Study of the transcriptional regulation of Mlxipl by RFX6 and identification of target genes in pancreatic beta cells

Julia Grans

UNIVERSITÉ DE STRASBOURG

ÉCOLE DOCTORALE 414 – SCIENCES DE LA VIE ET DE LA SANTÉ

Institut de Génétique et de Biologie Moléculaire et Cellulaire

(CNRS UMR 7104 INSERM U964)

THÈSE présentée par :

Julia GRANS

soutenue le : **5 avril 2019**

pour obtenir le grade de : **Docteur de l'Université de Strasbourg**

discipline – spécialité : **Science de la vie et de la santé – Aspects moléculaires et cellulaires de la biologie**

Étude de la régulation transcriptionnelle de *Mlxipl* par RFX6 et identification des gènes cibles dans les cellules bêta pancréatiques

THÈSE dirigée par :

M. GRADWOHL Gérard	Directeur de recherches, Université de Strasbourg

RAPPORTEURS :

M. HEIMBERG Harry	Professeur des Universités, Université de Bruxelles
M. RAVASSARD Philippe	Directeur de recherches, Université Pierre et Marie Curie

---

AUTRES MEMBRES DU JURY :

Mme RHINN Muriel	Chargé de recherches, Université de Strasbourg
Mme GUILMEAU Sandra	Maître de conférences, Université de Paris V

UNIVERSITY OF STRASBOURG

GRADUATE SCHOOL 414 – LIFE AND HEALTH SCIENCES

Institute of Genetics and of Molecular and Cellular Biology

(CNRS UMR 7104 INSERM U964)

THESIS presented by:

Julia GRANS

defended on the: **5th of April 2019**

to be awarded a: **PhD of the University of Strasbourg**

discipline – speciality: **Life and health sciences – Molecular and cellular aspects of biology**

Study of the transcriptional regulation of *Mlxipl* by RFX6 and identification of target genes in pancreatic beta cells

THESIS directed by:

Mr GRADWOHL Gérard Research Director, University of Strasbourg

ASSESSMENT COMMITTEE MEMBERS:

Mr HEIMBERG Harry University Professor, University of Brussels

Mr RAVASSARD Philippe Research Director, University Pierre et Marie Curie

---

ADDITIONAL MEMBERS:

Mrs RHINN Muriel Researcher, University of Strasbourg

Mrs GUILMEAU Sandra Associate professor, University Paris V

« Vous éprouvez le désir d’aller voir ailleurs,

estimant avoir fait le tour de la question.

De quelle question ? »

Preface

The experimental work depicted in this thesis has been carried out in the research group "Differentiation and Physiopathology of Endocrine Cells in the Pancreas and Intestine" headed by Gérard Gradwohl. His group is part of the "Development and Stem Cells" department of the "Institute of Genetics and of Molecular and Cellular Biology" (IGBMC) that belongs to the University of Strasbourg. The thesis was funded by the *Initiative d'Excellence* (IdEx) program of the University of Strasbourg and the NovoNordisk foundation. The experiments described in the result's part were performed with the help of Céline Ziegler-Birling and Céline Lapp and the bioinformatical analysis of the RNA sequencing data done by Constance Vagne.

The obtained results will partially be published in the following manuscript: "Genome-wide binding of the transcription factor RFX6 and identification of regulated genes revealed *Mlxipl* as a novel direct target of RFX6." (article in preparation) by Perrine Strasser, Julia Grans, Tao Ye, Vikash Chandra, Julie Piccand, Aline Meunier, Martine Vaxillaire, Raphaël Scharfmann and Gérard Gradwohl (this is probably not the final list of the authors).

Acknowledgments

Gérard, merci de m'avoir donné la possibilité de faire une thèse dans ton labo et de m'avoir guidée jusqu'à la fin de ma thèse. Elle m'a beaucoup appris. Pendant ces trois ans et les quelques mois avant et après, tu m'as donné beaucoup de conseils, et j'aimerais profiter de ces remerciements pour me permettre de t'en donner un également : essaye de faire plus souvent des compliments (à Florence, dans le futur, par exemple), parce que c'est motivant et que ça donne envie de continuer. Et à part « pas mal », essaie peut-être « bien » ou « très bien » pour changer un peu. Pour utiliser ton compliment préféré : Tu es « pas mal » comme chef.

Perrine et **Julie**, vous avez initié ce projet sur RFX6 et MLXIPL que j'ai continué pendant ma thèse. Nous avons réussi à trouver plein de nouvelles questions qui peuvent rendre fous les futurs doctorant(e)s.

Merci à toi, **Constance**, d'avoir analysé mes données de séquençage et d'avoir supplementé mon travail avec toutes les données de séquençage de l'ARN et séquençage ChIP.

Je remercie **Céline** et **Céline**, ou, plus précisément, Céline Ziegler-Birling et Céline Lapp, de m'avoir prêté vos mains pour le travail au labo, la culture cellulaire, les clonages, … Merci de n'avoir jamais perdu le moral quand rien ne fonctionnait.

Marie, du warst meine erste Praktikantin und ich hätte mir keine sympathischere wünschen können! Es ist schön zu wissen, dass du deinen Weg gefunden hast, und demnächst die Welt erobern wirst. **Haley**, it was a nice three months you were my intern, even though nothing really worked out of what we should have done together. One day I would like to travel to Canada to find out if you were right, that I would fit in well.

Merci à **Sandra Guilmeau** et **Alain Lescure** d'avoir discuté de mon sujet avec moi pendant mes deux mi-thèses, et **Catherine Postic**, **Sandra Guilmeau**, **Raphaël Scharfmann**, **Paul Richards**, **Roméo Ricci, Kevin Vivot, Adèle De Arcangelis** et **Valérie Schreiber** pour les discussions sur les cellules bêta à Paris et à Illkirch.

Je tiens à remercier la plateforme **GenomEast** et surtout **Serge Vicaire** et **Céline Keime**, les filles du **service de culture**, et **Muriel Philipps** et **Claudine Ebel** du **service de cytométrie en flux** pour le travail effectué.

J'aimerais également remercier les membres de mon jury de thèse pour avoir accepté de lire mon manuscrit et de juger mon travail : **Harry Heimberg**, **Philippe Ravassard**, **Muriel Rhinn** et **Sandra**

Guilmeau. J'espère que nous aurons une belle discussion le jour de ma soutenance et que je n'aurai pas oublié tous les détails d'ici là...

Colette, danke, dass du die Höhen und Tiefen der Doktorarbeit mit mir durchgestanden hast. Schön, dass ich mit dir eine zweite Doktorandin in meiner Arbeitsgruppe hatte, die meine Einstellungen und Ideen teilt, nicht auszudenken, wie es mit jemandem gewesen wäre, der anders ist als du.

Merci à **Colette** et à **Céline**, de nouveau, ainsi qu'à **Florence**, pour toutes les pauses passées ensemble. C'est très pratique que les résultats coulent de la machine à café, sinon on n'aurait rien produit pendant ces trois ans … Vous allez vraiment me manquer. Je m'arrête là, sinon je commence à pleurer.

Et, bien sûr, un grand merci aux autres membres de l'équipe (j'espère que je n'ai oublié personne) : **Sabitri**, **Adèle**, **Valérie**, **Aline**, **Sara**, **Annabelle**, **Laure**, **Katia**, **Claire**, **Caroline**, **François** et **Jean-Marie**.

Um nun noch etwas persönlicher zu werden, bevor es mit dem wissenschaftlichen Teil anfängt, möchte ich mich an dieser Stelle erst einmal bei all meinen **Freunden** bedanken, insbesondere bei Lisa und Aniela. **Lisa**, danke, dass du mich moralisch die drei Jahre über unterstützt hast. Ich freue mich schon sehr auf unsere Reise nach England und Schottland, das gibt mir einen guten Grund, mich bei der Verteidigung anzustrengen. **Aniela**, ich hoffe, dass wir beide trotz des Biologiestudiums bald wissen werden, was wir mit diesem Leben anfangen sollen, denn im Moment hätte ich keine große Lust, meinem jüngeren Ich gegenüber zu treten und ihm zu erklären, was das Ganze sollte.

Ein großer Dank gilt natürlich auch meiner **Familie**, der ich in der ganzen Doktorarbeitszeit nicht erklären konnte, woran ich arbeite – falls ihr mal nicht schlafen könnt, könnt ihr ja die Arbeit lesen – und warum ich bei den Berufsaussichten überhaupt die Entscheidung getroffen habe einen Doktor zu machen. Das bleibt eine gute Frage. Ich bin jedenfalls gespannt, welcher Ort bald in der Frage meiner Nichte „Warum musst du immer in das blöde Frankreich zurück?" auftauchen wird.

Ach so, und bevor ich es vergesse: ein Augenzwinkern gilt meinen **Vierbeinern**, die bellen, maunzen und wiehern, immer da sind, und sich, während ich über zu komplizierte Formulierungen nachgegrübelt habe, selbstbewusst in den Vordergrund gedrängelt haben, auf ihre eigene charmante Art und Weise.

Je vous souhaite une bonne lecture, et ne vous endormez pas …

Table of contents

A) Introduction

The experimental work depicted in this thesis manuscript has been carried out in the laboratory of Gérard Gradwohl who heads a group at the Institute of Genetics and of Molecular and Cellular Biology (IGBMC) at the University of Strasbourg. As a part of the development and stem cells department of the institute, his team work on the differentiation and function of endocrine cells in the pancreas and enteroendocrine cells in the gut with a special focus on their transcriptomes.

My PhD thesis project was about the function of the winged-helix transcription factor RFX6 and the basic helix-loop-helix transcription factor MLXIPL in adult beta cells. Before the start of my project in 2015, the team has revealed that RFX6 is an essential regulator of beta cell identity and function in adult mice (Piccand *et al.,* 2014). A screen for genes directly regulated by RFX6 in pancreatic beta cells revealed that the gene *Mlxipl* is the most downregulated gene in the *Rfx6* knockout (Piccand *et al.,* 2014; unpublished data of Perrine Strasser). MLXIPL is a metabolite-activated transcription factor that regulates glycolytic and lipogenic genes in the liver and in adipose tissues (Postic *et al.,* 2007, review) while its function in the pancreas still needs to be fully elucidated (Filhoulaud *et al.,* 2013, review; Iizuka *et al.,* 2013, review; Richards *et al.,* 2017, review). The work of this thesis shall confirm and characterize the regulation of *Mlxipl* expression by RFX6 and help to understand their respective roles in adult beta cells.

1. The pancreas is both an exocrine and endocrine gland and part of the gastrointestinal system

The pancreas is a glandular organ in the abdomen behind the stomach and part of the gastrointestinal system. It comprises an exocrine part releasing digestive enzymes into the intestine, and an endocrine part secreting hormones into the bloodstream, thus on the one hand, it is important for the digestion and, on the other hand, for the hormonal regulation of glucose metabolism and energy homoeostasis (**FIGURE A1**).

The exocrine part is constructed like a tree (**FIGURE A1A**): it contains a branched duct system, whose ducts are formed by epithelial duct cells. The acini are located like "leaves" at the tips of the ducts"; they are bladder-like structures formed by apical acinar cells. The acinar cells (**FIGURE A1B**) secrete various proenzymes and bicarbonate into the lumen of the acini. The pancreatic juice flows through the ducts into the pancreatic duct, which opens into the duodenum. In the duodenum, the proenzymes get activated, and they enzymatically cleave dietary carbohydrates, proteins, lipids and nucleic acids.

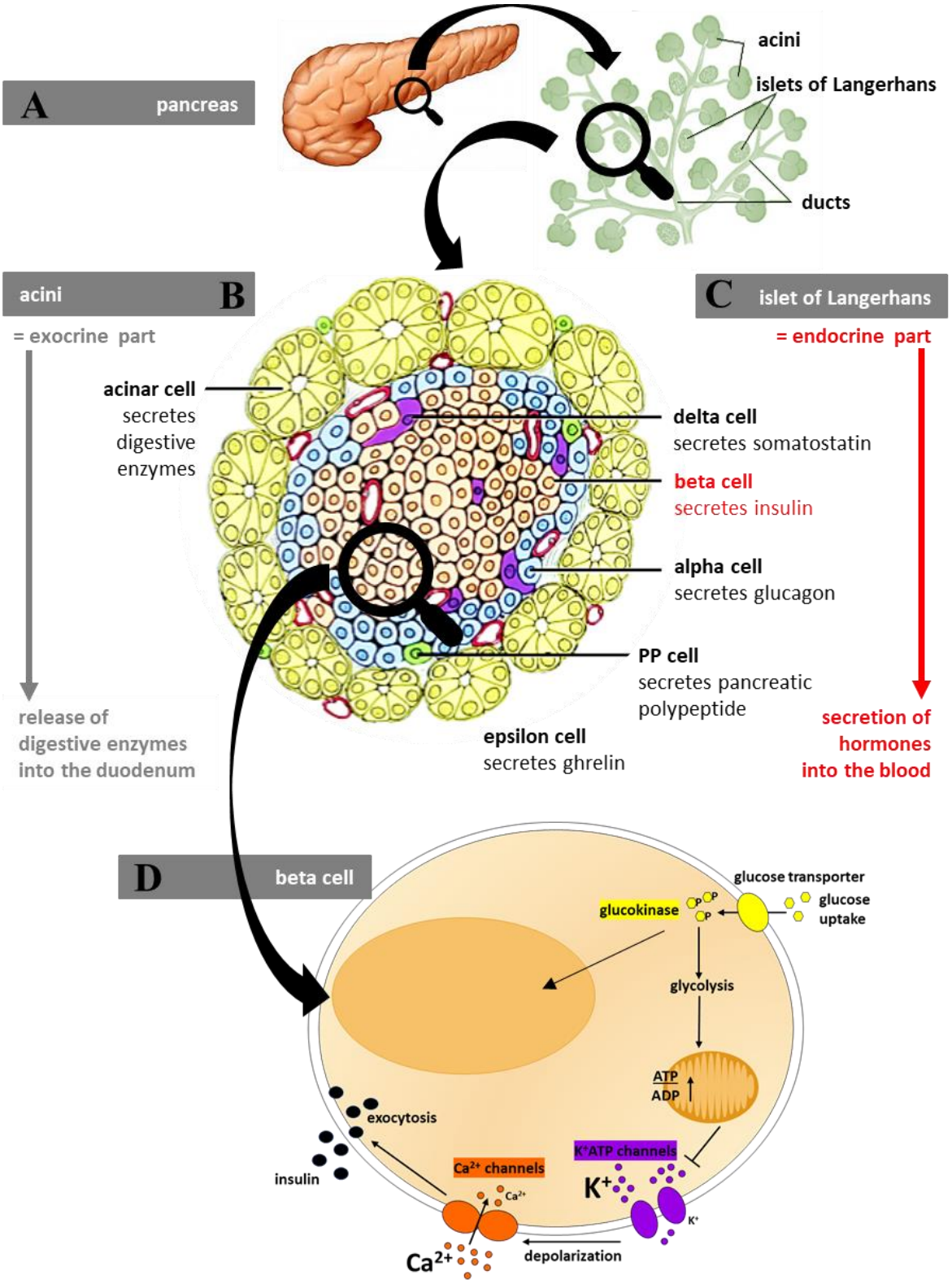


FIGURE A1: Beta cells are endocrine cells of the pancreas and secrete insulin in response to glucose. **(A)** Macroscopic structure of the exocrine part (acini and ducts) and the endocrine part (islets of Langerhans) of the pancreas (figures adapted from the University of Leeds / Cano *et al.*, 2013, review). **(B, C)** Microscopic structure of a rodent islet of Langerhans **(B)** and the surrounding acini **(C)** (figure adapted from Campbell University). **(D)** Schematic illustrating insulin secretion by pancreatic beta cells (figure adapted from Piccand *et al.*, 2014).

The islets of Langerhans (**FIGURE A1C**), which consist of endocrine cells, sit like fruits between the ducts and acini. The human pancreas contains up to one million of islets, but they account for only 1-2 % of the pancreatic mass; however, 10% of the blood flow through the pancreas supplies the islet cells (Rorsman and Braun, 2013, review). Endocrine cell types are insulin-producing beta cells (50 %), glucagon-producing alpha cells (35-40 %), somatostatin-producing delta cells (5 %), pancreatic polypeptide-producing PP cells and ghrelin-producing epsilon cells (percentages of endocrine cells found in human pancreatic islets, Rorsman and Braun, 2013, review). In rodents, the islets consist of a core of beta cells surrounded by a ring of alpha cells (**FIGURE A1C**). By contrast, in humans, the different endocrine cell types are distributed evenly throughout the islets (Rorsman and Braun, 2013, review).

Pancreatic endocrine function is critical for glucose homeostasis. Key regulators are the hormones glucagon and insulin. Glucagon stimulates the release of glucose by the liver, thus increasing the blood glucose level in the case of hypoglycaemia. In contrast, insulin promotes the uptake and metabolism of glucose by liver, fat and skeletal muscle cells in hyperglycaemia (Quesada *et al.,* 2008, review). Paracrine signalling via somatostatin, which is produced by neighbouring delta cells, inhibits the secretion of glucagon and insulin (Hauge-Evans *et al.,* 2009; Rorsman and Huising, 2018, review).

The hormones pancreatic polypeptide and ghrelin that are produced by PP cells and epsilon cells, respectively, control the food intake. Pancreatic polypeptide is released by PP cells after food intake and it reduces appetite and food ingestion (Batterham *et al.,* 2003). Ghrelin, which is secreted by epsilon cells, is the hunger hormone and the antagonist to the satiety hormone leptin. Leptin communicates to the brain that enough energy is available, and no food needs to be ingested while ghrelin is increasingly produced during starvation and stimulates food intake (Klok *et al.,* 2007, review).

1.1. The use of isolated islets and beta cell lines to study beta cell (patho)physiology

Possibilities to study the physiology and function of beta cells are isolated islets or immortalized beta cell lines (**FIGURE A2**). Rodent islets are usually isolated from mice or rats (**FIGURE A2A**) using islet isolation protocols. Human islets prepared from a donated human pancreas (**FIGURE A2A**) can be obtained via distribution programmes but the access to human islets is difficult since they are preferentially used for transplantation and the demand of islets for research is increasing (Nano *et al.,* 2015). The use of whole islets has the advantage that the beta cells are in their natural environment, so they have contact with the other islet cell types, blood vessels and nerves. By the enzymatic digestion of the islets, primary beta cell cultures can be produced but they can not be maintained in culture (Skelin *et al.,* 2010, review; **FIGURE A2A**).

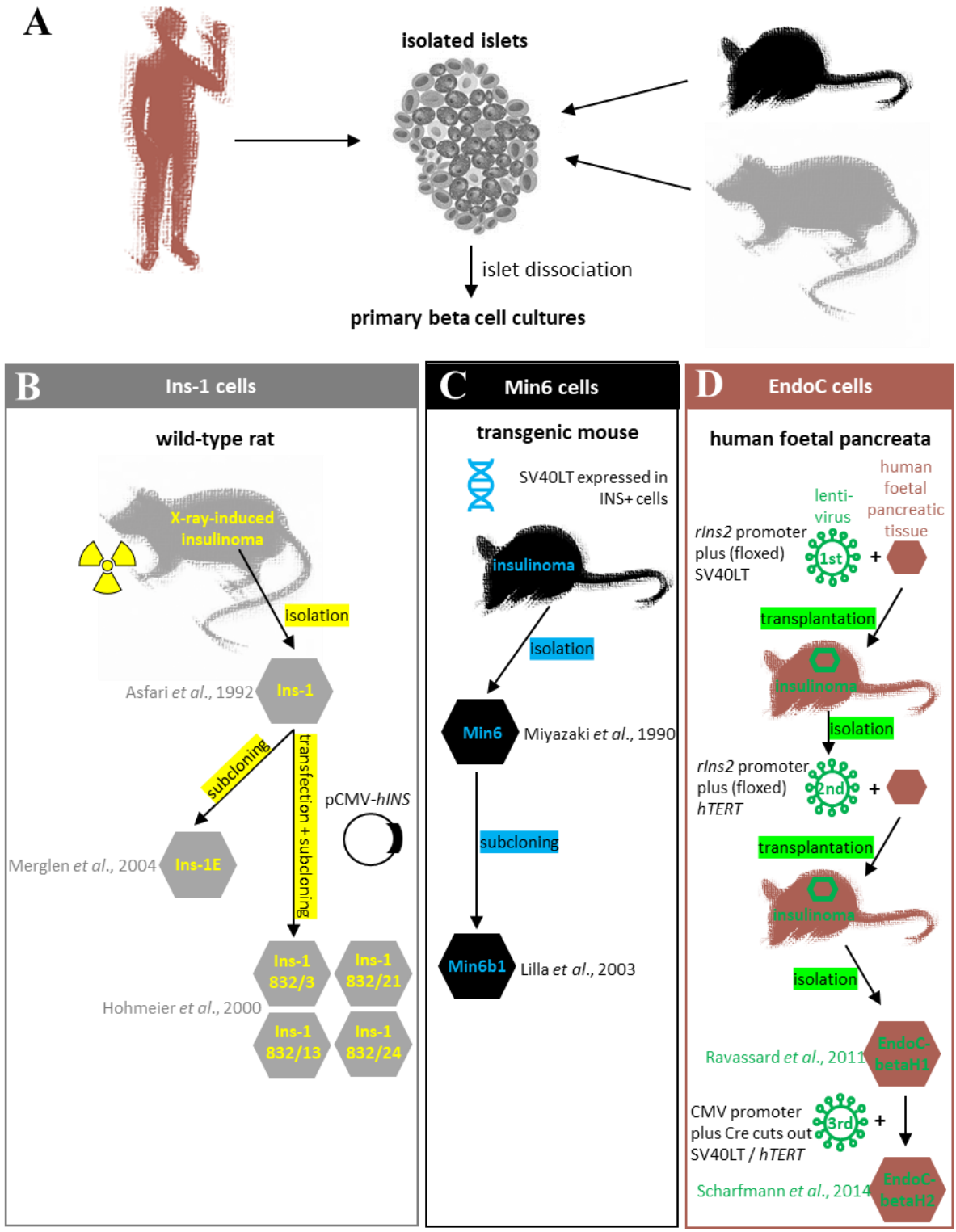


FIGURE A2: Commonly used model systems to study adult beta cell function include isolated islets and beta cell lines. (A) Isolated islets from human donors, mice and rats as a model system to study the function of beta cells (islet figure adapted from http://www.symmation.com/portfolio/media/ghrelin-cells-in-the-pancreatic-islet/99). **(A, B, C)** Experimental strategies used to generate the rat Ins-1 cells and the Ins-1E and Ins-1 832/13 subclone **(B)**, the murine Min6 cells and the Min6b1 subclone **(C)** and the human beta cell lines EndoC-betaH1 and EndoC-betaH2 **(D)**.

An alternative to islets and primary beta cell cultures is the use of stable beta cell lines. The first attempts to generate beta cell lines date back to the 1970s. Chick *et al.* transplanted rat cells isolated from an X-ray-induced insulinoma into a host rat, and they observed that the tumour cells had a normal islet cell morphology and expressed and secreted insulin (Chick *et al.,* 1977). In 1980, Gazdar *et al.* generated the RIN-r and RIN-m cell line. They used cells from an X-ray-induced insulinoma of an inbred albino rat strain called NEDH (for New England Deaconess Hospital where they have been bred initially), and they maintained the tumour by serial transplantations into rats. To produce stable cell lines, they transplanted the tumour cells into rats (RIN-r cells) and mice (RIN-m cells), re-isolated the tumours after one and four passages, respectively, and cultured the cells *in vitro*. Both cell lines could be maintained in culture and co-expressed insulin and somatostatin. The authors further showed that the cells were hypodiploid and heterogenous since subclones showed a various glucose-stimulated insulin secretion (Gazdra *et al.,* 1980).

The most commonly used beta cell lines are the rat Ins-1 cells and the murine Min6 cells (Skelin *et al.,* 2010, review; **FIGURE A2B+C**). The Ins-1 cell line was also isolated from an X-ray induced insulinoma of an NEDH rat in 1992; the authors found that beta-mercaptoethanol in the culture medium improved the proliferation and the maintenance of the differentiated state (Asfari *et al.,* 1992). Like the RIN cells, Ins-1 cells are a heterogeneous population consisting of cells that produce different amounts of insulin and that are glucose-responsive or glucose-unresponsive (Hohmeier *et al.,* 2000). In 2000, Hohmeier *et al.* stably transfected the parental Ins-1 cell line with the human insulin gene under the control of the cytomegalovirus (CMV) promoter, subcloned the transfected cells and screened the clones for a high insulin expression and glucose-stimulated insulin secretion. They obtained four cell lines with a poor insulin content and four cell lines with a high insulin content named Ins-1 832/3, 832/13, 832/21 and 832/24 (Hohmeier *et al.,* 2000; **FIGURE A2B**). Merglen *et al.* also used the Ins-1 cell line to generate more glucose-responsive beta cell line called Ins-1E cells marked by a high glucose-stimulated insulin secretion (Merglen *et al.,* 2004; **FIGURE A2B**). Unlike Hohmeier *et al.,* they did not transfect the cells, but only subcloned them (Hohmeier *et al.,* 2000; Merglen *et al.,* 2004).

The Min6 cell line has been isolated from an insulinoma of a transgenic mouse expressing the SV40 large T antigen and they secrete insulin in response to glucose in the presence of nicotinamide (Miyazaki *et al.,* 1990). The murine pancreatic beta cell line Min6b1 was derived from Min6 cells. It is a subclone of the heterogenous parental Min6 cell line and secretes insulin in a glucose-dependent manner. Like Ins-1 cells, Min6b1 cells are maintained in medium containing beta-mercaptoethanol (Lilla *et al.,* 2003; **FIGURE A2C**). In addition to the Min6 and Ins-1 cell lines, other rodent beta cell lines have been produced over the years, but Ins-1 and Min6 cells remain the most widely used cells.

The development of a human beta cell line was only possible in 2011. Since the availability of human islets is low, there is a great need to have an alternative model system to study human beta cells. Ravassard *et al.* published the human beta cell line EndoC-betaH1. To produce this cell line, they transduced human foetal pancreatic tissues with lentiviruses leading to the expression of the SV40 large T antigen under the control of the *rIns2* promoter. The transduced cells were transplanted into mice where they differentiated and formed an insulinoma. The insulinoma cells were isolated, and, in a second transduction, stably transfected with the human telomerase reverse transcriptase (*hTERT*) to prevent senescence and re-transplanted into mice. After their isolation, the tumour cells could be maintained in culture, secreted insulin in response to glucose and had a transcriptome comparable of human islets (Ravassard *et al.,* 2011; **FIGURE A2D**). In 2014, Scharfmann *et al.* published an improved EndoC cell line named EndoC-betaH2 in which the transgenes can be removed after cellular expansion by means of the Cre recombinase to obtain a more beta cell-like proliferation rate together with a more differentiated state (Scharfmann *et al.,* 2014; **FIGURE A2D**).

In summary, the main characteristics of these permanent cell lines are that they can be kept in culture over a long time, that they maintain the differentiated state in culture and that they secrete insulin in response to glucose. However, the physiology of most if not all beta cell lines does not perfectly mimic that of islet cells (Skelin *et al.,* 2010, review), e.g. Ins-1 832/13 cells start to secrete insulin from 3 mM on and they secrete the maximal amount of insulin at 8 mM glucose whereas insulin secretion by rat islets only starts at 5.5 mM and reaches its maximum at 16 to 20 mM glucose (Hohmeier *et al.,* 2000). One should also not forget that beta cell lines are tumour cells that have a huge proliferation rate. Therefore, one should be careful in interpreting the data. Nevertheless, beta cell lines remain a very useful tool to study beta cell physiology and pathophysiological without the need of animals or human donors (Skelin *et al.,* 2010, review).

1.2. The pancreatic beta cells secrete the hormone insulin in response to glucose

The secretion of insulin by pancreatic beta cells is regulated by nutrient availability and hormones. The import of glucose but also of other monosaccharides, amino acids and fatty acids, as well as hormone signalling can stimulate insulin secretion, but glucose is the most potent stimulator of insulin secretion in rodents and humans (Rutter *et al.*, 2015, review; Fu *et al.*, 2013, review).

Glucose transporters (GLUT1 and GLUT3 encoded by *SLC2A1* and *SLC2A3* in humans, GLUT2 encoded by *Slc2a2* in rodents) import glucose into the cytoplasm where it is directly phosphorylated by the enzyme glucokinase (GCK) and enters the glycolysis. The final product of the glycolysis is pyruvate that is converted into the TCA cycle substrate acetyl-CoA. The mitochondrial TCA cycle produces NADH and $FADH_2$, reducing agents that are needed in the respiratory chain in the

mitochondria to produce ATP. A raise in the ATP amount causes the closure of ATP-dependent potassium channels. These channels are formed by the channel subunit $K_{ir}6.2$ encoded by *Kcnj11* and the associated ATP-binding cassette transporter ABCC8 / SUR1 (Rutter *et al.,* 2004, review). The closure of them stops the efflux of potassium ions whereby the number of positive charges within the cell exceeds that outside the cell causing the local depolarization of the plasma membrane. The depolarization triggers the opening of voltage-gated calcium channels that are multi-subunit channels encoded by various genes (e.g. *Cacna1a*, *Cacna1c*, *Cacna1d*, *Cacnb2*). The calcium ions flow with their concentration gradient into the cell and induce the fusion of insulin vesicles with the plasma membrane and the release of insulin into the blood (**FIGURE A1D**). The fusion of insulin vesicles is mediated by various SNARE proteins on the outer surface of the vesicle and the inner site of the plasma membrane (Rorsman and Braun, 2013, review).

Insulin secretion is biphasic. This is because beta cells contain two pools of insulin vesicles: a readily releasable pool of primed insulin vesicles that can directly fuse with the plasma membrane and release their content when the intracellular Ca^{2+} concentration raises. Vesicles that were not yet in contact with the plasma membrane at the time of the stimulus belong to the vesicles that might fuse with the plasma membrane in the second phase of the insulin secretion. These vesicles thus release their contents into the blood later than the first vesicles. Whether the second vesicle pool is used or not depends on the duration of the stimulus (Rorsman and Braun, 2013, review).

In addition to insulin, beta cells secrete other molecules such as C-peptide (Fu *et al.,* 2013, review) and amylin / islet amyloid polypeptide (IAPP; Westermark *et al.,* 2011, review). For a long time, it has been argued that the C-peptide is only a waste product of insulin synthesis (Hills and Brunskill, 2008, review). The insulin mRNA is translated into preproinsulin, which is composed of a signal peptide, a B chain, the C peptide and a A chain. At the endoplasmic reticulum, , the signal peptide is removed before the folding of the protein. The A chain and the B chain are connected to each other via the formation of disulphide bridges, producing proinsulin. In the Golgi, the prohormone convertases PC1 and PC2 encoded by *Pcsk1* and *Pcsk2* and the carboxypeptidase E (*Cpe*) cut off the C-peptide, resulting in the final insulin. Insulin is stored as hexamers that coordinate two Zn^{2+} ions, in large dense-core vesicles in the cytoplasm. These vesicles also contain the C-peptide and IAPP that are secreted simultaneously with insulin (Fu *et al.,* 2013, review). It has been proposed that the C-peptide is a biologically active peptide that binds to a G-protein coupled receptor (GPCR) and induces signalling cascades, but its physiological role has yet to be uncovered (Hills and Brunskill, 2008, review). IAPP is also thought to bind to a GPCR, but again the physiological role is unknown. Studies have shown that it may function as a para- and autocrine molecule regulating glucagon and insulin secretion and may control gastric emptying. Furthermore, it aggregates in cytotoxic fibrils when its expression is increased (Westermark *et al.,* 2011, review).

After its release into the bloodstream, insulin binds to insulin receptors on the surface of target cells. The binding of insulin to its receptor causes the phosphorylation of insulin receptor substrate proteins (IRS1, IRS2, IRS3 and IRS4) and leads to kinase cascades within the cell. Due to the diversity of the kinases that are involved in the downstream signalling, the insulin signalling pathway is highly complex. The activation of the PI3K-AKT-PKB pathway induces metabolic changes while the Ras-MAPK pathway is involved in the control of cell growth and differentiation (Taniguchi *et al.,* 2006, review). Insulin stimulates the glucose uptake by muscle cells, adipose tissues and hepatocytes by inducing the translocation of the glucose transporter GLUT4 to the plasma membrane, thus increasing glucose import rates (Huang and Czech, 2007, review). Moreover, insulin increases the glycogen synthesis and lipogenesis in the liver (Michael *et al.,* 2000).

1.3. The transcriptome of beta cells is adapted to their function

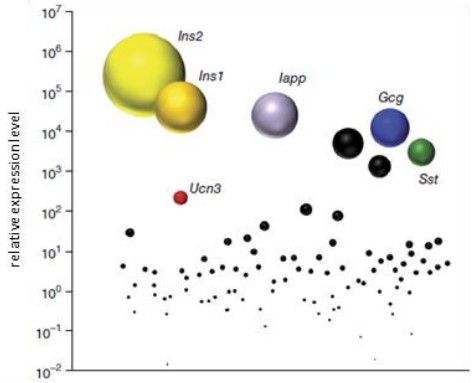
Adult beta cells highly express genes important for insulin production and secretion such as insulin, insulin processing enzymes, glucose transporters, enzymes of the glycolysis and the mitochondrial metabolism (TCA cycle and respiratory chain), ion channels and associated transmembrane proteins, components of insulin vesicles and proteins involved in the exocytosis. To ensure that insulin secretion is adapted to other external stimuli apart from glucose, they express a high number of G protein-coupled receptors and other hormone and neurotransmitter receptors. All these proteins are necessary for the maintenance and regulation of glucose-stimulated insulin secretion (Rutter *et al.,* 2015, review; Skelin Klemen *et al.,* 2017, review).

Beta cells share many common genes with the other endocrine islet cell types (**FIGURE A3**). The proximity of the different endocrine cell types within islets makes it difficult to distinguish between alpha, beta, delta, PP, and epsilon cell genes, especially when one analyses the transcriptome and transcriptional changes in whole islets. Nonetheless, insulin-expressing beta cells represent the largest population of the islets (**FIGURE A3A**).

In 2012, Ku *et al.* analysed the beta cell transcriptome in mice by using a mouse line that expresses GFP under the control of the mouse insulin promoter allowing the purification of GFP-positive beta cells by Fluorescence Activated Cell Sorting (FACS). They found 12,082 genes to be expressed in beta cells, 43 to be 4-fold enriched and 1,400 to be 4-fold decreased in beta cells compared to whole islets extracts (Ku *et al.,* 2012; Blodgett *et al.,* 2014, review). They further mentioned that 16 genes were exclusively expressed in beta cells but not in the other endocrine cells. Among these genes were the pancreatic transcription factors *Pdx1* and *Rfx6* (Ku *et al.,* 2012).

In 2013, Bramswig *et al.* studied both the epigenome and the transcriptome in sorted populations of human alpha and beta cells (**FIGURE A3B**) stained with cell-surface antibodies to identify alpha

A SECRETED FACTORS BY ISLET CELLS

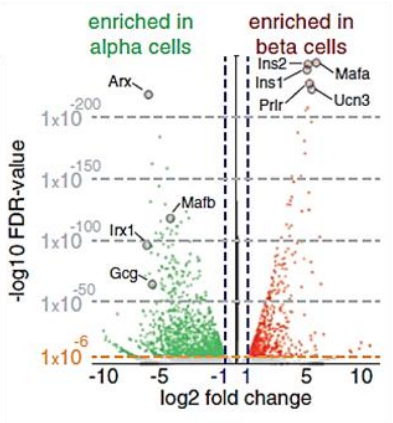


B ISLET CELL SUBTYPE MARKER GENES

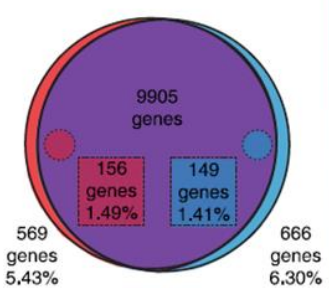
pan-endocrine: *PAX6, MORF4L1, NEUROD1, INSM1, NKX2.2*

β-cell	α-cell	ε-cell	δ-cell	γ/PP-cell
PAX4	*ARX*	*GHRLOS*	*HHEX*	*ARX*
PDX1	*MAFB*	*NPY1R*	*ETV1*	*ETV1*
MAFA	*IRX2*	*ASGR1*	*MAFB*	*UCN3*
MAFB	*GC*	*ARX*	*LEPR*	*FEV*
BMI1	*FEV*		*GHSR*	
DLK1	*TM4SF4*		*GPR120*	
SIX2/3	*GPR119*		*POUF3F1*	
ID1	*UCN3*			
IAPP	*FEV*			
UCN3	*DPP4*			
OLIG1				

C ALPHA AND BETA CELL GENES



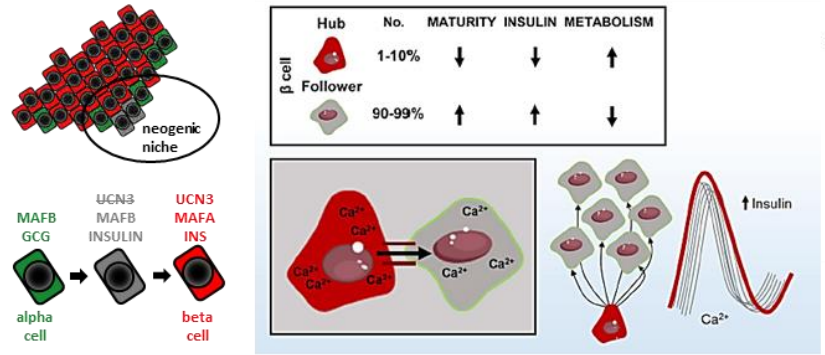
D HUMAN AND MURINE BETA CELL GENES

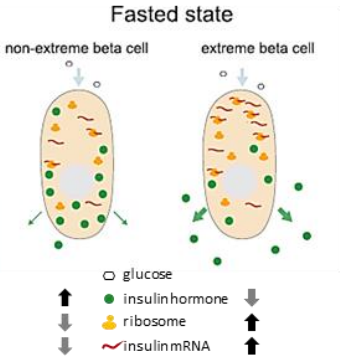


mouse beta-enriched	common beta	human beta-enriched
Iapp*	Ins2/Ins	Igf2*
Prlr*	Ucn3	Mafb*
Ghr	Glp1r	Gad2*
Cntfr	Mafa	Adcyap1
Il1r1*	Nkx6.1	Smad9
Pparg	Pdx1	Timp1
Slc2a2*	Pcsk1	Timp2*
Kcnip1	Ero1lb	Kcnk3
Bace2*	Slc2a1	Prss23*
Maob	G6pc2	Eef1a2*
Sytl4	Abcc8	Tuba1b*
Prss53*	Kcnj11	
Prkcb*		

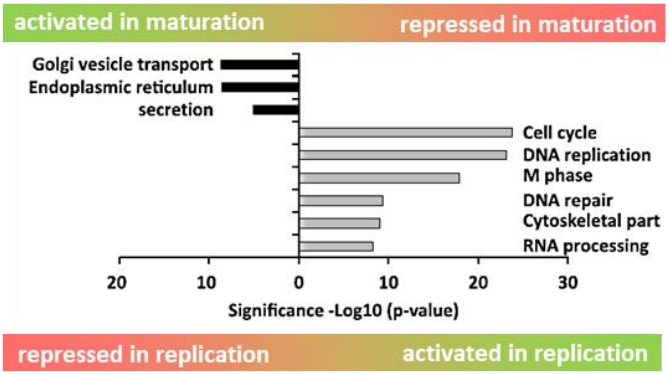
* significantly enriched in, but not exclusive to beta cells of this species

E BETA CELL SUBPOPULATIONS

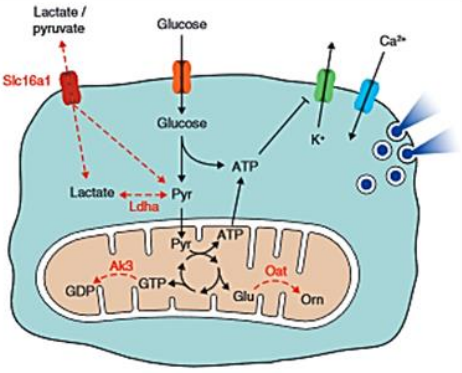




F PATHWAYS ACTIVATED IN BETA CELL PROLIFERATION



G BETA CELL DISALLOWED GENES



← **FIGURE A3: Beta cell transcriptome: insulin secretion gene, beta cell transcription factors, disallowed genes. (A)** Bulk RNA sequencing data illustrating the relative expression level of secreted factors in mouse islet cells (Van der Meulen *et al.,* 2015). **(B)** Single cell RNA sequencing data illustrating the marker genes of the five islet cell subtypes in human islets (Tritschler *et al.,* 2017, review). Genes written in blue show interspecies differences in their expression. **(C)** Bulk RNA sequencing data highlighting the genes enriched in alpha (in green) and in beta (in red) cells in mouse islet cells (Benner *et al.,* 2014). **(D)** Bulk RNA sequencing data comparing the beta cell-specific genes in murine (in red) and in human (in blue) beta cells (Benner *et al.,* 2014). **(E)** On the left: Schematic illustrating the transdifferentiation of alpha cells to beta cells within the neogenic niche (figure adapted from Van der Meulen *et al.,* 2017). In the middle: Schematic illustrating the characteristics of hub cells that orchestrate the insulin secretion (Johnston *et al.,* 2016). Schematic comparing insulin production and secretion in highly metabolic extreme beta cells and non-extreme beta cells (Farack *et al.,* 2018). **(F)** Gene ontology analysis of genes upregulated or downregulated in replicating murine beta cells compared to quiescent beta cells (Klochendler *et al.,* 2016). **(G)** Schematic showing examples for the disallowed interfering with the beta cell metabolism (Pullen and Rutter, 2013, review).

and beta cell-specific clusters. Among the beta cell-specific genes, they highlighted the transcription factors *MAFA*, *NKX6.1* and *PDX1* and genes involved in insulin biosynthesis and secretion such as *INS*, *IAPP*, *PCSK1*, *KCNJ11* and *SLC30A8*. They further reported that 35 % of the beta cell genes were marked by the activating histone modification H3K4me3 (histone 3 is trimethylated at lysine 4) and 36 % by the silencing histone modification H3K27me3 (histone 3 is trimethylated at lysine 27), while 26 % of the genes were bivalent with both histone marks. Interestingly, beta cell signature genes often had both H3K4me3 and K3K27me3 marks in alpha cells, suggesting the potential of transdifferentiation (Bramswig *et al.,* 2013).

In a study performed at the same time, Nica *et al.* identified 526 beta cell-specific genes and 614 non-beta cell genes in the beta cell population and the non-beta cell population obtained by FACS of dissociated human pancreatic islets. An analysis of the enriched pathways among the beta cell genes revealed that these genes are mostly neuron-related genes. Since neurons and beta cells are both excitable cells and the mechanism of insulin secretion shares many common features with the neurotransmitter release, this seems evident. Repressed genes in beta cells encode proteins involved in cell-cell communication, signal peptides and extracellular matrix components (Nica *et al.,* 2013).

In 2014, Benner *et al.* focussed on the transcriptional differences in the murine alpha and beta cells (**FIGURE A3C**). They generated a transgenic mouse line in which alpha and beta cells are fluorescently labelled with EGFP and mCherry, respectively. They isolated and dissociated the islets, and, after FACS, sequenced the transcriptomes of both populations, thus identifying alpha cell-enriched and beta cell-enriched genes. They determined 1075 alpha cell-specific and 1472 beta cell-specific genes. Alpha cells highly express *Gcg*, *Mafb*, *Arx*, *Irx1* and *Irx2* while beta cells express *Ins2*, *Ins1*, *Mafa*, *Prlr*, *Kcn3, Pdx1* and *Nkx6.1* (**FIGURE A3C**). The genes encoding the transcription factors *Pax6*, *Rfx6*, *Nkx2.2*, *Neurod1*, *Isl1*, *Myt1*, *Foxa2* and *Foxo1* and genes encoding ATP-sensitive potassium channels and voltage-gated calcium channels are expressed at comparable levels in both alpha and

beta cells, whereas the glucose transporter gene *Slc2a2*, the glycolytic gene *G6pc2*, the prohormone convertase *Pcsk1* and the vesicle protein *Sytl4* are enriched in beta cells (Benner *et al.*, 2014). In the second part of their study, Benner *et al.* compared the species-specific gene expression of mouse and human alpha and beta cells by comparing published RNA sequencing data of human pancreatic islets (Nica *et al.*, 2013) with their mouse data. They realized that *Ucn3* are exclusively expressed in murine beta cells and *Mafb* in murine alpha cells but co-expressed by human alpha and beta cells. Comparing mouse and human beta cells (**FIGURE A3D**), they found 666 genes that are only expressed in human beta cells, 149 genes enriched in human beta cells and 156 genes enriched in mouse beta cells. Common genes include the genes encoding insulin, UCN3, MAFA, NKX6.1, PDX1, G6PC2, ABCC8 and KCNJ11, two examples of genes enriched in mouse beta cell are *Iapp* and *Slc2a2*, while *MAFB* is an example of a gene enriched in human beta cells (Benner *et al.*, 2014).

There are not only interspecies differences between beta cells but also intraspecies differences within beta cell populations. Beta cells can differ in insulin content, insulin secretion capacity, maturity and in the proliferation rate. The existence of this heterogeneity has been known for a long time (50 years ago), but more precise studies about the heterogeneity have only been possible in recent years thanks to the new imaging, genomics and proteomics technologies and, from 2016 on, the new single cell technology permitted to extensive study of the beta cell heterogeneity. Multiple groups analysed the transcriptomes of human islets of healthy children and adults and patients with diabetes type 1 or type 2, of murine islets during embryogenesis and in adult mice and in rats (Carrano *et al.*, 2017, review; Tritschler *et al.*, 2017, review; Nasteska and Hodson, 2018, review).

In 2016, Johnston *et al.* imaged hub cells within murine islets that correspond to 1-10 % of the islets' beta cells (**FIGURE A3E**). Hub cells are metabolically highly active beta cells that dictate the insulin secretion after a glucose stimulus. They are less mature than the surrounding beta cells and only express low levels of *Pdx1* (Johnston *et al.*, 2016). In 2018, Farack *et al.* defined a minor group of so-called extreme beta cells in murine islets that highly transcribe insulin but that contain lower levels of insulin protein than non-extreme beta cells (**FIGURE A3E**). These extreme beta cells express high levels of *Ucn3* and *Pdx1* and highly secrete insulin at low glucose levels (Farack *et al.*, 2018). A third study identified *Ucn3*-negative immature beta cells accounting for 2 % of the beta cells in murine islets. These cells arise from the transdifferentiation of alpha cells into beta cells and form a neogenic beta cell niche (**FIGURE A3E**) important for beta cell regeneration (Van der Meulen *et al.*, 2017).

By single cell RNA sequencing, Dorrell *et al.* identified four beta cell populations in human islets and found that the two populations secreting the lowest levels of insulin, increase in number in diabetes type 2 (Dorrell *et al.*, 2016; Nasteska and Hodson, 2018, review). By single cell proteomics, Wang *et al.* found three populations in human islets. The PDX1/INS double positive cells were more abundant the higher the age and the weight of the donor (Wang *et al.*, 2016b). Bader *et al.* analysed the cell surface

marker FLTP and subdivided the beta cell population into mature FLTP+ and immature FLTP- cells, the latter increasing in metabolic stress (Bader *et al.,* 2016).

Overall, the single cell studies revealed that the number of immature proliferating beta cells increases, and the number of mature post-proliferative insulin-secreting beta cells decreases in diabetes type 2. This leads to a higher regeneration rate of the beta cells maintaining beta cell mass, but to a concomitant reduction in insulin secretion (Nasteska and Hodson, 2018, review).

1.4. Beta cells are mainly post-proliferative cells

In healthy mice and humans, most beta cells are post-proliferative due to the high degree of specialisation (Meier *et al.,* 2008; Teta *et al.,* 2005). In 2011, Porat *et al.* demonstrated that the metabolism of glucose by glucokinase and the depolarization of the membrane stimulate the proliferation of beta cells *in vivo* in mice. Since glucose metabolism in beta cells depends on blood sugar levels and blood sugar levels on secreted insulin, the authors concluded that beta cell proliferation is regulated by the workload of beta cells, that is the demand for insulin and its secretion. In fact, in their study, the grafting of islets in mice led to a significant decrease in beta cell proliferation in the pancreas itself due to the presence of additional insulin-secreting beta cells (Porat *et al.,* 2011).

The initiation of proliferation in pancreatic beta cells causes transcriptional changes (**FIGURE A3F**). Klochendler *et al.* compared the transcriptomes of replicating and quiescent beta cells in mouse islets. In their study, they used a mouse line that expresses GFP under the control of the cell cycle gene promoter *Ccnb1*, thus replicating beta cells are GFP-positive while quiescent beta cells are GFP-negative. They found that replicating beta cells express cell cycle regulators, proteins involved in DNA synthesis and components of the cytoskeleton to favour cell division. Simultaneously, the cells repress genes involved in the insulin secretion pathway as they must dedifferentiate to be able to proliferate (Klochendler *et al.,* 2016; **FIGURE A3F**).

In 2016, Schmidt *et al.* came to a similar conclusion when they studied the glucose-induced transcriptional changes in the rat Ins-1E cell line. They analysed the transcriptomes of these cells after 0 h, 1 h, 2 h, 4 h and 12 h in high glucose medium and they observed a biphasic glucose response marked by increased expression of insulin secretion genes shortly after the glucose stimulus and a loss of the beta cell identity at a later point when the cells prepared cell division (Schmidt *et al.,* 2016).

1.5. Disallowed genes are a group of genes specifically silenced in pancreatic beta cells

Genes coding for proteins that are not needed in quiescent insulin-secreting pancreatic beta cells are selectively silenced in beta cells (Quintens *et al.,* 2008; Pullen *et al.*, 2010; Thorrez *et al.,* 2011; Pullen and Rutter, 2013, review; **FIGURE A3G**). These genes are involved in multiple cellular

processes such as metabolism, proliferation, oxidative stress, vesicle-mediated transport and the formation of the cytoskeleton (Pullen and Rutter, 2013, review) and, if expressed, the encoded proteins would interfere with proper beta cell function.

Examples for silenced genes are the ones encoding the enzyme lactate dehydrogenase A (LDHA) that catalyses the conversion of pyruvate to lactate, thus lowering the conversion of pyruvate to Acetyl-CoA (Sekine *et al.*, 1994) and the transporter SLC16A1 that imports monocarboxylates such as lactate and pyruvate (Zhao *et al.*, 2001). The genes encoding adenylate kinase 3 (AK3) and ornithine aminotransferase (OAT) are also only expressed at low levels in beta cells. AK3 converts mitochondrial GTP of the TCA cycle into GDP, and OAT converts glutamine to ornithine. Both GTP and glutamine levels were proposed to regulate insulin secretion (Pullen *et al.*, 2010). Genes that are specifically silenced in beta cells, are called disallowed genes (Quintens *et al.*, 2008; Pullen *et al.*, 2010; Thorrez *et al.*, 2011).

1.6. Glucolipotoxicity, long-term damage of beta cells due to high lipid and glucose levels

A prolonged exposure to high glucose levels in chronic hyperglycaemia, however, leads to phenomena called glucotoxicity, marked by beta cell dysfunction and apoptosis via the production of reactive oxygen species causing oxidative stress. Moreover, the increased metabolism of glucose and the subsequent production of fatty acid metabolism leads to the accumulation of lipids in the cytoplasm of beta cells, a process referred to as glucolipotoxicity (Poitout and Robertson, 2008, review).

Beta cells are very sensitive to oxidative stress because they do not express high levels of antioxidant enzymes. Tiedge *et al.* reported that the expression of enzymes superoxide dismutase, catalase and glutathione peroxidase are very low in the rat beta cell line RINm5. Moreover, RINm5 cells and rat pancreatic islets do not upregulate antioxidant genes in response to stress (Tiedge *et al.*, 1997). Welsh *et al.* showed that human islets are slightly better protected against oxidative stress than rodent islets since they express higher levels of superoxide dismutase, catalase and heat-shock protein 70 (Welsh *et al.*, 1995).

Due to their high sensitivity to stress, the addition of the reducing agent beta-mecaptoethanol to the medium of beta cell lines prevents the oxidation of proteins. It was first added to the medium of Ins-1 cells by Asfari *et al.* who found that it is necessary to maintain the differentiated state and the proliferation of Ins-1 cells (Asfari *et al.*, 1992). This was before the discovery that beta cells are very sensitive to stress, but the concept of glucotoxicity inducing beta cell stress in high glucose and the fact that a reducing agent such as beta-mercaptoethanol is beneficial for the culture of beta cell lines, are consistent.

1.7. The expression of *Txnip* is highly induced by glucose in pancreatic beta cells

The gene most strongly activated by glucose in beta cells is *Txnip* encoding the thioredoxin-binding protein. This was demonstrated by Shalev *et al.* in 2005 who analysed the glucose-induced genes in human pancreatic islets using a microarray. The expression of *Txnip* was 11-fold induced after 24 h in high glucose (Shalev *et al.,* 2002). Minn *et al.* demonstrated that *Txnip* is also induced by glucose in murine islets and the pancreatic beta cell line Ins-1 and overexpressed in islets of diabetic mice (Minn *et al.,* 2005). In 2012, Poungvarin *et al.* showed that the exposure of isolated islets to high glucose for 48 h induced *Txnip* expression and a 3-fold induction of caspase 3/7 (Poungvarin *et al.,* 2012).

However, one should mention that the transcription of *Txnip* is also induced by ER stress (Lerner *et al.,* 2012; Oslowski *et al.,* 2012). TXNIP binds and represses thioredoxin in the cytoplasm. Thioredoxin reduces oxidized proteins and its inhibition causes the accumulation of oxidized proteins, leading to oxidative stress and apoptosis. Moreover, TXNIP binds to mitochondrial thioredoxin that forms a complex with ASK1. The release of ASK1 and its subsequent phosphorylation further induces apoptosis via the release of cytochrome C from the mitochondria and the cleavage of caspase 3 (Shalev, 2014, review).

1.8. The transcriptional control of insulin depends on the glucose level

Probably insulin itself is the most prominent glucose-activated gene. In 1991, Efrat *et al.* demonstrated that glucose increases the transcription of both insulin genes by 3-fold in a murine pancreatic beta cell line after 30 min in high glucose. They further postulated that the same pathway that induces insulin secretion, also controls insulin transcription because they observed that a calcium channel blocker inhibited glucose-mediated insulin upregulation (Efrat *et al.,* 1991). In 1998, Leibiger *et al.* tried a shorter exposure to high glucose of 15 min and they reported that insulin mRNA levels were increased by 5-fold in rat islets 90 min after the stimulus. In a second study, they showed that the release of insulin after a glucose stimulus leads to autocrine insulin actions that result in increased insulin transcription (Leibiger *et al.,* 1998a; Leibiger *et al.,* 1998b). Taken together, glucose and insulin signalling positively influence insulin transcription. In the follow up of their studies in 1998, Leibiger *et al.* further demonstrated that both pathways also stimulate insulin biosynthesis (Leibiger *et al.,* 2000). Thus, not only the transcription of insulin but also its translation is regulated by glucose and insulin. This control network is necessary to efficiently replenish the insulin supplies and to have enough vesicles available.

As mentioned, there is one insulin gene in humans (*INS*) and two in mice and rat (*Ins1* and *Ins2*). It has been reported that the two rodent genes are co-ordinately regulated (Giddings and Carnaghi, 1988; Koranyi *et al.,* 1989; Melloul *et al.,* 2002, review). Multiple transcription factor binding sites have

been identified in the 340 bp long insulin promoter sequence that may be bound by the respective transcription factors (Melloul *et al.,* 2002, review; Andrali *et al.,* 2008, review). It has been postulates that PDX1, NEUROD1 and MAFA are the most important factors regulating insulin transcription binding to the A3, E1 and C1 binding element (Andrali *et al.,* 2008, review).

Phosphorylated PDX1 recruits the histone acetyl transferase p300 and the histone methyltransferase SET9 to the insulin locus in high glucose. The combination of histone acetylation and H3K4 dimethylation by these histone modifying enzymes favours transcription. By contrast, in low glucose, dephosphorylated PDX1 mediates histone deacetylation by recruiting HDAC to the insulin locus. The phosphorylation of PDX1 is glucose-dependent and regulated by several kinase signalling pathways (Andrali *et al.,* 2008, review). The translocation of NEUROD1 from the cytoplasm to the nucleus is induced by post-transcriptional modifications (phosphorylation and O-linked N-acetyl-glycosylation) in high glucose. In the nucleus, NEUROD1 binds to the insulin gene as a dimer with the basic helix-loop-helix protein E47 and they interact with the transcriptional activator histone acetyl transferase p300 that competes with the transcriptional repressor SHT (Andrali *et al.,* 2008, review). The expression of the third transcription factor *Mafa* is increased in high glucose and MAFA activates insulin transcription in synergy with PDX1 and NEUROD1 (Andrali *et al.,* 2008, review).

These three transcription factors are only the major factors involved in insulin transcriptional regulation. The regulation of insulin is a highly complex process controlled by multiple transcription factors and signalling cascades, with glucose having a positive effect on insulin transcript levels, biosynthesis and secretion.

In conclusion, the beta cell transcriptome is perfectly adapted to their main task – the glucose-stimulated insulin secretion – and is tightly regulated in response to extracellular stimuli. Transcriptional control is achieved by a complex regulatory network build of transcription factors, chromatin modifying enzymes and signalling cascades induced by external stimuli with glucose playing a major role in gene expression regulation in pancreatic beta cells (Schuit *et al.,* 2002, review; Rutter *et al.,* 2015, review).

2. The metabolic disorder diabetes is caused by disturbances in the insulin signalling pathway

The secretion of the peptide hormone insulin is indispensable for the survival. Under hyperglycaemia, beta cells release insulin into the blood to enable the glucose uptake by peripheral tissues. In the case that beta cells do not provide enough insulin or that peripheral tissues become resistant to insulin action, the blood glucose level stays elevated leading to diabetes mellitus. According to the world health organization (WHO), 422 million of people in the world were diagnosed as having diabetes. At long term, an increased blood glucose level damages endothelial cells in small

blood vessels due to the increased osmotic potential of the blood. Furthermore, hyperglycaemia favours infections. Possible long-term consequences of hyperglycaemia are blindness, kidney failure, heart attacks, stroke, nerve damage and necrosis in limbs (WHO, 2018; Rask-Madsen and King, 2013, review).

2.1. Monogenic diabetes is a rare subtype of the pandemic diabetes mellitus

Diabetes mellitus is divided into different classes according to its causes and symptoms. The most common form of diabetes is type II diabetes (90 % of all cases). Patients with type II diabetes have a relative lack of insulin since prolonged demand of increased insulin levels due to both lifestyle factors and genetic components results in insulin resistance of peripheral tissues. In contrast, type I diabetes is an autoimmune disease; the immune system attacks pancreatic beta cells causing an absolute lack of insulin due to beta cell death (Mastracci and Sussel, 2012, review). Type I diabetes is known as insulin-dependent diabetes because patients need to be treated with insulin. On the other hand, the treatment of patients with type II diabetes can be either with drugs that increase insulin sensitivity or inhibit gluconeogenesis in the liver, or with insulin injections, or both (WHO, 2018; Ashcroft and Rorsman, 2012). Gestational diabetes, as the word implies, occurs during pregnancy and stops after giving birth (Coustan, 2013, review).

Besides type I and II diabetes, there are rare subtypes of diabetes called monogenic diabetes that are caused by mutations in a single gene. Since it is an inherited disease, monogenic diabetes has an early onset; it affects neonates (neonatal diabetes) or children and young adults (Maturity Onset Diabetes of the Young, MODY). Affected genes encode proteins that are essential for beta cell development or function and the disease is caused by beta cell lack or dysfunction. Genes that has been associated with monogenic diabetes encode pancreatic transcription factors, e.g. PDX1, PTF1A, NGN3, HNF1A, HNF4A, GLIS3, NEUROD1 and RFX6, and key genes of the insulin secretion pathway such as *INS*, *SLC2A2*, *GCK*, *KCNJ11* and *ABCC8* (Schwitzgebel, 2014, review; Ashcroft and Rorsman, 2012, review). Mutations in the *RFX6* gene cause a syndrome called Mitchell-Riley syndrome, a rare form of monogenic diabetes with a neonatal onset (Mitchell *et al.,* 2004; Smith *et al.,* 2010).

2.2. Homozygous or compound heterozygous *RFX6* mutations cause Mitchell-Riley syndrome

In 2004, Mitchell *et al.* described five cases of a syndrome characterized by an underdevelopment or a malformation of the pancreas, the intestine and the bile duct system as well as a missing or underdeveloped gallbladder, diarrhea and intrauterine growth restriction. In medical terms, the observed developmental defects included agenesis (absence of an organ) or hypoplasia (below-average number of cells in an organ) of the pancreas or annular pancreas (a ring of pancreatic tissue

surrounds the duodenum); intestinal atresia (a passage within the intestinal tube is closed), stenosis (a passage within the intestinal tube is narrowed) or malrotation; cholestasis (the bile flow from the liver to the duodenum is impaired); agenesis or hypoplasia of the gall bladder. Among the diseased newborns were three girls and two boys. Although the malformations of the colon were surgically corrected, and the neonatal diabetes was treated by insulin therapy, four of the five children died within the first months of their lives due to multi organ or liver failure. The authors figured that the observed phenotypes were an autosomal recessive syndrome and they suspected that is was due to an inherited mutation in a gene involved in the embryonic development of the pancreas and the gastrointestinal system, but they failed to identify the exact genetic cause (Mitchell *et al.,* 2004). The phenotype partially overlapped with the Martinez-Frias syndrome (OMIM #601346) described in 1992 (Martinez-Frias *et al.,* 1992). The main differences were neonatal diabetes, which was found in Mitchell *et al.* syndrome, and the possible occurrence of tracheoesophageal fistula in Martinez-Frias syndrome (Mitchell *et al.,* 2004; Martinez-Frias *et al.,* 1992).

In fact, in 1969 (or even earlier), a few clinical cases were described that had the same symptoms as the patients from the Mitchell study (Verwest *et al.,* 2000) but the phenotype remained rare as expected for a recessive disease. In 2007, 2008 and 2009 three groups reported cases like those described in 2004 (Galán-Gómez *et al.,* 2007; Chappell *et al.,* 2008; Martinovici *et al.,* 2009), again without knowledge of the genetic cause.

It was not until 2010 that Michael German's laboratory was able to identify the genetic cause of the disease (Smith *et al.,* 2010): an analysis of the DNA from patients in the Mitchell study (case 2 and 5, Mitchell *et al.,* 2004), the Chappell study (Chappell *et al.,* 2008), the Martinovici study (Martinovici *et al.,* 2009), and a new case they published (Smith *et al.,* 2010), revealed that in all cases the gene RFX6 was mutated (**FIGURE A4**, **TABLE A1**). They called the phenotype Mitchell-Riley syndrome (OMIM #615710). At the same time, it was demonstrated by the same group (Smith *et al.,* 2010) and our group (Soyer *et al.,* 2010) that, during mouse embryogenesis, RFX6 is expressed broadly expressed in the primitive gut tube including the pancreatic buds before it becomes restricted to the stomach, intestine and pancreas. Within the pancreas, the expression of RFX6 is restrained to the NGN3-positive endocrine progenitors and maintained in all developing and mature islet cells (Smith *et al.,* 2010; Soyer *et al.,* 2010). These findings contradict a result of Ku *et al.* who reported that *Rfx6* is exclusively expressed in beta cells and not in the other endocrine cell types (Ku *et al.,* 2012). In the intestine, RFX6 is expressed in NGN3-postive enteroendocrine progenitor cells (unpublished data of Julie Piccand). A global knockout of *Rfx6* in mice led to the death of newborn mice within the first two to three days after birth. No glucagon-expressing alpha, insulin-producing beta, somatostatin-producing delta and ghrelin-producing epsilon cells could be found in these mice whereas the number of pancreatic polypeptide-secreting cells increased while pancreatic exocrine function was not impaired (Smith *et*

al., 2010; Piccand *et al.,* 2014; **FIGURE A7A**). Immunostainings on a pancreas from a deceased newborn with Mitchell-Riley syndrome showed a total absence of insulin-, glucagon- and somatostatin-positive cells and atrophy of the exocrine compartment of the pancreas (Mitchell *et al.,* 2004). In mice, all hormone-secreting enteroendocrine cell types of the gut except serotonin-producing cells were lost (unpublished data of Julie Piccand). Furthermore, like the human infants, the pups had intestinal atresia and malformations of the biliary system (Smith *et al.,* 2010). The intestinal atresia was not observed in our knockout mouse line (Piccand *et al.,* 2014). All these findings are consistent with the hypothesis of Mitchell *et al.* that a gene involved in the embryonic development of the pancreas and gastrointestinal system must be responsible for Mitchell-Riley syndrome (Mitchell *et al.,* 2004).

Following this breakthrough, eight more cases of Mitchell-Riley syndrome with mutations in RFX6 have been published to date (Spiegel *et al.,* 2011; Chandra *et al.,* 2014; Concepcion *et al.,* 2014; Cheung *et al.,* 2015; Zegre Amorim *et al.,* 2015; Huopio *et al.,* 2016; Khan *et al.,* 2016; Poidvin *et al.,* 2016; **FIGURE A4**, **TABLE A1**). The treatment of patients seems to have improved in recent years, as of the eight new cases only one child died (Concepcion *et al.,* 2014). The diabetes is treated with insulin injections, the malformations of the intestine are corrected surgically, and the children receive short-term or long-term parenteral nutrition to compensate for nutrient loss through diarrhea, malabsorption and the lack of digestive juice. Many children are also anemic and have high bilirubin levels due to cholestasis (**TABLE A1**). Moreover, different groups reported patches of gastric (Sansbury *et al.,* 2015; Skopkova *et al.,* 2016) or pancreatic (Zegre Amorim *et al.,* 2015; Skopkova *et al.,* 2016) tissue in the intestine, a phenotype called gastric or pancreatic heteroplasia. Interestingly, in our global RFX6 knockout mouse, we also observed gastric patches in the intestine of mouse embryos as well as an increased expression of gastric genes in an RNA sequencing experiment on intestinal tissues from E18.0 mice (unpublished data of Julie Piccand).

The identified *RFX6* mutations include point mutations, frameshift mutations and the loss or creation of splicing sites leading to aberrant splicing, and their location was not limited to a single domain of the protein but concerned all domains (DNA-binding domain, B and C domain, dimerization domain) as well as out-of-domain regions (**FIGURE A4**). Functional tests have been carried out by different groups to assess the DNA-binding (Smith *et al.,* 2010), dimerization (Smith *et al.,* 2010) and target gene regulation (Chandra *et al.,* 2014) of RFX6 proteins carrying point mutations, and their ability to rescue a *Rfx6* knockdown (Pearl *et al.,* 2011; Chandra *et al.,* 2014). These experiments and their results will be discussed in more detail in the results section of this thesis. The authors suspected that there was a correlation between the position of the mutation, the functionality of the mutant RFX6 protein and the severity of the disease suggesting that mutations in the DNA-binding domain and in the dimerization domain cause more severe phenotypes. In agreement with these findings, Sansbury

TABLE A1: Symptoms and survival of patients with Mitchell-Riley syndrome. Published cases of Mitchell-Riley syndrome (* = poster, E = exon, I = Intron). Lines filled in light pink mark deceased children, lines filled in light green children with a delayed onset of diabetes.

#	case report	sexe	age	zygosity	*RFX6* mutation
1	Mitchell *et al.*, 2004 Smith et al., 2010	F	† 6 months	homozygous	c.380+2T>C (I2, splicing)
2	Mitchell *et al.*, 2004 Smith et al., 2010	F	9 years	compound heterozygous	c.224-12A>G (I1, splicing) c.672+2T>G (I6, splicing)
3	Chappell *et al.*, 2008 Smith *et al.*, 2010 Spiegel *et al.*, 2011 Pearl *et al.*, 2011	F	6 years	homozygous	c.649T>C (E6, p.S217P)
4	Martinovici *et al.*, 2009 Smith *et al.*, 2010 Pearl *et al.*, 2011	M	† 3 months	homozygous	c.542G>A (E4, p.R181Q)
5	Smith *et al.*, 2010	M	† 2.5 months	homozygous	c.776_780+8del (E7/I7, p.R215VfsX9)
6	Spiegel *et al.*, 2011	M	1.9 years	homozygous	c.779_787del; c.778_779insG (I7/E8, splicing)
7	Artuso *et al.*, 2014	F	14 years	heterozygous	c.1678G>A (E15, p.D560N)) + GCKR, HNF1A, KCNJ11, WFS1 mutations
8	Artuso *et al.*, 2014	M	19 years	heterozygous	c.1865C>A (E16, p.T622K) + HNF1A, KCNJ11 mutations
9	Artuso *et al.*, 2014	F	13 years	heterozygous	c.1558A>T (E15, p.S520T) + ABCC8, GCKR, HNF1A, KCNJ11, PPARG mutations
10	Chandra *et al.*, 2014 Poidvin *et al.*, 2016*	F	8 years	homozygous	c.1517T>G (E14, p.V506G)
11	Concepcion *et al.*, 2014	M	† 5 months	homozygous	c.779A>C (E7, p.K260T / splicing)
12	Cheung *et al.*, 2015*	F	1 year	homozygous	c.1556-40T>G (I14, splicing)
13	Sansbury *et al.*, 2015	F	9 years	compound heterozygous	c.2176C>T (E17, p.R726X) c.2596C>T (E18, p.R866X)
14	Sansbury *et al.*, 2015	M	9 years	compound heterozygous	c.2176C>T (E17, p.R726X) c.2596C>T (E18, p.R866X)
15	Zegre Amorim *et al.*, 2015 Poidvin *et al.*, 2016*	F	7 years	homozygous	c.541C>T (E4, p.R181W)
16	Huopio *et al.*, 2016	M	<1 month	homozygous	c.878_879del (E9, p.H293LfsX7)
17	Khan *et al.*, 2016	M	9 years	homozygous	c.1153C>T (E11, p.R385X)
18	Poidvin *et al.*, 2016* Zegre Amorim *et al.*, 2015	M	2.5 years	homozygous	c.541C>T (E4, p.R181W)
19	Skopkova *et al.*, 2016	F	13 years	compound heterozygous	c.1154G>A (E11, p.R385Q) c.1346_1319del (E13, p.I439TfsX13)
20	Skopkova *et al.*, 2016	F	8 years	compound heterozygous	c.1154G>A (E11, p.R385Q) c.1346_1319del (E13, p.I439TfsX13)

pancreas phenotype	intestinal phenotype	liver phenotype	other symptoms	treatment
neonatal diabetes, annular pancreas	duodenal and jejunal atresia, diarrhoea	gallbladder agenesis, cholestasis	low birth weight	surgery, insulin therapy, parenteral nutrition
neonatal diabetes, pancreas hypoplasia	duodenal atresia, duodenal stenosis, malrotation	gallbladder agenesis, cholestasis, hyperbilirubinemia	low birth weight	surgery, insulin therapy
neonatal diabetes	duodenal atresia, anal stenosis, malrotation, diarrhoea	gallbladder agenesis, cholestasis, hyperbilirubinemia	low birth weight	surgery, insulin therapy
neonatal diabetes, pancreas hypoplasia	duodenal and jejunal atresia, malrotation	gallbladder agenesis, cholestasis	low birth weight anemia	surgery, insulin therapy, parenteral nutrition
neonatal diabetes	duodenal atresia		low birth weight	surgery
neonatal diabetes, annular pancreas	duodenal and jejunal atresia, diarrhoea	gallbladder agenesis, cholestasis, hyperbilirubinemia	low birth weight anemia	surgery, insulin therapy, parenteral nutrition
diabetes at 6 years				
diabetes at 18 years				
diabetes at 10 years				insulin therapy
neonatal diabetes, pancreas hypoplasia	jejunal atresia, duodenal stenosis, diarrhoea	gallbladder agenesis	low birth weight anemia	surgery, insulin therapy, parenteral nutrition
neonatal diabetes, annular pancreas	duodenal atresia, malrotation, diarrhoea	gallbladder agenesis, cholestasis, hyperbilirubinemia	low birth weight anemia	surgery, insulin therapy, parenteral nutrition
neonatal diabetes, annular pancreas	duodenal atresia, diarrhoea	gallbladder agenesis, cholestasis, hyperbilirubinemia		surgery, insulin therapy, parenteral nutrition
diabetes at 2-3 years	duodenal atresia, jejunal stenosis, **gastric heteroplasia**	gallbladder agenesis	low birth weight	surgery, insulin therapy
diabetes at 5-6 years	duodenal atresia, malrotation		low birth weight	insulin therapy
neonatal diabetes, annular pancreas	duodenal atresia, jejunal stenosis, **pancreatic heteroplasia**, diarrhoea	gallbladder agenesis, cholestasis, hyperbilirubinemia	low birth weight	surgery, insulin therapy, parenteral nutrition
neonatal diabetes	duodenal atresia, diarrhoea		low birth weight	insulin therapy
neonatal diabetes, pancreas hypoplasia	duodenal atresia, diarrhoea	gallbladder agenesis, cholestasis, hyperbilirubinemia	low birth weight anemia	surgery, insulin therapy, parenteral nutrition
neonatal diabetes, pancreas agenesis	duodenal atresia, diarrhoea	gallbladder agenesis	low birth weight	surgery, insulin therapy, parenteral nutrition
diabetes at 2.9 years	duodenal atresia, **gastric heteroplasia**, diarrhoea	gallbladder hypoplasia	low birth weight anemia	surgery, insulin therapy
diabetes at 2.6 years, annular pancreas	duodenal atresia, **gastric heteroplasia**, **pancreatic heteroplasia**, diarrhoea	gallbladder hypoplasia	anemia	surgery, insulin therapy

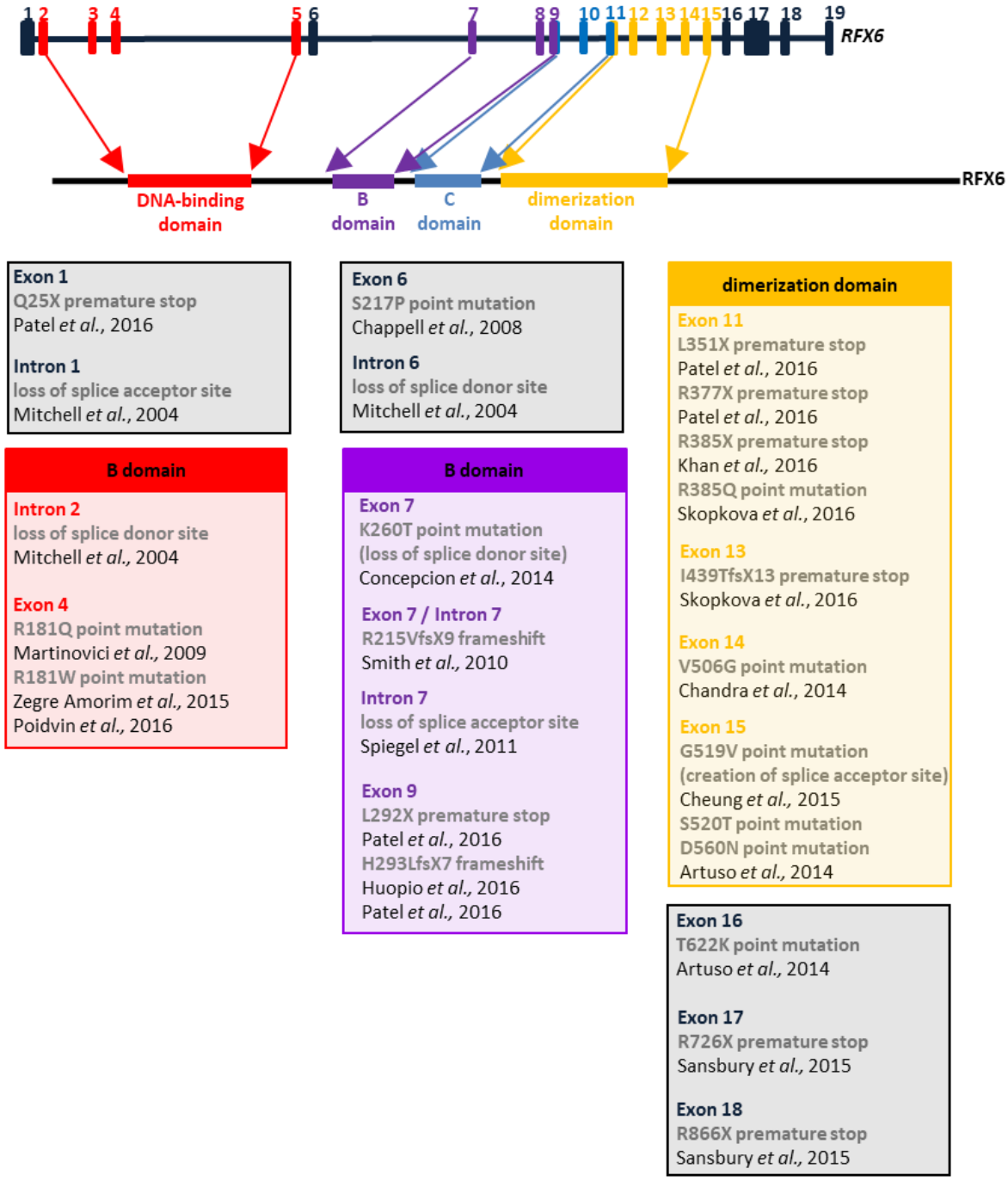


FIGURE A4: Human *Rfx6* mutation identified in Mitchell-Riley syndrome patients. Position and nature of mutations found in human *Rfx6* gene in patients with Mitchell-Riley syndrome (protein structure adapted from Aftab *et al.,* 2008). The figure supplements **TABLE A1**.

et al. report cousins with Mitchell-Riley syndrome with a delayed onset of diabetes at 3 or 6 years of age caused by compound heterozygous *RFX6* mutations identified in the C-terminal part of RFX6 downstream of the dimerization domain (Sansbury *et al.,* 2016; **FIGURE A4**, **TABLE A1**). In contrast, the siblings described by Skopkova *et al.* with RFX6 mutations in the dimerization domain also had a delayed onset of diabetes after completion of the second year of life (Skopkova *et al.,* 2016;

FIGURE A4, **TABLE A1**). Not only the location of the mutation is important, but also the chemical properties of the actual amino acid and those that it replaces, as well as the protein conformation that results. If amino acids that directly contact the DNA or participate in dimerization, are affected, the consequences for protein function are likely to be more severe, while substitution of another amino acid in the same protein domain may not have much impact. All these factors make it difficult to draw conclusions. However, the study of point mutations helps to better understand how the protein works and to identify important areas and amino acids.

2.3. Diabetes in carriers of heterozygous *RFX6* mutations has reduced penetrance

Since Mitchell-Riley syndrome is recessively inherited, the patients are carriers of a homozygous or a compound heterozygous mutation. Many publications report consanguinity (kinship of the parents; Mitchell *et al.,* 2004; Chappell *et al.,* 2008; Martinovici *et al.,* 2009; Chandra *et al.,* 2014; Zegre Amorim *et al.,* 2015; Khan *et al.,* 2016) explaining the frequent occurrence of homozygous mutations (**TABLE A1**). The expression of a phenotype in heterozygous gene carriers is controversial. Some parents of the patients, all heterozygous carriers of the *RFX6* mutations, have been described as asymptomatic (Mitchell *et al.,* 2004; Chappell *et al.,* 2008; Spiegel *et al.,* 2011; Concepcion *et al.,* 2014; Sansbury *et al.,* 2015; Huopio *et al.,* 2016), others as symptomatic. Either the mother, the father or both parents were affected. They suffered either from gestational diabetes (Smith *et al.,* 2010), type I diabetes (Smith *et al.,* 2010) or type II diabetes (Martinovici *et al.,* 2009; Sansbury *et al.,* 2015; Skopkova *et al.,* 2016) or they had impaired fasting glucose levels that had not yet been classified as diabetes (Martinovici *et al.,* 2009).

The authors of the Sansbury study that described cousins born with Mitchell-Riley syndrome (**TABLE A1**), measured the amount of glycated hemoglobin (HBA_{1C}, an indicator of the average of blood glucose levels over several months) in four non-diabetic heterozygous family members and in eight non-diabetic family members without *RFX6* mutation. The authors did not find any difference, but the sample size was very small (Sansbury *et al.,* 2015).

A larger cohort was tested in the Huopio study: They compared 7040 non-diabetic (group 1) with 7585 non-diabetic or newly diagnosed (group 2) participants. They identified 26 heterozygous carriers of a *RFX6* mutation (p.H293LfsX7; **FIGURE A4**, **TABLE A1**) in group 1 and 29 heterozygous carriers in group 2, and they measured the fasting plasma glucose concentration, the 2 h plasma glucose concentration after an oral glucose test (OGTT), the insulin secretion and insulin sensitivity. Only for the OGTT, they claim that there were significant differences between carriers and noncarriers (group 1: 6.04±1.68 mmol/l in non-carriers versus 6.80±1.71 mmol/l in carriers, p=0.025; group 2: 6.46±2.43

mmol/l in non-carriers versus 7.29±2.35 mmol/l in carriers, p=0.032). However, they mentioned that the possible influence of other mutations in other genes must be considered (Huopio *et al.,* 2016).

Artuso *et al.* described three cases of early onset diabetes at 6, 10 and 18 years of age caused by heterozygous RFX6 mutations, but the patients did not show any other typical symptoms of Mitchell-Riley syndrome, and the diabetes was rather mild, just one of them insulin-dependent. In addition to the *RFX6* mutation, mutations in other genes that have been classified as diabetes risk alleles, have been identified in the patients (Artuso *et al.,* 2014; **TABLE A1**).

In a recent study, Patel *et al.* showed that carriers of heterozygous *RFX6* mutations that cause a premature stop codon (**TABLE A1**), were enriched in maturity-onset diabetes of the young (MODY) cohorts (cohort 1: 2/38; cohort 2: 4/348; cohort 3: 6/80) compared to control cohorts (cohort 1: 15/33,346; cohort 2: 2/7,508; cohort 3: 26/7040) and they excluded mutations in known MODY risk genes. In total, they identified multiple new *RFX6* mutations during the analysis of the data sets (see supplementary data of Patel *et al.,* 2017), only the five mentioned in the article itself are shown in **FIGURE A4**. Next, they assessed the penetrance in a subgroup of 18 heterozygous individuals. At the study entry, five of them did not develop diabetes, five were diabetic before the age of 25 and 8 after the age of 25. The authors concluded that heterozygous *RFX6* mutations have a reduced penetrance. They also determined the proportion of carriers of heterozygous *RFX6* mutations in a diabetes type II cohort (12/8,373) and a control cohort (7/8466), but there was no enrichment in the diabetes type II cohort (Patel *et al.,* 2017).

In conclusion, all these studies do not show a clear link between heterozygous *RFX6* mutations and the onset of diabetes. Missing information about the relatives of the patients as well as possible false statements further complicate this evaluation. Also, the influence of the location of the mutation on the phenotype could not be clarified in both heterozygous and homozygous states in patients with Mitchell-Riley syndrome. Presumably both genetic and environmental factors play an important role in the expression of the phenotype and make a clear statement difficult.

3. The winged-helix transcription factor RFX6 belongs to the RFX transcription factor family

RFX6 is part of the RFX transcription factor family that has eight known members, RFX1-7, and since 2017, RFX8 as eights member (Aftab *et al.,* 2008; Sugiaman-Trapman *et al.,* 2018; **FIGURE A5**). RFX protein were identified in 1989 as transcription factors that bind to major histocompatibility complex (MHC) class II genes (Reith *et al.,* 1988; Reith *et al.,* 1989; Reith *et al.,* 1990). The first cloned family member was *RFX1*, followed by *RFX4*, *RFX2*, *RFX3* and *RFX4* (Aftab *et al.,* 2008; Reith *et al.,* 1989; Dotzlaw *et al.,* 1992; Reith *et al.,* 1994; Steimle *et al.,* 1995). Reith *et al.* cloned the *RFX1*, *RFX2* and *RFX3* mRNAs from human B lymphocytes (Reith *et al.,* 1989; Reith *et al.,* 1994). A partial *RFX4* cDNA

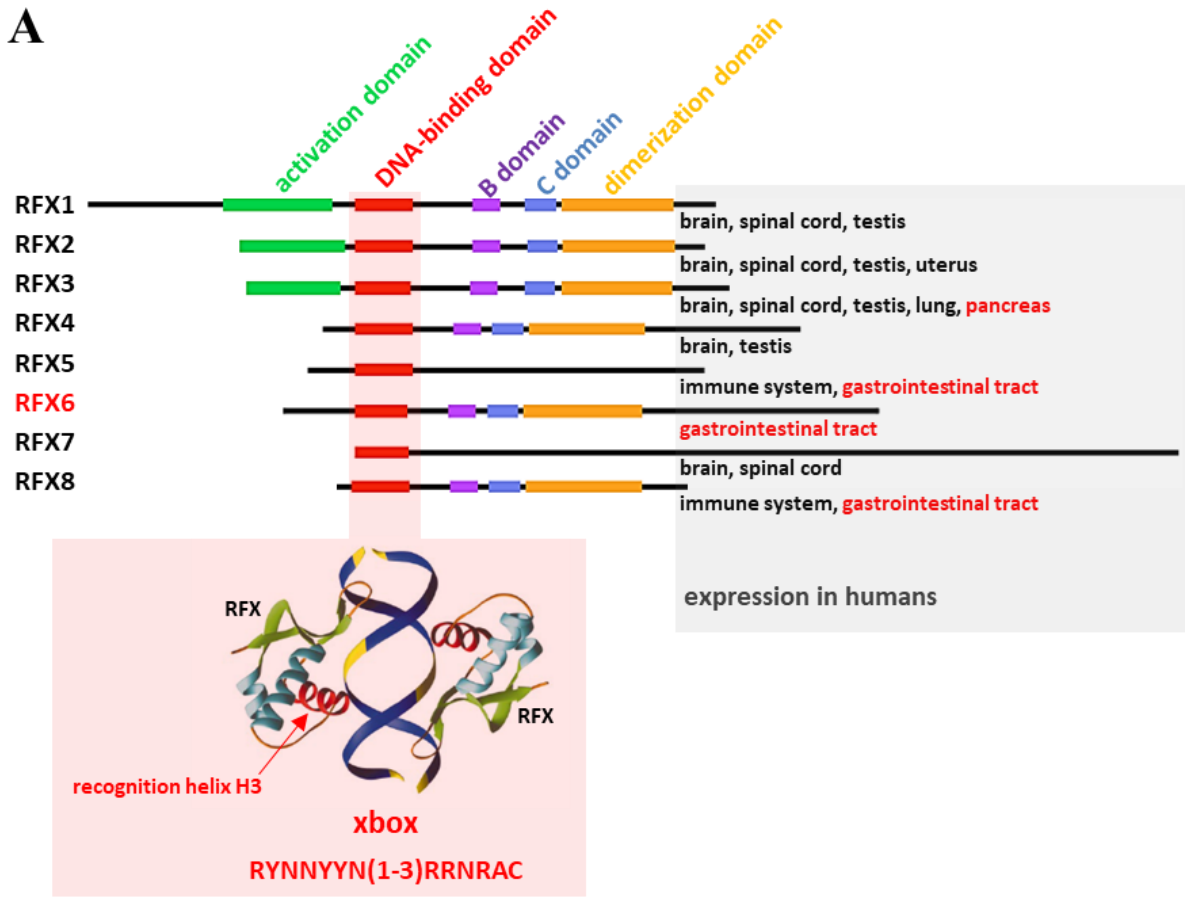


FIGURE A5: RFX6 belongs to the RFX transcription factor family. (A) RFX transcription factor family, expression in humans and DNA-binding mechanism (figures adapted from Aftab *et al.,* 2008; Sugiaman-Trapman *et al.,* 2018; Gajiwala *et al.,* 2000; Laurençon *et al.,* 2007). **(B)** Categories of RFX proteins based on their domains (Sugiaman-Trapman *et al.,* 2018). AD = activation domain, DBD = DNA-binding domain, B = B domain, C = C domain, DIM = dimerization domain.

was cloned from cDNA libraries of human breast cancer (Aftab *et al.,* 2008; Dotzlaw *et al.,* 1992), the full-length *RFX4* from a human testis cDNA library in 2002 (Morotomi-Yano *et al.,* 2002). Mutations in the *RFX5* gene were found in patients with bare lymphocyte syndrome, a syndrome that is caused by a dysregulation of MHC class II genes and is characterized by immunodeficiency, and the *RFX5* cDNA was cloned from a patient's B lymphocyte cell line (Steimle *et al.,* 1995). In 2008, Aftab *et al.*

bioinformatically identified the *RFX6* and *RFX7* gene (Aftab *et al.*, 2008), and Sugiaman-Trapman *et al.* found the *RFX8* gene in 2017 (Sugiaman-Trapman *et al.*, 2018). Not only mammals, but only other species have *Rfx* genes: one gene was found in *Saccharomyces cerevisiae*, one in *Caenorhabditis elegans* and three in *Drosophila melanogasta* (Choski *et al.*, 2014, review).

The members of the RFX family share a highly conserved winged-helix DNA binding domain (**FIGURE A5A**) that consist of three alpha helices, three beta strands and three loops (Reith *et al.*, 1989; Reith *et al.*, 1994; Aftab *et al.*, 2008; Gajiwala and Burley, 2010, review). RFX proteins bind to DNA as homo- or heterodimers and they recognize RFX binding sequences called xbox (Reith *et al.*, 1989; Reith *et al.*, 1994). In 2007, Laurençon *et al.* defined a xbox consensus sequence `5'RYNNYYN(1-3)RRNRAC` (with `R=A/G` and `Y=C/T`; Laurençon *et al.*, 2007; **FIGURE A5A**). Apart from the DNA-binding domain, RFX proteins might contain a N-terminal transactivation domain, conserved B and C domains and a dimerization domain (**FIGURE A5A**). The two conserved B and C domains participate together with the dimerization domain in the dimerization of RFX proteins and are called extended dimerization domains (Aftab *et al.*, 2008).

Based on their domains, the RFX proteins can be divided into four groups (**FIGURE A5B**): group 1 contains RFX1, RFX2 and RFX3 that possess all five domains; RFX4, RFX6 and RFX8 I group 2 lack the transactivation domain; and RFX5 and RFX7 in group 7 only have the DNA-binding domain but not the other domains. In addition, Sugiaman-Trapman *et al.* found two new *RFX4* and *RFX8* mRNAs in humans that do not contain the conserved DNA-binding domain sequence. The function of these two isoforms and if they are translated into proteins is unclear. Sugiaman-Trapman *et al.* defined a fourth group for these two *RFX* isoforms that includes RFX proteins without the characteristic DNA-binding domain (Sugiaman-Trapman *et al.*, 2018).

In humans, *RFX1*, *RFX2*, *RFX3*, *RFX5* and *RFX7* are broadly expressed while the expression of *RFX4*, *RFX6* and *RFX8* is more restricted. According to the publicly available FANTOM5 data, the four *RFX* mRNAs *RFX3*, *RFX5*, *RFX6* and *RFX8* could be detected in the pancreas (Sugiaman-Trapman *et al.*, 2018; **FIGURE A5A**). In mice, according to the SAGE data, *mRfx1*, *mRfx2*, *mRfx5* and *mRfx7* show a broad expression pattern with the highest expression in the brain, *mRfx3* in the brain, bladder, placenta, testis and in the gastrointestinal system, *mRfx4* in the brain, testis and pancreas and *mRfx6* in the heart, liver and pancreas (Aftab *et al.*, 2008).

RFX proteins have various functions in mice and humans. Since all of them share the DNA-binding domain, bind to xbox motifs and they might form heterodimers with each other (Reith *et al.*, 1994), elucidating the roles of co-expressed *RFX* proteins in a given tissue is not so evident, especially for the broadly expressed variants *RFX1*, *RFX2*, *RFX3*, *RFX5* and *RFX7* (Sugiaman-Trapman *et al.*, 2018).

The knockout of *mRfx1* in mice causes embryonic lethality at day 2.5 (Feng *et al.*, 2009). Male *mRfx2* knockout mice are sterile because mRFX2 is essential for the maturation of sperm cells and the

formation of the flagellum; the mice do not have any other phenotype and female *mRfx2* mice are fertile and healthy (Wu *et al.,* 2016). Consistently, the conditional knockout of *mRfx1* in murine testis caused infertility due to defects in the formation of the Sertoli cell, the nurse cell for the developing sperm cells (Wang *et al.,* 2016a). The generation of *mRfx3* knockout mice revealed a disturbed left-right asymmetry and growth retardation due to the loss of *mRfx3*. Three-quarter of the *mRfx3* -/- mice died during embryogenesis or at birth (Bonnafe *et al.,* 2004). The authors of this study demonstrated that mRFX3 is essential for the induction of ciliogenic genes in monociliated cells of the embryonic node. These cilia are important for the leftward flow in the embryo and the establishment of the left-right asymmetry in the developing embryo (Bonnafe *et al.,* 2004). In 2003, Blackshear *et al.* unintentionally inserted a transgene in the murine *mRfx4* allele, creating a null allele. The loss of a single or both *mRfx4* alleles severely affected brain development due to an incomplete neural tube closure, and the homozygous knockout mice died before or at birth (Blackshear *et al.,* 2003). This phenotype was confirmed by a conditional knockout of *mRfx4* in the mouse brain in 2018 (Xu *et al.,* 2018). The *mRfx5* -/- mice are immunodeficient due to the deregulation of MHC class II gene expression in the cortex of the thymus, an organ that is essential for the maturation of the immune system T cells, whereas the expression of MHC class II genes was not affected in the thymic medulla (Clausen *et al.,* 1998). The deletion of *mRfx7* specifically in immune cells of mice revealed that mRFX7 is involved in the control of the metabolism of natural killer cells; in the absence of mRFX7, the metabolic rate was higher, and the increased energy consumption shortened the live of natural killer cells (Castro *et al.,* 2018). In summary, RFX proteins are important for the brain development, the immunity and the formation of cilia.

3.1. RFX3 and RFX6 are important for the maturation of pancreatic beta cells

Smith *et al.* detected *mRfx1, mRfx2, mRfx3, mRfx5* and *mRfx6* mRNAs in the murine pancreas and human *RFX6* transcripts in pancreas, small intestine and colon (Smith *et al.,* 2010). So far, only two members of the RFX family, RFX3 and RFX6, have been studied in the developing pancreas (Ait-Lounis *et al.,* 2007; Ait-Lounis *et al.,* 2010; Soyer *et al.,* 2010; Smith *et al.,* 2010; Piccand *et al.,* 2014; Chandra *et al.,* 2014).

Pancreas organogenesis is initiated between embryonic day (E) E8.5 and E9.5 by two evaginations of the gut endoderm. Pancreatic buds later (E11.5-12.5) fuse to give rise to a unique organ that, after expansion and branching, becomes organized in tip and trunk domains characterized by a high rate of cell specification (E13.5). Three cell populations are formed: acinar progenitor cells in the tip domain and ductal and endocrine progenitors in the trunk domain. Endocrine progenitors next delaminate from the epithelium and differentiate into endocrine cells that aggregate and form clusters. These

events are tightly regulated through a cascade of transcriptional regulation. Initial key players are the transcription factors PDX1 and PTF1A that induces endodermal cells to become multipotent pancreatic progenitors. Trunk endocrine progenitors are determined by the transient expression of the proendocrine transcription factor NGN3. The NGN3-postive cells delaminate and are next specified to differentiate into islet subtypes such as alpha and beta cell lineages through the action of ARX and PAX4, respectively. The NGN3 expression is switched off after the delamination; adult islet cells do not express *Ngn3* (**FIGURE A6**; **FIGURE A7A**; Mastracci and Sussel, 2012, review; Cano *et al.,* 2013, review; Shih *et al.,* 2013, review).

In 2007, Ait-Lounis *et al.* performed immunostainings on pancreatic tissues from embryos and adult mice and they detected RFX3 in NGN3-positive cells from E13.5 on and in all mature islet cell types (Ait-Lounis *et al.,* 2007). RFX3 is an essential regulator of ciliogenic genes (Bonnafe *et al.,* 2004); the study of Ait-Lounis *et al.* revealed that RFX3 also directs ciliogenesis in the developing endocrine cells of the pancreas since the cilia were 5-fold shorter in islet cells of *mRfx3 -/-* mice. Moreover, adult *mRfx3 -/-* mice had small disorganized islets, were glucose intolerant due to a reduction of produced and secreted insulin (Ait-Lounis *et al.,* 2007). From E13.5 on, the number of PP-positive cells started to be higher in the knockout than in the wild type, and from E15.5 on, the number cells that stained positive for glucagon, insulin and ghrelin was lower in the islets of *mRfx3 -/-* embryos than in wild-type littermates. The number of somatostatin-expressing cells was not altered by the knockout throughout development. The exocrine part of the pancreas developed normally and NGN3- induction in trunk cells was not affected. At E15.5, when the alteration in endocrine cell type proportion was first detectable, the expression of transcription factors such as *Pdx1*, *Nkx6.1*, *Nkx2.2*, *Neurod1* and *Mafa* directing islet cell specification remained unchanged in the knockout cells. At E17.5 and E19.5, they found that 80 % of the cells were negative for insulin and NGN3 but positive for PDX1 and NKX6.1, two transcription factors that become restricted to the beta cell lineage in the final steps of pancreas organogenesis. The remaining 20 % were insulin-positive but expressed very low levels of the insulin secretion genes *Gck* and *Slc2a2*. They concluded that the beta cells were blocked between the NGN3- progenitor state and the mature state in which they should co-express insulin, PDX1 and NKX6.1 (Ait-Lounis *et al.,* 2007; Ait-Lounis *et al.*, 2010).

In 2010, the same group developed a pancreas-specific knockout mice with one floxed *mRfx3* allele and the Cre recombinase expressed under the control of the *Pdx1* promoter (*mRfx3* fl/-; Pdx1-Cre). The pancreas-specific knockout did not cause any embryonic or postnatal lethality compared to the full knockout, but the mice had the same pancreas phenotype as the *mRfx3 -/-* mice including shorter cilia and impaired glucose tolerance (Ait-Lounis *et al.*, 2010). In conclusion, their data proof that the maturation of the beta cells is blocked at an immature precursor state in the absence of mRFX3.

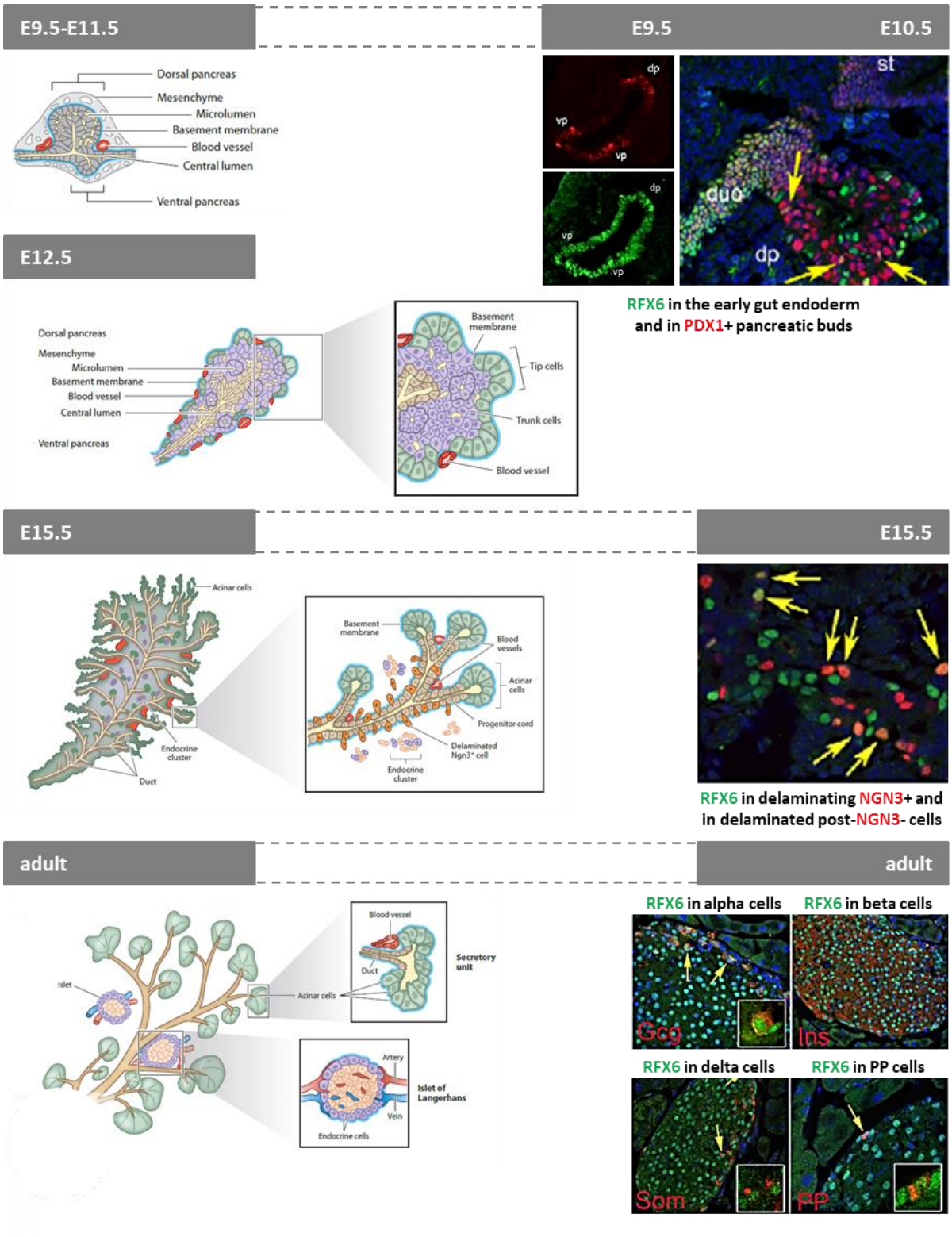


FIGURE A6: RFX6 is expressed in the early gut endoderm, pancreatic progenitors and adult islet cells. Schematic illustrating pancreas organogenesis in the mouse (on the left, figures adapted from Shih *et al.*, 2013, review) and immunostainings depicting RFX6 expression at the respective developmental stages (on the right, figures adapted from Soyer *et al.*, 2010). RFX6 is stained in green, PDX1, NGN3 and the pancreatic hormones in red and nuclei in blue with DAPI. duo = duodenum, vp = ventral pancreas, dp = dorsal pancreas.

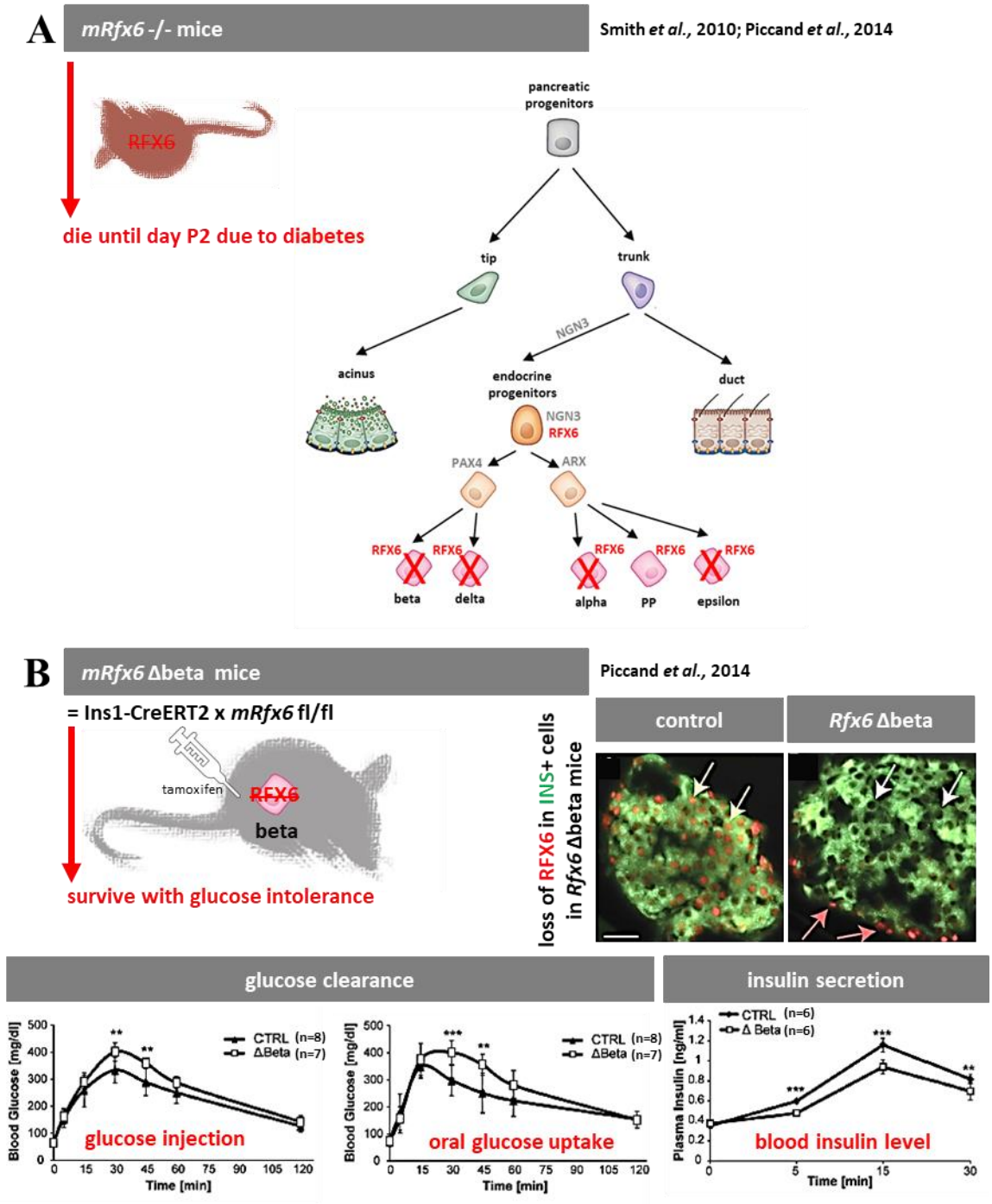


FIGURE A7: The full knockout of *Rfx6* is lethal whereas mice with a beta cell-specific *Rfx6* knockout survive. (A) *Rfx6* -/- mice (Smith *et al.,* 2010; Piccand *et al.,* 2014). The mice lack mature alpha, beta, delta and epsilon cells and die at day 2 or 3 after birth due to diabetes (schematic illustrating endocrine cell differentiation adapted from Shih *et al.,* 2013, review). **(B)** Beta cell-specific *Rfx6* knockout mice (*mRfx6* Δbeta, Piccand *et al.,* 2014). The mice have two floxed *mRfx6* alleles and express the tamoxifen-inducible CreERT2 recombinase under the control of the *Ins1* promoter. Above on the right: Immunostainings demonstrating the loss of RFX6 in insulin-expressing beta cells (RFX6 is stained in red and insulin in green). Below: intraperitoneal glucose tolerance test (IPGTT) after 16 h fasting on the left, oral glucose tolerance test (OGTT) after 16 h fasting in the middle and plasma insulin level on the right performed in wild-type control (*mRfx6* fl/fl) and knockout (*mRfx6* fl/fl; Ins1-CreERT2) mice 1 month (IPGTT and OGTT) or 5 days (insulin monitoring) after tamoxifen injection (figures adapted from Piccand *et al.,* 2014).

The idea that RFX6 might regulate the development and function of the pancreas was the result of a screen for genes enriched in NGN3-positive cells in the developing pancreas. This screen was performed independently in our lab and the lab of Micheal German at the same time (Soyer *et al.,* 2010; Smith *et al.,* 2010).

To stain the NGN3-postive cells, my group created a mouse line that expressed YFP under the control of the *Ngn3* promoter. Due to the stability of YFP, not only NGN3 expressing cells but also post-NGN3 cells are fluorescent (Mellitzer *et al.,* 2004). The group of Michael German used a reporter mice that expressed DsRed-E5 under the control of the *Ngn3* promoter. With time, DsRed-E5 changes its colour from red to green (Miyatzuka *et al.,* 2009). In both screens, *Rfx6* was found to be enriched in the sorted fluorescent cells (Soyer *et al.,* 2010; Smith *et al.,* 2010). Immunostainings and *in situ* hybridization revealed that *Rfx6* is initially expressed in the gut endoderm. After day 9.5, RFX6 expression is temporarily excluded from the pancreas before it can be detected in NGN3-positive endocrine progenitor at day 10.5 and its expression is maintained in the developing islet cells. In the adult, *Rfx6* is expressed in all mature islet cells (**FIGURE A6**; **FIGURE A7A**; Soyer *et al.,* 2010; Smith *et al.,* 2010).

In wholemount in situ hybridization experiments in the zebrafish *Danio rerio* performed in our group (Soyer *et al.,* 2010), *rfx6* was found to be expressed in pancreatic endocrine progenitor cells and in all mature endocrine cell types of the pancreas but not in the gut of the zebrafish. Its expression starts at 17 hours post fertilization (hpf) and is maintained until 72 hpf. The injection of morpholinos against *rfx6* into one- or two cell embryos at 0.75 hpf significantly increased the proportion of pancreatic progenitor cells while the proportion of differentiated cells was significantly reduced at 24 and 30 hpf. Since the total number of cells did not change, it seemed that the cells were blocked in their development at the progenitor stage. No glucagon- and ghrelin- expressing cells and only a few somatostatin-expressing cells could by found among the differentiated cells. The number of insulin-positive cells was only slightly reduced but they were no longer organized in compact islets (Soyer *et al.,* 2010).

The analysis of global *Rfx6* knockout mice also revealed, as already mentioned, a severe pancreas phenotype with the loss of alpha, beta, delta and epsilon cells leading to severe diabetes and postnatal death at day 2 or 3 of life (**FIGURE A7A**; Smith *et al.,* 2010; Piccand *et al.,* 2014). Immunostainings revealed that the endocrine cells were not lost since they stained positive for the endocrine marker chromogranin A and synaptophysin, but that they were negative for glucagon, insulin, somatostatin and ghrelin. The number of pancreatic polypeptide-positive cells was slightly increased. This suggests that the islet cells develop but that alpha, beta, delta and epsilon cells fail to maturate in the absence of RFX6 (Smith *et al.,* 2010; Piccand *et al.,* 2014). In the knockout, the expression of *Ngn3* was

unaffected while the transcript levels of transcription factors downstream of NGN3 such as *Pdx1*, *Pax6*, *Mafa*, *Neurod1*, *Irx* and *Arx* were reduced and *Pax4* expression was increased (Smith *et al.,* 2010).

The next step was the creation of an endocrine-specific *Rfx6* knockout mice called *mRfx6* Δendo (*mRfx6* fl/fl; Ngn3-Cre) in which *Rfx6* is deleted in NGN3-positive endocrine progenitor cells and thus in the endocrine lineage. The mice had a phenotype comparable with the full *mRfx6* -/- mice: no cells expressing glucagon, insulin, somatostatin and ghrelin and more cells expressing pancreatic polypeptide. The expression of *Ngn3* and *Pax4* was increased in *mRfx6* Δendo mice and the expression of *Pdx1*, *Mafa*, *Neurod1* and *Arx* was reduced. The mice died shortly after birth (Piccand *et al.,* 2014).

Taken together, these data and the human phenotype proof that RFX6 is essential for the maturation of islets cells including insulin-expressing beta cells, and that, in its absence, beta cells stay immature leading to severe diabetes. Like RFX6, RFX3 plays an important role in the maturation of beta cells.

3.2. RFX6 is important for the identity and function of adult pancreatic beta cells

In 2014, with the aim to study the role of RFX6 in adult mice, my group developed a conditional beta cell-specific *mRfx6* knockout mouse line (*mRfx6* fl/fl; Ins1-CreERT2) called *mRfx6* Δbeta in which they can induce the deletion of *mRfx6* by tamoxifen injections at any time point in insulin-positive cells (**FIGURE A7B**; Piccand *et al.,* 2014). The mice had mild glucose intolerance and slightly reduced insulin secretion, but, unlike *mRfx6* -/- and *mRfx6*Δendo mice, they survived. In a glucose tolerance test in control and knockout mice that had not eaten 16 hours before the test, injection or oral administration of glucose resulted in a prolonged increase in blood sugar levels in knockout mice compared to control mice. The plasma insulin levels of both control and knockout mice were similar but the glucose-stimulated insulin secretion was affected in *mRfx6* Δbeta mice. The expression of *mIns1* in the mice was reduced, whereas the expression of *mIns2* was not affected. However, since the insulin content and the number of insulin-positive cells were not altered, the phenotype likely resulted rather from a defect within the insulin secretion pathway than from a disturbed insulin production (Piccand *et al.,* 2014).

To elucidate the exact function of RFX6 in mature pancreatic beta cells, Julie Piccand analysed the islets' transcriptome of *mRfx6* Δbeta and control mice by RNA sequencing (**FIGURE A8A**; Piccand *et al.,* 2014) and identified 2,097 upregulated and 987 downregulated genes ($p<0.05$ and $-1>Log2FC>1$; **FIGURE A8A**). In parallel, Perrine Strasser, a PhD student of the lab, performed a ChIP sequencing experiment with an anti-HA antibody in the murine pancreatic beta cell line Min6b1 transfected with a plasmid encoding a 3HA-mRFX6 protein. She found 5,104 peaks corresponding to 3,594 RFX6-bound genes (**FIGURE A8B**; Piccand *et al.,* 2014). The comparison of the RNA and ChIP sequencing data sets

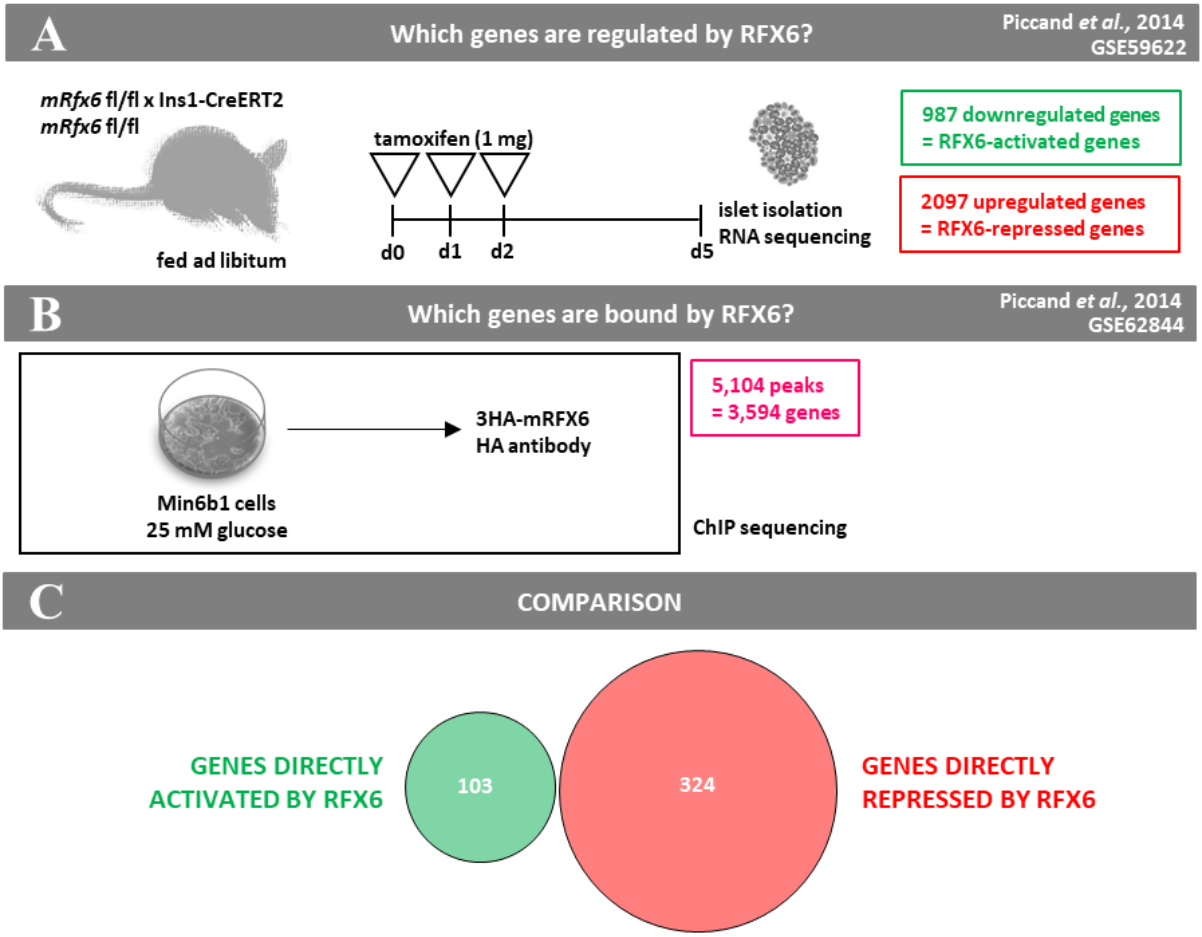


FIGURE A8: Identification of direct RFX6 target genes in adult murine beta cells. (A) RNA sequencing strategy to identify RFX6-regulated genes in mouse islets. In this experiments, RNA was prepared of islets of wild-type control (*mRfx6* fl/fl) and beta cell-specific knockout (*mRfx6* fl/fl; Ins1-CreERT2) mice isolated 5 days after tamoxifen injection. RFX6-regulated genes were identified the comparison of the RNA sequencing of wild-type and knockout samples (GSE59622, Piccand *et al.,* 2014). **(B)** ChIP sequencing (anti-HA ChIP on Min6b1 expressing 3HA-mRFX6, GSE62844, Piccand *et al.,* 2014) performed to identify RFX6-bound genes. **(C)** Diagram showing the proportion of directly upregulated and directly downregulated genes determined by the comparison of the RNA sequencing ad ChIP sequencing data ($p<0.05$ and $-1>Log2FC>1$).

revealed the direct targets of RFX6 (**FIGURE A8C**). It must be said that the number of genes resulting from the peaks and the number of regulated genes varies depending on the threshold for the definition of a peak and the assembly used to annotate the genes; **FIGURE A8C** shows one possible set of numbers for this analysis. From the data, it is obvious that RFX6 inhibits more genes than it induces.

Thanks to the target gene analysis, my group was able to explain the phenotype that it had observed in the mice because the data revealed that the *mRfx6* Δbeta mice had reduced transcript levels of the insulin secretion genes *Gck*, *Abcc8* and several calcium channels, e.g. *Cacna1a*, *Cacna1c*, *Cacna1d* and *Cacnb2* (**FIGURE A9**). Among these genes, *Gck*, *Abcc8*, *Cacna1c* and *Cacnb2* are direct targets of RFX6 and contain xbox elements (**FIGURE A9A**). Interestingly, RFX3 was also found to bind to *Gck* in Min6

cells and Min6 cells treated with siRfx3 had reduced transcript levels of *Ins2*, *Gck* and *Slc2a2*. Basal insulin release in siRfx3-treated Min6 cells were increased but glucose could not stimulate the insulin secretion (Ait-Lounis *et al.,* 2010). Both RFX3 and RFX6 bind to the promoter of *Gck* (Ait-Lounis *et al.,* 2010; Piccand *et al.,* 2014). Together with the findings of Smith *et al.* who had shown by Electrophoretic Mobility Shift Assays (EMSA) in 2010 that RFX3 and RFX6 can bind to DNA as a dimer (Smith *et al.,* 2010), these data suggest that RFX6 and RFX3 might coregulate a subset of their target genes as heterodimers. In conclusion, the knockout of *mRfx6* in adult beta cells reduces the insulin secretion via the reduced activation of insulin secretion genes in the absence of RFX6 (**FIGURE A9B**).

Moreover, the gene expression analysis revealed that RFX6 is important to repress disallowed genes in adult beta cells. In the *mRfx6* Δbeta mice, the transcription of 53 of the 66 disallowed genes (Pullen *et al.,* 2010; Thorrez *et al.,* 2011) was re-activated, 40 of 66 disallowed genes were bound by RFX6 and 32 of the 53 re-expressed genes were direct targets of RFX6 (Piccand *et al.,* 2014), among them the gene *Ldha* (**FIGURE A9A**). The repression of disallowed genes in adult beta cells is necessary to maintain the beta cell identity, maturity and function (Pullen and Rutter, 2013, review). Thus, RFX6 is not only essential for the beta cell maturation but also for the maintenance of the mature state in the adult (**FIGURE A9B**).

In the same year, also in 2014, Chandra *et al.* studied the function of RFX6 in human beta cells. They used their newly generated beta cell line EndoC-betaH2 (Scharfmann *et al.,* 2014). They first showed by immunostaining that RFX6 is expressed in their cell line. The inactivation of *RFX6* by RNA interference caused a reduction of *INS* transcription and transactivation and a decreased insulin content in the EncoC-betaH2 cells. The basal insulin secretion was not affected whereas the knockdown of *RFX6* almost abolished glucose-stimulated insulin secretion. Like in the mouse, this phenotype was due to a reduced expression of key players in the insulin secretion pathway such as *GCK*, *KCNJ11*, *ABCC8*, *CACNA1A*, *CACNA1C*, *CACNA1D*, *CACNA1H* and *CACNB2*. Consistently with the real time quantitative polymerase chain reaction (RT-qPCR) results, calcium imaging experiments revealed that the Ca^{2+}-induced insulin exocytosis was attenuated in the *RFX6* knockdown cells (Chandra *et al.,* 2014). The same experiment performed with isolated islets of our *mRfx6* Δbeta produced similar results.

In conclusion, RFX6 is an essential regulator of beta cell development, maturation and the maintenance of the mature state, and it is also key for insulin secretion in both mouse and human since it activates the expression of insulin and insulin secretion genes (Soyer *et al.,* 2010; Smith *et al.,* 2010; Piccand *et al.,* 2014; Chandra *et al.,* 2014).

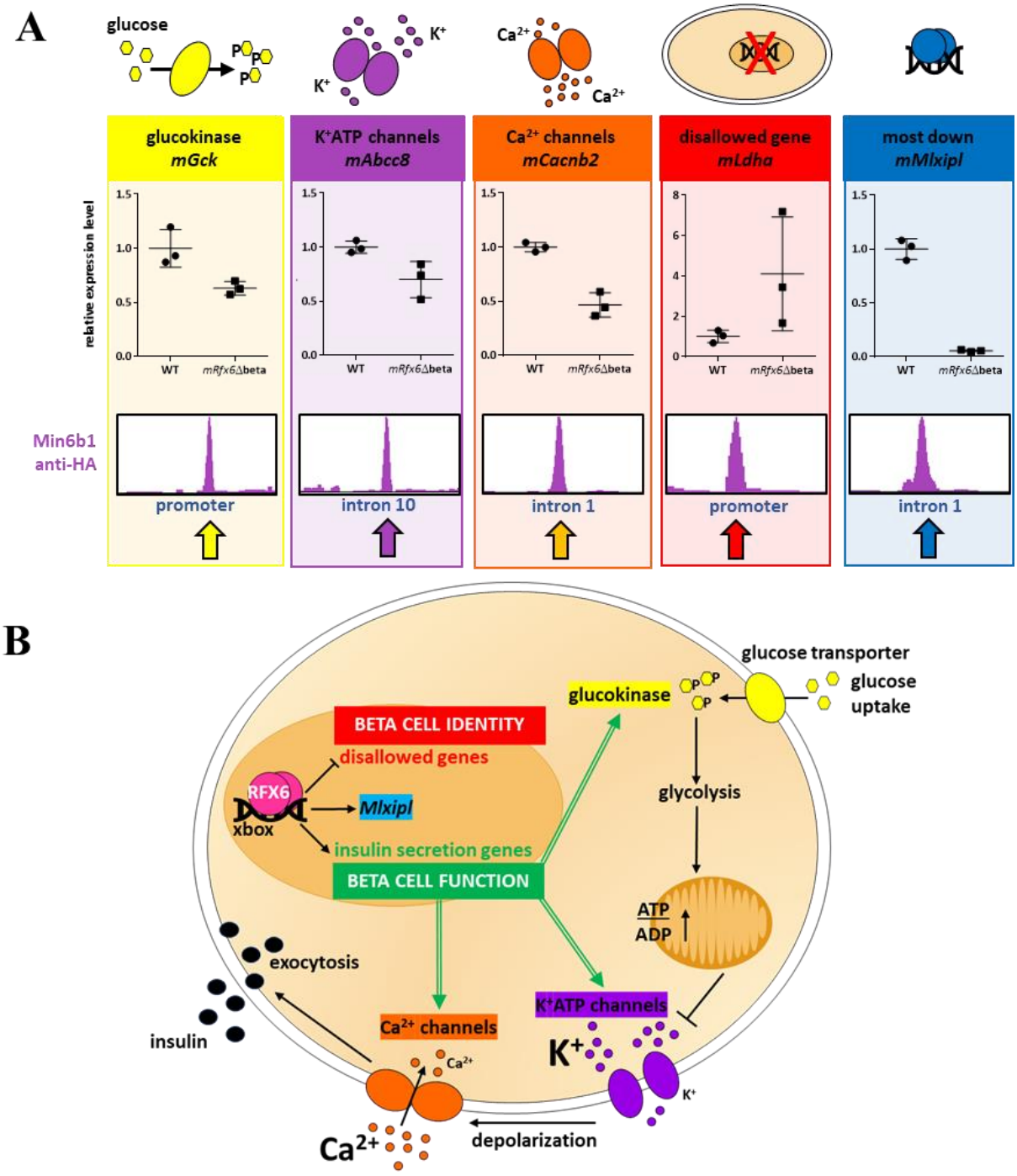


FIGURE A9: RFX6 is essential for beta cell function and identity. (A) Direct RFX6 target genes in mice (Piccand *et al.*, 2014). Above: Transcript levels of *mGck*, *mAbcc8*, *mCacnb2*, *mLdha* and *mMlxipl* measured by RNA sequencing of RNA isolated from islets of wild-type control (*mRfx6* fl/fl) and knockout (*mRfx6* fl/fl; Ins1-CreERT2) mice 5 days after tamoxifen injection (GSE59622, Piccand *et al.*, 2014). Below: ChIP sequencing data showing the binding of RFX6 on *mGck*, *mAbcc8*, *mCacnb2*, *mLdha* and *mMlxipl* (anti-HA ChIP on Min6b1 expressing 3HA-mRFX6, mm9, GSE62844, published in Piccand *et al.*, 2014). **(B)** Schematic illustrating RFX6 function in adult pancreatic beta cells (figure adapted from Piccand *et al.*, 2014).

3.3. The gene *mMlxipl* is the most downregulated gene in the *mRfx6* Δbeta mice

As mentioned, RFX6 is rather a transcriptional repressor than a transcriptional activator (**FIGURE A8C**). Nonetheless, the gene most affected by the beta cell-specific *mRfx6* knockout in mice was one of the downregulated genes, called *Mlxipl*. Its expression was 19-fold reduced in the knockout islets (**FIGURE A9A**); we detected 1,595 normalized reads in the wild type and only 84 normalized reads in the knockout (GSE59622, Piccand *et al.,* 2014). Moreover, the gene *Mlxipl* is not only regulated by RFX6 but also bound by RFX6. In the anti-HA ChIP sequencing experiment on Min6b1 cells expressing 3HA-mRFX6 (GSE62844, Piccand *et al.,* 2014), we found a predominant peak of RFX6 in the first intron of *Mlxipl* (**FIGURE A9A**), suggesting that *Mlxipl* is a direct target of RFX6.

4. MLXIPL is a metabolite-activated basic helix-loop-helix transcription factor

The gene *Mlxipl* was identified in 2001 as one of the genes that are deleted in the Williams-Beuren syndrome (OMIM #194050), a multi-symptomatic developmental disorder affecting the heart, brain, skin and other organs. In this syndrome, a 1.5 to 1.8 Mb long region is deleted on chromosome 7, arm q, region 11.23 (7q11.23). The hemizygous deletion causes the loss of one allele of approximately 28 genes. One of these genes is *WBSCR14* alias *MLXIPL* (Cairo *et al.,* 2001). Cairo *et al.* analysed the expression of *Mlxipl* in mouse embryos by *in situ* hybridization. In mice, *Mlxipl* is expressed in the spinal cord at E12.5, and at E14.5 and E16.5 in several tissues, including the nervous system, lung, liver, kidney, intestine, thymus and heart. They further showed that human *MLXIPL* is expressed in the brain and in the intestine (Cairo *et al.,* 2001). By northern blotting, Yamashita *et al.* detected MLXIPL in the liver, kidney and small intestine of adult rats (Yamashita *et al.,* 2001) and Iizuka *et al.* found the highest expression of *Mlxipl* in the intestine, liver, adipose tissues, skeletal muscle and kidney in adult mice (Iizuka *et al.,* 2004). In 2002, Wang and Wollheim revealed that *Mlxipl* is also expressed in rat pancreatic islets and in the rat beta cell line Ins-1 (Wang and Wollheim, 2002).

Because of its different functional domains, MLXIPL has three common names (**FIGURE A10A**): MLXIPL is called MLX-Interacting Protein-Like because it interacts with MLX via its basic helix-loop-helix leucine zipper and the leucine zipper-like domain to form a dimer (Cairo *et al.,* 2001). Cairo *et al.* defined the genomic DNA and coding DNA sequence of *MLXIPL* and classified it as basic helix-loop-helix leucine zipper (bHLH/LZ) transcription factor that is a dimerization partner of the bHLH/LZ transcription factor MLX (Cairo *et al.,* 2001), a member of the Myc/Mad/Max family that is involved in the transcriptional regulation of proliferation, differentiation and apoptosis (Grandori *et al.,* 2000, review). Later, it has been shown that MLXIPL can not only interact with MLX but also with other proteins (Richards *et al.,* 2017, review), such as FLII (Wu *et al.,* 2013), HNF4A (Adamson *et al.,* 2006), FXR that binds to the MLXIPL-HNF4A complex (Caron *et al.,* 2013), and PGC-1β (Chambers *et al.,* 2013).

MLXIPL is called Carbohydrate Response Element Binding Protein (ChREBP) because it binds to so-called Carbohydrate Response Elements (ChoRE), which consist of two E-boxes separated by five nucleotides (Yamashita *et al.,* 2001; **FIGURE A10A**). The bHLH/LZ proteins bind to E-box motifs with the consensus sequence 5'CACGTG. In 1995, Shih *et al.* had observed that glucose-regulated genes in the liver, for instance the gene *Pklr* encoding the liver pyruvate kinase that catalyzes the conversion of phosphoenolpyruvate to pyruvate and ATP, contain two close E-box motifs separated by five base pairs (Shih *et al.,* 1995). In 2001, Yamashita *et al.* used the *Pklr* ChoRE to identify proteins that bind to this motif. For this experiment, they carried out this experiment with nuclear extracts from liver tissues of rats that were on high-fat or carbohydrate-rich diet. Among the ChoRE-bound proteins, they found that the transcription factor MLXIPL interacted with the ChoRE in the liver nuclear extracts from rats fed with the high-carbohydrate diet and transactivated the *Pklr* gene via the ChoRE in a hepatocyte cell line in response to glucose (Yamashita *et al.,* 2001). ChoREs have the consensus sequence 5'CACGTGNNNNNCACGTG. The sequence of the two E-boxes at the left and right site is not necessarily stringent while the interspace between both E-boxes is critical for MLXIPL binding (Yamashita *et al.,* 2001). Ma *et al.* showed that two MLXIPL and MLX dimers bind simultaneously to the two E-boxes of one ChoRE (Ma *et al.,* 2006).

The third name, MondoB, was given to MLXIPL because it contains Mondo-Conserved Regions (MCR) in its N-terminal region that is shares its paralogue, the bHLH/LZ protein MondoA, also called MLXIP for MLX-Interacting Protein (Billin *et al.,* 2000; **FIGURE A10A**). The N-terminus of MLXIPL harbours not only the five MCR region I, II, III, IV and V but also two Nuclear Export Signals (NES), one Nuclear Localization Signal (NLS) and the Glucose-Sensing Module (GSM) that is composed of a Low-glucose Inhibitory Domain (LID) and a Glucose-Response Activation Conserved Element (GRACE). In combination with several post-translational modifications, these domains control the transcriptional activity of MLXIPL (Li *et al.,* 2006; Davies *et al.,* 2008; Davies *et al.,* 2010).

4.1. MLXIPL activity increases with the availability of nutrients

The mechanism of the regulation of MLXIPL protein activity is complex and was reviewed in multiple articles together with its function in different tissues (Dentin *et al.,* 2005, review; Postic *et al.,* 2007, review; Iizuka and Horikawa, 2008, review; Poupeau and Postic, 2011, review; Havula and Hietakangas, 2012, review; Leclerec *et al.,* 2012, review; Filhoulaud *et al.,* 2013, review; Iizuka *et al.,* 2013, review; Abdul-Wahed *et al.,* 2017, review; Richards *et al.,* 2017, review; Jois and Sleeman, 2017, review; Iizuka, 2017, review; Havula and Hietakangas, 2018, review; ...). We will give here only a summary of the regulation of MLXIPL protein activity (**FIGURE A10B**) and refer to the reviews for more information.

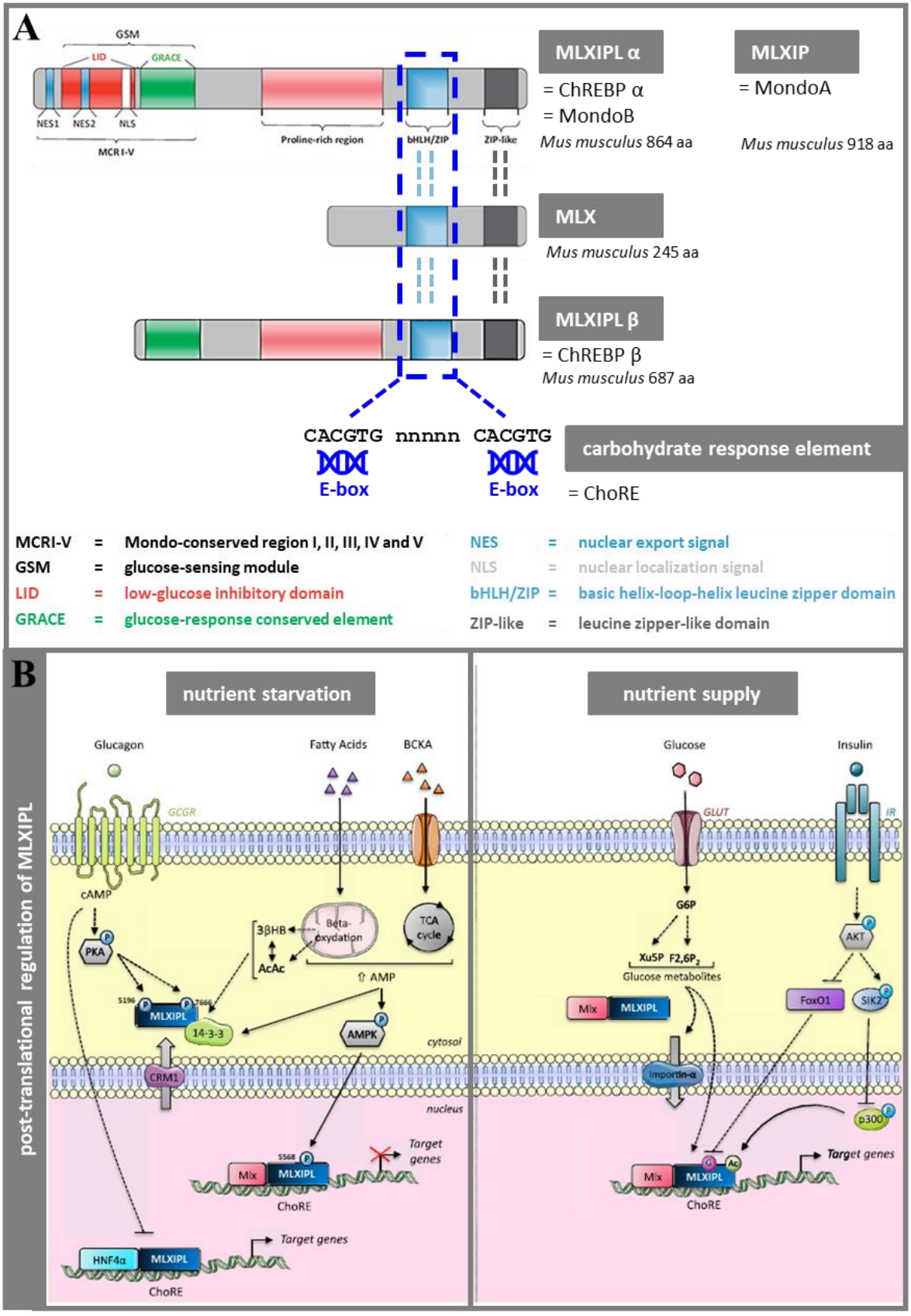


FIGURE A10: MLXIPL is a nutrient-activated basic helix-loop-helix transcription factor. (A) Structure of the basic helix-loop-helix leucine zipper transcription factors MLXIPL α and β, MLXIP and MLX (figure adapted from Filhoulaud *et al.,* 2013, review). MLXIPL forms a dimer with MLX and bind to carbohydrate response elements (ChoRE). **(B)** Regulation of MLXIPL activity (figure adapted from Richards *et al.,* 2017, review).

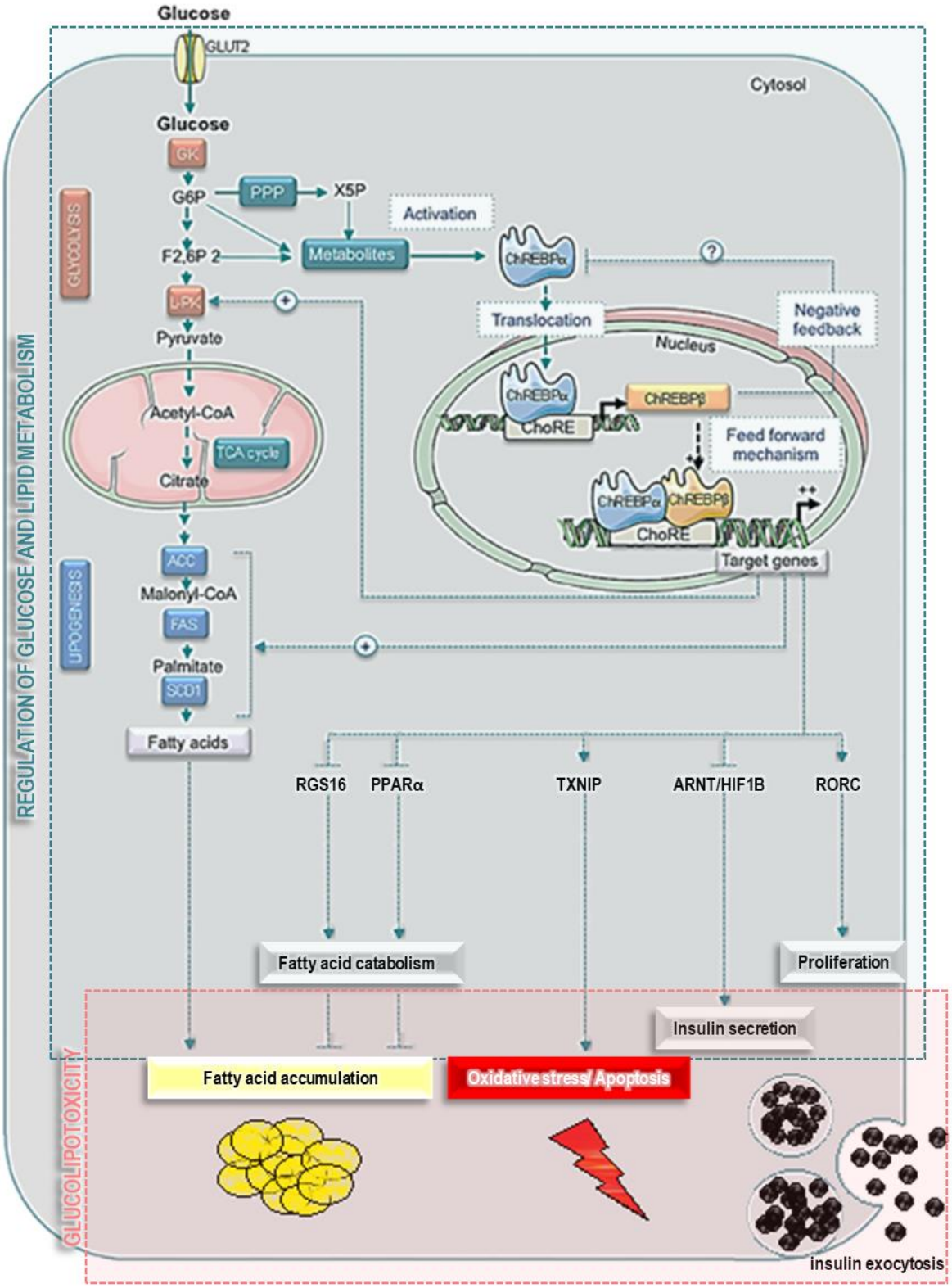


FIGURE A11: MLXIPL regulates glycolytic and lipogenic genes and mediates glucolipotoxicity. Schematic illustrating MLXIPL function in pancreatic beta cells (figure adapted from Abdul-Wahed *et al.,* 2017, review and Filhoulaud *et al.,* 2013, review). After its activation, MLXIPL induces the expression of glycolytic and lipogenic genes promoting both glycolysis and lipogenesis. The repression of RGS16 and PPARα further inhibits fatty acid catabolism. The inhibition of ARNT/HIF1B leads to decreased insulin secretion. Via RORC, MLXIPL activates the expression of cell cycle genes. In chronic hyperglycaemia, MLXIPL contributes to the glucolipotoxicity: the repression of RGS16 and PPRA α and the activation of lipogenic genes causes the accumulation of fatty acids; the activation of TXNIP induces oxidative stress and apoptosis.

One should remember above all that MLXIPL is activated by increased nutrient availability, while it is inactive at lower nutrient levels. The NLS is important for MLXIPL import into the nucleus via the nuclear import factor IMPORTINα and the NES mediate the interaction of MLXIPL with the nuclear export factor CRM1. During starvation periods, MLXIPL is retained in the cytoplasm. The NLS of MLXIPL is phosphorylated at serine 196 and threonine 666 by the kinase PKA that is activated through glucagon binding to its receptor and downstream cAMP signalling. In the cytoplasm, MLXIPL forms a complex with the 14-3-3 protein. A raise in the AMP level due to increased fatty acid and amino acid metabolism causes the retention of the MLXIPL/14-3-3 protein complex in the cytoplasm. The activation of the kinase AMPK by AMP leads to the phosphorylation of nuclear MLXIPL at serine 568, thus inhibiting its transcriptional activity. Metabolites of the beta oxidation inhibit the interaction of MLXIPL with IMPORTINα. When glucose becomes available, the glucose metabolites glucose-6-phosphate (G6P), xylulose-5-phosphate (Xu5P) and fructose-2,6-bisphosphate (F2,6P2) stimulate the activation of MLXIPL. Xu5P activates the phosphatase PP2A that dephosphorylates serine 196 and 568 and threonine 666, thus promoting the interaction of MLXIPL with IMPORTINα and the import of MLXIPL into the nucleus where it dimerizes with MLX and binds to ChoREs. G6P is supposed to bind directly to the GSM of MLXIPL and to induce a conformational change of the protein and its activation. Further post-translational modifications of MLXIPL include its acetylation by the histone acetyl transferase p300 and the O-linked N-acetyl-glycosylation by OGT that both increase the transcriptional activity of MLXIPL. The consequences of insulin signalling on MLXIPL activity are controversial since it inhibits its acetylation and promotes its O-linked N-acetyl-glycosylation (Filhoulaud *et al.,* 2013, review; Richards *et al.,* 2017, review; **FIGURE A10B**). In pancreatic beta cells, MLXIPL is bound to SORCIN in the cytoplasm. The influx of Ca^{2+}, a step of the insulin secretion pathway, triggers the release of MLXIPL and its translocation into the nucleus (Noordeen *et al.,* 2012).

In conclusion, MLXIPL gets activated in high glucose conditions and is a glucose-responsive transcription factor. The importance of the individual post-translational modifications and their *in vivo* significance have not yet been fully elucidated.

In 2006, Li *et al.* demonstrated that an N-terminal truncated MLXIPL protein without the LID domain (amino acid residues 1 to 196) is constitutively active in both low and high glucose conditions (Li *et al.,* 2006; Li *et al.,* 2008); it is therefore named CA-MLXIPL. In 2012, Herman *et al.* identified a natural isoform in murine adipose tissues that lacks the residues 1 to 177, thus corresponding to CA-MLXIPL. From its discovery on, the longer isoform is called MLXIPL α and the shorter isoform MLXIPL β (Herman *et al.,* 2012; **FIGURE A10A**). The transcription of *Mlxipl* β is initiated at exon 1b, that is localized 17 kb upstream of exon 1a, formerly known as exon 1 of *Mlxipl* α. Exon 1b contains two ChoREs and *Mlxipl* β transcription is induced by MLXIPL α in high glucose and autoregulated by MLXIPLβ after its production (Herman *et al.,* 2012; Zhang *et al.,* 2015). MLXIPL β is shorter than MLXIPL α because exon

1b does not contain an ATG and, in the splicing, exon 1b is linked to exon 2 bypassing exon 1a, and the first downstream ATG is in exon 4 (Herman *et al.,* 2012).

While the transcriptional regulation of *Mlxipl* β was largely clarified shortly after its discovery, the transcriptional regulation of *Mlxipl* α is still a mystery. So far, one only knows that *Mlxipl* α is expressed independently of the nutrient availability. The study by Herman *et al.* aimed to uncover the transcriptional regulation of *Mlxipl* α, as it was observed before the discovery of *Mlxipl* β that the *Mlxipl* transcript level increases to high glucose (Herman *et al.,* 2012). All the studies on *Mlxipl* transcriptional regulation have been performed before the discovery of *Mlxipl* β. In the liver, it has been shown that OCT-1 represses the transcription of *Mlxipl* in human hepatocytes via a POU binding site in the promoter region of *Mlxipl* upstream of exon 1a. Insulin lowers the repressive effect of OCT-1 (Sirek *et al.,* 2009). The activation of the nuclear factors LXR and RXR by their ligands induced their binding as an LXR/RXR heterodimer to two LXREs in the promoter of *Mlxipl* in the murine liver and the activation of its expression (Cha and Repa, 2007). Similarly, the thyroid hormone receptor (THR), also a nuclear factor, activates the transcription of *Mlxipl* by interacting with the LXREs as a dimer with RXR (Hashimoto *et al.,* 2009). Moreover, the rat *Mlxipl* promoter upstream of exon 1a contains one binding site for SREBP-1c, two for SP1 and two for NF-Y but each protein alone did not activate a reporter construct with these binding sites. The co-expression of SREBP-1c with SP1 or NF-Y mildly activated the construct (Satoh *et al.,* 2007). However, an important role of SREBP-1c in *Mlxipl* regulation seems unlikely since *Mlxipl* is not reduced in *Srebp-1c* knockout mice (Poupeau and Postic, 2011, review). In pancreatic beta cells, *Mlxipl* expression is known to be regulated by FOXA1 and FOXA2, which have been found to bind two conserved regions in the first intron of *Mlxipl* downstream of exon 1a. Knockout of both transcription factors causes a strong downregulation of *Mlxipl* expression (Gao *et al.,* 2010).

Taken together, MLXIPL has two major isoforms, MLXIPL α and MLXIPL β. MLXIPL α is constitutively expressed but its activity is regulated by nutrients. MLXIPL β is only produced in high glucose but its activity is constitutive and does not rely on nutrients. When we talk about MLXIPL in the following, we mean both isoforms together.

4.2. MLXIPL is not essential for embryogenesis since *Mlxipl* knockout mice develop normally

Not only the regulation of MLXIPL but also its function has been extensively reviewed in the same articles as mentioned above. The knockout of *Mlxipl* in mice leads to a relatively mild phenotype: The mice survive and have a normal appearance and body weight compared to wild-type littermates, and the knockout does not cause any embryonic lethality. The findings of Iizuka *et al.* do not support any role of MLXIPL in the development (Iizuka *et al.,* 2004).

Nonetheless, the *in situ* hybridization experiments of Cairo *et al.* showed that *Mlxipl* is expressed during development in different tissues (Cairo *et al.,* 2001). Soggia *et al.* demonstrated that the expression of *Mlxipl* in the embryonic pancreas of rats increases from E13.5 to E18.5 and the deletion of *Ngn3* causes the loss of *Mlxipl* expression in E18.5 pancreatic tissues, suggesting that *Mlxipl* is expressed in the endocrine lineage. This finding is in line with observations of my host lab that *Mlxipl* is 75-fold enriched in NGN3 descendants; this was the same screen for genes expressed in NGN3-positive cells and their descendants in which the team identified *Rfx6* (Soyer *et al.,* 2010). Soggia *et al.* used an *in vitro* protocol to differentiate rat pancreatic precursors into endocrine and exocrine cells, and they reported that MLXIPL enhances beta cell differentiation to external metabolic stimuli such as glucose and xylitol and that *Mlxipl* inhibitions decreased beta cell differentiation (Soggia *et al.,* 2012). However, an exact role of MLXIPL in pancreas organogenesis and beta cell differentiation is unknown. In their review, Filhoulaud *et al.* asked the question whether MLXIPL might only serve as energy provider during development (Filhoulaud *et al.,* 2013, review).

4.3. MLXIPL is a key regulator of glycolytic and lipogenic genes

In the adult, the function of MLXIPL has been studied most in the liver and in adipose tissues. In the liver, hyperglycaemia promotes the conversion of glucose into glycogen as well as fatty acids and their storage as triglycerides. To increase the metabolic rate, enzymes of the glycolysis and lipogenesis are activated by post-translational modifications, and the encoding genes are upregulated. One important transcription factor that induces the transcription of glycolytic and lipogenic genes in response to insulin signalling is SREBP-1c but the knockout of *Srebp-1c* in mice only results in the halving of fatty acid synthesis in the liver. Thus, several studies postulated that there must be another mechanism in the regulation of lipogenic genes in the liver that is independent of SREBP-1c (Dentin *et al.,* 2005, review).

In 2001, when MLXIPL was discovered, Yamashita already demonstrated that MLXIPL bind to the promoter of the metabolic gene *Pklr* (Yamashita *et al.,* 2001), and, *inter alia,* the study of siMlxipl-treated primary hepatocytes (Dentin *et al.,* 2004) and the analysis of *mMlxipl* -/- mice (Iizuka *et al.,* 2004) revealed that MLXIPL is key to control the glucose-induction of glycolytic and lipogenic gene in the liver, since, in its absence, the expression of several metabolic genes, e.g. *Pklr*, *Acaca*, *Fasn*, *Scd1* and *Elovl6*. Consequently, the liver fatty acid synthesis was reduced in the *mMlxipl* -/- mice; the glycolysis, however, was only inhibited in its penultimate step in the conversion of phosphoenolpyruvate to pyruvate and ATP by the reduction of pyruvate kinase (*Pklr*) and the amount of stored glycogen in the liver was increased in the *mMlxipl* -/- mice. The knockout mice had smaller

adipose tissues and a reduced level of free fatty acids in the blood, are sucrose-, fructose- and glucose-intolerant and have a reduced insulin sensitivity (Iizuka *et al.,* 2004).

An increase of the fatty acid synthesis over time leads to a pathology called Non-Alcoholic Fatty Liver Disease (NAFLD) or hepatic steatosis and is characterized by the accumulation of triglycerides in the cytoplasm of hepatocytes. The contribution of MLXIPL in insulin sensitivity or resistance complex because it involves multiple organs (liver, adipose tissues, skeletal muscle, pancreas, brain) and is a multisystemic effect (Filhoulaud *et al.,* 2013, review). The knockout of *Mlxipl* in obese mice and the liver-specific deletion of *Mlxipl* in obese mice diminishes hepatic steatosis and insulin resistance (Iizuka *et al.,* 2006; Dentin *et al.,* 2006). The overexpression in the murine liver improves insulin sensitivity in mice on high-fat diet due to the increased production of beneficial lipids (Benhamed *et al.,* 2012). By contrast, a recent study reported that the liver-specific knockout of *Mlxipl* led to insulin resistance regardless of the diet (Jois *et al.,* 2017). Taken together, the causative link between MLXIPL action and insulin sensitivity still needs further investigations.

4.4. MLXIPL represses insulin secretion and stimulates beta cell proliferation

Like in the liver, MLXIPL also regulates the expression of the glycolytic and lipogenic genes *Pklr*, *Fasn* and *Gpdh* in pancreatic beta cells, resulting in increased lipogenesis (Da Silva Xavier *et al.,* 2006; Boergesen *et al.,* 2011; Poungvarin *et al.,* 2012). The expression of a CA-MLXIPL in Ins-1 832/13 cells increased the expression of genes involved in the glycolysis and lipogenesis such as *Eno1*, *Pklr*, *Mdh1*, *Me1*, *Pdha1*, *Acly*, *Acaca*, *Fasn*, *Elovl6* and *Gpd1,* and the expression of a dominant negative MLXIPL isoform represses these genes (Sae-Lee *et al.,* 2016). Moreover, MLXIPL binds to and represses the transcription of *Pparα* and *Rgs16*. Both proteins promote the catabolism of lipids, thus their inhibition by MLXIPL increases the anabolism and decreases the catabolism of lipids in pancreatic beta cells (Boergesen *et al.,* 2011; Sae-Lee *et al.,* 2016; **FIGURE A11**).

The inhibition of *Mlxipl* by RNA interference increased the glucose-stimulated insulin secretion in Min6 cells (Da Silva Xavier *et al.,* 2006). The expression of CA-MLXIPL in Ins-1 832/13 cells inhibits the transcription of *Ins2* and completely abolishes insulin secretion in Ins-1 832/13 cells (Poungvarin *et al.,* 2012). In addition, MLXIPL directly represses *Arnt* (*Hif1b*) in Ins-1 832/13 cells and in mouse islets in low and high glucose (Noordeen *et al.,* 2010) and indirectly inhibits the transcription of *Pdx1* in Min6 cells in low glucose (Da Silva Xavier *et al.,* 2010), both important stimulators of insulin secretion (**FIGURE A11**). Taken together, MLXIPL is an inhibitor of insulin secretion from pancreatic beta cells.

In chronic hyperglycaemia, the prolonged activation of MLXIPL causes an increased fatty acid production and the accumulation of lipids in the cytoplasm of beta cells, the key phenotype of the above-mentioned glucolipotoxicity (Poungvarin *et al.,* 2012; Sae-Lee *et al.,* 2016; Boergesen *et al.,*

2011). Multiple studies suggest that MLXIPL also activates the transcription of *Txnip*, a key mediator of oxidative stress and apoptosis in beta cells (Shalev, 2014, review; **FIGURE A11**). Minn *et al.* identified a ChoRE in the promoter of *Txnip* and observed that *Txnip* induction is glucose-dependent (Minn *et al.,* 2005). Cha-Molstad *et al.* demonstrated that MLXIPL binds to the ChoRE in the *Txnip* promoter in Ins-1 cells, and that the binding is enhanced by glucose (Cha-Molstad *et al.,* 2009). The deletion of one E-box of this ChoRE abolished *Txnip* transactivation by MLXIPL (Poungvarin *et al.,* 2012). Kibbe *et al.* showed that the transcription factors FOXO1 and MLXIPL have close binding sites on the *Txnip* promoter and that they compete for the binding in human islets and Ins-1 cells (Kibbe *et al.,* 2013). The expression of CA-MLXIPL in Ins-1 832/13 cells increased *Txnip* transcription (Poungvarin *et al.,* 2012; Sae-Lee *et al.,* 2016), while the expression of a dominant negative MLXIPL in Ins-1 832/13 cells or the treatment of Ins-1 cells with an siMlxipl resulted in a decreased induction of *Txnip* in high glucose (Sae-Lee *et al.,* 2016; Cha-Molstad *et al.,* 2009). By contrast, an siRNA against *Mlxipl* β did not affect Txnip expression in Ins-1 832/13 cells, suggesting that *Txnip* is not a target of MLXIPL β (Zhang *et al.,* 2015). Furthermore, the expression of *Mlxipl* and *Txnip* is increased in islets of humans with diabetes and in murine islets after 48-h in high glucose (Poungvarin *et al.,* 2012). The expression of CA-MLXIPL in murine islets *in vivo* resulted in a disorganized islet structure, *Txnip* upregulation, oxidative stress and increased beta cell apoptosis when analysed 6 weeks after the virus transduction (Poungvarin *et al.,* 2012). Thus, MLXIPL seems to be an important cause of glucotoxicity and glucolipotoxicity, oxidative stress and apoptosis in beta cells.

By contrast, in 2012, Metukuri *et al.* demonstrated that MLXIPL is necessary for glucose-stimulated beta cell proliferation. The inhibition of *Mlxipl* by RNA interference in rat islets, human islets and Ins-1 832/13 cells diminished beta cell proliferation after a glucose stimulus. Even more convincing is the fact that they did not observe any cell division in islets of *mMlxipl* -/- mice. Consistently, the overexpression of MLXIPL in Ins-1 832/13 and rat islets promoted glucose-induced proliferation. Looking for the cause, they showed that the knockdown of *rMlxipl* in Ins-1 832/13 cells attenuated the glucose-induction of cell cycle regulators after 16 h in high glucose. The mRNA levels of *Ccnd2*, *Ccna2*, *Ccne1*, *Cdk1* and *Cdk2* and the protein amount of CCND2, CCNA, CDK4, CCNE and CDK6 was reduced in the knockdown cells compared to the control cells (Metukuri *et al.,* 2012). Considering the existence of the smaller isoform beta after its discovery in 2012, Zhang *et al.* proofed that the inhibition of this isoform with a specific siRNA significantly decreased glucose-induced beta cell proliferation in rat islets (Zhang *et al.,* 2015).

In 2016, Schmidt *et al.* revealed that the glucose response is biphasic in the rat beta cell line Ins-1E: the cells induce insulin secretion genes in the first wave of transcription early after the glucose stimulus and repress insulin secretion genes as well as beta cell genes in the later second wave while cell cycle regulators are induced. They investigated the role of MLXIPL in the two waves via its overexpression

and its downregulation in Ins-1E cells by transcriptome studies. The expression of CA-MLXIPL activated the insulin secretion and cell cycle genes that are induced after a glucose stimulus while the siMlxipl treatment inhibited their glucose-mediated induction. The identification of its direct target genes by ChIP sequencing on Ins-1E cells incubated for 0 h, 2 h or 12 h in high glucose and by RNA sequencing of siMlxipl-treated cells after 0 h and 12 h in high glucose showed that MLXIPL directly activates the expression of insulin secretion genes in the first wave of transcription while the induction of cell cycle regulators relays on the MLXIPL downstream target RORC (**FIGURE A11**). Consistently, the treatment of Ins-1E cells with siRorc decreased Ins-1E proliferation (Schmidt *et al.,* 2016).

In conclusion, as proposed since its discovery, MLXIPL is important for the glucose-induction of beta cell genes and the mediation of the glucose response in beta cells. It is implicated in both, beta cell survival and death (**FIGURE A11**). The study of Schmidt *et al.* suggests that the consequences of MLXIPL activation vary depending on the timing, with an early induction of insulin secretion and a subsequent inhibition of insulin secretion and the induction of beta cell proliferation (Schmidt *et al.,* 2016). Considering the aspect of timing, these data agree with the previous findings on MLXIPL's role in insulin secretion and beta cell division and clearly indicate that the duration of the glucose stimulus must not be disregarded in the evaluation of the data.

5. Research question

Previous, unpublished, findings from my host lab, suggest that the glucose-responsive transcription factor MLXIPL is a direct target gene of RFX6. Therefore, the first objective of my thesis was to study the regulation of *Mlxipl* expression by RFX6 by characterizing more precisely the enhancer activity of the *mMlxipl* intron 1 region. Our plans were to identify the RFX6-bound motifs (xbox) which are critical for *Mlxipl* α regulation by RFX6 and determine whether other transcription factors act in synergy to regulate *Mlxipl* α via intron 1 (**research question - FIGURE A12**). To address these questions, I used classical transactivation assays to investigate the interaction of different transcription factors including RFX6 with the *mMlxipl* intron 1 region and their role in *Mlxipl* transactivation.

Secondly, my plan was to determine the common and specific targets of RFX6 and MLXIPL in rodent beta cells. Indeed, given that *Mlxipl* is barely expressed in mouse beta cells lacking *Rfx6*, our working hypothesis was that a part of the deregulated genes in the *mRfx6* Δbeta mice must be genes that are normally regulated by MLXIPL (**research question - FIGURE A12**). Thus, identifying the respective targets of both genes should shed some light on their common and specific roles in beta cells. Moreover, considering that glucose plays a pivotal role in beta cell physiology and that MLXIPL is regulated by glucose, we planned to consider the influence of glucose level on RFX6- and MLXIPL-dependent transcriptomes. In this second part of the thesis, we used rodent beta cell lines as a model

system because they are easy to manipulate in defined glucose concentrations, and we considered RNA interference and CRISPR/Cas9 approaches to inactivate *Rfx6* and *Mlxipl* in these cells. To compare the genes differentially regulated in *Rfx6*-deficient, *Mlxipl*-deficient and wild-type cells at low and high glucose concentrations, we decided to sequence the transcriptomes to have a global view on the gene expression changes mediated by RFX6 and MLXIPL in low and high glucose.

RESEARCH QUESTIONs

PART 1 – transcriptional regulation of *Mlxipl α* by RFX6 in adult beta cells

PART 2 – genetic programs regulated by RFX6 and MLXIPL in adult beta cells

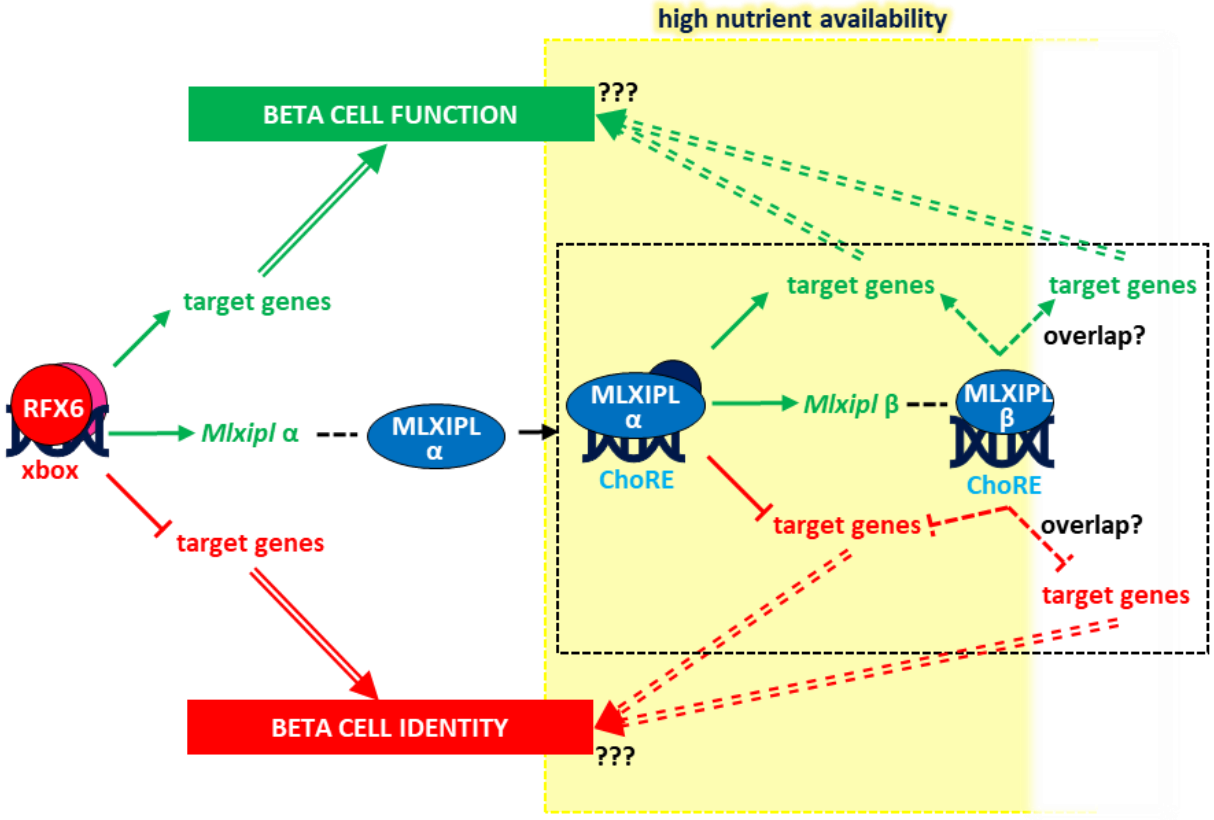


FIGURE A12: Research questions. This thesis deals with two related questions: The first is about the transcriptional regulation of *Mlxipl* α by RFX6 and the underlying mechanism behind it, which should be clarified. The ChIP experiment had shown that RFX6 binds to *Mlxipl* intron 1. This is the experimental starting point of the first part of the project. Secondly, the role of MLXIPL downstream of RFX6 in adult beta cells will be clarified by defining the genetic programs that are dependent of either one or both transcriptional factors.

B) Material and Methods

1. Materials

The respective provider of the used instruments, solutions, kits, enzymes, plasmids, antibodies, bacteria strains and cell lines is specified in the following parts. Unless otherwise stated, chemicals were purchased from Sigma-Aldrich, Merck or Thermo Fisher Scientific. Tissue culture treated flasks, plates and dishes were ordered from Corning or Falcon. Cell culture medium and solutions were prepared by the cell culture facility of the institute. They also take care about the freezing and thawing of cells as wells as the storage of frozen cell lines. Buffers and media and plates for bacteria cultures were prepared by the institute's media preparation service. The origin of general lab equipment (e.g. freezers, fridges, incubators, laminar flow hoods, optical microscopes, shakers, vortex, centrifuges, heat blocks, Bunsen burners, pipet aids and graduated pipettes, automatic pipettes and filter tips, tubes, petri dishes, glass ware) is not further specified.

2. Molecular biology methods

2.1. Genomic DNA extraction from adherent cell lines

For phenol-chloroform extractions, the cultured cells were washed with 1x PBS (137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2HPO_4, pH 7.4) and mechanically detached from the cell culture dish in 1x PBS. Cells were pelleted at 300 x g for 5 min at room temperature and resuspended in lysis buffer (100 mM Tris pH 8.5, 200 mM NaCl, 5 mM EDTA, 0.2 % (w/v) SDS, 0.1 mg/ml proteinase K). After an overnight incubation at 55 °C, proteinase K was inactivated by heating to 95 °C for 5 min and DNA was prepared using a standard phenol-chloroform protocol. In the end, DNA was resuspended in 1x TE buffer (10 mM Tris pH 8.0, 1 mM EDTA) and concentration and quality were measured using the NanoDrop 1000 (Thermo Scientific). Alternatively, genomic DNA was extracted using a one step lysis buffer-based extraction method. The cultured cells were detached using diluted or 2.5 ‰ (w/v) trypsin at 37 °C for 2 min. Trypsin action was quenched with an equal volume of cell culture medium, the cells were pelleted at 300 x g for 5 min at room temperature and resuspended in CAS lysis buffer (10 mM Tris pH 8.0, 2.5 mM EDTA, 0.2 mM NaCl, 0.15 % (w/v) SDS, 0.3 % (v/v) Tween 20). The lysate was incubated in the thermocycler using the following protocol: 30 s at 65 °C, 30 s at 8 °C, 1 min at 65 °C, 3 min at 97 °C, 1 min at 8 °C, 3 min at 65 °C, 1 min at 97 °C, 1 min at 65 °C, 10 min at 80 °C. In the end, the lysate was diluted with two volumes of water.

2.2. RNA extraction from adherent cell lines and cDNA preparation

Total RNA was extracted from adherent cell lines using TRIzol reagent (MRC). The cultured cells were washed with 1x PBS, placed on ice and covered with TRIzol reagent. The lysate was mixed with 0.2 volume chloroform-isoamyl alcohol (24:1) and centrifugated at 12,000 x g for 15 min at 4 °C. The aqueous phase was transferred into a new tube and supplemented with a volume of 100 % isopropanol equal to half the initial lysate volume. After an overnight incubation at -80 °C, the sample was centrifugated at 12,000 x g for 30 min at 4 °C, the RNA pellet was washed twice with 70 % (v/v) ethanol and dried at 37 °C. RNA was resolved in water at 60 °C for 5 min. Concentration and quality were measured using the NanoDrop 1000. Remaining DNA was eliminated by DNase digestion set-up as follows in a total volume of 20 µl: 0.5 to 2 µg of RNA, 1.5 U of DNase I (10 U/µl, Promega), 10 U of RNasin ribonuclease inhibitor (40 U/µl, Promega) and 1x DNase buffer (10x, Roche). The samples were incubated at 37 °C for 30 min and DNase I was inactivated at 75 °C for 10 min. 20 µl DNase I-treated RNA was mixed with 1 mM of each dNTP (10 mM each, Roche), 55 µM of $p(dN)_6$ primer (Roche) for quantitative PCR or a combination of 1 mM of $p(dN)_6$ primer plus 50 µM of $p(dT)_{15}$ primer (Roche) for cloning. Water was added to a total volume of 29 µl and the sample was incubated at 65 °C for 5 min. The sample was next split into two samples of 14.5 µl. One sample was supplemented with 40 U of RNasin (40 U/µl, Promega), 5U transcriptor reverse transcriptase (RT, 10 U/µl, Roche) and 1x RT buffer (5x, Roche) and water was added to a total volume of 20 µl; the RT was replaced by water in the other sample. Both samples were incubated in the thermocycler using the following protocol: 10 min at 25 °C, 30 min at 50 °C and 5 min at 95 °C.

2.3. Amplification, processing and analysis of linear and circular DNA

Polymerase chain reaction (PCR) was carried out using the Expand High Fidelity PCR System kit (Roche) containing the *Taq* DNA polymerase and the proofreading *Tgo* DNA polymerase. PCR reaction was set-up according to the manufacturer's instructions with either genomic DNA or cDNA as template. Enzymatic digestions of variable amounts of PCR products and plasmids were set-up in a total volume from 20 µl to 100 µl and restriction enzyme units, buffer, incubation time and temperature were chosen according to the manufacturer's indications (NEB). For the dephosphorylation, 10 µg of digested plasmid DNA, 10 U of calf intestinal alkaline phosphatase (CIP, 20 U/µl, NEB) were added to the digestion reaction and incubated at 37 °C for 30 min. PCR products and digested PCR products and plasmids were purified with the NucleoSpin Gel and PCR clean up kit (Macherey-Nagel). DNA concentration and quality were measured using the NanoDrop 1000 (Thermo Scientific). An aliquot of a DNA sample supplemented with 6 x loading dye (25 % (w/v) glycerol, 50 mM Tris-HCl pH 4.7, 50 mM EDTA, 0.05 % (w/v) bromophenol blue, 0.05 % (w/v) xylencyanol FF) was

analysed via agarose gel electrophoresis in 1x TAE buffer (40 mM Tris pH 8.0, 1 mM EDTA) at 8 V/cm for 20 to 60 min. Ligations of a vector with one or two inserts were carried out with the T4 DNA ligase. Vector and insert(s) were added in a molar ratio of 1:3 or 1:5. In a total volume of 20 µl, 50 ng of the linearized vector and the insert(s) were mixed with 800 U T4 DNA ligase (400 U/µl, NEB), 1 mM ATP (NEB) and 1x T4 DNA ligase buffer (10x, NEB). Samples were incubated overnight at 16 °C and the ligase was heat-inactivated at 70 °C for 10 min. Ligated circular DNA was introduced into bacteria via heat-shock transformation of $CaCl_2$-treated chemically competent *Escherichia coli* (*E. coli*) TOP10 (F- *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) Φ80*lac*ZΔM15 Δ*lac*X74 *rec*A1 *ara*D139 Δ(*araleu*)7697 *gal*U *gal*K *rps*L (StrR) *end*A1 *nup*G). Plasmid DNA was isolated using the NucleoSpin Plasmid kit or the NucleoBond Xtra Maxi kit (Macherey-Nagel) according to the manual instructions. DNA concentration and quality were measured using the NanoDrop 1000 (Thermo Scientific). Setting up of TOPO TA cloning reactions were carried out following the instructions of the TOPO TA cloning kit (Invitrogen). PCR products were cloned into the TOPO TA vector pCRII. Transformation of chemically competent *E. coli* TOP10 and blue white screens were performed as described in the manual. Sanger sequencing of purified PXR products and plasmid DNA was carried out by GATC Biotech and samples were prepared according to the instructions on their website (http://www.gatc-biotech.com). Analysis and manipulation of FASTA-formatted DNA sequences was performed with the application SerialCloner (version 2.6.1). Plasmid constructions were first built with SerialCloner and plasmid maps were drawn with the online tool http://www.rf-cloning.org/savvy.php.

2.4. Real time quantitative PCR

Real time quantitative PCR (RT-qPCR) was carried out in 96-well plates using the LightCycler 480 instrument II (Roche). LightCycler 480 SYBR Green I Master mix (Roche) was used for custom SYBR Green primers (Sigma-Aldrich) and the LightCycler 480 Probes Master mix (Roche) for commercial TaqMan assays (ThermoFisher) and Universal ProbeLibrary (UPL) assays (Roche). The used assays (SYBR Green, TaqMan, UPL) are listed in **TABLE B1**. Reactions were set-up in a total volume of 10 µl according to the manufacturer's instructions. Samples were run in technical duplicates. Relative changes in gene expression were determined by the $2^{-\Delta\Delta Ct}$ method using *Rplp0* as reference gene. For a grouped analysis with two factors, two-way ANOVA with Tukey's and Sidak's post-hoc multiple comparison test was used to determine whether the observed expression changes are affected significantly by the two tested factors ($p \geq 0.05$ not significant, $p < 0.05$ *, $p < 0.001$ **, $p < 0.0001$ ***, $p < 0.00001$ ****). For a column analysis with one factor, statistical significances were determined using an f test to compare variances and an unpaired two-tailed t test ($p \geq 0.05$ not significant, $p < 0.05$ *,

p<0.01 **, p<0.001 ***, p<0.0001 ****). Relative gene expression is represented as mean plus standard deviation.

TABLE B1: List of the SYBR Green, TaqMan and UPL assays used for RT-qPCR.

mRNA	fwd primer 5'3'	rev primer 5'3'	assay type	origin
mMlxipl	CAGAAGAGATCAGGGCAAGG	CGCACACACACATCTACATTCA	SYBR Green	Catherine Postic and Sandra Guilmeau
mMlxipl α	CGACACTCACCCACCTCTTC	TTGTTCAGCCGGATCTTGTC	SYBR Green	Catherine Postic and Sandra Guilmeau
mMlxipl β	TCTGCAGATCGCGTGGAG	CTTGTCCCGGCATAGCAAC	SYBR Green	Catherine Postic and Sandra Guilmeau
rMlxipl	TACTGTTCCCTGCCTGCTC	CTTGGAAACCTTCACCAGG	SYBR Green	Zhang *et al.,* 2015
rMlxipl α	TGCATCGATCACAGGTCATT	AGGCTCAAGCATTCGAAGAG	SYBR Green	Zhang *et al.,* 2015
rMlxipl β	TCTGCAGATCGCGCGGAG	CTTGTCCCGGCATAGCAAC	SYBR Green	Zhang *et al.,* 2015
mAbcc8	in exon 28	in exon 29	TaqMan	Mm00803450_m1
mGck	in exon 7	in exon 8	TaqMan	Mm00439129_m1
mIns1	in exon 1	in exon 2	TaqMan	Mm01259683_g1
mIns2	in exon 2	in exon 3	TaqMan	Mm00731595_gH
mPklr	in exon 9	in exon 10	TaqMan	Mm00443090_m1
mRfx3	in exon 15	in exon 16	TaqMan	Mm00803303_m1
mRplp0	in exon 3	in exon 3	TaqMan	Mm01974474_gH
rAbcc8	TAGTGTGGTTCGGTCCACTG	GGACAGGAACTCACTCAGCTTT	UPL #81	
mCacna1c	CCAACCTCATCCTCTTCTTCA	ACATAGTCTGCATTGCCTAGGAT	UPL #71	
rCacna1c	TGGCTCACAGAAGTGCAAGA	AGCATTTCTGCCGTGAAAAG	UPL #79	
mCx3cl1	CATCCGCTATCAGCTAAACCA	CAGAAGCGTCTGTGCTGTGT	UPL #80	
rCx3cl1	TCCACTATCAACTGAACCAGGA	TTGGGTCAGCACAGAAGTGT	UPL #80	
mCyb26b1	ACATCCACCGCAACAAGC	GGGCAGGTAGCTCTCAAGTG	UPL #41	
rCyp26b1	ACGGCAAGGAGATGACCA	GCATAGGCTGCGAAGATCA	UPL #17	
rGck	AATGTGAGGTCGGCATGATT	CACATTCTGCATTTCCTCCA	UPL #71	
rIns1	TCATAGACCATCAGCAAGCAG	CTTGGGCTCCCAGAGGAC	UPL #4	
rIns2	CGAAGTGGAGGACCCACA	TGCTGGTGCAGCACTGAT	UPL #10	
mLdha	GGCACTGACGCAGACAAG	TGATCACCTCGTAGGCACTG	UPL #12	
rLdha	CATGGCGACTCCAGTGTG	CCTGCTTGTGCACATCCTT	UPL #68	
mMlx	GTGTCTTCAGCTGGATTGAGG	GATGAAGGACACCGATCACA	UPL #100	
rMlx	CAGCCTTGAAGATAATGAAAGTGA	TTGAAAGACGTTGAACTTGACCT	UPL #55	
mMlxip	GCCCAACCCACGAGAAATA	GGGTTGCAAAGGGATCAG	UPL #20	
rMlxip	AGGCTGTCTCCTTGGTGTTG	TTTTGTGGTTGGCCTGTAAA	UPL #78	
mMlxipl	CAGAAGAGATCAGGGCAAGG	CGCACACACACATCTACATTCA	UPL #106	
rMlxipl	ACTGTTCCCTGCCTGCTCT	CCTTGTGGCTTGCTCAGG	UPL #89	
mNkx2.2	GCAGCGACAACCCCTACA	ATTTGGAGCTCGAGTCTTGG	UPL #20	
rNkx2.2	GCAGCGACAACCCCTACA	ACTTGGAGCTCGAGTCTTGG	UPL #20	
mRasgrp2	CACCCGTACAAGCCAACC	TGATACTGATCCAGGGACACC	UPL #62	
rRasgrp2	GGGACTTGGACCAGAACCA	GCGCAGGAAGTAGGAGATCA	UPL #46	
rRfx3	ACACTTCCCAGATCAACCAGA	TTGTCATCACACTGGCACAC	UPL #68	
mRfx6	GGCCATGGAGACTCAATTTAAC	GGCAGCTTTACTCGCATCC	UPL #15	
rRfx6	CGGAAGCAGAGCAACTTGT	TTGCAAATTGGAAGCCACT	UPL #55	
mRplp0	ACTGGTCTAGGACCCGAGAAG	TCCCACCTTGTCTCCAGTCT	UPL #9	
rRplp0	GATGCCCAGGGAAGACAG	GAAGCATTTTGGGTAGTCATCC	UPL #85	
mTm6sf2	AAGTTTCGGCGTTCTCACAG	CGTAGGTGATCTCCCCATGA	UPL #72	
rTm6sf2	GCTGGGGTCCTTCGCTAT	GCAAGGAAGAAGGTAGGTCTGA	UPL #79	
mTns1	CTCCTGTCATCGGAAATGC	CACCAGTTCATGGCTGGAC	UPL #83	
rTns1	GACAAGATCGTGCCCATTG	GCAGGCCGCTGAAGTAGT	UPL #82	
mTxnip	ATCCCAGATACCCCAGAAGC	TGAGAGTCGTCCACATCGTC	UPL #77	
rTxnip	TCTCCGAGTGCAGAAGATCA	GGAGCCAGGGACACTAACAT	UPL #26	

2.5. RNA sequencing

RNA sequencing was carried out by the GenomEast platform of the institute. RNA quantity was measured with the Qubit 3.0 fluorometer (Life Technologies), and RNA quality was assessed with the eukaryotic total RNA Nano assay run on the Agilent 2100 Bioanalyzer system (Agilent Technologies). The system calculates an RNA integrity number (RIN) from 1 (degraded RNA) to 10 (intact RNA) based

on the sample's electropherogram. The cDNA libraries were prepared using 1 µg of total RNA supplemented with spike-in RNAs and the TruSeq stranded mRNA low throughput sample preparation kit (Illumina) according to the manufacturer's instructions. The quality and quantity of the final cDNA libraries was assessed by capillary electrophoresis. The cDNA libraries were sequenced using the HiSeq 4000 (Illumina) with a read length of 50 bp (32 cDNA libraries were split onto 2 lanes of a flow cell). The analysis of the RNA sequencing data was done by Constance Vagne. In brief, FASTQ-formatted output data were preprocessed using cutadapt (version 1.10) to filter out adapter, polyA tails, low-quality sequences and reads shorter than 40 bp. Reads mapped to spike-in RNAs and rRNA sequences were removed using bowtie (version 2.2.8). The remaining reads were mapped on the rat genome (rn6 assembly) using STAR (version 2.5.3a). With the RNA quality control tool RSeQC (version 2.6.4.), the read coverage of all genes and the transcript integrity number (TIN) for all transcripts in all samples were determined. The TIN ranges from 0 (gene is not covered) to 100 (gene is completely covered). The median of the TIN score across all samples indicates the percentage of transcripts with a uniform read coverage per sample and is a good indicator for the RNA integrity of individual transcripts. Transcripts with less than 10 mapped reads were skipped from the analysis. Afterwards, the reads were sorted into uniquely mapped, multi-mapped and unmapped reads. Only the uniquely mapped reads were kept. The uniquely mapped reads were categorized into reads mapping to intergenic regions, introns or exons using the Ensembl genome browser 92. In the case that a read aligned to two types of genomic regions, reads were preferentially classified as exonic, then intronic and at last intergenic regions. In the following quantification step, the reads were assigned to annotated genes of the Ensembl genome browser 92 using htseq-count (version 0.6.4p1.) to quantify the number of reads per gene. Ambiguously assigned reads aligned to more than one annotated gene and unassigned reads that could not be aligned, were removed. The number of reads per sample was normalized with the DESeq2 median-of-ratio method (executed with R) to allow the comparison of samples. In this normalization method, the expression level of a given gene is divided by the expression level of the same gene in a fictive reference sample in which its expression is the geometric mean of the expression level of this gene in all samples. The median of the ratios calculated for all expressed genes of a given sample is used to normalize the expression level of all genes in this sample. The uniformity of the replicates was evaluated by the Simple Error Ratio Estimate (SERE) coefficient (calculated with R) to quantify global differences between RNA sequencing samples. A value of 0 means that there are no differences, a value of 1 indicates that the observed variance corresponds to the expected variance and values above 1 mean that the samples are different (the higher, the more different). The data sets of the different experimental conditions were compared using R and the Bioconductor package DESeq2 (version 1.16.1.) allowing the identification of differentially expressed genes. The p-value was assessed with the Wald test. Outliers with a high Cook's distance were removed from the analysis and

no p-value was assigned to these genes. An additional p-value called adjusted p-value was determined with the False Discovery Rate (FDR) Benjamini-Hochberg method. The adjusted p-value is necessary in the case that a high number of comparisons is performed, to minimize the rate of type-I-errors in null hypothesis testing, or, in other words, to decrease the number of non-significant differences that have been falsely classified as significant. For the calculation of the adjusted p-value, genes with a very low expression were excluded to increase the power of the statistical test.

2.6. Protein extraction from adherent cell lines and detection by western blot

For the standard freeze-thaw protein extraction, cells were washed with 1x PBS, mechanically detached from the cell culture dish in 1x PBS, pelleted at 300 x g for 5 min at 4 °C and resuspended in lysis buffer (1 % (v/v) NP40, 150 mM NaCl, 50 mM Tris pH 8.1, 1x complete EDTA-free protease inhibitor cocktail of Roche). The samples were frozen in liquid nitrogen, thawed on ice and vortexed. The freeze-thaw step was repeated once, and cell debris was pelleted at 15,000 x g for 15 min at 4°C. Protein concentration in the supernatant was determined by a Bradford assay. Alternatively, proteins were extracted using passive lysis buffer (PLB, Promega) allowing to extract proteins from confluent cells of a single well of a 96-well plate. Cells were washed with 1x PBS and covered with 1x PLB and incubated for 15 min at room temperature on a horizontal shaker. A standard wet blot protocol was used to detect proteins. 20 µg of proteins or 10 µl of lysate were mixed with 2 x Laemmli buffer (125 mM Tris pH 6.8, 0.1 % (w/v) SDS, 20 % (w/v) glycerol, 20 % (v/v) β-mercaptoethanol, 0.004 % (w/v) bromophenol blue) and heated to 99 °C for 10 min prior to gel electrophoresis. An upper 5 % (v/v) polyacrylamide gel a lower 8 % (v/v) polyacrylamide gel was used to separate the proteins according to their size by sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE). The proteins were transferred to a nitrocellulose membrane via electroblotting and 5 % (w/v) dried milk solution was used for the blocking step. Primary antibodies and secondary antibodies used for protein detection are listed in **TABLE B2**.

TABLE B2: List of primary and secondary antibodies used for western blotting.

Antibodies	Antigen	Host	Dilution	Company
Primary	ActinB	mouse	1:1,000	IGBMC antibody facility (reference 1ACT207)
	MLXIPL	rabbit	1:1,000	Novus (reference NB400-135)
	MLXIPL	rabbit	1:1,000	Abcam (reference AB92809)
	NKX2.2	mouse	1:1,000	Gift of Thomas Jessell
	RFX6	rabbit	1:1,000	IGBMC antibody facility (reference 2766)
	RFX6	rabbit	1:1,000	IGBMC antibody facility (reference 2766)
	RFX6	rabbit	1:1,000	Sigma (reference SAB1402062)
Secondary coupled to horseraddish peroxidase	mouse	goat	1:20,000	IGBMC antibody facility
	rabbit	goat	1:20,000	IGBMC antibody facility

3. Plasmids

3.1. Cloning of *mMlxipl* intron 1 luciferase reporter vectors

Three wild-type *mMlxipl* intron 1 luciferase reporter vectors were cloned by introducing either a 1165 bp (Chr5:135,110,073-135,111,237, mm10), an 821 bp (Chr5:135,110,215-135,111,035, mm10) or a 606 bp (Chr5:135,110,215-135,110,822, mm10) long sequence of the first intron of *mMlxipl* into the commercial firefly luciferase reporter vector pGL4.23-minP-luc2 (Promega). Three additional mutant reporter vectors were created by deleting the xbox 2 motif. In total, six *mMlxipl* intron 1 luciferase reporter vectors were produced (**FIGURE B1**; see **FIGURE C1E** and **FIGURE C2A** for more information about the fragments). A general workflow is shown in **FIGURE B1A** starting with the cloning of the reporter vector with the wild-type 1165 bp fragment (**FIGURE B1B**). The 1165 bp *mMlxipl* intron 1 fragment was amplified from mouse genomic DNA of Min6b1 cells using a primer pair that incorporated a *Kpn*I site on the 5'end and a *Nhe*I site on the 3'end of the PCR product (**TABLE B3**). The PCR product was purified, digested with *Kpn*I and *Nhe*I and re-purified by a second PCR clean up (**FIGURE B1C**). The reporter vector pGL4.23 was digested with *Kpn*I and *Nhe*I, dephosphorylated and the 4270 bp long plasmid backbone was purified by a PCR clean up (**FIGURE B1C**). Vector and insert were next ligated and correct insert integration was verified by *Not*I / *Ppu*MI digestion (**FIGURE B1C**) and sequencing (**TABLE B3**). The xbox 2 motif was deleted by site-directed mutagenesis using the QuickChange XL Site-Directed Mutagenesis kit (Agilent Technologies). Primers (**TABLE B3**) were designed using the online primer design tool http://www.agilent.com/genomics/qcpd. PCR, *Dpn*I digestion and transformation of XL10-Gold chemically competent cells (Tet^{r}Δ (*mcrA)183 Δ(mcrCB-hsdSMR-mrr)173 endA1 supE44 thi-1 recA1 gyrA96 relA1 lac* Hte [F′ *proAB lacI^{q}ZΔM15* Tn*10* (Tet^r) Amy Cam^r]) was performed following the instructions of the manual. The deletion of the xbox 2 motif was verified by sequencing (**FIGURE B1D**). The 821 bp and 606 bp fragments were amplified from the pGL4.23-minP-luc2 mMlxipl intron 1 1165 bp with or without the xbox 2 motif (**FIGURE B1A**) using primer pairs that incorporated a *Kpn*I site on the 5'end and a *Nhe*I site on the 3'end of the PCR product (**FIGURE B1E**, **TABLE B3**). The four PCR products were purified, digested with *Kpn*I and *Nhe*I and re-purified by a second PCR clean up. The 4270 bp long plasmid backbone and the newly produced inserts were ligated (**FIGURE B1F+G**) and correct insert integration was verified by *Not*I / *Ppu*MI digestion (**FIGURE B1H**) and sequencing (**TABLE B3**).

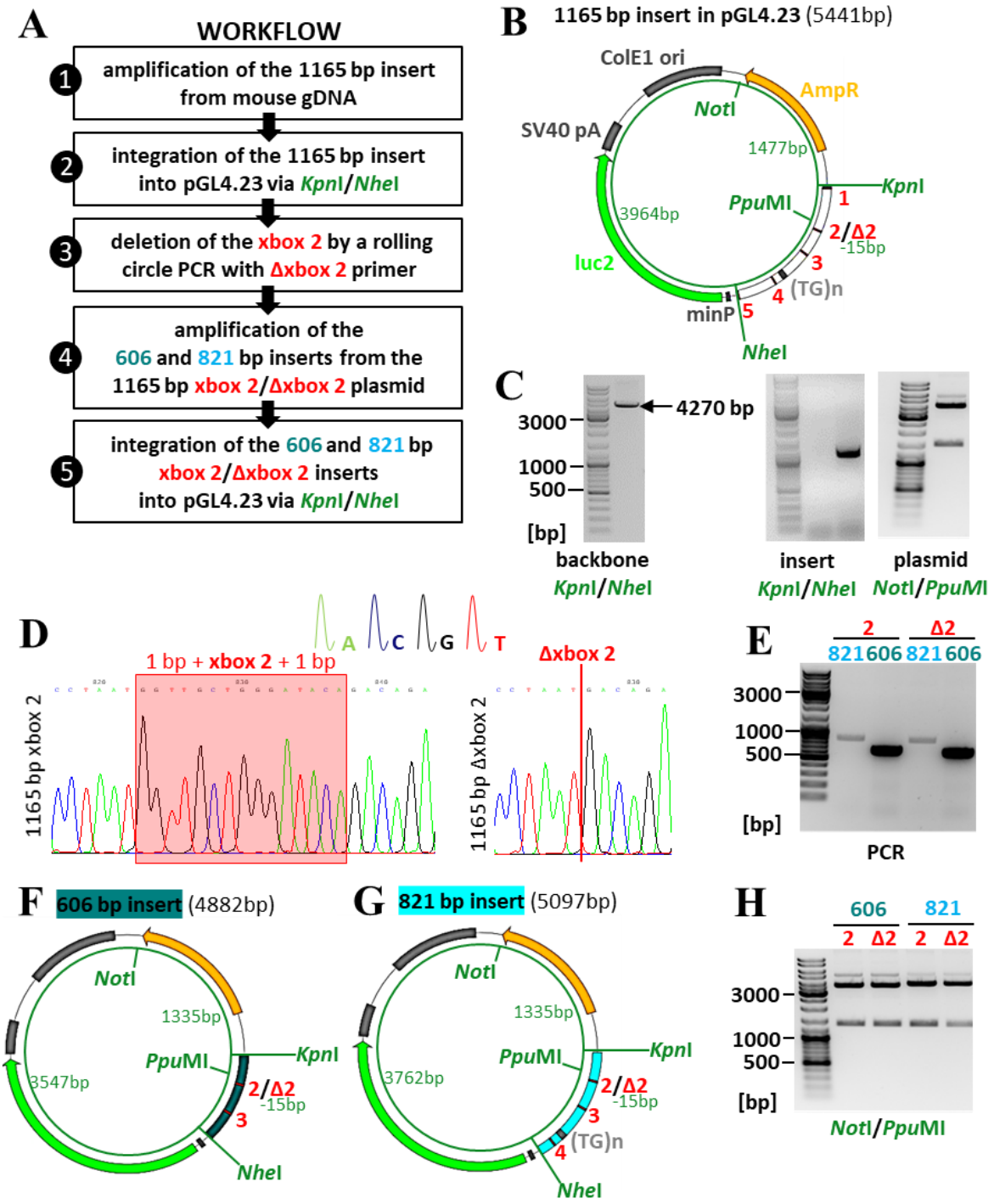


FIGURE B1: Cloning of *mMlxipl* intron 1 luciferase reporter vectors. (A) Overview of the workflow of the reporter vectors' cloning process (see **FIGURE C1E** and **FIGURE C2A** for more information about the fragments). **(B)** Map of the reporter vector pGL4.23 with the *mMlxipl* intron 1 1165 bp insert (restriction sites used for the cloning process are shown in green) upstream of a minimal promoter (minP) and the luciferase gene (*luc2*). **(C)** Agarose gel showing the *Kpn*I and *Nhe*I digested backbone pGL4.23 (on the left), the *Kpn*I and *Nhe*I digested *mMlxipl* intron 1 1165 bp insert amplified from Min6b1 genomic DNA (in the middle) and the control *Not*I and *Ppu*MI digestion of the ligated plasmid (on the right). **(D)** Sequencing proving the xbox 2 deletion by site-directed mutagenesis. **(E)** Agarose gel showing the wild-type and xbox 2 mutated *mMlxipl* intron 1 606 bp and 821 bp PCR products amplified from the wild-type and xbox 2 mutated *mMlxipl* intron 1 1165 bp plasmid, respectively. **(F, G)** Map of the reporter vector pGL4.23 with the *mMlxipl* intron 1 606 bp **(F)** and 821 bp **(G)** insert (restriction sites used for the cloning process are shown in green) upstream of a minimal promoter (minP) and the luciferase gene (*luc2*). **(H)** Agarose gel showing the control *Not*I and *Ppu*MI digestion of the ligated pGL4.23 plasmids with the wild-type and mutated *mMlxipl* intron 1 606 bp and 821 bp fragments.

TABLE B3: List of primers used for plasmid cloning. The restriction sites are highlighted in green, the coding sequence in red and the start / stop codons in orange.

primer (for cloning)	sequence 5'3'	plasmid
mMlxipl intron 1 1156bp KpnI fwd	CCGGTACCGAGAGACTGTCTCAAAAACG	pGL4.23-minP-luc2 with mMlxipl intron 1 606/821/1165 bp (xbox 2 or Δxbox 2)
mMlxipl intron 1 1156bp NheI rev	CCGCTAGCTCAAGGTTCCTCTGAGAACA	
mMlxipl intron1 606/821bp KpnI fwd	CGGGTACCTTTTGAACCTTG	
mMlxipl intron 1 821bp NheI rev	CTTTGCTAGCGCCATCCAG	
mMlxipl intron 1 606bp NheI rev	CTGCTAGCCTGCTCTCTGGGTAGCTC	
mMlxipl intron 1 Δxbox2 fwd	CCCTGGGGGAACAGAACCTAATGACAGAGGTCAC	
mMlxipl intron 1 Δxbox2 rev	GTGACCTCTGTCATTAGGTTCTGTTCCCCCAGGG	
mMlxipl exon 1b KpnI fwd	GGAAAGGTACCGCAGAAGGTGATTGGCAGGCTCC	pGL4.23-minP-luc2 with mMlxipl exon 1b
mMlxipl exon 1b NheI rev	GGATCGCTAGCTCCTCTGCGAGGCATCTATGTCC	
mMlxipl α NotI fwd	CCAGGCGGCCGCGGCGACAATAGTGGCCATGGC	pcDNA3-mMlxipl α
mMlxipl α XbaI rev	GAAGCAGTCCAGGTCTAGAAGC	
mMlxipl β NotI fwd	CCAGGCGGCCGCGGCGCATCGAGGTGGTGATG	pcDNA3-mMlxipl β
mMlxipl β ApaI rev	GCAAGGGCCCTTATAATGGTCTCCCCAGGGTG	
mMlx NotI fwd	GAATGCGGCCGCGGTAGGTTCACGATGACGGAG	pcDNA3-mMlx
mMlx XbaI rev	CTCCTCTAGATCAGTAGAGTTGGTTTTTCAACTG	
mNkx2.2 BamHI fwd	GGCATGGATCCGTCCGGAACCATGTCGCTGACC	pcDNA3-mNkx2.2
mNkx2.2 NotI rev	GGGTTGCGGCCGCTCACCAAGTCCACTGCTGGGCC	
hRfx6 NotI fwd	ACACTGGCGGCCGCCACCATGGCCAAGGTCCCGGAGCTGG	pcDNA3-hRfx6
hRfx6 XhoI rev	GAGATTATGCATTACTCGAGCACCAA	
hRfx6 XhoI fwd	TTGGTGCTCGAGTAATGCATAATCTC	
hRfx6 XbaI rev	CCGCTTTCTAGATTAAGTGCCTCCAGCTGCTGTTCC	
primer (for sequencing)	**sequence 5'3'**	**plasmid**
T7 fwd/rev	TAATACGACTCACTATAGGG	pCRII and pcDNA3
SP6 fwd/rev	ATTTAGGTGACACTATAGAA	
pGL4.23 fwd (1)	GCAAGTGCAGGTGCCAG	pGL4.23-minP-luc2
pGL4.23 fwd (2)	ACAACACCGCGCCACATAG	
pGL4.23 rev	AACAGTACCGGATTGCCAAG	
pCMV fwd	CGCAAATGGGCGGTAGGCGTG	pcDNA3
pcDNA3 rev	CAAACAACAGATGGCTGGC	
mMlxipl fwd	GGTGCGAACAGCTCTTCTCCAG	pcDNA3-mMlxipl α+β
mMlxipl rev	GAGACAGGTGGACACTCAGTGC	
hRfx6 fwd (1)	ACGACAAACATCTTTCTTAC	pcDNA3-hRfx6
hRfx6 fwd (2)	TAATTCACCAAATGGATACT	

3.2. Cloning of *mMlxipl* exon 1b luciferase reporter vector

The *mMlxipl* exon 1b luciferase reporter vector (**FIGURE B2**) was cloned by introducing a 407 bp (Chr5:135,089,259-135,090,113, mm10) *mMlxipl* exon 1b sequence (see **FIGURE C18** for more information about the fragment) into the commercial firefly luciferase reporter vector pGL4.23-minP-luc2 (Promega; **FIGURE B2A**). The fragment was amplified from mouse genomic DNA of Min6b1 cells using a primer pair that incorporated a *Kpn*I site on the 5'end and a *Nhe*I site on the 3'end of the PCR product (**TABLE B3**). The PCR product was purified, digested with *Kpn*I and *Nhe*I and re-purified by a second PCR clean up (**FIGURE B2B**). The reporter vector pGL4.23 was digested with *Kpn*I and *Nhe*I, dephosphorylated and the 4270 bp long plasmid backbone was purified by a PCR clean up (**FIGURE B2B**). Vector and insert were next ligated and correct insert integration was verified by *Not*I / *Ppu*MI digestion and sequencing (**FIGURE B2C**, **TABLE B3**).

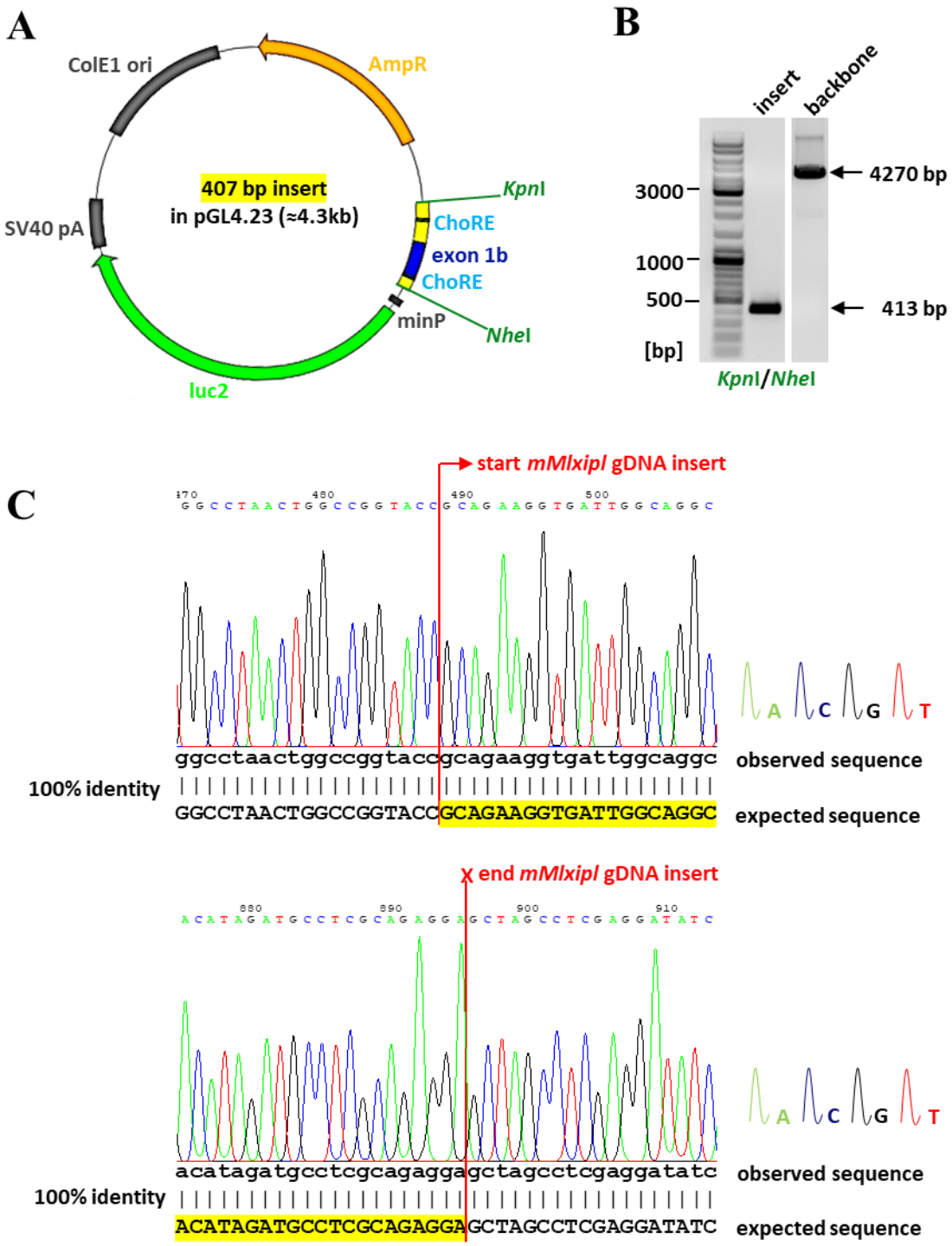


FIGURE B2: Cloning of mMlxipl exon 1b luciferase reporter vector. (A) Map of the reporter vector pGL4.23-minP-luc2 with the *mMlxipl* exon 1b 407 bp insert (see **FIGURE C18** for more information about the fragment, restriction sites used for the cloning process are shown in green) upstream of a minimal promoter (minP) and the luciferase gene (luc2). **(B)** Agarose gel showing the *Kpn*I and *Nhe*I digested *mMlxipl* exon 1b 407 bp insert amplified from Min6b1 genomic DNA and the *Kpn*I and *Nhe*I digested backbone pGL4.23. **(C)** Sequencing proving the correct integration of the *mMlxipl* exon 1b 407 bp insert into pGL4.23 at the 5' end and the 3' end of the inserted sequence.

3.3. Cloning of *mMlxipl* α, *mMlxipl* β and *mMlx* expression vectors

To produce *mMlxipl* expression vectors (**FIGURE B3**), the coding sequence of *mMlxipl* β (NM_021455) was amplified from cDNA prepared from a total RNA extract of mouse liver tissue (gift of Catherine Postic and Sandra Guilmeau) and cloned into the eukaryotic expression vector pcDNA3 (**FIGURE B3A**). The used primer pair (**TABLE B3**) integrated a *Not*I site on the 5'end and an *Apa*I site on the 3'end of the PCR product. The PCR product was purified, digested with *Not*I and *Apa*I and re-purified (**FIGURE B3B**). The plasmid pcDNA3 was digested with *Not*I and *Apa*I, dephosphorylated and the 5421 bp long plasmid backbone was purified by a PCR clean up (**FIGURE B3B**). Vector and insert were ligated, and correct insert integration was verified by *Kpn*I digestion (**FIGURE B3F**) and sequencing (**TABLE B3**). The expression vector pcDNA3-mMlxipl β was next used to clone pcDNA3-mMlxipl α (**FIGURE B3C**). The 3'end of the coding sequence of the shorter isoform *mMlxipl* β corresponds to the *mMlxipl* α coding sequence (NM_021455) that contains an additional 530 bp long part at the 5'end (**FIGURE B3D**). This additional sequence at the 5'end was amplified by PCR from mouse liver cDNA using a primer pair that added a *Not*I site at the 5'end of the PCR product (**TABLE B3**). The 3'end stopped at an internal *Xba*I site (**FIGURE B3D**, **TABLE B3**). The PCR product was purified, digested with *Not*I and *Xba*I and re-purified (**FIGURE B3E**). The expression vector pcDNA3-mMlxipl β was cut with *Not*I and *Xba*I into a longer 7291 bp fragment and a shorter 219 bp fragment. The 7291 bp fragment was purified by gel extraction, both fragments were ligated, and the integrity of the construct was verified by *Kpn*I digestion (**FIGURE B3F**) and sequencing (**TABLE B3**). For the preparation of the expression vector pcDNA3-mMlx (**FIGURE B3G**), the coding sequence of *mMlx* isoform 2 (NM_001159385) was amplified from the mouse liver cDNA with primers adding a *Not*I site on 5'end and a *Xho*I site on the 3'end. The PCR product was purified and digested with *Not*I and *Xho*I (**FIGURE B3H**). The plasmid backbone pcDNA3 was linearized with *Not*I and *Xba*I (**FIGURE B3H**). Both fragments were purified by PCR clean up and ligated. The correct insertion was checked by *Drd*I digestion (**FIGURE B3H**) and sequencing (**TABLE B3**).

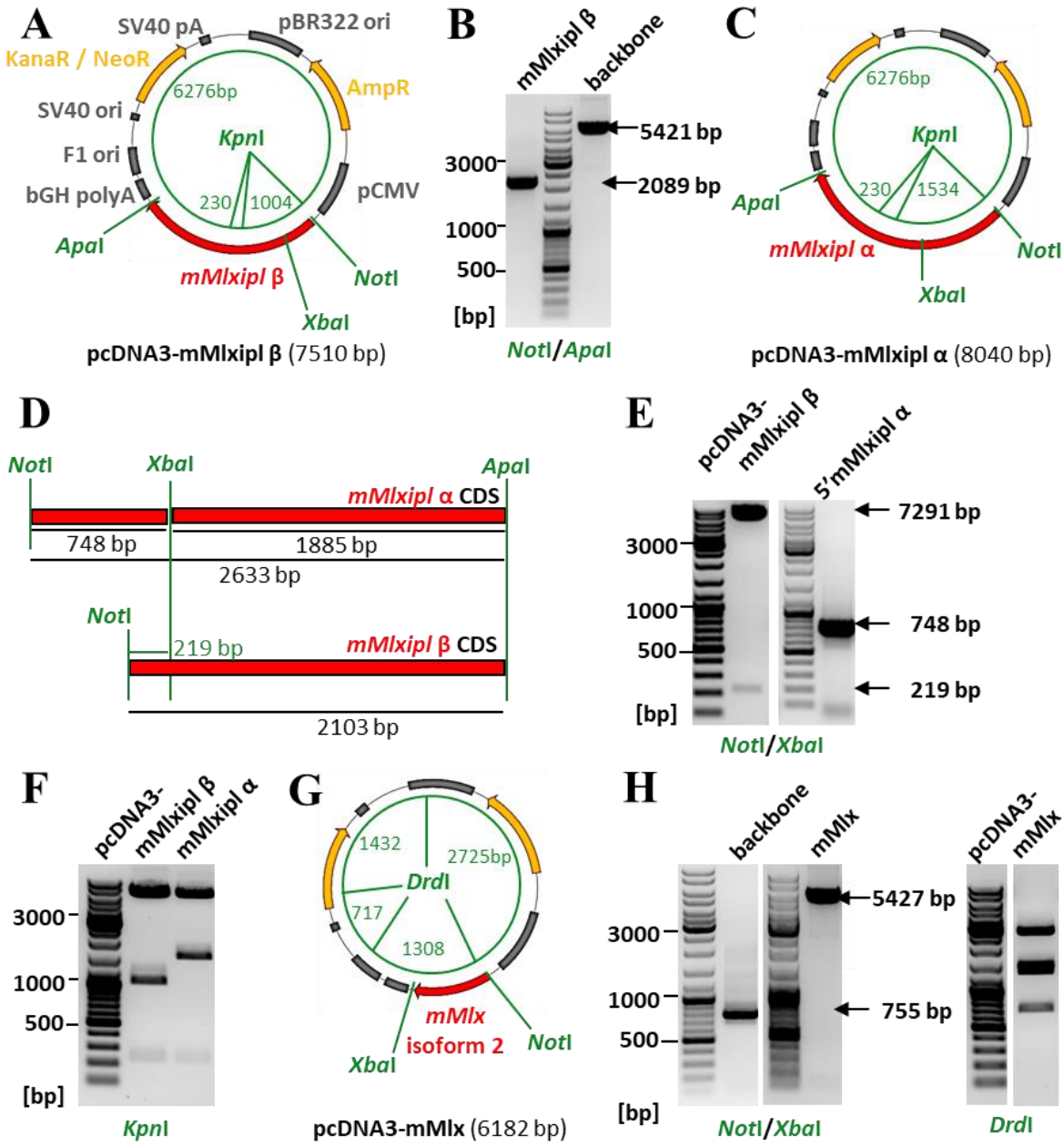


FIGURE B3: Cloning of a *mMlxipl* α and β and a *mMlx* expression vector. (A) Map of the pcDNA3-mMlxipl β expression vector (restriction sites used for the cloning process are shown in green). **(B)** Agarose gel showing the *Not*I and *Apa*I digested PCR product containing the *mMlxipl* β coding sequence amplified from mouse liver cDNA and the *Not*I and *Apa*I digested backbone pcDNA3. **(C)** Map of the pcDNA3-mMlxipl α expression vector (restriction sites used for the cloning process are shown in green). **(D)** Map showing the identical and different parts of the coding sequence of *mMlxipl* α and β and the restriction sites used for the cloning process. To integrate *mMlxipl* α into pcDNA3, the fact was used that the *mMlxipl* α and β coding sequences are identical in their 3'part. The 5'end of *mMlxipl* α from the NotI to the XbaI site was added into pcDNA3-mMlxipl β to produce pcDNA3-mMlxipl α. **(E)** Agarose gel showing the *Not*I and *Xba*I digested pcDNA3-mMlxipl β and the *Not*I and *Xba*I digested *mMlxipl* α 5'end of the coding sequence that was amplified from mouse liver cDNA. **(F)** Agarose gel showing the control *Kpn*I digestion of the pcDNA3-mMlxipl α and pcDNA3-mMlxipl β. **(G)** Map of the pcDNA3-mMlx expression vector (restriction sites used for the cloning process are shown in green). **(H)** Agarose gel showing the digested PCR product containing the *mMlx* coding sequence amplified from mouse liver cDNA and the *Not*I and *Xba*I digested backbone pcDNA3 (on the left) and the control *Drd*I digestion of the pcDNA3-mMlx (on the right).

3.4. Cloning of a *mNkx2.2* expression vector

To prepare the eukaryotic expression vector pcDNA3-mNkx2.2 (**FIGURE B4**), the coding sequence of *mNkx2.2* isoform 1 (NM_001077632; **FIGURE B4A**) was amplified by PCR on cDNA prepared from murine Min6b1 cells using a primer pair that introduced a *Bam*HI site and a *Not*I site at the 5' and 3'end, respectively. The plasmid pcDNA3 and the purified PCR product were cut with *Bam*HI and *Not*I and purified by a PCR clean up (**FIGURE B4B**). After ligation, the correct insertion of the insert into the plasmid backbone was verified by sequencing (**FIGURE B4C**, **TABLE B3**).

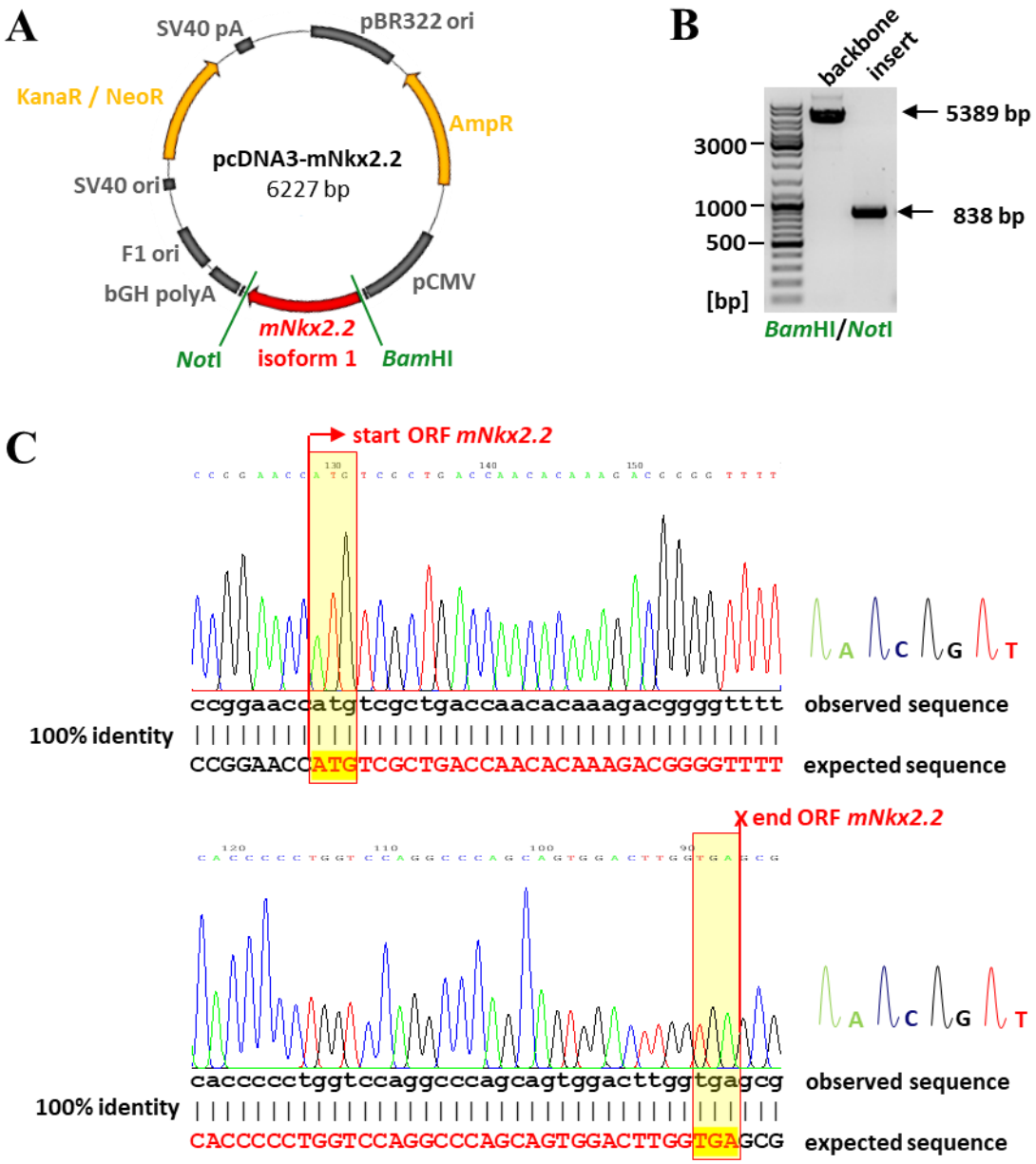


FIGURE B4: Cloning of a *mNkx2.2* expression vector. (A) Map of the expression vector pcDNA3-mNkx2.2 (restriction sites used for the cloning process are shown in green). **(B)** Agarose gel showing the *Bam*HI and *Not*I digested backbone pcDNA3 and the *Bam*HI and *Not*I digested PCR product containing the *mNkx2.2* coding sequence amplified from Min6b1 cDNA. **(C)** Sequencing proving the correct integration of the *mNkx2.2* coding sequence into pcDNA3 at the 5' end and the 3' end of the inserted sequence.

3.5. Cloning of a *hRFX6* expression vector

For the cloning of a human *RFX6* expression vector (**FIGURE B5**), a cDNA prepared from a total RNA extract of human EndoC-βH2 cells (a gift from Raphaël Scharfmann) was used as template for the PCR. The coding sequence of *hRFX6* (NM_173560) was divided into two fragments to facilitate its amplification by PCR and its integration into the eukaryotic expression vector pcDNA3 (**FIGURE B5A+B**). A *Not*I site and a *Xho*I site were inserted at the 5' and the 3'end of the 5'part of *hRFX6* during the amplification, respectively. The forward primer used to amplify the 3'part of *hRFX6* was complementary to the reverse primer of the 5'part and the reverse primer of the 3'part added a *Xba*I site to the 3'end of the PCR product (**FIGURE B5B**). Both PCR products were purified, digested with the respective restriction enzymes and re-purified. The plasmid pcDNA3 was cut with *Not*I and *Xba*I and purified. Both inserts were integrated simultaneously into the plasmid backbone and the correct integration was examined by sequencing (**FIGURE B9C**, **TABLE B9**).

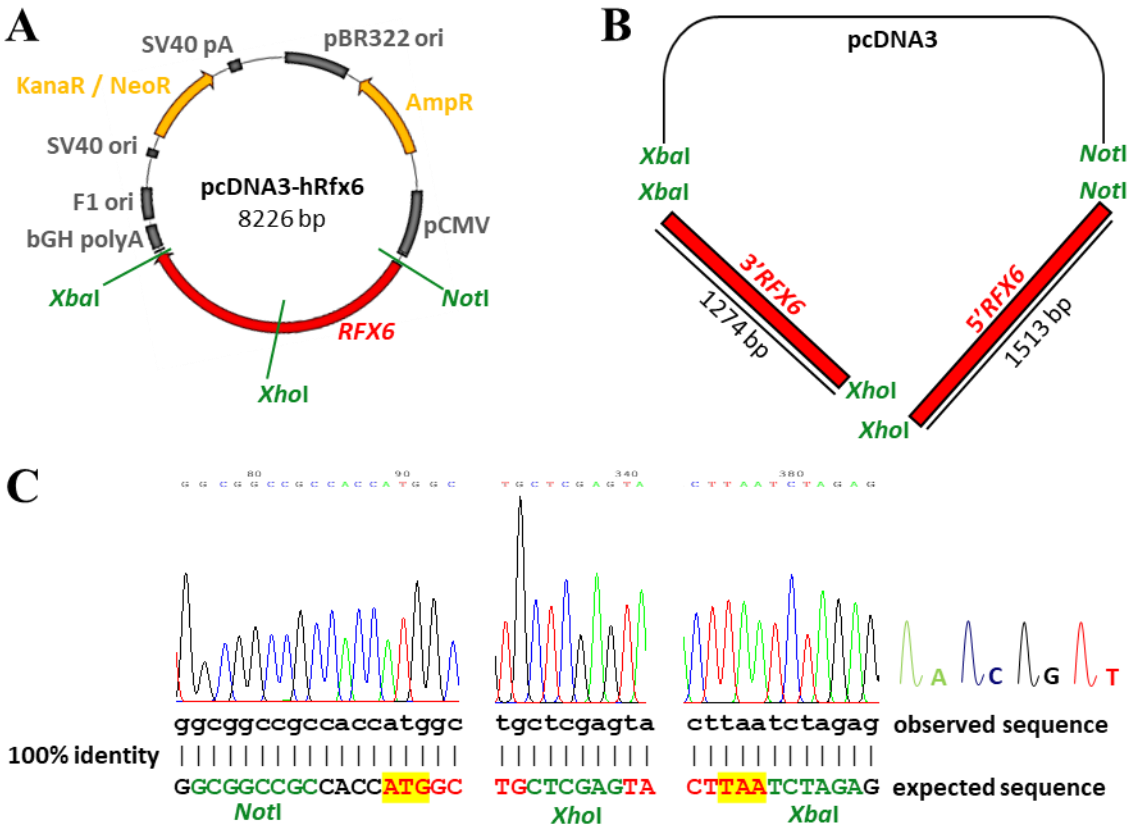


FIGURE B5: Cloning of a new *RFX6* expression vector. (A) Map of the expression vector pcDNA3-hRfx6 (restriction sites used for the cloning process are shown in green). The *hRFX6* cDNA was amplified from cDNA of the human beta cell line EndoC-βH2, contains the wild-type *hRFX6* cDNA without the point mutation of the previously cloned pcDNA3-hRfx6 (see **FIGURE C4C**). **(B)** Schematic illustrating the strategy for the cloning of the *hRFX6* coding sequence into pcDNA3. The coding sequence was split into two parts: a 5'part from a *Not*I site to an internal *Xho*I site and a 3'part from the internal *Xho*I site to a XbaI site. Both fragments were cloned simultaneously into pcDNA3 linearized with *Not*I and *Xba*I. **(C)** Sequencing proving the correct integration of the *hRFX6* coding sequence into pcDNA3 at the 5' end and the 3' end of the inserted sequence and the internal ligation site.

4. Cell culture methods

4.1. Cultivation of the human embryonic kidney cell line HEK293T

HEK293T cells (DuBridge *et al.,* 1987) were cultivated at 37 °C and 5 % CO_2 in Dulbecco's Modified Eagle Medium (DMEM, Life Technologies) containing 1 g/l glucose and supplemented with 10 % (v/v) foetal calf serum (lot # 407), 1 % (w/v) penicillin and 1 % (w/v) streptomycin. Medium was changed every two or three days. At 90 % confluence, cells were split using a diluted trypsin solution. Cells were covered with trypsin and incubated for 2 min at 37 °C. Trypsin action was quenched with an equal volume of cell culture medium and the cell suspension was mixed with fresh medium in a 1:19 ratio. For a glucose stimulation, DMEM containing 2 mM glucose or 20 mM glucose was used as low or high glucose medium, respectively.

4.2. Cultivation of the murine pancreatic beta cell line Min6b1

Min6b1 cells (Lilla *et al.,* 2003) were cultivated at 37 °C and 5 % CO_2 in Dulbecco's Modified Eagle Medium (DMEM, Life Technologies) containing 25 mM glucose, 15 % (v/v) foetal calf serum (lot # 3355), 71 µM β-mercaptoethanol, 1 % (w/v) penicillin and 1 % (w/v) streptomycin. Medium was changed every two or three days. At 90 % confluence, cells were split using a trypsin solution. Cells were covered with 2.5 ‰ (w/v) trypsin and incubated for 2 min at 37 °C. Trypsin action was quenched with an equal volume of cell culture medium and the cell suspension was mixed with fresh medium in a 1:3 ratio. For a glucose stimulation, standard cell culture medium supplemented with 3 mM or 30 mM glucose was used as low and high glucose medium, respectively. Incubation times in low and high glucose media for samples intended to be used for the RT-qPCR analysis are specified in the respective results' chapters.

4.3. Cultivation of the rat pancreatic beta cell line Ins-1 832/13

Ins-1 832/13 (Hohmeier *et al.,* 2000) were cultivated at 37 °C and 5 % CO_2 in RPMI 1640 medium supplemented with 0.3 mg/l L-glutamine, 2 g/l $NaHCO_3$, 10 mM HEPES, 11 mM glucose, 10 % (v/v) foetal calf serum (lot #3355), 50 µM β-mercaptoethanol and 40 µg/ml gentamycin. Medium was changed every two or three days. At 90 % confluence, cells were split using a trypsin solution. Cells were covered with 2.5 ‰ (w/v) trypsin and incubated for 2 min at 37 °C. Trypsin action was quenched with an equal volume of cell culture medium and the cell suspension was mixed with fresh medium in a 3:17 ratio. For a glucose stimulation, either DMEM or standard cell culture medium supplemented with 2 mM or 20 mM glucose was used as low and high glucose medium, respectively. Incubation times

in low and high glucose media for samples intended to be used for the RT-qPCR analysis are specified in the respective results' chapters.

4.4. RNA interference

For the knockdown experiments, confluent Ins-1 832/13 cells were transfected utilizing Lipofectamine RNAiMAX reagent (Invitrogen) according to the manufacturer's protocol. Dharmacon siRNAs (**TABLE B4**) were used at 1 pmol, 5 pmol and 25 pmol per well of a 96-well, 24-well and 6-well plate, respectively. The siMlxipl ordered from Sigma (**TABLE B4**) was either used at 6.25 pmol, 12.5 pmol or 25 pmol per well of a 6-well plate. Non-targeting control siRNAs (siControl) were ordered from Dharmacon (ON-TARGET plus non-targeting pool; **TABLE B4**) and Sigma (Mission siRNA Universal Negative Control #1 SIC001; **TABLE B4**) and used at the same concentrations as the targeting siRNAs. The incubation times and used cell culture media in the different experiments are stated in the respective results' parts.

TABLE B4: List of small interfering RNAs used in RNA interference experiments.

mRNA	targeted exon	siRNA sequence 5'3'	origin
non-targeting	none	UGGUUUACAUGUCGACUAA	Dharmacon ON-TARGET plus SMARTpool
		UGGUUUUACAUGUUGUGUGA	
		UGGUUUACAUGUUUUCUGA	
		UGGUUUACAUGUUUUCCUA	
rMlxipl (15 exons)	10	GAUCUCAACUCCAUUCAAC	Dharmacon ON-TARGET plus SMARTpool
	11	GCAGAAGAGGCGUUUCAAU	
	12	GCAACUGAGGGAUGAAAUA	
	14	CAUCCGACCUUUAUUUGAG	
rNkx2.2 (2 exons)	1	GGUUUUCAGUCAAGGACAU	Dharmacon ON-TARGET plus SMARTpool
	2	ACGCAGGCAAGAAGCGGAA	
	2	CCACGGUUGUUAUUAUUAU	
	2	ACAGAAUGUUUGUGCAGCU	
rRfx3 (16 exons)	2+3	CAAUGGAGCAAUCCGAACA	Dharmacon ON-TARGET plus SMARTpool
	3+4	CGACAAUUGAAAUGGCGAU	
	8+9	AGGCAAUAUUGGACGUUGU	
	11	GACCCAAGCCAUUCGAAAU	
mRfx6 (19 exons)	11	CGACAGACAUCUUUCUUGC	Dharmacon ON-TARGET plus SMARTpool
	12	UGAAUGCCAUGGUAUCUGA	
	12+13	GAACAUGACUCUAUUACUG	
	15	CAUCUGAUCCGAAUGCUUC	
rRfx6 (19 exons)	7	GAUGCAUUGCUAAGGAUAA	Dharmacon ON-TARGET plus SMARTpool
	12	GCACAGACACAGAAUCCGA	
	13	CCUUAAGAAGAAUGCGACA	
	17	CGUCAAGUCUCUAUGCUCA	
non-targeting	none	Mission siRNA Universal Negative Control #1 SIC001	Sigma
rMlxipl (15 exons)	12	GCAACUGAGGGAUGAAAUATT	Sigma custom siRNA Schmidt *et al.*, 2016

4.5. Transactivation assay

Luciferase assays were performed using the Dual Luciferase Reporter Assay System kit (Promega). The *Photinus pyralis* luciferase (*luc2*) referred to as firefly luciferase below reports the activation of a genomic DNA fragments that is cloned upstream of a minimal promoter and the firefly luciferase coding sequence into the reporter vector pGL4.23-minP-luc2 (Promega). The *Renilla reniformis* (sea pansy) luciferase (*hRluc*) is used as an internal control for cell density, transfection efficiency, lysis efficiency, lysate amount and measurement efficiency and is expressed under the control of a cytomegalovirus (CMV) promoter in the reporter vector pGL4.75-CMV-hRluc (Promega).

HEK293T cells were used to test the transactivation of the firefly luciferase reporter constructs by selected cDNAs while luciferase assays in the pancreatic beta cell lines aimed at the cell-intrinsic activation of the reporter constructs by the endogenous proteins. HEK293T, Min6b1 or Ins-1 832/13 cells were grown their respective cell culture medium in 96-well plates until they reached 70 % confluency. In combined RNA interference and transactivation assays, the cells were treated with the siRNAs (Dharmacon) 48 h before their transfection with the reporter vectors (see **PART B3.1+3.2**). The transfection with the reporter vectors was done in total volume of 100 µl per well using 50 µl of standard cell culture medium (see **PART B4**) and 50 µl of transfection mix per well. The transfection mix contained the plasmids, Opti-MEM Reduced Serum Medium (Life Technologies) and 0.5 µl of Lipofectamine 2000 reagent (Invitrogen) and was prepared following the manufacturer's recommendations. The used reporter and expression vectors are listed in **TABLE B5**. HEK293T cells were transfected with 0.185 ng of pGL4.75, 125 ng of a pGL4.23 construct and 10 ng per used expression vector. Min6b1 cells and Ins-1 832/13 cells were transfected with 0.2 ng of pGL4.75 and 125 ng of the pGL4.23 construct. In *mMlxipl* intron 1 transactivation assays, the cells were cultivated for 48 h before the analysis. For *mMlxipl* exon 1b transactivation assays, the cell culture medium part in the transfection was replaced by low glucose medium and the medium was changed to low or high glucose medium 24 h after transfection (see **PART B4**) and 24 h prior to the analysis.

On the analysis day, the cells were washed once with 1x PBS, and lysed with 20 µl of 1x PLB buffer (Promega) for 15 min at RT. 10 µl of the lysate was transferred to a 96-well microplate (Berthold Technologies). Luminescence measurement was conducted using the plate-reading luminometer Centro XS³ LB960 (Berthold Technologies) equipped with two reagent injectors. 25 µl of Luciferase Assay Reagent II (LARII) was injected into the well, solution was mixed for 2 s and the light was measured for 10 s. Subsequently, 25 µl of Stop & Glo Reagent was injected into the same well, solution was mixed for 2 s and the light was measured for 10 s. Samples were run in technical triplicates. The *Renilla* luminescence values were used to normalize the firefly luminescence values. Two-way ANOVA with Tukey's and Sidak's post-hoc multiple comparison test was used to determine whether the

observed expression changes are affected significantly by the two tested factors (p≥0.05 not significant, p<0.05 *, p<0.001 **, p<0.0001 ***, p<0.00001 ****). Relative luciferase activity is represented plus standard deviation.

TABLE B5: List plasmid backbones, expression and reporter vectors used in transactivation assays.

plasmid backbone		**...produced by...**
pcDNA1		commercial
pcDNA3		commercial
pCMV-Tag2A		commercial
pHD		Kaestner *et al.,* 1999
expression vector	**...that encodes...**	**...produced by...**
pcDNA3-mE2F1	*Mus musculus* E2F1	Fajas *et al.,* 2002
pHD-mFoxa2	*Mus musculus* FOXA2	Kaestner *et al.,* 1999
pcDNA3-mMlxipl α	*Mus musculus* MLXIPL α	Julia Grans
pcDNA3-mMlxipl β	*Mus musculus* MLXIPL β	Julia Grans
pcDNA3-mMlx	*Mus musculus* MLX	Julia Grans
pcDNA1-HA-mNeuroD1	*Mus musculus* NEUROD1 (N-terminal HA-tag)	Perrine Strasser
pcDNA3-mNkx2.2	*Mus musculus* NKX2.2	Céline Lapp, Julia Grans
pCMV-Tag2A-Flag-3HA-mRfx6	*Mus musculus* RFX6 (N-terminal Flag-, 3HA-tag)	Perrine Strasser
pcDNA3-hRfx6	*Homo sapiens* RFX6	Céline Ziegler-Birling, Julia Grans
pcDNA3-hRfx6	*Homo sapiens* RFX6 (p.E6K, p.T688A)	Mégane Denu, Aline Meunier
pcDNA3-hRfx6 p.F294V	*Homo sapiens* RFX6 (p.E6K, p.F294V, p.T688A)	Mégane Denu, Aline Meunier
pcDNA3-hRfx6 p.R181Q	*Homo sapiens* RFX6 (p.E6K, p.R181Q, p.T688A)	Mégane Denu, Aline Meunier
pcDNA3-hRfx6 p.R578P	*Homo sapiens* RFX6 (p.E6K, p.R578P, p.T688A)	Mégane Denu, Aline Meunier
pcDNA3-hRfx6 p.S217P	*Homo sapiens* RFX6 (p.E6K, p.S217P, p.T688A)	Mégane Denu, Aline Meunier
pcDNA3-hRfx6 p.V506G	*Homo sapiens* RFX6 (p.E6K, p.V506G, p.T688A)	Mégane Denu, Aline Meunier
reporter vector	**...that encodes...**	**...produced by...**
pGL4.75-CMV-hRluc	*Renilla reniformis* luciferase	commercial (Promega)
pGL4.23-minP-luc2	*Photinus pyralis* luciferase	commercial (Promega)
pGL4.23-minP-luc2 mMlxipl intron 1 606bp	*Photinus pyralis* luciferase	Julia Grans
pGL4.23-minP-luc2 mMlxipl intron 1 606bp Δxbox 2	*Photinus pyralis* luciferase	Julia Grans
pGL4.23-minP-luc2 mMlxipl intron 1 821bp	*Photinus pyralis* luciferase	Julia Grans
pGL4.23-minP-luc2 mMlxipl intron 1 821bp Δxbox 2	*Photinus pyralis* luciferase	Julia Grans
pGL4.23-minP-luc2 mMlxipl intron 1 1165bp	*Photinus pyralis* luciferase	Julia Grans
pGL4.23-minP-luc2 mMlxipl intron 1 1165bp Δxbox 2	*Photinus pyralis* luciferase	Julia Grans
pGL4.23-minP-luc2 mMlxipl exon 1b	*Photinus pyralis* luciferase	Julia Grans

5. CRISPR/Cas9 method

5.1. Integration of sgRNAs into the Cas9n expression vectors pX461 and pX462

The used CRISPR (Clustered Regularly Interspaced Short Palindromic Sequences)/Cas9 (CRISPR associated protein 9) eukaryotic expression vectors pX461-Cas9n-EGFP (pSpCas9n(BB)-2A-GFP, Ran *et al.,* 2013a; Ran *et al.,* 2013b) and pX462-Cas9n-Puro (pSpCas9n(BB)-2A-Puro version 1.0, Ran *et al.,* 2013a; Ran *et al.,* 2013b) ordered from Addgene (**FIGURE B6**) encode a mutated version of the *Streptococcus pyogenes* endonuclease Cas9 called Cas9 nickase (Cas9n, Jinek *et al.,* 2012). The Cas9 contains an HNH endonuclease domain cleaving the DNA strand that is complementary to the guide RNA (gRNA), and a RuvC-like endonuclease domain that cuts the non-complementary strand. The point mutation c.29A>C p.D10A in the Cas9 coding sequence (**FIGURE B6A+B**) inactivates the RuvC-like endonuclease domains that cannot cut the DNA anymore (Jinek *et al.,* 2012).This fact is used in the

double nicking RNA-guided Crispr/Cas9 technique (Jinek *et al.,* 2012; Ran *et al.,* 2013a; Ran *et al.,* 2013b) in which two single guide RNAs (sgRNA) direct two Cas9n proteins to two tightly positioned target sequences on opposite strands of the target gene. Both Cas9n molecules cut the complementary strand, thus producing two single strand breaks. The proximity of the two single strand breaks induces a double strand break-like repair mechanism by non-homologous end joining creating insertions and deletions (INDEL mutations). This method is supposed to minimize the off-target effect because two sgRNAs must recognize simultaneously their target sequences to induce Cas9n cleavage (Ran *et al.,* 2013a; Ran *et al.,* 2013b).

The Cas9n encoded by pX461 is fused to an N-terminal 3xFlag tag and a nuclear localization signal (NLS) and a C-terminal self-cleaving 2A peptide and EGFP (**FIGURE B6A**). In pX462, the EGFP is replaced by a puromycin resistance gene (**FIGURE B6B**). It is important to note that the puromycin resistance gene in the used version of pX462 contains the point mutation c.497G>A p.R165H (**FIGURE B6B+C**) that makes the puromycin selection less efficient in some cell lines. The Cas9n is expressed under the control of a CMV enhancer, the chicken β actin promoter and a chicken β actin / mouse minute virus hybrid intron in bot pX461 and pX462.

The gene-specific sgRNA pairs (**TABLE B6**) were designed with the online tools http://crispr.mit.edu or http://crispor.org. The genomic sequence that is complementary to the sgRNA, borders on an NGG motif named protospacer adjacent motif (PAM, Jinek *et al.,* 2012). The sgRNA can be inserted in-between the U6 promoter and the gRNA scaffold with PolIII terminator site into pX461 and pX462 using the restriction enzyme *Bbs*I as described in the protocol by Ran *et al.,* 2013b. In brief, the sgRNA sequence was ordered as a sense and an anti-sense primer with *Bbs*I complementary overhangs at the 5'end of the forward primer and at the 3'end of the reverse primer (**TABLE B7**, **FIGURE B6D**). The primers were annealed, and the duplex was ligated with the *Bbs*I linearized plasmid. Correct insertion was verified by sequencing (forward primer: `5'gagggcctatttcccatgat`; reverse primer: `5'gggcgtacttggcatatgat`).

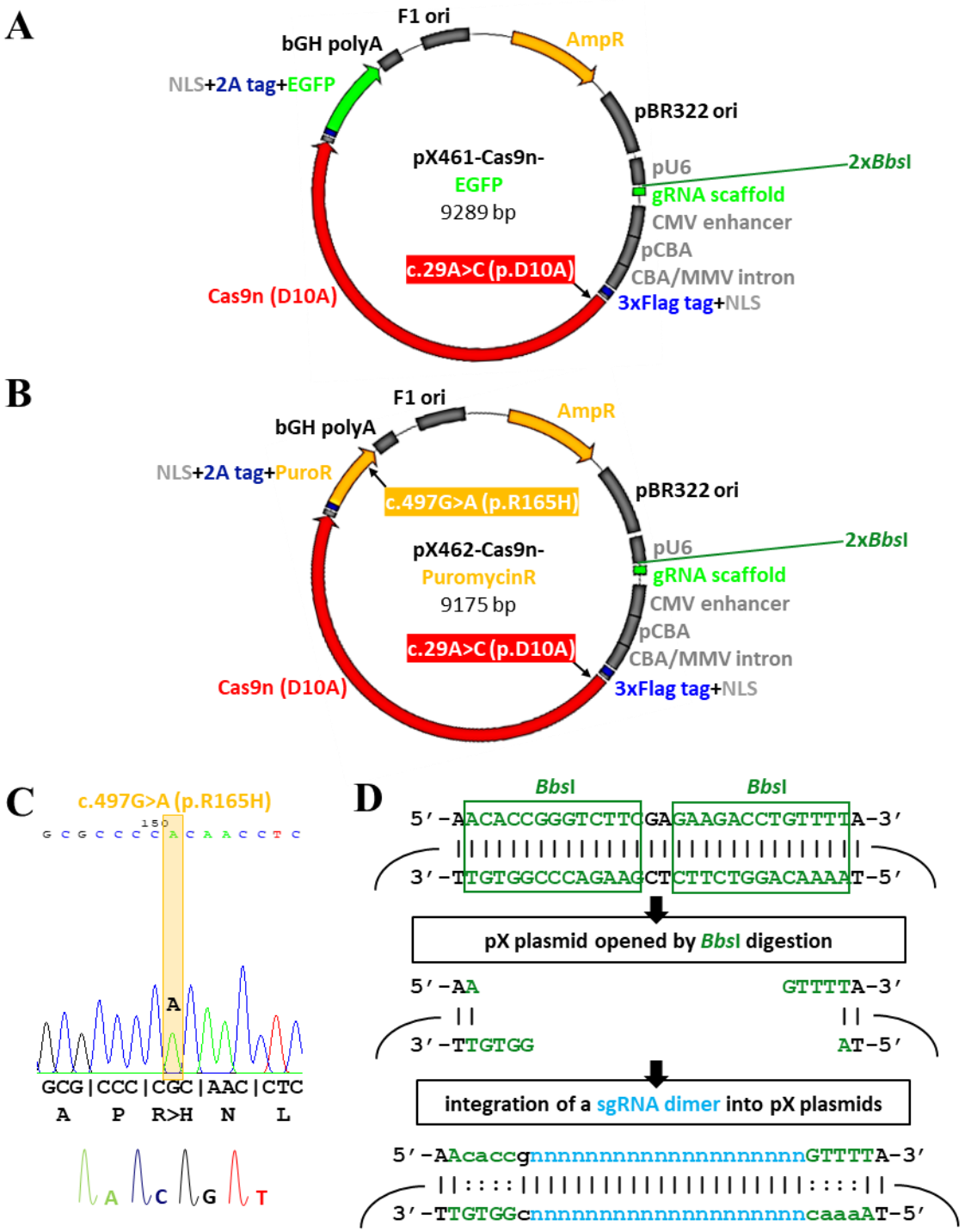


FIGURE B6: Integration of sgRNAs into pX461 and pX462. (A, B) and Map of the commercial CRISPR/Cas9 vectors pSpCas9n(BB)-2A-GFP **(A)** named pX461 and pSpCas9n(BB)-2A-Puro version 1.0 **(B)** named pX462 hereafter, both published in Ran *et al.*, 2013a and Ran *et al.*, 2013b. Both plasmids encode a mutated Streptococcus pyogenes Cas9n (c.29A>C, p.D10A) fused to an N-terminal 3xFlag tag and an NLS and a C-terminal NLS, 2A tag and EGFP (pX461) or a puromycin selection gene (PuroR, pX462). The Cas9n is expressed under the control of the chicken beta actin promoter (pCBA). **(C)** Point mutation in pX462 in puromycin resistance gene (c.497G>A, p.R165H) that makes puromycin selection less efficient in some cell lines. **(D)** *Bbs*I restriction sites in pX461 and pX462 allowing the ligation with a sgRNA primer dimer with compatible overhangs.

TABLE B6: List of single guide RNA pairs used in the CRISPR/Cas9 experiments. The single guide RNA (sgRNA) sequence is highlighted in light blue and the protospacer adjacent motif (PAM) in orange. (+) and (-) indicate that the sgRNA binds to the 5'3' strand and 3'5' strand, respectively. The x in the columns pX461 and pX462 means that the sgRNAs in this line were cloned into this plasmid (one sgRNA per plasmid).

target	sgRNA pair (number + sequence 5'3' + strand)						pX461	pX462	designed with
m/rMlxipl exon 1a	#10	ATCCGCCAGCGCGCGCGCCAtgg	(-)	#15	CCGTGAACTTGCAGGTCCCCcgg	(+)	x	x	http://crispr.mit.edu
mMlxipl exon 1a	#2	CCGGGCTAGGGACGACCCGGggg	(-)	#6	CTCGGACTCGGATACGGACTtgg	(+)		x	http://crispr.mit.edu
mMlxipl exon 6	#2	AGCTGTTCGCACCATCGCTCtgg	(-)	#10	CGTGGTGCCTGTGCTGCTTGggg	(+)		x	http://crispr.mit.edu
mMlxipl exon 13	#5	CTAGGATTCGACACCCTGCAcgg	(+)	#6	GAACCGCCTCTTCTGCTCCGcgg	(-)		x	http://crispr.mit.edu
rMlxipl exon 6	#7	CCTCTTCAGCGGTGGGATCCtgg	(+)	#15	TCAGCCATGGGGCCAAAGCTggg	(-)	x	x	http://crispr.mit.edu
rMlxipl exon 6	#26	TTGAAGATGTGGGCGTCTGAggg	(-)	#79	GCCCCCCAGCTTTGGCCCCAtgg	(+)		x	http://crispor.org
rMlxipl exon 10	#7	CCTCAAACTGTCCTAAGCCGggg	(+)	#18	GGTGGACACTCAGTGCCCCCggg	(-)	x	x	http://crispr.mit.edu
mRfx6 exon 1	#1	ACTTCCCTCCGAGATCGCGGagg	(+)	#28	CGAAAGCTTCCTCCAGTTCCcgg	(-)	x	x	http://crispr.mit.edu
mRfx6 exon 1	#4	CACTCCTCCGCGATCTCGGAggg	(-)	#10	TGCGCGCAGCTGCTTGGCAAggg	(+)		x	http://crispr.mit.edu
m/rRfx6 exon 3	#2	CATAAAGAATGCACCGTGGGagg	(-)	#19	TTTCTGCAGGAAAGAGAAACtgg	(+)		x	http://crispr.mit.edu
rRfx6 exon 2	#2	GAGCATCACCCAGATCATGAagg	(+)	#3	GAGCTGTGGGTCGGTGGCTTtgg	(-)	x	x	http://crispr.mit.edu
rRfx6 exon 3	#1	GAGCATAAAGAATGCACCGTggg	(-)	#10	TTTCTGCAGGAAAGAGAAACtgg	(+)	x	x	http://crispr.mit.edu
rRfx6 exon 3				#118	CCAGCCTGTGCAGCCACCTTtgg	(+)		x	http://crispor.org

TABLE B7: List of primer sequences used to clone the single guide RNAs into pX461 and pX462. The single guide RNA (sgRNA) sequence is highlighted in light blue and the *Bbs*I complementary overhangs in green.

target	sgRNA	fwd primer 5'3'	rev primer 5'3'
m/rMlxipl exon 1a	#10	caccgatccgccagcgcgcgcgcca	aaactggcgcgcgcgctggcggatc
	#15	caccgccgtgaacttgcaggtcccc	aaacggggacctgcaagttcacggc
mMlxipl exon 1a	#2	caccgccgggctagggacgacccgg	aaacccgggtcgtccctagcccggc
	#6	caccgctcggactcggatacggact	aaacagtccgtatccgagtccgagc
mMlxipl exon 6	#2	caccgagctgttcgcaccatcgctc	aaacgagcgatggtgcgaacagctc
	#10	caccgcgtggtgcctgtgctgcttg	aaaccaagcagcacaggcaccacgc
mMlxipl exon 13	#5	caccgctaggattcgacaccctgca	aaactgcagggtgtcgaatcctagc
	#6	caccggaaccgcctcttctgctccg	aaaccggagcagaagaggcggttcc
rMlxipl exon 6	#7	caccgcctcttcagcggtgggatcc	aaacggatcccaccgctgaagaggc
	#15	caccgagctttggccccatggctga	aaactcagccatggggccaaagctc
rMlxipl exon 6	#26	caccgtcagacgcccacatcttcaa	aaacttgaagatgtgggcgtctgac
	#79	caccggcccccccagctttggcccca	aaactggggccaaagctgggggccc
rMlxipl exon 10	#7	caccgcctcaaactgtcctaagccg	aaaccggcttaggacagtttgaggc
	#18	caccgggggcactgagtgtccacc	aaacggtggacactcagtgccccc
mRfx6 exon 1	#1	caccgacttcccctccgagatcgcgg	aaaccccgcgatctcggagggaagtc
	#28	caccgcgaaagcttcctccagttcc	aaacggaactggaggaagctttcgc
mRfx6 exon 1	#4	caccgcactcctccgcgatctcgga	aaactccgagatcgcggaggagtgc
	#10	caccgtgcgcgcagctgcttggcaa	aaacttgccaagcagctgcgcgcac
m/rRfx6 exon 3	#2	caccgcataaagaatgcaccgtggg	aaaccccacggtgcattctttatgc
	#19	caccgtttctgcaggaaagagaaac	aaacgtttctctttcctgcagaaac
rRfx6 exon 2	#2	caccggagcatcacccagatcatga	aaactcatgatctgggtgatgctcc
	#3	caccgaagccaccgacccacagctc	aaacgagctgtgggtcggtggcttc
rRfx6 exon 3	#1	caccgacggtgcattctttatgctc	aaacgagcataaagaatgcaccgtc
	#10	caccgtttctgcaggaaagagaaac	aaacgtttctctttcctgcagaaac
	#118	caccgccagcctgtgcagccacctt	aaacaaggtggctgcacaggctggc

5.2. Establishment of stable knockout Min6b1 and Ins-1 832/13 cell lines

To generate Min6b1 or Ins-832/13 knockout cell lines (**FIGURE B7**, **FIGURE B8**, **FIGURE B9**), the cells were seeded into 6-well plates, and, at 70 % confluency, transfected with 1.5 µg DNA per Cas9n sgRNA expression vector (3 µg in total for a sgRNA pair). Control cells were transfected with 3 µg of the Cas9n expression vector without sgRNA. The transfection was carried out following the manufacturer's recommendations using 500 µl of Opti-MEM Reduced Serum Medium (Life Technologies), 2 ml standard cell culture medium (see **PART B4**) and 10 µl of Lipofectamine 2000 reagent (Invitrogen) per well. After 48 h, transfected cells were selected by adding puromycin or

sorting the EGFP-positive cells depending on the used plasmid (**FIGURE B7A**, **FIGURE B8A**, **FIGURE B9A**).

Min6b1 or Ins-1 832/13 cells transiently transfected with pX462 were incubated for 72 h with 1.75 µg/ml or 3 µg/ml of puromycin, respectively (**FIGURE B7A+B**, **FIGURE B8A+B**). To remove the dead cells, the cells were washed several times with 1x PBS and the surviving cells (**FIGURE B7B**, **FIGURE B8B**) were cultured in standard cell culture medium for two weeks with a medium refreshment every three to four days. After 14 days, single colonies were hand-picked using the stereomicroscope SMZ645 (Nikon) and transferred into a 10-cm cell culture dish (**FIGURE B7B**, **FIGURE B8B**). Min6b1 and Ins-1 832/13 colonies were cultured for three and two weeks, respectively. Well grown colonies were hand-picked (**FIGURE B7B**, **FIGURE B8B**), transferred into 24-well plates and cell aggregates were separated mechanically by pipetting up and down. The remaining colonies in the 10-cm dish were used for phenol-chloroform DNA extraction (see **PART B2.1**) and T7EI assay (see **PART B5.3**). After one week, the isolated clones were split into a new 24-well plate using trypsin, and, after another week, the cells could be split into two or three wells for cell maintenance and analysis (**FIGURE B7A** and **FIGURE B8A**). To subclone heterogenous cell lines, 50,000 Min6b1 cells and 5,000 Ins-1 832/13 cells were plated per 35 mm dish, cultured for two weeks, and the cloning procedure was resumed at the step the cells were first picked by hand in the cloning protocol described above.

Ins-1 832/13 cells transfected with pX461 were selected by Fluorescence Activated Cell Sorting (FACS) of live cells (**FIGURE B9**). Cells were dissociated with 0.5 ml 2.5 ‰ trypsin for 5 min at 37 °C. Trypsin was inactivated with 1 ml cell culture medium. Cells were pelleted at 300 x g for 2 min and carefully resuspended in 750 µl of cell culture medium in a 5 ml FACS tube (BD) to prepare a single cell solution (3,000,000 to 4,000,000 cells / ml). FACS was performed by the flow cytometry facility of the institute. Non-transfected cells were used as a negative control for the GFP-positive population. GFP-positive cells (**FIGURE B9B**) were sorted into 96 well plates with 100 µl cell culture medium / well under sterile conditions using the FACS Aria Fusion or Aria II (BD) with a 100 µm nozzle (**FIGURE B9C**). Additionally, 20,000 GFP-positive cells were collected in a tube and plated into a 35-mm dish that was used for phenol-chloroform DNA extraction (see **PART B2.1**) and T7EI assay (see **PART B5.3**). In the post sorting (**FIGURE B9C**), it was verified that the correct cells had been selected in the actual sorting. The clones in the 96-well plates were split into a new 96-well plate using trypsin after 14 days of culture and into a 24-well plate after another 7 days. From day 48 onwards, could be split as needed to ensure cell maintenance and analysis (**FIGURE B9A**).

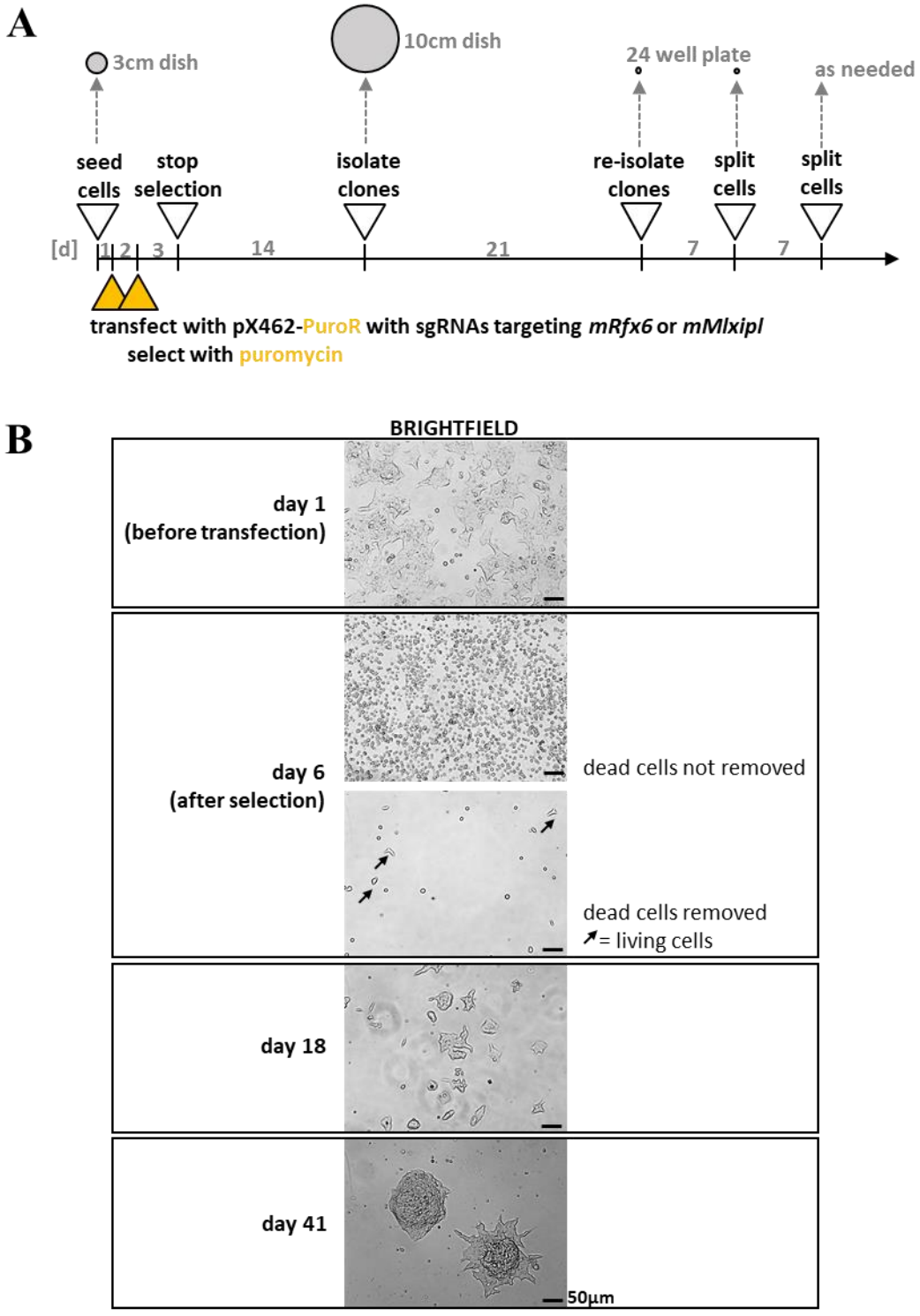


FIGURE B7: Protocol to establish stable Min6b1 knockout clones using puromycin. **(A)** Cloning protocol to obtain homozygous Min6b1 knockout cell lines. On day 0, cells are seeded in 3 cm dishes. They are transfected on the first day with two pX462 plasmids (puromycin selection gene) each containing one sgRNA of a locus-specific pair on day 1. Puromycin selection is induced on day 3 and stopped on day 6. On day 20, clones are picked and transferred into a 10 cm dish to form well isolated colonies. On day 41, clones are re-picked and transferred into 24 well plates. In this step, cell aggregates are separated by pipetting. On day 48, the clones are split again into a 24 well plate. From day 55 onwards, clones can be split as needed. **(B)** Brightfield images of the cells during the cloning process on day 1, 6, 18 and 41.

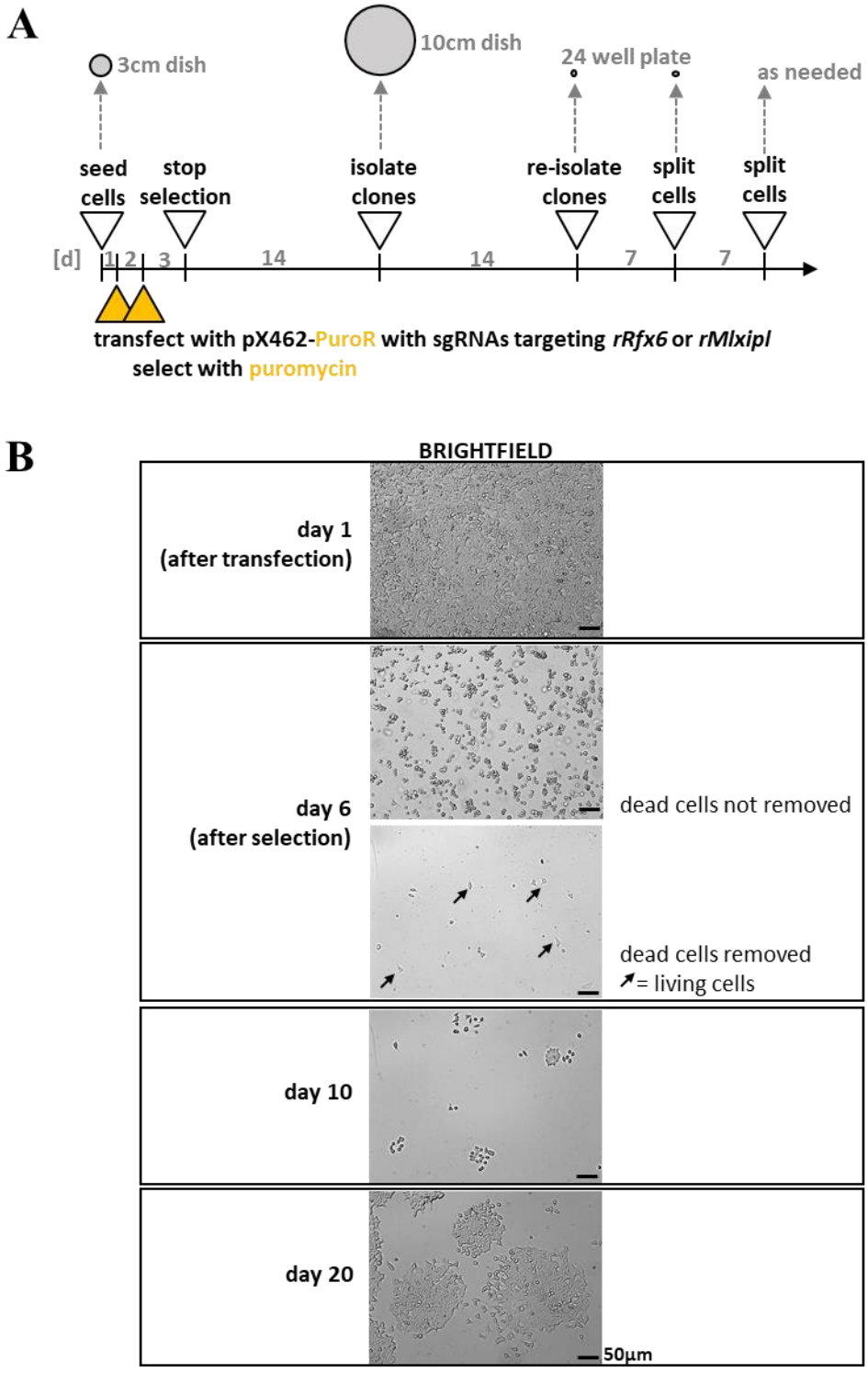


FIGURE B8: Protocol to establish stable Ins-1 832/13 knockout clones using puromycin. (A) Cloning protocol to obtain homozygous Ins-1 832/13 knockout clones. On day 0, cells are seeded in 3 cm dishes. They are transfected with two pX462 plasmids (puromycin selection gene) each containing one sgRNA of a locus-specific pair on day 1. Puromycin selection is induced on day 3 and stopped on day 6. On day 20, clones are picked and transferred into a 10 cm dish to form well isolated colonies. On day 34, clones are re-picked and transferred into 24 well plates. In this step, cell aggregates are separated by pipetting. On day 41, the clones are split again into a 24 well plate. From day 48 onwards, clones can be split as needed. **(B)** Brightfield images of the cells during the cloning process on day 1, 6, 10 and 20.

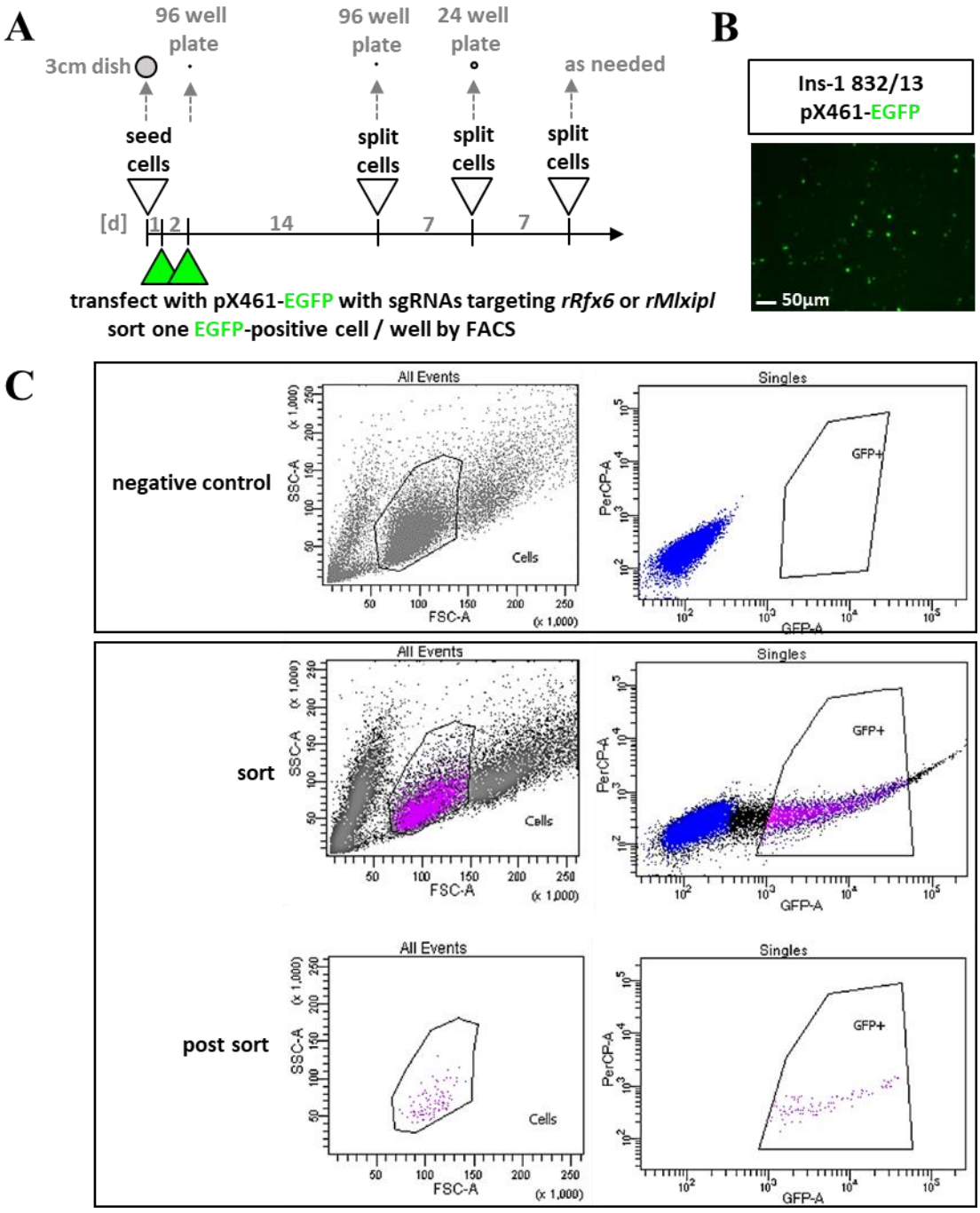


FIGURE B9: Protocol to establish stable Ins-1 832/13 knockout clones using EGFP. (A) Cloning protocol to obtain homozygous Ins-1 832/13 knockout clones. On day 0, cells are seeded in 3 cm dishes. They are transfected with two pX461 plasmids (EGFP selection gene) each containing one sgRNA of a locus-specific pair on day 1. On day 3, the cells are detached, resuspended and the EGFP-positive cells are sorted into 96 well plates by FACS. On day 17, the cells are split for the first time into a second 96 well plat, and, on day 24, into a 24 well plate. From day 31 on, cells can be split as needed. **(B)** Fluorescence image (green emission, excision at 488 nm, emission at 510 nm) of EGFP-positive Ins-1 832/13 cells transfected with pX461 plasmids (EGFP selection gene) on day 3 (before FACS). **(C)** FACS profiles (representative example samples for all used FACS samples) of the negative control (above), EGFP-positive sample (in the middle) and the post sort of this sample (below). The profiles on the left show the different cell populations using the FSC (forward scatter – cell size) and the SSC (side scatter – internal complexity of the cell). The line surrounds the single cells. The profiles on the right represent the EGFP-positive cells on the x-axis. The y-axis gives the absorption of PerCP-Cy5 as a negative control. For experiments, only cells with a medium EGFP signal were chosen.

5.3. T7EI assay to assess the efficiency of sgRNA/Cas9n-mediated INDEL production

The efficiency of the designed sgRNA pairs to induce target cleavage was assessed by the T7 Endonuclease I (T7EI) assay. T7EI binds to mismatches and extrahelical loops in double stranded DNA and cuts the DNA at these sites (Sadowski, 1971). This property is used in the T7EI assay to detect INDELs induced by the CRISPR/Cas9 technique. To perform this test, the targeted loci of the wild-type controls and the putative knockout samples were amplified by PCR (see **PART B2.3**, primers are listed in **TABLE B8**) and 250 ng of the purified PCR products were mixed with 1x NEB buffer 2 in a total volume of 10 µl. Two samples were prepared per PCR reaction. The DNA was denatured and hybridised in the following protocol: 10 min at 95 °C, 5 cycles at 95 °C (Δ2°C per cycle) for 1 s, 1 min at 85 °C, 34 cycles at 85 °C (Δ0.3°C per cycle) for 1 s, 1 min at 75 °C, 34 cycles at 75 °C (Δ0.3°C per cycle) for 1 s, 1 min at 65 °C, 34 cycles at 65 °C (Δ0.3°C per cycle) for 1 s, 1 min at 55 °C, 34 cycles at 55 °C (Δ0.3°C per cycle) for 1 s, 1 min at 45 °C, 34 cycles at 45 °C (Δ0.3°C per cycle) for 1 s, 1 min at 35 °C, 34 cycles at 35 °C (Δ0.3°C per cycle) for 1 s, 1 min at 25 °C. The T7EI+ sample was supplemented with 10 U of T7EI (10 U/µl, NEB). The T7EI was replaced by water in the T7EI- sample. Both samples were incubated for 1 h at 37 °C and the digestion was stopped with 20 mM EDTA and the fragmented DNA was analysed by agarose gel electrophoresis. Efficient target site cleavage and INDEL production are marked by the presence of cleavage products besides the PCR product itself.

TABLE B8: List of primer used for the T7EI assay.

gene	fwd primer	sequence 5'3'	rev primer	sequence 5'3'	T7EI assay	PCR product size
mRfx6	5' of exon 1	CTGATACTCAAGTCAAGGCTGACC	intron 1	GTTAACAGTGTTCCTGCTGAAACG	exon 1	823 bp
mRfx6	intron 2	AAGTCCATGTGTATCTACTCATTAAC	intron 3	GGTATACACATAGAGCAGTTCCAG	exon 3	408 bp
mMlxipl	5' of exon 1a	GCTCCAGGATCAGTGGTTTCACAC	intron 1	GATGGGACACAAATGATTCATGAGG	exon 1a	645 bp
mMlxipl	intron 5	GATAGGATGACCTTAGCCTTCCAG	intron 6	CAACATAAGCTGGAGGGACAGTG	exon 6	456 bp
mMlxipl	intron 12	GTGTAGACAACAACAAGGTACGTG	intron 13	GGTTGCTTTGCTCACCTGCGAG	exon 12	366 bp
rRfx6	intron 1	GGTGTGGAGGTCACTGGACTG	intron 2	CGTCTTTCATCAGGGTCAGAATAG	exon 2	409 bp
rRfx6	intron 2	CAACCAGAACAAAGAAAGTCTCATG	intron 3	TGAGCTCACATTAGATCTGGCAGC	exon 3	479 bp
rMlxipl	5' of exon 1a	TGGTTTCACACCAGGCTCATTGG	intron 1	GGAGCTAATGCCTAAGTTGTATCC	exon 1a	586 bp
rMlxipl	intron 5	GTCTCTTGTAGAAGCCTTCCACC	intron 6	GATAAGATATGCTCACCTGGAGAC	exon 6	323 bp
rMlxipl	intron 9	CAGCGGTAAATAGGAGCGAGAAG	intron 10	GTTCGTTCGGGTGAGGCTAAG	exon 10	332 bp

5.4. Screening methods

To screen the knockout clones, the targeted loci were amplified by PCR, analysed by agarose gel electrophoresis and sequenced (see **PART B2.3**, primers are listed in **TABLE B9**). Western blots were used to compare the level of the protein encoded by the targeted gene in wild-type and knockout clones (see **PART B2.6**, antibodies are listed in **TABLE B2**). As functional tests, the transactivation of the *mMlxipl* intron 1 reporter construct and the *mMlxipl* exon 1b reporter construct (see **PART B4.5**) was measured in putative *Rfx6* and *Mlxipl* knockout clones, respectively, and the expression of

RFX6 and MLXIPL target genes was determined by RT-qPCR (see **PART B2.2+2.4**, assay types are listed in **TABLE B1**).

TABLE B9: List of primer used for agarose gel electrophoresis and Sanger sequencing.

gene	fwd primer	sequence 5'3'	rev primer	sequence 5'3'	PCR product size	sequencing	sequence 5'3'
mRfx6	5' of exon 1	CTGATACTCAAGTCAAGGCTGACC	intron 1	GTTAACAGTGTTCCTGCTGAAACG	823 bp	fwd	CTGATACTCAAGTCAAGGCTGACC
						rev	GTTAACAGTGTTCCTGCTGAAACG
mRfx6	5' of exon 1	CTGATACTCAAGTCAAGGCTGACC	exon 2	TCCTTCATCATCTGCGTGATGCTC	1070 bp	fwd	CTGATACTCAAGTCAAGGCTGACC
						rev	TCCTTCATCATCTGCGTGATGCTC
mRfx6	5' of exon 1	GGTACCTGTTCACCCTGTCC	exon 1	CAGGTAAGCACTGTCCTCAG	302 bp	fwd	GGTACCTGTTCACCCTGTCC
						rev	CAGGTAAGCACTGTCCTCAG
mRfx6	intron 2	AAGTCCATGTGTATCTACTCATTAAC	intron 3	GGTATACACATAGAGCAGTTCCAG	408 bp	fwd	AAGTCCATGTGTATCTACTCATTAAC
mMlxipl	5' of exon 1a	GCTCCAGGATCAGTGGTTTCACAC	intron 1	GATGGGACACAAATGATTCATGAGG	645 bp	fwd	GCTCCAGGATCAGTGGTTTCACAC
						rev	GATGGGACACAAATGATTCATGAGG
mMlxipl	5' of exon 1a	TGTCTAGGAATACATATGGATCTCAG	intron 1	GATGGGACACAAATGATTCATGAGG	1091 bp	fwd	TGTCTAGGAATACATATGGATCTCAG
mMlxipl	5' of exon 1a	CGAGTGTTACACTTACAGAATCACC	intron 1	GATGGGACACAAATGATTCATGAGG	826 bp	fwd	CGAGTGTTACACTTACAGAATCACC
						fwd	GGACCTATCAGGAGAACAGG
						rev	GTGTCCGCTGTGGATGACC
mMlxipl	5' of exon 1a	GGACCTATCAGGAGAACAGG	exon 1a	GTGTCCGCTGTGGATGACC	253 bp	fwd	GGACCTATCAGGAGAACAGG
						rev	GTGTCCGCTGTGGATGACC
rRfx6	intron 1	GGTGTGGAGGTCACTGGACTG	intron 2	CGTCTTTCATCAGGGTCAGAATAG	409 bp	fwd	GGTGTGGAGGTCACTGGACTG
rRfx6	intron 2	CAACCAGAACAAAGAAAGTCTCATG	intron 3	TGAGCTCACATTAGATCTGGCAGC	479 bp	fwd	CAACCAGAACAAAGAAAGTCTCATG
rRfx6 mRNA	exon 1	GAACTGGAGGAAACCTTCGTGC	exon 4	GACCTCTTGTTCCAAGTCTTCTTG	529 bp	fwd	GAACTGGAGGAAACCTTCGTGC
rMlxipl	5' of exon 1a	TGGTTTCACACCAGGCTCATTGG	intron 1	GGAGCTAATGCCTAAGTTGTATCC	586 bp	fwd	TGGTTTCACACCAGGCTCATTGG
						rev	GGAGCTAATGCCTAAGTTGTATCC

6. Data sets

The following additional data sets were used in this thesis: i) RNA sequencing data of islets isolated from 8 to 10 weeks old wild-type (*mRfx6* fl/fl) and mice with a beta cell-specific knockout of *mRfx6* (*mRfx6* fl/fl; Ins1-CreERT2 mice), both fed ad libitum, five days after tamoxifen treatment (GSE59622, Piccand *et al.,* 2014); ii) ChIP sequencing data with an anti-HA antibody on Min6b1 cells transfected with a 3HA-mRfx6 cDNA (GSE62844, Piccand *et al.,* 2014), with an anti-RFX6 antibody on Min6b1 cells and with an anti-RFX6 antibody on islets isolated from wild-type mice (Perrine Strasser, unpublished data); iii) RNA sequencing data of Ins-1E cells incubated for 0, 1, 2, 4 or 12 h in 25 mM glucose medium and of Ins-1E cells treated with siControl or siMlxipl and incubated for 0 or 12 h in 25 mM glucose medium (GSE81628, Schmidt *et al.,* 2016); iv) ChIP sequencing data with an anti-MLXIPL antibody on Ins-1E cells incubated for 0, 2 or 12 h in 25 mM glucose (GSE81628, Schmidt *et al.,* 2016).

C) Results and discussion

1. RFX6 activates the transcription of *mMlxipl* via a conserved xbox in *mMlxipl*'s first intron

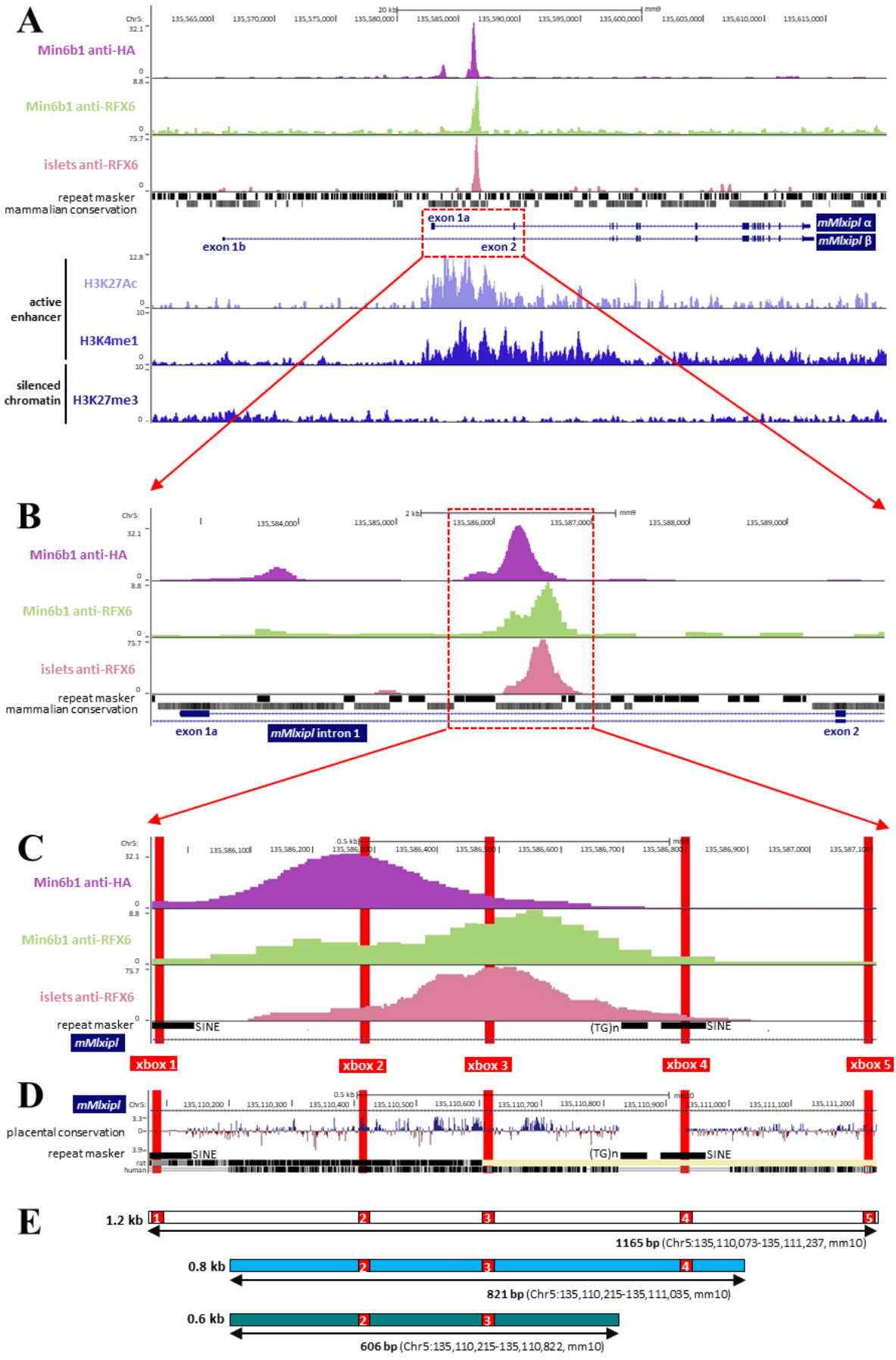
Previous work of my host lab revealed that the gene *mMlxipl* is a direct target gene of mRFX6 in mouse pancreatic beta cells. In the ChIP sequencing experiments performed to identify RFX6-bound genes in the pancreatic beta cell line Min6b1 (GSE62844, Piccand *et al.,* 2014), binding of mRFX6 occurred on the first intron of *mMlxipl,* 3500 bp downstream of the transcription start site and the core promoter region. Furthermore, RNA sequencing experiments on islets from *mRfx6* Δbeta mice demonstrated that the beta cell-specific knockout of *mRfx6* in mice caused the loss of *mMlxipl* expression. In wild-type mice, an average of 1595 normalized reads were measured, while there were only 84 normalized reads in the knockout (Log2 FC=-4, GSE59622, Piccand *et al.,* 2014).

Based on these findings, the first task of my thesis was to characterize more precisely the enhancer activity of the *mMlxipl* intron 1 region by identifying the RFX6 binding sites critical for *mMlxipl* regulation by RFX6. In addition, we wanted to find out if other transcription factors also interact with *mMlxipl* intron 1 and could thus be involved in *Mlxipl* regulation by RFX6 causing a synergistic or antagonistic effect. To address all these questions, we designed expression vectors and classical transactivation assays with wild-type and mutated versions of the *mMlxipl* intron 1 region.

1.1. RFX6 binds to *mMlxipl* intron 1 in murine islets and in the murine beta cell line Min6b1

Perrine Strasser, a former student and Postdoc of the lab, carried out three different ChIP sequencing experiments: i) an anti-HA ChIP on Min6b1 cells transfected with a 3HA-mRfx6 cDNA (Piccand *et al.,* 2014); ii) an anti-RFX6 ChIP on non-transfected Min6b1 cells (Perrine Strasser, unpublished data) and iii) an anti-RFX6 ChIP on islets of CD1 mice (Perrine Strasser, unpublished data). All three ChIP experiments revealed a peak on the first intron of *mMlxipl* (**FIGURE C1**).

The comparison of the ChIP sequencing data with published data of activating and silencing histone marks in pancreatic islets of Black6 mice (Avrahami *et al.,* 2015) revealed that the histones 3 in the peak region are predominantly acetylated at lysine 27 (H3K27Ac) and monomethylated at lysine 4 (H3K4me1) but not trimethylated at lysine 27 (H3K27me3). H3K27me3 marks transcriptionally silenced chromatin (Schubert *et al.,* 2006) whereas H3K27Ac and H3K4me1 are found on active enhancers (Avrahami *et al.,* 2015; Heintzman *et al.,* 2007). The combination of H3K27Ac and H3K4me1 at the peak region suggests that this region is an active enhancer region (**FIGURE C1A**).


A
Min6b1 anti-HA
Min6b1 anti-RFX6
islets anti-RFX6
repeat masker
mammalian conservation
exon 1a
exon 1b
exon 2
mMlxipl α
mMlxipl β
active enhancer
H3K27Ac
H3K4me1
silenced chromatin
H3K27me3
B
Min6b1 anti-HA
Min6b1 anti-RFX6
islets anti-RFX6
repeat masker
mammalian conservation
exon 1a
mMlxipl intron 1
exon 2
C
Min6b1 anti-HA
Min6b1 anti-RFX6
islets anti-RFX6
repeat masker
SINE
(TG)n
SINE
mMlxipl
xbox 1
xbox 2
xbox 3
xbox 4
xbox 5
D
mMlxipl
placental conservation
repeat masker
SINE
(TG)n
SINE
E
1.2 kb
1165 bp (Chr5:135,110,073-135,111,237, mm10)
0.8 kb
821 bp (Chr5:135,110,215-135,111,035, mm10)
0.6 kb
606 bp (Chr5:135,110,215-135,110,822, mm10)

← **FIGURE C1: RFX6 binds to the first intron of *mMlxipl* gene. (A)** ChIP sequencing data showing the binding of RFX6 on *mMlxipl* (anti-HA ChIP on Min6b1 expressing 3HA-mRFX6, published in Piccand *et al.,* 2014; anti-RFX6 ChIP on Min6b1 and anti RFX6 ChIP on islets of CD1 mice, unpublished data of Perrine Strasser) and the active (H3K27Ac and H3K4me1) and poised (H3K27me3) histone marks (published data generated in islets of Black6 mice, Avrahami *et al.,* 2015) in the *mMlxipl* gene (UCSC genome browser alignment done by Tao Ye). **(B)** Zoom in on *mMlxipl* intron 1. **(C)** Zoom in on RFX6 peak region in *mMlxipl* intron 1 (Chr5:135,585,943-135,587,107, mm9, 1165 bp) and position of the five RFX binding motifs (xbox, highlighted in red). **(D)** Conservation of the 1165 bp region in *Mlxipl* intron 1 (Chr5:135,110,073-135,111,237, mm10) with the rat genome (rn5) and the human genome (hg19). **(E)** Maps showing the position and length of the genomic DNA fragments that were cloned into pGL4.23-minP-luc (see **FIGURE C2**). The 821 bp fragment corresponds to the islet anti-RFX6 peak.

TABLE C1: The second of the five xbox motifs identified in a 1165 bp window in *Mlxipl* intron 1 is conserved between mouse, rat and human. First line: Sequence of the five xbox motifs (mm10) based on the published xbox consensus sequence of Laurençon *et al.,* 2007 and the corresponding sequences in the rat (rn5) and human (hg19) genome. Second line: Direction of the xbox motif and number of bases in the middle of the xbox (N1-3). Third to sixth line: Comparison of the five xbox motifs with the RFX6 binding motif defined in the anti-HA or anti-RFX6 Min6b1 ChIP sequencing data, the anti-RFX6 murine islets ChIP sequencing data (Perrine Strasser, unpublished data) and a human RFX6 binding motif identified in the human islet-like cell line TC-YIK (Lizio *et al.,* 2015).

	xbox 1	xbox 2	xbox 3	xbox 4	xbox 5
xbox conservation mouse (mm10) rat (rn5) human (hg19)					
consensus motif Laurençon *et al.,* 2007 RYNNYYN (1-3) RRNRAC R=A/G and Y=C/T	N(2) 5′-3′	N(1) 3′-5′	N(3) 3′-5′	N(2) 3′-5′	N(2) 5′-3′ 3′-5′
de novo motif (mouse) Min6b1 anti-HA ChIP		3′-5′ direction			
de novo motif (mouse) islets anti-RFX6 ChIP		3′-5′ direction			
de novo motif (mouse) Min6b1 anti-RFX6 ChIP		3′-5′ direction			
de novo motif (human) TC-YIK anti-RFX6 ChIP Lizio *et al.,* 2015		3′-5′ direction			

1.2. One of the five xbox motifs in *mMlxipl* intron 1 is conserved in mouse, rat and human

In a 1165 bp long region around the anti-RFX6 peaks (Chr5:135,585,943-135,587,107 on mm9 and Chr5:135,110,073-135,111,237 on mm10), we searched for RFX binding motifs, called xbox, with the consensus sequence 5'RYNNYYN(1-3)RRNRAC (with R=A/G and Y=C/T; Laurençon *et al.,* 2007) and we identified five different xbox motifs (**FIGURE C1C+D**, **TABLE C1**) which we numbered from 1 to 5. The first xbox is oriented in the forward direction, xbox 2, 3 and 4 are oriented in the reverse direction. The fifth xbox is a bidirectional motif (**TABLE C1**). According to the literature, the orientation of an enhancer element does not seem to have any influence on gene regulation (Lis and Walther, 2016).

The main peak sequence of the anti-HA and of the two anti-RFX6 ChIP does not colocalize with any repeat sequences, while the 5'extension of the anti-HA peak contains a Short Interspersed Nuclear Element (SINE), and the 3'extensions of the anti-RFX6 peaks a (TG)n repeat and a SINE. The first and the fourth xbox motif colocalize with the upstream and downstream SINE, respectively. Interestingly, compared to the peaks in the anti-RFX6 ChIP experiments, the peak maximum in the anti-HA ChIP was shifted by approximately 250 bp towards *mMlxipl* exon 1a (**FIGURE C1B+C**). The Min6b1 cells for the anti-HA ChIP expressed at the same time a mRFX6 protein with an N-terminal 3HA-tag as well as the untagged endogenous protein. The HA tag is derived from a glycoprotein found on the surface of human influenza virus (Green *et al.,* 1982; Wilson *et al.,* 1984) and it is very small (HA: 9 aa, 3HA: 27 aa) to prevent steric hindrance. The second xbox is centrally located under the maximum of the anti-HA peak and xbox 3 under the maxima of the anti-RFX6 peaks (**FIGURE C1C+D**). The shift of the peak could be because the tagged protein has bound at a different site than the untagged protein. This suggests that the 3HA-mRFX6 preferentially binds to xbox 2 while xbox 3 is the preferred fixation site of endogenous mRFX6. The N-terminal tag in the 3HA-mRFX6 fusion protein, although it is very small, could have affected proper protein function leading to altered binding. Furthermore, RFX proteins bind to xbox motifs as homo- or heterodimers with a second RFX transcription factor (Reith *et al.,* 1994), and the tag could have impaired the dimerization of the 3HA-mRFX6 protein with endogenous RFX proteins or its homodimerization. Since all RFX proteins recognize xbox motifs on the DNA, up to five RFX dimers could have bound simultaneously to the five xbox motifs in *mMlxipl* intron 1 region. In cells in which no foreign protein is overexpressed, xbox 2 could be occupied by non-RFX6 RFX dimers. The high abundance of the tagged protein due to its expression under a strong cytomegalovirus promoter could have driven the endogenous RFX proteins out of their natural binding sites leading to a shifted ChIP sequencing peak.

For each ChIP experiment, a *de novo* consensus xbox sequence was determined by comparing the 100 bp sequences under the ChIP sequencing peaks (1000 out of 5104 anti-HA peaks, 1000 out of 1552

anti-RFX6 peaks in Min6b1 cells and 863 anti-RFX6 peaks in islets). The sequence of xbox 2 is identical to these three *de novo* motifs whereas the other four xbox motif sequences are different from the *de novo* motifs (**TABLE C1**). Interestingly, xbox 2 also matches the *de novo* motif of a recently published anti-RFX6 ChIP experiment in human islet cell-like TC-YIK cells (Lizio *et al.*, 2015; **TABLE C1**). Moreover, among the five xbox motifs, only xbox 2 is conserved to 100 % in the mouse, rat and human genomes (**FIGURE C1C+D**, **TABLE C1**). Therefore, we decided to focus on the role of xbox 2 in the regulation of *mMlxipl* expression by RFX6 although the untagged RFX6 seems to preferentially bind to xbox 3.

1.3. RFX6-mediated transactivation of *mMlxipl* depends on xbox 2

To assess the enhancer activity of the RFX6-bound *mMlxipl* intron 1 region, we cloned the whole 1165 bp long region (Chr5:135,110,073-135,111,237 on mm10), an 821 bp long sequence (Chr5:135,110,215-135,111,035 on mm10) corresponding to the anti-RFX6 peak in islets, and a shorter 626 bp fragment (Chr5:135,110,215-135,110,882 on mm10) into the firefly luciferase reporter vector pGL4.23 (**FIGURE C2**). The 1165 bp fragment contains all five xbox motifs, the 821 bp fragment xbox 2, 3 and 4 and the 606 bp fragment only xbox 2 and 3 (**FIGURE C1E**, **FIGURE C2A**). The shortest fragment was cloned because we encountered difficulties to correctly amplify the (TG)n repeat by PCR. Fortunately, we finally succeeded in cloning all three fragments without mutations. The genomic fragments were inserted upstream of a minimal promoter and the firefly luciferase, so that the expression of the firefly luciferase was controlled by the genomic sequences of *mMlxipl* intron 1 (**FIGURE C2A**). Subsequently, we deleted the xbox 2 motif and one additional nucleotide up- and downstream of xbox 2 by site-directed mutagenesis (**FIGURE C2A**, **FIGURE B1D**). Hereafter, the reporter plasmids are named *mMlxipl* intron 1 1.2 kb, 0.8 kb and 0.6 kb construct, and ∆xbox 2 means that xbox 2 has been deleted.

First, we tested the activation of all constructs in the human embryonic kidney cell line HEK293T that we co-transfected with the reporter plasmids and *mRfx6* cDNA encoding a Flag-3HA-mRFX6 fusion protein (**FIGURE C3**). In cells transfected with *mRfx6* cDNA and the wild-type *mMlxipl* intron 1 1.2 kb, 0.8 kb and 0.6 kb constructs, firefly luciferase was strongly induced compared to cells transfected with a no cDNA control plasmid (**FIGURE C3**). A construct expressing the firefly luciferase under the control of the minimal promoter without genomic DNA only caused basal firefly luciferase activity in cells co-transfected with *mRfx6* cDNA, thus the activation clearly depended on the *mMlxipl* intron 1 genomic DNA. Intriguingly, the shorter the genomic DNA insert was, the higher was the activation (6-fold for the 1.2 kb, 14-fold for the 0.8 kb and 43-fold for the 0.6 kb construct; **FIGURE C3A**). This points again on the preferential binding of RFX6 to xbox 2 and xbox 3 because the 0.6 kb construct only

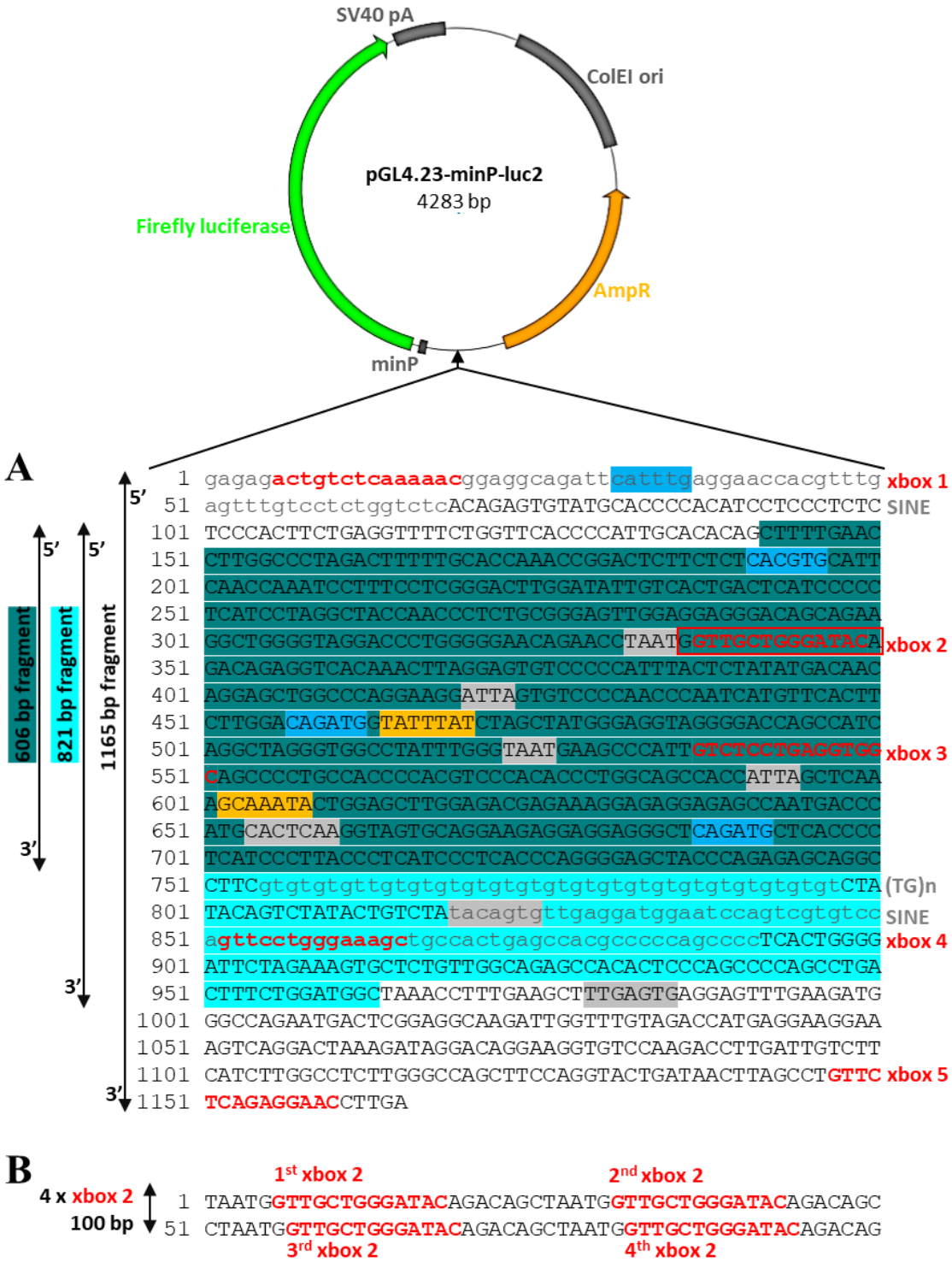


FIGURE C2: Firefly luciferase *mMlxipl* intron 1 reporter vectors. (A) Sequence of the 0.6 kb, 0.8 kb and 1.2 kb genomic *mMlxipl* intron 1 fragments (see **FIGURE C1**) integrated upstream of a minimal promoter and the firefly luciferase gene into the commercial reporter vector pGL4.23. The red box frames the deleted sequence in the mutant reporter vectors (called Δxbox 2). E-box motifs are marked in blue, FOXA1/2 binding sites in orange and putative NKX2 binding motifs in grey. **(B)** Sequence of a construct with four repetitions of the xbox 2 motif cloned into pGL4.23-minP-luc2 (named pGL4.23-quadruple xbox2-minP-luc2 and cloned by Perrine Strasser).

contains these two xbox motifs. It is possible that the additional xbox motifs in the 1.2 kb and 0.8 kb constructs capture RFX6 molecules that can no longer contribute to activation, or that the length of the fragment has an influence on the interaction of the transcription factors bound to the genomic DNA insert and the transcription initiation complex bound to the minimal promoter.

The deletion of the conserved xbox 2 motif in the three constructs significantly reduced RFX6-mediated transactivation. For the 1.2 kb, 0.8 kb and 0.6 kb construct, we observed a 3-fold, 4-fold and 9-fold reduction of the induction, respectively (**FIGURE C3A**). By using a reporter construct containing four repetitions of this xbox, we further showed that the xbox 2 motif is sufficient for RFX6-mediated *mMlxipl* transactivation as the construct was significantly activated in cells co-transfected with *mRfx6* cDNA in HEK293T cells (**FIGURE C3B**). Together, these data show that the xbox 2 is essential and sufficient for *mMlxipl* transactivation by RFX6.

Next, we wondered whether the endogenous transcription factors of the murine pancreatic beta cell line Min6b1 (Lilla *et al.*, 2003) and the rat cell line Ins-1 832/13 (Hohmeier *et al.*, 2000) can induce our *mMlxipl* intron 1a reporter constructs. We transfected these two cell lines with the 0.6 kb reporter construct and measured firefly luciferase activity 48 h later. We observed a significant 25-fold induction of firefly luciferase activity in both cell lines compared to cells transfected with the no gDNA control construct (**FIGURE C3C**). Like in HEK293T cells, the deletion of the evolutionary conserved xbox 2 motif significantly reduced the induction although the decrease was weaker (2-fold) than the one in HEK293T cells (9-fold; **FIGURE C3C**). As mentioned above, the 0.6 kb construct contains xbox 2 and 3. The Flag-3HA-mRFX6 protein that was introduced into the HEK293T cell line, might preferentially bind xbox 2 whereas the endogenous RFX proteins of Min6b1 and Ins-1 832/13 cells might equally bind xbox 2 and xbox 3. This could explain why the deletion of xbox 2 decreased the firefly luciferase activity only by half in the pancreatic beta cell lines. On the other hand, endogenous transcription factors of the Min6b1 and the Ins-1 832/13 cell line that did not bind this xbox motif, could have participated in *mMlxipl* intron 1 transactivation resulting in less severe reduction.

We further investigated the role of RFX6 in the transcriptional regulation of *Mlxipl* in Min6b1 and Ins-1 832/13 cells by reducing *Rfx6* expression via RNA interference. We found that the knockdown of *Rfx6* by RNA interference in Min6b1 cells (Perrine Strasser, unpublished data) or Ins-1 832/13 cells halved the *Mlxipl* transcript level (**FIGURE C3D+E**). These results confirm the observations made in the beta cell-specific *mRfx6* knockout mice mentioned above. Furthermore, RNA interference experiments in the human pancreatic beta cell line EndoC-betaH2 and in human islets revealed that a reduction of *RFX6* in these models also caused a decrease of *MLXIPL* expression (Vikash Chandra from the Cochin institute in Paris, unpublished data). Thus, the regulation of *Mlxipl* by RFX6 seems to be conserved in rodent and human cells as the results of the performed knockdown experiments are consistent among species.

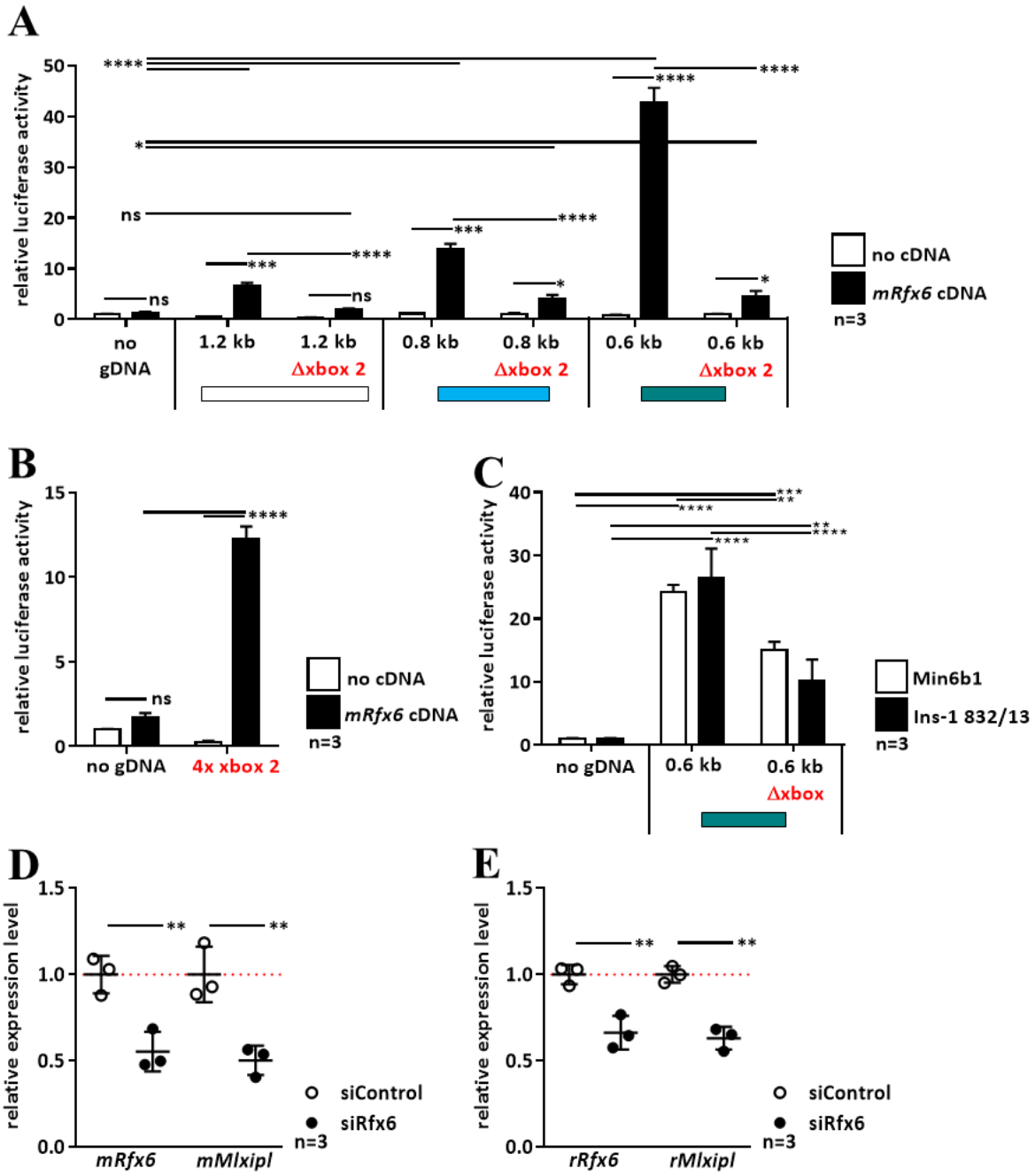


FIGURE C3: RFX6 activates *Mlxipl* expression via the conserved xbox 2 motif in *mMlxipl* intron 1. (A) Transactivation in HEK293T cells analysing the activation of the wild-type and Δxbox 2 *mMlxipl* intron 1 1.2 kb, 0.8 kb and 0.6 kb Firefly luciferase reporter constructs by mRFX6. **(B)** Transactivation assay in HEK293T cells analysing the activation of a Firefly luciferase reporter construct containing four xbox 2 motifs by mRFX6. **(C)** Transactivation assay in Min6b1 cells and Ins-1 832/13 cells using the wild-type or Δxbox 2 *mMlxipl* intron 1 0.6 kb Firefly luciferase reporter construct. **(D)** RT-qPCR analysing the expression of *mRfx6* and *mMlxipl* in siRfx6-treated Min6b1 cells (75 nM siRNA, standard cell culture medium at 25 mM glucose, analysis was done 48 h after siRNA treatment, unpublished data of Perrine Strasser). **(E)** RT-qPCR analysing the expression of *rRfx6* and *rMlxipl* in siRfx6-treated Ins-1 832/13 cells (11 nM siRNA, standard cell culture medium at 11 mM glucose, analysis was done 48 h after siRNA treatment). Data in panels **(A)** to **(E)** are shown as mean plus standard deviation (n=3). Statistical significances (p<0.01 **, p<0.001 ***, p<0.0001 ****) were determined by two-way ANOVA **(A, B, C)** or unpaired two-tailed t test **(D, E)**.

1.4. Human *RFX6* point mutations affect RFX6-mediated *mMlxipl* transactivation

We used our *mMlxipl* intron 1 transactivation assay as a model to test the effect of the three published point mutations p.R181Q, p.S217P and p.V506G (Smith *et al.,* 2010; Chandra *et al.,* 2014) and two unpublished *RFX6* point mutations p.F294V and p.R578P (communicated by Martine Vaxillaire from the Pasteur institute in Lille) identified in patients with Mitchell-Riley or a related syndrome, on RFX6 transcriptional function (**FIGURE C4**).

The phenotypes of the two patients carrying the homozygous point mutations p.R181Q (c.542C>A in exon 4) in the DNA-binding domain and p.S217P (c.649T>C in exon 6) located between the DNA-binding domain and the B domain (**FIGURE C4A**) were described in 2009 (Martinovici *et al.,* 2009) and 2008 (Chappell *et al.,* 2008), and the *RFX6* mutations were identified in 2010 (Smith *et al.,* 2010). Both patients showed typical symptoms of Mitchell-Riley syndrome (**TABLE A1**). The disease course of the child with the homozygous p.R181Q mutation was more severe; the child died in the third month of his life (Martinovici *et al.,* 2009). The child with the homozygous p.S217P mutation had a milder phenotype (Chappell *et al.,* 2008). In 2011, she was 6 years old and developed normally. She had no noticeable diarrhea since she was 18 months old and the diabetes is well controlled by an insulin pump (Spiegel *et al.,* 2011).

Smith *et al.* tested the capacity of both mutant proteins to bind the hepatitis B virus xbox motif in Electrophoretic Mobility Shift Assays (EMSA). The p.R181Q point mutation completely abolished DNA binding whereas the p.S217P point mutation mildly reduced DNA binding (Smith *et al.,* 2010). Pearl *et al.* tried to rescue a morpholino-induced knockdown of *rfx6* in the clawed frog *Xenopus laevis* by injecting these mutant proteins (Pearl *et al.,* 2014). The treatment of Xenopus with *rfx6* morpholinos led to the reduction of endocrine (insulin, glucagon, somatostatin) and exocrine gene expression and an abnormal gut formation. The injection of hormone-inducible wild-type *rfx6* partially rescued the phenotype when it was activated at the exact moment when pancreatic development was initiated. However, the injection of *rfx6* carrying the p.R181Q or the p.S217P mutation did not lead to any improvement in the phenotype (Pearl *et al.,* 2014).

The p.V506G point mutation (c.1517T>G in exon 14) in the dimerization domain (**FIGURE C4A**) was described by Chandra *et al.* in 2014. The girl had all symptoms of Mitchell-Riley syndrome (**TABLE A1**). During the first year of her life, she received parenteral nutrition. In 2016, she was 8 years old and on normal diet (Chandra *et al.,* 2014; Poidvin *et al.,* 2016). To assess the effect of the point mutation on the protein function, Chandra *et al.* transfected the human pancreatic beta cell line EndoC-betaH2 with a human insulin promoter luciferase reporter construct containing three xbox motifs and wild-type or mutant *RFX6* cDNAs. The wild-type protein activated the reporter while the mutant RFX6 failed to induce luciferase expression. Similarly, the expression of wild-type but not mutant RFX6 rescued the

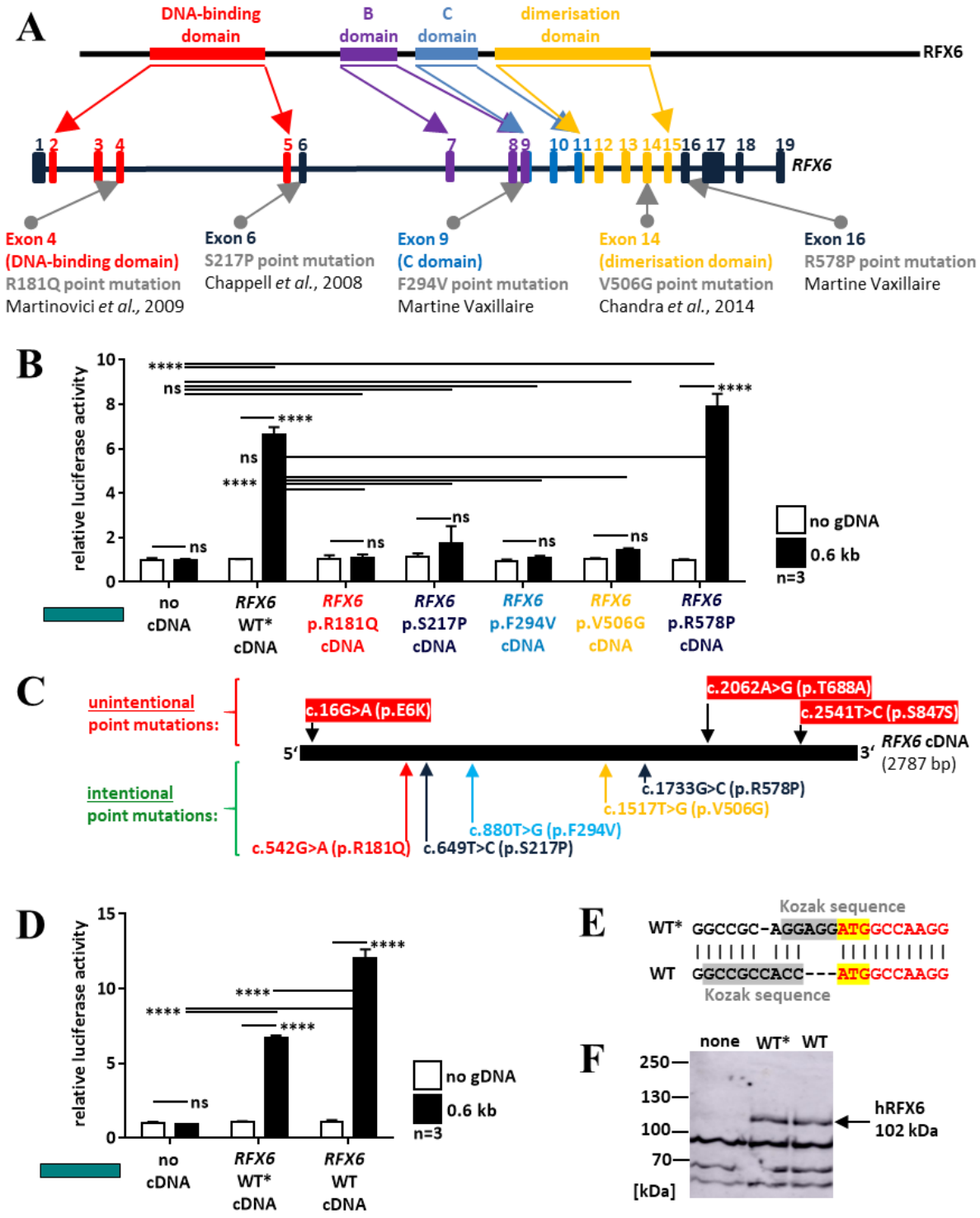


FIGURE C4: Human *RFX6* point mutations impair RFX6-mediated *mMlxipl* transactivation. (A) Position of the tested *RFX6* point mutations in the *RFX6* gene and RFX6 protein. **(B)** Transactivation assay in HEK293T analyzing the activation of the *mMlxipl* intron 1 0.6 kb Firefly luciferase reporter construct by wild-type and mutated RFX6 proteins. **(C)** Position of unintended point mutations present in the *RFX6* coding sequence of all used *RFX6* expression vectors. **(D)** Transactivation assay in HEK293T comparing the activation of the *mMlxipl* intron 1 0.6 kb Firefly luciferase reporter construct by RFX6 proteins with (WT*) and without (WT) the unintended point mutations. **(E)** Kozak sequence of the two expression vectors encoding *RFX6* WT* and WT. **(E)** Anti-RFX6 western blot detecting RFX6 WT and WT* expressed in HEK293T cells (rabbit anti-RFX6, Sigma). Data in panels **(B)** and **(D)** are shown as mean plus standard deviation (n=3). Statistical significances (p≥0.05 not significant (ns), p<0.0001 ****) were determined by two-way ANOVA.

loss of insulin promoter transactivation in EndoC-betaH2 cells treated with siRfx6. Next, they transfected with wild-type or mutant RXF6-IRES-EGFP, sorted the GFP-positive EndoC-betaH2 cells by FACS and measured the expression of the RFX6 target genes *INS* and *CACNA1A*, *CACNB2* and *CACNA1D* encoding calcium channels. Wild-type RFX6 but not the one with the point mutation increased the expression of the calcium channel genes, so the authors concluded that the mutant protein can not induce the expression of its target genes. For both proteins, there was no difference in the expression level of insulin. They suggested that the endogenous RFX6 had activated maximally insulin transcription (Chandra *et al.,* 2014).

In addition to the three published mutations, we tested two new mutations that our collaboration partner Martine Vaxillaire from the Pasteur institute in Lille had found in patients. Her group sequenced the *RFX6* locus in 22 maturity-onset diabetes of the young (MODY) patients and 86 patients with a diabetes onset during adulthood, and they identified the point mutations p.F294V (c.880T>G in exon 9) in the C domain and p.R578P (c.1733G>C in exon 16) C-terminal of the dimerization domain (**FIGURE C4A**) in two MODY patients.

In our *mMlxipl* intron 1 transactivation assay in HEK293T cells, we observed that the mutant proteins with the point mutations p.R181Q, p.S217P, p.F294V and p.V506G did not induce any firefly activity whereas the reporter induction by the RFX6 protein with the point mutation p.R578P was indistinguishable from that obtained with the wild-type protein (**FIGURE C4B**). We used the 0.6 kb reporter construct because it showed the strongest activation in combination with the *mRfx6* cDNA (**FIGURE C3A**). Our findings for the point mutations p.R181Q, p.S217P and p.V506G were consistent with the described inactivity of these mutant *RFX6* variants (Smith *et al.,* 2010; Pearl *et al.,* 2014; Chandra *et al.,* 2014).

Among the novel mutations, only p.F294V severely affected RFX6 function. This mutation was found in the conserved C domain. The mutation p.R578P located in the C-terminus of RFX6 had no impact on the transcriptional activity of RFX6 in our assay. As mentioned in the introduction, it is quite difficult to establish a link between the location of the point mutation and the severity of the disease but, in general, mutations in the C-terminal region of the protein do not severely affect RFX6 function (Sansbury *et al.,* 2015), and patients carrying those mutations have a delayed onset of diabetes.

In summary, our *Mlxipl* intron 1 transactivation assay is another useful functional test to evaluate the severity of *RFX6* mutations identified in Mitchell-Riley patients on protein function, and we add two new point mutations to the growing list of *RFX6* mutations.

Unfortunately, after completing the experiment, we found that our *RFX6* expression vector contains two missense and one silent point mutations in the *RFX6* coding sequence (p.E6K, c.16G>A; p.T688A, c.2062A>G; p.S847S, c.2541T>C; **FIGURE C4C**). The human *RFX6* CODING SEQUENCE encoded by our expression vector was amplified from the cDNA clone proposed by the Mammalian

Gene Collection (MGC cDNA clone # 33442, IMAGE # 5272028). In fact, the presence of the point mutations p.E6K and p.T688A was already noted on their website. To evaluate the importance of these point mutations on RFX6 function, we decided to clone a new expression vector. We amplified the human *RFX6* coding sequence from a cDNA that was prepared from a total RNA extract of human EndoC-betaH2 cells (a gift from Raphaël Scharfmann; **FIGURE B5**).

We compared the induction of the *mMlxipl* intron 1 0.6 kb reporter construct in HEK293T cells transfected with our old and the new RFX6 expression vector, and we found that the new one better induced firefly luciferase activity (**FIGURE C4D**). Possible causes include an influence of the two missense mutations on protein function or differences in expression because we also modified the Kozak sequence in the new expression vector (**FIGURE C4E**). However, we did not observe any difference in protein expression in the performed western blot, but we did not quantify the protein amounts (**FIGURE C4F**). Consequently, we could not draw any conclusions about the influence of these unintentional point mutations on protein function except that the protein remained functional (**FIGURE C4D**). Since all wild-type and mutant *RFX6* cDNAs that we used in our previous experiments, contained the unintended point mutations, our observations should be correct. Furthermore, they were consistent with the literature. However, we can not exclude that the unintended point mutations disturb intra-conformational changes that might be caused by the point mutations p.R181Q, p.S217P, p.F294V, p.V506G and p.R578P and that might alter protein function.

1.5. Involvement of cofactors in RFX6-mediated *mMlxipl* transactivation

The transcriptional regulation of a gene is usually not controlled by a single transcription factor, but by several – either multiple copies of the same transcription factor or different transcription factors – that bind to proximal regulatory elements within the promoter and to distal regulatory elements outside of the promoter region, and that can either work as activator or repressor binding to enhancer and silencer leading to the activation and repression of the transcription, respectively. Some transcription factors may also recruit chromatin remodelers to the DNA leading to the opening of the chromatin (Marton *et al.*, 2006, review). Most transcription factors are unable to access their specific binding sites when the chromatin is compact. Pioneer transcription factors such as GATA and FOXA are proteins that can still reach their binding sites in certain regions of compacted chromatin. They are the first detectable transcription factors at this locus, and their binding precedes the binding of other transcription factors (Zaret and Carroll, 2011, review).

Depending on the nature of the bound transcription factors, they synergistically or antagonistically regulate the activity of the pre-initiation complex at the core promoter and thus the transcription rate. The spatiotemporal expression pattern of a gene is determined by the transcription factors binding to

this gene at a given moment in a given cell type. In the following, we tested different candidates that might cooperatively regulate *Mlxipl* with RFX6.

1.5.1. E2F motifs colocalize with xbox motifs in the RFX6-bound genomic DNA regions

To identify potential cofactors of RFX6, a bioinformatic analysis has been executed to identify enriched sequence motifs in a 500 bp long sequence under the summit of the 10,143 peak sequences found in the islets anti-RFX6 ChIP experiment (data of Perrine Strasser; occurrence analysis by Stéphanie Legras; **FIGURE C5A**). As expected, RFX binding sites were strongly enriched in the peak sequences. The analysis further revealed an enrichment of binding sites of the DNA-binding proteins NRF1, CTCF, SP2, KLF5 and E2F in RFX-bound regions. A similar analysis with the 3,300 peak sequences of the anti-HA ChIP experiments on 3HA-mRfx6 expressing Min6b1 cells also showed the strong enrichment of RFX binding sites as well as binding sites of KLF5, SP2, E2F1 and NRF1 (data of Perrine Strasser; analysis by Stéphanie Legras; **data not shown**). In a subsequent analysis, the most abundant combinations of binding sites were identified (data of Perrine Strasser; analysis by Stéphanie Legras; **FIGURE C5B**). RFX motifs often co-occur with NRF1, CTCF, SP2, KLF5 and E2F sites in islets (**FIGURE C5B**) and with SP2, NRF1, KLF5 and E2F1 sites in Min6b1 cells (**data not shown**). Mutations in the zinc finger protein KLF11 cause MODY (Neve *et al.,* 2005; Fernandez-Zapico *et al.,* 2009) while mutations in the other *KLF* genes, including *KLF5*, could no been associated with an increased risk of diabetes type 2 (Gutiérrez-Aguilar *et al.,* 2007).

However, the occurrence and the co-occurrence report the distribution of binding sites in the genome relative to the RFX6 binding sites, but both analyses do not provide information on the use of these binding sites in a given cell type at a given stage of differentiation under the given physiological conditions. The increased appearance of a binding motif in the RFX6 peak sequences does not necessarily mean that the protein also binds to this site or that it is even expressed in the tested cell type, and the frequent co-occurrence of a binding site with an RFX binding motif does not inevitably indicate that both proteins bind simultaneously to their binding sites.

Consistently with our co-occurrence analysis, Varshney *et al.* found that binding sites of the CCCTC-binding factor CTCF often co-localize with xbox motifs in human islets (Varshney *et al.,* 2017). In their study, they investigated the relevance of Single Nucleotide Polymorphisms (SNPs) that have been identified by Genome-Wide Association Studies (GWAS) in diabetes type 2 patients, for the transcriptional regulation of islet cell genes in human pancreatic islets. They generated data about i) the genome sequence to identify SNPs in promoter and enhancer regions; ii) gene expression level (RNA sequencing); iii) histone marks (ChIP sequencing); and iv) chromatin accessibility (Assay of Transposase-Accessible Chromatin (ATAC)). They combined these data in a bioinformatical approach

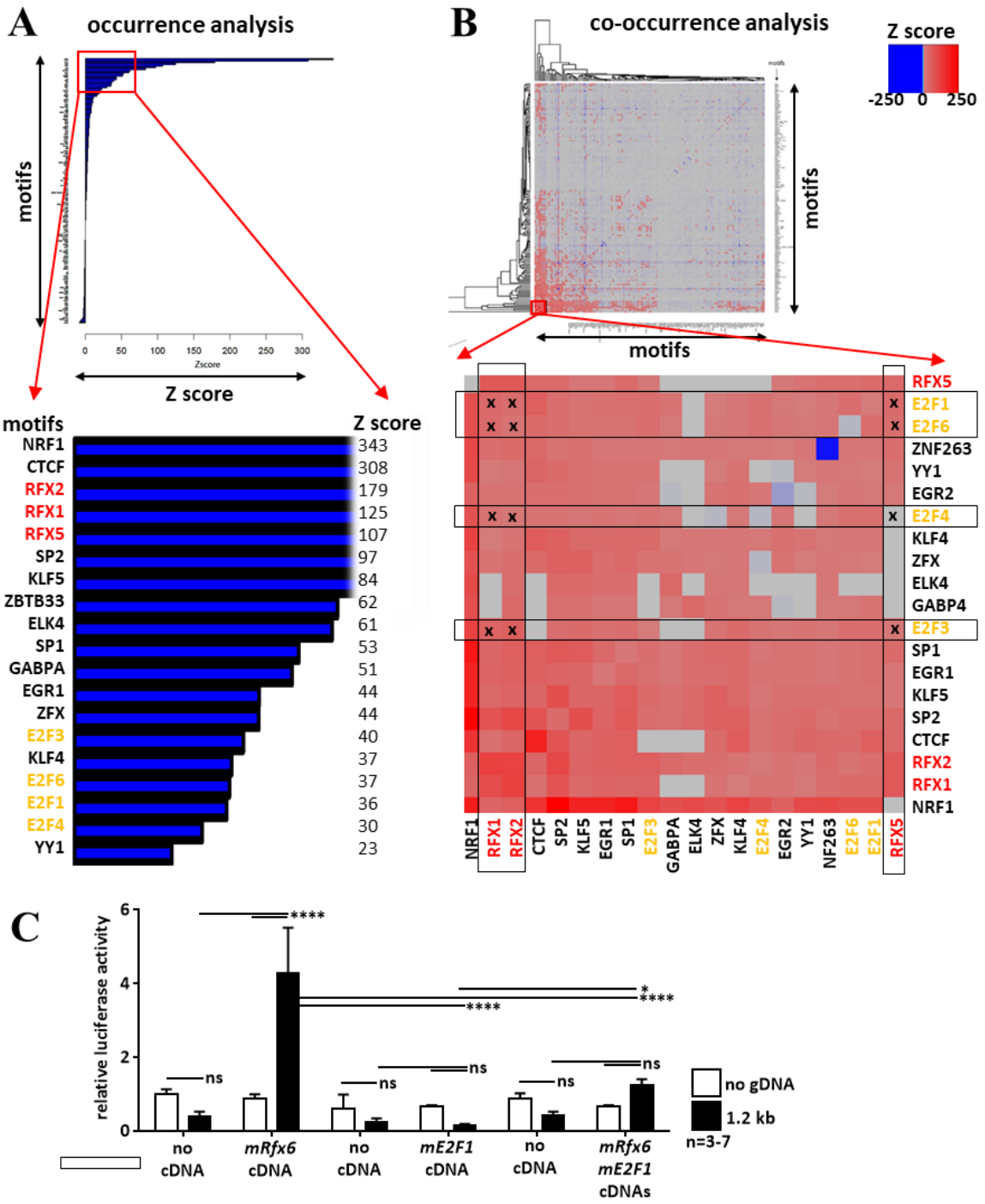


FIGURE C5: E2F sites often colocalize with xbox motifs in the murine genome. (A) Occurrence analysis determining the incidence of DNA motifs in the 10,143 peak sequences (500 bp window) of the islets anti-RFX6 ChIP experiment (done by Stéphanie Legras). **(B)** Co-occurrence analysis determining the frequency that two DNA motifs occur in the same peak sequences (500 bp window) of the islets anti-RFX6 ChIP experiment (done by Stéphanie Legras). The dimensionless Z value in **(A)** and **(B)** indicates whether a motif is more abundant or less abundant than it would be expected from a perfectly random distribution. The higher the value, the more abundant is the motif. **(C)** Transactivation assay in HEK293T cells analysing the activation of the *mMlxipl* intron 1 1.2 kb Firefly luciferase reporter construct by mRfx6, mE2F1 or both proteins. Data are shown as mean plus standard deviation (n=3-7). Statistical significances ($p \geq 0.05$ not significant (ns), $p < 0.05$ *, $p < 0.0001$ ****) were determined by two-way ANOVA.

with information about the transcription factor foot printing using the algorithm CENTIPEDE that identifies binding sites and predicts if these are bound by a transcription factor or not. Using this complicate bioinformatical approach, they found that the xbox motifs are often interrupted by the SNPs in diabetes type 2 risk alleles and that this likely prevents RFX binding and they showed a co-occurrence of xbox motifs with CTCF binding motifs, for instance in the RFX6 gene, thus suggesting an autoregulatory mechanism (Varshney *et al.,* 2017). In conclusion, two independent studies in mouse and in human pancreatic islets revealed a co-occurrence of RFX and CTCF binding motifs. This strongly supports the idea of a contribution of CTCF in the regulation of RFX6 target genes in beta cells.

NRF1 and NRF2 are both implicated in the transcriptional regulation of the mitochondrial transcription factor 1 (mtTFA) and are thus important for mitochondrial biogenesis and metabolism that plays an important role in insulin secretion pathway (Virbasius and Scarpulla, 1994; Viña *et al.,* 2009). The mtTFA gene also contains binding sites for SP1. SP1 belongs to the same family as SP2 and SP3. They are all zinc finger transcription factors. SP2 binds to GT-rich motifs whereas SP1 and SP3 binds to GT- as well as GC-rich motifs (Kingsley and Winoto, 1992). GC-rich binding motifs of SP1 have been found in the promoter of the *mMlxipl* gene (Satoh *et al.,* 2007). The deletion of the two GC boxes in the *mMlxipl* promoter significantly reduced the transactivation of a *mMlxipl* reporter construct by the endogenous proteins of the human liver cancer cell line HepG2, and both SP1 and SP3 could transactivate the wild-type *mMlxipl* reporter construct when co-expressed with the transcription factor NF-Y in the *Drosophila melanogaster* cell line SL2; by contrast, the expression of SP1 or SP3 alone did not induce the reporter. Finally, by ChIP experiments, the group demonstrated that SP1 binds to the *rMlxipl* gene in rat hepatocytes *in vivo* (Satoh *et al.,* 2007). SP1 has been shown to be a functional interaction partner of the transcription factor E2F1 co-regulating target genes that contain either SP1 bindings sites or E2F bindings sites (Lin *et al.,* 1996; Karlseder *et al.,* 1996). Three years later, Rotheneder *et al.* proofed by immunoprecipitation experiments with an anti-E2F1 antibody in HEK293T cells transfected with an E2F1 cDNA and either a SP1, SP2, SP3 or SP4 cDNA that all four SP proteins can interact with E2F1 (Rotheneder *et al.,* 1999). Taken together, the published data and the results of the occurrence and co-occurrence analysis suggest that SP1, SP2, SP3 and E2F1 might all be involved in *Mlxipl* transcriptional regulation.

E2F1 is a major activator of cell cycle progression and proliferation (Dyson, 1998, review). E2F1 knockout mice have a decreased pancreas size due to impaired pancreatic growth after birth, and a reduced beta cell number (Fajas *et al.,* 2004). The beta cells do not produce enough insulin due to a downregulation of the E2F1 target gene *Pdx1* (Iwata *et al.,* 2013), a master regulator of insulin gene transcription. This leads to insufficient insulin secretion in a glucose challenge (Fajas *et al.,* 2004). E2F transcription factors can form heterodimers with DP transcription factors and bind to the consensus sequence `5'TTTCGCGG` (Tao *et al.,* 1997) either as free heterodimers acting as transcriptional

activators or in complexes with retinoblastoma proteins and cyclin / cyclin-dependent kinases (CDK) repressing transcription (Dyson, 1998, review; Harbour and Dean, 2000, review). We decided to focus on E2F1 because it is a transcription factor with known functions in the pancreas and in insulin-secreting beta cells, but all the other proteins found in the co-occurrence analysis are also interesting targets and require further attention.

Using our *mMlxipl* intron 1 transactivation assay, we tested whether there might be a synergistic or antagonistic effect of mE2F1 and mRFX6 in the transcriptional regulation of *mMlxipl*. We observed that E2F1 alone could not activate the *mMlxipl* 1.2 kb reporter construct in HEK293T cells (**FIGURE C5C**). The lack of activation can not be due to the fact that we did not overexpressed its dimerization partner DP-1 or DP-2 as DP-1 is expressed in HEK293T cells according to the Human Protein Atlas (https://www.proteinatlas.org). In the *mMlxipl* intron 1 1165 bp region, we did not find an E2F binding motif using the published E2F1 consensus sequence (Tao *et al.,* 1997). However, this does not necessarily mean that E2F1 might not bind to a motif slightly different from this sequence. Even if there is no E2F binding motif in our construct, this does not exclude an involvement of E2F1 in the following co-transfection experiment with mRFX6. Rabinovich *et al.* found that E2F transcription factors can interfere in gene regulation without binding to a consensus motif by a simple protein-protein interaction with DNA-bound transcription factors (Rabinovich *et al.,* 2008), in our case mRFX6. They further reported that used E2F binding sites are normally localized in the core promoter region around the transcription start site of an actively transcribed gene. We use an enhancer region in the first intron of *mMlxipl* in our transactivation assays; the promoter region of *mMlxipl* is located upstream of exon 1. Thus, it is not surprising that we did not find a E2F binding motif in our construct. Moreover, Lin *et al.* reported that E2F1 can interact with DNA-bound SP1, a protein found to bind to the promoter of *mMlxipl* (Lin *et al.,* 1996). Taken together, the possible absence of an E2F1 binding motif is not a reason to exclude any participation of E2F1 in *mMlxipl* regulation by intron 1.

Indeed, the co-transfection of HEK293T cells with *mE2F1* and *mRfx6* cDNAs abrogated the activation we normally observe with the *mRfx6* cDNA (**FIGURE C5C**), suggesting that mE2F1 inhibited RFX6-mediated transactivation of *mMlxipl* and acted as a repressor. The repression mechanism could be a direct protein interaction between E2F1 and RFX6, or the above-mentioned E2F1 / DP-1 plus retinoblastoma and cyclin / CDK protein complex supposed to repress transcription. However, we can not exclude that the lack of activation was mainly due to a stronger expression of E2F1 compared to the one of RFX6, and that the measured luciferase activity largely reflected the non-activation of the reporter by E2F1 while its activation by RFX6 was underrepresented. Furthermore, we can not draw any conclusions about the role of E2F1 on *mMlxipl* regulation in pancreatic islets or beta cells because we used the HEK293T-intrinsic transcription factors and transcription machinery. We should perform a *mMlxipl* intron 1 transactivation assay in pancreatic beta cells in which we downregulate E2F1 and

compare the activation in siControl- and siE2F1-treated cells to see if we can reproduce the same inhibitory effect in a less artificial context.

1.5.2. The RFX6 peak sequence in *mMlxipl* intron 1 contains FOXA binding sites

Another putative RFX6 cofactor is the pioneer transcription factor FOXA2 (Gao *et al.,* 2010; **FIGURE C6**). The first intron of *mMlxipl* contains three FOXA binding sites. Two of them were included in our *mMlxipl* intron 1 1.2 kb reporter construct (**FIGURE C2A**). The third is localized 1,897 bp upstream of the 5'end of our 1.2 kb construct. FOXA2 binding could be detected during pancreas development as well as in mature islet cells (**FIGURE C6A**), and *mMlxipl* transcript and protein levels were significantly reduced in the embryonic pancreas and in the islets of adult mice with an inducible beta cell-specific double knockout of *mFoxa1* and *mFoxa2* (Gao *et al.,* 2010). The beta cell-specific knockout of *mFoxa1* or *mFoxa2* alone did not affect *mMlxipl* expression, suggesting redundant functions of these transcription factors (Gao *et al.,* 2010). Duncan *et al.* proposed that FOXA1 and FOXA2 bind to DNA as monomers and that the change in transcription might result from their competition for the available binding sites (Duncan *et al.,* 1998). Together, all these observations suggest that FOXA1, FOXA2 and RFX6 co-regulate the expression of *mMlxipl*.

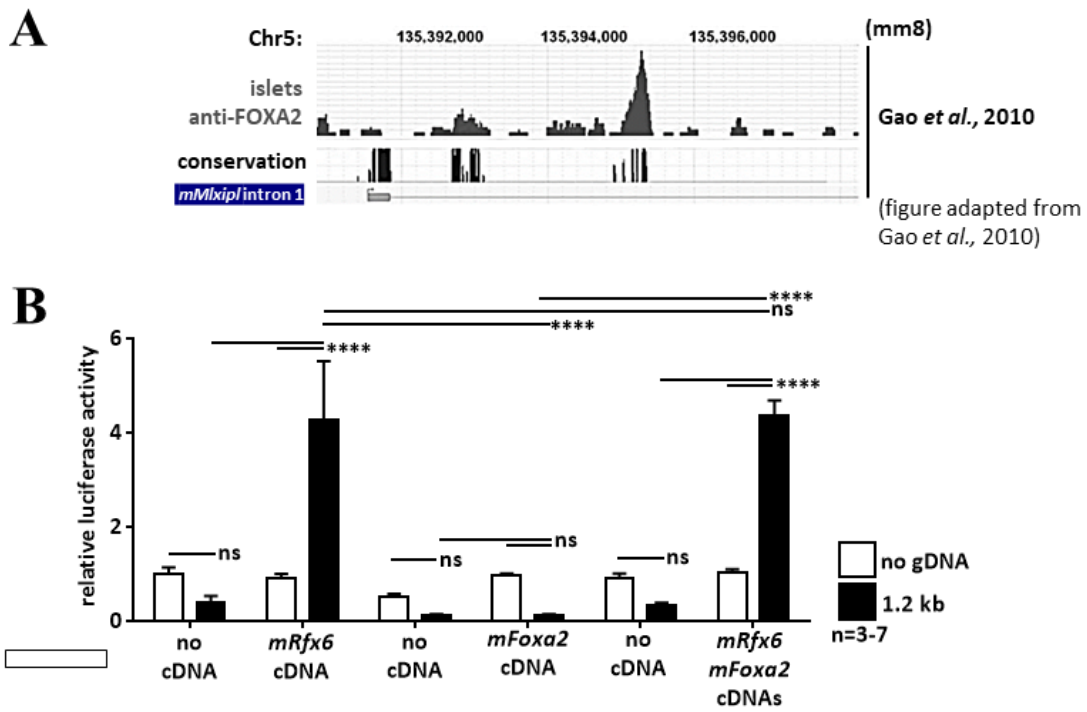


FIGURE C6: FOXA2 binds to the first intron of *mMlxipl*. (A) ChIP sequencing data showing the binding of mFOXA2 on *mMlxipl* intron 1 in mouse islets (Gao *et al.,* 2010). **(B)** Transactivation assay in HEK293T cells analysing the activation of the *mMlxipl* intron 1 1.2 kb Firefly luciferase reporter construct by mRFX6, mFOXA2 or both proteins. Data are shown as mean plus standard deviation (n=3-9). Statistical significances (p≥0.05 not significant (ns), p<0.0001 ****) were determined by two-way ANOVA.

Surprisingly, we did not observe any *mMlxipl* transactivation upon addition of *mFoxa2* cDNA in our *mMlxipl* intron 1 transactivation assay in HEK293T cells using the 1.2 kb construct, and the activation of the construct in cells co-transfected with *mFoxa2* and *mRfx6* cDNAs was not higher than the one obtained with the *mRfx6* cDNA alone (**FIGURE C6B**). As mentioned above, FOXA transcription factors are pioneer factors that can bind to dense chromatin regions before these become accessible to other transcription factors. This might not be reproducible in a plasmid-based assay. Although plasmid DNA can be organized in nucleosome-like particles (Mladenova *et al.,* 2009), it is unlikely that their structure matches exactly that of chromatin. By contrast, published results of a transactivation assay with a glucagon reporter construct demonstrate that FOXA2 can induce the firefly luciferase in this kind of assay (Kaestner *et al.,* 1999), so the plasmid-based strategy should not cause any problems.

In conclusion, we showed that the expression of *mFoxa2* alone did not induced the *mMlxipl* intron 1 1.2 kb reporter construct (**FIGURE C6B**). The fact that we did not observe any amplification of the luciferase activity in cells co-transfected with *mFoxa2* and *mRfx6* cDNAs does not necessarily exclude that both proteins bind to the construct. If both proteins bind to the *mMlxipl* genomic DNA in the plasmid, the assay shows that the induced transcription of the firefly luciferase gene exclusively relays on RFX6. In the natural genomic context, binding of FOXA1 and FOXA2 might be necessary to enable RFX6 binding on the first intron of *mMlxipl*, and, in the absence of FOXA1 and FOXA2, the chromatin accessibility of this *mMlxipl* region might be reduced. This would prevent RFX6 binding and transcriptional activation of *mMlxipl* expression by RFX6 leading to the observed loss of *mMlxipl* expression in the *mFoxa1/2* knockout, but with RFX6 playing a role in this process.

1.5.3. The RFX6 peak in *mMlxipl* intron 1 overlaps with NKX2.2 and NEUROD1 binding

Churchill *et al.* showed that the RFX6 peak sequence also overlaps with ChIP sequencing peaks of the homeobox transcription factor NKX2.2 and the basis helix-loop-helix (bHLH) transcription factor NEUROD1 (Churchill *et al.,* 2017; **FIGURE C7**). The anti-NKX2.2 ChIP analysed the NKX2.2 occupancy in the Min6 cell line (Gutiérrez *et al.,* 2017), and the anti-NEUROD1 ChIP experiment the binding of NEUROD1 in islets of adult mice (Tennant *et al.,* 2013). Churchill *et al.* aligned these two ChIP data sets with our Min6b1 anti-HA ChIP data set (Piccand *et al.,* 2014). They found that all three proteins share a prevalent peak in the first intron of *mMlxipl* (Churchill *et al.,* 2017; **FIGURE C1**; **FIGURE C7A**). They further reported that 72 genes were common target genes of all three transcription factors (Churchill *et al.,* 2017).

The knockout of *mNkx2.2* in the endocrine progenitors of mice resulted in a decreased expression of *mMlxipl* in the embryo (Churchill *et al.,* 2017), and the beta cell-specific knockout of *mNkx2.2* in adult mice reduced *mMlxipl* transcript level (Gutiérrez *et al.,* 2017). In the human islet cell-like TC-YIK

cells, a cell line that expresses three-quarter of the human islet-specific genes, the knockdown of *NEUROD1* diminished the expression of *MLXIPL* (Lizio *et al.,* 2015). Like in the mouse, human NEUROD1 also binds to the *MLXIPL* gene in this human cell line. They further reported that NEUROD1 and RFX6 simultaneously bind to 18 different genes in the human genome (Lizio *et al.,* 2015).

Moreover, the global knockout of *mNkx2.2* (Sussel *et al.,* 1998), *mNeurod1* (Naya *et al.,* 1997) and *mRfx6* (Smith *et al.,* 2010; Piccand *et al.,* 2014) in mice caused severe diabetes and the pups died shortly after birth. The phenotypes of beta cell-specific *mNkx2.2* (Gutiérrez *et al.,* 2017), *mNeurod1* (Gu *et al.,* 2010) and *mRfx6* knockout mice (Piccand *et al.,* 2014) are very similar: the beta cells in all three conditional knockout mice resemble more to immature than to adult beta cells. Thus, all three genes are important for the maintenance of the adult beta cell stage. In the absence of mNKX2.2, mNEUROD1 and mRFX6, the expression of disallowed genes was re-induced, and the expression of beta cell genes reduced (Gutiérrez *et al.,* 2017; Gu *et al.,* 2010; Piccand *et al.,* 2014). The diminished expression of *mIns1* in *mNeurod1* and *mRfx6* Δbeta mice (Gu *et al.,* 2010; Piccand *et al.,* 2014) and *mIns2* in *mNkx2.2* Δbeta mice (Gutiérrez *et al.,* 2017) caused a reduction of the insulin content in the islets of the knockout mice (Gutiérrez *et al.,* 2017; Gu *et al.,* 2010; Piccand *et al.,* 2014). The total number of islet cells remained unchanged in all three mouse models (Gutiérrez *et al.,* 2017; Gu *et al.,* 2010; Piccand *et al.,* 2014), but the loss of *mNkx2.2* and *mNeurod1* led to the appearance of bi-hormonal cells producing insulin and somatostatin (Gutiérrez *et al.,* 2017; Gu *et al.,* 2010). Physiologically, the three mouse models exhibited reduced insulin secretion and impaired glucose clearing (Gutiérrez *et al.,* 2017; Gu *et al.,* 2010; Piccand *et al.,* 2014).

In summary, the comparable role of the NEUROD1, NKX2.2 and RFX6 as well as the experiments showing that all three transcription factors bind to *Mlxipl* and that *Mlxipl* transcription is reduced in their absence, suggest that NEUROD1 and NKX2.2 likely co-regulate *Mlxipl* together with RFX6. Therefore, we expressed *mNkx2.2* and *mNeurod1* cDNAs either alone or in combination with *mRfx6* in HEK293T cells, and we tested the effect on the activation of our *mMlxipl* intron 1 1.2 kb, 0.8 kb and 0.6 kb reporter constructs (**FIGURE C7B+C+D**).

Like mRFX6, mNKX2.2 was able to significantly activate the *mMlxipl* intron 1 reporter constructs. This activation was not affected of the xbox 2 deletion (**FIGURE C7B+C+D**). We observed the strongest activation with the 0.8 kb construct (12-fold, **FIGURE C7C**) and a 5-fold activation with the 1.2 kb and 0.6 kb construct (**FIGURE C7B+D**). The induction of the 1.2 kb and the 0.8 kb construct by mNKX2.2 was equal to the one obtained with mRFX6 (**FIGURE C7B+C**) whereas the activation of the 0.6 kb was much stronger in cells transfected with the *mRfx6* cDNA (30-fold, **FIGURE C7D)**. The heights of the NKX2.2-mediated activation of the different constructs could be due to the distribution of the NKX binding sites. Other members of the NKX2 family, NKX2.1 and NKX2.5, can bind to the sequences `5'TNNAGTG`, `5'TAAT` and `5'CAAGTG` (Chen and Schwartz, 1995; Damante *et al.,* 1994;

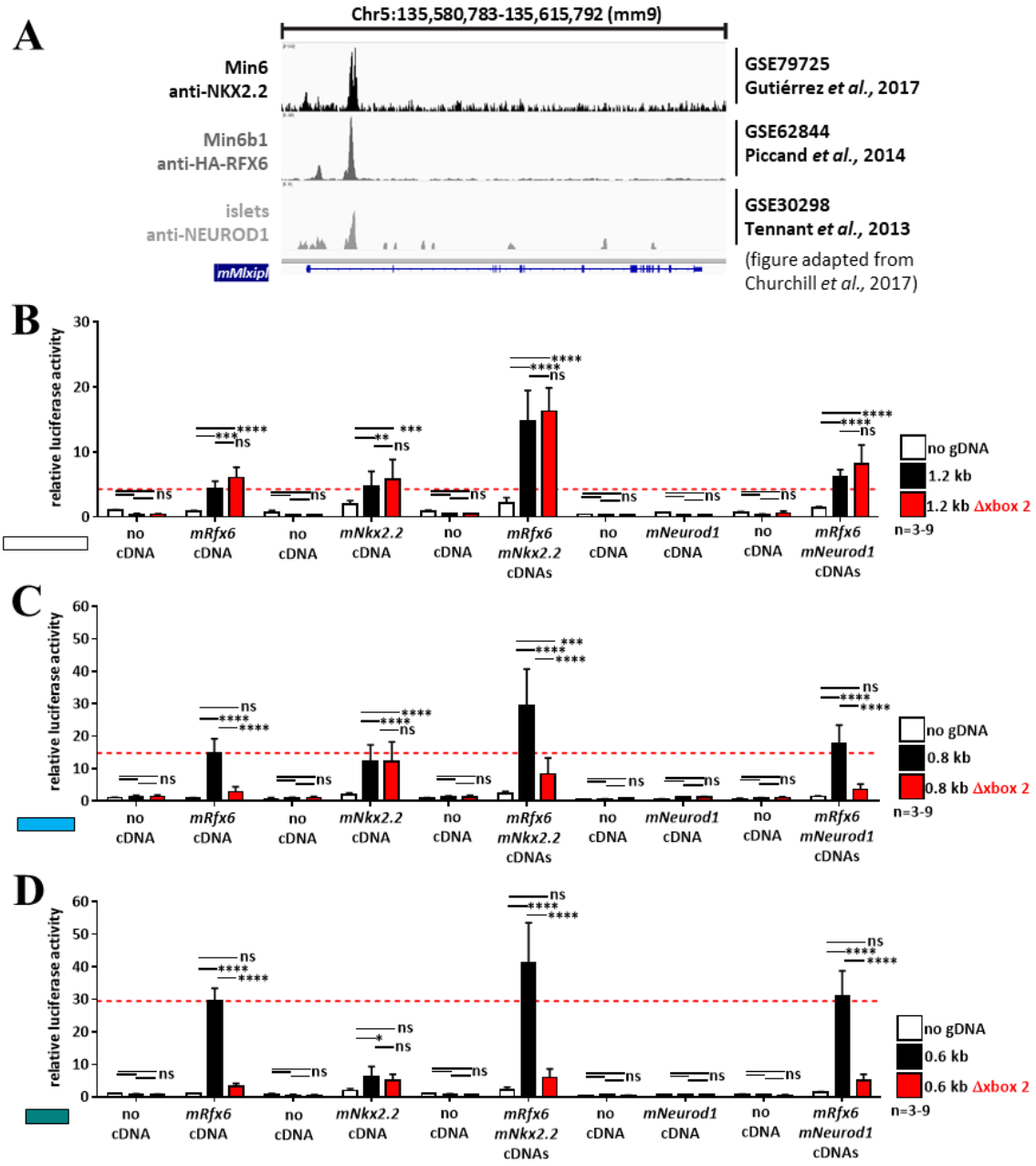


FIGURE C7: NKX2.2, NEUROD1 and RFX6 synergistically transactivate *mMlxipl*. (A) ChIP sequencing data showing the binding of mNKX2.2, mRFX6 and mNEUROD1 on *mMlxipl* intron 1 in Min6 cells (Churchill *et al.*, 2017), Min6b1 cells (Piccand *et al.*, 2014) and mouse islets (Churchill *et al.*, 2017), respectively. **(B, C, D)** Transactivation assay in HEK293T cells analysing the activation of the wild-type and Δxbox2 *mMlxipl* intron 1 1.2 kb **(B)**, 0.8 kb **(C)** and 0.6 kb **(D)** Firefly luciferase reporter constructs by mRFX6, mNEUROD1 and mNKX2.2 alone or by mRFX6 together with mNEUROD1 and mNKX2.2. Data are shown as mean plus standard deviation (n=3-9). Statistical significances (p≥0.05 not significant (ns), p<0.0.5 *, p<0.01 **, p<0.001 ***, p<0.0001 ****, additional statistics in the table below) were determined by two-way ANOVA.

	1.2 kb	1.2 kb Δxbox 2	0.8 kb	0.8 kb Δxbox 2	0.6 kb	0.6 kb Δxbox 2
no cDNA vs *mRfx6*	***	****	****	ns	****	ns
no cDNA vs *mNkx2.2*	****	****	****	****	*	ns
no cDNA vs mRfx6 + mNkx2.2	****	****	****	ns	****	ns
no cDNA vs *mNeurod1*	ns	ns	ns	ns	ns	ns
no cDNA vs mRfx6 + mNeurod1	****	****	****	ns	****	ns
mRfx6 vs mRfx6 + mNkx2.2	****	****	****	**	****	ns
mRfx6 vs mRfx6 + mNeurod1	****	ns	ns	ns	ns	ns

Amendt *et al.,* 1999). There are three 5′TNNAGTG in the 1.2 kb construct, two in the 0.8 kb construct and only one in the 0.6 kb construct and all constructs contain four 5′TAAT motifs (**FIGURE C2A**). If the mentioned motifs are important for the transactivation of *mMlxipl* by mNKX2.2, our results would suggest that maximum activation can be achieved if two of these binding sites are present, while the addition or removal of a binding site reduces transactivation.

The co-expression of *mNkx2.2* and *mRfx6* cDNAs doubled the induction of the 0.8 kb construct (**FIGURE C7C**) and tripled the induction of the 1.2 construct (**FIGURE C7B**) while the induction of the 0.6 kb construct was only slightly increased by their co-expression (**FIGURE C7D**). The deletion of the xbox motif had no influence on the increased induction of the 1.2 kb construct as expected because neither its induction by mRFX6 nor mNKX2.2 alone was affected by its deletion (**FIGURE C7B**). By contrast, there was no synergy in transactivation assays with the 0.8 kb and 0.6 kb Δxbox 2 constructs (**FIGURE C7C+D**). One would have expected that the value in these two cases would at least equal the sum of the individual values of activation, but it was more akin to activation by mNKX2.2 alone. This observation suggests that simultaneous binding of mRFX6 to xbox 2 and binding of mNKX2.2 to its binding site is necessary to achieve a synergistic effect.

The expression of *mNeurod1* alone did not induce luciferase activity regardless of the used construct (**FIGURE C7B+C+D**), and its combined expression with *mRfx6* only slightly increased the RFX6-mediated activation of the 1.2 kb reporter (**FIGURE C7B**), but this weak synergistic effect could not be reproduced with the 0.8 kb and the 0.6 kb construct (**FIGURE C7C+D**).

NEUROD1 binds to E-boxes with the consensus sequence 5′CANNTG (Naya *et al.,* 1995). We found four E-box motifs in our 1.2 kb construct and three E-box motifs in the 0.8 kb and the 0.6 kb construct (**FIGURE C2A**). The three E-box motifs present in the 0.8 kb and 0.6 kb fragments are located 139 bp upstream of xbox 2, between xbox 2 and xbox 3 and 135 bp downstream of xbox 3. The additional E-box in the 1.2 kb construct is 11 bp downstream of xbox 1. As we did not observe any additional activation of the 0.8 kb and 0.6 kb constructs in HEK293T cells co-transfected with *mNeurod1* and *mRfx6* cDNAs, one could imagine that there is a steric hindrance caused by the binding of 3HA-mRFX6 to xbox 2 and xbox 3. On the other hand, the deletion of xbox 2 made no difference (**FIGURE C7C+D**). Other possible explanations for the at most only weak involvement of mNEUROD1 in our *mMlxipl* transactivation assay are the N-terminal HA-tag fused to the *mNeurod1* coding sequence in our expression vector that could affect its functionality.

NEUROD1 belongs to the tissue-specific bHLH transcription factors that heterodimerize with ubiquitously expressed bHLH transcription factors. In the hamster pancreatic beta cell line HIT, NEUROD1 was found to form a heterodimer with TCF3 (Naya *et al.,* 1995). According to the Human Protein Atlas (https://www.proteinatlas.org/), TCF3 is only weakly expressed in HEK293T cells. However, it might be possible that NEUROD1 could also heterodimerize with another bHLH transcription factors

expressed in HEK293T cells, even if it is the human isoform. Thus, the fact that we did not overexpressed a putative dimerization partner does not necessarily mean that NEUROD1 could not bind to the plasmid DNA and activate the constructs, but the absence of mNKX2.2 or of other pancreatic transcription factors could preclude a contribution of mNEUROD1.

1.5.4. NKX2.2 does not participate in *rMlxipl* regulation in Ins-1 832/13 cells

In comparison to all the other factors that we tested, mNKX2.2 was the only one that could transactivate *mMlxipl* intron 1 in absence of mRFX6 and that gave rise to a strong synergetic effect in combination with mRFX6. Therefore, we decided to assess more precisely the role of NKX2.2 in *Mlxipl* transactivation and regulation in a beta cell context (**FIGURE C8**).

We downregulated *rRfx6* and *rNkx2.2* by RNA interference in the rat pancreatic beta cell line Ins-1 832/13, transfected the cells after 48 h with the *mMlxipl* intron 1 1.2 kb, 0.8 kb and 0.6 kb reporter construct with or without xbox 2, and another 48 h later, we measured luciferase activity (**FIGURE C8A**). As observed before, the maximal induction of the firefly luciferase reporter in the siControl-treated cells was also negatively correlated with the length of the genomic DNA fragment. We got the lowest activation obtained with the 1.2 kb construct and the highest with the 0.6 kb construct. The deletion of xbox 2 strongly decreased the activation suggesting again that an xbox binding protein plays a major role in *mMlxipl* transactivation. The downregulation of *rRfx6* strongly decreased the induction of each reporter construct (**FIGURE C8A**). This reaffirms our findings that RFX6 is the key factor in the transcriptional regulation of *Mlxipl*.

Surprisingly, the downregulation of *rNkx2.2* had no effect on *mMlxipl* intron 1 transactivation in Ins-1 832/13 cells although NKX2.2 had strongly activated the same constructs in HEK293T cells. There seems to be a large discrepancy between what can be obtained as results with an overexpressed protein in HEK293T cells and with an endogenous protein in beta cells. The fact that we used the mouse genomic *mMlxipl* intron 1 sequence in combination with the rat endogenous transcription factor should not cause a problem in the case of NKX2.2 because 270 out of 273 amino acid residues are identical in the mouse and rat protein.

We next tested the consequences of *rNkx2.2* knockdown on *rMlxipl* expression by western blotting (**FIGURE C8B+C**). Although the knockdown of rNKX2.2 was very efficient (**FIGURE C8B**), we did not observe any reduction in rMLXIPL protein amount (**FIGURE C8C**). As expected, the knockdown of *rRfx6* decreased rMLXIPL protein (**FIGURE C8B+C**). These results are in perfect agreement with those of the transactivation assay in Ins-1 832/13 cells treated with siNkx2.2 or siRfx6.

To asses whether rNKX2.2 might participate in *rMlxipl* regulation in a glucose challenge (**FIGURE C9**), we treated Ins-1 832/13 with siControl and siNkx2.2 for 24 h, starved the cells overnight in low

glucose medium and incubated them subsequently either in low or high glucose medium for 6 h prior to RT-qPCR analysis (**FIGURE C9A**). We verified that the glucose induction had worked, by analysing the induction of *rTxnip*, a gene whose expression is known to be activated by stress but also by high glucose concentrations (Shalev, 2014, review). The expression of *rTxnip* was highly increased in Ins-1 832/13 cells that were incubated in high glucose medium compared to the ones in the low glucose medium (**FIGURE C9B**).

We next tested the downregulation of *rNkx2.2*. The treatment with siNkx2.2 efficiently reduced *rNkx2.2* expression in both low and high glucose concentrations. Interestingly, we observed a strong decrease of the *rNkx2.2* transcript level in siControl-treated cells after an incubation in high glucose medium. This must be a cell-intrinsic mechanism induced by glucose. We also detected less *rNkx2.2* transcripts in siNkx2.2-treated cells that were incubated in high glucose medium than in the ones kept in low glucose medium. The glucose-induced downregulation of *rNkx2.2* apparently also takes place when the *rNkx2.2* level is already reduced (**FIGURE C9A+B)**.

The downregulation of *rNkx2.2* slightly increased *Mlxipl* levels in high glucose while *Mlxipl* expression remained unchanged in low glucose. An analysis of the isoform-specific expression revealed that this probably resulted from a weak increase in *rMlxipl* β expression. When the cells are exposed to high glucose, MLXIPL α activates the expression of *Mlxipl* β by binding to two carbohydrate response elements (ChoRE) in the promoter region of *Mlxipl* β in exon 1b, the first exon of *Mlxipl* β and an exon that is not common between both isoforms. The initiation of *Mlxipl* β expression prevents the transcription of *Mlxipl* α. As a result, the transcript level of *Mlxipl* α drops while that of *Mlxipl* β increases. This regulatory mechanism is conserved between mouse, rat and human (Herman *et al.,* 2012; Zhang *et al.,* 2015). Furthermore, it has been suggested that MLXIPL β autoregulates its own transcription by binding to the two ChoREs in its promoter region (Herman *et al.,* 2012). The inhibition of *rMlxipl* β due to the knockdown of *rNkx2.2* likely resulted from a mechanism independent of *rMlxipl* α regulation because *rMlxipl* α remained unchanged in siNkx2.2-treated cells.

Taken together, the results of the transactivation assay, the western blot and the qPCR suggest that NKX2.2 does not regulate the transcription of *Mlxipl* α by interacting with *mMlxipl* intron 1 in adult pancreatic beta cells. These conclusions contradict the results of Gutierrez *et al.* and Churchill *et al.,* who demonstrated in their ChIP sequencing and RNA sequencing experiments that MLXIPL is a direct target gene of NKX2.2 (Gutiérrez *et al.,* 2017; Churchill *et al.,* 2017). A possible explanation for this contradiction could be that NKX2.2 regulates *Mlxipl* exclusively in immature beta cells. The anti-NKX2.2 ChIP experiment was performed on Min6 cells (Gutiérrez *et al.,* 2017). There is some evidence that the heterogenous Min6 cell line consists of PDX1-postive / insulin-low immature cells and PDX1-positive / insulin-high mature beta cells (Szabat *et al.,* 2009). The study further showed that *Nkx2.2* expression was higher in the more immature Min6 cells than in the mature Min6 cells. In the *mNkx2.2* Δbeta mice,

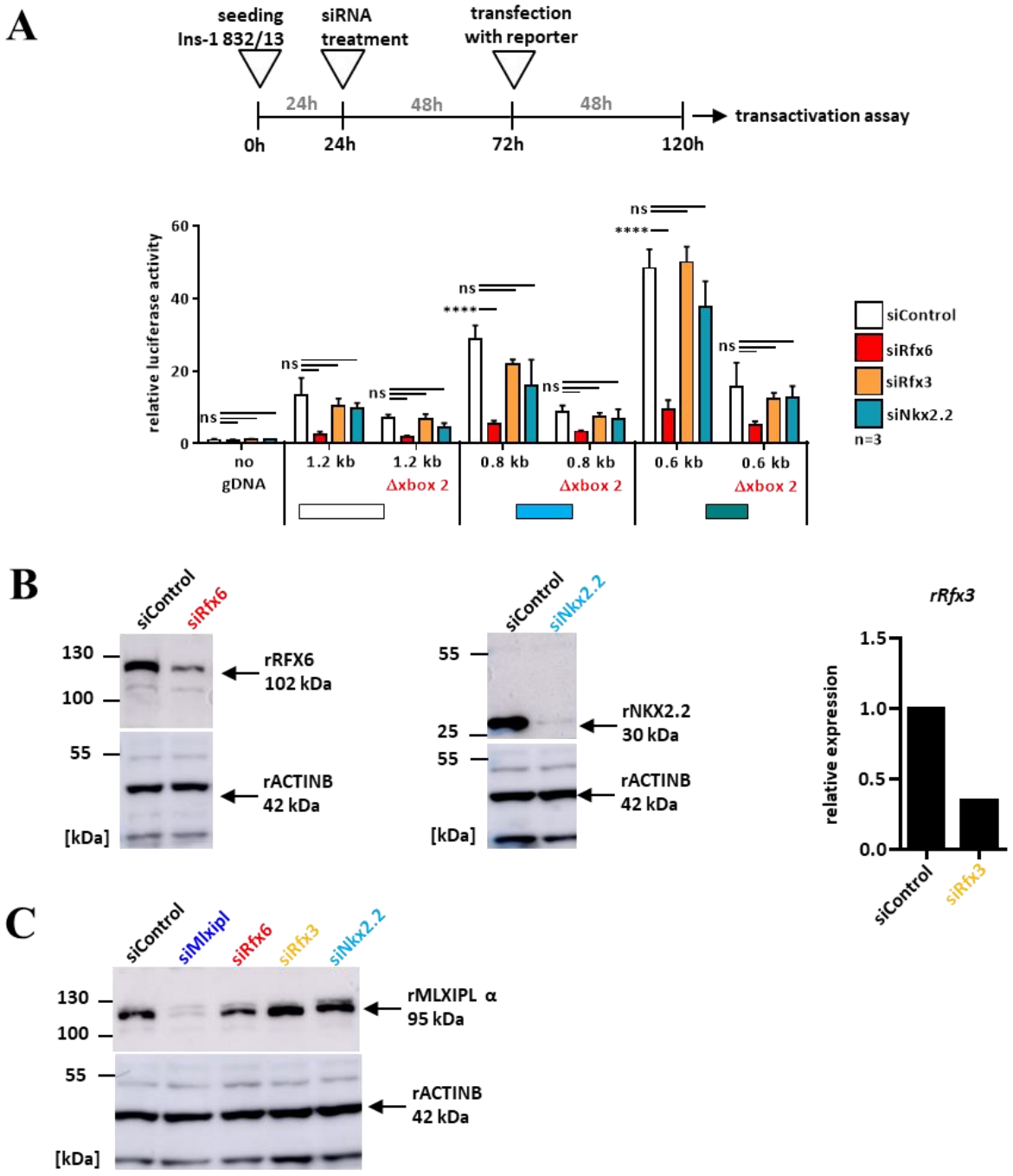


FIGURE C8: The knockdown of *rRfx6* but not of *rRfx3* and *rNkx2.2* affects *rMlxipl* protein transactivation and expression in Ins-1 832/13 cells. (A) Transactivation assay analysing the activation of the wild-type or Δxbox 2 *mMlxipl* intron 1 1.2 kb, 0.8 kb or 0.6 kb Firefly luciferase reporter constructs in Ins-1 832/13 cells treated with siControl, siRfx3 or siNkx2.2.48 h after the siRNA treatment, the cells were transfected with the reporter constructs and luciferase activity was measured after another 48 h. Data are shown as mean plus standard deviation (n=3). Statistical significances (p≥0.05 not significant (ns), p<0.0001 ****) were determined by two-way ANOVA. **(B)** Validation of the knockdown of *rRfx6* and *rNkx2.2* by western blot (rabbit anti-RFX6 2766 and mouse anti-NKX2.2) and of *rRfx3* by RT-qPCR after siRNA treatment of Ins-1 832/13 cells. **(C)** Western blot measuring the expression of rMLXIPL in siControl, siMlxipl, siRfx6, siRfx3 and siNkx2.2-treated Ins-1 832/13 cells (SmartPool siRNAs, rabbit anti-MLXIPL, Novus). RFX6 and MLXIPL **(B, C)** migrate at a higher molecular weight than expected probably due to posttranslational modifications.

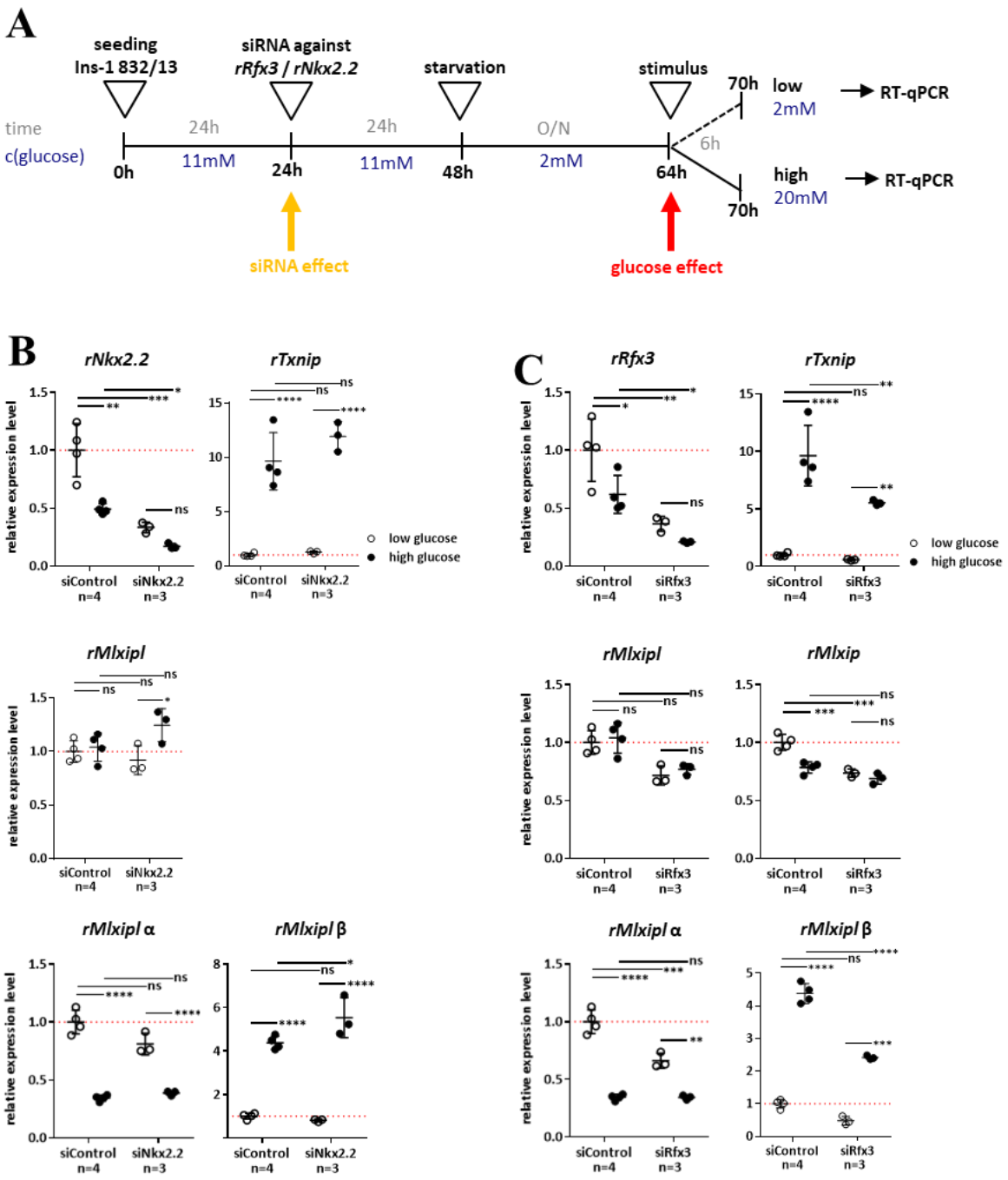


FIGURE C9: The knockdown of *rRfx3* but not of *rNkx2.2* reduces *rMlxipl* expression in a glucose challenge in Ins-1 832/13 cells. (A) Experimental strategy. Ins-1 832/13 cells were transfected with siControl, siRfx3 or siNkx2.2. After 24 h, the transfection medium was replaced by low glucose medium (standard cell culture medium with 2 mM glucose), and, after an overnight incubation, with low or high glucose medium (standard cell culture medium with 2 mM or 20 mM glucose, respectively). RNA was isolated after 6 h and transcript levels were measured by RT-qPCR. **(B, C)** RT-qPCR analysis of *rNkx2.2*, *rRfx3*, *rTxnip*, *rMlxipl*, *rMlxipl* and *rMlxipl* β in siNkx2.2-treated **(B)** or siRfx3-treated **(C)** Ins-1 832/13 cells in low and high glucose. Data are shown as mean plus standard deviation (n=4 for siControl, n=3 for siRfx3 and siNkx2.2). Statistical significances (p≥0.05 not significant (ns), p<0.05 *, p<0.01 **, p<0.001 ***, p<0.0001 ****) were determined by two-way ANOVA.

the knockout of *mNkx2.2* leads to an increase of immature beta cells (Gutiérrez *et al.,* 2017). It might be possible that NKX2.2 regulates *Mlxipl* exclusively in these de-differentiated cells that have newly emerged after tamoxifen injection. Interestingly, we also found a downregulation of *mMlxipl* in E15.5 embryos in the pancreas of *mRfx6* -/- mice and *mRfx6* Δendo mice (*mRfx6* fl/fl; Ngn3-Cre; unpublished data of Julie Piccand). In summary, the regulation of *Mlxipl* by NKX2.2 could be limited to differentiating islet cells and might not occur in mature beta cells. During embryogenesis, *Mlxipl* could be coregulated by NKX2.2 and RFX6 whereas NKX2.2 might not be involved in adult beta cells.

1.5.5. RFX3 activates *Mlxipl* transcription in a glucose challenge

We also tested the involvement of a putative RFX6 dimerization partner in *Mlxipl* regulation. As mentioned above, RFX proteins bind to DNA as homo- or heterodimers with another RFX protein (Reith *et al.,* 1994). An EMSA with the hepatitis B virus xbox motif and *in vitro* translated hRFX6 and hRFX3 proteins revealed that RFX6 and RFX3 can bind to DNA as a homodimer or by forming an RFX6 / RFX3 heterodimer (Smith *et al.,* 2010). To date, there is no proof of these *in vitro* data *in vivo*. In the future, it will be essential to discover direct interaction partners of RFX6 using mass spectrometry to gain further insights into how RFX6 works.

The role of the RFX3 in the pancreas has been discussed in the introduction part of this thesis. In brief, *Rfx3* is expressed in the developing and in all mature endocrine cells of the pancreas. In the global *mRfx3* knockout mice, the islets are disorganized and contain less glucagon-, insulin and ghrelin-producing cells while the number of PP-positive cells increases. Somatostatin-producing cells are not affected. Insulin secretion and glucose clearance are impaired in adult mice (Ait-Lounis *et al.,* 2008). Like RFX6, NKX2.2 and NEUROD1 (Piccand *et al.,* 2014; Gutiérrez *et al.,* 2017; Gu *et al.,* 2010), RFX3 plays a crucial role in the maturation of beta cells as 80 % of the developing beta cells in the adult *mRfx3* knockout mice are blocked at a progenitor state and express only low amounts of beta cell genes (Ait-Lounis *et al.,* 2010).

To gain further insights into the role of RFX3 in *Mlxipl* regulation in pancreatic beta cells and especially in a possible involvement in *mMlxipl* intron 1 transactivation in cooperation with RFX6 (**FIGURE C8**), we tested the effect of *rRfx3* downregulation on the induction of the *mMlxipl* intron 1 reporter constructs in Ins-1 832/13 cells (**FIGURE C8A**). We did not observe any difference between siControl- and siRfx3-treated cells. Thus, an involvement of RFX3 in *mMlxipl* intron 1 transactivation and its binding to one of the five xbox motifs in the RFX6 peak region seem unlikely.

We further tested whether the knockdown of *rRfx3* affected rMLXIPL protein level in Ins-1 832/13 cells (**FIGURE C8B+C**). As we do not have an antibody that detects the endogenous RFX3 in western

blotting, we verified the knockdown by RT-qPCR (**FIGURE C8B**). The efficient downregulation of *rRfx3* did not altered rMLXIPL protein level (**FIGURE C8C**). These findings are consistent with the results of the transactivation assay indicating no role of RFX3 in *Mlxipl* regulation.

However, when we tested the consequences of *rRfx3* knockdown in a glucose challenge in Ins-1 832/13 cells (**FIGURE C9**), we observed that the downregulation of *rRfx3* caused a reduction in *rMlxipl* α transcript levels in low and a decrease of *rMlxipl* β transcripts in high glucose concentrations (**FIGURE C9A+C**). Either RFX3 directly regulates *rMlxipl* β or the observed decrease in *rMlxipl* β at high glucose levels could be a result of its diminished transcriptional activation by the reduced number of MLXIPL α molecules during the 6 h in high glucose medium after the starvation phase. If *Mlxipl* α is indeed a direct target gene of RFX3, the xbox motifs are either not in intron 1 or RFX3 regulates *Mlxipl* only during starvation, but not when more glucose is available, explaining why we had not observed any difference in siControl- and siRfx3-treated cells in the transactivation assay performed in Ins-1 832/13 cells in 11 mM glucose medium.

Interestingly, a high concentration of glucose lead to a reduction in *rRfx3* levels in Ins-1 832/13 cells treated with siControl or siRfx3 (**FIGURE C9C**). Furthermore, the induction of *rTxnip* was weaker in siRfx3-treated cells than in siControl-treated cells. It has been suggested that *Txnip* is a direct target gene of MLXIPL (Cha-Molstad *et al.,* 2009), but a recent paper showed that it is rather regulated by MLXIPL's paralog MLXIP (Richards *et al.,* 2018). MLXIPL and MLXIP have the same protein structure and, like MLXIPL, MLXIP forms a dimer with MLX and binds to ChoREs (Billin *et al.,* 2000; Havula and Hietakangas, 2008, review). The human, murine and rat promoter of *Txnip* contains a ChoRE (Minn *et al.* 2005), thus it could be a target of MLXIPL / MLX or MLXIP / MLX dimers. Richards *et al.* showed that *Txnip* expression was not affected in islets of *mMlxipl* -/- mice and that the knockdown of *MLXIPL* did not altered *TXNIP* expression in the human EndoC-betaH2 cell line while its expression was almost lost in siMlxip- or siMlx-treated cells (Richards *et al.,* 2018).

Given that in our experiment, *rTxnip* expression was reduced in siRfx3-treated cells in high glucose, we wondered whether this could also be due to a change in *rMlxip* or *rMlx* expression. Indeed, we observed a reduction in *rMlxip* transcript levels in starved Ins-1 832/13 cells treated with siRfx3 (**FIGURE C9C**). The expression of *rMlx* remained unchanged (**data not shown**). The reduction of *rTxnip* in high glucose could be the consequence of a reduction of *rMlxip* as proposed for the decrease of *rMlxipl* β in high glucose that could be the consequence of *rMlxipl* α reduction in low glucose. In conclusion, our results suggest that *rMlxipl* α and *rMlxip* could be target genes of RFX3 in hypoglycaemia.

1.5.6. The binding of MLXIPL to its first intron suggests an autoregulatory mechanism

Binding of MLXIPL α and β to *Mlxipl* exon 1b has recently been described (Herman *et al.,* 2012; Zhang *et al.,* 2015). Yet unknown is the fact that MLXIPL also binds to its first intron (**FIGURE C10**). During the analysis of the published anti-MLXIPL ChIP sequencing data produced in the rat pancreatic beta cell line Ins-1E (GSE81628, Schmidt *et al.,* 2016), we figured out that, apart from the important binding peak of rMLXIPL on exon 1b, there was an additional smaller binding peak of rMLXIPL on the

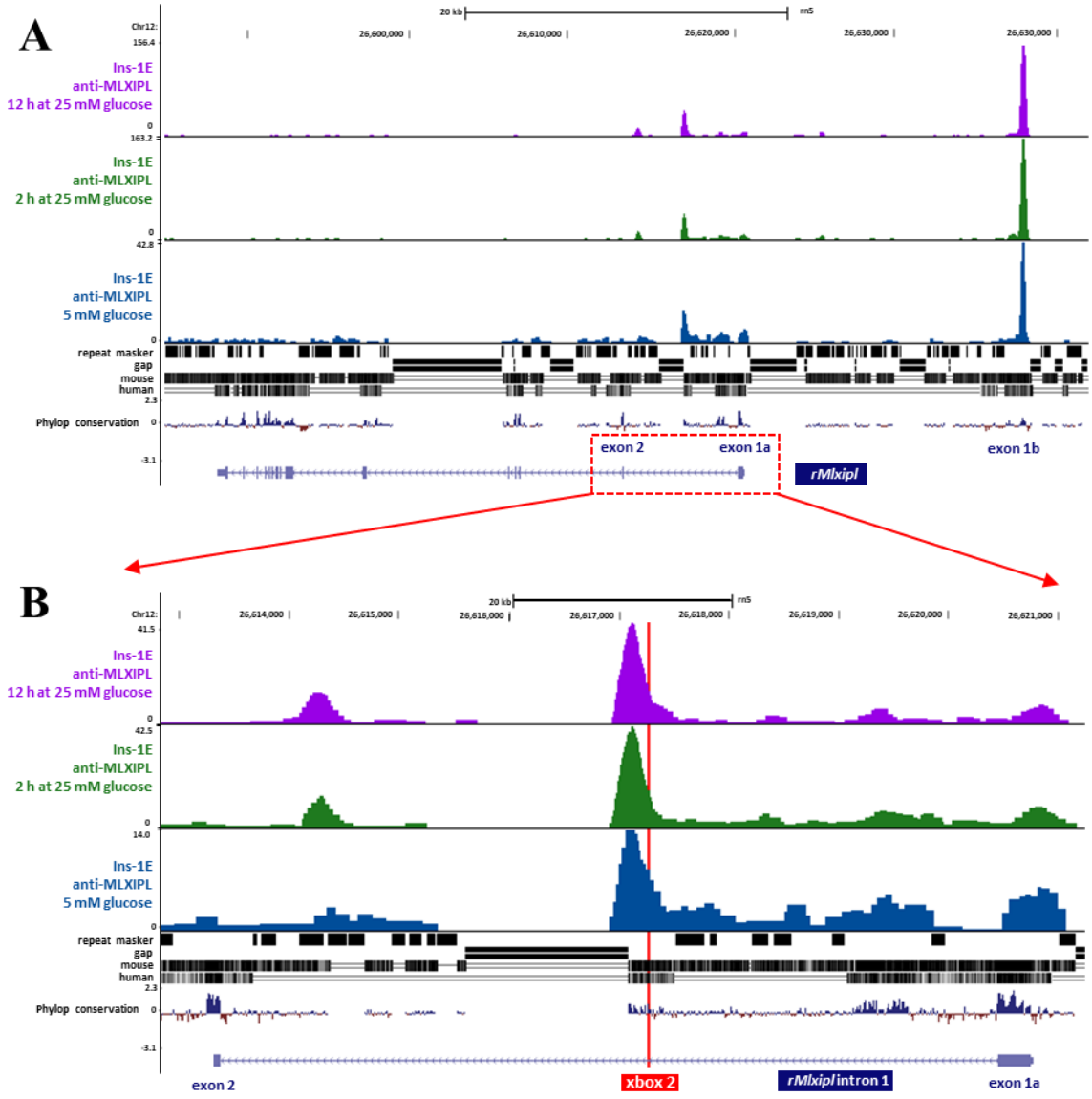


FIGURE C10: MLXIPL binds to the first intron of *rMlxipl* in Ins-1E cells. (A) ChIP sequencing data (GSE81628, Schmidt *et al.,* 2016) showing the binding of rMLXIPL on *rMlxipl* intron 1 in Ins-1E cells starved for 24 h in 5 mM glucose (blue profile), as well as after a subsequent incubation in 25 mM glucose for 2 h (green profile) or 12 h (purple profile). UCSC alignment (rn5) was done by Constance Vagne. **(B)** Zoom in on intron 1. The position of the xbox 2 motif is highlighted in red.

first intron of *rMlxipl* (**FIGURE C10A**). Based on the position of xbox 2 that is conserved in the genome of mice and rats, this peak seems to be exactly at the same place where the binding of mRFX6, mFOXA2, mNKX2.2 and mNEUROD1 occurs (**FIGURE C10B**; **FIGURE C1**; **FIGURE C6A**; **FIGURE C7A**). The binding of rMLXIPL to this region takes place in low and high glucose concentrations, but it is stronger in high glucose.

However, we were not able to identify any ChoRE (two E-box motifs separated by five nucleotides, consensus sequence 5'CACGTGNNNNNCACGTG, Shih *et al.,* 1995; Cairo *et al.,* 2001; Yamashita *et al.,* 2001) in the 1165 bp *mMlxipl* and the corresponding *rMlxipl* intron 1 region. The present E-box motifs are too wide apart (**FIGURE C2**). However, there might be an E-box-like motif forming a ChoRE wit one of the E-box motifs as it is also the case for one of the two ChoREs in *mMlxipl* exon 1b (Zhang *et al.,* 2015).

The antibody used by Schmidt *et al.* is directed against the C-terminus and recognizes both proteins (Schmidt *et al.,* 2016), thus *rMlxipl* intron 1 might be bound by rMLXIPL α or rMLXIPL β or both. In 2016, Jing *et al.* confirmed that the incubation of the Ins-1 cell line in high glucose medium triggered the translocation of rMLXIPL α into the nucleus, the transcriptional activation of *rMlxipl* β after 6 h and a detectable downregulation of *rMlxipl* α after 24 h, and they further showed that the treatment of Ins-1 cells with an siRNA directed against *rMlxipl* β lead to a drastic increase in rMLXIPL α nuclear shuttling (Jing *et al.,* 2016). Zhang *et al.* had postulated that the transcriptional initiation of *rMlxipl* β precluded the transcription of *rMlxipl* α, but Jing *et al.* concluded from their RNA interference assay that it is more likely a direct effect of rMLXIPL β on the activity of rMLXIPL α in a negative feedback-loop (Jing *et al.,* 2016). Given that rMLXIPL binds to its first intron, rMLXIPL β might directly inhibit the transcription of *rMlxipl* α causing the observed reduction in *rMlxipl* α transcript level. This might be the missing link between *rMlxipl* α and *rMlxipl* β regulation.

1.6. Conclusion of part 1 and graphical summary

When I started my thesis in 2015, no publication had ever reported about the transcriptional regulation of *Mlxipl* α by RFX6. In fact, little was known about the transcriptional control of *Mlxipl* α, whereas the regulation of *Mlxipl* β was already well described (Herman *et al.,* 2012; Zhang *et al.,* 2015). Before the discovery of the second isoform in 2012, different groups had studied the regulation of *Mlxipl* in the liver, and they identified the transcription factors OCT-1 as potential repressor and LXR, RXR and THR as activators of *Mlxipl* transcription (Sirek *et al.,* 2009; Cha and Repa, 2007; Hashimoto *et al.,* 2009). In pancreatic beta cells, *Mlxipl* expression is known to be regulated by FOXA1 and FOXA2 that bind to two conserved regions in the first intron of *Mlxipl* α. Knockout of both transcription factors causes a strong downregulation of *Mlxipl* expression (Gao *et al.,* 2010). The anti-RFX6 ChIP sequencing

experiments (Piccand *et al.,* 2014; unpublished data of Perrine Strasser) revealed that mRFX6 binds to the first intron of *mMlxipl* α and suggests that RFX6 controls the transcription of *Mlxipl* α. Given that *mMlxipl* α and β transcripts were almost lost in the islets of *mRfx6* Δbeta mice and that MLXIPL is important for both beta cell function and survival, we reasoned that the loss of mMLXIPL in the *mRfx6* knockout might largely contribute to the *mRfx6* Δbeta phenotype. Therefore, we started this thesis project in 2015 with the aim to understand the mechanism of *Mlxipl* regulation by RFX6. It was the sequel of the work published in 2014 (Piccand *et al.,* 2014) and the two thesis projects of Julie Piccand and Perrine Strasser, two former PhD students of the lab.

In this first part of the thesis, we confirmed our previous findings that mRFX6 is important for the transcriptional regulation of *mMlxipl* in murine beta cells and we further demonstrated for the first time that this mechanism is conserved in rat pancreatic beta cell line Ins-1 832/13.

We showed that mouse, rat and human RFX6 interacts with a highly conserved xbox within an active enhancer region in *Mlxipl* intron 1. A diminished *Rfx6* expression in murine pancreatic islets and murine and rat beta cell lines causes a reduction of *Mlxipl* transcript levels. A downregulation of *Rfx6* in the pancreatic beta cell line Ins-1 832/13 almost abolished *mMlxipl* intron 1 transactivation. The transactivation of *mMlxipl* intron 1 by RFX6 was affected when the conserved xbox motif was deleted or when point mutations in crucial RFX6 protein regions disturbed proper RFX6 function. Also, we described two novel RFX6 mutations, p.F294V and p.R578P. Taken together, these data proof that RFX6 is essential for the induction of *Mlxipl* transcription.

Moreover, with the discovery of the peak of MLXIPL on *Mlxipl* intron 1, we have most likely finally discovered a missing link that has been reported in the studies of Herman *et al.* and Zhang *et al.*: Although they had found that the transcription of *Mlxipl* α is inhibited in high glucose and MLXIPL β plays a crucial role (Herman *et al.,* 2012; Zhang *et al.,* 2015), they could not explain the exact mechanism. Since we knew that intron 1 is crucial for the transcriptional regulation of *Mlxipl* α, it was easy for us to find the peak in *Mlxipl* intron 1. Thus, in an auto-regulatory mechanism, MLXIPL induces the transcription of *Mlxipl* β in high glucose and represses simultaneously or staggered the transcription of *Mlxipl* α.

We further tested the potential of the four transcription factors E2F1, FOXA2 (Gao *et al.,* 2010), NEUROD1 and NKX2.2 (Churchill *et al.,* 2017) to coregulate *Mlxipl* together with RFX6. Transactivation assays in which we co-expressed each of these factors with RFX6, revealed that E2F1 worked as a co-repressor and NKX2.2 as a co-activator in *mMlxipl* intron 1 transactivation by RFX6, while we did not observe any contribution of FOXA2 and NEUROD1 in these assays although both are known to bind to *Mlxipl* intron 1 in mouse islets and the double knockout of *mFoxa1* and *mFoxa2* in murine islets significantly reduced *mMlxipl* transcript levels (Gao *et al.,* 2010; Churchill *et al.,* 2017). This suggests that *in vivo* data are not necessarily reproducible by the results of transactivation assays.

GRAPHICAL SUMMARY

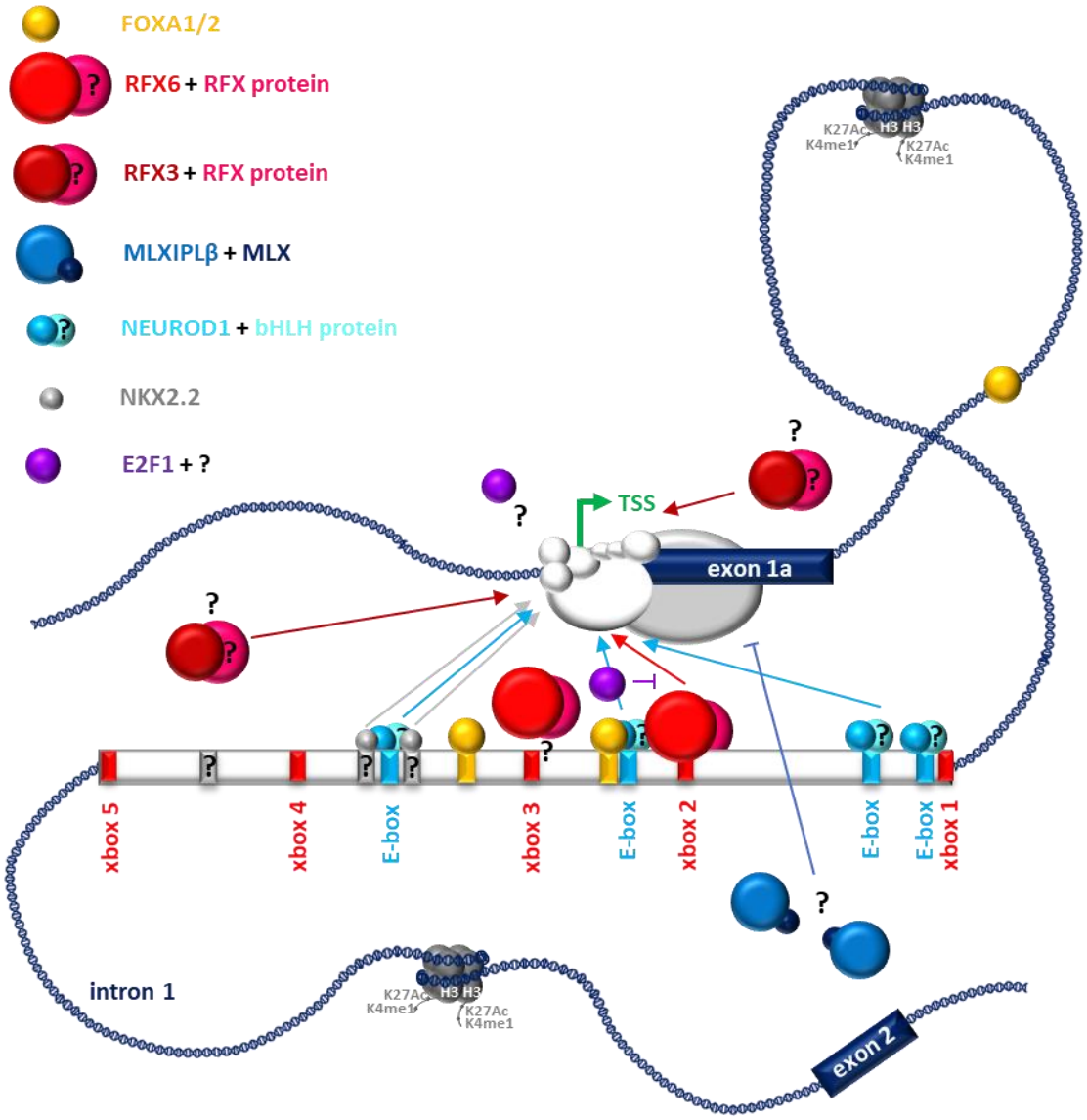


FIGURE C11: Hypothetical model of the transcription factor network responsible for *Mlxipl* regulation by intron 1. Graphical summary illustrating the results described in the first part of the thesis as well as published facts about the transcription factors that may be involved in the transcriptional regulation of *Mlxipl*. The model does not consider developmental and physiological effects.

Using Ins-1 832/13 cell, we determined the consequences of the *rRfx3* and *rNkx2.2* knockdown on *Mlxipl* expression and transactivation. Compared to *rRfx6* whose knockdown resulted in a strong decrease of *mMlxipl* intron 1 transactivation and *rMlxipl* α transcript and protein levels, the knockdown of *rRfx3* and *rNkx2.2* had no effect on the transactivation. Moreover, in siRfx3-treated cells, the expression of *rMlxipl* α was only slightly reduced in low glucose, while siNkx2.2 treatment did not cause

any consequences on *rMlxipl* α expression. Taken together, our data suggest that rRFX6 is the limiting factor for transcriptional activation and expression of *rMlxipl* in ins-1 832/13 cells.

From the co-occurrence analysis on our ChIP sequencing data to identify proteins that may be involved in the RFX6-mediated regulation of *Mlxipl* α transcription, we have derived further candidates that might co-regulate *Mlxipl* together with RFX6. Apart from E2F1, these factors include NRF1, CTCF, SP2, and KLF5 (unpublished data of Perrine Strasser and Stéphanie Legras) but we had no time to clone and test them.

Based on the published results and observations in the literature and our findings, one could imagine that FOXA2 together with FOXA1 opens the chromatin at the *Mlxipl* locus, thus allowing further transcription factors including RFX6 to bind to their respective sites. While NKX2.2, NEUROD1 and RFX3 more likely act in synergy with RFX6 enhancing *Mlxipl* transcription, E2F1 and MLXIPL β seem to be co-repressors (**graphical summary - FIGURE C11**). However, we still lack information about the beta cell-specific involvement of these transcription factors *in vivo*, especially regarding their role at different developmental stages and under different physiological conditions. and it is not a complete view of all involved transcription factors and DNA-binding proteins.

2. Identification of RFX6- and MLXIPL-regulated genetic programs in mature beta cells

In the following second part of my thesis, I focussed on the respective roles of RFX6 and MLXIPL in the regulation of pancreatic beta cells' genetic programs. The loss of *Mlxipl* expression in islets of adult *mRfx6* Δbeta mice (Piccand *et al.,* 2014) suggests that direct or indirect targets of MLXIPL constitute one subset of the dysregulated genes in the RFX6 knockout. Thus, we hypothesize that the RFX6 genetic program in beta cells is formed by genes that are regulated by RFX6 via its downstream target MLXIPL, genes regulated by RFX6 without any involvement of MLXIPL and genes that are controlled by RFX6 and MLXIPL. Identifying the respective targets of both genes should shed some light on their common and specific roles in adult beta cells and the contribution of MLXIPL to the *Rfx6 -/-* phenotype.

Our strategy was to produce knockout cells for *Rfx6* and *Mlxipl* in beta cell lines, to compare the deregulated genes in the *Rfx6* knockout with the deregulated genes in the *Mlxipl* knockout and to filter out those genes that were affected by either none, one, or both knockouts. This analysis will allow us to determine the number and nature of the specific and common target genes of both transcription factors and thus provide some information on the function of both genes at the transcriptional level. Moreover, given the fact that glucose plays a pivotal role in beta cell physiology and transcriptional regulation, and that MLXIPL α's activation and nuclear translocation and the expression of *Mlxipl* β are regulated by glucose, the influence of the glucose level on RFX6- and MLXIPL-dependent transcriptomes is a very important factor to be considered in the transcriptomes' analysis.

To answer these questions, we decided to use rodent beta cell lines as model system because they are easy to manipulate, the used medium has a defined glucose concentration, and we do not have the bias of the surrounding islet cells in the analysis. To invalidate *Rfx6* and *Mlxipl* in these cells, we developed a CRISPR / Cas9 strategy as main approach as well as RNA interference strategy as back-up approach. The analysis of the transcriptomes will be done by RNA sequencing to get information about the whole transcriptome.

2.1. Generation of *mRfx6* and *mMlxipl* knockout cell lines in the murine beta cell line Min6b1

We started this part of the project with the Min6b1 cell line because, in previous projects, Min6b1 cells were a useful tool to study RFX6 function (Piccand *et al*, 2014) and we had lists of RFX6-bound genes in Min6b1 cells (Piccand *et al.,* 2014; unpublished data of Perrine Strasser). In addition, we were more experienced with the use of the murine Min6b1 cell line than with the above-mentioned rat Ins-1 832/13 cells, and our laboratory has been working on mouse genes for a long time, so we had well-functioning qPCR primers for mouse transcripts and antibodies against mouse proteins. Thus, we thought using the mouse cell line would speed up our experiments.

To invalidate *mRfx6* and *mMlxipl* in Min6b1 cells, we decided to use the Cas9 nickase (Cas9n) with

two single guide RNAs (sgRNAs), a technique that is supposed to reduce off-target effects compared to knockout strategies based on the Cas9 with a single guide RNA (Jinek *et al.,* 2012; Ran *et al.,* 2013a; Ran *et al.,* 2013b).

2.1.1. T7EI assay to test the efficiency of *mRfx6* and *mMlxipl* sgRNA pairs in Min6b1 cells

We designed several sgRNA pairs to knockout *mRfx6* and *mMlxipl* in the murine beta cell line Min6b1 targeting either the translation start site (ATG) or essential protein domains (**FIGURE C12**). The three sgRNA pairs for *mRfx6* were directed against the ATG in the first exon (#2+18, #4+10) and against exon 3 (#2+19) that encodes the DNA-binding domain (**FIGURE C12A**). In *mMlxipl*, we targeted the ATG in the first exon of *mMlxipl* α (#10+15, #2+6), exon 6 (#2+10) which is the first common exon of *mMlxipl* α and *mMlxipl* β long enough to place two sgRNAs in its sequence, and exon 13 (#5+6) encoding the DNA binding domain (**FIGURE C12B**).

By a T7EI assay (see **PART B5.3**), we assessed the cleavage efficiency of the sgRNA pairs in Min6b1 cells. The Cas9n technique introduces insertions and deletions (INDEL) in the targeted loci. The re-hybridization of denatured PCR products that were amplified from these loci, leads to the formation of heteroduplexes with mismatches that are going to be cleaved after the addition of the T7EI enzyme that detects mismatches and extrahelical loops in double stranded DNA and cuts the DNA at these sites (Sadowski, 1971). The analysis of the fragmented PCR products on an agarose gel indicates if the amplified genomic DNA locus contained mutations or not.

All sgRNAs that we designed to target *mRfx6* and *mMlxipl* in Min6b1 cells, caused very efficiently the formation of mutations (**FIGURE C12A+B**). We detected even more cleavage products than uncut PCR products, with *mRfx6* sgRNA pair #2+19 in exon 3 as the only exception. Taken together, the designed sgRNAs were optimal tools to move on in our experiments and to start the generation of *mRfx6* and *mMlxipl* knockout cell lines in Min6b1 cells.

2.1.2. Summary of the attempts to knockout *mRfx6* and *mMlxipl* in Min6b1 cells

In the first attempt to knockout *mRfx6* and *mMlxipl* in Min6b1 cells (**TABLE C2**), we transfected wild-type Min6b1 cells with Cas9n expression vectors (pX462 with a puromycin selection gene) containing either the sgRNA pair #1+28 targeting *mRfx6* exon 1 to generate *mRfx6* knockout cell lines, or the sgRNA pair #10+15 targeting *mMlxipl* exon 1a to generate *mMlxipl* knockout cell lines.

After the transfection, we selected the transfected cells with 1.75 µg/ml of puromycin for 72 h and, afterwards, cultured the surviving cells for two weeks (see **PART B5.2**, **FIGURE B7**). We picked 24 clones per transfection and 13 putative *mRfx6* and 11 *mMlxipl* clones grew out of the 48 plated clones (24 for *mRfx6* sgRNA #1+28, 24 for *mMlxipl* sgRNA #10+15, **TABLE C2**).

To screen the clones, we amplified the targeted loci by PCR and sequenced the PCR products. Every clone contained INDEL mutations. We had not expected this to happen, and, as we had not transfected an additional vial with a Cas9n control vector without sgRNAs, we lacked an appropriate wild-type control cell lines for the following experiments. Instead, we used non-transfected wild-type cells of a comparable passage number as controls.

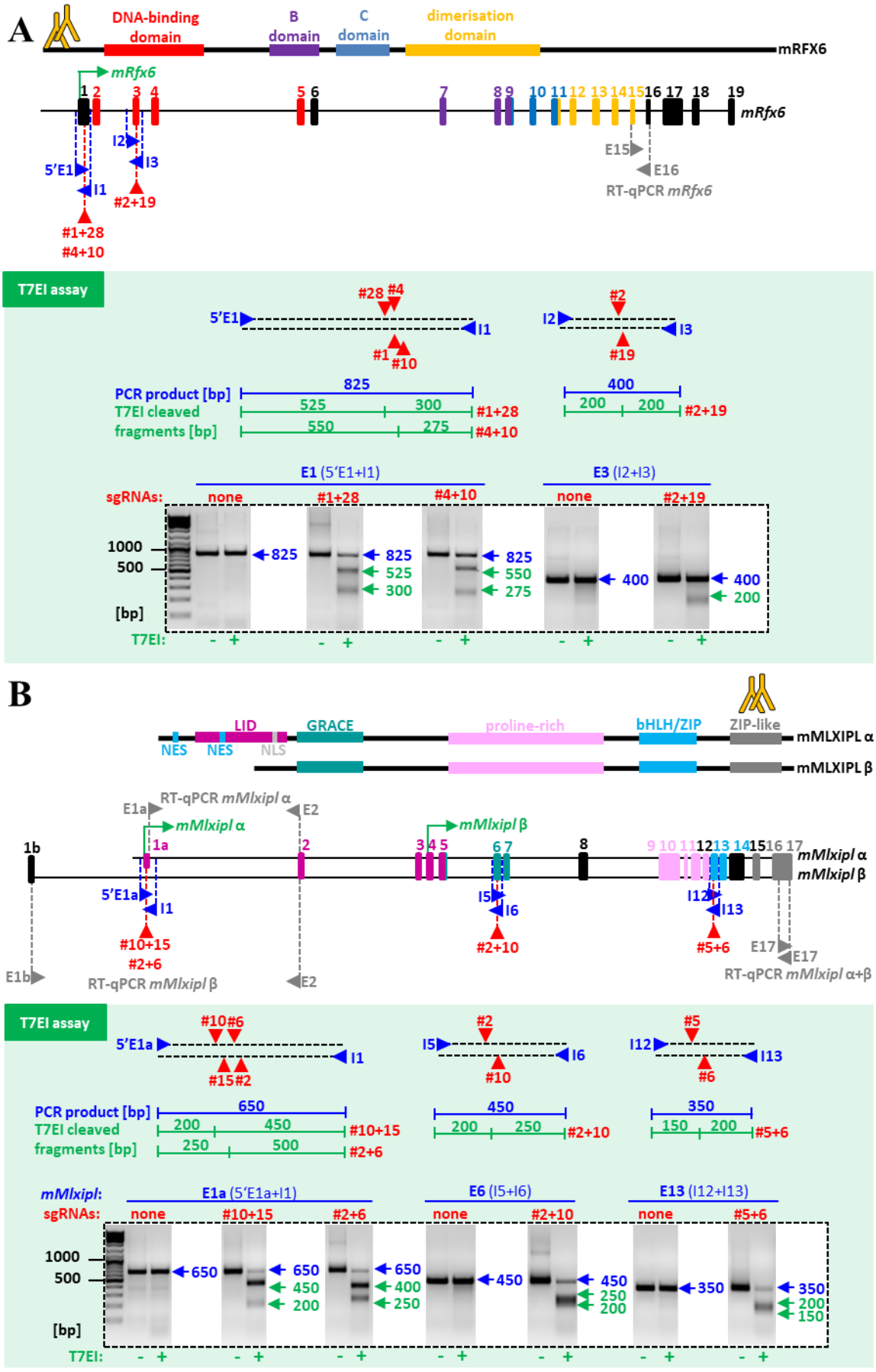
Only by performing multiple PCRs and sequencing reactions in the forward and reverse direction, we managed to properly analyse the genomic DNA mutations in all clones. Moreover, some clones were heterogeneous consisting of a mixture of cells with different genotypes that we had to subclone to separate the cell populations and to figure out if one of these subpopulations corresponds to a homozygous knockout clone. In the end, we identified one knockout clone for *mRfx6* with out-of-frame mutations in exon 1 on each allele (named KO mR-1; **TABLE C2**; **FIGURE C13A**) and one knockout clone for *mMlxipl* with out-of-frame mutations in exon 1a on each allele (named KO mM-1; **TABLE C2**; **FIGURE C14A**).

In parallel to the analysis of the INDEL in the 13 putative *mRfx6* and the 11 putative *mMlxipl* knockout clones, we also checked the expression of mRFX6 and mMLXIPL by western blotting using two different antibodies for each protein (rabbit anti-RFX6 #2766 and #2767; rabbit anti-MLXIPL of Novus and Abcam). The homozygous *mRfx6* knockout clone KO mR-1 and the homozygous *mMlxipl* knockout clone KO mM-1 were the only ones in which the mRFX6 and mMLXIPL band was completely lost (**FIGURE C13B**; **FIGURE C14B**). Thus, we confirmed that we produced a homozygous knockout cell line for each gene by analysing the DNA sequence as well as the protein expression.

After this first successful trial to generate homozygous *mRfx6* and *mMlxipl* knockout cell lines in Min6b1 cells, in two further attempts that will not be described in detail in this thesis, we tried to produce further knockout cell lines sgRNA pairs listed above (**FIGURE C12**) targeting exon 1 (sgRNA pair #1+28 and sgRNA pair #4+10) and exon 3 (sgRNA pair #2+19) of *mRfx6,* and exon 1a (sgRNA pair #10+15), exon 6 (sgRNA pair #2+10) and exon 13 (sgRNA pair #5+6) of *mMlxipl*, but we only obtained wild-type or heterozygous cell lines for both genes.

TABLE C2: Summary of the attempt to generate *mRfx6* and *mMlxipl* knockout cell lines in Min6b1 cells. Summary of the strategy used to generate knockout cell lines, the number of plated and grown clones, the obtained knockout and wild-type control cell lines, and the efficiency to generate knockout cell lines in proportion to the number of plated and analysed clones.

cell line	selection	gene	targeted exon	sgRNA pairs	plated clones	grown clones	knockout (-/-) cell lines	control (+/+) cell lines	# of KO lines / plated clones	# of KO lines / analysed clones
Min6b1	puromycin	*mRfx6*	1	#1+28	24	13	KO mR-1		1/24	1/13
Min6b1	puromycin	*mMlxipl*	1a	#10+15	24	11	KO mM-1		1/24	1/11


A
DNA-binding domain
B domain
C domain
dimerisation domain
mRFX6
mRfx6
mRfx6
5'E1
I1
I2
I3
#2+19
#1+28
#4+10
E15
E16
RT-qPCR mRfx6
T7EI assay
PCR product [bp]
T7EI cleaved fragments [bp]
825
525
300
550
275
400
200
E1 (5'E1+I1)
E3 (I2+I3)
sgRNAs:
none
#1+28
#4+10
#2+19
1000
500
[bp]
T7EI:
B
LID
GRACE
proline-rich
bHLH/ZIP
ZIP-like
NES
NES
NLS
mMLXIPL α
mMLXIPL β
RT-qPCR mMlxipl α
mMlxipl α
mMlxipl β
E1a
E2
1b
5'E1a
I1
I5
I6
I12
I13
#10+15
#2+6
#2+10
#5+6
E17
E1b
RT-qPCR mMlxipl β
RT-qPCR mMlxipl α+β
mMlxipl α
mMlxipl β
650
450
350
250
200
150
E1a (5'E1a+I1)
E6 (I5+I6)
E13 (I12+I13)

← FIGURE C12: T7EI assay to determine the efficiency of the *mRfx6* and *mMlxipl* sgRNAs to induce INDEL mutations at their target sites in murine Min6b1 cells. (A, B above) Structure of the *mRfx6* gene and mRFX6 protein **(A)** and *mMlxipl* gene and mMLXIPL α and β protein **(B)** pinpointing the position of the sgRNA pairs (red), PCR primer (blue), RT-qPCR primer (grey) and antibody binding sites (orange). The colour code of the exons indicates the encoded protein domain. **(A, B below)** Results of the T7EI assay assessing the CRISPR / Cas9-induced INDEL formation in the targeted *mRfx6* **(A)** and *mMlxipl* **(B)** loci. The cleavage efficiency in the T7EI test depends on the number of INDEL mutations in the amplified alleles produced by the Cas9n and the locus-specific sgRNAs. PCR products that differ in their sequences, form heteroduplexes containing mismatches. The endonuclease T7EI cleaves these heteroduplexes at the site where the mismatches are located, thereby producing smaller fragments (the expected sizes are indicated in the schemata).

2.1.3. The knockout of *mRfx6* in Min6b1 cells has no effect on its target gene *mMlxipl*

The homozygous *mRfx6* Min6b1 knockout cell line KO mR-1 was generated in the first attempt to generate *mRfx6* knockout cell lines in Min6b1 cells. In this attempt, we used the sgRNA pair #1+28. One allele in the KO mR-1 cells contains the mutation c.18_60del (Δ43 bp, p.E6DfsX57) and the other allele the mutation c.70_83del (Δ14bp, p.24_27del;p.C29LfsX14), both leading to a premature stop codon (**TABLE C2**; **FIGURE C13A**). In the western blot, we did not detect any residual protein in the KO mR-1 clone (**FIGURE C13B**). The transcript level of *mRfx6* was increased in the KO mR-1 clone compared to wildtype cells of a comparable passage number (**not shown**), but this is not surprising because we neither deleted the ATG in exon 1 nor a splicing site (**FIGURE C13A**) and the cells could have upregulated *mRfx6* transcription to compensate for the loss of the protein. Taken together, the nature of the mutations and the analysis of the mRFX6 protein expression both pointed to the fact that we produced a homozygous *Rfx6* knockout clone.

However, when we analysed the expression of the mRFX6 target gene *mMlxipl*, it was not affected in the *mRfx6* -/- clone compared to the wild-type control (**FIGURE C13C**). From the RNA interference assays described in results' part 1 (see **PART C1.3**), we know that a reduction in *mRfx6* in the Min6b1 cell line diminishes *mMlxipl* transcript level (**FIGURE C3D**). This should be reproducible in a knockout system unless the cells have implemented alternative regulatory mechanism to maintain *mMlxipl* expression in the absence of mRFX6 in the time between the generation of the knockout and the functional analysis. For instance, mNKX2.2 is known to bind *mMlxipl* in Min6 cells (Gutiérrez *et al.,* 2017; Churchill *et al.,* 2017; see **PART C1.5.3+1.5.4**; **FIGURE C7A**) and could thus have taken over for mRFX6.

Besides, one should also keep in mind that we targeted exon 1. Although we did not delete the ATG in exon 1, the cells be able to initiate the transcription from a downstream ATG. The first downstream in-frame ATG is at the beginning of exon 2. In fact, in the second attempt, we used sgRNA pairs directed against exon 3 that encodes the DNA-binding domain, to circumvent the problem that we might produce N-terminally truncated proteins that still bind to DNA, but we did not produce any

homozygous *mRfx6* knockout clone in this attempt, so we could not test this hypothesis. The N-terminal region of RFX6 has no known function (Aftab *et al.,* 2008; Sugiaman-Trapman *et al.,* 2018), thus a RFX6 variant with a shorter N-terminus might function as the full-length protein. There are two isoforms listed in NCBI, NM_001159389 we work with, and NM_177306 whose translation is initiated in exon 5 and that lacks the DNA-binding domain. The latter is probably not transcriptionally active and could at best act as coactivator or corepressor. We cannot detect these shorter N-terminally truncated isoforms in western blots using our antibodies because they are directed against the N-terminus of RFX6 and there is no available antibody that has its epitope at the C-terminus of RFX6. In summary, we do not think that the *mRfx6* Min6b1 knockout clone KO mR-1 is an adequate model to study the role of MLXIPL downstream of RFX6.

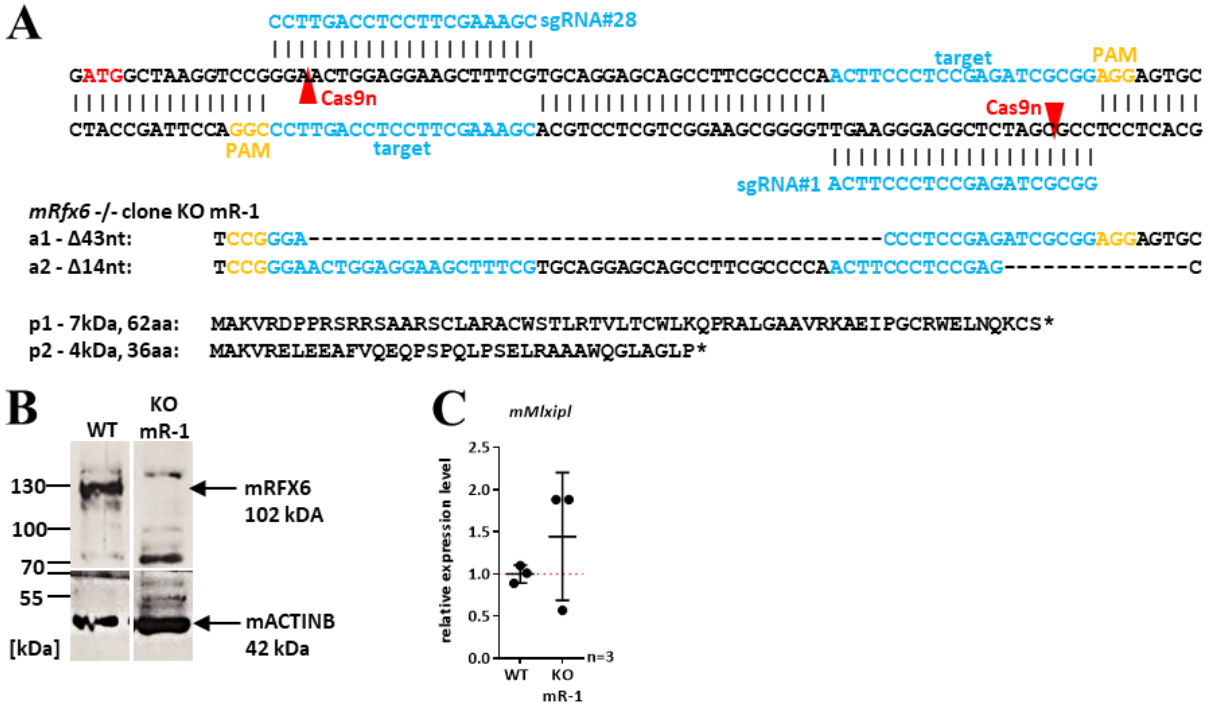


FIGURE C13: The expression of the mRFX6 target gene *mMlxipl* is not affected in the homozygous *mRfx6* -/- Min6b1 cell line. (A) Sequence of the mutated alleles in the homozygous *mRfx6* knockout Min6b1 cell line KO mR-1. **(B)** Western blot with an anti-RFX6 antibody (rabbit anti-RFX6 2767) on the homozygous *mRfx6* knockout Min6b1 cell line KO mR-1 and on wild-type Min6b1 cells (WT). **(C)** RT-qPCR analysis of the expression of the mRFX6 target gene *mMlxipl* in the homozygous *mRfx6* knockout Min6b1 cell line KO mR-1 and on wild-type Min6b1 cells (WT) in standard cell culture medium (25 mM glucose). Data are shown as mean plus standard deviation (n=3, technical triplicates).

2.1.4. The knockout of *mMlxipl* in Min6b1 cells has no effect on its target gene *mPklr*

The homozygous *mMlxipl* Min6b1 knockout clone KO mM-1 was generated in the same attempt in which we generated the *mRfx6* Min6b1 knockout cell line KO mR-1 (**TABLE C2**). We targeted the ATG in *mMlxipl* α's first exon 1a with the sgRNA pair #10+15. The selected clone KO mM-1 carries the mutations c.6_22del (Δ17 bp, p.A2AfsX13) and c.9_10insA; c.10_30del (Δ20 bp, p.4_10del; p.V11KfsX10) which cause the incorporation of a premature stop codon on both alleles (**TABLE C2**; **FIGURE C14A**). On a western blot, we did not detect any mMLXIPL band in the clone KO mM-1 (**FIGURE C14B**).

We measured the expression of the known mMLXIPL target gene *mPklr* to test if the expression of mMLXIPL target genes is affected by the loss of *mMlxipl* in the clone KO mM-1. It is important to keep in mind that the activity of rMLXIPL is regulated by glucose, thus we need to combine the analysis of the target gene expression with a glucose stimulus. The gene *mPklr* encodes an enzyme of the glycolysis and its expression is 2-fold upregulated in high glucose concentrations in siControl-treated Min6 cells and this glucose-dependent upregulation is lost in siMlxipl-treated Min6 cells (Da Silva Xavier *et al.,* 2006). By contrast, we did not observe an increased *mPklr* expression in the wild-type control cells in high glucose medium, and the expression of *mPklr* in the *mMlxipl* Min6b1 knockout clone KO mM-1 equalled the one in the wild-type control Min6b1 cells (**FIGURE C14C**).

We decided to examine the expression of two other genes, *Cx3cl1* and *Rasgrp2*, that we extracted from RNA sequencing data of a publication of Schmidt *et al*. published in 2016. They examined transcriptional changes in siMlxipl-treated Ins-1E cells in low (5 mM) glucose medium and after 12 h in high (25 mM) glucose medium and they compared the results with the ones measured in Ins-1E cells transfected with a scrambled siRNA (GSE81628, Schmidt *et al.,* 2016). The expression of *rCx3cl1* and *rRasgrp2* was strongly induced in high glucose concentrations in Ins-1E cells (10-fold for *rCx3cl1*, 50-fold for *rRasgrp2*) and the glucose-dependent induction was diminished in siMlxipl-treated Ins-1E cells compared to siControl-treated Ins-1E cells (6-fold for *rCx3cl1*, 20-fold for *rRasgrp2*). We hoped it would be easier to detect such large differences in the *mMlxipl* Min6b1 knockout cell line KO mM-1 than the slight differences in *mPklr* regulation (2-fold induction by glucose in siControl-treated Min6 cells, no glucose-induced upregulation in siMlxipl-treated Min6b1 cells, Da Silva Xavier *et al.,* 2006), but, in our Min6b1 control cells, the expression of *mCx3cl1* and *mRasgrp2* was neither induced by glucose in wild-type Min6b1 cells nor affected in the *mMlxipl* knockout clone KO mM-1 (**FIGURE C14C**). The non-activation by glucose could be due to cell line-specific differences and neither *mCx3cl1* nor *mRasgrp2* might be targets of mMLXIPL in Min6b1 cells.

Given that neither *mPklr* nor *mCx3cl1* or *mRasgrp2* were induced by glucose, we wanted to exclude the possibility that the *mMlxipl* Min6b1 knockout clone KO mM-1 and the wild-type control cells that

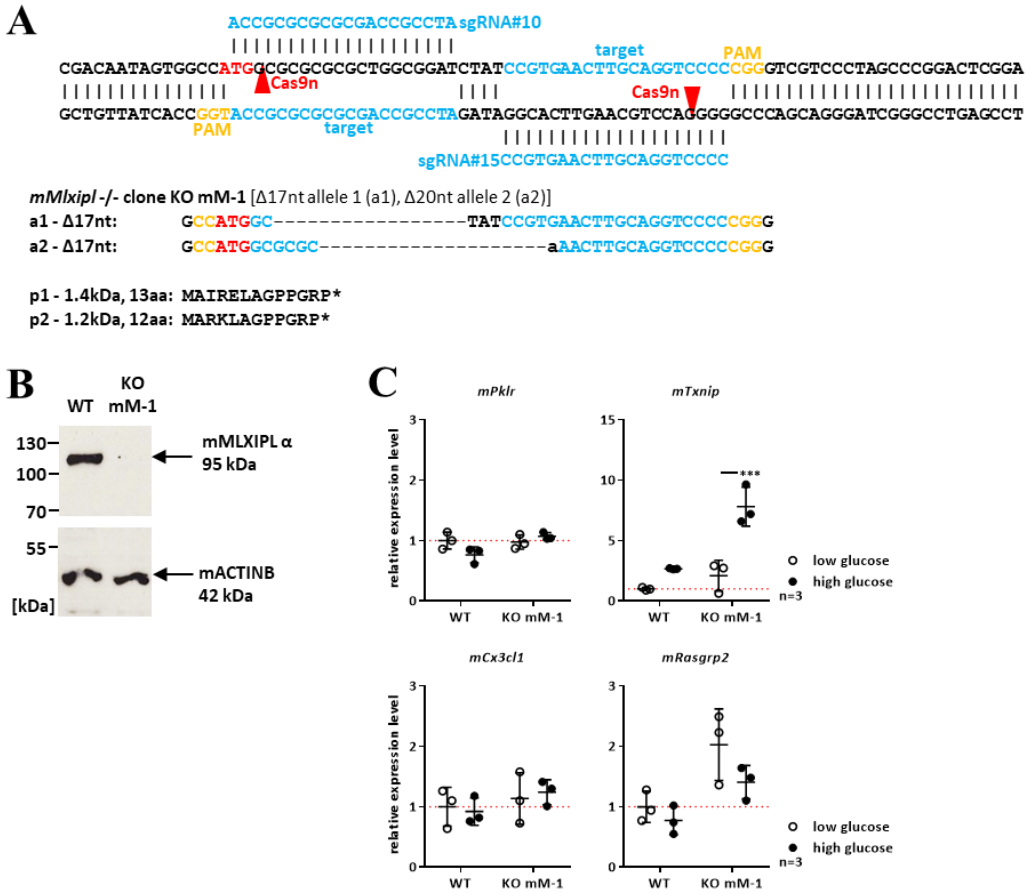


FIGURE C14: The expression of mMLXIPL target gene *mPklr* is not affected in the homozygous *mMlxipl* -/- Min6b1 cell line. (A) Sequence of the mutated alleles in the homozygous *mMlxipl* knockout Min6b1 cell line KO mM-1. **(B)** Western blot with an anti-MLXIPL antibody (rabbit anti-MLXIPL, Abcam) on the homozygous *mMlxipl* knockout Min6b1 cell line KO mM-1 and on wild-type Min6b1 cells (WT). **(C)** RT-qPCR analysis of the mMLXIPL target gene *mPklr,* the glucose-induced gene *rTxnip* and the putative mMLXIPL targets *mCx3cl1* and *mRasgrp2* in the homozygous *mMlxipl* knockout Min6b1 cell line KO mM-1 and on wild-type Min6b1 cells (WT) at 3 mM (low) and 30 mM (high) glucose. Control and *mMlxipl* -/- knockout cells were starved overnight in 3 mM glucose medium and incubated for 16 h in 3 mM or 30 mM glucose medium prior to analysis. Data are shown as mean plus standard deviation (n=3, technical triplicates). Statistical significances ($p<0.001$ ***) were determined by two-way ANOVA.

we used, had lost their glucose sensitivity during the cloning process. Therefore, we examined the expression of *Txnip*, a gene known to be upregulated by glucose (Shalev, 2014, review). Its expression was indeed upregulated in high glucose in the wild-type and in the *mMlxipl* knockout cells confirming that, despite their age, the cells were still able to sense the glucose level in the medium. Interestingly, the expression of *mTxnip* was higher in the *mMlxipl* Min6b1 knockout clone KO mM-1 than in the wild-type control (**FIGURE C14C**). Although we compared cells of comparable passage numbers, the wild-type cells were only in culture for 11 weeks at the time, while the clones had already been cultured for 9 months. Our strategy to generate Min6b1 knockout cell lines is based on the puromycin selection and limiting dilutions technique and necessitates at least five passages and 55 days in culture (see

PART B5.2, **FIGURE B7**). The clone KO mM-1 was passaged more than 20 times before the analysis (passage number > 40). Rani *et al.* showed that an increased expression of *mTxnip* is a main characteristic of Min6b1 cells that have lost their Glucose-Stimulated Insulin Secretion (GSIS) feature (Rani *et al.,* 2010). According to them, *mTxnip* expression increases already from passage 19 to 23 resulting in reduced GSIS. Cheng *et al.* analysed high passage Min6 cells and they revealed huge discrepancies between low passage and high passage cells: The high passage cells showed an increased glucose uptake rate, but the metabolism of glucose and the subsequent generation of ATP was drastically reduced. Consistently, they observed a decreased expression of glycolytic genes in these cells (Cheng *et al.*, 2012). Both findings explain the reduced GSIS observed in high passage cells by Rani *et al.* (Rani *et al.,* 2010). Considering these observations, we began to wonder if the Min6b1 knockout clones we can obtain at the end of our cloning process are still functional insulin-secreting beta cells, or if they are just immortalized cells without beta cell properties.

In summary, we managed to produce one *mRfx6* -/- (KO mR-1) and one *mMlxipl* -/- (KO mM-2) clone in Min6b1 cells, but we could not confirm the complete inactivation of both genes by the functional target gene analysis. Since there was no effect at all on the mRFX6 and mMLXIPL target gene expression in the clones KO mR-1 and KO mM-1, respectively, we do not think that they are the optimal tool to study RFX6- and MLXIPL-regulated genes.

2.2. Min6b1 cells, an adequate model to study glucose-dependent RFX6 and MLXIPL targets?

During the analysis of the *mMlxipl* Min6b1 knockout clone KO mM-1, we had noticed that the addition of glucose hardly influenced the transcript levels in the wild-type Min6b1 cells that we used as a control (**FIGURE C14C**). However, this may only be due to the high passage number of the cells (> 40 passages) and not to the cell line itself. In younger cells, glucose induction might work as expected. Therefore, we decided that it was necessary to figure out whether Min6b1 cells are a good model system to study glucose-dependent RFX6 and MLXIPL genetic programs.

2.2.1. Transcriptional adaptation after a glucose stimulus in Ins-1 832/13 and in Min6b1 cells

By the following experiments, we wanted to clarify which transcriptional changes we can expect in the Min6b1 cell line after a glucose stimulus. At the same time, we benefited from this series of experiments to test the glucose response in the Ins-1832/13 cell line, which we have already used successfully in the lab, and to find out whether this cell line may be an alternative to the Min6b1 cells in this part of the project. To address this question, we examined the glucose-induced expression changes in low passage number Min6b1 and Ins-1 832/13 cells after a glucose stimulus by RT-qPCR (**FIGURE C15**; **FIGURE C16**).

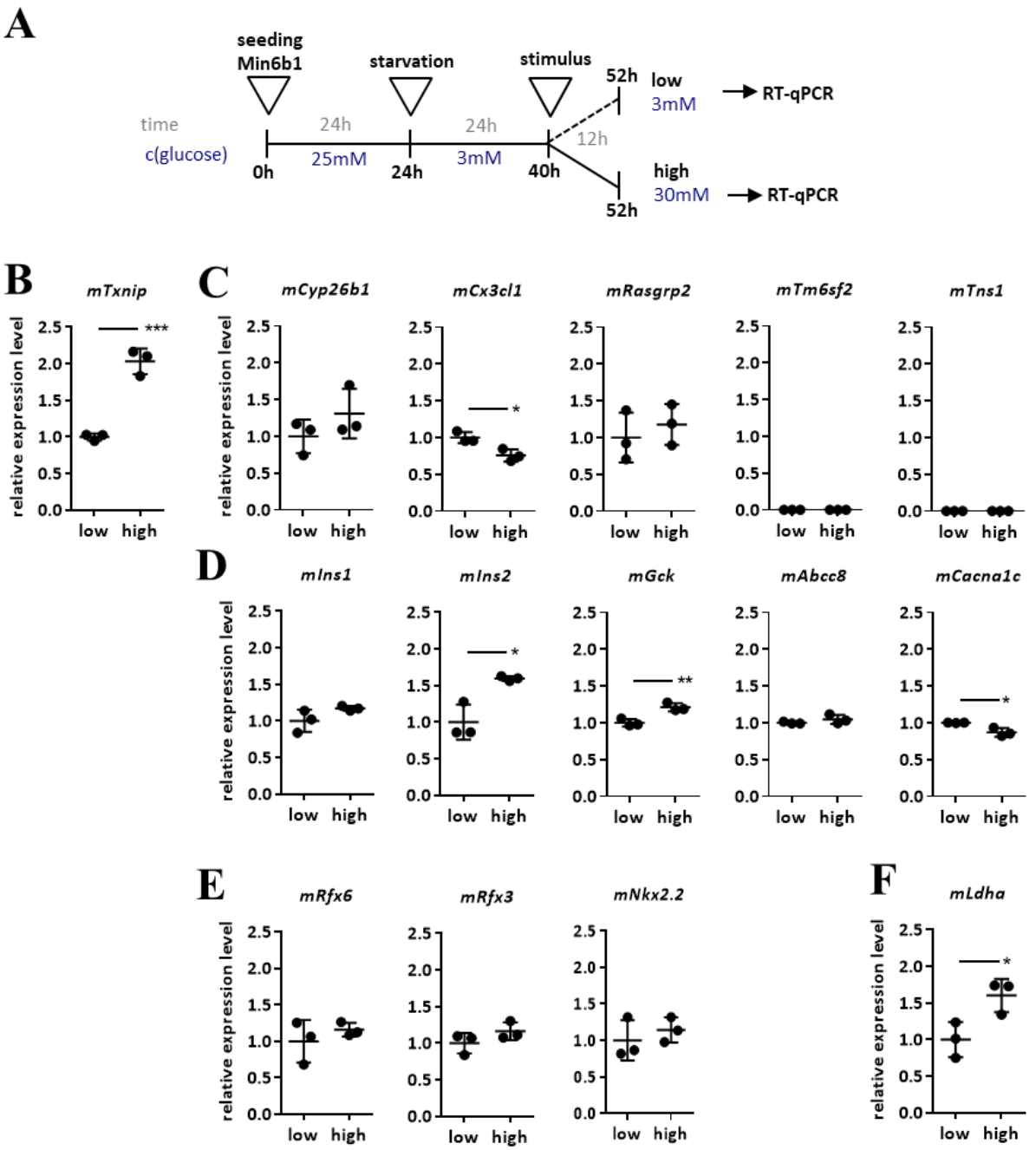


FIGURE C15: Gene expression changes in Min6b1 cells in response to glucose. (A) Experimental strategy. Min6b1 cells were seeded in standard cell culture medium (25 mM glucose). After 24 h, cells were starved for 24 h in standard cell culture medium supplemented with 3 mM glucose and incubated overnight (12 h) in low or high glucose medium (standard cell culture medium with 3 mM and 30 mM glucose, respectively) prior to RNA extraction and RT-qPCR. **(B-F)** RT-qPCR analysis determining the expression of the glucose-induced gene *mTxnip* **(B)**, , the genes *mCyp26b1*, *mCx3cl1*, *mRasgrp2*, *mTm6sf2* and *mTns1* all potentially induced by glucose **(C)**, the insulin secretion pathway genes *mIns1*, *mIns2*, *mGck*, *mAbcc8* and *mCacna1c* **(D)**, the transcription factors *mRfx6*, *mRfx3* and *mNkx2.2* **(E)**, and the disallowed gene *mLdha* **(F)**. Data are shown as mean plus standard deviation (n=3). Statistical significances (p<0.05 *, p<0.01 **, p<0.001 ***) were determined by unpaired two-tailed t tests.

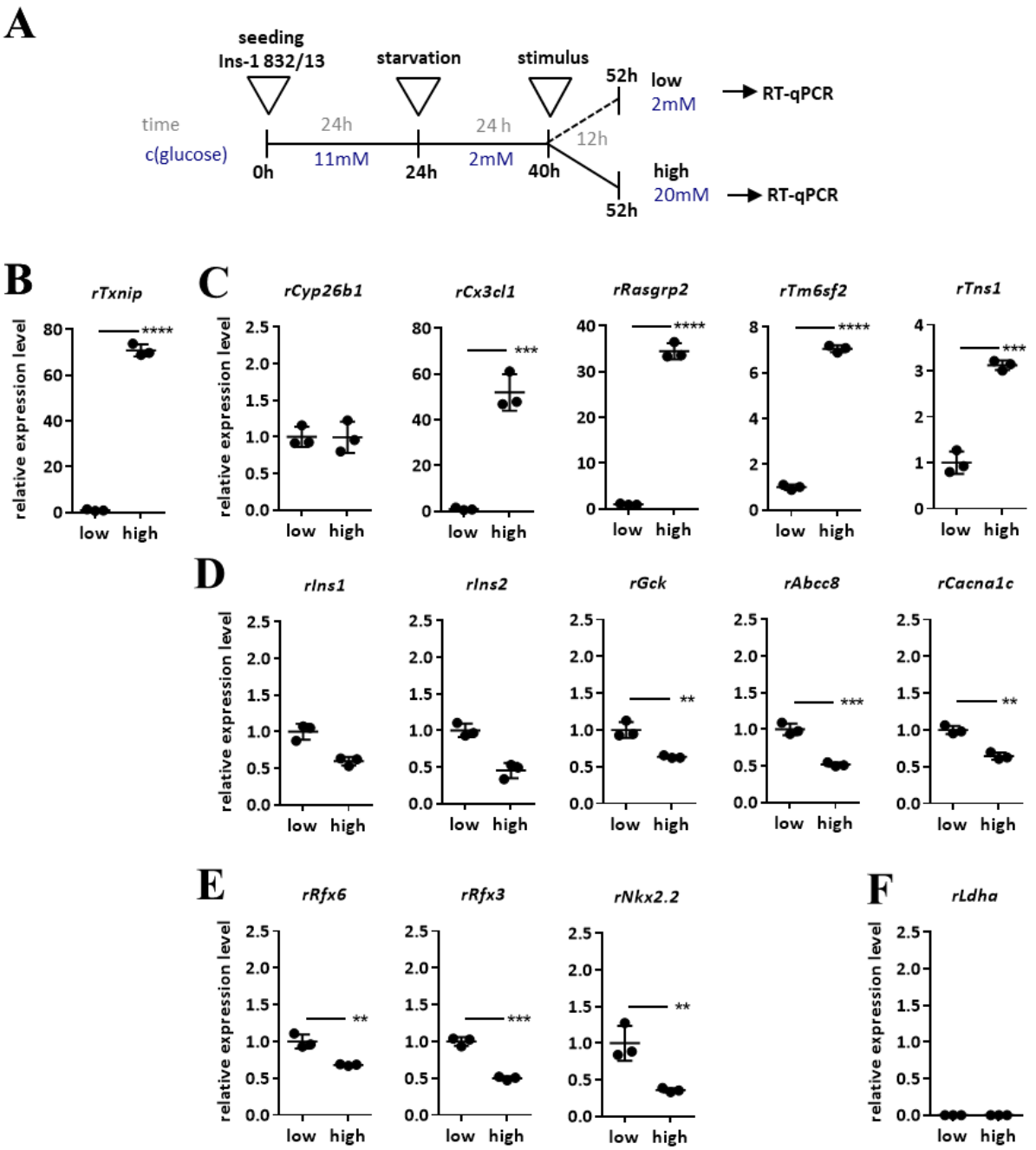


FIGURE C16: Gene expression changes in Ins-1 832/13 cells in response to glucose. (A) Experimental strategy. Ins-1 832/13 cells were seeded in standard cell culture medium (11 mM glucose). After 24 h, cells were starved for 24 h in DMEM supplemented with 2 mM glucose and incubated overnight (12 h) in low or high glucose medium (DMEM with 2 mM and 20 mM glucose, respectively) prior to RNA extraction and RT-qPCR. **(B-F)** RT-qPCR analysis determining the expression of the glucose-induced gene *rTxnip* **(B)**, the genes *rCyp26b1*, *rCx3cl1*, *rRasgrp2*, *rTm6sf2* and *rTns1* all potentially induced by glucose **(C)**, the insulin secretion pathway genes *rIns1*, *rIns2*, *rGck*, *rAbcc8* and *rCacna1c* **(D)**, the transcription factors *rRfx6*, *rRfx3* and *rNkx2.2* **(E)** and the disallowed gene *rLdha* **(F).** Data are shown as mean plus standard deviation (n=3). Statistical significances (p<0.05 *, p<0.01 **, p<0.001 ***) were determined by unpaired two-tailed t tests.

We starved the cells during 24 h in low glucose medium, incubated them overnight (12 h) in low or high glucose concentrations, isolated the RNA and measured transcript levels of selected candidates by RT-qPCR (**FIGURE C15A**; **FIGURE C16A**). We chose the 12 h for the incubation in high glucose medium to be able to compare the Ins-1 832/13 data with the published Ins-1E expression data generated by Schmidt *et al.* (GSE81628, Schmidt *et al.,* 2016). In fact, in their study, they not only investigated the consequences of the knockdown of *rMlxipl* on the transcriptome of Ins-1E cells in low and high glucose (2 h and 12 h incubation), but they also examined glucose-induced transcriptional changes in wild-type Ins-1E cells incubated for 0 h, 1 h, 2 h, 4 h and 12 h in high glucose medium (GSE81628, Schmidt *et al.,* 2016).

As candidate genes for the RT-qPCR analysis, we chose different genes whose expression, according to the literature, might be adapted in response to the glucose level. We re-tested the glucose-induced gene *Txnip* (Shalev *et al.,* 2002), and the genes *Cx3cl1* and *Rasgrp2* mentioned above, and we extracted three additional genes named *Cyp26b1*, *Tm6sf2*, and *Tns1* from the Schmidt *et al.* RNA sequencing data, all of which were more strongly expressed after the glucose stimulus (GSE81628, Schmidt *et al.,* 2016). Since the key point of this analysis is that the expression of these genes is potentially activated by glucose, their function will not be explained in more detail here. Moreover, we chose the insulin secretion pathway genes *Abcc8*, *Cacna1c*, *Gck*, *Ins1* and *Ins2*, since, after a glucose stimulus, beta cells upregulate the transcription of genes involved in insulin production and secretion (Schuit *et al.,* 2002, review) starting with the transcription of insulin itself (Efrat *et al.,* 1991). This enhances insulin secretion capacities and restores cellular insulin content. However, when beta cells start to proliferate, they need to dedifferentiate. Therefore, they downregulate insulin pathway genes, and, at the same time, they upregulate the expression of cell cycle regulators, proteins involved in DNA synthesis and components of the cytoskeleton to favour cell division (Schmidt *et al.,* 2016; Klochendler *et al.,* 2016). Webb *et al.* showed that Min6 cells increased the expression of secretory and metabolic genes after an 24 h incubation in 25 mM glucose medium and that these genes accounted for three-quarter of the upregulated genes (Webb *et al.,* 2000). Schmidt *et al.* showed that Ins-1E cells upregulate glycolytic and lipogenic genes in 2 hours after the stimulus, and that they subsequently repress the transcription of these metabolic genes and beta cell transcription factors and activate the transcription of proliferative genes in a second wave of transcriptional adaptation (Schmidt *et al.,* 2016). Since, in Ins-1 832/13 cells, we had already observed that the transcript levels of the transcription factors *rRfx3* and *rNkx2.2* are decreased after 6 h in high glucose medium (**FIGURE C9B+C**), a finding that is consistent with the observations of Schmidt *et al.* (GSE81628, Schmidt *et al.,* 2016), we wondered whether this might also be the case in Min6b1 cells. We also wanted to test if the expression of *Rfx6* is regulated in the same way as *Rfx3* and *Nkx2.2* in Ins-1 832/13 cells. Finally, we also added the disallowed gene *Ldha* to our list that belongs to a group of genes whose members are ubiquitously expressed but specifically

repressed in mature beta cells and should thus not be expressed in the Min6b1 and Ins-1 832/13 cells (Sekine *et al.,* 1994; Pullen *et al.,* 2010; Thorrez *et al.,* 2011).

In the Min6b1 cell line (**FIGURE C15B-F**), the only glucose-induced genes among the tested genes were *mTxnip* (2-fold induced, **FIGURE C15B**), *mCx3cl1* (slightly decreased, **FIGURE C15C**), *mIns2* and *mGck* (slightly induced, **FIGURE C15D**) and *mLdha* (2-fold induced, **FIGURE C15F**). The high expression of *mLdha* in Min6b1 cells is alarming since it is a gene that should not be transcribed by mature beta cells (Sekine *et al.,* 1994). However, in the Min6b1 cells we have got in the lab, *mLdha* was already highly expressed in starved Min6b1 cells and enriched after a glucose stimulus (**FIGURE C15F**).

In the Ins-1 832/13 cell line, almost all the tested genes were differentially expressed in cells incubated high glucose compared to the ones incubated in low glucose (**FIGURE C16B-F**). Glucose-induced genes were *rTxnip* (70-fold, **FIGURE C16B**), and *rCx3cl1* (50-fold), *rRasgrp2* (30-fold), *rTm6sf2* (7-fold) as well as *rTns1* (3-fold, all **FIGURE C16C**). The strength of the expression changes did not necessarily correspond to that measured in the Schmidt *et al.* data. This was likely due to technical differences (RNA sequencing versus RT-qPCR) and cell line-specific differences (Ins-1E versus Ins-1 832/13). The transcripts of the insulin secretion pathway genes *rAbcc8*, *rCacna1c*, *rGck*, *rIns1* and *rIns2* were halved in Ins-1 832/13 cells after 12 h in high glucose (**FIGURE C16D**), suggesting that they execute the second wave of transcriptional changes in which the expression of proliferative genes is stimulated, and the transcription of genes involved in the insulin pathway and beta cell transcription factors is inhibited (Schmidt *et al.,* 2016). Consistently, the expression of the transcription factors *rRfx6*, *rRfx3* and *rNkx2.2* was 2-fold lower in Ins-1 832/13 cells after 12 h in high glucose compared to cells kept in low glucose medium (**FIGURE C16E**). Last, the disallowed gene *rLdha* could not be detected in Ins-1 832/13 cells, neither at low nor at high glucose (**FIGURE C16F**).

In summary, the glucose-induced changes in Ins-1 832/13 cells are much stronger than in Min6b1 cells; the effect on *Txnip* for instance is 35 x higher in Ins-1 832/13 cells (70-fold) than in Min6b1 cells (2-fold). Considering the high expression of *mLdha* in Min6b1, one could further elaborate the theory that Min6b1 resemble more to immature than mature beta cells as proposed for the parental Min6 cell line (Szabat *et al.,* 2009).

2.2.2. The transactivation of *mMlxipl* exon 1b is glucose-independent in Min6b1 cells

Next, we analysed the glucose-induced transactivation of *Mlxipl* β by MLXIPL that is well described in mice, rats and humans (Herman *et al.,* 2012; Zhang *et al.,* 2015; Jing *et al.,* 2016). In fact, the transactivation of *mMlxipl* exon 1b is a glucose-dependent transcriptional regulation mechanism and *Mlxipl* β seems to be a bullet-proof target gene of MLXIPL, thus we thought that a *mMlxipl* exon 1b

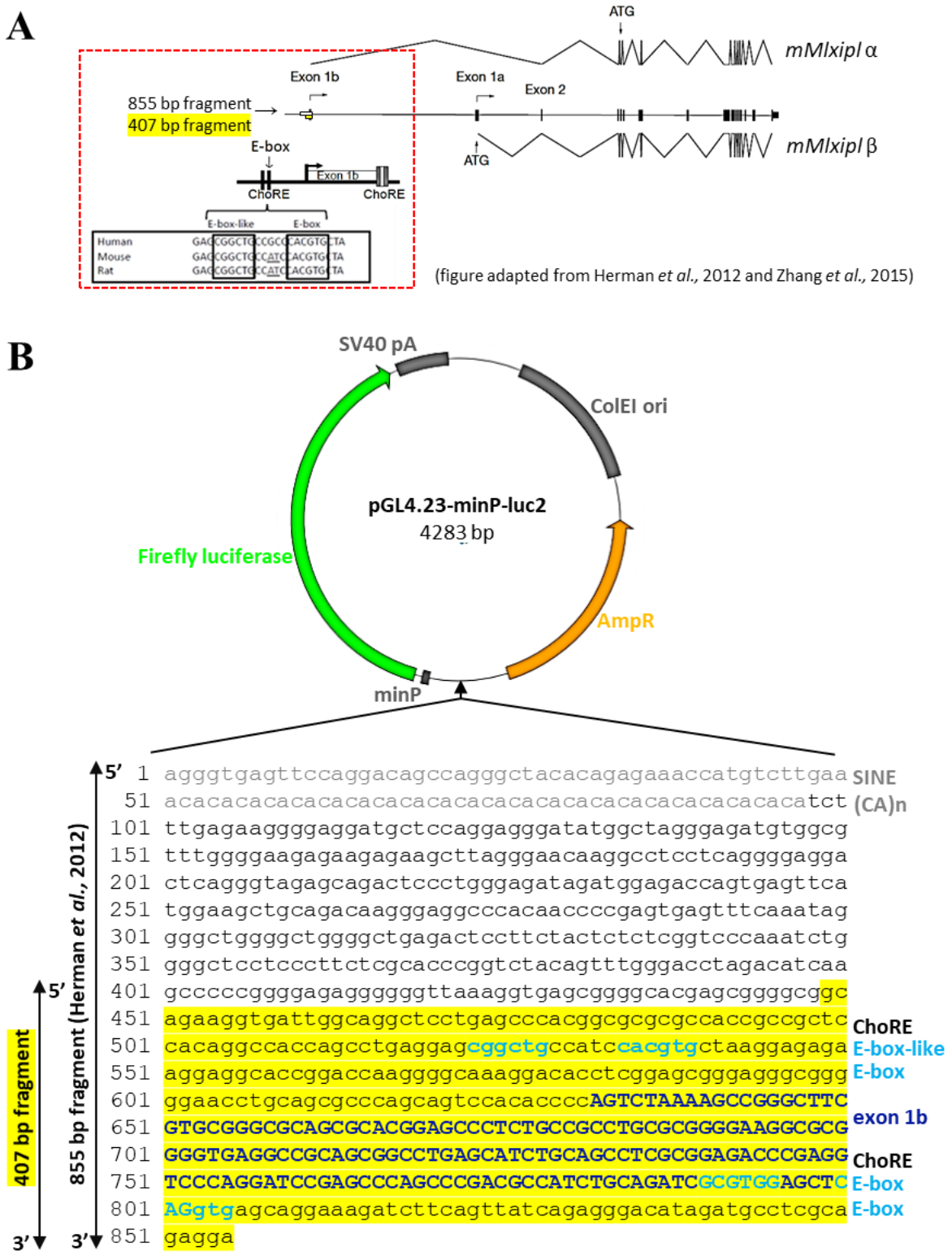


FIGURE C17: Firefly luciferase *mMlxipl* exon 1b reporter vector. (A) Model representing the structure and the splicing of *mMlxipl* α and β and the two carbohydrate response elements (ChoRE) upstream of and in *Mlxipl* exon 1b (figures modified after Herman *et al.,* 2012 and Zhang *et al.,* 2015). The first ChoRE is composed of an E-box-like motif and an E-box, the second consists of two E-boxes. **(B)** Sequence of the cloned *mMlxipl* exon 1b fragment cloned into a Firefly luciferase expression vector in Herman *et al.,* 2012 (855 bp, Chr5:135,089,259-135,090,113, mm10, cloned into pGL3) and in this thesis (407 bp, Chr5: 135,089,707-135,090,113, mm10, highlighted in yellow, cloned into pGL4.23). The E-boxes and the E-box-like motif are highlighted in light blue, the exon 1b in dark blue.

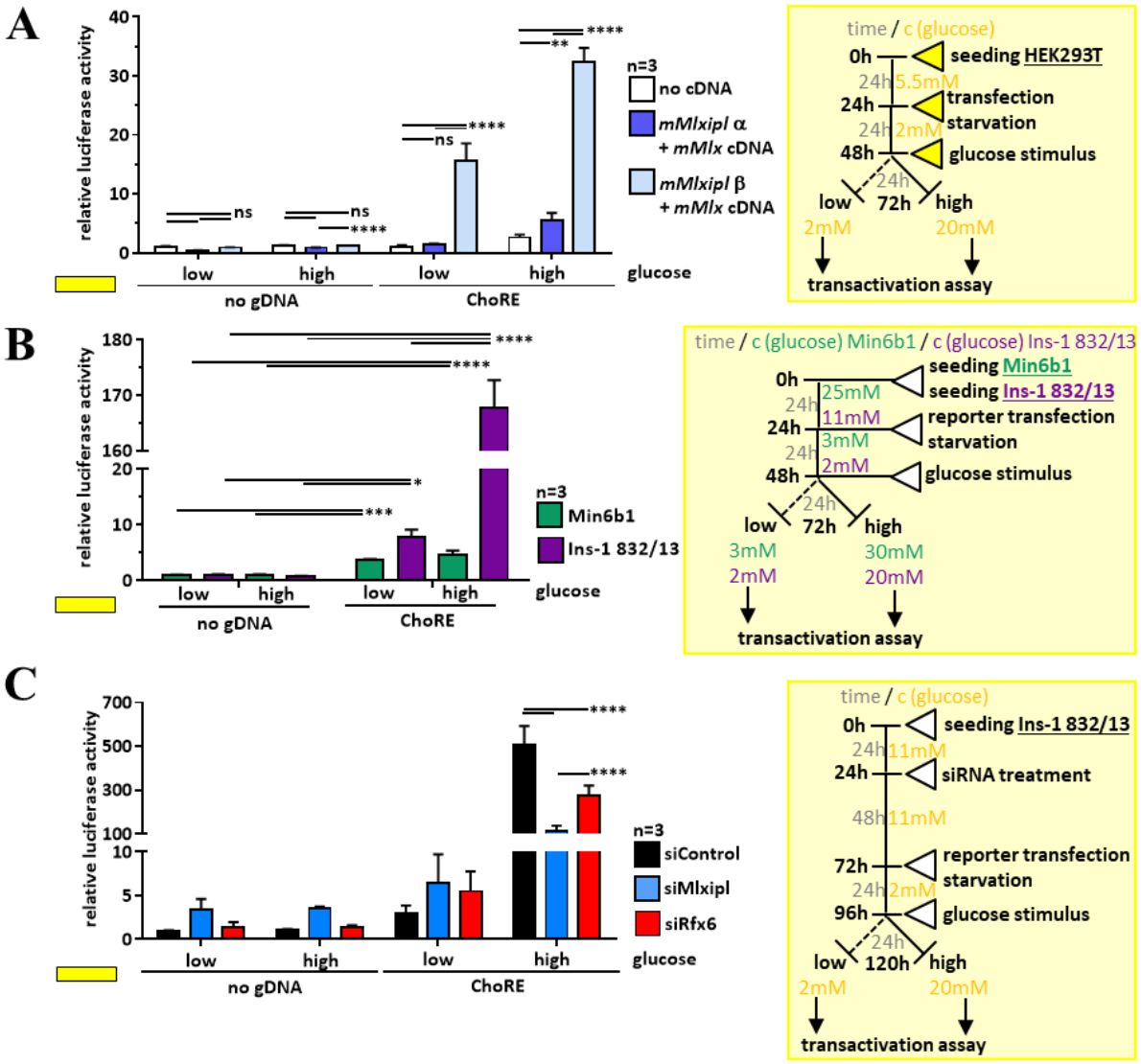


FIGURE C18: The *mMlxipl* exon 1b transactivation is glucose-independent in Min6b1 and glucose-dependent in Ins-1 832/13 cells. (A) Transactivation assay analysing the activation of the *mMlxipl* exon 1b (ChoRE) Firefly luciferase reporter construct at low or high glucose in HEK293T cells transfected with either a *mMlxipl* α or an *mMlxipl* β with a *mMlx* cDNA. Experimental strategy on the right: HEK293T cells were transfected with the expression vectors and the reporter constructs in DMEM with 2 mM glucose. After 24 h, medium was changed to DMEM containing 2 mM or 20 mM glucose and cells were cultured for 24 h before the test. **(B)** Transactivation assay analysing the activation of the *mMlxipl* exon 1b (ChoRE) Firefly luciferase reporter construct in Min6b1 cells at low or high glucose and in Ins-1 832/13 cells at low or high glucose. Experimental strategy on the right: Min6b1and Ins-1 832/13 cells were transfected with the reporter constructs in low glucose medium (standard cell culture medium with 3 mM glucose for Min6b1 cells, DMEM with 2 mM glucose for Ins-1 832/13 cells). After 24 h, medium was changed to low or high glucose medium (standard cell culture medium with 3 mM or 30 mM glucose for Min6b1 cells, DMEM with 2 mM or 20 mM glucose for Ins-1 832/13 cells) and cells were cultured for 24 h before the test. **(C)** Transactivation assay analysing the activation of the *mMlxipl* exon 1b (ChoRE) Firefly luciferase reporter construct in siControl-, siMlxipl- or siRfx6-treated Ins-1 832/13 cells at low or high glucose. Experimental strategy on the right: Ins-1 832/13 cells were transfected with SmartPool siRNAs in standard cell culture medium at 11 mM glucose. 48 h later, the cells were transfected with the reporter constructs in standard cell culture medium at 2 mM glucose. After 24 h, medium was changed to DMEM with 2 mM or 20 mM glucose and cells were cultured for 24 h before the test. Data in **(A, B, C)** are shown as mean plus standard deviation (n=3). Statistical significances (p≥0.05 not significant (ns), p<0.0.5 *, p<0.01 **, p<0.001 ***, p<0.0001 ****) were determined by two-way ANOVA.

transactivation assay would be the perfect test to investigate glucose response in Min6b1 and in Ins-1 832/13 cells on the transcriptional level (**FIGURE C17**; **FIGURE C18**). Furthermore, for our knockout experiments, the *mMlxipl* exon 1b transactivation assay would be the ideal functional test to distinguish *Mlxipl* -/- clones from wild-type clones.

The regulation of *Mlxipl* β transcription was first described in 2012 by Herman *et al.*: In the murine genome, they found a ChoRE in *mMlxipl* exon 1b and an additional E-box upstream of exon 1b (**FIGURE C17A**) suggesting that a ChoRE-binding protein regulates *Mlxipl* β transcription. To test the enhancer activity of these elements, they cloned an 855 bp long genomic region containing this ChoRE and the upstream E-box (Chr5:135,089,259-135,090,113, mm10; **FIGURE C17B**) into a firefly luciferase reporter vector. They could show that the construct was significantly activated after a glucose stimulus in HEK293T cells transfected with the construct and *mMlxipl* α / *mMlx* cDNAs (Herman *et al.,* 2012). In 2015, Zhang *et al.* reported that the upstream E-box forms a ChoRE with an E-box-like element and that the upstream ChoRE is conserved in human, mouse and rat (**FIGURE C17A**). They further demonstrated that the construct can strongly be activated in Ins-1 832/13 cells in 20 mM glucose medium by the endogenous transcription factors (Zhang *et al.,* 2015).

In this thesis, we cloned a shorter *mMlxipl* genomic fragment (407 bp) that includes the upstream and downstream *mMlxipl* ChoRE (Chr5: 135,089,707-135,090,113, mm10; **FIGURE C17B**), into the firefly reporter vector pGL4.23 because Herman *et al.* used a forward primer binding to a repetitive genomic sequence and they amplified the *mMlxipl* exon 1b genomic region from a bacterial artificial chromosome that contains the *mMlxipl* gene (Herman *et al.,* 2012). To be able to clone the genomic region from mouse genomic DNA, we decided to shorten the fragment and to use a forward primer downstream of the one used by Herman *et al.* (see **PART B3.2**, **FIGURE B2**).

We first verified that the shorter 407 bp long construct was activated in HEK293T cells transfected with *mMlxipl* α / *mMlx* cDNAs in high glucose medium, and, indeed, we observed a 5-fold induction (**FIGURE C18A**). The co-transfection of HEK293T cells with the reporter construct and *mMlxipl* β / *mMlx* cDNAs resulted in a 15-fold and 30-fold activation in low and high glucose medium, respectively (**FIGURE C18A**). These data proof that our *mMlxipl* exon 1b firefly luciferase reporter construct is functional.

We next tested its activation by the endogenous transcription factors of Min6b1 cells Ins-1 832/13 cells in high glucose (**FIGURE C18B**). In Min6b1 cells, we only observed a weak 3-fold induction that was completely glucose-independent (**FIGURE C18B**). By contrast, in Ins-1 832/13 cells, we got a weak 7-fold induction in low glucose and a strong 170-fold induction in high glucose (**FIGURE C18B**), thus the construct activation in this cell line was clearly glucose-dependent as expected. It seems that the murine Min6b1 cell line is not able to enhance the transactivation after a glucose stimulus. This could a cell-intrinsic problem or a problem related to the transfection. In fact, we already observed

that Min6b1 cells that had been transfected with a non-targeting siRNA, did not secrete insulin after a glucose stimulus (**not shown**). It could be that the cells are not sensitive to glucose directly after a transfection because it destabilizes the cell membrane (unpublished observations from our collaborator Paul Richards of the Cochin institute in Paris). If this is true, this would clearly limit our experimental possibilities because we would not be able to perform any rescue assays or transactivation assays in combination with a glucose stimulus in Min6b1 cells and clones.

In the Ins-1 832/13 cell line in which the transactivation was greatly enhanced by glucose, we further wanted to know whether this activation was down to rMLXIPL. To address this question, we treated the cells with siMlxipl and with siRfx6 of which we know that it reduces *Mlxipl* level, and we re-performed the *mMlxipl* exon 1b transactivation assay in the siRNA-treated cells (**FIGURE C18C**). We observed a 5-fold and a 2-fold reduction of the firefly luciferase activity in cells treated with siMlxipl and siRfx6, respectively (**FIGURE C18C**). Thus, targeting *rMlxipl* directly with an siRNA or indirectly by reducing its activator rRFX6 worked to affect *mMlxipl* exon 1b transactivation.

In conclusion, the results obtained in the Ins-1 832/13 cells come up to our expectations – the transactivation of the *mMlxipl* exon 1b reporter construct is extremely strong after a glucose stimulus and depends on rMLXIPL. Furthermore, the results demonstrate that we can use the *mMlxipl* exon 1b transactivation assay to identify *rMlxipl* -/- clones in Ins-1 832/13 cells, as reduction of *rMlxipl* by RNA interference also results in decreased luciferase activity. In contrast to the results for the Ins-1 832/13 cells, activation of the *mMlxipl* exon 1b firefly luciferase reporter construct in Min6b1 cells is very weak and, more importantly, not inducible by glucose. Together, these observations suggest that we should favour the Ins-1 832/13 cell line for this part of the project.

2.2.3. The isoform *mMlxipl* β is not expressed in Min6b1 cells

After we found that Min6b1 cells could not activate an artificial construct with the *mMlxipl* exon 1b genomic DNA, we wondered if this problem also involved the cell-intrinsic activation of *mMlxipl* β, or if it was indeed just a problem caused by the transfection. The expression of *Mlxipl* β is known to be induced in high glucose medium by MLXIPL α (Herman *et al.*, 2012; Zhang *et al.*, 2015). So far, *Mlxipl* β has been described in adipose and liver tissues of mice (Herman *et al.*, 2012) and in murine pancreatic islets (Jing *et al.*, 2016), in human adipose tissues (Kursawe *et al.*, 2013) and human pancreatic islets (Zhang *et al.*, 2015) as well as in rat pancreatic beta cells and in the Ins-1 832/13 cell line (Zhang *et al.*, 2015). The initial amount of *Mlxipl* β and its transcription rate are cell type-specific; the ratio of *Mlxipl* β to α before a glucose stimulus is 1 to 8 in Ins-1 832/13 cells, 1 to 1000 in rat islets and 1 to 30,000 in human islets (Zhang *et al.*, 2015). The glucose-induced increase of *Mlxipl* β transcript levels is detectable in Ins-1 832/13 cells after 2 h (Zhang *et al.*, 2015) and in human islets after 24 h (Jing *et al.*,

2016). The transcriptional activation of *Mlxipl* β leads to a reduction in the *Mlxipl* α transcript (Zhang *et al.,* 2015; Jing *et al.,* 2016) but the total *Mlxipl* amount does not change in some hours but only after 4 days (Zhang *et al.,* 2015). In human and rat islets, the expression of *Mlxipl* β does still not exceed the expression of *Mlxipl* α after 4 days in high glucose, whereas, in Ins-1 832/13 cells, the initial level of *Mlxipl* β is higher and it takes several hours to first equal and then exceed the expression level of *Mlxipl* α (Zhang *et al.,* 2015). By RT-qPCR, we measured the transcript levels of *Mlxipl*, *Mlxipl* α and *Mlxipl* β in both Min6b1 cells and Ins-1 832/13 cells after 12 h in low or high glucose medium (**FIGURE C19**; see **FIGURE C15A** and **FIGURE C16A** for information about the experimental strategy). Our results in Ins-1 832/13 cells were like those reported in the literature: after 12 h in high glucose, *rMlxipl* α was downregulated and *rMlxipl* β upregulated as expected (Herman *et al.,* 2012; Zhang *et al.,* 2015; **FIGURE C19D**). Astonishingly, in Min6b1 cells, the transcript level of *mMlxipl* α remained unchanged and *mMlxipl* β was not at all expressed after 12 h in high glucose (**FIGURE C19B**).

To rule out that the absence of any *mMlxipl* β transcript in Min6b1 cells was due to our qPCR primers (**FIGURE C19A**), we tested them on liver samples of wild-type and *mMlxipl* -/- mice (Iizuka *et al.,* 2004) that we got from our collaborators Catherine Postic and Sandra Guilmeau from the Cochin institute in Paris. In the wild-type liver cDNA sample, we could detect both isoforms, *mMlxipl* α and *mMlxipl* β, without problems. In the knockout, *mMlxipl* exons 12 to 14 were replaced by a neomycin cassette, leading to the loss of the basic helix-loop-helix leucine zipper domain encoded by exon 13 and 14 (Iizuka *et al.,* 2004). The expression of *mMlxipl* α in the *mMlxipl* -/- liver sample was reduced and the expression of *mMlxipl* β almost abolished (**FIGURE C19C**). Thus, we confirmed that our primers can detect *mMlxipl* β, and we concluded that there is no *mMlxipl* β transcript in Min6b1 cells.

After the analysis of the RNA level, we wondered whether we could detect MLXIPL β on a western blot to confirm the RT-qPCR results. In our lab, we use two different rabbit anti-MLXIPL antibodies (produced by Novus and Abcam, **TABLE B2**) that bind both to the C-terminus of MLXIPL and can thus detect both isoforms (**FIGURE C12B**; **FIGURE C20B**). MLXIPL α (95 kDa) migrates in our experiments at a higher molecular weight than expected (between 100 and 130 kDa, **FIGURE C8C**; **FIGURE C14B**), probably due to posttranslational modifications. The smaller isoform MLXIPL β has a molecular weight of 75 kDa (Herman *et al.,* 2012; **FIGURE C12B**; **FIGURE C20B**).

On the western blot performed with the anti-MLXIPL antibody of Novus, we detected a distinct band between 100 and 130 kDa in Min6b1 cells, in a protein extract from a liver sample of wild-type mice, Ins-1 832/13 cells and in HEK293T cells transfected with a *mMlxipl* α expression vector (**FIGURE C19E**). The band was lost in the sample from *mMlxipl* -/- mice and reduced in siMlxipl-treated Ins-1 832/13 cells (**FIGURE C19E**), confirming that this band is formed by MLXIPL and that it corresponded

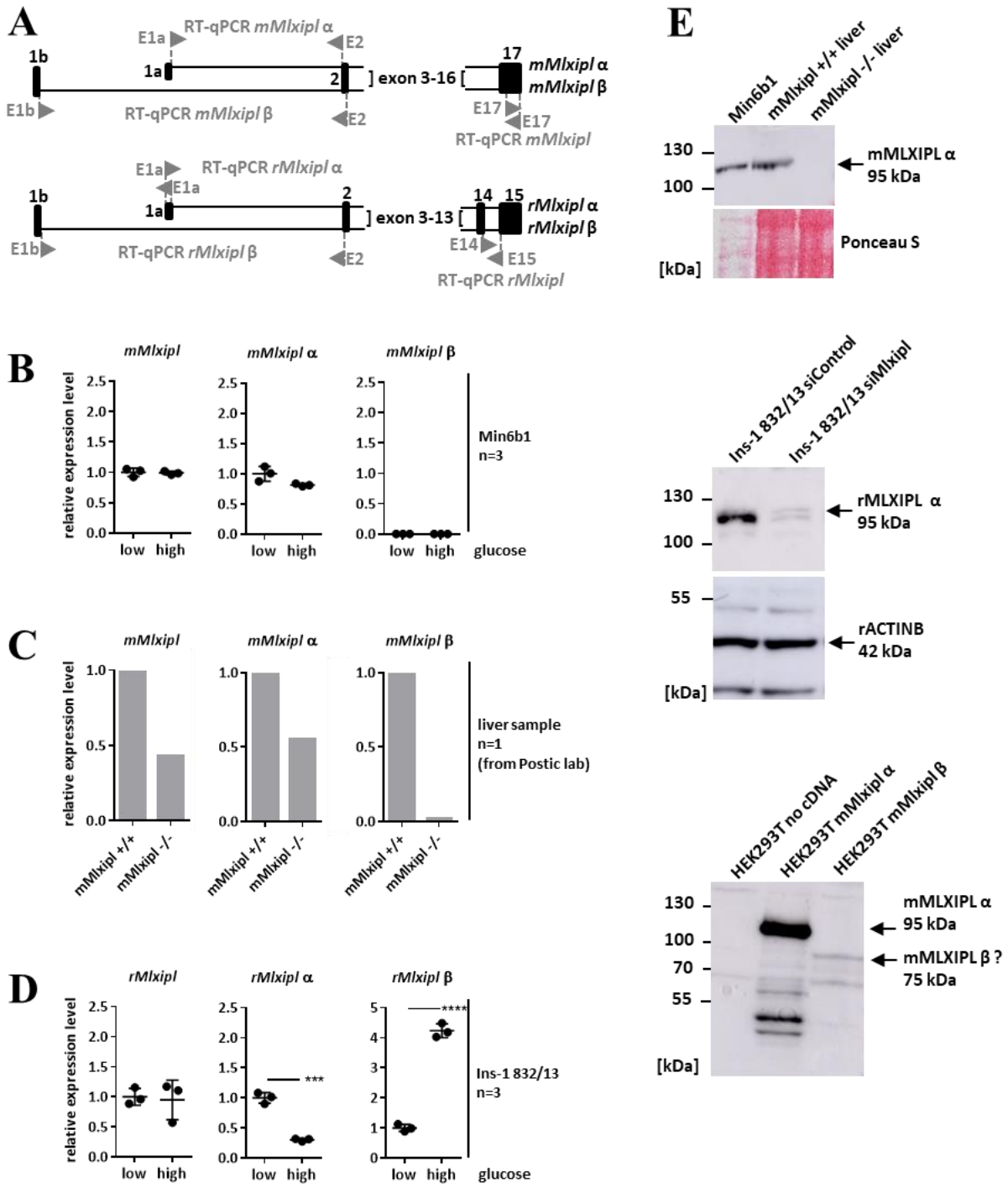


FIGURE C19: Min6b1 cells do not express *mMlxipl* β. (A) Position of the *Mlxipl*, *Mlxipl* α and *Mlxipl* β RT-qPCR primer in the murine (above) and rat (below) *Mlxipl* gene. **(B, C, D)** RT-qPCR analysis of *Mlxipl*, *Mlxipl* α and *Mlxipl* β in Min6b1 cells at 3 mM (low) and 30 mM (high) glucose **(B)**, in the liver of wild-type and *mMlxipl* -/- mice from Sandra Guilmeau of Catherine Postic's group **(C)** and in Ins-1 832/13 cells at 2 mM (low) and 20 mM (high) glucose **(D)**. Data in **(B)** and **(D)** are shown as mean plus standard deviation (n=3). Statistical significances ($p<0.001$ ***, $p<0.0001$ ****) were determined by unpaired two-tailed t test. See **FIGURE C15** and **FIGURE C16** for details about the experimental strategy. **(E)** Western blot determining MLXIPL expression (rabbit anti-MLXIPL, Novus) in the following samples: above: Min6b1 cells, liver extracts of *mMlxipl* -/- and wild-type mice (gift of Sandra Gulimeau); middle: siControl and siMlxipl-treated (SmartPool) Ins-1 832/13 cells; below: HEK293T transfected with an empty vector or an expression vector encoding either mMLXIPL α or β.

to MLXIPL α. In parallel, we tested the expression of mMLXIPL β in HEK293T cells transfected with a *mMlxipl* β expression vector. However, there were only two very weak bands of lower molecular weights; one of them could eventually correspond to mMLXIPL β (**FIGURE C19E**), but it is surprising that we did not detect a stronger band since we expressed *mMlxipl* β under the control of the strong CMV promoter like *mMlxipl* α (**FIGURE B3A+C**). The fact that we can not detect mMLXIPL β on a western blot is a handicap, since we cannot verify that we have both inactivated both isoforms in the generated *Mlxipl* knockout cell lines.

In summary, we revealed that *mMlxipl* β is not expressed by Min6b1 cells although it is expressed in murine pancreatic islets (Jing *et al.,* 2016). This is an important deficit of our Min6b1 cell line that makes it different from mature islet cells and not suitable for our study of mMLXIPL target genes in adult beta cells because we would lose the non-negligible contribution of mMLXIPL β to the transcriptional regulation of MLXIPL target genes after a glucose stimulus.

2.2.4. Why should we switch from Min6b1 cells to Ins-1 832/13 cells in our project?

To summarize, the most important arguments against the Min6b1 cell line we have in the lab, and the respective arguments for the use of the Ins-1 832/13 cell line are:

Cons Min6b1:

i. the transcriptional adaptation in Min6b1 cells after a glucose stimulus is weak and the cells are not responsive to changes in glucose concentrations after a transfection;
ii. it has been described that Min6b1 cells lose their beta cell characteristics including glucose-stimulated insulin secretion over four passages (Rani *et al.,* 2010; Cheng *et al.,* 2012), thus, we could not even finish the cloning process before the cells start to degenerate;
iii. the cells resemble to immature beta cells, highly express the disallowed gene *Ldha* and their morphology is highly heterogenous after cloning;
iv. they do not express *mMlxipl* β although mature islet cells do.

Pros Ins-1 832/13:

i. in Ins-1 832/13 cells show a strong transcriptional adaptation after a glucose stimulus;
ii. it has been reporter that Ins-1 832/13 cells maintain their glucose-stimulated insulin secretion feature over at least until passage 92 (Hohmeier *et al.*, 2000);
iii. Ins-1 832/13 cells do not express *Ldha*;
iv. they highly express *Mlxipl* β and activate its transcription after a glucose stimulus like rat islets.

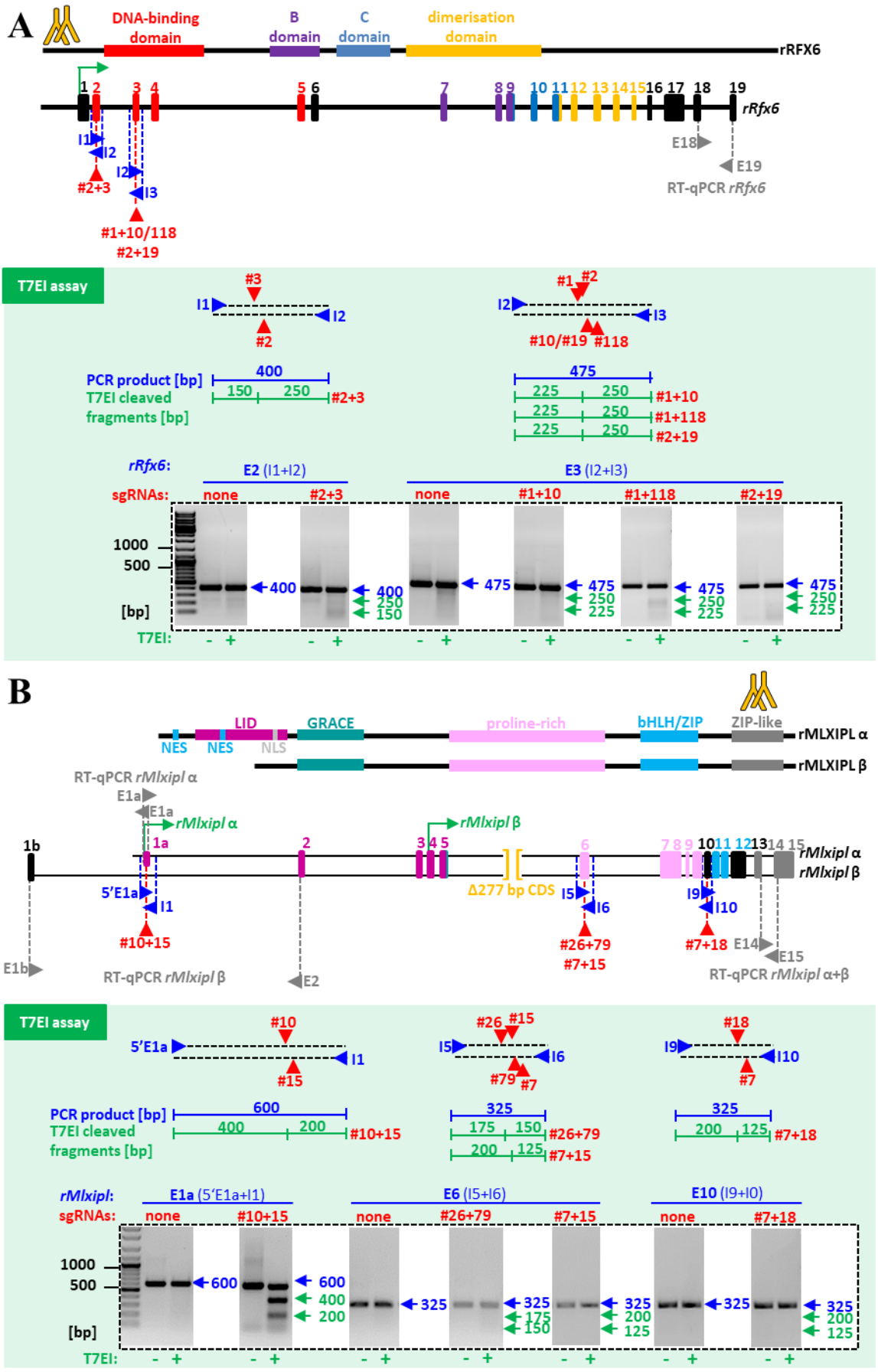
2.3. Generation of *Rfx6* and *Mlxipl* knockout clones in the rat Ins-1 832/13 beta cell line

After we had decided to restart the generation of *Rfx6* and *Mlxipl* knockout clones with the Ins-1 832/13 cell line, we had to redesign sgRNA pairs and to adapt our strategies to the rat cell line. Again, we chose to use the Cas9n in combination with two sgRNAs to limit the off-target effects (Jinek *et al.,* 2012; Ran *et al.,* 2013a; Ran *et al.,* 2013b).

2.3.1. T7EI assay to test the efficiency of *rRfx6* and *rMlxipl* sgRNA pairs in Ins-1 832/13 cells

To knockout *rRfx6* and *rMlxipl* in Ins-1 832/13 cells, we designed several sgRNA pairs targeting the rat genes this time and we tested the efficiency of the INDEL induction at the targeted loci by the above-described T7EI assay (see **PART B5.3** and **PART C2.1.1**; **FIGURE C20**). The sgRNA pairs against *rRfx6* targeted exon 2 (#2+3) and exon 3 (#1+10, #2+19, #1+118), both encoding the DNA-binding domain. The sgRNA pairs #2+19 was a sgRNA pair that we already used successfully in the Min6b1 cell line (**FIGURE C12A**); we could re-use it because the sgRNA sequence and the PAM are identical in both the mouse and the rat locus. Surprisingly, the T7EI assay revealed that only the *rRfx6* sgRNA pair #1+118 targeting exon 3 weakly induced INDEL mutations, while non the other three sgRNA pairs did not induce any detectable DNA alterations at the target site loci (**FIGURE C20A**). We did not detect any T7EI cleavage products for the other three sgRNA pairs against *rRfx6* (**FIGURE C20A**), even with the one that we had used in Min6b1 cells in which it efficiently worked.

For *rMlxipl*, we designed four sgRNA pairs (**FIGURE C20B**): #10+15 targets the ATG in *rMlxipl* exon 1a and is a sgRNA pair that we also used in the Min6b1 cell line (**FIGURE C12B**); the two sgRNA pairs #26+79 and #7+15 bind to *rMlxipl* exon 6 being the first exon downstream of the *rMlxipl* β ATG in exon 4 long enough to fit two sgRNAs into its sequence to target both *rMlxipl* isoforms α and β; last, #7+18 targets exon 10 and was chosen to mutate the basic helix-loop-helix domain encoded by exon 11 and exon 12. We assessed the cleavage efficiency of the sgRNA pairs by a T7EI test. The *rMlxipl* sgRNA pair #10+15 greatly induced INDEL mutations but none of the other three sgRNA pairs worked (**FIGURE C20B**). Although, according to the T7EI assay, the designed sgRNA pairs do not work well with one exception (*rMlxipl* sgRNA pair #10+15), due to time constraints, we nevertheless decided to try to use them for the generation of *rRfx6* and *rMlxipl* knockout cells in Ins-1 832/13 cells.


A
DNA-binding domain
B domain
C domain
dimerisation domain
rRFX6
rRfx6
E18
E19
RT-qPCR rRfx6
#2+3
#1+10/118
#2+19
T7EI assay
PCR product [bp]
T7EI cleaved fragments [bp]
#10/#19 #118
#1+10
#1+118
rRfx6:
E2 (I1+I2)
E3 (I2+I3)
sgRNAs:
none
[bp]
T7EI:
B
LID
GRACE
proline-rich
bHLH/ZIP
ZIP-like
NES
NLS
rMLXIPL α
rMLXIPL β
RT-qPCR rMlxipl α
rMlxipl α
rMlxipl β
Δ277 bp CDS
5'E1a
#10+15
#26+79
#7+15
#7+18
RT-qPCR rMlxipl β
RT-qPCR rMlxipl α+β
rMlxipl:
E1a (5'E1a+I1)
E6 (I5+I6)
E10 (I9+I0)

← **FIGURE C20: T7EI assay to determine the efficiency of the *rRfx6* and *rMlxipl* sgRNAs to induce INDEL mutations at their target sites in rat Ins-1 832/13 cells. (A, B above)** Structure of the *rRfx6* gene and rRFX6 protein **(A)** and *rMlxipl* gene and rMLXIPL α and β protein **(B)** pinpointing the position of the sgRNA pairs (red), PCR primer (blue), RT-qPCR primer (grey) and antibody binding sites (orange). The colour code of the exons indicates the encoded protein domain. **(A, B below)** Results of the T7EI assay assessing the CRISPR / Cas9-induced INDEL formation in the targeted *rRfx6* **(A)** and *rMlxipl* **(B)** loci. The cleavage efficiency in the T7EI test depends on the number of INDEL mutations in the amplified alleles produced by the Cas9n and the locus-specific sgRNAs. PCR products that differ in their sequences, form heteroduplexes containing mismatches. The endonuclease T7EI cleaves these heteroduplexes at the site where the mismatches are located, thereby producing smaller fragments (the expected sizes are indicated in the schemata).

TABLE C3: Summary of the first attempt to generate *rRfx6* and *rMlxipl* knockout cell lines in Ins-1 832/13 cells. Summary of the strategy used to generate knockout cell lines, the number of plated and grown clones, the obtained knockout and wild-type control cell lines, and the efficiency to generate knockout cell lines in proportion to the number of plated and analysed clones.

cell line	attempt	selection	gene	targeted exon	sgRNA pairs	plated clones	grown clones	knockout (-/-) cell lines	knockout (+/-) cell lines	control (+/+) cell lines	# of KO lines / plated clones	# of KO lines / analysed clones
Ins-1 832/13	1	puromycin				8	8			WT rM-1/2		
Ins-1 832/13	1	puromycin	*rRfx6*	3	#1+118	10	10				0/20	0/20
Ins-1 832/13	1	puromycin	*rRfx6*	3	#2+19	10	10					
Ins-1 832/13	1	puromycin	*rMlxipl*	1a	#10+15	10	10		KO rM-1/2		2/20	2/20
Ins-1 832/13	1	puromycin	*rMlxipl*	6	#26+79	10	10					

2.3.2. Transactivation assays, a tool to screen Ins-1 832/13 *rRfx6* and *rMlxipl* knockout cells?

To generate *rRfx6* and *rMlxipl* knockout cell lines in Ins-1 832/13 cells, we transfected wild-type Ins-1 832/13 cells with Cas9n expression vectors (pX462 with a puromycin selection gene) with sgRNAs. For *rRfx6*, for this first trial, we chose to test the sgRNA pairs #1+118 and #2+19, both targeting exon 3. For *rMlxipl*, we tried sgRNA pair #10+15 (exon 1a) and #26+79 (exon 6). As control, we transfected an additional vial of cells with pX462 without sgRNAs. In total, this corresponds to five transfections: no sgRNAs (control), *rRfx6* #1+118, *rRfx6* #2+19, *rMlxipl* #10+15, *rMlxipl* #26+79 (**TABLE C3**).

48 h after the transfection, we selected the transfected cells with 3 µg/ml of puromycin for 72 h. Then, we removed the dead cells and kept the living cells in culture for two weeks (see **PART B5.2**, **FIGURE B8**). We hand-picked 10 colonies per sgRNA pair (*rRfx6* #1+118, *rRfx6* #2+19, *rMlxipl* #10+15, *rMlxipl* #26+79) and 8 colonies for the control (48 colonies in total, **TABLE C3**). All cell clones survived the isolation and continued to grow.

We screened the clones using the functional *mMlxipl* intron 1 transactivation to screen the 20 putative *rRfx6* knockout clones and the *mMlxipl* exon 1b transactivation assay to screen the 20 putative *rMlxipl* knockout clones (**FIGURE C21**; **FIGURE C22**). In the experiments described above, we had demonstrated that a reduction in *rRfx6* in Ins-1 832/13 cells lowers the *mMlxipl* intron 1 transactivation (**FIGURE C8A**), and that a reduced level of *rMlxipl* causes a reduced activation of the *mMlxipl* exon

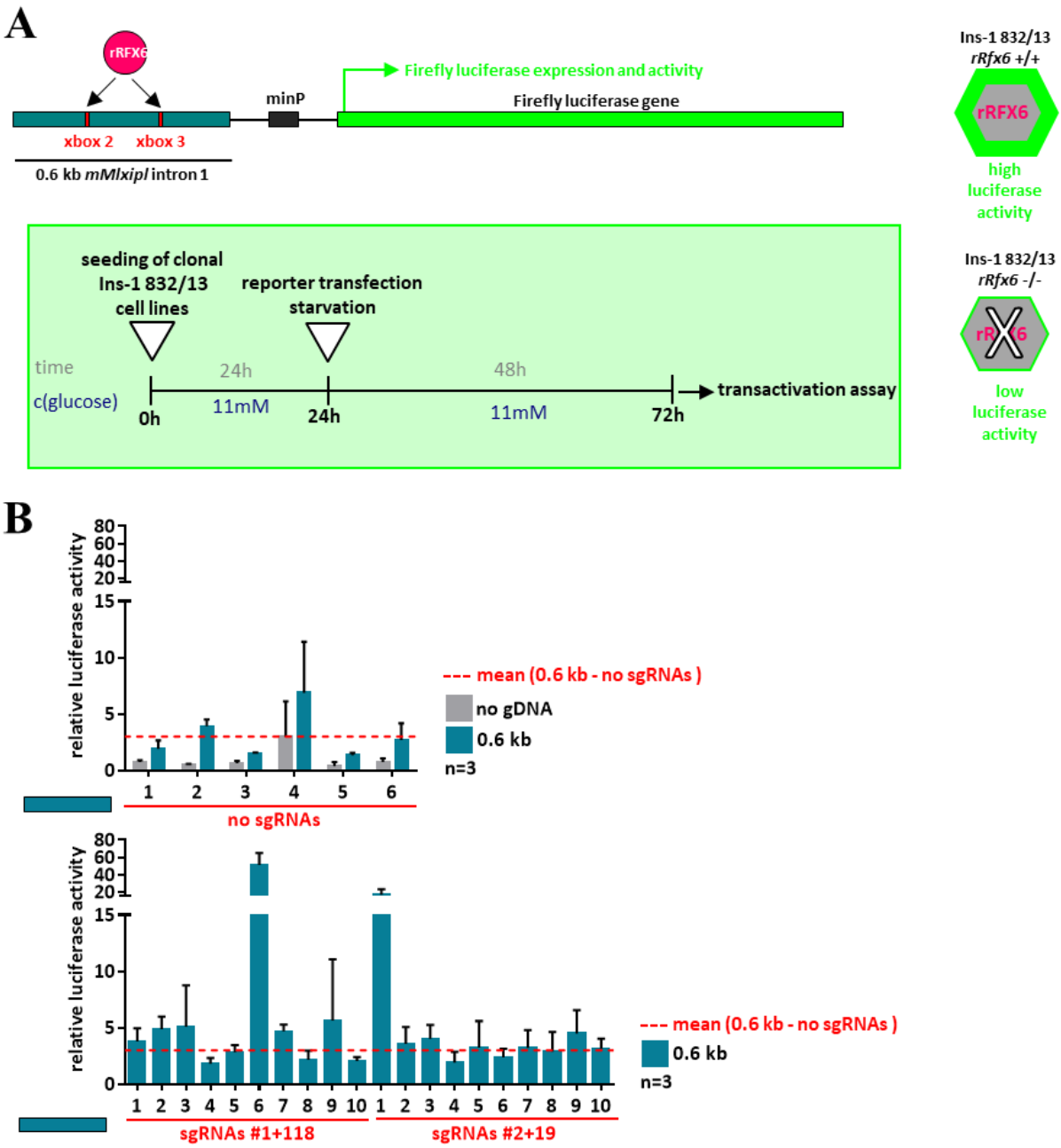


FIGURE C21: Screen for *rRfx6* knockout clones in the Ins-1 832/13 cell line. (A) Experimental strategy. The screen is based on the transactivation of the *mMlxipl* intron 1 0.6 kb Firefly luciferase reporter construct by rRFX6 in Ins-1 832/13 cells. Since the transactivation is dependent on rRFX6, the activation of the reporter construct should be higher in in *rRfx6* -/- Ins-1 832/13 cell lines than in wild-type Ins-1 832/13 cell lines. To perform the screen, the clonal Ins-1 832/13 cell lines (wild-type controls and putative knockout clones) were transfected with the reporter in standard cell culture medium (11 mM glucose) and the transactivation assay was done 48 h later. **(B)** Transactivation assay analysing the activation of the *mMlxipl* intron 1 0.6 kb Firefly luciferase reporter construct in the clonal Ins-1 832/13 wild-type control cell lines (above) and in the putative *rRfx6* knockout cell lines (below). Data are shown as mean plus standard deviation (n=3, technical triplicates). The red dashed line corresponds to the mean of the wild-type controls measured with the 0.6 kb construct.

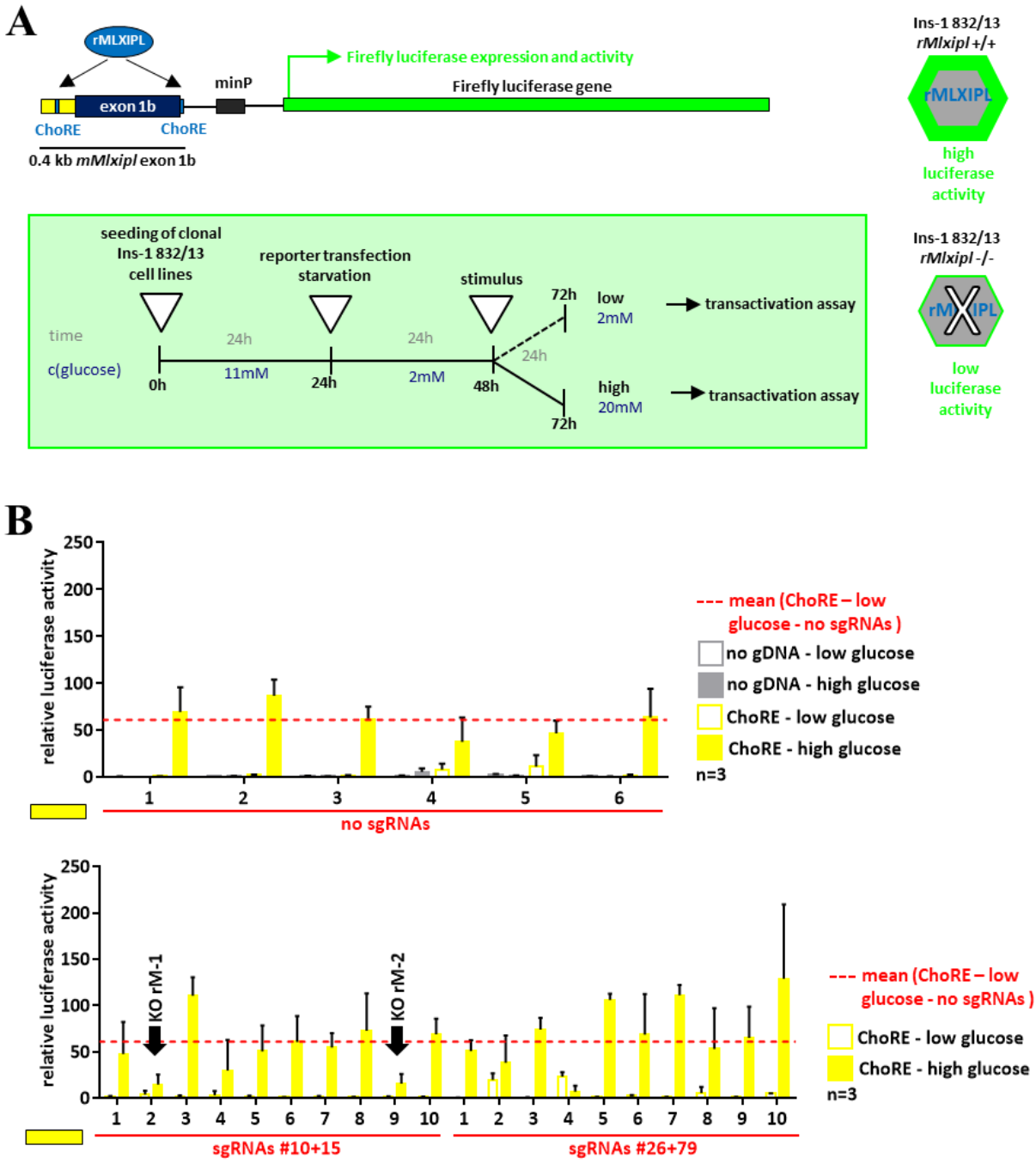


FIGURE C22: Screen for *rMlxipl* knockout clones in the Ins-1 832/13 cell line. (A) Experimental strategy. The screen is based on the transactivation of the *mMlxipl* exon 1b (ChoRE) Firefly luciferase reporter construct by rMLXIPL in high glucose in Ins-1 832/13 cells. Since the transactivation is dependent on rMLXIPL, the activation of the reporter construct should be lower in in *rMlxipl* -/- Ins-1 832/13 cell lines than in wild-type Ins-1 832/13 cell lines. To perform the screen, the clonal Ins-1 832/13 cell lines (wild-type controls and putative knockout clones) were transfected with the reporter construct, starved in standard cell culture medium with 2 mM glucose for 24 h and incubated in standard cell culture medium containing 2 (low) or 20 mM (high) glucose for another 24 h prior to analysis. **(B)** Transactivation assay analysing the activation of the *mMlxipl* exon 1b (ChoRE) Firefly luciferase reporter construct in the clonal Ins-1 832/13 wild-type control cell lines (above) and in the putative *rMlxipl* knockout cell lines (below). Data are shown as mean plus standard deviation (n=3, technical triplicates). The red dashed line corresponds to the mean of the wild-type controls measured with the ChoRE construct at low glucose. The second and the ninth clones of the knockout attempt with the *rMlxipl* sgRNAs #10+15 were selected for further analysis and are named KO rM-1 and KO rM-2.

1b luciferase reporter construct (**FIGURE C18C**). Thus, we thought these transactivation assays would be the perfect indicators to search for Ins-1 832/13 clones with a reduced rRFX6 and rMLXIPL expression.

For the *rRfx6* knockout screen, we transfected the 20 putative *rRfx6* knockout clones and six control clones with the *mMlxipl* intron 1 0.6 kb firefly luciferase construct and measured the firefly luciferase activity 48 h later. We assumed that, compared to wild-type control cells and wild-type clones, we should observe a significantly reduced firefly luminescence in clones in which we successfully inactivated one or both *rRfx6* alleles (**FIGURE C21A**). In all six wild-type control clones, the *mMlxipl* intron 1 0.6 kb firefly luciferase construct was activated, but the amount of activation was variable with a mean activation of 3. The activation of the construct in the putative *rRfx6* knockout clones was comparable with that in the control cells with two exceptions, *rRfx6* sgRNA pair #1+118 number 6 and *rRfx6* sgRNA pair #2+19 number 1, in which the activation was much higher (**FIGURE C21B**). Overall, we did not find any Ins-1 832/13 *rRfx6* knockout cell line.

For the second screen to identify *rMlxipl* knockout cells in Ins-1 832/13 cells, we used the *mMlxipl* exon 1b firefly luciferase construct whose activation is glucose-dependent in Ins-1 832/13 cells (**FIGURE C18C**). We transfected 20 putative *rMlxipl* knockout clones and six control clones with the construct, starved the cells for 24 h in low glucose medium and incubated them either in low or in high glucose medium for 24 h prior to the luminescence measurement. In the case that we inactivated one or both *rMlxipl* alleles in a clone, we expected to see a significantly reduced luciferase activity in high glucose after the transfection with the reporter construct (**FIGURE C22A**). In the wild-type control clones, we got a stable 60-fold induction of the firefly luciferase in high glucose medium. Among the putative *rMlxipl* knockout clones, we got three clones (*rMlxipl* sgRNA pair #10+15 number 2 and 9, *rMlxipl* sgRNA pair # 26+79 number 4) with a strongly reduced reporter activation in high glucose, but the one generated with the sgRNA pair # 26+79 showed an increased reporter induction at low glucose, thus we decided to focus on the other two clones. They will be named KO rM-1 (*rMlxipl* sgRNA pair #10+15 number 2) and KO rM-2 (*rMlxipl* sgRNA pair #10+15 number 9) hereafter.

Taken together, the *mMlxipl* intron 1 and *mMlxipl* exon 1b transactivation assays allowed us to perform a fast screen for *rRfx6* and *rMlxipl* knockout clones in Ins-1 832/13 cells, respectively, and to directly get a functional read-out concerning the rRFX6 and rMLXIPL functionality in the putative knockout clones.

2.3.3. Functional analysis of the Ins-1 832/13 *rMlxipl* knockout clones KO rM-1 and KO rM-2

In the attempt to generate *rMlxipl* knockout clones in Ins-1 832/13 cells, we produced the two *rMlxipl* clones KO rM-1 and KO rM-2 with a reduced *mMlxipl* exon 1b transactivation in high glucose

(**FIGURE C22B**). We confirmed the results of the *mMlxipl* exon 1b transactivation assay screen (**FIGURE C23**) by repeating the transactivation assay with the two clones KO rM-1 and KO rM-2 and two wild-type control clones WT rM-1 and WT rM-2 chosen from the eight clones transfected with the pX462 plasmid without sgRNAs (**TABLE C3**). In this second assay, we got the same results as in the original screen: The luciferase activity was significantly reduced in high glucose in KO rM-1 and KO rM-2 cells transfected with the *mMlxipl* exon 1b firefly luciferase reporter construct compared to the control cells WT rM-1 and WT rM-2 transfected with the same construct (**FIGURE C23A**).

To determine whether the reduced reporter construct activation is due to the inactivation of one or both *rMlxipl* alleles in the two clones KO rM-1 and KO rM-2, we sequenced the targeted loci (exon 1a). Intriguingly, we encountered the problem that, in both clones, we detected four sequences in the sequencing reactions instead of two sequences as one would have expected since there should be two *rMlxipl* alleles. To find out how many and which alleles were present in the sequencing profiles, we cloned the PCR products into the TOPO TA vector pCRII and tested 24 colonies per clone by sequencing. The results for clone KO rM-2 are shown in **FIGURE C24**. We found nine different alleles in clone KO rM-1 and seven in clone KO rM-2 (**FIGURE C24A+B+C**). In both clones, we found the wild-type allele, and six and two alleles with in-frame deletions and two and four alleles with out-of-frame mutations in clone KO rM-1 and KO rM-2, respectively.

We next checked the rMLXIPL expression by western blotting (rabbit anti-MLXIPL, Novus), but it was unchanged in the clones KO rM-1 and KO rM-2 compared to the wild-type WT rM-1 and WT rM-2 (**FIGURE C23B**). To test whether the expression of rMLXIPL target genes might be affected in two heterogenous clones KO rM-1 and KO rM-2, we starved the clones and the two control lines WT rM-1 and WT rM-2 overnight in low glucose medium, incubated them for 6 h in low or high glucose and analysed the expression of *rMlxipl* β, *rCx3cl1* and *rRasgrp2* by RT-qPCR (**FIGURE C23C+D+E**). We found a high variability in the glucose-dependent upregulation of all three genes in the wild-type clones. The huge difference between both controls made it impossible to draw any conclusions about the impact of the *rMlxipl* mutations on the expression of rMLXIPL targets (**FIGURE C23A**).

Since the amount of rMLXIPL seems to be unaffected in the *rMlxipl* clones KO rM-1 and KO rM-2, the reduced transactivation of the *mMlxipl* exon 1b reporter in the knockout clones cannot be caused by a reduced expression of rMLXIPL in the clones but could rather be due to the loss of some amino acid residues in exon 1a in the alleles with the in-frame mutations. These mostly concern the Nuclear Export Signal (NES) 1 upstream of the LID and the Mondo-Conserved Region (MCR) II (**FIGURE C24C**).

It has been shown that the loss of the NES2 in MCRII leads to the inactivation of MLXIPL. Although its deletion caused the enrichment of MLXIPL in the nucleus, the nuclear MLXIPL was transcriptionally inactive suggesting that the nuclear translocation is not enough to activate MLXIPL but that further

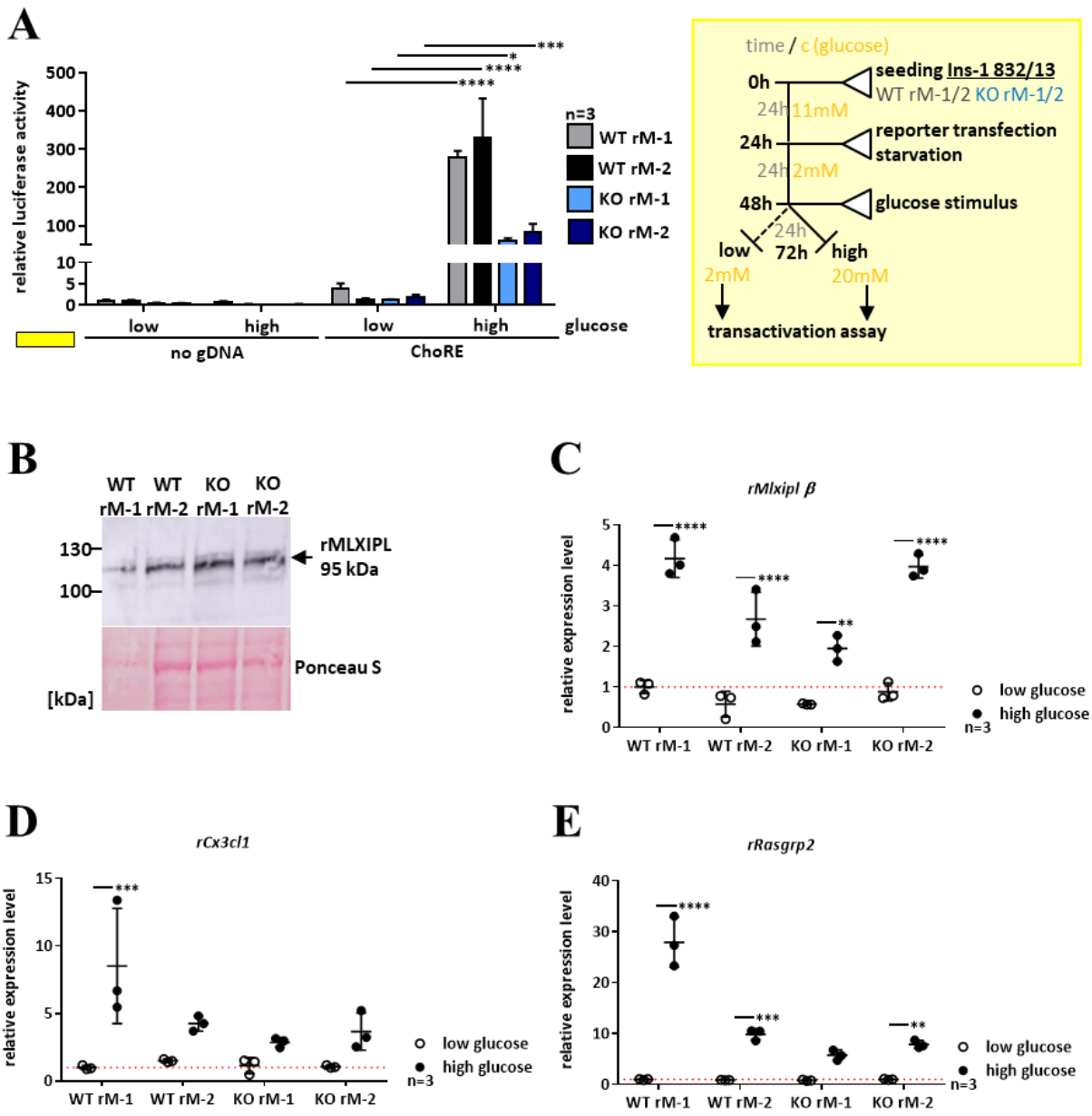


FIGURE C23: The mutations in the Ins-1 832/13 *rMlxipl* knockout clones KO rM-1 and -2 impair *mMlxipl* exon 1b transactivation but have no effect on target gene expression. (A) Transactivation assay analysing the activation of the *mMlxipl* exon 1b (ChoRE) Firefly luciferase reporter construct in the Ins-1 832/13 wild-type clones WT rM-1 and -2 and the *rMlxipl* knockout clones KO rM-1 and -2. Experimental strategy on the left: The cell lines were transfected with the reporter construct, starved in standard cell culture medium with 2 mM glucose for 24 h and incubated in standard cell culture medium containing 2 (low) or 20 mM (high) glucose for another 24 h prior to analysis. **(B)** Western blot analysing rMLXIPL expression in the Ins-1 832/13 wild-type clones WT rM-1 and -2 and the *rMlxipl* knockout clones KO rM-1 and -2 (rabbit anti-MLXIPL, Novus). **(C-E)** RT-qPCR analysis of the expression of the rMLXIPL target genes *rMlxipl* β **(C)**, *rCx3cl1* **(D)**, and *rRasgrp2* **(E)** in the Ins-1 832/13 wild-type clones WT rM-1 and -2 and the *rMlxipl* knockout clones KO rM-1 and -2. Prior to analysis, the cell lines were starved overnight in standard cell culture medium at 2 mM glucose and incubated for 6 h in standard cell culture medium supplemented with 2 mM (low) or 20 mM (high) glucose. Data are shown as mean plus standard deviation (n=3, technical triplicates). Statistical significances ($p<0.05$ *, $p<0.01$ **, $p<0.001$ ***, $p<0.0001$ ****) were determined by two-way ANOVA.

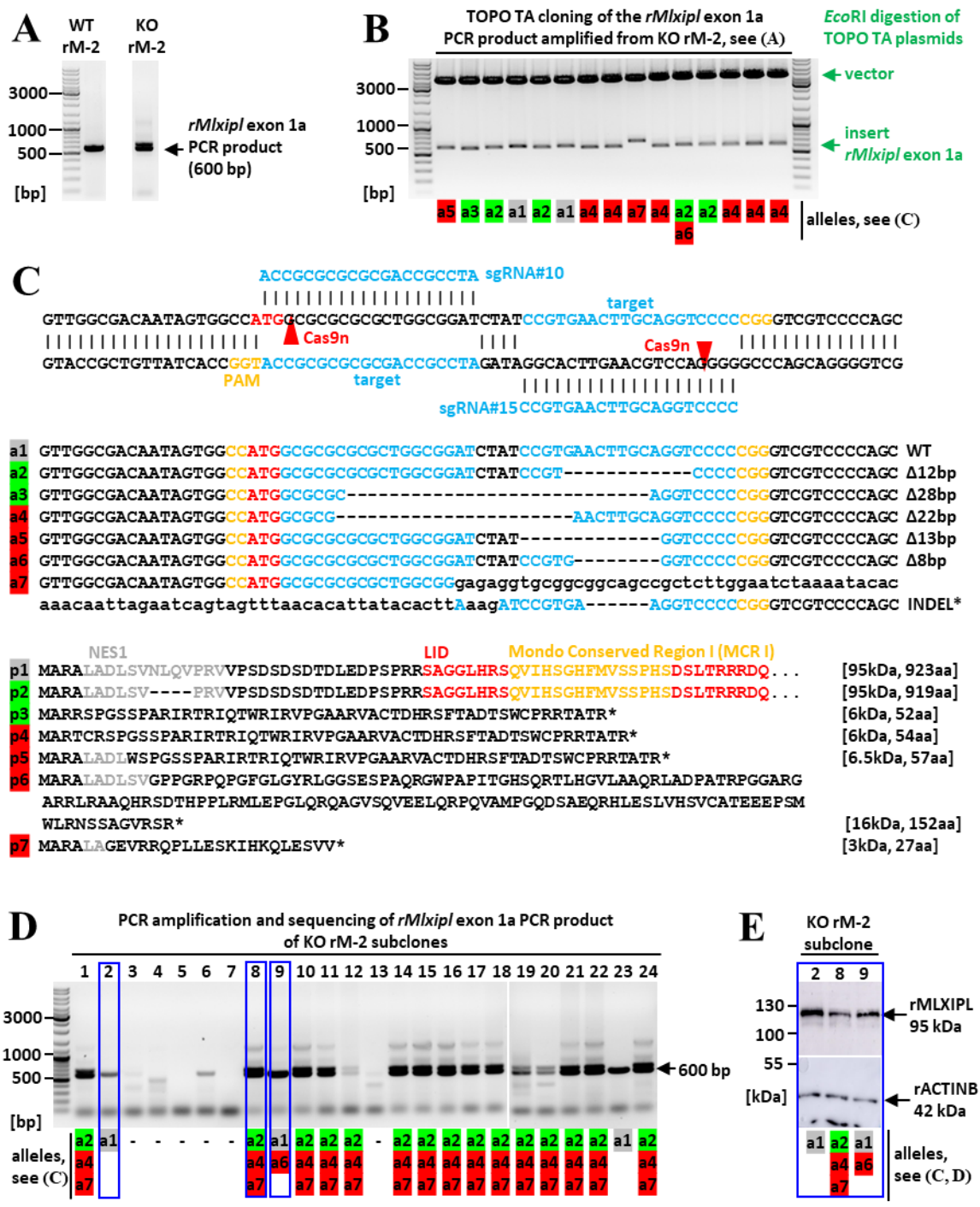


FIGURE C24: Abnormal allele number in Ins-1 832/13 *rMlxipl* knockout clone KO rM-2 before and after subcloning. (A) PCR amplification of *rMlxipl* exon 1a (target site of sgRNA#10+15 and Cas9n) in the Ins-1 832/13 wild-type clone WT rM-2 and the *rMlxipl* knockout clone KO rM-2. **(B)** Restriction analysis (*Eco*RI digestion) of the TOPO TA-cloned *rMlxipl* exon 1a PCR products amplified from KO rM-2 genomic DNA. **(C)** Sequencing of the KO rM-2 *rMlxipl* exon 1a TOPO TA plasmids (INDEL* = insertion of 79 bp, insertion of 3 bp, deletion of 6 bp). Wild-type alleles are marked in grey, alleles with in-frame mutations in green, alleles with out-of-frame deletions in red. **(D)** PCR amplification of *rMlxipl* exon 1a in subclones of the Ins-1 832/13 *rMlxipl* knockout clone KO rM-2. **(E)** Western blot detecting rMLXIPL (rabbit anti-MLXIPL, Novus) in the KO rM-2 subclones 2, 8 and 9 (framed in blue in **(D)**) representing all identified genotypes in the KO rM-2 subclones in **(D)**.

stimuli are necessary to activate the protein (Li *et al.,* 2008; Davies *et al.,* 2010). In fact, MLXIPL α activity and subcellular localization is regulated by the combination of posttranslational modifications (phosphorylation, O-linked N-acetylglucosamine modification, acetylation) and the action of the nuclear import factor IMPORTIN α, the nuclear export factor CMR1 and the cytoplasmic retention protein 14-3-3 (Havula and Hietakangas, 2008, review). Taken together, it is not unlikely that the in-frame deletions of amino acid residues in the NES1 affect rMLXIPL transcriptional activity.

2.3.4. Subcloning of the heterogenous Ins-1 832/13 *rMlxipl* cell line KO rM-2

As mentioned above, the *rMlxipl* knockout clones KO rM-1 and KO rM-2 contain multiple alleles. Via TOPO TA cloning, we determined the number of alleles per clone (nine for KO rM-1, seven for KO rM-2, **FIGURE C24A+B+C**). We thought that they might be a mixture of different clones of different genotypes, either wild type or mutated, forming what we call a heterogenous clone. In the cloning process to generate knockout clones (see **PART B5.2**, **FIGURE B8**), we used the limiting dilutions technique to produce clonal cell lines, but it is not excluded that several single cells attached close to each other to the cell culture surface and formed one colony. Therefore, we subcloned the heterogenous *rMlxipl* clone KO rM-2, amplified the CRISPR/Cas9n target site in exon 1a by PCR and sequenced it, expecting to see two sequences in the sequencing profile. Unexpectedly, the majority of the subclones (15 out of 24) still contained three different alleles. We compared the new sequencing profiles with the alleles identified in the TOPO TA cloning and we found that the 15 subclones, in which we found the three sequences, contained one allele with an in-frame mutation and two alleles with an out-of-frame mutation. Two further subclones were wild-type cells and one clone contained a wild-type allele and one mutant allele with an out-of-frame mutation (**FIGURE C24D**). We analysed the protein expression in three representative subclones (a wild-type clone, the clone with the wild-type and the nonsense allele, and one of the 15 subclones with the three alleles) on a western blot and we saw a decreasing level of rMLXIPL expression from wild type cells across the subclone with two alleles to the subclone with the three alleles (**FIGURE C24E**).

On one hand, one could imagine that we were not able to produce pure clones; this would mean that the clones with the three mutations were formed by two clones of different genotypes, one clone containing two heterozygous mutations and the other a homozygous mutation. If each of the fifteen subclones continued to be formed from two cells of the two genotypes, this is clearly a point against the further use of the limiting dilution technique. On the other hand, there could be a problem on the genomic level as the cells we work with, are derived from cancer cells. The Ins-1 832/3 cell line that was generated together with the Ins-1 832/13 cell line (Hohmeier *et al.*, 2000), has an abnormal karyotype (Naylor *et al.,* 2016), suggesting that this might also be the case for Ins-1 832/13 cells.

Thinking further, this could mean that the wild-type cells could contain three wild-type copies of *rMlxipl* and the clone with the two alleles two wild-type and one mutant allele. The latter would hardly be visible in the sequencing profiles. Most of the original clone KO rM-2 would have been formed by the cells with the three different alleles poorly contaminated with wild-type cells and cells of other genotypes. To find out if there really is a problem from a genetic point of view, we should karyotype the Ins-1 832/13 cells to see how far their karyogram deviates from a healthy rat karyogram, and to learn more about the chromosomes encoding our genes of interest *rRfx6* and *rMlxipl*. Second, we should determine the exact allele number by quantitative PCR or by southern blot.

2.3.5. FACS, a possibility to generate homogenous Ins-1 832/13 knockout cell lines?

As mentioned, an important point at which we can improve the strategy of generating Ins-1 832/13 knockout clones, is the step at which we need to get single cells that form colonies. Instead of trying it further with the limiting dilutions technique, we opted for FACS and to sort one cell per well in 96 well plates.

First, we tried to knockout *rMlxipl* with this new strategy. We transfected wild-type Ins-1 832/13 cells with the Cas9n expression vector pX461 encoding Cas9n-2A-EGFP. We used the sgRNA pairs #10+15 (exon 1a), #7+15 (exon 6) and #7+18 (exon 10). 48 h after the transfection, we used FACS to plate one cell per well in 96 well plates (see **PART B5.2**, **FIGURE B9**). We prepared two 96 well plates for the sgRNA pair #10+15, one for #7+15, one for #7+18 and one for the control that was transfected with the Cas9n expression vector without sgRNAs (five plates in total, **TABLE C4**). We got 36 control clones and 82, 19 and 21 putative *rMlxipl* knockout clones for the sgRNA pairs #10+15, #7+15 and #7+18, respectively (**TABLE C4**). To screen the clones, we amplified the targeted exonic sequences by PCR, analysed them by agarose gel electrophoresis and sequencing. For the sgRNA pairs #7+15 and #7+18, we only got wild-type clones. For the sgRNA pair #10+15, we sequenced the PCR products amplified from 22 clones. In 20 of them, we detected three sequences. We performed a western blot on these clones and we did not observe the loss of rMLXIPL protein (rabbit anti-MLXIPL of Novus) in any of them, thus we did not produce homozygous *rMlxipl* knockout clones and FACS could not solve the problem of the additional allele, suggesting that it might indeed be a genetic problem.

Using the new cloning strategy, we tried for a second time to knockout *rRfx6* in Ins-1 832/13 cells. We targeted *rRfx6* exon 2 (#2+3) and exon 3 (#1+10), we sorted the transfected cells into one 96 well plate per sgRNA pair and, after two weeks, we obtained 48 putative *rRfx6* knockout clones for sgRNA pair #2+3 and 34 for sgRNA pair #1+10 (**TABLE C4**). Since, in the previous trial to knockout *rRfx6* in Ins-1 832/13 cells, we did not sequence the *rRfx6* locus, we decided to start the screen for *rRfx6*

TABLE C4: Summary of the second attempt to generate *rRfx6* and *rMlxipl* knockout cell lines in Ins-1 832/13 cells. Summary of the strategy used to generate knockout cell lines, the number of plated and grown clones, the obtained knockout and wild-type control cell lines, and the efficiency to generate knockout cell lines in proportion to the number of plated and analysed clones.

cell line	attempt	selection	gene	targeted exon	sgRNA pairs	plated clones	grown clones	knockout (-/-) cell lines	knockout (+/-) cell lines	control (+/+) cell lines	# of KO lines / plated clones	# of KO lines / analysed clones
Ins-1 832/13	2	EGFP				96	36					
Ins-1 832/13	2-1	EGFP	*rRfx6*	2	#2+3	96	48		KO rR-1/2	WT rR-1	2/192	2/82
Ins-1 832/13	2-1	EGFP	*rRfx6*	3	#1+10	96	34					
Ins-1 832/13 KO rR-1	2-2	EGFP	*rRfx6*	2	#2+3	198	22		KO rR-3	WT rR-2	2/588	2/133
Ins-1 832/13 KO rR-1	2-2	EGFP	*rRfx6*	3	#1+118	96	16					
Ins-1 832/13 KO rR-2	2-2	EGFP	*rRfx6*	2	#2+3	198	69					
Ins-1 832/13 KO rR-2	2-2	EGFP	*rRfx6*	3	#1+118	96	26		KO rR-4	WT rR-3		
Ins-1 832/13 KO rR-1	2-3	EGFP	*rRfx6*	2	#2+3	96	13				0/384	0/44
Ins-1 832/13 KO rR-2	2-3	EGFP	*rRfx6*	2	#2+3	96	8					
Ins-1 832/13 KO rR-3	2-3	EGFP	*rRfx6*	2	#2+3	96	16					
Ins-1 832/13 KO rR-4	2-3	EGFP	*rRfx6*	3	#1+118	96	7					
Ins-1 832/13	2	EGFP	*rMlxipl*	1a	#10+15	198	82				0/390	0/122
Ins-1 832/13	2	EGFP	*rMlxipl*	6	#7+15	96	19					
Ins-1 832/13	2	EGFP	*rMlxipl*	10	#7+18	96	21					

knockout clones with a PCR and sequencing to exclude the possibility that there might also be a problem at the genetic level for this gene. We identified two heterozygous clones with a large deletion in the targeted exon 2 (Δ200 bp and Δ177 bp) that included the 3' exon-intron junction and splicing donor site of exon 2 (**FIGURE C25A+B**), while the second allele remained a wild-type allele. Thus, we cannot be sure that there are exactly two alleles, since one of the alleles in both clones is a wild-type allele and could be present in more than one copy.

We decided to focus on these two clones that we named KO rR-1 and KO rR-2. We thought it might be a good strategy to try to knockout the second allele (if there are only two) in these clones. The deletions in the clones KO rR-1 and KO rR-2 have eliminated the binding sites for the sgRNAs, thus we uniquely target the intact wild-type allele(s) in the heterozygous *rRfx6* clones. We re-transfected KO rR-1 and KO rR-2 with Cas9n and sgRNA pairs targeting either exon 2 (#2+3) or exon 3 (#1+10) and we sorted 196 and 98 cells per clone for the sgRNA pairs #2+3 and #1+10, respectively. In total, 133 cells grew. Again, we analysed the target loci by PCR and sequencing. Interestingly, we not only got clones that resembled to the initial heterozygous clones KO rR-1 and KO rR-2 with the Δ200 bp and Δ177 bp deletion in exon 2 but also wild-type clones and two new heterozygous knockout clones with a Δ119 bp in exon 2 and a Δ58 bp deletion in exon 3 (**FIGURE C25A+B**), that did not carry the two original mutations, and that we named KO rR-3 and KO rR-4. We concluded that we must have contaminated our initial heterozygous clones KO rR-1 and KO rR-2 with wild-type cells during the clonal expansion.

We performed again a western blot to analyse rRFX6 protein (rabbit anit-RFX6 #2766) and the protein level was halved in all four heterozygous clones (**FIGURE C25D**). This observation strongly suggests that there are only two *rRfx6* alleles in Ins-1 832/13 cells because the loss of one of them reduces the rRFX6 expression by 50 %.

We re-tried for a second time to knockout the second allele by re-transfecting the clones KO rR-1, KO rR-2 and KO rR-3 with the sgRNA pair #2+3 targeting exon 2 and KO rR-4 with the sgRNA pair #1+118 targeting exon 3. For the first time, we observed a huge cell death after the transfection. We sorted the surviving cells into 96 well plates (one plate per clone, four plates in total) and, after two weeks, we obtained 13, 8, 16 and 7 clones for the transfected and sorted KO rR-1, KO rR-2, KO rR-3 and KO rR-4, respectively. By western blotting, we searched for clones that do not express rRFX6, but we could still detect rRFX6 in all clones (rabbit anit-RFX6 #2766). After this, we stopped to try to inactivate the second allele and decided to focus on the *rRfx6* heterozygous clones.

2.3.6. The expression of *rMlxipl* is not affected in heterozygous *rRfx6* +/- Ins-1 832/13 clones

As mentioned above, we generated four different *rRfx6* clones KO rR-1, KO rR-2, KO rR-3 and KO rR-4 carrying heterozygous mutations in exon 2 and exon 3. The mutations c.223_365+57del (Δ200 bp, KO rR-1), c.240_365+51del (Δ177 bp, KO rR-2) and c.299_365+52del (Δ119 bp, KO rR-3) in exon 2 and the c.449_457del;c.459_489+18del (Δ58 bp, KO rR-4) mutation in exon 3 caused the loss of the 3' splice donor site of exon 2 and exon 3, respectively, while the 5' splice acceptor site of both exons remained intact (**FIGURE C25A+B**). To figure out whether the loss of the splicing sites caused aberrant splicing, we amplified the 5' end of the mRNA from exon 1 to exon 4. On the agarose gel, we detected two distinct bands, one formed by the wild-type transcript and smaller ones corresponding to the transcript without exon 2 or exon 3 (**FIGURE C25C**). We further confirmed these findings by sequencing. The loss of exon 2 or exon 3 causes the incorporation of a premature stop codon (p.E212GfsX24 for Δexon 2 and p.W122X for Δexon 3). The predicted proteins are 92 aa or 72 aa long (**FIGURE C25B**) and do not contain the DNA-binding domain. On a western blot, the full-length rRFX6 protein amount was halved (**FIGURE C25D**). Taken together, we confirmed the heterozygosity of the four Ins-1 832/13 cells *rRfx6* +/- clones on the DNA and protein level.

We wanted to use the heterozygous clones to find out if the loss of a *Rfx6* allele led to an effect on rRFX6 target genes in rat cells. To address this question, we analysed the expression of the rRFX6 target gene *rMlxipl* in the four *rRfx6* heterozygous knockout lines KO rR-1, KO rR-2, KO rR-3 and KO rR-4 and in three control lines WT rR-1, WT rR-2 and WT rR-3 (**FIGURE C26**), but there was no significant difference between *rRfx6* +/+ and *rRfx6* +/- cells (**FIGURE C26A**). To figure out whether there might be an effect after a glucose stimulus, we starved the cells overnight in low glucose medium, incubated

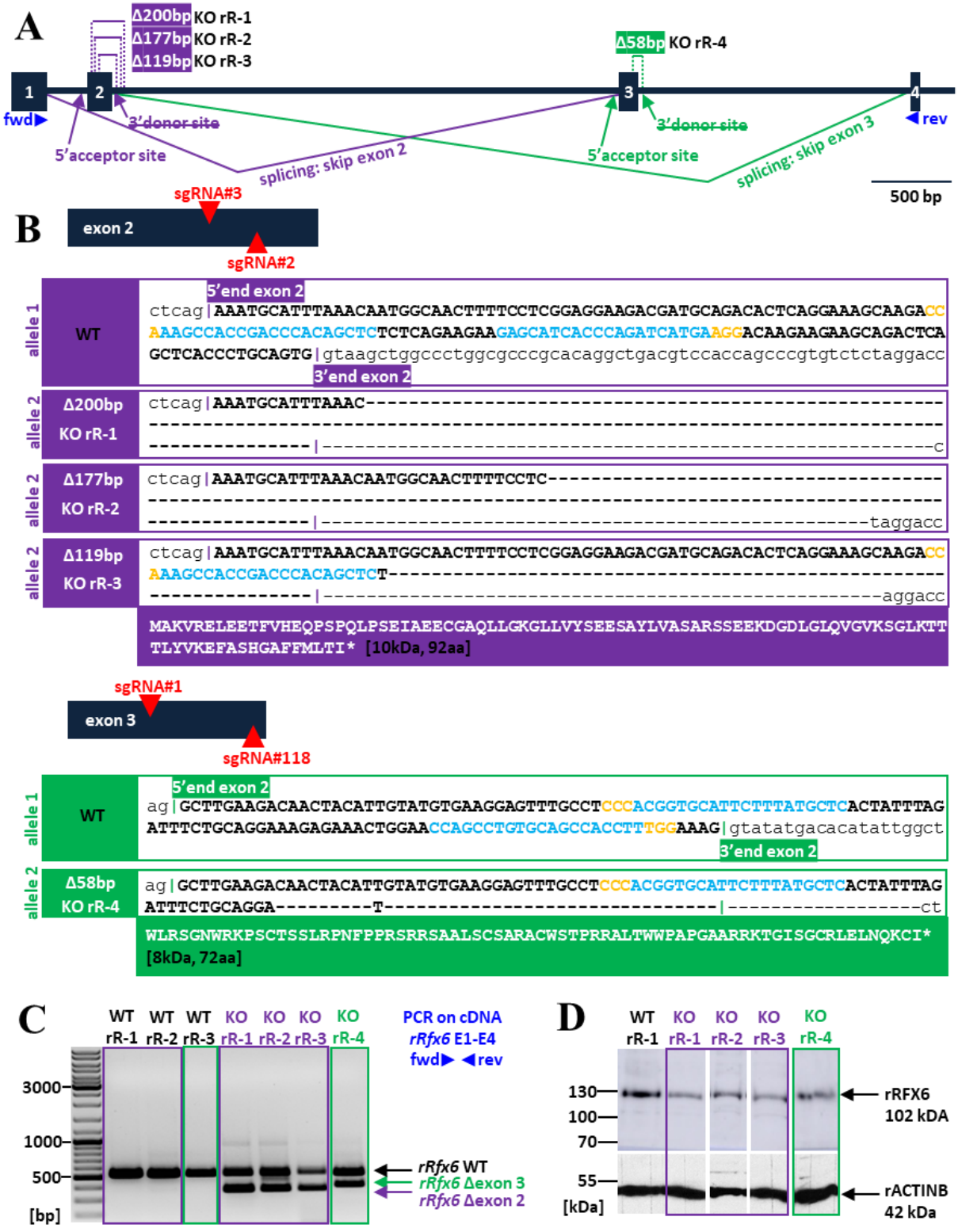


FIGURE C25: The *rRfx6* +/- Ins-1 832/13 cell lines carry large deletions on one *rRfx6* allele causing aberrant splicing. (A) Scheme illustrating the positions of the deletions in and the loss of the 3' splicing donor sites of *rRfx6* exon 2 and exon 3 in the *rRfx6* heterozygous Ins-1 832/13 cell lines KO rR-1/2/3/4. **(B)** Sequence of the mutated alleles in the heterozygous *rRfx6* Ins-1 832/13 cell lines KO rR-1/2/3/4 and the respective wild-type allele. **(C)** Analysis of the mRNA splicing in the four heterozygous rRfx6 Ins-1 832/13 cell lines KO rR-1/2/3/4 and the three wild-type control Ins-1 832/13 cell lines WT rR-1/2/3 by PCR on cDNAs amplifying exon 1 to 4 (primer positions are marked in **(A)**). **(D)** Western blot with an anti-RFX6 antibody (rabbit anti-RFX6 2766) on the four heterozygous *rRfx6* Ins-1 832/13 clones KO rR-1/2/3/4 and on one wild-type control clone WT rR-1.

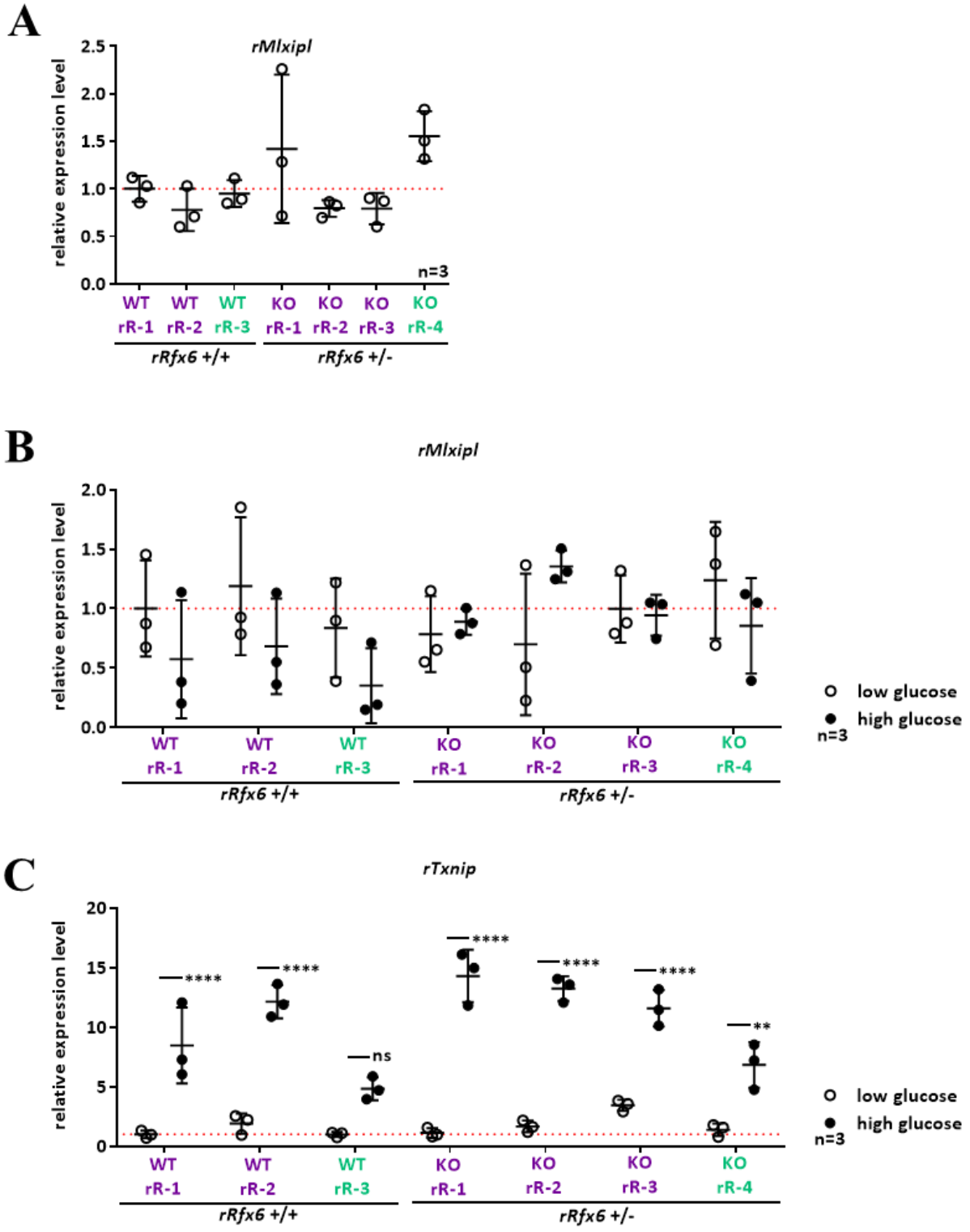


FIGURE C26: The expression of *rMlxipl* is not affected in *rRfx6* +/- Ins-1 832/13 cells. (A) RT-qPCR analysis of *rMlxipl* expression in Ins-1 832/13 wild-type control cell lines WT rR-1/2/3 and in heterozygous *rRfx6* +/- cell lines KO rR-1/2/3/4 incubated in standard cell culture medium (11 mM glucose). **(B, C)** RT-qPCR analysis of *rMlxipl* **(B)** and *rTxnip* **(C)** expression in Ins-1 832/13 wild-type control cell lines WT rR-1/2/3 and in heterozygous *rRfx6* +/- cell lines KO rR-1/2/3/4 starved overnight in standard cell culture medium with 2 mM glucose medium and incubated for 6 h in standard cell culture medium supplemented with 2 mM (low) or 20 mM (high) glucose. Data are shown as mean plus standard deviation (n=3, technical triplicates). Statistical significances (p≥0.05 not significant (ns), p<0.01 **, p<0.0001 ****) were determined by two-way ANOVA.

them for 6 h in low or high glucose medium and analysed the expression of *rMlxipl* by RT-qPCR. We observed a highly variable *rMlxipl* expression among clones of the same genotype and among replicates, but no effect of the *rRfx6* heterozygosity (**FIGURE C26B**). We further examined glucose-stimulated *rTxnip* expression, and we also found that the induction of *rTxnip* also varied a lot among *rRfx6* +/+ clones and *rRfx6* +/- clones (**FIGURE C26C**).

From our experiments, we concluded that the amount of rRFX6 in the heterozygous cells is enough to properly regulate *rMlxipl* in the rat beta cell line Ins-1 832/13. This finding is consistent with the observations in mice where the loss of one *mRfx6* allele did not cause any obvious phenotype (Smith *et al.,* 2010; Piccand *et al.,* 2014) while the consequences of heterozygous *RFX6* mutations in humans are not clear and controversially discussed. The parents of Mitchell-Riley syndrome patients sometimes, but not always, have diabetes themselves, although they are all heterozygous carriers of *RFX6* mutations (Mitchell *et al.,* 2004; Chappell *et al.,* 2008; Martinovici *et al.,* 2009; Smith *et al.,* 2010; Spiegel *et al.,* 2011; Concepcion *et al.,* 2014; Sansbury *et al.,* 2015; Huopio *et al.,* 2016; Skopkova *et al.,* 2016). Huopio *et al.* reported no predisposition for diabetes in heterozygous carriers (Huopio *et al.,* 2016) whereas Patel *et al.* found an enrichment of heterozygous carriers in maturity-onset diabetes of the young (MODY) patients (Patel *et al.,* 2017). Zhu *et al.* produced heterozygous and homozygous *RFX6* mutants in human embryonic stem cells and they differentiated them *in vitro* into glucose-responsive beta-like cells but they do not mention any effect of heterozygous *RFX6* mutations on the differentiation capacity while the formation of pancreatic progenitor cells is impaired in *RFX6* -/- mutants (Zhu *et al.,* 2016). In summary, the loss of a *Rfx6* allele in rodents seems to have no effect, while the consequences in humans are not clear.

2.4. Is the production of knockout clones in beta cell lines worth the effort?

For now, it was not in the bound of possibility to produce homozygous *rRfx6* and *rMlxipl* knockout clones. The generation of knockout clones is very time consuming and necessitates many analysis before one can be sure to have a homozygous knockout. To generate the knockout clones, we used the Cas9n with two sgRNAs and we tested two cloning strategies, one based on puromycin selection and limiting dilutions technique, and the other one on GFP selection and FACS. Beta cells usually grow in cluster, so it is complicate to get single surviving cells. The Ins-1E cell line was produced from the parenteral Ins-1 cell line using the limiting dilutions technique (Merglen *et al.,* 2004). Ins-1 832/3, 832/13, 832/21 and 832/24 cells were generated from Ins-1 cells by transfection and neomycin selection (Hohmeier *et al.,* 2000).

To our knowledge, there are only four studies about CRISRPR/Cas9 in Ins-1 cells (Naylor *et al.,* 2016; Bompada *et al.,* 2016; Merriman *et al.,* 2018; Peterson *et al.,* 2018) published to date. Naylor *et al.*

knocked out GIP and GLP-1 receptor in Ins-1 832/3 cells using Cas9 with one gRNA, GFP selection and FACS (Naylor *et al.,* 2016). Bompada *et al.* produced knockouts of the histone acetyltransferase EP300 in Ins-1 832/13 cells. They used the Cas9 with two gRNAs and the limiting dilutions technique (Bompada *et al.,* 2016). Merriman *et al.* targeted the zinc transporter ZnT8 with Cas9 and one gRNA in Ins-1E cells and they selected transfected cells by FACS (Merriman *et al.,* 2018). Peterson *et al.* worked with Ins-1 832/13 cells. They produced knockouts of the mitochondrial protein deacetylase SIRT3 using the Cas9n and two sgRNAs, puromycin selection and limiting dilutions technique.

In their study, Naylor *et al.* mentioned the problem of the copy number variants and they tested it for their genes before they started the generation of knockout clones (Naylor *et al.,* 2016). Since Ins-1 cells and the sublines are tumour cells, it is a good suggestion to keep in mind that there might be abnormal gene copy numbers. As described above, we suppose that there might be three *rMlxipl* alleles in Ins-1 832/13 cells.

Moreover, the study of Peterson *et al*. described a clonal variation after Ins-1 832/13 cloning and suggests that caution should be exercised in the analysis of monoclonal Ins-1 832/13 knockout clones (Peterson *et al.,* 2018). In their study, they knocked out the mitochondrial deacetylase *Sirt3* using the CRISPR/Cas9 technique, but when they analysed glucose-induced insulin secretion, they got a variable response among the knockout clones with complete SIRT3 loss. They next subcloned wild-type cells and tested the glucose-induced insulin secretion in these subclones and again, they observed a clonal variability. They speculated that this might be the case for most if not all insulinoma cell lines. In the Min6b1 cells, we also observed a high variability in the transcript levels (**FIGURE C13C**, **FIGURE C14C**) in the clones KO mR-1 and KO mM-1. This suggests that the heterogeneity of clonal cell lines might indeed not be exclusive to Ins-1 832/13 cells. We also observed this huge variability in our RT-qPCR results in the *rRfx6* +/- and *rMlxipl* clones and control cell lines, and we wondered where this problem came from. We did not have similar issues with data heterogeneity in the past when we analysed non-treated or treated wild-type Ins-1 832/13 cells. A knockout has the advantage that the protein is completely removed from the cell. In knockdown experiments, the amount of protein is only reduced so that there is still a residual activity. Depending on the activity of the respective protein, this may have a strong influence on the phenotype. However, the observations of Petersen *et al.* suggest that the analysis of a knockout phenotype in beta cell lines is not much simpler than that of a knockdown phenotype unless the loss of a protein has such a strong impact that variability becomes negligible.

In conclusion, the published results in the literature proof that it is possible to generate knockout clones in Ins-1 cells using the Cas9 or Cas9n and the limiting dilutions technique or a FACS-based strategy. However, before starting, one should first check if there are genomic alterations at the

targeted loci, and, once the clones are produced, the existing clonal variability must absolutely be considered in the analysis.

To knockout *rRfx6* and *rMlxipl* in Ins-1 832/13 cells, we should first check the copy numbers of both genes, and than re-start the knockout. We would need to generate more optimal sgRNA pairs targeting *rRfx6* and *rMlxipl* to have better tools to knockout both genes and we must screen higher numbers of clones after the transfection, selection and isolation. Since this would be very time consuming and the time for a thesis is limited (and we run out of time at this point), we decided to stop all attempts to generate knockout clones and we continued with a backup RNA interference strategy. Although RNA interference is not optimal, it should allow us to knockdown *rRfx6* and *rMlxipl* in Ins-1 832/13 cells and to find the rRFX6 and rMLXIPL target genes in this cell line.

2.5. Identification of RFX6- and MLXIPL-regulated genetic programs in Ins-1 832/13 cells

The aim of this second part of the thesis project was to figure out to what extend the loss of *Mlxipl* contributes to the *Rfx6* loss-of-function phenotype in adult beta cells and to identify the genes that are directly regulated by RFX6 or regulated via MLXIPL. The experiments described so far suggest that we should stay with Ins-1 832/13 cells as a model, but that instead of a knockout we should prefer a knockdown. The use of the siRNAs should allow us to reduce the levels of *rRfx6* and *rMlxipl*, and, by RNA sequencing, we should be able to identify the genes regulated by RFX6 and MLXIPL in low and high glucose conditions in Ins-1 832/13 cells.

2.5.1. The knockdown of *rRfx6* efficiently diminishes *rMlxipl* expression in Ins-1 832/13 cells

In previous experiments, we and others never managed to reduce *Rfx6* transcript levels by more than 50 %, regardless of whether we tried this in Min6b1 cells (**FIGURE C3D**), Ins-1 832/13 cells (**FIGURE C3E**), EndoC-betaH2 cells or in human islets (Vikash Chandra from the Cochin institute in Paris, unpublished data). By contrast, the knockdown of *rRfx3* and *rNkx2.2* worked fine in Ins-1 832/13 in our hands (**FIGURE C9B+C**) while the knockdown of *rRfx6* and *rMlxipl* in the same cells transfected on the same day with the same reagents and the same parameters was less efficient. These observations suggest that an efficient knockdown of *rRfx6* and *rMlxipl* could be detrimental to Ins-1 832/13 cells, and they fit the fact that we were unable to produce homozygous knockout clones.

In a time-course experiment, we examined the knockdown of *rRfx6* in Ins-1 832/13 cells after 24 h, 48 h and 72 h of siRNA treatment. As usual, we got a maximal downregulation of 50 %. The level of transcription rose slightly between the 24-h and the 48-h point and returned to the normal expression level after 72 h (**FIGURE C27A**), likely due to an increased *rRfx6* transcription in response to the decrease of the *rRfx6* mRNA and rRFX6 protein level by the siRNA treatment. The protein level,

however, decreased continuously from the 24-h point to the 72-h point (**FIGURE C27B**) suggesting a prolonged effect on the translation. The best effect on the rRFX6 target *rMlxipl* was observed after 48 h (**FIGURE C27C**), and rMLXIPL protein was not affected after 24 h but strongly diminished after 48 h and 72 h (**FIGURE C27D**). From this experiment, we concluded that we should wait more than 24 h before we start the incubation in low or high glucose medium because the best downregulation at both the transcript and protein level happened between 24 h and 48 h.

2.5.2. An efficient knockdown of *rMlxipl* α is cell-toxic for Ins-1 832/13 cells

For the knockdown of *rMlxipl*, we decided to compare the efficiency of a siRNA that was used in the Schmidt *et al.* study and that targets exon 12 (called exon 12 siRNA hereafter; Schmidt *et al.*, 2016), with the ON-TARGETplus SmartPool siRNA of Dharmacon (called SmartPool siRNA hereafter) that we already used successfully in the transactivation assays (**FIGURE C28A**). The latter contains four siRNA molecules targeting exon 10, 11, 12 and 14 of *rMlxipl*.

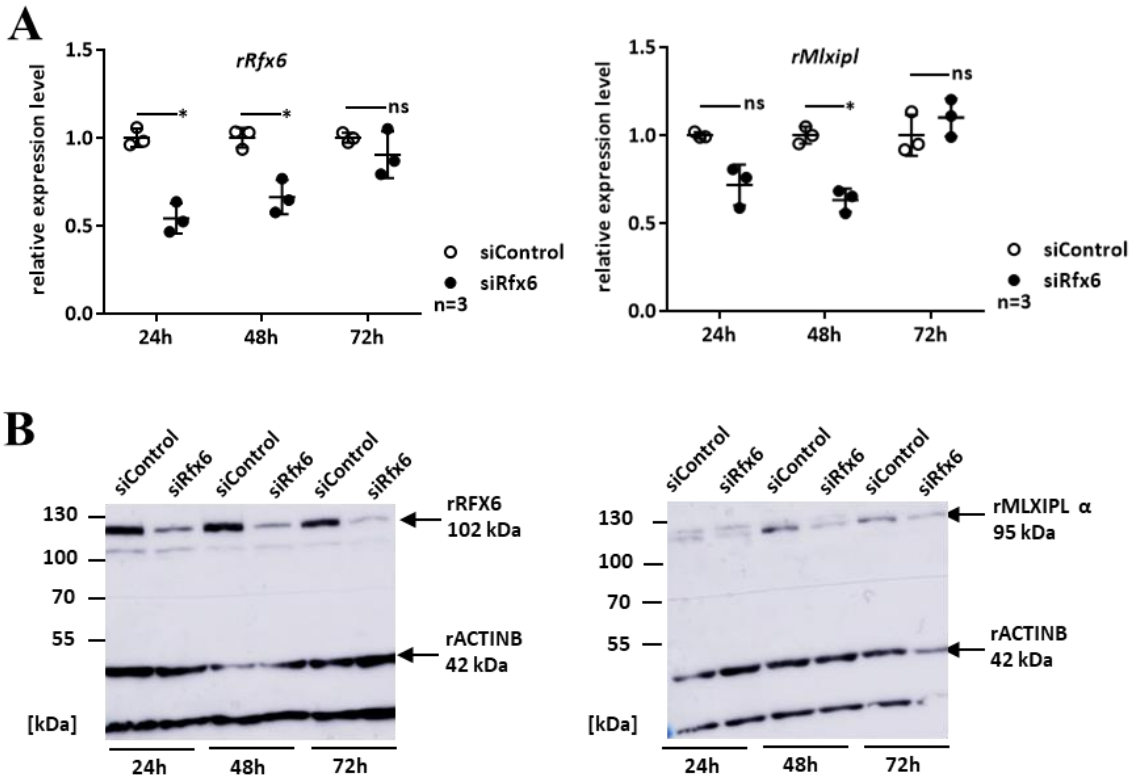


FIGURE C27: Time course of the knockdown of *rRfx6* by RNA interference and of the concomitant reduction of its target gene *rMlxipl* in Ins-1 832/13 cells. (A) RT-qPCR analysis of *rRfx6* (on the left) and *rMlxipl* (on the right) expression in siControl- or siRfx6-treated Ins-1 832/13 cells 24 h, 48 h or 72 h after the transfection with the siRNA. Data are shown as mean plus standard deviation (n=3). Statistical significances ($p \geq 0.05$ not significant (ns), $p < 0.05$ *) were determined using an unpaired two-tailed t test. **(B)** Western blot analysing rRFX6 (on the left, rabbit anti-RFX6 2766) and rMLXIPL (on the right, rabbit anti-MLXIPL, Novus) in siControl- or siRfx6-treated Ins-1 832/13 cells 24 h, 48 h or 72 h after the transfection with the siRNA.

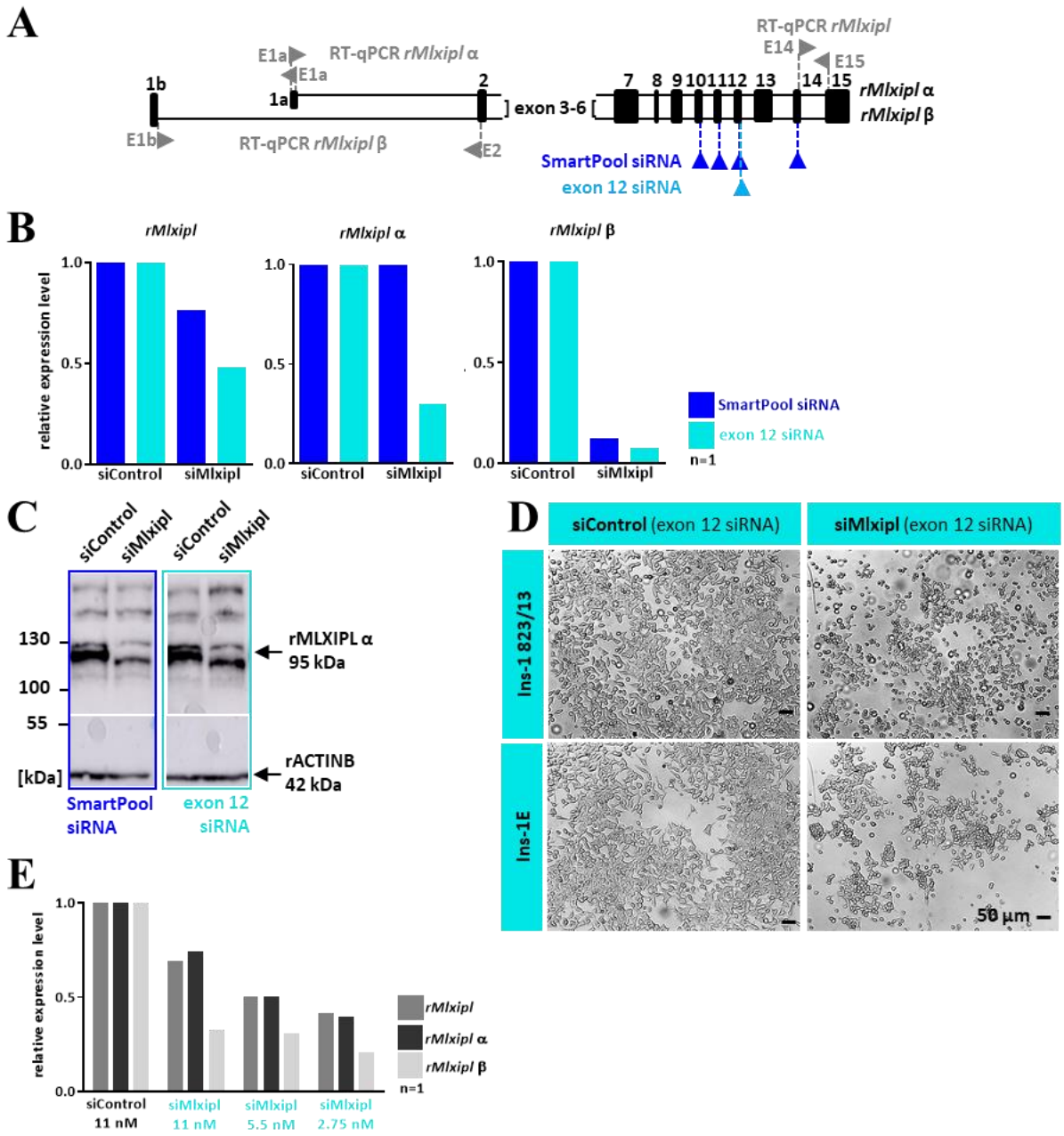


FIGURE C28: The exon 12 siRNA against *rMlxipl* efficiently reduces *rMlxipl* expression but it is cell toxic. (A) Position of the siRNA sequences of a SmartPool siRNA (Dharmacon) and a single exon 12 siRNA (Schmidt *et al.*, 2016) against *rMlxipl* as well as of the *rMlxipl*, *rMlxipl* α and *rMlxipl* β RT-qPCR primer in rat *rMlxipl* gene. **(B)** RT-qPCR analysis of *rMlxipl*, *rMlxipl* α and *rMlxipl* β expression in Ins-1 832/13 cells treated with the SmartPool siRNA and the exon 12 siRNA targeting *rMlxipl* and the corresponding non-targeting control siRNAs (all at 11 nM) for 48 h in standard cell culture medium (at 11 mM glucose). **(C)** Western blot with an anti-MLXIPL antibody (rabbit anti-MLXIPL, Novus) on Ins-1 832/13 cells treated with the SmartPool siRNA and the exon 12 siRNA targeting *rMlxipl* and the corresponding non-targeting control siRNAs (all at 11 nM) for 48 h in standard cell culture medium (at 11 mM glucose). **(D)** Brightfield images of Ins-1 832/13 (above) and Ins-1E (below) cells transfected with the exon 12 siRNA targeting *rMlxipl* and the corresponding non-targeting control siRNA (both at 11 nM) in standard cell culture medium at 11 mM glucose and, after 24 h, starved for 24 h in DMEM at 2 mM glucose. **(E)** RT-qPCR analysis of *rMlxipl*, *rMlxipl* α and *rMlxipl* β expression in Ins-1 832/13 cells transfected with the exon 12 siRNA targeting *rMlxipl* at 2.75 nM, 5.5 nM or 11 nM and the corresponding non-targeting control siRNA at 11 nM in standard cell culture medium at 11 mM glucose and, after 24 h, starved for 24 h in DMEM at 2 mM glucose.

We tested the downregulation of *rMlxipl* and its transcript variants *rMlxipl* α and *rMlxipl* β by the exon 12 and the SmartPool siRNA. Both siRNAs had a very good effect on *rMlxipl* β but the SmartPool siRNA did not affect *rMlxipl* α (**FIGURE C28B**). On the protein level, the effect of both siRNAs was similar (**FIGURE C28C**). Unexpectedly, we observed a cytotoxic effect of exon 12 siRNA, which was even worse in low glucose medium. In the study in which this siRNA was published, it was applied to Ins-1E cells and they did not report any toxic effect (Schmidt *et al.,* 2016). To find out whether these are cell line differences, we treated Ins-1 832/13 and Ins-1E cells with this siRNA and found increased cell death in both cell lines in low (**FIGURE C28D**) and standard (**not shown**) glucose medium. In fact, we used this siRNA at 11 nM, a much lower concentrations than in their study (150 nM), so we were quite surprised that they had no problems with this siRNA. It seems unlikely that a reduction of *Mlxipl* leads to massive beta cell death since *mMlxipl* -/- mouse do not have diabetes (Iizuka *et al.,* 2004). Nevertheless, we once again have indications that Ins-1 832/13 cells cannot survive without *rMlxipl* α. The main difference between the cytotoxic exon 12 siRNA and the SmartPool siRNA is that the former more efficiently targeted *rMlxipl* α.

To minimize the cytotoxic effect, we tested lower concentrations of the exon 12 siRNA (2.75 nM and 5.5 nM instead of 11 nM). With all three concentrations, we were able to knockdown *rMlxipl, rMlxipl* α and *rMlxipl* β in Ins-1 832/13 cells. In cells treated with 2.75 nM and 5.5 nM exon 12 siRNA, we observed a downregulation of *rMlxipl* α by about 50 % and of *rMlxipl* β by about 75 % (**FIGURE C28E**) and a reduced cell death compared to cells treated with 11 nM exon 12 siRNA. The survival rate was better, the less of the exon 12 siRNA was used (**not shown**). Taken together, the exon 12 siRNA seemed to work better than the SmartPool siRNA.

In summary, we have siRNAs targeting *rRfx6* and *rMlxipl* in Ins-1 832/13 cells that allow us to reduce the expression of both genes, but we will have to deal with the fact that we achieve a maximal downregulation of 50 %.

2.5.3. Validation of the siRNA and glucose treatment by RT-qPCR prior to RNA sequencing

To analyse the rRFX6- and rMLXIPL-dependent transcriptomes in low and high glucose in Ins-1 832/13 cells, we treated Ins-1 832/13 cells with siControl and siRfx6 (SmartPool and the respective control, both at 11 nM) and siControl and siMlxipl (exon 12 siRNA plus the respective control, both at 2.75 nM). After 24 h, the siRNA transfection medium was replaced by 2 mM glucose medium to starve the cells. After an overnight incubation, cells were changed to 2 mM or 20 mM glucose medium and RNA was isolated after 6 h (**FIGURE C29A**). To validate the samples (n=4 per condition), we measured the transcript levels of *rRfx6, rMlxipl* and *rTxnip* by RT-qPCR (**FIGURE C29B+C**) prior to RNA sequencing.

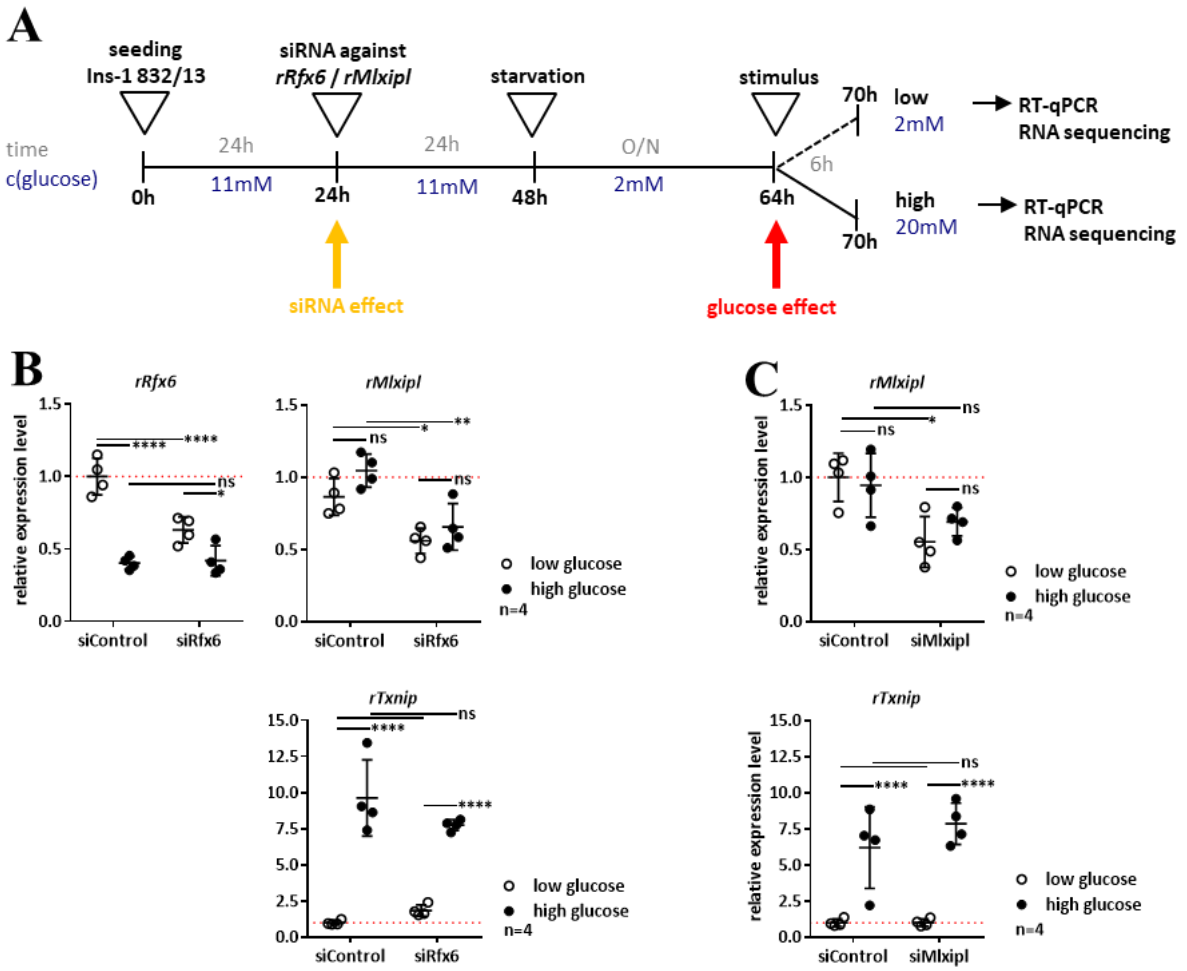


FIGURE C29: Validation of the *rRfx6* and *rMlxipl* knockdown in Ins-1 832/13 cells by RT-qPCR prior to RNA sequencing. (A) Experimental strategy. Ins-1 832/13 cells were transfected with siControl, siRfx6 or siMlxipl. After 24 h, the transfection medium was replaced by low glucose medium (standard cell culture medium with 2 mM glucose), and, after an overnight incubation, with low or high glucose medium (standard cell culture medium with 2 or 20 mM glucose, respectively). RNA was isolated after 6 h and transcript levels were measured by RT-qPCR. **(B)** RT-qPCR analysis of *rRfx6*, *rMlxipl* and *rTxnip* expression in siControl- or siRfx6-treated Ins-1 832/13 cells (both are SmartPool siRNAs, used at 11 nM). **(C)** RT-qPCR analysis of *rMlxipl* and *rTxnip* expression in siControl- or siMlxipl-treated Ins-1 832/13 cells (exon 12 siRNA against *rMlxipl* and corresponding control siRNA, used at 2.75 nM). Data are shown as mean plus standard deviation (n=4). Statistical significances ($p \geq 0.05$ not significant (ns), $p<0.05$ *, $p<0.01$ **, $p<0.0001$ ****) were determined by two-way ANOVA.

In the siRfx6- and siMlxipl-treated cells, we got a knockdown level of *rRfx6* and *rMlxipl* by about 50 % in both the cells incubated for 6 h in low or high glucose medium (**FIGURE C29B+C**). However, the 6 h incubation of siControl-treated cells in high glucose medium resulted in a significant *rRfx6* downregulation by about 50 % (**FIGURE C29B**); *rMlxipl* expression was not altered by the high glucose level in siControl-treated cells (**FIGURE C29C**). We wondered whether the glucose-induced reduction in *rRfx6* in siControl-treated cells might have consequences on rRFX6 target genes in these cells. This would mean that we cannot use the data sets of the siControl- and siRfx6-treated cells in high glucose because we would not be able to distinguish between gene expression changes caused

by the glucose-mediated reduction of rRFX6 and the siRNA-mediated reduction of rRFX6. Fortunately, *rMlxipl* expression remained unchanged in cells treated with siControl after 6 h in high glucose medium although *rRfx6* transcript levels were reduced (**FIGURE C29B**), suggesting that the time was not long enough to reduce rRFX6 protein levels and to change the expression of rRFX6 target genes. By contrast, the siRfx6 treatment reduced the transcript levels of *rMlxipl* in low and high glucose medium as observed before (**FIGURE C29B**). Based on this observation, we decided that we can keep the samples for the RNA sequencing.

Interestingly, no glucose-induced reduction of *rRfx6* (**FIGURE C16E**) could be observed in the siRfx6-treated cells; the decrease of *rRfx6* after the siRfx6 treatment was the same in low and high glucose medium (**FIGURE C29B**). From the *rNkx2.2* and *rRfx3* knockdown assays (**FIGURE C9B+C**), we know that it is possible to have an additive effect of glucose in siRNA-treated cells. This might indicate that Ins-1 832/13 cells cannot tolerate a reduction of *rRfx6* by more than half.

The incubation in high glucose increased the expression of the glucose-induced gene *rTxnip* by 10-fold in all siControl-, siRfx6- and siMlxipl-treated cells (**FIGURE C29B+C**), further confirming that we can use the samples to compare rRFX6 and rMLXIPL targets in low and high glucose. We also noticed that neither the knockdown of *rRfx6* nor the one of *rMlxipl* affected *rTxnip* induction. This is consistent with the data of Richards *et al.* who had shown that *mTxnip* expression was unaffected in islets of adult *mMlxipl* knockout mice after a glucose stimulus and that the treatment of the human EndoC-betaH1 cell line with siMlxipl does not alter *TXNIP* expression. They concluded that *Txnip* is more likely an MLXIP target than a MXLIPL target gene (Richards *et al.,* 2018). Our data show that *rTnxip* is also not a target gene of rMLXIPL in Ins-1 832/13 cells.

In conclusion, the RT-qPCR analysis allowed us to verify the knockdown of *rRfx6* and *rMlxipl* and to confirm that the 6-h incubation in high glucose was enough to alter the expression of glucose-sensitive genes. Thus, we decided to sequence the transcriptomes of these samples to identify rRFX6 and rMLXIPL target genes in low and high glucose in Ins-1 832/13 cells.

2.5.4. The gene *rRfx6* is a direct targets of rMLXIPL in Ins-1 832/13 cells

From the RNA sequencing data in siMlxipl-treated Ins-1 832/13 cells and published Ins-1E anti-MLXIPL ChIP sequencing data from Schmidt *et al.* (GSE81628, Schmidt *et al.,* 2016), we deduced that the reduction of *rRfx6* that we observed in high glucose in siControl-treated cells (**FIGURE C29B**), is due to a direct regulation of *rRfx6* by rMLXIPL (**FIGURE C30**). In fact, we always tested the expression of *rMlxipl* in siRfx6-treated cells, but we never did in the other direction, that means testing the expression of *rRfx6* in siMlxipl-treated cells. The knockdown of *rMlxipl* in Ins-1 832/13 cells prevented the glucose-mediated downregulation of *rRfx6* in high glucose, but it did not affect *rRfx6* expression in

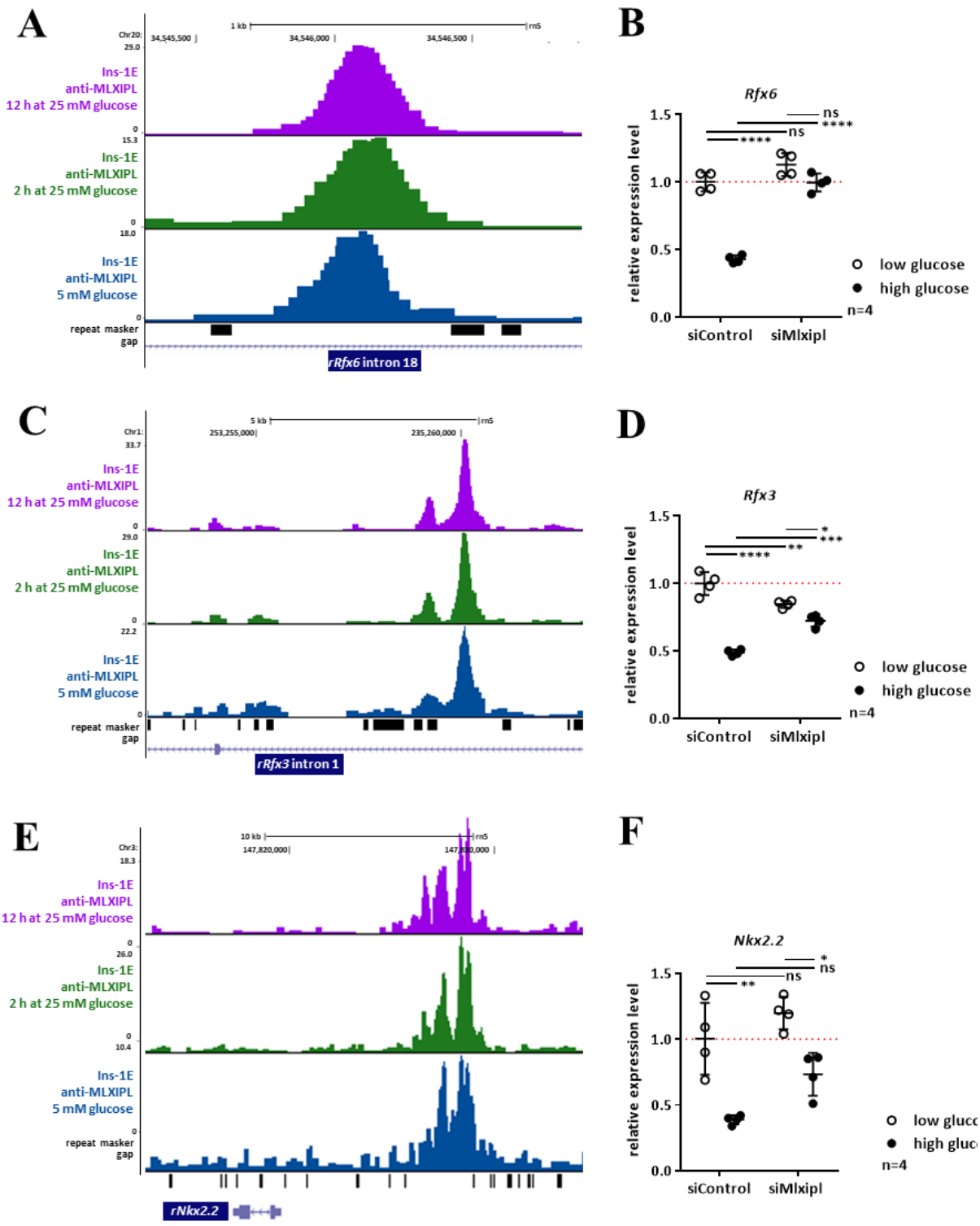


FIGURE C30: MLXIPL binds *rRfx6, rRfx3* and *rNkx2.2* in Ins-1E cells and inhibits their expression in high glucose in Ins-1 832/13 cells. (A, C, E) ChIP sequencing data (Schmidt *et al.*, 2016, GSE81628) showing the binding of rMLXIPL on *rRfx6* intron 18 **(A)**, *rRfx3* intron 1 **(C)** and in the *rNkx2.2* promoter **(E)** in Ins-1E cells starved for 24 h in 5 mM glucose (blue profile), as well as after a subsequent incubation in 25 mM glucose for 2 h (green profile) or 12 h (purple profile). UCSC alignment (rn5) was done by Constance Vagne. **(B, D)** Relative expression of *rRfx6* **(B)**, *rRfx3* **(D)** and *rNkx2.2* **(E)** measured by RNA sequencing of siControl- and siMlxipl-treated Ins-1 832/13 cells after an overnight starvation in low glucose medium (2 mM glucose) and a 6-h incubation in low or high glucose medium (2 mM and 20 mM glucose, respectively).

low glucose (**FIGURE C30B**), suggesting that rMLXIPL represses *rRfx6* in high glucose. Moreover, the ChIP sequencing data reveal that rMLXIPL binds to the last intron of *rRfx6* (**FIGURE C30A**), thus the regulation is likely direct.

Similarly, we also found that rMLXIPL binds to *rRfx3* (intron 1, **FIGURE C30C**) and *rNkx2.2* (promoter, **FIGURE C30E**) in Ins-1E cells and that the knockdown of *rMlxipl* in Ins-1 832/13 cells significantly lowers the glucose-induced downregulation of *rRfx3* (**FIGURE C30D**) and slightly diminishes the decrease of *rNkx2.2* in high glucose (**FIGURE C30F**).

In conclusion, these data show that *rMlxipl* is not only regulated by rRFX6, but that *rRfx6* is also a direct target gene of rMLXIPL. Since we have knocked down both *rMlxipl* α and *rMlxipl* β and since the antibody in ChIP can bind to both proteins, we cannot comment on whether the regulation of *rRfx6, rRfx3* and *rNkx2.2* is done by rMLXIPL α or rMLXIPL β or both. Nevertheless, one can certainly say that this is a feedback control. It was already known that MLXIPL α and β activate the expression of *Mlxipl* β in high glucose in a feedforward control (Herman *et al.,* 2012; Zhang *et al.,* 2015; Jing *et al.,* 2016). In the first part of the results, we further showed that rMLXIPL actively inhibits the expression of *rMlxipl* α. With these new findings, we now show that the *rMlxipl* α activator rRFX6 is also inhibited, which ensures on an additional level that the expression of *rMlxipl* α decreases in high glucose conditions after the activation of *rMlxipl* β transcription. These new results certainly require further research to be able to draw clear conclusions.

2.5.5. Does the knockdown of *rRfx6* and *rMlxipl* affect known targets in Ins-1 832/13 cells?

To validate our RNA sequencing data, we took a closer look on known RFX6 and MLXIPL target genes to find out whether they are differentially expressed after the siRfx6 and siMlxipl treatment (**TABLE C5**).

RFX6 is known to activate genes involved in insulin secretion (*Gck*, *Abcc8*, calcium channels; **FIGURE A9**) in adult murine beta cells and in the human beta cell line EndoC-betaH2 (Piccand *et al.,* 2014; Chandra *et al.,* 2014) and to repress disallowed genes (**FIGURE A9**; Piccand *et al.,* 2014). In the siRfx6-treated cells, we could confirm that rRFX6 is important for the expression of *Gck* and *Abcc8* in low and high glucose, ant the downregulation of 10 calcium channels in low and 11 in high glucose, for example *Cacn1b* (**TABLE C5**). From the disallowed genes that should specifically silenced in mature beta cells (Quintens *et al.,* 2008; Pullen *et al.*, 2010; Thorrez *et al.,* 2011), we found six genes (*Fgf1*, *Gas6*, *Gsta4*, *Igfbp4*, *Ndrg2*, *Ndrg4*) to be upregulated in low glucose and seven genes to be upregulated in high glucose (*Fgf1*, *Gsta4*, *Igfbp4*, *Lmo4*, *Ndrg2*, *Nfib*, *Parp3*) in the siRfx6-treated Ins-1 832/13 cells (**TABLE C5**). The other disallowed genes are either upregulated, not affected by the knockdown or not expressed in Ins-1 832/13 cells (e.g. *Ldha*). We also checked the expression of cilio-

TABLE C5: Regulation of known RFX6 and MLXIPL target genes in siRfx6- and siMlxipl-treated Ins-1 832/13 cells. List of published RFX6 and MLXIPL target genes and their expression in the siRfx6-treated (in purple) or in siMlxipl-treated (in blue) cells at 2 mM (low) and 20 mM (high) glucose identified by RNA sequencing. Significantly upregulated genes are marked in green (Log2 FC>0 and p<0.05) and significantly downregulated genes in red (Log2 FC<0 and p<0.05). See **FIGURE C29A** for more information about the samples.

	gene name	normalized expression siControl - low glucose	normalized expression siRfx6 - low glucose	Log2 FC	Adjusted p-value	normalized expression siControl - high glucose	normalized expression siRfx6 - high glucose	Log2 FC	Adjusted p-value	normalized expression siControl - low glucose	normalized expression siMlxipl - low glucose	Log2 FC	Adjusted p-value	normalized expression siControl - high glucose	normalized expression siMlxipl - high glucose	Log2 FC	Adjusted p-value
INSULIN SECRETION GENES	Slc2a2	15272	18452	0,27	9,99E-08	20359	21522	0,08	ns	16484	28083	0,76	9,34E-57	18189	31845	0,80	4,75E-63
	Gck	2030	1436	-0,50	1,57E-26	2046	1483	-0,46	1,11E-22	2036	1925	-0,08	ns	1945	2005	0,05	ns
	Abcc8	7634	6892	-0,15	4,68E-02	8151	7253	-0,17	1,60E-02	5270	3976	-0,40	4,15E-10	5465	4673	-0,22	7,91E-04
	Kcnj11	4995	4460	-0,16	7,84E-03	3693	4005	0,12	ns	3313	2474	-0,42	1,27E-13	2248	2303	0,03	ns
	Ins1	76134	106489	0,48	5,05E-13	62590	94814	0,59	1,04E-19	59460	116136	0,96	5,70E-50	47374	114215	1,26	3,00E-86
	Ins2	1013	4252	2,04	9,38E-160	662	3282	2,27	2,27E-189	1289	10575	3,00	0,00E+00	684	8375	3,56	0,00E+00
	Iapp	32549	24692	-0,40	5,64E-14	21435	18764	-0,19	5,76E-04	22713	32671	0,52	7,80E-24	15253	30347	0,99	2,48E-83
	Pcsk1	3922	1799	-1,11	9,31E-68	2484	1368	-0,85	4,24E-38	2613	5404	1,04	1,67E-60	1592	4859	1,59	1,30E-137
	Pcsk2	15980	18641	0,22	3,60E-08	13261	16660	0,33	5,15E-17	12062	15792	0,39	2,83E-23	9015	15726	0,80	2,16E-96
	Chga	177122	141753	-0,32	1,03E-16	171386	136829	-0,32	3,26E-17	164946	158194	-0,06	ns	152714	154631	0,02	ns
	Chgb	98438	140030	0,50	2,21E-17	92994	133872	0,52	1,13E-18	75689	51132	-0,56	1,69E-21	71350	55861	-0,35	4,89E-09
	Slc30a8	6607	6129	-0,11	4,43E-03	3783	4627	0,29	8,99E-16	5745	7898	0,46	6,28E-42	2847	6955	1,28	7,22E-298
CALCIUM CHANNELS	Cacna1a	1936	1286	-0,59	1,87E-20	4683	2543	-0,87	1,31E-48	2705	974	-1,46	4,94E-121	4819	2299	-1,06	1,60E-70
	Cacna1b	308	348	0,17	ns	310	331	0,10	ns	367	510	0,46	6,13E-10	383	440	0,20	1,58E-02
	Cacna1c	1734	1275	-0,44	2,26E-10	2288	1609	-0,50	1,24E-13	1993	1551	-0,36	2,14E-07	2367	1641	-0,52	8,72E-15
	Cacna1d	2061	1815	-0,18	2,58E-03	2304	1877	-0,29	3,95E-07	2104	2147	0,02	ns	2195	2244	0,03	ns
	Cacna1g	647	326	-0,98	6,74E-38	375	277	-0,43	7,58E-07	551	565	0,03	ns	294	388	0,40	2,39E-06
	Cacna1h	1891	1662	-0,18	1,05E-02	3877	2658	-0,54	3,45E-18	1980	1279	-0,62	3,23E-22	3364	2263	-0,57	7,96E-20
	Cacna2d1	2955	1637	-0,84	3,09E-36	3322	1653	-1,00	4,35E-50	3343	2077	-0,68	3,28E-24	3790	2359	-0,68	3,00E-24
	Cacna2d2	462	487	0,07	ns	679	549	-0,31	9,19E-07	656	555	-0,25	1,04E-04	929	657	-0,49	1,63E-17
	Cacnb1	490	273	-0,84	5,60E-34	298	208	-0,50	1,30E-09	535	467	-0,20	3,27E-03	306	340	0,15	ns
	Cacnb2	282	136	-1,03	7,50E-29	189	106	-0,81	2,43E-14	268	254	-0,08	ns	152	224	0,54	3,98E-08
	Cacnb3	9237	7387	-0,32	8,30E-16	13087	9062	-0,53	2,68E-42	9682	7538	-0,36	1,02E-19	12819	8864	-0,53	1,17E-42
	Cacng2	205	219	0,09	ns	199	195	-0,02	ns	201	182	-0,14	ns	182	165	-0,14	ns
	Cacng4	864	724	-0,25	8,45E-06	1089	850	-0,36	2,10E-11	801	584	-0,45	2,89E-15	1098	618	-0,82	3,77E-50
	Cacng6	60	96	0,62	1,25E-03	22	62	1,30	2,96E-10	70	88	0,34	ns	26	78	1,39	2,67E-12
	Cacng7	1137	1350	0,25	3,88E-06	962	1266	0,40	9,52E-14	1228	1376	0,16	4,40E-03	1015	1204	0,24	8,27E-06
DISALLOWED GENES	Cxcl12	49	17	-1,29	4,42E-08	22	14	-0,53	ns	60	22	-1,27	1,85E-08	23	16	-0,44	ns
	Fgf1	125	413	1,69	6,50E-66	150	493	1,69	2,23E-70	70	37	-0,84	5,12E-08	100	49	-0,92	1,78E-10
	Gas6	2483	3649	0,55	8,29E-18	7297	5793	-0,33	2,50E-07	3530	1941	-0,85	2,94E-41	8848	4285	-1,04	2,73E-64
	Gsta4	151	347	1,17	9,34E-32	151	340	1,14	8,23E-30	112	101	-0,11	ns	115	87	-0,37	6,91E-03
	Igfbp4	1291	3754	1,52	2,30E-120	952	3275	1,76	6,09E-155	1139	754	-0,59	2,81E-16	816	694	-0,23	3,66E-03
	Lmo4	94	99	0,09	ns	64	102	0,66	4,46E-06	94	170	0,82	3,71E-11	69	146	1,04	3,16E-15
	Ndrg2	177	329	0,88	7,13E-19	94	237	1,31	1,55E-31	99	88	-0,17	ns	57	68	0,25	ns
	Ndrg4	1479	1930	0,38	9,29E-14	2350	2107	-0,16	3,26E-03	1616	3547	1,13	9,60E-122	2996	3813	0,35	5,12E-13
	Nfib	16	27	0,58	ns	13	30	1,05	1,58E-04	15	9	-0,47	ns	6	9	0,41	ns
	Parp3	379	336	-0,17	4,36E-02	190	238	0,30	8,57E-04	297	402	0,43	1,31E-08	150	319	1,07	2,58E-35
	Plec	1973	1280	-0,62	3,84E-21	1369	1228	-0,15	3,98E-02	2022	2688	0,40	4,80E-10	1079	2205	1,02	2,03E-55
	Tgm2	840	734	-0,19	1,95E-03	2158	1215	-0,83	8,87E-60	698	323	-1,08	4,35E-62	1697	677	-1,31	5,58E-126
	Tns1	18	28	0,51	ns	148	62	-1,17	3,49E-10	18	8	-0,78	3,19E-03	353	73	-2,16	8,42E-35
MLXIPL TARGETS	Acaca	1868	1916	0,03	ns	8675	4958	-0,80	3,61E-42	2429	1880	-0,37	8,77E-09	10115	3876	-1,37	8,01E-122
	Arnt	576	658	0,18	2,98E-03	613	654	0,09	ns	654	498	-0,39	5,60E-11	647	504	-0,36	4,58E-09
	Fasn	9817	9931	0,02	ns	20078	13556	-0,56	3,73E-23	14726	10551	-0,48	7,66E-17	32205	14333	-1,16	1,33E-96
	Pklr	2846	2923	0,04	ns	8631	5647	-0,61	7,55E-46	3068	1854	-0,72	3,53E-55	8405	4594	-0,87	7,78E-91
	Ppara	19	19	-0,04	ns	4	7	0,35	ns	16	15	-0,06	ns	4	4	-0,06	ns
	Rgs16	814	980	0,27	3,47E-05	3027	2365	-0,35	1,32E-10	1050	294	-1,79	1,20E-153	3023	1483	-1,02	3,35E-77
	Rorc	20	27	0,33	ns	102	62	-0,68	3,06E-05	33	18	-0,67	2,45E-03	94	51	-0,83	7,22E-07
	Sirt1	314	287	-0,12	ns	274	260	-0,07	ns	323	369	0,19	ns	285	351	0,29	8,13E-03
	Txnip	2150	3493	0,70	4,04E-63	20315	17138	-0,24	3,40E-10	2916	2486	-0,23	1,80E-07	21345	19073	-0,16	4,50E-05
	Ccnd2	1290	1910	0,56	1,32E-13	5050	3746	-0,43	5,15E-09	1354	812	-0,73	2,73E-21	5187	2171	-1,24	1,32E-68
	Ccna2	1451	1305	-0,15	8,70E-04	1574	1263	-0,32	1,61E-13	1312	1106	-0,25	4,13E-08	1465	1163	-0,34	4,37E-14
	Ccne1	684	780	0,19	3,30E-04	862	900	0,06	ns	473	821	0,78	2,28E-52	586	804	0,45	1,35E-18
	Cdk1	2295	1948	-0,23	1,69E-06	2363	1930	-0,29	1,83E-09	2377	1744	-0,44	1,35E-20	2646	1805	-0,55	1,34E-31
	Cdk2	713	783	0,13	4,10E-02	587	730	0,31	1,41E-07	514	654	0,35	5,72E-09	442	584	0,39	5,51E-10
	Cdk4	3966	4566	0,20	1,83E-10	4014	4477	0,16	1,30E-06	3805	3149	-0,27	1,04E-16	3947	3181	-0,31	1,60E-21

genic genes in the *rRfx6* knockdown cells, because another member of the RFX family, RFX3, is a major regulator of ciliogenic genes in the pancreas (Ait-Lounis *et al.,* 2007; Ait-Lounis *et al.,* 2010). RFX3 is a putative dimerization partner of RFX6 (Smith *et al.,* 2010), thus it might be possible that they coregulate gene expression. For instance, *Gck* is bound and regulated by both transcription factors (Ait-Lounis *et al.,* 2010; Piccand *et al.,* 2014). However, we did not find any involvement of rRFX6 in the regulation of the ciliogenic genes (Thomas *et al.,* 2010, review), suggesting that they do not play a role in the proper formation of cilia.

Since MLXIPL is known to regulate glycolytic and lipogenic genes in the liver (Iizuka *et al.,* 2004; Ishii *et al.,* 2004; Dentin *et al.,* 2004), we looked at genes from both groups and we found the known rMLXIPL target genes *Acaca*, *Fasn* and *Pklr* (Filhoulaud *et al.,* 2013, review) to be repressed in siMlxipl-treated Ins-1 832/13 cells (**TABLE C5**). Moreover, Metukuri *et al.* had reported that MLXIPL is important for the glucose-stimulated proliferation of Ins-1 832/13 cells. After 16 h in high glucose, they found a reduced upregulation of the cell cycle regulators *Ccnd2*, *Ccna2*, *Ccne1*, *Cdk1* and *Cdk2* and a reduced protein expression of CCND2, CCNA, CDK4, CCNE and CDK6 in siMlxipl-treated Ins-1 832/13 cells (Metukuri *et al.,* 2012). Our data confirmed the inhibition of *Ccnd2, Ccna2, Cdk1* and *Cdk4* (**TABLE C5**). Schmidt *et al.* suggested that the regulation of proliferation after a glucose stimulus is induced by RORC that acts as a downstream regulator of MLXIPL and is a direct target of rMLXIPL in Ins-1E cells. In the siMlxipl-treated Ins-1 832/13 cells, the transcript level of *Rorc* was indeed decreased (**TABLE C5**). Noordeen *et al.* had shown that siMlxipl treatment of Ins-1 832/13 cells increased the expression of the hypoxic stress gene *Arnt* after 20 h in high glucose (Noordeen *et al.,* 2010). In our experiment, we found that *rArnt* was downregulated in siMlxipl-treated cells in low glucose and after 6 h in high glucose (**TABLE C5**). It is possible that the direction of *rArnt* regulation by rMLXIPL changes with the incubation time in high glucose, meaning that rMLXIPL might repress *rArnt* in low glucose or after a short time in high glucose (6 h) whereas a prolonged exposure to high glucose (20 h) triggers the activation of *rArnt* by rMLXIPL. Likewise, we found that the glucose-mediated repression of the gene *Ppara* that encodes a master regulator of fatty acid metabolism known to protect the cell against lipid-induced beta cell function, was not affected in the siMlxipl Ins-1 832/13 cells in low glucose and after 6 h in high glucose (**TABLE C5**), although the overexpression of MLXIPL in Ins-1E cells, rat islets and human islets repressed the glucose-induced activation of *Ppara* after 24 h in high glucose (Boergesen *et al.,* 2011).

Taken together, our new RNA sequencing data confirm the regulation of many, but not all published RFX6 and MLXIPL target genes. The overlap indicates that we have generated data of a good quality that can be used for further analysis.

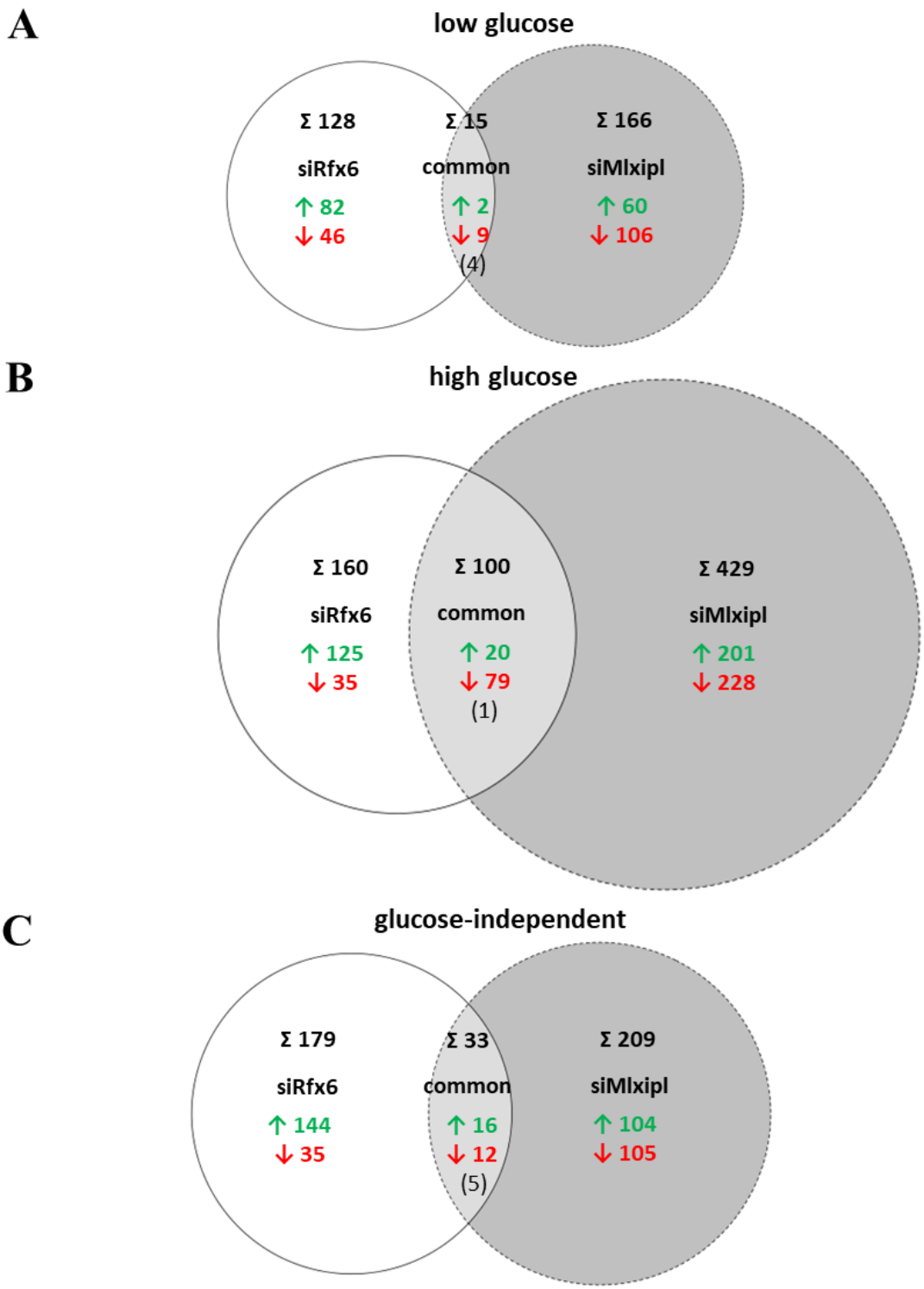


FIGURE C31: Glucose-dependent and -independent RFX6- and MLXIPL-specific and common targets in Ins-1 832/13 cells. Euler diagrams showing the genes differentially expressed at low glucose **(A)**, the genes differentially expressed at high glucose **(B)** and the differentially expressed glucose-independent genes **(C)** with -1>Log2 FC>1 and p<0.05 (see **FIGURE C29A** for more information about the samples). The number of regulated genes (Σ) is marked in black. Upregulated genes (↑) are marked in green and downregulated genes (↓) in red. The black numbers in parentheses represent those genes that are significantly affected in both datasets but not in the same direction. All diagrams were drawn with eulerr.co (Larsson, 2018. eulerr: Area-Proportional Euler and Venn Diagrams with Ellipses. R package version 4.1.0, https://cran.r-project.org/package=eulerr.).

2.5.6. Glucose-dependent and -independent rRFX6 and rMLXIPL targets in Ins-1 832/13 cells

In total, in our Ins-1 832/13 knockdown experiment, we sequenced the transcriptomes of eight different treatments in technical quadruplicates: control cells for the *rMlxipl* knockdown, siMlxipl-treated cells, control cells for the *rRfx6* knockdown and siRfx6-treated cells, all at low and high glucose. The use of two different controls was necessary because we used a SmartPool siRNA against *rRfx6* that contains four siRNAs, and the exon 12 siMlxipl that contains a single siRNA. Thus, we needed a control for the SmartPool siRNA with four different non-targeting siRNAs and a control for the single siRNA that was composed of a single non-targeting siRNA. Overall, the technical replicates were very similar. The two controls in low and high glucose clustered together, indicating that the two different control siRNAs cause little difference. The most strongly expressed genes in all data sets are mitochondrial genes, chromogranin A and B (*Chga*, *Chgb*), the eukaryotic translation elongation factor 2 (*Eef2*), the insulin-encoding gene *Ins1* and the ATP-citrate synthase *Acyl*.

To identify the rRFX6- and rMLXIPL-regulated genes, we compared the data of the siRfx6- and siMlxipl-treated cells with the respective controls either at low or at high glucose and we subdivided the differentially expressed genes into the following groups (**FIGURE C31**): i. RFX6-specific, MLXIPL-specific and common targets in low glucose; ii. RFX6-specific, MLXIPL-specific and common targets in high glucose; iii. glucose-independent RFX6-specific, MLXIPL-specific and common targets. Applying the parameters p<0.05 and -1>Log2 FC>1, we found 128 RFX6-specific, 166 MLXIPL-specific and 15 common genes at low glucose (**FIGURE C31A**) and 160 RFX6-specific, 429 MLXIPL-specific and 100 common genes at high glucose (**FIGURE C31B**), and 179 RFX6-specific, 209 MLXIPL-specific and 33 common genes whose regulation was glucose-independent (**FIGURE C31C**). **TABLE C6**, **TABLE C7** and **TABLE C8** show the Top 10 of the up- and downregulated genes of the different groups.

Overall, rRFX6 regulates the same number of target genes in low glucose, high glucose and both glucose concentrations (**FIGURE C31**) with the most upregulated gene being *G0s2*, *Ets1* and *Pax4* and the most downregulated genes being *Dnah12*, *Duox2* and *Balap3* or *Dnah9* in the three conditions (**TABLE C6**; **TABLE C7**; **TABLE C8**). Since nothing is known about whether the activity of RFX6 is somehow regulated by glucose, it is not surprising that there are no differences between low and high glucose. However, there are more genes upregulated than downregulated in the absence of *rRfx6* (**FIGURE C31**). This suggests that rRFX6 is rather a transcriptional inhibitor than a transcriptional activator. In fact, we had made similar observations in the mouse islets, where we also found more RFX6-activated than -repressed genes (unpublished data of Aline Meunier; Piccand *et al.,* 2014).

Unlike rRFX6, the activity of rMLXIPL is glucose-dependent as expected. The top upregulated genes were *Pknox2*, *Ptch2* and *Trim17* in low glucose, high glucose and in both glucose concentrations, respectively, and the top downregulated genes *Th*, *Islr* and *Pkp1* or *Ccnd1* (**TABLE C6**; **TABLE C7**;

TABLE C8). Most of the rMLXIPL-specific genes were found in high glucose (**FIGURE C31B**), while the number of glucose-independent rMLXIPL-specific genes balances with those regulated at low glucose (**FIGURE C31A+C**). The data demonstrate that MLXIPL both inhibits and activates equally its target genes. We can also deduce from our data that MLXIPL is more potent transcription factor than RFX6 because more genes were downregulated in the *rMlxipl* knockdown cells (**FIGURE C31**). Among the common targets in high glucose, we also found *Cx3cl1* and *Rasgrp2* (**TABLE C7**), two genes that were strongly downregulated in the siMlxipl-treated Ins-1E cells compared to siControl-treated Ins-1E cells after a 12-h glucose stimulus (GSE81628, Schmidt *et al.,* 2016). We had already assumed that they are rMLXIPL target genes (**FIGURE C23**) and the RNA sequencing data confirm that they are direct rMLXIPL targets downregulated in siRfx6- and siMlxipl-treated Ins-1 832/13 cells.

Overall, the number of common target genes is relatively low (**FIGURE C31**). We have the highest number of common target genes of RFX6 and MLXIPL in high glucose. This seems logical since MLXIPL is activated by glucose (Li *et al.,* 2006; Herman *et al.,* 2012). Most of the commonly regulated genes at low glucose were downregulated genes (**FIGURE C31**). It suggests that MLXIPL in high glucose might play the role of a downstream activator for RFX6.

In a knockout situation, we would have expected that all the deregulated genes in the *Mlxipl* -/- cells in the *Rfx6* -/- cells would be deregulated by the complete omission of MXLIPL in *Rfx6* -/- and in *Mlxipl* -/- cells. Moreover, in the *Rfx6* -/- cells it would have to have other deregulated genes that are not regulated by *Mlxipl* but are direct RFX6 target genes or regulated by other transcription factors. However, in our knockdown situation, where we were only able to reduce both RFX6 and MLXIPL by 50 %, the residual activity of both proteins results in the numbers not meeting expectations.

To validate our data, we compared the regulated genes with our islet anti-RFX6 ChIP sequencing data (Perrine Strasser, unpublished data) and the published Ins-1E anti-MLXIPL ChIP sequencing data (GSE81628, Schmidt *et al.,* 2016). Among the RFX6-specific genes, 30 % were bound by RFX6 regardless of the glucose level (**FIGURE C32**). For rMLXIPL-specific genes, on the other hand, we noticed a clear difference between low and high glucose: in low glucose, only 7 % of the genes are direct MLXIPL target genes, while 50 % of the deregulated genes are bound by MLXIPL in high glucose (**FIGURE C32**). Consistently, we had also found more MLXIPL-specific genes in high glucose than in low glucose (**FIGURE C31**). Schmidt *et al.* claimed in their study that MLXIPL does not bind to glucose-repressed genes (Schmidt *et al.,* 2016). We cannot confirm this in our data since 200 of the 307 genes that were inhibited in high-glucose in siMlxipl-treated cells, had a MLXIPL peak. Since we used their ChIP sequencing data to identify MLXIPL-bound genes, the different conclusions must rather be caused by differences in RNA sequencing that are probably due to the different incubation times, than by differences in the evaluation of ChIP sequencing data. In our opinion, MLXIPL is both an activator and a repressor. Our RNA sequencing data support this conclusion, as there was no preference neither

(**FIGURE C31**).

A quarter of the common target genes that were deregulated in siRfx6 and siMlxipl-treated cells in high glucose or regulated independently of glucose levels were bound by both transcription factors (**FIGURE C32**). Almost all common target genes bound by RFX6 and MLXIPL, apart from a single gene (*Btbd17*), were synergistically activated or inhibited.

We also identified 24 genes, that were regulated by RFX6 in low glucose and by MLXIPL in high glucose (e.g. *G0s2, Ahrr*, *Tenm4*, *Dpysl3*, *Igdcc4*, *Ihh*, *Unc5c*, *Car8*, **TABLE C6**) but only 7 genes that were regulated by MLXIPL in low glucose and by RFX6 in high glucose (*Fam149a*, *Slc26a7*, *Lhfp26*, *Slc17a3*, *Edil3*, *Rnf128*, *Th*, **TABLE C6**; **TABLE C7**). MLXIPL thus seems to take over RFX6-regulated genes in high glucose concentrations.

In summary, the results in knockout cells would probably have been clearer. Nevertheless, we were able to show that the activity of RFX6 is not affected by glucose and that it is more of a transcriptional repressor, whereas MLXIPL is activated by glucose and acts equally as transcriptional activator and repressor.

2.5.7. Could the reduction of *rPklr* in siMlxipl-treated Ins-1 832/13 cells cause cell death?

From the RNA sequencing data sets, we tried to derive signalling pathways that were particularly affected in the different samples. We performed a gene ontology analysis using the Reactome Pathway Browser (https://reactome.org/), but this analysis did not give us any further evidence of particularly affected signalling pathways. Since most genes are involved not only in one signal pathway, but in several, it is difficult to draw conclusions from the lists of deregulated genes on affected signalling pathways. It is easier to explain a phenotype by means of the RNA sequencing data. The only phenotype we had noticed, was the high cell death in siMlxipl treated cells, which was more potent in low glucose than in normal medium. One might suspect that siMlxipl-treated cells may have difficulties producing enough energy in low glucose medium, while the higher glucose concentration in the standard medium might has a mitigating effect. Iizuka *et al.* had shown that the knockout of *mMlxipl* affects the glycolysis rate in the mouse liver because the expression of the glucose transporter GLUT2 and the liver pyruvate kinase *mPklr* were decreased due to the loss of *mMlxipl* (Iizuka *et al.,* 2004). Indeed, we observed that the *rMlxipl* knockdown lowered the expression of *rPklr* in Ins-1 832/13 cells in both low and high glucose (**TABLE C5**). In siRfx6-treated cells in which we did not observe any increased cell death, *rPklr* is only affected in high glucose but not in low glucose (**TABLE C5**). The reduction of PKLR in siMlxipl-treated cells might be a limiting factor in the energy production especially in low glucose and might thus explain why we had this cell toxic effect after the siRNA treatment.

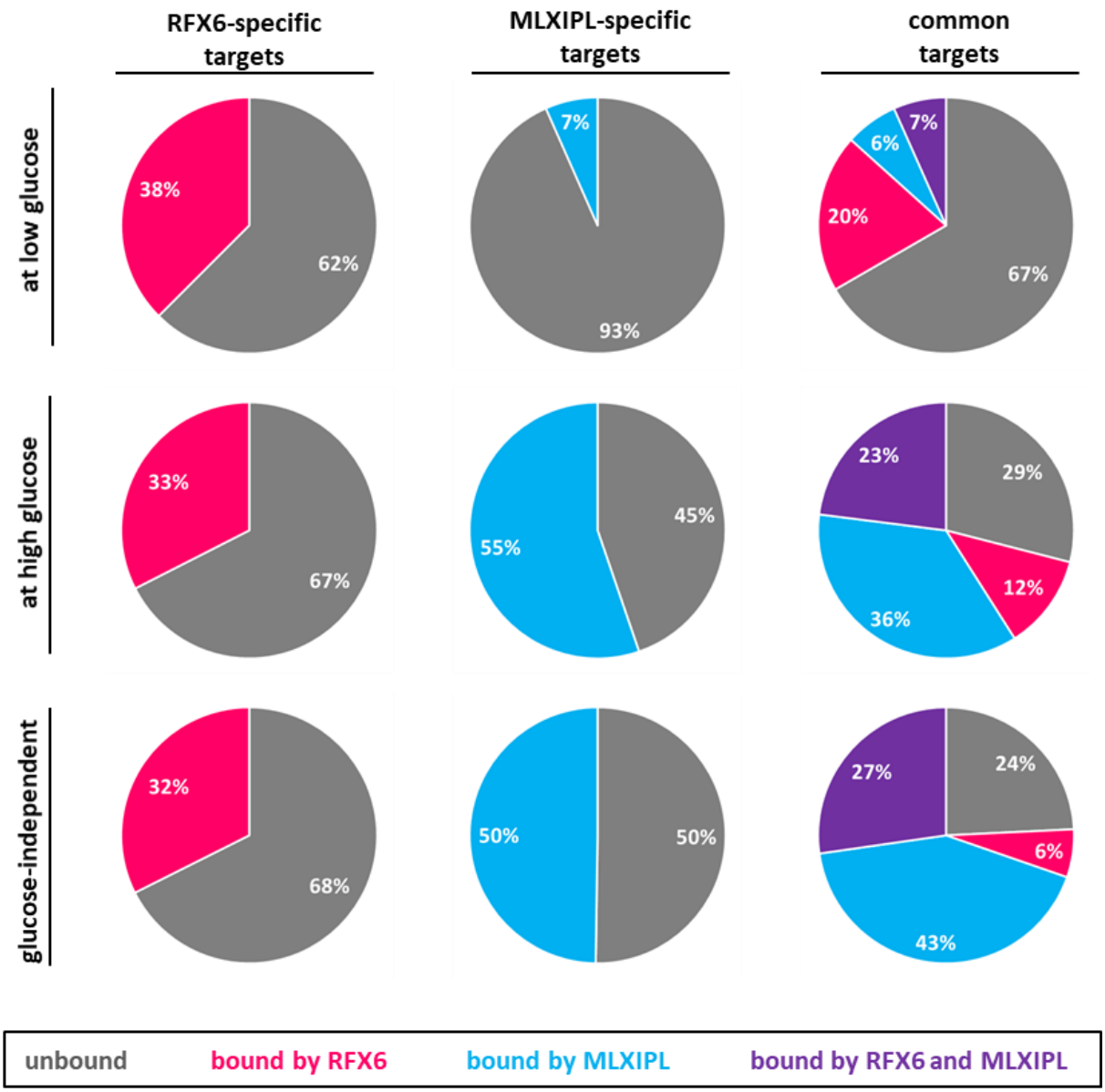


FIGURE C32: Direct target genes as a percentage of the RFX6- and MLXIPL-regulated genes in Ins-1 832/13 cells. Pie charts illustrating the percentage of genes bound by RFX6 (in pink), MLXIPL (in blue), both (in purple) or none (in grey) among the genes regulated by RFX6, MLXIPL or by both only at low glucose, only at high glucose or under both glucose conditions in Ins-1 832/13 cells (with -1>Log2 FC>1 and p<0.05). To produce the pie charts, lists of the regulated genes identified in the RNA sequencing (see **FIGURE C29A** for more information about the samples) have been compared with the anti-RFX6 islet ChIP sequencing data (unpublished data of Perrine Strasser) and the anti-MLXIPL Ins-1E ChIP data set (GSE81628, Schmidt *et al.*, 2016). The latter contains three sets of data. Genes regulated in low glucose have been compared with the ChIP profile after 24 h starvation in 5 mM glucose, genes regulated in high glucose with the combined data of the profiles after an incubation in 25 mM glucose for 2 h and 12 h, and for glucose-independent genes, all three data sets have been considered.

TABLE C6: Top 10 of the RFX6-specific, MLXIPL-specific and common targets in Ins-1 832/13 cells at low glucose. Tables showing the ten most upregulated (green, Log2 FC>1 and p<0.05) and downregulated (red, Log2 FC<-1 and p<0.05) genes in the siRfx6-treated (above on the left) or in siMlxipl-treated (above on the right) cells or in both knockdowns (below, the table shows the total number of common genes at low glucose) at 2 mM (low) glucose identified by RNA sequencing. A gene name's case filled in pink, in blue and in purple indicates a rRFX6-bound gene, rMLXIPL-bound gene and a gen bound by both transcription factors, respectively.

gene name	normalized expression siControl - low glucose	normalized expression siRfx6 - low glucose	Log2 FC	adjusted p-value
G0s2	126	567	2,11	1,86E-137
Ahrr	6	25	1,60	2,58E-09
Tenm4	63	198	1,57	9,61E-40
Dpysl3	309	841	1,41	8,82E-65
Igdcc4	39	110	1,39	4,77E-20
Kcna6	24	67	1,37	1,80E-12
Fam102b	21	60	1,33	2,28E-10
Lin28a	11	33	1,32	1,25E-06
Pde7b	9	28	1,31	1,65E-06
Ihh	2	13	1,30	2,20E-05
Unc5c	2194	911	-1,26	4,42E-81
Abhd15	151	60	-1,28	7,43E-20
Eif4e3	1129	461	-1,28	6,68E-125
Car8	2086	851	-1,29	9,30E-174
Ptk2b	2621	1052	-1,31	7,53E-297
Wdr49	16	3	-1,37	1,34E-05
Gch1	2220	847	-1,37	1,80E-77
Ttll10	37	10	-1,39	1,63E-06
Ccdc129	30	7	-1,49	4,84E-09
Dnah12	101	30	-1,64	1,45E-19

gene name	normalized expression siControl - low glucose	normalized expression siMlxipl - low glucose	Log2 FC	adjusted p-value
Pknox2	73	351	2,17	1,73E-89
Inhbe	1	14	1,86	8,26E-09
Kcnj12	12	55	1,79	1,86E-15
Skor1	13	59	1,64	3,98E-10
Cabp7	62	198	1,61	2,87E-31
Ephx4	6	32	1,55	9,78E-08
Grb10	33	106	1,53	5,09E-12
Phtf2	261	719	1,44	3,33E-83
Syt12	47	132	1,41	5,34E-22
Spdya	4	17	1,37	8,11E-06
Gnal	1732	625	-1,44	1,15E-79
Jhy	135	45	-1,48	8,92E-24
Slc17a3	75	24	-1,50	1,10E-14
B3galt5	11	1	-1,53	5,29E-07
Aldh3b1	16	3	-1,57	5,48E-07
Sec1	32	7	-1,57	1,81E-08
Corin	37	8	-1,66	1,16E-11
Nrp2	430	133	-1,67	1,75E-59
Myc	1335	345	-1,92	1,08E-192
Th	35	5	-1,93	4,68E-11

gene name	normalized expression siControl - low glucose	normalized expression siRfx6 - low glucose	Log2 FC	adjusted p-value	normalized expression siControl - low glucose	normalized expression siMlxipl - low glucose	Log2 FC	adjusted p-value
Mycn	14	49	1,43	1,56E-08	12	32	1,11	2,71E-05
Baiap2l2	9	30	1,30	1,39E-05	13	4	-1,02	1,43E-03
Cck	124	328	1,18	2,64E-06	191	54	-1,46	4,62E-09
Dtx4	9	24	1,10	4,84E-05	14	34	1,07	2,19E-05
Kif26b	19	44	1,09	2,01E-06	24	6	-1,35	2,80E-07
Cacna2d4	3	11	1,01	3,68E-03	11	2	-1,06	1,57E-03
Krt80	28	11	-1,10	2,48E-05	41	16	-1,23	3,40E-07
Pdzd3	24	9	-1,14	4,25E-05	48	14	-1,62	6,28E-11
Hr	200	87	-1,16	3,71E-22	278	136	-1,00	9,44E-20
Crip3	21	7	-1,19	3,82E-05	25	10	-1,08	1,04E-04
Lynx1	556	226	-1,28	1,35E-46	689	319	-1,08	9,21E-37
Cxcl12	49	17	-1,29	4,42E-08	60	22	-1,27	1,85E-08
Dysf	17	3	-1,49	3,51E-06	17	5	-1,08	8,28E-04

TABLE C7: Top 10 of the RFX6-specific, MLXIPL-specific and common targets in Ins-1 832/13 cells at high glucose. Tables showing the ten most upregulated (green, Log2 FC>1 and p<0.05) and downregulated (red, Log2 FC<-1 and p<0.05) genes in the siRfx6-treated (above on the left) or in siMlxipl-treated (above on the right) cells or in both knockdowns (below) at 20 mM (high) glucose identified by RNA sequencing. A gene name's case filled in pink, in blue and in purple indicates a rRFX6-bound gene, rMLXIPL-bound gene and a gen bound by both transcription factors, respectively.

gene name	normalized expression siControl - high glucose	normalized expression siRfx6 - high glucose	Log2 FC	adjusted p-value
Ets1	74	320	2,04	1,75E-83
Mx2	42	161	1,83	4,43E-38
Boc	7	29	1,63	8,69E-10
Tmed6	2	14	1,57	1,60E-06
Ddah2	202	577	1,47	1,74E-64
Slitrk2	9	34	1,45	1,71E-07
Fam149a	13	45	1,45	1,52E-09
Glrb	16	50	1,43	4,02E-11
Kmt2a	7	26	1,40	3,43E-07
Fcgbp	19	59	1,40	2,52E-11
Lypd1	184	76	-1,22	2,63E-24
Fstl4	1686	712	-1,23	3,24E-71
Th	16	3	-1,27	5,66E-05
Klf2	41	14	-1,29	5,47E-06
Cdc42ep5	30	10	-1,30	1,18E-05
Foxn4	170	66	-1,34	3,67E-21
Il10ra	15	3	-1,38	1,87E-05
Slc2a4	947	312	-1,58	5,85E-55
Col5a2	11	1	-1,58	3,36E-07
Duox2	24	3	-1,77	4,78E-09

gene name	normalized expression siControl - high glucose	normalized expression siMlxipl - high glucose	Log2 FC	adjusted p-value
Ptch2	27	124	1,99	4,28E-31
Cxxc4	317	1103	1,78	6,45E-151
Mnx1	107	361	1,67	4,67E-39
Synpo2	678	2195	1,67	1,72E-113
Notch1	211	670	1,64	1,09E-105
Epha5	11	43	1,63	3,73E-12
Slc16a10	320	1001	1,63	2,96E-122
Amigo2	1103	3337	1,59	1,08E-298
Adgrl3	83	259	1,59	4,07E-43
Pclo	1144	3496	1,57	8,21E-56
Cabp2	14	0	-1,95	1,90E-12
Trim29	13	0	-1,98	8,42E-12
Gdpd5	323	76	-1,99	1,60E-37
Fbln2	71	13	-2,00	2,80E-14
Sv2c	241	53	-2,03	2,92E-26
Cldn15	633	150	-2,05	3,48E-117
Gna15	22	3	-2,05	7,87E-11
G6pc	18	2	-2,10	1,38E-11
Grhl2	59	9	-2,13	2,78E-14
Islr	24	1	-2,34	3,41E-14

gene name	normalized expression siControl - high glucose	normalized expression siRfx6 - high glucose	Log2 FC	adjusted p-value	normalized expression siControl - high glucose	normalized expression siMlxipl - high glucose	Log2 FC	adjusted p-value
Tmem173	38	132	1,69	6,65E-32	75	197	1,33	6,62E-30
Dusp10	116	339	1,51	3,52E-66	31	80	1,27	2,06E-14
Nuak2	20	61	1,45	2,28E-14	26	78	1,39	2,67E-12
Pik3r3	280	760	1,41	4,79E-74	6	21	1,29	1,19E-05
Lrrn1	90	239	1,36	4,92E-37	91	293	1,63	7,43E-64
Cacng6	22	62	1,30	2,96E-10	18	44	1,10	5,32E-07
Vdr	690	1703	1,29	8,12E-139	25	58	1,09	5,69E-09
Pcdh9	28	75	1,26	9,71E-13	14	33	1,10	5,67E-06
Adcy8	118	286	1,25	5,58E-34	91	221	1,23	1,92E-29
Sytl4	891	2111	1,23	4,57E-102	9	59	2,27	4,32E-26
Fam212a	190	53	-1,77	4,86E-26	177	25	-2,70	5,94E-43
Cped1	34	7	-1,77	1,48E-09	247	65	-1,89	1,10E-36
Slc12a4	804	230	-1,78	7,75E-78	63	27	-1,16	1,66E-06
Prok2	512	121	-1,78	3,41E-12	1233	459	-1,42	1,37E-85
Lcat	139	37	-1,79	2,63E-24	71	20	-1,74	2,43E-14
Lamc3	32	7	-1,86	4,60E-11	8705	4050	-1,10	1,80E-85
Rasgrp2	594	155	-1,90	1,01E-57	353	73	-2,16	8,42E-35
Acer1	199	49	-1,96	3,19E-34	623	237	-1,37	6,63E-52
Cx3cl1	1147	196	-2,42	1,61E-56	8927	2997	-1,56	5,95E-133
Il2rb	235	40	-2,50	5,26E-60	167	50	-1,67	2,44E-21

TABLE C8: Top 10 of the RFX6-specific, MLXIPL-specific and common glucose-independent targets in Ins-1 832/13 cells. Tables showing the ten most upregulated (green, Log2 FC>1 and p<0.05) and downregulated (red, Log2 FC<-1 and p<0.05) genes in the siRfx6-treated (above on the left) or in siMlxipl-treated (in the middle on the right) cells or in both knockdowns (below) at 2 mM (low) and 20 mM (high) glucose identified by RNA sequencing. A gene name's case filled in pink, in blue and in purple indicates a rRFX6-bound gene, rMLXIPL-bound gene and a gen bound by both transcription factors, respectively.

gene name	normalized expression siControl - low glucose	normalized expression siRfx6 - low glucose	Log2 FC	adjusted p-value	normalized expression siControl - high glucose	normalized expression siRfx6 - high glucose	Log2 FC	adjusted p-value
Pax4	6	358	5,34	6,06E-143	4	279	5,28	7,40E-128
Thbs1	2	136	4,74	1,65E-79	2	166	5,09	4,03E-91
Gulo	3	57	3,44	1,59E-33	1	46	3,47	9,40E-33
Gstp1	1	27	3,03	1,01E-22	0	23	2,86	3,52E-20
Bmf	47	409	3,02	9,63E-132	87	450	2,32	2,22E-100
Tmem270	1	33	2,99	1,05E-21	2	21	2,11	4,15E-11
Zfp385c	5	51	2,78	8,40E-23	2	35	2,70	3,44E-20
Gstm1	87	577	2,68	7,68E-139	66	520	2,90	9,81E-145
Plg	6	50	2,45	1,13E-20	15	51	1,53	9,99E-10
Gfpt2	5	40	2,43	1,50E-17	2	32	2,81	2,07E-21
Mageh1	642	183	-1,79	9,33E-110	320	122	-1,33	6,89E-44
Apod	26	4	-1,88	8,08E-12	14	3	-1,25	2,66E-05
Drc1	133	33	-1,89	9,07E-35	100	31	-1,52	2,75E-20
Slc2a5	149	36	-1,90	1,42E-38	98	34	-1,39	4,94E-18
Trank1	206	51	-1,92	1,33E-47	84	35	-1,14	2,28E-12
Mtnr1a	133	32	-1,93	1,42E-28	24	7	-1,17	3,22E-06
Dcdc5	46	8	-1,94	2,58E-16	41	6	-1,92	6,93E-15
Pnma8a	222	50	-2,05	8,44E-66	147	49	-1,47	2,20E-29
Dnah9	2526	531	-2,22	8,84E-220	2194	470	-2,19	3,26E-199
Baiap3	4865	987	-2,28	0,00E+00	3015	672	-2,15	0,00E+00

gene name	normalized expression siControl - low glucose	normalized expression siMlxipl - low glucose	Log2 FC	adjusted p-value	normalized expression siControl - high glucose	normalized expression siMlxipl - high glucose	Log2 FC	adjusted p-value
Trim17	20	244	3,23	4,11E-56	15	159	2,94	4,56E-43
Hist1h1c	207	1735	2,75	1,18E-43	129	996	2,62	4,77E-39
Kcns2	5	49	2,53	1,04E-19	5	27	1,68	8,08E-09
Rasgef1c	36	214	2,39	1,16E-57	65	190	1,48	6,10E-26
Ascl1	54	300	2,36	7,84E-62	29	186	2,51	1,60E-53
Hist1h2bk	9	61	2,33	2,70E-21	7	56	2,43	1,37E-21
Kremen1	124	622	2,31	7,28E-155	112	482	2,05	8,58E-111
Hist2h4a	7	47	2,31	3,14E-17	4	48	2,67	1,27E-21
Wif1	2334	11043	2,23	0,00E+00	1861	9339	2,31	0,00E+00
Il20ra	19	97	2,16	1,00E-29	15	73	1,99	1,79E-22
Ramp1	22	2	-1,94	1,64E-11	65	13	-2,03	6,66E-16
Bhlhe40	1161	265	-2,00	1,40E-46	5313	2237	-1,19	8,43E-19
Tmem51	668	158	-2,03	2,16E-124	1360	332	-2,01	1,50E-158
Ambp	3463	770	-2,10	6,04E-89	5536	2319	-1,22	6,71E-32
Fam205a	35	5	-2,11	4,23E-15	20	5	-1,29	6,09E-06
Ccnd1	533	116	-2,12	7,84E-126	2147	403	-2,39	1,82E-273
S100a4	311	48	-2,54	4,51E-91	564	124	-2,12	7,59E-94
Cpne9	98	10	-2,59	1,13E-44	518	134	-1,91	6,87E-76
Plcxd3	829	107	-2,81	4,77E-124	2462	832	-1,52	2,95E-50
Pkp1	43	2	-2,85	4,45E-22	21	2	-1,98	9,04E-11

gene name	normalized expression siControl - low glucose	normalized expression siRfx6 - low glucose	Log2 FC	adjusted p-value	normalized expression siControl - high glucose	normalized expression siRfx6 - high glucose	Log2 FC	adjusted p-value	normalized expression siControl - low glucose	normalized expression siMlxipl - low glucose	Log2 FC	adjusted p-value	normalized expression siControl - high glucose	normalized expression siMlxipl - high glucose	Log2 FC	adjusted p-value
F3	158	834	2,37	2,14E-191	60	495	2,94	4,12E-178	22	52	1,07	2,78E-08	13	43	1,38	1,06E-10
Cyp1a1	26	135	2,29	2,30E-38	18	114	2,48	3,72E-39	7	41	2,05	7,69E-16	5	34	1,99	6,09E-14
Bmp2	34	156	2,09	5,01E-39	23	134	2,37	7,45E-43	36	260	2,68	5,59E-61	9	214	3,85	3,14E-83
Ins2	1013	4252	2,04	9,38E-160	662	3282	2,27	2,27E-189	58	178	1,57	4,80E-38	31	159	2,22	1,43E-53
Gria2	38	164	1,97	1,45E-35	39	147	1,79	4,24E-29	202	801	1,97	2,69E-141	53	637	3,45	1,13E-223
Cckar	31	123	1,85	8,96E-27	11	90	2,53	4,60E-36	50	126	1,26	7,94E-17	19	111	2,28	3,05E-37
Rd3	31	112	1,75	4,71E-26	33	119	1,72	5,97E-26	28	146	2,25	1,21E-37	12	151	3,14	1,52E-57
Btbd17	34	112	1,63	1,53E-21	20	69	1,63	4,94E-17	32	147	2,04	3,22E-35	37	142	1,82	7,67E-29
Chrm3	72	220	1,56	6,96E-43	40	180	2,07	1,66E-55	1289	10575	3,00	0,00E+00	684	8375	3,56	0,00E+00
Gad2	41	130	1,55	2,37E-20	17	80	1,95	1,08E-23	245	1142	2,18	3,77E-173	182	1106	2,55	1,57E-212
Camk2a	149	66	-1,12	1,97E-20	595	227	-1,37	2,91E-54	394	91	-2,01	1,54E-47	305	89	-1,70	1,73E-32
Lgi3	23	7	-1,13	2,00E-06	235	51	-2,14	2,80E-54	38	12	-1,28	1,24E-09	268	105	-1,33	1,78E-28
S100a5	852	377	-1,15	4,57E-44	1028	419	-1,27	7,94E-54	38	10	-1,42	2,09E-11	409	112	-1,86	2,04E-58
Adamts8	28	8	-1,16	4,33E-07	326	64	-2,29	6,41E-62	76	18	-1,86	3,06E-22	49	12	-1,59	1,34E-13
Galnt9	4868	1994	-1,28	1,45E-205	12547	4314	-1,54	0,00E+00	2625	561	-2,20	3,36E-224	1827	601	-1,58	6,17E-113
Gas7	118	31	-1,71	1,75E-32	1398	301	-2,19	9,69E-150	180	18	-2,94	2,37E-60	155	34	-2,05	7,35E-34
S100a3	73	18	-1,74	6,56E-20	106	26	-1,87	3,87E-25	912	190	-2,20	8,67E-134	1039	295	-1,78	4,14E-96
Gast	209	47	-2,01	6,74E-39	174	36	-2,07	1,25E-35	871	280	-1,60	9,22E-63	823	261	-1,61	3,11E-61
Rnd1	1543	345	-2,13	7,17E-188	1356	395	-1,76	3,67E-125	121	20	-2,41	7,95E-32	31	13	-1,00	5,34E-05
Slc6a19	62	8	-2,42	4,66E-25	24	4	-1,48	3,38E-08	277	68	-1,92	5,14E-67	2207	695	-1,65	4,01E-123

2.5.8. Insulin transcription is regulated by rRFX6 and rMLXIPL in Ins-1 832/13 cells

The most important common target of rRFX6 and rMLXIPL might be insulin. We found a strongly increased expression of *rIns1* and *rIns2* in the siRfx6- and siMlxipl-treated cells in both low and high glucose conditions (**TABLE C5**; **TABLE C8**). Furthermore, using the ChIP sequencing data of Schmidt *et al.,* we found that MLXIPL binds to both *rIns1* and *rIns2* (Schmidt *et al.,* 2016, **FIGURE C33A+B**) while RFX6 binds to *mIns2* in murine islets (unpublished data of Perrine Strasser, **FIGURE C33C**). The expression of *rIns2* seems to be synergistically regulated by RFX6 and MLXIPL, whereas the transcription of *rIns1* is only dependent on MLXIPL. To our knowledge, this is the first time that a direct regulation of insulin transcription by RFX6 and MLXIPL is described.

In humans, it has been shown that the transcription of insulin was reduced in siRfx6-treated EndoC-betaH2 cells (Chandra *et al.,* 2014). In *mRfx6* Δbeta mice, the expression of *mIns1* was reduced while *mIns2* expression remained unchanged. The insulin content was not altered, and the decreased insulin was likely due to a deregulation of *Gck*, *Abcc8* and calcium channels (Piccand *et al.,* 2014). Regarding MLXIPL, the plasma insulin levels of *mMlxipl* -/- mice are normal (Iizuka *et al.,* 2004). However, the treatment of Min6 with siMlxipl slightly increased *mIns2* levels (Da Silva Xavier *et al.,* 2010). The same study reported that the overexpression of MLXIPL in murine islets caused a decreased expression of *mIns1* and *mIns2* at low glucose but not at high glucose, suggesting that MLXIPL inhibits insulin transcription in low glucose. Indeed, the insulin secretion in siMlxipl-treated Min6 cells was increased (Da Silva Xavier *et al.,* 2006), suggesting that MLXIPL inhibits not only insulin transcription but also insulin secretion.

Our data demonstrate that *rIns1* is a directly repressed by rMLXIPL and that rRFX6 and rMLXIPL co-repress the transcription of *rIns2* in Ins-1 832/13 cells. Further experiments are needed to elucidate the regulation of insulin transcription at different time points after the glucose stimulus and to find out whether the knockdown of *rRfx6* and *rMlxipl* affect insulin content and insulin secretion not only in Ins-1 832/13 cells, in other cell lines and in islets.

2.5.9. The glucose response is impaired in siRfx6- and siMlxipl-treated Ins-1 832/13 cells

Finally, we compared the number of glucose-activated and glucose-repressed genes in siControl-, siRfx6- and siMlxipl-treated cells to find out whether rRFX6 and / or rMLXIPL are important for the transcriptional adaptation to glucose in Ins-1 832/13 cells (**FIGURE C34**).

We found 498 and 583 glucose-induced genes and 184 and 260 glucose-repressed genes ($p<0.05$, $-1>\text{Log2 FC}>1$) in the cells transfected with the SmartPool control and the single siRNA control, respectively (**FIGURE C34A+B**). Thus, approximately 200 genes more are affected in the cells treated with the single siRNA control, which is a large difference. In retrospect, we should better have

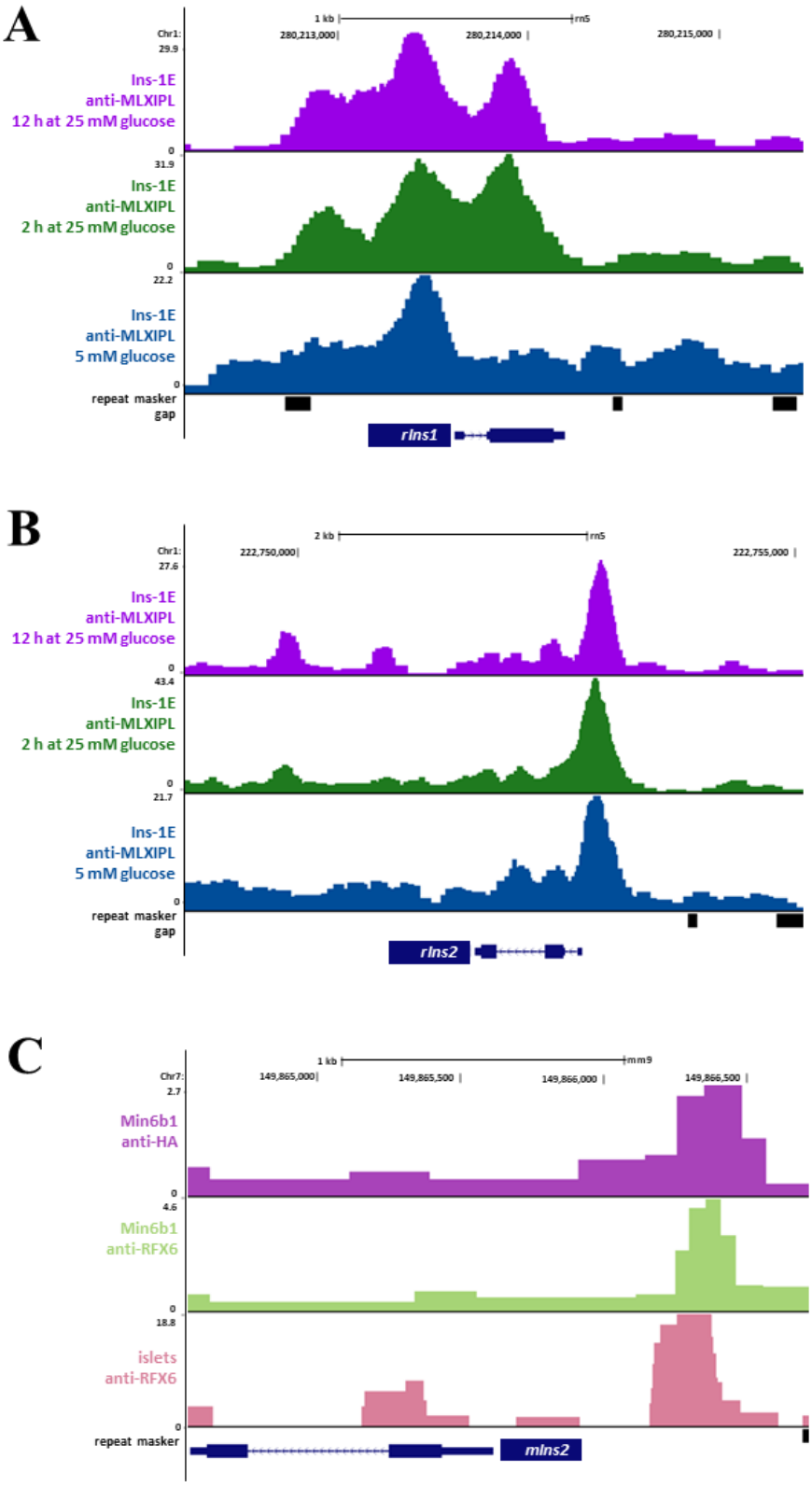


FIGURE C33: Binding of MLXIPL and RFX6 on *Ins1* and *Ins2*. (A, B) ChIP sequencing data (GSE81628, Schmidt *et al.,* 2016) showing the binding of rMLXIPL on *rIns1* **(A)** and *rIns2* **(B)** in Ins-1E cells starved for 24 h in 5 mM glucose (blue profile), as well as after a subsequent incubation in 25 mM glucose for 2 h (green profile) or 12 h (purple profile). UCSC alignment (rn5) was done by Constance Vagne. **(C)** ChIP sequencing data showing the binding of RFX6 on *mIns2* (anti-HA ChIP on Min6b1 expressing 3HA-mRFX6, published in Piccand *et al.,* 2014; anti-RFX6 ChIP on Min6b1 and anti RFX6 ChIP on islets of CD1 mice, unpublished data of Perrine Strasser). UCSC genome browser alignment (mm9) done by Tao Ye.

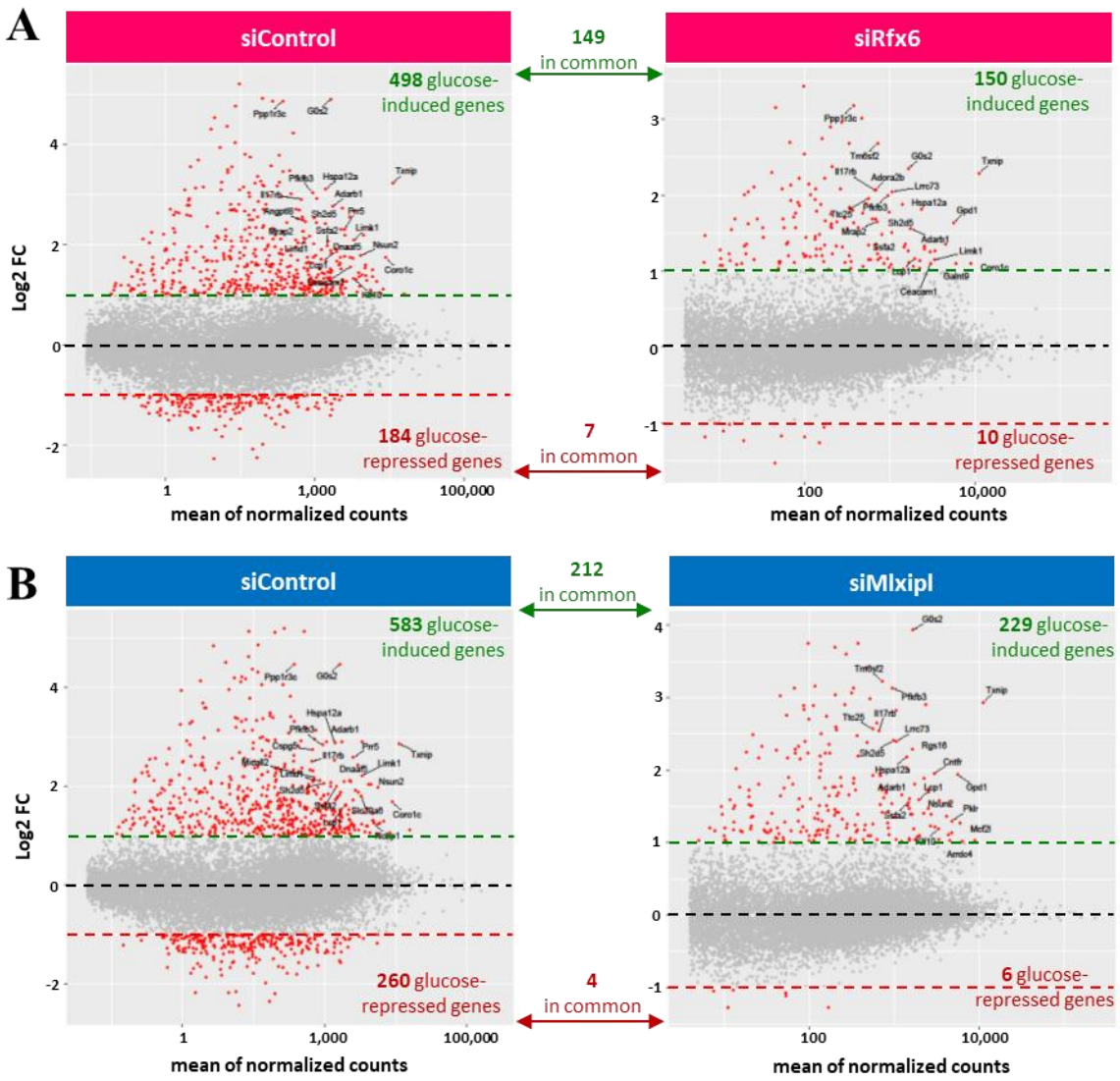


FIGURE C34: The loss of *rRfx6* and *rMlxipl* reduces the number of glucose-induced and -repressed genes in Ins-1 832/13 cells. MA plots representing the Log2 FC (high glucose versus low glucose) as a function of the mean of normalized counts for the following RNA sequencing samples: Ins-1 832/13 siControl and siRfx6 (SmartPool at 11 nM) in 2 or 20 mM glucose **(A)** and Ins-1 832/13 siControl and siMlxipl (exon 12 siRNA and corresponding control at 2.75 nM) at 2 mM (low) or 20 mM (high) glucose **(B)**. See **FIGURE C29A** for more information about the samples. Grey dots (●) are not significantly affected genes (p≥0.05 and/or -1≤Log2 FC≤1) and red dots (●) significantly upregulated (p<0.05 and Log2 FC>1) or downregulated (p<0.05 and Log2 FC<-1) genes. The analysis was done by and the figures were drawn by Constance Vagne.

used the SmartPool siRNA against *rRfx6* and *rMlxipl* to have only one control and to minimize the background of the siRNA treatment. Furthermore, it seems that there are more genes induced by glucose than repressed by glucose. However, when we used a less stringent filter (p<0.05), we got 3461 and 3649 upregulated and 3529 and 3568 downregulated genes for the *rRfx6* control and the *rMlxipl* control, respectively. The number of glucose-activated and repressed genes seems to be balanced, with the overall activation of the genes being greater than their repression.

In non-transfected Ins-1E cells, Schmidt *et al.* found 2042 glucose-induced and 1700 glucose-repressed genes after 12 h in high glucose with a higher number of strongly activated genes than of strongly repressed genes ($p<0.05$; Schmidt *et al.,* 2016). In their study, they also analysed the gene expression changes after 1 h, 2 h and 4 h in high glucose and they reported a significantly higher number of deregulated genes after 12 h in high glucose than after 4 h in high glucose. We incubated the Ins-1 832/13 cells for 6 h in high glucose, thus, our time point is in-between their 4-h and 12-h time point. Their observation suggests that there is a veritable wave of transcriptional changes between the 4-hour time point and the 12-hour time point, and we might be amid this wave with our 6-hour time point.

In the siRfx6- and siMlxipl-treated Ins-1 832/13 cells compared to siControl-treated Ins-1 832/13 cells, we found significantly less glucose-induced and -repressed genes ($p<0.05$, $-1>\text{Log2 FC}>1$; **FIGURE C34A+B**). There were 150 and 229 glucose-induced genes and 10 and 6 glucose-repressed genes in the siRfx6- and siMlxipl-treated cells, respectively (**FIGURE C34A+B**). That means that we lost 70 % and 60 % of the glucose-activated and 95 % and 98 % of the glucose-repressed genes via the knockdown of *rRfx6* and *rMlxipl*, respectively. All the genes that were still significantly deregulated in the knockdown cells, could also be found in the control cells, albeit some of same with a slightly reduced fold change below $|1|$.

Since we observed the decrease in glucose-regulated genes in both *Rfx6* experiments as well as in the siMlxipl-treated Ins-1 832/13 cells, and since we know that *Mlxipl* is a target gene of RFX6 in the mouse and rat, we hypothesized that the decreased glucose response in all cases is due to the reduction of MLXIPL. In addition, MLXIPL is a glucose-activated transcription factor (Li *et al.,* 2006; Herman *et al.,* 2012), suggesting that it is involved in the glucose response. In conclusion, the role of MLXIPL downstream of RFX6 is to mediate transcriptional changes following a glucose stimulus.

2.6. Conclusion of part 2 and graphical summary

In this second part of the thesis, we wanted to elucidate the transcriptional networks regulated by RFX6 and MLXIPL in adult pancreatic beta cells and to attribute clear roles to RFX6 and its downstream target MLXIPL. We had tried three different strategies to achieve this goal: a knockout of *mRfx6* and *mMlxipl* in the murine pancreatic beta cell line Min6b1, a knockout of *rRfx6* and *rMlxipl* in the rat pancreatic beta cell line Ins-1 832/13 and a knockdown of these two genes in the same cell line. Unfortunately, neither our attempts to generate Min6b1 knockout cell lines nor the attempts to produce Ins-1 832/13 knockout clones were successful. In the Min6b1 cells, the results of the functional analysis revealed that known mRFX6 and mMLXIPL target genes were not affected by the knockout, thus we decided not to work with these cells. In Ins-1 832/13 cells, we only obtained

heterozygous or heterogenous clones. We finally downregulated *rRfx6* and *rMlxipl* by RNA interference in the Ins-1 832/13 cell line and sequenced the transcriptome of knockdown cells that have been incubated for 6 h in low or high glucose medium.

The RNA sequencing data revealed that rMLXIPL is important for transcriptional adaptation to high glucose concentrations, as it was severely impaired in the *rRfx6* and *rMlxipl* knockdown cells. Furthermore, we found that RFX6 is more of a transcriptional repressor than an activator, whereas rMLXIPL equally activates and inhibits its target genes and seems to be a more potent transcriptional factor in high glucose than rRFX6 (**graphical summary - FIGURE C35**). The activity of rRFX6 was not altered by glucose. Target genes of both transcription factors could be found in almost all pathways, suggesting that both proteins have a relatively broad range of roles in Ins-1 832/13 cells. The number of common targets was relatively low. However, it must be said that our knockdown efficiency was not optimal; it only halved the expression of both genes. In knockout cells, the overlap would probably have been significantly greater, since MLXIPL is switched off in both the *Mlxipl* -/- cells and *Rfx6* -/- cells.

The most prominent common targets of both genes are the insulin encoding genes *rIns1* and *rIns2* (**graphical summary - FIGURE C35**) that were both strongly upregulated in the *rRfx6* and *rMlxipl* knockdown cells at low and high glucose. While *rIns2* is a direct target of both transcription factors, *rIns1* is only bound by rMLXIPL, thus its expression is regulated via rMLXIPL. The direct repression of *rIns2* by rRFX6 and rMLXIPL has not been reported yet and suggests an important involvement of rRFX6 and rMLXIPL in the control of the insulin content in Ins-1 832/13 cells.

Our project has further allowed us to demonstrate that the relation between RFX6 and MLXIPL is more complicated than expected (**graphical summary - FIGURE C35**): we demonstrated that rRFX6 activates the expression of *rMlxipl* α in low and high glucose. As already known, rMLXIPL α subsequently induces the transcription of *rMlxipl* β especially in in high glucose, since rMLXIPL α is activated by glucose. In the first part of the thesis, we further proofed that the transcription of *rMlxipl* α is inhibited in high glucose either directly by rMLXIPL β or by the activated rMLXIPL α. In this second part, we now showed that rMLXIPL α and / or rMLXIPL β repress the transcription of *rRfx6* in high glucose. This complex regulatory network makes it difficult to assign clear roles to rRFX6, rMLXIPL α and rMLXIPL β. Further experiments will be needed to understand their exact roles.

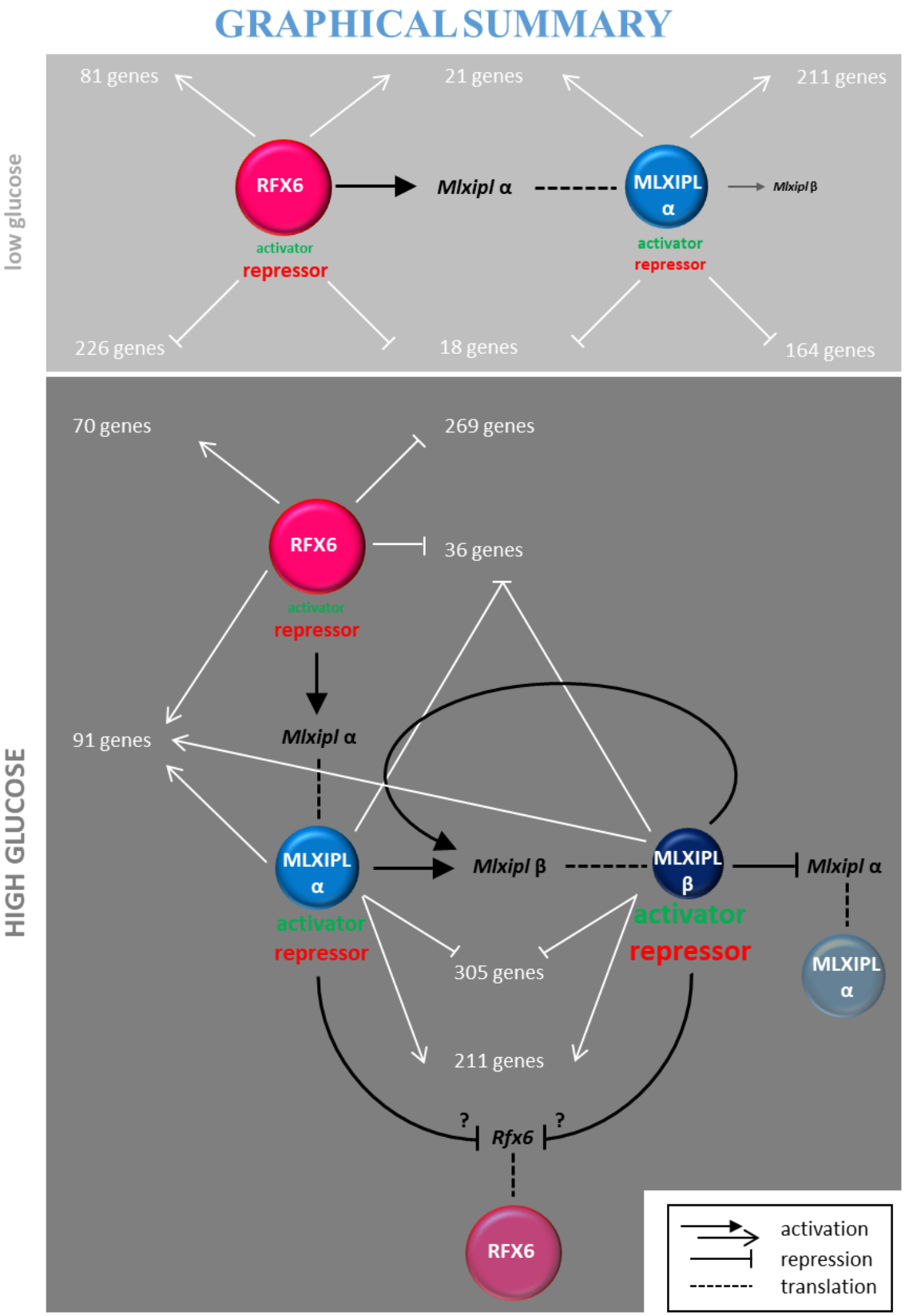


FIGURE C35: Model of the transcriptional network regulated by RFX6 and MLXIPL in Ins-1 832/13 cells at high and low glucose. Graphical summary illustrating the results of the RNA sequencing experiment and the transcriptional regulation of *Rfx6* and *Mlxipl.*

D) General conclusion and future perspectives

In 2010, it was discovered that the winged-helix transcription factor RFX6 is implicated in the proper development of the murine pancreas (Soyer *et al.,* 2010; Smith *et al.,* 2010), and, moreover, that *Rfx6* mutations cause Mitchell-Riley syndrome in humans (OMIM #601346; Smith *et al.,* 2010) characterized by neonatal diabetes and developmental defects in the formation of the digestive system (Mitchell *et al.,* 2004). In 2014, we and our collaborators showed that RFX6 is not only important during embryogenesis but also essential for the maintenance of the adult beta cell identity and function (Piccand *et al.,* 2014; Chandra *et al.*, 2014). Loss of RFX6 in beta cells of adult mice leads to reduced glucose tolerance and glucose-stimulated insulin secretion (Piccand *et al.,* 2014). Studying the *mRfx6* Δbeta mice allowed us to gain profound insights into the role of mRFX6 in the adult beta cells and in its target genes. Among the mRFX6-bound and downregulated genes, the transcription factor *mMlxipl* was the most affected in the beta cell-specific *mRfx6* knockout mice.

1. What was done in the first part of the project?

The first objective of my thesis was to precisely characterize the regulation of *Mlxipl* α by RFX6 via *Mlxipl* intron 1. In a 1200 kb long sequence around the ChIP peak, we found five RFX binding sites, called xbox (Laurençon *et al.,* 2007) and that the sequence of the second xbox in *mMlxipl* intron 1 is conserved in the genome of mice, rats and humans. By transactivation assays, we demonstrated that the *mMlxipl* intron 1 region has enhancer activity critical for *Mlxipl* α transactivation and that xbox 2 is a necessary and sufficient xbox motif. We further used our *Mlxipl* intron 1 transactivation assay as a model to test the effect on RFX6 transcriptional function of three published (Martinovici *et al.,* 2009; Chappell *et al.,* 2008; Chandra *et al.,* 2014) and two yet unpublished mutations (communicated by Martine Vaxillaire from the Pasteur institute in Lille) identified in the human *RFX6* gene of patients with Mitchell-Riley syndrome or in maturity-onset diabetes of the young (MODY) patients; four of these mutations severely affected RFX6 function. To elucidate *Mlxipl* regulation specifically in beta cells, we tested whether the endogenous transcription factors of the murine pancreatic beta cell line Min6b1 (Lilla *et al.,* 2003) and the rat cell line Ins-1 832/13 (Hohmeier *et al.,* 2000) can induce the *mMlxipl* intron 1 reporter constructs. We got a strong transactivation in both cell lines, and the transactivation was affected by the deletion of xbox 2, suggesting that an xbox-binding transcription factor such as RFX6 plays a pivotal role in the transactivation. We studied the cofactors that might be involved in *Mlxipl* regulation by RFX6 and we elaborated the following theoretical model about the transcriptional regulation of *Mlxipl* in beta cells: The pioneer transcription factors FOXA1 and FOXA2 opens the chromatin at the *Mlxipl* locus and enable other transcription factors such as RFX6, NKX2.2, E2F1, RFX3, NEUROD1 and MLXIPL to bind. RFX6, NKX2.2 and RFX3 activate the transcription of *Mlxipl*

α while E2F1 and MLXIPL act as inhibitors. Overall, however, our data in mice, rats and humans clearly demonstrated that RFX6 is the essential transcription factor that ensures *Mlxipl* regulation.

2. What was done in the second part of the project?

In the second part of the thesis, we aimed to determine the respective roles of RFX6 and MLXIPL in adult beta cells. We tested three different strategies to inactivate RFX6 and MLXIPL in beta cell lines: First, we tried to knockout *mRfx6* and *mMlxipl* in the murine pancreatic beta cell line Min6b1 (Lilla *et al.,* 2003) using the Cas9 nickase in combination with two single guide RNAs (Jinek *et al.,* 2012; Ran *et al.,* 2013a; Ran *et al.,* 2013b). We managed to produce one homozygous knockout clone for each gene, but the functional analysis revealed that RFX6 and MLXIPL target genes were not affected in the knockout cell lines. Next, we targeted *rRfx6* and *rMlxipl* in the rat pancreatic beta cell line Ins-1 832/13 (Hohmeier *et al.,* 2000) by a similar strategy. We only obtained heterozygous *rRfx6* and *rMlxipl* clones, but no homozygous knockout clone. Finally, we inactivated *rRfx6* and *rMlxipl* by RNA interference in Ins-1 832/13 cells starved the knockdown cells overnight, stimulated them for 6 h in low or high glucose medium and sequenced the transcriptomes.

We revealed that RFX6 is rather a transcriptional repressor than a transcriptional activator, and that its activity is not altered by glucose. Interestingly, we further found that rMLXIPL directly inhibits the transcription of *rRfx6* in high glucose, indicating a feedback control, since RFX6 is the transcriptional activator of *Mlxipl* α. MLXIPL is both a transcriptional activator and repressor, and, consistently with the published studies, its activity was higher in high glucose than in low glucose (Li *et al.,* 2006; Herman *et al.,* 2012). Unexpectedly, the siRNA that we used to knockdown *rMlxipl,* had a cell-toxic effect that was intensified in starvation. We deduced that the downregulation of the glycolytic gene *rPklr* in the knockdown cells impaired glycolysis and energy production, and that the provided energy in low glucose was not enough to ensure cell survival.

Overall, our data showed that MLXIPL downstream of RFX6 is important for the execution of the glucose response. The direct inhibition of MLXIPL by siMlxipl treatment or its indirect repression by the *rRfx6* knockdown strongly reduced the number of glucose-activated and -repressed genes. Thus, we concluded that MLXIPL controls the activation and silencing of genes after a glucose stimulus. Moreover, we demonstrated that *rIns1* and *rIns2* were directly repressed in Ins-1 832/13 cells by rMLXIPL alone and by rRFX6 and rMLXIPL together, respectively. To our knowledge, the direct regulation of insulin genes by RFX6 and MLXIPL has not been revealed before. The fact that both *rIns1* and *rIns2* are strongly upregulated in the knockdown cells suggest that RFX6 and MLXIPL are involved in the control of the beta cell insulin content.

PERSPECTIVES

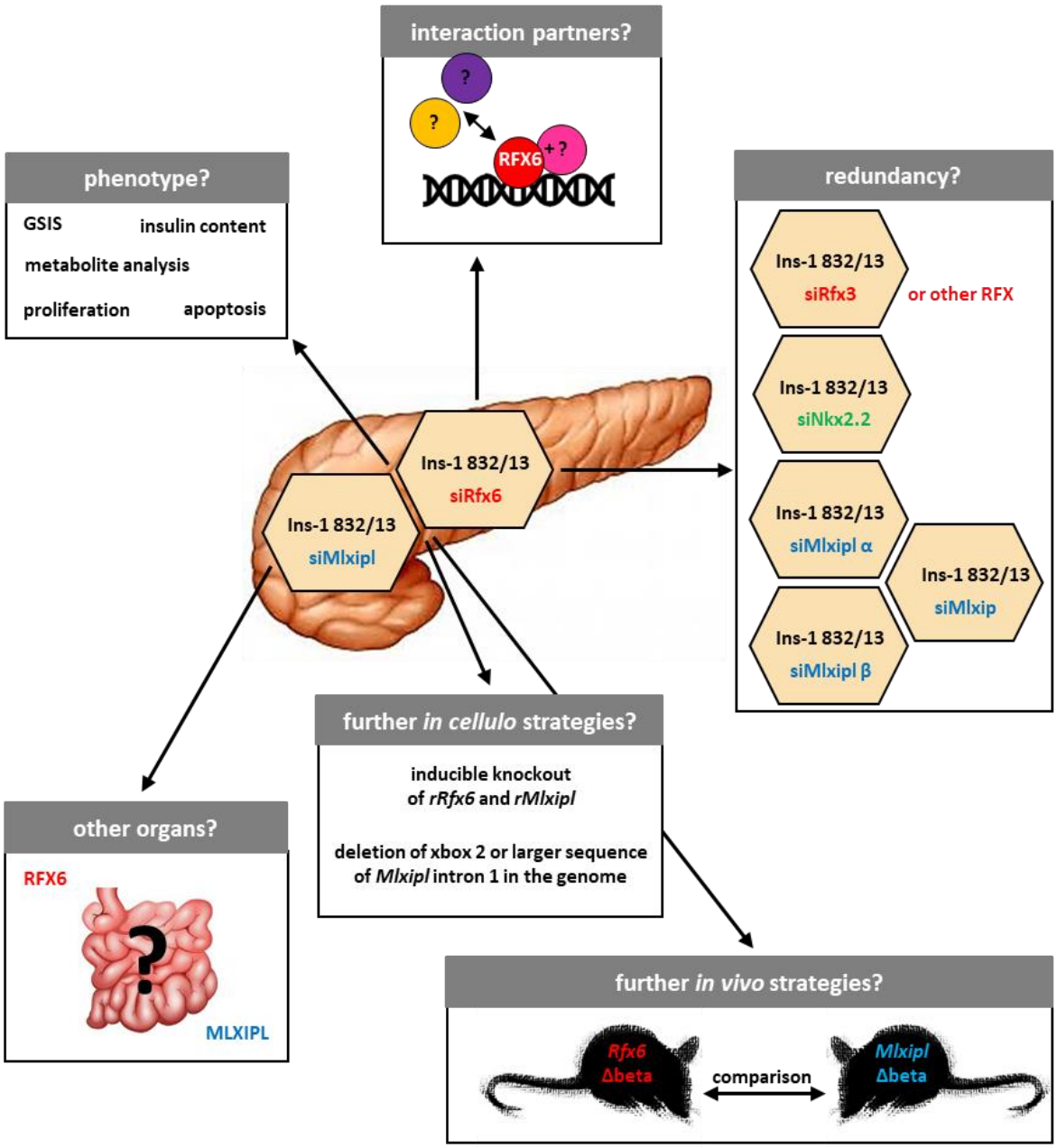


FIGURE D1: Perspectives. Follow-up projects mentioned in the thesis resulting from open questions in this project. These include near-term analyses such as those of the phenotype, which will help to better understand RNA sequencing data, the identification of interaction partners, redundancies with other transcription factors, the role of MLXIPL and RFX6 *in vivo* in genetically modified mice and in other organs, and alternative knockout strategies in Ins-1 832/13 cells (images of the organs from the University of Leeds).

3. RFX6 and MLXIPL – activators or inhibitors of insulin production and secretion?

For an optimal evaluation of the RNA sequencing data, it will be necessary to study the phenotype of siRfx6- and siMlxipl-treated Ins-1 832/13 cells more closely. A bioinformatic search for particularly

affected pathways in the two knockdown cells had provided no concrete evidence. Moreover, it is difficult to draw conclusions about the phenotype from the altered expression levels of genes since the mRNA level gives no indication of the amount of protein and the proportion of active protein and because most proteins are not only involved in one but in different cellular processes.

Obvious analyses would include a determination of the insulin content and the Glucose-Stimulated Insulin Secretion (GSIS), as we had found that the insulin-encoding genes *rIns1* and *rIns2* belong to the RFX6- and MLXIPL-repressed genes. Published studies in Min6b1 cells, in Ins-1 832/13 cells and in mouse islets already indicated that MLXIPL inhibits the expression of insulin genes (Da Silva Xavier *et al.,* 2006; Da Silva Xavier *et al.,* 2010; Noordeen *et al.,* 2010; Poungvarin *et al.,* 2012) and that GSIS is increased in the absence of MLXIPL (Da Silva Xavier *et al.,* 2010). By contrast, a reduction of *RFX6* in human EndoC-betaH2 cells reduced *INS* transcription (Chandra *et al.,* 2014), the *mRfx6* Δbeta mice had decreased *mIns1* transcript levels (Piccand *et al.,* 2014) and the GSIS was strongly affected in both models due to a reduced activation of insulin secretion genes in the absence of RFX6 (Chandra *et al.,* 2014; Piccand *et al.,* 2014). In summary, the literature suggests that insulin production and secretion is positively affected by RFX6 and negatively by MLXIPL.

However, our experiments show that the insulin transcripts were highly elevated in both the siRfx6 and siMlxipl-treated Ins-1 832/13 cells. Our experiments are the first ever to investigate the target genes of RFX6 and MLXIPL under the same experimental conditions. We should use our model system to determine the consequences of *rRfx6* and *rMlxipl* knockdown on insulin production by determining the mRNA and protein level and on insulin secretion by insulin ELISA and calcium imaging (**FIGURE D1**). Both insulin production and secretion should be analysed at different time points after the glucose stimulus to better understand the role of RFX6 and MLXIPL in the early and late glucose response and in glucolipotoxicity. This analysis could help to understand the respective roles of RFX6 and MLXIPL in insulin secretion and to explain the discrepancies.

4. RFX6 induces *Mlxipl* and MLXIPL inhibits *Rfx6* – a necessary genetic regulatory circuit?

One of the major achievements of this thesis was to demonstrate that RFX6 activates the transcription of *Mlxipl* α. So far, it was only known that *Mlxipl* α is constitutively produced, so its expression is not dependent on a stimulus, and that several transcription factors such as FOXA1, FOXA2, LXR, RXR and THR may play a role in its transcriptional regulation (Gao *et al.,* 2010; Cha and Repa, 2007; Hashimoto *et al.,* 2009). However, of the above five factors, the three nuclear factors are only activated by the binding of their respective ligands, so constitutive regulation of *Mlxipl* α by these three factors seems unlikely. The discovery of the transcriptional control of *Mlxipl* α by RFX6 could be the mechanism that maintains the expression of *Mlxipl* α in low glucose ensuring that, during a glucose

stimulus, there is enough MLXIPL protein to increase glycolysis and lipogenesis. This regulation does not occur in the promoter region, as would be the case for LXR, RXR, and THR (Cha and Repa, 2007; Hashimoto *et al.,* 2009), but in a more downstream intron 1 region, which is also bound by FOXA1 and FOXA2 (Gao *et al.,* 2010). In addition to RFX6, FOXA1 and FOXA2, we have evidence that NKX2.2 and NEUROD1 (Churchill *et al.,* 2017) as well as RFX3 and E2F1 are involved in the regulation of *Mlxipl* α. Unfortunately, Churchill *et al.* not only showed the ChIP sequencing peaks of NKX2.2 (Gutiérrez *et al.,* 2017) and NEUROD1 (Tennant *et al.,* 2013) on the *Mlxipl* intron 1, but also extracted the RFX6 peak from our Min6b1-3HA mRFX6 anti-HA data set (Piccand *et al.,* 2014), meaning that they hereby announce a possible involvement of RFX6 in the regulation of *Mlxipl* before we can do so (Churchill *et al.,* 2017). However, our study is a functional characterization of *Mlxipl* α regulation by RFX6 and cofactors including NKX2.2 and goes far beyond ChIP and RNA sequencing data. The next step should be to definitively prove the importance of the *Mlxipl* intron 1 region for the transcriptional activation of *Mlxipl* by deleting or changing xbox motifs (e.g. xbox 2 or xbox 3) or larger sequences in this region using the CRISPR / Cas9 method and to analyse the consequences on *Mlxipl* expression (**FIGURE D1**).

So far, we focussed on the transcriptional regulation of *Mlxipl* in pancreatic beta cells. We have further evidence that RFX6 also regulates *Mlxipl* in the mouse intestine. While *mMlxipl* is activated by mRFX6 in mouse islets, the data from mice with an intestine-specific knockout of *mRfx6* show that *mMlxipl* is upregulated in the absence of mRFX6 in this organ (unpublished data of Julie Piccand). The regulation of *Mlxipl* by RFX6 in other organs could be a good opportunity to learn more about their functions in different organs (**FIGURE D1**).

Besides, our study also suggests that not only the transcriptional activation of *Mlxipl* but also its inhibition in high glucose are important. Since persistent activation of MLXIPL and its nuclear localization is associated with lipid accumulation in the cytoplasm, oxidative stress and apoptosis, a situation referred to as glucolipotoxicity (Poungvarin *et al.,* 2012), it seems evident that the cell has implemented genetic circuits to transcriptionally and post-translationally limit the action of MLXIPL α and MLXIPL β when they are no longer needed. On the protein level, the inactivation of MLXIPL α in low glucose is mediated by phosphorylation and its cytoplasmic retention by the 14-3-3 protein or SORCIN (Filhoulaud *et al.,* 2013, review), while *Mlxipl* β is not expressed or only at very low levels (Herman *et al.,* 2012; Zhang *et al.,* 2015). In high glucose, MLXIPL α is activated and it induces the transcription of *Mlxipl* β that further promotes its own transcription (Herman *et al.,* 2012). Nothing is known about how the constitutive active isoform MLXIPL β is inactivated after its production.

Interestingly, in high glucose, MLXIPL inhibits the transcription of its activators. Iizuka *et al.* reported that MLXIPL induces the transcription of the glucagon receptor (*Gcgr*) in Ins-1E cells (Iizuka *et al.,* 2012). Glucagon signalling triggers the phosphorylation of MLXIPL and it inhibits the transcriptional activity of a MLXIPL / HNF4A dimer in the nucleus (Richards *et al.,* 2017, review). In this thesis, we identified a

peak of rMLXIPL on the first intron of *rMlxipl* α. This finding, in combination with the observations of Jing *et al.,* who showed that the inhibition of *rMlxipl* β led to a rise in rMLXIPL α activity (Jing *et al.,* 2016), suggests that rMLXIPL β directly represses *rMlxipl* α. However, we were also unable to find the ChoRE needed for the binding of MLXIPL. The fact that ChoREs do not necessarily consist of two E-box motifs separated by five nucleotides, but also E-box-like motifs (Yamashita *et al.,* 2001), makes it difficult to find them in genomic DNA. Despite the non-identification of the ChoRE, our new data provides a bridge between *Mlxipl* α and *Mlxipl* β, which has been sought since the discovery of *Mlxipl* β (Richards *et al.,* 2017, review).

We further demonstrated that rMLXIPL directly represses the transcription of its activator *rRfx6*. Like the regulation of *Mlxipl* α, the transcriptional regulation of *Rfx6* is a mystery that still needs to be solved. So far, it is only known that *Rfx6* is expressed in NGN3-positive cells, suggesting that NGN3 or a downstream effector of NGN3 contribute to the transcriptional activation of RFX6 (Soyer *et al.,* 2010). However, *Rfx6* is also expressed earlier in the development in the gut endoderm and in the pancreatic buds before its expression gets temporarily excluded from the pancreatic progenitors and then restricted to the NGN3-positive endocrine progenitor cells (Soyer *et al.,* 2010; Smith *et al.,* 2010). The mechanism by which the early expression of *Rfx6* and its short-term inhibition is regulated is unknown. Similarly, nothing is known about its regulation in adult islet cells. The inhibition of *rRfx6* in ins-1 832/13 cells in high glucose, which is presumably caused by rMLXIPL through its binding to *rRfx6* intron 18, is a first indication of genomic regions that might be important for the regulation of *Rfx6*.

5. Could redundancy mitigate the *Rfx6* and *Mlxipl* knockout phenotype?

Another important issue related to the role of RFX6 and MLXIPL in beta cell survival is the compensation for their elimination by other transcription factors. RFX6 belongs to the RFX transcription factors family. In the Ins-1 832/13 cells, we detected a high expression of *Rfx1, Rfx2*, *Rfx3*, *Rfx5*, *Rfx6* and *Rfx7*, a low expression of *Rfx8* and no expression of *Rfx4*; murine islets weakly express *Rfx1*, *Rfx2, Rfx7* and *Rfx8* and strongly express *Rfx3*, *Rfx5* and *Rfx6* (Piccand *et al.,* 2014; unpublished data of Aline Meunier). So far, only the function of RFX3 and RFX6 has been reported in the pancreas (Ait-Lounis *et al.,* 2007; Ait-Lounis *et al.,* 2010; Smith *et al.,* 2010; Soyer *et al.,* 2010; Piccand *et al.,* 2014; Chandra *et al.,* 2014), which, however, excludes no importance of the other six RFX proteins in pancreas development and function. In theory, each of them could eventually take over for RFX6. Notwithstanding, the full knockout phenotype of *Rfx6* in mice leads to the loss of glucagon-, insulin-, somatostatin- and ghrelin-expressing cells (Smith *et al.,* 2004; Piccand *et al.,* 2014) and does not suggest that any of the other RFX proteins completely fulfils the role of RFX6. There is some evidence that RFX3 and RFX6 may have redundant roles in adult pancreatic beta cells since, for instance, they

co-regulate *Gck* expression (Ait-Lounis *et al.,* 2010; Piccand *et al.,* 2014), and that they might form a dimer (Smith *et al.,* 2010), thus RFX3 should be the first candidate to be tested, for instance via siRfx3 treatment of Ins-1 832/13 cells (**FIGURE D1**) in a similar way that allowed us to determine RFX6 and MLXIPL targets in this cell line. An analysis of the dimerization and interaction partners of RFX6 (**FIGURE D1**) could further help to understand the respective roles of the RFX proteins in adult beta cells.

An important paralogue of MLXIPL is MLXIP; the redundancy between the two proteins might explain the mild phenotype of *mMlxipl* knockout mice (Iizuka *et al.,* 2004). The pancreas is one organ that co-express MLXIPL and MLXIP, and a recent study suggests that MLXIP is more important for beta cells than MLXIPL (Richards *et al.,* 2018). Nevertheless, it is believed that MLXIPL is important for beta cells as it is highly expressed in this cell type. To better distinguish between the functions of MLXIPL and MLXIP in beta cells, it would be useful to compare the knockdown phenotype of both genes in beta cell lines (**FIGURE D1**) and the beta cell-specific knockout in mice.

No study, including ours, has so far focused specifically on the transcriptomes that depend on the two MLXIPL isoforms in mature beta cells. For the sake of simplicity, it is generally assumed that *Mlxipl* β can not be formed in the absence of MLXIPL α and that siRNAs efficiently target both transcript variants. However, one could imagine that the ChoRE in *Mlxipl* β could also be bound by MLXIP / MLX, whereby the expression of MLXIPL β could be partly independent of MLXIPL α. While specific siRNAs are available against *Mlxipl* α and *Mlxipl* β, unfortunately, there is no antibody that could bind only to MLXIPL β and not to MLXIPL α, since MLXIPL β is a truncated isoform of MLXIPL α (Herman *et al.,* 2012). This makes it impossible to distinguish between MLXIPL α- and MLXIPL β-linked genes in ChIP sequencing data. Nonetheless, analysing the specific roles of MLXIPL α and MLXIPL β not only in beta cells (**FIGURE D1**) but also in other tissues is the next important step to understand the function of MLXIPL.

6. Beta cell lines – what are they worth?

Beta cell lines provide a simple model system to study the mechanisms underlying insulin secretion. Most notably, they were generated to exhibit maximal glucose-induced insulin secretion and minimal insulin secretion in low glucose, but they do not perfectly mimic the physiology of natural beta cells (Skelin *et al.,* 2010, review). Moreover, while natural beta cells are largely post-proliferative (Meier *et al.*, 2008; Teta *et al.*, 2005), beta cell lines have a high doubling rate because they are tumour cells (Skelin *et al.,* 2010, review). Scharfmann *et al.* considered this in the development of their human EndoC-betaH2 cell line in which they can limit the proliferation by the excision of the transgenes (Scharfmann *et al.,* 2014).

Beta cell lines might be the ideal model to study the mechanism of insulin secretion in detail and to test the influence of drugs on insulin secretion. Their benefit in studies beyond this should be critically scrutinized. Especially, experiments involving proliferation or apoptosis may give erroneous results due to the tumoral properties of beta cell lines. But even other studies that aim at gene expression changes or protein regulation, could be adulterated. The high variance between the results of similar studies in different beta cell model systems, which is evident in reading the literature, should motivate researchers to analyse their results with the idea in mind that the results are often cell line-dependent and do not necessarily reproduce *in vivo* data.

To investigate the role of RFX6 and MLXIPL in a more physiological model than an immortalized beta cell line, we should compare the consequences of their loss in beta cells of adult mice using a tamoxifen-inducible knockout. In contrast to *mRfx6* which is an essential gene for embryogenesis and the development of the pancreas and intestine – mice with a whole-body *mRfx6* knockout die shortly after birth due to diabetes (Smith *et al.,* 2010; Soyer *et al.,* 2010; Piccand *et al.,* 2014) –, *mMlxipl* knockout mice develop normally (Iizuka *et al.,* 2004). Adult *mMlxipl* -/- mice are mildly glucose-intolerant and insulin-resistant (Iizuka *et al.,* 2004). Given that the expression of *mMlxipl* was almost lost in the islets of *mRfx6* Δbeta mice, one could imagine that the phenotype of the *mRfx6* Δbeta mice resembles the one of *mMlxipl* -/- mice. Indeed, glucose intolerance and insulin resistance have been reported in both mouse lines (Piccand *et al.,* 2014; Iizuka *et al.,* 2004), but the insulin resistance in *mMlxipl* -/- mice is a systemic effect that involves multiple organs and is different in lean and obese mice (Iizuka *et al.,* 2006; Dentin *et al.,* 2006). We propose the use of two inducible beta cell-specific knockouts rather than the comparison of the existing *mMlxipl* -/- mice (Iizuka *et al.,* 2004) with the *mRfx6* Δbeta mice (Piccand *et al.,* 2014) to circumvent the problem of the systemic effect and to be able to compare the pancreas phenotypes (**FIGURE D1**).

7. Could the regulation of *Mlxipl* by RFX6 play a role in Mitchell-Riley syndrome?

Mutations in the *RFX6* gene cause Mitchell-Riley syndrome, which is a developmental disorder affecting the pancreas, intestine, and the bile ducts as well as the gallbladder (Smith *et al.,* 2010; Mitchell *et al.,* 2004). In this thesis, we demonstrated that RFX6 regulates *Mlxipl* in pancreatic beta cells. This regulatory mechanism might not be restricted to beta cells but might also take place in other cells types, including the other four endocrine cell types of the pancreas, in the enteroendocrine cells of the gut, and perhaps in other tissues. Both *Rfx6* and *Mlxipl* are expressed in the gut. The pancreatic phenotype of Mitchell-Riley syndrome patients, apart from the development of an artificial pancreas, does not provide a target for seeking further therapeutic treatments because pancreatic developmental disorders can not be corrected and (neonatal) diabetes can be efficiently treated with

insulin injections. By contrast, the patients also often have diarrhoea. Research aimed at the role of RFX6 and MLXIPL in the gut could help to develop a patient-specific diet that improves this symptom.

Heterozygous carriers of *RFX6* mutations sometimes but not always suffer from diabetes (Mitchell *et al.,* 2004; Chappell *et al.,* 2008; Martinovici *et al.,* 2009; Smith *et al.,* 2010; Spiegel *et al.,* 2011; Concepcion *et al.,* 2014; Sansbury *et al.,* 2015; Huopio *et al.,* 2016; Skopkova *et al.,* 2016; Patel *et al.,* 2017). The haploinsufficiency of *rRfx6* in the rat Ins-1 832/13 cells did not cause a reduction in *rMlxipl* expression levels, suggesting that half of the rRFX6 amount is still enough to properly induce the expression of *rMlxipl* in rat beta cells. This could but must not be true in human cells. In many research projects, including the beta cell field, mice and rats are used as model organisms to understand basic principles. This is necessary because research on humans is not possible, and little biological material is available from humans. However, the transfer of these results to humans is difficult due to species differences. For the moment, it is impossible to predict if the haploinsufficiency of *RFX6* and the concomitant reduction in *MLXIPL* in humans attenuates or aggravates the development of diabetes in heterozygous carriers or if there are no consequences at all.

8. Concluding remarks

We started this thesis project with two goals: the first was the characterization of the *Mlxipl* regulation by RFX6 that we have achieved. We proved that RFX6 regulates *Mlxipl* via a specific xbox in *Mlxipl* intron 1 and we further identified involved cofactors. The second goal was the identification of the respective and common target genes of RFX6 and MLXIPL, a task that we have solved through our RNA sequencing of siRfx6- and siMlxipl-treated Ins-1 832/13 cells and the comparison of these data with published and unpublished ChIP sequencing data. As always, this project like all the other research projects has raised more open questions than it has answered, leaving room for further experimentation and curiosity.

E) Résumé en français

Ce résumé en français récapitule brièvement le contexte scientifique, les objectifs, la stratégie expérimentale ainsi que les résultats obtenus qui ont été décrits en détail dans les parties A à D rédigées en anglais.

1. Introduction

1.1. La cellule bêta pancréatique sécrète de l'insuline en réponse au glucose

Les cellules bêta pancréatiques, localisées dans les îlots de Langerhans, sont spécialisées dans la production et la sécrétion endocrine de l'hormone polypeptidique insuline qui est libérée dans des conditions hyperglycémiques. L'insuline est indispensable pour la survie parce qu'elle permet l'absorption des nutriments par les cellules. Dans le cas où les cellules bêta ne sécréteraient pas suffisamment d'insuline ou les tissus périphériques seraient résistants à l'action de l'insuline, l'hyperglycémie devient chronique. Ce trouble du métabolisme est une pandémie connue sous le nom de diabète sucré (WHO, 2018).

Le glucose est le stimulant le plus puissant de la sécrétion d'insuline. Après son import et sa phosphorisation par l'enzyme glucokinase, il entre dans la glycolyse et le métabolisme mitochondrial, ce qui augmente le rapport ATP / ADP. Cette augmentation induit la fermeture des canaux potassiques sensibles aux variations d'ATP, ce qui a pour effet d'empêcher la sortie des ions potassium entrainant la dépolarisation de la membrane cellulaire. Par conséquent, des canaux calciques, sensibles au voltage, s'ouvrent et l'influx des ions calcium déclenche l'exocytose des vésicules d'insuline (Rorsman et Braun, 2013, revue).

Le transcriptome des cellules bêta est adapté à leur tâche (Skelin Klemen *et al.,* 2017, revue) : D'une part, elles expriment les gènes impliqués dans la production et la sécrétion d'insuline comme les transporteurs de glucose, les enzymes de la glycolyse (par exemple *Gck*) et du métabolisme mitochondrial ainsi que les composants des canaux potassiques (*Abcc8, Kcnj11*) et calciques (par exemple *Cacna1c, Cacnb2*) et des granules d'insuline. D'autre part, la haute spécialisation des cellules bêta nécessite une répression des gènes d'autres voies métaboliques qui interfèrent avec les voies métaboliques dont elle a besoin. Ces gènes sont désignés *disallowed genes* (Quintens *et al.,* 2008 ; Pullen *et al.*, 2010 ; Thorrez *et al.,* 2011). Ainsi le gène encodant la lactate déshydrogénase A (*Ldha*) qui réduit l'efficience du métabolisme du glucose, est réprimé (Sekine *et al.,* 1994).

Le maintien du transcriptome des cellules bêta est contrôlé par un réseau complexe constitué des facteurs de transcription, des enzymes remodelant la chromatine et des voies de signalisation induites par des stimuli externes avec le glucose jouant un rôle majeur dans la régulation de l'expression (Schuit

et al., 2002, revue ; Rutter *et al.,* 2015, revue). Il a été démontré, par exemple, que le glucose induit la transcription de l'insuline en activant les facteurs de transcription clé PDX1, MAFA et NEUROD1 (Andrali *et al.,* 2008, revue).

1.2. Le facteur de transcription RFX6 maintien la fonctionnalité et l'identité de la cellule bêta

En 2014, notre équipe a démontré que le facteur de transcription RFX6 est essentiel pour la sécrétion d'insuline et le maintien de l'identité de la cellule bêta en activant des gènes impliqués dans la sécrétion d'insuline et en réprimant un grand nombre de *disallowed genes* (Piccand *et al.,* 2014). RFX6 est un membre de la famille *regulatory factor x* et se fixe sur le motif xbox RYNNYYN(1-3)RRNRAC (Laurençon *et al.,* 2007) par un domaine de liaison à l'ADN de type *winged-helix* (Reith *et al.,* 1989 ; Reith *et al.,* 1994 ; Aftab *et al.,* 2008 ; Gajiwala and Burley, 2010, revue) en tant qu'homo- ou hétérodimère avec un deuxième RFX (Reith *et al.,* 1989 ; Reith *et al.,* 1994). L'expression de *Rfx6* est restreinte au système gastro-intestinal. Pendant le développement, *Rfx6* est exprimé dans l'endoderme intestinal avant que son expression ne soit confinée aux progéniteurs pancréatiques puis endocrines. Chez l'adulte, l'expression de *Rfx6* est maintenue dans toutes les cellules endocrines des îlots de Langerhans dont les cellules bêta et alpha sécrétrices respectivement d'insuline et de glucagon (Smith *et al.,* 2010 ; Soyer *et al.,* 2010).

La fonction de RFX6 dans la cellule bêta a été déterminée dans l'équipe grâce à un modèle de souris où il était possible d'inactiver le gène *Rfx6* spécifiquement dans les cellules bêta adultes (les souris *mRfx6* Δbêta ; Piccand *et al.,* 2014 ; **FIGURE E1A**). Contrairement aux souris avec un knockout constitutif de RFX6, qui meurent deux jours après leur naissance (Smith *et al.,* 2010 ; Piccand *et al.,* 2014), ces souris survivent avec une tolérance au glucose et une sécrétion d'insuline réduites (Piccand *et al.,* 2014). Pour expliquer ce phénotype, l'équipe a identifié les cibles de RFX6 en réalisant des séquençages d'ARN des îlots des souris sauvages et mutées (Piccand *et al.,* 2014 ; **FIGURE E1A**). En parallèle, elle a recherché les sites de fixation de RFX6 par séquençage ChIP dans la lignée cellulaire bêta murine Min6b1 en détectant soit la protéine endogène avec un anticorps anti-RFX6 (données non publiées de Perrine Strasser ; **FIGURE E1A**) soit la protéine de fusion 3HA-mRFX6 qui a été introduite de manière transitoire dans les Min6b1, avec un anticorps anti-HA (Piccand *et al.,* 2014 ; **FIGURE E1A**). Un troisième ChIP dans l'îlot murin a permis d'identifier les cibles de RFX6 pas seulement dans les cellules bêta mais dans l'ensemble des cellules endocrines pancréatiques (données non publiées de Perrine Strasser ; **FIGURE E1A**). La comparaison des résultats du séquençage de l'ARN et du séquençage ChIP a révélé les cibles directes de RFX6 dans les cellules bêta pancréatiques (**FIGURE E1A+B**). Le phénotype des souris *mRfx6* Δbêta résulte probablement de l'expression réduite des cibles *Gck*, *Abcc8* et des canaux potassiques sensibles à l'ATP (par exemple *Cacna1c*), tous impliqués

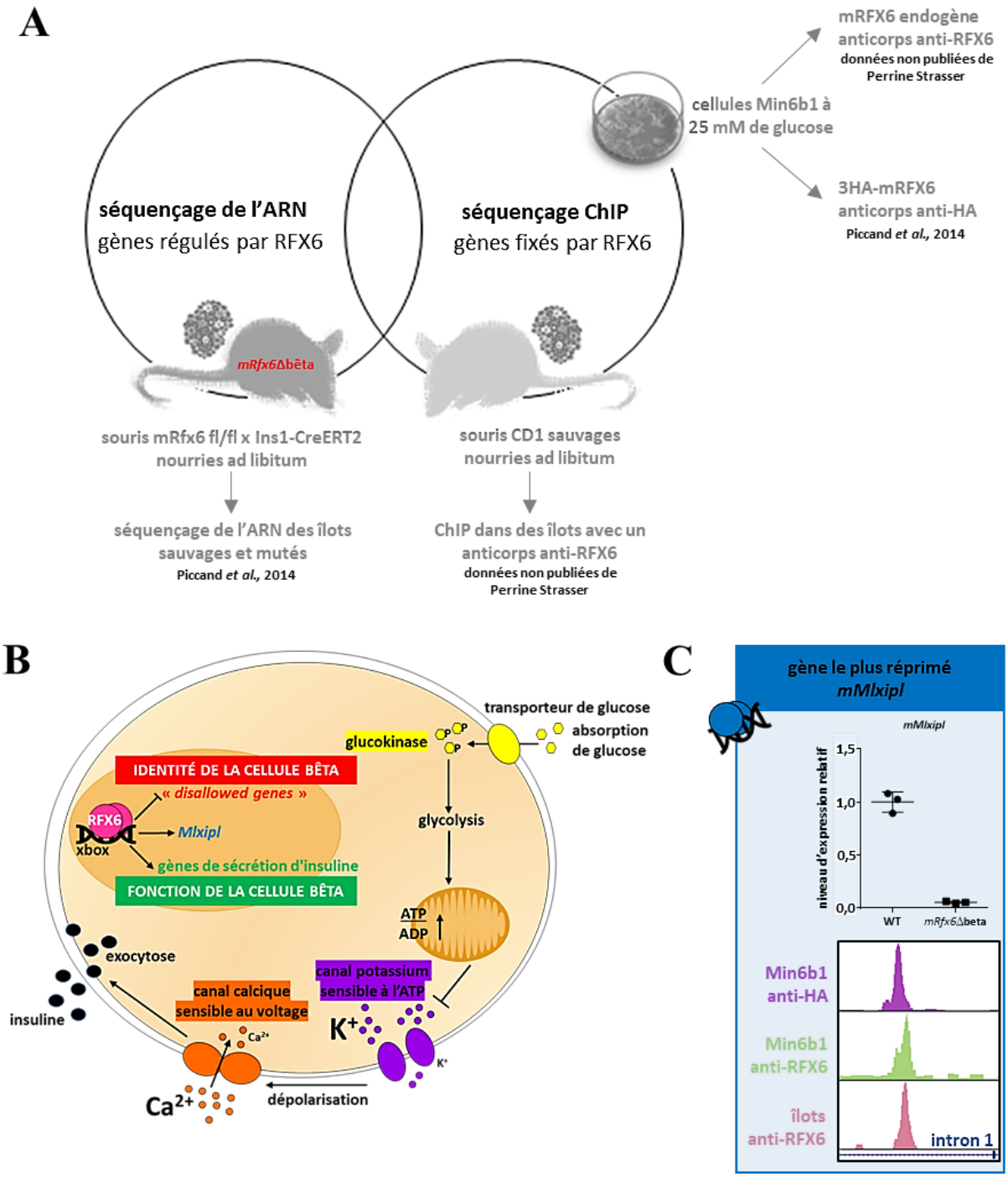


FIGURE E1 : L'identification des cibles directes de mRFX6 dans les cellules bêta pancréatiques murines a révélé *mMlxipl* comme cible la plus affectée par le knock-out. (A) Stratégie expérimentale pour identifier les gènes régulés par mRFX6 dans les cellules bêta pancréatiques murines : comparaison des données de séquençage de l'ARN analysant le transcriptome des îlots isolés des souris avec un knock-out de *mRfx6* spécifique aux cellules bêta (GSE59622, Piccand *et al.*, 2014) et des données de séquençage ChIP issues de trois manipulations différentes : un ChIP anti-HA dans les cellules Min6b1 exprimant 3HA-mRFX6 (GSE62844, Piccand *et al.*, 2014), un ChIP anti-RFX6 dans des cellules Min6b1 non-transfectées ainsi qu'un ChIP anti-RFX6 dans les îlots de souris CD1 (données non publiées de Perrine Strasser). **(B)** Schéma montrant la fonction de RFX6 dans les cellules bêta pancréatiques (figure adaptée de Piccand *et al.*, 2014). **(C)** Niveau d'expression de *mMlxipl* dans les îlots des souris *mRfx6* Δbêta (expression mesurée par le séquençage de l'ARN) et fixation de mRFX6 sur *mMlxipl* dans les Min6b1 et les îlots de souris (détectée par séquençage ChIP).

dans la sécrétion d'insuline. D'autre part, l'augmentation de l'expression des *disallowed genes* comme *Ldha* suggère une perte d'identité des cellules bêta (Piccand *et al.,* 2014 ; thèse de Perrine Strasser ; **FIGURE E1B**).

Chez l'homme, il a été démontré que des mutations dans *RFX6* causent le syndrome de Mitchell-Riley (Smith *et al.,* 2010) caractérisé par un diabète néonatal et des malformations du système gastro-intestinal, notamment de l'intestin. Ce phénotype (Mitchell *et al.,* 2004) est un trouble du développement rare à transmission autosomique récessive. Une seule étude rapporte un diabète de type MODY (*maturity onset diabetes of the young)* qui s'est déclaré chez l'adolescent portant, parmi d'autre, une mutation hétérozygote de *RFX6* (Patel *et al.,* 2017). Néanmoins, les effets des mutations RFX6 hétérozygotes sur l'expression du diabète sont discutables et nécessitent des recherches supplémentaires. Chez l'adulte, des gènes identifiés par des études d'association pangénomique du diabète de type 2, ont souvent des mutations dans les sites de fixation de *RFX6* (Varshney *et al.,* 2017).

1.3. Le facteur de transcription *Mlxipl* est une cible directe de RFX6

Nos études de transcriptome des îlots de souris *mRfx6* Δbêta ont révélé que l'expression du gène *Mlxipl* est fortement diminuée par la perte de RFX6 (données non publiées de Perrine Strasser ; **FIGURE E1B+C**). De plus, grâce aux expériences de séquençage ChIP réalisées dans les cellules Min6b1 et les îlots de Langerhans, l'équipe a pu identifier, que RFX6 se fixait à une région de 1,2 kb dans l'intron 1 de *Mlxipl* qui contient des motifs xbox (données non publiées de Perrine Strasser ; **FIGURE E1C**). Pris dans leur ensemble, ces résultats suggèrent que *Mlxipl* est une cible directe de RFX6.

MLXIPL est un facteur de transcription avec deux isoformes principales : α et β. L'expression de *Mlxipl α* est constitutive mais son activité est régulée par un domaine N-terminal qui est sensible au glucose. Le glucose induit sa translocation dans le noyau où le motif hélice-boucle-hélice permet sa dimérisation avec MLX et la fixation sur la séquence consensus dite *carbohydrate response element* (ChoRE ; Cairo *et al.,* 2001 ; Yamashita *et al.,* 2001). MLXIPL β est active dès sa production, dû à l'absence partielle du domaine sensible au glucose, mais, sa transcription est régulée par deux ChoRE par lesquelles MLXIPL α active sa transcription en présence du glucose (Herman *et al.,* 2012). Le mécanisme de la régulation transcriptionnelle de *Mlxipl* α n'est pas connu. Avant la découverte de la deuxième isoforme en 2012, différents groupes avaient étudié la régulation de *Mlxipl* dans le foie, identifiant les facteurs de transcription OCT-1 en tant que répresseur et LXR, RXR et THR en tant qu'activateur de la transcription de *Mlxipl* (Sirek *et al.,* 2009 ; Cha et Repa, 2007 ; Hashimoto *et al.,* 2009). Dans le pancréas, on sait seulement que l'expression de *Mlxipl* est régulée par FOXA1 et FOXA2 qui se lient à deux régions conservées du premier intron de *Mlxipl* α, la même région génomique

fréquentée par RFX6. L'inactivation de ces deux facteurs de transcription provoque une forte diminution de l'expression de *Mlxipl* (Gao *et al.,* 2010).

La fonction de MLXIPL a été caractérisée en détail dans le foie et dans les adipocytes où il contrôle les gènes de la glycolyse et de la lipogenèse (Postic *et al.,* 2007, revue). L'inactivation constitutive de *Mlxipl* chez la souris ne cause pas un phénotype sévère ; les souris ont une modeste intolérance au glucose et sont légèrement résistantes à l'insuline (Iizuka *et al.,* 2004). Dans les cellules bêta pancréatiques, MLXIPL est impliqué dans la régulation de la glycolyse et de la lipogenèse, mais il induit également la transcription des régulateurs de la sécrétion d'insuline (Da Silva Xavier *et al.,* 2006 ; Boergesen *et al.,* 2011 ; Poungvarin *et al.,* 2012 ; Schmidt *et al.,* 2016). Dans un second temps, MLXIPL active l'expression des gènes de la prolifération (Metukuri *et al.,* 2012 ; Schmidt *et al.,* 2016). Une exposition prolongée au glucose provoque un phénomène de glucolipotoxicité amenant le stress oxydant, l'accumulation intracellulaire des lipides et la mort cellulaire, et est aggravé par MLXIPL (Poungvarin *et al.,* 2012 ; Cha-Molstad *et al.,* 2009). Le rôle de MLXIPL dans le développement du pancréas n'est pas connu.

2. Objectif

Les expériences précédentes de l'équipe ont démontré, d'une part, que l'expression de *Mlxipl* est très fortement abaissée dans les cellules bêta chez la souris en absence de *Rfx6*, et d'autre part, que la protéine RFX6 se fixe au niveau de l'intron 1 de *Mlxipl*. Ces résultats suggèrent que *Mlxipl* est une cible directe de RFX6. Dans la continuité de ce travail, le premier objectif de ma thèse a été d'étudier la régulation transcriptionnelle de *Mlxipl* par RFX6 en identifiant les éléments régulateurs (xbox), ou *enhancer,* dans l'intron 1 de *Mlxipl,* responsables de l'activité transactivatrice de RFX6. Dans un second temps, j'ai testé si d'autres facteurs de transcription corégulent l'expression de *Mlxipl.* Enfin, j'ai tiré parti des tests de transactivation de *Mlxipl* que j'ai établi, pour évaluer la fonctionnalité des protéines RFX6 portant des mutations identifiées chez des patients souffrant du syndrome de Mitchell-Riley.

Afin de mieux comprendre les rôles respectifs de RFX6 et MLXIPL dans la cellule bêta, l'objectif principal de ma thèse a été de déterminer, parmi les gènes régulés par RFX6, lesquels sont dépendants de MLXIPL. L'activité physiologique de la cellule bêta et la translocation nucléaire de MLXIPL α étant régulées par le glucose, il a été prévu d'étudier l'influence du glucose sur les programmes génétiques régulés par RFX6 et MLXIPL. Pour répondre à ces questions, nous avons décidé d'utiliser des lignées des cellules bêta de rongeurs car leur manipulation est facile dans des conditions de glucose définies. Afin d'invalider *Rfx6* et *Mlxipl* dans ces lignées, les techniques de CRISPR/Cas9 et d'interférence par ARN ont été utilisées. Les conséquences de l'inactivation de *Rfx6* et de *Mlxipl* et l'influence du glucose sur les transcriptomes seront analysées par séquençage de l'ARN.

3. Résultats et discussion

3.1. RFX6 contrôle la transcription de *Mlxipl* via une xbox dans l'intron 1

Les expériences de séquençage ChIP réalisées afin d'identifier les cibles de RFX6 dans la lignée murine de cellules bêta Min6b1 (Piccand *et al.,* 2014 ; données non publiées de Perrine Strasser), ont révélé un pic de 800 pb dans le premier intron de *Mlxipl*. Dans une région plus large de 1,2 kb, englobant ce pic, nous avons retrouvé cinq motifs xbox (Laurençon *et al.,* 2007), dont un seul était conservé à l'identique dans les génomes humain, murin et de rat (**FIGURE E2A**). Cette xbox est située à proximité du maximum du pic. Afin de tester la fonction de l'intron 1, et notamment de cette xbox dans la régulation transcriptionnelle de *Mlxipl*, nous avons généré plusieurs constructions, contenant des fragments de l'intron 1 sauvage ou supprimé de la xbox conservée, pour des tests de transactivation à la luciférase (**FIGURE E2A**). Nos résultats montrent que l'expression forcée de *Rfx6* murin ou humain dans la lignée humaine rénale embryonnaire HEK293T induisait fortement l'activité de la luciférase (**FIGURE E2B+C**). Nous avons observé l'activation la plus forte pour la construction contenant 0.6 kb de l'intron 1. La délétion de la xbox conservée diminue fortement cette induction (**FIGURE E2B**). De plus, en utilisant une construction avec quatre répétitions de ce motif xbox en amont du gène de la luciférase, nous avons également pu observer une induction significative du rapporteur (**FIGURE E2B**). Il en résulte que la xbox conservée de l'intron 1 est nécessaire et suffisante pour la transactivation de *Mlxipl* par RFX6.

Par la suite, nous avons pensé tirer parti de la mise en place de ces essais de transactivation de *Mlxipl* pour tester l'effet des mutations de *RFX6,* identifiées chez des patients, sur la fonction de la protéine. Des rapports cliniques suggèrent que la sévérité du syndrome de Mitchell-Riley pourrait être en corrélation avec la position de la mutation et qu'une déclaration retardée du diabète, observée dans certains cas, peut être expliquée par une fonction résiduelle de RFX6 (Sansbury *et al.,* 2015 ; Skopkova *et al.,* 2016). Nous avons choisi trois mutations publiées (Smith *et al.,* 2010 ; Chandra *et al.,* 2014) et notre collaboratrice Martine Vaxillaire de l'institut Pasteur à Lille nous a communiqué deux nouvelles mutations de *RFX6*. L'activité transactivatrice de RFX6 a été diminuée par quatre mutations sur cinq (**FIGURE E2C**) ; elles sont localisées soit dans le domaine de liaison à l'ADN soit dans le domaine de dimérisation qui sont tous les deux essentiels pour le fonctionnement du facteur de transcription. En conclusion, notre test de transactivation est un outil optimal pour évaluer l'impact des mutations de *RFX6* sur sa fonction.

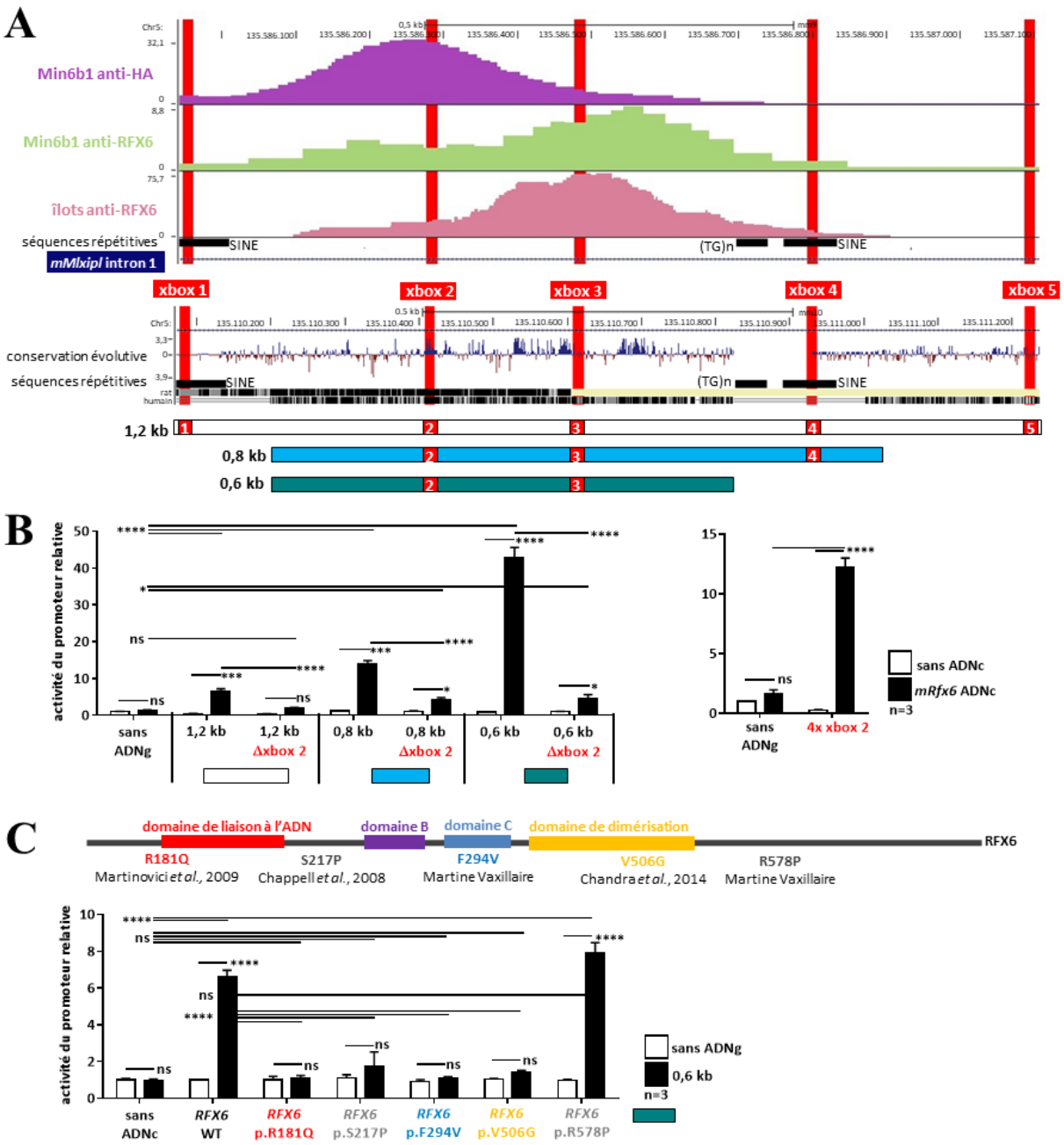


FIGURE E2 : RFX6 se fixe sur le premier intron de *mMlxipl* et active sa transcription. (A) La région intronique de *mMlxipl* où RFX6 se fixe, a été identifiée dans trois manipulations de ChIP indépendantes : un ChIP anti-HA dans les cellules Min6b1 exprimant 3HA-mRFX6 (Piccand *et al.,* 2014), ChIP anti-RFX6 dans des cellules Min6b1 non-transfectées ainsi qu'un ChIP anti-RFX6 dans les îlots murins (données non publiées de Perrine Strasser). Cette région, dans le premier intron de *mMlxipl,* contient cinq motifs de liaison RFX nommés xbox 1 à 5 (les positions sont surlignées en rouge). Les trois cartes sous l'alignement montrent les fragments d'ADN génomique qui ont été clonés dans le vecteur pGL4.23-minP-luc encodant une luciférase. **(B)** Essai de transactivation testant l'activation des plasmides rapporteurs *mMlxipl* intron 1 1,2 kb / 0,8 kb / 0,6 kb avec ou sans la xbox 2 (à gauche) et un construit contenant quatre répétitions de la xbox 2 (à droite) par la protéine RFX6 de souris dans les cellules HEK293T. **(C)** Essai de transactivation testant l'activation du plasmide rapporteur *mMlxipl* intron 1 0,6 kb par la protéine RFX6 humaine sauvage et avec cinq mutations ponctuelles (la carte indique leur position dans la protéine) dans les cellules HEK293T. Les résultats **(B, C)** sont exprimés en moyenne plus l'écart type. Les différences statistiques (ns non significative, p <0,05 *, p <0,001 ***, p <0,0001 ****) ont été déterminées par ANOVA.

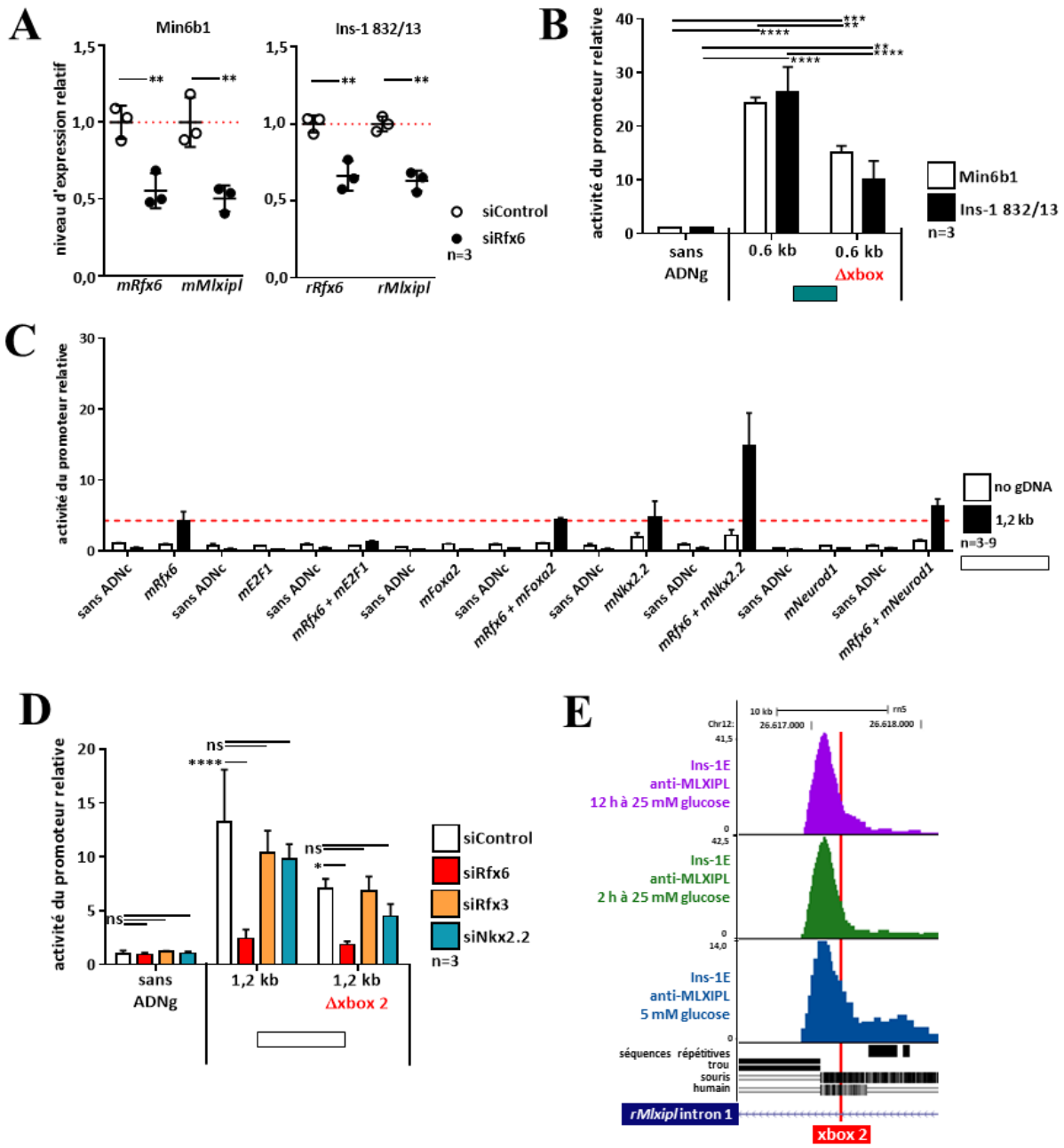


FIGURE E3 : Analyse des cofacteurs qui pourraient être impliqués dans la régulation transcriptionnelle de *mMlxipl* par RFX6 dans les cellules bêta. (A) Niveau d'expression (analyse par RT-qPCR) de *Mlxipl* dans les cellules Min6b1 (Perrine Strasser) et Ins-1 832/13 traitées avec un siRfx6. **(B)** Essai de transactivation testant l'activation du plasmide rapporteur *mMlxipl* intron 1 0,6 kb sans ou avec la xbox 2 dans les cellules Min6b1 et Ins-1 832/13. **(C)** Essai de transactivation testant l'activation du plasmide rapporteur *mMlxipl* intron 1 1.2 kb par mE2F1, mFOXA2, mNKX2.2 et mNEUROD1 en présence ou absence de mRFX6 dans les cellules HEK293T. **(D)** Essai de transactivation testant l'activation du plasmide rapporteur *mMlxipl* intron 1 1,2 kb sans ou avec la xbox 2 dans les cellules Ins-1 832/13 traitées avec un siRfx6, un siRfx3 ou un siNkx2.2. Les résultats **(A-D)** sont exprimés en moyenne plus l'écart type. Les différences statistiques (ns non significative, p <0,05 *, p< 0,01 **, p <0,001 ***, p <0,0001 ****) ont été déterminées par ANOVA **(A, D)** ou test t **(B)**. **(E)** Fixation de mMLXIPL sur *mMlxipl* dans les Ins-1E détectée par séquençage ChIP (données générées dans les cellules Ins-1E après une privation de 24 h dans du glucose à 5 mM (en bleu), ainsi qu'après une incubation ultérieure dans du glucose à 25 mM pendant 2 h (en vert) ou 12 h (en violet), GSE81628, Schmidt *et al.*, 2016, alignement par Constance Vagne). La position de la xbox 2 est marquée en rouge.

Étant donné que tous les essais de transactivation ont été réalisés en introduisant l'ADNc *Rfx6* dans la lignée HEK293T, la question se posait si ce mécanisme de transactivation de *Mlxipl α* par RFX6 pouvait être reproduit dans une lignée des cellules bêta qui expriment *Rfx6* de façon endogène. Dans notre laboratoire, nous avons vérifié que les lignées cellulaires bêta de souris Min6b1 (Lilla *et al.,* 2003) et de rat Ins-1 832/13 (Hohmeier *et al.,* 2000) expriment le facteur de transcription RFX6 et que la réduction de *Rfx6* par RNA interférence cause une diminution de *Mlxipl* (**FIGURE E3A**) comme observé dans les îlots *mRfx6* Δbêta (données non publiées de Perrine Strasser). L'activité luciférase des plasmides rapporteurs contenant les fragments de l'intron 1 a effectivement été activée de manière significative après transfection dans ces deux lignées. De plus, la délétion de la xbox conservée diminuait l'activité de la luciférase (**FIGURE E3B**).

D'autres études ont montré que plusieurs facteurs de transcription se fixaient également dans la région de l'intron 1. Il s'agit des facteurs de transcription FOXA1 et FOXA2 (Gao *et al.,* 2010), NKX2.2 et NEUROD1 (Churchill *et al.,* 2017). Par ailleurs, des analyses d'occurrence de motif avaient été mises en place pendant la thèse de Perrine Strasser afin d'identifier des motifs de liaison qui occurrent fréquemment dans les pics trouvés dans les trois manipulations de séquençage ChIP, suggérant que les protéines associées avec ces motifs corégulent des gènes avec RFX6. Un facteur de transcription identifié grâce à ces analyses, parmi d'autre, était E2F1 qui est impliqué dans le développement du pancréas et la régulation de la production d'insuline par les cellules bêta (Fajas *et al.,* 2004). En outre, nous avons considéré RFX3, un autre membre de la famille RFX, qui forme des dimères *in vitro* avec RFX6 (Smith *et al.,* 2010), comme cofacteur ou même partenaire de dimérisation de RFX6. Tous ces résultats et observations suggèrent une co-régulation de la transcription de *Mlxipl*, une hypothèse que nous avons testée.

Dans un premier temps, des tests de transactivation dans lesquels nous avons coexprimé chacun de ces facteurs avec RFX6 (**FIGURE E3C**) nous ont permis de révéler qu'E2F1 fonctionnait en tant que corépresseur et NKX2.2 en tant que coactivateur dans la régulation transcriptionnelle de *mMlxipl* par RFX6, alors que nous n'avons observé aucune contribution de FOXA2 et NEUROD1 dans ces tests, bien que les deux soient connus pour se lier à l'intron 1 de *Mlxipl* dans les îlots de souris et que la double inhibition de mFoxa1 et de mFoxa2 dans les îlots murins ait considérablement réduit le niveau de transcription de *mMlxipl* (Gao *et al.,* 2010 ; Churchill *et al.,* 2017). Ceci suggère que les données produites *in vivo* ne sont pas nécessairement reproductibles par les résultats des tests de transactivation.

En utilisant la lignée cellulaire Ins-1 832/13 comme modèle, nous avons déterminé les conséquences de l'inhibition de *rRfx6*, *rRfx3* et de *rNkx2.2* sur l'expression et la transactivation de *mMlxipl*. Comparé à *rRfx6*, dont l'inactivation a entraîné une forte diminution de la transactivation de

mMlxipl (**FIGURE E3D**) et de l'expression de *rMlxipl* α, l'invalidation de *rRfx3* et de *rNkx2.2* n'a pas eu d'effet sur la transactivation (**FIGURE E3D**). De plus, dans les cellules traitées par le siRfx3, l'expression de *rMlxipl* α n'était que légèrement réduite à un taux de glucose faible, tandis que le traitement par siNkx2.2 n'avait aucune conséquence sur l'expression de *rMlxipl* α. Dans l'ensemble, nos données suggèrent que rRFX6 est le facteur limitant de l'activation de la transcription et de l'expression de *rMlxipl* dans des cellules ins-1 832/13. Cependant, nous manquons des données supplémentaires pour tirer des conclusions sur d'autres lignées cellulaires, îlots et espèces.

De plus, avec la découverte du pic de rMLXIPL sur l'intron 1 de *rMlxipl* (**FIGURE E3E**), nous avons très probablement découvert un chaînon manquant dans le mécanisme de l'autorégulation transcriptionnelle de *Mlxipl*. En effet, l'induction de *Mlxipl* β par MLXIPL α après un stimulus au glucose est suivi par une diminution de la transcription de *Mlxipl* α mais aucune étude a pu expliquer le mécanisme exact (Herman *et al.,* 2012 ; Zhang *et al.,* 2015 ; Jing *et al.,* 2016). Sachant que l'intron 1 est crucial pour la régulation transcriptionnelle de *Mlxipl* α, il nous était facile de trouver le pic dans le premier intron de *Mlxipl* dans des données de séquençage ChIP publiées (Schmidt *et al.,* 2016).

En conclusion, à partir de l'analyse combinée de nos nouvelles données et des résultats publiés, nous avons élaboré le modèle théorique suivant concernant la régulation transcriptionnelle de *Mlxipl* dans les cellules bêta : Les facteurs de transcription pionniers FOXA1 et FOXA2 ouvrent la chromatine au locus *Mlxipl* permettant la fixation d'autres facteurs de transcription tels que RFX6, NKX2.2, E2F1, RFX3, NEUROD1 et MLXIPL. RFX6, NKX2.2 et RFX3 qui activent la transcription de *Mlxipl* α tandis que E2F1 et MLXIPL agissent en tant qu'inhibiteurs. Dans l'ensemble, toutefois, nos données chez la souris, le rat et l'homme ont clairement démontré que RFX6 est le facteur de transcription essentiel qui assure la régulation de *Mlxipl*.

3.2. Inactivation de *Rfx6* et *Mlxipl* par CRISPR/Cas9 dans la lignée cellulaire Min6b1

Afin d'identifier les cibles respectives de RFX6 et MLXIPL dans les cellules bêta adultes, notre plan expérimental consistait à invalider *Rfx6* et *Mlxipl* par la technique CRISPR/Cas9 dans une lignée cellulaire bêta. Dans un deuxième temps, le transcriptome des lignées « sauvages » et mutées devait être étudié par séquençage de l'ARN. Nous avons initialement choisi de travailler avec la lignée murine Min6b1 car c'est dans cette lignée que nous avions au préalable déterminé par séquençage ChIP les régions du génome qui lient RFX6 (Piccand *et al.,* 2014).

Nous avons réussi à établir des clones knockout (KO) pour les deux gènes. L'utilisation de la Cas9 nickase avec deux ARNs guides ciblant le codon d'initiation, permet l'introduction de petites insertions et délétions dans le locus voulu. Ce qui mènent à un décalage du cadre de lecture produisant une protéine tronquée par la création d'un codon stop prématuré (Jinek *et al.,* 2012 ; Ran *et al.,* 2013a ;

Ran *et al.,* 2013b). L'invalidation de deux allèles a été vérifiée par séquençage de Sanger et western blot. Cependant, l'inactivation de RFX6 ne diminuait pas l'expression de *Mlxipl* comme attendu et comme observé notamment lors d'expériences d'ARN interférence ciblant *Rfx6* dans la même lignée (**FIGURE E4A**). A ce jour, nous n'avons pas d'explication pour ce résultat, mais dans ces conditions, les lignées *Rfx6* KO générées ne pouvaient pas être utilisées pour répondre à la question posée. Concernant les clones KO pour *Mlxipl*, il nous manquait une cible irréfutable de ce facteur dans les cellules bêta pour tester l'effet de son invalidation. La plupart des cibles de MLXIPL ont en effet été identifiées dans le foie et les adipocytes. De plus, il a été découvert récemment que des gènes proposés initialement comme cibles de MLXIPL, sont en fait régulés par son paralogue MLXIP dans les cellules bêta (Richards *et al.,* 2018). Nous avons finalement testé le gène *Pklr* encodant une enzyme de la glycolyse. Alors que MLXIPL active l'expression de *Pklr* en présence d'une concentration élevée de glucose dans les Min6 (Da Silva Xavier *et al.,* 2006), nous n'avons pas pu révéler une telle régulation dans nos clones (**FIGURE E4B**). En résumé, nous n'avons pas pu confirmer les knockouts de *Rfx6* et *Mlxipl* par une analyse fonctionnelle des gènes cibles candidats.

Les résultats inattendus décrits ci-dessus nous ont fait suspecter que la lignée Min6b1 utilisée avait pu dériver et perdre ses caractéristiques. Par conséquent, nous avons testé la réponse transcriptionnelle au glucose. Nous avons constaté que le seul transcrit, glucose-dépendant, parmi les transcrits testés, était *Txnip* dont l'expression était doublée. Le glucose, mais également le stress oxydant, sont connus pour activer la transcription de *Txnip* (Shalev, 2014, revue). En revanche, le gène *disallowed Ldha*, normalement réprimé dans les cellules bêta matures (Sekine *et al.,* 1994), était fortement exprimé dans les Min6b1. En parallèle, nous avons réalisé le même test dans la lignée bêta de rat, Ins-1 832/13, et observé une forte induction de *Txnip* de 70 fois et l'absence du transcrit *Ldha* suggérant une meilleure réponse au glucose (**FIGURE E5A**).

Afin d'évaluer plus avant l'utilisation de la lignée Ins-1 832/13 plutôt que Min6b1 pour notre étude, nous nous sommes penchés sur la régulation transcriptionnelle de *Mlxipl* β par le glucose. En effet, *Mlxipl* β a été décrit comme une cible de MLXIPL α (Herman *et al.,* 2012) et pourrait donc servir d'indicateur pour évaluer l'effet d'un knockout de *Mlxipl* α. À cette fin, nous avons mimé l'essai de transactivation décrit par Herman et collègues (Herman *et al.,* 2012) qui est basé sur l'intéraction de MLXIPL avec deux ChoRE dans l'exon 1b de *Mlxipl* β, en clonant un fragment de 400 pb en amont de la luciférase (**FIGURE E5B**). Dans les HEK293T, nous avons pu reproduire l'activation de cette construction par les ADNc *Mlxipl α* ou *Mlxipl* β plus *Mlx.* Comme attendu, cette activation est dépendante du glucose pour la combinaison *Mlxipl α* / *Mlx* (Herman *et al.,* 2012). Dans les Ins-1 832/13, la forte induction glucose-dépendante du rapporteur a été parfaitement reproductible (Zhang *et al.,* 2015). De façon surprenante, la transactivation du rapporteur luciférase dans les Min6b1 était indépendante du glucose (**FIGURE E5B**). Par ailleurs, *Mlxipl* β n'a pas pu être détecté dans les

Min6b1 et l'expression de *Mlxipl α* n'est pas affectée par le glucose. En revanche, dans les Ins-1 832/13, nous avons bien observé une diminution de *Mlxipl α* et une augmentation de *Mlxipl* β dans les cellules cultivées dans le milieu riche en glucose (**FIGURE E5A**) comme décrit dans la littérature (Zhang *et al.,* 2015).

L'ensemble de ces tests nous a amené à conclure que le modèle Min6b1 n'était pas adéquat pour l'identification des gènes cibles de RFX6 et MLXIPL et notamment pour l'analyse du rôle du glucose dans leur activité transcriptionnelle. Par ailleurs, il a été montré qu'une perte de l'identité cellulaire des Min6b1 ainsi qu'une diminution de la capacité de sécrétion de l'insuline se déclarait après cinq passages (Rani *et al.,* 2010) ; l'établissement des lignées monoclonales après l'inactivation d'un gène nécessitait au moins quatre passages sans compter l'amplification des cellules. Contrairement aux Min6b1, les Ins-1 832/13 semblent maintenir leur identité pour 90 passages voire une période encore plus longue (Hohmeier *et al.,* 2000). Donc, nous avons pris la décision de recommencer la génération des clones knockout dans cette deuxième lignée.

3.3. Inactivation de *Rfx6* et *Mlxipl* par CRISPR/Cas9 dans la lignée cellulaire Ins-1 832/13

Pour l'invalidation de *rRfx6* et *rMlxipl* dans les Ins-1 832/13, nous avons adapté la stratégie expérimentale que nous avions utilisée dans les Min6b1 et qui avait été très efficace dans cette première lignée, or elle ne l'était plus dans les Ins-1 832/13. En effet, nous avons réussi à inactiver une copie de *Rfx6* mais le deuxième allèle restait intact malgré plusieurs tentatives d'inactivation. Il se peut que la perte totale de *Rfx6* soit létale dans les Ins-1 832/13. Des mutations hétérozygotes de *RFX6* ont été associées à un diabète chez les jeunes (Patel *et al.,* 2017), suggérant un phénotype hétérozygote chez l'homme, qui n'a pas été rapporté chez la souris. Nous avons néanmoins testé l'effet de l'hétérozygotie *Rfx6* dans les Ins-1 832/13 sur l'expression de *Mlxipl*. Aucun clone ne présentait une réduction de *Mlxipl* (**FIGURE E4C**). Il semblerait donc que la quantité de RFX6 dans les hétérozygotes soit suffisante pour éviter une altération de la régulation transcriptionnelle de *Mlxipl*.

Concernant l'inactivation de *Mlxipl*, nous avons constaté que les profils de séquençage du locus muté comportaient trois (au lieu de deux) séquences voire plus qui se chevauchaient. Par ailleurs, la protéine MLXIPL est détectée par western blot dans tous les clones testés (**FIGURE E4D**). En résumé, nous n'avons obtenu aucun clone homozygote. Une étude récente décrit un caryotype anormal de la lignée Ins-1 832/3 (Naylor *et al.,* 2016), donc il n'est pas exclu qu'il y ait une altération génomique du locus *Mlxipl* ce qui pourrait expliquer nos résultats, combinés à une difficulté à obtenir des clones purs. Une étude récemment publiée visant à obtenir des clones knockout par une approche CRISPR/Cas9 dans les Ins-1 832/13 décrit de fortes variations phénotypiques pour le même génotype (Peterson *et al.,* 2018). Les auteurs en concluent que la lignée Ins-1 832/13 reste une population hétérogène même

après plusieurs sous clonages. En conclusion, devant les difficultés rencontrées à générer des clones KO pour *Rfx6* et *Mlxipl* par la méthode CRISPR/Cas9, nous avons décidé d'inactiver ces gènes par ARN interférence dans les cellules Ins-1 832/13 pour identifier leurs cibles.

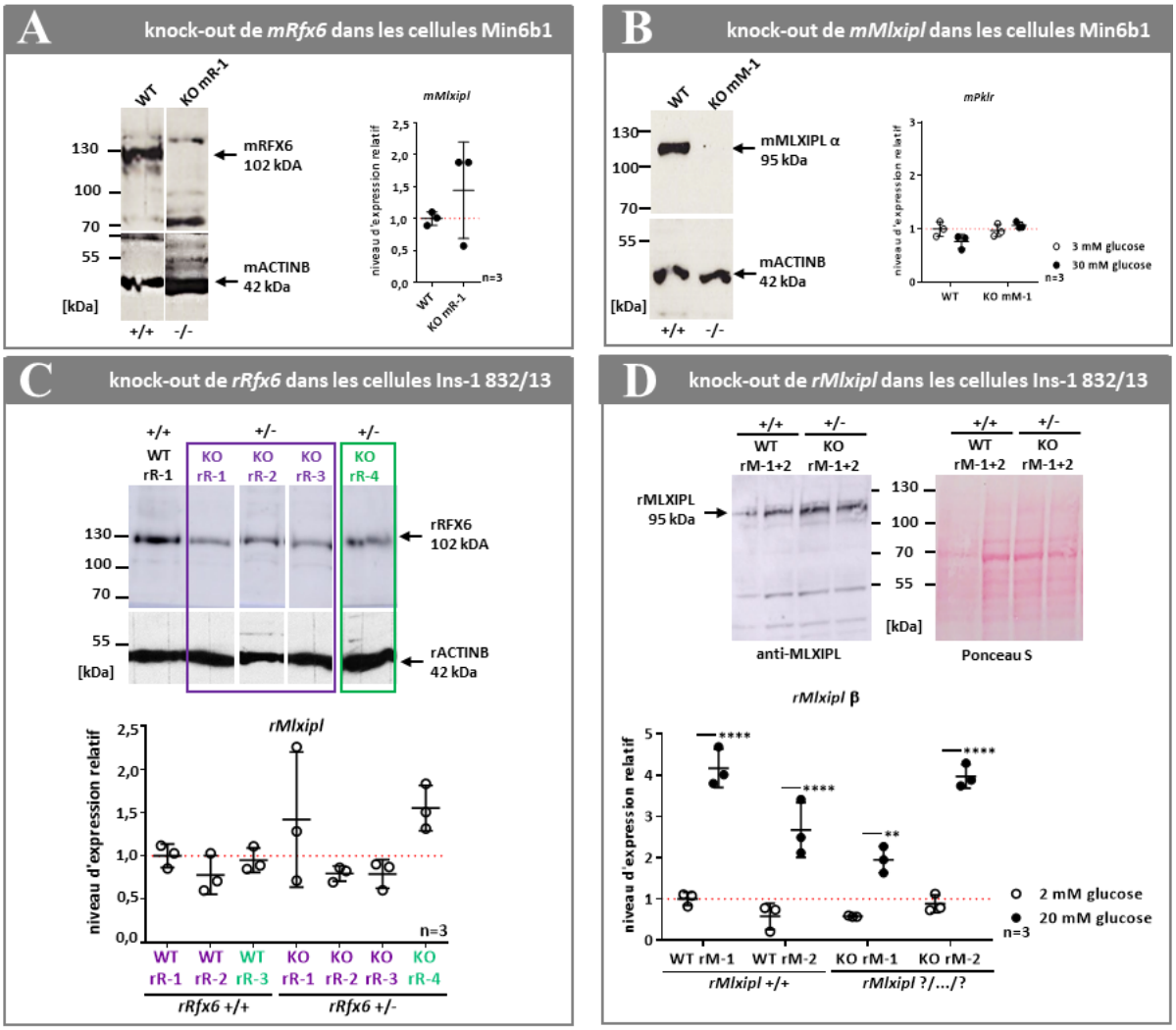


FIGURE E4 : L'expression des gènes cibles de RFX6 et MLXIPL n'est pas affectée par l'inhibition homozygote dans les cellules Min6b1, ni par l'inhibition hétérozygote dans les cellules Ins-1 832/13. (A, B) À gauche : Western blot testant l'expression de mRFX6 dans le clone Min6b1 avec un knock-out homozygote du gène *mRfx6* KO mR-1 **(A)** et mMLXIPL dans le clone Min6b1 avec un knock-out homozygote du gène *mMlxipl* KO mM-1 **(B)**. À droite : Niveau d'expression (analyse par RT-qPCR) de *mMlxipl* dans le clone Min6b1 avec un knock-out homozygote du gène *mRfx6* KO mR-1 **(A)** et de *mPklr* dans le clone Min6b1 avec un knock-out homozygote du gène *mMlxipl* KO mM-1 après une incubation de 16 h dans un milieu soit faible (3 mM) soit riche (30 mM) en glucose **(B)**. **(C, D)** En haut : Western blot testant l'expression de rRFX6 dans les clones Ins-1 832/13 avec des knock-out hétérozygotes du gène *rRfx6* KO rR-1 à KO rR-4 **(A)** et rMLXIPL dans les clones Ins-1 832/13 avec des knock-out hétérozygotes / hétérogènes du gène *rMlxipl* KO rM-1 et KO rM-2 **(B)**. En bas : Niveau d'expression (analyse par RT-qPCR) de *rMlxipl* dans les clones Ins-1 832/13 avec des knock-out hétérozygotes du gène *rRfx6* KO rR-1 à KO rR-4 **(A)** et de *rMlxipl* β dans les clones Ins-1 832/13 avec des knock-out hétérozygotes / hétérogènes du gène *rMlxipl* KO rM-1 et KO rM-2 après une incubation de 6 h dans un milieu soit faible (2 mM) soit riche (20 mM) en glucose **(B)**. Les résultats **(A, B, C, D)** sont exprimés en moyenne plus l'écart type. Les différences statistiques (p <0,001 ***, p <0,0001 ****) ont été déterminées par ANOVA **(D)**.

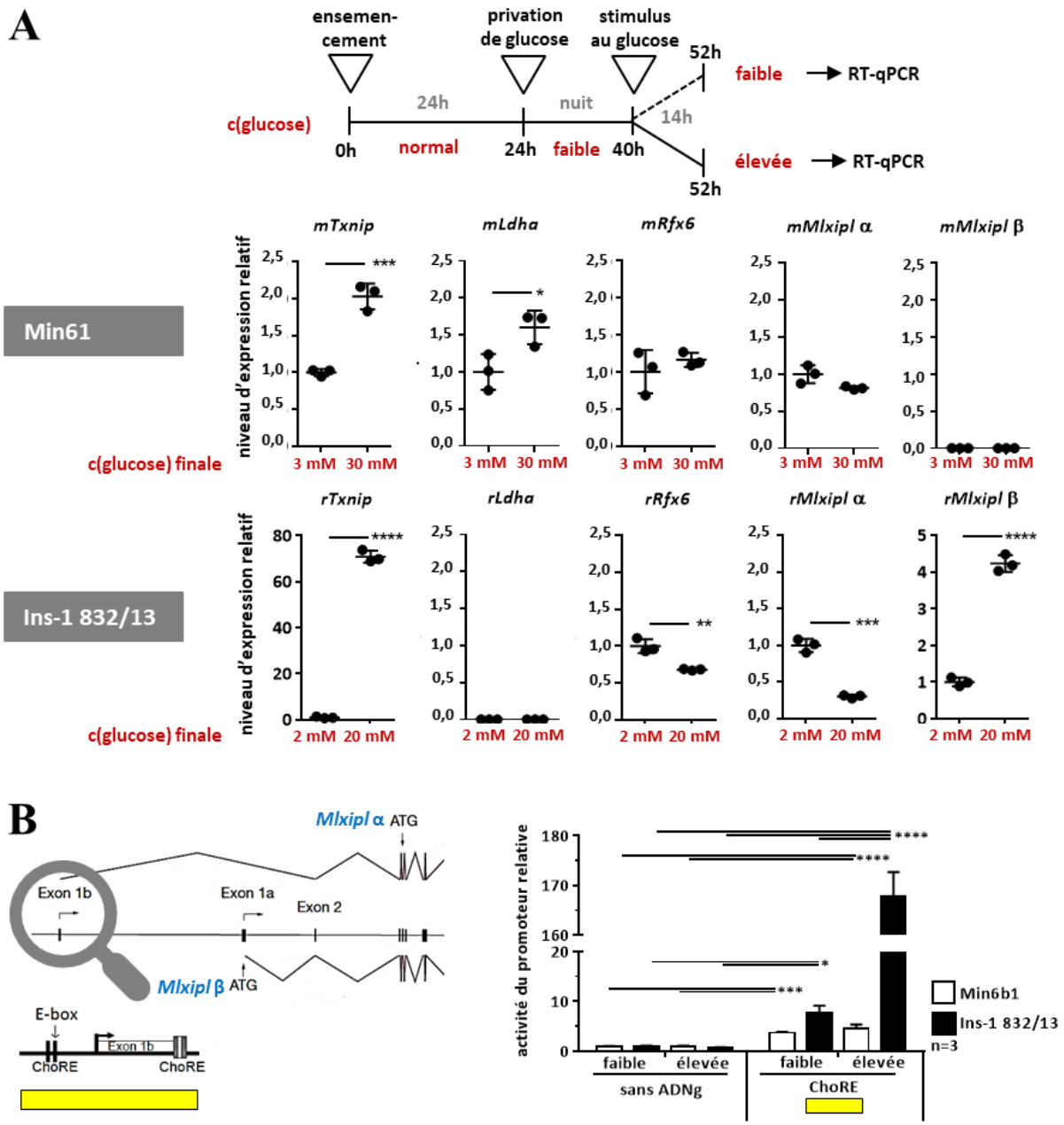


FIGURE E5 : La transcription dépendante du glucose et l'activation transcriptionnelle de *Mlxipl* β sont différentes entre les cellules Min6b1 et Ins-1 832/13. (A) En haut : Stratégie expérimentale. Les cellules Min6b1 et Ins-1 832/13 ont été incubées dans un milieu faible en glucose (3 mM pour le Min6b1, 2 mM pour les Ins-1 832/13) pendant une nuit avant l'incubation dans un milieu soit faible (3 mM pour le Min6b1, 2 mM pour les Ins-1 832/13) soit riche (30 mM pour le Min6b1, 20 mM pour les Ins-1 832/13) en glucose pendant 14 h. L'ARN a été extrait et l'expression a été analysée par RT-qPCR. En bas : Niveaux d'expression de *Txnip*, *Ldha*, *Rfx6*, *Mlxipl* α et *Mlxipl* β dans les cellules Min6b1 et Ins-1 832/13. **(B)** À gauche : Modèle représentant l'épissage de *mMlxipl* α et β et les deux éléments de réponse glucidiques (ChoRE) en amont et dans *Mlxipl* exon 1b sur lesquelles MLXIPL se fixe (figures modifiées d'après Herman *et al.*, 2012 et Zhang *et al.*, 2015). À droite : Essai de transactivation testant l'activation du plasmide rapporteur *mMlxipl* exon 1b dans les cellules Min6b1 et Ins-1 832/13 incubées dans un milieu faible en glucose (3 mM pour le Min6b1, 2 mM pour les Ins-1 832/13) pendant 24 h et dans un milieu soit faible (3 mM pour le Min6b1, 2 mM pour les Ins-1 832/13) soit riche en glucose (30 mM pour le Min6b1, 20 mM pour les Ins-1 832/13) pendant 24 h. Les résultats **(A, B)** sont exprimés en moyenne plus l'écart type. Les différences statistiques (p<0,05 *, p <0,01 **, p <0,001 ***, p <0,0001 ****) ont été déterminées par test t **(A)** et ANOVA **(B)**.

3.4. Inactivation de *Rfx6* et *Mlxipl* par l'ARN interférence dans la lignée cellulaire Ins-1 832/13

L'ARN interférence nous a permis de diminuer l'expression de *Rfx6* de 50 % au maximum dans la lignée Ins-1 832/13. Ce niveau de réduction correspond aux expériences réalisées précédemment dans l'équipe, dans les lignées bêta murines Min6b1 et humaines EndoC-bêtaH2 (données non-publiées de Perrine Strasser). D'autre part, cette réduction de *Rfx6* est suffisante pour diminuer l'expression de *Mlxipl* suggérant que les cibles de *Mlxipl* pourraient aussi être affectées. Pour *Mlxipl*, nous avons testé deux siRNA qui ciblent les isoformes α et β. Un de ce deux siRNA a été décrit par Schmidt et collègues (Schmidt *et al.,* 2016) ; il était plus efficace et des réductions de l'ordre de 50% des transcrits *Mlxipl* ont été obtenues. Néanmoins, ce siRNA s'est avéré très toxique pour la cellule (mort cellulaire ; **FIGURE E6B**) à des concentrations 10 fois plus faibles que celles utilisées par Schmidt et collègues, nécessitant une optimisation de sa concentration. En résumé, l'approche ARN interférence a permis une diminution de l'ordre de 50% pour les deux gènes ciblés.

Nous avons préparé les ARN des cellules Ins-1 832/13 traitées avec les siRfx6 et siMlxipl et stimulées au glucose (2 et 20 mM de glucose). Nous avons validé nos échantillons en mesurant les transcrits *rRfx6*, *rMlxipl* et *rTxnip* par RT-qPCR. Ainsi, nous avons pu confirmer une réduction de *rRfx6* et *rMlxipl* par l'ARN interférence et une induction significative de *rTxnip* par le glucose comme attendu. Comme observé auparavant, nous avons pu à nouveau constater dans cette manipulation que l'expression de *rRfx6* est inhibée par le glucose (**FIGURE E6A**).

3.5. Identification des gènes cibles de RFX6 et MLXIPL dans les Ins-1 832/13

Le séquençage de l'ARN a été accompli par la plateforme GenomEast de l'IGBMC et l'analyse bio-informatique a été réalisée dans l 'équipe afin d'identifier les gènes régulés respectivement par *Mlxipl* et *Rfx6* dans ce modèle de cellules bêta et de déterminer l'effet du glucose.

Premièrement, grâce aux nouvelles données et aux résultats du séquençage ChIP publiés par Schmidt *et al.* (GSE81628, Schmidt *et al.,* 2016), nous avons pu déduire que l'expression diminuée de *rRfx6* après une incubation dans un milieu riche en glucose (**FIGURE E6A**) est due à une inhibition directe de *rRfx6* par rMLXIPL (**FIGURE E6C**), indiquant un contrôle de rétroaction, puisque RFX6 est l'activateur de la transcription de *Mlxipl* α. Deuxièmement, l'analyse nous a permis de comprendre pourquoi le siRNA qui cible *rMlxipl* pourrait avoir un effet toxique sur les cellules, qui s'intensifie lorsque la concentration en glucose est faible. La réduction de *rMlxipl* causait une forte diminution du gène glycolytique *rPklr* altérant probablement négativement l'efficacité de la glycolyse et de la production d'énergie. Il en résulte que l'énergie fournie dans un milieu pauvre en glucose n'était pas suffisante pour assurer la survie des cellules.

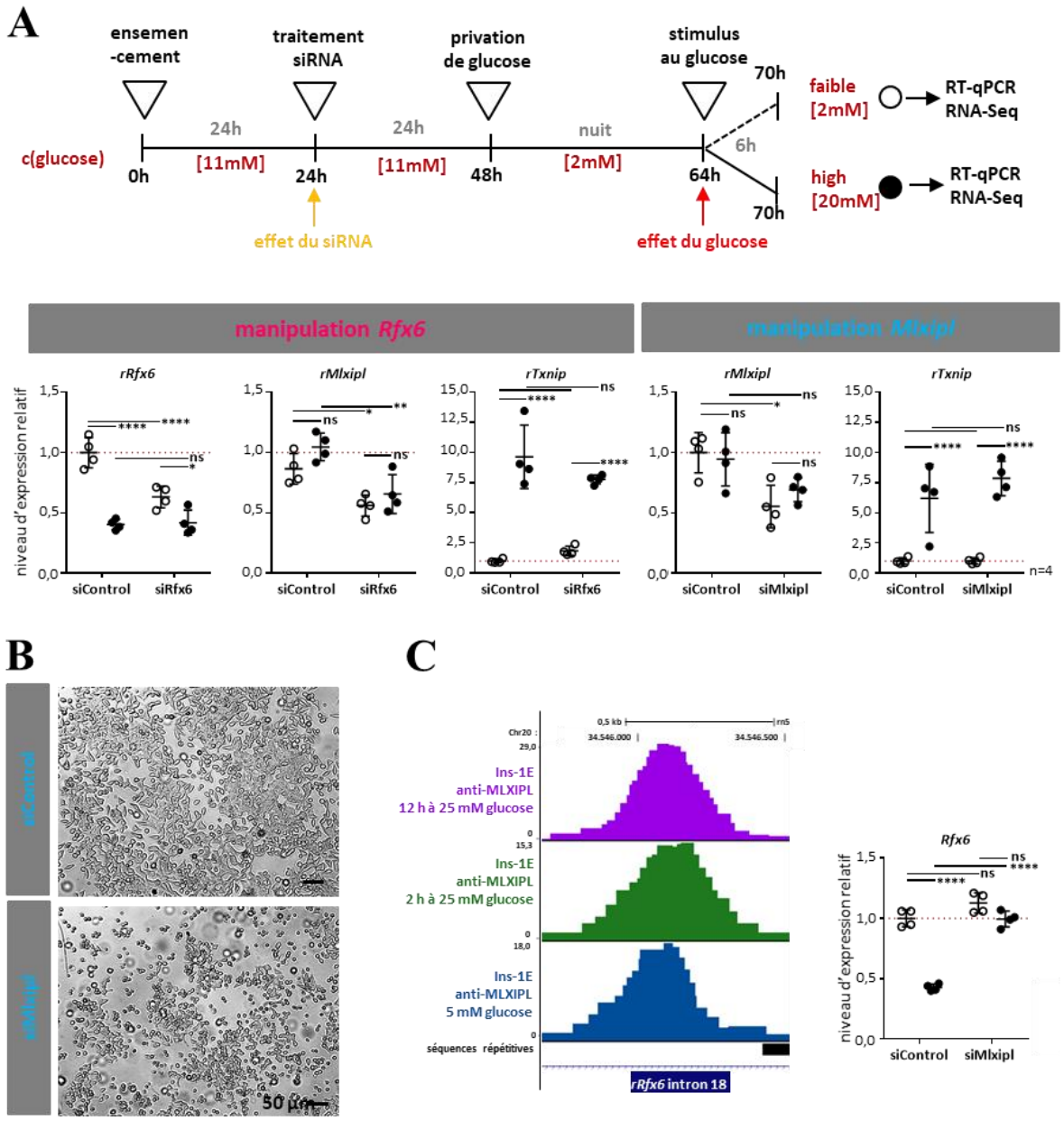


FIGURE E6 : La méthode d'ARN interférence permet une réduction de moitié de l'expression de *rRfx6* et de *rMlxipl* dans les cellules Ins-1 832/13. (A) En haut : Stratégie expérimentale. Les cellules Ins-1 832/13 ont été traitées par un siRfx6 ou par un siMlxipl, et après 48 h, ont été incubées dans un milieu faible en glucose (2 mM) pendant une nuit avant l'incubation dans un milieu soit faible (2 mM) soit riche (20 mM) en glucose pendant 6 h. L'ARN a été extrait et l'expression a été analysée par RT-qPCR et séquençage de l'ARN. En bas : Niveaux d'expression de *rRfx6*, *rMlxipl* et *rTxnip* dans les cellules Ins-1 832/13 traitées par siRfx6 (à gauche) ou siMlxipl (à droite). Les résultats sont exprimés en moyenne plus l'écart type. Les différences statistiques (ns non-significatif, p<0,05 *, p <0,01 **, p <0,001 ***, p <0,0001 ****) ont été déterminées par ANOVA. **(B)** Photos montrant la mortalité cellulaire après le traitement des Ins-1 832/13 avec le siMlxipl dans un milieu faible en glucose (2 mM). **(C)** À gauche : Fixation de mMLXIPL sur *mRfx6* dans les Ins-1E détectée par séquençage ChIP (données générées dans les cellules Ins-1E après une privation de 24 h dans du glucose à 5 mM (en bleu), ainsi qu'après une incubation ultérieure dans du glucose à 25 mM pendant 2 h (en vert) ou 12 h (en violet), GSE81628, Schmidt *et al.*, 2016, alignement par Constance Vagne). À droite : Niveau d'expression de *mRfx6* dans les Ins-1 832/13 traitées avec un siMlxipl et incubées pendant 6 h dans un milieu soit faible (2 mM) soit riche (20 mM) en glucose (expression mesurée par le séquençage de l'ARN).

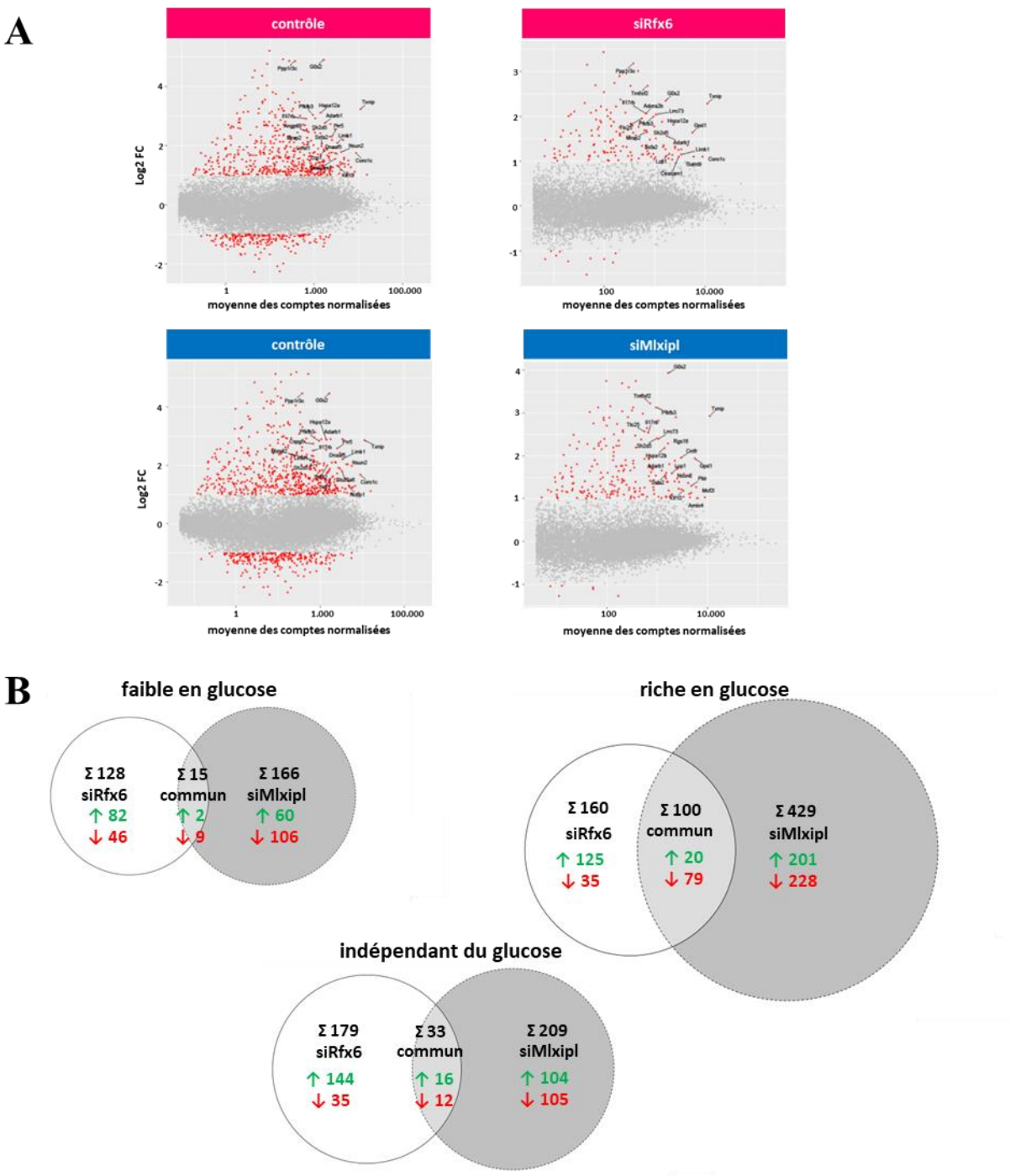


FIGURE E7 : Nombre de gènes exprimés de manière différentielle dans les cellules Ins-1 832/13 contrôles et traitées par siRfx6 ou siMlxipl en fonction de la concentration en glucose. (A) Diagrammes MA (Constance Vagne) représentant le Log2 FC (glucose élevé par rapport au glucose faible) en fonction de la moyenne des comptes normalisés des données de séquençage de l'ARN générées dans les Ins-1 832/13 traitées avec un siControl, siRfx6 ou siMlxip et incubées pendant 6 h dans un milieu soit faible (2 mM) soit élevé (20 mM) en glucose (● signifie un gène dont l'expression ne change pas de manière significative (p≥0,05 et / ou -1≤Log2 FC≤1), ● gène significativement activé (p<0,05 et Log2 FC>1) ou réprimé (p<0,05 et Log2 FC<-1)). **(B)** Schémas (dessinés avec eulerr.co) montrant les gènes différentiellement exprimés (-1< Log2 FC <1 et p <0,05, données de séquençage de l'ARN) dans les Ins-1 832/13 traitées avec un siControl, siRfx6 ou siMlxip et incubées pendant 6 h dans un milieu soit faible (2 mM) soit élevé (20 mM) en glucose et les gènes qui ont été dérégulés indépendamment du glucose. Le nombre de gènes dérégulés est marqué en noir. Les gènes régulés positivement (↑) sont marqués en vert et les gènes régulés négativement (↓) en rouge.

Globalement, nos données ont montré que MLXIPL en aval de RFX6 est important pour l'exécution de la réponse au glucose. L'inhibition directe de MLXIPL par le traitement siMlxipl ou sa répression indirecte par l'inactivation de *rRfx6* ont fortement réduit le nombre de gènes activés et réprimés par le glucose (**FIGURE E7A**). Ainsi, nous avons conclu que MLXIPL contrôle l'activation et l'inhibition des gènes après un stimulus au glucose. D'une part, ce résultat semble plausible, puisque MLXIPL est un facteur de transcription activé par le glucose (Li *et al.,* 2006 ; Herman *et al.,* 2012), d'autre part, une étude publiée en 2018 a suggéré que MLXIP pourrait être plus important que MLXIPL dans les cellules bêta adultes (Richards *et al.,* 2018). Par exemple, ils ont rapporté que le gène *Txnip* activé par le glucose et soupçonné depuis longtemps d'être un gène directement ciblé par MLXIPL (Cha-Molstad *et al.,* 2009 ; Poungvarin *et al.,* 2012), est plutôt régulé par MLXIP (Richards *et al.,* 2018). Bien que nous ayons confirmé que *rTxnip* n'est pas régulé par rMLXIPL dans les cellules Ins-1 832/13, nos données suggèrent toutefois que rMLXIPL joue un rôle important dans la régulation des gènes dont l'expression est dépendante du glucose dans les cellules bêta adultes.

En outre, les données de séquençage de l'ARN ont révélé que RFX6 est plutôt un répresseur transcriptionnel qu'un activateur transcriptionnel (**FIGURE E7B**). L'activité de la protéine ne semble pas être altérée par le glucose, mais, comme mentionné, son expression diminue lorsque la concentration de glucose est élevée (**FIGURE E7A**). MLXIPL est à la fois un activateur et un répresseur de transcription et, conformément aux études publiées, son activité était plus élevée en hyperglycémie qu'en hypoglycémie (Li *et al.,* 2006 ; Herman *et al.*, 2012 ; **FIGURE E7B**). Le nombre de gènes cibles communs a été beaucoup plus faible qu'attendu (**FIGURE E7B**), ce qui est probablement dû à la stratégie expérimentale et l'inactivation incomplète des deux facteurs de transcription. Notre étude était la première à tester les effets de l'inactivation de *Rfx6* et de *Mlxipl* dans des conditions comparables. Il se peut que la perte de RFX6 n'ait pas les mêmes conséquences directement ou longtemps après un stimulus au glucose. Ceci reste à analyser.

4. Conclusion

Nos résultats démontrent que le facteur de transcription RFX6 est essentiel pour la régulation transcriptionnelle directe de *Mlxipl* dans les cellules bêta pancréatiques et qu'un motif xbox dans l'intron 1 de *Mlxipl* est nécessaire et suffisant pour cette régulation. L'inactivation de *rRfx6* et de *rMlxipl* par RNA interférence dans les Ins-1 832/13 et les résultats du séquençage de l'ARN nous ont permis de déterminer les gènes directement régulés par RFX6 et ceux contrôlés par MLXIPL, et d'évaluer l'effet du glucose. Les résultats nous ont aidés à mieux comprendre les fonctions respectives de RFX6 et de MLXIPL et ont ouvert de nombreuses possibilités de recherche pouvant être explorées.

F) References

Abdul-Wahed, A., Guilmeau, S., and Postic, C. (2017). Sweet Sixteenth for ChREBP: Established Roles and Future Goals. Cell Metabolism 26, 324–341.

Adamson, A.W., Suchankova, G., Rufo, C., Nakamura, M.T., Teran-Garcia, M., Clarke, S.D., and Gettys, T.W. (2006). Hepatocyte nuclear factor-4α contributes to carbohydrate-induced transcriptional activation of hepatic fatty acid synthase. Biochemical Journal 399, 285–295.

Aftab, S., Semenec, L., Chu, J., and Chen, N. (2008). Identification and characterization of novel human tissue-specific RFX transcription factors. BMC Evolutionary Biology 8, 226.

Ait-Lounis, A., Baas, D., Barras, E., Benadiba, C., Charollais, A., Nlend Nlend, R., Liegeois, D., Meda, P., Durand, B., and Reith, W. (2007). Novel Function of the Ciliogenic Transcription Factor RFX3 in Development of the Endocrine Pancreas. Diabetes 56, 950–959.

Ait-Lounis, A., Bonal, C., Seguin-Estevez, Q., Schmid, C.D., Bucher, P., Herrera, P.L., Durand, B., Meda, P., and Reith, W. (2010). The Transcription Factor Rfx3 Regulates β-Cell Differentiation, Function, and Glucokinase Expression. Diabetes 59, 1674–1685.

Amendt, B.A., Sutherland, L.B., and Russo, A.F. (1999). Transcriptional Antagonism between Hmx1 and Nkx2.5 for a Shared DNA-binding Site. Journal of Biological Chemistry 274, 11635–11642.

Andrali, S.S., Sampley, M.L., Vanderford, N.L., and Özcan, S. (2008). Glucose regulation of insulin gene expression in pancreatic β-cells. Biochemical Journal 415, 1–10.

Artuso, R., Provenzano, A., Mazzinghi, B., Giunti, L., Palazzo, V., Andreucci, E., Blasetti, A., Chiuri, R.M., Gianiorio, F.E., Mandich, P., et al. (2015). Therapeutic implications of novel mutations of the RFX6 gene associated with early-onset diabetes. The Pharmacogenomics Journal 15, 49–54.

Asfari, M., Janjic, D., Meda, P., Li, G., Halban, P.A., and Wollheim, C.B. (1992). Establishment of 2-mercaptoethanol-dependent differentiated insulin-secreting cell lines. Endocrinology 130, 167–178.

Ashcroft, F.M., and Rorsman, P. (2012). Diabetes Mellitus and the β Cell: The Last Ten Years. Cell 148, 1160–1171.

Avrahami, D., Li, C., Zhang, J., Schug, J., Avrahami, R., Rao, S., Stadler, M.B., Burger, L., Schübeler, D., Glaser, B., et al. (2015). Aging-Dependent Demethylation of Regulatory Elements Correlates with Chromatin State and Improved β Cell Function. Cell Metabolism 22, 619–632.

Bader, E., Migliorini, A., Gegg, M., Moruzzi, N., Gerdes, J., Roscioni, S.S., Bakhti, M., Brandl, E., Irmler, M., Beckers, J., et al. (2016). Identification of proliferative and mature β-cells in the islets of Langerhans. Nature 535, 430–434.

Batterham, R.L., Le Roux, C.W., Cohen, M.A., Park, A.J., Ellis, S.M., Patterson, M., Frost, G.S., Ghatei, M.A., and Bloom, S.R. (2003). Pancreatic Polypeptide Reduces Appetite and Food Intake in Humans. The Journal of Clinical Endocrinology & Metabolism 88, 3989–3992.

Benhamed, F., Denechaud, P.-D., Lemoine, M., Robichon, C., Moldes, M., Bertrand-Michel, J., Ratziu, V., Serfaty, L., Housset, C., Capeau, J., et al. (2012). The lipogenic transcription factor ChREBP dissociates hepatic steatosis from insulin resistance in mice and humans. Journal of Clinical Investigation 122, 2176–2194.

Benner, C., van der Meulen, T., Cacéres, E., Tigyi, K., Donaldson, C.J., and Huising, M.O. (2014). The transcriptional landscape of mouse beta cells compared to human beta cells reveals notable species differences in long non-coding RNA and protein-coding gene expression. BMC Genomics 15, 620.

Billin, A.N., Eilers, A.L., Coulter, K.L., Logan, J.S., and Ayer, D.E. (2000). MondoA, a Novel Basic Helix-Loop-Helix-Leucine Zipper Transcriptional Activator That Constitutes a Positive Branch of a Max-Like Network. Molecular and Cellular Biology 20, 8845–8854.

Blackshear, P.J., Graves, J.P., Stumpo, D.J., Cobos, I., Rubenstein, J.L.R., and Zeldin, D.C. (2003). Graded phenotypic response to partial and complete deficiency of a brain-specific transcript variant of the winged helix transcription factor RFX4. Development 130, 4539–4552.

Blodgett, D.M., Cura, A.J., and Harlan, D.M. (2014). The pancreatic β-cell transcriptome and integrated-omics: Current Opinion in Endocrinology & Diabetes and Obesity 21, 83–88.

Boergesen, M., Poulsen, L. la C., Schmidt, S.F., Frigerio, F., Maechler, P., and Mandrup, S. (2011). ChREBP Mediates Glucose Repression of Peroxisome Proliferator-activated Receptor α Expression in Pancreatic β-Cells. Journal of Biological Chemistry 286, 13214–13225.

Bompada, P., Atac, D., Luan, C., Andersson, R., Omella, J.D., Laakso, E.O., Wright, J., Groop, L., and De Marinis, Y. (2016). Histone acetylation of glucose-induced thioredoxin-interacting protein gene expression in pancreatic islets. The International Journal of Biochemistry & Cell Biology 81, 82–91.

Bonnafe, E., Touka, M., AitLounis, A., Baas, D., Barras, E., Ucla, C., Moreau, A., Flamant, F., Dubruille, R., Couble, P., et al. (2004). The Transcription Factor RFX3 Directs Nodal Cilium Development and Left-Right Asymmetry Specification. Molecular and Cellular Biology 24, 4417–4427.

Bramswig, N.C., Everett, L.J., Schug, J., Dorrell, C., Liu, C., Luo, Y., Streeter, P.R., Naji, A., Grompe, M., and Kaestner, K.H. (2013). Epigenomic plasticity enables human pancreatic α to β cell reprogramming. The Journal of Clinical Investigation 123, 1275–1284.

Cairo, S. (2001). WBSCR14, a gene mapping to the Williams-Beuren syndrome deleted region, is a new member of the Mlx transcription factor network. Human Molecular Genetics 10, 617–627.

Cano, D.A., Soria, B., Martín, F., and Rojas, A. (2014). Transcriptional control of mammalian pancreas organogenesis. Cellular and Molecular Life Sciences 71, 2383–2402.

Caron, S., Huaman Samanez, C., Dehondt, H., Ploton, M., Briand, O., Lien, F., Dorchies, E., Dumont, J., Postic, C., Cariou, B., et al. (2013). Farnesoid X Receptor Inhibits the Transcriptional Activity of Carbohydrate Response Element Binding Protein in Human Hepatocytes. Molecular and Cellular Biology 33, 2202–2211.

Carrano, A.C., Mulas, F., Zeng, C., and Sander, M. (2017). Interrogating islets in health and disease with single-cell technologies. Molecular Metabolism 6, 991–1001.

Castro, W., Chelbi, S.T., Niogret, C., Ramon-Barros, C., Welten, S.P.M., Osterheld, K., Wang, H., Rota, G., Morgado, L., Vivier, E., et al. (2018). The transcription factor Rfx7 limits metabolism of NK cells and promotes their maintenance and immunity. Nature Immunology 19, 809–820.

Cha, J.-Y., and Repa, J.J. (2007). The Liver X Receptor (LXR) and Hepatic Lipogenesis: The Carbohydrate-Response Rlement-Binding Protein is a Target Gene of LXR. Journal of Biological Chemistry 282, 743–751.

Chambers, K.T., Chen, Z., Lai, L., Leone, T.C., Towle, H.C., Kralli, A., Crawford, P.A., and Finck, B.N. (2013). PGC-1β and ChREBP partner to cooperatively regulate hepatic lipogenesis in a glucose concentration-dependent manner. Molecular Metabolism 2, 194–204.

Cha-Molstad, H., Saxena, G., Chen, J., and Shalev, A. (2009). Glucose-stimulated Expression of Txnip Is Mediated by Carbohydrate Response Element-binding Protein, p300, and Histone H4 Acetylation in Pancreatic Beta Cells. Journal of Biological Chemistry 284, 16898–16905.

Chandra, V., Albagli-Curiel, O., Hastoy, B., Piccand, J., Randriamampita, C., Vaillant, E., Cavé, H., Busiah, K., Froguel, P., Vaxillaire, M., et al. (2014). RFX6 Regulates Insulin Secretion by Modulating Ca2+ Homeostasis in Human β Cells. Cell Reports 9, 2206–2218.

Chappell, L., Gorman, S., Campbell, F., Ellard, S., Rice, G., Dobbie, A., and Crow, Y. (2008). A further example of a distinctive autosomal recessive syndrome comprising neonatal diabetes mellitus,

intestinal atresias and gall bladder agenesis. American Journal of Medical Genetics Part A 146A, 1713–1717.

Chen, C.Y., and Schwartz, R.J. (1992). Identification of Novel DNA Binding Targets and Regulatory Domains of a Murine Tinman Homeodomain Factor, nkx-2.5. The Journal of Biological Chemistry 270, 15628–15633.

Cheng, K., Delghingaro-Augusto, V., Nolan, C.J., Turner, N., Hallahan, N., Andrikopoulos, S., and Gunton, J.E. (2012). High Passage MIN6 Cells Have Impaired Insulin Secretion with Impaired Glucose and Lipid Oxidation. PLoS ONE 7, e40868.

Cheung, M., Chapman, S., Hunt, K., Makin, E., Hickey, A., Hind, J., Ellard, S., Buchanan, C., and Kapoor, R. (2015). Clinical Characterisation of a novel RFX6 mutation – a rare cause of neonatal diabetes syndrome. Poster presented at the 54th annual European Society for Paediatric Endocrinology (ESPE) conference, 01-03/10/2015, Barcelona, Spain.

Chick, W.L., Warren, S., Chute, R.N., Like, A.A., Lauris, V., and Kitchen, K.C. (1977). A transplantable insulinoma in the rat. Proceedings of the National Academy of Sciences 74, 628–632.

Choksi, S.P., Lauter, G., Swoboda, P., and Roy, S. (2014). Switching on cilia: transcriptional networks regulating ciliogenesis. Development 141, 1427–1441.

Churchill, A.J., Gutiérrez, G.D., Singer, R.A., Lorberbaum, D.S., Fischer, K.A., and Sussel, L. (2017). Genetic evidence that Nkx2.2 acts primarily downstream of Neurog3 in pancreatic endocrine lineage development. ELife 6.

Clausen, B.E., Waldburger, J.-M., Schwenk, F., Barras, E., Mach, B., Rajewsky, K., Förster, I., and Reith, W. (1998). Residual MHC Class II Expression on Mature Dendritic Cells and Activated B Cells in RFX5-Deficient Mice. Immunity 8, 143–155.

Concepcion, J.P., Reh, C.S., Daniels, M., Liu, X., Paz, V.P., Ye, H., Highland, H.M., Hanis, C.L., and Greeley, S.A.W. (2014). Neonatal diabetes, gallbladder agenesis, duodenal atresia, and intestinal malrotation caused by a novel homozygous mutation in RFX6: Neonatal diabetes syndrome from an RFX6 mutation. Pediatric Diabetes 15, 67–72.

Coustan, D.R. (2013). Gestational Diabetes Mellitus. Clinical Chemistry 59, 1310–1321.

Da Silva Xavier, G., Rutter, G.A., Diraison, F., Andreolas, C., and Leclerc, I. (2006). ChREBP binding to fatty acid synthase and L-type pyruvate kinase genes is stimulated by glucose in pancreatic β-cells. Journal of Lipid Research 47, 2482–2491.

Da Silva Xavier, G., Sun, G., Qian, Q., Rutter, G.A., and Leclerc, I. (2010). ChREBP regulates Pdx-1 and other glucose-sensitive genes in pancreatic β-cells. Biochemical and Biophysical Research Communications 402, 252–257.

Damante, G., Fabbro, D., Pellizzari, L., Civitareale, D., Guazzi, S., Polycarpou-Schwartz, M., Cauci, S., Quadrifoglio, F., Formisano, S., and Lauro, R.D. (1994). Sequence-specific DNA recognition by the thyroid transcription factor-1 homeodomain. Nucleic Acids Research 22, 3075–3083.

Davies, M.N., O'Callaghan, B.L., and Towle, H.C. (2008). Glucose Activates ChREBP by Increasing Its Rate of Nuclear Entry and Relieving Repression of Its Transcriptional Activity. Journal of Biological Chemistry 283, 24029–24038.

Davies, M.N., O'Callaghan, B.L., and Towle, H.C. (2010). Activation and repression of glucose-stimulated ChREBP requires the concerted action of multiple domains within the MondoA conserved region. American Journal of Physiology-Endocrinology and Metabolism 299, E665–E674.

Dentin, R., Pégorier, J.-P., Benhamed, F., Foufelle, F., Ferré, P., Fauveau, V., Magnuson, M.A., Girard, J., and Postic, C. (2004). Hepatic Glucokinase Is Required for the Synergistic Action of ChREBP and SREBP-1c on Glycolytic and Lipogenic Gene Expression. Journal of Biological Chemistry 279, 20314–20326.

Dentin, R., Girard, J., and Postic, C. (2005). Carbohydrate responsive element binding protein (ChREBP) and sterol regulatory element binding protein-1c (SREBP-1c): two key regulators of glucose metabolism and lipid synthesis in liver. Biochimie 87, 81–86.

Dentin, R., Benhamed, F., Hainault, I., Fauveau, V., Foufelle, F., Dyck, J.R.B., Girard, J., and Postic, C. (2006). Liver-Specific Inhibition of ChREBP Improves Hepatic Steatosis and Insulin Resistance in ob/ob Mice. Diabetes 55, 2159–2170.

Dorrell, C., Schug, J., Canaday, P.S., Russ, H.A., Tarlow, B.D., Grompe, M.T., Horton, T., Hebrok, M., Streeter, P.R., Kaestner, K.H., et al. (2016). Human islets contain four distinct subtypes of β cells. Nature Communications 7.

Dotzlaw, H., Alkhalaf, M., and Murphy, L.C. (1992). Characterization of Estrogen Receptor Variant mRNAs from Human Breast Cancers. Molecular Endrocrinology 6, 773–785.

DuBridge, R.B., Tang, P., Hsia, H.C., Leong, P.M., Miller, J.H., and Calos, M.P. (1987). Analysis of mutation in human cells by using an Epstein-Barr virus shuttle system. Molecular and Cellular Biology 7, 379–387.

Duncan, S.A. (1998). Regulation of a Transcription Factor Network Required for Differentiation and Metabolism. Science 281, 692–695.

Dyson, N. (1998). The regulation of E2F by pRB-family proteins. Genes & Development 12, 2245–2262.

Efrat, S., Suranav, M., and Fleischer, N (1991). Glucose Induces Insulin Gene Transcription a Murine Pancreatic β-Cell Line. The Journal of Biological Chemistry 266, 11141–11143.

Fajas, L., Landsberg, R.L., Huss-Garcia, Y., Sardet, C., Lees, J.A., and Auwerx, J. (2002). E2Fs Regulate Adipocyte Differentiation. Developmental Cell 3, 39–49.

Fajas, L., Annicotte, J.-S., Miard, S., Sarruf, D., Watanabe, M., and Auwerx, J. (2004). Impaired pancreatic growth, β cell mass, and β cell function in E2F1 –/– mice. Journal of Clinical Investigation 113, 1288–1295.

Farack, L., Golan, M., Egozi, A., Dezorella, N., Bahar Halpern, K., Ben-Moshe, S., Garzilli, I., Tóth, B., Roitman, L., Krizhanovsky, V., et al. (2019). Transcriptional Heterogeneity of Beta Cells in the Intact Pancreas. Developmental Cell 48, 115-125.e4.

Feng, C., Xu, W., and Zuo, Z. (2009). Knockout of the regulatory factor X1 gene leads to early embryonic lethality. Biochemical and Biophysical Research Communications 386, 715–717.

Fernandez-Zapico, M.E., van Velkinburgh, J.C., Gutiérrez-Aguilar, R., Neve, B., Froguel, P., Urrutia, R., and Stein, R. (2009). MODY7 Gene, KLF11 , Is a Novel p300-dependent Regulator of Pdx-1 (MODY4) Transcription in Pancreatic Islet β Cells. Journal of Biological Chemistry 284, 36482–36490.

Filhoulaud, G., Guilmeau, S., Dentin, R., Girard, J., and Postic, C. (2013). Novel insights into ChREBP regulation and function. Trends in Endocrinology & Metabolism 24, 257–268.

Fu, Z., Gilbert, E.R., and Liu, D. (2013). Regulation of Insulin Synthesis and Secretion and Pancreatic Beta-Cell Dysfunction in Diabetes. Current Diabetes Reviews 9, 25–53.

Gajiwala, K.S., and Burley, S.K. (2000). Winged helix proteins. Current Opinion in Structural Biology 10, 110–116.

Gajiwala, K.S., Chen, H., Cornille, F., Roques, B.P., Reith, W., Mach, B., and Burley, S.K. (2000). Structure of the winged-helix protein hRFX1 reveals a new mode of DNA binding. Nature 403, 916–921.

Galán-Gómez, E., Sánchez, E.B., Arias-Castro, S., and Cardesa-García, J.J. (2007). Intrauterine growth retardation, duodenal and extrahepatic biliary atresia, hypoplastic pancreas and other intestinal

anomalies: Further evidence of the Martínez-Frías syndrome. European Journal of Medical Genetics 50, 144–148.

Gao, N., Le Lay, J., Qin, W., Doliba, N., Schug, J., Fox, A.J., Smirnova, O., Matschinsky, F.M., and Kaestner, K.H. (2010). Foxa1 and Foxa2 Maintain the Metabolic and Secretory Features of the Mature β-Cell. Molecular Endocrinology 24, 1594–1604.

Gazdar, A.F., Chick, W.L., Oie, H.K., Sims, H.L., King, D.L., Weir, G.C., and Lauris, V. (1980). Continuous, clonal, insulin- and somatostatin-secreting cell lines established from a transplantable rat islet cell tumor. Proceedings of the National Academy of Sciences 77, 3519–3523.

Giddings, S., and Carnaghi, L.R. (1988). The Two Nonallelic Rat Insulin mRNAs and Pre-mRNAs Are Regulated Coordinatelyin Viuo. The Journal of Biological Chemistry 263, 3845–3849.

Grandori, C., Cowley, S.M., James, L.P., and Eisenman, R.N. (2000). The Myc/Max/Mad Network and the Transcriptional Control of Cell Behavior. Annual Review of Cell and Developmental Biology 16, 653–699.

Green, N., Alexander, H., Olson, A., Alexander, S., Shinnick, T.M., Sutcliffe, J.G., and Lerner, R.A. (1982). Immunogenic structure of the influenza virus hemagglutinin. Cell 28, 477–487.

Gu, C., Stein, G.H., Pan, N., Goebbels, S., Hörnberg, H., Nave, K.-A., Herrera, P., White, P., Kaestner, K.H., Sussel, L., et al. (2010). Pancreatic β Cells Require NeuroD to Achieve and Maintain Functional Maturity. Cell Metabolism 11, 298–310.

Gutiérrez, G.D., Bender, A.S., Cirulli, V., Mastracci, T.L., Kelly, S.M., Tsirigos, A., Kaestner, K.H., and Sussel, L. (2016). Pancreatic β cell identity requires continual repression of non–β cell programs. Journal of Clinical Investigation 127, 244–259.

Gutiérrez-Aguilar, R., Benmezroua, Y., Vaillant, E., Balkau, B., Marre, M., Charpentier, G., Sladek, R., Froguel, P., and Neve, B. (2007). Analysis of KLFtranscription factor family gene variants in type 2 diabetes. BMC Medical Genetics 8, 53.

Harbour, J.W. (2000). The Rb/E2F pathway: expanding roles and emerging paradigms. Genes & Development 14, 2393–2409.

Hashimoto, K., Ishida, E., Matsumoto, S., Okada, S., Yamada, M., Satoh, T., Monden, T., and Mori, M. (2009). Carbohydrate Response Element Binding Protein Gene Expression Is Positively Regulated by Thyroid Hormone. Endocrinology 150, 3417–3424.

Hauge-Evans, A.C., King, A.J., Carmignac, D., Richardson, C.C., Robinson, I.C.A.F., Low, M.J., Christie, M.R., Persaud, S.J., and Jones, P.M. (2009). Somatostatin Secreted by Islet β-Cells Fulfills Multiple Roles as a Paracrine Regulator of Islet Function. Diabetes 58, 403–411.

Havula, E., and Hietakangas, V. (2012a). Glucose sensing by ChREBP/MondoA–Mlx transcription factors. Seminars in Cell & Developmental Biology 23, 640–647.

Havula, E., and Hietakangas, V. (2012b). Glucose sensing by ChREBP/MondoA–Mlx transcription factors. Seminars in Cell & Developmental Biology 23, 640–647.

Havula, E., and Hietakangas, V. (2018). Sugar sensing by ChREBP/Mondo-Mlx — new insight into downstream regulatory networks and integration of nutrient-derived signals. Current Opinion in Cell Biology 51, 89–96.

Hedge, M., Lortz, S., Drinkgern, J., and Lenzen, S. (1997). Relation Between Antioxidant Enzyme Gene Expression and Antioxidative Defense Status of Insulin-Producing Cells. Skeletal Muscle 46, 10.

Heintzman, N.D., Stuart, R.K., Hon, G., Fu, Y., Ching, C.W., Hawkins, R.D., Barrera, L.O., Van Calcar, S., Qu, C., Ching, K.A., et al. (2007). Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nature Genetics 39, 311–318.

Herman, M.A., Peroni, O.D., Villoria, J., Schön, M.R., Abumrad, N.A., Blüher, M., Klein, S., and Kahn, B.B. (2012). A novel ChREBP isoform in adipose tissue regulates systemic glucose metabolism. Nature 484, 333–338.

Hills, C.E., and Brunskill, N.J. (2008). Intracellular Signalling by C-Peptide. Experimental Diabetes Research 2008, 1–8.

Hohmeier, H.E., Mulder, H., Chen, G., Henkel-Rieger, R., Prentki, M., and Newgard, C.B. (2000). Isolation of INS-1-derived cell lines with robust ATP-sensitive K+ channel-dependent and -independent glucose-stimulated insulin secretion. Diabetes 49, 424–430.

Huang, S., and Czech, M.P. (2007). The GLUT4 Glucose Transporter. Cell Metabolism 5, 237–252.

Huopio, H., Miettinen, P.J., Ilonen, J., Nykänen, P., Veijola, R., Keskinen, P., Näntö-Salonen, K., Vangipurapu, J., Raivo, J., Stančáková, A., et al. (2016). Clinical, Genetic, and Biochemical Characteristics of Early-Onset Diabetes in the Finnish Population. The Journal of Clinical Endocrinology & Metabolism 101, 3018–3026.

Iizuka, K. (2017). The transcription factor carbohydrate-response element-binding protein (ChREBP): A possible link between metabolic disease and cancer. Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease 1863, 474–485.

Iizuka, K. (2013). Recent progress on the role of ChREBP in glucose and lipid metabolism. Endocrine Journal 60, 543–555.

Iizuka, K., and Horikawa, Y. (2008). ChREBP: A Glucose-activated Transcription Factor Involved in the Development of Metabolic Syndrome. Endocrine Journal 55, 617–624.

Iizuka, K., Miller, B., and Uyeda, K. (2006). Deficiency of carbohydrate-activated transcription factor ChREBP prevents obesity and improves plasma glucose control in leptin-deficient (ob/ob) mice. American Journal of Physiology-Endocrinology and Metabolism 291, E358–E364.

Iizuka, K., Tomita, R., Takeda, J., and Horikawa, Y. (2012). Rat glucagon receptor mRNA is directly regulated by glucose through transactivation of the carbohydrate response element binding protein. Biochemical and Biophysical Research Communications 417, 1107–1112.

Iizuka, K., Bruick, R.K., Liang, G., Horton, J.D., and Uyeda, K. (2004). Deficiency of carbohydrate response element-binding protein (ChREBP) reduces lipogenesis as well as glycolysis. Proceedings of the National Academy of Sciences 101, 7281–7286.

Ishii, S., IIzuka, K., Miller, B.C., and Uyeda, K. (2004). Carbohydrate response element binding protein directly promotes lipogenic enzyme gene transcription. Proceedings of the National Academy of Sciences 101, 15597–15602.

Iwata, T.N., Cowley, T.J., Sloma, M., Ji, Y., Kim, H., Qi, L., and Lee, S.S. (2013). The Transcriptional Co-Regulator HCF-1 Is Required for INS-1 β-cell Glucose-Stimulated Insulin Secretion. PLoS ONE 8, e78841.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier, E. (2012). A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity. Science 337, 816–821.

Jing, G., Chen, J., Xu, G., and Shalev, A. (2016). Islet ChREBP-β is increased in diabetes and controls ChREBP-α and glucose-induced gene expression via a negative feedback loop. Molecular Metabolism 5, 1208–1215.

Johnston, N.R., Mitchell, R.K., Haythorne, E., Pessoa, M.P., Semplici, F., Ferrer, J., Piemonti, L., Marchetti, P., Bugliani, M., Bosco, D., et al. (2016). Beta Cell Hubs Dictate Pancreatic Islet Responses to Glucose. Cell Metabolism 24, 389–401.

Jois, T., and Sleeman, M.W. (2017). The regulation and role of carbohydrate response element-binding protein in metabolic homeostasis and disease. Journal of Neuroendocrinology 29, e12473.

Jois, T., Chen, W., Howard, V., Harvey, R., Youngs, K., Thalmann, C., Saha, P., Chan, L., Cowley, M.A., and Sleeman, M.W. (2017). Deletion of hepatic carbohydrate response element binding protein (ChREBP) impairs glucose homeostasis and hepatic insulin sensitivity in mice. Molecular Metabolism 6, 1381–1394.

Kaestner, K.H., Katz, J., Liu, Y., and Drucker, D.J. (1999). Inactivation of the winged helix transcription factor HNF3α affects glucose homeostasis and islet glucagon gene expression in vivo. Genes & Development 13, 495–504.

Karlseder, J., Rotheneder, H., and Wintersberger, E. (1996). Interaction of Sp1 with the growth- and cell cycle-regulated transcription factor E2F. Molecular and Cellular Biology 16, 1659–1667.

Khan, N., Dandan, W., Hassani, N.A., and Had, S. (2016). A Newly-Discovered Mutation in the RFX6 Gene of the Rare Mitchell-Riley Syndrome. Journal of Clinical Research in Pediatric Endocrinology 8, 246–249.

Kibbe, C., Chen, J., Xu, G., Jing, G., and Shalev, A. (2013). FOXO1 Competes with Carbohydrate Response Element-binding Protein (ChREBP) and Inhibits Thioredoxin-interacting Protein (TXNIP) Transcription in Pancreatic Beta Cells. Journal of Biological Chemistry 288, 23194–23202.

Kingsley, C., and Winoto, A. (1992). Cloning of GT Box-Binding Proteins: A Novel Spl Multigene Family Regulating T-Cell Receptor Gene Expression. Molecular and Cellular Biology 12, 4251–4261.

Klochendler, A., Caspi, I., Corem, N., Moran, M., Friedlich, O., Elgavish, S., Nevo, Y., Helman, A., Glaser, B., Eden, A., et al. (2016). The Genetic Program of Pancreatic β-Cell Replication In Vivo. Diabetes 65, 2081–2093.

Klok, M.D., Jakobsdottir, S., and Drent, M.L. (2007). The role of leptin and ghrelin in the regulation of food intake and body weight in humans: a review. Obesity Reviews 8, 21–34.

Koranyi, L., Permutt, M.A., Chirgwin, J.M., and Giddings, S.J. (1989). Proinsulin I and II Gene Expression in Inbred Mouse Strains. Molecular Endocrinology 3, 1895–1902.

Ku, G.M., Kim, H., Vaughn, I.W., Hangauer, M.J., Oh, C.M., German, M.S., and McManus, M.T. (2012). Research Resource: RNA-Seq Reveals Unique Features of the Pancreatic β-Cell Transcriptome. Mol Endocrinol 26, 1783–1792.

Kursawe, R., Caprio, S., Giannini, C., Narayan, D., Lin, A., Shaw, M., Pierpont, B., Cushman, S.W., and Shulman, G.I. (2013). Decreased Transcription of ChREBP-α/β Isoforms in Abdominal Subcutaneous Adipose Tissue of Obese Adolescents With Prediabetes or Early Type 2 Diabetes. Associations With Insulin Resistance and Hyperglycemia. Diabetes 62, 837–844.

Laurençon, A., Dubruille, R., Efimenko, E., Grenier, G., Bissett, R., Cortier, E., Rolland, V., Swoboda, P., and Durand, B. (2007). Identification of novel regulatory factor X (RFX) target genes by comparative genomics in Drosophila species. Genome Biology 8, R195.

Leclerc, I., Rutter, G.A., Meur, G., and Noordeen, N. (2012). Roles of Ca2+ ions in the control of ChREBP nuclear translocation. Journal of Endocrinology 213, 115–122.

Leibiger, B., Moede, T., Schwarz, T., Brown, G.R., Kohler, M., Leibiger, I.B., and Berggren, P.-O. (1998a). Short-term regulation of insulin gene transcription by glucose. Proceedings of the National Academy of Sciences 95, 9307–9312.

Leibiger, B., Wåhlander, K., Berggren, P.-O., and Leibiger, I.B. (2000). Glucose-stimulated Insulin Biosynthesis Depends on Insulin-stimulated Insulin Gene Transcription. Journal of Biological Chemistry 275, 30153–30156.

Leibiger, I.B., Leibiger, B., Moede, T., and Berggren, P.-O. (1998b). Exocytosis of Insulin Promotes Insulin Gene Transcription via the Insulin Receptor/PI-3 Kinase/p70 s6 Kinase and CaM Kinase Pathways. Molecular Cell 1, 933–938.

Lerner, A.G., Upton, J.-P., Praveen, P.V.K., Ghosh, R., Nakagawa, Y., Igbaria, A., Shen, S., Nguyen, V., Backes, B.J., Heiman, M., et al. (2012). IRE1α Induces Thioredoxin-Interacting Protein to Activate the NLRP3 Inflammasome and Promote Programmed Cell Death under Irremediable ER Stress. Cell Metabolism 16, 250–264.

Li, M.V., Chang, B., Imamura, M., Poungvarin, N., and Chan, L. (2006). Glucose-Dependent Transcriptional Regulation by an Evolutionarily Conserved Glucose-Sensing Module. Diabetes 55, 1179–1189.

Li, M.V., Chen, W., Poungvarin, N., Imamura, M., and Chan, L. (2008). Glucose-Mediated Transactivation of Carbohydrate Response Element-Binding Protein Requires Cooperative Actions from Mondo Conserved Regions and Essential Trans -Acting Factor 14-3-3. Molecular Endocrinology 22, 1658–1672.

Lilla, V., Webb, G., Rickenbach, K., Maturana, A., Steiner, D.F., Halban, P.A., and Irminger, J.-C. (2003). Differential Gene Expression in Well-Regulated and Dysregulated Pancreatic β-Cell (MIN6) Sublines. Endocrinology 144, 1368–1379.

Lin, S.Y., Black, A.R., Kostic, D., Pajovic, S., Hoover, C.N., and Azizkhan, J.C. (1996). Cell cycle-regulated association of E2F1 and Sp1 is related to their functional interaction. Molecular and Cellular Biology 16, 1668–1675.

Lis, M., and Walther, D. (2016). The orientation of transcription factor binding site motifs in gene promoter regions: does it matter? BMC Genomics 17.

Lizio, M., Ishizu, Y., Itoh, M., Lassmann, T., Hasegawa, A., Kubosaki, A., Severin, J., Kawaji, H., Nakamura, Y., The FANTOM Consortium, et al. (2015). Mapping Mammalian Cell-type-specific Transcriptional Regulatory Networks Using KD-CAGE and ChIP-seq Data in the TC-YIK Cell Line. Frontiers in Genetics 6.

Ma, L., Sham, Y.Y., Walters, K.J., and Towle, H.C. (2006). A critical role for the loop region of the basic helix-loop-helix/leucine zipper protein Mlx in DNA binding and glucose-regulated transcription. Nucleic Acids Research 35, 35–44.

Martínez-Frías, M.-L., Frías, J.L., Galán, E., Domingo, R., Paisán, L., and Blanco, M. (1992). Tracheoesophageal fistula, gastrointestinal abnormalities, hypospadias, and prenatal growth deficiency. American Journal of Medical Genetics 44, 352–355.

Martinovici, D., Ransy, V., Vanden Eijnden, S., Ridremont, C., Pardou, A., Cassart, M., Avni, F., Donner, C., Lingier, P., Mathieu, A., et al. (2010). Neonatal hemochromatosis and Martinez-Frias syndrome of intestinal atresia and diabetes mellitus in a consanguineous newborn. European Journal of Medical Genetics 53, 25–28.

Maston, G.A., Evans, S.K., and Green, M.R. (2006). Transcriptional Regulatory Elements in the Human Genome. Annual Review of Genomics and Human Genetics 7, 29–59.

Mastracci, T.L., and Sussel, L. (2012). The endocrine pancreas: insights into development, differentiation, and diabetes: Pancreas development and diabetes. Wiley Interdisciplinary Reviews: Developmental Biology 1, 609–628.

Meier, J.J., Butler, A.E., Saisho, Y., Monchamp, T., Galasso, R., Bhushan, A., Rizza, R.A., and Butler, P.C. (2008). β-Cell Replication Is the Primary Mechanism Subserving the Postnatal Expansion of β-Cell Mass in Humans. Diabetes 57, 1584–1594.

Mellitzer, G., Martín, M., Sidhoum-Jenny, M., Orvain, C., Barths, J., Seymour, P.A., Sander, M., and Gradwohl, G. (2004). Pancreatic Islet Progenitor Cells in Neurogenin 3-Yellow Fluorescent Protein Knock-Add-On Mice. Molecular Endocrinology 18, 2765–2776.

Melloul, D., Marshak, S., and Cerasi, E. (2002). Regulation of insulin gene transcription. Diabetologia 45, 309–326.

Merglen, A., Theander, S., Rubi, B., Chaffard, G., Wollheim, C.B., and Maechler, P. (2004). Glucose Sensitivity and Metabolism-Secretion Coupling Studied during Two-Year Continuous Culture in INS-1E Insulinoma Cells. Endocrinology 145, 667–678.

Merriman, C., Huang, Q., Gu, W., Yu, L., and Fu, D. (2018). A subclass of serum anti-ZnT8 antibodies directed to the surface of live pancreatic β-cells. Journal of Biological Chemistry 293, 579–587.

Metukuri, M.R., Zhang, P., Basantani, M.K., Chin, C., Stamateris, R.E., Alonso, L.C., Takane, K.K., Gramignoli, R., Strom, S.C., Stewart, A.F., et al. (2012). ChREBP Mediates Glucose-Stimulated Pancreatic β-Cell Proliferation. Diabetes 61, 2004–2015.

Michael, M.D., Kulkarni, R.N., Postic, C., Previs, S.F., Shulman, G.I., Magnuson, M.A., and Kahn, C.R. (2000). Loss of Insulin Signaling in Hepatocytes Leads to Severe Insulin Resistance and Progressive Hepatic Dysfunction. Molecular Cell 6, 87–97.

Minn, A.H., Hafele, C., and Shalev, A. (2005). Thioredoxin-Interacting Protein Is Stimulated by Glucose through a Carbohydrate Response Element and Induces β-Cell Apoptosis. Endocrinology 146, 2397–2405.

Mitchell, J., Punthakee, Z., Lo, B., Bernard, C., Chong, K., Newman, C., Cartier, L., Desilets, V., Cutz, E., Hansen, I.L., et al. (2004). Neonatal diabetes, with hypoplastic pancreas, intestinal atresia and gall bladder hypoplasia: search for the aetiology of a new autosomal recessive syndrome. Diabetologia 47, 2160–2167.

Miyatsuka, T., Li, Z., and German, M.S. (2009). Chronology of Islet Differentiation Revealed By Temporal Cell Labeling. Diabetes 58, 1863–1868.

Miyazaki, J.I., Araki, K., Yamato, E., Ikegami, H., Asano, T., Shibasaki, Y., Oka, Y., and Yamamura, K.I. (1990). Establishment of a pancreatic beta-cell line that retains glucose-inducible insulin secretion – special reference to expression of glucose transporter isoforms. Endocrinology 127, 126–132.

Mladenova, V., Mladenov, E., and Russev, G. (2009). Organization of Plasmid DNA into Nucleosome-Like Structures after Transfection in Eukaryotic Cells. Biotechnology & Biotechnological Equipment 23, 1044–1047.

Morotomi-Yano, K., Yano, K., Saito, H., Sun, Z., Iwama, A., and Miki, Y. (2002). Human Regulatory Factor X 4 (RFX4) Is a Testis-specific Dimeric DNA-binding Protein That Cooperates with Other Human RFX Members. Journal of Biological Chemistry 277, 836–842.

Nano, R., Bosco, D., Kerr-Conte, J.A., Karlsson, M., Charvier, S., Melzi, R., Ezzouaoui, R., Mercalli, A., Hwa, A., Pattou, F., et al. (2015). Human islet distribution programme for basic research: activity over the last 5 years. Diabetologia 58, 1138–1140.

Nasteska, D., and Hodson, D.J. (2018). The role of beta cell heterogeneity in islet function and insulin release. Journal of Molecular Endocrinology 61, R43–R60.

Naya, F.J., Huang, H.-P., Qiu, Y., Mutoh, H., DeMayo, F.J., Leiter, A.B., and Tsai, M.-J. (1997). Diabetes, defective pancreatic morphogenesis, and abnormal enteroendocrine differentiation in BETA2/NeuroD-deficient mice. Genes & Development 11, 2323–2334.

Naya, F.J., Stellrecht, C.M.M., and Tsai, M.-J. (1995). Tissue-specific regulation of the insulin gene by a novel basic helix-loop-helix transcription factor. Genes & Development 9, 1009–1019.

Naylor, J., Suckow, A.T., Seth, A., Baker, D.J., Sermadiras, I., Ravn, P., Howes, R., Li, J., Snaith, M.R., Coghlan, M.P., et al. (2016). Use of CRISPR/Cas9-engineered INS-1 pancreatic cells to define the pharmacology of dual GIPR/GLP-1R agonists. Biochemical Journal 473, 2881–2891.

Neve, B., Fernandez-Zapico, M.E., Ashkenazi-Katalan, V., Dina, C., Hamid, Y.H., Joly, E., Vaillant, E., Benmezroua, Y., Durand, E., Bakaher, N., et al. (2005). Role of transcription factor KLF11 and its diabetes-associated gene variants in pancreatic beta cell function. Proceedings of the National Academy of Sciences 102, 4807–4812.

Nica, A.C., Ongen, H., Irminger, J.-C., Bosco, D., Berney, T., Antonarakis, S.E., Halban, P.A., and Dermitzakis, E.T. (2013). Cell-type, allelic, and genetic signatures in the human pancreatic beta cell transcriptome. Genome Research 23, 1554–1562.

Noordeen, N.A., Khera, T.K., Sun, G., Longbottom, E.R., Pullen, T.J., da Silva Xavier, G., Rutter, G.A., and Leclerc, I. (2010). Carbohydrate-Responsive Element-Binding Protein (ChREBP) Is a Negative Regulator of ARNT/HIF-1 Gene Expression in Pancreatic Islet -Cells. Diabetes 59, 153–160.

Noordeen, N.A., Meur, G., Rutter, G.A., and Leclerc, I. (2012). Glucose-Induced Nuclear Shuttling of ChREBP Is Mediated by Sorcin and Ca2+ Ions in Pancreatic β-Cells. Diabetes 61, 575–585.

Oslowski, C.M., Hara, T., O'Sullivan-Murphy, B., Kanekura, K., Lu, S., Hara, M., Ishigaki, S., Zhu, L.J., Hayashi, E., Hui, S.T., et al. (2012). Thioredoxin-Interacting Protein Mediates ER Stress-Induced β Cell Death through Initiation of the Inflammasome. Cell Metabolism 16, 265–273.

Patel, K.A., Kettunen, J., Laakso, M., Stančáková, A., Laver, T.W., Colclough, K., Johnson, M.B., Abramowicz, M., Groop, L., Miettinen, P.J., et al. (2017). Heterozygous RFX6 protein truncating variants are associated with MODY with reduced penetrance. Nature Communications 8.

Pearl, E.J., Jarikji, Z., and Horb, M.E. (2011). Functional analysis of Rfx6 and mutant variants associated with neonatal diabetes. Developmental Biology 351, 135–145.

Peterson, B.S., Campbell, J.E., Ilkayeva, O., Grimsrud, P.A., Hirschey, M.D., and Newgard, C.B. (2018). Remodeling of the Acetylproteome by SIRT3 Manipulation Fails to Affect Insulin Secretion or β Cell Metabolism in the Absence of Overnutrition. Cell Reports 24, 209-223.e6.

Piccand, J., Strasser, P., Hodson, D.J., Meunier, A., Ye, T., Keime, C., Birling, M.-C., Rutter, G.A., and Gradwohl, G. (2014). Rfx6 Maintains the Functional Identity of Adult Pancreatic β Cells. Cell Reports 9, 2219–2232.

Poidvin, A., Chandra, V., Fauret-Amsellem, A.-L., Cavé, H., Beltrand, J., Tubiana-Rufi, N., Carel, J.-C., Polak, M., and Scharfmann, R. (2016). Clinical Management of the Mitchell-Riley Syndrome Due to RFX6 Gene Mutations: Aggressive Support Results in Improved Outcome. Poster presented at the 55th annual European Society for Paediatric Endocrinology (ESPE) conference, 10-12/09/2016, Paris, France.

Poitout, V., and Robertson, R.P. (2008). Glucolipotoxicity: Fuel Excess and β-Cell Dysfunction. Endocrine Reviews 29, 351–366.

Porat, S., Weinberg-Corem, N., Tornovsky-Babaey, S., Schyr-Ben-Haroush, R., Hija, A., Stolovich-Rain, M., Dadon, D., Granot, Z., Ben-Hur, V., White, P., et al. (2011). Control of Pancreatic β Cell Regeneration by Glucose Metabolism. Cell Metabolism 13, 440–449.

Postic, C., Dentin, R., Denechaud, P.-D., and Girard, J. (2007). ChREBP, a Transcriptional Regulator of Glucose and Lipid Metabolism. Annual Review of Nutrition 27, 179–192.

Poungvarin, N., Lee, J.K., Yechoor, V.K., Li, M.V., Assavapokee, T., Suksaranjit, P., Thepsongwajja, J.J., Saha, P.K., Oka, K., and Chan, L. (2012). Carbohydrate response element-binding protein (ChREBP) plays a pivotal role in beta cell glucotoxicity. Diabetologia 55, 1783–1796.

Poupeau, A., and Postic, C. (2011). Cross-regulation of hepatic glucose metabolism via ChREBP and nuclear receptors. Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease 1812, 995–1006.

Pullen, T.J., and Rutter, G.A. (2013). When less is more: the forbidden fruits of gene repression in the adult β-cell. Diabetes, Obesity and Metabolism 15, 503–512.

Pullen, T.J., Khan, A.M., Barton, G., Butcher, S.A., Sun, G., and Rutter, G.A. (2010). Identification of genes selectively disallowed in the pancreatic islet. Islets 2, 89–95.

Quesada, I., Tudurí, E., Ripoll, C., and Nadal, Á. (2008). Physiology of the pancreatic α-cell and glucagon secretion: role in glucose homeostasis and diabetes. Journal of Endocrinology 199, 5–19.

Quintens, R., Hendrickx, N., Lemaire, K., and Schuit, F. (2008). Why expression of some genes is disallowed in β-cells. Biochemical Society Transactions 36, 300–305.

Rabinovich, A., Jin, V.X., Rabinovich, R., Xu, X., and Farnham, P.J. (2008). E2F in vivo binding specificity: Comparison of consensus versus nonconsensus binding sites. Genome Research 18, 1763–1777.

Ran, F.A., Hsu, P.D., Lin, C.-Y., Gootenberg, J.S., Konermann, S., Trevino, A.E., Scott, D.A., Inoue, A., Matoba, S., Zhang, Y., et al. (2013a). Double Nicking by RNA-Guided CRISPR Cas9 for Enhanced Genome Editing Specificity. Cell 154, 1380–1389.

Ran, F.A., Hsu, P.D., Wright, J., Agarwala, V., Scott, D.A., and Zhang, F. (2013b). Genome engineering using the CRISPR-Cas9 system. Nature Protocols 8, 2281–2308.

Rani, S., Mehta, J.P., Barron, N., Doolan, P., Jeppesen, P.B., Clynes, M., and O'Driscoll, L. (2010). Decreasing Txnip mRNA and Protein Levels in Pancreatic MIN6 Cells Reduces Reactive Oxygen Species and Restores Glucose Regulated Insulin Secretion. Cellular Physiology and Biochemistry 25, 667–674.

Rask-Madsen, C., and King, G.L. (2013). Vascular Complications of Diabetes: Mechanisms of Injury and Protective Factors. Cell Metabolism 17, 20–33.

Ravassard, P., Hazhouz, Y., Pechberty, S., Bricout-Neveu, E., Armanet, M., Czernichow, P., and Scharfmann, R. (2011). A genetically engineered human pancreatic β cell line exhibiting glucose-inducible insulin secretion. The Journal of Clinical Investigation 121, 3589–3597.

Reith, W., Satola, S., Sanchez, C.H., Amaldi, I., Lisowska-Grospierre, B., Griscelli, C., Hadam, M.R., and Mach, B. (1988). Congenital immunodeficiency with a regulatory defect in MHC class II gene expression lacks a specific HLA-DR promoter binding protein, RF-X. Cell 53, 897–906.

Reith, W., Barras, E., Satola, S., Kobr, M., Reinhart, D., Sanchez, C.H., and Mach, B. (1989). Cloning of the major histocompatibility complex class II promoter binding protein affected in a hereditary defect in class II gene regulation. Proceedings of the National Academy of Sciences 86, 4200–4204.

Reith, W., Ucla, C., Barras, E., Gaud, A., Durand, B., Herrero-Sanchez, C., Kobr, M., and Mach, B. (1994). RFX1, a transactivator of hepatitis B virus enhancer I, belongs to a novel family of homodimeric and heterodimeric DNA-binding proteins. Molecular and Cellular Biology 14, 1230–1244.

Reith, W., Herrero-Sanchez, C., Kobr, M., Silacci, P., Berte, C., Fey, S., and Mach, B. (1990). MHC class II regulatory factor RFX has a novel DNA-binding domain and a functionally independent dimerization domain. Genes & Development 4, 1528–1540.

Richards, P., Ourabah, S., Montagne, J., Burnol, A.-F., Postic, C., and Guilmeau, S. (2017). MondoA/ChREBP: The usual suspects of transcriptional glucose sensing; Implication in pathophysiology. Metabolism 70, 133–151.

Richards, P., Rachdi, L., Oshima, M., Marchetti, P., Bugliani, M., Armanet, M., Postic, C., Guilmeau, S., and Scharfmann, R. (2018). MondoA Is an Essential Glucose-Responsive Transcription Factor in Human Pancreatic β-Cells. Diabetes 67, 461–472.

Rorsman, P., and Braun, M. (2013). Regulation of Insulin Secretion in Human Pancreatic Islets. Annual Review of Physiology 75, 155–179.

Rorsman, P., and Huising, M.O. (2018). The somatostatin-secreting pancreatic δ-cell in health and disease. Nature Reviews Endocrinology 14, 404–414.

Rotheneder, H., Geymayer, S., and Haidweg, E. (1999). Transcription Factors of the Sp1 Family: Interaction with E2F and Regulation of the Murine Thymidine Kinase Promoter. Journal of Molecular Biology 293, 1005–1015.

Rutter, G.A. (2004). Visualising insulin secretion. The Minkowski Lecture 2004. Diabetologia 47, 1861–1872.

Rutter, G.A., Pullen, T.J., Hodson, D.J., and Martinez-Sanchez, A. (2015). Pancreatic β-cell identity, glucose sensing and the control of insulin secretion. Biochemical Journal 466, 203–218.

Sadowski, D. (1971). Bacteriophage T7 Endonuclease I. Properties of the enzyme purified from T7 phage-infected Escherichia coli B. The Journal of Biological Chemistry 246, 209–216.

Sae-Lee, C., Moolsuwan, K., Chan, L., and Poungvarin, N. (2016). ChREBP Regulates Itself and Metabolic Genes Implicated in Lipid Accumulation in β–Cell Line. PLOS ONE 11, e0147411.

Sansbury, F.H., Kirel, B., Caswell, R., Lango Allen, H., Flanagan, S.E., Hattersley, A.T., Ellard, S., and Shaw-Smith, C.J. (2015). Biallelic RFX6 mutations can cause childhood as well as neonatal onset diabetes mellitus. European Journal of Human Genetics 23, 1744–1748.

Satoh, S., Masatoshi, S., Shou, Z., Yamamoto, T., Ishigure, T., Semii, A., Yamada, K., and Noguchi, T. (2007). Identification of cis-regulatory elements and trans-acting proteins of the rat carbohydrate response element binding protein gene. Archives of Biochemistry and Biophysics 461, 113–122.

Scharfmann, R., Pechberty, S., Hazhouz, Y., von Bülow, M., Bricout-Neveu, E., Grenier-Godard, M., Guez, F., Rachdi, L., Lohmann, M., Czernichow, P., et al. (2014). Development of a conditionally immortalized human pancreatic β cell line. Journal of Clinical Investigation 124, 2087–2098.

Schmidt, S.F., Madsen, J.G.S., Frafjord, K.Ø., Poulsen, L. la C., Salö, S., Boergesen, M., Loft, A., Larsen, B.D., Madsen, M.S., Holst, J.J., et al. (2016). Integrative Genomics Outlines a Biphasic Glucose Response and a ChREBP-RORγ Axis Regulating Proliferation in β Cells. Cell Reports 16, 2359–2372.

Schubert, D., Primavesi, L., Bishopp, A., Roberts, G., Doonan, J., Jenuwein, T., and Goodrich, J. (2006). Silencing by plant Polycomb-group genes requires dispersed trimethylation of histone H3 at lysine 27. The EMBO Journal 25, 4638–4649.

Schuit, F., Flamez, D., De Vos, A., and Pipeleers, D. (2002). Glucose-Regulated Gene Expression Maintaining the Glucose-Responsive State of β-Cells. Diabetes 51, S326–S332.

Schwitzgebel, V.M. (2014). Many faces of monogenic diabetes. Journal of Diabetes Investigation 5, 121–133.

Sekine, N., Cirulli, V., Regazzi, R., Brown, L.J., Gine, E., Tamarit-Rodriguez, J., Girotti, M., Marie, S., MacDonald, M.J., Wollheim, C.B., and Rutter, G.A. (1994). Low Lactate Dehydrogenase and High Mitochondrial Glycerol Phosphate Dehydrogenase in Pancreatic β-Cells. The Journal of Biological Chemistry 269, 4895–4902.

Shalev, A. (2014). Minireview: Thioredoxin-Interacting Protein: Regulation and Function in the Pancreatic β-Cell. Molecular Endocrinology 28, 1211–1220.

Shalev, A., Pise-Masison, C.A., Radonovich, M., Hoffmann, S.C., Hirshberg, B., Brady, J.N., and Harlan, D.M. (2002). Oligonucleotide Microarray Analysis of Intact Human Pancreatic Islets: Identification of Glucose-Responsive Genes and a Highly Regulated TGFβ Signaling Pathway. Endocrinology 143, 3695–3695.

Shih, H.-M., Liu, Z., and Towle, H.C. (1995). Two CACGTG Motifs with Proper Spacing Dictate the Carbohydrate Regulation of Hepatic Gene Transcription. Journal of Biological Chemistry 270, 21991–21997.

Shih, H.P., Wang, A., and Sander, M. (2013). Pancreas Organogenesis: From Lineage Determination to Morphogenesis. Annual Review of Cell and Developmental Biology 29, 81–105.

Sirek, A.S., Liu, L., Naples, M., Adeli, K., Ng, D.S., and Jin, T. (2009). Insulin Stimulates the Expression of Carbohydrate Response Element Binding Protein (ChREBP) by Attenuating the Repressive Effect of Pit-1, Oct-1/Oct-2, and Unc-86 Homeodomain Protein Octamer Transcription Factor-1. Endocrinology 150, 3483–3492.

Skelin, M. (2010). Pancreatic beta cell lines and their applications in diabetes mellitus research. ALTEX 105–113.

Skelin Klemen, M., Dolenšek, J., Slak Rupnik, M., and Stožer, A. (2017). The triggering pathway to insulin secretion: Functional similarities and differences between the human and the mouse β cells and their translational relevance. Islets 9, 109–139.

Skopkova, M., Ciljakova, M., Havlicekova, Z., Vojtkova, J., Valentinova, L., Danis, D., Murgas, D., Szepeova, R., Stanik, J., Banovcin, P., et al. (2016). Two novel RFX6 variants in siblings with Mitchell-Riley syndrome with later diabetes onset and heterotopic gastric mucosa. European Journal of Medical Genetics 59, 429–435.

Smith, S.B., Qu, H.-Q., Taleb, N., Kishimoto, N.Y., Scheel, D.W., Lu, Y., Patch, A.-M., Grabs, R., Wang, J., Lynn, F.C., et al. (2010). Rfx6 directs islet formation and insulin production in mice and humans. Nature 463, 775–780.

Soggia, A., Flosseau, K., Ravassard, P., Szinnai, G., Scharfmann, R., and Guillemain, G. (2012). Activation of the transcription factor carbohydrate-responsive element-binding protein by glucose leads to increased pancreatic beta cell differentiation in rats. Diabetologia 55, 2713–2722.

Soyer, J., Flasse, L., Raffelsberger, W., Beucher, A., Orvain, C., Peers, B., Ravassard, P., Vermot, J., Voz, M.L., Mellitzer, G., et al. (2010). Rfx6 is an Ngn3-dependent winged helix transcription factor required for pancreatic islet cell development. Development 137, 203–212.

Spiegel, R., Dobbie, A., Hartman, C., de Vries, L., Ellard, S., and Shalev, S.A. (2011). Clinical characterization of a newly described neonatal diabetes syndrome caused by RFX6 mutations. American Journal of Medical Genetics Part A 155, 2821–2825.

Steimle, V., Durand, B., Barras, E., Zufferey, M., Hadam, M.R., Mach, B., and Reith, W. (1995). A novel DNA-binding regulatory factor is mutated in primary MHC class II deficiency (bare lymphocyte syndrome). Genes & Development 9, 1021–1032.

Sugiaman-Trapman, D., Vitezic, M., Jouhilahti, E.-M., Mathelier, A., Lauter, G., Misra, S., Daub, C.O., Kere, J., and Swoboda, P. (2018). Characterization of the human RFX transcription factor family by regulatory and target gene analysis. BMC Genomics 19.

Sussel, L., Kalamaras, J., Hartigan-O'Connor, D.J., Meneses, J.J., Pedersen, R.A., Rubenstein, J.L.R., and German, M.S. (1998). Mice lacking the homeodomain transcription factor Nkx2.2 have diabetes due to arrested differentiation of pancreatic β cells. Development 125, 2213–2221.

Szabat, M., Luciani, D.S., Piret, J.M., and Johnson, J.D. (2009). Maturation of Adult β-Cells Revealed Using a Pdx1/Insulin Dual-Reporter Lentivirus. Endocrinology 150, 1627–1635.

Taniguchi, C.M., Emanuelli, B., and Kahn, C.R. (2006). Critical nodes in signalling pathways: insights into insulin action. Nature Reviews Molecular Cell Biology 7, 85–96.

Tao, Y., Kassatly, R.F., Cress, W.D., and Horowitz, J.M. (1997). Subunit composition determines E2F DNA-binding site specificity. Molecular and Cellular Biology 17, 6994–7007.

Tennant, B.R., Robertson, A.G., Kramer, M., Li, L., Zhang, X., Beach, M., Thiessen, N., Chiu, R., Mungall, K., Whiting, C.J., et al. (2013). Identification and analysis of murine pancreatic islet enhancers. Diabetologia 56, 542–552.

Teta, M., Long, S.Y., Wartschow, L.M., Rankin, M.M., and Kushner, J.A. (2005). Very Slow Turnover of β-Cells in Aged Adult Mice. Diabetes 54, 2557–2567.

Thomas, J., Morlé, L., Soulavie, F., Laurençon, A., Sagnol, S., and Durand, B. (2010). Transcriptional control of genes involved in ciliogenesis: a first step in making cilia. Biology of the Cell 102, 499–513.

Thorrez, L., Laudadio, I., Van Deun, K., Quintens, R., Hendrickx, N., Granvik, M., Lemaire, K., Schraenen, A., Van Lommel, L., Lehnert, S., et al. (2011). Tissue-specific disallowance of housekeeping genes: The other face of cell differentiation. Genome Research 21, 95–105.

Tritschler, S., Theis, F.J., Lickert, H., and Böttcher, A. (2017). Systematic single-cell analysis provides new insights into heterogeneity and plasticity of the pancreas. Molecular Metabolism 6, 974–990.

Van der Meulen, T., Donaldson, C.J., Cáceres, E., Hunter, A.E., Cowing-Zitron, C., Pound, L.D., Adams, M.W., Zembrzycki, A., Grove, K.L., and Huising, M.O. (2015). Urocortin3 mediates somatostatin-dependent negative feedback control of insulin secretion. Nature Medicine 21, 769–776.

Van der Meulen, T., Mawla, A.M., DiGruccio, M.R., Adams, M.W., Nies, V., Dólleman, S., Liu, S., Ackermann, A.M., Cáceres, E., Hunter, A.E., et al. (2017). Virgin Beta Cells Persist throughout Life at a Neogenic Niche within Pancreatic Islets. Cell Metabolism 25, 911-926.e6.

Varshney, A., Scott, L.J., Welch, R.P., Erdos, M.R., Chines, P.S., Narisu, N., Albanus, R.D., Orchard, P., Wolford, B.N., Kursawe, R., et al. (2017). Genetic regulatory signatures underlying islet gene expression and type 2 diabetes. Proceedings of the National Academy of Sciences 114, 2301–2306.

Verwest, A.M., Poelman, M., Dinjens, W.N.M., Batstra, M.R., Oostra, B.A., Lequin, M.H., Larsson, L.-I., Aanstoot, H.-J., Bruining, G.J., and de Krijger, R.R. (2000). Absence of a PDX-1 mutation and normal gastroduodenal immunohistology in a child with pancreatic agenesis. Virchows Archiv 437, 680–684.

Viña, J., Gomez-Cabrera, M.C., Borras, C., Froio, T., Sanchis-Gomar, F., Martinez-Bello, V.E., and Pallardo, F.V. (2009). Mitochondrial biogenesis in exercise and in ageing. Advanced Drug Delivery Reviews 61, 1369–1374.

Virbasius, J.V., and Scarpulla, R.C. (1994). Activation of the human mitochondrial transcription factor A gene by nuclear respiratory factors: a potential regulatory link between nuclear and mitochondrial gene expression in organelle biogenesis. Proceedings of the National Academy of Sciences 91, 1309–1313.

Wang, H., and Wollheim, C.B. (2002). ChREBP Rather than USF2 Regulates Glucose Stimulation of Endogenous L-pyruvate Kinase Expression in Insulin-secreting Cells. Journal of Biological Chemistry 277, 32746–32752.

Wang, B., Qi, T., Chen, S.-Q., Ye, L., Huang, Z.-S., and Li, H. (2016a). RFX1 maintains testis cord integrity by regulating the expression of Itga6 in male mouse embryos. Molecular Reproduction and Development 83, 606–614.

Wang, Y.J., Schug, J., Won, K.-J., Liu, C., Naji, A., Avrahami, D., Golson, M.L., and Kaestner, K.H. (2016b). Single-Cell Transcriptomics of the Human Endocrine Pancreas. Diabetes 65, 3028–3038.

Webb, G.C., Akbar, M.S., Zhao, C., and Steiner, D.F. (2000). Expression profiling of pancreatic β cells: Glucose regulation of secretory and metabolic pathway genes. Proceedings of the National Academy of Sciences 97, 5773–5778.

Welsh, N., Margulis, B., Borg, L.A.Ha., Wiklund, H.J., Mello, M.A., Andersson, A., Pipeleers, D.G., Hellerstrom, C., and Eizirik, D.L. (1995). Differences in the Expression of Heat- Shock Proteins and Antioxidant Enzymes between Human and Rodent Pancreatic Islets: Implications for the Pathogenesis of Insulin-Dependent Diabetes Mellitus. Molecular Medicine 1, 806–820.

Westermark, P., Andersson, A., and Westermark, G.T. (2011). Islet Amyloid Polypeptide, Islet Amyloid, and Diabetes Mellitus. Physiological Reviews 91, 795–826.

Wilson, I.A., Niman, H.L., Houghten, R.A., Cherenson, A.R., Connolly, M.L., and Lerner, R.A. (1984). The structure of an antigenic determinant in a protein. Cell 37, 767–778.

Wu, L., Chen, H., Zhu, Y., Meng, J., Li, Y., Li, M., Yang, D., Zhang, P., Feng, M., and Tong, X. (2013). Flightless I homolog negatively regulates ChREBP activity in cancer cells. The International Journal of Biochemistry & Cell Biology 45, 2688–2697.

Wu, Y., Hu, X., Li, Z., Wang, M., Li, S., Wang, X., Lin, X., Liao, S., Zhang, Z., Feng, X., et al. (2016). Transcription Factor RFX2 Is a Key Regulator of Mouse Spermiogenesis. Nature Scientific Reports 6, 20435.

Xu, P., Morrison, J.P., Foley, J.F., Stumpo, D.J., Ward, T., Zeldin, D.C., and Blackshear, P.J. (2018). Conditional ablation of the RFX4 isoform 1 transcription factor: Allele dosage effects on brain phenotype. PLOS ONE 13, e0190561.

Yamashita, H., Takenoshita, M., Sakurai, M., Bruick, R.K., Henzel, W.J., Shillinglaw, W., Arnot, D., and Uyeda, K. (2001). A glucose-responsive transcription factor that regulates carbohydrate metabolism in the liver. Proceedings of the National Academy of Sciences 98, 9116–9121.

Zaret, K.S., and Carroll, J.S. (2011). Pioneer transcription factors: establishing competence for gene expression. Genes & Development 25, 2227–2241.

Zegre Amorim, M., Houghton, J.A.L., Carmo, S., Salva, I., Pita, A., and Pereira-da-Silva, L. (2015). Mitchell-Riley Syndrome: A Novel Mutation in RFX6 Gene. Case Reports in Genetics 2015, 1–3.

Zhang, P., Kumar, A., Katz, L.S., Li, L., Paulynice, M., Herman, M.A., and Scott, D.K. (2015). Induction of the ChREBPb Isoform Is Essential for Glucose-Stimulated β-Cell Proliferation. 64, 4158–4170.

Zhao, C., Wilson, M.C., Schuit, F., Halestrap, A.P., and Rutter, G.A. (2001). Expression and Distribution of Lactate/Monocarboxylate Transporter Isoforms in Pancreatic Islets and the Exocrine Pancreas. Diabetes 50, 361–366.

Zhu, Z., Li, Q.V., Lee, K., Rosen, B.P., González, F., Soh, C.-L., and Huangfu, D. (2016). Genome Editing of Lineage Determinants in Human Pluripotent Stem Cells Reveals Mechanisms of Pancreatic Development and Diabetes. Cell Stem Cell 18, 755–768.

Université de Strasbourg

Julia Grans

École Doctorale des Sciences de la Vie et de la Santé STRASBOURG

Étude de la régulation transcriptionnelle de *Mlxipl* par RFX6 et identification des gènes cibles dans les cellules bêta pancréatiques

Résumé

La fonction **endocrine** du pancréas est essentielle pour l'homéostasie du glucose parce que les îlots pancréatiques contiennent le seul type des cellules endocrines, nommées **cellules bêta**, qui sont capable de produire et sécréter de l'**insuline**. Le **facteur de transcription** RFX6, maintenu dans toutes les cellules endocrines matures, est essentiel pour le développement, l'identité et la fonction des cellules bêta. Chez l'homme, des mutations de *RFX6* causent le **syndrome de Mitchell-Riley**, un trouble du développement caractérisé par un **diabète néonatal** et des malformations du système gastro-intestinal. La recherche des cibles de RFX6 dans les îlots murins a révélé que le facteur de transcription *Mlxipl* est directement régulé par RFX6. Dans cette thèse, nous avons étudié le mécanisme de la **régulation transcriptionnelle** de *Mlxipl* par RFX6 ainsi que les rôles de RFX6 et MLXIPL dans les cellules bêta adultes. Nous avons démontré que RFX6 se lie au premier intron de *Mlxipl* qui contient un **motif de liaison** (xbox) critique, et nous avons identifié les cofacteurs de ce processus. En comparant l'effet de la répression de *Rfx6* et *Mlxipl* dans des milieux riches ou faibles en glucose dans la **lignée cellulaire bêta** Ins-1 832/13 sur le **transcriptome**, nous avons déterminé les **programmes génétiques** contrôlés par RFX6 et MLXIPL.

mots-clés : cellules bêta pancréatiques, diabète néonatal, régulation transcriptionnelle, expression

Study of the transcriptional regulation of *Mlxipl* by RFX6 and identification of target genes in pancreatic beta cells

Summary

Pancreatic **endocrine** function is critical for glucose homeostasis because pancreatic islets contain the only cells of the body, the **beta cells**, capable of producing and secreting **insulin**. The **transcription factor** RFX6 is maintained in all mature islet cells and is as an essential regulator of beta cell development, identity and function. In humans, *RFX6* mutations cause **Mitchell-Riley syndrome**, a developmental disorder characterized by **neonatal diabetes** and malformations of the digestive tract. The search for RFX6 targets in murine islets revealed that the transcription factor *Mlxipl* is directly regulated by RFX6. In this thesis, we investigated the mechanism of *Mlxipl* **transcriptional regulation** by RFX6, and the respective roles of RFX6 and its downstream target MLXIPL in adult beta cells. We demonstrated that RFX6 binds to the first intron of *Mlxipl* that contains a critical RFX **binding motif** (xbox), and we identified **cofactors** of this process. By comparing the changes in the **transcriptomes** linked to the loss of RFX6 or MLXIPL in the pancreatic **beta cell line** Ins-1 832/13 and the glucose level, we determined the genetic programs controlled by RFX6 and MLXIPL.

key words: pancreatic beta cell, neonatal diabetes, transcriptional regulation, expression