TECHNISCHE UNIVERSITÄT MÜNCHEN

Institut für Diabetes und Krebs
(Helmholtz Diabetes Center, Helmholtz Zentrum München)
Chair for Metabolic Programming
(TUM School of Life Sciences)

Multi-omics data integration approaches to study glucocorticoid receptor function

Kinga Balázs

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

Doktors der Naturwissenschaften

genehmigten Dissertation.

Vorsitzender: Prof. Dr. Dmitrij Frischmann

Prüfende/-r der Dissertation:

1. Prof. Dr. Nina Henriette Uhlenhaut
2. Prof. Dr. Hans-Werner Mewes

Die Dissertation wurde am 23.12.2020 bei der Technischen Universität München eingereicht und durch die TUM School of Life Sciences am 19.04.2021 angenommen.

Acknowledgments

I would like to take this opportunity to express my gratitude to those who have supported me during my research years. First and foremost, Prof. N. Henriette Uhlenhaut, for her guidance, scientific input and support throughout this research and the writing of this thesis. Her insights and expertise have been invaluable to the advancement of my work. I would like to thank my thesis committee members, Prof. H. Werner Mewes and Dr. Matthias Heinig. Prof. Mewes, who believed in me and gave me the opportunity to join his research group and provided excellent feedback along the way. Dr. Matthias Heinig has been an invaluable collaborator and his advice and guidance have been instrumental to my work.

My special thanks go to Fabiana, who taught me to think and work in a biologist friendly way. Fabi, you are a great colleague! Mille Grazie! I would like to thank the senior lab members (Franzi, Ken, Céline) who have provided great advice and feedback on many aspects of my work. Many thanks to my former boss Dr. Gabi Kastenmüller, who still considers me an IBIS-ler. Gabi, our scientific discussions significantly contributed to this work. I would like to thank all lab members (Kostas, Laura, Michael, Afzal, Katerina, Suhail, Teresa) that they were all great colleagues. It was a pleasure to work with you and to spend time together. Thank you Sybille and Ivonne for the coordination during my time as a student in the HELENA program.

Biggest thanks to my husband, Zoltán, who supported me always when I doubted that it was “too late” for doing research. I learned that it is never too late to do what you like. It’s worth every minute that you invest in going your way. Szeretlek! Thanks to my girls, who are the biggest teachers in my life. The hardest lesson was to learn patience and the joy of the moment. I would like to thank my parents who supported us during this whole period. I love you both. Thank you for caring and loving us so much. Mom, I still think that “aki mer, az nyer” and not “jobb félni, mint megijedni”. Dad, you are the only person I know who really knows what unconditional love means. You gave me a great lesson for life.

Table of contents

Abstract

Recent technological advances in genomics, metabolomics, proteomics and bioinformatics have greatly accelerated biomedical research and enabled unprecedented mechanistic insights into regulatory networks.

In this thesis I undertook a multi-omics data integration approach to study transcriptional regulation of hepatic metabolism by the glucocorticoid receptor (GR) in mice. My specific focus was the effect of high-fat diet on GR's genomic actions and its role in rhythmic gene expression during the day/night cycle.

GR is a ligand-gated transcription factor that is activated by glucocorticoid hormones (GCs), which are secreted with a prominent diurnal rhythm, and by synthetic steroids such as dexamethasone. To functionally characterize GR binding events at promoters and enhancers and to understand its contribution to rhythmic cycles of gene expression, we performed ChIP-seq, RNA-seq and metabolomics.

Specifically, I analyzed mouse liver tissues from lean and obese mice collected every 4 hours during the day/night cycle, either wild type or liver-specific GR mutants. We modeled the effects of diet, time, genotype and their interactions and identified time-point specific sets of GR targets. By using a non parametric algorithm, we detected rhythmic transcripts.

Integrating ChIP-seq with nearby transcript expression by linear proximity and functionally annotating GR target genes into pathways and gene ontology (GO) categories, we found that the majority of oscillatory genes are bound by and depend on GR to sustain their rhythmicity and amplitude stability.

Moreover, by including targeted metabolomics data, we identified metabolite changes that mirror the effects of altered biological processes associated with a prolonged high-fat diet consumption.

Finally, I performed differential transcript usage isoform analyses to study a potential effect of GR action on alternative splicing. We detected a set of high-confidence RNAs that show dexamethasone-induced differential transcript usage, suggesting that GCs have post-transcriptional functions.

Overall, we demonstrate GR's contribution to the orchestration of rhythmic metabolism by integrating multi-omics datasets from mouse liver samples.

Zusammenfassung

Neueste technologische Fortschritte im Bereich der Genomik, Metabolomik, Proteomik und Bioinformatik haben die biomedizinische Forschung stark vorangetrieben. Dadurch können beispiellose mechanistische Einblicke in regulatorische Netzwerke ermöglicht werden.

Ziel dieser Arbeit war es, anhand eines Multi-Omics Datenintegrationsansatzes die Glukokortikoidrezeptor (GR) vermittelte transkriptionelle Regulation des Leberstoffwechsels in der Maus zu untersuchen. Dabei lag der besondere Fokus auf dem Einfluss von fettreicher Ernährung auf das genomische Bindungsprofil des GR und dessen Rolle in der rhythmischen Genexpression während des Tag/ Nachtzyklus.

GR ist ein ligandenabhängiger Transkriptionsfaktor, der durch Glukokortikoid Hormone (GCs), die dem täglichen Tag/Nachtrhytmus folgen ausgeschüttet werden, sowie durch synthetische Steroide, wie Dexamethason aktiviert wird.

Um die Bindung des GR an “Promotor” und “Enhancer” Regionen funktional zu charakterisieren und dessen Beitrag in der rhythmischen Abfolge der Genexpressionen zu verstehen, haben wir Methoden des Next Generation Sequencing (NGS), wie ChIP-seq, RNA-seq, sowie Metabolomics angewendet.

Dabei untersuchte ich insbesondere Lebergewebe von schlanken und fettleibigen Wildtyp oder leberspezifischen GR knockout Mäusen, die in einer Abfolge von jeweils 4 Stunden während eines gesamten Tag/Nacht Zyklus gesammelt wurden. Danach modellierten wir die Effekte von Ernährung, Zeit und Genotyp, sowie deren Wechselwirkungen und identifizierten dabei eine zeitpunktspezifische Reihe von GR Zielgenen. Durch Anwendung eines nicht parametrischen Algorithmus konnten wir die rhythmischen Veränderungen der Genexpression ermitteln.

Durch die Integration von ChIP-seq mit angrenzender Transkriptionsexpression mittels linearer Näherung und funktioneller Annotierung von GR-Zielgenen und deren Einteilung in Signalwege und Gen-Ontologie (GO), fanden wir heraus, dass die Mehrheit der oszillatorischen Gene von GR gebunden werden. Damit ist die Anwesenheit von GR wichtig um deren Rhythmik und Amplitudenstabilität aufrecht zu erhalten.

Zusätzlich konnten wir durch Berücksichtigung gezielter Metabolomics Daten Veränderungen in Stoffwechselprodukten identifizieren, die die Auswirkungen veränderter biologischer Prozesse widerspiegeln, die mit einer Langzeit Hochfettdiät verbunden sind.

Schließlich führte ich differentielle „Transcript-Usage-Isoform"-Analysen durch, um einen möglichen Effekt des GR auf alternatives Spleißen zu untersuchen. Dabei entdeckten wir eine Reihe von „high-confidence RNAs", die eine Dexamethason-induzierte differentielle Transcript-Nutzung zeigten, welches auf posttranskriptionelle Funktionen der GCs hindeutet.

Insgesamt konnten wir die Rolle von GR bei der Orchestrierung des rhythmischen Stoffwechsels durch die Integration von multi-omics-Datensätzen aus Maus-Leberproben nachweisen.

List of abbreviations

A5/A3	alternative 5'/3' splice site
ACTH	adrenocorticotropin hormone
AF	activation function domain
AF/AL	alternative first/last exon
ANOVA	analysis of variance
AR	androgen receptor
AS	alternative splicing
BH	Benjamini-Hochberg
cDNA	complementary DNA
ChIP	chromatin immunoprecipitation
CRH	corticotropin-releasing hormone
CTRL	control
Da	Dalton
DBD	DNA-binding domain
Dex	dexamethasone
DGE	differential gene expression
DNA	deoxyribonucleic acid
DTU	differential transcript usage
ER	estrogen receptor
ER	exclusion reads
FDR	false discovery rate
GC	glucocorticoid
GLM	generalized linear model
GO	Gene Ontology
GR	glucocorticoid receptor
GR-LKO	liver-specific GR knock out
GRE	glucocorticoid response elements
GTEx	Genotype-Tissue Expression Project
GTF	gene transfer format
H	hinge region
HFD	high-fat diet
HPA	hypothalamic-pituitary-adrenal
HRE	hormone response element
IDR	irreproducible discovery rate
IR	inclusion reads

JTK	Jonckheere-Terpstra-Kendall
KO	knock out
LBD	ligand-binding domain
LM	linear model
LoD	limit of detection
LPC	lysophosphatidylcholine
LRT	likelihood ratio test
MAR	missing at random
MCAR	missing completely at random
mHG	minimum hypergeometric
ML	maximum likelihood
MNAR	missing not at random
MR	mineralocorticoid receptor
MS	mass spectrometry
MX	mutually exclusive exon
NB	negative binomial
NGS	next-generation sequencing
NL	nuclear localization signal
NR	nuclear receptor
NTD	N-terminal transactivation domain
OFDR	overall false discovery rate
PC	phosphatidylcholine
PCA	principal component analysis
PR	progesterone receptor
PSI	percent spliced in
QQ	quantile-quantile
RI	retained intron
RNA	ribonucleic acid
SCN	suprachiasmatic nucleus
SE	skipping exon
SEM	standard error of the mean
SM	sphingomyelin
TF	transcription factor
UNTR	untreated
VST	variance stabilizing transformation
WT	wild type
ZT	zeitgeber time

1. Introduction

1.1 The mammalian circadian clock

Almost all light-sensitive organisms, from cyanobacteria to humans, have acquired during evolution a timing system called the circadian clock. It consists of an autonomous and endogenous oscillator with a period of ca. 24 hours (therefore the name "circadian", meaning approximately one day) which is responsible for maintaining the rhythmicity in behavior, physiology, and metabolism.

While the molecular machinery of the clock is contained in all cells, the circadian clock is hierarchically organized. Indeed, in mammals, a master circadian clock is located in the suprachiasmatic nucleus (SCN) of the hypothalamus (Klein et al., 1991). The SCN receives light input (its major Zeitgeber, i.e. a stimulus that can entrain the circadian clock to the 24-hour cycle on Earth) via the retina, and regulates peripheral clocks through endocrine and systemic cues. In addition, other environmental cues such as feeding time and body temperature appear to play a role in the resetting of peripheral timekeepers. The most striking difference between the master and peripheral clocks lies in the fact that while SCN is not drifting out of phase when perturbations are coming from internal cues, peripheral clocks adapt to reflect their local metabolic environment.

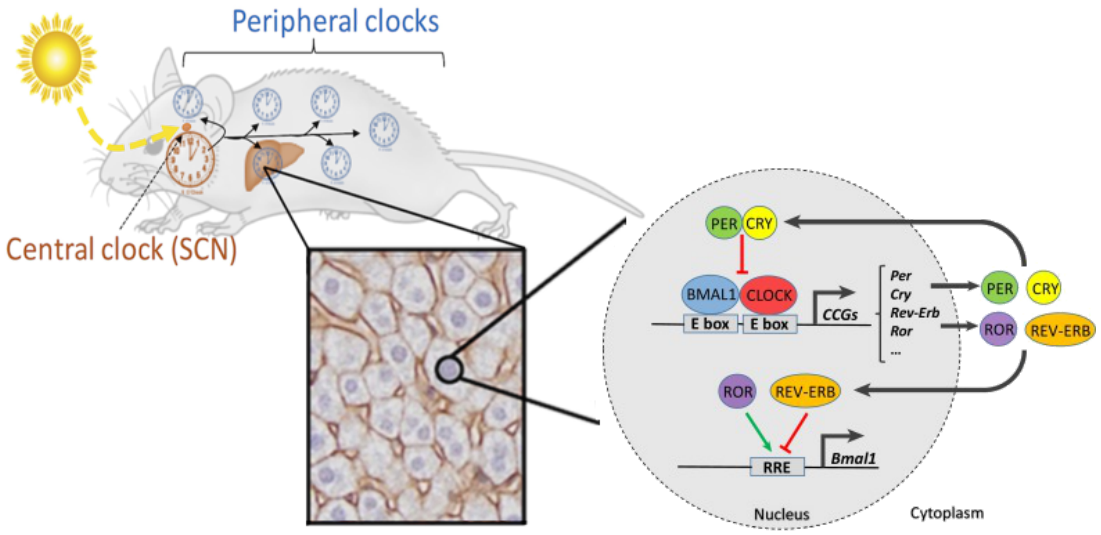


Figure 1: Hierarchical organization of the circadian clock in mammals.
The master clock resides in the SCN and entrains peripheral clocks, for example in the liver. The molecular clock machinery is a transcription/translation feedback loop oscillating with an approximately 24 hour period. Picture modified from (Atger et al., 2017) & (Gaucher et al., 2018).

At the molecular level, the circadian clock is formed by two interlocking transcription/translation feedback loops (Figure 1). In the first loop, the transcriptional activators BMAL1/CLOCK heterodimerize and then bind to conserved E-boxes at promoters of target genes (Gekakis et al., 1998). This way they induce their own repressors PER1/2/3 and CRY1/2 during the circadian day, which form a protein complex that translocates back to the nucleus and inhibits CLOCK:BMAL1-induced gene expression (Lee et al., 2001). The PER/CRY heterodimer is degraded during the night, which allows a new cycle of transcription to take place, as CLOCK/BMAL1 is no longer repressed. The second loop is formed by the nuclear receptors REV-ERBα/β and ROR α/β/γ which regulate the cyclic *Bmal1* transcription (Takahashi, 2016). The transcription of *Rev-erbα* and *β* is activated by the heterodimer BMAL1/CLOCK during the day and transrepressed by CRY/PER, which leads to circadian oscillation of *Rev-erbα* and *β*. REV-ERBα represses *Bmal1* and *Clock* transcription (Guillaumond et al., 2005). RORα and REV-ERBα share DNA binding elements, the ROREs in the *Bmal1* promoter, and have to compete for binding. As a result, the expression of *Bmal1* is repressed by REVERBα and activated by RORα.

1.2 Nuclear hormone receptors

Nuclear receptors (NRs) comprise a superfamily of structurally conserved, ligand-dependent transcription factors (TFs). There are 48 human (Figure 2) and 49 mouse nuclear receptors in total. With their conserved DNA-binding domains (DBD), they are able to recognize and bind to hormone response elements (HRE) as homo- or heterodimers and regulate gene expression both positively and negatively by recruiting co-regulators and histone-modifying enzymes (Mangelsdorf et al., 1995).

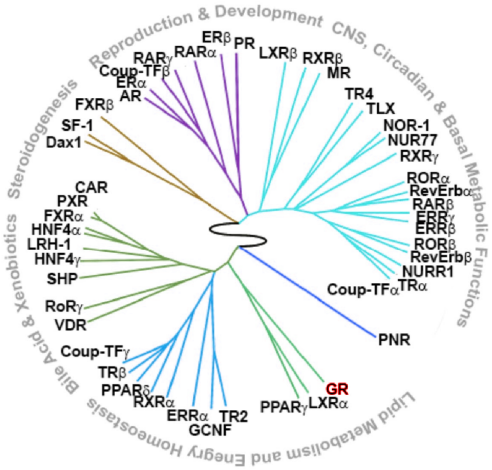


Figure 2: The superfamily of nuclear hormone receptors.
Clustering of nuclear hormone receptors (with GR highlighted in red) based on their distribution in tissues, showing their connection to physiological pathways like lipid metabolism and energy homeostasis. Picture modified from (Bookout et al., 2006).

In my thesis, I focus on one member of the NR family, namely the glucocorticoid receptor (GR). GR together with the estrogen receptor (ER), the mineralocorticoid receptor (MR), the androgen receptor (AR), and the progesterone receptor (PR) are known as the "classic" steroid receptors, all of them being named after the endogenous steroid hormone that they recognize. Since the complete cDNA of GR and ER were the first to be isolated in the 1980s, they are the "founding members" of the NR family (Greene et al., 1986; Hollenberg et al., 1985).

1.2.1 The glucocorticoid receptor (GR)

The glucocorticoid receptor (GR gene or *NR3C1*, i.e. nuclear receptor subfamily 3, group C, member 1) is a ligand-activated transcription factor that belongs to the steroid/thyroid hormone receptor superfamily. It resides in the cytoplasm together with chaperone proteins (hsp90, hsp70, and p23) (Rajapandi et al., 2000) and immunophilins of the FK506 family (FKBP51 and FKBP52) forming a multi-protein complex (Pratt and Toft, 1997). Upon activation, by glucocorticoid hormones (GCs, mainly cortisol in humans and corticosterone in mice), it can translocate to the nucleus, where it induces or represses the transcription of target genes (Figure 3).

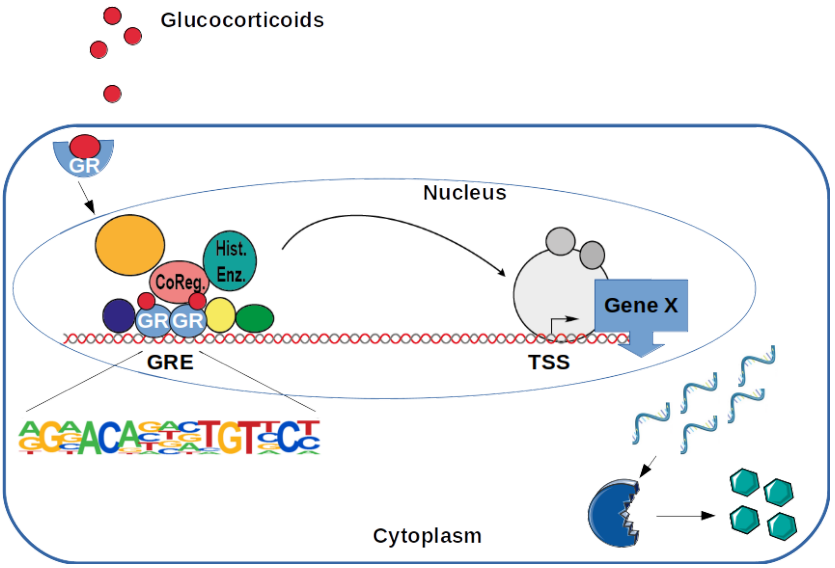


Figure 3: Gene regulation by the glucocorticoid receptor.
The dimeric glucocorticoid receptor (GR) binds Glucocorticoid Response Elements (GREs) to regulate target genes.

As a modular protein, it has three major domains: an N-terminal transactivation domain (NTD), a central DNA-binding domain (DBD), and a C-terminal ligand-binding domain (LBD) (Figure 4) (Oakley and Cidlowski, 2013). A flexible hinge region (H) separates the DBD and LBD. The 2 zinc finger motifs within the DBD, which is the most conserved region, play a role in the recognition and binding of target DNA sequences, i.e. glucocorticoid response elements (GREs). A transcriptional activation function domain (AF1) within the NDT is necessary for interactions with coregulators and the basal transcription machinery. The LBD harbors a hydrophobic pocket needed for ligand binding, and also contains an activation function domain (AF2). The transport into the nucleus is facilitated by the nuclear localization signals NL1 and NL2. NL1 is located at the DBD/hinge region junction, while NL2 is localized within the LBD (Kumar and Thompson, 2005).

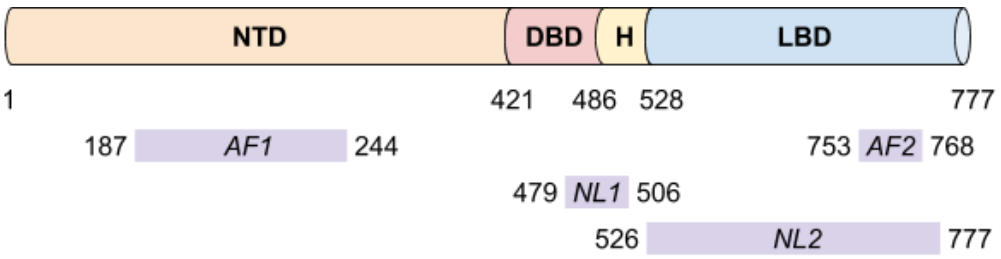


Figure 4: Schematic representation of the glucocorticoid receptor structure.
The N-terminal transactivation domain (NTD), the central DNA-binding domain (DBD), the hinge region (H), and the C-terminal ligand-binding domain (LBD) forming the glucocorticoid receptor. Transcriptional activation function domains (AF1 and AF2) and nuclear localization signals (NL1 and NL2) are indicated in purple. Numbers represent amino acids for human GR. Picture based on (Oakley and Cidlowski, 2013).

1.2.2 Glucocorticoids: activators of the glucocorticoid receptor

Since their discovery in 1949 by Philip Showalter Hench (Hench et al., 1949), when he successfully treated patients suffering from rheumatoid arthritis by administering them the “compound E”, glucocorticoids became one of the most widely used anti-inflammatory and immunosuppressive drugs.

Glucocorticoids are hormones produced mainly by the adrenal gland cortex that are released with both pulsatile and circadian rhythmicity, controlled by the hypothalamic-pituitary-adrenal (HPA) axis (Chrousos, 1998; Walker et al., 2010). After it is released from the hypothalamus, the corticotropin-releasing hormone (CRH) initiates the production of adrenocorticotropin (ACTH) by the

anterior pituitary. ACTH then signals to the cortical layer of the adrenal gland to produce cortisol from cholesterol. In a negative feedback loop, circulating glucocorticoids block the release of CRH and ACTH, thus inhibiting glucocorticoid secretion. Although also non-adrenal GC production was observed for example in the brain, thymus, or epithelial barriers, it seems that this does not affect the levels of systematically circulating GCs but results in high spatial specificity of steroid action (Taves et al., 2011).

Endogenous GC levels reach their peak prior to the active period (which is in the early morning in humans, and early evening in nocturnal animals, like mice), and decrease gradually at the onset of the inactive period. GCs are powerful regulators exerting their pleiotropic effects on metabolism, immune response, and cognitive and emotional functions. Their successful use in the clinics is due to the fact that they can effectively suppress inflammation not only by activating anti-inflammatory genes but also by suppressing pro-inflammatory and immune genes (Coutinho and Chapman, 2011).

1.2.3 Glucocorticoids: link between the clock and metabolism

The HPA axis plays a crucial role in the body's stress response. To meet increased energy demands and restore homeostasis in periods of acute or chronic stress, the adrenal cortex induces GC secretion (Herman et al., 2016). Glucocorticoids act as signaling molecules that stimulate hepatic gluconeogenesis. This glucose represents the major source of energy for the brain and peripheral tissues in times of fasting (Sacta et al., 2016). The release of glucocorticoid hormone is controlled by the SCN and GCs are able to influence peripheral clocks, as for example in the liver (Reddy et al., 2007).

Previous studies demonstrated that when cells (e.g fibroblasts) and mice were treated with glucocorticoids, it induced oscillation and circadian gene expression in the liver, kidney, and heart (Balsalobre et al., 2000; Oishi et al., 2005). Also, both physical and genetic interactions between the core clock machinery and the glucocorticoid receptor have been shown. For example, in the regulatory region of *Per1* and *Per2* functional GREs were identified, GR was found to be able to repress *Rev-erbα* promoter activity, and GR while interacting with CRY1 and CRY2 in a ligand-dependent manner repressed the activation of gluconeogenic genes like *Pck1* (Lamia et al., 2011; So et al., 2009; Torra et al., 2000). Additionally, physical interaction between CLOCK and GR was shown, where CLOCK suppressed the binding of GR to DNA (Nader et al., 2009).

1.2.4 Glucocorticoids signaling in the liver

There are only some genes that are commonly activated by GR between different tissues (e.g. *Dsip1* and *Sgk1*), as GC-induced transcription shows cell-type specificity. In the liver, *phosphoenolpyruvate carboxykinase 1* (*Pck1*) and *glucose-6-phosphatase* (*G6pc*), the two main rate limiting enzymes involved in hepatic gluconeogenesis, are direct targets of GR (Imai et al., 1990). Functional GREs were found in their promoters and GR binds the enhancer and promoter regions of both in the liver (Grøntved et al., 2013). Therefore, in liver-specific GR knockout mice, reduced transcriptional activation of these genes was observed in response to prolonged fasting (Opherk et al., 2004).

Another GR target gene is *Hes1*, involved in hepatic lipid metabolism. By recruiting histone deacetylases to its promoter, GR represses *Hes1* expression. As a consequence, in liver-specific GR knockout mice, derepression of *Hes1* is observed, and in fatty liver mouse models hepatic steatosis is improved (Lemke et al., 2008). *Fatty acid synthase (FASN)*, an enzyme involved in lipogenesis, that is expressed in the liver, adipose tissues, and the lung is also regulated by glucocorticoids (Sul and Wang, 1998). *Fasn* is repressed by GR, meaning that liver-specific GR knockout mice are at risk of developing hepatic steatosis, as they show increased triglyceride and lipid accumulation. Moreover, it has been shown that *Dgat1* (expressed in many tissues) and *Dgat2* (mostly highly expressed in liver and adipose tissue), both genes encoding enzymes in triglyceride synthetic pathways, are regulated by GR (Dolinsky et al., 2004). Additionally, a study conducted by Cheng and colleagues could show that dexamethasone (a synthetic GR agonist) induced the expression of the mouse gene *Ntcp* (sodium-taurocholate cotransporting polypeptide*, Slc10a1*), suggesting that it could be GR mediated, as the mouse *Ntcp* promoter harbors functional GREs (Cheng et al., 2007). *Ntcp* is a bile acid transporter gene, suggesting that liver-specific GR knockout mice are prone to bile acid accumulation.

1.2.5 Clock and GR in disease

Perturbations of the clock, that are either environmentally-induced or genetic, were linked to metabolic dysfunctions (Froy, 2010). Early lesion experiments (pre-1950) show how disruptions in the different parts of the hypothalamus can cause both hyperphagia and obesity, or cessation of food consumption (Dietrich and Horvath, 2009). Therefore, the interaction of the circadian clock with hypothalamic function is of special interest (Eckel-Mahan and Sassone-Corsi, 2013), as the disruption of metabolic homeostasis causes obesity, insulin resistance, and even type 2 diabetes. Previous studies have shown that a prolonged high-fat diet feeding can alter the mammalian circadian clock, however, the underlying mechanism is not yet elucidated (Kohsaka et al., 2007).

Disruption of either clock function or GC release was linked to severe metabolic implications. For example, the continual increase of GC levels was shown to lead to an approximately two-fold risk of developing metabolic syndrome (Chandola et al., 2006). Also, the excessive constant GC production (as in the case of patients suffering from Cushing's disease) induced all clinical phenotypes like obesity, diabetes mellitus, dyslipidemia, hypertension, and osteoporosis (Feelders et al., 2012).

However, it is not only the endogenous GC release that can be dysregulated and lead to health problems, but their use as commonly prescribed drugs is associated with both immediate and long-term complications and comorbidities. Although side effects depend on the type of GC, length and dosage of treatment, as well as the mode of application, up to 90% of patients who take them for more than sixty days develop adverse effects (Ericson-Neilsen and Kaye, 2014). GCs are administered in case of allergies and asthma to treat overreaction by stopping the inflammation process, they can be used after an organ transplant to prevent the rejection of the donor organ but also in the treatment of hematological cancers (Oakley and Cidlowski, 2013). However, the long-term side effects of these drugs are sometimes very damaging ranging from skin atrophy (Coondoo et al., 2014), muscle atrophy (Rose and Herzig, 2013), osteoporosis (Briot and Roux, 2015) to steroid-induced psychoses (Janes et al., 2019).

Detrimental side effects on liver metabolism are mainly attributed to GCs' ability to induce metabolic genes, which in acute stress response does not cause damage, as it is meant to be a short-term event. However, prolonged administration of GCs simulates permanent stress, and constantly elevated GC levels lead to perturbations in glucose and lipid metabolism. Administration of GC treatment over a longer period of time increases the risk of developing hyperglycemia and insulin resistance in peripheral tissue, known as "steroid diabetes" (Hwang and Weiss, 2014; Tamez-Pérez et al., 2015). Corticosteroid treatment also increases insulin resistance in patients with diabetes mellitus (Liu et al., 2013). Previous studies have linked prolonged GC treatment with the development of obesity and redistribution of white adipose tissue (increased abdominal adiposity) (Dallman et al., 2004). Additionally, GCs can induce lipolysis and *de novo* lipogenesis in peripheral adipose tissue. Moreover, high levels of circulating free fatty acids lead to an overall state of dyslipidemia (Carmena, 2005), which in turn affects the cardiovascular system. Corticosteroids have major effects on the liver, as they trigger lipid accumulation in the liver, hepatic enlargement, and even steatosis by inducing genes involved in triglyceride synthesis and lipogenesis (Vegiopoulos and Herzig, 2007; Wang et al., 2012).

Deciphering the regulation of energy metabolism by GR remains challenging and of great interest, as the benefits of these anti-inflammatory drugs are shaded by the major metabolic complications.

1.3 Next-generation sequencing (NGS) and omics data

Next-generation or high-throughput sequencing technology (e.g. Illumina (Bentley et al., 2008), Biotechnologies/SOLiD (Metzker, 2010)) has revolutionized genomic research. The first generation or Sanger sequencing technology required over a decade to finish the first version of the human genome. In contrast, NGS enables the sequencing of thousands or even millions of short reads in parallel, very fast (even less than 24 hours), at a low error rate, and affordable price (Wadapurkar and Vyas, 2018). Its applications range from DNA-sequencing and assembly of unknown genomes, exome sequencing, RNA- or transcriptome-sequencing, identification of DNA regions of DNA binding proteins (ChIP-seq), to the discovery of mutations.

Even though progress has been made, processing and analysis of NGS data can still be challenging as data are generally processed through multiple steps, and each of them needs different programs. The available web-based analysis platforms, although practical, user-friendly, and fast, cannot cover the broad range of existing programs and are not easily customizable. As for every step in the analysis, there are numerous alternatives to choose from, wet-lab scientists require bioinformatics support for optimal analysis.

1.3.1 RNA-seq data analysis

Whole transcriptome shotgun or RNA-seq is an application of NGS that enables the screening of the entire transcriptome of any organism or even single cells. Transcriptome analysis allows *de novo* assembly of transcriptome and gene discovery, as well as differential expression analysis and isoform analysis (Morozova et al., 2009).

Figure 5: Graphical representation of the experimental steps in RNA-seq.

RNA-seq identifies and quantifies RNA species at a given time point (as RNA abundance is not stable over time) in biological samples. The experimental protocol of RNA-seq has the following steps: (i)

RNA extraction and random fragmentation; (ii) conversion of RNAs into cDNAs, while attaching adapters to one or both ends; (iii) polymerase chain reaction amplification and sequencing of cDNAs (Figure 5). The raw sequencing data consists of a list of reads with associated quality scores for each sample, which are then subjected to RNA-seq data analysis.

In the scope of this thesis, only the application of RNA-seq for differential gene expression and isoform expression analysis is considered. The analysis essentially consists of mapping, quantification, filtering and normalization, detection of differentially expressed genes/isoforms, and finally the biological interpretation of the results (Figure 6) .

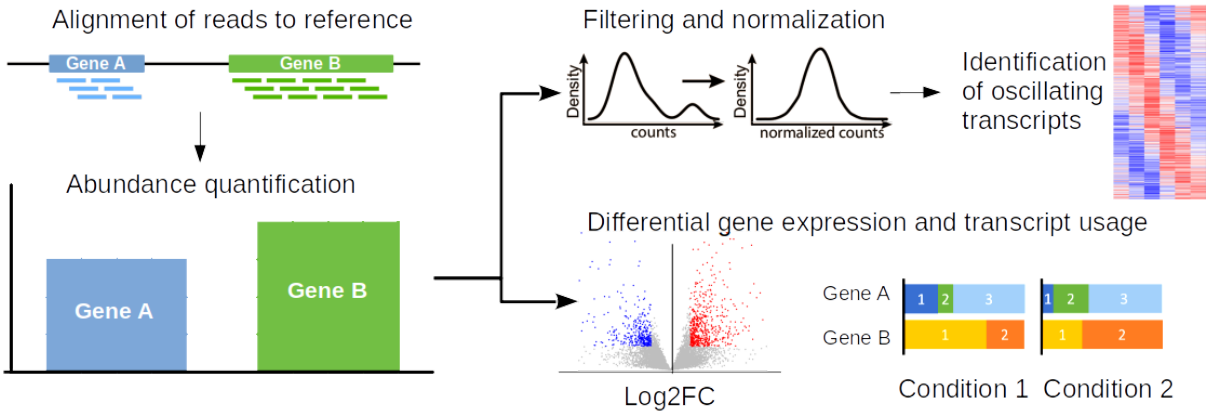


Figure 6: Graphical representation of the main steps in RNA-seq data analysis.

1.3.1.1 Quality check of the sequencing reads

Prior to downstream analysis, the quality of the raw data has to be assessed. This includes checking the following metrics: the quality of the base calls, GC content, duplication rate, adapter content, etc. A very popular tool is FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/), which evaluates all the above-mentioned criteria, and returns an overview of the metrics for each sample. While poor-quality data (across samples or a large percent of sequences with low mean quality) indicate problems in the sequencing, an unexpected %GC for the organism suggests sequence contamination. Having a high level of sequence duplication can be a consequence of having too many PCR cycles, not enough starting material, or a low complexity library. Depending on the quality, removal of adapter sequences and of low-quality bases by read trimming might be necessary prior to mapping.

1.3.1.2 Read mapping

Assigning reads to their best matching location in the reference is called mapping. Fragments, as a result of NGS sequencing, can either be mapped to a reference transcriptome or genome. When mapping to the transcriptome all isoforms of a gene are considered separately, whereas in the latter, reads are aligned to the underlying genes, regardless of what isoform the reads stem from (Babarinde et al., 2019).

STAR (Dobin et al., 2013), TopHat (Trapnell et al., 2009), TopHat2 (Kim et al., 2013), and Bowtie2 (Langmead and Salzberg, 2012) are the most popular splice-aware (can handle reads that span across introns) alignment tools. While these rely on a reference genome, an alternative is to use efficient alignment-free tools, which quantify the transcripts directly by mapping them to a transcriptome, as it is done for example by Kallisto (Bray et al., 2016) and Salmon (Patro et al., 2017). This quantification method works by fragmenting the reads into all possible k-mers, filtering out the unique ones, and then mapping them to the transcriptome, which is pre-indexed for efficiency.

STAR (Spliced Transcripts Alignment to Reference) (Dobin et al., 2013) is a high accuracy method that was shown to be around 50 times faster than its alternatives. It comes with the tradeoff of being overly memory-intensive because it employs an uncompressed suffix tree of the reference genome, that facilitates very efficient searches for exact matches. The alignment is based on finding “Maximal Mappable Prefixes”, i.e. the longest possible exact matches of partial reads, and using these as seeds for the alignment of the unmapped portions. Doing this sequentially over the length of the read results in splitting it into several portions at the supposed splice sites.

1.3.1.3 Quality check of the aligned data

Evaluating the quality of the aligned data includes for example checking the total percent of reads aligned to the genome, the percent of uniquely mapped reads, percent of unmapped reads, and even percent of multimappers. However, possible problems can also be detected, when comparing the sequencing depth between samples. Additionally, in case of paired-end sequencing, the percent of properly paired reads is another important quality metric that has to be checked. MultiQC (Ewels et al., 2016) and Qualimap (García-Alcalde et al., 2012) are often used tools for alignment evaluation.

1.3.1.4 Gene and transcript level quantification

Reads can be counted and clustered over different genomic features, such as transcripts or genes. Most commonly quantification refers to the estimation of the gene level abundances, by counting the number of reads/fragments overlapping the exons of the gene. However, a significant amount of the reads will map outside annotated exons, even for the best-annotated human or mouse data (Pickrell et al., 2010).

Due to a high similarity among isoforms, quantifying the abundances of individual transcripts is a more complex task than gene level quantification. Isoforms of a gene are composed in a big proportion from the same exons, which implicitly means that a significant fraction of the reads are so-called multi-mappers or multireads (i.e reads mapping to multiple locations, multiple transcripts). There are several approaches to handle multimappers, the easiest being to discard these and keep only uniquely mapped reads. While this approach is often used in gene-level quantification, when it comes to transcript-level quantification it is not recommended to exclude them, because we would lose a significant proportion of meaningful reads. Of note, by this way a bias would be introduced, as for genes that have more active isoforms the abundance would be underestimated. Another modality to handle them is to keep all matches, which leads to a higher number of mapped reads than the number of raw reads. However, in practice, the most applicable approach is to rescue these reads by allocating ambiguous reads in relative proportion based on probabilistic inference, as for example in RSEM (Li and Dewey, 2011) and TopHat (Trapnell et al., 2009).

CuffLinks (Roberts et al., 2011), featureCounts (Liao et al., 2014), Kallisto (Bray et al., 2016), and Salmon (Patro et al., 2017) are the most widely used quantification tools. Of these, Kallisto and Salmon are transcript based approaches, which estimate transcript abundances using Expectation Maximization. The featureCounts algorithm is an example of an “exon-based” approach, which assigns reads to features if the feature overlaps with the read or fragment on at least one basepair. It is calling a hit for a meta-feature if any of the component features are overlapping with the read or fragment. The method takes into account any gaps (i.e. insertions, deletions, exon-exon junctions, or structural variants) in the read while it maps each base in the genomic region spanned by the features. These methods finally output a matrix of read/fragment counts, in which the features of interest are in the rows and the different samples in the columns.

1.3.1.5 Filtering and normalization of count data

Normalization of the count data, that was generated in the previous step, is the most crucial step in the analysis. Its impact on the results is higher than that of the mapping method or the test statistics used for finding differentially expressed genes (Bullard et al., 2010; Li et al., 2017).

One can distinguish between two modalities: within and between sample normalization. In the first case by correcting for gene length and sequence composition, for example, GC-content (Benjamini and Speed, 2012), comparisons between the features of a single sample are enabled. In the second case, when features of different samples have to be compared, normalization is performed to adjust for the library size, i.e. sequencing depth (Sims et al., 2014; Tarazona et al., 2011). In this case it is assumed that under different conditions the total expression stays similar, therefore the same amount of mRNA is expected per cell over the different conditions (Evans et al., 2017). Therefore, the total count is used for normalization, i.e. the read counts are divided by the sum of the reads in each sample (Dillies et al., 2013). In order to account also for the length of the gene, the RPKM method for example extends the total count normalization by computing the reads per kilobase per million mapped reads (Mortazavi et al., 2008).

Alternative methods derive the normalization parameters from genes that are deemed to be non-differentially expressed between samples. One such example is the Trimmed Mean of the M-values approach of the edgeR package, which excludes differentially expressed genes from the computation of the normalization factor, under the assumption that most of the genes are not changing (Robinson et al., 2010). A reference sample is selected, logarithm count ratios are computed relative to it, and their mean is used for normalizing read counts. A similar normalization strategy is employed by DESeq (Anders and Huber, 2010), with the difference that for each gene of a reference sample the count ratio is computed relative to the geometric mean of all other samples, then the median of these is used for scaling.

For sample filtering Principal Component Analysis and the ComplexHeatmaps (Gu et al., 2016) R package are useful as they help to visually identify outlier samples (Robinson and Oshlack, 2010). Outliers can be considered for removal in order to not distort downstream analyses.

1.3.1.6 Differential gene expression (DGE) analysis

Early differential gene expression analysis methods approximated the RNA-seq count data by a Poisson distribution, assuming that reads are the result of a random sampling process (Audic and Claverie, 1997). However, this would require that the variance and mean of RNA-seq counts to be equal (to match the properties of the Poisson distribution), which is not the case. Therefore, the Negative Binomial (NB) distribution proved to be more appropriate for the modeling (Hu et al., 2012; Hulse and Cai, 2013), and it is used by the most popular methods, including the DESeq2 package (Love et al., 2014). The NB distribution has two parameters: the dispersion and the mean. In contrast to the Poisson distribution, the mean (μ_{ij}) and the variance (Var K_{ij}), i.e. within-group variability of read counts per gene, are not equal, but linked by the dispersion parameter α_i through a second order polynomial equation, as shown in the formula: Var $K_{ij} = \mu_{ij} + \alpha_i \mu^2_{ij}$. Note that this is a more general formulation than the Poisson distribution used earlier, as for $\alpha_i = 0$ one gets μ_{ij} = Var K_{ij} whereas the relationship changes to a nonlinear one for $\alpha_i \neq 0$.

DESeq is estimating the dispersion of the data in three steps. First, a gene-wise dispersion value is fitted using maximum likelihood (ML). Second, these values are modeled by fitting a smoothed curve that is a function of the mean expression level. Finally, this function is used to estimate the gene-wise dispersion. Instead of this regression approach, in DESeq2 an empirical Bayes method is applied to move the observed (i.e. ML) dispersion values closer to the modeled (i.e regressed) ones.

For the DGE analysis itself, the distribution of counts between different groups is compared. For example, in DESeq2 a generalized linear model (GLM) is fitted to each gene. The coefficients of the so-called design matrix (i.e. the linear function of the free variables) represent a log2 fold change in simple case-control experiments, but more complex relations can also be modeled. After estimating the coefficients for each gene, the null hypothesis of them being 0 is tested, i.e. whether there is no effect of the variables differentiating the groups. There are two tests implemented in DESeq2: (i) the likelihood ratio test that compares the full with a reduced model (where certain coefficients are left out, in order to test their importance in explaining the data); (ii) and the Wald test which also allows the contrasting of coefficients in order to formulate more advanced queries.

1.3.1.7 Analysis of RNA-seq time-series data

Time-series gene expression data is key to study and model dynamic biological processes and allows for finding genes with non-permanent expression changes. Although, not only gene expression can be

measured over time, still the majority of available time-series data is gene expression data (Bar-Joseph et al., 2012).

The sampling density of such experiments depends on the type of experiment. It was shown that in perturbation-response experiments (e.g. the immune response to a vaccine) having higher sampling rates at the beginning of the experiments is more important (Gaucher et al., 2008). In contrast, when circadian rhythms are of interest it is recommended to use uniform sampling and covering multiple cycles.

The trade-off between adding more time-points vs. more replicates depends again on the question to be answered. If the goal is to detect differentially expressed genes at certain time-points, then adding replicates is a better investment, whereas if we want to detect all changing transcripts over time, including those, that have a rise in expression only for a short period of time, than we should opt for more dense sampling (Bar-Joseph et al., 2012).

The classical way to analyze such data would be to use Fourier Transform methods, which would require over 10 cycles of the target periodicity (Takahashi and Menaker, 1982). Therefore, as these experiments are very costly, methods were developed that are suitable for the analysis of data sets that cover a shorter time span, that are still able to estimate the period and the amplitude. The most known programs for analyzing time-series RNA-seq data are JTK Cycle (Hughes et al., 2010), ARSER (Yang and Su, 2010), Lomb-Scargle (Glynn et al., 2006), and COSOPT (Panda et al., 2002).

1.3.1.8 Differential transcript usage (DTU) analysis

The majority of human genes express multiple isoforms and alternative exons show tissue-specific regulation (Wang et al., 2008). Specifically, it was found that more than 90% of human multi-exon genes are alternatively spliced (Carninci, 2009). The most common splicing events include exon skipping, mutually exclusive exons, alternative 3' and 5' donor sites, and alternative exons that contribute to the diversity of RNA isoforms and implicitly increase protein diversity. Although RNA-seq made possible a better characterization and quantification of the transcriptome, current technologies are still limited by the read length and thus cannot reveal full-length transcripts.

Splicing analysis tools can be of three different categories: (i) exon based e.g. DEXseq (Anders et al., 2012) and limma (Ritchie et al., 2015); (ii) splicing based e.g. rMATS (Shen et al., 2014) and SUPPA2 (Trincado et al., 2018); and (iii) transcript based e.g. DEXseq and SUPPA2. In the exon based methods, the exons of genes are considered. For each exon, it is tested if an exon is differentially used

compared to all other exons in that gene. In the splice based approach, differences between conditions are searched for at each splice event. In the transcript-based methods, the transcript level quantification of the RNASeq data is used (instead of gene level quantification) the goal being to detect isoform switches between conditions.

Originally DEXseq was developed for analyzing differential exon usage, hence the acronym. However, now it can be used also for DTU analysis, by considering the counts not being associated with exonic parts, but rather to different transcripts, resulting from the alternative splicing of RNA (Soneson et al., 2016b). In an intermediary step transcript-level abundance estimates are transformed to generate roughly count-scaled data. This way the counts follow a negative binomial distribution, and a generalized linear model (GLM) can be applied to analyze the different counts obtained for the transcripts of the same gene. The GLM allows for non-Gaussian error distributions (in our case NB) while retaining the advantage of the use of multiple coefficients that can then be interpreted and combined in various ways, and allows for the introduction of additional covariates if necessary.

DTU analysis in SUPPA2 relies on the "percent spliced in" (PSI or Ψ) metric, which is calculated based on the inclusion reads (IR) and exclusion reads (ER): PSI = IR / (IR + ER). It represents the fraction of mRNA of the inclusion isoform (Wang et al., 2008). A PSI value of 100% would indicate that the exon is included in all expressed isoforms. The statistical significance of each difference in PSI (ΔPSI) between two conditions is assessed by building a null distribution of the ΔPSI values between replicates. To make sure the null hypothesis is applicable to the testing of the observed ΔPSI, only transcripts or events with similar expression levels are considered. If the observed ΔPSI is more extreme than 1-α (where α is the significance level) of the no-effect values, then the null hypothesis can be rejected. This approach has the advantage of being orders of magnitude faster than DEXseq, up to 100 times in the tests (Trincado et al., 2018). When used as an event-based method, the splicing events are quantified such that the PSI value for each event is calculated. In this case, the PSI measures the fraction of mRNA from a gene that contains a specific form of that event. Additionally, SUPPA2 can also cluster the transcripts or events based on their PSI values in different conditions.

1.3.1.9 Biological interpretation of the results

After a set of genes is identified in an exploratory analysis, enrichment analysis is performed. Widely used tools for identifying significant over- or underrepresentation of annotated Gene Ontology terms (Smith et al., 2003), are Gene Set Enrichment Analysis (Subramanian et al., 2005) and GOrilla: **G**ene **O**ntology en**RI**chment ana**L**ysis and visua**L**is**A**tion tools (Eden et al., 2009, 2007). GOrilla assigns a

p-value to observe a GO term in the ranked list at a non-equiprobable distribution of occurrences, based on the minimum hypergeometric (mHG) score function (Eden et al., 2007). This has the advantage that the gene ranking can be done based on biological data (e.g. log2 fold change in expression levels), rather than defining a threshold (e.g. significantly differentially expressed) and treating every member of the resulting target set equally. Another online annotation platform is the PANTHER (Protein ANalysis THrough Evolutionary Relationships) Classification System (Mi et al., 2019a, 2019b). Functional annotation of genes can also be done by pathway annotation using the Kyoto Encyclopedia of Genes and Genomes (Kanehisa and Goto, 2000), Reactome (Fabregat et al., 2017), or WikiPathways (Kelder et al., 2009).

1.3.2 ChIP-seq data analysis

Chromatin Immunoprecipitation followed by Sequencing (ChIP-seq) is a method for genome-wide profiling of protein-DNA interactions (e.g. transcription factor binding and histone modifications) (Landt et al., 2012). For this thesis, only the transcription factor binding sites are of interest. For the identification of binding sites, cross-linked chromatin is sheared by sonication, and bead-attached antibodies are added to immunoprecipitate the target protein. Then the DNA is unlinked from the protein and the purified DNA is sequenced. It is recommended to include biological replicates in order to ensure reproducibility. Usually, two replicates proved to be sufficient and little information can be gained by adding more replicates (Rozowsky et al., 2009).

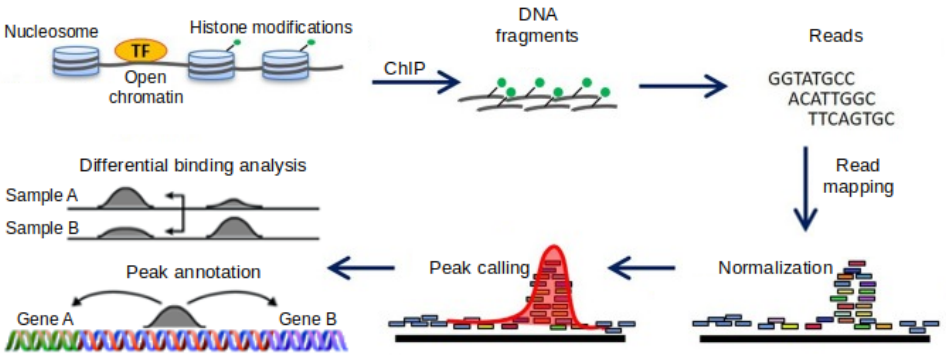


Figure 7: Experimental and data analysis steps for ChIP-seq.
Figure modified from (Nakato and Sakata, 2020) and (Höllbacher et al., 2020).

A typical ChIP-seq data analysis has the following steps: (i) quality assessment of the sequenced reads and read alignment, (ii) normalization, (iii) peak calling, (iv) differential binding analysis, and (v) peak annotation (Figure 7).

Quality assessment can be performed as already described in Chapter 1.3.1.1. Since the aligner does not have to be splice aware, in addition to the ones used for RNA-seq, also the Burrows-Wheeler Aligner can be employed (Li and Durbin, 2009). Another round of quality control is performed after this step, using MultiQC (Ewels et al., 2016). Normalization is performed to account for differences in sequencing depth, by "down-sampling" the reads to the lowest number of uniquely mapped reads. In contrast to RNA-seq, here duplicates are removed for example with Picard MarkDuplicates http://broadinstitute.github.io/picard/. After alignment, peak calling is performed, for which the most widely used algorithm is MACS: Model-based Analysis of ChIP-seq data (Zhang et al., 2008), which is optimized for identifying narrow peaks, mostly in TF binding data (Thomas et al., 2017). Software tools for the next step, i.e. identification of differential binding, were reviewed in (Tu and Shao, 2017). As in the case of DGE analysis in RNA-seq data, DESeq2 is a commonly used tool for this task and is often integrated in packages that were developed specifically for ChIP-seq data analysis, such as DiffBind (Ross-Innes et al., 2012; Stark and Brown, 2020).

Finally, inferring genetic function for these regulatory regions consists of peak annotation. Most commonly peaks are annotated by linear proximity to the closest transcription start site, with the caveat that the three-dimensional character of chromatin is not considered. The Bioconductor R package ChIPpeakAnno (Zhu et al., 2010) is a popular tool for peak annotation.

1.3.3 Metabolomics data analysis

The metabolome is defined as the collection of all metabolites: starting, intermediary, or end products of the metabolic processes in the organism. Metabolites can be produced by the host organism, but they can also derive from other exogenous sources like for example xenobiotics or food (Johnson et al., 2012). Moreover, the metabolome is highly dynamic and varies in time, as metabolites are sensitive to environmental changes. Metabolite concentrations can indicate phenotypic changes and in the clinics, they are routinely checked for diagnosing congenital or acquired diseases like phenylketonuria, diabetes, or even chronic kidney disease (Arneth et al., 2019; Williams et al., 2008).

Metabolomics aims to capture, measure, and characterize low molecular weight compounds (with molecular masses below 1500 Da) in biofluids, cells, and tissues, by using different metabolomics

platforms in the context of external stimuli or in disease states (Nicholson et al., 1999). There are two classical methodologies for metabolite identification: untargeted and targeted metabolomics. In the untargeted approach, the goal is to measure the broadest range of metabolites that is present in the sample in an unbiased way. It presents a global profiling of the metabolome and it is used for example for biomarker discovery (Vinayavekhin and Saghatelian, 2010). In targeted metabolomics *a priori* knowledge is needed and a predefined list of compounds is measured (e.g. amino acids, lipids, sugars, fatty acids), based on the specific question that is of interest (Roberts et al., 2012). Additionally, targeted metabolomics can also be used to validate results from the untargeted approach.

1.3.3.1 Identifying metabolites with differential abundance

There are many similarities between finding differentially expressed genes and differentially abundant metabolites. In both cases measured levels are compared between groups, using a linear model. In the case of DGE the choice of the underlying distribution for representing the data was clearly given by the nonlinear relationship, but still, an assumed direct dependence, between the mean and the variance. In the more commonly used Gaussian distribution the mean and variance are completely independent parameters. Therefore, when it comes to DGE, a standard ANOVA analysis, which assumes normally distributed data is not applicable, instead, a GLM is used.

In contrast, metabolite concentration levels are typically represented as real values (with a logarithmic transformation, thus covering both positive and negative intervals). Thus, the use of a Gaussian distribution is a natural choice, making ANOVA applicable. While there is no proof of normality of a distribution, there are statistical tests to reject the normality assumption and quantile-quantile plots (QQ plots) can be used for visual inspection of the probability density function. Using the R package *car* (Companion to Applied Regression), the statistical testing's confidence interval of the null hypothesis can be visualized around the QQ plot. Moreover, the ANOVA implementation of R computes the Cook's Distances of the residuals to verify the quality of the fit (Cook, 1977).

Nonetheless, using a generalized linear model (GLM) offers specific advantages, even when assuming normally distributed data, as more complex interactions can be tested in a straightforward way. For the post-hoc analysis both Welch's t-test (for the Gaussian distribution case) and Mann-Whitney U test (a nonparametric alternative) can be employed (Sawilowsky, 2005).

1.4 Omics data integration

"Omics" data, such as genomics, transcriptomics, proteomics, metabolomics, etc. are all interconnected: DNA, mRNA, proteins, metabolites, etc. are part of one complex network and they can interact with and have an impact on the other type of molecules. As such, multiple levels of omics data must be analyzed and understood simultaneously.

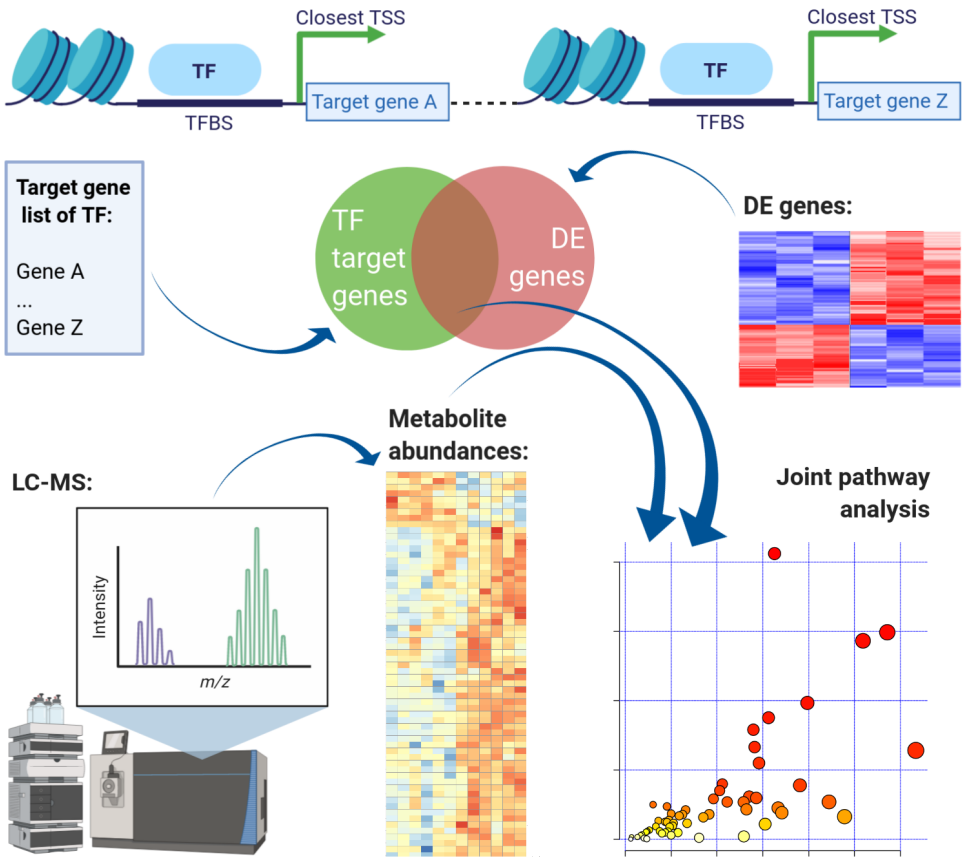


Figure 8: Data integration of ChIP-seq, RNA-seq, and metabolomics.
DE - differentially expressed; LC-MS - liquid chromatography coupled with mass-spectrometry; TSS - transcription start site; TF - transcription factor; TFBS - transcription factor binding site; WT - wild type. This figure was created with BioRender.com

Most commonly omics data integration combines ChIP- and RNA-seq data of matched samples. This way the subset of genes that are possibly direct targets of a transcription factor can be determined. Having the peaks of a transcription factor, one starts annotating the peaks by assigning genes to peaks. Most commonly genes are assigned by linear proximity to the closest transcription start site.

Afterwards, these genes are further screened to check whether they are also differentially expressed. The overlap between these datasets is then visualized in either a contingency table or a Venn diagram (Figure 8).

As metabolites have a wide range of functions, and are considered as a “silent” phenotypical readout of a perturbation, there is a growing motivation to understand their role and integrate metabolite measurements together with other omics datasets. A popular approach is to link metabolite measurements with gene expression data (RNA-seq). Pathway-mapping, i.e. placing metabolites into context with upstream/downstream genes and proteins is wileadly used to facilitate understanding of mechanistic aspects (Figure 8).

An early example of such an approach was the INtegrative Meta-analysis of Expression data (INMEX) tool, which took the position of different compounds in a pathway into account, and assigned weights to them that reflect their importance (Xia et al., 2013). This has evolved into Network Analyst 3.0 (Zhou et al., 2019), a platform for comprehensive analysis of expression data, from raw reads to protein-protein interaction network analysis.

An approach that is more decoupled from *a priori* defined pathways was devised in (Zimmermann et al., 2017). They build a network of the studied compounds by mining the KEGG database for reaction pairs. These connections create different metabolic subnetworks that can be filtered by the number of differentially occurring compounds, including metabolites and proteins.

The online platform MetaboAnalyst (Chong et al., 2019) offers a joint pathway analysis tool. Here the list of differentially expressed genes and the list of differentially expressed metabolites of matched samples can be analyzed and visualized together. Both lists can include their associated log2FCs, allowing an appropriate weighting of the results. Such a joint analysis has the potential to capture causative relations that contribute to complex conditions and disease, as it covers multiple omics levels involved in the regulation of a biological system. Given the complexity of interactions at an organism or even organ level, collecting supporting evidence from multiple omics sources gives more credence to biologically meaningful interpretations.

2. Scope of the thesis

The hypothesis of this Ph.D. thesis is that a high-fat diet and its associated pathological condition alter, in a time-dependent manner, the genome-wide distribution of GR chromatin landscape in the liver, and consequently remodel GR-dependent gene expression.

To explore multiple facets of the glucocorticoid receptor's action in response to a dietary challenge, genotype, time, and ligand availability, the main aims for this project were:

1. Identify diurnal changes in gene expression dependent on GR transcriptional regulation that are affected by the nutritional challenge.

We characterized differentially expressed genes in time and between diets from wild type (WT) and liver-specific GR knock out (GR-LKO) mice after 12 weeks of nutritional challenge (high-fat diet versus control diet). Additionally, as GR activity is tightly connected with circadian rhythms, we analyzed changes in transcripts rhythmicity again in both WT and GR-LKO mice on both diets. By integrating our findings with the results from ChIP-seq data analysis, we were able to identify direct GR target genes around the clock.

2. Characterize the response to GC treatment in the context of diet-induced obesity.

By treating WT mice on both high-fat and control diets with dexamethasone in two different timepoints we could demonstrate, that the increase in DNA-bound GR on HFD was not a ligand-dependent response but a consequence of the nutritional challenge

3. Investigate significant differences in the metabolic profiles of mice under different dietary conditions and genotype.

Using measurements from a targeted metabolomics platform we identified significant changes in the metabolic profiles of WT and GR-LKO mice on both high-fat and control diets at Zeitgeber Time (ZT) 12, where ZT0 and ZT12 correspond to the switch on and off of the light, respectively. Afterwards we coupled our findings with results from the transcriptomic data analysis of matched samples.

4. Characterize the dexamethasone dependent alternative splicing.

As our preliminary proteomics data suggested that GR could be related to splicing as it was pulled down with other core spliceosome members, we hypothesized that dexamethasone induces not only differential gene expression but also influences a differential splicing.

Therefore, we performed differential transcript usage analysis as well as event based analysis in the protein coding mouse liver transcriptome after dexamethasone injection. This way we were able to identify candidate genes that show dexamethasone-dependent transcript usage, that require further experimental validation.

The overall scope of this Ph.D. project is to gain a deeper understanding of the diurnal glucocorticoid hormones response to high-fat diet by multi-omics data integration.

3. Materials and methods

Partial results of this work have been published in (Quagliarini et al., 2019). Reused figures are properly attributed in their caption. The ChIP-seq and RNA-seq datasets were deposited in the GEO (NCBI) database under the SuperSeries accession number GEO: GSE108690.

3.1 Animal experiments

Ethical approval for all animal procedures was received from the regional animal welfare committee of the state of Bavaria (District government of upper Bavaria ROB-55.2-2532.Vet_02-14-33 and AZ_55.2-1-54-2532-158-14) - in accordance with the rules and guidelines established by the Institutional Animal Committee at the Helmholtz Zentrum München – Deutsches Forschungszentrum für Gesundheit und Umwelt (HMGU).

3.1.1 RNA isolation and sequencing

From 50mg of liver the total RNA was isolated using QIAZOL (QIAGEN). The RNA quality was assayed using the Agilent RNA 6000 Nano Kit in a 2100 Bioanalyzer (Agilent). Using 1 mg total RNA, libraries were prepared with the Illumina TruSeq RNA library prep kit v2 chemistry in an automated system (Agilent Bravo liquid handling platform). Libraries were run on a HighSeq4000 sequencer (Illumina).

3.1.2 Description of the used cohorts

For all experiments male C57BL/6J mice were maintained on a 12-h light-dark cycle at 23 ℃ under ad libitum feeding condition. Mice were fed with either a control diet (11 kcal% fat w/sucrose, Research Diets D12329) or a high-fat diet (58 kcal% fat w/sucrose, Research Diets D12331) for a period of 12 weeks. Details on the two diets can be found in the supplementary material (Figure S1), and the cohorts are summarized in Table 1.

Our first cohort consists of 36 wild type (WT) mice split into two equal groups. Half of them were fed control diet, while the other received high-fat diet for 12 weeks in total. At 18 weeks of age, mice were sacrificed by cervical dislocation, and livers were collected at 4 h intervals. It was a long time

course (> 5 time-points) experiment which was sampled uniformly (every 4 hours, starting at 7 am = ZT0), i.e. six time-points in total (ZT0, ZT4, ZT8, ZT12, ZT16, ZT20).

The second cohort consists of hepatocyte-specific Alb-Cre x GRflox/flox (Opherk et al., 2004) mice (liver-specific GR knockout = GR-LKO) and their controls which were subjected to the same dietary protocol. Again, at 18 weeks of age, mice were sacrificed by cervical dislocation, and livers were collected at 4 h intervals (ZT0, ZT4, ZT8, ZT12, ZT16, ZT20). For each diet/genotype/time-point group we had at least 2 biological replicates.

Additionally, we had mice treated with dexamethasone, to study the effect of an acute response to an exogenous GR ligand. In this third cohort wild type mice after 12 weeks of control or a high-fat diet were administered dexamethasone (Sigma, D2915) at ZT0 (when the endogenous cortisone levels are the lowest) and at ZT12 (highest levels) as a single i.p. injection of 10 mg/kg. Mice were sacrificed by cervical dislocation 4 hours later, at ZT4 and ZT16, respectively. We had at least two biological replicates per category.

For the quantitative metabolomics analysis, we used livers from GR-LKO and WT mice fed both diets; animals were sacrificed at ZT12. Frozen liver samples were sent to Biocrates and metabolite measurements were performed in collected liver tissue using the Absolute*IDQ*® p400 HR kit (BIOCRATES Life Sciences AG, Innsbruck, Austria).

Our last cohort used for the study of the dexamethasone-induced differential transcript usage consists of only WT mice on control diet, again injected with dexamethasone at ZT0 and ZT12, respectively. Livers were collected 4 hours later. This cohort is a subset of cohort 3, with additional modifications: we add one more control for the day time-point and analysed 2 dexamethasone-injected replicates instead of three in the night group.

Cohort	I	II	III	IV	V
Nr. of samples	36	67	20	12	10
Diets	CTRL and HFD	CTRL and HFD	CTRL and HFD	CTRL and HFD	CTRL
Genotype	WT	WT and GR-LKO	WT	WT and GR-LKO	WT
Time-points	ZT0, ZT4, ZT8, ZT12, ZT16, ZT20	ZT0, ZT4, ZT8, ZT12, ZT16, ZT20	ZT0→ ZT4; ZT12→ ZT16	ZT12	ZT0→ ZT4; ZT12→ ZT16
Performed analysis	Differential expression analysis and rhythmic transcript detection	Differential expression analysis and rhythmic transcript detection	Differential expression analysis after dexamethasone treatment	Differential metabolite abundance analysis	Differential transcript usage analysis

Table 1: Summary of the five used cohorts.

3.2 RNA-seq data analysis: pipeline for finding differentially expressed genes and rhythmic transcripts

The steps employed for the RNA-seq data processing are shown in Figure 9.

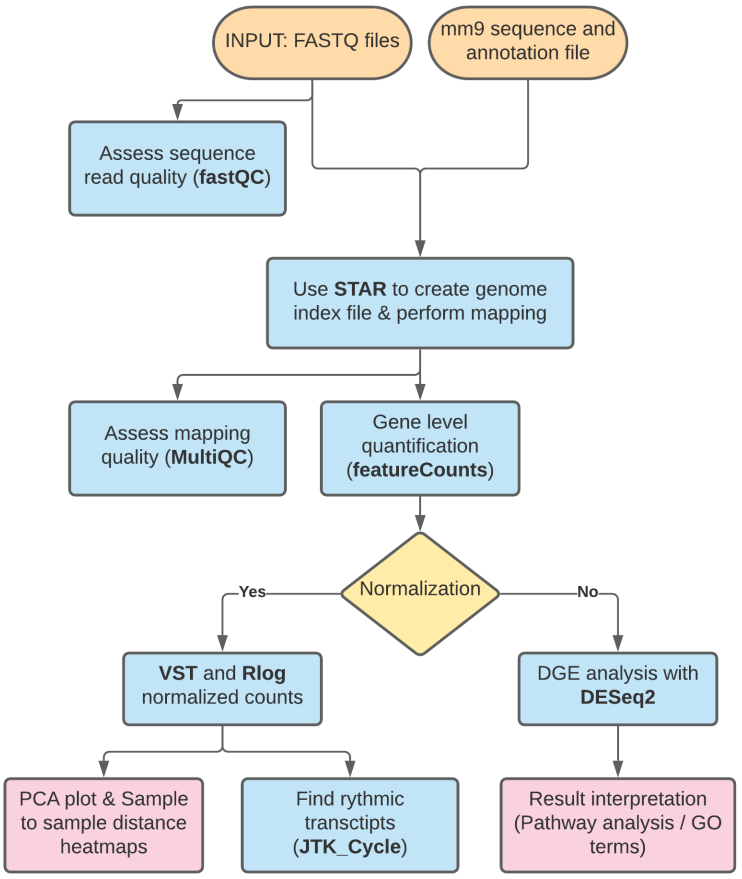


Figure 9: RNA-seq data processing steps for differential gene expression analysis and for rhythmic transcript identification.

3.2.1 Quality check of the sequencing reads

Before mapping the reads to the reference genome, checking the read quality was crucial in order to detect possible problems that could have arisen during sequencing. For that, the tool FastQC v0.11.8 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) was used, which provides an overview of basic quality control metrics for every sample. I checked for example the per base sequence quality, per base sequence content, duplication rate, GC content, etc.

3.2.2 Read mapping and quality assessment of mapped reads

First, the genome index file was generated once for the genome/annotation combination using the GENCODE mm9 reference genome sequences (FASTA file, version number NCBIM37) and the annotation (GTF file, gencode.vM1.annotation.gtf). The index allows to quickly search the genome for possible mapping positions for each read. In a second step, RNA-seq reads were mapped using the splice-aware STAR v2.4.2a aligner (Dobin et al., 2013) as it aligns reads with considerable gaps in between without penalizing the alignment. As we had paired-end data, and sequenced on more lanes the files from all lanes per read1 and analogously per read2 were merged before the mapping. We used the default parameters for running STAR, as these are optimized for mammalian genomes. The output files from the mapping step were sorted, then indexed, and format conversions were performed using Samtools v1.3.1 (Li and Durbin, 2009). Based on the *.Log* output files generated during the mapping step, a report was created with MultiQC v1.9 (Ewels et al., 2016), which provides an overview of the mapping statistics, like the number of mapped reads, percent of uniquely mapped reads, percent multimappers, etc.

3.2.3 Gene level quantification

In a gene-level analysis, genes are considered single transcripts that contain all exons of the gene. For gene-level quantification, the featureCounts v1.4.6 read summarization program (Liao et al., 2014) was used. I counted all reads which were:

- located within an exon
- do not overlap multiple features (because any single fragment must originate from only one of the target genes but the identity of the true target gene cannot be confidently determined)
- are not chimeric, meaning that I discard those fragments that have their two ends aligned to different chromosomes
- have a minimum mapping quality of 4 (MAPQ >= 4, meaning that we exclude multimappers).

Duplicates were not removed, as it was shown that this improves neither accuracy nor precision but can *"actually worsen the power and the False Discovery Rate (FDR) for differential gene expression"* (Parekh et al., 2016). The final output was a count matrix, where the rows correspond to the genes, while the columns represent the different samples.

3.2.4 Filtering and normalization of count data

First of all, I filtered out zero and low count genes, where the mean of the counts in the different samples was below 50. Next, data normalization was performed, which is a crucial step in the workflow, as it makes it possible to account for biological or technical biases.

As previously mentioned the choice of normalization method has a higher impact on the results than the mapping method or the test statistics used for finding differentially expressed genes, therefore deciding what method to use is important. I normalized the data using the variance stabilizing transformation (VST) and also Rlog method, both available through the DESeq2 R package. Both methods enable us to overcome the overdispersion problem, by reducing the dependence of the variance on the mean. Furthermore, the VST method also normalizes with respect to the library size.

The normalized count data was used as input for the JTK Cycle program, which was employed for finding rhythmic transcripts. Additionally, methods applied for exploratory analysis for multidimensional data like principal component analysis (PCA) or sample to sample heatmaps work better on homoscedastic data. When calculating the sample to sample distance, I used the default, Euclidean distance.

3.2.5 Differential gene expression analysis with DESeq2

The DESeq2 (v1.23.10) R package (Love et al., 2014) was used for differential expression analysis. We defined our model as: (i) for the first dataset, where we had only WT mice, 6 time-points and 2 different diets: design(dds) ~ diet + time + time:diet; (ii) for the second cohort, where we had WT and GR-LKO mice on both diets and measurement in 6 different time-points: design (dds) ~ diet + genotype + time + genotype:diet + time:genotype + time:diet, meaning that the expression level of a gene is expected to depend on the diet, genotype and time and combinations thereof. A time series analysis was performed, to find those genes that react in a condition-specific manner over time, compared to a set of baseline samples. For this purpose, the likelihood ratio test (LRT) was used, which compares the difference in the explanatory power of a full model to a reduced model. To obtain the comparisons of interests, Wald tests were performed on the coefficients (and combinations thereof) to contrast different time-point, different diets, and different genotype combinations. This experiment was run in two ways. First with the individual time-points separately (6 in total), and second with the three day time points and the three night time points pulled together. When using DESeq2 for the Dex-treated cohort, we were interested in the diet-dependent treatment effect, and the

used model was: design(dds) ~ diet + treatment + diet:treatment. In this case, the data for the day and night were analyzed separately. Of note, we had to make sure to reset the base level and the significance cutoff. This is important as by default the base level is considered to be the first term in alphabetical order. Also, the default cutoff value for significance is 0.1. We considered only genes below an FDR cutoff of 0.05 to be significant, i.e. differentially expressed. We did not apply a cutoff value for the log2FC.

3.2.6 Identification of oscillating transcripts in time course data with JTK Cycle

Genes whose mRNA levels follow a circadian rhythm with a periodicity of approximately 24 hours are called rhythmic or oscillatory genes. For their detection, RNA-seq was performed at several time points, at small temporal resolution (every 4 hours) over a 24-hour cycle. I used the JTK Cycle v3 (Jonckheere-Terpstra-Kendall = JTK) (Hughes et al., 2010) non-parametric algorithm for cycling transcript detection, which not only distinguishes between rhythmic and non-rhythmic transcript but also reports the period, phase, and amplitude of the oscillating genes. As we were interested in only "purely circadian" transcripts, I set the PERIOD parameter to exactly 24 hours. I used FDR < 0.05 as significance cutoff.

3.2.7 Result interpretation: pathway analysis, GO term enrichment

Enrichment analyses can attribute biological meaning once a gene set of interest has been defined. For this purpose, I checked for significant over- or underrepresentation of annotated gene sets, such as Gene Ontology terms, within the provided gene lists. I used the interactive web-based tool GOrilla that is able to identify enriched GO terms in ranked lists of genes without the need to provide explicit target and background sets (Eden et al., 2009). Besides looking at the overrepresentation of GO terms, gene lists were mapped onto pathways, and statistically significant associations were retrieved. For this purpose, I used the Reactome pathway analysis (Fabregat et al., 2017) through the ReactomePA v3.12 Bioconductor package.

3.3 ChIP-seq data analysis and integration with RNA-seq

ChIP-seq data were processed as described in our recent publication (Quagliarini et al., 2019). Briefly, the pipeline consists of the classical ChIP-seq analysis steps: data quality check, preprocessing, read alignment, peak calling, peak annotation, and motif analysis. Reads were aligned to the mouse mm9

reference genome using BWA-MEM v0.7.12 (Li and Durbin, 2010) and duplicates and multimappers were removed using Samtools and BamTools v2.4 (Barnett et al., 2011). Reads were downsampled to the lowest read count of the replicate. Peaks were called using MACS v2.1.1 (Zhang et al., 2008) with FDR<0.05, and peaks falling in blacklisted peak regions were left out using BEDTools v2.25 (Quinlan and Hall, 2010). The irreproducible discovery rate (IDR) was controlled by comparing peaks between replicates (Li et al., 2011) again using BEDTools. Peak union tables were created for GR, STAT5, PPAR, GR-Dex, and H3K27ac and used for IP efficiency normalization using the THOR package (Allhoff et al., 2016). Peaks were annotated and motif discovery was performed using HOMER v4.8 (Heinz et al., 2010).

As our goal was to identify possible direct targets of the activated GR in the liver (i.e. their expression level changed), we determined the set of genes that were bound and regulated by GR. We checked not only how many and which GR-bound genes were in the set of DEG between the WT and GR-LKO mice, but also identified rhythmic transcripts that were bound by GR. For every GR peak, we searched for the closest transcription start site (TSS) and we considered the TSS-associated genes as being bound by GR.

3.4 RNA-seq data analysis: pipeline for identifying differential transcript usage

The implemented steps for differential transcript usage analysis are presented in Figure 10.

3.4.1 Preprocessing and transcript level quantification

Data preprocessing was performed as described previously in section 3.2.1. For transcript level quantification I used the RSEM v1.3.3 (RNA-seq by Expectation Maximization) Bioconductor package (Li and Dewey, 2011) because it can estimate gene and isoform expression levels in one run. Additionally, it provides the option to generate BAM and Wiggle files in both transcript-coordinates and genomic-coordinates that can be used for visualization. This time the reads were mapped to the mm10 mouse genome and protein-coding transcriptome obtained from GENCODE version vM24. RSEM runs in two spets. First, RSEM was used to build the reference, and for this step sequence and annotation files were needed in FASTA (GRCm38) and GTF file format respectively. Second, the RNA-seq reads were aligned to the reference and based on the alignment, abundances were estimated and credibility intervals were calculated. RSEM by default employs the Bowtie (Langmead et al., 2009) aligner, however, this can be changed. As such, I used the *--star* parameter, which allows

employing the STAR aligner instead of Bowtie.

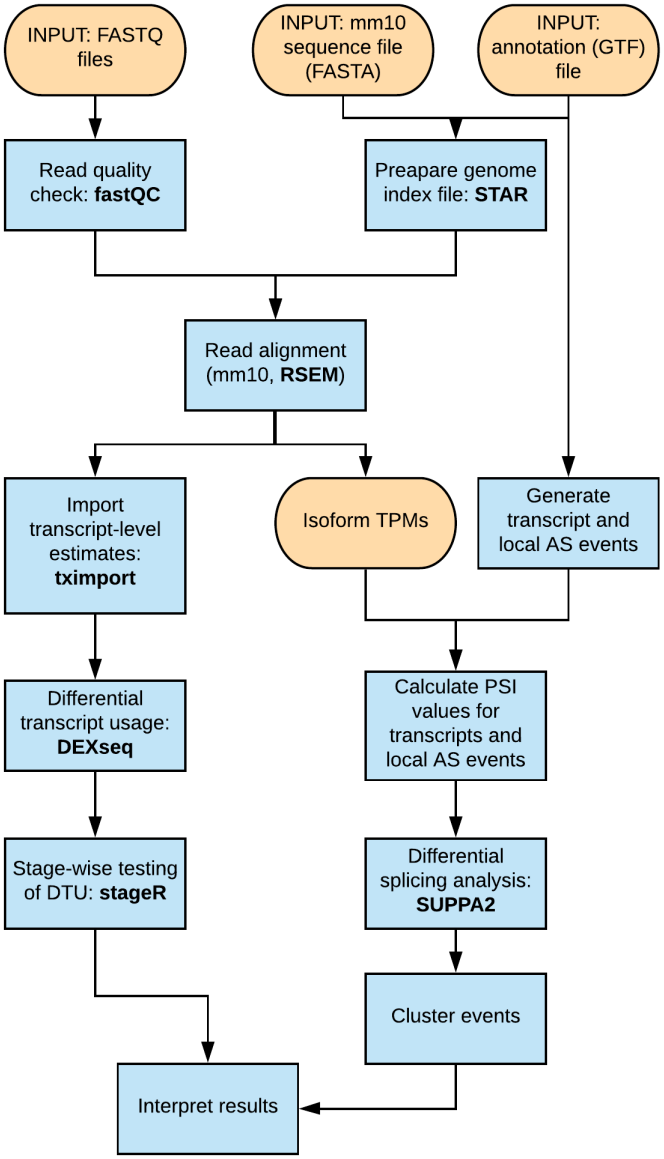


Figure 10: Workflow of the isoform analysis.
On the left: DTU analysis with DEXSeq. On the right: splicing analysis at the transcript isoform or at the local alternative splicing event level with SUPPA2.

3.4.2. Filtering and normalization of abundance estimates

Transcripts with zero abundance estimates were removed from the downstream analysis. I used the tximport v1.14 Bioconductor package (Soneson et al., 2016a) to import transcript-level abundance estimates generated by RSEM. This can be done by setting the type argument to *"rsem"*. As differential transcript usage analysis is a count-based workflow, the developers recommend using the

scaledTPM method to generate the counts from abundance estimates:

```
library(tximport)
txi <- tximport(files, type="rsem", txOut=TRUE, countsFromAbundance="scaledTPM",
geneIdCol="gene_id", txIdCol="transcript_id")
```

This is an intermediary step in which the TPM abundances are scaled such that their sum equals the number of mapped reads. Thus the data become roughly on the scale of counts, enabling the use of the Negative Binomial (NB) distribution to model them in the subsequent steps.

3.4.3 Differential transcript usage analysis

The transcript-level counts can be directly used by packages like DEXSeq (Anders et al., 2012) or DRIMSeq (Nowicka and Robinson, 2016) for the statistical analysis of differential transcript usage (DTU). This transcript-level analysis needs to be integrated with the gene-level detection of DTU. If a gene shows statistically significant evidence for DTU, the identification of the transcripts that confirm the presence of DTU can be identified in a post hoc analysis. I used DEXSeq v1.32 for the DTU analysis, where the applied linear model was: design ~ sample + exon + treatment:exon (in our case exon refers to the provided transcripts). DEXSeq will account for the total gene expression for each sample and transcript, and then detect dex-induced changes in transcript proportions by performing a likelihood ratio test with respect to the reduced model of sample + exon.

For controlling the overall error rate for the gene and transcript level hypothesis testing, I used the stageR v1.8 package (Van den Berge et al., 2017), that applies a target error rate (5% in our case) to control the overall false discovery rate (OFDR) introduced in (Heller et al., 2009).

Complementary to DEXseq, I also used another powerful tool, namely SUPPA2 (v2.3), that is suitable for splicing analysis (Trincado et al., 2018). SUPPA2 was employed to study splicing at the transcript isoform and at the local alternative splicing event level. While identifying differentially occurring splicing events, SUPPA2 tested the following local event types: (i) Skipping Exon (SE); (ii) Alternative 5'/3' Splice Sites (A5/A3); (iii) Mutually Exclusive Exons (MX); (iv) Retained Intron (RI); (v) Alternative First/Last Exons (AF/AL).

First, starting from a GTF annotation file the transcript events and local alternative splicing (AS) events were generated. For this, SUPPA2 uses only the “exon” lines from the GTF file. In a second

step, the percent spliced in (PSI) values for transcript and local alternative splicing events were calculated from the previously generated files and a transcript expression file that holds the transcript abundances. Third, the magnitude of splicing change (for events or transcripts) and its significance across multiple biological conditions was calculated.

DTU analysis was performed with both DEXSeq and SUPPA2 separately for the "day" (ZT0, ZT4, ZT8) and "night" (ZT12, ZT16, ZT20) groups. Afterward, I checked also for overlapping genes and transcripts, between the two groups. In the case of DEXseq, as previously mentioned, a gene might show DTU but when looking at the individual transcripts that contribute to this, only those genes were considered, where both gene and transcript pass the significance cutoff of OFDR < 0.05. The list of genes that show significant Dex-induced DTU was further analyzed. The list of DTU genes that were found by both methods was uploaded to Panther Classification System for functional classification test. Additionally GO biological process annotation was performed with GOrilla (Eden et al., 2009, 2007). As background, all expressed genes, i.e. the ones having non-zero expected counts reported by RSEM, were used.

3.5 Metabolomics data analysis pipeline

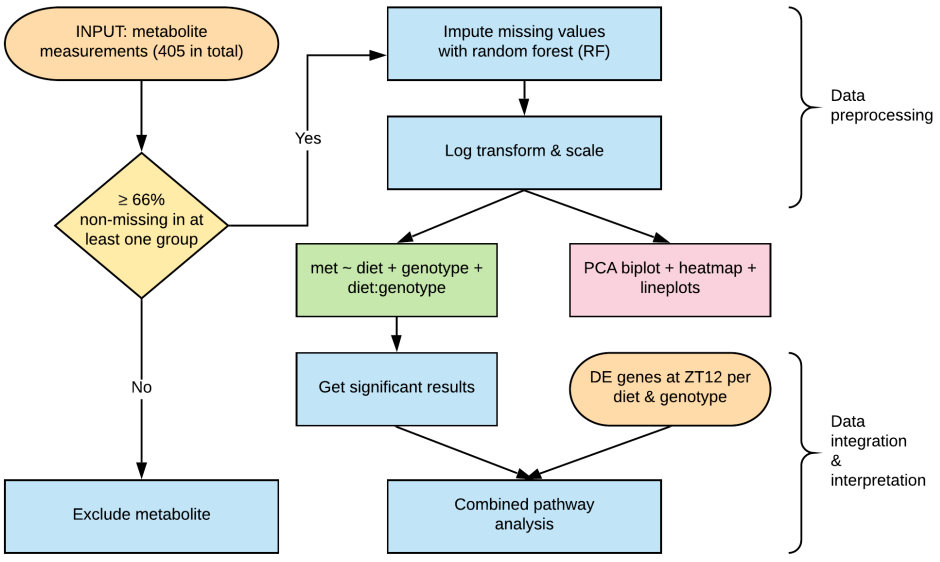


Figure 11: Flowchart of the applied computational steps for the metabolomics data analysis.

Applying a targeted metabolomics approach (using the Absolute*IDQ*® p400 HR kit from Biocrates), we investigated 405 metabolites in WT and GR-LKO mouse livers, after 12 weeks on a high-fat or

control diet. Measurements were taken at ZT12, having three biological replicates per category. This kit is a targeted metabolomics platform that is suitable for broad lipid and metabolic profiling and allows quantification of up to 408 metabolites and lipids from 11 compound classes. The predefined classes are shown in Table 2.

Small molecules	Polar lipids	Neutral lipids
Small molecules	Phosphatidylcholines (172)	Acylcarnitines (55)
Amino acids (21)	Lysophosphatidylcholines (24)	Cholesteryl esters (14)
Biogenic amines (21)	Sphingomyelins (31)	Diglycerides (18)
Hexose (1)	Ceramides (9)	Triglycerides (42)

Table 2: Compound classes of the Absolute*IDQ*® p400 HR kit from Biocrates.
(BIOCRATES Life Sciences AG, Innsbruck, Austria)

The overview of the employed steps for metabolomics data analysis are presented in Figure 11. To ensure data quality, before applying statistical methods, it was necessary to preprocess the data. Typically this includes filtering and missing value imputation. In our case this consists of the following two steps: (i) measurements below the limit of detection (LoD) were considered as "missing"; (ii) considering each of the four groups separately (WT CTRL, KO CTRL, WT HFD, KO HFD), a metabolite was kept if it had a non-missing value in at least 2/3 of the replicates in any of the 4 categories. Those metabolites that failed the above-mentioned criteria were excluded from the analysis. After filtering the MetImp v1.2 web-based tool was used to perform missing value imputation (Wei et al., 2018). Missing values were assumed to be missing at random (or completely at random), therefore the random forest method was applied. The missing not at random (MNAR) case can be verified for example by performing a Chi-square test on the distribution of missing values between the groups, however, in our case, we have too few samples to make this feasible. Nonetheless, the test was performed and the random case could not be rejected.

To check the results, I visualized sample (dis)similarities in a PCA score plot and a sample-to-sample heatmap. I mean normalized each metabolite, using the average of all samples (i.e. for each metabolite, I took the mean metabolite abundance, and divided all replicate metabolite levels with it). Data points outside three times the standard deviation for each metabolite were considered as outliers and were removed from the visualization. Afterward, data was also log-transformed in order to allow good visualization of both small and large values on the same color scale.

In order to assess significant effects, a linear regression model was fit to each metabolite. The applied model was: metabolite_level ~ diet + genotype + diet:genotype. When the interaction term did not obtain a significant effect size it was discarded and the simplified model metabolite_level ~ diet + genotype was employed. To correct for multiple testing, the false discovery rate was controlled at 5 % using the Benjamini–Hochberg procedure. The distribution of metabolite levels in the groups was additionally visualized in heatmaps. For better visualization of between-group differences, the metabolites were clustered based on Euclidean distance and Ward's clustering criterion (Murtagh and Legendre, 2014).

3.5.1 Integrative pathway analysis

The crosstalk between genes and metabolites is of great interest, as changes in metabolite levels are considered not only the phenotypical readouts of gene regulation, but they can in turn influence gene expression. Our goal was to interpret the findings from the metabolomics data analysis by integrating metabolomics and RNA-seq data results. Therefore, we used the web-based tool MetaboAnalyst (Chong et al., 2019), which offers a statistical, functional, and integrative analysis of metabolomics data. I employed the joint pathway analysis that offers exactly this option. Here the user has to provide two lists: one with gene names and their associated fold changes, the second one has to contain the significantly expressed metabolites. Because after the Benjamin-Hochberg correction we had only 29 significant metabolites that were showing either significant diet or genotype effect, for the joint pathway analysis we considered all metabolites that had a significant p-value < 0.05. This analysis was performed separately for the genes/metabolites that show genotype effect at ZT12 and those with significant diet effect. We used the default parameter setting, i.e. for the enrichment analysis we employed the hypergeometric test, for the topology analysis degree centrality, as a pathway database we chose gene-metabolite pathway, and as the integration method, the combined queries approach. The latter avoids the need of combining separate gene- and metabolite-based p-values, and the choice for a weighting method between the two, by treating the two sources in combination and doing a single search and p-value estimation.

3.6 Contribution from collaborators

Mouse experiments were conducted by Dr. Fabiana Quagliarini, Dr. Kenneth A. Dyar and Konstantinos Makris. Samples were sequenced by Elisabeth Graf, Sandy Lösecke and Thomas

Schwarzmayr at the Genomic Core Facility at Helmholtz Zentrum München. ChIP-seq data processing was performed by Dr. Ashfaq Ali Mir. ChIP-MS data was generated by Michael Wierer.

4. Results

It is known that there is a strong connection between a prolonged HFD, perturbation of circadian rhythms, and impaired metabolic function. Perturbations of the clock were linked to metabolic dysfunction, whereas a prolonged HFD feeding could alter the mammalian circadian clock and has detrimental effects on metabolism. Having accumulated evidence that the glucocorticoid receptor plays an important role in circadian biology, we wanted to uncover GR's contribution to diurnal rhythms and characterize how a prolonged HFD reprograms the cistromic glucocorticoid hormone response. Therefore, we studied the circadian hepatic GR cistrome and transcriptome in mice after 12 weeks of high-calorie feeding.

We applied a step-by-step approach: (i) First, we characterized the diet's effect on rhythms and the time-dependent diet effect on gene expression in livers from WT mice. (ii) Second, using livers from GR-LKO mice and their littermates upon control diet, we could assess the genotype effect, i.e. how the loss of GR influences rhythms and gene expression in a time-dependent manner. (iii) Third, in analogous experiments conducted in GR-LKO and WT mice upon HFD, we could detect the combined diet and genotype effects. (iv) Forth, in order to show that the gained GR binding was indeed induced by HFD and did not depend on GR ligand availability, we used dexamethasone-treated cohorts and checked the diet-dependent treatment effect. (v) Fifth, having metabolite measurements at ZT12 of WT and GR-LKO mice on both diets we detected metabolites that have significantly different levels in the two diets or genotypes. By combining these findings with RNA-seq data of matched samples we could identify possible GR-dependent metabolic pathways. (vi) Finally, in a subgroup of our Dex-treated cohort, we were able to identify genes that show differential transcript usage between conditions. Additionally, in steps one to three, we integrated our findings with the ChIP-seq data, meaning that we intersected the list of deregulated transcripts and the list of oscillating transcripts with those genes that harbored a nearby GR peak. Our applied multi-omics data analysis approach served to better understand the glucocorticoid's receptor action.

In the following subsections, the findings from each of the above-mentioned steps are further detailed.

4.1 Characterization of the diet's effect in RNA-seq data from WT mice after 12 weeks of high-fat diet

Our fist RNA-seq was performed in livers of WT mice put on either a control or a high-fat diet for 12 weeks in total. Livers were collected every four hours throughout the day, starting at 7 a.m. = ZT0. We

had 3 biological replicates per time point and diet, i.e. 36 samples in total (Figure 12). Based on this data we were able to assess the effects of a prolonged HFD on the circadian rhythm and on time-dependent gene expression.

Figure 12: Graphical overview of the experiment.

4.1.1 Quality metrics of the raw data and mapping statistics

The quality of the raw data was assessed with the FastQC tool. All samples passed quality control. From the final *.Log* files that are outputted by the STAR aligner (Dobin et al., 2013) I also generated a report with MultiQC (Ewels et al., 2016) that holds the mapping statistics. A good quality sample will have at least 75% of the reads uniquely mapped. As we can see in Table S1 in the supplementary material, this criterion applies to all our samples. The range being between ~76% to 91% confirms high-quality data. Also, the number of sequenced reads is very high, on average over 49 million reads per sample, meaning that we have good coverage.

The PCA (Figure 13), performed to get an overview of how the samples are clustered together based on their similarity, shows that one sample (control diet, time-point 4 = ZT12, replicate 1) is a clear outlier. This happened despite enough sequencing reads (54618150) and a high percent of uniquely mapped reads (91%). The same was confirmed by the sample-to-sample distance heatmap (Figure 14). We decided to discard this one sample from the downstream analysis. As seen on the PCA plot, the first two principal components together captured 21% of the variance, and we can already see a rough division of the samples according to the diet.

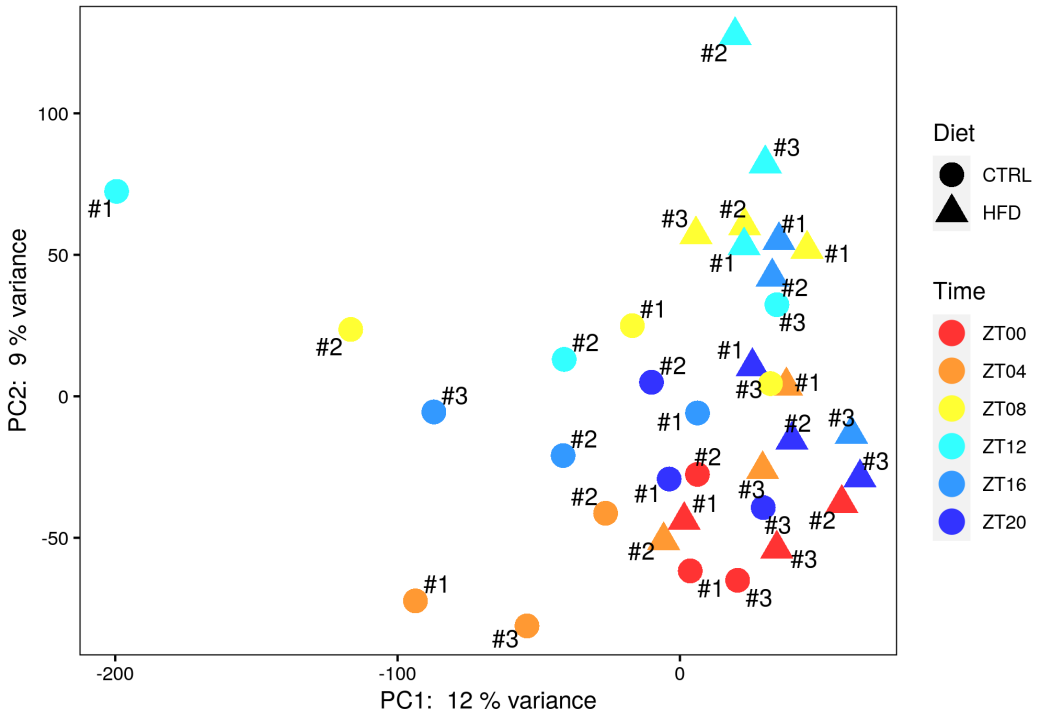


Figure 13: PCA plot based on the VST-normalized counts of WT mice after 12 weeks of nutritional challenge.

Half of the mice were fed control and the other half high-fat diet for 12 weeks in total. Livers were collected every 4 hours throughout the day starting at 7 am = ZT0. We had three biological replicates per time/diet category (marked as #1, #2, #3).

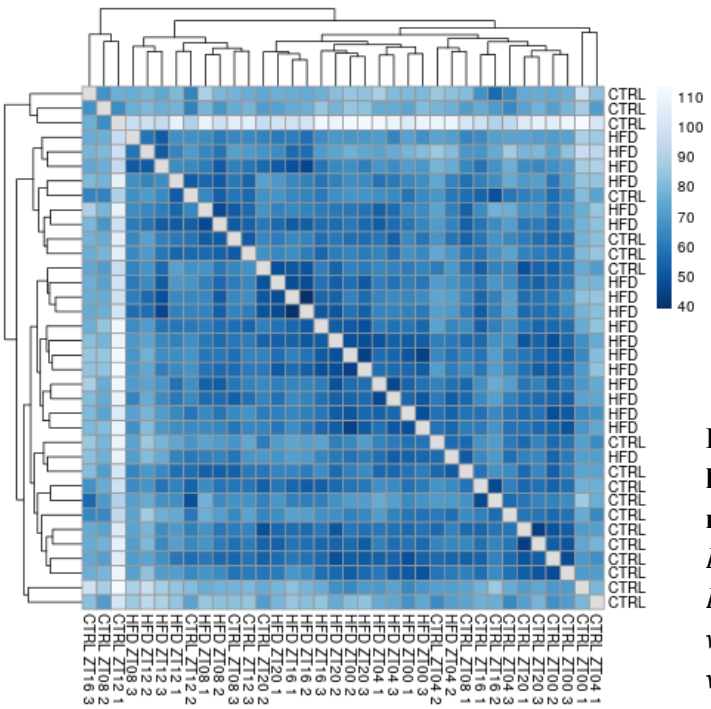


Figure 14: Sample to sample distance heatmap of WT mice after 12 weeks of nutritional challenge.

Based on the VST-normalized counts the Euclidean distance between the samples was calculated and in a heatmap visualized.

4.1.2 Prolonged high-fat diet feeding leads to increased number of oscillating transcripts

To test the hypothesis that a prolonged high-fat diet has a detrimental effect on the molecular clock, we investigated how transcript oscillation changes between the control and high-fat diet fed group. With JTK Cycle (Hughes et al., 2010) we identified the set of circadian genes in both diets (setting the period to exactly 24 hours). Then we calculated the overlap of these gene sets and saw that 275 genes were oscillating only in the control diet group, 675 in both, and strikingly the majority of genes, namely 2207, gained oscillation in the high-fat diet group (Figure 15A). The group of overlapping genes was further characterized to detect diet-specific differences. As shown in Figure 15B, we observe that only 39% of the genes stay in phase, while in 6% HFD induces phase delay. Interestingly, we found that the majority of the overlapping genes, 55%, were phase advanced in the high-fat diet fed group. Of note, we did not detect significant differences in the amplitude for most of these genes. 95% of amplitude changes were smaller than ±0.26, with a mean of -0.005 and standard deviation 0.12.

In the set of genes that preserved their oscillation, we could detect the core clock genes (*Clock, Bmal1, Per1, Per2, Rev-erbα, Rev-erbβ,* and *Rorc*). *Bmal1, Per1, and Clock* were phase advanced in the HFD group, while *Rev-erbα, Rev-erbβ,* and *Clock* displayed a change in amplitude that was more extreme than that of 95% of the genes. While *Rev-erbα*'s and *Rev-erbβ*'s amplitude increased (by 38% and 23%, respectively), *Clock*'s amplitude decreased by roughly half (43%). Two examples are shown in Figure 15C and 15D.

Overall our data suggest that HFD drives rhythmic transcription and induces phase advancement. In addition, we could show that the core clock machinery is very robust and clock genes stay rhythmic in the livers of HFD-fed mice.

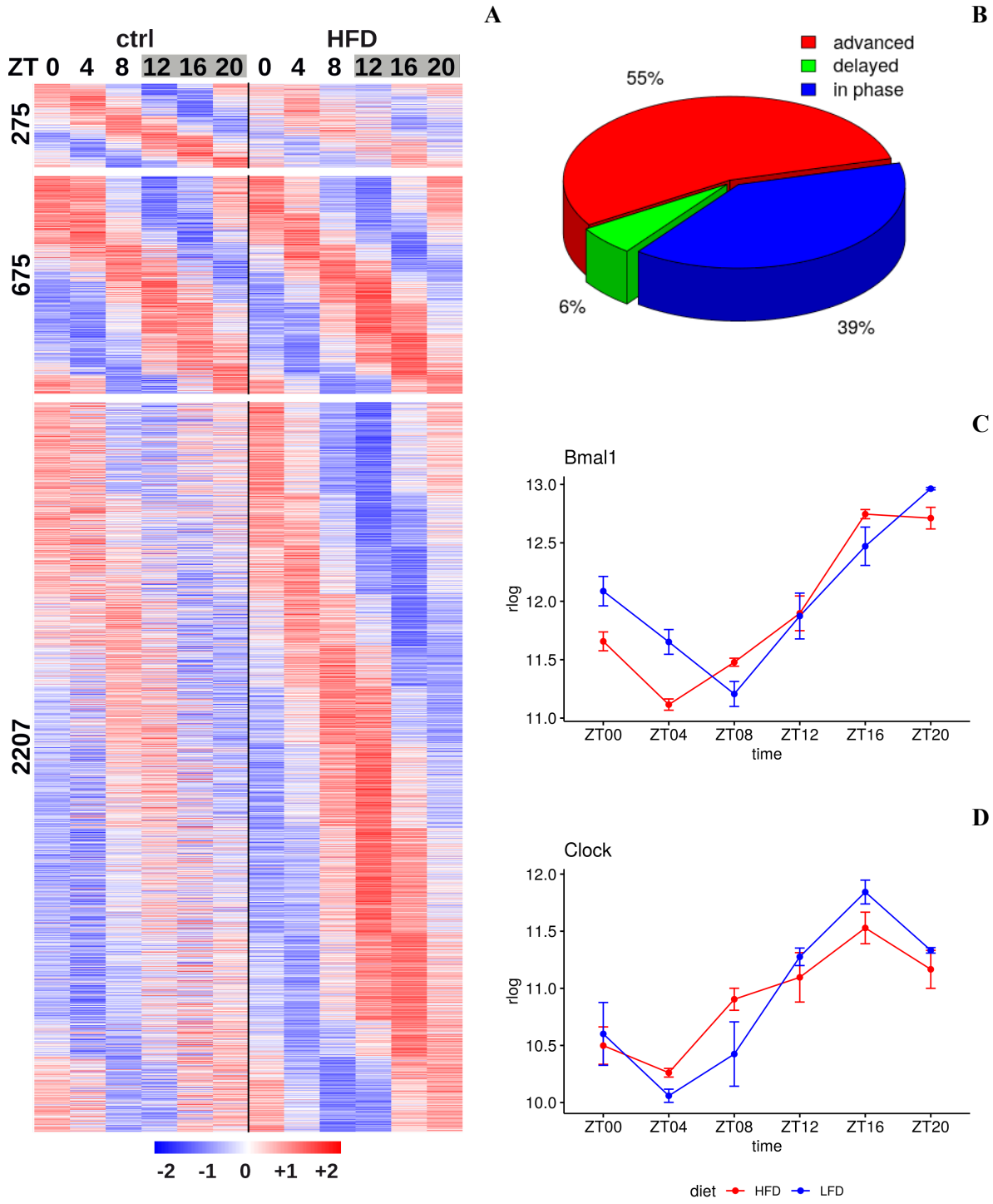


Figure 15: Prolonged HFD-feeding leads to increased number of oscillating transcripts.
(A) Phase sorted heatmap of all oscillating transcripts in liver from WT mice in control diet, both diets, and HFD for ZT0-20. Rhythmicity was assessed with JTK Cycle (period = 24, BH q-value < 0.05, n=3) (B) Pie chart of the phase differences in transcripts that preserve oscillation in both diets. (C&D) Normalized read counts (Rlog) for two rhythmic core clock genes in WT mice fed control (here as LFD) or a high-fat diet.

4.1.3 HFD induces more genes during the night and deregulates hepatic gene expression

To gain more insight into the altered gene expression after prolonged HFD we analyzed the transcriptome of mice to identify the set of genes and the pathways that were disturbed by this nutritional challenge. We mapped the reads to the mm9 mouse genome, and used DESeq2 (Love et al., 2014) for differential expression analysis. We grouped our sample into two categories, day (ZT0, ZT4, and ZT8) and night (ZT12, ZT16, and ZT20), respectively, and performed DGE analysis in each group separately. We defined our model as expression_level ~ diet + time + diet:time, meaning that we expect the expression levels to be influenced by the diet, time, and their interactions. Due to the fact that mice are active and feed during the night, we saw more deregulated genes during the night (1178) than during the day (497), as mice are nocturnal animals. However, almost 50% (213) of genes of the day group were also deregulated during the night (Figure 16).

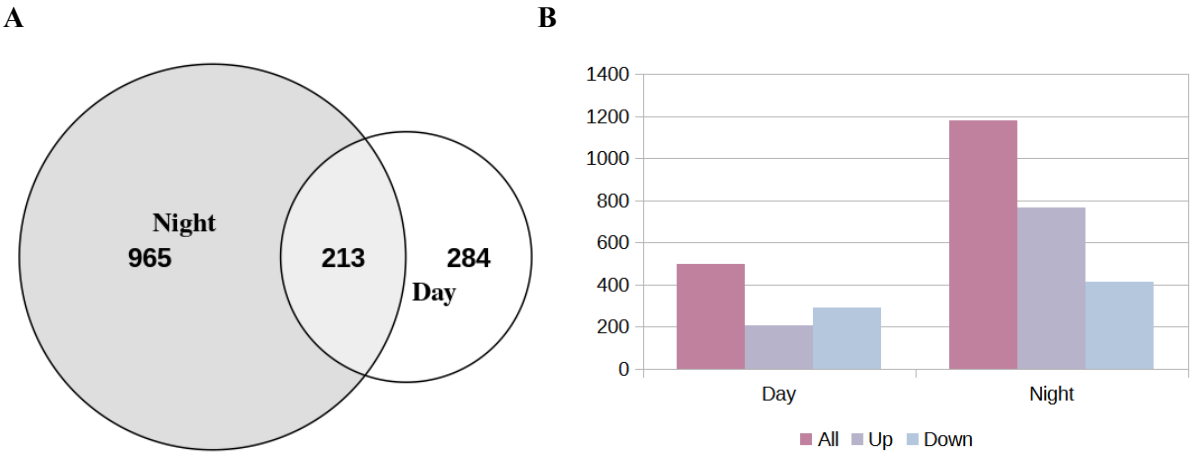


Figure 16: HFD induces more genes during the night.
(A) Venn-diagram showing the number of deregulated genes between control and high-fat diet fed WT mice during the "day" (ZT0, ZT4, and ZT8) and during the "night" (ZT12, ZT16, and ZT20). (B) Bar plot showing the distribution of up- and down-regulated genes during the "day" and during the "night". Overlapping genes are included in both.

The analysis identified 128 up- and 85 downregulated genes that showed a significant effect both during the day and the night. Their effect sizes during the two parts of the day were almost identical (with a Pearson correlation coefficient of 0.98). The results from the Reactome pathway analysis are presented in Figure 17. As shown, the metabolism of amino acids and derivatives was found to be significantly enriched for all upregulated gene subsets (both at day, night, and their overlap).

Additionally, the up-regulated genes unsurprisingly revealed enrichment in fatty acid metabolism, the metabolism of vitamins and cofactors, as well as protein localization. Down-regulated genes were linked to metabolism of carbohydrates and glucose metabolism.

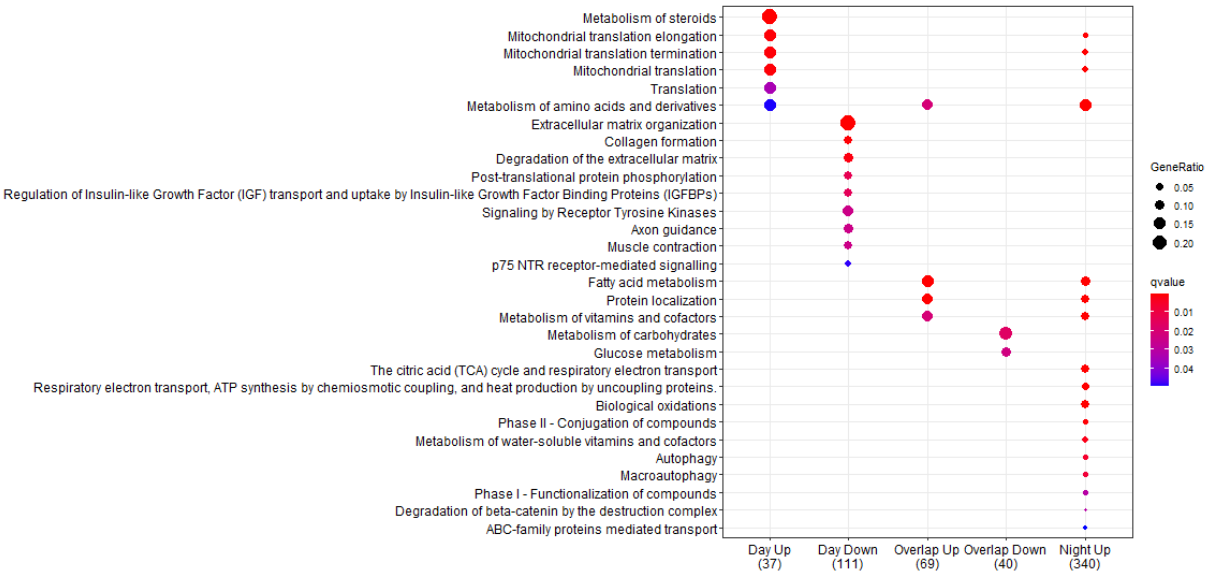


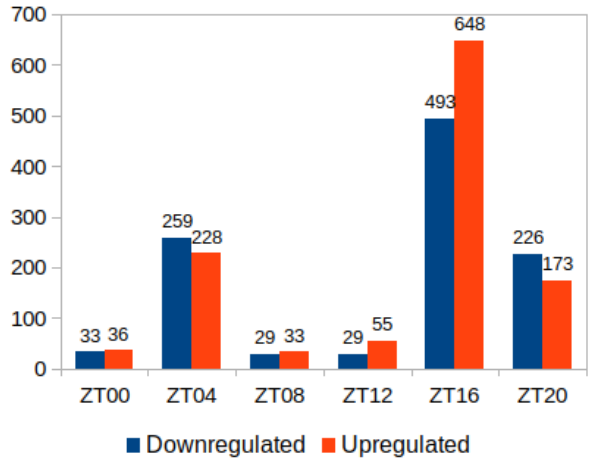
Figure 17: Reactome pathway analysis of up- and down-regulated genes for the "day", "night" and their overlap.

Exclusively during the day, we had an additional 77 upregulated and 207 downregulated genes. Again we performed Reactome pathway analysis on the up- and down-regulated set of genes separately. As shown in Figure 17, activated genes were enriched for pathways involved in the metabolism of steroids, (mitochondrial) translation, and as stated above, the metabolism of amino acids and derivatives.

There were considerably more genes that were deregulated exclusively during the night, as we had 637 activated and 328 repressed genes in this subset of the significant results. The activated genes were enriched for several pathways also associated with the daytime (and the overlapping) gene sets, as well as TCA cycle, biological oxidations, etc. However, the set of repressed genes did not show any significant enrichment in Reactome pathways after FDR correction, therefore they are not present in Figure 17.

We repeated the DGE analysis also per time-point and the results are presented in Figure 18. Of note, the number of deregulated genes does not represent unique genes, as some of them pop up as being

differentially expressed in more than one time-point. The number of activated and repressed genes is roughly the same for each time point. Overall, we observe much more deregulated transcripts during the night time-points. Mice, being nocturnal animals, wake up around ZT12 and have their meal, meaning that this shift towards ZT16, showing the most prominent changes in expression, is normal and expected. After ZT16 the numbers of differentially expressed genes decreased. Interestingly, we could observe a rise at ZT4, presumably because mice were fed ad libitum and had access to food non-stop, meaning that we can not exclude the possibility that they had a "snack" at ZT0, right before they go to sleep.



Description	P-value	FDR q-value
cell cycle process	9.08E-10	4.78E-06
mitotic cell cycle process	1.33E-09	3.51E-06
cell cycle	1.04E-08	1.82E-05
cell division	1.65E-07	2.18E-04
chromosome organization	3.45E-06	3.63E-03

Examples: *Kifc1, Cdca8, Cdt1, Smc2 etc.*

Figure 18: HFD induces more genes during the night time-points.
Bar plot showing the number of up- and downregulated genes for each time-point separately (left). GO biological process annotation of DEGs at ZT16 (right).

GO biological process analysis revealed that downregulated genes at ZT16 were associated with cell cycle process, cell division, and chromosome organization. As insulin stimulates anabolic processes this is in line with downregulation of the mentioned pathways. The same analysis for the upregulated genes found no significant enrichment after FDR correction. However, the best p-value was 5.96E-6, and most of the genes, although not significantly, were enriched for signaling pathways involved mainly in immune response, suggesting that the inflammatory mechanisms are stimulated by HFD. This finding is not surprising as prolonged HFD was linked to inflammation.

Taken together, we found that HFD induces more deregulated transcripts during the night, with metabolism of amino acids and derivatives, and fatty acid metabolism as being most prominently up-regulated.

4.1.4 ChIP and RNA-seq data integration identified HFD-induced GR target genes

We next examined how many of these genes are also bound by GR (gained GR peaks on HFD, from the ChIP-data set, Figure 19). We saw more HFD-induced GR targets during the night than during the day (Figure 20). For the day group almost 30% of the genes (=148) and for the night group, ~42% of the DE genes (=469) were bound by GR. During the inactive phase, HFD-induced GR target genes were associated with down-regulation of metabolism of carbohydrates and glucose, and up-regulation of cholesterol biosynthesis, fatty acid, and lipid metabolism. During the night, when the corticosteroid levels are the highest in nocturnal mice, the differentially expressed genes between the two diets, that harbored a nearby gained GR peak, were again assigned to glucose, amino acid, lipid, and fatty acid metabolism.

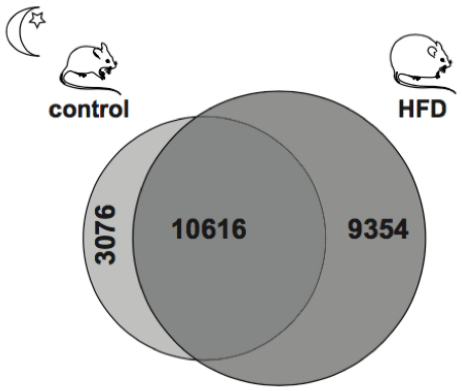


Figure 19: Venn diagram showing the number of GR peaks during the night in the control diet fed and HFD fed mice.
9354 peaks were newly gained in the HFD group. Figure already published in (Quagliarini et al., 2019)

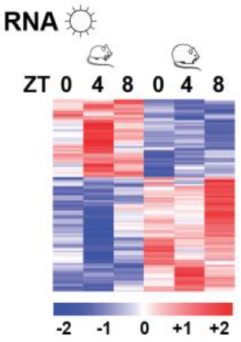
Altogether, our data show that a prolonged HFD reprograms hepatic transcription primarily during the active/feeding phase of the animals. In line with the increased number of GR ChIP peaks during the night in HFD, the number of HFD induced GR targets was also higher during the night.

A

B

Reactome: down-regulated	Adj p-val
interferon signaling	1.8E-03
peptide chain elongation	2.1E-03
metabolism of carbohydrates	3.8E-03
glucose metabolism	4.9E-03
metabolism of amino acids and derivatives	3.3E-02
Reactome: up-regulated	**Adj p-val**
metabolism of lipids and lipoproteins	1.4E-09
fatty acid, triacylglycerol & ketone body metabolism	1.7E-08
metabolism of amino acids and derivatives	2.8E-07
peroxisomal lipid metabolism	8.6E-07
branched-chain amino acid catabolism	4.6E-06
TCA cycle and respiratory electron transport	1.1E-04
mitochondrial fatty acid beta-oxidation	2.7E-04
fatty acyl-CoA biosynthesis	2.3E-03
pentose phosphate pathway	8.9E-03
pyrimidine metabolism	1.3E-02

C



D

Reactome: down-regulated	Adj p-val
metabolism of carbohydrates	4.0E-03
glucose metabolism	8.8E-03
Reactome: up-regulated	**Adj p-val**
cholesterol biosynthesis	4.5E-06
metabolism of lipids and lipoproteins	3.4E-04
fatty acid, triacylglycerol & ketone body metabolism	2.3E-03
PPARα activated gene expression	5.6E-03
metabolism of amino acids and derivatives	1.8E-02
α-linolenic (ω3) & linoleic (ω6) acid metabolism	3.0E-02

Figure 20: Expression of HFD-induced GR targets is more prominent during the night.
(A) Heatmap of transcripts deregulated by HFD during the night (ZT12, ZT16, and ZT20) intersected with the genes associated with the 9354 gained GR binding (see Figure 19). (B) Pathway annotation for transcripts either up- or downregulated in HFD livers. (C) Heatmap of transcripts deregulated by HFD during the day (ZT0,4,8) and associated with gained GR binding. (D) Pathway annotation was performed for transcripts either up- or down-regulated in HFD livers. Differential gene expression in livers from HFD fed and control wildtype mice was calculated by using DESeq2 (n=3, adj p<0.05) as described. Figure already published in (Quagliarini et al., 2019)

4.2 Characterization of the genotype effect in RNA-seq data from WT and GR-LKO mice after 12 weeks on control diet

Integration of our results from the GR ChIP-seq data (Quagliarini et al., 2019) with published liver cistromes for the core clock machinery revealed that the promoter or enhancer regions of all core clock genes harbor GR peaks. To uncover GR's contribution to diurnal rhythms, we profiled mRNA

expression in liver-specific GR knockout mice and their controls throughout the day (24 hours, 4-hour resolution starting at ZT0 = 7 am) after 12 weeks on a control diet (Figure 21).

Figure 21: Graphical overview of the experiment.

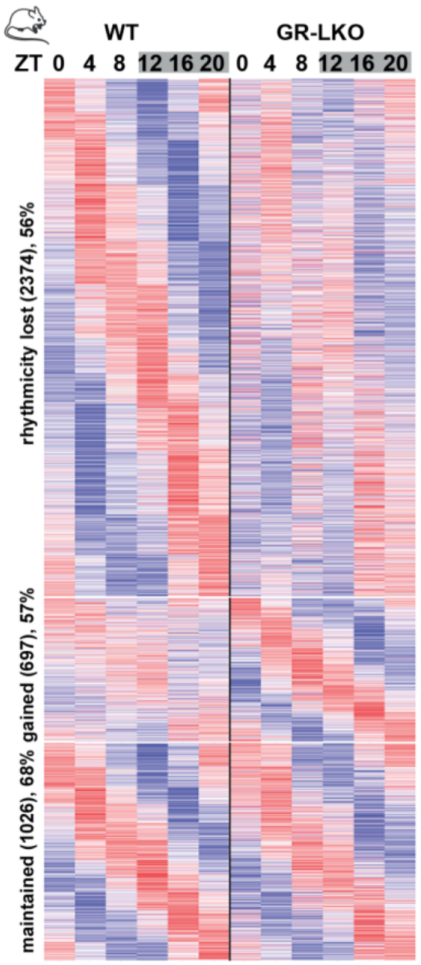
4.2.1 Quality metrics of the raw data and mapping statistics

The quality of the raw data was again assessed with the fastQC tool and all samples passed the quality control. Table S2 contains the mapping statistics and can be found in the supplement. As presented, all samples had high percent of uniquely mapped reads (>73%). This time the average number of reads per sample was 16 million.

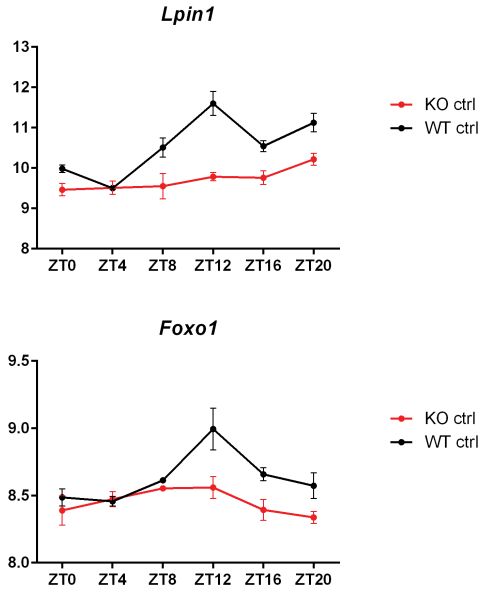
4.2.2 Loss of GR alters rhythmicity, but core clock factors maintain oscillation

Similarly to the previous cohort, where we tested if the HFD influences gene oscillation, here we tested for the effect of the genotype, i.e. what is the effect of GR loss on circadian rhythm. Having RNA-seq data from WT and liver-specific GR knockout mice on a control diet, using the JTK Cycle program, I was able to identify purely circadian genes, i.e. those that show a 24-hour oscillation. We detected oscillating genes in each group separately. When comparing the two lists of rhythmic genes in WT and GR-LKO, I could see that 1026 genes preserved, 2374 lost, and 697 gained oscillation in the absence of GR (Figure 22A). Reactome pathway annotation revealed that genes that lost oscillation in the absence of GR were mainly associated with metabolism of lipids and metabolism of amino acids and derivatives (Figure 22C). Two examples of genes that lose oscillation are *Lpin1* and *Foxo1*. This is not surprising as both are known GR targets and their phase coincides with the peak of glucocorticoids. *Lpin1* is involved in triglyceride synthesis and the transcription factor *Foxo1*, plays a critical role in hepatic glucose and lipid metabolism (Figure 22B) (Matsumoto et al., 2006; Reue and Zhang, 2008).

A



B



C

Reactome pathway analysis	Adjusted P-value
Metabolism of lipids	2.08E-04
Metabolism of amino acids and derivatives	2.20E-03
Cellular responses to external stimuli	2.20E-03
Biological oxidations	2.30E-03
MAP kinase activation	6.20E-03
Toll Like Receptor Cascade	1.12E-02
Cellular responses to stress	1.12E-02

Figure 22: Loss of GR alters rhythmicity.
(A) Phase sorted heatmap of all oscillating transcripts in livers from WT and GR-LKO mice throughout the day (ZT0, ZT4, ZT8, ZT12, ZT16, ZT20). Plotted are the VST normalized counts, averaged over replicates and normalized per row. Rhythmicity was determined using JTK Cycle (period 24, adj p<0.05, n=3). (B) Example of genes that lose oscillation. Plotted are the Rlog values (normalized counts) over time in the WT and GR-LKO (n=3). (C) Reactome pathway analysis of genes that lose oscillation in the absence of GR (FDR < 0.05). Figure (22A) already published in (Quagliarini et al., 2019)

In the group of more than 1000 genes that were oscillating in both genotypes, I was able to identify all members of the core clock machinery, meaning that the circadian clock remains intact despite the loss of GR (Figure 23). Only Per1, another GR target gene, showed an amplitude dampening in the absence of GR.

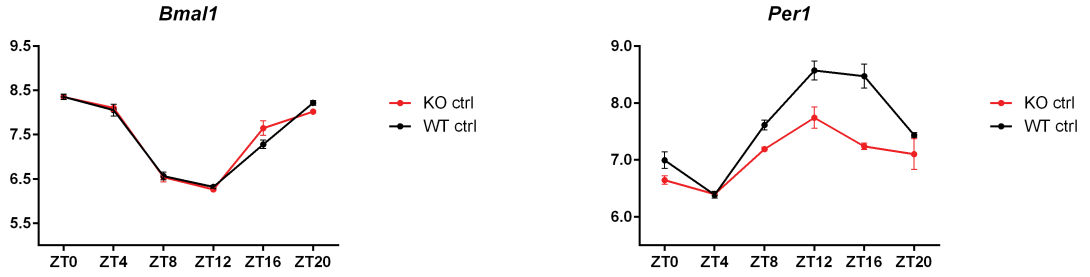


Figure 23: Core clock not affected by GR loss, except Per1.
Example of two clock genes. Plotted are the Rlog values (normalized counts) over time in WT and GR-LKO (n=3).

Reactome pathway annotation revealed that genes that preserved their rhythm in both genotypes were involved in the following pathways: metabolism of lipids and lipoproteins, circadian clock, fatty acid, triacylglycerol, and ketone body metabolism, and regulation of cholesterol biosynthesis by SREBP.

4.2.3 Majority of oscillating genes are bound by GR

By intersecting the lists of rhythmic transcripts with the list of genes that harbor a nearby GR peak (peak list again from the ChIP-seq result) we saw that the majority of oscillating genes are bound and regulated by GR. More precisely, 56.4% of genes that lost oscillation and 68.5 % of those that maintain rhythmicity (Figure 24), were GR targets.

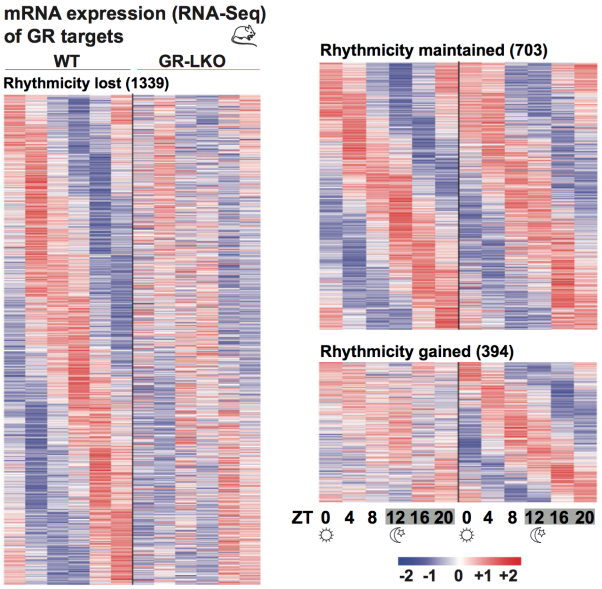


Figure 24: The majority of the oscillating transcripts are bound by GR.
Phase-sorted heatmap of oscillating transcripts in wild type (control) and GR-LKO livers for ZT0-20 of GR targets. Rhythmicity was determined using JTK Cycle (period 24, adj p<0.05, n=3). Figure already published in (Quagliarini et al., 2019)

4.2.4 Loss of GR causes amplitude dampening

As previously mentioned JTK Cycle outputs not only the list of rhythmic genes and their associated significance value but also the phase, period and amplitude. Therefore, I checked how the amplitude is affected by loss of GR over time. As seen in Figure 25, we found that the absence of GR in the liver influences the amplitude of the oscillating transcripts. More exactly, we could observe a significant amplitude dampening of rhythmic gene expression across all six time-points.

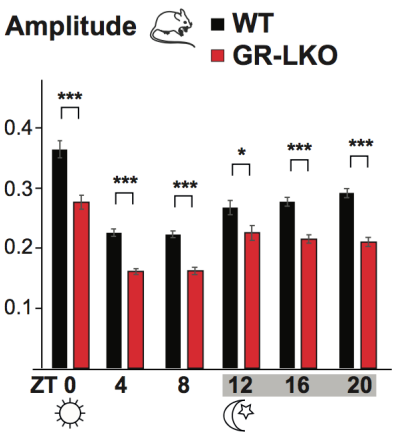
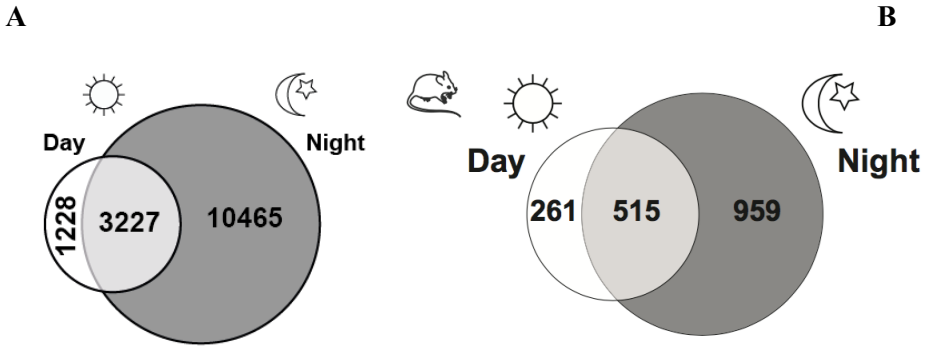


Figure 25: Loss of GR causes amplitude dampening across all time-points.
*Distribution of the amplitude over time in livers from GR-LKO compared to their WT controls. 3,400 genes cycling in control mice were binned according to peak time point (4 h). Values are represented as mean ± SEM (n = 3 per group). *p < 0.05, ***p < 0.001 (two-tailed t-test). Figure already published in (Quagliarini et al., 2019)*

4.2.5 More differential expression at night

To further characterize the impact of GR loss, the DESeq2 Bioconductor R package was used to detect differentially expressed genes during the day (ZT0, ZT4, ZT12) and during the night (ZT12, ZT16, ZT20) between the WT and GR-LKO both on a control diet. Our findings were in line with the results from the ChIP-seq, where we saw that GR genomic occupancy was higher during the night, meaning that the ChIP-seq signal reflected the endogenous ligand availability (Figure 26A) Accordingly, we detected more differential gene regulation during the night (1474) than during the day (776) (Figure 26B). Interestingly, the majority (two-thirds, 515) of the transcripts deregulated during the day were also deregulated during the night. We performed functional annotation of the up- and down-regulated gene lists for both “day” and “night” groups. Genes involved in glucose metabolism, gluconeogenesis, and synthesis of bile acids and bile salts were downregulated, while triglyceride biosynthesis, fatty acid, and lipid metabolism were upregulated at night. The differentially expressed genes during the day were assigned to down-regulation of the urea cycle and glucocorticoid

biosynthesis and up-regulation of cell cycle, DNA replication metabolism of proteins.



C

Reactome: down-regulated	Adj p-val
urea cycle	5.6E-03
regulation of complement cascade	1.1E-02
glucocorticoid biosynthesis	3.0E-02
metabolism of amino acids and derivatives	4.3E-02
Reactome: up-regulated	**Adj p-val**
cell cycle	1.7E-06
DNA replication	2.2E-06
nonsense mediated decay	8.8E-06
infectious disease	3.0E-04
translation	4.0E-04
metabolism of proteins	8.1E-03
glutathione conjugation	2.3E-02

Reactome: down-regulated	Adj p-val
glucose metabolism	6.8E-02
retinoid metabolism and transport	6.8E-02
gluconeogenesis	7.5E-02
lipoprotein metabolism	7.5E-02
metabolism of fat soluble vitamins	7.5E-02
synthesis of bile acids and bile salts	1.0E-01
Reactome: up-regulated	**Adj p-val**
triglyceride biosynthesis	5.0E-06
fatty acid, triacylglycerol, and ketone body metabolism	5.1E-05
metabolism of lipids and lipoproteins	1.3E-04
extracellular matrix organization	1.3E-04
cell cycle	1.5E-04
DNA repair	2.1E-03
glycerophospholipid biosynthesis	7.0E-03
collagen formation	7.2E-03

Figure 26: More differentially expressed genes during the night than during the day between WT and GR-LKO on control diet.
(A) Venn diagram showing the number of GR peaks during the day (ZT0, ZT4, and ZT8) and night (ZT12, ZT16, and ZT20) (based on data from ChIP experiments). (B) Venn diagram illustrating the number of differentially regulated genes in GR-LKO during the day (ZT0, ZT4, and ZT8) and night (ZT12, ZT16, and ZT20). Pathway annotation was performed separately for transcripts either up- or downregulated in GR-LKO. Figures already published in (Quagliarini et al., 2019)

Differential gene expression analysis was also performed for each time-point separately. As seen in Figure 27, there were more genes deregulated during the night time points than during the day time points. The highest number of differentially expressed genes was found again at ZT16, 4 hours after the beginning of the active/feeding phase of the animals and zenith of corticosterone levels. At ZT4 and ZT8 These numbers drop and reach the lowest amount during the day.

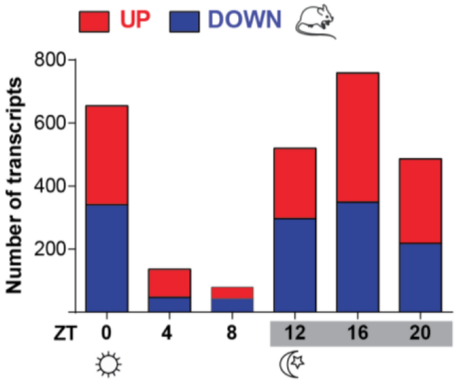


Figure 27: More differentially expressed genes during the night time-points between WT and GR-LKO on control diet.
Distribution of up and downregulated transcripts over time. Differential gene expression between GR-LKO and littermates was calculated per time point with DESeq2 (n=3, adj p<0.05). Figure already published in (Quagliarini et al., 2019)

4.3 Characterization of the diet and genotype effect based on RNA-seq data from WT and GR-LKO mice after 12 week on HFD

In the previous two datasets, we were able to characterize first the diet (HFD) and then the genotype (WT vs. GR-LKO) effect on rhythmicity and checked their time-dependent effect on gene expression. As our main goal was to understand how GR transcriptional regulation is affected by a long term nutritional challenge, we performed analogous RNA-seq experiments in mouse livers of hepatocyte-specific GR knockouts and their littermates after 12 weeks on a high-fat diet. Again livers were collected in biological triplicates, in 6 time points (ZT0, ZT4, ZT8, ZT12, ZT16, ZT20) (Figure 28).

Figure 28: Graphical overview of the experiment.

4.3.1 GR loss causes amplitude dampening also in HFD

By analyzing changes in transcripts rhythmicity after 12 weeks on a high-fat diet with JTK Cycle, we again could see that hepatic GR depletion caused significant transcriptional dampening of circadian amplitude across all six time-points (Figure 29). Almost the same number of genes lost and maintained rhythmicity in the absence of GR in the HFD samples (Figure 30A). These observed changes occurred despite similar mRNA expression of the core clock machinery itself (except for *Per1*) (Figure 30B).

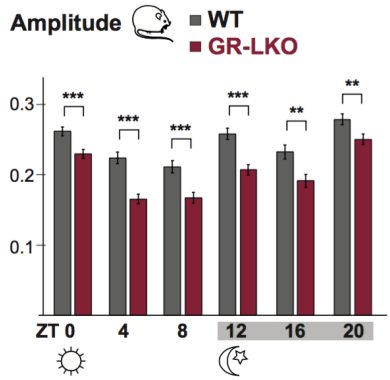


Figure 29: GR loss causes amplitude dampening also in HFD across all six time-points.
Amplitude distribution for ZT0–ZT20 in livers from GR-LKO compared to WT controls on HFD. Values are represented as mean ± SEM (n = 3 per group). Figure already published in (Quagliarini et al., 2019)

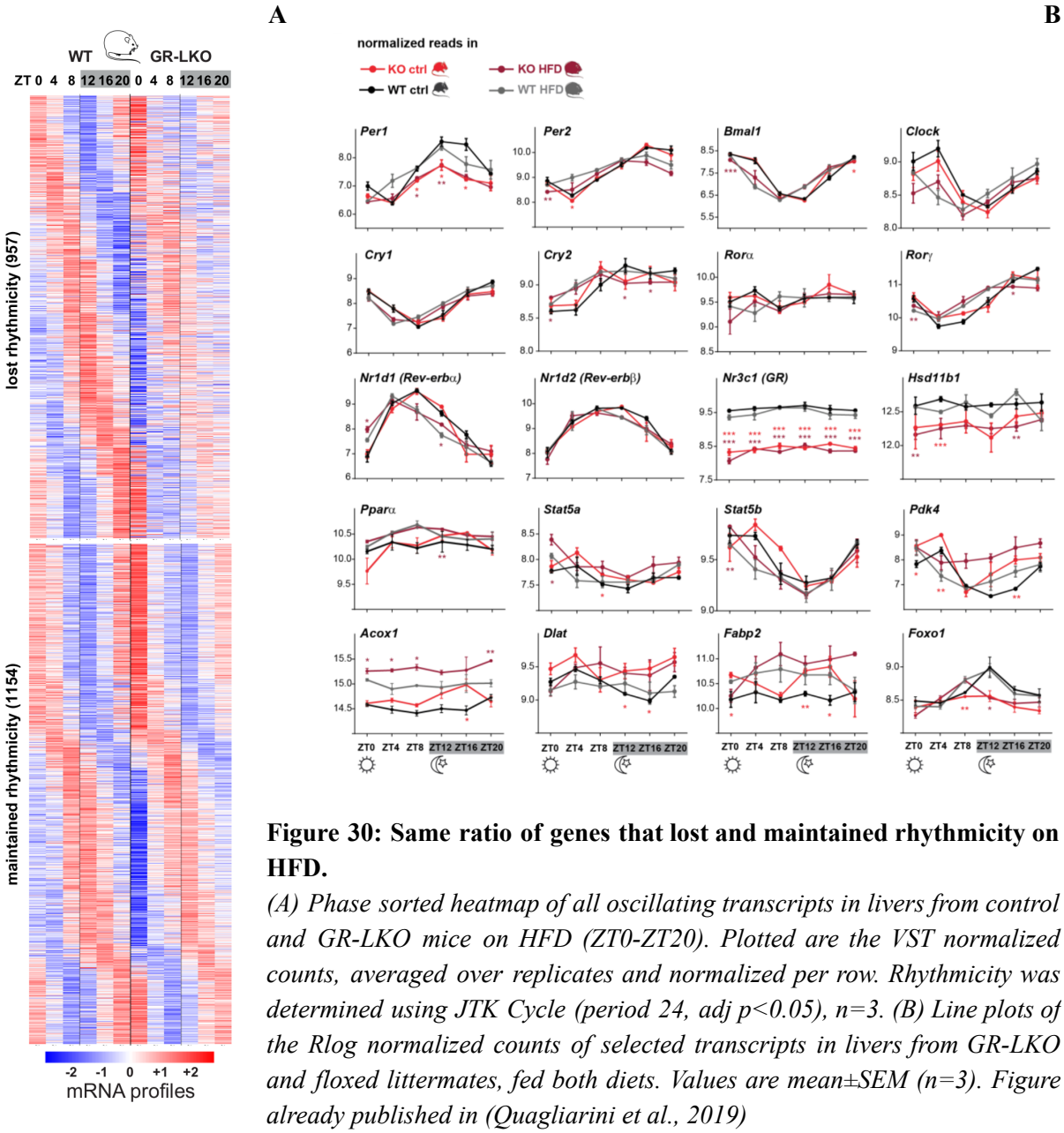
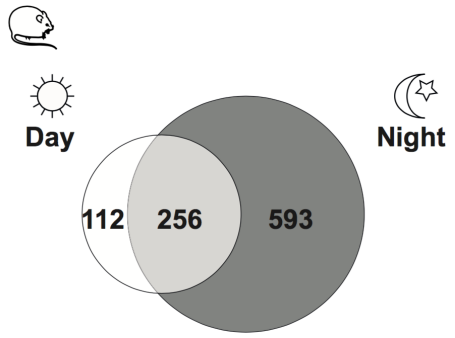


Figure 30: Same ratio of genes that lost and maintained rhythmicity on HFD.
(A) Phase sorted heatmap of all oscillating transcripts in livers from control and GR-LKO mice on HFD (ZT0-ZT20). Plotted are the VST normalized counts, averaged over replicates and normalized per row. Rhythmicity was determined using JTK Cycle (period 24, adj p<0.05), n=3. (B) Line plots of the Rlog normalized counts of selected transcripts in livers from GR-LKO and floxed littermates, fed both diets. Values are mean±SEM (n=3). Figure already published in (Quagliarini et al., 2019)

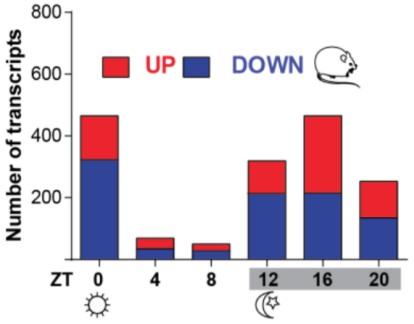
4.3.2 HFD-induced binding is reflected in the transcriptome

Differential gene expression was repeated again for the day and night groups separately. Once more, we observed the same trend as in the control diet-fed group, that there were more deregulated transcripts in GR-LKO during the night than during the day (Figures 31A). The analysis was repeated also for each time-point separately (Figure 31B). The differentially expressed genes in the night group were enriched for pathways including lipid, fatty acid, lipoprotein, and amino acid metabolism. In the day group, the up-regulated genes could be associated with metabolism of lipids and lipoproteins, fatty acid, lipid, and carbohydrate metabolism, while the down-regulated genes were linked to urea cycle, lipoprotein metabolism, steroid hormones, and amino acid metabolism.

A



B



C

Reactome: down-regulated	Adj p-val
urea cycle	5.1E-06
regulation of complement cascade	5.9E-03
metabolism of fat-soluble vitamins	5.9E-03
lipoprotein metabolism	1.2E-02
steroid hormones	1.4E-02
metabolism of amino acids and derivatives	2.5E-02
biological oxidations	4.5E-02

Reactome: up-regulated	Adj p-val
metabolism of lipis and lipoproteins	1.8E-01
synthesis of 12-eicosatetraenoic acid derivatives	1.8E-01
synthesis of PE	5.2E-01
hyaluronan metabolism	7.0E-01
fatty acid, triacylglycerol & ketone body metabolism	7.0E-01
metabolism of carbohydrates	7.0E-01
synthesis of DNA	7.0E-01
post-translational protein modification	7.0E-01

Reactome: down-regulated	Adj p-val
lipoprotein metabolism	1.5E-04
sulfur amino acid metabolism	5.6E-02
metabolism of fat-soluble vitamins	8.3E-02
phospholipase C-mediated cascade FGFR	1.9E-01
metabolism of amino acids and derivatives	1.9E-01

Reactome: up-regulated	Adj p-val
metabolism of lipids and lipoproteins	8.9E-05
fatty acid, triacylglycerol, and ketone body metabolism	2.5E-04
DNA strand elongation	9.9E-02
synthesis of ketone bodies	2.4E-01
pyruvate metabolism and citric acid (TCA) cycle	2.7E-01

Figure 31: More differentially expressed genes during the night between WT and GR-LKO after 12 weeks of HFD.

(A) Venn diagram showing the number of differentially regulated genes in GR-LKO during the day (ZT0, ZT4, and ZT8) and night (ZT12, ZT16, and ZT20). (B) Distribution of up and downregulated transcripts per time-point. Differential gene expression between GR-LKO and littermates was calculated using DESeq2 for each time point separately (n=3, adj p<0.05). (C) Pathway annotation for transcripts either up- or down-regulated in GR-LKO during the day and during the night as defined in (A). Figure already published in (Quagliarini et al., 2019)

4.3.3 HFD-induced binding is functionally linked to gene expression

As we were interested if the HFD-induced binding events were also functionally relevant, we checked the differential expression of the genes which harbored a nearby gained GR peak during the night. We found that the absence of GR caused deregulation in over 400 gained targets. These genes could be linked mainly to upregulation of lipid and fatty acid metabolism. The expression of gluconeogenic genes like *Pck1* and *Pfkfb3* was reduced, while *Cd36* and *PPARγ*, both known for their role in lipid and fatty acid utilization and storage, showed increased expression (Figure 32).

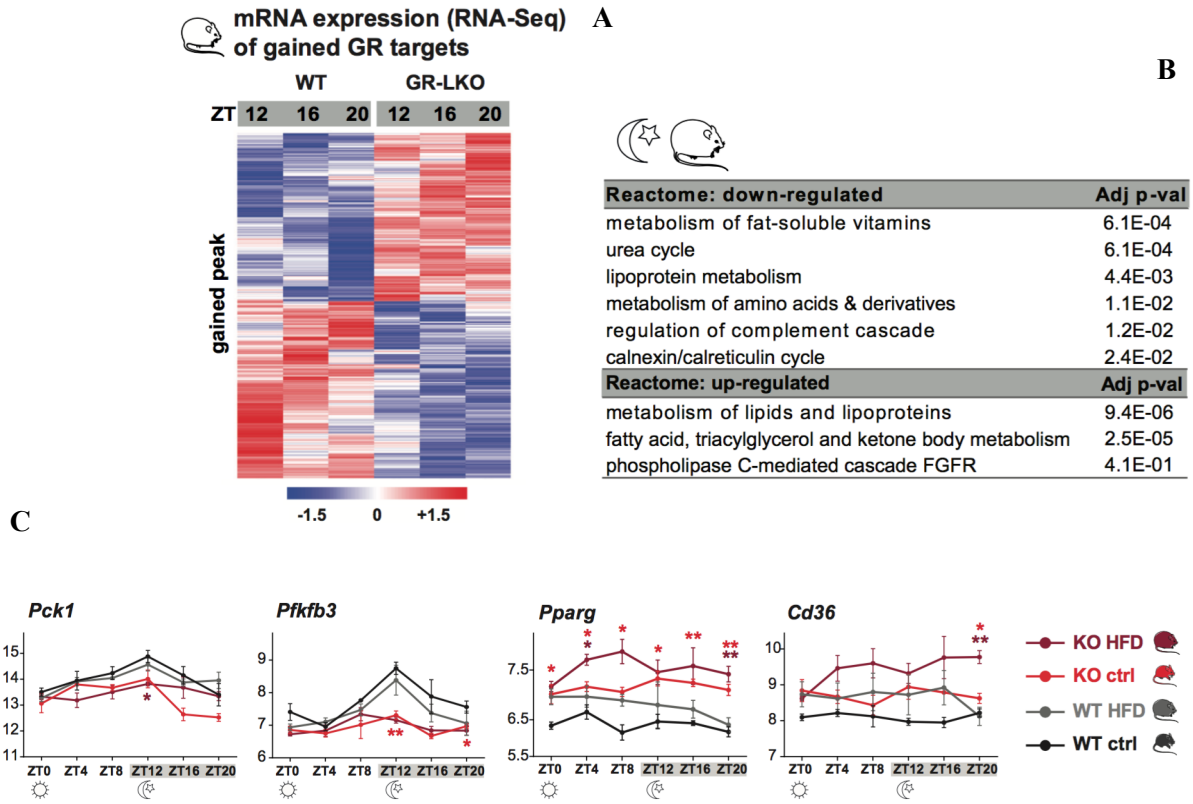


Figure 32: HFD-induced binding is functionally linked to gene expression.
*(A) Heatmap of differentially expressed transcripts in GR-LKO during the night (ZT12, 16, 20) associated with "gained" (9,354) GR peaks from Figure 19. (B) Pathway annotation of the transcripts either up- or downregulated in GR-LKO livers. (C) Examples of HFD-induced genes. Plotted are the Rlog normalized counts for deregulated gluconeogenic (Pck1 and Pfkfb3) and lipid metabolism (PPARγ and Cd36) transcripts in GR-LKO mice and controls fed both diets. Values are represented as mean ± SEM (n = 3 per group); *p < 0.05, **p < 0.01, ***p < 0.001 (two-tailed t-test). Figure already published in (Quagliarini et al., 2019)*

4.4 Characterization of RNA-seq data from dexamethasone-treated WT mice after 12 weeks of nutritional challenge

To further demonstrate that the increase in DNA-bound GR was a consequence of the nutritional challenge or a related pathogenic effect and not dependent on the GR ligand availability, mice on both diets were injected with a single dose of dexamethasone, an exogenous GR ligand. The Dex-treatment was administered in two time points, at ZT0 (day, lowest endogenous GC levels) and ZT12 (night, highest endogenous GC levels) respectively, and for ChIP-seq livers were harvested one, and for RNA-seq 4 hours later (Figure 33).

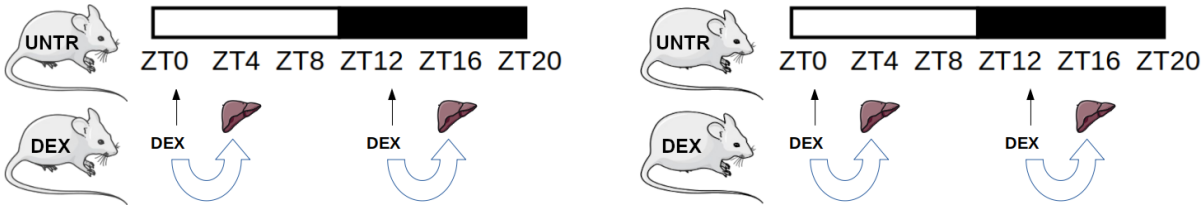


Figure 33: Graphical representation of the experiment.
WT mice on both diets were injected with dexamethasone (an exogenous GR agonist) in two time-points (ZT0 and ZT12), and livers were collected 4 hours later.

4.4.1 Quality metrics of the raw data and mapping statistics

Once again all samples passed the quality check. The mapping statistics are listed in Table S3 in the supplement. To make sure that the Dex-treatment was successful, we first compared the injected mice on both diets with their untreated littermates. The PCA was performed separately for the two groups, day and night respectively. If we look at the group treated during the day with dexamethasone, we can see that the first principal component, which captures 36% of the variance clearly separates the groups by treatment. The second principal component with 24% separates them by diet (Figure 34). The sample to sample distance heatmap confirmed the separation into 4 groups (seen along the diagonal of the matrix) (Figure 34). When looking at the group treated during the night (when mice woke up), we again observe similar trends. There is clear separation based on the treatment, however, when it comes to diet, although PC2 captures again 24% of the variance and the low-fat diet group tends to be positioned lower, and this separation is not so clear (Figure 35). This separation by the diet is more evident at ZT4 when endogenous corticosterone levels are low. The two treatment-based clusters can

be clearly recognized in the sample to sample distance heatmap, where we can distinguish the two squares in the top left and bottom right part of the matrix (Figure 35).

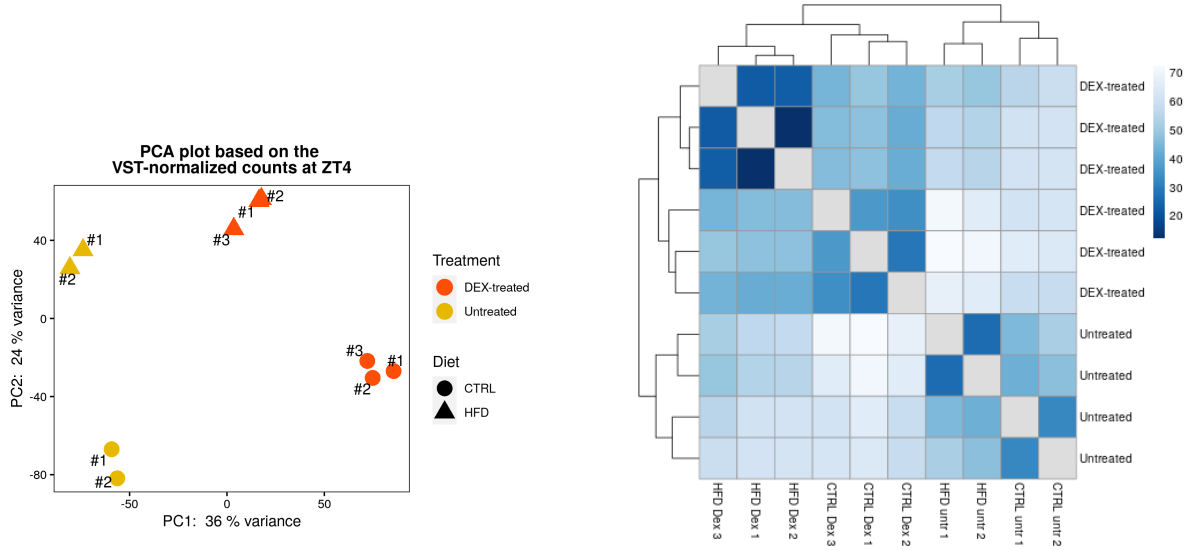


Figure 34: Visualization of sample similarities.
On the left: PCA plot based on the VST-normalized counts of the Dex-treated and untreated samples of both diets at ZT4. Mice were injected with dexamethasone at ZT0, livers were harvested 4 hours later. On the right: sample to sample distance heatmap calculated as Euclidean distance between the VST-normalized count vectors of the same samples.

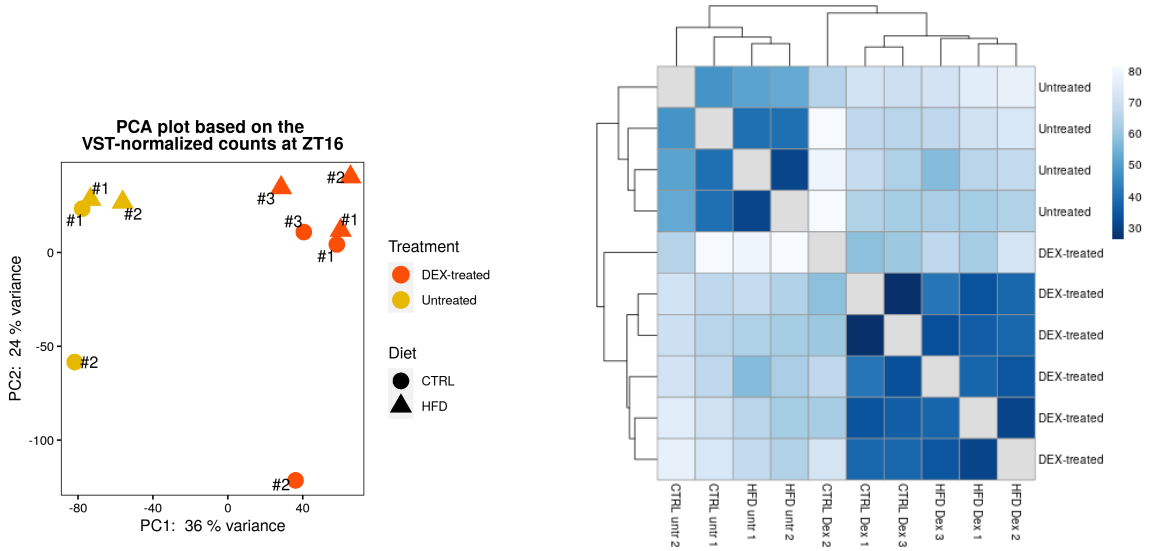


Figure 35: Visualization of sample similarities.
On the left: PCA plot based on the VST-normalized counts of the Dex-treated and untreated samples of both diets at ZT16. Mice were injected with dexamethasone at ZT12, livers were harvested 4 hours later. On the right: sample to sample distance heatmap calculated as Euclidean distance between the VST-normalized count vectors of the same samples.

4.4.2 Ligand-independent reprogramming by HFD

The dexamethasone-treated mice showed increased GR ChIP signal intensity compared to untreated controls at ZT0 as a consequence of increased occupancy in response to the ligand. In the HFD livers this signal was even stronger (Figure 36A and 36B). Of note, both the number of GR ChIP-seq peaks and the signal strength were increased upon HFD, confirming ligand level independent increased GR occupancy.

In both the control and HFD fed groups, I also checked how many deregulated transcripts are between the untreated and Dex-treated samples. As shown in Figure 36C there are more differentially expressed genes in the HFD group than in the control diet fed group. This again supports the finding from the ChIP experiment presented in Figure 36A and 36B.

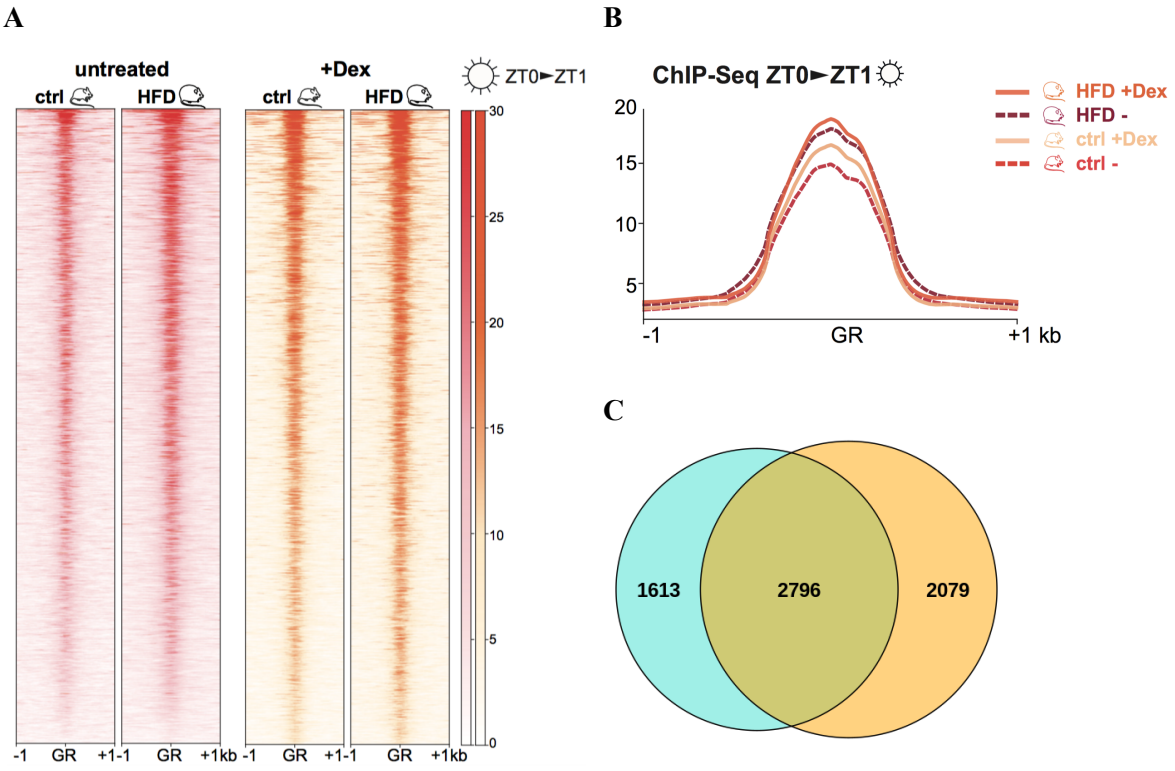
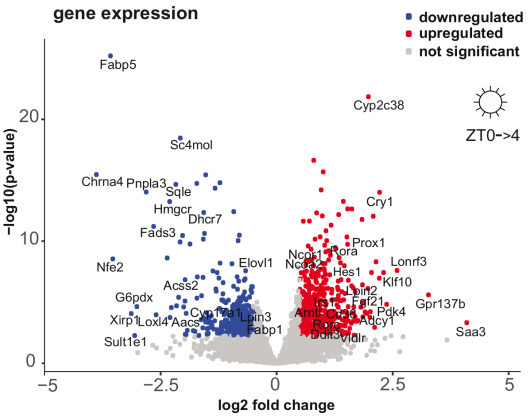


Figure 36: Ligand-independent increase in GR occupancy.
(A) Heatmap showing the genome-wide GR binding in both, control and Dex-treated livers, on control and HFD. Dex was injected at ZT0, livers were collected 1 hour later. (B) Normalized distribution of GR ChIP-seq tag density, corresponding to A. (C) Dex-induced differential expression in the control (blue) and high-fat (orange) diets, and the overlap between the two sets shown in a Venn diagram. Figure 36A and 36B were already published in (Quagliarini et al., 2019)

When analyzing the RNA-seq data of the Dex-treated mice our goal was to compare the response to treatment of the control and HFD fed mice, i.e. identify those genes whose log fold change due to treatment is significantly different for the two diets. The analysis revealed several hundreds of deregulated transcripts that responded differently to GR ligands on HFD during the day (Figure 37 upper part). As seen in Figure 37 (bottom part) the lists of down- and upregulated genes were annotated separately. Genes involved in cholesterol biosynthesis, metabolism of lipids, and lipoproteins were downregulated, while circadian rhythms, PPARα signaling were upregulated during the day. Among the genes showing reduced expression were *Elovl1*, involved in fatty acid elongation, and *Fabp5* which participates in fatty acid uptake, transport, and metabolism. Other genes like for example the clock gene *Cry1*, the nuclear receptor *Rora*, the nuclear corepressor *Ncor1* were upregulated.



Reactome: down-regulated	Adj p-value
cholesterol biosynthesis	1.9E-14
regulation of cholesterol biosynthesis by SREBP	8.4E-07
metabolism of lipids and lipoproteins	1.3E-04
processing of capped intron-containing pre-mRNA	4.3E-04
metabolism of non-coding RNA	2.6E-02
snRNP assembly	2.6E-02
mRNA splicing	2.7E-02

Reactome: up-regulated	Adj p-value
BMAL1:CLOCK,NPAS2 circadian gene expression	4.2E-05
signaling by TGF-beta receptor complex	2.4E-04
ER quality control compartment	2.1E-03
PPARα activated gene expression	2.9E-03
cellular responses to stress	1.3E-02
COPII mediated vesicle transport	1.3E-02
HDMs histone demethylation	1.4E-02

Figure 37: Ligand-independent genomic response on HFD during the day.
Volcano plot illustrating genes differentially responding to Dex treatment in HFD versus control (adj.p-value < 0.05). Dex was injected at ZT0 and livers collected four hours later. Pathway annotation of the up- and downregulated genes marked blue and red in the volcano plot. Figure already published in (Quagliarini et al., 2019)

In the analogous experiment performed during the night, although the GR ChIP-seq signal was again increased (Figure 38A), we did not observe major HFD-specific transcript changes in the Dex-response, presumably because GR ligand levels are the highest at night (Figure 38B).

Importantly, this dataset shows little change in gene expression, and did not reveal any functionally enriched pathways.

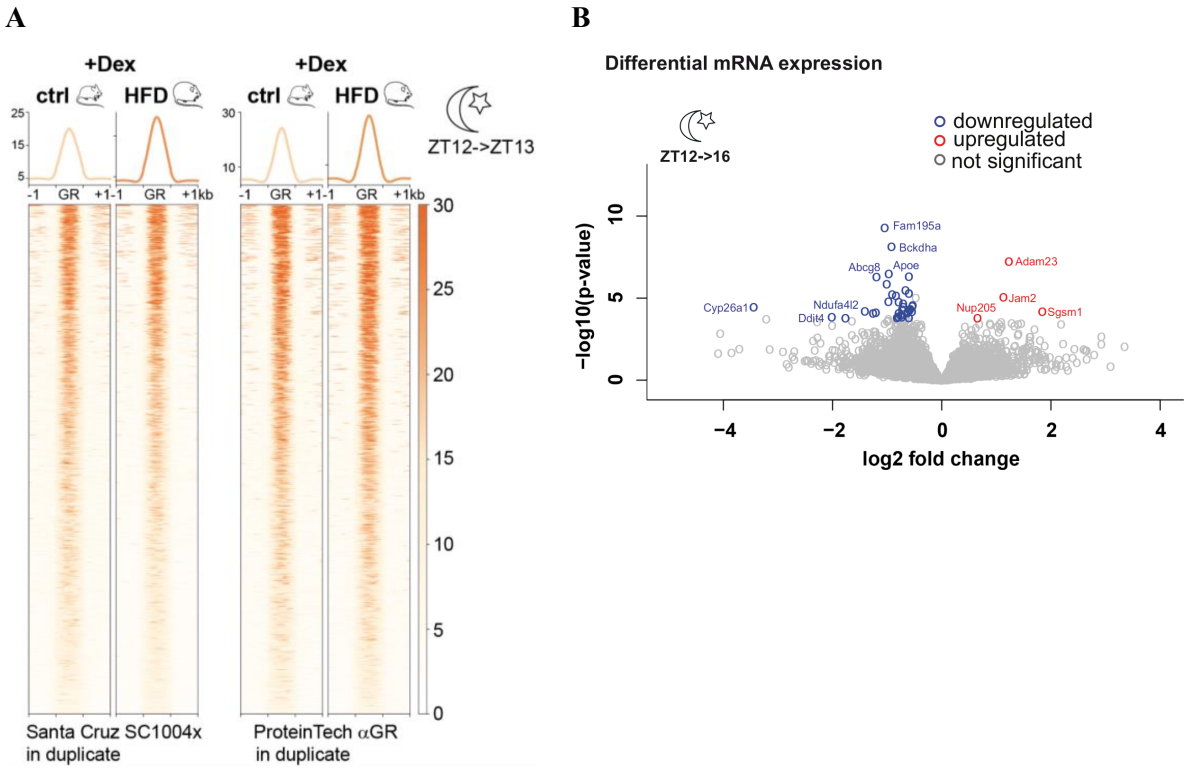


Figure 38: Ligand-independent genomic response on HFD during the night.
(A) Heatmaps illustrating GR genomic binding (using two different GR antibodies) in Dex-treated livers on control and high-fat diet, with Dex injection occurring at ZT12 and chromatin being processed 1 hour later. (B) Volcano plot illustrating genes differentially responding to Dex treatment in HFD versus control (adj.p-value < 0.05), for mice injected with Dex at ZT12 and livers collected four hours later. Figure already published in (Quagliarini et al., 2019)

However, we analysed the effect of HFD in the Dex-treated mice, by comparing the Dex-treated mice on control diet with the Dex-treated on HFD, while controlling for the sampling time. We found 1352 genes that were upregulated and 1142 downregulated with respect to the control diet.

In conclusion, our data suggest that HFD is able to reprogram the cistrome in a ligand independent manner.

4.5 Characterization of the metabolomics data

To investigate significant differences in the metabolic profiles of mice under different dietary conditions and genotypes, targeted metabolomics analysis on liver-specific GR knockout and WT mice on both control and high-fat diet was performed. Measurements were taken in one time-point, at ZT12, having three biological replicates per category (Figure 39). In total 405 metabolites were measured using the Absolute*IDQ*® p400 HR kit from Biocrates (Table 2).

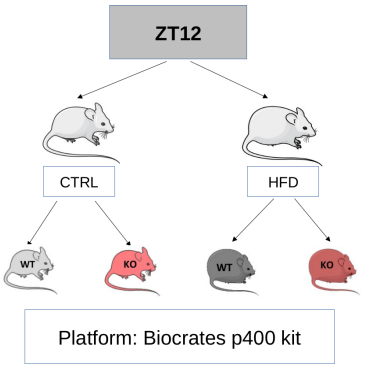


Figure 39: Graphical representation of the experiment.
Metabolite levels were measured in livers of WT and GR-LKO mice on both diets at ZT12 with the targeted p400 kit from Biocrates.

Out of the 405 metabolites, 247 metabolites passed the quality control. The excluded metabolite concentrations were split roughly equally between missing values and measurements below the limit of detection. After data transformation, results were visualized in a PCA biplot (Figure 40), where we can see that the first two principal components capture 24.7% and 19.4% of the variation. Replicates cluster together and we observed a clear separation by diet, along the metabolites with the highest loadings: PC(29:0) and SM(32:2).

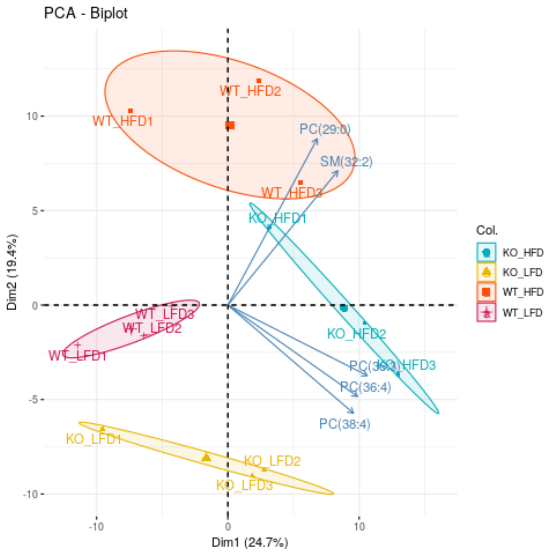


Figure 40: PCA biplot of the metabolomics samples.
The top 5 loadings are shown with blue arrows. Here the control diet is denoted LFD.

4.5.1 Metabolic adaptation to prolonged HFD feeding

In order to find significant metabolite expression differences between the groups, similarly to differential gene expression analysis, I applied a linear model. The model was defined as: metabolite ~ diet + genotype + diet:genotype, meaning that we expect the metabolite level to depend on the diet, genotype, and their combination. The FDR < 0.05 was used as a significance cutoff. The interaction term did not obtain a statistically significant effect size for any metabolite, therefore the simplified model was applied metabolite ~ diet + genotype. The list of metabolites that were significantly differentially expressed is listed in Table 3 below. As we can see, 29 metabolites were identified as significantly different, of which only 5 showed a genotype effect (PC(41:5), LPC(20:0), SM(39:2), SM(34:1) and PC(31:1)), while the rest were significantly different between the two diets.

Metabolite Name	Genotype adj. p-value	Diet adj. p-value
PC(41:5)	**1.19E-09**	6.80E-01
LPC(20:0)	**1.48E-03**	6.02E-02
SM(39:2)	**2.79E-02**	1.14E-01
SM(34:1)	**2.79E-02**	3.88E-01
PC(31:1)	**3.03E-02**	3.62E-01
PC(30:1)	1.34E-01	**2.53E-05**
PC-O(36:3)	2.83E-01	**6.67E-05**
SM(36:1)	8.31E-01	**7.03E-05**
PC(32:1)	4.74E-01	**8.90E-04**
PC(30:0)	3.44E-01	**1.26E-03**
AC(4:0-OH)	8.31E-01	**1.46E-03**
AC(4:0-DC)	5.61E-02	**1.89E-03**
LPC(20:3)	1.05E-01	**1.89E-03**
PC(34:1)	1.84E-01	**5.38E-03**
SM(32:1)	8.84E-01	**5.38E-03**
PC(29:0)	1.45E-01	**1.75E-02**
Histamine	4.88E-01	**2.19E-02**
CE(20:5)	1.05E-01	**2.33E-02**
t4-OH-Pro	3.96E-01	**2.33E-02**
PC-O(30:0)	9.55E-01	**3.39E-02**
PC(33:0)	5.34E-01	**3.49E-02**
PC(40:3)	4.02E-01	**4.40E-02**
PC(34:4)	6.56E-01	**4.40E-02**
LPC(16:0)	4.03E-01	**4.45E-02**
SM(32:2)	7.68E-02	**4.69E-02**
SM(30:1)	1.34E-01	**4.69E-02**
Cer(40:1)	1.80E-01	**4.69E-02**
Creatinine	8.09E-01	**4.69E-02**
DG(32:1)	9.69E-01	**4.69E-02**

Table 3: List of metabolites that display either a significant genotype or a diet effect.

Two examples from each significance category are also visualized in the line plots below. Figure 41 for LPC(20:0), a member of lysophosphatidylcholines, and PC(41:5) represent significant examples for the genotype effect. When we compare the measured metabolite levels in the control and HFD within the two genotypes, we can not observe any significant differences.

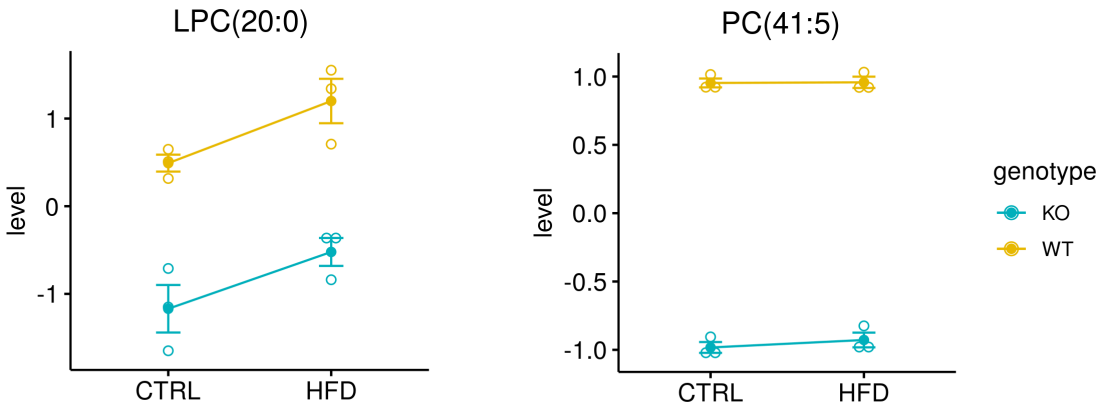


Figure 41: Line plots showing two examples of metabolites having a significant genotype effect. *Plotted are the transformed metabolite levels in each sample.*

Analogously, Figure 42 shows two examples of the diet effect. The measured metabolite levels are significantly higher in the control diet for creatinine and in the HFD for histamine. However, there is no significant difference between the genotypes in none of the diets. While there was a tendency to show a genotype-dependent diet effect in both cases, the interaction was not statistically significant.

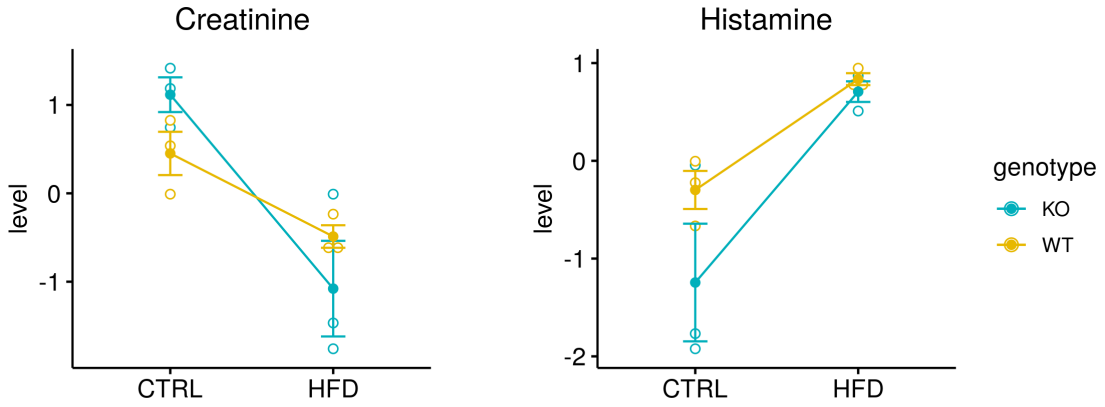


Figure 42: Line plots showing two examples of metabolites that have a significant diet effect. *Plotted are transformed metabolite levels for creatinine ($C_4H_7N_3O$) and histamine ($C_5H_9N_3$).*

To better understand the observed changes, as metabolites can be considered the phenotypic readout of gene expression, we aimed to couple our findings with the results from the RNA-seq data of matched samples. In order to overlay metabolite and transcript data, the MetaboAnalyst platform was used, which enables integrative pathway analysis by coupling differentially expressed genes and differentially expressed metabolites.

Because after FDR multiple testing correction only 29 metabolites were showing either significant diet or genotype effect (Table 3), for the joint pathway analysis we considered all metabolites that had a significant p-value < 0.05 (these metabolites are visualized in the heatmaps in Figure 43).

Although the integrative pathway analysis revealed four significantly enriched pathways (Table 4), they contained hits only from the RNA-seq data, involving none of the metabolites. A reason for this might be that before the pathway mapping is done, metabolite symbols are first converted into KEGG, HMDB, or PubChem IDs, and roughly half of the metabolites had no corresponding IDs.

Pathway	Total	Expected	Hits	FDR
Retinol metabolism	44	1.70	8	1.42E-02
Terpenoid backbone biosynthesis	36	1.39	7	1.42E-02
Nitrogen metabolism	10	0.39	4	1.92E-02
Arginine biosynthesis	27	1.05	6	1.92E-02

Table 4: Significantly impacted gene-metabolite pathways.
Integrative pathway analysis on genes and metabolites, showing significant diet (upper pair) and genotype (lower pair) effect at ZT12.

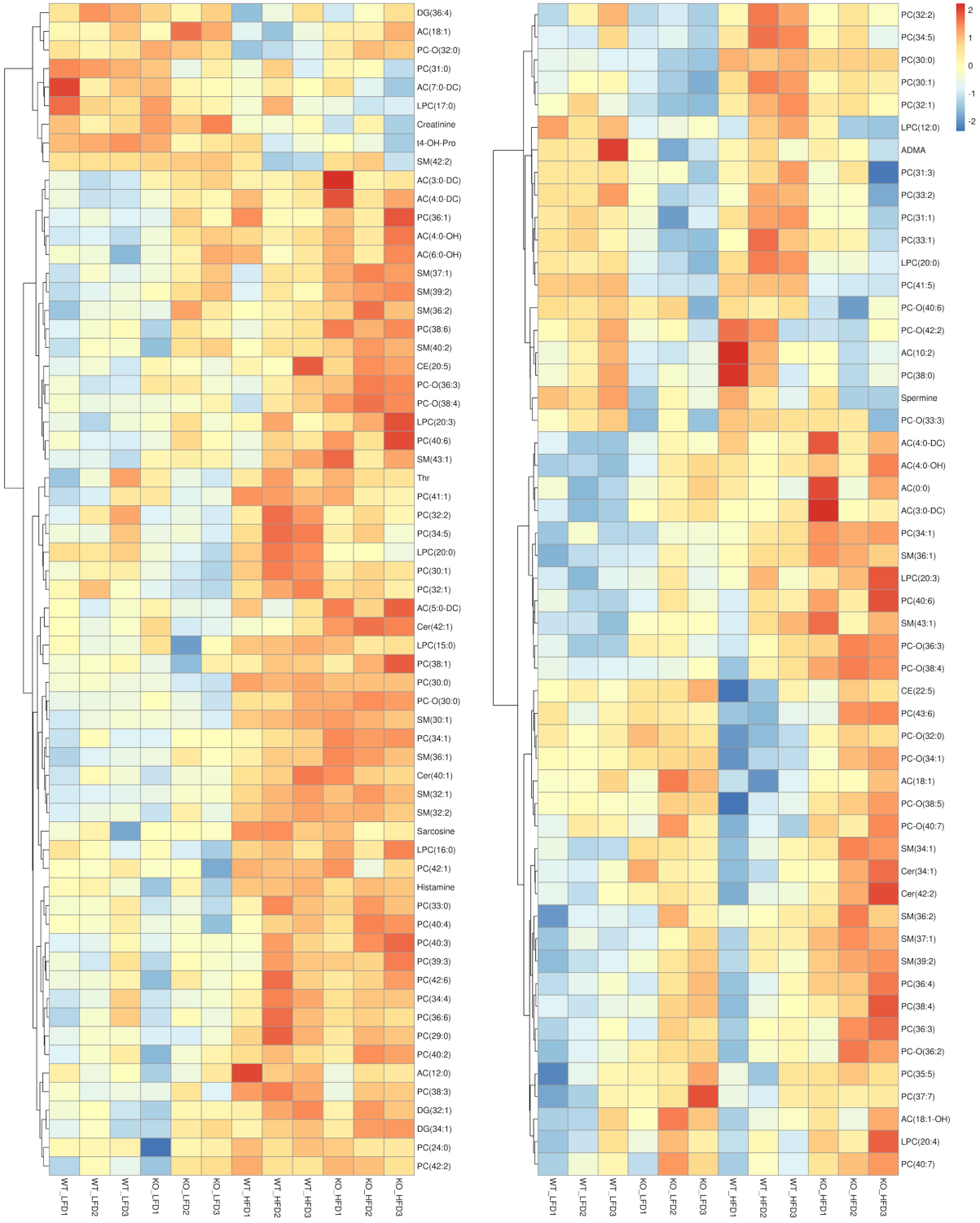


Figure 43: Heatmaps of metabolites with a significant p-values < 0.05 for diet (left) and genotype (right).

4.6 Characterization of RNA-seq data from dexamethasone-treated WT mice for differential transcript usage detection

The tissue-specific transcriptional regulation through GR is influenced by multiple interacting factors. In order to get insights into GR interactome at chromatin level, we performed Chromatin Immunoprecipitation coupled with mass spectrometry (ChIP-MS) experiments in mouse liver tissue (Figure 44). The proteomics analysis showed a prominent enrichment of core spliceosomal components and other associated factors.

In the last decades, it became clear that transcription factors can control gene expression not only through transcription initiation, but also through other gene regulatory layers, such as splicing, stability, and transport (Rambout et al., 2018). However, GR binding to mRNA or the spliceosome and its control over the processing of the transcriptional targets is still unexplored. We investigated GR's effect on differential transcript usage.

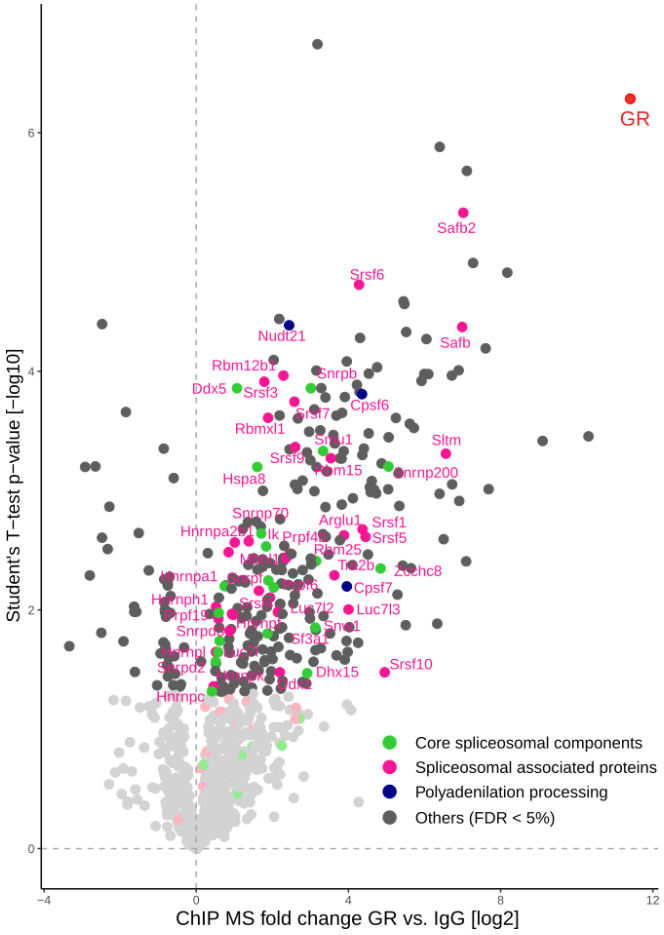


Figure 44: The Glucocorticoid Receptor and the splicing machinery.
GR transcriptional complex was immunoprecipitated and analyzed by ChIP-MS; components of the spliceosome were functionally annotated using KEGG categories.

In order to test our hypothesis that dexamethasone induces not only differential gene expression but also influences splicing, we analyzed a subgroup of the RNA-seq data from our dexamethasone-treated cohort. We aimed to identify whether there are differences in the composition of gene isoform abundances between conditions.

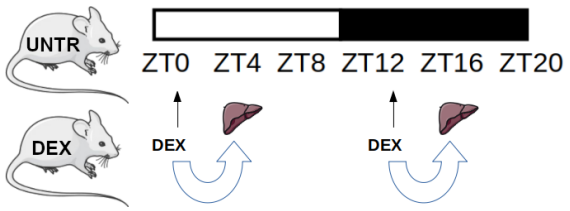


Figure 45: Graphical representation of the experiment.
Wild type mice after 12 weeks on a control diet were injected with dexamethasone in two time-points and liver samples were collected 4 hours later.

The mice used for this analysis were all WT mice on a control diet, untreated controls and their littermates injected with dexamethasone in two time-points at ZT0 and ZT12, respectively. Livers were collected 4 hours later (Figure 45). We had three biological replicates for the day group and two for the night group. The data were analyzed first with the DEXseq Bioconductor R package and then with SUPPA2, developed in Python. While both programs are able to detect DTU, SUPPA2 can also be used to calculate differential splicing for alternative splicing events, where the local alternative splicing events are standard local splicing variations.

4.6.1 Dexamethasone-induced differential transcript usage is more prominent during the night

As already mentioned, differential transcript usage refers to changes in the abundance of individual transcripts, compared only to transcripts of the same gene.

As presented in Figure 46A, when running DEXseq we detected over 300 genes that show evidence for DTU in the day group, and three times more, around 900, in the night group. More precisely, there were over 300 genes whose individual transcripts showed significantly different abundance between the control and the Dex-treated group.

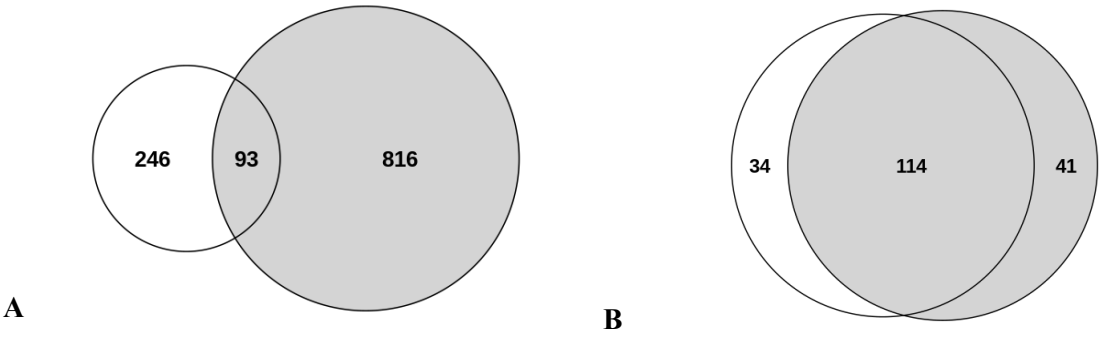


Figure 46: More Dex-induced DTU during the night than during the day detected with DEXseq. *(A) Venn-diagram showing the number of genes that show Dex-induced DTU during the day (ZT0, ZT4, ZT8, marked in white) and night (ZT12, ZT16, ZT20, marked in light gray). (B) Venn-diagram showing the distribution of the transcripts of the overlapping (93) genes.*

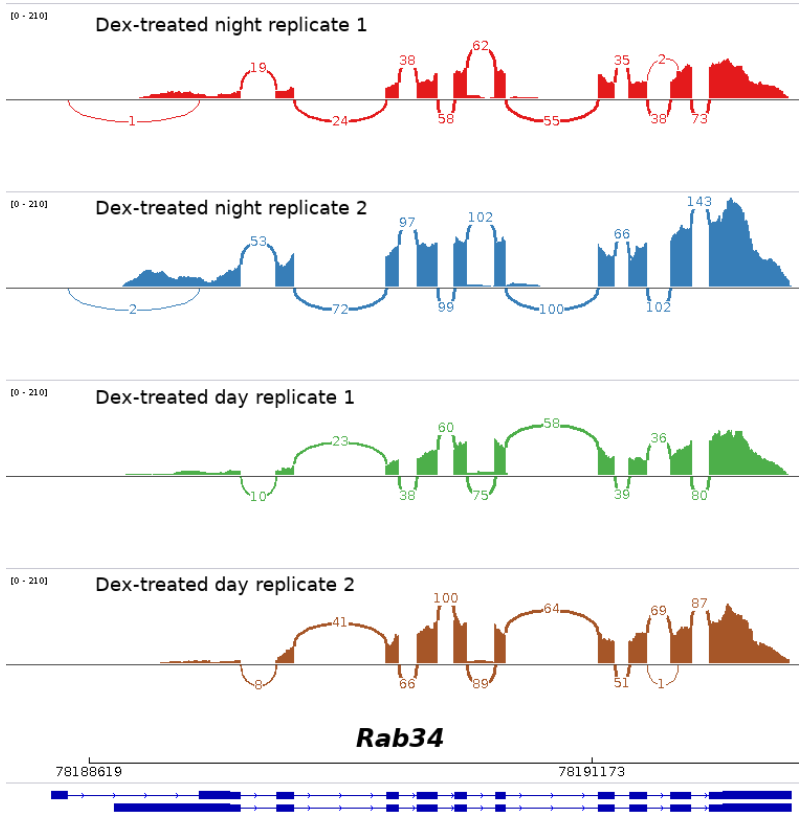


Figure 47: Representative Sashimi plot of the gene *Rab34*.
Dex induces other transcripts of the Rab34 gene during the night than during the day. Four Dex-treated samples are shown (2 night, 2 day). Numbers represent reads across different splice junctions.

When we further checked how many of these genes are present in both day and night groups, we found 93 genes overlapping between the two datasets. Out of the 93 genes that show time-independent Dex-induced DTU, there were 18 genes that differ in their transcripts, meaning they expressed other transcripts differentially during the day than during the night. As seen in Figure 46B, these 18 genes

have 34 transcripts that show DTU during the day and 41 during the night. Figure 47 presents an example gene, *Rab34*, which is from the list of 18 genes described before.

Figure 48: More dexamethasone-induced DTU during the night than during the day detected with SUPPA2. *Venn-diagram showing the number of genes that show Dex-induced DTU during the day (ZT0, ZT4, ZT8, marked in white) and night (ZT12, ZT16, ZT20, marked in light gray).*

Additionally to DEXseq I also run SUPPA2 on the same datasets, to study splicing at the transcript isoform and at the local alternative splicing event level. When running the alternative DTU-detection program, we observe again that there are significantly more genes that show DTU during the night than during the day as presented in Figure 48. We saw that almost half of the "day" Dex-induced DTU genes (262) are shared with the night group.

4.6.2 A shortlist of high-confidence genes was identified, where dexamethasone induced differential transcript usage

By intersecting the results of DEXSeq and SUPPA for both day and night, the aim was to detect those highest confidence genes which show a consistent DTU response to dexamethasone treatment. From the list of the 46 genes in the overall intersection (see Figure 49), I subtracted those which used different transcripts during the day and the night (according to both DEXSeq and SUPPA). This way we found a set of 42 genes that show consistent Dex-induced DTU. GO biological process analysis of these genes identified enriched terms involved in metabolic process, organic substance metabolic process, etc.

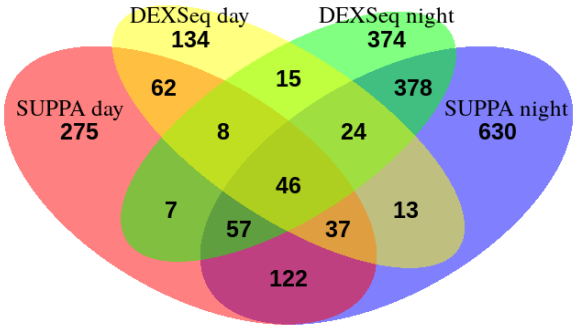


Figure 49: Overlaps and differences between genes identified by SUPPA and DEXSeq as showing DTU during the day and the night.
46 genes were found by both methods as showing Dex-induce DTU in a time independent manner.

Description	FDR q-value
metabolic process	2.60E-04
organic substance metabolic process	4.13E-03
cellular metabolic process	9.68E-03
small molecule metabolic process	1.32E-02
primary metabolic process	1.06E-02
suckling behavior	1.25E-02
catabolic process	2.14E-02
cellular catabolic process	2.78E-02
carboxylic acid metabolic process	3.69E-02
small molecule catabolic process	3.60E-02
oxoacid metabolic process	4.65E-02

Table 5: Significantly enriched GO categories of the 46 overlapping genes found by both DEXSeq and SUPPA that showed Dex-induced differential transcript usage.

By performing the functional classification of these genes through the PANTHER web service, we can see that a majority of them fall into the metabolite interconversion enzyme protein class (Figure 50). Only 28 of the genes were annotated in the ontology.

As seen in Table 5 most genes (e.g. *Ppm1b* and *Mat1a*) were involved in a broad range of metabolic processes. Examples of genes from the list include *Dek*, *Nqo2*, and *Plin5*. *Dek* is a known onco target gene, and it was shown that its loss induces genome instability and sensitizes cells to DNA double-strand breaks. *Nqo2* is mainly involved in detoxification pathways (Mandal et al., 2012), while *Plin5* was shown to have a protective role against hepatic lipotoxic injuries induced by the hepatitis C virus (Zhang et al., 2019).

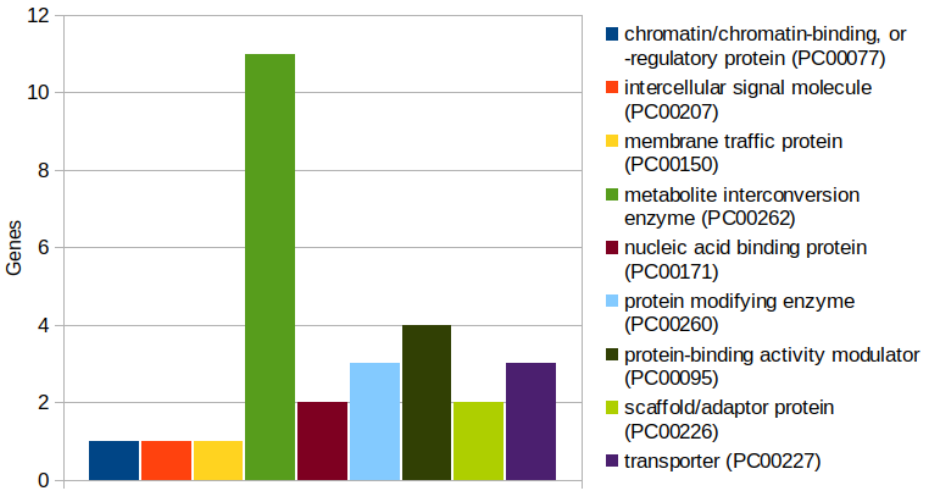


Figure 50: Distribution of the 46 overlapping genes over protein classes from the functional classification performed by PANTHER.
Only 28 of them were annotated in the ontology.

4.6.3 Event-based analysis shows the same ratio of event types during the day and during the night

When running SUPPA on the splicing events instead of a transcript-level analysis, differentially occurring event types were identified in 562 genes during the day (an example result for the *Rhbdd2* gene is shown in Figure 51) and 1162 during the night. We compared the results between day and night and checked the proportions of different splicing events. We observed no significant differences in the distribution of events between the two groups, suggesting that dexamethasone-induced splicing influences the same events in a time-independent manner (see Figure 52).

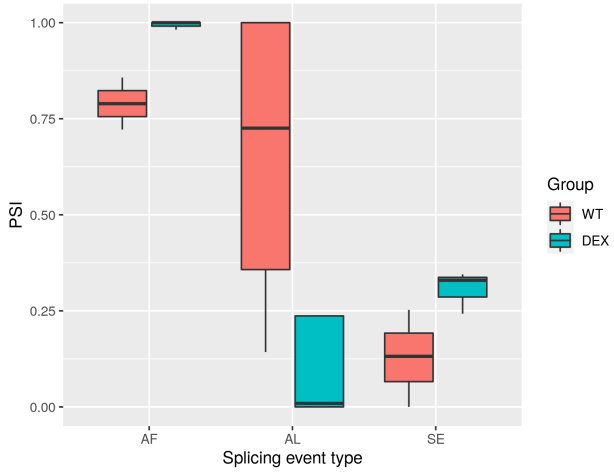


Figure 51: Boxplot representing the PSI values of different splicing event types identified to be different between WT and Dex-treated mice in the *Rhbdd2* gene during the day.
SUPPA2 events: alternative first/last exons (AF/AL), skipping exon (SE).

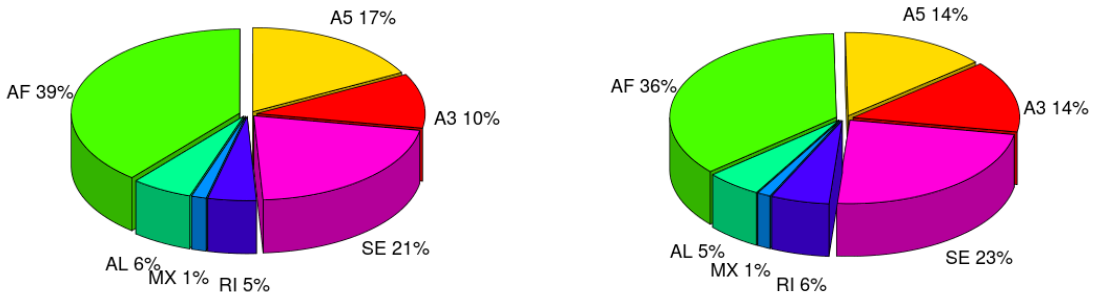


Figure 52: Differentially occurring splicing events during the day (left) and during the night (right) show a similar distribution of event categories.
SUPPA2 events: skipping exon (SE), alternative 5'/3' splice sites (A5/A3), mutually exclusive exons (MX), retained intron (RI), alternative first/last exons (AF/AL).

The distribution of event categories follows the expected proportions between the events. In higher eukaryotes, selection of alternative 5'/3' splice sites (A5/A3) occurs in 18% and 8% of cases, respectively (Mehmood et al., 2019). Intron retention, while common in plants, fungi, and metazoa, constitutes only 5% of the alternative splicing events in higher eukaryotes (Kim et al., 2008). Exon skipping (including SE, AF, and AL) is rarely observed in lower eukaryotes, but it is the most frequent one in higher eukaryotes, representing around 40% of all event types (Alekseyenko et al., 2007; Sugnet et al., 2003).

5. Discussion

5.1 Biological interpretation of the results

In this thesis, I analyzed and integrated multiple types of omic data in an attempt to better understand GR's contribution to diurnal rhythms in the context of prolonged high-fat and control diet feeding. Using circadian hepatic GR ChIP-seq and RNA-seq data coupled with targeted metabolomics data at ZT12 we were able to identify components relevant in the circadian hepatic response to HFD consumption. We could show that rhythms and amplitude stability depend on GR, as hepatic GR depletion leads to loss of oscillation in the majority of the genes and causes an amplitude dampening of rhythmic gene expression across all six time-points. However, the rhythmicity of clock genes was not affected in the liver-specific GR knockout mice, except for *Per1*. Interestingly we found that more than half of the oscillating genes harbored a nearby GR peak. Additionally, we could show that the increased DNA-binding of GR in the HFD group was ligand-independent and diet-induced. Moreover, we detected significant metabolic adaptation to prolonged high-fat diet feeding. Finally, we could link GR to splicing, by detecting dexamethasone-induced differential transcript usage.

5.1.1 Prolonged HFD feeding reprograms the hepatic circadian clock

The liver is considered as being the body's metabolic hub, which is able to connect metabolically different tissues, from adipose tissue, brain to skeletal muscle. Transcription factors and nuclear proteins tightly regulate the hepatic energy metabolism that involves mainly carbohydrate, lipid, and protein metabolism (Rui, 2014). Endogenous dysregulation as well as exogenous perturbations can cause metabolic imbalance. Accumulating evidence supports the fact that nutritional stress, like prolonged HFD feeding, altered feeding patterns, etc., can cause circadian disruption in hepatic peripheral clocks, which results in obesity, type 2 diabetes, insulin resistance, etc. (Arble et al., 2009; Eckel-Mahan et al., 2013).

Feeding a HFD to mammalian models for several weeks leads to the reprogramming of the liver's transcriptional and metabolic response and mice develop diet-induced obesity and even hepatosteatosis (Wang et al., 2016). The adaptive changes to nutritional cues include reprogramming of genes involved in glucose and lipid metabolism, elevated levels of circulating free fatty acids, elevated TG levels in the liver, increased inflammation, disruption of the circadian clock in the liver.

Using circadian RNA-seq data from livers of WT mice after 12 weeks on a control or a HFD, we explored the effects of the prolonged nutritional challenge on the circadian transcriptome. We have demonstrated that HFD-induced transcriptional remodeling consists of the following changes: (i) HFD induces a massive increase of de novo oscillating transcripts; (ii) majority of genes that preserved oscillation show phase advancement in the HFD group; (iii) the rhythmicity of clock genes is preserved.

Published data are only partially in accordance with our results. While (Eckel-Mahan et al., 2013) found approximately the same amount of genes that preserved oscillation as we did (775 vs. 675 in our dataset), they had much more genes that lost oscillation after a HFD feeding than genes that gained oscillation (1394 and 654). We found the opposite, the number of newly oscillating genes was strikingly high (2207) in comparison to those that lost oscillation (275). However, a thorough look at the genes that kept oscillation in both groups revealed the same: the majority of oscillating genes were phase-shifted by HFD. Eckel-Mahan and colleagues found that 66% of genes that preserved oscillation were phase-shifted, while we observed the same in 61% of the cases. Also, we both found that this phase shift is mostly a phase advancement. Pathway annotation linked these genes to "circadian clock". This finding, together with the fact that the circadian clock genes preserved their oscillation, underscores the robustness of the core clock genes in face of external perturbations.

Although the main findings of the two studies were in line, there was still a high discrepancy in the number of genes that were newly oscillating in our HFD group. However, one can identify many aspects that were distinct in the two approaches and could lead to the observed differences. First of all, they employed microarrays while we used RNA-seq as quantitative technology. They fed animals with a normal chow diet, while we had a control diet. Moreover, the used HFDs were not identical. Their mice were on this diet for 10 weeks, while ours for 12 weeks in total. Additionally, the parameter settings used in the analysis were different. While we allowed only strict periods of 24 hours, they set the period window to be between 20-28 hours. Interestingly, I also tested a larger range for the period, and this resulted in less oscillating transcripts in each category (gained, lost and maintained oscillation). Another important difference was the used significance cutoff. We defined our significance cutoff as < 0.05 after Benjamini-Hochberg correction, while they used $p < 0.01$.

All in all, we can conclude that while both groups studied the same research question, a dearth of methodological differences can accumulate to large differences in the results, not to mention the random effects resulting in false discoveries. Not only does the experimental setup influence the output, but also the performed analysis steps, involving different parameter settings, used thresholds, number of replicates, batch effect corrections, and sometimes the version of the used program as well;

with the latter variable being very hard to detect and judge for its impact. While exact reproducibility can not be achieved, careful documentation of the study's steps and performing meta-analyses over multiple studies are the only way to ensure high-certainty results.

5.1.2 HFD increases the number of deregulated transcripts during the night and affects fatty acid metabolism

To investigate the high-fat diet-induced dysregulation of hepatic transcriptome the same cohort was further analyzed. We aimed to understand the effect of HFD on gene expression and to identify disturbed metabolic pathways. Our analysis found twice as many genes that were differentially expressed during the night than during the day, in accordance with the ChIP-seq results on increased GR binding during the night. This result is expected, given the fact that mice are nocturnal animals and the circadian peak in GCs release is bound to the active phase of the animal. Physiologically, GCs play a crucial role in maintaining blood glucose during periods of stress or fasting and the entrainment of metabolic programs in peripheral tissues. The liver, being a central organ in the handling of energy homeostasis, is a major target for GCs signaling.

Almost half of the deregulated transcripts from the day group were shared with the night group. Pathway annotation revealed that the up-regulated genes in the overlap were associated with fatty acid metabolism (e.g. *Acsl1* - involved in lipid synthesis, and *Acox1* - the first enzyme of the fatty acid beta-oxidation pathway), metabolism of amino acids and derivatives metabolism of vitamins and cofactors, and protein localization.

5.1.3 The increased GR binding induced by HFD is functionally relevant

In our circadian GR ChIP-seq experiments in mouse livers after 12 weeks of control and high-fat diet, we observed a circadian pattern for GR binding aligned with its ligand availability (i.e. much more binding during the night). Interestingly, in the HFD-fed group, this general pattern was preserved, and additionally, we saw a massive increase of de novo binding sites during the night (Quagliarini et al., 2019), as shown in Figure S2. RNA-seq data analysis also showed much more deregulated HFD-induced transcripts during the night (Figure 10). Interestingly, we found that almost half (~42%) of the HFD-induced deregulated transcripts during the night were direct GR targets.

Genes involved in the following pathways were upregulated during the night:

- lipoprotein secretion and lipid and cholesterol efflux (*Abca1, Abcb11, Apoa1, Cyp7b1, Saa4*),
- fatty acid beta-oxidation (*Acat1*),
- intracellular fatty acid transport (*Fabp5*),
- and fatty acid and sterol synthesis, modification and elongation (*Acsl5, Elovl3, Fads2, Hsd17b12, Sc5d*).

Additionally, we identified the following HFD-induced GR targets that were downregulated:

- stearoyl coenzyme A desaturase 1 (*Scd1*), encoding the catalyst of a rate-limiting step during unsaturated fatty acid synthesis,
- polyunsaturated fatty acid (PUFA) epoxygenases (*Cyp2c29*, *Cyp2c37*, *Cyp2c38*, *Cyp2c39*),
- enzymes of the urea cycle (*Arg1*, *Asl*, *Ass1*).

Of note, all the above-reported changes were also detected in a study conducted by Toye and colleagues (Toye et al., 2007), who studied liver transcriptomics with the aim to identify changes in gene expression associated with HFD-induced non-alcoholic fatty liver disease (NAFLD) in 129S6 mice. Taken together, our findings underscore the link between HFD-feeding, GR, and potentially NAFLD which could be an interesting direction for future research, as it was shown that both the down- and up-regulation of GR has an aggravating effect on NAFLD pathology (Koorneef et al., 2018).

5.1.4 Rhythm and amplitude stability depend on GR, while the core clock is unaffected

Metabolic processes in the cell are regulated by the molecular circadian clock and metabolic homeostasis requires the alignment of these clocks between tissues. Environmental cues, like for example high-fat diet feeding, fasting, etc., are able to induce misalignment in peripheral clocks as they are very sensitive to nutritional changes. Altered rhythms, mainly phase shifts, can cause the disruption of the normal circadian cycle and have severe clinical implications on metabolism, resulting mainly in cardiometabolic disorders.

Although it is known that the glucocorticoid receptor plays a crucial role in the synchronization of peripheral clocks, our group was the first to study its contribution to diurnal rhythms on a genomic scale.

By profiling mRNA expression in both GR-LKO and WT mice throughout the day after 12 weeks on a control diet we could assess the effect of hepatic GR depletion on circadian rhythm. Our analysis

revealed that loss of GR leads to rhythmicity loss in the majority of genes. A thorough look at the amplitude of these genes showed that hepatic GR is required for amplitude stability. Data integration of the list of GR ChIP peaks with the RNA-seq data revealed that the majority of diurnally oscillating genes are direct GR targets. More than half of the genes that lost oscillation were bound by GR, therefore it was not surprising to see lost oscillation in genes like e.g. *Lpin1* or *Foxo1* that are known GR targets.

Detecting all members of the core clock in the list of genes that maintain rhythmicity, suggests that GR contributes to circadian rhythm synchronization of downstream target genes together with core clock factors.

5.1.5 More differential expression at night

The same samples as before were also used to detect how GR loss influences the transcriptional outcome in a time-dependent manner. Differential expression analysis for the day and night samples was performed separately. Once more our results were in line with the findings from the ChIP-seq data analysis, where we observed much more binding during the night than during the day. Functional annotation of the up and downregulated gene lists showed that genes upregulated at night were associated with lipid, fatty acid, and triglyceride metabolism, while downregulated genes were involved in glucose metabolism. Our findings suggest that glucocorticoids govern nutrient partitioning and substrate utilization during daily cycles of feeding and fasting.

5.1.6 HFD-induced cistromic reprogramming of the GR response is reflected in the transcriptome

Previous studies have investigated the link between HFD and epigenome and found that HFD increased H3K27 acetylation and DNaseI hypersensitivity (Siersbæk et al., 2017). Our ChIP-seq profiles for H3K27ac also showed an increased signal in the HFD group and although the difference was minimal, yet associated with gained binding. This increased chromatin accessibility might contribute to the observed increase in GR DNA occupancy upon HFD feeding.

By studying the relationship between GR, circadian rhythm, and HFD, in analogous experiments in WT and GR-LKO mice after 12 weeks on HFD, we observed the same trend as in the control diet fed group. Again, there were more deregulated genes between the GR-LKO and WT during the night than

during the day. Downregulated genes in the night group were enriched for pathways including lipoprotein metabolism, metabolism of amino acids, and derivatives, while upregulated genes were mainly associated with metabolism of lipids and lipoproteins. By testing whether the HFD-induced binding events are functionally relevant we detected more than 400 genes that were deregulated in the absence of GR and harbored a nearby GR peak. Upregulated genes were associated with lipid and fatty acid metabolism and downregulated ones with metabolism of fat-soluble vitamins and urea cycle.

Investigation of the combined genotype-diet (i.e. GR loss and HFD) effect on transcripts rhythmicity revealed once more that the core clock genes remained unperturbed. We saw a significant amplitude dampening in all six time-points for those genes that lost oscillation in the absence of GR. However, this time the number of genes that lost and preserved oscillation was roughly the same.

Overall our results underscore the role of GR in generating rhythmic outputs and emphasize the link between the mammalian circadian clock and metabolism.

5.1.7 Cistromic reprogramming by HFD is ligand-independent

By injecting mice on both diets with dexamethasone (an exogenous GR agonist) at ZT0 and ZT12, the nadir and zenith of the endogenous GC levels, we could demonstrate that the observed increase in DNA-bound GR is indeed a consequence of the HFD and did not depend on ligand availability.

Acute dexamethasone injection in WT mice during the day lead to a significant number of deregulated transcripts. For example the mRNA expression of genes like *Fabp5*, *Elovl1* was downregulated while *Hes1, Lpin2, and Cry1* were upregulated between the Dex-treated control and HFD-fed groups. Lipin2 (*Lpin2*) is involved in controlling the metabolism of fatty acids at different levels (Csaki and Reue, 2010). Fatty acid-binding protein 5 (*Fabp5*) participates in fatty acid uptake, transport, and metabolism by binding long-chain fatty acids (Senga et al., 2018), while *Elovl1* encodes the elongation of very long chain fatty acids protein 1 (Ofman et al., 2010).

Our findings might have broad consequences for patients on glucocorticoid treatment, as one could expect to see a different response to the treatment in obese individuals.

5.1.8 Metabolic adaptation to prolonged high-fat diet feeding

Performing targeted metabolomics in livers of WT and GR-LKO mice on both diets we were able to detect differences in the metabolic profile of mice. We captured the significant diet and genotype effects and could show that the most prominent changes are induced by the long-term HFD feeding, the most changes being detected in the levels of phosphatidylcholines followed by sphingomyelins.

Our findings are in line with the literature, as it was shown that HFD leads to elevated levels of sphingomyelin in the liver, skeletal muscle, adipose, and cardiovascular tissues (Choi and Snider, 2015). Moreover, HFD induces de novo sphingolipid synthesis that contributes to systemic insulin resistance and altered lipid accumulation (Turpin et al., 2014). Additionally, abnormal sphingomyelin levels were found in obesity-induced atherosclerosis (Li et al., 2005). Besides sphingomyelins also Cer(40:1) was found as significantly different. Interestingly, the literature review in (Choi and Snider, 2015) suggests species-dependent roles of ceramides in the regulation of energy metabolism.

Another metabolite that was found to have significantly different levels in the control and HFD was histamine. Histamine is a biogenic amine known to play an important role in the inflammatory response and allergies (Wood, 2006). The enzyme histidine decarboxylase (HDC) decarboxylates the amino acid L-histidine to form histamine. Accordingly, studies showed that plasma histamine levels were decreased in *Hdc* gene-deficient mice (DeChiara et al., 1995). Moreover, it has been shown that histmanine can regulate appetite, and promote biliary damage and hepatic fibrosis (Francis et al., 2008, 2007). Interestingly, Kennedy and colleagues showed that by knocking out the *Hdc* gene, mice receiving a high-fat diet were protected from hepatic fibrosis and cholangiocyte damage (Kennedy et al., 2018).

As we saw that adaptation to HFD included changes in genes associated with glucose and lipid metabolism, we expected to observe changes also in the metabolic profiles of these mice. By integrating our metabolomics results with our RNA-seq data of matched samples we found pathways that were significant, however, none of them contained significant gene/metabolite combinations. However, this could be a limitation of the annotation complexity and enrichment computation of MetaboAnalyst, since it did not even identify the simple Histamine-HDC connection described above.

On the other hand, a major drawback of this study was that we investigated a small cohort, only 3 biological replicates for each diet/genotype category. Larger sample size would definitely increase the reliability of the analysis. To verify if results are reproducible, the measurements could be repeated, and a meta-analysis performed. Furthermore, our analysis was limited by the measurement method. Using a targeted metabolomics kit, that is suitable for capturing changes in the lipids, introduces a

strong bias in the detection of metabolite changes. Further investigations should aim at using an untargeted approach that would uncover more significant differences. Additionally, as we know that metabolite levels change rapidly, it would be interesting to see measurements at a better time resolution, not only at ZT12.

Despite these limiting factors, our results confirm that metabolic adaptation occurs as a response to prolonged HFD feeding.

5.1.9 Results of the differential transcript usage analysis were verified experimentally

Differential transcript usage (DTU) can be seen as isoform switches when different isoforms are expressed in different conditions. In some cases, changes can not be seen at the gene level, as the total gene expression level may not change, however, the DTU is of biological importance. Most commonly DTU is observed when comparing expression across cell types. It was shown in (Reyes and Huber, 2018) that half of all expressed genes had tissue-specific isoforms, based on the analysis of the Genotype-Tissue Expression Project (GTEx) dataset (GTEx Consortium et al. 2017). Some DTU patterns seem to be more common for specific diseases. For example, protein domain losses (mainly those involved in protein-protein interactions) were found to be a specific DTU pattern in cancer (Vitting-Seerup and Sandelin, 2017). While in the case of this cancer study the identified genes were also more likely to be mutated, the resulting DTU patterns were indicative of survival, suggesting functional results of the transcript usage. Therefore, identifying whether there are differences in the composition of a gene's isoform abundances between conditions is a great way to complement the "classical" differential gene expression analysis, even in the case of "natural" variation induced by HFD, as this could result in biologically relevant consequences as well.

Using a small cohort of dexamethasone-treated mice and their controls we were able to find a list of genes that were "high confidence". This means that these genes were found to show dexamethasone-induced differential transcript usage in both day and night group and were detected by both methods, DEXseq and SUPPA2.

As part of a bachelor's thesis (Zwang, 2020), we tested several candidates (*Dek, Nqo2, Ndfip1, Ppm1b,* and *Derl2*) from our high-confidence gene list among others found only during the day by DEXseq. Our student was able to experimentally validate *Ppm1b* as showing dexamethasone-induced differential transcript usage (Figure 53). However, this is only a first step before attempting to infer its biological meaning. Unfortunately, *Dek, Nqo2, Ndfip1,* and *Derl2* ended up having disqualified transcripts after the quality control of primers.

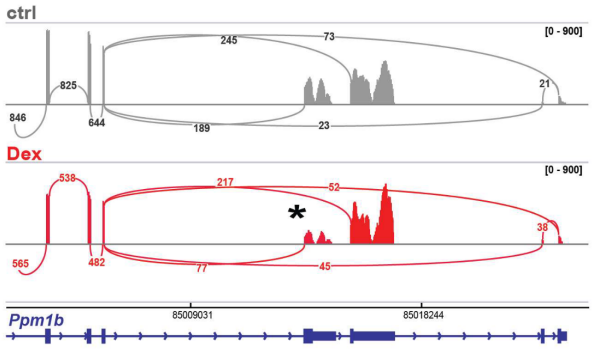


Figure 53: Representative Sashimi plot of liver RNA-seq coverage, showing the numbers of reads across different splice junctions of the *Ppm1b* gene.
The asterisk denotes an exon which is differential in response to Dex.

Nonetheless, understanding the link between the *Derl2* gene and GR would be interesting. GO biological process annotation of the "high-confidence" list found *Derl2* (Derlin-2) to be significantly associated with organic substance metabolic process, cellular metabolic process, primary metabolic process, and the (cellular) catabolic process subcategories. Interestingly, *Derlin-2*$^{-/-}$ mice were found to not develop hypoglycemia after starvation and refused to feed (Dougan et al., 2011). Additionally, *Derl2* is upregulating the protein quality control pathway required for the development of hepatocytes. However, conditional knockout experiments conducted by Dougan and colleagues showed unaffected hepatocyte function after liver-specific deletion of Derlin-2. On the other hand, this gene is conserved from yeast to human, and its human ortholog (97.9% similarity) was found to be upregulated in human liver cancer (Ying et al., 2001). Further biological studies to elucidate the interplay between *Derl2*, HFD, and GR is an interesting avenue for follow-up research.

5.2 Computational biology considerations

Increasingly, omic approaches are being recognized for their utility as each type of omic data provides a glimpse on a particular regulatory level. However, the integration of these heterogeneous results into a sound statistical framework, for testing a large number of hypotheses, can be challenging. To increase the sensitivity of the analysis methods and the robustness of the results, more dense sampling of the timepoints and more replicates would be needed for some of the methods. However, the presented procedures ensured that the results were statistically significant, and the multi-omics integration gave strong clues about biological relevance and causative relationships, as discussed above. Here I will discuss four specific computational biology issues that were relevant during the interpretation of the results.

5.2.1 Computational bottlenecks

During processing the data I encountered several limiting obstacles, due to the large file sizes and complex processing steps. While most steps could be run on a standard laptop, having access to a multi-core server or cluster provides a great speed-up through allowing for parallel processing of the samples. The one step where using a non-standard PC was required was the alignment, which required excessive amounts of memory (around 30GB RAM per sample along with hundreds of GB of storage) and parallel processing for practical runtimes.

In cases where sample sizes are large (in the hundreds or thousands), DGE and DTU analysis can also take a long time to process, and efficient k-mer based approaches like iMOKA might be a more suitable choice (Lorenzi et al., 2020). While big data collections are becoming commonplace, the circadian dataset sizes are strongly lagging this trend (Sun et al., 2020). However, future research is expected to produce larger datasets in this domain as well, making the analysis steps over all samples at all time-points challenging with standard hardware. This is especially true for novel deep learning approaches, that currently require specialized GPU clusters to train, and high-end GPU cards to use the learned models for inference.

5.2.2 Statistical significance versus biological relevance

Best practices on the use of statistical testing are well established in the computational biology field, and are implemented and performed automatically in most software tools. In the case of combining results over multiple steps or experiments, however, the task of correctly accounting for multiple hypothesis tests lies with the users. Therefore, special care has to be taken when integrating results from multi-omics data.

Generally, when identifying significant differences across time-points and conditions, the data-specific testing procedures evaluate the null hypothesis of no difference between groups, in accordance with the assumed underlying distributions of the data. The obtained p-values can be thresholded to filter out tests where the log fold-changes were not significantly different than zero, thus limiting the probability of false positives. But in order to account for the multiple tests performed in parallel, the number of type I errors (i.e. incorrectly rejected null hypotheses) has to be controlled, limiting the family-wise error rate (the probability of making at least one type I error).

One such correction method is the Bonferroni correction (Bland and Altman, 1995), which works by adjusting the significance threshold to ensure that their sum over all performed tests adds up to the significance level α. While this approach is directly controlling the family-wise error rate, it is too conservative, and, in practice, false discovery rate (FDR) controlling procedures are used instead. The FDR also represents the number of false positives (falsely rejected null hypotheses), but as a proportion among all results that were declared significant (all rejected null hypotheses).

The most common method to control the FDR is the Benjamini–Hochberg procedure (BH step-up procedure), which is also implemented in edgeR and DESeq2 (Benjamini and Hochberg, 1995). The resulting adjusted p-values (i.e. q-values) can then be thresholded to a desired level of significance. However, the interpretation of q-values has an important distinction with regard to that of the regular p-values. As discussed above, a 5% cutoff in the case of p-value means that 5% of the total number of tests will result in a false positive, while applying this filter on the q-values would mean that 5% of the tests that are deemed to be significant are in fact false positives (so according to the definition, the rate of false discoveries is 5%).

This raises the question of whether the usual 5% p-value cutoff is still the right choice in the case of q-values. In order to make sure that results are reproducible across experiments, and are not just statistical flukes, the Sequencing Quality Control consortium (Su et al., 2014) recommended that filters for p-values, but also fold-changes and measurement-levels are needed in order to exclude results with effect sizes that are statistically significant but at the same time biologically insignificant. Such a case can occur especially when sample sizes are small, and the variance in the data is underestimated (possibly also due to underlying assumptions that are not met in practice), resulting in small differences becoming statistically significant (necessitating reproduction studies and meta-analyses to exclude such results). Moreover, the consortium recommends applying these thresholds in a pipeline-specific manner, based on the specific consideration of the experiment performed. However, one must consider that applying additional filters apart from the q-value cutoff is distorting the FDR statistics of the final subset, and could both increase or decrease the FDR (depending on whether true positives or false positives are removed by the subsequent filters).

The above considerations apply to both RNA-seq and ChIP-seq data, as well as metabolomics analysis, and we consistently used the BH step-up procedure for correction within the method and for post-hoc analysis where applicable. We did not apply any additional filters to not distort the false discovery rates, and this enabled us and subsequent users of the results to compute the FDR of the different combinations of the results. For example, when intersecting RNA-seq and ChIP-seq results, if we limit the false discovery rate to 5% in both sets, the probability of the overlap containing false

discoveries is $1-(1-5\%)^2=9.75\%$ because the discovery is correct only if it is correctly identified in both input sets. For such multi-omics experiments, it is necessary to select a higher q-value cutoff in the data integration pipeline, than in the constituent ones.

When analyzing the results, a search for biological relevance is key, in order to be able to derive actionable conclusions for the design of treatments, but also as a causal underpinning of the identified correlation in order to validate its biological relevance. While some connections are straightforward, like the differential abundance in both histamine and HDC identified above, in most cases the causational chains in biological systems are complex, with different positive and negative feedback loops, trans-regulatory elements, and adaptations on a system level. Such complex relations can be elucidated only in a step by step manner, integrating large bodies of knowledge, as they do not lend themselves to reductionist approaches.

5.2.3 Meta-analysis of alternative methods

Another consideration for combining the results of multiple tests is the integration of p-values obtained by different alternative methods. This can be exemplified through the identification of cyclical genes. JTK Cycle is more reliable for analysis of large data sets because i) it is more resistant to outliers, ii) has increased power in case of multiple replicates, iii) can handle input data covering only a single cycle, and iv) is several orders of magnitude faster than its “competitor” COSOPT. On the other hand, there are other methods with different pros and cons that could be applied, like ARSER and Lomb-Scargle.

The low sampling density can pose problems for JTK Cycle as well as the other methods, but due to the different assumptions and computations, the adverse effects (that usually result in an increased false-negative rate) can cancel each other out in an “ensemble” approach. Therefore, the MetaCycle package combines the p-values obtained by running two or three of these methods in order to avoid the shortcomings of every single method (Wu et al., 2016). For example, JTK Cycle has a known bias towards cosine waves, whose effect can be mitigated by employing ARSER and Lomb-Scargle. It uses Fischer’s method to combine the method-wise p-values, under the assumption that these are independent. While this is obviously not completely true, studies show that using the more accurate Brown method actually reduces the accuracy of the results (Hutchison and Dinner, 2017).

While in this thesis I only employed JTK Cycle, and using different approaches are not expected to dramatically change the trends observed in the result, future work might include a meta-analysis of multiple methods for this problem.

5.2.4 Handling of missing measurements in metabolomics data analysis

Finally, a difficulty in modeling the data accurately is that certain omics measurements (for technical reasons), will produce incomplete observations, or some data points will be filtered out as part of the quality control. The analysis methods are, however, mostly unsuited for handling such inputs, and using pseudocounts or similar simple solutions can distort the results. I encountered this problem in the case of the metabolomics data, and had to select an appropriate missing value imputation method, in order to predict probable metabolite level measurements for sample-metabolite pairs where these were missing.

The occurrence of missing values is a well known phenomenon in mass-spectrometry based metabolomics measurements and can arise because the metabolite is not present in the sample, or it is present but below the limit of detection, or is present but remains undetected due to technical issues (Taylor et al., 2013). Having fully defined data matrices is a prerequisite of multivariate methods like PCA, thus imputation of missing data is needed before downstream analysis. During imputation missing data are replaced with values based on the information from the existing data.

There are several methods to handle missing values and the selection can have a significant impact on the results. The literature distinguishes between missing values of three types: missing completely at random (MCAR), missing at random (MAR), and missing not at random (MNAR). Whereas MCAR is the simplest case, where the propensity for missing data is uncorrelated to the measured variables (and MNAR is the clearly non-random case), MAR is describing a conditional probability distribution for missing data, dependent on some other variable, which can be controlled for in order to obtain a random distribution (Little and Rubin, 2014).

It was shown that for mass spectrometry based metabolomics data the random forest (RF) method outperformed all tested methods - mean, median, singular value decomposition and k-nearest neighbors - for MCAR/MAR data (Wei et al., 2018). In another recent study conducted by (Kokla et al., 2019) the authors tested different types and rates of missingness in LC-MS metabolomics data and finally recommended using random forest-based imputation for metabolomics data even when the type of missingness is not known in advance.

That a trained machine learning model is showing the most robust results was to be expected, since there is no analytical solution to the missing data problem. In the future, more refined bioinformatic tools based on machine learning, specifically deep learning might help decipher the underlying

structuring of the data, and aid tasks like missing value imputation for example. For such tasks, generative models are especially useful, as they learn a prior distribution from labeled or unlabeled datasets, and use this to generate novel realistic samples. The most basic form of these kinds of approaches are (variational) autoencoders, embedding the data in a low-dimensional distribution. This is, conceptually, a learned non-linear PCA analysis, and has been used already to impute missing data in a time course of metabolite measurements (Scholz et al., 2005). However, it must be noted that any biases present in the training data, or introduced by enforcing a fixed distribution in the learned latent space for variational autoencoders, will be reflected in the generated samples. Therefore, such approaches became feasible only after large datasets, and the hardware necessary to process them, became available. Different techniques, like drop-out (US9406017B2 patent by Google LLC), are commonly employed to reduce the possibility of overfitting, however general guarantees are impossible to give.

6. Concluding remarks and future perspectives

The continuously increasing amount of data, fueled by new emerging technologies and large investments in the field, lead to a need for a scalable integration methodology of the heterogeneous sources of information. Such a task poses fundamental challenges to traditional human-designed processing methods that rely on specific a priori assumptions and heuristics for parameter tuning. However, the availability of large public datasets (annotated and not), and generic tools for high-performance computing, opens the way of a broader deployment of deep learning approaches (Eraslan et al., 2019). The advantage of training computer programs in a data-driven way, instead of manually defining the computations to be done, becomes clear in the recent success of such approaches in variant calling (Luo et al., 2019; Poplin et al., 2018), denoising ChIP-seq data (Koh et al., 2017) and improving Hi-C data resolution (Zhang et al., 2018).

These deep artificial neural networks are generic tools that can be readily applied to a wide range of problems, from natural language processing, image recognition, and automatic translation through different omics analysis tasks. Therefore, it is expected that they will penetrate the field to encompass and bring together the functionality of many specific bioinformatics algorithms into a unified modeling and inference framework. This integration will ensure a holistic interpretation of the data, instead of the piecewise processing pipelines used today.

An early example of such an approach is BPNet (Avsec et al., 2020), which directly predicts ChIP-nexus binding profiles signals at single base resolution from a DNA sequence, and enables the identification of soft syntax rules for cooperative TF binding interactions. The integration of multiple data sources is an important aspect of multi-omics analysis (e.g. ChIP-seq, RNA-seq, proteomics, etc) which can be achieved with specialized network architectures and training protocols (Amodio and Krishnaswamy, 2018). For example, FactorNet combines gene sequences and expression, their annotations, as well as single-nucleotide resolution signals to predict cell type specific transcription factor binding (Quang and Xie, 2019).

Future developments are expected in the area of Graph Convolutional Networks for learning on structured data, e.g. gene interaction networks (Dutil et al., 2018), as well as advancements in the modeling and inference of causal effects. Both research areas will be aided by the continuing increase in perturbation data generation using systematic CRISPR screens and massively parallel reporter assays (Eraslan et al., 2019).

In summary, the added power of multi-omics is evident, and deep learning techniques will contribute to managing the complexity of such multi-dimensional data.

References

Alekseyenko, A.V., Kim, N., Lee, C.J., 2007. Global analysis of exon creation versus loss and the role of alternative splicing in 17 vertebrate genomes. RNA 13, 661–670. https://doi.org/10.1261/rna.325107

Allhoff, M., Seré, K., F. Pires, J., Zenke, M., G. Costa, I., 2016. Differential peak calling of ChIP-seq signals with replicates with THOR. Nucleic Acids Res. 44, e153–e153. https://doi.org/10.1093/nar/gkw680

Amodio, M., Krishnaswamy, S., 2018. MAGAN: Aligning Biological Manifolds, in: International Conference on Machine Learning. Presented at the International Conference on Machine Learning, PMLR, pp. 215–223.

Anders, S., Huber, W., 2010. Differential expression analysis for sequence count data. Genome Biol. 11, R106. https://doi.org/10.1186/gb-2010-11-10-r106

Anders, S., Reyes, A., Huber, W., 2012. Detecting differential usage of exons from RNA-seq data. Genome Res. 22, 2008–2017. https://doi.org/10.1101/gr.133744.111

Arble, D.M., Bass, J., Laposky, A.D., Vitaterna, M.H., Turek, F.W., 2009. Circadian Timing of Food Intake Contributes to Weight Gain. Obes. Silver Spring Md 17, 2100–2102. https://doi.org/10.1038/oby.2009.264

Arneth, B., Arneth, R., Shams, M., 2019. Metabolomics of Type 1 and Type 2 Diabetes. Int. J. Mol. Sci. 20. https://doi.org/10.3390/ijms20102467

Atger, F., Mauvoisin, D., Weger, B., Gobet, C., Gachon, F., 2017. Regulation of Mammalian Physiology by Interconnected Circadian and Feeding Rhythms. Front. Endocrinol. 8. https://doi.org/10.3389/fendo.2017.00042

Audic, S., Claverie, J.M., 1997. The significance of digital gene expression profiles. Genome Res. 7, 986–995. https://doi.org/10.1101/gr.7.10.986

Avsec, Ž., Weilert, M., Shrikumar, A., Krueger, S., Alexandari, A., Dalal, K., Fropf, R., McAnany, C., Gagneur, J., Kundaje, A., Zeitlinger, J., 2020. Base-resolution models of transcription factor binding reveal soft motif syntax. bioRxiv 737981. https://doi.org/10.1101/737981

Babarinde, I.A., Li, Y., Hutchins, A.P., 2019. Computational Methods for Mapping, Assembly and Quantification for Coding and Non-coding Transcripts. Comput. Struct. Biotechnol. J. 17, 628–637. https://doi.org/10.1016/j.csbj.2019.04.012

Balsalobre, A., Brown, S.A., Marcacci, L., Tronche, F., Kellendonk, C., Reichardt, H.M., Schütz, G., Schibler, U., 2000. Resetting of Circadian Time in Peripheral Tissues by Glucocorticoid Signaling. Science 289, 2344–2347. https://doi.org/10.1126/science.289.5488.2344

Bar-Joseph, Z., Gitter, A., Simon, I., 2012. Studying and modelling dynamic biological processes using time-series gene expression data. Nat. Rev. Genet. 13, 552–564. https://doi.org/10.1038/nrg3244

Barnett, D.W., Garrison, E.K., Quinlan, A.R., Strömberg, M.P., Marth, G.T., 2011. BamTools: a C++ API and toolkit for analyzing and managing BAM files. Bioinformatics 27, 1691–1692.

https://doi.org/10.1093/bioinformatics/btr174

Benjamini, Y., Hochberg, Y., 1995. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J. R. Stat. Soc. Ser. B Methodol. 57, 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Benjamini, Y., Speed, T.P., 2012. Summarizing and correcting the GC content bias in high-throughput sequencing. Nucleic Acids Res. 40, e72. https://doi.org/10.1093/nar/gks001

Bentley, D.R., Balasubramanian, S., Swerdlow, H.P., Smith, G.P., Milton, J., Brown, C.G., Hall, K.P., Evers, D.J., Barnes, C.L., Bignell, H.R., Boutell, J.M., Bryant, J., Carter, R.J., Keira Cheetham, R., Cox, A.J., Ellis, D.J., Flatbush, M.R., Gormley, N.A., Humphray, S.J., Irving, L.J., Karbelashvili, M.S., Kirk, S.M., Li, H., Liu, X., Maisinger, K.S., Murray, L.J., Obradovic, B., Ost, T., Parkinson, M.L., Pratt, M.R., Rasolonjatovo, I.M.J., Reed, M.T., Rigatti, R., Rodighiero, C., Ross, M.T., Sabot, A., Sankar, S.V., Scally, A., Schroth, G.P., Smith, M.E., Smith, V.P., Spiridou, A., Torrance, P.E., Tzonev, S.S., Vermaas, E.H., Walter, K., Wu, X., Zhang, L., Alam, M.D., Anastasi, C., Aniebo, I.C., Bailey, D.M.D., Bancarz, I.R., Banerjee, S., Barbour, S.G., Baybayan, P.A., Benoit, V.A., Benson, K.F., Bevis, C., Black, P.J., Boodhun, A., Brennan, J.S., Bridgham, J.A., Brown, R.C., Brown, A.A., Buermann, D.H., Bundu, A.A., Burrows, J.C., Carter, N.P., Castillo, N., Chiara E. Catenazzi, M., Chang, S., Neil Cooley, R., Crake, N.R., Dada, O.O., Diakoumakos, K.D., Dominguez-Fernandez, B., Earnshaw, D.J., Egbujor, U.C., Elmore, D.W., Etchin, S.S., Ewan, M.R., Fedurco, M., Fraser, L.J., Fuentes Fajardo, K.V., Scott Furey, W., George, D., Gietzen, K.J., Goddard, C.P., Golda, G.S., Granieri, P.A., Green, D.E., Gustafson, D.L., Hansen, N.F., Harnish, K., Haudenschild, C.D., Heyer, N.I., Hims, M.M., Ho, J.T., Horgan, A.M., Hoschler, K., Hurwitz, S., Ivanov, D.V., Johnson, M.Q., James, T., Huw Jones, T.A., Kang, G.-D., Kerelska, T.H., Kersey, A.D., Khrebtukova, I., Kindwall, A.P., Kingsbury, Z., Kokko-Gonzales, P.I., Kumar, A., Laurent, M.A., Lawley, C.T., Lee, S.E., Lee, X., Liao, A.K., Loch, J.A., Lok, M., Luo, S., Mammen, R.M., Martin, J.W., McCauley, P.G., McNitt, P., Mehta, P., Moon, K.W., Mullens, J.W., Newington, T., Ning, Z., Ling Ng, B., Novo, S.M., O'Neill, M.J., Osborne, M.A., Osnowski, A., Ostadan, O., Paraschos, L.L., Pickering, L., Pike, Andrew C., Pike, Alger C., Chris Pinkard, D., Pliskin, D.P., Podhasky, J., Quijano, V.J., Raczy, C., Rae, V.H., Rawlings, S.R., Chiva Rodriguez, A., Roe, P.M., Rogers, John, Rogert Bacigalupo, M.C., Romanov, N., Romieu, A., Roth, R.K., Rourke, N.J., Ruediger, S.T., Rusman, E., Sanches-Kuiper, R.M., Schenker, M.R., Seoane, J.M., Shaw, R.J., Shiver, M.K., Short, S.W., Sizto, N.L., Sluis, J.P., Smith, M.A., Ernest Sohna Sohna, J., Spence, E.J., Stevens, K., Sutton, N., Szajkowski, L., Tregidgo, C.L., Turcatti, G., vandeVondele, S., Verhovsky, Y., Virk, S.M., Wakelin, S., Walcott, G.C., Wang, J., Worsley, G.J., Yan, J., Yau, L., Zuerlein, M., Rogers, Jane, Mullikin, J.C., Hurles, M.E., McCooke, N.J., West, J.S., Oaks, F.L., Lundberg, P.L., Klenerman, D., Durbin, R., Smith, A.J., 2008. Accurate whole human genome sequencing using reversible terminator chemistry. Nature 456, 53–59. https://doi.org/10.1038/nature07517

Bland, J.M., Altman, D.G., 1995. Multiple significance tests: the Bonferroni method. BMJ 310, 170. https://doi.org/10.1136/bmj.310.6973.170

Bookout, A.L., Jeong, Y., Downes, M., Yu, R.T., Evans, R.M., Mangelsdorf, D.J., 2006. Anatomical Profiling of Nuclear Receptor Expression Reveals a Hierarchical Transcriptional Network. Cell 126, 789–799. https://doi.org/10.1016/j.cell.2006.06.049

Bray, N.L., Pimentel, H., Melsted, P., Pachter, L., 2016. Near-optimal probabilistic RNA-seq quantification. Nat. Biotechnol. 34, 525–527. https://doi.org/10.1038/nbt.3519

Briot, K., Roux, C., 2015. Glucocorticoid-induced osteoporosis. RMD Open 1.

https://doi.org/10.1136/rmdopen-2014-000014

Bullard, J.H., Purdom, E., Hansen, K.D., Dudoit, S., 2010. Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. BMC Bioinformatics 11, 94. https://doi.org/10.1186/1471-2105-11-94

Carmena, R., 2005. Type 2 diabetes, dyslipidemia, and vascular risk: Rationale and evidence for correcting the lipid imbalance. Am. Heart J. 150, 859–870. https://doi.org/10.1016/j.ahj.2005.04.027

Carninci, P., 2009. Is sequencing enlightenment ending the dark age of the transcriptome? Nat. Methods 6, 711–713. https://doi.org/10.1038/nmeth1009-711

Chandola, T., Brunner, E., Marmot, M., 2006. Chronic stress at work and the metabolic syndrome: prospective study. BMJ 332, 521–525. https://doi.org/10.1136/bmj.38693.435301.80

Cheng, X., Buckley, D., Klaassen, C.D., 2007. Regulation of hepatic bile acid transporters Ntcp and Bsep expression. Biochem. Pharmacol. 74, 1665–1676. https://doi.org/10.1016/j.bcp.2007.08.014

Choi, S., Snider, A.J., 2015. Sphingolipids in High Fat Diet and Obesity-Related Diseases [WWW Document]. Mediators Inflamm. https://doi.org/10.1155/2015/520618

Chong, J., Wishart, D.S., Xia, J., 2019. Using MetaboAnalyst 4.0 for Comprehensive and Integrative Metabolomics Data Analysis. Curr. Protoc. Bioinforma. 68, e86. https://doi.org/10.1002/cpbi.86

Chrousos, G.P., 1998. Editorial: Ultradian, Circadian, and Stress-Related Hypothalamic-Pituitary-Adrenal Axis Activity—A Dynamic Digital-to-Analog Modulation. Endocrinology 139, 437–440. https://doi.org/10.1210/endo.139.2.5857

Cook, R.D., 1977. Detection of Influential Observation in Linear Regression. Technometrics 19, 15–18. https://doi.org/10.2307/1268249

Coondoo, A., Phiske, M., Verma, S., Lahiri, K., 2014. Side-effects of topical steroids: A long overdue revisit. Indian Dermatol. Online J. 5, 416–425. https://doi.org/10.4103/2229-5178.142483

Coutinho, A.E., Chapman, K.E., 2011. The anti-inflammatory and immunosuppressive effects of glucocorticoids, recent developments and mechanistic insights. Mol. Cell. Endocrinol. 335, 2–13. https://doi.org/10.1016/j.mce.2010.04.005

Csaki, L.S., Reue, K., 2010. Lipins: Multifunctional Lipid Metabolism Proteins. Annu. Rev. Nutr. 30, 257–272. https://doi.org/10.1146/annurev.nutr.012809.104729

Dallman, M.F., la Fleur, S.E., Pecoraro, N.C., Gomez, F., Houshyar, H., Akana, S.F., 2004. Minireview: Glucocorticoids—Food Intake, Abdominal Obesity, and Wealthy Nations in 2004. Endocrinology 145, 2633–2638. https://doi.org/10.1210/en.2004-0037

DeChiara, T.M., Vejsada, R., Poueymirou, W.T., Acheson, A., Suri, C., Conover, J.C., Friedman, B., McClain, J., Pan, L., Stahl, N., Ip, N.Y., Yancopoulos, G.D., 1995. Mice lacking the CNTF receptor, unlike mice lacking CNTF, exhibit profound motor neuron deficits at birth. Cell 83, 313–322. https://doi.org/10.1016/0092-8674(95)90172-8

Dietrich, M.O., Horvath, T.L., 2009. Feeding signals and brain circuitry. Eur. J. Neurosci. 30, 1688–1696. https://doi.org/10.1111/j.1460-9568.2009.06963.x

Dillies, M.-A., Rau, A., Aubert, J., Hennequet-Antier, C., Jeanmougin, M., Servant, N., Keime, C., Marot, G., Castel, D., Estelle, J., Guernec, G., Jagla, B., Jouneau, L., Laloë, D., Le Gall, C., Schaëffer, B., Le Crom, S., Guedj, M., Jaffrézic, F., French StatOmique Consortium, 2013. A comprehensive evaluation of normalization methods for Illumina high-throughput RNA sequencing data analysis. Brief. Bioinform. 14, 671–683. https://doi.org/10.1093/bib/bbs046

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., Gingeras, T.R., 2013. STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29, 15–21. https://doi.org/10.1093/bioinformatics/bts635

Dolinsky, V.W., Douglas, D.N., Lehner, R., Vance, D.E., 2004. Regulation of the enzymes of hepatic microsomal triacylglycerol lipolysis and re-esterification by the glucocorticoid dexamethasone. Biochem. J. 378, 967–974. https://doi.org/10.1042/BJ20031320

Dougan, S.K., Hu, C.-C.A., Paquet, M.-E., Greenblatt, M.B., Kim, J., Lilley, B.N., Watson, N., Ploegh, H.L., 2011. Derlin-2-Deficient Mice Reveal an Essential Role for Protein Dislocation in Chondrocytes. Mol. Cell. Biol. 31, 1145–1159. https://doi.org/10.1128/MCB.00967-10

Dutil, F., Cohen, J.P., Weiss, M., Derevyanko, G., Bengio, Y., 2018. Towards Gene Expression Convolutions using Gene Interaction Graphs. ArXiv180606975 Cs Q-Bio Stat.

Eckel-Mahan, K., Sassone-Corsi, P., 2013. Metabolism and the circadian clock converge. Physiol. Rev. 93, 107–135. https://doi.org/10.1152/physrev.00016.2012

Eckel-Mahan, K.L., Patel, V.R., de Mateo, S., Orozco-Solis, R., Ceglia, N.J., Sahar, S., Dilag-Penilla, S.A., Dyar, K.A., Baldi, P., Sassone-Corsi, P., 2013. Reprogramming of the circadian clock by nutritional challenge. Cell 155, 1464–1478. https://doi.org/10.1016/j.cell.2013.11.034

Eden, E., Lipson, D., Yogev, S., Yakhini, Z., 2007. Discovering Motifs in Ranked Lists of DNA Sequences. PLOS Comput. Biol. 3, e39. https://doi.org/10.1371/journal.pcbi.0030039

Eden, E., Navon, R., Steinfeld, I., Lipson, D., Yakhini, Z., 2009. GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. BMC Bioinformatics 10, 48. https://doi.org/10.1186/1471-2105-10-48

Eraslan, G., Avsec, Ž., Gagneur, J., Theis, F.J., 2019. Deep learning: new computational modelling techniques for genomics. Nat. Rev. Genet. 20, 389–403. https://doi.org/10.1038/s41576-019-0122-6

Ericson-Neilsen, W., Kaye, A.D., 2014. Steroids: Pharmacology, Complications, and Practice Delivery Issues. Ochsner J. 14, 203–207.

Evans, C., Hardin, J., Stoebel, D.M., 2017. Selecting between-sample RNA-Seq normalization methods from the perspective of their assumptions. Brief. Bioinform. 19, 776–792. https://doi.org/10.1093/bib/bbx008

Ewels, P., Magnusson, M., Lundin, S., Käller, M., 2016. MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics 32, 3047–3048. https://doi.org/10.1093/bioinformatics/btw354

Fabregat, A., Sidiropoulos, K., Viteri, G., Forner, O., Marin-Garcia, P., Arnau, V., D'Eustachio, P., Stein, L., Hermjakob, H., 2017. Reactome pathway analysis: a high-performance in-memory approach. BMC Bioinformatics 18, 142. https://doi.org/10.1186/s12859-017-1559-2

Feelders, R.A., Pulgar, S.J., Kempel, A., Pereira, A.M., 2012. MANAGEMENT OF ENDOCRINE

DISEASE: The burden of Cushing's disease: clinical and health-related quality of life aspects. Eur. J. Endocrinol. 167, 311–326. https://doi.org/10.1530/EJE-11-1095

Francis, H., Franchitto, A., Ueno, Y., Glaser, S., DeMorrow, S., Venter, J., Gaudio, E., Alvaro, D., Fava, G., Marzioni, M., Vaculin, B., Alpini, G., 2007. H3 histamine receptor agonist inhibits biliary growth of BDL rats by downregulation of the cAMP-dependent PKA/ERK1/2/ELK-1 pathway. Lab. Investig. J. Tech. Methods Pathol. 87, 473–487. https://doi.org/10.1038/labinvest.3700533

Francis, H., Glaser, S., DeMorrow, S., Gaudio, E., Ueno, Y., Venter, J., Dostal, D., Onori, P., Franchitto, A., Marzioni, M., Vaculin, S., Vaculin, B., Katki, K., Stutes, M., Savage, J., Alpini, G., 2008. Small mouse cholangiocytes proliferate in response to H1 histamine receptor stimulation by activation of the IP3/CaMK I/CREB pathway. Am. J. Physiol. - Cell Physiol. 295, C499–C513. https://doi.org/10.1152/ajpcell.00369.2007

Froy, O., 2010. Metabolism and Circadian Rhythms—Implications for Obesity. Endocr. Rev. 31, 1–24. https://doi.org/10.1210/er.2009-0014

García-Alcalde, F., Okonechnikov, K., Carbonell, J., Cruz, L.M., Götz, S., Tarazona, S., Dopazo, J., Meyer, T.F., Conesa, A., 2012. Qualimap: evaluating next-generation sequencing alignment data. Bioinforma. Oxf. Engl. 28, 2678–2679. https://doi.org/10.1093/bioinformatics/bts503

Gaucher, D., Therrien, R., Kettaf, N., Angermann, B.R., Boucher, G., Filali-Mouhim, A., Moser, J.M., Mehta, R.S., Drake, D.R., Castro, E., Akondy, R., Rinfret, A., Yassine-Diab, B., Said, E.A., Chouikh, Y., Cameron, M.J., Clum, R., Kelvin, D., Somogyi, R., Greller, L.D., Balderas, R.S., Wilkinson, P., Pantaleo, G., Tartaglia, J., Haddad, E.K., Sékaly, R.-P., 2008. Yellow fever vaccine induces integrated multilineage and polyfunctional immune responses. J. Exp. Med. 205, 3119–3131. https://doi.org/10.1084/jem.20082292

Gaucher, J., Montellier, E., Sassone-Corsi, P., 2018. Molecular Cogs: Interplay between Circadian Clock and Cell Cycle. Trends Cell Biol. 28, 368–379. https://doi.org/10.1016/j.tcb.2018.01.006

Gekakis, N., Staknis, D., Nguyen, H.B., Davis, F.C., Wilsbacher, L.D., King, D.P., Takahashi, J.S., Weitz, C.J., 1998. Role of the CLOCK protein in the mammalian circadian mechanism. Science 280, 1564–1569. https://doi.org/10.1126/science.280.5369.1564

Glynn, E.F., Chen, J., Mushegian, A.R., 2006. Detecting periodic patterns in unevenly spaced gene expression time series using Lomb-Scargle periodograms. Bioinforma. Oxf. Engl. 22, 310–316. https://doi.org/10.1093/bioinformatics/bti789

Greene, G.L., Gilna, P., Waterfield, M., Baker, A., Hort, Y., Shine, J., 1986. Sequence and expression of human estrogen receptor complementary DNA. Science 231, 1150–1154. https://doi.org/10.1126/science.3753802

Grøntved, L., John, S., Baek, S., Liu, Y., Buckley, J.R., Vinson, C., Aguilera, G., Hager, G.L., 2013. C/EBP maintains chromatin accessibility in liver and facilitates glucocorticoid receptor recruitment to steroid response elements. EMBO J. 32, 1568–1583. https://doi.org/10.1038/emboj.2013.106

Gu, Z., Eils, R., Schlesner, M., 2016. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. Bioinforma. Oxf. Engl. 32, 2847–2849. https://doi.org/10.1093/bioinformatics/btw313

Guillaumond, F., Dardente, H., Giguère, V., Cermakian, N., 2005. Differential control of Bmal1

circadian transcription by REV-ERB and ROR nuclear receptors. J. Biol. Rhythms 20, 391–403. https://doi.org/10.1177/0748730405277232

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., Glass, C.K., 2010. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol. Cell 38, 576–589. https://doi.org/10.1016/j.molcel.2010.05.004

Heller, R., Manduchi, E., Grant, G.R., Ewens, W.J., 2009. A flexible two-stage procedure for identifying gene sets that are differentially expressed. Bioinforma. Oxf. Engl. 25, 1019–1025. https://doi.org/10.1093/bioinformatics/btp076

Hench, P.S., Kendall, E.C., Slocumb, C.H., Polley, H.F., 1949. The effect of a hormone of the adrenal cortex (17-hydroxy-11-dehydrocorticosterone: compound E) and of pituitary adrenocortical hormone in arthritis: preliminary report. Ann. Rheum. Dis. 8, 97–104. https://doi.org/10.1136/ard.8.2.97

Herman, J.P., McKlveen, J.M., Ghosal, S., Kopp, B., Wulsin, A., Makinson, R., Scheimann, J., Myers, B., 2016. Regulation of the hypothalamic-pituitary-adrenocortical stress response. Compr. Physiol. 6, 603–621. https://doi.org/10.1002/cphy.c150015

Höllbacher, B., Balázs, K., Heinig, M., Uhlenhaut, N.H., 2020. Seq-ing answers: Current data integration approaches to uncover mechanisms of transcriptional regulation. Comput. Struct. Biotechnol. J. 18, 1330–1341. https://doi.org/10.1016/j.csbj.2020.05.018

Hollenberg, S.M., Weinberger, C., Ong, E.S., Cerelli, G., Oro, A., Lebo, R., Brad Thompson, E., Rosenfeld, M.G., Evans, R.M., 1985. Primary structure and expression of a functional human glucocorticoid receptor cDNA. Nature 318, 635–641. https://doi.org/10.1038/318635a0

Hu, M., Zhu, Y., Taylor, J.M.G., Liu, J.S., Qin, Z.S., 2012. Using Poisson mixed-effects model to quantify transcript-level gene expression in RNA-Seq. Bioinforma. Oxf. Engl. 28, 63–68. https://doi.org/10.1093/bioinformatics/btr616

Hughes, M.E., Hogenesch, J.B., Kornacker, K., 2010. JTK_CYCLE: an efficient non-parametric algorithm for detecting rhythmic components in genome-scale datasets. J. Biol. Rhythms 25, 372–380. https://doi.org/10.1177/0748730410379711

Hulse, A.M., Cai, J.J., 2013. Genetic variants contribute to gene expression variability in humans. Genetics 193, 95–108. https://doi.org/10.1534/genetics.112.146779

Hutchison, A.L., Dinner, A.R., 2017. Correcting for Dependent P-values in Rhythm Detection. bioRxiv 118547. https://doi.org/10.1101/118547

Hwang, J.L., Weiss, R.E., 2014. Steroid-induced diabetes: a clinical and molecular approach to understanding and treatment. Diabetes Metab. Res. Rev. 30, 96–102. https://doi.org/10.1002/dmrr.2486

Imai, E., Stromstedt, P.E., Quinn, P.G., Carlstedt-Duke, J., Gustafsson, J.A., Granner, D.K., 1990. Characterization of a complex glucocorticoid response unit in the phosphoenolpyruvate carboxykinase gene. Mol. Cell. Biol. 10, 4712–4719.

Janes, M., Kuster, S., Goldson, T.M., Forjuoh, S.N., 2019. Steroid-induced psychosis. Proc. Bayl. Univ. Med. Cent. 32, 614–615. https://doi.org/10.1080/08998280.2019.1629223

Johnson, C.H., Patterson, A.D., Idle, J.R., Gonzalez, F.J., 2012. Xenobiotic Metabolomics: Major

Impact on the Metabolome. Annu. Rev. Pharmacol. Toxicol. 52, 37–56. https://doi.org/10.1146/annurev-pharmtox-010611-134748

Kanehisa, M., Goto, S., 2000. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 28, 27–30. https://doi.org/10.1093/nar/28.1.27

Kelder, T., Pico, A.R., Hanspers, K., van Iersel, M.P., Evelo, C., Conklin, B.R., 2009. Mining Biological Pathways Using WikiPathways Web Services. PLoS ONE 4. https://doi.org/10.1371/journal.pone.0006447

Kennedy, L., Hargrove, L., Demieville, J., Bailey, J.M., Dar, W., Polireddy, K., Chen, Q., Nevah Rubin, M.I., Sybenga, A., DeMorrow, S., Meng, F., Stockton, L., Alpini, G., Francis, H., 2018. Knockout of l-Histidine Decarboxylase Prevents Cholangiocyte Damage and Hepatic Fibrosis in Mice Subjected to High-Fat Diet Feeding via Disrupted Histamine/Leptin Signaling. Am. J. Pathol. 188, 600–615. https://doi.org/10.1016/j.ajpath.2017.11.016

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., Salzberg, S.L., 2013. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. 14, R36. https://doi.org/10.1186/gb-2013-14-4-r36

Kim, E., Goren, A., Ast, G., 2008. Alternative splicing: current perspectives. BioEssays News Rev. Mol. Cell. Dev. Biol. 30, 38–47. https://doi.org/10.1002/bies.20692

Klein, D., Moore, R., Reppert, S., 1991. The Suprachiasmatic Nucleus: The Mind's Clock. Oxford University Press, New York.

Koh, P.W., Pierson, E., Kundaje, A., 2017. Denoising genome-wide histone ChIP-seq with convolutional neural networks. Bioinformatics 33, i225. https://doi.org/10.1093/bioinformatics/btx243

Kohsaka, A., Laposky, A.D., Ramsey, K.M., Estrada, C., Joshu, C., Kobayashi, Y., Turek, F.W., Bass, J., 2007. High-Fat Diet Disrupts Behavioral and Molecular Circadian Rhythms in Mice. Cell Metab. 6, 414–421. https://doi.org/10.1016/j.cmet.2007.09.006

Kokla, M., Virtanen, J., Kolehmainen, M., Paananen, J., Hanhineva, K., 2019. Random forest-based imputation outperforms other methods for imputing LC-MS metabolomics data: a comparative study. BMC Bioinformatics 20. https://doi.org/10.1186/s12859-019-3110-0

Koorneef, L.L., van den Heuvel, J.K., Kroon, J., Boon, M.R., 't Hoen, P.A.C., Hettne, K.M., van de Velde, N.M., Kolenbrander, K.B., Streefland, T.C.M., Mol, I.M., Sips, H.C.M., Kielbasa, S.M., Mei, H., Belanoff, J.K., Pereira, A.M., Oosterveer, M.H., Hunt, H., Rensen, P.C.N., Meijer, O.C., 2018. Selective Glucocorticoid Receptor Modulation Prevents and Reverses Nonalcoholic Fatty Liver Disease in Male Mice. Endocrinology 159, 3925–3936. https://doi.org/10.1210/en.2018-00671

Kumar, R., Thompson, E.B., 2005. Gene regulation by the glucocorticoid receptor: Structure:function relationship. J. Steroid Biochem. Mol. Biol. 94, 383–394. https://doi.org/10.1016/j.jsbmb.2004.12.046

Lamia, K.A., Papp, S.J., Yu, R.T., Barish, G.D., Uhlenhaut, N.H., Jonker, J.W., Downes, M., Evans, R.M., 2011. Cryptochromes mediate rhythmic repression of the glucocorticoid receptor. Nature 480, 552–556. https://doi.org/10.1038/nature10700

Landt, S.G., Marinov, G.K., Kundaje, A., Kheradpour, P., Pauli, F., Batzoglou, S., Bernstein, B.E., Bickel, P., Brown, J.B., Cayting, P., Chen, Y., DeSalvo, G., Epstein, C., Fisher-Aylor, K.I.,

Euskirchen, G., Gerstein, M., Gertz, J., Hartemink, A.J., Hoffman, M.M., Iyer, V.R., Jung, Y.L., Karmakar, S., Kellis, M., Kharchenko, P.V., Li, Q., Liu, T., Liu, X.S., Ma, L., Milosavljevic, A., Myers, R.M., Park, P.J., Pazin, M.J., Perry, M.D., Raha, D., Reddy, T.E., Rozowsky, J., Shoresh, N., Sidow, A., Slattery, M., Stamatoyannopoulos, J.A., Tolstorukov, M.Y., White, K.P., Xi, S., Farnham, P.J., Lieb, J.D., Wold, B.J., Snyder, M., 2012. ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. Genome Res. 22, 1813–1831. https://doi.org/10.1101/gr.136184.111

Langmead, B., Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. Nat. Methods 9, 357–359. https://doi.org/10.1038/nmeth.1923

Langmead, B., Trapnell, C., Pop, M., Salzberg, S.L., 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 10, R25. https://doi.org/10.1186/gb-2009-10-3-r25

Lee, C., Etchegaray, J.-P., Cagampang, F.R.A., Loudon, A.S.I., Reppert, S.M., 2001. Posttranslational Mechanisms Regulate the Mammalian Circadian Clock. Cell 107, 855–867. https://doi.org/10.1016/S0092-8674(01)00610-9

Lemke, U., Krones-Herzig, A., Diaz, M.B., Narvekar, P., Ziegler, A., Vegiopoulos, A., Cato, A.C.B., Bohl, S., Klingmüller, U., Screaton, R.A., Müller-Decker, K., Kersten, S., Herzig, S., 2008. The Glucocorticoid Receptor Controls Hepatic Dyslipidemia through Hes1. Cell Metab. 8, 212–223. https://doi.org/10.1016/j.cmet.2008.08.001

Li, B., Dewey, C.N., 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics 12, 323. https://doi.org/10.1186/1471-2105-12-323

Li, H., Durbin, R., 2010. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinforma. Oxf. Engl. 26, 589–595. https://doi.org/10.1093/bioinformatics/btp698

Li, H., Durbin, R., 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics 25, 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Li, Q., Brown, J.B., Huang, H., Bickel, P.J., 2011. Measuring reproducibility of high-throughput experiments. Ann. Appl. Stat. 5, 1752–1779. https://doi.org/10.1214/11-AOAS466

Li, X., Brock, G.N., Rouchka, E.C., Cooper, N.G.F., Wu, D., O'Toole, T.E., Gill, R.S., Eteleeb, A.M., O'Brien, L., Rai, S.N., 2017. A comparison of per sample global scaling and per gene normalization methods for differential expression analysis of RNA-seq data. PLOS ONE 12, e0176185. https://doi.org/10.1371/journal.pone.0176185

Li, Z., Basterr, M.J., Hailemariam, T.K., Hojjati, M.R., Lu, S., Liu, J., Liu, R., Zhou, H., Jiang, X.-C., 2005. The effect of dietary sphingolipids on plasma sphingomyelin metabolism and atherosclerosis. Biochim. Biophys. Acta BBA - Mol. Cell Biol. Lipids 1735, 130–134. https://doi.org/10.1016/j.bbalip.2005.05.004

Liao, Y., Smyth, G.K., Shi, W., 2014. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinforma. Oxf. Engl. 30, 923–930. https://doi.org/10.1093/bioinformatics/btt656

Little, R.J.A., Rubin, D.B., 2014. Statistical Analysis with Missing Data. John Wiley & Sons.

Liu, D., Ahmet, A., Ward, L., Krishnamoorthy, P., Mandelcorn, E.D., Leigh, R., Brown, J.P., Cohen, A., Kim, H., 2013. A practical guide to the monitoring and management of the complications

of systemic corticosteroid therapy. Allergy Asthma Clin. Immunol. Off. J. Can. Soc. Allergy Clin. Immunol. 9, 30. https://doi.org/10.1186/1710-1492-9-30

Lorenzi, C., Barriere, S., Villemin, J.-P., Dejardin Bretones, L., Mancheron, A., Ritchie, W., 2020. iMOKA: k-mer based software to analyze large collections of sequencing data. Genome Biol. 21, 261. https://doi.org/10.1186/s13059-020-02165-2

Love, M.I., Huber, W., Anders, S., 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15. https://doi.org/10.1186/s13059-014-0550-8

Luo, R., Sedlazeck, F.J., Lam, T.-W., Schatz, M.C., 2019. A multi-task convolutional deep neural network for variant calling in single molecule sequencing. Nat. Commun. 10, 998. https://doi.org/10.1038/s41467-019-09025-z

Mandal, R.K., Nissar, K., Mittal, R.D., 2012. Genetic variants in metabolizing genes NQO1, NQO2, MTHFR and risk of prostate cancer: a study from North India. Mol. Biol. Rep. 39, 11145–11152. https://doi.org/10.1007/s11033-012-2023-z

Mangelsdorf, D.J., Thummel, C., Beato, M., Herrlich, P., Schütz, G., Umesono, K., Blumberg, B., Kastner, P., Mark, M., Chambon, P., Evans, R.M., 1995. The Nuclear Receptor Superfamily: The Second Decade. Cell 83, 835–839.

Matsumoto, M., Han, S., Kitamura, T., Accili, D., 2006. Dual role of transcription factor FoxO1 in controlling hepatic insulin sensitivity and lipid metabolism. J. Clin. Invest. 116, 2464–2472. https://doi.org/10.1172/JCI27047

Mehmood, A., Laiho, A., Venäläinen, M.S., McGlinchey, A.J., Wang, N., Elo, L.L., 2019. Systematic evaluation of differential splicing tools for RNA-seq studies. Brief. Bioinform. https://doi.org/10.1093/bib/bbz126

Metzker, M.L., 2010. Sequencing technologies — the next generation. Nat. Rev. Genet. 11, 31–46. https://doi.org/10.1038/nrg2626

Mi, H., Muruganujan, A., Ebert, D., Huang, X., Thomas, P.D., 2019a. PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. Nucleic Acids Res. 47, D419–D426. https://doi.org/10.1093/nar/gky1038

Mi, H., Muruganujan, A., Huang, X., Ebert, D., Mills, C., Guo, X., Thomas, P.D., 2019b. Protocol Update for large-scale genome and gene function analysis with the PANTHER classification system (v.14.0). Nat. Protoc. 14, 703–721. https://doi.org/10.1038/s41596-019-0128-8

Morozova, O., Hirst, M., Marra, M.A., 2009. Applications of New Sequencing Technologies for Transcriptome Analysis. Annu. Rev. Genomics Hum. Genet. 10, 135–151. https://doi.org/10.1146/annurev-genom-082908-145957

Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L., Wold, B., 2008. Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat. Methods 5, 621–628. https://doi.org/10.1038/nmeth.1226

Murtagh, F., Legendre, P., 2014. Ward's Hierarchical Agglomerative Clustering Method: Which Algorithms Implement Ward's Criterion? J. Classif. 31, 274–295. https://doi.org/10.1007/s00357-014-9161-z

Nader, N., Chrousos, G.P., Kino, T., 2009. Circadian rhythm transcription factor CLOCK regulates the transcriptional activity of the glucocorticoid receptor by acetylating its hinge region lysine

cluster: potential physiological implications. FASEB J. 23, 1572–1583. https://doi.org/10.1096/fj.08-117697

Nakato, R., Sakata, T., 2020. Methods for ChIP-seq analysis: A practical workflow and advanced applications. Methods. https://doi.org/10.1016/j.ymeth.2020.03.005

Nicholson, J.K., Lindon, J.C., Holmes, E., 1999. "Metabonomics": understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data. Xenobiotica 29, 1181–1189. https://doi.org/10.1080/004982599238047

Nowicka, M., Robinson, M.D., 2016. DRIMSeq: a Dirichlet-multinomial framework for multivariate count outcomes in genomics. F1000Research 5. https://doi.org/10.12688/f1000research.8900.2

Oakley, R.H., Cidlowski, J.A., 2013. The Biology of the Glucocorticoid Receptor: New Signaling Mechanisms in Health and Disease. J. Allergy Clin. Immunol. 132, 1033–1044. https://doi.org/10.1016/j.jaci.2013.09.007

Ofman, R., Dijkstra, I.M.E., van Roermund, C.W.T., Burger, N., Turkenburg, M., van Cruchten, A., van Engen, C.E., Wanders, R.J.A., Kemp, S., 2010. The role of ELOVL1 in very long-chain fatty acid homeostasis and X-linked adrenoleukodystrophy. EMBO Mol. Med. 2, 90–97. https://doi.org/10.1002/emmm.201000061

Oishi, K., Amagai, N., Shirai, H., Kadota, K., Ohkura, N., Ishida, N., 2005. Genome-wide Expression Analysis Reveals 100 Adrenal Gland-dependent Circadian Genes in the Mouse Liver. DNA Res. 12, 191–202. https://doi.org/10.1093/dnares/dsi003

Opherk, C., Tronche, F., Kellendonk, C., Kohlmüller, D., Schulze, A., Schmid, W., Schütz, G., 2004. Inactivation of the Glucocorticoid Receptor in Hepatocytes Leads to Fasting Hypoglycemia and Ameliorates Hyperglycemia in Streptozotocin-Induced Diabetes Mellitus. Mol. Endocrinol. 18, 1346–1353. https://doi.org/10.1210/me.2003-0283

Panda, S., Antoch, M.P., Miller, B.H., Su, A.I., Schook, A.B., Straume, M., Schultz, P.G., Kay, S.A., Takahashi, J.S., Hogenesch, J.B., 2002. Coordinated Transcription of Key Pathways in the Mouse by the Circadian Clock. Cell 109, 307–320. https://doi.org/10.1016/S0092-8674(02)00722-5

Parekh, S., Ziegenhain, C., Vieth, B., Enard, W., Hellmann, I., 2016. The impact of amplification on differential expression analyses by RNA-seq. Sci. Rep. 6, 25533. https://doi.org/10.1038/srep25533

Patro, R., Duggal, G., Love, M.I., Irizarry, R.A., Kingsford, C., 2017. Salmon provides fast and bias-aware quantification of transcript expression. Nat. Methods 14, 417–419. https://doi.org/10.1038/nmeth.4197

Pickrell, J.K., Marioni, J.C., Pai, A.A., Degner, J.F., Engelhardt, B.E., Nkadori, E., Veyrieras, J.-B., Stephens, M., Gilad, Y., Pritchard, J.K., 2010. Understanding mechanisms underlying human gene expression variation with RNA sequencing. Nature 464, 768–772. https://doi.org/10.1038/nature08872

Poplin, R., Chang, P.-C., Alexander, D., Schwartz, S., Colthurst, T., Ku, A., Newburger, D., Dijamco, J., Nguyen, N., Afshar, P.T., Gross, S.S., Dorfman, L., McLean, C.Y., DePristo, M.A., 2018. A universal SNP and small-indel variant caller using deep neural networks. Nat. Biotechnol.

36, 983–987. https://doi.org/10.1038/nbt.4235

Pratt, W.B., Toft, D.O., 1997. Steroid receptor interactions with heat shock protein and immunophilin chaperones [WWW Document]. Endocr. Rev. https://doi.org/10.1210/edrv.18.3.0303

Quagliarini, F., Mir, A.A., Balazs, K., Wierer, M., Dyar, K.A., Jouffe, C., Makris, K., Hawe, J., Heinig, M., Filipp, F.V., Barish, G.D., Uhlenhaut, N.H., 2019. Cistromic Reprogramming of the Diurnal Glucocorticoid Hormone Response by High-Fat Diet. Mol. Cell 76, 531-545.e5. https://doi.org/10.1016/j.molcel.2019.10.007

Quang, D., Xie, X., 2019. FactorNet: A deep learning framework for predicting cell type specific transcription factor binding from nucleotide-resolution sequential data. Methods, Deep Learning in Bioinformatics 166, 40–47. https://doi.org/10.1016/j.ymeth.2019.03.020

Quinlan, A.R., Hall, I.M., 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26, 841–842. https://doi.org/10.1093/bioinformatics/btq033

Rajapandi, T., Greene, L.E., Eisenberg, E., 2000. The molecular chaperones Hsp90 and Hsc70 are both necessary and sufficient to activate hormone binding by glucocorticoid receptor. J. Biol. Chem. 275, 22597–22604. https://doi.org/10.1074/jbc.M002035200

Rambout, X., Dequiedt, F., Maquat, L.E., 2018. Beyond Transcription: Roles of Transcription Factors in Pre-mRNA Splicing. Chem. Rev. 118, 4339–4364. https://doi.org/10.1021/acs.chemrev.7b00470

Reddy, A.B., Maywood, E.S., Karp, N.A., King, V.M., Inoue, Y., Gonzalez, F.J., Lilley, K.S., Kyriacou, C.P., Hastings, M.H., 2007. Glucocorticoid signaling synchronizes the liver circadian transcriptome. Hepatology 45, 1478–1488. https://doi.org/10.1002/hep.21571

Reue, K., Zhang, P., 2008. The lipin protein family: dual roles in lipid biosynthesis and gene expression. FEBS Lett. 582, 90–96. https://doi.org/10.1016/j.febslet.2007.11.014

Reyes, A., Huber, W., 2018. Alternative start and termination sites of transcription drive most transcript isoform differences across human tissues. Nucleic Acids Res. 46, 582–592. https://doi.org/10.1093/nar/gkx1165

Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., Smyth, G.K., 2015. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 43, e47. https://doi.org/10.1093/nar/gkv007

Roberts, A., Trapnell, C., Donaghey, J., Rinn, J.L., Pachter, L., 2011. Improving RNA-Seq expression estimates by correcting for fragment bias. Genome Biol. 12, R22. https://doi.org/10.1186/gb-2011-12-3-r22

Roberts, L.D., Souza, A.L., Gerszten, R.E., Clish, C.B., 2012. Targeted Metabolomics. Curr. Protoc. Mol. Biol. CHAPTER, Unit30.2. https://doi.org/10.1002/0471142727.mb3002s98

Robinson, M.D., McCarthy, D.J., Smyth, G.K., 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26, 139–140. https://doi.org/10.1093/bioinformatics/btp616

Robinson, M.D., Oshlack, A., 2010. A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biol. 11, R25. https://doi.org/10.1186/gb-2010-11-3-r25

Rose, A.J., Herzig, S., 2013. Metabolic control through glucocorticoid hormones: An update. Mol.

Cell. Endocrinol., Glucocorticoids 380, 65–78. https://doi.org/10.1016/j.mce.2013.03.007

Ross-Innes, C.S., Stark, R., Teschendorff, A.E., Holmes, K.A., Ali, H.R., Dunning, M.J., Brown, G.D., Gojis, O., Ellis, I.O., Green, A.R., Ali, S., Chin, S.-F., Palmieri, C., Caldas, C., Carroll, J.S., 2012. Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. Nature 481, 389–393. https://doi.org/10.1038/nature10730

Rozowsky, J., Euskirchen, G., Auerbach, R.K., Zhang, Z.D., Gibson, T., Bjornson, R., Carriero, N., Snyder, M., Gerstein, M.B., 2009. PeakSeq: Systematic Scoring of ChIP-Seq Experiments Relative to Controls. Nat. Biotechnol. 27, 66–75. https://doi.org/10.1038/nbt.1518

Rui, L., 2014. Energy Metabolism in the Liver. Compr. Physiol. 4, 177–197. https://doi.org/10.1002/cphy.c130024

Sacta, M.A., Chinenov, Y., Rogatsky, I., 2016. Glucocorticoid Signaling: An Update from a Genomic Perspective. Annu. Rev. Physiol. 78, 155–180. https://doi.org/10.1146/annurev-physiol-021115-105323

Sawilowsky, S., 2005. Misconceptions Leading to Choosing the t Test Over the Wilcoxon Mann-Whitney Test for Shift in Location Parameter. Theor. Behav. Found. Educ. Fac. Publ.

Scholz, M., Kaplan, F., Guy, C.L., Kopka, J., Selbig, J., 2005. Non-linear PCA: a missing data approach. Bioinformatics 21, 3887–3895. https://doi.org/10.1093/bioinformatics/bti634

Senga, S., Kobayashi, N., Kawaguchi, K., Ando, A., Fujii, H., 2018. Fatty acid-binding protein 5 (FABP5) promotes lipolysis of lipid droplets, de novo fatty acid (FA) synthesis and activation of nuclear factor-kappa B (NF-κB) signaling in cancer cells. Biochim. Biophys. Acta Mol. Cell Biol. Lipids 1863, 1057–1067. https://doi.org/10.1016/j.bbalip.2018.06.010

Shen, S., Park, J.W., Lu, Z., Lin, L., Henry, M.D., Wu, Y.N., Zhou, Q., Xing, Y., 2014. rMATS: Robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. Proc. Natl. Acad. Sci. 111, E5593–E5601. https://doi.org/10.1073/pnas.1419161111

Siersbæk, M., Varticovski, L., Yang, S., Baek, S., Nielsen, R., Mandrup, S., Hager, G.L., Chung, J.H., Grøntved, L., 2017. High fat diet-induced changes of mouse hepatic transcription and enhancer activity can be reversed by subsequent weight loss. Sci. Rep. 7, 40220. https://doi.org/10.1038/srep40220

Sims, D., Sudbery, I., Ilott, N.E., Heger, A., Ponting, C.P., 2014. Sequencing depth and coverage: key considerations in genomic analyses. Nat. Rev. Genet. 15, 121–132. https://doi.org/10.1038/nrg3642

Smith, B., Williams, J., Steffen, S.-K., 2003. The Ontology of the Gene Ontology. AMIA. Annu. Symp. Proc. 2003, 609–613.

So, A.Y.-L., Bernal, T.U., Pillsbury, M.L., Yamamoto, K.R., Feldman, B.J., 2009. Glucocorticoid regulation of the circadian clock modulates glucose homeostasis. Proc. Natl. Acad. Sci. 106, 17582–17587. https://doi.org/10.1073/pnas.0909733106

Soneson, C., Love, M.I., Robinson, M.D., 2016a. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. F1000Research 4, 1521. https://doi.org/10.12688/f1000research.7563.2

Soneson, C., Matthes, K.L., Nowicka, M., Law, C.W., Robinson, M.D., 2016b. Isoform prefiltering improves performance of count-based methods for analysis of differential transcript usage.

Genome Biol. 17, 12. https://doi.org/10.1186/s13059-015-0862-3

Stark, R., Brown, G., 2020. DiffBind: Differential Binding Analysis of ChIP-Seq Peak Data. Bioconductor version: Release (3.12). https://doi.org/10.18129/B9.bioc.DiffBind

Su, Z., Łabaj, P.P., Li, Sheng, Thierry-Mieg, J., Thierry-Mieg, D., Shi, W., Wang, C., Schroth, G.P., Setterquist, R.A., Thompson, J.F., Jones, W.D., Xiao, W., Xu, W., Jensen, R.V., Kelly, R., Xu, J., Conesa, A., Furlanello, C., Gao, Hanlin, Hong, H., Jafari, N., Letovsky, S., Liao, Y., Lu, F., Oakeley, E.J., Peng, Z., Praul, C.A., Santoyo-Lopez, J., Scherer, A., Shi, T., Smyth, G.K., Staedtler, F., Sykacek, P., Tan, X.-X., Thompson, E.A., Vandesompele, J., Wang, M.D., Wang, Jian, Wolfinger, R.D., Zavadil, J., Auerbach, S.S., Bao, W., Binder, H., Blomquist, T., Brilliant, M.H., Bushel, P.R., Cai, W., Catalano, J.G., Chang, C.-W., Chen, T., Chen, G., Chen, R., Chierici, M., Chu, T.-M., Clevert, D.-A., Deng, Y., Derti, A., Devanarayan, V., Dong, Z., Dopazo, J., Du, T., Fang, H., Fang, Y., Fasold, M., Fernandez, A., Fischer, M., Furió-Tari, P., Fuscoe, J.C., Caimet, F., Gaj, S., Gandara, J., Gao, Huan, Ge, W., Gondo, Y., Gong, B., Gong, M., Gong, Z., Green, B., Guo, C., Guo, L., Guo, L.-W., Hadfield, J., Hellemans, J., Hochreiter, S., Jia, M., Jian, M., Johnson, C.D., Kay, S., Kleinjans, J., Lababidi, S., Levy, S., Li, Q.-Z., Li, L., Li, L., Li, P., Li, Y., Li, H., Li, J., Li, Shiyong, Lin, S.M., López, F.J., Lu, X., Luo, H., Ma, X., Meehan, J., Megherbi, D.B., Mei, N., Mu, B., Ning, B., Pandey, A., Pérez-Florido, J., Perkins, R.G., Peters, R., Phan, J.H., Pirooznia, M., Qian, F., Qing, T., Rainbow, L., Rocca-Serra, P., Sambourg, L., Sansone, S.-A., Schwartz, S., Shah, R., Shen, J., Smith, T.M., Stegle, O., Stralis-Pavese, N., Stupka, E., Suzuki, Y., Szkotnicki, L.T., Tinning, M., Tu, B., van Delft, J., Vela-Boza, A., Venturini, E., Walker, S.J., Wan, L., Wang, W., Wang, Jinhui, Wang, Jun, Wieben, E.D., Willey, J.C., Wu, P.-Y., Xuan, J., Yang, Y., Ye, Z., Yin, Y., Yu, Y., Yuan, Y.-C., Zhang, J., Zhang, K.K., Zhang, Wenqian, Zhang, Wenwei, Zhang, Y., Zhao, C., Zheng, Y., Zhou, Y., Zumbo, P., Tong, W., Kreil, D.P., Mason, C.E., Shi, L., SEQC/MAQC-III Consortium, 2014. A comprehensive assessment of RNA-seq accuracy, reproducibility and information content by the Sequencing Quality Control Consortium. Nat. Biotechnol. 32, 903–914. https://doi.org/10.1038/nbt.2957

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., Mesirov, J.P., 2005. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. U. S. A. 102, 15545–15550. https://doi.org/10.1073/pnas.0506580102

Sugnet, C. w., Kent, W. j., Ares, M., Haussler, D., 2003. Transcriptome and genome conservation of alternative splicing events in humans and mice, in: Biocomputing 2004. WORLD SCIENTIFIC, pp. 66–77. https://doi.org/10.1142/9789812704856_0007

Sul, H.S., Wang, D., 1998. NUTRITIONAL AND HORMONAL REGULATION OF ENZYMES IN FAT SYNTHESIS: Studies of Fatty Acid Synthase and Mitochondrial Glycerol-3-Phosphate Acyltransferase Gene Transcription. Annu. Rev. Nutr. 18, 331–351. https://doi.org/10.1146/annurev.nutr.18.1.331

Sun, L., Ma, J., Turck, C.W., Xu, P., Wang, G.-Z., 2020. Genome-wide circadian regulation: A unique system for computational biology. Comput. Struct. Biotechnol. J. 18, 1914. https://doi.org/10.1016/j.csbj.2020.07.002

Takahashi, J.S., 2016. Molecular Architecture of the Circadian Clock in Mammals, in: Sassone-Corsi, P., Christen, Y. (Eds.), A Time for Metabolism and Hormones. Springer, Cham (CH).

Takahashi, J.S., Menaker, M., 1982. Role of the suprachiasmatic nuclei in the circadian system of the house sparrow, Passer domesticus. J. Neurosci. 2, 815–828.

https://doi.org/10.1523/JNEUROSCI.02-06-00815.1982

Tamez-Pérez, H.E., Quintanilla-Flores, D.L., Rodríguez-Gutiérrez, R., González-González, J.G., Tamez-Peña, A.L., 2015. Steroid hyperglycemia: Prevalence, early detection and therapeutic recommendations: A narrative review. World J. Diabetes 6, 1073–1081. https://doi.org/10.4239/wjd.v6.i8.1073

Tarazona, S., García-Alcalde, F., Dopazo, J., Ferrer, A., Conesa, A., 2011. Differential expression in RNA-seq: A matter of depth. Genome Res. 21, 2213–2223. https://doi.org/10.1101/gr.124321.111

Taves, M.D., Gomez-Sanchez, C.E., Soma, K.K., 2011. Extra-adrenal glucocorticoids and mineralocorticoids: evidence for local synthesis, regulation, and function. Am. J. Physiol. - Endocrinol. Metab. 301, E11–E24. https://doi.org/10.1152/ajpendo.00100.2011

Taylor, S.L., Leiserowitz, G.S., Kim, K., 2013. Accounting for undetected compounds in statistical analyses of mass spectrometry 'omic studies. Stat. Appl. Genet. Mol. Biol. 12, 703–722. https://doi.org/10.1515/sagmb-2013-0021

Thomas, R., Thomas, S., Holloway, A.K., Pollard, K.S., 2017. Features that define the best ChIP-seq peak calling algorithms. Brief. Bioinform. 18, 441–450. https://doi.org/10.1093/bib/bbw035

Torra, I.P., Tsibulsky, V., Delaunay, F., Saladin, R., Laudet, V., Fruchart, J.-C., Kosykh, V., Staels, B., 2000. Circadian and Glucocorticoid Regulation of Rev-erbα Expression in Liver. Endocrinology 141, 3799–3806. https://doi.org/10.1210/endo.141.10.7708

Toye, A.A., Dumas, M.E., Blancher, C., Rothwell, A.R., Fearnside, J.F., Wilder, S.P., Bihoreau, M.T., Cloarec, O., Azzouzi, I., Young, S., Barton, R.H., Holmes, E., McCarthy, M.I., Tatoud, R., Nicholson, J.K., Scott, J., Gauguier, D., 2007. Subtle metabolic and liver gene transcriptional changes underlie diet-induced fatty liver susceptibility in insulin-resistant mice. Diabetologia 50, 1867–1879. https://doi.org/10.1007/s00125-007-0738-5

Trapnell, C., Pachter, L., Salzberg, S.L., 2009. TopHat: discovering splice junctions with RNA-Seq. Bioinformatics 25, 1105–1111. https://doi.org/10.1093/bioinformatics/btp120

Trincado, J.L., Entizne, J.C., Hysenaj, G., Singh, B., Skalic, M., Elliott, D.J., Eyras, E., 2018. SUPPA2: fast, accurate, and uncertainty-aware differential splicing analysis across multiple conditions. Genome Biol. 19, 40. https://doi.org/10.1186/s13059-018-1417-1

Tu, S., Shao, Z., 2017. An introduction to computational tools for differential binding analysis with ChIP-seq data. Quant. Biol. 5, 226–235. https://doi.org/10.1007/s40484-017-0111-8

Turpin, S.M., Nicholls, H.T., Willmes, D.M., Mourier, A., Brodesser, S., Wunderlich, C.M., Mauer, J., Xu, E., Hammerschmidt, P., Brönneke, H.S., Trifunovic, A., LoSasso, G., Wunderlich, F.T., Kornfeld, J.-W., Blüher, M., Krönke, M., Brüning, J.C., 2014. Obesity-Induced CerS6-Dependent C16:0 Ceramide Production Promotes Weight Gain and Glucose Intolerance. Cell Metab. 20. https://doi.org/10.1016/j.cmet.2014.08.002

Van den Berge, K., Soneson, C., Robinson, M.D., Clement, L., 2017. stageR: a general stage-wise method for controlling the gene-level false discovery rate in differential expression and differential transcript usage. Genome Biol. 18, 151. https://doi.org/10.1186/s13059-017-1277-0

Vegiopoulos, A., Herzig, S., 2007. Glucocorticoids, metabolism and metabolic diseases. Mol. Cell. Endocrinol., Glucocorticoid Receptor Action and Selective Glucocorticoid Receptor Agonists

(SEGRAs) 275, 43–61. https://doi.org/10.1016/j.mce.2007.05.015

Vinayavekhin, N., Saghatelian, A., 2010. Untargeted Metabolomics. Curr. Protoc. Mol. Biol. 90, 30.1.1-30.1.24. https://doi.org/10.1002/0471142727.mb3001s90

Vitting-Seerup, K., Sandelin, A., 2017. The Landscape of Isoform Switches in Human Cancers. Mol. Cancer Res. 15, 1206–1220. https://doi.org/10.1158/1541-7786.MCR-16-0459

Wadapurkar, R.M., Vyas, R., 2018. Computational analysis of next generation sequencing data and its applications in clinical oncology. Inform. Med. Unlocked 11, 75–82. https://doi.org/10.1016/j.imu.2018.05.003

Walker, J.J., Terry, J.R., Lightman, S.L., 2010. Origin of ultradian pulsatility in the hypothalamic–pituitary–adrenal axis. Proc. R. Soc. B Biol. Sci. 277, 1627–1633. https://doi.org/10.1098/rspb.2009.2148

Wang, C., Tao, Q., Wang, Xinghe, Wang, Xiurong, Zhang, X., 2016. Impact of high-fat diet on liver genes expression profiles in mice model of nonalcoholic fatty liver disease. Environ. Toxicol. Pharmacol. 45, 52–62. https://doi.org/10.1016/j.etap.2016.05.014

Wang, E.T., Sandberg, R., Luo, S., Khrebtukova, I., Zhang, L., Mayr, C., Kingsmore, S.F., Schroth, G.P., Burge, C.B., 2008. Alternative isoform regulation in human tissue transcriptomes. Nature 456, 470–476. https://doi.org/10.1038/nature07509

Wang, J.-C., Gray, N.E., Kuo, T., Harris, C.A., 2012. Regulation of triglyceride metabolism by glucocorticoid receptor. Cell Biosci. 2, 19. https://doi.org/10.1186/2045-3701-2-19

Wei, R., Wang, J., Su, M., Jia, E., Chen, S., Chen, T., Ni, Y., 2018. Missing Value Imputation Approach for Mass Spectrometry-based Metabolomics Data. Sci. Rep. 8, 663. https://doi.org/10.1038/s41598-017-19120-0

Williams, R.A., Mamotte, C.D., Burnett, J.R., 2008. Phenylketonuria: An Inborn Error of Phenylalanine Metabolism. Clin. Biochem. Rev. 29, 31–41.

Wood, J.D., 2006. Histamine, mast cells, and the enteric nervous system in the irritable bowel syndrome, enteritis, and food allergies. Gut 55, 445–447. https://doi.org/10.1136/gut.2005.079046

Wu, G., Anafi, R.C., Hughes, M.E., Kornacker, K., Hogenesch, J.B., 2016. MetaCycle: an integrated R package to evaluate periodicity in large scale data. Bioinformatics 32, 3351–3353. https://doi.org/10.1093/bioinformatics/btw405

Xia, J., Fjell, C.D., Mayer, M.L., Pena, O.M., Wishart, D.S., Hancock, R.E.W., 2013. INMEX—a web-based tool for integrative meta-analysis of expression data. Nucleic Acids Res. 41, W63–W70. https://doi.org/10.1093/nar/gkt338

Yang, R., Su, Z., 2010. Analyzing circadian expression data by harmonic regression based on autoregressive spectral estimation. Bioinformatics 26, i168–i174. https://doi.org/10.1093/bioinformatics/btq189

Ying, H., Yu, Y., Xu, Y., 2001. Cloning and Characterization of F-LANa, Upregulated in Human Liver Cancer. Biochem. Biophys. Res. Commun. 286, 394–400. https://doi.org/10.1006/bbrc.2001.5390

Zhang, J., Gao, X., Yuan, Y., Sun, C., Zhao, Y., Xiao, L., Yang, Y., Gu, Y., Yang, R., Hu, P., Zhang, L.,

Wang, C., Ye, J., 2019. Perilipin 5 alleviates HCV NS5A-induced lipotoxic injuries in liver. Lipids Health Dis. 18, 87. https://doi.org/10.1186/s12944-019-1022-7

Zhang, Y., An, L., Xu, J., Zhang, B., Zheng, W.J., Hu, M., Tang, J., Yue, F., 2018. Enhancing Hi-C data resolution with deep convolutional neural network HiCPlus. Nat. Commun. 9, 750. https://doi.org/10.1038/s41467-018-03113-2

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., Liu, X.S., 2008. Model-based Analysis of ChIP-Seq (MACS). Genome Biol. 9, R137. https://doi.org/10.1186/gb-2008-9-9-r137

Zhou, G., Soufan, O., Ewald, J., Hancock, R.E.W., Basu, N., Xia, J., 2019. NetworkAnalyst 3.0: a visual analytics platform for comprehensive gene expression profiling and meta-analysis. Nucleic Acids Res. 47, W234–W241. https://doi.org/10.1093/nar/gkz240

Zhu, L.J., Gazin, C., Lawson, N.D., Pagès, H., Lin, S.M., Lapointe, D.S., Green, M.R., 2010. ChIPpeakAnno: a Bioconductor package to annotate ChIP-seq and ChIP-chip data. BMC Bioinformatics 11, 237. https://doi.org/10.1186/1471-2105-11-237

Zimmermann, M., Kogadeeva, M., Gengenbacher, M., McEwen, G., Mollenkopf, H.-J., Zamboni, N., Kaufmann, S.H.E., Sauer, U., 2017. Integration of Metabolomics and Transcriptomics Reveals a Complex Diet of Mycobacterium tuberculosis during Early Macrophage Infection. mSystems 2. https://doi.org/10.1128/mSystems.00057-17

Zwang, S., 2020. THE GLUCOCORTICOID RECEPTOR AND ITS IMPACT ON ALTERNATIVE SPLICING (Bachelor's Thesis). Technischen Universität München.

Supplementary data

Product #D12329		gm%	kcal%
Protein		16.8	16.4
Carbohydrate		74.3	73.1
Fat		4.8	10.5
	Total		100
	kcal/gm	4.07	
Ingredient		**gm**	**kcal**
Casein, 30 Mesh		228	912
DL-Methionine		2	0
Maltodextrin 10		170	680
Corn Starch		0	0
Sucrose		835	3340
Soybean Oil		25	225
Coconut Oil, Hydrogenated		40	360
Mineral Mix S10001		40	0
Sodium Bicarbonate		10.5	0
Potassium Citrate, 1 H2O		4	0
Vitamin Mix V10001		10	40
Choline Bitartrate		2	0
FD&C Blue Dye #1		0.1	0
Total		**1366.6**	**5557**

Product #D12331		gm%	kcal%
Protein		23.0	16.4
Carbohydrate		35.5	25.5
Fat		35.8	58.0
	Total		100
	kcal/gm	5.56	
Ingredient		**gm**	**kcal**
Casein, 30 Mesh		228	912
DL-Methionine		2	0
Maltodextrin 10		170	680
Corn Starch		0	0
Sucrose		175	700
Soybean Oil		25	225
Coconut Oil, Hydrogenated		333.5	3001.5
Mineral Mix S10001		40	0
Sodium Bicarbonate		10.5	0
Potassium Citrate, 1 H2O		4	0
Vitamin Mix V10001		10	40
Choline Bitartrate		2	0
FD&C Red Dye #4		0.1	0
Total		**1000.1**	**5558.5**

Figure S1: Formula for the used control (left) and high-fat diet (right).

By Research Diets, Inc. (20 Jules Lane, New Brunswick, NJ 08901 USA).

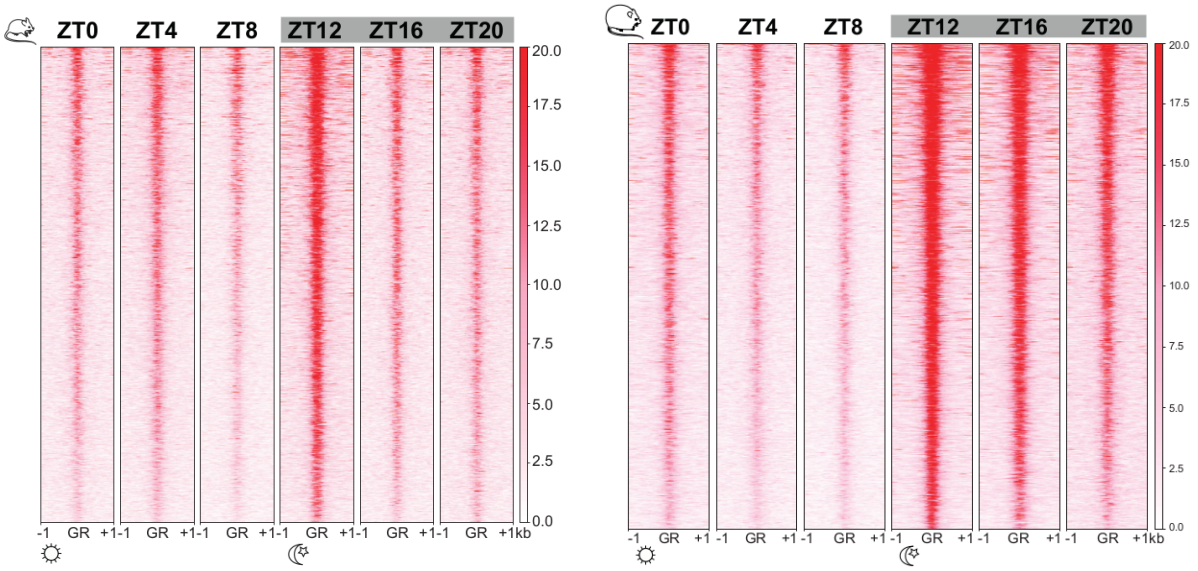


Figure S2: Heatmaps of GR genome-wide binding for all six time-points in control (left) and HFD (right).

Rows show the normalized unique tag counts for GR ChIP peaks, sorted according to signal strength. Image already published in (Quagliarini et al., 2019).

NAME: Diet_TP_Replicate	% Uniquely mapped	Uniquely mapped	Unmapped too short	Unmapped other	Mapped to too many loci	Multimapped	Total reads
HFD_ZT0_R1	82.62	44796696	867608	3557193	135202	4865579	54222278
HFD_ZT0_R2	81	34464623	1152338	3176370	109565	3648069	42550965
HFD_ZT0_R3	81.7	35038853	1090356	3176629	111029	3472583	42889450
HFD_ZT4_R1	80.1	40032830	764352	4256393	128041	4795059	49976675
HFD_ZT4_R2	79.35	31336053	924071	3862145	99930	3266419	39488618
HFD_ZT4_R3	79.56	34181517	1095388	3986352	111313	3586793	42961363
HFD_ZT8_R1	82.21	37693886	719318	3349180	101738	3989122	45853244
HFD_ZT8_R2	85.66	49139220	992844	2295594	112490	4828582	57368730
HFD_ZT8_R3	86.03	37565684	1048726	1809051	79857	3164539	43667857
HFD_ZT12_R1	79.53	41544904	752221	4884210	135011	4921723	52238069
HFD_ZT12_R2	86.59	47167498	854828	1764103	99746	4584830	54471005
HFD_ZT12_R3	86.93	33649680	1084942	1061693	69263	2843413	38708991
HFD_ZT16_R1	87.08	49477749	789736	1795371	86246	4669695	56818797
HFD_ZT16_R2	84.48	52347176	1122252	3025742	142042	5326607	61963819
HFD_ZT16_R3	80.65	30470265	1606366	2422777	90530	3192923	37782861
HFD_ZT20_R1	86.65	33582179	1011264	1131375	64919	2968054	38757791
HFD_ZT20_R2	81.59	47771821	907650	4227895	143614	5497442	58548422
HFD_ZT20_R3	82.85	33383233	1031498	2373250	82669	3422354	40293004
CTRL_ZT0_R1	76.51	40157951	803537	6349522	176169	5000058	52487237
CTRL_ZT0_R2	84.34	47676134	994008	2648807	124261	5083389	56526599
CTRL_ZT0_R3	83.71	36575135	1162674	2211702	101495	3640906	43691912
CTRL_ZT4_R1	85.67	45122811	847336	2210440	122522	4367766	52670875
CTRL_ZT4_R2	85.46	46291678	855895	2746448	123006	4152232	54169259
CTRL_ZT4_R3	86.64	35417035	1043746	1399847	76073	2941908	40878609
CTRL_ZT8_R1	81.22	49810799	987829	4552605	164600	5814853	61330686
CTRL_ZT8_R2	89.01	46497240	909428	898975	82231	3849469	52237343
CTRL_ZT8_R3	83.91	28813505	1071733	1641951	71199	2738588	34336976
CTRL_ZT12_R1	91.67	50066960	1044348	54678	90735	3361429	54618150
CTRL_ZT12_R2	86.43	53615847	1011262	2245870	115053	5043293	62031325
CTRL_ZT12_R3	82.18	33531670	917960	2810998	95902	3447344	40803874
CTRL_ZT16_R1	87.86	52070899	1015607	1146269	115051	4917160	59264986
CTRL_ZT16_R2	88.34	52969185	989500	1205390	109568	4688805	59962448
CTRL_ZT16_R3	89.57	36695087	974627	389032	77734	2833726	40970206
CTRL_ZT20_R1	83.41	45254206	830463	3386984	124424	4661403	54257480
CTRL_ZT20_R2	87	50656386	973821	1702730	108187	4787172	58228296
CTRL_ZT20_R3	84.86	34117174	959975	1831584	74393	3221201	40204327

Table S1: Mapping statistics of the first cohort generated with MultiQC.

NAME: TP_Diet_Genotype_ Replicate	% Uniquely mapped	Uniquely mapped	Unmapped too short	Unmapped other	Mapped to too many loci	Multimapped	Total reads
ZT00-HFD-KO-1	83.94	14036696	127081	1615262	35452	906888	16721379
ZT00-HFD-KO-2	90.92	15159086	136679	748401	32473	597100	16673739
ZT00-HFD-KO-3	86.15	12664836	129374	835054	35733	1036289	14701286
ZT00-HFD-WT-1	86.12	12712236	147693	781296	28101	1092467	14761793
ZT00-HFD-WT-2	80.11	10765865	127715	1492255	39962	1013668	13439465
ZT00-HFD-WT-3	88.11	12339920	123274	837700	29134	675319	14005347
ZT00-CTRL-KO-1	86.46	13482027	317939	782379	29640	981202	15593187
ZT00-CTRL-KO-2	79.09	13369704	187643	1903478	43021	1401610	16905456
ZT00-CTRL-KO-3	80.44	13069867	212728	1599522	42308	1322926	16247351

ZT00-CTRL-WT-1	83.56	14538328	144427	1614794	37842	1064149	17399540
ZT00-CTRL-WT-2	83.15	13703939	464826	1066461	37519	1209115	16481860
ZT00-CTRL-WT-3	81.45	13097017	364748	1298309	42677	1276855	16079606
ZT04-HFD-KO-1	77.53	9911702	423067	1327994	26673	1094459	12783895
ZT04-HFD-KO-2	82.73	13745027	204280	1355227	27576	1281941	16614051
ZT04-HFD-KO-3	81.1	14411718	202702	1861660	30257	1264127	17770464
ZT04-HFD-WT-1	83.57	12134920	329504	863679	24635	1168084	14520822
ZT04-HFD-WT-2	82.28	12339468	154495	1241961	27335	1234004	14997263
ZT04-HFD-WT-3	90.54	13249719	112678	659968	24160	587845	14634370
ZT04-CTRL-KO-1	89.02	16827944	360712	430589	31610	1251656	18902511
ZT04-CTRL-KO-2	78.6	11741233	376414	1626646	40702	1152519	14937514
ZT04-CTRL-KO-3	80.89	13577611	513733	1388422	37436	1268324	16785526
ZT04-CTRL-WT-1	81.57	12648747	426404	1049729	33395	1349194	15507469
ZT04-CTRL-WT-2	80.5	13212308	283819	1469952	38879	1406862	16411820
ZT04-CTRL-WT-3	78.93	11026340	335207	1282169	35079	1290166	13968961
ZT08-HFD-KO-1	76.01	12627995	342365	2250299	40943	1352070	16613672
ZT08-HFD-KO-2	79.46	14495844	434217	1857281	38617	1417054	18243013
ZT08-HFD-WT-1	73.76	11812327	467589	2310724	44680	1379358	16014678
ZT08-HFD-WT-2	80.54	13782782	210531	1656860	36257	1426572	17113002
ZT08-CTRL-KO-1	84.54	14158503	142439	1176376	30791	1238903	16747012
ZT08-CTRL-KO-2	73.14	11561460	303428	2557009	52720	1332000	15806617
ZT08-CTRL-WT-1	79.56	13319649	465435	1495083	44555	1417078	16741800
ZT08-CTRL-WT-2	81.75	11858558	336753	1066867	33351	1210232	14505761
ZT12-HFD-KO-1	79.5	12481760	342223	1521168	38663	1315943	15699757
ZT12-HFD-KO-3	78.59	11101799	373082	1447104	36218	1167461	14125664
ZT12-HFD-KO-4	73.62	12621396	437185	2547673	49993	1488828	17145075
ZT12-HFD-WT-1	84.94	15673298	627370	719630	36934	1394641	18451873
ZT12-HFD-WT-2	77.19	12460613	434480	1762147	39972	1444959	16142171
ZT12-HFD-WT-3	81.06	13554256	290740	1396891	37251	1442851	16721989
ZT12-CTRL-KO-1	84.94	14903122	518009	790184	30651	1302872	17544838
ZT12-CTRL-KO-2	88.78	13052839	465860	224847	25412	933886	14702844
ZT12-CTRL-KO-3	80.06	12047396	335366	1419665	34171	1211445	15048043
ZT12-CTRL-WT-1	78.85	12249356	462873	1486475	39899	1296022	15534625
ZT12-CTRL-WT-2	77.33	11900101	358665	1797940	42730	1289169	15388605
ZT12-CTRL-WT-3	80.12	13468689	480761	1464135	36283	1360233	16810101
ZT16-HFD-KO-1	85.14	14499756	411955	726879	31429	1360054	17030073
ZT16-HFD-KO-2	83.99	14175072	533042	816432	32331	1319416	16876293
ZT16-HFD-KO-3	85.29	12920653	413880	683735	32977	1098120	15149365
ZT16-HFD-WT-1	84.52	14305738	487015	779562	32621	1321276	16926212
ZT16-HFD-WT-2	82.43	13925037	478177	1013802	34949	1441748	16893713
ZT16-HFD-WT-3	79.4	12994192	410981	1493284	42777	1423349	16364583
ZT16-CTRL-KO-1	85.54	13075166	432595	539598	28646	1210295	15286300
ZT16-CTRL-KO-2	84.07	11586367	399457	713513	31601	1051107	13782045
ZT16-CTRL-WT-1	82.4	11321266	339297	938218	38078	1103029	13739888
ZT16-CTRL-WT-2	81.32	11999100	214100	1315609	30179	1195642	14754630
ZT16-CTRL-WT-3	82.19	16143099	538367	1304656	47811	1607716	19641649
ZT20-HFD-KO-1	82.85	13968601	485724	1028790	34730	1342150	16859995
ZT20-HFD-KO-2	80.66	13436715	351739	1371950	42332	1455113	16657849
ZT20-HFD-KO-3	85.26	15021747	492889	785103	32881	1286813	17619433
ZT20-HFD-WT-1	83.02	14805437	304854	1315684	35454	1371559	17832988
ZT20-HFD-WT-2	84.27	13591390	406492	774271	30466	1325155	16127774
ZT20-HFD-WT-3	80.26	11791299	408697	1271664	36296	1184083	14692039
ZT20-CTRL-KO-1	84.54	12397904	432728	726104	33644	1073958	14664338
ZT20-CTRL-KO-2	87.51	13309285	360159	518203	29713	991221	15208581

ZT20-CTRL-KO-3	81.22	11747781	354473	1163251	28478	1170591	14464574
ZT20-CTRL-WT-1	79.97	13709650	404327	1473397	34179	1522122	17143675
ZT20-CTRL-WT-2	82.89	11991154	372938	810922	31629	1259952	14466595
ZT20-CTRL-WT-3	80.17	12977630	433867	1408449	24070	1343833	16187849

Table S2: Mapping statistics of the second cohort generated with MultiQC.

NAME: Treatment_Diet_TP_ Replicate	% Uniquely mapped	Uniquely mapped	Unmapped too short	Unmapped other	Mapped to too many loci	Multimapped	Total reads
UNTR_HFD_day_R1	77.59	40599189	1569160	5460679	104188	4594956	52328172
UNTR_HFD_day_R2	76.63	48676669	1962860	7489358	133455	5259685	63522027
UNTR_HFD_night_R1	85.78	47370994	950951	2609585	97330	4193538	55222398
UNTR_HFD_night_R2	84	48515654	560215	4025459	123407	4529908	57754643
UNTR_CTRL_day_R1	84.74	44513471	1637517	2550747	100758	3724441	52526934
UNTR_CTRL_day_R2	82.34	44603026	867219	2858762	79806	5760446	54169259
UNTR_CTRL_day_R3	86.14	48004653	1553406	2171428	90185	3908322	55727994
UNTR_CTRL_night_R1	86.03	46641128	1469981	1686953	83628	4330448	54212138
UNTR_CTRL_night_R2	88.38	48485893	1798804	630678	91875	3853810	54861060
DEX_HFD_night_R1	81.15	42364386	757325	3603825	192044	5289828	52207408
DEX_HFD_night_R2	80.5	45284622	1485415	3342184	157424	5982897	56252542
DEX_HFD_night_R3	81.28	43165160	1847365	2548090	148465	5400251	53109331
DEX_HFD_day_R1	83.11	40318809	1513634	2047287	109548	4523790	48513068
DEX_HFD_day_R2	81.93	24148696	1270992	1262145	67992	2725941	29475766
DEX_HFD_day_R3	83.85	44213619	1612943	2213844	113535	4574013	52727954
DEX_CTRL_night_R1	83.69	48556359	1421751	1340508	183811	6516328	58018757
DEX_CTRL_night_R2	88.47	47607569	857170	318070	168838	4857801	53809448
DEX_CTRL_night_R3	85.46	46291678	855895	2746448	123006	4152232	54169259
DEX_CTRL_day_R1	87.1	42841546	1968516	310041	118887	3946995	49185985
DEX_CTRL_day_R2	85.33	45770883	2053171	745145	132275	4940374	53641848
DEX_CTRL_day_R3	86.69	45277522	2014167	491771	127389	4321350	52232199

Table S3: Mapping statistics of the third cohort generated with MultiQC.

Marked lines are the samples used in the differential transcript usage analysis. The line marked in green shows the mapping statistics when aligned to the mm10 genome. All other samples show the results when mapped to mm9. All marked samples were later also mapped to mm10 when used for DTU analysis.

Metabolite name	Padj. genotype	Padj. diet	Coef. genotype	Coef. diet
Ala	0.495	0.368	-0.608	0.840
Asn	0.474	0.511	0.726	0.635
Asp	0.955	0.307	-0.051	-0.971
Cit	0.944	0.546	0.090	-0.655
Gln	0.822	0.472	-0.283	-0.746
Glu	0.322	0.913	0.996	-0.124
Gly	0.907	0.554	0.142	-0.635
His	0.608	0.467	-0.517	-0.743
Ile	0.730	0.659	0.410	0.467
Lys	0.488	0.448	0.637	0.754
Met	0.490	0.725	0.670	-0.362

Orn	0.584	0.961	0.586	0.077
Phe	0.831	0.618	0.271	0.514
Pro	0.831	0.618	-0.280	0.511
Ser	0.732	0.578	-0.396	-0.584
Thr	0.902	0.159	0.122	1.205
Trp	0.925	0.554	-0.120	-0.645
Tyr	0.944	0.467	0.091	-0.772
Val	0.902	0.471	0.144	0.764
xLeu	0.542	0.596	0.601	0.531
ADMA	0.180	0.962	1.243	-0.048
alpha-AAA	0.402	0.659	0.852	-0.431
Creatinine	0.955	0.034	-0.037	-1.567
Histamine	0.396	0.023	0.537	1.543
Putrescine	0.322	0.682	-0.981	-0.375
Sarcosine	0.925	0.082	0.091	1.355
SDMA	0.410	0.372	0.781	-0.803
Serotonin	0.807	0.681	-0.328	0.439
Spermidine	0.344	0.313	0.860	-0.857
Spermine	0.163	0.618	1.262	-0.393
t4-OH-Pro	0.403	0.045	0.566	-1.455
Taurine	0.335	0.472	-0.921	-0.660
AC(0:0)	0.180	0.266	-1.132	0.827
AC(2:0)	0.498	0.322	-0.590	-0.901
AC(3:0)	0.740	0.249	0.336	1.037
AC(3:0-DC)	0.147	0.159	-1.160	0.956
AC(4:0)	0.474	0.245	-0.645	-0.998
AC(4:0-DC)	0.145	0.018	-0.872	1.459
AC(4:0-OH)	0.134	0.047	-1.025	1.273
AC(4:1-DC)	0.402	0.578	-0.857	0.536
AC(5:0)	0.488	0.472	-0.646	0.709
AC(5:0-DC)	0.402	0.061	-0.625	1.349
AC(5:0-OH)	0.484	0.847	-0.729	-0.211
AC(6:0)	0.624	0.926	-0.538	-0.115
AC(6:0-OH)	0.281	0.159	-0.918	1.069
AC(7:0)	0.740	0.488	-0.370	0.719
AC(7:0-DC)	0.403	0.179	0.714	-1.090
AC(9:0)	0.847	0.962	-0.255	0.057
AC(10:0)	0.960	0.511	0.049	0.693
AC(10:2)	0.105	0.511	1.395	0.470
AC(12:0)	0.304	0.107	0.830	1.177
AC(12:0-DC)	0.488	0.800	0.707	-0.292
AC(14:0)	0.876	0.982	0.204	0.027
AC(14:0-OH)	0.884	0.913	0.181	0.145
AC(14:1-DC)	0.322	0.660	1.007	0.401
AC(14:1-OH)	0.871	0.847	-0.218	0.223
AC(14:2)	0.630	0.554	0.504	0.617
AC(16:0)	0.609	0.307	-0.485	-0.944
AC(16:0-OH)	0.488	0.800	0.685	-0.289
AC(16:1)	0.455	0.197	-0.657	-1.079
AC(16:2)	0.676	0.684	-0.462	-0.420
AC(17:0)	0.571	0.807	-0.592	0.280
AC(18:0)	0.488	0.971	-0.698	0.040
AC(18:1)	0.161	0.179	-1.136	-0.945
AC(18:1-OH)	0.180	0.807	-1.224	-0.225
AC(18:2)	0.566	0.413	-0.550	-0.798
AC(19:0)	0.922	0.913	0.135	0.150
LPC(12:0)	0.134	0.511	1.334	-0.499
LPC(15:0)	0.781	0.066	0.247	1.398
LPC(16:0)	0.809	0.047	0.207	1.490
LPC(16:1)	0.376	0.264	0.810	0.934
LPC(17:0)	0.283	0.060	0.771	-1.330
LPC(17:1)	0.484	0.426	0.673	0.766
LPC(18:0)	0.692	0.390	-0.412	0.837
LPC(18:1)	0.322	0.362	-0.926	0.789
LPC(18:2)	0.773	0.554	0.349	0.623
LPC(20:0)	0.001	0.060	1.691	0.678

LPC(20:1)	0.455	0.618	-0.770	0.483
LPC(20:3)	0.180	0.047	-0.891	1.328
LPC(20:4)	0.163	0.661	-1.263	0.352
LPC(22:6)	0.467	0.238	-0.658	1.018
LPC-O(18:0)	0.871	0.758	-0.216	-0.355
LPC-O(18:1)	0.831	0.396	-0.251	-0.842
LPC-O(18:2)	0.621	0.574	-0.518	0.582
PC(24:0)	0.477	0.184	0.610	1.104
PC(25:0)	0.283	0.962	-1.074	0.051
PC(27:1)	0.831	0.240	-0.232	-1.065
PC(29:0)	0.344	0.001	0.388	1.752
PC(29:1)	0.467	0.236	0.661	1.023
PC(29:2)	0.389	0.574	0.878	0.538
PC(30:0)	0.134	0.000	0.376	1.842
PC(30:1)	0.056	0.002	0.980	1.514
PC(30:2)	0.449	0.295	0.705	0.922
PC(31:0)	0.402	0.159	0.723	-1.123
PC(31:1)	0.030	0.362	1.596	0.503
PC(31:2)	0.322	0.776	1.003	0.291
PC(31:3)	0.163	0.962	1.295	0.056
PC(32:0)	0.474	0.511	-0.712	-0.639
PC(32:1)	0.105	0.023	1.006	1.363
PC(32:2)	0.180	0.105	1.015	1.113
PC(32:4)	0.979	0.309	-0.022	0.965
PC(33:0)	0.488	0.022	0.406	1.581
PC(33:1)	0.105	0.307	1.360	0.684
PC(33:2)	0.076	0.979	1.533	-0.020
PC(33:4)	0.608	0.472	-0.521	0.720
PC(33:5)	0.871	0.448	-0.197	0.792
PC(34:1)	0.184	0.005	-0.671	1.601
PC(34:2)	0.938	0.372	-0.096	0.876
PC(34:3)	0.322	0.937	1.014	-0.087
PC(34:4)	0.656	0.044	0.308	1.509
PC(34:5)	0.187	0.060	0.909	1.268
PC(35:1)	0.498	0.996	0.671	-0.008
PC(35:2)	0.595	0.214	0.476	-1.078
PC(35:3)	0.719	0.989	0.433	-0.018
PC(35:4)	0.432	0.511	-0.786	0.626
PC(35:5)	0.143	0.714	-1.341	0.286
PC(36:1)	0.343	0.082	-0.729	1.261
PC(36:2)	0.283	0.245	-0.938	0.927
PC(36:3)	0.207	0.514	-1.134	0.549
PC(36:4)	0.119	0.571	-1.377	0.426
PC(36:5)	0.539	0.322	0.555	0.905
PC(36:6)	0.648	0.062	0.351	1.398
PC(37:1)	0.488	0.295	0.610	0.942
PC(37:2)	0.283	0.648	1.061	-0.409
PC(37:3)	0.474	0.863	-0.759	-0.191
PC(37:4)	0.259	0.245	-0.996	-0.909
PC(37:5)	0.635	0.833	-0.521	-0.247
PC(37:6)	0.872	0.901	0.212	-0.169
PC(37:7)	0.161	0.926	-1.313	0.087
PC(38:0)	0.226	0.839	1.157	0.195
PC(38:1)	0.955	0.060	0.042	1.433
PC(38:2)	0.402	0.578	0.841	0.533
PC(38:3)	0.539	0.060	0.424	1.420
PC(38:4)	0.077	0.593	-1.497	0.355
PC(38:5)	0.283	0.714	-1.080	-0.328
PC(38:6)	0.322	0.118	-0.807	1.160
PC(38:7)	0.944	0.818	0.092	0.273
PC(39:1)	0.438	0.843	-0.810	-0.215
PC(39:2)	0.860	0.919	0.239	-0.136
PC(39:3)	0.875	0.082	-0.146	1.360
PC(39:4)	0.488	0.937	-0.686	-0.098
PC(39:5)	0.773	0.413	-0.336	-0.818
PC(39:6)	0.822	0.362	-0.271	-0.896

PC(39:7)	0.955	0.493	0.058	-0.724
PC(40:2)	0.938	0.060	0.071	1.461
PC(40:3)	0.534	0.035	-0.385	1.528
PC(40:4)	0.928	0.060	0.081	1.434
PC(40:5)	0.809	0.962	-0.330	0.067
PC(40:6)	0.163	0.066	-1.002	1.201
PC(40:7)	0.166	0.818	-1.271	0.206
PC(40:8)	0.488	0.839	-0.690	-0.227
PC(41:1)	0.801	0.060	0.229	1.430
PC(41:2)	0.438	0.666	-0.797	0.420
PC(41:4)	0.376	0.832	0.925	0.228
PC(41:5)	0.000	0.680	1.911	0.030
PC(42:0)	0.955	0.264	0.061	1.033
PC(42:1)	0.401	0.090	0.682	1.250
PC(42:2)	0.676	0.088	-0.348	1.322
PC(42:4)	0.467	0.593	-0.749	0.522
PC(42:5)	0.979	0.805	-0.023	0.300
PC(42:6)	0.847	0.082	0.182	1.354
PC(42:7)	0.272	0.908	-1.115	0.129
PC(43:6)	0.193	0.567	-1.163	-0.491
PC(44:1)	0.435	0.923	0.824	0.116
PC(44:3)	0.676	0.472	-0.441	0.735
PC(44:12)	0.955	0.786	-0.061	-0.331
PC(46:1)	0.322	0.800	1.008	0.261
PC(46:2)	0.736	0.603	0.393	0.534
PC-O(28:0)	0.740	0.593	-0.384	-0.562
PC-O(28:1)	0.672	0.681	-0.473	0.427
PC-O(30:0)	0.474	0.001	-0.274	1.786
PC-O(31:0)	0.823	0.587	-0.288	0.578
PC-O(31:3)	0.882	0.618	0.188	0.519
PC-O(32:0)	0.076	0.082	-1.333	-0.981
PC-O(32:1)	0.322	0.603	-1.004	0.465
PC-O(32:2)	0.488	0.489	0.668	0.684
PC-O(33:3)	0.171	0.456	1.202	0.614
PC-O(34:0)	0.482	0.923	0.739	0.120
PC-O(34:1)	0.076	0.201	-1.431	-0.757
PC-O(34:2)	0.402	0.786	-0.874	-0.292
PC-O(36:0)	0.884	0.368	0.157	-0.883
PC-O(36:2)	0.105	0.681	-1.423	0.296
PC-O(36:3)	0.077	0.047	-1.190	1.175
PC-O(36:4)	0.322	0.847	-0.995	0.190
PC-O(36:5)	0.488	0.800	-0.680	-0.290
PC-O(36:6)	0.488	0.240	-0.594	1.021
PC-O(37:6)	0.417	0.591	-0.814	0.524
PC-O(38:4)	0.187	0.082	-0.950	1.191
PC-O(38:5)	0.124	0.593	-1.371	-0.396
PC-O(38:6)	0.698	0.368	-0.401	-0.874
PC-O(40:3)	0.283	0.307	0.977	-0.834
PC-O(40:4)	0.831	0.472	-0.251	0.743
PC-O(40:5)	0.884	0.818	-0.184	-0.277
PC-O(40:6)	0.156	0.511	1.276	-0.519
PC-O(40:7)	0.161	0.618	-1.282	-0.392
PC-O(42:1)	0.998	0.509	-0.004	0.707
PC-O(42:2)	0.207	0.901	1.180	-0.134
PC-O(42:6)	0.322	0.207	0.847	1.010
Cer(34:0)	0.822	0.708	0.299	-0.402
Cer(34:1)	0.077	0.930	-1.511	0.070
Cer(38:1)	0.871	0.201	0.180	1.130
Cer(40:1)	0.884	0.005	0.082	1.711
Cer(42:1)	0.322	0.114	-0.800	1.170
Cer(42:2)	0.076	0.693	-1.527	0.256
Cer(43:1)	0.416	0.214	0.716	1.036
SM(30:1)	0.831	0.000	-0.068	1.855
SM(32:1)	0.283	0.000	-0.302	1.839
SM(32:2)	0.831	0.001	-0.105	1.776
SM(33:1)	0.831	0.699	-0.278	-0.412

SM(34:1)	0.028	0.388	-1.624	0.456
SM(34:2)	0.815	0.906	-0.322	0.161
SM(35:1)	0.822	0.368	-0.280	-0.876
SM(36:0)	0.955	0.998	0.066	0.002
SM(36:1)	0.105	0.002	-0.787	1.605
SM(36:2)	0.143	0.120	-1.145	1.019
SM(37:1)	0.105	0.109	-1.230	0.996
SM(38:2)	0.732	0.554	-0.395	0.622
SM(39:1)	0.488	0.199	0.575	1.091
SM(39:2)	0.028	0.114	-1.528	0.775
SM(40:1)	0.283	0.259	-0.962	-0.900
SM(40:2)	0.435	0.159	-0.665	1.128
SM(41:1)	0.656	0.735	-0.495	0.364
SM(41:2)	0.339	0.970	-0.975	-0.041
SM(42:1)	0.676	0.659	0.458	-0.462
SM(42:2)	0.477	0.159	-0.601	-1.144
SM(42:3)	0.884	0.362	-0.158	-0.903
SM(43:1)	0.180	0.060	-0.929	1.257
H1	0.402	0.598	0.847	0.497
CE(17:2)	0.975	0.565	0.031	0.622
CE(20:5)	0.402	0.044	-0.564	1.463
CE(22:5)	0.180	0.276	-1.127	-0.820
CE(22:6)	0.376	0.962	-0.930	0.054
DG(32:1)	0.969	0.047	0.026	1.495
DG(32:2)	0.871	0.493	0.208	0.717
DG(34:1)	0.474	0.077	-0.553	1.326
DG(34:3)	0.656	0.462	0.459	0.758
DG(36:2)	0.488	0.962	-0.721	-0.054
DG(36:3)	0.488	0.596	-0.678	-0.523
DG(36:4)	0.953	0.126	0.063	-1.260
DG(41:1)	0.998	0.618	-0.002	-0.516
DG(42:2)	0.882	0.996	-0.193	0.007
DG-O(32:2)	0.860	0.239	-0.200	1.074
DG-O(34:1)	0.416	0.869	-0.845	0.180
DG-O(36:4)	0.335	0.554	-0.942	-0.557
TG(44:4)	0.610	0.818	-0.549	0.268
TG(50:1)	0.488	0.839	-0.701	0.230
TG(50:4)	0.283	0.847	-1.074	0.187

Table S4: List of metabolites that passed the quality control.

Shown are the adjusted p-values and the coefficients when running the linear model.

Index of figures and tables

Figures

Tables

Supplementary data

List of publications

[1] Fabiana Quagliarini, Ashfaq Ali Mir, **Kinga Balazs,** Michael Wierer, Kenneth Allen Dyar, Celine Jouffe,Konstantinos Makris, Johann Hawe, Matthias Heinig, Fabian Volker Filipp, Grant Daniel Barish, N. Henriette Uhlenhaut: ***Cistromic Reprogramming of the Diurnal Glucocorticoid Hormone Response by High-Fat Diet***, Molecular Cell Volume 76, Issue 4, 21 November 2019, Pages 531-545.e5, doi: 10.1016/j.molcel.2019.10.007

[2] **Kinga Balázs**[*], Barbara Höllbacher[*], Matthias Heinig and N. Henriette Uhlenhaut: ***Seq-ing answers: Current data integration approaches to uncover mechanisms of transcriptional regulation***, Comput Struct Biotechnol J. 2020; 18: 1330–1341. doi: 10.1016/j.csbj.2020.05.018, *These authors contributed equally to this work

Conference attendances and poster presentations

[1] **Kinga Balázs,** Fabiana Quagliarini, Ashfaq Ali Mir and N. Henriette Uhlenhaut: ***Integration of multi-omics data reveals diurnal control of hepatic metabolism by glucocorticoid hormones,*** *RECOMB 2020* (Padua, Italy, 23rd of June, 2020: due to the pandemic was held online)