universidade de aveiro
theoria poiesis praxis
dqua departamento de química

SDU
SYDDANSK UNIVERSITET

Catarina Mendes Correia
Nº Mec. 67742

Identificação de *enhancers* da transcrição regulados pela insulina em tecido hepático de ratinho

Identification of transcriptional enhancers regulated by insulin signalling in mouse liver tissue

Tese de Mestrado

Mestrado em Bioquímica
Ramo de Especialização em Bioquímica Clínica

Ano Letivo 2016/2017
Junho de 2017

universidade de aveiro
theoria poiesis praxis
dqua departamento de química

SDU
SYDDANSK UNIVERSITET

Catarina Mendes Correia
Nº Mec. 67742

Identificação de *enhancers* da transcrição regulados pela insulina em tecido hepático de ratinho

Identification of transcriptional enhancers regulated by insulin signalling in mouse liver tissue

Tese de Mestrado

Dissertação apresentada à Universidade de Aveiro para cumprimento dos requisitos necessários à obtenção do grau de Mestre em Bioquímica (Especialização em Bioquímica Clínica), realizada sob a orientação científica do Doutor Lars Grøntved, Professor Associado do Departamento de Bioquímica e Biologia Molecular (BMB) da Universidade do Sul da Dinamarca, Odense (SDU), e da Doutora Majken Siersbæk, investigadora post-Doc no BMB-SDU. Orientação interna do Doutor Francisco Amado, Professor Associado com Agregação do Departamento de Química (DQ) da Universidade de Aveiro (UA).

Apoio financeiro de: Novo Nordisk Foundation, the Danish Diabetes Academy, VILLUM FONDEN (part of The Velux Foundations) and SDU Horizon2020.

novo nordisk fonden
Danish Diabetes Academy
VILLUM FONDEN

the jury

president

Dr. Lars Grøntved, PhD

associate professor at the Department of Biochemistry and Molecular Biology (BMB) of the University of Southern Denmark

Dr. Cathy Mitchelmore, PhD

associate professor at the Department of Science and Environment of Roskilde University

Acknowledgements

First, a big thank you to Lars, my main supervisor, for taking a leap of faith in accepting me in his group and for his invaluable guidance and insight into my project work. Thank you also to Majken, Adrija and Stine for their availability in helping with lab procedures and discussion of results, and for their constant encouragement. Finally, a word of appreciation to everyone in the Mandrup Group and in the Functional Genomics lab for the great working environment and for showing me the best that Denmark has to offer. You were my second family during this year. Tak for tilliden og for hjælpen!

A sincere thank you also to Professor Francisco Amado, Associate Professor with Aggregation at the Department of Chemistry of the University of Aveiro and my internal supervisor, who provided me with help and support with administrative procedures and guidelines specific to my home university. Recognition is also due to the Erasmus+ programme, which allowed me to conduct my thesis work at SDU from September 2016 to June 2017.

Lastly, I would like to acknowledge the importance of my family's support, despite being so far away. These words are especially directed to my parents, for accepting and encouraging my decision and providing the financial and emotional support that made this exchange period possible, and to my sister, for always keeping my focus on the bright side and for putting her artistic skills at my service to help me build the scheme I idealized. Obrigada!

Keywords Insulin resistance, feeding response, type 2 diabetes, S961, enhancer, Mediator complex, transcriptional regulation, circadian rhythms

Abstract Type 2 diabetes mellitus (T2DM) affects about 8% of the adult population worldwide, accounting for 90% of diabetes cases. This disease is caused by impairment of insulin signalling, leading to hyperglycaemia and compensatory hyperinsulinemia that result in a decrease in life quality and expectancy of patients. Even though it has mostly been associated with dietary and lifestyle risk factors, genetic factors of T2DM have been proposed, with increasing evidence.

In order to study the transcriptional mechanisms involved in insulin signalling, I have analysed the livers of mice subjected to a fasting/refeeding program, after injection with insulin receptor antagonist S961 (to mimic insulin resistance) or PBS (control). To probe for feeding and insulin-regulated enhancers, Mediator (MED) complex recruitment was analysed.

Data obtained suggest that Mediator complex occupancy is increased in the vicinity of feeding-upregulated genes. However, no novel enhancers were identified in association with recruited Mediator and further analyses are necessary to improve data quality. In parallel, S961 was assessed regarding its suitability as an insulin receptor antagonist. Although systemic and transcriptional effects of insulin signalling were blocked by S961, phosphorylation of Akt and GSK, downstream targets of the signalling pathway, suggested a partial agonist behaviour that should be clarified in future studies.

Palavras-Chave Resistência à insulina, resposta à alimentação, diabetes de tipo 2, S961, *enhancer*, complexo Mediador, regulação transcripcional, ritmos circadianos

Resumo

A diabetes mellitus de tipo 2 (DMT2) afeta cerca de 8% da população adulta mundial, representando 90% de todos os casos de diabetes. Esta doença é causada por uma resposta diminuída à ação da insulina, levando a hiperglicemia e hiperinsulinemia compensatória, que resultam numa diminuição severa da qualidade e esperança de vida dos pacientes. Apesar de ter sido associada à dieta e estilo de vida como fatores de risco principais, foram já propostos fatores genéticos da doença, com crescente apoio científico.

Para estudar os mecanismos da transcrição envolvidos na resposta à insulina, analisei os fígados de ratinhos sujeitos a um programa de jejum/realimentação, após injeção do antagonista do recetor de insulina S961 (para simular resistência à insulina) ou PBS (usado como controlo). Para investigar e identificar *enhancers* regulados pela ingestão de alimentos e pela insulina, foi analisado o recrutamento do complexo Mediador (MED).

Os dados obtidos sugerem que o complexo Mediador tem maior ocupação na proximidade de genes sobre-regulados pela ingestão de alimentos. No entanto, não foi possível identificar novos *enhancers* associados ao recrutamento do Mediador, e mais análises serão necessárias para melhorar a qualidade dos dados. Paralelamente, a conformidade do S961 enquanto antagonista do recetor de insulina foi avaliada. Apesar de os efeitos sistémicos e transcricionais da insulina serem bloqueados por este composto, os estados de fosforilação do Akt e da GSK, alvos da via de sinalização da insulina, são indicativos de um comportamento parcialmente agonista que deverá ser elucidado em estudos futuros.

Nøgleord Insulin resistens, fodringsrespons, type 2 diabetes, S961, *enhancer*, mediator komplekset, transkriptionel regulering, døgnrytme

Resumé Type 2 diabetes mellitus (T2DM) rammer omkring 8% af alle voksne verden over og udgør endvidere 90% af alle diabetes tilfælde. Sygdommen er forårsaget af forringelse af insulin signaleringen, hvilket fører til hyperglycæmia og kompensatorisk hyperinsulinemia, resulterende i forværring af livskvaliteten samt formindskelse af den forventede levealder hos patienterne. Selvom T2DM mest er associeret med diæt og livstil, er der i stigende grad beviser for at genetiske faktorer også er involveret.

For at undersøge de transkriptionelle mekanismer involveret i insulin signalering har jeg analyseret levere udtaget fra mus, der har været udsat for et faste/fodringsprogram og inden da er blevet injiceret med en insulin receptor antagonist S961 (for at simulere insulin resistens) eller PBS (kontrol). For desuden at identificere fodrings- og insulin-regulerede enhancere blev rekruttering af mediator (MED) komplekset analyseret.

Det indsamlede data antyder, at forekomsten af mediator komplekset er øget i nærheden af fodrings-opreguleret gener. Der er dog ikke identificeret nogle nye enhancere, der er associeret med rekrutteret mediator, og yderligere analyser vil være nødvendige for at forbedre kvaliteten af det indsamlede data. Desuden blev S961 undersøgt for sin evne som en insulin receptor antagonist. Selvom systematiske og transkriptionelle effekter af insulin signaleringen var blokeret af S961, så antyder fosforyleringen af Akt og GSK, der indgår i insulin signaleringskaskaden, at S961 opfører sig som en partiel agonist. Dette bør blive belyst yderligere i fremtidige studier.

Abbreviations

Ab	*antibody*
mAb/pAb	*monoclonal/polyclonal antibody*
Akt/PKB	*protein kinase B*
pAkt	*phosphorylated Akt*
APS	*ammonia persulphate*
bp/kb	*base pairs / kilobases (of DNA/RNA nucleotides)*
BLAST	*Basic Local Alignment Search Tool*
BSA	*bovine serum albumin*
Cp	*qPCR threshold cycle*
CTD	*C-terminal domain (RNA polymerase)*
ddH_2O/milliQ	*double-distilled water*
DEPC	*diethyl pyrocarbonate*
DMSO	*dimethyl sulfoxide*
(c)DNA	*(complementary) deoxyribonucleic acid*
DSG	*disuccinimidyl glutarate*
DTT	*dithiothreitol*
EDTA	*ethylenediaminetetraacetic acid*
ELISA	*enzyme-linked immunosorbent assay*
FDR	*false discovery rate (statistics)*
FoxO	*Forkhead box (O subclass)*
ΔG	*Gibbs free energy*
G6Pc	*glucose-6-phosphatase, catalytic subunit*
Gck	*glucokinase*
GLUT	*glucose transporter*
(p)GSK	*(phosphorylated) glycogen synthase kinase*
H3K27ac	*histone H3 lysine-27 acetylation*
HOMER	*Hypergeometric Optimization of Motif EnRichment*
HRP	*horseradish peroxidase*
IGF-IR	*insulin-like growth factor 1 receptor*
(Ch)IP	*(chromatin) immunoprecipitation*
(p)IR	*(phosphorylated) insulin receptor*
(L)IRKO	*(liver) insulin receptor knockout*
IRS	*insulin receptor substrate*

MED1	*mediator 1, mediator of RNA polymerase II transcription subunit 1*
mTOR	*mechanistic target of rapamycin*
NE	*nuclei extract(ion)*
NG	*"No Gene" (negative control primer)*
(d)NTP	*(DNA) nucleoside triphosphate*
ob/ob	*genotype ob/ob; obese mouse*
PAGE	*polyacrylamide gel electrophoresis*
PBS	*phosphate buffered saline*
PCI	*Phenol Chloroform Isoamylalcohol*
Pck1	*phosphoenolpyruvate carboxykinase 1*
PI	*protease inhibitor*
PI3K/PIK	*phosphatidylinositol 4,5-bisphosphate 3-kinase family*
PMSF	*phenylmethylsulfonyl fluoride*
Pol II	*RNA polymerase II*
PVDF	*polyvinylidene fluoride*
(m)RNA	*(messenger) ribonucleic acid*
RT	*room temperature*
(RT-)qPCR	*(reverse-transcription) quantitative polymerase chain reaction*
SDS	*sodium dodecyl sulphate*
-seq	*-sequencing*
Ser	*Serine*
SREBP/Srebf	*sterol regulatory element-binding protein*
STAR	*Spliced Transcripts Alignment to a Reference (software)*
T2DM	*type 2 diabetes mellitus*
TBS(-T)	*Tris-Buffered Saline (and Tween 20)*
TEMED	*tetramethylethylenediamine*
TGS	*Tris-Glycine-SDS*
TF	*transcription factor(s), e.g. TFIIB – transcription factor II B*
TRAP	*thyroid hormone receptor-associated polypeptides*
TSS	*transcription start site*
UM/LM	*upper marker / lower marker*
(e/i)WAT	*(epididymal/inguinal) white adipose tissue*
WC	*whole-cell (extraction)*
WHO	*World Health Organisation*

Table of Contents

1. Introduction

Type 2 Diabetes Mellitus (T2DM) is a chronic and progressive metabolic disease that, according to the World Health Organisation (WHO), affects about 8% of adults worldwide, with prevalence increasing at an alarming rate. T2DM comprises about 90% of all diabetes cases and is characterized by hyperglycaemia and compensatory hyperinsulinemia due to impairment of insulin signalling (insulin resistance)[1]. Due to the high prevalence of this disease, expected to almost double by 2030, and to its impact in life quality and expectancy, elucidating its underlying mechanisms and trying to find a cure for T2DM has become a high priority in research and healthcare[1, 2]. Although type 2 diabetes has been mostly associated with dietary and lifestyle risk factors, increasing evidence points to a genetic component dictating the development of the disease. Therefore, the identification of genetic mechanisms regulated by feeding and insulin signalling may be a crucial step in the path to make T2DM a more manageable, if not curable, disease, as genes showing impaired transcription in the insulin resistant state could be targets for treatment.

1.1. Transcriptional machinery and epigenetic regulation in eukaryotes

The transcription of all protein-coding genes is controlled by a common multiprotein complex that makes up the eukaryotic transcription machinery, RNA polymerase II (Pol II). DNA transcription is naturally repressed as a result of its organisation in nucleosomes, where the DNA helix is wrapped around histone octamers, but also of active protein complexes that maintain the repressive structure. This organisation forms a complex called chromatin[3]. Epigenetics is a term used to describe the regulation of gene expression through chromatin modifications not involving the sequence of DNA[4]. For a gene to become active, it has to be signalled by positive regulatory mechanisms, such as histone modifications (acetylation, phosphorylation, methylation), that open the chromatin structure and make the enhancer and promoter regions more accessible to the transcription enzymes[3]. Although phosphorylation and methylation have varying effects on the structure of chromatin, histone acetylation of enhancers (namely at H3K27) is usually associated with open chromatin and activated transcription[5]. While this is true for some modifications, in some cases histone-modifying enzymes are recruited through binding to components of the transcription initiation complex and therefore histone modification occurs after assembly of the transcriptional machinery[6]. Enhancer sites are dynamic regions of chromatin that drive transcription of target genes

through recruitment of activator proteins and transcription factors; they may be close or distant to the gene promoter[5]. Enhancers also differ in strength, which affects the frequency of transcriptional bursts at their target gene[7]. Initially, it was considered that the alteration in chromatin structure leading to the activated promoter was brought about through sliding of the nucleosomes, leaving the transcription start site (TSS) unoccupied, but with the modified histone octamers still in the chromatin structure in a different region[3]. However, more recent data have showed that removal and reassembly of the modified histones is a more accurate mechanism. This opening of the chromatin structure allows for the binding of RNA polymerase II and the remaining components of the transcription initiation complex: the general transcription factors (TFIIB, -D, -E, -F and -H) and the Mediator complex (described in section 1.3.)[3, 8]. Enhancer and promoter chromatin is enriched in histone variants that make the nucleosomes less stable and more prone to displacement from the DNA and to recruitment of DNA-binding proteins[5]. RNA Pol II is the enzyme responsible for promoting the high-fidelity synthesis of an mRNA molecule complementary to the DNA template strand on the active gene and proofreading the newly-synthesised mRNA transcript[3, 8]. The largest subunit of RNA polymerase II has a tail region consisting of multiple tandem repeats of a conserved heptapeptide rich in proline and serine residues, named carboxy-terminal domain (CTD)[9]. This 'tail' can be modified by multiple phosphorylations and is likely to have an essential role in regulating the transcriptional machinery and the assembly of the transcription initiation complex by interacting with the transcription factors[10]. Phosphorylation of the CTD may be necessary for the polymerase to go from transcription initiation to elongation and the Mediator complex has been reported to bind to the RNA Pol II through the CTD and stimulate its phosphorylation by the TFIIH kinase[11, 12]. Specifically, the polymerase is reported to be recruited to gene promoters where the preinitiation complex is assembling with the CTD hypophosphorylated; the CTD is then increasingly phosphorylated at the Ser5 residue during initiation and RNA Pol II proceeds to the elongation step within the gene, during which the CTD becomes additionally phosphorylated at Ser2[6, 13].

The transcription factors recognise the promoter region and facilitate its unwinding: a subunit of TFIID helps the TATA box sequence of the promoter associate with TFIIB, which positions the DNA TSS near the active centre of the polymerase, stabilizes the newly-formed RNA and binds TFIIE; TFIIE then recruits TFIIH, that has helicase activity and unwinds the

target DNA; finally, TFIIF binds the non-template strand to avoid renaturing of the DNA helix while the template strand is transcribed at the active site, resulting in initiation of transcription. After synthesis of a sequence of 10 or more residues, the RNA molecule becomes stable enough for the transcription to progress; TFIIB is displaced and the polymerase can advance further downstream (promoter escape)[3].

1.2. The Mediator complex and MED1 activity

The Mediator complex is a signal processor protein that can directly interact with RNA polymerase II and thus assists activator proteins bound at upstream enhancers in stimulating transcription[14]. It is a structure unique to the eukaryotic organisms that is required for transcription initiation at almost all eukaryotic Pol II promoters[3, 8, 15]. Despite its transversal function across eukaryote organisms, the primary structures (aminoacid sequence) of the Mediator subunits show extensive divergence; however, the 3D structure of the complex seems to be highly conserved. The subunits of this complex are organized into four modules: head and middle (bind to the polymerase), tail (interacts with the activator proteins) and Cdk8 (involved in the repression or activation of transcription due to its protein kinase activity that phosphorylates either the CTD or enhancer-bound activators, respectively)[14, 16, 17]. Figure 1 shows a proposed structure for the Mediator complex, taking into consideration the established interactions between subunits.

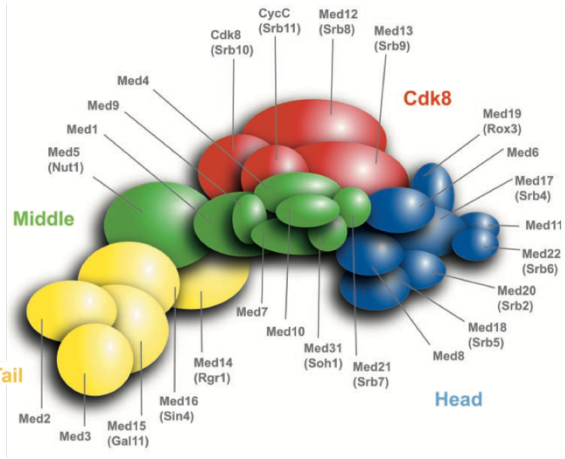


Figure 1. *Structure of the Mediator complex, inferred from protein-protein interactions detected between its subunits (adapted from [16]).*

Mediator acts as a bridge that interacts directly with an activator or repressor at transcriptional enhancer sites and with the transcription initiation complex at the gene promoter to elicit positive or negative regulation of gene expression[14]; it can act as a co-

activator or co-repressor depending on the proteins that it binds to[8, 15]. Hence, activator proteins promote transcription through indirect interaction with the TFs and Pol II requiring the presence of Mediator; direct binding to transcription factors happens only in a few cases at a low extent. The effect of the Mediator complex on transcription regulation likely results from binding to Pol II through the CTD[12, 18]. Together, the Mediator complex and Pol II in association with the transcription factors make up the RNA polymerase holoenzyme[12, 15, 19]. Mediator is recruited to enhancers first and then promoters before binding of Pol II; also, unfolding of the complex to a crescent-like shape upon contact with activators seems to be required for interaction with the polymerase[14, 15]. RNA Pol II only binds to a Cdk8-free complex; Cdk8 is recruited later[18]. The promotion of transcription by Mediator results from TFIIH-dependent phosphorylation of the CTD of Pol II elicited by the binding of Mediator to an activator protein, followed by loss of the Cdk8 module and promoter escape of Pol II[14, 18, 20]; binding to a repressor may lead to alterations of the CTD phosphorylation status producing the opposite effect on transcription[11, 12]. Mediator may also have additional roles in transcriptional regulation. In fact, it has been shown that Mediator interacts with a single nucleosome, and its protein Nut1 is likely involved in the acetylation of histone H3[21]. Mediator was shown to be a crucial piece of the promoter-dependent transcriptional mechanism as no transcription occurs in its absence, even when the activator is present[12, 15]. Figure 2 illustrates the mechanism of action of the Mediator complex through indirect binding to the target DNA, by interacting with both activator and Pol II.

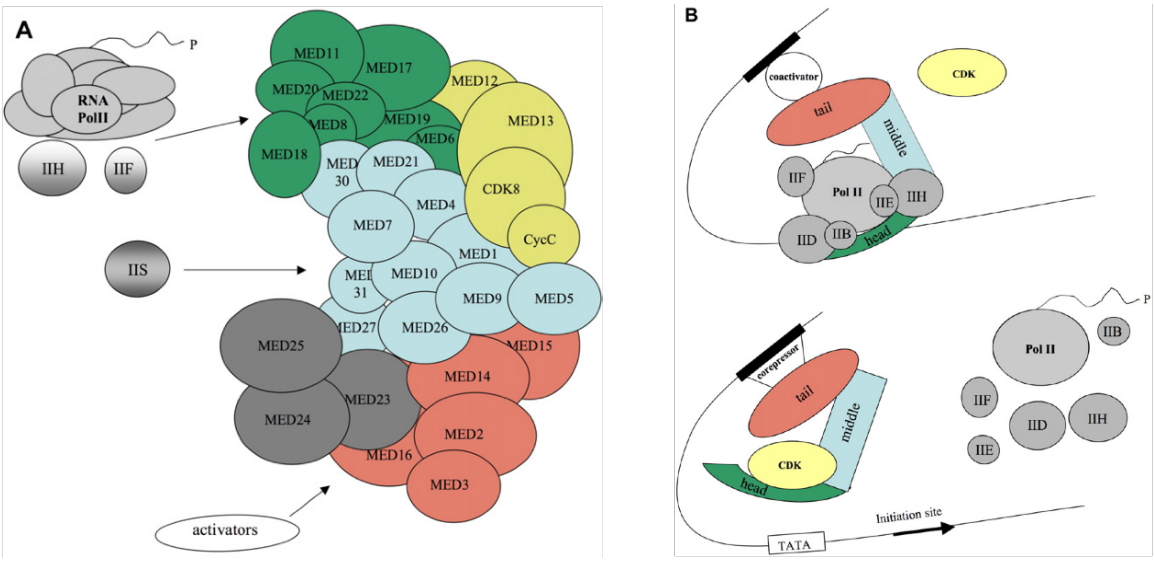


Figure 2. *Interactions established between the Mediator complex and the remaining components of the transcriptional machinery, as well as the activator/repressor proteins (**A**), and mechanism of action of the complex by bridging the enhancer and promoter sites (**B**) (from[18]).*

Mediator of RNA polymerase II transcription subunit 1, or MED1, also known as TRAP220, is part of the Mediator complex. MED1 seems to promote both activation and repression of gene expression and may also be involved in the interaction between the Med2 subunit and the remaining subunits of the Pol II holoenzyme complex[19]. MED1 regulates transcriptional activation through binding to different nuclear receptors and other activators[18, 22]. As shown in figures 1 and 2, MED1 is located in the middle module of the Mediator complex and interacts with the subunit MED14 in the tail region, thus connecting these two modules[16]. Phosphorylation of the MED1 subunit by a member of the MAPK family leads to its stabilisation, association with the Mediator complex (through the MED7 subunit in the middle module) and activation to promote transcription[23]. Genome-wide ChIP-seq based experiments have shown that MED1 occupies active enhancers and particularly high enrichment of MED1 binding to chromatin has been associated with super-enhancers[24].

1.3. Circadian regulation of metabolism

Most organisms have mechanisms to adapt to and predict periodic environmental changes, such as the 24h light/dark cycles, seasons or food availability, adjusting their activities and metabolic responses as a result of the time of the day[25]. One of those mechanisms is the 24-hour internal circadian clock, which regulates sleep, feeding and metabolism, hormone secretion, among other processes. This allows for synchronism of cellular events at an organism level in anticipation of environmental cycles to maintain homeostasis. Disruption of the normal circadian and metabolic rhythms, for example due to shift work, sleep deprivation or night eating can have a severe health impact and result in metabolic diseases[26, 27]. In fact, a rat rest-work model showed that night-shift work leads to a shift in feeding times and patterns that results in short-term altered metabolic gene expression in the liver, with some effects still prevailing after eight days of recovery[28]. As mentioned above, the epigenome (such as histone modifications and chromatin remodelling) alters and regulates gene expression, conferring it with plasticity to respond accordingly to the environmental conditions. Liver metabolism is highly influenced by these factors, as insulin is responsive to them (including nutrition and stress), and key metabolic enzymes are subject to epigenetic transcriptional regulation controlled by external signals[26]. There is not only regulation of metabolism by the epigenome, but the metabolic context of cells may also have an effect on epigenetic chromatin modifications[4].

The internal clock machinery is composed of a central clock located in the hypothalamus that entrains several peripheral clocks located in most organs, including the liver, through neuronal and hormonal signals[25, 26]. The main central clock regulator is light, and peripheral clocks are quite susceptible to fasting/feeding cycles; as a consequence, metabolic tissues are especially sensitive to regulation by the peripheral clocks[27]. The protein components of the clock machinery are transcription regulators that modulate the oscillatory and tissue-specific expression of their target genes and transcription factors, but may also affect their activity or stability or mediate a response by coactivators[26]. Clock proteins also recruit RNA Pol II and chromatin-modifying enzymes in a circadian pattern, modulating rhythmic histone acetylation and methylation, as well as promote chromatin remodelling, at both promoter and enhancer sites[6, 13, 25]. However, circadian regulation of cellular events is also a result of extensive post-transcriptional control, such as processing and degradation of mRNA[6, 13]. Genes involved in rhythmic hepatic function seem to be mostly regulated by the circadian regulation of transcriptional mechanisms described above, including Pol II recruitment and chromatin remodelling[6].

Metabolism is highly regulated by circadian clocks, and glucose homeostasis is particularly affected by the clock machinery; for example, the gluconeogenic key enzymes phosphoenolpyruvate kinase (Pck) and glucose 6-phosphatase (G6P) are direct targets of the clock machinery[26]. The circadian regulation of metabolism in the liver has been shown to depend on both fasting/feeding cycles (which depend on sleep/wake cycles, ultimately regulated by the central clock) and the action of peripheral clocks (also synchronised by the central clock mechanism)[26, 29]. Conversely, the availability of food and caloric intake also affects the fasting/feeding cycles, which can then alter the metabolic behaviour of the organism and regulate the peripheral clocks, specifically in the liver, and even uncouple them from the central mechanism[25-27, 30]. Feeding seems to be the main determinant of rhythmic gene-transcription in the liver, with the clock machinery enhancing the amplitude of oscillation and fine-tuning its timing, in anticipation of feeding patterns, which improves the efficiency of metabolic regulation[28, 29, 31].

1.4. Feeding response through the insulin signalling pathway

As mentioned above, feeding is one of the main regulators of metabolism, through activation or repression of expression of central metabolic genes, due to the increase in available

nutrients after feeding. The liver has a central function in the regulation of metabolism, and in the adaptation to the transitions between feeding and fasting. Fasting stimulates glycogenolysis, gluconeogenesis, β-oxidation and ketogenesis, while refeeding quickly restores glycogen content, with glycolysis only taking place later on; hepatic glycogen is thus fundamental in maintaining glucose homeostasis[32]. Insulin is a polypeptide hormone secreted by the pancreatic β-cells in response to feeding (increased blood glucose), and acting mainly in the liver, muscle and fat tissues[33]. Although it is primarily regulated by feeding patterns, insulin secretion also exhibits a weak circadian rhythm and is stimulated by proteins of the intrinsic clock machinery[26, 34]. Besides maintaining glucose homeostasis, insulin is also involved in the regulation of lipid and protein metabolism as well as cell growth and differentiation[2].

Insulin exerts its cellular effects upon binding to the insulin receptor (IR), a glycosylated transmembrane tyrosine kinase receptor composed of two αβ subunits ($\alpha_2\beta_2$ tetramer). The α-chains from each subunit are linked to the β-chain on the same subunit and to each other by disulphide bonds[35-37]. Insulin binds at the extracellular α-chains and elicits activation of the enzymatic portion of the β-chains[38]. It has been proposed that there are two interacting binding sites that exhibit negative cooperativity[36]. Negative cooperativity is a concept that refers to the property of a receptor in which the binding affinity at a particular binding site is lowered when the remaining site(s) is (are) occupied by the ligand, as a result of cooperative site-site interactions that increase the dissociation rate constants of other sites. In fact, in the IR, sites occupied by insulin lower the affinity for insulin of the free sites[37, 39]. The accepted model of binding of insulin to the receptor is shown below, in figure 3.

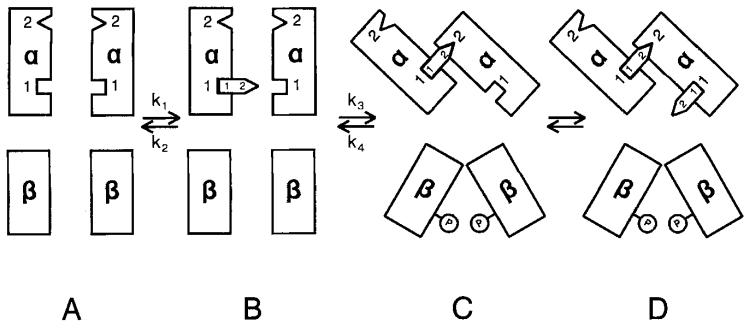


Figure 3. *Model for the binding of insulin to the insulin receptor. The receptor is initially empty, in its inactive conformation (**A**); insulin binds first to binding site 1 on one of the monomers (**B**); then, upon binding to the second binding site, insulin promotes crosslinking between the IR monomers (high-affinity binding), thus activating the IR complex and promoting its auto-phosphorylation (**C**); finally, a second insulin molecule might bind to the second pair of binding sites, but cannot promote its crosslinking (low-affinity binding) and therefore the receptor remains in the active conformation (**D**), (taken from [36]).*

Insulin has two regions involved in binding to the receptor, located on opposite sides of the molecule, named binding sites 1 and 2[36]. Some of the residues on insulin's binding sites have also been shown to be involved in negative cooperativity, making up a "cooperative" site within it[37]. Similarly, each α-chain of the IR has two different binding sites, also termed binding sites 1 and 2. It then becomes apparent that activation of the IR by insulin requires interaction between the two subunits of the receptor, through insulin-mediated bridging of opposite binding sites from each α-chain, resulting in a high-affinity complex that leads to a conformational change of the IR[36]; the conformation of insulin also changes upon binding to the receptor[37]. The free binding site 1 in one of the α-chains could then bind to a second insulin molecule, but with significantly lower affinity and not leading to interaction with the free binding site 2 in the opposite chain[36].

Upon binding to its receptor, insulin elicits a change in its conformation and, subsequently, promotes auto-transphosphorylation of the tyrosine kinase portions of the IR on the internal part of the β-monomers, on multiple tyrosine residues, resulting on their activation[36, 38], which leads to the propagation of the signalling pathway schematized in figure 4. Since the insulin-like growth factor 1 receptor (IGF-IR) is closely related to the IR, it also undergoes autophosphorylation following the binding of its ligand (IGF) and leads to the activation of the same cascade[40]. IR binds insulin with higher affinity than the IGF-1 receptor but, at high concentrations, insulin can also activate IGF-IR[36, 37].

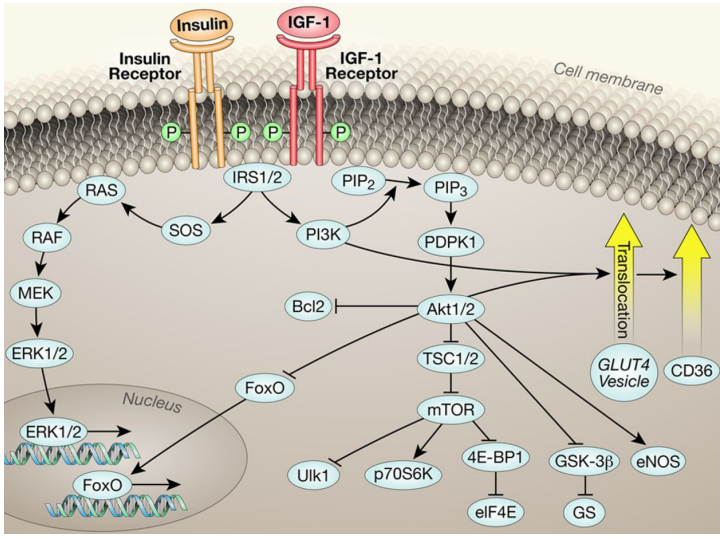


Figure 4. *Simplified schematic representation of the insulin signalling pathway, showing the central second messengers involved in the signal transduction after insulin binding to*

the IR. Activation of the IR, a tyrosine kinase receptor, results in its auto-phosphorylation, which in turn elicits a phosphorylation cascade that ultimately leads to phosphorylation of Akt. Akt is a central signalling molecule that becomes active upon phosphorylation and proceeds to activate or inactivate downstream effector molecules (such as FOXO, mTOR, GSK) through phosphorylation, resulting in the cellular effects promoted by feeding and insulin signalling. There is also an alternative transduction pathway involving the MAPK cascade, which regulates various genes' expression (adapted from [40]).

The active tyrosine kinase phosphorylates tyrosine residues of the insulin receptor substrate (IRS) proteins, which subsequently leads to the activation of two major phosphorylation cascades involved in different signalling events: the mitogen-activated protein kinases (MAPK) cascade, which ultimately promotes the expression of genes involved in the regulation of cell differentiation and proliferation, and the central Akt pathway[40]. Akt is phosphorylated by the kinase PDK1, through a mechanism initiated by PI3K enzymes phosphorylated by IRS proteins[2]. Phosphorylated Akt can then catalyse the phosphorylation of FOXO proteins, inhibiting their translocation to the nucleus and proapoptotic gene expression, and mTOR, promoting growth and protein synthesis. Phosphorylation by active Akt also promotes translocation of the muscle GLUT4 receptor to the cell membrane leading to increased glucose uptake, and induces inactivation of glycogen synthase kinase (GSK), allowing for hepatic synthesis of glycogen when glucose and insulin levels are high[40].

It should be noted that there are three different isoforms of Akt that seem to mediate distinct cellular processes: Akt1 has been associated with the regulation of somatic growth, while Akt2 appears to be involved in metabolism regulation, and specifically glucose homeostasis, and Akt3 may have a role in brain development[2, 40]. The diversity of cellular effects mediated by Akt seems to arise from the distinct functions of the three isoforms, despite their homology, which may be a result of, among others, different tissue distribution or different subcellular localization of the isoforms that conditions access to specific substrates, as appears to be the case of Akt2, in the regulation of glucose transport[41]. In fact, the different compartmentalization of each isoform within the cell, which seems to be independent of the phosphorylation status, has been verified: Akt1 is located mostly in the cytoplasm, Akt2 at the mitochondria and Akt3 in the nucleus[42].

Insulin resistance is a condition in which the cells are incapable of eliciting a normal response to the presence of insulin, resulting in generalized decreased glucose uptake and hyperglycaemia, as well as in increased synthesis of glucose, due to impairment of its suppression, and decreased glycogenesis in the liver. Compensatory hyperinsulinemia results from adaptation of the pancreatic β-cells to the high levels of blood glucose[40]. Besides

being hallmarks of type 2 diabetes, low insulin sensitivity (insulin resistance) and low glucose effectiveness (disposal mechanisms independent of high insulin) accurately predict the development of the disease and can precede diagnosis in prediabetic individuals by more than 10 years, independently of bodyweight[43]. A study involving Pima Indians (an ethnic group with high prevalence of T2DM) led to similar observations: insulin resistance was the strongest predictor of T2DM, but central distribution of body fat and high levels of fasting plasma glucose and insulin also preceded the onset of the disease[44]. However, insulin resistance can reveal selectiveness, and present repression of only some targets of insulin signalling, while others remain sensitive, which is evidenced by high levels of circulating lipids (free fatty acids and triglycerides)[40]. This can also be an effect of hyperinsulinemia on tissues and/or pathways not involved in glucose metabolism that may retain some sensitivity to insulin, such as the liver and the lipogenesis pathway. The phenotypical diversity of insulin resistance and T2DM may depend on the tissues and pathways affected[2].

1.5. Experimental models of insulin resistance and type 2 diabetes

As type 2 diabetes is becoming more and more prevalent in nowadays society, use of accurate experimental models for insulin resistance and type 2 diabetes is essential for research on the molecular basis and pathogenesis of the disease, for the development of an efficient treatment and, eventually, discovery of a cure[45]. Biological and clinical research have had major developments largely due to the use of model organisms, the ones yielding high efficiencies being favoured for their short generation times, easy breeding or propagation, simple manipulation, inexpensiveness, wide availability and homology to human systems[46]. Model organisms have allowed identification of conserved core biochemical principles and processes and include vertebrate and invertebrate animals, plants, bacteria and fungi[46]. Rodent models, including the mouse (*Mus musculus*), are preferred in studies about human type 2 diabetes because of their biological similarities to human characteristics and disease, be it monogenic or polygenic; however, the multitude of clinical presentations of T2DM makes it hard to find the ideal model for any particular study[45]. Several models for insulin resistance have been proposed and put into practice, either based on a dietary, genetic, surgical or chemical/pharmaceutical approach[2, 45, 47, 48]. High-fat diet, for example, can affect the circadian period and clock rhythmicity, therefore also disrupting metabolism regulation[26], and these effects can be exacerbated by frequent feeding or partly

reversed by time-restricted feeding and/or calorie restriction[27, 29-31]. *In vivo* models have focused mainly on genetic manipulation of the insulin signalling pathway to clarify its mechanism of action. Tissue-specific knockouts of the insulin receptor (IR) have been particularly useful in determining the role of many tissues on insulin resistant organisms; liver insulin receptor knockout (LIRKO) mice have exhibited hyperglycaemia due to the inability of suppressing glucose production in the liver, compensatory hyperinsulinemia (aggravated by decreased hepatic insulin degradation), a decrease in lipid synthesis and secretion and in SREBP expression, and overexpression of the gluconeogenesis key enzymes phosphoenolpyruvate carboxykinase and glucose 6-phosphatase[2, 49]. The leptin deficient *ob*/*ob* mouse is another used model for insulin resistance, showing hyperinsulinemia, hyperglycaemia and a decrease in IRS and PI3K activity and Akt phosphorylation in the liver, but in this model the effects can't be directly traced to changes in insulin signalling since the mice are hyperphagic and thus morbidly obese as well[2]. Therefore, different models lead to distinct changes in insulin signalling. In this project, we decided to take a pharmaceutical approach, using an antagonist molecule to the IR. Insulin receptor antagonism is commonly used to induce insulin resistance and can be achieved by binding of either antibodies or small peptides to specific sites that regulate the receptor's activity[47, 48]. Small peptides that have been artificially synthesized either agonize or antagonize the binding of insulin to its receptor, based on the structure of the IR and on its binding model. IR activation was only achieved using heterodimer peptides of the sites 1 and 2, not with homodimers that would promote site 1-site 1 or site 2-site 2 crosslinking. But orientation was also fundamental in determining the cellular effects of these molecules: site 2-1 dimers had mostly agonistic properties while site 1-2 dimers were antagonists[50]. S961 is a biosynthetic 43-aminoacid single chain site 1-2 peptide which shows insulin-antagonistic activity *in vitro* and *in vivo* through high-affinity and high-selectivity binding to the insulin receptor. This antagonist shows similar insulin receptor affinity and higher selectivity for IR over IGF-1 receptor, when compared to insulin itself[51]. Sprague-Dawley rats injected with S961 (10 or 30 nM/kg either subcutaneous, intravenous or intraperitoneally) exhibited a reversible increase in plasma glucose, which peaked 6h after injection, as well as a compensatory rise in serum insulin levels, and a long-term depletion of liver glycogen and body lipid storage, after prolonged exposure to S961 (injection of 10 nM/kg/day for 15 days)[52]. These results illustrate the role of insulin signalling in glucose metabolism

regulation and provide significant evidence that S961 could be used as an experimental model for insulin resistance and for induction of type 2 diabetes hallmark features (hyperglycaemia, hyperinsulinemia, glucose intolerance, among others)[51, 52]. For these reasons, injection with S961 was the chosen method to induce insulin resistance in this project. However, there has been evidence that S961 is not a full antagonist of the insulin receptor, but rather a mixed agonist/antagonist, acting as a partial agonist at lower concentrations (1 to 10 nM) and as an antagonist at concentrations of 10 nM or higher. These results were observed *in vitro* only for IR and Akt phosphorylation and mitogenicity; glycogen synthesis and lipogenesis were fully antagonized by S961[48]. In a different study, the antagonistic effects of S961 were compared to those of the specific allosteric monoclonal antibody IRAB-B. This newly-developed antibody was shown to induce acute and prolonged insulin resistance *in vitro* and *in vivo* by decreasing IR and Akt phosphorylation and increasing glycaemia, but, in contrast to S961, doesn't require repeat administration and binds even in the presence of insulin. S961 binds orthosterically and competes with insulin for the binding site, resulting in less expressive antagonistic effects at higher insulin concentrations[47]. The proposed model for binding of S961 to the IR shown in figure 5 may help explain these findings.

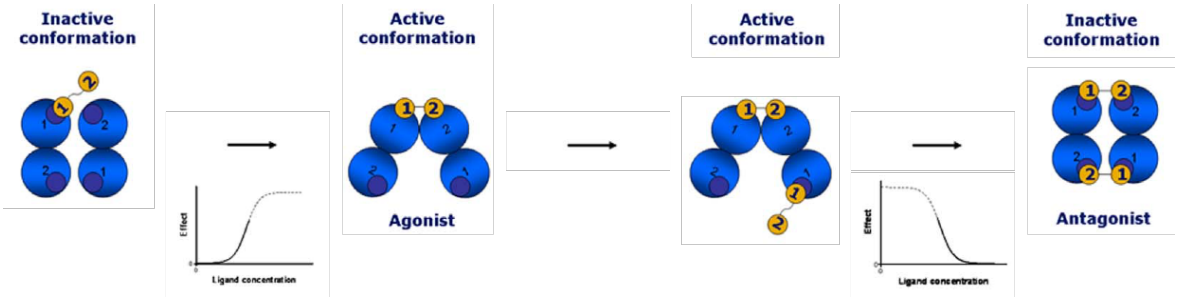


Figure 5. *Plausible model for the mechanism of binding of S961, shown as two connected yellow circles, to the active site of the insulin receptor (represented by the four blue circles), explaining how S961 may lead to different effects (agonistic or antagonistic behaviour) according to its concentration. In this model, low concentrations of S961 (1–10 nM) result in crosslinking of the receptor subunits in a similar way to insulin binding, mimicking its action, while higher concentrations allow for simultaneous crosslinking of the second pair of binding sites promoting an inactive conformation of the receptor and blocking insulin signalling. Insulin, as a less flexible molecule, can't bring together the second pair of binding sites on the IR and thus only elicits the active conformation upon binding (adapted from [48]).*

This model proposes that the IR has two identical monomers each containing a pair of partial binding sites (site 1 and 2) on the α-subunits, that are arranged antiparallelely, instead of symmetrically, as shown in figure 3. Insulin binding to a site 1 first, and then to the reciprocal one (site 2) on the other monomer would lead to crosslinking of the two α-subunits of the

receptor, which would become active, and the two-site binding model explains how S961 could elicit different effects according to its concentration, with a bell-shaped dose-response curve[48]. However, this still doesn't explain the fact that the agonistic effects only show in some cell processes but not in others. S961 was still chosen as the inducer of insulin resistance since high concentrations warrant antagonism of the insulin receptor and no other natural or modified ligands show antagonism of the IR. Novel monoclonal antibodies such as the IRAB may pose a viable alternative, but this antagonist was only reported when this project was already well underway.

1.6. Project aims

This project's main goal was to identify insulin-regulated enhancers in liver tissue and potentially find transcription factors that respond to insulin signalling, regulate activity of enhancers and thus control transcription of insulin-responsive genes. Preliminary work from the Grøntved group suggested that the activity of several enhancers in the liver is modulated by feeding, but the effect of insulin signalling in the regulation of the activity of these enhancers should be clarified. This was done by using a mouse model of insulin resistance induced by a synthetic antagonist (S961) that allows to study the regulation of transcriptional plasticity in the liver in response to feeding. Besides, identification of genes, enhancers and transcription factors affected by the insulin resistant state was also a priority along this project, due to the high prevalence of type 2 diabetes in nowadays society.

2. Materials and methods

In order to meet the goals set out for this project the following experimental setup was conceptualised and made into a scheme, as shown in figure 6. On the day of the animal experiment, mice were fed at their usual time (ZT11) and were subjected to S961 for 3 hours until sacrifice at ZT14. This scheme represents the initial plan for my project work; however, due to time limitations, it was not possible to conduct any experiments probing for H3K27ac.

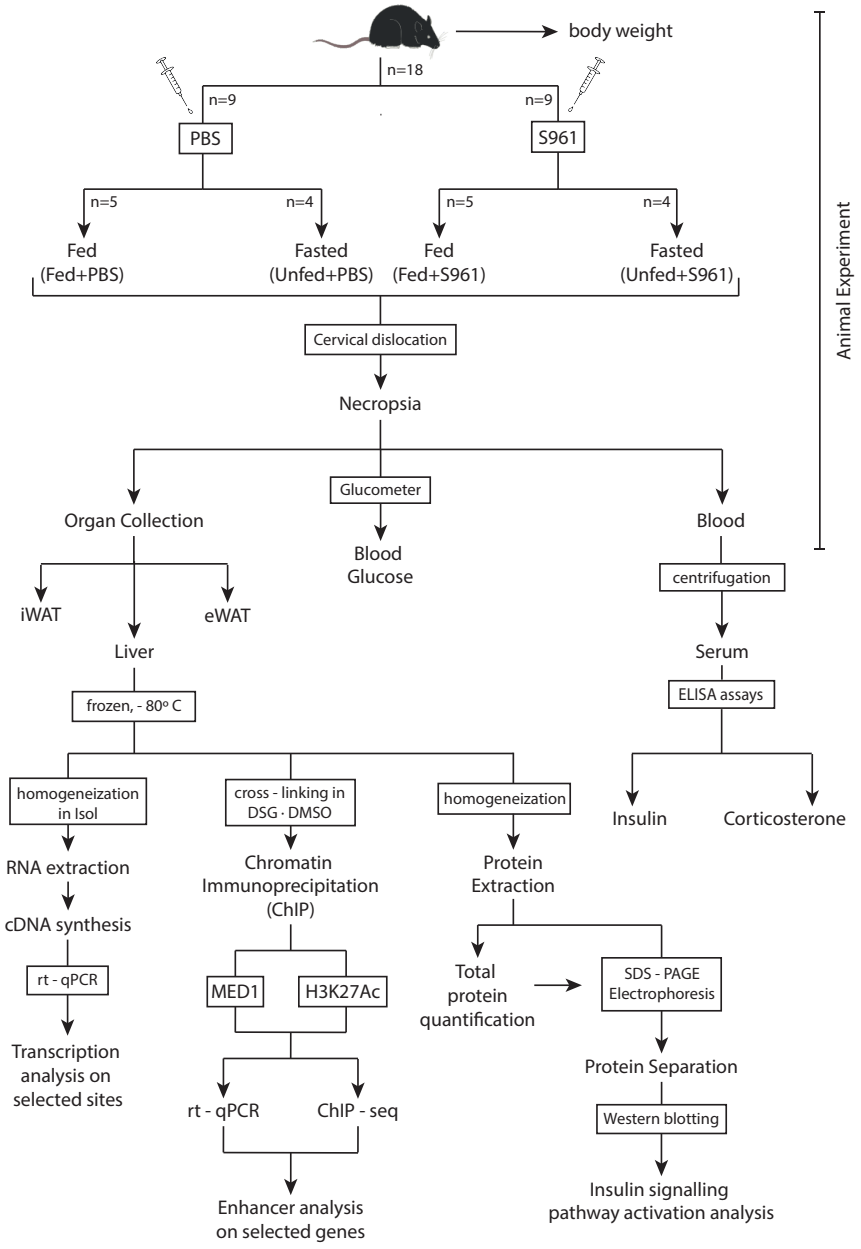


Figure 6. *Experimental setup designed to identify enhancers involved in feeding and insulin response in the liver. Mice were trained to a fasting/feeding program for 3 days. On the day of the experiment, PBS/S961 was intraperitoneally injected at ZT11 (40 nM/kg). Then mice were fed/fasted and sacrificed at ZT14.*

2.1. Animal experiment

Eighteen C57BL/6N male mice were purchased aged 8 weeks old from Taconic Biosciences, Inc. The mice were kept in controlled temperature (21 ± 2 °C), humidity (40 - 65%) and light (12 h light / 12 h dark) settings. Experimental procedures were approved by the Danish Animal Experiments Inspectorate (Dyreforsøgstilsynet).

After one week of acclimatization, mice were divided into single cages. Three days later, they were subjected to a fasting/feeding training program for three days, in which mice were fed (standard chow) only during the night period, with food being taken away at 8 a.m. and given back at 5.30 p.m. On the day of the experiment, at the end of the fasting period (ZT11), mice were weighed and randomly separated into two experimental groups. Mice from the first group were injected with PBS buffer (PBS, n=9) and mice from the second group with an S961 solution, an insulin antagonist (S961, n=9). Right after injection, food was given back only to 5 randomly selected mice from each group, resulting in four different treatment groups. Fed mice made up groups Fed+PBS and Fed+S961, while fasted mice made up groups Unfed+PBS and Unfed+S961. At ZT14, mice were sacrificed by cervical dislocation. After sacrificing the mice, blood glucose was measured (using a glucometer), blood was immediately collected and subsequently centrifuged for biochemical analyses in the serum, stomach and intestines were checked for abnormalities and the liver, epididymal white adipose tissue (eWAT) and inguinal white adipose tissue (iWAT) were collected, frozen in liquid nitrogen, preserved in dry ice and stored at -80 °C.

2.2. Analysis of biochemical parameters in the serum

Total blood was centrifuged at 10,000 rcf for 10 minutes, at room temperature, to obtain the serum. Corticosterone and insulin concentrations were measured in the serum of all the animals from every experimental group, by an ELISA assay (Ultra Sensitive Mouse Insulin ELISA kit, Crystal Chem Inc., cat. no. 90080; Corticosterone ELISA kit, Enzo Life Sciences Inc., cat. no. ADI-900-097).

2.3. Liver sample preparation for protein extraction

A small piece of liver, of approximately 30 mg, was shortly homogenized in 1mL of lysis buffer (containing Tris-HCl, glycerol, SDS, β-glycerophosphate, NaF, sodium orthovanadate and water) to which PMSF (200 μM, Sigma-Aldrich) and a protease inhibitor

cocktail (cOmplete tablet, EDTA-free, Roche) were added, using an IKA T10 standard Ultra-Turrax homogenizer. Samples were then subjected to a benzonase treatment, by adding 3 μL of benzonase to 1 mL of sample, mixing well (without vortexing) and leaving at room temperature (RT) for 15 minutes. After spinning down, the supernatant was kept and stored at -80 °C until further analysis.

2.4. Total protein quantification in liver extracts

Total protein concentration in the samples obtained (see 4.3.) was determined by the spectrophotometric BCA method modified by Lars Grøntved. The calibration curve was obtained by making measurements with bovine serum albumin (BSA) standards with different well-established concentrations. Samples and standards were loaded in 96-well microplates (10 μL), to which 200 μL of working solution (Pierce BCA Protein Assay Kit, Thermo Fisher Scientific) were added. Afterwards, plates were incubated for 30 minutes at 37 °C and optical density values were determined at 562 nm in a FLUOstar® Omega microplate reader (BMG LabTech). An example calibration curve is presented as an appendix (Appendix C).

2.5. Protein separation by SDS-PAGE electrophoresis

Polyacrylamide gels (10%) were prepared by cross-polymerization of acrylamide and bisacrylamide (30%, 29:1, Carl Roth GmbH) in the presence of ammonium persulphate (APS, Sigma-Aldrich) and tetramethylethylenediamine (TEMED, Sigma-Aldrich) in denaturing conditions provided by the addition of the anionic surfactant sodium dodecyl sulphate (SDS, Invitrogen, 0.4%). Each gel was cast with 15 wells. Pre-cast SDS-PAGE gels could be stored at 4 °C for up to 2–3 weeks if kept in a wet setting. Samples and a molecular weight marker were loaded into the gels. The loaded volume of sample was calculated so that total protein would amount to 80 μg in each well. Samples from the same experimental group were pooled into the same well so that less gels were required. Total volume loaded into each well was 20 μL, after adding protein lysis buffer, bromophenol blue (to confer colour to the samples) and mercaptoethanol (to denature proteins by breaking disulphide bonds). The loading mix containing sample and mercaptoethanol was boiled for 3 minutes in a heating block (95 °C) and centrifuged (maximum speed, 3 minutes) before being loaded into the gels, to complete protein denaturation. After loading the gels with the denatured samples, the gels were placed in the electrophoresis apparatus, covered with running buffer

(1x TGS, BioRad Laboratories) and run under a constant electrical current (20 mA per gel, for 2–2.5 hours) until the tracking dye reached the bottom of the gel, to induce proper protein separation in each sample/track, according to its molecular weight.

2.6. Protein analysis by Western blotting

Proteins separated in the electrophoresis gels were then transferred to polyvinylidene fluoride (PVDF) membranes (Amersham Hybond P 0.45 PVDF, GE Healthcare Life Sciences), by putting each gel and its corresponding membrane into contact and applying an electrical current, under transfer buffer, to induce the transfer (1h run at 400 mA, 100 V). To confirm protein transfer from the gels to the membranes, the latter were stained with an Amido Black 10B solution. After scanning and washing with ddH_2O to remove excess staining solution, the membranes (blots) were blocked with a 5% BSA solution in TBS (milk was not used due to probing of phosphoproteins), for 1 hour at RT or overnight at 4 °C, in gentle agitation, in order to avoid non-specific binding of the primary antibody to the membrane and minimize false positives. The membranes were then incubated overnight with a solution of the primary antibody specific for the protein to be analysed. Analysed proteins were β-tubulin as a positive control and native and phosphorylated β-IGF-IR, Akt, FOXO1A and GSK-3β. Antibodies used to probe for those proteins are described in table 1. Before and after incubating the membranes with the primary and secondary (1h at RT) antibodies, they were washed 3x in 1x TBS, and then with 1x TBS-T (3x 10 minutes, under gentle agitation) to remove leftover blocking solution or primary/secondary antibodies. Finally, membranes were incubated with a solution containing luminol and hydrogen peroxide (Pierce ECL Western Blotting Substrate, Thermo Fisher Scientific), which, upon reaction with the HRP (horseradish peroxidase) conjugated with the secondary antibody, results in the production of light from the protein bands recognised by the antibodies. The light emitted was digitally registered by an ImageQuant LAS 500 scanner (GE Healthcare Life Sciences) or a ChemiDoc XRS+ scanner (BioRad Laboratories). After exposure and scanning of the membrane, it was possible to identify and relatively quantify the proteins associated with their respective antibodies.

Table 1. *Specifications of the antibodies used to probe for the target proteins analysed by Western blotting.*

Analysed Protein	Primary antibody	Secondary antibody
β-tubulin	1:2500, mouse mAb, #05-661, Merck Millipore	1:2000 HRP-anti-mouse Ab, #7076, Cell Signaling Technology **(1)**
β-IGF-IR	1:1000, rabbit pAb C-20, sc-713, Santa Cruz Biotechnology	1:3000 HRP-goat anti-rabbit Ab, #7074, Cell Signaling Technology **(2)**
p-β-IGF-IR (phospho-Ser256)	1:1000, rabbit pAb, #3024S, Cell Signaling Technology	(2)
Akt	1:1000, rabbit pAb, #9272S, Cell Signaling Technology	(2)
pAkt (phospho-Ser473)	1:1000, rabbit pAb, #9271S, Cell Signaling Technology	(2)
FOXO1A	1:500, rabbit pAb, ab39670, abcam	(2)
pFOXO1A (phospho-Ser256)	1:500, rabbit pAb, ab131339, abcam	(2)
GSK-3β	1:1000, rabbit pAb, H-76, sc-9166, Santa Cruz Biotechnology	(2)
pGSK-3β (phospho-Ser9)	1:1000, mouse mAb F-2, sc-373800, Santa Cruz Biotechnology	(1)

2.7. RNA extraction from liver samples

RNA was extracted from liver samples using an in-house column purification procedure. A small portion of mouse liver (kept in dry ice to avoid thawing and tissue degradation), of approximately 10 mg, was weighed into a 2 mL Eppendorf tube for each sample and immediately homogenised in 700 µL of Isol-RNA Lysis Reagent (5PRIME), using an IKA T10 standard Ultra-Turrax probe. Samples were subsequently always kept on ice. Chloroform (0.2x volume of Isol) was added to the homogenised liver samples to allow for nucleic acids extraction. After an incubation period of 15 minutes (RT), the samples were centrifuged at 4 °C, 10,000 xg for 10 minutes. The upper aqueous phase, containing the RNA, was carefully retrieved (260 µL) to new 1.5 mL Eppendorf RNase-free tubes, which were then placed on ice to minimize endogenous RNase activity. The interphase (containing DNA) and lower organic phase (containing proteins) were discarded. Ethanol (96%, 0.6x volume of Isol) was added to the samples and this mix was transferred to EconoSpin columns (Epoch Life Science), which was centrifuged for 30 seconds at 13,000 xg, RT. Afterwards, the columns were washed 3 times with RPE buffer (Qiagen, 500 µL). Each wash was

followed by a spin at 13,000 xg for 30 seconds. Finally, the column was spun dry for 2 minutes at 13,000 xg, RT and the ethanol was let evaporate at RT. The RNA trapped in the column was then eluted in DEPC-treated water (≥ 97%, Sigma-Aldrich, 30 μL) by centrifuging at 16,000 xg, RT, for 1 minute. The RNA solutions were transferred to 1.5 mL Eppendorf tubes. RNA concentration was measured on a 2 μL sample on the NanoDrop ND-1000 Spectrophotometer (Thermo Fisher Scientific Inc.), and RNA samples were stored at -80 °C.

2.8. cDNA synthesis (reverse transcription)

Using the RNA samples obtained from step 4.3. as template, I obtained cDNA samples suitable for real-time quantitative PCR (qPCR). For each sample analysed, the volume of sample amounting to 1 μg of RNA (according to the quantification of RNA described previously), 2 μL of 5x DNase buffer, 1 μL of DNase I (Invitrogen, 10 U/μL) and DEPC-treated water to a total of 10 μL were mixed and incubated for 15 minutes at 37 °C to ensure that all remaining DNA was destroyed. Samples were placed on ice to minimize RNase activity (and thus RNA degradation) and random hexamers were added (Roche, 3 μg). After 5 minutes of incubation at 85 °C, to inactivate the DNase enzyme and allow for base pairing between the hexamers and the RNA, samples were put back on ice and were added 5 μL First Strand buffer (Invitrogen), 2.5 μL DTT (Invitrogen, 100 mM), 1 μL DEPC-treated water, 2.5 μL dNTP mix (10 mM of each dNTP) and 1 μL Reverse Transcriptase (Invitrogen, 200 U/μL). This mix was left at RT for 10 minutes and then incubated for 1 hour at 37 °C to ensure that the reverse transcription reaction was complete. At the end of the incubation time, double-distilled water (ddH_2O, 290 μL) was added to each sample tube. cDNA samples were stored at -20 °C.

2.9. Primer design for analysis of cDNA by PCR

To perform qPCR on the DNA samples obtained from the isolated RNA, it was necessary to design some of the primers used to select the analysed sites. First, the gene location under analysis was found on the Ensembl Genome Browser, filtered for the mouse genome. Then, for the pre-mRNA primers, the full gene sequence (including both intron and exons) was obtained and copied to the Primer3 online software (v. 0.4.0) to pick primers from the DNA sequence. It was defined that primers should have a melting temperature in the range of 59–

61 °C and that the product size should be between 75 and 125 bp in length. The Santa Lucia salt correction formula was applied and complementarity and 3' complementarity were limited to 1.0. Finally, the primer possibilities returned by the Primer3 software were checked for their melting temperatures, ΔG value (should be between -3 and -6 kcal/mol) and number of hairpins, primer dimers and self-dimers they could form on the OligoAnalyzer online tool (from IDT, Integrated DNA Technologies). The best candidate sets were selected and its sequences were BLAST'ed on the UCSC Genome Browser to assure they identified the target sequence. The primers were tested on a series of dilutions of a pooled cDNA sample from the Fed+PBS group (because their target genes were feeding-upregulated) to obtain a Cp value *vs.* dilution curve and the primer set with the best efficiency (closest to 100%) was chosen. The details for the designed primers are shown in table 2.

Table 2. *Specifications of the primer set designed to analyse SREBP-1c mRNA expression in cDNA samples obtained from mouse livers through rt-qPCR.*

Oligo	Sequence	Start	Tm (°C)	% GC	Any compl.	3' compl.
Left Primer	GGGGTAAAGGTCGGCAAGGCA	16553	59.03	61.90	2.00	0.00
Right Primer	CAGCGGCAGGCTAGATGGTGG	16655	59.58	66.67	4.00	0.00

The resulting primer set created an amplification product located in the SREBP-1c gene transcript with a length of 103 bp. The primer pair was then tested in a series of dilutions (1x, 2x, 4x, 8x, 16x) prepared from a cDNA sample belonging to the experimental group Fed+PBS (since this should be the group with higher SREBP-1c mRNA levels considering it is a feeding upregulated gene). The primer set had an efficiency of 91.2%. The way the primer efficiency is calculated after testing is demonstrated in Appendix B.

2.10. Chromatin Immunoprecipitation (ChIP)

Firstly, a DSG-DMSO-PBS (2 mM) crosslinking solution was prepared by diluting 400 μL of DSG-DMSO solution (0.5 M) per 100 mL of DSG-DMSO-PBS solution. DSG-DMSO solution was added to PBS very slowly, with extreme caution, at RT on full speed on the magnetic stirrer, to avoid DSG polymerization and subsequent precipitation. DSG-DMSO solution was made by dissolving DSG (ProteoChem, MW = 326,26 g/mol) in the appropriate amount of DMSO (Sigma-Aldrich, ≥ 99.5%) to achieve a concentration of 0.5 M.

Liver samples (approx. 100 mg) were weighed and homogenised in 7 mL of a DSG-DMSO-PBS solution, and then left rotating for 30 minutes at RT, to allow for crosslinking of protein-

protein interactions between co-factors and transcription factors (TF) associated with the DNA. This first crosslinking step can be skipped if doing TF or histone modification ChIP, but it should always be included if doing co-factor ChIP. Afterwards, the pellet was spun down at 1300 rpm, 4 °C for 10 minutes and re-suspended in 7 mL PBS + 1% Formaldehyde (Sigma-Aldrich, 36.5–38%) for 10 minutes while rotating, to allow for crosslinking of DNA-proteins (transcription factors) interactions. Then, 1 mL of glycine (1 M) was added to quench the crosslinking reaction and samples were centrifuged for 10 minutes at 1700 rpm, 4 °C. The supernatant was discarded and the pellet was washed twice in 10 mL cold PBS. Each wash was followed by a centrifugation at 1700 rpm, 4 °C (10 minutes).

Proceeding to the isolation of nuclei and preparation of chromatin, samples were diluted in 5 mL ChIP lysis buffer 1 (1% SDS, 20 mM EDTA, 50 mM Tris-HCl pH 8) and incubated for 1.5 hours at RT, while rotating, to allow for cell lysis and nuclei release through membrane disruption. The nuclei were then centrifuged at 1700 rpm, 8-10 °C, for 10 minutes to precipitate. At proper nuclei release, the pellet should be cloudy. The pellet was washed twice in 10 mL cold PBS; each wash was followed by a centrifugation at 1700 rpm, 8-10 °C (10 minutes). Samples were again diluted, this time in 2 mL ChIP lysis buffer 2 (0.1% SDS, 1% Triton, 150 mM NaCl, 1 mM EDTA, 20 mM Tris-HCl pH 8.0) + 1x PI + 1 μg/μL BSA, and then sonicated for 90 seconds on the Covaris ME220 focused ultra-sonicator (peak power 75; 10% duty factor; 200 cycles) to fragment the DNA crosslinked to TFs and co-factors and centrifuged at 10,000 xg, 4 °C, for 1 minute to precipitate cell debris. Supernatant was transferred to new tubes, to which enough ChIP lysis buffer 2 + 1x PI + 1 μg/μL BSA was added to total 400 μL of sample. 5% of the volume of each sample was saved for input, enabling input-output analysis later on.

In the meantime, while waiting for nuclei extraction, 2x 30 μL (for pre-clearing and for immunoprecipitation, IP) of raw Prot A/B beads (Santa Cruz, sc-2003) per sample were prepared. Since the beads come in 4x ethanol, 240 μL of bead solution was used for each sample. After being centrifuged at 700 xg for 1 minute to remove the ethanol, the beads were washed three times in 1 mL ChIP lysis buffer 2 and centrifuged at 700 xg for 1 minute after each wash, removing the supernatant. Finally, 2x 70 μL of ChIP lysis buffer 2 + 1x PI + 1 μg/μL BSA per ChIP reaction was added to the beads, which were then incubated for at least 30 minutes at 4 °C, while rotating, until needed for the next step. BSA ensures low non-specific binding between beads and antibody.

For the chromatin IP step, the samples were first precleared by adding 100 μL of the beads solution previously prepared to each of the samples and letting them incubate for 30 minutes at 4 °C, while rotating, to avoid non-specific binding between some proteins and the beads in the absence of antibody and thus minimize background signal. Samples were then centrifuged at 700 xg for 1 minute and the beads (pellet) were discarded. Next, the IP reaction took place by adding the specific antibody (2 μg) and BSA (5 μL) to the cleared samples containing the target-protein, letting them incubate overnight at 4 °C, while rotating, and then adding 100 μL of beads solution to each sample, followed by an incubation time of 3h at 4 °C, while rotating, to allow for binding of the protein-bound antibodies' heavy chain to the beads. This part of the procedure, as well as the next one, should be executed in a cold room set at 4 °C. The antibody targeted MED1 (M-255, sc-8998, Santa Cruz Biotechnology).
A washing step followed, with a total of 6 separate washes, to eliminate non-specific binding and remove DNA fragments not associated with beads and antibody (meaning it wasn't originally bound to the target-protein). Before the first wash, samples should be centrifuged at 700 xg, 4 °C, for 1 minute, and supernatant should be discarded. All washes were done with 500 μL buffer at 4 °C for 5 minutes, while rotating, and followed by a spin-down at 700 xg, 4 °C, for 1 minute. The sequence of washing buffers used is 2x #1; #2; #3; 2x #4, decreasing in concentration.
Afterwards, samples were eluted and decrosslinked. Firstly, 500 μL of elution buffer (1% SDS, 0.1M $NaHCO_3$) was added to each of the samples, but also to the input samples previously saved and stored at -20 °C (on these samples, the sample volume should be considered to the total volume, so only 480 μL were added). The elution buffer aids in the decrosslinking reaction and in the denaturation of proteins. Samples and input were let to rotate at RT for about 30 minutes. Samples were then spun-down at maximum speed to allow deposition of the protein/antibody-bound beads, now separated from the DNA, and the supernatant (containing the DNA still bound to some proteins) was transferred to new tubes, to which 5M NaCl (20 μL) was added to assist decrosslinking. The same amount of 5M NaCl was also added to the input tubes. After vortexing, all samples were let to decrosslink overnight at 65 °C.
IP and input samples were then subjected to phenol-chloroform extraction of the DNA, by generation of an organic-aqueous phase, that separates DNA (water-soluble) from proteins and cell debris. 1 volume (500 μL) of Phenol Chloroform Isoamylalcohol (PCI, Sigma-

Aldrich, 25:24:1 saturated with 10 mM Tris, pH 8.0, 1 mM EDTA) was added to all tubes (ChIP/input) to cause protein denaturation leading to a conformational change towards a hydrophobic-behaving structure (phenol) and allow for a sharper separation of the phases (chloroform). Isoamylalcohol is present to reduce the foaming caused by SDS added in previous steps. After thoroughly vortexing, mixing by inversion and centrifuging for 5 minutes at maximum speed and RT, the aqueous upper phase was transferred to new tubes, to which 1 volume (approx. 460 µL) of chloroform was added. All samples were again thoroughly vortexed, mixed by inversion and centrifuged (5 minutes, max. speed, RT) and the aqueous upper phase was transferred to new tubes, to which 2 µL glycogen (2 µg/µL), 40 µL sodium acetate (3M, pH 4.5) and 1 mL ice cold ethanol (100%) were added. Sodium acetate, as well as ethanol, neutralizes the negative charges in DNA, making it insoluble in water and allowing it to precipitate. Glycogen is insoluble in ethanol, precipitating along with DNA and making it easier to see the pellet. Samples were thoroughly vortexed, mixed by inversion and incubated for 1h at -20 °C, after which the DNA was precipitated by centrifuging at full speed, 4 °C, for 25 minutes. The supernatant was removed and the DNA-containing pellet was washed with 200 µL ethanol (70%), to remove any remaining salts, and spun-down at maximum speed for 5 minutes at 4 °C. Supernatant was again removed and the ethanol was left to evaporate (samples were completely dry when the pellet became transparent and no longer white). The pellet was ressuspended in milli-Q water (ddH_2O), with the volume depending on the intended analysis to be undertook. Samples were stored at -20 °C until further analysis by RT-qPCR or ChIP-seq.

2.11. Primer design for ChIP-qPCR

To perform rt-qPCR on the DNA samples obtained either from ChIP, it was necessary to design the primers used to select the analysed sites. First, the gene location under analysis was found on the UCSC Genome Browser, filtered for the mouse genome. From the tracks showing H3K27 acetylation on fed and unfed animals, I zoomed in the areas of the gene most likely to be transcriptionally active (seen as a valley flanked by peaks, indicating the location of a promoter or enhancer). The DNA sequence was viewed (with masking of repeats as N) and copied to the Primer3 online software (v. 0.4.0). Thereon, the process was similar to the one used for mRNA primers (melting temperature: 59–61 °C; product size: 75–125 bp; Santa Lucia salt correction formula; complementarity and 3’ complementarity:

1.0). Again, melting temperatures, ΔG value (should be between -3 and -6 kcal/mol) and number of hairpins, primer dimers and self-dimers of each possibility was checked with the OligoAnalyzer tool (IDT, Integrated DNA Technologies) to select the most suitable sets. The sequences were BLAST'ed on the UCSC Genome Browser to make sure the target site was recognized and to get the exact chromosome location. The primers were tested on a series of dilutions of a pooled ChIP DNA input sample to trace a Cp value *vs.* dilution curve. The primer set with the best efficiency (closest to 100%) was chosen.

2.12. Real-time qPCR (RT-qPCR)

Real-time qPCR was performed on DNA samples obtained either from ChIP or from cDNA synthesis from RNA samples as described on section 4.4. or 4.5., respectively. Samples were loaded on 384-well plates. Firstly, a mastermix containing SYBR Green Mastermix (Fast Start Essential DNA Green Master, Roche, 3.5 μL), 5' and 3' primers (10 pmol/μL, 0.5 μL of each) specific for the analysed gene and ddH_2O (0.5 μL) was prepared for each analysed gene. Then, prepared mastermixes were distributed in the plate (5 μL in each well) and samples were added to individual wells (2 μL, in duplicate, for each gene) after vortexing. The plate was carefully covered with a sealing foil and spun for 2 minutes at 1200 rpm. PCR reactions were run on the LightCycler® 480 Instrument II PCR machine (Roche) under the "SYBR Long Program" (95 °C/2 min and 45 cycles 95 °C/10 s; 60 °C/15 s; 72 °C/15 s). Melting curve analysis was performed by the machine while the samples were heated from 60 to 95 °C for 10 minutes. Data was saved in *.txt* format showing Cp values for each sample/well.

2.13. Evaluation of sonication efficiency on the Fragment Analyzer

Efficiency of sonication was assessed on ChIP input samples through an automated capillary electrophoresis on the AATI (Advanced Analytical Technologies, Inc.) Fragment Analyzer™ to check if the ChIP samples had enough quality for sequencing. A 96-well non-skirted plate was loaded with 30 μL of each sample, 30 μL of 1 kb plus DNA ladder FS-SLR930-U100 (on the last well of the last used row) and 30 μL of milliQ-water on every unused well of the partially-loaded rows. All loaded wells were then covered with one droplet of mineral oil, to avoid evaporation of the samples, and the plate was spun down at 1200 rpm, RT, for 1 minute. The gel used (Gel2) was dsDNA 930 separation gel, to which 2 μL of intercalating

dye were added per 20 mL of gel, and the marker was a combination of 75 bp and 20,000 bp markers (FS-SMK930-003). With the waste tray empty, the buffer and marker trays full and all solutions and plates loaded on the appropriate trays of the Fragment Analyzer, the samples were run under the method "DNF 930-33-DNA 75-20000 bp_sonication test", with a sample injection time of 3 to 10 seconds.

2.14. Quantification of DNA in ChIP samples

Before sequencing, it is necessary to know the concentrations of DNA in each ChIP sample so that enough amount is prepared. DNA concentrations were determined using the PicoGreen assay. 2 µL of each sample was loaded into a black 96-well plate with opaque bottoms and 98 µL of TE buffer (10 mM Tris + 1 mM EDTA, pH 8.0) was added to each used well. λ-DNA standards were prepared by dilution of a 2 ng/µL stock solution to concentrations of 10, 5, 2.5, 1.25, 1, 0.5, 0.25 and 0 ng/100µL and 100 µL of each was loaded in the plate, in duplicate. Next, 17.5 µL of PicoGreen were mixed with 3.5 mL of previously prepared 1x TE buffer and 100 µL of this solution was added to each of the already loaded wells. Fluorescence was then measured in the FLUOstar Galaxy luminometer (BMG LabTech), using the "Fast Start" setup named "Pico Assay MS". An example calibration curve for determination of DNA concentration is presented as an appendix (Appendix D).

2.15. Sequencing of ChIP samples (ChIP-seq)

Two pooled samples per experimental group (Fed+PBS and Unfed+PBS) were prepared so that each got material from three ChIP samples (corresponding to 2 individuals per pooled sample), amounting to 5 ng of DNA in total, and milli-Q water was added to total 30 µL. Samples were numbered SM2659 to SM2662. The samples showing outlier results in ChIP-qPCR were excluded from sequencing. These pooled ChIP samples were then subjected to library preparation for Illumina® sequencing, conducted by Majken Siersbæk, post-Doc with the Grøntved group, which I observed.

The first step consists of adding 20 µL of End Repair Mix to each sample and incubating at 30 °C for 30 minutes, to blunt unpaired nucleotides (overhangs) into phosphorylated ends. Then, the DNA fragments present in the sample were size selected to isolate only the fragments with a length of approximately 200 to 300 bp. The Agencourt AMPure XP beads (Beckman Coulter) that size-selectively immobilise the fragments need to be equilibrated to

RT and vortexed to resuspend before being used in a series of clean up steps. By first adding 37.5 μL of beads to each sample and letting them incubate at RT for 5 minutes (0.75x clean up), the fragments larger than 300 bp become immobilised in the beads and can be discarded by placing the tubes in a magnetic stand for 5 minutes (RT) until samples look clear and transferring the supernatant to new tubes. Next, adding 50 μL of new beads to the supernatant (followed by 5 minutes of incubation at RT) allows for immobilization of fragments longer than 200 bp in the beads (> 1x clean up). These fragments can subsequently be recovered by keeping the beads and discarding the supernatant after 5 minutes on the magnetic stand at RT. Without removing the tubes from the magnetic stand, the beads should be washed twice with 175 μL of fresh 80% ethanol, leaving to incubate for 30 seconds each time and letting the beads dry at RT for 5 minutes after removing all ethanol on the second wash. DNA can then be eluted from the beads by dissolving and incubating them in 20 μL of Resuspension buffer for 2 minutes (RT). Another 5 minutes on the magnetic stands allows to save the supernatant containing the size-selected DNA fragments and discard the beads.

Afterwards, 3'-ends are adenylated by adding 12.5 μL of A-tailing mix to the samples, which then are incubated for 30 minutes at 37 °C. This process leads to the addition of an adenosine monophosphate residue to the 3'-ends of the DNA fragments, which will facilitate ligation of an adapter sequence (hairpin). After incubation, 20 μL of Resuspension buffer are added, making up a total volume of 50 μL, and a new clean up step follows by adding 37.5 μL of beads to each sample, letting incubate for 5 minutes (RT) and discarding the supernatant (containing fragments shorter than 200 bp) after being left at the magnetic stand for 5 minutes (1x clean up). Beads should then be washed in 175 μL of 80% ethanol. After removing the ethanol, beads were let to dry. DNA fragments were eluted from the beads by suspending and incubating them in 22 μL of Resuspension buffer for 2 minutes (RT). Supernatant was subsequently transferred to new tubes and beads were discarded.

At this point, the ligation of a hairpin adapter to the 3'-adenylated and 5'-phosphorylated DNA fragments was promoted by incubation of the samples, at RT for 15 minutes, in an adapter ligation mix containing the ligase enzyme master mix and the hairpin adapters. After incubation, 1 μL of User Enzyme (New England Biolabs, 1000 U/mL) was added to each sample, and a new incubation period followed (37 °C, 15 minutes). Resuspension buffer (24 μL, total volume of 50 μL) was again added, leading to another clean up step: 45 μL of beads, incubation at RT for 5 minutes (0.9x clean up), placement on the magnetic stand for

5 minutes and discarding of the supernatant, resulting in removal of adapter remains since only fragments longer than 200 bp had been immobilised in the beads. Samples were washed twice in 175 µL of 80% ethanol and then let to dry; then, DNA was eluted from the beads by adding 52 µL of Resuspension buffer and incubating the mixture for 2 minutes (RT). The supernatant was saved and transferred to new tubes and a new 0.9x clean up step followed (45 µL of beads, elution in 22 µL of Resuspension buffer). At the end, 20 µL of the eluted DNA solution should be transferred to new tubes.

The DNA samples selected and purified by this procedure were then amplified by PCR, which selectively enrichs the sample in adapter-ligated DNA fragments. The samples (10 µL) were added to 10 µL of a reaction mix consisting of: Q5 reaction mix (New England Biolabs), dNTPs (New England Biolabs, 10 mM of each), Q5 polymerase (New England Biolabs, 2000 U/mL), ultra-pure water, universal PCR primer (New England Biolabs) and index primer (New England Biolabs). 16 PCR cycles were run due to the low amount of starting material (5 ng of DNA), with the following settings: 98 °C/30 sec, 16 cycles 98 °C/10 s; 65 °C/30 sec; 72 °C/30 sec and a final cycle at 72 °C for 5 minutes. At the end of the PCR reactions, 28 µL of Resuspension buffer were added to each sample and a 0.8x clean up step followed: 40 µL of beads; 5 minutes' incubation at RT; magnetic stand for 5 minutes; discarded supernatant; twice washing with 80% ethanol (175 µL); removal of ethanol; and drying of the beads. At last, the DNA fragments immobilised in the beads (> 250 bp) were eluted in 22 µL of Resuspension buffer, incubated at RT for 2 minutes, placed on the magnetic stand for 5 minutes. 20 µL of supernatant were saved and transferred to new tubes; beads were discarded.

Quality control of the sample library was done on the Fragment Analyzer, where a test was performed to guarantee the absence of adaptor dimers and of amplified fragments not belonging to the target size range. The samples were run under the setup "DNA 35-1500 bp_ChIP_QC". Besides that, the library concentration was determined using KAPA.

2.16. Data analysis and statistics

Analysis of RNA-seq and ChIP-seq data was conducted using the programs Terminal, Filezilla, Microsoft Excel and RStudio. RNA-seq data files were retrieved from the server and converted using Terminal and Filezilla. Alignment of sequences was achieved by applying the STAR algorithm, using mm9 as animal model (reference genome) to map the data. Analysis of differential expression of sequenced genes was done with the open source software DESeq2 (from Bioconductor)[53]; sequencing analysis and motif discovery was performed with HOMER (Hypergeometric Optimization of Motif EnRichment, from Salk Institute, UCSD)[54]. Illustrations were made resorting to the program RStudio, using R language code. Data analysis of systemic parameters and RT-qPCR results was done in Microsoft Excel. Statistical significance of the differences between experimental groups was assessed by a Student's t-test (two-tailed) and a p-value (or FDR) below 0.05 was considered significant. Results are shown as average ± standard deviation. Western blots were only qualitatively analysed.

3. Results and Discussion

3.1. RNA-seq characterisation of feeding-regulated genes

To provide a characterisation of the genes my research project is focused on, I started by analysing pre-existing RNA sequencing data from mice subjected to the same treatments as the ones applied on my animal experiment (see experimental setup in figure 6). The R-coding script created to analyse this data is shown in Appendix A. Sequenced reads from cDNA libraries were quantified at annotated genes using HOMER and analysed by DESeq2 to call statistically significant regulated genes by multiple testing, providing the false discovery rate (FDR). Genes with FDR<0.05 were considered differentially regulated. Figures 7 and 8 show the results of the preliminary analysis of sequenced mRNA from all annotated genes, highlighting the feeding-regulated (figure 7, up-regulation in green and down-regulation in red) and insulin-regulated (figure 8, in blue) sites.

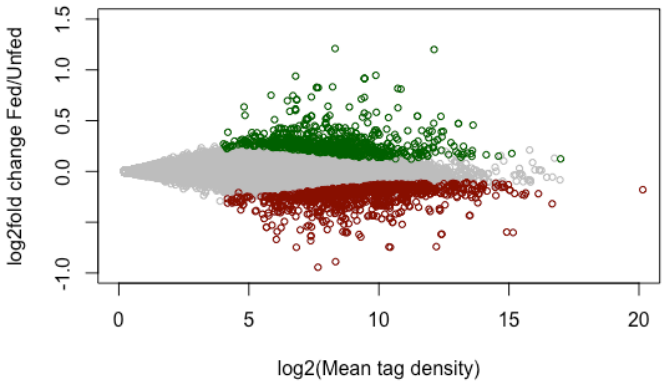


Figure 7. *Distribution of all genes analysed by RNA-sequencing (RNA-seq), with feeding-regulated genes highlighted (FDR<0.05 for Fed* vs. *Unfed groups). Feeding upregulated genes, located at the top of the plot, are shown in green, while feeding down-regulated genes, located at the bottom of the plot, are shown in red.*

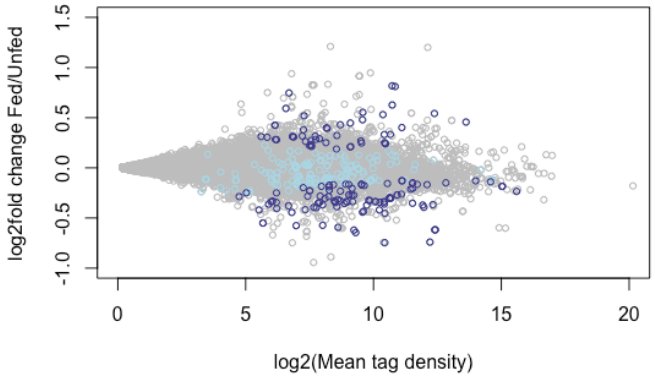


Figure 8. *Distribution of all genes analysed by RNA-sequencing (RNA-seq), with insulin-regulated genes highlighted (FDR<0.05 for Fed* vs. *Fed+S961 groups): shown in light blue (insulin regulation not associated with feeding response) and dark blue (overlap with feeding regulation).*

From these scatter plots, it is possible to see that both feeding- and insulin-regulation occur mostly on highly-expressed genes, demonstrated by the high tag densities (log_2(mean tag density) > 5), and that significant differential expression in response to both feeding and insulin signalling is more prevalent at large fold changes between fed and fasted states. From a total of 31,645 mRNA sequences, 1,884 genes are regulated by feeding (about 6%), and 846 of those show up-regulation. My project's focus is on these positive regulated genes, which comprise about 45% of the feeding-regulated genes and 2.7% of all sequenced genes. Furthermore, I am looking to identify genes and transcriptional enhancers associated with insulin regulation. 325 genes are affected by S961 injection immediately before feeding, which corresponds to 1% of the total analysed genes. From further analysis (data partially shown to avoid excessive overlapping of points and crowding of the plot; see appendix A for extra plots), it was also possible to observe that only about half of the S961-regulated sites are located in the feeding-regulated areas of the plot, showing that some feeding-independent genes are regulated by S961 (which could or could not correspond to regulation by insulin). This could point to some artefacts created by this drug or to activation of alternative signalling pathways by S961. For example, S961 may affect insulin signalling in adipose tissue which may affect gene expression in the liver. S961/insulin regulation overlaps with feeding-mediated regulation of transcription in 151 genes, which corresponds to 0.5% of all genes, or 8% of all feeding-regulated genes.

Afterwards, since the focus of this project lies on the feeding up-regulated genes, I decided to analyse only these genes and highlight the insulin-regulated genes (figure 9).

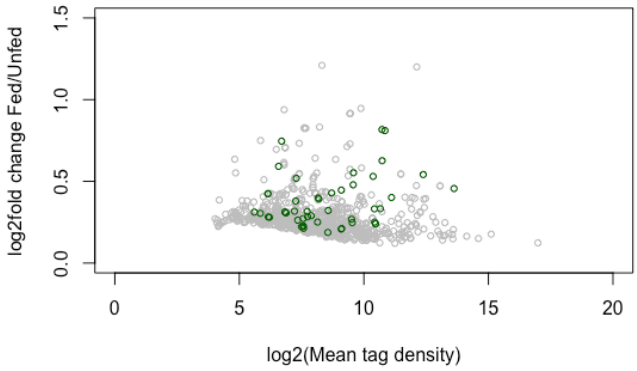


Figure 9. *Focus on the feeding up-regulated genes, with S961-regulated genes (FDR<0.05 for Fed* vs. *Fed+S961 groups) highlighted in green.*

Forty-eight of the 846 genes upregulated in response to feeding are also regulated by S961 and thus presumably by insulin. This corresponds to 5.7% of feeding-upregulated genes, or

2.5% of all feeding-regulated genes (0.2% of the total of analysed genes). Therefore, only a minority of the feeding regulated genes in the liver are deregulated upon systemic insulin resistance (2.5%). This apparently low percentage of insulin regulation of cellular effects of feeding could either indicate that there's significant regulation by alternative mehanisms (such as direct regulation by glucose) or be a result of artefacts of the action of S961 (the antagonist), as it may not antagonize all of insulin's effects on transcription. It can also derive from the systemic effects of S961, such as hyperinsulinemia, which might activate other regulatory pathways, independently of activation of the receptor.

Finally, the same data was organised in a way to show the effect of administrating the insulin antagonist S961 to fed mice on the feeding-upregulated genes (figure 10).

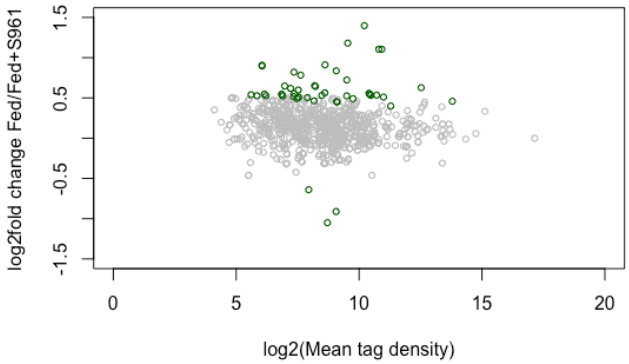


Figure 10. *Distribution of feeding up-regulated genes in terms of insulin regulation. S961 up- and down-regulated genes are highlighted in green.*

From this plot, it is possible to see that there are 3 genes that, though being up-regulated by feeding, are down-regulated by the action of insulin. Moreover, 2 of these 3 genes (Klf10 and E2f8) encode for transcription factors, which can show that insulin regulation may also have an indirect action on gene expression by altering the expression of transcription factors that will later modulate the expression of other feeding-regulated genes. The third gene (Slc25a30) encodes for a mitochondrial solute carrier protein. The 45 feeding upregulated genes that are repressed by S961 (thus presumably upregulated by insulin as well) include Srebf1 (sterol regulatory element-binding protein), Sqle (squalene epoxidase), Gck (glucokinase) and Pik3r1 (phosphatidylinositol 3-kinase regulatory subunit alpha) as some of the most feeding and insulin upregulated genes of the set. All of these genes are metabolism regulators. Srebf1 and Sqle encode for proteins involved in sterol biosynthesis, while Gck encodes for a key enzyme of glucose metabolism and Pik3r1 is a central messenger molecule in the insulin signalling pathway. Also, Gck, Srebf1 and Pik3r1 are well known targets of the insulin receptor signalling pathway.

3.2. Systemic evaluation of sacrificed mice

In order to characterise the systemic insulin resistant state of the animals injected with the insulin antagonist S961, each mouse's blood glucose was measured immediately after sacrifice. The mice were also weighed prior to injection with either PBS or S961 to account for their initial state after acclimatisation and before the experiment was carried out. The results are shown in table 3.

Table 3. *Mice body weight at the beginning of the experiment (after acclimatization) and effect of feeding and/or injection with insulin antagonist S961 in blood glucose (results are shown as average ± standard deviation). Experimental groups (for future reference):* Fed+PBS *(normal fed animals);* Fed+S961 *(insulin-resistant fed animals);* Unfed+PBS *(normal fasted animals);* Unfed+S961 *(insulin-resistant fasted animals).*

Experimental Groups	**Body weight** (g)	**S961 injection**	**PBS injection**	**Blood glucose** (mM)
ZT14 Unfed + PBS	21.18 ± 4.36	–	√	6.12 ± 1.34
ZT14 Unfed + S961	22.08 ± 1.84	√	–	11.59 ± 2.92*
ZT14 Fed + PBS	23.12 ± 1.66	–	√	6.34 ± 1.04
ZT14 Fed + S961	22.30 ± 1.23	√	–	17.65 ± 2.99+

* $p < 0.05$ *vs* Unfed+PBS*;* + $p < 0.05$ *vs* Fed+PBS

The initial body weight did not show any significant differences between experimental groups, as seen on the data shown in table 3, which suggests a uniform acclimatization across all individuals and random attribution of mice to experimental groups and thus rules out the possible influence of body weight on insulin action. As expected, the administration of S961, a known antagonist of insulin, caused a significant increase in blood glucose relatively to the relevant control group, injected with PBS, as described in previous studies[52]. This increase reflects the blocking of the insulin receptor and decreased glucose uptake, and is a good indicator that insulin resistance has been induced on the S961-injected mice, since insulin signalling has been blocked and thus glucose is not incorporated into the cells.

3.3. Analysis of serum biochemical parameters

The serum samples analysed in this project were obtained by centrifuging the clotted blood and saving the serum (liquid portion). The concentrations of insulin and corticosterone were determined in the serum of the mice from every experimental group, by means of ELISA assays, to test the suitability of the insulin resistance animal model through evaluation of the systemic effects of feeding and insulin antagonist S961. The results are shown in figure 11.

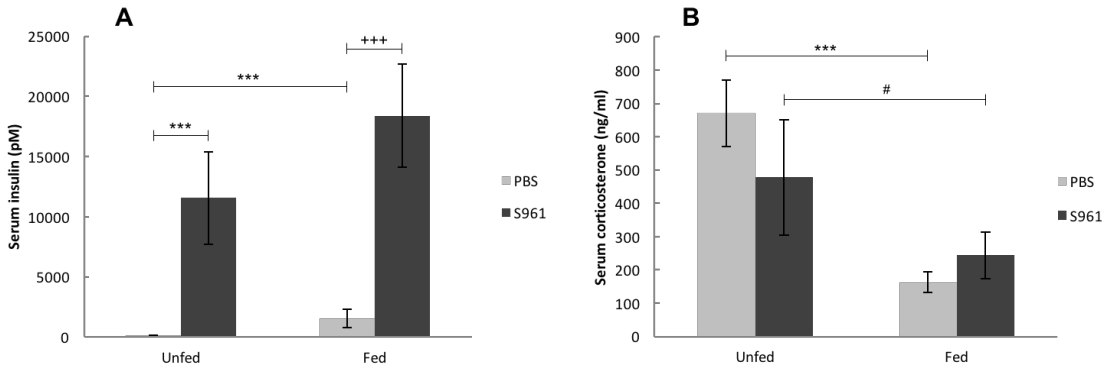


Figure 11. *Effect of feeding and/or injection with insulin antagonist S961 on the levels of serum insulin (**A**) and serum corticosterone (**B**). Results are shown as average ± standard deviation.*

*** $p < 0.001$ *vs.* Unfed+PBS; +++ $p < 0.001$ *vs.* Fed+PBS; # $p < 0.05$ *vs.* Unfed+S961

As can be seen in figure 11-A, both feeding and administration of the insulin antagonist S961 cause an increase in the levels of serum insulin, as expected and previously shown[52]. The increase in serum insulin after feeding on control groups (injected with PBS) can be explained by the normal physiological response to the increased levels of blood glucose achieved after a meal, where the pancreatic β-cells produce and release insulin in order to stimulate the tissues to take up glucose from the blood and metabolise it to produce energy for the body or to create stores as glycogen. On the other hand, the increase in insulin levels after administration of S961 in both Fed and Unfed groups should indicate the proper blocking of the insulin receptor on the target tissues by the insulin antagonist (S961), since this would inactivate the insulin signalling pathway, thus limiting the uptake of glucose by the cells and leading to an accumulation of glucose on the blood, which would then trigger a compensatory mechanism in which pancreatic β-cells over-produce insulin to try to lower blood glucose levels, resulting in hyperinsulinemia[40].

Figure 11-B shows a decrease in corticosterone levels in the serum with feeding in both control and S961 groups. Being a glucocorticoid hormone, corticosterone is a stress-reponse hormone and it is also involved in glucose metabolism regulation. Its release is stimulated by stress and by low blood glucose, which conforms with the observed results. Glucocorticoid hormones, best represented by cortisol, have a contrary effect to insulin and are known to promote gluconeogenesis in the liver and inhibit glucose uptake in the muscle and adipose tissue during fasting, to restore blood glucose levels. Cortisol also indirectly stimulates glycogenolysis in the liver through glucagon activation. Besides that, previous results from the Grøntved group have shown an increase in corticosterone levels just before feeding time, and other studies have suggested a diurnal variation in the levels of cortisol[55],

with peaks at the start of the active period, just before feeding. Therefore, the increased corticosterone levels observed in the fasted groups in figure 11-B seem to result from a normal response to lowered blood glucose and, possibly, from a circadian anticipation of feeding. Also, the higher levels of corticosterone in the fasted groups in contrast with the absence of a significant difference with administration of S961 in either Fed or Unfed groups could show that all animals were subjected to the same basal level of stress during handling and euthanasia, and that the fasting period was the only relevant source of stress during the experiment, minimising the influence of stress levels on the metabolism regulation and associated signalling pathways.

3.4. RT-qPCR analysis of RNA expression

Besides characterising the feeding- and insulin-regulated genes by statistical analysis of RNA-seq data, and analysing anthropometric and serum parameters, I also evaluated the differential expression of canonical feeding-regulated genes (involved in metabolism regulation) by qPCR analysis on samples of cDNA synthesised from RNA extracted from the liver. I intended to check if the inhibition of insulin signalling was producing the expected results on gene expression in the liver, and thus validate the RNA-seq data. Results are shown in figures 12 to 14.

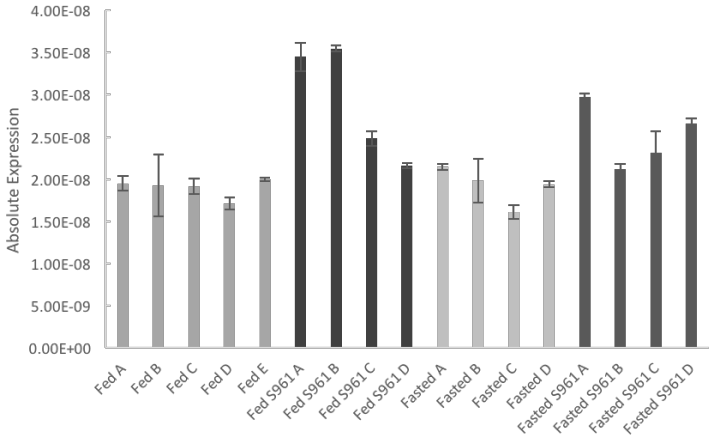


Figure 12. *Effect of feeding and/or insulin antagonist S961 on the mRNA expression levels of transcription factor IIB, or TFIIB, evaluated by rt-qPCR in liver tissue. Results are shown as replicate average ± standard deviation.*

From figure 12, it is possible to verify that there is little variation in the mRNA levels of transcription factor IIB (TFIIB, a ubiquitously expressed gene) among individuals, even belonging to different experimental groups. This not only serves as a positive control, but

also shows that cDNA synthesis was fairly uniform among all samples, and thus provides a reliable basal level against which other genes' expression can be normalized. Therefore, in this experiment, TFIIB can be used as a reference gene.

In order to evaluate the normal functioning of the chosen animal model, and thus assess the differential expression of well-known feeding-regulated genes in the liver, mRNA expression levels of specific sequences (transcripts) were evaluated via RT-qPCR. I focused on well-known and highly regulated genes involved in glucose (Gck, G6Pc, Pck1) and lipid metabolism (SREBP-1c). Even though the focus of my study lies on feeding-upregulated genes, for this purpose I chose two genes that are upregulated and two that are downregulated by feeding to get a better insight into the way the antagonist works and the effect of feeding and insulin on the expression of such genes. Results obtained are shown in figures 13 and 14. Figure 13 focuses on the differential expression of two feeding-upregulated genes (Gck and SREBP-1c) in response to feeding and/or insulin antagonist S961, while figure 14 shows the effect of the same treatments on two feeding-downregulated genes (G6Pc and Pck1).

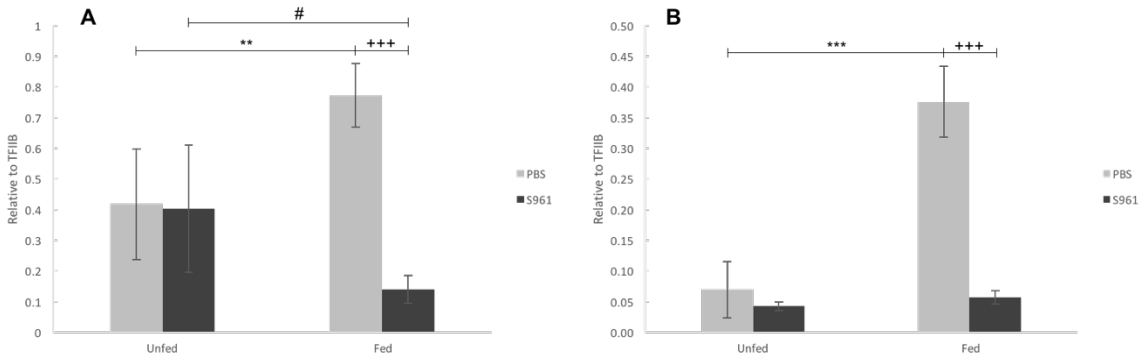


Figure 13. *Effect of feeding and/or insulin antagonist S961 on the mRNA expression levels of two canonical feeding-upregulated genes, coding for glucokinase, or Gck (**A**), and the catalytic subunit of the sterol regulatory element-binding protein, or SREBP-1c (**B**), in the liver, evaluated by rt-qPCR. Results are shown as group average ± standard deviation.*

*** $p < 0.001$ *vs.* Unfed+PBS; ** $p < 0.01$ *vs.* Unfed+PBS;
+++ $p < 0.001$ *vs.* Fed+PBS; # $p < 0.05$ *vs.* Unfed+S961

The plots in figure 13 show two feeding up-regulated genes. Glucokinase (Gck) is the enzyme that catalyses the first reaction of glycolysis, a key regulatory step in glucose metabolism, and of glycogen synthesis. Glucokinase is therefore highly expressed after meals to enable storage and, later, breakdown of the available glucose. Sterol regulatory element-binding proteins (SREBPs) are transcription factors that bind to the sterol regulatory element sequence in DNA in response to low levels of sterols to stimulate expression of enzymes involved in sterol synthesis, which are essential for fat anabolism (namely,

cholesterol and fatty acid synthesis). SREBP-1c, the liver-specific isoform, which responds to insulin signalling, is particularly important in regulating lipogenesis after feeding[49]. From the results, it is clear that both Gck and SREBP-1c are upregulated in the fed state, when compared to the fasted group, and also that the administration of the insulin receptor antagonist on the fed groups leads to a huge decrease in mRNA levels of both genes. This confirms RNA-seq data, validates the hypothesis that there is feeding regulation of these genes, and confirms that this effect is due to the action of insulin, since blocking of insulin signalling by S961 inhibits the normal increase in transcription of these genes as a result of feeding. In the case of Gck, mRNA levels even drop below the basal levels observed in the fasted groups when S961 is administered to a fed group.

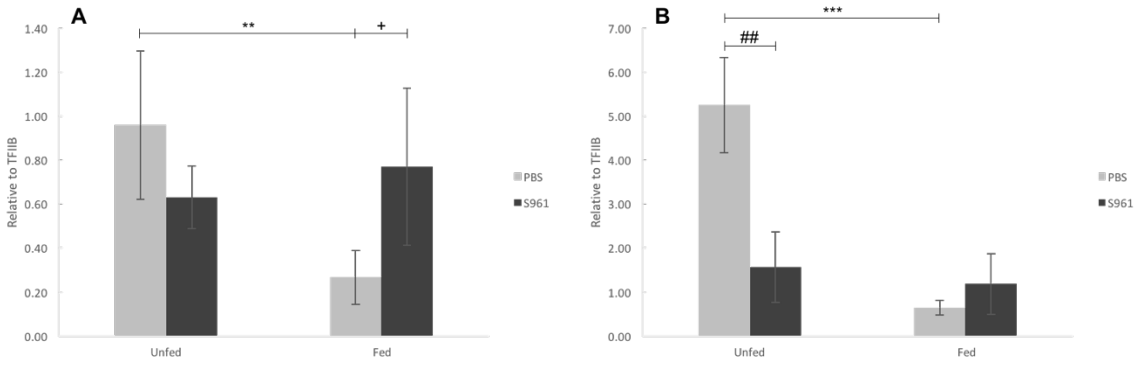


Figure 14. *Effect of feeding and/or S961 on the mRNA expression levels of two feeding-downregulated genes – the catalytic subunit of glucose-6-phosphatase, or G6Pc (**A**), and phosphoenolpyruvate carboxykinase 1, or Pck1 (**B**), in the liver, evaluated by rt-qPCR. Results are shown as group average ± standard deviation.*

*** $p < 0.001$ *vs.* Unfed+PBS; ** $p < 0.01$ *vs.* Unfed+PBS;
+ $p < 0.05$ *vs.* Fed+PBS; ## $p < 0.01$ *vs.* Unfed+S961

Finally, figure 14 shows the effect of the same treatments on two feeding downregulated genes encoding enzymes of gluconeogenesis. Both glucose-6-phosphatase (G6P) and phosphoenolpyruvate carboxykinase 1 (Pck1) are key enzymes in the gluconeogenesis metabolic pathway, and are thus upregulated during fasting periods, when blood glucose levels drop and the liver needs to produce glucose in compensation. G6P also catalyses the final step of glycogenolysis that leads to production of glucose through breakdown of hepatic glycogen stores. Results show a normal mRNA expression pattern in response to feeding and to blocking of insulin signalling by S961: feeding repressed both genes' mRNA expression. For G6Pc (figure 14-A), S961 caused an increase in mRNA levels in fed animals, which suggests that insulin was repressing G6Pc in the fed state and blocking of this pathway led to reversion of the feeding expression pattern back to fasting levels even after feeding.

In the case of Pck1 (figure 14-B), however, S961 only caused a decrease in mRNA levels in fasted animals, where insulin signalling should not be prevalent, and no significant difference was observed between the fed groups, which may mean that regulation of this enzyme by glucagon prevails over insulin.

Thus, the animal model seems to be working properly, since blood glucose, serum insulin and mRNA expression levels of feeding upregulated genes support S961 as an inducer of insulin resistance, a hallmark of type II diabetes.

3.5. ChIP experiment

PCR results of mRNA levels showed expected expression patterns in response to feeding and insulin signalling (demonstrated by IR antagonism induced by S961). Systemic effects were also as expected. We thus considered the insulin resistance animal model was working properly and was suitable for the study, and a MED1 Chromatin Immunoprecipitation (ChIP) experiment was designed to probe for feeding- and insulin-regulated enhancers.

3.5.1. Primer design for ChIP

To perform qPCR on the ChIP samples to assess MED1 recruitment to the selected feeding upregulated transcriptional enhancers, it was necessary to design the primer sets targeting putative feeding regulated enhancers near the the Gck (2 sites), SREBP-1c and PIK3R1 genes (DNA). For that purpose, I first had to select the sites on which I wanted to check for enrichment in MED1. I tried to choose possible regulatory regions (enhancers) that show an increase in H3K27 acetylation in response to feeding (unpublished H3K27Ac data from the Grøntved group), since this is associated with active enhancers and H3K27 acetylation shows some correlation with MED1 recruitment. Even though distant regions from the target genes can bridge to the promoter through looping of the chromatin strand, I focused on sites close to the gene promoters (transcription start site, TSS) because these are also associated with enrichment in H3K27ac and are easier to recognise. The Mediator complex acts as a molecular bridge between the activator proteins bound to enhancer sites and the transcription initiation complex (including RNA polymerase) bound to the promoter, and thus the promoter is the one gene region to which I was sure MED1 would be in close proximity in case of recruitment. With that in mind, I used the H3K27Ac Genome Browser tracks shown in figure 15 to identify putative regulatory elements.

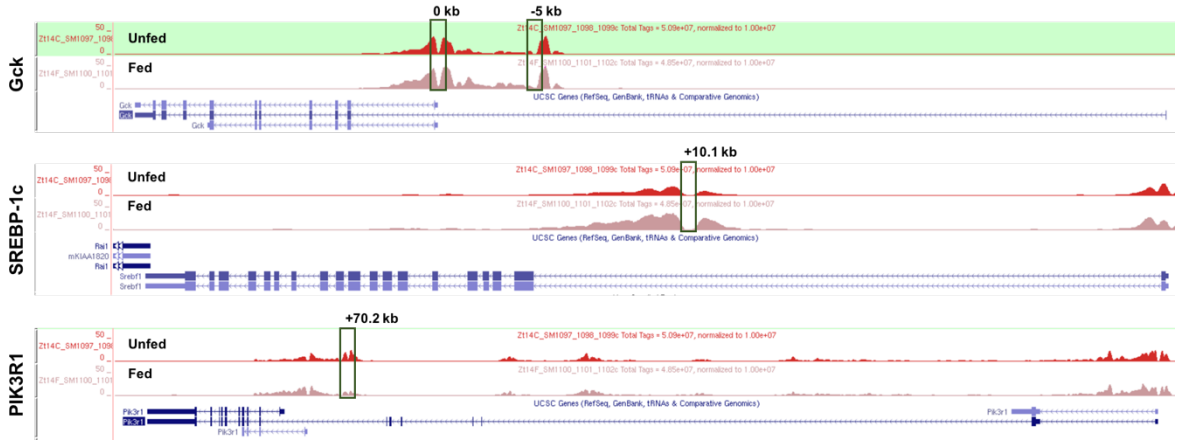


Figure 15. *UCSC genome browser tracks showing H3K27 acetylation across three feeding-upregulated genes (Gck, SREBP-1c, PIK3R1) on unfed and fed states (1st and 2nd track on each gene, respectively), with the selected primer sites highlighted for each gene.*

These tracks show acetylation of the lysine-27 residue of histone H3 (H3K27ac) along the genes under analysis (Gck, SREBP-1c and PIK3R1). This illustrates one of the modifications in chromatin landscape that elicit transcriptional activation. Peaks of H3K27ac point to transcriptionally active sites and are usually located flanking active promoters (at the TSS) or more distant enhancer sites. Histone molecules are temporarily dislocated from promoter and enhancer sites during active transcription (histone sliding) to allow for binding of RNA polymerase or activator proteins[3], originating valleys of acetylation on the Genome Browser tracks that correspond to open chromatin. Therefore, I tried to locate my primers on the valley areas of the tracks, surrounded by peaks of acetylation. The sites chosen for Gck and SREBP-1c do show an enrichment in H3K27ac upon feeding, but the primer site for PIK3R1 seems to have reduced H3K27 acetylation on the fed state, even though it is also strongly upregulated by feeding according to RNA-seq data. The other valley region seen on the tracks was also tried for primer design, but no options had enough quality to be further tested. Details for the designed primers that were ultimately selected are shown below in table 4.

Table 4. *Specifications of the primer sets designed to analyse MED1 enrichment near the Gck, SREBP-1c and PIK3R1 genes in ChIP DNA samples (from mouse livers) through qPCR.*

Gene	Oligo	Sequence	Tm (°C)	% GC	Any compl.	3' compl.
Gck 0kb	Left primer	CTGCCCTCCAGGTCTCCCCC	59.96	75.00	4.00	0.00
	Right primer	CCCCAACGACCCCTGCTTATCCT	59.99	60.87	2.00	0.00
Gck -5kb	Left primer	GGCCTGGTGAGTGGGAGGTGT	60.11	66.67	4.00	0.00
	Right primer	TGTCAGGGTGGGTTCTGCTGTCT	59.49	56.52	3.00	0.00
Srebf1	Left primer	CCGGGGTTACTAGCGGGCGT	60.67	70.00	4.00	0.00
	Right primer	GGGGATGGTTGCCTGTGCGG	60.32	70.00	2.00	1.00
Pik3r1	Left primer	AGGGAACAGGGGCTGAAGAACACA	59.75	54.17	3.00	0.00
	Right primer	ACCAGTCCCGGTGTGCTTTCTCT	59.80	56.52	4.00	0.00

The resulting primer sets created amplification products located near the Gck (2 sites), SREBP-1c and PIK3R1 genes with lengths of 125, 110, 108 and 77 bp, respectively. The primer pairs were then tested similarly to the primer pair designed for RT-qPCR in a series of dilutions (4x, 16x, 64x, 256x, 1024x, 4096x and 16348x) prepared from a pooled ChIP DNA input sample. The selected primer sets had efficacies of 95.0%, 100.5%, 99.9% and 118%, respectively. Calculation of primer efficacies is exemplified in Appendix B.

In addition, I used a primer for a non-coding region of the albumin gene ("Albumin no DHS") and a primer located in a heterochromatic intergenic region, that doesn't select for any gene regions ("No Gene"), as negative controls.

3.5.2. MED1 antibody testing and optimization of the sonication procedure

In order to maximize recoveries of DNA obtained after chromatin immunoprecipitation with a MED1-specific antibody, it was necessary to optimize the sonication procedure done with the Covaris ME220 focused-ultrasonicator. We wanted to check if the time-consuming nuclei extraction procedure that was in use could be replaced with a simpler whole-cell extraction procedure using higher sonication times. Additionally, due to discontinuation of polyclonal antibodies by Santa Cruz Biotechnology, Inc., I started testing a new monoclonal anti-MED1 antibody from the same company to check if it was a suitable alternative.

The results for the optimization and testing experiment are shown in figure 16, below.

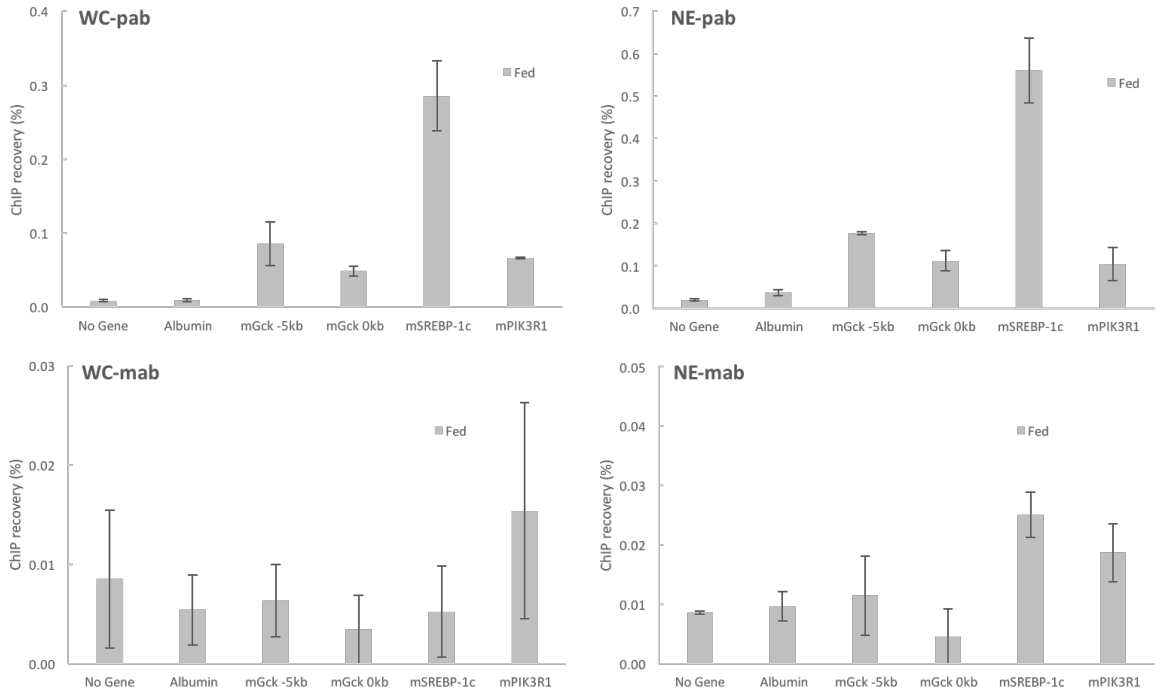


Figure 16. *ChIP recovery rates obtained with different extraction/sonication procedures and antibodies, as part of the first ChIP experiment (figure 11): whole-cell (WC) extraction with the polyclonal antibody (**WC-pAb**); nuclear extraction (NE) with the polyclonal antibody (**NE-pAb**); WC extraction with the monoclonal antibody (**WC-mAb**) and NE extraction with the monoclonal antibody (**NE-mAb**). Results are shown as group average ± standard deviation. No Gene was the reference site and albumin (no DHS) the negative control.*

From figure 16, there are two comparisons one can make: WC *vs.* NE (comparing recoveries obtained through two different DNA extraction methods and sonication procedures: WC refers to extraction from a whole-cell preparation, while NE refers to nuclear extraction, after isolation of the nuclei from the homogenised liver samples) and pAb *vs.* mAb (comparing the suitability of the two antibodies – currently used and potential alternative).

Firstly, comparing the two extraction and sonication procedures, WC and NE (top row of figure 16), NE seems to give better quality results than WC, since the nuclei extraction method shows higher recovery rates than the ones obtained with the whole-cell method and fold induction over "No Gene" is also higher using NE (data not shown), indicating lower background. Therefore, I decided to keep using the NE method with shorter sonication times in all subsequent ChIP experiments.

Regarding the antibody testing (pAb *vs.* mAb), it was concluded that the tested MED1 monoclonal antibody (H-7, sc-74475) should not replace the polyclonal MED1 antibody (M-255, sc-8998) currently in use, as it gives much lower recovery rates than those obtained with the polyclonal antibody and the analysed sites show about the same enrichment as the

negative controls, making them indistinguishable from background signals. The monoclonal antibody was then discarded. Further testing of alternatives will be conducted at a later point.

3.5.3. Analysis of MED1 ChIP data (Fed *vs.* Unfed experiment)

As mentioned above, a MED1 Chromatin Immunoprecipitation (ChIP) experiment was designed, where Mediator (MED) complex recruitment was analysed. In a first approach, DNA samples obtained from this procedure were analysed by qPCR to evaluate potential MED1 enrichment in specific sites near feeding upregulated genes. The genes chosen were Gck and SREBP-1c (see above), but also phosphatidylinositol 3-kinase regulatory subunit alpha (PIK3R1), an enzyme involved in the insulin signalling pathway, as part of the phosphorylation cascade; it is activated by IRS proteins and leads to phosphorylation of Akt[2]. Primers were designed so as to be located as close to the promoter region as possible, in transcriptionally active regions (see sections 1.3. and 3.5.1.).

It was decided to do a MED1 ChIP experiment on Unfed+PBS and Fed+PBS groups to check if feeding regulation of MED1 recruitment could be observed. With that in mind, I performed a ChIP experiment only on the fed and unfed control groups (injected with PBS). On this experiment, it was decided to test only two of the genes; Gck (only the -5kb region) and SREBP-1c were chosen for their higher fold changes between fed and unfed states on both RNA-seq and qPCR results, which led us to predict stronger regulation by feeding. Results from this experiment are shown below in figure 17.

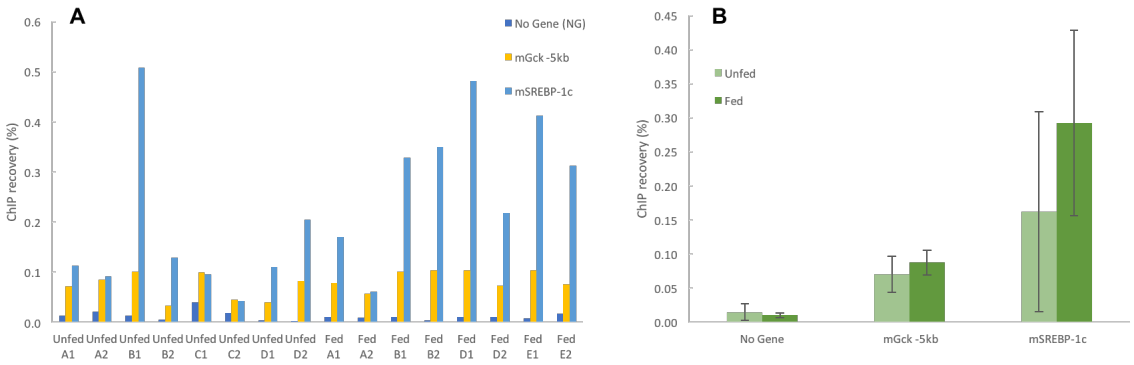


Figure 17. *Effect of feeding on the MED1 enrichment in four sites of three expected feeding-upregulated genes – Gck, SREBP-1c and phosphatidylinositol 3-kinase regulatory subunit alpha (PIK3R1) in the liver, evaluated by ChIP-qPCR. Results are shown as individual recoveries (**A**) and as group average ± standard deviation (**B**). No Gene was used as a reference site. Different letters indicate samples from different animals, while samples with the same letter and different numbers correspond to technical replicates split from the same sonicated sample and subject to separate, but parallel, IPs.*

From the DNA recovery rates obtained in the individual samples (figure 17-A), it is possible to identify two clear outliers, one in each experimental group (samples Unfed B1 and Fed A2). However, these outliers are not observed on both technical replicates of the biological sample they belong to, and therefore should only reflect some random experimental error (probably during the IP or DNA purification steps), not a biological deviation. Looking past those two technical outliers, there is a clear tendency for MED1 enrichment on both genes on fed animals. However, this tendency does not translate into a significant difference when considering the group averages including all technical replicates (figure 17-B) due to the presence of the outliers. Variation on MED1 enrichment on "No Gene" should be disregarded as DNA concentrations for this region were very low and therefore the high threshold cycle (Cp) values can lead to variation in the LightCycler's readings. That is the reason why results are shown as simple recoveries, and not normalized to "No Gene" levels (fold induction over NG). These results point to a potential feeding-regulation of Mediator recruitment to feeding-upregulated genes' enhancers. To try to identify some of these enhancers, we decided to sequence the DNA samples obtained (see section 3.6.).

3.5.4. Confirmation of sonication efficiency on the Fragment Analyzer

To check the efficiency of sonication and the quality of the ChIP DNA samples for subsequent sequencing, I ran a fragment analysis through an automated capillary electrophoresis (on the Fragment Analyzer) of the input samples from the MED1 ChIP experiment (Unfed *vs.* Fed). In figure 18, I include an example of the electropherogram obtained for each of the experimental groups.

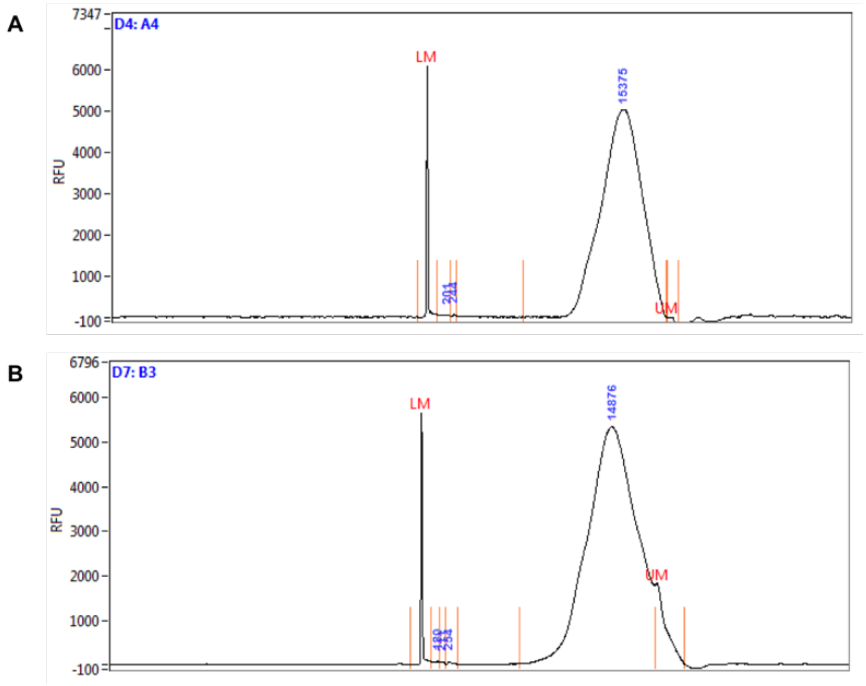


Figure 18. *Electropherogram showing size and amount (intensity) of the fragments of DNA formed after ChIP (due to sonication) of input samples A4 (**A**, Fed+PBS) and B3 (**B**, Unfed+PBS). The ladder used was "1kb plus DNA ladder FS-SLR930-U100" and 75 bp (lower marker, LM) and 20,000 bp (upper marker, UM) markers were run along with the samples.*

It is visible that all the DNA fragments obtained after sonication belong in the range limited by the two markers used (between 75 bp and 20,000 bp long). There is a big sonication-resistant peak corresponding to fragments of about 15,000 bp in size. However, a clear sonication peak consisting of fragments around 10 to 500 bp long would be expected and these plots don't clearly show such a peak. Although the amount of sonicated product in this size range was too low, this also shows that no over-sonication occurred, and therefore most, if not all, DNA fragments potentially isolated with a length of 10 to 500 bp should result from sonication of the regions bound to MED1, selected by the antibody during ChIP, and not be an artefact caused by excessive sonication leading to random short fragments. For that reason, it was decided to proceed to sequencing of the ChIP samples.

3.6. ChIP-seq data analysis

As mentioned above, sequencing was performed on pooled samples from the Fed vs. Unfed MED1 ChIP experiment. The outlier samples observed in figure 17 were not included in this analysis. I then made 2 pooled samples for each experimental group (Unfed+PBS and Fed+PBS), by combining 3 samples to obtain a total of 5 ng of DNA in each pooled sample, from which libraries were subsequently prepared for sequencing. Table 5 shows the samples that are a part of each pooled sample used for sequencing.

Table 5. *Samples prepared for ChIP-seq, from the MED1 ChIP Fed* vs *Unfed experiment discussed on 3.5.3.*

Pooled sample ID	Samples used	Mice included	Tissue	DNA (ng)
SM2659	Unfed A1; A2; B2	ZT14 Unfed+PBS A+B	Liver	5.0
SM2660	Unfed C1; C2; D2	ZT14 Unfed+PBS C+D	Liver	5.0
SM2661	Fed A1; B1; B2	ZT14 Fed+PBS A+B	Liver	5.0
SM2662	Fed D1; E1; E2	ZT14 Fed+PBS D+E	Liver	5.0

3.6.1. Library preparation quality control

After library preparation and before proceeding to sequencing, the samples were run on the Fragment Analyzer® to evaluate PCR efficiency and check if the fragments in the target size range (200-300 bp) were properly isolated to assess if the libraries generated were of the correct size. The results obtained are shown in figure 19.

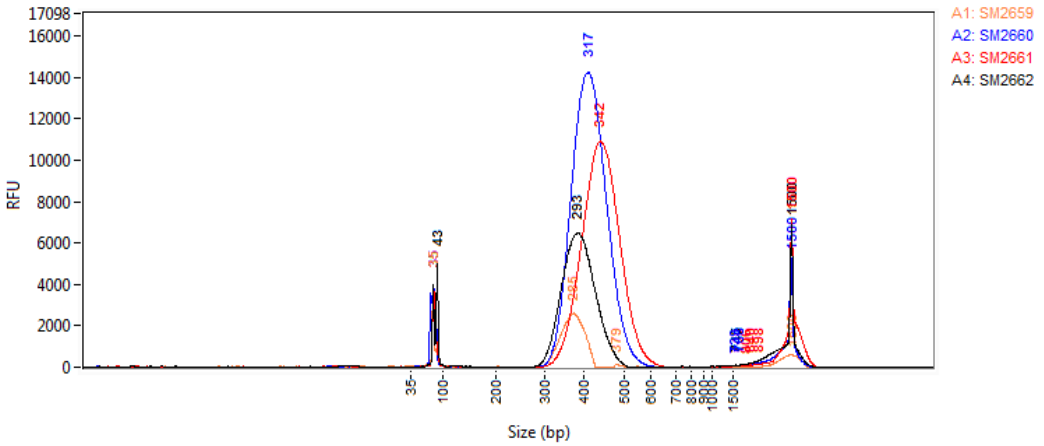


Figure 19. *Electropherogram showing the size of the DNA fragments selected after library preparation for ChIP-seq (where fragments in the range of 200 to 300 bp were isolated) of pooled samples SM2661 and SM2662 (Fed+PBS) and SM2659 and SM2660 (Unfed+PBS).*

From the electropherogram presented above, it's possible to verify that, for every sample, the peaks correspond to DNA fragments within the intended size range, and they also show quite high intensity, indicating that the amount of DNA obtained from PCR amplification

should be enough for sequencing. The peak on the left side of the graph indicating small-sized fragments (approximately 35–43 bp) may point to the presence of non-ligated adapter sequences that haven't been properly removed from the samples during library preparation.

3.6.2. Analysis of the MED1-enriched regions

The DNA fragments obtained from MED1 ChIP and subsequently selected for library preparation were sequenced to allow for characterisation of the DNA regions associated with MED1, through binding with the Mediator-bound activator protein, and possibly identifying some of them as novel enhancer sites upregulated by feeding and responsive to insulin. After sequencing, all reads were aligned to the mm9 genome using STAR. Unfortunately, sequencing provided a relatively low number of uniquely aligned peaks, as less than 5 million reads per replicate were attained (as opposed to the normal minimum value of 10 to 15 million reads). This low sequencing yield may be due to the presence of seq-adapters in the libraries (that may not have been removed entirely during library preparation). Therefore, to perform preliminary data analysis, it was decided to pool sequencing data from the two replicates sequenced. To identify MED1 enriched regions of the genome, all sequence data from all conditions were pooled to increase the number of sequencing reads used for peak calling. This combined data provided approximately 15 million reads in total. Enriched regions (peaks) were subsequently identified using HOMER[54].

The distribution of the identified fragments is shown as an MA plot in figure 20. MED1-enriched regions regulated by feeding are highlighted, with upregulation shown in green, on the top part of the plot, and down regulation in red on the bottom.

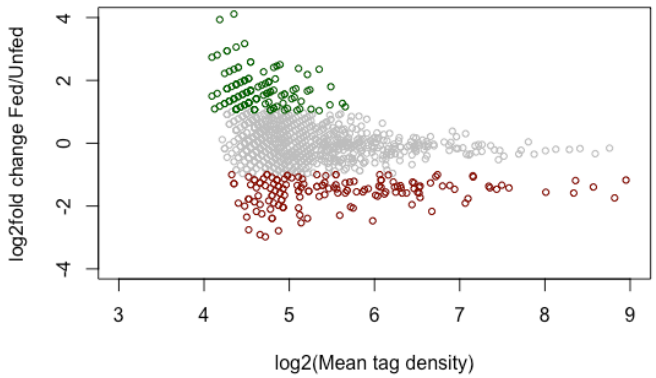


Figure 20. *Distribution of all regions analysed by ChIP-sequencing (ChIP-seq), with feeding-regulated genes highlighted (fold change > 2). Feeding upregulated genes are shown in green at the top of the plot, while feeding down-regulated genes are shown in red at the bottom of the plot.*

In total, 1156 MED1 peaks were identified. Of those, 134 showed significantly higher MED1 occupancy in the fed state than in the fasted state (2-fold enrichment), indicative of upregulation by feeding (11.6%), while 158 had less MED1 binding on the fed state, compared to the unfed, showing downregulation by feeding (13.7%). Two-fold enrichment was considered significant, since the low number of reads didn't allow for adequate statistical treatment. Some of the regions differentially enriched in MED1 in response to feeding show large fold changes between fed and fasted states. The R-script used to analyse ChIP-seq data is shown in Appendix E.

Examples of MED1 enriched regions are shown in figure 21, as UCSC Genome Browser tracks (regions near Gck and Srebp-1c genes are shown). MED1 occupancy along the DNA fragments recovered after ChIP (from my previous MED1 ChIP Fed *vs.* Unfed experiment), as well as H3K27 acetylation and DNase accessible regions (DNase hypersensitive sites, DHS) profiles from the same genome locations, obtained from previous experiments from the Grøntved group, are presented on the tracks below.

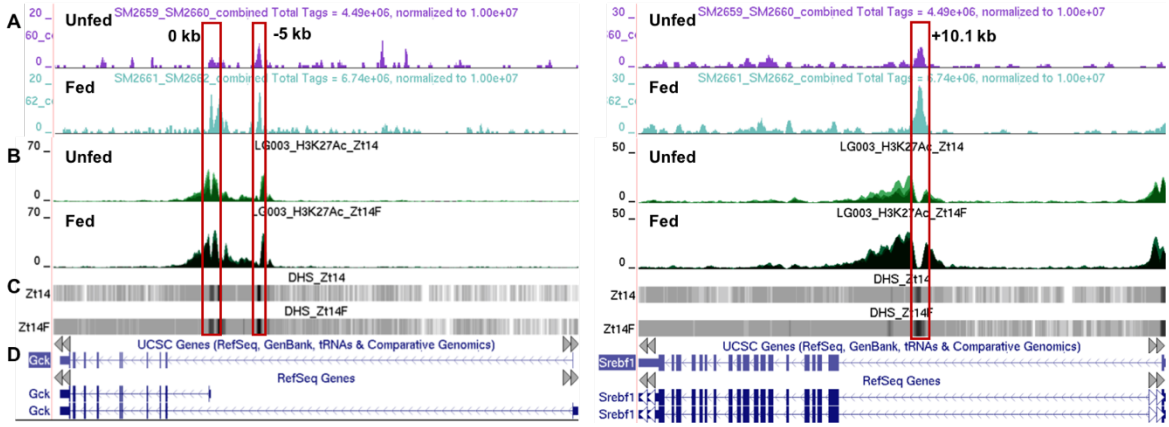


Figure 21. *UCSC genome browser tracks showing MED1 occupancy (tracks), as well as previously observed H3K27 acetylation (tracks **D**) and DNase I hypersensitive sites (DHS, tracks **E**) across the two previously probed feeding-upregulated genes (Gck, on the left, and SREBP-1c, on the right) on unfed and fed states (as annotated on each track). **A** shows combined reads from both samples of each experimental group; **D** shows the location and organization of the gene.*

As seen in these tracks, the MED1 occupancy pattern correlates well with the previously identified feeding upregulated genes, namely Gck and SREBP-1c (the location of the regions probed by ChIP-qPCR is marked on figure 21). These two regions show a clear enrichment in MED1 occupancy, even considering the low read count. This is in accordance with the gene expression patterns expected from the RNA-seq results and also with the tendencies observed by qPCR on the Fed *vs.* Unfed ChIP experiment, that pointed towards enrichment of MED1 in the fed state. It is also visible that the gene regions showing highest feeding

upregulation in MED1 recruitment coincide with the regions previously chosen for ChIP-qPCR probing, for which the primer sets were designed (see section 3.5.1.). The PIK3R1 gene, however, did not show any clear differences in MED1 occupancy between the two experimental groups (tracks not shown here). This is quite likely related to the fact that the PIK3R1 region selected by the designed primer set showed an abnormal H3K27 acetylation pattern on the fed state (see details for the designed primer in section 3.5.1.). The H3K27ac (tracks **D**) show a similar pattern to MED1 occupancy, with the more acetylated regions flanking the sites of highest MED1 recruitment, which therefore overlap with the acetylation valleys, as explained above due to the mechanism of chromatin remodelling and activation. The DNase I hypersensitive sites (DHS), regions of open chromatin that are especially susceptible to enzymatic cleavage and associated with active transcription, also concentrated on the regions of highest MED1 recruitment.

Motif analysis was also performed on the differentially regulated MED1 regions. Because of the low number of reads, the sequenced fragments were only probed for previously identified motif sequences, curated by the HOMER database, as opposed to a *de novo* analysis, where a motif would be derived from the most frequent DNA sequences present in the fragments. Tables 6 and 7 shows the top five most enriched sequence motifs identified in feeding up- and downregulated MED1-bound fragments, respectively.

Table 6. *Top five identified DNA motifs that show upregulation by feeding. These motifs were recognised by comparing DNA sequences from the fragments isolated by MED1 ChIP with previously identified motifs.*

Motif	Name	p-value	% target sequences with motif	% background sequences with motif
	RXR(NR), DR1	10^{-9}	35.16	13.59
	HNF4a(NR), DR1	10^{-8}	25.00	8.33
	FXR(NR), IR1	10^{-7}	17.97	4.88
	PPARE(NR), DR1	10^{-7}	29.69	11.81
	FOXA1(Forkhead)	10^{-6}	19.53	6.20

Feeding-upregulated regions show a clear pattern of motifs showing similar sequences. These motifs are usually bound by nuclear factors, namely the factors/activators RXR (retinoid X receptor), HNF4 (hepatocyte nuclear factor 4), FXR (farnesoid X receptor, or

bile acid receptor), PPAR (peroxisome proliferator-activated receptor) and FOXA1 (forkhead box protein A1, or hepatocyte nuclear factor 3). The proteins that bind to the identified motifs are all nuclear receptors, and can either recruit coactivator or corepressor proteins upon binding to their consensus sequences on target genes. Both FXR and PPAR interact and form heterodimers with RXR. The fact that the consensus sequences these proteins recognise and bind to consist of inverted repeats also points to binding of the receptors as dimers. FOXA1 is a transcriptional activator that acts as a pioneer factor to direct chromatin remodelling at its target enhancers[5]. These motifs are in accordance with the ones previously identified by the Grøntved group associated with H3K27ac (unpublished, data not shown here), and therefore do not represent any novel activator-binding sites (enhancers) upregulated by feeding and associated with the Mediator complex. However, the presence of the previously identified HNF4, FXR and PPAR binding sites validates the H3K27ac motifs previously identified, and suggests that these motifs could indeed constitute feeding-regulated enhancer sites. With confirmation of these preliminary results, implying HNF4, FXR and PPAR as activator proteins involved in feeding- and insulin-regulation of gene transcription, targeting of these proteins on new ChIP experiments could allow for identification of new enhancers associated with feeding-regulated genes.

As mentioned above, this analysis was also done for feeding down-regulated fragments, and the identified motifs are listed below, in table 7.

Table 7. *Identified DNA motifs that show downregulation by feeding. These motifs were recognised by comparing DNA sequences from the fragments isolated by MED1 ChIP with previously identified motifs.*

Motif	Name	p-value	% target sequences with motif	% background sequences with motif
	GRE(NR), IR3	10^{-14}	20.29	1.01
	GRE(NR), IR3	10^{-10}	15.49	1.09
	ARE(NR)	10^{-7}	17.39	2.33
	PGR(NR)	10^{-4}	14.49	2.72
	PR(NR)	10^{-2}	27.54	16.01

Once more, and even more expressively, it is possible to observe that the identified DNA motifs are quite similar to each other and are recognised and bound by nuclear receptors:

GRE (glucocorticoid response elements), which is bound by the glucocorticoid receptor, ARE (AU-rich elements) and PGR/PR (progesterone receptor). As suggested by the sequence on the motifs, some of these activators bind to the consensus sequences in a dimerized form, similarly to what was observed for the feeding-upregulated motifs. The observed motifs suggest regulation of feeding down-regulated gene enhancers by the glucocorticoid receptor, which agrees well with previously obtained data from the Grøntved group, providing preliminary validation for those results.

3.7. Analysis of MED1 ChIP data (Fed vs. Fed+S961 experiment)

A preliminary setup was defined, to check whether insulin signalling influenced MED1 recruitment to the genes identified to be feeding and insulin-regulated by RNA-seq and qPCR. For this purpose, I performed an identical experiment to the one previously described, but with the samples from both fed groups (injected with either PBS or S961). Results obtained are shown in figure 22.

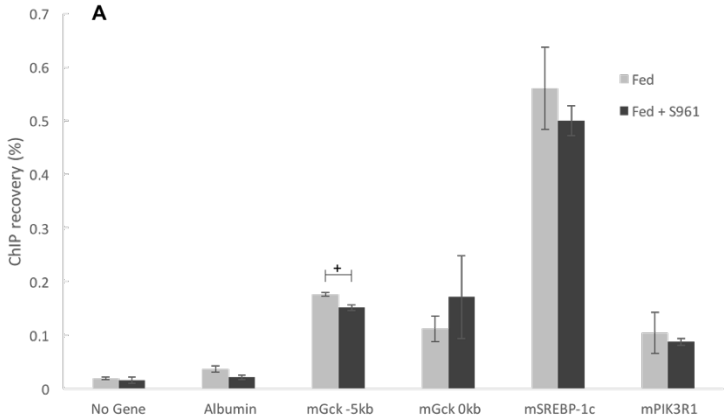


Figure 22. *Effect of insulin antagonist S961 on the MED1 enrichment in four sites of three feeding-upregulated genes – Gck, SREBP-1c and phosphatidylinositol 3-kinase regulatory subunit alpha (PIK3R1) in the liver, evaluated by ChIP-qPCR. Results are shown as group average ± standard deviation. No Gene was the reference site and albumin (no DHS) the negative control.* + $p < 0.05$ *vs.* Fed+PBS.

Figure 22 shows a significant difference between the two groups only on the upstream enhancer site of Gck, but not in any of the other sites. It is possible to see that background is not affecting the results since the negative control region (identified as No Gene, or NG, a region not used as an enhancer in liver tissue, and thus showing no H3K27Ac) shows a DNA recovery rate that is quite low and not comparable to the recoveries obtained with the designed primers. Regarding the enhancer of Gck located 5 kb upstream of the gene

promoter, one can see that there is enrichment in MED1 on the fed state, where insulin signalling is active, compared to the Fed+S961 group, where insulin signalling is blocked, mimicking insulin resistance. It should be noted that, even though this difference in MED1 recruitment is significant, the effect is very small. We hypothesize that Mediator recruitment may be regulated by insulin signalling and could therefore be a useful probe for identifying transcriptional enhancers subjected to insulin regulation, in a subsequent ChIP-seq experiment on these samples.

3.8. Expression analysis of proteins involved in the insulin signalling pathway

Expression and activation analysis of proteins involved in the insulin signalling pathway was also carried out in liver protein extracts to evaluate the activation of the insulin signalling pathway by feeding and the effectiveness of insulin receptor blocking by S961. Results are shown in figure 23.

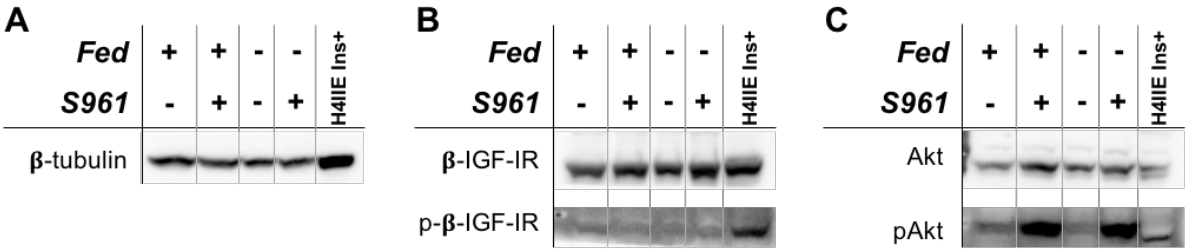


Figure 23. *Effect of feeding and/or injection with insulin antagonist S961 on the expression and activation (shown as phosphorylation) of insulin receptor subunit β (**B**) and Akt (**C**), evaluated by Western blotting. Expression of tubulin (**A**) was used as a loading control. The last lane on each blot corresponds to a positive control of the insulin signaling pathway activation by treatment of H4IIE cells with insulin.*

The blots reproduced in figure 23 show that loading of all samples on the electrophoresis gels was uniform, since the amount of β-tubulin, which should depend only on the amount of total protein, seems to be approximately the same on all lanes (figure 23-A), independently of the applied treatment. Similarly, in figures 23-B and 23-C, one can observe that amounts of insulin receptor (IR) and Akt, respectively, are approximately the same in all experimental groups, as expected. Therefore, differences in the phosphorylated forms of these proteins result only from different activation states due to insulin signalling, and are not due to differential expression of the native proteins. On the same figures, IR seems to be more phosphorylated on the fed state (Fed+PBS) than on either group of the unfed state, which results from the increase in insulin levels in the blood after feeding, leading to binding to the IR and its consequent activation through phosphorylation. However, there's no clear

effect of S961 on the phosphorylation of this protein. Moreover, if one looks at the results for Ser473-phosphorylated Akt (pAkt), it appears quite unexpected. Even though Akt is more phosphorylated in the fed state (Fed+PBS *vs.* Unfed+PBS), which is normal, it is also evident that S961 is strongly promoting Akt phosphorylation, in both the fed and fasted groups. This observation conflicts with the systemic and transcriptional effects of S961 observed in the previous experiments as insulin signalling should be blocked in response to S961 binding to the IR. It has previously been reported that S961 could show a partial agonist behaviour at low concentrations *in vitro* (1–10 nM), by weakly inducing IR and Akt phosphorylation[48], but this should not be an issue as the amount of S961 injected should be high enough to assure full antagonism of the IR (40 nM/kg, intraperitoneally). In fact, an *in vivo* study in rats showed that a concentration of 10 or 30 nM/kg was enough to elicit the systemic effects of insulin resistance (hyperglycaemia, hyperinsulinemia and depletion of hepatic glycogen)[52]. On the other hand, a more recent study focusing on the development of a monoclonal antibody alternative to insulin resistance models has now suggested that in the presence of increased concentrations of insulin (at 10 and 100 nM), the ability of S961 to block phosphorylation of IR and Akt becomes compromised, and levels of both pIR and pAkt increase above basal levels[47]. As seen in figure 11-A, mice injected with S961 became hyperinsulinemic, with insulin concentrations reaching 11.5 to 18.4 nM, which surpass the 10 nM shown to counteract the effects of S961 on Akt phosphorylation. The competitive antagonism model could explain that the higher levels of insulin in the blood in response to S961 administration would lead to dislocation of S961 and to higher activation of the IR. Nevertheless, this should also affect blood glucose and the mRNA expression patterns, but these still support the occurrence of insulin resistance induced by S961.

Thus, the mechanism of action of S961, which leads to blockage of insulin signalling, needs to be clarified. With that purpose, I performed further Western blots on the phosphorylation of two signalling molecules downstream of pAkt (FOXO1A and glycogen synthase kinase, or GSK) to check if their activation was affected in the same way or if it showed the expected response to the presence of S961. The results for GSK3β are shown below, in figure 24.

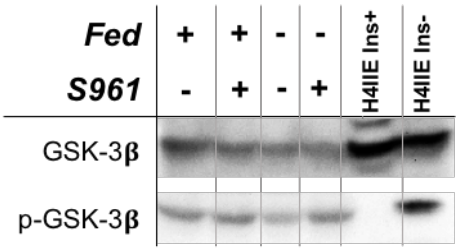


Figure 24. *Effect of feeding and/or injection with insulin antagonist S961 on the expression and activation (shown as phosphorylation) of glycogen synthase kinase 3 isoform β (GSK-3β), evaluated by Western blotting. The last two lanes on the blot correspond, in order, to a positive and a negative control of the insulin signalling pathway activation by treatment of H4IIE cells with insulin or from input samples of fasted animals, respectively.*

Figure 24 shows protein expression levels of native GSK and its activation state reflected by phosphorylation levels. Native and phosphorylated FOXO1A were also assessed but results are not shown here because no bands were detected on the blotted membrane. Once more, loading of the wells on the electrophoresis was uniform for all samples, as native GSK expression exhibits the same pattern across all samples. Looking at the control lanes for the phosphorylated GSK assay, the phosphorylation pattern is not according to expected. GSK is an enzyme that inactivates glycogen synthase by promoting its phosphorylation. Therefore, since glycogen synthesis is up-regulated in the liver by feeding, when insulin signalling is most active, glycogen synthase (GS) is expected to be active after feeding, and GSK should be inactive at this time, and active during fasting, inhibiting GS activity when available blood glucose is low. GSK is also inactivated by phosphorylation promoted by upstream signalling molecules and thus the expected pattern (GSK phosphorylation induced by insulin) does not correspond to the results seen on the controls where insulin signalling was promoted (positive control, no phosphorylated GSK) and repressed (negative control, where high GSK phosphorylation is observed). A troubleshooting experiment was performed on the same membrane, by reprobing with pAkt antibody, but no conclusive result was obtained, and therefore there are two possible explanations for this observation: although unlikely, there may have been a switch of controls during loading, or the unexpected phosphorylation pattern could result from an unknown mechanism or response to insulin by the cultured cells that may differ from cell behaviour *in vivo*. However, the results for the samples under analysis appear to match the ones observed for pAkt: GSK is more phosphorylated in the fed than in the fasted state, as would be expected, but S961 also seems to promote phosphorylation of this enzyme, instead of repressing it.

Therefore, this additional experiment did not allow to clarify the mechanism by which S961 seems to agonize phosphorylation patterns of the insulin signalling pathway, but still antagonize the effects of insulin on metabolism. A possible mechanism could involve different isoforms or subcellular localizations of one or more messenger molecules eliciting different cellular effects, as could be the case of Akt isoforms[41]. These results, in combination with the low percentage of S961-sensitive feeding regulated genes observed on the RNA-seq data, lead to questioning on whether S961 is the adequate model for full IR antagonism and mimicking of insulin resistance and therefore further research on the topic is needed.

4. Conclusions

Blood glucose, serum insulin and mRNA expression levels of feeding upregulated genes support S961 (an antagonist of the insulin receptor) as an inducer of insulin resistance. However, Akt phosphorylation conflicted with systemic and transcriptional effects of S961 as it stimulated phosphorylation of Akt when insulin signalling should be blocked. Previous studies have shown that both low concentrations of S961 and hyperinsulinemia may lead to partial agonism of the insulin receptor by S961, due to the competitive binding mechanism. Western blots on phosphorylation of downstream signalling molecules (GSK and FOXO1) were either inconclusive or showed the same phosphorylation pattern as Akt and did not help in clarifying the mechanism of action of S961. Another possibility for explaining the partial agonism of S961 at the phosphorylation patterns of insulin-activated second messengers, contrasting with its antagonist effects on the metabolic genes, may relate to the different isoforms of Akt showing distinct tissue distribution and subcellular localization that can affect its function. However, further research would be necessary to elucidate this odd behaviour exhibited in response to S961. Future work could then include additional Western blotting analyses, complemented by fluorescence microscopy studies of tagged native and phosphorylated messengers. This might provide clearer mechanisms to the transduction of insulin signalling and to the action of S961 as an antagonist of the insulin receptor.

MED1 ChIP-qPCR results showed a potential feeding and insulin regulation of recruitment of the Mediator complex (a crucial transcriptional coactivator) to enhancers. Sequencing of the DNA samples obtained from one of those ChIP experiments (ChIP-seq) allowed to preliminarily confirm feeding upregulation of Mediator occupancy near some canonical feeding upregulated genes. The identification of DNA motifs within these regions led to no new information regarding bound activators, but allowed to validate previously identified enhancer motifs for genes upregulated by feeding. However, the number of sequenced reads obtained was quite low and thus optimization as well as analysis of a larger number of samples (or recovery of a higher amount of DNA) would be important to validate and improve the quality of these results. Moreover, due to time limitations, it was not possible to undertake further ChIP experiments investigating the regulation of these genes by insulin/S961. It would be very interesting to proceed with these analyses and to perform supplementary ChIP experiments probing for different transcription factors or activators.

5. Bibliography

1. WHO. "Global report on diabetes". Geneva, 2016. http://www.who.int/diabetes/global-report/en/. (accessed on 10 April 2017)

2. Biddinger SB, Kahn CR. From mice to men: Insights into the insulin resistance syndromes. Annual Review of Physiology. Annual Review of Physiology. 68. Palo Alto: Annual Reviews; 2006. p. 123-58.

3. Boeger H, Bushnell DA, Davis R, Griesenbeck J, Lorch Y, Strattan JS, Westover KD, Kornberg RD, 2005. "Structural basis of eukaryotic gene transcription". *FEBS Lett*. **579**(4): 899-903.

4. Ost A, Pospisilik JA, 2015. "Epigenetic modulation of metabolic decisions". *Current Opinion in Cell Biology*. **33**: 88-94.

5. Calo E, Wysocka J, 2013. "Modification of Enhancer Chromatin: What, How, and Why?". *Molecular Cell*. **49**(5): 825-37.

6. Le Martelot G, Canella D, Symul L, Migliavacca E, Gilardi F, Liechti R, Martin O, Harshman K, Delorenzi M, Desvergne B, Herr W, Deplancke B, Schibler U, Rougemont J, Guex N, Hernandez N, Naef F, Cycli XC, 2012. "Genome-Wide RNA Polymerase II Profiles and RNA Accumulation Reveal Kinetics of Transcription and Associated Epigenetic Changes During Diurnal Cycles". *Plos Biology*. **10**(11): 16.

7. Fukaya T, Lim B, Levine M, 2016. "Enhancer Control of Transcriptional Bursting". *Cell*. **166**(2): 358-68.

8. Kornberg RD, 2007. "The molecular basis of eukaryotic transcription". *Proceedings of the National Academy of Sciences*. **104**(32): 12955-61.

9. Nonet M, Sweetser D, Young RA, 1987. "Functional redundancy and structural polymorphism in the large subunit of rna polymerase-ii". *Cell*. **50**(6): 909-15.

10. Corden JL, 1990. "Tails of RNA polymerase-II". *Trends BiochemSci*. **15**(10): 383-7.

11. Carlson M, 1997. "Genetics of transcriptional regulation in yeast: Connections to the RNA polymerase II CTD". *Annu Rev Cell Dev Biol*. **13**: 1-23.

12. Kim YJ, Bjorklund S, Li Y, Sayre MH, Kornberg RD, 1994. "A Multiprotein Mediator Of Transcriptional Activation And Its Interaction With The C-Terminal Repeat Domain Of Rna-Polymerase-II". *Cell*. **77**(4): 599-608.

13. Koike N, Yoo SH, Huang HC, Kumar V, Lee C, Kim TK, Takahashi JS, 2012. "Transcriptional Architecture and Chromatin Landscape of the Core Circadian Clock in Mammals". *Science*. **338**(6105): 349-54.

14. Petrenko N, Jin Y, Wong Koon H, Struhl K, 2016. "Mediator Undergoes a Compositional Change during Transcriptional Activation". *Molecular Cell*. **64**(3): 443-54.

15. Kornberg RD, 2005. "Mediator and the mechanism of transcriptional activation". *Trends BiochemSci*. **30**(5): 235-9.

16. Guglielmi B, van Berkum NL, Klapholz B, Bijma T, Boube M, Boschiero C, Bourbon HM, Holstege FCP, Werner M, 2004. "A high resolution protein interaction map of the yeast Mediator complex". *Nucleic Acids Res*. **32**(18): 5379-91.

17. Biddick R, Young ET, 2005. "Yeast Mediator and its role in transcriptional regulation". *C R Biol*. **328**(9): 773-82.

18. Casamassimi A, Napoli C, 2007. "Mediator complexes and eukaryotic transcription regulation: An overview". *Biochimie*. **89**(12): 1439-46.

19. Balciunas D, Gälman C, Ronne H, Björklund S, 1999. "The Med1 subunit of the yeast mediator complex is involved in both transcriptional activation and repression". *Proceedings of the National Academy of Sciences*. **96**(2): 376-81.

20. Jeronimo C, Langelier M-F, Bataille Alain R, Pascal John M, Pugh BF, Robert F, 2016. "Tail and Kinase Modules Differently Regulate Core Mediator Recruitment and Function In Vivo". *Molecular Cell*. **64**(3): 455-66.

21. Lorch Y, Beve J, Gustafsson CM, Myers LC, Kornberg RD, 2000. "Mediator-nucleosome interaction". *Molecular Cell*. **6**(1): 197-201.

22. Belakavadi M, Pandey PK, Vijayvargia R, Fondell JD, 2008. "MED1 phosphorylation promotes its association with Mediator: Implications for nuclear receptor signaling". *Mol Cell Biol*. **28**(12): 3932-42.

23. Pandey PK, Udayakumar TS, Lin XJ, Sharma D, Shapiro PS, Fondell JD, 2005. "Activation of TRAP/Mediator subunit TRAP220/Med1 is regulated by mitogen-activated protein kinase-dependent phosphorylation". *Mol Cell Biol*. **25**(24): 10695-710.

24. Pott S, Lieb JD, 2015. "What are super-enhancers?". *Nature Genet*. **47**(1): 8-12.

25. Aguilar-Arnal L, Sassone-Corsi P, 2013. "The circadian epigenome: how metabolism talks to chromatin remodeling". *Current Opinion in Cell Biology*. **25**(2): 170-6.

26. Feng D, Lazar MA, 2012. "Clocks, Metabolism, and the Epigenome". *Molecular Cell*. **47**(2): 158-67.

27. Sunderram J, Sofou S, Kamisoglu K, Karantza V, Androulakis IP, 2014. "Time-restricted feeding and the realignment of biological rhythms: translational opportunities and challenges". *Journal of Translational Medicine*. **12**: 9.

28. Marti AR, Meerlo P, Gronli J, van Hasselt SJ, Mrdalj J, Pallesen S, Pedersen TT, Henriksen TEG, Skrede S, 2016. "Shift in Food Intake and Changes in Metabolic Regulation and Gene Expression during Simulated Night-Shift Work: A Rat Model". *Nutrients*. **8**(11): 13.

29. Vollmers C, Gill S, DiTacchio L, Pulivarthy SR, Le HD, Panda S, 2009. "Time of feeding and the intrinsic circadian clock drive rhythms in hepatic gene expression". *Proceedings of the National Academy of Sciences of the United States of America*. **106**(50): 21453-8.

30. Eckel-Mahan KL, Patel VR, de Mateo S, Orozco-Solis R, Ceglia NJ, Sahar S, Dilag-Penilla SA, Dyar KA, Baldi P, Sassone-Corsi P, 2013. "Reprogramming of the Circadian Clock by Nutritional Challenge". *Cell*. **155**(7): 1464-78.

31. Hatori M, Vollmers C, Zarrinpar A, DiTacchio L, Bushong EA, Gill S, Leblanc M, Chaix A, Joens M, Fitzpatrick JAJ, Ellisman MH, Panda S, 2012. "Time-Restricted Feeding without Reducing Caloric Intake Prevents Metabolic Diseases in Mice Fed a High-Fat Diet". *Cell Metabolism*. **15**(6): 848-60.

32. Geisler CE, Hepler C, Higgins MR, Renquist BJ, 2016. "Hepatic adaptations to maintain metabolic homeostasis in response to fasting and refeeding in mice". *Nutr Metab*. **13**: 13.

33. Czech MP, 1977. "Molecular-basis of insulin action". *Annual Review of Biochemistry*. **46**: 359-84.

34. la Fleur SE, Kalsbeek A, Wortel J, Buijs RM, 1999. "A suprachiasmatic nucleus generated rhythm in basal glucose concentrations". *J Neuroendocrinol*. **11**(8): 643-52.

35. Pilch PF, Czech MP, 1980. "Subunit structure of the high-affinity insulin-receptor – Evidence for a disulfide-linked receptor complex in fat-cell and liver plasma-membranes". *Journal of Biological Chemistry*. **255**(4): 1722-31.

36. Schaffer L, 1994. "A model for insulin binding to the insulin-receptor". *Eur J Biochem*. **221**(3): 1127-32.

37. De Meyts P, 1994. "The structural basis of insulin and insulin-like growth factor-I receptor binding and negative co-operativity, and its relevance to mitogenic versus metabolic signalling". *Diabetologia*. **37**(2): S135-S48.

38. Kasuga M, Zick Y, Blithe DL, Crettaz M, Kahn CR, 1982. "Insulin stimulates tyrosine phosphorylation of the insulin-receptor in a cell-free system". *Nature*. **298**(5875): 667-9.

39. De Meyts P, Roth J, Neville DM, Gavin JR, Lesniak MA, 1973. "Insulin interactions with its receptors – Experimental evidence for negative cooperativity". *Biochem Biophys Res Commun*. **55**(1): 154-61.

40. Riehle C, Abel ED, 2016. "Insulin Signaling and Heart Failure". *Circulation Research*. **118**(7): 1151-69.

41. Gonzalez E, McGraw TE, 2009. "The Akt kinases Isoform specificity in metabolism and cancer". *Cell Cycle*. **8**(16): 2502-8.

42. Santi SA, Lee H, 2010. "The Akt isoforms are present at distinct subcellular locations". *Am J Physiol-Cell Physiol*. **298**(3): C580-C91.

43. Martin BC, Warram JH, Krolewski AS, Bergman RN, Soeldner JS, Kahn CR, 1992. "Role of glucose and insulin resistance in development of type-2 diabetes-mellitus - results of a 25-year follow-up-study". *Lancet*. **340**(8825): 925-9.

44. Lillioja S, Mott DM, Spraul M, Ferraro R, Foley JE, Ravussin E, Knowler WC, Bennett PH, Bogardus C, 1993. "Insulin-resistance and insulin secretory dysfunction as precursors of non-insulin-dependent diabetes-mellitus - prospective studies of pima-indians". *N Engl J Med*. **329**(27): 1988-92.

45. Chatzigeorgiou A, Halapas A, Kalafatakis K, Kamper E, 2009. "The Use of Animal Models in the Study of Diabetes Mellitus". *In Vivo*. **23**(2): 245-58.

46. Edison AS, Hall RD, Junot C, Karp PD, Kurland IJ, Mistrik R, Reed LK, Saito K, Salek RM, Steinbeck C, Sumner LW, Viant MR, 2016. "The Time Is Right to Focus on Model Organism Metabolomes". *Metabolites*. **6**(1): 7.

47. Cieniewicz AM, Kirchner T, Hinke SA, Nanjunda R, D'Aquino K, Boayke K, Cooper PR, Perkinson R, Chiu ML, Jarantow S, Johnson DL, Whaley JM, Lacy ER, Lingham RB, Liang Y, Kihm AJ, 2017. "Novel Monoclonal Antibody Is an Allosteric Insulin Receptor Antagonist That Induces Insulin Resistance". *Diabetes*. **66**(1): 206-17.

48. Knudsen L, Hansen BF, Jensen P, Pedersen TA, Vestergaard K, Schaffer L, Blagoev B, Oleksiewicz MB, Kiselyov VV, De Meyts P, 2012. "Agonism and Antagonism at the Insulin Receptor". *PLoS One*. **7**(12): 10.

49. Miao J, Haas JT, Manthena P, Wang YN, Zhao EP, Vaitheesvaran B, Kurland IJ, Biddinger SB, 2014. "Hepatic insulin receptor deficiency impairs the SREBP-2 response to feeding and statins". *J Lipid Res*. **55**(4): 659-67.

50. Schaffer L, Brissette RE, Spetzler JC, Pillutla RC, Ostergaard S, Lennick M, Brandt J, Fletcher PW, Danielsen GM, Hsiao KC, Andersen AS, Dedova O, Ribel U, Hoeg-Jensen T, Hansen PH, Blume AJ, Markussen J, Goldstein NI, 2003. "Assembly of high-affinity insulin receptor agonists and antagonists from peptide building blocks". *Proceedings of the National Academy of Sciences of the United States of America*. **100**(8): 4435-9.

51. Schaffer L, Brand CL, Hansen BF, Ribel U, Shaw AC, Slaaby R, Sturis J, 2008. "A novel high-affinity peptide antagonist to the insulin receptor". *Biochem Biophys Res Commun*. **376**(2): 380-3.

52. Vikram A, Jena G, 2010. "S961, an insulin receptor antagonist causes hyperinsulinemia, insulin-resistance and depletion of energy stores in rats". *Biochem Biophys Res Commun*. **398**(2): 260-5.

53. Love MI, Huber W, Anders S, 2014. "Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2". *Genome Biol*. **15**(12): 38.

54. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK, 2010. "Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities". *Molecular Cell*. **38**(4): 576-89.

55. Clow A, Thorn L, Evans P, Hucklebridge F, 2004. "The awakening cortisol response: Methodological issues and significance". *Stress*. **7**(1): 29-37.

Appendix A: R-script for RNA-seq data analysis

C-2 samples RNA-seq data analysis

Catarina M Correia

09/02/2017

Loading data

I want to choose my .txt file with the data I need to analyse.

```
# set working directory
# command is: setwd("folder name")

setwd("~/Documents/Escola/Universidade/Mestrado Bioquímica Clínica – UA/2º Ano,
Tese, Ano Lectivo 2016:2017/RNA-seq/C-2 samples")


# define the file on use and set a name for the data contained
# command is: set_name = read.delim("file name within the defined directory", h
eader = TRUE, dec = ".", etc.)

seqdata = read.delim("SM1981-SM2001_diffExp output.txt", header = TRUE, dec = "
.")


# view your data set
View(seqdata)     # opens a table with all your data
names(seqdata)    # gives a list with the names corresponding to each column nu
mber
```

```
##  [1] "Transcript.RepeatID..cmd.analyzeRepeats.pl.rna.mm9..strand.both..count
.exons..d..storage.Grontved.Homer_TagDir.SM1981...storage.Grontved.Homer_TagDir
.SM1982...storage.Grontved.Homer_TagDir.SM1983...storage.Grontved.Homer_TagDir.
SM1987...storage.Grontved.Homer_TagDir.SM1988...storage.Grontved.Homer_TagDir.S
M1989...storage.Grontved.Homer_TagDir.SM1993...storage.Grontved.Homer_TagDir.SM
1994...storage.Grontved.Homer_TagDir.SM1995...storage.Grontved.Homer_TagDir.SM1
999...storage.Grontved.Homer_TagDir.SM2000...storage.Grontved.Homer_TagDir.SM20
01...noadj...cmd.getDiffExpression.pl..data.home.correia.SM1981.SM2001_raw.txt.
ZT14F.ZT14F.ZT14F.ZT14FS961.ZT14FS961.ZT14FS961.ZT14unfed.ZT14unfed.ZT14unfed.Z
T14unfedS961.ZT14unfedS961.ZT14unfedS961..DESeq2."
##  [2] "chr"
##  [3] "start"
##  [4] "end"
##  [5] "strand"
##  [6] "Length"
##  [7] "Copies"
##  [8] "Annotation.Divergence"
##  [9] "X.storage.Grontved.Homer_TagDir.SM1981..reads"
## [10] "X.storage.Grontved.Homer_TagDir.SM1982..reads"
## [11] "X.storage.Grontved.Homer_TagDir.SM1983..reads"
## [12] "X.storage.Grontved.Homer_TagDir.SM1987..reads"
## [13] "X.storage.Grontved.Homer_TagDir.SM1988..reads"
## [14] "X.storage.Grontved.Homer_TagDir.SM1989..reads"
## [15] "X.storage.Grontved.Homer_TagDir.SM1993..reads"
## [16] "X.storage.Grontved.Homer_TagDir.SM1994..reads"
## [17] "X.storage.Grontved.Homer_TagDir.SM1995..reads"
```

```
## [18] "X.storage.Grontved.Homer_TagDir.SM1999..reads"
## [19] "X.storage.Grontved.Homer_TagDir.SM2000..reads"
## [20] "X.storage.Grontved.Homer_TagDir.SM2001..reads"
## [21] "ZT14F.vs..ZT14FS961.baseMean"
## [22] "ZT14F.vs..ZT14FS961.log2FoldChange"
## [23] "ZT14F.vs..ZT14FS961.lfcSE"
## [24] "ZT14F.vs..ZT14FS961.stat"
## [25] "ZT14F.vs..ZT14FS961.pvalue"
## [26] "ZT14F.vs..ZT14FS961.padj"
## [27] "ZT14F.vs..ZT14unfed.baseMean"
## [28] "ZT14F.vs..ZT14unfed.log2FoldChange"
## [29] "ZT14F.vs..ZT14unfed.lfcSE"
## [30] "ZT14F.vs..ZT14unfed.stat"
## [31] "ZT14F.vs..ZT14unfed.pvalue"
## [32] "ZT14F.vs..ZT14unfed.padj"
## [33] "ZT14F.vs..ZT14unfedS961.baseMean"
## [34] "ZT14F.vs..ZT14unfedS961.log2FoldChange"
## [35] "ZT14F.vs..ZT14unfedS961.lfcSE"
## [36] "ZT14F.vs..ZT14unfedS961.stat"
## [37] "ZT14F.vs..ZT14unfedS961.pvalue"
## [38] "ZT14F.vs..ZT14unfedS961.padj"
## [39] "ZT14FS961.vs..ZT14unfed.baseMean"
## [40] "ZT14FS961.vs..ZT14unfed.log2FoldChange"
## [41] "ZT14FS961.vs..ZT14unfed.lfcSE"
## [42] "ZT14FS961.vs..ZT14unfed.stat"
## [43] "ZT14FS961.vs..ZT14unfed.pvalue"
## [44] "ZT14FS961.vs..ZT14unfed.padj"
## [45] "ZT14FS961.vs..ZT14unfedS961.baseMean"
## [46] "ZT14FS961.vs..ZT14unfedS961.log2FoldChange"
## [47] "ZT14FS961.vs..ZT14unfedS961.lfcSE"
## [48] "ZT14FS961.vs..ZT14unfedS961.stat"
## [49] "ZT14FS961.vs..ZT14unfedS961.pvalue"
## [50] "ZT14FS961.vs..ZT14unfedS961.padj"
## [51] "ZT14unfed.vs..ZT14unfedS961.baseMean"
## [52] "ZT14unfed.vs..ZT14unfedS961.log2FoldChange"
## [53] "ZT14unfed.vs..ZT14unfedS961.lfcSE"
## [54] "ZT14unfed.vs..ZT14unfedS961.stat"
## [55] "ZT14unfed.vs..ZT14unfedS961.pvalue"
## [56] "ZT14unfed.vs..ZT14unfedS961.padj"
```

Filtering data (feeding-regulated sites)

Now I want to create subsets of data containing only the genome sites that are regulated by feeding, which consists of the genes which are significantly differently expressed (FDR <0.05) between the Fed and Unfed groups (ZT14F vs. ZT14unfed – padj).

```
# create the subsets you need
diff_exp = seqdata[ seqdata$ZT14F.vs..ZT14unfed.padj < 0.05, ]
diff_up = diff_exp[ diff_exp$ZT14F.vs..ZT14unfed.log2FoldChange < 0, ]
diff_down = diff_exp[ diff_exp$ZT14F.vs..ZT14unfed.log2FoldChange > 0, ]

# next three subsets are inserted in the feeding upregulated genes

ins_reg = diff_up[ diff_up$ZT14F.vs..ZT14FS961.padj < 0.05, ]

ins_up = diff_up[ diff_up$ZT14F.vs..ZT14FS961.padj < 0.05 & diff_up$ZT14F.vs..Z
T14FS961.log2FoldChange < 0, ]
```

```
ins_down = diff_up[ diff_up$ZT14F.vs..ZT14FS961.padj < 0.05 & diff_up$ZT14F.vs.
.ZT14FS961.log2FoldChange > 0, ]
```

Plotting feeding-regulated data, MA plot

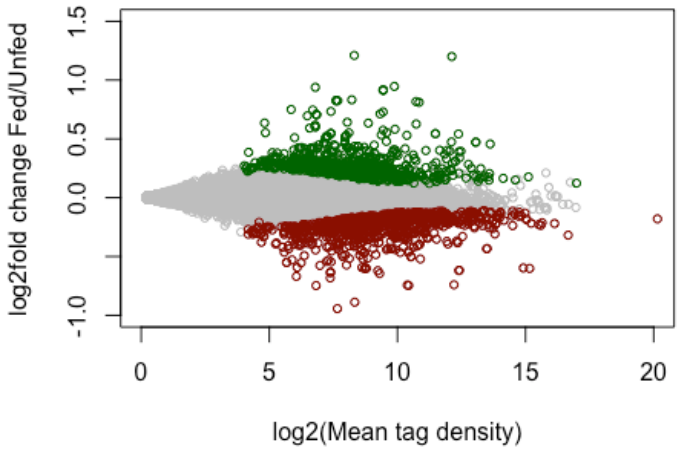
Now, it's time to plot all sequencing data, and highlighting the genes that are actually differentially expressed as a consequence of the feeding/fasting entrainment, in order to show the distribution of down- and upregulated genes by feeding. I want downregulated genes to be shown at the bottom of the plot and in red and upregulated genes to be shown at the top of the plot and in green.

```
# make an MA plot with the unfiltered data
plot(log2(seqdata$ZT14F.vs..ZT14unfed.baseMean + 1), -seqdata$ZT14F.vs..ZT14unf
ed.log2FoldChange, xlab="log2(Mean tag density)", ylab="log2fold change Fed/Unf
ed", pch=1, cex=0.7, col="gray", ylim=c(-1,1.5), xlim=c(0,20))


# highlight the feeding-regulated sites

points(log2(diff_up$ZT14F.vs..ZT14unfed.baseMean + 1), -diff_up$ZT14F.vs..ZT14u
nfed.log2FoldChange, xlab="log2(Mean tag density)", ylab="log2fold change Fed/U
nfed", pch=1, cex=0.7, col="darkgreen", ylim=c(-1,1.5), xlim=c(0,20))

points(log2(diff_down$ZT14F.vs..ZT14unfed.baseMean + 1), -diff_down$ZT14F.vs..Z
T14unfed.log2FoldChange, xlab="log2(Mean tag density)", ylab="log2fold change F
ed/Unfed", pch=1, cex=0.7, col="darkred", ylim=c(-1,1.5), xlim=c(0,20))
```



Plotting insulin-regulated data, MA plot

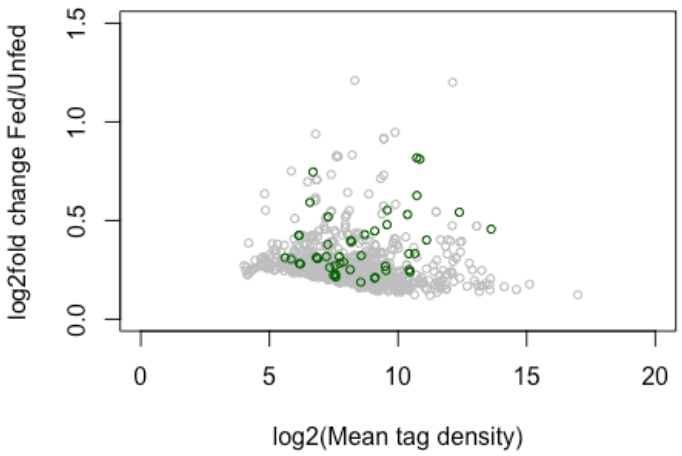
Since my main focus is the set of genes upregulated by feeding, now I want to plot only those and highlight the ones that are also differentially expressed as a consequence of insulin signaling (regulated by S961, an insulin antagonist). With that in mind, I decided to have two different kinds of plots: one that "zooms in" the feeding-upregulated region of the main plot, using the same axes, and highlights the genes regulated by insulin, in green; and another that shows the same data, but with the differential expression being plotted in terms of S961 effect, allowing for distinction between up- and down-regulation by S961/insulin.

```
# make an MA plot with only the feeding up-regulated data
plot(log2(diff_up$ZT14F.vs..ZT14unfed.baseMean + 1), -diff_up$ZT14F.vs..ZT14unf
ed.log2FoldChange, xlab="log2(Mean tag density)", ylab="log2fold change Fed/Unf
```

```
ed", pch=1, cex=0.7, col="gray", ylim=c(0,1.5), xlim=c(0,20))


# highlight the insulin-regulated sites

points(log2(ins_reg$ZT14F.vs..ZT14unfed.baseMean + 1), -ins_reg$ZT14F.vs..ZT14u
nfed.log2FoldChange, xlab="log2(Mean tag density)", ylab="log2fold change Fed/U
nfed", pch=1, cex=0.7, col="darkgreen", ylim=c(0,1.5), xlim=c(0,20))
```

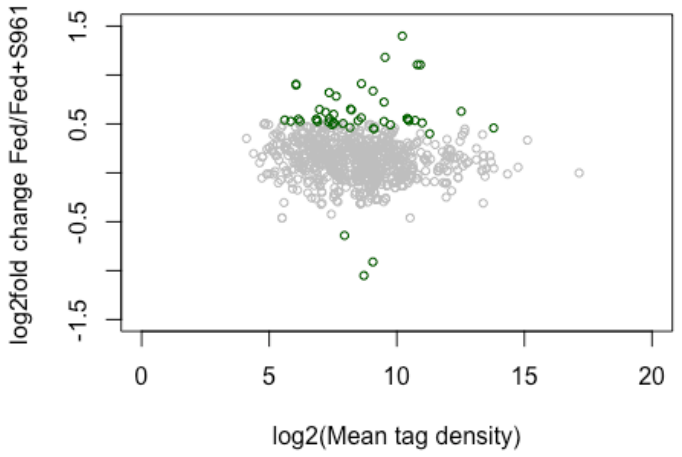



```
## ANOTHER VIEWING OPTION (to distinguish up- and down-regulation by S961)

# make an MA plot with only the feeding up-regulated data
plot(log2(diff_up$ZT14F.vs..ZT14FS961.baseMean + 1), -diff_up$ZT14F.vs..ZT14FS9
61.log2FoldChange, xlab="log2(Mean tag density)", ylab="log2fold change Fed/Fed
+S961", pch=1, cex=0.7, col="gray", ylim=c(-1.5,1.5), xlim=c(0,20))

# highlight the insulin-regulated sites

points(log2(ins_reg$ZT14F.vs..ZT14FS961.baseMean + 1), -ins_reg$ZT14F.vs..ZT14F
S961.log2FoldChange, xlab="log2(Mean tag density)", ylab="log2fold change Fed/F
ed+S961", pch=1, cex=0.7, col="darkgreen", ylim=c(-1.5,1.5), xlim=c(0,20))
```




```
# identify the 3 genes that are upregulated by S961 ("downregulated" by insulin
)

View(na.omit(ins_up[,7:8]))  # ins_up[,8] is the same as ins_up$Annotation.Dive
```

```
rgence
View(na.omit(ins_down[,7:8]))
```

Note: Result of the list of genes that are insulin-down-regulated but feeding-up-regulated

- Klf10|AI115143|EGR[a]|Egral|Gdnfif|TIEG-1|Tieg|Tieg1|mGIF|-|15 B3.1|15|protein-coding
- Slc25a30|4933433D23Rik|AV025504|AV221431|AW319655|KMCP1|-|14|14 D2|protein-coding
- E2f8|4432406C08Rik|AA410048|-|7 B4|7|protein-coding

Extra Plots on Insulin Regulation

I want to illustrate the insulin regulated genes now, independently of up- or down-regulation by feeding and also check whether any insulin-regulated genes are independent from feeding (which could be due to activation of alternative signalling pathways).

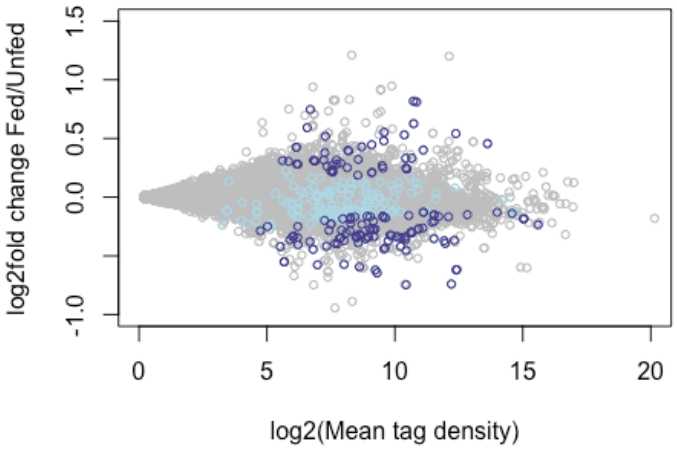
```
# make an MA plot showing all genes that are insulin-regulated on the main plot

plot(log2(seqdata$ZT14F.vs..ZT14unfed.baseMean + 1), -seqdata$ZT14F.vs..ZT14unf
ed.log2FoldChange, xlab="log2(Mean tag density)", ylab="log2fold change Fed/Unf
ed", pch=1, cex=0.7, col="gray", ylim=c(-1,1.5), xlim=c(0,20))

all_ins_reg = seqdata[ seqdata$ZT14F.vs..ZT14FS961.padj < 0.05, ]
ins_feed_reg = diff_exp[ diff_exp$ZT14F.vs..ZT14FS961.padj < 0.05, ]

points(log2(all_ins_reg$ZT14F.vs..ZT14unfed.baseMean + 1), -all_ins_reg$ZT14F.v
s..ZT14unfed.log2FoldChange, xlab="log2(Mean tag density)", ylab="log2fold chan
ge Fed/Unfed", pch=1, cex=0.7, col="lightblue", ylim=c(-1,1.5), xlim=c(0,20))

points(log2(ins_feed_reg$ZT14F.vs..ZT14unfed.baseMean + 1),-ins_feed_reg$ZT14F.
vs..ZT14unfed.log2FoldChange, xlab="log2(Mean tag density)", ylab="log2fold cha
nge Fed/Unfed", pch=1, cex=0.7, col="slateblue4", ylim=c(-1,1.5), xlim=c(0,20))
```



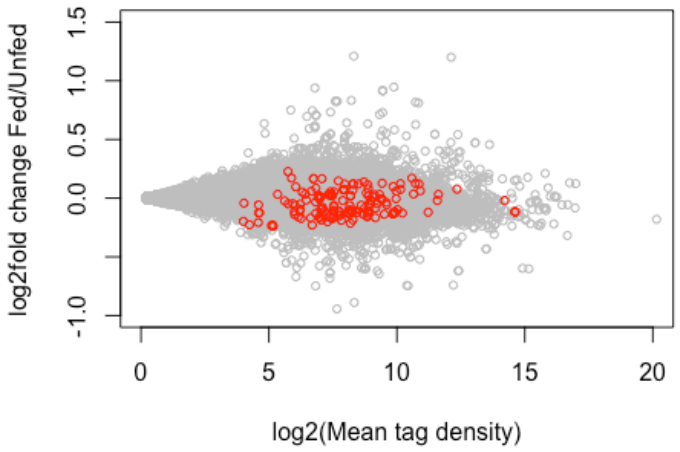

```
# make an MA plot showing insulin-regulated genes not feeding-regulated

plot(log2(seqdata$ZT14F.vs..ZT14unfed.baseMean + 1), -seqdata$ZT14F.vs..ZT14unf
ed.log2FoldChange, xlab="log2(Mean tag density)", ylab="log2fold change Fed/Unf
ed", pch=1, cex=0.7, col="gray", ylim=c(-1,1.5), xlim=c(0,20))

ins_reg_not_feed = seqdata[ seqdata$ZT14F.vs..ZT14FS961.padj < 0.05 & seqdata$Z
T14F.vs..ZT14unfed.padj > 0.05, ]

points(log2(ins_reg_not_feed$ZT14F.vs..ZT14unfed.baseMean + 1),-ins_reg_not_fee
```

```
d$ZT14F.vs..ZT14unfed.log2FoldChange,xlab="log2(Mean tag density)", ylab="log2fold change Fed/Unfed", pch=1, cex=0.7, col="red", ylim=c(-1,1.5), xlim=c(0,20))
```



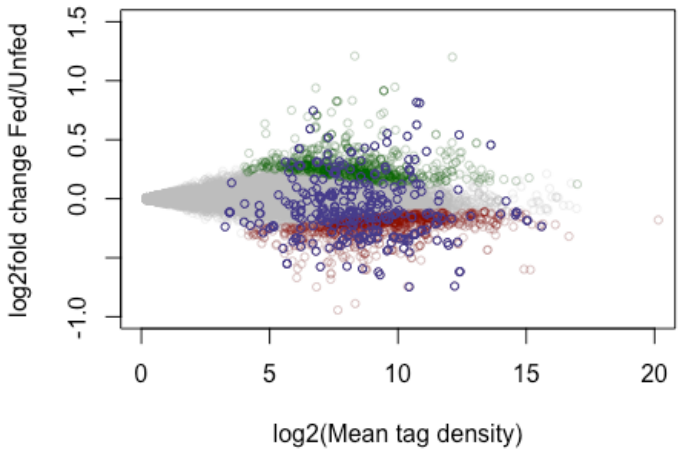

```
# make a merged version of the plots that show feeding-regulation and insulin-regulation

plot(log2(seqdata$ZT14F.vs..ZT14unfed.baseMean + 1), -seqdata$ZT14F.vs..ZT14unfed.log2FoldChange, xlab="log2(Mean tag density)", ylab="log2fold change Fed/Unfed", pch=1, cex=0.7, col=adjustcolor("gray", alpha.f = 0.2), ylim=c(-1,1.5), xlim=c(0,20))

points(log2(diff_up$ZT14F.vs..ZT14unfed.baseMean + 1), -diff_up$ZT14F.vs..ZT14unfed.log2FoldChange, xlab="log2(Mean tag density)", ylab="log2fold change Fed/Unfed", pch=1, cex=0.7, col=adjustcolor("darkgreen", alpha.f = 0.2), ylim=c(-1,1.5), xlim=c(0,20))

points(log2(diff_down$ZT14F.vs..ZT14unfed.baseMean + 1), -diff_down$ZT14F.vs..ZT14unfed.log2FoldChange, xlab="log2(Mean tag density)", ylab="log2fold change Fed/Unfed", pch=1, cex=0.7, col=adjustcolor("darkred", alpha.f = 0.2), ylim=c(-1,1.5), xlim=c(0,20))

points(log2(all_ins_reg$ZT14F.vs..ZT14unfed.baseMean + 1), -all_ins_reg$ZT14F.vs..ZT14unfed.log2FoldChange, xlab="log2(Mean tag density)", ylab="log2fold change Fed/Unfed", pch=1, cex=0.7, col="slateblue4", ylim=c(-1,1.5), xlim=c(0,20))
```



We can see from the last two graphs that there are quite a few genes that are not differentially expressed by feeding but still regulated by insulin, since on the there is a set of plotted points shown in red, and on the second not one all dark blue points overlap with red or green points.

Extra Plots on Feeding-downregulated genes

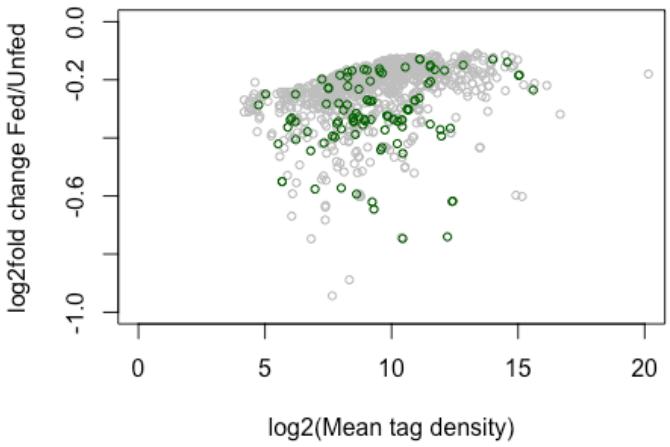
In this last section, I want to make the same kind of plots I had for feeding upregulated genes, but on the downregulated subset.

```
# make an MA plot with only the feeding down-regulated data
plot(log2(diff_down$ZT14F.vs..ZT14unfed.baseMean + 1), -diff_down$ZT14F.vs..ZT1
4unfed.log2FoldChange, xlab="log2(Mean tag density)", ylab="log2fold change Fed
/Unfed", pch=1, cex=0.7, col="gray", ylim=c(-1,0), xlim=c(0,20))


# highlight the insulin-regulated sites

ins_reg_feed_down = diff_down[ diff_down$ZT14F.vs..ZT14FS961.padj < 0.05, ]

points(log2(ins_reg_feed_down$ZT14F.vs..ZT14unfed.baseMean + 1), -ins_reg_feed_
down$ZT14F.vs..ZT14unfed.log2FoldChange, xlab="log2(Mean tag density)", ylab="l
og2fold change Fed/Unfed", pch=1, cex=0.7, col="darkgreen", ylim=c(-1,0), xlim=
c(0,20))
```

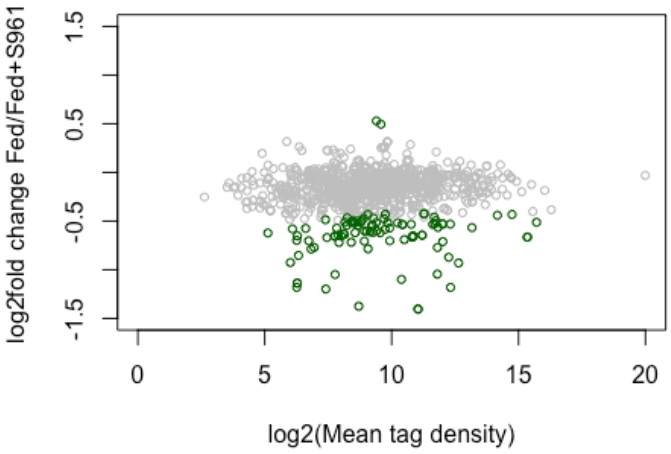



```
## ANOTHER VIEWING OPTION (to distinguish up- and down-regulation by S961)

# make an MA plot with only the feeding down-regulated data
plot(log2(diff_down$ZT14F.vs..ZT14FS961.baseMean + 1), -diff_down$ZT14F.vs..ZT1
4FS961.log2FoldChange, xlab="log2(Mean tag density)", ylab="log2fold change Fed
/Fed+S961", pch=1, cex=0.7, col="gray", ylim=c(-1.5,1.5), xlim=c(0,20))

# highlight the insulin-regulated sites

points(log2(ins_reg_feed_down$ZT14F.vs..ZT14FS961.baseMean + 1), -ins_reg_feed_
down$ZT14F.vs..ZT14FS961.log2FoldChange, xlab="log2(Mean tag density)", ylab="l
og2fold change Fed/Fed+S961", pch=1, cex=0.7, col="darkgreen", ylim=c(-1.5,1.5)
, xlim=c(0,20))
```




```
# identify the 2 genes that are downregulated by S961 ("upregulated" by insulin
)

ins_up_feed_down = diff_down[ diff_down$ZT14F.vs..ZT14FS961.padj < 0.05 & diff_
down$ZT14F.vs..ZT14FS961.log2FoldChange < 0, ]
ins_down_feed_down = diff_down[ diff_down$ZT14F.vs..ZT14FS961.padj < 0.05 & dif
f_down$ZT14F.vs..ZT14FS961.log2FoldChange > 0, ]

View(na.omit(ins_up_feed_down[,7:8]))
View(na.omit(ins_down_feed_down[,7:8]))
```

Note: Result of the list of genes that are insulin-up-regulated but feeding-down-regulated

- Tsc22d3|DIP|Dsip1|Gilz|TSC-22R|Tilz3|RP23-440B21.1|X F1|X|protein-coding
- Tsc22d3|DIP|Dsip1|Gilz|TSC-22R|Tilz3|RP23-440B21.1|X F1|X|protein-coding

Appendix B: Example calculation of primer efficiency (primer testing)

Primer set mSREBP-1c + 10.1 kb

Primer test, 08/12/2016
Test material: AK100 (ChIP), pooled input samples

Table B-1. *qPCR threshold cycle (Cp) values for a series of dilutions of a pooled ChIP DNA input sample, used to trace a calibration curve for determining the efficiency of the tested primer set.*

Sample	Cp1	Cp2	Average	Dilution factor	Log (dil. factor)
4x	24.46	24.46	24.46	0.2500	-0.6021
16x	26.44	26.23	26.34	0.0625	-1.2041
64x	28.19	28.25	28.22	0.0156	-1.8062
256x	30.44	30.57	30.51	0.0039	-2.4082
1024x	31.62	31.82	31.72	0.0010	-3.0103
4096x	34.32	34.01	34.17	0.0002	-3.6124
16348x	37.09	36.41	36.75	0.0001	-4.2135
NTC	40.00				

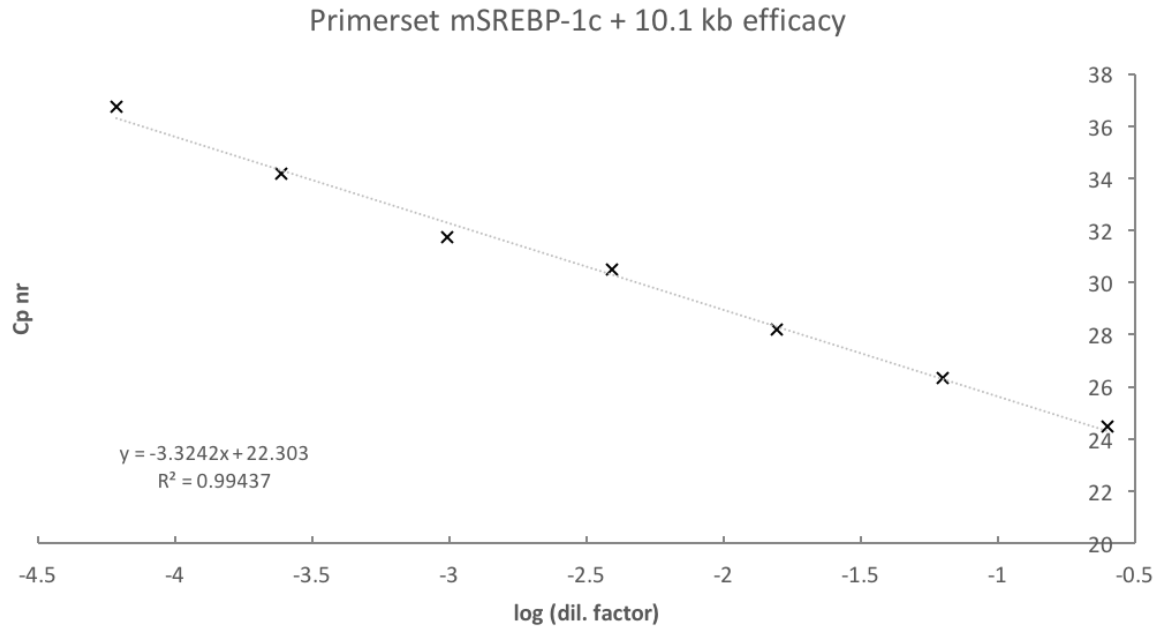


Figure B-1. *Calibration curve for the determination of qPCR primer efficiency, using the values presented in Table B-1.*

Primer efficiency can then be calculated from the slope of the plotted curve:

$$10^{-\frac{1}{\text{slope}}} - 1 = 10^{-\frac{1}{-3.3242}} - 1 = 0.9990 = 99.9\%$$

Appendix C: Example calibration curve for total protein quantification

Table C-1. *Absorbance values (at 562 nm) for standard solutions with different BSA concentrations, used to trace a calibration curve for quantifying total proteins in liver extracts.*

Standards	*BSA Conc. (ug/uL)*	*OD (562 nm)*
A	4.0	0.881
	4.0	0.869
B	2.0	0.528
	2.0	0.518
C	1.0	0.345
	1.0	0.343
D	0.5	0.244
	0.5	0.241
E	0.25	0.193
	0.25	0.178
F	0.125	0.151
	0.125	0.145
G	0.0625	0.129
	0.0625	0.128
H	0.0	0.100
	0.0	0.097

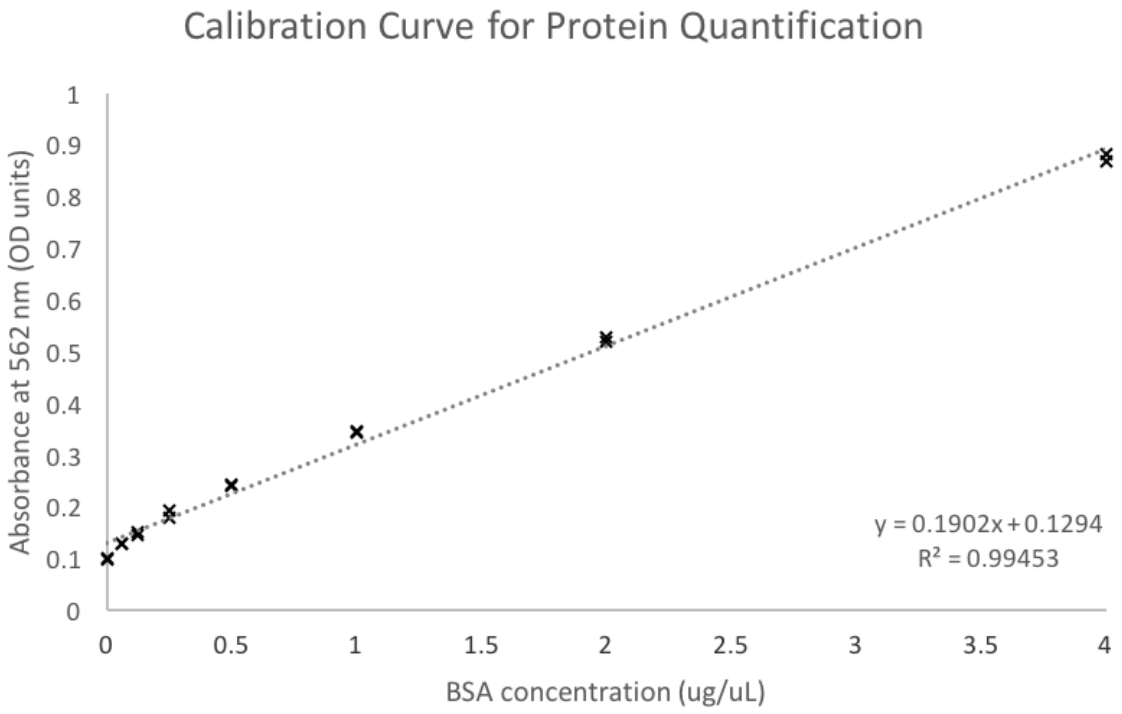


Figure C-1*. Calibration curve for the quantification of total proteins in liver extracts, using the values presented in Table C-1.*

Appendix D: Example calibration curve for DNA quantification (picogreen)

Table D-1. *Fluorescence values for standard solutions with different λ-DNA concentrations, used to trace a calibration curve for quantifying DNA in liver extracts subjected to ChIP.*

Standard	*Conc. (ng/ul)*	*Fluorescence 1*	*Fluorescence 2*	*Average*	*Excl. background*
A	0.00	304	307	305.5	0
B	0.25	542	551	546.5	241.0
C	0.50	775	746	760.5	455.0
D	1.00	1312	1134	1223	917.5
E	1.25	1145	1252	1198.5	893.0
F	2.50	2538	2181	2359.5	2054.0
G	5.00	4712	4368	4540	4234.5
H	10.00	9086	8732	8909	8603.5

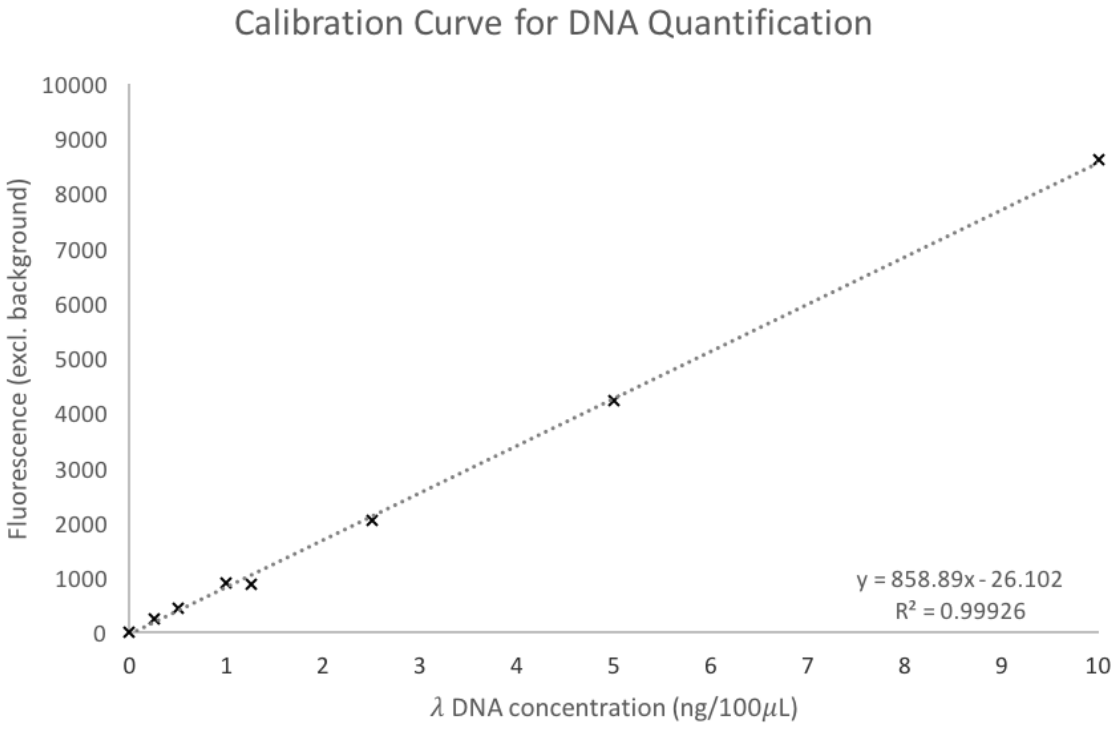


Figure D-1. *Calibration curve for the quantification of DNA in samples obtained after performing ChIP on liver extracts, using the values presented in Table D-1.*

Appendix E: R-script for ChIP-seq data analysis

MUS12 MED1 ChIP-seq

Catarina M Correia

26/05/2017

Loading data

I want to choose my .txt file with the data I need to analyse.

```
# set working directory
setwd("~/…")

# define the file on use and set a name for the data contained
data= read.delim("MED1_combined_peaks_MED1_tagcount.txt", header=TRUE, dec=".")

# format and view your data set
rownames(data) <- data[,1]
data[,1] <- NULL
data <- subset(data, select=c(1:3,19:20))
colnames(data)=c("chr","start","end","unfed","fed")
data$FC <- log2(data$fed/data$unfed)
data$mean <- rowMeans(data[,4:5])

View(data)    # opens a table with all your data
names(data)   # gives a list with the names corresponding to each column number

## [1] "chr"   "start" "end"   "unfed" "fed"   "FC"    "mean"
```

Filtering data (feeding-regulated sites)

Create subsets of data containing only the genome sites that are regulated by feeding, which consists of the genes which are significantly differently expressed (fed/unfed > 2).

```
# create the subsets you need

diff_up = data[ data$FC > 1, ]
diff_down = data[ data$FC < -1, ]
```

Plotting feeding-regulated data, MA plot

Plot all sequencing data, highlighting differentially expressed genes as a result of feeding/fasting entrainment, to show the distribution of down- and upregulated genes by feeding. Downregulated genes are to be shown at the bottom of the plot in red; upregulated genes at the top in green.

```
# make an MA plot with the unfiltered data
plot(log2(data$mean), data$FC, xlab="log2(Mean tag density)", ylab="log2fold ch
ange Fed/Unfed", pch=1, cex=0.7, col="gray", xlim=c(3,9), ylim=c(-4,4))

# highlight the feeding-regulated sites
points(log2(diff_up$mean),diff_up$FC, xlab="log2(Mean tag density)", ylab="log2
fold change Fed/Unfed", pch=1, cex=0.7, col="darkgreen", xlim=c(3,9), ylim=c(-4
,4))

points(log2(diff_down$mean),diff_down$FC, xlab="log2(Mean tag density)", ylab="
log2fold change Fed/Unfed", pch=1, cex=0.7, col="darkred", xlim=c(3,9), ylim=c(
-4,4))
```