

UC Irvine

UC Irvine Previously Published Works

Title

Systems genetics approaches for understanding complex traits with relevance for human disease.

Permalink

<https://escholarship.org/uc/item/3kp5t92d>

Authors

Allayee, Hooman

Farber, Charles

Williams, Evan

et al.

Publication Date

2023-11-14

DOI

10.7554/eLife.91004

Peer reviewed

Systems genetics approaches for understanding complex traits with relevance for human disease

Hooman Allayee^{1,2†}, Charles R Farber^{3,4,5†}, Marcus M Seldin⁶, Evan Graehl Williams⁷, David E James^{8,9,10}, Aldons J Lusis^{11,12,13*}

¹Departments of Population & Public Health Sciences, University of Southern California, Los Angeles, United States; ²Biochemistry & Molecular Medicine, Keck School of Medicine, University of Southern California, Los Angeles, United States; ³Center for Public Health Genomics, University of Virginia School of Medicine, Charlottesville, United States; ⁴Departments of Biochemistry & Molecular Genetics, University of Virginia School of Medicine, Charlottesville, United States; ⁵Public Health Sciences, University of Virginia School of Medicine, Charlottesville, United States; ⁶Department of Biological Chemistry, University of California, Irvine, Irvine, United States; ⁷Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Luxembourg, Luxembourg; ⁸School of Life and Environmental Sciences, University of Sydney, Camperdown, Australia; ⁹Faculty of Medicine and Health, University of Sydney, Camperdown, Australia; ¹⁰Charles Perkins Centre, University of Sydney, Camperdown, Australia; ¹¹Departments of Human Genetics, University of California, Los Angeles, Los Angeles, United States; ¹²Medicine, University of California, Los Angeles, Los Angeles, United States; ¹³Microbiology, Immunology, & Molecular Genetics, David Geffen School of Medicine of UCLA, Los Angeles, United States

*For correspondence:
JLusis@mednet.ucla.edu

†These authors contributed equally to this work

Competing interest: The authors declare that no competing interests exist.

Funding: See page 19

Received: 14 July 2023

Accepted: 16 October 2023

Published: 14 November 2023

Reviewing Editor: Detlef Weigel, Max Planck Institute for Biology Tübingen, Germany

© Copyright Allayee, Farber et al. This article is distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use and redistribution provided that the original author and source are credited.

Abstract Quantitative traits are often complex because of the contribution of many loci, with further complexity added by environmental factors. In medical research, systems genetics is a powerful approach for the study of complex traits, as it integrates intermediate phenotypes, such as RNA, protein, and metabolite levels, to understand molecular and physiological phenotypes linking discrete DNA sequence variation to complex clinical and physiological traits. The primary purpose of this review is to describe some of the resources and tools of systems genetics in humans and rodent models, so that researchers in many areas of biology and medicine can make use of the data.

Introduction

Complex traits, including common diseases, result from the combined effects of multiple genetic variations together with environmental factors. Genome-wide association studies (GWAS) have revealed that most complex traits are highly heterogeneous and can involve contributions from hundreds of genetic variants, each explaining a tiny fraction of susceptibility. Although GWAS have been very successful in identifying thousands of causal genetic loci, the underlying mechanisms and interactions remain, in most cases, poorly understood. The regions identified in GWAS usually contain blocks of credible SNPs, and one challenge is to pinpoint the causal gene in that locus of interest. A further challenge is that the majority of fine-mapped SNPs are found in non-coding regions of the genome with no obvious link to protein function. It is often assumed that such SNPs will regulate transcription of a

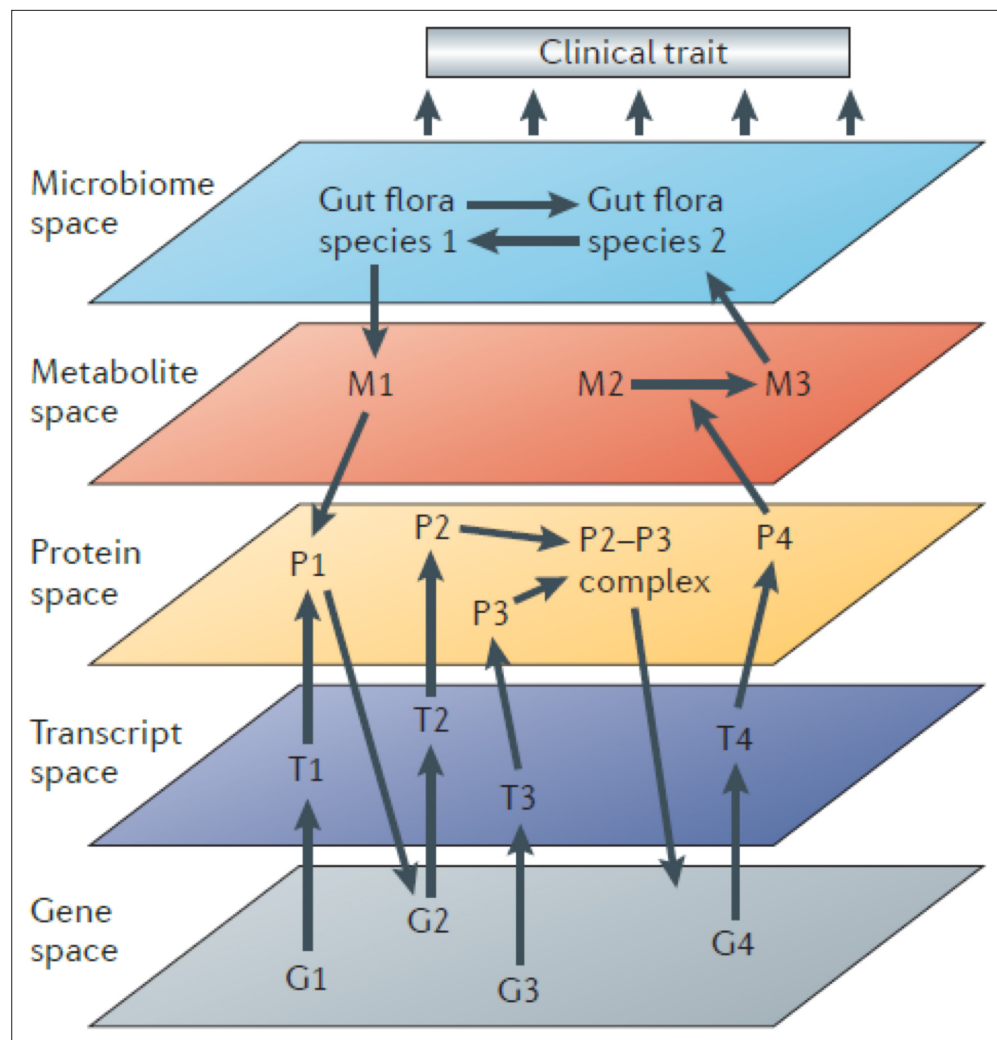


Figure 1. Systems genetics strategy for integration of clinical (and other complex) traits with molecular traits. In this cartoon, individuals in a cohort are examined for clinical or other complex traits of interest. Tissues from the same individuals are also examined using various omics technologies that quantitate to molecular traits. Genetic and environmental variations among the individuals will perturb the clinical and molecular traits. The relationships among the traits can then be statistically modelled using genetic mapping, correlation structure, causal inference, and network modelling. Figure adapted from *Civelek and Lusis, 2014*.

proximal gene but this can be difficult to prove and, in some cases, such as FTO (*Clausnitzer et al., 2015*), has proven not to be the case. A third challenge is to understand how the genetic variation perturbs the system to affect molecular, physiological, or clinical traits.

Systems genetics is an approach that seeks to understand the molecular and physiological phenotypes linking DNA sequence variation to complex clinical and physiological traits. It does this by using high-throughput ‘omics’ technologies to examine how intermediate molecular phenotypes, such as RNA, protein, and metabolite levels, are perturbed by natural genetic variation among individuals in the population and then attempting to relate these variations, using statistical methods, to physiological or clinical traits (**Figure 1**). Systems genetics approaches have proven very useful in identifying causal genes, pathways, and interactions underlying complex traits. We refer readers to a number of reviews of systems genetics strategies that cover historical developments, omics technologies, experimental designs, statistical analyses, and applications in detail (*Ehrenreich et al., 2009; Civelek and Lusis, 2014; Baliga et al., 2017; Schughart and Williams, 2017; Seldin et al., 2019; Li and Auwerx, 2020; Molendijk and Parker, 2021*).

The papers in this issue of eLife provide examples of its many applications across a broad spectrum of model systems. Vast amounts of systems genetics data (i.e., omics-level data collected in diverse populations of humans, mice, rats, flies, or yeast) are now compiled in publicly accessible databases or are available through collaborations. The primary purpose of this review is to describe some of the resources and tools of systems genetics, so that researchers in all areas of biology and medicine can make use of the data. Our focus is limited to studies in mammals although we note that many of the most important studies utilized yeast and flies. We begin with a brief overview, followed by descriptions of human and rodent tools and resources and end with thoughts about future directions.

Overview of systems genetics

Design of systems genetics studies

A typical systems genetics study involves the following steps: (1) Identification of an important question, or set of questions, that could be addressed with a systems genetics study. (2) Selection of an appropriate cohort exhibiting variation in the traits of interest and of sufficient statistical power. (3) Phenotyping of the population for physiological, clinical, and omics traits of interest. (4) Integration of the resulting data through genetic mapping and statistical methods (*Seldin et al., 2019*). (5) Validation of hypotheses using methods such as gain/loss-of-function experimental studies. (6) For studies of disease, validation in human cohorts if possible. Developing a study de novo represents a substantial effort and cost, whether the study involves human or rodent populations. Thus, if possible, the study should be designed such that it can address many different questions and provide a resource for other researchers.

Human studies versus animal models

Complex traits such as common diseases typically involve not only many genes and environmental factors, but also complications such as lifestyle factors, aging-related changes, ethnic differences, sex differences, and in some cases gut microbiome interactions. It is important to design the study with these factors in mind. Several large human biomedical data and research resources have recently emerged, most notably UK Biobank. While such databases will play an invaluable role in future biomedical discoveries, they still currently possess at least two major caveats that are noteworthy. First, access to relevant tissues or organs may be limited and this limits systems genetics analyses. Second, the function of many genes is dramatically influenced by the environment but this is almost impossible to control and/or quantify in humans. For these and other reasons, studies of model systems such as genetically diverse cohorts of mice or rats have proven especially valuable for systems genetics. Notably, the ability to control the environment, age and sex of the animals, and other variables greatly reduce the number of individuals required for the study. As discussed below, a variety of useful mouse and rat population resources have been developed.

Flow of information

Systems genetics can be used to follow the flow of information from DNA variation to intermediate phenotypes for the disease of interest. Thus, it is in some cases useful to interrogate the relevant tissues with different omics platforms (multi-omics) such as transcriptomics, proteomics, and metabolomics. Connecting the dots in this way, from DNA variation to molecular traits to the clinical trait, has been invaluable in identifying causal genes and relevant biological pathways. Information can, of course, also flow in reverse, and statistical methods can be useful in determining the direction of information flow (*Figure 1*).

Statistical analyses

Since information flow from DNA is unidirectional, causal pathways can be modeled and 'mediators' identified. For example, if both a clinical trait and the levels of a transcript are correlated and map to the same locus, one can condition on the transcript levels and ask whether a significant association between the locus and the clinical trait remains. If the association is eliminated, the results suggest that the effect on the clinical trait is not mediated by the transcript. Various causal inference tests have been developed and are typically referred to as mediation analysis (*Schadt et al., 2005; Zhu et al., 2016*). Mendelian randomization (MR) (*Zhu et al., 2016*) is one form of mediation analysis that has

been particularly informative in dissecting causal influences between intermediate physiologic traits and disease pathophysiology in human studies. Higher-order interactions in biological systems are also conveniently modeled as network graphs using systems genetics data (**Figure 1; Huang et al., 2018**).

Reductionist versus systems genetics approaches

Biological research is now dominated by reductionist approaches, such as gain- or loss-of-function studies in mice. These approaches are powerful in that they establish causality, but they have some important limitations that hinder a more complete understanding of the architecture of complex traits, as discussed below. In contrast, systems genetics studies must generally be combined with experimental studies to conclusively establish causality. One constraint with the purely reductionist approach is that it usually involves perturbation of a single gene in a single genetic background and thus is unlikely to detect genetic interactions, such as modifier genes (**Riordan and Nadeau, 2017**). In other words, a genetic variation acts in the context of the genetic background, and by examining the effect of a gain or loss of function in only a single genetic background, an incomplete view of the function of the gene will be obtained. For example, engineered mutations in mice often exhibit strikingly different phenotypes when examined in different strains (**Sittig et al., 2016**).

Another important feature of systems genetics approaches is that they are relatively unbiased. Reductionist scientists usually generate hypotheses based on results from previous studies, and thus some genes or pathways are explored in great depth, whereas others are ignored. A recent study (**Stoeger et al., 2018**) found that more than one-quarter of coding genes have never been the subject of a single paper, whereas approximately 2000 genes (<10% of the coding genome) have dominated the literature. Systems genetics hypotheses, in contrast, are driven by natural variation paired with global measures of omics data and are therefore relatively unbiased. The power of natural variation derives from the multitude of genetic perturbations that occur in all combinations in a population (**Heinz et al., 2013; Battle et al., 2017**).

Omics technologies

The different omics technologies used in systems genetics have been reviewed in detail (**Hasin et al., 2017; Molendijk and Parker, 2021**). In addition to genotyping, technologies have been developed that allow broad characterization of DNA modifications (epigenomics), gene expression (transcriptomics), protein expression (proteomics), small molecules (metabolomics), gut bacteria (metagenomics and metatranscriptomics). Below, we briefly review each of these technologies and their integration into systems genetics analyses of complex traits.

Epigenomics

In addition to variation in genomic sequence, chemical modifications to the nucleotide bases or DNA-bound histones, such as methylation, acetylation, and phosphorylation, can also vary between individuals (and cell types). Epigenomics is the study of such modifications and how they regulate gene expression (**Allis and Jenuwein, 2016**). The most well characterized of these modifications is the methylation of cytosine residues, which tend to cluster at repetitive CG dinucleotides in the proximal promoters of genes in regions referred to as CpG islands. Compared to unmethylated cytosines, methylated cytosine residues are protected from sodium bisulfite-induced deamination to uracil. Thus, it is possible to accurately quantitate the degree of methylation at individual cytosines by comparing the same DNA sequence before and after sodium bisulfite treatment.

Analogous to microarrays or SNP arrays, high-throughput methylation arrays have been developed that allow the simultaneous quantitative evaluation of >800,000 methylation sites throughout the genome. While such arrays do not cover all potential methylation sites, they have been designed to capture the proximal promoters of genes and other known important regulatory regions. Coupled with their relatively low cost, methylation arrays have allowed epigenome-wide association studies (EWAS) to be carried out on a large-scale (**Campagna et al., 2021**). In this approach, the degree of methylation at cytosines is tested for association with clinical traits or outcomes on a genome-wide level. Because of the large numbers of samples that have been archived or the ease with which they can be collected, most EWAS have been carried out with methylation levels in blood cell-derived DNA. Importantly, analytical methods have been developed that use the methylation data itself to

deconvolute individual leukocyte populations (**Houseman et al., 2012; Salas et al., 2022**), which can be used as covariates in the analyses given the heterogeneity in blood cell number among individuals. Furthermore, databases that have curated EWAS datasets are publicly available (**Li et al., 2019**) and provide efficient access to published results for integration with other omics data, such as those described below.

An example of the application of systems genetics to understand epigenetic marks is the study by **Orozco et al., 2015**. The paper examined the contribution of DNA methylation to several complex traits relevant to heart disease, diabetes, and osteoporosis. DNA methylation in liver was examined in 90 mouse inbred strains from the Hybrid Mouse Diversity Panel (HMDP) (see below) using bisulfite sequencing. The DNA methylation levels in the liver were integrated with global transcript levels as well as proteomic and metabolomic data using correlation, mapping, and modeling. Many associations between epigenetic marks and clinical traits were identified, and causal inference tests were performed using the R statistical package CIT (**Millstein et al., 2009**). CIT performs a series of conditional probability tests to determine if the associations between a genetic locus and a trait are mediated by DNA methylation. About 25% of clinical trait associations were predicted to be causal. The study also revealed an example of how natural genetic variation can influence methylation levels. Mapping analysis of CpG methylation identified several hotspots in the genome in which a single locus was associated with variations in DNA methylation in many locations in the genome. One locus, on mouse chromosome 13, was associated with hundreds of CpGs. A candidate gene in the locus, methionine synthase reductase (*Mtrr*), was regulated in cis and its expression was highly correlated with methylation levels of CpGs mapping to the locus. A causal role of *Mtrr* in the methylation was demonstrated using genetically engineered mice (**Orozco et al., 2015**).

Transcriptomics

Most systems genetics studies performed to date have utilized transcriptomic data, largely because global gene expression can be measured at scale in a cost-effective manner using RNA sequencing (RNA-seq) (**Stark et al., 2019**). Additionally, most (>90%) disease associations identified by GWAS implicate non-coding variation, making the identification of trait-associated variation altering gene regulation (e.g., expression, splicing, etc.) a logical starting point to identify causal genes and mechanisms (**Buniello et al., 2019**). Although the original transcriptomic studies used gene expression arrays, the transcriptomic data used in contemporary systems genetics studies is almost exclusively generated by RNA-seq. There are different ‘flavors’ of RNA-seq analyses (e.g., total RNA-seq, mRNA-seq, small RNA-seq, etc.) but the general principle involves a multi-step process of RNA isolation from cells or tissues, conversion of RNA to cDNA, and next-generation sequencing. Gene expression is then quantified by aligning the resulting reads to a reference genome or transcriptome and counting the number of reads aligning to a particular gene, which equates to a digital readout of gene expression (**Conesa et al., 2016**).

One of the most useful systems genetics analyses involving transcriptomic data is the identification of expression quantitative trait loci (eQTLs) (**Farber and Lusis, 2008**). eQTLs are sets of genetic variants influencing gene expression. There are two types of eQTL, local (also referred to as *cis*-eQTL) and distant (also referred to as *trans*). Local eQTL are loci located in close proximity to the gene they regulate, whereas distant eQTL, as their name implies, are generally defined as being >~1 Mbp from the gene they regulate. Local eQTLs have been widely used to connect variants to genes in genetic studies. Another approach is transcriptome-wide association studies (TWASs) (**Li and Ritchie, 2021**). TWASs use eQTLs from a reference population to predict or impute gene expression in a much larger population. Imputed (genetically regulated) gene expression is then correlated with a complex trait to identify genes potentially driving disease.

RNA-seq is most commonly performed in heterogeneous tissues comprised of multiple cell types. However, genetic influences on gene regulation are often cell-type specific (**Kim-Hellmuth et al., 2020**). Recently, single-cell RNA-seq (scRNA-seq) has enabled the generation of complete transcriptomes from individual cells (**Olsen and Baryawno, 2018**). This provides the opportunity to study the genetics of gene regulation in individual cells or homogenous groups of cells. Generation of scRNA-seq data across a variety of organisms, tissues, and disease states has fueled an explosion of potential cell-specific contributions to complex outcomes. In a recent study, Chatterjee and colleagues created a *Drosophila* model which causes hyper-invasive multilayering of the follicular epithelia (**Chatterjee**

et al., 2022). By comparing scRNA-seq profiles from ovaries of this cancerous strain to normal flies, the authors identified a link between *Keap1* and *Nrf2* as upstream components of tumorigenesis. However, scRNA-seq has technical limitations since it generally requires generating single-cell suspensions with fresh tissue. Therefore, scientists have also developed single-nuclei RNA-seq, which as the name implies involves sequencing RNA from nuclei of individual cells instead of whole-cell RNA. A variety of methods have been developed to analyze available scRNA-seq data to support or refute specific hypotheses. For example, a recent study developed Scallop, a statistical tool which enables quantification of transcriptional noise derived from scRNA-seq data (*Ibañez-Solé et al., 2022*). Here, the authors applied this method to normal or aged datasets in mice and humans to demonstrate that previously inferred transcriptional noise could be attributed to shifting cellular identities and thus not entirely stochastic. In another recent study, single-nuclei RNA-seq of brain tissue in 192 individuals identified 7607 genes with a local eQTL, 46% of which demonstrated cell-type-specific effects. The identified local eQTLs were found to colocalize with many loci identified in GWAS for Alzheimer's disease, Parkinson's disease, and schizophrenia, often in a cell-type-specific manner (*Bryois et al., 2022*). One caveat of scRNA analysis is the very low number of transcripts per gene in an average cell. Some estimates indicate that there are somewhere between 1 and 10 copies of RNA per gene per cell and this might be even lower in non-dividing cells (*Eberwine et al., 2014; Mund et al., 2022*). Hence, the accurate quantification using this method is likely restricted to abundant transcripts.

In the context of RNA-seq, the vast majority of transcriptomics data generated to date are based on short-read sequencing (reads of 75–150 bp). This has proven to be a powerful approach for quantifying expression at the level of individual genes. However, alternative splicing is a major gene regulatory mechanism that underlies disease states, including variation in disease-associated complex traits (*Castaldi et al., 2022*). While excellent for quantifying gene expression, short-read RNA-seq is not ideal for quantifying gene expression at the level of individual transcripts generated via alternative splicing. Long-read RNA-seq (reads of 10,000 bp or more) can address this limitation by enabling the identification and quantification of full-length transcripts. Long-read RNA-seq using technologies such as PacBio (*Rhoads and Au, 2015*) and Oxford Nanopore (*Steinbock and Radenovic, 2015*) promises to improve transcript quantification and our understanding of alternative splicing (*Abood et al., 2023*). Additionally, long-read RNA-seq coupled with short-read data in the short term should improve the identification and interpretation of genetically regulated alterations in splicing (i.e., splicing QTL) involved in disease (*Abood et al., 2023*). Ultimately, it is likely that, as the cost of generating long-read RNA-seq data continues to decrease, it will become the primary approach for characterizing and quantifying transcriptomes for systems genetics studies.

Proteomics

Advances over the past decade in sample preparation, fractionation methods, and mass spectrometry have led to the large-scale implementation of proteomics methods. Many studies now report quantitative tissue or cellular proteome measurements for >5000 proteins using minimalist sample preparation methods. Further, by using much deeper fractionation approaches, researchers can measure even more complete proteomes of 10,000–13,000 proteins in a single sample (*Wang et al., 2019*). These latter methods are not yet feasible on a larger scale due to the long run times and cost, but they presage a near future when scientists can measure the proteome as comprehensively and affordably as nucleotide sequencing has made transcriptomics. Several systems genetics studies have already utilized tissue proteomics to great affect and this approach is becoming a valuable part of the systems genetics armamentarium (*Mitok et al., 2018; Parker et al., 2019; Xiao et al., 2022; Yu et al., 2023*).

Other technologies capable of broadly surveying protein levels include the SOMAscan aptamer-based multiplexed platform based on the development of single-stranded DNA aptamers that show high affinity for protein targets (*Gold et al., 2010*) and the proximity extension assay developed by Olink (*Assarsson et al., 2014*). However, these methods are most commonly used on human plasma samples and their application to tissue samples is yet to be realized. The latest platform can now measure over 5000 proteins in a sample and has been used to consistently quantify plasma proteins from tens of thousands of human subjects. This method quantifies proteins when two antibodies to which barcodes are attached bind to the same protein, thus bringing the barcodes into close proximity with each other and allowing nucleic acid-based amplification. There are other promising techniques under development as well, including single-cell proteomics (*Molendijk and Parker, 2021*).

Similar to genetic studies of transcript levels, genetic variations affecting proteins can be classified into those in which the locus maps near the coding gene or those that map elsewhere in the genome. The former are termed 'local' protein QTL (pQTL) and presumably act in *cis*, affecting only the expression of the gene on the contiguous DNA strand. The latter are termed *trans* pQTL. The correlation between transcript and protein levels in genetic studies tends to be relatively modest ([Liu et al., 2016](#)). Thus, the information derived from a proteomics analysis will be quite different from that of a transcriptomics systems genetics study. Likewise, genetic loci affecting protein levels are often distinct from those that regulate the corresponding mRNA, even for variants acting in *cis* ([Williams et al., 2022](#)), for several reasons. First, the turnover of individual proteins varies from minutes to many days in cells; for instance, mitochondrial proteins exhibit particularly long half-lives ([Minard et al., 2016](#)). Moreover, protein turnover is under intense control by various post-translational modifications such as ubiquitination, phosphorylation, or acetylation. This is likely to be a major contributor to the non-linear relationship between mRNA and protein expression in cells. Second, many proteins form multi-protein complexes and this often leads to complex co-regulation of all the proteins that exist in such complexes. The archetypal example of this is the T cell receptor, which comprises six subunits. In T cells, surface expression of this receptor requires efficient expression and assembly of all six subunits in stoichiometric amounts and failure to make one single subunit results in rapid degradation of the remaining subunits ([Weissman, 1994](#)). Third, protein levels and functions can be affected by missense and nonsense variations that affect translation or processing. Thus, the genetic control of cellular protein levels is highly complex and requires more detailed analyses. Proteomics analyses are also capable of examining certain protein modifications that may be of functional relevance. Protein phosphorylation, in particular, is of great importance in intracellular signaling ([Needham et al., 2022](#)). In addition, other protein modifications have recently been quantified in systems genetics reference panels, such as UFMylation, which was demonstrated to impact muscle physiology ([Molendijk et al., 2022](#)).

Metabolomics

Similar to proteomics, high-throughput technologies have been applied to the characterization of the metabolome, often referred to as the pool of metabolites in biological systems, such as cells or tissues ([Wishart, 2019](#)). Two of the most commonly used methods include MS, coupled with either LC or gas chromatography (GC), or nuclear magnetic resonance (NMR). When done in an unbiased fashion, LC/GC-MS can detect and quantify thousands of small molecules in a particular sample. However, this approach is limited by not being able to provide absolute quantification of the features detected unless known standards are spiked into the sample being analyzed. By comparison, NMR-based metabolomics approaches can provide absolute levels with minimal or no pre-processing of samples, thus increasing efficiency, reducing time, and decreasing costs. However, NMR is typically not as sensitive as MS methods. Both metabolomics approaches have been applied in thousands of subjects from numerous human cohorts and in relation to multiple common diseases ([Wishart, 2019](#)). By being distal to mRNA and proteins but proximal to disease outcomes, metabolomics also offers the advantage of providing a level of biological data situated between DNA sequence variation and clinical traits.

Aside from characterizing the repertoire of small molecules, the coupling of metabolomics with genetic data has also allowed large-scale GWAS to be carried out simultaneously for thousands of metabolites, analogous to eQTL and pQTL analyses ([Kastenmüller et al., 2015](#)). Such metabolomic QTL (mQTL) have been very informative by revealing genetic determinants of small molecules derived from cellular and physiological processes and have provided insight into the intermediate mechanisms that could underlie association of loci with disease traits. mQTL summary statistics are also publicly available for download similar to other omics data through searchable web-based resources, thus facilitating efficient systems genetics analyses.

Gut microbiome

Systems genetics can also be applied to the gut microbiota. The various microbes can be identified and quantitated using shotgun sequencing of microbial DNA or sequencing of informative amplicons of the bacterial 16S gene. Whole microbial DNA sequencing frequently permits reads to be specifically assigned to individual species or genera. In contrast, 16S sequencing is much less expensive, but

it generates data which are often analyzed at a much broader scale, for example at the phylum level. Gut microbe compositions are determined by the diet, host genetics, and other characteristics of the host such as aging. Usually, in humans, bacterial DNA is isolated from stool samples, whereas in mice, cecal contents or fecal pellets can be used. Large-scale population studies of gut microbes, which include fungi as well as bacteria, have been performed in an effort to relate gut bacterial composition to clinical outcomes or other traits (*Kurilshikov et al., 2021*). Because there are thousands of species in a single individual's gut microbiome, and because the microbes exist as communities with other species, it has been difficult to identify specific relationships, although broad categories, such as the ratio of Firmicutes to Bacteroidetes phyla, have been related to traits such as obesity. Such studies are simplified in animal models, such as the mouse. For example, in a systems genetics study of microbes in a diverse set of ~100 inbred strains of mice, the species *Roseburia intestinalis* was found to be associated with resistance to atherosclerosis. Further studies with germ-free mice confirmed this and identified butyrate, a metabolite produced by *Roseburia*, as the intermediate (*Kasahara et al., 2018*). In this regard, analysis of bacteria-derived metabolites has been particularly informative (*Jie et al., 2017*). For example, Hazen and colleagues identified the metabolite trimethylamine-N-oxide, derived from dietary choline and carnitine, as a very significant cause of atherosclerosis (*Tang et al., 2019*). Systems genetics approaches have also successfully identified some host genetic loci with large effects on microbe compositions. The most significant of these is the lactase locus, which regulates the ability to hydrolyze dietary lactose. If an individual is deficient in the intestinal enzyme, the availability of lactose in the gut will lead to an increased abundance of bacteria that can utilize this sugar as an energy source.

Population resources

Mouse and rat populations

The mouse and rat have played critical roles in systems genetics studies of complex traits. The popularity of rodents for systems genetics studies is due to many reasons, including the similarity between rodent and human physiology, the relatively lower cost compared to human studies, access to critical tissues and disease phenotypes, and the availability of populations designed for genetic mapping. In particular, rodent populations are critical for systems genetics studies investigating clinically relevant conditions (e.g., genetic × diet effects in obesity [*Parks et al., 2013*] or biomechanical bone strength [*AlBarghouthi and Farber, 2019*]) that are difficult or impossible to study in humans.

There are two types of rodent populations that are commonly used for systems genetics studies - genetic reference populations (GRPs) and outbred stocks. GRPs can be recombinant inbred lines derived from two or more inbred strains (*Williams and Williams, 2017*) or collections of inbred strains, such as the HMDP (*Bennett et al., 2010*) and the Hybrid Rat Diversity Panel (HRDP) (*Tabakoff et al., 2019*). Outbred stocks can begin from a variety of starting points, including initial crosses between inbred founder strains, with the resulting progeny forming a randomly mating population. The most commonly used genetic reference populations and outbred populations are listed in *Table 1*.

The two different populations have unique advantages and disadvantages (*Jurjens et al., 2023*). The main advantage of GRPs is that animals within each strain in the population are identical and replicable. This eliminates the need for costly genotyping for each new study and provides a stable set of genotypes on which phenotype data can be accumulated. GRPs are ideal for investigating gene × environment (GXE) interactions given that phenotypes can be measured in strains exposed to different environmental conditions. They are also useful for the genetic analysis of traits with low heritability since having replicates of genetically identical individuals from the same strain decreases noise and increases heritability (*Bennett et al., 2010*; *Keele, 2023*). Heritability estimates provide a valuable metric to quantify the genetic and/or environmental contributions to complex outcomes. However, these have been difficult to accurately assess in humans, and studies have suggested differences in heritability estimates depending on methodological approaches used (*Mayhew and Meyre, 2017*). Several improvements to human heritability estimates have been proposed such as gene network-based approaches which have the power to inform context-specific estimates of complex traits (*Feng et al., 2022*). However, one disadvantage for rodent studies is that most GRPs consist of a relatively small number of strains (~25–140), limiting statistical power to detect subtle genetic effects on clinical and molecular traits.

Table 1. Commonly used genetic reference populations (GRPs) and outbred populations in mice and rats.

GRPs	Species	Inbred or outbred	# of strains	Description	Data repository
BXD	Mouse	Inbred	140	Recombinant inbred lines generated from C57BL/6J and DBA/2J founders	https://genenetwork.org/
HMDP	Mouse	Inbred	~100	A set of ~100 classical laboratory inbred strains and multiple recombinant inbred line panels	http://systems.genetics.ucla.edu
CC	Mouse	Inbred	~50–75	A panel of ~75 recombinant inbred lines derived from eight genetically diverse inbred founders	http://csbio.unc.edu/CCstatus/index.py
HRDP	Rat	Inbred	99	A set of ~100 classical laboratory inbred strains and recombinant inbred line panels	http://phenogen.org https://genenetwork.org/
HXB/BXH	Rat	Inbred	30	Recombinant inbred lines generated from the spontaneously hypertensive rat (SHR/Ola ^{pcv}) and Brown Norway (BN.Lx/Cub) founders	http://phenogen.org https://genenetwork.org/
DO	Mouse	Outbred		An outbred population derived from eight genetically diverse inbred founders	https://genenetwork.org/

By comparison, outbred stocks are not limited by sample size, enabling the detection of more subtle genetic effects and relationships among traits. This additional power can be important when exploring the polygenicity of complex traits. Additionally, outbred stocks accumulate recombinations at every generation, increasing mapping resolution over time, as compared to GRPs which have a fixed set of recombinations and, thus, fixed mapping precision. The main disadvantage of outbred stocks is the uniqueness of each animal, which requires de novo genotyping. This limits the number of disease phenotypes that can be captured on the same set of mice. Also, while GXE interactions can be studied in outbred stocks, it is much more straightforward with GRPs. Below, we discuss the most popular GRPs and outbred stocks used in both mice and rats. We note that, while not reviewed here, advanced intercross lines have also been extensively used to map complex traits. For example, a recent intercross between inbred QSi5 and 129T2/SvEms mouse lines identified 37 new QTLs which associated with cardiac interatrial septation (*Moradi Marjaneh et al., 2023*).

Widely used GRPs and outbred stocks

BXD

The BXD GRP is a set of recombinant inbred strains derived from crosses between the C57BL/6J and DBA/2J founders. The initial set of BXD strains was developed in the early 1970s by Dr. Benjamin Taylor at the Jackson Laboratory (*Taylor et al., 1973*). This set was subsequently expanded to approximately 140 strains (*Peirce et al., 2004; Ashbrook et al., 2021*). As stated above, one of the main advantages of inbred GRPs is the ability to accumulate data on the same genotypes over time. The BXD GRP has been used to map a wide range of complex traits (reviewed in *Ashbrook et al., 2021*). As an example, a recent study took advantage of the accumulation of de novo mutations over 50 years in the BXD to identify alleles in *Mutyh*, a known driver of colorectal cancer, that modulate specific germline mutagenesis signatures (*Sasani et al., 2022*). The BXDs are one of the most well-characterized GRP in any species. To date, over 7500 phenotypes, in disease categories such as cardiovascular disease (CAD), behavior, addiction, and cancer, just to name a few, have been collected. Additionally, over 100 ‘-omics’ (transcriptome, metabolome, DNA methylation, and proteome) datasets on multiple tissues have been generated in the BXD GRP, making it a powerful systems genetics resource. A recent integrative systems genetics study of bile acid metabolism in the BXD used different diets, bile acid profiling, liver transcriptomics, and metabolic phenotyping to identify *Ces1c* as a master regulator of plasma tauroursodeoxycholic acid, a bile acid known to have health-promoting actions (*Li et al., 2022*). All of these data are stored on the GeneNetwork2 database (<https://genenetwork.org/>) and can be used for interactive systems genetics studies.

HMDP

The HMDP is a GRP comprised of a set of commercially available recombinant inbred strains (including the BXD strains discussed above) and classical laboratory strains (*Bennett et al., 2010*). Most

experiments using the HMDP have included approximately 100 strains, though the number of strains that could be used is larger. The HMDP was designed to use available resources to address two issues, the limited statistical power of most strain panels and the lower mapping resolution of RILs. In the HMDP, recombinant inbred strains increase sample size and power, while the small haplotype blocks found in classical inbred strains increase mapping resolution. The HMDP has been used to study a wide range of systems genetics studies of disease phenotypes such as plasma lipids ([Bennett et al., 2010](#)), obesity ([Parks et al., 2013](#)), bone mass ([Farber et al., 2011](#)), fatty liver disease ([Hui et al., 2015](#); [Hui et al., 2018](#)), atherosclerosis ([Bennett et al., 2015](#)), heart failure ([Rau et al., 2015](#)), cellular phenotypes ([Davis et al., 2013](#); [Zhou et al., 2015](#)), and various omics datasets ([Ghazalpour et al., 2011](#); [Ghazalpour et al., 2014](#); [Hartiala et al., 2014](#); [Orozco et al., 2015](#)). Furthermore, the HMDP is ideal for GXE experiments and has been used to identify loci and genes responsible for differences in obesity and atherosclerosis between chow and high-fat diets and response to environmental exposures ([Lavinsky et al., 2016](#); [Maazi et al., 2019](#); [Tuominen et al., 2021](#)). Recently, the HMDP was also used to identify coagulation factor XI as an endocrine factor produced in the liver impacting cardiac fibrosis, inflammation, and heart failure ([Cao et al., 2022](#)). Extensive genetic, phenotypic, and omics (transcriptomic and proteomic) data collected in the HMDP can be found at <https://systems.genetics.ucla.edu/HMDP/>.

CC and DO mice

The collaborative cross (CC) and diversity outbred (DO) mice are multi-parental populations created by ‘combining’ eight genetically diverse inbred founders: A/J, C57BL/6J, 129S1/SvImJ, NZO/H1LtJ, NOD/LtJ, WSB/EiJ, PWK/PhJ, and CAST/EiJ ([Churchill et al., 2004](#); [Churchill et al., 2012](#); [Svenson et al., 2012](#); [Saul et al., 2019](#); [Solberg Woods and Palmer, 2019](#)). After the founders were intercrossed in a balanced design, the resulting progeny were used to create a set of recombinant inbred lines (CC) or randomly bred to create DO mice. The genomes of the CC/DO founders have been sequenced and from these data we know that at least ~45 million SNPs and INDELs/structural variants are segregating in the CC and DO mice ([Keane et al., 2011](#)). This level of variation is significantly higher than existing mouse GRPs (e.g., ~5 million variants are segregating in the BXD created from classical laboratory strains [[Wang et al., 2016](#)]) and this level of variation is more than the common variation seen across diverse human populations.

The original goal for the CC was the generation of 1000 recombinant inbred strains ([Churchill et al., 2004](#)); however, genetic incompatibilities among the founders caused most strains to be lost during development. Today there are roughly 50–70 strains that are accessible in the United States. The relatively small number of strains and the increased genetic diversity (which reduces statistical power to identify individual effects) has limited the use of the CC for genetic mapping of complex traits.

The DO has all of the advantages and disadvantages of outbred stocks that are described above. A consideration when designing systems genetics studies using the DO is the increased sample size required over crosses between classical strains to identify QTL in the DO due to the increased genetic variation ([Svenson et al., 2012](#)). Despite this, the DO provides a powerful platform to integrate polygenic traits and has been used to perform GWAS and systems genetics studies for stem cell biology ([Skelly et al., 2020](#); [Aydin et al., 2023](#)), insulin secretion ([Keller et al., 2019](#)), and a wide variety of other disease-associated traits ([Chick et al., 2016](#); [AlBarghouthi et al., 2021](#); [Xiao et al., 2022](#)).

HXB/BXH rats

The HXB/BXH GRP is a set of 30 RILs generated by intercrossing the spontaneously hypertensive rat with the normotensive Brown Norway rat strain ([Pravenec et al., 1989](#); [Pravenec et al., 2014](#)). The HXB/BXH GRP has been used for genetic and systems genetic studies of blood pressure ([Pravenec et al., 1989](#)), alcohol-related traits ([Vanderlinden et al., 2014](#); [Saba et al., 2015](#); [Lusk et al., 2018](#)), metabolic and cardiovascular traits ([Morrissey et al., 2011](#); [Pravenec et al., 2018](#)), among others. The HXB/BXH GRP was recently used to identify a pleiotropic QTL affecting hippocampal neurogenesis and serum glucose ([Senko et al., 2022](#)). Using a systems genetics approach integrating transcriptomics data the Telo2-interacting protein 2 (Tti2) gene was identified as a strong candidate and subsequently validated using a rat Tti2 knockout.

HRDP

The HRDP is a rat GRP similar to the HMDP. Like the HMDP, the HRDP is a panel consisting of both RILs and classical inbred strains, with the goal of maximizing power and mapping resolution using existing resources. Specifically, the HRDP is comprised of 35 inbred strains, 30 strains from the HXB/BXH RIL, and 34 strains from the LEXF/FXLE RIL (Voigt *et al.*, 2008; Tabakoff *et al.*, 2019). Simulations with the HRDP has suggested that an experiment with the full panel has 80% power to detect QTL accounting for approximately 12% of the variance for complex traits (Tabakoff *et al.*, 2019). While few studies have mapped QTL in the HRDP, mapping resolution should be quite good as the GRP has a median haplotype block in of approximately 150 kbp (Tabakoff *et al.*, 2019).

Human populations

Large-scale GWAS, since their first implementation 20 years ago, have been instrumental in elucidating the genetic architecture of complex traits. More recently, this approach has been expanded to include meta-analyses using data from multiple datasets and measurement types (e.g., phenome-wide association studies). This has led the international scientific community to build on this success and develop multi-ancestry biobanks with tens and hundreds of thousands of subjects for systems genetics studies that leverage the vast amount of clinical data and archived blood and tissue samples available in health-care systems around the world. Some of the datasets provide the advantage of including individuals of non-European ancestry and can be used to determine whether epidemiological and genetic findings are applicable across ethnicities. Importantly, this addresses one criticism of human genetic studies that subjects of European ancestry have been overrepresented in GWAS. Furthermore, a collaborative network comprised of 24 international biobanks has been formed, termed the Global Biobank Meta-analysis Initiative (GBMI), with the express goal of facilitating large-scale genetic analyses of selected common diseases (<https://www.globalbiobankmeta.org/>; Zhou *et al.*, 2022). Thus far, over 2.2 million genotyped multi-ancestry subjects from both population-based and hospital/health center-based cohorts have been included in the GBMI, which will undoubtedly increase as more groups join this initiative. While a description of every available biobank is beyond the scope of this review, below we highlight some of those that have extensive resources available and have been used widely (Table 2).

Table 2. Commonly used human biobanks for epidemiological and genetics studies.

Cohort	Ancestry (N)	Disease traits	Biomarkers	Genomics	Transcriptomics	Proteomics	Metabolomics	Data repository
UK Biobank	European, Asian, African, Other (502,492)	✓	✓	✓		✓	✓	https://www.ukbiobank.ac.uk/
FinnGEN	European (500,000*)	✓		✓				https://www.finnngen.fi/en
Biobank Japan	Asian (260,000)	✓	✓	✓		✓	✓	https://biobankjp.org/en/
China Kadoorie Biobank	Asian (512,000)	✓	✓	✓		✓	✓	https://www.ckbiobank.org/
TOPMed	European, African, Hispanic, Asian (205,092)	✓	✓	✓	✓	✓	✓	https://topmed.nhlbi.nih.gov/
BioVU	European (300,000 [†])	✓	✓	✓				https://victor.vumc.org/biovu-description/
Millions Veteran Program	European, African, Hispanic, Asian (950,000 [†])	✓	✓	✓				https://www.mvp.va.gov/pwa/
Geisinger MyCode	European (300,000 [†])	✓		✓				https://www.geisinger.org/precision-health/mycode

*Indicates goal of subject recruitment.

[†]Recruitment still ongoing. Citations for the studies are provided in the text.

The Genotype-Tissue Expression (GTEx) project (<https://gtexportal.org/home/>)

The GTEx project consists of >500 postmortem samples from ~50 tissues in humans. The project was initiated with the goal of providing a resource of organs-specific actions of eQTL variants (**Consortium, 2013**) where intersection with common disease association SNPs could implicate or disregard a specific tissue of action. These data have been particularly useful in providing researchers with a fairly comprehensive paired tissue RNA-seq and genome datasets among the same individuals. More recent efforts have expanded this resource to encompass measures of chromatin accessibility (**Zhang et al., 2021a**), long-read RNA-seq (**Glinos et al., 2022**), and single-nuclear RNA-seq (**Jones et al., 2022a**) in the same individuals.

The UK Biobank

The UK Biobank is a large, multi-site cohort that recruited a total of 503,325 individuals in the UK between 2006 and 2010 (**Bycroft et al., 2018**). Data and health-related information for ~6700 phenotypes are available on each participant, including biological measurements, lifestyle indicators, environmental exposures, imaging, blood and urinary biomarkers, metabolomics, proteomics, and genetic variants and exome and whole-genome sequencing. Since making the data available through an application-based approval process, hundreds of research groups have published thousands of studies using the UK Biobank cohort. Of all biobanks, access to primary level data in UK Biobank is least restrictive and can be obtained through a proposal and application review process.

FinnGen

The goal of the FinnGen biobank is to recruit ~500,000 subjects drawn from epidemiological and disease-based cohorts as well as hospital biobank samples in Finland (**Tabassum et al., 2019**). These data will be combined with registry data that record healthcare events over the entire lifespan. As of December 2022, over 356,000 participants have been recruited and genotyped on SNP arrays and the goal is to finish recruitment by the end of 2023. Results of GWAS analyses for over 2200 disease phenotypes are made publicly available twice year, which can be leveraged for meta-analyses and investigation of candidate loci/genes. However, plans to extend the scope of FinnGen to characterization of biomarkers and other omics data in participants have not been announced.

Biobank Japan

Biobank Japan is a national biobank of ~260,000 patients who were enrolled at 12 hospitals throughout Japan between 2003 and 2008 (**Hirata et al., 2017; Nagai et al., 2017**). The Biobank Japan study examines 51 common diseases in its participants and whole-genome SNP array data is already available in over 215,000 subjects, with plans to extend genetic analysis to exome and whole-genome sequencing. Other omics data, such as metabolomics and proteomics in serum, has also been collected on a subset of the cohort. Similar to UK Biobank and FinnGen, summary statistics are publicly available and can be leveraged for various systems genetics analyses.

China Kadoorie Biobank

The China Kadoorie Biobank has recruited over 500,000 adults from 10 geographically diverse (urban and rural) areas of China from 2004 to 2008 (**Chen et al., 2011**). Extensive data was collected at baseline but the China Kadoorie Biobank is also focused on prospective data collection with surveys of participants carried out every 4–5 years. In addition to genetic (whole-genome genotyping and sequencing) and disease outcome data, biological samples have been collected for biomarker, metabolomics, and proteomics analyses.

TOPMed

The Trans-Omics for Precision Medicine (TOPMed) program is an NIH-sponsored precision medicine initiative consisting of over 200,000 participants drawn from over 85 cohorts. The primary goal is to integrate clinical, genetic (whole-genome sequencing), and other omics data (e.g., epigenomics, transcriptomics, proteomics, and metabolomics) to gain a better understanding of heart, lung, blood, and sleep disorders (**Brody et al., 2017**). One notable advantage of TOPMed is the broad diversity in

participants' ancestries, including those of African, Hispanic/Latino, and Asian backgrounds. However, access to individual-level data and summary statistics is more restrictive than, for example, UK Biobank but can be gained through an application review process.

BioVU

BioVU is an electronic health record (EHR)-based biobank in which over 300,000 participants have been recruited at Vanderbilt University (*Roden et al., 2008*). It is one of the largest single-site biobanks and also leverages whole-genome genotyping, EHR data, and archived plasma samples to investigate the genetic basis of thousands of disease phenotypes. However, access to the data is also more restrictive than some other biobanks and requires being a Vanderbilt University faculty member. BioVU is also part of the large Electronic Medical Records and Genomics (eMERGE) national network of cohorts and biobanks (<https://emerge-network.org/>), which is organized and funded by the NIH to carry out large-scale, high-throughput genetic research with EHR-based data.

MVPMillion Veteran Program

The Million Veteran Program (MVP) is a national research program organized and sponsored by the Veterans Administration Office of Research and Development. The goals of the MVP are to better understand the interrelationships between health outcomes, genetic determinants, behavioral traits, and environmental factors, with a particular focus on improving the care of US Veterans (*Gaziano et al., 2016*). It is one of the largest biobanks and, as of mid 2023, has recruited over 950,000 subjects of different ancestries who are representative of the Veteran population. Thus, the MVP represents one of the largest research programs in the world studying genes and health outcomes.

Geisinger

The Geisinger MyCode Initiative is another EMR-based biobank that is enrolling participants from a rural, integrated health system serving central and northeastern Pennsylvania (*Carey et al., 2016; Jones et al., 2022b*). SNP genotyping and exome sequencing is being carried out on all participants with the same overall goal as other biobanks of discovering novel genetic associations with disease and therapeutic targets. However, one unique distinction between Geisinger MyCode and other biobanks is the decision to return clinically actionable genetic findings (e.g., BRCA1 mutation status) back to subjects. This is based on surveys of participants who generally indicated they are in favor of receiving such information (*Schwartz et al., 2018*).

Analytic approaches

Correlation structure and causal modeling

Correlation is a simple procedure for relating molecular and clinical/physiological traits. Of course, the traits can be correlated for multiple reasons. However, since information from DNA is unidirectional, causal pathways can be modeled. For example, if a clinical trait and the levels of a transcript are correlated and map to the same locus, researchers can condition on the transcript levels and ask whether a significant association between the locus and the clinical trait remains. If so, the results suggest that the effect on the clinical trait is not mediated by the transcript. Furthermore, SNPs are often associated with multiple clinical traits, which can be due to either pleiotropy, direct or indirect genetic effects, or environmental/non-genetic factors. Mediation analysis can therefore be used to quantify the proportion of direct and indirect effects of the exposure on the outcome of interest via the intermediate biomarker as the mediator. One straightforward application of this approach would be to use causal mediation analysis to determine whether association of a variant with insulin resistance is mediated through an effect on BMI. MR (see below) is another form of mediation analysis that has especially strict criteria in that the mediator (such as the levels of a protein) is required to explain all of the association between an SNP and a complex trait. Various forms of causal inference tests have been developed for mediation analysis (*Schadt et al., 2005; Zhu et al., 2016; Zeng et al., 2021; Crouse et al., 2022*).

Mendelian randomization

The gold standard for determining causality with biomarkers has relied on randomized clinical interventions. However, MR has emerged as an alternative and highly efficient strategy for leveraging

genetic data to make causal inferences (*Smith and Ebrahim, 2003; Didelez and Sheehan, 2007; Lawlor et al., 2008; Burgess et al., 2017*). In this approach, MR uses Mendel's laws of inheritance to treat DNA variants as instrumental variables that mimic the randomization of individuals in clinical trials to two 'treatment' groups. Thus, genetic variants that have been associated with biomarkers or intermediate traits of interest (exposures) are then tested for association for disease outcomes (ideally in independent datasets) to infer a causal relationship. This approach has been successfully applied to numerous biomarkers, with the best results being related to classic cardiometabolic risk factors, such as elevated blood pressure, LDL, and triglycerides, that have been confirmed as causal drivers of atherosclerosis (*Jansen et al., 2014*). However, there are inherent limitations to MR that require careful consideration when evaluating causality of intermediate traits. For example, MR assumes that selected genetic instruments are valid and only associated with the biomarker being tested but not other related pathways and traits or confounding factors. In other words, the variants do not have pleiotropic effects. Second, variants may be weak genetic instruments for the exposure, which can lead to imprecise estimates and/or require larger sample sizes to boost power for detecting associations. Various statistical modifications to MR analyses have also been developed to address these limitations, such as two-sample MR (*Hemani et al., 2018*), multivariate MR (*Sanderson et al., 2019*), and TwoStepCisMR (*Woolf et al., 2022*). Thus, the large amount of publicly available summary statistics for numerous biomarkers and disease outcomes from published GWAS analyses renders MR an efficient and important analytical tool in systems genetic studies. One intriguing recent study developed an integrative MR framework to analyze gene-metabolite-phenotype associations from large-scale data (*Auwerx et al., 2023*). Application of this method uncovered new molecular links between omics datatypes, such as *SLC6A12* correlating with serum creatinine through modulation of the levels of the renal osmolyte betaine.

Colocalization

Colocalization approaches have become instrumental in deciphering the genetic basis of GWAS associations by integrating eQTL, meQTL, and pQTL data (*Li and Ritchie, 2021*). The premise of colocalization analysis is to determine if two associations, one to a molecular phenotype (such as gene expression) and the other to a complex trait or disease (i.e., GWAS), are driven by the same causal variants. Thus, colocalization approaches can identify candidate causal genes and prioritize functional variants that may be driving the observed associations. By leveraging molecular data reflecting different layers of biological regulation (e.g., gene expression, DNA methylation, and protein abundance), colocalization studies provide valuable insights into the potential biological mechanisms underlying complex traits and diseases. These approaches aid in translating GWAS findings by generating testable hypotheses (e.g., variant *x* influences the expression of gene *y* which in turn influences a complex trait) for experimental validation. In an elegant example of colocalization analysis, Al-Barghouthi and colleagues performed colocalization of tissue-wide eQTLs from GTEx and GWA significant loci for bone mineral density (BMD) (*AlBarghouthi et al., 2022*). These analyses uncovered >500 new candidate genes mediating variance in human BMD and validated *PPP6R3* as a causal driver by ablating the gene in mice. In another study, Aberra and colleagues applied colocalization to multi-omic data to identify key genes and pathways which would distinguish BMI, WHR, and T2D (*Aberra et al., 2023*).

Several colocalization methods have been developed to facilitate the integration of molecular QTL data and GWAS, each with their strengths and limitations. Bayesian colocalization methods, such as COLOC (*Giambartolomei et al., 2014*), estimate the posterior probabilities of distinct colocalization hypotheses based on the observed summary statistics. Additional methods such as eCAVIAR (*Hormozdiari et al., 2014*) accommodate multiple causal variants. Most of the available methods do not require individual-level data, making them suitable for large-scale integrative analyses with publicly available summary statistics. Other approaches, such as summary-data-based Mendelian randomization (SMR) (*Zhu et al., 2016*) and joint likelihood mapping (*Chun et al., 2017*), rely on distinct statistical frameworks to assess colocalization. These methods have also been successfully applied to a wide range of complex traits and diseases, revealing novel candidate genes and functional variants.

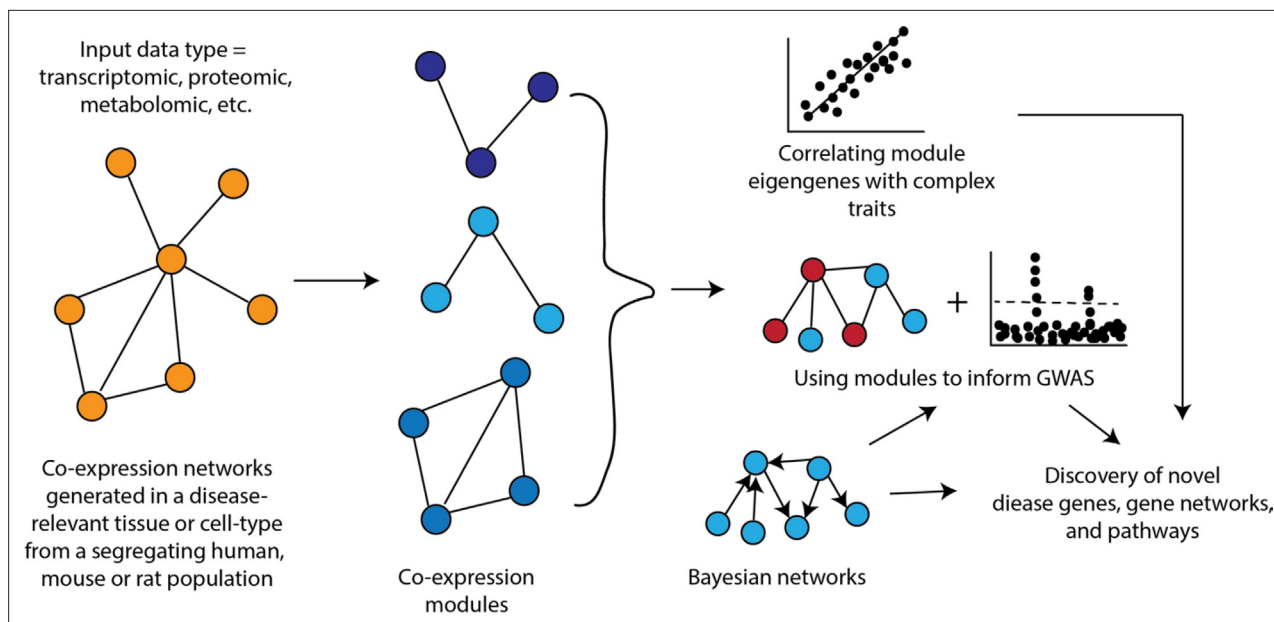


Figure 2. How networks can be used in systems genetics studies of disease.

Cross-tissue interactions

Systems genetics data from multiple tissues can be used to identify potential endocrine circuits mediating crosstalk. The method, first developed by *Seldin et al., 2018*, utilized correlation structure between the expression of individual genes in one tissue and all genes in a second tissue. If a given gene in the first tissue perturbs pathways in the second tissue, it may exhibit an overall correlation that rises above the noise of other gene correlations. The methods work best on animal models where environmental effects can be minimized, although it can also be applied to human datasets that are sufficiently large (*Koplev et al., 2022*). Recently, the method was used in mice to identify coagulation factor 11 as a novel mediator of liver-heart crosstalk with a significant impact on heart function and one form of heart failure (*Cao et al., 2022*).

Network modeling

One of the major premises of systems genetics is that biological components (transcripts, proteins, metabolites, etc.) do not act in isolation, but rather participate in complex networks that operate at the cellular, tissue, and organismal levels (*Nadeau and Dudley, 2011; Civelek and Lusis, 2014*). While reductionist approaches focus on linear events (e.g., kinase phosphorylates protein), systems genetics seeks to understand how genetic variation alters network homeostasis and ultimately disease (*AlBarghouthi and Farber, 2019*). As a result, a key component of systems genetics studies is the reconstruction of cellular networks.

Co-expression networks

Co-expression networks are one of the most popular types of networks used in systems genetics studies (*Figure 2*). In a co-expression network, nodes are genes and edges between genes represent a measure of the strength of their co-expression. Weighted gene co-expression network analysis (WGCNA) is a widely used tool for constructing co-expression networks (*Langfelder and Horvath, 2008*). WGCNA provides a way to organize biology by using the variation in gene expression generated by a series of perturbations (e.g., genetic background, treatments, etc.) to group genes into modules based on similarity in expression patterns. Studies across a wide range of species have demonstrated that genes whose expression co-varies often share similar functions and are members of the same biological pathway or process (*McCarroll et al., 2004; Singer et al., 2005; Ala et al., 2008*).

An important aspect of co-expression networks is that, compared to other biological networks, such as protein-protein interaction networks, they retain tissue or cell-type identity. Additionally, because these networks are generated across a set of samples, it is possible to relate the behavior of modules to other characteristics of the sample (**Figure 1**). For example, one of the first studies utilizing WGCNA identified liver tissue modules that correlated with differences in body weight in a mouse cross (**Ghazalpour et al., 2006**). Once modules are linked to complex traits, one can integrate the specific genes in the module for their involvement by integrating additional data (e.g., genetics data) or their place in the sub-network defined by the module, such as looking at module hub genes (**Figure 1**). Co-expression networks have been used to provide insight into a wide range of complex traits and diseases in rodents and humans such as chronic fatigue syndrome (**Presson et al., 2008**), CAD (**McDermott-Roe et al., 2011**), and diabetes (**Keller et al., 2008**).

Another use of co-expression networks has been to inform GWAS (**Figure 1**). The rationale behind this work is that causal genes from GWAS are often functionally similar and co-expression networks group genes with similar functions into modules. Therefore, it follows that using co-expression networks can help pinpoint potentially causal genes. The way some studies have implemented this strategy is to generate modules from disease-relevant tissues and then scan these modules to identify ones that are enriched for genes implicated by GWAS (located with associated loci) or for disease heritability. Pan et al. recently constructed a co-expression network from human adipose tissue and identified a module associated with waist-to-hip ratio adjusted for BMI (WHRadjBMI) (**Pan et al., 2021**). Additionally, regions around the 347 module genes were enriched for WHRadjBMI heritability. A transcription factor in this module was implicated as important in the context of WHRadjBMI through orthogonal data and its knockdown in human preadipocytes altered a large fraction of module genes (**Pan et al., 2021**). Calabrese et al. used a co-expression network generated from mouse bone to identify the 'osteoblast functional module' and use it to predict that *MARK3* regulated osteoblast activity and BMD (**Calabrese et al., 2017**), which was validated using an osteoblast-specific *Mark3* knockout mouse (**Zhang et al., 2021b**). In another recent example, Li et al. surveyed intestinal transcriptomics and cytokine abundances in 54 BXD strains fed a chow or HF diet (**Li et al., 2023**). The authors performed WGCNA on gene expression and applied association mapping to the module eigengenes to identify potentially causal drivers of intestinal bowel disease (IBD) phenotypes in mice. These modQTLs were intersected with IBD associations in UKBB, where *MUC4* and *EPHA6* were prioritized as key genes. Co-expression network principles have also been widely repurposed to evaluate molecular interactions beyond gene expression. For example, a recent *Caenorhabditis elegans* screen for argonaut protein localization was used to segregate members of the protein family into discrete modules which tracked with stress response and age (**Seroussi et al., 2023**).

Bayesian networks

A limitation of co-expression networks is that the direction of effect between genes is missing, and making definitive conclusions regarding causality is nearly impossible from a statistical perspective. Bayesian network reconstruction tools, on the other hand, have been developed to begin to address this limitation.

Bayesian networks are graphical models that represent probabilistic relationships among variables (**Koller and Friedman, 2009**). Bayesian networks have been widely used in various domains, such as artificial intelligence, machine learning, bioinformatics, and biology due to their ability to model complex relationships among variables and make inferences under uncertainty. They have also been widely used in the context of systems genetics studies. One of the first examples was the generation of Bayesian networks using transcriptomic data from segregating mouse populations to identify gene networks contributing to a wide range of metabolic phenotypes (**Zhu et al., 2004; Mehrabian et al., 2005; Chen et al., 2008**). These studies demonstrated the power of using networks to disentangle the molecular basis of complex traits.

Bayesian networks have been used in other ways in systems genetics studies (**Figure 2**). For example, **Zhao et al., 2016**, used tissue-specific Bayesian networks integrated with the results of GWAS for CAD to identify key driver genes for CAD. In this example, key driver genes were defined as those implicated by GWAS that were connected within a Bayesian network to known disease genes (**Zhao et al., 2016**). Their systems genetics framework identified tissue-specific relations among GWAS candidate genes and prioritized many genes as likely causal drivers of CAD. A similar approach

was recently used to identify likely causal genes for osteoporosis (*AlBarghouthi et al., 2021*). In this analysis, the authors identified key drivers in a Bayesian network generated from mouse cortical bone transcriptomics data and integrated with GWAS data on BMD. A total of 66 high-priority likely causal genes for BMD were identified including *Glt8d2*, which was experimentally demonstrated to be involved in the regulation of BMD.

One particularly powerful Bayesian network approach that incorporates biological information (such as eQTL) is Mergeomics (*Arneson et al., 2016; Shu et al., 2016*), which builds on the explicit hypothesis that multi-omics modalities are functionally related and together can provide information on interconnected biological processes. Mergeomics uses only summary-level multi-omics data, which can be derived from different studies or even species. Briefly, multi-layer disease association signals are mapped to pathways or networks comprising interacting molecules to reveal pathogenic processes perturbed by individual omics variants as well as those affected by multiple omics layers. Recent applications of Mergeomics have yielded substantial insights into the tissue-specific biological processes and regulatory genes involved in individual diseases and those shared between diseases (*Shu et al., 2017; von Scheidt et al., 2017; Chella Krishnan et al., 2018*).

Cross-species integration

While GWAS approaches have proven powerful in humans or model organisms, identification of conserved association signals at syntenic loci has not been widely observed. However, several notable successful examples of cross-species integration have been recently published. In one study, 60 loci were identified for ex vivo secretagogue-induced insulin secretion from isolated pancreatic islets of several hundred DO mice. Genes at these loci were significantly enriched for genes that localized to loci identified in human GWAS for diabetes-related traits (*Keller et al., 2019*). In a second study, integration of human lipid GWAS data with mouse liver co-expression networks from DO mice and the HMDP led to the identification of *SESN1* (Sestrin1) as a gene associated with cholesterol levels (*Li et al., 2020*). Prior to this study, Sestrin1 had not been known to be involved in lipid metabolism. More recent analyses have also demonstrated how incorporation of molecular networks can improve translation of GWAS results across species. For example, GWAS analyses in humans and rats identified hundreds of BMI-associated genes, of which 29 were found to overlap between the two species (*Wright et al., 2023*). However, by integrating molecular networks into the analyses, even greater convergence of GWAS signals was observed. Taken together, such studies highlight how cross-species systems genetics studies can identify novel candidate causal genes for complex traits and the need to develop additional computational tools to better carry out bi-directional translation of findings between humans and model organisms.

Future directions

There will undoubtedly continue to be technical and computational innovations in multi-omics endeavors and this will speed progress in the systems genetics field. For example, above we discuss the value of incorporating long-read RNA-seq into systems genetics analyses (*Glinos et al., 2022*). A comprehensive view of common genetic variation of isoform-specific changes from these data has the potential to refine the definition of causal gene impacts compared to changes in expression alone from eQTL mapping. Further, there has been rapid progress in single-cell modalities. For example, recent technological advances have enabled widespread application of spatially resolved single-cell quantification (*Moses and Pachter, 2022*). While testing the robustness of these data in the context of common genetic variation presents significant challenges (discussed above with scRNA-seq), these measures expand our view of the regulatory landscape of single cells and their interactions with surrounding tissue. In addition, technological advances in methods such as proteomics and metabolomics continue to evolve. Lastly, artificial intelligence is already beginning to be applied to systems genetics data. Databases will continue to increase in size and number, and in silico analyses will become increasingly powerful.

Over the past decade, researchers have significantly refined mass-spectrometry-based methods to detect utilization of specific substrates using stable isotope tracing (*Jang et al., 2018*). Knowledge of these small molecule usages at a genetic scale presents substantial appeal in defining metabolism and associated disease mechanisms. In a recent study, Akingbesote and colleagues measured liver isotope flux, transcriptomics, and proteomics in five species (*Akingbesote et al., 2023*). They

observed ~30,000-fold variation in the usage of substrates and suggested metabolic adaptations which differ between species, for example between mice and rats.

Systems genetics data also present a unique appeal to quantify complex interactions such as gene-by-diet responses (*Nelson et al., 2022*). Without measuring outcomes across diverse genetic backgrounds and environmental conditions, these interactions would be missed entirely which could prove relevant. In a recent example, Hodel and colleagues analyzed two longitudinal human patient cohorts for risk factors mediating coronary heart disease (*Hodel et al., 2023*). Here, the authors identified a unique interaction between polygenic risk of cardiovascular incidents and infection via *Fusobacterium nucleatum*, highlighting the complexities of GXE.

An important challenge will be to develop better methods for curating, storing, sharing, and integrating systems genetics data. We feel that a particularly useful goal would be to combine data from the various systems genetics groups into one searchable database that is accessible to biologists. One mechanism in which this type of database is implemented is available through the Alliance of Genome resources alliancegenome.org (*Alliance of Genome Resources Consortium, 2022*). This web-based API contains all data from seven model organism repositories (Saccharomyces Genome Database, WormBase, FlyBase, Mouse Genome Database, the Zebrafish Information Network, Rat Genome Database, and the Gene Ontology Resource for humans), which can be queried simultaneously at the level of genes, alleles, or disease models. Beyond centralizing these data and curating for streamlined comparisons, significant efforts will have to be applied to provide all researchers the ability to query and understand outputs from systems genetics analyses. This issue spans many areas of computational analyses related to biological data. Several efforts to centralize data analysis toolkits such as Docker, Github, and Jupyter have proved successful in this area; however, still require some knowledge of computational code. Undoubtedly, natural language processing will accelerate these improvements. For example, recent guidelines and considerations for utilization of chatGPT for coding optimization and generation have been summarized (*Piccolo et al., 2023; Shue et al., 2023*).

Major hurdles in applying unified analyses to common datatypes is differences in data structure, available pre-processing steps, and associated meta-data. Efforts have been proposed to reconcile these, such as the implementation of FAIR data principles (*Wilkinson et al., 2016*). These principles were first proposed in 2016 to suggest data standards, which maximize Findability, Accessibility, Interoperability, and Reusability. Given the utility of these principles in data management, these guidelines have been adopted by organizations such as the Research Data Alliance and been proposed as critical for the success of future data mining efforts such as machine learning (*Scheffler et al., 2022*). The approach should also be enhanced by the incorporation of additional data modalities, such as high-resolution clinical imaging data, single-cell gene expression data, and long non-coding RNA. In addition to quantitating the steady-state levels of macromolecules in populations, it may be possible to determine their rates of synthesis and degradation and how this is regulated.

One potentially fertile area of investigation is drug action using systems genetics data. An elegant example of this application was provided by Masson and colleagues who surveyed muscle proteomics from a diverse outbred population of mice exhibiting substantial variation in glucose metabolism profiles (*Masson et al., 2023*). The authors generated a pQTL footprint of proteins which co-mapped to MATSUDA index and linked these signatures to select compounds which were then subjected to two high-throughput drug screening platforms. Using these approaches, they identified thiostrepton as a compound which impacted glucose metabolism independent of the canonical insulin signaling pathway.

Systems genetics approaches are, of course, limited by that natural variation that is captured in the cohorts that are studied. In the case of human studies, certain ethnic groups have been disproportionately studied and, thus, some results may not apply to the less studied groups. Clear racial, ethnic, and gender disparities have been described for most common diseases, emphasizing the need to incorporate more diverse populations in current research gold standards.

Additional information

Funding

Funder	Grant reference number	Author
National Institutes of Health	R01HL133169	Hooman Allayee
National Institutes of Health	R01HL148110	Hooman Allayee
National Institutes of Health	R01HL168493	Hooman Allayee
National Institutes of Health	R01AR079839	Charles R Farber
National Institutes of Health	R01AR077992	Charles R Farber
National Institutes of Health	R01AR071657	Charles R Farber
National Institutes of Health	R01AR079179	Charles R Farber
National Institutes of Health	R01HL148110	Aldons J Lusis
National Institutes of Health	R01HL144651	Aldons J Lusis
National Institutes of Health	R01DK117850	Aldons J Lusis
Australian Research Council	Laureate Fellow	David E James
National Institutes of Health	DP1DK130640	Marcus M Seldin
National Institutes of Health	R21AA030358	Marcus M Seldin

The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

Author contributions

Hooman Allayee, Charles R Farber, Marcus M Seldin, Evan Graehl Williams, David E James, Aldons J Lusis, Conceptualization, Writing – original draft, Writing – review and editing

Author ORCIDs

Hooman Allayee  <https://orcid.org/0000-0002-2384-5239>

Charles R Farber  <https://orcid.org/0000-0002-6748-4711>

Marcus M Seldin  <https://orcid.org/0000-0001-8026-4759>

Evan Graehl Williams  <https://orcid.org/0000-0002-9746-376X>

David E James  <https://orcid.org/0000-0001-5946-5257>

Aldons J Lusis  <https://orcid.org/0000-0001-9013-0228>

References

- Aberra YT**, Ma L, Björkegren JLM, Civelek M. 2023. Predicting mechanisms of action at genetic loci associated with discordant effects on type 2 diabetes and abdominal fat accumulation. *eLife* **12**:e79834. DOI: <https://doi.org/10.7554/eLife.79834>, PMID: 37326626
- Abood A**, Mesner LD, Jeffery ED, Murali M, Lehe M, Saquing J, Farber CR, Sheynkman GM. 2023. Long-Read Proteogenomics to Connect Disease-Associated sQTLs to the Protein Isoform Effectors of Disease. *bioRxiv*. DOI: <https://doi.org/10.1101/2023.03.17.531557>, PMID: 36993769

- Akingbesote ND**, Leitner BP, Jovin DG, Desrouleaux R, Owusu D, Zhu W, Li Z, Pollak MN, Perry RJ. 2023. Gene and protein expression and metabolic flux analysis reveals metabolic scaling in liver ex vivo and in vivo. *eLife* **12**:e78335. DOI: <https://doi.org/10.7554/eLife.78335>, PMID: 37219930
- Ala U**, Piro RM, Grassi E, Damasco C, Silengo L, Oti M, Provero P, Di Cunto F. 2008. Prediction of human disease genes by human-mouse conserved coexpression analysis. *PLOS Computational Biology* **4**:e1000043. DOI: <https://doi.org/10.1371/journal.pcbi.1000043>, PMID: 18369433
- AlBarghouthi BM**, Farber CR. 2019. Dissecting the genetics of osteoporosis using systems approaches. *Trends in Genetics* **35**:55–67. DOI: <https://doi.org/10.1016/j.tig.2018.10.004>, PMID: 30470485
- AlBarghouthi BM**, Mesner LD, Calabrese GM, Brooks D, Tommasini SM, Bouxsein ML, Horowitz MC, Rosen CJ, Nguyen K, Haddox S, Farber EA, Onengut-Gumusc S, Pomp D, Farber CR. 2021. Systems genetics in diversity outbred mice inform BMD GWAS and identify determinants of bone strength. *Nature Communications* **12**:3408. DOI: <https://doi.org/10.1038/s41467-021-23649-0>, PMID: 34099702
- AlBarghouthi BM**, Rosenow WT, Du KP, Heo J, Maynard R, Mesner L, Calabrese G, Nakasone A, Senwar B, Gerstenfeld L, Larner J, Ferguson V, Ackert-Bicknell C, Morgan E, Brautigan D, Farber CR. 2022. Transcriptome-wide association study and eQTL colocalization identify potentially causal genes responsible for human bone mineral density GWAS associations. *eLife* **11**:e77285. DOI: <https://doi.org/10.7554/eLife.77285>, PMID: 36416764
- Alliance of Genome Resources Consortium**. 2022. Harmonizing model organism data in the Alliance of Genome Resources. *Genetics* **220**:iyac022. DOI: <https://doi.org/10.1093/genetics/iyac022>
- Allis CD**, Jenuwein T. 2016. The molecular hallmarks of epigenetic control. *Nature Reviews. Genetics* **17**:487–500. DOI: <https://doi.org/10.1038/nrg.2016.59>, PMID: 27346641
- Arneson D**, Bhattacharya A, Shu L, Mäkinen V-P, Yang X. 2016. Mergeomics: a web server for identifying pathological pathways, networks, and key regulators via multidimensional data integration. *BMC Genomics* **17**:722. DOI: <https://doi.org/10.1186/s12864-016-3057-8>, PMID: 27612452
- Ashbrook DG**, Arends D, Prins P, Mulligan MK, Roy S, Williams EG, Lutz CM, Valenzuela A, Bohl CJ, Ingels JF, McCarty MS, Centeno AG, Hager R, Auwerx J, Lu L, Williams RW. 2021. A platform for experimental precision medicine: The extended BXD mouse family. *Cell Systems* **12**:235–247. DOI: <https://doi.org/10.1016/j.cels.2020.12.002>, PMID: 33472028
- Assarsson E**, Lundberg M, Holmquist G, Björkstén J, Thorsen SB, Ekman D, Eriksson A, Rennel Dickens E, Ohlsson S, Edfeldt G, Andersson A-C, Lindstedt P, Stenvang J, Gullberg M, Fredriksson S. 2014. Homogenous 96-plex PEA immunoassay exhibiting high sensitivity, specificity, and excellent scalability. *PLOS ONE* **9**:e95192. DOI: <https://doi.org/10.1371/journal.pone.0095192>, PMID: 24755770
- Auwerx J**, Sadler MC, Woh T, Reymond A, Kutalik Z, Porcu E. 2023. Exploiting the mediating role of the metabolome to unravel transcript-to-phenotype associations. *eLife* **12**:e81097. DOI: <https://doi.org/10.7554/eLife.81097>, PMID: 36891970
- Aydin S**, Pham DT, Zhang T, Keele GR, Skelly DA, Paulo JA, Pankratz M, Choi T, Gygi SP, Reinholdt LG, Baker CL, Churchill GA, Munger SC. 2023. Genetic dissection of the pluripotent proteome through multi-omics data integration. *Cell Genomics* **3**:100283. DOI: <https://doi.org/10.1016/j.xgen.2023.100283>, PMID: 37082146
- Baliga NS**, Björkegren JLM, Boeke JD, Boutros M, Crawford NPS, Dudley AM, Farber CR, Jones A, Levey AI, Lusk AJ, Mak HC, Nadeau JH, Noyes MB, Petretto E, Seyfried NT, Steinmetz LM, Vonesch SC. 2017. The state of systems genetics in 2017. *Cell Systems* **4**:7–15. DOI: <https://doi.org/10.1016/j.cels.2017.01.005>, PMID: 28125793
- Battle A**, Brown CD, Engelhardt BE, Montgomery SB, GTEx Consortium, Laboratory, Data Analysis & Coordinating Center (LDACC)—Analysis Working Group, Statistical Methods groups—Analysis Working Group, Enhancing GTEx (eGTEx) groups, NIH Common Fund, NIH/NCI, NIH/NHGRI, NIH/NIMH, NIH/NIDA, Biospecimen Collection Source Site—NDRI, Biospecimen Collection Source Site—RPCI, Biospecimen Core Resource—VARI, Brain Bank Repository—University of Miami Brain Endowment Bank, Leidos Biomedical—Project Management, ELSI Study, Genome Browser Data Integration & Visualization—EBI, et al. 2017. Genetic effects on gene expression across human tissues. *Nature* **550**:204–213. DOI: <https://doi.org/10.1038/nature24277>, PMID: 29022597
- Bennett BJ**, Farber CR, Orozco L, Kang HM, Ghazalpour A, Siemers N, Neubauer M, Neuhaus I, Yordanova R, Guan B, Truong A, Yang W, He A, Kayne P, Gargalovic P, Kirchgesner T, Pan C, Castellani LW, Kostem E, Furlotte N, et al. 2010. A high-resolution association mapping panel for the dissection of complex traits in mice. *Genome Research* **20**:281–290. DOI: <https://doi.org/10.1101/gr.099234.109>, PMID: 20054062
- Bennett BJ**, Davis RC, Civelek M, Orozco L, Wu J, Qi H, Pan C, Packard RRS, Eskin E, Yan M, Kirchgesner T, Wang Z, Li X, Gregory JC, Hazen SL, Gargalovic PS, Lusk AJ. 2015. Genetic architecture of atherosclerosis in mice: a systems genetics analysis of common inbred strains. *PLOS Genetics* **11**:e1005711. DOI: <https://doi.org/10.1371/journal.pgen.1005711>, PMID: 26694027
- Brody JA**, Morrison AC, Bis JC, O’Connell JR, Brown MR, Huffman JE, Ames DC, Carroll A, Conomos MP, Gabriel S, Gibbs RA, Gogarten SM, Gupta N, Jaquish CE, Johnson AD, Lewis JP, Liu X, Manning AK, Papanicolaou GJ, Pitsillides AN, et al. 2017. Analysis commons, a team approach to discovery in a big-data environment for genetic epidemiology. *Nature Genetics* **49**:1560–1563. DOI: <https://doi.org/10.1038/ng.3968>, PMID: 29074945
- Bryois J**, Calini D, Macnair W, Foo L, Urich E, Ortmann W, Iglesias VA, Selvaraj S, Nutma E, Marzin M, Amor S, Williams A, Castelo-Branco G, Menon V, De Jager P, Malhotra D. 2022. Cell-type-specific cis-eQTLs in eight human brain cell types identify novel risk genes for psychiatric and neurological disorders. *Nature Neuroscience* **25**:1104–1112. DOI: <https://doi.org/10.1038/s41593-022-01128-z>, PMID: 35915177

- Buniello A**, MacArthur JAL, Cerezo M, Harris LW, Hayhurst J, Malangone C, McMahon A, Morales J, Mountjoy E, Sollis E, Suveges D, Vrousou O, Whetzel PL, Amode R, Guillen JA, Riat HS, Trevanion SJ, Hall P, Junkins H, Flicek P, et al. 2019. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Research* **47**:D1005–D1012. DOI: <https://doi.org/10.1093/nar/gky1120>, PMID: 30445434
- Burgess S**, Small DS, Thompson SG. 2017. A review of instrumental variable estimators for Mendelian randomization. *Statistical Methods in Medical Research* **26**:2333–2355. DOI: <https://doi.org/10.1177/0962280215597579>, PMID: 26282889
- Bycroft C**, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, Motyer A, Vukcevic D, Delaneau O, O'Connell J, Cortes A, Welsh S, Young A, Effingham M, McVean G, Leslie S, Allen N, Donnelly P, Marchini J. 2018. The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**:203–209. DOI: <https://doi.org/10.1038/s41586-018-0579-z>, PMID: 30305743
- Calabrese GM**, Mesner LD, Stains JP, Tommasini SM, Horowitz MC, Rosen CJ, Farber CR. 2017. Integrating GWAS and Co-expression Network Data Identifies Bone Mineral Density Genes SPTBN1 and MARK3 and an Osteoblast Functional Module. *Cell Systems* **4**:46–59. DOI: <https://doi.org/10.1016/j.cels.2016.10.014>, PMID: 27866947
- Campagna MP**, Xavier A, Lechner-Scott J, Maltby V, Scott RJ, Butzkueven H, Jokubaitis VG, Lea RA. 2021. Epigenome-wide association studies: current knowledge, strategies and recommendations. *Clinical Epigenetics* **13**:214. DOI: <https://doi.org/10.1186/s13148-021-01200-8>, PMID: 34863305
- Cao Y**, Wang Y, Zhou Z, Pan C, Jiang L, Zhou Z, Meng Y, Charugundla S, Li T, Allayee H, Seldin MM, Lusis AJ. 2022. Liver-heart cross-talk mediated by coagulation factor XI protects against heart failure. *Science* **377**:1399–1406. DOI: <https://doi.org/10.1126/science.abn0910>, PMID: 36137043
- Carey DJ**, Fetterolf SN, Davis FD, Faucett WA, Kirchner HL, Mirshahi U, Murray MF, Smelser DT, Gerhard GS, Ledbetter DH. 2016. The Geisinger MyCode community health initiative: an electronic health record-linked biobank for precision medicine research. *Genetics in Medicine* **18**:906–913. DOI: <https://doi.org/10.1038/gim.2015.187>, PMID: 26866580
- Castaldi PJ**, Abood A, Farber CR, Sheynkman GM. 2022. Bridging the splicing gap in human genetics with long-read RNA sequencing: finding the protein isoform drivers of disease. *Human Molecular Genetics* **31**:R123–R136. DOI: <https://doi.org/10.1093/hmg/ddac196>, PMID: 35960994
- Chatterjee D**, Costa CAM, Wang XF, Jevitt A, Huang YC, Deng WM. 2022. Single-cell transcriptomics identifies Keap1-Nrf2 regulated collective invasion in a *Drosophila* tumor model. *eLife* **11**:e80956. DOI: <https://doi.org/10.7554/eLife.80956>, PMID: 36321803
- Chella Krishnan K**, Kurt Z, Barrere-Cain R, Sabir S, Das A, Floyd R, Vergnes L, Zhao Y, Che N, Charugundla S, Qi H, Zhou Z, Meng Y, Pan C, Seldin MM, Norheim F, Hui S, Reue K, Lusis AJ, Yang X. 2018. Integration of multi-omics data from mouse diversity panel highlights mitochondrial dysfunction in non-alcoholic fatty liver disease. *Cell Systems* **6**:103–115. DOI: <https://doi.org/10.1016/j.cels.2017.12.006>, PMID: 29361464
- Chen Y**, Zhu J, Lum PY, Yang X, Pinto S, MacNeil DJ, Zhang C, Lamb J, Edwards S, Sieberts SK, Leonardson A, Castellini LW, Wang S, Champy MF, Zhang B, Emilsson V, Doss S, Ghazalpour A, Horvath S, Drake TA, et al. 2008. Variations in DNA elucidate molecular networks that cause disease. *Nature* **452**:429–435. DOI: <https://doi.org/10.1038/nature06757>, PMID: 18344982
- Chen Z**, Chen J, Collins R, Guo Y, Peto R, Wu F, Li L, China Kadoorie Biobank (CKB) collaborative group. 2011. China Kadoorie Biobank of 0.5 million people: survey methods, baseline characteristics and long-term follow-up. *International Journal of Epidemiology* **40**:1652–1666. DOI: <https://doi.org/10.1093/ije/dyr120>, PMID: 22158673
- Chick JM**, Munger SC, Simecek P, Huttlin EL, Choi K, Gatti DM, Raghupathy N, Svenson KL, Churchill GA, Gygi SP. 2016. Defining the consequences of genetic variation on a proteome-wide scale. *Nature* **534**:500–505. DOI: <https://doi.org/10.1038/nature18270>, PMID: 27309819
- Chun S**, Casparino A, Patsopoulos NA, Croteau-Chonka DC, Raby BA, De Jager PL, Sunyaev SR, Cotsapas C. 2017. Limited statistical evidence for shared genetic effects of eQTLs and autoimmune-disease-associated loci in three major immune-cell types. *Nature Genetics* **49**:600–605. DOI: <https://doi.org/10.1038/ng.3795>, PMID: 28218759
- Churchill GA**, Airey DC, Allayee H, Angel JM, Attie AD, Beatty J, Beavis WD, Belknap JK, Bennett B, Berrettini W, Bleich A, Bogue M, Broman KW, Buck KJ, Buckler E, Burmeister M, Chesler EJ, Cheverud JM, Clapcote S, Cook MN, et al. 2004. The Collaborative Cross, a community resource for the genetic analysis of complex traits. *Nature Genetics* **36**:1133–1137. DOI: <https://doi.org/10.1038/ng1104-1133>, PMID: 15514660
- Churchill GA**, Gatti DM, Munger SC, Svenson KL. 2012. The diversity outbred mouse population. *Mammalian Genome* **23**:713–718. DOI: <https://doi.org/10.1007/s00335-012-9414-2>, PMID: 22892839
- Civelek M**, Lusis AJ. 2014. Systems genetics approaches to understand complex traits. *Nature Reviews. Genetics* **15**:34–48. DOI: <https://doi.org/10.1038/nrg3575>, PMID: 24296534
- Claussnitzer M**, Dankel SN, Kim KH, Quon G, Meuleman W, Haugen C, Glunk V, Sousa IS, Beaudry JL, Puvion-Vandier V, Abdennur NA, Liu J, Svensson PA, Hsu YH, Drucker DJ, Mellgren G, Hui CC, Hauner H, Kellis M. 2015. FTO Obesity Variant Circuitry and Adipocyte Browning in Humans. *The New England Journal of Medicine* **373**:895–907. DOI: <https://doi.org/10.1056/NEJMoa1502214>, PMID: 26287746
- Conesa A**, Madrigal P, Tarazona S, Gomez-Cabrero D, Cervera A, McPherson A, Szczesniak MW, Gaffney DJ, Elo LL, Zhang X, Mortazavi A. 2016. A survey of best practices for RNA-seq data analysis. *Genome Biology* **17**:13. DOI: <https://doi.org/10.1186/s13059-016-0881-8>, PMID: 26813401
- Consortium GT**. 2013. The Genotype-Tissue Expression (GTEx) project. *Nature Genetics* **45**:580–585.

- Crouse WL**, Keele GR, Gastonguay MS, Churchill GA, Valdar W. 2022. A Bayesian model selection approach to mediation analysis. *PLOS Genetics* **18**:e1010184. DOI: <https://doi.org/10.1371/journal.pgen.1010184>, PMID: [35533209](https://pubmed.ncbi.nlm.nih.gov/35533209/)
- Davis RC**, van Nas A, Bennett B, Orozco L, Pan C, Rau CD, Eskin E, Lusis AJ. 2013. Genome-wide association mapping of blood cell traits in mice. *Mammalian Genome* **24**:105–118. DOI: <https://doi.org/10.1007/s00335-013-9448-0>, PMID: [23417284](https://pubmed.ncbi.nlm.nih.gov/23417284/)
- Didelez V**, Sheehan N. 2007. Mendelian randomization as an instrumental variable approach to causal inference. *Statistical Methods in Medical Research* **16**:309–330. DOI: <https://doi.org/10.1177/0962280206077743>, PMID: [17715159](https://pubmed.ncbi.nlm.nih.gov/17715159/)
- Eberwine J**, Sul JY, Bartfai T, Kim J. 2014. The promise of single-cell sequencing. *Nature Methods* **11**:25–27. DOI: <https://doi.org/10.1038/nmeth.2769>, PMID: [24524134](https://pubmed.ncbi.nlm.nih.gov/24524134/)
- Ehrenreich IM**, Gerke JP, Kruglyak L. 2009. Genetic dissection of complex traits in yeast: insights from studies of gene expression and other phenotypes in the BYxRM cross. *Cold Spring Harbor Symposia on Quantitative Biology* **74**:145–153. DOI: <https://doi.org/10.1101/sqb.2009.74.013>, PMID: [19734204](https://pubmed.ncbi.nlm.nih.gov/19734204/)
- Farber CR**, Lusis AJ. 2008. Integrating global gene expression analysis and genetics. *Advances in Genetics* **60**:571–601. DOI: [https://doi.org/10.1016/S0065-2660\(07\)00420-8](https://doi.org/10.1016/S0065-2660(07)00420-8), PMID: [18358333](https://pubmed.ncbi.nlm.nih.gov/18358333/)
- Farber CR**, Bennett BJ, Orozco L, Zou W, Lira A, Kostem E, Kang HM, Furlotte N, Berberyan A, Ghazalpour A, Suwanwela J, Drake TA, Eskin E, Wang QT, Teitelbaum SL, Lusis AJ. 2011. Mouse genome-wide association and systems genetics identify *Asxl2* as a regulator of bone mineral density and osteoclastogenesis. *PLOS Genetics* **7**:e1002038. DOI: <https://doi.org/10.1371/journal.pgen.1002038>, PMID: [21490954](https://pubmed.ncbi.nlm.nih.gov/21490954/)
- Feng Z**, Duren Z, Xin J, Yuan Q, He Y, Su B, Wong WH, Wang Y. 2022. Heritability enrichment in context-specific regulatory networks improves phenotype-relevant tissue identification. *eLife* **11**:e82535. DOI: <https://doi.org/10.7554/eLife.82535>, PMID: [36525361](https://pubmed.ncbi.nlm.nih.gov/36525361/)
- Gaziano JM**, Concato J, Brophy M, Fiore L, Pyarajan S, Breeling J, Whitbourne S, Deen J, Shannon C, Humphries D, Guarino P, Aslan M, Anderson D, LaFleur R, Hammond T, Schaa K, Moser J, Huang G, Muralidhar S, Przygodzki R, et al. 2016. Million Veteran Program: A mega-biobank to study genetic influences on health and disease. *Journal of Clinical Epidemiology* **70**:214–223. DOI: <https://doi.org/10.1016/j.jclinepi.2015.09.016>, PMID: [26441289](https://pubmed.ncbi.nlm.nih.gov/26441289/)
- Ghazalpour A**, Doss S, Zhang B, Wang S, Plaisier C, Castellanos R, Brozell A, Schadt EE, Drake TA, Lusis AJ, Horvath S, Gibson G. 2006. Integrating genetic and network analysis to characterize genes related to mouse weight. *PLOS Genetics* **2**:e130. DOI: <https://doi.org/10.1371/journal.pgen.0020130>, PMID: [16934000](https://pubmed.ncbi.nlm.nih.gov/16934000/)
- Ghazalpour A**, Bennett B, Petyuk VA, Orozco L, Hagopian R, Mungrue IN, Farber CR, Sinsheimer J, Kang HM, Furlotte N, Park CC, Wen PZ, Brewer H, Weitz K, Camp DG, Pan C, Yordanova R, Neuhaus I, Tilford C, Siemers N, et al. 2011. Comparative analysis of proteome and transcriptome variation in mouse. *PLOS Genetics* **7**:e1001393. DOI: <https://doi.org/10.1371/journal.pgen.1001393>, PMID: [21695224](https://pubmed.ncbi.nlm.nih.gov/21695224/)
- Ghazalpour A**, Bennett BJ, Shih D, Che N, Orozco L, Pan C, Hagopian R, He A, Kayne P, Yang W, Kirchgessner T, Lusis AJ. 2014. Genetic regulation of mouse liver metabolite levels. *Molecular Systems Biology* **10**:730. DOI: <https://doi.org/10.15252/msb.20135004>, PMID: [24860088](https://pubmed.ncbi.nlm.nih.gov/24860088/)
- Giambartolomei C**, Vukcevic D, Schadt EE, Franke L, Hingorani AD, Wallace C, Plagnol V, Williams SM. 2014. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLOS Genetics* **10**:e1004383. DOI: <https://doi.org/10.1371/journal.pgen.1004383>, PMID: [24830394](https://pubmed.ncbi.nlm.nih.gov/24830394/)
- Glinos DA**, Garborcauskas G, Hoffman P, Ehsan N, Jiang L, Gokden A, Dai X, Aguet F, Brown KL, Garimella K, Bowers T, Costello M, Ardlie K, Jian R, Tucker NR, Ellinor PT, Harrington ED, Tang H, Snyder M, Juul S, et al. 2022. Transcriptome variation in human tissues revealed by long-read sequencing. *Nature* **608**:353–359. DOI: <https://doi.org/10.1038/s41586-022-05035-y>, PMID: [35922509](https://pubmed.ncbi.nlm.nih.gov/35922509/)
- Gold L**, Ayers D, Bertino J, Bock C, Bock A, Brody E, Carter J, Cunningham V, Dalby A, Eaton B, Fitzwater T, Flather D, Forbes A, Foreman T, Fowler C, Gawande B, Goss M, Gunn M, Gupta S, Halladay D, et al. 2010. Aptamer-based multiplexed proteomic technology for biomarker discovery. *Nature Precedings* **5**:e15004. DOI: <https://doi.org/10.1038/npre.2010.4538.1>
- Hartiala J**, Bennett BJ, Tang WHW, Wang Z, Stewart AFR, Roberts R, McPherson R, Lusis AJ, Hazen SL, Allayee H, CARDIoGRAM Consortium. 2014. Comparative genome-wide association studies in mice and humans for trimethylamine N-oxide, a proatherogenic metabolite of choline and L-carnitine. *Arteriosclerosis, Thrombosis, and Vascular Biology* **34**:1307–1313. DOI: <https://doi.org/10.1161/ATVBAHA.114.303252>, PMID: [24675659](https://pubmed.ncbi.nlm.nih.gov/24675659/)
- Hasin Y**, Seldin M, Lusis A. 2017. Multi-omics approaches to disease. *Genome Biology* **18**:83. DOI: <https://doi.org/10.1186/s13059-017-1215-1>, PMID: [28476144](https://pubmed.ncbi.nlm.nih.gov/28476144/)
- Heinz S**, Romanoski CE, Benner C, Allison KA, Kaikkonen MU, Orozco LD, Glass CK. 2013. Effect of natural genetic variation on enhancer selection and function. *Nature* **503**:487–492. DOI: <https://doi.org/10.1038/nature12615>, PMID: [24121437](https://pubmed.ncbi.nlm.nih.gov/24121437/)
- Hemani G**, Zheng J, Elsworth B, Wade KH, Haberland V, Baird D, Laurin C, Burgess S, Bowden J, Langdon R, Tan VY, Yarmolinsky J, Shihab HA, Timpson NJ, Evans DM, Relton C, Martin RM, Davey Smith G, Gaunt TR, Haycock PC. 2018. The MR-Base platform supports systematic causal inference across the human phenome. *eLife* **7**:e34408. DOI: <https://doi.org/10.7554/eLife.34408>, PMID: [29846171](https://pubmed.ncbi.nlm.nih.gov/29846171/)
- Hirata M**, Kamatani Y, Nagai A, Kiyohara Y, Ninomiya T, Tamakoshi A, Yamagata Z, Kubo M, Muto K, Mushiroda T, Murakami Y, Yuji K, Furukawa Y, Zembutsu H, Tanaka T, Ohnishi Y, Nakamura Y, Hospital GBJC, Matsuda K. 2017. Cross-sectional analysis of BioBank Japan clinical data: A large cohort of 200,000 patients

- with 47 common diseases. *Journal of Epidemiology* **27**:S9–S21. DOI: <https://doi.org/10.1016/j.je.2016.12.003>, PMID: 28190657
- Hodel F**, Xu ZM, Thorball CW, de La Harpe R, Letang-Mathieu P, Brenner N, Butt J, Bender N, Waterboer T, Marques-Vidal PM, Vollenweider P, Vaucher J, Fellay J. 2023. Associations of genetic and infectious risk factors with coronary heart disease. *eLife* **12**:e79742. DOI: <https://doi.org/10.7554/eLife.79742>, PMID: 36785929
- Hormozdiari F**, Kostem E, Kang EY, Pasaniuc B, Eskin E. 2014. Identifying causal variants at loci with multiple signals of association. *Genetics* **198**:497–508. DOI: <https://doi.org/10.1534/genetics.114.167908>, PMID: 25104515
- Houseman EA**, Accomando WP, Koestler DC, Christensen BC, Marsit CJ, Nelson HH, Wiencke JK, Kelsey KT. 2012. DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics* **13**:86. DOI: <https://doi.org/10.1186/1471-2105-13-86>, PMID: 22568884
- Huang JK**, Carlin DE, Yu MK, Zhang W, Kreisberg JF, Tamayo P, Ideker T. 2018. Systematic evaluation of molecular networks for discovery of disease genes. *Cell Systems* **6**:484–495. DOI: <https://doi.org/10.1016/j.cels.2018.03.001>, PMID: 29605183
- Hui ST**, Parks BW, Org E, Norheim F, Che N, Pan C, Castellani LW, Charugundla S, Dirks DL, Psychogios N, Neuhaus I, Gerszten RE, Kirchgessner T, Gargalovic PS, Lusa AJ. 2015. The genetic architecture of NAFLD among inbred strains of mice. *eLife* **4**:e05607. DOI: <https://doi.org/10.7554/eLife.05607>
- Hui ST**, Kurt Z, Tuominen I, Norheim F, C Davis R, Pan C, Dirks DL, Magyar CE, French SW, Chella Krishnan K, Sabir S, Campos-Pérez F, Méndez-Sánchez N, Macías-Kauffer L, León-Mimila P, Canizales-Quinteros S, Yang X, Beaven SW, Huertas-Vazquez A, Lusa AJ. 2018. The genetic architecture of diet-induced hepatic fibrosis in mice. *Hepatology* **68**:2182–2196. DOI: <https://doi.org/10.1002/hep.30113>, PMID: 29907965
- Ibañez-Solé O**, Ascensión AM, Araúzo-Bravo MJ, Izeta A. 2022. Lack of evidence for increased transcriptional noise in aged tissues. *eLife* **11**:e80380. DOI: <https://doi.org/10.7554/eLife.80380>, PMID: 36576247
- Jang C**, Chen L, Rabinowitz JD. 2018. Metabolomics and Isotope Tracing. *Cell* **173**:822–837. DOI: <https://doi.org/10.1016/j.cell.2018.03.055>
- Jansen H**, Samani NJ, Schunkert H. 2014. Mendelian randomization studies in coronary artery disease. *European Heart Journal* **35**:1917–1924. DOI: <https://doi.org/10.1093/eurheartj/ehu208>, PMID: 24917639
- Jie Z**, Xia H, Zhong SL, Feng Q, Li S, Liang S, Zhong H, Liu Z, Gao Y, Zhao H, Zhang D, Su Z, Fang Z, Lan Z, Li J, Xiao L, Li J, Li R, Li X, Li F, et al. 2017. The gut microbiome in atherosclerotic cardiovascular disease. *Nature Communications* **8**:845. DOI: <https://doi.org/10.1038/s41467-017-00900-1>, PMID: 29018189
- Jones RC**, Karkanas J, Krasnow MA, Pisco AO, Quake SR, Salzman J, Yosef N, Bulthaupt B, Brown P, Harper W, Hemenez M, Ponnusamy R, Salehi A, Sanagavarapu BA, Spallino E, Aaron KA, Concepcion W, Gardner JM, Kelly B, Neidlinger N, et al. 2022a. The Tabula Sapiens: A multiple-organ, single-cell transcriptomic atlas of humans. *Science* **376**:eabl4896. DOI: <https://doi.org/10.1126/science.abl4896>, PMID: 35549404
- Jones LK**, Strande NT, Calvo EM, Chen J, Rodriguez G, McCormick CZ, Hallquist MLG, Savatt JM, Rocha H, Williams MS, Sturm AC, Buchanan AH, Glasgow RE, Martin CL, Rahm AK. 2022b. A RE-AIM Framework Analysis of DNA-Based Population Screening: Using Implementation Science to Translate Research Into Practice in A Healthcare System. *Frontiers in Genetics* **13**:883073. DOI: <https://doi.org/10.3389/fgene.2022.883073>, PMID: 35692820
- Jurjens AW**, Seldin MM, Giles C, Meikle PJ, Drew BG, Calkin AC. 2023. The potential of integrating human and mouse discovery platforms to advance our understanding of cardiometabolic diseases. *eLife* **12**:e86139. DOI: <https://doi.org/10.7554/eLife.86139>, PMID: 37000167
- Kasahara K**, Krautkramer KA, Org E, Romano KA, Kerby RL, Vivas EI, Mehrabian M, Denu JM, Bäckhed F, Lusa AJ, Rey FE. 2018. Interactions between *Roseburia intestinalis* and diet modulate atherogenesis in a murine model. *Nature Microbiology* **3**:1461–1471. DOI: <https://doi.org/10.1038/s41564-018-0272-x>, PMID: 30397344
- Kastenmüller G**, Raffler J, Gieger C, Suhre K. 2015. Genetics of human metabolism: an update. *Human Molecular Genetics* **24**:R93–R101. DOI: <https://doi.org/10.1093/hmg/ddv263>, PMID: 26160913
- Keane TM**, Goodstadt L, Danecek P, White MA, Wong K, Yalcin B, Heger A, Agam A, Slater G, Goodson M, Furlotte NA, Eskin E, Nellåker C, Whitley H, Cleak J, Janowitz D, Hernandez-Pliego P, Edwards A, Belgard TG, Oliver PL, et al. 2011. Mouse genomic variation and its effect on phenotypes and gene regulation. *Nature* **477**:289–294. DOI: <https://doi.org/10.1038/nature10413>, PMID: 21921910
- Keele GR**. 2023. Which mouse multiparental population is right for your study? The Collaborative Cross inbred strains, their F1 hybrids, or the Diversity Outbred population. *G3* **13**:jkad027. DOI: <https://doi.org/10.1093/g3journal/jkad027>, PMID: 36735601
- Keller MP**, Choi Y, Wang P, Davis DB, Rabaglia ME, Oler AT, Stapleton DS, Argmann C, Schueler KL, Edwards S, Steinberg HA, Chaibub Neto E, Kleinhanz R, Turner S, Hellerstein MK, Schadt EE, Yandell BS, Kendziorski C, Attie AD. 2008. A gene expression network model of type 2 diabetes links cell cycle regulation in islets with diabetes susceptibility. *Genome Research* **18**:706–716. DOI: <https://doi.org/10.1101/gr.074914.107>, PMID: 18347327
- Keller MP**, Rabaglia ME, Schueler KL, Stapleton DS, Gatti DM, Vincent M, Mitok KA, Wang Z, Ishimura T, Simonett SP, Emfinger CH, Das R, Beck T, Kendziorski C, Broman KW, Yandell BS, Churchill GA, Attie AD. 2019. Gene loci associated with insulin secretion in islets from non-diabetic mice. *The Journal of Clinical Investigation* **129**:4419–4432. DOI: <https://doi.org/10.1172/JCI129143>, PMID: 31343992
- Kim-Hellmuth S**, Aguet F, Oliva M, Muñoz-Aguirre M, Kasela S, Wucher V, Castel SE, Hamel AR, Viñuela A, Roberts AL, Mangul S, Wen X, Wang G, Barbeira AN, Garrido-Martín D, Nadel BB, Zou Y, Bonazzola R, Quan J,

- Brown A, et al. 2020. Cell type-specific genetic regulation of gene expression across human tissues. *Science* **369**:eaaz8528. DOI: <https://doi.org/10.1126/science.aaz8528>, PMID: 32913075
- Koller D**, Friedman N. 2009. *Probabilistic Graphical Models: Principles and Techniques* Cambridge, Massachusetts, London, England: MIT Press.
- Koplev S**, Seldin M, Sukhvasi K, Ermel R, Pang S, Zeng L, Bankier S, Di Narzo A, Cheng H, Meda V, Ma A, Talukdar H, Cohain A, Amadori L, Argmann C, Houten SM, Franzén O, Mocchi G, Meelu OA, Ishikawa K, et al. 2022. A mechanistic framework for cardiometabolic and coronary artery diseases. *Nature Cardiovascular Research* **1**:85–100. DOI: <https://doi.org/10.1038/s44161-021-00009-1>, PMID: 36276926
- Kurilshikov A**, Medina-Gomez C, Bacigalupe R, Radjabzadeh D, Wang J, Demirkan A, Le Roy CI, Raygoza Garay JA, Finnicum CT, Liu X, Zhernakova DV, Bonder MJ, Hansen TH, Frost F, Rühlemann MC, Turpin W, Moon J-Y, Kim H-N, Lüll K, Barkan E, et al. 2021. Large-scale association analyses identify host factors influencing human gut microbiome composition. *Nature Genetics* **53**:156–165. DOI: <https://doi.org/10.1038/s41588-020-00763-1>, PMID: 33462485
- Langfelder P**, Horvath S. 2008. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**:559. DOI: <https://doi.org/10.1186/1471-2105-9-559>, PMID: 19114008
- Lavinsky J**, Ge M, Crow AL, Pan C, Wang J, Salehi P, Myint A, Eskin E, Allayee H, Lusi AJ, Friedman RA. 2016. The genetic architecture of noise-induced hearing loss: evidence for a gene-by-environment interaction. *G3: Genes, Genomes, Genetics* **6**:3219–3228. DOI: <https://doi.org/10.1534/g3.116.032516>, PMID: 27520957
- Lawlor DA**, Harbord RM, Sterne JAC, Timpson N, Davey Smith G. 2008. Mendelian randomization: using genes as instruments for making causal inferences in epidemiology. *Statistics in Medicine* **27**:1133–1163. DOI: <https://doi.org/10.1002/sim.3034>, PMID: 17886233
- Li M**, Zou D, Li Z, Gao R, Sang J, Zhang Y, Li R, Xia L, Zhang T, Niu G, Bao Y, Zhang Z. 2019. EWAS Atlas: a curated knowledgebase of epigenome-wide association studies. *Nucleic Acids Research* **47**:D983–D988. DOI: <https://doi.org/10.1093/nar/gky1027>, PMID: 30364969
- Li H**, Auwerx J. 2020. Mouse systems genetics as a prelude to precision medicine. *Trends in Genetics* **36**:259–272. DOI: <https://doi.org/10.1016/j.tig.2020.01.004>, PMID: 32037011
- Li Z**, Votava JA, Zajac GJM, Nguyen JN, Leyva Jaimes FB, Ly SM, Brinkman JA, De Giorgi M, Kaul S, Green CL, St Clair SL, Belisle SL, Rios JM, Nelson DW, Sorci-Thomas MG, Lagor WR, Lamming DW, Eric Yen C-L, Parks BW. 2020. Integrating Mouse and Human Genetic Data to Move beyond GWAS and Identify Causal Genes in Cholesterol Metabolism. *Cell Metabolism* **31**:741–754. DOI: <https://doi.org/10.1016/j.cmet.2020.02.015>, PMID: 32197071
- Li B**, Ritchie MD. 2021. From GWAS to Gene: Transcriptome-Wide Association Studies and Other Methods to Functionally Understand GWAS Discoveries. *Frontiers in Genetics* **12**:713230. DOI: <https://doi.org/10.3389/fgene.2021.713230>, PMID: 34659337
- Li Hao**, Perino A, Huang Q, Von Alvensleben GVG, Banaei-Esfahani A, Velazquez-Villegas LA, Gariani K, Korbilius M, Bou Sleiman M, Imbach J, Sun Y, Li X, Bachmann A, Goeminne LJE, Gallart-Ayala H, Williams EG, Ivanisevic J, Auwerx J, Schoonjans K. 2022. Integrative systems analysis identifies genetic and dietary modulators of bile acid homeostasis. *Cell Metabolism* **34**:1594–1610. DOI: <https://doi.org/10.1016/j.cmet.2022.08.015>, PMID: 36099916
- Li X**, Morel JD, Benegiamo G, Poisson J, Bachmann A, Rapin A, Williams E, Perino A, Schoonjans K, Sleiman MB, Auwerx J. 2023. Genetic and Dietary Modulators of the Inflammatory Response in the Gastro-Intestinal Tract of the BXD Mouse Genetic Reference Population. *eLife* **12**:RP87569. DOI: <https://doi.org/10.7554/eLife.87569.1>
- Liu Y**, Beyer A, Aebersold R. 2016. On the Dependency of Cellular Protein Levels on mRNA Abundance. *Cell* **165**:535–550. DOI: <https://doi.org/10.1016/j.cell.2016.03.014>, PMID: 27104977
- Lusk R**, Saba LM, Vanderlinden LA, Zidek V, Silhavy J, Pravenec M, Hoffman PL, Tabakoff B. 2018. Unsupervised, statistically based systems biology approach for unraveling the genetics of complex traits: a demonstration with ethanol metabolism. *Alcoholism, Clinical and Experimental Research* **42**:1177–1191. DOI: <https://doi.org/10.1111/acer.13763>, PMID: 29689131
- Maazi H**, Hartiala JA, Suzuki Y, Crow AL, Shafiei Jahani P, Lam J, Patel N, Rigas D, Han Y, Huang P, Eskin E, Lusi AJ, Gilliland FD, Akbari O, Allayee H. 2019. A GWAS approach identifies Dapp1 as A determinant of air pollution-induced airway hyperreactivity. *PLOS Genetics* **15**:e1008528. DOI: <https://doi.org/10.1371/journal.pgen.1008528>, PMID: 31869344
- Masson SWC**, Madsen S, Cooke KC, Potter M, Vegas AD, Carroll L, Thillainadesan S, Cutler HB, Walder KR, Cooney GJ, Morahan G, Stöckli J, James DE. 2023. Leveraging genetic diversity to identify small molecules that reverse mouse skeletal muscle insulin resistance. *eLife* **12**:RP86961. DOI: <https://doi.org/10.7554/eLife.86961>, PMID: 37494090
- Mayhew AJ**, Meyre D. 2017. Assessing the heritability of complex traits in humans: methodological challenges and opportunities. *Current Genomics* **18**:332–340. DOI: <https://doi.org/10.2174/1389202918666170307161450>, PMID: 29081689
- McCarroll SA**, Murphy CT, Zou S, Pletcher SD, Chin CS, Jan YN, Kenyon C, Bargmann CI, Li H. 2004. Comparing genomic expression patterns across species identifies shared transcriptional profile in aging. *Nature Genetics* **36**:197–204. DOI: <https://doi.org/10.1038/ng1291>, PMID: 14730301
- McDermott-Roe C**, Ye J, Ahmed R, Sun X-M, Serafin A, Ware J, Bottolo L, Muckett P, Cañas X, Zhang J, Rowe GC, Buchan R, Lu H, Braithwaite A, Mancini M, Hauton D, Martí R, García-Arumí E, Hubner N, Jacob H, et al. 2011. Endonuclease G is a novel determinant of cardiac hypertrophy and mitochondrial function. *Nature* **478**:114–118. DOI: <https://doi.org/10.1038/nature10490>, PMID: 21979051

- Mehrabian M**, Allayee H, Stockton J, Lum PY, Drake TA, Castellani LW, Suh M, Armour C, Edwards S, Lamb J, Lusis AJ, Schadt EE. 2005. Integrating genotypic and expression data in a segregating mouse population to identify 5-lipoxygenase as a susceptibility gene for obesity and bone traits. *Nature Genetics* **37**:1224–1233. DOI: <https://doi.org/10.1038/ng1619>, PMID: 16200066
- Millstein J**, Zhang B, Zhu J, Schadt EE. 2009. Disentangling molecular relationships with a causal inference test. *BMC Genetics* **10**:23. DOI: <https://doi.org/10.1186/1471-2156-10-23>, PMID: 19473544
- Minard AY**, Wong MKL, Chaudhuri R, Tan S-X, Humphrey SJ, Parker BL, Yang JY, Laybutt DR, Cooney GJ, Coster ACF, Stöckli J, James DE. 2016. Hyperactivation of the insulin signaling pathway improves intracellular proteostasis by coordinately up-regulating the proteostatic machinery in adipocytes. *The Journal of Biological Chemistry* **291**:25629–25640. DOI: <https://doi.org/10.1074/jbc.M116.741140>, PMID: 27738101
- Mitok KA**, Freiburger EC, Schueler KL, Rabaglia ME, Stapleton DS, Kwicien NW, Malec PA, Hebert AS, Broman AT, Kennedy RT, Keller MP, Coon JJ, Attie AD. 2018. Islet proteomics reveals genetic variation in dopamine production resulting in altered insulin secretion. *The Journal of Biological Chemistry* **293**:5860–5877. DOI: <https://doi.org/10.1074/jbc.RA117.001102>, PMID: 29496998
- Molendijk J**, Parker BL. 2021. Proteome-wide systems genetics to identify functional regulators of complex traits. *Cell Systems* **12**:5–22. DOI: <https://doi.org/10.1016/j.cels.2020.10.005>, PMID: 33476553
- Molendijk J**, Blazev R, Mills RJ, Ng YK, Watt KI, Chau D, Gregorevic P, Crouch PJ, Hilton JBW, Lisowski L, Zhang P, Reue K, Lusis AJ, Hudson JE, James DE, Seldin MM, Parker BL. 2022. Proteome-wide systems genetics identifies UFMylation as a regulator of skeletal muscle function. *eLife* **11**:e82951. DOI: <https://doi.org/10.7554/eLife.82951>, PMID: 36472367
- Moradi Marjaneh M**, Kirk EP, Patrick R, Alankarage D, Humphreys DT, Del Monte-Nieto G, Cornejo-Paramo P, Janbandhu V, Doan TB, Dunwoodie SL, Wong ES, Moran C, Martin ICA, Thomson PC, Harvey RP. 2023. Quantitative trait and transcriptome analysis of genetic complexity underpinning cardiac interatrial septation in mice using an advanced intercross line. *eLife* **12**:e83606. DOI: <https://doi.org/10.7554/eLife.83606>, PMID: 37272612
- Morrissey C**, Grieve IC, Heinig M, Atanur S, Petretto E, Pravenec M, Hubner N, Aitman TJ. 2011. Integrated genomic approaches to identification of candidate genes underlying metabolic and cardiovascular phenotypes in the spontaneously hypertensive rat. *Physiological Genomics* **43**:1207–1218. DOI: <https://doi.org/10.1152/physiolgenomics.00210.2010>, PMID: 21846806
- Moses L**, Pachter L. 2022. Museum of spatial transcriptomics. *Nature Methods* **19**:534–546. DOI: <https://doi.org/10.1038/s41592-022-01409-2>, PMID: 35273392
- Mund A**, Brunner AD, Mann M. 2022. Unbiased spatial proteomics with single-cell resolution in tissues. *Molecular Cell* **82**:2335–2349. DOI: <https://doi.org/10.1016/j.molcel.2022.05.022>, PMID: 35714588
- Nadeau JH**, Dudley AM. 2011. Genetics: systems genetics. *Science* **331**:1015–1016. DOI: <https://doi.org/10.1126/science.1203869>, PMID: 21350153
- Nagai A**, Hirata M, Kamatani Y, Muto K, Matsuda K, Kiyohara Y, Ninomiya T, Tamakoshi A, Yamagata Z, Mushiroda T, Murakami Y, Yuji K, Furukawa Y, Zembutsu H, Tanaka T, Ohnishi Y, Nakamura Y, Kubo M, BioBank Japan Cooperative Hospital Group. 2017. Overview of the BioBank Japan project: study design and profile. *Journal of Epidemiology* **27**:S2–S8. DOI: <https://doi.org/10.1016/j.je.2016.12.005>, PMID: 28189464
- Needham EJ**, Hingst JR, Parker BL, Morrison KR, Yang G, Onsliev J, Kristensen JM, Højlund K, Ling NXY, Oakhill JS, Richter EA, Kiens B, Petersen J, Pehmøller C, James DE, Wojtaszewski JFP, Humphrey SJ. 2022. Personalized phosphoproteomics identifies functional signaling. *Nature Biotechnology* **40**:576–584. DOI: <https://doi.org/10.1038/s41587-021-01099-9>, PMID: 34857927
- Nelson ME**, Madsen S, Cooke KC, Fritzen AM, Thorius IH, Masson SWC, Carroll L, Weiss FC, Seldin MM, Potter M, Hocking SL, Fazakerley DJ, Brandon AE, Thillainadesan S, Senior AM, Cooney GJ, Stöckli J, James DE. 2022. Systems-level analysis of insulin action in mouse strains provides insight into tissue- and pathway-specific interactions that drive insulin resistance. *Cell Metabolism* **34**:227–239. DOI: <https://doi.org/10.1016/j.cmet.2021.12.013>, PMID: 35021042
- Olsen TK**, Baryawno N. 2018. Introduction to Single-Cell RNA Sequencing. *Current Protocols in Molecular Biology* **122**:e57. DOI: <https://doi.org/10.1002/cpmb.57>, PMID: 29851283
- Orozzo LD**, Morselli M, Rubbi L, Guo W, Go J, Shi H, Lopez D, Furlotte NA, Bennett BJ, Farber CR, Ghazalpour A, Zhang MQ, Bahous R, Rozen R, Lusis AJ, Pellegrini M. 2015. Epigenome-wide association of liver methylation patterns and complex metabolic traits in mice. *Cell Metabolism* **21**:905–917. DOI: <https://doi.org/10.1016/j.cmet.2015.04.025>, PMID: 26039453
- Pan DZ**, Miao Z, Comenho C, Rajkumar S, Koka A, Lee SHT, Alvarez M, Kaminska D, Ko A, Sinsheimer JS, Mohlke KL, Mancuso N, Muñoz-Hernandez LL, Herrera-Hernandez M, Tusié-Luna MT, Aguilar-Salinas C, Pietiläinen KH, Pihlajamäki J, Laakso M, Garske KM, et al. 2021. Correction to: Identification of TBX15 as an adipose master trans regulator of abdominal obesity genes. *Genome Medicine* **13**:139. DOI: <https://doi.org/10.1186/s13073-021-00954-3>, PMID: 34461981
- Parker BL**, Calkin AC, Seldin MM, Keating MF, Tarling EJ, Yang P, Moody SC, Liu Y, Zerenturk EJ, Needham EJ, Miller ML, Clifford BL, Morand P, Watt MJ, Meex RCR, Peng K-Y, Lee R, Jayawardana K, Pan C, Mellett NA, et al. 2019. An integrative systems genetic analysis of mammalian lipid metabolism. *Nature* **567**:187–193. DOI: <https://doi.org/10.1038/s41586-019-0984-y>, PMID: 30814737
- Parks BW**, Nam E, Org E, Kostem E, Norheim F, Hui ST, Pan C, Civelek M, Rau CD, Bennett BJ, Mehrabian M, Ursell LK, He A, Castellani LW, Zinker B, Kirby M, Drake TA, Drevon CA, Knight R, Gargalovic P, et al. 2013. Genetic control of obesity and gut microbiota composition in response to high-fat, high-sucrose diet in mice. *Cell Metabolism* **17**:141–152. DOI: <https://doi.org/10.1016/j.cmet.2012.12.007>, PMID: 23312289

- Peirce JL**, Lu L, Gu J, Silver LM, Williams RW. 2004. A new set of BXD recombinant inbred lines from advanced intercross populations in mice. *BMC Genetics* **5**:7. DOI: <https://doi.org/10.1186/1471-2156-5-7>, PMID: 15117419
- Piccolo SR**, Denny P, Luxton-Reilly A, Payne S, Ridge PG. 2023. Many Bioinformatics Programming Tasks Can Be Automated with ChatGPT. *arXiv*. <https://arxiv.org/abs/2303.13528>
- Pravenec M**, Klír P, Kren V, Zicha J, Kunes J. 1989. An analysis of spontaneous hypertension in spontaneously hypertensive rats by means of new recombinant inbred strains. *Journal of Hypertension* **7**:217–221 PMID: 2708818.
- Pravenec M**, Křen V, Landa V, Mlejnek P, Musilová A, Šilhavý J, Šimáková M, Zídek V. 2014. Recent progress in the genetics of spontaneously hypertensive rats. *Physiological Research* **63**:S1–S8. DOI: <https://doi.org/10.33549/physiolres.932622>, PMID: 24564651
- Pravenec M**, Saba LM, Zídek V, Landa V, Mlejnek P, Šilhavý J, Šimáková M, Strnad H, Trnovská J, Škop V, Hüttl M, Marková I, Oliyarnyk O, Malínská H, Kazdová L, Smith H, Tabakoff B. 2018. Systems genetic analysis of brown adipose tissue function. *Physiological Genomics* **50**:52–66. DOI: <https://doi.org/10.1152/physiolgenomics.00091.2017>, PMID: 29127223
- Presson AP**, Sobel EM, Papp JC, Suarez CJ, Whistler T, Rajeevan MS, Vernon SD, Horvath S. 2008. Integrated weighted gene co-expression network analysis with an application to chronic fatigue syndrome. *BMC Systems Biology* **2**:95. DOI: <https://doi.org/10.1186/1752-0509-2-95>, PMID: 18986552
- Rau CD**, Wang J, Avetisyan R, Romay MC, Martin L, Ren S, Wang Y, Lusic AJ. 2015. Mapping genetic contributions to cardiac pathology induced by Beta-adrenergic stimulation in mice. *Circulation. Cardiovascular Genetics* **8**:40–49. DOI: <https://doi.org/10.1161/CIRCGENETICS.113.000732>, PMID: 25480693
- Rhoads A**, Au KF. 2015. PacBio sequencing and its applications. *Genomics, Proteomics & Bioinformatics* **13**:278–289. DOI: <https://doi.org/10.1016/j.gpb.2015.08.002>, PMID: 26542840
- Riordan JD**, Nadeau JH. 2017. From peas to disease: modifier genes, network resilience, and the genetics of health. *American Journal of Human Genetics* **101**:177–191. DOI: <https://doi.org/10.1016/j.ajhg.2017.06.004>, PMID: 28777930
- Roden DM**, Pulley JM, Basford MA, Bernard GR, Clayton EW, Balsers JR, Masys DR. 2008. Development of a large-scale de-identified DNA biobank to enable personalized medicine. *Clinical Pharmacology and Therapeutics* **84**:362–369. DOI: <https://doi.org/10.1038/clpt.2008.89>, PMID: 18500243
- Saba LM**, Flink SC, Vanderlinden LA, Israel Y, Tampier L, Colombo G, Kiiamaa K, Bell RL, Printz MP, Flodman P, Koob G, Richardson HN, Lombardo J, Hoffman PL, Tabakoff B. 2015. The sequenced rat brain transcriptome--its use in identifying networks predisposing alcohol consumption. *The FEBS Journal* **282**:3556–3578. DOI: <https://doi.org/10.1111/febs.13358>, PMID: 26183165
- Salas LA**, Zhang Z, Koestler DC, Butler RA, Hansen HM, Molinaro AM, Wiencke JK, Kelsey KT, Christensen BC. 2022. Enhanced cell deconvolution of peripheral blood using DNA methylation for high-resolution immune profiling. *Nature Communications* **13**:761. DOI: <https://doi.org/10.1038/s41467-021-27864-7>, PMID: 35140201
- Sanderson E**, Davey Smith G, Windmeijer F, Bowden J. 2019. An examination of multivariable Mendelian randomization in the single-sample and two-sample summary data settings. *International Journal of Epidemiology* **48**:713–727. DOI: <https://doi.org/10.1093/ije/dyy262>, PMID: 30535378
- Sasani TA**, Ashbrook DG, Beichman AC, Lu L, Palmer AA, Williams RW, Pritchard JK, Harris K. 2022. A natural mutator allele shapes mutation spectrum variation in mice. *Nature* **605**:497–502. DOI: <https://doi.org/10.1038/s41586-022-04701-5>, PMID: 35545679
- Saul MC**, Philip VM, Reinholdt LG, Chesler EJ, Center for Systems Neurogenetics of Addiction. 2019. High-diversity mouse populations for complex traits. *Trends in Genetics* **35**:501–514. DOI: <https://doi.org/10.1016/j.tig.2019.04.003>, PMID: 31133439
- Schadt EE**, Lamb J, Yang X, Zhu J, Edwards S, Guhathakurta D, Sieberts SK, Monks S, Reitman M, Zhang C, Lum PY, Leonardson A, Thieringer R, Metzger JM, Yang L, Castle J, Zhu H, Kash SF, Drake TA, Sachs A, et al. 2005. An integrative genomics approach to infer causal associations between gene expression and disease. *Nature Genetics* **37**:710–717. DOI: <https://doi.org/10.1038/ng1589>, PMID: 15965475
- Scheffler M**, Aeschlimann M, Albrecht M, Bereau T, Bungartz H-J, Felsler C, Greiner M, Groß A, Koch CT, Kremer K, Nagel WE, Scheidgen M, Wöll C, Draxl C. 2022. FAIR data enabling new horizons for materials research. *Nature* **604**:635–642. DOI: <https://doi.org/10.1038/s41586-022-04501-x>, PMID: 35478233
- Schughart K**, Williams RW. 2017. Systems Genetics. *Systems Genetics: Methods and Protocols* New York, NY: Springer. DOI: <https://doi.org/10.1007/978-1-4939-6427-7>
- Schwartz MLB**, McCormick CZ, Lazzeri AL, Lindbuchler DM, Hallquist MLG, Manickam K, Buchanan AH, Rahm AK, Giovanni MA, Frisbie L, Flansburg CN, Davis FD, Sturm AC, Nicastro C, Lebo MS, Mason-Suares H, Mahanta LM, Carey DJ, Williams JL, Williams MS, et al. 2018. A model for genome-first care: returning secondary genomic findings to participants and their healthcare providers in a large research cohort. *American Journal of Human Genetics* **103**:328–337. DOI: <https://doi.org/10.1016/j.ajhg.2018.07.009>, PMID: 30100086
- Seldin MM**, Koplev S, Rajbhandari P, Vergnes L, Rosenberg GM, Meng Y, Pan C, Phuong TMN, Gharakhanian R, Che N, Mäkinen S, Shih DM, Civelek M, Parks BW, Kim ED, Norheim F, Chella Krishnan K, Hasin-Brumshtein Y, Mehrabian M, Laakso M, et al. 2018. A strategy for discovery of endocrine interactions with application to whole-body metabolism. *Cell Metabolism* **27**:1138–1155. DOI: <https://doi.org/10.1016/j.cmet.2018.03.015>, PMID: 29719227
- Seldin M**, Yang X, Lusic AJ. 2019. Systems genetics applications in metabolism research. *Nature Metabolism* **1**:1038–1050. DOI: <https://doi.org/10.1038/s42255-019-0132-x>, PMID: 32259026

- Senko AN**, Overall RW, Silhavy J, Mlejnek P, Malínská H, Hüttl M, Marková I, Fabel KS, Lu L, Stuchlik A, Williams RW, Pravenec M, Kempermann G. 2022. Systems genetics in the rat HXB/BXH family identifies *Tti2* as a pleiotropic quantitative trait gene for adult hippocampal neurogenesis and serum glucose. *PLOS Genetics* **18**:e1009638. DOI: <https://doi.org/10.1371/journal.pgen.1009638>, PMID: 35377872
- Seroussi U**, Lugowski A, Wadi L, Lao RX, Willis AR, Zhao W, Sundby AE, Charlesworth AG, Reinke AW, Claycomb JM. 2023. A comprehensive survey of *C. elegans* argonaute proteins reveals organism-wide gene regulatory networks and functions. *eLife* **12**:e83853. DOI: <https://doi.org/10.7554/eLife.83853>, PMID: 36790166
- Shu L**, Zhao Y, Kurt Z, Byars SG, Tukiainen T, Kettunen J, Orozco LD, Pellegrini M, Lusi AJ, Ripatti S, Zhang B, Inouye M, Mäkinen V-P, Yang X. 2016. Mergeomics: multidimensional data integration to identify pathogenic perturbations to biological systems. *BMC Genomics* **17**:874. DOI: <https://doi.org/10.1186/s12864-016-3198-9>, PMID: 27814671
- Shu L**, Chan KHK, Zhang G, Huan T, Kurt Z, Zhao Y, Codoni V, Trégouët D-A, Yang J, Wilson JG, Luo X, Levy D, Lusi AJ, Liu S, Yang X, Cardiogenics Consortium. 2017. Shared genetic regulatory networks for cardiovascular disease and type 2 diabetes in multiple populations of diverse ethnicities in the United States. *PLOS Genetics* **13**:e1007040. DOI: <https://doi.org/10.1371/journal.pgen.1007040>, PMID: 28957322
- Shue E**, Liu L, Li B, Feng Z, Li X, Hu G. 2023. Empowering beginners in bioinformatics with ChatGPT. *Quantitative Biology* **11**:105–108. DOI: <https://doi.org/10.15302/j-qb-023-0327>, PMID: 37378043
- Singer GAC**, Lloyd AT, Huminiecki LB, Wolfe KH. 2005. Clusters of co-expressed genes in mammalian genomes are conserved by natural selection. *Molecular Biology and Evolution* **22**:767–775. DOI: <https://doi.org/10.1093/molbev/msi062>, PMID: 15574806
- Sittig LJ**, Carbonetto P, Engel KA, Krauss KS, Barrios-Camacho CM, Palmer AA. 2016. Genetic background limits generalizability of genotype-phenotype relationships. *Neuron* **91**:1253–1259. DOI: <https://doi.org/10.1016/j.neuron.2016.08.013>, PMID: 27618673
- Skelly DA**, Czechanski A, Byars C, Aydin S, Spruce C, Olivier C, Choi K, Gatti DM, Raghupathy N, Keele GR, Stanton A, Vincent M, Dion S, Greenstein I, Pankratz M, Porter DK, Martin W, O'Connor C, Qin W, Harrill AH, et al. 2020. Mapping the effects of genetic variation on chromatin state and gene expression reveals loci that control ground state pluripotency. *Cell Stem Cell* **27**:459–469. DOI: <https://doi.org/10.1016/j.stem.2020.07.005>, PMID: 32795400
- Smith GD**, Ebrahim S. 2003. “Mendelian randomization”: can genetic epidemiology contribute to understanding environmental determinants of disease? *International Journal of Epidemiology* **32**:1–22. DOI: <https://doi.org/10.1093/ije/dyg070>, PMID: 12689998
- Solberg Woods LC**, Palmer AA. 2019. Using heterogeneous stocks for fine-mapping genetically complex traits. *Methods in Molecular Biology* **2018**:233–247. DOI: https://doi.org/10.1007/978-1-4939-9581-3_11, PMID: 31228160
- Stark R**, Grzelak M, Hadfield J. 2019. RNA sequencing: the teenage years. *Nature Reviews. Genetics* **20**:631–656. DOI: <https://doi.org/10.1038/s41576-019-0150-2>, PMID: 31341269
- Steinbock LJ**, Radenovic A. 2015. The emergence of nanopores in next-generation sequencing. *Nanotechnology* **26**:074003. DOI: <https://doi.org/10.1088/0957-4484/26/7/074003>, PMID: 25643284
- Stoeger T**, Gerlach M, Morimoto RI, Nunes Amaral LA, Freeman T. 2018. Large-scale investigation of the reasons why potentially important genes are ignored. *PLOS Biology* **16**:e2006643. DOI: <https://doi.org/10.1371/journal.pbio.2006643>, PMID: 30226837
- Svenson KL**, Gatti DM, Valdar W, Welsh CE, Cheng R, Chesler EJ, Palmer AA, McMillan L, Churchill GA. 2012. High-resolution genetic mapping using the Mouse Diversity outbred population. *Genetics* **190**:437–447. DOI: <https://doi.org/10.1534/genetics.111.132597>, PMID: 22345611
- Tabakoff B**, Smith H, Vanderlinden LA, Hoffman PL, Saba LM. 2019. Networking in biology: the hybrid rat diversity panel. *Methods in Molecular Biology* **2018**:213–231. DOI: https://doi.org/10.1007/978-1-4939-9581-3_10, PMID: 31228159
- Tabassum R**, Rämö JT, Ripatti P, Koskela JT, Kurki M, Karjalainen J, Palta P, Hassan S, Nunez-Fontarnau J, Kiiskinen TTJ, Söderlund S, Matikainen N, Gerl MJ, Surma MA, Klose C, Stitzel NO, Laivuori H, Havulinna AS, Service SK, Salomaa V, et al. 2019. Genetic architecture of human plasma lipidome and its link to cardiovascular disease. *Nature Communications* **10**:4329. DOI: <https://doi.org/10.1038/s41467-019-11954-8>, PMID: 31551469
- Tang WHW**, Bäckhed F, Landmesser U, Hazen SL. 2019. Intestinal Microbiota in Cardiovascular Health and Disease: JACC State-of-the-Art Review. *Journal of the American College of Cardiology* **73**:2089–2105. DOI: <https://doi.org/10.1016/j.jacc.2019.03.024>, PMID: 31023434
- Taylor BA**, Heiniger HJ, Meier H. 1973. Genetic analysis of resistance to cadmium-induced testicular damage in mice. *Experimental Biology and Medicine* **143**:629–633. DOI: <https://doi.org/10.3181/00379727-143-37380>
- Tuominen I**, Fuqua BK, Pan C, Renaud N, Wroblewski K, Civelek M, Clerkin K, Asaryan A, Haroutunian SG, Loureiro J, Borawski J, Roma G, Knehr J, Carbone W, Franch S, Parks BW, Hui ST, Mehrabian M, Magyar C, Cantor RM, et al. 2021. The genetic architecture of carbon tetrachloride-induced liver fibrosis in mice. *Cellular and Molecular Gastroenterology and Hepatology* **11**:199–220. DOI: <https://doi.org/10.1016/j.jcmgh.2020.08.010>, PMID: 32866618
- Vanderlinden LA**, Saba LM, Printz MP, Flodman P, Koob G, Richardson HN, Hoffman PL, Tabakoff B. 2014. Is the alcohol deprivation effect genetically mediated? Studies with HXB/BXH recombinant inbred rat strains. *Alcoholism, Clinical and Experimental Research* **38**:2148–2157. DOI: <https://doi.org/10.1111/acer.12471>, PMID: 24961585

- Voigt B**, Kuramoto T, Mashimo T, Tsurumi T, Sasaki Y, Hokao R, Serikawa T. 2008. Evaluation of LEXF/FXLE rat recombinant inbred strains for genetic dissection of complex traits. *Physiological Genomics* **32**:335–342. DOI: <https://doi.org/10.1152/physiolgenomics.00158.2007>, PMID: 18042833
- von Scheidt M**, Zhao Y, Kurt Z, Pan C, Zeng L, Yang X, Schunkert H, Lusic AJ. 2017. Applications and limitations of mouse models for understanding human atherosclerosis. *Cell Metabolism* **25**:248–261. DOI: <https://doi.org/10.1016/j.cmet.2016.11.001>, PMID: 27916529
- Wang X**, Pandey AK, Mulligan MK, Williams EG, Mozhui K, Li Z, Jovaisaite V, Quarles LD, Xiao Z, Huang J, Capra JA, Chen Z, Taylor WL, Bastarache L, Niu X, Pollard KS, Ciobanu DC, Reznik AO, Tishkov AV, Zhulin IB, et al. 2016. Joint mouse-human phenome-wide association to test gene function and disease risk. *Nature Communications* **7**:10464. DOI: <https://doi.org/10.1038/ncomms10464>, PMID: 26833085
- Wang D**, Eraslan B, Wieland T, Hallström B, Hopf T, Zolg DP, Zecha J, Asplund A, Li L-H, Meng C, Frejno M, Schmidt T, Schnatbaum K, Wilhelm M, Ponten F, Uhlen M, Gagneur J, Hahne H, Kuster B. 2019. A deep proteome and transcriptome abundance atlas of 29 healthy human tissues. *Molecular Systems Biology* **15**:e8503. DOI: <https://doi.org/10.15252/msb.20188503>, PMID: 30777892
- Weissman AM**. 1994. The T-cell antigen receptor: a multisubunit signaling complex. *Chemical Immunology* **59**:1–18 PMID: 7945923.
- Wilkinson MD**, Dumontier M, Aalbersberg IJJ, Appleton G, Axton M, Baak A, Blomberg N, Boiten J-W, da Silva Santos LB, Bourne PE, Bouwman J, Brookes AJ, Clark T, Crosas M, Dillo I, Dumon O, Edmunds S, Evelo CT, Finkers R, Gonzalez-Beltran A, et al. 2016. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data* **3**:160018. DOI: <https://doi.org/10.1038/sdata.2016.18>, PMID: 26978244
- Williams RW**, Williams EG. 2017. Resources for Systems Genetics. *Methods in Molecular Biology* **1488**:3–29. DOI: https://doi.org/10.1007/978-1-4939-6427-7_1, PMID: 27933518
- Williams EG**, Pfister N, Roy S, Statzer C, Haverty J, Ingels J, Bohl C, Hasan M, Čuklina J, Bühlmann P, Zamboni N, Lu L, Ewald CY, Williams RW, Aebersold R. 2022. Multiomic profiling of the liver across diets and age in a diverse mouse population. *Cell Systems* **13**:43–57. DOI: <https://doi.org/10.1016/j.cels.2021.09.005>, PMID: 34666007
- Wishart DS**. 2019. Metabolomics for investigating physiological and pathophysiological processes. *Physiological Reviews* **99**:1819–1875. DOI: <https://doi.org/10.1152/physrev.00035.2018>, PMID: 31434538
- Woolf B**, Zagkos L, Gill D. 2022. TwoStepCisMR: A Novel Method and R Package for Attenuating Bias in cis-Mendelian Randomization Analyses. *Genes* **13**:1541. DOI: <https://doi.org/10.3390/genes13091541>, PMID: 36140709
- Wright SN**, Leger BS, Rosenthal SB, Liu SN, Jia T, Chitre AS, Poleskaya O, Holl K, Gao J, Cheng R, Garcia Martinez A, George A, Gileta AF, Han W, Netzley AH, King CP, Lamparelli A, Martin C, St Pierre CL, Wang T, et al. 2023. Genome-wide association studies of human and rat BMI converge on synapse, epigenome, and hormone signaling networks. *Cell Reports* **42**:112873. DOI: <https://doi.org/10.1016/j.celrep.2023.112873>, PMID: 37527041
- Xiao H**, Bozi LHM, Sun Y, Riley CL, Philip VM, Chen M, Li J, Zhang T, Mills EL, Emont MP, Sun W, Reddy A, Garrity R, Long J, Becher T, Vitas LP, Laznik-Bogoslavski D, Ordonez M, Liu X, Chen X, et al. 2022. Architecture of the outbred brown fat proteome defines regulators of metabolic physiology. *Cell* **185**:4654–4673. DOI: <https://doi.org/10.1016/j.cell.2022.10.003>, PMID: 36334589
- Yu Q**, Liu X, Keller MP, Navarrete-Perea J, Zhang T, Fu S, Vaites LP, Shuken SR, Schmid E, Keele GR, Li J, Huttlin EL, Rashan EH, Simcox J, Churchill GA, Schweppe DK, Attie AD, Paulo JA, Gygi SP. 2023. Sample multiplexing-based targeted pathway proteomics with real-time analytics reveals the impact of genetic variation on protein expression. *Nature Communications* **14**:555. DOI: <https://doi.org/10.1038/s41467-023-36269-7>, PMID: 36732331
- Zeng P**, Shao Z, Zhou X. 2021. Statistical methods for mediation analysis in the era of high-throughput genomics: Current successes and future challenges. *Computational and Structural Biotechnology Journal* **19**:3209–3224. DOI: <https://doi.org/10.1016/j.csbj.2021.05.042>, PMID: 34141140
- Zhang Q**, Mesner LD, Calabrese GM, Dirckx N, Li Z, Verardo A, Yang Q, Tower RJ, Faugere MC, Farber CR, Clemens TL. 2021a. Genomic variants within chromosome 14q32.32 regulate bone mass through MARK3 signaling in osteoblasts. *The Journal of Clinical Investigation* **131**:e142580. DOI: <https://doi.org/10.1172/JCI142580>, PMID: 33792563
- Zhang K**, Hocker JD, Miller M, Hou X, Chiou J, Poirion OB, Qiu Y, Li YE, Gaulton KJ, Wang A, Preissl S, Ren B. 2021b. A single-cell atlas of chromatin accessibility in the human genome. *Cell* **184**:5985–6001. DOI: <https://doi.org/10.1016/j.cell.2021.10.024>, PMID: 34774128
- Zhao Y**, Chen J, Freudenberg JM, Meng Q, Rajpal DK, Yang X. 2016. Network-based identification and prioritization of key regulators of coronary artery disease loci. *Arteriosclerosis, Thrombosis, and Vascular Biology* **36**:928–941. DOI: <https://doi.org/10.1161/ATVBAHA.115.306725>, PMID: 26966275
- Zhou X**, Crow AL, Hartiala J, Spindler TJ, Ghazalpour A, Barsky LW, Bennett BJ, Parks BW, Eskin E, Jain R, Epstein JA, Lusic AJ, Adams GB, Allayee H. 2015. The genetic landscape of hematopoietic stem cell frequency in mice. *Stem Cell Reports* **5**:125–138. DOI: <https://doi.org/10.1016/j.stemcr.2015.05.008>, PMID: 26050929
- Zhou W**, Kanai M, Wu K-HH, Rasheed H, Tsuo K, Hirbo JB, Wang Y, Bhattacharya A, Zhao H, Namba S, Surakka I, Wolford BN, Lo Faro V, Lopera-Maya EA, Läll K, Favé M-J, Partanen JJ, Chapman SB, Karjalainen J, Kurki M, et al. 2022. Global biobank meta-analysis initiative: powering genetic discovery across human disease. *Cell Genomics* **2**:100192. DOI: <https://doi.org/10.1016/j.xgen.2022.100192>, PMID: 36777996

- Zhu J**, Lum PY, Lamb J, GuhaThakurta D, Edwards SW, Thieringer R, Berger JP, Wu MS, Thompson J, Sachs AB, Schadt EE. 2004. An integrative genomics approach to the reconstruction of gene networks in segregating populations. *Cytogenetic and Genome Research* **105**:363–374. DOI: <https://doi.org/10.1159/000078209>, PMID: [15237224](https://pubmed.ncbi.nlm.nih.gov/15237224/)
- Zhu Z**, Zhang F, Hu H, Bakshi A, Robinson MR, Powell JE, Montgomery GW, Goddard ME, Wray NR, Visscher PM, Yang J. 2016. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nature Genetics* **48**:481–487. DOI: <https://doi.org/10.1038/ng.3538>, PMID: [27019110](https://pubmed.ncbi.nlm.nih.gov/27019110/)