TUM School of Life Sciences
Technische Universität München

Identifying novel genetic determinants of Indirect Genetic Effects

Archana Tomar

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

Doktorin der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitz:

Prof. Dr. Angelika E. Schnieke

Prüfer*innen der Dissertation:

1. Prof. Dr. Dr. h.c. Martin Hrabe de Angelis
2. Prof. Dr. Eckhard Wolf

Die Dissertation wurde am 20.06.2022 bei der Technischen Universität München eingereicht und durch die TUM School of Life Sciences am 16.01.2023 angenommen.

Acknowledgement

Life is not about the destination, its about the Journey . . . . . .

The completion of my Ph.D. studies is unquestionably a milestone moment in my life. The last four years were a journey of new experiences, new knowledge, better understanding, and being independent. From the frustration when things don't work to the excitement when the problems get solved and you see results. From the lab lunchtime discussions to the time spent with colleagues and friends, it was an amazing journey. I'd want to thank everyone who shared those times with me, who supported me in whatever way they could through hardships but also enjoyed my achievements with me. I am grateful to have you all in my life!

First and foremost I am extremely grateful to my supervisor, *Dr. Raffaele Teperino* for his continuous and unconditional support during my Ph.D. study, for his patience, motivation, invaluable advice, discussions, and immense knowledge. Thank you for always being available as a mentor and as a friend. I am especially thankful to *Prof. Dr. Dr. h.c. Martin Hrabe de Angelis* for giving me this opportunity to work on this project and for scientific inputs and discussions. I would also like to thank Dr. Antonio Scialdone for the fruitful discussions on my study and for being a member of the thesis committee. Special mention to thank Prof. Dr. Eckhard Wolf for agreeing to review my thesis and Prof. Dr. Angelika E. Schnieke for taking the chair of the examining committee.

I appreciate the help and support provided by the GMC and IMPC centers without which this project would not have been possible.

Acknowledgement

I would like to thank all the members of the lab, IEG and GMC Jonatan Darr, Raffaele Gerlini, Maximilian Lassi, Fabienne Scheid, Gemma Comas-Armangué, Melisa Gómez-Velázquez, Lela Makharadze, Frederike Schäfer, Elida Schneltzer, Markus Kraiger, Anna-Lena Amend and many more. Cherished the time we spent together in the lab, and in social settings. Thanks for the never-ending food supply to the office. It is all your kind help and support that have made my study and life in Munich a wonderful time.

I'd like to thank all the friends with whom I traveled and enjoyed during these years, thus keeping balance in life. Thanks to amazing Raf and Debjani for being a family away from home, your presence made life much easier and enjoyable in Munich. Special mention to the friends who were away but always there to check on me.

Finally, I would like to express my gratitude to my family who supported and encouraged me. Especially my mom talking to you regularly makes me feel at home and thank you for your tremendous understanding and encouragement in every step of the way.

Abstract

Thought to be directly and uniquely dependent on genotypes, the ontogeny of individual phenotypes is much more complicated. Individual genetics, environmental exposures, and their interaction are the three main determinants of an individual's phenotype. This picture has been further complicated a decade ago when the Lamarckian theory of acquired inheritance has been rekindled with the discovery of epigenetic inheritance, according to which acquired phenotypes can be transmitted through fertilization and affect phenotypes across generations. These findings, together with the important degree of missing heritability in genetics highlighted by Genome-Wide Association Studies, suggest that not only acquired phenotypes, but also individual's genotypes may affect phenotypes intergenerationally through Indirect Genetic Effects. Here, I explored the genotype-phenotype association resource of the International Mouse Phenotyping Consortium (IMPC) with the aim of understanding whether Indirect Genetic Effects are detectable and how common they are in mammalian genetics, what are the underlying genetic determinants of Indirect Genetic Effects in mammals and which relevance they may have for human physiology and susceptibility to complex diseases. My results demonstrate that **Indirect Genetic Effects are common to mammalian genetics and influence intergenerational physiology across several layers spanning from metabolic to neurological and cardiovascular health**. Interestingly, functional annotation of the underlying genetic determinants indicate a tight clustering to proteins involved in protein ubiquitination and neuroactive signal transduction, despite the absence of clear genomic and topological clustering. Altogether, my results propose Indirect Genetic Effects as a new common feature of mammalian genetics, which

controls physiology across generations in a gene-dependent, genotype-independent manner; highlight new functions for known genes and gene families; and provide a hook to start looking for missing heritability in human genetics and the pathogenesis of complex diseases.

Zusammenfassung

Die Ontogenese individueller Phänotypen, von der man bisher annahm,, dass sie direkt und eindeutig von den Genotypen abhängt, hat sich alswesentlich komplexer als bisher gedacht herausgestellt. Die individuelle Genetik, Umwelteinflüsse und ihre Wechselwirkung sind die drei wichtigsten Determinanten des Phänotyps einess Individuums. Dieses Bild hat sich vor einem Jahrzehnt als die Lamarcksche Theorie der erworbenen Vererbung durch die Entdeckung der epigenetischen Vererbung neu belebt wurde, der zufolge erworbene Phänotypen durch die Befruchtung weitergegeben werden und den Phänotyp über Generationen hinweg beeinflussen können, weiter verkompliziert,. Diese Erkenntnisse, zusammen mit der Tatsache, dass Genomweite Assoziationsstudien gezeigt haben das es in der Genetik ein hohes Maß an fehlender Erblichkeit gibt, lassen vermuten, dass nicht nur erworbene Phänotypen, sondern auch individuelle Genotypen durch indirekte genetische Effekte Phänotypen über Generationen hinweg beeinflussen können. Für diese Arbeit habe ich die Genotyp-Phänotyp-Assoziation Ressource des International Mouse Phenotyping Consortium (IMPC) untersucht, um zu verstehen, ob indirekte genetische Effekte nachweisbar sind und wie häufig sie in der Säugetiere Genetik vorkommen. Zudem habe ich untersucht welche genetischen Determinanten den indirekten genetischen Effekten bei Säugetieren zugrunde liegen und welche Bedeutung diese für die menschliche Physiologie und die Anfälligkeit für komplexe Krankheiten haben könnten.

Meine Ergebnisse zeigen, **dass indirekte genetische Effekte in der Säugetiere Genetik weit verbreitet sind und die intergenerationale Physiologie auf mehreren Ebenen beeinflussen.** Dies reicht vom Stoffwechsel bis zur neurologischen und kar-

diovaskulären Gesundheit. Interessanterweise zeigt die funktionelle Annotation der zugrundeliegenden genetischen Determinanten ein enges Clustering mit Proteinen, die an der Ubiquitinierung von Proteinen und der neuroaktiven Signaltransduktion beteiligt sind, obwohl es kein eindeutiges genomisches und topologisches Clustering gibt. Insgesamt zeigen meine Ergebnisse, dass indirekte genetische Effekte ein neues gemeinsames Merkmal der Säugetiere Genetik sind, das die Physiologie über Generationen hinweg auf Gen-abhängige und Genotyp-unabhängige Weise steuert. Desweiteren heben meine Ergebnisseneue Funktionen für bekannte Gene und Genfamilien hervor und bieten einen Ausgangspunkt für die weitere Suche nach der fehlendeen Erblichkeit in der Humangenetik und der Pathogenese komplexer Krankheiten.

Publications

List of Authored Publications related to thesis

Genetic control of non-genetic inheritance in mammals: state-of-the-art and perspectives

A Tomar , R Teperino

Mamm Genome 2020, 31(5-6):146-156

Abstract

Thought to be directly and uniquely dependent from genotypes, the ontogeny of individual phenotypes is much more complicated. Individual genetics, environmental exposures, and their interaction are the three main determinants of individual's phenotype. This picture has been further complicated a decade ago when the Lamarckian theory of acquired inheritance has been rekindled with the discovery of epigenetic inheritance, according to which acquired phenotypes can be transmitted through fertilization and affect phenotypes across generations. The results of Genome-Wide Association Studies have also highlighted a big degree of missing heritability in genetics and have provided hints that not only acquired phenotypes, but also individual's genotypes affect phenotypes intergenerationally through indirect genetic effects. Here, we review available examples of indirect genetic effects in mammals, what is known of the underlying molecular mechanisms and their potential impact for our understanding of missing heritability, phenotypic variation. and individual disease risk.

List of Authored and Co-authored Publications not Discussed in this Thesis

iTAG-RNA Isolates Cell-Specific Transcriptional Responses to Environmental Stimuli and Identifies an RNA-Based Endocrine Axis

Darr J, **Tomar A**, Lassi M, Gerlini R, Berti L, Hering A, Scheid F, Hrabe de Angelis M, Witting M, Teperino R

Cell Rep 2020, 30(9):3183-3194 e3184.

Summary

Biofluids contain various circulating cell-free RNAs (ccfRNAs). The composition of these ccfRNAs varies among biofluids. They constitute tantalizing biomarker candidates for several pathologies and have been demonstrated to be mediators of cellular communication. Little is known about their function in physiological and developmental settings, and most works are limited to in vitro studies. Here, we develop iTAG-RNA, a method for the unbiased tagging of RNA transcripts in mice in vivo. We use iTAG-RNA to isolate hepatocytes and kidney proximal epithelial cell-specific transcriptional responses to a dietary challenge without interfering with the tissue architecture and to identify multiple hepatocyte-secreted ccfRNAs in plasma. We also identify specific transfer of liver-derived ccfRNAs to adipose tissue and skeletal muscle, where they likely constitute a buffering system to maintain lipid homeostasis under acute high-fat-diet feeding. Our findings directly demonstrate in vivo transfer of RNAs between tissues and highlight its implications for endocrine signaling and homeostasis.

Author Contributions

Conceptualization, Y.D.; Methodology, Y.D. and M.W.; Investigation, Y.D., M.L., R.G., A.T., F.S., L.B., A.H., and M.W.; Data Analysis, J.D., A.T., M.W., and R.T.; Writing – Original Draft, Y.D., M.W., and R.T.; Writing – Review & Editing, Y.D., M.H.d.A., M.W., and R.T.; Funding Acquisition, M.H.d.A. and R.T.; Resources, M.H.d.A., M.W., and R.T.; Supervision, R.T.

Disruption of paternal circadian rhythm affects metabolic health in male offspring via nongerm cell factors

Lassi M, **Tomar A**, Comas-Armangue G, Vogtmann R, Dijkstra DJ, Corujo D, Gerlini

R, Darr J, Scheid F, Rozman J, Aguilar-Pimentel A, Koren O, Buschbeck M, Fuchs H, Marschall S, Gailus-Durner V, Martin Hrabe de Angelis, Plösch T, Gellhaus A, Teperino R

Sci Adv 2021, 7(22).

Abstract

Circadian rhythm synchronizes each body function with the environment and regulates physiology. Disruption of normal circadian rhythm alters organismal physiology and increases disease risk. Recent epidemiological data and studies in model organisms have shown that maternal circadian disruption is important for offspring health and adult phenotypes. Less is known about the role of paternal circadian rhythm for offspring health. Here, we disrupted circadian rhythm in male mice by night-restricted feeding and showed that paternal circadian disruption at conception is important for offspring feeding behavior, metabolic health, and oscillatory transcription. Mechanistically, our data suggest that the effect of paternal circadian disruption is not transferred to the offspring via the germ cells but initiated by corticosterone-based parental communication at conception and programmed during in utero development through a state of fetal growth restriction. These findings indicate paternal circadian health at conception as a newly identified determinant of offspring phenotypes.

Author contributions:

R.T. conceptualized the project, supervised the experiments, analyzed and interpreted the results, wrote the manuscript, and acquired funds. M.L., G.C.-A., A.T., R.V., D.J.D., D.C., R.G., J.D., A.A.-P., and F.S. performed the experiments and analyzed the results. J.R., O.K., M.B., H.F., S.M., V.G.-D., M.H.d.A., T.P., and A.G. supervised experiments, acquired funds, and reviewed and edited the manuscript.

Uncovering the molecular identity of cardiosphere-derived cells (CDCs) by single-cell RNA sequencing

Kogan PS, Wirth F, **Tomar A**, Darr J, Teperino R, Lahm H, Dressen M, Puluca N, Zhang Z, Neb I, Beck N, Luzius T, Rosa LO, Gärtner K, Hüls C, Zeidler R, Ramanujam D, Engelhardt S, Wenk C, Holdt, LM, Mononen M, Sahara M, Cleuziou J, Hörer J,

Lange R, Krane M and Doppler SA

Basic Res Cardiol 2022, 117(1):11.

Abstract

Cardiosphere-derived cells (CDCs) generated from human cardiac biopsies have been shown to have disease-modifying bioactivity in clinical trials. Paradoxically, CDCs' cellular origin in the heart remains elusive. We studied the molecular identity of CDCs using single-cell RNA sequencing (sc-RNAseq) in comparison to cardiac non-myocyte and non-hematopoietic cells (cardiac fibroblasts/CFs, smooth muscle cells/SMCs and endothelial cells/ECs). We identified CDCs as a distinct and mitochondria-rich cell type that shared biological similarities with non-myocyte cells but not with cardiac progenitor cells derived from human-induced pluripotent stem cells. CXCL6 emerged as a new specific marker for CDCs. By analysis of sc-RNAseq data from human right atrial biopsies in comparison with CDCs we uncovered transcriptomic similarities between CDCs and CFs. By direct comparison of infant and adult CDC sc-RNAseq data, infant CDCs revealed GO-terms associated with cardiac development. To analyze the beneficial effects of CDCs (pro-angiogenic, anti-fibrotic, anti-apoptotic), we performed functional in vitro assays with CDC-derived extracellular vesicles (EVs). CDC EVs augmented in vitro angiogenesis and did not stimulate scarring. They also reduced the expression of pro-apoptotic Bax in NRCMs. In conclusion, CDCs were disclosed as mitochondria-rich cells with unique properties but also with similarities to right atrial CFs. CDCs displayed highly proliferative, secretory and immunomodulatory properties, characteristics that can also be found in activated or inflammatory cell types. By special culture conditions, CDCs earn some bioactivities, including angiogenic potential, which might modify disease in certain disorders.

Contributions P-SK: Conception and design, collection and/or assembly of data, data analysis and interpretation, manuscript writing, final approval of the manuscript. FW: collection and/or assembly of data, data analysis and interpretation, manuscript writing, final approval of the manuscript. AT: Data analysis and interpretation, final approval of the manuscript. JD: Collection and/or assembly of data, final approval of the

manuscript. RT: Data analysis and interpretation, final approval of the manuscript. HL: Collection and/or assembly of data, provision of study material or patients, final approval of the manuscript. MD: Collection and/or assembly of data, provision of study material or patients, final approval of the manuscript. NP: Collection and/or assembly of data, final approval of the manuscript. ZZ: Collection and/or assembly of data, final approval of the manuscript. IN: Collection and/or assembly of data, final approval of the manuscript. NB: Collection and/or assembly of data, provision of study material or patients, final approval of the manuscript. TL: Collection and/or assembly of data, final approval of the manuscript. LOR: Collection and/or assembly of data, final approval of the manuscript. KG: Collection and/or assembly of data, final approval of the manuscript. CH: Collection and/or assembly of data, final approval of the manuscript. RZ: Collection and/or assembly of data, final approval of the manuscript. DR: Collection and/or assembly of data, final approval of the manuscript. SE: Collection and/or assembly of data, final approval of the manuscript. CW: Collection and/or assembly of data, final approval of the manuscript. LMH: Collection and/or assembly of data, final approval of the manuscript. MM: Collection and/or assembly of data, final approval of the manuscript. MS: Collection and/or assembly of data, final approval of the manuscript. JC: provision of study material or patients, final approval of the manuscript. JH: provision of study material or patients, final approval of the manuscript. RL: Provision of study material or patients, administrative support, financial support, final approval of the manuscript. MK: Conception and design, provision of study material or patients, administrative support, financial support, final approval of the manuscript. SAD: Conception and design, data analysis and interpretation, manuscript writing, final approval of the manuscript.

Contents

Acknowledgement iii

Abstract v

Zusammenfassung vii

Publications ix

Contents xv

1 Introduction 1

1.1 Epigenetics . . . 3

1.2 Epigenetic inheritance . . . 3

1.3 Intergenerational and transgenerational epigenetic inheritance in animals . 5

1.4 General mechanisms of epigenetic inheritance . . . 6

1.4.1 Methylation . . . 6

1.4.2 Histone modifications . . . 9

1.4.3 Small RNA . . . 10

1.4.4 Prions . . . 11

1.5 Genome-Wide Association Studies and the missing heritability . . . 12

1.6 Epigenetic regulation of genomic imprinting . . . 14

1.7 Paramutation . . . 16

1.8 Sperm contribution to epigenetic inheritance . . . 19

1.9 Transgenerational epidemiological human studies . . . 22

1.10 Indirect Genetic Effects 26
1.11 Relevance of large scale IMPC data 28
1.12 Aim and scope of this thesis 31

2 Methodology **33**

2.1 The International Mouse Phenotyping Consortium (IMPC) 33
2.2 IMPC Data Collection 33
2.3 Data verification, filtration, and quality check 35
2.4 Principal Component Analysis (PCA) and correlation Analysis 37
2.5 Strategy for candidate genes selection 37
2.6 Gene Annotations and pathway analysis 40
2.7 Publicly available data 40
2.7.1 Genomic Location 40
2.7.2 Topologically associating domains (TADs) 41
2.7.3 Genome-wide association studies (GWAS) 41
2.7.4 mouse embryonic stem cells (mESCs) ChIP-seq and ATAC-seq data analysis 42
2.7.5 Chromatin segmentation analysis 42
2.7.6 CpG islands Identification 42
2.7.7 Human diseases 44
2.7.8 Phenotype based gene interaction Network 44

3 Results **47**

3.1 IMPC phenotyping data reveals center-based clustering 47
3.2 Wild-type offspring shows phenotypic variation 49
3.3 Functional Annotation of the identified 571 candidate genes 51
3.4 Inferring information about IGEs by exploring publicly available datasets 53
3.5 Human Disease Association 57
3.6 Phenotype gene association network 58

4 Discussion **61**

4.1 IMPC as a valuable resource to study Indirect Genetic Effects 62

4.2 Potential candidates of indirect genetic effects from IMPC resource 62

4.3 Quest for common features in IGE-inducing genes 63

5 Conclusion & Outlook **65**

6 Supplementary Figures **67**

List of Figures **87**

List of Tables **97**

Bibliography **99**

1 Introduction

The basic principle for the continuity of life is the faithful transmission of genotype, and hence phenotype, to the next generation. It is now known that not only the genome but also the epigenome plays a significant role in phenotype transmission. While the transmission of parental factors which involves DNA sequence is known as genetic inheritance, other forms of non-DNA based inheritance are collectively known as **Non-Genetic Inheritance**. At the core of non-genetic inheritance are epigenetic mechanisms, which modify gene function without changing the DNA sequence and most importantly are heritable through mitosis. Typically, Non-Genetic Inheritance is associated with parental experience and environmental exposure, but relatively recent findings have indicated that parental genetics can influence phenotypic trajectories also in non-carrier offspring in a non-genetic manner. These phenomena can therefore no longer be explained by the traditional Mendelian laws of inheritance, by which each inherited trait is defined by a pair of genes.

The conventional concept of inheritance clearly rejects the possibility that environmental effects and experiences can be transferred to offspring. Studies have revealed that parental history and experience also have epigenomic impacts, including sperm and oocyte cytosine methylation and chromatin patterning, noncoding RNAs and mitochondria. These epigenetic mechanisms broadly deviate developmental trajectories, thereby affecting adult phenotyping. Paternal paradigms can broadly be categorized into three groups in mammals: dietary interventions, stress conditions, and exposures to toxins [1, 2]. Dietary interventions focus mainly on high-fat diets, low-protein diets, and caloric constraints that primarily affect metabolic parameters such as glucose control,

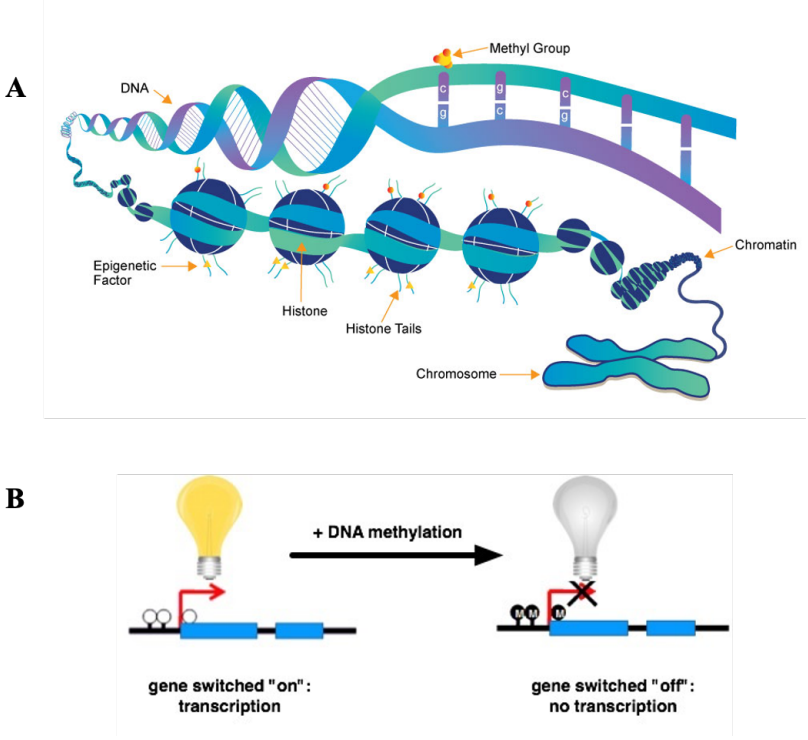


Figure 1.1: Graphical representation of Chromatin organization and epigenetic regulation. A. Chromatin is the subunit that creates a chromosome, which is made up of DNA and protein. B. DNA methylation on the promoter can lead to switching ON or OFF of the gene transcription.

and lipid metabolism in offspring [3, 4, 5, 6]. Paternal disorders including social defeat, maternal separation, and chronic unpredictable stress, have been linked to the modified release of cortisol, metabolism, and blood-brain barrier function in the next generation [7, 8, 9]. Finally, toxins and bioactive compounds used in paternal-effect studies range from endocrine disruptors (vinclozolin, BPA, etc.), to carbon tetrachloride, to drugs of abuse including nicotine and cocaine [10, 11, 12, 13]. It has been shown that environmental toxicants such as DDT cause epigenetic transgenerational disease inheritance (e.g. obesity) via the germline [14].

Currently, there is a big debate in the field on whether epigenetics modifications can be transgenerationally inherited and passed down to truly unexposed generations, such as the third generation (F3). F3 is focused on as it is not subjected to any environmental

influence of grandparents (F0), as the embryo (F1) and the germ cells of the embryo (F2) can be exposed. As the research in the field of epigenetic inheritance is expanding, so is our knowledge of the field.

1.1 Epigenetics

The word "Epi" in **epigenetics** is derived from Greek and it means on, above, at, upon or to. Epigenetic modifications are alterations on the DNA that control whether genes are switched on or off. The DNA in our cells is not in its purest form, several small chemical groups can be found attached to it in specific regions. It is found wrapped around special protein molecules called histones. These histones further pack and order the DNA into nucleosomes. When the chemical groups are attached or removed from the DNA or the associated proteins, it alters the expression of nearby genes which also changes the cell function (Figure. 1.1). Sometimes, if these chemical modifications are placed or taken off during critical stages of development, the patterns can stay for the rest of our lives and can even passed on to the next generations.

Epigenetics is defined as the mechanism that controls gene activity and phenotype ontogeny without altering the DNA sequence. Conrad H. Waddington introduced the term "epigenetic" in 1942 to describe the intricate, dynamic interactions between the developing environment and the genome that resulted in phenotypic formation. The field of epigenetics is rapidly expanding, as is our awareness that the environment and individual lifestyle choices may both directly interact with the DNA to drive epigenetic modification.

1.2 Epigenetic inheritance

The transmission of these epigenetic marks to offspring is referred to as epigenetic inheritance (Figure. 1.2). Epigenetic marks when transferred from one generation to the next are known as an intergenerational epigenetic inheritance; however, the transmis-

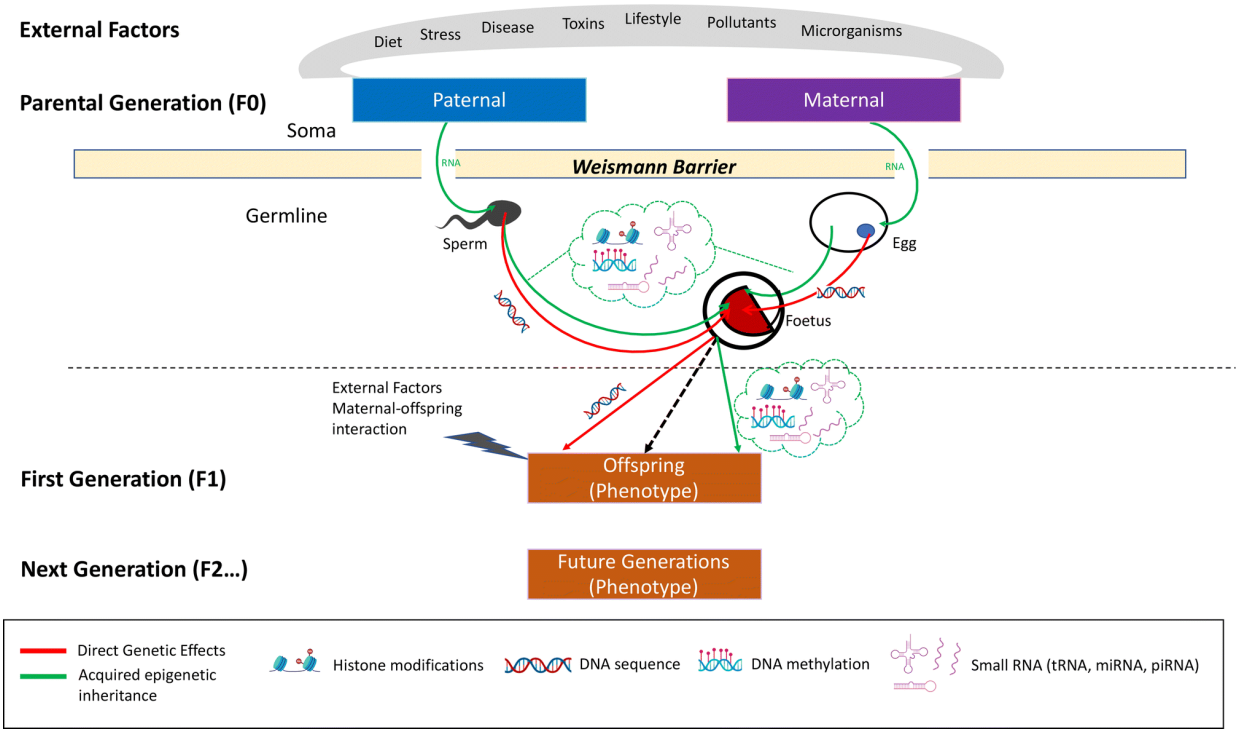


Figure 1.2: Systematic representation of sequence of events that characterize acquired epigenetic inheritance. Phenotypes acquired by the parental generation as a result of exposure to various environmental stressors are detected by the soma and produce epimutations in the germ cells, which are passed onto the offspring via fertilization and dictate their developmental and phenotypic trajectories[15].

sion of information from grandparents to grandchildren is known as "transgenerational" epigenetic inheritance .

Research in the past decade has shown that in addition to transmitting their genetic material at fertilization, parents also transfer a molecular memory of previous environmental events, including nutritional status, to their offspring via epigenetic processes. In a study conducted in 1990s, it was hypothesized that breast cancer originates in utero. Since then, numerous studies have been conducted in human cohorts and animal models that support the hypothesis of epigenetic inheritance. Family history is now known to be an important risk factor for breast cancer and several other diseases. Mutations in genes like BRCA1 and BRCA2 do account for breast cancer, but that's only in a small

percentage of cases [16]. Large-scale studies, such as The Cancer Genome Atlas TCGA, have shown that epigenetic components are frequently mutated in cancer.

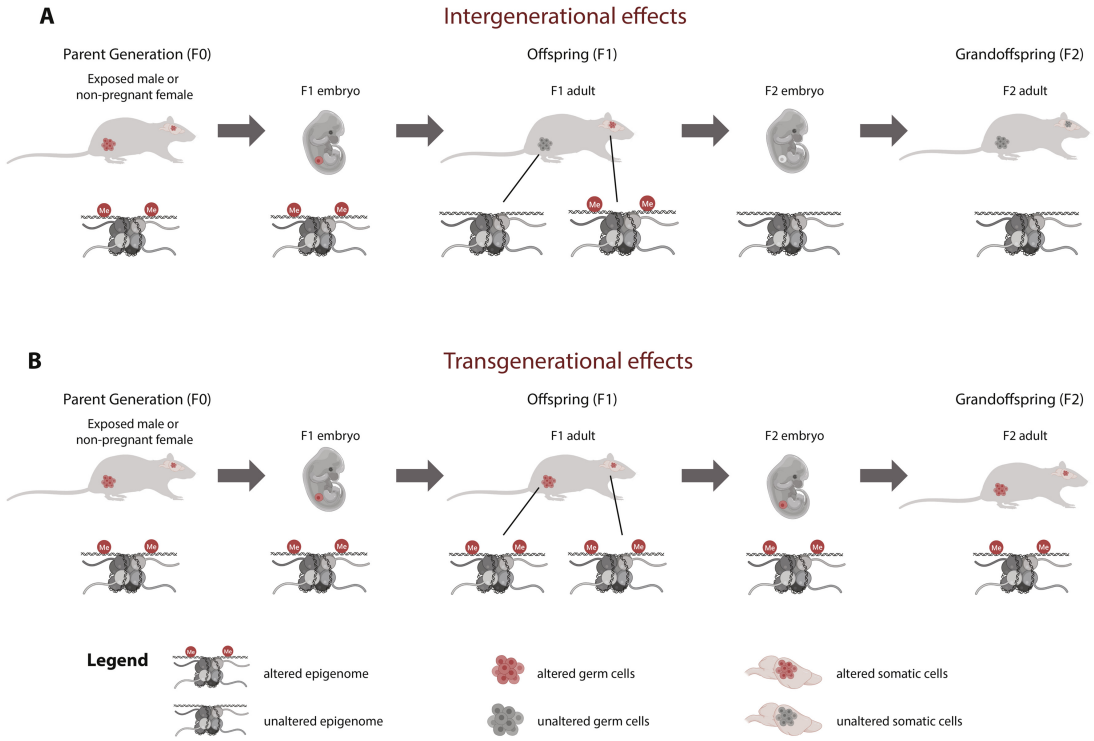


Figure 1.3: Diagrammatic representation of Intergenerational vs. Transgenerational epigenetic Inheritance

1.3 Intergenerational and transgenerational epigenetic inheritance in animals

Intergenerational epigenetic inheritance occurs when the parent generation (F0, male or non-pregnant female) is exposed to an environmental factor or external stimuli (e.g., drugs of abuse, high fat diet, stress) that causes an epigenetic change in the parent and parental germline cells (F1; e.g., sperm or oocytes). The epigenetic changes can be seen in F1 adult somatic tissues, but they do not survive in the F2 generation. To be termed transgenerational, epigenetic effects acquired in the F0 generation must be demonstrated

in the F1 and F2 generations, and maybe beyond . The pregnant dam (F0), fetus (F1), and fetal germline cells (F2) are all exposed to the external stimulation in pregnant females who undergo environmental-induced epigenetic alterations. In this example, intergenerational effects include transmission from F0 to F2, but transgenerational effects include just persistence to the F3 generation (Figure. 1.3) [17].

Numerous examples of intergenerational and transgenerational impacts in animals have been reported, for example, using model organisms like Caenorhabditis elegans, which reproduce rapidly and allow for straightforward management of genomic variation. In most cases the mechanism of inheritance is not fully understood. But with advancement in research and technologies and carefully controlled experiments a strong opinion for the involvement of epigenetics in the heritance is evident.

1.4 General mechanisms of epigenetic inheritance

In modern biology, epigenetic inheritance is a fast-emerging field. Most studies are focused on general mechanisms of epigenetic inheritance like DNA methylation/demethylation, histone modifications, competition of transcription factors, RNA-mediated post-transcriptional silencing and amyloid prionization (Figure. 1.4) . Basically, any regulatory mechanism that engages in gene expression or gene-product regulation, under certain settings may contribute to epigenetic inheritance [19].

1.4.1 Methylation

DNA methylation is an epigenetic process in which a methyl group is transferred to the C5 position of the cytosine (5mC). It plays an important role in the development, differentiation, and maintenance of cellular function. In vertebrates, DNA methylation is mostly limited to CpG sites, however non-CpG methylation has also been described. It is not fully understood to what extent DNA methylation contributes to epigenetic inheritance and in what biological contexts. Recent genome-wide studies suggest that the genome undergoes two phases of global demethylation/remethylation, first in the

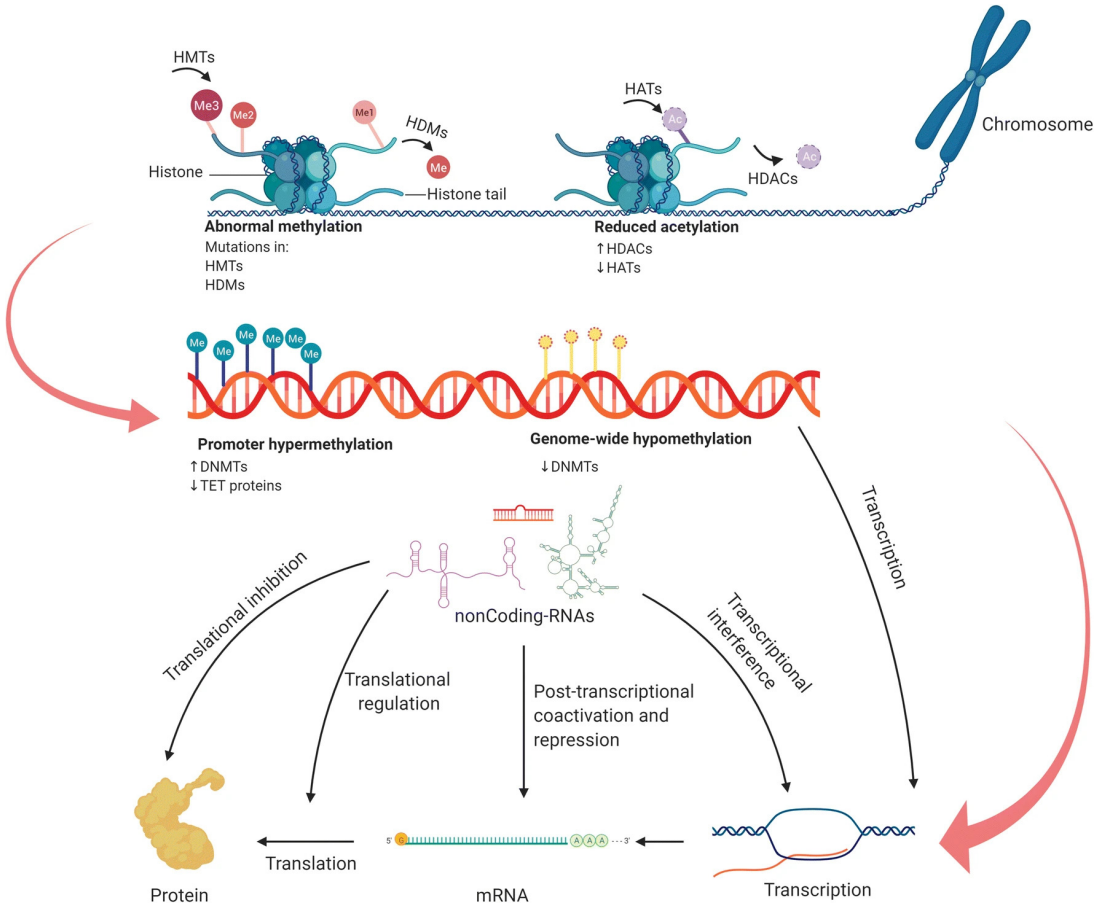


Figure 1.4: Schematic representation of three general mechanisms of epigenetic inheritance. Taken from [18]

germline, where methylation marks are globally erased in the primordial germ cells and end up with the establishment of sex-specific methylation patterns during later stages of germ cell development (Figure. 1.5). The second reprogramming occurs after fertilization, where most methylation marks inherited from the gametes are erased (with the exception of imprinted regions, see below) and the subsequent organization of the embryonic methylation landscape occurs [20].

In mammals, the mutant agouti viable yellow (A^{vy}) mouse model is extensively studied and has provided evidence of DNA methylation associated with transgenerational effects of various genetic and environmental perturbations, where methylation state in-

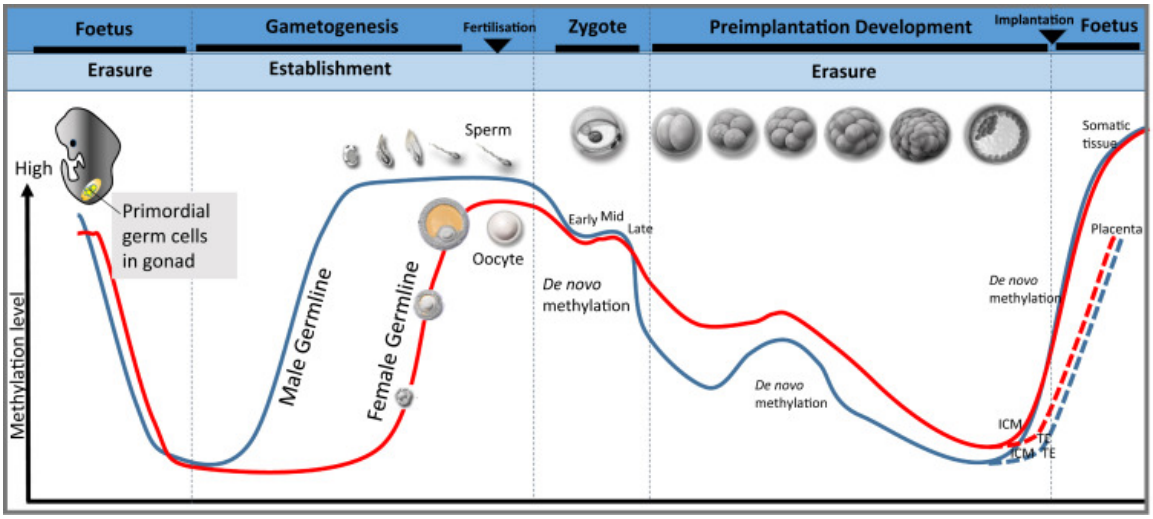


Figure 1.5: Epigenetic reprogramming in early mammalian development During preimplantation development, epigenetic reprogramming erases gametic epigenetic patterns, allowing the embryo to generate an epigenetic profile appropriate for early development and crucial for the developing conceptus [21].

versely correlates with transcriptional activity. Studies have shown that DNA methylation changes associated with parental exposure to toxins, maternal care, diet, etc can be transmitted to the grandchildren [22]. In mice, epigenetic profiling of offspring livers coming from males fed on a low-protein diet showed ∼20% changes in cytosine methylation patterns, increased expression of cholesterol biosynthesis genes, and decreased cholesterol esters. The phenotype was partially linked to altered methylation at a putative enhancer of Ppara [4]. Data on DNA methylation have evolved into a crucial source of information for biomarker development because, unlike static genetic risk estimations, DNA methylation fluctuates dynamically in reaction to different exogenous and endogenous variables, including environmental risk factors and complicated disease pathophysiology. Epigenome-wide association studies (EWAS) are a step forward in this direction to investigate the association between phenotype and epigenetic variants, mostly DNA methylation.

1.4.2 Histone modifications

Histone modifications contribute to a wider dynamic process that regulates access to DNA and transcription. Several histone modifications including methylation, acylation, phosphorylation, and ubiquitination are key players that contribute to classical epigenetic phenomena [23]. A histone mark must be somewhat stable to qualify as a vital epigenetic mark that may be transferred during mitotic division. Histone lysine methylation appears to be the most stable of the histone modifications. Among others histone lysine methylation states, H3K9 and H3K27 methylation are most likely to be heritable, as they are key regulators of epigenetic phenomena, position-effect variegation [24, 25, 26], Polycomb silencing [27, 28, 29, 30] and X inactivation [31, 32].

During cell division, the nucleosomes are evenly distributed among the two daughter chromosomes and retained close to the locus from where they are removed during replication. DNA replication results in hemimethylated CpG in daughter genomes which are acted upon by the maintenance enzyme DNA methyltransferases DNMT1 in mammals. A recent study has revealed Adenine-6N methylation as another factor for epigenetic memory. However, this mechanism doesn't work during trans-generational inheritance of the genome because all such modifications are erased during gamete formation. During spermatogenesis, the nucleosomes are removed and replaced by protamines, which are highly basic and allow super compaction of the chromosomes in the sperm head. Exceptions to this rule are zebrafish and Arabidopsis where nucleosomes are retained during the gametogenesis [33, 34]. Chromatin architecture is less distinctive in oocytes with respect to sperms as nucleosomes are retained; however, chromatin packaging is still different when compared to somatic cells. During oogenesis, oocytes chromatin is replaced by germline-specific histones such as H1oo in mammals [35]. Upon fertilization, the sperm and egg chromatin need to be decompacted which initiates extensive chromatin remodeling, and maternal and paternal genomes for several cell divisions remain distinctly bundled [36]. Protamines in sperm chromatin are replaced by replication-independent histone H3.3. In mammals, H3.3 histone is also involved in pericentric chromatin and surprisingly in the nuclear pore formation [37, 38]. These chromatin

changes guide Zygotic Genome Activation (ZGA), which characterizes the transition from maternally supplied RNAs to zygotic genome transcription. ZGA occurs as early as in the two-cell stage in mice to as late as 12 cleavage cycle in Drosophila [39].

In C. elegans, deficiencies in members of the COMPASS complex, responsible for the trimethylation of histone H3 at lysine 4, exhibit transgenerational inheritance of increased longevity upon reintroduction of functional COMPASS subunits [40]. This identifies a crucial role of histone methylation in aging and reveals communication between germline and soma for the transgenerational regulation of lifespan.

In mice, it is shown that chlordecone exposure leads to increased prostatic intraepithelial neoplasia phenotype (PIN) in both F1 and F3 generations which is mechanistically linked to a global increase in H3K4 trimethylation (H3K4me3) and a decrease in H3K27 trimethylation (H3K27me3) histone modifications [39]. It is also shown that H3K4me3 from the father is passed down to the embryo, influencing gene expression and development. It also implies that epigenetic mistakes can accumulate in sperm, resulting in poor developmental outcomes in children [41]. Chromatin may act as a mediator of molecular memory in transgenerational inheritance.

1.4.3 Small RNA

Small RNAs have long been implicated in the specification and stabilization of different chromatin marks/heterochromatin. It is now known that small RNAs including miRNA, piRNA, tRNA-derived small RNAs (tRNAs), and repeat-associated sRNAs have the ability to silence genes [42, 43, 44] and are important in the post-fertilization zygote [45, 46]. A burgeoning number of studies have shown that alterations in these sRNA have an impact on inter and/or transgenerational transmission of induced effects through the male germline [47, 48, 49]. Several studies in rodents showed altered sperm miRNAs expression associated with high-fat diet feeding and further leading to transmission of different miRNA profiles in the zygote, with repercussions on embryo development [45, 50]. When pregnant mice are exposed to toxins like vinclozolin, it leads to dysregulation of miRNAs in primordial germ cells (PGCs), which has downstream consequences on PGC differ-

entiation and lasts three generations. A recent study has interestingly shown the role of tRNA-derived fragments (tRFs), a new class of non-coding RNA, in the intergenerational inheritance of myocardial hypertrophy [51]. tRFs in sperm could function as a paternal epigenetic factor, and they could mediate the intergenerational inheritance of paternal disease [47]. These tRFs are produced by stress-released ribonuclease that cleaves mature tRNA into fragments [52]. Small RNAs are key players in RNA interference (RNAi)-related pathways, which function both in the cytoplasm of eukaryotic cells and in the nucleus where they often correlate with changes in chromatin modifications. As the involvement of sRNAs in epigenetic processes is becoming evident, we need to redefine our concepts of heredity. Breakthrough research in the field is changing the picture of the molecular pathways and mechanisms that drive epigenetic inheritance in animals.

1.4.4 Prions

Prions are an exceptional form of epigenetics. They are misfolded proteins and have the ability to infect normal forms of the same protein with their misfolded structure. Prions function outside of the basic dogma of molecular biology and are protein-based components of inheritance. They don't work by transcribing or translating genetic information; rather they adopt the final stage of central dogma – protein folding. They are the hallmarks of several deadly and transmissible neurodegenerative illnesses (such as Kuru and mad cow disease) [52] that affect humans and a variety of other species.

Its already proven that prions provide cells the ability to reorganize their metabolism [53, 54]. An increasing amount of evidence implies that prion-like conformational conversion is prevalent both within and between proteomes. Recent study reveals that protein-based epigenetic element—the [ESI+] prion can promote the transgenerational inheritance of an active chromatin state, which can confer an adaptive advantage [55]. This provides compelling evidence for the ability of this protein-based form of epigenetics to cause heritable diversity of phenotypic landscapes. Despite growing evidences of the protein-based inheritance in the past few decades, still there is lot to be explored.

1.5 Genome-Wide Association Studies and the missing heritability

In 1875 Francis Galton was the first to investigate twins to determine the relative strength of heredity and environment. To further explore heritability in human populations Genome-Wide Association Studies (GWAS) were initiated and have shown a significant step forward in understanding the heritability and association between disease-causing genetic variants and common human traits. Based on the hypothesis of "common disease, common variant" (CDCV), assuming that the same variants are responsible for the disease across the population, GWAS studies have identified more than 70,000 Single Nucleotide Polymorphisms (SNPs) associated with diseases and traits [56, 57, 58]. In fact, these studies have identified patterns for multiple sets of variants, inclusive of those associated with height, body mass index (BMI), prostate cancer risk, and many other complex traits. Altogether, these genetic variants could only explain a small percentage of the heritability and failed to detect disease risk for all variants involved. As an example, for height, by 2010 around 40 variants had been identified that collectively explained around 5% of the variation in height, compared to a twin heritability of around 80%. This gap between heritability estimates from genotype data and heritability estimates from twin data became labeled the "Missing Heritability" problem (Figure. 1.6). Several explanations were proposed to explain it, such as epigenetics, epistasis, RNAs, heritability overestimations, small size effect variants, experimental limitations, and many others. However,the complete heritability is still not fully explained [59, 60, 61].

One possible elucidation for understanding the missing heritability is that the majority of complex traits are polygenic, which means that traits are influenced by more than one gene. To address this concern, researchers proposed a risk scoring system known as Polygenic Risk Score (PRS), which is calculated by the weighted sum of risk alleles in an individual and the corresponding effect sizes obtained from the GWAS statistics summary. It enables a more accurate assessment of a person who is at risk of contracting the disease. PRS is widely used in neurodegenerative diseases/disorders, such as

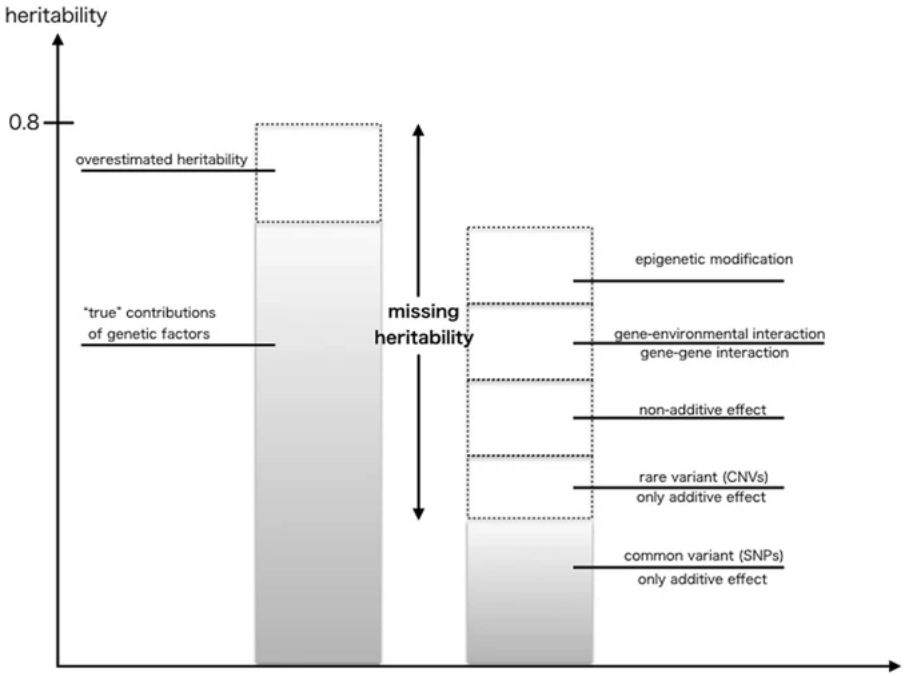


Figure 1.6: Possible mechanisms of Missing heritability. This diagram depicts the alleged mechanism of missing heritability. The proportion of each ingredient will vary between common disorders. [18].

schizophrenia, bipolar disorder, and Alzheimer and can be applied in clinical care, to identify individuals at risk and prescribe preventive measures. But studies have identified pitfalls in PRS construction that can affect its prediction efficiency in the real world, like lack of diversity in the population included in the studies, linkage disequilibrium-based pruning for construction of PRS may lead to bias due to limited reference haplotype panels for various populations. There have been recommendations for reducing the PRS bias concerning their implementation in populations with varying or mixed ancestries. Clearly, GWAS needs to include diverse populations to reduce biases and address health discrepancies [62, 63, 64]. As the cost of sequencing is reducing, Whole Genome Sequencing (WGS) is the new proposed approach to fill the gap between heritability estimates from monozygotic twin studies. The latest research suggests that much of the missing heritability is due to rare genetic variants that can be captured via WGS data [65, 66].

1.6 Epigenetic regulation of genomic imprinting

Genomic imprinting is an epigenetic process that determines whether genes are expressed or not based on whether they are inherited from the mother or the father. Genes can be partly imprinted as well. Imprinting, parent-of-origin-specific expression of one of two alleles in a diploid organism, is a classic example of intergenerational epigenetic inheritance [67, 68, 69] (Figure. 1.7). Paternally expressed genes tend to drive increased provisioning to offspring, whereas maternally expressed genes prevent excessive investment in any one child and it is this tussle that is supposed to be the driving force behind the evolution of imprinting [70, 71]. These imprinted genes occur in clusters surrounding a differentially methylated region (DMR), whose methylation status controls local gene expression. The majority of imprinting control regions are methylated on the maternal allele and unmethylated on the paternal allele— the reason being the near-complete demethylation of the paternal genome following fertilization. The regulation is different in plants where imprinting relies on allele-specific differences in the repressive H3K27 methylation mark [72, 73]. Interestingly, H3K27 methylation has also been implicated in certain cases of imprinting in mammals but are oocyte derived [74].

Igf2 is one of the best-studied paternally expressed imprinting genes that positively regulate fetal growth. Abnormal biallelic expression of Igf2 leads to embryonic overgrowth, whereas its decrease causes growth restriction. Grb10 is another maternally expressed imprinted gene that shows developmental effects (Figure. 1.8). Initially, imprinted genes were thought to be only important for prenatal development but over the years studies in mammals have shown their role in the regulation of metabolism, neurological, and physiological adaptations [75].

It appears that imprinted genes are not uniformly distributed over the whole genome, but were found clustered around a single imprinting control region (ICR) in certain genomic areas. These imprinting control regions [77], exhibit distinct epigenetic states in male and female gametes and are epigenetically reset each generation during germline development. Approximately 120 imprinted genes have been found and verified in humans and mice by 2014 [78]. But with a recent discovery of 71 new imprinted genes

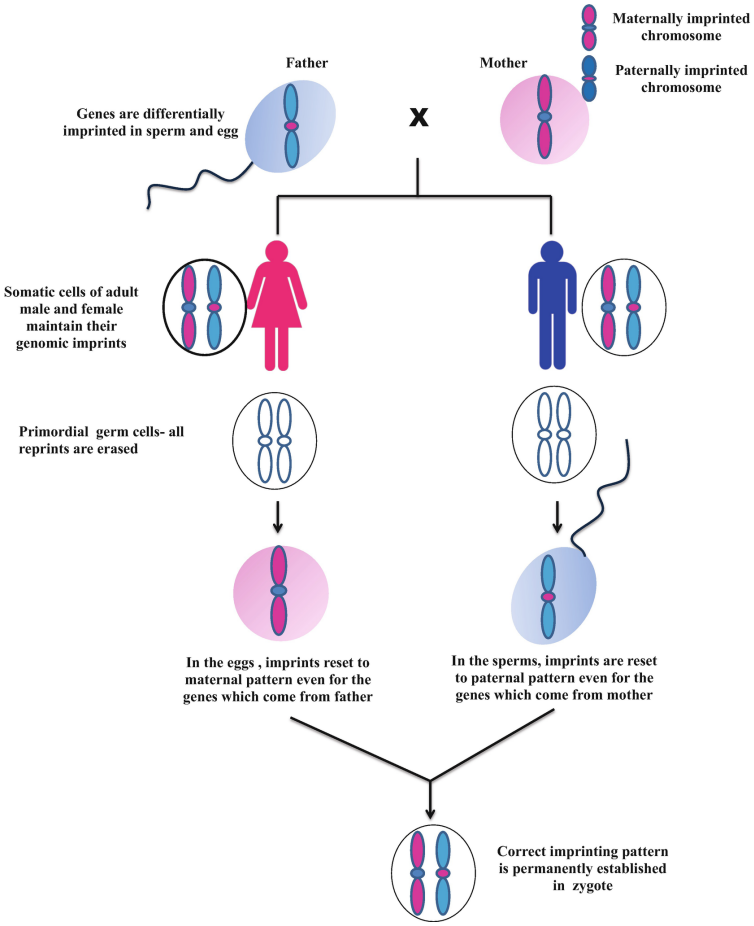


Figure 1.7: The image shows how imprinted genes are inherited across the generations (From Book Encyclopedia of Animal Cognition and Behavior).

in the mouse genome by the Biologists at the University of Bath and the University of Vienna, this number is about 200 now [79]. Another highlight from this study is that switching on and off imprinted genes is not always related to DNA methylation, but many of the newly discovered genes seem to be connected with histone 3 lysine 27 (H3K27me3) alterations, with only a few being associated with DNA methylation.

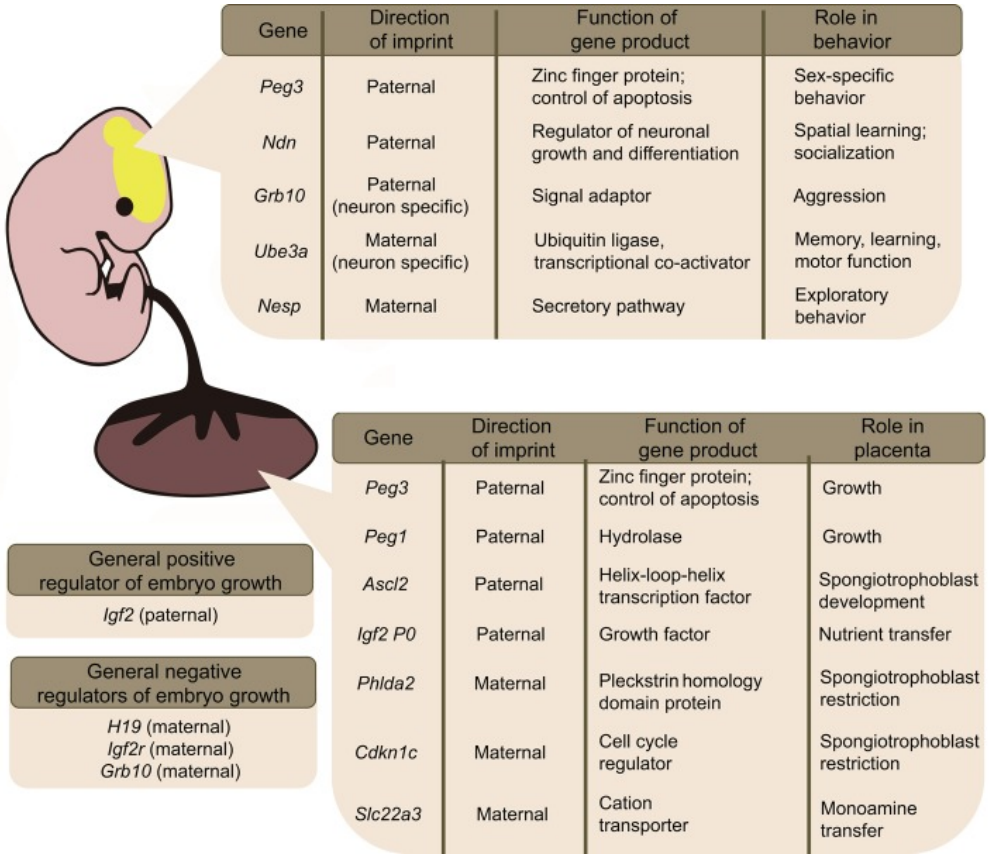


Figure 1.8: A summary of imprinted genes and their functions in the brain and placenta during embryogenesis. [76]

1.7 Paramutation

Paramutation is another example of genetically controlled heritable epigenetic variation which defies Mendel's first law that states that alleles are transmitted unchanged. This was first reported by Brink and Coe and they were baffled to find it at odds with Mendelian rules and, perhaps, for this reason, paramutation remained mysterious for decades to come. The basic tenet of paramutation is trans-homologues interactions between alleles namely 'paramutagenic' and 'paramutable'. The "paramutagenic" allele in heterozygotes transmits the phenotype to the wild-type allele ("paramutant") in a manner that is maintained through multiple generations. A universal hallmark of paramutation is that para mutable alleles become paramutagenic following exposure to

another paramutagenic allele in trans [80, 81, 82, 83, 84, 85]. In other words, these alleles are meta-stable. Thus, paramutation can be identified as the heritable silencing of one allele by the other. This crosstalk between the two alleles is mediated by short RNAs which act in trans and establish a transcriptionally silent chromatin state which is meiotically heritable through several generations viz. in the worm, C. elegans, bacterial avoidance behavior can be epigenetically maintained through multiple generations [86]. Epigenetic states assigned by paramutagenic alleles are occasionally permanent [87] and found in all future generations, and some are reversed after a few generations [83, 84, 88, 89, 90, 91, 92], as they show less than 100% heritability [93].

Another feature of paramutation is that it can be initiated in a parent of origin fashion. e.g. in mice, the transmission of the Rasgrf1$^{\mathrm{tm3.1Pds}}$ allele from father to daughter modified the daughter's Rasgrf1$^{+\mathrm{d}}$ wild-type allele in a manner that allowed it to affect the expression of both parental alleles in the grandchildren [94]. This example shows that paternal transmission of the mutated allele also induced methylation and expression in trans of the normally unmethylated and silent wild-type maternal allele.

Further, paramutation can be induced in an artificial manner like by inserting a transgene in one allele. For example, in mice a paramutation model was created by inserting a lacZ reporter gene in the first intron of Kit-oncogene which resulted in silencing of the wild type allele in subsequent three generations (Figure. 1.9) [95, 96, 97]. The same was also observed in the Igf2r locus. However, it remains to be discovered whether initiation of paramutation by trans-gene insertion, involved physical interactions between homologous alleles, the sequences in region 2 implicated in de novo methylation and allele discrimination, from the endogenous Igf2r locus, or the extensive inverted repeat structure in region 2 [94]. This suggested that despite the artificial means of initiating the trans allelic interactions, such interactions are normal genomic events.

The mechanism of paramutation in different organisms differs extensively. However, despite these fundamental differences in the gene regulatory mechanisms, the role of small RNA (sRNA) molecules is emerging as a unifying concept. These small RNAs result from misprocessing of different RNA molecules and the role of RNA-dependent

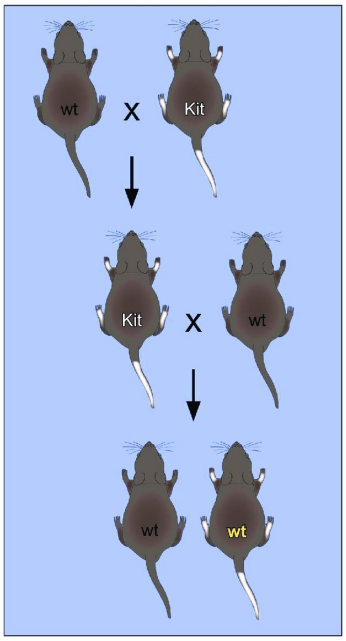


Figure 1.9: Kit paramutation in mice Male heterozygous mice with one Kit^{tm1Alf} allele and one normal wild-type (+) allele exhibit a speckled white tail tip and generate aberrant Kit mRNAs from the paramutagenic Kit^{tm1Alf} allele. These are packed in sperm and transferred to the embryo once the egg is fertilized. Progeny have the heterozygous father's wild-type Kit allele, but action of the transmitted aberrant RNAs still gives rise to the spotted tail and perpetuates its production, allowing paramutation of the wild-type Kit allele and continued transmission of the spotted tail.

RNA polymerase (RdRp) has also been shown [98, 99, 100]. While non-coding RNAs were reported to control epigenetic states in plants, as in other organisms, including in the Drosophila germline [101, 102], transgenerational determination of an epigenetic state by gametic RNA is, so far, unique to the mouse paramutation [95, 103].

These sRNAs are involved in directing RNA processing and/or chromatin modifying proteins to nascent RNA transcripts or DNA in a sequence specific manner [103, 104, 105, 106]. Self-reinforcing feedback loops that involve chromatin modifications and sRNA biogenesis that is mediated by Argonaute proteins [107] provide a persistent and heritable source of regulatory information [105], and the germline transmission. Recent studies

highlight paramutation as a byproduct of constitutive mechanisms that use small RNAs (sRNAs) to regulate epigenetic states even though their interaction with chromatin is still to be worked out.

Our knowledge and understanding of paramutations-like events are expanding and will be significant to understand the concept of genetic control of the non-genetic inheritance.

1.8 Sperm contribution to epigenetic inheritance

The gamete (sperm and egg) is one of the most critical developing cell types in any living process. The process of gametogenesis is largely responsible for the transfer of phenotypic and ideally suited physiology to following generations. How this information is transferred is largely unknown. For this reason, the area of sperm epigenetics has experienced a huge surge in attention and advancement during the last decade (Figure. 1.10). Several studies have shown that males contribute little more than the sperm upon mating, thus making mechanistic dissection of paternal effects relatively straightforward. During fertilization, sperm DNA methylation and histone profiles, nucleo-protamine distribution pattern, and small non-coding RNA composition create a unique epigenetic backdrop that could be transferred to the egg along with its haploid genome and can influence offspring health. Growing research in the field suggests that paternal environmental exposures, diet, and lifestyle can alter the sperm epigenome, affecting the embryonic development and health of the future generations and even purified gametes, have highlighted the role of sperm epigenome in reprogramming offspring. Males can even influence their offspring phenotype via non-germline mechanisms including seminal fluid, cryptic maternal effects, transfer of the microbiome, etc. It is known that towards the end of gametogenesis, 90% of sperm CpG are methylated, whereas oocytes have only 40% CpG methylated [108]. Transposons and intergenic regions have the most DNA alterations in sperms, while gene bodies and CGI are very sparsely methylated [109]. The methylation pattern of sperm DNA was found to be distinct from that of somatic cells but comparable to that of Embryonic Stem cells (ESC). The promoters of transcription and signaling factors of genes like Hox, Fox, Sox, or GATA family, which are involved in

early development shows similar methylation pattern both in sperm and ESC [110]. Epigenetic alterations—including changes to sperm chromatin, cytosine methylation, and small RNA payload—have been reported to occur in response to paternal exposures. In fact, microinjection of either purified sperm RNAs payload or synthetic RNA mixtures has been used in certain paradigms to partially reproduce paternally induced phenotypes in offspring [46]. However, the robust intergenerational effects of the paternal environment on the F1 generation get diluted in F2 and F3 generations.

Several research groups working on mouse models to study intergenerational effects have demonstrated that feeding male mice with different diets, like high-fat diet (HFD), low protein diet, and folate-deficient diet might result in metabolic abnormalities in the progeny, which were directly connected to alterations in sperm epigenome. A study by Skinner and colleagues has demonstrated that a mixture of plastic derived endocrine disruptor compounds bisphenol-A (BPA), bis(2-ethylhexyl)phthalate (DEHP) and dibutyl phthalate (DBP) all cause a significant increase in the occurrences of diseases/abnormalities (Pubertal abnormalities, testis disease, obesity, and ovarian disease) in the F3 generation [111]. The underlying mechanism remains unclear; however, altered sperm DNA methylation, non-coding RNA, and histone retention are found to be associated with transgenerational transmission of phenotype. It was interestingly observed that direct or ancestral exposures induce distinct reprogramming of the sperm epigenome [14].

Sperm epigenome is found to be altered by lifestyle factors like exercise before conception and could hold responsible for altered metabolism in the future offspring. It was shown that 6 weeks of endurance training altered the small RNA payload (with particular effect on the piRNA fraction). Exercise training was even found to influence DNA methylation of brain-related genes [112], whereas in humans even 3 months of exercise training can alter the sperm DNA methylations, many methylation changes occurred in genes related to brain diseases [113]. In humans, it appears that the offspring of fat males are more likely to develop obesity. An interesting preliminary study conducted in the Danish population (Obese = 10, Lean = 13, after surgery = 6) shows the

first epigenetic mapping of sperm in obese men and identifies the small RNA and DNA methylation changes. They showed that following bariatric surgery, sperm DNA methylation patterns shift, demonstrating that the epigenetic landscape of human sperm is dynamic and susceptible to environmental changes. These alterations were mostly seen in genes related to appetite regulation and Piwi-interacting RNAs [114]. To add to the complication, a recent study that looked at the miRNA profile of spermatozoa at the single-cell level found that spermatozoa from the same individual had varying miRNA profiles [115].

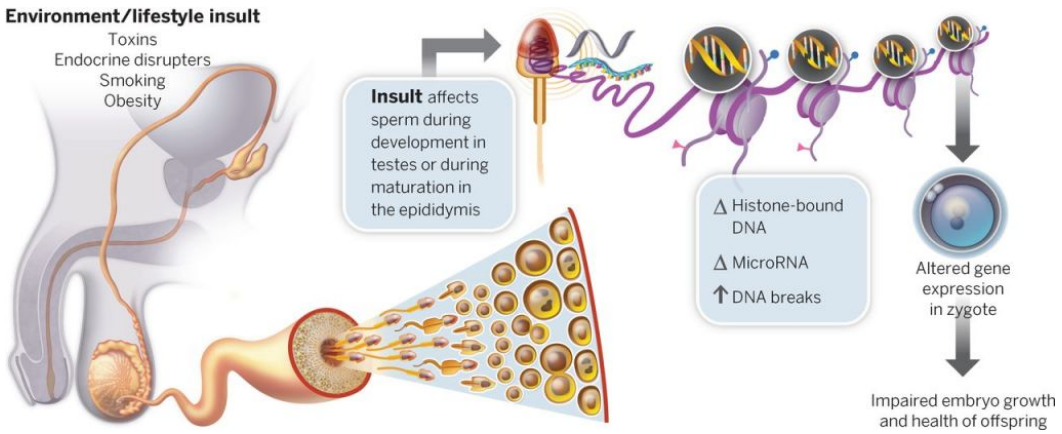


Figure 1.10: Diagram depicting environmental changes and the effects on paternally inherited non-genetic contributions (http://studentblogs.med.ed.ac.uk/reproductive-systems-group-2/mechanisms-of-epigenetics/)

In 2021 an interesting study about the effects of spaceflight on mice was published by Japanese scientists. After 35 days of spaceflight, male mice experience alterations in the binding of transcription factor ATF7, a regulator of heterochromatin formation, to promoter regions in the testis, as well as changes in small RNA production in spermatozoa. The offspring of these mice show a rise in hepatic expression of genes involved in the DNA replication [116].

Many studies across decades of research support the idea that the small RNA profile and methylation pattern of sperm is easily altered by any lifestyle change, diet,

stress, chemicals, etc, especially during embryonic development. During embryonic development, sperm RNAs may trigger a transcriptional cascade of actions that induces a paternally acquired phenotype in children. The mechanism by which the early impacts of sperm RNAs are changed to a permanent form of information to allow transgenerational inheritance remains a mystery, although it may include a complex interaction of transposable elements, DNA methylation, and chromatin structure. Future research is needed to provide insight on the impact of this post-testicular transmitted sncRNA, particularly how they may influence embryo development, thereby contributing to another level of paternal inheritance. Until we know more, would-be parents should just strive to be as healthy as possible at the time of conception, rather than being tempted to fad diets or other activities in order to impact the health of their children in ways we don't fully understand.

1.9 Transgenerational epidemiological human studies

If we look at human studies there are very few that suggest the existence of sex-specific transgenerational inheritance. The Överkalix cohorts in northern Sweden is one such historical dataset [117, 118, 119, 120]. Although difficult to interpret, the data suggest that excess consumption of food by paternal grandfather in pre-puberty can have adverse effects on the later generations of males and not females. The sons and grandsons had a high risk of death through cardiovascular diseases or diabetes, in response to an environmental challenge they themselves had never experienced. The results are particularly impressive when one considers that dietary interventions happened when the boys were pre-pubescent and so not even started to produce sperm. Still, they were able to pass an effect on to the future male generations. To test the hypothesis based on Överkalix cohorts another human cohort study was conducted on a 40 times larger dataset, in the Uppsala Multigeneration Study. Three generations were traced and studied and found support for the major Överkalix observations, the food availability of a paternal grandpa during pre-puberty predicts the all-cause mortality of his male grandchildren but not of his female grandchildren. Cancer mortality adds significantly to this pattern

in the study but it didn't replicate earlier diabetic and cardiovascular death outcomes [121]. Another fascinating study was conducted in Iceland to understand the influence of diet on disease. The disease in concern is Hereditary cystatin C amyloid angiopathy (HCCAA), a rare autosomal dominant genetic disease, which causes premature death by brain hemorrhages. In Iceland, the condition is caused by L68Q mutation in the cystatin C gene. The genealogies revealed that until around 1820, individuals with the mutations had a life span of around 60 years, but between 1820 and 1900, the life expectancy for people with the same disorder dropped to about 30 years and still stays the same. The group speculated in the research article that an environmental change in the period after 1820 modified the way cells respond to and control the effects of the mutation. The shift could also be possible due to a change in the diet from a traditional to a more mainstream European diet, such as an increase in the consumption of more carbohydrates or salt (for food preservation).

Another remarkable study on the human cohort is from Dutch Hunger Winter in 1944-45 after World War II which includes individuals who were prenatally exposed to famine [122, 123]. Women pregnant during the period gave birth to babies who were affected by health problems like being unusually small in size and prone to obesity and diabetes throughout their lives. In 2008, decades later it was observed that the imprinted IGF2 gene was hypomethylated in offspring as compared to their unexposed, same-sex siblings [122]. Further studies revealed minor but consistent differences in DNA methylation at five differentially methylated regions (DMRs) that regulate the imprinted status of the IGF2/H19 regions regulating genes involved in growth and metabolism [124]. In Newborn Epigenetics Study (NEST), DNA from umbilical cord blood leukocytes from 79 newborns was examined. The findings from the data showed an increase in DNA methylation at the IGF2 and H19 DMRs in babies born to obese moms, which aligned with the previous findings. These data contribute to the hypothesis that environmental exposures in early life can lead to lifelong epigenetic changes in humans [125].

A very fascinating study on the inhabitants of the Azorean Island showed a transgenerational epigenetic inheritance in humans caused by fetal Thyroid Hormone (TH)

exposure. The people from the Azorean Island of Sao Manuel have Thyroid hormone receptor (THR) autosomal dominant mutation and it occurs at a high frequency. Wild-type fetuses of heterozygous THRbeta mothers were exposed to high amounts of TH in utero without being harmed by maternal hyperthyroid illness outcome. The children of F1 males which were exposed to high TH levels during fetal life showed Reduced Sensitivity to TH (RSTH), while the children of the females didn't show the same phenotype. The F3 generation, whose great-grandmother had the heterozygous THR mutation and whose grandfather was exposed to raised TH levels in utero exhibit RSTH. Prolactin levels which are known to be associated with TH levels showed no differences in these individuals which suggests a novel approach by which this condition is inherited without the use of DNA mutations [126]. This is one of the most compelling examples of how unique phenotypes may be inherited in response to specific ancestral experiences. In mice, similar behavior has been reported [127].

We know very well that diabetes has a major effect on our health, it even affects sperm quality and more during reproductive years. Metformin is a known primary diabetes drug used for decades, Its effects on offspring's health were never studied. A very recent study on the Danish population (1997 to 2016) published in 2022 in the Annals of Internal Medicine looked at the offspring of men who were taking diabetes medication during the development of fertilizing sperm. What they found is that the sons born to those men were more than three times as likely as unexposed newborns to suffer a genital birth defect. Encouragingly, the researchers found no effect on the progeny of men who used the drug earlier in life or the year before or after the 90-day sperm production window. Metformin has been earlier shown in fish and mice studies to affect the development of male reproductive organs. As the authors said this is the first study and more studies should be done to replicate these findings and determine the cause [128]. Another study on the Japanese mother-child cohort showed that pregnant women's usage of disinfectants may put their children at risk for asthma and eczema. As the use of disinfectants is essential in today's world for the prevention of

various infectious diseases, this work should be replicated, and more investigation into the processes is needed [129].

A quite troubling pattern emerges through all these studies from human beings to animals, from famine to feast. Maybe the old aphorism 'We are what we eat' no more stands true in the present light of research. We are also what our parents and grandparents ate, not even that we are also what they were exposed to. This could lead us to wonder if the advice on a healthy life is any more relevant or not.

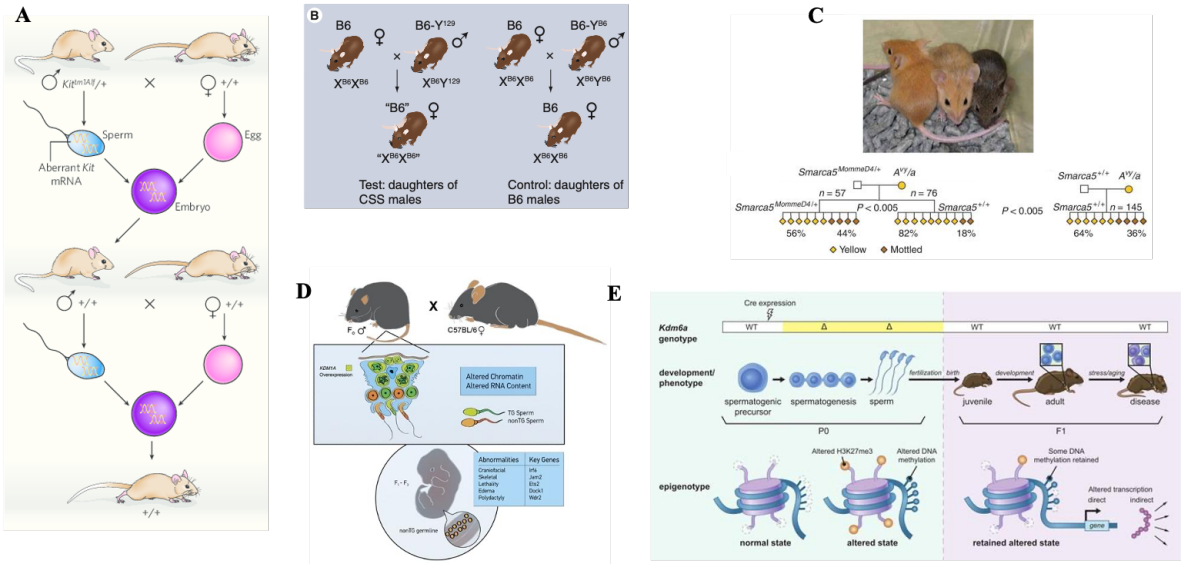


Figure 1.11: Studies on Non-genetic inheritance in mice. **A.** Non-Mendelian inheritance of mouse paramutations. **B.** The study design examined the effect of the paternal Y chromosome on the phenotypes of daughters. **C.** Study shows Smarca5 and Dnmt1, modifiers of epigenetic reprogramming. Isogenic mice with the agouti viable yellow allele have a variety of coat colors (yellow, mottled and agouti). **D.** The disruption of histone methylation in developing sperm caused by KDM1A transgene overexpression from a single generation substantially affected embryo development and child survival. [130] **E.** Model for intergenerational epigenetic inheritance following the deletion of Kdm6a in the male germline [131]

1.10 Indirect Genetic Effects

Yet most studies focus solely on genetic variants. Evidence for non-genetic mechanisms of transmission between parents and offspring (i.e., heredity) is accruing at a fast pace. Nongenetic mechanisms of inheritance include cultural, ecological and epigenetic inheritance, as well as parental effects and niche construction, all of which potentially contribute to heredity [132, 133, 134, 135, 136, 137]. Indirect genetic effects (IGEs) occur when parental genotype impacts the phenotype of non-carrier offspring irrespective of environmentally acquired phenotypes [138]. IGEs are genotype-driven effects and are genetically independent. IGEs of parents on their children can also be considered a special case of nongenetic inheritance. One interesting feature of IGEs is the increased variability and partial penetrance in offspring phenotypes.

During the last decade, few cases of IGEs have been reported in mammals and lower organisms. The "Kit paramutation," which depicts a persistent alteration of Kit gene expression, was the first mouse model for a non-Mendelian mechanism of inheritance [95]. When heterozygous Kit $^{tm1Alf/+}$ (Kit) mice are bred to wild-type (wt) mice, they produce Kit$^{tm1Alf/+}$ (Kit) pups with characteristic white tails phenotype. When these mice are mated to wild-type mice again, a portion of the progeny maintains the spotted white-tail phenotype, even if the genotype is wild (Figure. 1.11A). This phenotype may also be generated by microinjecting RNA into fertilized oocytes, indicating that RNA plays a key part in the inheritance pathway. Maternal miRNAs and piRNAs seemed to have an inhibitory influence on the effectiveness of paramutation germline transfer [139]. Biologically, we know that daughters do not inherit the Y chromosome and therefore should not share the phenotype of fathers. The study published in 2010 took advantage of chromosome substitution strains (CSSs) of mice in which the host strain's Y chromosome has been swapped with the donor strain's Y chromosome. Daughters from the CSS-Y men and host strain females are genetically identical and should be phenotypically equivalent in the absence of transgenerational genetic effects of the fathers' Y chromosome on the phenotypes of the daughters (Figure. 1.11B). Surprisingly, the results from this study found that although daughters were genetically identical to fe-

males from the host, they show phenotypic similarities with the father but with different intensities [140].

Ng et al. 2010 published in Nature the first report of non-genetic, intergenerational transfer of metabolic consequences (impaired glucose-insulin homeostasis) of an HFD from father to their female offspring in rats. Chronic HFD consumption in fathers showed an increase in body weight, adiposity, impaired glucose tolerance, and insulin sensitivity. The daughters of these fathers had an early onset of impaired insulin secretion and glucose tolerance, that aggravated with time and induces common changes in the transcriptomes of retroperitoneal adipose and pancreatic islet tissues. Differential expression analysis pinpoints to *Il13ra2* gene which is part of the Jak-Stat signaling pathway, it even shows reduced methylation in HFD offspring [6, 141].

Smarca5 and *Dnmt1* genes are known as modifiers of epigenetic reprogramming and have shown parental intergenerational effects in the mouse (Figure. 1.11C). Parental heterozygous autosomal mutations in these chromatin regulators influence the phenotype of their offspring and are consistent with the assumption that epigenetic information that controls gene expression is transmitted in mammals over generations. Histone modifier dosage has been shown able to regulate intergenerational effects [142]. Alterations in H3K4 methylation in the germline, caused by overexpression of the H3K4 demethylase KDM1A (also known as LSD1), also affect future generations in mammals as it caused reduced survival and developmental abnormalities not only in F1 but also in the F2 generation (Figure. 1.11D). These results demonstrate that histone demethylase activity in developing sperm can trigger epigenetic inheritance of abnormal development without affecting DNA methylation at CpG-rich locations [130] (Figure. 1.11E). These studies in mice have shown that the untransmitted genotype of male parents can influence the offspring phenotype.

In another study, the *Kdm6a* gene, present on the X chromosome, deleted in the paternal germline resulted in an increased incidence of cancer in subsequent generations which indicates transgenerational inheritance of the sperm epigenome. Interestingly, they observed that the heterozygote offspring had epigenetic features similar to the

mutant sperm. The epigenomic study in these mice unveiled that the majority of the modifications were erased during cellular reprogramming, but some of the methylation marks escaped the reprogramming machinery and passed on to the offspring [131]. The mechanism that defies reprogramming is not yet known. But these results raise questions regarding drugs used for cancer treatments, which target DNA methylation. Whether the cancer patients exposed to these drugs could transmit these modifications if they have children during or soon after their treatment. A basic understanding of how epigenetic changes affect hereditary diseases could provide answers to families affected by cancer.

What was pretty remarkable from these studies is that heterozygous mutations in the parental generation create germline epimutations, which are passed to children during fertilization and shape their developmental and phenotypic trajectories independently of the inherited genotype (Figure. 1.12). And these phenotypes we believe are gene-dependent, genotype independent.

Since paternal and maternal background can affect the transmission of traits through non-genetic means, this calls into question the role of sexual selection as well because we know that sexual selection is a major force, stronger than natural selection, in shaping evolution.

1.11 Relevance of large scale IMPC data

The International Mouse Phenotyping Consortium (IMPC) is an international establishment, to provide access to comprehensive and standardized mouse phenotypic data for the purpose of identifying and characterizing phenotypic abnormalities associated with each protein-coding gene knockout in the mouse genome [143]. To date, the IMPC has generated 8457 mutant lines and 7824 genes have been phenotyped from the mouse genome. Importantly, all mutants and their respective wild-type control animals were generated exclusively from heterozygous breeders on a coisogenic C57BL/6N background. Phenotypic data covers a wide range of system areas, including neurological, behavioral, metabolism, cardiovascular, pulmonary, reproductive, sensory, and musculoskeletal functions. The IMPC pipeline's broad range of phenotypes has enabled

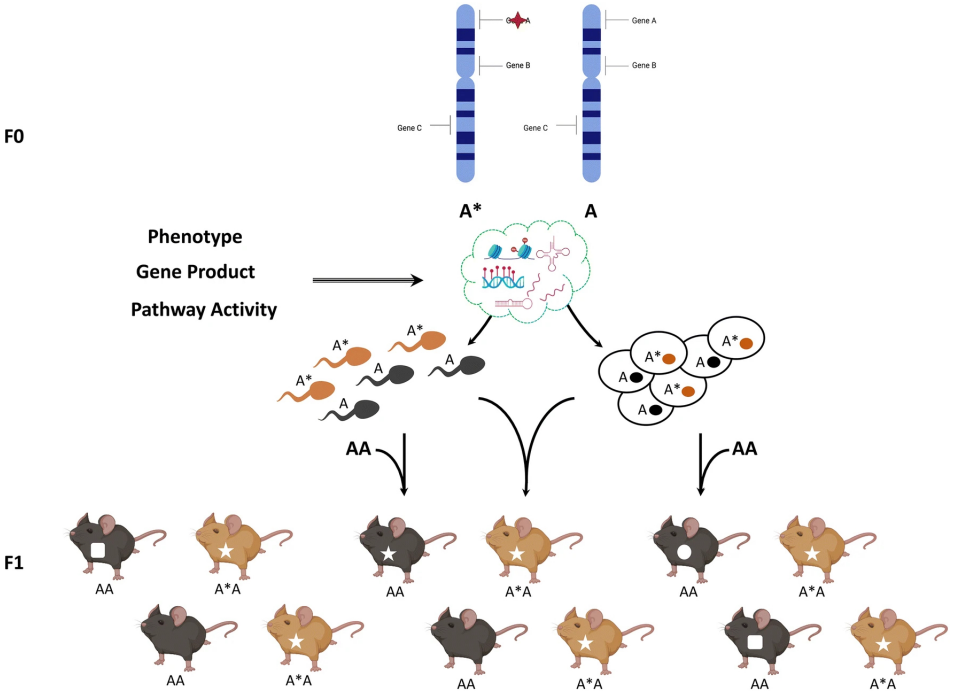


Figure 1.12: Diagrammatic representation of Indirect Genetic Effects (IGEs). This diagram depicts an example of what we mean by indirect genetic influences. Heterozygous mutations in the parental generation create germline epimutations, which are passed to children during fertilization and shape their developmental and phenotypic trajectories independently of the inherited genotype. The increased diversity and partial penetrance in offspring phenotypes is an intriguing aspect of IGEs.

research of putative genes implicated in certain disease areas that were previously unknown or unexplored [144].

We are well aware that data collected from mouse models is becoming biologically relevant for human clinical studies, increasing the value of an extensive phenotyping approach for a broader analysis of multidimensional data sets. These large-scale and multidimensional gene-phenotype datasets will shed new light on our understanding of the mammalian genome landscape and reveal many unknown dimensions of gene functions. An extensive study on the IMPC provides extensive novel insights into gene function along with numerous new disease models [144]. Systematic mouse phenotype data

analysis by different groups has identified candidate genes for metabolism abnormalities [145], eye development [146], auditory dysfunction [147], and bone mineral density [148]. To date, more than 2,000 international publications have been published which are benefited from the IMPC resources. It is phenomenal that 360 IMPC lines (40%) have phenotypic similarity with 889 human disease genes, and the majority (279, 78%) of lines are the first reported mouse model for these diseases. A recent study used the IMPC data to identify 486 genes that have never been associated with cardiac diseases in humans [149].

Such a valuable resource can assist in defining candidate genes to evaluate a condition of interest, provide information on the mechanisms involved or provide support for predictions of gene function that can play a role in the adaptation [143]. Mouse genetics is an effective tool for establishing links between genes and diseases and will shed some light on human physiology. Together, these methods provide unprecedented opportunities to analyze in vivo processes and systems to better understand pathophysiology and disease.

1.12 Aim and scope of this thesis

Our study focuses on the non-DNA-based inheritance and the wide spectrum of phenomena/mechanisms by which parents influence phenotypic variation in future generations.

A lot is done on how genetic changes like gene deletions, SNP, mutations etc directly impact the phenotype or disease, one such large-scale studies in humans are GWAS where population-based genetic modifications are linked to diseases and phenotypes. Another large-scale project focused on the mouse model system and called IMPC (International Mouse Phenotyping Consortium), aims at deciphering the functions of every mouse gene by knocking out and systemic phenotyping. . These are examples of Direct Genetic Effects (DGEs) studies where an individual's phenotype is directly impacted by its own genes.

Parents influence offspring phenotypes in many different ways. What we know is that other than parental genetics, their environmental exposure, either pre-conceptionally, during gestation, or post-delivery often influences offspring phenotypic trajectory. Recent studies have shown that parental genetic variation induces phenotypic variation also in non-carrier, wild-type, offspring - a phenomenon known as Indirect Genetic Effects (IGEs). Indirect genetic effects lead to complex inheritance pathways in which genetic and environmental sources of variation can be passed down over generations and hence contribute to phenotypic variation, adaptation and evolutionary change. IGEs not only constitute a paradigm-shift in our ken of genotype/phenotype relations across generations, but their full knowledge also holds the potential to provide a novel and full picture of evolutionary processes, modes of inheritance and complex disease (epi)genetics.

Up to now, the vast majority of studies on IGEs have focused on either canonical epigenetic modifiers, or on large parental genetic or genomic manipulations. What my thesis aims at is to understand whether this phenomenon is peculiar to epigenetic modifier genes or common in mammalian genetics, and to identify the underlying genetic determinants.

To achieve the above objective first, we will use existing systemic phenotypic data provided by the IMPC to isolate genetic determinants of phenotypic variation in non-carrier individuals (Figure. 1.13).

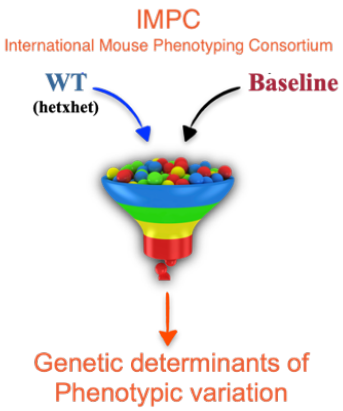


Figure 1.13: Model of the study to identify genetic determinants of Indirect Genetic Effects(IGEs).

Second, we will characterize gene or gene family-dependent molecular determinants of the parental effect(s).

Our results will provide a resource for deciphering IGEs in detail and will be valuable for understanding how much of complex phenotypes heritability is hidden by indirect genetic effects and epigenetic mechanisms, and to what extent these phenomena contribute to the pathogenesis of complex diseases.

2 Methodology

2.1 The International Mouse Phenotyping Consortium (IMPC)

The IMPC program entails the objective and systematic characterization of 20,000 known and predicted mouse genes knockout mouse strains using standard operating procedures (SOPs). The phenotyping is accomplished at major research centers across Europe, North America, and Asia. Standardized techniques have been devised to maintain uniformity and data quality throughout the centers. All mouse procedures are carried out in compliance with the various member centers and animal welfare authorities. The SOPs are detailed in the International Mouse Phenotyping Resource of Standardized Screens (IMPReSS) http://www.mousephenotype.org/impress (Figure. 2.1). Each IMPC center is free to nominate and therefore prioritize genes for systemic phenotyping, mostly based on their own research focus. The only guideline is to avoid duplication by not nominating genes more than once. The main objective of the consortium is to focus on choosing poorly understood genes for which there is little to no knowledge available. All mouse lines for the phenotyping are born and raised on a C57BL/6N genetic background, with support animals descended from C57BL/6NJ, C57BL/6NTac, or C57BL/6NCrl. Phenotyping data are collected between the age of 4 and 16 weeks following approved animal ethics protocols in every institution.

2.2 IMPC Data Collection

The IMPC provides a unified point of access to mouse phenotypic data from 19 research centers across the globe. For this study, we worked on the dataset from Data release 11.

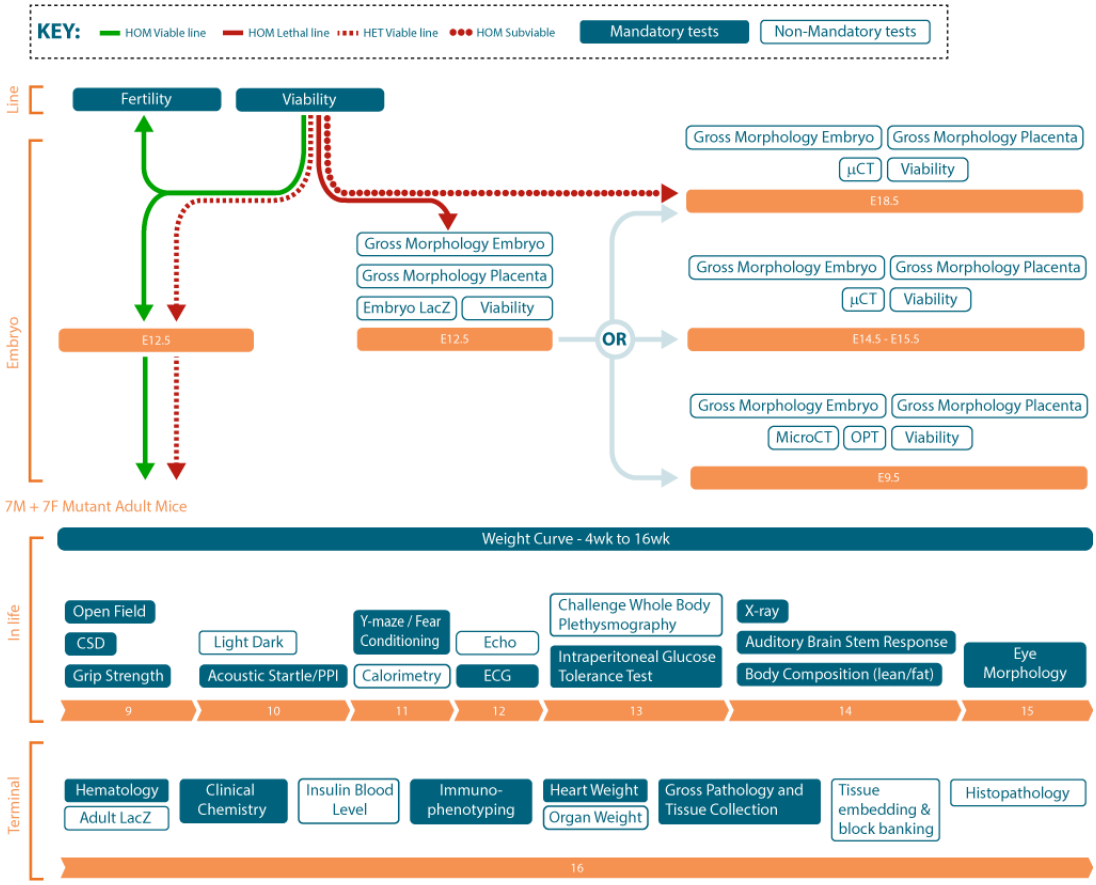


Figure 2.1: IMPRESS phenotyping pipeline The IMPC fundamental pipelines illustrate the consortium-wide phenotyping workflow.

The centers do systematic phenotyping of Baseline Control animals (C57BL/6N) and homozygous knockout mice which are viable. When homozygous mice are not viable, more homozygous animals are introduced into the embryo pipeline, and adult heterozygous mice, when viable, are phenotyped. Otherwise, centers also phenotype heterozygous animals sometimes. Our interest for this study is two populations of isogenic background: pure C57BL/6N wild-type lineage, named **Baseline (Ctrl)** animals; and C57BL/6N animals with a high degree of inbreeding derived from mutant parents that are heterozygous, named **wildtype (WT)** animals (Figure. 2.2A). We obtained from the IMPC an exclusive access to the data from these two cohorts, as they are not directly accessible from the IMPC website.

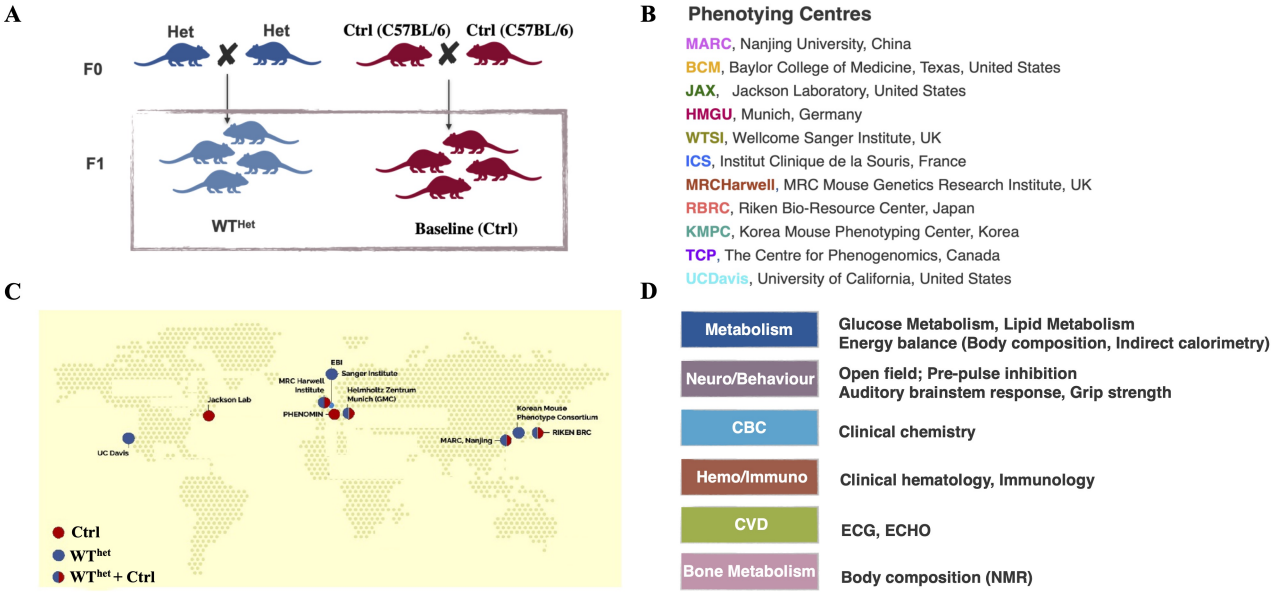


Figure 2.2: Representation of IMPC Data used in this study. A. Graphic representation of the breeding scheme of Baseline (Ctrl) animals; and C57BL/6N animals with a high degree of inbreeding derived from mutant parents that are heterozygous, named wildtype (WT) animals. **B.** List of the phenotypic centers which contributed data for this study. **C.** Geographical distribution of the phenotyping centers and their respective contribution of cohorts (WT, Baseline or Both) for the study. **D.** Representation of five broad physiological categories based on 129 phenotyping parameters used in the study.

2.3 Data verification, filtration, and quality check

In IMPC the WT and Baseline animals are mostly used interchangeably as Controls in the publicly available dataset, that's why it was difficult to differentiate between these animals from the online accessible data release. On special request, IMPC provided cohorts of wild-type offspring (WT) coming from mutant parents. As the parental status of this data was not fully confirmed by IMPC we further approached the respective phenotypic centers for the additional information and to identify parents of origin. In total we collected phenotyping data for 51608 animals from which 28637 were WT, 22971 Baseline from different phenotyping centers (Figure. 2.2B-C and Table. 2.1)

. For this study, we only focused on Baseline and WT animals. The whole dataset is a collection of 786 phenotyping parameters. The non-numerical, image data and no data parameters were excluded. After exclusion and filtering, we scaled-down to 129 phenotyping parameters, which were further categorized into 5 broad physiological categories (Figure. 2.2D).

Phenotyping Center	Total WT Animal	Total Baseline
BCM	4882	849
HMGU	1336	1146
MRC Harwell	2387	4634
ICS	17	2664
JAX	0	11254
KMPC	529	802
MARC	1949	0
RBRC	2038	77
TCP	6951	1303
UCD	6868	0
WTSI	1680	242

Table 2.1: Details of number of animals collected from different phenotyping centers.

After data collection, we performed a few initial data summaries and visualization tasks to understand data distribution. The complete dataset was then arranged in a multidimensional data matrix for further analysis. Animals with at least 3 replicates per group and sex were included for further steps and the rest were removed. Initial data visualization revealed that at least 48 % of data was missing if we pool the data in the single data matrix. Imputing such a high percentage of missing data is statistically not recommended, so we applied stringent filters to reduce the percentage of missing data but only to the extent that we do not lose too much information. For multidimensional data analysis, we performed data imputation with a random forest imputation algorithm

using *missForest* R package [150]. *MissForest* is a non-parametric method and can be applied to mixed data types. It first imputes all missing data using the mean/mode, then for each variable with missing values, it fits a random forest on the seen portion and predicts the missing part. This training and prediction procedure is repeated iteratively until a stopping condition is satisfied or a maximum number of user_specified iterations is achieved. The imputed data matrix was z—score normalised and plotted for data distribution visualization grouped by sex across all the parameters included in the study.

2.4 Principal Component Analysis (PCA) and correlation Analysis

The primary analysis plan is to identify the major source of variability and relation between the animals for which we performed PCA and correlation analysis. The data was mean averaged per sample category. The data matrix was then scaled and the *prcomp* function in R was used to determine the principal components of the complete dataset. To quantify the difference between the genes based on phenotypes we use the Spearman (rank) correlation method using *get_dist* from the *factoextra* R package. Spearman's rank correlation is a nonparametric measure using the rank values of the two variables. These correlation coefficients were calculated to identify similarity patterns in gene-phenotype pairs and visualized using a heatmap generated by using the *ComplexHeatmap* package from R. We used PCA based 2-dimensional (2D) visualizations to visualize the single animal of the cohorts by using t-Distributed Stochastic Neighbor Embedding (t-SNE) pca = TRUE, perplexity=50, theta=0.5, dims=2, epoch = 1000.

2.5 Strategy for candidate genes selection

For candidate gene selection we performed a phenotyping center based analysis which means WT animals were compared against the Baseline animals (Figure. 2.3A) coming from their respective phenotyping center. Groups with at least 4 animals/sex were included for the comparison for statistical power. Males and females were analyzed

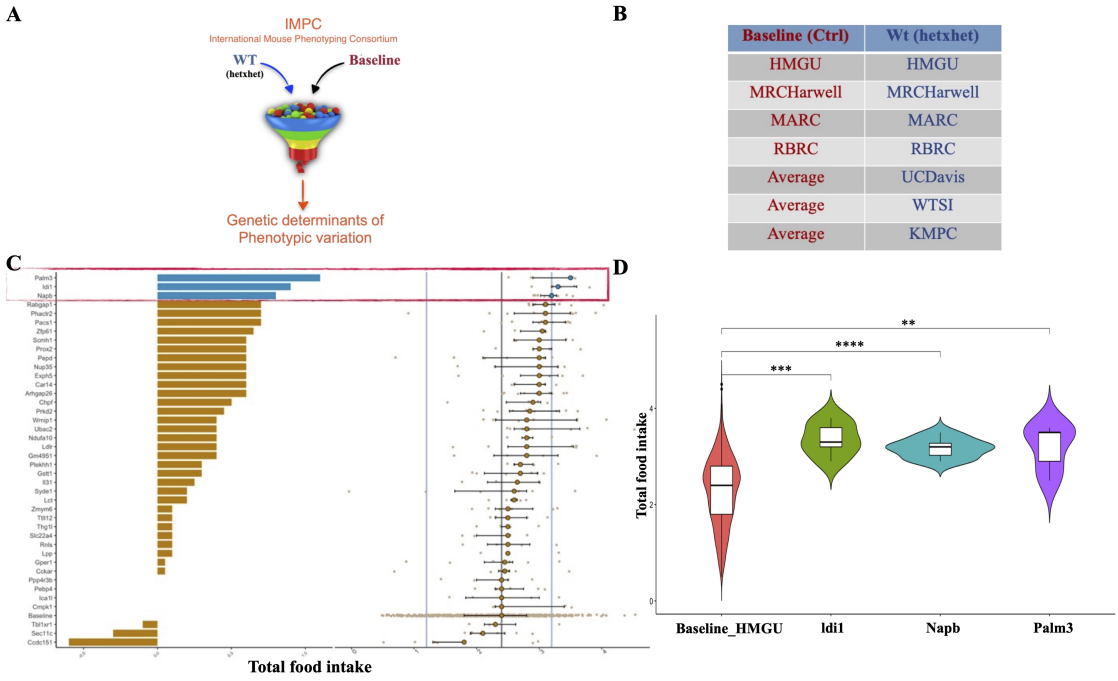


Figure 2.3: Detailed strategy for identification of candidate genes responsible for IGEs. A. Schematic diagram of the methodology followed in the study, **B.** Table shows center based comparison of Baseline and WT animals. **C.** Quantile based analysis, where WT animals which fall above quantile 90 (Q90) (shown in blue within the red outline) and below quantile 10 (Q10) are selected. **D.** Selected candidates further confirmed if they showed statistically significant differences (wilcoxon test) between the Baseline and WT animals.

separately to account for the sexual dimorphism typical of complex phenotypes [151]. In case the center doesn't have its own baseline animals, they were compared against the average baseline which includes all the baseline animals (Figure. 2.3B). We compared every WT animal group for each phenotyping parameter against the baseline animals of the same and selected only those whose median value lies above quantile 90 and below quantile 10 of the baseline animals data range (Figure. 2.3C). These comparisons were again statistically tested using the wilcoxon test to validate if the groups are different from one another in a statistically significant manner using the *compare_means* command from the *ggpubr* R package with default settings (Figure. 2.3D).

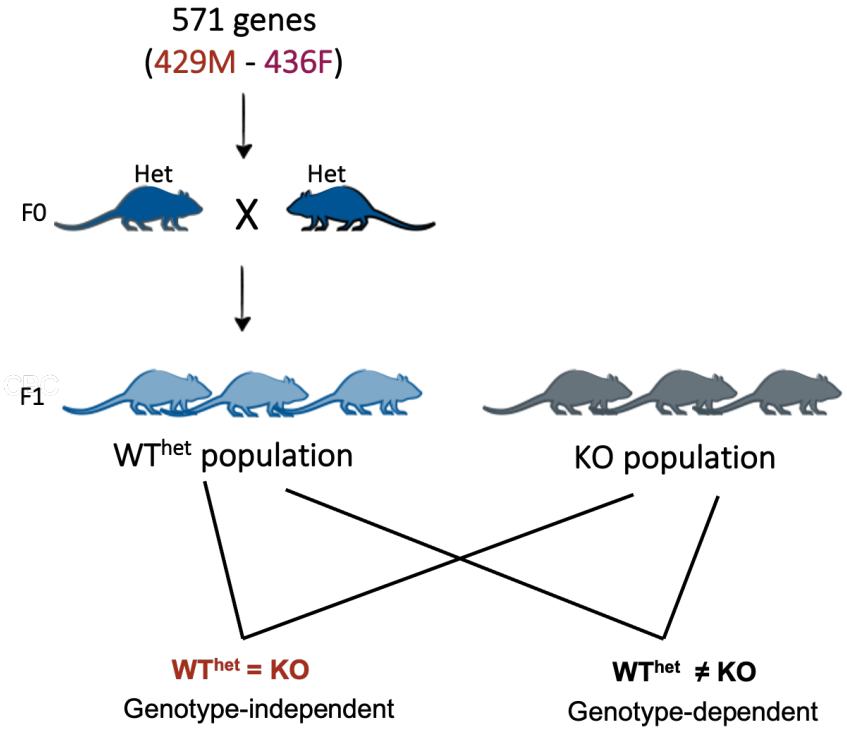


Figure 2.4: Identification of genotype dependant and independent phenotypes. Comparison between WT and HOM (homozygous KO mutant) offspring of the same heterozygous mating to differentiate between gene-dependent and genotype-independent candidates.

IGEs are defined as gene-dependent (as they depend on the mutated gene in the parental generation) and genotype-independent (because their phenotypic consequences are manifested in wild-type offspring of mutant parents). This definition entails that potential epimutations induced in the parental generation by the gene mutation have consequences in the next generation in a genotype-independent manner. Therefore, WT and HOM (homozygous mutant) offspring of the same heterozygous mating should have similar phenotypic characteristics when compared to the Baseline population (Figure. 2.4). To test this hypothesis, we downloaded HOM phenotyping data from IMPC using API and compared them with their WT littermates using wilcoxon test to validate if the groups are different from one another in a statistically significant manner.

2.6 Gene Annotations and pathway analysis

The functional annotation for Gene Ontology (GO), including biological process, cellular component, and molecular function, was performed using the open-access *WebGestalt* tool (http://www.webgestalt.org) [152]. The same tool was also used to implement the KEGG pathway enrichment analysis and for Over-representation analysis of the genes. Top results with the false discovery rate (FDR) $\leq$ 0. 05 were considered significant. Next, we conducted *Gene Set Enrichment Analysis (GSEA)* to uncover the signaling pathways and biological processes for which the geneset were enriched (http://software.broadinstitute.org/gsea/). The *STRING* (Search Tool for the Retrieval of Interacting Genes/Proteins) was used for protein-protein interaction (PPI) network analysis, a web-based visualization resource. The network analysis was carried out using a confidence interaction threshold of 0.40. We used the q value, which is the adjusted P value using the Benjamini–Hochberg FDR method with a 5% cutoff for correction for multiple hypotheses testing. The resulting diagram depicts the participation and interaction of hallmark genes in the PPI network.

2.7 Publicly available data

2.7.1 Genomic Location

The genes were plotted based on their chromosomal location on the mouse genome using *RIdeogram* package from R. We used *Cluster Locator* to determine number, size, and position of all the clusters formed by the genes of interest and statistically analyze the distribution of those genes along the reference genome and the percentage of gene clustering found (http://clusterlocator.bnd.edu.uy/). Analysis was performed on 555 genes with a max gap of 5. The gap between two given genes is the number of other genes located between them, defined by the location of their starting points in the reference genome.

2.7.2 Topologically associating domains (TADs)

We used publicly available Hi-C data from mouse embryonic stem cells available from 4DN Data Portal (https://data.4dnucleome.org). They performed Micro-C on JM8.N4 mouse embryonic stem cells (Pettitt et al., 2009) (male mESCs; Research Resource Identifier: RRID:CVCL_J962; obtained from the KOMP Repository at UC Davis) [165]. Micro-C is a variant of Hi-C that improves resolution and signal-to-noise ratio, which were possible limitations of classic Hi-C owing to the chemicals employed in its workflow. We downloaded the available processed file which includes TAD boundaries data in bed format. We used the Bedtools intersect function to find overlaps between the genomic coordinates of IGE genes with regions of TAD and TAD boundaries across the TAD landscape. TAD boundaries are regions bordering TADs. A 100 kb TAD boundaries are defined as sections 100 kb upstream of the TAD start and 100 kb downstream of the TAD end. If a TAD existed at chr1: 2,000,000–3,000,000, we would describe its TAD bounds as chr1: 1,900,000–2,000,000 (start boundary) and chr1: 3,000,000–3,100,000 (end boundary) (boundary around the end) [153]. Any borders that overlapped with genomic gaps (UCSC table browser centromeric/telomeric repeats) were removed. Genes of interest were checked if they overlap with these identified TAD boundaries.

2.7.3 Genome-wide association studies (GWAS)

GWAS examines hundreds of thousands of genetic variations across populations with different ethnicities and regions to uncover those that are statistically related to a given trait or illness. The GWAS catalogue hosted at https://www.ebi.ac.uk/gwas/ website gives free access to these association studies. The *All associations v1.0* was downloaded from the GWAS Catalog (https://www.ebi.ac.uk/gwas/docs/file-downloads). It includes 47232 associations between genetic variations like SNP to diseases and phenotypes. To compare our genes of interest to GWAS we converted mouse genes to human and then matched them to gwas association data to understand which genes have correspondent genetic alterations in humans and are associated with known disease or phenotypes according to GWAS data.

2.7.4 mouse embryonic stem cells (mESCs) ChIP-seq and ATAC-seq data analysis

Several publicly available databases for chip-seq and atac-seq from mouse embryonic stem cells (mESCs) were downloaded [154] Table 2.2. Reads from ChIP-seq experiments were mapped to the mouse genome (mm10) using the Bowtie software [155]. Only those reads that matched to a unique site with no more than two sequence mismatches were kept for further analysis. Peaks were called using the MACS2 software [156] using a bandwidth parameter of 150bp. Peaks with q-value cut-off <0.005 and fold >= 4-fold were retained. Peak annotation has been performed using HOMER (http://homer.salk.edu/homer/). For ATAC-seq we used the peaks data file available from the published article.

2.7.5 Chromatin segmentation analysis

Mapped dataset from ChIP-seq experiments of histone modifications and Transcription Factors (TFs) was used for chromatin segmentation analysis by the *EpiCSeg* software [169]. The software splits the genome into a regular grid and assigns a state to each bin based on abundance and co-occurrence of histone marks and TFs. EpiCSeg *getcounts* function generates a count matrix from a list of bam alignment files, and the count matrix is normalized using *normalizecounts* function. The bin size was default set to 200 bp. The segment function produced the segmentation, the number of states is a free parameter and was set to twenty five (–nstates 25) after trying different states from 8 to 30. The segmentation was calculated only for the genomic regions of interest. Based on the matrix of histone mark counts, the states are labelled after segmentation.

2.7.6 CpG islands Identification

To annotate CpGs with respect to whether these genes reside in CpG islands, we downloaded the mm10 cpgIslandExt table from the UCSC table browser

(https://hgdownload.soe.ucsc.edu/goldenPath/mm10/database/cpgIslandExt.txt.gz). The cpgIslandExt table contains annotations of CpG islands, where a genomic region is a CpG island if it meets the following criteria: having >50% GC content, a length

GEO accession	Target	Marks	Type
GSM2417080	H3K4me3	Histone Marks	ChIP-Seq
GSM2417084	H3K4me2	Histone Marks	ChIP-Seq
GSM2417088	H3K4me1	Histone Marks	ChIP-Seq
GSM2417092	H3K9ac	Histone Marks	ChIP-Seq
GSM2417096	H3K27ac	Histone Marks	ChIP-Seq
GSM2417100	H3K27me3	Histone Marks	ChIP-Seq
GSM2417100	H3K27me3	Histone Marks	ChIP-Seq
GSM2417104	H3K79me2	Histone Marks	ChIP-Seq
GSM2417112	H3K36me3	Histone Marks	ChIP-Seq
GSM2417112	H3K36me3	Histone Marks	ChIP-Seq
GSM2417116	H3.3	Histone Marks	ChIP-Seq
GSM2417120	H3	Histone Marks	ChIP-Seq
GSM2417124	Input native MNase	Input	ChIP-Seq
GSM2417127	WCE	Transcription Factor	ChIP-Seq
GSM2417142	Oct4	Transcription Factor	ChIP-Seq
GSM2417143	Sox2	Transcription Factor	ChIP-Seq
GSM2417144	Klf4	Transcription Factor	ChIP-Seq
GSM2417145	cMyc	Transcription Factor	ChIP-Seq
GSM2417169	p300	Transcription Factor	ChIP-Seq
GSM2417173	Hdac1	Transcription Factor	ChIP-Seq
GSM2417177	Brg1	Transcription Factor	ChIP-Seq
GSM2417187	Nanog	Transcription Factor	ChIP-Seq
GSM2417188	Esrrb	Transcription Factor	ChIP-Seq
GSM2417076	ATAC	Chromatin	ATAC-seq

Table 2.2: Details of the mESCs Chip-seq and ATAC-seq sequencing data downloaded from NCBI.

of >200bp and a ratio of observed to expected CG dinucleotides of >0.6. We assigned each CpG island a unique identifier, and examined overlaps between CpGs and IGEs candidate genes using the intersect function of *Bedtools* (v2.29.0, [157]).

2.7.7 Human diseases

We used *disgenet2r* R package to retrieve human disease association with our genes of interest. Functional analyses were performed using the *Enrichr* package.

2.7.8 Phenotype based gene interaction Network

To build a knowledge-based network with a reduced set of high-relevant genes, we built a pheontype-based co-correlation matrix and centered the 5 previously described parameter sets (Figure. 2.2D) around what we called core phenotypic terms. For example, IMPC_IPG_012_001 (AUCipgtt) is the core phenotypic term for glucose metabolism, as well as IMPC_CBC_015_001 (Total Cholesterol) is for Cholesterol homeostasis (see Supplementary Figure. 8 - 19 for the entire set of core phenotypic terms). The phenotype-based co-correlation matrix is a Spearman-based co_correlation matrix calculated using *Prism 8* (using the correlation function and a two-tailed p-value threshold of 0.05) on the complete and imputed phenotype dataset, which includes data from WT and Baseline animals. Around the core phenotypic terms and using the phenotype-based co-correlation matrix, we built a network of co-occurring phenotypes (p-value <10-4 / -0.5 <r >0.5), which we used to select relevant genes, showing significant phenotypic variation in WT offspring for the core phenotypic term and >3 parameters of the selected parameter set.

(Figure. 2.5) shows an example of gene selection using the Glucose Tolerance parameter set centered around the AUCipgtt as core phenotypic term. The same procedure was applied to Activity (IMPC_OFD_009_001), Adiposity (IMPC_DXA_002_001), Anemia (IMPC_HEM_003_001), Anxiety (IMPC_OFD_012_001), Cholesterol Homeostasis (IMPC_CBC_015_001), Exploratory Behavior (IMPC_OFD_019_001), Hearing (IMPC_ABR), Heart Function (IMPC_ECG_002_001), Immune function (IMPC_HEM_001_001), Liver

Function (IMPC_CBC_013_001), and Sensory motor gating (IMPC_ACS_037_001) phenotyping catagories to select relevant genes for respective terms.

After selecting phenotype and associated gene we used *network* package [158] from R to draw associations and used log2FC as weight. Network was created using the R package *igraph* [159] and Cytoscape was used to draw the network figure [160].

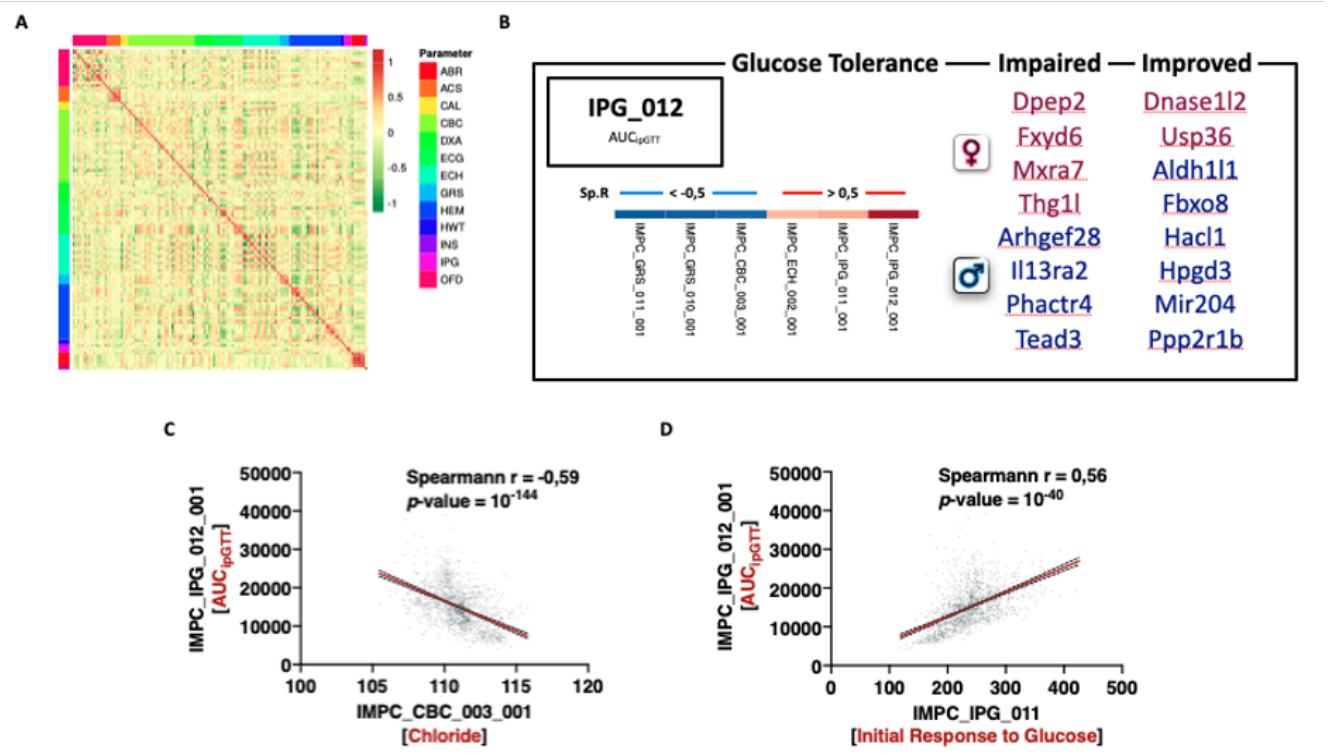


Figure 2.5: Steps to identify genes for Phenotype based gene interaction Network. **A.** Phenotype-based co-correlation matrix is a Spearman-based co_correlation matrix based on imputed phenotype dataset from WT and Baseline animals. **B.** IMPC_IPG_012_001 (AUCipgtt) is the core phenotypic term for glucose metabolism, phenotype-based Spearman co-correlation was calculated to identify relevant genes. **C.** AUCipgtt shows negative correlation with chlorine levels. **D.** wheres shows positive correlation with initial response to glucose.

Note: All data processing, analysis and plotting was performed using R version 4.1.2

3 Results

3.1 IMPC phenotyping data reveals center-based clustering

We recently defined Indirect Genetic Effects (IGEs) as gene-dependent (as they are determined by the parental genetic alterations), genotype-independent (as their manifestation is independent from the carrier genotype) control of phenotypic variation across generations [15]. To estimate the influence of parental genetics on the overall phenotypic variation among isogenic individuals (C57BL6/N mice) discordant for parental genotypes (either heterozygous mutants of pure wild-type), we uniformly compared IMPC phenotypic-data from wildtype animals generated from heterozygous breeders (WT), to Baseline (Ctrl) animals, from IMPC which have never seen any mutation, across 129 phenotypic parameters. As a dimensional reduction algorithm, spearman correlation between the animals demonstrated that following covariate adjustment, animals clustered closely based on the phenotyping center. Heatmaps were formulated with distance between rows and columns calculated by Euclidean distance.

The heatmap (Figure. 3.1A) showed 5 distinct clusters, of HMGU, WTSI, MARC, RBRC and MRC Harwell phenotyping centers. Within these clusters the animals form sub-clusters based on gender. The heatmaps and t-SNE plot of the individual phenotyping center clearly shows the sexual dimorphism (Supplementary Figure. 6.1-6.7) As the correlation analysis is based on averaged data we performed t-SNE where each data point represented as a single animal. We used PCA for t-SNE embedding initialization. PCA is a feature reduction method to project high-dimensional data into a lower-dimensional space that can explain the most variance of the input data. PCA initialization is more globally stable than random initialization. Consistent with our correlation observation,

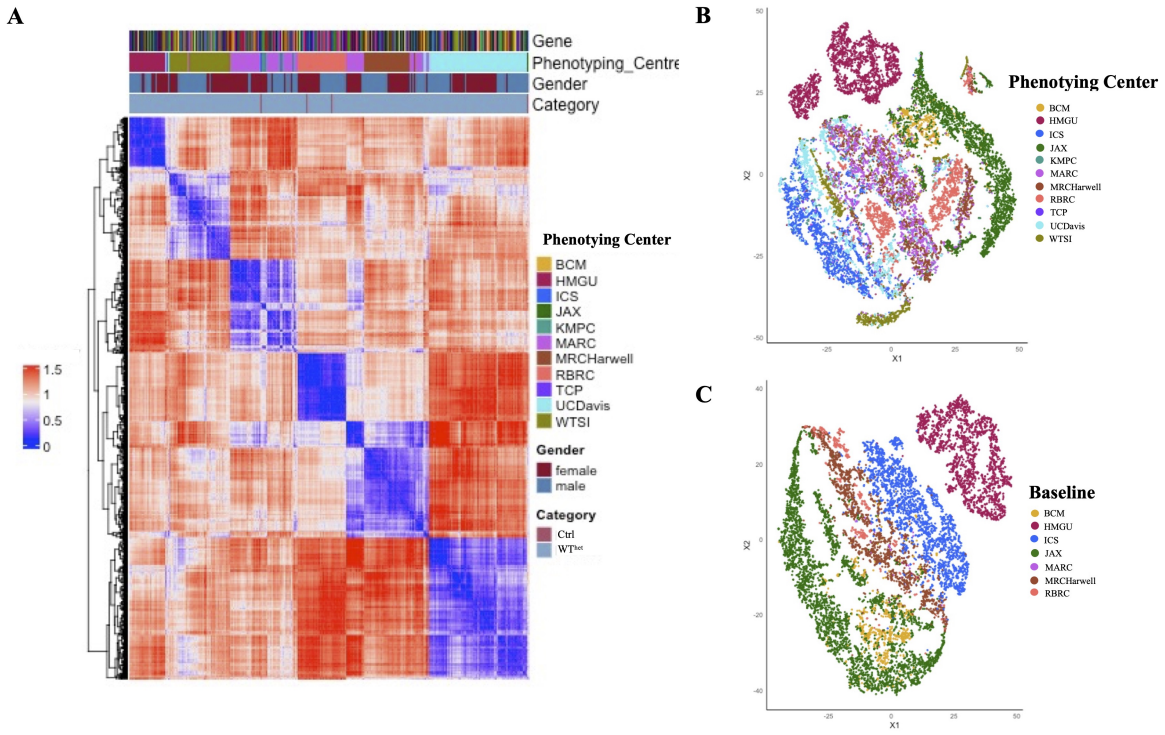


Figure 3.1: Correlation analysis among animals shows center-based phenotypic clustering A. The heatmap matrix shows the Spearman correlation coefficient between animals for all the genes across 129 phenotyping parameters. Samples cluster by phenotypic centers. Colors (blues) represent relationships between samples that are most similar; warmer colors (reds) represent samples that are more dissimilar with lower coefficients. **B**. t-SNE plot for both WT and baseline animals **C**. t-SNE plot for visualizing only the Baseline animals clusters in a projected 2D metric map.

the largest number of animals in the cohort showed center based clustering (Figure. 3.1B). Importantly, restricting the same analysis to Baseline animals highlighted a similar center-based clustering, suggesting a determinant environmental influence on mouse phenotypes (Figure. 3.1C).

The same t-SNE plots when differentiated based on the sex and type of animals (Baseline and WT) (Figure. 3.2) revealed a certain degree of differentiation in the data, which was importantly also evident when performing single-center correlation and t-SNE-based analysis. These results, while warning us on a strong center-based clustering

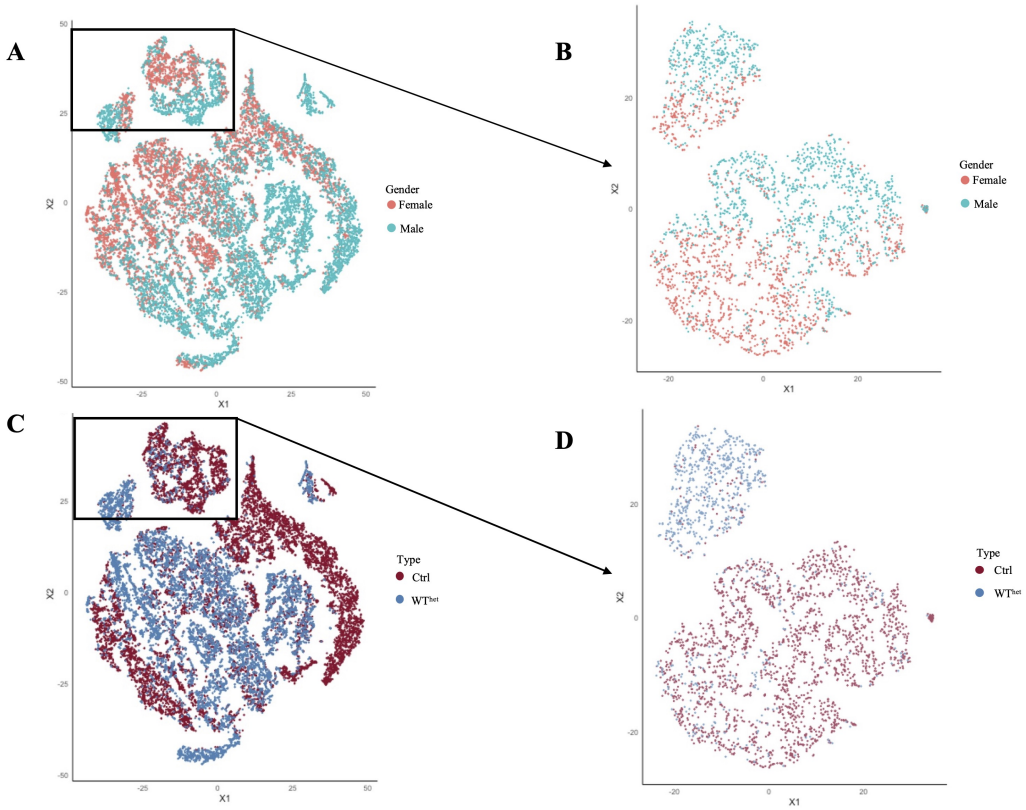


Figure 3.2: Sexual Dimorphism and WT, Baseline animals distribution. **A.** Sex-specific distribution of animals shown as a t-SNE plot. **B.** Sex-based distribution of HMGU phenotypic centers. **C.** WT and Baseline animals distribution shown as a t-SNE plot, **D.** WT and Baseline animals distribution of HMGU phenotypic center.

of the data, also highlighted a certain degree of phenotypic variation between WT and Baseline animals (and therefore IGEs) within individual centers, which encouraged us on further exploring of the IMPC dataset for the identification of novel genetic determinants of indirect genetic effects.

3.2 Wild-type offspring shows phenotypic variation

After the observation that the animal phenotypes are biased by a strong center-based clustering and to normalize for center effects, we adopted a quantile-based, wilcoxon

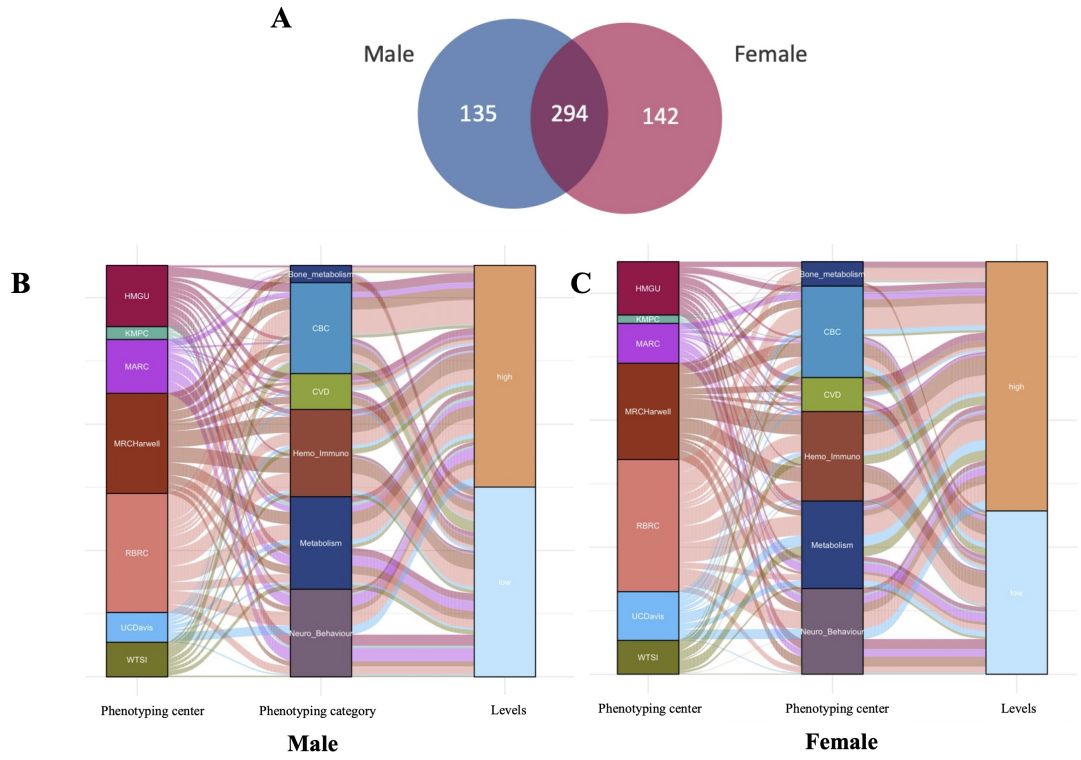


Figure 3.3: Distribution of 571 candidate genes of IGEs **A.** Venn Diagram of candidate genes to show overlap between male and female animals. **B. C.** Alluvial plot shows male and female dynamics of the candidate genes of IGEs. It shows candidates' genes distribution through phenotyping center, phenotype category and phenotype levels to be high or low. Each line represent a single gene connects it to phenotyping center to the respective phenotype categories and levels.

corrected statistical pipeline to compare WT animals coming from a phenotyping center to the respective baseline animals coming from the same center, or to the averaged baseline for centers without own baseline data (Figure. 2.3B). After comparisons, we identified 1275 statistically significant phenotypic differences in the male and 1623 in the female data. These phenotypes reduce down to 571 genes in total, from which 135 were found to be unique to males, 142 unique to females and 294 common to both sexes (Figure. 3.3A). As a quality check, the *alluvialchord* plot shows that the selected

candidate genes are not biased towards any phenotypic center or phenotyping category (Figure. 3.3B and C).

3.3 Functional Annotation of the identified 571 candidate genes

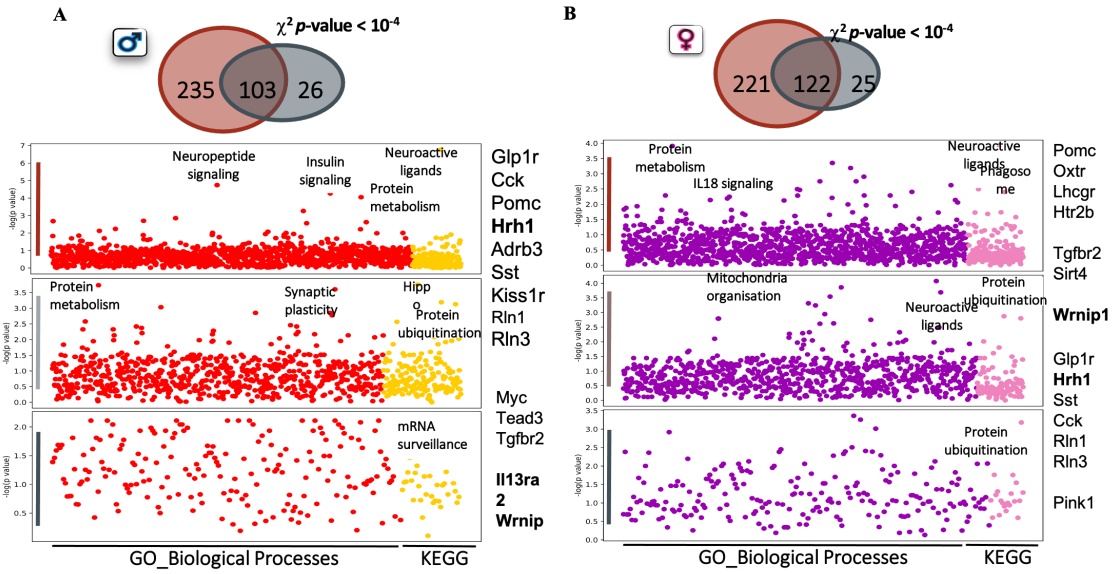


Figure 3.4: Functional annotation of A. male and B. female gene set that are not significantly different between WT and HOM phenotypes from the identified 571 candidate genes.

Indirect Genetic Effects underlie gene-dependent / genotype-independent mechanisms that control phenotypic variation across generations [15]. This definition entails that potential epimutations induced in the parental generation by the gene mutation have consequences in the next generation in a genotype-independent manner. Therefore, WT and HOM (homozygous mutant) offspring of the same heterozygous mating should have similar phenotypic characteristics when compared to the Baseline population. To test this hypothesis, we compared WT and HOM phenotypes from the identified 571 candidate genes with available HOM data (i.e. 368/436 for females and 364/429 for males).

Importantly, the vast majority of phenotypes are not significantly different between littermates (eg. 1095/1319 for females and 1105/1340 for males) highlighting 346/368 genes for females and 338/364 genes for males, which fit the definition of indirect genetic effects. Functional annotation of these genes using Gene Ontology and KEGG Pathway analysis shows enrichment for terms underlying neuropeptide signaling and cellular protein metabolism (via ubiquitination) in both male and female datasets (Figure. 3.4).

Importantly, extending the functional annotation to the entire set of 571 candidate (Figure 3.5) genes by GeneSet Enrichment Analysis, through the online WebGestalt tool (http://www.webgestalt.org/) reinforced the enrichment of similar terms (Supplementary Figure. 6.20), in line with the fact that more than 90% of the identified candidate genes induce proven IGEs. Furthermore, STRING-based Protein-protein interaction (PPI) analysis identified a network with 557 nodes and 1033 edges and showed that the identified 571 genes are highly interconnected and cluster to the previously identified pathways (Supplementary Figure. 6.21A and B). Of note, the PPI enrichment p-value 2.44e-15 shows that the network has significantly more interactions than expected for a random set of proteins of the same size and degree distribution drawn from the genome. Such an enrichment indicates that the proteins are at least partially biologically connected, as a group.

Altogether, these results - generated by three independent analysis methods/tools - indicate that:

1. Indirect Genetic Effects are a common feature of mammalian genes (and therefore not restricted to canonical epigenetic modifiers, as previously suggested);

2. Genetic determinants of IGEs are functionally interconnected; and

3. These results hint to an important role for neurophysiology, as well as protein and mRNA metabolism in non-genetic inheritance.

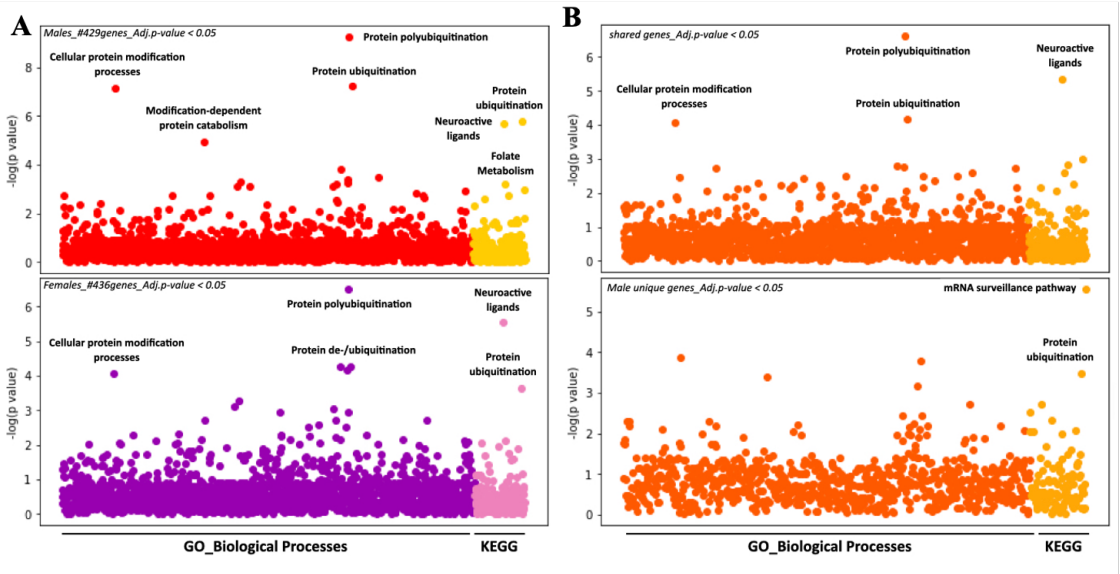


Figure 3.5: Function annotation of A. male and B. female 571 candidate genes identified of IGEs.

3.4 Inferring information about IGEs by exploring publicly available datasets

Genes are known to be non randomly organized in the human genome, both within and across chromosomes [161]. Genes that have comparable functions and evolutionary origins, as well as genes with similar expression patterns, are frequently found to be grouped allowing for coordinated regulation. While there is plenty of evidence to support this, there are even interesting cases of the conserved grouping of genes with seemingly unrelated roles [162, 163]. In more general terms, genomic topology and gene chromatin environment are important determinants of gene regulation, function, heritability and relevance for human diseases. Altogether, these pieces of evidence prompted us to look at the genomic location of the identified IGE-inducing genes and understand their genomic neighbourhood and regulatory landscape. The goal is to identify common features, which together with the functional and phenotypic clustering, would provide a set of theoretical characteristics of IGE-inducing genes.

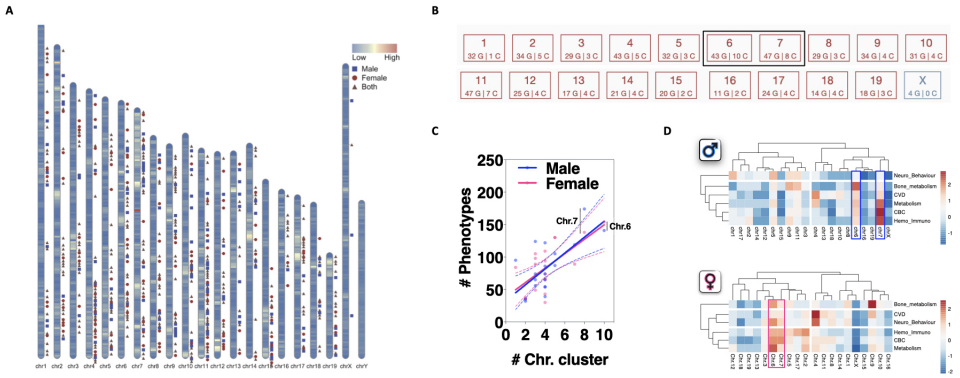


Figure 3.6: Genomic location assessment of 571 candidate genes of IGEs. A. Ideogram illustrating distribution and localization of 571 candidate genes along mouse chromosomes. The position of each candidate gene is represented with different shapes and colors based on their sex and relative position on its respective autosome and sexual chromosomes. **B.** Gene cluster distribution on the mouse chromosomes. **C.** Number of clusters on chromosome correlation with number of phenotypes. Chr 6 and 7 shows the highest correlation. **D.** Heatmap shows chromosome wise phenotypic enrichment.

We therefore mapped the 571 IGE-inducing genes on the mouse chromosomes. We identified that they are uniformly distributed along the 21 mouse autosomes (with an average of 29 genes/autosome) (Figure. 3.6A). Interestingly, 32.07% of the 571 genes (178 genes) form 80 clusters of 2-4 genes per cluster distributed along the chromosomes (Figure. 3.6B). This clustering is statistically significant (p-value <0.05) as compared to the clustering found in 1,000 lists of 571 genes randomly picked from the mouse genome Table. 3.1.

Although cluster density does not correlate with gene density on chromosomes and individual clusters do not seem to underlie common intergenerational effects, cluster density is significantly associated with the pleiotropy of the intergenerational phenotypic variation (Figure.3.6C) and the chromosomal location is associated to specific F1 phenotypes in a gender-dependent manner (Figure.3.6D). For example, neuro-behavioural phenotypes are associated with parental manipulations of genes located to chromosome

Cluster size	# Clusters	# Genes	% Genes
2	66	132	23.78%
3	10	30	5.41%
4	4	16	2.88%
	80	178	32.07%

Table 3.1: Cluster Locator results shows number and size of clusters found and number and percentage of genes on the list forming clusters.

1 and 4 in male and female offspring, respectively (Figure.3.6D). Conversely, metabolic phenotypes are intergenerationally induced by manipulations of genes located to chromosome 7 and 6, respectively in male and female offspring (Figure.3.6D) and interestingly, chromosomes 7 and 6 are also the ones with the highest cluster density (respectively 8 and 10 clusters) and the highest number of phenotypes showing significant variation in the WT offspring of mutant parents (respectively 174 and 141 in males and 138 and 154 in females) (Figure.3.6C).

These findings further support the strong gender dimorphism in mammalian physiology and reinforce the association between gene chromosomal location and physiological function, importantly, across generations and in a genotype-independent manner.

Importantly, these findings are in line with studies using mouse Chromosome Substitution Strains (CSS) showing how chromosome-specific information - both genetic and topological - is important for physiology, epigenetic stability and non-genetic inheritance [140, 164, 165]. We therefore sought to further explore the topology and the chromatin environment for genes on chromosomes 6 and 7, by overlapping IGE-inducing genes to publicly available Hi-C-seq signals in mouse ESCs (to avoid any tissue-specific signals) to annotate them to Topologically Associated Domains (TADs); and by performing chromatin-state segmentation analysis on a publicly available full epigenome in mouse ESCs to provide single classifications to individual loci and therefore simplify the many chromatin attributes of the locus.

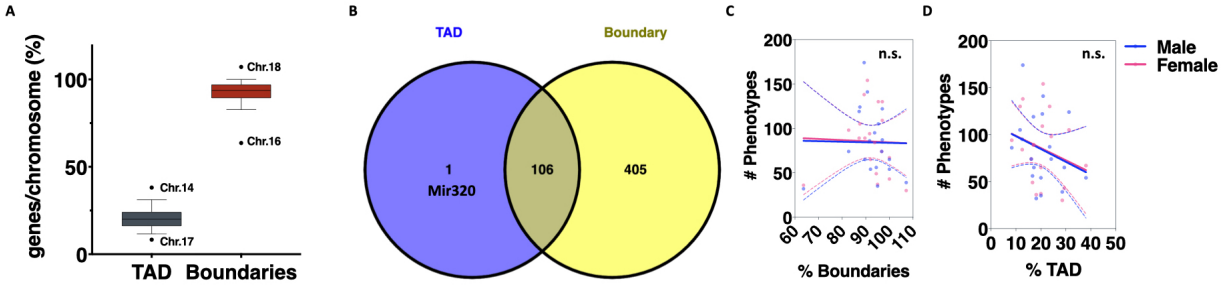


Figure 3.7: TAD boundaries overlap and association with identified IGE-inducing genes. A. Male and females genes/chromosomes(%) association with TAD boundaries. **B.** Overlap between IGEs genes identified in TAD and TAD boundaries. **C and D** Correlation between percentage of TAD and TAD boundaries with number of phenotypes in males and females.

On average 92% of the identified IGE-inducing genes overlap with TAD boundaries (Figure. 3.7A and B), which have been associated with traits heritability [153]. In contrast with these data, we could not detect any significant association between genes either located or spanning TAD boundaries with intergenerational phenotypic pleiotropy (Figure. 3.7C and D), suggesting non-canonical heritability mechanisms.

We further performed chromatin-state segmentation analysis of the 571 IGE-inducing genes across 21 histone modifications, transcription factors binding and ATAC signals. We clustered them to 25 individual states differentiating active (states 1-3), repressed (state 18) and transcribing (states 8-9) genes, as well as those targets of the three Yamanaka factors (states 22-24) (Figure. 3.8A and B). Interestingly, the vast majority of genomic bins, as well as the vast majority of IGE-inducing genes, cluster to states 16 and 17, which are not very well defined (Figure. 3.8C and D).

These findings support the previously presented data on TAD and reinforce the notion that IGEs are induced by non-canonical mechanisms of heritability.

Certain sections of the genome, known as CpG islands or CGIs, are devoid of DNA methylation and are generally located directly upstream of gene promoters . About 68%

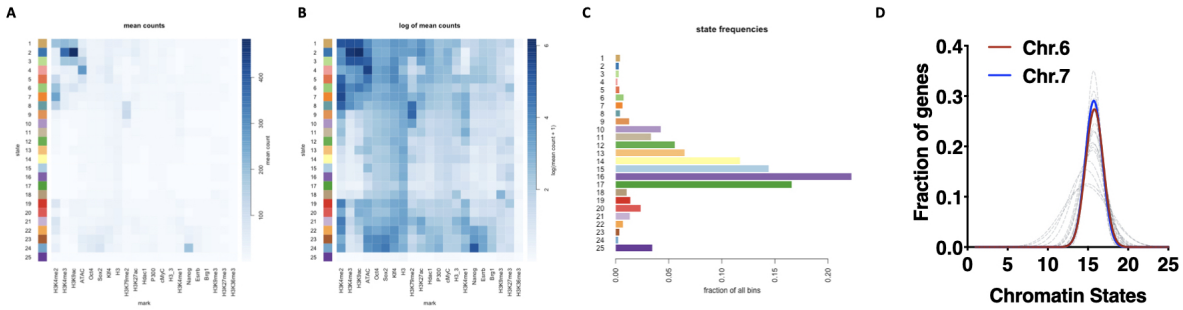


Figure 3.8: EpiCSeg segmentation analysis **A.** The dataset's average mark intensities per state. The heatmaps depict the average level of a certain mark in bins labeled with a specific state. **B.** The averages have been log-transformed for presentation reasons after adding a pseudocount of 1. **C.** Bar plot refers to a chromatin state, and its length proportional to the frequency of the state. **D.** State distribution around fractions of genes, highlighted chr 6 and 7.

of genes were found to have overlap with CpG island sites (389 out of 571). 250 genes out of 571 are found to be associated with diseases within GWA Studies.

3.5 Human Disease Association

DisGeNET is a search platform that has one of the most comprehensive publicly available libraries of genes and variations linked to human disorders. It combines information from expert-curated archives, GWAS libraries, animal models, and scientific literature. Data in DisGeNET is annotated uniformly with controlled vocabularies and community-driven ontologies. We used 26 genes that are known to be neuropeptides, and 11 genes involved in Ubiquitination pathways from 571 genelist as input for DisGeNET enrichment as performed using EnrichR. The results reveal that the neuropeptides genes are enriched in disease like obesity, depression, hypertension, anorexia etc, (Figure. 3.9) while the genes involved in Ubiquitination shows enrichment for neuromuscular diseases (Figure. 3.10). GWAS data shows 250 genes out of 571 IGEs genes are found to be associated with diseases/phenotypes.

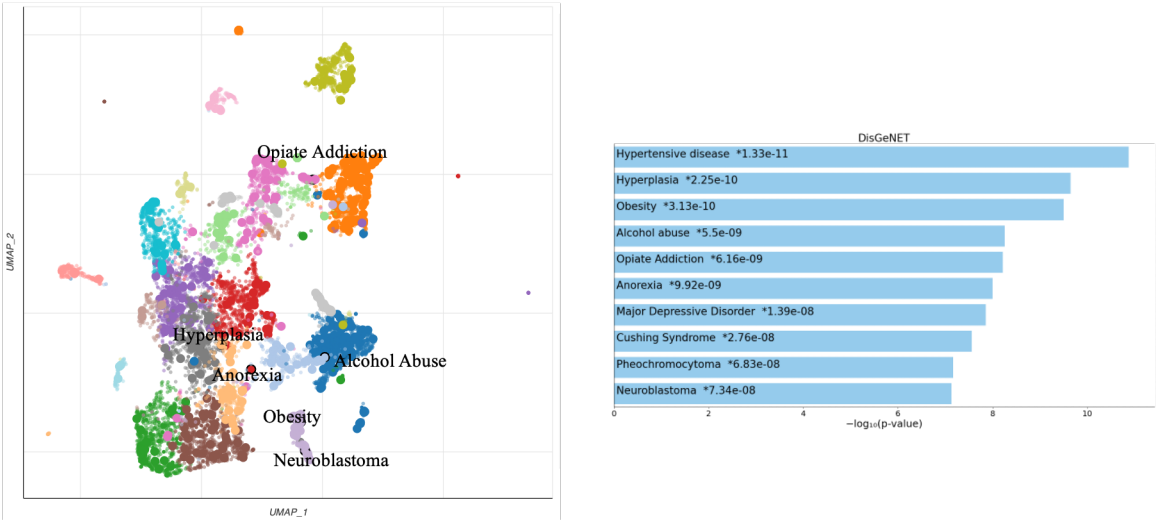


Figure 3.9: Human disease enrichment analysis for neuropeptides genes. The scatter plot visualizes gene cluster based on their similarity on the first two UMAP dimensions. The bar chart displays the top ten enriched words in the selected library, along with their p-values. Colored bars represent words with p-values less than 0.05. A phrase with an asterisk (*) next to it has a significant adjusted p-value (0.05).

3.6 Phenotype gene association network

The network is developed for 12 core phenotype categories encompassing the 5 main physiological areas highlighted before Figure. 2.2D. Interestingly, while there is evident association of terms like Exploratory behavior and anxiety (which belongs to neurobehavioral phenotypes and are quantified using the Open Field Test by the IMPC centers) as well as adiposity and Cholesterol Homeostasis, which are expected, some more interesting and unexpected associations pop up from our analysis. For example, there is an interesting association between hyperactivity and sensory motor gating (a proxy for schizophrenia-like phenotypes) through genes like trpc6, which belongs to the family of the TRP (Transient Receptor Potential) channels implicated in the pathogenesis of psychiatric disorders. Another interesting association is between hyperactivity and adiposity, through genes like Hrh1 (Histamine Receptor H1) and Atn1 (Atrophin 1).

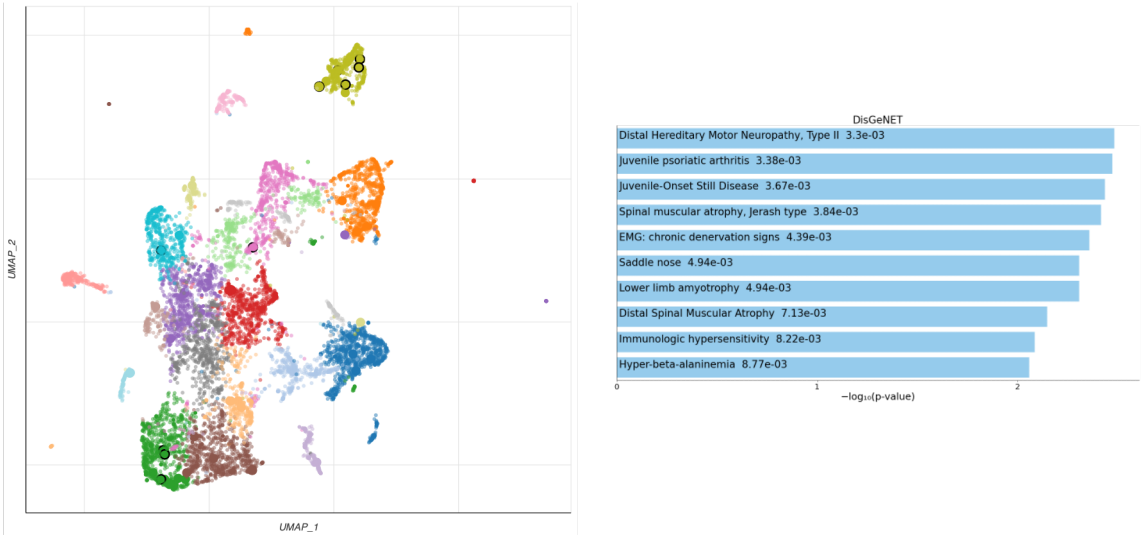


Figure 3.10: Human disease enrichment analysis for Ubiquitination pathways related genes. The scatter plot visualizes gene cluster based on their similarity in the first two UMAP dimensions. The bar chart displays the top ten enriched words in the selected library, along with their p-values. Colored bars represent words with p-values less than 0.05. A phrase with an asterisk (*) next to it has a significant adjusted p-value (0.05).

Interestingly, activity disorders in humans (such as ADHD syndrome) are strongly associated with being overweight - especially in children and young adults, most likely through rewiring of brain circuits controlling feeding behavior. Activity phenotypes are also associated with immune depression and impaired liver function through the Emc10 (ER Membrane Protein Complex subunit 10) gene, involved in angiogenesis, a process involved in a plethora of human diseases, involving developmentally programmed phenotypes (Figure. 3.11).

These findings highlight the power of our analysis and of IGEs to uncover genetic and phenotypic associations, as well as suggest (epi)genetic mechanisms by which complex phenotypes might be influenced by parental genetics, environmental exposures and gene x environment interactions.

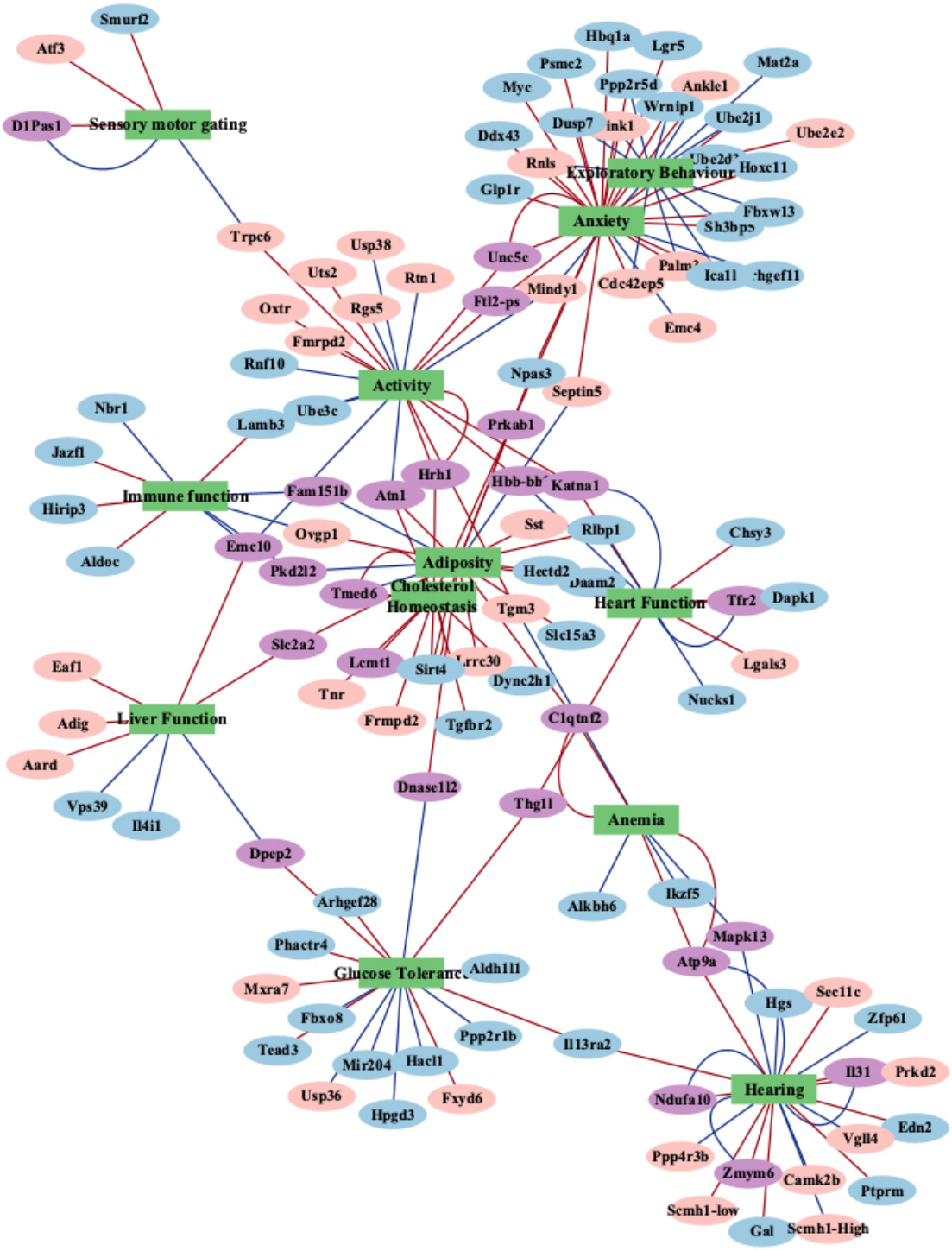


Figure 3.11: Phenotype based gene association Network The network has 128 nodes, 173 edges, and 13 multi-edge node pairs. The green square boxes are phenotypic category, ellipse shaped are the genes, blue color represent genes having phenotype in only male, pink represents female and purple show if genes have a phenotype in both male and female. The red connecting lines show increased levels, while the blue one show decreased levels.

4 Discussion

The overarching goal of this study is to investigate indirect genetic effects, i.e how parental genetics influence phenotypic trajectories in non-carrier, wild-type offspring. Indirect Genetic Effects (IGEs) are dependent on parental genetics, but independent from offspring genotype. While previously reported for canonical epigenetic modifier genes or as a result of large genetic/genomic manipulations, we aim at understanding whether Indirect Genetic Effects exist, how common they are in general mammalian genetics, and what their impact is on individual phenotypic trajectories. As such, Indirect Genetic Effects can constitute an important source of phenotypic variation in the general population and explain, at least partly, the discrepancy between the identified genetic variation and the observed phenotypic variation (a phenomenon known as the missing heritability problem). To study the general impact of Indirect Genetic Effects we looked for a multivariate, rich, and available dataset which would provide us with unbiased and systemic phenotypic data from isogenic mice, discordant for parental genetics. We found such a valuable data source in the data of the International Mouse Phenotyping Consortium (IMPC), which allowed us to study 2000 genes across 129 phenotypes and 2 genders, and identify 571 IGE-inducing genes spanning 5 main physiological areas such as glucose and lipid metabolism; neurology and behavior; immunology and hematology; cardiovascular health and bone metabolism.

4.1 IMPC as a valuable resource to study Indirect Genetic Effects

IMPC is a valuable resource for the scientific community as evident from more than 2000 peer-reviewed and published research articles to date. The systematic phenotyping of gene KO cohorts in IMPC helped the community to unravel gene function and its relevance to diseases that were not known before. To fulfill the aim of this project IMPC is a perfect resource, first it focuses on laboratory mice that are bred in a controlled environment, which are genetically controlled and have standard protocols for phenotyping, secondly, phenotyping data is available for vast categories like metabolism, cardiovascular, Immunological, clinical chemistry, neurological, etc. As every multicenter dataset, the IMPC resource has its own drawbacks. For example,the phenotypic data reveals that there is phenotypic variation between different centers, and that the data cluster well within the center but not between the centers. Another problem associated with the IMPC resource is the relatively high level of data scattering and need for data imputation. Among the different options to overcome these problems (including inter-center normalization, data scaling and extensive imputation to name some) and proceed with the data processing and analysis, we decided to take into account the differences among the centers and exploit the fact that almost every center could provide data from both wild-type mouse populations. We therefore proceeded to a center- and parameter-based analysis, which includes a Wilcoxon corrected quantile-based statistics to isolate top-varying phenotypes among the two wild-type populations and associate them to the genes mutated in their siring parents.

4.2 Potential candidates of indirect genetic effects from IMPC resource

After comparing WT and baseline animals we identified a total of 571 genes that induce phenotypic variation and these could be potential candidates for indirect genetic effects.

Notably, this number constitutes roughly of the entire starting dataset, and most likely this is due to the fact that we considered even the genes which induce phenotypic variation in single phenotypic parameters (Supplementary Figure. 6.22). Indeed, by filtering genes that induced variation in at least 2 phenotypic parameters the number of candidates shrinks down to 456 genes and 247 by filtering for at least 4 phenotypic parameters. Another option could have been to filter for effect-size (or log2FC) and only consider parameters with the strongest effect size. In both cases, though, we opted against the filtering strategy to avoid missing knowledge, for example on essential parameters for a given physiological function (eg. Area Under the Curve of a Glucose Tolerance Test - IPG_012 - for whole body glucose tolerance) or on essential parameter with intrinsically low variation (eg. the Respiratory Exchange Ratio - CAL_017).

As a functional quality control for the 571 identified genes, we tested whether they induce intergenerational genotype-independent effects (therefore matching the definition of IGEs) and therefore divided them into two categories one which induces phenotype variation in a genotype-independent manner (WThet = KO) and the other one in a genotype-dependent manner (WThet $\neq$ KO). Importantly, >90 % of the identified genes induce intergenerational genotype-independent effects, suggesting a high degree of true positive signals.

4.3 Quest for common features in IGE-inducing genes

While GWAS often neglect gene-gene interactions due to a lack of statistical power to identify them, mouse chromosomal substitution strains (CSSs) offer an alternative strategy for discovering epistasis due to their low allelic variation. CSS is a tool for identifying quantitative trait loci (QTL) associated with disease and phenotypes. It is a unique model for studying genetic architecture of complex traits and uncovering phenotypic hot spots on the chromosomes. It seems that different chromosomes are enriched for specific phenotypes. QTL on chromosome 6 regulates the onset of puberty in mice and chromosome 7 is linked to organ weight and obesity. CSS-10, -11, -6, and -Y show elevated PPI and decreased PPI in CSS-4 [165, 166, 167, 168, 169]. Our

results also show chromosomes 6 and 7 as hotspots for phenotypes and non-genetic programming of phenotypic trajectories across generations. Furthermore and in keeping with CSS studies, our data also show phenotypic clustering to specific chromosomes (Figure. 3.3D), importantly for the first time across generations. Genome topology has been lately associated with transcriptional regulation, gene function and disease etiology.

While these two pieces of evidence would suggest genomic/structural determinants for IGEs, data from TAD and EpiCSeg do not highlight any common genomic/structural features for IGEs. In particular, TAD data and Epicseg analysis on mESCs chromatin/histone marks show that most of the genes are present in not well-defined chromatin domains and span TAD boundaries, suggesting that they are constitutively expressed in mESCs (and most likely in adult tissues) and are essential for development. No particular divergence for genes located to chromosome 6 and 7 are identified. TAD and Epic therefore do not show any specific genomic cue that could help explain the observed IGEs and suggest non-canonical mechanisms of heritability.

Altogether, these data would suggest that what determines the likelihood and the direction of IGEs is the function of the gene itself and the signaling pathway(s) it is involved in. Indeed, the 571 genes we identified cluster to 3 discrete pathways: neuropeptide signaling, ubiquitin and mRNA surveillance. Interestingly, genes clustering on Chr.6 are involved in Neuropeptide Signaling, while those on Chr.7 are enriched for immune function (which includes components of both mRNA surveillance and ubiquitin pathways)

5 Conclusion & Outlook

The study developed an unbiased, systematic, and phenotype-based strategy to investigate the role of parental genetics in phenotypic diversity and adaptability in wild-type progeny. It established and identified a certain degree of indirect genetic effects (intergenerational phenotypes).

Key points from the study:

- With such controlled and guided phenotyping protocols certain degree of phenotypic variation is observed between the centers.
- Strong gender dimorphism in mammalian physiology
- We identified significant phenotypic variation in wild-type offspring of 571 mutants across several using available, high-quality, and systemic mouse phenotype data.
- The identified genes show functional clustering to Neuroactive genes, genes involved in protein catabolism (through ubiquitination), mRNA metabolism.
- Genomic location reveals an association between gene chromosomal location and physiological
- IGEs are induced by non-canonical mechanisms of heritability.
- Strong interconnections between induced phenotypes and gene functionality

The findings of this intensive characterization effort reveal new information about the candidate genes and their ability to generate indirect genetic effects and alter animal physiology over generations. Furthermore, while not providing conclusive mechanistic proof, the results of the molecular profiling will define a comprehensive dataset of

molecular signatures associated with the parental contribution to offspring variation in phenotypes and a potential starting point for mechanistic dissection of indirect genetic effects in mammals. Definitely, it opens more questions than answers, but answering these questions will help us understand a different aspect of inheritance.

It will interesting to do in-depth phenotyping to isolate parent-of-origin effects. One aspect of IMPC production is that the phenotyping cohorts (mutant animals and their wild-type littermates) are created solely from heterozygous breeders. This means that for this project, any parental influence might be supplied by either parent or arise from a combined parental contribution. To thoroughly explore the molecular foundations, it will be necessary to first identify parent-of-origin specific contributions to the reported parental effects.

Another aspect to be looked upon is to dissect germline-dependent or independent transmission of phenotypes. We know gametes are the primary, but not the only, means of information transmission from parents to offspring. In vitro fertilization (IVF) tests are the greatest technique to disentangle germline dependent and independent effects since they exclude non-gametic components at conception (mainly seminal fluid and maternal tract variables). Starting with a list of potential genes, constructing experimental wild-type cohorts via IVF and assess their respective phenotype observed in our data, with the parent(s) primarily contributing to the offspring phenotype.

The incidence of complex diseases like diabetes, obesity and neurodevelopmental disorders among others is on the rise worldwide and part of this is due to acquired (epigenetic) inheritance. The obtained information will shed light on the contribution of indirect genetic effects to disease susceptibility and epigenetic inheritance.

6 Supplementary Figures

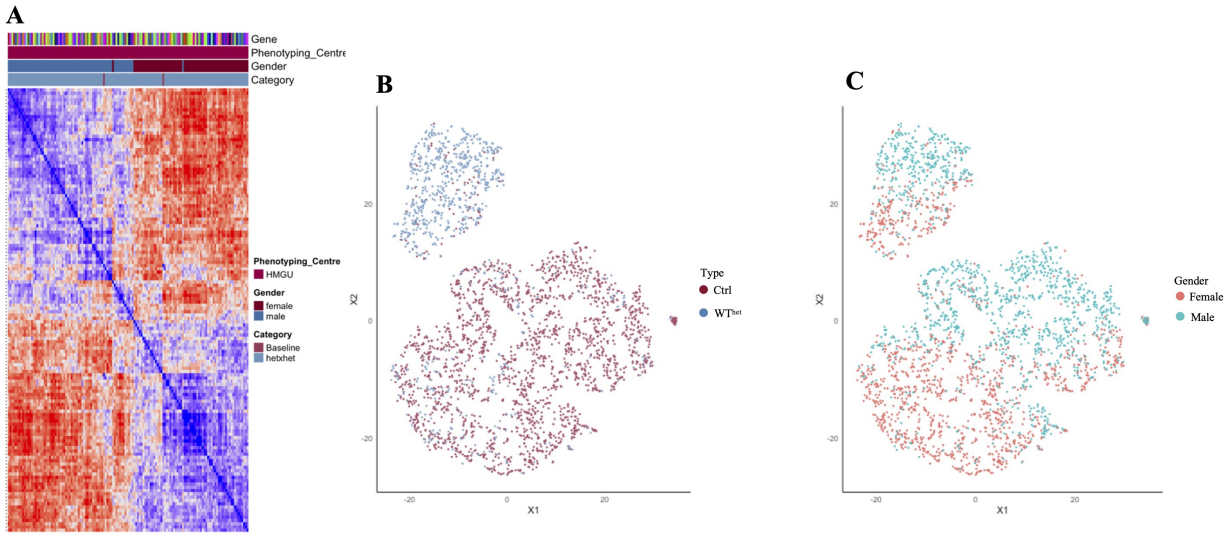


Figure 6.1: Correlation and PCA based analysis of HMGU phenotyping center A. The heatmap matrix shows the Spearman correlation coefficient between animals for all the genes across 129 phenotyping parameters. **B**. Gender based distribution of animals shown as a t-SNE plot both WT and baseline animals **C**. t-SNE plot for visualizing WT and Baseline animals distribution in a projected 2D metric map.

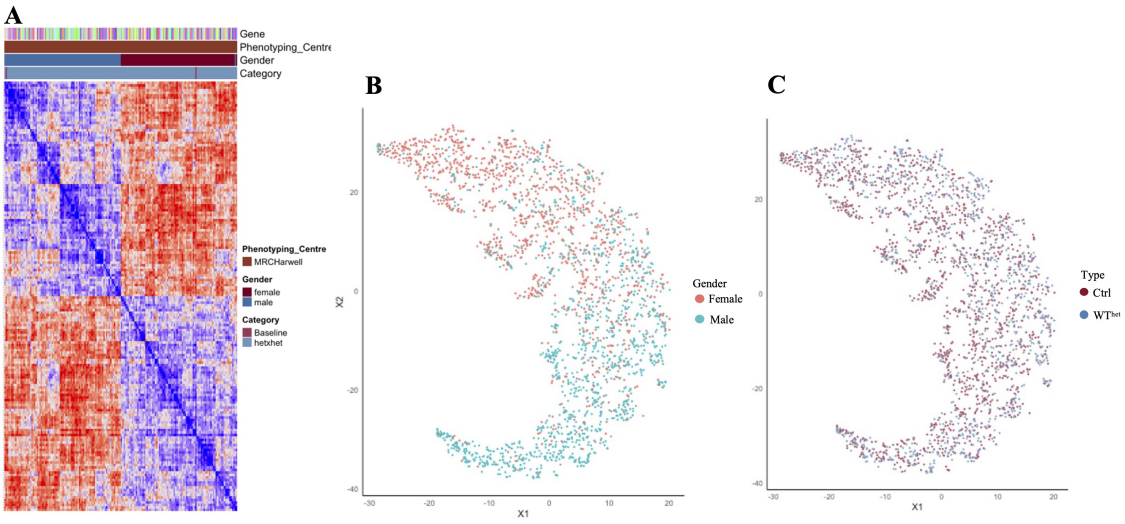


Figure 6.2: Correlation and PCA based analysis of MRC Harwell phenotyping center

A. The heatmap matrix shows the Spearman correlation coefficient between animals for all the genes across 129 phenotyping parameters. **B**. Gender based distribution of animals shown as a t-SNE plot both WT and baseline animals **C**. t-SNE plot for visualizing WT and Baseline animals distribution in a projected 2D metric map.

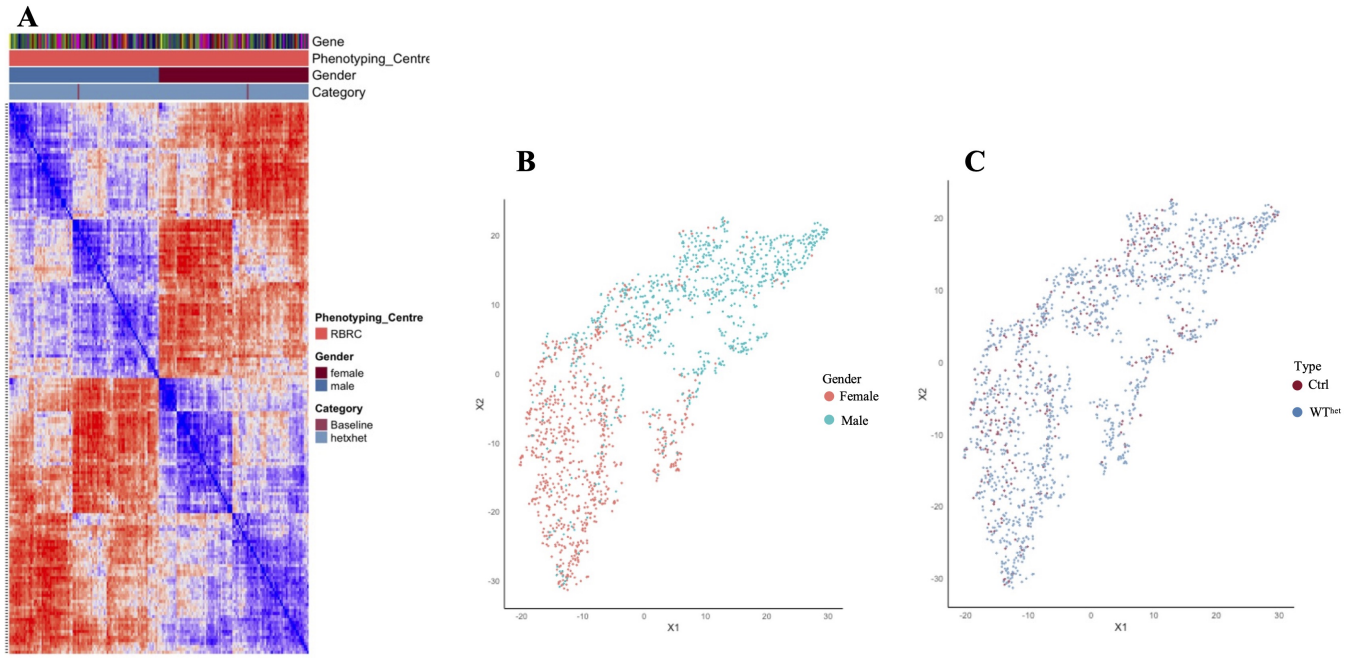


Figure 6.3: Correlation and PCA based analysis of RBRC phenotyping center A. The heatmap matrix shows the Spearman correlation coefficient between animals for all the genes across 129 phenotyping parameters. **B**. Gender based distribution of animals shown as a t-SNE plot both WT and baseline animals **C**. t-SNE plot for visualizing WT and Baseline animals distribution in a projected 2D metric map.

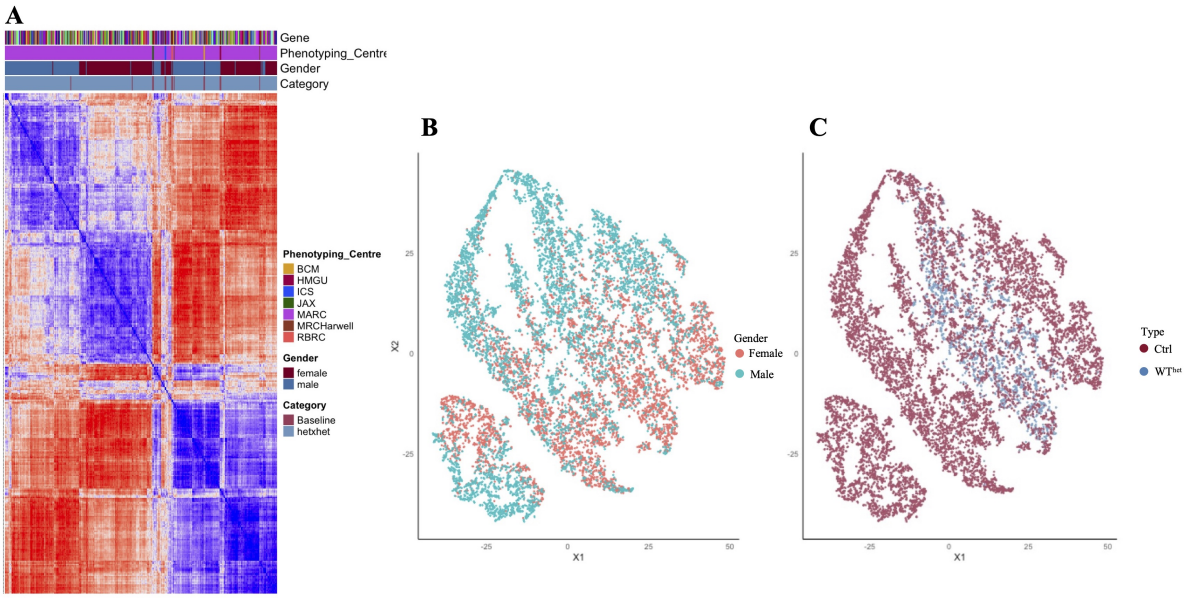


Figure 6.4: Correlation and PCA based analysis of MARC phenotyping center A. The heatmap matrix shows the Spearman correlation coefficient between animals for all the genes across 129 phenotyping parameters. **B**. Gender based distribution of animals shown as a t-SNE plot both WT and baseline animals **C**. t-SNE plot for visualizing WT and Baseline animals distribution in a projected 2D metric map.

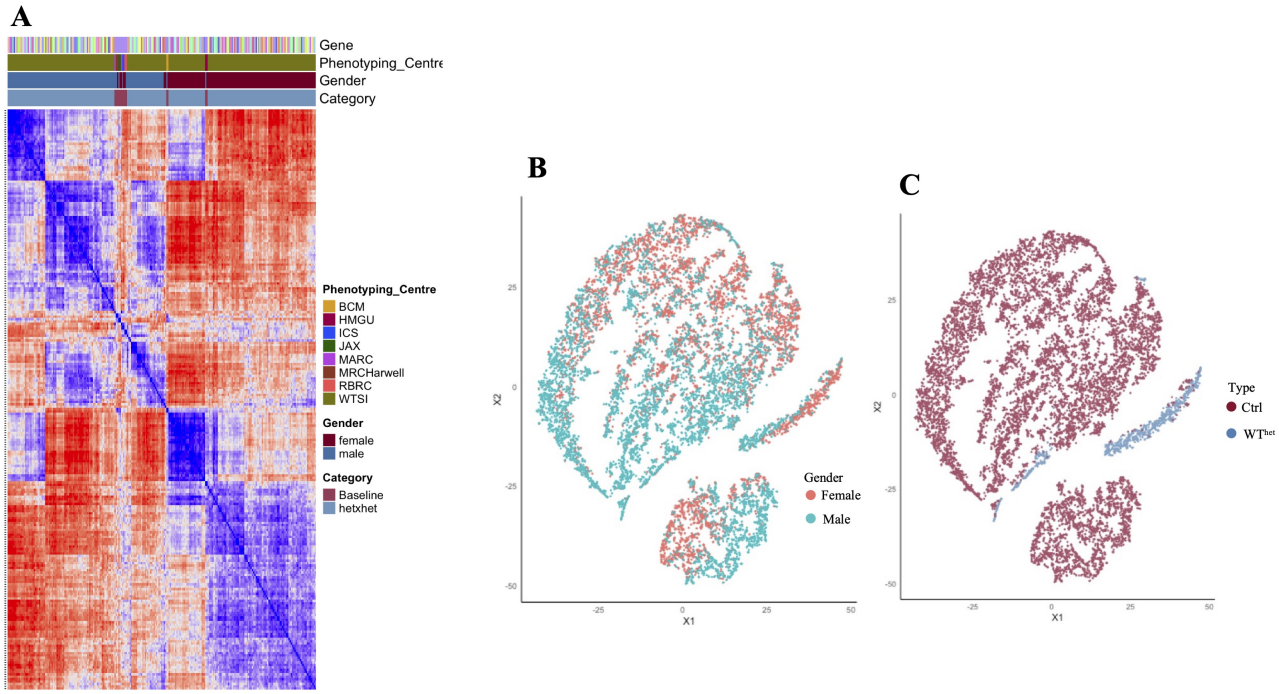


Figure 6.5: Correlation and PCA based analysis of WTSI phenotyping center A. The heatmap matrix shows the Spearman correlation coefficient between animals for all the genes across 129 phenotyping parameters. **B**. Gender based distribution of animals shown as a t-SNE plot both WT and baseline animals **C**. t-SNE plot for visualizing WT and Baseline animals distribution in a projected 2D metric map.

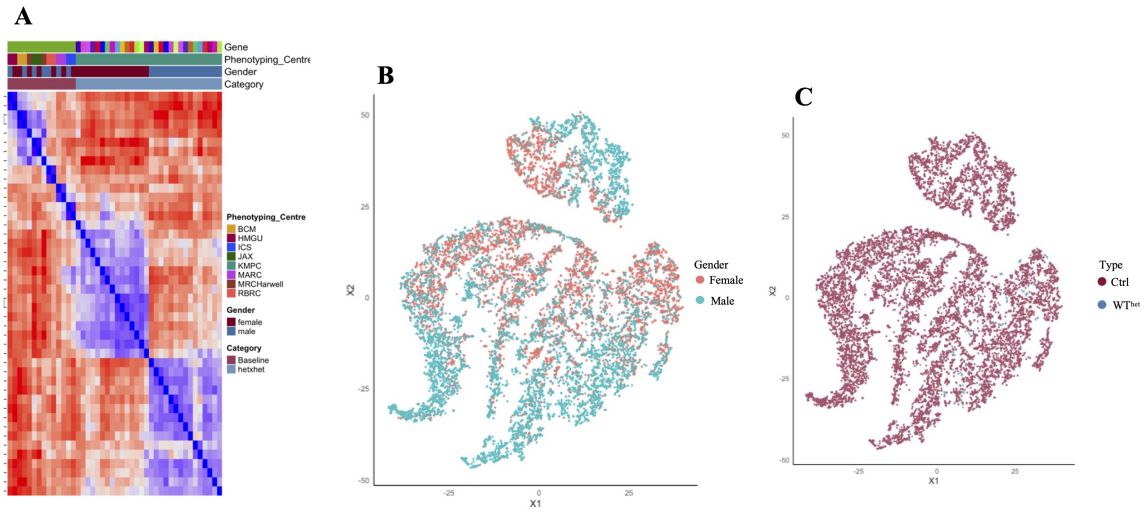


Figure 6.6: Correlation and PCA based analysis of KMPC phenotyping center A. The heatmap matrix shows the Spearman correlation coefficient between animals for all the genes across 129 phenotyping parameters. **B**. Gender based distribution of animals shown as a t-SNE plot both WT and baseline animals **C**. t-SNE plot for visualizing WT and Baseline animals distribution in a projected 2D metric map.

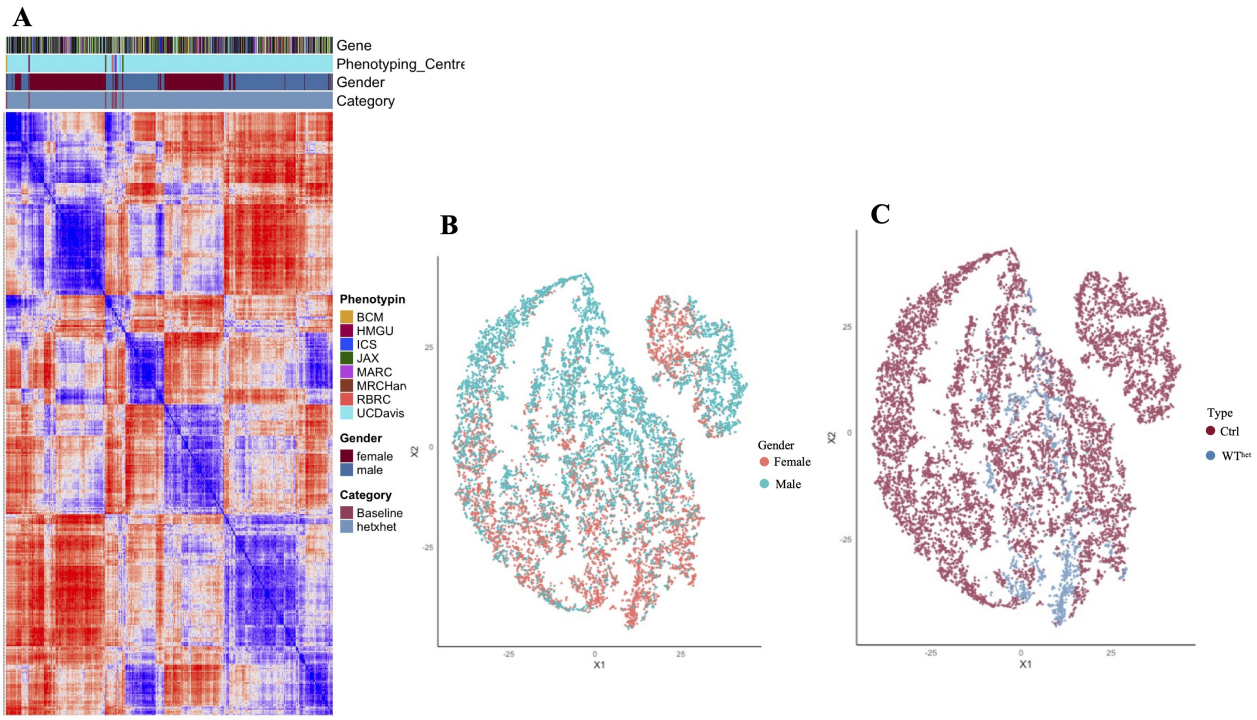


Figure 6.7: Correlation and PCA based analysis of UCDavis phenotyping center A. The heatmap matrix shows the Spearman correlation coefficient between animals for all the genes across 129 phenotyping parameters. **B**. Gender based distribution of animals shown as a t-SNE plot both WT and baseline animals **C**. t-SNE plot for visualizing WT and Baseline animals distribution in a projected 2D metric map.

Figure 6.8: Density plots showing the distribution of Auditory Brain stem response parameters(female:red; male:blue).

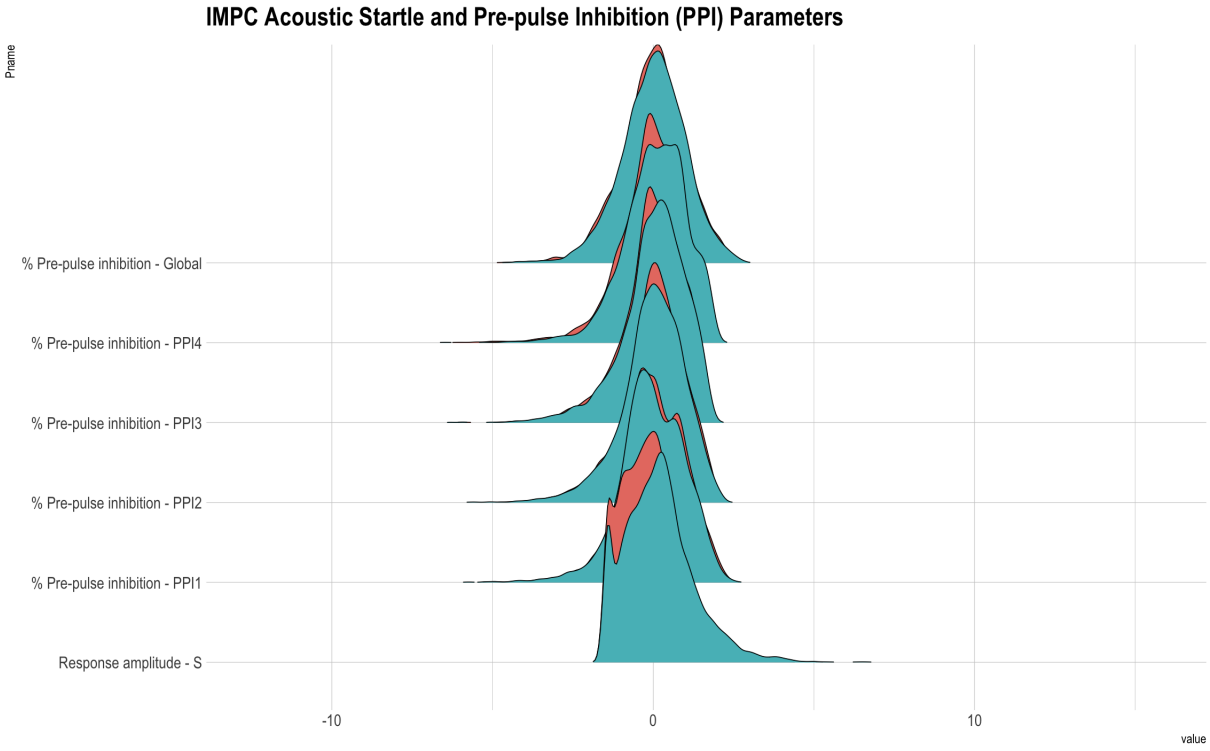


Figure 6.9: Density plots showing the distribution of Acoustic Startle and Pre-pulse Inhibition (PPI) parameters(female:red; male:blue).

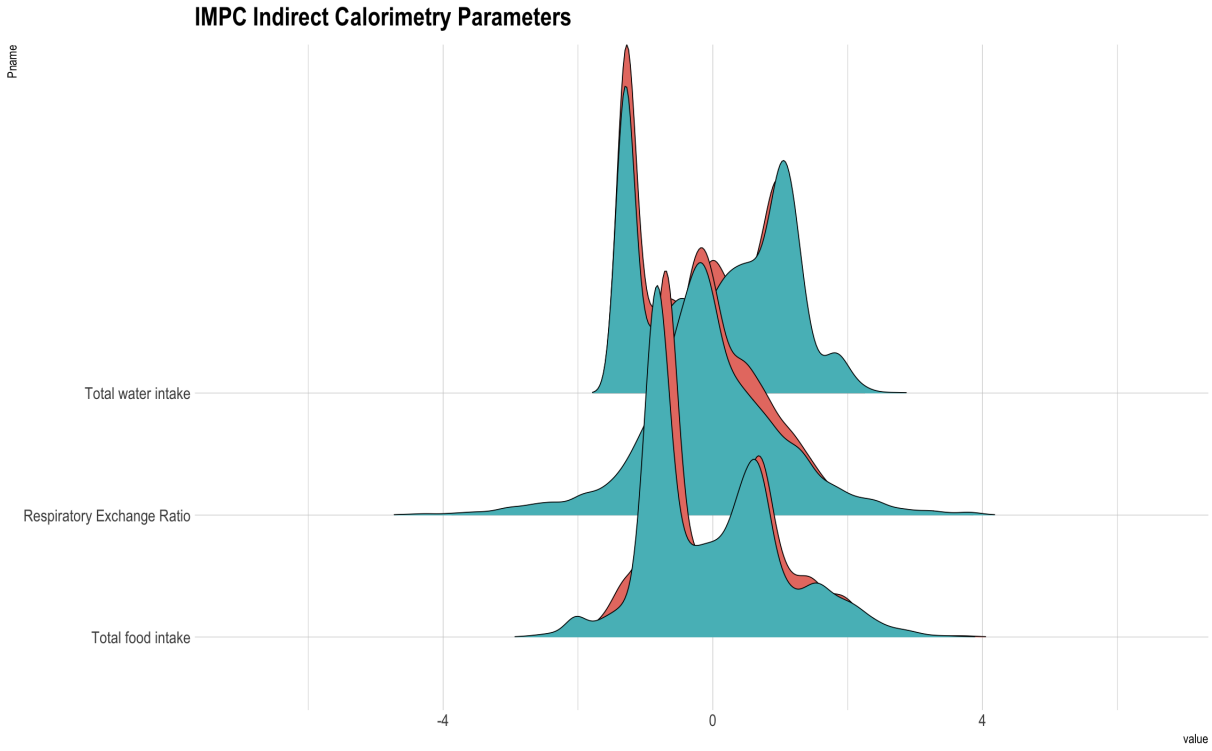


Figure 6.10: Density plots showing the distribution of Indirect Calorimetry parameters(female:red; male:blue).

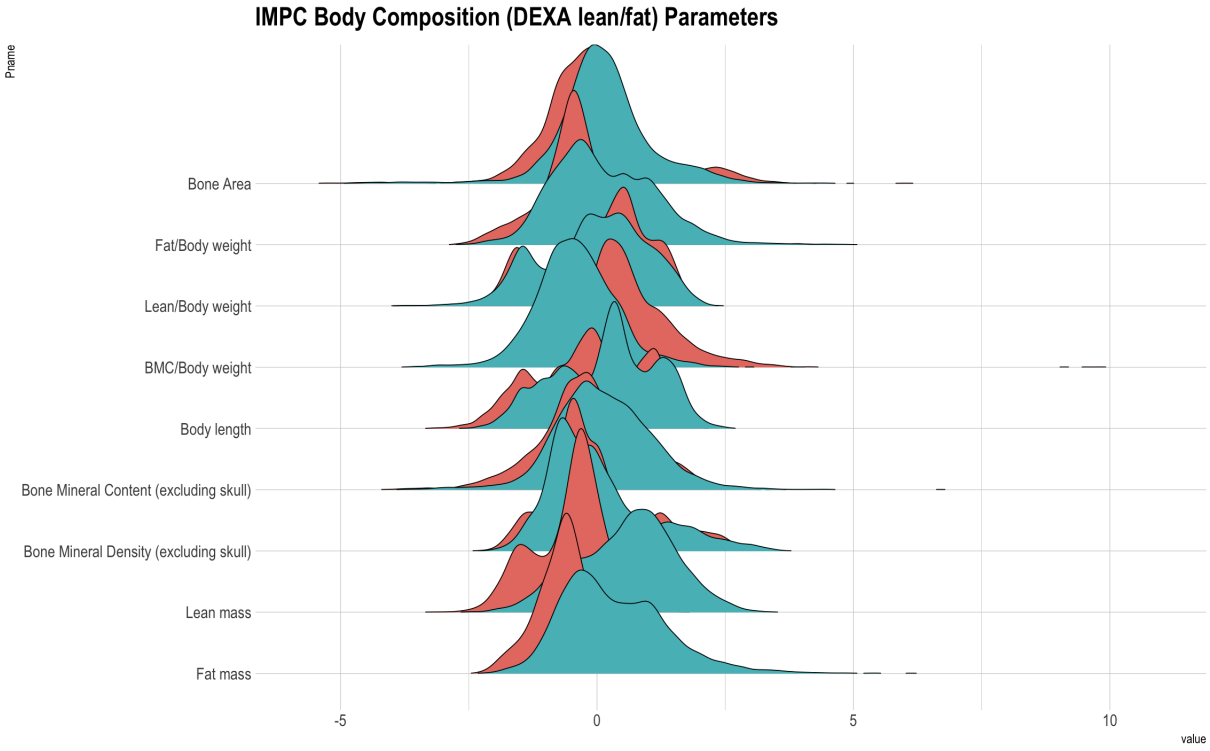


Figure 6.11: Density plots showing the distribution of Body Composition (DEXA lean/fat)parameters(female:red; male:blue).

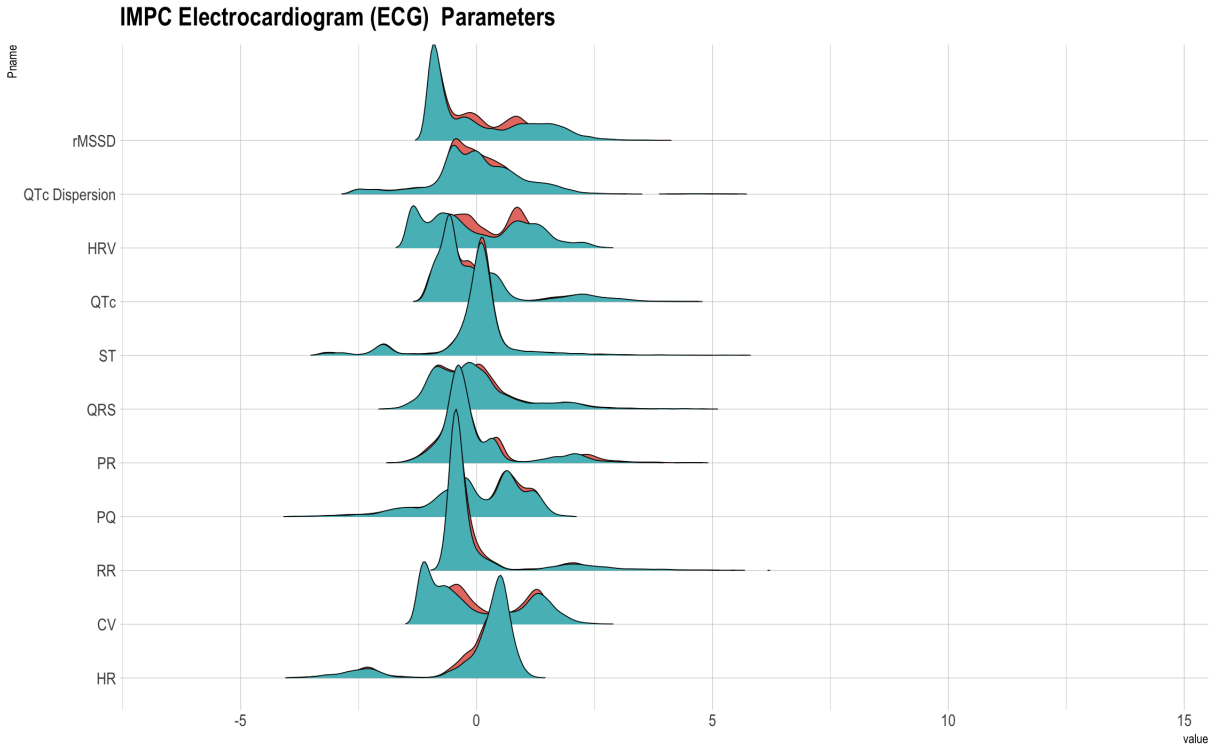


Figure 6.12: Density plots showing the distribution of Electrocardiogram (ECG) parameters(female:red; male:blue).

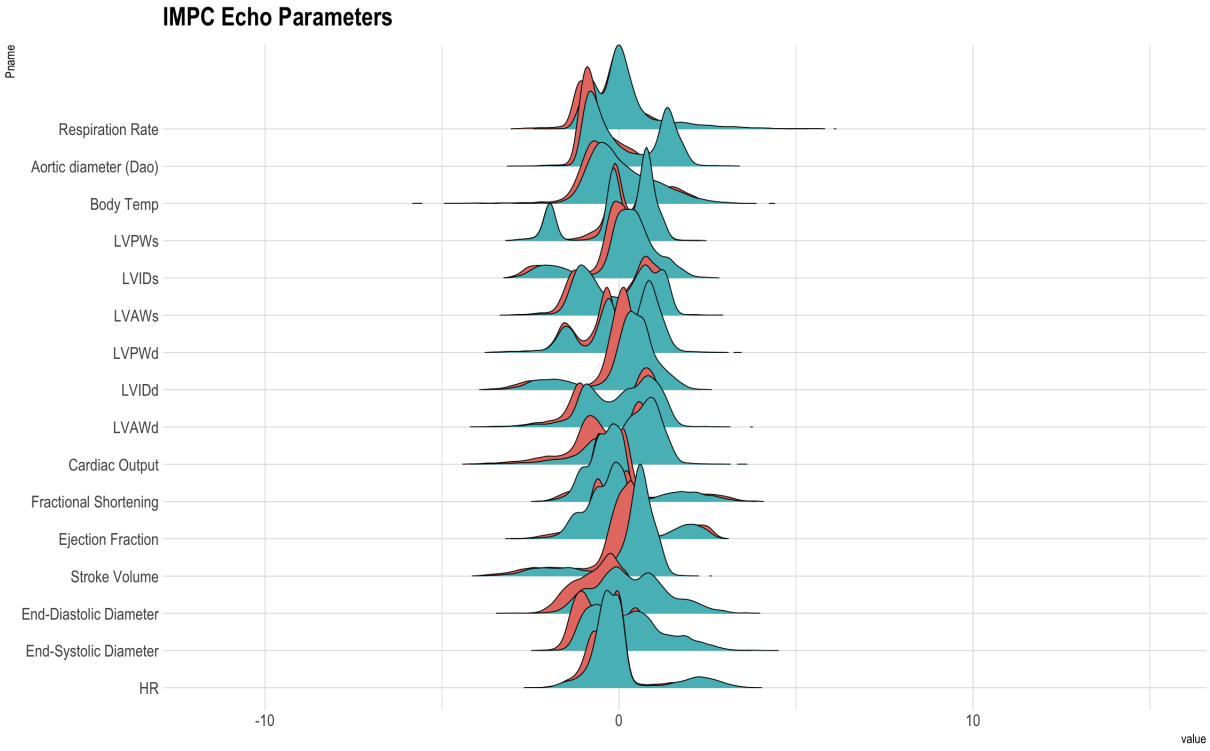


Figure 6.13: Density plots showing the distribution of ECHO parameters(female:red; male:blue).

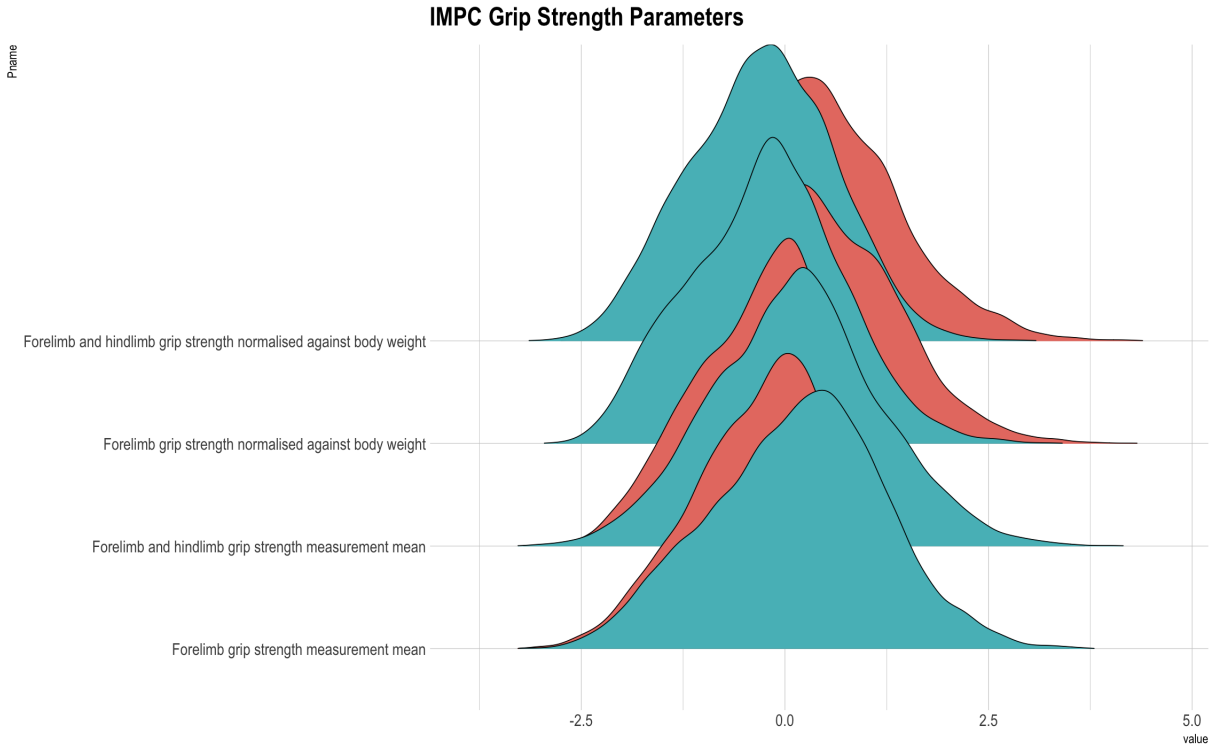


Figure 6.14: Density plots showing the distribution of Grip Strength parameters(female:red; male:blue).

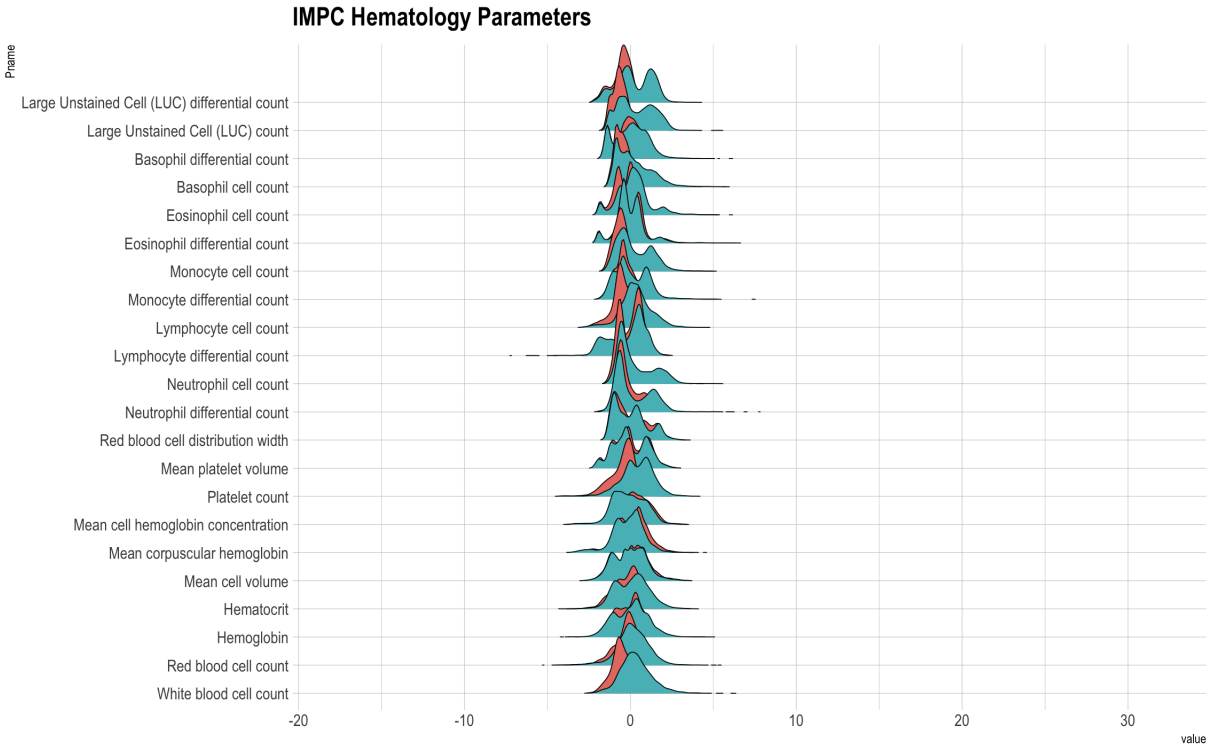


Figure 6.15: Density plots showing the distribution of Hematology parameters(female:red; male:blue).

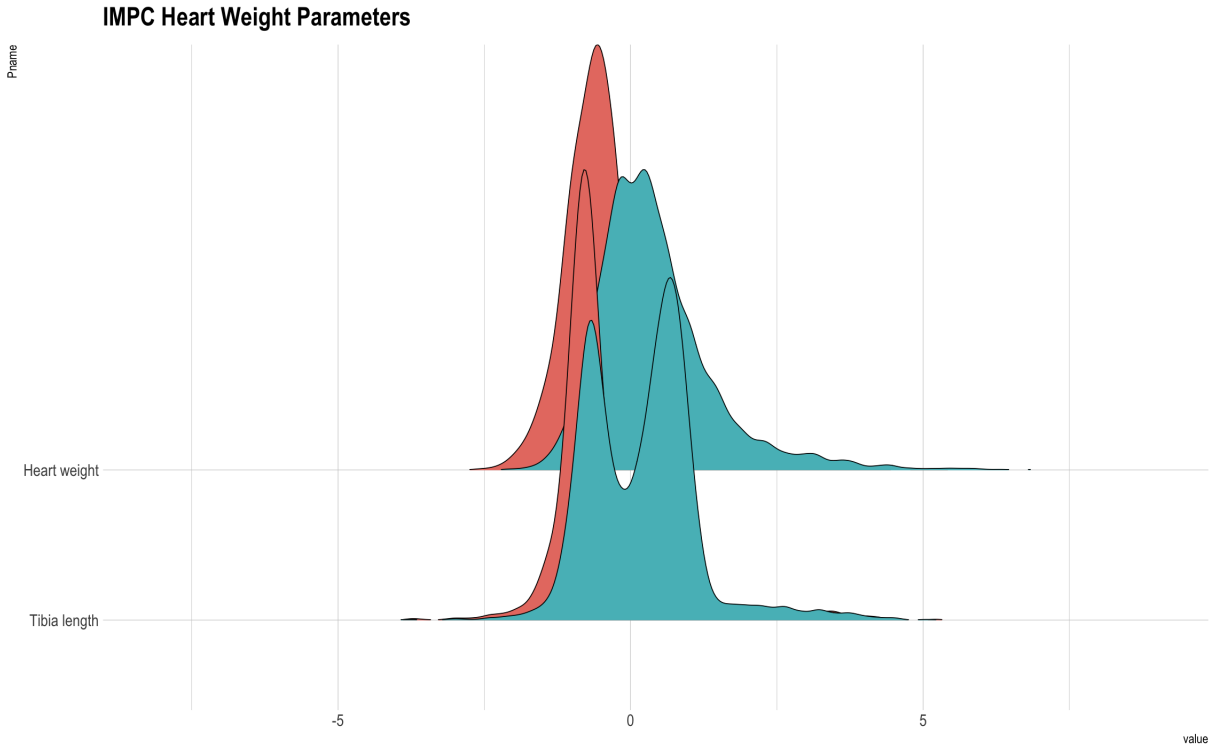


Figure 6.16: Density plots showing the distribution of Heart Weight parameters(female:red; male:blue).

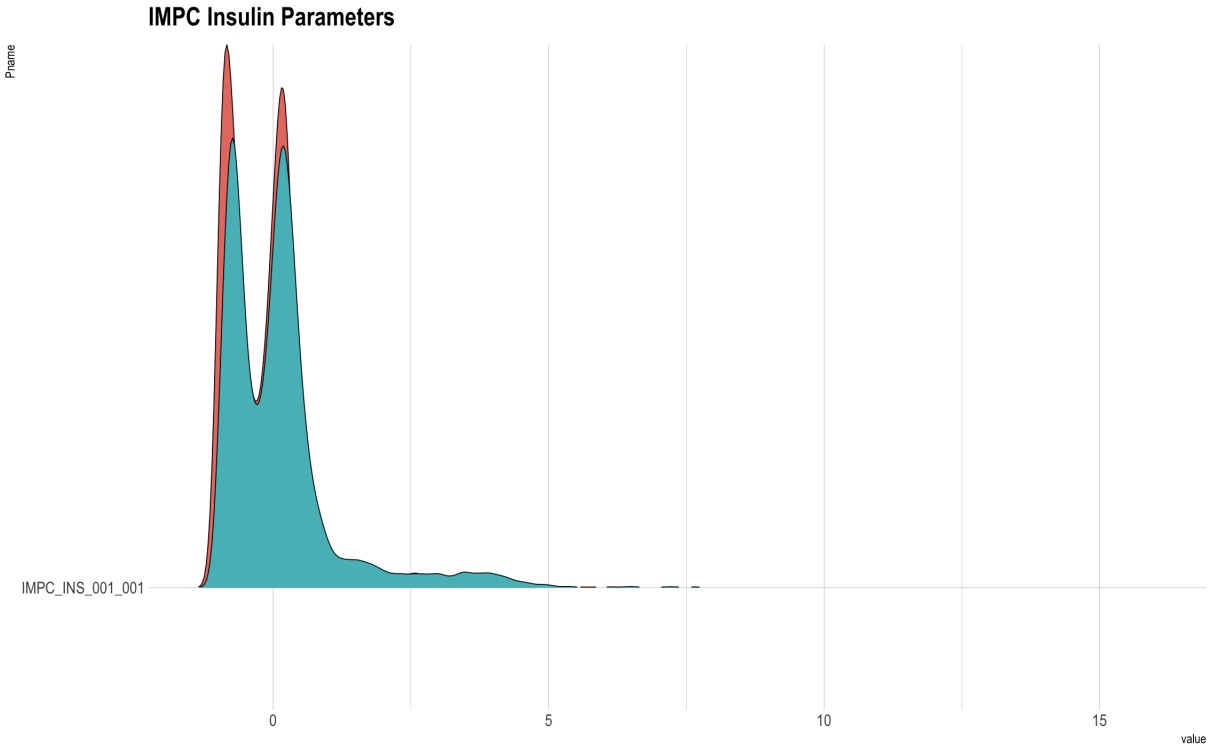


Figure 6.17: Density plots showing the distribution of Insulin parameters(female:red; male:blue).

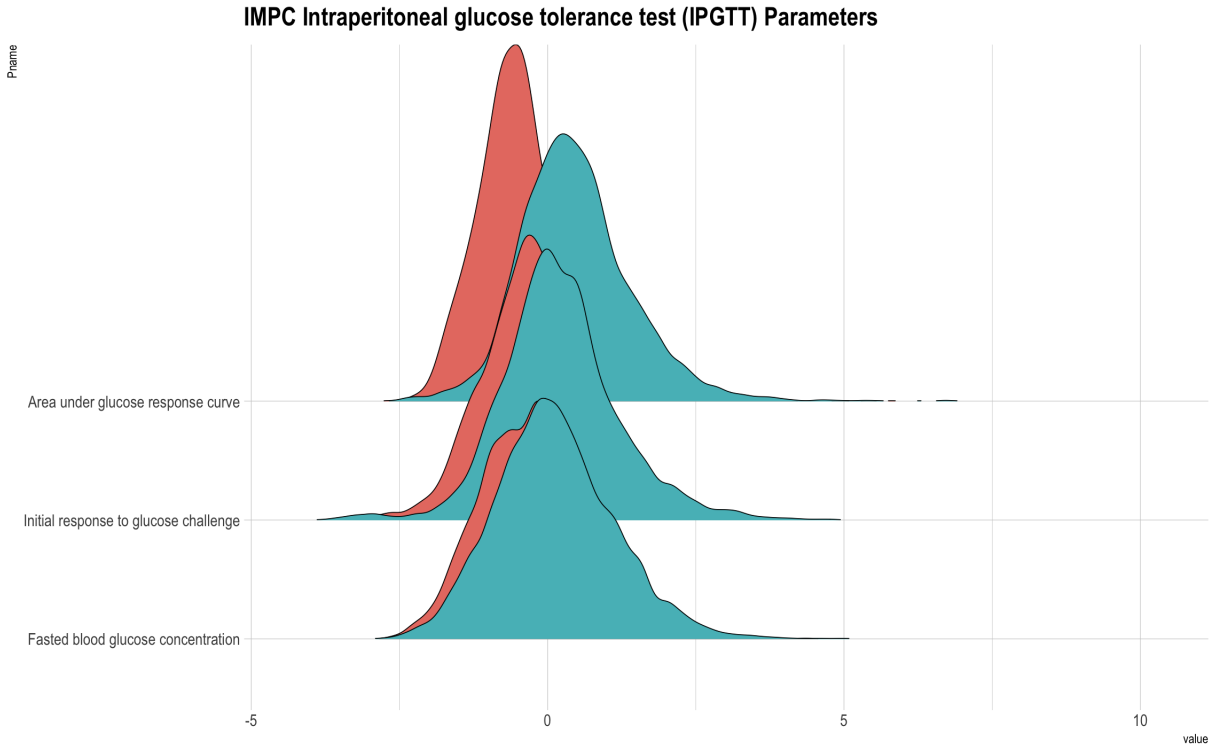


Figure 6.18: Density plots showing the distribution of Intraperitoneal glucose tolerance test (ipGTT) parameters(female:red; male:blue).

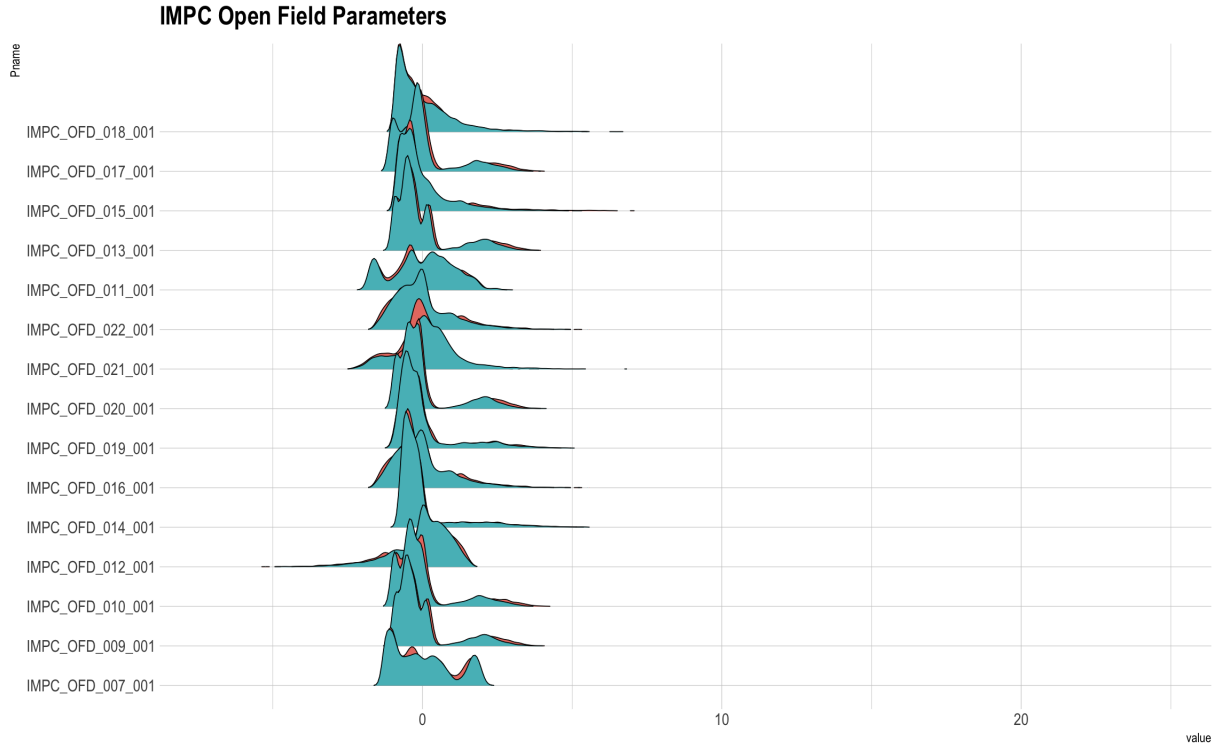
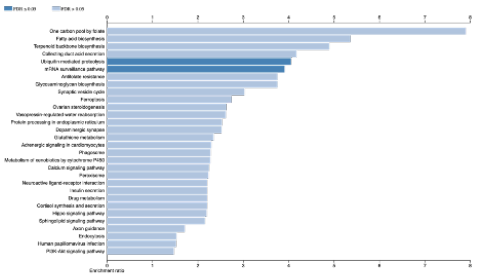


Figure 6.19: Density plots showing the distribution of Open Field parameters(female:red; male:blue).



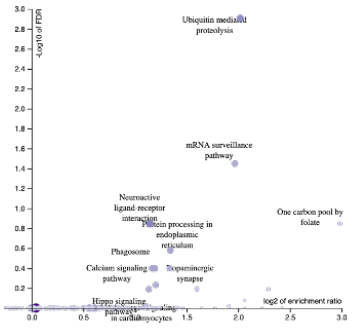


Figure 6.20: webgesalt analysis results

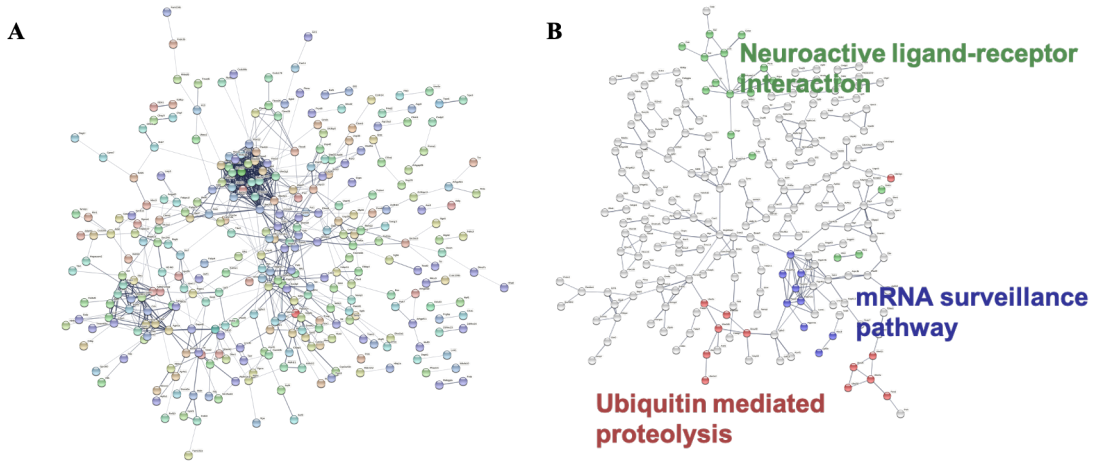


Figure 6.21: STRING Network analysis A. Network of 571 candidate genes. **B.** Highlighted significant KEGG pathway terms.

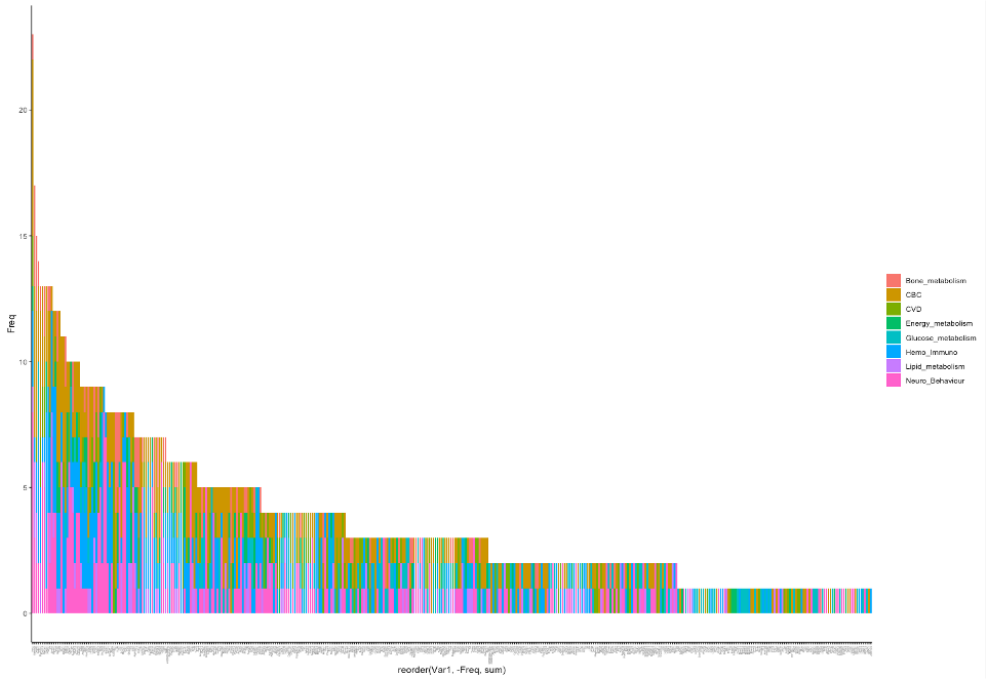


Figure 6.22: Bar plot for occurrence of phenotype term per gene.

List of Figures

1.1 **Graphical representation of Chromatin organization and epigenetic regulation.** A. Chromatin is the subunit that creates a chromosome, which is made up of DNA and protein. B. DNA methylation on the promoter can lead to switching ON or OFF of the gene transcription. 2

1.2 **Systematic representation of sequence of events that characterize acquired epigenetic inheritance.** Phenotypes acquired by the parental generation as a result of exposure to various environmental stressors are detected by the soma and produce epimutations in the germ cells, which are passed onto the offspring via fertilization and dictate their developmental and phenotypic trajectories[15]. 4

1.3 **Diagrammatic representation of Intergenerational vs. Transgenerational epigenetic Inheritance** 5

1.4 **Schematic representation of three general mechanisms of epigenetic inheritance.**Taken from [18] 7

1.5 **Epigenetic reprogramming in early mammalian development** During preimplantation development, epigenetic reprogramming erases gametic epigenetic patterns, allowing the embryo to generate an epigenetic profile appropriate for early development and crucial for the developing conceptus [21]. 8

1.6 **Possible mechanisms of Missing heritability.** This diagram depicts the alleged mechanism of missing heritability. The proportion of each ingredient will vary between common disorders. [18]. 13

1.7 **The image shows how imprinted genes are inherited across the generations** (From Book Encyclopedia of Animal Cognition and Behavior). 15

1.8 **A summary of imprinted genes and their functions in the brain and placenta during embryogenesis. [76]** 16

1.9 **Kit paramutation in mice** Male heterozygous mice with one Kit^{tm1Alf} allele and one normal wild-type (+) allele exhibit a speckled white tail tip and generate aberrant Kit mRNAs from the paramutagenic Kit^{tm1Alf} allele. These are packed in sperm and transferred to the embryo once the egg is fertilized. Progeny have the heterozygous father's wild-type Kit allele, but action of the transmitted aberrant RNAs still gives rise to the spotted tail and perpetuates its production, allowing paramutation of the wild-type Kit allele and continued transmission of the spotted tail. 18

1.10 **Diagram depicting environmental changes and the effects on paternally inherited non-genetic contributions** (http://studentblogs.med.ed.ac.uk/repro systems-group-2/mechanisms-of-epigenetics/) 21

1.11 **Studies on Non-genetic inheritance in mice. A.** Non-Mendelian inheritance of mouse paramutations. **B.** The study design examined the effect of the paternal Y chromosome on the phenotypes of daughters. **C.** Study shows Smarca5 and Dnmt1, modifiers of epigenetic reprogramming. Isogenic mice with the agouti viable yellow allele have a variety of coat colors (yellow, mottled and agouti). **D.** The disruption of histone methylation in developing sperm caused by KDM1A transgene overexpression from a single generation substantially affected embryo development and child survival. [130] **E.** Model for intergenerational epigenetic inheritance following the deletion of Kdm6a in the male germline [131] 25

1.12 **Diagrammatic representation of Indirect Genetic Effects (IGEs).** This diagram depicts an example of what we mean by indirect genetic influences. Heterozygous mutations in the parental generation create germline epimutations, which are passed to children during fertilization and shape their developmental and phenotypic trajectories independently of the inherited genotype. The increased diversity and partial penetrance in offspring phenotypes is an intriguing aspect of IGEs. 29

1.13 **Model of the study to identify genetic determinants of Indirect Genetic Effects(IGEs).** 32

2.1 **IMPRESS phenotyping pipeline** The IMPC fundamental pipelines illustrate the consortium-wide phenotyping workflow. 34

2.2 **Representation of IMPC Data used in this study. A.** Graphic representation of the breeding scheme of Baseline (Ctrl) animals; and C57BL/6N animals with a high degree of inbreeding derived from mutant parents that are heterozygous, named wildtype (WT) animals. **B.** List of the phenotypic centers which contributed data for this study. **C.** Geographical distribution of the phenotyping centers and their respective contribution of cohorts (WT, Baseline or Both) for the study. **D.** Representation of five broad physiological categories based on 129 phenotyping parameters used in the study. 35

2.3 **Detailed strategy for identification of candidate genes responsible for IGEs. A.** Schematic diagram of the methodology followed in the study, **B.** Table shows center based comparison of Baseline and WT animals. **C.** Quantile based analysis, where WT animals which fall above quantile 90 (Q90) (shown in blue within the red outline) and below quantile 10 (Q10) are selected. **D.** Selected candidates further confirmed if they showed statistically significant differences (wilcoxon test) between the Baseline and WT animals. 38

2.4 **Identification of genotype dependant and independent phenotypes.** Comparison between WT and HOM (homozygous KO mutant) offspring of the same heterozygous mating to differentiate between gene-dependent and genotype-independent candidates. 39

2.5 **Steps to identify genes for Phenotype based gene interaction Network. A.** Phenotype-based co-correlation matrix is a Spearman-based co_correlation matrix based on imputed phenotype dataset from WT and Baseline animals. **B.** IMPC_IPG_012_001 (AUCipgtt) is the core phenotypic term for glucose metabolism, phenotype-based Spearman co-correlation was calculated to identify relevant genes. **C.** AUCipgtt shows negative correlation with chlorine levels. **D.** wheres shows positive correlation with initial response to glucose. 45

3.1 **Correlation analysis among animals shows center-based phenotypic clustering A.** The heatmap matrix shows the Spearman correlation coefficient between animals for all the genes across 129 phenotyping parameters. Samples cluster by phenotypic centers. Colors (blues) represent relationships between samples that are most similar; warmer colors (reds) represent samples that are more dissimilar with lower coefficients. **B**. t-SNE plot for both WT and baseline animals **C**. t-SNE plot for visualizing only the Baseline animals clusters in a projected 2D metric map. 48

3.2 **Sexual Dimorphism and WT, Baseline animals distribution. A.** Sex-specific distribution of animals shown as a t-SNE plot. **B.** Sex-based distribution of HMGU phenotypic centers. **C.** WT and Baseline animals distribution shown as a t-SNE plot, **D.** WT and Baseline animals distribution of HMGU phenotypic center. 49

3.3 **Distribution of 571 candidate genes of IGEs A.** Venn Diagram of candidate genes to show overlap between male and female animals. **B. C.** Alluvial plot shows male and female dynamics of the candidate genes of IGEs. It shows candidates' genes distribution through phenotyping center, phenotype category and phenotype levels to be high or low. Each line represent a single gene connects it to phenotyping center to the respective phenotype categories and levels. 50

3.4 **Functional annotation of A. male and B. female gene set that are not significantly different between WT and HOM phenotypes from the identified 571 candidate genes.** 51

3.5 **Function annotation of A. male and B. female 571 candidate genes identified of IGEs.** 53

3.6 **Genomic location assessment of 571 candidate genes of IGEs. A.** Ideogram illustrating distribution and localization of 571 candidate genes along mouse chromosomes. The position of each candidate gene is represented with different shapes and colors based on their sex and relative position on its respective autosome and sexual chromosomes. **B.** Gene cluster distribution on the mouse chromosomes. **C.** Number of clusters on chromosome correlation with number of phenotypes. Chr 6 and 7 shows the highest correlation. **D.** Heatmap shows chromosome wise phenotypic enrichment. 54

3.7 **TAD boundaries overlap and association with identified IGE-inducing genes. A.** Male and females genes/chromosomes(%) association with TAD boundaries. **B.** Overlap between IGEs genes identified in TAD and TAD boundaries. **C and D** Correlation between percentage of TAD and TAD boundaries with number of phenotypes in males and females. 56

3.8 **EpiCSeg segmentation analysis A.** The dataset's average mark intensities per state. The heatmaps depict the average level of a certain mark in bins labeled with a specific state. **B.** The averages have been log-transformed for presentation reasons after adding a pseudocount of 1. **C.** Bar plot refers to a chromatin state, and its length proportional to the frequency of the state. **D.** State distribution around fractions of genes, highlighted chr 6 and 7. 57

3.9 **Human disease enrichment analysis for neuropeptides genes.** The scatter plot visualizes gene cluster based on their similarity on the first two UMAP dimensions. The bar chart displays the top ten enriched words in the selected library, along with their p-values. Colored bars represent words with p-values less than 0.05. A phrase with an asterisk (*) next to it has a significant adjusted p-value (0.05). 58

3.10 **Human disease enrichment analysis for Ubiquitination pathways related genes.** The scatter plot visualizes gene cluster based on their similarity in the first two UMAP dimensions. The bar chart displays the top ten enriched words in the selected library, along with their p-values. Colored bars represent words with p-values less than 0.05. A phrase with an asterisk (*) next to it has a significant adjusted p-value (0.05). 59

3.11 **Phenotype based gene association Network** The network has 128 nodes, 173 edges, and 13 multi-edge node pairs. The green square boxes are phenotypic category, ellipse shaped are the genes, blue color represent genes having phenotype in only male, pink represents female and purple show if genes have a phenotype in both male and female. The red connecting lines show increased levels, while the blue one show decreased levels. 60

6.1 **Correlation and PCA based analysis of HMGU phenotyping center A.** The heatmap matrix shows the Spearman correlation coefficient between animals for all the genes across 129 phenotyping parameters. **B**. Gender based distribution of animals shown as a t-SNE plot both WT and baseline animals **C**. t-SNE plot for visualizing WT and Baseline animals distribution in a projected 2D metric map. 67

6.2 **Correlation and PCA based analysis of MRC Harwell phenotyping center A.** The heatmap matrix shows the Spearman correlation coefficient between animals for all the genes across 129 phenotyping parameters. **B**. Gender based distribution of animals shown as a t-SNE plot both WT and baseline animals **C**. t-SNE plot for visualizing WT and Baseline animals distribution in a projected 2D metric map. 68

6.3 **Correlation and PCA based analysis of RBRC phenotyping center A.** The heatmap matrix shows the Spearman correlation coefficient between animals for all the genes across 129 phenotyping parameters. **B**. Gender based distribution of animals shown as a t-SNE plot both WT and baseline animals **C**. t-SNE plot for visualizing WT and Baseline animals distribution in a projected 2D metric map. 69

6.4 **Correlation and PCA based analysis of MARC phenotyping center A.** The heatmap matrix shows the Spearman correlation coefficient between animals for all the genes across 129 phenotyping parameters. **B**. Gender based distribution of animals shown as a t-SNE plot both WT and baseline animals **C**. t-SNE plot for visualizing WT and Baseline animals distribution in a projected 2D metric map. 70

6.5 **Correlation and PCA based analysis of WTSI phenotyping center A.** The heatmap matrix shows the Spearman correlation coefficient between animals for all the genes across 129 phenotyping parameters. **B**. Gender based distribution of animals shown as a t-SNE plot both WT and baseline animals **C**. t-SNE plot for visualizing WT and Baseline animals distribution in a projected 2D metric map. 71

6.6 **Correlation and PCA based analysis of KMPC phenotyping center A.** The heatmap matrix shows the Spearman correlation coefficient between animals for all the genes across 129 phenotyping parameters. **B**. Gender based distribution of animals shown as a t-SNE plot both WT and baseline animals **C**. t-SNE plot for visualizing WT and Baseline animals distribution in a projected 2D metric map. 72

6.7 **Correlation and PCA based analysis of UCDavis phenotyping center A.** The heatmap matrix shows the Spearman correlation coefficient between animals for all the genes across 129 phenotyping parameters. **B**. Gender based distribution of animals shown as a t-SNE plot both WT and baseline animals **C**. t-SNE plot for visualizing WT and Baseline animals distribution in a projected 2D metric map. 73

6.8 **Density plots showing the distribution of Auditory Brain stem response parameters(female:red; male:blue).** 74

6.9 **Density plots showing the distribution of Acoustic Startle and Pre-pulse Inhibition (PPI) parameters(female:red; male:blue).** . 75

6.10 **Density plots showing the distribution of Indirect Calorimetry parameters(female:red; male:blue).** 76

6.11 **Density plots showing the distribution of Body Composition (DEXA lean/fat)parameters(female:red; male:blue).** 77

6.12 **Density plots showing the distribution of Electrocardiogram (ECG) parameters(female:red; male:blue).** 78

6.13 **Density plots showing the distribution of ECHO parameters(female:red; male:blue).** 79
6.14 **Density plots showing the distribution of Grip Strength parameters(female:red; male:blue).** 80
6.15 **Density plots showing the distribution of Hematology parameters(female:red; male:blue).** 81
6.16 **Density plots showing the distribution of Heart Weight parameters(female:red; male:blue).** 82
6.17 **Density plots showing the distribution of Insulin parameters(female:red; male:blue).** 83
6.18 **Density plots showing the distribution of Intraperitoneal glucose tolerance test (ipGTT) parameters(female:red; male:blue).** . . . 84
6.19 **Density plots showing the distribution of Open Field parameters(female:red; male:blue).** 85
6.20 **webgesalt analysis results** 85
6.21 **STRING Network analysis A.** Network of 571 candidate genes. **B.** Highlighted significant KEGG pathway terms. 86
6.22 **Bar plot for occurrence of phenotype term per gene.** 86

List of Tables

2.1 **Details of number of animals collected from different phenotyping centers.** 36

2.2 **Details of the mESCs Chip-seq and ATAC-seq sequencing data downloaded from NCBI.** 43

3.1 **Cluster Locator results shows number and size of clusters found and number and percentage of genes on the list forming clusters.** 55

Bibliography

[1] M. Lane, R. L. Robker, and S. A. Robertson. Parenting from before conception. *Science*, 345(6198):756–760, 2014.

[2] O. J. Rando. Daddy issues: paternal effects on phenotype. *Cell*, 151(4):702–708, 2012.

[3] L. M. Anderson, L. Riffle, R. Wilson, G. S. Travlos, M. S. Lubomirski, and W. G. Alvord. Preconceptional fasting of fathers alters serum glucose in offspring of mice. *Nutrition*, 22(3):327–31, 2006.

[4] B. R. Carone, L. Fauquier, N. Habib, J. M. Shea, C. E. Hart, R. Li, C. Bock, C. Li, H. Gu, P. D. Zamore, A. Meissner, Z. Weng, H. A. Hofmann, N. Friedman, and O. J. Rando. Paternally induced transgenerational environmental reprogramming of metabolic gene expression in mammals. *Cell*, 143(7):1084–96, 2010.

[5] J. C. Jimenez-Chillaron, E. Isganaitis, M. Charalambous, S. Gesta, T. Pentinat-Pelegrin, R. R. Faucette, J. P. Otis, A. Chow, R. Diaz, A. Ferguson-Smith, and M. E. Patti. Intergenerational transmission of glucose intolerance and obesity by in utero undernutrition in mice. *Diabetes*, 58(2):460–468, 2009.

[6] S. F. Ng, R. C. Lin, D. R. Laybutt, R. Barres, J. A. Owens, and M. J. Morris. Chronic high-fat diet in fathers programs beta-cell dysfunction in female rat offspring. *Nature*, 467(7318):963–6, 2010.

[7] T. L. Bale. Epigenetic and transgenerational reprogramming of brain development. *Nat Rev Neurosci*, 16(6):332–44, 2015.

[8] K. Gapp, A. Jawaid, P. Sarkies, J. Bohacek, P. Pelczar, J. Prados, L. Farinelli, E. Miska, and I. M. Mansuy. Implication of sperm rnas in transgenerational inheritance of the effects of early trauma in mice. *Nature Neuroscience*, 17(5):667–+, 2014.

[9] A. B. Rodgers, C. P. Morgan, N. A. Leu, and T. L. Bale. Transgenerational epigenetic programming via sperm microrna recapitulates effects of paternal stress. *Proc Natl Acad Sci U S A*, 112(44):13699–704, 2015.

[10] M. D. Anway, A. S. Cupp, M. Uzumcu, and M. K. Skinner. Epigenetic transgenerational actions of endocrine disruptors and male fertility. *Science*, 308(5727):1466–9, 2005.

[11] M. P. Vallaster, S. Kukreja, X. Y. Bing, J. Ngolab, R. Zhao-Shea, P. D. Gardner, A. R. Tapper, and O. J. Rando. Paternal nicotine exposure alters hepatic xenobiotic metabolism in offspring. *Elife*, 6, 2017.

[12] F. M. Vassoler, S. L. White, H. D. Schmidt, G. Sadri-Vakili, and R. C. Pierce. Epigenetic inheritance of a cocaine-resistance phenotype. *Nat Neurosci*, 16(1):42–7, 2013.

[13] M. Zeybel, T. Hardy, Y. K. Wong, J. C. Mathers, C. R. Fox, A. Gackowska, F. Oakley, A. D. Burt, C. L. Wilson, Q. M. Anstee, M. J. Barter, S. Masson, A. M. Elsharkawy, D. A. Mann, and J. Mann. Multigenerational epigenetic adaptation of the hepatic wound-healing response. *Nat Med*, 18(9):1369–77, 2012.

[14] M. K. Skinner, M. Ben Maamar, I. Sadler-Riggleman, D. Beck, E. Nilsson, M. McBirney, R. Klukovich, Y. Xie, C. Tang, and W. Yan. Alterations in sperm dna methylation, non-coding rna and histone retention associate with ddt-induced epigenetic transgenerational inheritance of disease. *Epigenetics Chromatin*, 11(1):8, 2018.

[15] A. Tomar and R. Teperino. Genetic control of non-genetic inheritance in mammals: state-of-the-art and perspectives. *Mamm Genome*, 31(5-6):146–156, 2020.

[16] A. C. Antoniou and D. F. Easton. Models of genetic susceptibility to breast cancer. *Oncogene*, 25(43):5898–905, 2006.

[17] J. J. Tuscher and J. J. Day. Multigenerational epigenetic inheritance: One step forward, two generations back. *Neurobiol Dis*, 132:104591, 2019.

[18] T. H. Akbaba, E. Sag, B. Balci-Peynircioglu, and S. Ozen. Epigenetics for clinicians from the perspective of pediatric rheumatic diseases. *Curr Rheumatol Rep*, 22(8):46, 2020.

[19] O. N. Tikhodeyev. Heredity determined by the environment: Lamarckian ideas in modern molecular biology. *Sci Total Environ*, page 135521, 2019.

[20] Y. Zeng and T. Chen. Dna methylation reprogramming during mammalian development. *Genes (Basel)*, 10(4), 2019.

[21] J. Huntriss. Chapter 6 - epigenetic reprogramming in the embryo. In T. Tollefsbol, editor, *Epigenetics and Reproductive Health*, volume 21 of *Translational Epigenetics*, pages 97–116. Academic Press, 2021. URL: `https://www.sciencedirect.com/science/article/pii/B9780128197530000064`, `doi:https://doi.org/10.1016/B978-0-12-819753-0.00006-4`.

[22] A. Sen, N. Heredia, M. C. Senut, S. Land, K. Hollocher, X. Lu, M. O. Dereski, and D. M. Ruden. Multigenerational epigenetic inheritance in humans: Dna methylation changes associated with maternal exposure to lead can be transmitted to the grandchildren. *Sci Rep*, 5:14466, 2015.

[23] C. Huang, M. Xu, and B. Zhu. Epigenetic inheritance mediated by histone lysine methylation: maintaining transcriptional states without the precise restoration of marks? *Philosophical Transactions of the Royal Society B-Biological Sciences*, 368(1609), 2013.

[24] S. Rea, F. Eisenhaber, D. O'Carroll, B. D. Strahl, Z. W. Sun, M. Schmid, S. Opravil, K. Mechtler, C. P. Ponting, C. D. Allis, and T. Jenuwein. Regula-

tion of chromatin structure by site-specific histone h3 methyltransferases. *Nature*, 406(6796):593–9, 2000.

[25] M. Lachner, D. O'Carroll, S. Rea, K. Mechtler, and T. Jenuwein. Methylation of histone h3 lysine 9 creates a binding site for hp1 proteins. *Nature*, 410(6824):116–20, 2001.

[26] A. J. Bannister, P. Zegerman, J. F. Partridge, E. A. Miska, J. O. Thomas, R. C. Allshire, and T. Kouzarides. Selective recognition of methylated lysine 9 on histone h3 by the hp1 chromo domain. *Nature*, 410(6824):120–4, 2001.

[27] R. Cao, L. Wang, H. Wang, L. Xia, H. Erdjument-Bromage, P. Tempst, R. S. Jones, and Y. Zhang. Role of histone h3 lysine 27 methylation in polycomb-group silencing. *Science*, 298(5595):1039–43, 2002.

[28] B. Czermin, R. Melfi, D. McCabe, V. Seitz, A. Imhof, and V. Pirrotta. Drosophila enhancer of zeste/esc complexes have a histone h3 methyltransferase activity that marks chromosomal polycomb sites. *Cell*, 111(2):185–196, 2002.

[29] J. Muller, C. M. Hart, N. J. Francis, M. L. Vargas, A. Sengupta, B. Wild, E. L. Miller, M. B. O'Connor, R. E. Kingston, and J. A. Simon. Histone methyltransferase activity of a drosophila polycomb group repressor complex. *Cell*, 111(2):197–208, 2002.

[30] A. Kuzmichev, K. Nishioka, H. Erdjument-Bromage, P. Tempst, and D. Reinberg. Histone methyltransferase activity associated with a human multiprotein complex containing the enhancer of zeste protein. *Genes Development*, 16(22):2893–2905, 2002.

[31] K. Plath, J. Fang, S. K. Mlynarczyk-Evans, R. Cao, K. A. Worringer, H. Wang, C. C. de la Cruz, A. P. Otte, B. Panning, and Y. Zhang. Role of histone h3 lysine 27 methylation in x inactivation. *Science*, 300(5616):131–5, 2003.

[32] I. Okamoto, A. P. Otte, C. D. Allis, D. Reinberg, and E. Heard. Epigenetic dynamics of imprinted x inactivation during early mouse development. *Science*, 303(5658):644–9, 2004.

[33] M. Ingouff, Y. Hamamura, M. GourgueS, T. Higashiyama, and F. Berger. Distinct dynamics of histone3 variants between the two fertilization products in plants. *Current Biology*, 17(12):1032–1037, 2007.

[34] T. Okada, M. Endo, M. B. Singh, and P. L. Bhalla. Analysis of the histone h3 gene family in arabidopsis and identification of the male-gamete-specific variant atmgh3. *Plant Journal*, 44(4):557–568, 2005.

[35] S. L. Ooi and S. Henikoff. Germline histone dynamics and epigenetics. *Curr Opin Cell Biol*, 19(3):257–65, 2007.

[36] K. L. Arney, S. Bao, A. J. Bannister, T. Kouzarides, and M. A. Surani. Histone methylation defines epigenetic asymmetry in the mouse zygote. *Int J Dev Biol*, 46(3):317–20, 2002.

[37] A. Inoue and Y. Zhang. Nucleosome assembly is required for nuclear pore complex assembly in mouse zygotes. *Nature Structural Molecular Biology*, 21(7):609–616, 2014.

[38] C. Zierhut, C. Jenness, H. Kimura, and H. Funabiki. Nucleosomal regulation of chromatin composition and nuclear assembly revealed by histone depletion. *Nature Structural Molecular Biology*, 21(7):617–625, 2014.

[39] A. A. Amodeo, D. Jukam, A. F. Straight, and J. M. Skotheim. Histone titration against the genome sets the dna-to-cytoplasm threshold for the xenopus midblastula transition. *Proc Natl Acad Sci U S A*, 112(10):E1086–95, 2015.

[40] E. L. Greer, T. J. Maures, D. Ucar, A. G. Hauswirth, E. Mancini, J. P. Lim, B. A. Benayoun, Y. Shi, and A. Brunet. Transgenerational epigenetic inheritance of longevity in caenorhabditis elegans. *Nature*, 479(7373):365–71, 2011.

[41] A. Lismer, V. Dumeaux, C. Lafleur, R. Lambrot, J. Brind'Amour, M. C. Lorincz, and S. Kimmins. Histone h3 lysine 4 trimethylation in sperm is transmitted to the embryo and associated with diet-induced phenotypes in the offspring. *Dev Cell*, 56(5):671–686 e6, 2021.

[42] E. Heard and R. A. Martienssen. Transgenerational epigenetic inheritance: myths and mechanisms. *Cell*, 157(1):95–109, 2014.

[43] R. M. Alcazar, R. Lin, and A. Z. Fire. Transmission dynamics of heritable silencing induced by double-stranded rna in caenorhabditis elegans. *Genetics*, 180(3):1275–88, 2008.

[44] A. Grishok, H. Tabara, and C. C. Mello. Genetic requirements for inheritance of rnai in c. elegans. *Science*, 287(5462):2494–7, 2000.

[45] S. Yuan, A. Schuster, C. Tang, T. Yu, N. Ortogero, J. Bao, H. Zheng, and W. Yan. Sperm-borne mirnas and endo-sirnas are important for fertilization and preimplantation embryonic development. *Development*, 143(4):635–47, 2016.

[46] C. C. Conine, F. Sun, L. Song, J. A. Rivera-Perez, and O. J. Rando. Small rnas gained during epididymal transit of sperm are essential for embryonic development in mice. *Dev Cell*, 46(4):470–480 e3, 2018.

[47] Q. Chen, M. Yan, Z. Cao, X. Li, Y. Zhang, J. Shi, G. H. Feng, H. Peng, X. Zhang, Y. Zhang, J. Qian, E. Duan, Q. Zhai, and Q. Zhou. Sperm tsrnas contribute to intergenerational inheritance of an acquired metabolic disorder. *Science*, 351(6271):397–400, 2016.

[48] V. Grandjean, S. Fourre, D. A. De Abreu, M. A. Derieppe, J. J. Remy, and M. Rassoulzadegan. Rna-mediated paternal heredity of diet-induced obesity and metabolic disorders. *Sci Rep*, 5:18193, 2015.

[49] U. Sharma, C. C. Conine, J. M. Shea, A. Boskovic, A. G. Derr, X. Y. Bing, C. Belleannee, A. Kucukural, R. W. Serra, F. Sun, L. Song, B. R. Carone, E. P.

Ricci, X. Z. Li, L. Fauquier, M. J. Moore, R. Sullivan, C. C. Mello, M. Garber, and O. J. Rando. Biogenesis and function of trna fragments during sperm maturation and fertilization in mammals. *Science*, 351(6271):391–396, 2016.

[50] T. de Castro Barbosa, L. R. Ingerslev, P. S. Alm, S. Versteyhe, J. Massart, M. Rasmussen, I. Donkin, R. Sjogren, J. M. Mudry, L. Vetterli, S. Gupta, A. Krook, J. R. Zierath, and R. Barres. High-fat diet reprograms the epigenome of rat spermatozoa and transgenerationally affects metabolism of the offspring. *Mol Metab*, 5(3):184–197, 2016.

[51] L. Shen, M. Gan, Z. Tan, D. Jiang, Y. Jiang, M. Li, J. Wang, X. Li, S. Zhang, and L. Zhu. A novel class of trna-derived small non-coding rnas respond to myocardial hypertrophy and contribute to intergenerational inheritance. *Biomolecules*, 8(3), 2018.

[52] Y. S. Lee, Y. Shibata, A. Malhotra, and A. Dutta. A novel class of small rnas: trna-derived rna fragments (trfs). *Genes Dev*, 23(22):2639–49, 2009.

[53] D. F. Jarosz, J. C. S. Brown, G. A. Walker, M. S. Datta, W. L. Ung, A. K. Lancaster, A. Rotem, A. Chang, G. A. Newby, D. A. Weitz, L. F. Bisson, and S. Lindquist. Cross-kingdom chemical communication drives a heritable, mutually beneficial prion-based transformation of metabolism. *Cell*, 158(5):1083–1093, 2014.

[54] D. F. Jarosz, A. K. Lancaster, J. C. S. Brown, and S. Lindquist. An evolutionarily conserved prion-like element converts wild fungi from metabolic specialists to generalists. *Cell*, 158(5):1072–1082, 2014.

[55] Z. H. Harvey, A. K. Chakravarty, R. A. Futia, and D. F. Jarosz. A prion epigenetic switch establishes an active chromatin state. *Cell*, 180(5):928–940 e14, 2020.

[56] N. J. Schork, S. S. Murray, K. A. Frazer, and E. J. Topol. Common vs. rare allele hypotheses for complex diseases. *Curr Opin Genet Dev*, 19(3):212–9, 2009.

[57] K. Hemminki, A. Forsti, and J. L. Bermejo. The 'common disease-common variant' hypothesis and familial risks. *Plos One*, 3(6), 2008.

[58] B. Liu, M. J. Gloudemans, A. S. Rao, E. Ingelsson, and S. B. Montgomery. Abundant associations with gene expression complicate gwas follow-up. *Nat Genet*, 51(5):768–769, 2019.

[59] E. Lopez-Cortegano and A. Caballero. Inferring the nature of missing heritability in human traits using data from the gwas catalog. *Genetics*, 212(3):891–904, 2019.

[60] A. I. Young. Solving the missing heritability problem. *PLoS Genet*, 15(6):e1008222, 2019.

[61] M. Trerotola, V. Relli, P. Simeone, and S. Alberti. Epigenetic inheritance and the missing heritability. *Hum Genomics*, 9:17, 2015.

[62] V. Tam, N. Patel, M. Turcotte, Y. Bosse, G. Pare, and D. Meyre. Benefits and limitations of genome-wide association studies. *Nat Rev Genet*, 20(8):467–484, 2019.

[63] L. Duncan, H. Shen, B. Gelaye, J. Meijsen, K. Ressler, M. Feldman, R. Peterson, and B. Domingue. Analysis of polygenic risk score usage and performance in diverse human populations. *Nat Commun*, 10(1):3328, 2019.

[64] A. R. Martin, M. Kanai, Y. Kamatani, Y. Okada, B. M. Neale, and M. J. Daly. Clinical use of current polygenic risk scores may exacerbate health disparities. *Nature Genetics*, 51(4):584–591, 2019.

[65] P. Wainschtein, D. Jain, Z. Zheng, T. O. A. W. Group, N. T.-O. f. P. M. Consortium, L. A. Cupples, A. H. Shadyab, B. McKnight, B. M. Shoemaker, B. D. Mitchell, B. M. Psaty, C. Kooperberg, C. T. Liu, C. M. Albert, D. Roden, D. I. Chasman, D. Darbar, D. M. Lloyd-Jones, D. K. Arnett, E. A. Regan, E. Boerwinkle, J. I. Rotter, J. R. O'Connell, L. R. Yanek, M. de Andrade, M. A. Allison, M. N. McDonald, M. K. Chung, M. Fornage, N. Chami, N. L. Smith, P. T. Ellinor,

R. S. Vasan, R. A. Mathias, R. J. F. Loos, S. S. Rich, S. A. Lubitz, S. R. Heckbert, S. Redline, X. Guo, Y. I. Chen, C. A. Laurie, R. D. Hernandez, S. T. McGarvey, M. E. Goddard, C. C. Laurie, K. E. North, L. A. Lange, B. S. Weir, L. Yengo, J. Yang, and P. M. Visscher. Assessing the contribution of rare variants to complex trait heritability from whole-genome sequence data. *Nat Genet*, 2022.

[66] A. I. Young. Discovering missing heritability in whole-genome sequencing data. *Nat Genet*, 2022.

[67] S. C. Barton, M. A. Surani, and M. L. Norris. Role of paternal and maternal genomes in mouse development. *Nature*, 311(5984):374–6, 1984.

[68] J. McGrath and D. Solter. Completion of mouse embryogenesis requires both the maternal and paternal genomes. *Cell*, 37(1):179–83, 1984.

[69] R. J. Scott, M. Spielman, J. Bailey, and H. G. Dickinson. Parent-of-origin effects on seed development in arabidopsis thaliana. *Development*, 125(17):3329–41, 1998.

[70] D. Haig. Genomic imprinting and kinship: How good is the evidence? *Annual Review of Genetics*, 38:553–585, 2004.

[71] D. Haig and J. F. Wilkins. Genomic imprinting, sibling solidarity and the logic of collective action. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences*, 355(1403):1593–1597, 2000.

[72] T. Kawashima and F. Berger. Epigenetic reprogramming in plant sexual reproduction. *Nat Rev Genet*, 15(9):613–24, 2014.

[73] M. T. Raissig, C. Baroux, and U. Grossniklaus. Regulation and flexibility of genomic imprinting during seed development. *Plant Cell*, 23(1):16–26, 2011.

[74] A. Inoue, L. Jiang, F. Lu, T. Suzuki, and Y. Zhang. Maternal h3k27me3 controls dna methylation-independent imprinting. *Nature*, 547(7664):419–424, 2017.

[75] M. A. Cleaton, C. A. Edwards, and A. C. Ferguson-Smith. Phenotypic outcomes of imprinted gene models in mice: elucidation of pre- and postnatal functions of imprinted genes. *Annu Rev Genomics Hum Genet*, 15:93–126, 2014.

[76] R. N. Plasschaert and M. S. Bartolomei. Genomic imprinting in development, growth, behavior and stem cells. *Development*, 141(9):1805–13, 2014.

[77] W. W. Tang, T. Kobayashi, N. Irie, S. Dietmann, and M. A. Surani. Specification and epigenetic programming of the human germ line. *Nat Rev Genet*, 17(10):585–600, 2016.

[78] X. Wang and A. G. Clark. Using next-generation rna sequencing to identify imprinted genes. *Heredity (Edinb)*, 113(2):156–66, 2014.

[79] L. Santini, F. Halbritter, F. Titz-Teixeira, T. Suzuki, M. Asami, X. Ma, J. Ramesmayer, A. Lackner, N. Warr, F. Pauler, S. Hippenmeyer, E. Laue, M. Farlik, C. Bock, A. Beyer, A. C. F. Perry, and M. Leeb. Genomic imprinting in mouse blastocysts is predominantly associated with h3k27me3. *Nat Commun*, 12(1):3804, 2021.

[80] R. A. Brink, D. F. Brown, J. Kermicle, and W. H. Weyers. Locus dependence of the paramutant r phenotype in maize. *Genetics*, 45(9):1297–312, 1960.

[81] D. F. Brown and R. A. Brink. Paramutagenic action of paramutant r and r alleles in maize. *Genetics*, 45(10):1313–6, 1960.

[82] E. H. Coe. A regular and continuing conversion-type phenomenon at the b locus in maize. *Proc Natl Acad Sci U S A*, 45(6):828–32, 1959.

[83] J. B. Hollick, G. I. Patterson, J. Coe, E. H., K. C. Cone, and V. L. Chandler. Allelic interactions heritably alter the activity of a metastable maize pl allele. *Genetics*, 141(2):709–19, 1995.

[84] W. Goettel and J. Messing. Paramutagenicity of a p1 epiallele in maize. *Theor Appl Genet*, 126(1):159–77, 2013.

[85] L. V. Sidorenko and T. Peterson. Transgene-induced silencing identifies sequences involved in the establishment of paramutation of the maize p1 gene. *Plant Cell*, 13(2):319–35, 2001.

[86] R. Posner, I. A. Toker, O. Antonova, E. Star, S. Anava, E. Azmon, M. Hendricks, S. Bracha, H. Gingold, and O. Rechavi. Neuronal small rnas control behavior transgenerationally. *Cell*, 177(7):1814–1826 e15, 2019.

[87] E. H. Coe. The properties, origin, and mechanism of conversion-type inheritance at the b locus in maize. *Genetics*, 53(6):1035–63, 1966.

[88] R. A. Brink. A genetic change associated with the r locus in maize which is directed and potentially reversible. *Genetics*, 41(6):872–89, 1956.

[89] J. B. Hollick and V. L. Chandler. Epigenetic allelic states of a maize transcriptional regulatory locus exhibit overdominant gene action. *Genetics*, 150(2):891–7, 1998.

[90] E. D. Styles and R. A. Brink. The metastable nature of paramutable r alleles in maize. iv. parallel enhancement of r action in heterozygotes with r and in hemizygotes. *Genetics*, 61(4):801–11, 1969.

[91] S. M. Gross and J. B. Hollick. Multiple trans-sensing interactions affect meiotically heritable epigenetic states at the maize pl1 locus. *Genetics*, 176(2):829–839, 2007.

[92] C. L. Belele, L. Sidorenko, M. Stam, R. Bader, M. A. Arteaga-Vazquez, and V. L. Chandler. Specific tandem repeats are sufficient for paramutation-induced transgenerational silencing. *PLoS Genet*, 9(10):e1003773, 2013.

[93] J. B. Hollick. Paramutation and related phenomena in diverse species. *Nat Rev Genet*, 18(1):5–23, 2017.

[94] H. Herman, M. Lu, M. Anggraini, A. Sikora, Y. Chang, B. J. Yoon, and P. D. Soloway. Trans allele methylation and paramutation-like effects in mice. *Nat Genet*, 34(2):199–202, 2003.

[95] M. Rassoulzadegan, V. Grandjean, P. Gounon, S. Vincent, I. Gillot, and F. Cuzin. Rna-mediated non-mendelian inheritance of an epigenetic change in the mouse. *Nature*, 441(7092):469–74, 2006.

[96] V. Grandjean, P. Gounon, N. Wagner, L. Martin, K. D. Wagner, F. Bernex, F. Cuzin, and M. Rassoulzadegan. The mir-124-sox9 paramutation: Rna-mediated epigenetic control of embryonic and adult growth. *Development*, 136(21):3647–55, 2009.

[97] K. D. Wagner, N. Wagner, H. Ghanbarian, V. Grandjean, P. Gounon, F. Cuzin, and M. Rassoulzadegan. Rna induction and inheritance of epigenetic cardiac hypertrophy in the mouse. *Dev Cell*, 14(6):962–9, 2008.

[98] H. C. Lee, W. Gu, M. Shirayama, E. Youngman, J. Conte, D., and C. C. Mello. C. elegans pirnas mediate the genome-wide surveillance of germline transcripts. *Cell*, 150(1):78–87, 2012.

[99] A. Ashe, A. Sapetschnig, E. M. Weick, J. Mitchell, M. P. Bagijn, A. C. Cording, A. L. Doebley, L. D. Goldstein, N. J. Lehrbach, J. Le Pen, G. Pintacuda, A. Sakaguchi, P. Sarkies, S. Ahmed, and E. A. Miska. pirnas can trigger a multigenerational epigenetic memory in the germline of c. elegans. *Cell*, 150(1):88–99, 2012.

[100] A. Sapetschnig, P. Sarkies, N. J. Lehrbach, and E. A. Miska. Tertiary sirnas mediate paramutation in c. elegans. *PLoS Genet*, 11(3):e1005078, 2015.

[101] C. Hermant, A. Boivin, L. Teysset, V. Delmarre, A. Asif-Laidin, M. van den Beek, C. Antoniewski, and S. Ronsseray. Paramutation in drosophila requires both nuclear and cytoplasmic actors of the pirna pathway and induces cis-spreading of pirna production. *Genetics*, 201(4):1381–+, 2015.

[102] A. de Vanssay, A. L. Bouge, A. Boivin, C. Hermant, L. Teysset, V. Delmarre, C. Antoniewski, and S. Ronsseray. Paramutation in drosophila linked to emergence of a pirna-producing locus. *Nature*, 490(7418):112–5, 2012.

[103] F. Cuzin and M. Rassoulzadegan. Non-mendelian epigenetic heredity: gametic rnas as epigenetic regulators and transgenerational signals. *Essays in Biochemistry: Epigenetics, Disease and Behaviour*, 48:101–106, 2010.

[104] B. Wiedenheft, S. H. Sternberg, and J. A. Doudna. Rna-guided genetic silencing systems in bacteria and archaea. *Nature*, 482(7385):331–8, 2012.

[105] D. Holoch and D. Moazed. Rna-mediated epigenetic regulation of gene expression. *Nat Rev Genet*, 16(2):71–84, 2015.

[106] F. Borges and R. A. Martienssen. The expanding world of small rnas in plants. *Nat Rev Mol Cell Biol*, 16(12):727–41, 2015.

[107] L. Peters and G. Meister. Argonaute proteins: mediators of rna silencing. *Mol Cell*, 26(5):611–23, 2007.

[108] C. Popp, W. Dean, S. Feng, S. J. Cokus, S. Andrews, M. Pellegrini, S. E. Jacobsen, and W. Reik. Genome-wide erasure of dna methylation in mouse primordial germ cells is affected by aid deficiency. *Nature*, 463(7284):1101–5, 2010.

[109] Z. D. Smith, M. M. Chan, T. S. Mikkelsen, H. Gu, A. Gnirke, A. Regev, and A. Meissner. A unique regulatory phase of dna methylation in the early mammalian embryo. *Nature*, 484(7394):339–44, 2012.

[110] A. Champroux, J. Cocquet, J. Henry-Berger, J. R. Drevet, and A. Kocer. A decade of exploring the mammalian sperm epigenome: Paternal epigenetic and transgenerational inheritance. *Front Cell Dev Biol*, 6:50, 2018.

[111] M. Manikkam, R. Tracey, C. Guerrero-Bosagna, and M. K. Skinner. Plastics derived endocrine disruptors (bpa, dehp and dbp) induce epigenetic transgenerational inheritance of obesity, reproductive disease and sperm epimutations. *PLoS One*, 8(1):e55387, 2013.

[112] L. R. Ingerslev, I. Donkin, O. Fabre, S. Versteyhe, M. Mechta, P. Pattamaprapanont, B. Mortensen, N. T. Krarup, and R. Barres. Endurance training remodels

sperm-borne small rna expression and methylation at neurological gene hotspots. *Clin Epigenetics*, 10:12, 2018.

[113] J. Denham, B. J. O'Brien, J. T. Harvey, and F. J. Charchar. Genome-wide sperm dna methylation changes after 3 months of exercise training in humans. *Epigenomics*, 7(5):717–31, 2015.

[114] I. Donkin, S. Versteyhe, L. R. Ingerslev, K. Qian, M. Mechta, L. Nordkap, B. Mortensen, E. V. Appel, N. Jorgensen, V. B. Kristiansen, T. Hansen, C. T. Workman, J. R. Zierath, and R. Barres. Obesity and bariatric surgery drive epigenetic variation of spermatozoa in humans. *Cell Metab*, 23(2):369–78, 2016.

[115] T. Yu, C. Tang, Y. Zhang, R. Zhang, and W. Yan. Microfluidics-based digital quantitative pcr for single-cell small rna quantification. *Biol Reprod*, 97(3):490–496, 2017.

[116] K. Yoshida, S. I. Fujita, A. Isotani, T. Kudo, S. Takahashi, M. Ikawa, D. Shiba, M. Shirakawa, M. Muratani, and S. Ishii. Intergenerational effect of short-term spaceflight in mice. *iScience*, 24(7):102773, 2021.

[117] L. O. Bygren, G. Kaati, and S. Edvinsson. Longevity determined by paternal ancestors' nutrition during their slow growth period. *Acta Biotheoretica*, 49(1):53–59, 2001.

[118] G. Kaati, L. O. Bygren, and S. Edvinsson. Cardiovascular and diabetes mortality determined by nutrition during parents' and grandparents' slow growth period. *European Journal of Human Genetics*, 10(11):682–688, 2002.

[119] M. E. Pembrey, L. O. Bygren, G. Kaati, S. Edvinsson, K. Northstone, M. Sjostrom, J. Golding, and A. S. Team. Sex-specific, male-line transgenerational responses in humans. *European Journal of Human Genetics*, 14(2):159–166, 2006.

[120] G. Kaati, L. O. Bygren, M. Pembrey, and M. Sjostrom. Transgenerational response to nutrition, early life circumstances and longevity. *European Journal of Human Genetics*, 15(7):784–790, 2007.

[121] D. Vagero, P. R. Pinger, V. Aronsson, and G. J. van den Berg. Paternal grandfather's access to food predicts all-cause and cancer mortality in grandsons. *Nat Commun*, 9(1):5124, 2018.

[122] M. V. Veenendaal, R. C. Painter, S. R. de Rooij, P. M. Bossuyt, J. A. van der Post, P. D. Gluckman, M. A. Hanson, and T. J. Roseboom. Transgenerational effects of prenatal exposure to the 1944-45 dutch famine. *BJOG*, 120(5):548–53, 2013.

[123] L. H. Lumey, A. D. Stein, H. S. Kahn, K. M. van der Pal-de Bruin, G. J. Blauw, P. A. Zybert, and E. S. Susser. Cohort profile: the dutch hunger winter families study. *Int J Epidemiol*, 36(6):1196–204, 2007.

[124] E. W. Tobi, P. E. Slagboom, J. van Dongen, D. Kremer, A. D. Stein, H. Putter, B. T. Heijmans, and L. H. Lumey. Prenatal famine and genetic variation are independently and additively associated with dna methylation at regulatory loci within igf2/h19. *Plos One*, 7(5), 2012.

[125] A. Soubry, J. M. Schildkraut, A. Murtha, F. Wang, Z. Huang, A. Bernal, J. Kurtzberg, R. L. Jirtle, S. K. Murphy, and C. Hoyo. Paternal obesity is associated with igf2 hypomethylation in newborns: results from a newborn epigenetics study (nest) cohort. *BMC Med*, 11:29, 2013.

[126] J. Anselmo, N. H. Scherberg, A. M. Dumitrescu, and S. Refetoff. Reduced sensitivity to thyroid hormone as a transgenerational epigenetic marker transmitted along the human male line. *Thyroid*, 29(6):778–782, 2019.

[127] M. E. Martinez, C. W. Duarte, J. P. Stohn, A. Karaczyn, Z. Wu, V. E. DeMambro, and A. Hernandez. Thyroid hormone influences brain gene expression programs

and behaviors in later generations by altering germ line epigenetic information. *Mol Psychiatry*, 25(5):939–950, 2020.

[128] M. J. Wensink, Y. Lu, L. Tian, G. M. Shaw, S. Rizzi, T. K. Jensen, E. R. Mathiesen, N. E. Skakkebaek, R. Lindahl-Jacobsen, and M. L. Eisenberg. Preconception antidiabetic drugs in men and birth defects in offspring : A nationwide cohort study. *Ann Intern Med*, 2022.

[129] R. Kojima, R. Shinohara, M. Kushima, S. Horiuchi, S. Otawa, H. Yokomichi, Y. Akiyama, T. Ooka, K. Miyake, Z. Yamagata, E. Japan, and G. Children's Study. Prenatal occupational disinfectant exposure and childhood allergies: the japan environment and children's study. *Occup Environ Med*, 2022.

[130] K. Siklenka, S. Erkek, M. Godmann, R. Lambrot, S. McGraw, C. Lafleur, T. Cohen, J. Xia, M. Suderman, M. Hallett, J. Trasler, A. H. Peters, and S. Kimmins. Disruption of histone methylation in developing sperm impairs offspring health transgenerationally. *Science*, 350(6261):aab2006, 2015.

[131] B. J. Lesch, Z. Tothova, E. A. Morgan, Z. Liao, R. T. Bronson, B. L. Ebert, and D. C. Page. Intergenerational epigenetic inheritance of cancer susceptibility in mammals. *Elife*, 8, 2019.

[132] O. Bossdorf, C. L. Richards, and M. Pigliucci. Epigenetics for ecologists. *Ecology Letters*, 11(2):106–115, 2008. URL: `https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1461-0248.2007.01130.x`, `arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1461-0248.2007.01130.x`, `doi:https://doi.org/10.1111/j.1461-0248.2007.01130.x`.

[133] E. Danchin, A. Charmantier, F. A. Champagne, A. Mesoudi, B. Pujol, and S. Blanchet. Beyond dna: integrating inclusive inheritance into an extended theory of evolution. *Nat Rev Genet*, 12(7):475–86, 2011.

[134] E. Danchin, B. Pujol, and R. H. Wagner. The double pedigree: a method for studying culturally and genetically inherited behavior in tandem. *PLoS One*, 8(5):e61254, 2013.

[135] R. Bonduriansky, A. J. Crean, and T. Day. The implications of nongenetic inheritance for evolution in changing environments. *Evol Appl*, 5(2):192–201, 2012.

[136] P. Huneman and D. M. Walsh. *Challenging the modern synthesis : adaptation, development, and inheritance*. Oxford University Press, New York, NY, United States of America, 2017.

[137] Y. Wang, H. Liu, and Z. Sun. Lamarck rises from his grave: parental environment-induced epigenetic inheritance in model organisms and humans. *Biol Rev Camb Philos Soc*, 92(4):2084–2111, 2017.

[138] J. B. Wolf, E. D. Brodie Iii, J. M. Cheverud, A. J. Moore, and M. J. Wade. Evolutionary consequences of indirect genetic effects. *Trends Ecol Evol*, 13(2):64–9, 1998.

[139] S. Yuan, D. Oliver, A. Schuster, H. Zheng, and W. Yan. Breeding scheme and maternal small rnas affect the efficiency of transgenerational inheritance of a paramutation in mice. *Sci Rep*, 5:9266, 2015.

[140] V. R. Nelson, S. H. Spiezio, and J. H. Nadeau. Transgenerational genetic effects of the paternal y chromosome on daughters' phenotypes. *Epigenomics*, 2(4):513–21, 2010.

[141] S. F. Ng, R. C. Lin, C. A. Maloney, N. A. Youngson, J. A. Owens, and M. J. Morris. Paternal high-fat diet consumption induces common changes in the transcriptomes of retroperitoneal adipose and pancreatic islet tissues in female rat offspring. *FASEB J*, 28(4):1830–41, 2014.

[142] S. Chong2007, N. Vickaryous, A. Ashe, N. Zamudio, N. Youngson, S. Hemley, T. Stopka, A. Skoultchi, J. Matthews, H. S. Scott, D. de Kretser, M. O'Bryan,

M. Blewitt, and E. Whitelaw. Modifiers of epigenetic reprogramming show paternal effects in the mouse. *Nat Genet*, 39(5):614–22, 2007.

[143] S. D. M. Brown, C. C. Holmes, A. M. Mallon, T. F. Meehan, D. Smedley, and S. Wells. High-throughput mouse phenomics for characterizing mammalian gene function. *Nat Rev Genet*, 19(6):357–370, 2018.

[144] T. F. Meehan, N. Conte, D. B. West, J. O. Jacobsen, J. Mason, J. Warren, C. K. Chen, I. Tudose, M. Relac, P. Matthews, N. Karp, L. Santos, T. Fiegel, N. Ring, H. Westerberg, S. Greenaway, D. Sneddon, H. Morgan, G. F. Codner, M. E. Stewart, J. Brown, N. Horner, C. International Mouse Phenotyping, M. Haendel, N. Washington, C. J. Mungall, C. L. Reynolds, J. Gallegos, V. Gailus-Durner, T. Sorg, G. Pavlovic, L. R. Bower, M. Moore, I. Morse, X. Gao, G. P. Tocchini-Valentini, Y. Obata, S. Y. Cho, J. K. Seong, J. Seavitt, A. L. Beaudet, M. E. Dickinson, Y. Herault, W. Wurst, M. H. de Angelis, K. C. K. Lloyd, A. M. Flenniken, L. M. J. Nutter, S. Newbigging, C. McKerlie, M. J. Justice, S. A. Murray, K. L. Svenson, R. E. Braun, J. K. White, A. Bradley, P. Flicek, S. Wells, W. C. Skarnes, D. J. Adams, H. Parkinson, A. M. Mallon, S. D. M. Brown, and D. Smedley. Disease model discovery from 3,328 gene knockouts by the international mouse phenotyping consortium. *Nat Genet*, 49(8):1231–1238, 2017.

[145] J. Rozman, B. Rathkolb, M. A. Oestereicher, C. Schutt, A. C. Ravindranath, S. Leuchtenberger, S. Sharma, M. Kistler, M. Willershauser, R. Brommage, T. F. Meehan, J. Mason, H. Haselimashhadi, I. Consortium, T. Hough, A. M. Mallon, S. Wells, L. Santos, C. J. Lelliott, J. K. White, T. Sorg, M. F. Champy, L. R. Bower, C. L. Reynolds, A. M. Flenniken, S. A. Murray, L. M. J. Nutter, K. L. Svenson, D. West, G. P. Tocchini-Valentini, A. L. Beaudet, F. Bosch, R. B. Braun, M. S. Dobbie, X. Gao, Y. Herault, A. Moshiri, B. A. Moore, K. C. Kent Lloyd, C. McKerlie, H. Masuya, N. Tanaka, P. Flicek, H. E. Parkinson, R. Sedlacek, J. K. Seong, C. L. Wang, M. Moore, S. D. Brown, M. H. Tschop, W. Wurst, M. Klingenspor, E. Wolf, J. Beckers, F. Machicao, A. Peter, H. Staiger, H. U. Haring,

H. Grallert, M. Campillos, H. Maier, H. Fuchs, V. Gailus-Durner, T. Werner, and M. Hrabe de Angelis. Identification of genetic elements in metabolism by high-throughput mouse phenotyping. *Nat Commun*, 9(1):288, 2018.

[146] B. A. Moore, B. C. Leonard, L. Sebbag, S. G. Edwards, A. Cooper, D. M. Imai, E. Straiton, L. Santos, C. Reilly, S. M. Griffey, L. Bower, D. Clary, J. Mason, M. J. Roux, H. Meziane, Y. Herault, C. International Mouse Phenotyping, C. McKerlie, A. M. Flenniken, L. M. J. Nutter, Z. Berberovic, C. Owen, S. Newbigging, H. Adissu, M. Eskandarian, C. W. Hsu, S. Kalaga, U. Udensi, C. Asomugha, R. Bohat, J. J. Gallegos, J. R. Seavitt, J. D. Heaney, A. L. Beaudet, M. E. Dickinson, M. J. Justice, V. Philip, V. Kumar, K. L. Svenson, R. E. Braun, S. Wells, H. Cater, M. Stewart, S. Clementson-Mobbs, R. Joynson, X. Gao, T. Suzuki, S. Wakana, D. Smedley, J. K. Seong, G. Tocchini-Valentini, M. Moore, C. Fletcher, N. Karp, R. Ramirez-Solis, J. K. White, M. H. de Angelis, W. Wurst, S. M. Thomasy, P. Flicek, H. Parkinson, S. D. M. Brown, T. F. Meehan, P. M. Nishina, S. A. Murray, M. P. Krebs, A. M. Mallon, K. C. K. Lloyd, C. J. Murphy, and A. Moshiri. Identification of genes required for eye development by high-throughput screening of mouse knockouts. *Commun Biol*, 1:236, 2018.

[147] M. R. Bowl, M. M. Simon, N. J. Ingham, S. Greenaway, L. Santos, H. Cater, S. Taylor, J. Mason, N. Kurbatova, S. Pearson, L. R. Bower, D. A. Clary, H. Meziane, P. Reilly, O. Minowa, L. Kelsey, C. International Mouse Phenotyping, G. P. Tocchini-Valentini, X. Gao, A. Bradley, W. C. Skarnes, M. Moore, A. L. Beaudet, M. J. Justice, J. Seavitt, M. E. Dickinson, W. Wurst, M. H. de Angelis, Y. Herault, S. Wakana, L. M. J. Nutter, A. M. Flenniken, C. McKerlie, S. A. Murray, K. L. Svenson, R. E. Braun, D. B. West, K. C. K. Lloyd, D. J. Adams, J. White, N. Karp, P. Flicek, D. Smedley, T. F. Meehan, H. E. Parkinson, L. M. Teboul, S. Wells, K. P. Steel, A. M. Mallon, and S. D. M. Brown. A large scale hearing loss screen reveals an extensive unexplored genetic landscape for auditory dysfunction. *Nat Commun*, 8(1):886, 2017.

[148] A. L. Swan, C. Schutt, J. Rozman, M. Del Mar Muniz Moreno, S. Brandmaier, M. Simon, S. Leuchtenberger, M. Griffiths, R. Brommage, P. Keskivali-Bond, H. Grallert, T. Werner, R. Teperino, L. Becker, G. Miller, A. Moshiri, J. R. Seavitt, D. D. Cissell, T. F. Meehan, E. F. Acar, C. J. Lelliott, A. M. Flenniken, M. F. Champy, T. Sorg, A. Ayadi, R. E. Braun, H. Cater, M. E. Dickinson, P. Flicek, J. Gallegos, E. J. Ghirardello, J. D. Heaney, S. Jacquot, C. Lally, J. G. Logan, L. Teboul, J. Mason, N. Spielmann, C. McKerlie, S. A. Murray, L. M. J. Nutter, K. F. Odfalk, H. Parkinson, J. Prochazka, C. L. Reynolds, M. Selloum, F. Spoutil, K. L. Svenson, T. S. Vales, S. E. Wells, J. K. White, R. Sedlacek, W. Wurst, K. C. K. Lloyd, P. I. Croucher, H. Fuchs, G. R. Williams, J. H. D. Bassett, V. Gailus-Durner, Y. Herault, A. M. Mallon, S. D. M. Brown, P. Mayer-Kuckuk, M. Hrabe de Angelis, and I. Consortium. Mouse mutant phenotyping at scale reveals novel genes controlling bone mineral density. *PLoS Genet*, 16(12):e1009190, 2020.

[149] N. Spielmann, G. Miller, T. I. Oprea, C.-W. Hsu, G. Fobo, G. Frishman, C. Montrone, H. Haseli Mashhadi, J. Mason, V. Munoz Fuentes, S. Leuchtenberger, A. Ruepp, M. Wagner, D. S. Westphal, C. Wolf, A. Görlach, A. Sanz-Moreno, Y.-L. Cho, R. Teperino, S. Brandmaier, S. Sharma, I. R. Galter, M. A. Östereicher, L. Zapf, P. Mayer-Kuckuk, J. Rozman, L. Teboul, R. K. A. Bunton-Stasyshyn, H. Cater, M. Stewart, S. Christou, H. Westerberg, A. M. Willett, J. M. Wotton, W. B. Roper, A. E. Christiansen, C. S. Ward, J. D. Heaney, C. L. Reynolds, J. Prochazka, L. Bower, D. Clary, M. Selloum, G. Bou About, O. Wendling, H. Jacobs, S. Leblanc, H. Meziane, T. Sorg, E. Audain, A. Gilly, N. W. Rayner, J. A. Aguilar-Pimentel, L. Becker, L. Garrett, S. M. Hölter, O. V. Amarie, J. Calzada-Wack, T. Klein-Rodewald, P. da Silva-Buttkus, C. Lengger, C. Stoeger, R. Gerlini, B. Rathkolb, D. Mayr, J. Seavitt, A. Gaspero, J. R. Green, A. Garza, R. Bohat, L. Wong, M. L. McElwee, S. Kalaga, T. L. Rasmussen, I. Lorenzo, D. G. Lanza, R. C. Samaco, S. Veeraragaven, J. J. Gallegos, P. Kašpárek, S. Petrezsélyová, R. King, S. Johnson, J. Cleak, Z. Szkoe-Kovacs, G. Codner, M. Mackenzie, A. Caul-

der, J. Kenyon, W. Gardiner, H. Phelps, R. Hancock, C. Norris, M. A. Moore, A. M. Seluke, R. Urban, C. Kane, L. O. Goodwin, K. A. Peterson, M. Mckay, J. J. Cook, J. P. Lowy, M. McFarland, J. A. Wood, B. J. Willis, H. Tolentino, T. Tolentino, M. Schuchbauer, J. Salazar, J. Johnson, R. Munson, A. Ayadi, G. Pavlovic, M.-C. Birling, S. Jacquot, D. Ali-Hadji, P. Charles, P. Andre, M.-F. Champy, F. Riet, I. Vukobradovic, Z. Berberovic, D. Qu, R. Guo, A. D'Souza, Z. Huang, S. Camilleri, M. Ganguly, H. Adissu, M. Eskandarian, X. Shang, K. Duffin, C. Xu, K. Roberton, V. Laurin, Q. Lan, G. Sleep, A. Creighton, L. Lintott, M. Gertsenstein, M. Pereira, S. Tondat, A. Patel, M. Cruz, A. Bezginov, D. Miller, W. Hy, A. Yoshiki, N. Tanaka, M. Tamura, Z. Liu, O. Ermakova, A. Ferrara, P. Fruscoloni, C. Seisenberger, A. Bürger, F. Giesert, J. C. Ambrose, P. Arumu gam, R. Bevers, M. Bleda, F. Boardman-Pretty, C. R. Boustred, H. Brittain, M. J. Caulfield, G. C. Chan, T. Fowler, A. Giess, A. Hamblin, S. Henderson, T. J. P. Hubbard, R. Jackson, L. J. Jones, D. Kasperaviciute, M. Kayikci, A. Kousathanas, L. Lahnstein, S. E. A. Leigh, I. U. S. Leong, F. J. Lopez, F. Maleady-Crowe, M. McEntagart, F. Minneci, L. Moutsianas, M. Mueller, N. Murugaesu, A. C. Need, P. O'Donovan, C. A. Odhams, C. Patch, D. Perez-Gil, M. B. Pereira, J. Pullinger, T. Rahim, A. Rendon, T. Rogers, K. Savage, K. Sawant, R. H. Scott, A. Siddiq, A. Sieghart, S. C. Smith, A. Sosinsky, A. Stuckey, M. Tanguy, A. L. Taylor-Tavares, E. R. A. Thomas, S. R. Thompson, A. Tucci, M. J. Welland, E. Williams, K. Witkowska, S. M. Wood, M.-P. Hitz, E. Zeggini, E. Wolf, R. Sedlacek, S. A. Murray, K. L. Svenson, R. E. Braun, J. K. White, L. Kelsey, X. Gao, T. Shiroishi, Y. Xu, J. K. Seong, F. Mammano, G. P. Tocchini-Valentini, A. L. Beaudet, T. F. Meehan, H. Parkinson, D. Smedley, A.-M. Mallon, S. E. Wells, H. Grallert, W. Wurst, S. Marschall, H. Fuchs, S. D. M. Brown, A. M. Flenniken, L. M. J. Nutter, C. McKerlie, Y. Herault, K. C. K. Lloyd, M. E. Dickinson, V. Gailus-Durner, M. Hrabe de Angelis, I. consortium, and G. E. R. Consortium. Extensive identification of genes involved in congenital and structural heart disorders and cardiomyopathy. *Nature Cardiovascular Research*,

1(2):157–173, Feb 2022. URL: https://doi.org/10.1038/s44161-022-00018-8, doi:10.1038/s44161-022-00018-8.

[150] D. J. Stekhoven and P. Buhlmann. Missforest–non-parametric missing value imputation for mixed-type data. *Bioinformatics*, 28(1):112–8, 2012.

[151] N. A. Karp, J. Mason, A. L. Beaudet, Y. Benjamini, L. Bower, R. E. Braun, S. D. M. Brown, E. J. Chesler, M. E. Dickinson, A. M. Flenniken, H. Fuchs, M. H. Angelis, X. Gao, S. Guo, S. Greenaway, R. Heller, Y. Herault, M. J. Justice, N. Kurbatova, C. J. Lelliott, K. C. K. Lloyd, A. M. Mallon, J. E. Mank, H. Masuya, C. McKerlie, T. F. Meehan, R. F. Mott, S. A. Murray, H. Parkinson, R. Ramirez-Solis, L. Santos, J. R. Seavitt, D. Smedley, T. Sorg, A. O. Speak, K. P. Steel, K. L. Svenson, C. International Mouse Phenotyping, S. Wakana, D. West, S. Wells, H. Westerberg, S. Yaacoby, and J. K. White. Prevalence of sexual dimorphism in mammalian phenotypic traits. *Nat Commun*, 8:15475, 2017.

[152] Y. Liao, J. Wang, E. J. Jaehnig, Z. Shi, and B. Zhang. Webgestalt 2019: gene set analysis toolkit with revamped uis and apis. *Nucleic Acids Res*, 47(W1):W199–W205, 2019.

[153] E. McArthur and J. A. Capra. Topologically associating domain boundaries that are stable across diverse cell types are evolutionarily constrained and enriched for heritability. *Am J Hum Genet*, 108(2):269–283, 2021.

[154] C. Chronis, P. Fiziev, B. Papp, S. Butz, G. Bonora, S. Sabri, J. Ernst, and K. Plath. Cooperative binding of transcription factors orchestrates reprogramming. *Cell*, 168(3):442–459 e20, 2017.

[155] B. Langmead, C. Trapnell, M. Pop, and S. L. Salzberg. Ultrafast and memory-efficient alignment of short dna sequences to the human genome. *Genome Biol*, 10(3):R25, 2009.

[156] A. Mammana and H. R. Chung. Chromatin segmentation based on a probabilistic model for read counts explains a large portion of the epigenome. *Genome Biol*, 16:151, 2015.

[157] A. R. Quinlan and I. M. Hall. Bedtools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, 26(6):841–2, 2010.

[158] C. T. Butts. network: A package for managing relational data in r. *Journal of Statistical Software*, 24(2):1–36, 2008. URL: `https://www.jstatsoft.org/index.php/jss/article/view/v024i02`, `doi:10.18637/jss.v024.i02`.

[159] G. Csardi and T. Nepusz. The igraph software package for complex network research. *InterJournal*, Complex Systems:1695, 2006. URL: `https://igraph.org`.

[160] P. Shannon, A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, N. Amin, B. Schwikowski, and T. Ideker. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*, 13(11):2498–504, 2003.

[161] D. N. Cooper, E. V. Ball, and M. Mort. Chromosomal distribution of disease genes in the human genome. *Genet Test Mol Biomarkers*, 14(4):441–6, 2010.

[162] E. V. Koonin. Evolution of genome architecture. *Int J Biochem Cell Biol*, 41(2):298–306, 2009.

[163] I. B. Rogozin, K. S. Makarova, J. Murvai, E. Czabarka, Y. I. Wolf, R. L. Tatusov, L. A. Szekely, and E. V. Koonin. Connected gene neighborhoods in prokaryotic genomes. *Nucleic Acids Res*, 30(10):2212–23, 2002.

[164] S. N. Yazbek, S. H. Spiezio, J. H. Nadeau, and D. A. Buchner. Ancestral paternal genotype controls body weight and food intake for multiple generations. *Hum Mol Genet*, 19(21):4134–44, 2010.

[165] J. H. Nadeau, J. Forejt, T. Takada, and T. Shiroishi. Chromosome substitution strains: gene discovery, functional analysis, and systems studies. *Mamm Genome*, 23(9-10):693–705, 2012.

[166] T. D. Krewson, P. J. Supelak, A. E. Hill, J. B. Singer, E. S. Lander, J. H. Nadeau, and M. R. Palmert. Chromosomes 6 and 13 harbor genes that regulate pubertal timing in mouse chromosome substitution strains. *Endocrinology*, 145(10):4447–51, 2004.

[167] B. M. Nathan, C. A. Hodges, P. J. Supelak, L. C. Burrage, J. H. Nadeau, and M. R. Palmert. A quantitative trait locus on chromosome 6 regulates the onset of puberty in mice. *Endocrinology*, 147(11):5132–8, 2006.

[168] J. B. Singer, A. E. Hill, J. H. Nadeau, and E. S. Lander. Mapping quantitative trait loci for anxiety in chromosome substitution strains of mice. *Genetics*, 169(2):855–62, 2005.

[169] M. P. Leussis, M. L. Frayne, M. Saito, E. M. Berry, K. A. Aldinger, G. N. Rockwell, J. Hammer, R. P., A. E. Baskin-Hill, J. B. Singer, J. H. Nadeau, P. Sklar, and T. L. Petryshen. Genomic survey of prepulse inhibition in mouse chromosome substitution strains. *Genes Brain Behav*, 8(8):806–16, 2009.