A Systems Approach to Identify Genetic and Environmental Regulators of Metabolism

THÈSE N^{O} 6486 (2015)

PRÉSENTÉE LE 13 MARS 2015

À LA FACULTÉ DES SCIENCES DE LA VIE
CHAIRE NESTLÉ EN MÉTABOLISME ÉNERGÉTIQUE
PROGRAMME DOCTORAL EN BIOTECHNOLOGIE ET GÉNIE BIOLOGIQUE

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

POUR L'OBTENTION DU GRADE DE DOCTEUR ÈS SCIENCES

PAR

Evan Graehl WILLIAMS

acceptée sur proposition du jury:

Prof. M. Lütolf, président du jury
Prof. J. Auwerx, directeur de thèse
Prof. R. Aebersold, rapporteur
Prof. G. Barsh, rapporteur
Prof. A. Levine, rapporteur
Prof. J. Lusis, rapporteur

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Suisse
2014

Acknowledgements

I came to EPFL shortly after graduating from college with an engineering degree and only a cursory understanding of genetics and the intricacies of biochemistry and metabolism. My initial position here was only a year, with the goal of completing my thesis project for a master's degree in bioinformatics. I soon found the work and the environment here to be excellent, and jumped at the possibility of staying to pursue a doctoral degree. At the end of my internship, I performed a 180° turn away from *in vitro* biochemical work, and into a new project in the laboratory on metabolic phenotypes and how they are affected by diet and genetics. I gradually took the reins of this project and became progressively more interested in the "big picture" of genetics. To this end, I am tremendously grateful to Professor Johan Auwerx, and I am surprised (yet delighted) he chose to give such a large project to me and another junior Ph.D. student in the lab, Pénélope Andreux, with whom I worked closely for the beginning of my thesis. After more than five years at the EPFL, I have to thank a tremendous number of people who have helped me over the years, many of whom I lack the space to thank here.

Thank you to Rob for getting me into genetics in the first place, and it's lucky for me that family businesses have relaxed child labor laws. Thanks to Yasmin for translating the abstract, providing moral support and making me particularly lucky to have stayed in Switzerland for my doctoral work. Thanks to Kathy for encouraging me to come out here, and to Alex for encouraging me to stay.

Thank you very much to all of my previous teachers and mentors who have helped me develop from a student into a researcher—Lu Lu, Klaus Schughart, and Jake Lusis for welcoming me into your laboratories as a junior student and learn about genetics and getting me interested in the subject before I even knew myself—Tomasz Tkacyk, Ann Saterbak, and Maria Oden for developing my engineering background, and Stephan Morgenthaler for advising me throughout my doctoral work, both in statistics, and in more general aspects of professional and academic life. Thank you also to Profs. Johan Auwerx, Rob Williams, Bart Deplancke, and Kristina Schoonjans for believing me capable enough to take lead positions on project aims for major (and successful!) grant applications, and to Profs Klaus Schughart, Xavier Montagutelli, and Leo Schalkwyk for allowing me to participate in scientific debates and allowing me to teach courses, whether through the SysGenet program, or at their home institutions.

In the laboratory, thank you to Dr. Raffaele Teperino, who managed to teach me laboratory technique almost from scratch during my internship year at the EPFL, and Dr. Hiroyasu Yamamoto, who kindly put up with me while I transitioned from bioengineer to biochemist to bioinformatician. Thanks also to Dr. Riekelt Houtkooper, who helped transition me to the BXD project and who showed me how the "systems" project could be applied at a smaller scale. Also a thank you to Patricia (et Ludovic) who hosted me for my first two years in Switzerland, and Arash (and Umlaut) the year after. Thanks to everyone who helped organize, establish, and support the entire BXD phenotyping project, particularly Dr. Charles Thomas, Cristina Cartoni and Sebastien Lamy, and to Sabrina Bichet, Norman Moullan, Marie-Laure Denereaz, Giuseppe Lo Sasso, Hongbo Zhang, Thibaud Clerc, and everyone who filled in when needed. Thanks to Tarik Ouhmad and Tiffany Amariuta, my first trainees, and a tremendous thank you to the facility managers and animal caretakers who have helped over the years—Dr. Xavier Warot, Dr. Philippe Cettour-Rose, Dr. Roy Comb, Dr. Raphaël Doenlen, Gisèle Ferrand, Saadia Mutanda Soussi, Pascal Guéla, Aurore Bony, and Christine Pehm.

During my thesis, I have also had the pleasure of establishing at least a dozen strong collaborations inside and outside the EPFL, allowing me to learn from the best in the field for a tremendous range of specialties. In particular, thank you to Dr. Ruedi Aebersold and Dr. Yibo Wu for their knowledge of proteomics, to Dr. Nicola Zamboni and Sébastien Dubuis for their knowledge of metabolomics, to Lorne Rose for transcriptomics, to Dr. Dan Peterson, Dr. Daniel Ciobanu, and Maria Elisa Perez-Munoz for microbiomics, Dr. Zoltán Kulatnik for human translation, Dr. Sander Houten and Dr. Carman Argmann for plasma analysis and validation of results, and to Dr. Paul Franken, Dr. Bernard Thorens, and Shanaz

Diessler for their corresponding work on the BXDs and getting the Sinergia project together. Thank you to Michael Frochaux for introducing me to the world of *Drosophila*, Dr. Xu Wang for introducing me to the *Arabidopsis* world, and Dr. Laurent Mouchiroud for introducing me to the world of *C. elegans.* And finally a big thank you to Valerie Stengel for organizing *everything* in the lab and keeping us all running as smoothly as possible!

Throughout the thesis, I will use the pronoun "I" to describe the work in which I was directly involved or responsible, but in most cases a rotating cast of "we" would be more appropriate, including many of those thanked above.

Evan Williams
Lausanne, September 2014

Abstract

For more than a century, it has been recognized that our genetic inheritance and our environment interact to shape who we are and how we act ("nature vs. nurture"). The study of genetics has allowed us to explain why traits can vary dramatically between individuals (i.e. trait *variance*), and yet often be strongly shared within families (i.e. trait *heritability*). Scientists, statisticians, and physicians can calculate heritability and can observe how both genes and environment influence health, but relatively few interventions or treatments have stemmed from this expanded knowledge. In large part this is because testing biomedical hypotheses is difficult, because we are all so different, and because environmental factors are so hard to control.

In 2001 the first (nearly) complete DNA sequence of the human genome was generated after years of technical development and brute force effort. This accomplishment has ushered in a new field of personalized medicine, as prevention and treatment can both be tailored based on genome and environment. Unfortunately, the science of genomic sequencing has greatly outpaced our ability to actually understand the genetic code, and it remains difficult to make accurate predictions of an individual's characteristics and susceptibilities except in a few clear-cut cases such as eye color or risk of Huntington's disease. There are several reasons for this disconnect: (1) many traits and diseases are driven by complex interactions among environmental causes and genetic risks, (2) there are many aspects of genetics which we do not fully understand and that cannot be easily observed (e.g. non-coding RNAs, epigenetic modifications, and complex metagenomic cohorts that we all carry), and (3) gene products (such as mRNA and protein) change dramatically over time and across cell types in complex and even unpredictable ways.

Today, myriad developments promise to improve our capacity to predict resistance and susceptibility to diseases based on our individual genome. This technical capability has also provided new avenues for developing therapeutic agents and/or lifestyle changes. However, no matter our developments in scientific understanding, personalized medicine will never allow perfect prediction—environmental influences and personal choices will always affect our health, and medical treatments will still require information gained from standard medical checkups. Rather than perfect prediction, personalized medicine will instead provide more accurate predictions about health than those previously possible. Thus, people at high risk for (e.g.) diabetes will know better to watch their diet, while those at low risk should remember that biology is never completely predictable.

In this thesis, I have analyzed 45 strains of mice from a genetic reference population called the BXD with the goal of identifying major gene regulators of metabolic phenotypes, such as exercise capacity and glucose response. Each member of this BXD family, which contains approximately 150 distinct but related lines (or "strains"), has a unique genetic makeup that has been fixed by inbreeding. Each family member is thus available as an unlimited supply of "identical twins" which may be studied over years and among laboratories. With this population, it is possible to both (1) test what would occur to a single individual in different environmental conditions and (2) analyze how much environmental influences vary across genetically diverse individuals. In my study, which examines 10 individuals from each of 45 strains, it would be as if two parents had 450 children, split into 45 pairs of identical ten-tuplets. Each individual (in this case) thus has nine genetically identical twins, and 440 other siblings who share on average half the genome.

My thesis seeks to address two main questions: (1) how do genetics, sex, and environment interact to shape our metabolome? And (2) how can large systems-scale datasets be applied to other research projects outside my own direct interests? For the first question, in the first study (chapter 2) I examined the metabolic differences between males and females who share the same genome (excepting X and Y chromosomes), and I evaluated how traits vary between identical male/female twins and how traits compare between related male/female siblings. In the second study on this question (chapter 4), I examined the differences between a population of male BXD strains fed a healthy chow diet (**CD**, 6% calories from fat, 74% from carbohydrates, and 20% from protein), and their identical quintuplet populations fed a high fat diet (**HFD**, 60% calories from fat, 20% from carbohydrates, and 20% from protein). This study likewise highlighted dramatic differences due to diet and genotype. For example, most strains gained 25% body mass

when fed the HFD, but two strains remained completely lean regardless of diet! In the second, much more massive study, I obtained detailed molecular data on the same individuals, including their gene expression and metabolite levels in many different tissues. With this "systems" data, not only have I been able to determine to what extent traits are sensitive to changes in environment (between diets) and differences in genetics (between strains), but I have also been able to identify *how* and *why* many of these differences occur, most importantly including a new diabetes-related gene, *Dhtkd1* driving changes in the metabolite 2-aminoadipate.

For the second case (chapter 3), I used the data acquired in chapters 2 and 4, but applied these data to extant hypotheses generated by other scientists. Here, the fundamental aim is to make as broad use as possible of the data generated. The population dataset is too large and diverse for any one researcher to completely exhaust its utility; thus I shared my findings with other scientists to see if their hypotheses could be confirmed and expanded using the BXD data. In this case, there is no single fundamental question, but instead many independent questions can be investigated using the same data to answer questions about sex, genotype, diet, and metabolism.

Through both studies, I identified differences among the lines in nearly every trait measured: both the overt "clinical" traits, such as body weight, and molecular traits, such as gene expression. In these cases, we could attribute this variation to differences in (1) genetics, (2) sex, and/or (3) diet. Genetic variance influences nearly all traits, while the sex of the animal affected about half of the traits, and diet affected only about a third. While attributing sex and diet to the causal phenotypic effect is simple enough, identifying the precise underlying genetic differences behind trait variance is a greater challenge. To this end, we were able to identify causal genetic regions which influence heart rate, the capacity to exercise, response to glucose, standard ambulatory movement, and many metabolic parameters. However, other parameters such as body weight were influenced by too many disparate genetic variants for us to isolate any singular effect. For some of these genetic regions, such as glucose response and movement activity, we were able to identify precisely which gene variant was causing the phenotypic variant.

In chapter 2, I helped to define two genes causal of major phenotypic effects and their related sex-linked and dietary differences. For the first study (section 2.1), I established that both mutations and sex-linked effects on the gene *alkaline phosphatase* (*Alpl*) result in changes in bone density and size. In the second study (section 2.2), I established that mutations in the gene *aryl hydrocarbon receptor* (*Ahr*), sex, and diet all influence spontaneous movement activity. In chapter 3, we had previously identified *NCoR1* and *Fra-1* as genes which influence muscle strength and liver metabolism, respectively. Using the BXD data, we were able to discern *how* these two genes can influence their associated phenotypes. In chapter 4, we examined the molecular characteristics of the liver in the BXD population in great detail, and we determined that genetic changes in the gene *Dhtkd1* influence development of the disease 2-aminoadipic aciduria, which in turn affects glucose homeostasis and diabetes. Furthermore, we were able to apply our links between *Alpl, Ahr,* and *Dhtkd1* to human population studies, and we are confident that these links found in mice will be validated in humans and may lead to medical advances.

In ongoing studies, I am examining additional genetic factors underlying metabolic variance in the BXDs, and continuing research on the leads into glucose homeostasis and diabetes so far uncovered. However, the data obtained during this thesis far exceeds the capacity of a single researcher or even a single research group to fully exploit. Critically, one of the primary benefits of working with strain families like the BXDs is that, due to their "infinite" reproducibility, exceptionally detailed medical records can be obtained for this population over many years, and results from different research groups and even research fields can be assembled together. Therefore, the scope of my project will extend beyond the focus presented in this thesis. While my collaborators and I will continue to find explanations and leads or treatments for metabolic diseases using this approach, these data may also inform upon critical pathways involved in (for example) cancer or neurological disorders. By making the data and methods publicly accessible, this work will find more utility in explaining and treating diseases beyond what we (as a laboratory or even direct collaborative network) could accomplish alone. As we are already beginning to see, the communal assembly of massive medical records of phenotypes, gene expression data, metabolite levels, and so forth, provides a platform with which we will actually be able to accurately predict disease susceptibility, ushering us into the era of personalized medicine.

Keywords: BXDs, systems genetics, genetic reference populations, holistic genetics, reductive genetics, energy metabolism, diabetes, QTL mapping, ALPL, AHR, DHTKD1, NCoR1, AP-1, PPARg

Résumé

Voilà plus d'un siècle, il a été démontré que notre héritage génétique interagit avec notre environnement pour déterminer qui nous sommes et comment nous agissons (inné/acquis). L'étude de la génétique nous permet d'expliquer pourquoi les caractères phénotypiques diffèrent autant entre individus (i.e. variance de trait) mais peuvent pourtant être très similaires entre les membres d'une famille (héritage de traits). Les scientifiques, les statisticiens et les physiciens ont mesuré les effets d'héritage et d'environnement, mais peu d'interventions ou de traitements ont émergé de cette grand connaissance. Ceci est en grande partie dû au fait qu'il est difficile de tester des hypothèses biomédicales, que nous sommes tous très différents et parce que les facteurs de l'environnement sont très difficiles à contrôler.

En 2001, la première séquence ADN quasi complète du génome humain a été générée, après des années de développement technique et de travail acharné. Cet accomplissement a permis la création d'un nouveau domaine de recherche, la médecine personnalisée, où la prévention et les traitements peuvent être conçus à partir du génome et de l'environnement des patients. Malheureusement, la génomique a largement dépassé notre capacité à comprendre le code génétique et il est encore difficile de prédire correctement les caractéristiques (et susceptibilités, risques) d'un individu, hormis dans des cas flagrant tel que la couleur des yeux ou le risque de développer la maladie de Huntington. Il y a plusieurs raisons à cet écueil : (1) beaucoup de traits et de maladies sont les conséquences d'interactions complexes entre causes environnementales et risques génétiques ; (2) beaucoup d'aspects de la génétique sont difficiles à comprendre et à observer (par ex. : ARN non codants, modifications épigénétiques et cohortes métagénomiques complexes que nous possédons tous) et (3) les produits de gènes (ARN codants et protéines) changent radicalement au fil du temps et des types de cellules, de manière complexe et imprévisible.

Aujourd'hui, une combinaison de développements pourrait permettre d'améliorer notre capacité à prédire résistance et vulnérabilité aux maladies, en se basant sur notre génome individuel. Cette possibilité technique ouvre la voie au développement de nouvelles approches thérapeutiques ou des changements de style de vie. Cependant, malgré toutes les avancées scientifiques, la médecine personnalisée ne pourra jamais permettre une prédiction parfaite, car l'environnement et nos choix personnels auront toujours une influence sur notre santé; les traitements médicaux nécessiteront encore des informations obtenues lors de contrôles médicaux standards. Plutôt qu'une prédiction parfaite, la médecine personnalisée apportera une prédiction de notre santé plus correcte que celles possibles auparavant. Ainsi, les patients présentant un haut risque de (par ex.) diabète sauront mieux surveiller leur alimentation, alors que ceux à faible risque devraient garder en tête que la biologie n'est jamais totalement prévisible.

Durant cette thèse, j'ai analysé 45 souches d'une population de référence génétique nommée BXD, dans le but d'identifier les gènes régulateurs de phénotypes métaboliques, comme la capacité à réaliser un exercice physique et la réponse au glucose. Chaque membre de cette famille de BXD (environ 150 lignées reliées mais distinctes) a un fond génétique unique, qui a été corrigé par rétrocroisement. Chaque membre est ainsi disponible comme une source illimitée de « jumeaux identiques », qui peuvent être étudiés au fil des ans et des laboratoires. Avec cette population, il est possible de (1) tester les effets sur un seul individu de divers environnements et (2) analyser à quel point les influences de l'environnement varient entre des individus génétiquement différents. Par exemple, une étude examinant 10 individus provenant de 100 souches équivaut au cas où deux parents auraient eu 1'000 enfants, divisés en 100 paires de 10 couples identiques. Dans ce cas, chaque individu a neuf jumeaux génétiquement identiques, et 990 autres frères et sœurs qui partagent la moitié du génome en moyenne.

Dans ma thèse, je cherche à répondre principalement à deux questions : (1) de quelle manière la génétique, le sexe et l'environnement interagissent pour former notre métabolome. Et (2) comment peut on exploiter un système de base de données de cette échelle pour d'autres projets de recherche que le mien ? Pour la première question (chapitre 2), j'ai examiné les différences métaboliques entre individus mâles et femelles qui partagent un même génome (sauf les chromosomes X et Y). J'ai évalué comment les traits entre mâles et femelles jumeaux varient, ainsi qu'entre frères et sœurs. Dans la deuxième étude concernant cette question (chapitre 4), j'ai examiné les différences entre une popula-

tion mâle de souche BXD ayant une alimentation pauvre en graisses (dite « chow diet » **CD**, 6% de calories provenant de graisses, 74% de carbohydrates et 20% de protéines) et la population de leur jumeaux monozygotes, nourris avec une alimentation riche en graisses (dite « high fat diet » **HFD**, soit 60% de calories provenant de graisses, 20% de carbohydrates et 20% de protéines). De même, cette étude a permis d'observer d'importantes différences dues à l'alimentation et au génotype. Par exemple, la plupart des souches nourries au HFD ont vus leur masse corporelle augmenter de 25%, sauf deux souches qui sont restées minces, indépendamment de leur alimentation ! Dans cette deuxième étude, d'une plus grande portée, j'ai collecté des données moléculaires détaillées des mêmes individus, comprenant l'expression de leurs gènes et leurs niveaux de métabolite dans divers tissus. Grâce à ces « systèmes » de données, j'ai pu non seulement déterminer à quel point les traits phénotypiques sont sensibles au changement de l'environnement (différentes alimentations) et les différences génétiques (entre souches) mais également identifier *comment* et *pourquoi* nombre de ces différences existent, incluant notamment l'identification d'un nouveau gène de susceptibilité au diabète appelé DHTKD1 et impliqué dans le métabolisme de l'acide 2-aminoadipadic.

Pour le deuxième cas (chapitre 3), j'ai utilisé les données obtenues dans les chapitres 2 et 4, mais en les appliquant cette fois-ci des hypothèses générées par d'autres scientifiques. Le but étant ici de faire un usage le plus large possible des données produites. La population étudiée est trop grande et trop diverse pour qu'un seul chercheur puisse utiliser tout son potentiel ; j'ai donc partagé mes découvertes avec d'autres chercheurs, afin de voir si leurs hypothèses pouvaient être confirmées et étendues, en utilisant les données BXD. Dans ce cas, il n'y a pas une seule grande questions, mais plutôt plusieurs questions indépendantes, auxquelles on peut répondre en utilisant les mêmes données que celles générées pour répondre aux questions de sexe, génotype, métabolisme et alimentation.

A travers ces deux études, j'ai pu identifier des différences entre les souches pour pratiquement chaque caractéristique mesuré : tant pour les traits/caractéristiques « cliniques » tels que le poids, que pour des traits/caractéristiques « moléculaires », tel que l'expression des gènes. Dans ce cas, on pourrait attribuer la variation à des différences (1) génétiques, (2) de sexe et/ ou (3) d'alimentation. La génétique affecte pratiquement toutes les caractéristiques/traits, le sexe des animaux influence environ la moitié des traits et l'alimentation agit environ un tiers des traits. Si attribuer le changement de sexe et d'alimentation à une fonction causale est relativement simple, le principal défi est bien d'identifier précisément *quelles* variantes génétiques causent le changement. Nous avons pu identifier des régions de causalité génétiques, dont découlent des différences majeures dans le rythme cardiaque des animaux, leur capacité d'exercice, leur réponse au glucose, les mouvements de base et beaucoup de paramètres métaboliques. D'autres paramètres, tel que la prise de poids, sont affectés par trop de variantes génétiques disparates pour pouvoir identifier une cause particulière. Pour certaines de ces régions génétiques, nous avons pu identifier avec précision quelle variante d'un gène causait la variante phénotypique.

Dans le chapitre 2, je suis parvenu à définir deux gènes responsable d'effets phénotypiques majeurs, ainsi que les différences de sexe et d'alimentation liés. Pour la première étude (chapitre 2.1), j'ai pu établir que les effets liés aux mutations ainsi qu'au sexe sur le gène de l'*alkaline phosphatase* (Alpl) provoquent des changement de densité et de taille des os. Dans la deuxième étude (chapitre 2.2), j'établis que l'activité physique spontanée est influencée par les mutations dans le gène *aryl hydrocarbon receptor* (*Ahr*), le sexe et l'alimentation. Dans le chapitre 3, nous avions précédemment identifié deux gènes, NCoR1 et Fra-1, qui influencent respectivement la force musculaire et le métabolisme du foie. En utilisant les données BXD, nous avons pu voir *comment* ces deux gènes peuvent indfluencer leurs phénotypes associés. Dans le chapitre 4, nous avons examiné dans le détail les caractéristiques moléculaires du foie dans la population BXD, et nous pouvons affirmer que les changements génétiques dans le gène *Dhtkd1* a une influence sur le développement de la maladie 2-aminoadipic aciduria, qui affecte elle-même l'homéostasie du glucose et le diabète. De plus, nous avons pu appliquer les liens entre Alpl, Ahr et Dhtkd à des études de populations humaines. Nous pensons que ces liens trouvés chez la souris seront validés chez l'être humain et pourront amener des avancées pour la médecine.

Dans des recherches en cours, j'observe d'autres facteurs génétiques causant des variances métaboliques chez les BXD et je continue la recherche sur l'homéostasie du glucose. Cependant, les données obtenues dans le cadre de cette thèse excèdent de loin la capacité d'un seul chercheur ou même d'un groupe de chercheurs à en exploiter tout le potentiel. Un des avantage à travailler avec des famille souches comme les BXD est leur « reproductibilité infinie », ainsi que les données médicales incroyablement détaillées que l'on peut obtenir au fil des années. Les résultats de divers groupes de recherches et même de divers domaines de recherches peuvent être rassemblés. Mon projet pourra ainsi

servir au-delà du sujet abordé dans cette thèse, tandis que mes collaborateurs et moi continuerons à trouver des explications et des traitements aux maladies métaboliques, en utilisant la même approche. Il se peut que ces données puissent révéler des causes de cancer ou de problèmes neurologiques. En rendant les données et les méthodes accessibles au public, l'utilité de ce travail s'en trouve décuplée. Nous pouvons déjà l'observer : la collecte commune massive de données médicales sur les phénotypes, l'expression des gènes, les niveaux de métabolites, etc. offre une plateforme qui pourra prévoir de manière précise la sensibilité aux maladies, nous propulsant dans l'ère de la médecine personnalisée.

Mot clés : BXD, systèmes génétiques, population génétique de référence, génétique holistique, génétique réductive, métabolisme énergétique, diabète, cartographie des QTLs, ALPL, AHR, DHTKD1, NCoR1, AP-1, PPARg

Contents

List of Figures

Chapter 1 Introduction

Modern scientific studies using animals to model human diseases began in the 19th century, driven initially by the development of vaccines and drug treatments for transmissible diseases [1]. A few decades thereafter, mice and *Drosophila* began to be used in research on fundamental genetics [2, 3]. The potential of using such organisms as models for complex human processes was quickly realized [4], yet the key genetic and environmental factors behind risk or resistance to diseases remained elusive. Nevertheless, dramatic strides were made in both genetic understanding and in scientific techniques and technologies, which gradually illuminated the full complexity of physiology. In tandem, it became clear that more numerous, complex, and controlled model populations would facilitate deciphering the human genetic puzzle. Today, a wide array of such models have been developed from every branch of life: prokaryotes (e.g. *Escherichia*), single-cell eukaryotes (e.g. *Saccharomyces*), plants (e.g. *Arabidopsis*), roundworms (e.g. *Caenorhabditis*), insects (e.g. *Drosophila*), fish (e.g. *Danio*), mammals (e.g. *Mus*), and so forth.

The first approach used to establish links between the genome and phenome was *forward genetics*, wherein a divergent phenotype is observed, then the causative gene identified (Fig. 1:1A). The earliest such studies, such as those on eye color in *Drosophila* [2], would perform multigenerational backcrossing for the "target" strain (e.g. a white-eyed fly) against a "reference" strain (e.g. a red-eyed fly). Through this approach, geneticists were able to determine the mechanics of genetic recombination, link causative genetic loci to phenotypic traits, and gradually develop linkage maps of the genome [5]. Decades later, following the development of mutagenic techniques, the field of *reverse genetics* arose (Fig. 1:1B), whereby a gene of interest could be rationally modified, and the associated phenotypic change(s) sought out. This approach is today driven largely by gain-of-function or loss-of-function (G/LOF) models, and has been indispensible in the mechanistic delineation of many molecular pathways. These two approaches still serve today as the fundamental basis for genetics studies: "top down" phenotype-to-genes forward genetics, and "bottom up" gene-to-phenotypes reverse genetics.

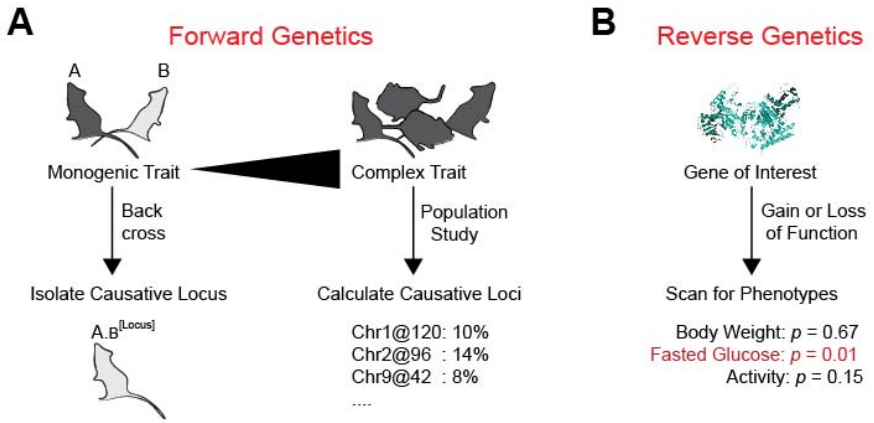


Figure 1:1 **Principal Genetic Approaches**

Summary of genetics approaches. **(A)** The forward genetics approach relies on identifying divergent phenotypes between two groups and then searching for the causative gene. The causative gene region, called the locus (plural loci), is identified either through backcrossing or through QTL or SNP mapping analysis. Backcrossing can also be performed in population studies in GRPs, but is much slower and typically no longer used. **(B)** The reverse genetics approach relies on identifying a gene of interest and then scanning mutant versions of this gene for impact on traits, such as phenotypes or on gene expression for known targets.

As with the contrast between forward and reverse genetics, there is a similar polarity between the *systems genetics* (or also *holistic*) and *reductive* approaches (Fig. 1:2A–B). Until recently, most studies in model organisms relied on the

reductive approach, which aims to understand biology by focusing on genetic or environmental variants one at a time. This technique has succeeded in identifying many genes and environmental effects that influence complex phenotypes, such as the influence of AP-1 and C/EBP [6] or of the E2F family of transcription factors on adipogenesis [7]. Such results were built upon painstakingly over decades, and thousands of genetic interactions have now been proven by this approach. Thanks to such efforts today hundreds of biological pathways are well-detailed [8]. Despite such achievements, few medical advances have yet come out of this knowledge. This is due in large part because the most common metabolic disorders are driven by dozens of gene-by-environment (GXE) variants in single individuals, i.e. the actions of single genes or even pathways are not primarily additive [9, 10]. While the reductionist approach can explain how one gene affects several diverse phenotypes (pleiotropy), it is less inefficient for examining how many genes interact to affect one phenotype (epistasis). Furthermore, the most common reductionist approach typically relies on G/LOF mutants such as knockouts, which are a poor model for the occurrence of common complex diseases in humans [11], and moreover such studies are typically performed on only a single genetic background (e.g. the C57BL/6J mouse).

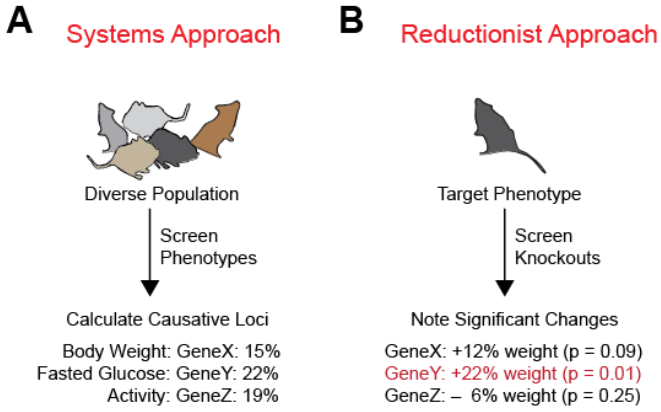


Figure 1:2 **Principal Study Designs**

(A) The systems genetics approach relies on using genetically diverse populations, phenotyping, genotyping, and acquiring other information on the cohort, and then identifying genes or loci, which contribute to variance in the trait(s) of interest. **(B)** The reductionist approach typically already has a single phenotype of interest, then mutants (e.g. knockouts, ENU mutants) are screened to observe which knockouts influence the phenotype. These diagrams are simplified, and while they describe the approach used in the majority of studies, these are not the only possibilities (e.g. please read [12] to view alternative approaches).

Such limitations to the reductive approach have been long recognized [13], but a systems genetic approach using genetic reference populations (GRPs) was long impeded by technical hurdles which have only recently begun to clear, stemming initially from improvements in "omics" technologies, namely genomics [14] and transcriptomics [15]. With this approach, completely novel regulatory genes or GXE effects can be identified *de novo* in the context of the whole organism through the use of quantitative trait locus (QTL) mapping or genome-wide association studies (GWAS). Furthermore, systems genetics analyses can identify entire sets of regulatory genes affecting complex phenotypes (i.e. gene-by-gene (GXG) interaction, called epistasis) as well as GXE effects in a single study [16-18], rather than the singular results from the reductive approach. While mechanistic validations are more difficult to perform in populations, doing so is not impossible, and today due to the increase in omics datasets, hypotheses can frequently be immediately tested *in silico* by analyzing similar, but independent, populations (e.g. from GEO [19]). Alternately—or subsequently—these hypotheses can be fed back into reductive approach for more detailed mechanistic examination, even for experiments on epistasis [20].

In tandem with research in model organisms, studies in human cohorts have also sought to identify novel genetic, environmental, and GXE risk factors for disease. In contrast to research in model organisms, human genetics has been driven primarily by the systems approach, although reductive studies on twins have also been extensively performed [21]. Early, large population studies such as the Framingham Heart Study successfully identified dozens of major environmental factors influencing complex traits and led to a number of major improvements in health, e.g. by linking

smoking to atherosclerosis [22]. However, while heritability in complex traits could be calculated in these studies [23], the underlying variants driving the inheritance could not typically be identified, largely due to an inability to obtain extensive sequence data on the patients. As with GRPs, this changed in the early 2000s with the development and application of improved sequencing technology, and briefly led to high hopes of personalized medicine, whereby the genetic variants underlying disease could be directly identified and attributed in specific patients. Unfortunately, despite the identification of a few major SNP–phenotype connections [24-27], it was readily apparent that linkage studies using human populations alone are generally insufficient for the identification, validation, and molecular characterization of causal variants behind heritable traits [28, 29]. This difficulty stems from many factors, but chief among them that (a) these diseases typically have a long-term progression caused by myriad GXE factors, which are particularly difficult to control in human studies, (b) it is difficult to obtain tissue biopsies in humans, and thus the ongoing cellular and molecular processes cannot be readily or comprehensively followed, and (c) mechanistic interactions cannot necessarily be validated *in vitro*.

In this review, we will discuss how a systems genetics approach using GRPs provides a platform that may be used to identify novel GXE elements leading to disease and establish hypotheses for the underpinning molecular mechanisms, which may then be used to create detailed experimental designs for reductive validation and subsequently treatment of disease in human patients.

1.1 Development and Application of Systems Genetics Approaches

1.1.1 Model Organisms in Disease Research

By the early 20th century, the use of specific model organisms, primarily mice and *Drosophila*, had become commonplace in laboratory research, and today dozens of model organisms are used, depending on the situational needs for particular aspects of biology research. These models vary dramatically in their similarity to humans and in their research applications—from mice, to which ~99% of human protein-coding genes have approximate homologs [30], to fruit flies with ~77% [31], *C. elegans* with ~65% [32], yeast with ~58% [33], and *A. thaliana* with ~48% [34]. While mammals serve as the closest proxy to humans, the primary mechanisms of genetic action, and even many basic cellular phenotypes, such as mitochondrial function, are well conserved throughout evolution. In such cases, examining a simpler organism may provide hypotheses that can be far more efficiently validated, though it is essential to consider potential cross-species complementarity [35], even for highly complex neurological phenotypes, e.g. *Drosophila* or *C. elegans* are commonly used in research on Parkinson's disease [36, 37]. For basic knowledge of metabolic pathways—which are relatively well conserved to invertebrates—much of our understanding has stemmed from such “down-thinking” approaches, e.g. a significant amount of our insight of insulin signaling stems from *C. elegans* research [38-40]. By examining such models for analogs and mechanisms to human diseases, the primary genetic, environmental, and GXE regulators of complex phenotypes such as energy metabolism and immune response may be identified, then tested in human populations for ultimate validation.

1.1.2 Scientific Development of Genetic Models

While it had been recognized since the 1860s that studies in identical twins provide the ability to separate environmental from genetic effects, pioneering studies on model organisms had no such way to readily obtain such equivalents. Thus, the often striking variability between individuals resulted in experiments which were difficult to reliably reproduce. It was recognized that inbreeding would (mostly) eliminate this unpredictable genetic variation, and the success of the earliest such approaches led to a massive and gradual process that has adapted thousands of “wildtype” organisms into controlled inbred strains (Fig. 1:3A). For the most commonly used model organisms, dozens of independent derivations have been performed from populations collected across the world [41, 42]. Thus, studies using inbred strains can take advantage of both genetic homogeneity (within strains) or genetic diversity (across strains). While inbred strains are most commonly used for the reductive research strategies, there are an increasing number of studies, which use multiple different inbred strains in tandem to create diversity panels (e.g. [43]), akin to GWAS.

Alternatively, inbred lines may be intercrossed to directly observe trait inheritance and heritability. The first such applications of this method relied on scanning diverse populations of second-generation (F2) recombinant crosses between inbred parental lines (Fig. 1:3B). For complex traits, F2s will display a gradual range of phenotypes, typically exceeding the differences between the parents. The causal loci underlying these effects may either be defined through genotyping and QTL mapping, or the strains of interest can be successively phenotyped and backcrossed with one of the parental strains in order to narrow down the causal region(s). However, as the F2 populations themselves are non-inbred, genotyping must be performed on every individual—until recently an expensive and imprecise process. Furthermore, results can be difficult to reproduce, and GXE factors cannot be readily deciphered. Lastly, the QTLs observed in F2 populations are often quite broad; an average mouse F2 study using ~500 genotypes can expect QTLs of 20–30 megabases, corresponding to an average of 200-300 candidate genes [44]. While such QTLs can be narrowed down and validated by backcrossing, this requires 10+ generations of successive phenotyping and genotyping, a time consuming process even in species with short generation times.

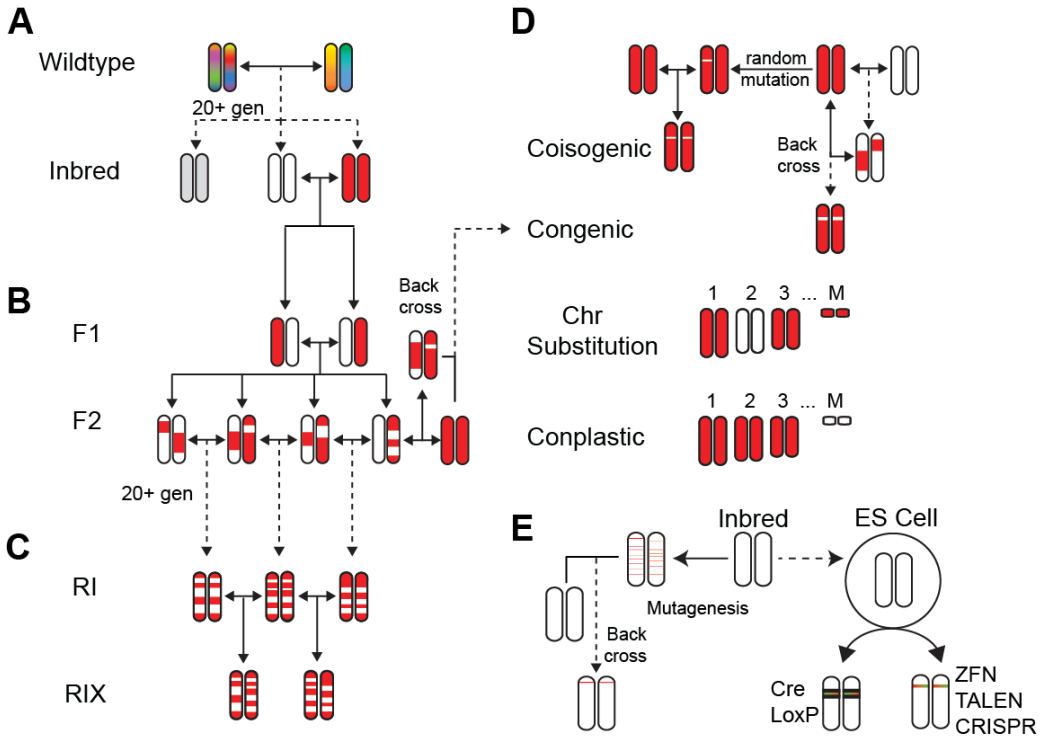


Figure 1:3 **Breeding Schematic for Common Types of Genetic Populations**

(A) Wildtype organisms can be collected in diverse locations, either randomly, or due to suspected environmental adaptations. After 20 generations of inbreeding, the lines are considered fully inbred with 99.5% homozygosity. **(B)** Inbred lines can then be crossed to form genetically diverse populations of F2 grandchildren, which can then be used for gene mapping studies. When a phenotype or QTL is located, the region of interest can be selected and backcrossed over 10+ generations for validation. **(C)** Alternatively, the F2 animals may be again inbred to form recombinant inbred (RI) lines, which provides biologically reproducible "clones" that may be used for more accurate QTL mapping. These RI lines may be crossed to F1s to create semi-heterozygotic recombinant intercrosses (RIX), or may even be outbred further to develop diversity outbred populations. **(D)** Alternative types of inbred strains exist, including a spectrum of similar approaches leading to coisogenic strains that have one or a few variants in only a single gene separating them from their "parent" line, congenic lines that have a larger gene-region separating the two lines, or chromosome substitution lines, with an entire chromosome replaced. Similarly, conplastic lines can be generated with the mitochondria of one inbred strain but the nuclear genome of another. **(E)** Inbred lines are also commonly rationally turned into G/LOF "strains", historically through chemical mutuagenesis, but now typically through ES cell modifications). Please note that this figure is a summary of the techniques most commonly applied in murine models, and is not exhaustively comprehensive.

Because of the shortcomings of F2s and backcrosses, development continued on recombinant GRPs, and in the late 1960s, the first recombinant inbred (RI) populations were developed (Fig. 1:3C) [45]. As with F2s, RI populations provide genetic variance, yet as each "individual" is an inbred strain, which can be controlled and replicated at will, outliers may be accounted for, heritability accurately calculated, and causative QTL regions calculated with higher precision (typically on the order of 1–5 megabases in mice [16, 46]). Furthermore, the fact that identical populations can be studied in multiple research groups provides economics of scale: higher quality sequencing/genotyping can be performed, more tissues can be studied, more phenotypes examined, better molecular records assembled, and so forth. Consequently, the RI concept has been implemented in a variety of organisms and has now been implemented in a wide variety of organisms, e.g. corn [47], rats [48], mice [49], *Arabidopsis* [50], yeast [51], *C. elegans* [52], and *Drosophila* [53]. More recently, recombinant intercross (RIX) lines have been developed, which are the F1 progeny of two RI lines [54]. These animals are semi-heterozygous, and can thus be examined for dominant and recessive traits, yet are also fully biologically reproducible, solving weaknesses of both F2 and RI GRPs.

A variety of other models are available as well (Fig. 1:3D), including coisogenics, which differ from their cousin strain by only a single point mutation, congenics, which differ by a somewhat larger genetic region, chromosome substitution strains with full chromosome(s) changed [55], and conplastic strains with mitochondria from one inbred strain and the nuclear genome of another. Finally, it is worth recalling that not even fully inbred organisms are truly genetically identical, due to differences in epigenetic modifications, gut microbiota, and the thousands of spontaneous mutations that arise over an individual's lifetime. These changes can be transmitted to progeny, and have led inbred strains to diverge over time when kept in separate facilities unless special attention is taken. While such spontaneous mutations have in some cases led to the fortuitous discovery of gene–phenotype links, such as the link between leptin and obesity [56] or between *Nnt* and glucose levels [57]. However, in general this genetic drift is an undesirable experimental variable.

1.1.3 Reductionist Genetics : G/LOF Variants

In 1927, it was discovered that ionizing radiation can induce DNA changes [58], a finding followed by the first proven chemical mutagenic chemical agents [59]. While early mutagenesis studies were unable to target any specific gene, they paved the way for the rational capability to modify an organism's genes at will. By the late 1970s, genetic modification (GM) techniques had advanced such that specific genes could be effectively targeted [60], and for the first time the reverse genetics studies on gene-to-phenotype could be feasibly undertaken. This field rapidly expanded, leading first to developments allowing arbitrary genome modifications in cell lines [61], and soon after by the creation of the viable transgenic organisms via pronuclear microinjection of transgenes into fertilized germ line cells [62] (Fig. 1:3A and Fig. 1:4A). These developments, combined with the advances in genetic engineering, provided a stable platform for rationally modifying model organisms, and by the 1990s, G/LOF models were possible for nearly any gene in many model organisms, and with temporal (i.e. can be activated at a given time) or spatial (i.e. tissue-specific) precision [63-65]. This targeted GM approach has since become not only the standard method for validation of novel gene–phenotype links and for the in-depth mechanistic examination of metabolic and cellular pathways, but is also used to screen for the effects of genes by the reductive approach. Such screening programs predominantly rely on using LOF modifications of genes one-by-one on a single inbred strain background, then phenotyping each mutant strain and hoping the relevant phenotype was examined. Nearly all genes have been examined using this approach in *C. elegans* [66], *D. melanogaster* [67], *A. thaliana* [68], and *S. cerevisiae* [69].

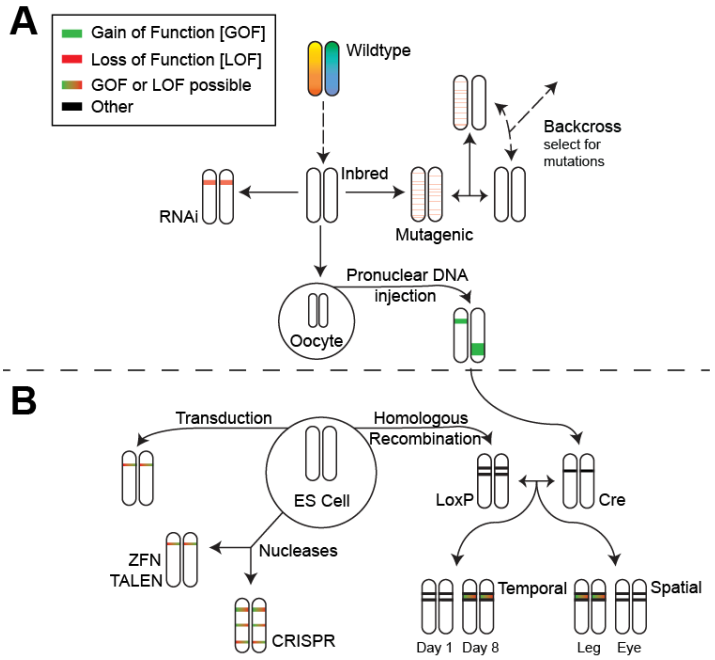


Figure 1:4 **Generation Schematic for Common Types of G/LOF Populations**

Preparing gain and loss of function model organisms usually starts with an inbred line. It is important to consider that most G/LOF models are done in one specific genotype in most model organisms. **(A)** Traditionally, inbred organisms were targeted by non-specific mutagens such as X-rays or the chemical mutagen ENU. The many resulting mutations would then be isolated to find the causative change by successive generations of backcrossing, with phenotyping performed at each generation. In *C. elegans*, RNAi treatment can be performed by direct feeding of bacterial clones expressing the RNAi; in other species, the technique requires genomic modifications. In the germ line, oocytes may be modified, e.g. by retroviruses or pronuclear injection, to include target DNA into an organism. **(B)** More precise techniques generally use ES cells or eggs as a platform providing improved reliability in complex organisms, e.g. mice, zebrafish, and *Drosophila*. These techniques rely on accurate targeting of a gene region, excision of the DNA, and then either insertion of new DNA, or proper repair of the broken ends. Several such tools are available, e.g. the older viral methods (adenoviruses and lentiviruses) or traditional homologous recombination as well as more modern nuclease-based techniques (e.g. ZFN, TALEN, CRISPR). Genetic modifications can all be applied in a highly specific (e.g. modifying a single phosphorylation site), spatially controlled (e.g. tissue-specific) or temporally controlled (e.g. drug activated) manner. For spatial and temporal control, several such systems are available, e.g. the Cre-lox and the FLP/FRT systems in mice and zebrafish, or the GAL4/UAS system in *Drosophila* [70].

However, the cost-benefit ratio of such hypothesis-null screening approaches came into question during the development of equivalent programs in mice, where even with a worldwide consortium, the entire genome cannot be feasibly screened (though more than half the genes are now available with targeted mouse ES cell lines [71-73]). Even in invertebrates, it is impossible to screen every gene for all major phenotypes of research interest. More importantly, it is clear that the reductive approach is ill-suited for decomposing complex traits into their major components [74], a difficulty arising from the geometrically increasing number of epistatic interactions, the pleiotropic effects of genes [13, 75], and the fact that even for Mendelian connections, gene action is not a simple on–off mechanism or even necessarily a linear response [76]. Consequently, most knockout consortiums have turned away from blind screening and moved towards serving as a resource platform for validation. This change has particularly been driven by the “omics” revolution, which has demonstrated the improved capability for GWAS and GRP studies to identify novel gene–phenotype links, which can then be examined with G/LOF models. The most recent developments in reductive genetics now allow relatively expedited modifications of several genes at once, such as through the TALEN or CRISPR-Cas9 systems (well covered in recent reviews such as [77] and [78], summarized in Fig. 1:4B), meaning even multi-gene networks identified through systems studies can be feasibly tested in a reductive setting.

1.1.4 Synthesis of Complex Trait Analysis

Recently, an increasing number of studies have been published using a hybrid approach combining systems genetics and reductionism to identify new regulators of complex metabolic processes and diseases [79, 80]. Furthermore, there is far more capability today to directly validate results in a cross-species approach [81-84]. Thanks to the above-described technical developments on model organisms, and the exponential increase in omics datasets and technologies, it is now feasible to combine large-scale genomic, transcriptomic, proteomic, metabolomic, and phenomic layers in a single study. In human studies too, it is becoming increasingly common to include metabolomic [85] and proteomic [86, 87] profiles, in addition to the SNP–transcriptome–phenotype layers that have become more common over the past decade. This multilayered approach has been recently shown for several complex diseases where novel genes and networks were found as causal: diabetes [18], obesity [88], general mitochondrial function [89], and drug response [90]. Lastly, even for the reductive approach, these technologies improve the ability to fully characterize the effects of modifications to target genes: i.e. there is less risk of an unknown/off-target pleiotropic effect to go missed when the full transcriptome is quantified.

Today, the large-scale availability of detailed omics datasets means that novel results can, in many cases, be validated completely *in silico*. For example, a recent observation linking variants in the gene *Dhtkd1* to 2-aminoadipate levels in the BXD mouse population was also observed in two independent mouse populations [18, 91], and then translated cross-species to humans using one of the dozens of large hospital studies [18, 85]. A similar, but converse, approach was also taken with 2-aminoadipate, where it was connected to glucose response in a human population study, then validated in mice [92]. To note, while these links were indeed highly significant in the original human data alone, the simple scale of omics datasets means that there is a frequently inseparable mixture of false positives and false negatives and such validation is necessary [93]. This noise particularly affects the study of complex traits, as the impact of any individual gene variant is often so small as to be almost negligible on the population scale [94]. Thus, the overlaid consistency across multiple populations brings noisy connections into focus [18, 95], or can indicate differences between populations [96, 97].

Once such links have been confirmed independently (and ideally translated to humans), it becomes of strong interest to examine the detailed molecular action of these new targets. The original datasets can then be used again to generate further hypotheses to be applied with the reductive approach—thus a factor which is known to be of interest, can be gradually built upon to define a *why* and *how* link to the phenotype. In these cases, G/LOF models provide a level of control not possible in population genetics, and minute perturbations can be made to assess and monitor every step of the pathway [98]. Thus, the reductionist and systems approaches complement, rather than compete, with one another.

1.1.5 Developments on the Horizon

Despite the tremendous gains in our ability to quantify and understand biological mechanisms, we remain far from obtaining a complete picture of physiology. Most omics datasets only make use of the genome and transcriptome, as the proteome, metabolome, epigenome, and other aspects of cellular systems can only be partially quantified and only by specialized research groups. Furthermore, while an individual's DNA is relatively fixed from cell-to-cell and day-to-day, nearly all other aspects of biology vary dramatically from cell to tissue and across time. Single-cell sequencing promises to address the cell-to-cell disparity of measurements [99], while fluorphoric techniques promise the ability to address day-to-day differences—at least in cell lines [100]. Other developments, such as SWATH proteomics [101], promise to expand the number of omics “layers” that may be examined. Unfortunately, these upcoming technical developments—and even more prospective theoretical technologies—do not herald a capability to bypass model research and fully characterize complex diseases directly in humans. Model organisms retain the capability of much stronger environmental control, controlled genetic variability, faster generation time and hypothesis testing, far more easily obtained tissue samples, biological replication of results, a better capability to examine GXE factors, and a safer way to examine potential treatments. Ultimately, however, the goal of disease research is to confidently define major risk factors in humans, provide early diagnosis and monitoring of disease progression, and establish effective treatments. It is increasingly clear that large population studies in model organisms will serve far beyond their initial scope, just as just as the Framingham Heart Study continues to yield results more than 70 years after its debut.

1.2 Perspectives on Holistic and Redudctive Genetics

The wealth of genetic information painstakingly generated over the past century, along with associated technological developments, has resulted in an increasing variety of viable scientific approaches towards understanding complex traits: through reverse and forward genetics, and through reductionist and systems designs. Both approaches necessitated the generation of a tremendous range of technological advances in order to better understand and explain the fundamental relationships between genotype and phenotype, and extensive development of model organisms. In the past thirty years, the reductive approach has become nearly perfected, with an ability to apply precise genetic modifications *in vitro* and *in vivo*, allowing the exact assessment of molecular reactions driving many biological pathways. Conversely, it is only within the past decade that the systems genetics approach became truly viable for the decomposition of complex genetic systems. This capability has come in part due to new population models, but moreover because the widespread use of transcriptomics, genomics, and a rising capability in metabolomics and proteomics has allows unprecedented understanding of the totality of an organism's biological machinery and thus a corresponding increase in our understanding of molecular interactions at the cellular level.

Despite these advances, our ability to predict or prevent the incidence of complex (and common) diseases such as cancer, diabetes, and atherosclerosis remains limited except for the most striking environmental influences (e.g. tobacco) or genetic variants with major impacts (e.g. leptin deficiency). These failures had led to a crossroads where many argue for one research approach by pointing to its merits, and to the deficiencies of other methods, and many such criticisms are valid. Within the context of reductionist approaches, it is clearly infeasible to combine genome wide G/LOF strategies with exhaustive phenotyping, even under “basal” non-challenged conditions for a simple model organism such as *C. elegans*, much less for the mouse. Likewise, systems genetics studies are frequently plagued by an excess of nominally significant findings, too many potentially causative variants, and/or an inability to convincingly pinpoint the genetic causation. While further technical developments will assist in the deconvolution of complex traits, the fundamental differences between reverse and forward, systems and reductive techniques can already be considered complementary with their reciprocal strengths rather than as competitors. This synergy, for example, allows general phenotype screening platforms to provide hypotheses for G/LOF experiments, providing a full understanding from the complex organism phenotype down to the molecular interactions at the cell level. The results from such hybrid approaches have begun to increasingly provide biologists and clinicians with an understanding of many fundamental relationships and interactions between genetic inheritance and environmental influences.

In this review, we have focused on the fundamental applications of population studies for the deconvolution of complex traits, yet this is only the first step towards improving human health. Gains in the understanding of molecular mechanisms underpinning diseases must be followed by new therapeutic strategies, accurate suggestions for lifestyle changes, and must be accurately tailored to the needs of each individual based on the minimal information obtainable for most human subjects (e.g. blood, urine, and DNA). The variation of individual susceptibility to illnesses is becoming increasingly recognized and tested, yet also creates complexities which must also be addressed in the development of new therapeutic agents; if a complex disease can be caused by ten independent mechanisms, it is unlikely for a single drug to treat all patients. While this situation is increasingly realized in human trials for cancer therapeutics [102], the preclinical success of many compounds is hinged largely on whether the compound works in a single genetic background of cells *in vitro* or animals *in vivo*. Doing so is risky: for example, had early researchers exclusively used the DBA/2J mouse strain for drug screening, they likely would have concluded that morphine is a non-addictive and inefficient painkiller [103]! Examining even a handful of diverse cell lines or inbred strains would allow a basic characterization of genetic diversity in drug response and inform greatly for early clinical trials. Combining the reductive and systems approaches will not allow identification of all the intricacies of complex genetic diseases, nor will they provide a silver bullet for rational drug development. However, this approach will allow further insights into complex traits, leading not only to additional potential drug targets, but an improved capability to understand the development of disease states and thus earlier diagnosis, better monitoring, and eventually prevention. With currently available tools and technology, complex mechanisms may now be rationally simplified to allow for more rapid and accurate identification of novel genetic or GXE factors that lead to disease and provide a more effective bridge from studies in model organisms to the clinic.

Chapter 2 Reducing Systems Genetics

Section 2.1 is adapted from Andreux P.A.*, **Williams E.G.*,** Koutnikova H., Houtkooper R.H., Champy M.F., Henry H., Schoonjans K., Williams R.W., and Auwerx, J. Systems genetics of metabolism : the use of the BXD murine reference panel for multiscalar integration of traits, *Cell*, 2012. (* : Co-first author with Andreux.)

Section 2.2 is adapted from **Williams, E.G.,** Mouchiroud L., Frochaux M., Pandey A., Andreux P.A., Deplancke B., and Auwerx J. A conserved role for the aryl hydrocarbon receptor in the regulation of movement, *PLoS Genetics*, 2014.

Over the past decades, study designs have been optimized to analyze genetic factors in large populations of naturally-divergent strains, chiefly in *Mus musculus*, *S. cerevisiae* [104, 105], *Drosophila melanogaster* [106], and more recently, *C. elegans* [107]. Murine genetic reference populations (GRPs) are among the best-established mammalian models with which to study gene-by-environment (GXE) interactions. These GRPs are typically sets of inbred strains which have been assembled to incorporate carefully titered levels of genetic complexity that model aspects of human populations. The recombinant inbred (RI) strain families enable tight experimental control while each genotype is represented by an entire isogenic line, thereby enabling extensive replication studies [108, 109]. The BXD family, currently the largest and best characterized mouse GRP, is composed of ~160 lines which descend from crosses between C57BL/6J (B) and DBA/2J (D) [110]. GRPs such as the BXDs have been bred for quantitative trait loci (QTL) analyses, which are statistical genetic techniques that define regions of the genome (intervals or loci) and their modulating effects on phenotype. The fixed genotypes of each strain can be combined with full sequence data of the parental strains to simplify QTL mapping and in some cases specifically identify causal sequence variants [111, 112]. Furthermore, owing to the relatively fixed genotypes of RI populations, massive databases of phenotypes and expression data can be assembled and shared across time, allowing for rapid multiscalar analyses. Over the last two decades the BXD family has been exploited mainly to study the genetics of immune function and infectious disease [113, 114] and in behavioral and neuropharmacological research [64, 115-117]. However, few metabolic phenotypes have been previously generated. In the present metabolic survey, we systematically generated quantitative data for 140 standardized phenotypes, including glucose response, body weight change, physical activity, and oxygen consumption across a large subset of the BXD family using adult males and females.

Energy homeostasis in metabolism is the result of a tight balance between energy intake and expenditure. Metabolic disorders, such as obesity and type 2 diabetes, often result when this equilibrium is disturbed by complex interactions between genetic and environmental factors [118]. Predisposition to complex diseases such as the metabolic syndrome is inherited in a non-Mendelian fashion, emphasizing genetic heterogeneity and complex GXE in pathogenesis. Genetically engineered mouse models are not ideal for dissecting polygenic networks or GXE interactions precisely because they have been optimized to study actions of single genes on single genetic backgrounds [72]. In contrast, studies in humans have identified risk factors for developing metabolic diseases with both environmental (e.g. lack of exercise) and genetic causes (e.g. mutations in the *FTO* locus [119]), but these studies typically fall short of defining GXE due to an inability to control environmental influences, cohort and admixture effects, difficulty in obtaining certain types of physiological and molecular data, and the inability to sample many individuals with identical genomes under different conditions. Effective population-based experimental methods to dissect intricate GXE effects, such as the BXDs, are needed to model complex genetically admixed human populations.

All data are publicly available as a resource to the scientific community and complement massive expression data for key cells, tissues, and organs that we and others have deposited in the GeneNetwork database (**www.genenetwork.org**). These essential baseline clinical phenotypes are widely variable, often highly heritable, and in many cases can be linked to genetic loci encompassing known and novel candidate genes. Heritability and sex effects were highly variable among traits. Variation in one key trait, alkaline phosphatase, was linked to hypophosphatasia in lines containing the C57BL/6 allelic variant. Downstream of this hypophosphatasia, differences in bone structure were noted, as well as a significant increase in pyridoxal-phosphate (PLP) levels and a corresponding decrease in pyrixodal levels. In the same study, we also observed major differences in glucose homeostasis and response to a glucose tolerance test across the BXDs, and across male and female mice. However, in both cases, a few key genes were observed as

responsive for different stages of glucose response. In a followup study (section 2.2), we examined how spontaneous activity varies across the population, and we uncovered a single gene, the aryl hydrocarbon receptor (*Ahr*) which is responsable for a major portion of this variability. These studies validate the use of the BXD GRP as a powerful resource to study metabolic homeostasis and identifies new candidate genes for disease, and highlight the importance of considering males and females separately in the study of most metabolic traits.

2.1 Sexual Dimorphism in General Metabolic Traits

2.1.1 Optimization of Study Design for Genetics of Metabolism

Mendelian traits such as coat color can be accurately measured with a sample size of one per genotype. In contrast, complex metabolic traits with lower heritability benefit greatly from larger sample sizes per genotype. One of the advantages of GRPs over non-inbred yet genetically diverse strain populations (such as an F2 intercross) is that the heritability of a trait can be accurately measured and effectively boosted by repeatedly resampling cohorts of isogenic individuals of the same genotype and sex (Fig. 2:1A). The observed variability within isogenic strains is thus primarily environmental, whereas differences among strains are largely genetic. A first step in experimental designs using GRPs is to estimate heritability and the sample size required to accurately measure a phenotype in a specific environment. Such calculations provide both a justifiable number of replicates per genotype and the total numbers of genotypes needed to reliably discriminate gene loci using QTL mapping methods. When traits have high heritability, increasing the number of genotypes (i.e. strains) greatly improves the power and precision with which QTLs can be mapped, while increasing the number of cases per genotype has almost no impact (Fig. 2:1B, top). In the case of a Mendelian trait such as coat color, mutations in the brown (*Tyrp1*) and dilute (*Myo5a*) genes explain nearly all variation in pigmentation in the BXD family and map with peak LOD scores above 25 at these genes [120]. In contrast, complex diseases, including virtually all metabolic disorders, are polygenic, more dependent on environmental influences, and typically have a nearly normal distribution across genotypes. Power analysis shows that a study using 20 BXD strains with 4 animals per genotype will often detect loci that explain 50% or more of the genetic variance of a moderately heritable trait, while a study using ~40 strains will typically detect loci responsible for ~30% of the variance (Fig. 2:1B, bottom).

We phenotyped a set of 43 strains—24 strains with male and female animals, 18 strains only with males, and 1 strain only with females (Fig. 2:1C). Phenotyping was initiated at 13 wks of age and all cases underwent six testing categories (Fig. 2:1C). Mapping power was estimated as a function of heritability and sample size. We calculated heritability as 1 minus the ratio of the within-strain variance to total sample variance (Fig. 2:1D). Most parameters have moderate to high heritability (h^2), averaging 0.67. The more heritable a trait, the higher the likelihood it will map to significant QTLs [121]. Under the tightly controlled environmental conditions in our vivarium, both glucose tolerance and blood pressure had h^2 ~0.7, indicating ~70% of the variance in these traits could be explained by genetic factors. Consistent with these moderately high values, each of these traits was associated with at least one significant locus. In the case of blood pressure, we have defined a strong candidate gene—*UBP1*—using a human genome-wide association study (GWAS) as a cross validation cohort [81]. The BXD GRP allows complementary forward and reverse genetic methods to analyze metabolic traits (Fig. 2:1E). Conventional forward mapping methods start with heritable differences in phenotype and define loci and sequence variants (Fig. 2:1E). Reverse genetics starts from defined candidate genes and identifies downstream phenotypic effects (Fig. 2:1E) [92, 122]. Because full sequence data are available for both parental strains of the BXD family, reverse methods can now be applied genome-wide using a standard mapping population. In this resource manuscript, we demonstrate both approaches to the data.

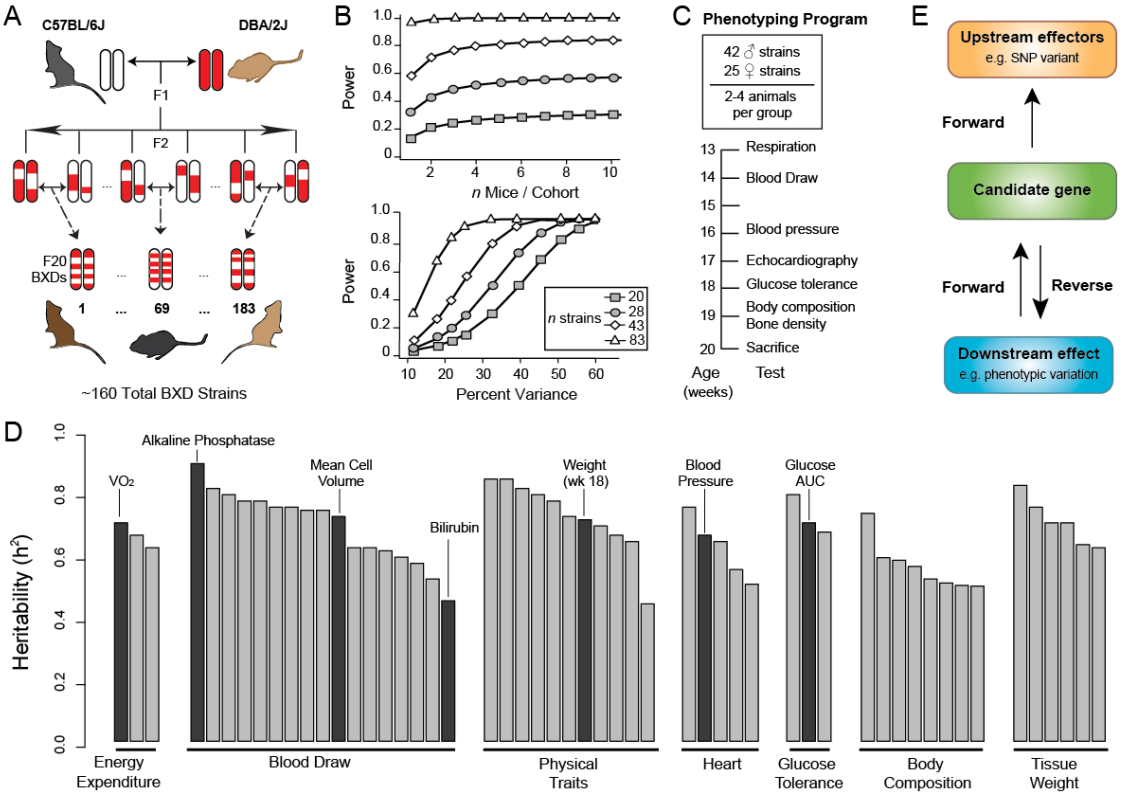


Figure 2:1 **Study Design for Identifying Genetic Loci Which Control Metabolic Traits**

(A) The BXD lines were created by crossing B and D parents. The resulting heterozygous F1 mice were again crossed to generate genetically diverse but non-reproducible F2 animals. These F2 animals were inbred until generation F20+, at which point the genome was 99.5+% isogenic and the strains are considered fully inbred and together constitute a GRP. The ~160 BXD strains are numbered 1–183. **(B)** Top panel: power depends mainly on strain number instead on number of mice/strain. Bottom panel: the power to detect a given fraction of variance with a single QTL. Power was calculated with cohort size fixed at 4. Both calculations were done for a trait with heritability of 0.67. **(C)** All animals underwent the same phenotyping programs as specified in the result section. The age corresponds to the timing of the phenotype experiment. BXD60 was only phenotyped in females—all other female strains overlapped with male strains. **(D)** Heritability of select traits from each phenotyping group. Traits discussed in-depth in this resource paper are indicated in black. Traits are grouped according to phenotyping test. "Body Composition" represents echocardiography phenotypes, bone size, bone density, and body weight. **(E)** The analysis flow-chart. The cause-to-consequence effect reads from top to bottom. A SNP induces a change in the expression or function of a candidate gene, which has an impact on a given downstream effect or phenotype.

2.1.2 The BXD Family Has Wide Variation in Metabolic Traits

Metabolic traits vary to a remarkable degree among the BXD strains, with every trait showing significant variation (Fig. 2:2A). The least variable trait, basal body temperature, varied significantly (from 35.8±0.3 °C in BXD34 to 38.9±0.2 °C in BXD95) despite an overall variation of less than 1.1 fold, and the most variable trait, alkaline phosphatase (ALPL) protein level, varied 6.5-fold between BXD31 (44±11 U/L) and BXD90 (287±18 U/L). For the majority of traits, the distribution of phenotypes is continuous and approximately normal in both sexes.

As expected, sex-linked differences are a significant source of variance [123], with the more variable traits somewhat more likely to exhibit significant sexual dimorphism (Fig. 2:2B), although nearly 60% of metabolic traits had no significant differences between the sexes. These metabolic traits can thus be divided into four categories in terms of the significance of variation and correlations between sexes. The first separation considers whether the trait is variant between the sexes, while the second separation considers traits which do not correlate between the male and female trait values (e.g. shortening fraction in Fig. 2:2C), and then those which are correlated (e.g. red blood cell volume in Fig.

2:2D). The first category, for traits with no significant sex differences but also without strong correlation between males and females, indicates a phenotype where variation cannot be reliably attributed to genetic factors (i.e. variation is due to environmental or technical influence) or that there is an inconsistent sex-by-strain interaction effect. In the case of the shortening fraction, we suspect the environmental or technical noise as the cause of this discrepancy due to the relatively low heritability (h^2 = 0.5) of this trait (Fig. 2:2C). In contrast, the traits with strong correlations between sexes indicate that the genetic variability is not affected greatly by the environment, and thus the slope of the correlation is close to 1 (Fig. 2:2D).

The remaining 40% of traits exhibit significant differences between males and females, and again thse traits can be with or without strong male-female correlation. One such striking example is body weight, where there are dramatic differences between males and females, but again without consistency across strains: e.g. BXD75 males are the leanest, whereas the BXD75 females are in the middle range of body weights (Fig. 2:2E). Lastly, a few traits had both significant sex-associated differences, but also strong genetic correlation, indicating both genetic and sex-by-strain interaction effects. One such trait, alkaline phosphatase, will be discussed further in Fig. 2:4.

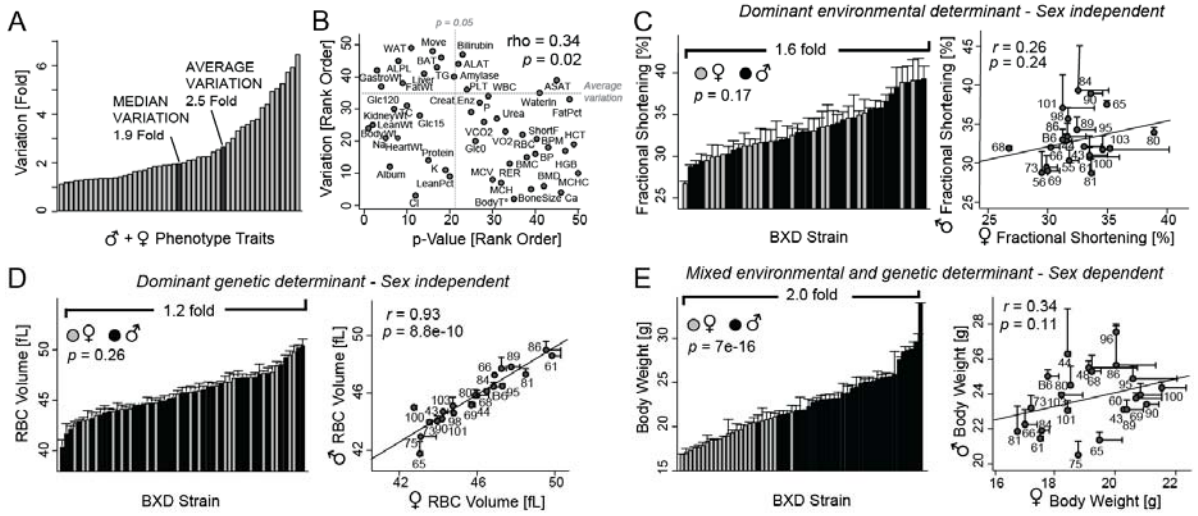


Figure 2:2 **Strain and Sex Influences Differ Widely Across Metabolic Parameters**

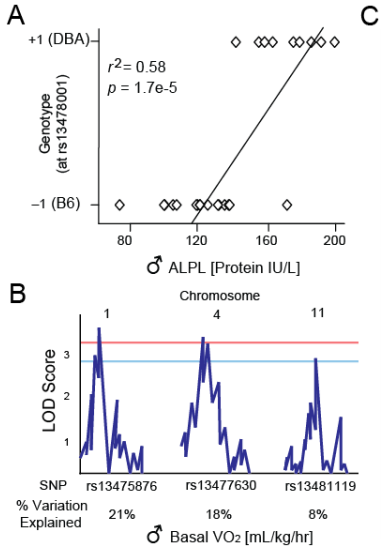
(A) Metabolic phenotype variation for all cohorts. **(B)** Traits with higher variation are more likely to have significant differences between males and females; 41% of traits are significantly different (traits on the *x*-axis to the left of rank 22). **(C)** Example of a sex-independent trait that is mostly determined by environment: the ventricular shortening fraction. **(D)** Example of a sex-independent trait that is mostly determined by genetics: the mean cell volume of red blood cells. **(E)** Example of a sex-influenced trait that is explained by a mix of genetic and environmental factors: body weight at 18 wks.

2.1.3 QTLs Link Phenotypes to Causal Loci

The data sets generated in this population consist of both general blood parameters, such as hematocrit and iron levels, and physical phenotypes, such as heart rate and oxygen consumption. Suggestive or significant QTLs were detected for nearly 40% of all phenotypes. Traits with the highest heritability and most significant QTLs usually mapped to a single locus. This type of trait is exemplified by the tissue-nonspecific (liver/bone/kidney) alkaline phosphatase (hereafter termed ALPL) protein activity, a trait for which 58% of the variance is explained by a single QTL (Fig. 2:3A). Many other traits, particularly complex phenotypes such as respiration, map to several differnt loci. In these cases, polygenic networks of unlinked loci influence trait variance. One such example is basal oxygen consumption (VO_2), a trait for which three loci independently contribute to phenotypic variation—21% for the most significant locus on Chr1, 18% for the locus on Chr4, and 8% for the locus on Chr11 (Fig. 2:3B).

Genes located within all QTLs were first examined for known links to function using PubMed and GeneRIF [124]. All but one of the top 10 QTLs contained at least one strong positional and functional candidate (Fig. 2:3C). In some cases, the positional candidate is well-established in literature as the causative gene behind the QTL, e.g. the QTL for hematocrit levels maps to a locus containing the hemoglobin gene [125]. In other cases, the candidates are linked more tenta-

tively. For example, the QTL for bone surface area contains the steroid 5 alpha-reductase 1 (*Srd5a1*) gene, which has been linked to bone mass in a recent study on *Srd5a1* knockout mice [126], while other phenotypes were not clearly linked to any candidate gene, suggesting that the QTL is either completely novel, or not well studied.



C

Trait Description	LOD	p-val	Chr@Mb	Possible Candidate
Alkaline Phosphatase (♂)	6.8	1.6e-5	Chr4@131.5	*Alpl* (Foreman et al., 2005)
Alkaline Phosphatase (♀)	6.6	2.3e-5	Chr4@139.3	
Blood Cell Volume (♂)	6.1	7.3e-5	Chr7@114.5	*Hbb (Popp, 1962)*
Blood Cell Volume (♀)	5.5	3.3e-4	Chr7@114.5	
White Blood Cell Count (♀)	5.5	3.3e-4	Chr10@120.1	*Irak-M* (Xie et al., 2007)
Hemoglobin Concentration (♂)	5.1	8.0e-4	Chr7@111.3	*Hbb (Popp, 1962)*
Hemoglobin Concentration (♀)	4.0	9.2e-3	Chr7@114.5	
CO2 Production (♂)	4.8	1.7e-3	Chr1@57.9	*Aox1 (Weigert et al., 2008)*
Hematocrit (♂)	4.8	3.1e-3	Chr7@114.5	*Hbb (Popp, 1962)*
Systolic Blood Pressure (♂)	4.7	3.2e-3	Chr9@113.2	*Ubp1* (Koutnikova et al., 2009)
Heart Rate (♀)	4.5	3.7e-3	Chr19@33.0	*Pten* (Zu et al., 2011)
Bone Surface Area (♀)	4.4	3.8e-3	Chr13@71.1	*Srd5a1* (Windahl et al., 2011)
Fasting Glucose (♀)	4.3	5.1e-3	Chr6@149.0	No clear candidates
RER (♀)	4.3	5.6e-3	Chr9@83.6	*Bckdh* (Joshi et al., 2006)

Figure 2:3 **Analysis of Peak Trait QTLs**

(A) The genetic variation attributable to a QTL is calculated by the correlation between the trait values (here, ALPL levels) and the genotype at the peak location. ALPL has only one QTL, which explains 58% of the genetic variation. **(B)** The genetic variation attributable to each of the three QTLs mapped to oxygen consumption. Each QTL has a smaller effect than for ALPL, but together account for a large amount of variation. Generalized linear modeling or ANOVA (used here) is necessary to calculate the variation independently attributable to each QTL for traits which map to multiple QTLs. **(C)** The LOD score, corrected *p*-value, and QTL peak location are given along with the trait for the top 10 distinct phenotypes. Of note, this is not all of the significant or suggestive peaks for each trait; only the most significant peak is listed. One positional candidate is given for traits when a literature analysis of every gene under each peak has yielded a linked result. Candidates not discussed in the text include the link between white blood cells and *Irak-M* [127], *Pten* and heart rate [128], and RER in females and *Bckdh* [16]. Systolic BP LOD score is for analysis as described in [81].

2.1.4 ALPL Variants Contribute to Changes in Bone and Vitamin B_6 Homeostasis

Serum ALPL levels are strongly variable among BXD strains, with females having significantly higher expression (but less variation) than males (Fig. 2:4A). Expression of *Alpl* mRNA in the liver is also highly variable, but does not differ by sex (Fig. 2:4A) [129]. Interestingly, both *Alpl* mRNA and ALPL protein activity are highly correlated by strain, regardless of the sex of the animal (Fig. 2:4B), suggesting strong genetic modulation and a sex effect shared by all strains. Data for ALPL activity in males and females mapped to a common QTL: an 11 Mb region of Chr4 (Fig. 2:4C) that has already been reported in an F2 cross between C57BL/6 (*B*) and DBA/2 (*D*) mice [130]. *Alpl* mRNA also mapped to the same region (Fig. 2:4C). This QTL harbors ~100 genes, including the *Alpl* gene itself. We hypothesized that differences in ALPL activity are primarily and perhaps almost exclusively caused by one or more allelic differences in the cognate *Alpl* gene, i.e that this is a so-called *cis*-regulated expression QTL (*cis*-eQTL). Rapid identification of all sequence variants in the region was performed using a mouse SNP database (**www.genenetwork.org**). Of 11 exonic SNPs that differentiate *B* and *D* haplotypes of *Alpl*, nine cause synonymous mutations and two cause missense mutations in exon 9, resulting in amino acid changes of R318Q and L324P (Fig. 2:4E). We then modeled the impact of these mutations on ALPL protein structure using Phyre2 (**www.sbg.bio.ic.ac.uk/phyre2/**) (Fig. 2:4D). Although these mutations are not located in the active site of ALPL (around S110) [131], they are predicted to change the β-sheet/α-helix structure (R318Q and L324P are in green in Fig. 2:4D, showing the described excerpt of ALPL ± 21 amino acids). Remarkably, mouse and human ALPL protein sequences share 93%/97% identity/homology. In particular, the region harboring the two SNPs is highly conserved in vertebrates—notably proline 324 is invariant in other vertebrates, except in C57BL/6 where it has mutated into leucine

(Fig. 2:4E, same excerpt as in Fig. 2:4D). Moreover, the high degree of conservation of the ALPL protein is also revealed by phylogenetic tree analysis—human ALPL has 77%/87% sequence identity/homology with zebrafish ALPL, 44%/61% with *D. melanogaster*, and 32%/45% with *E. coli* (Fig. 2:4E). ALPL orthologs separate into three clusters: vertebrates, insects, and bacteria (Fig. 2:4F). Given this genetic conservation, it is likely ALPL has functions in vertebrates that are wholly unrelated to bone structure.

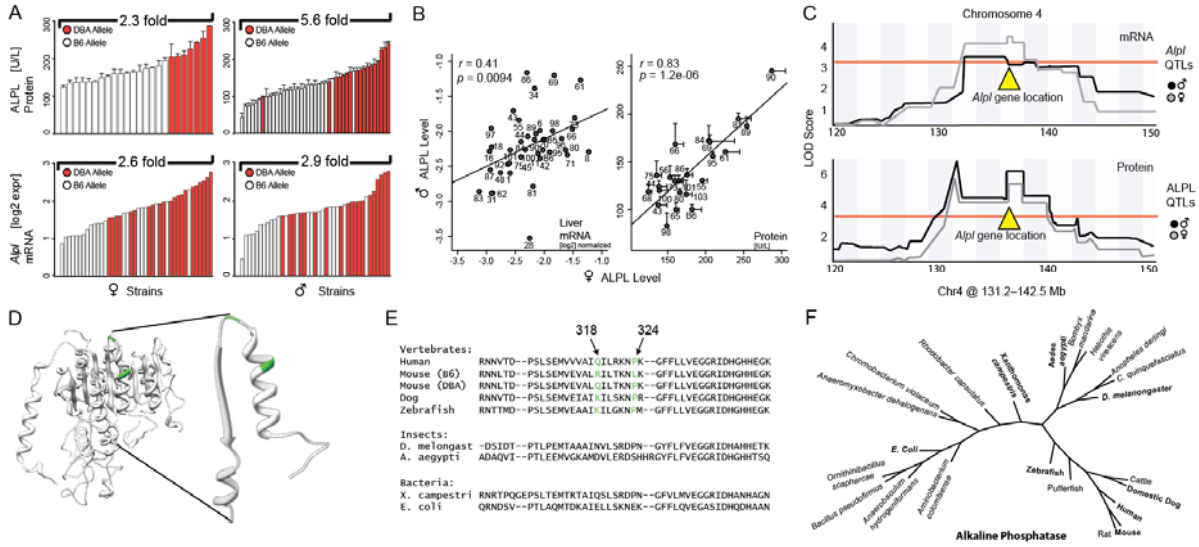


Figure 2:4 **Reducing ALPL to the Causative Allelic Variant**

(A) Expression variation of serum ALPL activity and mRNA levels from liver transcriptome data. Strains with the *D* allele (red) have significantly more ALPL protein and *Alpl* mRNA than strains with the *B* allele (white) for all comparisons ($p < 0.005$). Males have significantly less ALPL protein than females, but there is no difference in mRNA levels. **(B)** Correlation between male and female *Alpl* mRNA and ALPL protein levels demonstrate high trait heritability (Fig. 2:1D). **(C)** QTL map for liver *Alpl* mRNA and blood ALPL levels in males and females. All four QTLs (male and female traits are mapped separately) mapped to the same location and crossed the significance threshold of $p < 0.05$, indicated by the red horizontal line. The *Alpl* gene is located below the rectangular peak at right. **(D)** 3D model of ALPL from Phyre2, based on the crystal structure of the placental form of alkaline phosphatase. Enlarged on the right are the structure of the amino acids flanking and including the two missense mutations (noted in green). **(E)** 44 amino acid excerpt of sequence comparison of orthologs of ALPL, the amino acid sequence of the region modeled in panel D is shown here. The full protein sequence is highly conserved in mammals (>90% homology) and moderately from mammals to bacteria (>30% homology). **(F)** Phylogenetic tree of *Alpl*. The gene has orthologs in all known species, notably including invertebrates.

In humans, mutations in the *ALPL* gene can cause varying degrees of hypophosphatasia [132], a disorder nearly always associated with the accumulation of phosphoethanolamine (PEA) in the urine, extracellular pyridoxal-5'-phosphate (PLP) in the blood, and extracellular inorganic pyrophosphate (PP_i) in urine and blood. Downstream of these biomarkers are many pathways involved in bone remodeling [133], amino acid metabolism, and neurotransmitter regulation [134]. Therefore, we examined patterns of segregation of phenotypes among BXDs by testing several strains with either the *B* allele (C57BL/6, BXD48, BXD65, BXD68, BXD75, BXD103) or the *D* allele of *Alpl* (DBA/2, BXD89, BXD90, BXD95). We first reconfirmed that ALPL enzymatic activity was different among strains carrying the alleles (Fig. 2:5A). As predicted, extracellular PLP levels were strongly negatively correlated with ALPL in both sexes, supporting the hypothesis that the BXDs are a good model for hypophosphatasia (Fig. 2:5B). Furthermore, ALPL levels did not correlate with levels of calcium or phosphate (P_i), consistent with studies in humans differentiating hypophosphatasia from rickets and osteomalacia [133]. Femur bone structure evaluated via X-ray microtomography (μCT) indicated decreased bone area and volume in females with the B allele of *Alpl* (Fig. 2:5C), with no difference in overall animal weight between cohorts ($p = 0.76$). The same tendency was also noted for males. Together these data suggest that the BXD strains are a useful model for hypophosphatasia.

It is very plausible that the differential regulation of PLP (Fig. 2:5B) by ALPL drives further phenotypic changes (Fig. 2:5D). PLP acts as a cofactor in many important biochemical reactions: principally, it is a required prosthetic group for all aminotransferases, suggesting a critical role in amino acid metabolism [134], and it is also an essential coenzyme for biosynthesis of heme [135]. Confirmation of this link between ALPL over PLP to amino acid and hematological parameters warrants future work.

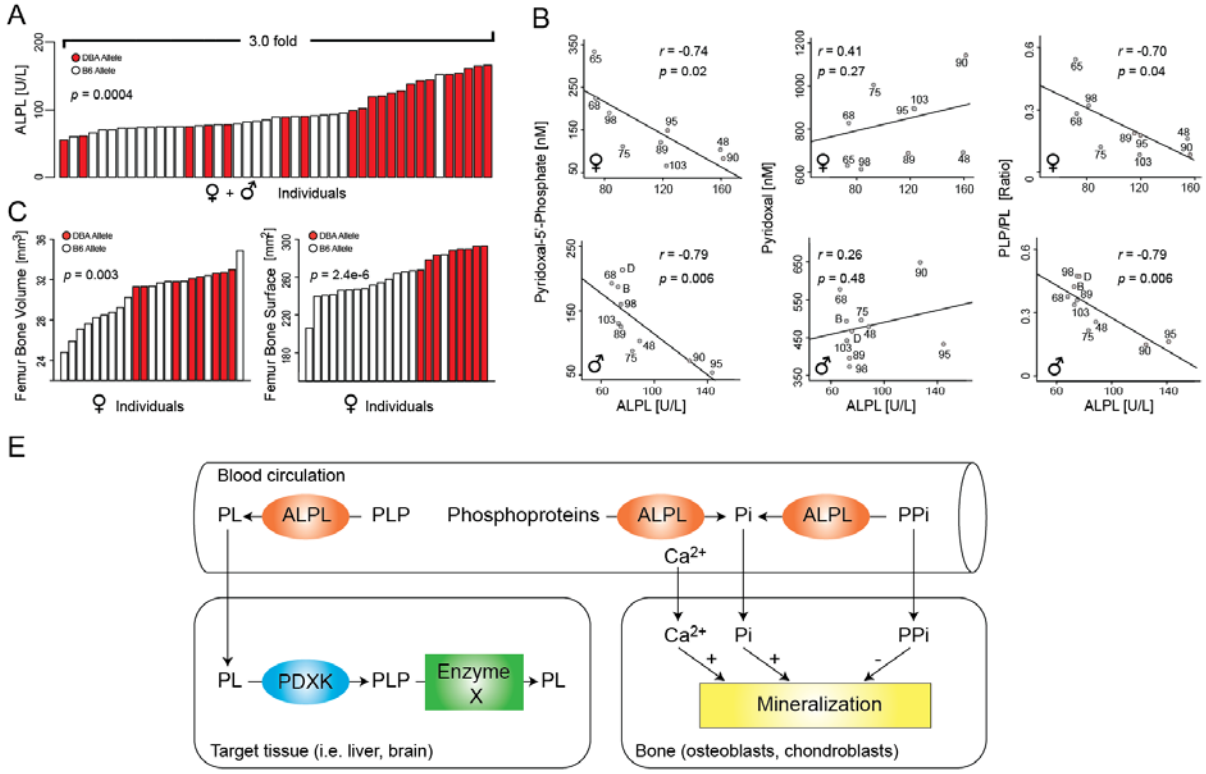


Figure 2:5 **Clinical Phenotyping of Lines with Extreme Alkaline Phosphatase Levels**

(A) ALPL levels in all individuals in a further in-depth study in male and female mice for individuals from the 11 strains described in the text. The effect observed on serum ALPL levels shown in Fig. 2:4A was highly reproducible in this independent study, even given a significant difference in the age of the animals in the two studies (19 weeks vs. ~46 weeks). **(B)** Extracellular PLP levels strongly negatively correlated with ALPL levels in males and females. No correlations were observed between ALPL and extracellular PL, but the ratio of extracellular PLP/PL was also strongly negatively correlated with ALPL. **(C)** Bone volume and surface area were significantly decreased in animals with the B allele of *Alpl*, consistent with the hypothesis of hypophosphatasia in the strains with the B allele. **(D)** Scheme summarizing the multiple roles of ALPL in metabolism. Before entering tissues, pyridoxal-5'-phosphate (PLP) in the plasma must be dephosphorylated by ALPL into pyridoxal (PL), which can traverse cell membranes. PL is converted back into PLP by the pyridoxine kinase (PDXK) in the target tissues, such as liver or brain. PLP is a cofactor in many enzymatic reactions. Circulating ALPL also converts pyrophosphate (PP_i) into inorganic phosphate (P_i). While PP_i inhibits bone mineralization, P_i and calcium (Ca^{2+}) stimulate it. Low levels of ALPL activity disrupt proper bone development and remodeling via this mechanism.

2.1.5 Variation in Glucose Response Indicates Genetic Factors Influencing Diabetes

Analysis of complex metabolic phenotypes also yielded QTLs with promising candidates. One such phenotype is glucose response during intraperitoneal glucose tolerance test (IPGTT), which records glycaemia levels before and 15, 30, 60 and 120 minutes after glucose injection and the overall glucose excursion, calculated as area under the curve (AUC). The AUC is significantly higher in males (Fig. 2:6A), but correlates between the sexes (Fig. 2:6B). The AUC is also linked to physical parameters such as body weight and composition in males, but surprisingly, not in females (Fig. 2:6B), perhaps partly explained by the fact that fewer female strains were studied. These results show that variation in glucose

response is significantly determined by genetics and partly sex independent. Consistent with these observations, mapping QTLs for glucose levels during IPGTT revealed both common and distinct loci in both sexes.

All genes under the four significant or suggestive QTLs for female glucose response on Chrs1, 2, 7 and 9 (Fig. 2:6C) were investigated. The locus on Chr1 was associated with late glucose response, with glycaemia at 120 minutes giving the highest LOD score (Fig. 2:6C). Within this locus, two genes are associated with type 2 diabetes susceptibility in humans: cytosolic phospholipase A2 group IV gene (*Pla2g4a*) and prostaglandin-endoperoxide synthase 2 (*Ptgs2*, also known as *Cox2*), both involved in prostaglandin metabolism [136, 137]. A second QTL for early glucose response mapped on Chr2 (Fig. 2:6C). Several genes known to be involved in insulin secretion or type 1 diabetes susceptibility are located under this peak, including paired box 6 (*Pax6*), *Cd44,* and catalase (*Cat*) [52, 138, 139]. When blood glucose is measured from mice under fasting conditions, only a single locus is mapped on Chr7 (Fig. 2:6C). Although there are 91 genes under the locus, no gene was previously linked to fasting glucose levels in mice or in humans. Finally, the global AUC is strongly associated with the QTL located on Chr9 (Fig. 2:6C). This region contains 31 genes, including one that has also been linked with insulin secretion: plasma membrane-related Ca^{2+}-ATPase-1 (*Atp2c1*) [140].

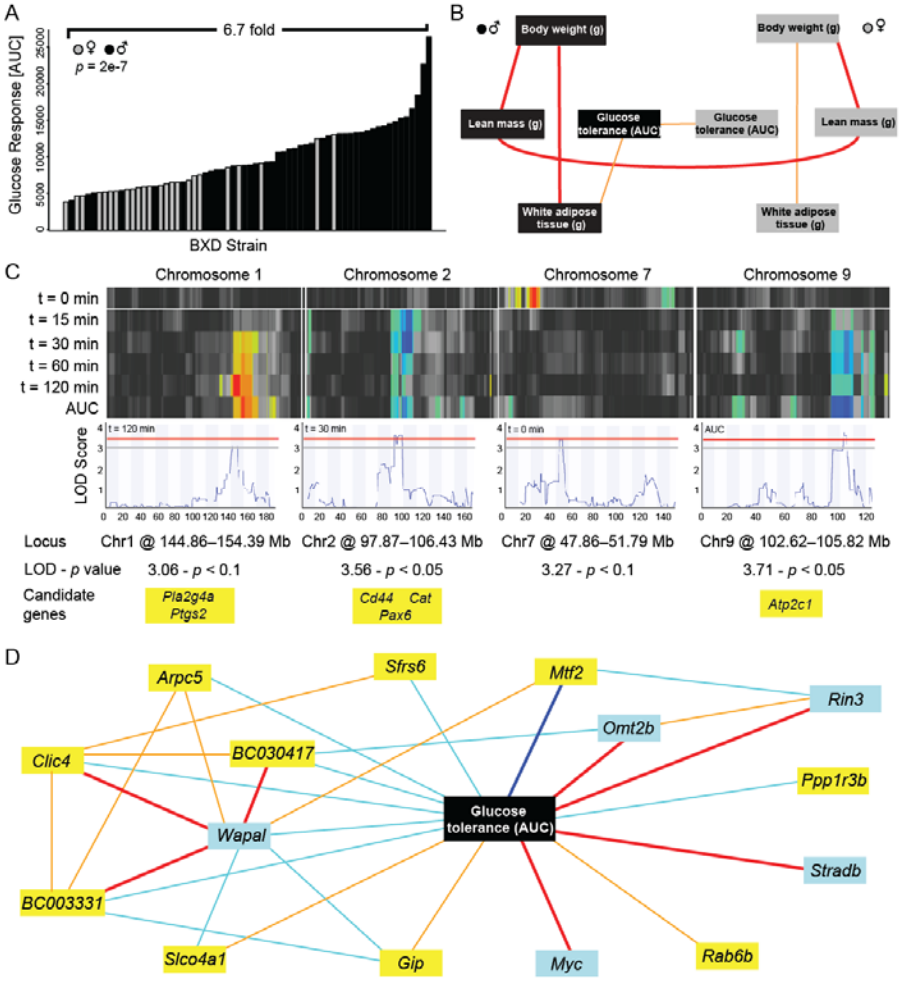


Figure 2:6 **Analysis of Response to an Intraperitoneal Glucose Tolerance Test**

(A) Variation of the AUC of the glucose levels from 0 to 120 minutes. **(B)** Network of glucose tolerance response determinants in males and females. Full details are given in Table S2. **(C)** Multiple and single QTL heat map for glucose levels at each time-point. Loci switch according to the feeding status of the mice. Fasting glucose (t = 0min) maps appear only on Chr7. The early response to glucose injection (t = 15min) maps on Chr2 only, while the rest of the time points also map on Chr1 and Chr9. Candidates under each QTL were selected according to existing literature showing their link with type 2 or type 1 diabetes. **(D)** Network built around overall glucose response using gene expression in the liver. Among the top 500 liver mRNA correlates with AUC, two were located on Chr1 QTL (*Arpc5* and *BC003331*), and one on Chr9 QTL (*Rab6b*) (yellow

boxes). One gene had a transQTL on the Chr1 QTL (*Ppp1r3b*), four had transQTLs on the Chr2 QTL (*Sfrs6, Clic4, Slco4a1* and *Mtf2*), and one had a transQTL on the Chr9 QTL (*Gip*) (yellow boxes). Blue boxes indicate the five genes best correlated with AUC (*Wapal, Rin3, Omt2b, Stradb* and *Myc*). Bold dark blue lines represent -1 < *r* < -0.7, light blues line -0.7 < *r* < -0.5, light orange lines 0.5 < *r* < 0.7 and bold red lines 0.7 < *r* < 1.

Functional analysis of the three QTLs related to glucose response on Chr1, 2, and 9 uncovered six genes known to be involved in glucose regulation. However, these QTLs contain altogether 131 genes, many of which are not yet characterized. Therefore, we highlighted potential new candidates by finding the strongest mRNA correlates in liver (Fig. 2:6D). From the list of the 500 best correlates, we selected candidates from three categories: (1) genes located under the QTLs on Chr1, 2 and 9; (2) genes that have *trans*-QTLs mapping to these loci; and (3) the five top gene correlates with glucose response. The most significant matches were integrated into the network (Fig. 2:6D). Protein phosphatase 1 regulatory subunit 3B (*Ppp1r3b*), a protein that regulates hepatic glycogen synthesis [141], *Myc*, which plays a role in hepatic glucose uptake and utilization [142], STE20-related kinase adaptor beta (*Stradb*), and *Rab6b*, correlated exclusively with AUC (Fig. 2:6D). In contrast, the rest of the candidates were entangled in an intricate network evolving around the AUC (*Mtf2, Slco4a1, BC030417, Srfs6, Gip, Wapal, Clic4, Arpc5, Rin3*, *Omt2b*, and *BC003331*) (Fig. 2:6D). Amongst these genes, *Sfrs6* and *Gip* have also been linked with glucose homeostasis [143, 144].

2.1.6 Energy Expenditure Linked to mRNA Expression Networks

The volumes of oxygen consumed (VO_2) and CO_2 exhaled (VCO_2) give an indirect measurement of mitochondrial function *in vivo*—another key complex metabolic phenotype [43, 118]. The respiratory exchange ratio (RER), defined as VCO_2 / VO_2, is an indirect measurement of the primary energy substrate used by the organism, e.g. fats, proteins, or carbohydrates. Together, these parameters provide a window on global energy homeostasis. In contrast to ALPL and glucose response, the overall variation in these three parameters was quite modest (Fig. 2:7A). Furthermore, these traits had only average heritability (Fig. 2:1D). Males and females did not have significantly different energy expenditure, and the phenotypes were not significantly correlated by strain. Despite these modest indicators of genetic influence, these traits mapped significantly and suggestively to five distinct QTLs (Fig. 2:7B).

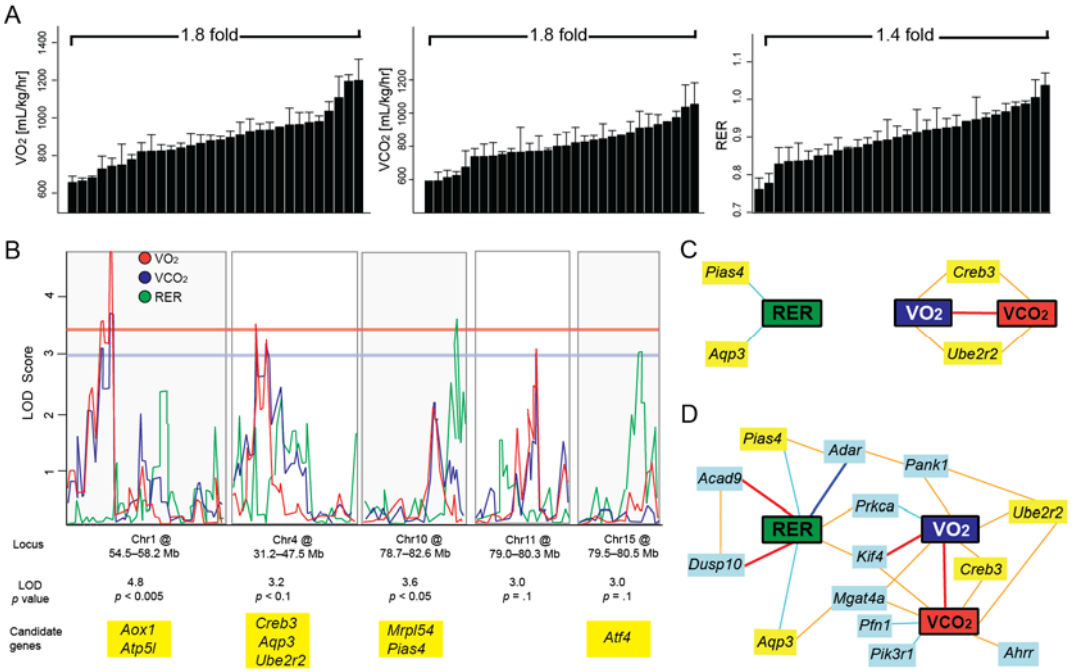


Figure 2:7 **Regulatory Network Underlying Differences in Respiration**

(A) Variation among male strains of BXD mice in three parameters of respiration: VO_2, VCO_2 and RER. **(B)** QTL graphs of the three respiratory parameters, VO_2 (blue), VCO_2 (red), and RER (green). Significance is shown by the red horizontal line, suggestive by the blue line (LOD > 3.3 & 3.0, respectively). Candidate genes with established links to energy expenditure are listed below the locus (*Aox1*, *Atp5l*, *Creb3*, *Ube2r2, Aqp3, Mrpl54*,

Pias4, *Tab1*, and *Atf4*). **(C)** Network graph showing all positional candidates with mRNA expression that correlates significantly with the phenotypes in whole eye tissue [120]. **(D)** Expanded network graph using the same dataset, including the top 10 mRNA correlates of the phenotypes (light blue boxes). Interestingly, RER had the strongest top mRNA correlates (*Acad9, Dusp10,* and *Adar*), while *Kif4* was the only significant correlate of all three parameters, and the only strong (|r|>0.7) mRNA correlate to VO_2 and none were observed for VCO_2. *Prkca* and *Mgat4a* were shared significantly between two parameters. Despite the strong correlation between VO_2 and VCO_2, most top correlates were not shared. Bold dark blue lines represent $-1 < r < -0.7$, light blues line $-0.7 < r < -0.5$, light orange lines $0.5 < r < 0.7$ and bold red lines $0.7 < r < 1$.

Similar to glucose response mapping, common QTLs were observed, with VO_2 and VCO_2 containing overlapping peaks on Chr1, 4, and 11. Somewhat surprisingly, RER maps to separate QTLs on Chr10 and Chr15 (Fig. 2:7B). All 441 genes under the five peaks were analyzed for functional relationships to energy expenditure. Several genes closely related to mitochondrial function were identified under the QTLs (Fig. 2:7B), for example ATP synthase H^+ transporting mitochondrial F_0 complex subunit G (*Atp5l*) and mitochondrial ribosomal protein 54 (*Mrpl54*) (Fig. 2:7B) [145]. Several candidates related to fat metabolism were also identified, including aldehyde oxidase 1 (*Aox1*) and activating transcription factor 4 (*Atf4*, also known as *Creb2*) [146, 147].

All positional candidates were subsequently analyzed for covariation with VO_2, VCO_2, and RER in order to identify potential candidates that were not yet reported in literature. This was done using an extensive BXD microarray database of whole eye mRNA [120], a tissue containing a variety of cell types, including neurons and both glycolytic and oxidative muscle. The network analysis of the 441 genes under the RER, VO_2, and VCO_2 QTLs narrowed the list to four genes tightly covarying with the phenotypes: aquaporin 3 (*Aqp3*), ubiquitin-conjugating enzyme E2R 2 (*Ube2r2*), cAMP responsive element binding protein 3 (*Creb3*), and protein inhibitor of activated STAT 4 (*Pias4*) (Fig. 2:7C). *Aqp3* and *Pias4* are negative correlates of RER, while *Ube2r2* and *Creb3* are positive correlates of VO_2 and VCO_2 (Fig. 2:7C). As of yet, none of these genes have been clearly linked to respiration or mitochondrial function. However, these candidates belong to gene families with clear ties, suggesting a functional, yet hitherto unknown relationship.

To expand the small network of positional candidates, the top 500 expression correlates with each trait were calculated and the top 10 correlates added into the network (Fig. 2:7D). This network indicated that while VO_2 and VCO_2 are tightly correlated and share QTLs, several of the most significantly correlated genes, like profilin 1 (*Pfn1*), phosphatidylinositol 3-kinase, regulatory subunit 1 (*Pik3r1*), and aryl hydrocarbon receptor repressor (*Ahrr*), were linked to VCO_2 exclusively (Fig. 2:7D). RER also presented exclusive correlates, such as dual specificity phosphatase 10 (*Dusp10*) and acyl-CoA dehydrogenase family, member 9 (*Acad9*). While on first thought it may be surprising that RER does not network closely with VO_2 and VCO_2, RER is a weight-independent measurement of metabolic flexibility. The common nodes that link VO_2 or VCO_2 with RER, such as protein kinase C alpha (*Prkca*), pantothenate kinase 1 (*Pank1*), and kinesin family member 4 (*Kif4*) have clear relevance to energy regulation [148]. Importantly, the only gene significantly linked to all energy expenditure phenotypes, *Kif4*, could affect metabolism possibly due to its regulation of poly ADP-ribose polymerase 1 (*Parp1*) [94, 149]. Similar network analyses using the same analytical technique in liver showed many similar top candidates, notably including *Pank1* and *Kif4*, which are also linked in the eye network to VO_2 or VCO_2 with RER.

2.1.7 Discussion

One of the main challenges in biomedical research is determining how genetic factors interact to influence human healthspan and how to develop effective strategies for the diagnosis, prevention, and treatment of chronic multifactorial metabolic diseases. Studies in model organisms often focus on single genes, which allows for precise mechanistic dissection of individual pathways yet lacks the same level of real-world genetic complexity behind variation in metabolic traits. This can compromise the generality and the translational utility of findings. To address this problem and to produce more robust models for systems genetics approaches on metabolism, we have shown the utility of the BXD GRP, which is currently the largest and best studied mammalian GRP. The BXDs constitute a reproducible and high resolution mapping panel with ~5 million segregating variants, similar to human populations [150]. When using GRPs, statistical power and mapping precision are mainly a function of the number of genotypes that are included in a study. For the metabolic phenotypes measured here, non-genetic variation within a genotype is generally low compared to variation across genotypes, demonstrating high heritability in a standard vivarium environment. Essentially all measured traits

have sufficient heritable variation to motivate detailed genetic and QTL analyses using only a moderate number of strains (~40) and mice (~180). Nearly half of the metabolic phenotypes differed significantly between sexes, thus metabolic traits do not generalize reliably from one sex to the other. While balanced and interleaved sampling of both sexes may offer more benefit than studying only one sex, though practical reasons often limit such designs.

As such, we analyzed 140 basic metabolic traits in both sexes, such as glucose response and body weight, of which approximately half displayed significant differences between males and females. For one such trait, activity, we examined it in closer depth in a cohort of male mice only. Across these measurements, we observed 40 significant peak QTLs and a further 16 had suggestive peak QTLs. The strongest QTLs are for expression phenotypes and Mendelian-like classical phenotypes, with the strongest mapping for ALPL enzyme levels and could be clearly linked to the *Alpl* gene. The BXDs have several variants in this gene, two of which cause amino acid changes in evolutionary conserved residues—R318Q and L324P—and are predicted to change the 3D structure and activity of the ALPL protein. Presence of these mutations in carriers of the B6 allele closely mimics downstream phenotypes of the human disorder hypophosphatasia, primarily the strong increase in extracellular PLP levels, which along with ALPL levels provides the most precise biomarker of hypophosphatasia [133]. One prominent downstream effect of ALPL is its effect on promoting bone mineralization by hydrolyzing inorganic pyrophosphate (PP_i), an inhibitor of hydroxyapatite formation, into inorganic phosphate. Consistent with this hypothesis, mice with the B allele had significantly reduced bone volume and surface area. These observation have important translational value, as *Alpl* coding variants are known to contribute to heritable differences in bone metabolism in humans [151].

Most metabolic traits are, however, complex and map to multiple loci and with lower strength. In one striking example, we identified that ~30% of the variability in spontaneous movement (itself 5-fold variable across the BXD population) is attributable to a single locus on chromosome 12. Through bioinformatic approaches, we were able to prioritize and rank genes in this locus and attribute *Ahr* as the most likely causal candidate. In a cross-species validation, we identified that reducing the activity of this transcription factor leads to a substantial increase in humanized mice, *C. elegans*, and *Drosophila*, a surprisingly conserved role for this gene in the regulation of movement (a relatively general phenotype) across such a wide range of species. For other complex traits, such as glucose levels, we were able to detect multiple loci. In this case, we confirmed six established positional candidates. Three of these candidates—*Cd44*, *Cat*, and *Pax6*—are located in the Chr2 locus and linked to insulin secretion or type I diabetes. In line with this linkage, a large F2 cross study of basal insulin levels also revealed a QTL on the same region [152]. Studies in humans have identified only a single insulin *sensitivity* gene (*PPAR* [153]), yet GWAS have identified several genes that control insulin *secretion* [44]. In addition to those identified genes that are clear candidates for glucose response, it is possible, even likely, that other genes under these loci also influence glucose/insulin response.

Similarly, genetic mapping of energy expenditure traits yielded five distinct QTLs containing over 400 positional candidate genes, including established genes in oxidative metabolism in two general categories: those tied to general mitochondrial function (e.g. *Mrpl54* and *Atp5l*) and those tied to the regulation of lipid metabolism (e.g. *Aox1* and *Atf4*). *Atp5l* is a component of the complex V ATP synthesis machinery in oxidative phosphorylation, while *Mrpl54* is part of the mitochondrial ribosome responsible for the translation of the 13 proteins encoded by the mitochondrial genome that are key components of the electron transport chain [145]. From the other angle, *Aox1* is involved in lipid metabolism and produces reactive oxygen species as byproducts [147] that are known to affect respiration [154]. *Atf4* (aka *Creb2*) is an established regulator of lipid and energy homeostasis [146]. However, in this case, the sheer number of candidate genes—and the likelihood that several are, together, involved in energy expenditure—precluded rapid validation of any particular candidate(s) by knockout or cross-species approaches, as used for *Ahr* (which mapped to a locus with only 9 candidate genes).

However, by combining QTL mapping with network building of transcriptomics, proteomics, genomics, and so forth, it becomes practical to expand the phenotype and positional candidate networks to highlight intriguing candidates genes that are good targets for mechanistic studies of glycaemia and energy expenditure. For example, the serine/arginine-rich splicing factor 6 (*Sfrs6*) is among the top correlates of glucose response, and it is associated with type 2 diabetes susceptibility in humans [143]. Many other relevant genes appear as top correlates when networks are extended further—the glucose-dependent insulinotropic polypeptide (*Gip*) is only vicariously connected to the phenotypes, but it has a strong functional relation: it promotes pancreatic islet function and exerts pro-survival actions in humans [144]. As with glucose response, network identification of the top mRNA correlates with the phenotype re-

vealed a melange of genes involved in different aspects of energy regulation. For example, the only significant correlate of all three parameters, *Kif4*, regulates Parp1, the major NAD^+ consuming enzyme in the cell [155]. *Parp1* critically controls oxidative metabolism, due to its competition with SIRT1 for a common and limiting NAD^+ pool in the cell [149]. *Acad9*, linked with RER, is essential for complex I assembly, via its impact on *Ndufa1* and *Ecsit* [93, 156]. However, other genes without known ties to lipid or glucose metabolism also appeared as top correlates, for example, *Ahrr* and *Pfn1*.

In terms of building and testing models of complex processes such as metabolism, even traits with low heritability or without strong QTLs provide great value. For instance, while VCO_2 is not linked to unique QTLs, its inclusion in network analysis not only strengthens the common interacting genes with VO_2, but also uncovers unique genes, which may be novel to the phenotype. Likewise, the integration of gene expression from liver to the glucose tolerance network increases the power to identify genes that are specific to one organ function—in this case hepatic glucose metabolism. For both analyses, it is likely that more novel genes involved in glucose response and energy expenditure remain under the loci described. Such network construction underscores the usefulness of the BXD GRP as a powerful resource for hypothesis generation and validation in a field as complex as metabolism. Further complementary phenotypic measurements of the BXDs under challenged conditions may help refine these regulatory pathways. Today, even the most striking connections derived bioinformatically must be validated mechanistically (a laborious and lengthy process), yet as models improve, old data may be revisited and reanalyzed, allowing such tentative linkages to be confirmed later due to improvements in technology or statistical approaches.

2.1.8 Methods

We acquired 25 BXD strains of females and 42 BXD strains of males from the University of Tennessee Health Science Center (Memphis, TN, USA). Strains and sexes were housed together with 2–4 animals per cage under 12h light, 12h dark cycle with *ad libitum* access to water and chow diet food (D04, SAFE, Augy, France) at all times. The studies shown in Fig. 2:5 were carried out in a subset of these strains at a significantly older age (40–52 weeks) than the animals in the primary study protocol (8–20 weeks). The general outline of the phenotyping study is described in the results section and summarized in Fig. 2:1C. All tests were carried out according to rigorous standard operating procedures (SOP) established and validated within the EUMODIC EMPReSS program [123, 157]. R was used for the analysis of phenotypic data, and GeneNetwork (**www.genenetwork.org**) was used for all genetic analyses [158]. Student's *t*-test was used to calculate significance between males and females. Pearson's *r* and *p*-values were calculated to determine magnitude and significance of correlations except when otherwise indicated. Correlation networks and QTL calculations are Bonferroni-corrected for multiple testing. Bar graphs and *X*-*Y* plots are expressed as mean+SEM. The *y*-axes of bar graphs are often cropped to highlight strain differences. Most analyses in this resource—including all of Fig. 2:6 and 2:7—can be generated entirely with public access to **www.genenetwork.org**. The phenotypes from this study can be found under BXD phenotypes by searching for "LISP1" and either analyzed on-site, or downloaded for separate analysis. A wide variety of public microarray data and phenotype data from other studies on the same BXD strains are also available for translational analyses on GeneNetwork. Together this makes the BXDs among the largest and best-categorized family of isogenic strains, and we expect resources like this to serve as the backbone for murine studies on complex disease for the coming decade.

2.1.9 Acknowledgments

I would like to thank Laurent Pouilly, Tania Sorg, Hongbo Zhang, and Ulrike Kettenberger for technical assistance with various aspects of mouse phenotyping. Discussions with Thomas Vogt (Merck Research Laboratories), Carmen Argmann and Sander Houten (University of Amsterdam), Vincent Mooser (CHUV), Stephan Morgenthaler (EPFL), Dominique Pioletti (EPFL), Bernard Thorens (UNIL), Paul Franken (UNIL), and Ioannis Xenarios (UNIL) and the team members of the Auwerx lab contributed to the outline and execution of this project. RHH was supported by a Rubicon fellowship of the NWO. JA was the Nestlé Chair in Energy Metabolism and the JA/KS laboratory was supported by grants of the Ecole Polytechnique Fédérale de Lausanne, the EU Ideas program (ERC-2008-AdG-23118), the Velux Stiftung, the Swiss National Science Foundation (31003A-124713, 31003A-125487, and the Sinergia grant CSRII3-136201). RWW and GeneNetwork were supported by NIH (P20-DA 21131 UO1AA13499 and U01AA14425) and the UT Center for Integrative and Translational Genomics.

2.2 Effects of Genetics on Spontaneous Movement and Exercise

2.2.1 Examination of Behavioral-Metabolic Traits: Movement

Due to the striking results obtained for metabolic variability in male and female BXD strains, we chose to examine further metabolic phenotypes, including traits influenced by behavior such as voluntary spontaneous activity, in new experimental programs. One such phenotyping program was developed to examine how two basic metabolic parameters—activity and weight—vary across the population, and how much of this variance may be attributed to genetic variants. In the previous study (chapter 2.1), we examined how many metabolic traits vary due to sex in the BXDs [16], but did not examine basal movement. We thus examined 68 female and 68 male animals from 22 BXDs strains at around 16 to 24 weeks of age. Animals were placed individually in normal housing cages and spontaneous movement recording over a 48 hour period. Females were slightly more active (Fig. 2:8A), and movement was highly variable for both sexes (~3-fold for ambulations and ~4-fold for rearing), but overall, movement was strongly correlated by strain between males and females (Fig. 2:8B), with the range of across genotypes (~5-fold range) dramatically outweighing the range across the sexes (~1.3-fold, Fig. 2:8A).

With this variance and sexual dimorphism in mind, we designed an enlarged phenotyping program to search for genetic regulators of activity, but this time only in male BXDs and using a larger population sample: 43 strains with 5 animals per cohort phenotyped at precisely 23 weeks of age, using the same diet and recording setup. In the expanded data, both rearing and ambulatory activity again varied dramatically across the population—5-fold and 8-fold respectively (Fig. 2:8C). The two aspects of movement were tightly correlated by strain (Fig. 2:8D) and highly consistent for all five biological replicates within each strain, yielding high estimates of heritability (h^2 = 0.59 for ambulations, not shown, and h^2 = 0.68 for rearing, shown in Fig. 2:8E). The BXD population likewise had highly variable body weights across the population (~2-fold range) with a similarly high estimate of heritability (h^2 = 0.74; data not shown). Surprisingly however, spontaneous activity had no association with body weight or food intake (Fig. 2:8F), indicating the strains' movement is driven primarily by internal motivating factors, rather than by access to food or water, despite that both require rearing to reach. Due to the strong heritability and wide and consistent cross-strain variance, we suspected that the movement variation may be linked to quantitative trait loci (QTLs), which could indicate the region(s) of the genome causing the genotypically-driven effects. Indeed, for both measurements, rearing and ambulation, we detected QTLs, and interestingly, both overlapped to a small narrow-sense peak on chromosome 12 from 35.5 to 37.6 Mb (broad-sense from 30.3 to 37.6 Mb) (Fig. 2:8G). For ambulatory activity, this locus explains 25% of the overall variance (not shown), corresponding to 1300 counts/hr, and for rearing the same locus explains 41% of the variance, or 400 counts/hr (Fig. 2:8H).

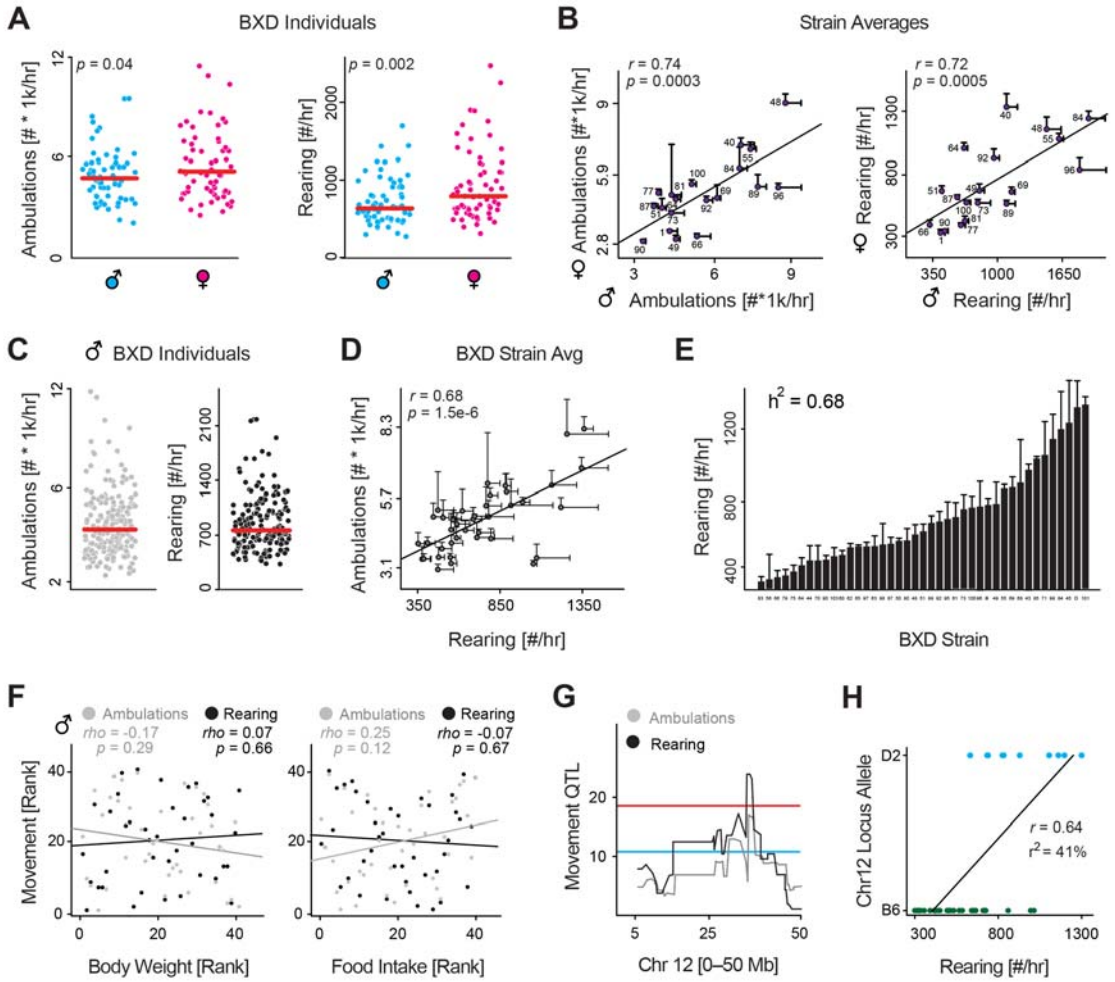


Figure 2:8 **Identification and Validation of a Movement QTL**

(A) Nighttime rearing and ambulatory activity for all 68 males and 68 females phenotyped across 22 (male) and 19 (female) strains. These 19 strains were phenotyped in both sexes. Females are slightly more active. **(B)** Despite moving somewhat more, female and male activity are strongly correlated by strain. **(C)** Nighttime rearing and ambulatory activities for all 196 animals across 43 strains. Each strain has ~5 biological replicates. **(D)** Ambulatory and rearing activity are tightly correlated, though the measurements are technically independent. **(E)** Nighttime rearing activity for all 43 strains, ordered by value. Activity varies by 3.9 fold across the population. The strong heritability (h^2) of 0.68 indicates that the majority of this variance can be attributed to genetic factors. **(F)** Body weight (Left) and food intake (Right) have no effect on ambulatory or rearing activity, suggesting movement is largely independent of the weight or the need to eat or drink. Animals must rear to reach the food basket or drink. **(G)** Rearing and ambulatory movement mapped to a common narrow 2 Mb locus on chromosome 12. **(H)** The target locus (chromosome 12 from 35.5–37.6 Mb) explains ~40% of variance (r^2) in rearing activity and ~25% of variance in ambulatory activity (not shown).

While the movement parameters mapped to several suggestive and two significant loci, the significant locus on chromosome 12 was the most striking and consistent (Fig. 2:9A), thus we prioritized it for validation. To establish the effect of the Chr12 locus, we examined a congenic strain of C57BL/6J (B6) with the DBA/2J (D2) locus at the region of interest (B6.D2N-*Ahr*^{d}) [122]. We sequenced this line and observed it carries a 6 Mb segment of D2 genome on Chr 12 between 34.60 and 40.48 Mb, while the rest of the genome is B6 (though several dozen individual SNPs—i.e. spontaneous mutations—are observed elsewhere in the genome). Ten males from all three cohorts were then entered into the same phenotyping platform until 23 weeks of age, at which point the movement experiment was performed. As predicted, the congenic line and D2 moved significantly more than the B6 animals, while the congenic line and D2 moved

the same amount (Fig. 2:9B), validating the QTL as causative of movement variance. Moreover, these increases matched the calculated effect size from the QTL: ambulatory activity increased by 1500 counts/hr, while rearing activity increased by 380 counts/hr. This analysis validated the QTL as influencing movement, though did not indicate which candidate gene(s) cause the effect (Fig. 2:9C). The broad-sense QTL, from 30.3 to 37.6 Mb, contains 38 genes, including 9 which are under the narrow-sense QTL (35.5 to 37.6 Mb). The congenic region, from 34.6–40.5 Mb, contains 17 genes, 13 of which are within the QTL bounds. We retained all 42 genes (38 within the QTL, 4 exclusively in the congenic region) for subsequent bioinformatic analyses, though with a particular eye for the 9 genes overlapping in the congenic and significant narrow-sense QTL.

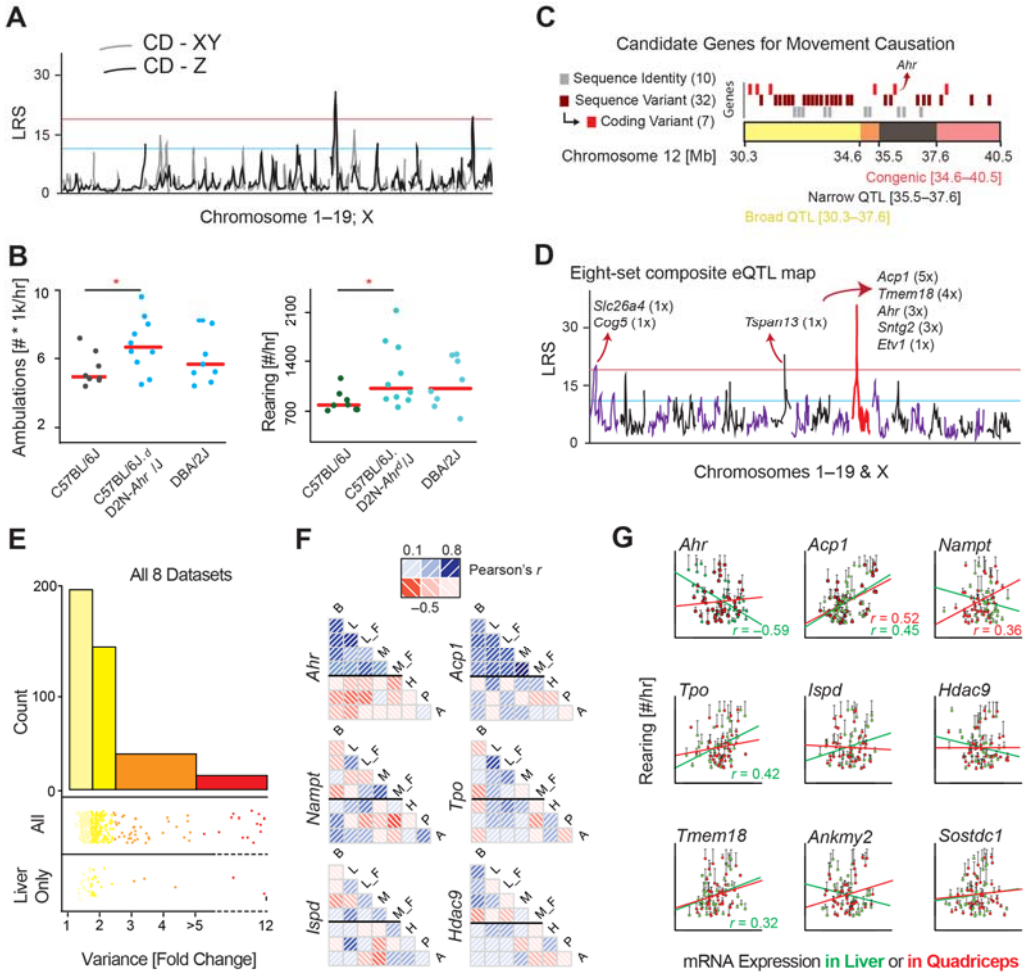


Figure 2:9 **Identification of the Movement QTG: *Ahr***

(A) The movement QTLs across all chromosomes indicates one consistently significant locus on chromosome 12, and one significant locus on chromosome 19 only for rearing activity. **(B)** A congenic strain on a B6 background with a D2 congenic interval at the region of the chromosome 12 QTL has significantly more spontaneous movement, validating that this locus impacts movement. One-way ANOVA is suggestive (p = 0.09), while the t-test between only the congenic and control groups is significant (p = 0.01). **(C)** Close-up of the congenic (red) and overlapping significant QTL (black) and suggestive (yellow) regions on chromosome 12. All genes are represented by rectangles at their approximate chromosomal positions. Genes that are sequence identical across the BXDs are marked in grey and were discounted. Genes with sequence variants of unknown effect (maroon) and genes with sequence variants with clear effects (e.g. coding differences, red) were considered more likely candidate QTGs. **(D)** eQTLs were plotted for all positional candidates in all eight datasets tissues. Five genes map to chromosome 12 (~25-45 Mb) as cis-eQTLs. As the movement QTL also maps to this locus, trans-eQTLs are of less interest. **(E)** Transcript variance was checked across all 9 datasets. Half of the transcripts are at least moderately variable (range > 1.75 fold) across the cohorts, while 9% of transcripts are highly variable (> 3.0 fold). Upper stripchart: Variance of all 42 candidate genes in all 8 tissues

(i.e. each gene is represented ~8 times). Lower stripchart: Variance of all 42 candidate genes in CD liver. **(F)** Few transcripts are consistently expressed across strains in different tissues (Brown Adipose: B; Liver: L; Muscle: M; Hypothalamus: H, Pituitary: P, Adrenal: A; F: High fat fed cohorts; all others are chow fed). Ahr covaries well for all five datasets taken from this study, but negatively with the three publicly available BXD datasets; Acp1 covaries positively in all datasets, although much stronger in the five paired sets. Most other genes (e.g. the four shown) have little consistency. **(G)** Gene expression correlates with movement inconsistently by tissue, with only a handful of genes yielding consistent (*Acp1*, *Tpo*, *Tmem18*) or significant (*Ahr*, *Acp1*) correlations. Nominally significant correlations ($p < 0.05$) are displayed on the chart. For brown adipose tissue, no genes correlate highly significantly ($p < 0.01$) with activity.

2.2.2 Identification of the Causative Gene, *Ahr*, in a Gene-Dense QTL

To select candidate genes for validation experiments, we used several established methods to prioritize candidate gene(s), which are most likely to influence movement [159]. First, the sequence variants were examined for all candidate genes, including within 5 kb of the 3' or 5' untranslated regions. 10 of the candidates are identical by descent across all of the BXD strains, making these genes unlikely to be causal for the QTL [160], including 3 of the 9 priority candidates. For the 32 genes with sequence variants, seven have protein-coding changes: peroxidasin homolog (*Pxdn*), thyroid peroxidase (*Tpo*), histone deacetylase 9 (*Hdac9*), mesenchyme homeobox 2 (*Meox2*), transmembrane proteins 18 and 195 (*Tmem18* and *Tmem195*), and *Ahr*. To further rank the candidate genes, we examined the transcriptional variance and regulation in eight diverse microarray datasets. Candidate genes with higher transcript variability, which have strong genetic variants at the same locus as the phenotype, and/or which associate with the phenotype are more likely to be the QTG under the QTL [159].

First, we examined the 42 genes in 8 datasets from 6 tissues. In three tissues—liver, brown adipose, and quadriceps—measurements were performed in the same BXD animals, which mapped to the movement QTL. Quadriceps and liver were also sampled in the same BXD strains on a high fat diet (explored in more depth in Fig. 2:11). The other three tissues—hypothalamus, pituitary, and adrenal—were collected previously and published by other research groups in the same BXD strains in similar conditions (i.e. age, sex, diet) [141]. All transcripts were detected in at least one tissue except *Slc26a3* (in the congenic interval) and *Prps1l1* (in the narrow QTL region). We then mapped all transcripts to identify the existence and location of significant expression QTLs (eQTLs; LRS ≥ 20, Fig. 2:9D). No *trans*-eQTLs were consistent across more than one dataset, while four genes gave consistent significant *cis*-eQTLs: acid phosphatase 1 (*Acp1*, in 5 datasets)*, Tmem18* (in 4), *Ahr* (in 3) and syntrophin gamma 2 (*Sntg2*, in 3). We then examined the transcript variance of all candidate genes in the eight datasets. Most genes had at significant variability across the strains in each tissue (50% have variance ≥ 1.75 fold; Fig. 2:9E), with three genes, *Sh3yl1*, *Prkar2b* and *Acp1*, being particularly highly variable (range ≥ 3.0 fold) in multiple tissues. However, most genes did not covary across the tissues, with only two having particularly consistent expression: *Ahr* and *Acp1* (Fig. 2:9F). We last examined how the expression of each gene associated with movement phenotypes in the BXDs, with particular focus on candidates under the significant QTL and congenic locus (e.g. *Ahr*, *Sostdc1*, *Ispd*) and those with major or consistent transcript variance (e.g. *Acp1*, *Prkar2b,* and again *Ahr*). Only two genes, *Ahr* and *Acp1* had significant correlations after multiple testing correction (Fig. 2:9G), though several other genes yielded consistent but non-significant correlations (e.g. *Tmem18*). Together, these bioinformatic analyses indicated several genes as potentially causative of movement variance, and with one top candidate: *Ahr*, which was prioritized as the first gene for validation as the quantitative trait gene (QTG), as the other strong candidate gene, *Acp1*, was not in the congenic region or under the peak QTL.

2.2.3 The Effect of *Ahr* On Movement is Highly Conserved Across Evolution

Ahr is strongly conserved throughout evolution (Fig. 2:10A), and acts as a bHLH-type transcription factor with impact on development and homeostasis in all species [142]. Given this consistency, we examined whether movement regulation may be another conserved physiological process regulated by the gene. In the BXDs, three particular SNPs have been established as causal for differences in *Ahr* by affecting its enzymatic activity (A375V), ligand binding and the *cis*-regulatory mechanism (L471P), and protein length (*805R, which adds 43 amino acids to the C terminus) [161, 162]. Strikingly, these three particular variants are conserved in humans, with the most common human allele (*hAHR*) humans matching the D2 allele at all three (Fig. 2:10B) [163]. Correspondingly, *hAHR* enzymatic activity is similar to that of D2 mice [163, 164]. Moreover, the valine at position 375 (or 381 in humans) is unique among mammalian reference

genomes to D2 and humans, and is not even found in macaques, which likewise have much higher *Ahr* activity than humans and D2-type mice [165, 166]. To examine a potential link between *hAHR* and activity, we phenotyped transgenic B6 mice with the *hAHR* allele replacing the murine *Ahr* [167]. In the same home cage monitoring experiments as those used before, we observed that the humanized animals were significantly more active than their wildtype counterparts (Fig. 2:10C), and again with an increase equal to that of the congenic line (Fig. 2:9A). This finding indicates both that *Ahr* was properly selected as the QTG, and that the effect may well be conserved cross species. To validate this hypothesis, we looked to lower organisms.

The key transcription factor motifs in *Ahr* (bHLH and PAS) are highly conserved in the *D. melanogaster* ortholog called *spineless* (*ss*) [168], and the *C. elegans* homolog called *ahr-1* [169], thus we hypothesized the regulation of movement may be further conserved to these simpler model organisms. We first examined movement in *D. melanogaster*, where we crossed the *w* strain with a loss of function allele $ss^{D115.7}$ [168], and examined this line as a heterozygous knockout, both in males and females. In both models, *ss* reduction resulted in a robust ~25% increase in movement (Fig. 2:10D). Given the conservation to *D. melanogaster*, we hypothesized that this connection may manifest also in *C. elegans*. As before, inhibition of *ahr-1* by RNAi resulted in a marked and robust increase in activity (Fig. 2:10E). Moreover, the data in *C. elegans* indicates that the effect of *Ahr* inhibition on activity is approximately linear, at least within the expression variation tested. Full knockouts in mice, while viable, have poor postnatal survival rates [170] and have dramatically smaller livers (< 50% size [171]); likewise, full knockouts in *Drosophila* of *ss* have notable morphological problems [142].

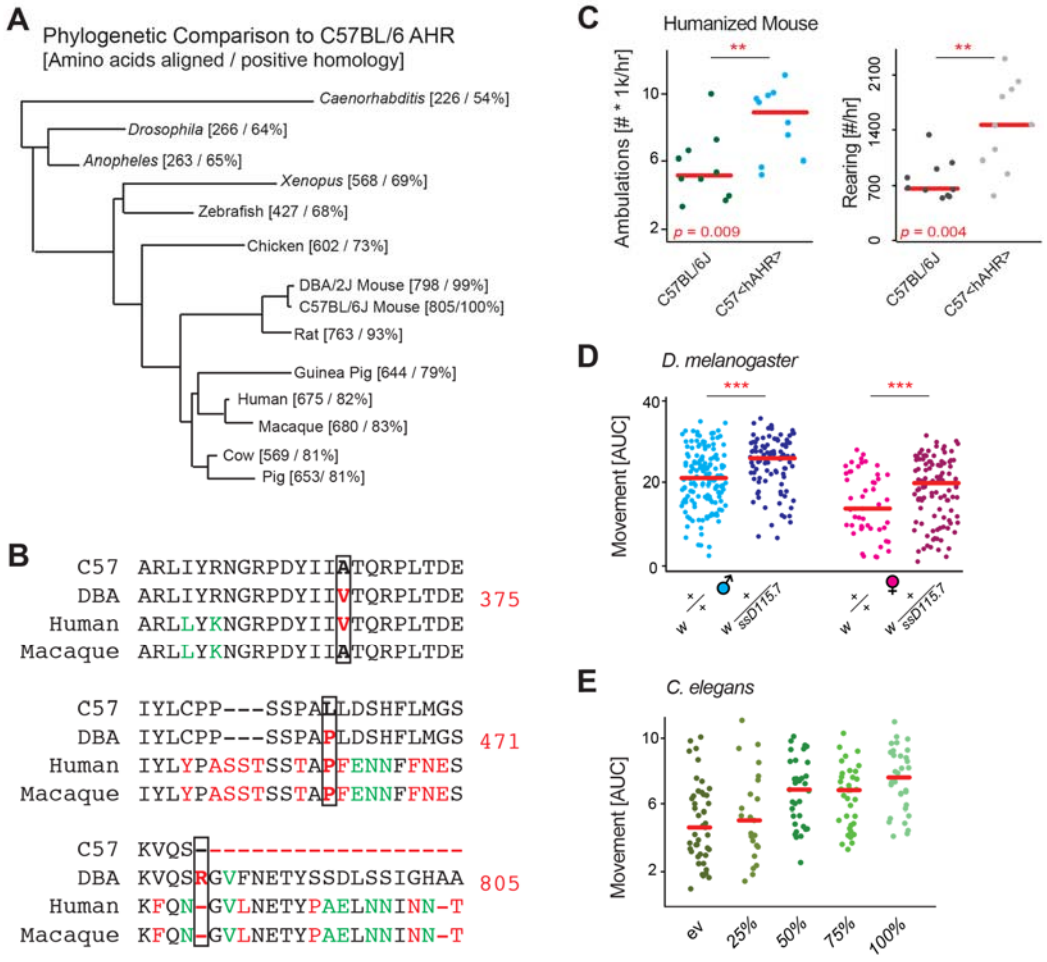


Figure 2:10 **Evolutionary Analysis Links *Ahr* to Movement**

(A) Phylogenetic BLAST analysis of mouse Ahr showed that the gene is highly conserved down to simple multicellular animals such as *C. elegans*, the gene likely has conserved basic metabolic functions. **(B)** Sequence

analysis of the three missense mutations of Ahr between B6 and D2 known to have an impact oh AHR activity (375, 471, 805). **(C)** B6 mice with the humanized AHR allele are nearly twice as active as controls. The humanized AHR allele is similar to the D2 allele in many tests of enzymatic activity, with a ~90%+ reduction in activity compared to the B6 allele [24]. **(D)** *D. melanogaster* with a heterozygous deletion allele of the *Ahr* ortholog *ss* are also significantly more active than controls. The 50% reduced expression appears to increase movement by about 20% in both males and females. Each comparison is a separate Welch's *t*-test with $p < 0.001$. Females are ~30% less active than males in both instances ($p < 0.001$). **(E)** *C. elegans* treated from early development with RNAi for *ahr-1* are nearly twice as active as worms treated with a control vector. Reduced doses of RNAi have intermediary effects on activity. p = 2.9e-6 for 100% vs. empty vector (*ev*).

Ahr inhibition robustly increased movement in all models examined, and the effect was observed in the absence of any clearly established *Ahr* ligand. In *Drosophila* and *C. elegans*, the *Ahr* orthologs are suspected to be exclusively constitutively active [172]. However, this does not indicate whether it is constitutive activity operating in the mice, or if it is an unknown dietary component of the chow diet common to all cohorts, which could influence movement in a ligand-dependent manner. As known *Ahr* ligands are either known to be non-selective and activate multiple signaling pathways (e.g. resveratrol or quercetin [147, 173]) or are highly toxic and poorly suited for normal physiological studies (i.e. TCDD), we chose to expose BXD cohorts to an environmental component that is known to influence movement: a high fat diet (HFD). We raised males from the same BXD strains, but now on a HFD from 8 until 23 weeks of age at which point we again measured spontaneous activity (Fig. 2:11A). HFD robustly increases body weight by this age, by about 12 grams or ~35% of body weight in each strain (Fig. 2:11A). Movement is similarly affected, with rearing decreasing by ~50% (Fig. 2:11B) and ambulatory movement by ~25% (not shown). Strikingly, both movement parameters again map to the same locus on Chromosome 12, indicating the genetic effect of this locus is independent of the dietary influence on movement (Fig. 2:11C). Moreover, the dietary effect is consistent across all strains (Fig. 2:11D, left) and is not directly due to the increase in body weight (Fig. 2:11D, right). Thus, while there may be a gene-by-environment effect of diet on movement in the BXDs, it is independent of *Ahr*, which is equally expressed in both dietary cohorts (not shown). As in the CD cohorts, *Ahr* expression has a strong negative correlation with spontaneous activity in the HFD population (Fig. 2:11E). This independent phenotype both confirms the locus identified using CD cohorts, but also indicates that *Ahr* can influence movement in mammals independently of incidental dietary or environmental effects.

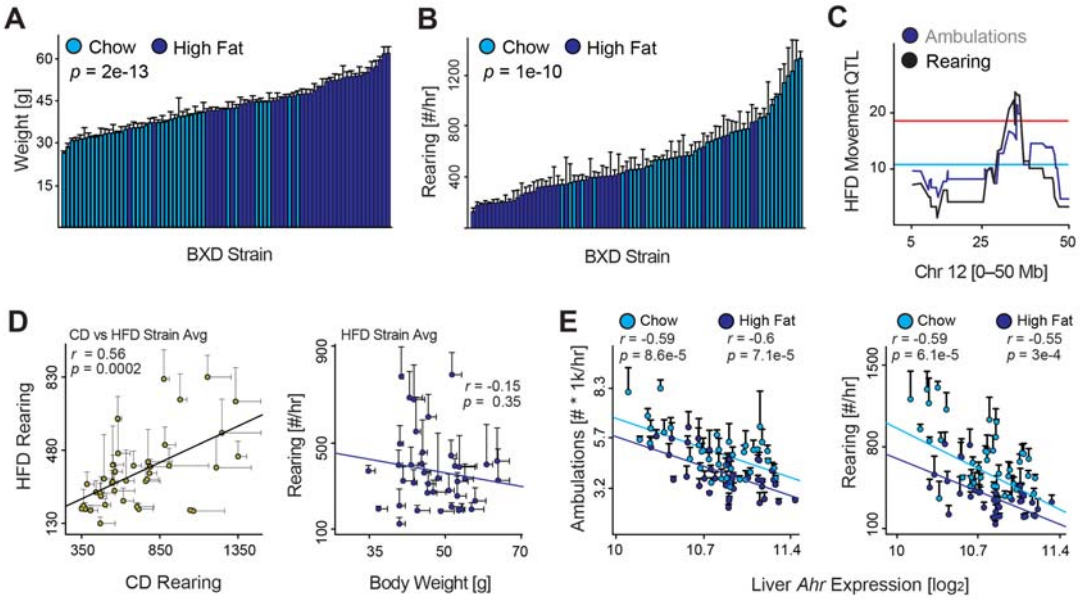


Figure 2:11 **Environmental Effects on Movement and *Ahr***

(A) Weight is significantly increased in cohorts fed HFD (dark blue) by 23 weeks of age—after 15 weeks on HFD—compared to the CD-fed cohorts (light blue). **(B)** Rearing activity and ambulatory activity (not shown) are significantly reduced by HFD feeding. **(C)** Both movement parameters also map significantly to the *Ahr* locus in BXDs fed HFD, indicating that the effect of HFD on movement is independent of this locus. **(D)** Left: Rearing and movement activity in HFD cohorts correlates well with the movement of chow diet (CD) cohorts run previously. Right: Body weight has no effect on spontaneous activity in mice on HFD, which is similar as

observed previously for animals on CD. **(E)** Ambulatory activity (left) and rearing activity (right) correlations with Ahr expression in CD and HFD BXD cohorts. Ahr expression is unaffected by HFD.

2.2.4 Discussion

In this study we characterized 43 strains of the BXDs genetic reference population to assess basal physiological parameters: movement and body weight. Both phenotypes are driven by numerous and complex interactions between genes and environment, by neurological/motivational states, and by physiological limitations. In the BXDs, body weight and physical activity are both sexually dimorphic, highly variable, and heritable. Surprisingly, body weight and food intake have no impact on standard spontaneous activity in either CD or HFD-fed cohorts despite a major decrease in movement in HFD cohorts, and despite that animals must be active to eat and drink. In both dietary groups, we identified a single common QTL causal of ~25-40% of the variance of movement across the population. Using a congenic line, we confirmed the effect of this locus and set out to establish the causal gene through a bioinformatics approach. By analyzing nine diverse tissues, five in the cohorts phenotyped and three from other published BXD studies, we were able to establish the aryl hydrocarbon receptor (*Ahr*) as the single best candidate gene for mechanistic validation.

Ahr is an evolutionary conserved transcription factor involved in development, signal transduction, and metabolism [174, 175]. *Ahr* has constitutive activity, but can also be activated by a variety of ligands such as the endogenous metabolite, kynurenine, and a wide variety of environmental chemicals. Moderate to strong (~95%) reductions in *Ahr* activity appear to have little negative effect an organism's health or viability, whether in D2 or humanized hAHR mice, or in heterozygotic fly mutants. Through cross-species analysis to *Drosophila* and *C. elegans*, we were able to confirm that reducing the expression of this gene consistently leads to an evolutionarily consistent increase in spontaneous movement.

After examining the same BXD strains on a HFD, we observed that despite a major decrease in movement in this population, the *Ahr* QTL is consistent and with a similar effect as calculated in the CD population, and further is independent of the major environmental perturbation caused by diet. This conserved effect in BXDs across different environmental conditions indicates that the role for *Ahr* in the regulation of movement is likely constitutive, i.e. innate and not reliant on dietary ligands. This hypothesis is further supported by the observation that reduction of *Ahr* orthologs in invertebrates has a consistent effect on movement, as the *Drosophila* ortholog (*spineless)* and *C. elegans* ortholog (*ahr-1*) are both constitutively active [176] and are not known to be affected by any exogenous ligands [172]. In humans, a large number of *AHR* polymorphisms have been identified in diverse population studies [177, 178], though it remains to be seen if these variants lead to variation in locomotion and/or disposition to exercise in humans as in mice. However, as the movement link is conserved in mice with a humanized *Ahr* allele, it seems likely that natural variation in *Ahr* or of its ligands may explain part of the natural variation in human proclivity for activity. Furthermore, while our data indicate constitutive *Ahr* activation as a regulator of movement, it is conceivable that this role may be *further* modulated by specific ligands in mice and humans. In combination, our study expands the phenotypic roles of AHR, endowing it with a commanding role in the control of movement that is conserved across evolution.

2.2.5 Methods

All animals were communally housed by strain until phenotyping and fed a chow diet (CD; (Harlan 2018; 6% kCal/fat, 20% kCal/protein, 74% kCal/carbohydrate) throughout life after weaning. All BXD strains (BXD43–103) were originally sourced from the vivarium at the University of Tennessee Health Science Center (Memphis, TN, USA) then bred for two or more generations until progeny entered the phenotyping colony.

Male versus Female Phenotyping: 136 retired breeders (68 female, 68 male) from 22 strains (male) or 19 strains (female; all 19 overlap) were taken at 20±4 weeks of age from a breeding colony at the EPFL facility and transferred to the phenotyping unit. Males and females were separated for 3+ weeks to ensure pregnant females were not phenotyped. Males and females were phenotyped on separate days, with 10–16 animals entered into the phenotyping program every 2 days.

CD Male Phenotyping: 196 male mice from 43 strains of the BXD family were bred at the EPFL facility and transferred to the phenotyping unit at 8 weeks of age. Each cohort was communally housed (3–5 animals per cage) under 12

h light, 12 h dark cycle with *ad libitum* access to food and water at all times. Animals were in solitary cages only for the movement phenotyping test (48 hours) and were sacrificed 5 weeks after at ~28 weeks of age.

HFD Male Phenotyping: 186 animals from 42 strains (all but 1 overlapping) were entered into the colony as before, with HFD starting at 8 weeks of age (Harlan 06414; 60% kCal/fat, 20% kCal/protein, 20% kCal/carbohydrate). Animals were in solitary cages only for the movement phenotyping test (48 hours) and were sacrificed 5 weeks after at ~28 weeks of age.

Congenic AHR mice were purchased from The Jackson Laboratory (stock number 002921), with animals delivered at 8 weeks of age along with control B6 and control D2 mice. The congenic mice were generated by crossing B6 with D2, followed by successive backcrossing (N13) to return the B6 genome except in the region of *Ahr* [122]. The congenic region was genotyped independently to confirm the size of the interval. We then sequenced the DNA of the congenic strain using an Ion Proton PI Chip at 1.5x depth (i.e. ~15 million 200 bp reads) and aligned it against the C57BL/6J reference. We confirmed the reported congenic interval is the only region that retains a D2 background. Several dozen SNPs were observed throughout the genome outside the reported Chr12 interval (~34.6 to 40.5), but were distributed evenly across the chromosomes and consequently represent spontaneous mutations or sequencing and assembly errors, rather than residual D2 genotype. 10 humanized AHR mice were ordered from Taconic (model 9165), delivered at 8 weeks of age along with B6 controls [164]. The transgenic mice have exons 3–11 (of 11) replaced with hAHR, while exons 1–2 retain the B6 sequence, retaining 2 amino acids unique to B6 not present in hAHR. The key mutations (375, 471, and the lost stop codon) are present in the transgenic animal.

For tissue collection on CD and HFD BXD cohorts, animals were sacrificed under isoflurane anesthesia and cardiac perfusion after an overnight fast. High fat diet treatment and two day isolation for the recording experiment were considered as having low impact on the animals' welfare, while all other measurements and conditions were considered as having no negative impact. All research was approved by the Swiss cantonal veterinary authorities of Vaud under licenses 2257.0 and 2257.1.

Home cage monitoring was performed at 23±1 weeks of age for all mice except retired breeders (23±4 weeks), using a laser detection grid developed by TSE Systems (Bad Homburg, Germany) and used in the animals' standard housing cages. The detection grid has two layers: one for detecting X-Y movement ("ambulations") the other for Z movement ("rearings"). Both measurements are technically independent, though the measurements of movement are strongly correlated (r ~ 0.70, see Fig. 2:8B). Animals were housed individually for the 48-hour experiment starting at about 10am, with the night cycles (7pm–7am with 30 minutes of both dawn and dusk) used for movement calculations [179].

D. melanogaster lines containing a null mutation in the *spineless* gene, the *D. melanogaster* ortholog of *Ahr*, *w; $ss^{D115.7}$ / TM3*, were obtained courtesy of Ian Duncan's laboratory and passaged for two generations in a standard incubator. This line was crossed with w^- and the progeny segregated accordingly (i.e. *w* vs. *w;$^{ssD115.7}$* and *w;TM3, hb-LacZ* vs. *$w^{ssD115.7}$ / TM3, hb-LacZ*). Movement was recorded by placing flies in a sealed chamber, tapping the chamber, and recording their movement as they naturally climb towards the top. For the tapping test, 1–2 day old flies were recorded using a standard SLR camera with a Leica macro objective. The experiment was performed four times for each cohort with one minute recordings each, with a "tap" sending the flies to the bottom of the chamber every 10 seconds. The speed with which flies reached the top of the chamber was measured using the Parallel Worm Tracker for MATLAB, which we modified slightly to work with *D. melanogaster* [180]. This speed was converted into distance by taking the area-under-the-curve (AUC) integral of their velocity.

C. elegans movement was recorded for 45 seconds at days 2, 3, and 4 of adulthood using a Nikon DS-L2 / DS-Fi1 camera and controller setup, attached to a computerized Nikon bright field microscope. Seven plates of worms, with 10 worms per plate, were measured in each condition. The movement of worms during this time was calculated by following the worm centroids using the same modified version of the freely-available for the Parallel Worm Tracker as above.

R was used for basic analysis of phenotypic data. GeneNetwork (**www.genenetwork.org**) was used for correlation and genetic analyses. The original phenotypes published in this paper and all microarray data generated in these cohorts are available for public analysis or download using the GeneNetwork database (Species: Mouse, Group: BXD, Type: Adipose mRNA, Liver mRNA, or Muscle mRNA, then select the EPFL datasets). The three historical BXD mRNA datasets, for adrenals, pituitary, and hypothalamus, are also available here [181].

Phenotype data were checked for normality using the Shapiro-Wilk test, with a W-value ≥ 0.80 accepted as approximately normal. Heritability was calculated by one-way ANOVA—the aov() function in R—taking the sum of squares of

within-strain variance divided by the total sum of squares variance. Dot plots are represented as individual measurements, or mean + SEM depending on the figure panel. Dot plots with error bars (e.g. Fig. 2:8D) indicate each dot is a strain average of ~5 individuals. Individual QTL plots consider a suggestive LRS ≥ 12 and significant LRS ≥ 18. Large scale QTL plots (Fig. 2:9D) use LRS ≥ 20 for significance due to multiple testing. Welch's *t*-tests were performed for two-way comparisons between phenotype data, as variances were typically unequal in these comparison groups. Student's *t*-tests were performed for array data, as all data are normally distributed with equal variance. Pearson's *r* is calculated for correlation plots as no outliers were observed. A *p*-value of less than 0.05 was considered the significance threshold for all analyses, except in QTL mapping when correction for multiple testing was used. All BXD phenotype data can be found on **www.genenetwork.org** under the "Type: Phenotype" entry then by searching for "Lisp3".

2.2.6 Acknowledgments

We would like to thank Cristina Cartoni, Sébastien Lamy, and Charles Thomas at the Center of Phenogenomics (CPG, EPFL) for help in establishing and phenotyping the BXD mice, and Jesse Ingels at the University of Tennessee Health Science Center (UTHSC) who genotyped the congenic AHR line. We also thank Lorne Rose and the Molecular Resource Center of Excellence at UTHSC for processing all microarrays. Thanks to Ian Duncan for supplying the mutant *ss D. melanogaster* lines. Discussions with Prof. Stephan Morgenthaler (EPFL) are also acknowledged. E.G.W. was supported by a fellowship from the Fondation Romande pour la Recherche sur le Diabète. J.A. is the Nestlé Chair in Energy Metabolism and the J.A. laboratory is supported by grants from the École Polytechnique Fédérale de Lausanne, the EU Ideas program (Sirtuins; AdG-231138), the Velux Stiftung, the Swiss National Science Foundation (31003A-140780, and CSRII3-136201), NIH (R01AG043930), and the AgingX program of the Swiss Initiative for Systems Biology (SystemsX 51RTP0-151019).

Chapter 3 Expanding Reductive Genetics

This chapter is adapted from three separate projects :

• Muscle : Yamamoto H., **Williams E.G.**, Mouchiroud L., Cantó C., Fan W., Downes M., Héligon C., Barish G.D., Desvergne B., Evans R.M., Schoonjans K., Auwerx J. NCoR1 Is a Conserved Physiological Modulator of Muscle Mass and Oxidative Function, *Cell*, 2011.

• Muscle : Pirinen E., Cantó C**.,** Jo Y.S., Morato L., Zhang H., Menzies K.J., **Williams E.G.,** Mouchiroud L., Moullan N., Hagberg C., Li W., Timmers S., Imhof R., Verbeek J., Pujol A., van Loon B., Viscomi C., Zeviani M., Schrauwen P., Sauve A.A., Schoonjans K., Auwerx J. Pharmacological Inhibition of Poly(ADP-Ribose) Polymerases Improves Fitness and Mitochondrial Function in Skeletal Muscle, *Cell Metabolism*, 2014.

• Liver : Hasenfuss S.C., Bakiri L., Thomsen M.K., **Williams E.G.**, Auwerx J., and Wagner E.F. Regulation of Steatohepatitis and PPARg Signaling by Distinct AP-1 Dimers, *Cell Metabolism*, 2014.

In the following chapter, derived from three independent papers, I applied the BXD genetic reference population to delve further into specific aspects of biomolecular mechanisms which had been identified through prior works. While none of these studies were designed with the BXDs in mind, for all, the mechanism could be better explained and more convincingly validated by applying the results from a large, genetically-diverse population. In the first two sections, on muscle, we observed that either *NCoR1* skeletal muscle knockouts or treatment with PARP inhibitors can improve muscle function, but the precise mechanisms of action were unknown. By examining transcripts covarying with *NCoR1* and *PARP* across the BXDs and similar F2 populations, we could associate a number of molecular signatures of mitochondrial homeostasis (and muscle function), then re-validate these targets in the knockout and PARP-treated populations. In the third section, I used the BXDs in a more exploratory fashion to choose the "best" target gene to examine in knockout and transgenic models. After observing an expected metabolic link in the BXDs—between *PPARg* and fitness (particularly liver health), with a magnified effect in animals on the high fat diet—I was able to select related and novel genes for further mechanistic development.

From these analyses, my colleagues H. Yamamoto, E. Pirinen, S. Hasenfuss, and L. Bakiri were able to generate and analyze the mutant forms of target key genes to perform detailed mechanistic work on their pathways of interest. In these projects, the systems approach was built on top of years of study by many research groups, thus the BXDs were used to (A) validate and (B) extend findings made by GEMM studies, for example by associating new genes with established pathways. Such information not only highlights genes involved in pathways of interest, but it also proves that the mechanisms are not isolated to work in only a single cell line or genotype of mouse, and could be reasonably extrapolated to work in all mice, in turn greatly improving the odds that these pathways are conserved across species, particularly to humans. These three papers have been reduced to the "core content" necessary to understand the scope of the project and my contribution. Please see the above references for full details on the works by Yamamoto (public access), Hasenfuss (public access in January 2015), and Pirinen (public access in June 2015).

3.1 Genetic Alterations to Improve Muscle Function

3.1.1 NCoR1: A Physiological Modulator of Muscle Mass & Oxidative Function

Transcriptional coregulators control the activity of many transcription factors and are thought to have wide-ranging effects on gene expression patterns. In this section, we show that muscle-specific loss of nuclear receptor corepressor 1 (NCoR1) in mice leads to enhanced exercise endurance due to an increase of both muscle mass and of mitochondrial number and activity. The activation of selected transcription factors that control muscle function, such as MEF2, PPARb/d, and ERRs, underpins these phenotypic alterations. NCoR1 levels are decreased in conditions that require fat oxidation, resetting transcriptional programs to boost oxidative metabolism. Knockdown of gei-8, the sole *C. elegans* NCoR1 homolog, also robustly increased muscle mitochondria and respiration, suggesting strong evolutionary conservation of NCoR1 function. Collectively, the data suggest that NCoR1 plays an adaptive role in muscle physiology and

that interference with NCoR1 action could be used to improve muscle function.

Transcription factors are key mediators in homeostatic circuits, as they process environmental signals into transcriptional changes [182, 183]. Transcriptional coregulators have recently emerged as equally important modulators of such adaptive transcriptional responses. The fact that the activity of coactivators and corepressors is tightly regulated through the spatial and temporal control of their expression and activity levels opens another avenue to adapt transcription to environmental cue [184-187]. Interestingly, many of these coregulators do not operate in isolation but are part of large multiprotein complexes that integrate complex signaling pathways. The convergence of an elaborate coregulator network on the peroxisome proliferator-activated receptor (PPAR) coactivator a (PGC)-1a illustrates this principle well, as its activity depends on several other coregulators, including the steroid receptor coactivators, NR-interacting protein 1 or RIP140, CREB-binding protein, p300, protein arginine methyltransferase 1, general control of amino acid synthesis 5, and SIRT1 [188, 189].

The corepressor NCoR1 and the silencing mediator for retinoid and thyroid hormone receptor (SMRT or NCoR2) also act as cofactor scaffolding platforms. NCoR1 and SMRT hardwire corepressor pathways that incorporate several deacetylases (including class I [HDAC3], class II [HDAC4, 5, 7, and 9], and class III [SIRT1] HDACs), transducin beta-like 1 (TBL1) and TBLR1, two highly related F box/WD40-containing factors, and the G protein pathway suppressor 2 (reviewed in [190]). Because germline *NCoR1*$^{-/-}$ and *SMRT*$^{-/-}$ mice are embryonically lethal [191, 192], information on the role of these proteins in adult physiology is limited. Studies of mice with mutations in the NR interaction domains (RIDs) 1 and 2 of SMRT (SMRTmRID), which solely disrupts its interaction with NRs, indicated that lethality of *SMRT*$^{-/-}$ mice is caused by non-NR transcription factors [193]. Work in 3T3-L1 cells in which NCoR1 or SMRT expression was reduced by RNA interference demonstrated that they repress adipogenesis by inhibiting PPARg [194]. In line with this, adipogenesis was enhanced in mouse embryonic fibroblasts (MEFs) from SMRTmRID mice [193]. Interestingly, SIRT1 is also part of the NCoR1/SMRT complex and contributes to the inhibition of PPARg [195].

Contrary to adipose tissue, the function of NCoR1/SMRT in skeletal muscle has not yet been established. We here report the generation and characterization of muscle-specific *NCoR1*$^{-/-}$ (*NCoR1*$^{skm-/-}$) mice, which display a remarkable enhancement of exercise capacity. This was the result of increased muscle mass and a muscle fiber type shift toward more oxidative fibers, coordinated by the induction of genes involved in mitochondrial biogenesis and function, ensuing from the activation of PPARb/d, the estrogen-related receptors (ERRs), and myocyte-specific enhancer factor 2 (MEF2). Worms with a muscle-selective knockdown of *gei-8*, the sole *C. elegans* NCoR1 homolog, also had improved mitochondrial activity. These data combined with the specific reduction in the expression levels of NCoR1, but not SMRT, in situations of enhanced fat oxidation establish NCoR1 as a key physiological regulator of muscle mass and function.

3.1.2 NCoR1$^{skm-/-}$ Mice Have Improved Muscle Function and Exercise Performance

Given the embryonic lethality of germline *NCoR1*$^{-/-}$ mice [191], we generated a floxed NCoR1 mouse line in which exon 11 of the *NCoR1* gene [196] was flanked with LoxP sites, priming it for subsequent deletion using the Cre-LoxP system. These mice, bearing floxed *NCoR1* L2 alleles, were then bred with a skeletal muscle (skm)-specific Cre driver (human a-skeletal actin promoter) [197] to yield *NCoR1*$^{skm-/-}$ and *NCoR1*$^{skm+/+}$ mice. As expected, *NCoR1* mRNA expression was significantly decreased in soleus, gastrocnemius, and quadriceps and modestly reduced in the heart muscle of *NCoR1*$^{skm-/-}$ mice, but not altered in other tissues (Fig. 3:1A). No compensatory induction of the related corepressor SMRT/NCoR2 [198] was observed (Fig. 3:1A). We also tried to determine NCoR1 protein levels in muscle but failed to detect the endogenous protein with the currently available NCoR1 antibodies.

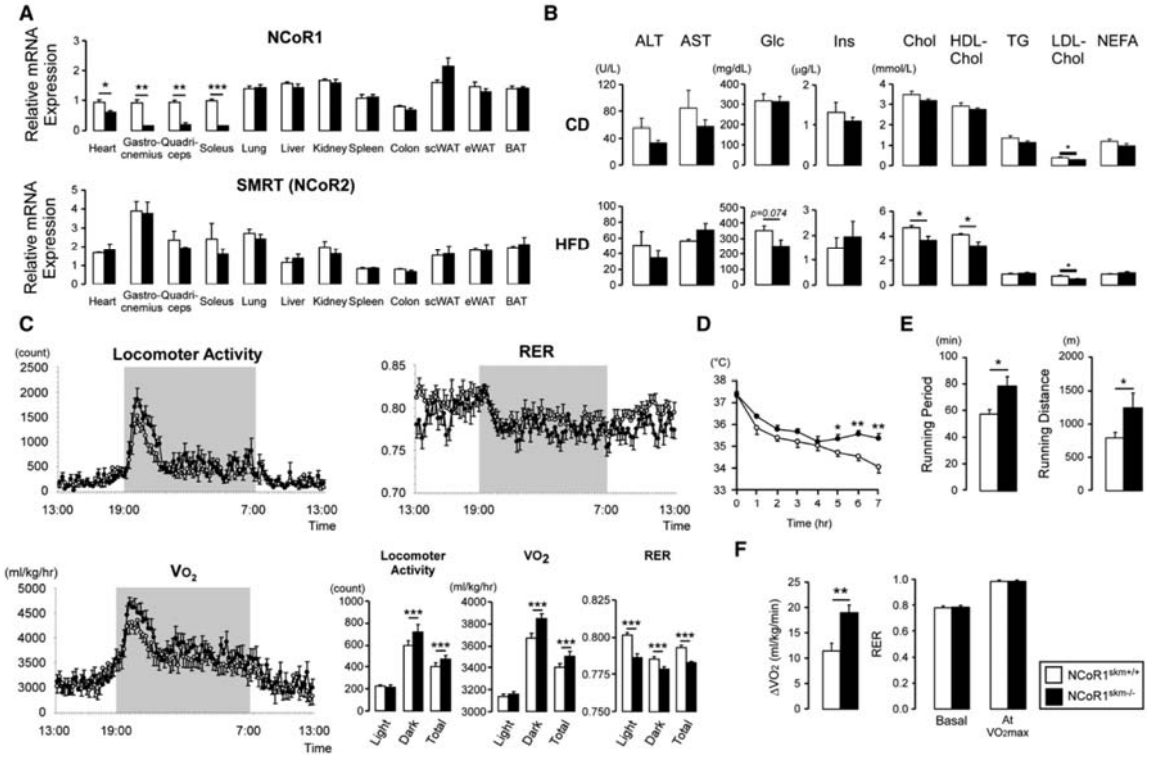


Figure 3:1 **Metabolic Phenotypes of *NCoR1*$^{skm-/-}$ Mice**

(A) mRNA levels of *NCoR1* and *Smrt* in different tissues, determined by qRT-PCR. Values were normalized to 36B4. *n* = 8–10/group. **(B)** Biochemical analysis of plasma from *NCoR1*$^{skm+/+}$ and $^{skm-/-}$ mice after 6 hr fasting (*n* = 8) either fed CD (top) or HFD (bottom). **(C)** Circadian activity, measured as total locomotor activity, and energy expenditure were evaluated by the measurement of oxygen consumption (VO_2) and by the calculation of the RER over a 24 hr period after 12 weeks of HFD. *n* = 12. **(D)** Body temperature was measured for 7 hr in mice exposed to 4°C after 18 weeks of HFD. *n* = 7–8. **(E and F)** Exercise experiments, measuring running time and distance until exhaustion (endurance exercise) **(E)** and the increment of VO_{2max} during exercise and RER levels at basal and VO_{2max} condition **(F)** were performed after 14 and 17 weeks HFD. All bar graphs represent the average + SEM for each group.

NCoR1$^{skm-/-}$ mice were indistinguishable from *NCoR1*$^{skm+/+}$ mice upon visual inspection, and no gross organ anomalies were revealed upon autopsy. The relative mass of the soleus muscle was higher, whereas the mass of the gastrocnemius showed a trend toward an increase, though did not reach statistical significance. The soleus was also more intensely red, and there were larger sections with reddish color in the gastrocnemius in *NCoR1*$^{skm-/-}$ mice. Body weight evolution and food intake of male *NCoR1*$^{skm-/-}$ and *NCoR1*$^{skm+/+}$ mice after weaning were comparable for the cohorts fed either CD or HFD. On CD, carbohydrate and lipid profiles were similar except for LDL cholesterol, which was reduced in *NCoR1*$^{skm-/-}$ mice (Fig. 3:1B). In addition to the lower LDL cholesterol on CD, total and HDL cholesterol levels were also reduced in *NCoR1*$^{skm-/-}$ mice on HFD (Fig. 3:1B). Furthermore, glucose edged down ($p = 0.074$) in the wake of similar insulin levels on HFD. The slightly reduced area under the curve (AUC) in intraperitoneal glucose tolerance test (IPGTT) and the delayed recovery from hypoglycemia during intraperitoneal insulin tolerance test (IPITT) in mutant mice on HFD may also suggest a discrete improvement in insulin sensitivity but without a clear impact on glucose tolerance.

We next evaluated energy expenditure by indirect calorimetry and actiometry in CD- and HFD-fed mice (Fig. 3:1C). Total locomotor activity was significantly higher in *NCoR1*$^{skm-/-}$ mice. Consistent with this, O_2 consumption (VO_2) was increased under both CD and HFD. Interestingly, the *NCoR1*$^{skm-/-}$ mice displayed a marked decrease in the respiratory exchange ratio (RER) on a HFD (Fig. 3:1C), indicating an enhanced use of fat as a main energy source. *NCoR1*$^{skm-/-}$ mice were also more cold tolerant, as they maintained their body temperature better when exposed to 4°C over a 7 hour period (Fig. 3:1D). Exercise performance was also strikingly improved in *NCoR1*$^{skm-/-}$ mice (Fig. 3:1E and 3:1F). In endurance exercises, *NCoR1*$^{skm-/-}$ mice ran for a significantly longer time and distance before exhaustion (Fig. 3:1E). The in-

crease of the VO_2 values (ΔVO_2) during exercise and the maximal ability to utilize oxygen during exercise (VO_{2max}), which critically determines the endurance performance of skeletal muscle, was slightly higher in *NCoR1*$^{skm-/-}$ mice on both CD and HFD (Fig. 3:1F). Despite the moderate reduction in *NCoR1* mRNA levels in cardiac muscle of *NCoR1*$^{skm-/-}$ mice, heart rate, blood pressure, cardiac morphology, and function were not changed.

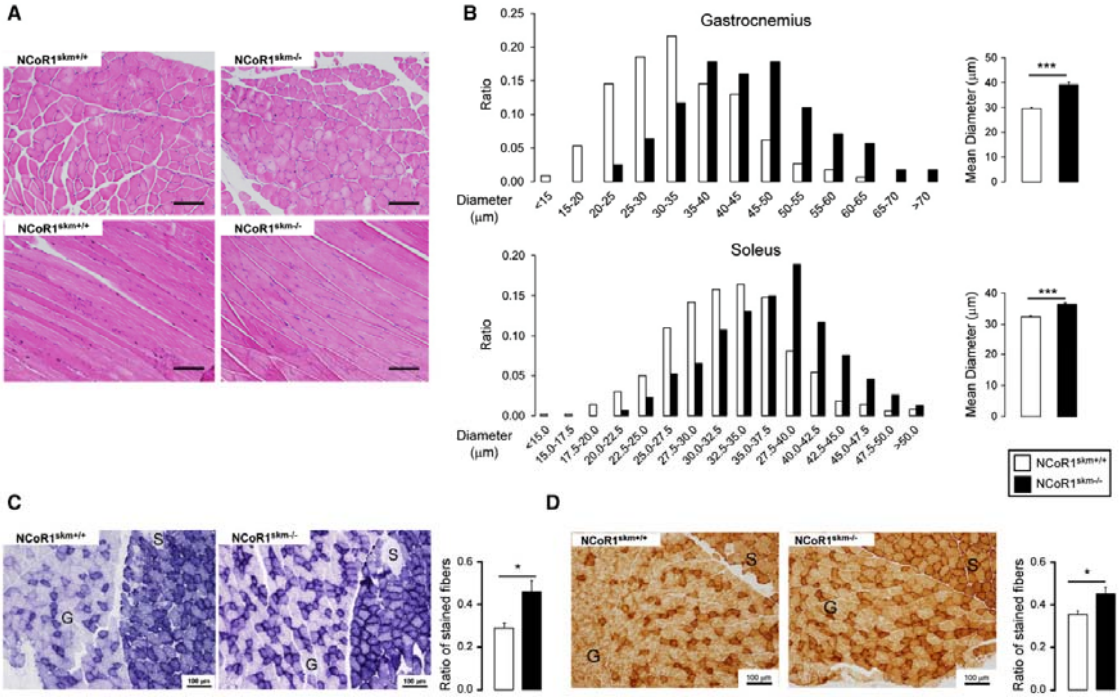


Figure 3:2 **Histological Analyses of the Muscles of Control and *NCoR1*$^{skm-/-}$ Mice**

(A) Histological analysis of gastrocnemius sections stained with hematoxylin and eosin. **(B)** Distribution and mean diameter of muscle fibers in gastrocnemius and soleus. **(C and D)** Histological analysis of gastrocnemius and soleus sections by succinate dehydrogenase **(C)** and cytochrome C oxidase staining **(D)**. S, soleus; G, gastrocnemius. The ratio of the stained fibers is indicated in the graph at right. Data are expressed as mean ± SEM.

The enhanced exercise capacity, associated with the increase in overall muscle mass and change in muscle appearance, led us to examine muscle morphology. Upon staining muscles with hematoxylin/eosin or toluidine blue, not only was the diameter of single muscle fibers larger, but also the connective tissue between muscle bundles was less abundant in *NCoR1*$^{skm-/-}$ mice (Fig. 3:2A and 3:2B). The increased number of intensely stained fibers upon succinate dehydrogenase (SDH) and cytochrome oxidase (COX) (Fig. 3:2C and 3:2D) staining further testified to increased mitochondrial activity in the *NCoR1*$^{skm-/-}$ gastrocnemius. Two mitochondrial DNA markers, cyclooxygenase 2 (*Cox2*) and 16S ribosomal RNA, normalized by genomic DNA markers (uncoupling protein 2 [*Ucp2*] and hexokinase 2 [*Hk2*]) were both significantly higher in *NCoR1*$^{skm-/-}$ muscle, indicative of increased mitochondrial content (Fig. 3:3B). This observation was also underscored by electron microscopy, which revealed more abundant and larger mitochondria with normal structure (Fig. 3:3A). Immunohistochemical analysis of the myosin heavy-chain (MyHC) isoforms [199] demonstrated a decreased number of the more glycolytic MyHC2b fibers, with a concomitant increase in the number of more oxidative MyHC2x and 2a fibers in the *NCoR1*$^{skm-/-}$gastrocnemius (Fig. 3:3C). This observation was consolidated by analysis of MyHC isoform mRNAs, which indicated an increased expression of the mRNAs of *MyHC2x* and *2a* (more oxidative fibers) compared to that of *MyHC2b* (more glycolytic) in both *NCoR1*$^{skm-/-}$ gastrocnemius and quadriceps (Fig. 3:3D). In quadriceps, but not gastrocnemius, the expression of *MyHC1* mRNA was also increased. Finally, staining of platelet-endothelial cell adhesion molecule (PECAM)-1, an endothelial cell marker of angiogenesis and tissue vascularization that contributes to enhanced myocellular aerobic capacity, also increased in *NCoR1*$^{skm-/-}$ muscle.

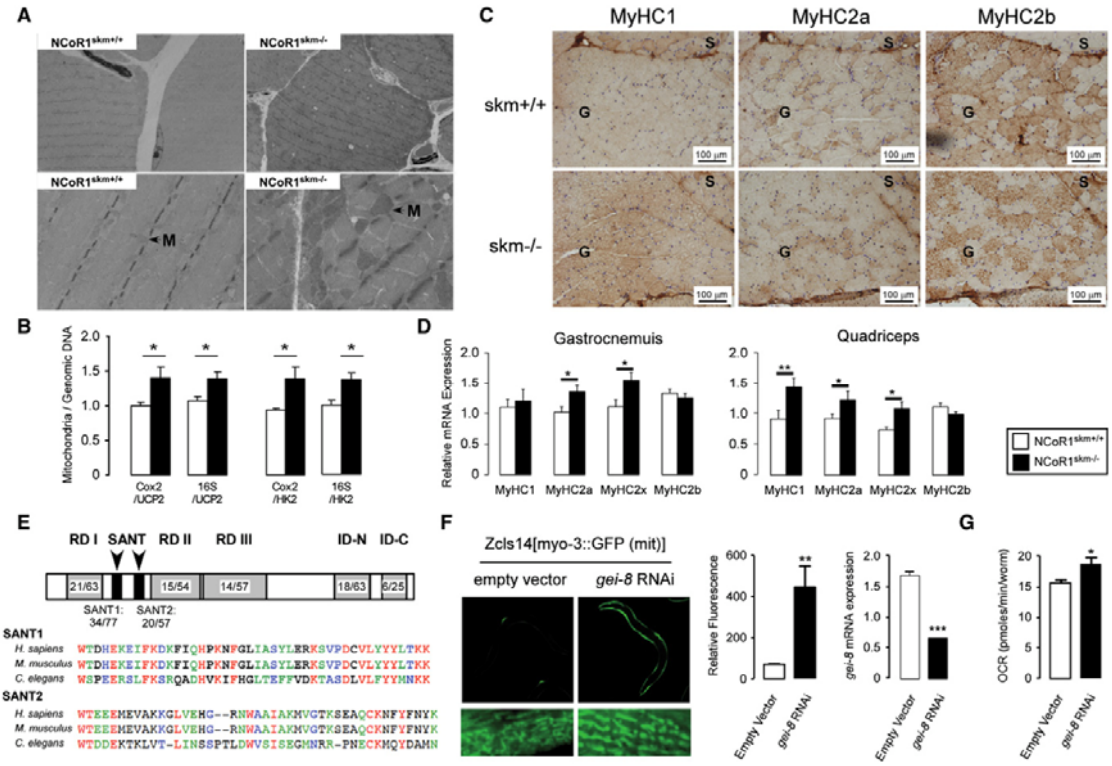


Figure 3:3 **Histological Analyses of the Muscles of Mice and *C. elegans***

(A) Transmission electron microscopy of nonoxidative fibers of *NCoR1*$^{skm+/+}$ and $^{skm-/-}$ gastrocnemius. M, mitochondria. **(B)** Relative mitochondrial DNA content (*Cox2* or *16S*) in gastrocnemius was measured and normalized by genomic DNA content (*Ucp2* and *Hk*). *n* = 6. **(C)** Representative MyHC1, 2a, and 2b immunohistochemical detection on serial sections of the soleus and gastrocnemius. **(D)** Subtypes of MyHCs in gastrocnemius and quadriceps were analyzed by qRT-PCR in *NCoR1*$^{skm+/+}$ and $^{skm-/-}$ mice. *n* = 6. **(E)** Identity/similarity (%) in the sequences of *gei-8* with mammalian NCoR1. Multiple alignments of SANT domains of NCoR1 homologs. Color identifies similarity. **(F)** Representative pictures of the effects of RNAi-mediated knockdown of *gei-8* on mitochondrial morphology and number in a *C. elegans* strain carrying a mitochondrial GFP reporter driven by the muscle-specific *myo-3* promoter (left). Quantification of the mitochondrial induction by fluorescence upon *gei-8* knockdown (middle) and of the efficacy of the RNAi-mediated *gei-8* knockdown by qRT-PCR analysis (right). **(G)** Muscle-specific RNAi inhibition of *gei-8* enhances respiration in *C. elegans*. Data are expressed as mean ± SEM.

To investigate whether the effects of NCoR1 deficiency are evolutionary conserved, we took advantage of the power of *C. elegans* genetics. A protein blast search indicated that *GEX-interacting protein family member 8* (*gei-8*) is the only putative NCoR1 homolog in the *C. elegans* genome. Further analysis showed that the total amino acid sequence of *gei-8* is 43% homologous to mouse NCoR1 and contained conserved SANT domains (switching-defective protein 3 [Swi3], adaptor 2 [Ada2], nuclear receptor corepressor [N-CoR], transcription factor [TF] IIIB; 34% identical/77% similar for SANT1; 20% identical/ 57% similar for SANT2) (Fig. 3:3E). Other important functional domains (repressor domain [RD] and nuclear receptor interaction domain [ID]) were also conserved (Fig. 3:3E). Upon the robust *gei-8* knockdown in worms expressing a mitochondrial GFP reporter driven by the muscle-specific *myo-3* promoter, a striking enlargement of the mitochondria was observed in body wall muscle (Fig. 3:3F). This result is not due to an indirect effect on transcriptional activity through the *myo-3* promoter because no increase in GFP expression is observed with another strain carrying the Pmyo-3::GFP reporter. We also measured O_2 consumption in NR350 transgenic worms fed with *gei-8* dsRNA. NR350 worms lack *rde-1*, an essential component of the RNAi machinery encoding a member of the PIWI/STING/Argonaute family, in all tissues except the body wall muscle in which the wild-type *rde-1* gene has been rescued using the *hlh-1* promoter [200]. Consistent with the effects observed in the mouse, also the muscle-specific knockdown of *gei-8* enhanced O_2 consumption in these NR350 worms (Fig. 3:3G), suggesting that the function of *gei-8*

in controlling mitochondrial metabolism is conserved through evolution.

3.1.3 NCoR1 Is Associated With Muscle Function in Mouse GRPs

After establishing these striking mitochondrial effects of NCoR1 in mice and worms, we exploited a complementary systems genetics approach to evaluate *NCoR1* molecular coexpression partners in the BXD and BH F2 mouse populations [201, 202]. Expression of *NCoR1* in these two panels of genetically heterogeneous mice varied by ± 1.5-fold between cases in both lung and muscle (Fig. 3:4A). A large number of transcripts covaried significantly with *NCoR1* in the different mice lines belonging either to the BH cross or BXD strains. Most distinctively, a large fraction of the top covariates were negative (~50% of top 500 correlates, compared to ~20% for a random gene). In skeletal muscle from the BH F2s (*n* = 124 females), strong covariates of *NCoR1* include *Mef2d*, myoglobin (*Mb*), muscle creatine kinase (*Mck*), and glucose transporter type 4 (*Glut4*) [203]. A similar analysis of lung tissue from the BXD (*n* = 51 strains) includes genes such as cytochrome c (*Cycs*), citrate synthase (*Cs*), pyruvate dehydrogenase kinase 4 (*Pdk4*), uncoupling protein 3 (*Ucp3*), vascular endothelial growth factor b (*Vegfb*), and long-chain acyl-CoA dehydrogenase (*Lcad*) (Fig. 3:4B) [204]. These analyses significantly extend the number of *NCoR1* targets and covariates, with several of them being consistent with increased mass and mitochondrial biogenesis observed in *NCoR1*$^{skm-/-}$ muscle, and provided hypotheses for further mechanistic validation. This initial set of *NCoR1* covariates (Fig. 3:4B) was then included together with other potential candidates for qRT-PCR analysis in mixed fiber muscle such as the gastrocnemius (Fig. 3:4C).

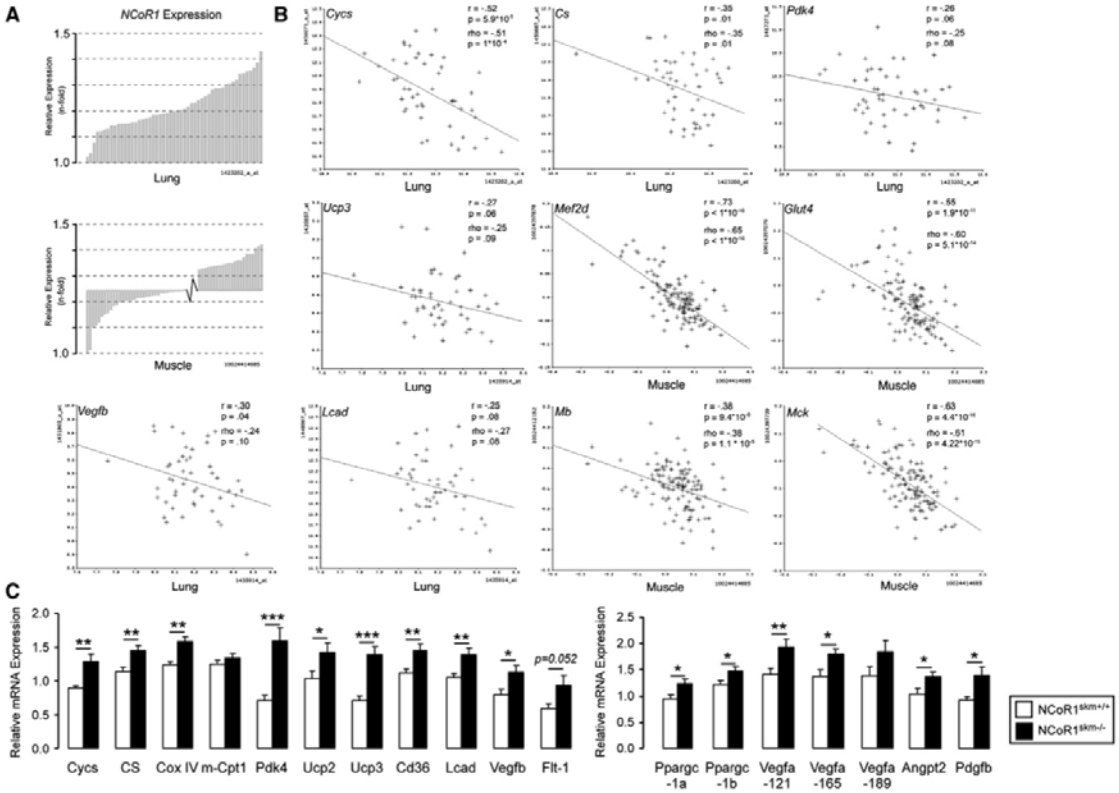


Figure 3:4 **Identification of *NCoR1*-Correlated Genes**

(A) Expression of *NCoR1* mRNA in lung tissue of the different BXD strains (top) and in muscle tissue from an F2 intercross between C57BL/6J and C3H/HeJ (bottom). Natural expression variation across the animals is ~ 1.5-fold in each tissue. **(B)** Pearson's *r* and Spearman's rank correlation coefficient, *rho*, were calculated with corresponding *p* values for the mRNA covariation between *NCoR1* and genes involved in oxidative phosphorylation (*Cycs*, *Cs*, and *Pdk4*), mitochondrial uncoupling (*Ucp3*), fatty acid metabolism (*Lcad*), angiogenesis (*Vegfb*), glucose uptake (*Glut4*), and myogenesis (*Mef2d*, *Mb*, and *Mck*). The tissue from which data were generated is indicated. **(C)** Gene expression analysis by qRT-PCR in *NCoR1* $^{skm+/+}$ and $^{skm-/-}$ gastrocnemius. *n* = 10. Data are expressed as mean ± SEM.

While the mRNAs of most relevant NRs were unchanged, mRNA levels of *PGC-1a* and *b* (*Ppargc1a* and *1b*) increased. Several genes involved in mitochondrial function, including those encoding for proteins involved in TCA cycle and oxidative phosphorylation (*Cs*, cytochrome c oxidase subunit IV [*CoxIV*], *Pdk4*), uncoupling (*Ucp2* and *Ucp3*), fatty acid uptake, and metabolism (*Cd36* and *Lcad*), were robustly induced in *NCoR1*$^{skm-/-}$ muscle. In addition, mRNA levels of *Vegfb* and its receptor *Flt1*, which regulates *trans*-endothelial fatty acid transport [205], were also induced. Interestingly, the expression of hypoxia-inducible factor (*Hif*) 1a and of its targets, glucose transporter 1 (*Glut1*), fibroblast growth factor (*Fgf*), and *Fgf*-receptor 2 (*Fgfr2*), were unchanged, whereas all three *Vegfa* isoforms, i.e., *Vegfa*-121, -165, and -189, were induced in *NCoR1*$^{-/-}$ quadriceps, gastrocnemius, and soleus (Fig. 3:4C and 3:5B). Together with this increase in *Vegfa*, both *Angpt2* and *Pdgfb* mRNA levels were induced (Fig. 3:4C), suggesting that myocellular aerobic capacity is facilitated by an HIF1a-independent angiogenic pathway in *NCoR1*$^{skm-/-}$ mice [206]. Several genes whose expression is changed in the absence of NCoR1 are PPARb/d and/or ERR targets (Fig. 3:4). Because the expression of PPARb/d and/or ERR was unchanged in *NCoR1*$^{skm-/-}$ mice, a direct effect of NCoR1 on the expression of these targets through the activation of these NRs was expected. As cases in point to demonstrate the recruitment of NCoR1 to these genes, we selected the murine *Ucp3* and *Pdk4* promoters, which contain three PPAR responsive elements (PPREs) (Fig. 3:5A) and extended NR half-sites (NR1/2), known to bind members of the ERR subfamily (Fig. 3:5E) [207], respectively. We first used NIH 3T3 cells in which an epitope-tagged version of NCoR1 (NCoR1-FLAG) was expressed. The two PPREs adjacent to the *Ucp3* transcription start site recruited NCoR1 more efficiently, compared to two control sequences in the *Gapdh* and *Ucp3* promoter that lack PPREs (Fig. 3:5B, left). Likewise, NCoR1 bound avidly to the mouse *Pdk4* promoter NR1/2 site in transfected NIH 3T3 cells (Fig. 3:5F, left). Although there is a two nucleotide difference in NR1/2 site of the human *Pdk4* promoter (Fig. 3:5E), NCoR1 and ERRa were also recruited to this site in human HEK293 cells (Fig. 3:5H).

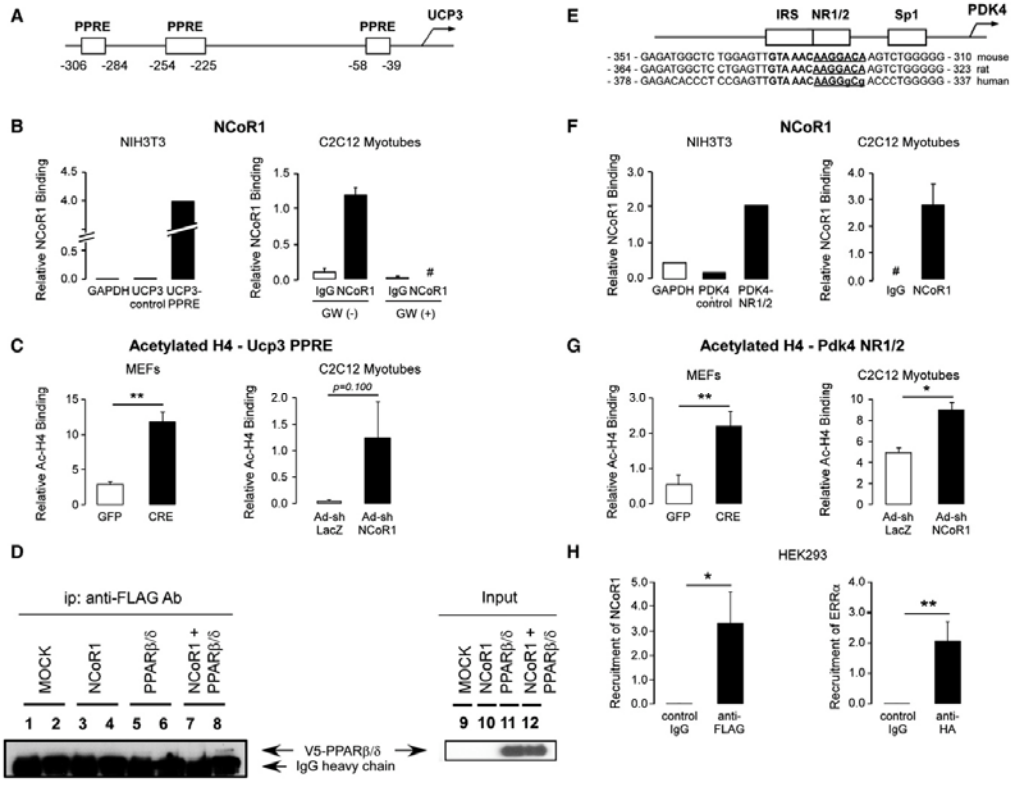


Figure 3:5 **Increased PPARB/D and ERR Activity in *NCoR1*$^{skm-/-}$ Muscle**

(A, B, E, F, and H) NCoR1 recruitment to the PPREs on mouse *Ucp3* promoter **(B)** and to the ERR-RE on human **(H)** and mouse **(F)** *Pdk4* promoter determined by ChIP in NIH 3T3 cells transfected with an NCoR1-FLAG vector or in C2C12 myotubes. A schematic of the promoters of the *Ucp3* **(A)** and *Pdk4* **(E)** genes and the sequence alignment of the mouse, rat, and human *Pdk4* promoter is also shown to highlight the conservation of the NR1/2 (or ERR-RE) **(E)**. Boxes indicate putative PPREs in the *Ucp3* and the NR1/2, IRS, and Sp1 in the *Pdk4* promoter. ChIP experiments for the *Ucp3* promoter were performed in C2C12 myotubes both before and 6 hr after addition of a PPARb/d agonist (100 nM GW501516). #, not detected. ChIP experiments in

HEK293 cells transfected with FLAG-NCoR1 and HA-ERRa vector **(H)**. **(C and G)** Binding of acetylated histone 4 (H4) to the PPREs on the *Ucp3* and to the NR1/2 on the *Pdk4* promoters in ChIP assays, using either immortalized *NCoR1*$^{L2/L2}$ MEFs, infected with an adenovirus either expressing GFP or Cre recombinase, or C2C12 myotubes infected with the Ad-shNCoR1 virus. Representative data are shown from three experiments. **(D)** Interaction between PPARb/d and NCoR1 determined by in vitro co-IP experiments from HEK293 cells in which NCoR1-FLAG and/or V5-PPARb/d are expressed. IP was performed with control IgG (lanes 1, 3, 5, and 7) or anti-FLAG antibody (lanes 2, 4, 6, and 8), and the immunoblot was developed with an anti-V5 antibody. PPARb/d coimmunoprecipitated by the anti-FLAG antibody is indicated by an arrow. Input samples are shown in lanes 9–12. Data are expressed as mean ± SEM.

The increased muscle mass observed in *NCoR1*$^{skm-/-}$ mice indicated that the absence of *NCoR1* not only induced oxidative metabolism, but also stimulated myogenesis. In line with this, mRNA levels of two markers of myogenesis, *Mb* and *Mck*, were increased in the quadriceps of *NCoR1*$^{skm-/-}$ animals (Fig. 3:6A). Among several myogenic regulatory factors, only the expression of two *Mef2* family members, i.e., *Mef2c* and *Mef2d*, negatively correlated with *NCoR1* expression in our systems genetics analysis (Fig. 3:4). The selective induction of *Mef2c* and *Mef2d* mRNA was furthermore confirmed by qRT-PCR of *NCoR1*$^{skm-/-}$ gastrocnemius and quadriceps, whereas no changes were found in *MyoD*, *Myf5*, and *myogenin* mRNA (Fig. 3:6A). The activity of MEF2 family members is not only controlled by their expression levels, but is also modulated by their acetylation status. MEF2 is acetylated and activated by p300, whereas it is deacetylated by HDAC3 and HDAC4, which are part of the NCoR1 corepressor complex [208, 209]. Because the expression of the *Mef2d* isoform is most prominently correlated with *NCoR1* expression, we investigated MEF2D acetylation in gastrocnemius and found that its acetylation levels were enhanced in *NCoR1*$^{skm-/-}$ mice (Fig. 3:6A and 3:6B).

We then compared the acetylation of MEF2D in floxed *NCoR1*$^{L2/L2}$ MEFs infected with an adenovirus expressing either GFP as control or Cre-recombinase to reduce NCoR1 protein expression (Fig. 3:6C, left). Whereas MEF2D protein levels were stable in *NCoR1*$^{-/-}$ MEFs, perhaps due to the more acute nature of the deletion, MEF2D was robustly hyperacetylated when NCoR1 levels were attenuated (Fig. 3:6C, right). Likewise, a slight but consistent MEF2D hyperacetylation was observed in C2C12 myotubes infected with Ad-shNCoR1 to knock down *NCoR1* expression (Fig. 3:6D), further underscoring the importance of MEF2D deacetylation by the NCoR1 complex. Given the induction of MEF2D expression and its hyperacetylation and consistent with our systems genetics analysis (Fig. 3:4B), mRNA levels of the MEF2 targets, *Mb* and *Mck*, were robustly induced in *NCoR1*$^{skm-/-}$ gastrocnemius (Fig. 3:6A). Silencing of *NCoR1* in C2C12 myotubes also resulted in a similar induction of several MEF2 target genes, including *Mb*, *Mck*, *Glut4*, *c-Jun*, *Nur77*, *PGC-1*a, and *PGC-1*b (Fig. 3:6E). In line with these data, endogenous NCoR1 was readily detected on MEF2-binding sites on these target promoters in C2C12 myotubes, as illustrated for the *Mb* promoter (Fig. 3:6F). The induction of these MEF2 targets by NCoR1 knockdown was furthermore accompanied by H4K16 and global H4 hyperacetylation on their promoters (e.g., *Mb*, *Glut4*, and *Mck*) (Fig. 3:6G).

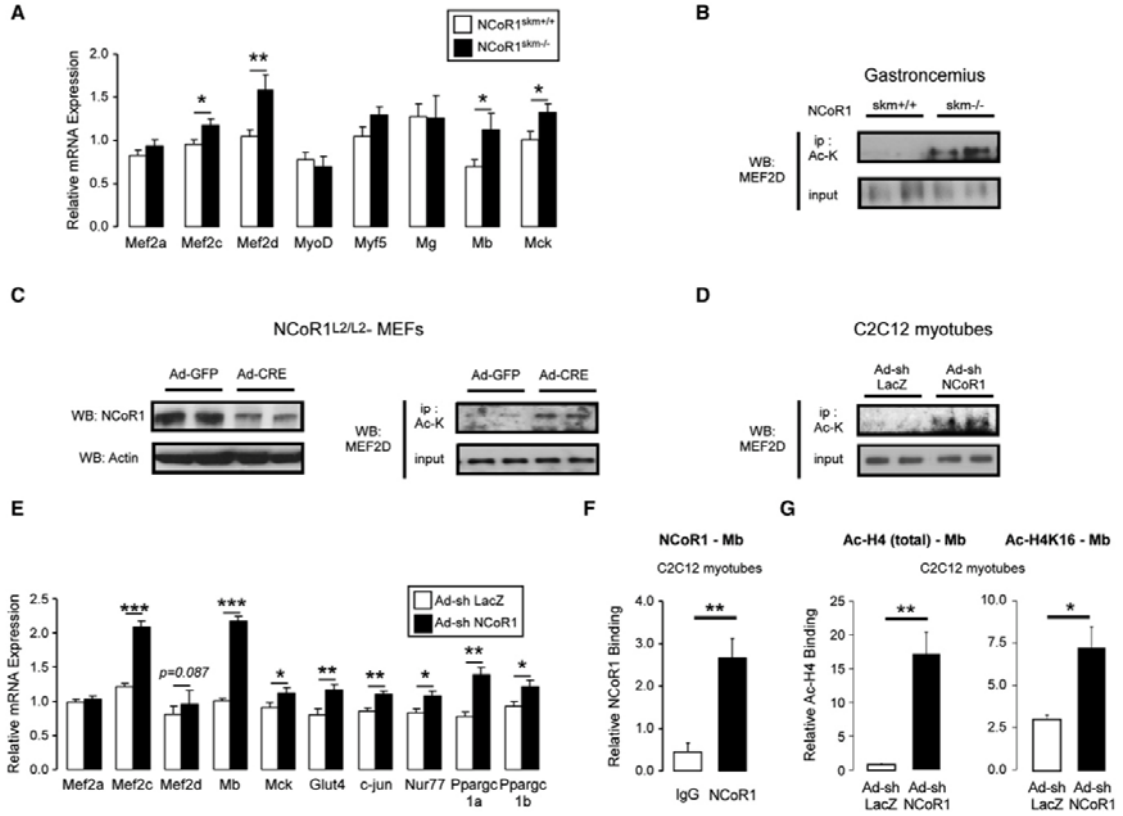


Figure 3:6 **Enhanced MEF2 Activity in NCoR1$^{skm-/-}$ Muscle**

(A) Gene expression of myogenesis-related genes was measured by qRT-PCR in *NCoR1*$^{skm+/+}$ and $^{skm-/-}$ quadriceps. *n* = 10. (B–D) Acetylation levels of MEF2D were determined by western blot after immunoprecipitation with an Ac-Lys Ab from gastrocnemius **(B)**, from *NCoR1*$^{L2/L2}$ MEFs infected with Ad-GFP or Ad-Cre recombinase **(C, right)**, and from C2C12 myotubes infected with Ad-shLacZ or Ad-shNCoR1 **(D)**. MEF2D expression in total protein extracts was shown in the bottom panels. The expression of NCoR1 and actin in *NCoR1*$^{L2/L2}$ MEFs was also shown **(C, left)**. **(E)** MEF2 target mRNAs determined by qRT-PCR in C2C12 myotubes infected with either Ad-shLacZ or Ad-shNCoR1. *n* = 6. **(F)** NCoR1 recruitment to the MEF2 site of the mouse *Mb* promoter determined by ChIP in C2C12 myotubes. **(G)** Binding of either global acetylated histone 4 (H4) or H4 acetylated on K16 (H4K16) to the MEF2 site of the *Mb* gene was evaluated by ChIP from C2C12 myotubes infected as in **(E)**. Data are expressed as mean ± SEM.

3.1.4 NCoR1 Muscle Discussion

The increased muscle mass, associated with a strikingly improved exercise capacity, is the most prominent phenotypic outcome of the muscle-specific *NCoR1* gene deletion. The enhanced exercise capacity is associated with a reprogramming of glycolytic to more oxidative muscle fibers and a corresponding stimulation of oxidative mitochondrial metabolism, indicative of an improved intrinsic quality of the muscles. The increased muscle quantity and oxidative profile in *NCoR1*$^{skm-/-}$ mice also contribute to the slight improvement in metabolic parameters after HFD feeding and to the cold resistance subsequent to shivering thermogenesis [210]. Combined, these properties suggest that NCoR1 acts as a master modulator of mitochondrial metabolism in the muscle, a hypothesis bolstered by the fact that the inhibition of the single worm NCoR homolog, *gei-8*, also robustly boosts muscle oxidative mitochondrial metabolism in *C. elegans*. The evolutionary conservation of the structure and function of this corepressor makes it tempting to speculate that NCoR may have evolved to facilitate metabolic adaptation of the mitochondria to energy availability, as has been described for other cofactors as SIRT1 [210].

As to how the absence of NCoR1 in the muscle achieves these remarkable effects, it is important to recall that NCoR1 docks histone deacetylases, such as HDAC3 [211] and SIRT1 [195]. The use of mice with a point mutation in the

NCoR1 deacetylase activation domain, incapacitating its interaction with HDAC3, indicated that this interaction is a nodal point in epigenetic regulation [211, 212]. The physiological alterations in the *NCoR1*$^{skm-/-}$ muscle together with the study of the expression correlates of *NCoR1* suggested that, rather than the generalized transcriptional activation expected upon ablation of a corepressor platform protein, only a small set of transcriptional pathways was selectively affected.

Several transcription factors control muscle differentiation and development, including *MyoD*, myogenin, *Myf5*, *Myf6*, and *Mef2* [213-216]. *Mef2* especially caught our attention, as only its expression negatively correlated with *NCoR1* (Fig. 3:4B) and was induced in *NCoR1*$^{skm-/-}$ muscle (Fig. 3:6A). *Mef2* is key for muscle development and also participates in muscle stress response and remodeling in adulthood, such as occurs during muscle fiber type switch [216, 217]. *Mef2* activity is not only controlled at the level of its expression, but also by a wide range of intracellular signaling pathways and interacting coregulator molecules (reviewed in [216]). The histone acetyltransferases, p300/CBP, bind, acetylate, and activate MEF2 [208, 214], whereas class I (HDAC3 [218]), class II (HDAC4, 5, 7, and 9 [214, 219]), and class III HDACs (SIRT1 [220]) all are reported to interact with MEF2 and prevent the activation of its target genes [214]. The absence of NCoR1 would hence favor the acetylation and activation of MEF2. The hyperacetylation of MEF2D and histone 4 in *NCoR1*-deficient MEFs, C2C12 myotubes, and muscles, which translates into the induction of *Mef2* targets and the gain of muscle mass in *NCoR1*$^{skm-/-}$ mice, is fully in line with this idea (Fig. 3:6).

The increased expression of genes related to fatty acid catabolism and mitochondrial respiration underlies the oxidative mitochondrial changes observed in the *NCoR1*$^{skm-/-}$ mice. In the muscle, the expression of these gene sets is tightly controlled by NRs belonging to the PPAR and ERR families [221]. PPARb/d, the predominant PPAR isoform in oxidative fibers, regulates oxidative capacity and enhances slow fiber-type function, resulting in improved exercise capacity and metabolic protection [222-224]. Among the ERRs, mainly ERRa and ERRg seem to be involved in the coordination of muscle energy homeostasis (reviewed in [225, 226]). In line with this, a genome-wide location analysis of ERRa and ERRg identified binding sites in genes of a large number of mitochondrial proteins [227], and studies in mouse models show that they coordinate many aspects of muscle oxidative metabolism, including endurance capacity [225, 226, 228]. Several features of the *NCoR1*$^{skm-/-}$ mice are suggestive of the activation of PPARb/d and ERR. The demonstration that NCoR1 is recruited to the PPREs in the *Ucp3* and the NR1/2 in *Pdk4* promoters and that histone 4 is hyperacetylated on these promoters when NCoR1 is absent suggests that coactivators now activate the transcription of these genes in a fashion unopposed by the NCoR1 corepressor platform. Given that PGC-1a is a key coactivator of PPARb/d and ERR transcriptional programs (reviewed in [188]), it is no surprise that several phenotypic features, ranging from similarities in gene expression patterns over the induction of mitochondrial oxidative metabolism and exercise capacity to HIF1a-independent angiogenesis, are shared between mice that lack NCoR1 or that overexpress PGC-1a [206, 229].

NCoR1 action is tightly regulated by various physiological challenges, and this is achieved through at least two different mechanisms. First, nuclear levels of NCoR1 are regulated. Our results show how insulin, which stimulates glucose oxidation at the expense of fatty acid oxidation, increases NCoR1 levels in the nucleus, enabling it to subsequently repress lipid oxidation genes. Interestingly, this effect of insulin on nuclear NCoR1 accumulation is consistent with the positive effect of mTORC1 on nuclear NCoR1 accumulation, recently reported in hepatocytes [230]. The second mechanism involves the modulation of NCoR1 expression. Exposing cells to media with low glucose and/or high fatty acid levels reduces specifically *NCoR1* (but not *SMRT*) mRNA and protein levels, ultimately derepressing genes that control oxidative lipid metabolism. Likewise, endurance exercise, fasting, high-fat feeding, and aging—conditions paired with increased fat oxidation—also are characterized by attenuated muscle *NCoR1* mRNA expression. Although we are unable to detect NCoR1 protein in the muscle with the currently available antibodies, our data in adipose tissue unequivocally show that the specific reduction in *NCoR1* mRNA observed in that tissue after HFD is matched with a spectacular drop in NCoR1 protein. If the reduced *NCoR1* mRNA levels in muscle also translate to corresponding changes in NCoR1 protein, they will prime the muscle for mitochondrial oxidation. These selective effects of NCoR1 to repress muscle fatty acid oxidation hence suggest that changes in NCoR1 levels adapt transcriptional outcomes to physiological energy needs.

In conclusion, we demonstrated here that NCoR is an evolutionary conserved negative regulator of both muscle mass and mitochondrial oxidative metabolism in nematode and mammals. In the mouse, NCoR1 achieves these effects through controlling a rather selected set of functional pathways, which are governed by MEF2, PPARb/d, and the ERRs, and can be seen in diverse genetic populations. The *NCoR1*$^{skm-/-}$ muscle phenotype furthermore mirrors many features

of the stimulation of PGC-1a, a coactivator, whose action is less constrained by the absence of the NCoR1 corepressor scaffold. Our work also provides evidence that NCoR1 expression is regulated in a dynamic fashion and, as such, could play a role similar to PGC-1a in transcriptional adaptation to physiological challenges. Moreover, pharmacological inhibition of NCoR1 and/or its interaction with deacetylases may be a viable approach to improve muscle mass and oxidative metabolism. The fact that the inhibition of HDACs increases muscle cell size supports this concept [231]. It is also tempting to speculate that the beneficial effects of the inhibition of mTORC1 and insulin signaling on health and life span may, in part, rely on the attenuation of NCoR1 activity and the subsequent induction of oxidative metabolism in the muscle (reviewed in [202]).

3.1.5 Methods

NCoR1 floxed (*NCoR1*$^{L2/L2}$), *NCoR1*$^{skm+/+}$, and $^{skm-/-}$ mice were generated at the Mouse Clinical Institute (Strasbourg, France) and phenotyped according to standard procedures [123, 157]. *C. elegans* O_2 consumption was measured using 200 2-day-old worms using a Seahorse XF24. Staining of muscles with hematoxylin and eosin, immunohistochemical and EM analysis, and analysis of enzymatic activity of SDH and COX were carried out as described [232]. mRNA expression levels were measured in cells and tissues using qRT-PCR [232]. The GeneNetwork program (**www.genenetwork.org**) was used to generate a broad range of NCoR1-correlated genes that may contribute to the phenotype of *NCoR1*$^{skm-/-}$ mice. Skeletal muscle mRNA expression was analyzed using 124 females from a classic F2 intercross between C57BL/6J and C3H/HeJ (UCLA BHHBF2 Muscle; GEO GSE12795; [203]). *NCoR1* (10024414685, 3' UTR) was compared across all transcripts to find muscle covariates. Lung mRNA in a recombinant inbred intercross between C57BL/6J and DBA/2J was analyzed across 51 strains (HZI BXD Lung M430v2 [Apr08] RMA) [204]. Four *NCoR1* probe sets from this microarray were analyzed (1423200_at, 3' UTR; 1435914_at, 3' UTR; 1423202_a_at, exonic and 3' UTR; 1423201_at, exonic). For all probe sets, the top correlates were calculated. Strong or interesting correlates were selected for validation by qRT-PCR. Statistical analyses were performed with a Student's *t*-test for independent samples. Data are expressed as mean ± SEM, and *p* values smaller than 0.05 were considered as statistically significant. **p* < 0.05; ***p* < 0.01; ****p* < 0.001.

3.1.6 Acknowledgments

We acknowledge M. Lazar (University of Pennsylvania, Philadelphia), A. Kralli (Scripps Research Institute, San Diego), J.-S. Annicotte and L. Fajas (Institut de Génétique Moléculaire de Montpellier, France), A.J. Lusis (University of California, Los Angeles), K. Schughart (Helmholtz Zentrum, Hannover, Germany), R.W. Williams (University of Tennessee Health Science Center, Memphis), and the Caenorhabditis Genetics Center (CGC) for generous sharing of research reagents and data. We thank N. Messadeq (Institut Clinique de la Souris, Strasbourg, France) for EM analysis and the Center for PhenoGenomics (CPG) at the EPFL for help with mouse phenotyping. This work was supported by the École Polytechnique Fédérale de Lausanne, Swiss National Science Foundation, NIH (DK059820 to J.A., DK062434 to R.M.E., 1K08HL092298 to G.D.B., HD027183 to R.M.E., and DK057978 to R.M.E.), the EU ideas program (ERC-2008-AdG-23118), the Helmsley Charitable Trust, the Glenn Foundation, and the Howard Hughes Medical Institute (HHMI). H.Y. was supported by an FRM fellowship. R.M.E. is an Investigator of the HHMI and the March of Dimes Chair in Molecular and Developmental Biology. J.A. is the Nestle Chair in Energy Metabolism. We thank the members of the Auwerx lab and R. Williams for discussions.

3.2 Drug Treatments to Improve Muscle Function

3.2.1 PARP Inhibition Enhances Energy Expenditure and Improves Muscle Function

Reductions in mitochondrial number and activity are a hallmark of several inherited mitochondrial diseases, as well as a number of age-related neurodegenerative and metabolic disorders [233]. The sirtuin enzymes have emerged as critical regulators of mitochondrial and oxidative metabolism [234]. Several studies have shown that NAD^+ levels can be rate limiting for the deacetylase activity of SIRT1, the best-characterized sirtuin [235]. Boosting intracellular NAD^+ levels has hence become an attractive approach to activate SIRT1 and mitochondrial metabolism. In line with this, NAD^+ avail-

ability might even determine circadian shifts in oxidative capacity [236]. One strategy to increase NAD^+ levels consists of the inhibition of alternative NAD^+-consuming enzymes, such as poly(ADP-ribose) polymerase-1 (PARP-1) [131], which is involved in DNA damage detection and repair. Upon activation, PARP-1 can deplete intracellular NAD^+ levels by 80%. Genetic ablation of *Parp-1* increases NAD^+ availability and SIRT1 activity in tissues, such as skeletal muscle and brown adipose tissue [149]. As a consequence, mitochondrial oxidative capacity is enhanced in muscle of *Parp-1*$^{-/-}$ mice, protecting them against high-fat diet (HFD)-induced insulin resistance. Therefore, PARP inhibition constitutes a plausible strategy to improve metabolic homeostasis. PARP inhibitors (Paribs) already exist and are clinically tested in the cancer field [150]. The role of PARPs in genome maintenance, however questions whether Paribs can be used therapeutically over longer periods of time to prevent the detrimental metabolic consequences of mitochondrial dysfunction. Here, we demonstrate that long-term Parib treatment is sustainable and improves mitochondrial function in skeletal muscle, enhancing endurance performance and protecting against HFD-induced metabolic complications. We furthermore show that *Parp-1* expression is negatively correlated with energy expenditure in mouse genetic reference populations. Finally, we demonstrate how acquired and genetic mitochondrial defects can be improved by Paribs. Therefore, our work sets the stage for the possible clinical use of Paribs in situations of defective mitochondrial function.

As defects in mitochondrial metabolism are a hallmark for many diseases, Paribs could in principle also be used for these nononcological indications. However, such use could be overshadowed by several concerns. First, chronic Parib treatment could induce genomic instability [150]. Second, Paribs would affect not just PARP-1 but also PARP-2, and the combined reduction of both activities could be detrimental for long-term viability [237]. Hence, we characterized the impact of long-term MRL-45696 treatment in mice. We initially determined that dietary admixture achieved higher MRL-45696 levels in plasma and muscle than oral gavage, hence choosing this route for further studies. We then fed HFD admixed with MRL-45696 (50 mg/kg/day) to 10-week-old male C57BL/6J mice. MRL-45696 blunted HFD-induced body weight gain (Fig. 3:7A) due to reduced fat accumulation (Fig. 3:7B) and was associated with higher energy expenditure (Fig. 3:7C), without affecting activity or food intake (Fig. 3:7D and 3:7E). Although PARP-1 has been shown to impact genome stability and cell viability, no evidence for toxicity on genomic DNA or cellular damage was found, as liver 8-oxo-dG and muscle lipid peroxidation levels were similar between the groups (Fig. 3:7F and 3:7G). This is in line with the fact that *Parp-1*$^{-/-}$ mice are viable and do not show signs of DNA damage unless challenged with cytotoxic stimuli [238].

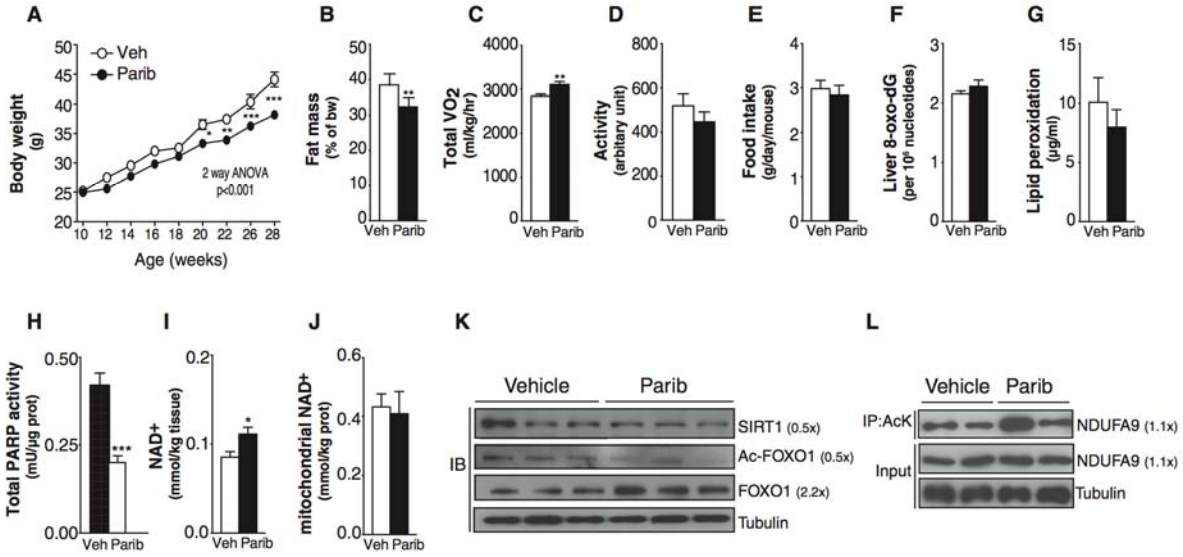


Figure 3:7 **Paribs Protect from HFD-Induced Metabolic Complications**

Ten-week-old male C57BL/6J mice were challenged with HFD supplemented with either vehicle (DMSO; Veh) or MRL-45696 (50 mg/kg/day) (*n* = 10/group). **(A)** Body weight gain during 18 weeks of HFD. **(B)** Fat mass was measured using Echo-MRI. (C and D) A comprehensive laboratory animal monitoring system was used to evaluate VO2 **(C)** and activity **(D)** after 7 weeks of HFD. **(E)** Food intake measured by averaging weekly food consumption during HFD. **(F)** Liver 8-oxo-dG, an indicator of DNA damage, **(G)** muscle lipid peroxidation-derived aldehyde, 4-hydroxy-2-nonenal, and **(H)** muscle total poly(ADP-ribose) (PAR) contents were measured in vehicle and MRL-45696-treated mice (*n* = 7/group). **(I and J)** Total intracellular **(I)** and mito-

chondrial **(J)** NAD^+ levels in gastrocnemius of refed vehicle and MRL-45696-treated mice (n = 5–10/group). **(K)** SIRT1, acetylated FOXO1, and total FOXO1 protein levels were assessed in total homogenates from quadriceps of CD-fed mice. **(L)** The acetylation status of NDUFA9 immunoprecipates was tested as a marker of SIRT3 activity. Values are shown as mean ± SEM. Asterisk indicates statistical significant difference versus respective Veh group. *p < 0.05; **p < 0.01; ***p < 0.001.

MRL-45696 reduced total PARP activity in skeletal muscle (~50%; Fig. 3:7H) and in other tissues tested, including brain and liver (not shown). In line with PARP inhibition, intracellular NAD^+ content was higher in muscle from treated mice (Fig. 3:7I). The higher NAD^+ levels were restricted to the nuclear and/or cytosolic compartments, since mitochondrial NAD^+ content was similar between both groups (Fig. 3:7J). Reductions in PARP activity translated also into SIRT1 activation, as reflected by lower FOXO1 acetylation, despite slightly lowered SIRT1 levels (Fig. 3:7K). Conversely, the unaltered mitochondrial NAD^+ levels did not change SIRT3 activity, as manifested in the acetylation of NDUFA9 (Fig. 3:7L). These observations are in line with observations in *Parp-1*$^{-/-}$ mice, where SIRT1, but not SIRT3, activity was enhanced [149].

3.2.2 PARP Inhibition Enhances Endurance and Mitochondrial Function

Given that MRL-45696 accumulates and inhibits PARP in the skeletal muscle (Fig. 3:7H), we next examined the impact of Paribs on muscle performance and mitochondrial function in CD-fed mice. Despite similar body weight and composition (not shown), MRL-45696-treated mice ran for a significantly longer distance (Fig. 3:8A). Maximal O_2 consumption capacity (VO_{2max}), which critically determines endurance performance, was significantly higher with MRL-45696 (not shown). As cardiovascular parameters, including ejection fraction, fractional shortening, ventricular mass, and diameter (not shown), were unchanged, the higher endurance was most likely due to improved muscle mitochondrial function.

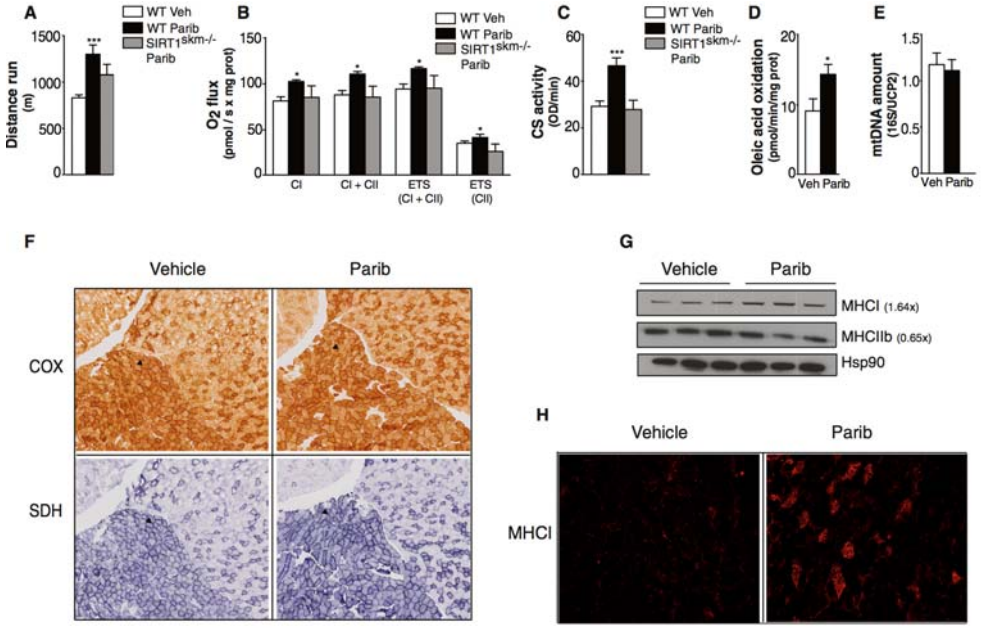


Figure 3:8 **Paribs Enhance Exercise Capacity and Muscle Mitochondrial Function**

Chow-fed male C57BL/6J and congenic SIRT1$^{skm-/-}$ mice (n = 5–10/group) treated with vehicle (DMSO; Veh) and/or MRL-45696 (50 mg/kg/day) were subjected to **(A)** endurance treadmill test after 13 weeks of treatments, **(B)** respirometry analysis of permeabilized EDL muscle fibers (CI/II, complex I/II; ETS, electron transport system), and **(C)** CS activity measurement. **(D)** Oleic acid oxidation rate in the muscles of Veh and MRL-45696-treated mice after 18 weeks of treatment (n = 5–7/group). **(E)** Mitochondria DNA abundance in quadriceps of Veh and MRL-45696-treated mice (n = 8/group). Results are expressed as mitochondrial DNA amount (16S) relative to genomic DNA (UCP2). **(F)** Cytochrome c oxidase (COX) and succinate dehydrogenase (SDH) stainings in gastrocnemius of Veh and MRL-45696-treated mice. Soleus is indicated by an arrow. **(G)** Protein levels of MHCI and IIb were evaluated using heat-shock protein 90 (Hsp90) as a loading control.

(H) Myosin heavy chain I staining (MHCI) of gastrocnemius of Veh and MRL-45696-treated mice. Values are shown as mean ± SEM. Asterisk indicates statistical significant difference versus respective Veh group. *p < 0.05; **p < 0.01; ***p < 0.001.

To substantiate this hypothesis, we performed respirometry tests in permeabilized EDL muscle fibers. Maximal respiration in the coupled state, with electron input through either complex I alone—using malate, pyruvate, and glutamate—or complex I+II—using, in addition, succinate—was markedly higher in muscles after MRL-45696 (Fig. 3:8B). Maximum electron transport system capacity was also higher after PARP inhibition, even after the addition of the complex I inhibitor rotenone (Fig. 3:8B). These results are consistent with the marked increase in citrate synthase (CS) activity after MRL-45696 (Fig. 3:8C). Finally, Paribs increased oleic acid oxidation rates by ~50% (Fig. 3:8D). Mitochondrial DNA abundance was, however, not changed (Fig. 3:8E), suggesting that MRL-45696 does not influence muscle mitochondrial function by increasing mitochondrial number.

Histological analysis further certified that Paribs enhanced the oxidative profile of skeletal muscle, as testified by more intense cytochrome *c* oxidase and succinate dehydrogenase stainings in gastrocnemius (Fig. 3:8F). Remarkably, an increase in slow oxidative MHCI, a marker for type 1 fibers, was observed, whereas MHCIIb levels, a type 2 fiber marker, were decreased by MRL-45696 (Fig. 3:8G). Furthermore, MHCI staining of gastrocnemius confirmed the higher type 1 fiber content after MRL-45696 (Fig. 3:8H). The higher oxidative capacity after Paribs was accompanied by increased insulin sensitivity, as reflected by the higher glucose infusion rate during a hyperinsulinemic-euglycemic clamp (not shown). We next evaluated the possible contribution of SIRT1 in mediating the effects of Paribs by using a muscle-specific SIRT1-deficient mouse line (*SIRT1*$^{skm-/-}$ mice). In line with previous reports [239, 240], endurance or VO_{2max} was not impaired in *SIRT1*$^{skm-/-}$ mice (data not shown). Interestingly, Paribs were unable to enhance endurance, respiratory capacity, and CS activity in the *SIRT1*$^{skm-/-}$mice (Fig. 3:8A–C). Of note, the expression of *Sirt1*, *Tfam*, and *c-myc* was also increased in Parib-treated muscles (not shown), which indicates that a recently identified SIRT1-dependent pathway [241] by which NAD^+ controls metabolic health might also be at play. Therefore, SIRT1 seems key to the actions of Paribs on muscle metabolism.

3.2.3 Parp-1 Expression Negatively Correlates with Energy Expenditure

Given the effects of Paribs on mitochondrial function in the C57BL/6J mouse strain, we next aimed to elucidate whether *Parp-1* could determine energy expenditure and oxidative capacity in genetically complex mouse populations such as the BXD mouse GRP [110]. *Parp-1* expression varied by ~1.3-fold in skeletal muscle across the BXD strains, enabling the study of gene expression correlations (Fig. 3:9A). These changes in *Parp-1* expression are not related to natural variations in fiber type composition, as *Parp-1* expression is comparable in muscles having different fiber type content, such as gastrocnemius or soleus (data not shown). *Parp-1* expression in muscles of the BXDs strains was negatively correlated with night VO_2 (i.e. during the mouse active phase), VO_{2max} improvement after 10 days' training on an activity wheel, and with expression of the troponin I type 1 slow-twitch muscle isoform (*Tnni1*) (Fig. 3:9B). In contrast, *Parp-1* positively correlated with body weight, troponin C type 2 fast-twitch muscle isoform (*Tnnc2*) and troponin I type 2 fast-twitch muscle isoform (*Tnni2*) (Fig. 3:9B).

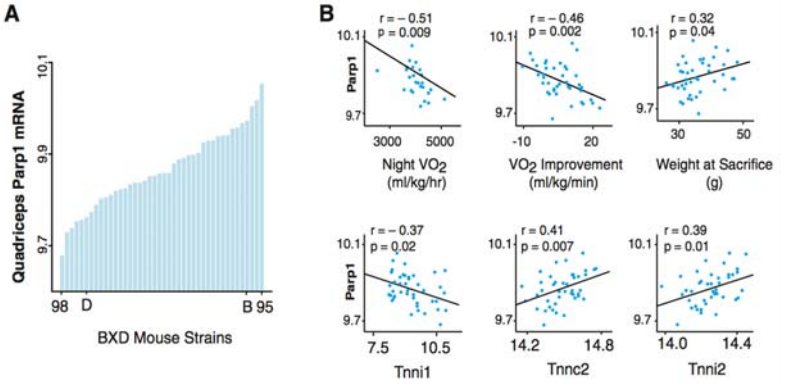


Figure 3:9 ***Parp* Negatively Correlates with Mitochondrial Function**

(A) Expression of *Parp-1* in the quadriceps of 37 BXD strains. Bar represent pools of ~5 animals. Extremes and parentals are labeled. **(B)** (Top) Correlation between muscle *Parp-1* expression and BXD phenotypes. VO_2 was measured though indirect calorimetry. VO_2 improvement represents the change in VO_{2max} after 10 days of voluntary exercise. (Bottom) Correlation between quadriceps *Parp-1* and fiber type genes in the same data set.

This robust genetic correlation suggests a physiological influence of PARP-1 activity on oxidative metabolism in more diverse populations. Indeed,

previous results indicated that PARP activity is increased upon HFD [149], likely favoring fat storage over burning [242]. Consistent with this, PARP-1 activity was found to enhance adipogenesis [242]. Oppositely, PARP activity was decreased in muscle upon fasting [149], when lipid oxidation is favored. Collectively, this indicates that physiological changes in PARP-1 activity might help balancing whole-body energy metabolism.

3.2.4 Conclusions and Perspectives on PARP Inhibitors

In sum, this work provides evidence that drugs used at present in the cancer field could be retooled to treat metabolic dysfunction linked to impaired mitochondrial activity, even in a long-term fashion. Paribs are currently being tested for oncology indications, with positive outcomes in tumors with dysfunctional homologous DNA recombination repair [150]. The success of Paribs is thought to rely on the high dependency of these cancer cells on PARP enzymes for DNA repair [150], but could also involve their capacity to reprogram cells toward oxidative metabolism [149]. Notably, Paribs are in general well tolerated by patients, with few reported side effects [135, 243, 244]. Within the time frame of this study, and in line with findings in *Parp-1*$^{-/-}$ mice [149] and in patients treated with Olaparib [135, 243, 244], no toxicity or genomic instability was seen in MRL-45696-treated mice. In contrast, marked metabolic effects were observed, such as improved mitochondrial function and protection against diet-induced obesity. Importantly, PARP inhibition rescued mitochondrial activity in situations of genetically determined mitochondrial dysfunction. Therefore, our results open the path of using Paribs to improve invalidating diseases caused by impaired mitochondrial function. While this could also set the stage for Paribs to impinge on other metabolic complex diseases, such as seen during type 2 diabetes, further work must be done to ensure the safety and feasibility of these treatments in non-life-threatening diseases.

3.2.5 Methods

To compare administration routes, male C57BL/6J mice were given MRL-45696 by gavage or food admix at a dose of 50 mg/kg/day (mkd) for 5 days. Mice were sacrificed either 6 hr after the last gavage or in the random fed state, respectively. MRL-45696 concentration in plasma and muscle was measured by mass spectrometry. For the long-term animal studies, 10-week-old male C57BL/6J mice (Charles River) or *SIRT1*$^{skm-/-}$mice were fed pellets containing vehicle and/or PARP inhibitor (50 mkd) for 18 weeks. During the experiment, mice were housed under a 12 hr dark-light cycle and had ad libitum access to water and food. Clinical tests were carried out according to standard operational procedures. Mitochondrial function in permeabilized EDL muscle fibers was evaluated using the Oxygraph-2k respirometer (Oroboros, Austria). Correlations in the BXD mouse genetic reference population were identified using the GeneNetwork database (**www.genenetwork.org**). All mouse experiments were performed according to Swiss regulations and approved by the ethical commission of the Canton de Vaud, Switzerland, under licenses 2465 and 2465.1. Statistical analyses were performed with Prism software (GraphPad). Differences between two groups were analyzed using Student's *t*-test (two-tailed), and multiple comparisons were analyzed by ANOVA with a Bonferroni post hoc test. A *p*-value less than 0.05 was considered significant. Data are expressed as means ± SEM.

3.2.6 Acknowledgments

We thank the members of the Auwerx lab and M. Hottiger for discussions, and P. Cettour-Rose, B. Rochat, D. Dahlmans, N. Ziegler, and S.J. Sturla for technical help. We also thank the Molecular Resource Center of Excellence at The University of Tennessee Health Science Center for running the BXD microarrays. C.C. is an employee of the Nestlé Institute of Health Sciences. J.A. is the Nestlé Chair in Energy Metabolism and a founder and scientific advisor to Mitokyne. A.A.S. has intellectual property in the field of NAD^+ enhancers. E.P. was funded by the Academy of Finland, the Saastamoinen Foundation, the Finnish Cultural Foundation, and the Finnish Diabetes Foundation. A.A.S. was supported by the NIH (R01 GM106072-01). The work in the Auwerx lab was supported by the École Polytechnique Fédérale de Lausanne, the EU Ideas program (AdG-23138 and AdG-322424), the NIH (R01HL106511-01A and R01AG043930), and the Swiss National Science Foundation (31003A-124713 and 31003A-125487 and CSRII3-1362).

3.3 AP-1-Related Regulation of Steatohepatitis and PPARg Signaling

3.3.1 Introduction to AP-1, PPARγ Signaling, and NAFLD

Given their high energy-to-weight ratio compared to carbohydrates and proteins, lipids are the most efficient energy substrate in mammals. The adipose tissue is the major lipid storage organ and it is essential for controlling metabolic homeostasis [245]. In the healthy state, tissues such as muscle and liver store only minor quantities of lipids [246]. However, metabolic stress, as occurring in obese or alcohol-abusing patients, can cause massive ectopic lipid deposition, leading to a disease state, termed "steatosis" or "fatty liver disease". Depending on the etiology, this disease can be further subgrouped into alcoholic or non-alcoholic fatty liver disease (AFLD and NAFLD respectively). NAFLD is the most common liver disorder in industrialized countries and it frequently leads to severe liver inflammation and damage, a disease state termed "non-alcoholic steatohepatitis" (NASH) [247]. Moreover, NAFLD contributes to hepatic insulin resistance in diabetes [248] and is a risk factor for liver dysfunction and cancer development [249]. Understanding the cellular and molecular mechanisms leading to NAFLD, as well as the identification of novel targets for NAFLD therapy has therefore become a priority [250, 251].

The Activator Protein-1 (AP-1) (Fos/Jun) protein complex is a dimeric leucine zipper (bZIP) transcription factor. Three different Jun proteins (c-Jun, JunB and JunD) and four different Fos proteins (c-Fos, FosB, Fra-1 and Fra-2) can form AP-1 dimers. Jun proteins can either form homodimers, such as c-Jun/c-Jun or c-Jun/JunB, or heterodimers, such as c-Jun/c-Fos. In contrast, Fos proteins exclusively form heterodimers [252]. Jun and Fos proteins also form heterodimers with other bZIP transcription factors, such as specific MAF and ATF family members [253]. Thus, a vast combinatorial variety of AP-1 dimers with likely different molecular and biological functions exists [254-256]. Studies using genetically modified mice have unravelled essential roles of AP-1-forming proteins in development, inflammation and cancer [253]. Moreover, AP-1 modulates the response to acute cellular stresses, such as reactive oxygen species or DNA damage [257]. Cellular stresses typically activate AP-1 by augmenting transcription, protein stability and transactivation potential of Jun and Fos family members [258]. In the liver, the genetic inactivation of single *Jun* or *Fos* genes in hepatocytes does not compromise organ homeostasis [253, 259]. However, AP-1 is critical for the liver's response to acute stress. For example, c-Jun protects hepatocytes from injury [260, 261] and is essential for liver regeneration [262] and carcinogenesis [263-265]. More recently, we have documented that Fra-1, but not Fra-2, protects hepatocytes from acetaminophen overdose, a paradigm for xenobiotic-mediated acute liver failure [266]. In contrast, little is known about the role of AP-1 in chronic stress conditions and the potential contribution of AP-1 to the development of hepatic metabolic disease. Here we combined system genetics with gain- and loss-of-function (G/LOF) mouse models to study the function of AP-1 in hepatic lipid metabolism and NAFLD development. We show that, depending on dimer composition, AP-1 can either repress or activate the transcription of the pro-steatotic nuclear receptor Peroxisome Proliferator-Activated Receptor γ (PPARγ), which promotes hepatic lipid uptake and lipid droplet formation. Some AP-1 proteins, such as Fra-1 and Fra-2, inhibit the PPARγ pathway and reduce hepatic lipid content, while other AP-1 proteins, such as c-Fos and JunD induce hepatic PPARγ signaling and lipid accumulation. We also show that AP-1 regulates the PPARγ pathway through direct regulation of the *Pparg2* promoter. Using a mouse model for inducible hepatocyte-restricted Fra-1 expression, we demonstrate that the Fra-1-induced suppression of the PPARγ pathway can revert established NAFLD. For the first time, liver-specific single chain Jun~Fos forced dimer mice were employed, in which dimerization of a Fos protein is restricted to a single Jun partner [267]. The analyses of these mouse models provide *in vivo* evidence that distinct AP-1 dimers regulate the PPARγ pathway in an antagonistic fashion. Finally, we show that JunD is essential for efficient PPARγ signaling and NAFLD formation. Overall, this study identifies AP-1 as a link between dietary obesity, hepatic lipid metabolism and NAFLD.

3.3.2 AP-1-Mediated Regulation of Hepatic Lipid Metabolism and NAFLD

To identify a possible function of AP-1 in metabolism, we analyzed 42 genetically diverse mouse strains from the BXD mouse genetic reference population (GRP) [268]. 10 animals for each strain were split evenly into two cohorts fed either chow diet (CD) or high-fat diet (HFD) for five months. Hepatic AP-1 mRNA expression was then analyzed using genome-wide expression profiles from the BXD strains. *Fra-1* mRNA levels were found to be significantly reduced in the

HFD-fed cohort, while the expression of *c-fos, fosB, fra-2, c-jun, junb* and *jund* were not affected by the diet (Fig. 3:10A). To explore whether *Fra-1* could causally contribute to HFD-associated metabolic changes in the liver, we analyzed hepatic metabolism in Fra-1^{hep} mice, a previously established model of Doxycycline (Dox)-controllable hepatocyte-restricted Fra-1-overexpression, which does not display any obvious phenotype under basal conditions [266] (for details on mouse strains see Table S1). After HFD feeding, the livers appeared less pale on the macroscopic level and weighed significantly less in Fra-1^{hep} mice compared to HFD-fed littermate controls (Fig. 3:10B, C). Liver histology indicated a reduction in lipid droplets in mutant mice (Fig. 3:10B), which was confirmed by the quantitation of Oil-RedO (ORO)-positive lipid droplets and liver triglyceride (TG) content analysis (Fig. 3:10C).

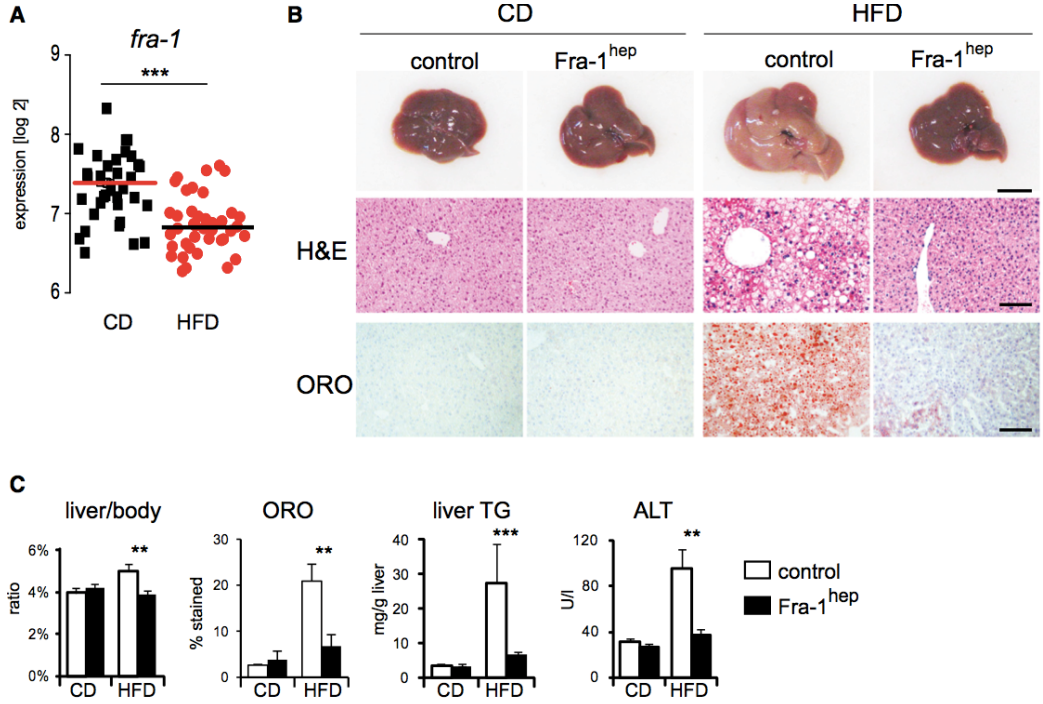


Figure 3:10 **Fra-1 is Regulated by HFD and Inhibits NAFLD and PPARγ Expression**

(A) Hepatic *fra-1* expression in CD and HFD (for 5 months; 60% kCal/fat) in 42 BXD inbred strains. Each data point represents the mean expression of 5 mice. (**B,C)**, Fra-1^{hep} mice and control littermates were on CD or HFD (for 5-9 months; 45% kCal/fat); *n*≥5/cohort. **(B)** Representative liver pictures and histology in Fra-1^{hep} and control mice; ORO=Oil-RedO; bars=1cm and 100µm. **(C)** Quantitation of ORO-positive areas, liver/body ratio, liver TG content and serum ALT levels. Bar graphs are presented as mean±SEM. ** $p < 0.01$, *** $p < 0.001$.

We next addressed the effect of Fra-1 expression on NAFLD-associated liver damage and inflammation. Augmented serum levels of the liver damage marker alanine aminotransferase (ALT) and increased hepatic inflammation marker expression were observed in controls after HFD-feeding, but not in Fra-1^{hep} mice (Fig. 3:10C). Moreover, immunohistochemistry (IHC) for the pan-lymphocyte marker CD45 and the macrophage marker F4/80 revealed a significant reduction in immune cell infiltrates in HFD-fed mutants compared to diet-matched controls. In the HFD-fed state, serum IL-6 levels were also reduced in HFD-fed Fra-1^{hep} mice compared to controls (data not shown). We next explored the effects of hepatic Fra-1 expression on circulating metabolite and hormone levels. Serum TG and cholesterol were mildly elevated in HFD-fed Fra-1^{hep} compared to control mice in the fasted and/or fed states, while other serum parameters were not affected. Despite decreased NAFLD and liver damage, glucose tolerance and insulin tolerance tests (GTT and ITT) revealed that glucose metabolism was not improved, but rather slightly worsened in HFD-fed mutants as compared to controls. Similar effects of Fra-1 on NAFLD development were also observed in Fra-1^{hep} mice on a C57BL/6J background or using 60% kCal/fat HFD (data not shown). These data collectively suggest that hepatocyte-specific Fra-1 expression protects from dietary-induced NAFLD and secondary liver damage and inflammation, but has little impact on systemic obesity and glucose metabolism.

3.3.3 Mechanism of Action: Fra-1 Regulates the PPARg Pathway

We next analyzed the molecular mechanisms underlying reduced NAFLD in Fra-1^{hep} mice. Genome-wide hepatic gene expression analyses in CD- and HFD-fed Fra-1^{hep} mice demonstrated that the expression of ~3000 genes was changed by at least 1.5 fold in Fra-1^{hep} livers. The vast majority of these genes were regulated in a similar fashion in both dietary conditions, and many PPARγ targets such as *adipsin* and *cidea* were among the most downregulated genes (Fig. 3:11A). KEGG pathway analysis [269] of the top 2000 most changed genes established PPARγ signaling among the most significantly affected pathways in both dietary cohorts (Fig. 3:11B). A highly significant fraction of mRNAs, which were reduced in Fra-1^{hep} mice, were encoded by genes with promoters containing putative AP-1 sites (p=9.0E-13) and PPARγ response elements (PPREs) (p=3.3E-11), as revealed by the UCSC TFBS conserved tracks analyses on DAVID (**http://david.abcc.ncifcrf.gov**) [270] (data not shown). qRT-PCR, immunoblot and IHC analyses confirmed decreased hepatic *pparγ* mRNA/PPARγ protein expression in CD- and HFD-fed Fra-1^{hep} mice (Fig. 3:11C and 3:11D).

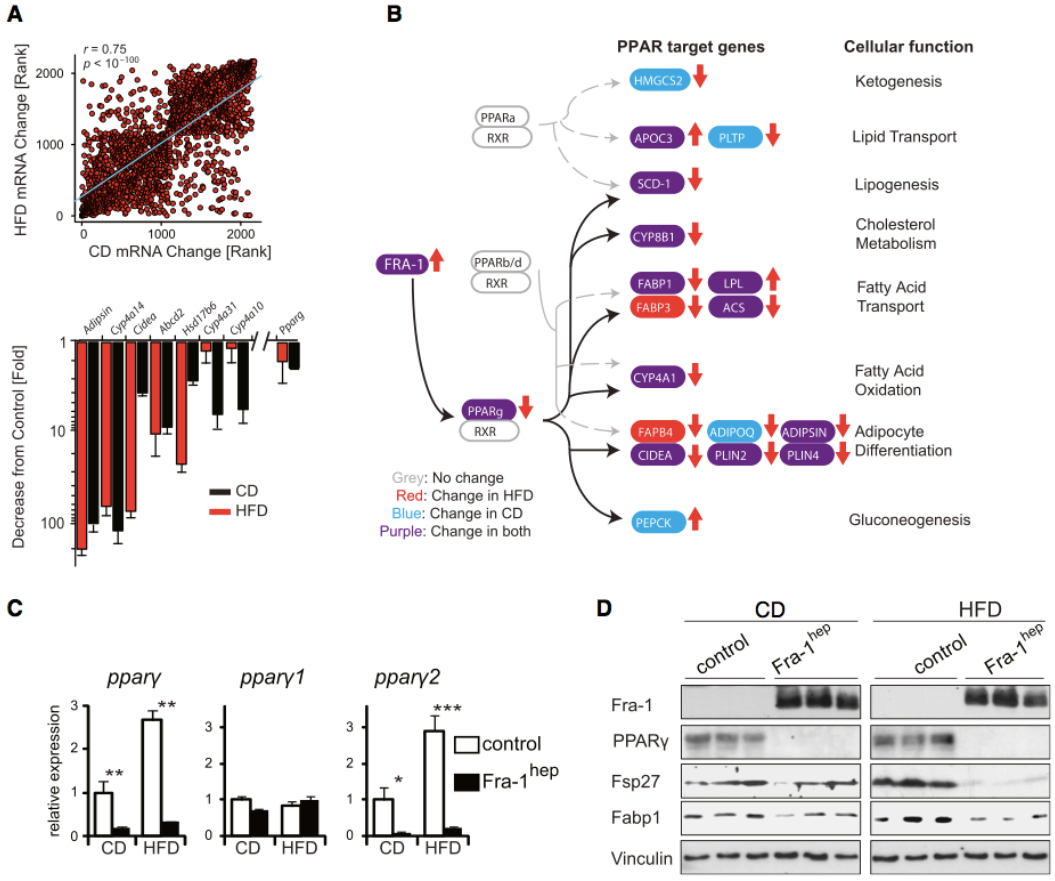


Figure 3:11 **Fra-1 Regulates the PPARγ Pathway**

(A–D) Fra-1^{hep} mice and control littermates on CD or HFD (for 5-9 months; 45 % kCal/fat); n≥5/condition. **(A)** Top: Spearman correlation of ≥ 1.5 fold-changed hepatic transcripts in Fra-1^{hep} and control littermates (C57BL/6J) in CD (x-axis) and HFD (y-axis). Bottom: Top common downregulated genes in Fra-1^{hep} livers. **(B)** KEGG pathway analyses for the top 2000 most changed transcripts: PPAR target genes and their cellular functions are indicated. Transcripts changed in CD, HFD or both are highlighted in blue, red and purple respectively. Arrows indicate up- or down-regulation. **(C)** qRT-PCR analyses of *pparγ* and its isoforms. **(D)** Immunoblot analysis; vinculin is the loading control. Bar graphs are presented as mean±SEM.

Decreased *pparγ* mRNA levels were due to reduced expression of *pparγ2* mRNA, the main *pparγ* isoform in the liver (Fig. 3:11C) [271]. Among other metabolic regulators *Nr0b2*, a potential PPARγ target [272] and regulator [273], which encodes the orphan nuclear receptor SHP, was also found reduced in HFD-fed Fra-1^{hep} mice. Moreover, we confirmed reduced mRNA expression for several PPARγ target genes, such as *fabp1* and *lpl,* involved in hepatic lipid uptake, and

plin2, cidea, fitm1, fitm2, g0s2, involved in lipid droplet formation. Notably, the Fra-1-induced reduction of *pparγ2* expression was reversible, as *pparγ2* levels reverted to baseline levels upon switching off the transgene. Kinetic analyses of another inducible *Fra-1* mouse model (Fra-1^{tetON} mice)[266] revealed that hepatic *pparγ2* mRNA decreased as early as 4 days after *Fra-1* induction.

G/LOF studies previously established that expression of PPARγ in the hepatocytes is both essential and sufficient for NAFLD formation [271, 274-278]. To determine whether reduced NAFLD development in Fra-1^{hep} mice is directly due to decreased PPARγ levels, HFD-fed Fra-1^{hep} and control mice were intravenously injected with either Adeno-PPARγ or Adeno-GFP control virus 8-10 days prior to sacrifice. Adeno-PPARγ had no overt effect on liver macroscopy in steatotic control mice (Fig. 3:12A). In contrast, PPARγ expression increased ORO-positive lipid droplets, liver TG content, and liver/body weight ratio in HFD-fed Fra-1^{hep} mice (Fig. 3:12A and 3:12B). Moreover, Adeno-PPARγ increased PPARγ target gene expression in the livers of Fra-1^{hep} mice (Fig. 3:12C and 3:12D). These data demonstrate that the short-term induction of PPARγ signaling restores hepatic fat accumulation in HFD-fed Fra-1^{hep} mice, supporting its central function in the hepatic phenotype of Fra-1^{hep} mutant mice.

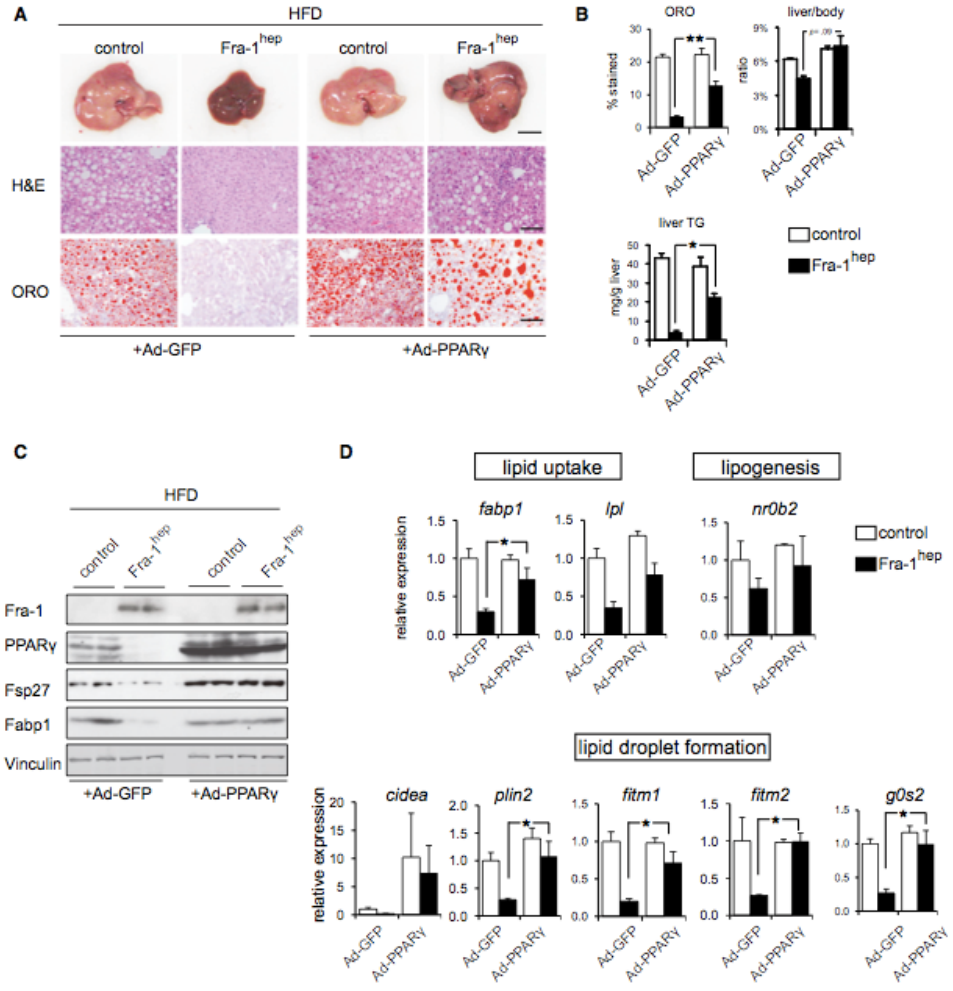


Figure 3:12 **PPARγ Delivery Restores NAFLD Development in Fra-1^{hep} Mice**

(A–D) Fra-1^{hep} and control littermates on HFD (for 4–5 months, 45 % kCal/fat) were injected with Adenoviruses expressing PPARγ (Ad-PPARγ) or GFP (Ad-GFP) 8-10 days prior to sacrifice. *n* = 3 for control genotype/cohort; *n* = 4 for Fra-1^{hep} mice/cohort. **(A)** Liver macroscopy and histology; bars=1cm and 100µm. **(B)** Quantitation of ORO-positive areas, liver/body ratio and liver TG. **(C)** Immunoblot analyses in Fra-1^{hep} and control mice. Vinculin served as loading control. **(D)** qRT-PCR analyses of PPARγ target genes involved in lipid uptake and lipid droplet formation; Ad-GFP controls are set to 1. Bar graphs are presented as mean±SEM.

To examine whether *Fra-1* induction in steatotic livers ameliorates disease symptoms, Fra-1^{hep} and control littermates were generated in the "Fra-1 off" state and HFD-feeding was started at 1 month of age (Fig. 3:13A). As expected, control and mutant mice were indistinguishable at 7 months of age in the absence of transgene expression (Fig. 3:13B-D). Immunoblot analyses confirmed comparable PPARγ, Fsp27 and Fabp1 levels between control and mutant mice in the "Fra-1 off" state (Fig. 3:4E). A cohort of Fra-1^{hep} and control mice were kept on HFD, but Fra-1 expression was switched on in mutant mice at 7 months of age (Fra-1 off-on). After 2 months of *Fra-1* induction, serum ALT was significantly lower in Fra-1^{hep} mice than in control littermates and continued to improve 3 months later, even though the mice were maintained on HFD (Fig. 3:13C). At this point the liver was collected for macroscopy, histology, ORO-quantitation, liver/body weight ratio and liver TG content analysis, revealing an almost complete reversion of NAFLD in Fra-1^{hep} mutants (Fig. 3:13B-D). Immunoblotting and qRT-PCR analyses confirmed transgene induction, as well as the repression of PPARγ, targets of PPARγ and inflammatory markers after switching on *Fra-1* expression (Fig. 3:13E and 3:13F). These data suggest that *Fra-1*-mediated inhibition of the PPARγ pathway can reverse established NAFLD and liver damage.

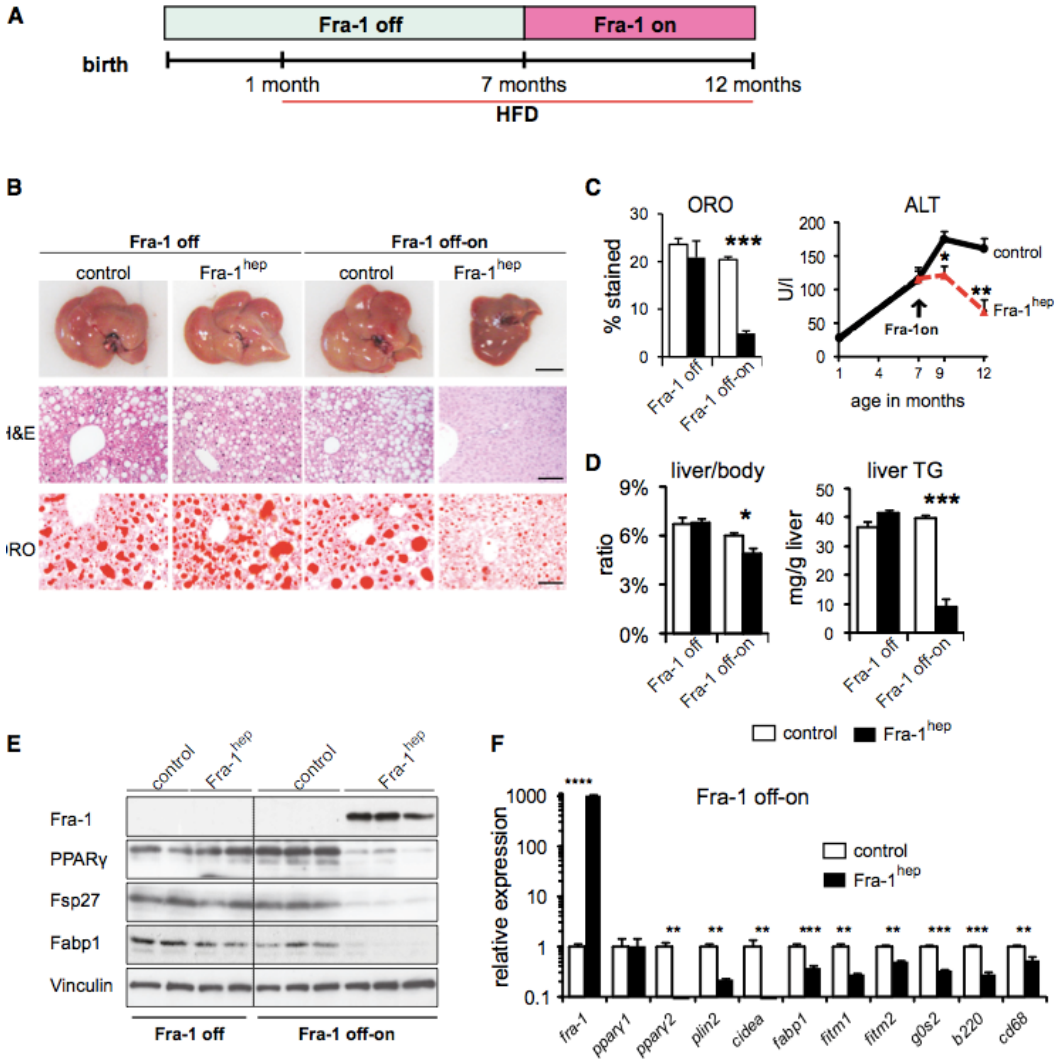


Figure 3:13 **Fra-1 Expression Reverts NAFLD and Liver Damage**

(A) Fra-1^{hep} and control littermates were maintained in the "Fra-1 off" state and HFD (45 % kCal/fat) was supplied from 1 month of age. Mice were analyzed at 7 months (Fra-1 off, *n* = 2/cohort) or kept on HFD until 12 months, while transgene expression was induced (Fra-1 off-on, *n* = 6/cohort). **(B)** Liver macroscopy and histology; bars=1cm and 100µm. **(C)** Serum ALT. **(D)** Quantitation of ORO-positive areas, liver/body ratio and liver TG content. **(E)** Immunoblot analyses in Fra-1^{hep} and control mice. Vinculin served as loading control. **(F)** qRT-PCR analyses of *fra-*

1, *pparγ* isoforms, PPARγ target genes and inflammation markers (Fra-1 off-on). Bar graphs are presented as mean±SEM.

We next searched for correlations between AP-1 genes, PPARγ and PPARγ targets in gene expression arrays from the diverse wildtype BXD family of inbred mice. This analysis revealed that hepatic *pparγ* expression, assessed with a probe detecting both *pparγ* isoforms, is significantly upregulated in HFD-fed cohorts. As expected, *pparγ* mRNA levels strongly correlated with the expression of PPARγ targets, such as *cidea* and *adipsin*. Notably, and consistent with our findings in Fra-1^{hep} mice, a significant inverse correlation was found between *fra-1* and *pparγ* and *fra-1* and *cidea* regardless of the diet (Fig. 3:14A). Other PPARγ targets, such as *adipsin* followed a similar trend, though without reaching statistical significance (Fig. 3:14A). Interestingly, *junb,* a potential dimerization partner for *Fra-1* also negatively correlated with *pparγ* (Fig. 3:14B), indicating that several AP-1 proteins consistently regulate PPARγ signaling even in genetically diverse populations.

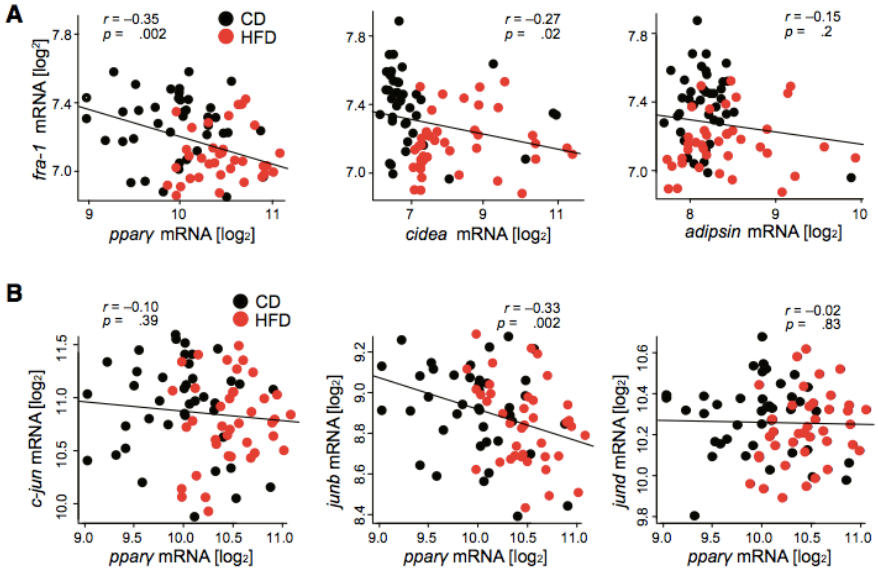


Figure 3:14 **Several AP-1 Proteins Regulate the PPARγ Pathway**

Correlation plots for *fra-1* with *pparγ, cidea* and *adipsin* **(A)** and for *Jun* members with *pparγ* **(B)** in CD and HFD (5months; 60 % kCal/fat) in the BXD inbred family. Each data point represents the average expression from 5 pooled mice. Pearson's *r* was used to analyze correlations and *p*-values are indicated.

3.3.4 Discussion

Combined forward and reverse genetic approaches have a strong potential for discovering new regulators of metabolism. The initial identification of *Fra-1* as a potential obesity-related gene in livers from the BXD population of inbred mouse strains prompted us to further explore the role of each component in the AP-1 dimer makeup. Subsequent mechanistic studies led to the discovery that AP-1 can function as a molecular link between obesity and liver metabolism. First, this study established AP-1 as a potent regulator of lipid metabolism and NAFLD development. Second, gene pathway analysis and BXD population genetics highlighted the AP-1 complex as a novel regulator of hepatic PPARγ signaling. Third, we demonstrate that Fra-1 repressed the PPARγ-dependent expression of genes involved in lipid uptake/lipid droplet formation, thereby efficiently improved established steatosis, liver damage and inflammation.

PPARγ promotes lipid uptake by increasing the expression of lipid transporters, such as fatty acid binding proteins (Fabps), and by promoting lipid storage in lipid droplets. Lipid droplet proteins (LDPs) inhibit TG lipolysis thereby preventing lipid-droplet breakdown [279-281]. Several LDPs are regulated by PPARγ at the transcriptional level [reviewed in 282] and promote NAFLD in mice, including via *cidea*, *plin2*, and *fsp27* [275, 280, 283-286]. Similarly, deletion of the fatty acid transporter *Fabp1* reduced the dietary induction of NAFLD [287, 288]. Previous studies suggested that hepatic PPARγ also promotes hepatic lipogenesis [274, 278]. In Fra-1^{hep} mice, which display a dramatic reduction in PPARγ

levels, decreased expression of the Stearoyl-CoA desaturase-1 (SCD-1), a key enzyme in the generation of unsaturated fatty acids, was observed. However, no consistent changes in the expression of SREBP-1/2, the main transcriptional regulators of *de novo* lipogenesis, nor of the SREBP-1/2 targets FAS and ACC, were observed. Since FAS and ACC catalyze the rate-limiting steps in fatty acid synthesis, altered lipogenesis likely does not play a major role in the Fra-1-mediated repression of NAFLD. Instead, decreased hepatic lipid uptake and lipid droplet formation is most likely the primary cause for reduced steatosis formation in Fra-1^{hep}, Fra-2^{hep} and c-Jun~Fra-2^{hep} mice. Previous reports have shown that PPARγ induces the expression of *Nr0b2* [272]. In line with this, HFD-fed Fra-1^{hep} mice displayed reduced *pparγ2* and *Nr0b2* levels, which appeared normalized after Adeno-PPARγ treatment. More recently, *Nr0b2* was shown to be required for hepatic PPARγ expression and NAFLD [273]. However, we did not observe a consistent correlation between *pparγ2* and *Nr0b2* expression across dietary conditions in our AP-1 mutant mouse models. Therefore, AP-1 likely regulates *pparγ2* and NAFLD independently of *Nr0b2*.

Hepatocyte-specific *Pparg* deletion, like Fra-1 overexpression, has been shown to reduce liver TG and to increase serum TG levels in the obese ob/ob mice and the AZIP lipodystrophy model, likely due to decreased hepatic lipid uptake [274, 276]. Moreover, PPARγ2-dependent hepatic steatosis has been suggested to buffer systemic TG levels [278]. Similar to mice with hepatocyte-restricted *Pparg* deletion, Fra-1^{hep} mice displayed worsened glucose metabolism after HFD-feeding, despite a reduction in NAFLD. Thus, hepatic Fra-1 overexpression largely phenocopies the effects of hepatocyte-specific *Pparg* deletion on lipid and glucose metabolism. As elevated serum TG levels are associated with diabetes development, increased serum TG levels likely contribute to the deterioration of glucose metabolism in HFD-fed Fra-1^{hep} and *Pparg*-deficient mice.

Our data suggest a functional antagonism between activating c-Fos/AP-1 and repressing Fra/AP-1 dimers. Previous studies in other organs suggested overlapping functions of c-Fos and Fra-1/2 [289, 290]. Thus, we here identify *Pparg* as the first gene to be antagonistically regulated by different Fos proteins via different dimeric composition of AP-1. This finding raises the intriguing question: How do structurally similar protein complexes, such as c-Fos/AP-1 and Fra/AP-1 dimers, have opposite effects on the same promoter? While further research is required to address this question, several AP-1 co-repressors, such as Sirt1 [291] and HDAC3 [212, 280, 292] are involved in hepatic lipid metabolism. We speculate that specifically Fra/AP-1 dimers might interact with such co-repressors to inhibit *Pparg2* promoter activity.

The described data from G/LOF mouse models, together with the correlations between AP-1 components and the PPARγ pathway in the BXD cohort, establish AP-1 as an important regulator of PPARγ signaling and NAFLD. HFD affects a plethora of cellular signaling cascades, such as the Insulin [293], the JNK [294] and the PKC [295] pathways. As these pathways are also known regulators of AP-1 expression and activity, exploring how they affect AP-1 levels and dimerization during obesity is certainly an important challenge for future experiments. Moreover, extensive cross-talk between AP-1 and transcription factors of the NF-κB [296] and the nuclear receptor family [297-299] has been described. Future studies should further reveal the molecular interplay of these pathways with AP-1 signaling in the context of NAFLD in both mice and humans.

3.3.5 Methods

Mice were maintained in a 12 hour light/12 hour dark cycle with food and water *ad libitum*. Chow (D8604, Harlan), 45% kCal/fat HFD (D12451, Research diets), and 60% kCal/fat HFD (D12492, Research diets) were used as specified in the figure legends. If not indicated otherwise, male mice were used and HFD feeding was started between 4-8 weeks of age. Dox (1g/l) was supplied in sucrose-containing (100g/l) drinking water. The BXD mice were sacrificed in the morning (9–11 am) after overnight fasting, while in other experiments, the mice were sacrificed in CO_2-chambers between 2pm and 5pm in the fed state. In the BXDs, cholesterol and TG measurements were performed using serum from overnight fasted mice at the time of sacrifice. For intra-peritoneal glucose tolerance (GTT) and insulin tolerance tests (ITT), mice were fasted for 6 hours (GTT) or 8 hours (ITT) and intra-peritoneally injected with 1 mg glucose/kg body weight (GTT) or 0.5U insulin/kg body weight (ITT). Glucose and insulin were diluted in PBS to an injectable volume. Blood glucose was determined by tail puncture for all time points. All mouse experiments were performed in accordance with local and institutional regulations. Blood was collected from the submandibular vein, by tail puncture or by cardiac puncture at experimental endpoints. Unless indicated otherwise specified, parameters were analyzed in the fed state. Serum ALT, TG and cholesterol levels were determined using a Reflovet blood chemistry analyzer and glucose using an Accucheck glucose analyzer (Aviva). Serum Leptin, Resistin, Adiponectin and IL-6 were measured using

Quantikine ELISA kits (R&D) and serum Insulin was determined with an ultra-sensitive ELISA (Mercodia). Serum β-HB and FFA were measured using enzymatic assays (Cayman Chemicals). Frozen liver tissue (25-75mg) was homogenized in chloroform/methanol (8:1 v/v; 500 µL per 25 mg tissue) and shaken at RT for 8–16 hours. H_2SO_4 was added to a final concentration of 0.28M. After centrifugation, the lower phase was collected, dried, and TG content was measured using an enzymatic assay (Caymen Chem).

qRT-PCR was performed using the GoTaq® qPCR Master Mix and an Eppendorf light cycler. Expression levels were calculated using the ΔCt-method. Data were normalized to a housekeeping gene (rps27 or rpl0). Immunoblot analysis was performed using standard protocols and following antibodies: ACC, PPARγ, CEBPβ, c-Jun, phospho-CREBP, total CREB (Cell Signaling), Vinculin (Sigma), PPARγ, Parp-1, CEBPα, c-Fos, Fra-1 (Santa Cruz), HNF4, FAS, Fabp1 (Abcam), and Fsp27 (Novus Biologicals), SREBP-1/2 (BD Bioscience). Nuclear extracts from liver tissue were obtained using the NE-PER Nuclear Protein Extraction Kit (Pierce). For histology, H&E- and ORO-staining were performed using standard procedures. ORO-positive areas were quantified as previously described [300]. IHC was performed as described [266] using following antibodies: PPARγ (Cell Signaling), CD45 (Abcam), F4/80 *(AbD Serotec).*

For generating array data, the RNA was first isolated using the RNEasy Midi kit (Qiagen) and RNA integrity was evaluated using an Agilent 2100 Bioanalyzer. Samples of RNA integrity score above 7.8 were used for microarray analysis. 100ng of RNA was labeled with Cy3 (RNA pool from at least 5 control mice, which were either fed CD or HFD) or Cy5 (RNA samples from individual mutants, which were either fed CD or HFD) using the Low Input Quick Amp Labeling Kit Version 6.5 (Agilent). Labeled RNAs were purified using RNeasy spin columns (Qiagen) and hybridized to a mouse gene expression array G3 8x60K (Agilent microarray design ID 028005, P/N G4852A). On each array, the Cy3-labeled control pool and one Cy5 labeled mutant sample were hybridized at 65°C for 17 hours. The microarray was scanned on a 2505C DNA microarray scanner (Agilent) and images were analyzed using the Feature Extraction Software Version 10.7 (Agilent). Multiple testing correction was performed using the Benjamin-Hochberg procedure. Array data are deposited in NCBI's Gene Expression Omnibus and are accessible through GEO Series accession number GSE52275 (G/LOF mice) or GSE60149 (BXD mice). Hepatic gene expression of the BXD strains was analyzed using Mouse Gene 1.0 ST Arrays (Affymetrix) and is also accessible on **www.genenetwork.org**. Standard array analysis methods were used, e.g. RMA normalization, as described elsewhere [301].

For analyzing the array data, microarray data were analyzed separately in CD and HFD conditions by comparing control to mutant livers. All nominally significant changes with fold change ≥ 1.5 were retained. Gene sets were then winnowed using multiple testing correction (Benjamin-Hochberg) and entered independently into the Web Gestalt (**bioinfo.vanderbilt.edu/webgestalt/**) resource. Enriched pathways were generated based on KEGG gene ontology annotations. The PPARγ signaling pathway was found significantly modulated in both dietary conditions. The two independently-generated pathways were then overlaid and redrawn to generate the pathway diagram.

For the *in vitro* experiments, HuH7 and 293T cells were cultured in DMEM/10%FCS at 37°C and 5% CO_2. For reporter assays, 0.8x 10^5 HuH7 or 293T cells were plated per well of a 24-well plate. 24 hours later, 0.01µg Renilla vector, 0.2µg *PPARG2*-luc vector [302] and 0.6µg pCMV-AP-1 or pCMV-empty control vector were transfected using Lipofectamine 2000 (Invitrogen). Cells were harvested 48h after transfection and luciferase activity was analyzed using the Dual-Glo Luciferase Assay (Promega). ChIP was performed using following antibodies: Flag (F3165, Sigma), Fra-1 (SC-183, Santa Cruz), Fra-2 (rat, CNIO polyclonal), c-Fos (PC-05, Calbiochem), c-Jun (BD), JunB (SC-73, Santa Cruz), JunD (CS5000, Cell Signaling). Pan Jun ChIP with HuH7 cells has been performed with a mixture of 2 antibodies raised against an epitope present in all Jun proteins. Statistical significance was calculated using Student's two-tailed *t*-test if not indicated otherwise, *: $p < 0.05$; **: $p < 0.01$; ***: $p < 0.001$, ****: $p < 0.0001$.

3.3.6 Acknowledgments

We thank Drs. N. Djouder, M. Perez-Moreno, R. Ricci, M. Serrano and G. Sumara for critical reading of the manuscript and valuable suggestions; the CNIO Transgenics Unit and G. Luque and G. Medrano for technical help with mouse procedures. Erwin F. Wagner was supported by the Banco Bilbao Vizcaya Argentaria Foundation (F-BBVA), a grant from the Spanish Ministry of Economy (BFU2012-40230) and an ERC-Advanced grant ERC-FCK/2008/37. Martin K. Thomsen was supported by a Juan de la Cierva postdoctoral fellowship. Sebastian C. Hasenfuss received a Boehringer Ingelheim Fonds (BIF) PhD fellowship and an EMBO short term fellowship (ASTF 198–2012).

Chapter 4 Holistic Omics Analysis

Adapted from Wu Y.*, **Williams E.G.*,** Dubuis S., Mottis A., Jovaisaite V., Houten S.M., Argmann C.A., Faridi P., Wolski W., Kutalik Z., Zamboni N., Auwerx J., Aebersold R. Multilayered genetic and omics dissection of mitochondrial activity in a mouse genetic reference population, *Cell*, 2014. *Co-first author

The central dogma of molecular biology states that genetic information encoded in DNA is first transcribed by RNA polymerase, then translated by ribosomes into proteins [303]. However, knowing the DNA sequence of a gene is insufficient to predict the expression levels of its RNA and protein products, the key parameters that will ultimately determine phenotypic variation. Consequently, to link genomic variants to phenotypic traits, RNA transcript and protein levels need to be precisely quantified. Since the advent of microarray technology in the 1990s [15], and more recently by next generation sequencing [304], comprehensive gene expression patterns—i.e. the transcriptome—can be precisely quantified across tissues and conditions. Unfortunately, transcript levels generally have only modest correlation with the levels of corresponding proteins [305-307], and genetic variants similarly affecting both the transcript and peptide levels of a gene are relatively uncommon [308, 309]. As proteins in most cases are more directly responsible than transcripts in the regulation of cellular pathways—and ultimately phenotypic traits—there is a critical need for efficient, large-scale and accurately quantitative proteomics methods to complement transcriptomic datasets

Over the past decade, the development of discovery mass spectrometry ("shotgun") has allowed the first large-scale studies on quantitative proteomics. In this approach, protein extracts are cleaved into short peptide sequences, which are then chromatographically separated and analyzed by tandem mass spectrometry. This allows the untargeted discovery of thousands of peptides, but if the number of unique peptide fragments in a sample significantly exceeds the number of available sequencing cycles (as in whole proteome extracts), any individual peptide will be inconsistently sampled across repeat analyses. This reduces the technical reproducibility, and moreover means that the number of peptides consistently quantified across all (or most) samples decreases as the study size increases [151]. Consequently, discovery mass spectrometry strategy has yielded mixed results in large population studies [305, 310], particularly as specific peptides of interest cannot be targeted, and the most consistently identified peptides are biased towards those of higher abundance [311]. To overcome these hurdles, selected reaction monitoring (SRM) was developed, which perfects technical reproducibility and allows consistent multiplexed quantitation of target proteins by deploying a mass spectrometric measurement assay that is specific for each targeted peptide [312]. Thus, hundreds of target peptides can be consistently and accurately quantified across large populations of samples. Recent studies in yeast have shown that the proteins and transcripts of genes are typically controlled by different, distinct mechanisms [309, 313]. However, these hypotheses have not been well tested in mammalian genetic reference populations (GRPs) through multilayered transcriptomic and proteomic strategies.

Large GRPs are frequently used to determine the extent to which phenotypic variation is driven by genetic variants (i.e. heritability), and to subsequently identify genes driving such variation. These genes can be identified by genome-wide association (GWA) or by quantitative trait locus (QTL) mapping, approaches that have been applied to various species and have led to the successful identification of dozens of major allelic variants [16, 26, 314, 315]. In mammals, the murine BXD family is the largest and best studied GRP, consisting of ~150 recombinant inbred strains descended from C57BL/6J (B6) and DBA/2J (D2) [16]. Using 40 strains of this population on both chow (CD) and high fat (HFD) diets, we have obtained major metabolic phenotypes and established a multilayered dataset focused on 192 metabolic genes expressed in the liver. For all genes, we know the sequence variants, transcript levels, and protein levels in all cohorts. These data are further supplemented with targeted metabolite analysis in liver and serum, generating the first large-scale multilayered quantitative picture of any cellular process in the BXD population.

4.1 Results

4.1.1 Protein Targeting Across a Genetically and Environmentally Diverse Murine GRP

We first selected 192 metabolic proteins for study, with particular focus on genes regulating mitochondria and general energy metabolism. For each gene, synthetic peptides were generated based on established assays [316] (Fig. 4:1A) to accurately quantify each protein across all cohorts. To validate peptide measurements, we compared the coefficients of variation (CV) among technical and biological peptide replicates. Both technical and biological replicates showed a high degree of reproducibility (CV ~ 0.09 and 0.12 respectively), indicating the peptides are sufficiently accurate for the quantification of targeted peptides across large, diverse populations. More importantly, the results for biological replicates reveal nearly equally high overall reproducibility ($r > 0.98$; Fig. 4:1B), suggesting low biological variance within each cohort, particularly relative to cross-cohort variance (CV ~ 0.33). 82.1% of the quantified peptides are highly variable (CV > 0.20), indicating that the biological error (i.e. variation within a cohort) is much smaller than the variation induced by differences in genotypes and diet. We observed nearly complete quantification of all 192 proteins across all cohorts, with only ~60 missing peptide counts out of ~15,000 measurements (i.e. 99.6% completion; Fig. 4:1C, top). This completion is similar to the level of completeness achieved by microarrays (Fig. 4:1C, bottom) and a major contrast to shotgun-acquired proteomics, which typically have completeness of ~70% [151].

It has been well-established that transcriptomic networks of many metabolic processes covary quite well, e.g. within the electron transport chain or within the citric acid cycle [139]. On protein level, proteins which function in common biological processes or which localize to the same functional modules are also reported to be subject to similar regulatory process and generally covary [108]. To validate and identify which of these 192 proteins vary synchronously, we computed the robust Spearman correlation network for all protein pairs using the full SRM dataset (Fig. 4:1D). The resulting network contained 82 correlated nodes (proteins) with 211 edges in 3 main enrichment clusters. As expected, many of the mitochondrial proteins and proteins involved in lipid metabolism are highly correlated (Fig. 4:1D, cluster a and b). Within cluster a. are five of the six measured proteins involved in mitochondrial unfolded protein response (UPR^{mt}) (HSPD1, HSPE1, HSPA9, CLPP, and LONP1 indicated in red), along with three of the four measured NADH dehydrogenase genes (NDUFA1, NDUFB3 and NDUFS6—in blue), and four of the eight measured mitochondrial-encoded proteins (MT-CYB, MT-CO2, MT-CO3, and MT-ND3 in black). Meanwhile, proteins involved in carbohydrate metabolism are enriched in the same cluster (Fig. 4:1D, cluster c). These results show that functionally-related proteins tend to be coordinately regulated at a protein level, and that coregulation of protein abundance is strongly maintained for certain processes. To validate biological significance of these function-based covariation clusters, we further investigated one, the UPR^{mt} network (elaborated in Fig. 4:6).

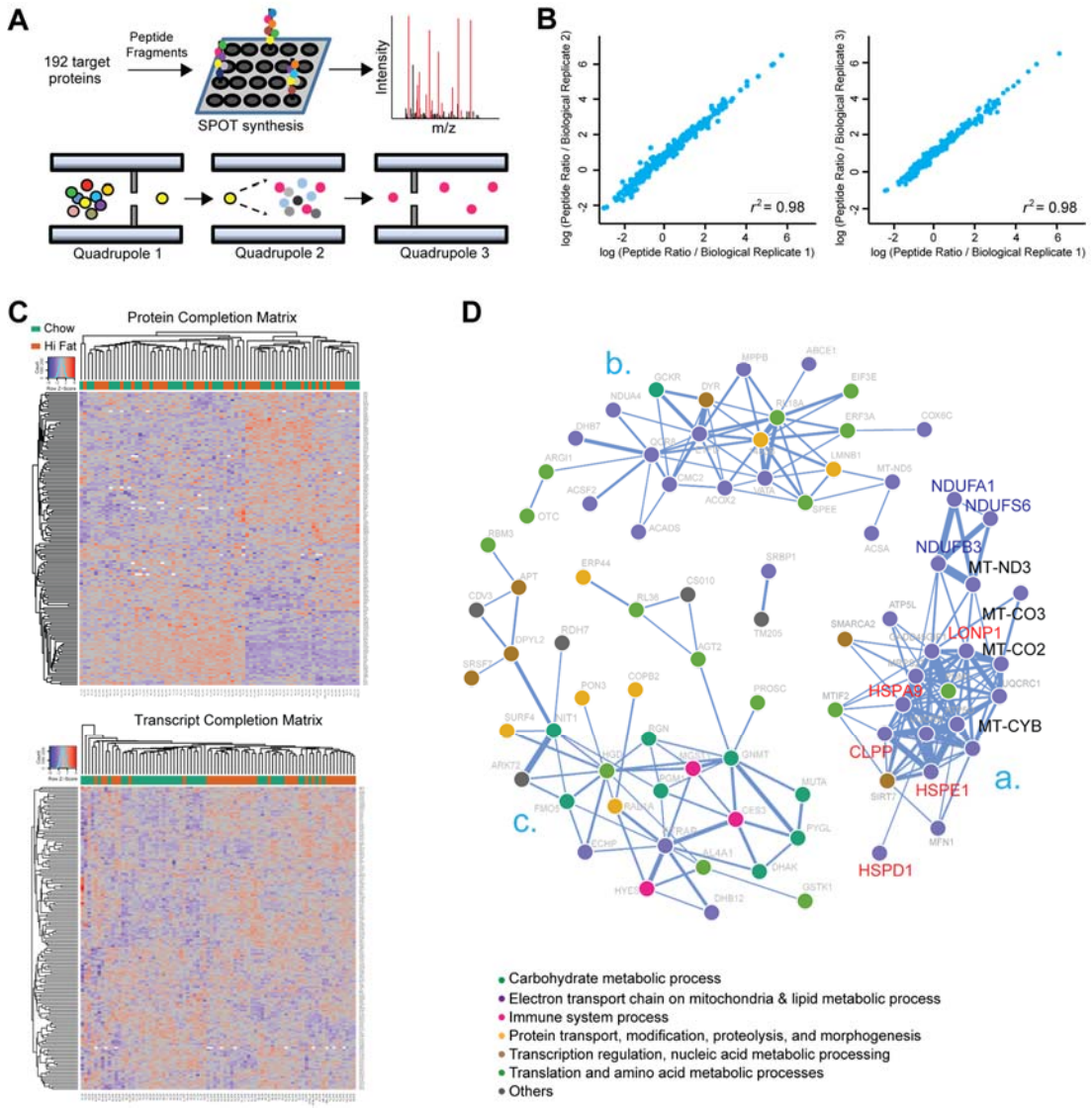


Figure 4:1 **SRM-based Protein Quantification and Covariation Network**

(A) SRM assay development for targeted proteomic measurements. 309 peptides corresponding to 192 genes were designed and synthesized via SPOT synthesis. Fragment ion spectra were generated on a triple-quadrupole MS with SRM-triggered MS2 mode, then ions were selected based on their relative intensities. Dot colors indicate different amino acids. m/z, mass-to-charge ratio. Mouse liver homogenate combined with the heavy reference proteome was analyzed with SRM on a triple-quadrupole MS. Different dot colors represent different peptides (Quadrupole 1) or product ions (Quadrupole 3). **(B)** Biological replicates had near-perfect reproducibility (r ~ 0.98), shown here for all three biological replicates of the BXD60 CD cohort. **(C)** Two-way cluster analysis of protein (upper) and transcript (lower) abundances in all 77 cohorts (40 CD, 37 HFD). Columns are clustered based on samples, and rows are clustered based on gene-product abundances. Protein and transcript abundances are colored in a red-blue scale. Red: high abundance. Blue: low abundance. White: missing data. **(D)** Protein association network based on robust Spearman correlation measures for all protein pairs. Statistically significant and strong positive associations ($p < 0.01$ & $r > 0.6$) are edges. The largest correlation clusters are labeled “a.”, “b.”, and “c.”. Nodes are labeled with protein names and colored according to their biological process, as reported by DAVID [270]

4.1.2 Protein and mRNA Gene Products Generally Do Not Correlate

With the general protein measurements validated, we generated a global overview of how genotype and diet influence differential transcript and protein expression (Fig. 4:2A). At the genetic level, transcripts and proteins map to an equivalent number of significant QTLs: we detected 65 significant transcript QTLs (eQTLs—blue lines at center) and 57 significant protein QTLs (pQTLs—red lines at center). However, though the total number of eQTLs and pQTLs are roughly equivalent, the predominant type of regulation was very different: 74% of the eQTLs are *cis*-mapping (ratio of solid to dashed blue lines), versus only 31% of pQTLs (ratio of solid to dashed red lines). This indicates a closer connection between a transcript and its gene than a protein and its gene. In general, *trans*-mapping proteins and transcripts mapped evenly across the genome, not yielding any clear "hot spots" for these metabolic genes (sample magnification of QTLs mapping to chromosome 5 in Fig. 4:2B).

Within each diet, we found proteins and transcripts to be nominally correlated for ~25% of genes (i.e. Spearman ρ correlation *p*-value < 0.05; Fig. 4:2C, diets considered separately), similar to findings in other species and populations [305, 306, 308, 317]. Of the correlated transcript–peptide pairs (46 in CD, 55 in HFD), 31 correlate significantly in both diets (16%), while 70 pairs (36%) correlate in at least one diet. Correspondingly, while ~50% of genes were affected by diet, transcripts were more frequently influenced (84 of 189) than proteins (37 of 192), with only 21 genes affected at both the transcript and protein level (Fig. 4:2D). Genes which were the most strongly affected by diet at the transcript level (Fig. 4:2E) tend to be similarly affected at the protein level (Fig. 4:2F)—e.g. *Cyp3a11* is higher in CD—though exceptions are frequent, e.g. *Srebf1* mRNA is induced by HFD, but unaffected at the protein level, while in another counter-example, the HFD cohorts have more ETFDH protein, but less of the transcript. Thus, while transcripts and proteins are moderately covarying estimations of their gene's activity and typically have covarying responses to external factors (e.g. diet), these trends are too weak to support the measurement of any one particular transcript to serve as a proxy for the protein (or vice versa) without prior knowledge.

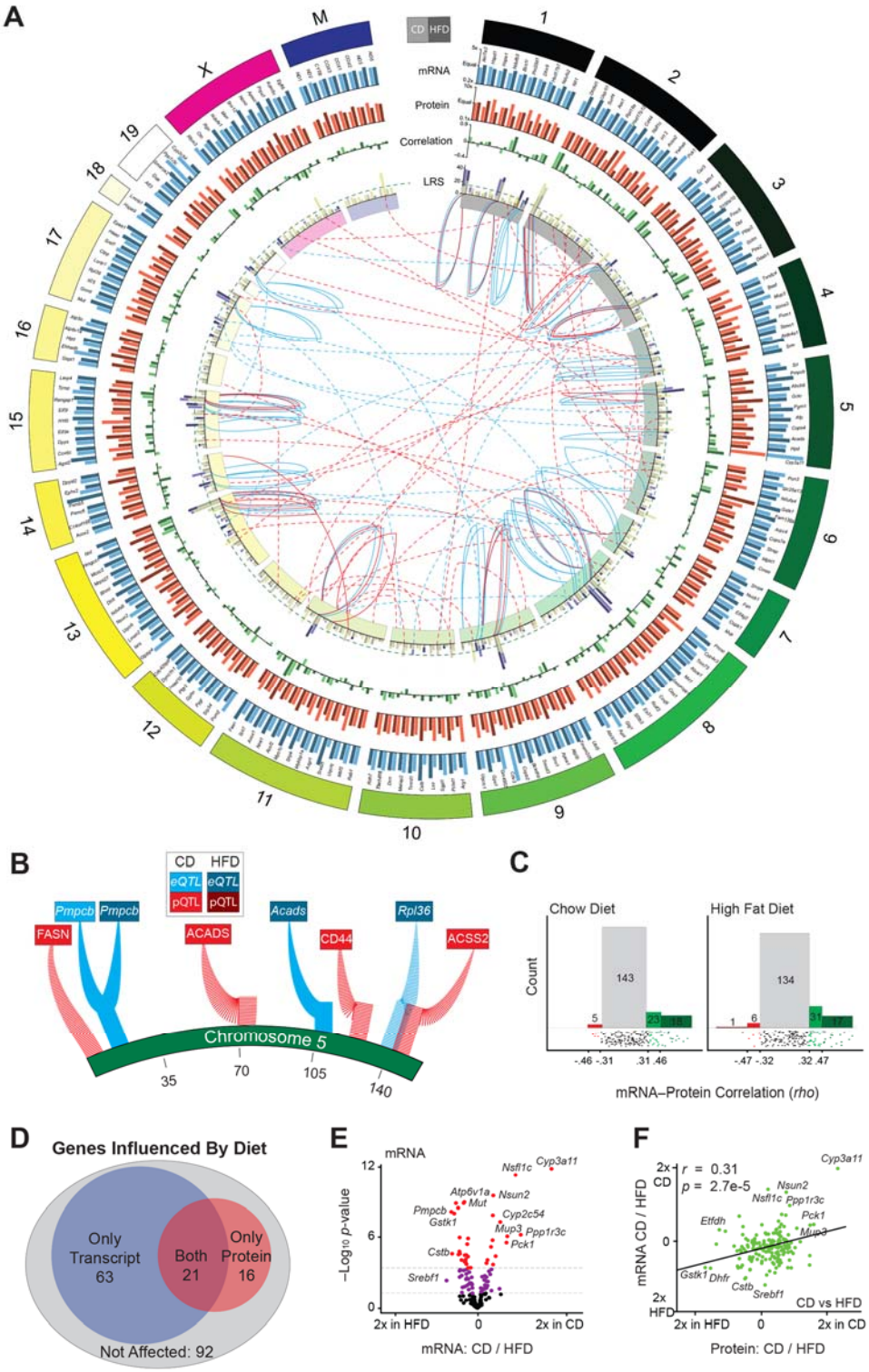


Figure 4:2 **mRNA and Protein Overview**

(A) Circos plot of mRNA and protein data for all 192 genes, labeled on outer edge. Genes are represented by two bars: light for CD, dark for HFD. Genes are arranged by relative chromosome position; chromosome length is according to number of genes measured. **Blue Bars**: Transcript relative expression CD vs. HFD; **Orange Bars**: Protein relative expression CD vs. HFD. Bars with more unequal heights indicate diet has a larger

impact. **Green Bars**: Correlation between transcript and protein for each gene within each diet. **Purple/Yellow Bars**: Fuchsia bars represent the strength of the peak eQTL. Yellow bars represent the strength of the peak pQTL. The two bars are overlaid with transparency. The dashed green line represents the simplified significance cutoff (LRS ≥ 18). **Inner Ring:** Chromosome location. **Central Lines:** Blue lines for significant eQTLs, red for significant pQTLs. Solid lines represent *cis*-QTLs, dashed lines represent *trans*-QTLs. QTL lines stem from the LRS bar graph, and terminate on the inner side of the chromosome ring at the approximate QTL location. **(B)** Magnified view of eQTLs/pQTLs mapping to chromosome 5. **(C)** In CD ~25% (left) and in HFD ~30% (right) of transcripts correlate nominally significantly with their protein. The lower strip charts show correlation distribution. Spearman correlation values corresponding to nominal significance ($p < 0.05$) and corrected significance ($p < 0.0002$) are displayed on the axis. **(D)** Venn diagram of genes which are differentially regulated between CD and HFD as transcripts (blue), proteins (red), both (purple), or neither (grey). **(E)** Volcano plot for mRNA showing magnitude of dietary effect versus significance. ~45% vary with nominal significance ($p < 0.05$) between the dietary conditions. ~19% vary with corrected significance (raw-p < 0.0003). Some extreme genes are labeled. **(F)** Plot of the effect of diet on transcripts vs. the effect of diet on proteins. In general, transcripts and proteins are similarly affected by diet

4.1.3 Most Transcript and Protein QTLs Do Not Overlap

Of the 192 target genes, 79 map to a significant eQTL or pQTL in at least one dietary condition. A strong majority of significantly mapped QTLs, ~80%, are unique to either the transcript or protein level (Fig. 4:3A). At the transcript level, 28 genes map to *cis*-eQTLs, 20 of which in both diets, while 17 transcripts map as *trans*-eQTLs, none of which are observed in both diets. Together, we observed 65 significant eQTLs stemming from 45 distinct genes (Fig. 4:3B left; the 20 distinct genes which map *cis* in both diets contribute 40 significant eQTLs). The range of transcript variance within a diet was a strong predictive factor for observing an eQTL. Transcripts in the least variable quartile (range < 1.5 fold from the lowest to the highest expressing BXD strain) contained only 10% of the significant QTLs. In contrast, the 2nd quartile (range 1.5–1.65 fold) contained 17%, the 3rd quartile (range 1.65–2.0 fold) contained 29%, and the top quartile (range ≥ 2.0 fold) contained 41% of the significant QTLs.

At the protein level, 57 significant pQTLs stem from 48 distinct proteins (Fig. 4:3B right). In striking contrast to transcript regulation, only 13 distinct proteins map to *cis*-pQTLs, five in both diets, while 36 proteins map as *trans*-pQTLs. No *trans*-pQTLs are consistent across diets, although three proteins significantly map to separate *trans*-pQTLs in both diets (*Cd44, Acss2,* and *Ndufs6*), and one (*Hmgcs1*) maps to both *cis* and *trans*-pQTLs. Together with transcript data, we could moderately predict pQTLs: of the 28 genes with *cis*-eQTLs, 12 correlated well ($r > 0.40$) between transcript and protein. Of these 12 genes, five (*Dhtkd1*, *Nnt*, *Tymp*, *Gclm,* and *Bckdhb*) map significantly as pQTLs and eQTLs in both diets (Fig. 43C–D; *Bckdhb* in Fig. 4:4B), and four (*Acox1, Mfn1, Mri1, Pm20d1*) map as pQTLs and eQTLs in at least one diet. Interestingly, three proteins mapped as *cis*-pQTLs and correlated with their transcript, but did not have associated eQTLs (*Hmgcs1*, *Car3*, and *Acsf2*), suggesting the genetic variant driving these changes manifests more prominently at the protein level than for the transcript (e.g. Fig. 4:3E).

Diet also plays a clear role in the consistent identification of *cis*-QTLs. For the 20 genes with *cis*-eQTLs identified in both diets, HFD significantly affects the levels of only one (*Pmpcb*; Fig. 4:3F). Conversely, for the 8 genes mapping to an eQTL in only one diet, the levels of only three are affected (*Ndufs2, Acads,* and *Aldh4a1*). Similar trends remain are observed at the protein level: of the five *cis*-pQTLs found in both diets, none are affected by diet, while for the 8 genes with *cis*-pQTLs in only one diet, only the levels of one (*Hmgcs1*) is affected. This relationship is similar for *trans*-eQTLs: of the 17 genes with *trans*-eQTLs, only four are influenced by diet (whereas ~50% of total transcripts are influenced by diet). However, diet does not appear to strongly influence *trans*-pQTLs, where 11 of the 39 are affected by diet, a similar proportion as the overall effect, recalling that ~20% of proteins have levels influenced by diet (Fig. 4:2D). Together this suggests diet can change the genetic factors regulating these genes without changing their overall levels.

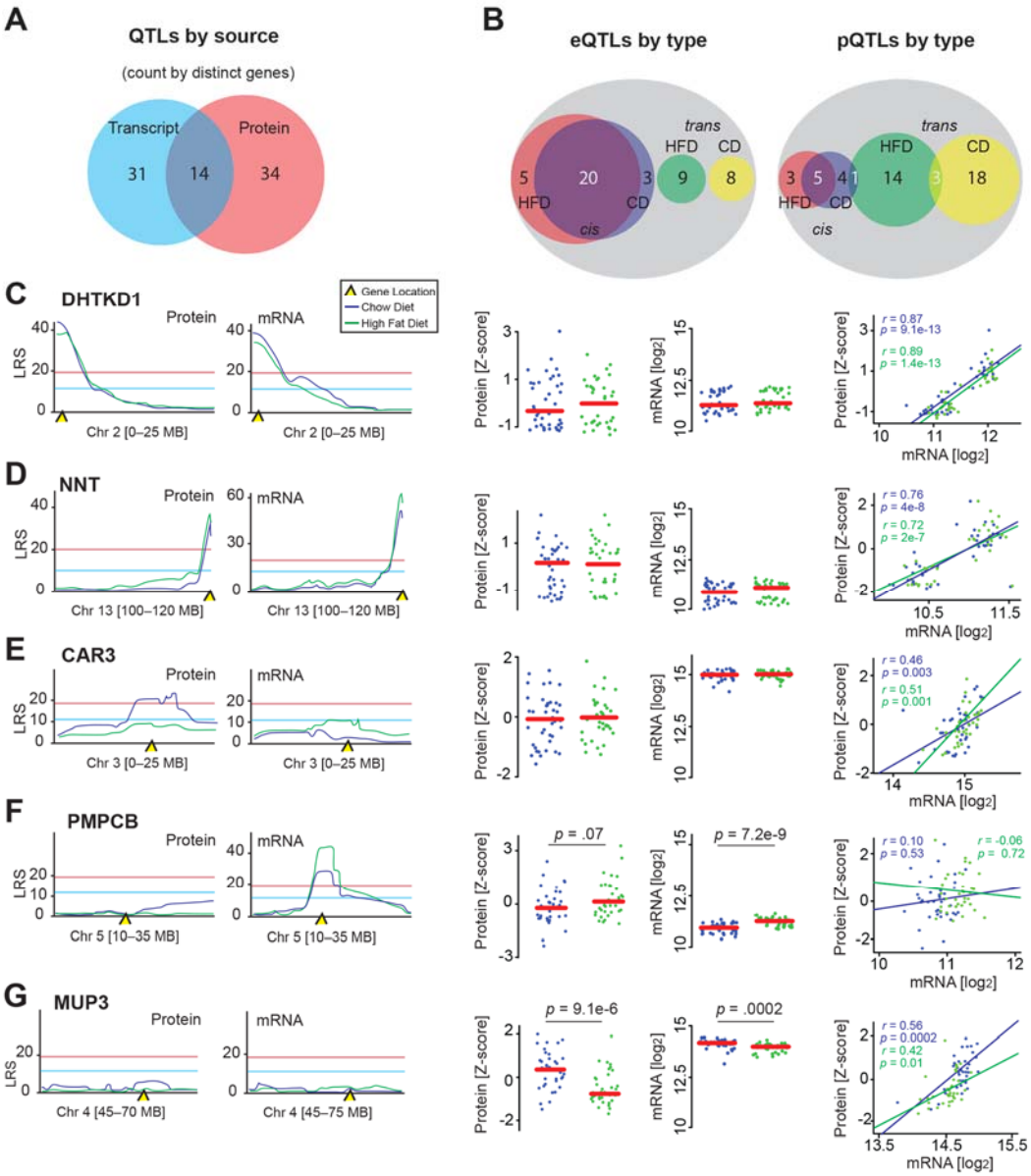


Figure 4:3 **QTL Overview**

(A) Venn diagram separating all genes with distinct QTLs based on provenance. **(B)** Venn diagrams separating eQTLs or pQTLs by dietary source and regulatory mechanism. Overlapping regions indicates genes giving an eQTL or pQTL in both diets; white numbers are counted twice for QTL count, but once for distinct gene count (e.g. there are 48 significant *cis*-eQTLs which stem from 28 distinct genes). **(C)** In both diets, DHTKD1 and *Dhtkd1* share a common *cis*-QTL, are unaffected by diet, and strongly correlate. **(D)** NNT and *Nnt* display a similar pattern. **(E)** *Car3* has only one significant pQTL despite an absence of dietary effect and a strong transcript–protein correlation. **(F)** PMPCB does not map to a significant pQTL, however *Pmpcb* maps to a significant *cis*-eQTL in both diets, despite a strong transcriptional upregulation by HFD. The transcript and protein levels do not correlate. **(G)** *Mup3* and MUP3 do not map to significant QTLs, despite having high levels of variation and a strong transcript–protein correlation

While high expression variability is associated with stronger QTLs, it is not a definitive predictor, e.g., *Mup3* is highly variable, strongly affected by diet, and consistently expressed between mRNA and protein, but does not map to a QTL in any measurement (Fig. 4:3G). Therefore, while the ability to predict peptide levels based on transcript measurements on a systems scale is quite powerful (the ~25%–~37% of correlated transcript–peptide pairs is much better than the

~5% expected by chance), the probability to fail on any one particular gene is quite high. These chances can be adjusted somewhat—perturbations dramatically affecting transcript levels are more likely to manifest at the protein level and vice-versa—but even so, prior research must be established before gene expression can be confidently considered a proxy for protein levels in targeted genetic studies. This also indicates that applying quantitative proteomic data to pathways established at the transcript level can indicate new links that were previously obscured.

4.1.4 Functional Relationships of pQTLs to Phenotypes

To characterize the cellular function and potential physiological relevance of the pQTLs, we first collated all Entrez [318] and UniProt [319] entries for genes with the most significant pQTLs (LRS ≥ 20) (Table 4:1). As the BXDs have extensive historical phenotype and metabolite data available on GeneNetwork [181], we performed a phenome-wide association study to determine if any collected phenotype data mapped at least suggestively (LRS ≥ 12) as clinical QTLs (cQTL) to the same loci. A handful of phenotypic connections in the BXDs were supported by literature, including a link between *Nnt* and insulin [320], and between *Car3* and subcutaneous adipose mass [321]. However, for the majority of pQTLs, no established cQTLs mapped to the same loci. We thus selected the two genes with the most significant and novel pQTLs in both diets for follow-up analysis and validation—*Bckdhb* and *Dhtkd1*.

Protein	Diet	LRS	QTL Area [Mb]	Gene	Function	Phenotype
DHTKD1	CD	43	Chr2: 3–8	cis	Lysine Catabolism	Reye Syndrome, Glucose Regulation
DHTKD1	HFD	41	Chr2: 3–8	cis	Lysine Catabolism	Reye Syndrome, Glucose Regulation
NNT	CD	37	Chr13: 119–120	cis	NADPH Regeneration	Insulin Secretion
GCLM	HFD	34	Chr3:118–124	cis	Glutathione Synthesis	Detoxification, Ischemia, Anemia
NNT	HFD	31	Chr13: 119–120	cis	NADPH Regeneration	Insulin Secretion
PRPS1	CD	31	Chr6: 61–66	73	Nucleotide Synthesis	Arts Syndrome
GPX1	CD	30	Chr11: 119–122	70	Glutathione Peroxidase	Detoxification, Cardiomyopathy
PM20D1	HFD	29	Chr1: 134–137	cis	Zinc Binding (Putative)	Tuberculosis Sensitivity
MT-ND3	CD	28	Chr3: 11–18	21	Complex I Subunit	Leigh Syndrome
CMAS	CD	28	Chr7: 95–105	70	Sialic Acid Adhesion	Fibrosarcoma, Dermatomyositis
MRI1	CD	28	Chr8: 88–90	cis	Tyrosine Phosphorylation	Cell Invasion Regulation
NDUFS6	CD	27	Chr3: 11–18	21	Complex I Subunit	Lactic Acidosis
PUM2	HFD	27	Chr10: 106–109	7	RNA Binding	Development, Cell Differentiation
BCKDHB	CD	26	Chr9: 83–85	cis	BCAA Degradation	Maple Syrup Urinary Disease
BCKDHB	HFD	24	Chr9: 83–83	cis	BCAA Degradation	Maple Syrup Urinary Disease
TYMP	CD	24	Chr15: 89–90	cis	Thymidine Phosphorylation	Angiogenesis, mtDNA Reg.
RPL18A	CD	24	Chr3: 25–31	33	Ribosome 60S Subunit	Hepatitis Susceptibility
CAR3	CD	24	Chr3: 14–21	cis	Carbon Dioxide Hydration	Laryngeal Disease, Adipogenesis
DAK	HFD	24	Chr4: 128–130	34	Dihydroxyacetone Kinase, FMN	Glycerolipid Metabolism
NDUFA4	HFD	24	Chr11: 33–40	5	Complex I Subunit	Glucose Regulation
ACADS	CD	23	Chr5: 67–76	42	Beta Oxidation, Short Chain	SCAD Deficiency
HMGCS1	HFD	23	Chr4: 80–84	10	Cholesterol Synthesis	Cholesterolemia
FASN	CD	23	Chr5: 137–139	38	Fatty Acid Synthesis	Obesity
GCLM	CD	23	Chr3: 118–124	cis	Glutathione Synthesis	Ischemia, Anemia
ABCE1	CD	23	Chr13: 114–121	28	Ribonuclease L Inhibitor	Viral Load, RNA Stability
ABCB8	CD	23	Chr5: 23–25	cis	Mito Membrane Transporter	Iron Homeostasis
ARG1	HFD	22	Chr11: 64–66	13	Urea Cycle	Argininemia, Hyperammonemia
DCN	HFD	22	Chr11: 51–52	11	Collagen Assembly	Cell Growth and Proliferation
AKR7A2	HFD	22	Chr1: 23–24	cis	Succinic Semialdehyde Redox	Detox of Aldehydes and Ketones
AFP	HFD	21	Chr6: 86–88	29	Fetal Serum Albumin (Putative)	Copper, Fatty Acid, Bilirubin Binding
COPB1	CD	21	Chr16: 8–15	56	Golgi Complex Transporter	Lipid Storage, LDL Levels
ACSS2	HFD	21	Chr11: 115–121	101	Acetate Activation	Lipid Synthesis, Energy Generation
TYMP	HFD	21	Chr15: 89–90	cis	Thymidine Phosphorylation	Angiogenesis, mtDNA Reg.
COX6C	CD	20	Chr7: 141–144	14	Cytochrome C Electron Transfer	Energy Generation
MYBBP1A	HFD	20	Chr14: 112–116	1	Regulation of Ribosomal DNA	Nucleolar Stress Response
GSTK1	HFD	20	Chr4: 67–70	3	Glutathione Regulation	Cellular Detoxification
EPAS1	HFD	20	Chr10: 106–109	7	O2-responsive Txn Factor	Erythryocytosis, Hypoxia Response
ACSF2	CD	20	Chr11: 116–118	cis	Fatty Acid Metabolism	Tuberculosis Sensitivity
SRSF7	HFD	20	Chr15: 40–48	11	Spliceosome, mRNA Export	Immune Response

Table 4:1 **pQTLs Link to Phenotypes**

A list of all genes with highly significant (LRS ≥ 20) pQTLs. For each gene, the significant peak QTL width considers a 5 LRS dropoff on either side of the peak. A count of all genes under the wide peak is displayed, unless it is a *cis*-pQTL as the candidate gene can be easily identified (itself). The cellular function and overt physiological effect described by literature for each gene is given.

Bckdhb is a subunit of the branched chain amino α-keto acid dehydrogenase (BCKD, EC 2.7.11.4) complex (Fig 4:4A) and maps to the two of the most significant eQTLs and among the most significant consistent pQTLs (Fig. 4:4B and

Table 1; eQTLs not all shown). *Bckdhb* transcript and protein levels are highly variable across strains and are unaffected by diet (Fig. 4:4B), but no cQTL mapped to this locus. In humans, variant alleles of *BCKDHB* are known to cause Type IB maple syrup urinare disease (MSUD), an inborn error of metabolism characterized by a buildup of branched chain amino acids (BCAAs) and branched chain α-keto acids (BCKAs). MSUD manifests in newborns and is associated with neurological problems and delayed development, and if untreated it can be fatal [322]. This pathway is a multi-tissue system driven particularly by the muscle and liver, and consequently is diagnosed by significant increases of BCAAs and BCKAs in both serum and urine, a decrease in alanine, and/or by detection of L-alloisoleucine [322, 323]. Genetically engineered mouse models of MSUD have similar phenotypic progression as humans, including the same metabolite buildups and related phenotypes such as movement difficulties and reduced body growth [324, 325]. We thus examined the BXDs with respect to these traits. While L-alloisoleucine was not detected, strains with the less functional B6 allele of *Bckdhb* had a marked increase in the BCAA/alanine ratio in serum and liver (Fig. 4:4C left and right, respectively). However, no association with movement, growth, or body weight was present in the BXDs (data not shown), suggesting that this *Bckdhb* variant carries only the subclinical metabolic effects of MSUD. This observation is also in line with the human disease, where MSUD symptoms are only apparent when BCKD activity is < ~30% of normal [326]. In contrast, strains with the B6 allele of *Bckdhb* have ~50% the enzymatic activity compared to those with the D2 allele. However, this raises the possibility that a sustained BCAA loading test in the BXDs may lead to clinical manifestation of the intermittent form of MSUD.

To continue scanning for physiological links with the proteins examined, we also investigated the strongest pQTL in both diets. This gene, *Dhtkd1*, encodes a mitochondrial protein operating as the E1 subunit of a dehydrogenase complex involved in lysine metabolism (EC 1.2.4.2, Fig. 4:4D). In humans, variants in *DHTKD1* have been linked to urinary buildup of 2-aminoadipate (2-AA) and 2-oxoadipate, yet the few studies on the physiological consequences of these variants are conflicted [327-329]. After measuring the metabolites involved in this pathway in the BXD liver and serum, we observed a striking correlation between DHTKD1 levels and 2-AA (Fig. 4:4E left) as well between metabolites in the pathway itself (Fig. 4:4E right). Moreover, 2-AA mapped significantly in both diets as an mQTL to the locus of *Dhtkd1* itself, indicating the BXDs share an inborn error of metabolism similar to humans (Fig. 4:4F). Due to the strength, and consistency of this correlation across diets, we hypothesized this variant may lead to phenotypic consequences and

inform for human validation.

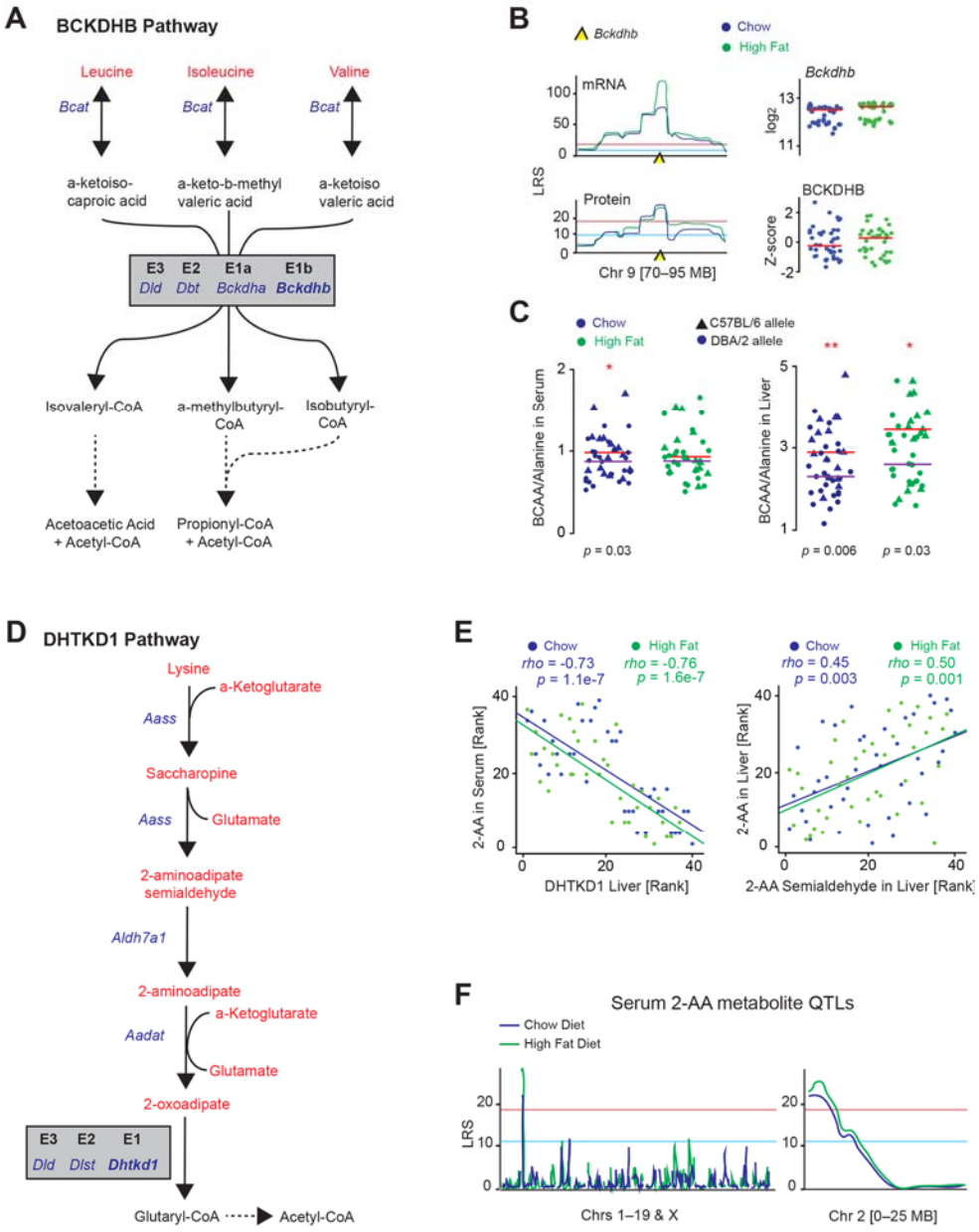


Figure 4:4 **Metabolic Consequences of BCKDHB and DHTKD1**

(A) BCKDHB is the E1b subunit of the BCKD complex, which irreversibly converts several BCKAs. **(B)** *Bckdhb* mRNA has the strongest eQTLs of the 192 target genes examined in the liver, and has among the strongest pQTLs. Neither transcript nor protein are affected by diet. **(C)** The BCAA/alanine ratio is significantly increased in animals with the dysfunctional C57BL/6 allele in CD serum and liver, and HFD liver measurements, in line with it acting as a risk for MSUD. Purple line = median for strains with the C57BL/6 allele, red line = median for strains with the DBA/2 allele. **(D)** DHTKD1 is the E1 subunit of the dehydrogenase complex that catalyzes the irreversible conversion of 2-oxoadipate to glutaryl-CoA. **(E)** Left: Serum 2-AA levels are strongly related to DHTKD1, as are liver levels (not shown). Right: Other upstream metabolites in the DHTKD1 pathway correlate strongly with one another, e.g. 2-AA and 2-AA semialdehyde, shown here in liver. **(F)** Serum 2-AA maps as an mQTL in both diets to proximal chromosome 2, the location of DHTKD1.

4.1.5 DHTKD1: A Regulator of Glucose Homeostasis

Inhibition of *DHTKD1* in human liver cells is known to diminish mitochondrial activity [330], and external administration of 2-AA improves insulin secretion in mice [92]. We thus examined whether the natural genetic variants in *Dhtkd1* may also influence glucose homeostasis in the BXDs. The BXDs have varying responses to HFD-induced diabesity and include some resistant strains, but most strains are affected, leading to general increases in liver weight, serum cholesterol, and fasted glucose (Fig. 4:5A). Strikingly, 2-AA is negatively associated with liver mass and fasted glucose levels in both diets, and with serum cholesterol and insulin in CD (Fig. 4:5B). Other metabolites in this pathway were also connected to the same phenotypes and with negative correlation (e.g. saccharopine, Fig. 4:5C). Interestingly, 2-AA levels are significantly decreased in the HFD cohorts (Fig. 4:5D), despite that the genetic regulation of 2-AA by *Dhtkd1* is equivalent between CD and HFD cohorts (Fig. 4:4E–F). However, this decrease is not directly linked to diet, but instead with the animals' progression towards diabetes. Based on HOMA insulin resistance indices [119], most HFD-fed BXD animals (108 of 180) became glucose intolerant / insulin resistant over the six month period of dietary challenge. Conversely, only 18% of CD animals (29 of 165) have poor insulin sensitivity. Strikingly, the changes in 2-AA levels go in tandem with the HOMA index: animals with worse insulin sensitivity have lower 2-AA, and animals with good insulin sensitivity have equivalent levels of 2-AA, with no effects linked directly to diet (Fig. 4:5E). With this in mind, we further examined 2-AA in two independent population studies. In the Hybrid Mouse Diversity Panel, a recent profile of liver metabolomes in 96 cohorts in the fasted state contained 2-AA measurements [91]. In this diverse population, which includes ~30 BXD strains (with only one strain overlapping with those in our study), 2-AA is diminished in cohorts with higher fasted glucose (Fig. 4:5F), and the two are again negatively correlated (not shown). Likewise, we observed the same trend in a human population study (CoLaus [331]), in which we recently analyzed the spot urine samples of 835 middle-aged or elderly individuals [85]. In this cohort, diabetic patients also displayed a marked decrease in 2-AA levels (Fig. 4:5F). Together, these findings indicate that *Dhtkd1* plays a major role in the regulation of 2-AA in both mice and humans, and that the modulation of this pathway may influence the development of diabetes.

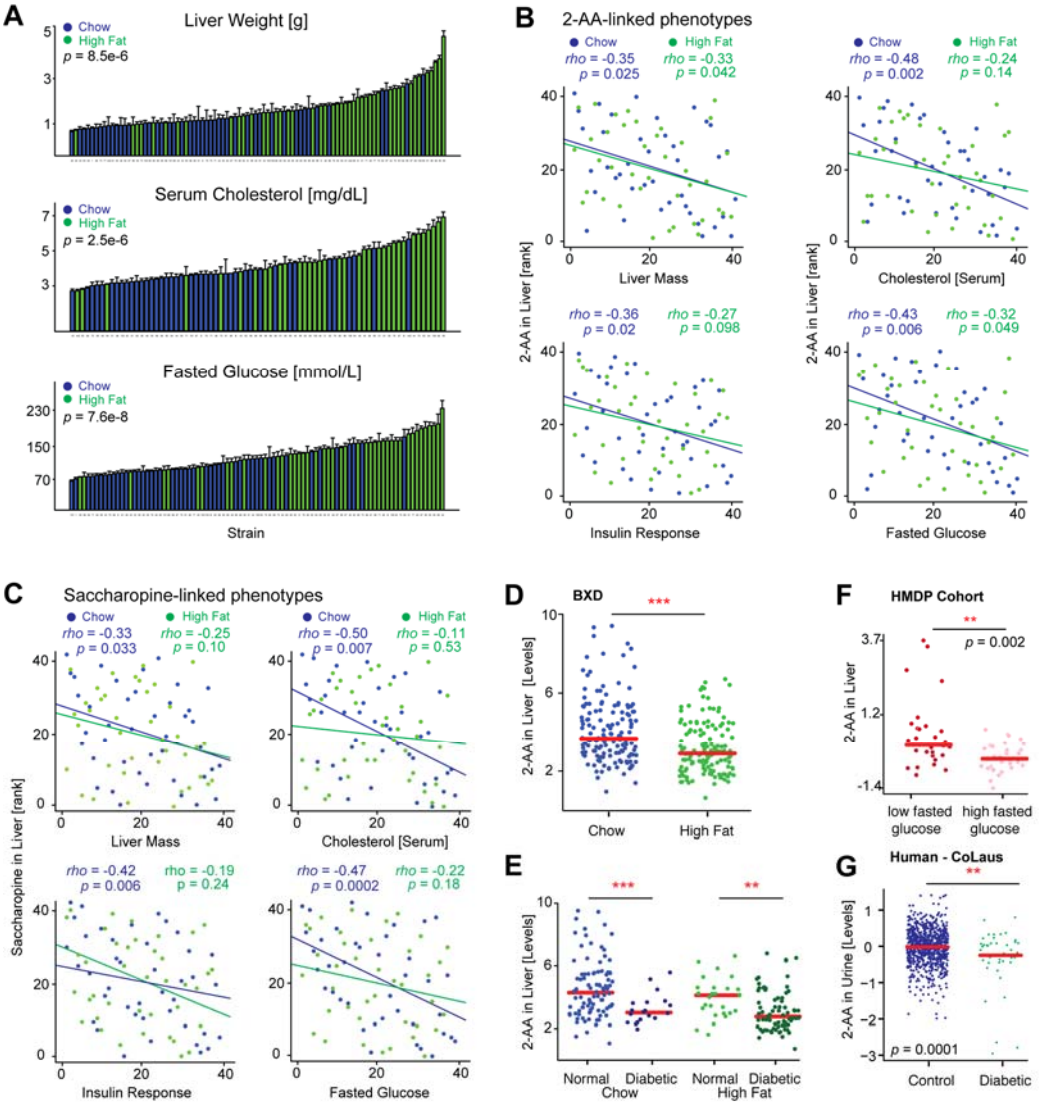


Figure 4:5 **Physiological Consequences of DHTKD1 Variants**

(A) Liver size, serum cholesterol and fasted glucose levels increase after HFD across the BXDs. **(B)** In CD livers, 2-AA is associated negatively with liver mass, serum cholesterol, insulin, and glucose. In HFD livers, 2-AA is associated negatively with liver mass and fasting glucose. **(C)** Liver levels of saccharopine, a metabolite upstream of 2-AA, are also associated negatively with the same phenotypes, although only in CD. **(D)** 2-AA levels in the liver are significantly decreased in HFD-fed BXD cohorts. **(E)** However, when correcting for diabetes status (HOMA-IR > 10 as diabetic, or HOMA-IR < 5 as healthy), there was no difference between CD and HFD—instead only between diabetic and non-diabetic cohorts. **(F)** Using publicly available data from a recent metabolomic profiling of livers from the HMDP [91], we can observe that the inverse relationship between 2-AA and fasted glucose is highly consistent in mouse populations. p = 0.01 if the four high non-outliers in the low group are suppressed. **(G)** Also in a human population study with urine metabolomics, diabetic patients had markedly lower levels of 2-AA.

4.1.6 The Mitochondrial Unfolded Protein Response (UPR^{mt})

The UPR^{mt} is a mitochondrial stress response pathway that is activated by proteostatic stress, such as by accumulation of unassembled or unfolded proteins in the mitochondria [332], by the presence of an imbalance between mitochondrial and nuclear encoded proteins [82], or by electron transport chain defects [200, 333]. The activation of UPR^{mt}

in turn leads to the transcription/translation of nuclear-encoded protective genes such as mitochondrial chaperones and proteases to reestablish mitochondrial proteostasis (reviewed in [134, 334, 335]). The bulk of research on UPR^{mt} has taken place using *C. elegans* and mammalian cell lines, thus little is known about when or how UPR^{mt} occurs *in vivo* in mammals. Furthermore, as the UPR^{mt} is a stress response tied to maintaining mitochondrial protein balance, we hypothesized that its protein correlation networks may be different than those generally examined at the transcriptional level.

In the worm, two "classical" approaches have been typically used to induce UPR^{mt}: loss-of-function of *cco-1*, a nuclear encoded component of the electron transport chain [200], or loss of function of *spg-7*, a mitochondrial protein quality-control protease [336]. We confirmed that the knockdown of either gene by RNAi triggers the UPR^{mt} response in *C. elegans*, by strong induction of the mitochondrial chaperone *hsp-6* and of the proteases *lonp-1* and *clpp-1* (Fig. 4:6A). Moreover, we linked this UPR^{mt} activation to specific phenotypes—a major reduction in size and mobility, as well as a decrease in oxygen consumption—which are consequences of mitochondrial stress (Fig. 4:6B). However, whether this coordinated regulation of UPR^{mt} genes is conserved in mammals *in vivo*, has not been previously shown.

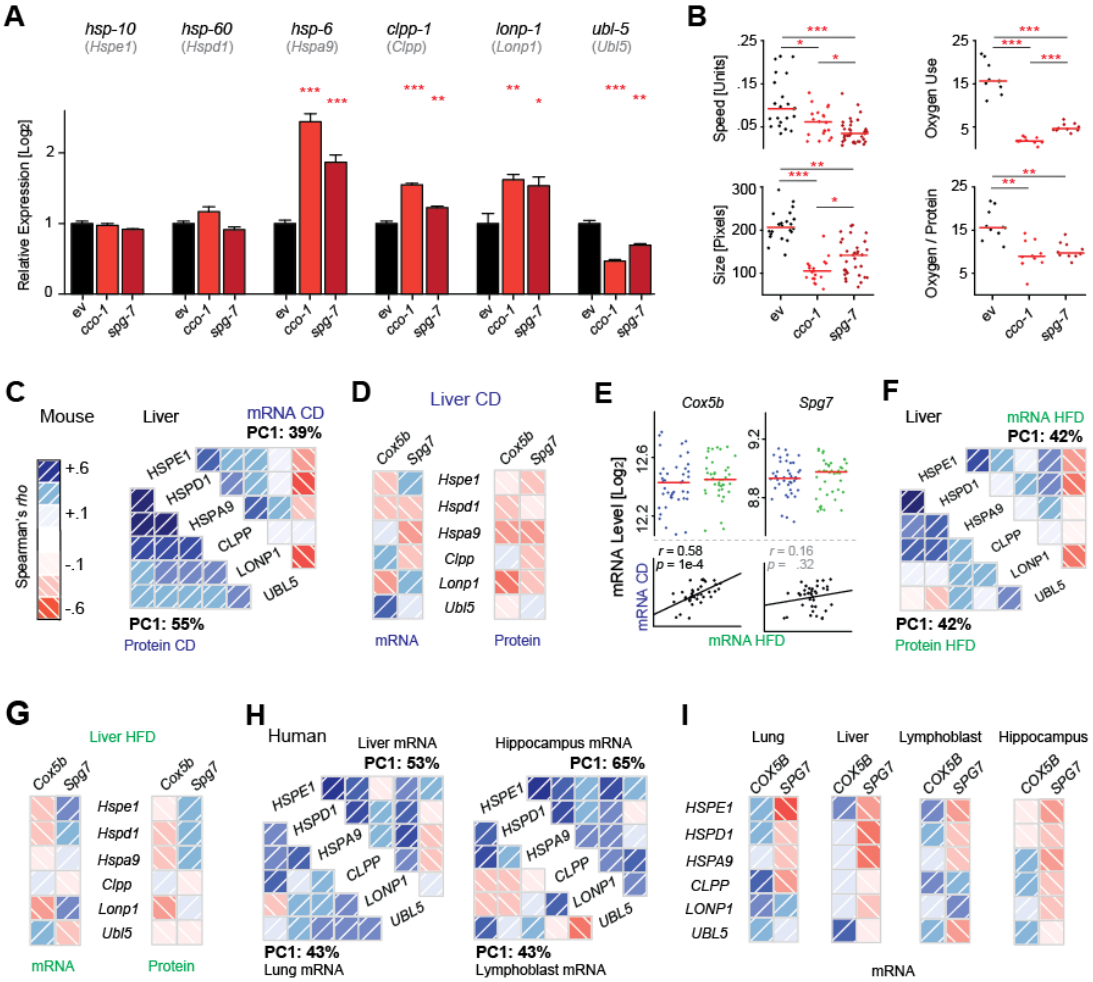


Figure 4:6 **The Mitochondrial Unfolded Protein Response**

(A) UPR^{mt} induction in *C. elegans* triggered by interference with ETC (RNAi of *cco-1*) or mitochondrial proteostasis (RNAi for *spg-7*). These triggers result in upregulation of UPR^{mt} effectors *hsp-6*, *clpp-1*, and *lonp-1* and a reduction in *ubl-5*. The orthologous mouse genes are indicated below the respective *C. elegans* gene symbol. Error bars represent mean + SEM. **(B)** UPR^{mt} induction in *C. elegans* decreases movement, size, and oxygen consumption. **(C)** UPR^{mt} genes and proteins form a network of coordinately expressed mRNAs and proteins *in vivo* in mice, which is stronger at the protein than at the mRNA level. **(D)** *Cox5b* and *Spg7* (orthologs of *C. elegans cco-1* and *spg-7*), are generally negatively associated with the levels of all UPR^{mt} genes in CD

cohorts, particularly at the protein level, in line with observations in the worm. **(E)** While the levels of *Cox5b* and *Spg7* are not affected by diet, expression is consistent by strain across the two diets only for *Cox5b*. **(F)** The UPR^{mt} network in HFD livers is similar to that observed in CD, but somewhat weaker. *Ubl5* remains a striking negative correlate at the mRNA level. **(G)** In HFD, *Cox5b* remains a negative correlate of UPRmt transcripts and proteins, while *Spg7* does not. (H) The features of the UPR^{mt} network are also conserved in 427 human liver biopsies [337], 405 lung biopsies [338], 180 lymphoblast lines [339], and 43 hippocampi [340]. **(I)** In humans, *SPG7* is a consistent negative correlate of the UPR^{mt} transcripts.

In the BXDs, we investigated the expression of six members of the UPR^{mt} pathway, which are well-conserved from *C. elegans*: mitochondrial chaperones (*Hspd1, Hspe1, Hspa9*), proteases (*Clpp, Lonp1*), and a transcriptional regulator involved in UPR^{mt} (*Ubl5*). These UPR^{mt} genes are also coordinately regulated at both mRNA and protein level in the BXDs, but with much stronger connections among proteins (Fig. 4:6C). Moreover, the UPR^{mt} network correlates negatively with *Cox5b* and *Spg7* (mouse orthologs of worm *cco-1* and *spg-7*, respectively), indicating that low abundance of these genes amplifies UPR^{mt} in mammals as in *C. elegans* (Fig. 4:6D). The network is also influenced in part by diet. While *Cox5b* expression patterns are similar between CD and HFD, *Spg7* covariation is disjointed between the dietary cohorts (Fig. 4:6E). This may explain why, despite a similar overall UPR^{mt} response in both diets (Fig. 4:6D & F), *Spg7* trends positively in HFD cohorts, while *Cox5b* remains consistent (Fig. 4:6G). Using four large transcriptional studies of human tissue biopsies, we observe similar transcriptional links, particularly including a strong network between all the UPR^{mt} genes (Fig. 4:6H). In humans, *SPG7* is a consistent negative correlate of this network, in contrast to *COX5B,* which generally has positive covariation with the UPR^{mt} response (Fig. 4:6I). Thus, while many of the overall regulators of UPR^{mt} remain coregulated across species—worm, mouse and human—particular nuances of its activation pathways appear variable dependent on species, environment, tissue, and likely other factors.

4.1.7 Discussion

Due to major differences in transcript and protein regulation, it has become increasingly clear that systems proteomics is essential for the analysis of complex systems such as metabolism [121, 308]. Traditionally this was attempted by semi-quantitative immunoassays such as ELISA or Western blotting, yet these techniques allow only a handful of proteins to be measured in parallel, and are moreover limited by the scarcity of quantitative immunoassays and the dubious quality of many antibodies [341]. Recent shotgun proteomics experiments on diverse populations have provided fundamental data on protein variance in populations of yeast [308, 309], mammals [305], and humans [133]. However, the inability to measure target proteins has limited the application of this approach in the study of disease pathways. New developments in mass spectrometry have led to SRM, a targeted quantitative systems proteomics technique in which choice proteins can be quantified based on *a priori* information. This technique has been recently applied in a moderate-sized yeast population study [316], but until now has not been applied to study genetic regulation in multicellular species nor for the study of a choice pathway.

In this study, we quantified 192 metabolism genes at the transcript and protein level in livers from 77 cohorts of the BXD GRP under two different dietary conditions. Along with sequence variants, basic metabolomics, and phenotype data, this combined multilayered population dataset enabled us to accurately estimate abundance changes in gene products due to genotype, diet, and gene-by-environment interactions (GXE). By complementing these layers, we were able to tentatively link dozens of pQTLs with phenotypes. Of the ~50 of significant pQTLs identified, we established links between four genes leading to phenotypic consequence: two novel, two confirmatory. For these genes, *Nnt*, *Car3*, *Dhtkd1*, and *Bckdhb*, coding differences between B6 and D2 alleles were linked to robust differences at both the transcript and protein level, leading to shared *cis*-regulatory eQTLs and pQTLs, and then to phenotypic consequences. While *Nnt* has been previously attributed as causal to differences in insulin secretion in the BXDs [320], the other pQTLs are novel, including for *Car3*, which was previously linked to adiposity only in knockout mice [321]. For *Bckdhb*, strains carrying the B6 allele have a ~50% reduction in enzyme levels which leads to a buildup of BCAA levels and decrease in alanine levels, as in the intermittent form of MSUD. In humans, MSUD is a rare inborn error of metabolism, and while BXDs do not suffer any overt physiological consequences of BCKDHB deficiency in the basal state, certain dietary stresses may magnify the symptoms and make the BXDs appropriate to model MSUD.

For *Dhtkd1*, the BXD strains map to highly significant and overlapping eQTLs and pQTLs with an approximately bimodal distribution with 2-fold variance between strains with the B6 (low) or the D2 allele (high) of the gene, and with

no effect from diet. Furthermore, the metabolite 2-AA maps as a significant mQTL to the *Dhtkd1* locus in both diets, despite a marked decrease of 2-AA levels in HFD cohorts. Interestingly the decrease in 2-AA is not directly due to the HFD, but rather due to the increased penetrance of insulin resistance after HFD. Moreover, prior works show that external administration of 2-AA can improve insulin sensitivity and glucose response in mice [92], indicating that variants in *Dhtkd1* and 2-AA levels may be driving the pathogenesis of diabetes, rather than being mere passive biomarkers of the disease. Strikingly, we replicated this association between 2-AA and high glucose levels in a recent and completely independent metabolomic dataset collected in the Hybrid Mouse Diversity Panel [91]. Furthermore and clinically relevant, we demonstrated in a human population-based sample, CoLaus, that diabetic patients likewise have lower urinary levels of 2-AA [85]. Thus, 2-AA levels are consistently linked to diabetes status in both mice and humans, and in the BXDs we can unequivocally attribute a causal part of this variance to variants in *Dhtkd1*. The BXD strains were necessary to bring the hypothesis to power a human study, and also confirm that the mouse link has relevance for human disease. Together, these findings further validate the BXD mouse population as a model for human metabolic diseases [16], and they indicate a critical need for multi-layered measurements to effectively analyze complex systems, particularly within the context of the study of GXE.

It is worth stressing that novel regulatory mechanisms can be found either through QTL analysis and their equivalent from GWAS, SNP analysis, or through network analyses, which are a complementary and powerful approach to dissect complex traits. The network approach is particularly viable when backed by high-depth multilayered datasets such as illustrated by our example on UPR^{mt}. UPR^{mt} is a reparative pathway activated by mitochondrial proteotoxic stress that has been primarily studied in the *C. elegans* and in cultured cells, but little is known about whether it occurs *in vivo* in mammals. We examined six genes that are known to be major regulators of *C. elegans* UPR^{mt} and which are conserved in mammals. These six genes form a robust coexpression network in both diets at the transcriptional and proteomic levels, with the proteomic connections being stronger, befitting the role of UPR^{mt} as a sensor and regulator of protein stress. One observation that stood out in the analysis of the UPR^{mt}, was the striking "contradiction" between the *Ubl5* transcript and UBL5 protein correlations to the UPR^{mt} network. *Ubl5* is a transcriptional regulator known to induce UPR^{mt}, yet in both worms and mice, its transcript levels decrease when UPR^{mt} is activated. Conversely, the UBL5 protein is increased with UPR^{mt} activation in the BXDs, an observation also previously reported in *C. elegans* [342]. This discordance in protein/transcript regulation suggests the existence of posttranscriptional mechanisms or a negative feedback loop, which could not be detected at the transcript or protein level alone. While there remains a great deal of this pathway left to be explored, it is clear that accurate, systems-scale protein measurements are essential to effectively model complex protein response networks like UPR^{mt}.

Today, there is an unprecedented capacity for accurate measurement quantification of nearly all aspects of biology. Surprisingly perhaps, proteins remain one of the most difficult aspects of biology to precisely measure at a systems scale, though it has been long clear that transcripts only serve as a weak proxy for protein levels [307]. Our data show that the integration of systems proteomics datasets with different layered systems measurements provides unprecedented insights into the mechanistic regulation of complex systems, such as exemplified by metabolism, that can eventually lead to the improved diagnosis and treatment of metabolic disease.

4.1.8 Methods

BXD strains were sourced from the University of Tennessee Health Science Center (Memphis, TN, USA) and bred at the EPFL animal facility for 2+ generations prior to incorporation into the study. We examined 40 strains of the BXD population—40 on CD, 37 on HFD—with ~10 male animals from each strain separated into two groups of ~five for each diet (3 strains on HFD were lost prior to tissue collection), with a total of 183 CD and 168 HFD animals. Strains were entered staggered one or two weeks apart in random order. Most strains entered with both dietary cohorts at the same time, with the exception of BXD50, 68, 69, 71, 84, 85, 89, 95, 96, 101, where CD cohorts entered before HFD cohorts. All cohorts consisted of littermates. HFD feeding started at 8 weeks of age. Cohorts were communally housed by strain and diet from 3 until 23 weeks of age, and were then individually housed until sacrifice at 29 weeks of age. Chow Diet (CD) is Harlan 2018 (6% kCal/fat, 20% kCal/protein, 74% kCal/carbohydrates), High Fat Diet (HFD) is Harlan 06414 (60% kCal/fat, 20% kCal/protein, 20% kCal/carbohydrates). All mice were housed under 12 h light, 12 h dark with *ad libitum* access to food and water at all times, except prior to sacrifice, when animals were fasted overnight. Body weight was measured weekly from 8 weeks of age until sacrifice. Sacrifices started at 9:00 am until 10:30 am, with

isoflurane anesthesia followed by a complete blood draw (~1 mL) from the vena cava, followed by perfusion with phosphate-buffered saline. Blood was placed into lithium-heparin coated tubes, shaken, and stored on ice, followed immediately by collection of the liver. Gall bladders were removed and the livers were cut into small pieces before freezing in liquid nitrogen until preparation into mRNA, protein, or metabolite samples. Blood was centrifuged at 4,500 rpm for 10 minutes at 4°C, and supernatant (serum) was placed in new tubes and flash frozen in liquid nitrogen. Liver and serum were then stored at -80°C until analysis. All research was approved by the Swiss cantonal veterinary authorities of Vaud under license 2257.

Serum analysis was performed using ultra performance liquid chromatrography (UPLC) electrospray tandem mass spectrometry (MS). Plasma was deproteinized using acetonitrile after the addition of internal standards. Amino acids were separated using reverse phase UPLC and heptafluorobutyric acid as an ion pairing agent, with detection on a Waters Quattro Premier XE.

For liver analyses, three ~100 mg pieces were taken from cold storage for each individual, then weighed and sorted for mRNA, protein, and metabolite measurements.

For mRNA, the ~100 mg pieces were suspended in TRIzol (Invitrogen) and homogenized with stainless steel beads using a TissueLyser II (Qiagen) at 30 Hz for 2 min, followed by a standard phase separation extraction using chloroform and precipitated by isopropanol. mRNA concentration was measured for all samples and then pooled equally for each cohort (i.e. 5 biological replicates for BXD103 CD became 1 mixed pool of BXD103 CD). Pooled RNA was cleaned up using RNEasy (Qiagen). The mRNA of all cohorts was prepared in direct series over a ~2 week period. 76 of the 77 cohorts had high quality mRNA based on RNA integrity numbers ≥ 8.0, indicating they are suitable for amplification and subsequent microarray analysis. Arrays were run for all cohorts in direct series over a three week period using the Affymetrix MouseGene 1.0 ST array at the Molecular Resource Center of Excellence in The University of Tennessee Health Science Center. Data were normalized using the RMA method [301], then used analyses in GeneNetwork and R. Three genes identified at the protein level were not measured on the array (*Srp54*, *mt-ND5*, and *mt-COX3*).

For liver protein, the ~100 mg liver samples were homogenized with 4 mL radioimmune precipitation assay-modified buffer (1% Nonidet P-40, 0.1% sodium deoxycholate, 150 mM NaCl, 1 mM EDTA, 50 mM Tris, pH 7.5, protease inhibitors EDTA-free, 10 mM NaF, 10 mM sodium pyrophosphate, 5 mM 2-glycerophosphate) and a glass-glass tight dounce homogenizer (Wheaton Science Products) at 4°C. After the homogenates were centrifuged (20,000 g at 4 °C for 15 min), the supernatant was collected and kept at 4°C. The pellets were resuspended with UREA-Tris buffer (50 mM Tris, pH 8.1, 75 mM NaCl, 8 M urea, EDTA-free protease inhibitors, 10 mM NaF, 10 mM sodium pyrophosphate, 5 mM 2-glycerophosphate) and sonicated for 5 min, then centrifuged at 20,000 g for 15 min at 4°C. The supernatant from the two steps were combined and protein concentration were determined with BCA Protein Assay (Thermo Fisher Scientific). Protein was prepared in two separate batches about six months apart (Batch 1: C57BL/6J, DBA/2J, BXD43, 44, 45, 48, 49, 51, 55, 61, 62, 64, 66, 68, 69, 70, 71, 73, 75, 80, 81, 83, 84, 85, 87, 90, 99, 100, 103; Batch 2: 50, 56, 60, 79, 89, 92, 95, 96, 97, 98, 101). Each batch was run separately on SRM, described in detail in the following section.

For liver metabolites, the ~100 mg liver pieces were homogenized in 1mL 70% Ethanol at -20°C. Metabolites were extracted by adding 7mL 70% Ethanol at 75°C for 2 min. Extracts were centrifuged for 10 minutes at 4,000 rpm at 4°C. Clean metabolites extracts were dried in a vacuum centrifuge and resuspended in double-distilled H_2O with volume according to the weight of the extracted liver piece. Quantification of metabolites was performed on an Agilent 6550 QTOF instrument by flow injection analysis time-of-flight mass spectrometry [343]. All samples were injected in duplicates. Ions were annotated based on their accurate mass and the Human Metabolome Database reference list [152] allowing a tolerance of 0.001 Da. Unknown ions and those annotated as adducts were discarded. The m/z ratios are reported in the online data. In particular, deprotonated α-ketoglutarate, 2-aminoadipic acid, saccharopine, and correspond to 144.0302, 160.0614 and 275.1249, respectively.

For all data, QTL mapping was performed using GeneNetwork [181].

For proteins previously observed in shotgun proteomics experiment, 3 to 4 proteotypic peptides most frequently observed by mass spectrometry were chosen based on the information present in the publicly accessible proteomic data repository PeptideAtlas [344] (**www.peptideatlas.org**). Peptide identifications deriving from isotope-coded affinity tag experiments were not considered. For proteins with less than three proteotypic peptides available from PeptideAtlas, additional proteotypic peptides amenable for mass spectrometry analysis were derived by bioinformatic prediction using the analysis software PeptideSieve [345] (**tools.proteomecenter.org**) and filtered by peptide length

and hydrophobicity. All peptides were synthesized without tags and lyophilized in a 96-well plate with ~50 nmol of peptide material per well (JPT Peptide Technologies). Peptides were resuspended in 20% acetonitrile, 1% formic acid, with 180 µL per well, vortexed for 15 min and sonicated for 10 min in the 96-well plate. Peptides were combined with each 5 µl and evaporated on a vacuum centrifuge to dryness, resolubilized in 2% acetonitrile, 0.1% formic acid. Eleven iRT peptides from Biognosys [346] with elution times spanning the whole solvent gradient were spiked into each mixture to facilitate the correlation of relative retention times between LC-MS/MS runs.

For each proteotypic peptide, the two best SRM transitions for both the double- and the triple-charged state were calculated, corresponding to fragment ions of the y-series. Fragment ions with a mass-to-charge ratio (m/z) greater than the precursor ion m/z precursor + 20 were prioritized. This selection process was processed through TIQAM [347]. The selected transitions were used to detect the peptides by SRM for synthetic peptides and to trigger acquisition of the full fragment ion spectra of the peptides. MS/MS spectra were acquired in enhanced product ion mode for the two highest SRM transitions, using the following settings: dynamic fill time, Q1 resolution low, scan speed 4000 Da/s, m/z range 380–1400. Raw MS/MS .wiff data was converted to .mzXML format with the program mzWiff and searched against the target proteins database using Mascot (**www.matrixscience.com**). A decoy database was generated by reversing the amino acid sequence and appended to the target database. Precursor mass tolerance was set at 0.8 Da. Data were searched with full tryptic cleavage except for K/R-proline. Cystein carbamidomethylation was set as the static modification and methionine oxidation was set as the variable modification. The search results were validated and filtered with *p*-value < 0.05 and expect value < 0.05. The resulting Mascot searching result were imported as the spectral library. For each peptide, the top 3 fragment ions resulting in the highest signals were selected from the QQQ MS/MS spectra. The corresponding SRM transitions associated with the ($^{13}C_6$, $^{15}N_4$)-Arginine, ($^{13}C_6$, $^{15}N_2$)-Lysine peptide analog of each peptide were calculated and combined with the endogenous transitions to generate the final SRM assay.

The final SRM assays were used to detect and quantify the proteins in the 77 samples of mouse liver tissue lysate using time-scheduled SRM/MRM acquisition (retention time window, 240 s; target scan time, 2.5 seconds, up to a maximum of 550 transitions per run) [348]. Heavy proteome using ($^{13}C_6$, $^{15}N_4$)-Arginine, ($^{13}C_6$, $^{15}N_2$)-Lysine cultured Hepa1-6 cells line was generated as previously described [349]. For peptides not detected in Hepa1-6 proteome, ($^{13}C_6$, $^{15}N_4$)-Arginine, ($^{13}C_6$, $^{15}N_2$)-Lysine peptides were synthesized and spiked into each sample with same amount prior to mass spectrometry measurements and mixed with equal amounts with the 77 liver homogenate samples respectively prior to protein digestion. Retention times were predicted using equation calculated from iRT peptides. β-gal runs (β-gal, 0.1% formic acid, injected) were performed prior to each 2 SRM measurements. 309 peptides were successfully quantified.

Wild-type Bristol N2 *C. elegans* provided by the *Caenorhabditis* Genetics Center (University of Minnesota) were cultured at 20°C and sustained on the OP50 *E. coli* strain. Bacterial feeding RNAi experiments were carried out as described [350]. *cco-1* (F26E4.9) and *spg-7* (Y47G6A.10) clones were purchased from GeneService and sequenced.

For qPCR analysis, five biological replicates for each condition were prepared, consisting of ~600 worms per sample in M9 minimal liquid medium. Before mRNA preparation, samples were washed twice with 5 mL M9 to eliminate residual bacteria. Total RNA was prepared using TRIzol (Invitrogen) according to the manufacturer's instructions. RNA was treated with DNase, and 1 µg of RNA was used for reverse transcription (RT). 15X diluted cDNA was used for RT-quantitative PCR (RT-qPCR) reactions. The RT-qPCR reactions were performed using the LightCycler 480 System (Roche Applied Science) and a qPCR Supermix (QIAGEN) with the indicated primers. *act-1* was used as normalization control. Three technical replicates were used for each biological replicate.

The primers used for the *C. elegans* genes were as follows:

cco-1 (F26E4.9):	Fwd: GCTCGTCTTGCTGGAGATGATCGTT
	Rvs: GGTCGGCGTCGACTCCCTTG
spg-7 (Y47G6A.10):	Fwd: AAGTATGCAGGACAAACGTGC
	Rvs: TGAGGTTTGGGATTTCGCGT
hsp-6 (C37H5.8):	Fwd: AACCACCGTCAACAACGCCG
	Rvs: AGCGATGATCTTATCTCCAGCGTCC
hsp-60 (Y22D7AL.5):	Fwd: TTCTCGCCAGAGCCATCGCC
	Rvs: TCTCTTCGGGGGTGGTGACCTTC

hsp-10(Y22D7AL.10): Fwd: GGGAAAAGTCCTTGAAGCCAC

Rvs: CTCCGAGAAGATCAGACTCGC

clpp-1 (ZK970.2): Fwd: TGCACAGGGAACCTGCTCGG

Rvs: TTGAGAGCTTCGTGGGCGCT

lonp-1 (C34B2.6): Fwd: CGATGATGGCCATTGTGCAG

Rvs: CGCTTTGAAACATCAATTTCATCCA

ubl-5 (F46F11.4): Fwd: ACGAATCAAGTGCAATCCATCAG

Rvs: GCTCGAAATTGAATCCCTCGTG

act-1 (T04C12.6): Fwd: GCTGGACGTGATCTTACTGATTACC

Rvs: GTAGCAGAGCTTCTCCTTGATGTC

For *C. elegans* phenotyping, oxygen consumption was measured using the Seahorse XF96 equipment (Seahorse Bioscience Inc.) as described previously [351]. Typically, 100 worms per condition were recovered from plates with Nematode Growth Medium (NGM), then washed three times in 2 mL of M9 liquid medium to eliminate residual bacteria, and resuspended again in 500 µL of M9. Worms were transferred in 96-well standard Seahorse plates (#100777-004) (10 worms per well) and oxygen consumption was measured 6 times. Respiration rates were normalized to the number of worms in each individual well.

Movement was recorded for 45 seconds at day 2 of adulthood using a Nikon DS-L2 / DS-Fi1 camera and controller setup, attached to a computerized Nikon bright field microscope. Five plates of worms, with 20 worms per plate, were measured in each condition. The movement of worms during this time was calculated by following the worm centroids using the same modified version of the freely-available for the Parallel Worm Tracker in MATLAB [180].

Proteins were prepared and measured in two separate batches: the first with 58 cohorts (29 of each diet), the second with 19 cohorts (11 CD, 8 HFD). Due to the presence of batch effects between the runs, the protein levels were converted to *Z*-scores based on the principle that the means and variation between the two groups should be roughly equal. This equality was proven by the identification and tracking of *cis*-pQTLs (particularly DHTKD1 and NNT), which were identified separately in both batches, but significance was much improved with the total set of 40. E.g. for CD-DHTKD1, the LRS was 31 for the batch of 29 strains, which increased to 43 when re-normalized and combined with the remaining 11 strains, indicating the combined data are improved. We measured 309 peptides to represent the 192 proteins, of which 94 proteins were measured by 1 peptide, 79 were measured by 2 peptides, and 19 were measured by 3 peptides. Of the 103 proteins with more than one measurement peptide, we were able to reduce 91 peptides, leaving us with 218 peptides measuring 192 proteins. Peptides were consolidated (averaged) if the peptides correlated ≥ 0.5 with one another. In other cases, one peptide was selected if (a) one peptide correlated ≥ 0.5 with its transcript and the other peptide(s) did not correlate at least nominally significantly (i.e. rho ≤ |0.31|), (b) if one peptide had a *cis*-pQTL, or (c) if a peptide had technical reproducibility < 0.95 it was removed. After this consolidation and removal, we were left with 218 peptides for all 192 genes, with 26 proteins retaining 2 independent peptide measurements. In such cases, it is not known if this independence is due to unidentified isoforms, unknown PTM sites, or other conflicts.

To further establish measurement reliability, we generated a heatmap quantification matrix (Fig. 4:1C, upper). This matrix displays protein abundance fold changes of condition (light) versus standard (heavy labeling). The peptide fold changes were computed from transition fold-changes using Skyline [352]. The association network (see Fig. 4:1D) was constructed from significantly nonzero ($p < 0.01$, multiple testing correction) and strong positive associations among proteins ($r[XY] > 0.6$). To compute the associations we used the Spearman rank correlation measure. The correlations were computed on peptide level and if multiple peptides were present per protein the median correlation was taken. In order to assess the stability and significance of the partial correlations obtained, we were resampling with replacement from the 77 cohorts and re-computing the associations. We also examined how robust the association measures are with respect to changes in scaling and normalization.

For a given target metabolite we selected its NMR shifts obtained from the HMDB database (www.hmdb.ca). Next, we excluded shifts with heights less than 20% of the maximal height and also those whose shifts were shared by more than 10 other metabolites in the database. The resulting shifts are approximately unique signatures of the target metabolite which were then further pruned to avoid counting highly correlated signals multiple times. Finally, the concentration of the target metabolite was approximated as the mean concentration of the filtered and pruned set of most representative NMR shifts. We applied (permutation-based) regression to estimate the effect of the derived metabolite concentration proxies on the three selected high level phenotypes: diabetes status, serum glucose, and serum choles-

terol levels. We have recently published tese data [85], which contain additional information and metabolite measurements.

All BXD data can be found on the **www.genenetwork.org** resource. To download the clinical phenotype data, change the type to “Phenotype” and enter **LISP2** to find all associated results. Select all and export to recover the data, or analyze it online. Microarray data and protein measurements can be found on the same resource; change type to “Liver Proteome” or “Liver mRNA” and navigate to the named EPFL datasets under CD or HFD (e.g. EPFL/LISP BXD CD Liver Affy Mouse Gene 1.0 ST (Apr13) RMA). To download these data, please click the “INFO” button at the right-hand side of the search page and download the raw dataset in the upper right box of the screen. All microarray data can be found on GEO as well, under the accession GSE60149.

4.1.9 Acknowledgments

We thank A. van Cruchten and W. Smit for the serum analysis of 2-AA. The CoLaus study received financial contributions from GlaxoSmithKline, the Faculty of Biology and Medicine of Lausanne, and the Swiss National Science Foundation (SNSF; 33CSCO-122661). The authors thank P. Vollenweider, G. Waeber, V. Mooser and D. Waterworth, Co-PIs of the CoLaus study. We also thank L. Gillet, H. Röst, and P. Picotti for insightful discussions, and R. Huttenhain, A. Maiolica, P. Kouvonen, T. Sajic, A. Leitner, and A. Bensimon for technical support. Special thanks to M. Bochud, Y. Barreau, M. Firmann, V. Mayor, A. Bastian, B. Ramic, M. Moranville, M. Baumer, M. Sagette, J. Ecoffey and S. Mermoud for data collection. YW was supported by the ERC (Proteomics v3.0; AdG-233226 to RA), and the LiverX program and EGW by a fellowship from the Fondation Romande pour la Recherche sur le Diabète. JA is the Nestlé Chair in Energy Metabolism. Research was supported by the EPFL, ETHZ, ERC (Sirtuins; AdG-231138 and Proteomics v3.0; AdG-233226), Velux Stiftung, LiverX and AgingX programs of the Swiss Initiative for Systems Biology (51RTP0-151019 and 2013/153), SNSF (31003A-140780, 31003A-143914, and CSRII3-136201), and the NIH (R01AG043930).

Chapter 5 Conclusion

In this thesis, I have described and implemented an approach using a complex, genetically diverse population of mice to study the effects of genetics, sex, diet, and exercise on broad metabolic phenotypes. Particularly, I have sought to identify and explain how genetic variance pertains to and drives our phenotypic traits, particularly those leading to metabolic diseases such as diabetes or osteoporosis. In studying the millions of DNA variants in the BXD mouse population, which occur spontaneously in nature just as in human populations, I was able to identify a few more pieces in the puzzle linking our genome to physiological outcomes. In chapter 2, the gene *alkaline phosphatase* (*Alpl*) was linked to major differences in the circulating levels of vitamin B_6 and differences in bone strength in mice—and very likely in humans as well. Similarly, the *aryl hydrocarbon receptor* (*Ahr*) was shown to be causal of differences in normal movement activity in a wide array of organisms: *C. elegans*, *Drosophila*, and mice, and again likely has similar effects in people. Next, in chapter 3, I used data from the same BXD population study to confirm and expand upon several hypotheses formed using gain or loss of function (i.e. genetically modified) organisms. First, using muscle expression data and related exercise phenotypes, we showed that the *nuclear receptor co-repressor 1* (*NCoR1*) is tied to mitochondrial function and muscle fitness across a broad population, also identified/confirmed one metabolic pathway by which PARP inhibitors can improve muscle fitness. Similar experiments using liver expression and liver-related phenotypes confirmed that in the BXDs, several different genes connect to form the *activator protein-1* (AP-1), and in turn can lead to opposite effects despite working on the same target genes. Here, as in the G/LOF cohorts, both diet (CD vs HFD) and genotype influence the development of fatty liver disease. In chapter 4, we explored data acquired through a recent development in mass spectrophotometry, called SRM, which allowed the accurate, high-throughput measurement of hundreds of proteins in a single sample. Using these data, I was able to identify the gene *dehydrogenase and transketolase domain containing 1* (*Dhtkd1*) as causal of major changes in 2-aminoadiate, which in turn is associated with resistance to diabetes, and is unaffected by our environmental perturbation, HFD. Further findings from this study include the observation that the BXD population has a mild form of Maple Syrup Urine Disease (MSUD), a rare genetic condition affecting amino acid metabolism in certain populations, and that the mitochondrial unfolded protein response (UPR^{mt}) is relatively well conserved from *C. elegans* to mice. The finding on 2-aminoadipate proposes a new potential route to treating diabetes (or preventing/slowing its development) in humans, while the finding on MSUD provides a diverse mouse population that can be used to test drugs or treatments for eventual human trials. The research on UPR^{mt} is fundamental biological research, and it provides provides critical further evidence for this recently-discovered pathway which is now thought to be one of the major regulators of overall metabolic well-being. In time and with further research, this too is poised to have profound implications on medicine.

The ability to perform these studies was all made possible thanks to technologies and techniques developed over the past decades, from the BXD population itself, to microarray technology and mass spectrometry. A reductive study (such as those examining the effects a single gene) can be successful by measuring a few transcripts and proteins by hand, while the systems approach relies much more on an ability for *all* transcripts and *all* proteins can be measured. In the past 10–15 years, technological advances, particularly in sequencing, have greatly aided the systems approach, yet there is still a long way to go, particularly regarding quantitative proteomics (as discussed in chapter 4). Even state-of-the-art protein quantification systems "only" allow the simultaneous quantification of ~2000–3000 proteins, or roughly 10–15% of the proteome. By taking advantage of these new developments, we were able to apply the BXD population to re-examine an old question but with new eyes—how do genotype, sex, and environment interact to influence metabolism, particularly complex metabolic diseases such as obesity and diabetes? Indeed, we were able to identify novel factors leading to changes in bone density, in liver damage, in muscle function, in glucose response, and in amino acid metabolism. These findings can be applied to the rational development of treatments as well: perhaps an *NCoR1* inhibitor can restore lost muscle function in the elderly, or drugs targeting AP-1 activity can be used to restore glucose tolerance in diabetic patients. In fact, this hypothesis is put to the test in chapter 3.2, where we examine the effects of PARP inhibitors on muscle function in mice. The findings in this thesis are largely directed related towards understanding the fundamental interactions of genetic and environmental influences, yet as shown they also provide new starting points

for the development of novel therapeutics, and provide further evidence for the removal of negative environmental stressors (e.g. high fat diet) and the addition of beneficial environmental influences (e.g. exercise).

It is my hope that, in the coming decades, the tremendous effort this thesis represents will seem as basic to future geneticists as we consider much research from the 1980s today. Major pieces of work 30 years ago, such as on sequencing and identifying genes, required months of effort originally but can now be done in hours. Taken alone, each of the findings in my thesis represents a small addition to the greater knowledge of genetic regulation of metabolic phenotypes, yet these painstakingly-validated relationships will in the future provide the basis for predictive gene–phenotype relationships, paving the way for real, effective personalized medicine. Even within the four year scope of this thesis, new technologies arose providing increased capabilities to understand molecular action, and the cost of such studies came down. Just as the pace of genetic discovery has exponentially increased over the past decades, the future outlook indicates that this rate will only continue to increase. The first pieces of a puzzle are the most difficult, and although we do not yet know how many pieces our genetic puzzle contains, we now have the corner pieces thanks in large part to DNA sequencing, RNA sequencing, and other "full scale" omics technologies. Already, we can clinically identify many diseases based on molecular measurements in the blood—fasting glucose is indicative of diabetes, high LDL cholesterol is indicative of atherosclerosis, and so on. By identifying variation in mice and assessing the physiological impact of gene and environmental effectors, we can further identify which genetic factors are associated with disease by using the same biomarkers (e.g. fasting glucose levels in response to gene knockouts). Thus, we can uncover genetic variants influencing fasted glucose—and therefore diabetes—and then translate these findings to human population studies. Gradually, such experiments fit together to provide an accurate prediction of an individual's risk for any particular metabolic trait based on their DNA sequence and basic clinical measurements. Already we can see the first steps into personalized medicine thanks to with commercial entities such as 23andMe. While the confidence in most results from such platforms is quite low at the moment, the quality is increasingly rapidly. However, we must also realize that genetic tests on complex diseases will never be 100% accurate, as most of us—fortunately—can in part control our own destiny through environmental modifications. Likewise, we are unlikely to discover a single drug or lifestyle change which will cure the entire world of any complex disease, diabetes, cancer, or so forth, as the diversity of causes leading to the diseases means there is no single cure. However, with a tempered understanding of the promises of "personalized medicine", we are already able to make better and more accurate lifestyle choices, tailored to our own genome.

5.1 Applied Systems Genetics : Broad and Narrow

In chapters 2 and 4, we saw broad applications of systems genetics which then focused in on novel, individual vignettes. In chapter 3, we saw three applied uses of population study to supplement already-established pathways. In each case, we were able to identify novel regulators and mechanisms underlying general metabolic dysfunctions (e.g. bone density, muscle function, liver damage). In the past several years, it has become increasingly clear that a reductionist application of genetics can gain significant power by examining the same hypotheses in larger populations, demonstrated most prominently with the AP-1 example in chapter 3.3. This necessity is due to the fact that the most critical molecular regulatory mechanisms have multiple redundancies—if *GeneX* is inactivated by a random mutation, *GeneY*, or *GeneZ* will become upregulated, thereby preventing the most severe effects of *GeneX* inactivation. Alternatively, while *GeneA* may improve muscle regeneration when modified, it is exceptionally unlikely to act entirely by itself. Instead, such genes are part of a large network with a multitude of other factors, which combined to provide a robust system of checks-and-balances. This concept, i.e. that the activity of the genome is not simply the individual activity of all genes summed together, is known as epistasis and is almost certainly the most complicated and critical missing point behind studies of complex diseases such as cancer and diabetes. While reverse engineering such multi-partner constructs is difficult, it is not impossible: this is rwhere broad applications of systems genetics comes in to focus. Large population studies, particularly in humans and mice, have enjoyed notable success in the past few years towards the identification of new gene networks and key regulatory genes, often through QTL mapping and GWAS. However, our current predictive capacity for complex gene networks is too weak to fully examine identified mechanisms *in silico*, so at present the links must still be validated with the narrow focus, often through the use of genetically modified organisms or cells. Fortunately, technical developments for performing genomic modifications have improved at least as

much as the "omics" technologies have in the past decade, and increasingly, we are able to quickly generate models to validate epistatic interactions using a small number of genes in tandem in a single study. In the coming years, by training predictive models by using relatively simple examples of epistasis, we may then be able to trust later findings which bioinformatically identify large-scale gene interaction networks for a more comprehensive understanding of genetics. As of yet, it is infeasible to test such large epistatic experiments, but it is highly likely that complex diseases stem from the interactions of dozens of major regulatory genes.

Today, systems biology approaches sit at a point between "too simple to model reality" and "too complex to understand". A close reading of this thesis may have identified the following somewhat contradictory point: the genes which were found through the systems genetics approach are largely monogenically linked to phenotypes (*Alpl, Dhtkd1, Bckdhb,* and to a lesser extent, *Ahr* in chapters 2 and 4), while the three pathways studied using traditional reductionist approaches (chapter 3) were expanded into large polygenic networks. This is not because such complex pathways were not identified (see the glucose and oxygen respiration examples in chapter 2.1), but rather due monogenic connections being "lower hanging fruit", i.e. they are far more easily identified and validated. Likewise, expanding existing networks is much easier than identifying and validating a system from scratch—and therefore quicker to publish. Continuing work on validating complex networks identified in this population study, mainly the UPR^{mt} discussed at the end of chapter 4, will be the short-term focus of my research following this thesis. As such, it is worth noting that this thesis represents the BXD systems metabolism project starting from the ground up, but it does not go to the end of the project: we are closer to the *midpoint*.

5.2 Forward Thinking

The majority of the findings in this thesis were developed in animal models, whether mouse or *C. elegans*, yet fundamentally these data must be translated back to human medicine to find utility in the medical world and justify the project. This process is not a simple one-way exchange from animal models to human cohorts, as simply identifying the cause(s) of disease is far from developing a treatment or cure. For example, while hypophosphatasia is well-identified and categorized in humans prior to mouse, the translational (and novel) finding of this disease existing in the BXDs has strong medical applications. Firstly, this disease is not truly monogenic—although 50% or more of hypophosphatasia may be attributable to *Alpl*, many other factors are known to influence vitamin B_6 metabolism and bone strength, and thus could be contributing to (or alleviating) patients with variable symptoms of the disease. Moreover, while hypophosphatasia is well-understood from a biochemical standpoint, there is no cure nor any particularly good treatment, thus appropriate animal models are necessary to treat the human affliction. While today there already some gene therapy treatments for humans to cure the most debilitating diseases, it is almost certain that at least for the next few decades, drug treatments will continue to serve the vast majority of patients. To properly screen potential drug treatments, compounds must be first tested *in vitro* (i.e. in cell lines grown in petri dishes) for negative signatures of toxicity and for positive signatures of influencing target genes or metabolites. Next, molecular signatures related to the disease must be identified *in vivo* (i.e. in animal models) and monitored. Traditionally, only a single genotype of mouse is used for compound screening, but as we move into an era where genetic variants are increasingly known to affect drug response, it is essential that we examine drugs in diverse afflicted populations, as the BXDs are for (e.g.) hypophosphatasia. Thus, identifying (or generating) a genetically diverse model population for examining a human disease is a critical component on the pathway for developing a new treatment. For complex diseases, this means that a new test compound can be tested not only across a range of afflictions (e.g. severe diabetes to mild diabetes) but also in a population that has a diverse response to the drug itself, independent of the disease state.

In my thesis work, I focused on identifying genetic molecular signatures behind disease, with a particular eye for metabolic disorders, and have translated clear causative cases from human to mouse (i.e. *Alpl* levels in conjunction with vitamin B_6 levels can diagnose hypophosphatasia) and also suggested new biomarkers to check for translating from mouse to human (i.e. an association between *Dhtkd1* levels and 2-aminoadipate to diabetes susceptibility). However, it is often possible to lose perspective in such lengthy work: no single person will work on the entire process from target identification to drug development to the clinical testing of a new compound. Fortunately, I have had incredible opportunities at the EPFL to work with dozens of other scientists and share and receive data with people working on all aspects of medical development, and the extensive data generated will be continued to used for at least the next several

years. This network will lead to improved understanding of metabolism, and indeed may directly contribute to better diagnosis and prevention of diabetes and other metabolic diseases, both through my own direct involvement (e.g. if *Dhtkd1* or 2-aminoadipate do indeed end up being directly relevant to diagnosing or treating diabetes), or by simply providing a general platform of data which others may build upon (e.g. if someone uses the BXD data I generated to uncover or validate a novel gene regulating energy metabolism). While this thesis represents the end of my direction in this project, such broad systems genetics studies do not lead to one particular "conclusion," and instead I represent one part of the interim progress of a massive collaborative effort, and in fact is it certain that you will see some datasets generated for this thesis directly used in other graduate theses and scientific papers in the coming years.

List of Abbreviations

AHR	Aryl hydrocarbon receptor
ALPL	Alkaline phosphatase
AUC	Area under the curve (i.e. integral)
B6	C57BL/6J, a common strain of inbred laboratory mouse. The mother of all BXDs.
BXD	A type of recombinant inbred mouse cross (C57BL/6J x DBA/2J)
CD	Chow diet (i.e. low in fat)
ChIP	Chromatin Immunoprecipitation
D2	DBA/2J, the first inbred mouse strain. The father of all BXDs.
DHTKD1	Dehydrogenase E1 and transketolase domain containing 1
DNA	Deoxyribonucleic acid
ES	Embryonic stem (cells)
GEMM	Genetically modified mouse model
GM	Genetically modified
GOF	Gain-of-function
GRP	Genetic reference population
GTT	Glucose tolerance test
GWAS	Genome-wide association study
GXE	Gene-by-environment
HFD	High fat diet
ITT	Insulin tolerance test
LOD	Logarithm of odds
LOF	Loss-of-function
LRS	Likelihood ratio statistic
mRNA	Messenger ribonucleic acid
MS	Mass spectrometry
NAFLD	Non-alcoholic fatty liver disease
NCOR1	Nuclear receptor corepressor 1
PARP	Poly ADP ribose polymerase
PPARG	Peroxisome proliferator-activated receptor gamma
QTL	Quantitative trait locus
RER	Respiratory exchange ratio
RI	Recombinant inbred
SRM	Selected reaction monitoring
TCDD	2,3,7,8-tetrachlorodibenzo-*p*-dioxin
VCO_2	Volume of carbon dioxide
VO_2	Volume of oxygen

Curriculum Vitae

EVAN G. **WILLIAMS**

evangw@gmail.com or +41 (0)79 176 97 85

Rue École-de-Commerce 1, 1004 Lausanne, Vaud, Switzerland

English native speaker. French proficiency.

Education

Ph.D. in Bioengineering, expected 2014.

École Polytechnique Fédérale de Lausanne. Lausanne, Switerlerand.

Laboratory of Johan Auwerx.

Bachelor of Science in Bioengineering, 2009.

Rice University. Houston, TX, USA

High School Diploma, 2005.

White Station High School, Memphis, TN, USA

Research Experience

École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland

Graduate Student, September 2010–November 2014: Laboratory of Johan Auwerx. Ph.D. thesis work mainly regarding functional aspects of genetics (e.g. variation in metabolites, protein, or mRNA expression due to natural sequence variants). Significant time spent on metabolic phenotyping of animal populations and translational research between mice and *C. elegans* and *Drosophila*.

École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland

Research Assistant, September 2009–August 2010: Laboratory of Johan Auwerx. Researching biochemical pathways for *Sirt1* activation in mouse and cell models. Studied one mechanism involved in prolonged longevity known to be caused by caloric restriction, mostly at the cellular level.

Rice University, Houston, Texas

Technician, January 2008–June 2009: Laboratory of Tomasz Tkaczyk. Created optical lenses using spin coating, diamond turning, and molding techniques for use in fiberoptics, particularly for a confocal reflectance microscope for endoscopy.

Thesis Project, August 2008–May 2009: Award-winning team design on a NASA project to develop a portable urinalysis system to non-invasively monitor bone mineral density loss aboard the International Space Station or in developing countries.

University of California, Los Angeles, Los Angeles, California

Intern, May–August 2008: Laboratory of Jake Lusis. Worked with microarray data on four F2 mouse populations each with hundreds of individuals. Performed MSSQL and MySQL database management, genetic correlation computations, and analysis for candidate genes involved in atherosclerosis and other metabolic disorders.

Helmholtz Zentrum für Infektionsforschung, Braunschweig, Germany

Intern, May–August 2007: Laboratory of Klaus Schughart. Set up server hardware and software for a Linux-Apache-MySQL-Python system and served as the systems administrator. Performed error checking and analysis of previously acquired data.

University of Tennessee Health Science Center, Memphis, Tennessee

Consultant, June–August 2009: Laboratory of Lu Lu. Improved Python coding for a SNP analysis tool; also implemented a feature for analyzing indels and other sequence variants. Aggregated public SNP information for dozens of mouse strains.

Consultant, May–August 2006: Managed MySQL database for GeneNetwork. Wrote documentation and tutorial for much of the structure and its use.

Lab assistant, May–August 2005: Acquired and updated data for GeneNetwork. Refined Python code and MySQL organization in parts of the system.

Professional Skills

Laboratory Techniques

Bench: Standard wetlab biology (e.g. Western, cell culture, transfection, transduction, qPCR)

Animal: Extensive mouse experience (e.g. VO_{2max}, cold response, glucose test, breeding)
Some experience with *C. elegans* and *Drosophila* (general handling)

Analytical: Significant experience with microarray analysis, sequence variant analysis, functional genomics. GSEA (& other Broad Institute tools), GeneNetwork (QTL), Cytoscape.

General Software Skills

Software: *Expert*: OS X, Windows, UNIX (general command line). Photoshop, Illustrator, Office, Keynote.
Proficient: Prism, Cytoscape, Canvas, Final Cut

Database: *Proficient*: MySQL, FileMaker, Dropbox For Business (administrator)

Code: *Expert*: R. *Proficieint*: HTML, PHP, MatLab, Java.

Awards

2011 Fondation Romande pour la Recherche sur le Diabete Fellowship
2009 NASA ESMD Competition: First Place
2009 NASA Texas Space Grant Consortium, First Place Design Project
2008 NASA Texas Space Grant Consortium, First Place Design Project
2005 National Merit Scholar, Advanced Placement Scholar with Distinction
2004 Tennessee Governor's School for the Sciences

Publications

First Author

2014, September. *PLoS Genetics*. An evolutionarily conserved role for the aryl hydrocarbon receptor in the regulation of movement. **Williams EG**, Mouchiroud L, Frochaux M, Andreux PA, Deplancke B, Auwerx J.

2014, September. *Cell*. Protein quantitative trait locus (pQTL) analysis based on targeted proteomics in a mouse genetic reference population. Wu Y*, **Williams EG ***, Dubuis S, Mottis A, Jovaisaite V, Houten SM, Argmann CA, Faridi P, Wolski W, Kutalik Z, Zamboni N, Auwerx J**, Aebersold R**. *co-first and **co-

corresponding authors.

2012, September. *Cell*. Systems genetics of metabolism - the use of the BXD murine reference panel for multiscalar integration of traits. Andreux PA*, **Williams EG** *, Koutnikova H, Houtkooper RH, Champy MF, Henry H, Schoonjans K, Williams RW, Auwerx J. *co-first author

Supporting Author
2014/2015 (publication date pending)**.** *Diabetes*. Evidence for a direct effect of the NAD^+ precursor Acipimox on mitochondrial function in humans. Van der Weijer T*, Phielix E*, Bilet L, **Williams EG**, Laufs A, Livingstone R, Nowotny P, Sparks LM, Paglialunga S, Szendroedi J, Havekes B, Moullan N, Pirinen E, Hwang JH, Schrauwen-Hinderling VB, Hesselink MKC, Auwerx J, Roden M**, Schrauwen P**. * co-first and ** co-corresponding authors.

2014, June. *Cell Metabolism*. Pharmacological inhibition of poly(ADP-ribose) polymerases improves fitness and mitochondrial function in skeletal muscle. Pirinen E, Canto C, Jo YS, Morato L, Zhang H, Menzies K, **Williams EG**, Mouchiroud L, Moullan N, Hagberg C, Li W, Timmers S, Imhof R, Verbeek J, Pujol A, van Loon B, Viscomi C, Zeviani M, Schrauwen P, Sauve A, Schoonjans K, Auwerx J.

2014, January. *Cell Metabolism*. Regulation of steatohepatitis and PPARg signaling by distinct AP-1 dimers. Hasenfuss SC, Bakiri L, Thomsen MK, **Williams EG**, Auwerx J, Wagner EF.

2012, June. *PLoS One.* Murine gut microbiota is defined by host genetics and modulates variation of metabolic traits. McKnite AM, Perez-Munoz ME, Lu L, **Williams EG**, Brewer S, Andreux P, Bastiaansen J, Wang X, Kachman S, Auwerx J, Williams RW, Benson A, Peterson D, Ciobanu D.

2011, November. *Cell*. NCoR1 is a conserved physiological modulator of muscle mass and oxidative function. Yamamoto H, **Williams EG,** Mouchiroud L, Cantó C, Fan W, Downes M, Héligon C, Barish GD, Desvergne B, Evans RM, Schoonjans K, Auwerx J.

Conferences

2015, January. Paris. *Annual mouse genetics course* at Institut Pasteur. Guest lecturer.
2014, May. Ascona. *Systems Genetics and Evolution*. Attendee.
2014, January. Paris. *Annual mouse genetics course* at Institut Pasteur. Guest lecturer.
2013, September. London. *SysGenet*. Guest lecturer.
2013, June. Prague. *SysGenet*. Roundtable discussions.
2013, February. Luxembourg. *EASL*. Invited speaker and session co-chair.
2013, January. Paris. *Annual mouse genetics course* at Institut Pasteur. Guest lecturer.
2012, December. Bilbao. *SysGenet*. Invited speaker.
2012, November. Geneva. *FRRD*. Invited speaker to present project for fellowship award.
2012, June. Paris. *CTC*. Invited speaker.
2009, December. Braunschweig. *BXD World*. Poster presentation.

References

1. Walsh, J.J., *Makers of modern medicine*. 1907, New York,: Fordham university press. viii, 2 l., 13-362 p.
2. Morgan, T.H., *Sex Limited Inheritance in Drosophila.* Science, 1910. **32**(812): p. 120-2.
3. Cuenot, L., *Mendel's law and heredity of pigmentation in mice.* Comptes Rendus Hebdomadaires Des Seances De L Academie Des Sciences, 1902. **134**: p. 779-781.
4. Murray, W.S. and C.C. Little, *The Genetics of Mammary Tumor Incidence in Mice.* Genetics, 1935. **20**(5): p. 466-96.
5. Sturtevant, A.H., *The linear arrangement of six sex-linked factors in Drosophila, as shown by their mode of association.* Journal of Experimental Zoology, 1913. **14**(1): p. 43-59.
6. Ross, S.R., et al., *A fat-specific enhancer is the primary determinant of gene expression for adipocyte P2 in vivo.* Proc Natl Acad Sci USA, 1990. **87**: p. 9590-9594.
7. Fajas, L., et al., *E2Fs regulate adipocyte differentiation.* Dev Cell, 2002. **3**(1): p. 39-49.
8. Kanehisa, M., et al., *Data, information, knowledge and principle: back to metabolism in KEGG.* Nucleic Acids Research, 2014. **42**(D1): p. D199-D205.
9. Clark, A.G., *Limits to prediction of phenotypes from knowledge of genotypes.* Evolutionary Biology, Vol 32, 2000. **32**: p. 205-224.
10. Bogardus, C., et al., *Identification of susceptibility genes for complex metabolic diseases.* Ann N Y Acad Sci, 2002. **967**: p. 1-6.
11. MacArthur, D.G., et al., *A systematic survey of loss-of-function variants in human protein-coding genes.* Science, 2012. **335**(6070): p. 823-8.
12. Lawson, N.D. and S.A. Wolfe, *Forward and reverse genetic approaches for the analysis of vertebrate development in the zebrafish.* Dev Cell, 2011. **21**(1): p. 48-64.
13. Williams, G.C., *Pleiotropy, Natural-Selection, and the Evolution of Senescence.* Evolution, 1957. **11**(4): p. 398-411.
14. Margulies, M., et al., *Genome sequencing in microfabricated high-density picolitre reactors.* Nature, 2005. **437**(7057): p. 376-80.
15. Schena, M., et al., *Quantitative monitoring of gene expression patterns with a complementary DNA microarray.* Science, 1995. **270**(5235): p. 467-70.
16. Andreux, P.A., et al., *Systems genetics of metabolism: the use of the BXD murine reference panel for multiscalar integration of traits.* Cell, 2012. **150**(6): p. 1287-99.
17. Orozco, L.D., et al., *Unraveling Inflammatory Responses using Systems Genetics and Gene-Environment Interactions in Macrophages.* Cell, 2012. **151**(3): p. 658-670.
18. Wu, Y., et al., *Multilayered genetics and omics dissection of mitochondrial activity in a mouse reference population.* Cell, 2014. **158**(6).
19. Barrett, T., et al., *NCBI GEO: archive for functional genomics data sets--update.* Nucleic Acids Res, 2013. **41**(Database issue): p. D991-5.
20. Wang, H., et al., *One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering.* Cell, 2013. **153**(4): p. 910-8.
21. Thorndike, E.L., *Measurements of twins*. Archives of philosophy, psychology and scientific methods ed by J M Cattell and F J E Woodbridge,. 1905, New York,: The Science press. vi, 64 p.
22. Gordon, T. and W.B. Kannel, *Predisposition to atherosclerosis in the head, heart, and legs. The Framingham study.* JAMA, 1972. **221**(7): p. 661-6.
23. Jensen, A.R., *Estimation of the limits of heritability of traits by comparison of monozygotic and dizygotic twins.* Proc Natl Acad Sci U S A, 1967. **58**(1): p. 149-56.

24. Friedrich, B., et al., *Variance of the SGK1 gene is associated with insulin secretion in different European populations: results from the TUEF, EUGENE2, and METSIM studies.* PLoS One, 2008. **3**(11): p. e3506.
25. McDermott, D.H., et al., *CCL2 polymorphisms are associated with serum monocyte chemoattractant protein-1 levels and myocardial infarction in the Framingham Heart Study.* Circulation, 2005. **112**(8): p. 1113-20.
26. Deeb, S.S., et al., *A Pro12Ala substitution in PPARgamma2 associated with decreased receptor activity, lower body mass index and improved insulin sensitivity.* Nat Genet, 1998. **20**(3): p. 284-7.
27. Yeo, G.S., et al., *A frameshift mutation in MC4R associated with dominantly inherited human obesity.* Nat Genet, 1998. **20**(2): p. 111-2.
28. Peltonen, L. and V.A. McKusick, *Genomics and medicine. Dissecting human disease in the postgenomic era.* Science, 2001. **291**(5507): p. 1224-9.
29. Manolio, T.A., *Genomewide association studies and assessment of the risk of disease.* N Engl J Med, 2010. **363**(2): p. 166-76.
30. Waterston, R.H., et al., *Initial sequencing and comparative analysis of the mouse genome.* Nature, 2002. **420**(6915): p. 520-62.
31. Reiter, L.T., et al., *A systematic analysis of human disease-associated gene sequences in Drosophila melanogaster.* Genome Res, 2001. **11**(6): p. 1114-25.
32. Sonnhammer, E.L. and R. Durbin, *Analysis of protein domain families in Caenorhabditis elegans.* Genomics, 1997. **46**(2): p. 200-16.
33. Peterson, T.A., D. Park, and M.G. Kann, *A protein domain-centric approach for the comparative analysis of human and yeast phenotypically relevant mutations.* BMC Genomics, 2013. **14 Suppl 3**: p. S5.
34. Arabidopsis Genome, I., *Analysis of the genome sequence of the flowering plant Arabidopsis thaliana.* Nature, 2000. **408**(6814): p. 796-815.
35. Baumeister, R., *Cross-species studies for target validation.* Brief Funct Genomic Proteomic, 2002. **1**(1): p. 53-65.
36. Feany, M.B. and W.W. Bender, *A Drosophila model of Parkinson's disease.* Nature, 2000. **404**(6776): p. 394-8.
37. van Ham, T.J., et al., *C. elegans model identifies genetic modifiers of alpha-synuclein inclusion formation during aging.* PLoS Genet, 2008. **4**(3): p. e1000027.
38. Kimura, K.D., et al., *Daf-2, an insulin receptor-like gene that regulates longevity and diapause in Caenorhabditis elegans.* Science, 1997. **277**: p. 942-946.
39. Lee, R.Y., J. Hench, and G. Ruvkun, *Regulation of C. elegans DAF-16 and its human ortholog FKHRL1 by the daf-2 insulin-like signaling pathway.* Curr Biol, 2001. **11**(24): p. 1950-7.
40. Lin, K., et al., *Regulation of the Caenorhabditis elegans longevity protein DAF-16 by insulin/IGF-1 and germline signaling.* Nat Genet, 2001. **28**(2): p. 139-45.
41. Barriere, A. and M.A. Felix, *Natural variation and population genetics of Caenorhabditis elegans.* WormBook, 2005: p. 1-19.
42. Beck, J.A., et al., *Genealogies of mouse inbred strains.* Nat Genet, 2000. **24**(1): p. 23-5.
43. Bennett, B.J., et al., *A high-resolution association mapping panel for the dissection of complex traits in mice.* Genome Res, 2010. **20**(2): p. 281-90.
44. Rocha, J.L., et al., *A large-sample QTL study in mice: I. Growth.* Mammalian Genome, 2004. **15**(2): p. 83-99.
45. Bailey, D.W., *Recombinant-inbred strains. An aid to finding identity, linkage, and function of histocompatibility and other genes.* Transplantation, 1971. **11**(3): p. 325-7.
46. Threadgill, D.W. and G.A. Churchill, *Ten years of the Collaborative Cross.* Genetics, 2012. **190**(2): p. 291-4.
47. Burr, B., et al., *Gene mapping with recombinant inbreds in maize.* Genetics, 1988. **118**(3): p. 519-26.
48. Pravenec, M., et al., *An analysis of spontaneous hypertension in spontaneously hypertensive rats by means of new recombinant inbred strains.* J Hypertens, 1989. **7**(3): p. 217-21.

49. Taylor, B.A., H.J. Heiniger, and H. Meier, *Genetic analysis of resistance to cadmium-induced testicular damage in mice.* Proc Soc Exp Biol Med, 1973. **143**(3): p. 629-33.
50. Lister, C. and C. Dean, *Recombinant inbred lines for mapping RFLP and phenotypic markers in Arabidopsis thaliana.* Plant J, 1993. **4**: p. 745–750.
51. Smith, E.N. and L. Kruglyak, *Gene-environment interaction in yeast gene expression.* PLoS Biol, 2008. **6**(4): p. e83.
52. Rockman, M.V. and L. Kruglyak, *Recombinational landscape and population genomics of Caenorhabditis elegans.* PLoS Genet, 2009. **5**(3): p. e1000419.
53. King, E.G., S.J. Macdonald, and A.D. Long, *Properties and power of the Drosophila Synthetic Population Resource for the routine dissection of complex traits.* Genetics, 2012. **191**(3): p. 935-49.
54. Zou, F., et al., *Quantitative trait locus analysis using recombinant inbred intercrosses: Theoretical and empirical considerations.* Genetics, 2005. **170**(3): p. 1299-1311.
55. Nadeau, J.H., et al., *Analysing complex genetic traits with chromosome substitution strains.* Nat Genet, 2000. **24**(3): p. 221-5.
56. Halaas, J.L., et al., *Weight-reducing effects of the plasma protein encoded by the obese gene.* Science, 1995. **269**: p. 543-546.
57. Freeman, H.C., et al., *Deletion of nicotinamide nucleotide transhydrogenase: a new quantitive trait locus accounting for glucose intolerance in C57BL/6J mice.* Diabetes, 2006. **55**(7): p. 2153-6.
58. Muller, H.J., *Artificial Transmutation of the Gene.* Science, 1927. **66**(1699): p. 84-7.
59. Auerbach, C. and J.M. Robson, *Chemical production of mutations.* Nature, 1946. **157**: p. 302.
60. Wigler, M., et al., *Transfer of purified herpes virus thymidine kinase gene to cultured mouse cells.* Cell, 1977. **11**(1): p. 223-32.
61. Folger, K.R., et al., *Patterns of integration of DNA microinjected into cultured mammalian cells: evidence for homologous recombination between injected plasmid DNA molecules.* Mol Cell Biol, 1982. **2**(11): p. 1372-87.
62. Palmiter, R.D., et al., *Dramatic growth of mice that develop from eggs microinjected with metallothionein-growth hormone fusion genes.* Nature, 1982. **300**(5893): p. 611-5.
63. Orban, P.C., D. Chui, and J.D. Marth, *Tissue- and site-specific DNA recombination in transgenic mice.* Proc Natl Acad Sci USA, 1992. **89**: p. 6861-6865.
64. Gossen, M. and H. Bujard, *Tight control of gene expression in mammalian cells by tetracycline-responsive promoters.* Proc Natl Acad Sci U S A, 1992. **89**(12): p. 5547-51.
65. Metzger, D., et al., *Conditional site-specific recombination in mammalian cells using a ligand-dependent chimeric Cre recombinase.* Proc Natl Acad Sci U S A, 1995. **92**(15): p. 6991-5.
66. Fraser, A.G., et al., *Functional genomic analysis of C. elegans chromosome I by systematic RNA interference.* Nature, 2000. **408**(6810): p. 325-30.
67. Ryder, E., et al., *The DrosDel deletion collection: a Drosophila genomewide chromosomal deficiency resource.* Genetics, 2007. **177**(1): p. 615-29.
68. Alonso, J.M., et al., *Genome-wide insertional mutagenesis of Arabidopsis thaliana.* Science, 2003. **301**(5633): p. 653-7.
69. Winzeler, E.A., et al., *Functional characterization of the S-cerevisiae genome by gene deletion and parallel analysis.* Science, 1999. **285**(5429): p. 901-906.
70. Brand, A.H. and N. Perrimon, *Targeted gene expression as a means of altering cell fates and generating dominant phenotypes.* Development, 1993. **118**(2): p. 401-15.
71. Skarnes, W.C., et al., *A conditional knockout resource for the genome-wide study of mouse gene function.* Nature, 2011. **474**(7351): p. 337-42.
72. Auwerx, J., et al., *The European dimension for the mouse genome mutagenesis program.* Nat Genet, 2004. **36**(9): p. 925-7.
73. Austin, C.P., et al., *The knockout mouse project.* Nat Genet, 2004. **36**(9): p. 921-4.
74. Capecchi, M.R., *Gene targeting in mice: functional analysis of the mammalian genome for the twenty-first century.* Nat Rev Genet, 2005. **6**(6): p. 507-12.
75. Wallace, D.C., *Mitochondrial diseases in man and mouse.* Science, 1999. **283**(5407): p. 1482-8.

76. Chakravarti, A., A.G. Clark, and V.K. Mootha, *Distilling pathophysiology from complex disease genetics.* Cell, 2013. **155**(1): p. 21-6.
77. Kim, H. and J.S. Kim, *A guide to genome engineering with programmable nucleases.* Nat Rev Genet, 2014. **15**(5): p. 321-34.
78. Sander, J.D. and J.K. Joung, *CRISPR-Cas systems for editing, regulating and targeting genomes.* Nat Biotechnol, 2014. **32**(4): p. 347-55.
79. Hasenfuss, S.C., et al., *Regulation of steatohepatitis and PPARγ signaling by distinct AP-1 dimers.* Cell Metabolism, 2014. **19**(1).
80. Pirinen, E., et al., *Pharmacological Inhibition of poly(ADP-ribose) polymerases improves fitness and mitochondrial function in skeletal muscle.* Cell Metab, 2014. **19**(6): p. 1034-41.
81. Koutnikova, H., et al., *Identification of the UBP1 locus as a critical blood pressure determinant using a combination of mouse and human genetics.* PLoS Genet, 2009. **5**(8): p. e1000591.
82. Houtkooper, R.H., et al., *Mitonuclear protein imbalance as a conserved longevity mechanism.* Nature, 2013. **497**(7450): p. 451-7.
83. Neely, G.G., et al., *A genome-wide Drosophila screen for heat nociception identifies alpha2delta3 as an evolutionarily conserved pain gene.* Cell, 2010. **143**(4): p. 628-38.
84. Williams, E.G., et al., *An evolutionarily conserved role for the aryl hydrocarbon receptor in the regulation of movement.* PLoS Genet, 2014. **10**(9).
85. Rueedi, R., et al., *Genome-wide association study of metabolic traits reveals novel gene-metabolite-disease links.* PLoS Genet, 2014. **10**(2): p. e1004132.
86. Wilhelm, M., et al., *Mass-spectrometry-based draft of the human proteome.* Nature, 2014. **509**(7502): p. 582-7.
87. Kim, M.S., et al., *A draft map of the human proteome.* Nature, 2014. **509**(7502): p. 575-81.
88. Parks, B.W., et al., *Genetic control of obesity and gut microbiota composition in response to high-fat, high-sucrose diet in mice.* Cell Metab, 2013. **17**(1): p. 141-52.
89. Jumbo-Lucioni, P., et al., *Nuclear genomic control of naturally occurring variation in mitochondrial function in Drosophila melanogaster.* BMC Genomics, 2012. **13**: p. 659.
90. Perlstein, E.O., et al., *Genetic basis of individual differences in the response to small-molecule drugs in yeast.* Nat Genet, 2007. **39**(4): p. 496-502.
91. Ghazalpour, A., et al., *Genetic regulation of mouse liver metabolite levels.* Mol Syst Biol, 2014. **10**: p. 730.
92. Wang, T.J., et al., *2-Aminoadipic acid is a biomarker for diabetes risk.* J Clin Invest, 2013. **123**(10): p. 4309-17.
93. Korte, A. and A. Farlow, *The advantages and limitations of trait analysis with GWAS: a review.* Plant Methods, 2013. **9**(1): p. 29.
94. Manolio, T.A., et al., *Finding the missing heritability of complex diseases.* Nature, 2009. **461**(7265): p. 747-53.
95. Makinen, V.P., et al., *Integrative genomics reveals novel molecular pathways and gene networks for coronary artery disease.* PLoS Genet, 2014. **10**(7): p. e1004502.
96. Flannick, J., et al., *Loss-of-function mutations in SLC30A8 protect against type 2 diabetes.* Nat Genet, 2014. **46**(4): p. 357-63.
97. Hartiala, J., et al., *Comparative Genome-Wide Association Studies in Mice and Humans for Trimethylamine N-Oxide, a Proatherogenic Metabolite of Choline and l-Carnitine.* Arteriosclerosis Thrombosis and Vascular Biology, 2014. **34**(6): p. 1307-1313.
98. Chen, Y., et al., *Variations in DNA elucidate molecular networks that cause disease.* Nature, 2008. **452**(7186): p. 429-35.
99. Nawy, T., *Single-cell sequencing.* Nat Methods, 2014. **11**(1): p. 18.
100. Marras, S.A., et al., *Real-time measurement of in vitro transcription.* Nucleic Acids Res, 2004. **32**(9): p. e72.
101. Rost, H.L., et al., *OpenSWATH enables automated, targeted analysis of data-independent acquisition MS data.* Nat Biotechnol, 2014. **32**(3): p. 219-23.

102. Barretina, J., et al., *The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity.* Nature, 2012. **483**(7391): p. 603-7.
103. Elmer, G.I., et al., *Qualitative differences between C57BL/6J and DBA/2J mice in morphine potentiation of brain stimulation reward and intravenous self-administration.* Psychopharmacology (Berl), 2010. **208**(2): p. 309-21.
104. Brem, R.B., et al., *Genetic dissection of transcriptional regulation in budding yeast.* Science, 2002. **296**(5568): p. 752-5.
105. Ehrenreich, I.M., et al., *Dissection of genetically complex traits with extremely large pools of yeast segregants.* Nature, 2010. **464**(7291): p. 1039-42.
106. Pellegrino, M.W., A.M. Nargund, and C.M. Haynes, *Signaling the mitochondrial unfolded protein response.* Biochim Biophys Acta, 2013. **1833**(2): p. 410-6.
107. Haynes, C.M. and D. Ron, *The mitochondrial UPR - protecting organelle protein homeostasis.* J Cell Sci, 2010. **123**(Pt 22): p. 3849-55.
108. Foster, L.J., et al., *A mammalian organelle map by protein correlation profiling.* Cell, 2006. **125**(1): p. 187-99.
109. Wing, R.R., et al., *Behavioral science research in diabetes: lifestyle changes related to obesity, eating behavior, and physical activity.* Diabetes Care, 2001. **24**(1): p. 117-23.
110. Peirce, J.L., et al., *A new set of BXD recombinant inbred lines from advanced intercross populations in mice.* BMC Genet, 2004. **5**(1): p. 7.
111. Wang, X., et al., *High-throughput sequencing of the DBA/2J mouse genome.* BMC Bioinformatics, 2010. **11**(Suppl 4): p. O7.
112. Hockenbery, D.M., et al., *Bcl-2 functions in an antioxidant pathway to prevent apoptosis.* Cell, 1993. **75**(2): p. 241-51.
113. Bystrykh, L., et al., *Uncovering regulatory pathways that affect hematopoietic stem cell function using 'genetical genomics'.* Nat Genet, 2005. **37**(3): p. 225-32.
114. Troyanskaya, O., et al., *Missing value estimation methods for DNA microarrays.* Bioinformatics, 2001. **17**(6): p. 520-5.
115. Chesler, E.J., et al., *Complex trait analysis of gene expression uncovers polygenic and pleiotropic networks that modulate nervous system function.* Nat Genet, 2005. **37**(3): p. 233-42.
116. Suhre, K., et al., *A genome-wide association study of metabolic traits in human urine.* Nat Genet, 2011. **43**(6): p. 565-9.
117. Gusella, J.F., et al., *A polymorphic DNA marker genetically linked to Huntington's disease.* Nature, 1983. **306**(5940): p. 234-8.
118. Auwerx, J., *Improving metabolism by increasing energy expenditure.* Nat Med, 2006. **12**(1): p. 44-5; discussion 45.
119. Lee, S., et al., *Comparison between surrogate indexes of insulin sensitivity and resistance and hyperinsulinemic euglycemic clamp estimates in mice.* American Journal of Physiology-Endocrinology and Metabolism, 2008. **294**(2): p. E261-E270.
120. Geisert, E.E., et al., *Gene expression in the mouse eye: an online resource for genetics using 103 strains of mice.* Mol Vis, 2009. **15**: p. 1730-63.
121. Khan, Z., et al., *Primate transcript and protein expression levels evolve under compensatory selection pressures.* Science, 2013. **342**(6162): p. 1100-4.
122. Poland, A., D. Palen, and E. Glover, *Analysis of the four alleles of the murine aryl hydrocarbon receptor.* Mol Pharmacol, 1994. **46**(5): p. 915-21.
123. Champy, M.F., et al., *Mouse functional genomics requires standardization of mouse handling and housing conditions.* Mamm Genome, 2004. **15**(10): p. 768-83.
124. Mitchell, J., et al., *Gene Indexing: Characterization and Analysis of NLM's GeneRIFs.* AMIA Annu Symp Proc., 2003: p. 460-464.
125. Mackay, T.F., *The genetic architecture of quantitative traits.* Annu Rev Genet, 2001. **35**: p. 303-39.
126. Chang, C.Y., et al., *10 Nucleotide Differences, 5 of Which Cause Amino-Acid Changes, Are Associated with the Ah Receptor Locus Polymorphism of C57bl/6 and Dba/2 Mice.* Pharmacogenetics, 1993. **3**(6): p. 312-321.

127. Bertazzi, P.A., et al., *The Seveso studies on early and long-term effects of dioxin exposure: a review.* Environ Health Perspect, 1998. **106 Suppl 2**: p. 625-33.
128. C, C.L. and A.B. Aceves, *Analysis of switching phenomena in a dense medium of two-level atoms.* Opt Lett, 1993. **18**(9): p. 687-9.
129. Consortium, U., *Reorganizing the protein space at the Universal Protein Resource (UniProt).* Nucleic Acids Research, 2012. **40**(D1): p. D71-D75.
130. Foreman, J., et al., *Serum alkaline phosphatase activity is regulated by a chromosomal region containing the alkaline phosphatase 2 gene (Akp2) in C57BL/6J and DBA/2J mice.* Physiol Genomics, 2005. **23**(3): p. 295-303.
131. Bai, P. and C. Canto, *The role of PARP-1 and PARP-2 enzymes in metabolic regulation and disease.* Cell Metab, 2012. **16**(3): p. 290-5.
132. Mornet, E., *Hypophosphatasia.* Best Pract Res Clin Rheumatol, 2008. **22**(1): p. 113-27.
133. Hwang, H., et al., *Proteomics analysis of human skeletal muscle reveals novel abnormalities in obesity and type 2 diabetes.* Diabetes, 2010. **59**(1): p. 33-42.
134. Wolff, S., J.S. Weissman, and A. Dillin, *Differential scales of protein quality control.* Cell, 2014. **157**(1): p. 52-64.
135. Tutt, A., et al., *Oral poly(ADP-ribose) polymerase inhibitor olaparib in patients with BRCA1 or BRCA2 mutations and advanced breast cancer: a proof-of-concept trial.* Lancet, 2010. **376**(9737): p. 235-44.
136. Andersen, J.S., et al., *Proteomic characterization of the human centrosome by protein correlation profiling.* Nature, 2003. **426**(6966): p. 570-4.
137. Lander, E.S., et al., *Initial sequencing and analysis of the human genome.* Nature, 2001. **409**(6822): p. 860-921.
138. Rual, J.F., et al., *Toward improving Caenorhabditis elegans phenome mapping with an ORFeome-based RNAi library.* Genome Res, 2004. **14**(10B): p. 2162-8.
139. Ihmels, J., et al., *Revealing modular organization in the yeast transcriptional network.* Nat Genet, 2002. **31**(4): p. 370-7.
140. Mitchell, K.J., T. Tsuboi, and G.A. Rutter, *Role for plasma membrane-related Ca2+-ATPase-1 (ATP2C1) in pancreatic beta-cell Ca2+ homeostasis revealed by RNA silencing.* Diabetes, 2004. **53**(2): p. 393-400.
141. Mozhui, K., et al., *Sex-specific modulation of gene expression networks in murine hypothalamus.* Front Neurosci, 2012. **6**: p. 63.
142. McMillan, B.J. and C.A. Bradfield, *The aryl hydrocarbon receptor sans xenobiotics: endogenous function in genetic model systems.* Mol Pharmacol, 2007. **72**(3): p. 487-98.
143. Lander, E.S. and D. Botstein, *Mapping mendelian factors underlying quantitative traits using RFLP linkage maps.* Genetics, 1989. **121**(1): p. 185-99.
144. Costa, C.C.G., et al., *Dynamic changes of plasma acylcarnitine levels induced by fasting and sunflower oil challenge test in children.* Pediatric Research, 1999. **46**(4): p. 440-444.
145. Kent, W.J., et al., *The human genome browser at UCSC.* Genome Res, 2002. **12**(6): p. 996-1006.
146. Suhre, K., et al., *Human metabolic individuality in biomedical and pharmaceutical research.* Nature, 2011. **477**(7362): p. 54-60.
147. Jeuken, A., et al., *Activation of the Ah receptor by extracts of dietary herbal supplements, vegetables, and fruits.* Journal of Agricultural and Food Chemistry, 2003. **51**(18): p. 5478-5487.
148. Illig, T., et al., *A genome-wide perspective of genetic variation in human metabolism.* Nat Genet, 2010. **42**(2): p. 137-41.
149. Bai, P., et al., *PARP-1 inhibition increases mitochondrial metabolism through SIRT1 activation.* Cell Metabolism, 2011. **13**(4): p. 461-468.
150. Curtin, N.J. and C. Szabo, *Therapeutic applications of PARP inhibitors: anticancer therapy and beyond.* Mol Aspects Med, 2013. **34**(6): p. 1217-56.
151. Karpievitch, Y.V., A.R. Dabney, and R.D. Smith, *Normalization and missing value imputation for label-free LC-MS analysis.* BMC Bioinformatics, 2012. **13 Suppl 16**: p. S5.

152. Wishart, D.S., et al., *HMDB 3.0--The Human Metabolome Database in 2013.* Nucleic Acids Res, 2013. **41**(Database issue): p. D801-7.
153. Deeb, S., et al., *A Pro 12 Ala substitution in the human peroxisome proliferator-activated receptor gamma2 is associated with decreased receptor activity, improved insulin sensitivity , and lowered body mass index.* Nature Genetics, 1998. **20**: p. 284-287.
154. Newgard, C.B., et al., *A branched-chain amino acid-related metabolic signature that differentiates obese and lean humans and contributes to insulin resistance.* Cell Metab, 2009. **9**(4): p. 311-26.
155. Houtkooper, R.H., et al., *The secret life of NAD+: an old metabolite controlling new metabolic signaling pathways.* Endocr Rev, 2010. **31**(2): p. 194-223.
156. Nouws, J., et al., *Acyl-CoA dehydrogenase 9 is required for the biogenesis of oxidative phosphorylation complex I.* Cell Metabolism, 2010. **12**(3): p. 283-294.
157. Champy, M.F., et al., *Genetic background determines metabolic phenotypes in the mouse.* Mamm Genome, 2008. **19**(5): p. 318-31.
158. Chesler, E.J., et al., *WebQTL: rapid exploratory analysis of gene expression and genetic networks for brain and behavior.* Nat Neurosci, 2004. **7**(5): p. 485-6.
159. Abiola, O., et al., *The nature and identification of quantitative trait loci: a community's view.* Nat Rev Genet, 2003. **4**(11): p. 911-6.
160. Wiltshire, T., et al., *Genome-wide single-nucleotide polymorphism analysis defines haplotype patterns in mouse.* Proc Natl Acad Sci U S A, 2003. **100**(6): p. 3380-5.
161. Ema, M., et al., *Dioxin Binding Activities of Polymorphic Forms of Mouse and Human Arylhydrocarbon Receptors.* J Biol Chem, 1994. **269**(44): p. 27337-27343.
162. Chang, C., et al., *Ten nucleotide differences, five of which cause amino acid changes, are associated with the Ah receptor locus polymorphism of C57BL/6 and DBA/2 mice.* Pharmacogenetics, 1993. **3**(6): p. 312-21.
163. Connor, K.T. and L.L. Aylward, *Human response to dioxin: aryl hydrocarbon receptor (AhR) molecular structure, function, and dose-response data for enzyme induction indicate an impaired human AhR.* J Toxicol Environ Health B Crit Rev, 2006. **9**(2): p. 147-71.
164. Flaveny, C.A., et al., *Ligand selectivity and gene regulation by the human aryl hydrocarbon receptor in transgenic mice.* Mol Pharmacol, 2009. **75**(6): p. 1412-20.
165. Gasiewicz, T.A. and G. Rucci, *Cytosolic receptor for 2,3,7,8-tetrachlorodibenzo-p-dioxin. Evidence for a homologous nature among various mammalian species.* Mol Pharmacol, 1984. **26**(1): p. 90-8.
166. Lesca, P., et al., *The pig as a model for studying AH receptor and other PAH-binding proteins in man.* Biochem Biophys Res Commun, 1994. **200**(1): p. 475-81.
167. Flaveny, C.A. and G.H. Perdew, *Transgenic Humanized AHR Mouse Reveals Differences between Human and Mouse AHR Ligand Selectivity.* Mol Cell Pharmacol, 2009. **1**(3): p. 119-123.
168. Duncan, D.M., E.A. Burgess, and I. Duncan, *Control of distal antennal identity and tarsal development in Drosophila by spineless-aristapedia, a homolog of the mammalian dioxin receptor.* Genes Dev, 1998. **12**(9): p. 1290-303.
169. Huang, X., J.A. Powell-Coffman, and Y. Jin, *The AHR-1 aryl hydrocarbon receptor and its co-factor the AHA-1 aryl hydrocarbon receptor nuclear translocator specify GABAergic neuron cell fate in C. elegans.* Development, 2004. **131**(4): p. 819-28.
170. Fernandez-Salguero, P., et al., *Immune system impairment and hepatic fibrosis in mice lacking the dioxin-binding Ah receptor.* Science, 1995. **268**(5211): p. 722-6.
171. Andreola, F., et al., *Aryl hydrocarbon receptor knockout mice (AHR-/-) exhibit liver retinoid accumulation and reduced retinoic acid metabolism.* Cancer Res, 1997. **57**(14): p. 2835-8.
172. Kudo, K., et al., *Characterization of the region of the aryl hydrocarbon receptor required for ligand dependency of transactivation using chimeric receptor between Drosophila and Mus musculus.* Biochim Biophys Acta, 2009. **1789**(6-8): p. 477-86.
173. Casper, R.F., et al., *Resveratrol has antagonist activity on the aryl hydrocarbon receptor: Implications for prevention of dioxin toxicity.* Molecular Pharmacology, 1999. **56**(4): p. 784-790.

174. Matikainen, T., et al., *Aromatic hydrocarbon receptor-driven Bax gene expression is required for premature ovarian failure caused by biohazardous environmental chemicals.* Nat Genet, 2001. **28**(4): p. 355-60.
175. Zhang, J., et al., *Neuronal target identification requires AHA-1-mediated fine-tuning of Wnt signaling in C. elegans.* PLoS Genet, 2013. **9**(6): p. e1003618.
176. Chang, C.Y. and A. Puga, *Constitutive activation of the aromatic hydrocarbon receptor.* Mol Cell Biol, 1998. **18**(1): p. 525-35.
177. Cauchi, S., et al., *Structure and polymorphisms of human aryl hydrocarbon receptor repressor (AhRR) gene in a French population: relationship with CYP1A1 inducibility and lung cancer.* Pharmacogenetics, 2003. **13**(6): p. 339-47.
178. Rowlands, C.J., et al., *The human AHR: identification of single nucleotide polymorphisms from six ethnic populations.* Pharmacogenet Genomics, 2010. **20**(5): p. 283-90.
179. Mandillo, S., et al., *Reliability, robustness, and reproducibility in mouse behavioral phenotyping: a cross-laboratory study.* Physiol Genomics, 2008. **34**(3): p. 243-55.
180. Ramot, D., et al., *The Parallel Worm Tracker: a platform for measuring average speed and drug-induced paralysis in nematodes.* PLoS One, 2008. **3**(5): p. e2208.
181. Wang, J.T., R.W. Williams, and K.F. Manly, *WebQTL - Web-based complex trait analysis.* Neuroinformatics, 2003. **1**(4): p. 299-308.
182. Desvergne, B., L. Michalik, and W. Wahli, *Transcriptional regulation of metabolism.* Physiol Rev, 2006. **86**(2): p. 465-514.
183. Francis, G.A., et al., *Nuclear receptors and the control of metabolism.* Annu Rev Physiol, 2003. **65**: p. 261-311.
184. Feige, J.N. and J. Auwerx, *Transcriptional coregulators in the control of energy homeostasis.* Trends Cell Biol, 2007. **17**(6): p. 292-301.
185. Rosenfeld, M.G., V.V. Lunyak, and C.K. Glass, *Sensors and signals: a coactivator/corepressor/epigenetic code for integrating signal-dependent programs of transcriptional response.* Genes Dev, 2006. **20**(11): p. 1405-28.
186. Smith, C.L. and B.W. O'Malley, *Coregulator function: a key to understanding tissue specificity of selective receptor modulators.* Endocr Rev, 2004. **25**(1): p. 45-71.
187. Spiegelman, B.M. and R. Heinrich, *Biological control through regulated transcriptional coactivators.* Cell, 2004. **119**(2): p. 157-67.
188. Handschin, C. and B.M. Spiegelman, *Peroxisome proliferator-activated receptor gamma coactivator 1 coactivators, energy homeostasis, and metabolism.* Endocr Rev, 2006. **27**(7): p. 728-35.
189. Fernandez-Marcos, P.J. and J. Auwerx, *Regulation of PGC-1alpha, a nodal regulator of mitochondrial biogenesis.* Am J Clin Nutr, 2011. **93**(4): p. 884S-90.
190. Perissi, V., et al., *Deconstructing repression: evolving models of co-repressor action.* Nat Rev Genet, 2010. **11**(2): p. 109-23.
191. Jepsen, K., et al., *Combinatorial roles of the nuclear receptor corepressor in transcription and development.* Cell, 2000. **102**(6): p. 753-63.
192. Jepsen, K., et al., *SMRT-mediated repression of an H3K27 demethylase in progression from neural stem cell to neuron.* Nature, 2007. **450**(7168): p. 415-9.
193. Nofsinger, R.R., et al., *SMRT repression of nuclear receptors controls the adipogenic set point and metabolic homeostasis.* Proc Natl Acad Sci U S A, 2008. **105**(50): p. 20021-6.
194. Yu, C., et al., *The nuclear receptor corepressors NCoR and SMRT decrease peroxisome proliferator-activated receptor gamma transcriptional activity and repress 3T3-L1 adipogenesis.* J Biol Chem, 2005. **280**(14): p. 13600-5.
195. Picard, F., et al., *Sirt1 promotes fat mobilization in white adipocytes by repressing PPAR-gamma.* Nature, 2004. **429**(6993): p. 771-6.
196. Hörlein, A.J., et al., *Ligand-independent repression by the thyroid hormone receptor mediated by a nuclear receptor co-repressor.* Nature, 1995. **377**: p. 397-404.

197. Miniou, P., et al., *Gene targeting restricted to mouse striated muscle lineage.* Nucleic Acids Res, 1999. **27**(19): p. e27.
198. Chen, J.D. and R.M. Evans, *A transcriptional co-repressor that interacts with nuclear hormone receptors [see comments].* Nature, 1995. **377**(6548): p. 454-7.
199. Schiaffino, S., et al., *Three myosin heavy chain isoforms in type 2 skeletal muscle fibres.* J Muscle Res Cell Motil, 1989. **10**(3): p. 197-205.
200. Durieux, J., S. Wolff, and A. Dillin, *The cell-non-autonomous nature of electron transport chain-mediated longevity.* Cell, 2011. **144**(1): p. 79-91.
201. Argmann, C.A., P. Chambon, and J. Auwerx, *Mouse phenogenomics: the fast track to "systems metabolism".* Cell Metab, 2005. **2**(6): p. 349-60.
202. Houtkooper, R.H., R.W. Williams, and J. Auwerx, *Metabolic networks of longevity.* Cell, 2010. **142**(1): p. 9-14.
203. van Nas, A., et al., *Expression Quantitative Trait Loci: Replication, Tissue- and Sex-Specificity in Mice.* Genetics, 2010. **185**(3): p. 1059-1068.
204. Alberts, R., et al., *Genome-wide analysis of the mouse lung transcriptome reveals novel molecular gene interaction networks and cell-specific expression signatures.* Respir Res, 2011. **12**: p. 61.
205. Hagberg, C.E., et al., *Vascular endothelial growth factor B controls endothelial fatty acid uptake.* Nature, 2010. **464**(7290): p. 917-21.
206. Arany, Z., et al., *HIF-independent regulation of VEGF and angiogenesis by the transcriptional coactivator PGC-1alpha.* Nature, 2008. **451**(7181): p. 1008-12.
207. Zhang, Y., et al., *Estrogen-related receptors stimulate pyruvate dehydrogenase kinase isoform 4 gene expression.* J Biol Chem, 2006. **281**(52): p. 39897-906.
208. Ma, K., et al., *Myocyte enhancer factor 2 acetylation by p300 enhances its DNA binding activity, transcriptional activity, and myogenic differentiation.* Mol Cell Biol, 2005. **25**(9): p. 3575-82.
209. Nebbioso, A., et al., *Selective class II HDAC inhibitors impair myogenesis by modulating the stability and activity of HDAC-MEF2 complexes.* EMBO Rep, 2009. **10**(7): p. 776-82.
210. Canto, C. and J. Auwerx, *Caloric restriction, SIRT1 and longevity.* Trends Endocrinol Metab, 2009. **20**(7): p. 325-31.
211. Alenghat, T., et al., *Nuclear receptor corepressor and histone deacetylase 3 govern circadian metabolic physiology.* Nature, 2008. **456**(7224): p. 997-1000.
212. Feng, D., et al., *A circadian rhythm orchestrated by histone deacetylase 3 controls hepatic lipid metabolism.* Science, 2011. **331**(6022): p. 1315-9.
213. Black, B.L. and E.N. Olson, *Transcriptional control of muscle development by myocyte enhancer factor-2 (MEF2) proteins.* Annu Rev Cell Dev Biol, 1998. **14**: p. 167-96.
214. McKinsey, T.A., C.L. Zhang, and E.N. Olson, *Control of muscle development by dueling HATs and HDACs.* Curr Opin Genet Dev, 2001. **11**(5): p. 497-504.
215. McKinsey, T.A., C.L. Zhang, and E.N. Olson, *MEF2: a calcium-dependent regulator of cell division, differentiation and death.* Trends Biochem Sci, 2002. **27**(1): p. 40-7.
216. Potthoff, M.J. and E.N. Olson, *MEF2: a central regulator of diverse developmental programs.* Development, 2007. **134**(23): p. 4131-40.
217. Zhang, C.L., et al., *Class II histone deacetylases act as signal-responsive repressors of cardiac hypertrophy.* Cell, 2002. **110**(4): p. 479-88.
218. Gregoire, S., et al., *Histone deacetylase 3 interacts with and deacetylates myocyte enhancer factor 2.* Mol Cell Biol, 2007. **27**(4): p. 1280-95.
219. Haberland, M., et al., *Regulation of HDAC9 gene expression by MEF2 establishes a negative-feedback loop in the transcriptional circuitry of muscle differentiation.* Mol Cell Biol, 2007. **27**(2): p. 518-25.
220. Zhao, X., et al., *Regulation of MEF2 by histone deacetylase 4- and SIRT1 deacetylase-mediated lysine modifications.* Mol Cell Biol, 2005. **25**(19): p. 8456-64.
221. Alaynick, W.A., *Nuclear receptors, mitochondria and lipid metabolism.* Mitochondrion, 2008. **8**(4): p. 329-37.

222. Luquet, S., et al., *Peroxisome proliferator-activated receptor delta controls muscle development and oxidative capability.* Faseb J, 2003. **17**(15): p. 2299-301.
223. Tanaka, T., et al., *Activation of peroxisome proliferator-activated receptor delta induces fatty acid beta-oxidation in skeletal muscle and attenuates metabolic syndrome.* Proc Natl Acad Sci U S A, 2003. **100**(26): p. 15924-9.
224. Wang, Y.X., et al., *Regulation of muscle fiber type and running endurance by PPARdelta.* PLoS Biol, 2004. **2**(10): p. e294.
225. Giguere, V., *Transcriptional control of energy homeostasis by the estrogen-related receptors.* Endocr Rev, 2008. **29**(6): p. 677-96.
226. Villena, J.A. and A. Kralli, *ERRalpha: a metabolic function for the oldest orphan.* Trends Endocrinol Metab, 2008. **19**(8): p. 269-76.
227. Dufour, C.R., et al., *Genome-wide orchestration of cardiac functions by the orphan nuclear receptors ERRalpha and gamma.* Cell Metab, 2007. **5**(5): p. 345-56.
228. Narkar, V.A., et al., *Exercise and PGC-1alpha-independent synchronization of type I muscle metabolism and vasculature by ERRgamma.* Cell Metab, 2011. **13**(3): p. 283-93.
229. Lin, J., et al., *Transcriptional co-activator PGC-1 alpha drives the formation of slow-twitch muscle fibres.* Nature, 2002. **418**(6899): p. 797-801.
230. Sengupta, S., et al., *mTORC1 controls fasting-induced ketogenesis and its modulation by ageing.* Nature, 2010. **468**(7327): p. 1100-U502.
231. Iezzi, S., et al., *Deacetylase inhibitors increase muscle cell size by promoting myoblast recruitment and fusion through induction of follistatin.* Dev Cell, 2004. **6**(5): p. 673-84.
232. Lagouge, M., et al., *Resveratrol improves mitochondrial function and protects against metabolic disease by activating SIRT1 and PGC-1alpha.* Cell, 2006. **127**(6): p. 1109-22.
233. Andreux, P.A., R.H. Houtkooper, and J. Auwerx, *Pharmacological approaches to restore mitochondrial function.* Nat Rev Drug Discov, 2013.
234. Houtkooper, R.H., E. Pirinen, and J. Auwerx, *Sirtuins as regulators of metabolism and healthspan.* Nature Reviews Molecular Cell Biology, 2012. **13**(4): p. 225-238.
235. Canto, C. and J. Auwerx, *Targeting sirtuin 1 to improve metabolism: all you need is NAD(+)?* Pharmacol Rev, 2012. **64**(1): p. 166-87.
236. Peek, C.B., et al., *Circadian clock NAD+ cycle drives mitochondrial oxidative metabolism in mice.* Science, 2013. **342**(6158): p. 1243417.
237. Menissier de Murcia, J., et al., *Functional interaction between PARP-1 and PARP-2 in chromosome stability and embryonic development in mouse.* EMBO J, 2003. **22**(9): p. 2255-63.
238. Menissier de Murcia, J., et al., *Requirement of poly(ADP-ribose) polymerase in recovery from DNA damage in mice and in cells.* Proc Natl Acad Sci U S A, 1997. **94**(14): p. 7303-7.
239. Menzies, K.J., et al., *Sirtuin 1-mediated effects of exercise and resveratrol on mitochondrial biogenesis.* J Biol Chem, 2013. **288**(10): p. 6968-79.
240. Philp, A., et al., *Sirtuin 1 (SIRT1) Deacetylase Activity Is Not Required for Mitochondrial Biogenesis or Peroxisome Proliferator-activated Receptor-gamma Coactivator-1 alpha (PGC-1 alpha) Deacetylation following Endurance Exercise.* Journal of Biological Chemistry, 2011. **286**(35): p. 30561-30570.
241. Gomes, A.P., et al., *Declining NAD(+) induces a pseudohypoxic state disrupting nuclear-mitochondrial communication during aging.* Cell, 2013. **155**(7): p. 1624-38.
242. Erener, S., et al., *ARTD1 deletion causes increased hepatic lipid accumulation in mice fed a high-fat diet and impairs adipocyte function and differentiation.* FASEB J, 2012. **26**(6): p. 2631-8.
243. Bundred, N., et al., *Evaluation of the pharmacodynamics and pharmacokinetics of the PARP inhibitor olaparib: a phase I multicentre trial in patients scheduled for elective breast cancer surgery.* Invest New Drugs, 2013. **31**(4): p. 949-58.
244. Audeh, M.W., et al., *Oral poly(ADP-ribose) polymerase inhibitor olaparib in patients with BRCA1 or BRCA2 mutations and recurrent ovarian cancer: a proof-of-concept trial.* Lancet, 2010. **376**(9737): p. 245-51.

245. Sethi, J.K. and A.J. Vidal-Puig, *Thematic review series: adipocyte biology. Adipose tissue function and plasticity orchestrate nutritional adaptation.* J Lipid Res, 2007. **48**(6): p. 1253-62.
246. Lara-Castro, C. and W.T. Garvey, *Intracellular lipid accumulation in liver and muscle and the insulin resistance syndrome.* Endocrinol Metab Clin North Am, 2008. **37**(4): p. 841-56.
247. Browning, J.D. and J.D. Horton, *Molecular mediators of hepatic steatosis and liver injury.* J Clin Invest, 2004. **114**(2): p. 147-52.
248. Farese, R.V., Jr., et al., *The problem of establishing relationships between hepatic steatosis and hepatic insulin resistance.* Cell Metab, 2012. **15**(5): p. 570-3.
249. Smedile, A. and E. Bugianesi, *Steatosis and hepatocellular carcinoma risk.* Eur Rev Med Pharmacol Sci, 2005. **9**(5): p. 291-3.
250. Cohen, J.C., J.D. Horton, and H.H. Hobbs, *Human fatty liver disease: old questions and new insights.* Science, 2011. **332**(6037): p. 1519-23.
251. Lazo, M. and J.M. Clark, *The epidemiology of nonalcoholic fatty liver disease: a global perspective.* Semin Liver Dis, 2008. **28**(4): p. 339-50.
252. Halazonetis, T.D., et al., *c-Jun dimerizes with itself and with c-Fos, forming complexes of different DNA binding affinities.* Cell, 1988. **55**(5): p. 917-24.
253. Eferl, R. and E.F. Wagner, *AP-1: a double-edged sword in tumorigenesis.* Nat Rev Cancer, 2003. **3**(11): p. 859-68.
254. Hess, J., P. Angel, and M. Schorpp-Kistner, *AP-1 subunits: quarrel and harmony among siblings.* J Cell Sci, 2004. **117**(Pt 25): p. 5965-73.
255. Verde, P., et al., *Deciphering AP-1 function in tumorigenesis: fra-ternizing on target promoters.* Cell Cycle, 2007. **6**(21): p. 2633-9.
256. Wagner, E.F., et al., *Psoriasis: what we have learned from mouse models.* Nat Rev Rheumatol, 2010. **6**(12): p. 704-14.
257. Shaulian, E. and M. Karin, *AP-1 as a regulator of cell life and death.* Nat Cell Biol, 2002. **4**(5): p. E131-6.
258. Wagner, E.F. and A.R. Nebreda, *Signal integration by JNK and p38 MAPK pathways in cancer development.* Nat Rev Cancer, 2009. **9**(8): p. 537-49.
259. Bakiri, L. and E.F. Wagner, *Mouse models for liver cancer.* Mol Oncol, 2013. **7**(2): p. 206-23.
260. Fuest, M., et al., *The transcription factor c-Jun protects against sustained hepatic endoplasmic reticulum stress thereby promoting hepatocyte survival.* Hepatology, 2012. **55**(2): p. 408-18.
261. Hasselblatt, P., et al., *Hepatocyte survival in acute hepatitis is due to c-Jun/AP-1-dependent expression of inducible nitric oxide synthase.* Proc Natl Acad Sci U S A, 2007. **104**(43): p. 17105-10.
262. Behrens, A., et al., *Impaired postnatal hepatocyte proliferation and liver regeneration in mice lacking c-jun in the liver.* EMBO J, 2002. **21**(7): p. 1782-90.
263. Eferl, R., et al., *Liver tumor development. c-Jun antagonizes the proapoptotic activity of p53.* Cell, 2003. **112**(2): p. 181-92.
264. Machida, K., et al., *c-Jun mediates hepatitis C virus hepatocarcinogenesis through signal transducer and activator of transcription 3 and nitric oxide-dependent impairment of oxidative DNA repair.* Hepatology, 2010. **52**(2): p. 480-92.
265. Min, L., et al., *Liver cancer initiation is controlled by AP-1 through SIRT6-dependent inhibition of survivin.* Nat Cell Biol, 2012. **14**(11): p. 1203-11.
266. Hasenfuss, S.C., et al., *The AP-1 transcription factor Fra-1 is dispensable for murine liver fibrosis, but modulates xenobiotic metabolism.* Hepatology, 2013.
267. Bakiri, L., et al., *Promoter specificity and biological activity of tethered AP-1 dimers.* Mol Cell Biol, 2002. **22**(13): p. 4952-64.
268. Peirce, J.L., et al., *A new set of BXD recombinant inbred lines from advanced intercross populations in mice.* BMC Genet, 2004. **5**: p. 7.
269. Kanehisa, M., et al., *KEGG for integration and interpretation of large-scale molecular data sets.* Nucleic Acids Res, 2012. **40**(Database issue): p. D109-14.
270. Huang da, W., B.T. Sherman, and R.A. Lempicki, *Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources.* Nat Protoc, 2009. **4**(1): p. 44-57.

271. Lee, Y.J., et al., *Nuclear receptor PPARgamma-regulated monoacylglycerol O-acyltransferase 1 (MGAT1) expression is responsible for the lipid accumulation in diet-induced hepatic steatosis.* Proc Natl Acad Sci U S A, 2012. **109**(34): p. 13656-61.
272. Kim, H.I., et al., *Transcriptional activation of SHP by PPAR-gamma in liver.* Biochem Biophys Res Commun, 2007. **360**(2): p. 301-6.
273. Kim, S.C., et al., *All-trans-retinoic acid ameliorates hepatic steatosis in mice via a novel transcriptional cascade.* Hepatology, 2013.
274. Matsusue, K., et al., *Liver-specific disruption of PPARgamma in leptin-deficient mice improves fatty liver but aggravates diabetic phenotypes.* J Clin Invest, 2003. **111**(5): p. 737-47.
275. Matsusue, K., et al., *Hepatic steatosis in leptin-deficient mice is promoted by the PPARgamma target gene Fsp27.* Cell Metab, 2008. **7**(4): p. 302-11.
276. Gavrilova, O., et al., *Liver peroxisome proliferator-activated receptor gamma contributes to hepatic steatosis, triglyceride clearance, and regulation of body fat mass.* J Biol Chem, 2003. **278**(36): p. 34268-76.
277. Moran-Salvador, E., et al., *Role for PPARgamma in obesity-induced hepatic steatosis as determined by hepatocyte- and macrophage-specific conditional knockouts.* FASEB J, 2011. **25**(8): p. 2538-50.
278. Medina-Gomez, G., et al., *PPAR gamma 2 prevents lipotoxicity by controlling adipose tissue expandability and peripheral lipid metabolism.* PLoS Genet, 2007. **3**(4): p. e64.
279. Fujimoto, T., et al., *Lipid droplets: a classic organelle with new outfits.* Histochem Cell Biol, 2008. **130**(2): p. 263-79.
280. Sun, Z., et al., *Hepatic Hdac3 promotes gluconeogenesis by repressing lipid synthesis and sequestration.* Nat Med, 2012. **18**(6): p. 934-42.
281. Puri, V., et al., *Cidea is associated with lipid droplets and insulin sensitivity in humans.* Proc Natl Acad Sci U S A, 2008. **105**(22): p. 7833-8.
282. Tontonoz, P. and B.M. Spiegelman, *Fat and beyond: the diverse biology of PPARgamma.* Annu Rev Biochem, 2008. **77**: p. 289-312.
283. Chang, B.H., et al., *Protection against fatty liver but normal adipogenesis in mice lacking adipose differentiation-related protein.* Mol Cell Biol, 2006. **26**(3): p. 1063-76.
284. Zhou, L., et al., *Cidea promotes hepatic steatosis by sensing dietary fatty acids.* Hepatology, 2012. **56**(1): p. 95-107.
285. Dalen, K.T., et al., *Adipose tissue expression of the lipid droplet-associating proteins S3-12 and perilipin is controlled by peroxisome proliferator-activated receptor-gamma.* Diabetes, 2004. **53**(5): p. 1243-52.
286. Varela, G.M., et al., *Inhibition of ADRP prevents diet-induced insulin resistance.* Am J Physiol Gastrointest Liver Physiol, 2008. **295**(3): p. G621-8.
287. Newberry, E.P., et al., *Protection against Western diet-induced obesity and hepatic steatosis in liver fatty acid-binding protein knockout mice.* Hepatology, 2006. **44**(5): p. 1191-205.
288. Newberry, E.P., et al., *Decreased hepatic triglyceride accumulation and altered fatty acid uptake in mice with deletion of the liver fatty acid-binding protein gene.* J Biol Chem, 2003. **278**(51): p. 51664-72.
289. Matsuo, K., et al., *Fosl1 is a transcriptional target of c-Fos during osteoclast differentiation.* Nat Genet, 2000. **24**(2): p. 184-7.
290. Fleischmann, A., et al., *Fra-1 replaces c-Fos-dependent functions in mice.* Genes Dev, 2000. **14**(21): p. 2695-700.
291. Purushotham, A., et al., *Hepatocyte-specific deletion of SIRT1 alters fatty acid metabolism and results in hepatic steatosis and inflammation.* Cell Metab, 2009. **9**(4): p. 327-38.
292. Knutson, S.K., et al., *Liver-specific deletion of histone deacetylase 3 disrupts metabolic transcriptional networks.* EMBO J, 2008. **27**(7): p. 1017-28.
293. Kim, S.J. and C.R. Kahn, *Insulin stimulates phosphorylation of c-Jun, c-Fos, and Fos-related proteins in cultured adipocytes.* J Biol Chem, 1994. **269**(16): p. 11887-92.
294. Hibi, M., et al., *Identification of an oncoprotein- and UV-responsive protein kinase that binds and potentiates the c-Jun activation domain.* Genes Dev, 1993. **7**(11): p. 2135-48.

295. Boyle, W.J., et al., *Activation of protein kinase C decreases phosphorylation of c-Jun at sites that negatively regulate its DNA-binding activity.* Cell, 1991. **64**(3): p. 573-84.
296. Fujioka, S., et al., *NF-kappaB and AP-1 connection: mechanism of NF-kappaB-dependent regulation of AP-1 activity.* Mol Cell Biol, 2004. **24**(17): p. 7806-19.
297. Ricote, M. and C.K. Glass, *PPARs and molecular mechanisms of transrepression.* Biochim Biophys Acta, 2007. **1771**(8): p. 926-35.
298. Wan, Y., L.W. Chong, and R.M. Evans, *PPAR-gamma regulates osteoclastogenesis in mice.* Nat Med, 2007. **13**(12): p. 1496-503.
299. Glass, C.K. and K. Saijo, *Nuclear receptor transrepression pathways that regulate inflammation in macrophages and T cells.* Nat Rev Immunol, 2010. **10**(5): p. 365-76.
300. Mehlem, A., et al., *Imaging of neutral lipids by oil red O for analyzing the metabolic status in health and disease.* Nat Protoc, 2013. **8**(6): p. 1149-54.
301. Irizarry, R.A., et al., *Exploration, normalization, and summaries of high density oligonucleotide array probe level data.* Biostatistics, 2003. **4**(2): p. 249-64.
302. Saladin, R., et al., *Differential regulation of peroxisome proliferator activated receptor gamma1 (PPARgamma1) and PPARgamma2 messenger RNA expression in the early stages of adipogenesis.* Cell Growth Differ, 1999. **10**(1): p. 43-8.
303. Crick, F., *Central dogma of molecular biology.* Nature, 1970. **227**(5258): p. 561-3.
304. Mardis, E.R., *The impact of next-generation sequencing technology on genetics.* Trends Genet, 2008. **24**(3): p. 133-41.
305. Ghazalpour, A., et al., *Comparative analysis of proteome and transcriptome variation in mouse.* PLoS Genet, 2011. **7**(6): p. e1001393.
306. Schwanhausser, B., et al., *Global quantification of mammalian gene expression control.* Nature, 2011. **473**(7347): p. 337-42.
307. Gygi, S.P., et al., *Correlation between protein and mRNA abundance in yeast.* Mol Cell Biol, 1999. **19**(3): p. 1720-30.
308. Skelly, D.A., et al., *Integrative phenomics reveals insight into the structure of phenotypic diversity in budding yeast.* Genome Res, 2013. **23**(9): p. 1496-504.
309. Albert, F.W., et al., *Genetics of single-cell protein abundance variation in large yeast populations.* Nature, 2014.
310. Holdt, L.M., et al., *Quantitative trait loci mapping of the mouse plasma proteome (pQTL).* Genetics, 2013. **193**(2): p. 601-8.
311. Callister, S.J., et al., *Normalization approaches for removing systematic biases associated with mass spectrometry and label-free proteomics.* J Proteome Res, 2006. **5**(2): p. 277-86.
312. Lange, V., et al., *Selected reaction monitoring for quantitative proteomics: a tutorial.* Mol Syst Biol, 2008. **4**: p. 222.
313. Picotti, P., et al., *A complete mass-spectrometric map of the yeast proteome applied to quantitative trait analysis.* Nature, 2013. **494**(7436): p. 266-70.
314. Yvert, G., et al., *Trans-acting regulatory variation in Saccharomyces cerevisiae and the role of transcription factors.* Nat Genet, 2003. **35**(1): p. 57-64.
315. De Luca, M., et al., *Dopa decarboxylase (Ddc) affects variation in Drosophila longevity.* Nat Genet, 2003. **34**(4): p. 429-33.
316. Picotti, P., et al., *High-throughput generation of selected reaction-monitoring assays for proteins and proteomes.* Nat Methods, 2010. **7**(1): p. 43-6.
317. Foss, E.J., et al., *Genetic basis of proteome variation in yeast.* Nat Genet, 2007. **39**(11): p. 1369-75.
318. Maglott, D., et al., *Entrez Gene: gene-centered information at NCBI.* Nucleic Acids Res, 2005. **33**: p. D54-8.
319. Magrane, M. and U. Consortium, *UniProt Knowledgebase: a hub of integrated protein data.* Database (Oxford), 2011. **2011**: p. bar009.
320. Wong, N., et al., *A novel mechanism regulating insulin secretion involving Herpud1 in mice.* Diabetologia, 2013. **56**(7): p. 1569-76.

321. Mitterberger, M.C., et al., *Carbonic anhydrase III regulates peroxisome proliferator-activated receptor-gamma 2.* Experimental Cell Research, 2012. **318**(8): p. 877-886.
322. Chuang, D.T. and V.E. Shih, *Maple syrup urine disease (branched-chain ketoaciduria)*, in *The Metabolic and Molecular Bases of Inherited Disease*. 2001, McGraw-Hill: New York. p. 1971–2005.
323. Haymond, M.W., E. Ben-Galim, and K.E. Strobel, *Glucose and alanine metabolism in children with maple syrup urine disease.* J Clin Invest, 1978. **62**(2): p. 398-405.
324. Koepp, P., et al., *Maple syrup urine disease variant: report on an infant.* Z Kinderheilkd, 1974. **116**(3): p. 177-84.
325. Wu, J.Y., et al., *ENU mutagenesis identifies mice with mitochondrial branched-chain aminotransferase deficiency resembling human maple syrup urine disease.* J Clin Invest, 2004. **113**(3): p. 434-40.
326. Skvorak, K.J., *Animal models of maple syrup urine disease.* J Inherit Metab Dis, 2009. **32**(2): p. 229-46.
327. Danhauser, K., et al., *DHTKD1 mutations cause 2-aminoadipic and 2-oxoadipic aciduria.* Am J Hum Genet, 2012. **91**(6): p. 1082-7.
328. Houten, S.M., et al., *Genetic basis of hyperlysinemia.* Orphanet J Rare Dis, 2013. **8**: p. 57.
329. Xu, W.Y., et al., *A nonsense mutation in DHTKD1 causes Charcot-Marie-Tooth disease type 2 in a large Chinese pedigree.* Am J Hum Genet, 2012. **91**(6): p. 1088-94.
330. Xu, W.Y., et al., *DHTKD1 is essential for mitochondrial biogenesis and function maintenance.* Febs Letters, 2013. **587**(21): p. 3587-3592.
331. Firmann, M., et al., *The CoLaus study: a population-based study to investigate the epidemiology and genetic determinants of cardiovascular risk factors and metabolic syndrome.* BMC Cardiovasc Disord, 2008. **8**: p. 6.
332. Zhao, Q., et al., *A mitochondrial specific stress response in mammalian cells.* EMBO J, 2002. **21**(17): p. 4411-9.
333. Runkel, E.D., et al., *Surveillance-activated defenses block the ROS-induced mitochondrial unfolded protein response.* PLoS Genet, 2013. **9**(3): p. e1003346.
334. Jovaisaite, V., L. Mouchiroud, and J. Auwerx, *The mitochondrial unfolded protein response, a conserved stress response pathway with implications in health and disease.* J Exp Biol, 2014. **217**(Pt 1): p. 137-43.
335. Haynes, C.M., C.J. Fiorese, and Y.F. Lin, *Evaluating and responding to mitochondrial dysfunction: the mitochondrial unfolded-protein response and beyond.* Trends Cell Biol, 2013. **23**(7): p. 311-8.
336. Yoneda, T., et al., *Compartment-specific perturbation of protein handling activates genes encoding mitochondrial chaperones.* J Cell Sci, 2004. **117**(Pt 18): p. 4055-66.
337. Schadt, E.E., et al., *Mapping the genetic architecture of gene expression in human liver.* PLoS Biol, 2008. **6**(5): p. e107.
338. Ding, L., et al., *A lung tissue bank for gene expression studies in chronic obstructive pulmonary disease.* COPD, 2004. **1**(2): p. 191-204.
339. Monks, S.A., et al., *Genetic inheritance of gene expression in human cell lines.* Am J Hum Genet, 2004. **75**(6): p. 1094-105.
340. Berchtold, N.C., et al., *Gene expression changes in the course of normal brain aging are sexually dimorphic.* Proc Natl Acad Sci U S A, 2008. **105**(40): p. 15605-10.
341. Marx, V., *Finding the right antibody for the job.* Nature Methods, 2013. **10**(10): p. 703-707.
342. Benedetti, C., et al., *Ubiquitin-like protein 5 positively regulates chaperone gene expression in the mitochondrial unfolded protein response.* Genetics, 2006. **174**(1): p. 229-39.
343. Fuhrer, T., et al., *High-throughput, accurate mass metabolome profiling of cellular extracts by flow injection-time-of-flight mass spectrometry.* Anal Chem, 2011. **83**(18): p. 7074-80.
344. Deutsch, E.W., H. Lam, and R. Aebersold, *PeptideAtlas: a resource for target selection for emerging targeted proteomics workflows.* EMBO Rep, 2008. **9**(5): p. 429-34.
345. Mallick, P., et al., *Computational prediction of proteotypic peptides for quantitative proteomics.* Nat Biotechnol, 2007. **25**(1): p. 125-31.

346. Escher, C., et al., *Using iRT, a normalized retention time for more targeted measurement of peptides.* Proteomics, 2012. **12**(8): p. 1111-21.
347. Lange, V., et al., *Targeted quantitative analysis of Streptococcus pyogenes virulence factors by multiple reaction monitoring.* Mol Cell Proteomics, 2008. **7**(8): p. 1489-500.
348. Fillatre, Y., et al., *Multiresidue analysis of multiclass pesticides in lavandin essential oil by LC/MS/MS using the scheduled selected reaction monitoring mode.* Anal Chem, 2011. **83**(1): p. 109-17.
349. Sabido, E., et al., *Targeted proteomics reveals strain-specific changes in the mouse insulin and central metabolic pathways after a sustained high-fat diet.* Mol Syst Biol, 2013. **9**: p. 681.
350. Kamath, R.S., et al., *Effectiveness of specific RNA-mediated interference through ingested double-stranded RNA in Caenorhabditis elegans.* Genome Biol, 2001. **2**(1): p. RESEARCH0002.
351. Yamamoto, H., et al., *NCoR1 is a conserved physiological modulator of muscle mass and oxidative function.* Cell, 2011. **147**(4): p. 827-39.
352. MacLean, B., et al., *Skyline: an open source document editor for creating and analyzing targeted proteomics experiments.* Bioinformatics, 2010. **26**(7): p. 966-8.