (Sequence length 120 bp)	similar	Exon 7 (though only 2bp) – exon 8 – exon 10
C3	190/200 (Sequence length 220 bp)	similar	Exon 7 (though only 2bp) – exon 8 - exon9-exon9b-exon 10
C4	205/210	Sequencing quality too poor to tell but has region of exon8; need to purify bands further to identify.	Unknown.
C5	240/250	Sequencing quality too poor to tell but has region	Unknown.

		of exon8b; need to purify bands further to identify.	
C6	290 (Sequence length 310 bp)/ Exon 10 did not go through the reverse primer.	similar	Exon 7 (though only 2bp) – exon 8 –exon8b-exon8c-exon9-exon9bexon 10 (part of e-10; reverse primer seq missing
C7	350	Sequencing quality too poor to tell but has region of exon8; need to purify bands further to identify.	Unknown.
C8	490	Sequence quality too low to tell anything.	Unknown.
G1	100/105 (Sequence length 119 bp)	Similar to C1	Exon 7 (though only 2bp) – exon 8 – exon 10 ie is lacking exon 9
G2	190-Predicted Band Size (Sequence length 193 bp)	Similar	Exon 7 (though only 2bp) – exon 8-exon9-exon 10
G3	200/205 (Sequence length 219 bp)	Identical to C3	Exon 7 (though only 2bp) – exon 8- exon9-exon9b-exon 10
G4	270	Sequencing quality too poor to tell but has region of exon8; need to purify bands	Unknown.

		further to identify (may be equivalent to C5 based on prod length and that clean sequence in both was exon8).	
G5	300	Sequencing quality too poor to tell but has region of exon8; need to purify bands further to identify (may be same band as G4).	Unknown.
G6	300/310 (Sequence length 340 bp)	Identical to C6, except that exon10 goes through to rev primer	Exon 7 (though only 2bp) – exon 8 –exon8b-exon8c-exon9-exon9b-exon 10

9.11 *Tcf7l2* splice variantrs on high sugar and free fatty acids and compare them with controls. A) Region 1 and B) Region 2

A)



B)