UNIVERSITÉ DU LUXEMBOURG

PhD-FSTM-2025-012

The Faculty of Sciences, Technology and Medicine

DISSERTATION

Defence held on 14/01/2025 in Luxembourg

to obtain the degree of

**DOCTEUR DE L'UNIVERSITÉ DU LUXEMBOURG EN BIOLOGIE**

by

**Ziyun ZHOU**

Born on 16 June 1992 in Beijing, (China)

# Decoding Metabolism: How Dynamic Molecular Processes React to Genetic, Environmental, and Gut Microbiota Interactions and Shape Our Health

## Dissertation defense committee

Dr Evan G. Williams, dissertation Supervisor
*Associate Professor, University of Luxembourg*

Dr Gunnar Dittmar
*Professor, Luxembourg Institute of Health*

Dr Xiaowei Zhan
*Associate Professor, UT Southwestern Medical Center, United States*

Dr Henry-Michel Cauchie
*Luxembourg Institute of Science and Technology*

Dr Thomas Sauter
*Professor, University of Luxembourg*


A Dissertation by

Ziyun ZHOU

submitted to the University of Luxembourg

in partial fulfilment of the requirements for the degree of

# DOCTEUR en BIOLOGIE

**Approved by the Dissertation Defense Committee:**

| | |
|---|---|
| Chair of committee: | Prof. Thomas Sauter |
| Committee members: | Prof. Gunnar Dittmar |
| | Assoc. Prof. Xiaowei Zhan |
| | Dr. Henry-Michel Cauchie |
| Supervisor: | Assoc. Prof. Evan G. Williams |

# Affidavit

I hereby confirm that the PhD thesis entitled "Decoding Metabolism: How Dynamic Molecular Processes React to Genetic, Environmental, and Gut Microbiota Interactions and Shape Our Health" has been written independently and without any other sources than cited. All necessary ethical approvals have been obtained in accordance with the European Directive 2010/63/EU law (on the use of clinical samples and on the Care and Use of laboratory animals, where applicable).

Luxembourg, 16/12/2024 Ziyun Zhou

# Acknowledgements

I would first like to express my deepest gratitude to my supervisor, **Prof. Evan G. Williams**. I still remember the excitement of the day I received his email offering me the PhD position, and even now, recalling that day brings back those feelings. The first few months were not easy for our group. As a newly established lab during the COVID-19 pandemic, it felt like everything was moving in slow motion. Nevertheless, I feel incredibly fortunate to be the first doctoral student in the GEM group, as it allowed me the unique opportunity to observe and learn how Evan was building his lab from the ground up. Throughout my PhD, Evan provided me with a great deal of positive feedback and supported me to the fullest extent in learning and exploring anything I was interested in. Reading Evan's previous publications has greatly motivated me to advance my own projects, as I aspire to produce work of equally high quality. Although, as his PhD student, my current work may not yet reach the level of excellence he achieved during his own doctoral studies, he will always remain a role model on my scientific journey!

Thank you to **Prof. Paul Wilmes** and **Prof. Gunnar Dittmar**, who, as members of my CET committee, graciously invited me to visit their laboratories at the early stage of my PhD. This opportunity allowed me to connect with their lab members and receive invaluable guidance in establishing and advancing my research project. I would like to extend my appreciation to **Laura Lebrun**, **Dr. Susheel Busi**, and **Dr. Rashi Halder** from Paul's group for their support. Laura provided essential help in conducting preliminary experiments related to microbiology, Susheel guided me in using the Kraken2 pipeline, and Rashi performed high-quality mRNA sequencing for our project. And **Dr. Marta Mendes** from the Gunnar group, who guided me through the protein sample digestion process for proteomics and demonstrated the robotic systems they use to assist with experiments.

Many thanks also go to **Prof. Xiaowei Zhan**, who warmly hosted me in his lab for six weeks. I am grateful for his openness to discuss my ideas at any time and for offering me opportunities to present my work to various audiences. I learned not only many new data analysis methods from him, but, more importantly, I also gained insights into how to quickly learn and apply different approaches. I would also like to thank the other members of my thesis jury, **Dr. Henry-Michel Cauchie** and **Prof. Thomas Sauter**, for participating in my defense and providing valuable feedback on my thesis. I am also grateful to **Prof. Pingsheng Liu** and **my former labmates**, with whom I collaborated on Chapter 5 included in this thesis. Working with them gave me my first real insight into how research is conducted.

I would like to express my heartfelt thanks to all current and former colleagues in the **Gene Expression & Metabolism group**. Special thanks to **Jack Haverty** and **Thamila Kerkour** for the joyful times exploring Luxembourg together. I am also deeply thankful to **Besma Boussoufa**, **Arianna Lamanna**, and **Emeline Pansart**, with whom I shared extended times of work. Traveling, having meals together, and discussing projects with them have all been an indispensable part of my PhD journey.

Finally, I owe special thanks to **my parents** and **Zishan** (my twin sister). Thank you for sharing in all my joys and sorrows and for always supporting me in pursuing what I truly want to do. And thanks also goes to **all the friends** who have supported and cared for me along the way, whether in China, Luxembourg, or other places.

Last but not least, I am thankful to the **LCSB Travel Grants** for supporting my participation in the RIKEN research conference, as well as the **Pelican Grants** for funding my exchange at UTSW.

**Ziyun Zhou | 周梓耘**
Belvaux, December 2024.

# Table of Contents

# List of Figures

# List of Tables

# Abbreviations

| | |
|---|---|
| AAV | Adeno-associated virus |
| ALD | Alcohol-associated liver disease |
| ApoA-IV | Apolipoprotein A-IV |
| B6 | C57BL/6J |
| BAs | Bile acids |
| BAT | Brown adipose tissue |
| BXD | A type of recombinant inbred mouse cross (C57BL/6J × DBA/2J) |
| cAMP | Cyclic AMP |
| CD | Chow diet (i.e. low in fat) |
| CREB | cAMP-response element binding protein |
| CVD | Cardiovascular diseases |
| D2 | DBA/2J |
| db/db | $Lepr^{db}$ |
| DNA | Deoxyribonucleic acid |
| DNN | Deep neural network |
| DNP | 2,4-dinitrophenol |
| DVC | Digital Ventilated Cage |
| EGC | Enteric glial cell |
| FMT | Fecal microbiota transplants |
| GNM | Gaussian network model |
| GPCRs | G protein-coupled receptors |
| GRP | Genetic reference population |

| | |
|---|---|
| GSIS | Glucose-stimulated insulin secretion |
| GVB | Gut vascular barrier |
| GWAS | Genome-wide association study |
| GXE | Gene-Environment Interactions |
| HCM | Home cage monitoring |
| HF | High fat diet |
| IBD | Inflammatory bowel disease |
| IgA | Immunoglobulin A |
| ILC3s | Group 3 innate lymphoid cells |
| IPGTT | Intraperitoneal glucose tolerance test |
| Logistic | Logistic regression |
| LRP1 | Lipoprotein receptor-related protein 1 |
| LSG | Laparoscopic sleeve gastrectomy |
| MG | Metagenome |
| mRNA | Messenger ribonucleic acid |
| MS | Mass spectrometry |
| MT | Metatranscriptome |
| NAFLD | Non-alcoholic fatty liver disease |
| NASH | Non-alcoholic steatohepatitis |
| ob/ob | $Lep^{ob}$ |
| QTL | Quantitative trait locus |
| RER | Respiratory exchange ratio |
| RF | Random forest |
| RI | Recombinant inbred |
| RYGB | Roux-en-Y gastric bypass |
| SCFAs | Short-chain fatty acids |
| SOPF | Specific and Opportunistic Pathogen Free |
| SPF | Specific Pathogen Free |

| | |
|---|---|
| STZ | Streptozotocin |
| T1D | Type 1 diabetes |
| T2D | Type 2 diabetes |
| $VO_2$ | Oxygen consumption |
| $VCO_2$ | Carbon dioxide production |
| WGCNA | Weighted correlation network analysis |
| WGS | whole genome sequencing |
| XGB | XGBoost |

# Abstract

Metabolic diseases, such as obesity and type 2 diabetes, are rapidly growing global health concerns. These conditions are influenced by a complex interplay of genetic predisposition, environmental factors, and gut microbiota, which together regulate host metabolism. Understanding these multifactorial influences is critical for identifying new therapeutic strategies for metabolic disorders. This thesis investigates the relationships between host genotype, environmental factors such as diet and exercise, gut microbiota, and specific proteins involved in metabolic regulation. The study is divided into three distinct but interconnected parts, each exploring a different aspect of this intricate system.

The first part of the thesis focuses on the cecum transcriptome (mRNA), metagenome (MG), and metatranscriptome (MT) of mice with varying genotypes, diets, and ages. The aim is to explore how these factors shape the gut microbiota and, consequently, influence host metabolism. Gut microbiota are known to play a crucial role in modulating various physiological processes, including nutrient absorption, immune function, and energy homeostasis. Variations in the composition of gut microbiota can lead to alterations in metabolic profiles, impacting susceptibility to obesity, insulin resistance, and other metabolic disorders. By analyzing the cecal content of mice across different conditions, we aim to identify specific microbial and genetic interactions that contribute to metabolic changes. Using machine learning models, we demonstrate that microbial species profiles can robustly predict host phenotypes such as body weight and age. Furthermore, we show that integrating microbiome and host liver proteomic features enhances prediction accuracy, highlighting the value of multi-omics integration for biomarker discovery.

In the second part of the study, we explore the effects of long-term voluntary physical activity on host metabolism, with a particular focus on glucose tolerance. Increased physical activity is well-known for its beneficial effects on metabolic health, including improved insulin sensitivity and glucose homeostasis. We observed a significant improvement in glucose tolerance in the exercise group compared to the sedentary group, regardless of genotype. Metagenomic analysis of the cecal content revealed distinct microbial profiles in the exercise group, with an increase in microbial diversity and the presence of specific bacterial taxa associated with enhanced metabolic function. These findings suggest that exercise not only improves glucose tolerance but also induces favorable shifts in gut microbiota composition, which may, in turn, contribute to its metabolic benefits. The identification of bacterial species linked to exercise offers promising avenues for future research

into microbiome-targeted therapies for metabolic diseases.

The third part of this thesis examines the role of the protein Apolipoprotein A-IV (apoA-IV) in glucose metabolism and its potential therapeutic application. Our findings indicate that apoA-IV significantly improves glucose tolerance in both db/db and ob/ob mice, supporting its role in enhancing insulin sensitivity. Additionally, through computational predictions, we identified functional peptides of apoA-IV that may be responsible for its glucose-regulatory effects. We synthesized these peptides and administered them to the same mouse models, observing further improvements in glucose tolerance. Furthermore, functional experiments in diabetic rodent models demonstrated that both full-length apoA-IV and its bioinformatically predicted peptide fragment (T55–121) significantly enhanced glucose tolerance and insulin secretion, partially through the G$\alpha$s-coupled GPCR/-cAMP signaling pathway. These results suggest that the glucose-regulating function of apoA-IV may be mediated, at least in part, by specific peptide fragments, which hold potential as therapeutic agents for treating metabolic diseases.

Together, these findings contribute to a growing body of evidence that host-microbiota interactions are essential in regulating metabolic health. They also highlight the therapeutic potential of targeting both microbiota and host proteins to improve glucose regulation and combat metabolic diseases. The insights gained from this research may inform the development of new strategies for preventing and treating conditions such as obesity and type 2 diabetes, offering hope for better management of these health challenges.

# Chapter 1

# Introduction

Contemporary human societies are increasingly grappling with chronic diseases induced by metabolic disorders, such as obesity, diabetes, and cardiovascular diseases (CVD) [1]. These metabolism-related diseases significantly impact human longevity and quality of life, and they have massively and globally increased in prevalence over the last thirty years. Thus, there has been an increasing scientific focus on understanding their molecular etiology and finding early biomarkers, prophylactics, and treatments. During the traditional Paleolithic period, individuals generally maintained good health into old age, with the primary causes of mortality being infectious diseases rather than metabolic disorders [2]. However, since the Paleolithic era, and particularly since industrialization and improvements in sanitation, vaccination, and antibiotics, human living environments and lifestyles have undergone rapid change. In contrast, human genomic evolution has proceeded at a significantly slower pace [3]. Additionally, in general, natural selection during evolution has not developed strategies to counteract excessive nutrition and energy but has instead prioritized phenotypes that ensure survival up to reproductive age, but not old age [4]. Thus, genetic adaptation is insufficient to effectively respond to dietary changes (e.g. processed and refined foods), reduced daily physical activity (e.g. a sedentary lifestyle), chronic stress and other factors, all of which have serious health consequences [5], but predominantly later in life (i.e. after 50 years of age).

For the past 140 years, clinicians and scientists have understood and mathematically quantified the relative contributions of environmental influences to genetic predisposition on the development of disease [6]. The relative contribution of the genome and the environment to an individual characteristic is called heritability in scientific terms, or "nature versus nurture" for the common lay term. Scientists have thus long understood that many traits are driven primarily by genetic factors, such as individuals' heights assuming proper childhood nutrition. More recently it has been recognized that many environmentally-influenced diseases such as type 2 diabetes (T2D) are also widely affected by genetic predisposition, and genetic factors explain up to half of all variation in T2D susceptibility [7]. Since the first draft of the human genome was assembled in 2001, population scientists have made massive inroads regarding our collective understanding of which specific variants in DNA

cause molecular effects which in turn affect overall physical characteristics. As of today (December 2024), just over 510,000 trait-associated variants have been identified and reported to the Human Gene Mutation Database [8]. However, despite this tremendous gain in genetic knowledge since the first human genome assembly in 2001, scientists remain generally unable to predict the genetic and cellular mechanisms by which a DNA variant lead to a trait outcome. For highly-heritable traits where many DNA variants have been causally identified as driving a specific outcome, like height, it also remains prohibitively difficult to predict the gene-gene *interactions* of these variants, as some variants may act only in epistasis, while others may be simply additive. Furthermore, mathematical models in genetics tend to not account for large-scale epistasis; nearly 12,000 DNA variants have been identified which causally affect height [9], but clearly a theoretical individual with all of the "taller height" DNA variants would not be dozens of meters tall.

To address the challenges of identifying the intermediate molecular mechanisms connecting a causal DNA variant to an eventual heritable phenotype, over the past 10 years research groups such as ours have increasingly taken a multi-omics approach to understanding genetic mechanisms. For a DNA variant to cause a systematic effect on an individual, whether increased height or increased propensity to develop diabetes, there must be specific cellular effects that arise via the transcriptome, proteome, and metabolome. However, unlike the DNA which is essentially constant across an individual's lifetime and all of their cells, these intermediate molecular "omics" layers vary widely across time, tissue, and cell type. Thus, identifying the molecular networks leading from DNA to phenotype requires first identifying large systemic molecular shifts, and then narrowing down their timing of activation and proximal causes. In the case of my studies during my PhD, I have followed three different cohorts of mice over long periods, including up to their entire natural lifespan. The traits studied - aging and obesity (Chapter 3), exercise and diabetes (Chapter 4), and the mechanisms of glucose control (Chapter 5) were studied in different environmental conditions (e.g. varying diets and exercise regimens), ages, and genetic backgrounds (more than 30 strains studied in Chapter 3). By incorporating genetic and environmental complexity in the studies of my doctoral thesis, we aimed to model human metabolic complexity in mice, while retaining the ability to reproduce key findings through the use of inbred strains.

The primary development during my thesis was on the large-scale whole genome sequencing of the cecal contents from more than 250 mice at different ages (Chapter 3), to identify and validate relationships between the microbiome and metabolic outcomes. That is, which taxa of the microbiome are predictive of future health outcomes, and which are simply a diagnostic that the host's physiology has already started to change? The rationale for focusing on the microbiome in my thesis is to build off a previous large multi-omics study of the transcriptome, proteome, and metabolome that was performed in 300 of the same exact mice and published at the start of my PhD [10]. By building off this existing dataset, I was able to extend analysis of the cellular mechanisms by which the gut microbiome "coordinates" with the host's own tissues. In addition to the previous analysis of the liver [10] and my microbiome analysis, I also sequenced the mouse's own host cecal transcriptome in 100 of the same individuals, thus creating a dataset with proximal tissues to the microbiome (i.e.

cecal transcriptome to the cecal microbiome) and more distal components of metabolism (i.e. the liver multi-omics data). In addition to this study which used prior biobanked samples (the cohort was finished in 2019), I also launched a new mouse phenotyping study with Prof. Williams on the long-term effects of voluntary exercise in conjunction with voluntary *ad libitum* high fat feeding Chapter 4. This study, for which the phenotyping arm completed on Nov 14, 2024, is still ongoing for molecular analysis, but the main phenotypes were reproducible: exercise has a substantial effect on mitigating poor glucose response and fatty liver disease, but due to increased food intake on animals which exercise regularly, obesity rates in fact remain constant regardless of sustained exercise. In this cohort, I also measured the microbiome in one of the four cohorts as a pilot to cross-check results from my main study, and to ensure that the study design works properly for the future. This study is nearly complete, and the final cohort will be completed together with a PhD student, Arianna Lamanna, in November 2025 and who I trained on this project when she joined the lab in 2023. Finally, I had several collaborative projects during my doctoral thesis, of which one has been recently published (Chapter 5).

In this introduction, I will provide further details on each of the main experimental and analytical components of my thesis, as most of these elements are shared across the three primary experimental chapters. An abbreviated introduction is also written for each subsequent chapter.

## 1.1 Factors Influencing Host Metabolic Diseases

Metabolism is a complex interplay of genetic and environmental factors that govern physiological and biochemical processes crucial to maintaining energy balance, growth, and overall health. Host metabolic function is intricately shaped by various influences, including genetic predispositions and environmental exposures that can enhance or impede metabolic homeostasis. Studying these factors is essential for understanding the pathways leading to metabolic diseases like diabetes and obesity.

Metabolic diseases are often thought of by existing in two categories. First, "textbook example" diseases that result entirely from genetic mutations like cystic fibrosis or musculuar dystrophy, and second those that are more sporadically acquired such as type II diabetes (T2D) or non-alcoholic fatty liver disease (NAFLD). This second category of disease results from a mixture of lifestyle factors such as diet and physical activity, environmental factors like pollution and agricultural chemicals. However, this classification is often misleading or improperly understood by the wider community; for instance, people with T2D are commonly considered to have the capacity to take the disease "within their own control" by simply "eating less and exercising" while those with type I diabetes are commonly considered to be unlucky individuals with no avenues for addressing their illness beyond judicious use of insulin pumps and glucose strips. This is a common misconception: T2D actually has similar heritability to type I diabetes (T1D), of up to 50-70% depending on the study [7]. That is, more than half of variation in susceptibility to T2D across human populations is due to inherited factors, not environmental ones. That does not mean T2D is an immutable, untreatable genetic disease, but rather than its severity and how difficult it is to treat will vary dramatically

depending on one's own inherited predispositions. In my studies, the primary disease of study is T2D, but as this disease does not exist in isolation, all three papers also consider related diseases such as NAFLD and obesity. The primary goal is to understand which and how genetic risk factors across a large and diverse population of mice interact with environmental/lifestyle factors, particularly diet and exercise. The following studies are entirely performed in mice, due to the increased ease with which we can study and reproduce gene-environment interactions, and because of the timelines required to study the disease's development in humans (at least ten years) are compressed in the mouse model (around six to ten months is sufficient).

The ultimate goal is to translate these findings to human cohorts. Between 1980 and 2014, the global prevalence of diabetes among adults nearly doubled [11]. Diabetes is a chronic metabolic disease caused by defects in $\beta$-cells [12]. In a healthy state, an increase in blood glucose stimulates $\beta$-cells to produce insulin, thereby maintaining blood glucose levels within a stable range [13]. By far the two most common types of diabetes are T1D and T2D, with ~90% of patients diagnosed with T2D and ~5-10% with T1D [14]. In T1D, the immune system mistakenly attacks healthy pancreatic $\beta$-cells, leading to an inability of the body to produce insulin and resulting a series of complications [15] (Fig. 1.1). Compared to T1D, which is caused by autoimmune disorder and often has sudden onset, T2D is primarily driven by environmental factors, and thus, it typically has a later age of onset, and onset is usually gradual. In patients with T2D, $\beta$-cell dysfunction leads to increased insulin secretion to compensate, but this increase eventually leads to insulin resistance and progressive $\beta$-cell failure [16]. Current research generally suggests that a high-sugar, high-fat diet and a sedentary lifestyle are the most influential factors in the development of T2D that can be modified practically (i.e. in contrast to e.g. gene editing of risk factors) [17].

Over the past decades, numerous drugs and treatments have come to market which have sought to address T2D and obesity. These drugs have often had significant success, but often also have resulted in severe side effects due to the many challenges inherent in modifying a core energy metabolism pathway. In the early 1900s, one of the first drugs was created and used for inducing weight loss - 2,4-dinitrophenol (DNP). When ingested also increases mitochondrial uncoupling, thus raising the body's temperature and substantially increasing energy usage. However, DNP has a narrow therapeutic index and numerous accidental overdoses occurred and the drug was withdrawn from most markets by 1940 - not to mention that it is also a chemical explosive with military uses. Due to these dangers, it was withdrawn by the 1930s for use in obesity. By 1960, doctors began using mixtures of amphetamine and methamphetamine for weight loss, most notably under the brand name Obetrol. Unlike DNP which increased energy expenditure, the various amphetamines will suppress appetite and thus decrease energy intake, with much lower risk of acute toxicity. While these various formulations were certainly effective for weight loss, the former (DNP) resulted in severe, acute, direct health risks, while the latter (amphetamine) resulted in high risk of addiction and eventual secondary health and neurological risks. Thus, for the following decades numerous other treatments for obesity and T2D have been developed, including surgical approaches such as gastric bypass surgery, which greatly reduces appetite in most patients. Most recently, drugs

based on glucagon-like peptide-1 (GLP1) have come to market, namely tirzepatide (Mounjaro) and semaglutide (Ozempic) appear to have had tremendous success in suppressing appetite without going through the central nervous system (as with amphetamine) or highly invasive surgery (as with gastric bypass). While these two new treatments, and countless others shortly behind, are providing what looks to be a revolutionary therapeutic for T2D with a limited side effect profile and extremely wide therapeutic index, these remain expensive (roughly $500/month per person) and they are only used as treatments for disease once it has occurred, rather than as financially-efficient and medically-effective prophylactics. In 2021, ~ 75% of American adults are now obese or overweight, and most of the rest of the wealthy world is not far behind [18]. Consequently, rather than treating billions of people with semaglutide at the cost of hundreds of dollars per person per month, it would be more effective to understand when the disease begins to develop, and mitigating this stage long before clinical manifestation. In my thesis, I am largely focused on the molecular and chronological timeline of T2D development, and how we can measure, understand, and avoid its further progression.

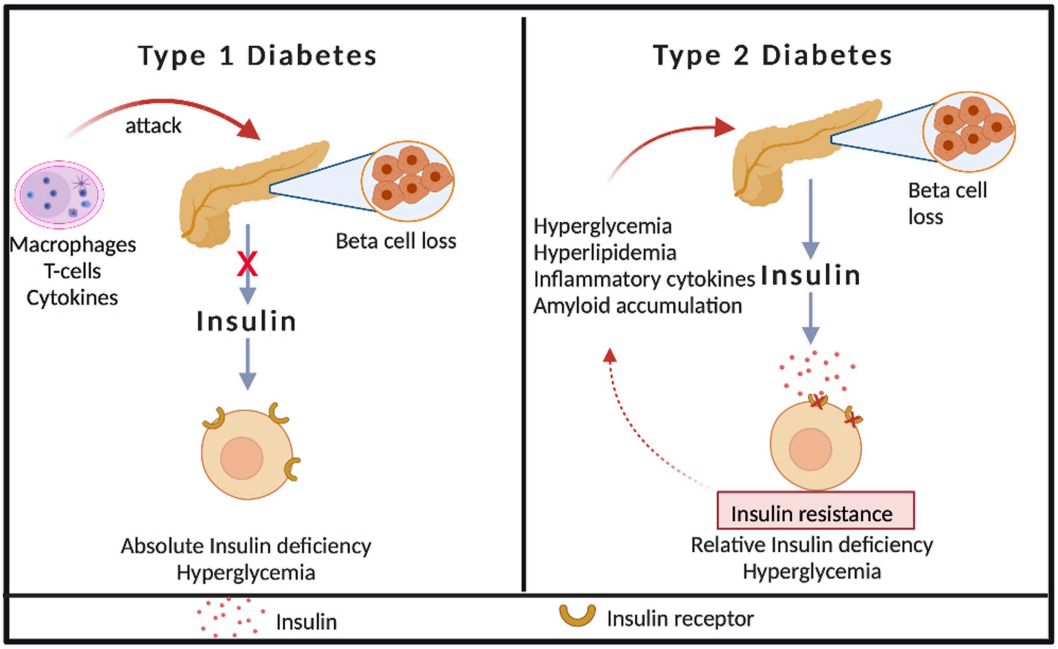


**Figure 1.1** Pathogenesis of Two Types of Diabetes [12]. In T1D, immune cells (e.g., T-cells and macrophages) and their cytokines destroy pancreatic $\beta$-cells, causing an absolute insulin deficiency and hyperglycemia. In T2D, pancreatic $\beta$-cells are impaired by factors like hyperglycemia, hyperlipidemia, etc. While insulin is produced, its level is insufficient to counteract insulin resistance, leading to a relative insulin deficiency and hyperglycemia.

### 1.1.1 The Impact of Host Genetics

Genetic makeup strongly influences host metabolism, contributing to variations in energy expenditure, insulin sensitivity, and lipid metabolism. Although dietary intake and lifestyle factors like exercise have a clear effect on the development of obesity diseases like T2D, genetic background plays a huge role and makes this a vastly uneven playing field across different people. The first "monogenic" cause of obesity in mice, the ob/ob mouse, was accidentally generated due to a spontaneous mutation that occurred in the 1940s. These mice were found to be deficient in the gene called leptin, which affects satiety. Essentially, this genetic variant causes these mice to constantly feel hunger. Such variants are extremely rare in clinical studies, but they do occur and cause similar effects in a few human patients with the same "monogenic" cause of obesity. While such individuals can, in theory or in a laboratory, keep their weight under control with precisely-managed diets, outside of laboratory conditions it is impractical (and indeed nearly impossible) to tell a patient whose hormones are constantly signalling a feeling of extreme hunger and a sense of starvation to "simply stop eating so much." These cases are rare and extreme, but they clearly indicate that the sense of satiety and control over food intake vs. energy expenditure is not a simple case of "mental fortitude" as is often considered as such by the lay community (and even some scientists), but that there is a large range in how addressable obesity and metabolic diseases are with lifestyle changes alone.

In my research, I am focused on more common genetic profiles that lead to propensity to obesity and T2D, particularly in conjunction with fatty diets and exercise. Using such complex multivariate study designs, I have modeled, measured, and analyzed the myriad interacting and changing pathways in different enzymatic and hormonal controls, rather than single gene variants in leptin or the leptin receptor. To do this, I am building off previous work of my PhD supervisor in the BXD mouse population whose previous studies profiled more than 50 different types of diversely-generated and non-genetically modified inbred mice to see their natural predisposition to obesity and T2D in various environments [19]. Specific genetic profiles, such as those observed in BXD mouse models, allow researchers to investigate the link between genotype and metabolic traits. This approach aids in identifying genetic biomarkers and pathways associated with metabolic diseases.

Family and twin studies indicate that up to 80% of the variance in body mass index (BMI) is attributable to genetic factors. Genetic factors explain approximately 50% of T2D risk, and the heritability rates of metabolic syndrome are likewise approximately 10%–30%[20]. In addition to genetic factors, metabolic syndrome is substantially influenced by environmental factors. In the BXDs, this heritability was around 70% within diet, or around 30% across diet [19]. However, like with the vast majority of family and twin studies on obesity, there was no single genetic cause identified that explained more than a tiny fraction of that 30-70% of heritability: the heritable variation in body weight must be driven by a complex interaction of many variants of relatively small impact. However, this means that both genome-wide association studies (GWAS) in those human populations and the conceptually-equivalent quantitative trait locus (QTL) mapping in mouse populations have reported relatively limited genetic understanding of obesity and T2D - even though these traits are highly genetically driven! If we have a greater understanding of potential gene-nutrient interac-

tions, then it may be possible to manipulate diet in such a way to minimize the metabolic risk of obesity, to attenuate insulin resistance and development of cardiometabolic disease. To do so, we can now quantify intermediate molecular datasets between DNA and phenotype, i.e. the transcriptome, proteome, metabolome, and - the main focus of my PhD research - the microbiome, which plays a critical and as-of-yet understudied role in metabolism.

### 1.1.2 The Role of Environmental Factors

Environmental factors, such as diet, physical activity, and microbiota exposure, significantly impact metabolic health. Dietary intake, particularly with high-fat and high-sugar, is a major contributor to metabolic diseases, both directly by affecting gene expression and more indirectly by altering microbiome composition and function. Other environmental influences, including stress and lifestyle, also affect the host's susceptibility to metabolic disorders. Despite the long-known, common, and scientifically correct knowledge that diet and exercise affect metabolism, there is a vast discrepancy in the knowledge of *which* diets and *how much* exercise are actually best suited for a given individual. Consequently, dozens (if not hundreds) of scientific, semi-scientific, and pseudo-scientific dietary regimens have been created: low carbohydrate/high fat or protein diets like the "keto-diet" and "paleo-diet", low sugar diets, diets with various focus on different compositions of amino acids, calorie counting, intermittent fasting, and so on. In addition to these direct dietary recommendations, there are numerous supplements available that supposedly affect one's metabolism and health. Many of these diets and supplements likely work as stated - at least under highly-specific laboratory conditions. However, scientific support for many dietary supplements is scant, even for supplements such as rapamycin and resveratrol that have long been known in animal studies to improve metabolic health [21, 22]. Several fundamental challenges stymie the generalization of these findings to "real world" applications. For example, in "calorie counting", the amount of caloric energy two people will get from the exact same quantity of the exact food can vary by multiples: the microbiome, intestinal absorption, as well as genetic factors, will influence *for a specific individual* how much energy is actually in their food. For instance, as an extreme example, mice and rats that are "germ-free" (i.e. born and raised in entirely sterile conditions, and which have no gut microbiome) have been known for decades to require higher food intake (20-30%) for their body weight, and to excrete more water and nutrients [23]. Similarly, in a widely known example, ruminants that were germ-free would quickly starve to death, as it is the bacteria in their stomachs that allows them to metabolize grass; it is not an innate ability in these species' own DNA.

Consequently, scientists today have almost *too many* environmental and lifestyle factors that are supposedly known to be helpful and can be changed to influence health and metabolic disease propensity and development, yet which are poorly substantiated in basic scientific literature, let alone clinical studies. The main challenge is to be certain that a given intervention is helpful for a particular individual. This is an issue not only for motivation to maintain a specific lifestyle change, but also because dietary changes typically take months, often years, before a clear and sustained effect is seen on traits like body weight, fasted glucose levels, blood pressure, or cholesterol levels.

While it is possible to examine tissues like the liver and muscle to see whether the molecular trajectory of a disease is improving in response to an environmental change [24, 25], in clinical medicine it is both impractical and unethical to take tissue biopsies for such information. Clearly, non-invasive readouts are necessary, such as from the feces or urine, or at least a minimally-invasive readout from blood. In my thesis, I focus on the analysis of fecal samples.

## 1.2 Mouse as a Model Organism for Studying Metabolic Diseases

Metabolic syndrome is a multifactorial disease with a high and continually-increasing prevalence nowadays across the globe. Therefore, establishing suitable experimental animal models that can simulate human metabolic syndrome is crucial for addressing issues related to this cluster of metabolic diseases [26]. Additionally, the long durations required for studying these diseases in humans means that refined hypotheses are necessary, and these hypotheses can often be generated in mice. Other practical benefits besides the rapid time of disease onset include a rapid breeding rate and minimal space requirements for housing, meaning that biological replicates can be easily tested to differentiate environmental versus genetic causes of disease. Additionally, in mice changes in genotype, diet, histology, function, biochemistry, and morphology can be easily monitored and analyzed, which is challenging to achieve in human populations [26]. All these factors have led to rodents, particularly mice, have been the most widely used animal models in metabolic disorder-related research over the past four decades [27, 28]. However, the metabolic characteristics of mouse models in some cases are quite different from humans, either in the genetic pathways such as large variations in cholesterol metabolism between mice and humans [29], or in specific hypotheses generated, e.g. only around 15% of the genera in the gut microbiome are the same in mice and humans [30]. Specifically to the wider scientific scope of my thesis work, mice can be raised in gnotobiotic facilities unlike humans, where in addition to being simply germ-free, their digestive tracts can be seeded with specific taxa of bacteria, and thus cause-and-effect can be identified, unlike human microbiome studies which are almost entirely observational. While my project did not investigate germ-free or gnotobiotic mice, we do have such a facility on campus and my findings in chapters 3 and 4 are refined with the eventual goal in mind of passing the project along to a future scientist in the research group who can test these hypotheses mechanistically, and perhaps to take some of these microbiome findings for my own future experimental career.

### 1.2.1 The BXD Mouse Model

The C57BL/6J (B6) mouse strain is the most commonly used mouse model in research; however, it is a single genetic background and thus effectively only simulates one individual, and it does not effectively simulate any real-world mammalian cohort. To its benefit however, because it is inbred, it allows much more reproducible experiments on gene-environment interaction, even though in some cases this may oversimplify how genetic diversity works in normal populations. Extensive genetic variation among individuals is a characteristic of human populations, and metabolic diseases

like T2D are rarely associated with individual gene variants. This aspect may limit the use of single inbred strains in studying metabolic syndrome [31]. As we have established previously in this chapter, metabolic diseases share some common factors across all individuals of a species (or even across mammals), but particular environmental recommendations need to be tailored or adjusted to each individual, both based on their genetic code and to other interacting environmental situations. Moreover, it is challenging to obtain ideal results in studying complex traits influenced by multiple genetic and environmental factors using a single inbred strain such as B6 [32]. To address these issues and obtain a multi-strain family, Benjamin A. Taylor began producing the BXD family in 1971 by multiple different crosses of female B6 with male DBA/2J (D2) mice, and currently, there are ~ 150 strains available [33] (Fig. 1.2). Furthermore, the BXD mouse model has been used in research on various diseases, including Alzheimer's disease [34], neurotoxicity induced by ethanol and cocaine [35], and developmental coordination disorder [36].

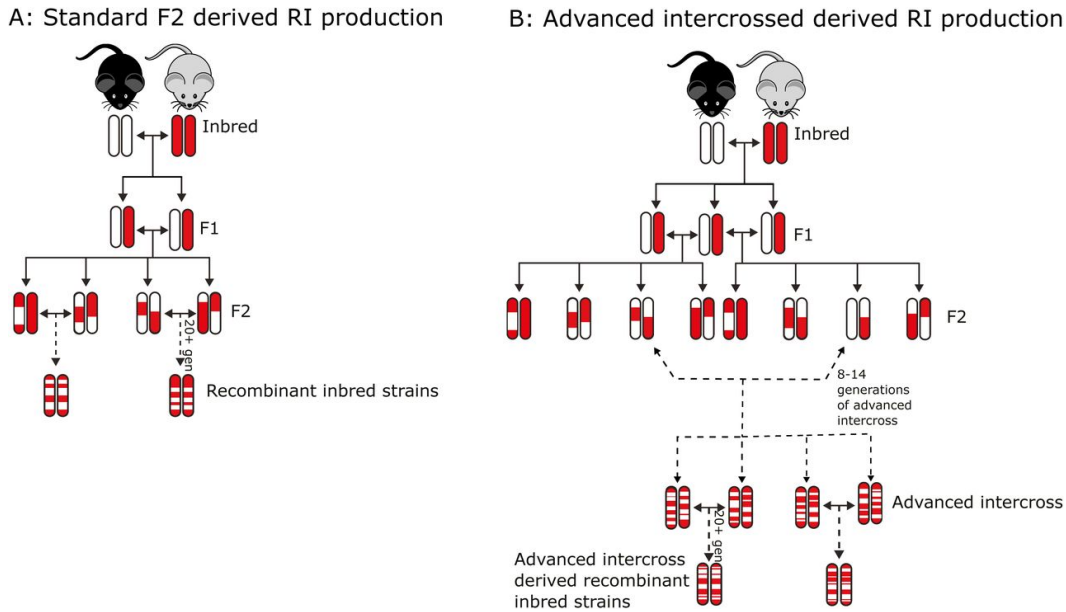


**Figure 1.2** Production process of BXD RI lines [33]. Standard F2 RI lines produced ~ 75 strains, while advanced intercross RI lines generated ~ 65 strains.

The interactions between gut microbiota and the host are complex and also influenced by multiple factors; therefore, using the BXD recombinant inbred (RI) mouse model allows for a better consideration of the impact of genetic background. For example, a recent study used dozens of strains from the BXD mouse family to examine the relationship between gut microbiota and host bile acids (BAs), revealing that the impact of a High fat dite (HF) on the gut microbiome varies depending on the strain [37]. The BXDs thus solve the primary limitation of using B6 mice: genetic diversity is reincorporated back into the population. Furthermore, it solves the main issue of using "normal" populations that are genetically diverse, where it is impossible to perfectly reproduce gene-environment interactions in independent follow-up cohorts. While each individual BXD strain is

itself inbred, there are further breeding schemes, particularly of crossing two different inbred strains with each other, to obtain individuals that are heterozygotic and yet also have perfectly reproducible genetic backgrounds (i.e. the first generation F1 cross of B6 with D2 will always yield a mouse with one full copy of every chromosome from each parent). While such "RI-F1s" of the BXDs were not used in my study, they are currently being generated at the same collaborating laboratory in the United States where the aging colony (Chapter 3) was performed. Thus, within the next 3 to 5 years, the hypotheses I have generated in the BXD aging colony will be cross-validated in the RI-F1 BXD aging colony.

### 1.2.2 The db/db and ob/ob Mouse Models

About one-third of the global population is affected by obesity, which significantly increases their risk of developing diabetes [38, 39]. Therefore, there is an urgent need to develop safe and effective therapeutic strategies [38]. Many of the current research findings related to obesity and diabetes have been derived from mouse models, particularly the db/db and ob/ob mouse models. As mentioned previously, ob/ob mice were first described in 1949 and have a mutation in the leptin gene (*Lep*) [40, 41]. Compared to control littermates, leptin mutant mice exhibit significant rapid weight gain [40]. They are typically obese throughout their lifespan and have mild insulin resistance, but they do not necessarily develop diabetes [42]. Subsequently, in 1966, the db/db mouse was first identified by Jackson Labs [43]. Similar to the ob/ob mouse, the db/db mouse also develops obesity due to impaired leptin signaling. However, the distinction lies in that the db/db mouse has a mutation in the gene encoding the leptin *receptor* [42]. This leads to the db/db mouse typically suffering from both obesity and diabetes with a loss of $\beta$-cells [42, 44]. There have also been studies that use these two mouse models in gut microbiota-related research. For example, one study found that changes in the microbial composition of ob/ob and db/db mice exhibited similar patterns to changes in blood glucose levels [45]. These models have been widely successful in proving various mechanisms of action between the gut-brain axis, hormonal control of food intake, and the effect of various diets on obesity. For instance, in both ob/ob and db/db mice, mice will get obese and diabetic even on a standard "healthy" control diet, so long as they are able to eat at will (*ad libitum*) [46]. This research proved that obesity and T2D are not simply an environmental consequence of "good" or "bad" lifestyle choices. Conversely, the specific genetic mechanisms in ob/ob and db/db mice are so rare in humans that these specific causal factors do not widely translate to a clinical setting. While the db/db and ob/ob mouse models were not the primary focus of our study, they were employed due to their unique genetic backgrounds in our investigation of the diabetes-related protein apoA-IV (Chapter 5). These models played a crucial role in validating our hypotheses.

### 1.2.3 The Germ-free Mouse Model

Beyond the use of mouse models with specific genetic backgrounds for metabolic studies, germ-free mice have also been widely employed as a common model. This is because the gut microbiota play a

crucial role in maintaining host metabolic homeostasis, and collectively the gut microbiome is often regarded as an organ-like entity of the host[47]. Consequently, due to the critical connection between the gut microbiota and host metabolism, germ-free mouse models have been widely utilized to investigate the relationship between the microbiome and metabolic diseases. As early as 1946, germ-free rat colonies were successfully established [48], and germ-free mice also emerged in 1959 [49]. The primary use of germ-free mouse models has been to clarify the role of gut microbiota in host functions or diseases by minimizing confounding variables and enabling more interpretable results [50]. And these studies have typically combined the use of germ-free mice and fecal microbiome transplantation (FMT) approach. For instance, FMT from monozygotic twin pairs discordant for obesity into germ-free mice has demonstrated that adiposity and related metabolic phenotypes are transmissible, underscoring the causal role of the gut microbiota in host metabolism [51]. While the germ-free mouse model has not yet been utilized in this thesis, hypotheses related to gut microbiota have been derived from both the BXD aging colony study (Chapter 3) and the exercise study (Chapter 4). Thus, in subsequent validations, germ-free mouse will play a crucial role.

## 1.3 Gut Microbiota and Host Metabolic Interaction

The composition of the gut microbiota co-develops with the host from birth. In humans, about 100 species of microbiota populate the infant colon at birth if naturally delivered and rapidly increase in number, becoming relatively stable in adulthood, yet undergoing another shift in composition during old age [52] (Fig. 1.3). However, not all gut microbial species have a positive impact on the host. Gut microbiota are often categorized as beneficial bacteria and harmful bacteria. However, these broad categorizations are widely considered incorrect by the microbiome research community today, as it is rarely that a bacteria is universally *bad*, but rather that its abundance may be excessively high or low, or that it may have negative interactions or ratios with other types of bacteria. Those bacteria, their abundances and ratios, that are considered beneficial bacteria are those with mutually beneficial relationships with their host, contributing positively to the host's fitness. In contrast, some bacteria are strictly harmful, e.g. bacteria which are parasitic and may exert negative effects on the host [53]. However, even for such species, bacteria once considered "harmful" are often later noted to have specific effects either in time, depending on the composition of the rest of the microbiome, or response to particular environmental signatures that may turn a "bad" bacteria into a "good" bacteria. For example, a observation found that stinkbugs, a type of insect, produces a "symbiont capsule" during oviposition, which vertically transmits specific gut bacteria beneficial to host health to their offspring. If symbionts are removed, it leads to stunted growth or even death in insects [53]. Conversely, some bacteria do remain seen as strictly harmful; studies have found that *Clostridium difficile* colonizes the human large intestine and releases two protein exotoxins, TcdA and TcdB, which may lead to colitis [54]. Studying and identifying gut bacteria that are beneficial or harmful to the host is essential for the detection and treatment of diseases. Moreover, the last decade of research into the microbiome has shown that it is not simply the presence or absence

of a bacterial taxa that can be determined as good or bad, but that the entire holistic view of the microbiome and a particular taxon's place within it must be understood.

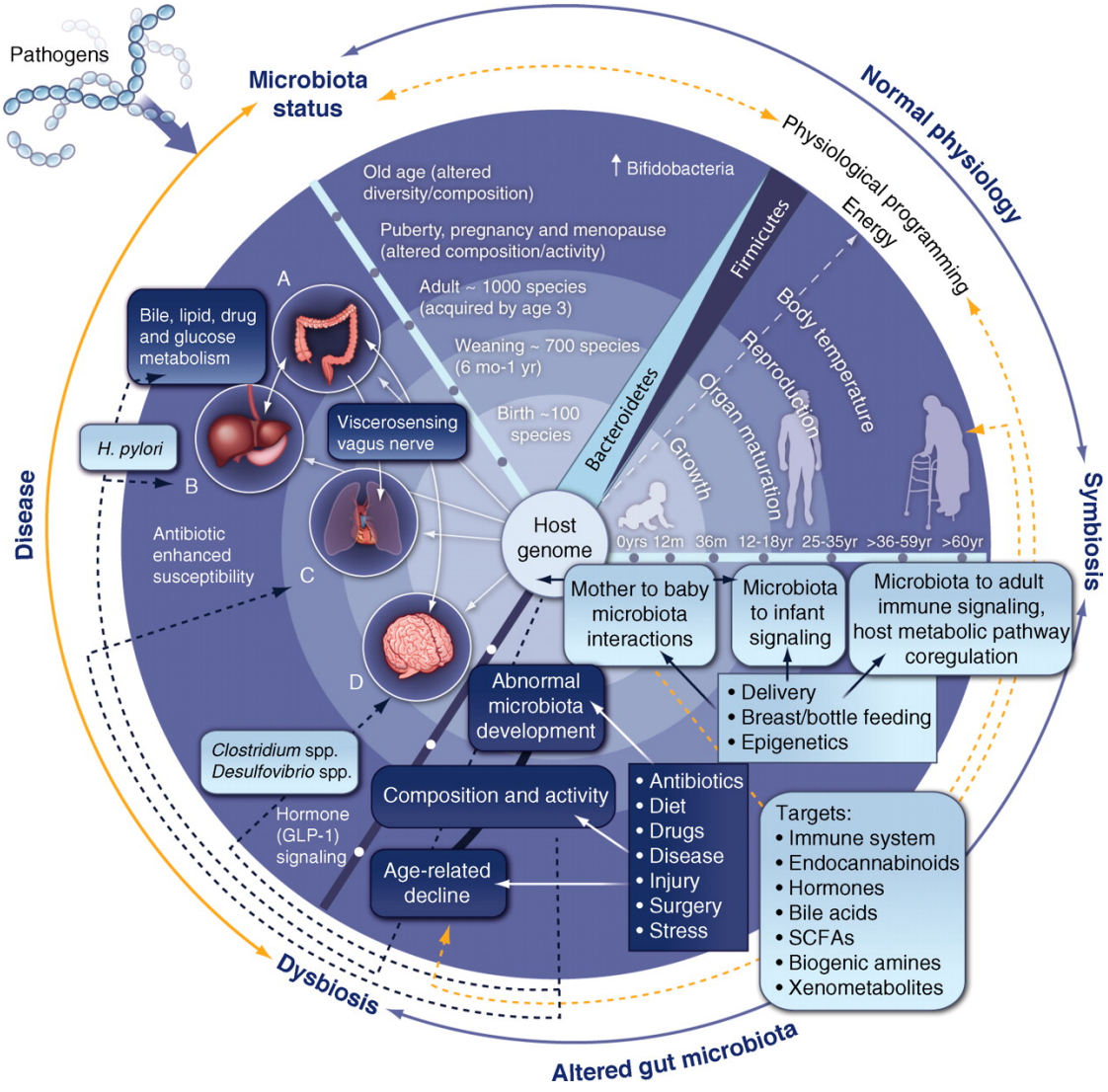


**Figure 1.3** Correlation of Gut Microbiota with Age and Disease [52]. Effects of gut microbiota on host health persist throughout the lifespan. In humans, over a hundred microbial species colonize the infant colon at birth. The number of gut microbes increases with age and eventually stabilizes, but their composition can change in older age. The gut microbiota influences host metabolism through interactions with organs such as the intestine, liver, brain, and lungs, and may become dysbiosis due to environmental factors or disease.

Similar to metabolism, the composition of gut microbiota also exhibits a complex interplay with host genetics and various environmental factors [52] (Fig. 1.3). The gut microbiota primarily participates in various metabolic pathways, such as signaling, and immune-inflammatory responses, through physiologically connected axes with different host tissues, including the gut, liver, muscle,

and brain [52]. The most commonly studied axes currently are the gut-liver axis and the gut-brain axis, and research on other signaling mechanisms between the gut and the rest of the organism is also increasingly growing. For instance, FMT from healthy donors to individuals with a wide range of metabolic diseases have been tried in clinical settings, ranging from *C. difficile* infections [55] to obesity [56]. However, the results of such attempts, while typically well-exceeding placebo, remain intermittent and difficult to predict [57]. Extensive further research is necessary to improve this to standardize the medical usage of FMT. For instance what should be the ideal composition of the donor's FMT for the target individual, and which parameters in the recipient can help predict whether they are likely to benefit from the process? While in my thesis I do not implement FMT in the mice, the goal of my primary research projects (Chapter 3 and Chapter 4) are to determine which bacteria can be selected for FMT in gnotobiotic mice to influence the development and reverse the occurrence of diet-and-genome-induced obesity, T2D, and NAFLD.

In a healthy state, the host has multiple barriers to prevent microbiota from translocating through the gut into other parts of the body [58] (Fig. 1.4). First, the mucus layer serves as the first line of defense in the gut, protecting epithelial cells from toxins and microbiota while also providing a niche for microbial colonization [59]. Disruptions to this balance can lead to infections and contribute to inflammatory conditions such as Crohn's disease and ulcerative colitis [60, 61, 62]. Second, the intestinal epithelial barrier functions as a selective, semipermeable layer that allows nutrient absorption while blocking harmful substances, supported by gut microbiota-derived metabolites that maintain barrier integrity and modulate immune responses [63]. Specialized epithelial cells and immune components further enhance its protective role, with IL-22 production by group 3 innate lymphoid cells (ILC3s) playing a crucial role in promoting antimicrobial peptide secretion and Immunoglobulin A (IgA) production to defend against microbial invasion [64, 65, 66](Fig. 1.4). Finally, the gut vascular barrier (GVB) serves as the deepest layer of intestinal defense, preventing the passage of microorganisms into the bloodstream and distant organs [67]. Similar to the blood-brain barrier, the GVB is maintained by endothelial cells, pericytes, and enteric glial cells, forming a "gut vascular unit" that controls permeability. Unlike the blood-brain barrier, it allows some larger molecules to pass, supporting tolerance to dietary antigens while blocking bacterial translocation under healthy conditions [59]. The gut microbiota also play a role in GVB development and maintenance, influencing both vascular density [68] and enteric glial cell (EGC) networks [69].

The gut-liver axis is defined as the bidirectional communication between the microbiota in the gut and the liver tissue [70]. In terms of anatomical and physiological connections, the liver influences the gut through the biliary tract, while the gut affects the liver via the portal vein. Additionally, a more complex connection between the two is also established through the systemic circulation [71]. According to current research, the liver regulates bacterial overgrowth in the gut and maintains intestinal eubiosis by releasing bile salts, antimicrobial molecules, and various bioactive mediators into the bile ducts and systemic circulation [71]. The microbiota in the gut also metabolizes various endogenous and exogenous substrates to produce products which are then transported

to the liver through the portal vein, where they further regulate hepatic bile acid synthesis as well as glucose and lipid metabolism [71].

In my studies, I also integrate the generated cecal transcriptomic, metagenomic, and metatranscriptomic data with the liver multi-omics data [10] previously generated by Prof. Evan G. Williams to explore the interactions among gut microbiota, the gut, and the liver. Please note that these samples were taken in the exact same individual mice, all at the exact time of sacrifice, and thus the molecular modeling of the gut-liver axis is expected to be as precise as possible.

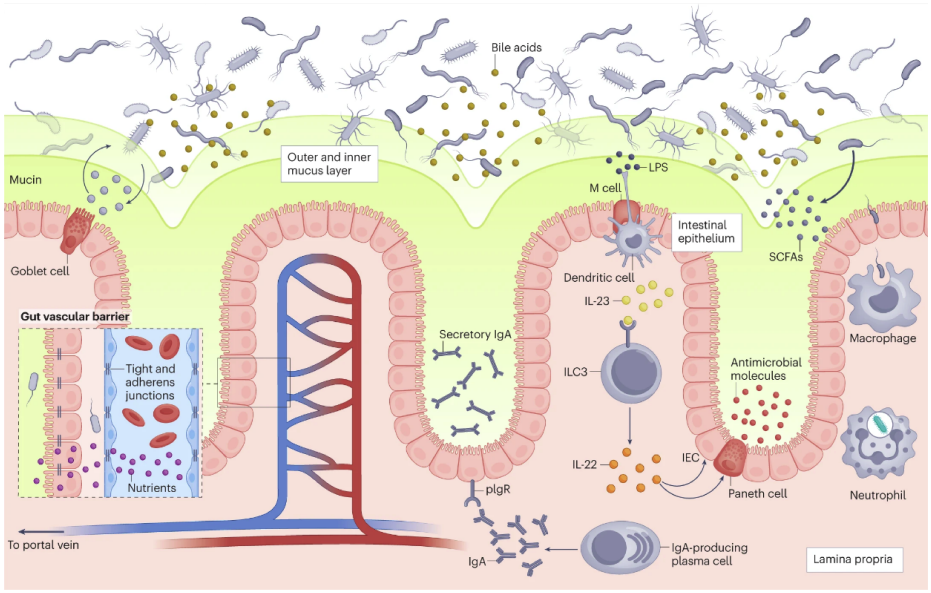


**Figure 1.4** Multiple layers of host defenses prevent the translocation of gut microbiota [58]. Hosts have three defense layers to protect against gut microbiota migrating throughout the body, including the mucus layer, the intestinal epithelium, and the gut vascular barrier. The mucus layer (green), produced by goblet cells, separates gut bacteria from the intestinal epithelial cells (IECs) beneath. The intestinal epithelium forms a barrier that limits bacterial passage, with whole bacteria that cross it being phagocytosed by immune cells. The gut vascular barrier (inset box) further regulates permeability, allowing nutrient passage while restricting viable bacteria and microbial antigens from entering the enteral circulation.

# Chapter 2

# Aim of the study

The primary aim of this thesis is to investigate how host genetics, environmental factors (such as diet and physical activity), and the gut microbiota interact to regulate metabolic health—particularly in relation to glucose tolerance and insulin sensitivity. By combining large-scale mouse models, multi-omics profiling, and functional experiments, this work seeks to uncover mechanisms underlying microbiota–host crosstalk and to identify potential molecular and microbial targets for metabolic disease interventions.

In the first part of the study, we examine how variation in genotype, diet, and age shapes the gut environment and host physiology. To this end, we integrate cecal metagenome (MG), metatranscriptome (MT), and host cecum transcriptome (mRNA) data from hundreds of mice across dietary and age groups. This approach allows us to systematically evaluate how environmental perturbations interact with genetic background to influence gut microbial composition and host gene expression.

The second part of the study focuses on the impact of voluntary physical activity on glucose metabolism. Using a high-fat diet mouse model, we compare sedentary and exercise groups to determine whether long-term physical activity improves glucose tolerance. We further investigate how these metabolic improvements are associated with specific alterations in gut microbiota, leveraging metagenomic profiling to identify bacterial taxa linked to exercise-induced benefits.

In the third part, we explore the mechanistic role of Apolipoprotein A-IV (apoA-IV) in glucose regulation. We assess how both the full-length protein and its bioactive peptide fragment (T55–121) modulate glucose tolerance in genetically diabetic mice. We also evaluate their capacity to stimulate insulin secretion, aiming to provide insight into their therapeutic potential in metabolic disorders such as type 2 diabetes.

Together, these three complementary studies aim to advance our understanding of microbiome–host interactions in the context of metabolic disease and aging, and to explore novel strategies for therapeutic modulation of glucose homeostasis.

# Chapter 3

# Manuscript 1

## Integrative Analysis of the Mouse Cecal Microbiome Across Diet, Age, and Metabolic State in the Diverse BXD Population

## Preface

**I** tested various protocols and developed methods for extracting high-quality biomolecules from complex tissues derived from cecal and fecal mixtures; DNA was extracted from about 270 samples and RNA was extracted from around 100 samples by **me**, followed by quality control assessments; **I** generated a MG dataset for 210 samples and both cecal transcriptomic and metatranscriptomic datasets for 80 samples; **I** processed the raw sequencing data to obtain gene expression data for the cecal mRNA, as well as taxonomic and functional profiles for both MG and metatranscriptomic data; **I** conducted bioinformatic analyses, and applying machine learning models to establish classification models and to identify biomarkers; **Prof. Evan G. Williams** and **Prof. Paul Wilmes** were responsible for the initial conceptualization and study design of the project; **Prof. Evan G. Williams** took charge of supervision and guidance throughout the research process; Animal dissection was performed by **Prof. Robert W. Williams**, **Kyhobeni Mozhui**, **Lu Lu**, **Suheeta Roy**; The RNA sequencing was performed by **Dr. Rashi Halder**; **Arianna Lamanna**, following the protocols established by **me**, generated a metagenomic dataset for 45 samples and both cecal mRNA and MT datasets for 20 samples (not used in this thesis); **Emeline Pansart** learned the extraction protocols and jointly performed the GSEA analysis of the cecal mRNA data with me; **Besma Boussoufa** helped with several practical aspects of the experiments, e.g. preparing logistics for extraction experiments; **Thamila Kerkour** ground approximately 40 samples in liquid nitrogen and performed exploratory initial analyses on the pilot mRNA data from 20 samples that **I** generated.

*In preparation for publication*

# Integrative Analysis of the Mouse Cecal Microbiome Across Diet, Age, and Metabolic State in the Diverse BXD Population

**Ziyun Zhou[1], Arianna Lamanna[1], Rashi Halder[1], Emeline Pansart[1], Besma Boussoufa[1], Thamila Kerkour[2], Paul Wilmes[1], Evan G. Williams[1,*]**

**1** Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Esch-sur-Alzette 4362, Luxembourg
**2** Department of Dermatology, Erasmus MC Cancer Institute, 3015 GD Rotterdam, The Netherlands

* Correspondence to evan.williams@uni.lu

**Keywords:**
Microbiota; Metagenome; Metatranscriptome; Cecum; Aging; Machine Learning; Biomarkers; GXE; Recombinant Inbred Strains

*In preparation for publication*

## 3.1 Abstract

The gut microbiota both adapts to, and shapes, the metabolic state of individuals. This bidirectional relationship is mediated via circulating metabolites and gene regulatory networks and interacts with many organs, including by the gut-brain axis. In this study, we have processed the cecum from 232 mice from a recent aging colony of ours [10], with varying age (6-24 months), diet (chow or high fat), and genetics (43 BXD strains) and sequenced their metagenome, metatranscriptome, and cecal transcriptome. We quantify changes in over 300 species in the metagenome and metatranscriptome caused by interactions between diet, age, and genetic background. Traditional bioinformatics approaches consequently linked particular microbes to physiology, while newer machine learning models based on clusters of microbial profiles far more accurately predicted host phenotype outcomes, including individual body weight (AUC = 0.92) and chronological age (AUC = 0.84). A compact multi-omics model combining our new microbiome data with recent liver proteomes from the same individuals further improved predictions (e.g. improving age AUC to 0.95). Mechanistically, integrative network analyses identified dozens of significant links between particular bacterial taxon and gene expression pathways, such as a strong negative correlation between host *Ido1* expression and short-chain fatty acid (SCFA)-producing *Lachnospiraceae*, indicating dietary fat can modulate host tryptophan metabolism via microbiota shifts. Moreover, as our study uses inbred mice sampled across time, we have identified signature sets of taxonomies that provide excellent predictive value for future metabolic outcomes driven by metabolic networks connecting the microbiome to the host organism's tissues (in this study, cecum and liver from the same mice). By better understanding the gut-liver axis, we can understand the cellular etiologies of metabolic disease and identify earlier and personalized diagnostic biomarkers that are attuned to the genetic background and environmental state of each individual.

## 3.2 Introduction

The typical, individual mammalian intestinal tract is colonized by trillions of bacteria, coming from thousands to tens of thousands of species, collectively called the gut microbiota [72]. Acting as a crucial link between diet and host metabolism, the gut microbiota play an essential role in digestion and nutrient absorption and can impact overall metabolic health [73]. For instance, "germ-free" mice born and raised in sterile environments do not become obese when fed a high-fat and high-carbohydrate diet [74], but they also have immunodeficiencies [75], and with age they tend to suffer from a variety of metabolic disorders such as a dramatic hypertrophy of the cecum [76]. In parallel with diet, aging also alters the diversity and stability of gut microbiota [77]. These aging-associated changes are linked to increased inflammation and a reduction in beneficial metabolite production, which may further accelerate metabolic and immune decline [78].

The cecum, located at the junction between the small intestine and the colon, is a notable region for microbial activity, as it serves as a primary fermentation site for indigestible fibers, leading to the production of short-chain fatty acids (SCFAs), such as butyrate [79]. The cecum is anatomically highly distinct, facilitating equivalent sample collection over long study periods (6 years in the case of our colony). The cecum is a "cul-de-sac" and has a relatively long residence time, with it filling by 6 hours after feeding and then emptying over the next 16 hours [80]. The relative ease of consistent sample collection over time and across variation in time since the last meal makes the cecum an ideal source for microbial sampling. Furthermore, and in contrast to collecting fresh feces from the rectum, cecal collection permits measurement of both the microbe content and the gene expression networks of immediately-surrounding mouse tissue, which can offer direct insights into systemic host-microbiota interactions that are essential for understanding the etiology of metabolic diseases [79, 81].

Given the high diversity of the microbiome and the colonial nature of microbe-microbe interactions, research in this field has broadly struggled to identify specific causal mechanistic relationships between particular microbial taxa to predictable phenotypic outcomes. However, numerous clinical studies over the past decade have shown that the gut microbiome as a whole can have a causal effect on both causing and curing complex metabolic diseases. For instance, mice with inflammatory bowel disease (IBD) can be cured by oral delivery of specific *Akkermansia* bacteria [82], and fecal microbiota transplants (FMT) have been widely proven to cure *Clostridioides difficile* infections in humans, which hitherto had been notoriously difficult to treat [83]. However, successful clinical studies using FMT result in whole-scale shifts affecting hundreds of billions of bacteria, meaning that the specific causal bacteria and their mechanisms are exceedingly difficult to identify, validate, standardize, and reproduce at scale. The causal connection between the gut microbiome to diseases like IBD and colon cancer are unsurprising, yet in recent years, numerous pioneering clinical studies have further found the microbiome to be causally associated with a dizzying variety of diseases far more distal to intestinal health, from Parkinson's disease [84] to type I diabetes (T1D) [85] and type II diabetes (T2D) [86]. However, performing FMT to treat such systemic diseases, and metabolic

phenotypes such as morbid obesity, have had unpredictable results [87]. More detailed mechanistic understanding of the relationship between the microbiome, host, and metabolic health is thus necessary to develop new, targeted treatments.

Therefore, in this study, we generated multi-omics data in the ceca from 232 mice varying according to the interactions between our three independent experimental variables: age (6 to 24 months), diet (low fat "chow" (CD) or high fat (HF)), and genotype (43 inbred strains from the BXD mouse genetic reference population (GRP)). In this cohort, we measured the metagenome (MG) and metatranscriptome (MT) from the cecal contents, and the transcriptome of the cecum tissue itself (abbreviated as "mRNA"). These newly generated data are then supplemented with our prior phenotype data on these mice, along with these individuals' liver transcriptome and proteomes [10]. We aim to first identify and quantify the abundance and activity of specific microbial taxa associated with diet and age, and then apply both traditional bioinformatics approaches and newer machine learning methods to calculate their relationships with key metabolic phenotypes like obesity and expression of key gene networks. We show that these multi-tissue, multi-omic data can be used to predict not only major phenotype outcomes, but also provide insight into the cellular activity of other tissues in the mouse. Thus, not only can we identify specific associations between the microbiome and metabolic health, but we can also infer the affected mechanistic pathways, and — by using the different ages of mice — suggest the causal direction, consistency, and timeline of the connection.

## 3.3 Results

### 3.3.1 Study Design

In 2021, we completed the phenotyping of an aging population of 2157 mice from 89 BXD strains [88], tested in two diets with calories obtained predominantly from carbohydrates (i.e. CD) or from fats (i.e. HF). 662 of these individuals were taken for tissue collection at four approximate intervals: 7, 12, 18, and 24 months of age. Our laboratory has recently completed a detailed molecular analysis of the paired transcriptome, proteome, and metabolome in 347 livers of these mice from a subset of 58 strains across diet and age [10]. In that study, we were able to identify predictive molecular signals of a mouse's future outcomes (such as lifespan), as well as regulatory genes affecting health (such as susceptibility to diet-induced obesity) and the control of molecular pathways affecting aging and diet related genes (such as cholesterol synthesis). From that aging study, we received the whole, flash-frozen cecum samples from 597 individuals in 56 different strains (all overnight fasted, CD n = 299, HF n = 298). For those 597 individuals, there are typically two biological replicates for the selected strain-diet-age matches. Thus, we selected a subset of 232 mice in order to study a wide range of interacting variables in a cost-conscious manner. For these 232 mice, we have sequenced the MG (n = 210), MT (n = 80), and the cecal mRNA (n = 80) (Fig. 3.1 A). All MT and mRNA mice were the same, while 48 overlapped between MT and MG, and overall 155 of these samples overlapped with our prior liver transcript and protein expression data. To ensure consistent extraction across all samples, the entire frozen cecal tissue was homogenized in liquid nitrogen and a powdered aliquot of ~30 mg was taken both for DNA and RNA extraction. The MT and cecal mRNA samples were co-extracted using the same tissue aliquot. MG, MT, and cecal mRNA samples were extracted and sequenced in two temporally distinct batches (November 2022 and November 2023). An overview of the experimental and analytical workflow, including extraction, sequencing, quality control, and downstream processing, is presented in Fig. 3.1 B. For the extracted DNA, samples were then cleaned up by size exclusion to remove large DNA fragments. As we are performing shotgun sequencing on the MG samples (rather than 16S), this step was necessary to remove most of the mouse DNA content of each sample, which was not of interest given that the BXDs are inbred and have recently been fully sequenced [89]. After cleanup, the resulting relative proportions of the microbial to host DNA content for each sample were estimated using RT-qPCR, and samples were subsequently ranked and filtered based on these ratios (Fig. S3.1 A-B). Samples that contained more than 40% microbe DNA were then retained for shotgun sequencing. For extracted total RNA, samples were first separated based on poly(A) enrichment, with the poly(A)-enriched sample being sequenced to obtain the cecal tissue mRNA. The fraction without poly(A) was then treated with restriction enzymes to deplete mouse ribosomal RNA, and the remaining sample was then sequenced to obtain the MT measurements (Fig. S3.1 C).

After read alignment and counting, we used ANOVA to systematically evaluated the relative contributions of the three independent variables — genotype ("strain"), diet, and age — to the observed variation in our multi-omics data, across the MG, MT, and cecal mRNA datasets. These results re-

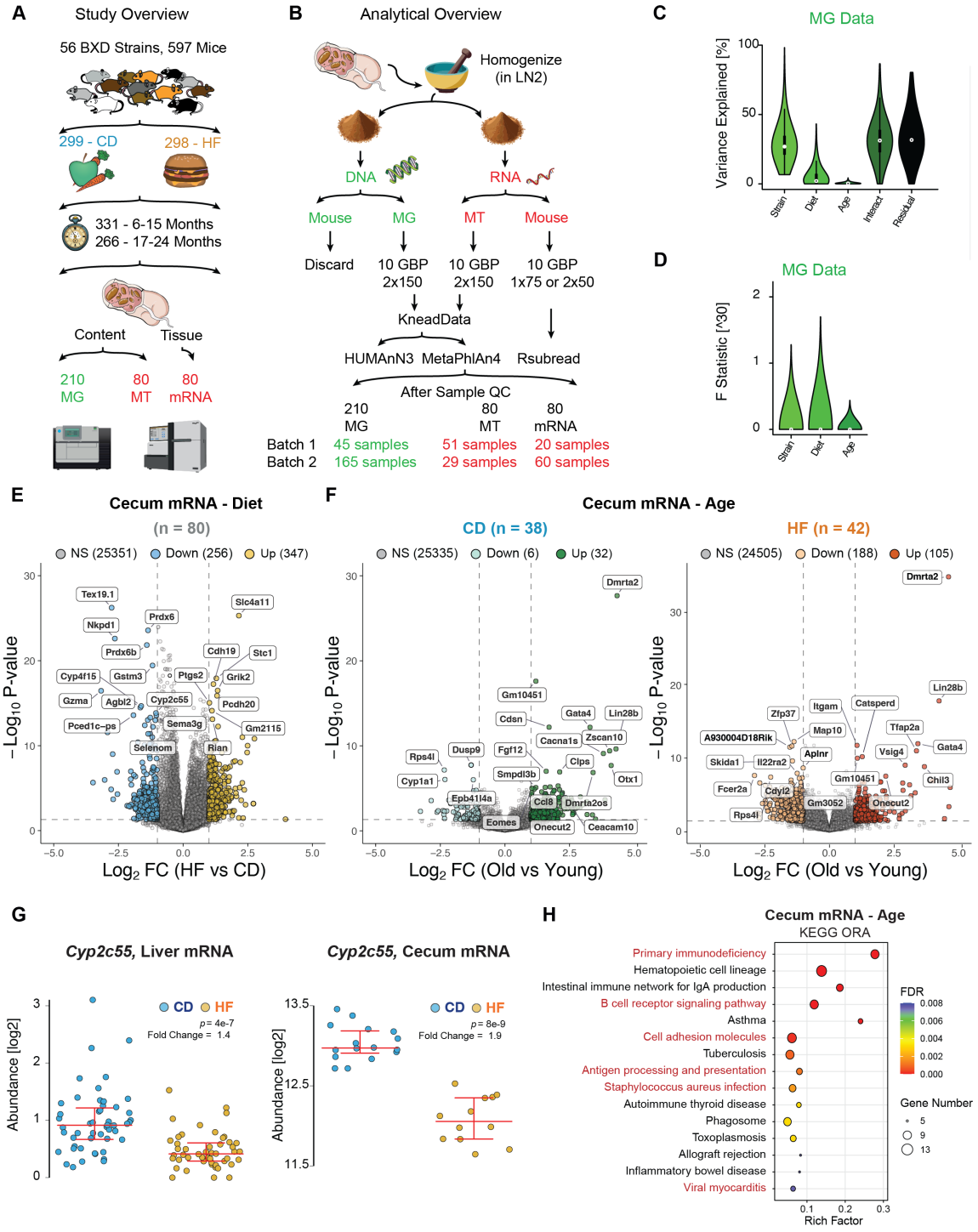


**Figure 3.1** Experimental workflow and analysis of multifactorial cecal mRNA responses. (**A**) Study design for mouse population analysis and data generation of three different omics types from the cecum. (**B**) Experimental process for extraction, sequencing, and analysis of the cecal samples. (**C**) The violin plot from ANOVA analysis. For variance explained, "Interact" represents the combined effects of all interactions, including Strain*Diet, Strain*Age, Diet*Age, Strain*Diet*Age. (**D**) F statistics, computed based on species-level abundance data, for the effects of Strain, Diet, and Age. (**E-F**) Volcano plots of DEGs influenced by dietary and age. (**G**) Dot plot showing the dietary effect on *Cyp2c55* expression in liver and cecum, going in the same direction. (**H**) The 15 most significantly age-associated KEGG pathways, based on DEGs from the cecal mRNA dataset.

vealed that genotype, diet, age, and their interactions together explained, on average, 60-70% of the observed variation in MG (Fig. 3.1 C), MT, and cecal mRNA datasets (Fig. S3.1 D). While strain had by far the largest impact on variance, it is important to note the difference in degrees of freedom between the independent variables. The samples spanned 43 BXD strains, but only two diets, and the ages of sacrificed animals were clustered specifically around four timepoints adulthood (i.e. around 6, 12, 18, and 24 months of age, with a margin of around 30 days). To address this imbalance in the degrees of freedom for each class of independent variable, we calculated the F-statistic, which is adjusted for that concern. This confirmed that diet individually exhibited the strongest adjusted association with gut microbiota composition, generally driving the greatest proportion of variance, although even once adjusted, strain still plays a major role (Fig. 3.1 D). The major effect of diet on the gut microbiome is of course well-documented, and to mitigate potential confounding effects arising from the strong diet-microbiota effect, subsequent analyses targeting diet utilized the full dataset, while age- or strain- related analyses were separately performed within the CD or HF dietary groups, as will be indicated for each analysis.

To further evaluate the internal consistency and reliability of the gut microbiome datasets, we examined genus-level correlations between MG and MT profiles for the 48 overlapping samples. The correlation distribution was strongly shifted toward positive values compared to a shuffled control (median correlation of around 0.75), indicating strong concordance between DNA- and RNA- based taxonomic estimates (Fig. S3.1 E). In addition, principal component analysis (PCA) of species-level MG profiles perfect overlap in between the two chronological extraction batches, indicating that batch control was effective (Fig. S3.1 F). Similarly, both batches of cecal mRNA showed no significant differences after both were transformed and variance stabilized (Fig. S3.1 G). Together, these results demonstrate high internal agreement and minimal batch effects, supporting the reliability of the MG, MT, and cecal mRNA datasets for downstream integrative analyses.

### 3.3.2 Diet and Aging Alter the Cecal mRNA

Host cecal mRNA data were aligned and quantified using Rsubread, and normalized gene counts were used for downstream analysis (Fig. 3.1 B). In this study, we aim to also connect the gene expression mechanisms to bacterial abundances and consequent metabolic outcomes, so we first analyzed the direct effect of our key independent variables of diet and age on the mRNA expression in the ceca (Fig. 3.1 E-F). Differential expression analysis of cecal mRNA revealed distinct transcriptional responses to both diet and age. A total of 603 significant differentially expressed genes (DEGs) were observed (FDR $< 0.05$, $|\log_2 \text{FC}| > 1$) between CD and HF groups, including 347 upregulated and 256 downregulated genes by HF, Fig. 3.1 E). Notably, genes such as *Slc4a11*, *Stc1*, and *Cdh19* were among the most strongly induced by HF feeding, while *Tex19.1*, *Nkpd1*, and *Prdx6b* were downregulated.

We next examined age-associated transcriptional changes, separately for each dietary condition. Under long-term CD, only 38 genes (32 upregulated, 6 downregulated) were differentially expressed with age, suggesting that aging has a relatively limited or non-uniform impact on cecal mRNA expression in the absence of dietary challenge (Fig. 3.1 F, left panel) - also in line with the

lower contribution of age to variance in the ANOVA analysis. In contrast, the HF group exhibited a comparatively enhanced age-related transcriptomic shift, with 293 genes significantly altered (105 upregulated, 188 downregulated; Fig. Fig. 3.1 F, right panel). To identify age-associated transcriptional signatures consistent across dietary conditions, we compared the 20 DEGs most affected by age from the CD and HF age subsets, including both upregulated and downregulated genes. Among the genes upregulated in old mice, five were shared in both two diets: *Dmrta2*, *Lin28b*, *Gata4*, *Onecut2*, and *Gm10451*. In contrast, only *Rps4l* was consistently upregulated in young mice under both diets. Several of these overlapping genes are known to change with age, such as *Gata4* [90] and *Rps4l* [91], whereas the roles of others in aging remain uncharacterized. However, this consistent link with age across diet and numerous diverse genotypes suggests an underlying cellular mechanism. Regarding diet, which has been far more extensively studied, many transcripts most affected by HF diet were in line with expectations from literature and from our prior liver dataset in these individuals. For example, the expression level of *Cyp2c55* is expected to decrease in the gut under HF feeding [92]. This trend was also observed in our dataset, which showed significantly lower *Cyp2c55* mRNA expression in both the liver (fold change = 1.4) and cecum (fold change = 1.9) in HF (Fig. 3.1 G).

To further explore functional pathways associated with age and diet, we applied both over-representation analysis (ORA) based on DEGs and gene set enrichment analysis (GSEA) using ranked differential expression values. For diet-associated changes, using an FDR threshold of $< 0.05$, ORA identified 24 significantly enriched pathways, while GSEA revealed 99 pathways. 16 pathways were consistently detected by both approaches. The top 15 enriched pathways from each method are shown in Fig. S3.2 A–B, with overlapping pathways highlighted in red. In addition, KEGG pathway analysis revealed that several metabolic pathways were significantly downregulated in the HF group, including Metabolism of xenobiotics by cytochrome P450, Chemical carcinogenesis –DNA adducts, Steroid hormone biosynthesis, and Retinol metabolism, among others. In contrast, the Neuroactive ligand –receptor interaction pathway was significantly upregulated in response to HF. Several pathways have been previously linked to HF intake. For instance, the Steroid hormone biosynthesis pathway was significantly suppressed under HF conditions [93]. This suppression may stem from HF-induced mitochondrial dysfunction that impairs cholesterol transport into mitochondria, the first step of steroid hormone biosynthesis [94]. This pathway was enriched with 15 DEGs (e.g., *Cyp2c55*, *Cyp3a11*, *Cyp1a1*, *Cyp2e1*, *Ugt1a9*) (Fig. S3.2 C, upper panel). Moreover, the Neuroactive ligand–receptor interaction pathway has also been reported to be upregulated by HF in the hypothalamic transcriptome of mice [95], potentially driven by the diet–gut–brain axis. In our result, this pathway was enriched with 12 DEGs, including *Lep*, *Grik1*, *Grik2*, and *Grm7*, among others (Fig. S3.2 C, lower panel).

Building on the same analytical framework, we next investigated how age influences pathway-level changes through both ORA and GSEA (Fig. 3.1 H, Fig. S3.2 D). ORA identified 27 significantly enriched pathways, while GSEA revealed 58, with 19 pathways consistently detected by both methods. Among these, the B cell receptor signaling pathway was significantly enriched in both analyses.

This finding aligns with previous reports in humans and animal models, which have shown that aging is often accompanied by a reduction in B cell numbers and functional impairments within B cell populations [96, 97]. Furthermore, this pathway included 10 DEGs (Fig. S3.2 E, left panel). Conversely, several pathways positively associated with aging have also been extensively studied, such as the NOD-like receptor signaling pathway (Fig. S3.2 E, right panel), which is upregulated with age but can be effectively suppressed by fecal microbiota transplantation from young donors [98]. These cecal mRNA results reveal distinct pathway alterations driven by diet and age, offering a potential framework for integration with microbial functional data.

### 3.3.3 MG and MT Analyses Reveal Cecal Microbiome Changes Associated with Dietary Change, Aging and Strains

We next examined the broad taxonomic profiles of the gut microbiota, using quality-controlled MG and MT reads processed with KneadData [99] and analyzed using MetaPhlAn4 [100] (Fig. 3.1 B). Compared to the CD group, HF diet reduced the variety of identified taxa by around 30 to 50% across all taxonomic levels in both MG and MT data (Fig. S3.3 A). In addition, $\alpha$-diversity remained consistently and significantly decreased throughout life, with an equivalent reduction in both MG and MT measurements (Fig. 3.2 A). These findings are consistent with previous widely reported observations that a HF diet reduces microbiome diversity, including at the earliest timepoint (roughly 90 days after the start of HF feeding), as the reduction in diversity is rapid, typically within less than a month [101]. However, the reduction in diversity with diet is an average. Given that our mouse population is highly genetically variable, we observed some overlap between CD mice with the lowest $\alpha$-diversity and HF mice with the highest $\alpha$-diversity, and this distinction remains stable over time. Examining how general patterns in our microbiome data align with established findings from the literature also served as a way to validate the reliability of the extraction, sequencing, and analytical procedures. In addition, $\beta$-diversity analyses (Bray-Curtis dissimilarity) demonstrated that both diet and age significantly influenced gut microbial community structure (Fig. S3.3 B). Diet explained a larger proportion of the variance in community composition and was associated with more pronounced shifts across samples. In contrast, age accounted for a smaller portion of the variance but still exhibited statistically significant effects.

We then checked whether these taxonomic signatures in the ceca will reliably indicate general health outcomes, rather than only recapitulate the known dietary state of the mice. To initially test this, we checked the ratio between the *Firmicutes* and *Bacteroidetes* phyla, which is consistently reported in literature to track with obesity as well as HF diets [102]. Here, we again observed a highly significant elevation in HF cohorts (Fig. 3.2 B), and both *Firmicutes* and *Bacteroidetes* abundances were significantly correlated with body weight (Fig. S3.3 C). However, when age was considered as a factor, no significant difference in the *Firmicutes* / *Bacteroidetes* ratio was observed in older mice (Fig. 3.2 C). This may be due to the general decrease in gut microbial diversity and the increase in individual variation with aging, which could render the relationship between this ratio and body weight less stable. Moreover, while the overall trends are significant, there is substantial overlap

for individual mice within the two diets, indicating that genetic background (i.e. strain) plays a substantial role in the strength of the connection between *Firmicutes* and *Bacteroidetes* abundances and dietary response. Moreover, both *Firmicutes* and *Bacteroidetes* are large and highly-diverse phyla of bacteria, and their lower taxonomic branches can have diverse effects. For instance, within the Firmicutes phylum, *Lactobacillus reuteri* has been associated with increased obesity, whereas *Lactobacillus gasseri SBT2055* exhibits anti-obesogenic properties [103]. This highlights the significance of generating MG data (or using long-read 16S sequencing), which allow precise analysis at more specific taxonomic rankings, down to the genus, species, and even strain level.

As a final quality control, we aimed to show that the specific taxonomies quantified also highlighted reliable, reproducible associations with host health. To this end, we analyzed genus-level relative abundance profiles derived from both MG and MT datasets (Fig. 3.2 D) to dissect microbial composition at a deeper level. Samples were grouped by diet, and within each dietary group, individuals were ordered from left to right by increasing age. We visualize the 15 most abundant genera identified independently in the MG and MT datasets, along with the top 10 shared genera, resulting in a total of 34 genera (plus "Others", which is the sum of all smaller genera). Here, we again observe marked differences in genus composition between the CD and HF groups, such as a dramatic rise in Lachnospiraceae in HF. Conversely, few patterns are visually evident when assessing based on the X-axis (age), consistent with the ANOVA results which showed that diet explains a greater proportion of the variance in microbial composition than age (Fig. 3.1 C, Fig. S3.1 D). Overall, 24 taxa were significantly regulated by diet in the MG data and 19 in the MT data, of which 18 were significant in both datasets.

As with the expected overall decrease in diversity with HF, specifically-affected taxonomic classifications also broadly matched expectations from literature. For instance, *Lachnospiraceae_unclassified*, *Lactobacillus*, and *Bacteroides* exhibited increased relative abundance in the HF group across both MG and MT datasets [104, 105], while *Lactococcus* [106], *Prevotella*, and *Lactobacillus* were less abundant in HF across both measurement types. However, certain highly-consistent and variable genera, such as *GGB62427*, were less well-characterized and have not been extensively studied in the context of diet. Together, these results provide a valuable resource for identifying candidate genera consistently associated with obesity-promoting diets independently of genetic background.

Given the complex influences of diet, age, and strain, we next examined species-level profiles to uncover more specific microbial associations. Based on the results of MaAsLin3 ([107]), we identified a total of 274 gut microbial species at the MG level that were significantly associated (FDR < 0.05) with diet, and 31 species correlated with age. The 25 most strongly and significantly diet-associated features are displayed for the MG and MT datasets in Fig. 3.2 E and Fig. S3.3 D, respectively. The $\beta$-coefficients from the multivariable models indicate both the direction and magnitude of association with the HF diet. Since we used relative abundance data, we considered both abundance and prevalence in our analysis. Positive values indicate higher abundance or prevalence in HF, while negative values indicate depletion. Of the top 25 species selected, 12 were detected as significantly

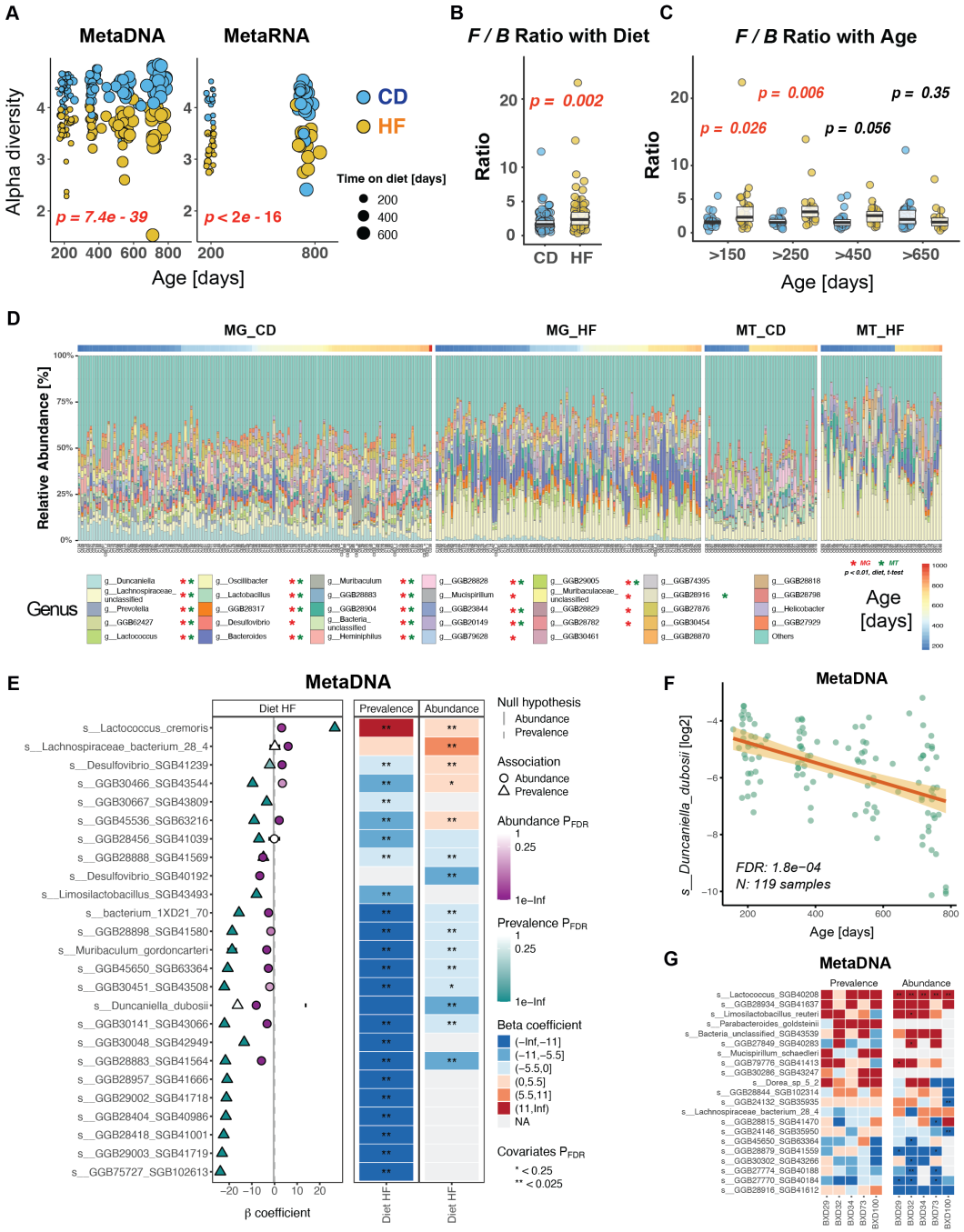


**Figure 3.2** Microbial Diversity and Composition Are Shaped by Diet, Age, and Genotype in MG and MT Datasets. (**A**) The $\alpha$-diversity at the MG and MT levels, calculated using the Shannon index, showing significantly reduced diversity under HF diet conditions. (**B**) The *Firmicutes* / *Bacteroidetes* ratio is significantly increased in HF cohorts. (**C**) This diet-associated difference in F/B ratio diminishes with aging. (**D**) Stacked bar plots showing the relative abundance of the 34 most abundant genera in the MG (left panel) and MT (right panel) datasets, ranked by diet-associated t-test significance. ⋆indicate genera with $p$-values < 0.01. (**E**) Top 25 species most significantly associated with diet in the MG dataset, identified using MaAsLin3. Direction and significance of associations are shown for both prevalence and abundance. (**F**) *s__Duncaniella_dubosii*, identified as the most strongly age-correlated taxon in the MaAsLin3 analysis, shows a significant negative correlation with age. (**G**) Correlation matrix showing strain-level differences in diet associations for selected species, using B6 as the reference.

altered in both abundance and prevalence models. While the results do not always overlap, the discrepancies and alignments help provide a more comprehensive interpretation of the relationship between species and diet. For example, among the top features, *Lactococcus_cremoris* has both increased abundance and prevalence in the HF group, suggesting not only a higher overall abundance but also a broader presence across individuals in the cohort. In comparison, *Desulfovibrio_SGB41239* exhibits increased abundance but decreased prevalence under HF, indicating that while this taxon becomes more abundant in certain individuals, its occurrence became more restricted across the population. This discrepancy may reflect ecological heterogeneity within the HF group, for instance *Desulfovibrio_SGB41239* had a negative prevalence change across the entire cohort, but it had an increased abundance in a subset of mice which may have distinct gut environments or host responses. In contrast, species such as *Muribaculum_gordoncarteri* and *Duncaniella_dubosii* were consistently depleted in both abundance and prevalence by HF, highlighting their general reduction in both quantity and distribution. The observed species-level shifts mirrored those at the genus or family level, but provided more precise identification of diet-associated gut microbes. Notably, four species appeared among the top 25 diet-associated taxa in both the MG and MT datasets (Fig. S3.3 D). For example, the abundance of *Duncaniella_dubosii* was significantly reduced by the HF in both the MG and MT datasets (Fig. S3.3 E). Their consistent detection across MG and MT levels suggests that these taxa are not only present, but also transcriptionally active in response to dietary changes. In addition to well-characterized species, several unclassified taxa showed concordant patterns between MG and MT. For example, *bacterium_1XD21_70* consistently decreased in both abundance and prevalence under the HF diet (Fig. S3.3 F-G), despite lacking functional annotation. These shared and reproducible species may serve as robust candidates for hypothesis generation and future functional validation.

To disentangle the effects of age and strain from dietary influences (Fig. 3.1 C), we performed separate species-level analyses using only the CD samples from the MG dataset. Using the same MaAsLin3 method, we identified several species that were significantly associated with age, such as *Duncaniella_dubosii*, which not only decreased under HF but also showed a marked decline with aging (Fig. 3.2 F). Additionally, to investigate strain-specific microbial patterns, we used B6 mice as the control group and included data from the 5 BXD strains with more than 10 measured individuals. Several species showed notable variation across different strains, highlighting genotype-specific microbial patterns that may reflect underlying host genetic influences on gut microbiota composition (Fig. 3.2 G). Notably, *GGB27849_SGB40283* exhibited elevated abundance in the BXD32, while *GGB28815_SGB41470* was more enriched in BXD100 compared to other strains.

We next focused on functional profiling with HUMANnN3 using the same MG and MT datasets to quantify microbial pathway abundance and coverage across all samples. Prior studies have shown that MT-derived data, which capture microbial activities, are often more informative for linking microbial functions to host phenotypes than MG-derived data, which reflect functional genetic potentials [108]. Nevertheless, MG profiling serves as a valuable reference to account for potential biases introduced during MT data generation and interpretation. A total of 447 MetaCyc pathways were

detected in the MG and 391 in the MT dataset. Among the 33 most abundant pathways across both datasets, 14 pathways in MG and 9 in MT showed significant differences between dietary groups (t-test, $p < 0.05$), with five pathways overlapping between the two datasets (Fig. 3.3 A).

In addition to these dominant pathways, several low-abundance pathways also exhibited significant associations with either diet or age, highlighting potentially subtle but biologically relevant functional shifts. For example, the peptidoglycan biosynthesis III (mycobacteria) pathway, associated with bacterial cell wall formation and host immune stimulation, was significantly upregulated in the HF group at both the MG and MT levels (Fig. 3.3 B). The components of this pathway are generated by a broad range of gut bacterial genera, including *Bacteroides*, *Clostridium*, *Lachnospiraceae_unclassified*, *Helicobacter*, and *Lactococcus*. In both datasets, the HF group exhibited higher total abundance of this pathway, indicating a potential upregulation of bacterial cell wall synthesis.

To comprehensively assess the associations between all pathways and host variables such as diet and age, we further performed detailed statistical comparisons using MaAsLin3, analyzing pathway abundance in MG and MT datasets (Fig. 3.3 C-D) and pathway prevalence in the corresponding datasets (Fig. S3.4 A-B) across different dietary groups. Overall, the pathway abundance analysis identified 210 diet-associated pathways in MG and 110 in MT, while the pathway prevalence analysis revealed 78 significant pathways in MG and 39 in MT. Of these, 91 pathways were shared between MG and MT in the abundance results, and 15 were shared in the prevalence results. Some pathways were primarily driven by specific species, such as Lactose and galactose degradation I (Fig. S3.4 C), whereas others, such as 4-deoxy-L-threo-hex-4-enopyranuronate degradation (Fig. S3.4 D), appeared to be largely attributable to unclassified species, making mechanistic investigation inherently more challenging. Notably, in both MG and MT, the Lactose and galactose degradation I pathway was significantly upregulated in the HF, primarily driven by the marked increase of *s__Lactococcus lactis* under HF conditions. In contrast, other studies have reported a downregulation of this pathway in HF-fed mice, which was largely attributable to a significant decrease in *s__Lactobacillus murinus* [109]. Given that the species driving a pathway may also contribute to multiple functional modules, these results should be interpreted cautiously and are best considered hypothesis-generating. Further experimental validation will be required to delineate the precise mechanisms. By contrast, the downregulation of the 4-deoxy-L-threo-hex-4-enopyranuronate degradation pathway under HF conditions was consistent with previous studies [110]. In addition, several pathways demonstrated robust results across both MG and MT (e.g., Superpathway of (R,R)-butanediol biosynthesis and UTP and CTP dephosphorylation I), although no related studies have been reported to date. Together, these results provide complementary insights into microbial functional alterations associated with dietary.

Beyond diet, age also influenced microbial functional profiles (Fig. S3.5 A-B). In the pathway abundance analysis (Fig. S3.5 A), the L-histidine biosynthesis and L-lysine biosynthesis III pathways were negatively associated with aging, consistent with observations from studies across different human age groups [111]. Conversely, the fucose and rhamnose degradation pathway showed a significant positive association with aging in the pathway prevalence analysis (Fig. S3.5 B), aligning

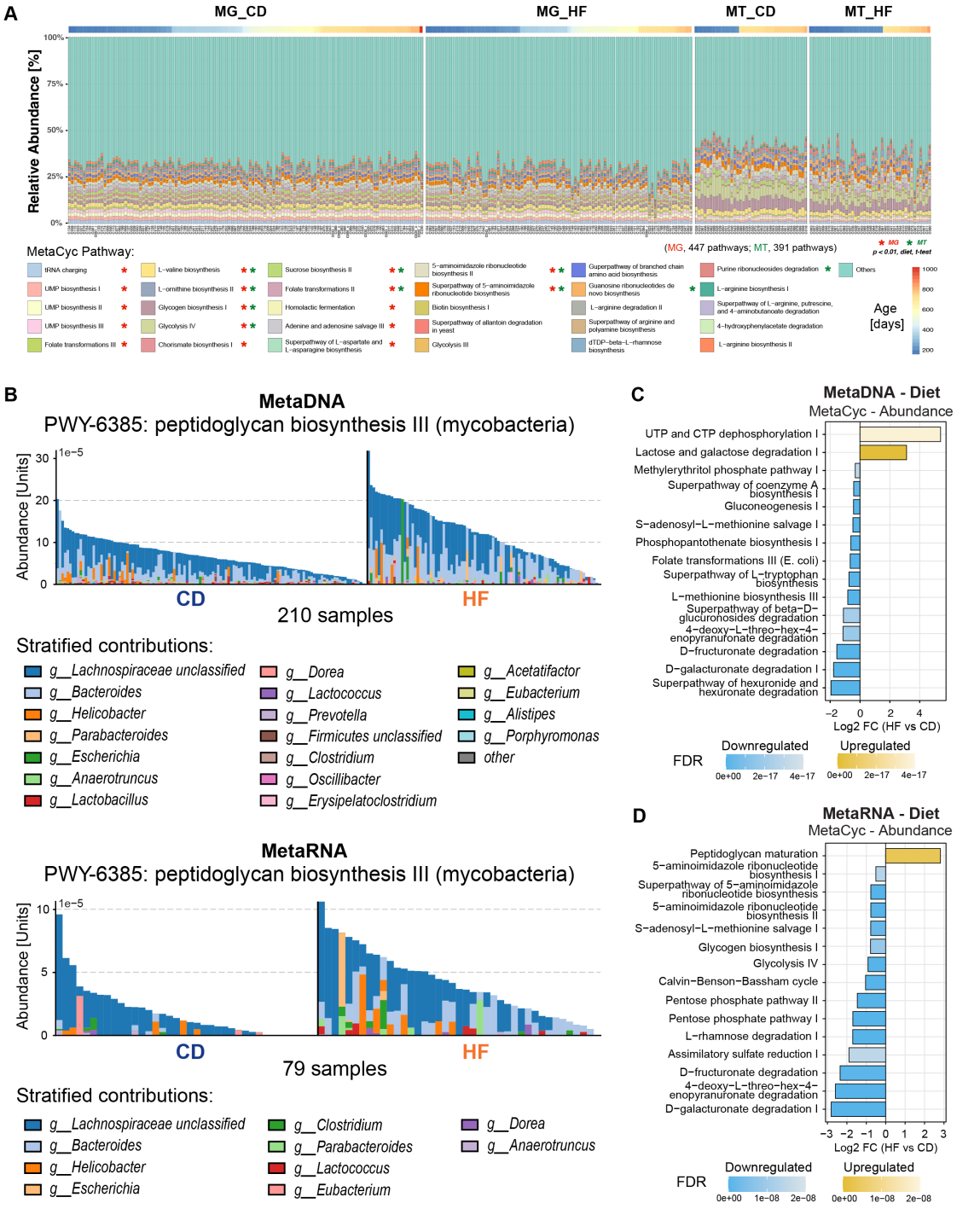


**Figure 3.3** Diet-associated Changes in Microbial Functional Profiles. (**A**) Stacked bar plots of the relative abundance of the 33 most abundant MetaCyc pathways in the MG (left panel) and MT (right panel) datasets, ranked by diet-associated t-test significance. ⋆indicates pathways with $p$-values < 0.01. (**B**) Dietary differences in the peptidoglycan biosynthesis III pathway based on MG (top) and MT (bottom) data. Contributions from "unclassified" bacterial genera were excluded from both panels. (**C**-**D**) Top 15 MetaCyc pathways with significant dietary differences identified by MaAsLin3. Abundance from MG (**C**) and MT (**D**) between CD and HF diets are shown. Bars indicate $\log_2$FC (HF vs CD).

with previous findings of its marked upregulation in older adults with sarcopenia [112]. Taken together, these results highlight how diet and age jointly modulate the gut microbial landscape at both taxonomic and functional levels, and illustrate the utility of integrating MG and MT datasets to capture microbial metabolic outputs.

### 3.3.4 Machine Learning-based Cecal Age Modeling to Identify Microbial Biomarkers Associated with Aging

Beyond conventional statistical analyses of taxonomic and functional profiles described in the previous section, we further employed supervised machine learning to model and predict host outcomes based on gut microbiome composition, starting with chronological age. This approach allows us to first evaluate whether microbial profiles contain sufficient predictive information about host aging, and then to identify key microbial species that contribute most to age classification. By training and validating multiple classifiers, we aimed to complement previous age-associated analyses and prioritize robust microbial biomarkers that may not be easily identified by traditional univariate methods. Given that ANOVA results (Fig. 3.1 C) indicated a much stronger influence of diet than age on microbial composition in the MG dataset, we trained age-prediction models separately within each dietary group. Moreover, the age distribution within our cohort was not continuous (Fig. S3.5 C), making classification more suitable than regression for downstream modeling. To ensure balanced group sizes, we defined samples younger than 350 days as the "Young" group and those older than 450 days as the "Old" group. After filtering, a total of 100 samples (Young n = 40, Old n = 60) were retained for modeling.

We trained four supervised learning models including Random Forest (RF), XGBoost (XGB), logistic regression (Logistic), and deep neural network (DNN) using 5-fold cross-validation to evaluate their ability to classify age groups based on MG species-level data. Among these, the RF classifier achieved the highest performance with an average area under the curve (AUC) of 0.84. To ensure that these selected features reflect biological aging rather than technical artifacts, such as cage- or facility-associated contamination during the extended sample collection period, we generated a combined plot of the relative abundances of the top 20 RF-ranked species across all samples. No evidence of batch-related shifts was observed (Fig. S3.5 D), supporting their relevance to aging rather than environmental confounders. These overlapping features may serve as prioritized candidates for generating aging-related hypotheses and for future experimental validation. In our initial test of age prediction, we achieved significant predictive capacity (AUC 0.84) within the training dataset (Fig. 3.4 A, left panel). To further assess the generalizability of these models, we conducted an independent batch-wise evaluation in which one MG data batch was used for training and the other for testing. Both RF and XGB models achieved AUCs exceeding 0.80 in this setting (Fig. 3.4 A, right panel), supporting their robustness across biological replicates. To improve this, we incorporated feature interaction (FI) terms to enhance model expressiveness when trained on smaller datasets. Inclusion of interacting features yielded modest improvements in prediction accuracy, suggesting potential benefits (e.g. AUC improved from 0.80 to 0.88 with RF) in capturing non-linear relation-

ships between microbial features and host age. In parallel, we also trained age classification models using MG data from the HF group and MT data from the CD group (Fig. S3.6 A–B). In both cases, the RF classifier outperformed other models, with cross-validated AUCs exceeding 0.80. Even with the smaller sample size in the MT dataset, the AUC of age prediction remained similar with RF (0.81 in both cases).

Building on the strong performance of the classification model for age, we next ranked all microbial features by their importance scores derived from the trained classifiers. To evaluate the contribution of top-ranked features, we retrained 5-fold cross-validation models using stepwise subsets of the top 50 ranked species, increasing the feature set in increments of five. The RF model achieved an AUC of 0.85 using only the top 5 features, and exceeded 0.90 when using the top 20 features. No improvement was observed with larger numbers of features selected, indicating that a small subset of microbial taxa carries sufficient information for accurate age prediction (Fig. 3.4 B). XGB models showed similar trends, although their predictive performance was slightly lower than that of RF. By ranking features according to the sum of SHAP value magnitudes across all samples, we found that 10 species overlapped among the top 20 features identified by both models (Fig. 3.4 C), highlighting their robustness as age-associated candidates, e.g. *GGB27866_SGB40300*, *Streptococcus danieliae*, *Paramuribaculum intestinale*, and *Duncaniella dubosii*.

To compare microbial features identified by different analytical strategies, we next examined the overlap between statistical significance from MaAsLin3 linear models and feature importance scores from RF classifiers. While there was substantial variation between the two approaches, a subset of species was consistently highlighted by both methods (Fig. 3.4 D). A moderate but significant positive correlation was observed between model-derived feature importance and the negative log-transformed q-values from MaAsLin3 (Spearman's: $R = 0.2$, $p = 1.3e-4$), suggesting that species strongly associated with age are more likely to be prioritized across analytical frameworks. Of particular note, *Streptococcus danieliae* exhibited the strongest positive correlation with aging, whereas *Duncaniella dubosii* showed the strongest negative association. To further explore the phylogenetic distribution of age- and diet- associated microbial features, we annotated a taxonomic tree of all detected species with significance results from the MaAsLin3 analysis (Fig. S3.6 C). This visualization enabled us to examine whether microbes responsive to host variables were clustered within specific taxonomic clades. However, no clear aggregation patterns were observed, suggesting that microbial responses to diet and aging are shaped by complex host–microbe interactions rather than being driven by a single functional pathway or metabolite class. Interestingly, several species highlighted by both machine learning methods and traditional bioinformatic approaches exhibited low mean abundance, reinforcing the importance of using complementary analytical strategies to identify robust features. This integrative approach allowed us to uncover microbial signatures that may otherwise be overlooked — particularly those bacteria with only sequential identifiers rather than formal species names — and offers a framework for identifying and selecting candidates for mechanistic experiments.

To further evaluate the reliability and biological relevance of our modeling framework, we ap-

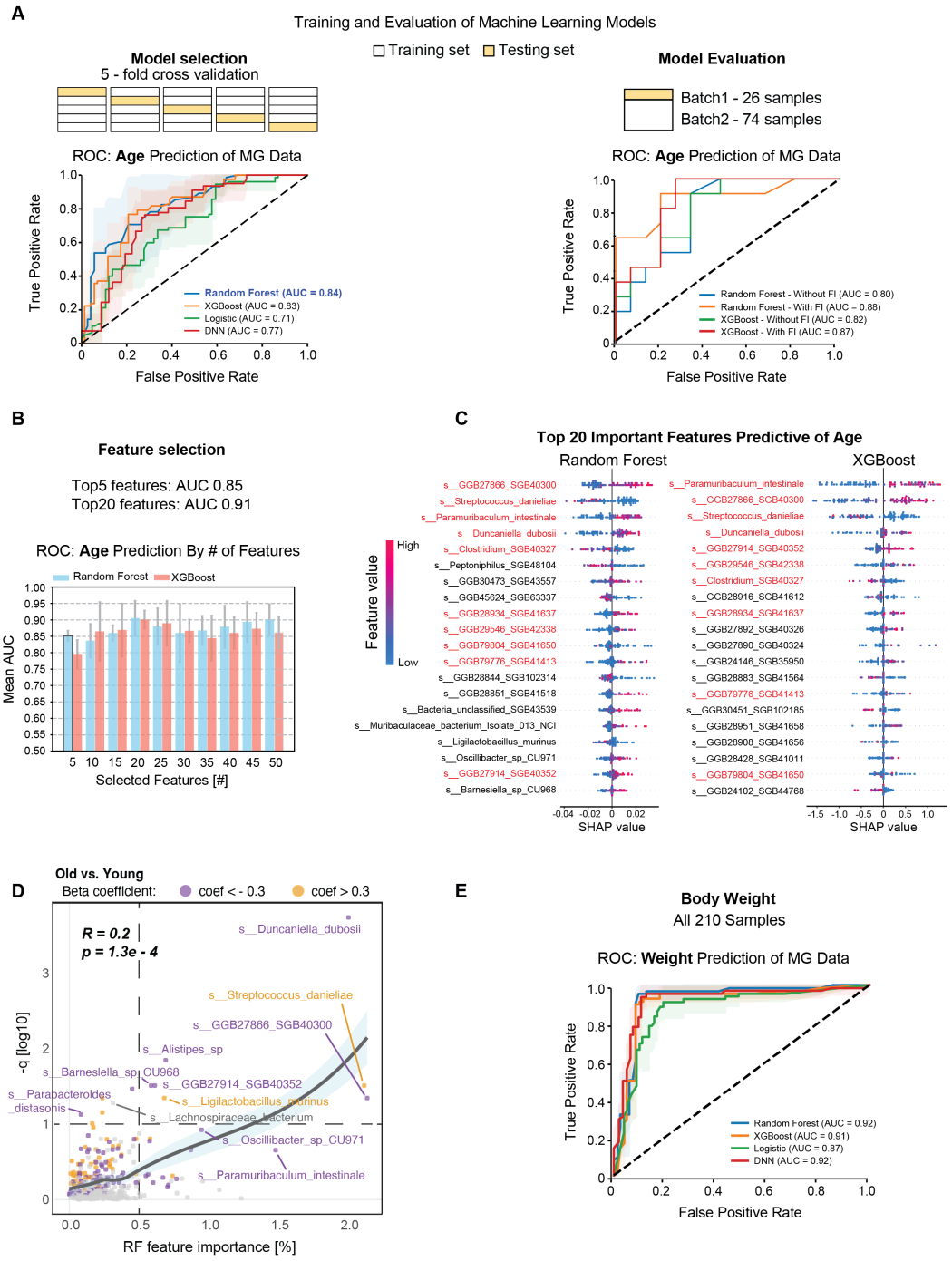


**Figure 3.4** Machine Learning Models Identify Microbial Species Predictive of Host Age. (**A**) Five-fold cross-validation for model training and evaluation of age prediction using MG data. Left: RF achieved the highest AUC (0.84) among tested models. Right: Evaluation on two independent batches demonstrated consistent model performance. (**B**) AUC values of models retrained using different numbers of top-ranked features with 5-fold cross-validation. (**C**) Feature importance was calculated using RF and XGB models. (**D**) Comparison of species associated with age identified by MaAsLin3 and RF. Each point represents one species. The scatter plot shows a positive correlation between RF feature importance and MaAsLin3 statistical significance. (**E**) MG data used to train the body weight prediction models have shown high accuracy.

plied the same machine learning pipeline to predict mouse body weight using MG data (Fig. 3.4 E). Given that diet exerts a much stronger effect than age in the MG dataset and is directly associated with body weight, the body weight prediction model — as expected — performed exceptionally well (RF: AUC = 0.92). We also compared the top 20 features ranked by RF and XGB models (Fig. S3.6 D), and observed a higher degree of agreement, with 10 overlapping species — greater than what was observed in the age prediction models. This result not only highlights the predictive capacity of microbiome-based features for host metabolic traits but also indirectly supports the robustness of our age prediction models, and that the approach can be applied to both dependent (body weight) or independent (age) variables.

### 3.3.5 Integrated Multi-omics Analyses Reveal Host-Microbiome Interactions Linked to Metabolism and Aging

Having performed a series of single-omics analyses on individual datasets in the previous sections, we identified multiple microbial and host features associated with diet, age, and strain. To further leverage the multi-omics datasets generated from the same BXD mouse cohort [10], we conducted integrative analyses across microbiome, cecum, and liver datasets. This enabled us to explore host–microbiome interactions more comprehensively and evaluate the added value of multi-omics integration in predicting biological traits and identifying functionally relevant microbial associations. We first examined the relationship between gut microbial composition and host metabolic gene expression by integrating MG species-level profiles with liver transcriptomic data. In particular, we tested the expression of *Fasn* (fatty acid synthase), a key enzyme in hepatic lipogenesis, in relation to the F/B ratio. In the CD group, *Fasn* expression showed a significant positive correlation with the F/B ratio ($r = 0.52$, $p = 0.001$), whereas no such association was observed in the HF group ($r = 0.26$, $p = 0.16$) (Fig. 3.5 A). These findings suggest that the relationship between microbial composition and hepatic lipid metabolism may be modulated by dietary context, with stronger microbiota–host interactions under CD conditions.

To further assess whether integrating multi-omics data could further enhance the performance of microbiome-based age classifiers, we focused on two datasets — MG species-level data and liver proteomics (LP) — which had the largest number of overlapping samples (n = 139) within our multi-omics collection. As a first step, we evaluated the predictive capacity of LP data alone. Using the same classification framework applied to the MG dataset, we trained classifiers on LP data from CD mice, binarized into Young (< 350 days) and Old (> 450 days) age groups. The RF model achieved an average AUC of 0.81 in 5-fold cross-validation (Fig. S3.7 A, left panel). Notably, several models reached AUCs of approximately 0.85 using only the top five features, indicating that a small subset of liver proteins was sufficient for robust age prediction (Fig. S3.7 A, right panel). Consequently, we next selected the overlapping samples between the MG and LP datasets and stratified them by age, retaining 61 samples that fell into distinct Young (< 350 days, n = 25) or Old (> 450 days, n = 36) categories. These samples were then used to independently train age-prediction classifiers on each dataset (Fig. 3.5 B). When using all available features from each dataset, model performance was

modest (MG: AUC = 0.59, LP: AUC = 0.76), likely due to the small sample size relative to high feature dimensionality. To address this, we applied prior feature selection based on the top-performing models from earlier analyses, retaining only the top 10 features from each dataset. This reduction led to a marked improvement in performance for both MG (AUC = 0.78) and LP (AUC = 0.92), demonstrating the utility of targeted feature selection in small-sample, high-dimensional settings. Finally, we combined the top five MG features with the top five LP features to train an integrated classifier. This resulted in a compact 10-feature model, allowing direct comparison with the earlier models trained using the top 10 features from each individual dataset. This minimal multi-omics model achieved a striking AUC of 0.95, suggesting that low-dimensional integration of key features across omics layers can substantially improve predictive performance. These results highlight the potential of compact, cross-modal biomarker panels to overcome limitations in multi-omics sample overlap, enabling more flexible study designs. The most predictive microbial and proteomic features identified here also offer promising targets for future mechanistic studies into host–microbiome interactions during aging.

Given the central role of the gut–liver axis in host metabolic and immune regulation, we next explored potential interactions between microbial and hepatic protein features identified through our machine learning models. Specifically, we used the HAllA framework [113] to assess associations between the top 30 LP features and the top 20 MG features. The number of LP features was expanded relative to MG to reflect the greater feature dimensionality in the proteomic dataset, with the top 10 ranked associations highlighted (Fig. 3.5 C). We observed that *Streptococcus danieliae*, one of the most predictive MG features for age, exhibited strong positive correlations with three liver proteins — *Por*, *Gsta2*, and *Cyp4a10*. These cross-omics associations provide preliminary evidence for microbiome–host molecular crosstalk and highlight candidate gut–liver pathways for future mechanistic investigation.

Lastly, to further investigate cecum–microbiome interactions at the systems level, we constructed cross-omics co-expression networks (Fig. S3.7 B-C, Fig. 3.5 D) using Weighted Gene Correlation Network Analysis (WGCNA) based on cecal mRNA and metagenomic species data. This analysis identified several modules containing tightly correlated host genes and microbial taxa. For example, *Ido1* expression in the cecum strongly associated with a group of HF-enriched species in the family *Lachnospiraceae*, such as *Lachnospiraceaebacterium M18-1*. Previous studies have shown that *Ido1* knockout in mice alters gut microbial composition, and that mice lacking *Ido1* exhibit significantly reduced white adipose tissue when fed a HF diet [114]. These findings support the intriguing possibility that, under HF conditions, *Ido1* may influence host metabolic outcomes through its interactions with specific *Lachnospiraceae* species. Consequently, this integrated network approach allowed us to link multi-layered biological signals across host and microbial compartments, highlighting candidate gene–microbe relationships that warrant further mechanistic exploration. Together, these results underscore the value of combining statistical modeling, machine learning, and network-based strategies to extract meaningful biological insights from multi-omics data. They also establish a foundation for future hypothesis-driven studies aimed at dissecting causal

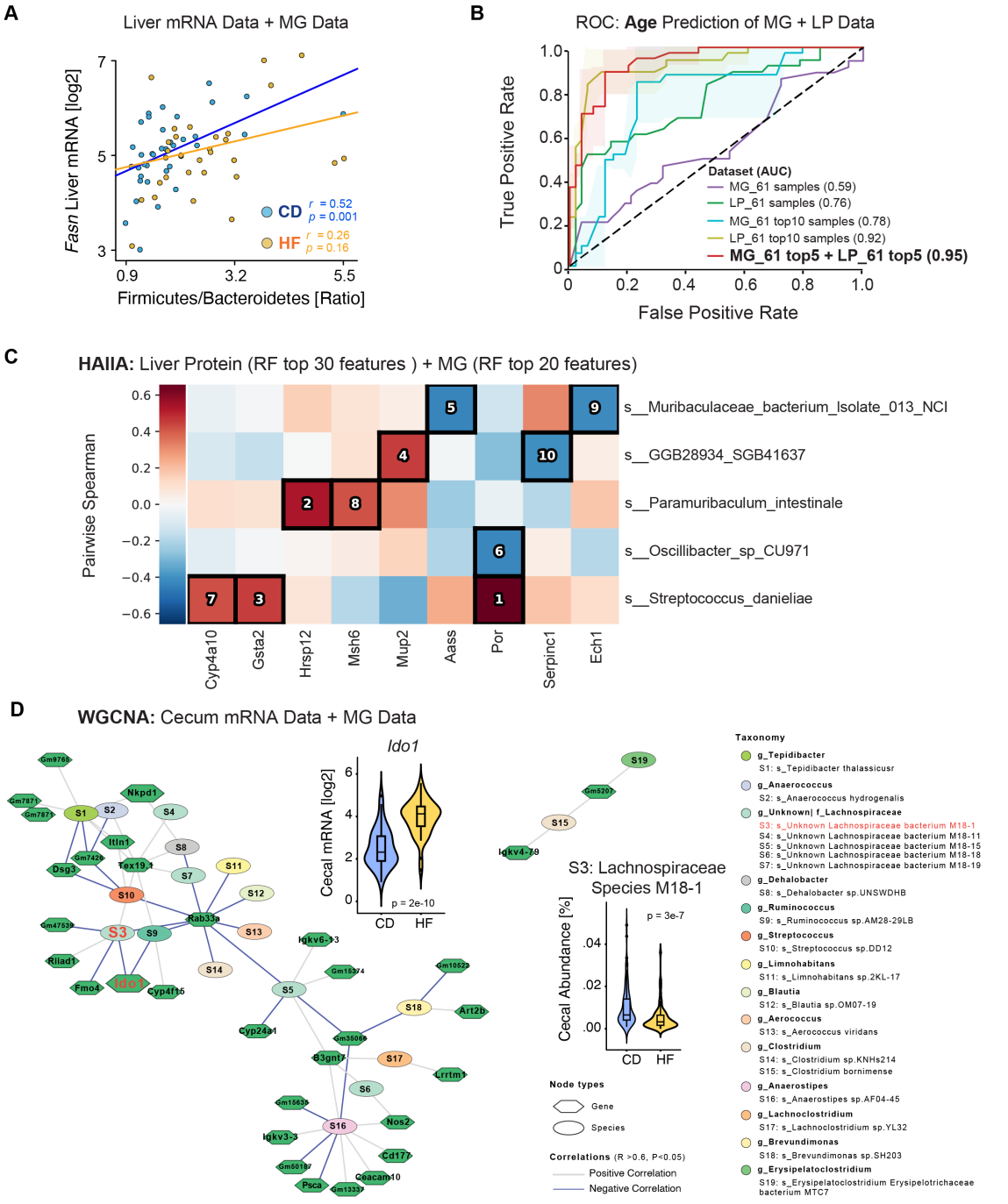


**Figure 3.5** Integrated Multi-omics Analysis Reveals Age Prediction Performance and Microbial-host Interactions. (**A**) Correlation between *Firmicutes / Bacteroidetes* ratio and liver expression of *Fasn* stratified by diet group. (**B**) ROC curves for age prediction models based on MG and liver proteomic (LP) datasets, showing improved performance with integrated top features (AUCs up to 0.95). (**C**) Pairwise Spearman correlation heatmap between liver proteins and MG species. Color indicates correlation strength, with red for positive and blue for negative correlations. Black boxes mark significant correlations, with numbers showing their significance ranks. (**D**) Network visualization of host–microbiome correlations. Nodes represent either genes (diamonds) or species (ovals); edges indicate significant correlations ($|r| > 0.6$, $p < 0.05$). Violin plots show the expression of *Ido1* and relative abundance of *Lachnospiraceae_species_M18-1*.

microbiome–host interactions in the context of diet and aging.

## 3.4 Discussion

Comprehensive studies analyzing the gut microbiota are showing new insights that researchers and clinicians can find to diagnose current health concerns, and for prognosticating the future metabolic health of individuals. Furthermore, as with blood and urine, gut microbiome samples can be collected longitudinally and minimally invasive. In this study, we have focused on the cecum rather than fecal samples due both to the fasting state of the aging mice, the larger sample size compared to fecal pellets, to maintain timing consistency with the previously-collected liver samples, and the ability to ensure consistent sample collection across an aging colony of several years. Over the past decade, rapid improvements in sequencing quality and orders of magnitude of cost decreases have opened the door for large-scale studies on the composition and activity of the microbiome. These technological gains now allow for studies, like this one, which can accurately identify and quantify hundreds to thousands of taxa reproducibly and accurately across hundreds of individuals. Here, we systematically integrated MG, MT, and host cecal mRNA data from a genetically heterogeneous mouse cohort to investigate how diet, aging, and genetic background collectively shape gut microbial communities and host molecular responses. We first showed that age-associated transcriptional changes in the ceca were relatively modest under long-term CD but became substantially more pronounced under HF conditions, highlighting a strong diet — by — age interaction at the host transcriptome level. At the taxonomic level, we quantified 614 microbial species in the MG dataset. Using supervised machine learning, we demonstrated that these species-level profiles can accurately predict key host phenotypes: body weight (AUC = 0.92) and age (AUC = 0.84), underscoring the predictive capacity of the gut microbiome. Furthermore, through integrative modeling and prior feature selection between this dataset and our recently-collected liver transcriptome and proteome in the same mice, we were able to perform cross-tissue multi-omics analyses to obtain more reliable results with fewer selected features. Given the dynamic responsiveness of the microbiome to environmental shifts and the convenience of sampling, this integrative approach offers powerful predictive capabilities for individualized assessments of future health trajectories, reflecting genetic and environmental contexts.

Through an integrated analytical approach utilizing both MaAsLin3 and RF modeling, several microbial species consistently associated with aging emerged, notably *Streptococcus danieliae* and *Duncaniella dubosii*. *S. danieliae*, identified as strongly positively correlated with age in our analysis, aligns with prior findings from fecal 16S rRNA sequencing studies indicating that increased abundance of the *Streptococcus* genera accelerates aging [115]. These studies suggest that certain *Streptococcus* species could promote cellular aging, such as by elevating oxidative stress levels. Our findings raise the possibility that *S. danieliae* may represent a species-level candidate contributing to the observed age-associated microbial shifts. Conversely, *D. dubosii* showed a significant negative correlation with aging, highlighting its potential protective role. Prior studies have linked microbial modulation of tryptophan metabolism to longevity and age-related disease resistance [116], and intriguingly, the survival benefits conferred by tryptophan were found to be microbiota-dependent.

While its precise role in aging remains to be determined, our findings nominate *D. dubosii* as a candidate taxon for further investigation into microbiome-based mechanisms supporting healthy aging.

Our network analysis using WGCNA identified significant associations between cecal expression of *Ido1* and specific *Lachnospiraceae* species, particularly *Lachnospiraceae bacterium M18-1*. Notably, under HF conditions, we observed a significant increase in *Ido1* expression alongside a marked reduction in the abundance of several *Lachnospiraceae* species. This inverse pattern was reflected in a negative correlation within the co-expression network. Although no prior studies have directly examined interactions between *Ido1* and individual species within the *Lachnospiraceae* family-level taxon, existing literature provides a plausible mechanistic link. Members of *Lachnospiraceae* are core constituents of the gut microbiota and are major producers of short-chain fatty acids (SCFAs) such as butyrate [117]. SCFAs have been shown to inhibit the expression of *Stat1*, thereby downregulating IFN$\gamma$-induced transcription of *Ido1* [118]. Our findings are consistent with this model, suggesting that a HF-induced decline in SCFA-producing microbes may relieve inhibitory control over *Ido1*, leading to its upregulation. Furthermore, *Ido1* knockout has been associated with altered microbiota composition under HF conditions [114], pointing to a bidirectional feedback between host tryptophan metabolism and microbial ecology. These observations underscore a potential regulatory axis between microbial SCFA production and host immune-metabolic responses via the *Ido1* pathway.

Our multi-omics predictive modeling, integrating the top five MG species and top five liver proteomic features, significantly enhanced predictive accuracy (AUC = 0.95), demonstrating the efficacy of compact multi-omics biomarker panels. Beyond improved predictive performance, this strategy effectively addresses the challenge posed by limited overlapping samples common in multi-omics studies, especially those with extended and heterogeneous sampling periods. By initially identifying robust biomarkers within individual omic datasets, researchers can subsequently leverage smaller, overlapping sample subsets for integrative modeling. This approach not only enhances study flexibility and resilience but is also highly applicable to human studies, where sample heterogeneity and logistical constraints often limit integrative analyses [119].

Several limitations of our study should be acknowledged. Firstly, the sample size per strain restricted detailed analysis of the genetic impact across diverse BXD strains. Only 5 strains had more than 10 individuals analyzed, thus limiting insights into microbiome interactions unique to individual outlier strains. Secondly, key identified microbial taxa, such as *"GGB28815 SGB41470"*, remain poorly characterized and their cultivation conditions are unknown, complicating downstream functional validations through traditional microbiological methods. Further development in culturomics and advanced molecular techniques are required to overcome these barriers. Additionally, our study relied on cross-sectional sampling at discrete age points, precluding longitudinal analyses that could offer deeper insights into microbiome dynamics and causal inference.

In conclusion, our integrative multi-omics approach highlights the influence of dietary patterns and aging on the gut microbiome composition and functional landscape, identifying key microbial

taxa and metabolic pathways intricately linked to host metabolic health. Our results further underscore the remarkable utility of microbiome profiling combined with machine learning for accurate host age prediction, underscoring its potential application as a reliable biomarker for biological aging and personalized health monitoring. These findings provide a robust foundation for future mechanistic explorations and translational applications aiming to harness microbiome insights for metabolic health interventions and age-related health assessments.

## 3.5 Materials and Methods

### 3.5.1 Mice Sample Collection and Processing

597 animals from 56 strains of the BXD family were sacrificed at specific ages (7, 12, 18, or 24 months of age, plus or minus one month) for tissue collection across cohort [10]. Diets were either Harlan Teklad 2018 (CD; 24% calories from protein, 18% from fat, 58% from carbohydrates) or Harlan Teklad 06414 (HF; 18.3% calories from protein, 60.3% from fat, 21.4% from carbohydrates). Animals were removed from the aging colony the night prior to sacrifice, but they retained access to food and water. Cecal tissue containing fecal matter was harvested and subsequently stored in -80°C freezers.

For the sample preprocessing, a subset of 231 of the above samples from 43 strains were selected. For these samples, the mixed cecum and fecal tissue samples were transferred from a -80°C freezer to dry ice. Prior to grinding, the mortar and pestle was pre-cooled with liquid nitrogen. The tissues were ground using the pre-cooled mortar and pestle until fine powders are obtained. After the sublimation of the liquid nitrogen, the powder was transferred to new tubes using a round spatula scraper and immediately flash-freeze in liquid nitrogen. The mixed cecum and fecal powder was stored in a -80°C freezer for subsequent DNA and RNA extraction.

### 3.5.2 Optimization of DNA extraction protocol

For DNA extraction, a modified extraction method employing the Wizard Genomic DNA Purification Kit was used to simultaneously extract DNA from both the host and the gut bacteria in mixed samples of cecum and feces. ~ 30 mg of tissue powder was suspended in 480 $\mu$L of 50 mM EDTA and homogenized with stainless steel milling beads using an Oscillating Mill MM400 (Retsch) at 25 Hz for 1 min, then incubated at 37°C for 60 min. After the homogenates were centrifuged for 2 min at 500 g, the supernatant was transferred into a fresh tube. This supernatant was then centrifuged for 2 min at 16,000 g. After centrifugation, the supernatant was removed and the pellets were collected. The pellets were lysed by adding 600 $\mu$L of chilled Nuclei Lysis Solution, gently mixed by pipetting, incubated at 80°C for 5 min, and subsequently cooled to room temperature. Next, 3 $\mu$L of RNase Solution was added, and the mixture was incubated at 37°C for 45 min before cooling to room temperature. Protein precipitation involved adding 200 $\mu$L of Protein Precipitation Solution (from the same Wizard kit), followed by vortexing and a 5 min incubation on ice. The sample was then centrifuged at 13,000 × g for 3 min. Subsequently, the supernatant was transferred to a fresh tube containing 600 $\mu$L of isopropanol for DNA precipitation and mixed gently by inversion. This was followed by a centrifugation at 16,000 × g for 2 min to remove the supernatant. Next, 600 $\mu$L of 70% ethanol was added, mixed, and the sample was centrifuged again for 2 min at 16,000 g. The ethanol was aspirated, and the pellet was air-dried for 15 min. The DNA pellet was rehydrated in 100 $\mu$L of Rehydration Solution for 1 hour at 65°C. For DNA purification, the Wizard DNA Clean-up System was utilized following the manufacturer's protocol.

For DNA quality assessment, the concentration and purity of each DNA sample were measured using a Spectrophotometer Nanodrop 2000C (Thermo Scientific). We established a method

for evaluating the proportion of gut bacterial DNA using RT-qPCR with the ratio of *Gapdh* primers (Forward: CAAGGAGACCTCAAGGTCAG; reverse: GTATGCACCTCACAACGCCATG) for the eukaryotic content compared to the universal prokaryotic 16S rRNA primers (Eurogentec, Forward: AAACTCAAAKGAATTGACGG; reverse: CTCACRRCACGAGCTGAC). Host DNA and intestinal bacterial DNA were extracted separately from liver and fecal samples and then diluted to a concentration of 10 ng/$\mu$L. Subsequently, different ratios of host DNA and gut microbiota DNA were mixed to serve as controls. The concentration of all samples was adjusted to 10 ng/$\mu$L for use in RT-qPCR assays. RT-qPCR was performed using the LightCycler 480 Instrument II (Roche). Each reaction had 2.5 $\mu$L of cDNA, 2.5 $\mu$L of primer mixture (forward and backward primers) (2 $\mu$M) and 5 $\mu$L of the iQ SYBR Green Supermix (BIO-RAD). The PCR reaction conditions were as follows: an initial denaturation step at 95°C for 10 min, followed by 40 cycles of denaturation at 95°C for 15 seconds and annealing/extension at 60°C for 1 min. The reaction was subsequently cooled to 20°C for 1 hour. Based on the Crossing Point (Cp) values obtained from the RT-qPCR results, we defined a %Expression index to estimate gut microbiota–to–host DNA ratios and used it to filter samples with higher bacterial DNA content. The %Expression index was defined as

$$\Delta C_p(16S - Gapdh) = (\text{16S\_average\_}C_p) - (Gapdh\text{\_average\_}C_p)$$

$$\Delta\Delta C_p(\Delta C_p\text{\_sample} - \Delta C_p\text{\_6L4F}) = \Delta C_p(\text{16S - }Gapdh)\text{\_sample} + \Delta C_p(\text{16S - }Gapdh)\text{\_6L4F}$$

$$\%\text{Expression} = 100 \times 2^{-\Delta\Delta C_p(\Delta C_p\text{\_sample} - \Delta C_p\text{\_6L4F})}$$

Where `16S_average_Cp` is the average Cp value from three technical replicates of each DNA sample extracted from mouse cecal contents, using universal prokaryotic 16S rRNA primers; and *Gapdh*`_average_Cp` is the average Cp value from the same samples using primers targeting the mouse reference gene *Gapdh*. The value of $\Delta C_p$_6L4F refers to the Cp difference calculated from a reference DNA mixture composed of 40% fecal DNA and 60% liver DNA (mixed at equal concentrations), which serves as a normalization control across samples.

Samples with a gut microbiota DNA proportion estimated to be less than 40%, based on the Cp values of the controls, were excluded from sequencing (Fig. S3.1 A-B).

### 3.5.3 RNA Extraction Protocol

For RNA extraction, ~ 30 mg of tissue powder was suspended in 1 mL chilled TRIzol (Thermo Scientific) and homogenized with stainless steel milling beads in an Oscillating Mill MM400 (Retsch) at 25 Hz for 30 seconds, followed by a phase separation using 200 $\mu$L of chloroform, and the mixture was vortexed before being left at room temperature for another 2 min. The mixture was centrifuged at 12,000 g for 15 min at 4°C, the upper aqueous phase was then transferred to a new RNase-free tube. For RNA precipitation, 400 $\mu$L of isopropanol was added and thoroughly mixed by pipetting, followed by incubated at room temperature for 2-5 min. This was followed by centrifugation at 12,000 g for 10 min at 4°C, with the supernatant was discarded. The RNA pellet was washed with

1 mL of 75% ethanol, centrifuged at 7,500 g for 5 min at 4°C, and the supernatant was subsequently discarded. Finally, the RNA pellet was air-dried at room temperature for 10 min before being re-suspended in 50 $\mu$L of RNase-free water. For RNA purification, the RNeasy MinElute Cleanup Kit was utilized following the manufacturer's protocol (Qiagen, March 2016 version). For RNA quality assessment, the concentration and purity of each RNA sample were measured using a Spectrophotometer Nanodrop 2000C (Thermo Scientific). RNA quality was evaluated using an Agilent RNA 6000 Nano Kit and Agilent 2100 Bioanalyzer. Given that the RNA samples contain both host RNA and gut bacterial RNA, the RNA Integrity Number (RIN) is not applicable for these mixed samples. Only samples exhibiting four distinct peaks were selected for subsequent sequencing (Fig. S3.1 C).

### 3.5.4 Sequencing and Profiling

#### Sequencing

The mRNA sequencing libraries were prepared using the TruSeq Stranded mRNA Library Prep Kit (Illumina). Sequencing was performed on the Illumina NextSeq 500 system, utilizing single-end sequencing with a read length of 75 base pairs. MT sequencing was conducted after library preparation using the Zymo-Seq RiboFree Total RNA Kit, and sequencing was carried out on the Illumina NextSeq 2000 system with 2×150 bp paired-end reads. MG sequencing used the Illumina HiSeq 1000, with roughly 25 million reads per sample, using 2×150 bp paired-end reads. Quality assessment of the MG and MT sequenced reads, along with their corresponding bases, was conducted using FastQC v0.12.1. In our case, all reads and bases were within an acceptable quality range, with Phred scores consistently exceeding 30, indicating high-quality sequencing data.

#### Cecal mRNA Data Processing

For gene expression analyses, the cecal mRNA sequencing data were aligned and quantified to generate per-gene unique read counts using the R package `Rsubread` v2.12.3 [120], resulting in a raw gene count matrix. This raw count matrix was used for DEG analysis in `DESeq2` v1.44.0 to identify transcriptional changes associated with diet and age, and the resulting $\log_2$(fold changes) were used for GSEA. In parallel, raw counts were also converted into a TPM (Transcripts Per Million) matrix using the package `tximport` v1.26.1 [121], followed by log transformation as $\log_2(\text{TPM} + 1)$ for use in machine learning models and WGCNA. This dual processing strategy enabled robust DEG identification while facilitating pathway-level interpretation and comparison across samples.

#### MG and MT Data Processing

All raw paired-end MG and MT sequence reads were quality-processed using KneadData v0.10.0 [99]. Sequence trimming was performed using KneadData, which ran Trimmomatic v0.39 to trim regions where base quality fell below Q20 within a 4-nucleotide sliding window, and to remove reads that were truncated by more than 50% (parameters: SLIDINGWINDOW:4:20, MINLEN:50).

Additionally, KneadData was used to run Tandem Repeat Finder (TRF) v4.09.1, to locate and remove reads characterized by tandem repeats. Bowtie2 v2.4.2, also employed within KneadData, was used to deplete host-derived sequences from the MG data using the mouse (C57BL) reference database and from the MT data using both the mouse (B6) reference database and the SILVA Ribosomal RNA reference database. After processing with KneadData, the remaining non-host reads were used as input to obtain the taxonomic profiles of the samples using MetaPhlAn v4.0.6 and the CHOCOPhlAn (mpa_vOct22_CHOCOPhlAnSGB_202212) database. Functional profiling was subsequently obtained through the HUMAnN v3.8 workflow.

### 3.5.5 Miscellaneous Bioinformatics and Statistical Analysis

All analyses were performed using R 4.2.2. For exploratory data visualization, differential expression analysis, statistical modeling, and multivariate analyses (e.g., ANOVA and PCA), we used standard R packages including `ggplot2`, `vegan`. To identify modules of co-varying host genes and microbial taxa, we performed WGCNA using the R package: `WGCNA`. Cross-omics edges were defined by Spearman correlation coefficients $|R| > 0.6$ and FDR-adjusted $p < 0.05$. Significant gene–microbe interactions were exported as edge lists and visualized using Cytoscape (v3.10.3) [122].

#### Multivariable and Cross-Omics Association Analysis

To identify microbial features associated with metadata variables such as diet, age, and strain, we applied MaAsLin3 (v1.10.0) [107], a multivariable linear modeling framework suitable for compositional microbiome data. Relative abundance tables (species- or genus- level) from MetaPhlAn4 were used as input. We applied arcsine square root transformation to the abundance data and used diet, age, and strain as fixed effects. False discovery rate (FDR) was controlled using the Benjamini-Hochberg procedure, and features with FDR $< 0.05$ were considered statistically significant.

To explore cross-modal associations between microbial species and host liver proteomic features, we used the HAllA framework [113]. HAllA identifies statistically significant associations between high-dimensional datasets by combining hierarchical clustering and association testing across feature blocks. Spearman correlations were calculated for each species–protein pair, and the significance of associations was determined using a permutation-based FDR threshold of 0.05.

### 3.5.6 Supervised Learning and Model Evaluation

All prediction models were implemented using Python 3.9.18 with standard open-source libraries, including `pandas` (v2.1.4), `numpy` (v1.23.5), `scikit-learn` (v1.3.0), and `matplotlib` (v3.8.0). We trained multiple supervised classifiers including RF, XGB, Logistic, and DNN to predict host phenotypes (e.g., age and body weight) based on gut microbial composition. Model performance was evaluated using five-fold cross-validation and ROC curves. Feature importance scores were extracted from tree-based models (RF and XGB) to rank the contribution of individual microbial species or protein features. To mitigate overfitting and improve interpretability, we applied iterative feature

selection based on model-derived importance scores. To further capture potential nonlinear or combinatorial effects between microbial features, polynomial feature interaction terms were generated using the `PolynomialFeatures` module from `scikit-learn`. In multi-omics integration analyses, classifiers were trained using top-ranked features from each omics layer and evaluated using a compact, 10-feature integrated model. Model optimization was guided by cross-validation accuracy and AUC metrics across independent data batches.

## Code Availability

Codes and scripts developed in this study are all available at the GitHub repository (`https://github.com/zzhougithub/GEM_Microbiome`).

## Data availability

The raw sequence data are available in the NCBI Bioproject repository under the accession number PRJNA953833 (data will be released upon article publication).

## Supplemental Information

Supplemental Information contains seven figures.

## Acknowledgments

We would like to acknowledge the helps of Kurt Lamour, Xiaowei Zhan, Jack Haverty, Susheel Bhanu, and Julien Schleich. The mRNA and MT sequencing experiments presented in this paper were carried out at the LCSB sequencing platform (RRID: SCR_021931) at the University of Luxembourg. The processing of sequencing raw data presented in this paper were carried out using the HPC facilities of the University of Luxembourg [123]. This project is implemented with the support of the Fondation du Pélican de Mie et Pierre Hippert-Faber, under the aegis of the Fondation de Luxembourg.

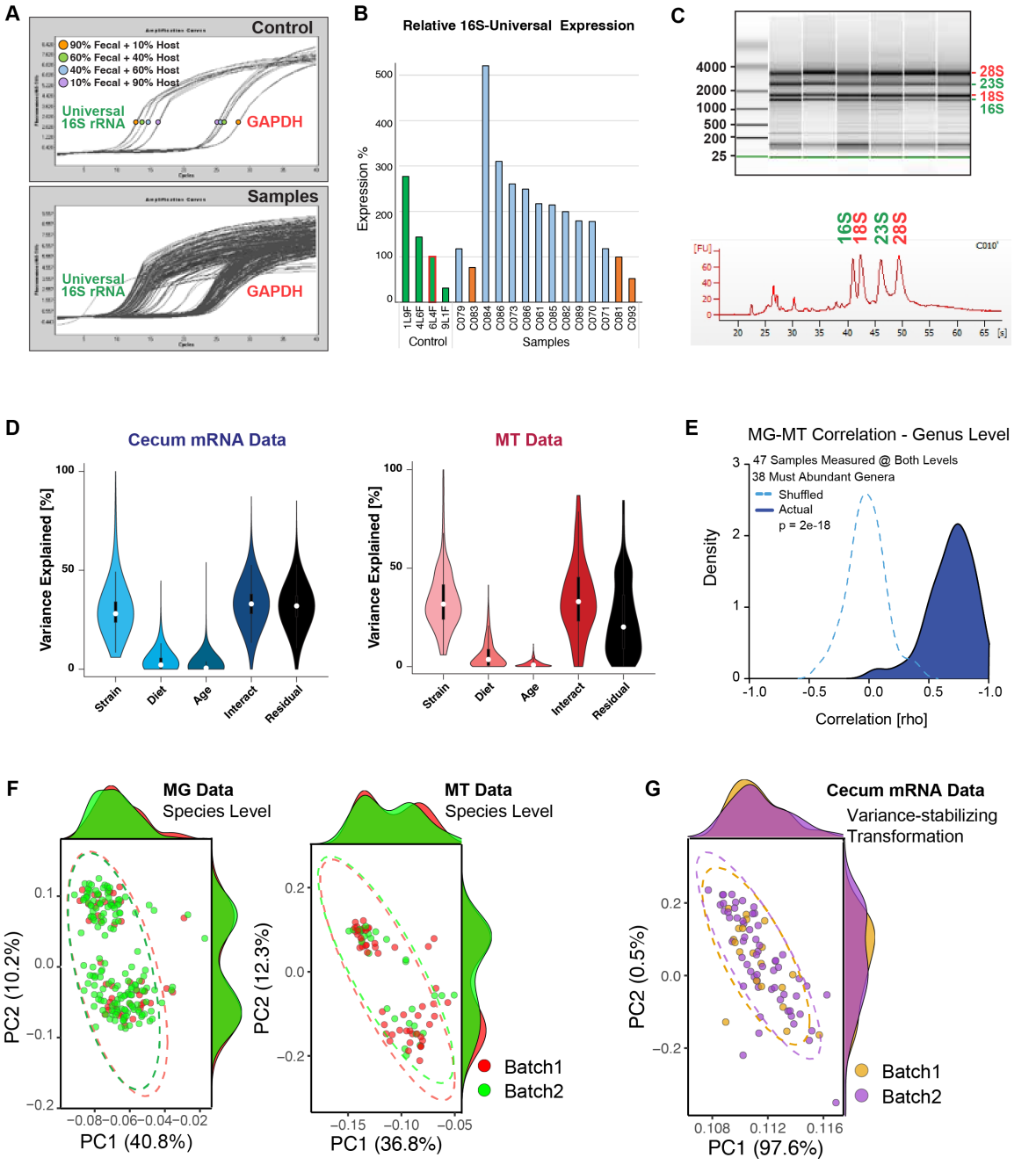


**Figure S3.1** Establishment of Experimental Methods and Quality Control. (**A**) A method based on RT-qPCR was established to assess the ratio of host to gut microbiota DNA in samples. (**B**) Samples with gut microbial DNA content estimated to be above ~ 40% were selected based on gradient controls. (**C**) Bioanalyzer results of RNA samples were used to select samples with four clear bands and no significant degradation for subsequent sequencing. (**D**) ANOVA analysis of cecal mRNA and MT data showing variation explained. HF feeding reduces gut microbial diversity across all levels of taxonomy. (**E**) Correlation analysis indicates a strong correlation between the MG and MT data. (**F**) PCA plots of MG and MT at the species level show no dominant batch effect. (**G**) PCA based on variance-stabilized cecal mRNA expression data shows that samples do not cluster by batch, indicating no evident batch effect. In this data, the impact of diet is not immediately evident from the data cloud.

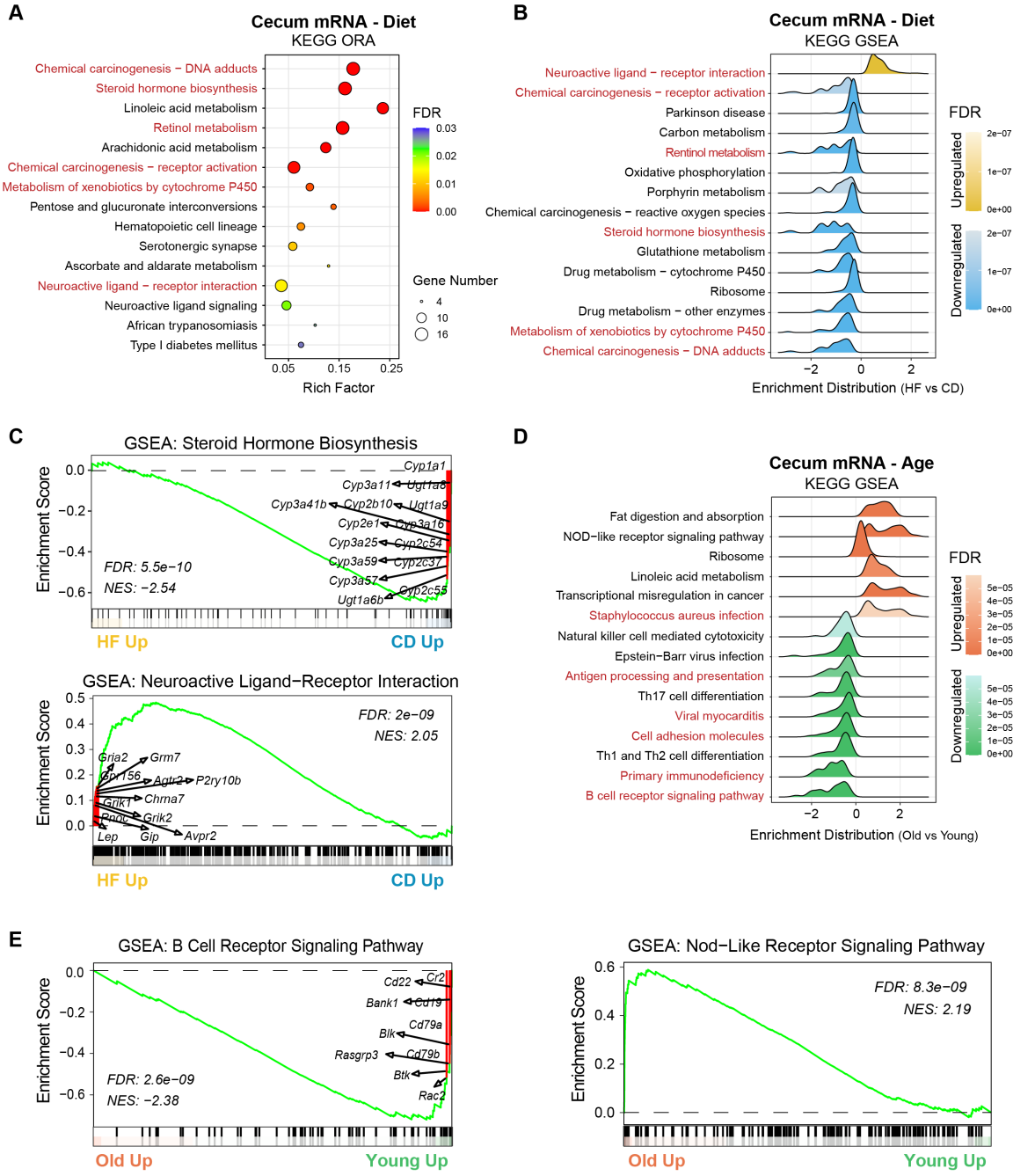


**Figure S3.2** Differential Pathway Enrichment in Cecal Transcriptomes by Diet and Age. (**A**) KEGG ORA results showing the top 15 enriched pathways. Bubble size indicates the number of DEGs per pathway; bubble color represents FDR; the x-axis shows Rich Factor. (**B**) KEGG GSEA for the same diet comparison, displaying the top 15 pathways ranked by FDR. (**C**) Representative GSEA enrichment plots for diet: Steroid hormone biosynthesis (upper; enriched in CD) and Neuroactive ligand–receptor interaction (lower; enriched in HF). Marked genes are DEGs present in the respective pathways. (**D**) KEGG GSEA results showing the top 15 age-associated pathways. (**E**) Two GSEA plots for age: B cell receptor signaling (left; enriched in Young) and NOD-like receptor signaling (right; enriched in Old).

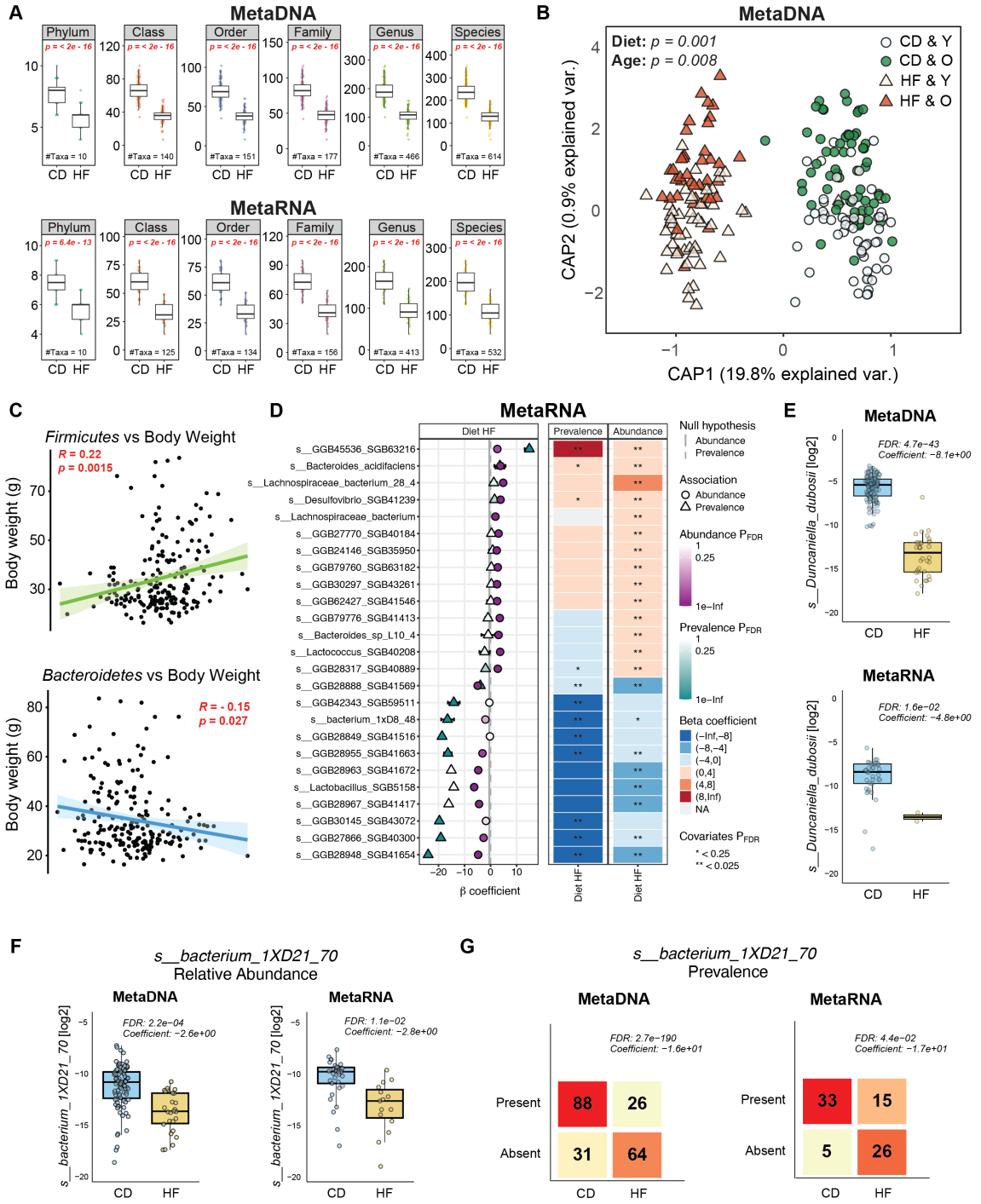


**Figure S3.3** Supplementary Analysis of MG and MT Data. (**A**) Total number of detected gut microbial taxa across all taxonomic levels. HF significantly reduced microbial richness at each taxonomic level. (**B**) Distance-based redundancy analysis(dbRDA) based on Bray–Curtis dissimilarity showing gut microbial community structure across diet and age groups (CD-Young, CD-Old, HF-Young, HF-Old). *P*-values from PERMANOVA (999 permutations). (**C**) *Firmicutes* and *Bacteroidetes* phyla are significantly associated with body weight. (**D**) Coefficient plot and heatmap of the most significant associated species in MT data between different diets. (**E**) *s__Duncaniella_dubosii*. The relative abundance (**F**) and prevalence (**G**) of *s__bacterium_1XD21_70* across different diets.

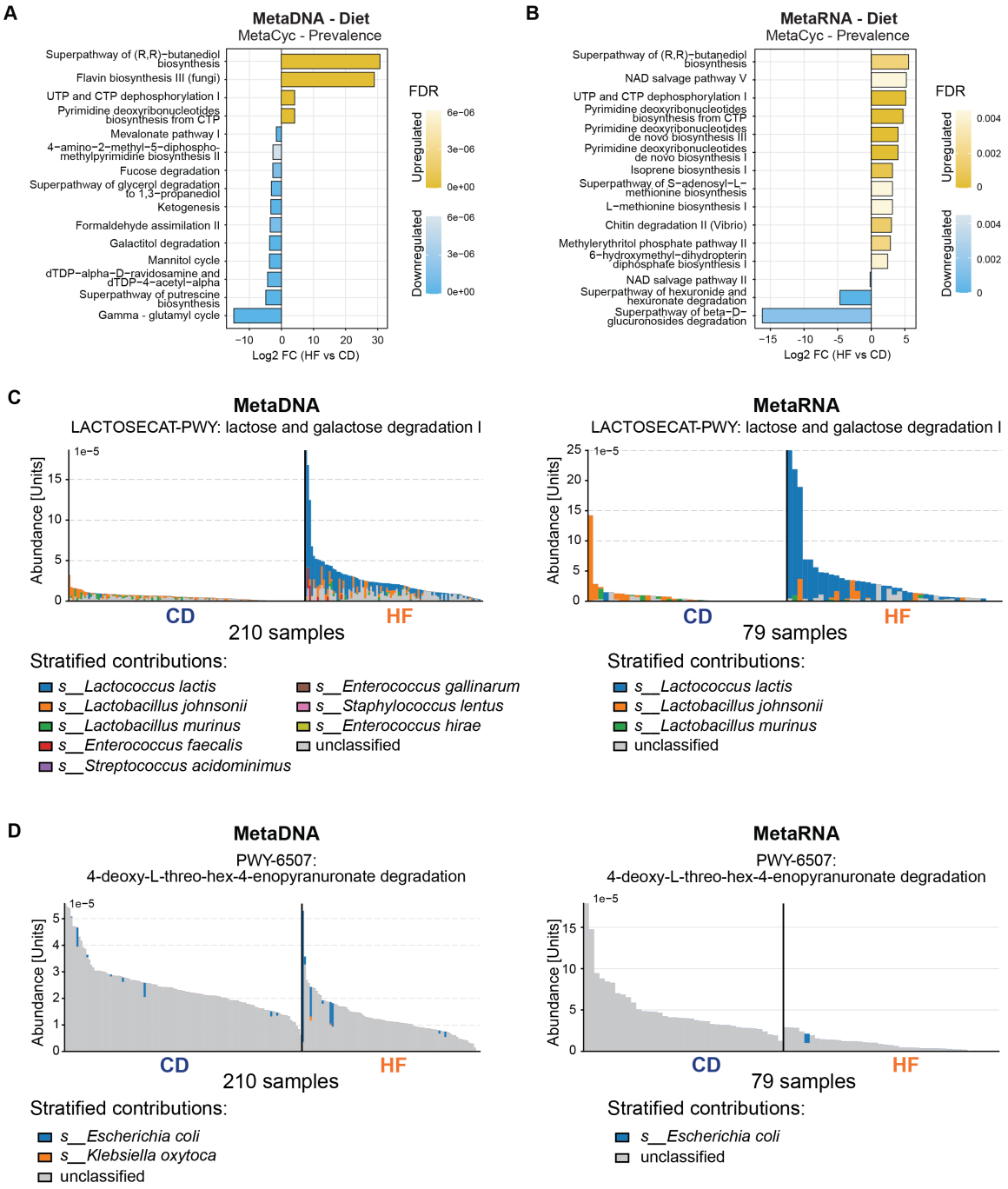


**Figure S3.4** Supplementary Details of Diet-associated Changes in Microbial Functional Profiles. (**A-B**) Top 15 MetaCyc pathways with significant dietary differences in pathway prevalence from MG data(**A**) and MT data(**B**). (**C-D**) Relative abundance distributions of the lactose and galactose degradation I pathway (LACTOSECAT-PWY) (**C**) and the 4-deoxy-L-threo-hex-4-enopyranuronate degradation pathway (PWY-6507) (**D**) in the MG and MT datasets, stratified by major contributing species.

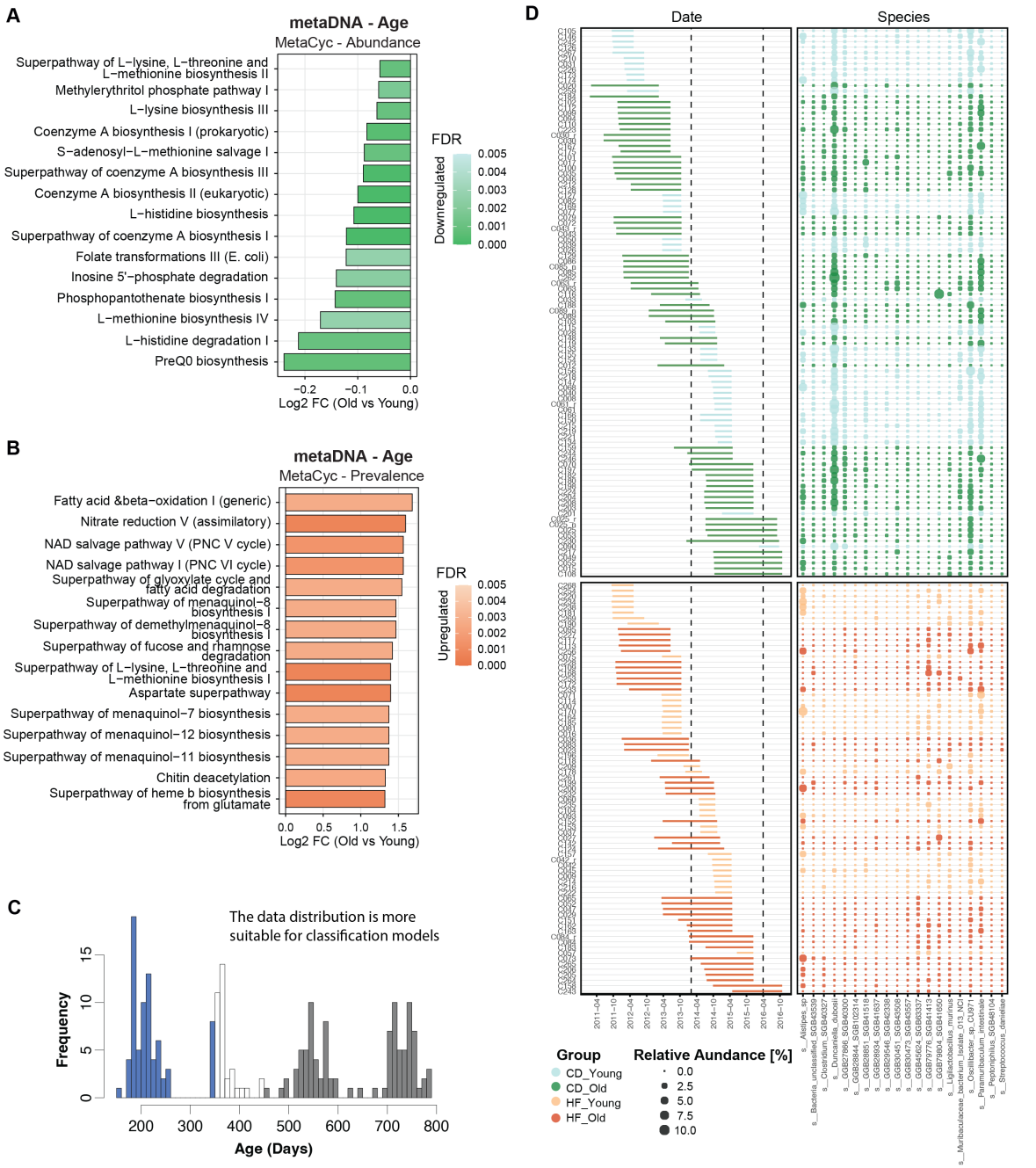


**Figure S3.5** Supplementary Details Related to Age-Associated Analyses. (**A-B**) Abundance and prevalence of the top 15 age-associated MetaCyc pathways in MG data. Bars show $\log_2$FC (Old vs Young). (**C**) Distribution of sample ages across the MG cohort. Young and old groups were defined to ensure comparable sample sizes. (**D**) Combined plot of the relative abundances of the top 20 species identified by Random Forest model and the experimental time points of the samples.

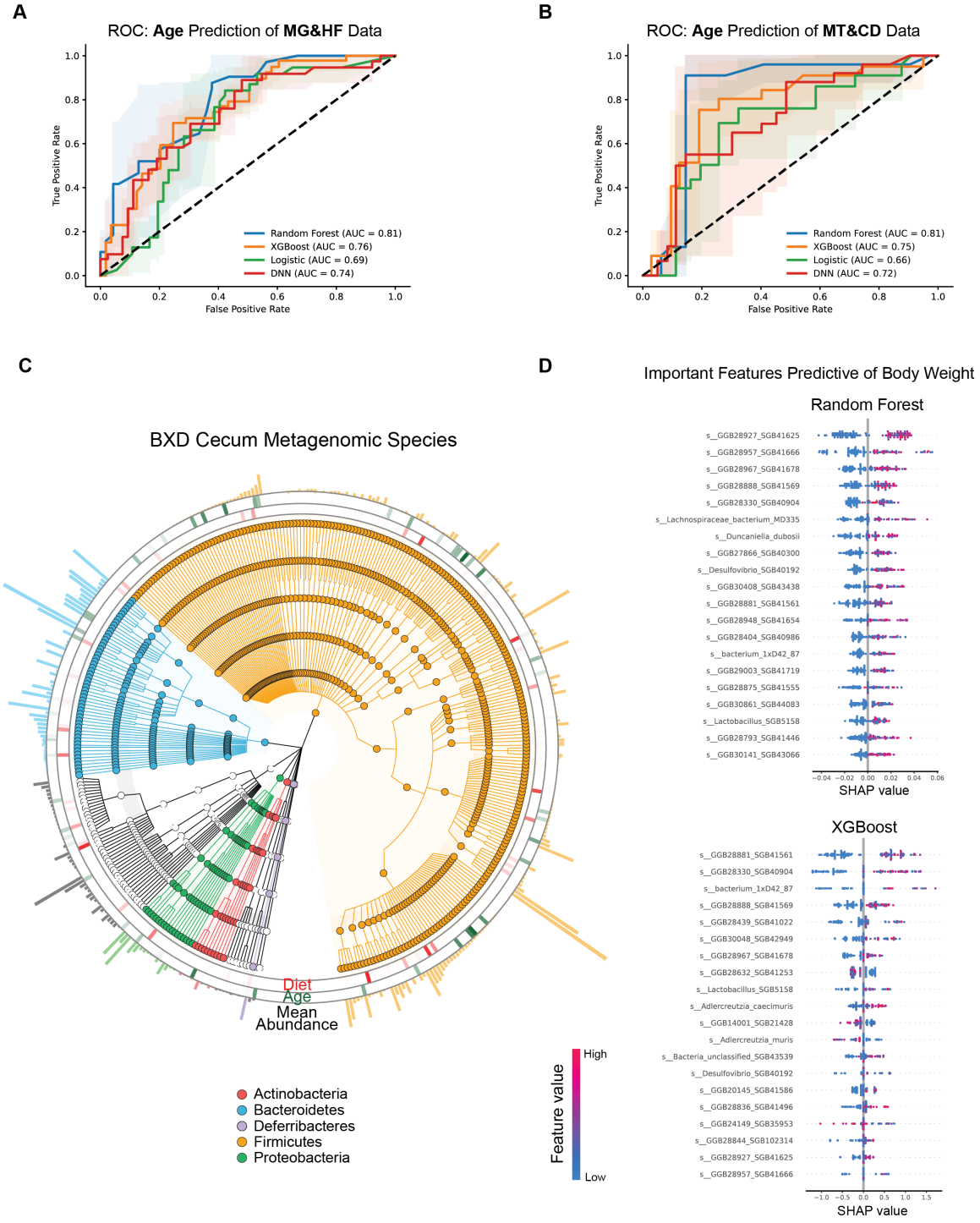


**Figure S3.6** Machine Learning Models Predicting Age or Body Weight Using Multi-Omics Data. (**A**) ROC curve of the age prediction model for the HF group in the MG data. (**B**) ROC curve of the age prediction model for the CD group in the MT data. (**C**) Cladogram of the MG. Outer ring 1 represents species with significant differences across different diets. Outer ring 2 represents species with significant differences across different ages. Outer ring 3 indicates the mean relative abundance of each species. (**D**) Top 20 biomarkers selected by RF and XGB models for body weight prediction using MG data.

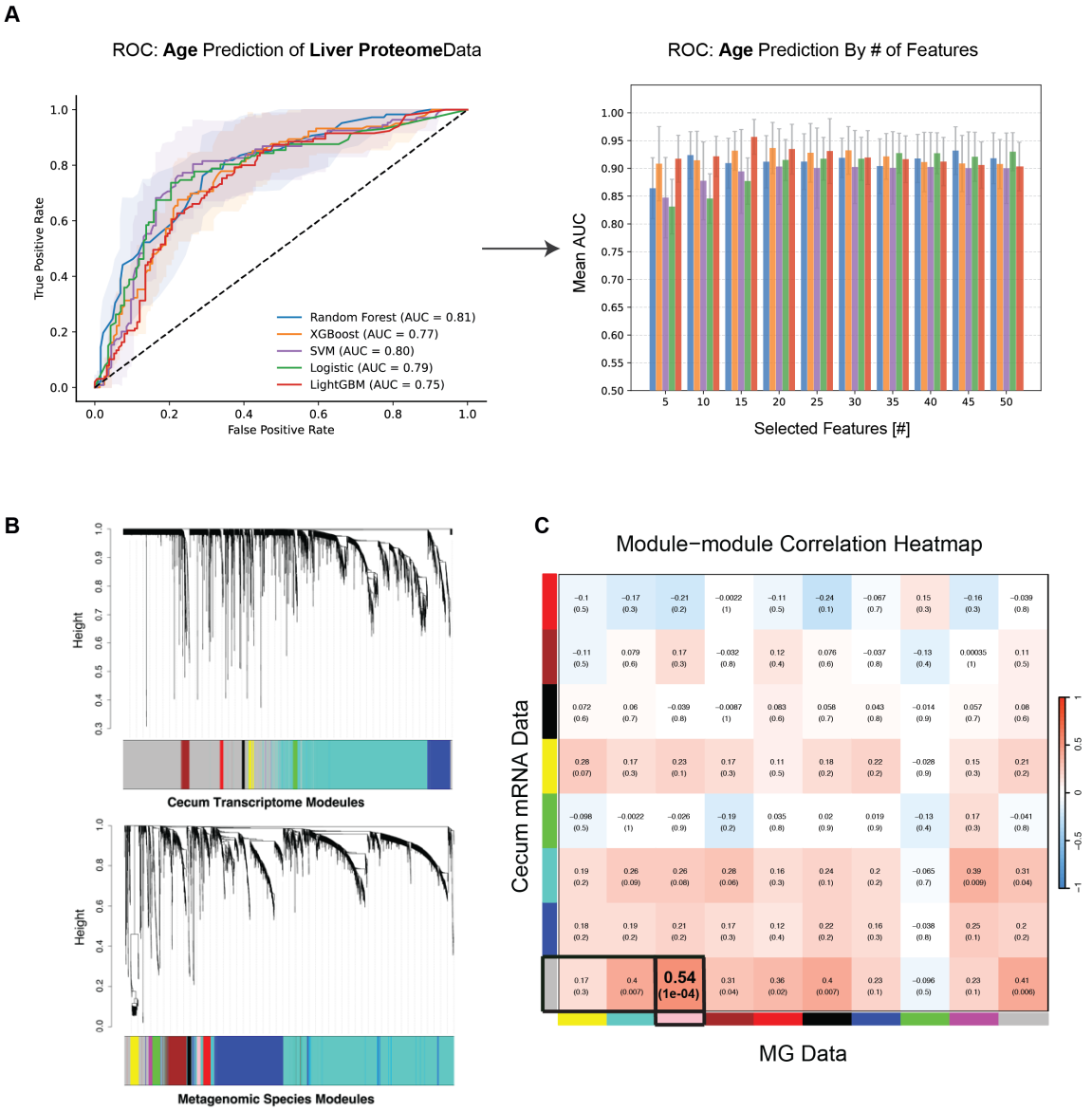


**Figure S3.7** Supporting Results from Integrated Multi-omics and Network Analyses. (**A**) ROC curve and feature selection bar plot of the age prediction model for the CD group based on liver proteomics data. (**B**) Hierarchical clustering dendrograms of cecal mRNA (left panel) and MG (right panel), showing co-expression modules identified by WGCNA. (**C**) Module–module correlation heatmap between cecal mRNA modules and MG species modules. The strongest positive correlation was observed between MEgrey (cecum) and MEpink (microbiome) modules (r = 0.54, FDR = 1e-04).

# Chapter 4

# Manuscript 2

## Voluntary Exercise Differentially Affects High-Fat Induced Metabolic Phenotypes Across BXD Mouse Strains

## Preface

**Prof. Evan G. Williams** and **Dr. Djalil Coowar** were responsible for the initial conceptualization and study design of the project; The mouse experiments were conducted by **Evan**, **Ziyun**, and **Arianna** (Batch1: **Evan** and **Ziyun**; Batch2: **Ziyun**; Batch3: **Ziyun** and **Arianna**; Batch4: **Arianna**); **Ziyun** developed the code for processing wheel output data; **Besma** developed the code for handling DVC data; **All authors** contributed to data analysis; The dissection of mice was performed by **Evan**, **Ziyun**, **Arianna**, **Manuel**, **Tony**, and **Pierre**; IPGTT was conducted by **Evan**, **Ziyun**, **Arianna**, **Jennifer**; 10 cecum DNA samples was extracted by **Arianna** using the protocols established by **Ziyun** in Chapter 3.

*In preparation for publication*

# Voluntary Exercise Differentially Affects High-Fat Induced Metabolic Phenotypes Across BXD Mouse Strains

**Ziyun Zhou[1,#], Arianna Lamanna[1,#], Besma Boussoufa[1], Pierre Garcia[1], Djalil Coowar[1], Evan G. Williams[1,*]**

**1** Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Esch-sur-Alzette 4362, Luxembourg

# These authors contributed equally

* Correspondence to evan.williams@uni.lu

**Keywords:**
Long-Term Voluntary Exercise; High Fat Diet; Diabetes; Obesity; Mouse Populations; Genetic Diversity; Home Cage Monitoring; Physical Endurance

*In preparation for publication*

## 4.1 Abstract

Obesity and metabolic diseases such as diabetes are driven by long-term interactions between key variables such as diet, lifestyle, and genetic predisposition. Due to the difficulty of measuring and controlling these parameters long-term in humans, mouse models are used to study the variables' effect sizes. However, modelling behavioral factors remains difficult, particularly long-term exercise. We previously phenotyped ~50 different strains of the BXD mouse population, identifying strains BXD 40, 43, and 100, that are predisposed to voluntarily use exercise wheels (~6 km/day at 24 weeks of age), and which have variable incidence of diet-induced obesity and diabetes. Here, we follow these strains from 8 to 32 weeks of age, fed 60% kcal/fat high fat diet in either standard home cage monitoring (without wheel), or with *ad libitum* access to monitored running wheels. Running cohorts had diminishing usage of the wheel over time (from ~9 km/day in the first two weeks to a steady ~5 km/day for the last 24 weeks), but this wheel usage was enough to improve response to an oral glucose tolerance test by 35% compared to sedentary controls by 24 weeks of age. While both sedentary and running cohorts had equivalent body weight, running cohorts increased food intake vs. body weight by 20%, indicating improved metabolic health despite higher caloric intake. Finally, in this study we have validated 3 inbred mouse strains that — together with the canonical C57BL/6 — can be used to further study interactions between diet, exercise and genetic background to influence the development of metabolic diseases.

## 4.2 Introduction

Decades of scientific research have shown a clear, causal connection between excess caloric intake and excess weight gain – often followed by numerous metabolic diseases such as type II diabetes (T2D), hypertension, and nonalcoholic fatty liver disease (NAFLD). The precise contribution and ratios of the different dietary components leading to these effects remains in debate (e.g. the intake of sucrose, saturated fats, carbohydrates, etc), as is the extent to which moderate exercise mitigates the development of metabolic diseases. In human cohorts, weight loss successfully achieved through dieting is typically reversed within 2-5 years [124], and while aerobic exercise is indeed beneficial for health, this tends to have only a rather small effect (< 5%) on individuals' body weights [125]. Weight gain towards obesity, – and the strongly-associated consequent metabolic diseases – are not simply driven by a simple, calculable equation between caloric intake versus calories burned, as complex genetic and molecular pathways interact to prevent simple lifestyle changes, such as sustainable diets and exercise regimens, from being sufficient for many people. Due to the long-term nature of metabolic syndrome and the myriad complex and difficult–to–control cofactors, we can use diverse sets of inbred mice strains to determine the effect size of interventions such as diet and long-term exercise on metabolic disease. Over nearly a century of such studies, researchers have identified and created numerous diets which lead to metabolic syndrome in standard inbred laboratory mouse strains, and likewise scientists have identified and created numerous genetic variants which lead to metabolic syndrome, regardless of the intake of a "healthy" or "unhealthy" diet. That is, obesity for most individuals comes from a mixture of factors both nature and nurture, referred to as gene–by–environment (GxE) interactions. In natural populations (such as humans), singular genetic or environmental causes of obesity and metabolic syndrome are rarely observed, thus it is essential to model the complexity of human metabolism using a controllable model system.

Studies in mice that do not have monogenic causes of obesity (i.e. not db/db or ob/ob mice) are overwhelmingly performed in the C57BL/6 (B6) strain, and they have broadly shown that medium-duration forced exercise on treadmills will induce mice to lose weight and improve glucose tolerance when subjected to ad libitum high fat diet (HF) (e.g. 20 minutes/day for a 12 week regiment [126]). However, treadmill experiments also subject mice to significant levels of psychological stress unlike voluntary wheel access [127], and psychological stress in mice also reduces body weight and improves glucose tolerance [128]. To avoid this confounding factor, studies on exercising mice have provided voluntary access to running wheels over similar 12-week periods, yet these studies have had conflicting results: some show that voluntary exercise does mitigate weight gain on HF (e.g. [129]), while others show that it does not (e.g. [130]). As we know that genetic background and the amount of exercise play a role in the development of obesity and diabetes, we hypothesized that other inbred strains may provide new insight into these interactions and their effects. We have previously performed metabolic phenotyping on around 50 different recombinant inbred strains from the BXD murine genetic reference population [19]. These strains have wide variance in their natural propensity to exercise, their body weight, susceptibility to HF-induced obesity and glucose intoler-

ance, and many other key parameters such as expected lifespan [88]. While mice which voluntarily exercised could reverse some of the weight gain induced by HF, our earlier study only permitted 10 days of running activity before mice were returned to standard (wheel-less) housing. Mice are widely known to have rapidly diminishing usage of running wheels after around 2 weeks, often dropping to a steady rate of around $1/3^{rd}$ their initial use when the wheel is first introduced (e.g. [130]).

One additional confounding factor into using mice to observe the contribution of voluntary exercise to metabolic effects is that an exercise wheel may not be strictly necessary for exercise: mice also have sufficient space in a normal housing cage to get low-intensity exercise by simply walking around. Developments in home cage monitoring (HCM) over the past 15 years have led to the implementation of mouse housing that can record mice without the necessity of surgically-implanted tracking devices. Early such technologies used invisible laser beam grids that recorded animal movement, but this resulted in limitations in the use of normal housing material, such as bedding/nesting, which could disrupt the sensors. More recent HCM devices, such as the Tecniplast Digital Ventilated Cage (DVC) used in this study, allow animal activity to be tracked via sensors detecting perturbations in electromagnetic fields caused by animals' movements [131]. These newer systems allow animals to be tracked without implants and in completely standard housing, including normal bedding and enrichment. Thus, it is now possible to continuously record the typical activity of mice throughout their lives.

In this study, we have followed 128 individual mice from 3 BXD strains: BXD40, 43, and 100, on *ad libitum* 60% kcal/fat (5.1 kcal/g) HF from 8 to 32 weeks of age, with cohorts in cages with access to monitored running wheels, or in HCM without access to wheels. Food intake and weight were monitored weekly, and oral glucose response was measured at 24 weeks of age. While both genetic background and exercise played a role in glucose response, we observed that in these lines, increased physical exercise led to an increased caloric intake, and little impact on body weight – although exercise did substantially improve glucose response. Identifying and studying these differences is key to understanding how exercise affects obesity and related fitness parameters in humans: these strains are different from one another and from the canonical B6 strain, just as are humans different from one-another. These three strains, which we have previously phenotyped in short-duration exercise, demonstrate models that are suited for identifying the diversity of cellular mechanisms and responses driving the variety of outcomes in genotype, diet, and exercise-dependent traits that occur across populations.

## 4.3 Results

### 4.3.1 Highly Heritable Fitness Parameters Used to Select BXDs 40, 43, and 100 For Long-Term Exercise Phenotyping

To select the strains used for this study, we re-examined the data from our large BXD screening study [19], where 57 strains were phenotyped on either Harlan Chow (CD) or Harlan 06414 60% HF diets from 8 to 29 weeks of age, with 10 days of access to a running wheel from 22-24 weeks of age while solitarily housed. We have observed that the heritability (h2) of these key metabolic traits is very high: body weight h2 = 0.76, voluntary exercise h2 = 0.64, and oral glucose tolerance test response h2 = 0.82. Given the high heritability of these traits, we expected to be able to reproduce and build on these prior findings. To do so, we selected three strains – BXD40, 43, and 100 – which had radically different propensity to develop obesity and metabolic disease when fed HF (Fig. 4.1 A). BXD40's body weight is nearly unaffected by HF (27 vs 32g), while BXD43 is lean on CD while obese on HF (33g vs. 50g), and BXD100 is large on CD and obese on HF (42g vs. 54g). Furthermore, all three strains were excellent runners: BXD100 runs around 4 km/night, BXD40 6.5 km/night, and BXD43 7.7 km/night (Fig. 4.1 B), and diet had no overall effect on exercise wheel usage. The mice in this prior study of ours only had 10 days of access to a running wheel over 22 weeks of HF, thus HF-fed BXD40s did not remain thin due to their wheel usage, so we first hypothesized that differences in normal home cage movement behavior (monitored over 48 hours at 24 weeks of age) may be compensating. However, both BXD40 and BXD43 were exceptionally active, while BXD100 (as with its wheel usage) was only somewhat more active than the average strain (Fig. 4.1 C). Despite these pronounced exercise traits, all three strains had very average glucose responses in both dietary conditions, with a mild propensity for HF-induced glucose intolerance at 17 weeks of age, after 9 weeks of HF feeding (Fig. S4.1 A). We also previously observed that 10 days of running can reduce body weight by around 10% in both dietary cohorts (Fig. S4.1 B). With these prior results in mind, we designed a study to allow us to continuously monitor normal movement in a "sedentary" HCM cohort, while an active "running" cohort would have ad libitum access to a recorded running wheel, on the same 60% kcal/fat HF, and over a similar range of ages (8-32 weeks old) (Fig. 4.1 D). In this study, we measured body weight and food intake weekly during this period, and the results of an intraperitoneal glucose tolerance test (IPGTT) at 24 weeks of age. We hypothesized that mice resistant to the negative effects of HF may not be directly genetically resistant to the diet, but instead may continue to exercise by running around their cage-albeit likely less than for mice with access to a running wheel.

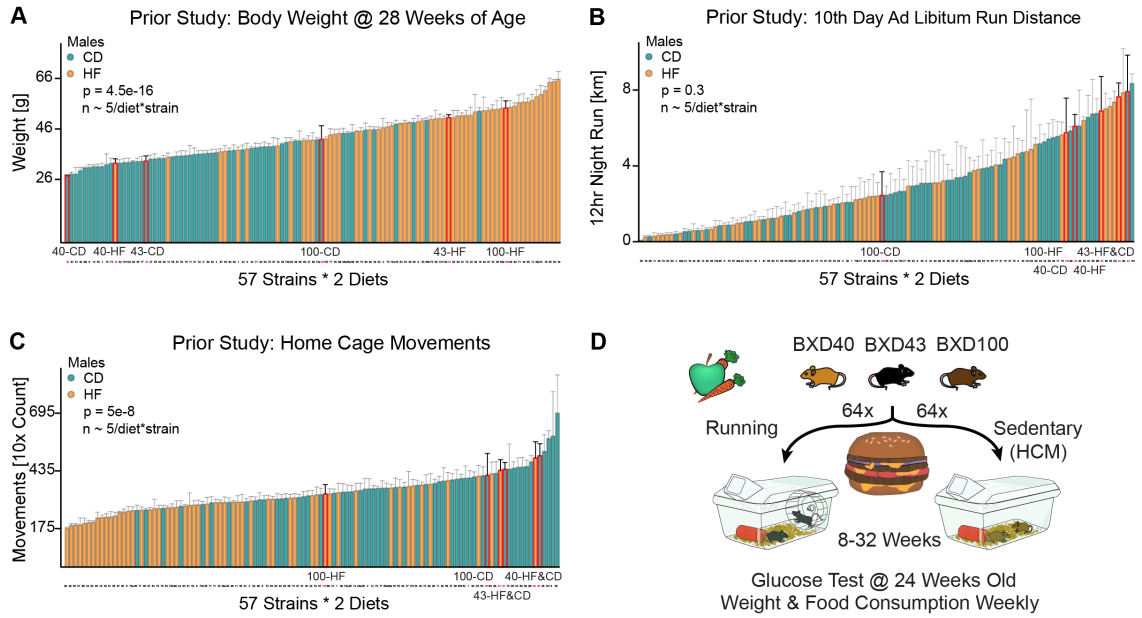


**Figure 4.1** BXD Strain Selection, Hypothesis Development, and Study Design. (**A**) Body weight at 28 weeks of age (20 weeks of CD or HF feeding) in a prior study on 57 BXD strains. (**B**) The final cumulative daily run for a solitary mouse with an exercise wheel after nine days of habituation, in the same BXD cohort at around 24 weeks of age. (**C**) HCM over 24 hours for a solitary mouse after one day of habituation, in the same cohort. (**D**) Study design using a subset of the previous strains on HF with continuous access to a running wheel or HCM activity monitoring.

### 4.3.2 Voluntary Exercise Only Slightly Improves Glucose Response and Lowers Body Weight Gain in HF Cohorts

While we knew from our prior 10-day exercise study that the BXD40, 43, and 100 (as well as B6) will use a running wheel voluntarily, we also knew from prior literature that B6 mice reduce their activity wheel usage over time, typically dropping by half from around ~10 km/day at ~10 weeks of age to ~5 km/day by 20 weeks of age [132]. Thus, we first needed to validate that our strains would continue to voluntarily exercise throughout the duration of the experiment. We indeed observe that average weekly running distance dropped by 60% from 64 to 25 km/week/cage between 9 weeks versus 29 weeks of age ($p = 3e\text{-}6$; Fig. 4.2 A). This decline in running is consistent for nearly all running mice: of the 46 running mice, all but 2 reduced their median running activity in the first 10 weeks of the study, compared to their median running in the following 10 weeks, dropping by 43% ($p = 2e\text{-}4$; Fig. 4.2 C). Due to space and equipment limitations, only 16 mice could be run per cohort, but we observed no difference in running distance between the three cohorts (ANOVA, $p = 0.16$, Fig. 4.2 D). Compared to the decrease in activity wheel usage over time, we expected a less significant effect in normal home cage over time, based on prior literature showing only a slight decline with time [133].

Indeed, we observed that our sedentary cohorts moved only 27% less between from 14 weeks to 29 weeks of age ($p = 0.01$, Fig. 4.2 B). (Note that 14 weeks is used as the baseline for HCM, as the first 6 weeks of DVC data were lost for Cohort 1.) This decline in movement at the young age also quickly stabilized, with mice not having any further decline in HCM activity between 17 ± 3 weeks to 26 ± 3 weeks of age ($p = 0.51$, Fig. 4.2 E). In contrast to the activity wheel, we did observe differences across cohorts for their overall home cage activity, with a significant elevation in Cohort 3, which moved nearly 50% more in HCM than did the first two cohorts ($p < 2e-16$, Fig. 4.2 F).

Given that the mice exercised, we expected to identify significant changes in the weight gain and early indicators of metabolic syndrome, such as hyperglycemia during the IPGTT. Indeed, during the IPGTT experiment at 24 weeks of age, sedentary mice had at least a 35% worse glucose response curve than the running cohort ($p = 2e-5$, Fig. 4.2 G). To note, this difference is likely higher, but the glucometer used has a maximum value of 600 mg/dL, which was reached at 30 minutes by 8 running mice, compared to 24 sedentary mice. We next checked the overall change in body weight from 8 to 32 weeks of age and observed only a 12% increase in body weight for sedentary mice compared to running cohorts ($p = 0.04$, Fig. 4.2 H). This was initially a surprise, given the significant use of the running wheel, and we hypothesized that the running mice may have increased their food intake. This was indeed the case, and running mice ate about 11% more than the sedentary cohort ($p = 0.01$, Fig. 4.2 I). Factoring the smaller size of the running mice, we observed that the running mice ate 20% calories compared to their body weight versus the sedentary mice ($p = 1e-6$, Fig. 4.2 J), in line with expectations from literature [134].

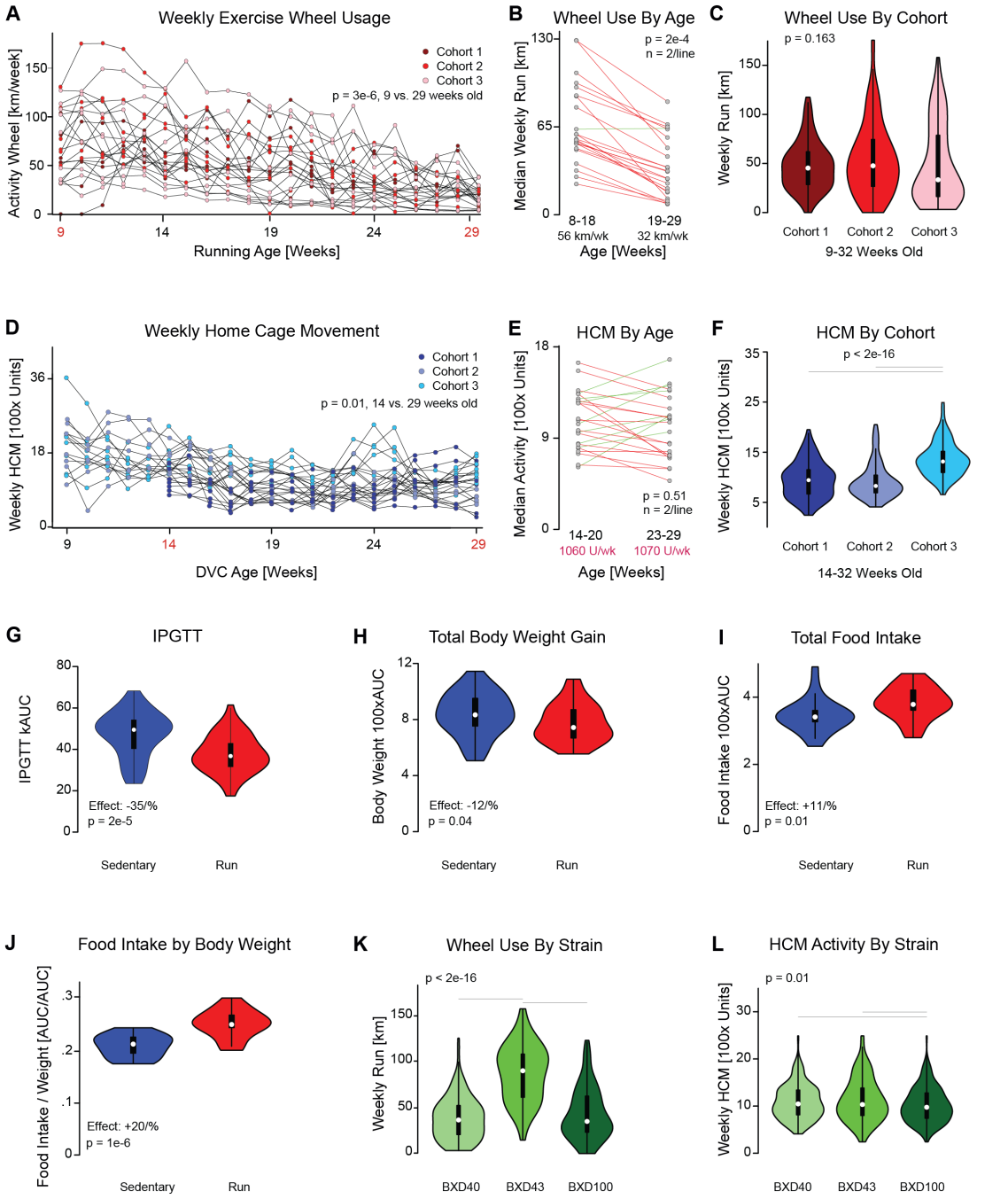


**Figure 4.2** Exercise and Home Cage Monitoring in the BXDs. (**A**) Weekly total activity wheel usage for all running cohorts across time. (**B**) Slopegraph comparison of median movement between weeks 8-18 versus 19-29 for all running cohorts. (**C**) Violin plot of running distance for all mice in each cohort. (**D**) HCM total activity for all sedentary cohorts across time. (**E**) Slopegraph comparison of median movement between weeks 8-18 versus 19-29 for all sedentary cohorts. (**F**) Violin plot of HCM activity for all mice in each cohort.(**G**) IPGTT glucose response decrease of 35% in running cohorts vs. sedentary cohorts. (**H**) Body weight decreased by 12% in running cohorts vs. sedentary cohorts. (**I**) Running mice ate around 11% more than sedentary mice, (**J**) or around 20% more than sedentary mice if corrected for the slightly decreased body weight of running mice. (**K**) BXD43 ran more than twice as much as the other two strains, but (**L**) HCM activity was roughly constant across all three lines (6% decrease for BXD100).

While *ad libitum* voluntary exercise led to a significant, although not massive, improvement in metabolic fitness, we expected that genetic background would play a role in these outcomes. For instance, BXD43 ran a median of 90 km/week throughout the experiment, far exceeding the average of the other strains' combined median of 38 km/week ($p = 1e - 20$, Fig. 4.2 K). Conversely, home cage activity of all three strains were nearly equivalent, with only BXD100 being very marginally less active than the other two strains (- 6%; $p = 0.1$, Fig. 4.2 L). Thus, we overall confirmed that BXD43 is a highly active mouse model for voluntary exercise, and that the high exercise wheel usage does not correspond to a higher basal activity in the absence of a wheel. We thus selected three avenues of analysis to further pursue this finding. First, do most active running mice perform metabolically better than the least active, and is there any connection between HCM activity and metabolic fitness? Second, by testing a fourth cohort using only BXD43, to increase the sample size and replicate the main finding, do we confirm that strain's particular connection between exercise, glucose response, weight, and food intake? Finally, can we identify cofactors which indicate the causal mechanisms, strength, or predictive variables connecting genetic background, diet, and exercise to metabolic disease outcomes?

### 4.3.3 Increased Activity Does Not Strongly Predict Improved Metabolic Outcomes

We have identified how genetic predisposition (Fig. 4.1 A-C) and voluntary exercise (Fig. 4.2 G-J) modulate the metabolic effects of HF diets. However, in human cohorts, it remains unclear whether short-duration daily exercise (e.g., 30 minutes per day) is sufficient to prevent metabolic syndrome, or whether there exists a dose-dependent relationship in which greater exercise duration linearly reduces the risk of obesity and diabetes. In our mouse model, despite substantial voluntary wheel running, we observed only a modest reduction in weight gain (Fig. 4.2 H) and glycemic improvement (Fig. 4.2 G). Additionally, we observed a increase in food intake among runners (Fig. 4.2 I-J), which likely reflects a compensatory response to elevated energy expenditure—a well-documented phenomenon in both rodent and human exercise studies [135]. Such hyperphagic responses may further blunt the weight-reducing effects of exercise, underscoring the need to consider energy balance holistically when evaluating the metabolic benefits of physical activity.

### 4.3.4 Validation Cohort: Expansion & Re-Testing of BXD43

Given that BXD43 had the most striking activity wheel usage in both our original large-scale screening study and in this long-duration exercise cohort (i.e. Fig. 4.1 A and Fig. 4.2 K), we decided to test this strain again, using all of the available running wheels (i.e. testing 32 BXD43 mice at once). As expected, BXD43 again demonstrated remarkable voluntary exercise, with an average running distance of 130 km/week at 9 weeks of age (Fig. 4.3 A). Over time, consistent with our previous observations, activity declined significantly to approximately 66 km/week by 19-29 weeks of age, reflecting a 49% reduction in voluntary running ($p = 0.002$, Fig. 4.3 B).

HCM was monitored in a matched sedentary cohort to assess spontaneous activity outside of

wheel usage. As with the previous cohorts, HCM activity remained stable from 15 to 29 weeks of age, with no detectable change over time ($p = 0.91$, Fig. 4.3 C), and the median HCM activity across the cohort showed no significant difference between early and late adult stages ($p = 0.38$, Fig. 4.3 D), suggesting that normal ambulation in the home cage is relatively stable throughout adulthood.

To validate our previous findings, we repeated key metabolic assays in this expanded BXD43 cohort. Specifically, we conducted an IPGTT at 24 weeks of age to assess whether sustained voluntary running recapitulated the improvements in glucose handling observed in our earlier cohorts. Running mice displayed significantly improved glucose tolerance, with a 38% reduction in area under the glucose curve compared to sedentary mice ($p = 0.001$, Fig. 4.3 E). Body weight gain over the 24-week period was also significantly attenuated in the running cohort, with a 22% reduction relative to sedentary controls ($p = 2e\text{-}4$, Fig. 4.3 F). Despite these improvements, total food intake was significantly higher in runners (+31%, $p = 0.001$, Fig. 4.3 G), and when normalized to body weight, this difference became even more pronounced, with a 68% increase in food consumption per gram of body mass ($p = 2e\text{-}5$, Fig. 4.3 H). These findings further support the presence of compensatory hyperphagia in response to elevated energy expenditure and emphasize the metabolic cost of sustained voluntary exercise. The consistency of these outcomes across independent cohorts underscores the robustness and reproducibility of our results.

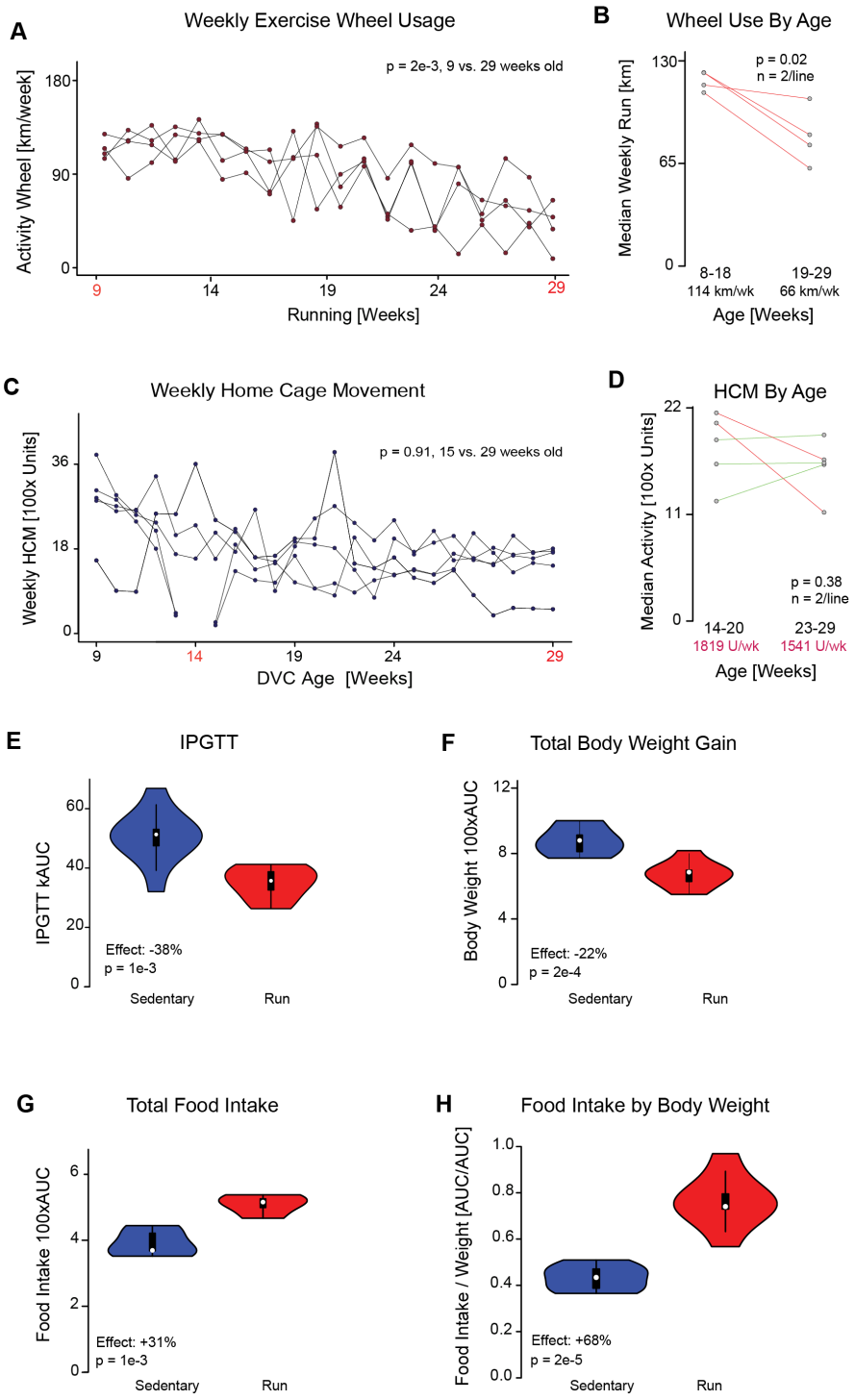


**Figure 4.3** BXD43: Cohort 4: Validation. (**A**) Weekly total activity wheel usage for cohort 4 females over time. (**B**) Slopegraph comparison of median movement between weeks 8-18 versus 19-29 for all running cohorts. (**C**) HCM total activity for all sedentary cohorts across time. (**D**) Slopegraph comparison of median movement between weeks 14-20 versus 23-29 for all sedentary cohorts. (**E**) IPGTT glucose response decrease of 38% in running cohorts vs. sedentary cohorts. (**F**) Body weight decreased by 22% in running cohorts vs. sedentary cohorts. (**G**) Running mice ate around 31% more than sedentary mice, (**H**) or around 68% more than sedentary mice if corrected for the slightly decreased body weight of running mice.

### 4.3.5 Causal Associations and Possible Molecular Mechanisms and Variables Predicting Metabolic Health Outcomes

In line with our findings that voluntary exercise mitigates metabolic decline, genus-level taxonomic analysis of the metagenome revealed several bacterial groups whose relative abundances were significantly altered by activity (Fig. 4.3 A). According to the results presented in Chapter 3, voluntary exercise led to a significant reduction in the abundance of *Lachnospiraceae_unclassified* (Fig. 4.4 B), in contrast to the substantial elevation observed in high-fat-fed animals across MG and MT profiles (Fig. 4.4 C–D). This suggests that sustained physical activity may partially reverse HF-induced microbiota remodeling at the genus level. Importantly, this exercise-induced decrease in *Lachnospiraceae_unclassified* was consistent across individual mice, supporting the robustness of this association (Fig. 4.4 A).

Two additional genera, *GGB28904* and *Bacteria_unclassified*, also demonstrated consistent patterns across datasets and experimental conditions. Both taxa showed significant increases in relative abundance under HF feeding (Chapter 3, Fig. S4.2 B–C, E–F) , yet were significantly depleted in the running cohort (Fig. S4.2 A and D). These findings mirror those of *Lachnospiraceae_unclassified* and suggest that voluntary activity not only alters host metabolism but also shapes the gut microbiota in ways that oppose HF-driven dysbiosis. Together, these data support a model in which exercise serves as a counter-regulatory force that partially restores the gut microbiome to a healthier state, thereby contributing to improved host metabolic outcomes.

To explore potential mechanistic links between gut microbiota and host metabolic function, we further analyzed correlations between the relative abundance of *Lachnospiraceae_unclassified* and the expression of key cecal metabolic genes. Significant positive associations were observed with *Cpt1a*, *Bdh1*, and *Ldha* (Fig. 4.4 E–G). These genes are involved in distinct aspects of energy metabolism: *Cpt1a* in mitochondrial fatty acid oxidation [136], *Bdh1* in ketone utilization [137], and *Ldha* in glycolytic lactate production under hypoxic or high-demand conditions [138]. These results suggest that the presence of *Lachnospiraceae_unclassified* may influence, or reflect, host energy metabolic processes. The attenuation of this taxon by exercise thus represents a plausible microbial contributor to the improved metabolic phenotype observed in the running cohorts.

To further refine our analysis beyond the genus level, we performed species-level differential abundance and prevalence modeling using MaAsLin3 ([107]), comparing running and sedentary cohorts. This analysis revealed a subset of microbial species that were significantly associated with voluntary exercise, either through increased prevalence, decreased abundance, or both (Fig. 4.5 A). Notably, species such as *s__Clostridium_butyricum* showed consistent depletion in the running cohort. However, it is worth noting that these results may also be partially influenced by the limited sample size in the validation cohort, which can amplify the apparent presence–absence contrasts between groups. Nonetheless, these findings offer valuable preliminary insights and warrant follow-up validation in larger cohorts. Together, these species-level associations extend and refine our genus-level observations, offering more precise taxonomic resolution and strengthening the link between voluntary exercise and the gut microbial ecosystem.

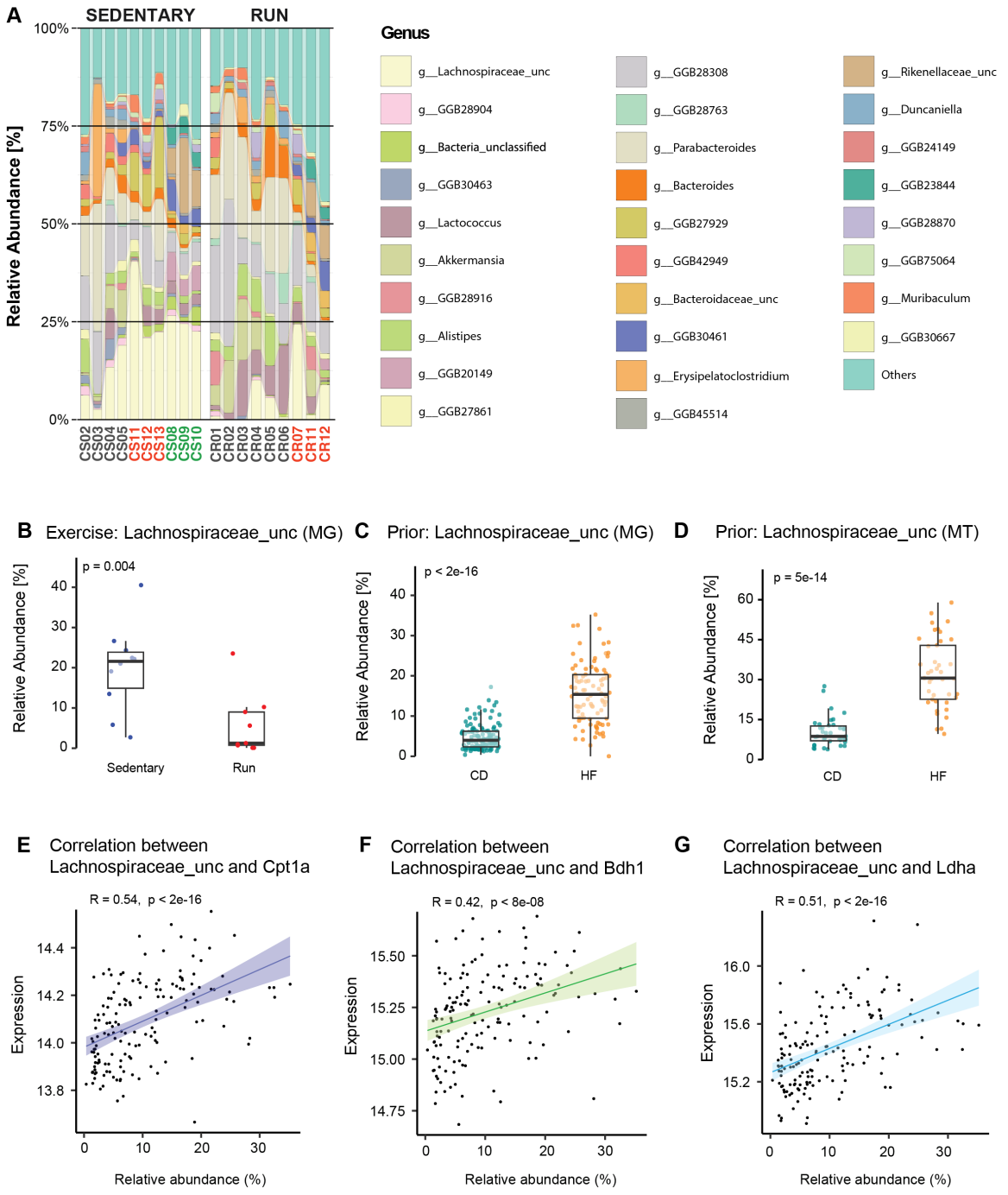


**Figure 4.4** Causal Effects and Associative Molecular Factors. (**A**) Abundance percentiles of the 28 most abundant genera of cecal microbiome measured in 10 sedentary and 10 running mice. (**B**) Relative abundance of the *Lachnospiraceae* family, showing a significant decrease in running individuals (p = 0.004). (**C**) This finding is consistent with a recent study of ours (unpublished) in the metagenome (MG) cecal samples of hundreds of mice recorded on either low fat chow (CD) or the same HF as used in this study, where HF individuals had a significant increase in *Lachnospiraceae* ($p < 2e{-}16$). (**D**) Equivalent findings in the metatranscriptome (MT) from the same cohorts. (**E-G**) Certain bacterial taxa also associated with gene expression in the cecum for particular gene sets, such as *Lachnospiraceae* with energy metabolism genes such as *Cpt1*, *Bdh1*, and *Ldha*.

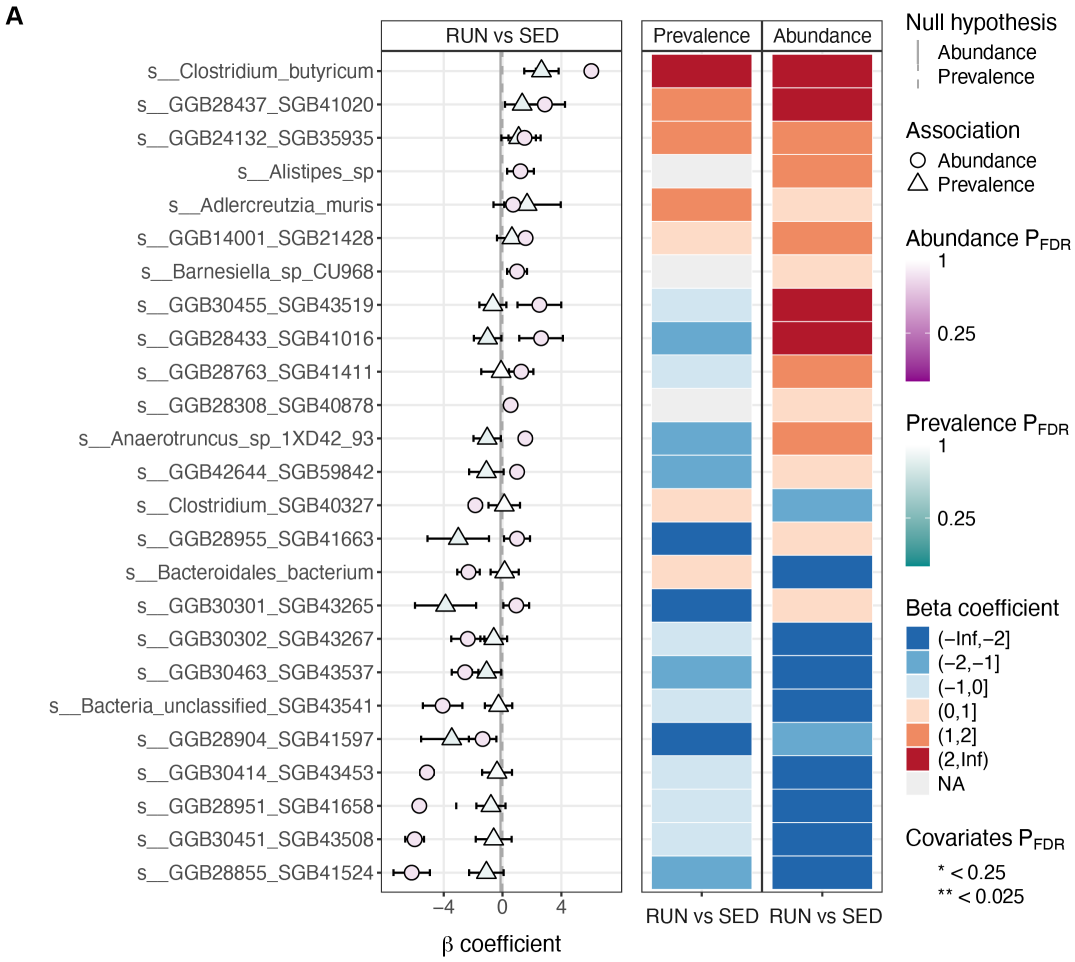


**Figure 4.5** Microbiome Species Altered by Exercise in HFD-fed Mice. (**A**) Species-level changes in the gut microbiome between exercised (RUN) and sedentary (SED) mice under high-fat diet conditions, identified using MaAsLin3. The plot shows associations in both abundance (circles) and prevalence (triangles), with effect sizes (β coefficients) and FDR-adjusted p-values. Species are sorted by the effect size in the abundance model.

## 4.4 Discussion

In this study, we used three highly active BXD strains to evaluate the long-term effects of voluntary exercise on metabolic outcomes under a high-fat (HF) dietary challenge. Despite sustained physical activity across 24 weeks, wheel access did not substantially prevent HF-induced weight gain but markedly improved glucose tolerance and metabolic flexibility, even in the presence of elevated caloric intake. Notably, running cohorts displayed significantly improved glucose responses despite similar or higher body weight compared to sedentary controls, suggesting that exercise conferred metabolic benefits independent of weight loss. This dissociation between weight and glucose control underscores the importance of assessing physiological outcomes beyond body mass, and supports the concept that physical activity enhances metabolic health even without changes in total adiposity. Our results further validate BXD strains 40, 43, and 100 as robust models for dissecting gene–exercise–diet interactions, with BXD43 showing exceptional voluntary activity across multiple cohorts.

A key finding in this study was that voluntary exercise significantly modulated gut microbiota composition at both the genus and species levels. In particular, the genus *Lachnospiraceae_unclassified*, which was consistently elevated in HF-fed animals, showed a marked reduction in abundance among exercising mice. This pattern was robust across cohorts and experimental batches, suggesting a reproducible association between voluntary activity and attenuation of HF-induced microbial shifts. The reduction of *Lachnospiraceae_unclassified* may reflect alterations in microbial metabolism, particularly related to SCFAs. Prior studies have shown that members of the *Lachnospiraceae* family contribute significantly to butyrate production in the gut [139], and that circulating levels of butyrate correlate positively with aerobic fitness in humans [140]. Butyrate has also been shown to enhance glucose-stimulated insulin secretion [141], providing a potential mechanistic link between microbial composition, metabolite profiles, and improved glucose regulation in our running cohorts. Thus, the observed microbial shifts may not be merely correlative but could reflect functional adaptations that enhance host metabolic resilience.

At the species level, we observed a significant depletion of *s__Clostridium_butyricum* in sedentary animals, which was partially restored in runners. *C.butyricum* is a known symbiotic bacterium with protective roles in intestinal health, and its abundance has been negatively associated with obesity and metabolic dysfunction. Recent evidence suggests that *C.butyricum* exerts anti-obesity effects by modulating microbial composition and intestinal metabolite profiles [142]. Its increased abundance in our exercised mice aligns with these findings and may contribute to the observed improvements in glucose handling. Taken together, these results underscore the potential of voluntary exercise to counteract HF-induced dysbiosis and promote the colonization of beneficial taxa linked to metabolic health.

Despite these insights, our study is not without limitations. First, the long duration required for each cohort (spanning nearly six months) constrained throughput and limited our ability to assess larger sample sizes or include more strains. Second, batch effects in diet formulation and microbiome sequencing may introduce variability that is difficult to control retrospectively, particularly when

dealing with subtle microbiota-related phenotypes. Third, our microbiome dataset, though informative, was relatively small and based on a limited number of samples, potentially reducing statistical power for detecting more subtle taxonomic shifts. Larger-scale studies incorporating time-course microbiome sampling, metabolomics, and metatranscriptomics would provide a more comprehensive view of microbial dynamics in response to exercise and diet.

In conclusion, our findings demonstrate that sustained voluntary exercise confers significant metabolic benefits under high-fat dietary conditions, including improved glucose tolerance and remodeling of the gut microbiota. By integrating long-term physical activity tracking with precise assessments of food intake and body composition, we establish a robust framework for dissecting gene–environment interactions that influence metabolic health. Notably, these benefits occurred independently of major weight loss, underscoring the importance of exercise-driven metabolic flexibility. Key microbial taxa, such as *Lachnospiraceae_unclassified* and *Clostridium_butyricum*, emerged as potential mediators of these effects, likely via mechanisms involving SCFA production and host energy homeostasis. Together, these results highlight a critical role for the gut microbiome in translating physical activity into metabolic resilience and point toward new microbiota-targeted strategies for preventing or mitigating obesity-related diseases.

## 4.5 Materials & Methods

### 4.5.1 Animal Handling and Cohort Selection

The 3 strains selected for this study were ordered from The Jackson Laboratory, and imported to our facility at the University of Luxembourg in fall 2020. All strains went through one to two rounds of breeding in Specific Pathogen Free (SPF) conditions to amplify the line and create sufficient embryos for embryo transfer to our Specific and Opportunistic Pathogen Free (SOPF) facility for long-term study maintenance. In March 2021, our study protocol, GEDI:CAPMED, was approved by the Luxembourgish Ministry of the Environment (protocol LUPA 2021/07) and the first of four cohorts, all starting with 32 mice, began on May 2022 and concluded in November 2024. The 32 mice of each cohort were separated evenly, i.e. 16 into those cages with *ad libitum* access to a recorded running wheel, and 16 into those cages with simple HCM, without a wheel. All mice were established in cages of 2 individuals/housing unit, which was intended to minimize conflicts over running wheel use and saturation of HCM sensors, without necessitating solitary housing, which would cause emotional distress over the 24 weeks of this study. Besides the consideration for improving animal welfare, we also specifically do not want to model depression-linked weight effects that could be affected by isolation [143, 144, 145]. Furthermore, the difference in activity and wheel usage between single-housed and dual-housed mice is approximately known from prior literature: dual-housed mice use wheels ~1.5x as much as single-housed mice [146], and dual-housed mice are ~1.8x as active in the home cage. All animals were fed Harlan Teklad 2018 "chow" diet from weaning until 8 weeks of age, at which point they started on our HF, Harlan Teklad TD.06414. Mice in exploratory cohorts 1, 2, and 3 were all female. Mice in the validation cohort, cohort 4 (exclusively BXD43), were male and female. All mice were on a 12h day/night cycle from 7am to 7pm. Mice were checked by a trained technician or veterinarian daily, and the scientific team checked the mice every Tuesday morning, weighing the mice, then weighing and replacing the food to calculate weekly food intake. Four mice died during the study. For one cage, both mice died shortly apart and data from that cage was suppressed, and in two other cages, one of the two died during the study. Those solitary mice were authorized to complete the study and their data are included in all analyses. They are flagged as potential confounding factors in the Supplemental Data Tables, but as only 2 "solitary" cages out of 64, their inclusion or suppression does not affect any overall conclusions.

### 4.5.2 Dietary Information

Our study was partially confounded by changes in the production line to Harlan Teklad 06414. While the diet used throughout the study was from the same provider and reference number, we could not order and stock 3 years of food all at once. Thus, three batches of diet were ordered.

### 4.5.3 Home Cage Monitoring (HCM)

HCM was performed with two DVC racks (Tecniplast). Cages were cleaned every three weeks throughout the study.

### 4.5.4 Exercise Wheel Monitoring

Continuously–recorded exercise wheels were set up in a separate sound–isolated and ventilated cabinet in the same room as the HCM rack, with 8 recorded wheels (Columbus Instruments, 4 inch diameter solid plastic wheel). The Columbus Instruments "CI Multi-Device Interface" recording software was used to generate the data tables, with the total number of wheel revolutions summed up continuously for each 10 minute interval throughout the duration of the study. The software occasionally crashed, resulting in a loss of 35 days in Cohort 1, 7 days in Cohort 3, and 2 days in Cohort 4. However, during this time the mice were still able to use the wheels, only the recording information was lost. Running wheel cages were changed every two weeks, as the cages had somewhat less bedding than the HCM cohort. The reduction in bedding was due to issues noticed during the study pilot, where mice could pile up bedding next to the wheel, which was avoided during this study.

### 4.5.5 Intraperitoneal Glucose Tolerance Test

Prior to the IPGTT, mice were moved to clean cages and fasted overnight (~6pm to ~9am). Body weights were recorded at both times, and fasted blood glucose was measured with a handheld glucometer (notably, with a sensor cap at 600 mg/dl). After all fasting glucose values were checked, mice were all intraperitoneally injected with 5 $\mu$L/g body weight of a 40% glucose solution in 0.9% sodium chloride pre-warmed on a hot plate to approximately 37°C. Blood glucose values were taken at 15, 30, 45, 60, 90, and 120 minutes using a pinprick on the distal end of the tail. At the end of the 120 minutes, the tails were treated with Bepanthen cream (an antiseptic / anti-inflammatory) and mice were returned to their cages with HF and water.

### 4.5.6 Sacrifice and Tissue Collection

At the end of the study, at approximately 33 weeks of age, mice were again moved to clean cages and fasted overnight (~6pm to ~9am). Mice were intraperitoneally injected with a 10:1 ketamine:xylazine mixture, injected at 160 mg/kg ketamine to 16 mg/kg xylazine. After mice were fully anesthetized, they were perfused with ice-cold saline for terminal euthanasia, and tissues were collected for future studies over the next ~10 minutes per mouse. Each cohort was sacrificed this way, i.e. over a ~3 hour period for all mice in the cohort.

### 4.5.7 Microbiome Analysis

Same as in Chapter 3.

### 4.5.8 Bioinformatics and Statistics

All analyses were performed using R. Tests of normality were performed with Shapiro-Wilk test (shapiro.test in Base R), with values of $W > 0.85$ considered as normally distributed. Significance tests between two groups were performed using Student's t-test (t.test function in base R), with paired tests used for comparing movement within cages over time. Significance tests between three or more groups were performed using ANOVA (aov function in base R). When ANOVA results were significant, Tukey post-hoc test was used to test for differences between groups (using TukeyHSD in base R).

## Supplementary Information

Raw data for the DVC and wheel are available in the article and/or Supplementary Materials. Materials are available upon request.

## Acknowledgements

Thank you to Giorgio Rosati and Tecniplast for initial piloting of the DVC equipment and data access. AL was supported by the Luxembourg National Research Fund (PRIDE21/16749720/NEXTIMMUNE2). Thanks also to Manuel Buttini, Pierre Garcia, and Tony Heurtaux for assistance in the tissue harvest at the end of each cohort. Thanks also to the lead animal facility caretakers, Julie Virolle and Cloé Ordrenneau, and veterinarian Jennifer Behm.

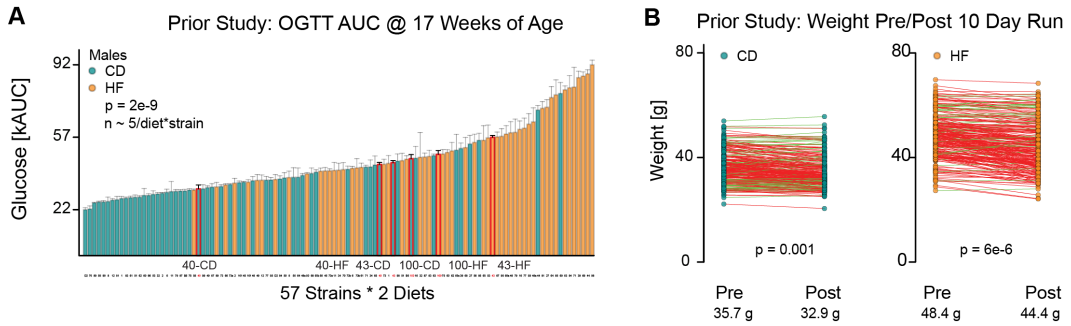


**Figure S4.1** Previous Results and Knowledge About BXD Metabolic Phenotypes. (**A**) Oral glucose response in a prior study on 57 BXD strains. (**B**) Difference in body weight before and after 10 days of voluntary exercise in a prior study on 57 BXD strains.

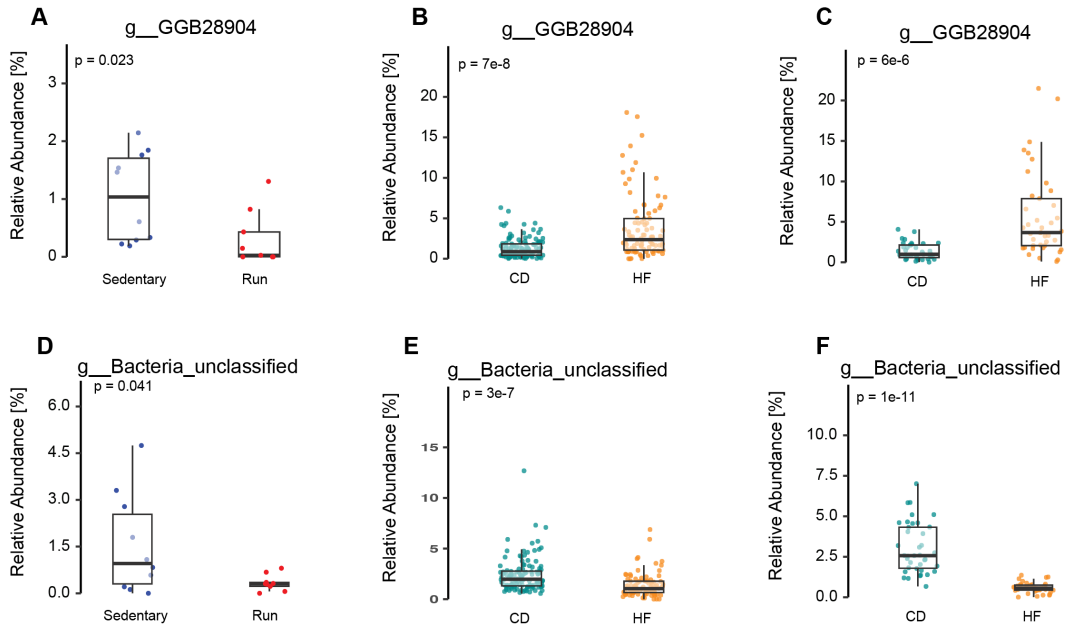


**Figure S4.2** Exercise-Associated Shifts in Gut Microbiota at the Genus Level. (**A, D**) Two bacterial genera—*GGB28904* and *Bacteria_unclassified*—were significantly depleted in the running cohort compared to the matched sedentary mice, reinforcing the association between exercise and microbial composition shifts observed in the main dataset. (**B, E**) In a large MG dataset (Chapter 3) comprising CD and HF mice, both genera were significantly more abundant under HF. (**C, F**) Similar patterns were confirmed in the MT data from the same cohort, indicating not only presence but transcriptional activity of these genera under HF.

# Chapter 5

# Manuscript 3

## Apolipoprotein A-IV and its derived peptide, T55-121, improve glycemic control and increase energy expenditure

## Preface

**I** conducted the prediction of Apolipoprotein A-IV functional peptide (T55-121), participated in all mouse experiments related to the Apolipoprotein A-IV and it functional peptides, and was responsible for drafting the manuscript sections related to the functional peptides. **I** participated in this project at the Chinese Academy of Sciences from 2018 to 2020 and contributed to manuscript writing in 2023.

*Life Metabolism, Volume 3, Issue 4, August 2024, loae010, doi: 10.1093/lifemeta/loae010.*

# Apolipoprotein A-IV and its derived peptide, T55-121, improve glycemic control and increase energy expenditure

**Zhen Cao[a,#], Lei Lei[a,c,#], Ziyun Zhou[d,#], Shimeng Xu[a], Linlin Wang[e], Weikang Gong[f], Qi Zhang[a,b], Bin Pan[a,b], Gaoxin Zhang[g], Quan Yuan[g], Liujuan Cui[a], Min Zheng[h], Tao Xu[a,b,e], You Wang[a,*], Shuyan Zhang[i,j,k,l,*], Pingsheng Liu[a,b,*]**

**a** National Laboratory of Biomacromolecules, CAS Center for Excellence in Biomacromolecules, Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, China
**b** University of Chinese Academy of Sciences, Beijing 100049, China
**c** The State Key Laboratory of Bioactive Substance and Function of Natural Medicines, Institute of Materia Medica, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100050, China
**d** Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Esch-sur-Alzette 4362, Luxembourg
**e** Guangzhou Regenerative Medicine and Health Guangdong Laboratory (Bioland Laboratory), Guangzhou 510005, China
**f** Department of Computer Science, School of Computing, National University of Singapore, 117417, Singapore
**g** School of Basic Medical Sciences, Southwest Medical University, Luzhou 646000, China
**h** The State Key Laboratory for Diagnosis and Treatment of Infectious Diseases, National Clinical Research Center for Infectious Diseases, Collaborative Innovation Center for Diagnosis and Treatment of Infectious Diseases, The First Affiliated Hospital, College of Medicine, Zhejiang University, Hangzhou 310003, China
**i** Beijing Key Laboratory of Emerging Infectious Diseases, Institute of Infectious Diseases, Beijing Ditan Hospital, Capital Medical University, Beijing 100015, China
**j** Beijing Institute of Infectious Diseases , Beijing, 100015, China
**k** National Center for Infectious Diseases, Beijing Ditan Hospital, Capital Medical University, Beijing 100015, China
**l** National Key Laboratory of Intelligent Tracking and Forecasting for Infectious Diseases, Beijing,100015, China

# These authors contributed equally

* Corresponding authors:

Pingsheng Liu, National Laboratory of Biomacromolecules, CAS Center for Excellence in Biomacromolecules, Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, China; University of Chinese Academy of Sciences, Beijing 100049, China. Phone number: 010-64888517. E-mail address: pliu@ibp.ac.cn
Shuyan Zhang, Institute of Infectious Diseases, Beijing Key Laboratory of Emerging Infectious Diseases, Beijing Ditan Hospital, Capital Medical University, Beijing 100015, China. Phone number: 010-84322620. E-mail address: syzhang@ibp.ac.cn
You Wang, National Laboratory of Biomacromolecules, CAS Center for Excellence in Biomacromolecules, Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, China. Phone number: 010-64888524. E-mail address: wangyou51@126.com

**Keywords:**
Bariatric/Metabolic Surgeries; Proteomics; Apolipoprotein A-IV; Glucose Tolerance; Glucose-Stimulated Insulin Secretion; Human Islets

*Life Metabolism, Volume 3, Issue 4, August 2024, loae010, doi: 10.1093/lifemeta/loae010.*

## 5.1 Abstract

It is crucial to understand the glucose control within our bodies. Bariatric/metabolic surgeries, including laparoscopic sleeve gastrectomy (LSG) and Roux-en-Y gastric bypass (RYGB), provide an avenue for exploring the potential key factors involving in maintaining glucose homeostasis since these surgeries have shown promising results in improving glycemic control among patients with severe type 2 diabetes (T2D). For the first time, a markedly altered population of serum proteins in patients after LSG was discovered and analyzed through proteomics. ApoA-IV was revealed to be increased dramatically in diabetic obese patients following LSG, and a similar effect was observed in patients after RYGB surgery. Moreover, recombinant protein apoA-IV treatment was proved to enhance insulin secretion in isolated human islets. These results show apoA-IV may play a crucial role in glycemic control in humans, potentially through enhancing insulin secretion in human islets. ApoA-IV was further shown to enhance energy expenditure and improve glucose tolerance in diabetic rodents, through stimulating glucose-dependent insulin secretion in pancreatic $\beta$ cells, partially via G$\alpha$s-coupled GPCR/cAMP signaling. Furthermore, T55-121, truncated peptide 55-121 of apoA-IV, was discovered to mediate the function of apoA-IV. These collective findings contribute to our understanding of the relationship between apoA-IV and glycemic control, highlighting its potential as a biomarker or therapeutic target in managing and improving glucose regulation.

## 5.2 Introduction

Over the past decades, there has been a striking surge in the global occurrence of obesity and type 2 diabetes (T2D), leading to significant social and economic consequences [147]. Bariatric surgery has been utilized as a means of weight reduction for individuals diagnosed with morbid obesity as well as an effective intervention that can lead to significant improvements in diabetes management [148]. Currently, laparoscopic sleeve gastrectomy (LSG) and laparoscopic Roux-en-Y gastric bypass (RYGB) are the two most commonly performed procedures in the field of bariatric surgery. Both LSG and RYGB have been found to yield enduring weight loss, long-term improvement in glycemic control, and other notable metabolic improvements in patients with morbid obesity and severe T2D [149, 150]. Exploring the significant altered factors associated with substantial improvement of type 2 diabetes after surgery would aid in understanding the mechanisms underlying glucose control.

Apolipoprotein A-IV (apoA-IV) is almost exclusively synthesized by the enterocytes of the small intestine, with dietary lipids serving as significant stimulants for its production [151, 152]. Of note, the production of apoA-IV accounts for 3% of total protein synthesis following oleic acid infusion [153, 154]. The protein exists in blood circulation as both lipoproteins (chylomicron and high-density lipoprotein) and in a lipoprotein-free form after metabolism of chylomicrons [155]. Adiposome-derived artificial lipoproteins bearing apoA-IV are established and exhibit functions in the improvement of glucose tolerance [156]. Importantly, apoA-IV is shown to rise after RYGB in humans, and simultaneously, there was a noticeable improvement in diabetes [157, 158]. Additionally, population-based genotype studies show that apoA-IV polymorphism, especially apoA-IV360, has a significant effect on serum glucose levels [159, 160]. These findings indicate that apoA-IV plays a role in the regulation of glucose levels.

Further, Fei Wang and colleagues reported that apoA-IV augments glucose-stimulated insulin secretion in wild type mice [161]. In addition, Xiaominng Li et al. found that apoA-IV reduces hepatic gluconeogenesis through interaction with nuclear receptor NR4A1 and NR1D1 to repress phosphoenolpyruvate carboxykinase (PEPCK) and Glc-6-Pase expression at the transcriptional level [162, 163]. Xiaominng Li and colleagues also found that apoA-IV contributes to improved insulin sensitivity in adipocytes via PI3K-Akt signaling [164]. However, it is unknown whether apoA-IV enhances insulin secretion in humans and the mechanism by which apoA-IV enhances insulin secretion is not fully understood. Furthermore, it is crucial to identify the specific functional peptide of apoA-IV that plays a role in modulating glucose homeostasis, which would potentially offer therapeutic benefits in managing obesity and T2D.

Here, for the first time, the patient serum proteins were compared before and after LSG through gel-based proteomics to reveal the potential proteins involved in glycemic control in humans. In the most changed bands, apoA-IV was shown to be the most abundant protein and revealed to increase extremely after LSG. Besides, we also found apoA-IV was increased significantly in patients following RYGB. All the data imply apoA-IV play a role in glucose control in humans. Further we found that administrating eukaryote-derived recombinant apoA-IV resulted in an enhanced energy

expenditure and improved glucose tolerance in both WT and diabetic rodents. Furthermore, our findings showed that apoA-IV stimulated glucose-stimulated insulin secretion (GSIS) in human islets, partially via G$\alpha$s-coupled GPCR/cAMP signaling. Moreover, we identified that T55-121 (truncated peptide 55-121 of murine apoA-IV) improved glucose tolerance and enhanced metabolic rate. These findings would expand our knowledge of the relationship between apoA-IV and glycemic control, underscoring its potential as a biomarker or therapeutic target for enhancing glucose regulation.

## 5.3 Results

### 5.3.1 ApoA-IV rises in sera of patients after bariatric surgeries

Bariatric surgeries, such as LSG and RYGB, have been identified as effective procedures for long-term enhancement in glycemic control and other significant metabolic improvements in patients suffering from morbid obesity and T2D. To gain a better understanding of the underlying mechanisms involved in glucose control, it is essential to identify the significant factors that undergo noticeable changes following bariatric surgeries. Thus in this study, we performed analysis of patient sera before and after undergoing either LSG or RYGB procedures.

Similar to previous reports [149], patients showed significant improvement in glycemic control one year after LSG, as evidenced by a substantial reduction in fasting blood glucose, insulin, C-peptide, and HbA1c levels (Fig. 5.1 a-d). Compared with sera before LSG, the post-surgery sera exhibited a noticeable reduction in turbidity (Fig. 5.1 e). To investigate the significant factors that undergo notable changes in patient sera after LSG, for the first time, we performed a comparative analysis of serum proteins before and after LSG through gel-based proteomics. Notably, similar protein bands, 1 and 2, around 45 kDa in SDS-PAGE, were remarkably increased in both patient sera after LSG (indicated as red arrowhead in Fig. 5.1 f), the proteins in which could potentially play crucial roles in the improvement of glycemic control. Then both bands were subjected to proteomic analysis using mass spectrometry (MS) to identify the major proteins. There were 26 proteins identified in band 1 (Table S5.1) and 23 proteins in band 2 (Table S5.5) with 17 proteins in common (Fig. 5.1 g). Among the 17 common proteins, apoA-IV achieved the highest ranking with the highest score, coverage and the largest number of unique peptide segments. More importantly, apoA-IV level in serum was revealed to be elevated substantially after LSG in both patients, through immunoblotting, using AGP-1, one of the 17 common proteins, as a loading control (Fig. 5.1 h).

Besides the effects of LSG, we conducted an analysis of the factors that experienced changes in the plasma of patients before and one year after RYGB as well. The analysis was performed using the reported proteomics data [165]. One year after RYGB, the fasting blood glucose levels of the 10 diabetic obese patients exhibited a notable decrease, transitioning from levels higher than 9 mM to levels lower than 7 mM [165]. Moreover, consistent to the effect of LSG, plasma apoA-IV level demonstrated a remarkable increase one year after RYGB surgery (Fig. 5.1 i; Table S5.3). Collectively, these data consistently demonstrated a significant increase in apoA-IV levels in patients, accompanied by long-term enhancement in glycemic control. Therefore, apoA-IV may be important in glycemic control in humans.

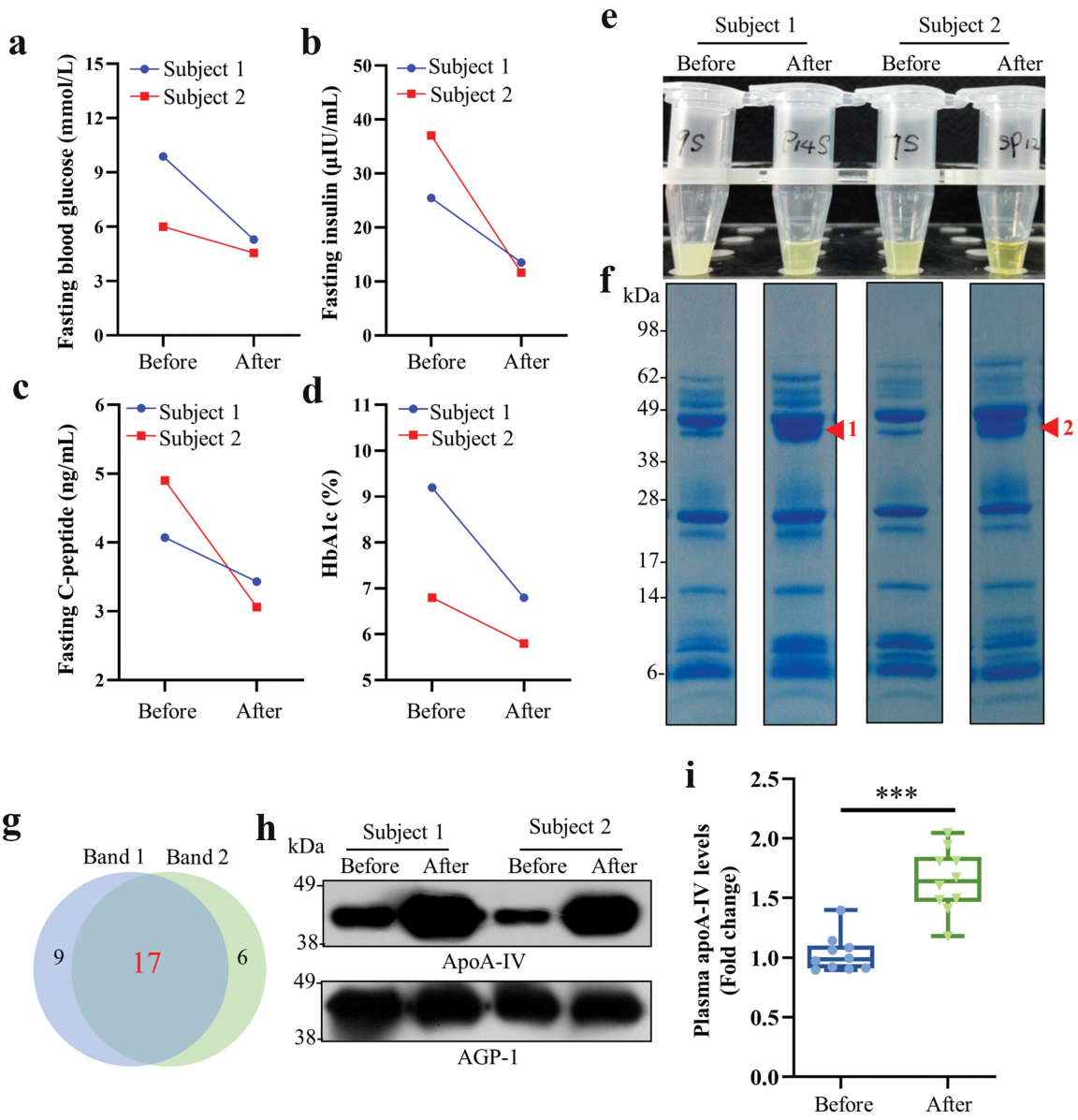


**Figure 5.1** The levels of apoA-IV increase in the sera of patients after bariatric surgeries. (**a-d**) Changes in serum biochemical parameters of patients with morbid obesity and T2D before and one year after LSG, including fasting blood glucose (**a**), fasting insulin (**b**), fasting C-peptide (**c**), and HbA1c (**d**). (**e and f**) Protein profiles in sera collected from patients before and after LSG (**e**) analyzed by SDS-PAGE (**f**). Protein bands 1 and 2 from post-surgery samples, indicated by red arrowheads, were subjected to MS-based proteomics. (**g and h**) Identification of proteins in bands 1 and 2 based on proteomics and immunoblot analysis. Proteomic analysis of bands 1 and 2 (**g**). Immunoblot analysis of apoA-IV and AGP-1 in sera of subjects before and after LSG (**h**). AGP-1 is a loading control. (**i**) The changes of plasma apoA-IV levels of diabetic obese patients (n = 10) before and one year after RYGB were analyzed based on the reported proteomics data [19]. Data are presented as mean ± SEM. Statistical significance was determined by the two-tailed Student's t-test. ***P < 0.001.

### 5.3.2 ApoA-IV improves glucose tolerance in rodents

In order to investigate the potential involvement of apoA-IV in improving glucose homeostasis, we administered exogenous apoA-IV protein directly to both wild-type and T2D rodents via intraperitoneal injection. Recombinant protein apoA-IV was expressed in HEK293 cells and purified for use (Fig. 5.2 a). The administration of apoA-IV significantly improved the glucose tolerance in wild type mice (Fig. 5.2 b). Then we tested whether apoA-IV has a beneficial effect on glucose tolerance in type 2 diabetic rodents. The results showed that the eukaryote-derived apoA-IV treatment led to improved glucose tolerance in both db/db mice and ob/ob mice in a dose-dependent manner (Fig. 5.2 c and d). Similarly, glucose tolerance was enhanced in apoA-IV-treated Goto-Kakizaki (GK) rats (Fig. 5.2 e).

In addition to studying the effect of apoA-IV through administration, we also investigated its physiological effects in both WT and diabetic mice by overexpressing apoA-IV in vivo using an adeno-associated virus (AAV). After 1 month of infection of apoA-IV-AAV, serum apoA-IV level was 70% higher than that observed in control-AAV treated mice (Fig. 5.2 f). Also the expression of exogenous apoA-IV was detected in tissues. The results indicated the presence of apoA-IV-Flag in the small intestine, pancreas, and liver (Fig. S5.1). Although no differences were evident in body weight between the two groups (Fig. S5.2), ob/ob mice overexpressing apoA-IV exhibited improved glucose tolerance compared to the control mice (Fig. 5.2 g). Similarly, apoA-IV overexpression also enhanced glucose tolerance of wild-type mice (Fig. 5.2 h). Notably, ob/ob mice overexpressing apoA-IV exhibited elevated levels of HDL and lowered levels of LDL ((Fig. 5.2 i), suggesting that apoA-IV overexpression is beneficial for maintaining cholesterol homeostasis. Indeed, previous research have showed that apoA-IV may protect against atherosclerosis [166] and cAMP-stimulated cholesterol efflux from macrophages is induced in transgenic mice overexpressing human apoA-IV [167].

Taken together, these results demonstrate that both administration of recombinant protein apoA-IV and overexpression of apoA-IV in vivo using AAV improve glucose tolerance in both WT and diabetic rodents.

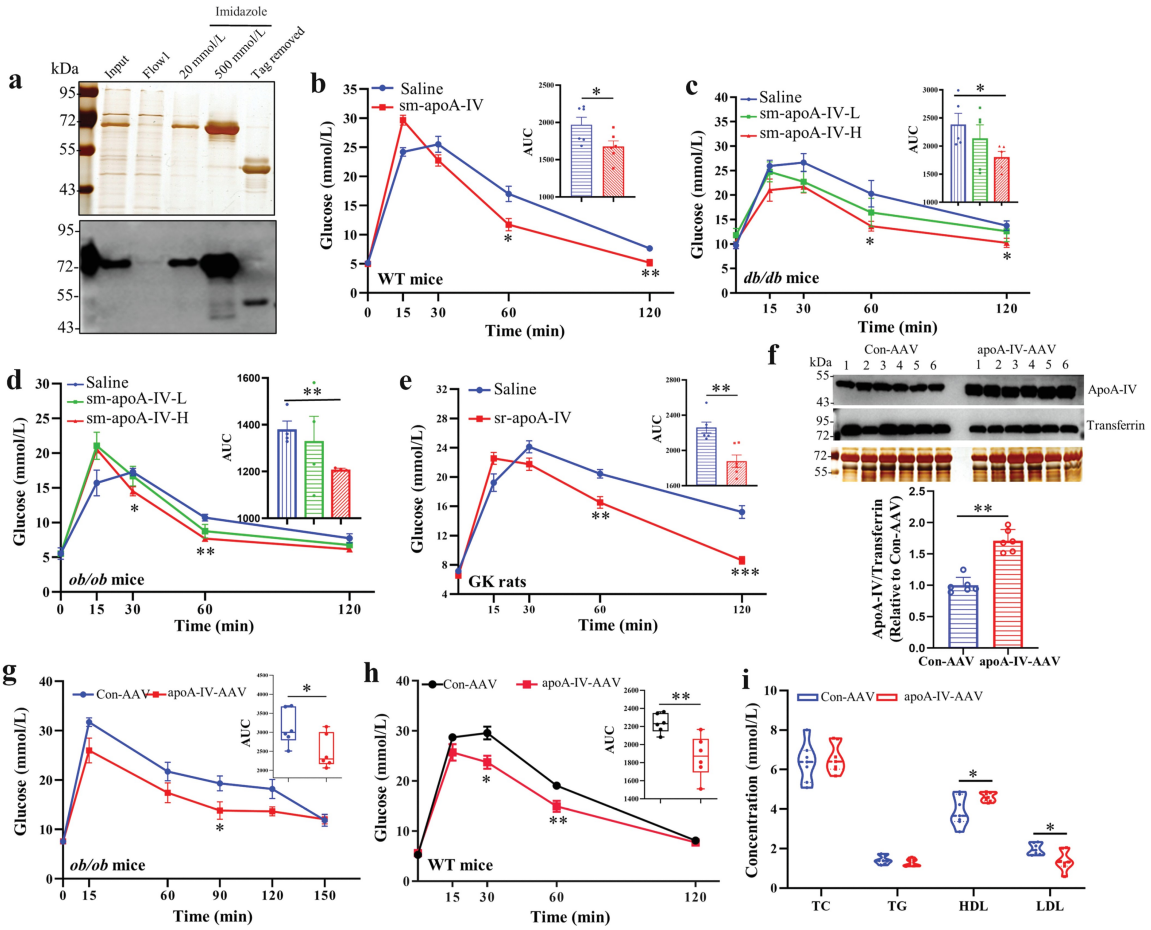


**Figure 5.2** Administration of apoA-IV and overexpression of apoA-IV improve glucose tolerance in rodents. (**a**) Silver staining (top) and immunoblot analysis (bottom) of recombinant apoA-IV protein purified from a eukaryotic expression system (HEK293F cells). (**b**) The IPGTT in WT mice (n = 6) with saline or sm-apoA-IV (6 mg/kg body weight) treatment. (**c**) The IPGTT in db/db mice (n = 5) with saline, sm-apoA-IV-L (3 mg/kg body weight), or sm-apoA-IV-H (6 mg/kg body weight) treatment. (**d**) The IPGTT in ob/ob mice (n = 4) with saline, sm-apoA-IV-L (3 mg/kg body weight), or sm-apoA-IV-H (6 mg/kg body weight) treatment. (**e**) The IPGTT in GK rats (n = 6) with saline or sr-apoA-IV treatment (6 mg/kg body weight) treatment. (**f**) Immunoblot analysis of apoA-IV (top) in sera from ob/ob mice infected with Con-AAV or apoA-IV-AAV, with transferrin as a loading control, and quantification of serum apoA-IV levels in both groups (bottom). (**g**) The IPGTT in ob/ob mice infected with Con-AAV or apoA-IV-AAV (n = 6). (**h**) The IPGTT in WT mice infected with Con-AAV or apoA-IV-AAV (n = 6). (**i**) Serum biochemical changes of AAV-infected ob/ob mice. Data are presented as mean ± SEM. Statistical significance was determined by the two-tailed Student's t-test. *P < 0.05, **P < 0.01. AUC is the area under the curve. sm-apoA-IV, signal peptide-removed mouse apoA-IV. sr-apoA-IV, signal peptide-removed rat apoA-IV. Con-AAV, adeno-associated virus encoding GFP. apoA-IV-AAV, adeno-associated virus encoding apoA-IV.

### 5.3.3 ApoA-IV enhances insulin secretion from human and rodent islets, partially via G$\alpha$s-coupled GPCR/cAMP signaling

As described above, apoA-IV may participate in glycemic control in humans and apoA-IV is proved to improve glucose tolerance in rodents. Since normal insulin secretion from pancreatic islets plays a crucial role in maintaining glucose homeostasis in the body, we wondered whether apoA-IV affects insulin secretion of pancreatic islets. Notably, apoA-IV significantly promoted glucose-dependent insulin secretion (GSIS) in isolated non-diabetic human primary islets (Fig. 5.3 a), indicating apoA-IV functions in glycemic control in humans through affecting insulin secretion of pancreatic islets. Similarly, following apoA-IV treatment, GSIS was observed to be enhanced in both isolated primary islets from spontaneous type 2 diabetic KKAy mice and mouse pancreatic $\beta$-cell line MIN6 (Fig. 5.3 b and c). Additionally, administered recombinant apoA-IV-Flag (Flag tagged to C-terminal of apoA-IV) was detected in mouse pancreas one hour after intraperitoneal administration but not detectable in mouse brain (Arrowhead shown in Fig. 5.3 d).

Fluorescent sections of pancreas and brain tissue were observed to ascertain the distribution of apoA-IV in these tissues. In alignment with the immunoblot results, apoA-IV-GFP was predominantly localized in the pancreas and was scarcely detectable in the brain (Fig. 5.3 e). These results demonstrate that one of the mechanisms by which apoA-IV exerts its glycemic control effects is by promoting insulin secretion in pancreatic $\beta$-cells. Indeed, in the case of type 1 diabetic mice, where islet $\beta$ cells are destroyed using streptozotocin (STZ) (Fig. 5.3 f), the administration of apoA-IV did not lead to improved glucose tolerance (Fig. 5.3 g).

Further we explored the mechanism by which apoA-IV enhances the secretion of insulin in pancreatic islets. It is known that cyclic AMP (cAMP) and Ca2+ play an important role in regulating glucose-stimulated insulin secretion from islet $\beta$ cells. Additionally, G$\alpha$q, G$\alpha$i and G$\alpha$s-coupled G protein coupled receptors (GPCRs) are known to mediate signaling pathways in islet $\beta$ cells [168]. Therefore, the [cAMP]i, [DAG]i, and [Ca2+]i in MIN6 cells were examined after apoA-IV treatment. As shown in Fig. 5.3 h, apoA-IV significantly increased cytoplasmic cAMP content in MIN6 cells. However, it had no effect on [DAG]i and [Ca2+]i in MIN6 cells (Fig. 5.3 i and j). Both G protein G$\alpha$s and adenylate cyclase (AC) are intimately involved in the formation of cAMP. We chose adenylate cyclase inhibitor SQ22536 and G$\alpha$s-selective antagonist NF449 to test the effect of apoA-IV on glucose-stimulated insulin secretion. Both SQ22536 and NF449 could partially block the stimulating effect evoked by apoA-IV in MIN6 cells (Fig. 5.3 k). Additionally, apoA-IV treatment slightly promoted the phosphorylation of cAMP-response element binding protein (CREB) (Fig. 5.3 l).

These results show that apoA-IV acts on islet $\beta$ cells to promote insulin secretion partially via the G$\alpha$s-coupled GPCR/cAMP pathway.

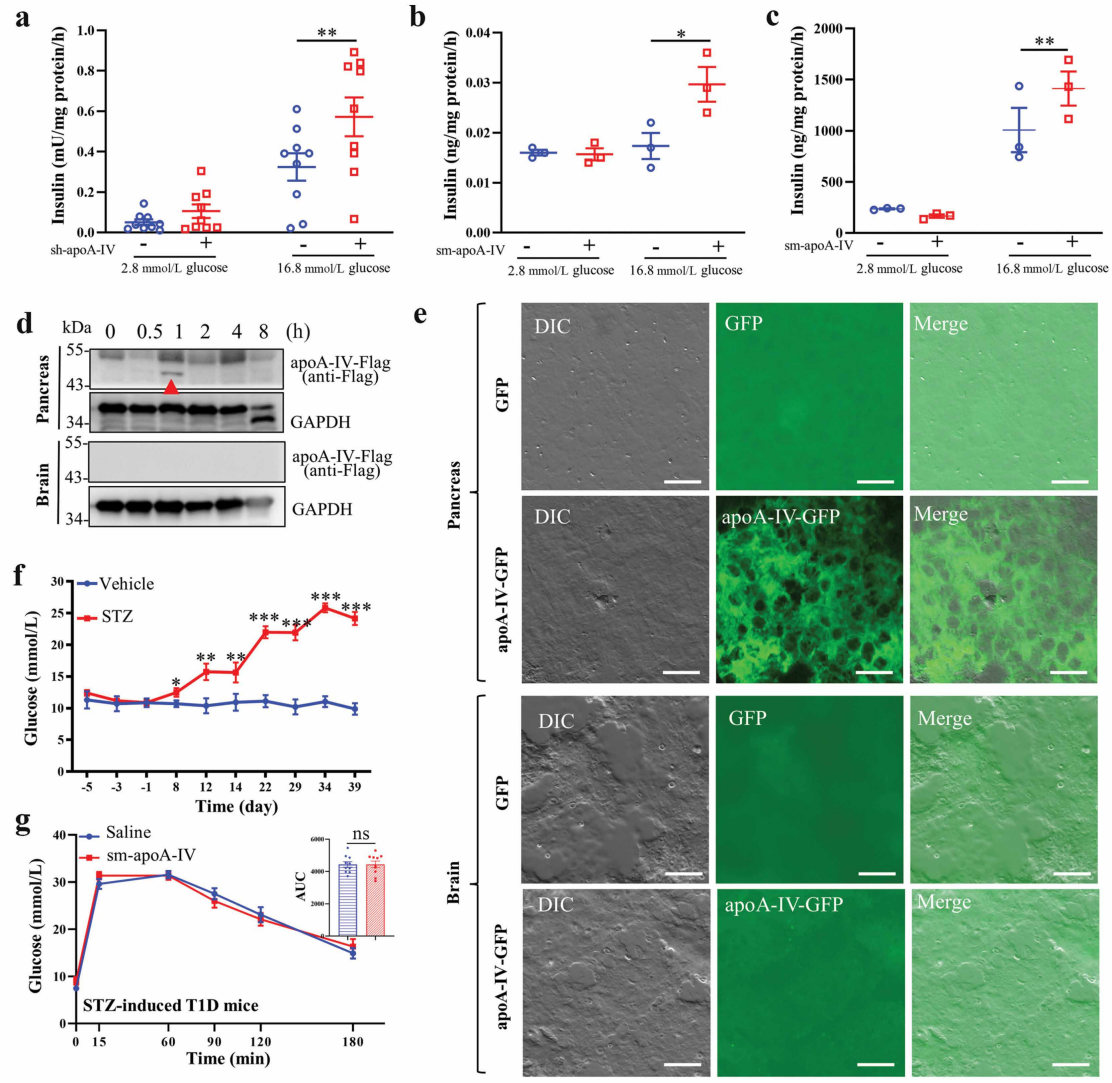


**Figure 5.3** ApoA-IV promotes GSIS in human primary islets, partially through G$\alpha$s-coupled GPCR/cAMP signaling. (**a**) GSIS in human primary islets treated with vehicle or 100 µg/mL sh-apoA-IV for 3 h. (**b**) GSIS in spontaneous T2D KKAy mouse primary islets treated with vehicle or 100 µg/mL sm-apoA-IV for 3 h. (**c**) GSIS assay in MIN6 cells treated with vehicle or 100 µg/mL sm-apoA-IV for 3 h. (**d**) Immunoblot analysis of apoA-IV-Flag in pancreas (top) and brain (bottom) at indicated times after intraperitoneal administration of recombinant apoA-IV-Flag protein. GAPDH is a loading control. (**e**) Fluorescent sections of pancreas and brain after 2 h incubation with 10 µg/mL apoA-IV-GFP or GFP. Scale bar = 20 µm. (**f**) Blood glucose levels in mice with vehicle or STZ treatment. (**g**) The IPGTT in STZ-induced T1D mice (n = 10) with the administration of saline or sm-apoA-IV (6 mg/kg body weight).

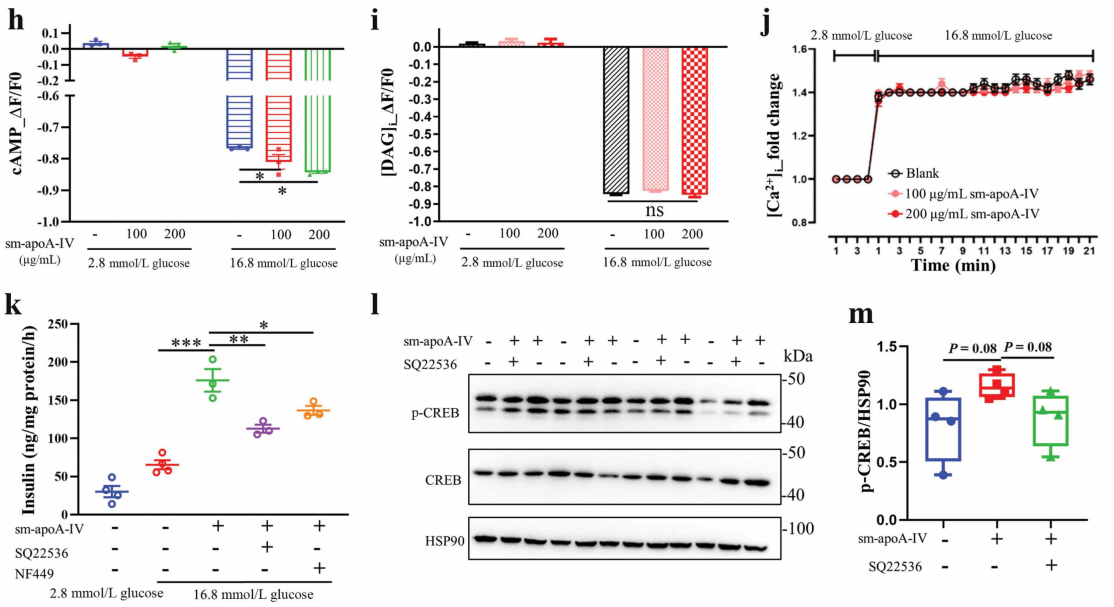


**Figure 5.3** (**h**) The cAMP sensor fluorescence changes in MIN6 cells with 0, 100, or 200 μg/mL sm-apoA-IV. (**i**) Fluorescent DAG sensor changes in MIN6 cells treated with 0, 100, or 200 μg/mL sm-apoA-IV. (**j**) Fluorescent Ca2+ sensor changes in MIN6 cells treated with 0, 100, or 200 μg/mL sm-apoA-IV. (**k**) GSIS in MIN6 cells treated with 100 μg/mL sm-apoA-IV alone, 100 μg/mL sm-apoA-IV together with 20 μmol/L SQ22536 (AC inhibitor), or 100 μg/mL sm-apoA-IV together with 10 μmol/L NF449 (G$\alpha$s-selective antagonist). (**l**) Immunoblot analysis of phosphorylated CREB (p-CREB) and CREB in MIN6 cells treated with 100 μg/mL sm-apoA-IV alone, or 100 μg/mL sm-apoA-IV together with 20 μmol/L of SQ22536. HSP90 is a loading control. (**m**) Ratio of p-CREB/HSP90 calculated from l. Data are presented as mean ± SEM. *P < 0.05, **P < 0.01, ***P < 0.001. ns, no significance. GSIS, glucose-stimulated insulin secretion. STZ, streptozocin. sh-apoA-IV, signal peptide-removed human apoA-IV. sm-apoA-IV, signal peptide-removed mouse apoA-IV.

### 5.3.4 ApoA-IV enhances energy expenditure

Since glucose metabolism plays a vital role in energy homeostasis in living organism, we then investigated whether apoA-IV would affect energy homeostasis in the body. ApoA-IV-treated mice were subjected to indirect calorimetry. There was no difference in body weight between the two groups (Fig. 5.4 a). Inconsistent with a previous report that apoA-IV inhibits appetite [169, 170], there was no discernible decrease in cumulative food intake after administration of eukaryote-derived apoA-IV (Fig. 5.4 b). Although there were no comparable differences in the respiratory exchange ratio (RER) and locomotor activity between the two groups (Fig. 5.4 c and d), apoA-IV-treated mice exhibited a dramatic increase in oxygen consumption ($VO_2$), carbon dioxide production ($VCO_2$), and heat expenditure (HE) (Fig. 5.4 e-g). Consistent with this observation, the metabolic rate of apoA-IV-treated db/db mice was also elevated (Fig. S5.3). Actually, it has been reported that diet-induced brown adipose tissue (BAT) thermogenesis and energy expenditure are reduced in apoA-IV knockout mice [171].

Collectively, these results show that apoA-IV enhances energy expenditure independent of food

intake, which may be associated with improved glucose tolerance.

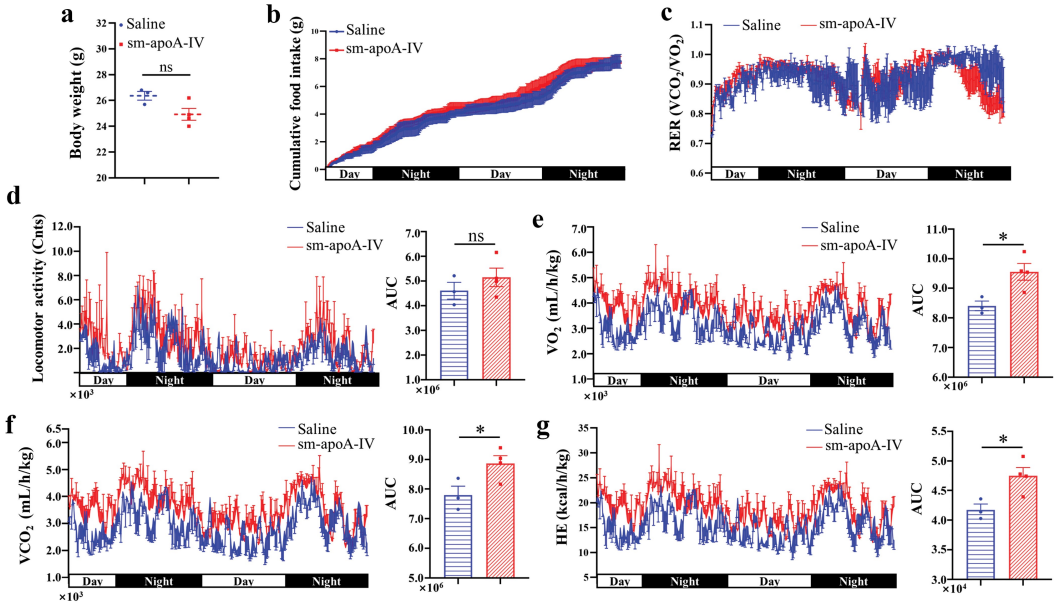


**Figure 5.4** ApoA-IV enhances energy expenditure. Analysis of indirect calorimetry of WT male mice following the administration of saline (n = 3) or sm-apoA-IV (n = 4). (**a**) Body weight. (**b**) Cumulative food intake. (**c**) Respiratory exchange ratio (RER). (**d**) Locomotor activity. (**e**) $O_2$ consumption ($VO_2$). (**f**) $CO_2$ production ($VCO_2$). (**g**) Heat expenditure (HE). The right panel is the area under the curve (AUC) in the left panel (**d–g**). Data are presented as mean ± SEM. Statistical significance was determined by the two-tailed Student's t-test. *P < 0.05, two-tailed t-test. ns, no significance. sm-apoA-IV, signal peptide-removed mouse apoA-IV.

### 5.3.5 T55-121 of murine apoA-IV improves glucose tolerance

Identifying the specific functional peptide of apoA-IV that is involved in modulating glucose homeostasis can have significant therapeutic implications for managing obesity and T2D. We employed Gaussian network model (GNM) to predict the potential functional peptide of apoA-IV. I-TASSER server was used to predict the mouse apoA-IV model and Model 1 in Fig. 5.5 a exhibited superior characteristics compared to the other models, as indicated by its C-score and Ramachandran plot (Fig. 5.5 b). Notably, Model 1 demonstrated satisfactory stereochemical quality, with only 3% of residues falling within the disallowed region of the Ramachandran plot (Fig. 5.5 b). Besides, numerous studies have documented the utilization of the slowest motion mode of GNM to identify crucial functional residues in catalytic sites [172] and ligand-binding sites [173]. The fluctuation profiles of mouse apoA-IV were exhibited in the second slowest mode. The significant fluctuations were considered to play a driving role in protein conformational transitions. As depicted in Fig. 5.5 c, the residue 55 resided near the minima and the residue 121 located at the peak of the second slowest mode. The calculation results indicate that the truncated peptide 55-121 (T55-121) may be

a potential functional peptide of mouse apoA-IV.

Thus, peptide T55-121 and other two peptides (T140-212 and T245-316) of mouse apoA-IV were synthesized using solid-phase synthesis methods (Fig. 5.5 d). The glucose tolerance test showed that only the treatment with T55-121 resulted in improved glucose tolerance while the other two peptides did not exhibit any effects (Fig. 5.5 e). Furthermore, T55-121 was revealed to promote insulin secretion in MIN6 cells (Fig. 5.5 f). In line with the effect of full-length apoA-IV on enhancing energy expenditure, T55-121 significantly increased oxygen consumption (Fig. 5.5 g).

These results show that the murine T55-121 of apoA-IV exhibits the ability to improve glucose tolerance and increase oxygen consumption.

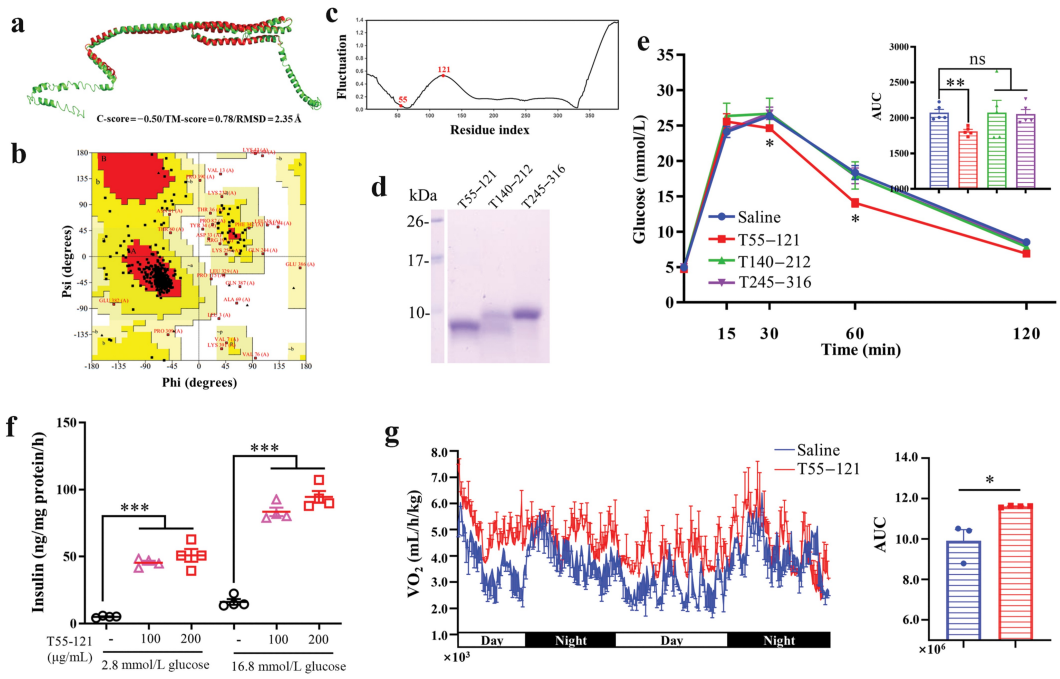


**Figure 5.5** T55–121 derived from apoA-IV improves glucose tolerance. (**a-c**) Bioinformatics analysis to predict functional peptides of apoA-IV. Model 1 of the 3D structure of mouse apoA-IV (**a**). Ramachandran plot shows the distributions of phi and psi angles with z-score of apoA-IV Model 1 (**b**). Fluctuation profiles in the second slowest mode for apoA-IV (**c**). (**d**) SDS-PAGE of the peptides synthesized by solid-phase synthesis methods. (**e**) The IPGTT in WT mice (n = 5) with saline, T55–121, T140–212, or T245–316 (10 mg/kg body weight) treatment. (**f**) GSIS assay in MIN6 cells treated with 0, 100, or 200 μg/mL T55–121. (**g**) $VO_2$ of WT mice following the administration of saline or T55–121. The right panel is the AUC in the left panel. Data are presented as mean ± SEM. *P < 0.05, **P < 0.01, ***P < 0.001.

## 5.4 Discussion

Bariatric/metabolic surgery is currently the most effective strategy for treating patients with severe obesity and type 2 diabetes. The procedure results in metabolic benefits through multiple mechanisms including caloric restriction, elevated GLP-1 levels, altered bile acids and an altered microbiome composition [174, 175]. Both RYGB and LSG have similar benefits of durable weight loss, reduced hyperlipidemia, and improved glucose homeostasis [176]. Therefore, it is crucial to identify the key factors that undergo significant changes following bariatric surgeries to enhance our understanding of the mechanisms involved in glucose control. In this study, through gel-based proteomics and immunoblotting analysis, a dramatic increase in serum apoA-IV levels in patients following LSG is demonstrated, consistent with results in other bariatric surgeries [157, 158, 165, 177]. Therefore, apoA-IV may play an important role in glycemic control in humans. Indeed, for the first time we have demonstrated that apoA-IV could enhance insulin secretion in human islets (Fig. 5.3).

In current study, we observed that both the administration of recombinant apoA-IV protein (Fig. 5.2; Fig. S5.4) and the in vivo overexpression of apoA-IV using AAV (Fig. 5.3) led to significant improvements in glucose tolerance in various diabetic rodent models. However, there appears to be a contradictory phenotype in apoA-IV-deficiency mice and rats. ApoA-IV-KO mice displayed impaired glucose tolerance [161], whereas apoA-IV-deficiency rats exhibited the opposite phenotype with improved glucose clearance [178]. The discrepancy could potentially be attributed to variations in host species and the specific gene knockout procedures. Further investigation is needed to clarify these contrasting outcomes.

Regarding the potential mechanism of apoA-IV in regulating glucose homeostasis, we demonstrated that apoA-IV promoted GSIS in islet cells, partially via a G$\alpha$s-coupled GPCR/cAMP pathway. Regarding the apoA-IV receptor, Xiaominng Li et al. found that apoA-IV directly interacts with nuclear receptors NR1D1 and NR4A1, to reduce gluconeogenesis in hepatic cells [162, 163]. In other work it was found that apoA-IV can bind to apoA-I or apoA-II recognition sites at the cell surface to regulate cholesterol efflux in adipose cells [179]. It was recently reported that apoA-IV binds low-density lipoprotein receptor-related protein 1 (LRP1) to facilitate glucose uptake in adipose tissue [180]. In summary, there may be multiple binding sites for apoA-IV, contributing to the regulation of metabolism.

Several mimetic peptides derived from apolipoproteins have been investigated and developed as potential therapeutic agents [181, 182]. ApoA-I mimetic peptides promote the formation of HDLs and facilitates the cholesterol efflux [183, 184]. ApoE mimetic peptides bind to LDL receptor-related protein and play a significant role in reducing inflammation and preventing cardiovascular diseases as well as neurodegenerative diseases [185, 186, 187]. Here, we discovered that T55-121 was a functional peptide of murine apoA-IV, which can improve glucose tolerance, providing a foundation for developing possible therapeutic interventions targeting glucose-related disorders. Furthermore, denatured apoA-IV showed a similar effect to the native form, leading to improved glucose toler-

ance. This suggests that apoA-IV and the identified T55-121 peptide may perform their biological functions independently of their secondary structure (Fig. S5.5). Further, we have performed an alignment analysis of the apoA-IV functional peptide T55-121 with other apolipoprotein mimetic peptides, including apoA-I, apoA-E, and apoC-II. The alignment result did not reveal any common features among these peptides (Fig. S5.6). Understanding the differences and unique characteristics of those mimetic peptides will contribute to a more comprehensive understanding of their functions and therapeutic potentials.

Bariatric/metabolic surgeries benefiting patients with severe T2D offer insight into factors involved in maintaining glucose homeostasis. In this study, serum proteomic analysis revealed a distinct protein profile in patients after LSG, with a significant increase in apoA-IV levels, and similar effects were observed in patients after RYGB. ApoA-IV may play a critical role in glycemic control, potentially by enhancing insulin secretion in human islets. ApoA-IV and its derived peptide, T55-121, were demonstrated to improve energy expenditure and glucose tolerance in diabetic rodents, partially through G$\alpha$s-coupled GPCR/cAMP signaling. These findings enhance our understanding of apoA-IV's connection to glycemic control, emphasizing its potential as a biomarker or therapeutic target for improving glucose regulation.

## 5.5 Materials and methods

### 5.5.1 Materials

The commercial materials and reagents used in the present study are listed in Table S5.1 and were applied as the manufacturers' procedures described. The peptides of apoA-IV were synthesized by GL Biochem (Shanghai) Ltd.

### 5.5.2 Human specimen

The blood samples were collected from two diabetic obese patients before and one year after LSG. Patient characteristics are listed in Table S5.2. The use of human blood samples was approved by the Ethics Committee of the First Affiliated Hospital, College of Medicine, Zhejiang University. Human islets from nondiabetic donors were isolated through Collagenase NB1 and Neutral Protease NB digestion followed by continuous density purification [188]. Then islets were picked up by hand and cultured in CMRL-1066 medium, supplemented with 10% human serum albumin at 37°C in 5% $CO_2$. The characteristics of nine non-diabetic subjects are shown in Table S5.3. The use of human islets was approved by the Ethics Committee of the Tianjin First Central Hospital. All procedures were performed in accordance with institutional guidelines and written informed consents were obtained from all the participants.

### 5.5.3 LC-MS/MS for protein identification

The sera collected from patients were precipitated by the addition of a 1.5-fold volume of acetonitrile followed by vortex mixing. The samples were centrifuged at 20,000 g for 5 min at 4°C and the supernatants were collected for evaporation using a CentriVap Benchtop Vacuum Concentrator (LABCONCO, USA). The samples were separated by SDS-PAGE and were stained with Coomassie Brilliant Blue according to manufacturers' procedure. The protein bands of interest were cut and subjected to enzymatic hydrolysis. The peptide mixture was analyzed by the Laboratory of Proteomics, Institute of Biophysics, Chinese Academy of Sciences.

### 5.5.4 Animals

Wild type C57BL/6J mice were obtained from Vital River Laboratories (Beijing, China). Diabetic $Lepr^{db}$ (db/db) mice and obese $Lep^{ob}$ (ob/ob) mice were purchased from Nanjing Biomedical Research Institute of Nanjing University. Spontaneous type 2 diabetic KKAy mice were obtained from Beijing HFK Bioscience Co. Ltd (Beijing, China), and Goto-Kakizaki (GK) diabetic rats were generously supplied by Prof. Tao Xu's Lab. All animals were maintained under a 12-h light-dark cycle at 22 ± 1℃. The KKAy mice were fed with a high-fat diet (Research Diets, Inc., D12451, USA), the others were fed with standard rodent chow. All mice were used between 6 and 24 weeks of age. All animal experimental procedures were approved by the Animal Care and Use Committee of the Institute of Biophysics, Chinese Academy of Sciences.

### 5.5.5 Construction of animal models

For the apoA-IV-overexpressing mice, 16-week-old ob/ob mice were infected with adeno-associated virus (AAV) encoding mouse apoA-IV (apoA-IV-AAV) or control adeno-associated virus encoding GFP (Con-AAV) via intraperitoneal injection. To mimic the expression of endogenous apoA-IV, an Adeno-associated virus 9 (AAV9) expressing apoA-IV-Flag under a CMV promoter was constructed, which was administered via intraperitoneal injection to infect the small intestine. The mice received a single dose of 2.0×1012 vg of AAV9. ApoA-IV-AAV and Con-AAV were produced by Hanbio Biotechnology Co., LTD. (Shanghai, China). The type 1 diabetic mouse model induced by streptozocin (STZ) was described previously [189]. STZ was dissolved to 0.1 mol/L in sodium citrate buffer (pH=4.2). 10-week-old wild-type C57BL/6J mice were injected once intraperitoneally with STZ (150 mg/kg) and were used when their random blood glucose level was constantly above 20 mM.

### 5.5.6 Intraperitoneal glucose tolerance test

For the intraperitoneal glucose tolerance test (IPGTT), mice were fasted for 16 h and injected intraperitoneally with saline or apoA-IV. Two hours later glucose was injected intraperitoneally (2 g/kg). Tail blood samples were collected and glucose levels were measured with a commercial glucometer (Ascensia ELITE; Bayer) at 0, 15, 30, 60, and 120 min post challenge.

### 5.5.7 Measurement of indirect calorimetry

Wild type mice or db/db mice, fasted overnight for 16 h, were administered saline or apoA-IV intraperitoneally. Two hours after injection, indirect calorimetry was performed using a TSE Systems instrument (Germany) following the manufacture's procedure. The data were analyzed using GraphPad Prism.

### 5.5.8 Primary islet preparation

The islets of KKAy mice were isolated via the collagenase V perfusion method and were cultured in RPMI-1640 medium supplemented with 10% FBS. The detailed procedure was reported previously [190].

### 5.5.9 Cell culture

The mouse $\beta$ cell line MIN6 was a gift of Prof. Xiao Han in Nanjing Medical University. The cells were cultured in DMEM supplemented with 15% FBS.

### 5.5.10 Expression and purification of recombinant protein apoA-IV

The construction of the prokaryotic expression system was based on methods described in a previous report [169]. The eukaryotic expression system was constructed to obtain endotoxin-free recombinant apoA-IV. The nucleotide sequence of human serum albumin signal peptide (SPHSA) was

inserted at the 5' end of HIS-SMT3-apoA-IV to improve yield. HEK293F suspension cells were transiently transfected with the pcDNA3.1-SPHSA-HIS-SMT3-apoA-IV plasmid. The secreted apoA-IV was purified with Ni Sepharose 6 Fast Flow chromatography according to the manufacturer's guidelines. The recombinant apoA-IV from eukaryotic expression system constructed by our laboratory was only used in glucose tolerance test in WT mice. Endotoxin-free recombinant apoA-IV used in other experiments were obtained from DetaiBio Co., Ltd (Nanjing, China).

### 5.5.11 Immunoblot

For the immunoblot, proteins were dissolved in 2 × sample buffer, denatured at 95°C for 5 min, separated by SDS-PAGE, and transferred to a PVDF membrane. The membrane was incubated in primary and then secondary antibodies (listed in Table S5.1). The image was detected by an ECL chemiluminescence system (SAGECREATION, China). The intensities of bands were analyzed by the software Image J.

### 5.5.12 Fluorescence section of tissues

The pancreas and brain of Goto-Kakizaki (GK) diabetic rats (4 weeks) were dissected and frozen in OCT Tissue Tek Compound, and cut into 5 $\mu$m and 15 $\mu$m thick sections, respectively. They were incubated with 10 $\mu$g/mL of apoA-IV-GFP or GFP at room temperature for 2 h, and then washed three times with PBS. The signals were detected by an inverted fluorescence microscope (Zeiss).

### 5.5.13 Insulin measurement

MIN6 cells, human primary islets, and mouse primary islets were used to test insulin secretion in response to different concentrations of glucose [190]. MIN6 cells were treated with vehicle, or 100 $\mu$g/mL of sm-apoA-IV for 3 h, or sm-apoA-IV together with SQ22536 or NF449. Isolated human primary islets and mouse primary islets were treated with vehicle or 100 $\mu$g/mL of sh-apopA-IV and sm-apoA-IV for 3 h. The glucose-stimulated insulin secretion assay was carried out in Kreb's buffer containing 2.8 or 16.8 mM glucose. The supernatant was collected for measurement of insulin by ELISA, and cells and islets were used for quantitative determination of protein by BCA protein assay kit.

### 5.5.14 Measurement of [cAMP]i, [DAG]i, and [Ca2+]i

For measurement of [cAMP]i and [DAG]i, MIN6 cells were transfected with a cAMP or DAG green downward fluorescent sensor and cultured for 24 h [191]. Then cells were treated with 0, 100, or 200 $\mu$g/mL of sm-apoA-IV for 3 h. The fluorescence of the cAMP sensor or DAG sensor was detected after treatment with 2.8 mM or 16.8 mM glucose using a PerkinElmer EnSpire2300 Multimode Microplate Reader (Synergy 2, BIO-TEK, USA). For measurement of [Ca2+]i, MIN6 cells in 96-well plates were cultured with 0, 100 or 200 $\mu$g/mL sm-apoA-IV for 1 h. Then they were loaded with the Ca2+ fluorescent sensor Fluo-4 AM for 30 min at 37°C. The cells in Krebs buffer containing 2.8 mM

glucose were the basic situation. For stimulation, the cells were perfused with Krebs buffer containing 16.8 mM glucose with sm-apoA-IV or vehicle. The changes in [Ca2+]i were recorded using a Multimode Microplate Reader through time-series recording of the fluorescent signal [192].

### 5.5.15 The prediction of the potential functional peptide of apoA-IV

The modeling approaches and the Gaussian network model (GNM) were used to predict the potential functional peptide of apoA-IV [193]. Protein Data Bank (PDB) was used for searching the structure of mouse apoA-IV [194]. Since the experimental 3D structure of mouse apoA-IV is unavailable, a 3D model of mouse apoA-IV was constructed using I-TASSER [195], with structure of human apoA-IV (PDB ID: 3S84) as a template [196]. The selection of the model was based on two criteria, including C-score indicated by confidence score provided by the I-TASSER server, and stereochemical properties evaluated using the PROCHECK tool. In the GNM, the 3D structure of a protein is considered as an elastic network of C$\alpha$ atoms, in which nodes are connected by harmonic springs with a force constant $\gamma$. The internal energy of the system in GNM are defined as

$$V = \frac{1}{2}\gamma \left[\Delta \mathbf{R}^T (\Gamma \otimes E) \Delta \mathbf{R}\right]$$

where $\Delta\mathbf{R}$ represents the $3N$-dimensional column vector of fluctuation $\Delta R_1, \Delta R_2, \ldots, \Delta R_N$ of the C$\alpha$ atoms, where $N$ is the number of residues; the superscript $T$ denotes the transpose; $E$ is the third-order identity matrix; $\otimes$ is the direct product; and $\Gamma$ is the $N \times N$ symmetric Kirchhoff matrix.

$$\Gamma_{ij} = \begin{cases} -1, & if \quad i \neq j \; and \; R_{ij} \leq r_c \\ 0, & if \quad i \neq j \; and \; R_{ij} > r_c \\ -\sum_{j,j\neq 1}^{N} \Gamma_{ij}, & if \quad i = j \end{cases}$$

where $r_c$ is the cutoff distance and $R_{ij}$ is the distance between the $i$th and $j$th nodes. The mean-square fluctuation of each residue is in proportion to the diagonal elements of the pseudoinverse of Kirchhoff matrix.

$$\langle \Delta R_i \cdot \Delta R_i \rangle = \frac{3k_B T}{\gamma} \left[\Gamma^{-1}\right]_{ii}$$

where $k_B$ is the Boltzmann constant, and $T$ is the thermodynamic temperature. The pseudoinverse of Kirchhoff matrix can be decomposed as

$$\Gamma^{-1} = \sum_{k=2}^{N} \lambda_k^{-1} \mu_k \cdot \mu_k^T$$

where $\lambda_k$ and $\mu_k$ are the $k$th eigenvalue and eigenvector of Kirchhoff matrix, respectively.

### 5.5.16 Statistical Analysis

Results are presented as mean ± SEM. Statistical significance was determined by the two-tailed Student's *t*-test between two groups or ANOVA in multiple groups. $p<0.05$ was considered to be a significant difference.

## Supplementary data

Supplementary material is available at Life Metabolism online.

## Acknowledgments

We thank Dr. John Zehmer for his critical reading and useful suggestions. We also thank Dr. Lemin Zheng for providing purified ApoA-I. This work was supported by the National Natural Science Foundation of China (92357302, 32170787, and 32100557) and the National Key Research and Development Program of China (2018YFA0800700, 2023YFA1801103, and 2018YFA0800900). Researches on human islets were supported by the National Natural Science Foundation of Tianjin Municipal Human Resources and Social Security Bureau (XB202011).

## Author's contribution

P.L. supervised the project. P.L. performed study concept and design. Z.C., L.L., and Z.Z. performed the experiments, analyzed and interpreted data, and wrote manuscript. P.L., S.Z., Y.W., and T.X. critically revised the manuscript. Z.Z. and W.G. conducted the prediction of the potential functional peptide of mouse apoA-IV. M.Z. provided sera of patients following surgery. S.X. and L.C. performed proteomic experiments. L.W. isolated mouse islets. Q.Z., B.P., G.Z., and Q.Y. helped conduct animal experiments.

## Ethics approval

The use of human blood samples was approved by the Ethics Committee of the First Affiliated Hospital, College of Medicine, Zhejiang University. The use of human islets was approved by the Ethics Committee of the Tianjin First Central Hospital. All procedures were performed in accordance with institutional guidelines and written informed consents were obtained from all the participants. All animal experimental procedures were approved by the Animal Care and Use Committee of the Institute of Biophysics, Chinese Academy of Sciences.

## Data availability

All study data are included in the article and/or Supplementary Materials. Materials are available upon request.

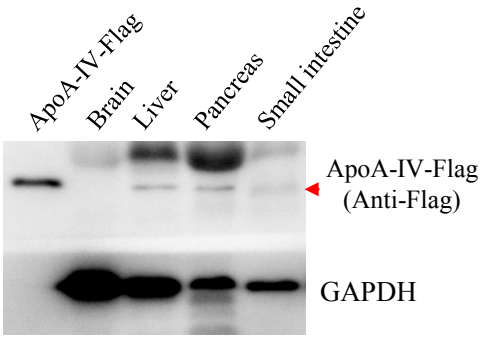


**Figure S5.1** The distribution of apoA-IV-Flag after AAV infection. Immunoblot analysis of apoA-IV-Flag in the brain, liver, pancreas, and small intestine in mouse infected with adeno-associated virus 9 expressing apoA-IV from a CMV promoter. Red arrowhead indicates apoA-IV-Flag.

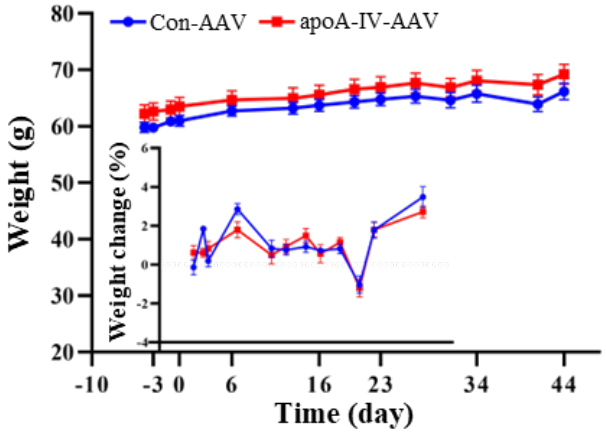


**Figure S5.2** Body weight in apoA-IV overexpressing ob/ob mice. Body weightand weight change (inlay on the lower left) of the Con-AAV infection group (n = 6) and apoA-IV-AAV infection group (n = 6).

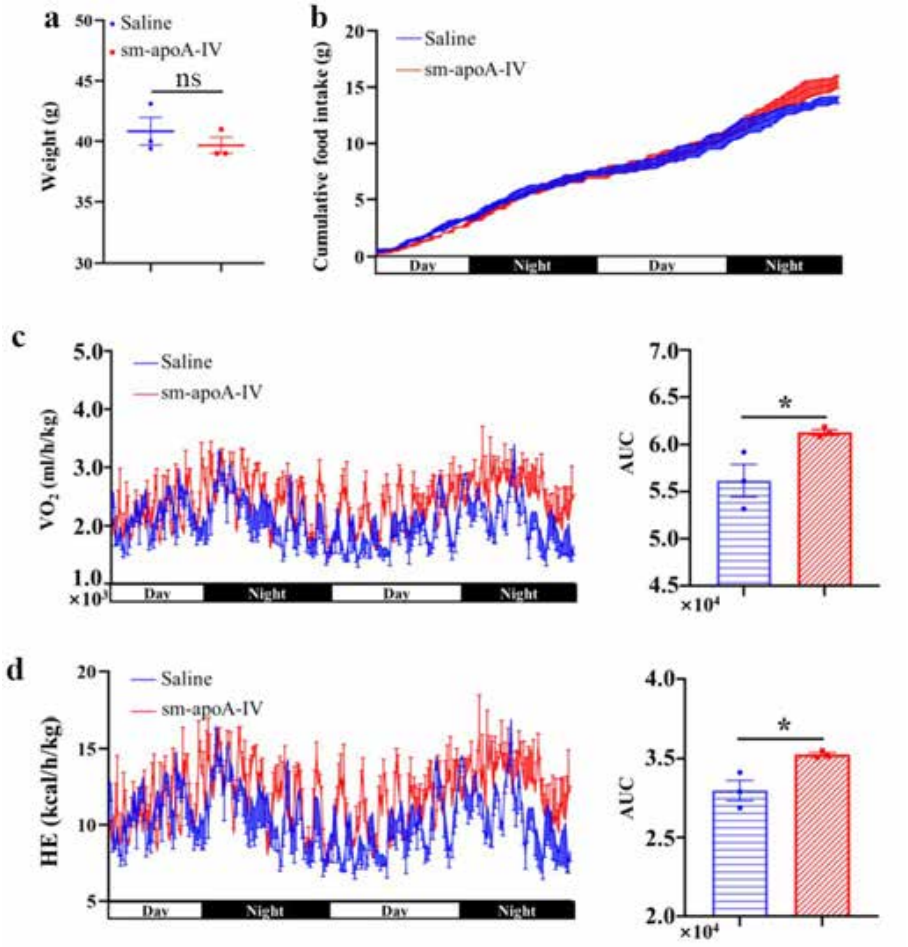


**Figure S5.3** ApoA-IV enhances energy expenditure in db/db mice. Analysis ofindirect calorimetry of db/db mice following the administration of saline or sm-apoA-IV (n = 3). **(a)** Bodyweight. **(b)** Cumulativefood intake. **(c)** $O_2$ consumption ($VO_2$). **(d)** Heat expenditure (HE). The right panel is the area under curve (AUC) (**c** and **d**).Data are presented as mean ± SEM. Statistical significance was determined by the two-tailedStudent's t-test. *P < 0.05.

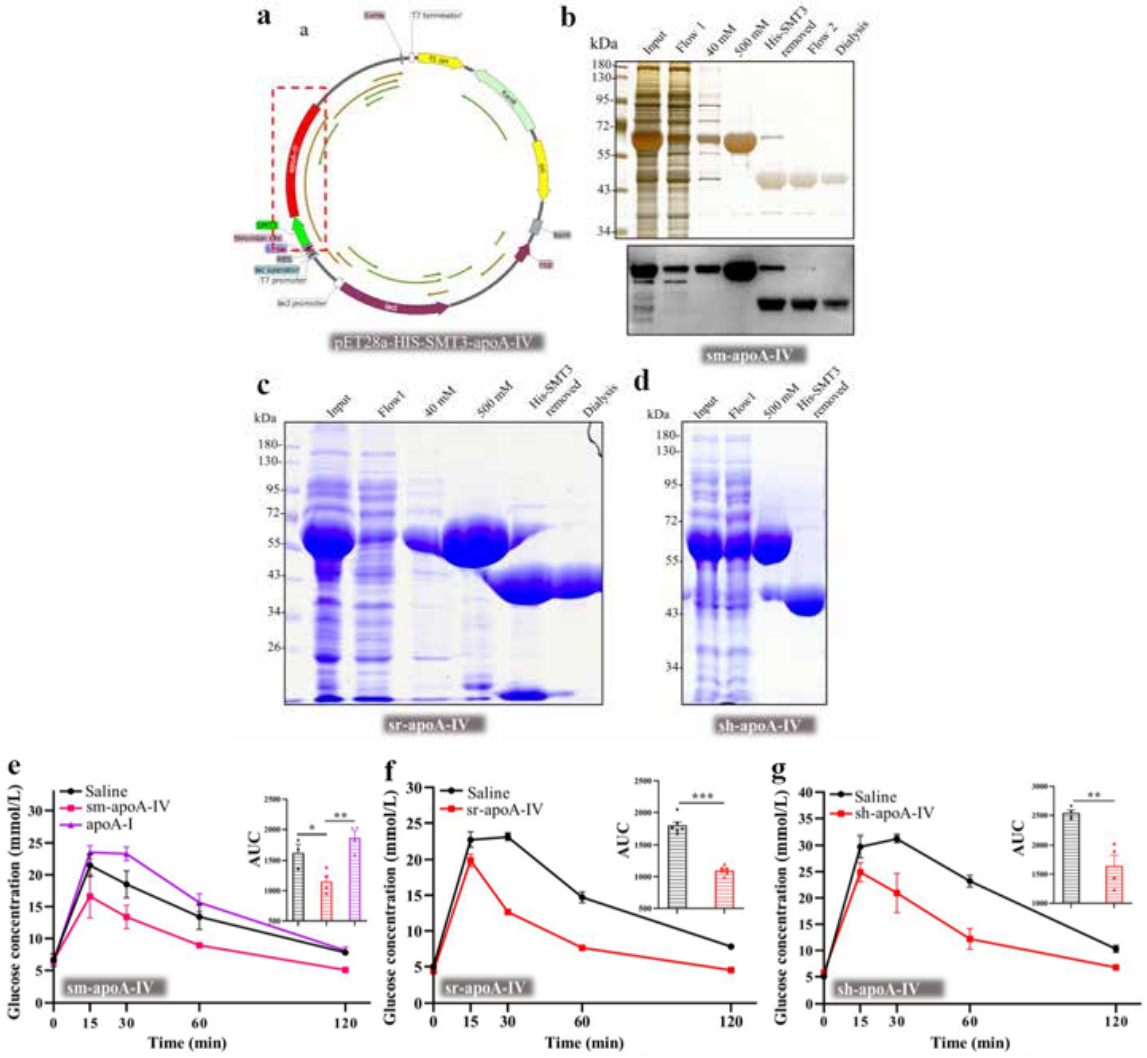


**Figure S5.4** Glucose tolerance is improved in WT mice with the administration of apoA-IV of different species. **(a–d)** Purification of different species of recombinant apoA-IV protein using a prokaryotic expression system. **(a)** Plasmid construct for recombinant apoA-IV protein expression system. The red dotted box represents the cloning site for insertion of apoA-IV nucleotide sequences of different species. **(b)** Purification of recombinant mouse apoA-IV. **(c)** Purification of recombinant rat apoA-IV. **(d)** Purification of recombinant human apoA-IV. **(e)** The IPGTT in WT mice (n = 3–4) with saline, sm-apoA-IV (1.5 mg/kg body weight) or apoA-I (1.5 mg/kg body weight) treatment. **(f)** The IPGTT in WT mice (n = 6) with saline or sr-apoA-IV (1.5 mg/kg body weight) treatment. **(g)** The IPGTT in WT mice (n = 4) with saline or sh-apoA-IV (1.5 mg/kg) treatment. Data are presented as mean ± SEM. Statistical significance was determined by the two-tailed Student's t-test. *P < 0.05, **P < 0.01, ***P < 0.001. IPGTT, intraperitoneal glucose tolerance test. AUC, area under the curve. sm-apoA-IV, signal peptide-removed mouse apoA-IV. sr-apoA-IV, signal peptide-removed rat apoA-IV. sh-apoA-IV, signal peptide-removed human apoA-IV.

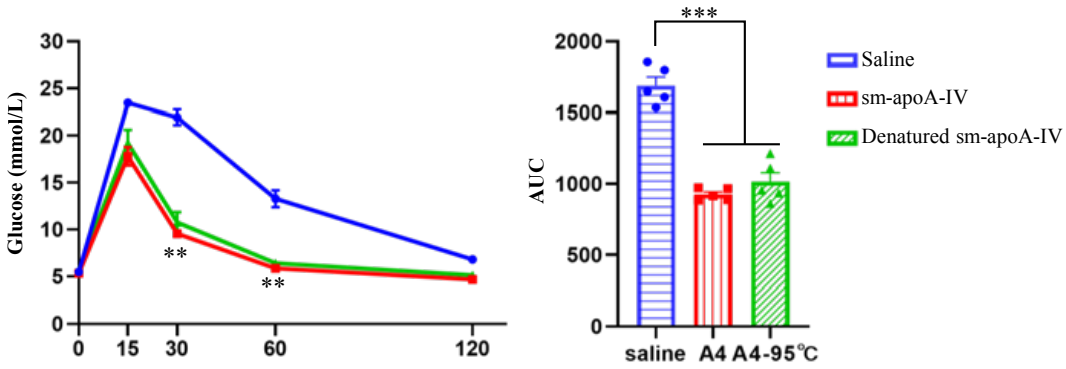


**Figure S5.5** TDenatured apoA-IV improves glucose tolerance. The recombinant sm-apoA-IV was subjected to denaturation by boiling at 95 °C for 10 min. Equal amount (6 mg/kg body weight) of denatured sm-apoA-IV or non-denatured sm-apoA-IV was administrated to mice (n = 5) for IPGTT, respectively. Data are presented as mean ± SEM. Statistical significance was determined by the two-tailed Student's t-test. **P < 0.01, ***P < 0.001. sm-apoA-IV, signal peptide-removed mouse apoA-IV.


**Figure S5.6** The alignment of apoA-IV functional peptide T55–121 with other apolipoprotein mimetic peptides. ApoA-IV functional peptide T55–121 was aligned with other apolipoprotein mimetic peptides, including apoA-I, apoA-E, and apoC-II via Clustal Omega website. ApoA-IV functional peptide T55–121 was aligned with other apolipoprotein mimetic peptides, including apoA-I, apoA-E, and apoC-II via Clustal Omega website.

| Number | Accession | Description | Score | MW [kDa] | Coverage | Proteins | Unique Peptides | Peptides | PSMs | AAs | calc. pI |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | P06727 | Apolipoprotein A-IV OS=Homo sapiens GN=APOA4 PE=1 SV=3 - [APOA4_HUMAN] | 184.9 | 45.4 | 61.36 | 2 | 28 | 28 | 110 | 396 | 5.38 |
| 2 | P02763 | Alpha-1-acid glycoprotein 1 OS=Homo sapiens GN=ORM1 PE=1 SV=1 - [A1AG1_HUMAN] | 118.55 | 23.5 | 48.26 | 1 | 8 | 11 | 112 | 201 | 5.02 |
| 3 | P19652 | Alpha-1-acid glycoprotein 2 OS=Homo sapiens GN=ORM2 PE=1 SV=2 - [A1AG2_HUMAN] | 63.15 | 23.6 | 35.82 | 1 | 5 | 8 | 61 | 201 | 5.11 |
| 4 | P0C0L4-2 | Isoform 2 of Complement C4-A OS=Homo sapiens GN=C4A - [CO4A_HUMAN] | 38.25 | 187.6 | 6.83 | 4 | 3 | 11 | 26 | 1698 | 7.12 |
| 5 | F5GXS0 | C4b-B OS=Homo sapiens GN=C4B PE=4 SV=1 - [F5GXS0_HUMAN] | 37.21 | 187.6 | 6.83 | 2 | 3 | 11 | 27 | 1698 | 7.33 |
| 6 | P25311 | Zinc-alpha-2-glycoprotein OS=Homo sapiens GN=AZGP1 PE=1 SV=2 - [ZA2G_HUMAN] | 30.22 | 34.2 | 36.24 | 3 | 12 | 12 | 19 | 298 | 6.05 |
| 7 | P60709 | Actin, cytoplasmic 1 OS=Homo sapiens GN=ACTB PE=1 SV=1 - [ACTB_HUMAN] | 29.93 | 41.7 | 33.6 | 21 | 4 | 10 | 34 | 375 | 5.48 |
| 8 | P02765 | Alpha-2-HS-glycoprotein OS=Homo sapiens GN=AHSG PE=1 SV=1 - [FETUA_HUMAN] | 21.47 | 39.3 | 17.44 | 3 | 5 | 5 | 9 | 367 | 5.72 |
| 9 | P01009-2 | Isoform 2 of Alpha-1-antitrypsin OS=Homo sapiens GN=SERPINA1 - [A1AT_HUMAN] | 14.94 | 40.2 | 24.79 | 7 | 7 | 7 | 13 | 359 | 5.47 |
| 10 | P02766 | Transthyretin OS=Homo sapiens GN=TTR PE=1 SV=1 - [TTHY_HUMAN] | 11.51 | 15.9 | 18.37 | 1 | 3 | 3 | 5 | 147 | 5.76 |
| 11 | F5H0C8 | Enolase OS=Homo sapiens GN=ENO2 PE=2 SV=1 - [F5H0C8_HUMAN] | 8.01 | 34.7 | 6.98 | 8 | 1 | 1 | 3 | 315 | 4.87 |
| 12 | P02750 | Leucine-rich alpha-2-glycoprotein OS=Homo sapiens GN=LRG1 PE=1 SV=2 - [A2GL_HUMAN] | 7.93 | 38.2 | 11.24 | 1 | 4 | 4 | 5 | 347 | 6.95 |
| 13 | P05090 | Apolipoprotein D OS=Homo sapiens GN=APOD PE=1 SV=1 - [APOD_HUMAN] | 7.26 | 21.3 | 13.23 | 4 | 2 | 2 | 4 | 189 | 5.15 |
| 14 | P68032 | Actin, alpha cardiac muscle 1 OS=Homo sapiens GN=ACTC1 PE=1 SV=1 - [ACTC_HUMAN] | 5.74 | 42 | 26.79 | 16 | 2 | 8 | 17 | 377 | 5.39 |
| 15 | P02647 | Apolipoprotein A-I OS=Homo sapiens GN=APOA1 PE=1 SV=1 - [APOA1_HUMAN] | 4.87 | 30.8 | 8.99 | 2 | 2 | 2 | 3 | 267 | 5.76 |
| 16 | E5RK62 | SPARC (Fragment) OS=Homo sapiens GN=SPARC PE=2 SV=1 - [E5RK62_HUMAN] | 4.18 | 13.4 | 16.52 | 3 | 2 | 2 | 3 | 115 | 7.25 |
| 17 | P81605 | Dermcidin OS=Homo sapiens GN=DCD PE=1 SV=2 - [DCD_HUMAN] | 3.54 | 11.3 | 10 | 2 | 1 | 1 | 2 | 110 | 6.54 |
| 18 | P12259 | Coagulation factor V OS=Homo sapiens GN=F5 PE=1 SV=4 - [FA5_HUMAN] | 2.33 | 251.5 | 0.49 | 1 | 1 | 1 | 1 | 2224 | 6.05 |
| 19 | P01019 | Angiotensinogen OS=Homo sapiens GN=AGT PE=1 SV=1 - [ANGT_HUMAN] | 2.29 | 53.1 | 7.63 | 1 | 3 | 3 | 3 | 485 | 6.32 |
| 20 | P02652 | Apolipoprotein A-II OS=Homo sapiens GN=APOA2 PE=1 SV=1 - [APOA2_HUMAN] | 2.04 | 11.2 | 21 | 1 | 3 | 3 | 3 | 100 | 6.62 |
| 21 | Q5VY30 | Plasma retinol-binding protein(1-182) OS=Homo sapiens GN=RBP4 PE=2 SV=1 - [Q5VY30_HUMAN] | 1.6 | 22.9 | 5.03 | 2 | 1 | 1 | 1 | 199 | 6.09 |
| 22 | H7C1V2 | RalBP1-associated Eps domain-containing protein 1 (Fragment) OS=Homo sapiens GN=REPS1 PE=2 SV=1 - [H7C1V2_HUMAN] | 0 | 16.5 | 23.13 | 1 | 1 | 1 | 1 | 147 | 5.02 |
| 23 | O75643 | U5 small nuclear ribonucleoprotein 200 kDa helicase OS=Homo sapiens GN=SNRNP200 PE=1 SV=2 - [U520_HUMAN] | 0 | 244.4 | 0.47 | 1 | 1 | 1 | 1 | 2136 | 6.06 |
| 24 | O95445-2 | Isoform 2 of Apolipoprotein M OS=Homo sapiens GN=APOM - [APOM_HUMAN] | 0 | 13 | 16.38 | 3 | 2 | 2 | 2 | 116 | 7.75 |
| 25 | Q562R1 | Beta-actin-like protein 2 OS=Homo sapiens GN=ACTBL2 PE=1 SV=2 - [ACTBL_HUMAN] | 0 | 42 | 12.77 | 1 | 1 | 4 | 6 | 376 | 5.59 |
| 26 | Q6UXD5-6 | Isoform 6 of Seizure 6-like protein 2 OS=Homo sapiens GN=SEZ6L2 - [SE6L2_HUMAN] | 0 | 86.8 | 1.24 | 5 | 1 | 1 | 1 | 809 | 4.82 |

**Table S5.1** MS results of the band 1 from patient after LSG.

| Number | Accession | Description | Score | MW [kDa] | Coverage | Proteins | Unique Peptides | Peptides | PSMs | AAs | calc. pI |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | P06727 | Apolipoprotein A-IV OS=Homo sapiens GN=APOA4 PE=1 SV=3 - [APOA4_HUMAN] | 123.85 | 45.4 | 57.32 | 2 | 25 | 25 | 84 | 396 | 5.38 |
| 2 | P02763 | Alpha-1-acid glycoprotein 1 OS=Homo sapiens GN=ORM1 PE=1 SV=1 - [A1AG1_HUMAN] | 120.72 | 23.5 | 45.27 | 1 | 6 | 9 | 96 | 201 | 5.02 |
| 3 | P19652 | Alpha-1-acid glycoprotein 2 OS=Homo sapiens GN=ORM2 PE=1 SV=2 - [A1AG2_HUMAN] | 55.85 | 23.6 | 35.82 | 1 | 5 | 8 | 46 | 201 | 5.11 |
| 4 | P0C0L4-2 | Isoform 2 of Complement C4-A OS=Homo sapiens GN=C4A - [CO4A_HUMAN] | 34.32 | 187.6 | 7.89 | 4 | 2 | 11 | 26 | 1698 | 7.12 |
| 5 | F5GXS0 | C4b-B OS=Homo sapiens GN=C4B PE=4 SV=1 - [F5GXS0_HUMAN] | 31.87 | 187.6 | 7.89 | 2 | 2 | 11 | 26 | 1698 | 7.33 |
| 6 | P25311 | Zinc-alpha-2-glycoprotein OS=Homo sapiens GN=AZGP1 PE=1 SV=2 - [ZA2G_HUMAN] | 24.21 | 34.2 | 33.22 | 2 | 10 | 10 | 17 | 298 | 6.05 |
| 7 | P60709 | Actin, cytoplasmic 1 OS=Homo sapiens GN=ACTB PE=1 SV=1 - [ACTB_HUMAN] | 20.27 | 41.7 | 32.8 | 23 | 4 | 9 | 16 | 375 | 5.48 |
| 8 | P02765 | Alpha-2-HS-glycoprotein OS=Homo sapiens GN=AHSG PE=1 SV=1 - [FETUA_HUMAN] | 14.28 | 39.3 | 9.26 | 2 | 2 | 2 | 6 | 367 | 5.72 |
| 9 | Q5T8M8 | Actin, alpha skeletal muscle OS=Homo sapiens GN=ACTA1 PE=2 SV=1 - [Q5T8M8_HUMAN] | 8.3 | 32 | 26.13 | 16 | 1 | 6 | 8 | 287 | 5.41 |
| 10 | P01009-2 | Isoform 2 of Alpha-1-antitrypsin OS=Homo sapiens GN=SERPINA1 - [A1AT_HUMAN] | 8.1 | 40.2 | 21.73 | 4 | 5 | 5 | 7 | 359 | 5.47 |
| 11 | O95445-2 | Isoform 2 of Apolipoprotein M OS=Homo sapiens GN=APOM - [APOM_HUMAN] | 6.2 | 13 | 11.21 | 3 | 1 | 1 | 3 | 116 | 7.75 |
| 12 | P05090 | Apolipoprotein D OS=Homo sapiens GN=APOD PE=1 SV=1 - [APOD_HUMAN] | 5.68 | 21.3 | 17.46 | 3 | 3 | 3 | 3 | 189 | 5.15 |
| 13 | C9JKR2 | Albumin, isoform CRA_k OS=Homo sapiens GN=ALB PE=4 SV=1 - [C9JKR2_HUMAN] | 5.61 | 47.3 | 6 | 7 | 1 | 3 | 5 | 417 | 6.35 |
| 14 | P15085 | Carboxypeptidase A1 OS=Homo sapiens GN=CPA1 PE=1 SV=2 - [CBPA1_HUMAN] | 4.8 | 47.1 | 5.49 | 3 | 2 | 2 | 2 | 419 | 5.76 |
| 15 | P02750 | Leucine-rich alpha-2-glycoprotein OS=Homo sapiens GN=LRG1 PE=1 SV=2 - [A2GL_HUMAN] | 4.59 | 38.2 | 11.24 | 1 | 3 | 3 | 4 | 347 | 6.95 |
| 16 | H3BUX1 | Mesothelin (Fragment) OS=Homo sapiens GN=MSLN PE=2 SV=1 - [H3BUX1_HUMAN] | 2.27 | 43.8 | 3.52 | 6 | 1 | 1 | 2 | 398 | 6.37 |
| 17 | P02652 | Apolipoprotein A-II OS=Homo sapiens GN=APOA2 PE=1 SV=1 - [APOA2_HUMAN] | 2.24 | 11.2 | 11 | 1 | 1 | 1 | 1 | 100 | 6.62 |
| 18 | P12259 | Coagulation factor V OS=Homo sapiens GN=F5 PE=1 SV=4 - [FA5_HUMAN] | 2.15 | 251.5 | 0.49 | 1 | 1 | 1 | 1 | 2224 | 6.05 |
| 19 | P02647 | Apolipoprotein A-I OS=Homo sapiens GN=APOA1 PE=1 SV=1 - [APOA1_HUMAN] | 1.96 | 30.8 | 4.87 | 1 | 1 | 1 | 1 | 267 | 5.76 |
| 20 | F5GXS5 | Apolipoprotein F OS=Homo sapiens GN=APOF PE=2 SV=1 - [F5GXS5_HUMAN] | 1.61 | 33.4 | 4.55 | 2 | 1 | 1 | 1 | 308 | 5.45 |
| 21 | H7C5W5 | Peripherin (Fragment) OS=Homo sapiens GN=PRPH PE=3 SV=1 - [H7C5W5_HUMAN] | 0 | 22.9 | 3 | 21 | 1 | 1 | 1 | 200 | 5.29 |
| 22 | P01019 | Angiotensinogen OS=Homo sapiens GN=AGT PE=1 SV=1 - [ANGT_HUMAN] | 0 | 53.1 | 2.47 | 1 | 1 | 1 | 2 | 485 | 6.32 |
| 23 | Q5VY30 | Plasma retinol-binding protein(1-182) OS=Homo sapiens GN=RBP4 PE=2 SV=1 - [Q5VY30_HUMAN] | 0 | 22.9 | 5.03 | 2 | 1 | 1 | 1 | 199 | 6.09 |

**Table S5.2** MS results of the band 2 from patient after LSG.

**Table S5.3** The change of plasma apoA-IV levels in the patients before and one year after surgery.

| Subject number | Before surgery | One year after surgery |
|---|---|---|
| 1 | 1.142 | 2.046 |
| 2 | 1.398 | 1.948 |
| 3 | 0.974 | 1.497 |
| 4 | 0.906 | 1.674 |
| 5 | 0.995 | 1.486 |
| 6 | 0.914 | 1.18 |
| 7 | 0.922 | 1.812 |
| 8 | 1.085 | 1.414 |
| 9 | 1.063 | 1.81 |
| 10 | 0.898 | 1.607 |

**Table S5.4** Antibodies and reagents.

| Antibody | Product number | Manufacturer |
|---|---|---|
| Anti-mouse apoA-IV | PAB967Mu01 | CLOUD-CLONE |
| Anti-human apoA-IV | 1D6B6 | Cell Signaling Technology |
| Transferrin | A1448 | ABclonal |
| Flag | F1804 | Sigma-Aldrich |
| CREB | 9192 | Cell Signaling Technology |
| pCREB (Ser133) | 9191 | Cell Signaling Technology |
| GAPDH | AB2302 | Merck Millipore |
| HRP-labeled goat anti-mouse IgG | ZB-2305 | ZSGB-BIO |
| HRP-labeled goat anti-rabbit IgG | ZB-2301 | ZSGB-BIO |
| Alexa Fluor 594-conjugated goat anti-rabbit IgG | ZF-0516 | ZSGB-BIO |

| Reagent or kit name | Product number | Manufacturer |
|---|---|---|
| Dulbecco's Modified Eagle's Medium (DMEM) | CM10017 | MACGENE |
| CMRL 1066 medium | 21530-027 | Invitrogen Life Technologies |
| RPMI-1640 medium | C22400500BT | Invitrogen Life Technologies |
| Fetal bovine serum (FBS) | 10437-028 | GIBCO |
| Penicillin-Streptomycin | CC004 | MACGENE |
| Collagenase NB1 | DS17455.03 | SERVA |
| Neutral Protease NB | DS30303.01 | SERVA |
| Collagenase V | C9263 | Sigma-Aldrich |
| Bovine serum albumin | A1933 | Sigma-Aldrich |
| Streptozotocin | S0130 | Sigma-Aldrich |
| Polybrene | S2267 | Sigma-Aldrich |
| Triton X-100 | T9284 | Sigma-Aldrich |
| Insulin | P3375 | Beyotime |
| Isopropyl $\beta$-D-1-thiogalactopyranoside (IPTG) | 487 | VWR AMRESCO |
| Ni Sepharose 6 Fast Flow | 17526801 | GE Healthcare |
| PVDF membrane | ISEQ00010 | Merck Millipore |
| Coomassie brilliant blue | PA101 | TIANGEN |
| Adenylate cyclase inhibitor SQ22536 | HY-100396 | MedChemExpress |
| Gsα-selective antagonist NF449 | 1391 | ROCRIS Bioscience |
| Green Down cADDis cAMP Assay Kit | 0200G | Montana Molecular |
| Rat/Mouse Insulin ELISA Kit | EZRMI-13K | Merck Millipore |
| Human Insulin ELISA Kit | 27365 | Mercodia |
| Pierce BCA Protein Assay Kit | 23227 | Thermo Fisher Scientific |
| $Ca^{2+}$ sensing fluorescent probe Fluo-4 AM | F14201 | Thermo Fisher Scientific |

**Table S5.5** Information of subjects for LSG.

| | **Subject 1** | **Subject 2** |
|---|---|---|
| Gender | Male | Male |
| Age (years) | 25 | 23 |
| Height (cm) | 176 | 175 |
| Weight (kg) | 135 | 115 |
| BMI (kg/m2) | 42.93 | 37.55 |

**Table S5.6** Characteristic of subjects for donating primary islets.

| | |
|---|---|
| **Age (years)** | 27-62 |
| **Body weight (kg)** | 78.3±18.2 |
| **BMI (**$kg/m^2$**)** | 26.1±5.2 |
| **C-peptide (ng/mL)** | 13.2±6.7 |
| **HbA1c (%)** | 5.3±0.3 |

# Chapter 6

# Summary and Perspectives

The three preceding chapters detail the large majority, but not entirety, of my research learning and progress over my doctoral work. The two primary chapters of my thesis, both of which I am the lead author of the work, focused on developing my wetlab and bioinformatic expertise in the study of the microbiome (Chapter 3) and on developing my ability to design, manage, and complete an *in vivo* experimental study (Chapter 4). The final chapter was in a collaborative study to which I had contributed experimental work directly prior to arriving in Luxembourg, and on which I continued to assist in analysis and writing after starting my doctoral thesis. Thus, across these three chapters in addition to the specific skills acquired and discoveries made, I have also seen the full scientific process of discovery, albeit split across three independent (although conceptually related) studies. Due to significant delays from COVID, my professor had moved to Luxembourg only shortly before I arrived (3.5 months prior, instead of 7 months prior) and that and various other COVID delays meant the laboratory remained largely non-operational when I arrived. While this had some logistical downsides (a roughly 4 month delay on the experimental start of my thesis), it also meant I was involved on the very earliest stages of laboratory setup, as well as the experimental setup and design for Chapter 4, from mouse ordering, to experimental protocol design, to the ethical application. This experimental work only finished in November 2024, so I have also obtained hands-on experience with passing along a substantially-completed project to a new generation of PhD students and scientists, who will help finish the molecular analysis of this cohort. Similarly, I myself started in a similar position regarding Chapter 3: the mouse experimental work had been finished in 2019 shortly prior to my professor's move to Luxembourg, but only cursory molecular experiments had been performed. We received the biobank of tissues from this cohort in January 2021, at which point I started my primary thesis project (i.e. Chapter 3) on the microbiome. For this project, I was involved in every step of the project design and execution except for the initial cohort design and phenotyping. Thus, my "Chapter 3" microbiome project starts off at almost exactly the same conceptual point where my "Chapter 4" mouse phenotyping project leaves off. Finally, in Chapter 5, I have worked on the mechanistic understanding of a specific protein, ApoA4, involved in a specific pathway of glucose metabolism, where we focused on understanding precise cellular processes. For

this paper, I was also involved as a co-first author contributing to the project from the earliest draft experiments to the final publication, although I was not the principle scientist working on it. Thus, across the three chapters, I have worked first-hand on the entire scientific discovery process, both in technical terms as well as in the social networks necessary for designing, running, and handing-off projects that take many years and many minds and sets of hands to complete.

Each of the preceding three chapters has a discussion section regarding the specific findings and conclusions of each study independently. In this overall discussion to my thesis, I will focus on how these studies contributes conceptually to the broader scientific community, and how they tie in to my future career and research direction.

## 6.1 Discussion Chapter 3: on MG, MT, and Holistic Approaches to the Molecular Science of Aging

When I arrived in Luxembourg I received a tremendously large and nearly-unique biobank of several thousand aging samples, taken from seven tissues and hundreds of mice and various ages and environmental conditions, from my professor's recently-completed large aging study. My initial project design discussed with Prof. Williams when I interview and joined the laboratory was in fact to run high-throughput proteomics and transcriptomics on at least two of these other tissues, in order to complement, cross-reference, and extend the results that he had acquired (but not yet published or analyzed) in liver tissues from the same cohort. While I tested out protocols and performed a small pilot study in collaboration with Prof. Gunnar Dittmar early on in this project - and I had previous experience running and analyzing proteomics in Beijing prior to arriving in Luxembourg - various practical factors induced me to adapt my thesis direction after six months away from the proteome and towards the microbiome. Primarily, the liver proteome data had not been fully analyzed and thus there was some risk that performing proteomics in additional tissues would have experimental flaws or be excessively redundant. This paper was only completed in September 2021, nearly a year after the start of my PhD. Two further reasons affected this pivot in topic: Prof. Williams and I wanted to considered other possible directions to take my own project such that it could not be considered derivative of recent work of his own. Two prior pilot experiments had been performed in the fecal microbiome of two different BXD cohorts with Prof. Williams ([197] and [106]), and while these studies were relatively cursory analysis of short-read 16S sequence data, they proved that a larger study would certainly yield unique insight. The second reason was practical: I joined a completely new laboratory during the height of COVID, in an annex building across campus that housed only a small subset of the biology teams. As scientific networks is critical, especially in-person communication during experimental work and early collaboration steps, I was fortunate that the laboratory of Prof. Paul Wilmes was in the same location and that our laboratories shared equipment and space which facilitated spontaneous and rapid face-to-face communication. Prof. Wilmes's laboratory is focused on the study of the microbiome, and many members of his team were instrumental in helping me independently establish my own project, which became Chapter 3.

Regarding the data and scientific discoveries from this study, my work in Chapter 3 represents one of the first large-scale, multi-omics analyses of the cecal ecosystem in aging mice, integrating MG, MT, and host cecal mRNA data. By analyzing samples from more than 200 BXD mice across two dietary conditions and a range of ages, we were able to systematically dissect how diet, aging, and host genetics independently and interactively shape gut microbial composition and host gene expression. One of the key findings from this chapter was that dietary intervention exerted the most dominant effect on gut microbiota diversity and composition, but that the effect of aging became significantly more pronounced under HF conditions—both at the taxonomic and transcriptomic levels. This highlights a potential synergy between environmental and intrinsic aging processes that may accelerate molecular dysregulation in the gut.

Furthermore, using machine learning approaches, I demonstrated that gut microbial species profiles alone can accurately predict host phenotypes such as age and body weight, with classification AUCs reaching 0.84 and 0.92, respectively. This result underscores the utility of the gut microbiome as a non-invasive and information-rich biosensor of host physiological status. More importantly, by integrating microbial and host liver proteomic data, I built a compact 10-feature classifier combining five microbial species and five liver proteins that achieved an AUC of 0.95, emphasizing the power of cross-omics modeling even in settings with limited sample overlap. This multi-omics integration also allowed me to identify biologically meaningful correlations between key taxa and host pathways —such as the association between *Streptococcus_danieliae* and age-related hepatic proteins like *Por* and *Cyp4a10*—which may reflect microbiome–host co-regulation of metabolic aging. Finally, co-expression network analysis linked HF-enriched microbial taxa to immune regulatory genes such as *Ido1*, supporting mechanistic hypotheses about the role of microbiota in modulating host inflammatory tone during dietary stress.

This project not only allowed me to acquire deep expertise in high-dimensional data integration and network analysis, but also provided the conceptual and technical foundation for future work seeking to define causal microbe–host relationships in aging and metabolism. Ultimately, the experience of independently designing, executing, and interpreting this multi-omics study has been one of the most formative components of my PhD training.

## 6.2 Discussion Chapter 4: on Voluntary Exercise, Genetics, and Metabolic Disease Phenotyping

As mentioned above, I was involved on the entire project conceptualization and experimental aspects of this project, with the only initial guidelines being from an older paper of Prof. Williams's, where he identified several inbred strains of mice that have high propensity for voluntary exercise [19], but that which despite their exercise, have highly-varying outcomes in terms of metabolic diseases. Together we selected four strains to phenotype, with the aim of identifying one particular strain that would strongly show a phenotype where mice which exercise are metabolically healthy, and those which do not exercise develop metabolic disorders. This initial concept was because of a

longstanding theory based on mouse research performed exclusively in C57BL/6J mice, which has said that voluntary exercise is not considered sufficient to mitigate metabolic disease if animals are maintained on an *ad libitum* HF diet [125]. While this may indeed be the case for some individuals - and we did notice a universally higher caloric intake on animals which exercised - we consider it an important finding to show that there is nothing particularly special about B6, and depending on the strain studied, the interpretation of the effect of diet and exercise on obesity and diabetes will change.

Thus far, the work on this study has shown that BXD43 is a particularly interesting strain to study, in that it develops NAFLD and diabetes on HF, but only when sedentary, despite it being a fairly obese mouse regardless of its exercising behavior. Furthermore, we show that this works in *voluntary* exercise, which makes the design of future studies vastly more simplified and cheaper - as well as more ethical - as compared to the many long-term exercise studies in B6 which required the forced exercise of mice on treadmills. Some challenges remain for the follow-up grant application and study based on the work in this chapter, for instance the male mice would viciously fight over access to a shared running wheel, but separating males and keeping them individually housed for 6+ months is also not recommended due to the social needs of these animals. That is, individual housing has ethical and scientific concerns (stereotypic exercise due to boredom or depression may not be the same as 'normal' exercise), while communal housing also has ethical and scientific concerns (severe fighting and injuries, and a cage with a tremendous difference between the dominant and submissive male will yield noisy data). Fortunately, the female component of this study proceeded without such issues, and we were able to carry out detailed molecular profiling of both the microbiome and host tissues.

Notably, initial microbiome analyses from these female cohorts revealed that voluntary exercise significantly reshaped the gut microbial landscape. In particular, we observed consistent shifts in the abundance of specific taxa, such as members of the *Lachnospiraceae* family and *Clostridium_butyricum* species, both of which have been previously linked to SCFA production and improved glycemic control. Their enrichment in active mice suggests that some of the observed metabolic benefits of exercise may be mediated, at least in part, through gut microbial remodeling. These findings not only complement the host phenotyping results, but also open a mechanistic window into the host–microbe interface, highlighting the potential for gut-derived signals to modulate systemic energy metabolism. This line of inquiry provides a compelling basis for follow-up work, including FMT experiments to directly test whether exercise-induced microbial shifts are sufficient to confer metabolic resilience in sedentary hosts.

Building on these findings, a small pilot dataset was already generated and is included in this chapter —work that I helped conceptualize and co-supervised alongside a new PhD student, Arianna Lamanna. Arianna has since taken over much of the day-to-day running of the project, managing and phenotyping the fourth and final cohort detailed here. Over the past year, I have successfully transferred the experimental pipeline and data workflows to her. She is now preparing a fifth cohort that will test the causal effects of microbiota transfer from exercised to sedentary mice, with the

aim of validating whether the same microbial taxa emerge as consistent predictors or mediators of improved metabolic outcomes in an independent cohort.

## 6.3 Discussion Chapter 5: on the Mechanistic Understanding of ApoA-IV and Mechanistic Approaches to the Molecular Science of Diabetes

I employed bioinformatics methods to conduct the prediction of the Apolipoprotein A-IV functional peptide (T55-121). Subsequent mouse and cell experiments confirmed that T55-121 can exert effects similar to ApoA4. Identifying this functional peptide substantially increases the potential for utilizing this work in diabetes treatment, as synthesizing short peptides significantly reduces both time and cost compared to purifying or synthesizing full-length proteins.

In addition to identifying T55–121 as a novel functional peptide, this study highlights the translational value of serum proteomics in uncovering endogenous molecules associated with improved metabolic health. By leveraging the physiological perturbation induced by bariatric surgery—an intervention already known to ameliorate glucose dysregulation—we were able to pinpoint apoA-IV as a candidate regulator of glycemic control. The observation that apoA-IV levels increase consistently in both LSG and RYGB patients, in parallel with improved glucose parameters, provides strong human-relevant validation. This approach exemplifies how longitudinal human data can be integrated with animal and cellular models to bridge mechanistic discovery with clinical relevance. Importantly, the ability of recombinant apoA-IV to stimulate insulin secretion in human islets for the first time not only expands the known biology of this protein, but also highlights its promise as a therapeutic modulator of $\beta$-cell function.

Beyond identifying a new peptide candidate, this project also underscores the utility of computational structural biology in guiding peptide-based discovery. By combining Gaussian network modeling with domain-focused truncation, we were able to rationally define and validate a short bioactive peptide capable of reproducing the metabolic benefits of full-length apoA-IV, including improved glucose tolerance and increased energy expenditure. This strategy opens new avenues for peptide drug development that may bypass the challenges of protein purification, stability, or immunogenicity. Moreover, the cAMP signaling mechanism identified here complements earlier reports of apoA-IV activity on transcriptional and lipid signaling axes, suggesting multifaceted effects depending on tissue and context. Moving forward, systematic structure–function analyses of T55–121 and its human homologs could aid in optimizing therapeutic efficacy while minimizing off-target effects. Together, these findings demonstrate the value of integrating omics-guided discovery, mechanistic validation, and bioinformatic design to uncover novel therapeutic strategies for metabolic disease.

Across the entirety of my doctoral work, I have developed an integrated view of host–micro-

biome interactions, from complex in vivo phenotyping to mechanistic molecular dissection. Each of the three chapters contributes a unique lens to this overarching theme: Chapter 3 offers systems-level insights into how environmental and genetic factors shape gut ecosystems and their impact on aging; Chapter 4 demonstrates the necessity of incorporating genetic diversity and behavioral variability in metabolic research; and Chapter 5 explores targeted interventions at the molecular level. Together, these chapters illustrate the power of combining diverse methodologies—including metagenomics, transcriptomics, machine learning, experimental modeling, and mechanistic validation—to approach biological complexity from multiple scales. Through this multifaceted training, I have not only gained proficiency in individual experimental techniques and computational pipelines, but also a broader strategic perspective on how to design impactful biological research.

Looking forward, the scientific perspective I have developed during this thesis emphasizes both interdisciplinarity and translational potential. As the field increasingly moves toward multi-omics, longitudinal, and intervention-based designs, I believe that the integration of computational modeling with mechanistic follow-up will be essential for generating actionable insights from complex datasets. I hope to further contribute to this space by building on my training in microbiome science, host physiology, and systems biology. In particular, I am interested in applying similar integrative frameworks to identify predictive microbial or metabolic markers of aging-related diseases, and to investigate causal pathways that could be targeted for therapeutic intervention. The intellectual foundation laid by this thesis has not only prepared me to pursue such questions, but also strengthened my commitment to collaborative and hypothesis-driven science.

# References


[1] Christopher B. Newgard. "Metabolomics and Metabolic Diseases: Where Do We Stand?" en. In: *Cell Metabolism* 25.1 (Jan. 2017), pp. 43–56. ISSN: 15504131. DOI: 10.1016/j.cmet.2016.09.018. URL: https://linkinghub.elsevier.com/retrieve/pii/S1550413116305034.

[2] Wolfgang Kopp. "How Western Diet And Lifestyle Drive The Pandemic Of Obesity And Civilization Diseases". en. In: *Diabetes, Metabolic Syndrome and Obesity: Targets and Therapy* Volume 12 (Oct. 2019), pp. 2221–2236. ISSN: 1178-7007. DOI: 10.2147/DMSO.S216791. URL: https://www.dovepress.com/how-western-diet-and-lifestyle-drive-the-pandemic-of-obesity-and-civil-peer-reviewed-article-DMSO.

[3] W. P. T. James et al. "Nutrition and its role in human evolution". en. In: *Journal of Internal Medicine* 285.5 (May 2019), pp. 533–549. ISSN: 0954-6820, 1365-2796. DOI: 10.1111/joim.12878. URL: https://onlinelibrary.wiley.com/doi/10.1111/joim.12878.

[4] Gökhan S. Hotamisligil. "Inflammation, metaflammation and immunometabolic disorders". en. In: *Nature* 542.7640 (Feb. 2017), pp. 177–185. ISSN: 0028-0836, 1476-4687. DOI: 10.1038/nature21363. URL: https://www.nature.com/articles/nature21363.

[5] Vicente Javier Clemente-Suárez et al. "New Insights and Potential Therapeutic Interventions in Metabolic Diseases". en. In: *International Journal of Molecular Sciences* 24.13 (June 2023), p. 10672. ISSN: 1422-0067. DOI: 10.3390/ijms241310672. URL: https://www.mdpi.com/1422-0067/24/13/10672.

[6] Francis Galton. *Natural Inheritance*. Macmillan and Co., 1889. URL: chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://galton.org/books/natural-inheritance/pdf/galton-nat-inh-1up-clean.pdf.

[7] Rashmi Prasad and Leif Groop. "Genetics of Type 2 Diabetes—Pitfalls and Possibilities". en. In: *Genes* 6.1 (Mar. 2015), pp. 87–123. ISSN: 2073-4425. DOI: 10.3390/genes6010087. URL: https://www.mdpi.com/2073-4425/6/1/87.

[8] Peter D. Stenson et al. "The Human Gene Mutation Database (HGMD®): optimizing its use in a clinical diagnostic or research setting". en. In: *Human Genetics* 139.10 (Oct. 2020), pp. 1197–1207. ISSN: 0340-6717, 1432-1203. DOI: 10.1007/s00439-020-02199-3. URL: https://link.springer.com/10.1007/s00439-020-02199-3.

[9] Loïc Yengo et al. "A saturated map of common genetic variants associated with human height". en. In: *Nature* 610.7933 (Oct. 2022), pp. 704–712. ISSN: 0028-0836, 1476-4687. DOI: 10.1038/s41586-022-05275-y. URL: https://www.nature.com/articles/s41586-022-05275-y.

[10] Evan G. Williams et al. "Multiomic profiling of the liver across diets and age in a diverse mouse population". In: *Cell Systems* 13.1 (Jan. 2022), 43–57.e6. ISSN: 2405-4712. DOI: 10.1016/j.cels.2021.09.005.

[11] B Zhou et al. "Worldwide trends in diabetes since 1980: a pooled analysis of 751 population-based studies with $4 \cdot 4$ million participants". en. In: *The Lancet* 387.10027 (Apr. 2016), pp. 1513–1530. ISSN: 01406736. DOI: 10.1016/S0140-6736(16)00618-8. URL: https://linkinghub.elsevier.com/retrieve/pii/S0140673616006188.

[12] Phyu Phyu Khin, Jong Han Lee, and Hee-Sook Jun. "Pancreatic Beta-cell Dysfunction in Type 2 Diabetes". en. In: *European Journal of Inflammation* 21 (June 2023), p. 1721727X231154152. ISSN: 1721-727X, 2058-7392. DOI: 10.1177/1721727X231154152. URL: https://journals.sagepub.com/doi/10.1177/1721727X231154152.

[13] Michael P Czech. "Insulin action and resistance in obesity and type 2 diabetes". en. In: *Nature Medicine* 23.7 (July 2017), pp. 804–814. ISSN: 1078-8956, 1546-170X. DOI: 10.1038/nm.4350. URL: https://www.nature.com/articles/nm.4350.

[14] American Diabetes Association. "Diagnosis and Classification of Diabetes Mellitus". en. In: *Diabetes Care* 33.Supplement_1 (Jan. 2010), S62–S69. ISSN: 0149-5992, 1935-5548. DOI: 10.2337/dc10-S062. URL: https://diabetesjournals.org/care/article/33/Supplement_1/S62/25777/Diagnosis-and-Classification-of-Diabetes-Mellitus.

[15] Marfa Blanter et al. "Genetic and Environmental Interaction in Type 1 Diabetes: a Relationship Between Genetic Risk Alleles and Molecular Traits of Enterovirus Infection?" en. In: *Current Diabetes Reports* 19.9 (Sept. 2019), p. 82. ISSN: 1534-4827, 1539-0829. DOI: 10.1007/s11892-019-1192-8. URL: http://link.springer.com/10.1007/s11892-019-1192-8.

[16] Ralph A. DeFronzo, Roy Eldor, and Muhammad Abdul-Ghani. "Pathophysiologic Approach to Therapy in Patients With Newly Diagnosed Type 2 Diabetes". en. In: *Diabetes Care* 36.Supplement_2 (Aug. 2013), S127–S138. ISSN: 0149-5992, 1935-5548. DOI: 10.2337/dcS13-2011. URL: https://diabetesjournals.org/care/article/36/Supplement_2/S127/30480/Pathophysiologic-Approach-to-Therapy-in-Patients.

[17] Sudesna Chatterjee, Kamlesh Khunti, and Melanie J Davies. "Type 2 diabetes". en. In: *The Lancet* 389.10085 (June 2017), pp. 2239–2251. ISSN: 01406736. DOI: 10.1016/S0140-6736(17)30058-2. URL: https://linkinghub.elsevier.com/retrieve/pii/S0140673617300582.

[18] Marie Ng et al. "National-level and state-level prevalence of overweight and obesity among children, adolescents, and adults in the USA, 1990–2021, and forecasts up to 2050". In: *The Lancet* 404.10469 (Dec. 2024), pp. 2278–2298. ISSN: 01406736. DOI: 10.1016/S0140-6736(24)01548-4. URL: https://linkinghub.elsevier.com/retrieve/pii/S0140673624015484.

[19] Evan G. Williams et al. "Systems proteomics of liver mitochondria function". en. In: *Science* 352.6291 (June 2016), aad0189. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.aad0189. URL: https://www.science.org/doi/10.1126/science.aad0189.

[20] C.M. Povel, J.M.A. Boer, and E.J.M. Feskens. "Shared genetic variance between the features of the metabolic syndrome: Heritability studies". en. In: *Molecular Genetics and Metabolism* 104.4 (Dec. 2011), pp. 666–669. ISSN: 10967192. DOI: 10.1016/j.ymgme.2011.08.035. URL: https://linkinghub.elsevier.com/retrieve/pii/S1096719211003118.

[21] Deborah J W Lee, Ajla Hodzic Kuerec, and Andrea B Maier. "Targeting ageing with rapamycin and its derivatives in humans: a systematic review". In: *The Lancet Healthy Longevity* 5.2 (Feb. 2024), e152–e162. ISSN: 26667568. DOI: 10.1016/S2666-7568(23)00258-1. URL: https://linkinghub.elsevier.com/retrieve/pii/S2666756823002581.

[22] Karen Brown et al. "Resveratrol for the Management of Human Health: How Far Have We Come? A Systematic Review of Resveratrol Clinical Trials to Highlight Gaps and Opportunities". In: *International Journal of Molecular Sciences* 25.2 (Jan. 2024), p. 747. ISSN: 1422-0067. DOI: 10.3390/ijms25020747. URL: https://www.mdpi.com/1422-0067/25/2/747.

[23] Wostmann BS et al. "Dietary intake, energy metabolism, and excretory losses of adult male germfree Wistar rats". In: *Lab Anim Sci* (Feb. 1983), pp. 46–50. URL: https://pubmed.ncbi.nlm.nih.gov/6834773/.

[24] Aqib B. Chowdhury and Kosha J. Mehta. "Liver biopsy for assessment of chronic liver diseases: a synopsis". en. In: *Clinical and Experimental Medicine* 23.2 (Feb. 2022), pp. 273–285. ISSN: 1591-9528. DOI: 10.1007/s10238-022-00799-z. URL: https://link.springer.com/10.1007/s10238-022-00799-z.

[25] Jonathan I. Quinlan et al. "Feasibility, Efficacy, and Safety of Percutaneous Muscle Biopsies in Patients With Chronic Liver Disease". en. In: *Frontiers in Physiology* 12 (Feb. 2022), p. 817152. ISSN: 1664-042X. DOI: 10.3389/fphys.2021.817152. URL: https://www.frontiersin.org/articles/10.3389/fphys.2021.817152/full.

[26] Sok Kuan Wong et al. "Animal models of metabolic syndrome: a review". en. In: *Nutrition & Metabolism* 13.1 (Dec. 2016), p. 65. ISSN: 1743-7075. DOI: 10.1186/s12986-016-0123-9. URL: http://nutritionandmetabolism.biomedcentral.com/articles/10.1186/s12986-016-0123-9.

[27] Lieselotte Herberg and Douglas L. Coleman. "Laboratory animals exhibiting obesity and diabetes syndromes". en. In: *Metabolism* 26.1 (Jan. 1977), pp. 59–99. ISSN: 00260495. DOI: 10.1016/0026-0495(77)90128-7. URL: https://linkinghub.elsevier.com/retrieve/pii/0026049577901287.

[28] D. A. Rees and J. C. Alcolado. "Animal models of diabetes mellitus". en. In: *Diabetic Medicine* 22.4 (Apr. 2005), pp. 359–370. ISSN: 0742-3071, 1464-5491. DOI: 10.1111/j.1464-5491.2005.01499.x. URL: https://onlinelibrary.wiley.com/doi/10.1111/j.1464-5491.2005.01499.x.

[29] Michele Vacca et al. "An unbiased ranking of murine dietary models based on their proximity to human metabolic dysfunction-associated steatotic liver disease (MASLD)". In: *Nature Metabolism* 6.6 (June 2024), pp. 1178–1196. ISSN: 2522-5812. DOI: 10.1038/s42255-024-01043-6. URL: https://www.nature.com/articles/s42255-024-01043-6.

[30] Lukasz Krych et al. "Quantitatively Different, yet Qualitatively Alike: A Meta-Analysis of the Mouse Core Gut Microbiome with a View towards the Human Gut Microbiome". In: *PLoS ONE* 8.5 (May 2013). Ed. by Stefan Bereswill, e62578. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0062578. URL: https://dx.plos.org/10.1371/journal.pone.0062578.

[31] Lyne Fellmann et al. "Murine models for pharmacological studies of the metabolic syndrome". en. In: *Pharmacology & Therapeutics* 137.3 (Mar. 2013), pp. 331–340. ISSN: 01637258. DOI: 10.1016/j.pharmthera.2012.11.004. URL: https://linkinghub.elsevier.com/retrieve/pii/S0163725812002380.

[32] Pénélope A. Andreux et al. "Systems Genetics of Metabolism: The Use of the BXD Murine Reference Panel for Multiscalar Integration of Traits". en. In: *Cell* 150.6 (Sept. 2012), pp. 1287–1299. ISSN: 00928674. DOI: 10.1016/j.cell.2012.08.012. URL: https://linkinghub.elsevier.com/retrieve/pii/S0092867412010070.

[33] David G. Ashbrook et al. *The expanded BXD family of mice: A cohort for experimental systems genetics and precision medicine*. en. June 2019. DOI: 10.1101/672097. URL: http://biorxiv.org/lookup/doi/10.1101/672097.

[34] Kristen M. S. O'Connell et al. "Genetic background modifies CNS-mediated sensorimotor decline in the AD-BXD mouse model of genetic diversity in Alzheimer's disease". en. In: *Genes, Brain and Behavior* 18.8 (Nov. 2019), e12603. ISSN: 1601-1848, 1601-183X. DOI: 10.1111/gbb.12603. URL: https://onlinelibrary.wiley.com/doi/10.1111/gbb.12603.

[35] Airton C. Martins et al. "BXD Recombinant Inbred Mice as a Model to Study Neurotoxicity". en. In: *Biomolecules* 11.12 (Nov. 2021), p. 1762. ISSN: 2218-273X. DOI: 10.3390/biom11121762. URL: https://www.mdpi.com/2218-273X/11/12/1762.

[36] Kamaldeep K. Gill et al. "Using a mouse model to gain insights into developmental coordination disorder". en. In: *Genes, Brain and Behavior* 19.4 (Apr. 2020), e12647. ISSN: 1601-1848, 1601-183X. DOI: 10.1111/gbb.12647. URL: https://onlinelibrary.wiley.com/doi/10.1111/gbb.12647.

[37] Johan Auwerx et al. *Genetic and dietary determinants of gut microbiome-bile acid interactions in the BXD recombinant inbred mouse population.* en. Aug. 2024. DOI: 10.21203/rs.3.rs-4782355/v1. URL: https://www.researchsquare.com/article/rs-4782355/v1.

[38] Maximilian Kleinert et al. "Animal models of obesity and diabetes mellitus". en. In: *Nature Reviews Endocrinology* 14.3 (Mar. 2018), pp. 140–162. ISSN: 1759-5029, 1759-5037. DOI: 10.1038/nrendo.2017.161. URL: https://www.nature.com/articles/nrendo.2017.161.

[39] Rexiati Ruze et al. "Obesity and type 2 diabetes mellitus: connections in epidemiology, pathogenesis, and treatments". en. In: *Frontiers in Endocrinology* 14 (Apr. 2023), p. 1161521. ISSN: 1664-2392. DOI: 10.3389/fendo.2023.1161521. URL: https://www.frontiersin.org/articles/10.3389/fendo.2023.1161521/full.

[40] Ann M. Ingalls, Margaret M. Dickie, and G. D. Snell. "Obese, a new mutation in the house mouse". In: *Journal of Heredity* 41.12 (Nov. 1950), pp. 317–318. ISSN: 12. DOI: 10.1093/oxfordjournals.jhered.a106073. URL: https://doi.org/10.1093/oxfordjournals.jhered.a106073.

[41] Daniel Ferguson et al. "Mouse Embryonic Fibroblasts Protect ob/ob Mice From Obesity and Metabolic Complications". en. In: *Endocrinology* 159.9 (Sept. 2018), pp. 3275–3286. ISSN: 1945-7170. DOI: 10.1210/en.2018-00561. URL: https://academic.oup.com/endo/article/159/9/3275/5061420.

[42] Pieter Giesbertz et al. "Metabolite profiling in plasma and tissues of ob/ob and db/db mice identifies novel markers of obesity and type 2 diabetes". en. In: *Diabetologia* 58.9 (Sept. 2015), pp. 2133–2143. ISSN: 0012-186X, 1432-0428. DOI: 10.1007/s00125-015-3656-y. URL: http://link.springer.com/10.1007/s00125-015-3656-y.

[43] Kumar Sharma, Peter McCue, and Stephen R. Dunn. "Diabetic kidney disease in the *db* / *db* mouse". en. In: *American Journal of Physiology-Renal Physiology* 284.6 (June 2003), F1138–F1144. ISSN: 1931-857X, 1522-1466. DOI: 10.1152/ajprenal.00315.2002. URL: https://www.physiology.org/doi/10.1152/ajprenal.00315.2002.

[44] Oanh H. Do et al. "Changes in beta cell function occur in prediabetes and early disease in the Lepr db mouse model of diabetes". en. In: *Diabetologia* 59.6 (June 2016), pp. 1222–1230. ISSN: 0012-186X, 1432-0428. DOI: 10.1007/s00125-016-3942-3. URL: http://link.springer.com/10.1007/s00125-016-3942-3.

[45] Mingsheng Yang et al. "Gut Microbiota Composition and Structure of the Ob/Ob and Db/Db Mice". en. In: *International Journal of Endocrinology* 2019 (Mar. 2019), pp. 1–9. ISSN: 1687-8337, 1687-8345. DOI: 10.1155/2019/1394097. URL: https://www.hindawi.com/journals/ije/2019/1394097/.

[46] Eunyoung Lee et al. "Diet-Induced Obese Mice and Leptin-Deficient Lepob/ob Mice Exhibit Increased Circulating GIP Levels Produced by Different Mechanisms". en. In: *International Journal of Molecular Sciences* 20.18 (Sept. 2019), p. 4448. ISSN: 1422-0067. DOI: 10.3390/ijms20184448. URL: https://www.mdpi.com/1422-0067/20/18/4448.

[47] F. Baquero and C. Nombela. "The microbiome as a human organ". In: *Clinical Microbiology and Infection* 18 (July 2012), pp. 2–4. ISSN: 1198743X. DOI: 10.1111/j.1469-0691.2012.03916.x. URL: https://linkinghub.elsevier.com/retrieve/pii/S1198743X14609587.

[48] Ervin Reyniers Trexiler. "Rearing germfree albino rats". In: *Lobund reports* 1 (Nov. 1946), pp. 1–84.

[49] Julian Pleasants. "Rearing germfree cesarean-born rats, mice, and rabbits through weaning". In: *Annals of the New York Academy of Sciences* 78.1 (May 1959), pp. 116–126. DOI: 10.1111/j.1749-6632.1959.tb53099.x.

[50] G. Hörmannsperger, M. Schaubeck, and D. Haller. "Intestinal Microbiota in Animal Models of Inflammatory Diseases". en. In: *ILAR Journal* 56.2 (Aug. 2015), pp. 179–191. ISSN: 1084-2020, 1930-6180. DOI: 10.1093/ilar/ilv019. URL: https://academic.oup.com/ilarjournal/article-lookup/doi/10.1093/ilar/ilv019.

[51] Vanessa K. Ridaura et al. "Gut Microbiota from Twins Discordant for Obesity Modulate Metabolism in Mice". en. In: *Science* 341.6150 (Sept. 2013), p. 1241214. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.1241214. URL: https://www.science.org/doi/10.1126/science.1241214.

[52] Jeremy K. Nicholson et al. "Host-Gut Microbiota Metabolic Interactions". en. In: *Science* 336.6086 (June 2012), pp. 1262–1267. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.1223813. URL: https://www.science.org/doi/10.1126/science.1223813.

[53] Takahiro Hosokawa et al. "Strict Host-Symbiont Cospeciation and Reductive Genome Evolution in Insect Gut Bacteria". en. In: *PLoS Biology* 4.10 (Oct. 2006). Ed. by Jonathan Eisen, e337. ISSN: 1545-7885. DOI: 10.1371/journal.pbio.0040337. URL: https://dx.plos.org/10.1371/journal.pbio.0040337.

[54] Daniel A. Leffler and J. Thomas Lamont. "*Clostridium difficile* Infection". en. In: *New England Journal of Medicine* 372.16 (Apr. 2015). Ed. by Dan L. Longo, pp. 1539–1548. ISSN: 0028-4793, 1533-4406. DOI: 10.1056/NEJMra1403772. URL: http://www.nejm.org/doi/10.1056/NEJMra1403772.

[55] Johan S. Bakken et al. “Treating Clostridium difficile Infection With Fecal Microbiota Transplantation”. en. In: *Clinical Gastroenterology and Hepatology* 9.12 (Dec. 2011), pp. 1044–1049. ISSN: 15423565. DOI: 10.1016/j.cgh.2011.08.014. URL: https://linkinghub.elsevier.com/retrieve/pii/S1542356511008913.

[56] Neha Alang and Colleen R. Kelly. “Weight Gain After Fecal Microbiota Transplantation”. en. In: *Open Forum Infectious Diseases* 2.1 (Jan. 2015), ofv004. ISSN: 2328-8957. DOI: 10.1093/ofid/ofv004. URL: https://academic.oup.com/ofid/article/doi/10.1093/ofid/ofv004/1461242.

[57] Li Zecheng et al. “Fecal microbiota transplantation in obesity metabolism: A meta analysis and systematic review”. en. In: *Diabetes Research and Clinical Practice* 202 (Aug. 2023), p. 110803. ISSN: 01688227. DOI: 10.1016/j.diabres.2023.110803. URL: https://linkinghub.elsevier.com/retrieve/pii/S0168822723005661.

[58] Cynthia L. Hsu and Bernd Schnabl. “The gut–liver axis and gut microbiota in health and liver disease”. en. In: *Nature Reviews Microbiology* 21.11 (Nov. 2023), pp. 719–733. ISSN: 1740-1526, 1740-1534. DOI: 10.1038/s41579-023-00904-3. URL: https://www.nature.com/articles/s41579-023-00904-3.

[59] Paola Brescia and Maria Rescigno. “The gut vascular barrier: a new player in the gut–liver–brain axis”. en. In: *Trends in Molecular Medicine* 27.9 (Sept. 2021), pp. 844–855. ISSN: 14714914. DOI: 10.1016/j.molmed.2021.06.007. URL: https://linkinghub.elsevier.com/retrieve/pii/S1471491421001544.

[60] Malin E V Johansson et al. “Bacteria penetrate the normally impenetrable inner colon mucus layer in both murine colitis models and patients with ulcerative colitis”. en. In: *Gut* 63.2 (Feb. 2014), pp. 281–291. ISSN: 0017-5749, 1468-3288. DOI: 10.1136/gutjnl-2012-303207. URL: https://gut.bmj.com/lookup/doi/10.1136/gutjnl-2012-303207.

[61] Kaushal Parikh et al. “Colonic epithelial cell diversity in health and inflammatory bowel disease”. en. In: *Nature* 567.7746 (Mar. 2019), pp. 49–55. ISSN: 0028-0836, 1476-4687. DOI: 10.1038/s41586-019-0992-y. URL: https://www.nature.com/articles/s41586-019-0992-y.

[62] Sjoerd Van Der Post et al. “Structural weakening of the colonic mucus barrier is an early event in ulcerative colitis pathogenesis”. en. In: *Gut* 68.12 (Dec. 2019), pp. 2142–2151. ISSN: 0017-5749, 1468-3288. DOI: 10.1136/gutjnl-2018-317571. URL: https://gut.bmj.com/lookup/doi/10.1136/gutjnl-2018-317571.

[63] Matthew A. Odenwald and Jerrold R. Turner. “The intestinal epithelial barrier: a therapeutic target?” en. In: *Nature Reviews Gastroenterology & Hepatology* 14.1 (Jan. 2017), pp. 9–21. ISSN: 1759-5045, 1759-5053. DOI: 10.1038/nrgastro.2016.169. URL: https://www.nature.com/articles/nrgastro.2016.169.

[64] D Rios et al. “Antigen sampling by intestinal M cells is the principal pathway initiating mucosal IgA production to commensal enteric bacteria”. en. In: *Mucosal Immunology* 9.4 (July 2016), pp. 907–916. ISSN: 19330219. DOI: 10.1038/mi.2015.121. URL: https://linkinghub.elsevier.com/retrieve/pii/S1933021922009151.

[65] Gregory F. Sonnenberg et al. “Innate Lymphoid Cells Promote Anatomical Containment of Lymphoid-Resident Commensal Bacteria”. en. In: *Science* 336.6086 (June 2012), pp. 1321–1325. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.1222551. URL: https://www.science.org/doi/10.1126/science.1222551.

[66] Kathrin Moor et al. “High-avidity IgA protects the intestine by enchaining growing bacteria”. en. In: *Nature* 544.7651 (Apr. 2017), pp. 498–502. ISSN: 0028-0836, 1476-4687. DOI: 10.1038/nature22058. URL: https://www.nature.com/articles/nature22058.

[67] Ilaria Spadoni et al. “A gut-vascular barrier controls the systemic dissemination of bacteria”. en. In: *Science* 350.6262 (Nov. 2015), pp. 830–834. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.aad0135. URL: https://www.science.org/doi/10.1126/science.aad0135.

[68] Christoph Reinhardt et al. “Tissue factor and PAR1 promote microbiota-induced intestinal vascular remodelling”. en. In: *Nature* 483.7391 (Mar. 2012), pp. 627–631. ISSN: 0028-0836, 1476-4687. DOI: 10.1038/nature10893. URL: https://www.nature.com/articles/nature10893.

[69] Panagiotis S. Kabouridis et al. “Microbiota Controls the Homeostasis of Glial Cells in the Gut Lamina Propria”. en. In: *Neuron* 85.2 (Jan. 2015), pp. 289–295. ISSN: 08966273. DOI: 10.1016/j.neuron.2014.12.037. URL: https://linkinghub.elsevier.com/retrieve/pii/S0896627314011465.

[70] Agustín Albillos, Andrea De Gottardi, and María Rescigno. “The gut-liver axis in liver disease: Pathophysiological basis for therapy”. en. In: *Journal of Hepatology* 72.3 (Mar. 2020), pp. 558–577. ISSN: 01688278. DOI: 10.1016/j.jhep.2019.10.003. URL: https://linkinghub.elsevier.com/retrieve/pii/S016882781930604X.

[71] Anupriya Tripathi et al. “The gut–liver axis and the intersection with the microbiome”. en. In: *Nature Reviews Gastroenterology & Hepatology* 15.7 (July 2018), pp. 397–411. ISSN: 1759-5045, 1759-5053. DOI: 10.1038/s41575-018-0011-z. URL: https://www.nature.com/articles/s41575-018-0011-z.

[72] Benjamin S. Beresford-Jones et al. “The Mouse Gastrointestinal Bacteria Catalogue enables translation between the mouse and human gut microbiotas via functional mapping”. en. In: *Cell Host & Microbe* 30.1 (Jan. 2022), 124–138.e8. ISSN: 19313128. DOI: 10.1016/j.chom.2021.12.003. URL: https://linkinghub.elsevier.com/retrieve/pii/S1931312821005680.

[73] Juliana Durack and Susan V. Lynch. "The gut microbiome: Relationships with disease and opportunities for therapy". In: *Journal of Experimental Medicine* 216.1 (Oct. 2018), pp. 20–40. ISSN: 1540-9538. DOI: 10.1084/jem.20180448.

[74] Sylvie Rabot et al. "Germ-free C57BL/6J mice are resistant to high-fat-diet-induced insulin resistance and have altered cholesterol metabolism". In: *The FASEB Journal* 24.12 (Dec. 2010), pp. 4948–4959. ISSN: 1530-6860. DOI: 10.1096/fj.10-164921.

[75] June L. Round and Sarkis K. Mazmanian. "The gut microbiota shapes intestinal immune responses during health and disease". In: *Nature Reviews Immunology* 9.5 (May 2009), pp. 313–323. ISSN: 1474-1741. DOI: 10.1038/nri2515.

[76] Elizabeth A. Kennedy, Katherine Y. King, and Megan T. Baldridge. "Mouse Microbiota Models: Comparing Germ-Free Mice and Antibiotics Treatment as Tools for Modifying Gut Bacteria". en. In: *Frontiers in Physiology* 9 (Oct. 2018). Publisher: Frontiers Media SA. ISSN: 1664-042X. DOI: 10.3389/fphys.2018.01534. URL: https://www.frontiersin.org/article/10.3389/fphys.2018.01534/full (visited on 07/17/2025).

[77] Maria Elisa Caetano-Silva et al. "Aging amplifies a gut microbiota immunogenic signature linked to heightened inflammation". en. In: *Aging Cell* 23.8 (Aug. 2024). Publisher: Wiley. ISSN: 1474-9718, 1474-9726. DOI: 10.1111/acel.14190. URL: https://onlinelibrary.wiley.com/doi/10.1111/acel.14190 (visited on 07/17/2025).

[78] Tarini Shankar Ghosh, Fergus Shanahan, and Paul W. O'Toole. "The gut microbiome as a modulator of healthy ageing". In: *Nature Reviews Gastroenterology & Hepatology* 19.9 (Sept. 2022), pp. 565–584. ISSN: 1759-5045, 1759-5053. DOI: 10.1038/s41575-022-00605-x.

[79] Ara Koh et al. "From Dietary Fiber to Host Physiology: Short-Chain Fatty Acids as Key Bacterial Metabolites". en. In: *Cell* 165.6 (June 2016), pp. 1332–1345. ISSN: 00928674. DOI: 10.1016/j.cell.2016.05.041. URL: https://linkinghub.elsevier.com/retrieve/pii/S009286741630592X.

[80] Parasuraman Padmanabhan et al. "Gastrointestinal transit measurements in mice with 99mTc-DTPA-labeled activated charcoal using NanoSPECT-CT". en. In: *EJNMMI Research* 3.1 (2013). Publisher: Springer Science and Business Media LLC, p. 60. ISSN: 2191-219X. DOI: 10.1186/2191-219x-3-60. URL: http://ejnmmires.springeropen.com/articles/10.1186/2191-219X-3-60 (visited on 07/17/2025).

[81] Giorgia Greter et al. *Acute targeted induction of gut-microbial metabolism affects host clock genes and nocturnal feeding*. en. Sept. 2024. DOI: 10.7554/elife.100640.1. URL: https://elifesciences.org/reviewed-preprints/100640v1 (visited on 07/17/2025).

[82] Maëva Meynier et al. "Pasteurized *Akkermansia muciniphila* improves irritable bowel syndrome-like symptoms and related behavioral disorders in mice". en. In: *Gut Microbes* 16.1 (Dec. 2024), p. 2298026. ISSN: 1949-0976, 1949-0984. DOI: 10.1080/19490976.2023.2

298026. URL: https://www.tandfonline.com/doi/full/10.1080/19490976.2023.2298026.

[83] Colleen R. Kelly et al. "Fecal Microbiota Transplantation Is Highly Effective in Real-World Practice: Initial Results From the FMT National Registry". en. In: *Gastroenterology* 160.1 (Jan. 2021). Publisher: Elsevier BV, 183–192.e3. ISSN: 0016-5085. DOI: 10.1053/j.gastro.2020.09.038. URL: https://linkinghub.elsevier.com/retrieve/pii/S0016508520352215 (visited on 07/18/2025).

[84] Dayoon Kwon et al. "Diet and the gut microbiome in patients with Parkinson's disease". en. In: *npj Parkinson's Disease* 10.1 (Apr. 2024), p. 89. ISSN: 2373-8057. DOI: 10.1038/s41531-024-00681-7. URL: https://www.nature.com/articles/s41531-024-00681-7.

[85] Xiaoxiao Yuan et al. "Functional and metabolic alterations of gut microbiota in children with new-onset type 1 diabetes". en. In: *Nature Communications* 13.1 (Oct. 2022), p. 6356. ISSN: 2041-1723. DOI: 10.1038/s41467-022-33656-4. URL: https://www.nature.com/articles/s41467-022-33656-4.

[86] Ge Yang et al. "Role of the gut microbiota in type 2 diabetes and related diseases". en. In: *Metabolism* 117 (Apr. 2021), p. 154712. ISSN: 00260495. DOI: 10.1016/j.metabol.2021.154712. URL: https://linkinghub.elsevier.com/retrieve/pii/S0026049521000123.

[87] Zhang et al. "Impact of Fecal Microbiota Transplantation on Obesity and Metabolic Syndrome—A Systematic Review". In: *Nutrients* 11.10 (Sept. 2019), p. 2291. ISSN: 2072-6643. DOI: 10.3390/nu11102291.

[88] S. Roy et al. "Gene-by-environmental modulation of lifespan and weight gain in the murine BXD family". In: *Nature Metabolism* (2021).

[89] David G. Ashbrook et al. "A platform for experimental precision medicine: The extended BXD mouse family". In: *Cell Systems* 12.3 (Mar. 2021), 235–247.e9. ISSN: 2405-4712. DOI: 10.1016/j.cels.2020.12.002.

[90] Chanhee Kang et al. "The DNA damage response induces inflammation and senescence by inhibiting autophagy of GATA4". en. In: *Science* 349.6255 (Sept. 2015), aaa5612. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.aaa5612. URL: https://www.science.org/doi/10.1126/science.aaa5612.

[91] Juri Kazakevych et al. "Transcriptome analysis identifies a robust gene expression program in the mouse intestinal epithelium on aging". en. In: *Scientific Reports* 9.1 (July 2019), p. 10410. ISSN: 2045-2322. DOI: 10.1038/s41598-019-46966-3. URL: https://www.nature.com/articles/s41598-019-46966-3.

[92] Jose Martinez-Lomeli et al. *Impact of Various High Fat Diets on Gene Expression and the Microbiome Across the Mouse Intestines*. en. Oct. 2023. DOI: 10.21203/rs.3.rs-3401763/v1. URL: https://www.researchsquare.com/article/rs-3401763/v1.

[93] Evert M. Van Schothorst et al. "Adipose Gene Expression Patterns of Weight Gain Suggest Counteracting Steroid Hormone Synthesis". en. In: *Obesity Research* 13.6 (June 2005). Publisher: Wiley, pp. 1031–1041. ISSN: 1071-7323, 1550-8528. DOI: 10.1038/oby.2005.121. URL: https://onlinelibrary.wiley.com/doi/10.1038/oby.2005.121 (visited on 07/28/2012).

[94] Zheng-Mei Lv et al. "Dysregulation of mitochondrial dynamics and mitophagy are involved in high-fat diet-induced steroidogenesis inhibition". en. In: *Journal of Lipid Research* 65.10 (Oct. 2024). Publisher: Elsevier BV, p. 100639. ISSN: 0022-2275. DOI: 10.1016/j.jlr.2024.100639. URL: https://linkinghub.elsevier.com/retrieve/pii/S0022227524001445 (visited on 07/28/2025).

[95] Y. Sun et al. "PYY-mediated appetite control and obesity alleviation through short-chain fatty acid–driven gut-brain axis modulation by Lacticaseibacillus rhamnosus HF01 isolated from Qula". en. In: *Journal of Dairy Science* 108.8 (Aug. 2025). Publisher: American Dairy Science Association, pp. 7960–7978. ISSN: 0022-0302. DOI: 10.3168/jds.2024-26193. URL: https://linkinghub.elsevier.com/retrieve/pii/S0022030225003947 (visited on 07/28/2025).

[96] Daniela Frasca and Bonnie B Blomberg. "Effects of aging on B cell function". en. In: *Current Opinion in Immunology* 21.4 (Aug. 2009). Publisher: Elsevier BV, pp. 425–430. ISSN: 0952-7915. DOI: 10.1016/j.coi.2009.06.001. URL: https://linkinghub.elsevier.com/retrieve/pii/S0952791509001319 (visited on 07/29/2025).

[97] Jiahao Wang et al. "Characterizing age-related features for assessing biological age and characteristics in Xinjiang Brown cattle". en. In: *BMC Genomics* 26.1 (May 2025). Publisher: Springer Science and Business Media LLC. ISSN: 1471-2164. DOI: 10.1186/s12864-025-11430-2. URL: https://bmcgenomics.biomedcentral.com/articles/10.1186/s12864-025-11430-2 (visited on 07/29/2025).

[98] Xinwei Jiao et al. "Microbial Reconstitution Improves Aging-Driven Lacrimal Gland Circadian Dysfunction". en. In: *The American Journal of Pathology* 191.12 (Dec. 2021). Publisher: Elsevier BV, pp. 2091–2116. ISSN: 0002-9440. DOI: 10.1016/j.ajpath.2021.08.006. URL: https://linkinghub.elsevier.com/retrieve/pii/S0002944021003552 (visited on 07/29/2025).

[99] Francesco Beghini et al. "Integrating taxonomic, functional, and strain-level profiling of diverse microbial communities with bioBakery 3". en. In: *eLife* 10 (May 2021), e65088. ISSN: 2050-084X. DOI: 10.7554/eLife.65088. URL: https://elifesciences.org/articles/65088.

[100] Aitor Blanco-Míguez et al. "Extending and improving metagenomic taxonomic profiling with uncharacterized species using MetaPhlAn 4". en. In: *Nature Biotechnology* 41.11 (Nov. 2023). Publisher: Springer Science and Business Media LLC, pp. 1633–1644. ISSN: 1087-0156, 1546-

1696. DOI: 10.1038/s41587-023-01688-w. URL: https://www.nature.com/articles/s41587-023-01688-w (visited on 07/18/2025).

[101] Yuan Liu et al. "Gut microbiome alterations in high-fat-diet-fed mice are associated with antibiotic tolerance". en. In: *Nature Microbiology* 6.7 (May 2021), pp. 874–884. ISSN: 2058-5276. DOI: 10.1038/s41564-021-00912-0. URL: https://www.nature.com/articles/s41564-021-00912-0.

[102] Xianping Li et al. "Lactobacillus plantarum prevents obesity via modulation of gut microbiota and metabolites in high-fat feeding mice". en. In: *Journal of Functional Foods* 73 (Oct. 2020), p. 104103. ISSN: 17564646. DOI: 10.1016/j.jff.2020.104103. URL: https://linkinghub.elsevier.com/retrieve/pii/S1756464620303273.

[103] Spase Stojanov, Aleš Berlec, and Borut Štrukelj. "The Influence of Probiotics on the Firmicutes/Bacteroidetes Ratio in the Treatment of Obesity and Inflammatory Bowel disease". en. In: *Microorganisms* 8.11 (Nov. 2020). Publisher: MDPI AG, p. 1715. ISSN: 2076-2607. DOI: 10.3390/microorganisms8111715. URL: https://www.mdpi.com/2076-2607/8/11/1715 (visited on 07/18/2025).

[104] María Bailén et al. "Microbiota Features Associated With a High-Fat/Low-Fiber Diet in Healthy Adults". en. In: *Frontiers in Nutrition* 7 (Dec. 2020), p. 583608. ISSN: 2296-861X. DOI: 10.3389/fnut.2020.583608. URL: https://www.frontiersin.org/articles/10.3389/fnut.2020.583608/full.

[105] Rebecca H. Gellman et al. "Hadza Prevotella require diet-derived microbiota-accessible carbohydrates to persist in mice". en. In: *Cell Reports* 42.11 (Nov. 2023), p. 113233. ISSN: 22111247. DOI: 10.1016/j.celrep.2023.113233. URL: https://linkinghub.elsevier.com/retrieve/pii/S2211124723012457.

[106] Maria Elisa Perez-Munoz et al. "Diet modulates cecum bacterial diversity and physiological phenotypes across the BXD mouse genetic reference population". In: *PLOS ONE* 14.10 (Oct. 2019), e0224100. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0224100.

[107] Himel Mallick et al. "Multivariable association discovery in population-scale meta-omics studies". en. In: *PLOS Computational Biology* 17.11 (Nov. 2021). Ed. by Luis Pedro Coelho, e1009442. ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1009442. URL: https://dx.plos.org/10.1371/journal.pcbi.1009442.

[108] Fuyong Li et al. "Comparative metagenomic and metatranscriptomic analyses reveal the breed effect on the rumen microbiome and its associations with feed efficiency in beef cattle". en. In: *Microbiome* 7.1 (Dec. 2019). Publisher: Springer Science and Business Media LLC. ISSN: 2049-2618. DOI: 10.1186/s40168-019-0618-5. URL: https://microbiomejournal.biomedcentral.com/articles/10.1186/s40168-019-0618-5 (visited on 07/22/2025).

[109] Boping Li et al. "In-depth multiomic characterization of the effects of obesity in high-fat diet-fed mice". en. In: *FEBS Open Bio* 14.5 (May 2024), pp. 771–792. ISSN: 2211-5463, 2211-5463. DOI: 10.1002/2211-5463.13788. URL: https://febs.onlinelibrary.wiley.com/doi/10.1002/2211-5463.13788 (visited on 07/30/2025).

[110] XuWen Mao et al. "Modulation of gut microbiota, up-regulation of ZO-1, and promotion of metabolism as therapeutic mechanisms of indole-3-carbinol against obesity in mice". en. In: *Frontiers in Pharmacology* 15 (Jan. 2025), p. 1499142. ISSN: 1663-9812. DOI: 10.3389/fphar.2024.1499142. URL: https://www.frontiersin.org/articles/10.3389/fphar.2024.1499142/full (visited on 07/30/2025).

[111] Jingjing Wang et al. "The landscape in the gut microbiome of long-lived families reveals new insights on longevity and aging –relevant neural and immune function". en. In: *Gut Microbes* 14.1 (Dec. 2022), p. 2107288. ISSN: 1949-0976, 1949-0984. DOI: 10.1080/19490976.2022.2107288. URL: https://www.tandfonline.com/doi/full/10.1080/19490976.2022.2107288 (visited on 07/30/2025).

[112] Yangli He et al. "Metabolites of the gut microbiota may serve as precise diagnostic markers for sarcopenia in the elderly". en. In: *Frontiers in Microbiology* 14 (Dec. 2023), p. 1301805. ISSN: 1664-302X. DOI: 10.3389/fmicb.2023.1301805. URL: https://www.frontiersin.org/articles/10.3389/fmicb.2023.1301805/full (visited on 07/30/2025).

[113] Andrew R Ghazi et al. "High-sensitivity pattern discovery in large, paired multiomic datasets". en. In: *Bioinformatics* 38.Supplement_1 (June 2022), pp. i378–i385. ISSN: 1367-4803, 1367-4811. DOI: 10.1093/bioinformatics/btac232. URL: https://academic.oup.com/bioinformatics/article/38/Supplement_1/i378/6617508.

[114] Ludivine Laurans et al. "Genetic deficiency of indoleamine 2,3-dioxygenase promotes gut microbiota-mediated metabolic health". en. In: *Nature Medicine* 24.8 (Aug. 2018), pp. 1113–1120. ISSN: 1078-8956, 1546-170X. DOI: 10.1038/s41591-018-0060-4. URL: https://www.nature.com/articles/s41591-018-0060-4.

[115] Chenglin Ye et al. "Association between Gut Microbiota and Biological Aging: A Two-Sample Mendelian Randomization Study". en. In: *Microorganisms* 12.2 (Feb. 2024), p. 370. ISSN: 2076-2607. DOI: 10.3390/microorganisms12020370. URL: https://www.mdpi.com/2076-2607/12/2/370.

[116] Hyeon Cheol Kim et al. "1 Title: Gut microbiota dysbiosis induced by brain tumor modulates the 2 efficacy of immunotherapy". en. In: *bioRxiv* (2024).

[117] Diana Ćesić et al. "Association of Gut Lachnospiraceae and Chronic Spontaneous Urticaria". en. In: *Life* 13.6 (May 2023), p. 1280. ISSN: 2075-1729. DOI: 10.3390/life13061280. URL: https://www.mdpi.com/2075-1729/13/6/1280.

[118] Yingjian Hou, Jing Li, and Shuhuan Ying. "Tryptophan Metabolism and Gut Microbiota: A Novel Regulatory Axis Integrating the Microbiome, Immunity, and Cancer". en. In: *Metabolites* 13.11 (Nov. 2023), p. 1166. ISSN: 2218-1989. DOI: 10.3390/metabo13111166. URL: https://www.mdpi.com/2218-1989/13/11/1166.

[119] Yehudit Hasin, Marcus Seldin, and Aldons Lusis. "Multi-omics approaches to disease". en. In: *Genome Biology* 18.1 (Dec. 2017), p. 83. ISSN: 1474-760X. DOI: 10.1186/s13059-017-1215-1. URL: https://genomebiology.biomedcentral.com/articles/10.1186/s13059-017-1215-1.

[120] Yang Liao, Gordon K Smyth, and Wei Shi. "The R package Rsubread is easier, faster, cheaper and better for alignment and quantification of RNA sequencing reads". en. In: *Nucleic Acids Research* 47.8 (May 2019), e47–e47. ISSN: 0305-1048, 1362-4962. DOI: 10.1093/nar/gkz114. URL: https://academic.oup.com/nar/article/47/8/e47/5345150.

[121] Charlotte Soneson, Michael I. Love, and Mark D. Robinson. "Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences". en. In: *F1000Research* 4 (Dec. 2015), p. 1521. ISSN: 2046-1402. DOI: 10.12688/f1000research.7563.1. URL: https://f1000research.com/articles/4-1521/v1.

[122] Paul Shannon et al. "Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks". en. In: *Genome Research* 13.11 (Nov. 2003), pp. 2498–2504. ISSN: 1088-9051. DOI: 10.1101/gr.1239303. URL: http://genome.cshlp.org/lookup/doi/10.1101/gr.1239303.

[123] S. Varrette et al. "Management of an Academic HPC & Research Computing Facility: The ULHPC Experience 2.0". In: *Proc. of the 6th ACM High Performance Computing and Cluster Technologies Conf. (HPCCT 2022)*. Fuzhou, China: Association for Computing Machinery (ACM), July 2022. ISBN: 978-1-4503-9664-6.

[124] Kevin D. Hall and Scott Kahan. "Maintenance of Lost Weight and Long-Term Management of Obesity". en. In: *Medical Clinics of North America* 102.1 (Jan. 2018), pp. 183–197. ISSN: 00257125. DOI: 10.1016/j.mcna.2017.08.012. URL: https://linkinghub.elsevier.com/retrieve/pii/S0025712517301360.

[125] Jean-Michel Oppert, Cécile Ciangura, and Alice Bellicha. "Physical activity and exercise for weight loss and maintenance in people living with obesity". en. In: *Reviews in Endocrine and Metabolic Disorders* 24.5 (Oct. 2023), pp. 937–949. ISSN: 1389-9155, 1573-2606. DOI: 10.1007/s11154-023-09805-5. URL: https://link.springer.com/10.1007/s11154-023-09805-5.

[126] Geng-Ruei Chang et al. "Exercise Affects Blood Glucose Levels and Tissue Chromium Distribution in High-Fat Diet-Fed C57BL6 Mice". en. In: *Molecules* 25.7 (Apr. 2020), p. 1658. ISSN: 1420-3049. DOI: 10.3390/molecules25071658. URL: https://www.mdpi.com/1420-3049/25/7/1658.

[127] Hiroyuki Sasaki et al. "Forced rather than voluntary exercise entrains peripheral clocks via a corticosterone/noradrenaline increase in PER2::LUC mice". en. In: *Scientific Reports* 6.1 (June 2016), p. 27607. ISSN: 2045-2322. DOI: 10.1038/srep27607. URL: https://www.nature.com/articles/srep27607.

[128] Kyle Hatton-Jones et al. "Stress-induced body weight loss and improvements in cardiometabolic risk factors do not translate to improved myocardial ischemic tolerance in western diet-fed mice". en. In: *Physiological Reports* 10.2 (Jan. 2022). ISSN: 2051-817X, 2051-817X. DOI: 10.14814/phy2.15170. URL: https://onlinelibrary.wiley.com/doi/10.14814/phy2.15170.

[129] Jasmin A. Hicks et al. "Voluntary wheel-running attenuates insulin and weight gain and affects anxiety-like behaviors in C57BL6/J mice exposed to a high-fat diet". en. In: *Behavioural Brain Research* 310 (Sept. 2016), pp. 1–10. ISSN: 01664328. DOI: 10.1016/j.bbr.2016.04.051. URL: https://linkinghub.elsevier.com/retrieve/pii/S0166432816302674.

[130] Alan P. Jung and David R. Luthin. "Wheel Access Does Not Attenuate Weight Gain in Mice Fed High-Fat or High-CHO Diets". en. In: *Medicine & Science in Sports & Exercise* 42.2 (Feb. 2010), pp. 355–360. ISSN: 0195-9131. DOI: 10.1249/MSS.0b013e3181a6d88f. URL: https://journals.lww.com/00005768-201002000-00017.

[131] Karin Pernold et al. "Towards large scale automated cage monitoring –Diurnal rhythm and impact of interventions on in-cage activity of C57BL/6J mice recorded 24/7 with a non-disrupting capacitive-based technique". en. In: *PLOS ONE* 14.2 (Feb. 2019). Ed. by Eric M. Mintz, e0211063. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0211063. URL: https://dx.plos.org/10.1371/journal.pone.0211063.

[132] P. A. Figueiredo et al. "Impact of Lifelong Sedentary Behavior on Mitochondrial Function of Mice Skeletal Muscle". en. In: *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences* 64A.9 (Sept. 2009), pp. 927–939. ISSN: 1079-5006, 1758-535X. DOI: 10.1093/gerona/glp066. URL: https://academic.oup.com/biomedgerontology/article-lookup/doi/10.1093/gerona/glp066.

[133] Karin Pernold, Eric Rullman, and Brun Ulfhake. "Major oscillations in spontaneous home-cage activity in C57BL/6 mice housed under constant conditions". en. In: *Scientific Reports* 11.1 (Mar. 2021), p. 4961. ISSN: 2045-2322. DOI: 10.1038/s41598-021-84141-9. URL: https://www.nature.com/articles/s41598-021-84141-9.

[134] Scott A. Kelly et al. "Exercise, weight loss, and changes in body composition in mice: phenotypic relationships and genetic architecture". en. In: *Physiological Genomics* 43.4 (Feb. 2011), pp. 199–212. ISSN: 1094-8341, 1531-2267. DOI: 10.1152/physiolgenomics.00217.2010. URL: https://www.physiology.org/doi/10.1152/physiolgenomics.00217.2010.

[135] Kyle D. Flack et al. "Energy compensation in response to aerobic exercise training in overweight adults". en. In: *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology* 315.4 (Oct. 2018), R619–R626. ISSN: 0363-6119, 1522-1490. DOI: 10.1152/ajpregu.00071.2018. URL: https://www.physiology.org/doi/10.1152/ajpregu.00071.2018.

[136] Joseph Choi et al. "Carnitine palmitoyltransferase 1 facilitates fatty acid oxidation in a non-cell-autonomous manner". en. In: *Cell Reports* 43.12 (Dec. 2024), p. 115006. ISSN: 22111247. DOI: 10.1016/j.celrep.2024.115006. URL: https://linkinghub.elsevier.com/retrieve/pii/S2211124724013573.

[137] Ashley S. Williams et al. "Ketone flux through BDH1 supports metabolic remodeling of skeletal and cardiac muscles in response to intermittent time-restricted feeding". en. In: *Cell Metabolism* 36.2 (Feb. 2024), 422–437.e8. ISSN: 15504131. DOI: 10.1016/j.cmet.2024.01.007. URL: https://linkinghub.elsevier.com/retrieve/pii/S155041312400007X.

[138] Anne Le et al. "Inhibition of lactate dehydrogenase A induces oxidative stress and inhibits tumor progression". en. In: *Proceedings of the National Academy of Sciences* 107.5 (Feb. 2010), pp. 2037–2042. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.0914433107. URL: https://pnas.org/doi/full/10.1073/pnas.0914433107.

[139] Marius Vital, Adina Chuang Howe, and James M. Tiedje. "Revealing the Bacterial Butyrate Synthesis Pathways by Analyzing (Meta)genomic Data". en. In: *mBio* 5.2 (May 2014). Ed. by Mary Ann Moran, e00889–14. ISSN: 2161-2129, 2150-7511. DOI: 10.1128/mBio.00889-14. URL: https://journals.asm.org/doi/10.1128/mBio.00889-14.

[140] Mehrbod Estaki et al. "Cardiorespiratory fitness as a predictor of intestinal microbial diversity and distinct metagenomic functions". en. In: *Microbiome* 4.1 (Dec. 2016), p. 42. ISSN: 2049-2618. DOI: 10.1186/s40168-016-0189-7. URL: http://microbiomejournal.biomedcentral.com/articles/10.1186/s40168-016-0189-7.

[141] Signe Schultz Pedersen et al. "Targeted Delivery of Butyrate Improves Glucose Homeostasis, Reduces Hepatic Lipid Accumulation and Inflammation in db/db Mice". en. In: *International Journal of Molecular Sciences* 24.5 (Feb. 2023), p. 4533. ISSN: 1422-0067. DOI: 10.3390/ijms24054533. URL: https://www.mdpi.com/1422-0067/24/5/4533.

[142] Jingyi Liao et al. "Clostridium butyricum Reduces Obesity in a Butyrate-Independent Way". en. In: *Microorganisms* 11.5 (May 2023), p. 1292. ISSN: 2076-2607. DOI: 10.3390/microorganisms11051292. URL: https://www.mdpi.com/2076-2607/11/5/1292.

[143] Alessandra Berry et al. "Social deprivation stress is a triggering factor for the emergence of anxiety- and depression-like behaviours and leads to reduced brain BDNF levels in C57BL/6J mice". en. In: *Psychoneuroendocrinology* 37.6 (June 2012), pp. 762–772. ISSN: 03064530. DOI: 10.1016/j.psyneuen.2011.09.007. URL: https://linkinghub.elsevier.com/retrieve/pii/S0306453011002770.

[144] Sharon S. Lander, Donna Linder-Shacham, and Inna Gaisler-Salomon. "Differential effects of social isolation in adolescent and adult mice on behavior and cortical gene expression". en. In: *Behavioural Brain Research* 316 (Jan. 2017), pp. 245–254. ISSN: 01664328. DOI: 10.1016/j.bbr.2016.09.005. URL: https://linkinghub.elsevier.com/retrieve/pii/S0166432816305897.

[145] Jin-Seok Lee, Ji-Yun Kang, and Chang-Gue Son. "A Comparison of Isolation Stress and Unpredictable Chronic Mild Stress for the Establishment of Mouse Models of Depressive Disorder". en. In: *Frontiers in Behavioral Neuroscience* 14 (Jan. 2021), p. 616389. ISSN: 1662-5153. DOI: 10.3389/fnbeh.2020.616389. URL: https://www.frontiersin.org/articles/10.3389/fnbeh.2020.616389/full.

[146] Zara Kanold-Tso and Dietmar Plenz. *Variability of wheel running behavior in mice is dependent on housing, sex, and genetic background*. en. Oct. 2022. DOI: 10.1101/2022.10.28.514233. URL: http://biorxiv.org/lookup/doi/10.1101/2022.10.28.514233.

[147] Josh Reed, Stephen Bain, and Venkateswarlu Kanamarlapudi. "A Review of Current Trends with Type 2 Diabetes Epidemiology, Aetiology, Pathogenesis, Treatments and Future Perspectives". en. In: *Diabetes, Metabolic Syndrome and Obesity: Targets and Therapy* Volume 14 (Aug. 2021), pp. 3567–3602. ISSN: 1178-7007. DOI: 10.2147/DMSO.S319895. URL: https://www.dovepress.com/a-review-of-current-trends-with-type-2-diabetes-epidemiology-aetiology-peer-reviewed-fulltext-article-DMSO.

[148] "Metabolic Surgery in the Treatment Algorithm for Type 2 Diabetes: A Joint Statement by International Diabetes Organizations". en. In: ().

[149] Khalid R. Murshid et al. "Laparoscopic sleeve gastrectomy for weight loss and treatment of type 2 diabetes mellitus". en. In: *Journal of Taibah University Medical Sciences* 16.3 (June 2021), pp. 387–394. ISSN: 16583612. DOI: 10.1016/j.jtumed.2020.12.018. URL: https://linkinghub.elsevier.com/retrieve/pii/S165836122100010X.

[150] Steven K. Malin and Sangeeta R. Kashyap. "Effects of various gastrointestinal procedures on β-cell function in obesity and type 2 diabetes". en. In: *Surgery for Obesity and Related Diseases* 12.6 (July 2016), pp. 1213–1219. ISSN: 15507289. DOI: 10.1016/j.soard.2016.02.035. URL: https://linkinghub.elsevier.com/retrieve/pii/S1550728916000800.

[151] Yalong Wang et al. "Single-cell transcriptome analysis reveals differential nutrient absorption functions in human intestine". en. In: *Journal of Experimental Medicine* 217.2 (Feb. 2020), e20191130. ISSN: 0022-1007, 1540-9538. DOI: 10.1084/jem.20191130. URL: https://rupress.org/jem/article/217/2/e20191130/132578/Single-cell-transcriptome-analysis-reveals.

[152] Alison B. Kohan et al. "ApoA-IV: current and emerging roles in intestinal lipid metabolism, glucose homeostasis, and satiety". en. In: *American Journal of Physiology-Gastrointestinal and Liver Physiology* 308.6 (Mar. 2015), G472–G481. ISSN: 0193-1857, 1522-1547. DOI: 10.11

52/ajpgi.00098.2014. URL: https://www.physiology.org/doi/10.1152/ajpgi.00098.2014.

[153] Theodore J Kalogeris. "Intestinal Synthesis and Lymphatic Secretion of Apolipoprotein A–IV Vary with Chain Length of Intestinally Infused Fatty Acids in Rats". en. In: ().

[154] T J Kalogeris, K Fukagawa, and P Tso. "Synthesis and lymphatic transport of intestinal apolipoprotein A-IV in response to graded doses of triglyceride." en. In: *Journal of Lipid Research* 35.7 (July 1994), pp. 1141–1151. ISSN: 00222275. DOI: 10.1016/S0022-2275(20)39959-4. URL: https://linkinghub.elsevier.com/retrieve/pii/S0022227520399594.

[155] C L Bisgaier et al. "Distribution of apolipoprotein A-IV in human plasma." en. In: *Journal of Lipid Research* 26.1 (Jan. 1985), pp. 11–25. ISSN: 00222275. DOI: 10.1016/S0022-2275(20)34400-X. URL: https://linkinghub.elsevier.com/retrieve/pii/S002222752034400X.

[156] Zhen Cao et al. "Construction and application of artificial lipoproteins using adiposomes". en. In: *Journal of Lipid Research* 64.10 (Oct. 2023), p. 100436. ISSN: 00222275. DOI: 10.1016/j.jlr.2023.100436. URL: https://linkinghub.elsevier.com/retrieve/pii/S0022227523001098.

[157] Derek M. Culnan et al. "Apolipoprotein A-IV, a Putative Satiety/Antiatherogenic Factor, Rises After Gastric Bypass". en. In: *Obesity* 17.1 (Jan. 2009), pp. 46–52. ISSN: 1930-7381, 1930-739X. DOI: 10.1038/oby.2008.428. URL: https://onlinelibrary.wiley.com/doi/10.1038/oby.2008.428.

[158] Marco Raffaelli et al. "Effect of Gastric Bypass Versus Diet on Cardiovascular Risk Factors". en. In: *Annals of Surgery* 259.4 (Apr. 2014), pp. 694–699. ISSN: 0003-4932. DOI: 10.1097/SLA.0b013e31829d6989. URL: https://journals.lww.com/00000658-201404000-00014.

[159] Hans-Jurgen Menzel et al. "Frequency and effect of human apolipoprotein A-IV polymorphism on lipid and lipoprotein levels in an Icelandic population". en. In: *Human Genetics* 84.4 (Mar. 1990). ISSN: 0340-6717, 1432-1203. DOI: 10.1007/BF00196231. URL: http://link.springer.com/10.1007/BF00196231.

[160] Ia Larson et al. "Effects of apolipoprotein A-IV genotype on glucose and plasma lipoprotein levels". en. In: *Clinical Genetics* 61.6 (June 2002), pp. 430–436. ISSN: 0009-9163, 1399-0004. DOI: 10.1034/j.1399-0004.2002.610606.x. URL: https://onlinelibrary.wiley.com/doi/10.1034/j.1399-0004.2002.610606.x.

[161] Fei Wang et al. "Apolipoprotein A-IV improves glucose homeostasis by enhancing insulin secretion". en. In: *Proceedings of the National Academy of Sciences* 109.24 (June 2012), pp. 9641–9646. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1201433109. URL: https://pnas.org/doi/full/10.1073/pnas.1201433109.

[162] Xiaoming Li et al. "Apolipoprotein A-IV Reduces Hepatic Gluconeogenesis through Nuclear Receptor NR1D1". en. In: *Journal of Biological Chemistry* 289.4 (Jan. 2014), pp. 2396–2404. ISSN: 00219258. DOI: 10.1074/jbc.M113.511766. URL: https://linkinghub.elsevier.com/retrieve/pii/S0021925820335924.

[163] Xiaoming Li et al. "Interaction of ApoA-IV with NR4A1 and NR1D1 Represses G6Pase and PEPCK Transcription: Nuclear Receptor-Mediated Downregulation of Hepatic Gluconeogenesis in Mice and a Human Hepatocyte Cell Line". en. In: *PLOS ONE* 10.11 (Nov. 2015). Ed. by Yanqiao Zhang, e0142098. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0142098. URL: https://dx.plos.org/10.1371/journal.pone.0142098.

[164] Xiaoming Li et al. "ApoA-IV improves insulin sensitivity and glucose uptake in mouse adipocytes via PI3K-Akt Signaling". en. In: *Scientific Reports* 7.1 (Jan. 2017), p. 41289. ISSN: 2045-2322. DOI: 10.1038/srep41289. URL: https://www.nature.com/articles/srep41289.

[165] Qing-Run Li et al. "Systems Signatures Reveal Unique Remission-path of Type 2 Diabetes Following Roux-en-Y Gastric Bypass Surgery". en. In: *EBioMedicine* 28 (Feb. 2018), pp. 234–240. ISSN: 23523964. DOI: 10.1016/j.ebiom.2018.01.018. URL: https://linkinghub.elsevier.com/retrieve/pii/S2352396418300227.

[166] R D Cohen et al. "Reduced aortic lesions and elevated high density lipoprotein levels in transgenic mice overexpressing mouse apolipoprotein A-IV." en. In: *Journal of Clinical Investigation* 99.8 (Apr. 1997), pp. 1906–1916. ISSN: 0021-9738. DOI: 10.1172/JCI119358. URL: http://www.jci.org/articles/view/119358.

[167] Natalie Fournier et al. "Human ApoA-IV Overexpression in Transgenic Mice Induces cAMP-Stimulated Cholesterol Efflux From J774 Macrophages to Whole Serum". en. In: *Arteriosclerosis, Thrombosis, and Vascular Biology* 20.5 (May 2000), pp. 1283–1292. ISSN: 1079-5642, 1524-4636. DOI: 10.1161/01.ATV.20.5.1283. URL: https://www.ahajournals.org/doi/10.1161/01.ATV.20.5.1283.

[168] Bo Ahrén. "Islet G protein-coupled receptors as potential targets for treatment of type 2 diabetes". en. In: *Nature Reviews Drug Discovery* 8.5 (May 2009), pp. 369–385. ISSN: 1474-1776, 1474-1784. DOI: 10.1038/nrd2782. URL: https://www.nature.com/articles/nrd2782.

[169] M Liu. "Expression of biologically active rat apolipoprotein AIV in Escherichia coli". en. In: *Physiology & Behavior* 78.1 (Jan. 2003), pp. 149–155. ISSN: 00319384. DOI: 10.1016/S0031-9384(02)00959-9. URL: https://linkinghub.elsevier.com/retrieve/pii/S0031938402009599.

[170] K. Fujimoto, J. A. Cardelli, and P. Tso. "Increased apolipoprotein A-IV in rat mesenteric lymph after lipid meal acts as a physiological signal for satiation". en. In: *American Journal of Physiology-Gastrointestinal and Liver Physiology* 262.6 (June 1992), G1002–G1006. ISSN: 0193-

1857, 1522-1547. DOI: 10.1152/ajpgi.1992.262.6.G1002. URL: https://www.physiology.org/doi/10.1152/ajpgi.1992.262.6.G1002.

[171] Sydney Pence et al. "Reduced Diet-induced Thermogenesis in Apolipoprotein A-IV Deficient Mice". en. In: *International Journal of Molecular Sciences* 20.13 (June 2019), p. 3176. ISSN: 1422-0067. DOI: 10.3390/ijms20133176. URL: https://www.mdpi.com/1422-0067/20/13/3176.

[172] Lee-Wei Yang and Ivet Bahar. "Coupling between Catalytic Site and Collective Dynamics: A Requirement for Mechanochemical Activity of Enzymes". en. In: *Structure* 13.6 (June 2005), pp. 893–904. ISSN: 09692126. DOI: 10.1016/j.str.2005.03.015. URL: https://linkinghub.elsevier.com/retrieve/pii/S096921260500167X.

[173] Anindita Dutta and Ivet Bahar. "Metal-Binding Sites Are Designed to Achieve Optimal Mechanical and Signaling Properties". en. In: *Structure* 18.9 (Sept. 2010), pp. 1140–1148. ISSN: 09692126. DOI: 10.1016/j.str.2010.06.013. URL: https://linkinghub.elsevier.com/retrieve/pii/S0969212610002674.

[174] Glenn S. Gerhard et al. "A Role for Fibroblast Growth Factor 19 and Bile Acids in Diabetes Remission After Roux-en-Y Gastric Bypass". en. In: *Diabetes Care* 36.7 (July 2013), pp. 1859–1864. ISSN: 0149-5992, 1935-5548. DOI: 10.2337/dc12-2255. URL: https://diabetesjournals.org/care/article/36/7/1859/33035/A-Role-for-Fibroblast-Growth-Factor-19-and-Bile.

[175] Lili Ding et al. "Vertical sleeve gastrectomy confers metabolic improvements by reducing intestinal bile acids and lipid absorption in mice". en. In: *Proceedings of the National Academy of Sciences* 118.6 (Feb. 2021), e2019388118. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.2019388118. URL: https://pnas.org/doi/full/10.1073/pnas.2019388118.

[176] Margaret A. Stefater et al. "All Bariatric Surgeries Are Not Created Equal: Insights from Mechanistic Comparisons". en. In: *Endocrine Reviews* 33.4 (Aug. 2012), pp. 595–622. ISSN: 0163-769X, 1945-7189. DOI: 10.1210/er.2011-1044. URL: https://academic.oup.com/edrv/article/33/4/595/2354851.

[177] Nicolai J. Wewer Albrechtsen et al. "Plasma Proteome Profiling Reveals Dynamics of Inflammatory and Lipid Homeostasis Markers after Roux-En-Y Gastric Bypass Surgery". en. In: *Cell Systems* 7.6 (Dec. 2018), 601–612.e3. ISSN: 24054712. DOI: 10.1016/j.cels.2018.10.012. URL: https://linkinghub.elsevier.com/retrieve/pii/S240547121830437X.

[178] Zhenguo Wang et al. "Apolipoprotein A-IV involves in glucose and lipid metabolism of rat". en. In: *Nutrition & Metabolism* 16.1 (Dec. 2019), p. 41. ISSN: 1743-7075. DOI: 10.1186/s12986-019-0367-2. URL: https://nutritionandmetabolism.biomedcentral.com/articles/10.1186/s12986-019-0367-2.

[179] A Steinmetz et al. "Human apolipoprotein A-IV binds to apolipoprotein A-I/A-II receptor sites and promotes cholesterol efflux from adipose cells." en. In: *Journal of Biological Chemistry* 265.14 (June 1990), pp. 7859–7863. ISSN: 00219258. DOI: 10.1016/S0021-9258(19)39010-6. URL: https://linkinghub.elsevier.com/retrieve/pii/S0021925819390106.

[180] Jie Qu et al. "Low-density lipoprotein receptor-related protein 1 (LRP1) is a novel receptor for apolipoprotein A4 (APOA4) in adipose tissue". en. In: *Scientific Reports* 11.1 (June 2021), p. 13289. ISSN: 2045-2322. DOI: 10.1038/s41598-021-92711-0. URL: https://www.nature.com/articles/s41598-021-92711-0.

[181] David O. Osei-Hwedieh et al. "Apolipoprotein mimetic peptides: Mechanisms of action as anti-atherogenic agents". en. In: *Pharmacology & Therapeutics* 130.1 (Apr. 2011), pp. 83–91. ISSN: 01637258. DOI: 10.1016/j.pharmthera.2010.12.003. URL: https://linkinghub.elsevier.com/retrieve/pii/S0163725810002305.

[182] C. Roger White, David W. Garber, and G.M. Anantharamaiah. "Anti-inflammatory and cholesterol-reducing properties of apolipoprotein mimetics: a review". en. In: *Journal of Lipid Research* 55.10 (Oct. 2014), pp. 2007–2021. ISSN: 00222275. DOI: 10.1194/jlr.R051367. URL: https://linkinghub.elsevier.com/retrieve/pii/S0022227520350057.

[183] C. Roger White et al. "Regulation of Pattern Recognition Receptors by the Apolipoprotein A-I Mimetic Peptide 4F". en. In: *Arteriosclerosis, Thrombosis, and Vascular Biology* 32.11 (Nov. 2012), pp. 2631–2639. ISSN: 1079-5642, 1524-4636. DOI: 10.1161/ATVBAHA.112.300167. URL: https://www.ahajournals.org/doi/10.1161/ATVBAHA.112.300167.

[184] Mohamad Navab et al. "High-Density Lipoprotein and 4F Peptide Reduce Systemic Inflammation by Modulating Intestinal Oxidized Lipid Metabolism: Novel Hypotheses and Review of Literature". en. In: *Arteriosclerosis, Thrombosis, and Vascular Biology* 32.11 (Nov. 2012), pp. 2553–2560. ISSN: 1079-5642, 1524-4636. DOI: 10.1161/ATVBAHA.112.300282. URL: https://www.ahajournals.org/doi/10.1161/ATVBAHA.112.300282.

[185] Johnny E. Croy, Theodore Brandon, and Elizabeth A. Komives. "Two Apolipoprotein E Mimetic Peptides, ApoE(130–149) and ApoE(141–155) $^{2}$ , Bind to LRP1". en. In: *Biochemistry* 43.23 (June 2004), pp. 7328–7335. ISSN: 0006-2960, 1520-4995. DOI: 10.1021/bi036208p. URL: https://pubs.acs.org/doi/10.1021/bi036208p.

[186] Eftaxia-Konstantina Valanti, Angeliki Chroni, and Despina Sanoudou. "The future of apolipoprotein E mimetic peptides in the prevention of cardiovascular disease". en. In: *Current Opinion in Lipidology* 30.4 (Aug. 2019), pp. 326–341. ISSN: 0957-9672, 1473-6535. DOI: 10.1097/MOL.0000000000000615. URL: https://journals.lww.com/00041433-201908000-00010.

[187] Sakeel Ahmed, Abhay H. Pande, and Shyam Sunder Sharma. "Therapeutic potential of ApoE-mimetic peptides in CNS disorders: Current perspective". en. In: *Experimental Neurology* 353 (July 2022), p. 114051. ISSN: 00144886. DOI: 10.1016/j.expneurol.2022.114051. URL: https://linkinghub.elsevier.com/retrieve/pii/S0014488622000760.

[188] Le Wang et al. "Mesenchymal stem cells ameliorate β cell dysfunction of human type 2 diabetic islets by reversing β cell dedifferentiation". en. In: *EBioMedicine* 51 (Jan. 2020), p. 102615. ISSN: 23523964. DOI: 10.1016/j.ebiom.2019.102615. URL: https://linkinghub.elsevier.com/retrieve/pii/S2352396419308308.

[189] M C Deeds et al. "Single dose streptozotocin-induced diabetes: considerations for study design in islet transplantation models". en. In: *Laboratory Animals* 45.3 (July 2011), pp. 131–140. ISSN: 0023-6772, 1758-1117. DOI: 10.1258/la.2010.010090. URL: https://journals.sagepub.com/doi/10.1258/la.2010.010090.

[190] Lei Lei et al. "Antidiabetic potential of a novel dual-target activator of glucokinase and peroxisome proliferator activated receptor-γ". en. In: *Metabolism* 64.10 (Oct. 2015), pp. 1250–1261. ISSN: 00260495. DOI: 10.1016/j.metabol.2015.06.014. URL: https://linkinghub.elsevier.com/retrieve/pii/S0026049515001754.

[191] Paul H. Tewson et al. "New DAG and cAMP Sensors Optimized for Live-Cell Assays in Automated Laboratories". en. In: *SLAS Discovery* 21.3 (Mar. 2016), pp. 298–305. ISSN: 24725552. DOI: 10.1177/1087057115618608. URL: https://linkinghub.elsevier.com/retrieve/pii/S2472555222070812.

[192] Lei Lei et al. "Morus alba L. (Sangzhi) Alkaloids Promote Insulin Secretion, Restore Diabetic β-Cell Function by Preventing Dedifferentiation and Apoptosis". en. In: *Frontiers in Pharmacology* 13 (Mar. 2022), p. 841981. ISSN: 1663-9812. DOI: 10.3389/fphar.2022.841981. URL: https://www.frontiersin.org/articles/10.3389/fphar.2022.841981/full.

[193] Weikang Gong et al. "Equally Weighted Multiscale Elastic Network Model and Its Comparison with Traditional and Parameter-Free Models". en. In: *Journal of Chemical Information and Modeling* 61.2 (Feb. 2021), pp. 921–937. ISSN: 1549-9596, 1549-960X. DOI: 10.1021/acs.jcim.0c01178. URL: https://pubs.acs.org/doi/10.1021/acs.jcim.0c01178.

[194] Stephen K Burley et al. "RCSB Protein Data Bank (RCSB.org): delivery of experimentally-determined PDB structures alongside one million computed structure models of proteins from artificial intelligence/machine learning". en. In: *Nucleic Acids Research* 51.D1 (Jan. 2023), pp. D488–D508. ISSN: 0305-1048, 1362-4962. DOI: 10.1093/nar/gkac1077. URL: https://academic.oup.com/nar/article/51/D1/D488/6845439.

[195] Jianyi Yang et al. "The I-TASSER Suite: protein structure and function prediction". en. In: *Nature Methods* 12.1 (Jan. 2015), pp. 7–8. ISSN: 1548-7091, 1548-7105. DOI: 10.1038/nmeth.3213. URL: https://www.nature.com/articles/nmeth.3213.

[196] Xiaodi Deng et al. “The Structure of Dimeric Apolipoprotein A-IV and Its Mechanism of Self-Association”. en. In: *Structure* 20.5 (May 2012), pp. 767–779. ISSN: 09692126. DOI: 10.1016/j.str.2012.02.020. URL: https://linkinghub.elsevier.com/retrieve/pii/S0969212612001001.

[197] Autumn M. McKnite et al. “Murine Gut Microbiota Is Defined by Host Genetics and Modulates Variation of Metabolic Traits”. en. In: *PLoS ONE* 7.6 (June 2012). Ed. by Bryan A. White, e39191. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0039191. URL: https://dx.plos.org/10.1371/journal.pone.0039191.