LUND UNIVERSITY

# Mechanisms by which variants in the TCF7L2 gene increase the risk of developing Type 2 diabetes

Zhou, Yuedan

2013

# Mechanisms by which variants in the *TCF7L2* gene increase the risk of developing Type 2 diabetes

Yuedan Zhou

LUNDS UNIVERSITET

DOCTORAL DISSERTATION

By due permission of the Faculty of Medicine, Lund University, Sweden.

To be defended at CRC Aulan, Jan Waldenströmsgatan 35,
Skåne Universitet Sjukhus in Malmö, Sweden
Friday 10th of January 2014 at 9 a.m. for the degree of Doctor of Philosophy

FACULTY OPPONENT:

Professor Pål R. Njølstad

Department of Clinical Science

University of Bergen

| Organization | Document name |
|---|---|
| LUND UNIVERSITY | Doctoral Dissertation |
| Department of Clinical Science | Date of issue December 13, 2013 |
| Author(s) | Sponsoring organization |
| Yuedan Zhou | |


Title and subtitle

Mechanisms by which variants in the *TCF7L2* gene increase the risk of developing Type 2 diabetes


## Abstract

Type 2 diabetes mellitus (T2DM) is a heterogeneous disease with a multifactorial aetiology comprising of genetic and environmental factors. The common variant most highly associated with T2DM known to date is a SNP rs7903146 in the *TCF7L2* gene. However, the role *TCF7L2* plays in the development of T2DM was unclear. Clinical, animal and *in vitro* studies have revealed an association of the risk T-allele of rs7903146 with impaired β cell function including glucose- and incretin-stimulated insulin secretion and β cell survival. The aim of this thesis is to elucidate the function of *TCF7L2* in the β cells and explain the mechanisms by which genetic variants in *TCF7L2* confer the increased risk of developing T2DM.

*TCF7L2* is a transcription factor of the WNT signalling pathway. Risk genotype carriers of rs7903146 displayed higher *TCF7L2* mRNA expression in islets of Langerhans, reduced insulin content and perturbed glucose stimulated insulin secretion. We have identified a large number of target genes (study I) and target networks (study II) of *TCF7L2* that regulate β cell survival, proinsulin expression and insulin maturation. The expression of many of these *TCF7L2* target genes, including T2DM associated genes/loci, correlated with *TCF7L2* mRNA expression in CC, but not in CT/TT genotype carriers in human pancreatic islets. These data indicate that *TCF7L2* has a central role in insulin synthesis and secretion, as well as in the regulation of other T2DM associated genes. We provided a possible explanation for the large impact *TCF7L2* has on the risk of T2DM. The molecular link between the T-risk allele and the diabetogenic action of *TCF7L2* in β cells is still unclear. *In vitro* and animal experiments have indicated that the 92kb around rs7903146 locus is an enhancer region, and that the T-allele has a higher activity compared to the C-allele. In study III, we identified a protein (HMGB1) that binds the rs7903146 locus, which potentially influence this enhancer activity. At last, we have also investigated the function of different *TCF7L2* splice variants and their influence on β cell function using Antisense Oligo Nucleotides (study IV). These results indicated that exon 4 of *TCF7L2* has an inhibitory function, influencing both insulin synthesis and β cell survival.



Key words

TCF7L2, rs7903146, Type 2 diabetes, β cell function, β cell survival, insulin secretion, proinsulin synthesis, proinsulin-to-insulin conversion, insulin maturation, open chromain, exon skipping, p53, TP53INP1, ISL1, MAFA, NEUROD1, PDX1, NKX6.1, PCSK1, PCSK2, SLC30A8, GWAS.

Classification system and/or index terms (if any)

| Supplementary bibliographical information | Language English |
|---|---|
| ISSN and key title<br>1652-8220 | ISBN<br>978-91-87651-30-4 |

| Recipient's notes | Number of pages | Price |
|---|---|---|
| | Security classification | |

# Mechanisms by which genetic variants in *TCF7L2* increase the risk of developing Type 2 diabetes

Yuedan Zhou


LUNDS
UNIVERSITET

Cover picture depicted the German violinmaker Jacob Stainer (1617-1683, left) and his wife (in the bed), who had acute diabetic symptom. The physician (right) tasted the urine in the chamber pot. The cover was painted by Mr. Zhengrong Zheng.



Faculty of Medicine, Department of Clinical Science
ISBN 978-91-87651-30-4
ISSN 1652-8220

Printed in Sweden by Media-Tryck, Lund University
Lund 2013

To my grandparents

# Contents

# Abstract

Type 2 diabetes mellitus (T2DM) is a heterogeneous disease with a multifactorial aetiology comprising of genetic and environmental factors. The common variant most highly associated with T2DM known to date is a SNP rs7903146 in the *TCF7L2* gene. However, the role *TCF7L2* plays in the development of T2DM was unclear. Clinical, animal and *in vitro* studies have revealed an association of the risk T-allele of rs7903146 with impaired β cell function including glucose- and incretin-stimulated insulin secretion and β cell survival. The aim of this thesis was to elucidate the function of *TCF7L2* in the β cells and explain the mechanisms by which genetic variants in *TCF7L2* confer the increased risk of developing T2DM.

*TCF7L2* is a transcription factor of the WNT signalling pathway. Risk genotype carriers of rs7903146 displayed higher *TCF7L2* mRNA expression in islets of Langerhans, reduced insulin content and perturbed glucose stimulated insulin secretion. We have identified a large number of target genes (study I) and target networks (study II) of *TCF7L2* that regulate β cell survival, proinsulin expression and insulin maturation. The expression of many of these *TCF7L2* target genes, including T2DM associated genes/loci, correlated with *TCF7L2* mRNA expression in CC, but not in CT/TT genotype carriers in human pancreatic islets. These data indicate that *TCF7L2* has a central role in insulin synthesis and secretion, as well as in the regulation of other T2DM associated genes. We provided a possible explanation for the large impact *TCF7L2* has on the risk of T2DM. The molecular link between the T-risk allele and the diabetogenic action of *TCF7L2* in β cells is still unclear. *In vitro* and animal experiments have indicated that the 92kb around rs7903146 locus is an enhancer region, and that the T-allele has a higher activity compared to the C-allele. In study III, we identified a protein (HMGB1) that binds the rs7903146 locus, which potentially influence this enhancer activity. At last, we have also investigated the function of different *TCF7L2* splice variants and their influence on β cell function using Antisense Oligo Nucleotides (study IV). These results indicated that exon 4 of *TCF7L2* has an inhibitory function, influencing both insulin synthesis and β cell survival.

# Studies included in this thesis

**1. Survival of pancreatic beta cells is partly controlled by a TCF7L2-p53-p53INP1-dependent pathway.** *Human Molecular Genetics. 2012 Jan 1;21(1):196-207.* **Zhou Y,** Zhang E, Berggreen C, Jing X, Osmark P, Lang S, Cilio CM, Göransson O, Groop L, Renström E, Hansson O.

**2. *TCF7L2* is a master regulator of insulin synthesis and processing. Zhou Y**, Park S-Y, Su J, Bailey K, Ottosson-Laakso E, Shcherbina L, Zhang E, Thevenin T, Fadista J, Oskolkov N, Bennet H, Wierup N, Fex M, Rung J, Wolheim C, Nobrega M, Renström E, Groop L and Hansson O. ***Manuscript submitted***

**3. The SNP rs7903146 is bound by HMGB1 protein in the open chromatin region of the *TCF7L2* gene. Zhou Y**, Oskolkov N, Matins B, Ratti J, Kock K-H, Jing S, Martin B, Oskolkova M, Osmark P, Göransson O, Bacon J, Li W, Bucciarelli S, Cilio C, Brazma A, Thatcher B, Rung J, Renström E, Groop L and Hansson O. ***Manuscript***

**4. The role of alternative splicing in the *TCF7L2* gene on islets function. Zhou Y,** Osmark P, Aartsma-Rus A, Ström K, van Ommen GJ, Renström E, Hansson O and Groop L. ***Ongoing project***

# Publications not included in the thesis

1. **Molecular function of TCF7L2: Consequences of TCF7L2 splicing for molecular function and risk for Type 2 diabetes.** *Current Diabetes Report. 2010 Dec;10(6):444-51.* Review. Hansson O, **Zhou Y**, Renström E, Osmark P.
2. **A Common Variant Upstream of the PAX6 Gene Influences Islet Function in Man.** *Diabetologia 2011 Aug* Ahlqvist E, Turrini F, Lang S, Taneera1 J, Zhou Y, Almgren P, Hansson O, Isomaa B, Tuomi T, Eriksson K, Eriksson J, Lyssenko V and Groop L.
3. **A systems genetics approach identifies novel genes and pathways for Type 2 diabetes in human islets**. *Cell Metabolism 2012 Jul 3;16(1):122-34.* Taneera J, Lang S, Sharma A, Fadista J, **Zhou Y,** Ahlqvist E, Jonsson A, Lyssenko V, Vikman P, Hansson O, Parikh H, Salehi A, Korsgren O, Soni A, Krus U, Zhang E, Jing X, Esguerra.J, Wollheim C, Salehi A, Rosengren A, Renström E and Groop L.
4. **First degree relatives of Type 2 diabetic patients have reduced expression of genes involved in fatty acid metabolism in skeletal muscle.** *Journal of Clinical Endocrinology & Metabolism 2012 Jul;97(7):E1332-7.* Elgzyri T, Parikh H, **Zhou Y**, Dekker-Nitert M, Rönn T, Segerström A, Ling C, Franks P, Wollmer P, Eriksson K, Groop L, Hansson O.
5. **Role of TCF7L2 risk variant and dietary fibre intake on incident Type 2 diabetes.** *Diabetologia 2012 Oct;55(10):2646-54.* Hindy G, Jing X, **Zhou Y**, Hansson O, Renström E, Wirfält E and Orho-Melander M.
6. **Secreted frizzled-related protein 4 reduces insulin secretion and is overexpressed in Type 2 diabetes.** *Cell Metabolism 2012 Nov 7;16(5):625-33* Mahdi T, Hänzelmann S, Salehi A, Muhammed Sj, Reinbothe TM, Tang Y, Axelsson AS, **Zhou Y**, Jing X, Almgren P, Krus U, Taneera J, Blom AM, Lyssenko V, Esguerra JL, Hansson O, Eliasson L, Derry J, Zhang E, Wollheim CB, Groop L, Renström E, Rosengren AH.

# Abbreviations

| | |
|---|---|
| 2OMePS | 2’-O-methyl RNA phosphorothioate modification |
| ADAMTS9 | A disintegrin and metalloproteinase with thrombospondin motifs 9 |
| AON | Antisense Oligo Nucleotides |
| AS | Alternative splicing |
| ASE | Allele-specific expression |
| AXIN2 | Axis inhibition protein 2 |
| CART | Cocaine- and amphetamine-regulated transcript |
| ChIP | Chromatin immunoprecipitation |
| CREB | cAMP response element-binding protein |
| CtBP | C-terminal protein |
| DGI | Diabetes Genetics Initiative |
| DHS | DNAseI hyper-sensitive site |
| DIAGRAM | DIAbetes Genetics Replication And Meta-analysis |
| DKK1 | Dickkopf WNT signalling pathway inhibitor 1 |
| DLS | Dynamic light scattering |
| ELISA | Enzyme-linked immunosorbent assay |
| EMSA | Electrophoretic mobility shift assay |
| eQTLs | Expression quantitative trait loci |
| ES | Exon skipping |
| ESE | Exonic splicing enhancer |
| ESS | Exonic splicing silencer |
| FAIRE | Formaldehyde-assisted isolation of regulatory elements |
| FDR | False discovery rate |
| FOXO | Forkhead box transcription factors sub group O |
| FTO | Fat mass and obesity-associated protein |
| FUSION | Finland-United States Investigation of NIDDM Genetics |
| FZL | Frizzled family G protein-coupled membrane receptors |
| G6PC | Glucose-6-phosphatase |

| | |
|---|---|
| GIP | Glucose dependent insulinotropic peptide |
| GLP-1 | Glucagon-like peptide-1 |
| GSIS | Glucose stimulate insulin secretion |
| GSK3b | Glycogen synthase kinase 3b |
| GWAS | Genome-Wide Association Study |
| HbA1c | Haemoglobin A1c |
| HCC | Hepatocellular carcinoma |
| HIPK2 | Homeodomain-interacting protein kinase 2 |
| HMGB1 | High-mobility group protein B1 |
| HMM | Hidden Markov Model |
| HNRNPs | Heterogeneous ribonucleoproteins |
| ISL1 | Insulin gene enhancer protein (Islets-1) |
| KO | Knock out |
| LD | Linkage disequilibrium |
| LinDA | Single-tube linear DNA amplification |
| LRP | Low-density lipoprotein-related protein |
| MAFA | V-maf musculoaponeurotic fibrosarcoma oncogene homolog A |
| MAGIC | Meta-Analyses of Glucose and Insulin-related traits Consortium |
| MODY | Maturity-onset diabetes of the young |
| NGS | Next generation sequencing |
| NEUROD1 | Neurogenic differentiation 1 |
| NFκB | Nuclear factor kappa-light-chain-enhancer of activated B cells |
| NKX2.2 | NK2 homeobox 2 |
| NKX6.1 | NK6 homeobox 1 |
| PCP | Planar cell polarity |
| PCSK1 | Code for prohormone convertase 1/3 |
| PCSK2 | Code for prohormone convertase 2 |
| PDX1 | Pancreatic and duodenal homeobox 1 |
| PEPCK1 | Phosphoenolpyruvate carboxykinase |
| SAXS | Small-angle X-ray scattering |
| sFrps | Secreted frizzled-related protein family |
| SNP | Single nucleotide polymorphism |
| snRNP | Small nuclear ribonucleoprotein particle |
| SPRK | Serine-arginine protein kinase |
| SR | Serine / arginine-rich splice factor |
| T1DM | Type 1 diabetes mellitus |

| | |
|---|---|
| T2DM | Type 2 diabetes mellitus |
| TCF7L2 | Transcription factor 7-like 2 |
| TIRFM | Total internal reflection fluorescence microscopy |
| TP53 | Tumour protein 53 |
| TP53INP1 | TP53 inducible nuclear protein 1 |
| TSS | Alternative transcript start sites |
| ZDF | Zucker diabetic fatty rat |
| ZnT8 | Zinc transporter 8 |

# Introduction

## Genetics of Type 2 Diabetes Mellitus

Modern genetic analysis is no doubt one of the most powerful tools in diabetes research (and other complex diseases). It has been estimated that 30% –70% of T2DM risk may be due to genetic factors [1]. Prior to the Genome-Wide Association Study (GWAS) era, linkage analyses facilitated the identification of several highly penetrant genetic variants for monogenic forms of diabetes *e.g.* Maturity-onset diabetes of the young (MODY) [2, 3]. The linkage study approach works well if the genetic composition of the disease involves only a few or a single variant with strong effect, but the method is not well-suited to identify genes involved in complex disease with cumulative effects of many components.

A SNP is a single nucleotide difference between individuals. It is estimated that SNPs occur at a frequency of approximately one in every thousand base pairs (bp) throughout the genome [4]. More than 14 million SNPs have been annotated so far. Therefore, SNPs have become by far the most abundant class of genetic variants and the most traceable to interrogate. These polymorphisms can be used to distinguish small differences both within a population and among different populations.

By selecting 'tag' SNPs that are in high linkage disequilibrium (LD) with other variants, many more SNPs in the genome can be imputed by genotyping a relatively small amount of tag SNPs. Therefore, novel genes and pathways involved in a complex disease can be identified in a hypothesis-free manner using GWAS. But the SNPs identified through the initial GWAS studies consortia such as DGI (Diabetes Genetics Initiative) and FUSION (Finland-United States Investigation of NIDDM Genetics) could only account small amounts of the estimated heritability of T2DM

[5]. Therefore larger consortia were established to meta-analyse data. DIAGRAM (DIAbetes Genetics Replication And Meta-analysis), DIAGRAM+ and MAGIC (Meta-Analyses of Glucose and Insulin-related traits Consortium) aim to discover more genetic variants with smaller effect size, not only associated with T2DM but also with other disease related traits using multi-ethnic case-control cohorts.

GWAS has identified over 65 common risk variants associated with T2DM during the last decade [6-14]. The MAGIC consortium focused on identifying genetic variants associated with glucose and insulin-related traits, including fasting glucose, fasting insulin, 2-hours circulating glucose and $HbA_{1c}$. In addition, meta-analysis of more sophisticated measures of insulin secretion and insulin sensitivity were performed. Most of the GWAS signals map to noncoding regions of the genome, which makes it difficult to establish functional links to specific genes and transcripts. Nevertheless, some of the GWAS genes associated with T2DM and insulin related phenotypes became very good indicators of molecular pathway analysis during my PhD study (Study I and II).

There is still missing heritability, which has made researchers doubt the "common variant for common disease" hypothesis. Therefore new consortia, such as GO-T2D have been formed aiming to conduct extensive sequencing to identify rare variants associated with increased risk of T2DM and related traits (not limited to SNPs, but also other types of genetic variation, such as structural variations). After the discovery of an associated genetic variant, the big question is:

## How does a specific genetic variation influence a phenotype or risk of disease?

If we return to the SNPs for a while, human genomes are mostly identical. Only a small portion of the genetic difference results in huge phenotypic variations observed among individuals of a species. The complex interaction between genes and the environment also makes the

understanding and quantification of human phenotypic variations difficult. A point mutation may or may not change the function of a gene or its expression. If the mutation provides greater fitness then the favourable allele will increase in frequency by positive selection in the population. But the majority of existing genetic variation is evolutionarily neutral and for these variants, genetic drift will be a main factor influencing its frequency.

If a variant influences a phenotype this may be through many different mechanisms, *e.g.* via influencing: (1) promoter and enhancer activity (*i.e.* expression), (2) mRNA stability and alternative splicing or (3) protein functions. These are also reasons why certain SNPs predispose an individual to a common disease. This is the task of functional studies after the identification of a common variant associated a complex disease.

## The association of genetic variants in *TCF7L2* with T2DM

The strongest common variant associated with T2DM identified to date is rs7903146 in the *TCF7L2* gene, which has a per allele odds ratio of ~1.35 for risk of the disease [8-10]. The first signal detected in the gene was a microsatellite, identified using the linkage study. In 2006, Grant *et al*. reported that a microsatellite (DG10S478) in *TCF7L2* was associated with an increased risk of T2DM [6]. The DG10S478 is located in intron 3 of *TCF7L2*. Two SNPs investigated within intron 3 and 4 of *TCF7L2* (rs7903146 and rs12255372) were in strong linkage disequilibrium with DG10S478 and showed similar robust associations with T2DM [6]. Subsequent studies in many other ethnic groups have confirmed that rs7903146 and rs12255372 are the two SNPs most strongly associated with T2DM [15, 16]. Several other SNPs in the 3'-end of *TCF7L2* have also been associated with T2DM in an East Asian population [17]. Since 2006, numerous studies have tried to functionally link the risk T-allele of rs7903146 with T2DM by investigating the *TCF7L2* expression level, alternative splicing (AS) and the physiological function of *TCF7L2* in

pancreas and liver. Most of the 65 common variant signals identified by GWAS are associated with impaired pancreatic islet function, indicating that pancreas could be the primary target tissue of the *TCF7L2*-associated risk of developing T2DM [14, 18]. The aim of my PhD thesis was to establish a molecular link between the risk T-allele of rs7903146 with T2DM, particularly by focusing on its role in pancreatic islets.

## The expression of *TCF7L2* in risk allele carriers and non-risk allele carriers

Using absolute quantification, the expression of *TCF7L2* mRNA in various human tissues has been studied in our laboratory [19]. *TCF7L2* is most highly expressed in pancreatic islets, followed by adipose tissues (visceral and subcutaneous) and blood. The expression in the skeletal muscle is very low compared to other tissues [19]. It was later also shown to be expressed in the liver [20, 21]. The influence of the risk T-allele of rs7903146 on *TCF7L2* expression is, however, an on-going debate.

An increased expression of *TCF7L2* in islets from both healthy risk-allele carriers and T2DM patients has been reported [22-24]. In non-diabetic individuals, homozygous risk-genotype carriers (TT) have 2 - 2.6 fold higher expression of *TCF7L2* mRNA in pancreatic islets compared to CC-genotype carriers [23, 24], while diabetic homozygous TT-genotype carriers display a 5-fold increased *TCF7L2* expression compared to CC-genotype carriers [22]. However, these findings have not been consistently replicated [25]. Expression quantitative trait loci (eQTLs) are genomic loci that regulate expression levels of mRNA or protein [26]. If the eQTL is located in the vicinity of a gene, it is considered to be a *cis*-eQTL and if the gene(s) regulated are physically far away from the SNP it is called *trans*-eQTL. A recent report demonstrated that most eQTLs influence expression over considerable genetic distances [27]. In the case of *TCF7L2*, unpublished data have shown that the rs7903146 is indeed a *cis*-eQTL of *TCF7L2* (study II). In obese ZDF rats (Zucker diabetic fatty (*fa/fa*) T2DM rat model), an increase of *Tcf7l2* mRNA expression in islets was also

observed [28]. The expression of *Tcf7l2* progressed together with the development of T2DM. An elevated expression was observed during the β cell compensation phase together with increased β cell mass, followed by a decrease after the onset of the diabetes [29]. It was also reported that the protein expression of TCF7L2 was reduced in islets from diabetic mice compared to the healthy controls whereas surprisingly, mRNA expression was higher in these mice [28]. Some researchers have suggested that the increase in *Tcf7l2* expression in concert with the development of the disease indicates that the increase of *Tcf7l2* expression might be a consequence, rather than a cause of defective insulin secretion [29]. Another possible explanation is that since there are multiple possible transcripts of *TCF7L2* due to alternative splicing, a genuine increase in mRNA levels may represent increased levels of transcripts encoding less active isoforms, or a transcript of *TCF7L2* with inhibitory effect on insulin secretion.

## TCF7L2 is a member of the WNT signalling pathway

TCF7L2 is a member of the WNT signalling pathway. Prior to its association with T2DM, *TCF7L2* was known for its involvement in cancer development [30-32]. Aberrant activation of the WNT pathway might lead to the development of colorectal cancer and other types of tumours.

A large family of cysteine-rich glycoprotein ligands (encoded by 19 different genes in mammals) secreted by undifferentiated or inflammatory cells and adipocytes activates the WNT signalling pathway. WNT ligands can act as autocrine or paracrine signals as well as distant endocrine signals [33-35] and coordinate cell fate during the development and developmental homeostasis [36]. WNT-ligands are also involved in normal cholesterol metabolism, glucose-induced insulin secretion [37] and the synthesis of the incretin hormone glucagon-like peptide-1 (GLP-1) [38-41]. WNT-ligands bind the "Frizzled family G protein-coupled membrane receptors" (Fzls), or the "low-density lipoprotein-related

protein" family of single-domain transmembrane receptors (Lrps). Fzl-Lrp co-receptors activate different signalling pathways, and are stimulated by WNT3a, WNT5a or inhibited by Dickkopf WNT signalling pathway Inhibitor 1 (Dkk1) and the "secreted frizzled-related protein" family (sFrps). The WNT-ligands WNT3a and WNT5a stimulate insulin secretion in wild-type mice but not in *Lpr5*$^{-/-}$ mice [37], suggesting that WNT-ligands require a functional LRP5 to regulate insulin secretion. The WNT pathway is also involved in regulating β cell development and proliferation and thereby pancreatic islet cell mass [42]. Purified WNT3a stimulates proliferation of both the mouse β cell line MIN6 and primary mouse pancreatic β cells, possibly through the cell cycle regulators cyclin D1, cyclin D2, cyclin-dependent kinase 4, and transcription factor Pitx2 [42]. Loss of β-catenin signalling lead to pancreatic hypoplasia [43]. However, it has also been reported that WNT activation by WNT3a is absent in pancreatic β cells from both obese and lean mice [44]. On the other hand, it is possible that WNT pathway and thereby TCF7L2 can be activated by other stimuli, such as incretin hormones.

TCF7L2 is one of the downstream transcription factors of the WNT pathway called as the β-catenin-dependent (or canonical) pathway, which uses the bipartite transcription factor β-catenin/TCF as the major effector, but also of β-catenin-independent pathways [45]. In the canonical pathway, WNT binding to Fzl-Lrp co-receptors induces Disheveled and Axin to dismantle the "destruction protein complex". This protein complex contains several kinases that act sequentially to mark proteins for proteosomal degradation. The most important of these kinases is Glycogen synthase kinase 3b (GSK3b), which constitutively phosphorylates β-catenin and other protein factors, thereby marking them for degradation. Disassembling of the "destruction complex" results in accumulation of β-catenin. WNT-induced β-catenin stabilization leads to its entry into the nucleus where it binds members of the T-cell factor/lymphoid enhancer factor (TCF/LEF) family of transcription factors. Thus, β-catenin acts as an intracellular cofactor that activates TCF/LEF factors in a WNT-dependent fashion. FOXO (Forkhead box transcription factors subgroup O) is a mediator of stress, which competes with TCF proteins for the

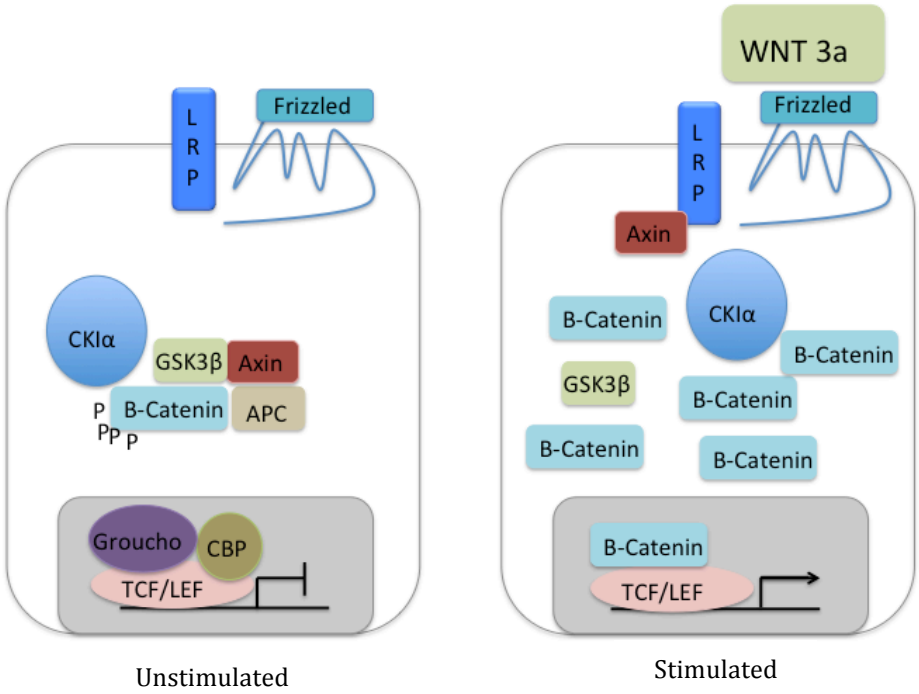


**Figure 1. WNT signalling pathway** Under no ligand stimulation (Left panel) and upon WNT ligand stimulation (WNT3a, Right panel), adopted from [49].

limited pool of free β-catenin [46]. There are four known TCF/LEF factors in mammals, *i.e.* TCF7 (also known as TCF1), LEF1, TCF7L1 (also known as TCF3), and TCF7L2 (formerly known as TCF4). In the absence of β-catenin, TCF proteins bind to an evolutionarily conserved motif A-C/G-A/T-T-C-A-A-A-G at target gene promoters and repress their activity [47]. The transcription of target genes is activated in the presence of β-catenin. The physiological effect of WNT signalling is, however, complicated for several reasons: (1) a cell can contain multiple effectors; (2) the diversity of transcription factors is multiplied by different promoter usages of the factors and (3) alternative splicing generates multiple LEF/TCF factor isoforms with distinct transcriptional properties and the relative proportions of the isoforms.

The two major β-catenin-independent non-canonical pathways are

the Wnt/Cafl2 pathway, dependent on G-protein association to Fzl receptors, and the planar cell polarity (PCP) pathway, dependent on LRP co-receptors other than LRP5/6 that regulates transcription-independent actin cytoskeleton reorganization [45, 48].

## TCF7L2 functions in the pancreatic β cells

The risk T-allele of rs7903146 has been associated with dysregulated pancreatic β cell function and T2DM [6, 8, 9]. The clinical data showed that individuals carrying the risk T-allele had (1) reduced insulin secretion [22, 50, 51], (2) elevated plasma proinsulin level and an increased proinsulin-to-insulin ratio, suggesting the possibility of perturbed proinsulin processing [12, 51-58] and (3) impaired incretin-potentiated insulin secretion [22, 59]. Many *in vitro* and animal studies have therefore focused on elucidating mechanisms underlying these observed β cell defects. In line with the clinical observations, *TCF7L2* also influences both glucose- and incretin-stimulated insulin secretion in isolated islets. Moreover, *TCF7L2* plays a pivotal role in β cell apoptosis, exocytosis (calcium handling) and islet cell morphology. For example, siRNA-mediated *TCF7L2* silencing impairs glucose-stimulated and incretin-potentiated insulin secretion, but not KCL-induced insulin secretion in isolated mouse islets, MIN6 or INS1 832/13 cells [60]. Silencing *Tcf7l2* did not alter the insulin content in mouse islets, but reduced preproinsulin expression [61]. Over-expression of *Tcf7l2* did not affect glucose-stimulated insulin secretion in mouse islets or calcium influx in INS1 832/13 cells, but an increased preproinsulin expression was observed in mouse islets [61]. In human pancreatic islets from cadaver donors, the T-allele of rs7903146 has been associated with reduced basal and glucose-stimulated insulin secretion. But insulin content was comparable in risk and non-risk carriers [23]. In contrast, we have later demonstrated that risk T-allele is associated with less insulin content in pancreatic islets (study II).

In line with results from clinical studies [59, 62], a reduction in GLP-1- and GIP-potentiated glucose-stimulated insulin secretion was

observed in isolated islets depleted of *TCF7L2* [60, 63]. Although the promoter of the preproinsulin gene contains a binding site for TCF7L2 and thereby could influence GLP-1 secretion [66], no defects in GLP-1 secretion has been reported in T-allele carriers [59, 62, 64]. Focus has later been switched to exploring causes of deranged incretin signalling in the β cells. Decreased TCF7L2 protein expression in diabetic mice and T2DM patients is associated with reduced pancreatic islet GLP-1 and GIP receptor expression, which contribute to impaired β cell function [28, 61]. Apart from its effect on insulin secretion, *TCF7L2* has also been shown to mediate GLP-1-stimulated β cell proliferation through activation of the WNT signalling pathway. Both TCF7L2 and β-catenin are required for GLP-1-stimulated proliferation of the rat pancreatic β cell line INS-1 832/13 [12]. Increased apoptosis was observed in *TCF7L2*-silenced islets [28, 60] and in individuals carrying the T-allele of rs7903146 [59, 65]. This is very likely one of the consequences of a lower expression of GIP and GLP-1 receptors due to *TCF7L2* dysregulation, possibly because of decreased incretin-stimulated AKT and FOXO phosphorylation [60]. We have also shown that TCF7L2 inhibits p53 and p53INP1 expression in non-apoptotic conditions and dysregulation of *TCF7L2* may fail to inhibit the p53-p53INP pathway leading to increased apoptosis [66]. Over-expression of *TCF7L2* protects islets from cytokine-induced cell apoptosis [60]. However, a change of cell viability was not observed in mouse islets depleted of *Tcf7l2* [63]. In support of this, no association with impaired cell viability was found in T-allele carriers [23].

Exocytosis is a key step in secretion of insulin. SiRNA-mediated *Tcf7l2* silencing in primary mouse islets, MIN6 and INS1 832/13 cells leads to defective insulin vesicle movement, vesicle fusion [63], glucose-stimulated ATP/ADP increment and calcium influx. But readily releasable insulin vesicle number was not affected by *TCF7L2* silencing [63]. In contrast, we have observed decreased readily releasable insulin vesicle numbers after *Tcf7l2* silencing in INS1 832/13 cells (Further discussion in study II). Overexpression of *Tcf7l2* had a slight inhibitory effect on glucose stimulated calcium influx [63].

A morphological analysis was carried out to compare the islet composition of risk and non-risk allele carriers [23]. The total mass of pancreatic tissue was not altered in pancreatic sections from CT/TT-genotype carriers compared with CC-genotype carriers. However, TT-genotype carriers had a lower number of islets and tended to have islets with larger diameters. Higher proportion of α-cells was also observed in the TT-genotype carriers, but not in CT-carriers compared to the CC-carriers [23].

In many clinical studies, the risk T-allele of *TCF7L2* has been associated with elevated plasma proinsulin level and increased proinsulin / insulin ratio [12, 51-58], but no molecular explanation for these observations has been provided. In study II of this thesis, we have also demonstrated that TCF7L2 regulates the expression of the prohormone convertase 1/3, convertase 2 and ZnT8 in pancreatic β cells and silencing *Tcf7l2* in INS1 832/13 cells leads to an increased proinsulin-to-insulin ratio, indicating aberrant insulin processing.

To summarize, many *in vivo* and *in vitro* studies have provided evidence for a crucial role of *TCF7L2* in pancreatic β cells. *TCF7L2* regulates glucose- and incretin-stimulated insulin secretion, in part through calcium handling, but more likely through transcriptional regulation (Study II), proinsulin-to-insulin conversion and probably development and survival of the islets.

## *Tcf7l2* mouse models revealed unprecedented metabolic functions in pancreas and liver

Several genetically engineered *Tcf7l2* knock-out (KO) mouse models have been generated and some KO models were manipulated in a cell-type-specific manner. But different models have not always been able to display reproducible phenotypes. In addition to whole-body KO models, specific disruption of *Tcf7l2* in pancreas, pancreatic β cells, hepatocytes as well as a transgenic *Tcf7l2* overexpression models were generated.

**Whole-body *Tcf7l2* knockout mouse models**

The first homozygous whole-body *Tcf7l2* knock out (*Tcf7l2*$^{-/-}$) mice died shortly after birth due to a lack of proliferative stem cells that should transform from intestinal endoderm into epithelium in the prospective gut crypt regions [67]. The pups died within 24 hours after birth with milk in the stomach. *TCF7L2* is thus apparently crucial for the development program of the small intestine. Constitutive expression of *TCF7L2* may contribute to the malignant transformation of cancer [67] and this mouse model was created years ago in order to study the pathogenesis of colon cancer.

Inducible homozygous KO of *Tcf7l2* after the crucial point of *Tcf7l2* expression during development results in viable pups [67-70]. Newborn *Tcf7l2*$^{-/-}$ mice have reduced body weight with significantly lower blood glucose directly after birth [69], and lower body weight when fed high fat diet during adulthood [68]. Heterozygote *Tcf7l2*$^{+/-}$ mice also display decreased body weight compared to wild-type littermates [69] and lower plasma fasting glucose, fasting insulin [68, 69], free fatty acids, triglycerides, and cholesterol during adulthood compared to wild type littermates [69]. Furthermore the heterozygote *Tcf7l2*$^{+/-}$ mice show increased insulin sensitivity, enhanced glucose tolerance [68-70] and reduced hepatic gluconeogenesis and lipogenesis [69, 70].

**Pancreas specific knockout mouse model (*pTcf7l2*$^{-/-}$)**

A pancreas-specific KO mouse (*pTcf7l2*$^{-/-}$) was generated using the *Pdx1* promoter Cre system [61] in order to delete *Tcf7l2* in all cells of pancreatic lineage. The *pTcf7l2*$^{-/-}$ mouse is viable, but displays defects during development, mainly in the accumulation of stem cells in the intestinal crypt [67]. Isolated islets from *pTcf7l2*$^{-/-}$ mice fed high fat diet have abnormalities in glucose- and GLP-1-induced insulin secretion. Glucose intolerance can be observed from week 12 when challenged with oral glucose. Until week 20, the mice were glucose intolerant when challenged both with oral and intraperitoneal glucose, suggesting an

impaired incretin effect. *PTcf7l2*$^{-/-}$ mice displayed thus a pronounced age-dependent glucose intolerance phenotype, pointing at an impaired incretin effect. In line with the previous report [28], GLP-1 receptor expression was significantly decreased in these mice. *PTcf7l2*$^{-/-}$ mice fed high fat diet also had decreased β cell mass after a glucose challenge [61]. This suggests that a reduced TCF7L2 activity in the pancreas might contribute to the β cell loss induced by gluco- and lipotoxicity, a link also later observed by us (study I). A reduction of *Ccnd1* expression could partly explain the reduction of cell proliferation [61]. No difference in circulating glucagon or GLP-1 concentrations was observed in *pTcf7l2*$^{-/-}$ mice [61]. The *pTcf7l2*$^{-/-}$ model clearly showed that *TCF7L2* has indispensable functions in the pancreas and islets of Langerhans.

**β cell-specific knockout mouse model**

An inducible β cell *Tcf7l2* knockout (*βTCF4KO*) mouse has been generated using the RIP2 promoter system by breeding a cre-ERT2 strain with *Tcf7l2*loxP/+ mice [69]. In this model, no difference in circulating insulin or glucose stimulated insulin secretion from isolated islets was observed. *βTCF4KO* mouse had normal intraperitoneal glucose tolerance. Notably, this mouse model showed a very different phenotype than the pancreas specific *pTcf7l2*$^{-/-}$ mouse model described above, suggesting that *TCF7L2* is not involved in β cell function. It cannot be excluded that the construction might influence the phenotype by affecting *TCF7L2* expression in hypothalamic neurons. The discrepancy between these two mouse models could possibly be explained by a compensatory up-regulation of other genes in the *βTCF4KO* mouse, *e.g.* TCF7L1, capable of similar functions as TCF7L2. In support of this hypothesis, the model showed a lack of regulation of well-established TCF7L2-target genes like Axin2 [69]. Another unpublished β cell specific KO mouse model using Ins1.Cre system had very similar phenotype as the pancreatic specific *pTcf7l2*$^{-/-}$ mouse model [71] (Ins1.Cre; J. Ferrer, B. Thorens, unpublished).

**Transgenic mouse models**

In transgenic mouse models created in the laboratory of our collaborator, Pr. Marcelo Nobrega at the University of Chicago, different regions centred on human *TCF7L2* gene (including introns, exons and potential non-coding regulatory regions around the gene) was inserted into bacterial artificial chromosomes (BACs). The enhancer activities, or the regulatory potential of different regions in human *TCF7L2* gene were assessed by measuring β-galactosidase activity coded by *LacZ* cassette in the BACs. The strongest enhancer activity was seen in the 92kb interval bearing rs7903146 locus [70]. Deleting the 92 kb genomic interval resulted in complete silencing of exogenic human LacZ expression in adult islets [70]. The suggested long-rang enhancing ability for *TCF7L2* expression is in fact in line with the finding of open chromatin around the rs7903146 locus in T-allele carriers [72], where a potential enhancer activity of the rs7903146 locus was described. This is also in line with findings in study III of this thesis. This transgenic mouse model confirmed that the rs7903146 locus might exert increased enhancer activity (Further discussion in study III).

Inserting multiple copies of mouse *Tcf7l2* into the genome resulted in a dose-dependent increase in the expression of *Tcf7l2*, and the mouse became increasingly diabetic [70]. In contrast, an inducible *Tcf7l2*$^{-/-}$ model, displayed the lowest glucose and insulin values suggesting that complete lack of *Tcf7l2* expression protects from T2DM. These findings contradict with studies of *TCF7L2* silencing in isolated human and rodent islets, but are in concordance with the observation of increased expression of *TCF7L2* in carriers of the T-allele [70].

**Liver-specific knockout and overexpression mouse models**

The liver-specific *Tcf7l2* (*LTcf7l2KO*) knockout mouse model [69] had similar birth weight compared to control littermates, but circulating fasting glucose levels in the KO mice were significantly lower. This was not caused by hyperinsulinemia. Besides, glucose-stimulated insulin secretion was similar in control and KO mice. This finding was ascribed to

TCF7L2-dependent regulation of crucial genes involved in liver glucose metabolism as the *LTcf7l2KO* mice displayed reduced hepatic glucose production. Opposite to *LTcf7l2KO* mice, liver-specific *Tcf7l2* overexpression using adenovirus induced a significant increase in serum glucose under fasting condition compared to controls [69]. In addition, overexpression of *Tcf7l2* in the liver led to an increased hepatic glucose production via pyruvate stimulation [69]. These findings will be further discussed in the following section.

## The influence of *TCFL2* on hepatic glucose and lipid metabolism

The *TCF7L2* expression in the liver has major effects on metabolism. The T-allele of rs7903146 has been associated with increased hepatic glucose production [62, 73], most likely due to impaired hepatic insulin sensitivity [74, 75]. However, the increase in hepatic glucose production has not been replicated in all studies [76-78].

The WNT/β-catenin pathway is a major determinant of the hepatocyte metabolism [79, 80]. Food intake and insulin stimulate the expression of *Tcf7l2* in mouse hepatocytes [81]. Furthermore, WNT activation by lithium or WNT3a, but not WNT11a reduces gluconeogenesis [21]. The *LTcf7l2KO* mice generated by Boj *et al*. suggested that the reduced circulating glucose level was a consequence of reduced hepatic gluconeogenesis. Glucagon-stimulated glucose release from hepatocytes was also reduced in the *LTcf7l2KO* mice compared to wild type mice [69]. *TCF7L2* seems to regulate expression of rate-limiting glucogeogenic enzymes [21, 68], such as phosphoenolpyruvate carboxykinase (PEPCK1, encoded by PCK1) and glucose-6-phosphatase (G6Pase, encoded by G6PC). Chromatin immunoprecipitation followed by sequencing (ChIP-seq) analysis revealed that TCF7L2 binds and regulate the expression of both these genes [21, 69, 81]. *G6pc* was identified as a direct target gene of TCF7L2 in all studies [21, 68, 69]. Its expression was up-regulated in *Tcf7l2*$^{+/-}$ mice [68] and decreased after *Tcf7l2*

overexpression [69]. However, when fed with high fat diet, the *Tcf7l2*$^{+/-}$ mouse exhibits significantly decreased *G6pc* expression. Furthermore, knocking down *Tcf7l2* by siRNA in primary hepatocytes resulted in decreased *G6pc* expression [68]. Overexpression of *Tcf7l2* in the liver increased *G6pc* expression [69]. *Pck1* was not identified as direct target gene in all studies [69], and changes in *Pck1* expression after *Tcf7l2* manipulation have been inconsistent between studies [68, 69].

Other genes in the liver regulated by TCF7L2 include *Aqp9*, *Acadl*, *Akr1c6*, *Cyp2e1*, *Cyp2c39*, *Slc25a21*, and *Hmgcs2a*, identified using both ChIP-seq and microarray mRNA expression profiling [69]. Silencing of *Tcf7l2* leads to increased basal glucose production in cultured hepatocytes [21, 81], as well as impaired insulin- and metaformin-suppressed glucose production [21]. As TCF7L2 is a down-stream effector of the WNT pathway, alteration of WNT signalling also leads to regulation of these rate-limiting gluconeogenic enzymes in hepatocytes [20]. This might partially explain the impaired hepatic insulin sensitivity observed in risk T-allele carriers. Taken together, TCF7L2 probably regulates gluconeogenesis by influencing the expression of key gluconeogenic enzymes. However, measurement of *TCF7L2* expression in hepatocytes from C- and T-allele carriers is clearly needed to firmly establish the link to the risk genotype.

To conclude this section, it is reasonable to deduce that perturbation of *TCF7L2* expression may influence processes in both pancreas and liver. However, as both organs are crucial for whole-body metabolic regulation and influence each other's functions, it is difficult to infer the causality. Especially as minute decreases in insulin secretion will be reflected by enhanced hepatic glucose production given the extreme sensitivity of hepatic glucose production to small amounts of insulin. In other words, it is difficult to conclude if the risk T-allele exerts its primary influence in the pancreas or the liver, or both.

## Open chromatin

Two studies have explored the potential regulatory function of the rs7903146 locus in pancreatic islets by profiling the chromatin state [72, 82]. Using formaldehyde-assisted isolation of regulatory elements (FAIRE) coupled with high-throughput sequencing (FAIRE-seq), Gaulton *et al.* have identified chromatin-free sites encompassing *cis*-regulatory elements in human islets from three non-diabetic donors. 45% of all identified open chromatin sites were regarded as islets cell specific (3348 clusters of open regulatory elements). The rs7903146 in *TCF7L2* overlapped with an islet-specific FAIRE region. In addition, the risk T-allele was in a more accessible chromatin conformation compared to the non-risk C-allele. This allelic difference was confirmed in 9 heterozygous genotype carriers. The authors advanced the hypothesis that the risk allele exerts its effect on T2DM risk by altering the accessibility of the DNA at this locus. Using a Luciferase assay in MIN6 cells, they demonstrated that the T-allele had greater enhancer activity than the C-allele. A further 10 T2DM-associated loci were located in similar open-chromatin regions.

The other study used a different method (DNAseI-Seq analysis), that identified a distinct set of chromatin regions hyper-sensitive to DNAseI (DHS). It was claimed that many transcription start sites are located in these open-chromatin regions [82]. Six T2DM-associated SNPs, including the rs7903146, were located in DHS sites, which represent sites of putative regulatory elements. The allele-specific enhancer activity of rs7903146 region was compared using a Luciferase assay. Again the risk T-allele had 3-fold higher enhancer activity compared to the non-risk C-allele in the rodent β cell line MIN6 (but not in HeLa cells) [82].

The two studies suggest that altered enhancer activity plays a key role in the molecular mechanism(s) underlying the diabetogenic effect of rs7903146. These findings prompted study III to identify proteins binding to this open chromatin. Although the increased enhancer activity of the T-allele is in accordance with the reported increased expression of *TCF7L2* in islets from patients with T2DM and healthy T-allele carriers, several

inconsistent results still need to be discussed (Further discussion in study III).

## Alternative splicing of *TCF7L2*

The interpretation of the physiological effects of *TCF7L2* regulation is further complicated by the fact that TCF7L2 shows different protein isoforms generated by alternative splicing (AS). *TCF7L2* splicing can contribute to cell metabolism by modifying WNT signalling [83]. AS has been suggested to explain the association between *TCF7L2* SNPs and impaired TCF7L2 function in T2DM. *TCF7L2* comprises of 17 expressed exons, five of which are alternative in several colorectal cancer cell lines (*i.e*. exons 4 and 13–16) [84], and one predicted but never detected exon 4a [72] (Figure 2). Several alternative splice acceptor sites are used in exon 7, 9, 16, and 17. Alternative splice acceptors may cause shifts in reading frames, and alter the length of the protein. The first study of *TCF7L2* AS by Duval *et al*. identified 8 coding variants in colon cancer cells [84]. The alternative exons are mostly gathered in the 3'-end of the gene, exon 16 is rarely present in the cell lines examined. Most splice isoforms are predicted to yield a protein lacking the C-terminal (CtBP) interacting part of the full-length transcript. Only the presence of full exon 17 introduces the binding sites for CtBP1 and CtBP2 in the protein after translation. CtBP1 and 2 are inhibitors of the WNT signalling pathway [84]. An issue is that two nomenclature systems are used to describe *TCF7L2* AS (indicated in Figure 2). Similar *TCF7L2* splicing patterns have been reported in other forms of cancers such as the hepatocellular carcinoma (HCC). The basic splice patterns in different tissues are similar. Different splice variants may vary significantly in their transcriptional activity and thus the regulation of TCF7L2 target genes [85].

Alternative transcript start sites (TSS) have also been found in the 5'-end of the gene in a pool of 8 human tissues [65]. Transcription starting from TSS1 results in proteins containing exon 1 and 2, which encode the β-catenin binding domain. Transcription starting from TSS2 or TSS3 will

result in proteins lacking the β-catenin binding domain [65]. Another shorter form, having only the 5' half of the gene has also been detected [86]. This isoform is predicted to repress TCF7L2-dependent target genes expression, as it possesses the β-catenin binding domain but not the HMG (high-mobility group) box DNA-binding domain [86]. Of note, more than 500 different splice variants of *TCF7L2* are theoretically possible!

In human pancreatic islets, the four predominant splice isoforms are with/without exon 4 at the 5'-end and with/without exon 15 at the 3'-end [19]. No association between rs7903146 and total expression or expression of unique splice forms of *TCF7L2* was observed in a previous study from our laboratory [19]. Several other studies have searched for an association between T2DM risk genotype and *TCF7L2* splice pattern. Most of them did not find any association between splicing and rs7903146 [19,88] with the exception of one study in adipose tissue [87].

The tissue-specific splice patterns of *TCF7L2* have been described in several human tissues (pancreas, pancreatic islets, colon, liver, monocytes, primary lymphocytes, skeletal muscle, subcutaneous and visceral adipose tissue, peripheral blood mononuclear cells, and lymphoblastoid cell lines) [19, 65, 87-89]. Splice variants containing exon 14 at the 3'-end are predominant in subcutaneous and visceral adipose tissue, while exon 15 containing isoforms predominate in liver cells. Exon 4 can be either present or absent at 5'-end of the transcripts [88].

Splice variants containing exon 16 have been described. The highest expression of variants containing exons 14-16 boundary was detected in several areas of the brain (hypothalamus/thalamus, occipital lobe) and in neuronal cell lines. The expression of isoforms containing the 14-16 boundary correlates with the expression of the gene for cocaine- and amphetamine-regulated transcript (CART, also known as CARTPT) in pancreatic islets and colon [90]. It addition, expression of this isoform also correlates with proinsulin expression in islets [65]. However, the level of exon 16-containing isoforms of *TCF7L2* is extremely low in human pancreatic islets [19]. *CARTPT* mRNA expression measured using RNA-sequencing in INS1 832/13 cells was also very low (10 counts / gene,

study II). A possible effect of this isoform on *CARTPT* is probably more important in the central nerve systems than in the pancreas, as CART is an important neurotransmitter in the CNS.

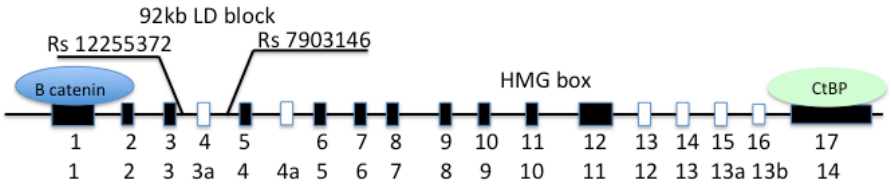


**Figure 2. Summary of the *TCF7L2* genomic organization and its alternative splicing**
Two nomenclatures are used in the studies of *TCF7L2* alternative splicing. Black boxes represent the constitutive exons, and white boxes represent the alternative exons.

Exonic and intronic sequences in pre-mRNA can exert the function of splicing enhancers and silencers. These sequences are recognized and bound by Serine / Arginine-rich splice factors (SR proteins family, SRSFs) and heterogeneous ribonucleoproteins (HNRNPs). Together with five small nuclear ribonucleoprotein particles (snRNPs), SR proteins and HNRNPs form the spliceosome. All exons contain splicing enhancers and / or silencers, which are needed for constitutive splicing before protein translation. AS relies on less efficient regulatory sequences compared to constitutive splicing enhancers. Splicing may depend on the interaction of SR proteins and HNRNPs with the regulatory elements. Insulin signalling leads to phosphorylation of SR and HNRNPs proteins [91, 92]. Insulin stimulates PI3K-dependent phosphorylation of AKT, which in turn phosphorylates members of the SR family directly or indirectly via activation of “serine-arginine protein kinases” (SPRKs) [93]. One of the best characterized AS events regulated by AKT is PKCβ after insulin stimulation. Insulin stimulation induces PI3K-dependent activation of AKT, which phosphorylates SRSF5 [94, 95]. Furthermore, AKT phosphorylates CLK1 and enhances its activity. In turn, CLK1

phosphorylates SRSF4 (previously named SRp75) and SRSF6 (previously named SRp55) [91].

There is some evidence that splicing of exon 4 and 15 in *TCF7L2* is influenced by insulin or glucose [87]. In adipose tissue, the risk T-allele of rs7903146 is associated with lower retention of exon 4 [87]. After significant weight loss, the proportion of exon 15-containing transcripts was reduced, and the proportion exon 4-containing transcripts increased in both liver and adipose tissues [88]. In adipose tissue, transcripts without exon 15 are associated with hyperglycaemia and impaired insulin action, increased plasma free fatty acids [88].

Since in risk CT/TT- but not in CC-genotype carrier, the percentage of exon 4 in pancreatic islets correlates with $HbA_{1c}$, we have focused on exon 4 at the 5'-end of the *TCF7L2* gene [19]. Moreover, rs7903146 resides in intron 3 of *TCF7L2*. Weise *et al*. reported that the exon 4 containing transcripts have an inhibitory effect on WNT target gene expression [96]. Exon 15 encodes a CRARF domain, which is a potent activator of the WNT signalling cascade [96]. *TCF7L2* transcripts containing exon 12-17 boundary are associated with decreased activation of WNT target genes expression. Overexpression of *TCF7L2* transcripts containing 12-15-17 boundary had a protective effect on β cell function and survival, whereas overexpression of transcript with 12-17 boundary resulted in induced impaired insulin secretion and increased apoptosis in human islets [83].

To summarize, exon 4 and 15 of *TCF7L2* are the two predominant alternative exons in pancreatic islets [19]. The mechanism behind the correlation between exon 4 usages and $HbA_{1c}$ needs to be clarified. The AS of both exons was examined in study IV of this thesis.

# Aim of the thesis

The main aim of my thesis is to understand how genetic variants in *TCF7L2* confer an increased risk of developing T2DM. My PhD thesis was divided into three parts in order to approach the problem from different angles.

## Study I and II

The first two studies are interconnected. Since TCF7L2 is a downstream transcription factor of WNT signalling, the aim of study I and II is to understand the normal physiological function of *TCF7L2* in β cells by identifying primary target genes (study I), secondary and tertiary target genes (study II), then reconstruct the molecular network through which *TCF7L2* regulates β cell survival (study I) and GSIS (study II).

## Study III

The chromatin around rs7903146 locus is in an islets-specific "open" conformational state, and reporter assays indicate that the risk-T-allele has a higher enhancing activity compared to the C-allele [72]. The aim of study III is to identify protein(s) binding to this locus and examine their effect on transcriptional regulation by TCF7L2.

## Study IV

Alternative splicing induced by the causative SNP rs7903146 has been hypothesized to produce TCF7L2 protein isoforms with inhibitory action on insulin secretion. Previously, four predominant forms of *TCF7L2* mRNA were described in pancreatic islets [19]. The aim of study III is to elucidate the function of exon 4 and 15 of the *TCF7L2* gene in pancreatic β cells. For this purpose, Antisense Oligo Nucleotides (AON) were used to selectively modify the alternative splicing of *TCF7L2* without altering the total expression of the gene. We will continue to examine the effect of skipping exon 4 or 15 on gene expression, insulin secretion, β cells apoptosis and incretin effects. In addition, potential regulators of exon 4 and 15 alternative splicing will be identified.

**To achieve these aims, different methods were used during my PhD.**

# Methods

During my thesis, I was highly encouraged by Ola and Leif to try out new experimental methods. As the aim of my thesis was to understand the function of one single gene and the functional consequences of its variants, the diversity of the methods became crucial in the attempt to explore novel molecular mechanisms explaining the diabetogenic effect of variants in the *TCF7L2* gene. This chapter will discuss the techniques used for each scientific papers included in this thesis plus some of the trial ideas, which are not yet included in the manuscripts. ***Comments are added for the tricks I have learned by trial and error. I hope some of these thoughts can be helpful to fellow PhD students who try to figure out how to make things work in the wet-lab, and find fun in doing so.**

## Most used methods during my thesis

### Nucleic acid transfection in cell lines, primary human and rodent islets

To deliberately introduce nucleic acids into eukaryotic cells using non-viral methods is defined as transfection [97]. The most common transfection agent I have used is Lipofectamine®, a lipid based cationic reagent, which can form liposomes in an aqueous environment. Liposomes capture negatively charged nucleic acid and fuse with the cellular lipid bilayer aided by a neutral co-lipid. Negatively charged nucleic acid trapped inside the liposomes will attract cationic charge to the external surface of the liposome therefore promote contact of liposome with negatively charged cell membrane. This is followed by a fusion between liposomes and cells due to the similarity of lipid bilayer structures.

Lipofectamine® was used to transfect siRNA, plasmids and Antisense Oligo Nucleotides (AON for exon skipping).

**siRNA-mediated gene knock down**

In order to understand the function of a gene, a useful way is to silence its expression, then compare the functional differences between the gene-depleted sample and control samples. Double-stranded siRNA, often 21 nucleotides, activates the RISC complex after being processed by DICER [98], hence interferes with the expression of its complementary gene on post-transcriptional level. Compared to knock out mice, siRNA is much cheaper and less time consuming. But cell line knock down does not reflect all the physiological consequences that gene silencing in a living organism entails.

***Double knock down** was performed simply by mixing the two siRNA against two different targets [66]. The knock down efficiency of the two knock down is independent of each other. The silencing efficiency is not proportional to the siRNA concentration used.

SiRNA-mediated gene silencing in human islets was performed with the aid of 'extra cellular matrix' (Paper II), a gift from Pr. Philippe Halban [99], and prepared by LUDC human tissue lab. Human islet knock down is a very unstable experiment due to the condition of human islets and the fact that the turnover of islets cells is very low, if any. SiRNA-mediated gene silencing requires an active proliferation so that the RNA molecules can be included into the nucleus after the reappearance of the nuclear membrane during mitosis. ***The trick:** One way to sustain the isolated islets is to perform the siRNA knock down using medium containing human serum and extra cellular matrix (available in Human tissue lab). After 24 hours incubation with siRNA, islet cells that are not attached to the plate are usually irresponsive (dead), and only cells that are attached to the plate should be harvested for further measurements.

**Real time-PCR**

(1) Relative quantification by TaqMan qPCR is the most used technique during my PhD. It is a PCR based measurement using a fluorescent probe specific to the coding part of mRNA, often exon boundary. The fluorescence intensity is proportional to the amount of mRNA target present in the sample. ***With the new reagent** from Life Technologies, using hot-start DNA polymerase, the reaction time has been shortened to 45 minutes instead of 1 hour 40 minutes. The triplicates are slightly less accurate (~5-10%) compared to the traditional reagent. Therefore the Advanced Fast Master Mix should not be recommended for absolute quantification.

(2) Absolute quantification is one important measurement I have learned at the beginning of my PhD, thanks to our lab manager Mrs. Mona Svärdh. Expressions of the same gene from two different tissues are not directly comparable by qPCR due to the potential differences in expression level of the house-keeping gene used. Gene expression measured using different TaqMan® assays are not comparable, because of differences in efficiency of the PCR primers and probes. The best way of comparing the mRNA level between genes and tissues is to perform absolute quantification real-time PCR [19]. A standard curve of a probe with known concentration will be measured at the same time as samples of interest The house-keeping genes are not necessary in absolute quantification of qPCR. ***The challenge** is not to contaminate the reagents/samples with the standard curve DNA oligos. The standard curve has a wide range. Even a slight contamination will ruin the experiment. ***The trick to detour this problem**, if one uses inexpensive and easily available cell line materials, is to increase the starting sample amount by 10 to 20 times without changing the range of the standard curve, *i.e* usually 10 ng of total cDNA is used in each reaction, now 100-200 ng of cDNA can be used in each reaction. In this case, the contamination becomes negligible.

**Western Blot**

Western Blot is likely the most used analytic technique for detecting

specific protein [100-102]. During my PhD study, Ms. Christine Berggreen and Docent Olga Göransson taught me how to perform Western Blot analysis correctly. It is a semi-quantitative quantitative. *The most crucial part of this method is the quality of the antibodies used. All other steps can be performed uniformly, reproducibly and relatively fast with commercially available materials. Technical variation in the Western Blot is more common than for nucleotide quantification. The linear range of the Western Blot is narrow, therefore overloading should be avoided.

**Transcription factor activity measurement**

Luciferase-based assays were used to measure the transcription activity of several proteins including TCF7L2, p53, HMGB1, NFκB and CREB. Known binding sites of target transcription factors are introduced into the plasmid containing the luciferase (*LUC*) gene. The protein binding sites are placed in the promoter region of the luciferase gene. The expression of Luciferase is proportional to the transcriptional activity of the protein measured.

**Chromatin ImmunoPrecipitation (ChIP)**

Chromatin ImmunoPrecipitation was used to determine the chromatin occupancy of DNA binding proteins. DNA cross-linked to nuclear proteins was co-precipitated using an antibody specific to the nuclear proteins. The DNA was then extracted using traditional methods, *i.e.* chloroform:phenol extraction followed by ethanol precipitation in high salt concentrations. The most important variation in this method depends on the quality of the antibody used.

The precipitation protocol I have used is very inefficient when it comes to small amount of starting material, for example human pancreatic islets. This is probably due to the large starting volume required (It's like using a bread knife to crack a walnut, wrong tool!). The AmpureBead® based protocols are much more suitable with limited starting material.

***A linear amplification protocol** When DNA precipitated is too

scarce for the sample preparation before sequencing, a linear amplification can be performed to obtain an adequate amount of starting material. A protocol based on RNA transcription instead of PCR, named LinDA [103, 104] is available. It is reported to be an efficient amplification method.

***Allelic imbalance ChIP**: In study III, we intended to measure differential binding of HMGB1 to the C- and T-alleles of rs7903146. The idea was to perform the ChIP using the islets from a heterozygous donor, CT-genotype carrier of rs7903146. If there is allele-preference of HMGB1 binding for rs7903146, then the precipitated proportion of C/T will be different from the starting material. The initial idea was to perform ChIP-sequencing in order to quantify the potential allelic difference in protein binding ability. But the sample preparation and data analysis took much longer time compared to a TaqMan-based method. A quantitative qPCR measurement using TaqMan genotyping assay should be able to measure the difference. A portion of the starting material is indispensible since the FAM and VIC fluorescence have very different efficiencies. Compared to the starting material, the difference in ChIP enrichment should be possible to calculate. In the study II, all three donors were CC genotype carriers. Islets from heterozygous donors are needed to optimize the method. The crucial point is that when only one or several loci are investigated, no-genome wide ChIP-sequencing is needed. The results can be obtained within 1 hour compared to 1 months using Next Generation Sequencing. Alternatively, Sanger sequencing can also be used for the same purpose.

**Enzyme-linked immunosorbent assay (ELISA)**

The antibody based quantitative method was used for apoptosis assessment (detecting caspase 3/7), proinsulin and insulin quantification [105].

**Electrophoretic Mobility Shift Assay (EMSA)**

EMSA is a common affinity electrophoresis method used to confirm

the DNA-protein binding or RNA-protein binding [107]. When retardation shift is observed during the electrophoresis, a binding between the nucleotide probes and nuclear proteins can be identified. In order to confirm the identity of the binding nuclear proteins, antibody specific to nuclear proteins under question is used. If a super retardation shift is observed, the identity of the binding protein can be confirmed.

Apart from these methods that are common to several studies, the following methods were used in individual studies.

## Methods used specifically in different studies

## Study I

### Chromatin ImmunoPrecipitation on Tiling array (ChIP-chip) and data analysis

After we performed ChIP using antibody against TCF7L2 in both high- and low-glucose conditions, Roche NimbleGen tiling array was used to identify all promoters bound by TCF7L2. The raw data from tiling array hybridization was first analysed using the Hidden-Markov method [108]. Later it was reanalysed using the NimbleGen Tiling Array package NTAP [109]. A binding peak was called when five adjacent probes were all significantly enriched in all three experiments performed.

Other methods used in study I were Luciferase reporter assays for TCF7L2 and p53 activity measurements, ELISA for cell apoptosis measurements and EMSA for the confirmation of TCF7L2 binding to the *Tp53* promoter.

# Study II

### RNA-sequencing and Data Analysis

During my thesis, Dr. Johan Rung and Dr. Jing Su of European Bioinformatics Institute, tutored me during 3 months in RNA-seq data processing and ChIP-seq data processing. Ms. Emilia Ottosson-Laakso tutored me during RNA-sequencing sample preparation.

RNA-seq can also be called "Whole Transcriptome Shotgun Sequencing". The recent developments of Next-Generation Sequencing (NGS) allow for increased base coverage of a cDNA sequence. In study II, transcriptome sequencing was performed on the Illumina HiSeq 2000 platform. Message RNA was prepared from INS1 832/13 cells. Libraries were constructed using Illumina TrueSeq™ RNA sample preparation kit for paired-end sequencing. Amplification was performed for 101 cycles. Six libraries were multiplexed on each lane of the flowcell. The average number of paired-end reads for each sample obtained was 72 million aligned reads. The RNA reads were trimmed from 101 bp to 85 bp and aligned using Top Hat 2.0.0 [110], Bowtie version: 2.0.0.5 with the parameters "-r 250 --mate-std-dev 150" to the following genome assembly: rat RGSC3.4. Aligned reads were counted using HTSeq with the parameters "-m intersection-nonempty -s no". 1.5 FPKM per gene was considered the threshold of the expressed gene, as it represents around 30 count per gene. The differential expression was computed using EdgeR [111] and CuffDiff [112]. The false discovery rate (FDR) was used to correct for multiple comparisons. Pathway analysis was performed using WEB-based Gene Set Analysis (WebGestalt) and Ingenuity IPA.

### Total internal reflection fluorescence microscopy (TIRFM)

The TIRFM analysis was performed by Dr. Enming Zhang to compare insulin granule number before and after *Tcf7l2* silencing in INS1 832/13 cells. Insulin was stained in INS1 832/13 cells. A total internal reflection fluorescence microscope was used to observe a thin region of a

specimen; usually around 100 nm can be observed [113]. A TIRFM use an evanescent wave to selectively illuminate fluorophores in the samples immediately adjacent to the glass-water interface. The evanescent wave is only generated when internally reflected at the glass-water interface. The electromagnetic field of the incident light decay exponentially from the interface, therefore the light can only penetrate to a depth of about 100 nm into the sample solution. It's noteworthy that the area visualized is a few nanometres wide. The principle of TIRFM is illustrated in Figure 3.

Other methods used in study II were: ELISA and WB for measurements of insulin content; ELISA and RIA for measuring glucose stimulated insulin secretion; ELISA for measuring proinsulin secretion.

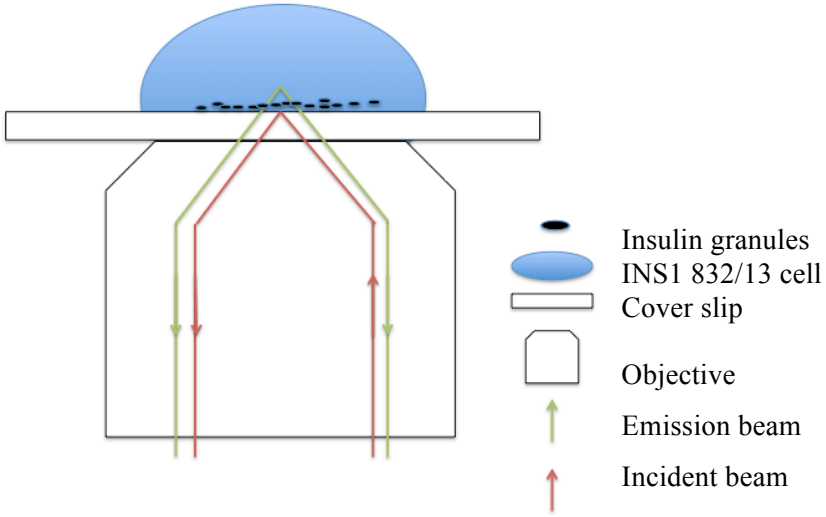


**Figure 3. Insulin granule counting using total internal reflection fluorescence microscopy (TIRFM)**

# Study III

### Affinity column protein purification

In this study, we have employed a 'home-made' affinity column to

purify and enrich soluble nuclear protein(s) that bind(s) the DNA. The binding buffer used was the same as for EMSA experiments. The principle of the purification is illustrated in Figure 4. A single strand DNA probe labelled at 3'-end with biotin was first annealed with a reverse complement probe, in order to obtain a double-strand biotin labelled DNA probe. The probe length varied from 20 nucleotides to 60 nucleotides. The DNA probe was then bound to monomeric avidine-coated agarose beads. Nuclear extract was incubated with DNA probes fixed on the agarose beads for around one hour. The DNA-protein complex was rinsed with binding buffer containing up to 150 mM NaCl. Depending on the binding affinity, unspecific binding could be eliminated by stringent wash. The DNA probes (together with the binding nuclear proteins) were finally eluted using 0.2 M biotin due to the reversible property of the agarose beads used. The collected elute should be submitted to a Mass Spectrometry or Edman sequencing in order to identify the enriched protein.

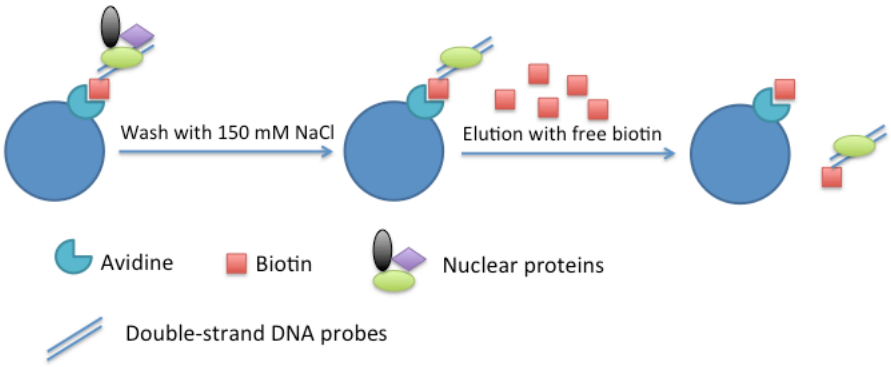


**Figure 4. DNA binding nuclear protein purification using affinity column**

## Edman Sequencing

Edman sequencing is a protein identification method using Edman degradation (Phenylthiohydantion). A minimum of 1 pmol of protein is needed in order to obtain a reliable identification. Edman degradation was

developed and improved by Pehr Victor Edman (1916-1977), a Swedish biochemist and physician. In fact, he was a docent at Lund University from 1947 to 1957, and part of the work of improving the method was done at Lund University.

**ChIP-sequencing**

ChIP-sequencing was used to replace ChIP-chip with Next Generation Sequencing. The library preparation was performed using Illumina ChIP-seq$^{TM}$ DNA sample preparation kit, paired-end. The sequencing was performed on the Illumina HiSeq 2000 platform. Amplification was performed for 101 cycles. Five to six libraries were multiplexed on each lane of the flowcell. The alignment of the reads was performed using BWA [114], SAM tool [115], and the peak calling was performed using MACS [116] and SWEMBL (Article in preparation).

**Dynamic Light Scattering (DLS)**

The DLS measurements were performed by Dr. Nikolay Oskolkov to compare the rigidity between the DNA probes representing C- and T-allele of the rs7903146 locus and the binding affinity of these two DNA probes with HMGB1.

The method was previously unfamiliar to me, but DLS is in fact one of the most popular methods used to characterize particles or molecules dispersed in liquid. The principle of the method is as follows. Shining a monochromatic light beam, such as a laser, onto a solution of particles in Brownian motion (random movement of particles suspended in a fluid), the electric field of the light induces an oscillating polarization of electrons in the molecules. This causes the molecules to provide a secondary source of light and resulting in scattering light. The Brownian motion of particles or molecules in suspension causes laser light to be scattered at different intensities. Analysis of the intensity fluctuation yields the diffusion coefficient of the Brownian motion and hence the particle sphere size distribution using the Stokes-Einstein relationship. This gives a description of the particle's motion in the medium: the diffusion coefficient of the

particle. The detection of DLS is illustrated in Figure 5.

The size, shape and molecular interactions in the scattering material is determined by the frequency shift of the light, angular distribution and the polarization of the light. It is also possible to obtain absolute measurements of several parameters of interest, like molecular weight, radius of gyration, translational diffusion coefficient etc.

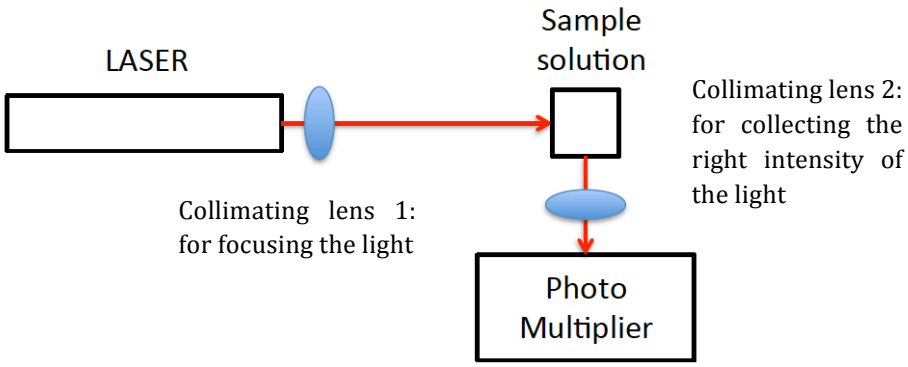


**Figure 5. Scattered light Detection in DLS method** A laser beam is shined on sample solution. Scattered light is collected at 90 degree using a photo multiplier.

## Small-angle X-ray scattering (SAXS)

The SAXS measurements were also performed by Dr. Nikolay Oskolkov in order to compare the rigidity of two 60 nucleotides DNA probes representing the C- and T-alleles of rs903146. In SAXS measurement, X-rays with wavelength of 0.1-0.2 nm are applied to molecular samples (DNA probes in our study).

Small-angle X-ray scattering (SAXS) is a powerful tool to explore the nanostructure of matter. Originally developed in the 1930s to study metal alloys [117], SAXS rapidly became useful for other objects,

including biological macromolecules. It is now a mature technique in the field of structural biology. The method is applicable to a broad range of molecular sizes and requires a small amount of material.

The principle of this method is illustrated in Figure 6. A monochromatic beam hits the samples; the electrons of the sample atoms become sources of secondary waves. The scattered photons are collected on a 2D detector. Since the scattering is isotropic, the recorded image is radially averaged to obtain the scattering curve. The scattered intensity I is a function of wave vector: $q = 4\pi \sin \theta / \lambda$. Here $\lambda$ is incident beam wavelength. The scattering of the solute and solvent are collected separately, the difference is proportional to the solute concentration. The scattered intensity can be plotted as a function of wave vector q (such as Figure 7, adopted from Study III).

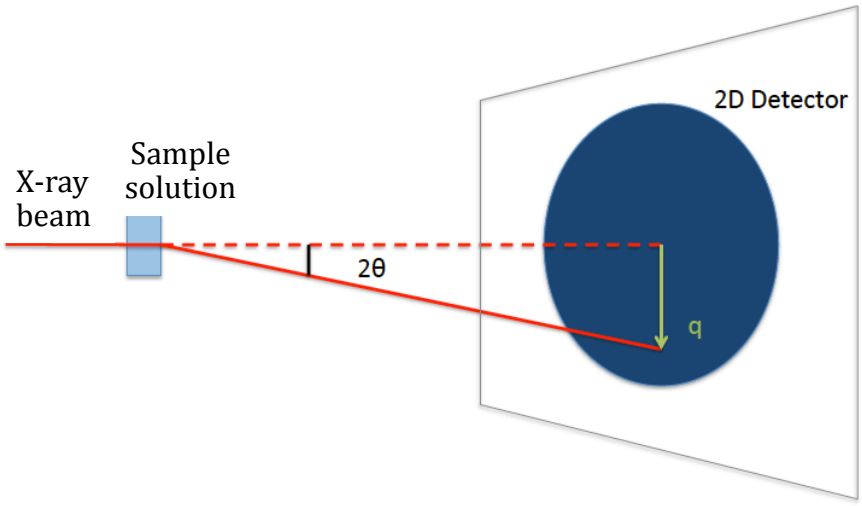


**Figure 6. Principle of the SAXS measurement** An X-ray beam is shined on sample solution. Scattered light intensity is collected on a 2D detector. Scattered intensity is integrated as a function of q.

However, in Study III, the Guinier analysis was used to determine the radius of gyration $R_g$. The definition of this is a length that represents the distance in a rotating system between the point about which it is

rotating and the point to or from which a transfer of energy has the maximum effect. To make it simpler, imagine an object that has condensed all its particles into a hollow sphere. The hollow sphere (with no thickness) has the same rotational inertia as the original object. The radius of the sphere is the radius of gyration of the original object. In polymer physics, the radius of gyration is used to describe the dimension of a polymer chain (such as a DNA molecule). Since the chain conformation and shape of a polymer sample (DNA) are infinite in number and constantly change over time, the radius of gyration is measured in an average over time. Radius gyration measured in the constant moving polymer samples can therefore be considered proportional to the stiffness and of the polymer.

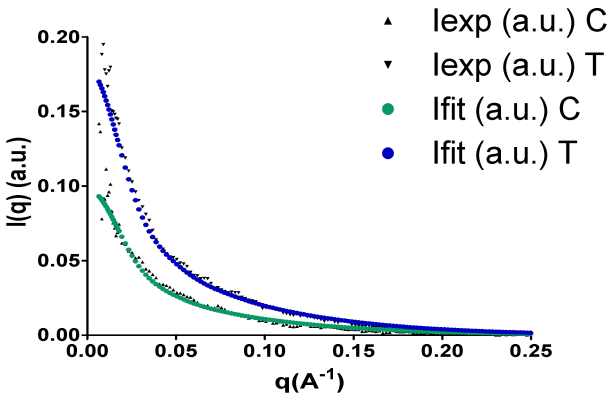


**Figure 7. Background corrected scattering intensity curve as a function of the wave vector for two DNA probes (Paper III, Figure 2B)**

ChIP-qPCR was also used to in study III to confirm the HMGB1 binding to the rs7903146 locus in human pancreatic islets.

# Study IV

### Exon Skipping

Exon skipping was first described and is currently under clinical trial in the treatment of Duchenne muscular dystrophy [118]. Antisense

Oligonucleotides (AONs) were used to induce effective exon skipping during pre-mRNA splicing. AONs have an optimal length of 20 nucleotides. It is usually chemically modified in order to be resistant to endo-, exonucleases and RNase H. Multiple chemical modifications are available, and the most commonly used are the 2'-O-methyl RNA phosphorothioate modification (2OMePS). The chemical modification also improves the affinity for target sequences. The sequences of AONs are designed to target pre-mRNA sequences essential for proper inclusion and exclusion of the exons into the mature mRNA.

The AONs manipulate splicing by steric hindrance of splicing factors to the target pre-mRNA, *i.e.* by occupying the target-binding site of splice factors. Therefore the affinity needs to be sufficient to form the complex and prevent binding. The AONs can target splice sites at the beginning, the end of exons or the exon-internal sequences called exonic splicing enhancers (ESEs). These sequences are bound by different Serine / Arginine rich proteins (SR proteins) that recruit U1 snRNP and U2AF to the 5'-end splice site and polypyrimidine to 3'-end splice site. In addition to the ESEs, there are also exonic splicing silencers (ESSs), which are bound by hnRNPs. HnRNPs are the counterparts of SR proteins that prevent binding of snRNPs, to the splice sites and lead to exon skipping. AONs blocking ESSs would increase exon inclusions [119].

The extensive trials made in the dystrophin gene exon skipping made it possible to compare the AONs targeting splice sites with those targeting exon-internal sequences. The results favoured the exon-internal targeting sequences [120]. The reason might be that higher GC contents inside the exons give higher affinity between AONs and target sites. The retrospective analysis also revealed the enrichment for different SR proteins binding sites in successful AONs [119, 121]. This reflected that the exons use different SR proteins for their inclusion in their mature mRNA.

Technically, using AONs-mediated exon skipping is very similar to siRNA-mediated gene silencing. Lipofectamine® was used for the

delivery of AONs. After 72 hours of incubation, the skipping efficacy can be measured using qPCR. ***The specific conditions** should be strictly optimized for each cell line, as AONs can be toxic for the cells. Therefore, minimum amount of AONs should be used without compromising the skipping efficiency.

**BeadArray® and Data Analysis**

The BeadArray® was one form of microarray provided by Illumina. The hybridization and detection were performed by Dr. Kristoffer Ström. The array was used to measure the genome-wide mRNA expression of HCT116 cells after exon skipping. The differentially expressed genes were identified using "beadarray" R package [122].

# Results and discussion

## Study I. Survival of pancreatic beta cells is partly controlled by a TCF7L2-p53-p53INP1-dependent pathway

TCF7L2 is a transcription factor in the WNT signalling pathway and functions through binding to its target genes. Promoter regions (-1000 to +150 pb from the transcription start site) bound by TCF7L2 were identified using Chromatin-Immunoprecipitation (ChIP-chip) on Roche NimbleGen rat promoter tiling array 2.1M in INS1 832/13 cells. The ChIP-chip experiments were performed at both 5.5 mM and 14.3 mM glucose condition. The chromatin occupancy of TCF7L2 in INS1 832/13 cells was identified using a Hidden Markov Model (HMM) method [108] and 3646 genes were identified as potential target genes of TCF7L2 in either 5.5mM or 14.3mM glucose. The strategy was to select potential TCF7L2 regulated genes for further study by the following: (1) loci previously associated with T2DM (genes located proximal to the associated SNPs) and (2) known genes involved in the pathogenesis of T2DM, such as *PTEN* in the insulin signalling pathway [60, 123].

By 2010, around 30 loci were reported to be associated with T2DM [14]. Among the genes near the associated loci, *Fto* and *Adamts9* were directly bound by TCF7L2. *Tp53inp1* (encoding p53 inducible nuclear protein 1), was selected despite of the fact that it was not directly bound by TCF7L2. *Tp53inp1* may induce cell apoptosis and its main regulator p53, coded by *Tp53*, was bound by TCF7L2 in high glucose condition. The direct binding of TCF7L2 in the promoter of *Tp53* (tumour protein 53) was confirmed using EMSA.

In order to exam the influence of the identified candidate pathway *Tp53-Tp53inp1* in β cells, a series of functional analysis was carried out. These results showed that the TCF7L2-p53-p53INP1 pathway is a key component in the mechanism whereby a dysregulation in *TCF7L2* expression may have a very strong impact in β cells apoptosis. We demonstrated that TCF7L2 represses the expression of p53INP1 and that p53 is a necessary intermediate transcription factor in this inhibition. Silencing *Tcf7l2* in INS1 832/13 cells leads to a slight increase of *Tp53* mRNA expression, but a substantially increased p53 transcriptional activity under both high and low glucose conditions. P53 is a main transcriptional activator of *Tp53inp1*. We observed increased mRNA and protein expression of p53INP1 under both glucose conditions after *Tcf7l2* silencing. In *Tp53*-depleted cells, *Tcf7l2* silencing could not alter *Tp53inp1* expression (Figure 8). This was demonstrated using double knock down of *Tcf7l2* and *Tp53* in INS1 832/13 cells.

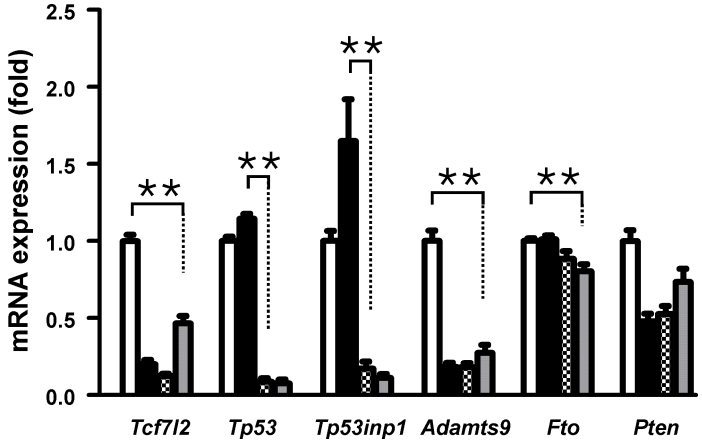


**Figure 8. p53 is necessary for the observed increase in *Tp53inp1* expression induced by *Tcf7l2* silencing in INS-1 cells.** Cells were incubated in 25 mM glucose either with scrambled siRNA (white bars), with siRNA-targeting *Tcf7l2* (black bars), with siRNA targeting *Tcf7l2* and *Tp53* (checked bars) or with siRNA targeting only *Tp53* (gray bars). *Tp53* silencing lead to a down-regulation of *Tcf7l2* by 54%, Adamts9 by 73% and Fto by 20% compared with scrambled control. Simultaneous silencing of *Tcf7l2* and *Tp53* lead to a down-regulation of *Tp53* by 94% and of *Tp53inp1* by 90% compared with *Tcf7l2* knock-down **(Paper I, Figure 5).**

Shu *et al*. reported that *TCF7L2*-depleted human islets exhibited increased cell apoptosis and decreased cell proliferation [60]. The question was if this was due to p53INP1 regulation. It has been previously reported that p53INP1 regulates p53-dependent cell apoptosis [122]. Increased p53INP1 activates Homeodomain-interacting protein kinase 2 (HIPK2) [124], which is a kinase that phosphorylates serine 46 of p53 [125], by binding to the p53 protein. Phosphorylation on serine 46 activates p53 to specifically induce the transcription of *Tp53inp1*. Our hypothesis was that *Tcf7l2* silencing via p53 activation, increased p53INP1 expression, which in turn induced the phosphorylation of p53 by interacting HIPK2, creating an auto-amplification loop to further promote the expression of p53INP1 and induction of apoptosis (Figure 11). This might be the cause of the observed increased cell apoptosis after *Tcf7l2* silencing.

In order to test this hypothesis, a "rescue experiment" was attempted. Double knock down of *Tcf7l2* and *Tp53*, or, *Tcf7l2* and *Tp53inp1* should prevent β cells from TCF7L2-induced apoptosis. In INS1 832/13 cells and isolated rat islets the hypothesis was confirmed (Figure 9). In line with the previous findings, an increased apoptosis was seen after *Tcf7l2* silencing. However, reversal of the increased *Tp53* or *Tp53inp1* mRNA expression by double knock down (*Tcf7l2-Tp53*, or *Tcf7l2-Tp53inp1*) protected from *Tcf7l2*-induced apoptosis in both the clonal cell line and the isolated primary rat islets. The measurements of apoptosis were confirmed using three different methods (Figure 10), including: (1) Immunohistochemistry staining of the apoptotic markers Annexin V and 7-AAD, (2) Caspase-3 degradation using WB and (3) apoptosis assessment measuring DNA degradation.

Study I has confirmed the previously observed increased cell apoptosis after *TCF7L2* silencing [60] and provided a novel molecular pathway of β cells apoptosis. The data also suggested a molecular link between the two T2DM-associated genes: *Tcf7l2* and *Tp53inp1*.

Genes involved in either glucose-stimulated or incretin-potentiated insulin secretion were not identified as direct targets of TCF7L2 in study I. The ChIP-chip alone is not a comprehensive method to identify the entire

down-stream regulatory network of TCF7L2 in the β cells. The reason is because Chromatin ImmunoPrecipitation can only identify the direct occupancy of TCF7L2 in the chromatin. From the TCF7L2-p53-p53INP1 pathway, we realized that the TCF7L2 can influence of its downstream target genes on the mRNA transcription level or on the post-translational level. ChIP experiments alone cannot reflect all the regulation caused by TCF7L2. It turns out that study I was just the prologue of study II.

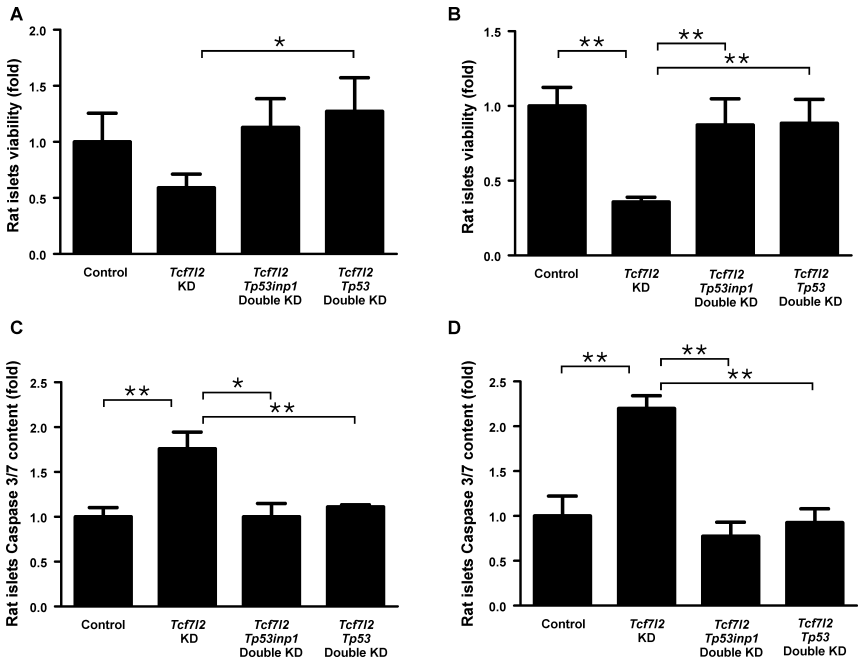


**Figure 9. Viability and apoptosis measurements after siRNA treatment in rat primary islets** After 24 hours of incubation with siRNA in either 5.5 mM glucose (A and C) or 14.3 mM glucose (B and D). *Tcf7l2* KD resulted in 41% decrease in viability in low glucose (A) and 65% in high glucose (B), and islets with a non-induced *Tp53inp1* or *Tp53* level had similar viability as control. Apoptosis was measured quantifying active caspase-3/7 in rat islets. *Tcf7l2* KD resulted in a 75% increase in apoptosis in low glucose (C), a 120% increase in apoptosis in high glucose (D). Islets with a non-induced *Tp53inp1* or *Tp53* level had similar apoptosis level as control. N = 5 **(Paper I, Figure 7).**

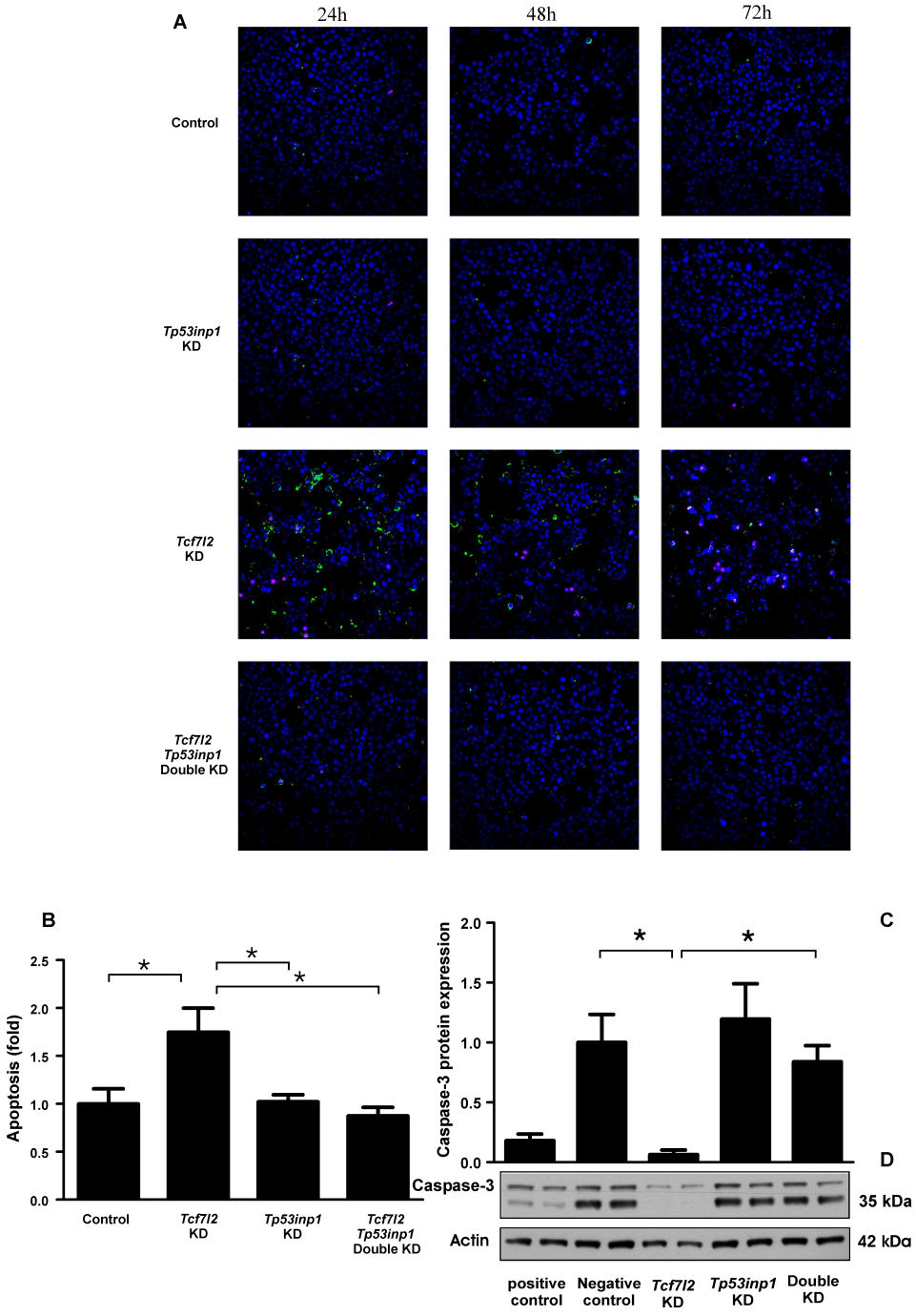
A
24h
48h
72h
Control
Tp53inp1 KD
Tcf7l2 KD
Tcf7l2 Tp53inp1 Double KD
B
Apoptosis (fold)
2.5
2.0
1.5
1.0
0.5
0.0
Control
Tcf7l2 KD
Tp53inp1 KD
Tcf7l2 Tp53inp1 Double KD
C
Caspase-3 protein expression
2.0
1.5
1.0
0.5
0.0
D
Caspase-3
35 kDa
Actin
42 kDa
positive control
Negative control
Tcf7l2 KD
Tp53inp1 KD
Double KD

**Figure 10. Apoptosis measurements after siRNA treatment in INS-1 cells** (A) Apoptosis measured using IHC. Cells were triple stained using: Hoechst (DNA, blue), Annexin 5 (early apoptosis marker, green) and 7AAD (late apoptosis marker, red). Images from three time points are shown: 24, 48 and 72 h after siRNA incubation. (B) Apoptosis measured using a DNA degradation-based kit after 72 h of incubation with siRNA. *Tcf7l2* silencing resulted in a 74% increased cell death and cells with a non-induced *Tp53inp1* level, i.e. (*Tp53inp1* KD) and (*Tcf7l2 Tp53inp1* double KD) had similar cell death levels as control (C) Apoptosis measured using Western Blot analysis of caspase-3 heavy chain degradation after 72 h of incubation with siRNA. A total of 15 ug of protein lysate was run in duplicate with β-actin used as loading control. Bars represent the quantification of three independent experiments, shown together with a representative blot. Cells treated with adriamycin for 4 h (positive control) or with scrambled siRNA (negative control) were used as controls. *Tcf7l2* silencing resulted in a 94% decreased caspase-3 heavy chain expression and cells with a non-induced *Tp53inp1* level, i.e. (*Tp53inp1* KD) and (*Tcf7l2 Tp53inp1* double KD) had similar cell death level as the negative control. Cells were incubated in 11 mM glucose **(Paper I, Figure 6).**

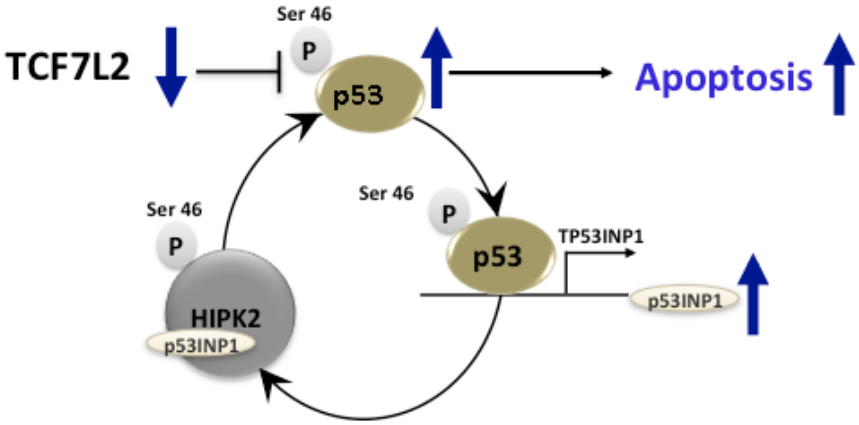


**Figure 11. Mechanistic illustration of TCF7L2-p53-p53INP1 pathway in cell apoptosis**

## Study II. *TCF7L2* is a master regulator of insulin synthesis and processing

In study II, we silenced *Tcf7l2* in INS1 832/13 cells, and identified differentially expressed genes using RNA-sequencing (3565 genes, after 5% FDR correction). The differentially expressed genes are a mixture of TCF7L2 primary, secondary and down-stream target genes. To delineate the regulatory network, several approaches of gene selection were adopted: (1) confirming the primary target genes of TCF7L2 identified in study I; (2) Pathway enrichment analyses (Ingenuity) on genes differentially expressed after *Tcf7l2* silencing and (3) exploring GWAS genes/loci associated with T2DM. Furthermore, we also analysed genes influencing circulating proinsulin and proinsulin-to-insulin ratio.

Firstly, primary target genes were defined as genes whose promoters were occupied by TCF7L2 (from ChIP-chip in study I), and whose mRNA expression was significantly affected by *Tcf7l2* silencing. 511 unique genes were thereby identified as primary target genes. Several genes with known function in T2DM belong to this category, such as *Gcg*, *Gipr*, *Iapp*, *Ins1*, *Isl1, Jazf1*, *Myc*, *NeuroD4*, *Nfya*, *Npy1r*, *Pak2*, *Sirt2* and *Sirt3*. Secondly, the pathway analyses identified several pathways that are potentially important in the pathogenesis of T2DM, including: **(1)** Voltage-dependent L-type calcium channels, (mainly alpha-subunit, together with *Cacna2d1* in the delta-subunit); **(2)** the GABA shunt pathway (including ABAT and GAD2) and **(3)** the insulin signalling pathway, together with *Ins1* and *Ins2*. Lastly, we also followed up published GWAS results. More than 65 loci have been associated with T2DM or circulating glucose/insulin [6-14] and around 12 loci with circulating proinsulin and proinsulin-to-insulin ratio [12, 51-58]. Of these loci, 26 genes were differentially expressed after *Tcf7l2* silencing (Figure 12 and in Paper II, Supplemental Table 2). Among these, ZnT8, coded by *SLC30A8*, is important for the formation of insulin crystals [126]. A common non-synonymous SNP in *SLC30A8* has been associated with both T2DM [9] and reduced proinsulin-to-insulin conversion [52]. *PCSK1* and *PCSK2*, code for convertase 1/3 and convertase 2 respectively, which participate in

the processing of proinsulin to form mature insulin and C-peptide. The expression of *Pcsk1*, *Pcsk2* and *Slc30a8* were down-regulated after *Tcf7l2* silencing. Therefore, **(4)** the proinsulin-to-insulin conversion and post-translational modifications turned out to be another important β cell function that *TCF7L2* might be involved in.

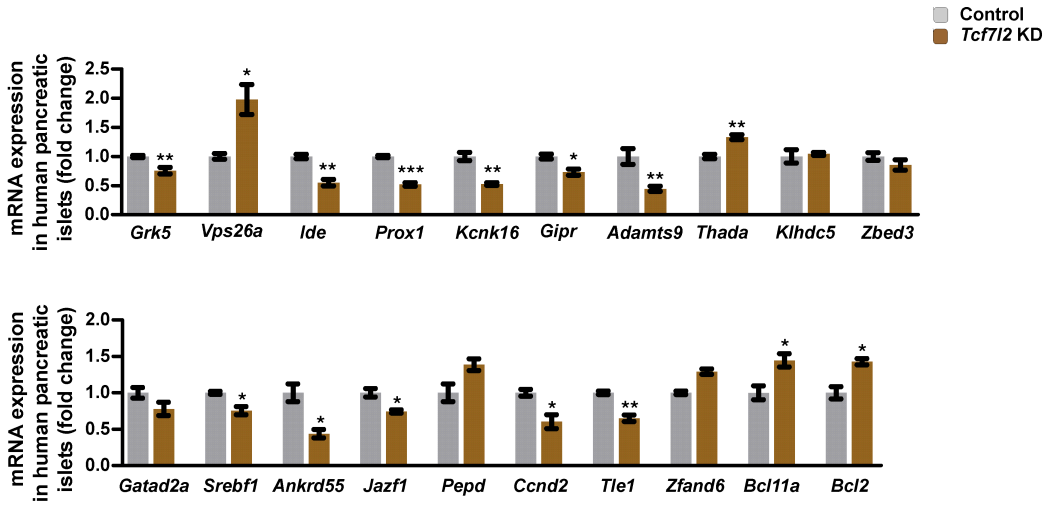


**Figure 12. *Tcf7l2* silencing affected other T2DM GWAS genes.** *Tcf7l2* silencing led to decreased mRNA expression of *Grk5*, *Ide*, *Prox1*, *Kcnk16*, *Gipr*, *Adamts9*, *Srebf1*, *Ankrd55*, *Jazf1*, *Ccnd2*, *Tle1* and increased mRNA expression of *Vps26a*, *Thada*, *Bcl2* and *Bcl11a* **(Paper II, Figure 5A).**

Around a dozen transcription factors regulate the expression of proinsulin in human β cells, including PAR1, PDX1, OCT1, ISF, PAX4, MAFA, USF1, USF2, CREB, CREM, E2A, NEUROD1 and HEB [127, 128]. After independent replications of the both mRNA and protein levels, PDX1, NEUROD1 and MAFA were confirmed as differentially regulated after *Tcf7l2* silencing. These three transcription factors are also among the most crucial factors for maintaining β cell identity and function [128]. In addition, it was previously reported that ISL1 binds and regulates the expression of *MAFA* [147]. The two identified pathways (TCF7L2-ISL1-MAFA-proinsulin and TCF7L2-NEUROD1-proinsulin) provided means by which TCF7L2 might influence proinsulin expression. *MAFA*

expression is also tightly regulated by about 10 transcription factors [129]. Among these, the mRNA expression of *Isl1, Prox1*, *Nkx6.1* and *Nkx2.2* were significantly influenced by *Tcf7l2* silencing. The differential expression of ISL1 and NKX6.1 was replicated on the protein level.

A series of functional analyses were carried out in order to delineate the potential regulatory cascade of transcription factors in INS1 832/13 cells. *Tc7fl2*, *Isl1*, *MafA*, *NeuroD1* and *Nkx6.1* were sequentially knocked down, then the mRNA and later protein expression of all target genes, including *Ins1*, *Ins2*, *Pcsk1*, *Pcsk2* and *Slc30a8* were measured. These results indicate that ISL1 is an upstream regulator of several of these genes. The influence of ISL1 on *Pdx1*, *Nkx6.1*, *Pcsk1* and *Pcsk2* expression has not been described previously. Direct binding of TCF7L2 to the *Isl1* promoter was also demonstrated in study I. The key finding of this study is that together with ISL1, TCF7L2 regulates (1) the proinsulin expression through PDX1, MAFA, NEUROD1 and NKX6.1; (2) proinsulin-to-insulin conversion through convertase1/3 (*PCSK1*), convertase2 (*PCSK2*) and ZnT8 (*SLC30A8*) and (3) many genes/loci associated with T2DM in published GWAS. In other words, a link between proinsulin synthesis and processing was established, as TCF7L2 and ISL1 seems to form a master-regulatory circuit influencing both processes.

In *Tcf7l2*$^{-/-}$ allele null mice [70], the expression of ISL1, MAFA, NEUROD1, ZnT8 and proinsulin expression were significantly lower compared to wild-type littermates. In addition, siRNA-mediated *TCF7L2* silencing in isolated human islets further confirmed the down-regulation of the above target genes.

The consequence of a dysregulation of this network is impaired insulin secretion. Knocking down *Tc7fl2*, *Isl1*, *MafA* and *NeuroD1* results in perturb GSIS (Figure 13). In addition, in the *Tc7fl2*-silenced cells, there is an increased proinsulin-to-insulin ratio in the secreted pool (Figure 14), indicating impaired insulin maturation. The molecular mechanisms outlined in study II are summarized in Figure 15.

Several important genotype effects of rs7903146 were observed in this study. Firstly, glucose stimulated insulin secretion was significantly

impaired in risk CT/TT-genotype carriers compared to non-risk CC-genotype carriers (Figure 16), which is consistent with previous findings [22, 50, 51, 130]. Secondly, total insulin content was significant lower in CT/TT genotype carriers compared to the CC-genotype carriers. A previous study did not detect the genotype effect on insulin content in the whole islets [23]. This might be due to method used for measuring insulin content. Insulin content in our study was measured without prior freezing, while the previous study used conserved human islets [23]. Thirdly, we confirmed in a large number of human islets that the T-allele is associated with higher expression of *TCF7L2* mRNA (Figure 16). Lastly, mRNA expression of the identified target genes (*INS*, *MAFA, ISL1*, *NKX6.1*) correlated with *TCF7L2* mRNA expression in CC-genotype carriers but not in risk CT/TT-genotype carriers; one example is the proinsulin gene (Figure 16). In addition, expression of the other GWAS genes that were regulated by TCF7L2 in human pancreatic islets, correlated with *TCF7L2* mRNA expression in CC- but not in CT/TT-genotype carriers. This suggests that altered TCF7L2 activity characteristic of risk T-allele will affect expression of genes in the insulin synthesis pathway as well as genes associated with T2DM in GWAS. However, the molecular mechanisms of how the genotype may affect TCF7L2 activity were not examined in study II.

To summarize, Study II highlights the central role of TCF7L2 in pancreatic β cells. We have demonstrated that ISL1 is a direct target of TCF7L2 and that ISL1, in turn, regulates proinsulin synthesis and processing via regulation of MAFA, PDX1, NKX6.1, PCSK1, PCSK2 and ZnT8. With these data, we provided an explanation for the large impact of *TCF7L2* on the risk of developing T2DM.

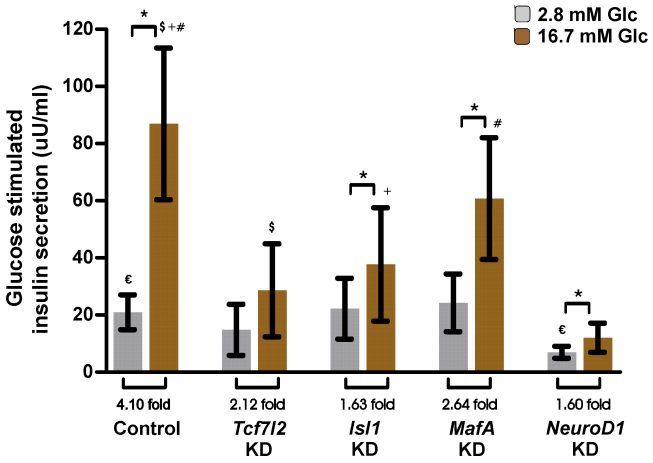


**Figure 13. Glucose stimulated insulin secretion in INS1 832/13 cells after *Tcf7l2*, *Isl1*, *MafA* and *NeuroD1* silencing (Paper II, Figure 3D).**

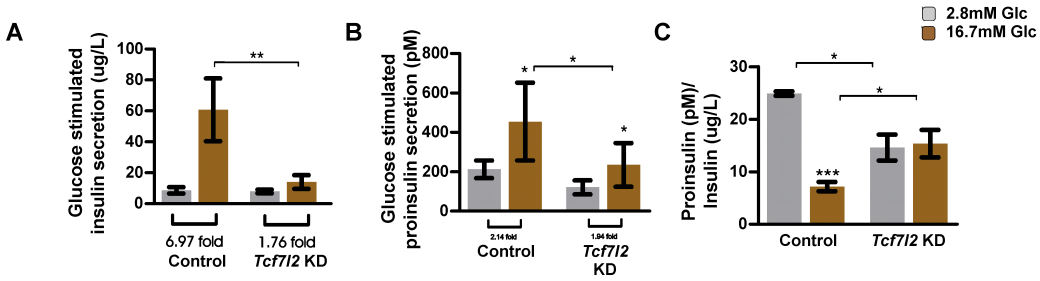


**Figure 14. Glucose stimulated insulin secretion, proinsulin secretion and proinsulin-to-insulin ratio in INS1 832/13 cells after *Tcf7l2* silencing (Paper II, Figure 1E-G)**

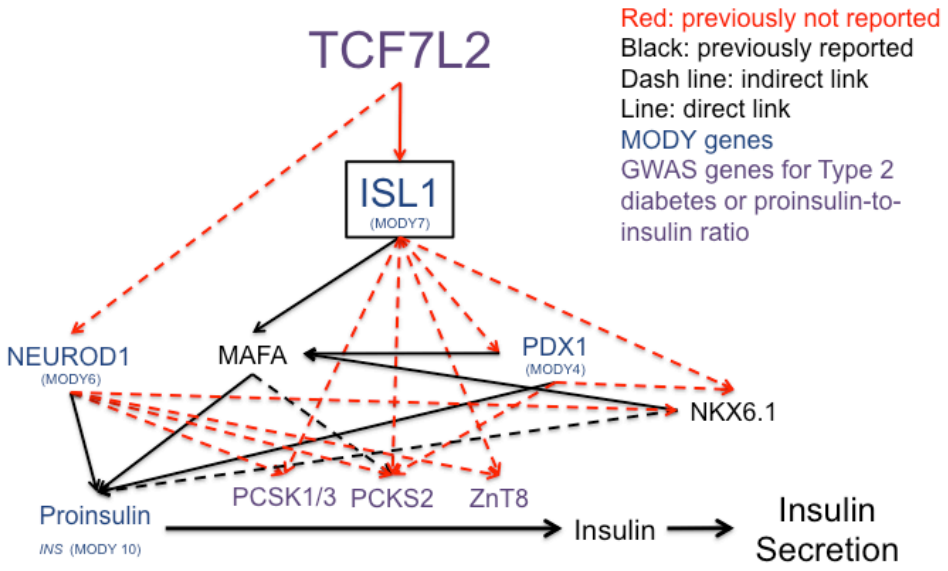


**Figure 15. Molecular mechanisms through which *TCF7L2* regulates proinsulin expression, proinsulin-to-insulin conversion and insulin maturation**

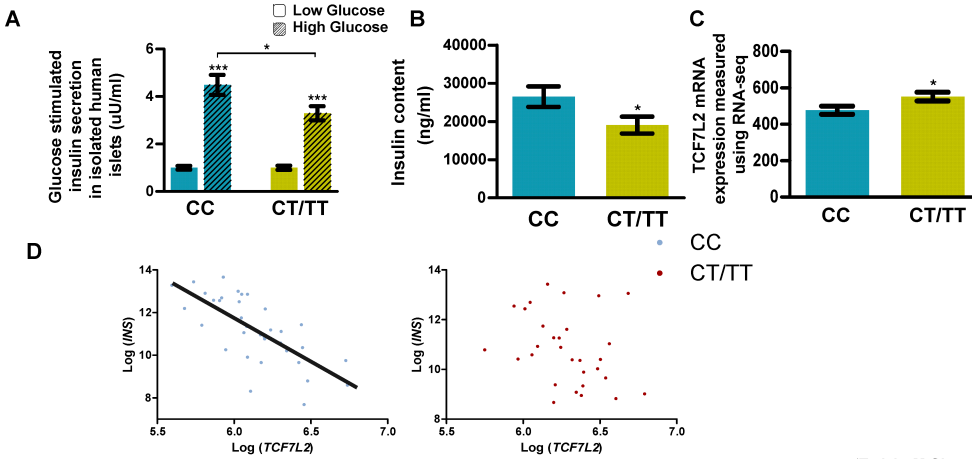


**Figure 16. Rs7903146 genotype and TCF7L2 influence on insulin content and glucose-stimulated insulin secretion in human pancreatic islets** (A) Glucose-stimulated insulin secretion in isolated human islets from 44 CC allele carriers of rs7903146 compared to 31 CT/TT allele carriers. (B) Insulin content measured from 45 CC allele carriers compared to 36 CT/TT carriers with purity superior than 70%, after correction for age, gender and BMI. (C) Regression analysis between *TCF7L2* and

rs7903146 genotype in pancreatic islets from 36 CC allele carriers and 30 CT/TT carriers after correction for age, gender and BMI. (D) Linear regression between *INS* and *TCF7L2* mRNA expressions separated by rs7903146 genotype from 36 CC allele carriers and 30 CT/TT carriers after correction for age, gender and BMI **(Paper II, Figure 1A-D).**

# Study III. The SNP rs7903146 is bound by HMGB1 protein in the open chromatin region of the *TCF7L2* gene

In study II, we demonstrated an influence of the risk-genotype on TCF7L2-target gene expression, but we did not provide a molecular explanation for this effect in human islets. The risk T-allele of rs7903146 locus is associated with a higher degree of open chromatin structure compared to the C-allele. This effect was restricted to islets [72, 82]. Reporter assays also showed that the T-allele displays markedly higher enhancer activity than the C-allele [72, 82]. In the two studies from Gaulton and Stitzel [72, 82], no increase in *TCF7L2* mRNA expression in human islets was observed, probably due to the lack of power. The lacking piece of information from these studies was: which protein(s) would bind to the open chromatin? It can be anticipated that the protein-DNA interactions should be influenced by the chromatin conformational state of the region [131]. The aim of study III is to isolate nuclear protein(s) that potentially bind to the open chromatin associated with the risk T-allele.

To address this question, we used an affinity column with 61 base pair long oligonucleotides to purify potential DNA-binding protein(s), followed by Edman sequencing for protein identification. Using this method, we enriched and isolated a nuclear protein, HMGB1 (high-mobility group protein B1) that binds to the rs7903146 locus. Small-angle X-ray scattering (SAXS) and Dynamic Light Scattering (DLS) measurements were applied to compare the affinity between HMGB1 binding to C- or T-oligonucleotides. HMGB1 is a small protein of 27kDa and consists of two consecutive L-shaped basic domains (HMG boxes), a short linker and a 30 amino acid-long acidic ‘tail’ [132]. In the nucleus, HMGB1 binds to the minor groove of DNA with little sequence specificity. The binding of HMGB1 distorts the double helix and induces a bending of 90° or more [133]. HMGB1 usually enhances the binding of other nuclear proteins [134], and targeting of HMGB1 to a specific sequence is achieved

by the interaction with other nuclear proteins. It is also well known that HMGB1 interacts with and enhances the activities of a number of transcription factors, such as p53, p73, RB and NF-κB family [135-138]. In the case of the rs7903146, HMGB1 is probably only part of the machinery and serves as the protein anchor for a larger regulatory complex, as HMGB1 distorts the DNA and facilitates the binding of other nucleoproteins. Further studies are needed to identify the entity of this protein machinery. In our experiments, we applied stringent washing step (using 150 mM of NaCl) to eliminate unspecifically bound proteins. Affinity column might not be the optimal way of purifying large protein complexes since the binding strength among the protein partners might differ largely depending on interactive properties.

The abundant expression and wide range of HMGB1 protein functions also raised the concern of binding specificity. Histone H1 and HMGB1 are the most abundant chromosomal proteins apart from the core histones. *In vivo* and *in vitro* data showed that both are able to organize the chromatin structure, *i.e.* binding of H1 will result in a more stable structure and HMGB1 binding in a less stable structure of the chromatin [139]. The binding sites for H1 and HMGB1 are often partially overlapping and replacement of H1 by HMGB1 could result in a fast switching of chromatin states.

*In vitro* confirmation of HMGB1 binding in this study was performed using human pancreatic islet nuclear proteins by three different methods; (1) Affinity column pull-down and detection with an antibody against HMGB1 (Figure 17), (2) ChIP-qPCR to and (3) super-shift EMSA (Figure 18). All three methods confirmed a specific binding of HMGB1 at the rs7903146 locus. Next, we compared the binding affinity of HMGB1 to the C- and T-allele. A potential allele binding difference was already observed during the confirmation step, where HMGB1 was more highly enriched in the C-allele than the T-allele purification in all three human islets donors (Figure 18). A series of physical measurements, including DLS and SAXS measurements, were carried out to compare the properties of the two sequences and the difference in HMGB1 binding strength.

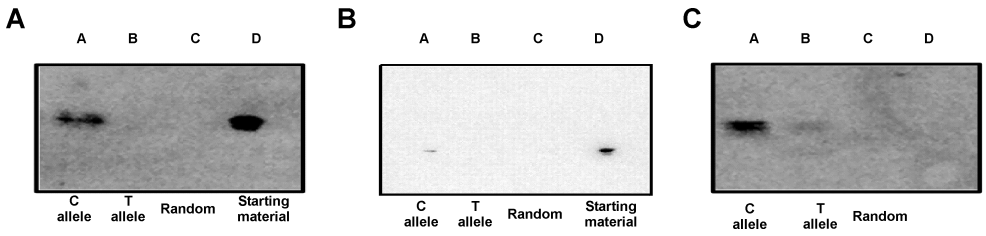


**Figure 17. Confirmation of HMGB1 binding on rs7903146 locus in human pancreatic islets from 3 donors (A-C)** Affinity column purified human pancreatic islets nucleoproteins using 61 nucleotides DNA probes. Lane A, HMGB1 was detected using DNA probes representing C-allele of *TCF7L2* in the purification. Lane B, nuclear proteins purified using DNA probes representing T-allele of *TCF7L2*. Lane C, Nuclear proteins purified using random DNA probes as negative control. Lane D, Starting material, human pancreatic islets nuclear extract **(Paper III, Figure 1A-C).**

The rigidity of the C- and T-allele was not significantly different and the interaction between HMGB1 and the C- and T-probes were further measured using DLS. The concentration of the HMGB1-C complex is higher than that of HMGB1-T complex in all molar-ratios measured, indicating that HMGB1 has a higher affinity to the C-allele compared to the T-allele. This finding is contrary to our initial hypothesis where the T-allele was anticipated to have a stronger enhancer activity if the open chromatin was the binding site of an enhancer protein complex. Our results revealed that the C-allele of rs7903146 might be more prone to HMGB1 binding and thereby regulates *TCF7L2* expression. One explanation of the discrepancy could be the method used in the initial report of an open chromatin state in the rs7903146 locus. In FAIRE-sequencing, DNA bound by proteins will be enriched in the organic phase during the extraction and the "naked" DNA in the water phase. If there had been binding proteins around T-allele, this DNA would not have been identified as an open-chromatin. In this case, HMGB1 fails to bind the T-allele due to a lower affinity and hence leaving the chromatin region exposed, therefore subsequently extracted in the organic phase during FAIRE. At the same time, the *in vitro* comparison of the binding affinity is still highly

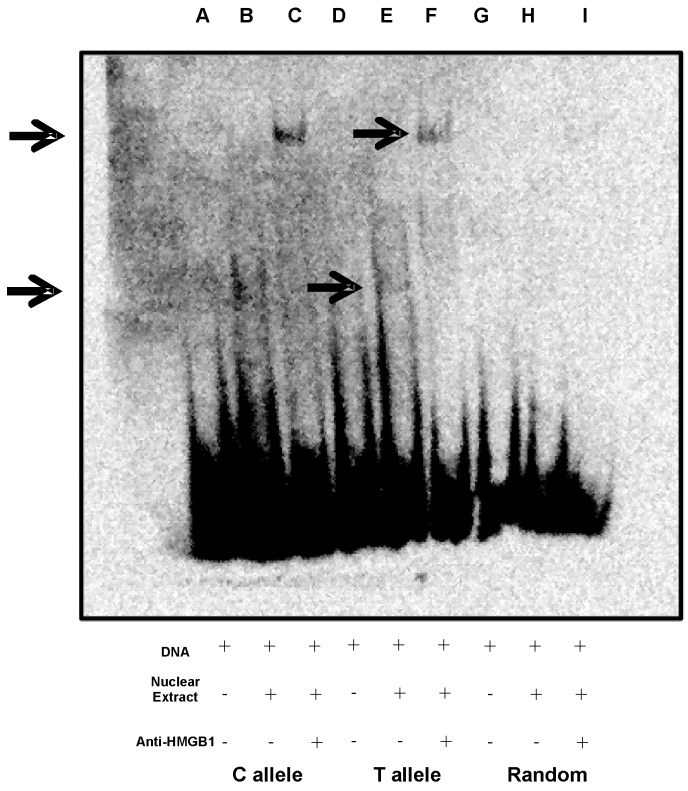


**Figure 18. Confirmation of HMGB1 binding on rs7903146 locus by Electrophoretic Mobility Shift Assay** Lane A, C-probe. Lane B, C-probe, plus HCT116 nuclear extract. Lane C, C-probe, plus HCT116 nuclear extract, plus antibody anti-HMGB1 Lane D, T-probe. Lane E, T-probe, plus HCT116 nuclear extract. Lane F, T-probe, plus HCT116 nuclear extract, plus antibody anti-HMGB1. Lane G, Random nucleotide probes. Lane H, Random nucleotide probes, plus HCT116 nuclear extract. Lane I, Random nucleotides probes, plus HCT116 nuclear extract, plus antibody anti-HMGB1 **(Paper III, Figure 1E).**

artificial. The chromatin state is for example not taken into consideration at all during DLS measurements.

In order to understand the consequences of HMGB1 binding at the rs7903146 locus, we silenced *HMGB1* in HCT116 cells (which have the CC-genotype of rs7903146). The expression of *TCF7L2* mRNA was reduced by 26% after *HMGB1* silencing. In addition, TCF7L2 activity measured using a Luciferase assay was also reduced by 44%. The

reduction of TCF7L2 transcriptional activity is very likely a result of a reduced total *TCF7L2* expression level. Activating the WNT signalling pathway resulted in increased TCF7L2 transcriptional activity by 7.83-fold in control cells and by 7.88-fold in the *HMGB1* silenced cells (Figure 19).

To compare the effect of HMGB1 in C- and T-allele carriers of rs7903146, an activity assay was performed in HCT116 cells using two plasmids carrying either C- or T-allele of rs7903146 (a generous gift from Pr. Francis Collins and Dr. Michael Stitzel) used in their previous work [82] (Figure 20).

However, the Luciferase activities that we have detected were comparable to the background activity in HCT116 cells. The activity level was similar in C- and T-allele carrying plasmids. The allele-specific binding of HMGB1 was only observed in human islets and not in HCT116 cells or HEK293 cells (study III, Figure 1). The experiments were performed in the HCT116 cells and not in human pancreatic islets due to technical challenges. In the previous publications, a higher enhancer activity was detected in the T-allele in two β cell lines MIN6, INS1 832/13 cells [72, 82] and in HeLa cells [82]. Therefore, the lack of activity can possibly be explained by a tissue-specific activation of HMGB1.

Many studies described the interaction between HMGB1 and p53 in different cancers [140-143]. Binding of p53 to its target sequences can be facilitated by HMGB1. P53, in turn, can regulate HMGB1 binding specificity [144]. HMGB1 might have different binding partners in human islets and in HCT116 cells. The activating protein complex containing HMGB1 might be able to bind the C-allele with an equal affinity as to the T-allele in HCT116 cells.

Previously, T-allele carriers were reported to have 2 – 5 fold higher expression of *TCF7L2* in human islets compared to C-allele carriers [23, 24, 62]. The higher expression was more prominent in islets from diabetic donors. We have also compared *TCF7L2* mRNA expression in 36 CC and 30 CT/TT non-diabetic donors. A 16% higher expression was observed in the CT/TT donors (Study II). But RNA-sequencing could not quantify the

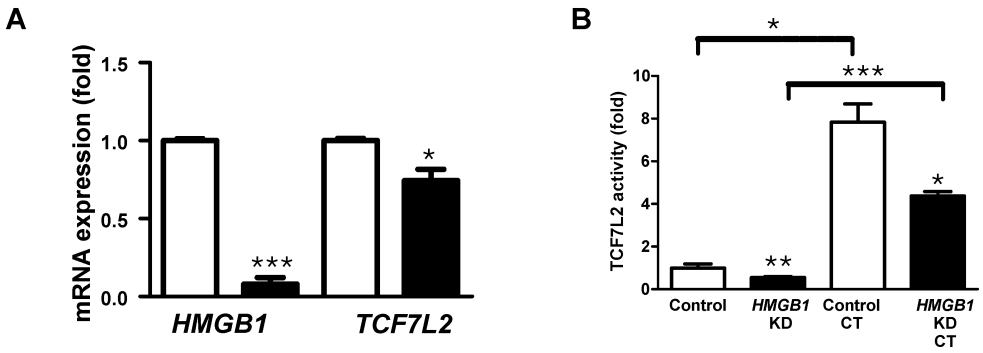


**Figure 19 HMGB1 binding to the open chromatin region of *TCF7L2* promoted the gene expression and transcriptional activity in HCT116 cells.** (A) *TCF7L2* mRNA expression in HCT116 cells was measured after *HMGB1* silencing. After 92% *HMGB1* down-regulation, the mRNA expression of *TCF7L2* was reduced by 26%. (B) TCF7L2 transcriptional activity after *HMGB1* silencing with and without CT99021 activation. With no external stimulation, the transcriptional activity of TCF7L2 was reduced by 45% compared to the scramble control siRNA treated cells. Upon CT 99021 stimulation, TCF7L2 transcriptional activity increased by 7.8 folds in scramble treated cells and 7.82 folds in *HMGB1* depleted HCT116 cells. The down-regulation of TCF7L2 transcriptional activity in *HMGB1* depleted HCT116 cells upon CT99021 stimulation was hence 44%. **(Paper III, Figure 5).**

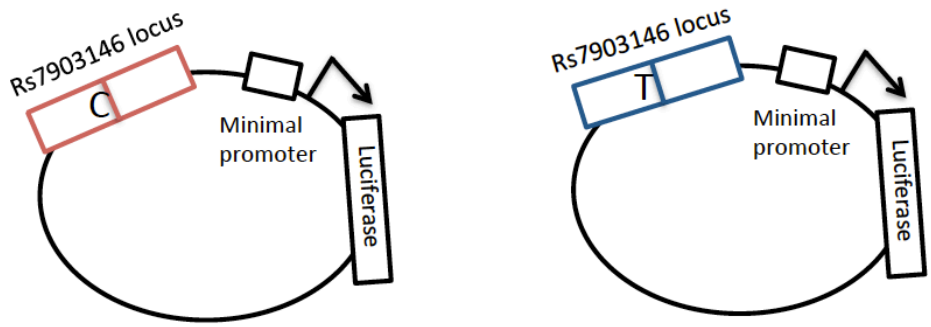


**Figure 20. Plasmids containing luciferase gene with rs7903146 locus as the enhancer [82]**

amount of mRNA that is transcribed from a specific allele, as the SNP is located in the intronic region. Therefore, exonic heterozygote "reporter SNPs", tagging the two alleles can be used to measure allele-specific expression (ASE) caused by an intronic variant. This experiment was conducted by Dr. Peter Osmark to measure the allele specific expression of *TCF7L2* in human islets. In this method, one additional exonic SNP needs to be identified in the mRNA of *TCF7L2*. The exonic SNP will reflect the genotype of rs7903146 on the mRNA level. The mRNAs transcribed from the two alleles are compared by measuring the ratio of exonic SNP using quantitative Sanger sequencing. Based on these results, no difference in expression between the C-allele and T-allele was observed (unpublished data). Therefore, up-regulation of *TCF7L2* mRNA expression might be secondary to altered glucose metabolism rather than a primary cause in T-allele carriers.

In summary, we have identified HMGB1 as a nuclear protein that binds the vicinity of rs7903146 in human pancreatic islets, possibly with a preference to the C-allele of rs7903146. However, it is not clear if HMGB1 binding can explain the higher enhancer activity of the T-allele in the rs7903146 locus.

# Study IV. The role of alternative splicing in the *TCF7L2* gene on islets function

Alternative splicing has been suggested to produce TCF7L2 protein isoforms with diverse biological functions and may thus be one mechanism whereby T2DM-associated genetic variants in the *TCF7L2* gene exert their diabetogenic action. We have previously described the four predominant splice variants of *TCF7L2* in human pancreatic islets. These four either include or not exon 4 in the 5'-end or exon 15 at the 3'-end [19]. The aim of study IV is to use Antisense Oligo Nucleotides (AONs) to selectively remove these exons in an attempt to enhance insulin secretion and glucose tolerance.

AONs are currently being tested in the clinical trials of Duchenne's muscle dystrophy [145]. AONs manipulate splicing by steric hindrance of splice factor binding to target pre-mRNA. The most effective AONs often target exonic splice enhancer sequences [120]. These enhancer sequences are bound by different Serine / Arginine rich proteins (SR proteins), which are the regulators of alternative splicing. By blocking enhancer sequences, SR proteins fail to bind and therefore inhibit the inclusion of the target exon into the mature mRNA. Using AONs, exon 4 and 15 of *TCF7L2* were successfully excluded from the mRNA in HCT116 cells (Figure 21) and in INS1 832/13 cells. A major advantage of AONs is that they altered the proportion of different splice variants without affecting the total expression level of the gene. AONs thus provided us with a possibility to test the function of almost any splice variant.

After exon 4 and 15 skipping at both basal and WNT-stimulated condition, transcriptome profiling was performed using Illumina BeadArray® in HCT116 cells. In total, the expression of 35013 genes was measured with the array. Exon 15 skipping caused few genes to be differently expressed (30 genes). At basal condition, skipping exon 4 lead

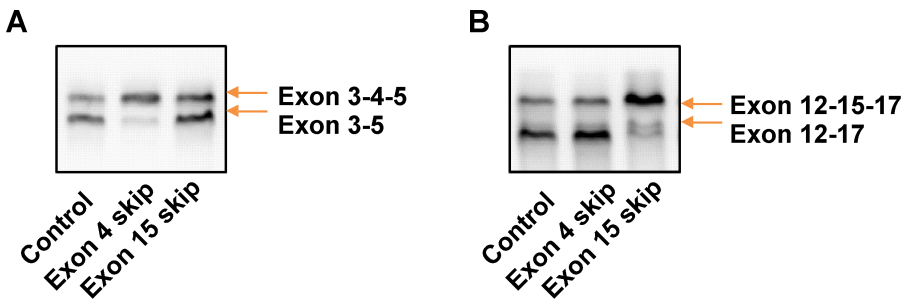


**Figure 21. Manipulation of *TCF7L2* splicing using AON in HCT116 cells visualized using RT-PCR on agarose gel** (A) PCR primers span over exon 3 to 5 of *TCF7L2*; (B) PCR primers span over exon 12 to 17 of *TCF7L2*.

to increased expression of 1058 genes and decreased expression of 234 genes. With WNT-stimulation, 1029 genes had increased expression and 115 genes had decreased expression after exon 4 skipping. With WNT-stimulation, the most enriched pathway was "insulin-signalling". "Cell apoptosis pathway" was also enriched both at basal and after WNT-stimulation. After exon 4 skipping, the expression of *TP53INP1* was significantly down-regulated, suggesting that exon 4 might have a pro-apoptotic function in HCT116 cells. Skipping exon 4 in INS1 832/13 cells led to increased proinsulin expression. TCF7L2 target genes identified in study II, such as *Isl1*, *Pdx1*, *Ins1* and *Pcsk2* were all up-regulated after exon 4 skipping. These results suggest that exon 4-containing splice variants can have inhibitory properties in pancreatic β cells. It is in line with a previous report [96], where exon 4-containing splice variants of *TCF7L2* displayed an inhibitory function, or a less activating function of TCF7L2 transcriptional activity in a colon cancer cell lines [96].

The AONs also provided us with additional information. Three AONs were tested for exon 4 and two for exon 15. The most efficient AON for skipping exon 4 targeted an exon-internal splice enhancer of SRSF1, SRSF2, SRSF5 and SRSF6. Therefore the four SR proteins became candidate splice factors responsible for exon 4 splicing. Similarly, SRSF2, SRSF5 and SRSF6 became candidate splice factors responsible for

exon 15 splicing. Sequential knock down of *SRSF1*, *SRSF2*, *SRSF5* and *SRSF6* revealed that SRSF1 and SRSF2 regulate the exclusion and the inclusion of exon 4 of *TCF7L2* respectively.

This observation might also shed some light on a previous observation of an association between $HbA_{1c}$ with exon 4 inclusion in risk T-allele carriers of rs7903146 [19]. It is known that insulin-signalling pathway leads to phosphorylation of SR proteins [91, 92]. There is also clinical evidence that splicing of exon 4 and 15 in *TCF7L2* is influenced by insulin or glucose [87]. In adipose tissue, the risk T-allele of rs7903146 is associated with lower retention of exon 4 [87]. In our previous study, the causality of the correlation between $HbA_{1c}$ and exon 4 usages in islets was largely unknown. A hypothesis of the causality between hyperglycaemia and exon 4 inclusion can be advanced: Hyperglycaemia might lead to slightly increased fasting glucose, which leads to elevated circulating insulin. Insulin alters the phosphorylation state of SR proteins and thereby the retention of exons in *TCF7L2* mRNA transcripts. The insulin hypothesis has not been tested in the previous study [19], as data on circulating insulin was not available from the cadaver donors. This hypothesis will be tested by incubating β cell line INS1 832/13 with 100 nM insulin and measure the exon 4 and 15 inclusion. On the other hand, TCF7L2 also regulates the *SRSF1* expression [146]. *Srsf2* is directly bound and regulated by TCF7L2 in INS1 832/13 cells according to two of our previous studies (Study I and II).

Taken together, this preliminary data supports the view that exon 4-containing *TCF7L2* transcripts might have inhibitory properties and that an increased expression of these transcripts could lead to increased apoptosis and reduced insulin synthesis. To confirm this hypothesis, glucose stimulated insulin secretion, insulin content and cell apoptosis needs to be measured in human islets after exon 4 skipping.

# Conclusions and future perspectives

To conclude the work in my PhD thesis,

- Pancreatic β cell survival is partly regulated by TCF7L2 though a p53-p53INP1 dependent pathway.

- *TCF7L2* is a key regulator of proinsulin synthesis and insulin processing, one of the most central mechanisms in the pathogenesis of T2DM. *ISL1* is a direct target of TCF7L2. ISL1 in turn regulates proinsulin synthesis and processing via regulation of *PCSK1*, *PCSK2*, *SLC30A8*, *MAFA*, *PDX1* and *NKX6.1*. *TCF7L2* target genes expressions are influenced by the rs7903146 genotype in pancreatic islets.

- HMGB1 was identified as a nuclear protein binding the rs7903146 locus in human pancreatic islets, probably with a preference to the C-allele of rs7903146. The binding of HMGB1 on rs7903146 locus regulates TCF7L2 expression and transcriptional activity.

- Exon 4 of *TCF7L2* seems to exert inhibitory properties in pancreatic β cells. An increased expression of exon 4-containing transcripts may lead to increased apoptosis and reduced insulin synthesis.

The culprit of T2DM is the inability to increase insulin secretion to compensate for the increased needs imposed by obesity and insulin resistance. *TCF7L2* interferes with all key steps of this protection against the westernized world. Several questions still need to be answered before we have a comprehensive understanding of *TCF7L2* risk genotype effect in all phenotypes observed. The current thesis provides several pieces of evidence explaining the dominant role of *TCF7L2* in the pathogenesis of T2DM. We should therefore no longer state that the mechanisms by which variants in the *TCF7L2* gene increase susceptibility to T2DM are unknown. Instead, *TCF7L2* is rather a gatekeeper of healthy β cell function.

# 概括

II型糖尿病的發展趨勢在全世界範圍內非常迅速。我國目前有超過九千萬II型糖尿病患者，另外有大量血糖不正常的糖尿病高危人群。其中只有三成病人對自己的病情有所了解。II型糖尿病的病人與日俱增。

說到糖尿病，自然就會想到胰島素，以及血糖和胰島素的關系。胰島素是人體中非常重要的激素（荷爾蒙）。人體所需要的糖分主要是從食物中的澱粉類食物（米飯，面食）中來的，肉類也含有一部分的糖分。糖分從小腸吸收進入血液循環。胰島素的功能是讓身體細胞把血液中的糖分吸收進入細胞（比如肌肉細胞）加以使用，同時降低血液中的血糖成份。如果胰島素分泌不足，或者人體細胞對胰島素的敏感性降低會引起血糖升高，長期高血糖會損害血管和臟器的健康，最終導致糖尿病並發癥的發生。糖尿病並發癥是一種常見的慢性並發癥，後果相當嚴重，足部壞疽、腎功能衰竭、尿毒癥、視力模糊不清、失明、腦血管病變（腦梗）、冠心病、心臟神經病變引起的心梗、皮膚病等都是糖尿病最常見的並發癥，也是導致糖尿病患者死亡的主要原因。胰島素是由人體胰腺貝塔細胞分泌產生的。所以了解貝塔細胞功能、知道貝塔細胞為何不正常分泌胰島素成為了防治糖尿病的關鍵。

糖尿病可以細分成四種類型。第一種，I型糖尿病。多數發生在兒童與青少年中。I型糖尿病是由於自身免疫系統在抵抗外來感染時，錯誤地摧毀自身胰腺貝塔細胞引起的。第二種，II型糖尿病。絕大多數人患的是II型糖尿病。II型糖尿病的發病是多因素的，有人類基因的內因因素加之與環境的外在因素相互作用引起II型糖尿病的發生。

由於長期缺乏鍛煉，加上不合理飲食，造成胰腺貝塔細胞功能減退，無法分泌人體所需要的胰島素，II型糖尿病在中老年階段發病的幾率很高。第三種，壬辰糖尿病，孕婦懷孕二十四周左右很可能發生的妊娠糖尿病，一般妊娠糖尿病在懷孕結束後會自動消除，但是妊娠糖尿病會造成胎兒過大，而且胎兒今後得糖尿病風險升高。第四種，罕見的“年青的成年發病型糖尿病”，這類糖尿病是由單個關鍵基因突變造成貝塔細胞病變。這類糖尿病在人類人口中的比重很少，經常會被誤診為I型糖尿病。

糖尿病的發病是要有遺傳與環境兩方面因素共同作用才會產生。從同卵雙胞胎的研究中得到證明，雙胞胎中的一人有糖尿病，另外一人的風險會上升，但是發病不是絕對的。II型糖尿病的遺傳性比I型糖尿病高，這可能是因為遺傳基因相近，加上親人之間生活飲食習慣相似的緣故。光有遺傳易感性是不夠致病的，一定要有環境因素觸發糖尿病發生。

我博士階段的論文課題是TCF7L2在糖尿病中的發病機理。我的博士論文的重要意義在於闡述TCF7L2在胰腺貝塔細胞中的重要作用。貳零零陸年隨著人類基因多樣性研究方法的快速發展，冰島遺傳學家用統計學方法從冰島民眾基因組中發現這個基因裏的遺傳信號與糖尿病密切相關，說明這個基因可能是II型糖尿病的重要致病基因。TCF7L2是第一個以高通量的醫學統計方法，對比糖尿病人群與非糖尿病人群的基因組成差異而確定的跟疾病有關的基因。這一結果在全世界不同的民族、不同地區的人群中相繼得到證實。貳零零陸年以後，有超過65個基因被確定為II型糖尿病致病基因。這些基因有著不同的功能，對糖尿病的影響也各不相同。

TCF7L2的功能原本只在癌癥中有所研究。在貳零零陸年之前，TCF7L2一直被公認為是導致直腸癌的致病基因，先前這個基因在導致II型糖尿病裏的角色並不被人所知。人體每個基因總是有兩個拷貝，

一個來自父親，另一個來自母親。很多臨床數據顯示，擁有一個缺陷TCF7L2基因的病人空腹時血液中的胰島素比帶正常基因拷貝的人低，而且進食後胰島素的分泌也比後者分泌得少，他們血液中不成熟、功能不全的胰島素比例升高很多。加之他們天生貝塔細胞就比其他人少。體外培養發現，如果貝塔細胞沒有TCF7L2，那麼細胞雕亡將大幅上升，胰島素分泌幾乎停止。

所以我博士論文的目的就是從分子水平上搞清為什麼缺陷TCF7L2可以引起這些病理癥狀。我的論文分為兩大部分，第一部分（兩篇論文）探討TCF7L2在貝塔細胞中正常的生理功能。了解正常生理功能是研究病理不可缺少的一步。我和課題組首先發現，TCF7L2基因可以調控一個重要的癌癥基因p53的表達（論文一）。p53的一個重要作用是促進細胞的程序性雕亡。在癌細胞裏，p53的活化可以減緩和防止癌癥發展，而它又似一把雙刃劍，在分泌胰島素的貝塔細胞裏，p53的活化卻會傷害這些細胞的存活。貝塔細胞對人類是十分珍貴的，人類的貝塔細胞在成年以後（孕婦除外）是不會再增殖分裂的。我的研究結果發現TCF7L2在正常情況下會抑制p53和p53INP1基因殺傷貝塔細胞，從而長期保護貝塔細胞。我於2011年十月將這一發現成果發表在《人類分子遺傳學》雜誌上。

我接下去發現，TCF7L2在貝塔細胞分泌胰島素的過程中，有著舉足輕重的作用（論文二）。這個基因在胰腺貝塔細胞中幾乎控制所有與胰島素分泌有關的分子通路的基因表達！TCF7L2控制三個其他重要的轉錄因子（MAFA，NEUROD1以及PDX1)，這三個轉錄因子是表達胰島素不可缺少的。胰島素是以前胰島素的形式在貝塔細胞中產生，並被包裝進入胰島素分泌囊中加工、成熟。這一過程也是受到TCF7L2的控制。前胰島素加工成胰島素，是由前荷爾蒙轉化酶來控制的。TCF7L2控制貝塔細胞中兩個最重要的轉化酶的表達。TCF7L2的功能如果出現問題，不僅胰島素的分泌總量下降，而且在分泌的胰島素中，不成熟沒有作用的前胰島素的比例大幅上升。這與臨床數據

顯示一致。除此之外，TCF7L2還控制著十幾個II型糖尿病致病基因的表達。貳零零陸年以後，有超過65個基因被確定為II型糖尿病致病基因。我們觀察發現，其中十幾個基因的表達可以直接或間接地由TCF7L2調控。TCF7L2與這些基因之間的聯系有待進一步研究。目前這方面的局限，很大一部分在於這些糖尿病致病基因本身的功能還不明確，從而難以判斷TCF7L2在這些基因致病機理中的作用。我的觀察起碼說明TCF7L2對糖尿病的作用是很廣泛的。

博士論文的第二部分是了解到底缺陷TCF7L2的缺陷是什麼。在論文三中，我們用了不同的生物化學方法，分離出一個蛋白質HMGB1, 這個蛋白質在貝塔細胞中會增加TCF7L2的數量。TCF7L2的數量增加原本是有益於人體防治糖尿病發生的。但是我們發現，HMGB1可能增加了缺陷TCF7L2的比例，這樣更進一步的加大了糖尿病的風險。在論文四中，我們進一步探索了缺陷TCF7L2。這個基因一共有十七部分組成，其中第四、十三、十四、十五、十六部分可以被選擇性地排除的。這個現象叫基因的選擇性剪接，在人類所有的基因中都普遍存在，目的是用一個基因產生好幾種不同作用的蛋白質。TCF7L2因此就有很多種不同的形式。其中我們發現，第四部分對貝塔細胞有著很大的影響。包含第四部分的TCF7L2可能會降低胰島素分泌，並且影響貝塔細胞存活。我們正在嘗試用分子生物學方法人工去除TCF7L2的第四部分的表達，從而防治糖尿病的發生。

胰島素在貝塔細胞中的分泌，是所有哺乳動物中共有的機理。我們所發現的機理是所有人類與哺乳動物中都有的機理。我希望我的博士論文可以增加對糖尿病致病機理的了解，對今後可能的治療方法有所幫助。雖然TCF7L2是最大的糖尿病基因，但是真正的影響也只有百分之四十，有規律地適量鍛煉是預防糖尿病的關鍵。

# Populäravetenskaplig Samanfattning

Diabetes är en sjukdom där kroppen har högre blodsockernivåer än normalt. Detta kallas för hyperglykemi. Blodsockret kommer från det vi äter och insulin är det hormon som är nödvändigt för att transportera socker från blodet till andra celler i kroppen där det ska användas. Om bukspottkörteln inte producerat tillräckligt med insulin, eller andra celler inte lyckas använda insulinet på rätt sätt, kommer blodsockernivån att öka. Det finns i princip tre typer av diabetes; Typ 1, typ 2 och graviditetsdiabetes. Typ 1 diabetes diagnostiseras vanligen i barn och ungdomar. Patienter med typ 1 diabetes förlorar de celler i bukspottskörteln som producerar insulin. Bara 5 % av alla människor med diabetes har denna form av sjukdomen. Med hjälp av insulininjektioner och andra behandlingar kan patienter med typ 1 diabetes leva långa och normala liv. Typ 2 diabetes är den vanligaste formen av diabetes. Om man har typ 2 diabetes använder inte kroppen insulin på rätt sätt. Detta kallas insulinresistens. I början av sjukdomsförloppet producerar bukspottkörteln extra insulin för att kompensera för detta. Men över tid klarar den inte av att hänga med, och kan inte längre producera tillräckligt med insulin för att hålla blodsockret på normala nivåer. Till slut är cellerna som producerar insulin uttömda och minskar i antal. Graviditetsdiabetes diagnostiseras i gravida kvinnor, oftast kring vecka 24. Vanligtvis försvinner den efter graviditeten, men graviditetsdiabetes ökar risken för barnet att utveckla typ 2 diabetes senare i livet. Diabetes ökar risken för att utveckla många allvarliga komplikationer så som diabetesfot, amputering, njursvikt, uremi, blindhet, högt blodtryck och hjärt- och kärlsjukdomar. Med rätt behandling och rekommenderade livsstilsändringar kan många människor med diabetes förhindra och fördröja uppkomsten av komplikationer.

Mitt arbete i denna avhandling har främst varit fokuserat på typ 2 diabetes. Två faktorer är viktiga för uppkomsten av typ 1 och typ 2

diabetes. Man ärver en kombination av gener som predisponerar för sjukdomen som sedan utlöses av en eller flera miljöfaktorer. Miljöfaktorerna kan bestå av dålig mathållning eller för lite fysisk aktivitet. Gener i sig är inte tillräckliga för att utveckla diabetes men typ 2 diabetes har en tydlig koppling till familjehistoria och arv, mer så än typ 1 diabetes. Förutom i några ovanliga former (MODY) kommer den genetiska risken från flera olika gener. I dagsläget har forskare upptäckt mer än 60 gener som är kopplade till uppkomsten av typ 2 diabetes. Jag arbetar med den första genen som identifierades; *TCF7L2*. Denna gen ger också den högsta risken bland de 60 generna. En person har två kopior av samma gen, en från varje förälder. Har man en defekt *TCF7L2* kopia ökar risken för att få diabetes med 40%. Min avhandling syftar till att förstå vad genen gör och varför den ökar risken för att utveckla typ 2 diabetes.

*TCF7L2* har tidigare visats ge ökad risk för koloncancer men dess roll i diabetes har varit okänd. Kliniker tog därför en närmare titt på effekten av den defekta *TCF7L2* genen. Har patienten en defekt gen orsakar det förhöjda insulinnivåer i blodet vid fasta och efter måltid och det finns mer av det omogna proinsulinet i blodet. Forskare i labbet upptäckte också att en defekt kopia av *TCF7L2* gör så att β-cellerna (de viktiga cellerna i bukspottkörteln som utsöndrar insulin) dör fortare. Patienter som har en defekt *TCF7L2* gen föds också med färre β celler än andra. Mitt arbete var att förstå varför och hur den defekta genen orsakar alla dessa skadliga effekter.

Denna avhandling är indelad i fyra delar. De två första studierna gjordes för att förstå den normala funktionen av TCF7L2 i β celler. Detta för att hjälpa oss förstå varför en defekt gen orsakar diabetes. Vi upptäckte att TCF7L2 förhindrar två cancergener att döda β celler. De två cancergenerna kallas p53 och p53INP. P53 skyddar vanligen människor från cancer genom att döda cancerartade celler innan en tumör kan bildas men om p53 dödar för många β celler kan det leda till typ 2 diabetes. TCF7L2 motverkar effekten av p53 för att balansera celldelning mot celldöd, men man har sett att om TCF7L2 är för aktivt kan det leda till uppkomsten av koloncancer. Min andra studie var kanske den viktigaste för den här avhandlingen. Vi upptäckte att TCF7L2 kontrollerar många viktiga proteiner i β cellerna. Det är som ett nätverk eller en pyramid av

gener; TCF7L2 kontrollerar vissa gener (*ISL1* och *NEUROD1*) och dessa i sin tur kontrollerar ytterligare gener (*MAFA*, *NKX6.1*, *PDX1*, *PCSK1*, *PCSK2* och *SLC30A8*). Vissa av dessa reglerar direkt insulinproduktionen (*INS1*, *INS2* och *INS3* gener), andra reglerar mognaden av insulin; Efter att insulin har producerats måste det bearbetas för att bli användbart (funktionellt) och TCF7L2 kontrollerar enzymer som är nödvändiga för detta. Det är därför man i patienter med en defekt kopia av *TCF7L2* hittar mer omoget och oanvändbart proinsulin i blodet vilket är den mest direkta effekten av TCF7L2 vid uppkomsten av typ 2 diabetes.

Den andra delen av min avhandling syftade till att förstå vad som är fel i den defekta kopian av *TCF7L2* genen genom att ta reda på vad defekten innebär och hur den påverkar funktionen av genen. I Studie II använde vi många olika biokemiska metoder och upptäckte att ett protein kallat HMGB1 ökar mängden TCF7L2 i β cellerna. Ökad mängd TCF7L2 borde vara fördelaktigt för att inte utveckla diabetes eftersom det reglerar effekten av *p53*, men förmodligen är det den defekta formen av TCF7L2 som ökar vilket i sin tur ökar risken för att utveckla diabetes. I Studie IV undersökte vi vilken form av proteinet som produceras från *TCF7L2* genen som kan skydda mot eller öka risken för diabetes. *TCF7L2* genen består av 17 olika delar, varav del 4, 13, 14, 15 och 16 kan vara antingen inkluderat eller exkluderat i det färdiga proteinet. Detta fenomen är mycket vanligt för alla mänskliga gener. Problemet med TCF7L2 är att om den 4e delen saknas är man skyddad från typ 2 diabetes, men om den är en del av proteinet kan detta vara en anledning till att β cellerna producerar mindre insulin och misslyckas med att skydda cellerna från att dö.

Jag tycker att min avhandling är väldigt värdefull för förståelsen av typ 2 diabetes. *TCF7L2* är den största typ 2 diabetes genen vi känner till i nuläget. Att förstå hur den predisponerar för typ 2 diabetes är fundamentalt för forskare och läkare för att i framtiden kunna hitta ett botemedel för en sjukdom som många idag drabbas av. Även om *TCF7L2* har den största genetiska risken för diabetes är den faktiska risken relativt liten. Att leva hälsosamt, ha goda matvanor och få tillräckligt med motion kanske är det bästa sättet att förhindra diabetes.

# Acknowledgements

Most people read this part first. Therefore it's the perfect occasion to express my gratitude towards the following people:

I would like to express my most sincere gratitude to **Leif** for giving me the opportunity to do my PhD in your lab. When I first time came from Lille to Malmö for the interview, I would never imagine how much knowledge I was going to learn and how much fun I was going to experience in the realm of science. Apart from giving me scientific ideas and guidance, you have always been incredibly tolerant to my spontaneity (a nicer word for craziness). I am grateful for your trust and your confidence in me. The trust, confidence and liberty meant a lot to me. It encouraged me to go further. I tried to use my excessive energy and moodiness to get more things done and switch from one failed idea to the next 'going-to-fail' idea. The work has already become my hobby. Plus, whenever I make an indifferent statement, you would just laugh about it. You are perfectly approachable, and genuinely nice to me. I am grateful and will never forget. Whatever I do later in my carrier (maybe comedian, who knows), I will do my best to live up to your expectations.

I thank **Erik** for all the constructive scientific advices you provided along the way. Whenever Ola and I have opposite opinions, you are always the best judge. Thank you for the continuous support during these years. **Claes,** thank you for all the good suggestions, the correction of several big mistakes in different projects during my PhD and all the support since 2010.

I thank the most kind **Ulrika (Blom-Nilsson)** for all the help throughout my PhD, which was a lot of extra work for you I guess... I sincerely thank you for the preparations for my dissertation. And **Mona,** for all the materials you have ordered for me, for being so reliable and

efficient all the time, for teaching me how to do my first qPCR and all the kindness.

I thank **Peter** for all the 'meat and mead' and the critical reading of this thesis. Several times during my PhD, when I am really mentally defeated, your kind words and philosophy can be surprisingly soothing and timely. Thank you, Peter. And **Kristoffer**, I know that I am the only person that can piss you off in just ten minutes during the last decade. I still remember one Sunday afternoon in BMC, when you saw me with my blots and said: 'Get a life, loser!' I thought you were very funny. Thank you for the good time I had in the group. **Nikolay**, when you worked over night or over the weekend, coming back to Malmö after you put your kids to bed, I am motivated more than ever. Thank you very much for the work experience. I started to like teamwork, and I enjoyed it quite a lot actually. Thank you for answering all the questions about the exotic methods. I love physics, and these methods are fun. You are incredibly devoted and meticulous in what you do. I loved the round table meeting every Friday with you guys!

**Emilia (Ottosson Laksso)**, the incredible friend. Apart from fashion and men, we did talk a lot about science. Thank you for all the good time we had together, and all the support you gave me. I do think you are intelligent, kind, reliable and sensible. I have learned a lot from you, and you know that. I thought I was independent, but you were kind of keeping an eye out for me a lot of the times. I am sure you will become extremely good at what you do, although you almost don't care about that. You, alone, made my moving to Sweden worth it.

**Anna (Jonsson)**, thank you for teaching me so much, from making scientific figures to cooking. You are a very tough girl, despite your look. "Yuedan, you should never make any mistake at any time." I do remember that. After you went to Denmark, I think I have become a bit sloppier, but you are right about that: one should not. It was you who showed me that when facing difficulties, sometimes just a little more persistence and the problems could be solved.

I thank **Maria, Gabriella, Mona, Helena, Anna (Berglund), Malin, Esa,** and **Ulrika** for all the good time we had in the lab and during the trips. These are simple good memories I will keep; the non-work-related fun I had in the lab. I think the little Swedish I know comes mostly from the daily conversations with all of you. Thank you for all your care and kindness. Besides, you all showed a great deal of caution, accuracy, persistence and professionalism during the daily work. I thank **Widet,** for showing me so much courage and support. You are one of the brave people I have met. **Helena,** thank you for the critical reading of the thesis.

**Britt-Marie** and **Anna-Maria,** LUDC would not function properly without you. You have always been so kind to me, no matter how much I asked for.

**Jalal,** thank you for always being so nice with me. I think you are a genuine nice person and a good friend. I hope all the best for you and Imane. **João** and **Rashmi**, I am so glad I share birthday with you both. It's kind of cool. You both are hard working and very good at what you do. It's inspiring. I hope it's a Capricorn thing. I will think of you both every time I turn older. **Amitabh**, I miss the football, I miss the burning spicy cooking and I miss the beer time. **Anders (Olsson),** thank you for the suggestions on thesis writing. It was very timely. And **Ulrika (Krus), Jacqueline, Stefan, Anne (Vähäniemi), Tord, Yang, Emma (Ahlqvist), Beatrice, Petr, Karl, Tina, Tasnim, Emma, Elin, Sasha, Mahboubeh, George, Céline, Malin (Svensson), Sammi, Wenny, Tarun, Jasmina, Petter, Jenny, Marloes, Marju, Valeriya, Bushra, Tereza, Sarah (Bonnetti), Maja and Carl (Ekman)** for the five years well spent! I specially thank **Jacqueline** for the critical reading of the thesis.

During my thesis, I had the honour to work with several brilliant collaborators, and basically one-way collaboration: they taught me a lot.

**Enming,** while everyone was telling me to relax, you were the only one that urged me to go further, faster and never give up. You have a way to make me work more when I was tired or lazy. You are the true friend. I

felt highly motivated when we were working together. You could exhaust until I got fever. But I thought it was not only me who was tired.

**Liliya,** you are not only a collaborator, but also my wedding witness. You are one of the rare loving, caring people I have met. Plus, you are very brave. Compare to you, I think I am way too whiny (Emilia would say 'That's not hard'). On top of that, you are very good at what you do. I do hope all the very best for you. I will be there for you, too (In fact, we will be there for you, Thomas and I). **Nils,** thank you for so much trust. Your encouragements were ego boost. **Mia, Anna-Helene** and **Andreas,** thank you for letting me spend a substantial amount of time in your lab. It was fun.

I thank **Johan, Jing, Mar** and **Alvis** for the most intense learning in Cambridge during my thesis. It was precious knowledge and great experience for me. I thank you for the good time I had in the UK and for showing me that so many good researchers are out there. **Jing**, thank you for late night chats and all the encouragements.

**Olga, Christine, Amélie, Elin, Emilia, Emma and Eva** for a wonderful summer in BMC. **Christine,** I am especially grateful that you showed me right way of running Western Blots. I used to think it is such a cumbersome technique, but you taught me the right way to do it. Whenever people complain that they hate WB, I really don't see why anymore. Learning from the best makes things much easier. **Johanna,** thank you for the good time in the lab. I really enjoyed it.

**Charlotte,** I haven't worked with you at all, but I admire you. During the five years I have seen how much your group has developed under your direction. What you have done is an inspiration. You manage to handle ~10 bright group members (together with data, funding and papers), 3 kids plus a husband with very high efficiency! I secretly want to become someone like you one day, the sooner the better.

**Malin,** CRC is not worth going without you. Thank you for all kinds of help, suggestions and support.

**My dear friends: Sayeh, Neivis, Bengt, Regina, Andreas, Anders (Littorin).** It's so good to have you. Thank you for the support, care and all the good time. Life has more fun with all of you around.

I want to thank **my parents**. You are definitely highly supportive for my work. **Dear father:** You are one of the hardest working people I have ever known. Thank you for the crucial support whenever I am truly in trouble. **Dear mother:** Thank you for being so patient about all my complaints. I took a lot for granted. But I do love you very much. I believe apart from my supervisors, you are the one who knows about my work the most.

To **my paternal grandparents and maternal grandfather:** It's devastating to recall that I can never see you again. I hope the fact that I put all my efforts into doing something serious and useful will appease you. I know my thesis would have made you all very proud. I miss you. To **my maternal grandmother:** The bravest woman I have ever known. I am a scientist now. I will be home soon.

I cannot express enough gratitude to **Thomas** for being so incredibly supportive all the time. I do believe a lot of people who read this acknowledgement wonder how you can put up with me. But the fact is: I think you can handle me very well. I am incredibly happy everyday living with you and Newton. You are so wonderfully nice and kind to everyone. I will also do my very best to make us happy. Thank you for the patient proof reading of my thesis (and the funny comments that made me laugh during the long writing process).

**Ola**,

I am very aware of the fact that my postdoc boss is not going to be like you, whoever that will be. It was great to be the 'only PhD student' for a 'single child'.

I will always remember some key sentences you said:

(1) Yuedan, think! Think! Think! Think!!

(2) Form your own hypothesis, test the idea in the lab, accept it or refute it.

(3) Yuedan, gut feeling is not valid. If you were Nobel Laureate and said, this is my gut feeling, some people would believe it. People like us need to provide evidences, present things with logic; otherwise people will very fast stop listening to you.

I know you sometimes were strict with me for my own good, so that I could develop into someone professional and upright in the not-so-near future. Only good tutors do this. I am sort of happy I didn't do genetics, as I wanted from the beginning with other PI. (I am still easily impressed by genetics, no doubt.) People told me it's difficult to be the first PhD student of someone, but I think I was lucky. I still have so much to learn from you in writing and scientific discussion.

The most important thing you taught me was to think first, and then think correctly. I am not good at that yet, but I think I start to see what I can't do. Next would be to take a step back and see a bigger picture.

The round table meetings were incredibly fun. I guess it was the first time I got interested in what other people are doing. It can be of great fun to see how other people's idea work.

I think since you have survived me, you will survive any future student. I am very sure soon enough I will not be your favourite PhD student, which is good for you and the group. But I will try very hard to become as good as you first, and then later better than you are, as all the

students should later know more than their teachers (still long long way to go for me). Otherwise, human race will not advance in science.

Saying thank you is not enough for all that you have done for me. Me too, I hope the best for you, boss.

# References


1. Poulsen, P., et al., *Heritability of type II (non-insulin-dependent) diabetes mellitus and abnormal glucose tolerance--a population-based twin study.* Diabetologia, 1999. **42**(2): p. 139-45.
2. Owen, K. and A.T. Hattersley, *Maturity-onset diabetes of the young: from clinical description to molecular genetic characterization.* Best Pract Res Clin Endocrinol Metab, 2001. **15**(3): p. 309-23.
3. Altshuler, D., et al., *The common PPARgamma Pro12Ala polymorphism is associated with decreased risk of type 2 diabetes.* Nat Genet, 2000. **26**(1): p. 76-80.
4. Brookes, A.J., *The essence of SNPs.* Gene, 1999. **234**(2): p. 177-86.
5. Billings, L.K. and J.C. Florez, *The genetics of type 2 diabetes: what have we learned from GWAS?* Ann N Y Acad Sci, 2010. **1212**: p. 59-77.
6. Grant, S.F., et al., *Variant of transcription factor 7-like 2 (TCF7L2) gene confers risk of type 2 diabetes.* Nat Genet, 2006. **38**(3): p. 320-3.
7. Saxena, R., et al., *Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels.* Science, 2007. **316**(5829): p. 1331-6.
8. Scott, L.J., et al., *A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants.* Science, 2007. **316**(5829): p. 1341-5.
9. Sladek, R., et al., *A genome-wide association study identifies novel risk loci for type 2 diabetes.* Nature, 2007. **445**(7130): p. 881-5.
10. Zeggini, E., et al., *Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes.* Science, 2007. **316**(5829): p. 1336-41.
11. Cho, Y.S., et al., *Meta-analysis of genome-wide association studies identifies eight new loci for type 2 diabetes in east Asians.* Nat Genet, 2012. **44**(1): p. 67-72.
12. Dupuis, J., et al., *New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk.* Nat Genet, 2010. **42**(2): p. 105-16.
13. Kooner, J.S., et al., *Genome-wide association study in individuals of South Asian ancestry identifies six new type 2 diabetes susceptibility loci.* Nat Genet, 2011. **43**(10): p. 984-9.

14. Voight, B.F., et al., *Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis.* Nat Genet, 2010. **42**(7): p. 579-89.
15. Saxena, R., et al., *Common single nucleotide polymorphisms in TCF7L2 are reproducibly associated with type 2 diabetes and reduce the insulin response to glucose in nondiabetic individuals.* Diabetes, 2006. **55**(10): p. 2890-5.
16. Helgason, A., et al., *Refining the impact of TCF7L2 gene variants on type 2 diabetes and adaptive evolution.* Nat Genet, 2007. **39**(2): p. 218-25.
17. Cho, Y.S., et al., *Genetics of type 2 diabetes in East Asian populations.* Curr Diab Rep, 2012. **12**(6): p. 686-96.
18. Florez, J.C., *Newly identified loci highlight beta cell dysfunction as a key cause of type 2 diabetes: where are the insulin resistance genes?* Diabetologia, 2008. **51**(7): p. 1100-10.
19. Osmark, P., et al., *Unique splicing pattern of the TCF7L2 gene in human pancreatic islets.* Diabetologia, 2009. **52**(5): p. 850-4.
20. Ip, W., Y.T. Chiang, and T. Jin, *The involvement of the wnt signaling pathway and TCF7L2 in diabetes mellitus: The current understanding, dispute, and perspective.* Cell Biosci, 2012. **2**(1): p. 28.
21. Norton, L., et al., *Chromatin occupancy of transcription factor 7-like 2 (TCF7L2) and its role in hepatic glucose metabolism.* Diabetologia, 2011. **54**(12): p. 3132-42.
22. Lyssenko, V., et al., *Mechanisms by which common variants in the TCF7L2 gene increase risk of type 2 diabetes.* J Clin Invest, 2007. **117**(8): p. 2155-63.
23. Le Bacquer, O., et al., *TCF7L2 rs7903146 impairs islet function and morphology in non-diabetic individuals.* Diabetologia, 2012. **55**(10): p. 2677-81.
24. Pang, D.X., A.J. Smith, and S.E. Humphries, *Functional analysis of TCF7L2 genetic variants associated with type 2 diabetes.* Nutr Metab Cardiovasc Dis, 2013. **23**(6): p. 550-6.
25. Elbein, S.C., et al., *Transcription factor 7-like 2 polymorphisms and type 2 diabetes, glucose homeostasis traits and gene expression in US participants of European and African descent.* Diabetologia, 2007. **50**(8): p. 1621-30.
26. Consoli, L., et al., *QTL analysis of proteome and transcriptome variations for dissecting the genetic architecture of complex traits in maize.* Plant Mol Biol, 2002. **48**(5-6): p. 575-81.
27. Elbein, S.C., et al., *Genetic risk factors for type 2 diabetes: a trans-regulatory genetic architecture?* Am J Hum Genet, 2012. **91**(3): p. 466-77.

28. Shu, L., et al., *Decreased TCF7L2 protein levels in type 2 diabetes mellitus correlate with downregulation of GIP- and GLP-1 receptors and impaired beta-cell function.* Hum Mol Genet, 2009. **18**(13): p. 2388-99.
29. Shu, L., et al., *TCF7L2 promotes beta cell regeneration in human and mouse pancreas.* Diabetologia, 2012. **55**(12): p. 3296-307.
30. Slattery, M.L., et al., *Transcription factor 7-like 2 polymorphism and colon cancer.* Cancer Epidemiol Biomarkers Prev, 2008. **17**(4): p. 978-82.
31. Saadeddin, A., et al., *The links between transcription, beta-catenin/JNK signaling, and carcinogenesis.* Mol Cancer Res, 2009. **7**(8): p. 1189-96.
32. Roose, J. and H. Clevers, *TCF transcription factors: molecular switches in carcinogenesis.* Biochim Biophys Acta, 1999. **1424**(2-3): p. M23-37.
33. Brack, A.S., et al., *Increased Wnt signaling during aging alters muscle stem cell fate and increases fibrosis.* Science, 2007. **317**(5839): p. 807-10.
34. Liu, H., et al., *Augmented Wnt signaling in a mammalian model of accelerated aging.* Science, 2007. **317**(5839): p. 803-6.
35. Tian, E., et al., *The role of the Wnt-signaling antagonist DKK1 in the development of osteolytic lesions in multiple myeloma.* N Engl J Med, 2003. **349**(26): p. 2483-94.
36. Mikels, A.J. and R. Nusse, *Wnts as ligands: processing, secretion and reception.* Oncogene, 2006. **25**(57): p. 7461-8.
37. Fujino, T., et al., *Low-density lipoprotein receptor-related protein 5 (LRP5) is essential for normal cholesterol metabolism and glucose-induced insulin secretion.* Proc Natl Acad Sci U S A, 2003. **100**(1): p. 229-34.
38. Doble, B.W. and J.R. Woodgett, *GSK-3: tricks of the trade for a multi-tasking kinase.* J Cell Sci, 2003. **116**(Pt 7): p. 1175-86.
39. Yi, F., P.L. Brubaker, and T. Jin, *TCF-4 mediates cell type-specific regulation of proglucagon gene expression by beta-catenin and glycogen synthase kinase-3beta.* J Biol Chem, 2005. **280**(2): p. 1457-64.
40. Ni, Z., et al., *Transcriptional activation of the proglucagon gene by lithium and beta-catenin in intestinal endocrine L cells.* J Biol Chem, 2003. **278**(2): p. 1380-7.
41. Yi, F., et al., *Cross talk between the insulin and Wnt signaling pathways: evidence from intestinal endocrine L cells.* Endocrinology, 2008. **149**(5): p. 2341-51.
42. Rulifson, I.C., et al., *Wnt signaling regulates pancreatic beta cell proliferation.* Proc Natl Acad Sci U S A, 2007. **104**(15): p. 6247-52.
43. Wells, J.M., et al., *Wnt/beta-catenin signaling is required for development of the exocrine pancreas.* BMC Dev Biol, 2007. **7**: p. 4.
44. Krutzfeldt, J. and M. Stoffel, *Regulation of wingless-type MMTV integration site family (WNT) signalling in pancreatic islets from wild-type and obese mice.* Diabetologia, 2010. **53**(1): p. 123-7.

45. Komiya, Y. and R. Habas, *Wnt signal transduction pathways.* Organogenesis, 2008. **4**(2): p. 68-75.
46. Manolagas, S.C. and M. Almeida, *Gone with the Wnts: beta-catenin, T-cell factor, forkhead box O, and oxidative stress in age-dependent diseases of bone, lipid, and glucose metabolism.* Mol Endocrinol, 2007. **21**(11): p. 2605-14.
47. Hatzis, P., et al., *Genome-wide pattern of TCF7L2/TCF4 chromatin occupancy in colorectal cancer cells.* Mol Cell Biol, 2008. **28**(8): p. 2732-44.
48. Gordon, M.D. and R. Nusse, *Wnt signaling: multiple pathways, multiple receptors, and multiple transcription factors.* J Biol Chem, 2006. **281**(32): p. 22429-33.
49. Eisenmann, D.M., *Wnt signaling.* WormBook, 2005: p. 1-17.
50. Florez, J.C., et al., *TCF7L2 polymorphisms and progression to diabetes in the Diabetes Prevention Program.* N Engl J Med, 2006. **355**(3): p. 241-50.
51. Loos, R.J., et al., *TCF7L2 polymorphisms modulate proinsulin levels and beta-cell function in a British Europid population.* Diabetes, 2007. **56**(7): p. 1943-7.
52. Kirchhoff, K., et al., *Polymorphisms in the TCF7L2, CDKAL1 and SLC30A8 genes are associated with impaired proinsulin conversion.* Diabetologia, 2008. **51**(4): p. 597-601.
53. Strawbridge, R.J., et al., *Genome-wide association identifies nine common variants associated with fasting proinsulin levels and provides new insights into the pathophysiology of type 2 diabetes.* Diabetes, 2011. **60**(10): p. 2624-34.
54. Gonzalez-Sanchez, J.L., et al., *Association of variants of the TCF7L2 gene with increases in the risk of type 2 diabetes and the proinsulin:insulin ratio in the Spanish population.* Diabetologia, 2008. **51**(11): p. 1993-7.
55. Stolerman, E.S., et al., *TCF7L2 variants are associated with increased proinsulin/insulin ratios but not obesity traits in the Framingham Heart Study.* Diabetologia, 2009. **52**(4): p. 614-20.
56. Dahlgren, A., et al., *Variants of the TCF7L2 gene are associated with beta cell dysfunction and confer an increased risk of type 2 diabetes mellitus in the ULSAM cohort of Swedish elderly men.* Diabetologia, 2007. **50**(9): p. 1852-7.
57. Silbernagel, G., et al., *Association of TCF7L2 SNPs with age at onset of type 2 diabetes and proinsulin/insulin ratio but not with glucagon-like peptide 1.* Diabetes Metab Res Rev, 2011. **27**(5): p. 499-505.

58. McCaffery, J.M., et al., *TCF7L2 polymorphism, weight loss and proinsulin:insulin ratio in the diabetes prevention program.* PLoS One, 2011. **6**(7): p. e21518.
59. Villareal, D.T., et al., *TCF7L2 variant rs7903146 affects the risk of type 2 diabetes by modulating incretin action.* Diabetes, 2010. **59**(2): p. 479-85.
60. Shu, L., et al., *Transcription factor 7-like 2 regulates beta-cell survival and function in human pancreatic islets.* Diabetes, 2008. **57**(3): p. 645-53.
61. da Silva Xavier, G., et al., *Abnormal glucose tolerance and insulin secretion in pancreas-specific Tcf7l2-null mice.* Diabetologia, 2012. **55**(10): p. 2667-76.
62. Lyssenko, V., *The transcription factor 7-like 2 gene and increased risk of type 2 diabetes: an update.* Curr Opin Clin Nutr Metab Care, 2008. **11**(4): p. 385-92.
63. da Silva Xavier, G., et al., *TCF7L2 regulates late events in insulin secretion from pancreatic islet beta-cells.* Diabetes, 2009. **58**(4): p. 894-905.
64. Schafer, S.A., et al., *Impaired glucagon-like peptide-1-induced insulin secretion in carriers of transcription factor 7-like 2 (TCF7L2) gene polymorphisms.* Diabetologia, 2007. **50**(12): p. 2443-50.
65. Prokunina-Olsson, L., et al., *Tissue-specific alternative splicing of TCF7L2.* Hum Mol Genet, 2009. **18**(20): p. 3795-804.
66. Zhou, Y., et al., *Survival of pancreatic beta cells is partly controlled by a TCF7L2-p53-p53INP1-dependent pathway.* Hum Mol Genet, 2012. **21**(1): p. 196-207.
67. Korinek, V., et al., *Depletion of epithelial stem-cell compartments in the small intestine of mice lacking Tcf-4.* Nat Genet, 1998. **19**(4): p. 379-83.
68. Yang, H., et al., *Reduction in Tcf7l2 expression decreases diabetic susceptibility in mice.* Int J Biol Sci, 2012. **8**(6): p. 791-801.
69. Boj, S.F., et al., *Diabetes risk gene and Wnt effector Tcf7l2/TCF4 controls hepatic response to perinatal and adult metabolic demand.* Cell, 2012. **151**(7): p. 1595-607.
70. Savic, D., et al., *Alterations in TCF7L2 expression define its role as a key regulator of glucose metabolism.* Genome Res, 2011. **21**(9): p. 1417-25.
71. da Silva Xavier, G., et al., *Animal models of GWAS-identified type 2 diabetes genes.* J Diabetes Res, 2013. **2013**: p. 906590.
72. Gaulton, K.J., et al., *A map of open chromatin in human pancreatic islets.* Nat Genet, 2010. **42**(3): p. 255-9.
73. Pilgaard, K., et al., *The T allele of rs7903146 TCF7L2 is associated with impaired insulinotropic action of incretin hormones, reduced 24 h profiles of plasma insulin and glucagon, and increased hepatic glucose production in young healthy men.* Diabetologia, 2009. **52**(7): p. 1298-307.

74. Wegner, L., et al., *Impact of TCF7L2 rs7903146 on insulin secretion and action in young and elderly Danish twins.* J Clin Endocrinol Metab, 2008. **93**(10): p. 4013-9.
75. Musso, G., et al., *Transcription factor 7-like 2 polymorphism modulates glucose and lipid homeostasis, adipokine profile, and hepatocyte apoptosis in NASH.* Hepatology, 2009. **49**(2): p. 426-35.
76. Faerch, K., et al., *Incretin and pancreatic hormone secretion in Caucasian non-diabetic carriers of the TCF7L2 rs7903146 risk T allele.* Diabetes Obes Metab, 2013. **15**(1): p. 91-5.
77. Alibegovic, A.C., et al., *The T-allele of TCF7L2 rs7903146 associates with a reduced compensation of insulin secretion for insulin resistance induced by 9 days of bed rest.* Diabetes, 2010. **59**(4): p. 836-43.
78. Vangipurapu, J., et al., *Association of indices of liver and adipocyte insulin resistance with 19 confirmed susceptibility loci for type 2 diabetes in 6,733 non-diabetic Finnish men.* Diabetologia, 2011. **54**(3): p. 563-71.
79. Cadoret, A., et al., *New targets of beta-catenin signaling in the liver are involved in the glutamine metabolism.* Oncogene, 2002. **21**(54): p. 8293-301.
80. Loeppen, S., et al., *Overexpression of glutamine synthetase is associated with beta-catenin-mutations in mouse liver tumors during promotion of hepatocarcinogenesis by phenobarbital.* Cancer Res, 2002. **62**(20): p. 5685-8.
81. Ip, W., et al., *The Wnt signaling pathway effector TCF7L2 is upregulated by insulin and represses hepatic gluconeogenesis.* Am J Physiol Endocrinol Metab, 2012. **303**(9): p. E1166-76.
82. Stitzel, M.L., et al., *Global epigenomic analysis of primary human pancreatic islets provides insights into type 2 diabetes susceptibility loci.* Cell Metab, 2010. **12**(5): p. 443-55.
83. Le Bacquer, O., et al., *TCF7L2 splice variants have distinct effects on beta-cell turnover and function.* Hum Mol Genet, 2011. **20**(10): p. 1906-15.
84. Duval, A., et al., *The human T-cell transcription factor-4 gene: structure, extensive characterization of alternative splicings, and mutational analysis in colorectal cancer cell lines.* Cancer Res, 2000. **60**(14): p. 3872-9.
85. Tsedensodnom, O., et al., *Identification of T-cell factor-4 isoforms that contribute to the malignant phenotype of hepatocellular carcinoma cells.* Exp Cell Res, 2011. **317**(7): p. 920-31.
86. Locke, J.M., et al., *An alternative polyadenylation signal in TCF7L2 generates isoforms that inhibit T cell factor/lymphoid-enhancer factor (TCF/LEF)-dependent target genes.* Diabetologia, 2011. **54**(12): p. 3078-82.

87. Mondal, A.K., et al., *Genotype and tissue-specific effects on alternative splicing of the transcription factor 7-like 2 gene in humans.* J Clin Endocrinol Metab, 2010. **95**(3): p. 1450-7.
88. Kaminska, D., et al., *Adipose tissue TCF7L2 splicing is regulated by weight loss and associates with glucose and fatty acid metabolism.* Diabetes, 2012. **61**(11): p. 2807-13.
89. Hansson, O., et al., *Molecular function of TCF7L2: Consequences of TCF7L2 splicing for molecular function and risk for type 2 diabetes.* Curr Diab Rep, 2010. **10**(6): p. 444-51.
90. Prokunina-Olsson, L. and J.L. Hall, *Evidence for neuroendocrine function of a unique splicing form of TCF7L2 in human brain, islets and gut.* Diabetologia, 2010. **53**(4): p. 712-6.
91. Jiang, K., et al., *Akt2 regulation of Cdc2-like kinases (Clk/Sty), serine/arginine-rich (SR) protein phosphorylation, and insulin-induced alternative splicing of PKCbetaII messenger ribonucleic acid.* Endocrinology, 2009. **150**(5): p. 2087-97.
92. Reardon, H.T., et al., *The polypyrimidine tract binding protein regulates desaturase alternative splicing and PUFA composition.* J Lipid Res, 2011. **52**(12): p. 2279-86.
93. Blaustein, M., et al., *Concerted regulation of nuclear and cytoplasmic activities of SR proteins by AKT.* Nat Struct Mol Biol, 2005. **12**(12): p. 1037-44.
94. Patel, N.A., et al., *Insulin regulates alternative splicing of protein kinase C beta II through a phosphatidylinositol 3-kinase-dependent pathway involving the nuclear serine/arginine-rich splicing factor, SRp40, in skeletal muscle cells.* J Biol Chem, 2001. **276**(25): p. 22648-54.
95. Patel, N.A., et al., *Molecular and genetic studies imply Akt-mediated signaling promotes protein kinase CbetaII alternative splicing via phosphorylation of serine/arginine-rich splicing factor SRp40.* J Biol Chem, 2005. **280**(14): p. 14302-9.
96. Weise, A., et al., *Alternative splicing of Tcf7l2 transcripts generates protein variants with differential promoter-binding and transcriptional activation properties at Wnt/beta-catenin targets.* Nucleic Acids Res, 2010. **38**(6): p. 1964-81.
97. Groskreutz, D. and E.T. Schenborn, *Reporter systems.* Methods Mol Biol, 1997. **63**: p. 11-30.
98. Timmons, L., et al., *Inducible systemic RNA silencing in Caenorhabditis elegans.* Mol Biol Cell, 2003. **14**(7): p. 2972-83.
99. Hammar, E., et al., *Extracellular matrix protects pancreatic beta-cells against apoptosis: role of short- and long-term signaling pathways.* Diabetes, 2004. **53**(8): p. 2034-41.

100. Renart, J., J. Reiser, and G.R. Stark, *Transfer of proteins from gels to diazobenzyloxymethyl-paper and detection with antisera: a method for studying antibody specificity and antigen structure.* Proc Natl Acad Sci U S A, 1979. **76**(7): p. 3116-20.
101. Towbin, H., T. Staehelin, and J. Gordon, *Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications.* Proc Natl Acad Sci U S A, 1979. **76**(9): p. 4350-4.
102. Burnette, W.N., *"Western blotting": electrophoretic transfer of proteins from sodium dodecyl sulfate--polyacrylamide gels to unmodified nitrocellulose and radiographic detection with antibody and radioiodinated protein A.* Anal Biochem, 1981. **112**(2): p. 195-203.
103. Shankaranarayanan, P., et al., *Single-tube linear DNA amplification for genome-wide studies using a few thousand cells.* Nat Protoc, 2012. **7**(2): p. 328-38.
104. Shankaranarayanan, P., et al., *Single-tube linear DNA amplification (LinDA) for robust ChIP-seq.* Nat Methods, 2011. **8**(7): p. 565-7.
105. Engvall, E. and P. Perlmann, *Enzyme-linked immunosorbent assay (ELISA). Quantitative assay of immunoglobulin G.* Immunochemistry, 1971. **8**(9): p. 871-4.
106. Engvall, E., *The ELISA, enzyme-linked immunosorbent assay.* Clin Chem, 2010. **56**(2): p. 319-20.
107. Garner, M.M. and A. Revzin, *A gel electrophoresis method for quantifying the binding of proteins to specific DNA regions: application to components of the Escherichia coli lactose operon regulatory system.* Nucleic Acids Res, 1981. **9**(13): p. 3047-60.
108. Ji, H. and W.H. Wong, *TileMap: create chromosomal map of tiling array hybridizations.* Bioinformatics, 2005. **21**(18): p. 3629-36.
109. He, K., et al., *NTAP: for NimbleGen tiling array ChIP-chip data analysis.* Bioinformatics, 2009. **25**(14): p. 1838-40.
110. Kim, D., et al., *TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions.* Genome Biol, 2013. **14**(4): p. R36.
111. Robinson, M.D., D.J. McCarthy, and G.K. Smyth, *edgeR: a Bioconductor package for differential expression analysis of digital gene expression data.* Bioinformatics, 2010. **26**(1): p. 139-40.
112. Trapnell, C., et al., *Differential analysis of gene regulation at transcript resolution with RNA-seq.* Nat Biotechnol, 2013. **31**(1): p. 46-53.
113. Ambrose, E.J., *A surface contact microscope for the study of cell movements.* Nature, 1956. **178**(4543): p. 1194.
114. Li, H. and R. Durbin, *Fast and accurate short read alignment with Burrows-Wheeler transform.* Bioinformatics, 2009. **25**(14): p. 1754-60.

115. Sakib, M.N., et al., *Improving transmission efficiency of large sequence alignment/map (SAM) files.* PLoS One, 2011. **6**(12): p. e28251.
116. Zhang, Y., et al., *Model-based analysis of ChIP-Seq (MACS).* Genome Biol, 2008. **9**(9): p. R137.
117. Guinier, A., *La diffraction des rayons X aux très petits angles; application a l'étude de phénomènes ultramicroscopiques* Annales de physique, 1939. **12**: p. 161-237.
118. Aartsma-Rus, A., et al., *Comparative analysis of antisense oligonucleotide analogs for targeted DMD exon 46 skipping in muscle cells.* Gene Ther, 2004. **11**(18): p. 1391-8.
119. Aartsma-Rus, A., et al., *Guidelines for antisense oligonucleotide design and insight into splice-modulating mechanisms.* Mol Ther, 2009. **17**(3): p. 548-53.
120. Aartsma-Rus, A., et al., *Exonic sequences provide better targets for antisense oligonucleotides than splice site sequences in the modulation of Duchenne muscular dystrophy splicing.* Oligonucleotides, 2010. **20**(2): p. 69-77.
121. Popplewell, L.J., et al., *Design of phosphorodiamidate morpholino oligomers (PMOs) for the induction of exon skipping of the human DMD gene.* Mol Ther, 2009. **17**(3): p. 554-61.
122. Dunning, M.J., et al., *beadarray: R classes and methods for Illumina bead-based data.* Bioinformatics, 2007. **23**(16): p. 2183-4.
123. Elghazi, L. and E. Bernal-Mizrachi, *Akt and PTEN: beta-cell mass and pancreas plasticity.* Trends Endocrinol Metab, 2009. **20**(5): p. 243-51.
124. Tomasini, R., et al., *TP53INP1s and homeodomain-interacting protein kinase-2 (HIPK2) are partners in regulating p53 activity.* J Biol Chem, 2003. **278**(39): p. 37722-9.
125. Oda, K., et al., *p53AIP1, a potential mediator of p53-dependent apoptosis, and its regulation by Ser-46-phosphorylated p53.* Cell, 2000. **102**(6): p. 849-62.
126. Scott, D.A., *Crystalline insulin.* Biochem J, 1934. **28**(4): p. 1592-1602 1.
127. Hay, C.W. and K. Docherty, *Comparative analysis of insulin gene promoters: implications for diabetes research.* Diabetes, 2006. **55**(12): p. 3201-13.
128. Bernardo, A.S., C.W. Hay, and K. Docherty, *Pancreatic transcription factors and their role in the birth, life and survival of the pancreatic beta cell.* Mol Cell Endocrinol, 2008. **294**(1-2): p. 1-9.
129. Vanderford, N.L., *Regulation of beta-cell-specific and glucose-dependent MafA expression.* Islets, 2011. **3**(1): p. 35-7.
130. Rosengren, A.H., et al., *Reduced insulin exocytosis in human pancreatic beta-cells with gene variants linked to type 2 diabetes.* Diabetes, 2012. **61**(7): p. 1726-33.

131. Groop, L., *Open chromatin and diabetes risk.* Nat Genet, 2010. **42**(3): p. 190-2.
132. Thomas, J.O. and A.A. Travers, *HMG1 and 2, and related 'architectural' DNA-binding proteins.* Trends Biochem Sci, 2001. **26**(3): p. 167-74.
133. Stott, K., et al., *Structure of a complex of tandem HMG boxes and DNA.* J Mol Biol, 2006. **360**(1): p. 90-104.
134. Najima, Y., et al., *High mobility group protein-B1 interacts with sterol regulatory element-binding proteins to enhance their DNA binding.* J Biol Chem, 2005. **280**(30): p. 27523-32.
135. Tang, D., et al., *High-mobility group box 1 and cancer.* Biochim Biophys Acta, 2010. **1799**(1-2): p. 131-40.
136. Stros, M., *HMGB proteins: interactions with DNA and chromatin.* Biochim Biophys Acta, 2010. **1799**(1-2): p. 101-13.
137. Agresti, A. and M.E. Bianchi, *HMGB proteins and gene expression.* Curr Opin Genet Dev, 2003. **13**(2): p. 170-8.
138. Muller, S., L. Ronfani, and M.E. Bianchi, *Regulated expression and subcellular localization of HMGB1, a chromatin protein with a cytokine function.* J Intern Med, 2004. **255**(3): p. 332-43.
139. Thomas, J.O. and K. Stott, *H1 and HMGB1: modulators of chromatin structure.* Biochem Soc Trans, 2012. **40**(2): p. 341-6.
140. Rowell, J.P., et al., *HMGB1-facilitated p53 DNA binding occurs via HMG-Box/p53 transactivation domain interaction, regulated by the acidic tail.* Structure, 2012. **20**(12): p. 2014-24.
141. Livesey, K.M., et al., *p53/HMGB1 complexes regulate autophagy and apoptosis.* Cancer Res, 2012. **72**(8): p. 1996-2005.
142. Stros, M., et al., *High-affinity binding of tumor-suppressor protein p53 and HMGB1 to hemicatenated DNA loops.* Biochemistry, 2004. **43**(22): p. 7215-25.
143. Jayaraman, L., et al., *High mobility group protein-1 (HMG-1) is a unique activator of p53.* Genes Dev, 1998. **12**(4): p. 462-72.
144. Uramoto, H., et al., *Physical interaction of tumour suppressor p53/p73 with CCAAT-binding transcription factor 2 (CTF2) and differential regulation of human high-mobility group 1 (HMG1) gene expression.* Biochem J, 2003. **371**(Pt 2): p. 301-10.
145. Goemans, N.M., et al., *Systemic administration of PRO051 in Duchenne's muscular dystrophy.* N Engl J Med, 2011. **364**(16): p. 1513-22.
146. Shitashige, M., et al., *Involvement of splicing factor-1 in beta-catenin/T-cell factor-4-mediated gene transactivation and pre-mRNA splicing.* Gastroenterology, 2007. **132**(3): p. 1039-54.

147. Du, A., et al., *Islet-1 is required for the maturation, proliferaion, and survival of the endocrine pancreas.* Diabetes, 2007. **58**(9): p. 2059-69.