Technische Universität München
TUM School of Life Sciences

# Identification of transcription factors and cofactors binding to the disease risk variant *rs7903146* that modulate gene expression and type 2 diabetes phenotypes

Lena Maria Hoerburger

Vollständiger Abdruck der von der TUM School of Life Sciences der Technischen Universität München zur Erlangung des akademischen Grades einer

Doktorin der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitz: Prof. Dr. Johann J. Hauner

Prüfer*innen der Dissertation:

1. Prof. Dr. Heiko Witt
2. Prof. Dr. Nina Henriette Uhlenhaut

Die Dissertation wurde am 10.05.2022 bei der Technischen Universität München eingereicht und durch die TUM School of Life Sciences am 07.07.2022 angenommen.

# 1 Abstract

Genome wide association studies (GWAS) revealed large numbers of variants that could explain the heritability of type 2 diabetes (T2D). Since the majority of variants associated with common diseases lie in the non-coding region of the genome, research is shifting its focus away from coding mutations towards gene regulation mechanisms. In this thesis, we focused specifically on a single nucleotide polymorphism (SNP) that is located in intron four of the Transcription factor 7-like 2 (*TCF7L2*) gene. Many studies in various tissues have provided mechanisms how changes in the TCF7L2 content can lead to impaired glucose metabolism. The variant *rs7903146* has been identified over 15 years ago and since then research groups have been looking for the molecular mechanism that explains the increased risk of T2D. The variant does not seem to affect splicing of *TCF7L2*. The main hypothesis arose that the variant leads to an altered binding of a transcription factor. The T risk allele is located in open chromatin and can result in increased *TCF7L2* expression in human pancreatic islets. This led to the assumption that transcriptional activators specifically bind to the T risk allele, which could explain the higher *TCF7L2* expression observed in carriers of this allele. Up to now, two research groups have found the C non-risk allele binding proteins PARP1 and HMGB1, both are ubiquitously expressed and known to have a rather low sequence-specificity. Here we were able to find a variety of potential activating proteins that bind to the T risk allele. Using affinity chromatography coupled with label-free LC MS/MS, we identified several candidates and narrowed them down further in liver and pancreatic cell lines by a large-scale transient knockdown approach. Next, using gene reporter assays, EMSA and ChIP, we verified proteins as allele-specific regulatory factors of gene expression leading to a disease-risk phenotype. We identified interesting candidate transcription factors, both with reported activating and inhibiting properties. The high number of identified regulators with small impact on gene expression might point towards a complex interaction of several proteins that needs to be further investigated.

# Contents


**1 Abstract** **3**

**2 Abbreviations** **7**

**3 Introduction** **11**

3.1 Pathophysiology of type 2 diabetes . . . 11

3.2 Genetics of type 2 diabetes . . . 13

3.2.1 Non-coding variants associated with type 2 diabetes . . . 14

3.3 The T2D risk locus *TCF7L2* . . . 14

3.3.1 Function of the TCF7L2 in the WNT signaling pathway . . . 14

3.3.2 Splicing of the *TCF7L2* gene locus . . . 16

3.3.3 Potential regulating mechanisms of *rs7903146* . . . 17

3.4 Aim of thesis . . . 19

**4 Material and Methods** **21**

4.1 Material . . . 21

4.1.1 Chemicals and consumables . . . 21

4.1.2 Buffers and prepared solutions . . . 27

4.2 Methods . . . 30

4.2.1 Cell culture . . . 30

4.2.2 Whole protein harvest . . . 31

4.2.3 Nuclear extraction . . . 31

4.2.4 Electrophoretic mobility shift assay (EMSA) . . . 31

4.2.5 Affinity chromatography . . . 33

4.2.6 Liquid chromatography coupled with tandem mass spectrometry (LC-MS/MS) 33

4.2.7 RNA isolation . . . 34

4.2.8 cDNA synthesis . . . 35

4.2.9 Overexpression of transcription factors of interest . . . 35

4.2.10 Real-time Polymerase Chain Reaction (RT-PCR) . . . 35

4.2.11 Genotyping . . . 36

4.2.12 Interferone γ stimulation . . . 36

4.2.13 Western blot . . . 36

4.2.14 Replication of vectors . . . 37

4.2.15 Luciferase reporter gene assays . . . 37

4.2.16 Transient siRNA knockdown . . . 38

4.2.17 Chromatin immunoprecipitation (ChIP) . . . 39

**5 Results** **41**

5.1 Identification of candidate *rs7903146* binding proteins . . . . . 41
5.1.1 Genotypes of selected cell lines . . . . . 41
5.1.2 Selection of regulated target genes . . . . . 41
5.1.3 Affinity chromatography coupled with LC MS/MS identifies candidate regulatory Transcription Factors . . . . . 41
5.1.4 Transient siRNA knockdown to evaluate effect of candidate Transcription Factors 42
5.1.5 Interferone γ stimulation . . . . . 43
5.2 *PDX1* . . . . . 46
5.2.1 Pdx1 was identified with AC LC MS/MS binding specifically to the C non-risk allele. . . . . 46
5.2.2 The binding of PDX1 were confirmed in EMSA, supershift EMSA and competition EMSA. . . . . 46
5.2.3 Knockdown of *Pdx1* in Ins1 cells only showed a slight effect on *Ascl5* expression. 51
5.2.4 Overexpression of *PDX1* in 1.1B4 did not show a relevant effect on mRNA expression of *TCF7L2* or *ACSL5*. . . . . 51
5.2.5 Overexpression of *PDX1* in a gene reporter assay showed no allele-specific effect on *TCF7L2* promotor function. . . . . 51
5.3 *CEBP* Family . . . . . 54
5.3.1 CEBPs were identified to be preferably binding to the T risk allele in liver cell lines in AC LC MS/MS. . . . . 54
5.3.2 The binding of the CEBPs to the T risk allele was confirmed in EMSA, supershift EMSA and competition EMSA. . . . . 55
5.3.3 CEBPs interact with each other *in vitro*. . . . . 57
5.3.4 Knockdown of *CEBPB* in 1.1B4 showed a significant decrease in *TCF7L2* . . . 57
5.3.5 Overexpression of *CEBPG* and *CEBPA* in a gene reporter assay showed no allele-specific effect on *TCF7L2* promotor function. . . . . 59
5.4 *STAT1* . . . . . 63
5.4.1 STAT1 binds 5.5 fold stronger to the T risk allele in HepG2 in AC LC-MS/MS. 63
5.4.2 Binding of STAT1 to the risk locus was confirmed in supershift EMSA. . . . . 63
5.4.3 Knockdown of *STAT1* in HepG2 and Huh7 shows no significant effect on target gene expression. . . . . 63
5.4.4 Knockdown of *STAT1* after IFNγ stimulation in HepG2 showed significant effects on target gene expression. . . . . 67
5.4.5 Knockdown of *STAT1* showed no difference in allele expression in HepG2. . . . 67
5.4.6 Overexpression of *STAT1* in a reporter gene assay showed no allele specific effect on gene expression. . . . . 67
5.4.7 STAT1 ChIP shows a binding to the *rs7903146* locus. . . . . 67
5.5 *HMGA2* . . . . . 69
5.5.1 HMGA2 was identified in AC LC-MS/MS to bind 2.9 fold stronger to the T risk allele in Huh7. . . . . 69
5.5.2 Binding of HMGA2 to the T risk allele was confirmed in EMSA and supershift EMSA in several cell lines. . . . . 69

5.5.3 Knockdown of *HMGA2* in 1.1B4 cells showed a significant increase in *ACSL5* expression. . . . 72
5.5.4 HMGA2 binds at the genomic region *rs7903146* in 1.1B4 cells . . . 72
5.6 Other transcription factors of interest . . . 72
5.6.1 Knockdown of a selection of potential regulators revealed promising candidates for further anaylsis. . . . 72
5.6.2 HMGA1 and SATB1 bind T allele specific, while LHX2 binds non allele specific to the region of interest *in vitro*. . . . 73

**6 Discussion** **75**
6.1 Unraveling molecular mechanisms underlying genotype-phenotype associations in type 2 diabetes . . . 75
6.1.1 Identifying potential regulating proteins with affinity chromatography coupled LC MS/MS . . . 76
6.1.2 Investigation of *in vitro* protein/DNA binding behavior via EMSA . . . 76
6.1.3 Transient siRNA knockdown as a tool to capture downstream effects . . . 77
6.1.4 Analysing regulatory effects of transcription factors with gene reporter assays . 78
6.1.5 Chromatin immunoprecipitation as a tool for investigating *in vivo* binding . . . 79
6.1.6 Choice of cell type . . . 79
6.2 The role of PDX1 as a potential regulator . . . 81
6.3 The role of CEBPs as a potential regulators . . . 81
6.4 The role of STAT1 as a potential regulator . . . 83
6.5 The role of HMGA2 as a potential regulator . . . 84
6.6 Other transcription factors as potential regulators . . . 85

**7 Conclusion and outlook** **87**

**Acknowledgements** **109**

**Appendix** **I**

# List of Figures


3.1 Schematic of glucose-stimulated insulin secretion . . . 12
3.2 Schemtaic of the *TCF7L2* gene . . . 16
3.3 Schematic of the enhancer harbouring *rs7903146* . . . 18

5.1 Determination of the best transient knockdown conditions. . . . 44
5.2 Gene expression after IFNγ stimulation in Huh7 and HepG2 cells. . . . 45
5.3 *In vitro* binding of PDX1 to the sequence of interest in EMSA. . . . 48
5.4 Different conditions for binding of PDX1 to the sequence of interest in EMSA. . . . 50
5.5 Different conditions for binding of PDX1 to the sequence of interest in EMSA with cold-labeled oligonucleotides. . . . 52
5.6 *Pdx1* knockdown in Ins1 cells and overexpression of *PDX1* in a gene reporter assay in 1.1B4 cells and HEK293 cells. . . . 53
5.7 *In vitro* binding of CEBPs in EMSA. . . . 56
5.8 *In vitro* interaction of CEBPs in EMSA. . . . 58
5.9 Transient knockdown of different *CEBPs* in SGBS, 1.1B4, HepG2 and Huh7. . . . 60
5.10 *TCF7L2* allele expression after transient *CEBPG* knockdown and gene reporter assays overexpressing *CEBPG* and *CEBPA* in HEK293 and 1.1B4 cells. . . . 62
5.11 *In vitro* binding of STAT1 in HepG2 after 24 h IFNγ treatment. . . . 64
5.12 *TCF7L2* and *ACSL5* expression after transient siRNA *STAT1* knockdown in HepG2 and Huh7 and gene reporter assays investigating effects of *STAT1* on promotor activity. 65
5.13 STAT1 ChIP in HepG2 cells . . . 68
5.14 EMSA with different cell nuclear extracts and supershift EMSA . . . 70
5.15 Supershift EMSA with different cell nuclear extracts. . . . 71
5.16 HMGA2-ChIP and *HMGA2* knockdown in 1.1B4 cells. . . . 72
5.17 Experiments investigating *in vitro* binding of different transcription factors with EMSA, supershift EMSA and competition EMSA. . . . 74

6.1 Schematic workflow of affinity chromatography coupled with LC MS/MS . . . 77
6.2 Schematic overview of the CEBPs. . . . 82
6.3 Overview of identified proteins and their regulatory role in type 2 diabetes . . . 86

A1 Expression analysis of *TCF7L2* and *ACSL5* in different cell lines . . . IX
A2 Western blot of HEK293T nuclear extracts overexpressing a variety of DYK-tagged proteins. . . . IX
A3 Analysis of *PPIB* expression after knockdown. . . . X
A4 Luciferase vector maps . . . XI
A5 EMSA PTPN11 . . . XII
A6 EMSA CEBP interactions. . . . XIII

# List of Tables


2.1 Abbreviations . . . . . 7

4.1 Material . . . . . 21
4.2 Equipment . . . . . 25

5.1 Significant results of the affinity chromatography . . . . . 42
5.2 CEBP Family in affinity chromatography . . . . . 54
5.3 HMGA2 in affinity chromatography 400 mM. . . . . 69

A1 p Values from **Figure 5.2** . . . . . I
A2 EMSA conditions from **Figure 5.14** . . . . . II
A3 Protein size in western blot . . . . . III
A4 Vectors . . . . . IV
A5 Antibodies . . . . . V
A6 Primer . . . . . V

# 2 Abbreviations

**Table 2.1:** Abbreviations

| Abbreviation | Full description |
|---|---|
| **18S** | 18S ribosomal RNA |
| **3C** | Chromosome conformation capture |
| **3T3-L1** | Mouse white pre-adipocyte cell line |
| **4C** | Circularised chromosome conformation capture |
| **5C** | Chromosome conformation capture carbon copy |
| **AC** | Affinity chromatography |
| **ACSL5** | Acyl-CoA synthetase long chain family member 5 |
| **ACTB** | Actin β |
| **AKT** | Protein kinase B |
| **Amp** | Synthetic β-lactamase coding region |
| **AmpliTaq** | Heatstable DNA polymerase modified from Thermus aquaticus |
| **APS** | Ammonium persulfate |
| **ATF1** | Activating Transcription Factor 1 |
| **ATP** | Adenosine triphosphate |
| **B&W** | Binding and washing |
| **BBX** | High Mobility Group Box Domain Containing |
| **bp** | Base pair |
| **BSA** | Bovine serum albumin |
| **bZIP** | Basic Leucine Zipper |
| **cDNA** | Complementary deoxyribonucleic acid |
| **CEBPA** | CCAAT/enhancer-binding protein α |
| **CEBPB** | CCAAT/enhancer-binding protein β |
| **CEBPG** | CCAAT/enhancer-binding protein γ |
| **CEBPZ** | CCAAT/enhancer-binding protein zeta |
| **CHAPS** | 3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate |
| **ChIP** | Chromatin immunoprecipitation |
| **ChIP-seq** | Chromatin immunoprecipitation-sequencing |
| **CREB** | cAMP response element-binding protein |
| **CRISPR** | Clustered regularly interspaced short palindromic repeats |
| **Ct** | Threshold cycle |
| **CtBP** | C-terminal-binding protein |
| **CUX1** | Cut Like Homeobox 1 |
| **CVD** | Cardiovascular disease |
| **Cy5** | Cyanine 5 dye |
| **Da** | Dalton |
| **ddH2O** | Double distilled water |
| **ddNTP** | Dideoxynucleotide |
| **DMEM** | Dulbecco's modified Eagle's medium |
| **dNTP** | Deoxynucleotide mix |

*Continued on next page*

Table 2.1 – *Continued from previous page*

| Abbreviation | Full description |
|---|---|
| **dsDNA** | Double stranded DNA |
| **DTT** | Dithiothreitol |
| **E. coli** | Escherichia coli |
| **EDTA** | Ethylene diamine tetra acetic acid |
| **EGTA** | Ethylene glycol tetraacetic acid |
| **EMSA** | Electrophoretic mobility shift assay |
| **ENCODE** | ENCyclopedia Of DNA Elements |
| **eQTL** | Expression quantitative trait loci |
| **FAIRE-Seq** | Formaldehyde-Assisted Isolation of Regulatory Elements-sequencing |
| **FBS** | Fetal bovine Serum |
| **FISH** | Fluorescence in situ hybridization |
| **Foxo1** | Forkhead box protein O1 |
| **GAPDH** | Glyceraldehyde-3-phosphate dehydrogenase |
| **GCK** | Glucokinase |
| **gDNA** | Genomic DNA |
| **GFP** | Green fluorescent protein |
| **GIP** | Gastric inhibitory polypeptide |
| **GLP-1** | Glucagon-like peptide 1 |
| **GLUT** | Glucose transporter |
| **gRNA** | guide RNA |
| **GSIS** | Glucose-stimulated insulin secretion |
| **GTF** | General transcription factor |
| **GTF2B** | General Transcription Factor IIB |
| **HapMap** | International Haplotype Map project |
| **HDAC** | Histone deacetylases |
| **HDR** | Homologous directed repair |
| **HEPES** | 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid |
| **Hi-C** | Chromosome conformation capture with high through-put sequencing |
| **HLTF** | Helicase-like transcription factor |
| **HMGA1** | High Mobility Group AT-Hook 1 |
| **HMGA2** | High-mobility group AT-Hook 2 |
| **HMGB2** | High mobility group box 2 |
| **HPLC** | High performance liquid chromatography |
| **HPRT1** | Hypoxanthine Phosphoribosyltransferase 1 |
| **IBMX** | 3-isobutyl-1-methylxanthine |
| **IFNγ** | Interferone γ |
| **IGF1** | Insulin growth factor 1 |
| **INS** | Insulin |
| **IRS1** | Insulin receptor substrate 1 |
| **ISL1** | Insulin gene enhancer protein |
| **JAK** | Janus kinase |
| **kb** | Kilobases |
| **KCL** | Potassium chloride |
| **kDa** | Kilo dalton |
| **KLF14** | Krueppel-like factor 14 |
| **LAP** | Liver activating protein |
| **LB medium** | Lysogeny broth medium |

*Continued on next page*

Table 2.1 – *Continued from previous page*

| Abbreviation | Full description |
|---|---|
| **LC-MS/MS** | Liquid chromatography–mass spectrometry/ mass spectrometry |
| **LD** | Linkage disequilibrium |
| **LHX2** | LIM/homeobox protein |
| **LIP** | Liver inhibiting protein |
| **lncRNA** | Long non-conding RNA |
| **MAF** | Minor allele frequency |
| **MAFA** | Mast cell function-associated antigen |
| **MED19** | Mediator Complex Subunit 19 |
| **miRNA** | Micro ribonucleic acid |
| **MYEF2** | Myelin Expression Factor 2 |
| **NE** | Nuclear extract |
| **NGS** | Next-generation sequencing |
| **NHEJ** | Non-homologous end joining |
| **NKRF** | NFKB Repressing Factor |
| **NR2F1** | Nuclear Receptor Subfamily 2 Group F Member 1 |
| **OptiMEM** | Optimal Minimal Essential Medium |
| **OR** | Odds Ratio |
| **PARP1** | Poly(ADP-Ribose) Polymerase 1 |
| **PBMCs** | Peripheral Blood Mononuclear Cells |
| **PBS** | Phosphate buffered medium |
| **PCR** | Polymerase chain reaction |
| **PCSK** | Proprotein convertase |
| **PDX1** | Pancreatic and duodenal homeobox 1 |
| **Pen/Strep** | Penicillin streptomycin |
| **PEPCK** | Phosphoenolpyruvate carboxykinase |
| **PI3K** | Phosphoinositide 3-kinase |
| **PIAS** | Protein inhibitors of activated STAT |
| **PIK3R1** | Phosphatidylinositol 3-kinase regulatory subunit $\alpha$ |
| **PMCA** | Phylogenetic module complexity analysis |
| **PMSF** | Phenylmethanesulfonyl fluoride |
| **PolII** | RNA polymerase II |
| **Poly(dIdc)** | Poly(deoxyinosinic-deoxycytidylic) acid sodium salt |
| **PPARG** | Peroxisome proliferator-activated receptor $\gamma$ |
| **PPIA** | Peptidylprolyl Isomerase A |
| **PPIB** | Peptidylprolyl Isomerase B |
| **PRDM16** | PRD1-BF-1-RIZ1 homologous domain containing protein-16 |
| **pTK6 vector** | Protein-thymidine kinase 6 vector |
| **PTP** | Protein tyrosine phosphatase |
| **PURB** | Purine Rich Element Binding Protein B |
| **RISC** | RNA interference silencing complex |
| **RNA** | Ribonucleic acid |
| **rpm** | Rounds per minute |
| **RPMI medium** | Roswell Park Memorial Institute medium |
| **SATB1** | Special AT-Rich Sequence Binding Protein 1 |
| **SCRT1** | Scratch Family Transcriptional Repressor 1 |
| **SDS** | Sodium dodecyl sulfate |
| **SFPQ** | Splicing Factor Proline And Glutamine Rich |

*Continued on next page*

Table 2.1 – *Continued from previous page*

| Abbreviation | Full description |
|---|---|
| **SGBS** | Simpsom golabi behmel syndrome cell line |
| **SHP** | SH2-containing phosphatase |
| **shRNA** | Short hairpin RNA |
| **siRNA** | Small interfering RNA |
| **SNP** | Single nucleotide polymorphism |
| **SOC medium** | Super optimal broth medium |
| **SUB1** | SUB1 Regulator Of Transcription |
| **T1D** | Type 1 diabetes |
| **T2D** | Type 2 diabetes |
| **TAD** | Topological-associated domain |
| **TAE buffer** | TRIS-acetate-EDTA buffer |
| **TAF10** | TATA-Box Binding Protein Associated Factor 10 |
| **TAF6** | TATA-Box Binding Protein Associated Factor 6 |
| **TBE buffer** | TRIS boric acid EDTA buffer |
| **TCF7L2** | Transcription factor 7-like 2 |
| **TE buffer** | TRIS-EDTA buffer |
| **TEMED** | Tetramethylethylenediamine |
| **TF** | Transcription factor |
| **TFBS** | Transcription factor binding site |
| **TK** | Thymidine kinase |
| **TLE1** | Transducin-like enhancer protein 1 |
| **TRIS** | tris(hydroxymethyl)aminomethane |
| **TSSs** | Transcription start sites |
| **UTR** | Untranslated region |
| **WD repeat** | repeats of tryptophan-aspartic acid |
| **WT** | Wild type |
| **YWHAZ** | Tyrosine 3-Monooxygenase/Tryptophan 5-Monooxygenase Activation Protein Zeta |
| **ZNF593** | Zinc Finger Protein 593 |

# 3 Introduction

Diabetes is a disease that can be documented back thousands of years. Early diagnosis in India was the excess attraction of ants and flies to the urine of the diseased. First treatment was prescribed 1000 AD by Greek physicians consisting of horseback riding. “Diabetes” means “flow”, and refers to the fact that the patients lost more fluid than they could take in. The word “mellitus”, which stands for “honey” indicates the sweet taste of the urine. Until the discovery of insulin in 1922, which later made treatment possible, patients had to go through an ordeal of various approaches, none of which really worked, from bleeding to opium, overfeeding to starvation. In the 1960s, we were finally able to distinguish between type 1 and type 2 diabetes [1]. According to estimates from 2017, the alarming number of adults with diabetes is around 451 million people worldwide. Health care already had a cost burden of 850 billion USD in 2017 due to the disease. Five million deaths worldwide can be attributed to diabetes, and scientists expect the number of people with diabetes to rise to 693 million by 2045. In Germany, the number of patients with diabetes increased to 7.5 million in 2017, making it the country with the second highest rate in Europe and the ninth worldwide [2]. These numbers emphasize the need to find new ways to prevent, treat, and cure the disease. Diabetes mellitus is an umbrella term for several pathophysiological conditions concerning a disturbed glucose metabolism. Roughly, diabetes mellitus is comprised of 90 % type 2 and 10 % type 1 and other rare entities. Type 1 diabetes is a condition, where through an unknown trigger, an autoimmune reaction towards β-cells type 1a is initiated. This condition leads to destruction of these cells and thus the inability to produce and secrete insulin. Type 1 diabetes has some genetic predispositions and commonly develops during childhood. These patients are rarely overweight and their only treatment option is insulin, once the majority of the β-cells have been destroyed. Type 2 diabetes is the result of a long-term deregulated glucose metabolism, mostly through environmental factors such as high caloric intake and low physical activity leading to insulin resistance. A variety of other types of diabetes are known, such as MODY, pancreatitis-induced diabetes, chemical- induced diabetes, which will not be discussed in the course of this thesis [3]. Type 2 diabetes is the most common form and is the result of many risk factors that accumulate over a long period of time. How come T2D is such an abundant disease? Glucose homeostasis is regulated in a complex manner, which will be addressed briefly in the next section.

## 3.1 Pathophysiology of type 2 diabetes

Glucose homeostasis is comprised of a sensitive feedback loop of insulin producing cells in the islets of Langerhans and the insulin sensitive tissues. When blood sugar rises, glucose is transported into the pancreatic β-cell by facilitated diffusion via GLUT2. The consecutive glycolysis and the associated ATP production lead to an inhibition of potassium channels. The diminished potassium influx results in an influx of extracellular calcium ions, which triggers the translocation of insulin containing granules into the membrane. This leads to the secretion of insulin and C-peptide into the blood stream (Figure 3.1). Insulin facilitates the uptake of glucose into the target tissues by signaling the translocation of GLUT4 transporters into the cell membrane. The decrease of blood glucose acts as a

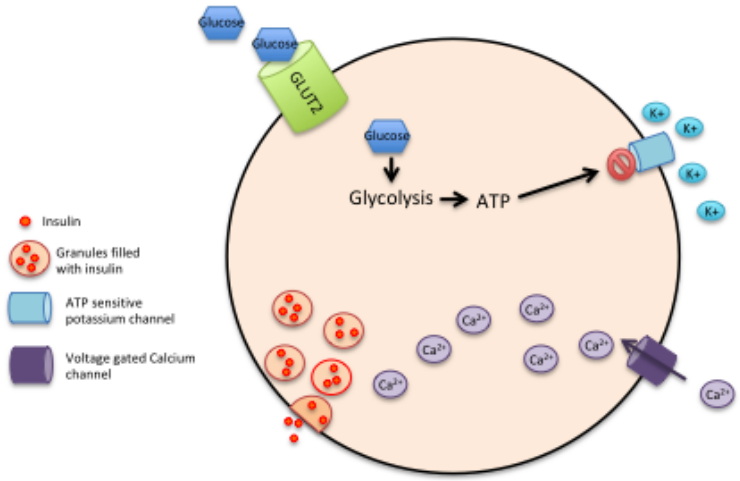


**Figure 3.1:** Schematic review on the mechanism of glucose-stimulated insulin secretion of the β-cell. Figure adapted from [4].

negative feedback to the glucose sensing β-cells. The major regulatory opponent is glucagon. Type 2 diabetes is a combination of both, decreased insulin action and secretion. While the decreased insulin action is found in the main target tissues such as muscle, white adipose tissue and liver, the decreased insulin secretion is caused by defects of the β-cells making makes T2D to a multi-organ disease. Other tissues such as kidney and brain are also involved, but will not be discussed in this thesis.. Insulin is an anabolic hormone, on the one hand upregulating metabolic processes such as hepatic glycogen and lipid synthesis, adipose glucose influx and lipogenesis, muscular glucose utilization and net glycogen synthesis, on the other hand inhibiting hepatic gluconeogenesis and adipose lipolysis. Apart from the metabolic actions, insulin controls mitogenic processes via the ERK pathway [5]. The insulin receptor comprises of two α and two β subunits linked by disulfide bonds. When insulin binds to the α subunit, the β subunit becomes phosphorylated on several tyrosines, allowing it to phosphorylate insulin receptor target proteins. Downstream proteins that can be phosphorylated are Casitas b-lineage Lymphoma gene (CBL), APS, SH2B, GRB2-associated-binding protein 1 and 2 (GAB1, GAB2) and Dedicator of cytokinesis 1 or 2 (DOCK1, DOCK2) [6]. The cascade continues either with activations of the PI3K or the ERK pathways. While the ERK pathway is responsible for controlling cell growth and differentiation together with other partners such as IGF1, the PI3K pathway controls metabolic actions. Insulin signaling has been intensively studied in the major tissues and detailed reviews are available [5, 7, 8]. Wu *et al.* demonstrated a binding of TCF7L2 to the *PIK3R1* promotor, which controls p85, as part of the PI3K/AKT pathway [9].

The most prominent risk factors of T2D are physical inactivity, obesity, excessive caloric intake and smoking, but also biological factors such as age, gender, ethnicity and genetic background (reviewed by Bellou et al., 2018 [10]). In obesity, target tissues have developed an insulin resistance, hence, β-cells have to release more insulin to maintain blood glucose levels. When the β-cells cannot meet this demand, hyperglycemia occurs. The excessive production of insulin leads to an insulinemia and eventually reduction in β-cell mass. Once the β-cells are destroyed, insulin needs to be supplemented externally. This stage of T2D is irreversible. Insulin resistance is derived from many environmental factors (such as high caloric diet, low physical activity, epigenetic programming *in utero*) and genetic factors. Insulin resistance is also caused by inflammation. Notably, in obesity a low-grade inflammation by increased secretion of pro-inflammatory adipokines from adipose tissue is a well investigated field [11, 12]. Genetic variants associated with insulin resistance have been found in genes involved in lipid metabolism such as *PPARG* or *KLF14*, but also directly in the insulin signaling pathway (*IRS1, GCKR*). The variants associated with insulin resistance are summarized in Brown *et al.* [13]. Other regulating mechanisms include GIP and GLP-1 secretion of intestinal L-cells after food intake, bile acids and the brain signal chains. A more extensive review of T2D pathophysiology was published by Kahn *et al.* [14]. Numerous drugs are available for treating T2D. If the bodys' own insulin production is no longer sufficient, insulin must be given. Treatment options range from lifestyle changes (dietary, physical activity) over metformin to sulfonylureas; summarized by Marín-Peñalver *et al.* [15]. Complications of poorly controlled diabetes include microvascular (including retinopathy, nephropathy and neuropathy) and macrovascular diseases (CVDs such as heart attack, stroke, peripheral artery disease etc.), severely reducing life quality and expectancy [16]. We are interested in genetic risk factors, in the hope of eventually finding novel approaches for therapy of T2D. The regulation of glucose metabolism is very complex, thus trying to find the mechanism how gene variants influence the network remains a challenge. In order to find a genotype-phenotype relationship, one has to consider the relevant tissues. Since T2D, as well as the expression of our genes of interest occurs in many tissues, we decided to investigate the major cell types involved in the disease: liver, β-cell, adipose and intestinal cells.

## 3.2 Genetics of type 2 diabetes

Type 2 diabetes has a relatively high heritability. The risk of developing the disease when one parent is affected lies at 40 %, and almost 70 % when both are affected [17, 18]. Meta-analysis of twin studies show a 70 % heritability [19]. There is a variety of protein altering variants in T2D-associated loci e.g. *PPARG* [20], *KCNJ11*, *ABCC8* [21],*GCKR* [22] and *SLC30A8* [23]. The existing genetic markers, however, only make up less than 15 % of the T2D heritability [24]. This missing heritability is likely due to the presence of common variants ($MAF \geq 0.01$). Since classical ways in finding causal variants was limited in the past, a new approach was initiated for the investigation of genetics of common diseases - the genome-wide association studies (GWAS) [25]. Before GWAS, new genotypes were identified via linkage analysis and candidate gene approaches, leading to the discovery of rather rare genetic loci strongly influencing the disease phenotype, explaining only a small portion of the heritability, particularly in common diseases such as T2D [26].

GWAS evolved out of the Humane Genome Project and the International HapMap project [27]. These projects identified many SNPs (Single nucleotide polymorphisms are common substitutions of a single nucleotide in the genome with an occurrence of at least 0.5 % in a population) and their haplotype structure, the prerequisite for subsequent microarray based association studies in disease populations.

According to Sachidanandam *et al.* approximately 1.42 million SNPs are found in the human genome, with only an estimated 60,000 SNPs falling in exonic regions. HapMap genotyped 3.9 million SNPs. While in the beginning, GWAS focused on diabetes versus nondiabetes as categories, later on, groups started looking into glycemic traits. This lead, among others, to the discovery and extensive analysis of *rs10830963* in the locus of *MTNR1B* [28, 29].

### 3.2.1 Non-coding variants associated with type 2 diabetes

Many noncoding risk variants can be found in eQTLs [30] and open chromatin [31], making them valuable loci to be further investigated. Up to today 143 risk variants and putative regulatory mechanisms for type 2 diabetes have been identified by a meta-analysis of GWAS data of T2D samples. Of these, 139 of these are common variants and four are rare [32]. Another research group investigated so called super enhancer SNPs associated with T2D, finding 286 potential functional T2D SNPs. They also used GWAS data, however included T2D tissue specific histone modification ChIP-seq data [33]. The most plausible underlying mechanisms explaining the associations between disease and variant are enhancer-promotor interactions [32]. Molecular mechanisms of SNPs in non-coding regions have been investigated in the field of T2D. The SNP *rs4684847* in the *PPARG* risk locus leads to a binding of PRRX1 - a transcriptional repressor - and thus influences lipid metabolism and systemic insulin sensitivity [34]. The non-risk allele of *rs7647481*, another non-coding SNP in the *PPARG* locus, shows a binding of YY1, which is diminished in the risk allele. These findings support the crucial role of cis-regulatory variants contributing to the disease pathophysiology of T2D [35]. Another example is the *KLF14* locus containing 29 SNPs in LD. These SNPs lead to a reduced expression of *KLF14* in adult adipose tissue. This in turn leads to fewer but larger adipocytes, increased pre-adipocyte proliferation but impaired lipogenesis and furthermore also modulates expression of 385 trans genes [36]. Intronic variants of *KCNQ1* are associated with insulin secretion [37] and a variant near *IRS1* is associated with insulin resistance and hyperinsulinemia [38] emphasizing the relevance of such variants. We hypothesize a similar regulatory mechanism behind the main associated risk locus for T2D: *TCF7L2*.

## 3.3 The T2D risk locus *TCF7L2*

The first genes that have been identified by linkage analysis to T2D were *calpain 10* [39, 40] and *TCF7L2* [41–43]. *TCF7L2*, more precisely a SNP in intron four of this gene [44], initially detected in the Icelandic population, was confirmed since then in a large variety of populations e.g. [41, 45–49]. The non-coding variants associated with T2D are often common variants with a MAF of $> 5$ %. *Rs7903146* risk allele, the lead SNP of the *TCF7L2* locus, has a MAF of 27 %. Odds ratio of this risk variant varies between populations, however a meta-analysis of Ding *et al.* and Lou *et al.* showed overall significant associations between the variant and T2D [50, 51]. Notably, TCF7L2 is involved in approximately 20 % of the T2D cases [52].

### 3.3.1 Function of the TCF7L2 in the WNT signaling pathway

It has been shown, that the variant *rs7903146* influences expression of *TCF7L2* [53]. How does this protein fit into the context of T2D? TCF7L2 is part of the WNT signaling pathway. The WNT signaling pathway was first recognized in developmental [54] and cancer research [55]. Consecutive research showed that it not only is involved in cell proliferation, morphology, fate, motility, organ development and axis formation [56, 57], but also plays a role in metabolic pathways as well, especially

in glucose homeostasis [58,59]. WNT components play a role in proliferation or growth of β-cells [60–62] and glucose-stimulated insulin secretion [63]. Furthermore, WNT inhibits adipogenesis [63–79]. An overview of these mechanisms was summarized by Jin in 2016 [80]. The term "WNT" stands for the wingless-type MMTV integration site, as it was originally called int-1 and is similar to the product of the wingless gene from Drosophila melanogaster [81]. WNT signaling is highly conserved, found in Drosophila and C. elegans up to highly developed species such as mammals [56]. WNT proteins are secreted glycoproteins that can trigger gene expression. There are a variety of WNT proteins (19 in humans) and the frizzled receptors also show ten different subclasses. In general, one can distinguish between the canonical and the non-canonical pathway. The canonical involves β-catenin binding to TCF/LEF, acting as a coactivator, hence inducing gene transcription. In the absence of WNT, TCF/LEF transcription factors bind to WNT response elements, resulting in the recruitment of WNT repressors such as Groucho, CtBP and HDACs to the genomic region [82]. In this state, β-catenin is phosphorylated by CKI α and GSK3 β forming a complex with axin and APC. Once β-catenin is phosphorylated, ubiquitination is initiated, which leads to the degradation of the protein [57]. When WNT binds to the frizzled receptor, β-catenin is transported to the nucleus, binds to TCF/LEF and recruits transcriptional coactivators such as SWI/SNF [82]. Posttranslational modifications regulate which coactivators or repressors or DNA interact with the TCF/β-catenin complex. TCF and LEF proteins are a subgroup of the high mobility group (HMG) box-containing superfamily. TCF binds as a monomer to HMG box, which consists of 80 amino acids and whose core consensus sequence is AGATCAAAGGG. HMG box-containing transcription factors can induce substantial bend of DNA, thereby facilitating the formation of large complexes. TCF proteins cannot modulate transcription themselves, but only through the recruitment of other transcription factors [83–85]. In order to study the function of TCF7L2 more in detail, several research groups performed knockouts or knockdowns. Many attempts to create a homozygous *Tcf7l2*-/- mouse failed, since the knockout becomes lethal within the first 24 h of birth, however heterozygous knockouts mice have been studied in the context of diabetes. Yang *et al.* described that haploinsufficiency has an impact on glucose and lipid homeostasis. A reduction of *Tcf7l2* led to a decrease in blood glucose with a simultaneous increase of glucose tolerance. The authors traced back this effect partially to a decreased hepatic glucose production. They found a significant decrease in glucose-6-phosphatase in both models, the *Tcf7l2*+/- and in primary mouse hepatocytes with a shRNA knockdown of *Tcf7l2*. *Tcf7l2*+/- mice had a protection against high fat diet induced fatty liver, which was explained by reduced levels of factors involved in *de novo* lipogenesis and an increased expression of factors involved in fatty acid oxidation. A reduction of visceral adipose tissue and lower serum triglyceride levels compared to WT mice was also observed. In this mouse model, the authors did not find any effects on pancreatic β-cell function [86]. *Tcf7l2* silencing in the β-cell lines Ins1 and Min6 had an effect on secretory granule fusion [87]. A siRNA mediated knockdown of *TCF7L2* in HepG2 cells increased glucose production and gluconeogenic gene expression [88]. In conclusion, investigations of the function of Tcf7l2 in mouse and rat showed involvement of the protein in liver, adipose and β-cell metabolism.

Patients carrying the *rs7903146* risk allele display a higher hepatic glucose production [89–91]. It has been extensively proven, that liver glucose metabolism plays a major role in the pathogenesis of T2D [92–101].

Oh *et al.* showed in 2012 that TCF7L2 acts on the hepatic glucose metabolism via CREB and FoxO1 transcriptional pathways [98–100]. An RNA-seq and ChIP-seq approach by Norton and co-workers showed that *TCF7L2* not only plays a role in gluconeogenesis, but also in other glucose

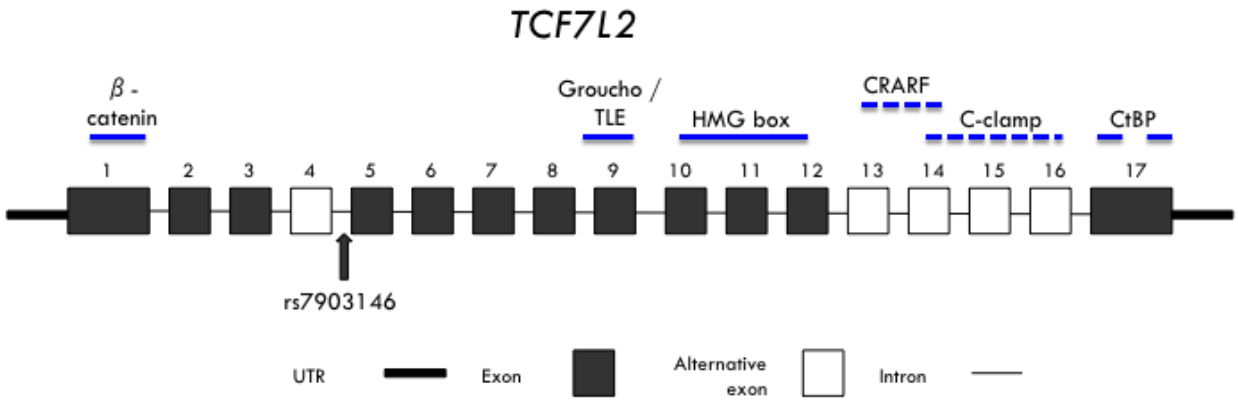


**Figure 3.2:** Scheme of the *TCF7L2* locus with exon and intron structure indicating the location of the risk variant *rs7903146*. The approximate location of potential binding domains are indicated with blue lines. CRARF, C-clamp and CtBP domains are not found in all isoforms. The β-catenin binding domain is not found in dominant negative isoforms of *TCF7L2*.

related pathways in the liver [101]. These findings support the hypothesis that T2D-related regulatory mechanisms of *TCF7L2* are tissue specific. Therefore, several potential target tissues need to be examined to decipher the role of *rs7903146* in the T2D risk phenotype.

### 3.3.2 Splicing of the *TCF7L2* gene locus

The transcription factor 7 like 2 (TCF7L2) gene is located on chromosome 10 (10:112,950,250-113,167,678). It consists of 17 (including alternative ones) exons and 32 splice variants [102, 103]. Duval *et al.* originally described the Human T-Cell Transcription Factor-4 gene in 2000. There it was investigated in the context of colon cancer. They were able to find 17 exons and reported 256 theoretical different splice forms of *TCF7L2*, whereas *in silico* showed over 500 possible mRNA transcripts. *TCF7L2* has a HMG domain in exon 10/11 [103].

Most human protein-coding genes are comprised of small coding exons separated from each other by very large sequences of noncoding introns [104]. This architecture makes it possible to create splice variants of a protein, which have different properties - even opposing roles in metabolism e.g. isoforms of CEBPA and CEBPB [105], but enhancer and lncRNAs are also located in intronic regions. Approximately 95 % of the human genes are spliced [106]. While during intial transcription, the full sequence is copied with UTR and introns (pre-mRNA), the so called spliceosome recognizes short specific conserved splice sites and is formed by five RNAs and more than 200 proteins [107]. Multiple mechanisms are responsible if and how splicing occurs [108]. Intronic mutations can disrupt transcriptional regulatory motifs e.g. enhancers [109] or splicing [110]. *TCF7L2* is abundantly expressed in all tissues, however some of its splice variants seem to be tissue specific. Alternative exon 16 (see **Figure 3.2**) is reported to be specific to pancreas and colon [111], however, others found it only in islet cells [102].

Several TSS exist in the 5' region of the gene, some of which lead to the elimination of the β-catenin binding domain and thus diminish the main function of TCF7L2 as a regulator of the WNT-signaling pathway [111]. The isoforms affecting the C-terminus of TCF7L2 can also potentially affect protein function. Long isoforms (with exons 14, 15 or 16) contain a binding site for the C-terminal binding protein CtBP [112], which is a negative regulator for the WNT signaling pathway [113]. Upon splicing of exon 13 and 14, a new binding domain is formed. This CRARF domain is reported to be a potent activator of the WNT pathway [114, 115]. The TLE binding domain in exon 9 appears to be present in all isoforms, as does the HMG box spanning exon 10-12. The alternative exon 4 is potentially influences the interaction with a Groucho repressor protein [116]. The binding domain of this repressor protein is located in exon 9 next to the TLE domain [117].
In pancreatic islets, four isoforms appear to be predominant. They contain either exon 4 or exon 15, both exons or neither of them [118]. Compared to other tissues, exon 4 is twice as abundant in pancreatic *TCF7L2* isoforms [119].
In adipose tissue, isoforms with retained exons 14 and 15 correlated significantly with obesity. However, this does not apply to the overall *TCF7L2* expression. Insulin suppresses all isoforms in SGBS cells. *TCF7L2* isoforms containing exon 15 seem to play a role in adipocyte differentiation, as they are predominantly expressed at day two of differentiation and then gradually decrease again. The longest isoforms were observed in total pancreas while the shortest were found in adipose and HepG2 cells [102].

It has been noted by Hansson *et al.* [119] that *rs7903146* is in close proximity to exon 4, which has been associated with hemoglobin A1c levels. In HepG2 cells, a protein interaction between HNF1α and specific medium and long isoforms (without exon 13 and the other with exon 15/14) of TCF7L2 has been shown. HNF1α is a transcriptional regulator of glucose-6-phosphatase, a key enzyme of hepatic glucose metabolism [120]. So far, however, none of the alternative splicings has been clearly associated with *rs7903146* [102, 111, 118, 121]. Even though literature shows, how different isoforms can be associated with T2D phenotypes, so far our variant of interest has not shown any influence on isoform expression.

### 3.3.3 Potential regulating mechanisms of *rs7903146*

The variant *rs7903146* lies within intron 4 of the *TCF7L2* gene. Intron 4 contains an enhancer region [122–124]. The hypothesis, that this variant modulates gene expression of nearby genes has been investigated as well. Especially its influence in the pancreatic β-cell has been in focus. The T risk allele of *rs7903146* leads to an impaired insulin secretion, reduced incretin sensitivity on the β-cell [121, 125] and enhanced endogenous glucose production. *TCF7L2* has been found 5 fold higher in islets of T2D patients. Overexpression of *TCF7L2* in human islets reduces glucose-stimulated insulin secretion [121].
Patients homozygous for *rs7903146* and *rs12255372* have a 2.5 fold higher expression of *TCF7L2* in PBMCs. These variants do not affect splicing of*TCF7L2*, which is supported by the fact that they are not located near a splice donor or acceptor site [53]. Interestingly, the research group of Pang *et al.* found in PBMC an equal amount of *TCF7L2* isoforms with exon 1 and 2 and without them, hence equal amounts of isoforms binding and not binding β-catenin. This would mean equimolar amounts of dominant negative and activating TCF7L2 proteins [53].
Gaulton *et al.* conducted FAIRE-seq on human pancreatic islets and found that the *rs7903146* T risk allele has a more open chromatin. They also showed an increased expression from the *rs7903146* T

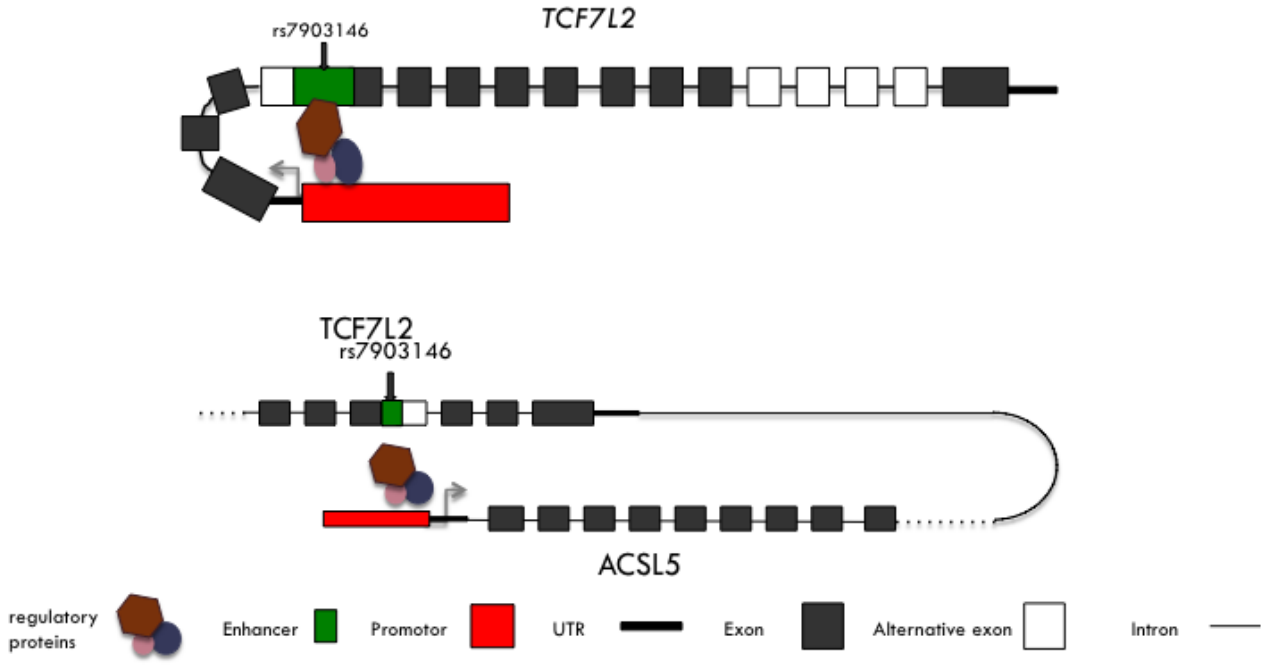


**Figure 3.3:** Schematic of the potential targets of the intronic enhancer harbouring *rs7903146*

allele by a luciferase assay in the established rat beta-cell line MIN6 [126].

One difficulty in studying the gene-regulatory functions of a variant is identifying its target genes. In addition to the neighboring genes, a variant in a certain gene locus may influence the expression of several other genes due to the complexity of the chromatin architecture. Thus, while *rs7903146* itself seems to influence *TCF7L2* expression, other genes should not remain unexplored. The research group around Xia *et al.* asked exactly this question and performed long-range chromatin interaction 4C and Capture C experiments in HCT116 cells with and without a 1.4 kb deletion harboring *rs7903146*. They concluded that *ACSL5*, a gene directly upstream of *TCF7L2*, is also regulated by the intronic region harboring *rs7903146*. Therefore, regulation of both *TCF7L2* and *ACSL5* was taken into account in our analysis. In addition to the regulation of downstream genes, a key issue in genetics is to decipher which gene-regulatory mechanisms are affected by a risk variant. In the case of *rs7903146*, it was identified that PARP-1 binds to the C non-risk allele in HCT116 cells with an oligo pull-down combined with MS. The researchers investigated the interaction of PARP-1 with other highly abundant proteins identified, DNA topoisomerase I and ATP-dependent RNA helicase A and could show, that both dimerized with TCF7L2 and each other [127]. Moreover, Zhou *et al.* identified HMGB1 binding to the locus using Edman-sequencing on HCT116 nuclear extracts and showed that HMGB1 binds prominently to the C non-risk allele by Western blots of human islets. Zhou argues that HMGB1 facilitates the binding of protein complexes and refers to PARP1 [128] binding. These two proteins, however, did not explain a possible role of *rs7903146* in liver or adipose tissue, since most data were generated in a colon cell line and confirmed in β-cells. We wanted to close this information gap by carrying out experiments in two liver cell lines (Huh7 and HepG2), one white adipose cell line (3T3L1) and additionally in two β-cell lines (Ins1 and 1.1B4) in order to find further regulators in this tissue.

## 3.4 Aim of thesis

The elucidation of the regulatory mechanisms of disease-associated SNPs is becoming more and more relevant for modern medicine. In personalized medicine, the genetic make-up of a patient and its impact on the outcome is crucial in order to find therapies with lower costs, higher efficacy and fewer side effects. Studies investigating *rs7903146* in the context of response to treatment already discovered a better response of TT risk carriers to metformin [129]. Another group was able to show that TT risk carriers have earlier treatment failure with sufonylureas compared to non-risk carriers [130]. In view of the increasing number of T2D cases worldwide and the mediocre treatment success so far, there is an urgent need to find new therapeutic approaches to stop the progression from impaired glucose homeostasis to T2D. Basic research is lying the groundwork for future therapies. For almost two decades, *TCF7L2* has been one of the most promising new candidates resolving the missing gap in the heritability of T2D. Associations in different populations are published almost monthly. However the "how" still needs to be elucidated. How does this single nucleotide exchange in an intron of the *TCF7L2* gene alter gene expression towards a risk phenotype? Which regulatory proteins are involved in this process? In which tissues are the strongest effects of this genotype observed? Can these effects be reversed? Hence, the aim of this thesis was to find an allele-specific transcription factor that binds to the *rs7903146* locus and thus regulates expression of *TCF7L2* or *ACSL5*. We investigated the relevance of this effect in the context of a T2D disease phenotype.

# 4 Material and Methods

## 4.1 Material

### 4.1.1 Chemicals and consumables

**Table 4.1:** Material

| Chemical | Company |
|---|---|
| 5x siRNA Buffer | Dharmacon, Lafayette,Colorado, USA |
| Accell siRNA Delivery Media | Dharmacon, Lafayette,Colorado, USA |
| AccuTaq LA DNA Polymerase | Sigma-Aldrich, St. Louis, USA |
| Acrylamid 37.5:1 30% | Carl Roth GmbH, Karlsruhe, Germany |
| Acrylamid 37.5:1 40% | Carl Roth GmbH, Karlsruhe, Germany |
| Actinomycin D | VWR, Darmstadt, Germany |
| Agarose | Carl Roth GmbH, Karlsruhe, Germany |
| alamarBlue Cell Viability Reagent | Thermo Fisher Scientific, Waltham, USA |
| Albumin Fraction V for western blotting | Applichem, Darmstadt, Germany |
| Ammonium persulfate | Bio-Rad Laboratories, Munich, Germany |
| Ammoniumacetate | Thermo Fisher Scientific, Waltham, USA |
| Ampicillin | Sigma-Aldrich, St. Louis, USA |
| AmpliTaq Gold with Gold Buffer | Thermo Fisher Scientific, Waltham, USA |
| AmpliTaq polymerase | Thermo Fisher Scientific, Waltham, USA |
| Antarctic Phosphatase | NEB, Ipswich, USA |
| ATP | PJK Biotech GmbH, Kleinblittersdorf, Germany |
| Betaine | Carl Roth GmbH, Karlsruhe, Germany |
| Bio-Rad Protein Assay | Bio-Rad Laboratories, Munich, Germany |
| Biotin (suitable for cell culture) | Sigma-Aldrich, St. Louis, USA |
| Boric acid | VWR, Darmstadt, Germany |
| Bovine serum albumin (fraction V) | AppliChem, Darmstadt, Germany |
| Bromphenolblue | VWR, Darmstadt, Germany |
| Cell Line Nucleofector Kit R | Lonza Group,Basel Switzerland |
| Cell Line Nucleofector Kit T | Lonza Group, Basel, Switzerland |
| Cell Line Nucleofector Kit V | Lonza Group, Basel, Switzerland |

*Continued on next page*

Table 4.1 – *Continued from previous page*

| Chemical | Company |
|---|---|
| CHAPS Bufferan | Carl Roth GmbH, Karlsruhe, Germany |
| ChIP-IT Control qPCR Kit-Human | Active Motif,La Hulpe, Belgium |
| ChIP-IT Express Enzymatic Kit | Active Motif,La Hulpe, Belgium |
| Chloroform | Applichem, Darmstadt, Germany |
| Coelenterazine (native-CTZ) | PJK Biotech GmbH, Kleinblittersdorf, Germany |
| Coenzyme A | PJK Biotech GmbH, Kleinblittersdorf, Germany |
| cOmplete mini tablets, EDTA free | Roche Pharma AG, Grenzach-Whylen, Germany |
| Coomassie | Carl Roth GmbH, Karlsruhe, Germany |
| Cortisol | Sigma-Aldrich, St. Louis, USA |
| D-Luciferin | PJK Biotech GmbH, Kleinblittersdorf, Germany |
| Biotin | Carl Roth GmbH, Karlsruhe, Germany |
| Dexamethasone | Sigma-Aldrich, St. Louis, USA |
| DharmaFECT 1 Transfection Reagent | Dharmacon, Lafayette,Colorado, USA |
| Dimethylsulfoxid | VWR, Darmstadt, Germany |
| DMEM low Glucose | Sigma-Aldrich, St. Louis, USA |
| DMEM- high glucose | Sigma-Aldrich, St. Louis, USA |
| DMEM/F-12 | Invitrogen, Darmstadt, Germany |
| DNAse | Qiagen, Hilden, Germany |
| dNTPs | Carl Roth GmbH, Karlsruhe, Germany |
| DTT | AppliChem, Darmstadt, Germany |
| Dulbecco‘s Phosphate Buffered Saline | Sigma-Aldrich, St. Louis, USA |
| EDTA | Carl Roth GmbH, Karlsruhe, Germany |
| EGTA | Merck, Darmstadt, Germany |
| Ethanol absolut | VWR, Darmstadt, Germany |
| Ethidium bromide | AppliChem, Darmstadt, Germany |
| Exonuclease I | NEB, Ipswich, USA |
| Fetal bovine serum Superior | Biochrom AG Biotechnologie Berlin, Germany |
| Ficoll | Sigma-Aldrich, St. Louis, USA |
| Formaldehyd 37% | Sigma-Aldrich, St. Louis, USA |
| Fugene HD | Promega, Madison, USA |
| Gene Ruler Mix | Thermo Fisher Scientific, Waltham, USA |
| Gene Ruler Ultra Low Range DNA Ladder | Thermo Fisher Scientific, Waltham, USA |
| Gentamicin | Sigma-Aldrich, St. Louis, USA |
| Glacial acetic acid | VWR, Darmstadt, Germany |
| Glutamine | Carl Roth GmbH, Karlsruhe, Germany |

*Continued on next page*

Table 4.1 – *Continued from previous page*

| Chemical | Company |
|---|---|
| Glycerol | Carl Roth GmbH, Karlsruhe, Germany |
| Glycine | AppliChem, Darmstadt, Germany |
| HEPES | Carl Roth GmbH, Karlsruhe, Germany |
| HEPES 1M | Thermo Fisher Scientific, Waltham, USA |
| HiFi Proof-Reading Polymerase | Nippon Genetics Europe, Germany |
| HiPerFect Transfection Reagent | Qiagen, Hilden, Germany |
| HPLC Water | Carl Roth GmbH, Karlsruhe, Germany |
| Human apo Transferrin | Sigma-Aldrich, St. Louis, USA |
| Human Insulin | Sigma-Aldrich, St. Louis, USA |
| IBMX | Sigma-Aldrich, St. Louis, USA |
| IFN-γ human | Sigma-Aldrich, St. Louis, USA |
| Isol-RNA Lysis Reagent | 5 Prime, Hilden, Germany |
| Isopropanol | Carl Roth GmbH, Karlsruhe, Germany |
| KCl | Carl Roth GmbH, Karlsruhe, Germany |
| LB Agar | Carl Roth GmbH, Karlsruhe, Germany |
| LB Medium | Carl Roth GmbH, Karlsruhe, Germany |
| Lipofectamin 2000 Transfection Reagent | Thermo Fisher Scientific, Waltham, USA |
| LookOut Mycoplasma PCR Detection Kit | Sigma-Aldrich, St. Louis, USA |
| Magnesium chloride | VWR, Darmstadt, Germany |
| Minimum Essential Medium Eagle | Sigma-Aldrich, St. Louis, USA |
| Molecular Grade RNase-free water | Dharmacon, Lafayette,Colorado, USA |
| MyOne Streptavidin C1 | Thermo Fisher Scientific, Waltham, USA |
| MyTaq DNA Polymerase | Bioline, Taunton, USA |
| Na-Pyruvate | Thermo Fisher Scientific, Waltham, USA |
| NaN3 | Carl Roth GmbH, Karlsruhe, Germany |
| NEB 5-α Competent E. coli | NEB, Ipswich, USA |
| Non essentiell amino acids 100x | Sigma-Aldrich, St. Louis, USA |
| Nonidet P 40 Substitute | Carl Roth GmbH, Karlsruhe, Germany |
| OptiMEM | Thermo Fisher Scientific, Waltham, USA |
| Orange G | Carl Roth GmbH, Karlsruhe, Germany |
| PAGE ruler prestained | Thermo Fisher Scientific, Waltham, USA |
| Panthotenate | Sigma-Aldrich, St. Louis, USA |
| Passive Lysis Buffer | Promega, Madison, USA |
| PCR Clean-up Kit | Promega, Madison, USA |

*Continued on next page*

Table 4.1 – *Continued from previous page*

| Chemical | Company |
|---|---|
| Penicillin/streptomycin | Sigma-Aldrich, St. Louis, USA |
| Phosphatase inhibitor cocktail | Roche Pharma AG, Grenzach-Whylen, Germany |
| PMSF | Carl Roth GmbH, Karlsruhe, Germany |
| Poly(dIdc) | Sigma-Aldrich, St. Louis, USA |
| Potassiumacetate | Carl Roth GmbH, Karlsruhe, Germany |
| Primer | TIB Molbiol, Berlin, Germany |
| Primer | Sigma-Aldrich, St. Louis, USA |
| Protease | Qiagen, Hilden, Germany |
| Protease inhibitor cocktail | Roche Pharma AG, Grenzach-Whylen, Germany |
| PureYield Plasmid Maxiprep System | Promega, Madison, USA |
| QIAmp DNA Mini Kit | Qiagen, Hilden, Germany |
| QIAshredder | Qiagen, Hilden, Germany |
| qPCRBIO SyGreen Mix Lo-ROX | Nippon Genetics Europe, Germany |
| QuantiTect Reverse Transcription Kit | Qiagen, Hilden, Germany |
| RNA isolation Kit 96 well | Qiagen, Hilden, Germany |
| RNase A | Qiagen, Hilden, Germany |
| RNeasy Mini Kit | Qiagen, Hilden, Germany |
| Rosiglitazone | Biomol GmbH, Hamburg, Germany |
| Roti Phenol Chloroform | Carl Roth GmbH, Karlsruhe, Germany |
| RPMI -1640 | Sigma-Aldrich, St. Louis, USA |
| SDS Pellets | Carl Roth GmbH, Karlsruhe, Germany |
| SensiFast Probe No-Rox Kit | Biocat, Heidelberg, Germany |
| SOC Medium | TAKARA BIO INC., Japan |
| Sodium hydroxide | VWR, Darmstadt, Germany |
| Stella Competent Cells | TAKARA BIO INC., Japan |
| Sucrose | AppliChem, Darmstadt, Germany |
| T4 DNA Ligase | Thermo Fisher Scientific, Waltham, USA |
| T4 PNK | NEB, Ipswich, USA |
| TA Cloning Kit, with pCR2.1 Vector, without competent cells | Thermo Fisher Scientific, Waltham, USA |
| TEMED | Bio-Rad Laboratories, Munich, Germany |
| Tricine | Carl Roth GmbH, Karlsruhe, Germany |
| Triiodothyronine | Sigma-Aldrich, St. Louis, USA |
| Tris | VWR, Darmstadt, Germany |
| TritonX-100 | Sigma-Aldrich, St. Louis, USA |
| Trypan Blue Solution, 0.4 % | Thermo Fisher Scientific, Waltham, USA |

*Continued on next page*

Table 4.1 – *Continued from previous page*

| Chemical | Company |
|---|---|
| Trypsin-EDTA | Biochrom AG Biotechnologie Berlin, Germany |
| Trypsin-EDTA | Sigma-Aldrich, St. Louis, USA |
| Tween 20 | Sigma-Aldrich, St. Louis, USA |
| UNIVERSAL PROBE LIBRARY | Sigma-Aldrich, St. Louis, USA |
| Wizard SV Gel and PCR Clean-Up | Promega, Madison, USA |
| Xylencyanol | Carl Roth GmbH, Karlsruhe, Germany |
| β-Mercaptoethanol | Merck, Darmstadt, Germany |

**Table 4.2:** Equipment

| Equipment | Company |
|---|---|
| 384-well plate | 4titude, Surrey, UK |
| 96-well flat bottom plate | Carl Roth GmbH, Karlsruhe, Germany |
| 96-well plate (4ti-0750) PCR | 4titude, Surrey, UK |
| 96-well plate (4ti-0770) Cycle Sequencing | 4titude, Surrey, UK |
| Autoclave | Wolf, Bad Ueberkingen, Germany |
| Berthold Detection System | SIRIUSBerthold Technologies GmbH& Co KG, Bad, Germany |
| Cannula 24´G | B. Braun, Melsungen, Germany |
| Cell culture flasks (25 $cm^2$, 75 $cm^2$, 175 $cm^2$),TC Dish 150 mm | Sarstedt AG & Co., Nuembrecht, Germany |
| Centrifuge 5417 C | Eppendorf AG, Hamburg, Germany |
| Centrifuge 5430 | Eppendorf AG, Hamburg, Germany |
| Centrifuge 5810/ 5415R | Eppendorf AG, Hamburg, Germany |
| Centrifuge A14 | Societe Jouan, Saint-Herblain, France |
| Centrifuge Rotina 420R | Andreas Hettich GmbH & Co. KG, Tuttlingen, Germany |
| Centrifuge X3R | Hereaus Holding, Hanau, Germany |
| Combitips for multipette | Eppendorf AG, Hamburg, Germany |
| DNA LoBind Tubes | Eppendorf AG, Hamburg, Germany |
| Electronic multi-channel pipette 30 µl | Thermo Scientific, Munich, Germany |
| Electrophoresis unit (Multiphore II) | GE Healthcare, Solingen, Germany |
| Electrophoresis, agarose- Gel | PEQLAB Biotechnologie, Erlangen, Germany |
| EMSA gel chamber | Biometra, Goettingen, Germany |
| EpiShear Probe Sonicator | Active Motif,La Hulpe, Belgium |
| Falcon flasks 175 $cm^2$ | VWR, Darmstadt, Germany |
| Falcon flasks 25 $cm^2$ | VWR, Darmstadt, Germany |
| Fluorescence microscope (DMI 4000 B) | Leica Biosystems, Wetzlar, Germany |
| GelBond PAG Film | Lonza Group,Basel Switzerland |
| Heating block | Eppendorf AG, Hamburg, Germany |

*Continued on next page*

Table 4.2 – *Continued from previous page*

| **Equipment** | **Company** |
|---|---|
| Horizontal laboratory shaker, HS 501 | IKA, Staufen, Germany |
| Incubator (37 °C) | Societe Jouan, Saint-Herblain, France |
| Incubator (E. coli) | Societe Jouan, Saint-Herblain, France |
| Incubator Heracell (cell culture) | Thermo Fisher Scientific, Waltham, USA |
| Inoculation tubes | Sarstedt AG & Co., Nuembrecht, Germany |
| Laminar flow hood Herasafe | Thermo Fisher Scientific, Waltham, USA |
| LightCycler 480 | Roche Pharma AG, Grenzach-Whylen, Germany |
| Luminometer tubes | VWR, Darmstadt, Germany |
| Magna-SepTM Magnetic Particle Separator | Invitrogen, Carlsbad, CA, USA |
| Magnetic stirrer with heating plate | IKA, Staufen, Germany |
| Nanodrop | Thermo Fisher Scientific, Waltham, USA |
| Neubauer counting chamber (0.0025 mm2) | Brand GmbH&Co. KG,Wertheim, Germany |
| Nitrocellulose membrane | GE Healthcare Europe, Freiburg, Germany |
| Odyssey Infrared Imager | LI-COR Biosciences-GmbH, Germany |
| Pasteurpipettes | VWR, Darmstadt, Germany |
| pH meter | Mettler-Toledo, Inc., OH, USA |
| Power supply (Powerpac HC) | Bio-Rad Laboratories, Hercules, USA |
| Reaction tubes (1.5, 2.0 ml) | Sarstedt AG & Co., Nuembrecht, Germany |
| Scales ATILON, max. 150 g | Acculab, Goettigen, Germany |
| Shaker cabinet, heatable | Edmund Buehler, Tuebingen, Germany |
| Sorvall Evolution RC Centrifuge | Thermo Fisher Scientific, Waltham, USA |
| Sterile pipettes (5, 10, 25 ml) | Sarstedt AG & Co., Nuembrecht, Germany |
| Syringes (1 ml, 5 ml, 20ml) | B. Braun, Melsungen, Germany |
| TECAN Infinite 200 NanoQuant | Tecan Group Ltd. Maennedorf, Switzerland |
| Thermo mixer, heatable | Eppendorf AG, Hamburg, Germany |
| Thermocycler (TProfessional) | Biometra, Goettingen, Germany |
| Typhoon TRIO+ Variable Mode Imager | Amersham Biosciences Europe GmbH, Freiburg, Germany |
| U-100 Insulin 30Gx1/2” | Braun, Bad Arolsen, Germany |
| Vacuum centrifuge | Societe Jouan, Saint-Herblain, France |
| Varioskan Flash, Multimode Reader | Thermo Fisher Scientific, Waltham, USA |
| Water bath | JULABO GmbH, Seelbach, Germany |
| Whatman paper | A. Hartenstein, Wuerzburg, Germany |

### 4.1.2 Buffers and prepared solutions

**Cell Culture**

Pan/Bio
3.3 mM biotin, 1.7 mM panthotenate in DMEM/F-12, sterile filtered
Growth medium for 1.1B4
RPMI-1640, 10 % FBS, 1 % P/S
Growth medium for Caco2
MEM, 10 % FBS, 1 % NEAA, 0.2 % Gentamicin
Growth medium for high glucose for HEK, Huh7, 3T3L1, HepG2
DMEM high glucose, 10 % FBS, 1 % P/S
Growth medium for HEK, Huh7, 3T3L1, HepG2
DMEM low glucose, 10 % FBS, 1 % NEAA, 1 % P/S
Induction medium for 3T3L1
DMEM high glucose, 10 % FBS, 1 % P/S, 5 µg/ml insulin, 0.25 µM dexamethasone, 0.5 mM IBMX
Differentiation medium for 3T3L1
DMEM high glucose, 10 % FBS, 1 % P/S, 5 µg/ml insulin
Growth medium for HIB1B
DMEM/F-12 (1:1), 10 % FBS, 16 µM Biotin, 0.1 % Gentamicin
Differentiation medium for HIB1B
DMEM/F-12 (1:1), 7 % FBS, 16 µM Biotin, 17 nM Insulin, 0.1 % Gentamicin
Growth medium for HT-29
RPMI-1640, 10 % FBS, 1 % Glutamine, 1 % P/S
Growth medium for Ins-1
RPMI-1640, 10 % FBS, 50 µM β-Mercaptoethanol, 1 % P/S
Growth medium for NCI H716
RPMI-1640, 10 % FBS, 2.5 g/L Glucose, 10 mM HEPES, 1 mM NaPyruvat, 1 % P/S
Growth medium for SGBS
DMEM/F-12 (1:1), 10 % FBS, 1 %Pan/Bio, 1 % P/S
Induction medium for SGBS
DMEM/F-12 (1:1), 0.33 mM biotin, 0.17 mM panthotenate, 1 %P/S, 10 µg/µl transferrin, 20 nM insulin, 100 nM cortisol, 0.2 nM T3, 25 nM dexamethasone, 250 µM IBMX, 2 µM rosiglitazone
Differentiation medium for SGBS
DMEM/F-12 (1:1), 0.33 mM biotin, 0.17 mM panthotenate, 1 %P/S, 10 µg/µl transferrin, 20 nM insulin, 100 nM cortisol, 0.2 nM T3
Growth medium for Jurkat
RPMI-1640, 10 % FBS h.i., 50 µM β-Mercaptoethanol, 1 % P/S

**EMSA and affinity chromatography**

10xLoading Buffer
250 mM Tris-HCl (pH 7.5), 40 % (v/v) glycerol, 0,2 % Orange G
2xBind and Wash
10 mM Tris-HCl (pH 7.5), 2 M NaCl, 1 mM EDTA
5x Gel Binding Buffer for EMSA
20 % glycerol, 5 mM MgCl2, 2.5 mM EDTA, 2.5 mM DTT, 250 mM NaCl, 50 mM Tris-HCl (pH 7.5)

5x GBB Ficoll Potassium for EMSA
20 % Ficoll, 5 mM EDTA, 5 mM DTT, 1 M KCl, 125 mM HEPES (pH 7.5)
5x Gel Binding Buffer Potassium for EMSA
20 % glycerol, 5 mM MgCl2, 2.5 mM EDTA, 2.5 mM DTT, 1 M KCl, 50 mM Tris-HCl (pH 7.5)
5x GBB w/o NaCl for AC
20 % glycerol, 5 mM MgCl2, 2.5 mM EDTA, 2.5 mM DTT, 50 mM Tris-HCl (pH 7.5)
5x TBE, pH 8.3
10 mM EDTA (pH 8.5), 445 mM tris base, 445 mM boric acid
Elution Buffers Affinity Chromatography
1x GBB for AC with NaCl concentrations ranging from 50 mM to 1 M
Nuclear extraction buffers
Buffer A: 10 mM HEPES (pH 8), 10 mM KCl, 0.1 mM EDTA, 0.1 mM EGTA, 0.5 mM PMSF, 1 mM DTT, Protease inhibitor 7x, Phosphatase inhibitor 7x
Buffer C: 20 mM HEPES (pH 8), 400 mM NaCl, 20 % glycerol (v/v), 1 mM EDTA, 1 mM EGTA, 1 mM PMSF, 1 mM DTT, Protease inhibitor 7x, Phosphatase inhibitor 7x
Nuclear extract for Affinity Chromatography
1x BB for AC, 0.045 ‰ CHAPS, x mg nuclear protein
Polyacrylamide Gel
4 %: 0.5x TBE, 4 ml acrylamide/bisacrylamide 37.5:1 (40 % w/v), 2.5 % glycerol (v/v), 300 µl APS (10 %), 20 µl TEMED
5.3 %: 0.5x TBE, 5.3 ml acrylamide/bisacrylamide 37.5:1 (40 % w/v), 2.5 % glycerol (v/v), 300 µl APS (10 %), 20 µl TEMED
12 %: 0.5x TBE, 12 ml acrylamide/bisacrylamide 37.5:1 (40 % w/v), 2.5 % glycerol (v/v), 300 µl APS (10 %), 20 µl TEMED
TE Buffer
10 mM Tris HCl, 1 mM disodium, pH 7.6
Washing Buffer Affinity Chromatography
1x GBB for AC, 10 mM NaCl

**PCR and vector cloning**

50x TAE Buffer
2 M TRIS, 64 mM EDTA, 5.7 % (v/v) acetic acid
Agarose gels
1x TAE, x % agarose, 0.05 ‰ ethidium bromide, alternatively 5 µl HD Green
Ampicillin Solution
100 mg/ml Ampicillin in 50 % Ethanol
dNTPs
10 µM dGTP, 10 µM dCTP, 10 µM dATP, 10 µM dTTP
LB-Agar + Amp (sterilized by autoclaving)
35 g LB-Agar solved in 1 l ddH2O with 0.1 mg/ml Ampicillin
LB-Medium (sterilized by autoclaving)
20 g LB-Medium solved in 1 l ddH2O
Miniprep solution 1
100 mM TRIS, 10 mM EDTA, 7 U/ml RNase A

Miniprep solution 2
200 mM NaOH, 1 % (v/v) SDS
Miniprep solution 3
3 M potassium acetate, 11.5 % (v/v) acetic acid
SOC medium
20 g/l Peptone, 5 g/l Yeast extract, 8.6 mM NaCl, 2.5 mM KCl. After autoclaving, sterile filtered 20 mM glucose and 10 mM MgCl2 were added.

**Western blot**

RIPA Buffer
10 mM Tris HCl, 150 mM NaCl, 5 mM EDTA, 1 % Triton X, 1x Protease inhibitor
10x PBS
1,37 M NaCl, 26.8 mM KCl, 100 mM Na2HPO4, 17.6 mM KH2PO4, pH 7.4
4x Laemmli
277 mM SDS; 75 mM TRIS, 20 % (v/v) glycerol, 20 % (v/v) β-mercaptoethanol, 0.4 % (w/v) bromphenol blue
Separation Gel
400 mM Tris (pH 8.8), 0.1 % SDS Acrylamid 37.5:1 (30 %) according to desired percentage of Gel, 50 µl APS (10 %), 2.5 µl TEMED ad 4.5 ml
Stacking Gel
139 mM Tris (pH 6.8), 0.11 % SDS, 2 % Acrylamid 37.5:1 (30 %), 12 µl APS (10 %), 1.5 µl TEMED ad 2 ml
5x Running Buffer
125 mM TRIS, 960 mM glycine, 17.3 mM SDS
2x Blotting Buffer
150 mM glycine, 19.5 mM TRIS, 20 % (v/v) methanol, 0.02 % (v/v) SDS
Blocking Buffer
5 % BSA in 1x PBS
PBS-T
1xPBS with 0.5 % TWEEN 20

**Luciferase Assay**

Firefly Luciferase Substrate Buffer
470 µM D-Luciferin, 530 µM ATP, 270 µM CoenzymeA, 33.3 mM DTT, 20 mM Tricine, 2.67 mM MgSO4 Mg(OH)2 *5H2O, 0.1 mM EDTA
Renilla Luciferase Substrate Buffer
1.43 µM Coelenterazine, 2.2 mM Na2EDTA, 0.22 M KxPO4 (pH 5.1), 0.44 mg/ml BSA, 1.1 M NaCl, 1.3 mM NaN3

## 4.2 Methods

### 4.2.1 Cell culture

All cells were regularly tested for mycoplasma infection with the LookOut Mycoplasma PCR Detection Kit. Huh7 [131], HEK 293T [132], HEK 293, HepG2 [133], Ins-1 [134], Caco-2 [135] and 1.1B4 [136] cells were cultured in their growth medium, with the following conditions: 37 °C with 5 % $CO_2$ in a T75 polystyrene flask filled with 12 ml medium. The cells were sub-cultured twice a week (until they reached 80-90 % confluency). Sub-culturing was performed by washing with 0.1 ml/cm$^2$ PBS and a trypsinisation with approximately 0.02 ml/cm$^2$ of trypsin for 5 min at 37 °C and 5 % $CO_2$. In order to produce large quantities of nuclear protein, cells were also transferred and cultured in 150 mm dishes.

**SGBS cell line**

SGBS cells are a human Simpson-Golabi-Behmel syndrome (SGBS) preadipocyte cell strain that is neither transformed nor immortalized, originating from a patient with SGBS [137]. SGBS cells were cultured in growth medium. The conditions were 37 °C with 5 % $CO_2$ in a T175 polystyrene flask filled with 25 ml medium. The cells were sub-cultured once a week (until they reached ca. 80-90 % confluency). Sub-culturing was performed by washing with 0.1 ml/cm$^2$ PBS and a trypsinisation with ca. 0.02 ml/cm$^2$ of trypsin for ca. 5 min at 37 °C and 5 % $CO_2$. The trypsinisation was stopped with growth medium and centrifuged at 200 g for 10 min. Cells were resuspended in growth medium and seeded in T25 (120,000 cells) or T175 (500,000 cells). For differentiation, cells were seeded in a density of 150,000 cells per 6 well to reach a confluency of 99 % on the day of induction, 3-4 days later. Cells were carefully washed once with 6 ml PBS per 6well and 2 ml induction medium was added. After 4 days medium was changed to 2 ml differentiation medium per well. Cells were harvested at different time points throughout differentiation up to d14.

**NCI H716 cell line**

NCI H716 are derived from an L-cell colorectal adenocarcinoma of a 33-year-old Caucasian male [138]. NCI H716 cells were cultured in their growth medium, with the conditions were 37 °C with 5 % $CO_2$ in a T25 polystyrene flask filled with 8 ml medium. Once a week they were subcultured. Therefore they were counted and cell viability was checked with trypan blue staining. They were then centrifuged at 200 g for 5 min and resuspended in a correct amount of new growth medium leading to a final cell count of 900,000 cells/ml.

**Jurkat cell line**

Jurkat cells [139] were cultured in their growth medium, at 37 °C with 5 % $CO_2$ in a T75 polystyrene flask filled with 20 ml medium. Twice a week they were subcultured. Therefore they were counted and cell viability was checked with trypan blue staining. They were then centrifuged at 200 g for 5 min and resuspended in a correct amount of new growth medium leading to a final cell count of 100,000 cells/ml.

**3T3L1 cell line**

3T3 L1 is a continuous strain of 3T3 developed through clonal isolation. 3T3L1 cells were cultured in their growth medium, with the conditions were 37 °C with 5 % $CO_2$ in a T75 polystyrene flask filled with 8 ml medium. The cells were sub-cultured twice a week, when they reached 80-90 % confluency. Sub-culturing was performed with a washing step with 0.1 ml/cm$^2$ PBS and a trypsinisation with approximately 0.02 ml/cm$^2$ of trypsin for ca. 5 min at 37 °C and 5 % $CO_2$. Approximately 1 million cells were transferred into a new T75 flask. In order to produce large quantities of nuclear protein, cells were also transferred and cultured in 150 mm dishes. For differentiation into adipocytes, cells were

plated in 6 well plates and induction medium was added when the cells reached 90 % confluency. On day four the induction medium was exchanged for differentiation medium. This medium was changed every three days until the cells were fully differentiated.

**HIB 1B cell line**

HIB 1B are derived from a brown fat tumor of a transgenic mouse [140]. HIB 1B cells were cultured in their growth medium, with the conditions were 37 °C with 5 % $CO_2$ in a T75 polystyrene flask filled with 8 ml medium. The cells were sub-cultured twice a week, when they reached 80-90 % confluency. Sub-culturing was performed with a washing step with 0.1 ml/cm$^2$ PBS and a trypsinisation with ca. 0.02 ml/cm$^2$ of trypsin for ca. 5 min at 37 °C and 5 % $CO_2$. Approximately 2 million cells were transferred into a new T75 flask. In order to produce large quantities of nuclear protein, cells were also transferred and cultured in 150 mm dishes. For differentiation into brown adipocytes, cells were plated in 6 well plates and induction medium was added as soon as the cells reached 80 % confluency. Differentiation medium was added and changed every two days until differentiation was achieved.

### 4.2.2 Whole protein harvest

Cells were plated in 6 well format and grown until 80-90 % confluency. RIPA buffer was prepared on ice and the cells were washed twice with 1 ml cold PBS. Cells were scraped in 0.5 ml cold PBS and transferred into a 1.5 ml tube. The wells were washed with an additional 0.5 ml cold PBS and transferred into the tube. After a centrifugation step (4 °C, 2 min, 2000 g) the supernatant was discarded. Thirty µl RIPA buffer per harvested 6 well was added to the tube, mixed, resuspended and then pipetted up and down in a 24'G syringe 15 times. This was centrifuged (4 °C, 3 min, 800 g) and the supernatant was recovered into a new tube. This step was repeated with 45 µl RIPA buffer per tube. The protein was frozen in -80 °C for further use.

### 4.2.3 Nuclear extraction

Nuclear extraction was performed with two buffers: Buffer A and Buffer C adapted from the original publication of Schreiber *et al.* [141]. The whole procedure was kept on ice (4 °C). After washing the cells once with PBS, scraping in 0.5 ml Buffer A/150 mm petri dish was performed. Another 0.5 ml Buffer A was added to the tube, resuspended and incubated 30 min. After an addition of Nonidet NP-40 final concentration of 0.3 % the tubes were mixed vigorously for 10 s and centrifuged at 21000 g for 5 min. Supernatant was collected as a cytoplasmic fraction control. The pellets were washed twice with 1 ml Buffer A with a centrifugation step in between (21000 g, 5 min). Buffer C in twice the amount of the pellet volume was added and incubated for 15-30 min on a shaker. After another centrifugation step (21000 g, 5 min) the supernatant containing the nuclear proteins could be recovered. Nuclear protein concentrations were measured with Bradford Protein Assay according to manufacturers' instructions. The standard curves as well as the samples were measured in triplicates. Measurements were taken at 595 nm. Aliquots of nuclear protein were shock-frozen in liquid nitrogen and stored in -80 °C until further use.

### 4.2.4 Electrophoretic mobility shift assay (EMSA)

Electrophoretic mobility shift assay visualizes protein/DNA interactions. Double stranded oligonucleotides containing a variant of interest are incubated with nuclear protein extracts and subsequently loaded on a polyacrylamid gel for electrophoresis. The protein/DNA binding leads to a slower migra-

tion in the electric field and is being visualized, in this case, by fluorescent dye [142].

**Annealing**

EMSA was performed with Cy5-labeled oligonucleotide probes; oligonucleotides were selected as adjacent regions of SNP of interest of the major as well as minor variant. The sequences are listed in the appendix (**Table A6**). Primers were diluted with $ddH_2O$ to 0.1 mM according to manufacturers' instructions. Preparation of the oligonucleotides was performed accordingly: Cy5-labeled forward strands were annealed with non-labeled reverse strands at 85 °C (depending on Tm of the primers) for 5 min with a slow overnight cool-down until 10 °C with 7 µl forward primer, 7 µl reverse primer and 6 µl of 1x TE buffer. The double-stranded probes were separated from single-stranded oligonucleotides on a 12 % polyacrylamide gel (together with 1x LB 4 °C, 200 V, ca. 3 h in running buffer 0.5x TBE buffer). These DNA bands were cut out of the gel and purified by adding 120 µl of TE buffer to each tube and incubated overnight on a shaker (37 °C, 1000 rpm). Next the oligonucleotides were run through a 1000 µl filtered pipette tip (centrifuged for 1 min at 3381 g). A washing step with 30µl of 1x TE buffer was included (centrifuged for 1 min at 3381 g). Double stranded DNA concentration was measured with Nanodrop . Each oligonucleotide was then diluted to a 40 ng/µl stock and a 1 ng/µl working solution.

**PCR**

EMSAs were also performed with cy5-labeled oligonucleotide probes with a length of 121 bp. Therefore a PCR was performed with a Cy5-labeled forward primer and an unlabeled reverse primer. Full sequences can be found in the appendix (**Table A6**). Two hundred ng of a linearized vector containing 2500 bp of the surrounding region of the *rs7903146* was used. PCR was performed with PCRBIO HiFi Polymerase according to manufacturers' protocol with the following program: 1 min 95 °C, 35 x (15 s 95 °C, 15 s 60 °C, 20 s 72 °C), 72 °C 2 min, 10 °C). PCR was performed in large quantities and cleaned up over a 2.3 % agarose gel (95 V, 90 min) with a subsequent clean-up with the Promega Wizard PCR clean-up kit. Each oligo was then diluted to a 40 ng/µl stock and a 1 ng/µl working solution.

**EMSA**

A 5.3 % polyacrylamide gel was used for EMSA. EMSA 1x GBB, the nuclear extract and the competitor poly(dIdC) were incubated for 10 min on ice, followed by the addition of the Cy5-labeled oligonucleotides and another 20 min incubation on ice. At 200 V in 0.5x TBE running buffer for 3 h the probe was separated and subsequently fluorescence was measured with Typhoon TRIO+ Variable Mode Imager (633 nm, high sensitivity, 650 V).

**Supershift assay**

The supershift was performed like an EMSA with the exception of adding an antibody right after addition of the nuclear extract and incubating for 40 min instead of 10 min. The antibody binds the protein of interest and leads to a change in band pattern.

**Competition assay**

In the competition assay, a consensus sequence of the transcription factor of interest was added to the EMSA reaction after the addition of nuclear extract and incubated for 20 min. Consensus sequences were selected and designed based on the Genomatix database and were ordered as forward and reverse

strand. These oligonucleotides were annealed the same as the Cy5-labeled oligonucleotides.

### 4.2.5 Affinity chromatography

**Annealing**

Affinity chromatography was performed with biotinylated oligonucleotide probes, which were selected as surrounding regions of SNPs of interest of the major as well as minor variant. The full sequences are listed in the appendix. Primers were diluted with dd$H_2O$ to 100 $\mu$M according to manufacturer's instructions. Preparation of the oligonucleotides was performed like the Cy5-labeled oligonucleotides. Double stranded DNA concentration was measured with Nanodrop. Oligonucleotides were diluted to 8.2 ng/µl as a working solution.

**Beads preparation**

Magnetic streptavidin-labeled Dynabeads were mixed and 50 µl was transferred into a DNA LoBind reaction tube. The tube was placed on the magnet stand and the supernatant was removed. The beads were washed twice with 100 µl of 1x B&W buffer. One hundred µl of 2x B&W was added and the suspension was split into two DNA LoBind tubes, adjoining the addition of 50 µl of biotinylated oligonucleotides and incubated for 1 h at room temperature on a rotator was performed. This was followed by an incubation at 4 °C overnight. The supernatant was removed and the beads were washed once with 100 µl 1x B&W, hence 100 µl of 1x B&W with 2 ng/µl biotin was added and incubated for 1 h at room temperature on a rotator to reduce unspecific binding. Again a washing step was performed, this time with 200 µl of washing buffer (WB).

**Trapping**

After removal of the WB, 200 µl of 1x BB was added and removed again, subsequently 1 mg of nuclear extract for affinity chromatography was added to the beads and incubated for 20 min, rotating at 4 °C. Hence a specific amount of Poly(dIdC) was added to the tubes and incubated rotating at 4 °C for another 10 min. The supernatant was collected (SN). Next the beads were washed three times with 200 µl WB and the supernatants were pooled (W1-3). Afterwards the elution was performed: 100 µl of elution buffers (EBs) with increasing NaCl concentrations (starting from 200 mM until 1 M NaCl) were incubated for 3 min and then collected in DNA LoBind tubes. The collected samples were frozen in liquid nitrogen and stored at -80 °C for further analysis with tandem MS/MS. Beforehand, a fraction of the sample was used to perform an EMSA.

### 4.2.6 Liquid chromatography coupled with tandem mass spectrometry (LC-MS/MS)

Sample treatment, measurement and LC-MS/MS data processing was performed by Dr. Juliane Merl-Pham at the Proteomic Core Facility of the Helmholtz centre Munich as previously described [35] with modifications. In brief, fractions were digested by filter-aided sample preparation (FASP) as described before with an FASP approach adaptation [143] using Vivacon 500 filters with a 30 kDa cut-off (Sartorius, Goettingen, Germany). LC-MS/MS analysis was performed on a Q-Exactive HF mass spectrometer (Thermo Scientific) online coupled to an Ultimate 3000 nano-RSLC (Dionex). Tryptic peptides were automatically loaded on a C18 trap column (300 µm inner diameter (ID) × 5 mm, Acclaim PepMap100 C18, 5 µm, 100 Å, LC Packings) at 30 µl/min flow rate prior to C18 reversed

phase chromatography on the analytical column (nanoEase MZ HSS T3 Column, 100Å, 1.8 µm, 75 µm x 250 mm, Waters) at 250 nl/min flow rate in a 95 min non-linear acetonitrile gradient from 3 to 40 % in 0.1 % formic acid. Profile precursor spectra from 300 to 1500 m/z were recorded at 60000 resolution with an automatic gain control (AGC) target of 3e6 and a maximum injection time of 50 ms. TOP10 fragment spectra of charges 2 to 7 were recorded at 15000 resolution with an AGC target of 1e5, a maximum injection time of 50 ms, an isolation window of 1.6 m/z, a normalized collision energy of 27 or 28 and a dynamic exclusion of 30 seconds. For protein identification and label-free relative quantification the RAW files (Thermo Scientific, Dreieich, Germany) were analyzed using the Progenesis QI for proteomics software (version 4.0, Nonlinear Dynamics, Waters, Eschborn, Germany), as described previously [144, 145], with the following changes: Dependent on the dataset and time of analysis, spectra were searched using the search engine Mascot (Matrix Science, London, UK) against the following protein databases: Ensembl Rat (release 80, 28611 sequences), Swissprot human (release 2017 02, 20235 sequences), Swissprot mouse (release 2017 02, 16871 sequences) or Swissprot rat (release 2017 02, 8024 sequences). False discovery rates were stringently kept below 1 % as calculated by a Mascot-integrated decoy database search using the percolator algorithm (significance threshold of $P < 0.05$). Peptide assignments were re-imported into Progenesis QI for proteomics. Normalized abundances of all unique peptides were summed up and allocated to the respective protein.

### 4.2.7 RNA isolation

**Method 1**

Cells were plated in 6 wells and grown until 80-90 % confluency. Cells were washed twice with 1 ml cold PBS. 0.5 ml cold IsolRNA was added to each well and resuspended. This was transferred into 1.5 ml tubes und frozen at -80 °C for further use. Samples were thawed and incubated at room temperature for 5 min. Two hundred µl chloroform per ml IsolRNA was added and mixed for 15 s. After 3 min incubation, the samples were centrifuged for 15 min at 4 °C and 12000 g. The sample consisted of three layers of which the upper most was recovered. Per 1 ml IsolRNA, 500 µl isopropanol was added. After a short vortex it was incubated at room temperature for 10 min. Subsequently the samples were centrifuged for 10 min, 4 °C, 12000 g and the supernatant was discarded. One ml 75 % ethanol was added per ml IsolRNA and centrifuged for 5 min, 4 °C, 7600 g. The supernatant was discarded and the sample was dried for 20 min. Finally 40 µl of HPLC water was added and concentrations were measured with Nanodrop.

**Method 2**

RNA was isolated with the RNeasy Mini Kit from Qiagen according to manufacturers' protocol. QIAshredder columns were used in addition to enhance homogenization of the samples. In cases of pre mRNA isolation an additional step with a DNAse digest was perfomed. Samples harvested in RLT buffer containing β-mercaptoethanol (1:100) were added to the QIAshredder columns and centrifuged for 2 min at 12000 g. The flow-through was mixed with 70 % ethanol in a ratio of 1:1. The solution was added to the RNA column and centrifuged for 30 s at 12000 g. The flow-through was discarded and 700 µl of RW1 was added, centrifuged for 30 s at 12000 g and the flow-through discarded again. Then, the column was washed once with 500 µl of RPE, while centrifuging for 30 s at 12000 g, and then again with 500 µl of RPE but centrifuging for 2 min at 12000 g. The column was placed in a new tube and centrifuged an additional time for 1 min at 12000 g. At least 30 µl of RNAse-free water was added to the column, which was placed in a 1.5 ml tube and was centrifuged for 1 min at 12000 g. RNA concentration was measured with Nanodrop and samples were either directly used for cDNA

synthesis or stored at -80 °C.

**Method 3**

RNA was isolated with the RNeasy 96 Kit from Qiagen according to manufacturers' protocol. Treated cells in 96 well plates were washed once with 150 µl PBS (4 °C) and 150 µl of RLT was added. After shaking horizontally for a few seconds every direction, 150 µl of 70 % ethanol was added and mixed by pipetting up- and down 3 times. The cell lysate was then transferred into the 96 well RNA column plate. The plate was centrifuged for 6 min, 4500 g at 16 °C. After the addition of 800 µl of RWI buffer to each well, the plate was centrifuged again for 6 min, 4500 g at 16 °C. The flow-through was discarded and 800 µl of RPE buffer was added and again centrifuged for 6 min, 4500 g at 16 °C. Another 800 µl of RPE buffer was added, but this time centrifuged for 15 min, 4500 g at 16 °C. Forty-five µl of RNAse free water was added to the membrane and centrifuged for 6 min, 4500 g at 16 °C and an additional 25 µl of RNAse free water was added and centrifuged the same way. RNA concentrations were measured with Tecan.

### 4.2.8 cDNA synthesis

RNA was thawed on ice and 500 ng or 100 ng were brought to a final volume of 12 µl. Quantitect Reverse Transcription Kit was used according to manufacturers' protocol. Therefore 2 µl of 7x gDNA Wipeout Buffer was added to the RNA and incubated for 5 min at 42 °C, hence immediately placed on ice. A reverse-transcription master mix was added to the RNA. Each reaction contained 1 µl RT Primer Mix, 4 µl 5x Quantiscript RT Buffer and 1 µl of Quantiscript Reverse Transcriptase reaching a total volume of 20 µl. This was incubated for 25 min at 42 °C and 2 min at 95 °C before frozen at -20 °C. A working solution with a dilution of 1:5 or 1:10 was prepared additionally.

### 4.2.9 Overexpression of transcription factors of interest

**Overexpression**

**Method 1**

Cells were transfected using electroporation with AMAXA transfection kits. Kit V, R, T were used for HepG2, Huh7, 1.1B4 and HEK293T, respectively. Seven days after electroporation cells were harvested for nuclear extraction, ChIP harvest and RNA isolation. This type of transfection was not successful in the HepG2 and Huh7 cells.

**Method 2**

Selected transcription factors were purchased as DYK-tagged ORF clones from Genscript and used for transfection. Therefore 4 µg of vector was mixed in 3 ml OptiMEM medium. Simultaneously 50 µl of Lipofectamine 2000 was added to 3 ml OptiMEM. These two solutions were mixed together and incubated for 15 min at room temperature. HEK293T cells were plated 4 million per 150 mm dish 48 h prior to transfection. They were washed once with 25 ml warm PBS and the 6 ml transfection mix was added to the dish. This was incubated for 2 h at 37 °C 5 % $CO_2$ before adding another 20 ml of OptiMEM. The cells were harvested for nuclear extraction and Western blot 48 h after transfection start.

### 4.2.10 Real-time Polymerase Chain Reaction (RT-PCR)

For TaqMan assays, 400 nM forward primer, 400 nM reverse primer and 100 nM probe were mixed with 1x SensiFast Probe No-Rox reagent to a final volume of 10 µl. The PCR program consisted

of an intial denaturation step at 95 °C for 2 min followed by 45 cycles of 95 °C for 10 s and 60 °C for 30 s. As reference genes, *ACTB, YWHAZ* and *PPIA* were used. Final values were calculated as mean $2^{(-ddCt)}$ plus SEM. For SYBR assays, 300 nM forward primer and 300 nM reverse primer were mixed with 1x qPCRBIO SyGreen Mix Lo-ROX reagent to a final volume of 10 µl The PCR program consisted of an intial denaturation step at 95 °C for 10 min followed by 45 cycles of 95 °C for 15 s, 60 °C for 20 s and 72 °C for 20 s followed by a melting curve. Final values were calculated as mean $2^{(-ddCt)}$ plus SEM. RT-PCR was always performed in technical duplicates or triplicates.

### 4.2.11 Genotyping

Cell lines were genotyped concerning the SNP *rs7903146*. Therefore, primers flanking the SNP were used, giving a product size of approximately 2600 bp. AmpliTaq Gold (0.15 µl) was used with 1x Gold buffer, 3.3 mM $MgCl_2$ and 1 mM dNTPs, 500 nM of TCF790seq1 Fa, 500 nM of TCF790E5XhoI Ra and 200 ng of gDNA. The PCR program consisted of a 12 min initial denaturation step at 95 °C followed by 50x (20 s at 95 °C, 40 s at 60 °C and 1 min 30 s at 72 °C) with a final elongation for 2 min at 72 °C. The PCR product was digested with exonuclease I and antarctic phosphatase according to manufacturers' protocol. This sample was send to sequencing with the internal primer (TCF790seq1 Fa). Electropherrograms were analysed by visual inspection.

### 4.2.12 Interferone γ stimulation

Huh7 and HepG2 cells were seeded at a density of 20,000 cells/48 well and 50,000 cells/48 well 24 h prior to stimulation. IFNγ was diluted to a working solution of 1 µg/ml in 0.1 % BSA. Cells were incubated in DM (delivery media) with a final IFNγ concentration of 25 ng/ml or 50 ng/ml. Cells were harvested for RNA isolation at the following time points: 2 h, 4 h, 6 h and 24 h. Downstream targets of IFNγ were investigated with RT-PCR. The following downstream genes were analysed: *STAT1, STAT2, STAT3, IRF1, DDB1* and *SOCS1*. Additonally, cells were treated with 50 ng/ml and harvested after 24 h for ChIP and nuclear extraction.

### 4.2.13 Western blot

Nuclear extracts from overexpression experiments were tested for protein content. Since all overexpressed proteins contained a DYK-Tag, only one antibody was necessary. Twenty µg of nuclear protein were incubated in 1x Laemmli buffer for 5 min at 90 °C and centrifuged for 5 min at 13000 g at 4 °C. For proteins larger than 60 kDa a 10 % acrylamid gel was used, all others were loaded on 15 % gels. Gels ran for 15 min, 120 V in 1x Running buffer, then 50 min at 160 V. For semi-dry blotting, nitrocellulose membranes were used. Membrane and gel were sandwiched in between two thick Whatman paper soaked in freshly prepared blotting buffer. The blotting was performed for 10 % gels at 300 mA, 25 V for 55 min, for the 15 % gels at 250 mA, 25 V for 45 min. Gels were stained with Coomassie to ensure full blot of proteins onto the membrane. The membrane was washed in 1x PBS and then blocked for 1 h at room temperature in 1x PBS containing 5 % BSA. Hence the first antibodies were added 1:1000 in blocking solution with 0.1 % Tween20 overnight at 4 °C rocking. The next day, the membrane was washed 3x for 10 min in PBS-T. The second antibodies were added 1:10000 in PBS-T and incubated for 2 h at room temperature rocking. Before measurement, the membrane was washed another two times for 10 min with PBS-T and once with PBS.

### 4.2.14 Replication of vectors

Vectors used for overexpression and luciferase assay experiments were replicated in Stellar Competent Cells. Therefore 1 $\mu$l of plasmid was incubated with 25 $\mu$l of *E.coli* for 30 min on ice. Then a heat shock was performed at 42 °C for 1 min followed by a 1 min cool-down on ice. After the addition of 250 $\mu$l of SOC medium, th ecells were incubated for 45 min at 37 °C on a shaker (165 rpm). The cells were then plated on a LB plate containing antibiotics ampicillin 100 $\mu$g/ml or kanamicin 50 $\mu$g/ml overnight at 37 °C. Clones were inoculated overnight in LB medium containing antibiotics (ampicillin 100 $\mu$g/ml or kanamicin 50 $\mu$g/ml) at 37 °C 165 rpm. Vector DNA was isolated with the Promga PureYield Plasmid Maxiprep System according to manufacturers' instructions or alternativley with Miniprep solutions. Therefore bacteria were centrifuged at 20000 g for 1 min for the removal of the LB medium. The pellet was re-suspended in 200 $\mu$l Miniprep solution 1, followed by the addition of 200 $\mu$l Miniprep solution 2 for cell lysis. After careful mixing the cells for 5 min, 200 $\mu$l Miniprep solution 3 was added for neutralization. This was again carefully mixed and then centrifuged for 5 min at 20000 g. The supernatant containing the plasmid DNA was collected into a new tube and 420 $\mu$l isopropanol was added and for 1 h for precipitation. Another centrifugation step of 10 min at 20000 g was performed in order to remove the supernatant. The DNA pellet was dried and cleaned with the Promega Wizard SV Gel and PCR Clean-Up kit according to manufacturers' instructions.

### 4.2.15 Luciferase reporter gene assays

In order to assess the effect of proteins on transcriptional activity, gene reporter assays were performed. Therefore, HEK293 cells or 1.1B4 were seeded in 48 well plates with a density of 40,000 or 20,000 cells per well 24 h prior to transfection. One hour before transfection medium was changed to OptiMEM medium. Transfection was performed with FuGene transfection reagent: per well, 90 ng of, each, the firefly vector and pcDNA overexpression vectors plus 50 ng of renilla vector were mixed with 0.625 µl of FuGene reagent in OptiMEM and incubated 24 h at 37 °C, 5 % $CO_2$. The cells were washed once with PBS and 40 µl of passive lysis buffer was added. After a 10 min incubation at 1050 rpm on a shaker plate at room temperature, the luciferase activity of firefly and renilla were measured in a single tube luminometer as follows: 5 µl of sample with 25 µl of firefly buffer and respectively 25 µl of renilla buffer. Tubes were blanked individually. Firefly values were substracted by the corresponding renilla values and normalized in regard to the control vector (containing only the promotor) with a pcDNA DYK vector without the overexpressed gene. Experiments were performed in triplicates on three separate days. Additional experiments were performed by Katharina Eiseler and Andrea Tóth with luciferase vectors containing only a 44 bp fragment of the SNP-surrounding area together with overexpression of selected transcription factors. Twenty-four hours prior transfection, 80,000 HEK293 cells were seeded in 48 well format. One hour prior to transfection, the medium was changed to OptiMEM. Transfection mixes contained 200 ng promotor vector, 100 ng Renilla control vector and 100 ng transcription factor vector together with 0.63 µl Fugene reagent. After 24 h incubation at 37 °C, 5 % $CO_2$, medium was changed to normal growth medium and in selected cases, IFN$\gamma$ stimulation was performed. After another 24 h cells were harvested and measured as described above. Vector maps are depicted in **Figure A4**.

### 4.2.16 Transient siRNA knockdown

**Implementation of cell transfection**

siRNAs were purchased from Dharmacon. Accel siRNA was used according to manufacturers' recommendations: Stock solutions of 100 $\mu$M were dissolved in siRNA buffer. Next, cells were seeded one day prior transfection to reach 80 % confluency. siRNA was mixed with either OptiMEM or Delivery Medium in a ratio of 1:100 to reach a final concentration of 1 $\mu$M. This mixture was added to the cells after removal of the growth medium and incubated for 72 h at 37 °C, 5 % $CO_2$. Alternatively medium was changed to growth medium and incubated additional 24 h. Cells were washed twice with 150 $\mu$l PBS and RNA was isolated as described above. As control cells with only OptiMEM or Delivery Medium (mock) and cells incubated with scrambled siRNA were included in the analysis. All experiments were performed in 96 or 48 well format. Final condition consisted of 72 h incubation of 1 $\mu$M siRNA in Delivery Medium, followed by a change to growth medium for an additional 24 h before harvest.

**Determination of mRNA stability of *TCF7L2* and *ACSL5* with actinomycin D**

Actinomycin D is an antibiotic and anticancer drug, which intercalates to DNA and stabilizes the topoisomerases I and II- DNA complexes, hence inhibiting mRNA expression [146]. Therefore, the stability of the existing mRNA in the nucleus can be determined. 1.1B4 cells were seeded 24 h prior to treatment. Cells were approximately 80 % confluent when actinomycin D was added in two different concentrations: 2 $\mu$g/ml and 5 $\mu$g/ml. Cells were harvested at the following time points: 3 h, 6 h and 24 h, each with a sham control (methanol). mRNA was isolated according to method 2. cDNA synthesis was performed as described above, 100 ng RNA was used per reaction. cDNA was diluted 1:5 and RT-PCR was performed in 384 well format with a final reaction volume of 10 $\mu$l. The reaction consisted of 2 $\mu$l of cDNA, 330 mM forward primer, 330 mM reverse primer and 1x SYBR reagent per reaction. The PCR program proceeded as followed: 10 min at 95 °C, 40x (15 s 95 °C, 30 s 60 °C, 30 s 72 °C), 15 s 95 °C, 15 s 40 °C and a melting curve. *18S* was used as reference gene.

**Cell viability assay**

Cells were transfected with siRNA as described above. *PPIB* siRNA, *scrambled* siRNA and mock transfected cells were tested in this assay. Therefore 5 % (v/v) of alamarBlue was added to the media and incubated for 1 h and 25 h light protected at 37 °C, 5 % $CO_2$ before measurement. Both absorbance and fluorescence were measured at the two time points. Absorbance was detected at 570 nm with a reference wavelength at 600 nm. For fluorescence detection the excitation wavelength was set at 560 nm and the emission wavelength at 590 nm. Values were divided by the mean mock value and mean and SEM was calculated.

**Pre mRNA RT-PCR**

We wanted to determine, whether a knockdown of a certain factor can influence the expression of the alleles of *TCF7L2*. Since *rs7903146* is located within an intron, it can be detected in the pre-mRNA. In order to determine allele-specific expression of *TCF7L2*, the pre-mRNA had to be pre-amplified. Herefore 2 $\mu$l of cDNA from the knockdown experiments were mixed with 16.75 $\mu$l water, 1x HiFi Buffer, 400 mM TCF7L2preampF, 400 mM TCF7L2preampR and 0.25 $\mu$l HiFi Polymerase and incubated the following way: 95 °C 1 min, 15x (95 °C 15 s, 62 °C 30 s, 72 °C 30 s), 72 °C 2 min. This was followed by a digest with antarctic phosphatase and exonuclease I according to manufacturer's protocol. The samples were diluted 1:10 before 2 $\mu$l was used for qPCR. Final primer concentrations was 0.33 $\mu$M. Allele- specific qPCR was performed under the following conditions: 10

min at 95 °C, then 45x (15 s 95 °C, 20 s 60 °C, 20 s 72 °C), 15 s 95 °C, 15 s 40 °C and a melting curve. A standard curve was added in every qPCR to determine the primer efficiency. Therefore, the gDNA of the cell line of interest was added to the reaction: 0.03 ng, 0.3 ng, 3 ng, 30 ng and 300 ng. From these Ct values a curve was generated and the slope was determined. The results were calculated as follows: $10^{((-1/\text{slope})-1)} * 100$. As reference, a 90 bp genomic region surrounding the SNP was used to determine the dCt. As ddCt the values were substracted from the C allele dCt values. Values were displayed as mean $2^{-\text{ddCt}}$ with SEM.

### 4.2.17 Chromatin immunoprecipitation (ChIP)

In this method, DNA-binding proteins are crosslinked to the DNA with formaldehyde and subsequently sonicated to obtain smaller DNA fragments (100-700 bp) before performing an immunoprecipitation with a protein-specific antibody. Cells of interest grew in 150 mm Petri dishes until they reached 80 % confluency. The ChIP-IT Express Enzymatic Kit was used.

**Cell harvest**

For the cell harvest, the following solutions were prepared for 1x 150 mm dish: fixation solution (15 ml medium, 1 % formaldehyde), 1x PBS and glycine stop fix solution (1x Glycine, 1x PBS). The medium of the cells was removed and 20 ml of fixation solution/dish was added. This was incubated for 10 min, rocking at room temperature. The fixation solution was discarded and 5 ml ice cold 1x PBS was added to wash the cells. After discarding the 1x PBS, 5 ml of glycine stop fix solution was added to the dish and incubated for 5 min, rocking at room temperature. The solution was discarded and the cells were washed another time with 5 ml ice cold 1x PBS before adding 2.5 ml of ice cold 1x PBS, scratching the cells and transferring them into a 15 ml tube. Hence, the tubes were centrifuged for 10 min at 4 °C, 720 g and the supernatant was discarded. One $\mu$l of PMSF (100 mM) and 1 $\mu$l of protease inhibitor cocktail was added to the cell pellets before shock freezing them in liquid nitrogen and storing them at -80 °C until further use.

**Chromatin shearing**

Pellets were resuspended in 1 ml 1x lysis buffer and incubated on ice for 30 min before physically lysing them with a Sub-Q-Syringe by pumping up and down 20 times. The lysate was centrifuged at 2400 g, 10 min at 4 °C, the supernatant was discarded. Pellets were resuspended in 350 $\mu$l 1x digestion buffer and incubated at 37 °C for 5 min, followed by the addition of ETDA to a final concentration of 10 mM and an incubation on ice for 10 min. The suspension was now sonified under the final condition: Amplitude: 25 %, 12x (pulse: 20 s, pause: 30 s). After centrifuging at 18000 g for 10 min at 4 °C the supernatant was collected in a new tube and frozen at -80 °C until further use.

**Chromatin check**

To determine the gDNA concentration and DNA fragment sizes in the sheared samples, a phenol/chloroform isolation was performed. Hereby 50 $\mu$l of the sheared gDNA was incubated overnight at 65 °C and a NaCl concentration of 240 mM. The following day, *RNAse A* was added in a final concentration of 47.4 ng/$\mu$l and incubated at 37 °C for 15 min. Next, Proteinase K was added with a final concentration of 22.6 ng/$\mu$l and incubated at 42 °C for 90 min. To this, 200 $\mu$l of phenol/chloroform (ratio 1:1.1) was added and mixed vigorously for 10 s, followed by a centrifugation step of 5 min at 21000 g. The top water phase was collected in a new DNA Lobind tube and sodium acetate (pH 5.5) was added to a final concentration of 0.22 M. Ice-cold ethanol was added to reach a final ethanol

concentration of 70 %, mixed and kept at -80 °C for 1.5 h. The samples were centrifuged with 21000 g for 15 min at 4 °C and the supernatant was removed. A washing step with ice-cold 70 % ethanol was performed before centrifuging at 21000g for 10 min at 4°C. The supernatant was discarded and the ethanol evaporated with a vacuum centrifuge for 10-15 min. The pellet was resuspended in 20 $\mu$l water and concentration was determined via Nanodrop. Ten $\mu$l of the chromatin was loaded on a 1.5 % agarose gel (together with 1x loading dye) and ran at 150 V for 1.5 h to determine gDNA fragment sizes.

**Immuno precipitation**

In this step, the chromatin incubates with antibody and is cleaned up over a bead system to enrich the genomic region where a specific transcription factor is binding. Ten $\mu$l of untreated chromatin is saved to use as an input control later in the qPCR. Beside the antibody for the protein of interest, an isotope (IgG) control as well as a positive control (ChIP-IT Express Enzymatic Kit) were used. For the reaction 10 $\mu$g of sheared gDNA was used, together with 2 $\mu$g of antibody according to manufacturers' recommendations. Therefore, 25 $\mu$l of Protein G Magnetic Beads were mixed with 10 $\mu$l ChIP Buffer I, 10 $\mu$g chromatin, 1 $\mu$l PIC, 2 $\mu$g of antibody and filled up to 100 $\mu$l with water. The samples were incubated overnight at 4 °C on a shaker. After a short spin-down, the samples were placed on the magnet and the beads were washed once with 800 $\mu$l of ChIP Buffer I and twice with 800 $\mu$l ChIP Buffer II. After removal of the supernatant, the beads were re-suspended with 50 $\mu$l of Elution Buffer AM2. This solution was incubated, rotating, for 15 min at room temperature. Hence, 50 $\mu$l of Reverse Cross-linking Buffer was added and mixed by pipetting up and down. The sample was put back onto the magnet and the supernatant was collected in a new tube. The input gDNA was mixed with 88 $\mu$l ChIP Buffer 2 and a final concentration of NaCl of 100 mM. Both samples were incubated at 65 °C overnight. Ten ng/$\mu$l proteinase K was added to the samples and incubated for 1 h at 37 °C, followed by the addition of 2 $\mu$l of pre-warmed proteinase K Stop Solution. The samples were spun down and stored at -20 °C until further use.

**Allele-specific qPCR**

First, primer pairs were checked for allele-specificity. Therefore, a qPCR with gDNA with genotypes C/C (Huh7), C/T (HepG2) and T/T (HEK) as templates was performed as follows: 40 ng gDNA, 0.33 $\mu$M forward primer, 0.33 $\mu$M reverse primer, 5 $\mu$l of SYBR reagent per reaction. Total volume was 10 $\mu$l. The reactions were incubated 10 min at 95 °C, then 40x (15 s 95 °C, 30 s 64 °C, 30 s 72 °C), 15 s 95 °C, 15 s 40 °C and a melting curve. The final primer combinations were CHiPTCF7L2Fb with either ChIPrs7903146C R7 or ChIPrs7903146T R7. The final PCR was performed as follows: 3 $\mu$l chromatin template, 0.33 $\mu\mu$M forward primer, 0.33 $\mu$M reverse primer, 5 $\mu$l SYBR reagent, filled up to 10 $\mu$l with water. qPCR was performed under these conditions: 10 min at 95 °C, then 45x (15 s 95 °C, 20 s 60 °C, 20 s 72 °C), 15 s 95 °C, 15 s 40 °C and a melting curve. A standard curve was added in every qPCR in order to determine the primer efficiency. Therefore input DNA of the cell line of interest was diluted to the concentrations: 0.005 ng, 0.05 ng, 0.5 ng, 5 ng and 50 ng. From these values a curve was generated and the slope was determined (Graphpad). The efficiency was calculated with a formula provide by active motif: $10^{((-1/\text{slope})-1)} * 100$. The amplification efficiency (AE) was calculated as follows: $10^{(-1/\text{slope})-1}$. To calculate the difference in allelic binding, the following calculations were performed: $100* (AE^{(\text{mean Input - mean sample})})$. The ChIP was seen as positive when the amplified DNA of the specific region of interest was at least four times higher than the amplification of a random intergenic region (supplied with Kit) with the same antibody immunoprecipitation material. All conditions were performed in biological triplicates and qPCR was additionally performed in technical triplicates.

# 5 Results

## 5.1 Identification of candidate *rs7903146* binding proteins

### 5.1.1 Genotypes of selected cell lines

We looked at expression of C and T allele of *TCF7L2* in experiments such as transient siRNA knockdowns and allele-specific *in vivo* binding of transcription factors with ChIP, so it was important to identify the correct cell model. A cell line that is heterozygous was a big advantage, since efforts to create a cell line with all three genotypes with the HDR CRISPR/Cas9 system failed during the course of this thesis (data not shown). Huh7, Caco2 and SGBS cells display the C/C wild type genotype. HepG2, HT29 and 1.1B4 cells display a C/T genotype, whereas HEK293 cells the T/T risk genotype. Thus, for the investigation of *in vivo* binding of transcription factors, HepG2 and 1.1B4 were chosen to determine allele-specificity.

### 5.1.2 Selection of regulated target genes

The target gene of choice, when looking at a SNP in an intronic region is, primarily, the gene containing this intron itself, in the case of *rs7903146*, *TCF7L2*. However, Xia *et al.* [147] revealed, *TCF7L2* might not be the only gene regulated by *rs7903146*. They claim, *ACSL5*, a gene directly upstream of *TCF7L2*, is also affected by the genomic region containing the SNP. They also found many more potential candidates. However, *ACSL5* was most promising, since deletion of the *rs7903146* containing region abolished significantly detectable chromatin contacts with the *ACSL5* promoter [147]. With this information, it was obvious not only to check whether *TCF7L2* is influenced by e.g. knockdown of potential regulating candidates, but *ACSL5* as well. Before starting, the expressions of both genes was checked via RT-PCR. *TCF7L2* is expressed well in all cell lines tested, while expression of *ACSL5* was low in 1.1B4, HEK293 and NCI H716 cells compared to HepG2 cells. Ct values were high, but still detectable (**Figure A1**).

### 5.1.3 Affinity chromatography coupled with LC MS/MS identifies candidate regulatory Transcription Factors

Affinity chromatography is an *in vitro* method, where nuclear extract is incubated with DNA sequences of interest, which are coupled to beads (solid phase). These beads are then washed with a salt gradient. Salt interferes with protein/DNA binding, which eventually elutes the binding proteins to the liquid phase. The different salt elutes are collected separately and later on measured in LC-MS/MS to identify the proteins. The higher the affinity of a protein to the DNA sequence of interest, the later it will be detected in the salt elutions. In this experiment we were interested in proteins binding to either C nonrisk or T risk allele with different affinities. We identified all proteins in the eluted fractions and quantified the abundance in each fraction, which enabled assessing allele-specificity. These experiments were performed in the liver cell lines HepG2, Huh7, the beta-cell lines 1.1B4, Ins1

and in the preadipocyte cell line 3T3L1. In these experiments, the range of total identified proteins ranged from 1300 to 10000 proteins per cell line extract. Of these approximately 16 % were annotated in MATBASE (Genomatix) as genes related to transcriptional regulation. Overall, in this gene-set 1 % - 6 % of proteins were found binding with a significantly ($p \leq 0.05$ , Wilcoxon test) higher affinity to one of the two alleles. The two liver cell lines HepG2 and Huh7, shared 450 MATBASE annotated proteins, of which 66 were annotated as TFs. The two β-cell lines Ins1 and 1.1B4 shared 110 MATBASE annotated proteins of which 26 were annotated as TFs. Overall all cell lines shared 800 proteins of which 140 were MATBASE annotated, however only 7 annotated as TFs. None of the TFs were significant in all cell lines. Analysing the dataset proved a challenge. Apart from quantified proteins, proteins solely identified but not quantified, limited statistical analysis. Final criteria for selecting relevant proteins were: (1) annotated as DNA binding TFs in the MATBASE database, (2) fold change between C to T allele either $\geq 2$ fold or $\leq$ 0.5 fold, (3) statistical significance with a Wilcoxon signed rank test. This lead to eleven proteins of interest, shown in **Table 5.1**. We want to note, that we also found two proteins already published as allele-specific binding: PARP1 [127] with FC = 4.35, $\mathbf{p = 0.03}$ in Ins1 cells and HMGB1 [128] with FC = 0.6, $\mathbf{p = 0.03}$ in 1.1B4 cells (both not annotated as TF and therefore not included in **Table 5.1**). Since HMGB1 was already extentsively investigated, we no longer included it in the following experiments.

**Table 5.1:** Significant results of the affinity chromatography

| **Protein** (gene symbol) | **Fold Change** (C/T allele) | **p value** | **Cell line** | **Fraction** [mM NaCl] | **Repl** |
|---|---|---|---|---|---|
| Atf1 | 8.32 | **0.03** | 3T3L1 | 200 | 4 |
| Bbx | 0.12 | **0.03** | 3T3L1 | 1000 | 4 |
| CEBPG | 0.32 | **0.03** | Huh7 | 300 | 6 |
| Gtf2b | 0.48 | **0.03** | 3T3L1 | 200 | 4 |
| HMGA2 | 0.34 | $\mathbf{1*10^{-3}}$ | Huh7 | 400 | 6 |
| HMGA2 | 0.10 | **0.03** | 1.1B4 | 1000 | 6 |
| Pdx1 | 2.42 | **0.03** | Ins1 | 400 | 4 |
| Purb | 0.00 | **0.03** | Ins1 | 200 | 4 |
| SCRT1 | 0.28 | **0.03** | 1.1B4 | 1000 | 6 |
| STAT1 | 0.18 | **0.03** | HepG2 | 400 | 4 |
| Taf6 | 19.40 | **0.03** | 3T3L1 | 200 | 4 |
| ZNF593 | 0.33 | **0.03** | HepG2 | 400 | 4 |

### 5.1.4 Transient siRNA knockdown to evaluate effect of candidate Transcription Factors

A transient siRNA knockdown is a fast method to see whether a certain protein influences the expression of the target gene of interest. The siRNA sequence binds the mRNA of the protein that is intended to be removed and this duplex is degraded by the RIS complex, inhibiting further translation. After the selection of 11 transcription factors identified in the AC-LC MS/MS, we initiated a large scale transient siRNA knockdown. To find the best conditions, a various experiments were carried out in advance. Firstly, the mRNA stability of our target genes was tested by actinomycin D, which inhibits mRNA expression [146]. This leaves only the already existing mRNA, which than can be investigated for its rates of degradation. This information is important to evaluate when the downstream effect

of a knockdown is measureable. If the target gene is very stable, the effect may not be detectable, if harvested too early. Two actinomycin D concentrations were tested (2 $\mu$g/ml and 5 $\mu$g/ml) and the cells were harvested after 3 h, 6 h and 24 h, each with a sham control (methanol). After 6 h, there was already so little mRNA left, that a reliable measurement with RT-PCR was not possible. **Figure 5.1 A** shows the relative mRNA content of *TCF7L2* and *ACSL5* with 2 $\mu$g/ml actinomycin D and a 3 h incubation. As reference gene, *18S* was used, since it is a very highly expressed gene among many cell types and *ACTB* was used as an additional control [148]. mRNA levels of *TCF7L2* was reduced to 45 % ($\mathbf{p = 1*10^{-3}}$), *ACSL5* to 43 % ($\mathbf{p = 6*10^{-3}}$) already after 3 h. *ACTB*, a highly expressed gene was reduced to 75% (p = 0.19). Cell viability is also a crucial factor in an experiment. Cell viability was only tested in regard to the medium, two time points and siRNA type, which required no lipofection reagent. AlamarBlue is a redox indicator for cellular health [149]. Overall, all conditions displayed a high cell viability compared to untreated control (100 %). Lowest cell viabilty was noted in 1.1B4 cells after 96 h knockdown with *PPIB* in DM medium (68 %), $\mathbf{p = 0.03}$). After 72 h only HepG2 *PPIB* DM knockdown showed a significant decrease of cell viability by 14 % ($\mathbf{p = 1*10^{-3}}$). After 96 h knockdown, Huh7 cells displayed in several conditions a significant reduction in cell viability. *PPIB* knockdown in OptiMEM revealed a reduction by 14 % ($\mathbf{p = 3*10^{-3}}$), scrambled in OptiMEM by 10 % ($\mathbf{p = 0.01}$) and *PPIB* knockdown in DM by 20 % ($\mathbf{p = 0.03}$). Knockdown efficiencies ae depicted in **Figure A3**. The efficiencies varied between 65 % and 79 %.

### 5.1.5 Interferone γ stimulation

STAT1 was one of the TFs identified and selected with AC LC MS/MS (**Table 5.1**). However, there was no binding visible in EMSA, even after successful overexpression of *STAT1* in HEK293T cells (data not shown). The hypothesis arouse that not enough activated STAT1 was present in the nucleus. IFNγ is an important activator of the JAK-STAT pathway playing a role in immunity and inflammation. The binding of IFNγ to its receptor activates Jak1 and Jak2, which in turn leads to phosphorylation of cytosolic inactive STAT at tyrosine 701. Phosphorylated STAT1 translocates to the nucleus, where it binds as dimers to its specific gene targets [150]. To take advantage of this mechanism, we stimulated Huh7 and HepG2 cells with IFNγ with two different concentrations (25 ng/µl or 50 ng/µl) and harvested at different time points (2 h, 4 h, 6 h and 24 h) to see whether pathway specific genes are being upregulated. Time points and concentrations were chosen based on a publication by Melén *et al.* [151]. We looked at a selection of downstream targets (*STAT1, STAT2, STAT3, DDB2, IRF1 and SOCS1*), reported by literature [151, 152], to test for succesful stimulation and therefore activation and translocation of STAT1 to the nucleus. Apart from these known regulated genes, we were also interested, whether our genes of interest, *TCF7L2* and *ACSL5*, were already influenced by the stimulation. Both IFNγ concentrations had similar effects, so in this thesis, only 50 ng/µl is shown in Figure **5.2**. For each time point, there was a stimulated and a sham control (methanol), which was then used for the calculation of the relative change in expression. Sham control of each time point is therefore set to 1 and not shown in the graphs. *TCF7L2* was in general not influenced by the stimulus. Due to low variability, significant effects were reached in 4 h, 6 h and 24 h in Huh7 and after 6 h in HepG2 (all p values can be found in Table **A1**). However, when looking at the difference in relative expression, it is rather low. *ACSL5* showed significant increases beginning 6 h of IFNγ stimulation in both cell lines. In Huh7 cells, expression was doubled after 4 h, then decreased slightly to a 68 % increase and after 24 h displayed a 4.6 fold increase. *STAT1* and *STAT2* display increased expression as well, already after 2 h of stimulus.

Figure A

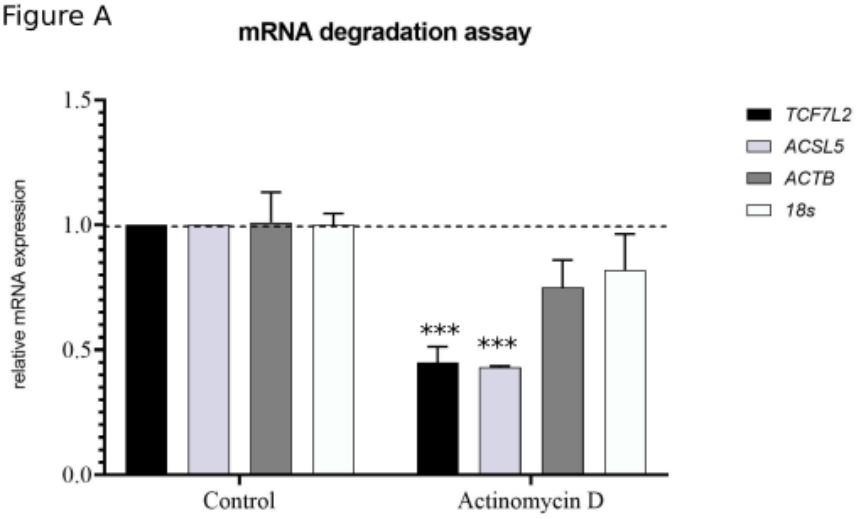


Figure B

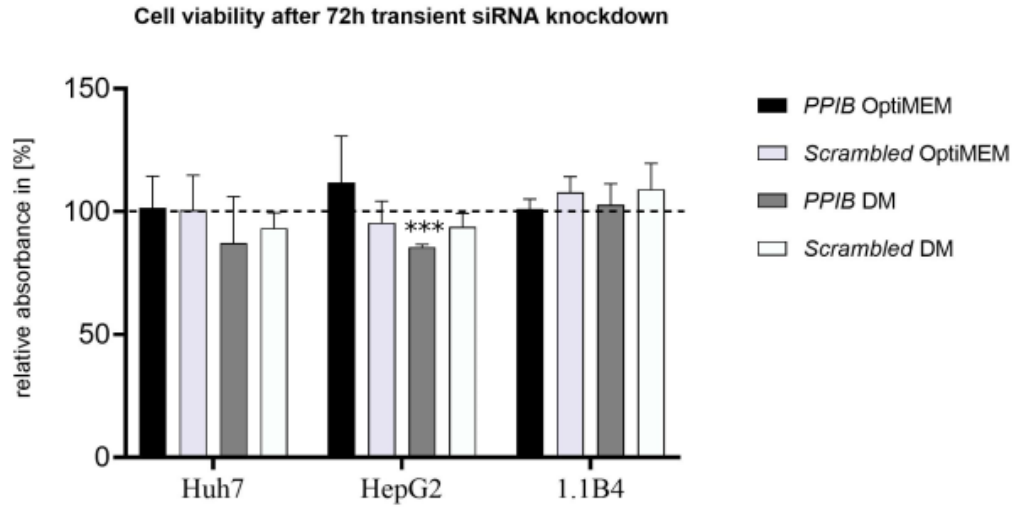


Figure C

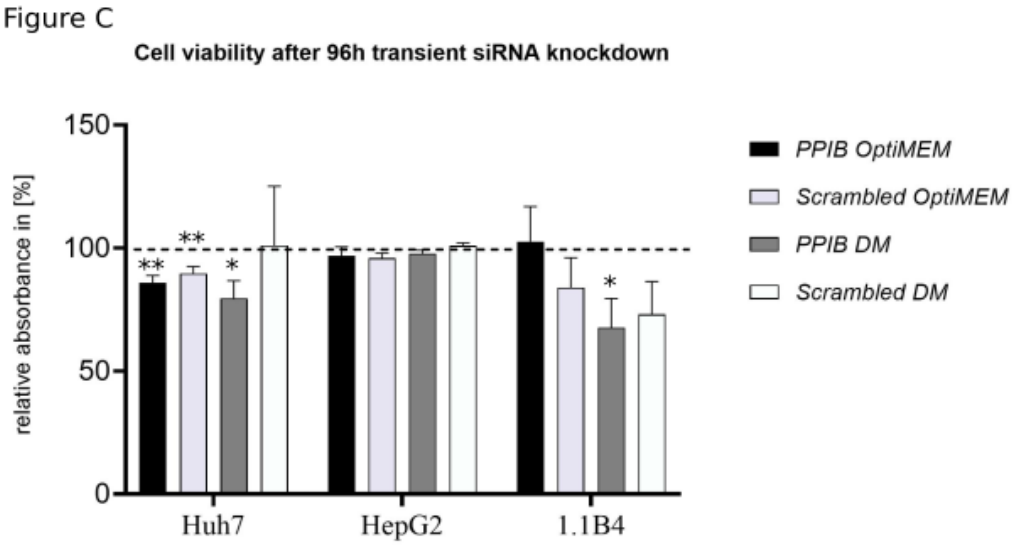


**Figure 5.1:** **A**: mRNA stability determination with actinomycin D. Cells were incubated for 3h with 2µg/ml actinomycin D before RNA isolation. Values are displayed mean relative expression of *TCF7L2, ACSL5* and *ACTB* compared to *18S* expression + SEM. **B** and **C**: Cell viability measurement of different transient siRNA knockdown conditions with Alamarblue. Conditions varied in time (72 h or 96 h), transfection medium (OptiMEM or DM) and siRNA (*PPIB* or scrambled) in Huh7, HepG2 and 1.1B4. Values are displayed as mean relative absorbance + SEM, n = 4. Unpaired t-test was used for analysis of the gene expressions compared to untreated, $p < 0.05 = *$, $p < 0.01 = **$, $p < 1*10^{-3} = ***$.

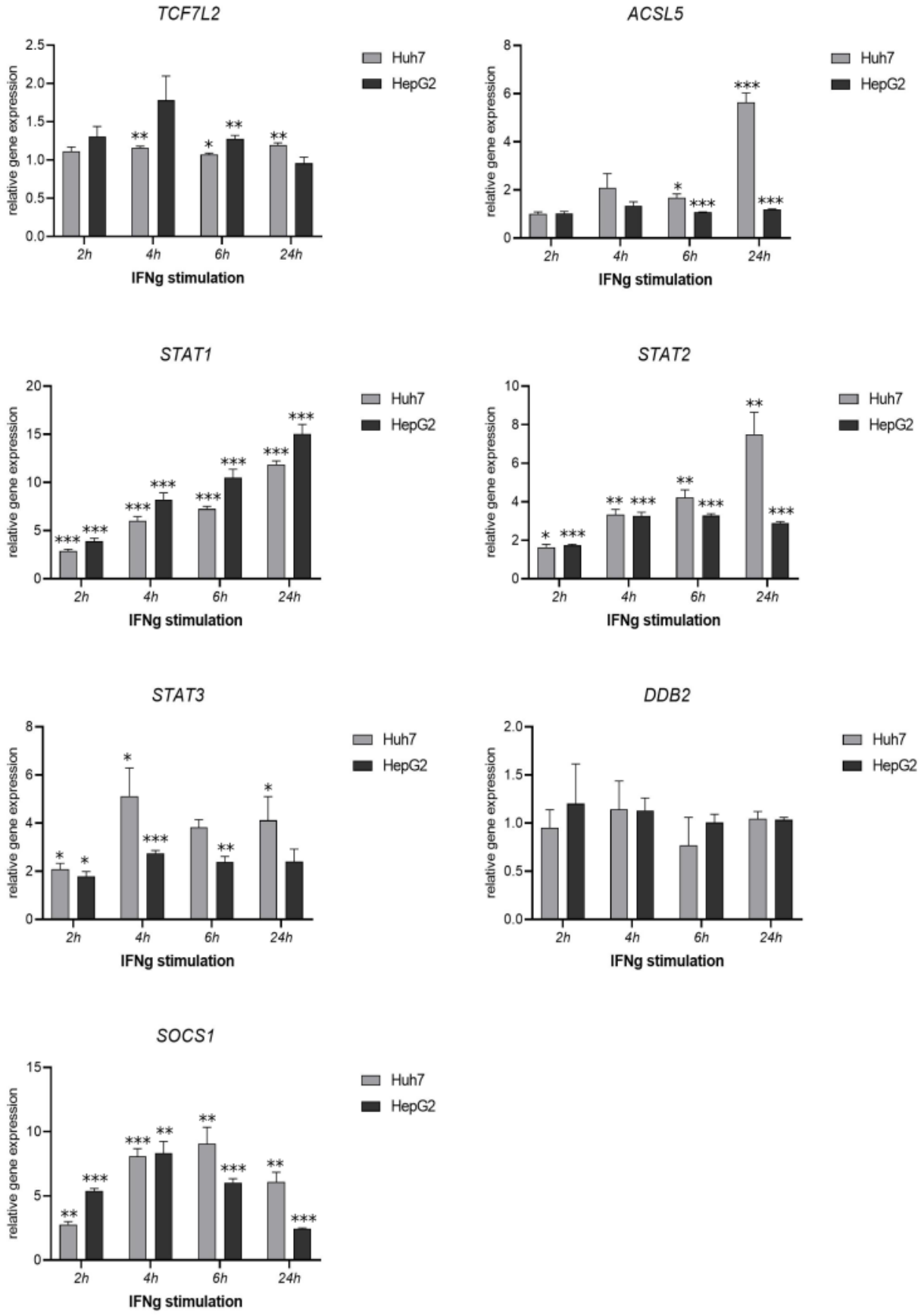


**Figure 5.2:** Huh7 and HepG2 stimulated with 50 ng/ml IFNγ for different time spans: 2 h, 4 h, 6 h and 24 h. Each time point had an unstimulated condition as control. Values are displayed as mean relative expression of each gene + SEM, n = 3. Unpaired t-test was used for analysis of the gene expressions compared to untreated, $p < 0.05$ = *, $p < 0.01$ = **, $p < 1*10^{-3}$ = ***.

## 5.2 *PDX1*

### 5.2.1 Pdx1 was identified with AC LC MS/MS binding specifically to the C non-risk allele.

Pdx1 is highly relevant for β-cell specificity and was only detected in the affinity chromatography of Ins1. It was found to bind 2.4 fold stronger to the C non-risk allele in the salt elution fraction 400 mM (Wilcoxon signed rank test $\mathbf{p = 0.03}$). The experiment was performed with 4 biological replicates (**Table 5.1**). Performing LC MS/MS of the nuclear extracts and cytosolic fractions revealed the presence on Pdx1 only in the nuclear fraction of the Ins1 cells and in none of the other cell lines. When we performed RT-PCR, *PDX1* was not detectable in 1.1B4, even though expression was reported in literature [136]. This problem was later circumvented by overexpression. According to SNPinspector (Genomatix, Germany), a PDX1 binding site is generated newly through the C to T transition. Hence a T risk allele specific binding of PDX1 was expected according to this analysis.

### 5.2.2 The binding of PDX1 were confirmed in EMSA, supershift EMSA and competition EMSA.

Since PDX1 is only expressed in Ins1, naturally only there we would find endogenous PDX1 binding to cy5-labeled oligonuecleotides representing our region of interest. In **Figure 5.3 A** it was shown that using Ins1 nuclear extract, there were several bands containing PDX1. The addition of unlabeled *PDX1* consensus sequence oligonucleotides (**Figure 5.3 B**) lead to a reduction in their band intensity. The two consensus sequences (**Figure 5.3 D**) were added in three increasing amounts (5 fold, 10 fold and 50 fold excess compared to the labeled oligonucleotide of interest), leading to a corresponding reduction in band intensity. **Figure 5.3 C**, confirmed this result with 5 fold addition of the two different cold-labeled PDX1 consensus sequences, where it showed four bands which distinctively reduce with the competition. The addition of a PDX1-specific antibody lead to the shift of three bands to the upper region of the gel. This effect could not be seen when adding an unpecific IgG antibody to the reaction. Since we wanted to see the binding behavior of human PDX1 in a human cell, we overexpressed *PDX1* in 1.1B4 and HEK293T before harvesting nuclear extract for EMSA. The human hybridoma cell line 1.1B4 overexpressing PDX1 showed a band, which vanished after the addition of the PDX1-specific antibody, and could not be found in the control 1.1B4 overexpressing the pcDNA3.1 DYK control vector. **Figure 5.3 D** shows HEK293T cells overexpressing *PDX1* and untransfected HEK293T as control. The overexpression of *PDX1* showed two very distinct additional bands appearing, while another seemed to be missing. The intensity of these newly appearing bands was very pronounced. All of these results layed prove, that Pdx1/PDX1 is binding the region surrounding the SNP, however no allele-specificity could be observed (quantification of band intensity: mean C/T = 0.97, SEM = 0.01, n = 8). Additional experiments were performed to find conditions, where PDX1 potentially binds allele-specifically (**Figures 5.4 and 5.5**). Conditions included varying detergent concentrations, unspecific competitors, buffer conditions, salt content, specific unlabeled sequence competition or different PDX1 protein concentrations. A summary on optimization of shift conditions can be found in Hellman *et al.* [153]. Adding detergent to an electrophoretic mobility shift assay can increase protein solubility. In our case, CHAPS, TWEEN20 and TRITONX were tested in two different concentrations each (0.1 % and 0.01 %). CHAPS and TWEEN20 showed no effect on band pattern, while TRITON X lead to the disappearance of several bands. However, none of the conditions led to an allele-specific binding pattern with Ins1 nuclear proteins (**Figure 5.4 A**). Next to Poly(dIdC), other proteins can be added

to reduce unspecific binding in EMSA reactions. We added BSA in four different amounts to our EMSA reactions (1 ng, 3 ng, 1000 ng and 3000 ng). The 1000 ng and 3000 ng conditions lead to a slight backround smear, which is rather counterproductive in this assay. No allele-specific binding patterns could be detected under these conditions (**Figure 5.4 B**). The gel binding buffer contains pH stabilizer, salt and glycerol. The variation of the concentration of its constituents can influence the protein/DNA complexes in the gel. Different buffer concentrations (0.06x, 0.12x, 0.5x, 0.75x, 1x) were tested with different amounts of protein present (3 µg and 1 µg) together with different amounts of oligonucleotide (0.8 ng, 1 ng or 1.5 ng), leading to no allele-specific binding patterns. When 1.5 ng of oligonucleotide was added, an additional band was visible (**Figure 5.4 C**). Buffers can vary in components e.g. glycerol can be replaced by Ficoll or NaCl can be replaced by KCl. Varying these conditions did not have any effect on the binding pattern, except for the addition of TRITON X, which again shifted some of the bands (**Figure 5.4 D**). NaCl is known to play a role in protein/DNA interaction [154], so in **Figure 5.4 E**, we used NaCl concentrations starting from 68 mM to 1 M. It could be shown, that the higher the salt concentration, the lower the binding of proteins in general. However no allele-specificity could be detected in these conditions. In the experiment seen in **Figure5.4 F**, HEK293T overexpressing *PDX1* were used instead of Ins1 nuclear extract. The PDX1 containing extract was mixed with non containing extract in different concentrations. While the band containing PDX1 vanished, when the overexpressing extract was less than 8 % of the total mix, no allele-specificity could be observed. In **Figure 5.5** conditions with unlabeled oligonucleotide was tested. Therefore the cold labeled oligonucleotide of each allele was added at an excess of 0x, 1x, 5x, 10x, 20x and 40x compared to the labeled oligonucleotide. It could be observed, that only after an high excess of unlabeled competitor there was a strong reduction in signal, however no allele-specificity could be observed. In summary, we could not find a condition, where PDX1 displayed allele-specific binding to our locus.

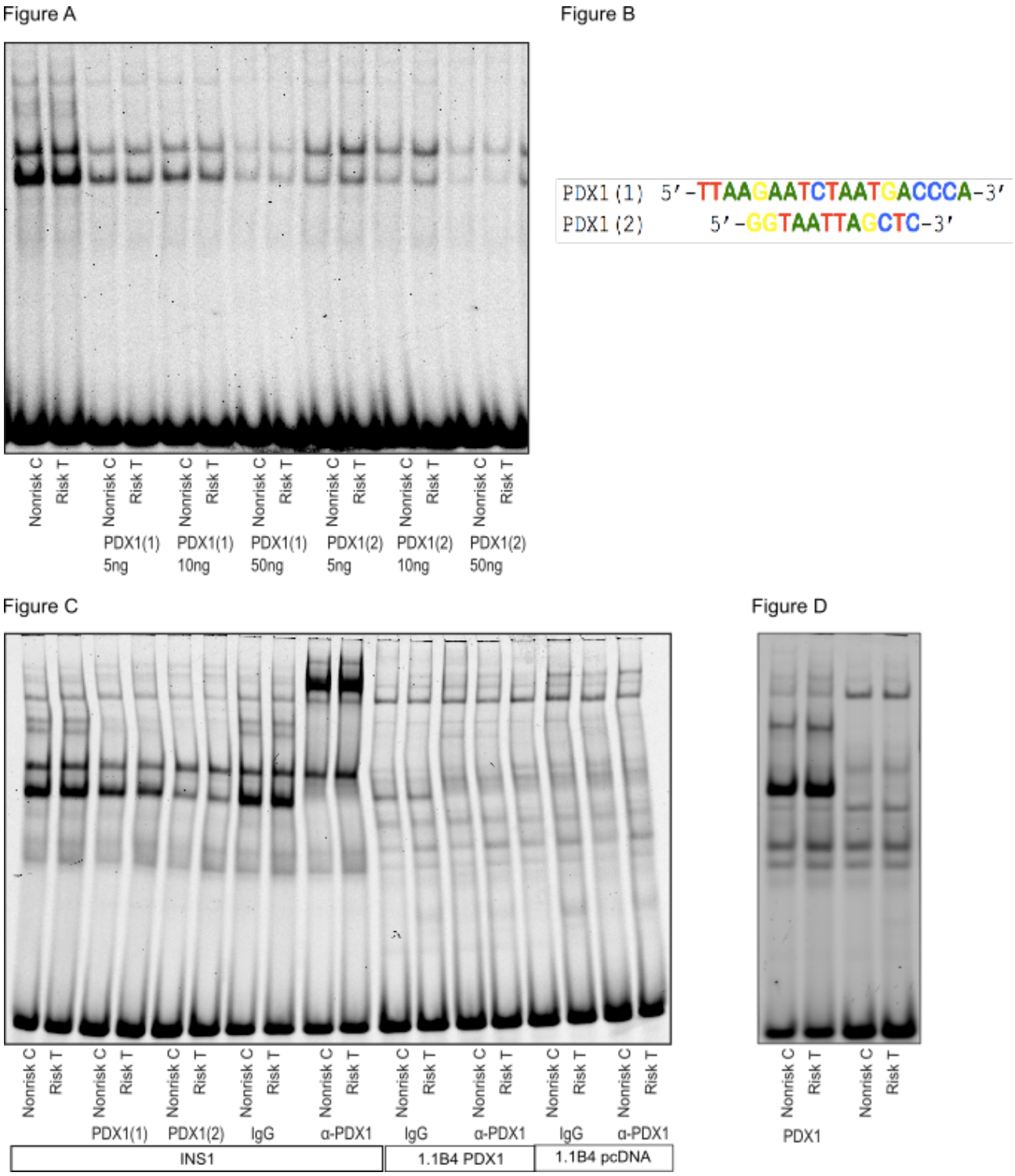


**Figure 5.3:** *In vitro* binding of PDX1 to the sequence of interest in EMSA.

**Figure 5.3 A**: EMSA incubating 3 µg of Ins1 nuclear extract with 700 ng of Poly(dIdC) and 1 ng of a 45 bp cy5-labeled oligonucleotide containing the SNP at mid-position. Two different cold-labeled consensus sequences were tested in 3 different quantities: 5 ng, 10 ng and 50 ng of oligonucleotide. **B**: For the competition EMSA, sequences were chosen based on annotated consensus sequences of the Matrix family of each transcription factor. Data was collected from the database of Genomatix.de. **C**: EMSA incubating 5 µg of Ins1 nuclear extract with 700 ng of Poly(dIdC) or 5 µg of 1.1B4 nuclear extract with 350 ng of Poly(dIdC) and 1 ng of a 45 bp cy5-labeled oligonucleotide containing the SNP at mid-position. To lane 3-6 each 5 ng of two different cold-labeled consensus sequences were added, while in lane 7-18 0.3 µg of antibody was incubated. 1.1B4 PDX1 are cells transfected with PDX1 ORF vector in order to overexpress this protein. 1.1B4 pcDNA are cells transfected with the empty control vector pcDNA3.1 DYK. **D**: EMSA incubating 3 µg of HEK293T nuclear extract with 350 ng of Poly(dIdC) and 1 ng of a 45 bp cy5-labeled oligonucleotide containing the SNP at mid-position. Lane 1-2 contain protein of HEK293T cells overexpressing *PDX1*, while samples in lane 3-4 contain untransfected HEK293T.

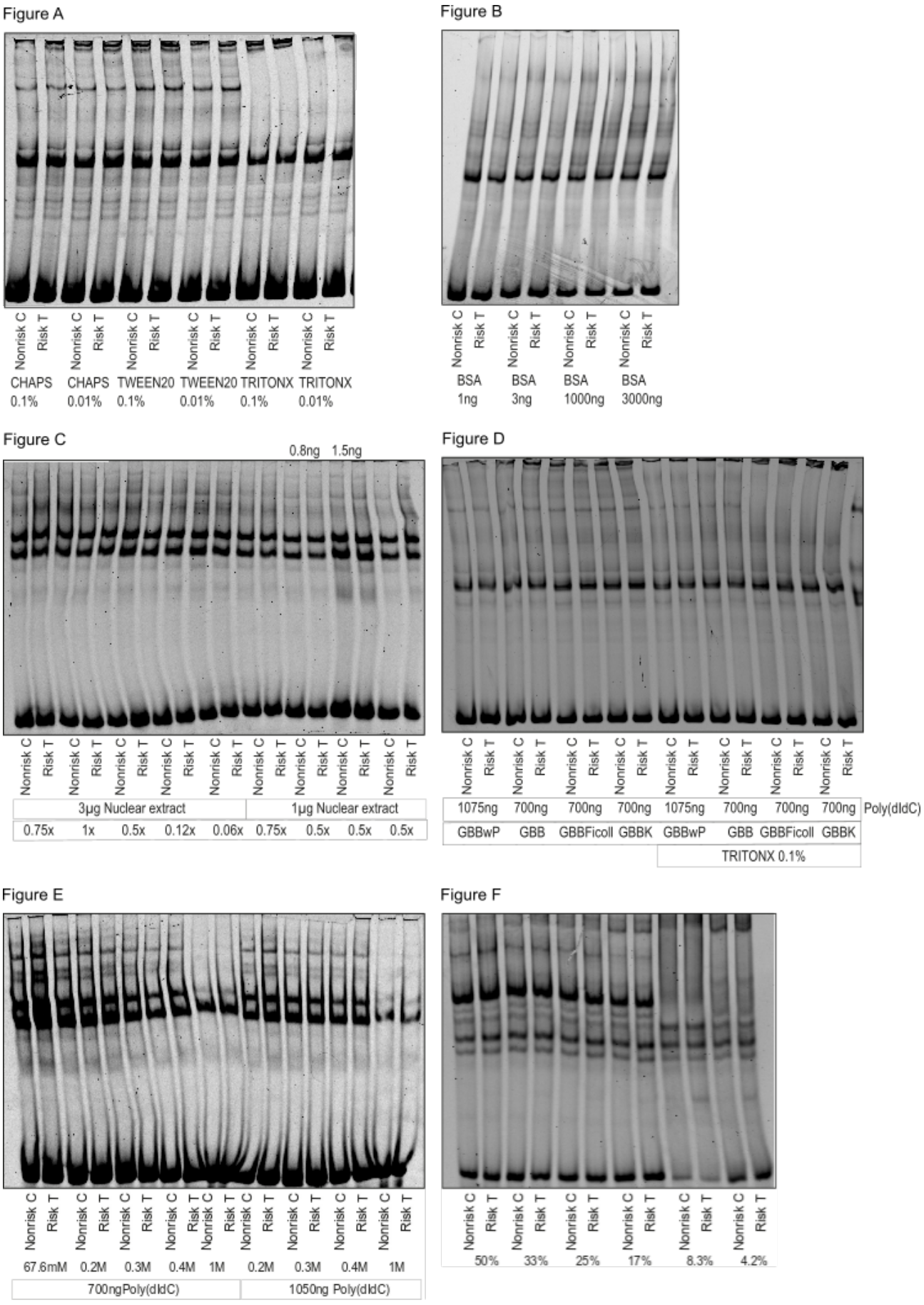


**Figure 5.4:** Different conditions for binding of PDX1 to the sequence of interest in EMSA.

**Figure 5.4 A**: EMSA incubating 1.3 µg of Ins1 nuclear extract with 1075 ng of Poly(dIdC) and a 45 bp cy5-labeled oligonucleotide containing the SNP at mid-position. Different detergent conditions were tested. **B**: EMSA incubating 2.15 µg of Ins1 nuclear extract with 700 ng of Poly(dIdC), 0.1 % TRITONX and a 121 bp cy5-labeled oligonucleotide containing the SNP at mid-position. Different BSA conditions were tested. This gel was 4 % instead of 5.3 %. **C**: EMSA incubating Ins1 nuclear extract with 700 ng of Poly(dIdC) and 1 ng (except lanes 13-16) 45 bp cy5-labeled oligonucleotide containing the SNP at mid-position. Additionally, different buffer concentrations were tested. **D**: EMSA incubating 2.15 µg of Ins1 nuclear extract with varying amounts of Poly(dIdC) and 1 ng 121 bp cy5-labeled oligonucleotide containing the SNP at mid-position. Additionally, different buffer and detergent conditions were tested. **E**: EMSA incubating 3 µg of Ins1 nuclear extract with different amounts of Poly(dIdC) and 1 ng of a 45 bp cy5-labeled oligonucleotide containing the SNP at mid-position. Different NaCl concentrations were tested. **F**: EMSA incubating 6 µg of HEK293T overexpressing PDX1 nuclear extract with 350 ng of Poly(dIdC) and 1.5 ng of a 45 bp cy5-labeled oligonucleotide containing the SNP at mid-position. The nuclear extract was diluted with HEK293T overexpressing pcDNA3.1 DYK nuclear extract to create different PDX1 concentrations.

### 5.2.3 Knockdown of *Pdx1* in Ins1 cells only showed a slight effect on *Ascl5* expression.

Since the rat β-cell line Ins1 was the only cell line, in this project, which expressed *Pdx1*, the knockdown was perfomed in these cells. The knockdown efficiency in all four biological replicates was on average 76 %, and showed no effect on *Tcf7l2* mRNA expression (p = 1.00). The reduction in *Pdx1* lead however to a mild 20 % decrease in *Acsl5* mRNA expression levels (**p = 0.02**) as shown in **Figure 5.6 A**.

### 5.2.4 Overexpression of *PDX1* in 1.1B4 did not show a relevant effect on mRNA expression of *TCF7L2* or *ACSL5*.

We overexpressed *PDX1* in 1.1B4 and harvested next to nuclear extract for EMSA also RNA, to be able to look at gene expression patterns. *PDX1* expression was 31 fold (SD = 18) higher compared to 1.1B4 transfected with pcDNA3.1 DYK or untransfected cells (n = 3). *TCF7L2* was upregulated by 15 % (SD = 0.08, **p = 9*10$^{-3}$**, n = 4) as a result of the overexpression of *PDX1*, and *ACSL5* was upregulated by 19 % (SD = 0.13, **p = 0.03**, n = 4) compared to untransfected cells.

### 5.2.5 Overexpression of *PDX1* in a gene reporter assay showed no allele-specific effect on *TCF7L2* promotor function.

In the course of this project, a luciferase vector construct containing a 3500 bp fragment of the surrounding region of *rs7903146* (Intron 4 of *TCF7L2*) followed by a 868 bp fragment of the *TCF7L2* promotor upstream of the *luciferase* ORF was generated and investigated. While the region containing the wild type and SNP did act as an enhancer in HEK293 cells, it rather decreased expression in 1.1B4 cells (**Figure 5.6 B, D**). In 1.1B4 cells, the T allele was significantly lowering relative expression of luciferase by 19 % (**p = 0.013**). In HEK293 cells, both wild type and SNP caused an average 8 fold increase in expression. When overexpressing *PDX1* in this assay, there was no distinct change in luciferase expression detectable in HEK293 cells or 1.1B4 cells (**Figure 5.6 B, D**), nor was there allele specificity in the influence of the enhancer on the promotor. What should be noted, is that PDX1 acted as an inhibitor on the *TCF7L2* promotor itself (**Figure 5.6 C**).

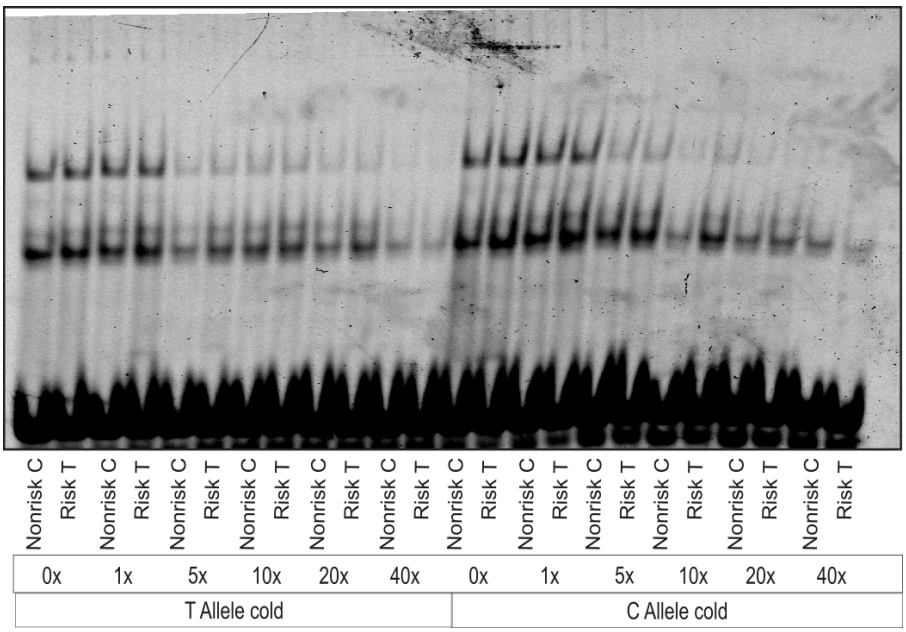


**Figure 5.5:** 5.5: EMSA incubating 3 µg of Ins1 nuclear extract with 700 ng of Poly(dIdC) and 1 ng of a 45 bp cy5- labeled oligonucleotide containing the SNP at mid-position. Different concentrations of cold-labeled oligonucleotide were tested.

Figure A

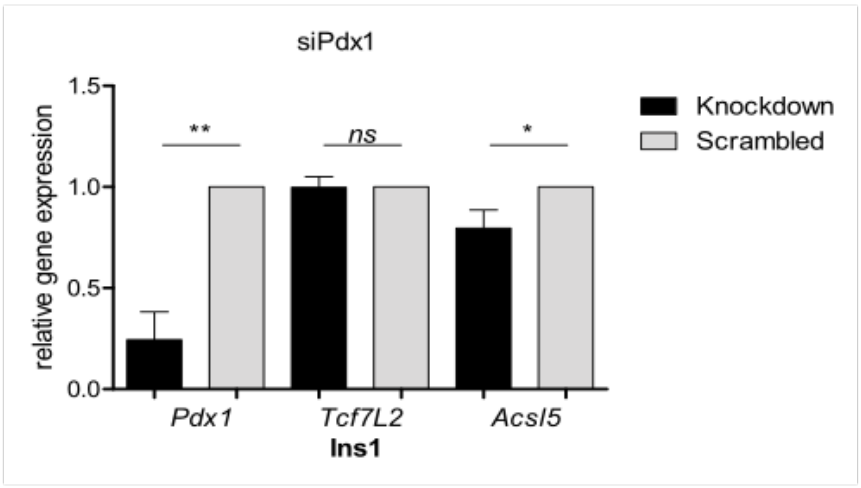


Figure B

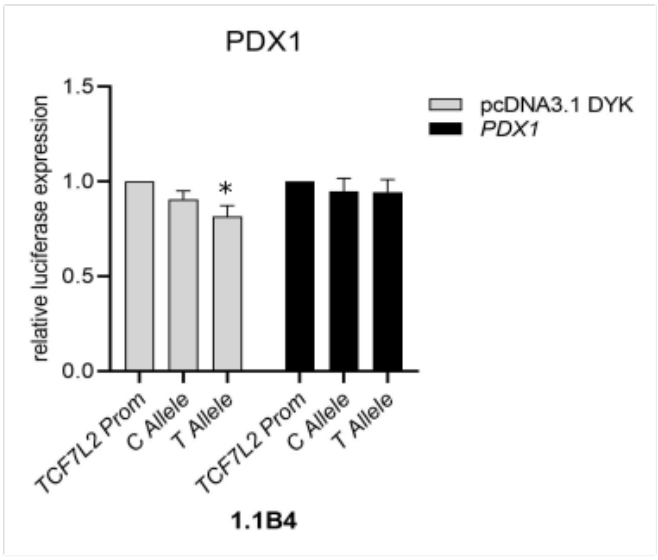


Figure C

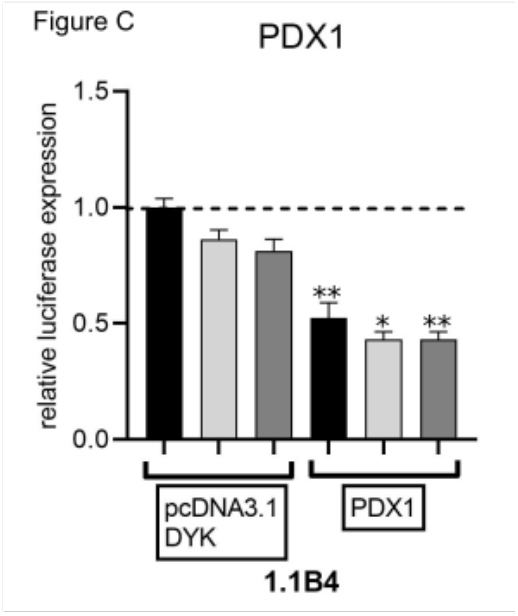


Figure D

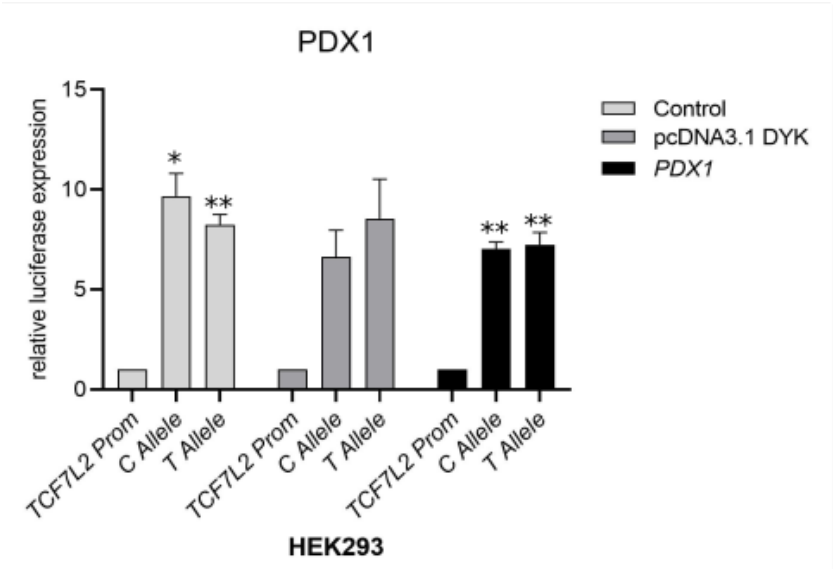


**Figure 5.6:** *Pdx1* knockdown in Ins1 cells and overexpression of *PDX1* in a gene reporter assay in 1.1B4 cells and HEK293 cells.

**Figure 5.6 A**: Transient siRNA knockdown of *Pdx1* in Ins1 cells. Included were experiments with knockdown efficiencies > 50 %. Displayed are mean relative expressions + SEM. One sample t-test was used for analysis, **p (Pdx1) = 2*10$^{-3}$**, n = 4. **B** and **C**: Gene reporter assay investigating the influence of PDX1 on promotor activity in 1.1B4 cells. *TCF7L2* Prom = pGL4.22 vector containing an upstream 868 bp promotor fragment of *TCF7L2*. C/T Allele= *TCF7L2* Prom vector containing an upstream 3500 bp fragment of *TCF7L2* enhancer region containing wild type or SNP. Each condition was calculated against each individual *TCF7L2* promotor as control (**B**) or to *TCF7L2* Prom with pcDNA3.1 DYK (**C**). Values are displayed as mean relative luciferase expression + SEM, n = 8. Black bars = pGL4.22 vector containing a 868 bp fragment of the *TCF7L2* promotor ("TCF7L2 prom"). Light grey bars = *TCF7L2* prom with 3500 bp *TCF7L2* enhancer fragment upstream containing the wild type C allele. Dark grey bars = *TCF7L2* prom with 3500 bp *TCF7L2* enhancer fragment upstream containing the risk T allele. **D**: Gene reporter assay investigating influence of PDX1 on promotor activity in HEK293 cells. Each condition was calculated against each *TCF7L2* promotor as control. Values are displayed as mean relative expression + SEM, n = 3. Unpaired t-test (allelic difference) and one sample t-test were used for analysis, p < 0.05 = *, p < 0.01 = **.

## 5.3 *CEBP* Family

### 5.3.1 CEBPs were identified to be preferably binding to the T risk allele in liver cell lines in AC LC MS/MS.

CEBPG was found significantly (**p = 0.03**) binding to the T risk allele (3.5 fold) in Huh7 at a salt elution of 300 mM (**Table 5.1**). CEBPs are known to form homodimers and heterodimers with each other in order to stabilize and bind to the DNA. Despite the fact, that only CEBPG was found to be significantly, we also included other members of the CEBP family in our anaylsis. Not all members of the CEBP family were found in all cell lines in our AC-MS/MS approach. Further information can be found in **Table 5.2** .

**Table 5.2:** CEBP Family in affinity chromatography

| Protein | Cell line | Fold change | Allele | p-value | replicates |
|---|---|---|---|---|---|
| CEBPG | Huh7 | 3.53 | T | 0.03 | 6 |
| CEBPG | HepG2 | 1.33 | T | 0.69 | 4 |
| CEBPG | 1.1B4 | 1.23 | C | 0.69 | 6 |
| CEBPA | Huh7 | 1.65 | T | 0.24 | 6 |
| CEBPB | HepG2 | 2.35 | T | 0.66 | 4 |
| CEBPB | 3T3L1 | 1.68 | C | 0.34 | 4 |
| CEBPZ | Huh7 | 1.16 | T | 0.39 | 6 |
| CEBPZ | HepG2 | 1.13 | T | 0.89 | 4 |
| CEBPZ | 3T3L1 | 1.41 | C | 0.49 | 4 |
| CEBPZ | 1.1B4 | 1.02 | C | 0.69 | 6 |

In 3T3L1 nuclear extracts only CEBPZ was identified, in Ins1 cells, none of the CEBPs were detectable. CEBPZ was furthermore found in the nuclear extracts of 1.1B4, Caco2, HEK293, HepG2, Huh7 and THP1 cells, while CEBPG was only detectable in Huh7 cells. CEBPB was only found in

HepG2 and 3T3L1 nuclear fractions. These measurements, however, were only performed once and should be interpreted with caution.

### 5.3.2 The binding of the CEBPs to the T risk allele was confirmed in EMSA, supershift EMSA and competition EMSA.

Like in the affinity chromatography, only in Huh7 a T allele-specific endogenous binding of CEBPG could be detected and confirmed with a supershift EMSA (**Figure 5.7 A**). This endogenous CEBPG band was rather unstable not appearing in every EMSA. Especially freeze thaw cycles of the nuclear extract seemed to decrease binding abilities. Since in the EMSAs it was soon discovered, that the only prominent allele-specific band contained HMGA2, it became necessary to overexpress the transcription factors of interest, to be able to properly investigate their binding affinity in this assay. T allele-specific binding of CEBPG was confirmed in HEK293T overexpressing *CEBPG*, as shown in **Figure 5.7 B**, where a shift of this specific band was achieved with two different CEBPG-specific antibodies, but not with the unspecific isotype control IgG. When overexpressing specific isoforms of CEBPA and CEBPB in HEK293T, a highly T-allele binding affinity of these factors was observed. **Figure 5.7 D** shows a supershift EMSA with a CEBPA-specific antibody. It produced a shift of the expected band. The long CEBPA isoform appeared as a weaker band compared to the medium CEBPA isoform. This was also seen in Western blot of the overexpressed proteins (**Figure A2**). If the protein was expressed less or degraded faster was not further investigated in this thesis. **Figure 5.7 E** shows after the addition of the cold-labeled consensus sequence (**Figure 5.7 G**) of CEBPG to HEK293T overexpressing CEBPG, the T risk specific bands shift away. Similar results can be achieved with CEBPA isoforms, when adding the unlabeled CEBPA consensus sequence (**Figure 5.7 F**). In this EMSA it was observed again, that the longer isoform of CEBPA had a less pronounced intensity compared to the medium CEBPA isoform. In this EMSA it was observed again, that the longer isoform of CEBPA had a less pronounced intensity compared to the medium CEBPA isoform A T allele-specific band could often be observed in the lower third of the gels. This band contains HMGA2, which is discussed in section 5.5.

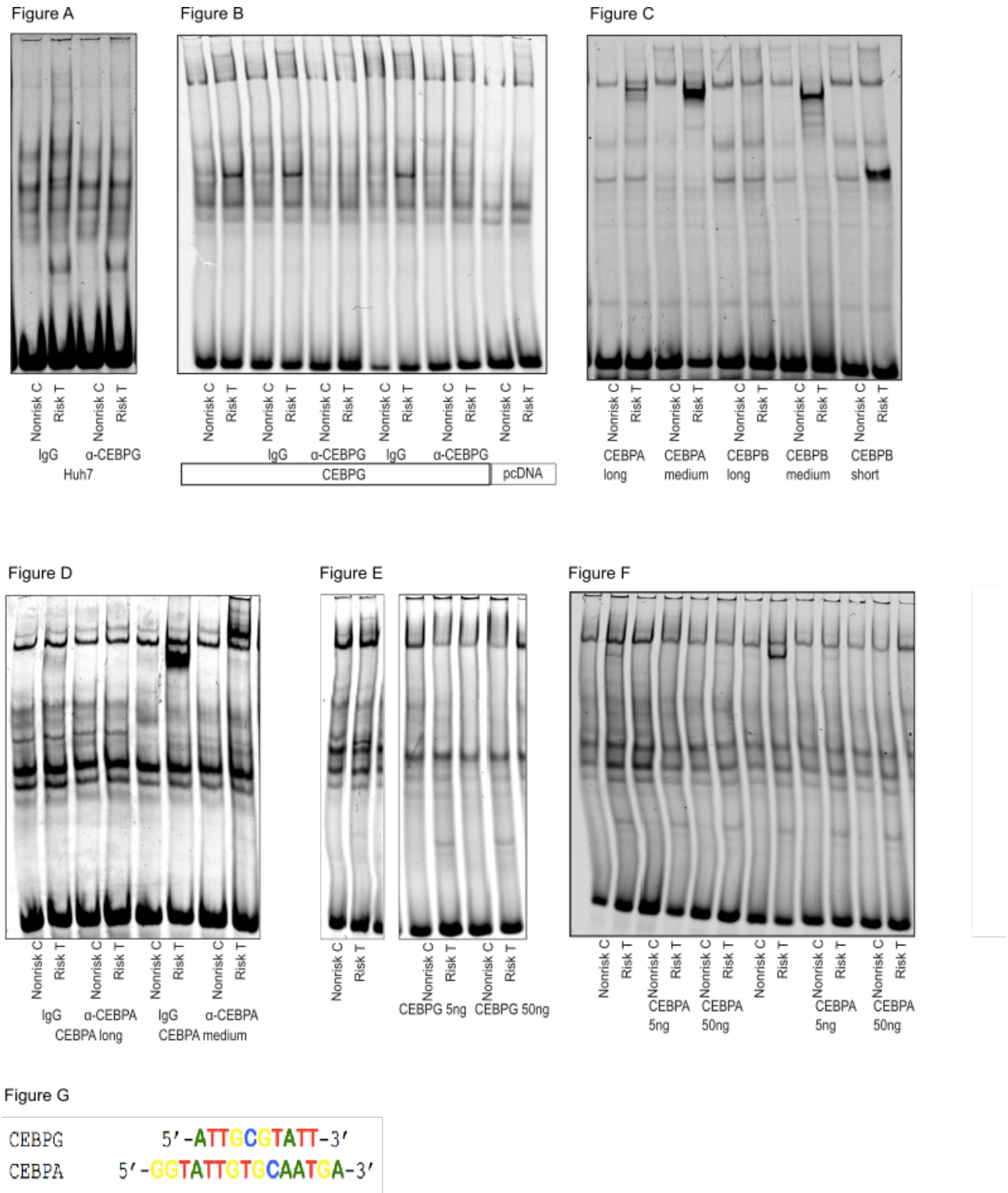


**Figure 5.7:** *In vitro* binding of CEBPs in EMSA.

**Figure 5.7 A**: Supershift EMSA incubating 0.4 µg of Huh7 nuclear extract with 700 ng of Poly(dIdC), 0.3 µg of antibody and 1 ng of a 45 bp cy5-labeled oligonucleotide containing the SNP at mid-position. **B**: 2 µg of HEK293T nuclear extracts overexpressing CEBPG or pcDNA3.1 DYK with 210 ng of Poly(dIdC) and 1 ng of a 45 bp cy5-labeled oligonucleotide. For supershift 3 µg of antibody was added and two different CEBPG antibodies were used. **C**: 3 µg of HEK293T nuclear extracts overexpressing different members of the CEBP family were incubated with 350 ng of Poly(dIdC) and 1.5 ng of a 45 bp cy5-labeled oligonucleotide. **D**: 3 µg of HEK293T nuclear extracts overexpressing different isoforms of CEBPA were incubated with 350 ng of Poly(dIdC) and 1.5 ng of a 45 bp cy5-labeled oligonucleotide. For supershift 0.5 µg of CEBPA antibody was added to the reaction. **E**: EMSA incubating 6 µg of HEK293T nuclear extract overexpressing CEBPG with 350 ng Poly(dIdC) and 1 ng of 45 bp cy5-labeled oligonucleotide. A cold-labeled consensus sequence for CEBPG was added to the reaction. **F**: EMSA incubating 6 µg of HEK293T nuclear extract overexpressing CEBPA long (lane 1-6) or CEBPA medium (lane 7-12) with 350 ng Poly(dIdC) and 1 ng of 45 bp cy5-labeled oligonucleotide. A cold-labeled consensus sequence for CEBPA was added to the reaction. **G**: Consensus sequences chosen for CEBPG and CEBPA based on the Genomatix database.

### 5.3.3 CEBPs interact with each other *in vitro*.

When nuclear extracts of HEK293T overexpressing CEBPs were mixed, there was an additional band appearing at the T risk allele. The band was located below the CEBPA and CEBPB bands and higher than the expected CEBPG band. The CEBPG band was often not visible anymore, at the expected height (**Figure 5.8 A**). This phenomenom lead to the hypothesis of heterodimer formation of the different CEBPs. To support the hypothesis a supershift with a CEBPG-specific antibody was performed (**Figure 5.8 B**). This additional T allele-specific band underneath the CEBPA band disappeared in the shift. The same supershift result could be shown when CEBPG and CEBPB isoforms are combined (**Figure 5.8 C**). When the CEBPB isoforms were combined with each other or with CEBPA isoforms, changes in band pattern were also observed, hence dimerization occurs between all the CEBPs investigated (**Figure A6 A**). The effects were independant of overexpressing the factors simultaneously in HEK293T or mixing the nuclear extracts of each individual overexpressed factor (**Figure A6 B**).

### 5.3.4 Knockdown of *CEBPB* in 1.1B4 showed a significant decrease in *TCF7L2*

To investigate whether the CEBPs have a regulating effect on our target genes *TCF7L2* and *ACSL5*, we carried out a transient siRNA knockdown. By removing the transcription factor a change in expression was expected, since it was shown in EMSA (**Figure 5.7**), that the protein bound specifically to the T risk allele. The knockdowns of *CEBPA, CEBPB, CEBPG* and *CEBPZ* were performed in five different cell lines (SGBS, HEK293, Huh7, HepG2 and 1.1B4), however inconsistent knockdown efficiencies and low expression of e.g. *CEBPA*, lead to only partial results. While the *CEBPG* knockdown was successfully performed three or more times in all cell lines, *CEBPA* and *CEBPB* was only succesful in 1.1B4 cells. *CEBPZ* expression was successfully decreased in HepG2 and 1.1B4 cells.

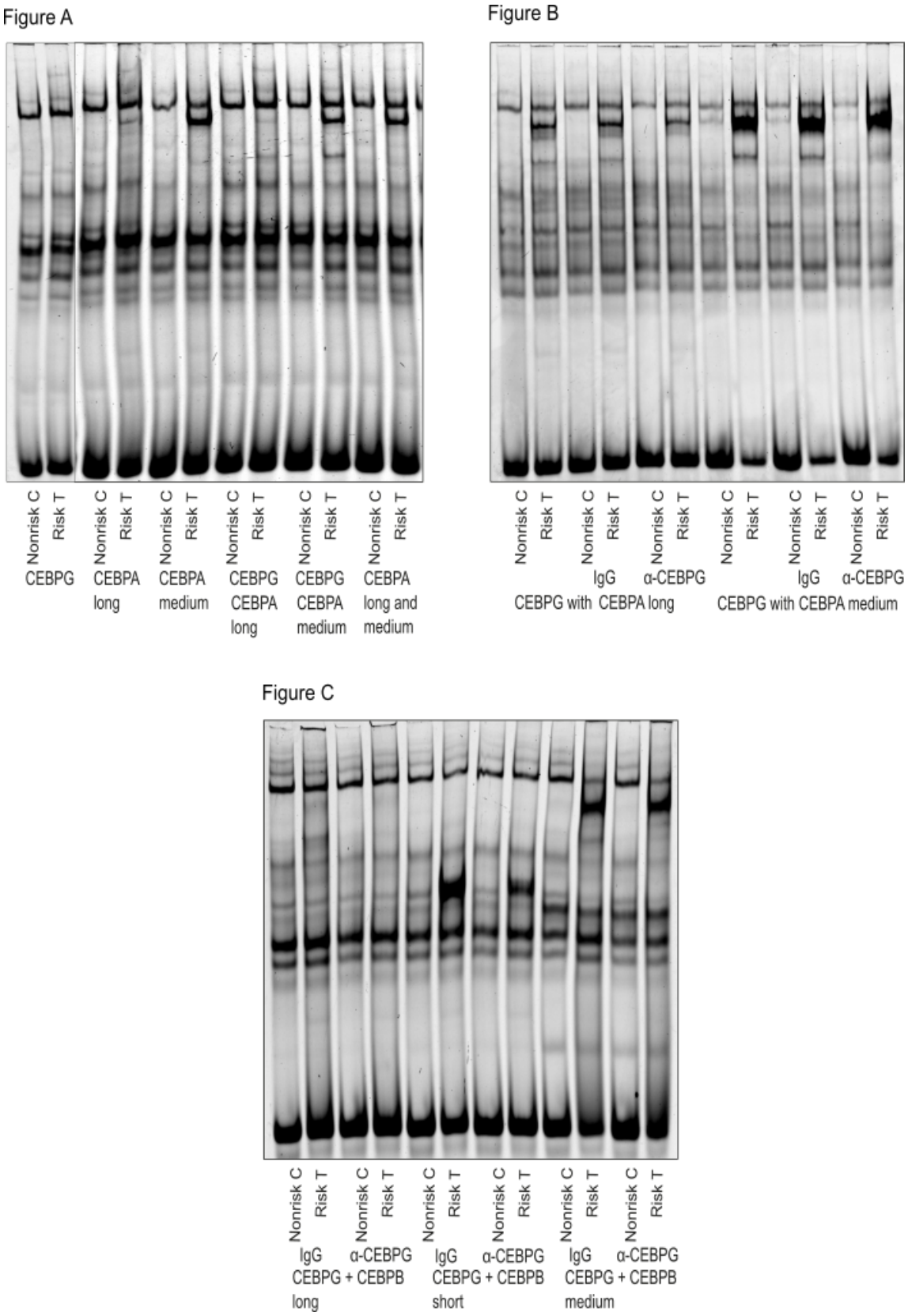


**Figure 5.8:** **A**: EMSA with 3 µg of HEK293T nuclear extracts overexpressing CEBPG, CEPBA long or CEBPA medium. Nuclear extracts were mixed together. **B**: EMSA with 4 µg of HEK293T nuclear extracts overexpressing different members of the CEBP family. For supershift 0.2 µg of CEBPG antibody was added to the reaction. **C**: EMSA with 4 µg of a mix of HEK293T nuclear extracts overexpressing different members of the CEBP family. For supershift 0.2 µg of CEBPG antibody was added to the reaction. All reactions contained 350 ng of Poly(dIdC) and 1.5 ng of a 45 bp cy5-labeled oligonucleotide containing the SNP at mid-position.

**CEBPG**

While in EMSA showing very high T allele specificity in Huh7 and HEK293T overexpressing *CEBPG*, the transient siRNA knockdown of *CEBPG* could not show any significant effects on *TCF7L2* or *ACSL5* in Huh7, HepG2, 1.1B4 or SGBS (**Figure 5.9 A-D**).

**CEBPA**

*CEBPA* was expressed very low in the cell lines used, which lead to difficulties in determining of the knockdown efficiency. In 1.1B4 an 85 % knockdown efficiency could be achieved, however no effect on *TCF7L2* or *ACSL5* expression was detectable (**Figure 5.9 E**).

**CEBPB**

*CEBPB* showed a 30 % decrease of *TCF7L2* ($\mathbf{p = 9*10^{-3}}$) in 1.1B4 with a mean knockdown efficiency of 64 % (**Figure 5.9 F**). Due to transfection difficulties, it was not possible to generate a biological triplicate in the other cell lines.

**CEBPZ**

*CEBPZ* was succesfully reduced in 1.1B4 and HepG2 (knockdown efficiencies: 81 % and 73 %), however this had no effect on expression of neither *TCF7L2* nor *ACSL5* (**Figure 5.9 G, H**).

We were not only interested in expression of total *TCF7L2*, but also whether changes of allele expression could be detected. *Rs7903146* is part of intron 4 of *TCF7L2*, which made it possible to investigate the allelic expression on pre- mRNA level. Since the pre - mRNA levels in the classical cDNA synthesis was below detection level, a preamplification of the region was necessary. A primer spanning over the region of the variant was used as control. There was no difference in *TCF7L2* allele expression between a *CEBPG* knockdown, scrambled siRNA or untransfected cells. This applies for both, HepG2 and 1.1B4 cells (data not shown).

### 5.3.5 Overexpression of *CEBPG* and *CEBPA* in a gene reporter assay showed no allele-specific effect on *TCF7L2* promotor function.

Gene reporter assays are widely used to test whether a genomic region has regulatory effects. Therefore a promotor is cloned upstream of a *luciferase* ORF and the relative light units can be measured. In this case, we were interested in the regulating properties of the genomic region containing the SNP, more precise the differences in regulating abilities between wild type and SNP on the endogenous *TCF7L2* promotor and *TK* promotor. Gene reporter assays concerning the CEBP family were performed with two different luciferase constructs and in two cell lines. For one a luciferase vector containing 3500 bp of WT/SNP-surrounding enhancer with 868 bp *TCF7L2* promotor, for the other a 40 bp WT/SNP-surrounding enhancer fragment, with SNP at mid-position upstream with a *TK* promotor. The *TCF7L2* promotor construct 'long enhancer' was tested in HEK293 and 1.1B4 cells , while the shorter construct 'short enhancer' was tested in HEK293 cells only.

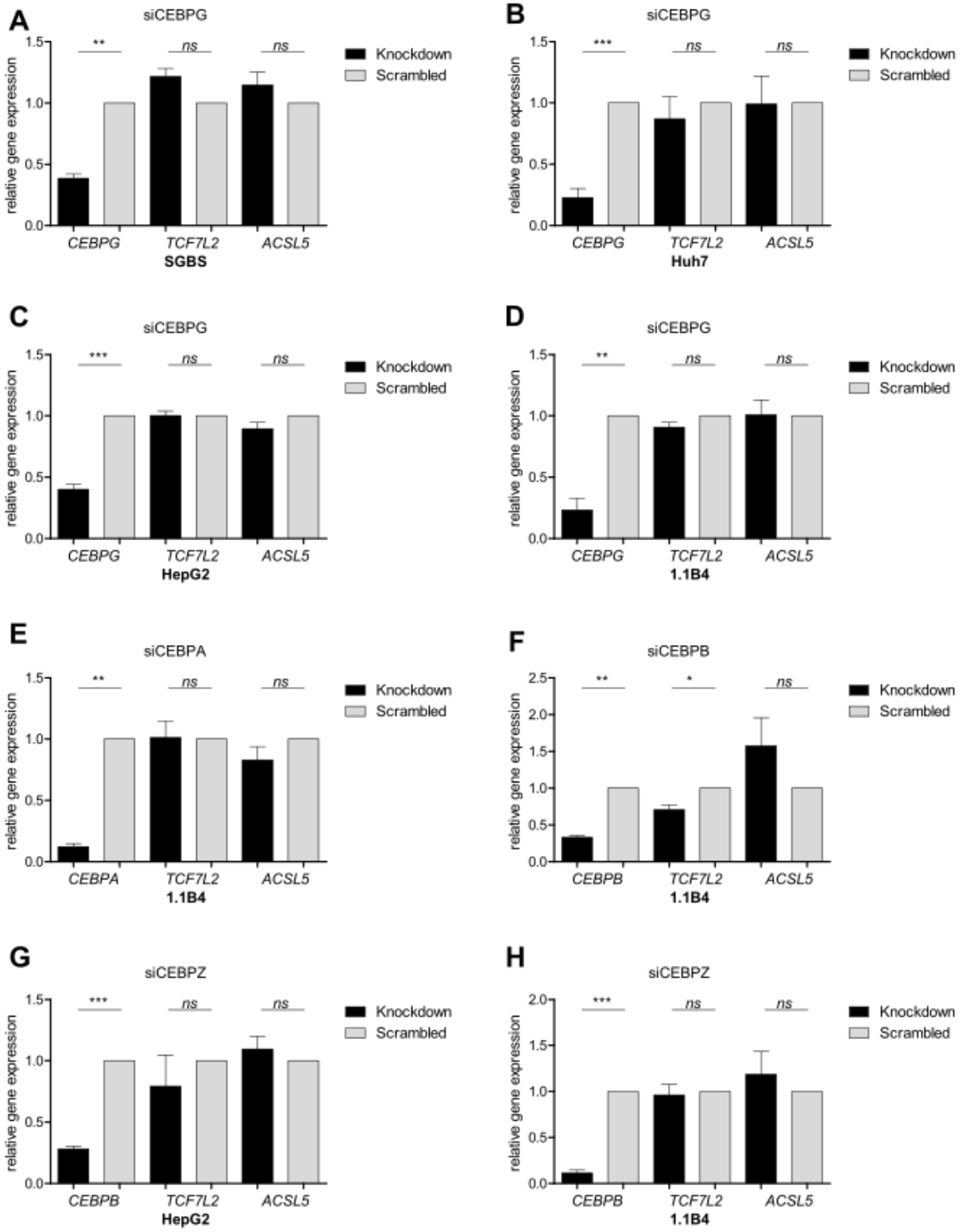


**Figure 5.9:** Transient siRNA knockdown of different CEBPs in SGBS, 1.1B4, HepG2 and Huh7 cells. Included were experiments with knockdown efficiencies > 50%. Displayed are mean relative expressions of *TCF7L2* and *ACSL5* + SEM. One sample t-test was used for analysis, $p < 0.05$= *, $p < 0.01$= **, $p = 1*10^{-3}$= *** . **A**: n = 3, **B**: n = 5, **C**: n = 6, **D**: n = 6, **E**: n = 3, **F**: n = 3, **G**: n = 4, **H**: n = 4.

Gene reporter assay with the long enhancer

**Figure 5.10 A** shows that the long enhancer had a significant activating effect on the *TCF7L2* promotor in HEK293 cells independant of allele. The enhancer increased the expression 2 - 2.7 fold compared to the promotor control, independant of the overexpression of *CEBPG* or *CEBPA*. In 1.1B4 cells, the long enhancer had no pronounced effect on the *TCF7L2* promotor activity, the overexpression of *CEBPG* or *CEBPA* again did not change the allelic expressions. The T allele showed a significant ($\mathbf{p = 0.05}$) reduction in transcriptional activity compared to the promotor control, however, when looking at the effect of 12 % less expression, it was rather low.
CEBPA significantly ($\mathbf{p < 1*10^{-3}}$) influenced the *TCF7L2* promotor itself rather then the enhancer fragment containing the genomic region of interest. This effect could not be rescued by the enhancer fragments (**Figure 5.10 C and D**).

Gene reporter assay with the short enhancer

HEK293 cells were transfected for 24 h with the short promotor constructs and the vectors: *pcDNA3.1 DYK, CEBPG* or *CEBPA*. The 40 bp *TCF7L2* enhancer fragment increased promotor activity by approximately 50 %, independant of wild type or SNP. Overexpression of either *CEBPG* or *CEBPA* showed no influence (**Figure 5.10 B**).

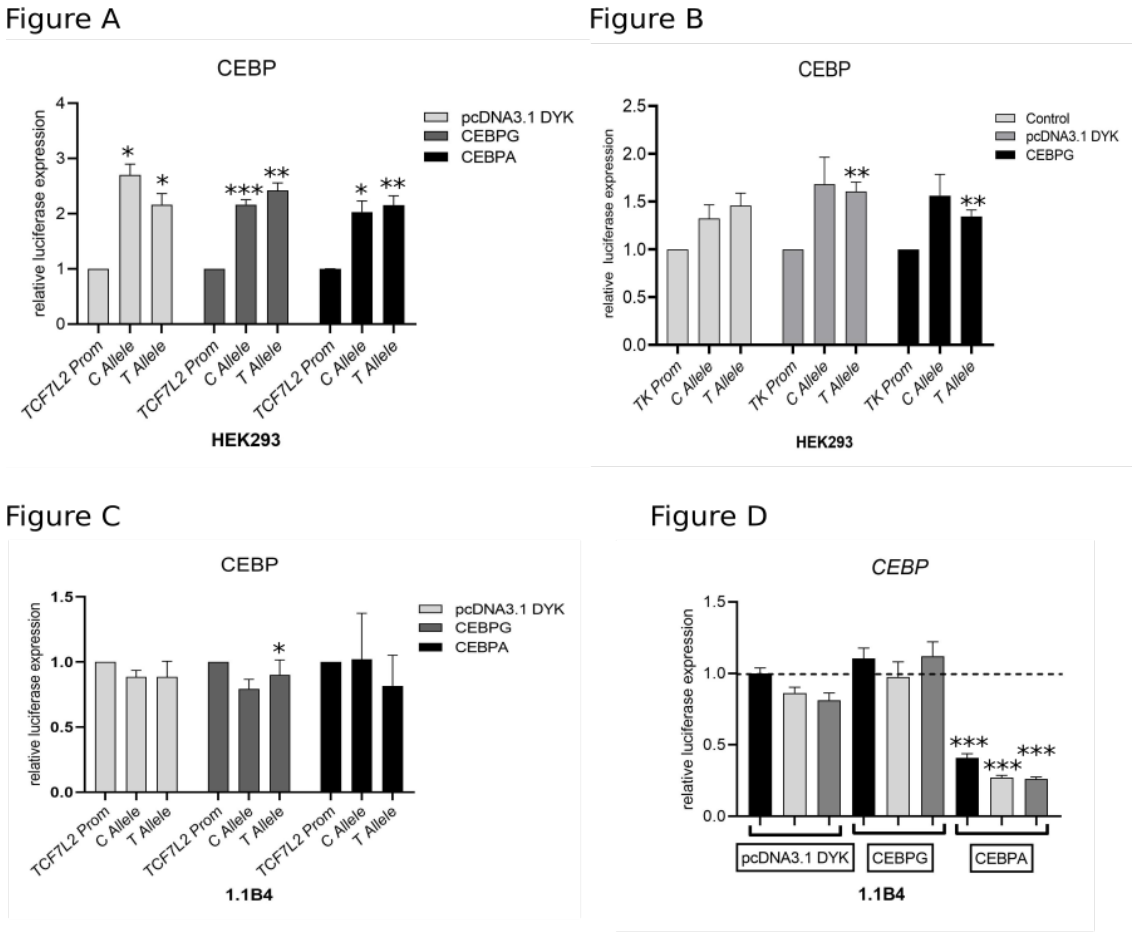


**Figure 5.10:** **A** and **B**: Gene reporter assay investigating the effect of CEBPG and CEBPA on promotor activity in HEK293 cells. *TCF7L2* prom = pGL4.22 vector containing a 868 bp promotor fragment of *TCF7L2*. C/T Allele *TCF7L2* Prom vector containing an upstream 3500 bp fragment of the *TCF7L2* enhancer region with wild type or SNP. *TK* Prom = pGL4.22 vector with an upstream texitTK promotor. C/T Allele= *TK* Prom with a 40 bp fragment of the *TCF7L2* enhancer containing wild type or SNP. Each condition was calculated against each individual *TCF7L2* Prom. **C**: Gene reporter assay in 1.1B4 cells. **D**: Values of the 1.1B4 cells assay compared to *TCF7L2* Prom pcDNA3.1 DYK. Black bars= pGL4.22 vector containing a 868 bp fragment of the *TCF7L2* promotor ("TCF7L2 prom"). Light grey bars or dark grey bars = *TCF7L2* prom with 3500 bp *TCF7L2* enhancer fragment upstream containing the wild type C allele or, respectivley risk T allele. Unpaired t-test (allelic difference) and one sample t-test were used for analysis, $p < 0.05$ = *, $p < 0.01$ = **, $p < 1*10^{-3}$ = ***.

## 5.4 *STAT1*

### 5.4.1 STAT1 binds 5.5 fold stronger to the T risk allele in HepG2 in AC LC-MS/MS.

To identify transcription factors that potentially bind our region of interest, affinity chromatography was performed with nuclear protein of different cell lines. STAT1 bound in the 400 mM salt elution fraction of the affinity chromatography of HepG2 cells to be significantly ($\mathbf{p = 0.03}$) stronger binding to the T risk allele (**Table 5.1**). STAT1 was also identified in the cytosolic and nuclear fraction of all cell lines tested in this thesis. In this assay the protein only appeared to bind the gene locus of interest in HepG2 cells.

### 5.4.2 Binding of STAT1 to the risk locus was confirmed in supershift EMSA.

Since a supershift EMSA did not reveal any shift of bands, *STAT1* was overexpressed in HEK293T, expecting additional bands. However, a simple overexpression of *STAT1* in HEK293T did not show any effect, even though overexpression was verified by Western blot (**Figure A2**). Since it is known that STAT1 has to be activated, in order to translocate into the cell nucleus and to bind to its target regions in the genome, IFNγ stimulation was established (Paragraph 5.1.5). HepG2 cells had to be stimulated with IFNγ for 24 h in order to see STAT1 bands in EMSA. This effect could not be seen in Huh7 cells (data not shown), even though IFNγ stimulation was effective in the cells (Paragraph 5.1.5). An EMSA revealed a non allele-specific band running in the upper region of the gel, which the specific STAT1 antibody shifted (**Figure 5.11**). Another C allele binding band was observed, which was not apparent in non IFNγ stimluated HepG2 nuclear extract. This band was unaffected of the STAT1-specific antibody.

### 5.4.3 Knockdown of *STAT1* in HepG2 and Huh7 shows no significant effect on target gene expression.

After showing binding of STAT1 to the gene locus, we examined the regulatory role. We performed a transient *STAT1* siRNA knockdown in order to find effects on our two target genes *TCF7L2* and *ACSL5*. A *STAT1* knockdown efficiency of 70 % ($\mathbf{p = 0.0011}$) showed no significant effects on *TCF7L2* or *ACSL5* expression in HepG2 cells. No significant regulation of *TCF7L2* or *ACSL5* expression was observed in Huh7 cells (knockdown efficiency 86 %, $\mathbf{p = 0.004}$ ) (**Figure 5.12 A**).

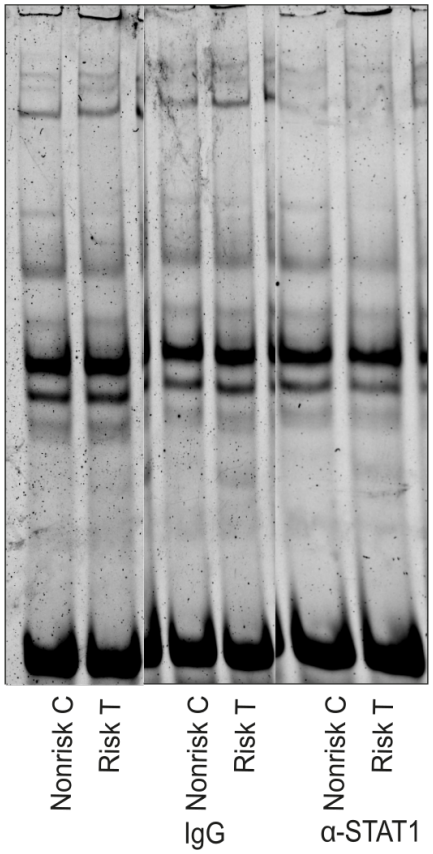


**Figure 5.11:** Supershift EMSA incubating 3 µg of HepG2 nuclear extract with 350 ng of Poly(dIdC), 0.5 µg of antibody and 1.5 ng of a 45 bp cy5-labeled oligonucleotide containing the SNP at mid-position. HepG2 cells were stimulated with IFNγ 24 h before harvest.

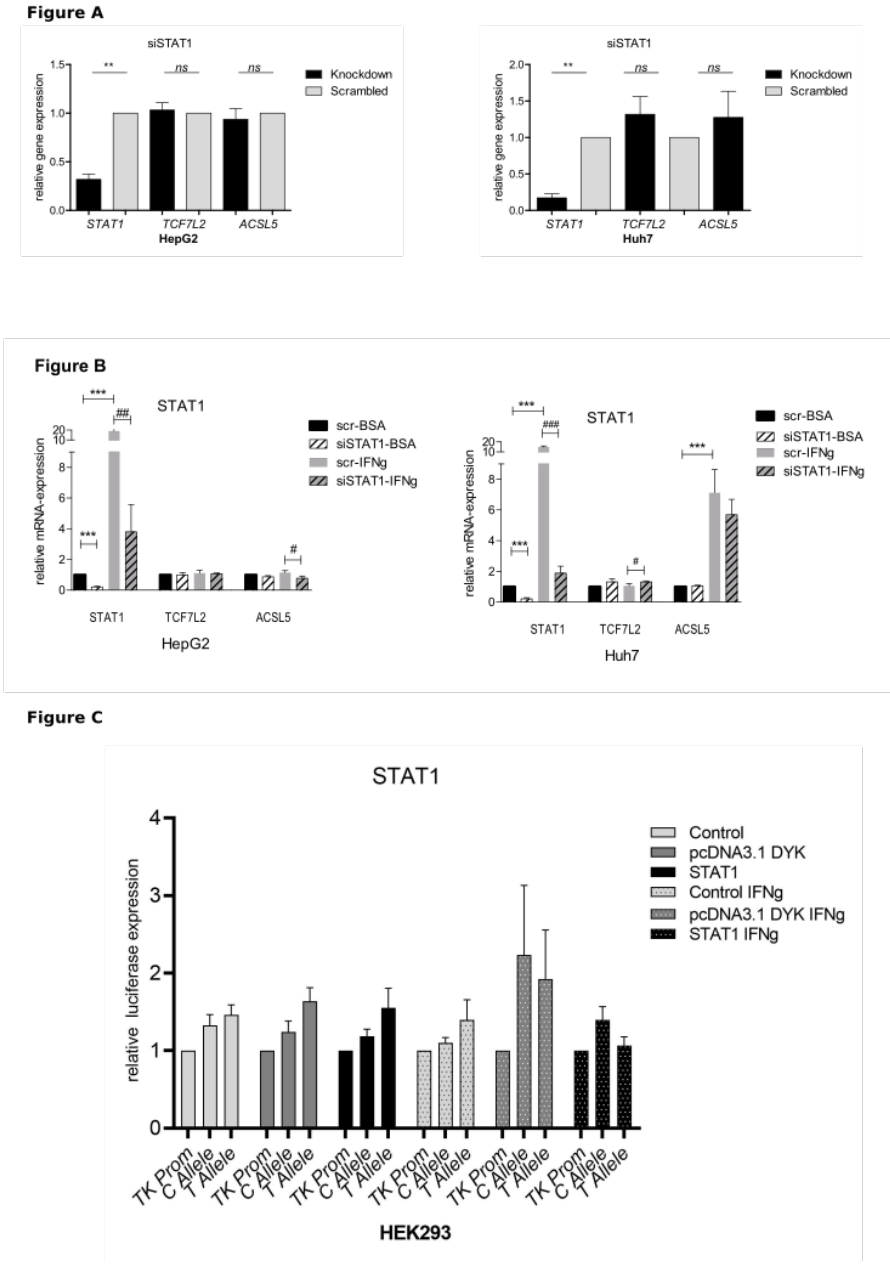


**Figure 5.12:** *TCF7L2* and *ACSL5* expression after transient siRNA *STAT1* knockdown in HepG2 and Huh7 and gene reporter assays investigating effects of *STAT1* on promotor activity.

**Figure 5.12 A**: Transient siRNA knockdown of *STAT1* in HepG2 and Huh7 cells. Included were experiments with knockdown efficiencies > 50 %. Displayed are mean relative expressions of *TCF7L2* and *ACSL5* + SEM. N (HepG2) = 4, n (Huh7) = 5. $p < 0.01$ = **, from one sample t-test. **B**: Transient siRNA knockdown of *STAT1* in HepG2 and Huh7 cells treated with and without IFNγ. Included were experiments with knockdown efficiencies > 50 %. Displayed are mean relative expressions of *TCF7L2* and *ACSL5* + SEM, n = 3. $p < 0.01$ = **, $p < 0.001$ = *** from one sample t-test, and $p < 0.05$ = #, $p < 0.01$ = ##, $p < 0.001$ = ### from students t-test / unpaired t-test. **C**: Gene reporter assay investigating the effect of STAT1 on promotor activity in HEK293 cells. *TK* Prom = pGL4.22 vector with an upstream TK promotor. C/T Allele = *TK* Prom with an upstream 40 bp fragment of the *TCF7L2* enhancer containing wild type or SNP. Each condition was calculated against each individual *TK* Prom control. Values are displayed as mean expression + SEM, n = 3. Unpaired t-test (allelic difference) and one sample t-test were used for analysis.

### 5.4.4 Knockdown of *STAT1* after IFNγ stimulation in HepG2 showed significant effects on target gene expression.

Since we could only find a binding affinity of STAT1 *in vitro* after stimulating the cells with IFNγ, we tried IFNγ stimulation in a transient *STAT1* siRNA knockdown. For this *STAT1* depletion HepG2 and Huh7 cells were treated with IFNγ for 24 h, while the control was not stimulated (BSA). The IFNγ stimulation increased *STAT1* expression 19 fold in HepG2 cells ($\mathbf{p = 0.02}$) and 15 fold in Huh7 cells ($\mathbf{p = 2* 10^{-3}}$). The knockdown signifinanctly diminished the *STAT1* expression in both cell lines and conditions (HepG2 cells unstimulated by 78 % $\mathbf{p = 2* 10^{-3}}$, HepG2 cells stimulated by 80 % $\mathbf{p = 5* 10^{-3}}$, Huh7 cells unstimulated by 80 % $\mathbf{p = 4* 10^{-3}}$, Huh7 cells stimulated by 88 % $\mathbf{p = 1* 10^{-4}}$). In HepG2 cells the knockdown had no effect on *TCF7L2* expression, independant of IFNγ stimulation. *ACSL5* expression, however, experienced a significant decrease by approximately 37 % ($\mathbf{p = 0.02}$). In Huh7 cells, *ACSL5* expression was unaffected by the knockdown independant of IFNγ stimulation, but *TCF7L2* expression was increased by 26 % ($\mathbf{p = 0.05}$) **Figure 5.12 B**.

### 5.4.5 Knockdown of *STAT1* showed no difference in allele expression in HepG2.

There was no significant change in *TCF7L2* allele expression after *STAT1* knockdown. This was independant of a IFNγ stimulation. C and T allele were expressed in similar amounts, leading to a C/T ratio close to 1 (data not shown).

### 5.4.6 Overexpression of *STAT1* in a reporter gene assay showed no allele specific effect on gene expression.

To see whether STAT1 regulates the promotor function via the SNP-containing enhancer, a gene reporter assay was perfomed. This assay was performed in HEK293 cells with short promotor constructs by Andrea Tóth (**Figure 5.12 C**). We included a 24 h IFNγ stimulation as well. No significant effects could be detected. The short enhancer caused an elevated activity of the *TK* promotor. The C allele increased expression by 32 %, while the T allele increased it by 46 %. This effect could also be observed when the backbone vector control pcDNA3.1 DYK was added to the equation. In this case, the C allele increased luciferase expression by 24 %, the T allele by 64 %. *STAT1* overexpression did not affect these effects significantly. While the C allele caused an 19 % increase in expression, the T allele showed an 55 % increase. When stimulating with IFNγ, the T allele showed an tendency to activate more than the C allele, in the stimulated cells overexpressing either pcDNA3.1 DYK backbone vector or *STAT1*, this effect appeared to be reversed. In the untransfected control, the C allele did not cause any change in expression compared to the promotor control, while the T allele led to a 40 % increase. When overexpressing the control pcDNA3.1 DYK, the C allele caused an 123 % increase in expression, while the T allele caused an increase of 93 %. The variation between the replicates was high. When overexpressing *STAT1*, the C allele caused a 40 % increased expression, while the T allele did not show any effect on promotor function.

### 5.4.7 STAT1 ChIP shows a binding to the *rs7903146* locus.

We performed chromatin immunoprecipitation in HepG2 cells. When stimulating the cells with IFNγ, an increase in binding of STAT1 to the locus was observed from 8 fold binding of STAT1 (unstimulated) to the locus to 38 fold increased binding of STAT1 to the *rs7903146* locus compared to isotope control ($\mathbf{p = 0.02}$) **Figure 5.13**.

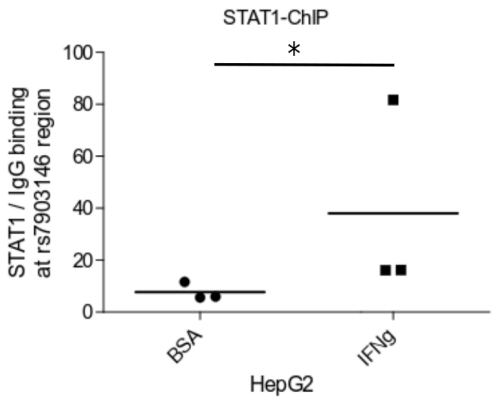


**Figure 5.13:** ChIP performed in HepG2 and IFNγ stimulated HepG2 cells with a STAT1 specific antibody. Displayed is the mean ratio of of anti-STAT1 / anti-IgG isotype control antibody binding at the *rs7903146* genomic region; + SD, n = 3. Unpaired t-test was used for analysis of significance, $p < 0.05 = *$.

## 5.5 *HMGA2*

### 5.5.1 HMGA2 was identified in AC LC-MS/MS to bind 2.9 fold stronger to the T risk allele in Huh7.

HMGA2 bound to the intronic TCF7L2 locus in HepG2, Huh7, 3T3L1 and 1.1B4. In the 400 mM salt elution fraction of the Huh7 trapping, there was a significant 2.9 fold increased binding to the T risk allele. In 1.1B4 cells, there was also a 2.5 fold increased binding and in HepG2 cells a 1.8 fold stronger binding to the T risk allele. In 3T3L1 cells, the binding to the risk allele was 1.64 fold increased but insignificant there (**Table 5.3**). HMGA2 was not identified in Ins1 cells, which coincides with EMSA data.

**Table 5.3:** HMGA2 in affinity chromatography 400 mM.

| **Cell line** | **Fold change** T/C allele | **Binding allele** | **p value** | **Repl** |
|---|---|---|---|---|
| Huh7 | 2.87 | T | **2*10$^{-3}$** | 6 |
| HepG2 | 1.79 | T | 0.89 | 4 |
| 1.1B4 | 2.52 | T | 0.057 | 6 |
| 3T3L1 | 1.64 | T | 0.34 | 4 |

### 5.5.2 Binding of HMGA2 to the T risk allele was confirmed in EMSA and supershift EMSA in several cell lines.

**Figure 5.14 A** shows EMSAs in 14 different cell types. In every cell type except for Ins1 and HEK293 there was a T risk specific band appearing in the lower third of the gel. However, EMSAs with HEK293T cells overexpressing transcription factors of interest later on, the HMGA2 band could not always detected (see other sections). Exact conditions can be found in **Table A2**. **Figure 5.14 B** shows similar results with a 121 bp long oligonucleotide. Due to its high instability, the 121 bp oligonucleotides were not further used in EMSA experiments. **Figure 5.15** shows a shift of the band containing HMGA2 in eight different cell lines. When overexpressing *HMGA2* in HEK293T cells, there were bands in both alleles visible, but more pronounced in the T risk allele. When performing a supershift with a HMGA2-specific antibody, the band in the C nonrisk allele was shifted, but not in the T risk allele. This could be due to an excessive amount of HMGA2 so that the concentration of the antibody was too low (**Figure 5.14 C**).

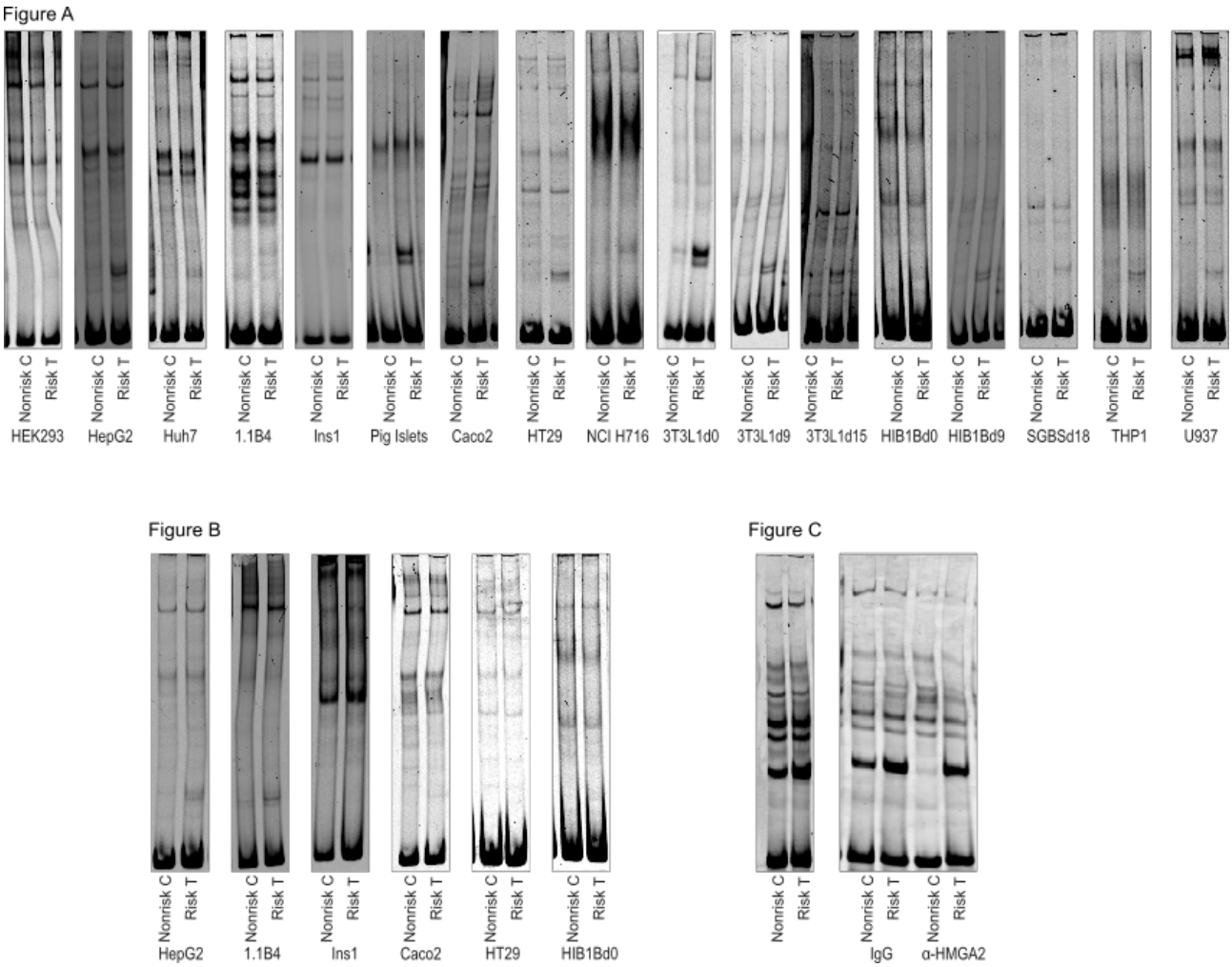


**Figure 5.14:** **A**: EMSA incubating nuclear extracts from different cells lines with a 45 bp cy5-labeled oligonucleotide containing the SNP at mid-position. **B**: EMSA incubating nuclear extracts from different cells lines with a 121 bp cy5-labeled oligonucleotide containing the SNP at mid-position. **C**: EMSA incubating 3 µg of HEK293T nuclear extracts overexpressing HMGA2 with 350 ng Poly(dIdC) and 1.5 ng of a 45 bp cy5-labeled oligonucleotide containing the SNP at mid-position. For supershift 0.5 µg of HMGA2 antibody was added.

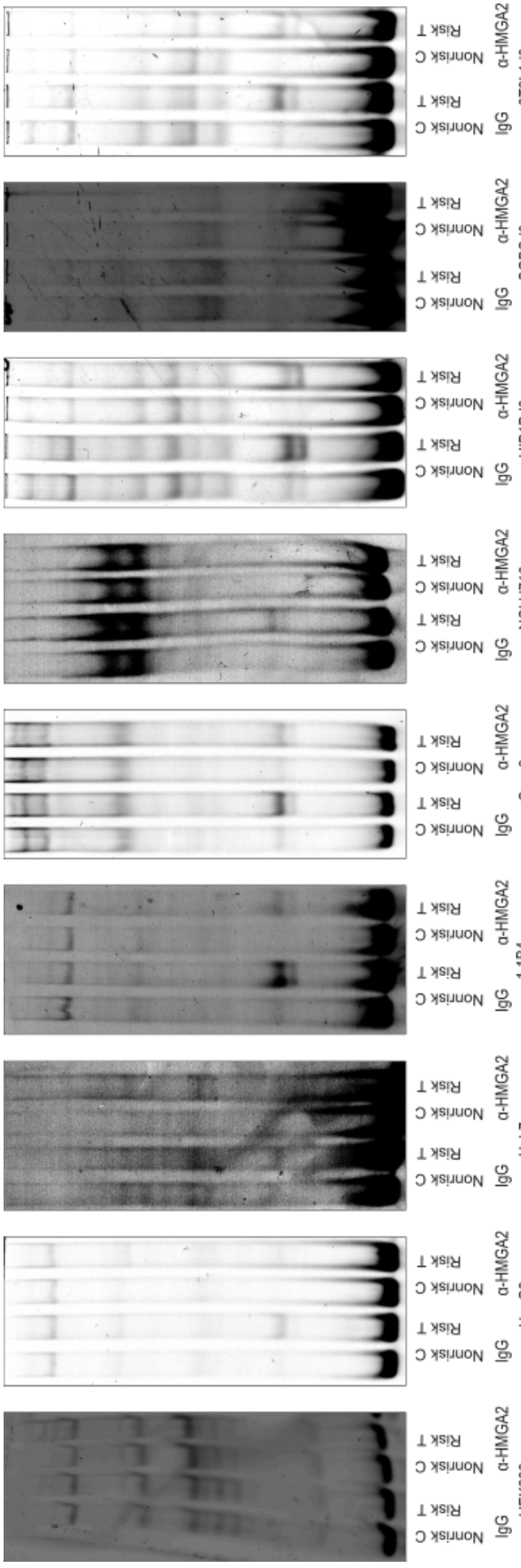


**Figure 5.15:** EMSA incubating 3 µg of various nuclear extracts with 350 ng of Poly(dIdC) and 1.5 ng of a 45 bp cy5-labeled oligonucleotide containing the SNP at mid-position. For supershift 0.4 µg of HMGA2 antibody was added.

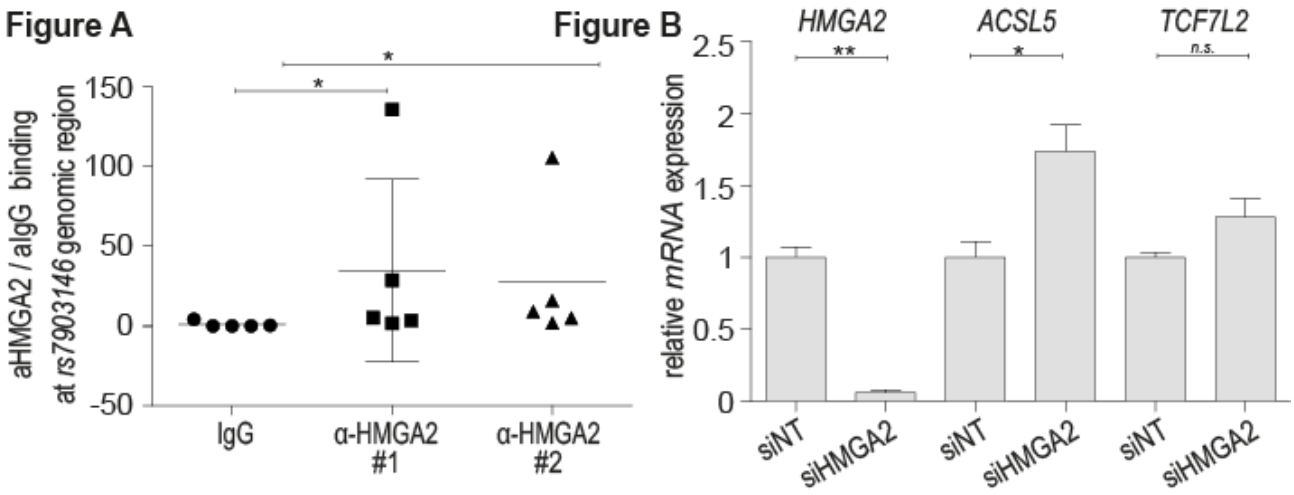


**Figure 5.16:** **A**: HMGA2-ChIP performed in 1.1B4 cells, n = 4. Values are displayed as mean ± SD. Mann Whitney test was used for analysis, $p \geq 0.05$ = *. **B**: Transient siRNA knockdown of *HMGA2* in 1.1B4 cells. Values are displayed as mean relative expression of *TCF7L2* and *ACSL5* + SEM, n = 3. One sample t-test was used for analysis, $p < 0.05$ = *, $p < 0.01$ = **.

### 5.5.3 Knockdown of *HMGA2* in 1.1B4 cells showed a significant increase in *ACSL5* expression.

To investigate regulatory functions of HMGA2, we applied a transient siRNA knockdown. In 1.1B4, an *HMGA2* knockdown showed a 74 % increase of *ACSL5* expression (**p = 0.03**), and no significant increase of *TCF7L2* expression (28 %, p = 0.08). Mean knockdown efficiency was determined as 93 % (SD = 1.9 %) (**Figure 5.16** ).

### 5.5.4 HMGA2 binds at the genomic region *rs7903146* in 1.1B4 cells

ChIP with two different HMGA2 antibodies revealed a significant binding at the genomic region of *rs7903146*. We observed a significant 35 - or 28 - fold ( $p \geq$ **0.05**) enrichment, respectively, as compared to the respective isotype controls (**Figure 5.16**).

## 5.6 Other transcription factors of interest

This section addresses transcription factors for which a significantly higher affinity for a *TCF7L2* allele has also been identified, or transcription factors that have been selected according to other criteria.

### 5.6.1 Knockdown of a selection of potential regulators revealed promising candidates for further anaylsis.

A large scale transient siRNA knockdown approach was performed after the the identification of potential regulating candidates in affinity chromatography coupled LC MS/MS. In total, 33 proteins / genes were analyzed, including the 11 proteins prioritized as described in **Table 5.1** and additionally other genes annotated for example as transcriptional cofactors, by knockdown in diverse cell lines including HepG2, Huh7, 1.1B4. However, none of these additionally analysed genes showed a significant influence on expression of the target genes *TCF7L2* or *ACSL5* (data not shown).

### 5.6.2 HMGA1 and SATB1 bind T allele specific, while LHX2 binds non allele specific to the region of interest *in vitro*.

A variety of potential transcription factor candidates from the AC LC-MS/MS experiments were overexpressed in HEK293T cells in order to investigate *in vitro* binding affinity towards the nonrisk and risk allele. Overexpression was verified over Western blot (**Figure A2**). **Figure 5.17 A** shows the binding behavior of the succesfully overexpressed proteins. MYEF2, PURB, NR2F1, PARP1, GTF2B, PAX6 and ZNF593 were all proteins that under these EMSA conditions did not bind to the region at all. HMGA1 appeared as a T allele specific band in the lower third of the gel, at about the same height as HMGA2 (see Paragraph 5.5). SATB1 displayed two intense bands in the upper region of the gel at the T allele. LHX2 showed pronounced bands in the middle part of the gel, without any apparent allele specificity ( mean C/T = 1.14, SEM = 0.035, n = 6). In **Figure 5.17 B** the LHX2 bands could be shifted with a LHX2-specific antibody. When adding two different consensus sequences of PAX6 (MATBASE annotated) to Ins1 extract, only PAX6(2) showed effects. These sequences competed away bands in a dose-dependant manner (**Figure 5.17 C**). In **Figure 5.17 D** the sequences of both PAX6 competitors are displayed.

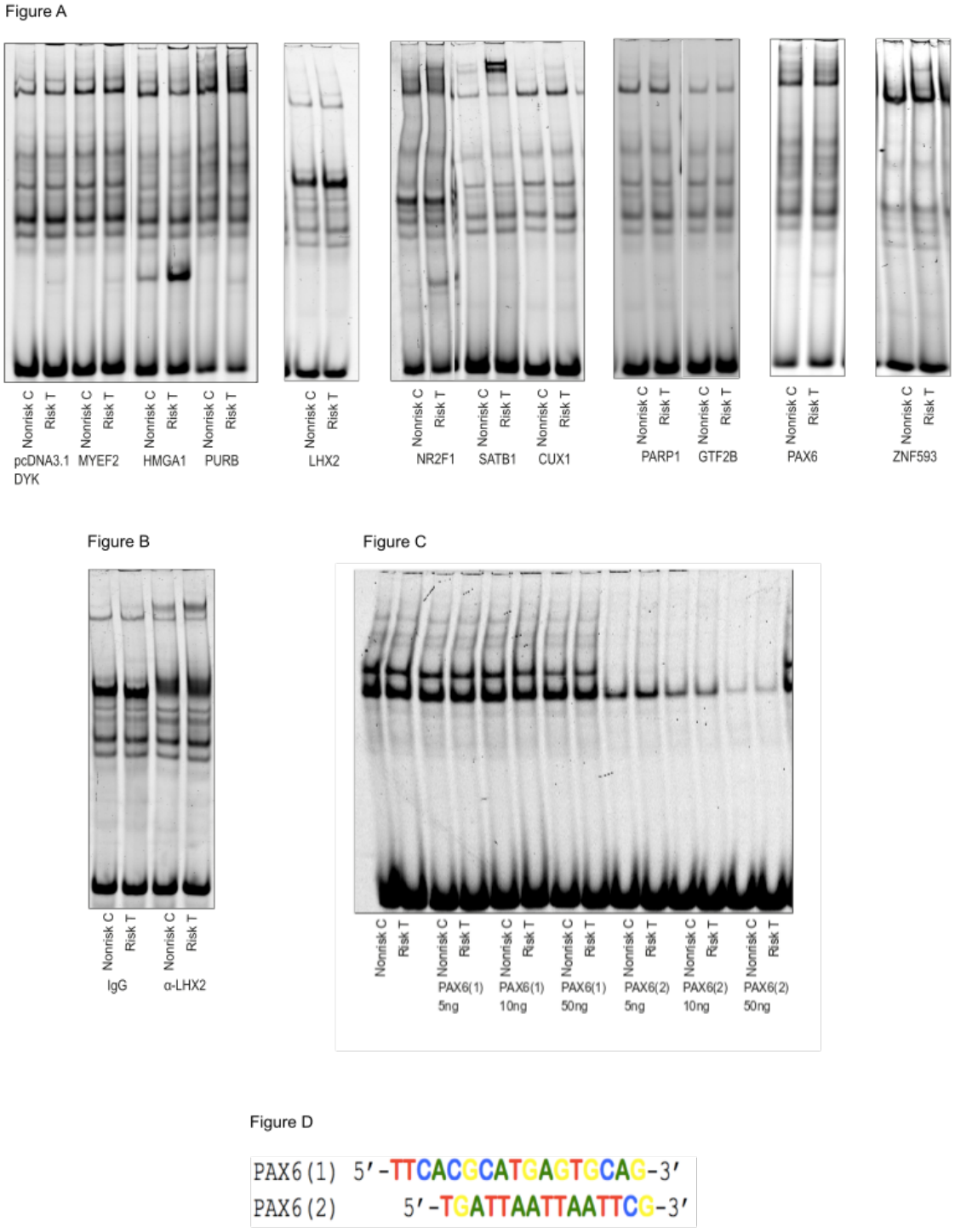


**Figure 5.17:** **A**: EMSA incubating 3 µg of HEK293T nuclear extracts overexpressing different proteins with 350 ng of Poly(dIdC) and 1.5 ng of a 45 bp cy5-labeled oligonucleotide. **B**: Supershift EMSA with the same conditions, except 0.2 µg of antibody was added to the reaction. **C**: EMSA incubating 3 µg of Ins1 nuclear extract with 700 ng of Poly(dIdC) and 1 ng of a 45 bp cy5-labeled oligonucleotide. Two different cold-labeled consensus sequences were tested in 3 different quantities: 5 ng, 10 ng and 50 ng.**D**: Consensus sequences of PAX6(1) and PAX6(2) used for the competition EMSA.

# 6 Discussion

## 6.1 Unraveling molecular mechanisms underlying genotype-phenotype associations in type 2 diabetes

Mendelian diseases are usually easily identified, since one defect gene is responsible for the phenotype. Linkage analysis with the help of family pedigrees are used to identify the causal genes. Very few genes in the context of T2D were identified over linkage analysis, due to small individual effects of many variants at different loci contributing to a risk phenotype. Due to this problem, the genes identified so far represent only a small part of the inheritance in common diseases [155]. The small individual effects are due to the location of the variants. The majority of variants associated with T2D were found in non-coding regions, so research needs to focus on finding better methods unraveling their regulatory role in the disease phenotype. While a coding variant is often responsible for a change in amino acids and therefore protein structure and function, variants in non-coding regions modulate genes in a more complex and subtle way. These variants can influence gene expression in manifold ways by e.g. (1) causing differential binding of transcription factors, hence changing the composition of the transcription initiation complex, (2) influencing chromatin structure and therefore accessibility of the DNA, (3) initiating different splicing patterns of genes or (4) generating or destroying long noncoding RNAs [156] or microRNAs [157]. The mammalian genome is comprised of around 23000 genes with approximately one million enhancer regions [158]. Many non-coding variants lie within these *cis*- and *trans*-regulatory elements of the genome. In these regions, transcription factor binding is one of the main mechanisms how gene expression is controlled. Enhancers and promotors assemble general transcription factors (GTFs), PolII and elongation factors leading ultimately to the correct formation of the transcription-initiation complex. Hence SNPs in these regions can cause different expression of genes, potentially shifting towards a higher disease risk. These mechanisms of protein/DNA binding are highly dependant on accessibility of the DNA. DNA accessibility again is dependant on chromatin architecture. Chromatin density is regulated over histones, which can be modified by enzymes and other proteins [159]. Therefore, proteins that govern the histone structure of the genome influence transcription factor binding, regulating the approachability to the nucleic acids. The DNA is not rigid, but can bend and interact with different regions of itself, making it more difficult to determine potential targets of non-coding mutations. Enhancer regions far away from transcriptional start sites of genes can bend onto their promotor to assemble the transcription-initiation complex. Methods to determine these 3D chromatin structures are among others: FISH, 3C, 4C, 5C and hiC (summarized in [160]). Some variants lie within splice donor or acceptor sites, leading to a different splicing pattern of the pre-mRNA. This can ultimately lead to different protein structures but also more or less protein content [161]. This complex and spacious interplay is not easy to grasp with todays methods available. Bioinformatics play a key role in the pre-selection of potential genetic candidates, the exact mechanism however needs to be still determined experimentally. One bioinformatic tool is PMCA, where cross-species analysis of transcription factor binding site modules can help identifing relevant

*cis*-regulatory variants. Instead of looking at the level of regulatory sequences of each TFBS, the method comprehends the phylogenetic conservation of TFBS co-occurrence [162]. Once *cis*-regulatory variants are identified, the target genes need to be evaluated. Many enhancer elements do not lie within the direct vicinity of the regulated gene. Potential regulated genes can be covered by RNA-seq analysis of experiments such as overexpression or knockdown. The gold standard to investigate the properties of a variant is homologous directed repair with CRISPR-Cas9, where a targeted single nucleotide exchange is created with the otherwise uniform genetic background. In this way, for an ideal comparison, an RNA-seq can be performed with a cell that is identical except for one nucleic acid exchange. While, in the course of this thesis, we were not able to generate a cell line with all three genotypes of our risk variant *rs7903146*, we took recourse to somewhat older and more established assays to investigate the gene-regulatory role of *rs7903146* in the context of a T2D risk phenotype. These methods will be critically dicussed in the following sections.

### 6.1.1 Identifying potential regulating proteins with affinity chromatography coupled LC MS/MS

We applied an affinity chromatography assay to identify novel proteins binding to the DNA sequence of interest [163]. We incubated these beads with nuclear protein extracts of the tissue of interest together with an unspecific DNA competitor to reduce unspecific binding. This method has an artificial setting, since we incubated only small DNA fragments with protein at the unphysiological temperature of 4 °C. This could pose an advantage, reducing the complexity. Since protein/DNA bindings are dependant on NaCl concentrations, we performed a salt elution to separate the specific proteins from the bead-coupled DNA. These elutes were measured in label-free quantification via LC-MS/MS (**Figure 6.1**). The label-free quantification holds various advantages over SILAC (stable isotope labeling in cell culture) or chemically-labeled quantification methods. Alternatives are iTRAQ (isobaric tags for relative and absolute quantification) or TMT (tandem mass tags). All these labeling techniques are time-intensive and expensive compared to label-free proteomics. For example, labeled proteomics needs a dual labeling approach for quality ensurance. This doubles the amount of samples being measured [164]. A comparison of SILAC and label-free proteomics was perfomed by the Merl *et al.* in 2012 [165]. For reviews about proteomics see Terzi *et al.* [166] and Pappireddi *et al.* [167]. Claussnitzer *et al.* and Lee *et al.* already successfully unraveled the molecular mechanisms of non-coding variants in the *PPARG* locus with label-free proteomics [162, 168]. Since this method was well evaluated and established, we chose it for our identification of DNA-binding proteins at our locus. **Figure 6.1** displays an overview of the workflow used in this thesis, as explained above.

### 6.1.2 Investigation of *in vitro* protein/DNA binding behavior via EMSA

The electrophoretic mobility shift assay is an artificial method to determine protein/ DNA binding. Therefore the DNA sequences of interest are labeled either radioactively or fluorenscently and incubated with a protein mix, usually nuclear or whole protein extracts. This is then applied on an acrylamidgel and the DNA/protein mixture is separated by electrical charge. After separation the labeled DNA can be visualized. To evade unwanted unspecific proteins blocking the DNA and thereby preventing the interaction of specific protein/DNA complexes, DNA competitors are added. For the stabilization of the pH, a gel binding buffer is used. Since the protein still holds its native structure, the migration pattern does not necessarily reflect protein size. We performed EMSAs at 4 °C, which

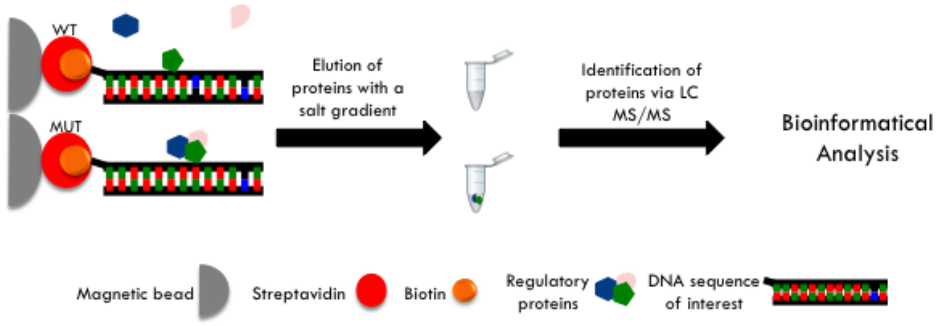


**Figure 6.1:** Schematic workflow of affinity chromatography coupled with LC MS/MS

is far off the physiological temperature of 37 °C. We choose this low temperature to slow down the kinetics of equilibrium reactions and to prevent degradation of protein. The ion composition of the reactions are also often not in physiological range, and the tertiary and quartiary structure of the DNA are completley neglected. In addition, the exact binding position of proteins cannot be determined. With nuclear extracts in which certain proteins are overexpressed, we have a high disbalance in the protein composition, which does not reflect *in vivo* conditions in the nucleus [153]. Even if the method has disadvantages, it delivers fast and cost-effective results. EMSA is a useful tool for getting preliminary results. One can easily test different DNA fragments with various nuclear protein extracts of different tissues. In our case, we also performed EMSA to find appropiate cell line extracts for our initial experiment - the affinity chromatography. Continuative methods such as the supershift assay are more valuable because the binding pattern of an identified protein can be demonstrated and determined. In addition, binding kinetics of proteins can be evaluated via EMSA.

### 6.1.3 Transient siRNA knockdown as a tool to capture downstream effects

In order to find genes that are regulated by a certain protein, it is possible to silence the potential regulator. Science has made use of an endogenous process of gene expression regulation: small interfering RNA. The so-called RIS complex (RNA interference silencing complex) uses the short, single-stranded antisense RNA as a template, recognizes the target mRNA and ultimately leads to its post-transcriptional degradation. Two methods are commonly used: The delivery of synthetic double-stranded RNA or plasmid DNA that encodes a short hairpin RNA (shRNA). While the shRNA needs to enter the nucleus for transcription, the siRNA can act directly in the cytoplasm. The advantage of shRNA is the high expression, which is not present when you add a ready translated RNA. McAnuff *et al.* found shRNA to be more potent [169]. The siRNAs can be ordered pre-tested for efficiency or designed individually with commercially available bioinformatic tools. The efficiency of a knockdown is strongly dependend on the target gene sequence, so most companies offer mixes of several siRNA to ensure success. The mRNA and protein stability of the regulator and the target are also an im-

portant factor that must be taken into account when defining incubation times. A big disadvantage of a transient knockdown is the high variability between the biological replicates, since efficiency can scatter. *siRNA* delivery is constantly optimized, especially for potential therapeutic use, summarized by Han *et al.* [170]. An alternative to the variable transient siRNA knockdown is the stable lentiviral knockdown. It has the advantage of a robust knockdown efficiency, which facilitates the execution of follow-up experiments. However, the lentiviral construct stably integrates randomly into the genome, which can potentially lead to unintended off-target effects [171]. A variant of the lentivrial knockdown is the inducible knockdown, which enables a gradual expression and thus gradual gene-regulatory effects of common variants on phenotypes [172]. Transient siRNA knockdown is fast, cost-effective and therefore suitable for screening purposes. In this PhD thesis, we tested almost 40 different transcription factors and cofactors in three cell lines in a transient siRNA knockdown to narrow down potential candidates. However, transfection efficiency varied between cell lines and proteins, making it impossible to generate three successful biological replicates for all. Nevertheless, the choice of method was suitable for our purposes.

### 6.1.4 Analysing regulatory effects of transcription factors with gene reporter assays

The luciferase assay principle is based on the ability of luciferase enzyme of the firefly *Photinus pyralis* to cleave D-luciferin and therefore emit light. A control luciferase of the sea pansy *Renilla reniformis* has a similar function with the substrate coelenterazine at another pH. In this way, it is possible to control for transfection efficiency within one sample. Since the light emission is relative to the gene expression, one can investigate promotor function and activity [173]. Therefore the DNA sequences of interest are cloned in a vector upstream of the *luciferase* gene. Cells are then transfected with this vector and the Renilla control vector and relative light units are measured and compared between the DNA fragments. In the case of gene variants lying within an enhancing region, a promotor of choice has to be downstream of the investigated enhancer fragment. The choice of promotor poses a first potential pitfall. Most of the time a strong promotor is used to get sufficient signals, but one has to ask whether the enhancer region might have promotor specific effects. The use of a promotor of the gene of interest might lead to low expression and hence low signal intensity. In our case, we tested both. We used the promotor of the *thymidine kinase*, since it has a strong and reproducible activity in various cell lines. Additionally, we used an 868 bp long promotor fragment of *TCF7L2*. The next difficulty with this type of assay is the missing chromatin structure. In our case, the variant of interest lay within the intron 4, suggesting a looping mechanism, which is not necessarily present in a classical vector construct. Bringing a vector into a cell is a very artifical concept, forcing the cell to express massively *luciferase*, which disturbs normal metabolism. The choice of cell model is also crucial, since missing or overrepresentation of transcription or co-factors could bias the results. However, the assay does give a first impression, whether a genomic region of interest has regulatory function, and whether genetic variants influence these properties. The assay can also be applied for validation of transcription factor influence on promotor/enhancer by overexpression or silencing of the factor. An alternative poses the minigene analysis, provided the variant of interest is in an intron or close to the TSS [174]. Although it had been reported that the T allele leads to increased expression, we were not able to see this effect in the luciferase assay with our constructs in our cell lines. The addition of transcription factors, which potentially regulate expression, did not induce any allele specificity either. We conclude that a gene reporter assay is not suitable for our specific research question.

### 6.1.5 Chromatin immunoprecipitation as a tool for investigating *in vivo* binding

Chromatin immunoprecipitation (ChIP) is a method to determine the *in vivo* DNA-binding behaviours of proteins. Cells need to be fixated with formaldehyde. After gDNA isolation and shearing of the chromatin, the sample is incubated with an antibody coupled to protein G beads. This way, genomic regions fixed with the protein of interest are bound to the antibody-coupled beads and can be washed. In the final step, the genomic DNA is dissipated from the beads and protein is digested. The gDNA can then be measured quantitatively by qPCR or gel electrophoresis. As control, a mixture of random antibodies of the same isotype is used (IgGs). Enrichment is compared to the input gDNA and the IgG [175]. Determination of allele-specific protein/DNA interactions via ChIP has become a common tool in the investigation of regulatory variants [168, 176]. ChIP is also ideal for studying epigenetics by determining histone modifications. ChIP has advanced into ChIP-seq to determine all of the loci bound by the protein of interest. A overview of the latest advances in ChIP technology has been summarized by Collas [177]. ChIP is very prone to methodological errors. The incubation time of formaldehyde is crucial because too short a time does not fix the protein of interest to the DNA, while too long leads to false positive results. The false positive results are due to a coincidental proximity of the protein to DNA segments. The fragmentation of the gDNA has to be highly standardized for each cell type in order to obtain reproducible fragment sizes. However, the most delicate task is to find a suitable antibody. For many epigenetic markers and common transcription factors, companies offer ChIP-trialed antibodies. Once all pitfalls are overcome, ChIP is an excellent tool to explore protein / DNA binding in a physiological setting.

### 6.1.6 Choice of cell type

Finding a good cell model for genotype - phenotype related mechansims is always a challenge. The best model is primary tissue material of patients and controls. However, the limited availability, the genetic heterogenitiy and the small amounts of cell material pose disadvantages towards this option. Large genetic heterogenity implicates the need for large sample sizes. In this thesis, we could not aquire such material. Cell lines have a known genetic background, an almost unlimited potential for cell division, and are easily accessible and managable. Especially with respect to nuclear protein for affinity chromatography, cell lines offer an advantage, as sufficient material can be generated within short time. The main disadvantages of cell lines are their abnormal chromosomal content, genetic mutations, altered metabolism and missing cell to cell interaction [178]. We wanted to investigate our variant in several different tissues: liver, β-cell, adipocytes and intestine. Since T2D affects many different tissues, we wanted to find out in which tissue our variant influences gene expression. In the course of the thesis, however, we decided to focus on β-cells and liver cells. β-cells are part of the Islets of Lagerhans found in the endocrine pancreas. They secrete the hormon insulin and, together with α-cells, primarily responsible for blood glucose homeostasis. Proteins that lead to a β-cell specific phenotype are among others PDX1, NKX6.1 (Homeobox protein Nkx-6.1), ISL1, NKX2.2 (Homeobox protein Nkx-2.2) and PAX6 [179]. A mature β-cell additionally expresses e.g. MAFA, urocortin 3 (UCN3) and estrogen-related receptor gamma (ESRRG) [180–188]. A β-cell couples nutrient metabolism with electrical processes. A glucose concentration higher 7 mM results in depolarization and consecutive insulin release [189, 190]. The glucose stimulated insulin secretion is one of the most important mechanisms in a β-cell and mandatory for the investigation of T2D related phenotypes. There are five mouse, eight rat and nine human β-cell lines commonly used [178]. Ins1 cells are responsive within normal

physiological range regarding GSIS, have a high insulin content and express glucokinase [134]. This makes the cell line a good model to study insulin-dependent processes. Unfortunately, the cell line is derived from the species *rattus norwegicus*, which calls into question the relevance of the results for a human disease. Human cell lines do not have the problem of the missing genetic background for T2D, but these cell lines all have little to no responsiveness to glucose in the context of insulin signaling. The cell line 1.1B4 was characterized by McCluskey *et al.* in the context of GSIS properties [191]. Further studies revealed the ability to stimulate with GLP-1 and GIP with increased expression of *INS, PCSK1, PCSK2, PDX1, GLP1R and GIPR* [192]. Vasu *et al.* reported, that prolonged exposure to high glucose levels (25 mM) resulted in decreased *INS, GCK and PDX1* levels [193]. Although this research group was able to detect *PDX1* expression, we were not. Of note, after finalizing both, the practical work and also drafting of the this thesis, a very recent report showed that in fact 1.1B4 cells do not retain the phenotype expected of human β-cells [194] and charges of the cell line may be contaminated with cell of rat origin. Thus, intriguingly we were not able to identify β-cell specific factors in this cell line. Yet, identification of binding proteins by the used in vitro ACMS approach and EMSA experiments still may find some relevant factors. However, it has to be stated that all functional follow up studies reported here for the 1.1B4 cell line have to be interpreted with caution! The novel human β-cell line EndoC-β H1 is a very good model, due to its low cell division rate, it is more suitable for experiments only with the most promising candidates or therapeutic research [195]. HepG2 is a cell line that has been used extensively in research of various liver-related topics. However, Nagarajan *et al.*analysed the lipid and glucose metabolism of hepatocyte cell lines and primary mouse hepatocytes and concluded that HepG2 is not a good model to explore these T2D-related metabolic pathways. They revealed striking differences in glucose incorporation rates, fatty acid oxidation and *de novo* lipogenesis [196]. Huh7 cells are more and more used in research, however also displays a typical cancer phenotype. HepG2 and Huh7 both secrete apoB100 and LDL [197]. A research group recently tried to restore normal metabolism in Huh7 cells by exchanging FBS with adult human serum and succesfully reduced the Warburg-like metabolism. More precise, the cells formed large lipid and glycogen stores, increased glycogenesis, β-oxidation and ketogenesis and decreased glycolysis. The cells were able to degrade xenobiotics and to secrete bile, VLDL and albumin . This questions the whole concept of keeping cell lines in fetal bovine serum for experiments concerning normal metabolism [198]. We decided to take both cell lines into all of our investigations, trying to compensate what each cell line lacks in liver tissue properties. To study the gene locus in adipose tissue, we wanted to find out which cell line was most suitable. We tested 3T3L1, HIB 1B and SGBS cells in EMSA (**Figure 5.14**). 3T3L1, a mouse preadipocyte cell line, remains the most used cell model for the investigation of adipogenesis, even though its results are not necessarily applicable on human phenotypes [199]. Nevertheless, we needed a large amount of nuclear proteins, which is why we chose 3T3L1 for affinity chromatography. Since we expected a T2D phenotype in white adipose tissue and not necessarily in brown adipose tissue, we excluded HIB 1B for later experiments. A very good model for human adipose tissue are SGBS cells. Originally isolated from the stromal cells fraction of subcutaneous adipose tissue of a patient with SGBS, these cells display after differentiation an expression profile similar to mature human fat cells [200]. However, we had difficulties with consistent differentiation efficiency of the cells. Additionally, the cells showed low knockdown efficiencies, excluding many experiments from analysis. Since we were more successful with our liver and β-cell lines, we decided to focus primarily on these tissues.

## 6.2 The role of PDX1 as a potential regulator

Pancreatic duodenal homeobox-1 (PDX1) was discovered by several research groups. Using an electrophoretic mobility shift assay, one group demonstrated that PDX1 binds to multiple sites on the human *insulin* promotor [201]. *PDX1* is expressed in β- and δ- cells of the pancreas and in endocrine cells of the duodenum [202]. Its 283 amino acids lead up to a molecular weight of 31 kDa. It is comprised of proline-rich regions separated by an antennapedia-like homeodomain, which is responsible for protein-protein interactions. PDX1 influences gene expression of *INS, GLUT2, GCK* and *islet amyloid polypeptide* (IAPP) (reviewed in [203]). PDX1/DNA binding can be induced by glucose [204], insulin [205] or GLP-1 [206]. In summary, PDX1 is a strong candidate for a T2D-driven phenotype. We were able to show binding of PDX1 to the *rs7903146* locus *in vitro* via EMSA. Wang *et al.* performed PDX1 ChIP-seq and found PDX1 occupancy in intronic regions of *TCF7L2* [207]. With a gene reporter assay, we showed that PDX1 acts as an inhibitor on the *TCF7L2* promotor in 1.1B4 cells. In Ins1 cells, silencing of *Pdx1* caused a slight reduction of *Acsl5* expression. Even though we undeniably proved a regulatory role of PDX1 on the gene locus - *rs7903146* does not have an effect on this role. The question remains whether we did not use the appropiate assay to determine effects, or whether we missed effects due to the lack of coregulators of PDX1. Proteins such as E47/Pan1 and HMGA1 can interact with PDX1 and generate an activation complex [208]. If one of these proteins was missing or was expressed in the wrong concentration in our cell model, this could have influenced the result. Zhou *et al.* elucidated the role of TCF7L2 as a master regulator of insulin production and processing. TCF7L2 influences ISL1, which is a regulator of PDX1 [209]. Does this pathway have a feedback loop, by PDX1 binding to the TCF7L2 risk variant *rs7903146*? A knockdown of *PDX1* in Ins1 cells had no effect on the *Tcf7l2* expression, but the overexpression of *PDX1* in 1.1B4 strongly inhibited the *TCF7L2* promotor activity (**Figure 5.6 C**). In HEK293 cells, PDX1 had no effect on the *TCF7L2* promotor activity (data not shown). This strongly indicates a human β-cell specific influence of PDX1 on *TCF7L2* promotor activity. The knockdown was performed in a rat cell line, so it lacks the region containing the variant. It is necessary to study these effects in primary human material. PAX6 and PDX1 bind and activate the human *GIP* promotor. The combinaton of both transcription factors leads to a coexpresssion of GLP-1 in intestinal L-cells [210]. TCF7L2 has also been reported to affect GLP-1 expression in intestinal L-cells [211]. This raises the question of whether the β-cell is the only cell type that has PDX1 binding to the TCF7L2 risk variant.

## 6.3 The role of CEBPs as a potential regulators

CEBPs are part of the larger bZIP protein family. They are bipartite and occur active in homo - and heterodimers. This dimerization is possible through the leucine zipper region, which is usually located adjacent to the basic region responsible for sequence-specific DNA binding. The third domain is the transactivation domain, which may contain activating or inhibiting domains. There are six CEBPs classified so far, named after order of discovery α, β, γ, δ, ε and ζ. In **Figure 6.2** adapted from Schrem *et al.* [212] the different forms of CEBPs examined in this thesis are displayed. We looked at two CEBPA isoforms: p42* and p42 (CEBPA(1)/CEBPA long and CEBPA(2)/CEBPA medium). They differ only in a 107 bp UTR sequence which is upstream of CEBPA(1) but not CEBPA(2). Their role in gene regulation is strongly influenced by their dimerization. For exanple, CEBPA and ATF2 bind to CRE sites as homodimers, but their heterodimers do not [213]. This formation of homodimers or heterodimers appears to be concentration-dependent in the case of C/ATF and CEBPB and therefore

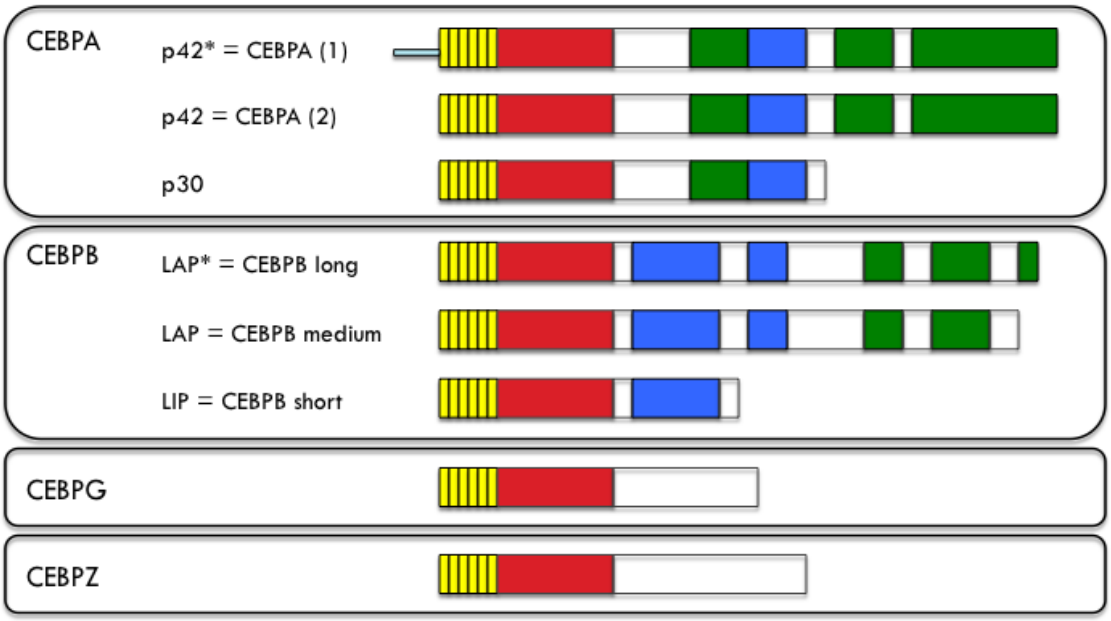


**Figure 6.2:** Schematic overview of the different CEBPs investigated in this thesis adopted from Schrem *et al.* [212]. Yellow = leucine zipper, red = basic region, green = activating domains, blue = negative regulatory domains, light blue = 5' UTR.

leads to the expression of different hormones. While heterodimers lead to increased expression of somatostatin, enkephalin and PEPCK, homodimers are more likely to bind to the angiotensinogen promotor [214].

*CEBPA* consists of only one exon with a length of 2783 bp. There are two common isoforms: p42 (42 kDa) and p30 (30 kDa). The isoform ratio is regulated by eIF2a/eIF4E. The full-length protein consists of 358 amino acids, whereas the N-terminal truncated protein lacks the first 117 amino acids due to an alternative translation start site at the third AUG codon. These missing 117 amino acids encode one transactivation domain (TAD). However, the second TAD and is present in both isoforms, as well as the basic region and the leucine zipper [215]. The transactivation domains contain recognition sites for transcription factors and co-activators such as TBP/TFIIB or CBP/ p300 and the chromatin remodeling complex SWI/SNF [216]. CEBPA is expressed in many tissues such as liver, lung, muscle, pancreas, GI, leukocytes, and in low levels also in brain and kidney [217]. Only CEBPA p42 is a potent inhibitor of cell proliferation. Key physiological features are described in more detail in Schrem *et al.* [212]. Briefly, CEBPA is involved in energy metabolism, cytochrom P450 gene expression, tumor biology, cellular differentiation, cell cycle control, liver regeneration, acute phase response, and detoxification of ammonia and bilirubin. Due to the low expression levels of *CEBPA* in the cell lines, it was difficult to generate results in knockdowns experiments. In 1.1B4, no effect of CEBPA on the expression of *TCF7L2* or *ACSL5* was detected, although we observed binding of both isoforms to the T risk allele . A possible explanation for the lack of effects in the transient knockdown could be the absence of cofactors, compensation by other CEBPs, or that we used the wrong cell line. CEBPB consists of four isoforms. The full-length isoform is 38 kDa, followed by the liver activating protein (LAP) with 35 kDa. The liver inhibiting protein (LIP) isoform is 21 kDa, while a fourth isoform is only 14 kDa. The larger isoforms (38 kDa and 35 kDa) are a product of alternative translation initiation [215]. The 21 kDa isoform appears to be regulated by alternative translation as well as proteolysis. The 14 kDa isoform, however, is generated solely through

proteolysis [218]. The main physiological characteristics are described in more detail in Schrem *et al.* [212]. Briefly, CEBPB, like CEBPA, is involved in energy metabolism, cytochrom P450 expression, acute phase response, liver regeneration, cellular differentiation and cell cycle control, as well as in apoptosis. Insulin has been reported to influence CEBPB expression in the liver [219]. In β-cells, CEBPB represses insulin expression at supraphysiological glucose levels [220]. Depletion of *TCF7L2* reduces insulin expression [221]. Carriers of the diabetic T risk allele were found to have an reduced insulin content in pancreatic islets, while increased *TCF7L2* levels [209]. Our results show that CEBPB activates TCF7L2 expression in 1.1B4 cells. In EMSA, CEBPB binds the T risk allele and thereby increases expression of *TCF7L2*. This could lead to decreased insulin levels. CEBPG is 16.4 kDa in size and has no TADs, which is why it acts as a transcription inhibitor by forming heterodimers with CEBPA and CEBPB [222]. CEBPG heterodimerizing with CEBPB does not affect DNA binding. It is not fully known how this heterodimerization inhibits translational activity, but it is hypothesized that this somehow affects the TADs of CEBPB. However, the inhibitory activity could not be confirmed in HepG2 cells and the authors suggested that the heterodimers are activating in these cells [223]. CEBPs are lower expressed in HepG2 cells compared to liver tissue [224]. Even though CEBPG has shown inhibitory function, we could not see this effect in our experiments. *CEBPG* depletion had no effect on gene expression in Huh7, HepG2 or 1.1B4 cells. Only in the preadipocyte cell line SGBS was a slight increase in *TCF7L2* expression. In EMSA we detected a clear T allele-specific band of endogenous CEBPG, so that the question if why this had no effect remains open. A heterodimerization with CEBPA and CEBPB isoforms could also be shown in EMSA assays (**Figure A6 C**). CEBPZ knockdowns had effects on *TCF7L2* and *ACSL5* expression in HepG2 and 1.1B4 cells. CEBPZ is a dominant negative regulator of the other CEBPs that are induced by growth-arresting and DNA-damaging agents [212]. The interplay of CEBPs with each other and various leucine zipper region containing proteins seems to be too complex to grasp with the methods applied in this thesis. Though their mutual high affinity for the T allele in EMSA and affinity chromatography is promising, we could prove a certain regulatory role of this transcription factor family in the context of *rs7903146*. To achieve this, the role of CEBPB in the β-cell needs to be closer investigated in the context of insulin expression, once *in vivo* binding to the T risk allele has been confirmed.

## 6.4 The role of STAT1 as a potential regulator

STAT1 is one of seven reported STATs (1, 2, 3, 4, 5, 5a, 5b, 6). STAT1 is a downstream effector of the IFNγ mediated signal cascade in inflammation. IFNγ acts mainly through the JAK/STAT pathway. IFNγ binds to the extracellular domain of its receptor. This leads to a phosphorylation of Janus kinase 1 and 2 at the receptor tyrosine residue Y440, which then phosphorylate STAT1 at tyrosine 701. Phosphorylated STAT1 can then form homodimers and translocate to the nucleus, where it binds to the target sequences (Gamma - Activated Sequences) [152]. STAT proteins are in general activators of gene expression, but act in 10 % of the cases as repressors. One target gene is *SOCS1*, a negative regulator of the STAT pathway that creates a negative feedback [225] ). Other inhibitors of the JAK/STAT signaling are SH2-containing phosphatases (SHPs), protein inhibitors of activated STATs (PIASs) and protein tyrosine phosphatases (PTPs) [226]. Non-alcoholic fatty liver disease - a common part of obestiy, is partially shifted towards non-alcoholic steato hepatitis by JAK/STAT signaling. JAK2 and STAT5 increase *insulin-like growth factor-1* expression by stimulation of growth hormone in the pituitary gland [227]. Deregulation of this pathway has been implicated in obesity and

steatosis [228]. High glucose levels promote cell apoptosis in MPC-5 cells, which is accompanied by high *STAT1* expression and decreased *FoxO1* expression. Silencing of STAT1 increases cell viability [229]. STAT1 knockout in mice showed that STAT1 is involved in energy consumption, lipid metabolism and mitochondrial biogenesis [230]. The role of STAT1 in obesity and T2D is complex and extensive, but we could not assign a regulatory role of STAT1 in the context of *rs7903146*. Originally identified in affinity chromatography with HepG2 nuclear extract, it showed no tremendous effects when depleted in HepG2 or Huh7 cells. In addition, the attempt to see substantial effects when stimulating the full IFNγ pathway failed. Effects on promotor function through the enhancer could not be detected (see **Figure 5.12 A-C**). We found a significantly higher binding affinity of STAT1 to the *rs7903146* locus. The question remains why this had no effect on gene expression. Additional stimuli or other transcription factors might be missing in the methodological setting.

## 6.5 The role of HMGA2 as a potential regulator

High mobility group proteins are architectual elements that modify chromatin. They recognize certain genomic regions and open the chromatin for the transcriptional machinery. They were isolated by Goodwin and Johns in 1973 [231]. Due to their small size (less than 30 kDa) they show a high mobility in an acrylamidgel, which is how they ascertained their name. In general, one can distinguish three types of HMGs: HMGAs, HMGBs and HMGNs. HMGA can again be subdivided into HMGA1 and HMGA2. HMGA can in turn be subdivided into HMGA1 and HMGA2. HMGA1 has three isoforms: HMGA1a, HMGA1b and HMGA1c. While HMGNs recognize nucleosome structure themself, HMGBs bind to chromatosomes. HMGAs recognize AT-rich regions on the genomes based on their AT-hook binding motifs. All HMGs contain an acidic tail that is thought to contain a domain that interacts with other proteins. In general, all HMG proteins have a low binding specificity [232]. HMGAs have little secondary structure in solution and only form a distinct conformation upon binding, which leads to enhanceosomes [233]. HMGAs are thought to pave the way for certain transcription factors to bind to DNA by unwinding inhibitory nucleosomes, but direct protein - protein interactions have also been reported [233, 234]. HMG proteins are highest expressed during embryogenesis and in cancer [235]. HMGAs have been reported to regulate insulin receptor expression [236]. Zhou *et al.* found HMGB1 binding to the locus using Edman-sequencing on HCT116 nuclear extracts. They confirmed HMGB1 binding prominently to the C nonrisk allele during Western blot of islets of three patients There was no allele specificity in HEK293 and HCT116 nuclear extracts. They used three donors with CC for a ChIP and were able to find binding of HMGB1 to the locus. This was confirmed with a supershift EMSA in HCT116. A transient *HMGB1* knockdown in HCT116 resulted in a 26 % reduction of *TCF7L2* expression. A GSK3b inhibitor was used to test WNT stimulated conditions. When depleting HMGB1 during WNT stimulation, *TCF7L2* increased 44 % less. Knockdown of *Hmgb1* in Ins1 cells reduced expression of *Ins* genes by 40-60 % and of *Tcf7l2* by 23 %. *Hmgb1* silencing abolished GSIS. Zhou and co-workers argued, that HMGB1 facilitates the binding of protein complexes and refers to PARP1, a factor found by Xia *et al.* [237], which also bound the C allele [238]. We identified HMGA2 binding to the T risk allele, which inhibited the expression of the nearby *ACSL5* . Long-chain acyl-CoA synthase 5 is part of the ACSL family that activates fatty acids to produce long chain fatty acyl-CoA. These are then used in different lipid pathways, e.g. to generate neutral lipids or phospholipids. *Acsl5* knockout mice show, among others, less obesity, improved insulin sensitivity, increased energy consumption and delayed triglyceride uptake [239]. Acetyl-CoA carboxylases catalyze the formation of malonyl-CoA.

Malonyl-CoA can be used for fatty acid synthesis, or in β-cells, for signaling purposes [240]. Inhibition of ACSL5 with ADIPO C in Ins1 β-cells had a protective effect under glucolipotoxic conditions [241]. As this factor is not very specific, interactions of this protein with other transcription factors in this locus should be investigated. This could be achieved by e.g. ChIP on ChIP.

## 6.6 Other transcription factors as potential regulators

In addition to our selection criteria ($p < 0.05$, C/T ratio $< 0.5$ or $> 2$) in affinity chromatography, we also selected other candidates, some as negative controls, others based on other criteria such as impressive fold change.

Atf1, Bbx, Prdm16 and Taf6 were proteins identified in the affinity chromatography with 3T3L1 nuclear extract. Bbx and Prdm16 proteins have been described to play a role in adipocyte biology and/or diabetes [242, 243]. All proteins except Prdm16 showed a significant enrichment at one allele, but we could not show any further effects of the proteins in the methods used. The knockdowns were not peformed in 3T3L1 due to time limitations, but solely in the HepG2 and 1.1B4 cell screenings.
We selected six additional proteins (SATB1, PAX6, LHX2, MYEF2, PURB and SUB1), that we identified in affinity chromatography performed with Ins1 nuclear extract. For SATB1, PAX6 and PURB also a role in the context of diabetes has been described [244–246]. Of these, SATB1 and PAX6 had no allele-specificity (SATB1 p=0.057, PAX6 p=0.057), LHX2 had an impressive 25 fold but insignificant (p=0.25) enrichment at the C nonrisk allele, and MYEF2 was insignificant (p=0.34). PURB and SUB1 bound to the T allele with a significantly higher affinity. However, PURB did not bind in EMSA when overexpressed in HEK293T cells. For this reason, the factor was not investigated further. SUB1 had no significant effects in knockdown experiments and was therefore not addressed further. SATB1 showed strong binding to the T allele *in vitro*, but no allelic effect in the gene reporter assay. PAX6 did not show any binding in EMSA and was not investigated further.
In 1.1B4 affinity chromatography, three additional identified factors (SCRT1, NR2F1 and TAF10) were examined in more detail, since a role in the context of diabetes has been described for SCRT1 and NR2F1 [247, 248]. SCRT1 bound significantly more strongly to the C allele, while the binding of NR2F1 to the C allele did not reach significance ($p = 0.057$). TAF10 showed a higher but insignificant affinity to the C nonrisk allele. SCRT1 binds to E-box enhancer motifs, which are targets of basic helix-loop-helix transcription factors, leading to gene suppresion [249]. However, we could not find this in our knockdown experiments. NR2F1 did not show *in vitro* binding to the risk locus, ultimately excluding the factor as a potential candidate. TAF10 had a 4-fold higher affinity for the C allele in 1.1B4 and was therefore tested in a knockdown. TAF10 depletion had no effect on the target genes. TAF10 is a component of TFIID and the TATA box-binding protein (TBP)-free TAF-containing complexes [250].
We tested two additional proteins identified in affinity chromatography with Huh7 nuclear extract. They were selected due to their significant binding to the T risk allele (HMGA1 and MED19). HMGA1 showed *in vitro* a strong T allele specific binding and it would be interesting to investigate downstream effects via transient knockdown and *in vivo* binding via ChIP. In *Hmga1* knockout mice, it was reported that the binding of PDX1 to the insulin receptor was reduced in the Hmga1-deficient mice, while *PDX1* protein expression was unaffected. These knockout mice showed reduced insulin receptor expression and reduced insulin signaling. Hypoinsulinemia of the knockout mice was caused by reduced β-cell mass and insulin secretion. The research group found four individuals with insulin resistance and type

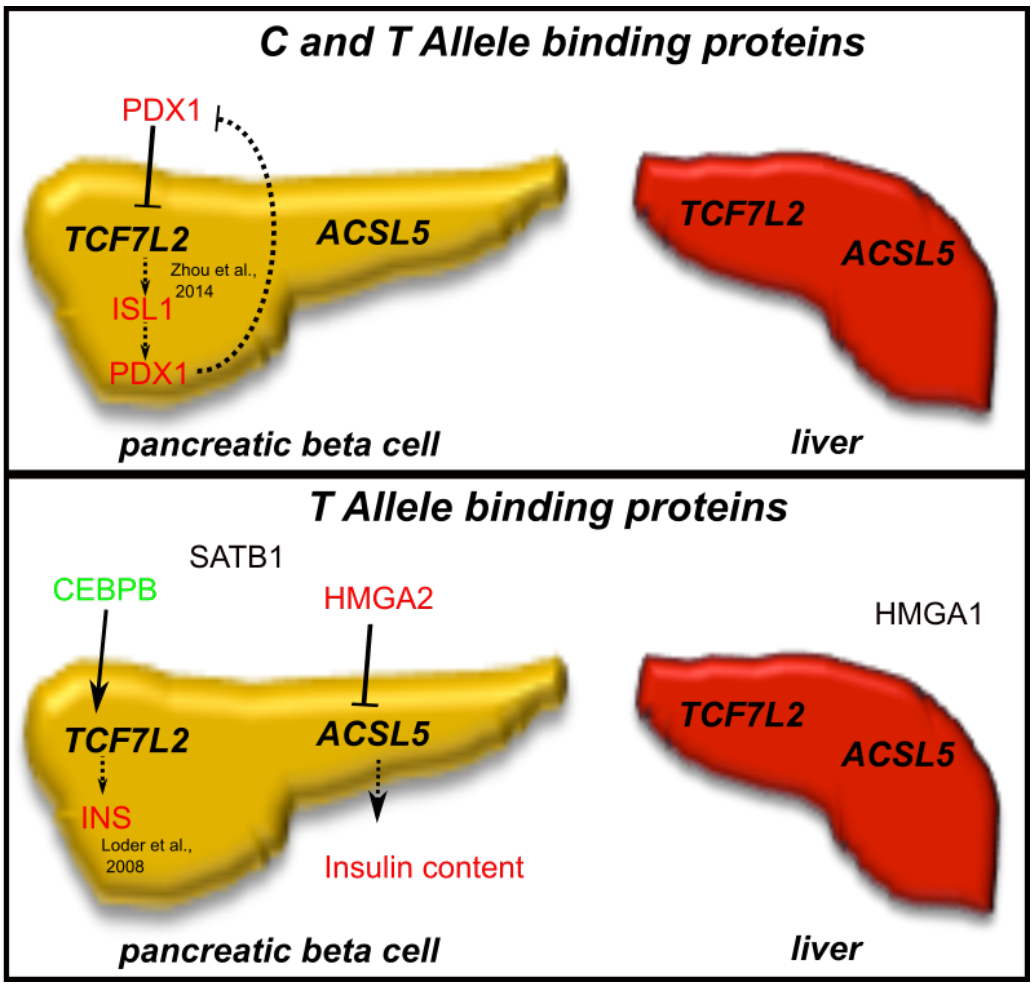


**Figure 6.3:** Identified and examined proteins of interest, regulating target genes (TCF7L2 and ACSL5) and potentially type 2 diabetes associated pathways. The candidates were categorized in unspecific allele binding or T allele binding. Proteins depicted in red are inhibiting target genes or are negatively influenced by the pathway. Green proteins are activators of the two target genes or positively regulated by a pathway. PDX1 inhibits TCF7L2, which again leads to a reduction of ISL1 and respectively *PDX1* expression [209]. CEBPB binds to the T allele in β-cells, leading to an increased TCF7L2 content. High TCF7L2 concentrations are associated with lower insulin levels in the β-cells [221].

2 diabetes, which showed no mutations in the *INSR* gene, however a markedly reduced *HMGA1* level. These patients showed not only a reduced expression of *HMGA1*, but also less occupancy of HMGA1 at the *INSR* locus in lymphoblasts [251]. The knockdown of *MED19* showed no effect, and therefore the protein was not investigated further.

We tested an additional protein (ZNF593) which was identified in affinity chromatography with HepG2 nuclear extract, but which after depletion in HepG2 and 1.1B4 showed no effect on *TCF7L2* and *ACSL5* expression. No binding was detected in EMSA, although it bound to the T allele with a 3-fold higher affinity in the affinity chromatography. Overall, the preliminary data genreated for these candidate proteins did not provide any evidence that supports a major role of these factors in gene regulation of the *rs7903146* genomic region.

# 7 Conclusion and outlook

We are still at the beginning of resolving the mystery around regulatory roles of SNPs in noncoding genomic regions. Only a tiny fraction of the DNA is comprised of actual coding sequences, nearly three billion bases had yet unknown function in 2012. The Encyclopedia of DNA Elements (ENCODE) project was launched to untangle this huge regulatory machinery. They were able to show that 80 % of the genome have a biochemical function that permanently supersedes the term "junk DNA" [252]. The term "junk DNA" was coined in the 1972 by Susumo Ohno [253] but mainly referred to pseudogenes. However, the term was later often used for all noncoding regions. The nongenic DNA is made up by a majority of transposable elements. These are, in humans, highly degraded, repetitive or repeat derived sequences, originally short and long interspersed nuclear elements, endogenous retroviruses and cut-and-paste transposons [254, 255]. However, it has been shown many times that these regions can have important functions (reviewed in Cowley and Oakey, 2013 [256]). The most relevant regulatory elements in the genome, which could influence a disease phenotype are promotor, enhancer, silencer, isolator, splicer regions and lncRNAs. However, a SNP associated with a disease does not necessarily have to be the causative one: Variants in linkage disequilibrium could also play a role and should not be left uninvestigated. Type 2 diabetes is not Mendelian, but a multigenic and environmentally driven disease. Until the GWAS studies were published, we had missed most of the underlying heritability. The complex genetics of T2D is still unresolved, but several potential candidates have now been identified and intensively investigated. But how to approach such a complicated topic as transcription factor binding? Transcription factors rarely act alone and form huge complexes to initiate or inhibit the transcription machinery. Finding a regulatory protein binding differentially to a region due to a single nucleotide exchange is like searching a needle in a haystack. Affinity chromatography has the disadvantage of looking only at a small window of the genome - in our case a 45 bp sequence surrounding the variant *rs7903146*. What we miss is the full capacity of chromatin plasticity if one ignores the involvement of additional cis-regulatory elements and transcription factors. An interesting method on the rise is the proximity-based labeling by *in vivo* biotinylation of proteins over a biotin ligase fused to a known protein binding the regions of interest [257]. Used biotin ligases are BirA, Turbo ID, Mini Turbo, BioID2 and BASU, which are all promisciously labeling all proteins within their vicinity. Another faster labeling alternative is APEX (ascorbate peroxidase catalyses the oxidation of biotin-phenol to a biotin-phenoxyl radical leading to a biotinylation of electron rich amino acids) [258]. The inclusion of CRISPR-mediated tagging with BirA enables the mapping of protein - protein interaction *in vivo* on specific target regions of the DNA [259]. This could contribute significantly to the elucidation of the transcriptional regulation at specific enhancer regions. The method can finally support the highly artificial methods of today and bring research to a next level. Until then, many steps can be taken to complete the data generated in this thesis. Many regulatory proteins identified in this thesis were more involved in chromatin remodeling than in direct binding and formation of transcription complexes. Since these did not meet intital criteria as a direct DNA-binding transcription factor, they were not investigated. So the experimental set up has to shift towards more epigenetic

approaches. HMGA1 recognizes AT-rich regions in the genome and acts as an antagonist of the linker histone H1. This linker histone H1 binds to the same DNA regions and keeps chromatin tightly packed and therefore transcriptionally inactive [260]. A chromatin accessibility assay based on nuclease treatment could be performed to investigate the exact effects of HMGA1 on the T allele. A more accessible chromatin on the T allele in pancreatic islets has been described in literature [126] - and HMGA2 and HMGA1, which were both identified in the human β-cell line 1.1B4 in this thesis could lead to such a state. A major topic that evolved in the thesis was the question of whether cell cycle plays a role in the whole transcription factor binding processes, which could explain the large discrepancies between *in vitro* and *in vivo* methods. A way to resolve this might be a cell cylce arrest before performing knockdown and ChIP experiments. This can be achieved chemically or by adding human adult serum (Chapter 6.1.6). Another step might be the bioinformatic analysis of the identified proteins in the context of protein - protein interactions. Creating networks and investigating the potential candiates in this setting might reveal further mechanisms. This type of anylsis has already started. One of the interesting networks was STAT1 interacting with PTPN11. We overexpressed *PTPN11* in HEK293T and wanted to see in EMSA, whether we could see any effects on the STAT1 binding to the risk locus, however no effects were seen in this assay (**Figure A5**). Also the potential interplay of SATB1 and STAT1 could also be evaluated. SATB1 and STAT1 both share an inhibitor called PIAS1 [261, 262] and are involved in inflammatory processes by regulating the MHC class I locus [263].

We looked at two potential targets (TCF7L2 and ACSL5) of the enhancer region harbouring *rs7903146*, but we might have missed out on other potentially regulated genes. Therefore RNA-Seq of regulator knockdowns might be a better choice. Completing the already performed knockdowns to biological triplicates will aid narrowing down potential targets, and effects of the transient knockdowns should be confirmed at the protein level. We found regulatory proteins involved in inflammation, chromatin remodeling, insulin expression and many more. All of them can be brought into the context of a T2D phenotype. There is still a lot of information, that can be generated from this data. Further investigation of the factors studied in this thesis could shed more light onto the role of *rs7903146* in the context of a T2D risk phenotype.

# Bibliography


[1] MacCracken J, Hoel D, Jovanovic L. From ants to analogues. Postgraduate Medicine. 1997;101(4):138–150.

[2] Cho NH, Shaw JE, Karuranga S, Huang Y, da Rocha Fernandes JD, Ohlrogge AW, et al. IDF Diabetes Atlas: Global estimates of diabetes prevalence for 2017 and projections for 2045 [Journal Article]. Diabetes Res Clin Pract. 2018;138:271–281.

[3] American Diabetes A. Diagnosis and classification of diabetes mellitus. Diabetes Care. 2009 Jan;32 Suppl 1(Suppl 1):S62–7.

[4] Castiello FR, Heileman K, Tabrizian M. Microfluidic perfusion systems for secretion fingerprint analysis of pancreatic islets: applications, challenges and opportunities. Lab Chip. 2016 Feb;16(3):409–431.

[5] Petersen MC, Shulman GI. Mechanisms of Insulin Action and Insulin Resistance. Physiol Rev. 2018 10;98(4):2133–2223.

[6] White MF. Insulin signaling in health and disease. Science. 2003 Dec;302(5651):1710–1711.

[7] Boucher J, Kleinridders A, Kahn CR. Insulin receptor signaling in normal and insulin-resistant states. Cold Spring Harb Perspect Biol. 2014 Jan;6(1).

[8] Taniguchi CM, Emanuelli B, Kahn CR. Critical nodes in signalling pathways: insights into insulin action. Nat Rev Mol Cell Biol. 2006 Feb;7(2):85–96.

[9] Wu HH, Li YL, Liu NJ, Yang Z, Tao XM, Du YP, et al. TCF7L2 regulates pancreatic beta-cell function through PI3K/AKT signal pathway. Diabetol Metab Syndr. 2019;11:55.

[10] Bellou V, Belbasis L, Tzoulaki I, Evangelou E. Risk factors for type 2 diabetes mellitus: An exposure-wide umbrella review of meta-analyses. PLoS ONE. 2018;13(3):e0194127.

[11] Hotamisligil GS, Shargill NS, Spiegelman BM. Adipose expression of tumor necrosis factor-alpha: direct role in obesity-linked insulin resistance. Science. 1993 Jan;259(5091):87–91.

[12] Kahn BB, Flier JS. Obesity and insulin resistance. J Clin Invest. 2000 Aug;106(4):473–481.

[13] Brown AE, Walker M. Genetics of Insulin Resistance and the Metabolic Syndrome. Curr Cardiol Rep. 2016 08;18(8):75.

[14] Kahn SE, Cooper ME, Del Prato S. Pathophysiology and treatment of type 2 diabetes: perspectives on the past, present, and future. Lancet. 2014 Mar;383(9922):1068–1083.

[15] Marín-Peñalver JJ, Martín-Timón I, Sevillano-Collantes C, Del Cañizo-Gómez FJ. Update on the treatment of type 2 diabetes mellitus. World J Diabetes. 2016 Sep;7(17):354–395.

[16] Papatheodorou K, Banach M, Bekiari E, Rizzo M, Edmonds M. Complications of Diabetes 2017. J Diabetes Res. 2018;2018:3086167.

[17] Groop L, Forsblom C, Lehtovirta M, Tuomi T, Karanko S, Nissen M, et al. Metabolic consequences of a family history of NIDDM (the Botnia study): evidence for sex-specific parental effects. Diabetes. 1996 Nov;45(11):1585–1593.

[18] Tillil H, Kobberling J. Age-corrected empirical genetic risk estimates for first-degree relatives of IDDM patients. Diabetes. 1987 Jan;36(1):93–99.

[19] Willemsen G, Ward KJ, Bell CG, Christensen K, Bowden J, Dalgard C, et al. The Concordance and Heritability of Type 2 Diabetes in 34,166 Twin Pairs From International Twin Registers: The Discordant Twin (DISCOTWIN) Consortium. Twin Res Hum Genet. 2015 Dec;18(6):762–771.

[20] Altshuler D, Hirschhorn JN, Klannemark M, Lindgren CM, Vohl MC, Nemesh J, et al. The common PPARgamma Pro12Ala polymorphism is associated with decreased risk of type 2 diabetes. Nat Genet. 2000 Sep;26(1):76–80.

[21] Gloyn AL, Weedon MN, Owen KR, Turner MJ, Knight BA, Hitman G, et al. Large-scale association studies of variants in genes encoding the pancreatic beta-cell KATP channel subunits Kir6.2 (KCNJ11) and SUR1 (ABCC8) confirm that the KCNJ11 E23K variant is associated with type 2 diabetes. Diabetes. 2003 Feb;52(2):568–572.

[22] Dupuis J, Langenberg C, Prokopenko I, Saxena R, Soranzo N, Jackson AU, et al. New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. Nat Genet. 2010 Feb;42(2):105–116.

[23] Sladek R, Rocheleau G, Rung J, Dina C, Shen L, Serre D, et al. A genome-wide association study identifies novel risk loci for type 2 diabetes [Journal Article]. Nature. 2007;445(7130):881–885.

[24] Ahlqvist E, Ahluwalia TS, Groop L. Genetics of type 2 diabetes. Clin Chem. 2011 Feb;57(2):241–254.

[25] Tam V, Patel N, Turcotte M, Bossé Y, Paré G, Meyre D. Benefits and limitations of genome-wide association studies. Nat Rev Genet. 2019 08;20(8):467–484.

[26] Billings LK, Florez JC. The genetics of type 2 diabetes: what have we learned from GWAS? Ann N Y Acad Sci. 2010 Nov;1212:59–77.

[27] Sachidanandam R, Weissman D, Schmidt SC, Kakol JM, Stein LD, Marth G, et al. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. Nature. 2001 Feb;409(6822):928–933.

[28] Prokopenko I, Langenberg C, Florez JC, Saxena R, Soranzo N, Thorleifsson G, et al. Variants in MTNR1B influence fasting glucose levels. Nat Genet. 2009 Jan;41(1):77–81.

[29] Lyssenko V, Nagorny CL, Erdos MR, Wierup N, Jonsson A, Spégel P, et al. Common variant in MTNR1B associated with increased risk of type 2 diabetes and impaired early insulin secretion. Nat Genet. 2009 Jan;41(1):82–88.

[30] Nica AC, Montgomery SB, Dimas AS, Stranger BE, Beazley C, Barroso I, et al. Candidate causal regulatory effects by integration of expression QTLs with complex trait genetic associations. PLoS Genet. 2010 Apr;6(4):e1000895.

[31] Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, et al. Systematic localization of common disease-associated variation in regulatory DNA. Science. 2012 Sep;337(6099):1190–1195.

[32] Xue A, Wu Y, Zhu Z, Zhang F, Kemper KE, Zheng Z, et al. Genome-wide association analyses identify 143 risk variants and putative regulatory mechanisms for type 2 diabetes. Nature Communications. 2018;9(1):2941.

[33] Sun W, Yao S, Tang J, Liu S, Chen J, Deng D, et al. Integrative analysis of super enhancer SNPs for type 2 diabetes. PLoS ONE. 2018;13(1):e0192105.

[34] Claussnitzer M, Dankel SN, Klocke B, Grallert H, Glunk V, Berulava T, et al. Leveraging cross-species transcription factor binding site patterns: from diabetes risk loci to disease mechanisms [Journal Article]. Cell. 2014;156(1-2):343–58.

[35] Lee H, Qian K, Toerne Cv, Hoerburger L, Claussnitzer M, Hoffmann C, et al. Allele-specific quantitative proteomics unravels molecular mechanisms modulated by cis-regulatory PPARG locus variation [Journal Article]. Nucleic acids research. 2017;45(6):3266–3279.

[36] Small KS, Todorcevic M, Civelek M, El-Sayed Moustafa JS, Wang X, Simon MM, et al. Regulatory variants at KLF14 influence type 2 diabetes risk via a female-specific effect on adipocyte size and body composition. Nat Genet. 2018 Apr;50(4):572–580.

[37] Yasuda K, Miyake K, Horikawa Y, Hara K, Osawa H, Furuta H, et al. Variants in KCNQ1 are associated with susceptibility to type 2 diabetes mellitus. Nat Genet. 2008 Sep;40(9):1092–1097.

[38] Rung J, Cauchi S, Albrechtsen A, Shen L, Rocheleau G, Cavalcanti-Proença C, et al. Genetic variant near IRS1 is associated with type 2 diabetes, insulin resistance and hyperinsulinemia. Nat Genet. 2009 Oct;41(10):1110–1115.

[39] Panico P, Salazar AM, Burns AL, Ostrosky-Wegman P. Role of calpain-10 in the development of diabetes mellitus and its complications. Arch Med Res. 2014 Feb;45(2):103–115.

[40] Horikawa Y, Oda N, Cox NJ, Li X, Orho-Melander M, Hara M, et al. Genetic variation in the gene encoding calpain-10 is associated with type 2 diabetes mellitus. Nat Genet. 2000 Oct;26(2):163–175.

[41] Reynisdottir I, Thorleifsson G, Benediktsson R, Sigurdsson G, Emilsson V, Einarsdottir AS, et al. Localization of a susceptibility gene for type 2 diabetes to chromosome 5q34-q35.2. Am J Hum Genet. 2003 Aug;73(2):323–335.

[42] Duggirala R, Blangero J, Almasy L, Dyer TD, Williams KL, Leach RJ, et al. Linkage of type 2 diabetes mellitus and of age at onset to a genetic location on chromosome 10q in Mexican Americans. Am J Hum Genet. 1999 Apr;64(4):1127–1140.

[43] Grant SFA, Thorleifsson G, Reynisdottir I, Benediktsson R, Manolescu A, Sainz J, et al. Variant of transcription factor 7-like 2 (TCF7L2) gene confers risk of type 2 diabetes. Nat Genet. 2006 Mar;38(3):320–323.

[44] Gaulton KJ, Ferreira T, Lee Y, Raimondo A, Magi R, Reschen ME, et al. Genetic fine mapping and genomic annotation defines causal mechanisms at type 2 diabetes susceptibility loci. Nat Genet. 2015 Dec;47(12):1415–1425.

[45] Tong Y, Lin Y, Zhang Y, Yang J, Zhang Y, Liu H, et al. Association between TCF7L2 gene polymorphisms and susceptibility to type 2 diabetes mellitus: a large Human Genome Epidemiology (HuGE) review and meta-analysis. BMC Med Genet. 2009 Feb;10:15.

[46] Thorsby PM, Midthjell K, Gjerlaugsen N, Holmen J, Hanssen KF, Birkeland KI, et al. Comparison of genetic risk in three candidate genes (TCF7L2, PPARG, KCNJ11) with traditional risk factors for type 2 diabetes in a population-based study–the HUNT study. Scand J Clin Lab Invest. 2009;69(2):282–287.

[47] Liu XH, Xie CG, An Y, Zhang XX, Wu WB. Meta-analysis of the association between the rs7903146 polymorphism at the TCF7L2 locus and type 2 diabetes mellitus susceptibility. Genet Mol Res. 2015 Dec;14(4):16856–16862.

[48] Salonen JT, Uimari P, Aalto JM, Pirskanen M, Kaikkonen J, Todorova B, et al. Type 2 diabetes whole-genome association study in four populations: the DiaGen consortium. Am J Hum Genet. 2007 Aug;81(2):338–345.

[49] Palmer ND, Lehtinen AB, Langefeld CD, Campbell JK, Haffner SM, Norris JM, et al. Association of TCF7L2 gene polymorphisms with reduced acute insulin response in Hispanic Americans. J Clin Endocrinol Metab. 2008 Jan;93(1):304–309.

[50] Ding W, Xu L, Zhang L, Han Z, Jiang Q, Wang Z, et al. Meta-analysis of association between TCF7L2 polymorphism rs7903146 and type 2 diabetes mellitus [Journal Article]. BMC medical genetics. 2018;19(1):38.

[51] Lou L, Wang J, Wang J. Genetic associations between Transcription Factor 7 Like 2 rs7903146 polymorphism and type 2 diabetes mellitus: a meta-analysis of 115,809 subjects. Diabetol Metab Syndr. 2019;11:56.

[52] Tong Y, Lin Y, Zhang Y, Yang J, Zhang Y, Liu H, et al. Association between TCF7L2 gene polymorphisms and susceptibility to type 2 diabetes mellitus: a large Human Genome Epidemiology (HuGE) review and meta-analysis. BMC Med Genet. 2009 Feb;10:15.

[53] Pang DX, Smith AJP, Humphries SE. Functional analysis of TCF7L2 genetic variants associated with type 2 diabetes. Nutr Metab Cardiovasc Dis. 2013 Jun;23(6):550–556.

[54] Moon RT, Brown JD, Torres M. WNTs modulate cell fate and behavior during vertebrate development. Trends Genet. 1997 Apr;13(4):157–162.

[55] Peifer M, Polakis P. Wnt signaling in oncogenesis and embryogenesis–a look outside the nucleus. Science. 2000 Mar;287(5458):1606–1609.

[56] Wodarz A, Nusse R. Mechanisms of Wnt signaling in development. Annu Rev Cell Dev Biol. 1998;14:59–88.

[57] Kikuchi A, Kishida S, Yamamoto H. Regulation of Wnt signaling by protein-protein interaction and post-translational modifications. Exp Mol Med. 2006 Feb;38(1):1–10.

[58] Yi F, Brubaker PL, Jin T. TCF-4 mediates cell type-specific regulation of proglucagon gene expression by beta-catenin and glycogen synthase kinase-3beta. J Biol Chem. 2005 Jan;280(2):1457–1464.

[59] Chiang YT, Ip W, Jin T. The role of the Wnt signaling pathway in incretin hormone production and function. Front Physiol. 2012;3:273.

[60] Papadopoulou S, Edlund H. Attenuated Wnt signaling perturbs pancreatic growth but not pancreatic function [Journal Article]. Diabetes. 2005;54(10):2844–51.

[61] Heiser PW, Lau J, Taketo MM, Herrera PL, Hebrok M. Stabilization of beta-catenin impacts pancreas growth [Journal Article]. Development. 2006;133(10):2023–32.

[62] Rulifson IC, Karnik SK, Heiser PW, ten Berge D, Chen H, Gu X, et al. Wnt signaling regulates pancreatic beta cell proliferation [Journal Article]. Proc Natl Acad Sci U S A. 2007;104(15):6247–52.

[63] Fujino T, Asaba H, Kang MJ, Ikeda Y, Sone H, Takada S, 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 USA. 2003 Jan;100(1):229–234.

[64] Papadopoulou S, Edlund H. Attenuated Wnt signaling perturbs pancreatic growth but not pancreatic function [Journal Article]. Diabetes. 2005;54(10):2844–51.

[65] Ross SE, Hemati N, Longo KA, Bennett CN, Lucas PC, Erickson RL, et al. Inhibition of adipogenesis by Wnt signaling. Science. 2000 Aug;289(5481):950–953.

[66] Bennett CN, Ross SE, Longo KA, Bajnok L, Hemati N, Johnson KW, et al. Regulation of Wnt signaling during adipogenesis. J Biol Chem. 2002 Aug;277(34):30998–31004.

[67] Vertino AM, Taylor-Jones JM, Longo KA, Bearden ED, Lane TF, McGehee RE, et al. Wnt10b deficiency promotes coexpression of myogenic and adipogenic programs in myoblasts. Mol Biol Cell. 2005 Apr;16(4):2039–2048.

[68] Kennell JA, MacDougald OA. Wnt signaling inhibits adipogenesis through beta-catenin-dependent and -independent mechanisms. J Biol Chem. 2005 Jun;280(25):24004–24010.

[69] Kang S, Bajnok L, Longo KA, Petersen RK, Hansen JB, Kristiansen K, et al. Effects of Wnt signaling on brown adipocyte differentiation and metabolism mediated by PGC-1alpha. Mol Cell Biol. 2005 Feb;25(4):1272–1282.

[70] Longo KA, Wright WS, Kang S, Gerin I, Chiang SH, Lucas PC, et al. Wnt10b inhibits development of white and brown adipose tissues [Journal Article]. J Biol Chem. 2004;279(34):35503–9.

[71] Bennett CN, Hodge CL, MacDougald OA, Schwartz J. Role of Wnt10b and C/EBPalpha in spontaneous adipogenesis of 243 cells [Journal Article]. Biochem Biophys Res Commun. 2003;302(1):12–6.

[72] MacDougald OA. Signaling pathway puts the break on fat cell formation. ScientificWorldJournal. 2001 May;1:188–189.

[73] Ross SE, Erickson RL, Gerin I, DeRose PM, Bajnok L, Longo KA, et al. Microarray analyses during adipogenesis: understanding the effects of Wnt signaling on adipogenesis and the roles of liver X receptor alpha in adipocyte metabolism. Mol Cell Biol. 2002 Aug;22(16):5989–5999.

[74] Christodoulides C, Lagathu C, Sethi JK, Vidal-Puig A. Adipogenesis and WNT signalling. Trends Endocrinol Metab. 2009 Jan;20(1):16–24.

[75] Gustafson B, Hammarstedt A, Hedjazifar S, Smith U. Restricted adipogenesis in hypertrophic obesity: the role of WISP2, WNT, and BMP4. Diabetes. 2013 Sep;62(9):2997–3004.

[76] Gustafson B, Smith U. The WNT inhibitor Dickkopf 1 and bone morphogenetic protein 4 rescue adipogenesis in hypertrophic obesity in humans. Diabetes. 2012 May;61(5):1217–1224.

[77] Wright WS, Longo KA, Dolinsky VW, Gerin I, Kang S, Bennett CN, et al. Wnt10b inhibits obesity in ob/ob and agouti mice. Diabetes. 2007 Feb;56(2):295–303.

[78] Prestwich TC, Macdougald OA. Wnt/beta-catenin signaling in adipogenesis and metabolism. Curr Opin Cell Biol. 2007 Dec;19(6):612–617.

[79] Aslanidi G, Kroutov V, Philipsberg G, Lamb K, Campbell-Thompson M, Walter GA, et al. Ectopic expression of Wnt10b decreases adiposity and improves glucose homeostasis in obese rats. Am J Physiol Endocrinol Metab. 2007 Sep;293(3):E726–736.

[80] Jin T. Current Understanding on Role of the Wnt Signaling Pathway Effector TCF7L2 in Glucose Homeostasis. Endocr Rev. 2016 06;37(3):254–277.

[81] Nusse R, Brown A, Papkoff J, Scambler P, Shackleford G, McMahon A, et al. A new nomenclature for int-1 and related genes: the Wnt gene family. Cell. 1991 Jan;64(2):231.

[82] Rao TP, Kuehl M. An updated overview on Wnt signaling pathways: a prelude for more. Circ Res. 2010 Jun;106(12):1798–1806.

[83] Hurlstone A, Clevers H. T-cell factors: turn-ons and turn-offs. EMBO J. 2002 May;21(10):2303–2311.

[84] Giese K, Amsterdam A, Grosschedl R. DNA-binding properties of the HMG domain of the lymphoid-specific transcriptional regulator LEF-1. Genes Dev. 1991 Dec;5(12B):2567–2578.

[85] Giese K, Cox J, Grosschedl R. The HMG domain of lymphoid enhancer factor 1 bends DNA and facilitates assembly of functional nucleoprotein structures. Cell. 1992 Apr;69(1):185–195.

[86] Yang H, Li Q, Lee JH, Shu Y. Reduction in Tcf7l2 expression decreases diabetic susceptibility in mice. Int J Biol Sci. 2012;8(6):791–801.

[87] da Silva Xavier G, Loder MK, McDonald A, Tarasov AI, Carzaniga R, Kronenberger K, et al. TCF7L2 regulates late events in insulin secretion from pancreatic islet beta-cells [Journal Article]. Diabetes. 2009;58(4):894–905.

[88] Ip W, Shao W, Chiang YT, Jin T. The Wnt signaling pathway effector TCF7L2 is upregulated by insulin and represses hepatic gluconeogenesis. Am J Physiol Endocrinol Metab. 2012 Nov;303(9):E1166–1176.

[89] Lyssenko V, Lupi R, Marchetti P, Del Guerra S, Orho-Melander M, Almgren P, et al. Mechanisms by which common variants in the TCF7L2 gene increase risk of type 2 diabetes. J Clin Invest. 2007 Aug;117(8):2155–2163.

[90] Pilgaard K, Jensen CB, Schou JH, Lyssenko V, Wegner L, Brons C, 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 [Journal Article]. Diabetologia. 2009;52(7):1298–307.

[91] Wegner L, Hussain MS, Pilgaard K, Hansen T, Pedersen O, Vaag A, et al. Impact of TCF7L2 rs7903146 on insulin secretion and action in young and elderly Danish twins [Journal Article]. J Clin Endocrinol Metab. 2008;93(10):4013–9.

[92] Nishida T. Diagnosis and Clinical Implications of Diabetes in Liver Cirrhosis: A Focus on the Oral Glucose Tolerance Test. J Endocr Soc. 2017 Jul;1(7):886–896.

[93] Cho H, Mu J, Kim JK, Thorvaldsen JL, Chu Q, Crenshaw r E B, et al. Insulin resistance and a diabetes mellitus-like syndrome in mice lacking the protein kinase Akt2 (PKB beta) [Journal Article]. Science. 2001;292(5522):1728–31.

[94] Taniguchi CM, Kondo T, Sajan M, Luo J, Bronson R, Asano T, et al. Divergent regulation of hepatic glucose and lipid metabolism by phosphoinositide 3-kinase via Akt and PKClambda/zeta. Cell Metab. 2006 May;3(5):343–353.

[95] Saltiel AR, Kahn CR. Insulin signalling and the regulation of glucose and lipid metabolism. Nature. 2001 Dec;414(6865):799–806.

[96] Wan M, Leavens KF, Saleh D, Easton RM, Guertin DA, Peterson TR, et al. Postprandial hepatic lipid metabolism requires signaling through Akt2 independent of the transcription factors FoxA2, FoxO1, and SREBP1c. Cell Metab. 2011 Oct;14(4):516–527.

[97] Loria P, Lonardo A, Anania F. Liver and diabetes. A vicious circle. Hepatol Res. 2013 Jan;43(1):51–64.

[98] Besse-Patin A, Jeromson S, Levesque-Damphousse P, Secco B, Laplante M, Estall JL. PGC1A regulates the IRS1:IRS2 ratio during fasting to influence hepatic metabolism downstream of insulin. Proc Natl Acad Sci USA. 2019 03;116(10):4285–4290.

[99] Oh KJ, Han HS, Kim MJ, Koo SH. CREB and FoxO1: two transcription factors for the regulation of hepatic gluconeogenesis. BMB Rep. 2013 Dec;46(12):567–574.

[100] Oh KJ, Park J, Kim SS, Oh H, Choi CS, Koo SH. TCF7L2 modulates glucose homeostasis by regulating CREB- and FoxO1-dependent transcriptional pathway in the liver. PLoS Genet. 2012 Sep;8(9):e1002986.

[101] Norton L, Chen X, Fourcaudot M, Acharya NK, DeFronzo RA, Heikkinen S. The mechanisms of genome-wide target gene regulation by TCF7L2 in liver cells. Nucleic Acids Res. 2014 Dec;42(22):13646–13661.

[102] Mondal AK, Das SK, Baldini G, Chu WS, Sharma NK, Hackney OG, et al. Genotype and tissue-specific effects on alternative splicing of the transcription factor 7-like 2 gene in humans [Journal Article]. J Clin Endocrinol Metab. 2010;95(3):1450–7.

[103] Duval A, Rolland S, Tubacher E, Bui H, Thomas G, Hamelin R. 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 Jul;60(14):3872–3879.

[104] Vaz-Drago R, Custódio N, Carmo-Fonseca M. Deep intronic mutations and human disease. Hum Genet. 2017 09;136(9):1093–1111.

[105] Calkhoven CF, Mueller C, Leutz A. Translational control of C/EBPalpha and C/EBPbeta isoform expression. Genes Dev. 2000 Aug;14(15):1920–1932.

[106] Wilkinson ME, Charenton C, Nagai K. RNA Splicing by the Spliceosome. Annual Review of Biochemistry. 2020;89(1):null. PMID: 31794245.

[107] Wahl MC, Will CL, Luehrmann R. The spliceosome: design principles of a dynamic RNP machine. Cell. 2009 Feb;136(4):701–718.

[108] Naftelberg S, Schor IE, Ast G, Kornblihtt AR. Regulation of alternative splicing through coupling with transcription and chromatin structure. Annu Rev Biochem. 2015;84:165–198.

[109] Antonellis A, Dennis MY, Burzynski G, Huynh J, Maduro V, Hodonsky CJ, et al. A rare myelin protein zero (MPZ) variant alters enhancer activity in vitro and in vivo. PLoS ONE. 2010 Dec;5(12):e14346.

[110] Seo S, Takayama K, Uno K, Ohi K, Hashimoto R, Nishizawa D, et al. Functional analysis of deep intronic SNP rs13438494 in intron 24 of PCLO gene. PLoS ONE. 2013;8(10):e76960.

[111] Prokunina-Olsson L, Kaplan LM, Schadt EE, Collins FS. Alternative splicing of TCF7L2 gene in omental and subcutaneous adipose tissue and risk of type 2 diabetes [Journal Article]. PloS one. 2009;4(9):e7231.

[112] Brannon M, Brown JD, Bates R, Kimelman D, Moon RT. XCtBP is a XTcf-3 co-repressor with roles throughout Xenopus development. Development. 1999 Jun;126(14):3159–3170.

[113] Fang M, Li J, Blauwkamp T, Bhambhani C, Campbell N, Cadigan KM. C-terminal-binding protein directly activates and represses Wnt transcriptional targets in Drosophila. EMBO J. 2006 Jun;25(12):2735–2745.

[114] Prokunina-Olsson L, Welch C, Hansson O, Adhikari N, Scott LJ, Usher N, et al. Tissue-specific alternative splicing of TCF7L2. Hum Mol Genet. 2009 Oct;18(20):3795–3804.

[115] Atcha FA, Syed A, Wu B, Hoverter NP, Yokoyama NN, Ting JHT, et al. A unique DNA binding domain converts T-cell factors into strong Wnt effectors. Mol Cell Biol. 2007 Dec;27(23):8352–8363.

[116] Brantjes H, Roose J, van De Wetering M, Clevers H. All Tcf HMG box transcription factors interact with Groucho-related co-repressors. Nucleic Acids Res. 2001 Apr;29(7):1410–1419.

[117] Arce L, Pate KT, Waterman ML. Groucho binds two conserved regions of LEF-1 for HDAC-dependent repression. BMC Cancer. 2009 May;9:159.

[118] Osmark P, Hansson O, Jonsson A, Rönn T, Groop L, Renström E. Unique splicing pattern of the TCF7L2 gene in human pancreatic islets [Journal Article]. Diabetologia. 2009;52(5):850–854.

[119] Hansson O, Zhou Y, Renström E, Osmark P. Molecular function of TCF7L2: Consequences of TCF7L2 splicing for molecular function and risk for type 2 diabetes [Journal Article]. Current diabetes reports. 2010;10(6):444–451.

[120] Neve B, Le Bacquer O, Caron S, Huyvaert M, Leloire A, Poulain-Godefroy O, et al. Alternative human liver transcripts of TCF7L2 bind to the gluconeogenesis regulator HNF4α at the protein level [Journal Article]. Diabetologia. 2014;57(4):785–796.

[121] Lyssenko V, Lupi R, Marchetti P, Del Guerra S, Orho-Melander M, Almgren P, et al. Mechanisms by which common variants in the TCF7L2 gene increase risk of type 2 diabetes [Journal Article]. J Clin Invest. 2007;117(8):2155–63.

[122] Gaulton KJ, Nammo T, Pasquali L, Simon JM, Giresi PG, Fogarty MP, et al. A map of open chromatin in human pancreatic islets. Nat Genet. 2010 Mar;42(3):255–259.

[123] Stitzel ML, Sethupathy P, Pearson DS, Chines PS, Song L, Erdos MR, et al. Global epigenomic analysis of primary human pancreatic islets provides insights into type 2 diabetes susceptibility loci. Cell Metab. 2010 Nov;12(5):443–455.

[124] Viñuela A, Varshney A, van de Bunt M, Prasad RB, Asplund O, Bennett A, et al. Influence of genetic variants on gene expression in human pancreatic islets – implications for type 2 diabetes. bioRxiv. 2019;.

[125] Schafer SA, Tschritter O, Machicao F, Thamer C, Stefan N, Gallwitz B, et al. Impaired glucagon-like peptide-1-induced insulin secretion in carriers of transcription factor 7-like 2 (TCF7L2) gene polymorphisms [Journal Article]. Diabetologia. 2007;50(12):2443–50.

[126] Gaulton KJ, Nammo T, Pasquali L, Simon JM, Giresi PG, Fogarty MP, et al. A map of open chromatin in human pancreatic islets [Journal Article]. NatGenet. 2010;42(3):255–259.

[127] Xia Q, Deliard S, Yuan CX, Johnson ME, Grant SF. Characterization of the transcriptional machinery bound across the widely presumed type 2 diabetes causal variant, rs7903146, within TCF7L2 [Journal Article]. Eur J Hum Genet. 2015;23(1):103–9.

[128] Zhou Y, Oskolkov N, Shcherbina L, Ratti J, Kock KH, Su J, et al. HMGB1 binds to the rs7903146 locus in TCF7L2 in human pancreatic islets [Journal Article]. Molecular and cellular endocrinology. 2016;430:138–145.

[129] Dujic T, Bego T, Malenica M, Velija-Asimi Z, Ahlqvist E, Groop L, et al. Effects of TCF7L2 rs7903146 variant on metformin response in patients with type 2 diabetes. Bosn J Basic Med Sci. 2019 Nov;19(4):368–374.

[130] Holstein A, Hahn M, Koerner A, Stumvoll M, Kovacs P. TCF7L2 and therapeutic response to sulfonylureas in patients with type 2 diabetes. BMC Med Genet. 2011 Feb;12:30.

[131] Nakabayashi H, Taketa K, Miyano K, Yamane T, Sato J. Growth of human hepatoma cells lines with differentiated functions in chemically defined medium. Cancer Res. 1982 Sep;42(9):3858–3863.

[132] Graham FL, Smiley J, Russell WC, Nairn R. Characteristics of a human cell line transformed by DNA from human adenovirus type 5. J Gen Virol. 1977 Jul;36(1):59–74.

[133] Knowles BB, Howe CC, Aden DP. Human hepatocellular carcinoma cell lines secrete the major plasma proteins and hepatitis B surface antigen. Science. 1980 Jul;209(4455):497–499.

[134] Asfari M, Janjic D, Meda P, Li G, Halban PA, Wollheim CB. Establishment of 2-mercaptoethanol-dependent differentiated insulin-secreting cell lines. Endocrinology. 1992 Jan;130(1):167–178.

[135] Jumarie C, Malo C. Caco-2 cells cultured in serum-free medium as a model for the study of enterocytic differentiation in vitro. J Cell Physiol. 1991 Oct;149(1):24–33.

[136] Vasu S, McClenaghan NH, McCluskey JT, Flatt PR. Cellular responses of novel human pancreatic beta-cell line, 1.1B4 to hyperglycemia. Islets. 2013 Jul-Aug;5(4):170–177.

[137] Fischer-Posovszky P, Newell FS, Wabitsch M, Tornqvist HE. Human SGBS cells - a unique tool for studies of human fat cell biology. Obes Facts. 2008;1(4):184–189.

[138] Park JG, Oie HK, Sugarbaker PH, Henslee JG, Chen TR, Johnson BE, et al. Characteristics of cell lines established from human colorectal carcinoma. Cancer Res. 1987 Dec;47(24 Pt 1):6710–6718.

[139] Schneider U, Schwenk HU, Bornkamm G. Characterization of EBV-genome negative "null" and "T" cell lines derived from children with acute lymphoblastic leukemia and leukemic transformed non-Hodgkin lymphoma. Int J Cancer. 1977 May;19(5):621–626.

[140] Klaus S, Choy L, Champigny O, Cassard-Doulcier AM, Ross S, Spiegelman B, et al. Characterization of the novel brown adipocyte cell line HIB 1B. Adrenergic pathways involved in regulation of uncoupling protein gene expression [Journal Article]. Journal of cell science. 1994;107 ( Pt 1):313–319.

[141] Schreiber E, Matthias P, Müller MM, Schaffner W. Rapid detection of octamer binding proteins with 'mini extracts', prepared from a small number of cells [Journal Article]. Nucleic acids research. 1989;17(15):6419.

[142] Hellman LM, Fried MG. Electrophoretic mobility shift assay (EMSA) for detecting protein-nucleic acid interactions. Nat Protoc. 2007;2(8):1849–1861.

[143] Wiśniewski J, Zougman A, Nagaraj N, Mann M. Universal sample preparation method for proteome analysis. Nature methods. 2009 05;6:359–62.

[144] Hauck S, Dietter J, Degroote R, Hofmaier F, Zipplies J, Amann B, et al. Deciphering Membrane-Associated Molecular Processes in Target Tissue of Autoimmune Uveitis by Label-Free Quantitative Mass Spectrometry. Molecular cellular proteomics : MCP. 2010 10;9:2292–305.

[145] von Toerne C, Kahle M, Schäfer A, Ispiryan R, Blindert M, Hrabe De Angelis M, et al. Apoe, Mbl2, and Psp plasma protein levels correlate with diabetic phenotype in NZO mice–an optimized rapid workflow for SRM-based quantification. Journal of proteome research. 2013 March;12(3):1331—1343.

[146] Koba M, Konopa J. [Actinomycin D and its mechanisms of action]. Postepy Hig Med Dosw (Online). 2005;59:290–298.

[147] Xia Q, Chesi A, Manduchi E, Johnston BT, Lu S, Leonard ME, et al. The type 2 diabetes presumed causal variant within TCF7L2 resides in an element that controls the expression of ACSL5 [Journal Article]. Diabetologia. 2016;59(11):2360–2368.

[148] Bas A, Forsberg G, Hammarstroem S, Hammarstroem ML. Utility of the Housekeeping Genes 18S rRNA, β-Actin and Glyceraldehyde-3-Phosphate-Dehydrogenase for Normalization in Real-Time Quantitative Reverse Transcriptase-Polymerase Chain Reaction Analysis of Gene Expression in Human T Lymphocytes. Scandinavian Journal of Immunology. 2004;59(6):566–573.

[149] Rampersad SN. Multiple applications of Alamar Blue as an indicator of metabolic function and cellular health in cell viability bioassays [Journal Article]. Sensors (Basel, Switzerland). 2012;12(9):12347–12360.

[150] Hu X, Ivashkiv LB. Cross-regulation of Signaling Pathways by Interferon-γ: Implications for Immune Responses and Autoimmune Diseases. Immunity. 2009;31(4):539 – 550.

[151] Melén K, Keskinen P, Lehtonen A, Julkunen I. Interferon-induced gene expression and signaling in human hepatoma cell lines. Journal of Hepatology. 2000;33(5):764 – 772.

[152] Gao B, Wang H, Lafdil F, Feng D. STAT proteins – Key regulators of anti-viral responses, inflammation, and tumorigenesis in the liver. Journal of Hepatology. 2012;57(2):430 – 441.

[153] Hellman LM, Fried MG. Electrophoretic mobility shift assay (EMSA) for detecting protein-nucleic acid interactions [Journal Article]. Nature protocols. 2007;2(8):1849–1861.

[154] Privalov PL, Dragan AI, Crane-Robinson C. Interpreting protein/DNA interactions: distinguishing specific from non-specific and electrostatic from non-electrostatic components. Nucleic Acids Research. 2010 11;39(7):2483–2491.

[155] Stranger BE, Stahl EA, Raj T. Progress and promise of genome-wide association studies for human complex trait genetics. Genetics. 2011 Feb;187(2):367–383.

[156] Guo X, Gao L, Wang Y, Chiu DK, Wang T, Deng Y. Advances in long noncoding RNAs: identification, structure prediction and function annotation. Brief Funct Genomics. 2016 Jan;15(1):38–46.

[157] Liu B, Li J, Cairns MJ. Identifying miRNAs, targets and functions. Brief Bioinformatics. 2014 Jan;15(1):1–19.

[158] de Laat W, Duboule D. Topology of mammalian developmental enhancers and their regulatory landscapes. Nature. 2013 Oct;502(7472):499–506.

[159] Mora A, Sandve GK, Gabrielsen OS, Eskeland R. In the loop: promoter-enhancer interactions and bioinformatics. Brief Bioinformatics. 2016 11;17(6):980–995.

[160] Barutcu AR, Fritz AJ, Zaidi SK, van Wijnen AJ, Lian JB, Stein JL, et al. C-ing the Genome: A Compendium of Chromosome Conformation Capture Methods to Study Higher-Order Chromatin Organization. J Cell Physiol. 2016 Jan;231(1):31–35.

[161] Rojano E, Seoane P, Ranea JAG, Perkins JR. Regulatory variants: from detection to predicting impact. Brief Bioinformatics. 2019 09;20(5):1639–1654.

[162] Claussnitzer M, Dankel SN, Klocke B, Grallert H, Glunk V, Berulava T, et al. Leveraging cross-species transcription factor binding site patterns: from diabetes risk loci to disease mechanisms. Cell. 2014 Jan;156(1-2):343–358.

[163] Jutras BL, Verma A, Stevenson B. Identification of novel DNA-binding proteins using DNA-affinity chromatography/pull down. Curr Protoc Microbiol. 2012 Feb;Chapter 1:Unit1F.1.

[164] Butter F, Davison L, Viturawong T, Scheibe M, Vermeulen M, Todd JA, et al. Proteome-wide analysis of disease-associated SNPs that show allele-specific transcription factor binding. PLoS Genet. 2012 Sep;8(9):e1002982.

[165] Merl J, Ueffing M, Hauck SM, von Toerne C. Direct comparison of MS-based label-free and SILAC quantitative proteome profiling strategies in primary retinal MÃ¼ller cells. Proteomics. 2012 Jun;12(12):1902–1911.

[166] Terzi F, Cambridge S. An Overview of Advanced SILAC-Labeling Strategies for Quantitative Proteomics. Meth Enzymol. 2017;585:29–47.

[167] Pappireddi N, Martin L, Wuehr M. A Review on Quantitative Multiplexed Proteomics. Chembiochem. 2019 05;20(10):1210–1224.

[168] Lee H, Qian K, von Toerne C, Hoerburger L, Claussnitzer M, Hoffmann C, et al. Allele-specific quantitative proteomics unravels molecular mechanisms modulated by cis-regulatory PPARG locus variation. Nucleic Acids Res. 2017 Apr;45(6):3266–3279.

[169] McAnuff MA, Rettig GR, Rice KG. Potency of siRNA versus shRNA mediated knockdown in vivo. J Pharm Sci. 2007 Nov;96(11):2922–2930.

[170] Han J, Cai J, Borjihan W, Ganbold T, Rana TM, Baigude H. Preparation of novel curdlan nanoparticles for intracellular siRNA delivery. Carbohydr Polym. 2015 Mar;117:324–330.

[171] Shearer RF, Saunders DN. Experimental design for stable genetic manipulation in mammalian cell lines: lentivirus and alternatives. Genes Cells. 2015 Jan;20(1):1–10.

[172] Matejka K, Stoeckler F, Salomon M, Ensenauer R, Reischl E, Hoerburger L, et al. Dynamic modelling of an ACADS genotype in fatty acid oxidation - Application of cellular models for the analysis of common genetic variants. PLoS ONE. 2019;14(5):e0216110.

[173] Smale ST. Luciferase assay. Cold Spring Harb Protoc. 2010 May;2010(5):pdb.prot5421.

[174] Kereszturi E, Király O, Sahin-Tóth M. Minigene analysis of intronic variants in common SPINK1 haplotypes associated with chronic pancreatitis. Gut. 2009 Apr;58(4):545–549.

[175] Gade P, Kalvakolanu DV. Chromatin immunoprecipitation assay as a tool for analyzing transcription factor activity. Methods Mol Biol. 2012;809:85–104.

[176] Heckman CA, Boxer LM. Allele-specific analysis of transcription factors binding to promoter regions. Methods. 2002 Jan;26(1):19–26.

[177] Collas P. The current state of chromatin immunoprecipitation. Mol Biotechnol. 2010 May;45(1):87–100.

[178] Skelin M, Rupnik M, Cencic A. Pancreatic beta cell lines and their applications in diabetes mellitus research [Journal Article]. Altex. 2010;27(2):105–13.

[179] Aguayo-Mazzucato C, DiIenno A, Hollister-Lock J, Cahill C, Sharma A, Weir G, et al. MAFA and T3 Drive Maturation of Both Fetal Human Islets and Insulin-Producing Cells Differentiated From hESC [Journal Article]. J Clin Endocrinol Metab. 2015;100(10):3651–9.

[180] Wang H, Brun T, Kataoka K, Sharma AJ, Wollheim CB. MAFA controls genes implicated in insulin biosynthesis and secretion [Journal Article]. Diabetologia. 2007;50(2):348–58.

[181] Matsuoka TA, Artner I, Henderson E, Means A, Sander M, Stein R. The MafA transcription factor appears to be responsible for tissue-specific expression of insulin [Journal Article]. Proc Natl Acad Sci U S A. 2004;101(9):2930–3.

[182] Aguayo-Mazzucato C, Koh A, El Khattabi I, Li WC, Toschi E, Jermendy A, et al. Mafa expression enhances glucose-responsive insulin secretion in neonatal rat beta cells [Journal Article]. Diabetologia. 2011;54(3):583–93.

[183] Aguayo-Mazzucato C, Zavacki AM, Marinelarena A, Hollister-Lock J, El Khattabi I, Marsili A, et al. Thyroid hormone promotes postnatal rat pancreatic beta-cell development and glucose-responsive insulin secretion through MAFA [Journal Article]. Diabetes. 2013;62(5):1569–80.

[184] Yoshihara E, Wei Z, Lin CS, Fang S, Ahmadian M, Kida Y, et al. ERRgamma Is Required for the Metabolic Maturation of Therapeutically Functional Glucose-Responsive beta Cells [Journal Article]. Cell Metab. 2016;23(4):622–34.

[185] Blum B, Hrvatin S, Schuetz C, Bonal C, Rezania A, Melton DA. Functional beta-cell maturation is marked by an increased glucose threshold and by expression of urocortin 3 [Journal Article]. Nat Biotechnol. 2012;30(3):261–4.

[186] van der Meulen T, Donaldson CJ, Caceres E, Hunter AE, Cowing-Zitron C, Pound LD, et al. Urocortin3 mediates somatostatin-dependent negative feedback control of insulin secretion [Journal Article]. Nat Med. 2015;21(7):769–76.

[187] Li C, Chen P, Vaughan J, Blount A, Chen A, Jamieson PM, et al. Urocortin III is expressed in pancreatic beta-cells and stimulates insulin and glucagon secretion [Journal Article]. Endocrinology. 2003;144(7):3216–24.

[188] Li C, Chen P, Vaughan J, Lee KF, Vale W. Urocortin 3 regulates glucose-stimulated insulin secretion and energy homeostasis [Journal Article]. Proc Natl Acad Sci U S A. 2007;104(10):4206–11.

[189] Dean PM, Matthews EK. Electrical activity in pancreatic islet cells [Journal Article]. Nature. 1968;219(5152):389–90.

[190] Ashcroft FM, Rorsman P. Electrophysiology of the pancreatic beta-cell [Journal Article]. Prog Biophys Mol Biol. 1989;54(2):87–143.

[191] McCluskey JT, Hamid M, Guo-Parke H, McClenaghan NH, Gomis R, Flatt PR. Development and functional characterization of insulin-releasing human pancreatic beta cell lines produced by electrofusion [Journal Article]. J Biol Chem. 2011;286(25):21982–92.

[192] Green AD, Vasu S, McClenaghan NH, Flatt PR. Pseudoislet formation enhances gene expression, insulin secretion and cytoprotective mechanisms of clonal human insulin-secreting 1.1B4 cells. Pflugers Arch. 2015 Oct;467(10):2219–2228.

[193] Vasu S, McClenaghan NH, McCluskey JT, Flatt PR. Cellular responses of novel human pancreatic β-cell line, 1.1B4 to hyperglycemia [Journal Article]. Islets. 2013;5(4):170–177.

[194] Chaffey JR, Young J, Leslie KA, Partridge K, Akhbari P, Dhayal S, et al. Investigation of the utility of the 1.1B4 cell as a model human beta cell line for study of persistent enteroviral infection [Journal Article]. Scientific Reports. 2021;11(1):15624.

[195] Ravassard P, Hazhouz Y, Pechberty S, Bricout-Neveu E, Armanet M, Czernichow P, et al. A genetically engineered human pancreatic Î² cell line exhibiting glucose-inducible insulin secretion. J Clin Invest. 2011 Sep;121(9):3589–3597.

[196] Nagarajan SR, Paul-Heng M, Krycer JR, Fazakerley DJ, Sharland AF, Hoy AJ. Lipid and glucose metabolism in hepatocyte cell lines and primary mouse hepatocytes: a comprehensive resource for in vitro studies of hepatic metabolism. Am J Physiol Endocrinol Metab. 2019 04;316(4):E578–E589.

[197] Meex SJ, Andreo U, Sparks JD, Fisher EA. Huh-7 or HepG2 cells: which is the better model for studying human apolipoprotein-B100 assembly and secretion? J Lipid Res. 2011 Jan;52(1):152–158.

[198] Steenbergen R, Oti M, Ter Horst R, Tat W, Neufeldt C, Belovodskiy A, et al. Establishing normal metabolism and differentiation in hepatocellular carcinoma cells by culturing in adult human serum. Sci Rep. 2018 08;8(1):11685.

[199] Ruiz-Ojeda FJ, Rupérez AI, Gomez-Llorente C, Gil A, Aguilera CM. Cell Models and Their Application for Studying Adipogenic Differentiation in Relation to Obesity: A Review. Int J Mol Sci. 2016 Jun;17(7).

[200] Fischer-Posovszky P, Newell FS, Wabitsch M, Tornqvist HE. Human SGBS Cells – a Unique Tool for Studies of Human Fat Cell Biology [Journal Article]. Obesity Facts. 2008;1(4):184–189.

[201] Boam DS, Docherty K. A tissue-specific nuclear factor binds to multiple sites in the human insulin-gene enhancer. Biochem J. 1989 Nov;264(1):233–239.

[202] Uhlén M, Fagerberg L, Hallstroem BM, Lindskog C, Oksvold P, Mardinoglu A, et al. Proteomics. Tissue-based map of the human proteome. Science. 2015 Jan;347(6220):1260419.

[203] McKinnon CM, Docherty K. Pancreatic duodenal homeobox-1, PDX-1, a major regulator of beta cell identity and function. Diabetologia. 2001 Oct;44(10):1203–1214.

[204] MacFarlane WM, Read ML, Gilligan M, Bujalska I, Docherty K. Glucose modulates the binding activity of the beta-cell transcription factor IUF1 in a phosphorylation-dependent manner. Biochem J. 1994 Oct;303 ( Pt 2):625–631.

[205] Wu H, MacFarlane WM, Tadayyon M, Arch JR, James RF, Docherty K. Insulin stimulates pancreatic-duodenal homoeobox factor-1 (PDX1) DNA-binding activity and insulin promoter activity in pancreatic beta cells. Biochem J. 1999 Dec;344 Pt 3:813–818.

[206] Buteau J, Roduit R, Susini S, Prentki M. Glucagon-like peptide-1 promotes DNA synthesis, activates phosphatidylinositol 3-kinase and increases transcription factor pancreatic and duodenal homeobox gene 1 (PDX-1) DNA binding activity in beta (INS-1)-cells. Diabetologia. 1999 Jul;42(7):856–864.

[207] Wang X, Sterr M, Burtscher I, Chen S, Hieronimus A, Machicao F, et al. Genome-wide analysis of PDX1 target genes in human pancreatic progenitors [Journal Article]. Molecular metabolism. 2018;.

[208] Ohneda K, Mirmira RG, Wang J, Johnson JD, German MS. The homeodomain of PDX-1 mediates multiple protein-protein interactions in the formation of a transcriptional activation complex on the insulin promoter. Mol Cell Biol. 2000 Feb;20(3):900–911.

[209] Zhou Y, Park SY, Su J, Bailey K, Ottosson-Laakso E, Shcherbina L, et al. TCF7L2 is a master regulator of insulin production and processing [Journal Article]. Human Molecular Genetics. 2014;23(24):6419–6431.

[210] Fujita Y, Chui JW, King DS, Zhang T, Seufert J, Pownall S, et al. Pax6 and Pdx1 are required for production of glucose-dependent insulinotropic polypeptide in proglucagon-expressing L cells. Am J Physiol Endocrinol Metab. 2008 Sep;295(3):E648–657.

[211] Yi F, Brubaker PL, Jin T. TCF-4 mediates cell type-specific regulation of proglucagon gene expression by beta-catenin and glycogen synthase kinase-3beta [Journal Article]. The Journal of biological chemistry. 2005;280(2):1457–1464.

[212] Schrem H, Klempnauer J, Borlak J. Liver-enriched transcription factors in liver function and development. Part II: the C/EBPs and D site-binding protein in cell cycle control, carcinogenesis, circadian gene regulation, liver regeneration, apoptosis, and liver-specific gene regulation [Journal Article]. Pharmacol Rev. 2004;56(2):291–330.

[213] Shuman JD, Cheong J, Coligan JE. ATF-2 and C/EBPalpha can form a heterodimeric DNA binding complex in vitro. Functional implications for transcriptional regulation [Journal Article]. J Biol Chem. 1997;272(19):12793–800.

[214] Vallejo M, Ron D, Miller CP, Habener JF. C/ATF, a member of the activating transcription factor family of DNA-binding proteins, dimerizes with CAAT/enhancer-binding proteins and directs their binding to cAMP response elements [Journal Article]. Proc Natl Acad Sci U S A. 1993;90(10):4679–83.

[215] Ossipow V, Descombes P, Schibler U. CCAAT/enhancer-binding protein mRNA is translated into multiple proteins with different transcription activation potentials [Journal Article]. Proc Natl Acad Sci U S A. 1993;90(17):8219–23.

[216] Reckzeh K, Cammenga J. Molecular mechanisms underlying deregulation of C/EBPα in acute myeloid leukemia [Journal Article]. International journal of hematology. 2010;91:557–68.

[217] Antonson P, Xanthopoulos KG. Molecular cloning, sequence, and expression patterns of the human gene encoding CCAAT/enhancer binding protein alpha (C/EBP alpha) [Journal Article]. Biochem Biophys Res Commun. 1995;215(1):106–13.

[218] Baer M, Johnson PF. Generation of truncated C/EBPbeta isoforms by in vitro proteolysis [Journal Article]. J Biol Chem. 2000;275(34):26582–90.

[219] Bosch F, Sabater J, Valera A. Insulin inhibits liver expression of the CCAAT/enhancer-binding protein beta [Journal Article]. Diabetes. 1995;44(3):267–71.

[220] Lu M, Seufert J, Habener JF. Pancreatic beta-cell-specific repression of insulin gene transcription by CCAAT/enhancer-binding protein beta. Inhibitory interactions with basic helix-loop-helix transcription factor E47. J Biol Chem. 1997 Nov;272(45):28349–28359.

[221] Loder MK, da Silva Xavier G, McDonald A, Rutter GA. TCF7L2 controls insulin gene expression and insulin secretion in mature pancreatic beta-cells. Biochem Soc Trans. 2008 Jun;36(Pt 3):357–359.

[222] Cooper C, Henderson A, Artandi S, Avitahl N, Calame K. Ig/EBP (C/EBP gamma) is a transdominant negative inhibitor of C/EBP family transcriptional activators [Journal Article]. Nucleic Acids Res. 1995;23(21):4371–7.

[223] Parkin SE, Baer M, Copeland TD, Schwartz RC, Johnson PF. Regulation of CCAAT/enhancer-binding protein (C/EBP) activator proteins by heterodimerization with C/EBPgamma (Ig/EBP) [Journal Article]. J Biol Chem. 2002;277(26):23563–72.

[224] Friedman AD, Landschulz WH, McKnight SL. CCAAT/enhancer binding protein activates the promoter of the serum albumin gene in cultured hepatoma cells [Journal Article]. Genes Dev. 1989;3(9):1314–22.

[225] O'Shea JJ, Murray PJ. Cytokine signaling modules in inflammatory responses. Immunity. 2008 Apr;28(4):477–487.

[226] Wormald S, Hilton DJ. Inhibitors of cytokine signal transduction. J Biol Chem. 2004 Jan;279(2):821–824.

[227] Gurzov EN, Stanley WJ, Pappas EG, Thomas HE, Gough DJ. The JAK/STAT pathway in obesity and diabetes. FEBS J. 2016 08;283(16):3002–3015.

[228] Berryman DE, Glad CA, List EO, Johannsson G. The GH/IGF-1 axis in obesity: pathophysiology and therapeutic considerations. Nat Rev Endocrinol. 2013 Jun;9(6):346–356.

[229] Wang H, Zhang Y, Xia F, Zhang W, Chen P, Yang G. Protective effect of silencing Stat1 on high glucose-induced podocytes injury via Forkhead transcription factor O1-regulated the oxidative stress response. BMC Mol Cell Biol. 2019 07;20(1):27.

[230] Sisler JD, Morgan M, Raje V, Grande RC, Derecka M, Meier J, et al. The Signal Transducer and Activator of Transcription 1 (STAT1) Inhibits Mitochondrial Biogenesis in Liver and Fatty Acid Oxidation in Adipocytes. PLoS ONE. 2015;10(12):e0144444.

[231] Goodwin GH, Johns EW. Isolation and characterisation of two calf-thymus chromatin non-histone proteins with high contents of acidic and basic amino acids [Journal Article]. Eur J Biochem. 1973;40(1):215–9.

[232] Catez F, Hock R. Binding and interplay of HMG proteins on chromatin: lessons from live cell imaging [Journal Article]. Biochim Biophys Acta. 2010;1799(1-2):15–27.

[233] Reeves R, Beckerbauer L. HMGI/Y proteins: flexible regulators of transcription and chromatin structure [Journal Article]. Biochim Biophys Acta. 2001;1519(1-2):13–29.

[234] Chin MT, Pellacani A, Wang H, Lin SSJ, Jain MK, Perrella MA, et al. Enhancement of Serum-response Factor-dependent Transcription and DNA Binding by the Architectural Transcription Factor HMG-I(Y) [Journal Article]. Journal of Biological Chemistry. 1998;273(16):9755–9760.

[235] Sgarra R, Rustighi A, Tessari MA, Di Bernardo J, Altamura S, Fusco A, et al. Nuclear phosphoproteins HMGA and their relationship with chromatin structure and cancer [Journal Article]. FEBS Letters. 2004;574(1):1–8.

[236] Brunetti A, Manfioletti G, Chiefari E, Goldfine ID, Foti D. Transcriptional regulation of human insulin receptor gene by the high-mobility group protein HMGI(Y) [Journal Article]. The FASEB Journal. 2001;15(2):492–500.

[237] Xia Q, Deliard S, Yuan CX, Johnson ME, Grant SF. Characterization of the transcriptional machinery bound across the widely presumed type 2 diabetes causal variant, rs7903146, within TCF7L2 [Journal Article]. Eur J Hum Genet. 2015;23(1):103–9.

[238] Zhou Y, Oskolkov N, Shcherbina L, Ratti J, Kock KH, Su J, et al. HMGB1 binds to the rs7903146 locus in TCF7L2 in human pancreatic islets [Journal Article]. Mol Cell Endocrinol. 2016;430:138–45.

[239] Bowman TA, O'Keeffe KR, D'Aquila T, Yan QW, Griffin JD, Killion EA, et al. Acyl CoA synthetase 5 (ACSL5) ablation in mice increases energy expenditure and insulin sensitivity and delays fat absorption. Mol Metab. 2016 Mar;5(3):210–220.

[240] Ansari IH, Longacre MJ, Stoker SW, Kendrick MA, O'Neill LM, Zitur LJ, et al. Characterization of Acyl-CoA synthetase isoforms in pancreatic beta cells: Gene silencing shows participation of ACSL3 and ACSL4 in insulin secretion. Arch Biochem Biophys. 2017 03;618:32–43.

[241] Qiu Y, Corkey B, Deeney J, Li C. Acyl-CoA Synthetase Inhibition Protects Clonal Pancreatic Beta-cell from Effects of Chronic Excess Nutrients. Current Developments in Nutrition. 2020 05;4(Supplement 2):652–652.

[242] Adissu HA, Estabel J, Sunter D, Tuck E, Hooks Y, Carragher DM, et al. Histopathology reveals correlative and unique phenotypes in a high-throughput mouse phenotyping screen. Disease Models & Mechanisms. 2014 05;7(5):515–524.

[243] Moreno-Navarrete JM, Ortega F, Moreno M, Xifra G, Ricart W, Fernández-Real JM. PRDM16 sustains white fat gene expression profile in human adipocytes in direct relation with insulin action. Molecular and Cellular Endocrinology. 2015;405:84–93.

[244] Jukic I, Kostic S, Filipović N, Ensor L, Ivandic M, Dukic J, et al. Changes in expression of special AT-rich sequence binding protein 1 and phosphatase and tensin homologue in kidneys of diabetic rats during ageing. Nephrology, dialysis, transplantation : official publication of the European Dialysis and Transplant Association - European Renal Association. 2018 06;33:1075.

[245] Gosmain Y, Katz LS, Masson MH, Cheyssac C, Poisson C, Philippe J. Pax6 Is Crucial for β-Cell Function, Insulin Biosynthesis, and Glucose-Induced Insulin Secretion. Molecular Endocrinology. 2012 04;26(4):696–709.

[246] Jia L, Jiang Y, Xinzhi L, Chen Z. Purβ promotes hepatic glucose production by increasing Adcy6 transcription. Molecular Metabolism. 2019 11;31.

[247] Chriett S, Lindqvist A, Shcherbina L, Edlund A, Abels M, Asplund O, et al. SCRT1 is a novel beta cell transcription factor with insulin regulatory properties. Molecular and Cellular Endocrinology. 2021;521:111107.

[248] Giguere V. Transcriptional Control of Energy Homeostasis by the Estrogen-Related Receptors. Endocrine Reviews. 2008 10;29(6):677–696.

[249] Nakakura EK, Watkins DN, Schuebel KE, Sriuranpong V, Borges MW, Nelkin BD, et al. Mammalian Scratch: a neural-specific Snail family transcriptional repressor. Proc Natl Acad Sci USA. 2001 Mar;98(7):4010–4015.

[250] Mohan WS, Scheer E, Wendling O, Metzger D, Tora L. TAF10 (TAF(II)30) is necessary for TFIID stability and early embryogenesis in mice. Mol Cell Biol. 2003 Jun;23(12):4307–4318.

[251] Foti D, Chiefari E, Fedele M, Iuliano R, Brunetti L, Paonessa F, et al. Lack of the architectural factor HMGA1 causes insulin resistance and diabetes in humans and mice [Journal Article]. Nat Med. 2005;11(7):765–73.

[252] Dunham I, Kundaje A, Aldred SF, Collins PJ, Davis CA, Doyle F, et al. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012 Sep;489(7414):57–74.

[253] Ohno S. So much "junk" DNA in our genome. Brookhaven Symp Biol. 1972;23:366–370.

[254] de Koning AP, Gu W, Castoe TA, Batzer MA, Pollock DD. Repetitive elements may comprise over two-thirds of the human genome. PLoS Genet. 2011 Dec;7(12):e1002384.

[255] Palazzo AF, Gregory TR. The case for junk DNA. PLoS Genet. 2014 May;10(5):e1004351.

[256] Cowley M, Oakey RJ. Transposable elements re-wire and fine-tune the transcriptome. PLoS Genet. 2013;9(1):e1003234.

[257] Sears RM, May DG, Roux KJ. BioID as a Tool for Protein-Proximity Labeling in Living Cells. Methods Mol Biol. 2019;2012:299–313.

[258] Trinkle-Mulcahy L. Recent advances in proximity-based labeling methods for interactome mapping. F1000Res. 2019;8.

[259] Long S, Brown KM, Sibley LD. CRISPR-mediated Tagging with BirA Allows Proximity Labeling in Toxoplasma gondii. Bio Protoc. 2018 Mar;8(6).

[260] Benecke AG, Eilebrecht S. RNA-Mediated Regulation of HMGA1 Function. Biomolecules. 2015;5(2):943–957.

[261] Tan JA, Song J, Chen Y, Durrin LK. Phosphorylation-dependent interaction of SATB1 and PIAS1 directs SUMO-regulated caspase cleavage of SATB1. Mol Cell Biol. 2010 Jun;30(11):2823–2836.

[262] Liu B, Gross M, ten Hoeve J, Shuai K. A transcriptional corepressor of Stat1 with an essential LXXLL signature motif. Proc Natl Acad Sci USA. 2001 Mar;98(6):3203–3207.

[263] Kumar PP, Bischof O, Purbey PK, Notani D, Urlaub H, Dejean A, et al. Functional interaction between PML and SATB1 regulates chromatin-loop architecture and transcription of the MHC class I locus. Nat Cell Biol. 2007 Jan;9(1):45–56.

# Acknowledgements

I want to thank **Katharina Eiseler, Andreas Schmidt, Luiza Kaneva** and **Lara Unger** for always supporting and motivating me.

Also, I would like to thank my husband **Gabor** for his emotional support throughout this project.

My gratitude goes to **Manuela Hubersberger** for her patience and helpfulness and **Dr. Juliane Merl-Pham** and **Fabian Gruhn** for providing the LC-MS/MS measurements and data analysis.
I am very thankful for having **Beate Rauscher** and **Daniela Kolmeder** helping me in my last stage of the thesis.

Furthermore, I like to thank **Prof. Hans Hauner** for providing lab facilities. Thank you **Prof. Dr. Jochen Seissler** (LMU) for providing pig islets to prepare native protein extracts.

I would like to express my graditute to the **Deutsche Forschungsgemeinschaft** for funding this project.

I want to thank my advisor **Prof. Dr. med. Heiko Witt** for giving me this opportunity, his humor and knowledge.

Last but not least, i would like to thank my mentor **Dr. Helmut Laumen** for the support during this project, for his relentless optimism, motivation and knowledge.

# Appendix

**Table A1:** p Values from **Figure 5.2**

| Gene | Cell line | time point | p Value |
|---|---|---|---|
| TCF7L2 | Huh7 | 2h | 0.125 |
| TCF7L2 | Huh7 | 4h | **1*10$^{-3}$** |
| TCF7L2 | Huh7 | 6h | **0.03** |
| TCF7L2 | Huh7 | 24h | **3*10$^{-3}$** |
| TCF7L2 | HepG2 | 2h | 0.082 |
| TCF7L2 | HepG2 | 4h | 0.0696 |
| TCF7L2 | HepG2 | 6h | **3*10$^{-3}$** |
| TCF7L2 | HepG2 | 24h | 0.62 |
| ACSL5 | Huh7 | 2h | 0.94 |
| ACSL5 | Huh7 | 4h | 0.13 |
| ACSL5 | Huh7 | 6h | **0.01** |
| ACSL5 | Huh7 | 24h | **3*10$^{-4}$** |
| ACSL5 | HepG2 | 2h | 0.83 |
| ACSL5 | HepG2 | 4h | 0.09 |
| ACSL5 | HepG2 | 6h | **8*10$^{-4}$** |
| ACSL5 | HepG2 | 24h | **7*10$^{-4}$** |
| STAT1 | Huh7 | 2h | **5*10$^{-4}$** |
| STAT1 | Huh7 | 4h | **4*10$^{-4}$** |
| STAT1 | Huh7 | 6h | **1*10$^{-4}$** |
| STAT1 | Huh7 | 24h | **1*10$^{-4}$** |
| STAT1 | HepG2 | 2h | **8*10$^{-4}$** |
| STAT1 | HepG2 | 4h | **5*10$^{-4}$** |
| STAT1 | HepG2 | 6h | **4*10$^{-4}$** |
| STAT1 | HepG2 | 24h | **1*10$^{-4}$** |
| STAT2 | Huh7 | 2h | **0.02** |
| STAT2 | Huh7 | 4h | **1*10$^{-3}$** |
| STAT2 | Huh7 | 6h | **1*10$^{-3}$** |
| STAT2 | Huh7 | 24h | **5*10$^{-3}$** |
| STAT2 | HepG2 | 2h | **3*10$^{-4}$** |
| STAT2 | HepG2 | 4h | **2*10$^{-4}$** |
| STAT2 | HepG2 | 6h | **1*10$^{-4}$** |
| STAT2 | HepG2 | 24h | **1*10$^{-4}$** |
| STAT3 | Huh7 | 2h | **0.01** |
| STAT3 | Huh7 | 4h | **0.02** |

*Continued on next page*

Table A1 – *Continued from previous page*

| Gene | Cell line | time point | p Value |
|---|---|---|---|
| STAT3 | Huh7 | 6h | **8*10$^{-4}$** |
| STAT3 | Huh7 | 24h | **0.03** |
| STAT3 | HepG2 | 2h | **0.02** |
| STAT3 | HepG2 | 4h | **2*10$^{-4}$** |
| STAT3 | HepG2 | 6h | **3*10$^{-3}$** |
| STAT3 | HepG2 | 24h | 0.05 |
| DDB2 | Huh7 | 2h | 0.8 |
| DDB2 | Huh7 | 4h | 0.65 |
| DDB2 | Huh7 | 6h | 0.47 |
| DDB2 | Huh7 | 24h | 0.57 |
| DDB2 | HepG2 | 2h | 0.65 |
| DDB2 | HepG2 | 4h | 0.38 |
| DDB2 | HepG2 | 6h | 0.94 |
| DDB2 | HepG2 | 24h | 0.23 |
| SOCS1 | Huh7 | 2h | **1*10$^{-3}$** |
| SOCS1 | Huh7 | 4h | **3*10$^{-4}$** |
| SOCS1 | Huh7 | 6h | **3*10$^{-3}$** |
| SOCS1 | Huh7 | 24h | **2*10$^{-3}$** |
| SOCS1 | HepG2 | 2h | **1*10$^{-4}$** |
| SOCS1 | HepG2 | 4h | **1*10$^{-3}$** |
| SOCS1 | HepG2 | 6h | **1*10$^{-4}$** |
| SOCS1 | HepG2 | 24h | **1*10$^{-4}$** |

**Table A2:** EMSA conditions from **Figure 5.14**

| Cell line | Nuclear extract [µg] | Poly(dIdC) [ng] | Oligonucleotide |
|---|---|---|---|
| HEK293 | 2.4 | 350 | 45 bp |
| HepG2 | 3.4 | 700 | 45 bp |
| Huh7 | 3 | 350 | 45 bp |
| 1.1B4 | 4 | 350 | 45 bp |
| Ins1 | 2.2 | 540 | 45 bp |
| Pig islets | 9 | 350 | 45 bp |
| Caco2 | 1.7 | 700 | 45 bp |
| HT29 | 3.7 | 540 | 45 bp |
| NCI H716 | 7 | 700 | 45 bp |
| 3T3L1 d0 | 4.2 | 700 | 45 bp |
| 3T3L1 d9 | 5 | 700 | 45 bp |
| 3T3L1 d15 | 3.7 | 700 | 45 bp |
| HIB1B d0 | 2.4 | 560 | 45 bp |
| HIB1B d9 | 5 | 700 | 45 bp |

*Continued on next page*

Table A2 – *Continued from previous page*

| **Cell line** | **Nuclear extract [µg]** | **Poly(dIdC) [ng]** | **Oligonucleotide** |
|---|---|---|---|
| SGBS d18 | 4 | 700 | 45 bp |
| THP1 | 1 | 540 | 45 bp |
| U937 | 1 | 540 | 45 bp |
| HepG2 | 4.2 | 700 | 121 bp |
| 1.1B4 | 6 | 350 | 121 bp |
| Ins1 | 4.8 | 280 | 121 bp |
| Caco2 | 2.6 | 280 | 121 bp |
| HT29 | 3.7 | 560 | 121 bp |
| HIB1B d0 | 2.4 | 560 | 121 bp |

**Table A3:** Protein size in western blot

| **Protein** | **Protein size [kDa]** | **Gel** |
|---|---|---|
| Prdm16 | 170 | 10 % |
| HLTF | 114 | 10 % |
| PARP1 | 113 | 10 % |
| NKRF | 100 | 10 % |
| SATB1 | 90 | 10 % |
| STAT1 | 90 | 10 % |
| CUX1 | 80 | 10 % |
| TCF7L2 | 68 | 10 % |
| YY1 | 68 | 10 % |
| Lamin B1 | 66 | 10 % |
| MYEF2 | 64 | 10 % |
| PAX6 | 47 | 15 % |
| LHX2 | 46 | 15 % |
| NR2F1 | 46 | 15 % |
| NKX6.1 | 38 | 15 % |
| CEBPB long | 38 | 15 % |
| CEBPA 1+2 | 38 | 15 % |
| GTF2B | 36 | 15 % |
| CEBPB medium | 34 | 15 % |
| PURB | 33 | 15 % |
| PDX1 | 31 | 15 % |
| MED19 | 26 | 15 % |
| CEBPB short | 20 | 15 % |
| HMGA2 | 18 | 15 % |
| HMGA1 | 17 | 15 % |
| CEBPG | 17 | 15 % |

*Continued on next page*

Table A3 – *Continued from previous page*

| Protein | Protein size [kDa] | Gel |
|---|---|---|
| ZNF593 | 15 | 15 % |

**Table A4:** Vectors

| Name | Company |
|---|---|
| pcDNA3.1 CEBPA Flag | Genscript #OHu21531 |
| pcDNA3.1 CEBPA2 Flag | Genscript #OHu20497 |
| pcDNA3.1 CEBPB Long Flag | Genscript #OHu27318 |
| pcDNA3.1 CEBPB Medium Flag | Genscript #OHu27612 |
| pcDNA3.1 CEBPB Short Flag | Genscript #OHu27597 |
| pcDNA3.1 CEBPG Flag | Genscript #OHu10827 |
| pcDNA3.1 Cux1 Flag | Genscript #OHu23887 |
| pcDNA3.1 DYK | Genscript pcDNA3.1+/C-(K)-DYK |
| pcDNA3.1 GTF2B Flag | Genscript #OHu02972 |
| pcDNA3.1 HLTF Flag | Genscript #OHu03219D |
| pcDNA3.1 HMGA1 Flag | Genscript #OHu10965D |
| pcDNA3.1 HMGA2 Flag | Genscript #OHu25597 |
| pcDNA3.1 LHX2 Flag | Genscript #OHu13181 |
| pcDNA3.1 MED19 Flag | Genscript #OHu07037D |
| pcDNA3.1 MYEF2 Flag | Genscript #OHu24268D |
| pcDNA3.1 NKRF Flag | Genscript #OHu04603D |
| pcDNA3.1 NKX6-1 Flag | Genscript #OHu05283D |
| pcDNA3.1 NR2F1 Flag | Genscript #OHu23866 |
| pcDNA3.1 PARP1 Flag | Genscript #OHu25551D |
| pcDNA3.1 PAX6 Flag | Genscript #OHu20414D |
| pcDNA3.1 PDX1 Flag | Genscript #OHu19441D |
| pcDNA3.1 PurB Flag | Genscript #OHu12284 |
| pcDNA3.1 SATB1 Flag | Genscript #OHu18985D |
| pcDNA3.1 STAT1 Flag | Genscript #OHu19484D |
| pcDNA3.1 ZNF593 Flag | Genscript #OHu30985D |
| pGL4.22 | Promega, Madison, USA |
| pGL-TCF7L2Prom-868 | Cloned by P. D'Albora, Paediatric Nutritional Medicine, TUM |
| pGL-TCF7L2Prom-868 EnhC | Cloned by P. D'Albora and K. Eiseler, Paediatric Nutritional Medicine, TUM, Germany |
| pGL-TCF7L2Prom-868 EnhT | Cloned by P. D'Albora and K. Eiseler, Paediatric Nutritional Medicine, TUM, Germany |
| pRL-Ubi | Promega, Madison, USA |
| pTK6 | provided by Nutritional Medicine, TUM, Germany |

Table A4 – *Continued from previous page*

| Name | Company |
|---|---|
| pTK6-TCF7L2-rs7903146 C | provided by H. Lee, Nutritional Medicine, TUM, Germany |
| pTK6-TCF7L2-rs7903146 T | provided by H. Lee, Nutritional Medicine, TUM, Germany |

**Table A5:** Antibodies

| Gene ID | Name | Source | Company |
|---|---|---|---|
| CEBPA | CEBPA | mouse | Santa Cruz Biotechnology 36531 |
| CEBPG | CEBPG | rabbit | Sigma-Aldrich HPA012024 |
| CEBPG | CEBPG | rabbit | Biomol 151647.50 |
| FLAG | anti-Flag M2 | mouse | Sigma-Aldrich F3165 |
| HMGA2 | HMGA2 | rabbit | Thermo Fisher PA5-21320 |
| IgG | IgG | goat | Sigma-Aldrich I5256 |
| IgG | IgG | mouse | Sigma-Aldrich I5381 |
| IgG | IgG | rabbit | Sigma-Aldrich I5006 |
| IgG | IgG rabbit X | rabbit | Santa Cruz Biotechnology |
| LHX2 | LHX2 | mouse | Santa Cruz Biotechnology 81311 |
| PDX1 | PDX-1 | mouse | Santa Cruz Biotechnology 390792 |
| PDX1 | PDX1(A-17) | goat | Santa Cruz Biotechnology 14664 |
| STAT1 | Stat1 | rabbit | Cell Signaling 9172S |
| STAT1 | Stat1 (D4Y6Z) | rabbit | Sigma-Aldrich 14995S |

**Table A6:** Primer

| Method | Name | Sequence | UPL Probe |
|---|---|---|---|
| Competition | CEBPA-C F | AGTTGAGGTATTGTGCAATGAAGGC | - |
| Competition | CEBPA-C R | GCCTTCATTGCACAATACCTCAACT | - |
| Competition | CEBPG-C F | AGTTGAATTGCGTAATAGGC | - |
| Competition | CEBPG-C R | GCCTATTACGCAATTCAACT | - |
| Competition | PAX6-C1 F | AGTTGATTCACGCATGAGTGCAGAGGC | - |
| Competition | PAX6-C1 R | GCCTCTGCACTCATGCGTGAATCAACT | - |
| Competition | PAX6-C2 F | AGTTGATGATTAATTAATTCGAGGC | - |
| Competition | PAX6-C2 R | GCCTCGAATTAATTAATCATCAACT | - |
| Competition | PDX1-C1 F | AGTTGATTAAGAATCTAATGACCCAAGGC | - |
| Competition | PDX1-C1 R | GCCTTGGGTCATTAGATTCTTAATCAACT | - |
| Competition | PDX1-C2 F | AGTTGAGGTAATTAGCTCAGGC | - |
| Competition | PDX1-C2 R | GCCTGAGCTAATTACCTCAACT | - |
| EMSA | rs7903146C45bpF Cy5 | Cy5-AGAGCTAAGCACTTTTTAGATAC TATATAATTTAATTGCCGTATG | - |
| EMSA | rs7903146C45bp R | CATACGGCAATTAAATTATATAGTAT CTAAAAAGTGCTTAGCTCT | - |
| EMSA | rs7903146T45bpF Cy5 | Cy5-AGAGCTAAGCACTTTTTAGATAT TATATAATTTAATTGCCGTATG | - |

Table A6 – *Continued from previous page*

| **Method** | **Name** | **Sequence** | **UPL Probe** |
|---|---|---|---|
| EMSA | rs7903146T45bp R | CATACGGCAATTAAATTATATAATATC TAAAAAGTGCTTAGCTCT | - |
| AC | rs7903146C45bpF Biotin | Biotin-AGAGCTAAGCACTTTTTAGATAC TATATAATTTAATTGCCGTATG | - |
| AC | rs7903146T45bpF Biotin | Biotin-AGAGCTAAGCACTTTTTAGATAT TATATAATTTAATTGCCGTATG | - |
| Sequencing | CMV-Seq-F | CGCAAATGGGCGGTAGGCGTG | - |
| Sequencing | M13for | GTCATAGCTGTTTCCTG | - |
| Sequencing | M13rev | CAGGAAACAGCTATGAC | - |
| Sequencing | pcDNA3.1-DYKR | GGGCAAACAACAGATGGCTGGC | - |
| Sequencing | SV40-Prom-SeqF | TATTTATGCAGAGGCCGAGG | - |
| Sequencing | T7F | TAATACGACTCACTATAGGG | - |
| Sequencing | T7R | CCCTATAGTGAGTCGTATTA | - |
| Sequencing | TCF790seq1 Fa | GCATTGTAATCCAGTGACACC | - |
| Sequencing | TCF790E5XhoI Ra | GGCCCAGAAAGGCAAAGTGACAGATCC TCGAGTTAACC | - |
| Sequencing | TCF790-Seq1 F | GATGTGATGAGATCTCTGCC | - |
| SYBR Assay | ACSL5 F2 | AGGCATGATAGTTTCTGGGAC | - |
| SYBR Assay | ACSL5 R3 | ACCTGACATCCCATTGCTG | - |
| SYBR Assay | ChIP TAP1 F | GTCTGTGTGATGAGTTGGTCC | - |
| SYBR Assay | ChIP TAP1 R | GGGATGGTGCAAAGAGATGAG | - |
| SYBR Assay | CHiPTCF7L2Fb | CTCAAAACCTAGCACAGCTGTTATTT ACTGA | - |
| SYBR Assay | ChIPrs7903146CR7 | GGGTGCCTCATACGGCAATTAAATTAT ACAG | - |
| SYBR Assay | ChIPrs7903146TR7 | GGGTGCCTCATACGGCAATTAAATTAT ACAA | - |
| SYBR Assay | CHiPTCF7L2Fa | CCTAGCACAGC TGTTATTTACTGA | - |
| SYBR Assay | CHiPTCF7L2Ra | ACTAAGGGTGCCTCATACGG | - |
| SYBR Assay | TCF7L2preampF | gggAgCCgTCAgATggTAATg | - |
| SYBR Assay | TCF7L2preampR | gTCCAgggCCCCTCTAACC | - |
| SYBR Assay | GAPDH F | GATCATCAGCAATGCCTCCTGC | - |
| SYBR Assay | GAPDH R | ACAGTCTTCTGGGTGGCAGTGA | - |
| SYBR Assay | h18S F | GATATGCTCATGTGGTGTTG | - |
| SYBR Assay | h18S R | AATCTTCTTCAGTCGCTCCA | - |
| SYBR Assay | hActin-1F | gCgCCCCAggCACCAgggCg | - |
| SYBR Assay | hActin-1R | AggTCTCAAACATgATCTgg | - |
| SYBR Assay | TCF7L2 F | CGTAGACCCCAAAACAGGAA | - |
| SYBR Assay | TCF7L2 R | ATCCTCCTGTCGTGATTGGG | - |
| SYBR Assay | R acsl5 F2 | AAACCAATGCCTCCGAACCC | - |
| SYBR Assay | R acsl5 R2 | CCTGGACAAGCCTCTCAAACA | - |
| SYBR Assay | R actb F1 | CGCGAGTACAACCTTCTTGC | - |
| SYBR Assay | R actb R1 | CGTCATCCATGGCGAACTGG | - |
| SYBR Assay | R gapdh F1 | AGTGCCAGCCTCGTCTCATA | - |
| SYBR Assay | R gapdh R1 | GATGGTGATGGGTTTCCCGT | - |
| SYBR Assay | R pdx1 F1 | CGGACCTTTCCCGAATGGAA | - |
| SYBR Assay | R pdx1 R1 | TATGCACCTCCTGCCCACT | - |
| SYBR Assay | R tcf7l2 F3 | TAGGCGCTAACGACGAACTG | - |
| SYBR Assay | R tcf7l2 R3 | TTGGCCGCTTCTTCCAAACT | - |
| TaqMan Assay | ACSL5 S | gAAgggTTCgTgTAATTgTCAC | - |

Table A6 – *Continued from previous page*

| Method | Name | Sequence | UPL Probe |
|---|---|---|---|
| TaqMan Assay | ACSL5 R | CTgTgCATTCTgTTTgACCATAAg | - |
| TaqMan Assay | ACSL5 TM | CCATgTCCACTTCAgTCATgACATTCTTCC | - |
| TaqMan Assay | ACTIN as | CggAACCgCTCATTgCC | - |
| TaqMan Assay | ACTIN se | ACCCACACTgTgCCCATCTA | - |
| TaqMan Assay | ACTIN TM | ATgCCCTCCCCCATgCCATCCTgCgTC | - |
| TaqMan Assay | CEBPG F | CACCCTTTCCAgACTTTAgAgC | - |
| TaqMan Assay | CEBPG A | TTTgAgTCATggAAATggACAAC | - |
| TaqMan Assay | CEBPG TM | CCACCgACACCACTCATgTCAATgg | - |
| TaqMan Assay | MED19 S | CCTCATgAgggAACTgCCA | - |
| TaqMan Assay | MED19 A | CACAgAATTTATTATAggCTTgTTCCA | - |
| TaqMan Assay | MED19 TM | CTgACAggCAgCACgAATCTgATCAC | - |
| TaqMan Assay | PPIAex4 F | CACTggAgAgAAAggATTTgg | - |
| TaqMan Assay | PPIAex4-5 R | gTgTgAAgTCACCACCCTgA | - |
| TaqMan Assay | PPIAex4 TM | AAACCCTggAATAATTCTgTgAAAgCAgg | - |
| TaqMan Assay | STAT1 S | CCAgAATgCCCTgATTAATgATg | - |
| TaqMan Assay | STAT1 A | gTgTATTTCTgTTCCAATTCCTCCA | |
| TaqMan Assay | STAT1 TM | CTTTTTAAgCTgCTgCCgAACTTgCT | - |
| TaqMan Assay | TAF10 S | CCACgCATAATTCggCTC | - |
| TaqMan Assay | TAF10 A | TgTAgggCATCATTggCAATA | - |
| TaqMan Assay | TAF10 TM | CTCCTTAgCTgCCCAgAAATTCATCTCA | - |
| TaqMan Assay | TCF7L2all S | CCgACgTAgACCCCAAAAC | - |
| TaqMan Assay | TCF7L2all A | CggTgCCAggCgATAg | - |
| TaqMan Assay | TCF7L2all TM | CACCCTCCAgATATATCCCCgTATTACCC | - |
| TaqMan Assay | YWHAZ S | AgAgTCATACAAAgACAgCACgCTAA | - |
| TaqMan Assay | YWHAZ A | TCTCCTgCTTCAgCTTCgTCTC | - |
| TaqMan Assay | YWHAZ TM | TTgggTATCCgATgTCCACAATgTCAA | - |
| TaqMan Assay | ZNF593 S | AAgAAAAggCTgAAgCAgCT | - |
| TaqMan Assay | ZNF593 A | ATACCCgCTgCCCTCTC | - |
| TaqMan Assay | ZNF593 TM | CgTCgAgCCCTACAgTCAggAAgAg | - |
| TaqMan Assay | HMGA2 S | CCTCTCCTAAgAgACCCAgg | - |
| TaqMan Assay | HMGA2 A | AgTggCTTCTgCTTTCTTTTgA | - |
| TaqMan Assay | HMGA2 TM | CCCAAAggCAgCAAAAACAAgAgTC | - |
| UPL Probe Assay | ATF1 F | gtggggaagtgggtagtgaa | #5 |
| UPL Probe Assay | ATF1 R | ttgtgggaatcttccataatca | |
| UPL Probe Assay | BBX F | gcgagctaatgttgacagagg | #82 |
| UPL Probe Assay | BBX R | tcttcattccaacacccttcat | |
| UPL Probe Assay | CEBPG F | ctcctcctcgctgactcg | #68 |
| UPL Probe Assay | CEBPG R | cacgctaagctgccaaaaa | |
| UPL Probe Assay | PDX1 F | aagctcacgcgtggaaag | #78 |
| UPL Probe Assay | PDX1 R | gccgtgagatgtacttgttgaa | |
| UPL Probe Assay | PURB F | agctttatgagcgacgtggt | #70 |
| UPL Probe Assay | PURB R | cctcctcaccctctgactctt | |
| UPL Probe Assay | SCRT1 F | aacttgacgcgttctcttcg | #80 |
| UPL Probe Assay | SCRT1 R | cccacgtagtcgctgagg | |
| UPL Probe Assay | STAT1 F | tgagttgatttctgtgtctgaagtt | #32 |
| UPL Probe Assay | STAT1 R | acacctcgtcaaactcctcag | |
| UPL Probe Assay | TAF6 F | acgacccctcttgtaaacc | #52 |

Table A6 – *Continued from previous page*

| **Method** | **Name** | **Sequence** | **UPL Probe** |
|---|---|---|---|
| UPL Probe Assay | TAF6 R | cttctcctcagccattctgg | |
| UPL Probe Assay | UPL1CEBPBF1 | ctggagacgcagcacaag | #1 |
| UPL Probe Assay | UPL1CEBPBR1 | acagctgctccaccttcttc | |
| UPL Probe Assay | UPL1PRDM16F1 | cggatgttccccaacaagta | #1 |
| UPL Probe Assay | UPL1PRDM16R1 | catttgtactcgcgctcct | |
| UPL Probe Assay | UPL1STAT3F1 | cctctgccggagaaacagt | #1 |
| UPL Probe Assay | UPL1STAT3R1 | cattgggaagctgtcactgtag | |
| UPL Probe Assay | UPL12SOCS1F1 | gagaacctggctcgcatc | #12 |
| UPL Probe Assay | UPL12SOCS1R1 | ctgccggtcaaatctgga | |
| UPL Probe Assay | UPL2CEBPAF1 | agttcctggccgacctgt | #2 |
| UPL Probe Assay | UPL2CEBPAR1 | cccgggtagtcaaagtcg | |
| UPL Probe Assay | UPL42STAT2F1 | aactggacaaaaggagaaagga | #42 |
| UPL Probe Assay | UPL42STAT2R1 | tggcagcagtagctcgatta | |
| UPL Probe Assay | UPL69CEBPZF1 | ataatactagtgaagccgagaatgg | #69 |
| UPL Probe Assay | UPL69CEBPZR1 | gccagcataaggtaatcttgct | |
| UPL Probe Assay | UPL69DDB2F1 | ggcatcagttcgcttaatgaat | #69 |
| UPL Probe Assay | UPL69DDB2R1 | tgtctctcacttccgtgtcct | |
| UPL Probe Assay | UPL84IRF1F1 | gccttcttccctcttccact | #84 |
| UPL Probe Assay | UPL84IRF1R1 | cgagtgatgggcatgttg | |
| UPL Probe Assay | ZNF593 F | agatagggctctcacctggtc | #41 |
| UPL Probe Assay | ZNF593 R | ccgcctcttcctgactgtag | |

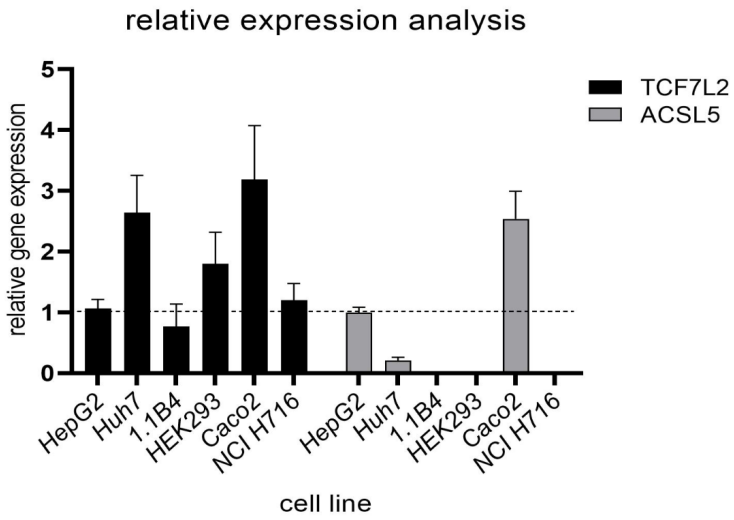


**Figure A1:** RNA from cells of HepG2, Huh7, 1.1B4, HEK293, Caco2 and NCI H716 was translated into cDNA and gene expression was examined. Housekeepers *ACTB* and *GAPDH* were used for normalization. Values are displayed relative to HepG2 gene expression as mean + SEM, n = 3-6.

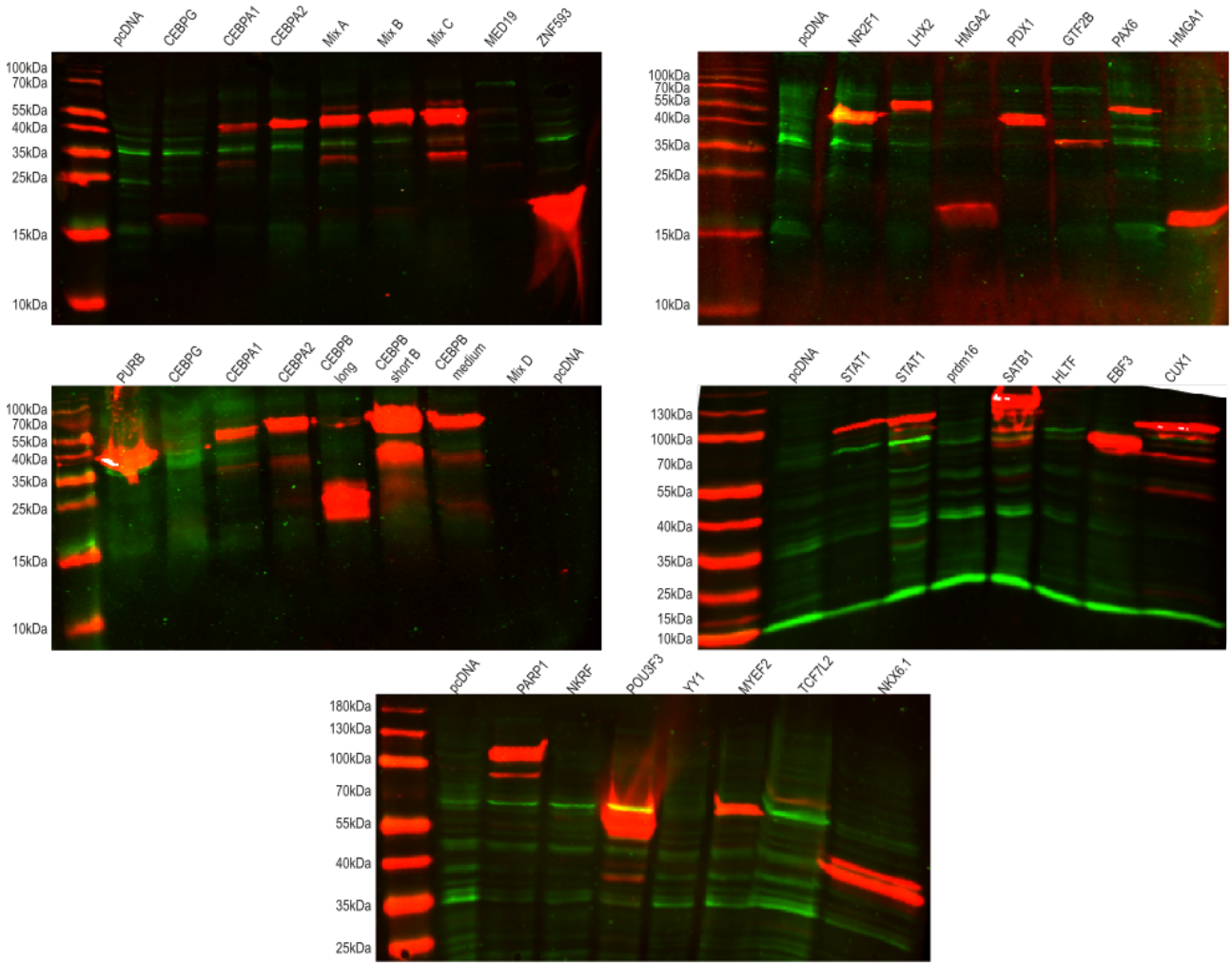


**Figure A2:** Western blot of 35 µg HEK293T nuclear cell extracts overexpressing transcription factors of interest. All overexpressed proteins contained a 5' DYK FLAG-Tag. Red bands represent the FLAG-Tag, while green represent the Housekeeper Lamin B1. PcDNA (pcDNA3.1 DYK) is a vector only containing the FLAG-Tag, serving as a negative control. 1st antibodies, anti-M2 FLAG and anti-Lamin B1 were incubated 1:1000 overnight at 4 °C.

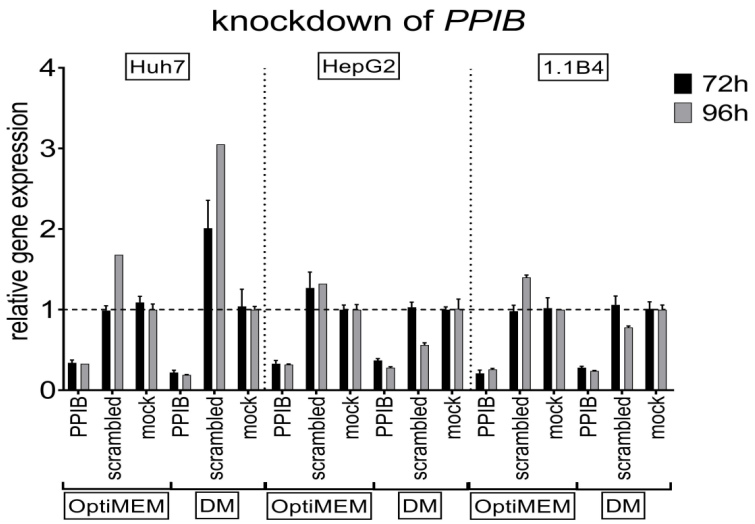


**Figure A3:** *PPIB* knockdown was tested in Huh7, HepG2 and 1.1B4 cells in different medium (OptiMEM or DM) and at different time points (72 h or 96 h).Values are displayed relative to the mock (untreated) gene expression as mean + SEM, n = 1-3.

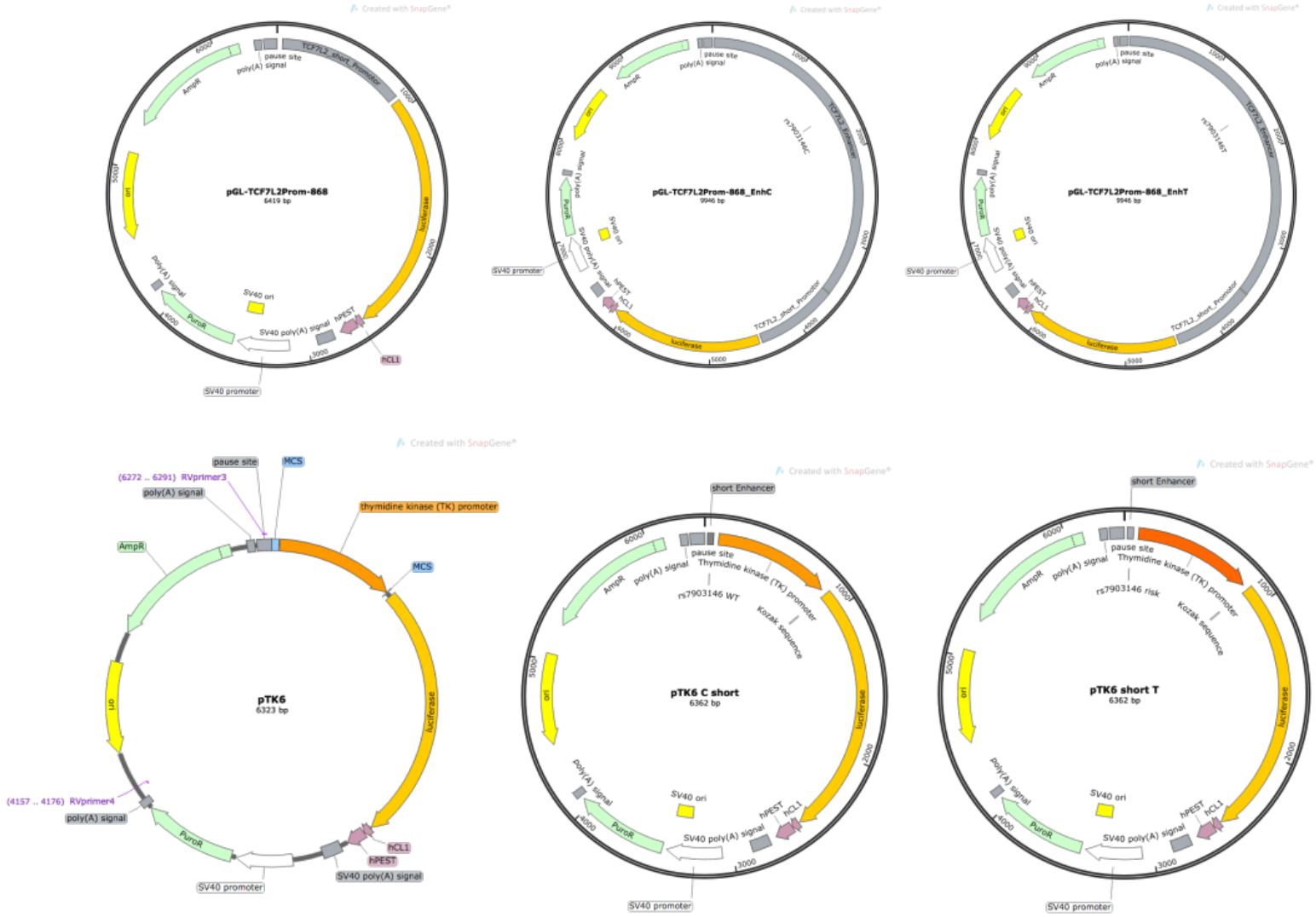


**Figure A4:** Luciferase vector constructs used in this thesis. pGL-TCF7L2Prom-868 = pGL4.22 vector with a 868 bp *TCF7L2* Promotor fragment upstream of the *Luciferase* ORF. pGL-TCF7L2-868 EnhC or T = pGL4.22 vector with a 3500 bp *TCF7L2* Enhancer fragment containing either WT or SNP and a 868 bp *TCF7L2* Promotor fragment upstream of the *Luciferase* ORF. pTK6 = pGL4.22 vector with a *Thymidine kinase* promotor upstream of the *Luciferase* ORF. pTK6 C/T short = pGL4.22 vector containing a 40 bp *TCF7L2* Enhancer fragment containing either WT or SNP and a *Thymidine kinase* promotor upstream of the *Luciferase* ORF.

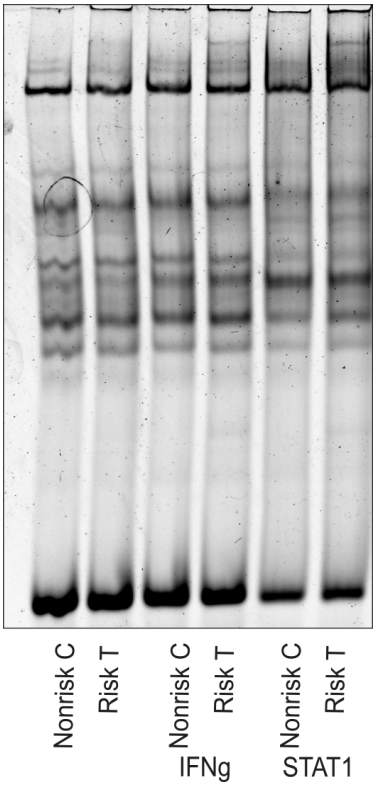


**Figure A5:** EMSA with HEK293T nuclear extract overexpressing *PTPN11*. IFNγ = PTPN11 overexpressing nuclear extract mixed 1:2 with HepG2 treated with IFNγ for 24 h. STAT1 = PTPN11 overexpressing nuclear extract mixed 1:2 with HEK293T cells overexpressing STAT1. Total nuclear extract per reaction was 1.5 µg in lane 1-2 and 4.5 µg in the mix conditions. Each reaction contained additionally 700 ng Poly(dIdC) and 1.5 ng of a 45 bp Cy5- labeled oligonucleotide.

Figure A

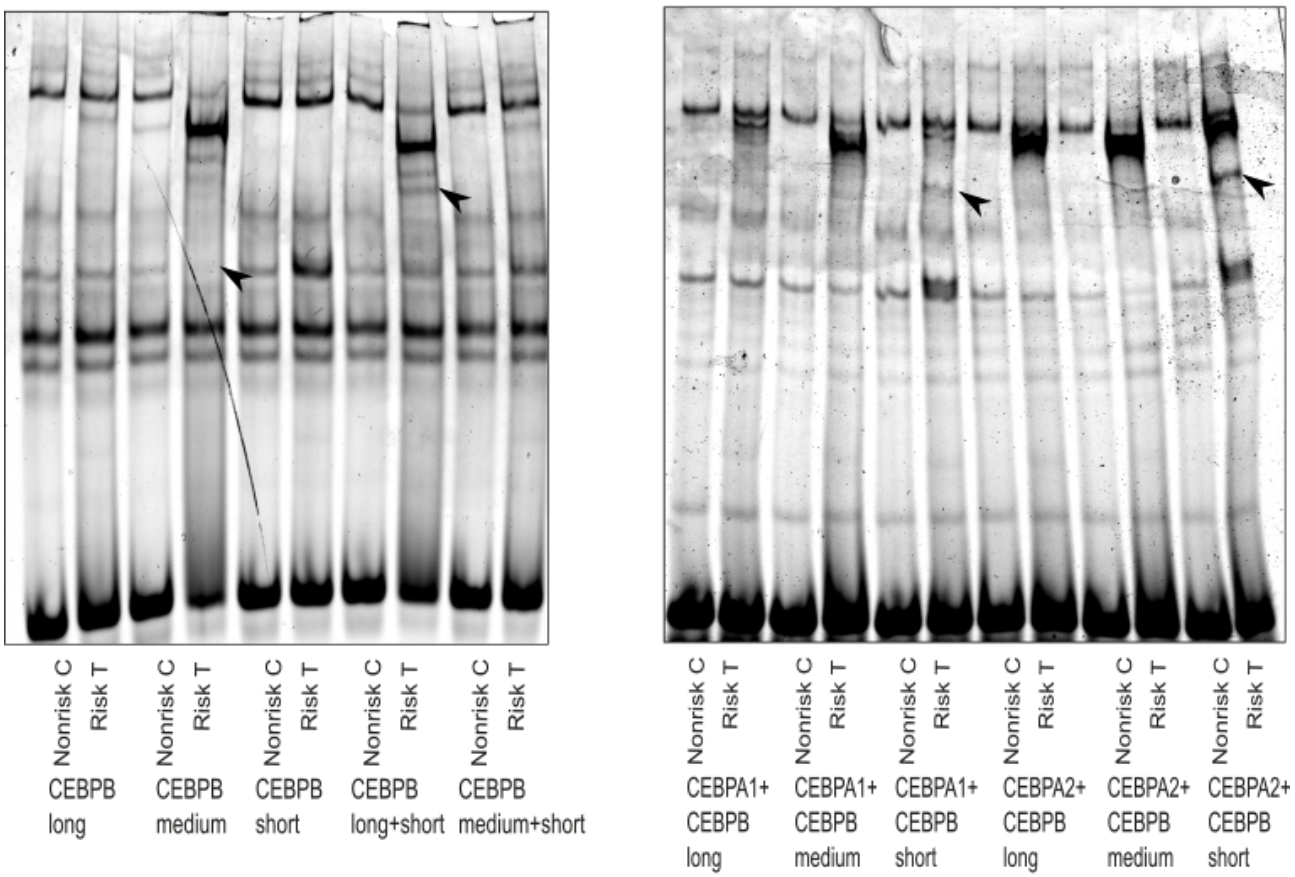


Figure B

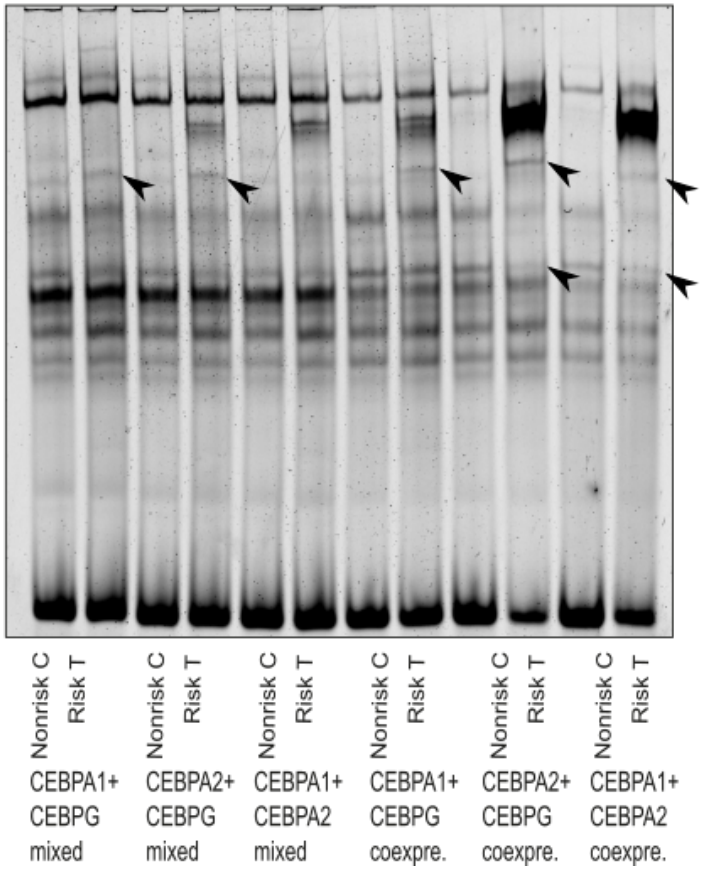


**Figure A6:** EMSA with HEK293T nuclear extract overexpressing different *CEBPs*. Total nuclear extract per reaction was 3 µg, with 700 ng Poly(dIdC) and 1.5 ng of a 45 bp Cy5- labeled oligonucleotide.**A**: Nuclear extracts of CEBPA1, CEBPA2 and the three CEBPB isoforms in different combinations mixed.**B**: Comparison of nuclear extracts of two different factors mixed together versus them coexpressed simultaneously in HEK293T.