Identification of new candidate genes associated with metabolic traits applying a multiomics approach in the obese mouse model BFMI-861

D i s s e r t a t i o n

zur Erlangung des akademischen Grades

doctor of philosophy

(Ph.D.)

eingereicht an der
Lebenswissenschaftlichen Fakultät
der Humboldt-Universität zu Berlin

von
Master of Science, Manuel Delpero

Präsidentin Prof. Dr. Julia Von Blumenthal
der Humboldt-Universität zu Berlin

Dekan Prof. Dr. Christian Ulrichs
der Lebenswissenschaftlichen Fakultät

Gutachterin/Gutachter

Prof. Dr. Timo Kautz
Prof. Dr. Gudrun Brockmann
Prof. Dr. Michael Schupp
Dr. Siham Rahmatalla
Dr. Paula Korkuc

Tag der mündlichen Prüfung: 24/02/2023

Table of contents

List of figures

List of tables

List of abbreviations

KEGG	Kyoto Encyclopedia of Genes and Genomes	LD	Linkage Disequilibrium
MQM	Multiple QTL mapping	LOD	Logarithm (Base 10) of Odds
GATK	Genome AnalysisToolkit	Q-Q plot	Quantile-Quantile Plot
BWA	Burrows-Wheeler Aligner	KASP	Kompetitive Allele Specific PCR
GEO	Gene ExpressionOmnibus	QTL	Quantitative TraitLocus
GO	Gene Ontology	RT-PCR	Real-Time PCR
GWAS	Genome-Wide AssociationStudy	RNA	Ribonucleic Acid
NCBI	National Center for Biotechnology Information	BFMI	Berlin Fat Mouse Inbred
Mb	Megabase	BMI	Body Mass Index
WGS	Whole Genome Sequencing	SNP	Single nucleotide polymorphism

Summary

Obesity increases the risk of developing common metabolic disorders such as the metabolic syndrome. This is a metabolic abnormality commonly associated with high body weight, ectopic fat storage, insulin resistance, high blood pressure, and mild chronic inflammation. Estimates of heritability for each trait of metabolic syndrome are high, with some estimates exceeding 50%. Nevertheless, identified loci from genome-wide association studies (GWAS) on features of the metabolic syndrome together explain only 1 to 7% of the variance in the human population. Further studies are needed to identify additional causative genes and to better understand their direct and interaction effects that contribute to the metabolic syndrome. These studies can be performed on experimental populations such as controlled mouse populations, allowing for easier correction for environmental factors or family structures compared to human populations.

The Berlin fat mouse population was selected for juvenile obesity. In a cross between the most obese inbred line BFMI860 and the lean control line C57BL/6NCrl, we previously identified a recessive genetic defect at a locus on chromosome 3 that accounts for 40% of the variance in body fat weight at 6 weeks of age (Neuschl et al., 2010a), (Arends, Heise, Kärst, Trost, & Brockmann, 2016). This juvenile obesity locus (*jObes1*) is fixed in all BFMI lines.

In the current studies, we aimed to identify additional genetic factors that contribute to obesity, the metabolic syndrome and fatty liver disease in the Berlin fat mouse. In the focus of the studies presented in this thesis, the BFMI861-S1 line served as a model for traits associated with obesity and the metabolic syndrome.

To investigate the cause of a metabolic disorder in BFMI861-S1, we used two different breeding partners and generated advanced intercross line (AIL) populations. In these AILs, we performed QTL mapping and then prioritization of candidate genes using multiple data sources such as whole-genome sequencing and gene expression data.

The BFMI861-S1 lineage exhibits high body weight, increased hepatic fat storage, low insulin sensitivity, and impaired glucose tolerance, while the BFMI861-S2 lineage is insulin sensitive despite obesity. These two lines are genetically very similar. Therefore, the remaining genetic diversity must be responsible for the phenotypic differences. To identify genetic loci responsible for the observed obesity and impaired glucose homeostasis in BFMI861-S1 mice independent of the *jObes1* locus, an advanced intercross line was generated, resulting from an initial cross between BFMI861-S1 and BFMI861 -S2.

The second AIL was generated from BFMI861-S1 and B6N (AIL (BFMI861-S1xB6N)), where B6N is a lean reference mouse line to BFMI861-S1.

Overlapping QTL for gonadal fat weight and blood glucose concentration on chromosome 3 and for gonadal fat weight, liver weight and blood glucose concentration on chromosome 17 were identified in the AIL (BFMI861-S1xBFMI861-S2). An additional QTL for gonadal fat weight is located on chromosome 15. In addition, two QTL with the time course of body weight development from weeks 9 to 25 were found on chromosomes 15 and 16.

In the second AIL (BFMI861-S1xB6N), three QTLs associated with liver weight, body weight and subcutaneous adipose tissue weight were identified. A highly significant QTL on chromosome 1 showed an association with liver weight. A QTL for body weight at 20 weeks of age overlapping with a QTL for subcutaneous fat weight was found at chromosome 3. In a multiple QTL mapping approach, another QTL influencing body weight at 16 weeks of age at chromosome 6 was identified. Interestingly, the top candidate genes for the identified QTL were previously associated with metabolic traits.

By combining QTL mapping with a detailed prioritization approach, we were able to identify new candidate genes in both cross populations and strengthen already known candidates associated with metabolic syndrome traits in the BFMI861-S1 line.

Zusammenfassung

Adipositas erhöht das Risiko, weit verbreitete Stoffwechselstörungen wie das metabolische Syndrom zu entwickeln. Hierbei handelt es sich um eine Stoffwechselanomalien, die meist mit hohem Körpergewicht, ektoper Fettspeicherung, Insulinresistenz, Bluthochdruck und leichten chronischen Entzündungen verbunden ist. Heritabilitätsschätzungen für jedes Merkmal des metabolischen Syndroms sind hoch, wobei einige Schätzungen 50 % überschreiten. Dennoch erklären identifizierte Loci aus genomweiten Assoziationsstudien (GWAS) zu Merkmalen des Metabolischen Syndroms zusammen nur 1 bis 7 % der Varianz in der menschlichen Population. Weitere Studien sind erforderlich, um zusätzliche ursächliche Gene zu identifizieren und um ihre direkten und Wechselwirkungseffekte, die zum metabolischen Syndrom beitragen, besser zu verstehen. Diese Studien können an experimentellen Populationen wie kontrollierten Mauspopulationen durchgeführt werden, was im Vergleich zu menschlichen Populationen eine leichtere Korrektur für Umweltfaktoren oder Familienstrukturen ermöglicht.

Die Population der Berliner Fettmaus wurde auf juvenile Adipositas selektiert. In einer Kreuzung zwischen der adipösesten Inzuchtlinie BFMI860 und der schlanken Kontrolllinie C57BL/6NCrl haben wir zuvor einen rezessiven genetischen Defekt an einem Ort auf Chromosom (Chr) 3 identifiziert, der 40 % der Varianz im Körperfettgewicht im Alter von 6 Wochen ausmacht (Neuschl et al., 2010a) (Arends et al., 2016). Dieser juvenile Adipositas-Locus (*jObes1*) ist in allen BFMI-Unterlinien fixiert.

In den aktuellen Studien zielten wir darauf ab, zusätzliche genetische Faktoren zu identifizieren, die zur Fettleibigkeit, dem metabolischen Syndrom und der Fettlebererkrankung bei der Berliner Fettmaus beitragen. Im Focus der in der vorliegenden Arbeit präsentierten Studien diente die BFMI861-S1-Linie als Modell für Merkmale, die mit Fettleibigkeit assoziiert sind.

Um die Ursache einer Stoffwechselstörung in BFMI861-S1 zu untersuchen, haben wir zwei verschiedene Zuchtpartner verwendet und vorangeschrittene Kreuzungspopulationen (advanced intercross line, AIL) generiert. In diesen AILs haben wir eine QTL-Kartierung und

anschließend eine Priorisierung von Kandidatengenen unter Verwendung mehrerer Datenquellen wie Whole-Genome Sequenz- und Genexpressionsdaten durchgeführt.

Die BFMI861-S1-Linie zeigt ein hohes Körpergewicht, erhöhte hepatische Fettspeicherung, geringe Insulinsensitivität und eine beeinträchtigte Glukosetoleranz, während die Linie BFMI861-S2 trotz Fettleibigkeit insulinsensitiv ist. Diese beiden Linien sind genetisch sehr ähnlich. Deshalb muss die verbleibende genetische Vielfalt für die phänotypischen Unterschiede verantwortlich sein. Um genetische Loci zu identifizieren, die für die beobachtete Fettleibigkeit und die beeinträchtigte Glukosehomöostase in BFMI861-S1-Mäusen unabhängig vom *jObes1* Locus verantwortlich sind, wurde eine fortgeschrittene Intercross-Linie (AIL) generiert, die aus einer anfänglichen Kreuzung zwischen BFMI861-S1 und BFMI861-S2 stammt.

Die zweite AIL wurde aus BFMI861-S1 und B6N (AIL (BFMI861-S1xB6N)) generiert, wobei B6N eine schlanke Referenzmauslinie zu BFMI861-S1 ist.

In der AIL (BFMI861-S1xBFMI861-S2) wurden überlappende QTL für das gonadale Fettgewicht und die Blutglukosekonzentration auf Chromosom 3 und für das gonadale Fettgewicht, Lebergewicht und Blutglukosekonzentration auf Chromosom 17 identifiziert. Ein zusätzlicher QTL für gonadales Fettgewicht befindet sich auf Chromosom 15. Darüber hinaus wurden zwei QTL mit dem zeitlichen Verlauf der Körpergewichtsentwicklung von Woche 9 bis 25 auf den Chromosomen 15 und 16 gefunden.

In der zweiten AIL (BFMI861-S1xB6N) wurden drei QTL identifiziert, die mit dem Lebergewicht, dem Körpergewicht und dem Gewicht des subkutanen Fettgewebes assoziiert sind. Ein hochsignifikanter QTL auf Chromosomen 1 zeigte eine Assoziation mit dem Lebergewicht. Ein QTL für das Körpergewicht im Alter von 20 Wochen, der mit einem QTL für das Gewicht des subkutanen Fettgewebes überlappte, wurde auf Chromosomen 3 gefunden. In einem multiplen QTL-Mapping-Ansatz wurde ein weiterer QTL mit Einfluss auf das Körpergewicht im Alter von 16 Wochen auf Chromosomen 6 identifiziert. Interessanterweise wurden alle Top-Kandidatengene für die identifizierten QTL zuvor mit metabolischen Merkmalen.

Durch die Verknüpfung des QTL-Mappings mit einem detaillierten Priorisierungsansatz konnten wir in beiden Kreuzungspopulationen neue Kandidatengene identifizieren und bereits bekannte Kandidaten stärken, die mit Merkmalen des metabolischen Syndroms in der BFMI861-S1-Linie assoziiert sind.

Chapter 1: Introduction

1.1 Obesity and metabolic syndrome

It is widely known that obesity is a complex multifactorial condition observed in all ages and sexes (Engin, 2017) where genetics is an important component. Overweight and obesity are proposed to reach levels of 89% and 85% in males and females, respectively by 2030 (Keaver et al., 2020). This will result in an increase in the obesity-related prevalence of coronary heart disease (CHD) by 97%, cancers by 61% and type 2 diabetes by 21%. In addition obesity is associated with decreased life expectancy by 3.3-18.7 years affecting the way we age (Keaver et al., 2020).

It is obvious that obesity increases the risk of developing different metabolic disorders. Around 70 % of obesity patients show the metabolically unhealthy obese phenotype where high bodyweight is related to metabolic abnormalities such as low insulin sensitivity and dyslipidemia (Blüher, 2019). Only 30% of obese patients are metabolically healthy with normal insulin sensitivity and low visceral fat content (Blüher, 2019).

Obesity and its related pathologies such as insulin resistance, type 2 diabetes, and fatty liver are associated with an imbalance of energy (Blüher, 2019). A sedentary lifestyle as well as the consumption of easily available energy-dense food contribute to this imbalance (Hebebrand et al., 2017). In addition to eating behavior, sedentary lifestyle, cultural influence and other environmental factors, the genetic constitution sets the stage for the phenotypic characteristics of obesity.

In the last two decades, genome-wide scans identified more than one hundred loci associated with obesity (Goodarzi, 2018). Nevertheless, identified loci and underlying genes explain only 3 % of the estimated heredity of obesity related phenotypes (Goodarzi, 2018). This situation suggests that additional genetic loci predisposing weight gain (including low/rare frequency alleles) remain to be discovered. Or other genetic variants such as copy number variations (CNVs) and structural variants (SVs).

1.2 Genetic association studies

The goal of genome-wide association studies (GWAS) is to identify genotype to phenotype associations through testing differences in the means between genotype classes of genetic variants such as single nucleotide polymorphisms (SNPs) between individuals that differ in phenotypes (Uffelmann et al., 2021).

In the last 15 years, many loci have been associated with diseases and traits (Uffelmann et al., 2021). More than 5,700 GWAS have now been conducted for more than 3,300 traits and a push for more statistical power has thrust GWAS sample sizes well beyond a million participants, yielding numerous associated and replicable variants for many heritable traits (Watanabe et al., 2019). Therefore, many genetic associations for different phenotypes are currently known. However, we are facing the next big challenge in the contest of genetic discoveries which is the interpretation of these associations in a biological and genomic context. Previous GWAS have shown that most traits are influenced by thousands of causal variants that individually confer very little risk. These variants are often associated with many other traits making direct biological, causal inferences complicated (Holland et al., 2020). In addition, in human populations genetic associations may differ across ancestries, complicating direct comparisons between groups of individuals drawing unambiguous conclusions about the biological meaning of GWAS results and limiting their utility to produce mechanistic insights or to serve as starting points for drug development (Uffelmann et al., 2021) (Figure 1.1). However, these several challenges in GWAS can be resolved by using controlled populations such as mouse inbred lines (Hao Li & Auwerx, 2020).

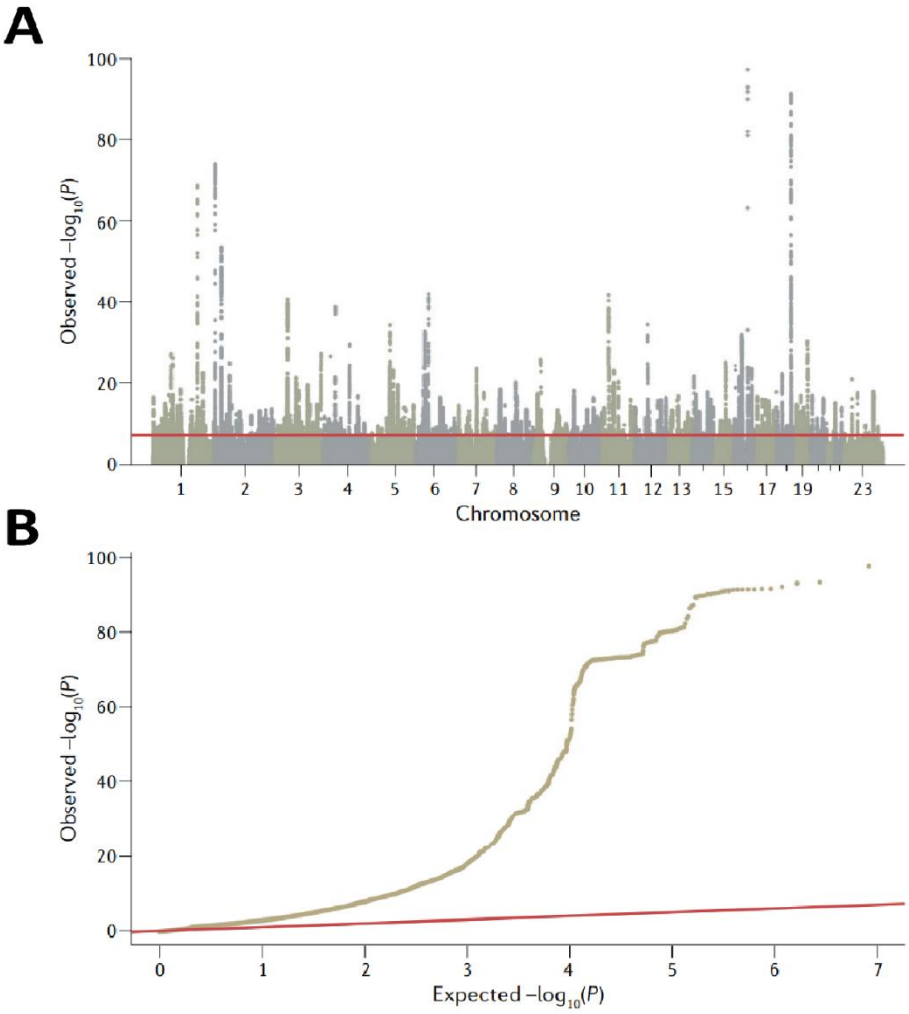


Figure 1.1. Manhattan plot and Q-Q plot that represent the results of a human GWAS for BMI.

(A) Manhattan plot of a human GWAS for BMI. **(B)** Q-Q plot showing an early separation of the observed from the expected, suggesting population stratification and limiting the utility to produce meaningful insights (Uffelmann et al., 2021).

1.3 Mouse populations for genetic studies

Mouse models are essential for genetic studies because they allow high control of environmental factors during experiments not possible in human studies (Darvasi & Soller, 1995). In detail, compared to human studies, mouse inbred lines allow rigorously control of husbandry and environmental conditions, including diet, activity, stress and the microbiome. In such conditions, relations between genotype and phenotype can be tested without the confounding effects of heterogeneous genetic backgrounds (Nadeau & Auwerx, 2019). In parallel, segregating crosses and outbred populations can be used to test effects of genetic heterogeneity. The mouse genome has also been studied in detail more than any other animal model, and with the high amount of freely available data sources from many different mouse strains genetic studies in mouse populations can be performed at accessible costs (Gurumurthy & Lloyd, 2019).

Although mouse models do not perfectly approximate human biology, when used properly, they can provide precise insights into the human condition and enable discovery of fundamental biological principles, prioritize human studies and test hypotheses (Yalcin, Flint, & Mott, 2005). For example, a variant of the *Ob* gene responsible for extreme obesity in the ob/ob mouse was discovered and shown to encode for Leptin, which is a peptide secreted by the adipocytes (Zhang et al., 1994). These findings led to the clarification that hunger and satiety are controlled by leptin and the Leptin receptor in the hypothalamus and that variants in human orthologs of these genes produced similar phenotypes (Joost & Schürmann, 2014). Animal models and in particular mouse models are therefore essential to advance in genetic discoveries and to reveal the genetic contributions of complex diseases.

1.4 The Berlin Fat Mouse

Mouse populations are essential powerful tools to perform genetic studies because as previously explained in mouse genetic studies environmental conditions can be perfectly controlled. The Berlin Fat Mouse Inbred (BFMI) line is a model organism for researching polygenic obesity (Arends et al., 2016), and is particularly used to understand the genetics behind complex traits such as obesity and associated metabolic traits. The origin of the different BFMI lines starts around 60 years ago when mice were bought in different pet shops in Berlin. After initial crossing experiments, the offspring generations were first selected for low protein content, followed by high body mass and high fat content and finally selected for high fatness and then inbred for 58 generations, which resulted in the generation of six BFMI strains (Wagener et al., 2006). All six of these strains are obese with high body mass, high total fat and reduced lean percentage (Heise et al., 2016). Using a QTL mapping approach in an AIL population between the most obese BFMI line (BFMI860-12) and B6N has discovered that 40 % of the obesity of these strains can be explained by the *jObes1* locus on chromosome 3 which is present in all BFMI strains and predicts the *Bbs7* gene as the most likely factor of variance in body weight (Arends et al., 2016). The cause of the remaining 60% of body weight variance and for the reduced insulin sensitivity has not been discovered yet.

The BFMI lines BFMI861-S1 and -S2 are two sublines that were generated during the process of inbreeding. They are 96.4% genetically identical and both carry the *jObes1* obesity locus. However, despite the high genetic similarity they differ in several metabolic traits such liver triglycerides, blood glucose, and plasma insulin (Figure 1.2). In addition, the BFMI861-S1 shows insulin resistance whereas the S2 line retains its insulin sensitivity (Heise et al., 2016). The BFMI861-S1 and BFMI861-S2 lines are therefore suitable mouse lines to perform genetic studies and to reveal genetics variants responsible for metabolic syndrome related traits.

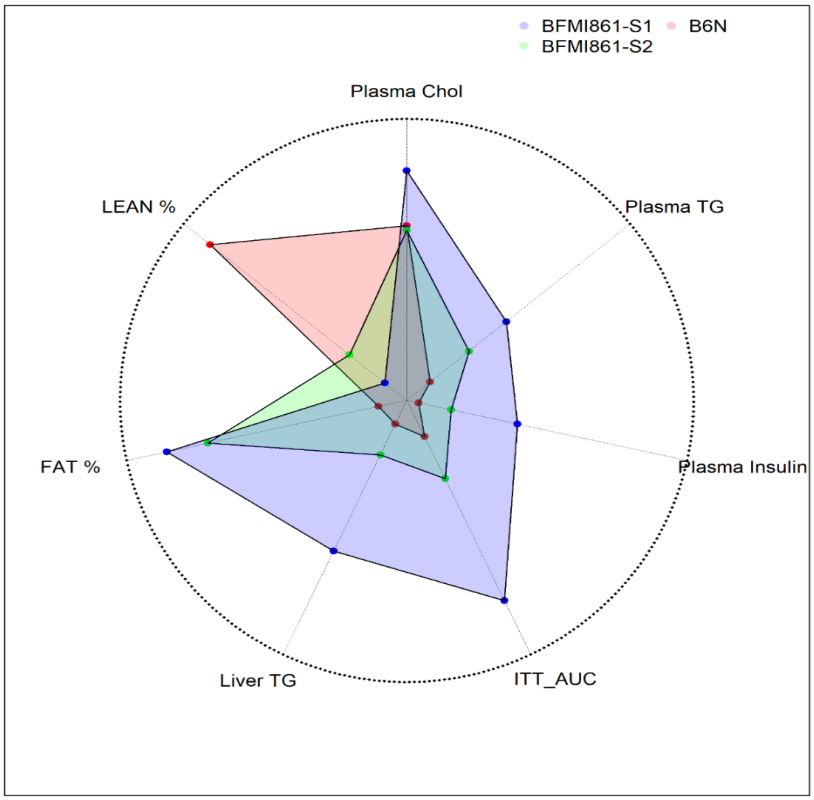


Figure 1.2. Differences in metabolic traits between the two BFMI lines BFMI861-S1 (blue), BFMI861-S2 (green), and the reference B6N (red). FAT %, fat mass percentage; LEAN %, lean mass percentage; TG, triglycerides; ITT_AUC, area under the curve for insulin tolerance test.

1.5 Mouse recombinant inbred lines and advanced intercross lines

In the last decades much effort has been spent to identify molecular variants responsible for complex traits such as obesity and the metabolic syndrome. Despite the effort the association between complex traits and genetic variants is still poorly resolved. The final goal of quantitative genetics is to map the relationship between genotypes and phenotypes to genomic regions and quantitative trait loci (QTL). However, the associated genomic regions are usually large and therefore the dissection of QTL into quantitative trait nucleotides is still a major challenge (Delpero et al., 2021).

Mouse populations are optimal tools for genetic studies, and different mouse populations

have been designed to improve the resolution of a QTL. For example, recombinant inbred lines (RILs) are produced by a series of continued brother-sister mating (Rockman & Kruglyak, 2008). Additional powerful population used for genetic studies have been developed such as advanced intercross lines (AILs). AILs are experimental populations that can provide more accurate estimates of QTL map location than conventional mapping populations such as RILs (Darvasi & Soller, 1995). This population is produced by randomly and sequentially intercrossing a population initially originated from a cross between two inbred lines or some variant. Random mating provides increasing probability of recombination between any two loci (Darvasi & Soller, 1995). Considering a set of RILs we can reach many of the map-expanding properties of an AIL. For reasonable statistical power in QTL mapping, however, a set consisting of large number of RILs is required. Consequently, a single AIL is efficient for fine mapping as a very large number of RILs resulting in a less expensive experimental approach.

AILs are particularly applicable to species with a short generation cycle that can easily be reproduced. Inbred lines of mice are the outstanding example for mammals, because these can readily be crossed and subsequently intercrossed without great effort. In addition, the ability of AILs to increase the number of recombinational events between any two loci, leads to a decrease in the confidence interval of a QTL and, therefore, serves as a general fine-mapping resource for the species in question.

1.6 Multiomics approaches for prioritization of candidate genes in a QTL region

To perform genetic association studies that lead to high mapping resolution, first we need to choose the right population, and second, accurately detect variants that are in high linkage disequilibrium (LD). Over 95% of the variants in high LD ($R^2 > 0.8$) are located outside of genes in the non-coding DNA and can be located several Kbs or even Mbs apart depending on the resolution of the used population (Broekema, Bakker, & Jonkers, 2020). Any of the significantly associated variants in a QTL could be the actual causal variant as well as any of the genes located in the confidence interval. To overcome these challenges in gene and nucleotides prioritization in a newly identified QTL region, computational methods that integrates multiple data sources (sequencing data, gene expression data, proteomics and literature information) help to narrow down the list of candidate genes and variants allowing the identification of the right candidate.

Computational gene prioritization strategies can be classified into four different types: text mining methods, network-based methods, machine learning methods, and hybrid strategies (Raj & Sreeja, 2018).

Text mining method applies published scientific literature to identify associations between genes and complex disorders. Network-based methods represent biological data as network and apply graph mining techniques to rank genes. The study and implementation of new algorithms are explored using machine learning approaches and therefore through the creation of new decision trees for achieving ranking and gene prioritization (Raj & Sreeja, 2018). A combination of any of these methods has been grouped as hybrid methods. Even though text mining, network based and machine learning techniques are widely used for computational gene prioritization, recent studies show a significant boost in the performance by integrating multiple approaches.

1.7 Aims of the study

This thesis has three goals: 1) Discover new QTL associated with metabolic traits and the metabolic syndrome in the Berlin Fat Mouse line BFMI861-S1, 2) Prioritize and detect candidate genes in the newly identified QTL, 3) Identify the causal tissue responsible for the metabolic syndrome in the BFMI861-S1 line. To reach these goals, two different AILs were generated with always BFMI861-S1 as one parental strain and BFMI861-S2 or C57BL/6N (B6N) as breeding partner, respectively.

To detect QTL for metabolic traits that act independently of the effect of the *jObes1* locus, an AIL from the initial cross between the lines BFMI861-S1 and BFMI861-S2 was generated. Both lines, BFMI861-S1 and BFMI861-S2, share the *jObes1* locus on Chr 3 which is responsible for the high body weight of all BFMI lines. Despite being genetically closely related these two lines differ in metabolic traits such as body weight, and insulin sensitivity. With this AIL we performed QTL mapping using traits collected at the end of experiment in week 25 and time series data on body weight.

To detect additional QTL for obesity and fatty liver disease in the BFMI861-S1 mouse model, we generated an AIL between BFMI861-S1 and C57BL/6NCrl (B6N).

To accurately identify the most likely candidate genes and variants in each QTL region that are responsible for the associated phenotypes a multiomics prioritization approach that includes information from different sources such as next generation sequencing (NGS), gene expression, and literature data was designed. Gene expression data collected in multiple tissues such as gonadal adipose tissue, liver, pancreatic islets, and skeletal muscle were also used to detect the driver tissue for the metabolic syndrome in the BFMI861-S1 line.

The combination of QTL mapping together with a specific prioritization method allows to identify novel candidate genes and provides additional evidence for the action of already known genes involved in common human metabolic diseases. Thereby, this study contributes to the increasing molecular understanding of obesity and related disorders and helps to set the stage for improved prevention and therapy.

Chapter 2: Identification of four novel QTL linked to the metabolic syndrome in the BFMI861 lines

Manuel Delpero[1], Danny Arends[1], Maximilian Sprechert[1], Florian Krause[1], Oliver Kluth[2,3], Annette Schürmann[2,3,4], Gudrun A. Brockmann[1], Deike Hesse[1]

Affiliation:
[1] Albrecht Daniel Thaer-Institut für Agrar- und Gartenbauwissenschaften, Humboldt-Universität zu Berlin, Berlin, Germany.
[2] Department für Experimentelle Diabetologie, Deutsches Institut für Ernährungsforschung Potsdam-Rehbrücke (DIfE), Nuthetal, Germany.
[3] German Center for Diabetes Research (DZD), München-Neuherberg, Germany
[4] University of Potsdam, Institute of Nutritional Science, Germany

The content of this chapter was published in: Delpero, M., Arends, D., Sprechert, M., Krause, F., Kluth, O., Schürmann, A., Brockmann G. A., Hesse, D. (2021). Identification of four novel QTL linked to the metabolic syndrome in the Berlin Fat Mouse. *International Journal of Obesity*.
DOI: https://doi.org/10.1038/s41366-021-00991-3

Background: The Berlin Fat Mouse Inbred line (BFMI) is a model for obesity and the metabolic syndrome. This study aimed to identify genetic variants associated with the impaired glucose metabolism using the obese lines BFMI861-S1 and BFMI861-S2, which are genetically closely related, but differ in several traits. BFMI861-S1 is insulin resistant and stores ectopic fat in the liver, whereas BFMI861-S2 is insulin sensitive.

Methods: In generation 10, 397 males of an advanced intercross line (AIL) BFMI861-S1 x BFMI861-S2 were challenged with a high-fat, high-carbohydrate diet and phenotyped over 25 weeks. QTL-analysis was performed after selective genotyping of 200 mice using the GigaMUGA Genotyping Array. Additional 197 males were genotyped for 7 top SNPs in QTL regions. For the prioritization of positional candidate genes whole genome sequencing and gene expression data of the parental lines were used.

Results: Overlapping QTL for gonadal adipose tissue weight and blood glucose concentration were detected on chromosome (Chr) 3 (95.8-100.1 Mb), and for gonadal adipose tissue weight, liver weight, and blood glucose concentration on Chr 17 (9.5-26.1 Mb). Causal modelling suggested for Chr 3-QTL direct effects on adipose tissue weight, but indirect effects on blood glucose concentration. Direct effects on adipose tissue weight, liver weight and blood glucose concentration were suggested for Chr 17-QTL. Prioritized positional candidate genes for the identified QTL were *Notch2* and *Fmo5* (Chr 3) and *Plg* and *Acat2* (Chr 17). Two additional QTL were detected for gonadal adipose tissue weight on Chr 15 (67.9–74.6) and for body weight on Chr 16 (3.9-21.4 Mb).

Conclusions: QTL mapping together with a detailed prioritization approach allowed us to identify candidate genes associated with traits of the metabolic syndrome. In addition, we provided evidence for direct and indirect genetic effects on blood glucose concentration in the insulin resistant mouse line BFMI861-S1.

2.1 Introduction

The metabolic syndrome is defined as a metabolic abnormality that leads to high body weight, ectopic fat storage, insulin resistance, high blood pressure, and chronic low-grade inflammation (Aguilar-Salinas & Viveros-Ruiz, 2019). Heritability estimates for each trait of the metabolic syndrome are high with some estimates exceeding 50 % (Goodarzi, 2018). Nevertheless, genome wide association studies (GWAS) on body mass index (BMI) and other traits of the metabolic syndrome identified loci, that combined, account for only 1 to 7 % of the variance in the examined population (Goodarzi, 2018). Therefore, studies on different populations are needed to identify additional causal genes to better understand their direct and interaction effects contributing to the metabolic syndrome.

The goal of the current study was to identify genetic factors contributing to obesity and glucose homeostasis in the Berlin Fat Mouse. Originally, the Berlin Fat Mouse population was selected for juvenile obesity. After 58 generations of selection, different Berlin Fat Mouse Inbred (BFMI) lines were generated through repeated brother-sister mating (Wagener et al., 2006). In a cross between the most obese inbred line BFMI860 and the lean control line C57BL/6NCrl, we have previously identified a recessive genetic defect at a locus on chromosome (Chr) 3 accounting for 40 % of the variance in adipose tissue weight at 6 weeks (Neuschl et al., 2010a) (Arends et al., 2016). This juvenile obesity locus (*jObes1*) is fixed in all BFMI sublines.

In the current study, we used the inbred lines BFMI861-S1 (S1) and BFMI861-S2 (S2). S1 and S2 are sublines created from the BFMI860, as such the BFMI860 is the predecessor of the S1 and S2. The S1 and S2 lines were conspicuously different with respect to metabolic traits (Heise et al., 2016). In particular, the S1 line showed high body weight, hepatic fat storage, low insulin sensitivity, and impaired glucose tolerance. In contrast, S2 is insulin sensitive despite being obese (Heise et al., 2016). This observation was particularly interesting, since these two lines were derived from one parental line that was divided into two sub-lines only after four generations of inbreeding. Therefore, these two lines are genetically highly similar, and the remaining genetic diversity is responsible for phenotypic differences. To

identify genetic loci accounting for the observed obesity, and glucose homeostasis in S1 mice, we performed a quantitative trait locus (QTL) mapping study in an advanced intercross line (AIL) which was generated from an initial cross between the BFMI861 lines S1 and S2. In this study, all AIL mice were challenged with a high-fat, high-carbohydrate diet.

2.2 Material and methods

Mouse population

We used male mice of the parental mouse lines BFMI861-S1 (S1) and BFMI861-S2 (S2) and generation 10 of an AIL population. The AIL population was generated from an initial cross between a BFMI861-S1 (S1) male and a BFMI861-S2 (S2) female followed by repeated random mating in every generation. For randomization of mating pairs, the program RandoMate (Schmitt, Bortfeldt, Neuschl, & Brockmann, 2009) was used. The BFMI861 lines S1 and S2 were generated as described in Heise et al., 2016 (Heise et al., 2016).

Animal husbandry

All experimental treatments of mice were approved by the German Animal Welfare Authorities (approval no. G0235/17). Mice were kept under conventional conditions with a 12:12 h light–dark cycle (lights on at 0600 hours) and at a temperature of 22 ± 2 °C. Mice had *ad libitum* access to food and water.

Experiment and phenotyping

Data from parental strains S1 and S2 were collected at 20 weeks on a standard diet containing 16.7 MJ/kg of metabolizable energy, 11 % from fat, 26 % from protein and 53 % from carbohydrates (V1534-000, ssniff EF R/M; Ssniff Spezialdiäten GmbH, Soest, Germany) and

blood glucose was measured at 25 weeks after five weeks exposure to a high-fat, high-carbohydrate diet containing 21.9 MJ/kg of metabolizable energy, 28% from fat, 20% from protein and 40% from carbohydrates (Kluth et al., 2015).

To emphasize the difference in glucose homeostasis, all AIL animals were challenged with a dietary regime that provides a gluco-lipotoxic environment for the ß-cells and thereby provokes differences in ß-cell resilience (Kluth et al., 2011). This dietary regime challenge was undertaken to provoke differences in the phenotypes studied. Until the age of 20 weeks, AIL mice were fed the rodent standard diet. In weeks 21 and 22, mice were fed a high-fat, low-carbohydrate diet, containing 16.9 MJ/kg of metabolizable energy, 34 % from fat, 19 % from protein and 47 % from carbohydrates (C1057; Altromin Spezialfutter GmbH & Co. KG, Lage, Germany) to increase obesity but to protect ß-cells. Afterwards, animals were fed for 3 weeks a high-fat, high-carbohydrate diet containing 21.9 MJ/kg of metabolizable energy, 28 % from fat, 20 % from protein and 40 % from carbohydrates (Oliver Kluth et al., 2015) to challenge ß-cells with carbohydrates and thereby increase differences in glucose metabolism.

AIL mice were phenotyped between the age of 3 (after weaning) and 25 weeks. Body mass was recorded weekly. To investigate the glucose metabolism, an oral glucose tolerance test (oGTT) was performed in week 18 and an intraperitoneal insulin tolerance test (ITT) in week 20 as described before (Heise et al., 2016). The area under the curve (AUC) for blood glucose concentration of oGTT and ITT was calculated. At 25 weeks, final blood glucose concentration was recorded after fasting for two hours. Afterwards, mice were anesthetized with isoflurane and sacrificed (Hesse et al., 2018). Gonadal adipose tissue (GonAT), subcutaneous adipose tissue, liver, and skeletal muscle (quadriceps) were dissected and weighed. Tissues were collected in liquid nitrogen and stored at -80°C. Protein content and triglycerides of homogenized liver samples were determined as described in Hesse et al., 2014 (Hesse et al., 2014).

Outliers, defined as individuals which have a measurement that deviates from the population mean by more than four standard deviations (SD), were removed from the data. Pearson's

correlation coefficients were calculated between normal distributed traits. For non-normal distributed traits, Spearman's correlation coefficients were calculated.

Genotyping

Out of the 397 males that were phenotyped, selective genotyping was performed; 200 mice representing the two tails of the phenotypic distributions of gonadal adipose tissue weight and liver weight were selected for genotyping with the Giga Mouse Universal Genotyping Array (GigaMUGA; Illumina, San Diego, CA, USA) (Morgan et al., 2016). Genotyping was done at Neogen GeneSeek (Lincoln, NE, USA). Due to high genetic similarity of the parental lines S1 and S2 of the AIL population, only 5,215 out of 143,259 SNPs on the array were informative and passed the quality control (Figure 2.1).

Remaining 197 males of the AIL population were genotyped for 7 top markers. For these markers, KASP genotyping assays were developed as described previously (Kreuzer, Reissmann, & Brockmann, 2013). The additional animals were genotyped to counteract any bias in the estimates of allele effect sizes introduced by selective genotyping.

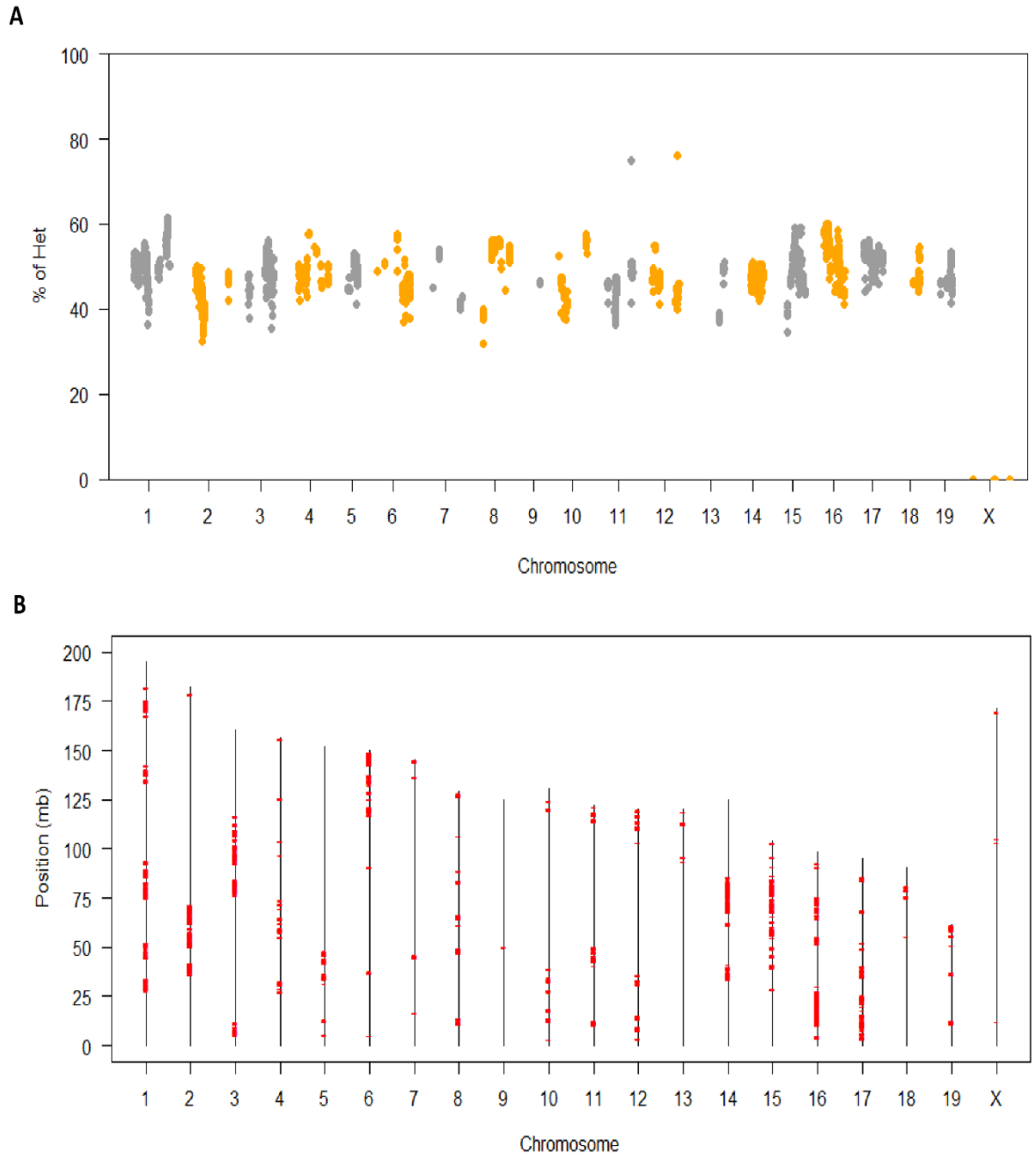


Figure 2.1. (A) Percentage (%) of heterozygosity for each informative SNP (5,215) in the AIL (BFMI861-S1xBFMI861-S2) at generation 10 for 397 males. (B) Distribution of informative SNPs across the genome in the AIL (BFMI861-S1xBFMI861-S2).

QTL mapping

QTL mapping was performed in two steps: First, a QTL scan was performed using the 200 males that were genotyped with the GigaMUGA Array. Afterwards, a final QTL scan was performed including all animals (genotyped by GigaMUGA and KASP).

Covariates (subfamily and litter size) were investigated for a significant influence on each phenotype. Covariate analysis showed that litter size significantly influenced liver weight ($p < 0.02$), as such, litter size was added as a covariate to the model when QTL mapping liver weight. No other significant covariates were found. Using pedigree information of the AIL population, we tested the sub-family effect on the phenotype, but no significant influence was found (code available upon request).

To confirm that QTL mapping models are valid, residuals of the models were tested for normality using a Shapiro-Wilk test. If residuals were found to not be normally distributed, a non-parametric Kruskal–Wallis one-way analysis of variance was performed to validate the top marker.

The number of independent statistical tests was estimated by simpleM (Gao, 2011) which determined the number of independent tests to be 849 (window size = 820, mEff = 849). Afterwards Bonferroni correction for multiple testing correction (Haynes, 2013) was performed using the number of independent SNPs as determined by simpleM. P-values were converted to LOD scores, using LOD = -log10(p-value). LOD scores above 4.9 and 4.2 were deemed to be genome-wide highly significant and significant, respectively. The 95% confidence interval of a QTL was determined by a 1.5 LOD drop from the top SNP position (Dupuis & Siegmund, 1999). Start and end positions were defined as the first SNP upstream or downstream of the 1.5 LOD-drop confidence interval.

Causal modeling

In case of an overlapping QTL between multiple traits, we applied pairwise causal modeling as previously described (Y. Li, Tesson, Churchill, & Jansen, 2010) (Brockmann, Tsaih, Neuschl, Churchill, & Li, 2009). In short when a common QTL is found for two (or more traits), we model the effect of the QTL on these traits in a pairwise manner. Causal modeling was performed by comparing the independent model (QTL directly affects both T1 and T2) with the causal/reactive model (QTL directly affects T1 which in turn affects T2). Of course, it can happen that none of the models fit the data satisfactory, we then assume causality is undetermined.

Direct QTL effects are defined as caused by a QTL which directly affects the variability of both traits (independent model fits best, QTL directly affect T1 and T2). Indirect effects were defined as effects on a trait (T1) caused by a QTL through another trait (T2) (causal model fits best). In this case the QTL is defined as having a direct effect on T2, and an indirect effect on T1. Causal modeling to determine direct and indirect effects of QTL on traits was performed for GonAT weight, liver weight, and blood glucose concentration on Gatlgq and for GonAT weight and blood glucose concentration on Gatq1.

Whole-genome sequencing

The two parental lines of the AIL (S1 and S2) were paired-end sequenced using the “Illumina HiSeq” (Illumina) platform. Obtained DNA reads were trimmed and aligned to the mouse genome (MM10, GRCm38.p3), sequence variants were called using BCFtools and annotated using the Ensembl Variant Effect Predictor (VEP) (H. Li et al., 2009) (McLaren et al., 2016). VEP provided information on the position of SNPs within known motifs such as promoters, regulatory sites and protein domains. DNA sequencing data were deposited at the NCBI Sequence Read Archive (SRA) under BioProject ID: PRJNA717237.

Gene expression analysis

RNA was isolated from gonadal adipose tissue (S1: n=7, S2: n=8), liver (S1: n=7, S2: n=8) and skeletal muscle (S1: n=7, S2: n=8) of males of the parental lines S1 and S2 at 10 weeks. Pancreatic islets (S1: n=6, S2: n=6) were isolated as described in Gotoh et al., 1985 (Gotoh et al., 1985) and RNA was extracted as described (Hesse et al., 2018). Gene expression was measured with the Clariom S assay for mouse (Thermo Fisher Scientific) using service (ATLAS Biolabs, Berlin, Germany). The intensity values of the arrays were transformed to the logarithm of base 2 and quantile normalized for each tissue separately. To test for expression differences between S1 and S2 mice, t-tests were performed for each probe on the array in each tissue. Benjamini-Hochberg correction was applied for multiple testing. R was used for statistical analysis and graphical presentation (Team, 2018). Quantitative real time PCR was performed as described in Heise et al., 2016 (Heise et al., 2016). Relative transcript amounts were calculated using the relative quantification method (ddCT-method) (Livak et al., 2001).

Candidate gene prioritization

Genomic DNA sequences of all protein coding positional candidate genes were downloaded using bioMART (Durinck, Spellman, Birney, & Huber, 2009). To include regulatory regions such as promoters, we considered additional 1,000 base pairs from the start and end position of each gene. Monomorphic genes without sequence variants between S1 and S2 were removed from the list of positional candidate genes. All other genes were scored for potential functional effects of sequence variants, gene expression differences between S1 and S2 in gonadal adipose tissue and liver, and their contribution to KEGG pathways (Kanehisa, 2000). Coding sequence variants leading to stop gain/stop loss codons and missense mutations located in functional protein domains were awarded 3 points to the gene score. A missense variant with either a deleterious or a tolerated SIFT (Sorting Intolerant From Tolerant) value obtained 3 or 1 point, respectively. Non-coding variants were scored based on their location in potential functional sites. If a non-coding

variant was located in the promoter or in a splice site, 3 points were awarded; if located in untranslated regions (UTRs), enhancers, or CTCF binding sites (involved in 3D structure of chromatin) 1 point was awarded. Genes differentially expressed in at least one tissue were awarded 2 points. Genes annotated in relevant KEGG metabolic pathways were awarded 1 point. Genes in KEGG pathways were downloaded using the R package "StarBioTrek" (Cava, Bertoli, & Castiglioni, 2015). To find further evidence for potential causality, highest scored candidate genes were screened for metabolic processes or diseases using Gene Ontology (GO), public literature, and databases such as Mouse Genome Informatics and the International Mouse Phenotyping Consortium.

2.3 Results

Response of parental lines S1 and S2, and AIL males to high-fat, high carbohydrate diet

According to SNP chip data (GigaMuga), S1 and S2 animals are 96.4% genetically identical (Figure 2.1b). However, with a standard diet S1 males had significantly higher body weight ($p < 0.001$, $n = 10$) and higher liver weight ($p < 0.001$, $n = 10$) compared to S2 males at 20 weeks of age (Figure 2.2). To challenge the glucose homeostasis, we fed 20 weeks-old S1 and S2 mice a high-fat, high-carbohydrate diet for five weeks and observed extreme high blood glucose concentration in S1 males (369 +/- 54 mg/dl) compared to S2 males (178 +/- 31 md/dl) (Figure 2.2). To elucidate the genetic impact on the response to this challenge the AIL population was exposed to a gluco-lipotoxic environment provoking differences in ß-cell resilience (O Kluth et al., 2011). Therefore, the diet was switched at 20 weeks from a standard diet to a lipotoxic high-fat, low-carbohydrate diet (two weeks) to increase obesity, followed by a gluco-lipotoxic high-fat, high-carbohydrate diet for additional three weeks to challenge ß-cells. At 25 weeks AIL mice showed an average blood glucose concentration of 210 +/- 79 mg/dl. In addition, gonadal adipose tissue weight was 1.72 +/- 0.69 g and liver weight was 3.07 +/- 0.65 g on average. Liver triglycerides/protein content was 124 +/- 64

ug/ug, and body weight was 47.17 +/- 4.00 g on average (Figure 2.2).

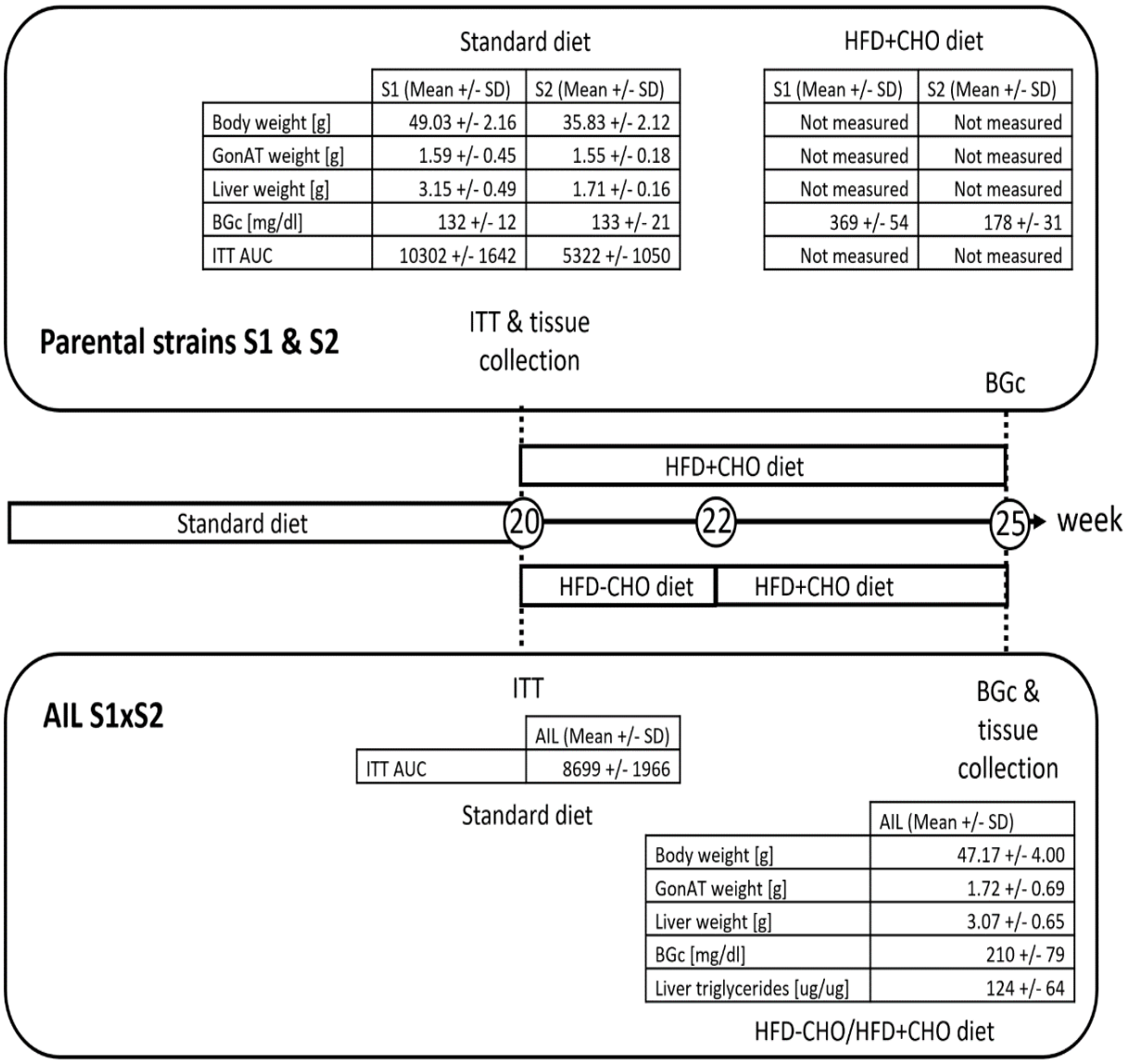


Figure 2.2. Response of parental lines S1 and S2, and AIL males to high-fat, high-carbohydrate diet. Abbreviations: GonAT, gonadal adipose tissue; BGc, blood glucose concentration; ITT, insulin tolerance test; AUC, area under the curve. HFD-CHO, High-fat/low-carbohydrate diet; HFD+CHO, high-fat/high-carbohydrate diet.

Correlation between traits

In metabolically healthy individuals of our AIL we would expect high body weights associated with high adipose tissue weight, unchanged liver weight and normal glucose clearance. In contrast, the correlation analysis showed no correlation of body weight with gonadal adipose tissue weight (r = 0.02, p = 0.65), but a positive correlation with liver weight (r = 0.65, p = 2.20E-16) and liver triglycerides (r = 0.39, p = 1.94E-15) (Table 2.1). Moreover, a negative correlation was found between gonadal adipose tissue and liver weight (r = -0.47, p = 2.2E-16), and gonadal adipose tissue weight and liver triglycerides (r = -0.33, p = 2.88E-11). Low adipose tissue weight together with high body weight and high liver weight was also associated with a large area under the curve for the blood glucose concentration in the ITT (r = -0.41, p = 2.2E-16) and high blood glucose concentration (r = -0.59, p = 2.2E-16). Consistent with the negative correlation coefficients for adipose tissue weights and the other parameters, positive correlations were found between liver weights and the same parameters (Table 2.1).

Table 2.1. Correlation coefficients between the collected traits in the AIL (BFMI861-S1xBFMI861-S2).

	Liver weight	Liver triglycerides/protein	BGc	ITT AUC	Body weight
	(r and P-value)	(r and P-value)	(r and P-value)	(r and P-value)	(r and P-value)
GonAT weight	-0.47, 2.20E-16	-0.33, 2.88E-11	-0.59, 2.20E-16	-0.41, 2.20E-16	0.02, 0.65
Liver weight		0.55, 2.20E-16	0.71, 2.20E-16	0.27, 3.98E-08	0.65, 2.20E-16
Liver triglycerides/protein			0.45, 2.20E-16	0.28, 1.41E-08	0.39, 1.94E-15
BGc				0.34, 9.55E-13	0.31, 2.55E-10
ITT AUC					0.21, 2.56E-05

Abbreviations: GonAT, gonadal adipose tissue; BGc, blood glucose concentration; ITT, insulin tolerance test; AUC, area under the curve.

QTL mapping

QTL mapping was performed for body weight, gonadal adipose tissue weight, liver weight, liver triglycerides, and blood glucose concentration at the end of the experiment and for ITT AUC at 20 weeks before the diet switch. The QTL analysis on selectively genotyped 200 AIL males revealed significant loci on Chr 3, 15, 16, and 17. The follow up analysis after KASP genotyping including all 397 males confirmed all four QTL and provided true estimates for the genetic effects (Table 2.2).

Three significant QTL for gonadal adipose tissue weight were identified on Chr 17 (Gatlgq) at 25.25 Mb (LOD = 7.3), Chr 3 (Gatq1) at 98.19 Mb (LOD = 6.3), and on Chr 15 (Gatq2) at 68.46 Mb (LOD = 4.2). Interestingly, for these three QTL the S1 allele always decreased the amount of adipose tissue. Gatlgq had also an effect on liver weight (LOD = 7.5) which could be caused by an increased hepatic fat storage. Indeed, liver triglycerides show a significant effect on Gatlgq based on genome wide multiple testing correction when we consider the selectively genotyping of the initial 200 selected animals (LOD = 4.8). Mapping liver triglycerides using the whole population there is still an effect (LOD = 2.4). However, this effect does not reach the threshold for genome wide significance (<0.05) but is still suggestive ($P < 0.1$). Furthermore, Gatlgq and Gatq1 affected the blood glucose concentration (LOD Gatlgq = 8, LOD Gatq1 = 4.2). For Gatlgq the allele of the insulin resistant S1 line was responsible for low adipose tissue weight, elevated liver weight, higher liver triglycerides and high blood glucose concentration. The S1-allele effects of the Gatlgq on gonadal adipose tissue weight and liver weight and on gonadal adipose tissue weight and liver triglycerides were in opposing direction, supporting the negative correlation between the traits. In contrast, for Gatq1 the S1-QTL allele decreased the adipose tissue weight, reduced the blood glucose concentration (Table 2.2, Figure 2.3b) and was associated with faster glucose clearance in the ITT (LOD = 3.8).

A QTL for body weight was mapped on Chr 16 (Bwq26) at 11.12 Mb (LOD = 7.1). At this locus, the S1 allele was increasing body weight (Table 2.2).

Since Gatlgq and Gatq1 showed pleiotropic effects on several traits, causal modeling was

performed. Causal modeling of Gatlgq suggests direct effects on gonadal adipose tissue, liver weight, and blood glucose concentration. Causal modeling of Gatq1 showed a direct effect of Gatq1 on gonadal adipose tissue weight, which in turn affects blood glucose concentration.

Table 2.2. Position and effects of QTL identified in the AIL population of 397 mice.

Traits	QTL confidence interval					LOD(BH)	Var %	Mean S1	Δ Mean S1-HET	Δ Mean S1-S2
	QTL name	Chr	StartPos	TopPos	StopPos					
GonAT weight [g]	Gatlgq	17	9 483 181	25 258 903	25 391 933	7.3	8.2	1.26	-0.54	-0.61
Liver weight [g]		17	9 483 181	25 258 903	25 391 933	7.5	8.3	3.32	0.26	0.43
BGc [mg/dl]		17	11 934 634	25 258 903	26 054 796	8	9	260	54	73
GonAT weight [g]	Gatq1	3	95 763 020	98 196 163	100 780 367	6.3	6.4	1.49	-0.27	-0.47
BGc [mg/dl]		3	95 763 020	98 196 163	100 543 098	4.2	4.2	194	-13	-44
Body weight [g]	Bwq26	16	3 892 297	11 120 784	21 355 904	7.1	7.2	48.28	1.73	3.15
GonAT weight [g]	Gatq2	15	67 855 285	68 461 862	74 582 319	4.2	4.1	1.49	-0.23	-0.37

Abbreviations: GonAT, gonadal adipose tissue; BGc, blood glucose concentration; QTL, quantitative trait locus; Chr, chromosome number; StartPos, TopPos, and StopPos, position of the start of the QTL confidence interval, position of the SNP with the highest LOD score, and position of the end of the QTL confidence interval in base pairs, respectively; Positions are given according to the Mouse Genome Version MM10, GRCm38.p3. SNP, single-nucleotide polymorphism. The confidence interval gives the 1.5 LOD drop region of the top SNP position. A LOD score above 4.9 was deemed to be 'genome-wide highly significant' and above 4.2 was deemed 'genome-wide significant'; BH, Bonferroni correction; LOD, logarithm (base 10) of odds; Var %, percentage of total variance.

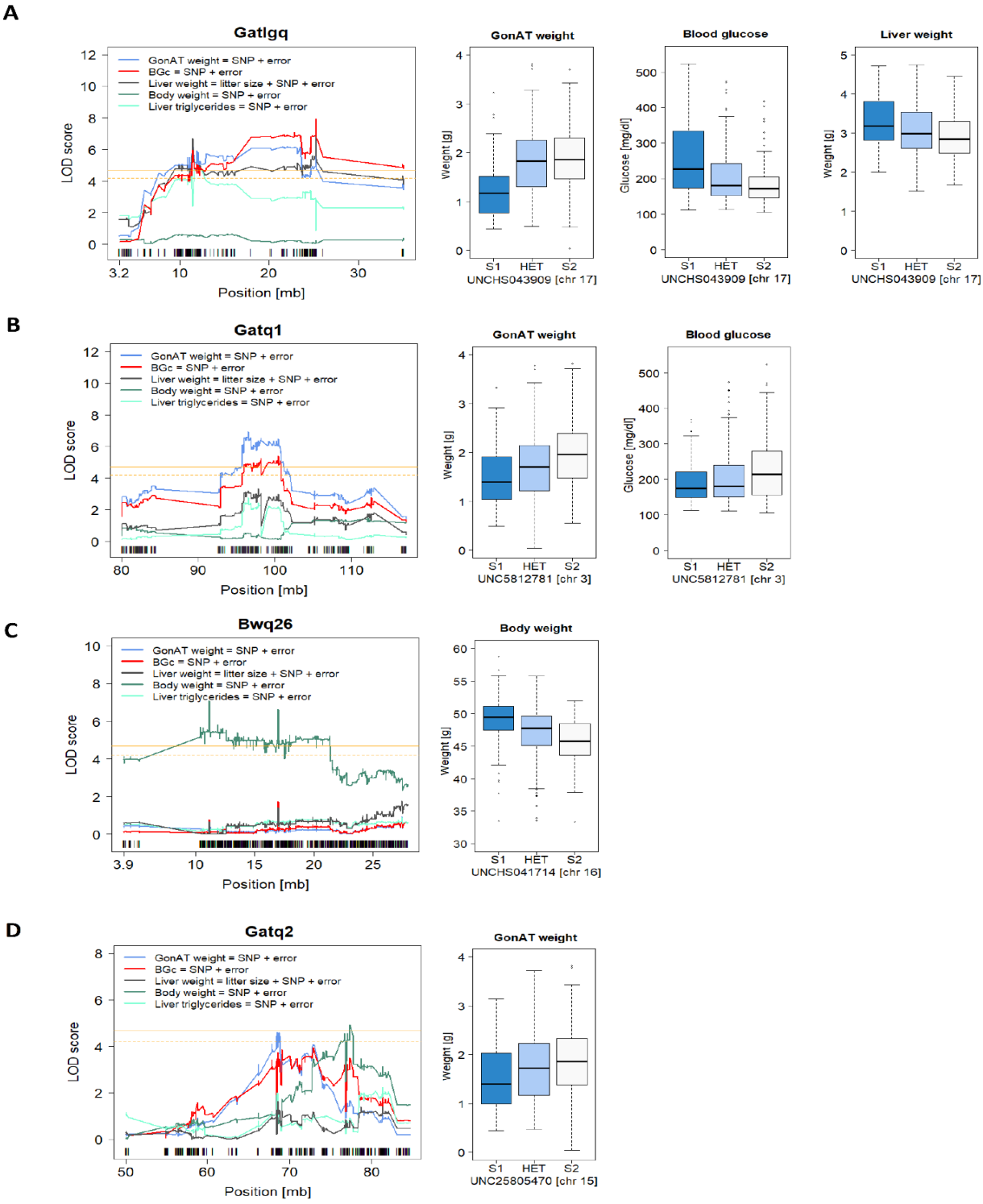


Figure 2.3. LOD score profiles and effect plots for top SNPs of significant traits (n =397) for (A) Gatlgq, (B) Gatq1, (C) Bwq26, (D) Gatq2. Blue line - gonadal adipose tissue weight, black line – liver weight, red line - blood glucose concentration, dark green line – body weight, light green line – liver triglycerides.

Abbreviations: QTL, quantitative trait locus; GonAT, gonadal adipose tissue; BGc, blood glucose concentration; SNP, single-nucleotide polymorphism; Chr, chromosome number; HET, heterozygous.

Candidate gene prioritization

The confidence intervals of the four significant QTL contain 534 protein-coding potential candidate genes. Sixty-two genes were polymorphic between S1 and S2; 27 in Gatlgq, 27 in Gatq1, 4 in Bwq26, 4 in Gatq2. Mutations in these genes were scored for their potential functional effects on the quality or expression level of the encoded protein according to the decision tree (Figure 2.5a). None of the genes carried a stop gain or stop loss mutation. Nevertheless, different mutations influencing protein sequence or gene regulation occurred. According to microarrays analysis, considering the 62 candidate genes, we found 37 genes differentially expressed between S1 and S2 in the gonadal adipose tissue, 8 in the liver, 8 in pancreatic islets and 3 in skeletal muscle.

Since correlation analysis of gene expression data between all examined animals showed that mice of the same mouse line clustered together only with gene expression data of the gonadal adipose tissue (Figure 2.5b), gonadal adipose tissue is suggested as the main tissue contributing to obesity, and glucose homeostasis in the S1 line.

Genes with the highest and second highest score in every QTL confidence interval were regarded as top candidates (Table 2.3). Differences in the expression of the top candidate genes for each QTL were confirmed by quantitative real time PCR in both gonadal adipose tissue and liver (Figure 2.4). *Plg* (Plasminogen) and *Acat2* (Acetyl-CoA acetyltransferase 2, cytosolic) are the top candidates in Gatlgq. *Plg* was not differentially expressed but contains one tolerated missense variant in the low-complexity region, one SNP in the promoter and additional SNPs in enhancers and CTCF binding sites in mice of the S1 line. *Acat2* was lower expressed in gonadal adipose tissue of S1 mice ($p = 7.45E-06$) and carries a deleterious missense variant in the thiolase, N-terminal domain. In Gatq1 *Fmo5* (Flavin-containing monooxygenase 5) and *Notch2* (Notch homolog 2) were prioritized. *Fmo5* was lower expressed in gonadal adipose tissue ($p = 3.83E-03$) and liver (9.11E-08) of S1 versus S2 mice. The Fmo5 gene in S1 mice carries one tolerated missense variant in the FMO-like domain, SNPs in the promoter and additional SNPs in enhancers and untranslated regions. *Notch2* ($p = 1.29E-03$) was higher expressed in gonadal adipose tissue of S1 mice and carries one

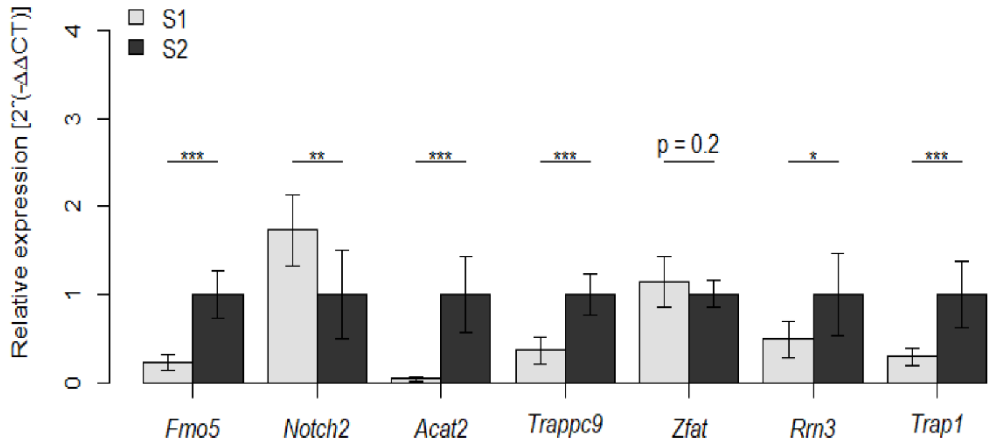
deleterious missense variant located in the EGF-like domain plus SNPs in untranslated regions in S1 mice. *Trap1* (TNF receptor associated protein 1) and *Rrn3 (*RRN3 homolog, RNA polymerase I transcription factor) ranked highest in Bwq26. Both candidate genes *Trap1* (p = 5.50E-05) and *Rrn3* (p = 2.05E-06) were lower expressed in gonadal adipose tissue, and *Rrn3* was additionally significantly lower expressed in the liver (p = 1.28E-06) of S1 mice. Both *Trap1* and *Rrn3* carry one tolerated missense variant. For Gatq2 *Trappc9* (Trafficking protein particle complex subunit 9) and *Zfat* (Zinc finger and AT hook domain containing) ranged as top candidates. *Trappc9* was lower expressed in S1 *versus* S2 mice in both gonadal adipose tissue (p = 5.89E-05) and liver (p = 1.92E-04). *Trappc9* possess variants in UTRs, CTCF binding sites, enhancer, and promoter. *Zfat* was higher expressed (p = 2.36E-03) in gonadal adipose tissue of S1 mice and it carries one deleterious missense variant in the low-complexity region in S1 mice.

Table 2.3. Expression differences of prioritized positional candidate genes between males of the parental lines BFMI861-S1 (n=7) and BFMI861-S2 (n=8). Bold indicates significant differences. The p-values are corrected according to Benjamini-Hochberg.

QTL name	Chr	Candidate Gene	Type of mutation	FC GonAT (S1/S2)	P-value GonAT	FC liver (S1/S2)	P-value liver	Gene score
Gatlgq	17	Plg	Tolerated domain missense, CTCF binds, enhancer and promoter variant	0.12	0.23	0.02	4.66E-03	10
	17	Acat2	Deleterious domain missense variant	-0.21	**7.45E-06**	-0.02	0.28	9
Gatq1	3	Fmo5	Tolerated domain missense, UTRs, enhancer, and promoter variant	-0.07	**3.83E-03**	-0.1	**9.11E-08**	12
	3	Notch2	Deleterious domain missense and UTRs variant	0.05	**1.29E-03**	-0.06	0.02	10
Bwq26	16	Trap1	Tolerated domain missense variant	-0.06	**5.50E-05**	-0.01	0.42	7
	16	Rrn3	Tolerated missense variant	-0.08	**2.05E-06**	-0.12	**1.28E-06**	4
Gatq2	15	Trappc9	UTRs, CTCF binds, enhancer, and promoter variant	-0.06	**5.89E-05**	-0.07	**1.92E-04**	9
	15	Zfat	Deleterious domain missense variant	0.05	**2.36E-03**	-0.03	0.25	8

Abbreviations: GonAT, gonadal adipose tissue; Chr, chromosome; FC, fold change.

A



B

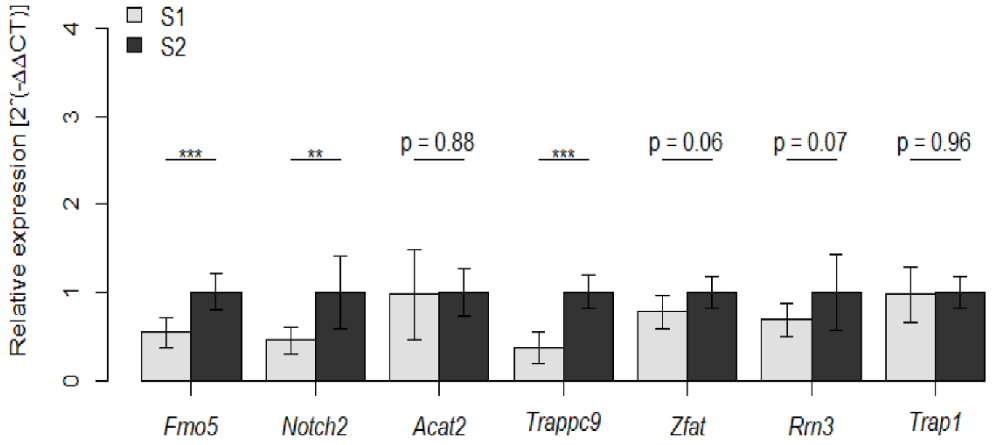


Figure 2.4. Validation of microarrays gene expression data for the top candidate genes through semi-quantitative real-time PCR in both (A) gonadal adipose tissue (S1: n=7, S2: n=8) and (B) liver (S1: n=7, S2: n=8). Given are transcript amounts in the S1 line relative to S2. Transcript amounts were normalized with the endogenous control *Gapdh*.

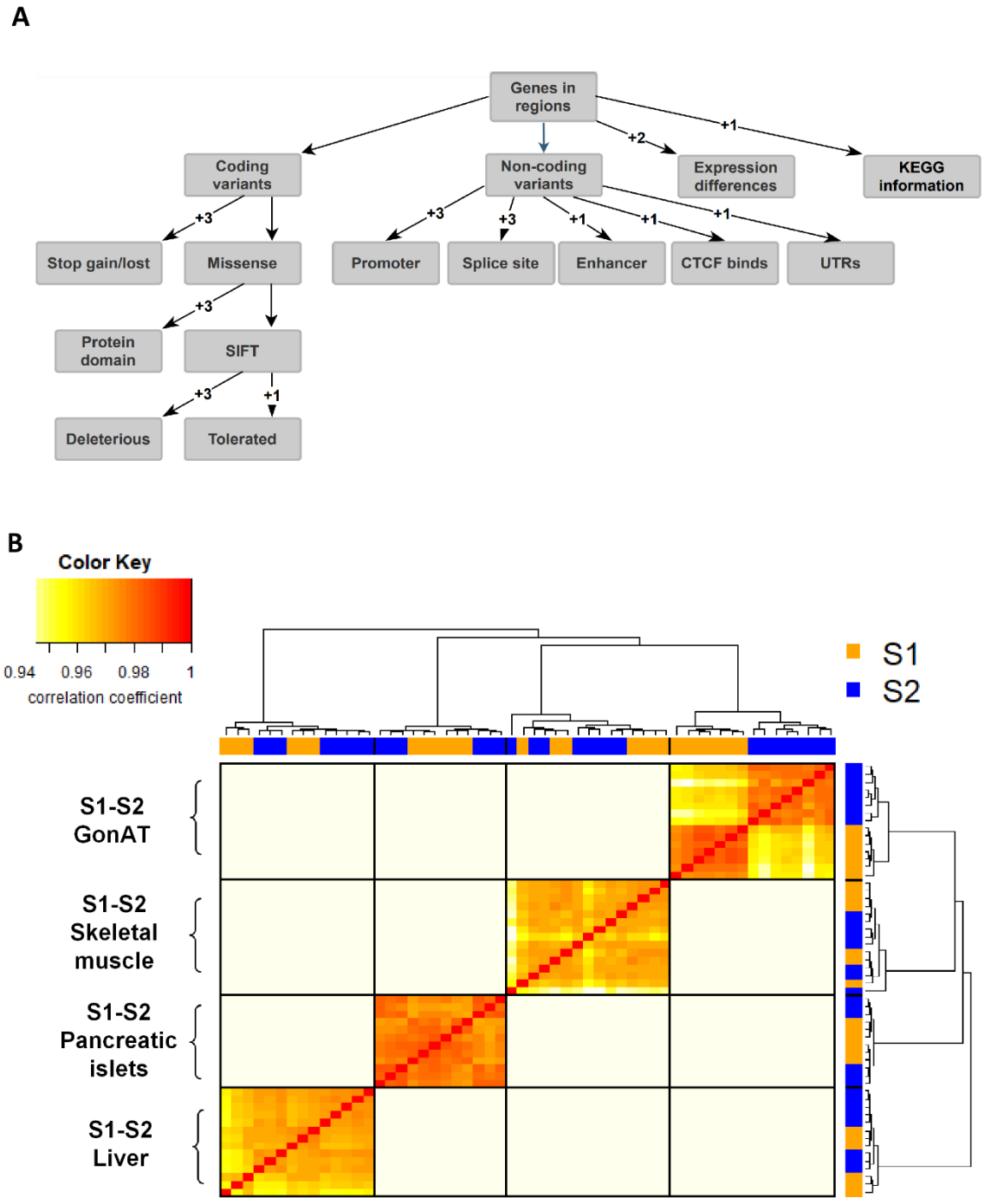


Figure 2.5. (A) Decision tree for prioritization of candidate genes located in a QTL region. Genes in a QTL region containing sequence variants between the parental lines S1 and S2 were ranked according to the sum of scores based on the functional annotation of coding and non-coding variants, gene expression data and KEGG information. (B) Heatmap and dendrogram of microarrays gene expression data from 4 different tissues (gonadal adipose tissue, skeletal muscle, pancreatic islets and liver) of the parental lines (S1 and S2).

2.4 Discussion and conclusions

To better understand the differences in insulin sensitivity in two sub-lines of the Berlin Fat Mouse independent of a major obesity QTL on Chr 3 (*jObes1*) and to unravel the genetic architecture underlying the observed aspects of the metabolic syndrome we investigated an advanced intercross population of the BFMI861 mouse lines S1 and S2. Besides being genetically closely related and sharing the known obesity locus on Chr3 (Arends et al., 2016), the two parental mouse lines differ extremely in their metabolic phenotype. The mouse line S1 showed clear features of the metabolic syndrome while S2 was also obese, but had a normal glucose homeostasis even under a high-fat, high-carbohydrate diet feeding. The extreme phenotypic data propose the examined mouse lines as an excellent model for studying the genetic determinants of traits of the metabolic syndrome. Due to the random mixture of the genomes of the BFMI861-S1 and -S2 lines, the AIL individuals showed a wide range of phenotypes. Different from expectations in metabolically healthy individuals, we found no correlation between body weight and gonadal adipose tissue weight and negative correlations between gonadal adipose tissue weight and all other traits in AIL males, while liver weight was positively correlated with all other traits. These findings indicate ectopic fat storage in the liver which was indeed confirmed by the assessment of hepatic triglycerides in our AIL. Hepatic fat storage is also present in individuals of the S1 line (Heise et al., 2016). Ectopic fat storage in the liver instead of storage in the adipose tissue as the major fat storage organ has been reported repeatedly as causal defect for later impaired glucose clearance (Parker, 2018) (Rosen & Spiegelman, 2006). Therefore, we suggest this shift as the likely driver for impaired glucose homeostasis in our mouse model. Gene expression data further supported the assumption of impaired adipose tissue function being causal for the observed phenotypes of the metabolic syndrome in S1 mice. For example, we found distinguished clusters of differentially expressed genes in gonadal adipose tissue between S1 and S2 animals but not in liver, muscle and pancreatic islets.

The overlap of QTL effects in some regions is consistent with the correlations that we found between the affected traits. For Gatlgq, the S1 allele reduces adipose tissue weight and

increases the liver weight and hepatic fat content. Moreover, by shifting fat storage from adipose tissue to ectopic storage in the liver blood glucose concentration is increased. In contrast, the S1 allele on Gatq1 contributes to lower adipose tissue weight, lower blood glucose concentration and increased insulin sensitivity.

To disentangle the direct and indirect genetic effects of the different QTL, we performed causal modelling. Using causal inference, we were able to provide evidence that out of the two QTL associated with blood glucose concentration in our population, likely only Gatlgq has a direct influence on blood glucose concentration. The second QTL, Gatq1 influences blood glucose concentration indirectly through the regulation of fat storage in adipose tissue, whose weight is directly affected by this QTL. A possible explanation for the discrepancy in the correlation of adipose tissue weight to blood glucose concentration of the two QTL could be that a reduced adipose tissue mass via the S1 allele of Gatlgq could indicate a shift towards ectopic fat storage which is reflected in elevated liver weight and liver triglycerides and thereby contributes to higher blood glucose concentration. In contrast, Gatq1 and Gatq2 could harbor S1 alleles protecting against obesity resulting in lower adipose tissue weight which for Gatq1 is accompanied by lower blood glucose concentration. However, the overall phenotype of the insulin resistant S1 line appears to be driven mainly by the larger effects of Gatlgq. These findings provide strong evidence for the importance of direct genetic effects on adipose tissue, which indirectly contribute to the etiology of impaired glucose homeostasis.

The AIL population used in this study has the advantage of having a high mapping resolution with respect to the call of positional candidate genes. Because the examined AIL accumulates chromosomal recombination over 10 generations, the physical length of the QTL regions is relatively short and thereby the number of positional candidate genes low. In our study, the number of positional candidate genes could be further reduced because long chromosomal stretches are identical between the closely related mouse lines S1 and S2 and, therefore, genes in these regions are not polymorphic and can be excluded from further studies, resulting in 62 out of 534 protein coding genes as positional candidate genes.

In our prioritization approach, *Plg* and *Acat2* are the top candidate genes for direct effects of Gatlgq on gonadal adipose tissue, liver weight, and blood glucose concentration. *Plg* possesses one tolerated missense variant in the low-complexity region of the protein in S1 mice, a region that is significant for the functionality of this protein (Ntountoumi et al., 2019). *Plg*-knockout mice have lower amounts of adipose tissue (Hoover-Plow, Ellis, & Yuen, 2002). During cell differentiation plasminogen binding is increasing in 3T3 cells and isolated adipocytes suggesting a role in adipose tissue development (Hoover-Plow & Yuen, 2001). In humans, *Plg* was reported to be relevant for the development of insulin resistance and diabetes (Ghosh & Vaughan, 2012) (Ajjan et al., 2013) (Qi, Workalemahu, Zhang, Hu, & Qi, 2012). Thus, *Plg* could be causal for the impaired glucose homeostasis by modifying adipose tissue in S1 mice. *Acat2* is involved in the biosynthesis of fatty acids and cholesterol and is mainly expressed in the liver and intestine (Fukao et al., 1997). In S1 mice, *Acat2* carries one deleterious mutation that leads to a Valine/Methionine substitution at amino acid position 216 located in the conserved N-terminal domain. This domain is important for the thiolase activity and a mutation in this region could have effects on the protein function. Thus, *Acat2* is a good candidate for the observed hepatic fat storage in S1 mice.

Fmo5 and *Notch2* were identified as the most likely candidate genes in Gatq1 affecting gonadal adipose tissue weight directly. S1 mice carry two SNPs in the promoter of *Fmo5*. According to the Ensembl database one SNP affects a transcription factor binding site for *Elf5*. The other SNP affects two transcription factors binding sites; one for *Rxra* and one for multiple transcription factors such as *Nr2f6*, *Rara*, *Rarb* and *Rarg*. According to Bgee database, all identified transcription factors that potentially bind to the promoter region of *Fmo5* are expressed in adipose tissue. The identified SNPs in the promoter of *Fmo5* could therefore be responsible for its lower expression in the gonadal adipose tissue and liver of S1 mice. Consistent with the QTL allele effect of S1 mice leading to lower adipose tissue weight and lower blood glucose concentration, *Fmo5* knockout mice store less fat in gonadal adipose tissue and have lower blood glucose concentration at 20 weeks (Gonzalez Malagon et al., 2015). *Notch2* is important for developmental processes by controlling cell fate

decisions (Siebel & Lendahl, 2017) and lipid storage (Bissonnette, Lane, & Chang, 2017). *Notch2* has been linked to type 2 diabetes in humans (Zeggini et al., 2008). S1 mice carry a deleterious mutation leading to a glycine/serine substitution at amino acid position 136. This mutation resides in the EGF-like domain which is important for Notch2 activation (Lai, 2004) and could be causal for the low-fat deposition in gonadal adipose tissue found in S1 mice.

Based on the findings of this genetic study, additional research is necessary to further validate the suggested candidate genes. This could be done by knockout of certain genes, or through continuation of the AIL to reduce the physical length of QTL regions and thereby the number of candidate genes. It is important to note that, although we have prioritized candidate genes using all available information, we cannot completely rule out that one of the polymorphic genes or even an unannotated gene was wrongly discarded.

The human metabolic syndrome is a complex disease with many actors, many still unknown, contributing to its expression. The identification of new potential partners in the network by QTL analysis and subsequent data analysis could help to replenish the gaps.

2.5 Acknowledgements

The project was funded by the Deutsche Forschungsgemeinschaft (DFG, HE8165/1-1). We thank Marion Bütow, Ulf Kiessling, and Ines Walter for assistance in animal care. We also thank Anett Helms and Franziska Gabler for pancreatic islet isolation and RNA extraction.

Chapter 3: QTL mapping on body weight time series data identifies two body weight QTL in the BFMI861 lines

To this work contributed:
Manuel Delpero[1], Danny Arends[1], Gudrun A. Brockmann[1], Deike Hesse[1]

Affiliation:
[1] Albrecht Daniel Thaer-Institut für Agrar- und Gartenbauwissenschaften, Humboldt-Universität zu Berlin, Berlin, Germany.

Background: The Berlin Fat Mouse Inbred line (BFMI) is a model for obesity and the metabolic syndrome. Sublines of BFMI, BFMI861-S1 and BFMI861-S2, differ in body weight despite being genetically very similar and sharing the known *jObes1* locus on Chr 3 which is responsible for 40% variance in the body weight in all BFMI lines. This study aimed to identify additional body weight QTL by using time series body weight measurements.

Methods: In generation 10, 397 males of an advanced intercross line (AIL) BFMI861-S1 x BFMI861-S2 were weighted weekly until week 25. Mice were challenged with a high-fat, high-carbohydrate diet for the last 3 weeks following 2 weeks of a high-fat, free-carbohydrate diet to enhance differences in body weight development. QTL-analysis on body weight time series data was performed after selective genotyping of 200 mice using the GigaMUGA Genotyping Array. Additional 197 males were genotyped for 2 top SNPs in QTL regions.

Results: One QTL for body weight from week 9 to week 25 was mapped on Chr 16 from 10.48 Mb to 21.5 Mb. At this locus, the BFMI861-S1 allele increased body weight by 11.6 %. An additional genome wide significant QTL for body weight from week 9 to week 20 was mapped on Chr 15 from 68.46 Mb to 77.77 Mb. The S1 allele on the QTL on Chr 15 increased body weight by 13.8 %. Prioritization of candidate genes identified *Trap1* and *Rrn3* for the QTL on Chr 16 and *Trappc9* and *Zfat* for the QTL on Chr 15 as most likely candidates for exerting the effect.

Conclusions: QTL mapping together with a detailed prioritization approach allowed us to identify two QTL (one novel on Chr 15) associated with body weight using two lines of the Berlin Fat Mouse. These results helped us to understand which additional genomic regions contribute to increased body weight in the Berlin Fat Mouse by using body weight data that are normally collected in every mouse study.

3.1 Introduction

More than one-third of the world's adult population is overweight, and the incidence is further increasing (Locke et al., 2015). Besides overeating and reduced activity associated energy expenditure, heritability plays a major role in the development of obesity.

Animal models and in particular mouse models are essential to enhance genetic discoveries and to reveal the genetic contributions of complex diseases. Weekly body weights are often collected during animal experiments including quantitative trait locus (QTL) studies to monitor animal health and are used as an additional phenotype for mapping. Usually, these time series data are not considered and only analyzed as endpoint measuring points. However, QTL mapping with body weight measurements collected weekly could help to identify further genes linked to obesity and related metabolic disorders that are only visible during the time course and could even loose effect at later time points.

Originally, the Berlin Fat Mouse population was selected for juvenile obesity. After 58 generations of selection, different Berlin Fat Mouse Inbred (BFMI) lines were generated through repeated brother-sister mating (Heise et al., 2016). In a cross between the most obese inbred line BFMI860 and the lean control line C57BL/6NCrl, we have previously identified a recessive genetic defect at a locus on chromosome (Chr) 3 accounting for 40 % of the variance in adipose tissue weight at 6 weeks (Arends et al., 2016). This juvenile obesity

locus (*jObes1*) is fixed in all BFMI sublines.

The goal of the current study was to identify additional genetic factors (besides the *jObes1* region) contributing to elevated body weight in the Berlin Fat Mouse.

To detect novel genetic factors contributing to body weight we used the inbred lines BFMI861-S1 (S1) and BFMI861-S2 (S2). S1 and S2 are sublines created from the BFMI860, as such the BFMI860 is the predecessor of the S1 and S2. These two lines were derived from one parental line that was divided into two sub-lines only after four generations of inbreeding. Therefore, these two lines are genetically highly similar, and the remaining genetic diversity is responsible for phenotypic differences. Although genetically close, the S1 and S2 lines are quite different with respect to metabolic traits (Heise et al., 2016). In particular, the S1 line shows higher body weight, hepatic fat storage, low insulin sensitivity, and impaired glucose tolerance. In contrast, S2 is insulin sensitive despite being obese (Heise et al., 2016). To identify more genetic loci accounting for the observed difference in body weight, in addition to the previously identified *jObes1* locus, we performed a quantitative trait locus (QTL) mapping study in an advanced intercross line (AIL) which was generated from an initial cross between the BFMI861 lines S1 and S2.

The advantage of this AIL is that by crossing two BFMI lines we can naturally correct for the large effect on body weight of the *jObes1* locus and potentially unravel additional, until now hidden, minor QTL that contribute to differences in body weight plus investigate additional differences in metabolic traits between the BFMI lines.

By performing QTL mapping in this AIL four novel QTL for traits of the metabolic syndrome (gonadal adipose tissue weight, liver weight, blood glucose concentration, liver triglycerides, and body weight) were already successfully identified using end point measurements (Delpero et al., 2021). In the current study we focused on time series body weight data that were collected in this population every week over a period of 25 weeks. This data allowed us to identify additional obesity QTL that contribute to the overall obese phenotype peculiar of the BFMI lines in addition to the known *jObes1* locus on Chr 3.

3.2 Material and Methods

Mouse population

We used male mice of the AIL population in generation 10 generated from an initial cross between a BFMI861-S1 male and a BFMI861-S2 female followed by repeated random mating in every generation. For randomization of mating pairs, the program RandoMate (Schmitt et al., 2009) was used. The BFMI861 lines S1 and S2 were generated as described in Heise et al., 2016 (Heise et al., 2016).

Animal husbandry

All experimental treatments of mice were approved by the German Animal Welfare Authorities (approval no. G0235/17). Mice were kept under conventional conditions with a 12:12 h light–dark cycle (lights on at 0600 hours) and at a temperature of 22 ± 2 °C. Mice had *ad libitum* access to food and water.

Experiment phenotyping

AIL mice were fed as described in Delpero et al. 2021 (Delpero et al., 2021).

AIL mice were phenotyped between the age of 3 (after weaning) and 25 weeks. Body mass was recorded weekly. At 25 weeks mice were anesthetized with isoflurane and sacrificed (Hesse et al., 2018).

Outliers, defined as individuals which have a measurement that deviates from the population mean by more than four standard deviations (SD), were removed from the data.

Genotyping

Out of the 397 males that were phenotyped, 200 mice were genotyped as described in Delpero et al. 2021 (Delpero et al., 2021). Due to high genetic similarity of the parental lines S1 and S2 of the AIL population, only 5,215 out of 143,259 SNPs on the array were informative and passed the quality control (Delpero et al., 2021).

Remaining 197 males of the AIL population were genotyped for 2 top markers identified to be associated with body weight in preliminary analysis (see QTL mapping section). The additional animals were genotyped to counteract any bias in the estimates of allele effect sizes introduced by selective genotyping. For these markers, KASP genotyping assays were developed as described previously (Kreuzer et al., 2013).

QTL mapping

QTL mapping was performed for each body weight time point separately in two steps: First, a QTL scan was performed using the 200 males that were genotyped with the GigaMUGA Array. Afterwards, a final QTL scan was performed including all animals (genotyped by GigaMUGA and KASP).

Covariates (subfamily and litter size) were investigated for a significant influence on body weight at each time point.

To confirm that QTL mapping models are valid, residuals of the models were tested for normality using a Shapiro-Wilk test. If residuals were found to not be normally distributed, a non-parametric Kruskal–Wallis one-way analysis of variance was performed to validate the top marker.

The number of independent statistical tests was estimated by simpleM (Gao, 2011) which determined the number of independent tests to be 849 (window size = 820, mEff = 849). Afterwards Bonferroni correction for multiple testing correction (Haynes, 2013) was performed using the number of independent SNPs as determined by simpleM. P-values were converted to LOD scores, using LOD = -log10(p-value). LOD scores above 4.9 and 4.2 were deemed to be genome-wide highly significant and significant, respectively. The 95% confidence interval of a QTL was determined by a 1.5 LOD drop from the top SNP position (Dupuis & Siegmund, 1999). Start and end positions were defined as the first SNP upstream or downstream of the 1.5 LOD-drop confidence interval.

Candidate genes prioritization

Prioritization of candidate genes in each QTL region was performed as described in Delpero et al. 2021 (Delpero et al., 2021).

3.3 Results

QTL mapping

QTL mapping was performed for body weight with data collected once a week from week 9 until the end of the experiment (week 25). QTL analysis on selectively genotyped 200 AIL males revealed significant loci on Chr 15 and 16 (Figure 3.1 left and Figure 3.1 right, respectively).

The follow up analysis after KASP genotyping including all 397 males confirmed the two QTL and provided true estimates for the genetic effects (Table 3.1).

In detail, a QTL for body weight was mapped on Chr 15 from 68.46 Mb to 77.77 Mb from week 9 to week 20. This region contains 129 protein-coding genes. At this locus the S1 allele increased body weight (Table 3.1). The most significant SNP of this region was UNCHS040893 at week 20 (78,820,549; LOD = 7.67). In detail, homozygous BFMI861-S1 mice showed 13.8 % higher body weight compared to homozygous BFMI861-S2 mice (mean BFMI861-S1 = 44.4 ± 4.04 g, mean BFMI861-S2 = 39 ± 3.76 g) and 7.8 % elevated compared to heterozygous mice (mean heterozygous mice = 41.2 ± 4.47 g) at week 20.

Another genome-wide significant QTL for body weight from week 9 to week 25 was mapped on Chr 16 from 10.48 Mb to 21.5 Mb. This region contains 158 protein-coding genes and the most significant SNP of this region was UNCHS041907 at week 18 (16,995,303; LOD = 11.84). At this locus, the S1 allele increased body weight (Table 3.1). Homozygous BFMI861-S1 mice showed 11.6 % higher body weight compared to homozygous BFMI861-S2 mice (mean BFMI861-S1 = 40.4 ± 3.15 g, mean BFMI861-S2 = 36.2 ± 3.84 g) and 6 % elevated compared to

heterozygous mice (mean heterozygous mice = 38.15 ± 3.39 g).

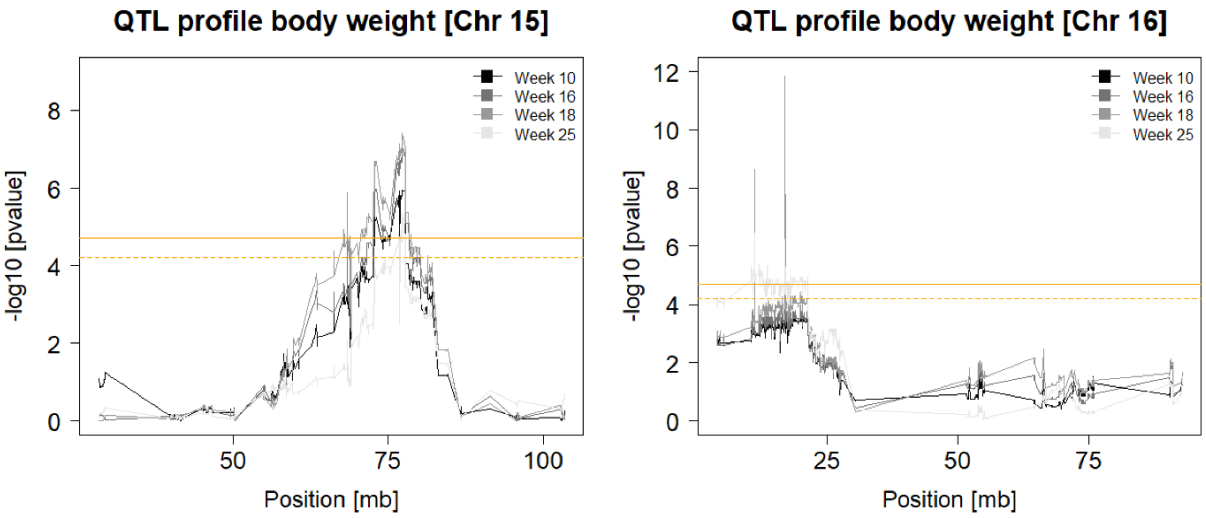


Figure 3.1. LOD curve across Chr 15 (left) and Chr 16 (right) for body weight collected at different time points. The orange (1%) and orange dashed (5%) horizontal lines mark the significance thresholds.

Table 3.1. Position and effects of body weight QTL identified in the AIL (BFMI861-S1xBFMI861-S2) of 397 mice.

QTL name	Age [weeks]	QTL region				LOD	Mean S1/S1	Mean S1/S2	Mean S2/S2
		Chr	Start [bp]	Top [bp]	Stop [bp]				
BwChr15	9	15	68,461,862	76,820,549	77,770,986	5.82	34	32.7	31.4
	10	15	68,461,862	76,820,549	77,770,986	5.92	35.7	34	32.7
	11	15	68,461,862	76,820,549	77,770,986	6.52	37.2	35.2	33.8
	12	15	68,461,862	76,820,549	77,770,986	6.6	37.8	36.4	35.3
	13	15	68,461,862	76,820,549	77,770,986	6.45	40.3	37.8	36.3
	14	15	68,461,862	76,820,549	77,770,986	6.75	41.5	38.9	37.2
	15	15	68,461,862	76,820,549	77,770,986	6.56	42.2	38.7	37.9
	16	15	68,461,862	76,820,549	77,770,986	7.02	43.4	40.1	38.8
	17	15	68,461,862	76,820,549	77,770,986	6.45	44.1	41.6	39.4
	18	15	68,461,862	76,820,549	77,770,986	7.15	40.9	38.4	36.3
	19	15	68,461,862	76,820,549	77,770,986	6.66	44.5	41	39.1
	20	15	68,461,862	76,820,549	77,770,986	7.67	44.4	41.2	39
BwChr16	9	16	3,892,297	16,995,303	21,355,904	8.08	33.6	32.5	31.3
	10	16	3,892,297	16,995,303	21,355,904	9.52	35.2	33.9	32.5
	11	16	3,892,297	16,995,303	21,355,904	9.91	36.5	35	33.5
	12	16	3,892,297	16,995,303	21,355,904	10.82	38.1	36.4	34.7
	13	16	3,892,297	16,995,303	21,355,904	10.85	39.4	37.6	35.7
	14	16	3,892,297	16,995,303	21,355,904	11.59	40.7	38.5	36.7
	15	16	3,892,297	16,995,303	21,355,904	10.91	41.6	39.4	37.6
	16	16	3,892,297	16,995,303	21,355,904	10.93	42.7	40.5	38.5
	17	16	3,892,297	16,995,303	21,355,904	10.5	43.6	41.3	39.3
	18	16	3,892,297	16,995,303	21,355,904	11.84	40.4	38.1	36.2
	19	16	3,892,297	16,995,303	21,355,904	9.97	43.3	41.1	38.8
	20	16	3,892,297	16,995,303	21,355,904	11.49	42.2	40.38	39
	21	16	3,892,297	16,995,303	21,355,904	6.13	43.2	41.2	39.4
	22	16	3,892,297	16,995,303	21,355,904	5.69	44.82	42.6	41
	23	16	3,892,297	16,995,303	21,355,904	6.32	47.3	45	43.5
	24	16	3,892,297	16,995,303	21,355,904	6.46	48.1	46	44.7
	25	16	3,892,297	16,995,303	21,355,904	6.25	49.3	47.2	45.2

Abbreviations: QTL, quantitative trait locus; Chr, chromosome number; StartPos, TopPos, and StopPos, position of the start of the QTL confidence interval, position of the SNP with the highest LOD score, and position of the end of the QTL confidence interval in base pairs, respectively; Positions are given according to the Mouse Genome Version MM10, GRCm38.p6. SNP, single-nucleotide polymorphism. The confidence interval gives the 1.5 LOD drop region of the top SNP position. A LOD score above 4.9 was deemed to be 'genome-wide highly significant' and above 4.2 was deemed 'genome-wide significant'; LOD, logarithm (base 10) of odds.

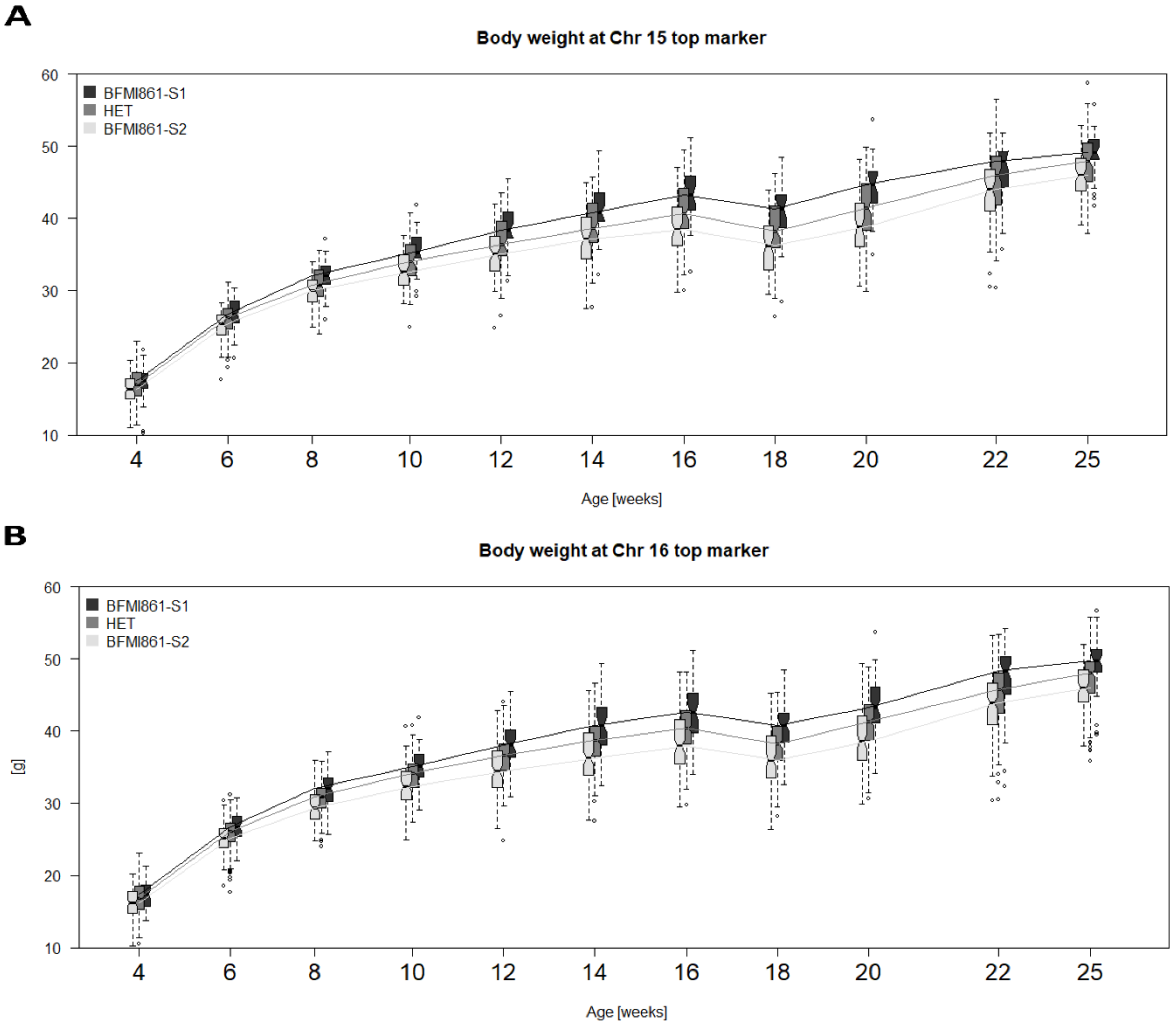


Figure 3.2. Boxplots for 397 mice of the AIL (BFMI861-S1xBFMI861-S2) in generation 10 aged 4-25 weeks and curves depicting body weight development. For every time point, boxplots for all three genotype classes (BFMI861-S1 homozygous; HET, heterozygous; BFMI861-S2 homozygous) are shown for SNP UNCHS040893 located at the top position on Chr 15 (A) and for SNP UNCHS041907 located at the top position on Chr 16 (B). Outliers are represented as dots.

Prioritization of positional candidate genes

The confidence intervals of the two significant QTL contain 339 protein-coding potential candidate genes. Eight were polymorphic between S1 and S2; four in the Chr 15 QTL, and four in Chr 16 QTL. Mutations in these genes were scored for their potential functional effects on the quality or expression level of the encoded protein according to the decision tree as

described in Delpero et al., 2021 (Delpero et al., 2021). None of the candidate genes carried a loss of function mutation. Nevertheless, different mutations influencing protein sequence or gene regulation occurred.

Considering the QTL on Chr 15, *Trappc9* (trafficking protein particle complex subunit 9) and *Zfat* (zinc finger and AT hook domain containing) ranked as top candidates (Table 3.2). *Trappc9* was lower expressed in S1 *versus* S2 mice in both gonadal adipose tissue (p = 5.89E-05) and liver (p = 1.92E-04). Furthermore, *Trappc9* possesses variants in UTRs, CTCF binding sites, enhancer, and promoter. *Zfat* was higher expressed (p = 2.36E-03) in gonadal adipose tissue of S1 mice and it carries one deleterious missense variant in the low-complexity region in S1 mice.

Trap1 (TNF receptor-associated protein 1) and *Rrn3 (*RRN3 homolog, RNA polymerase I transcription factor) ranked highest for the QTL on Chr 16 (Table 3.2). Both candidate genes *Trap1* (p = 5.50E-05) and *Rrn3* (p = 2.05E-06) were lower expressed in gonadal adipose tissue, and *Rrn3* was additionally significantly lower expressed in the liver (p = 1.28E-06) of S1 mice. Both *Trap1* and *Rrn3* carry one tolerated missense variant.

Table 3.2 Top candidate genes after applying the prioritization criteria.

QTL name	Chr	Candidate Gene	Type of mutation	FC GonAT (S1/S2)	P-value GonAT	FC liver (S1/S2)	P-value liver	Gene score
BwChr15	15	Trappc9	UTRs, CTCF binds, enhancer, and promoter variant	-0.07	**5.89E-05**	-0.07	**1.92E-04**	9
	15	Zfat	Deleterious domain missense variant	0.05	**2.36E-03**	-0.03	0.25	8
BwChr16	16	Trap1	Tolerated domain missense variant	-0.06	**5.50E-05**	-0.01	0.42	7
	16	Rrn3	Tolerated missense variant	-0.08	**2.05E-06**	-0.12	**1.28E-06**	4

Abbreviations: GonAT, gonadal adipose tissue; Chr, chromosome; FC, fold change. Bold indicates significant differences. The *p*-values are corrected according to Benjamini-Hochberg.

3.4 Discussion and conclusion

To better understand the differences in body weight in the two sublines of the Berlin Fat Mouse BFMI861-S1 and BFMI861-S2 that show 96.4 % of genetic similarity (Delpero et al., 2021), we investigated an advanced intercross population of the BFMI861 mouse lines S1 and S2. Besides being genetically closely related these two BFMI lines share the known juvenile obesity locus on Chr 3 which has a 40 % contribution to the overall variance in obesity in all BFMI lines (Arends et al., 2016).

Performing QTL mapping on time series body weight data, we identified a QTL for body weight on Chr 16 which explains 12% of body weight variance in the AIL population at 20 weeks. An additional QTL for body weight was identified on Chr 15 which explains 8.5% of the variance at week 20. The effects of the two QTL were highly significant during a long period from week 9 until week 25.

The QTL on Chr 16 had been identified in the AIL (BFMI861S1xB6N) before (Delpero et al., 2021), where we were mapping QTL for body weight at week 25. In the current study we associated this QTL also with body weight at younger age from week 9 until week 25. Yet the Chr 15 QTL had not been identified before, because the QTL effect is genome widely significant only before week 20 and the effect is lost afterwards. The time-series analyses of body weight facilitated the mapping of QTL which are hidden at later age but play a role for the evolvement of adult body weight.

The candidate genes for the novel QTL on Chr 15 associated with body weight are *Trappc9* and *Zfat*. These two genes have been previously identified as top candidate genes for a QTL associated with blood glucose concentration and gonadal adipose tissue weight (Gatq2) in the AIL (BFMI861-S1xBFMI861S2) (Delpero et al., 2021). The identification of the same candidate genes for the previously identified Gatq2 and for the novel body weight QTL is explained by the 6 Mb overlap between the two QTL. *Trappc9* and *Zfat* could therefore be responsible partially for the higher body weight observed in the BFMI861-S1 line, and at the same time for the differences in gonadal adipose tissue weight and blood glucose concentration between the AIL parental lines.

The selected prioritization approach allowed to identify *Trap1* and *Rrn3* as top candidate genes for the QTL on Chr 16 which were previously described in Delpero et al. 2021 (Delpero et al., 2021) to be associated with final body weight at week 25. This finding on schedule draw conclusions on the evolvement of the trait over time and the time pattern could then be brought together with puberty or maturity. While the juvenile obesity QTL on Chr 3 is responsible for juvenile obesity in all BFMI lines (Arends et al., 2016), this QTL on Chr 16 affects the persistence of obesity in the BBFMI-861-S1 mouse line at the later age.

In the current study we identified one novel QTL for body weight and confirmed one previously identified body weight QTL in our population by using time series data. The identification of these two QTL which are significant over a wide range of ontogenetic development help us to unravel the genetic puzzle that is driving the higher body weight observed in the BFMI lines over time.

Obesity is a complex trait driven by multiple genetic and environmental factors. While environmental factors are well known, many genetic factors are still unknown. QTL mapping and the subsequent identification of genomic regions and candidate genes responsible for obesity in both mice and humans are important to help understand the genetic component contributing to this common human disease.

3.5 Acknowledgements

DH was funded by the Deutsche Forschungsgemeinschaft (DFG, HE8165/1-1). The project was supported by the DDG (Deutsche Diabetes Gesellschaft). We thank Marion Bütow, Ulf Kiessling, and Ines Walter for assistance in animal care and Maximilian Sprechert for technical assistance.

Chapter 4: QTL mapping in the cross BFMI861-S1xB6N identifies additional candidate genes for obesity and fatty liver disease

Manuel Delpero[1], Danny Arends[1], Aimee Freiberg[1], Gudrun A. Brockmann[1], Deike Hesse[1]

Affiliation:
[1] Albrecht Daniel Thaer-Institut für Agrar- und Gartenbauwissenschaften, Humboldt-Universität zu Berlin, Berlin, Germany.

The content of this chapter was published in: Delpero, M., Arends, D., Freiberg A., Brockmann, G. A., Hesse, D., (2022). QTL-mapping in the obese Berlin Fat Mouse identifies additional candidate genes for obesity and fatty liver disease. *Scientific Reports*.
DOI: 10.1038/s41598-022-14316-5

Background: The Berlin Fat Mouse Inbred line (BFMI) is a model for obesity and the metabolic syndrome. This study aimed to identify genetic variants associated with liver weight, liver triglycerides, and body weight using the obese BFMI sub-line BFMI861-S1. BFMI861-S1 mice are insulin resistant and store ectopic fat in the liver.

Methods: In generation 10, 58 males and 65 females of the advanced intercross line (AIL) BFMI861-S1xB6N were phenotyped under a standard diet over 20 weeks. QTL-analysis was performed after genotyping with the MiniMUGA Genotyping Array. Whole genome sequencing and gene expression data of the parental lines was used for the prioritization of positional candidate genes.

Results: Three QTL associated with liver weight, body weight, and subcutaneous adipose tissue (scAT) weight were identified. A highly significant QTL on chromosome (Chr) 1 (157–168 Mb) showed an association with liver weight. A QTL for body weight at 20 weeks was found on Chr 3 (34.1 – 40 Mb) overlapping with a QTL for scAT weight. In a multiple QTL mapping approach, an additional QTL affecting body weight at 16 weeks was identified on Chr 6 (9.5-26.1 Mb). Considering sequence variants and expression differences, *Sec16b* and

Astn1 were prioritized as top positional candidate genes for the liver weight QTL on Chr 1; *Met* and *Ica1* for the body weight QTL on Chr 6. Interestingly, all top candidate genes have previously been linked with metabolic traits.

Conclusions: This study shows once more the power of an advanced intercross line for fine mapping. QTL mapping combined with a detailed prioritization approach allowed us to identify additional and plausible candidate genes linked to metabolic traits in the AIL (BFMI861-S1xB6N). By reidentifying known candidate genes in a different crossing population the causal link with specific traits is underlined and additional evidence is given for further investigations.

4.1 Introduction

Obesity and its related pathologies such as insulin resistance, type 2 diabetes, and fatty liver are symptoms of an imbalanced energy homeostasis (Blüher, 2019). A sedentary lifestyle as well as the (over)consumption of easily available energy-dense food contribute to this imbalance (Blüher, 2019). However, genetic constitution sets the stage for the phenotypic characteristics. Genome-wide association studies (GWAS) in humans have revealed more than 300 single-nucleotide polymorphisms (SNPs) associated with obesity-related phenotypes (Goodarzi, 2018). Nevertheless, identified loci and underlying genes explain only a minor proportion of the estimated heritability (Goodarzi, 2018) (Albuquerque, Nóbrega, Manco, & Padez, 2017).

Mice are ideal model organisms for studying genetic effects because the environmental conditions can be tightly controlled. Furthermore, different inbred mouse lines with distinct but well-defined genetic constitution are available, which can be used to improve our understanding of the genetic architecture of complex phenotypes like obesity. Crosses between inbred lines allow the generation of structured populations. These prerequisites make association studies between genetic loci and phenotypes feasible in relatively small populations with high statistical power (Darvasi & Soller, 1995). In particular, advanced intercross lines (AIL) allow high resolution QTL mapping by increasing recombination between any two loci (Darvasi & Soller, 1995) (Rockman & Kruglyak, 2008) (Arends et al., 2016) (Delpero et al., 2021).

The Berlin Fat Mouse with its different inbred lines (BFMI) were generated initially by crossing various mice from different pet shops, subsequent selection for high body weight and high fat mass (Hantschel, Wagener, Neuschl, Teupser, & Brockmann, 2011), and finally repeated inbreeding between pairs of full sibs to generate a mouse model for the investigation of body weight gain and body composition. The different BFMI sub-lines are genetically closely related. They are all obese, but show different features of the metabolic syndrome (Heise et al., 2016) (Wagener et al., 2006).

Due to their unique genetic background in combination with the distinct obese phenotypes, the BFMI sub-lines allow the identification of diverse genetic contributors to the metabolic syndrome. In a cross between the obese line BFMI860-12 and C57BL/6N (B6N) as a lean strain, a major QTL for total fat mass was mapped on chromosome (Chr) 3 (*jObes1*) (Neuschl et al., 2010b). The locus was further fine-mapped and characterized, which led to the identification of *Bbs7* as causal gene (Arends et al., 2016), a gene that contributes to the Bardet-Biedl syndrome in humans. In an AIL between BFMI861-S1 and BFMI861-S2, two genetically very similar lines, we identified QTL and candidate genes responsible for differences in liver weight, liver triglycerides, gonadal adipose tissue weight, and body weight (Delpero et al., 2021). The BFMI861-S1 line of this cross does not only carry the mutant *jObes1* allele, it also shows the highest liver weight and liver triglyceride (TG) concentration among all BFMI sub-lines (Heise et al., 2016). Therefore, the BFMI861-S1 line is an interesting mouse model to study the genetic architecture of hepatic fat deposition in the context of obesity. In this study, we have generated the advanced intercross of BFMI861-S1 and B6N, where B6N is a lean counterpart to BFMI861-S1 to identify additional genes contributing to the specific phenotype of the BFMI861-S1 sub-line, in particular its fatty liver.

4.2 Material and Methods

Mouse population

58 male and 65 female mice of the AIL BFMI861-S1xC57BL/6NCrl (AIL (BFMI861-S1xB6N)) in generation 10 were genotyped and phenotyped. The AIL population was generated from an F2 population between an obese BFMI861-S1 male and lean B6N females. Beginning in generation F1, individuals were randomly mated to mice from the same generation using the program RandoMate (Schmitt et al., 2009).

Animal husbandry

All animal experiments were approved by the German Animal Welfare Authorities (approval no. G0099/16). Mice were maintained under conventional conditions and a 12:12 h light–dark cycle (lights on at 06:00) at a temperature of 22 ± 2°C. Animals had *ad libitum* access to food and water. Animals were fed a standard rodent diet containing 16.7 MJ/kg of metabolizable energy, 11 % from fat, 26 % from protein and 53 % from carbohydrates (V1534-000, ssniff EF R/M; Ssniff Spezialdiäten GmbH, Soest, Germany).

Phenotyping

Animals were analyzed between the age of 4 (after weaning at 3 weeks) and 20 weeks with body weight being recorded weekly. To investigate glucose metabolism, an oral glucose tolerance test (GTT) (week 18) and an intraperitoneal insulin tolerance test (ITT) (week 20, 1 U insulin/kg body weight) were performed as previously described (Hesse et al., 2012). The area under the curve (AUC) for blood glucose for GTT and ITT was calculated. At 20 weeks, the mice were anesthetized with isofluorane after a fasting period of 2 hours and sacrificed by decapitation (in the morning until 12 AM). Several tissues including gonadal adipose tissue (gonAT), subcutaneous inguinal adipose tissue (scAT), and liver were

dissected, weighed, shock-frozen in liquid nitrogen and stored until further use at -80°C. Body length was measured (Hesse, Dunn, Heldmaier, Klingenspor, & Rozman, 2010) and the body mass index (BMI) was calculated using the DuBois equation (Gargiulo et al., 2014). Liver TG were assessed as previously described (Hesse et al., 2014). Plasma free fatty acids (FFA), cholesterol, and TG were measured as described in Schulz et al. 2011 (Schulz et al., 2011). Plasma insulin and skeletal muscle fat % were measured as previously described (Heise et al., 2016).

Genotyping

DNA isolation was done by salt extraction (5 M NaCl with ß-mercapthoethanol and proteinase K) and subsequent ethanol precipitation. Genotypes of all 123 mice were generated by Neogen GeneSeek (Lincoln, NE, USA) using the Mini Mouse Universal Genotyping Array (MiniMUGA; Illumina, San Diego, CA, USA). The MiniMUGA array contains probes targeting 10,171 known SNPs (markers) (Sigmon et al., 2020). Markers were removed when all genotypes were missing or when the marker was not segregating. In addition, to prevent spurious associations, we required that at least two of the genotype groups contained 10 observations each. In case one out of three genotype groups contained less than 10 individuals, these were set to N/A, but the marker was kept. After quality control 1,886 high quality markers were available and used for subsequent QTL analysis (Figure 4.1).

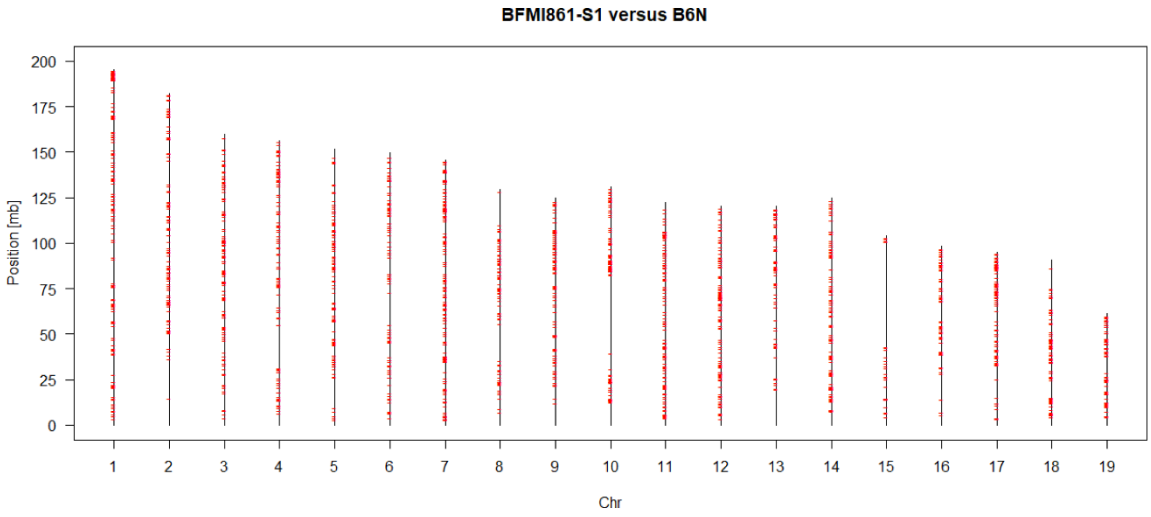


Figure 4.1. Distribution of informative SNPs across the genome in the AIL (BFMI861-S1xB6N).

QTL mapping

Linear models were used to investigate the influence of subfamily, litter size, and sex on each phenotype. Because the factors affected the phenotypes differently, different statistical models were used for mapping each phenotype, which included significant factors as fixed covariates accordingly.

To minimize the influence of population structure, the genomic inflation factor (λ) was computed. If the genomic inflation factor was above 1.05 (Devlin & Roeder, 1999), results were corrected using λ-correction. To account for multiple testing, significance thresholds were corrected using stringent Bonferroni correction. The number of independent SNPs was determined using the simpleM method (Gao, Becker, Becker, Starmer, & Province, 2010). The threshold for significance was set using the number of independent SNPs (1,365) as the total number of tests performed. This resulted in a LOD score (as defined by -log10(pvalue)) after λ-correction above 5.1 to be deemed 'genome-wide highly significant' and above 4.4 to be 'genome-wide significant'. QTL regions were defined by a 1.5 LOD drop from the top marker. Region start and end positions are defined by the first marker upstream and downstream, respectively, that have a drop of 1.5 from the LOD score of the top marker

(Dupuis & Siegmund, 1999).

To discover additional QTLs for body weight, which might be hidden by the known strong effect QTL of the *jObes1* locus, a variation of multiple QTL mapping (MQM) was used (Arends, Prins, Jansen, & Broman, 2010). The single QTL model was adjusted to compensate for the known effect of the *jObes1* locus by including the top marker from the Chr 3 region (SNP gUNC5036315) as an additional cofactor into the model:

Body weight = sex + mother + gUNC5036315 + marker genotype + error

Gene expression analysis

Gene expression was measured in RNA isolated from liver of BFMI861-S1 male mice (n=6) at 10 weeks. RNA was extracted as described in Hesse et al. 2018 (Hesse et al., 2018). Gene expression was measured with the Clariom™ S Assay for mouse (Thermo Fisher Scientific) using the service of ATLAS Biolabs, Berlin, Germany. Gene expression data of male B6N mice (n = 5) measured with the Clariom™ S assay for mouse were downloaded from Gene Expression Omnibus (Clough & Barrett, 2016). Probe intensities were $\log_2$ transformed and quantile normalized. To test for expression differences between BFMI861-S1 and B6N mice, two-tailed t-tests were performed. False positives due to multiple testing were minimized using a Benjamini-Hochberg correction. For statistical analysis and for graphical presentation R: A Language and Environment for Statistical Computing (R Team, 2018) was used.

Whole genome sequencing

The BFMI861-S1 parental genome was paired-end sequenced using the Illumina HiSeq (Illumina) platform. Obtained DNA reads were trimmed using trimmomatic (Bolger, Lohse, & Usadel, 2014) after which trimmed reads were aligned to the mouse genome (MM10, GRCm38.p6) using the Burrows–Wheeler Aligner (BWA) software (Heng Li & Durbin, 2009). Subsequently, SAM files were converted to BAM files, sorted, and indexed using Samtools

(Heng Li et al., 2009). (Optical) Duplicate reads were removed using Picard tools v2.19.0 after which indel realignment and base recalibration was done using the GATK v4.1.0.0 (McKenna et al., 2010), according to GATK best practices.

All sequence variants in BFMI861-S1 mice were called using BCFtools (Heng Li et al., 2009). Variants were further annotated using the Ensembl Variant Effect Predictor (VEP) (McLaren et al., 2016). DNA sequencing data were deposited at the NCBI Sequence Read Archive (SRA) under BioProject ID: PRJNA717237.

Candidate genes prioritization

Prioritization of candidate genes in each QTL region was performed as described in Delpero et al. 2021 (Delpero et al., 2021). In brief, genes in a QTL region containing sequence variants between the parental lines BFMI861-S1 and B6N were ranked according to the sum of scores for the functional annotation of coding and non-coding variants, gene expression data, and the Kyoto Encyclopedia of Genes and Genomes (KEGG). Coding sequence variants leading to stop gain/stop loss codons and missense mutations located in functional protein domains were awarded a score of 3 points. A missense variant with either a deleterious or a tolerated SIFT (Sorting Intolerant From Tolerant) value obtained a score of 3 or 1, respectively. Non-coding variants were scored based on their location in potential functional sites. If a non-coding variant was located in the promoter or in a splice site, a score of 3 was awarded; if located in untranslated regions (UTRs), enhancers, or CTCF binding sites (involved in 3D structure of chromatin) the score was 1. Genes differentially expressed in the liver were scored with 2; genes annotated in relevant KEGG metabolic pathways with 1.

4.3 Results

Phenotypic variation and correlation analysis in the AIL

Animals of the AIL population showed high standard deviation for the collected phenotypes which were expected and are needed for QTL analysis. In detail, body weight was on average 21.9 ± 7.27 g at the end of the experiment (week 20). GonAT weight, scAT, and liver weight were on average 1.53 ± 1.0 g, 0.65 ± 0.32 g, and 1.92 ± 0.55 g, respectively. The areas under the curve for blood glucose during GTT and ITT were 21,002 ± 12,415 and 7,308 ± 3,195, respectively. Liver triglycerides and plasma triglycerides were on average 112 ± 68 µg TG / µg protein and 896 ± 506 ug/ml, respectively. Additional plasma parameters such as plasma cholesterol, plasma FFA, and plasma insulin were on average 44 ± 9 mg/dl, 0.23 ± 0.06 mmol/l, and 7 ± 13 ng/mL, respectively. Skeletal muscle fat % also showed high standard deviation in the AIL population (mean = 19 ± 5 %) (Table 4.1).

In order to assess the relationship among the phenotypes measured in all AIL mice, Spearman correlation was computed between all the collected phenotypes. Most of the phenotypes (scAT weight, gonAT weight, liver weight, body weight, GTT AUC, and ITT AUC, plasma cholesterol, plasma insulin, BMI, body length, and skeletal muscle fat %) were positively correlated among each other (Table 4.2).

No significant correlation was detected between gonAT weight and liver weight, gonAT weight and plasma TG, and scAT weight and plasma TG. Liver TG showed a significant positive correlation with body weight at 20 weeks (r = 0.34), gonAT weight (r = 0.32), scAT weight (r = 0.44), and skeletal muscle fat % (r = 0.45). In addition, plasma FFA did not show any correlation with the collected phenotypes.

Table 4.1. Mean and SD for the collected traits in the AIL (BFMI861-S1xB6N).

	Mean (SD) AIL
Body weight [g]	21.9 ± 7.27
GonAT weight [g]	1.53 ± 1.0
ScAT weight [g]	0.65 ± 0.32
Liver weight [g]	1.92 ± 0.55
GTT auc	21,002 ± 12,415
ITT auc	7,308 ± 3,195
Liver triglycerides µg TG / µg protein	112 ± 68
Plasma triglycerides µg TG / µg protein	896 ± 506
Plasma cholesterol mg/dl	44 ± 9
Plasma FFA mmol/l	0.23 ± 0.06
Plasma insulin ng/mL	7 ± 13
Skeletal muscle fat %	19 ± 5

Abbreviations: GonAT, gonadal adipose tissue; ScAT, subcutaneous adipose tissue; TG, triglycerides; ITT, insulin tolerance test; GTT, glucose tolerance test; AUC, area under the curve; FFA, free fatty acids; BMI, body mass index; SMuscle fat %, skeletal muscle fat percentage.

Table 4.2. Spearman correlation coefficients between the collected phenotypes in the AIL (BFMI861-S1xB6N). Bold indicates significant correlation after multiple testing correction ($p < 9.10E-04$).

	Liver weight	GonAT weight	ScAT weight	Liver TG/Proteins	GTT AUC	ITT AUC	Plasma TG	Plasma FFA	Plasma cholesterol	Plasma insulin	BMI	Body length	SMuscle fat %
Body weight (20 weeks)	**0.74**	**0.64**	**0.81**	**0.34**	**0.78**	**0.84**	**0.34**	0.08	**0.66**	**0.42**	**0.98**	**0.63**	**0.63**
Liver weight		0.21	**0.43**	0.15	**0.61**	**0.66**	**0.56**	0.16	**0.55**	0.2	**0.68**	**0.62**	0.33
GonAT weight			**0.77**	**0.32**	**0.49**	**0.52**	-0.07	0.09	**0.36**	**0.35**	**0.66**	0.23	**0.74**
ScAT weight				**0.44**	**0.62**	**0.7**	0.12	0.02	**0.51**	**0.48**	**0.81**	**0.4**	**0.81**
Liver TG/Proteins					0.25	0.33	-0.29	-0.03	0.26	0.27	**0.35**	0.08	**0.45**
GTT AUC						**0.74**	**0.39**	0.02	**0.49**	**0.39**	**0.79**	**0.37**	**0.56**
ITT AUC							**0.33**	0.03	**0.62**	**0.39**	**0.84**	**0.44**	**0.59**
Plasma TG								0.28	**0.34**	0.25	0.3	**0.38**	0.05
Plasma FFA									0.25	0.01	0.03	0.14	-0.01
Plasma cholesterol										0.22	**0.65**	**0.45**	**0.39**
Plasma insulin											**0.43**	0.13	**0.48**
BMI												**0.48**	**0.65**
Body length													0.22

Abbreviations: gonAT, gonadal adipose tissue; scAT, subcutaneous adipose tissue; TG, triglycerides; ITT, insulin tolerance test; GTT, glucose tolerance test; AUC, area under the curve; FFA, free fatty acids; BMI, body mass index; SMuscle fat %, skeletal muscle fat percentage.

QTL mapping

For QTL analysis, different statistical models were used for mapping each phenotype (Table 4.3). The results revealed genome-wide significant loci on three different chromosomes (1, 3, and 6) associated with one or more of the investigated phenotypes (Table 4.4). Additionally, a suggestive QTL associated with liver TG was found on Chr 8.

In detail, the significant QTL on Chr 1 (157,132,066 - 168,495,457) with a LOD score of 4.96 was associated with liver weight (Figure 4.2a). This region contains 89 annotated protein-coding positional candidate genes. The most significant SNP for liver weight in this region was "gUNC2036998" (Chr1:158,663,689). Interestingly, this SNP showed only two genotype classes (homozygous BFMI861-S1 and heterozygous). The liver of homozygous mice carrying the BFMI861-S1 allele was 17 % heavier compared to the liver of heterozygous (Het) mice (mean BFMI861-S1 = 1.81 ± 0.25 g, mean Het = 1.55 ± 0.42 g) (Figure 4.2b).

The highly significant region for body weight on Chr 3 (34,066,622 - 40,043,158) corresponded with the *jObes1* locus that was identified in BFMI mice before (Arends et al., 2016). This QTL effect in the AIL BFMI861-S1xB6N persisted at all time points starting from week 9 until week 20 (Figure 4.3a). The most significant association (LOD = 8.89) was body weight at week 14 with the top marker gUNC5036315 (Chr3:35,986,311) (Figure 4.3b). This marker was 604 kbp away from the *Bbs7* gene that had been identified recently as causal gene for obesity in BFMI mice (Arends et al., 2016). At the top marker locus, 14 weeks-old mice homozygous for the BFMI861-S1 allele were 10.09 g heavier than homozygous B6N counterparts. The same region affected also scAT weight (LOD = 5.8), and BMI (LOD = 4.94) with homozygous BFMI861-S1 mice carrying 98 % more scAT than B6N homozygous mice (mean BFMI861-S1 = 0.95 ± 0.24 g, mean B6N = 0.48 ± 0.21 g) and 83 % compared to heterozygous mice (mean Het = 0.52 ± 0.27 g). In addition, homozygous BFMI861-S1 mice showed 14 % BMI increase compared to B6N homozygous mice (mean BFMI861-S1 = 4.20 ± 0.20 kg/m2, mean B6N = 3.67 ± 0.31 kg/m2) and 12 % compared to heterozygous mice (mean Het = 3.76 ± 0.33 kg/m2). This region contains 30 annotated protein-coding genes.

When correcting for the top marker of the *jObes1* locus on Chr 3 (gUNC5036315), an

additional region associated with body weight at 16 weeks was detected on Chr 6 (0 - 17,553,096) (Figure 4.2c). This region contains 59 protein-coding genes. The most significant SNP of this region was gUNC10595065 (3,919,413; LOD = 5.41). Heterozygous mice showed 9 % increase in body weight compared to homozygous B6N mice (mean Het = 45.14 ± 2.91 g, mean B6N = 41.47 ± 3.24 g) and 4.5 % increase compared to homozygous BFMI861-S1 mice (mean BFMI861-S1 = 43.33 ± 2.7 g (Figure 4.2d).

A suggestive QTL for liver TG was identified on Chr 8 (86,158,420-106,738,488). Due to the suggestive significance, the region is large containing 179 protein-coding genes. The top marker in this region was "S1H083826428" (Chr8:95,660,710; LOD = 3.93). On average, homozygous mice carrying the BFMI861-S1 allele at this marker showed 90 % increase amounts of liver TG compared to homozygous B6N mice (mean BFMI861-S1 = 192 ± 54 µg TG / µg protein, mean B6N = 101 ± 53 µg TG / µg protein) and 83 % increase compared to heterozygous mice (mean Het = 105 ± 61 µg TG / µg protein).

Table 4.3. P values for effects of covariates on each phenotype.

	Covariates		
	Sex	Subfamily	Litter size
Body weight (20 weeks)	**0.00001**	**0.00097**	0.17152
GTT AUC	**2.01E-06**	**0.01000**	0.05274
ITT AUC	**0.00025**	**0.00022**	0.12422
GonAT weight	**0.00087**	**0.00029**	**0.01736**
ScAT weight	0.55182	**3.35E-06**	0.11907
Liver weight	**1.07E-09**	**0.00077**	0.19891
Liver TG	0.28798	0.51456	0.48629
Plasma cholesterol	**0.00510**	**0.01481**	0.59906
Plasma FFA	**0.00543**	0.20579	0.50047
Plasma TG	**3.24E-10**	0.11351	0.98701
Plasma insulin	0.52115	**0.00252**	0.54256
Body length	**0.00012**	0.10646	0.08461
BMI	**0.00006**	**0.00062**	0.23395
SMuscle fat %	**0.03634**	**0.00003**	**0.03419**

Abbreviations: P values for effects of covariates on each phenotype. In bold are represented signifcant covariates that were included in the model for each trait. gonAT, gonadal adipose tissue; scAT, subcutaneous adipose tissue; Gluc, blood glucose concentration; ITT, insulin tolerance test; GTT, glucose tolerance test; AUC, area under the curve; FFA, free fatty acids; TG, triglycerides; BMI, body mass index; SMuscle fat %, skeletal muscle fat percentage.

Table 4.4. QTL identified for different phenotypes in the AIL (BFMI861-S1xB6N).

Phenotype	Age [weeks]	N	QTL region				LOD(BH)	% AIL Var	Mean BFMI/BFMI	Mean B6/BFMI	Mean B6/B6
			Chr	Start [bp]	Top [bp]	Stop [bp]					
Body weight [g]	9	123	3	34,066,622	35,986,311	40,043,158	7.45	8.1	34.21	29.08	28.02
	10	123	3	34,066,622	35,986,311	40,043,158	5.69	9.5	36.27	29.98	29.33
	11	123	3	34,066,622	38,187,507	40,043,158	7.24	9.6	38.04	30.75	29.85
	12	123	3	34,066,622	35,986,311	40,043,158	5.92	9.2	39.54	31.68	30.51
	13	123	3	34,066,622	35,986,311	40,043,158	7.08	11.2	40.93	32.57	31.86
	14	123	3	34,066,622	35,986,311	40,043,158	8.89	13	42.46	33.4	32.37
	15	123	3	34,066,622	38,187,507	40,043,158	7.22	10.4	42.93	33.86	33.2
	16	123	3	34,066,622	35,986,311	40,043,158	7.74	12.7	44.34	34.57	33.53
	16	123	6	0	13,919,413	17,553,096	5.41	8	43.33	45.14	41.47
	17	123	3	34,066,622	38,187,507	40,043,158	4.97	8.6	44.98	36.07	35.15
	18	123	3	34,066,622	38,187,507	40,043,158	5.24	9.1	45.6	36.21	35.81
	19	123	3	34,066,622	35,986,311	40,043,158	5.63	9.1	45.05	36.13	34.78
	20	123	3	34,066,622	35,986,311	40,043,158	5.07	9.8	46.06	37.54	35.91
ScAT weight [g]	20	121	3	34,066,622	38,187,507	40,043,158	5.8	9.1	0.95	0.52	0.48
BMI [kg/m2]	20	122	3	34,066,622	38,187,507	40,043,158	4.94	10.4	4.2	3.76	3.67
Liver weight [g]	20	123	1	157,132,066	158,663,689	168,495,457	4.96	5.6	1.81	1.55	-

Abbreviations: ScAT, subcutaneous adipose tissue; BMI, body mass index; QTL, quantitative phenotypes locus; Chr, chromosome number; Start, Top, and Stop, position of the start of the QTL confidence interval, position of the SNP with the highest LOD score, and position of the end of the QTL confidence interval in base pairs, respectively. Positions are given according to the Mouse Genome Version MM10, GRCm38.p6. SNP, single-nucleotide polymorphism. The confidence interval gives the 1.5 LOD drop region of the top SNP position. A LOD score above 5.1 was deemed to be 'genome-wide highly significant' and above 4.4 was deemed 'genome-wide significant'. BH, Bonferroni correction; LOD, logarithm (base 10) of odds; Var %, percentage of total variance in AIL explained. The Mean columns show the phenotypic mean adjusted for significant covariates of homozygous BFMI861-S1, heterozygous, and homozygous B6N animals, respectively.

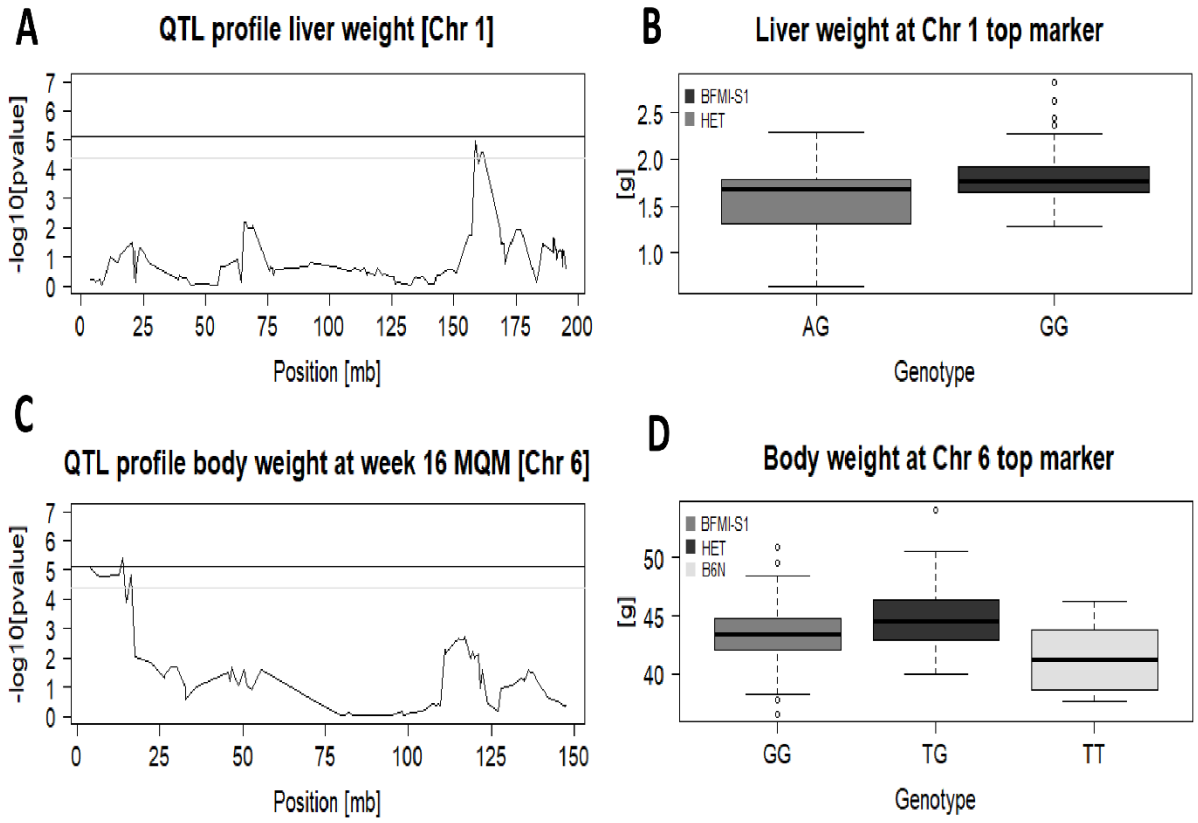


Figure 4.2. (A) QTL mapping curve of the locus on chromosome 1 for liver weight. The black (1%) and grey (5%) horizontal lines mark the significance thresholds; likelihood ratios above the black line are formally highly significant (LOD > 5.1), likelihood ratios above the grey line are formally significant (LOD > 4.4). (B) Boxplots for two genotype classes (BFMI-S1, BFMI861-S1 homozygous; HET, heterozygous) at SNP gUNC2036998 which is located at the top position for liver weight. (C) QTL mapping curve on chromosome 6 for body weight at week 16 after performing MQM. (D) Boxplots for all three genotype classes (BFMI-S1, BFMI861-S1 homozygous; HET, heterozygous; B6N, C57BL/6N homozygous) at SNP gUNC10595065 which is located at the top position for body weight at week 16 after performing MQM.

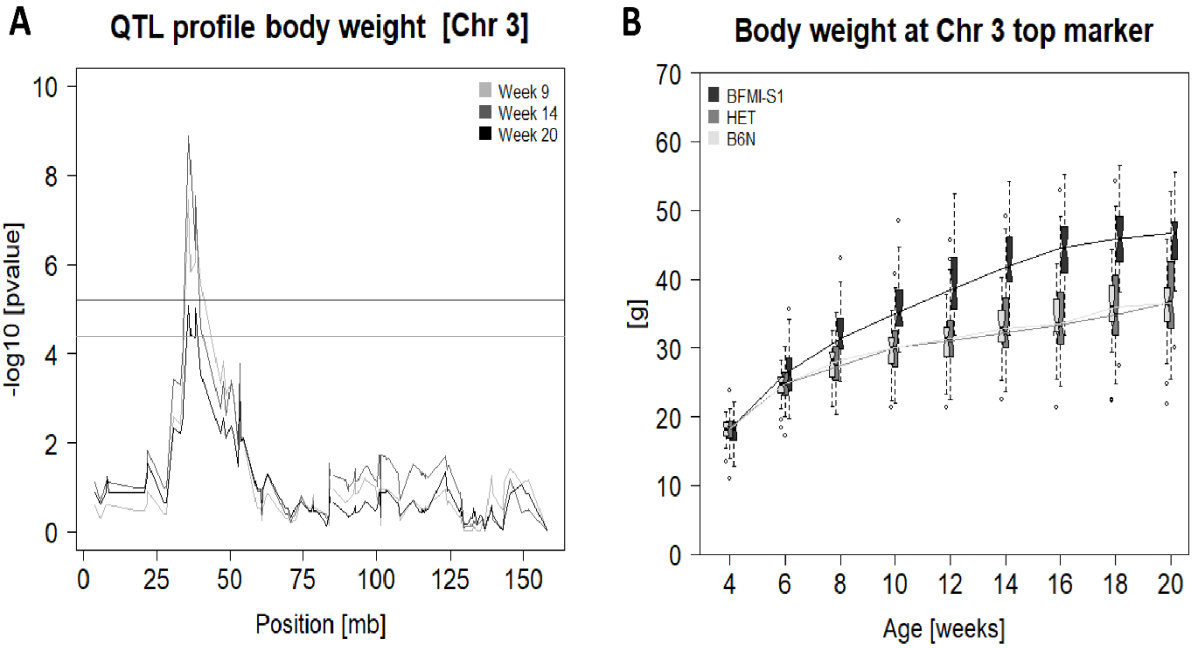


Figure 4.3. (A) QTL mapping curve of the *jObes1* locus on chromosome 3 for total body weight at week 9, 14, and 20. The black (1%) and grey (5%) horizontal lines mark the significance thresholds. (B) Boxplots for mice aged 4-20 weeks and curves depicting body weight development. For every time point, boxplots for all three genotype classes (BFMI-S1, BFMI861-S1 homozygous; HET, heterozygous; B6N, C57BL/6N homozygous) are shown at SNP gUNC5036315, which is located at the top position of the *jObes1* region.

Candidate gene prioritization

Within the confidence intervals of the four QTL (including the QTL on Chr 8 suggestively associated with liver TG) 357 protein coding positional candidate genes were located. 152 genes were polymorphic between BFMI861-S1 and B6N in protein-coding and/or regulatory regions; 29 on Chr 1, 38 on Chr 3, 22 on Chr 6, and 63 on Chr 8. To identify the most likely candidate genes for each QTL, the 152 polymorphic positional candidate genes were scored according to the decision tree. Tissue expression of top candidate genes was investigated using MGI (27). After applying the prioritization criteria, two genes (*Astn1* and *Sec16b*)

located in the region on Chr 1 associated with liver weight ranked with the highest score of 12 and 10, respectively (Table 4.5). *Astn1* and *Sec16b* carry deleterious missense variants according to the variant effect predictor and are widely expressed including the liver. In addition, both *Sec16b* and *Astn1* show variants in the promoter region and 5-prime and 3-prime UTRs. However, despite variants in regulatory regions *Sec16b* and *Astn1* did not show gene expression differences in the liver. Tree genes on Chr 3 (*Frem2*, *Bbs7*, and *Noct*), with a score of 13, 12, and 12, respectively ranked as top candidate genes. Among the candidate genes located in the region on Chr 3 (*jObes1*) associated with body weight from week 9 to 20, *Bbs7*, *Noct*, and *Frem2* all carried missense variants in domains and regulatory region variants and are all widely expressed across tissues including the liver and the nervous system. In addition, *Bbs7* and *Noct* were both downregulated (p=0.01346 and p=0.00876, respectively) in liver of BFMI861-S1 mice compare to B6N, while *Frem2* did not show differences in the expression. In the region on Chr 6 associated with body weight two genes (*Met* and *Ica1*) with scores of 10 and 9, respectively ranked as top candidate genes. *Met* carries a tolerated missense variant in the IPT (Ig-like, plexins, transcription factors) domain and variants in regulatory regions such as enhancers and untranslated regions. *Ica1* showed variants only in regulatory regions (promoter, CTCF binds, enhancers, and untranslated regions). According to Mouse Genome Informatics both genes are highly expressed in a large variety of tissues. *Met* and *Ica1* were both downregulated in the liver of BFMI861-S1 mice (p=8.2E−06 and 0.00200, respectively). In the region affecting liver TG on Chr 8, *Fto* and *Lpcat2* were identified as the top candidate genes with scores of 16 and 13, respectively. *Fto* carried a stop gain variant and additional variants in different regulatory regions (promoter, CTCF binds, and enhancer) and was downregulated in the liver of BFMI861-S1 mice (p=0.01092).

Lpcat2 instead carried one deleterious missense variant and regulatory region variants (promoter and enhancer) but did not show expression differences in the liver between BFMI861-S1 and B6N mice. According to the Mouse Genome Informatics database both *Fto* and *Lpcat2* are widely expressed across tissues including the liver.

Table 4.5. Top candidate genes after applying the prioritization criteria. Bold indicates differentially expressed in liver after Benjamini-Hochberg correction.

Gene score	Phenotype	Chr	Positional candidate gene	Type of mutation	P-value liver	FC liver (S1/B6N)
12	Liver weight	1	*Astn1*	Deleterious domain missense, CTCF binds, UTRs, enhancer and promoter variant	0.66493	-0.03
10		1	*Sec16b*	Deleterious missense, CTCF binds, UTRs, enhancer and promoter variant	0.54953	-0.01
13	Body weight, scAT weight, BMI	3	*Frem2*	Deleterious domain missense, CTCF binds, UTRs, enhancer and promoter variant	0.02092	0.12
12		3	*Bbs7*	Tolerated domain missense, CTCF binds, UTRs, enhancer and promoter variant	**0.01346**	-0.14
12		3	*Noct*	Tolerated domain missense, CTCF binds, UTRs, and promoter variant	**0.00876**	-0.17
10	Body weight	6	*Met*	Tolerated domain missense, CTCF binds, UTRs, and enhancer variant	**8.20E-06**	-0.1
9		6	*Ica1*	CTCF binds, UTRs, enhancer and promoter variant	**0.002**	-0.15
16	Liver TG	8	*Fto*	Tolerated domain missense, stop gained, CTCF binds, UTRs, enhancer and promoter variant	**0.01092**	-0.09
13		8	*Lpcat2*	Deleterious domain missense, CTCF binds, UTRs, enhancer and promoter variant	0.22981	-0.05

Abbreviations: scAT, subcutaneous adipose tissue; TG, triglycerides; BMI, body mass index; Chr, chromosome; FC, fold change.

4.4 Discussion and conclusion

Genome wide association studies (GWAS) on obesity-associated phenotypes identified loci that account in sum only for a small percentage of the total variance of the examined population (Goodarzi, 2018). Therefore, studies are needed that better allow the identification of genetic effects than most populations do.

In order to unravel the genetics behind obesity and hepatic fat deposition, we investigated an AIL population generated from a cross between the obese BFMI861-S1 mouse line and the lean reference strain B6N.

The AIL population used in this study has the advantage of having a high resolution for QTL mapping. Because the examined AIL accumulated recombination over 10 generations, the physical length of the QTL regions is relatively short and as such the number of positional candidate genes is low (Darvasi & Soller, 1995). In our population, the number of positional candidate genes could be further reduced by removing regions in the genome that are identical between BFMI861-S1 and B6N. Excluding non-polymorphic genes reduced the number of protein coding candidate genes from 357 to 152. In addition, the application of the decision tree led to the prioritization of the most likely candidate genes among the 152 polymorphic genes.

In the region on Chr 1 associated with liver weight, the top candidate genes are *Sec16b* and *Astn1*. *Sec16b* is required for secretory cargo traffic from the endoplasmic reticulum to the Golgi apparatus (Bhattacharyya & Glick, 2007). Previously, the gene has been linked to increased fat storage in both mice and humans. A human GWAS associated *Sec16b* with differences in body composition (Sahibdeen et al., 2018). In mice, dysfunctional *Sec16b* was associated with increased body weight (Bult et al., 2019). The BFMI861-S1 mice of our study carry a deleterious missense variant of *Sec16b* leading to an impaired protein variant and, in addition, SNPs in the promoter region and 3-prime and 5-prime UTRs. Since *Sec16b* was not differentially expressed between the parental lines BFMI861-S1 and B6N, we hypothesize that the detected deleterious missense variant is responsible for the dysfunction of the encoded protein. Therefore, we consider *Sec16b* as a very strong candidate responsible for

the increased liver weight in BFMI861-S1 mice. *Astn1,* the second prioritized candidate gene in the Chr 1 region, is a neuronal adhesion molecule required for the migration of young postmitotic neuroblasts along glial fibers (Zheng, Heintz, & Hatten, 1996). The gene has not been associated with obesity related phenotypes yet. In BFMI861-S1 mice, *Astn1* carries one deleterious missense variant in the fibronectin type 3 domain. This domain is known to be responsible for interactions with other extracellular matrix (ECM) or cell surface proteins (Kim, McKnite, Xie, & Christian, 2018). Therefore, the identified deleterious mutation could reduce the interaction ability of the encoded protein. Additional SNPs between the BFMI861-S1 and B6N in the promoter, 5-prime UTR, enhancers, and CTCF binding sites of *Astn1* could be responsible for the observed downregulation of the gene in the liver of BFMI861-S1 mice versus B6N. This finding confirms that the selected prioritization approach is useful to identify so far unknown candidate genes for the phenotype under investigation which should be considered for follow-up studies.

In the QTL interval for body weight on Chr 3, the prioritization approach identified *Frem2*, *Bbs7*, and *Noct* as the top candidate genes. *Bbs7* has been identify to be the causal gene for elevated fat mass and obesity in all BFMI lines before (Arends et al., 2016). Among diverse sequence variants between BFMI and B6N in the *Bbs7* gene, it has been clarified that a large intronic deletion is partially responsible for the high fat content in BFMI mice (Krause et al., 2021) This is additional evidence for the prioritization approach being suitable for the correct identification of positional candidate genes in a defined confidence interval.

Only if we accounted for the strong effect of *jObes1* by including the top marker from the Chr 3 region (gUNC5036315) as a cofactor into the model we detected another QTL for body weight on Chr 6. This QTL affecting body weight at 16 weeks partially overlaps with a previously identified QTL for body weight at 10 weeks that was identified in the F2 population BFMI860-12xB6N, which used BFMI860-12 instead of BFMI861-S1 from the current study (Neuschl et al., 2010b). In the confidence interval of the QTL identified in our cross, *Met* and *Ica1* are the highest scored top candidate genes. *Met* encodes a receptor tyrosine kinase involved in the transduction of signals from the extracellular matrix to the cytoplasm

by binding to hepatocyte growth factor/HGF ligand (Bladt, Riethmacher, Isenmann, Aguzzi, & Birchmeier, 1995). In BFMI861-S1 mice *Met* carries a tolerated missense variant in the Ig-like, plexins, transcription factors domain which is involved in the control of cell dissociation, motility, and invasion of extracellular matrices (Collesi, Santoro, Gaudino, & Comoglio, 1996). In addition, *Met* carries SNPs in regulatory regions (enhancers and CTCF binds) which could contribute to the observed downregulation in the liver of BFMI861-S1 mice compare to B6N. According to literature, *Met* is known to be involved in pancreatic-cell death and diabetes (Bult et al., 2019) (Mellado-Gil et al., 2011). Therefore, we consider this gene to be an interesting candidate for the increased body weight of heterozygous mice compared to homozygous BFMI861-S1 and homozygous B6N mice. *Ica1* which is the second prioritized gene of the Chr 6 region encodes for Islet cell autoantigen 1 and plays a role in neurotransmitter secretion (Pilon, Peng, Spence, Plasterk, & Dosch, 2000). *Ica1* carries several SNPs in the promoter and 5-prime UTR which might cause the downregulation in liver of BFMI861-S1 mice. *Ica1* is also known to be associated with type 1 diabetes mellitus in non-obese diabetic mice (Bonner et al., 2012) as well as with glucose homeostasis (Bult et al., 2019). Therefore, *Ica1* could cause the higher body weight that we observed in heterozygous mice compared to homozygous BFMI861-S1 and homozygous B6N mice.

In the confidence interval of the suggestive liver TG QTL on Chr 8, *Fto* and *Lpcat2* are the most likely candidate genes responsible for hepatic fat accumulation in BFMI861-S1 mice. The same genomic region has previously been associated with liver TG in the Collaborative Cross (Abu-Toamih Atamni, Botzman, Mott, Gat-Viks, & Iraqi, 2018). In this QTL, we suggested *Fto* to be responsible for the increased amount of TG in the liver of BFMI861-S1 mice. *Fto* has been associated with the body mass index in humans and therefore this gene has been examined extensively (Lan et al., 2020) (Zhao, Yang, Sun, Zhao, & Yang, 2014). *Fto* codes for an RNA demethylase. In mice, both *Fto* knockout and *Fto* missense variants are responsible for fat accumulation and hypertriglyceridemia (Church et al., 2009). The BFMI861-S1 allele carriers of our study carry a stop gain variant at FTO amino acid position 314, which leads to a premature stop codon, and thereby to a shortened protein.

Furthermore, BFMI861-S1 mice contain additional missense and promoter variants. The BFMI861-S1 *Fto* haplotype likely contributes as a whole to the observed downregulation in the liver of BFMI861-S1 mice. These findings led us to consider *Fto* as the main contributor to the hepatic fat accumulation in BFMI861-S1 mice. *Lpcat2,* the other prioritized candidate gene is an acetyltransferase (Shindou et al., 2007). In BFMI861-S1 mice *Lpcat2* carries a missense variant at amino acid position 59 located in *the* transmembrane helical domain which could affect the function of the protein. In addition, *Lpcat2* carries variants in regulatory regions but is not differentially expressed in liver of BFMI861-S1 mice compare to B6N. According to literature*, Lpcat* proteins are associated with polyunsaturated fatty acid accumulation (Harayama et al., 2014). Moreover, knockdown of both *Lpcat1* and *Lpcat2* leads to an increase in lipid droplets size (Moessinger et al., 2014). The occurrence of several SNPs in functional regions of *Lpcat2* in BFMI861-S1 mice and its known function in fat accumulation let us to consider this gene as another potential contributor to liver TG accumulation.

In conclusion, our approach identified strong candidate genes that are likely involved in the development of obesity and fatty liver disease in our Berlin Fat Mouse model. However, although we have prioritized candidate genes using the available information, we cannot completely rule out that one of the other polymorphic genes was wrongly discarded, or that non-protein coding regions might be causal. Nevertheless, the natural mutations found in this study in the Berlin Fat Mouse Inbred line BFMI861-S1 contribute to our understanding of which genes impact obesity and hepatic fat storage. These findings help to clarify and support the role of known candidates. The examined mouse model and the applied gene prioritization approach allow to unravel the effects of the identified QTL regions and to link genes with observed phenotypes. Additional studies on the candidate genes should be performed to discover by which molecular mechanisms they contribute to the development of obesity and liver associated diseases in mouse models but also in humans.

4.5 Acknowledgements

DH was funded by the Deutsche Forschungsgemeinschaft (DFG, HE8165/1-1). The project was supported by the DDG (Deutsche Diabetes Gesellschaft). We thank Marion Bütow, Ulf Kiessling, and Ines Walter for assistance in animal care and Maximilian Sprechert for technical assistance.

Chapter 5: General discussion

The overall goal of this thesis is to identify QTL and candidate genes that are responsible for the unhealthy metabolic phenotypes observed in one peculiar BFMI line (BFMI861-S1). The specific goals are: 1) Discover new QTL associated with metabolic traits in the Berlin Fat Mouse line BFMI861-S1, 2) Prioritize and detect candidate genes in the newly identified QTL, 3) Identify the causal tissue responsible for the metabolic syndrome in the BFMI861-S1 line. To detect QTL for metabolic traits responsible for the unhealthy BFMI861-S1 phenotypes one AIL was generated from the cross between the BFMI lines BFMI861-S1 and BFMI861-S2. Both BFMI861-S1 and BFMI861-S2 lines carry the BFMI obesity locus on chromosome 3. However, BFMI861-S1 shows the highest liver weight and liver triglycerides when compared to other BFMI lines while BFMI861-S2 is insulin sensitive despite being obese (Heise et al., 2016). Mating these two genetically close inbred lines we created the AIL (BFMI861-S1xBFMI861-S2). Using the AIL population, overlapping QTL for gonadal adipose tissue weight and blood glucose concentration were detected on chromosome (Chr) 3 (95.8-100.1 Mb), and for gonadal adipose tissue weight, liver weight, and blood glucose concentration on Chr 17 (9.5-26.1 Mb). Causal modelling suggested for Chr 3-QTL direct effects on adipose tissue weight, but indirect effects on blood glucose concentration. Direct effects on adipose tissue weight, liver weight and blood glucose concentration were suggested for Chr 17-QTL. Prioritized positional candidate genes for the identified QTL were *Notch2* and *Fmo5* (Chr 3) and *Plg* and *Acat2* (Chr 17). Two additional QTL were detected for gonadal adipose tissue weight on Chr 15 (67.9–74.6) and for body weight on Chr 16 (3.9-21.4 Mb). Using the AIL (BFMI861-S1xBFMI861-S2) we also performed time series analysis using body weight data.

To detect further genetic variants for obesity and liver disease, a cross between BFMI861-S1 and C57BL/6N (B6N) was generated in addition. This additional population was suitable for the identification of gene variants that contribute to obesity and liver associated phenotypes not detected in the AIL (BFMI861-S1xBFMI861-S2) due to the high genetic similarity of the parental lines (both BFMI861-S1 and BFMI861-S2 carry the obesity locus *jObes1* on Chr 3). In

this cross, three QTL associated with liver weight, body weight, and subcutaneous adipose tissue (scAT) weight were identified. A highly significant QTL on Chr 1 (157–168 Mb) showed an association with liver weight. A QTL for body weight at 20 weeks was found in the *jObes1* locus on Chr 3 (34.1 – 40 Mb) overlapping with a QTL for scAT weight. In a multiple QTL mapping approach using the top marker of the *jObes1* locus as a covariate, one additional QTL affecting body weight at 16 weeks was identified on Chr 6 (9.5-26.1 Mb). Considering sequence variants and expression differences, *Sec16b* and *Astn1* were prioritized as top positional candidate genes for the liver weight QTL on Chr 1; *Met* and *Ica1* for the body weight QTL on Chr 6.

Gene expression data collected in multiple tissues such as gonadal adipose tissue, liver, pancreatic islets, and skeletal muscle were used to detect the driver tissue for the metabolic syndrome in the BFMI861-S1 line. Applying correlation and cluster analysis on the gene expression data, the gonadal adipose tissue was suggested as causal for the metabolic syndrome in the BFMI861-S1 line.

5.1 AIL (BFMI861-S1xBFMI861-S2) vs AIL (BFMI861-S1xB6N)

According to literature, sample size is a critical factor that influences statistical power and, therefore, the power to detect QTL (Darvasi & Soller, 1995). The two generated AILs differed in size and therefore in the ability to detect hidden QTL associated with obesity and additional metabolic phenotypes. In particular, the AIL (BFMI861-S1xBFMI861-S2) contained in total 397 males while the AIL (BFMI861-S1xB6N) contained in total 123 mice (58 male and 65 female). In addition, due to the high genetic similarity of the parental lines, the AIL (BFMI861-S1x BFMI861-S2) was genotyped for 143,259 SNPs while the AIL (BFMI861-S1xB6N) was genotyped only for 10,171 known SNPs considering the high genetic differences between BFMI861-S1 and the reference B6N.

In the AIL (BFMI861-S1xBFMI861-S2), the large sample size increased the ability to detect

significant effects. In addition, individuals of this AIL showed a wide range of phenotypes despite being closely related, enhancing the power of QTL detection. The high genetic similarity of the two BFMI lines allowed better prioritization of positional candidate genes excluding monomorphic genes in the discovered QTL regions.

The AIL (BFMI861-S1xB6N), which is the second AIL used in this thesis had a smaller sample size compared to the AIL (BFMI861-S1xBFMI861-S2). Despite the small sample size even in this population we were able to detect significant effects due to large differences in phenotypes between the parental lines, especially in body weight and fat mass. However, the higher genetic differences between the parental lines BFMI861-S1 and B6N impacted the ability to exclude candidate genes in the QTL regions.

Considering differences in the genetic background, another point of divergence between the two AILs is that the BFMI861-S1 and BFMI861-S2 lines share the *jObes1* locus on chromosome 3 responsible for the high body weight of all BFMI lines. Therefore, the resulting AIL allowed to naturally correct for this major obesity locus and detect additional hidden obesity QTL. In contrast, in the AIL (BFMI861-S1xB6N) hidden obesity QTL were detected by performing multiple QTL mapping and thereby correcting for the top marker of the *jObes1* locus on Chr 3 when QTL mapping body weight. This well-established method allowed to correct for major effect and detect one hidden QTL for body weight.

5.2 QTL mapping in the AIL populations

To identify candidate gene variants that are responsible for the metabolic features of the BFMI861-S1 line, QTL mapping was performed in the two described AILs both generated using the BFMI861-S1 as one of the two parental lines. Due to the large sample size and the high phenotypic variance in the AIL (BFMI861-S1xBFMI861-S2) we identified four QTL associated with the collected metabolic phenotypes. High correlation between phenotypes leads to overlapping QTL (Y. Li et al., 2010). As expected, overlapping QTL for highly correlated traits

were identified. The main QTL identified in this population was located on Chr 17 (Gatlgq) and was associated with five collected traits (GonAT weight, liver weight, liver triglycerides, and blood glucose). For Gatlgq, the S1 allele reduces adipose tissue weight and increases the liver weight and hepatic fat content. Moreover, reduced uptake in AT and maybe other tissues increased gluconeogenesis in the liver leading to an increase in blood glucose concentration. This shows that the identified QTL on Chr 17 seems to be the main driver for ectopic fat deposition and abnormal blood glucose concentration in the S1 line. Indeed, hepatic fat storage had been observed already as responsible for high blood glucose concentration (Parker, 2018). The Gatlgq was not replicated in the AIL (BFMI861-S1xB6N). This could be for two main reasons:

1. The lower number of animals available in the AIL (BFMI861-S1xB6N) and the fact that both males and females were considered for the analysis.
2. The different diet regimes between the two AILs.

We can speculate that one reason that allowed us to identify the QTL on chromosome 17 in the AIL (BFMI861-S1xBFMI861-S2) is the high fat, high carbohydrate diet that was fed from week 22 until week 25 (end of the experiment). This diet increased the differences in metabolic traits such as adipose tissue weight and blood glucose concentration between the animals of the AIL (AILBFMI861-S1xBFMI861-S2). The AIL (BFMI861-S1xB6N) was kept under a standard diet for the entire experiment (from week 0 until week 20) and therefore differences in blood glucose and other metabolic traits associated with Gatlgq were not enhanced in this population.

Despite Gatlgq was not identified In the AIL (BFMI861-S1xB6N), additional QTL were identified in this population. Particularly, we replicated the known obesity QTL on chr 3 (*jObes1*) which was associated also with BMI and scAT weight. Furthermore, performing MQM we identified one additional obesity QTL on chromosome 6. The ability to identify major and minor QTL associated with multiple traits is an additional confirmation that the BFMI861-S1 line used with different breeding partners generates a powerful population to study genetics behind complex metabolic traits.

5.3 Prioritization of positional candidate genes

As previously introduced the most difficult challenge when performing QTL analysis is the ability to identify the correct candidate genes in a large QTL region. One QTL region can contain even hundreds of genes and the identification of the right candidate is usually not straight forward (Warwick Vesztrocy, Dessimoz, & Redestig, 2018).

To prioritize candidate genes in the QTL regions we designed a decision tree prioritization approach that takes into consideration data from different sources such as WGS, gene expression and literature data. Using these different types of information, we suggest potential candidate genes in each QTL region.

The application of a decision tree using different biological data sources, can have applications in drug discovery to prioritize the right candidate genes that are linked to diseases in a specific genomic region. Decision trees can be designed *de novo* (as the one used in this thesis) or through training of machine learning models using different datasets. However, the design and implementation of training datasets based on previous genetic knowledge is not an easy task (Vesztrocy et al., 2018).

Our prioritization approach includes enough details for a fine candidate gene prioritization by using multiple data sources such as WGS, gene expression data and literature. However, several limitations of this approach need to be considered. For example, we cannot rule out that one of the candidate genes is an unannotated gene in the identified QTL region. In addition, all the identified regions are large and include even hundreds of genes which increased chances of false positive identification. These chances can be reduced by performing QTL mapping with further generations of the current AIL. Another limitation of this approach is that we considered only protein coding genes. Doing so, we exclude e.g. non-coding RNAs that could be the correct candidate genes responsible for the associated traits. Integrating RNA sequencing data will be essential to not exclude non-coding RNAs such as microRNAs as potential candidates (Liu et al., 2017).

In the AIL (BFMI861-S1xBFMI861-S2) we identified 534 protein-coding potential candidate genes in total considering the four discovered QTL regions. After removing the genomic

regions that are identical between BFMI861-S1 and BFMI861-S2 only sixty-two genes remained as polymorphic between the two BFMI lines. Applying the prioritization approach the sixty-two polymorphic genes were scored according to sequencing, gene expression and literature data. The most promising candidates were therefore identified as the genes with the highest and second highest score in each region.

The same prioritization method was applied in the AIL (BFMI861-S1xB6N). Also in this population, with this approach we were able to identify the most likely candidates for each QTL region. To confirm the strength of the prioritization method we identified *Bbs7* as one of the top candidates in the *jObes1* locus that was identified in the AIL (BFMI861-S1xB6N).

Literature search applied after running the prioritization method confirmed that all the discovered candidate genes in both AILs are involved in metabolic pathways or had been previously linked to metabolic traits. These results indicate the ability of the applied prioritization method to discover accurately candidate genes.

With the studies that are part of this thesis we were able to identify candidate genes involved in the development of the metabolic syndrome which could help in the future to develop new treatments in humans for fatty liver disease, obesity or even type 2 diabetes. However, additional research will be needed to confirm *in vivo* and *in vitro* the identified candidate genes and to clarify their molecular mechanism and interactions. *In vivo* studies could involve direct knockout of certain genes or performing fine mapping studies with further generations of the presented AILs. *In vitro* assays allow insight into the mechanism and could include reporter gene assays or functional protein assays depending on the investigated candidate.

5.4 Correlation analysis and identification of the causal tissue

Phenotypic correlation is an important parameter to understand observational relationships between complex traits underlying complex diseases (T. Li, Ning, & Shen, 2021). Considering

the results of the correlation analysis, in the AIL (BFMI861-S1xBFMI861-S2) there is no correlation between body weight and gonadal adipose tissue weight and negative correlations between gonadal adipose tissue weight and all other traits in AIL males, while liver weight was positively correlated with all other traits. These results are totally opposite to our expectations in metabolically healthy individuals where liver weight and adipose tissue weight should show positive correlation. This indicates ectopic fat storage in the liver (O'Neill & O'Driscoll, 2015). Ectopic fat deposition in the liver is also present in individuals of the S1 line (Heise et al., 2016). When crossing the unhealthy S1 mice with healthy B6N mice, the mice of the AIL (BFMI861-S1xB6N) showed a positive correlation between body weight and gonadal adipose tissue weight together with a weak positive correlation between gonadal adipose tissue weight and liver weight. Therefore, by crossing BFMI861-S1 with B6N the dysfunction in the gonadal adipose tissue that leads to hepatic steatosis is not evident in the resulted AIL. These results make it clear that the genetic interaction of background genes is important and therefore many different strains should be used to understand the whole pattern of ectopic fat deposition.

Ectopic fat storage in the liver instead of storage in the adipose tissue as the major fat storage organ has been reported many times already to be linked with impaired glucose homeostasis (Parker, 2018) (Rosen & Spiegelman, 2006). Ectopic fat storage is a peculiarity of our BFMI861-S1 line which is even more evident in the AIL (BFMI861-S1xBFMI861-S2). Hepatic fat deposition likely is the main driver for impaired glucose homeostasis in the BFMI861-S1 line, which could be initiated by adipose tissue dysfunction. Gene expression data supported the assumption of impaired adipose tissue function being causal for ectopic fat storage in S1 mice due to the high number of differentially expressed genes identified in the gonadal adipose tissue between the BFMI861-S1 and BFMI861-S2 line. In addition, when performing hierarchical clustering analysis on the correlation matrix using gene expression data, were observed distinguished clusters of differentially expressed genes only in the gonadal adipose tissue and not in other analyzed tissues. Correlation and cluster analysis applied to gene expression data is widely used to identify patterns of differentially expressed

genes between healthy and unhealthy tissues both in mouse and humans (Kosti, Jain, Aran, Butte, & Sirota, 2016). We assume that the BFMI861-S1 line is an interesting mouse model in addition to the most common C57BL/6 (which shows a similar pattern) (Van Herck, Vonghia, & Francque, 2017) to investigate the genetic component behind adipose tissue dysfunction which in turns leads to ectopic fat deposition.

Chapter 6: Conclusions

In this study we successfully identified QTL associated with metabolic traits such as gonadal adipose tissue weight, liver weight, body weight, and blood glucose concentration using the AIL (BFMI861-S1xBFMI861-S2) and the AIL (BFMI861-S1xB6N). Applying a decision tree prioritization approach in all the QTL identified in both AILs we scored positional genes from the most likely candidate to the less likely candidate gene. Finally, performing correlation and cluster analysis on gene expression data, the gonadal adipose tissue was suggested as causal for the metabolic syndrome in the BFMI861-S1 line.

Based on the results obtained in these studies, the following conclusions are drawn:

1- BFMI lines are excellent mouse models not only to investigate obesity but also additional metabolic diseases, in particular traits associated with the metabolic syndrome. In this study, the BFMI861-S1 was successfully used as model for the fatty liver as a feature of the metabolic syndrome.

2- AILs obtained from closely related inbred lines are excellent tools that can be used to fine map QTL and for prioritization of candidate genes in a QTL. Due to high genetic similarity large genomic regions can be excluded.

3- Four QTL were identified in the AIL (BFMI861-S1xBFMI861-S2). The QTL on Chr 17 was associated with gonadal adipose tissue weight, liver weight, blood glucose concentration, and liver triglycerides and is the main driver of the metabolic syndrome in the BFMI861-S1. The genes *Plg* and *Acat2* were prioritized as top candidate genes. These genes had been previously linked to metabolic traits. These results add evidence on the detected candidate genes and could therefore be used for further studies.

4- The main driver tissue in the BFMI861-S1 line contributing to the metabolic syndrome and related traits (gonadal adipose tissue weight, liver weight, blood glucose concentration, liver triglycerides, and body weight) is the gonadal adipose tissue.

Correlation analysis of gene expression data showed that mice of the same mouse line clustered together only with gene expression data of the gonadal adipose tissue.

5- In the AIL (BFMI861-S1 x B6N) we identified four QTL associated with body mass, liver triglycerides and liver weight. The decision tree prioritization approach identified for each QTL region candidate genes and confirmed *Bbs7* as top candidate genes for high body weight in the BFMI lines. This additional cross was important to obtain an AIL with a different background which allowed us to discover QTL that had not been identified in the crosses BFMI861-S1xBFMI861-S2 and BFMI860-12xB6N.

6- A decision tree that includes multiomics data from AIL parental lines allow to prioritize candidate genes even in large QTL regions.

7- QTL mapping together with a detailed prioritization approach can be used for the identification of candidate genes linked to common human diseases.

References

Abu-Toamih Atamni, H. J., Botzman, M., Mott, R., Gat-Viks, I., & Iraqi, F. A. (2018). Mapping novel genetic loci associated with female liver weight variations using Collaborative Cross mice. *Animal Models and Experimental Medicine*, 1(3), 212–220. Retrieved from https://doi.org/10.1002/ame2.12036

Aguilar-Salinas, C. A., & Viveros-Ruiz, T. (2019). Recent advances in managing/understanding the metabolic syndrome. *F1000Research*, 8, 370. Retrieved from https://doi.org/10.12688/f1000research.17122.1

Ajjan, R. A., Gamlen, T., Standeven, K. F., Mughal, S., Hess, K., Smith, K. A., … Grant, P. J. (2013). Diabetes is associated with posttranslational modifications in plasminogen resulting in reduced plasmin generation and enzyme-specific activity. *Blood*, 122(1), 134–142. Retrieved from https://doi.org/10.1182/blood-2013-04-494641

Albuquerque, D., Nóbrega, C., Manco, L., & Padez, C. (2017). The contribution of genetics and environment to obesity. *British Medical Bulletin*, 123(1), 159–173. Retrieved from https://doi.org/10.1093/bmb/ldx022

Arends, D, Heise, S., Kärst, S., Trost, J., & Brockmann, G. A. (2016). Fine mapping a major obesity locus (jObes1) using a Berlin Fat Mouse × B6N advanced intercross population. *International Journal of Obesity (2005)*, 40(11), 1784–1788. Retrieved from https://doi.org/10.1038/ijo.2016.150

Arends, D., Heise, S., Kärst, S., Trost, J., & Brockmann, G. A. (2016). Fine mapping a major obesity locus (jObes1) using a Berlin Fat Mouse × B6N advanced intercross population. *International Journal of Obesity*, 40(11), 1784–1788. Retrieved from https://doi.org/10.1038/ijo.2016.150

Arends, D., Prins, P., Jansen, R. C., & Broman, K. W. (2010). R/qtl: high-throughput multiple QTL mapping: Fig. 1. *Bioinformatics*, 26(23), 2990–2992. Retrieved from https://doi.org/10.1093/bioinformatics/btq565

Bhattacharyya, D., & Glick, B. S. (2007). Two Mammalian Sec16 Homologues Have Nonredundant Functions in Endoplasmic Reticulum (ER) Export and Transitional ER Organization. *Molecular Biology of the Cell*, 18(3), 839–849. Retrieved from https://doi.org/10.1091/mbc.e06-08-0707

Bissonnette, M. L. Z., Lane, J. C., & Chang, A. (2017). Extreme Renal Pathology in Alagille Syndrome. *Kidney International Reports*, 2(3), 493–497. Retrieved from https://doi.org/10.1016/j.ekir.2016.11.002

Bladt, F., Riethmacher, D., Isenmann, S., Aguzzi, A., & Birchmeier, C. (1995). Essential role for the c-met receptor in the migration of myogenic precursor cells into the limb bud. *Nature*, 376(6543), 768–771. Retrieved from https://doi.org/10.1038/376768a0

Blüher, M. (2019). Obesity: global epidemiology and pathogenesis. *Nature Reviews Endocrinology*, 15(5), 288–298. Retrieved from https://doi.org/10.1038/s41574-019-0176-8

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15), 2114–2120. Retrieved from https://doi.org/10.1093/bioinformatics/btu170

Bonner, S. M., Pietropaolo, S. L., Fan, Y., Chang, Y., Sethupathy, P., Morran, M. P., … Pietropaolo, M. (2012). Sequence Variation in Promoter of Ica1 Gene, Which Encodes Protein Implicated in Type 1 Diabetes, Causes Transcription Factor Autoimmune Regulator (AIRE) to Increase Its Binding and Down-regulate Expression. *Journal of Biological Chemistry*, 287(21), 17882–17893. Retrieved from https://doi.org/10.1074/jbc.M111.319020

Brockmann, G. A., Tsaih, S.-W., Neuschl, C., Churchill, G. A., & Li, R. (2009). Genetic factors contributing to obesity and body weight can act through mechanisms affecting muscle weight, fat weight, or both. *Physiological Genomics*, 36(2), 114–126. Retrieved from https://doi.org/10.1152/physiolgenomics.90277.2008

Broekema, R. V., Bakker, O. B., & Jonkers, I. H. (2020). A practical view of fine-mapping

and gene prioritization in the post-genome-wide association era. *Open Biology*, 10(1), 190221. Retrieved from https://doi.org/10.1098/rsob.190221

Bult, C. J., Blake, J. A., Smith, C. L., Kadin, J. A., Richardson, J. E., Anagnostopoulos, A., … Zhu, Y. (2019). Mouse Genome Database (MGD) 2019. *Nucleic Acids Research*, 47(D1), D801–D806. Retrieved from https://doi.org/10.1093/nar/gky1056

Cava, C., Bertoli, G., & Castiglioni, I. (2015). Integrating genetics and epigenetics in breast cancer: biological insights, experimental, computational methods and therapeutic potential. *BMC Systems Biology*, 9(1), 62. Retrieved from https://doi.org/10.1186/s12918-015-0211-x

Church, C., Lee, S., Bagg, E. A. L., McTaggart, J. S., Deacon, R., Gerken, T., … Cox, R. D. (2009). A Mouse Model for the Metabolic Effects of the Human Fat Mass and Obesity Associated FTO Gene. *PLoS Genetics*, 5(8), e1000599. Retrieved from https://doi.org/10.1371/journal.pgen.1000599

Clough, E., & Barrett, T. (2016). The Gene Expression Omnibus Database (pp. 93–110). Retrieved from https://doi.org/10.1007/978-1-4939-3578-9_5

Collesi, C., Santoro, M. M., Gaudino, G., & Comoglio, P. M. (1996). A splicing variant of the RON transcript induces constitutive tyrosine kinase activity and an invasive phenotype. *Molecular and Cellular Biology*, 16(10), 5518–5526. Retrieved from https://doi.org/10.1128/MCB.16.10.5518

Darvasi, A., & Soller, M. (1995). Advanced intercross lines, an experimental population for fine genetic mapping. *Genetics*, 141(3), 1199–207. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8582624

Delpero, M., Arends, D., Sprechert, M., Krause, F., Kluth, O., Schürmann, A., … Hesse, D. (2021). Identification of four novel QTL linked to the metabolic syndrome in the Berlin Fat Mouse. *International Journal of Obesity*. Retrieved from https://doi.org/10.1038/s41366-021-00991-3

Devlin, B., & Roeder, K. (1999). Genomic control for association studies. *Biometrics*, 55(4),

997–1004. Retrieved from https://doi.org/10.1111/j.0006-341x.1999.00997.x

Dupuis, J., & Siegmund, D. (1999). Statistical methods for mapping quantitative trait loci from a dense set of markers. *Genetics*, 151(1), 373–86. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9872974

Durinck, S., Spellman, P. T., Birney, E., & Huber, W. (2009). Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nature Protocols*, 4(8), 1184–1191. Retrieved from https://doi.org/10.1038/nprot.2009.97

Engin, A. (2017). The Definition and Prevalence of Obesity and Metabolic Syndrome (pp. 1–17). Retrieved from https://doi.org/10.1007/978-3-319-48382-5_1

Krause, F., Mohebian, K., Delpero, M., Hesse, D., Kühn, R., Arends, D., Brockmann, G.A. (2021). A deletion containing a CTCF-element in intron 8 of the Bbs7 gene is partially responsible for juvenile obesity in the Berlin Fat Mouse. *Mammalian Genome*.

Fukao, T., Song, X.-Q., Mitchell, G. A., Yamaguchi, S., Sukegawa, K., Or, T., & Kondo, N. (1997). Enzymes of Ketone Body Utilization in Human Tissues: Protein and Messenger RNA Levels of Succinyl-Coenzyme A (CoA):3-Ketoacid CoA Transferase and Mitochondrial and Cytosolic Acetoacetyl-CoA Thiolases. *Pediatric Research*, 42(4), 498–502. Retrieved from https://doi.org/10.1203/00006450-199710000-00013

Gao, X. (2011). Multiple testing corrections for imputed SNPs. *Genetic Epidemiology*, 35(3), 154–158. Retrieved from https://doi.org/10.1002/gepi.20563

Gao, X., Becker, L. C., Becker, D. M., Starmer, J. D., & Province, M. A. (2010). Avoiding the high Bonferroni penalty in genome-wide association studies. *Genet Epidemiol.*, 34(1), 100–105. Retrieved from https://doi.org/10.1002/gepi.20430

Gargiulo, S., Gramanzini, M., Megna, R., Greco, A., Albanese, S., Manfredi, C., & Brunetti, A. (2014). Evaluation of Growth Patterns and Body Composition in C57Bl/6J Mice Using Dual Energy X-Ray Absorptiometry. *BioMed Research International*, 2014, 1–11. Retrieved from https://doi.org/10.1155/2014/253067

Ghosh, A. K., & Vaughan, D. E. (2012). PAI-1 in tissue fibrosis. *Journal of Cellular Physiology*,

227(2), 493–507. Retrieved from https://doi.org/10.1002/jcp.22783

Gonzalez Malagon, S. G., Melidoni, A. N., Hernandez, D., Omar, B. A., Houseman, L., Veeravalli, S., … Shephard, E. A. (2015). The phenotype of a knockout mouse identifies flavin-containing monooxygenase 5 (FMO5) as a regulator of metabolic ageing. *Biochemical Pharmacology*, 96(3), 267–277. Retrieved from https://doi.org/10.1016/j.bcp.2015.05.013

Goodarzi, M. O. (2018). Genetics of obesity: what genetic association studies have taught us about the biology of obesity and its complications. *The Lancet Diabetes & Endocrinology*, 6(3), 223–236. Retrieved from https://doi.org/10.1016/S2213-8587(17)30200-0

Gotoh, M., Maki, T., Kiyoizumi, T., Satomi, S., & Monaco, A. P. (1985). an improved method for isolation of mouse pancreatic islets. *Transplantation*, 40(4), 437. Retrieved from https://doi.org/10.1097/00007890-198510000-00018

Gurumurthy, C. B., & Lloyd, K. C. K. (2019). Generating mouse models for biomedical research: technological advances. *Disease Models & Mechanisms*, 12(1). Retrieved from https://doi.org/10.1242/dmm.029462

Hantschel, C., Wagener, A., Neuschl, C., Teupser, D., & Brockmann, G. A. (2011). Features of the metabolic syndrome in the Berlin Fat Mouse as a model for human obesity. *Obesity Facts*, 4(4), 270–7. Retrieved from https://doi.org/10.1159/000330819

Harayama, T., Eto, M., Shindou, H., Kita, Y., Otsubo, E., Hishikawa, D., … Shimizu, T. (2014). Lysophospholipid Acyltransferases Mediate Phosphatidylcholine Diversification to Achieve the Physical Properties Required In Vivo. *Cell Metabolism*, 20(2), 295–305. Retrieved from https://doi.org/10.1016/j.cmet.2014.05.019

Haynes, W. (2013). Bonferroni Correction. In *Encyclopedia of Systems Biology* (pp. 154–154). New York, NY: Springer New York. Retrieved from https://doi.org/10.1007/978-1-4419-9863-7_1213

Hebebrand, J., Holm, J.-C., Woodward, E., Baker, J. L., Blaak, E., Durrer Schutz, D., …

Yumuk, V. (2017). A Proposal of the European Association for the Study of Obesity to Improve the ICD-11 Diagnostic Criteria for Obesity Based on the Three Dimensions Etiology, Degree of Adiposity and Health Risk. *Obesity Facts*, 10(4), 284–307. Retrieved from https://doi.org/10.1159/000479208

Heise, S, Trost, J., Arends, D., Wirth, E. K., Schäfer, N., Köhrle, J., … Brockmann, G. A. (2016). High Variability of Insulin Sensitivity in Closely Related Obese Mouse Inbred Strains. *Experimental and Clinical Endocrinology & Diabetes : Official Journal, German Society of Endocrinology [and] German Diabetes Association*, 124(9), 519–528. Retrieved from https://doi.org/10.1055/s-0042-109261

Heise, S., Trost, J., Arends, D., Wirth, E., Schäfer, N., Köhrle, J., … Brockmann, G. A. (2016). High Variability of Insulin Sensitivity in Closely Related Obese Mouse Inbred Strains. *Experimental and Clinical Endocrinology & Diabetes*, 124(09), 519–528. Retrieved from https://doi.org/10.1055/s-0042-109261

Hesse, D., Jaschke, A., Kanzleiter, T., Witte, N., Augustin, R., Hommel, A., … Schurmann, A. (2012). GTPase ARFRP1 Is Essential for Normal Hepatic Glycogen Storage and Insulin-Like Growth Factor 1 Secretion. *Molecular and Cellular Biology*, 32(21), 4363–4374. Retrieved from https://doi.org/10.1128/MCB.00522-12

Hesse, D., Dunn, M., Heldmaier, G., Klingenspor, M., & Rozman, J. (2010). Behavioural mechanisms affecting energy regulation in mice prone or resistant to diet- induced obesity. *Physiology & Behavior*, 99(3), 370–380. Retrieved from https://doi.org/10.1016/j.physbeh.2009.12.001

Hesse, D., Radloff, K., Jaschke, A., Lagerpusch, M., Chung, B., Tailleux, A., …Schürmann, A. (2014). Hepatic trans-Golgi action coordinated by the GTPase ARFRP1 is crucial for lipoprotein lipidation and assembly. *Journal of Lipid Research*, 55(1), 41–52. Retrieved from https://doi.org/10.1194/jlr.M040089

Hesse, D., Trost, J., Schäfer, N., Schwerbel, K., Hoeflich, A., Schürmann, A., & Brockmann, G. A. (2018). Effect of adipocyte-derived IGF-I on adipose tissue mass and glucose

metabolism in the Berlin Fat Mouse. *Growth Factors*, 36(1–2), 78–88. Retrieved from https://doi.org/10.1080/08977194.2018.1497621

Holland, D., Frei, O., Desikan, R., Fan, C.-C., Shadrin, A. A., Smeland, O. B., … Dale, A. M. (2020). Beyond SNP heritability: Polygenicity and discoverability of phenotypes estimated with a univariate Gaussian mixture model. *PLOS Genetics*, 16(5), e1008612. Retrieved from https://doi.org/10.1371/journal.pgen.1008612

Hoover-Plow, J., Ellis, J., & Yuen, L. (2002). In vivo plasminogen deficiency reduces fat accumulation. *Thrombosis and Haemostasis*, 87(6), 1011–9. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12083480

Hoover-Plow, J., & Yuen, L. (2001). Plasminogen Binding Is Increased with Adipocyte Differentiation. *Biochemical and Biophysical Research Communications*, 284(2), 389–394. Retrieved from https://doi.org/10.1006/bbrc.2001.4984

Joost, H.-G., & Schürmann, A. (2014). The genetic basis of obesity-associated type 2 diabetes (diabesity) in polygenic mouse models. *Mammalian Genome : Official Journal of the International Mammalian Genome Society*, 25(9–10), 401–12. Retrieved from https://doi.org/10.1007/s00335-014-9514-2

Kanehisa, M. (2000). KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research*, 28(1), 27–30. Retrieved from https://doi.org/10.1093/nar/28.1.27

Keaver, L., Xu, B., Jaccard, A., & Webber, L. (2020). Morbid obesity in the UK: A modelling projection study to 2035. *Scandinavian Journal of Public Health*, 48(4), 422–427. Retrieved from https://doi.org/10.1177/1403494818794814

Kim, H.-S., McKnite, A., Xie, Y., & Christian, J. L. (2018). Fibronectin type III and intracellular domains of Toll-like receptor 4 interactor with leucine-rich repeats (Tril) are required for developmental signaling. *Molecular Biology of the Cell*, 29(5), 523–531. Retrieved from https://doi.org/10.1091/mbc.E17-07-0446

Kluth, O, Mirhashemi, F., Scherneck, S., Kaiser, D., Kluge, R., Neschen, S., … Schürmann, A. (2011). Dissociation of lipotoxicity and glucotoxicity in a mouse model of obesity

associated diabetes: role of forkhead box O1 (FOXO1) in glucose-induced beta cell failure. *Diabetologia*, 54(3), 605–16. Retrieved from https://doi.org/10.1007/s00125-010-1973-8

Kluth, Oliver, Matzke, D., Kamitz, A., Jähnert, M., Vogel, H., Scherneck, S., … Schürmann, A. (2015). Identification of Four Mouse Diabetes Candidate Genes Altering ß-Cell Proliferation. *PLOS Genetics*, 11(9), e1005506. Retrieved from https://doi.org/10.1371/journal.pgen.1005506

Kosti, I., Jain, N., Aran, D., Butte, A. J., & Sirota, M. (2016). Cross-tissue Analysis of Gene and Protein Expression in Normal and Cancer Tissues. *Scientific Reports*, 6(1), 24799. Retrieved from https://doi.org/10.1038/srep24799

Kreuzer, S., Reissmann, M., & Brockmann, G. A. (2013). Gene test to elucidate the ETEC F4ab/F4ac receptor status in pigs. *Veterinary Microbiology*, 162(1), 293–295. Retrieved from https://doi.org/10.1016/j.vetmic.2012.07.049

Lai, E. C. (2004). Notch signaling: control of cell communication and cell fate. *Development*, 131(5), 965–973. Retrieved from https://doi.org/10.1242/dev.01074

Lan, N., Lu, Y., Zhang, Y., Pu, S., Xi, H., Nie, X., … Yuan, W. (2020). FTO – A Common Genetic Basis for Obesity and Cancer. *Frontiers in Genetics*, 11. Retrieved from https://doi.org/10.3389/fgene.2020.559138

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., … Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, 25(16), 2078–2079. Retrieved from https://doi.org/10.1093/bioinformatics/btp352

Li, Hao, & Auwerx, J. (2020). Mouse Systems Genetics as a Prelude to Precision Medicine. *Trends in Genetics*, 36(4), 259–272. Retrieved from https://doi.org/10.1016/j.tig.2020.01.004

Li, Heng, & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics (Oxford, England)*, 25(14), 1754–60. Retrieved from https://doi.org/10.1093/bioinformatics/btp324

Li, Heng, Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., … 1000 Genome Project Data Processing Subgroup. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics (Oxford, England)*, 25(16), 2078–9. Retrieved from https://doi.org/10.1093/bioinformatics/btp352

Li, T., Ning, Z., & Shen, X. (2021). Improved Estimation of Phenotypic Correlations Using Summary Association Statistics. *Frontiers in Genetics*, 12. Retrieved from https://doi.org/10.3389/fgene.2021.665252

Li, Y., Tesson, B. M., Churchill, G. A., & Jansen, R. C. (2010). Critical reasoning on causal inference in genome-wide linkage and association studies. *Trends in Genetics*, 26(12), 493–498. Retrieved from https://doi.org/10.1016/j.tig.2010.09.002

Liu, Q., Yang, T., Yu, T., Zhang, S., Mao, X., Zhao, J., … Liu, B. (2017). Integrating Small RNA Sequencing with QTL Mapping for Identification of miRNAs and Their Target Genes Associated with Heat Tolerance at the Flowering Stage in Rice. *Frontiers in Plant Science*, 8. Retrieved from https://doi.org/10.3389/fpls.2017.00043

Livak, K. J., & Schmittgen, T. D. (2001). Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the 2−ΔΔCT Method. *Methods*, 25(4), 402–408. Retrieved from https://doi.org/10.1006/meth.2001.1262

Locke, A. E., Kahali, B., Berndt, S. I., Justice, A. E., Pers, T. H., Day, F. R., … Speliotes, E. K. (2015). Genetic studies of body mass index yield new insights for obesity biology. *Nature*, 518(7538), 197–206. Retrieved from https://doi.org/10.1038/nature14177

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., … DePristo, M. A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, 20(9), 1297–303. Retrieved from https://doi.org/10.1101/gr.107524.110

McLaren, W., Gil, L., Hunt, S. E., Riat, H. S., Ritchie, G. R. S., Thormann, A., … Cunningham, F. (2016). The Ensembl Variant Effect Predictor. *Genome Biology*, 17(1), 122. Retrieved from https://doi.org/10.1186/s13059-016-0974-4

Mellado-Gil, J., Rosa, T. C., Demirci, C., Gonzalez-Pertusa, J. A., Velazquez-Garcia, S., Ernst, S., … Garcia-Ocana, A. (2011). Disruption of Hepatocyte Growth Factor/c-Met Signaling Enhances Pancreatic -Cell Death and Accelerates the Onset of Diabetes. *Diabetes*, 60(2), 525–536. Retrieved from https://doi.org/10.2337/db09-1305

Moessinger, C., Klizaite, K., Steinhagen, A., Philippou-Massier, J., Shevchenko, A., Hoch, M., … Thiele, C. (2014). Two different pathways of phosphatidylcholine synthesis, the Kennedy Pathway and the Lands Cycle, differentially regulate cellular triacylglycerol storage. *BMC Cell Biology*, 15(1), 43. Retrieved from https://doi.org/10.1186/s12860-014-0043-3

Morgan, A. P., Fu, C.-P., Kao, C.-Y., Welsh, C. E., Didion, J. P., Yadgary, L., … Pardo-Manuel de Villena, F. (2016). The Mouse Universal Genotyping Array: From Substrains to Subspecies. *G3: Genes|Genomes|Genetics*, 6(2), 263–279. Retrieved from https://doi.org/10.1534/g3.115.022087

Nadeau, J. H., & Auwerx, J. (2019). The virtuous cycle of human genetics and mouse models in drug discovery. *Nature Reviews Drug Discovery*, 18(4), 255–272. Retrieved from https://doi.org/10.1038/s41573-018-0009-9

Neuschl, C., Hantschel, C., Wagener, A., Schmitt, A. O., Illig, T., & Brockmann, G. A. (2010a). A unique genetic defect on chromosome 3 is responsible for juvenile obesity in the Berlin Fat Mouse. *International Journal of Obesity (2005)*, 34(12), 1706–14. Retrieved from https://doi.org/10.1038/ijo.2010.97

Neuschl, C., Hantschel, C., Wagener, A., Schmitt, A. O., Illig, T., & Brockmann, G. A. (2010b). A unique genetic defect on chromosome 3 is responsible for juvenile obesity in the Berlin Fat Mouse. *International Journal of Obesity*, 34(12), 1706–1714. Retrieved from https://doi.org/10.1038/ijo.2010.97

Ntountoumi, C., Vlastaridis, P., Mossialos, D., Stathopoulos, C., Iliopoulos, I., Promponas, V., … Amoutzias, G. D. (2019). Low complexity regions in the proteins of prokaryotes perform important functional roles and are highly conserved. *Nucleic Acids Research*,

47(19), 9998–10009. Retrieved from https://doi.org/10.1093/nar/gkz730

O'Neill, S., & O'Driscoll, L. (2015). Metabolic syndrome: a closer look at the growing epidemic and its associated pathologies. *Obesity Reviews*, 16(1), 1–12. Retrieved from https://doi.org/10.1111/obr.12229

Parker, R. (2018). The role of adipose tissue in fatty liver diseases. *Liver Research*, 2(1), 35–42. Retrieved from https://doi.org/10.1016/j.livres.2018.02.002

Pilon, M., Peng, X.-R., Spence, A. M., Plasterk, R. H. A., & Dosch, H.-M. (2000). The Diabetes Autoantigen ICA69 and Its Caenorhabditis elegans Homologue, ric-19 , Are Conserved Regulators of Neuroendocrine Secretion. *Molecular Biology of the Cell*, 11(10), 3277–3288. Retrieved from https://doi.org/10.1091/mbc.11.10.3277

Qi, Q., Workalemahu, T., Zhang, C., Hu, F. B., & Qi, L. (2012). Genetic variants, plasma lipoprotein(a) levels, and risk of cardiovascular morbidity and mortality among two prospective cohorts of type 2 diabetes. *European Heart Journal*, 33(3), 325–334. Retrieved from https://doi.org/10.1093/eurheartj/ehr350

Raj, M. R., & Sreeja, A. (2018). Analysis of Computational Gene Prioritization Approaches. *Procedia Computer Science*, 143, 395–410. Retrieved from https://doi.org/10.1016/j.procs.2018.10.411

Rockman, M. V., & Kruglyak, L. (2008). Breeding Designs for Recombinant Inbred Advanced Intercross Lines. *Genetics*, 179(2), 1069–1078. Retrieved from https://doi.org/10.1534/genetics.107.083873

Rosen, E. D., & Spiegelman, B. M. (2006). Adipocytes as regulators of energy balance and glucose homeostasis. *Nature*, 444(7121), 847–853. Retrieved from https://doi.org/10.1038/nature05483

Sahibdeen, V., Crowther, N. J., Soodyall, H., Hendry, L. M., Munthali, R. J., Hazelhurst, S., … Lombard, Z. (2018). Genetic variants in SEC16B are associated with body composition in black South Africans. *Nutrition & Diabetes*, 8(1), 43. Retrieved from https://doi.org/10.1038/s41387-018-0050-0

Schmitt, A. O., Bortfeldt, R., Neuschl, C., & Brockmann, G. A. (2009). RandoMate: a program for the generation of random mating schemes for small laboratory animals. *Mammalian Genome : Official Journal of the International Mammalian Genome Society*, 20(5), 321–5. Retrieved from https://doi.org/10.1007/s00335-009-9185-6

Schulz, N., Himmelbauer, H., Rath, M., van Weeghel, M., Houten, S., Kulik, W., … Schürmann, A. (2011). Role of Medium- and Short-Chain L-3-Hydroxyacyl-CoA Dehydrogenase in the Regulation of Body Weight and Thermogenesis. *Endocrinology*, 152(12), 4641–4651. Retrieved from https://doi.org/10.1210/en.2011-1547

Shindou, H., Hishikawa, D., Nakanishi, H., Harayama, T., Ishii, S., Taguchi, R., & Shimizu, T. (2007). A Single Enzyme Catalyzes Both Platelet-activating Factor Production and Membrane Biogenesis of Inflammatory Cells. *Journal of Biological Chemistry*, 282(9), 6532–6539. Retrieved from https://doi.org/10.1074/jbc.M609641200

Siebel, C., & Lendahl, U. (2017). Notch Signaling in Development, Tissue Homeostasis, and Disease. *Physiological Reviews*, 97(4), 1235–1294. Retrieved from https://doi.org/10.1152/physrev.00005.2017

Sigmon, J. S., Blanchard, M. W., Baric, R. S., Bell, T. A., Brennan, J., Brockmann, G. A., … Manuel de Villena, F. P. (2020). Content and Performance of the MiniMUGA Genotyping Array: A New Tool To Improve Rigor and Reproducibility in Mouse Research. *Genetics*, 216(4), 905–930. Retrieved from https://doi.org/10.1534/genetics.120.303596

Uffelmann, E., Huang, Q. Q., Munung, N. S., de Vries, J., Okada, Y., Martin, A. R., … Posthuma, D. (2021). Genome-wide association studies. *Nature Reviews Methods Primers*, 1(1), 59. Retrieved from https://doi.org/10.1038/s43586-021-00056-9

Van Herck, M., Vonghia, L., & Francque, S. (2017). Animal Models of Nonalcoholic Fatty Liver Disease—A Starter's Guide. *Nutrients*, 9(10), 1072. Retrieved from https://doi.org/10.3390/nu9101072

Wagener, A., Schmitt, A. O., Aksu, S., Schlote, W., Neuschl, C., & Brockmann, G. A. (2006). Genetic, sex, and diet effects on body weight and obesity in the Berlin Fat Mouse Inbred

lines. *Physiological Genomics*, 27(3), 264–70. Retrieved from https://doi.org/10.1152/physiolgenomics.00225.2005

Warwick Vesztrocy, A., Dessimoz, C., & Redestig, H. (2018). Prioritising candidate genes causing QTL using hierarchical orthologous groups. *Bioinformatics*, 34(17), i612–i619. Retrieved from https://doi.org/10.1093/bioinformatics/bty615

Watanabe, K., Stringer, S., Frei, O., Umićević Mirkov, M., de Leeuw, C., Polderman, T. J. C., … Posthuma, D. (2019). A global overview of pleiotropy and genetic architecture in complex traits. *Nature Genetics*, 51(9), 1339–1348. Retrieved from https://doi.org/10.1038/s41588-019-0481-0

Yalcin, B., Flint, J., & Mott, R. (2005). Using progenitor strain information to identify quantitative trait nucleotides in outbred mice. *Genetics*, 171(2), 673–81. Retrieved from https://doi.org/10.1534/genetics.104.028902

Zeggini, E., Scott, L. J., Saxena, R., Voight, B. F., Marchini, J. L., Hu, T., … Altshuler, D. (2008). Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. *Nature Genetics*, 40(5), 638–645. Retrieved from https://doi.org/10.1038/ng.120

Zhang, Y., Proenca, R., Maffei, M., Barone, M., Leopold, L., & Friedman, J. M. (1994). Positional cloning of the mouse obese gene and its human homologue. *Nature*, 372(6505), 425–32. Retrieved from https://doi.org/10.1038/372425

Acknowledgements

My special thanks are expressed to:

My supervisor, Deike Hesse, for giving me the opportunity to pursue my PhD in your group. Thank you for supervising and helping me with patience, encouragement, and understanding throughout my PhD journey. Your intellectual guidance and steadfast support helped me develop myself in the academic field and be able to face future challenges with more confidence. Thank you for commenting on my manuscripts and dissertation patiently and carefully.

My supervisor, Gudrun A. Brockmann, for giving me the opportunity to pursue my PhD in your group. Thank you for supervising and helping me with patience, encouragement, and understanding throughout my PhD journey. Your intellectual guidance and steadfast support helped me develop myself in the academic field and be able to face future challenges with more confidence. Thank you for commenting on my manuscripts and dissertation patiently and carefully.

Danny Arends, for all the guidance, help and support. Thank you for your patience in answering my questions every time. Thank you also for teaching me how to do proper data analysis.

All my colleagues past and present at the Breeding Biology and Molecular Genetics group for giving me a comfortable atmosphere to discuss questions and exchange ideas. The time we spent together was short but remarkable. I will keep that in my heart.

Last but not least, none of this work would have been possible without the endless support and encouragement of my family, my girlfriend, and my friends. Thank you for your trust and company.