



OPEN ACCESS

EDITED BY

Juan Carlos Fernandez-Lopez,
National Institute of Genomic Medicine
(INMEGEN), Mexico

REVIEWED BY

Qian Zhang,
Wellcome Sanger Institute (WT),
United Kingdom
Karol Estrada,
Brandeis University, United States

*CORRESPONDENCE

Struan F. A. Grant,
✉ grants@chop.edu

RECEIVED 23 January 2024

ACCEPTED 24 September 2024

PUBLISHED 03 October 2024

CITATION

Bruner WS and Grant SFA (2024) Translation of
genome-wide association study: from genomic
signals to biological insights.
Front. Genet. 15:1375481.
doi: 10.3389/fgene.2024.1375481

COPYRIGHT

© 2024 Bruner and Grant. This is an open-
access article distributed under the terms of the
[Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/).
The use, distribution or reproduction in other
forums is permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication in this
journal is cited, in accordance with accepted
academic practice. No use, distribution or
reproduction is permitted which does not
comply with these terms.

Translation of genome-wide association study: from genomic signals to biological insights

Winter S. Bruner^{1,2} and Struan F. A. Grant^{1,2,3,4,5*}

¹Center for Spatial and Functional Genomics, Children's Hospital of Philadelphia, Philadelphia, PA, United States, ²Division of Human Genetics, Children's Hospital of Philadelphia, Philadelphia, PA, United States, ³Department of Genetics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, United States, ⁴Department of Pediatrics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, United States, ⁵Division of Endocrinology and Diabetes, Children's Hospital of Philadelphia, Philadelphia, PA, United States

Since the turn of the 21st century, genome-wide association study (GWAS) have successfully identified genetic signals associated with a myriad of common complex traits and diseases. As we transition from establishing robust genetic associations with diverse phenotypes, the central challenge is now focused on characterizing the underlying functional mechanisms driving these signals. Previous GWAS efforts have revealed multiple variants, each conferring relatively subtle susceptibility, collectively contributing to the pathogenesis of various common diseases. Such variants can further exhibit associations with multiple other traits and differ across ancestries, plus disentangling causal variants from non-causal due to linkage disequilibrium complexities can lead to challenges in drawing direct biological conclusions. Combined with cellular context considerations, such challenges can reduce the capacity to definitively elucidate the biological significance of GWAS signals, limiting the potential to define mechanistic insights. This review will detail current and anticipated approaches for functional interpretation of GWAS signals, both in terms of characterizing the underlying causal variants and the corresponding effector genes.

KEYWORDS

GWAS, genome-wide association study, snps (single nucleotide polymorphisms), variant to gene, complex traits genetics., function

Introduction

The pathogenesis of common, complex human traits and diseases emerge as a consequence of the interplay between environmental and genetic factors. To uncover the genetic underpinnings of such traits, studies have successfully employed genome-wide association study (GWAS) to identify susceptibility loci. When a GWAS is conducted, differences in allele frequencies across hundreds of thousands to millions of single nucleotide polymorphisms (SNPs) assayed in one experiment are assessed by comparing the genotypes of individuals with and without a trait of interest (dichotomous), such as asthma, or treating the trait as a continuous variable (quantitative), such as body mass index. One can identify loci associated with a specific disease or trait of interest by evaluating allelic frequency differences that remain statistically significant after correcting for the large degree of multiple comparisons across the genome. This method crucially relies on linkage disequilibrium (LD) to inform the analysis, which can readily aid in identifying associated genetic loci; however, this same factor can limit

implications of the actual underlying causal functional variants driving the pathogenesis of the phenotype of interest. For this reason, various tools and study designs have been leveraged to carry out GWAS follow-up studies to uncover which variants are causal for complex traits, along with implicating the corresponding effector genes.

Although GWAS has proven successful in uncovering trait-associated genetic susceptibility loci, ranging from breast cancer to migraine to type 2 diabetes (Bradfield et al., 2012; Bradfield et al., 2019; Easton et al., 2007; Papaemmanuil et al., 2009; Xue et al., 2018; Genome-wide association study identifies new, 2009), there are associated challenges with the overall study design. The ability to obtain statistical power of 80% or more for genetic associations stems from the ability to recruit a sufficient sample size for the GWAS study, which can often prove challenging (more information on GWAS sample size and cohort-based replication studies can be found elsewhere) (Uffelmann et al., 2021). Low sample size and its impact on statistical power contributes to type I and II errors, directly and negatively impacting downstream follow-up studies (Serdar et al., 2021; Banerjee et al., 2009; Krzywinski and Altman, 2013). Typically, collaboration is required to meet such high demands for appropriate sample sizes for statistical power and allowing for the opportunity to replicate initial findings within independent datasets. Additionally, large-scale collaboration efforts lend themselves to subgroup analysis, allowing for additional investigation of complex diseases and traits. With independent and worldwide genomic data collection sites, incorporating different ancestral data collectively can be accomplished through trans-ethnic meta-analysis. Further subgroup analysis can be accomplished based on age, sex, or other dichotomous characteristics to find novel loci for further function follow-up studies. This has been successfully carried out in complex diseases or disorders such as childhood obesity (Bradfield et al., 2012; Bradfield et al., 2019), body mass index (Akiyama et al., 2017), migraines (Anttila et al., 2013), to name a few.

Additionally, GWAS has to account for population-biased findings. Since the allele frequencies used for comparisons often originate from European ancestry, findings from GWAS efforts often need to be more representative across various ancestral groups, resulting in replication challenges across populations (Peterson et al., 2019). As such, combined with remaining power challenges, GWAS is still limited in addressing a large portion of the 'missing heritability' (Manolio et al., 2009; Matthews and Turkheimer, 2022) for common complex traits. Furthermore, GWAS are often performed with SNP array data heavily biased towards common variants ($MAF \geq 5\%$) (Momozawa and Mizukami, 2021; Gibson, 2012). This subsequently limits the potential findings of casual rare variants ($MAF < 1\%$) (Momozawa and Mizukami, 2021; Gibson, 2012; Wainschtein et al., 2022). As more studies include increasingly larger sample sizes from diverse ancestry and include better imputation panels, the degree of missing heritability remaining to be characterized should narrow (Momozawa and Mizukami, 2021; Gibson, 2012; Wainschtein et al., 2022).

The results of a GWAS are also limited to simply detecting genetic signals. Indeed, such signals themselves cannot pinpoint the true causal variant (s) in LD with the SNP producing the overall lowest *P*-value. This means that the causal variant is not typically

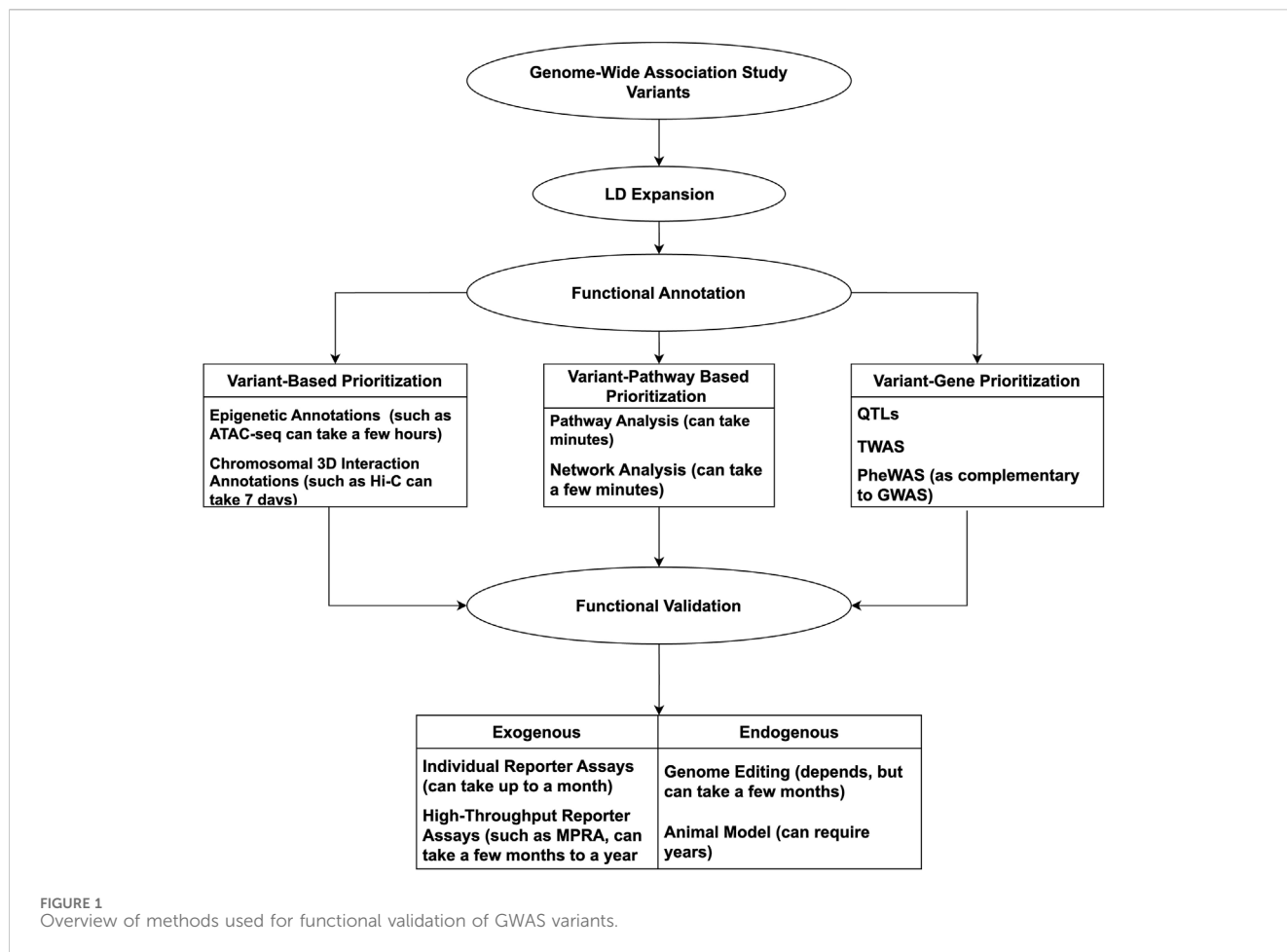
assayed directly in the given genotyping assay. Additionally, given the usual polygenicity of common complex traits, the magnitude of each signal is relatively small, with only the additive effects of loci driving the overall genetic etiology of the phenotype of interest. Furthermore, genetic effects are often cell-type specific. As such, determining which cell or tissue type is impacted by GWAS loci has often proven arduous. Together with phenotype heterogeneity, these features of GWAS make mechanistic follow-up analyses challenging.

Although multiple examples of GWAS functionalization attempts exist, one of the most noteworthy examples is at the *FTO* obesity locus (Frayling et al., 2007). This very robust association signal located within an intronic region of the *FTO* gene has been widely replicated across different studies involving different ethnicities (Hassanein et al., 2010; Okada et al., 2012; Wen et al., 2012; Loos and Yeo, 2014) and age groups (Bradfield et al., 2012; Bradfield et al., 2019; Grant et al., 2008; Felix et al., 2016; Elks et al., 2010; Elks et al., 2012). Although there are hundreds of studies validating this association signal with obesity risk, it is becoming clear that the *FTO* gene itself may not be the causal effector gene at this key associated signal. Research assessing the genomic interactions at this locus found a direct contact between the *FTO* intronic region harboring the genetic signal and the *IRX3* gene (Smemo et al., 2014). This led to the conclusion that there is an enhancer imbedded within the *FTO* intron directly influencing the regulation of the neighboring *IRX3* gene (Smemo et al., 2014). Additional work in primary adipocytes further showed that knockout of *IRX3*, together with the next gene along *IRX5*, directly impacted thermogenetic properties and, by extension, demonstrated their role in obesity (Claussnitzer et al., 2015). This is a key example of why GWAS follow-up studies can be time-consuming. Despite such difficulties, establishing casual genetic influences from a GWAS is attainable, especially when incorporating available public resources and cutting-edge techniques (discussed below) to enable important follow-up study designs to reveal crucial novel biological insights.

Considering the complexities of GWAS, from the plethora of data to the underlying complex gene-gene/gene-environment interactions, following up on potential leads can appear daunting. Computational tools and technologies can be incorporated to offset such constraints, having proven successful and timely. With so many advances, it is timely to review the available resources to conduct a GWAS mechanistic 'variant-to-function' (V2F) follow-up successfully. As such, we will highlight methods and techniques involved in GWAS V2F studies, emphasizing more recent high-throughput methods. An overview of the discussed tools and techniques used for GWAS follow-up can be seen in Figure 1.

Genetic signal follow-up strategies

Non-coding variants represent over 90% of GWAS reports, which is a large contributing factor to why GWAS follow-up can be so arduous (Schipper and Posthuma, 2022). Although variants can also be found within a gene coding region it potentially regulates, it is still imperative to consider variants in LD that may still reside in non-coding regions (McCarthy et al., 2008).



Challenges in understanding the genotype to phenotype relationship resulting from the putative regulation of a noncoding variant, requires connection between an associated variant to gene(s) regulation, and by extension tissue site(s) of mechanistic action. Based on the association alone, researchers cannot determine which variants are causal, the putative gene effector target(s), or tissue-specific involvement. This is where incorporating previously generated data, such as fine-mapping, functional annotations, and a combination with multi-omics approaches plays a role in elucidating the overall biological underpinnings from a GWAS-nominated variant. Given that variant-to-gene methods need to be conducted in a human setting in the first instance, animal models can only be leveraged subsequently, albeit often successfully, once such leads are determined (Palermo et al., 2023; Soleimanpour et al., 2014; Srivastava et al., 2019).

Variant/gene prioritization approaches and tools

With the myriad of signals discovered by GWAS, narrowing down variants with a higher probability of being causal becomes necessary. Signals detected by GWAS typically do not necessarily represent the causal variant for a given phenotype, but rather represent a tag-SNP in LD with the underlying causal variant(s).

Indeed, understanding the underlying LD structure is the initial step in making sense of GWAS signals. Comprehensively assessing both tag-SNPs and their LD proxies facilitates the acquisition of the true causal variant(s) in any functional follow-up approaches (Schaid et al., 2018; Raychaudhuri, 2011). This can be achieved by incorporating fine-mapping into the study design to aid in the prioritization of candidate causal variants by considering both LD patterns and association statistics.

Fine-mapping helps narrow down a list of GWAS signals through a combination of statistical approaches, in conjunction with functional annotations. Types of statistical models used for fine-mapping include Bayesian-based methods (Schaid et al., 2018; de los Campos et al., 2023), heuristic (Schaid et al., 2018), penalized regression (Schaid et al., 2018), and conditional association analysis (Kocarnik et al., 2018). When wanting to obtain more information about more than one SNP collectively producing an effect, the Bayesian method may prove most beneficial (Schaid et al., 2018). Many approaches are available that include the Bayesian-based method for fine-mapping, such as CARMA (Yang et al., 2023a), SUSIE (Wang et al., 2020), PAINTOR (Kichaev et al., 2014), CAVIAR (Wang et al., 2020), etc. Following fine-mapping efforts, an assessment of the genomic landscape can be conducted to determine which variants reside in genomic regions areas which accessible and potentially functional. This can be accomplished through various annotation methods described below.

Work was recently conducted in Alzheimer's Disease (AD) incorporating variant prioritization via three fine-mapping approaches (Bayesian, FINEMAP, and PAINTOR) in conjunction with annotation-based data from primary microglia and iPSC-derived macrophages (Schwartzentruber et al., 2021). This work revealed 21 variants being prioritized as most probable (>50%) for causality, and an additional 79 variants within the 10%–50% potential variants of interest (Schwartzentruber et al., 2021). Some prioritized SNPs were close to already established AD genes, like *BINI*(43). In addition to these known genes, there were new AD risk genes uncovered through leveraging fine-mapped SNPs (Schwartzentruber et al., 2021). Such variant-to-gene pairs included rs143080277 in *NCK2*, rs2830489 near *ADAMTS1*, and rs268120 in *SPRED2*(43). Similarly, fine-mapping-based work in chronic lymphocytic leukemia (CLL) led to successful GWAS V2F follow-up (Slager et al., 2013). Using genotype data from over 2000 participants, fine-mapping was conducted (Slager et al., 2013). Results revealed a functional connection between rs1044873 and the *IRF8* gene (Slager et al., 2013).

Functional annotation methods

Epigenetic and chromosomal-based techniques

Variant annotation methods that incorporate epigenetic data have become a standard approach to elucidating the functional consequences of genetic variants. Assessment of DNA accessibility can be carried out in multiple cell types to reveal underlying gene regulatory roles of chromatin organization in a given cell type, which in turn suggests how such activity confers its trait effects. Integrating various publicly available resources can further aid the prioritization of GWAS-identified variants. Assay for transposase-accessible chromatin sequencing (ATAC-seq) (Buenrostro et al., 2015) has proven to be an efficacious assay in assessing GWAS variants and functionality for complex diseases. A study focusing on type 2 diabetes (T2D) determined that the open chromatin landscape in human pancreatic islets cells differed between samples obtained from individuals with and without T2D (46). A total of 13 T2D associated SNPs were found in regions marked within open chromatin accessibility sites which were located near genes such as *TCF7L2*, *ADCY5*, and *GCH* (Bysani et al., 2019). In addition to the 13 SNPs, there were 67 SNPs that were in LD with T2D associated SNPs and further annotated to known T2D genes (such as *PPARG*, *FTO*, and *KCNJ11*) (Bysani et al., 2019). Another study focused on T2D used a combination of pancreatic cell expression data, chromatin accessibility, and network analysis methods to prioritize the gene *RFX6* from GWAS results for the disease (Walker et al., 2023). This study took a different approach for incorporating ATAC-seq by using this technique to assess the chromatin architecture following the knockdown of the putative T2D causal gene (Walker et al., 2023). Knockdown of *RFX6* in beta-pancreatic cells not only resulted in substantial variation in gene expression, but also provided evidence of dysregulation of regulatory elements harboring T2D GWAS variants via changes in genome-wide chromatin states (Walker et al., 2023).

As such, these annotations help determine a variant's context with respect to a typically non-coding role as a potential regulatory

element. Identifying whether a genetic variant resides in a regulatory element (such as promoter, enhancer, or transcription binding site) can be highly informative when determining subsequent biological consequences with respect to gene expression. Furthermore, assessing epigenetic markers in a cell type-dependent manner can help determine gene regulation specificity, adding important context to cell and tissue involvement for a given complex trait or disease. A popular resource for assessing variant annotations includes the Encyclopedia of DNA Elements (ENCODE) Consortium, which consists of histone modification, expression, and chromatin conformation data across different cell types (Luo et al., 2020).

Histone modifications are an essential aspect of gene regulation by influencing how tightly or loosely DNA is packaged, which indicates areas of the genome that are accessible and available for gene transcription. Acetylation and methylation patterns are histone modifications often leveraged to assess poised or inactive cell-specific chromatin states (Karlič et al., 2010). Histone marks used to determine overall DNA accessibility help determine whether a specific region is accessible to transcription factor (TF) binding, and therefore a potentially functional element. Regions with histone marks like H3K4me3 and H3K27ac are typically defined as promoter/open chromatin regions (Barski et al., 2007; Creighton et al., 2010), while marks like H3K27me3 indicate closed chromatin states (Cai et al., 2021). Such histone modification patterns can be assessed by a number of available techniques, which are explored in greater detail elsewhere (Mansidor and Risca, 2022). Examples of some techniques used to assess chromatin accessibility include not only ATAC-seq (discussed in previous paragraph), but also formaldehyde-Assisted Isolation of Regulatory Elements sequencing (FAIRE-seq) (Giresi et al., 2007), DNase I hypersensitive sites sequencing (DNase-seq) (Song and Crawford, 2010) and sequencing of micrococcal nuclease sensitive sites (MNase-seq) (Deng et al., 2022; Wong et al., 2023). Although many options are available, subtle nuances between the available techniques make some techniques more appealing than others depending on the specific application. Regarding experimental time and output success among the available options, ATAC-seq has become a gold standard, with an approximate three-hour protocol for preparation (Buenrostro et al., 2015; Grandi et al., 2022), where transposases are used with DNA-associated adaptors for subsequent high-throughput sequencing (Buenrostro et al., 2015).

Another significant component of the genomic landscape to consider is chromatin 3D interactions, which offer insight into physical connections between GWAS-implicated candidate causal variants and putative effector genes. Understanding which gene(s) are regulated by non-coding GWAS associated variants is essential to understanding complex traits fully. This is where the power of incorporating chromosomal capture techniques become apparent.

Chromosomal capture techniques involve crosslinking interacting genomic loci to each other, followed by high throughput sequencing (Dixon et al., 2012). The sequencing results generate a map of interacting segments from the genome, which can help nominate candidate effector genes controlled by regulatory elements harboring GWAS associated variants. Various techniques have been developed that can be used for such a purpose,

TABLE 1 Resource examples used in the GWAS V2F follow-up.

Resource	Data access/Usage
Functional ANnotation Of the Mammalian genome (FANTOM) (Lizio et al., 2019; Lizio et al., 2015)	Project FANTOM is committed to producing an atlas of human transcripts and is currently on its 6th iteration with a focus on long noncoding RNAs (lncRNAs) (Ramilowski et al., 2020). Previous FANTOM5 work focused on collecting transcriptome data of cells under different states. Data includes transcriptomics of different cell types over various time course and exposures (Noguchi et al., 2017). https://fantom.gsc.riken.jp/
ENCODE Consortium (Luo et al., 2020)	Available annotations include open chromatin, histone marks, transcription factor binding, gene expression, transcription start site, RNA binding protein occupancy, DNA methylation, 3D chromatin interactions, and topologically associated domains data. https://www.encodeproject.org
Roadmap Epigenomics Mapping Centers (REMC) (Kundaje et al., 2015)	DNase, DNase-seq, histone modification, and RNA-seq data from relatively normal human cell and tissue samples. https://egg2.wustl.edu/roadmap/web_portal/index.html
International Human Epigenome Consortium (IHEC) (Stunnenberg and Hirst, 2016)	Epigenetic profiles of both diseased and unremarkable states with collected mRNA-seq, DNase-seq, WGBS, and histone modification data. Data collected has originated from worldwide projects and include BLUEPRINT, CEEHRC Epigenomic Platform Project, CREST, and DEEP. https://epigenomesportal.ca/ihec/
Human Genome Browser (Kent et al., 2002)	Web-based tool that allows integration of annotated data across entire genomes to more readily visualize. https://genome.ucsc.edu
Galaxy (Galaxy Community, 2022)	Bioinformatics web-based platform with many computational tools for analysis for a variety of reason (e.g., RNA-seq, ChIP-seq, variant calling, etc.) Allows users to upload own data to analyze, as well as incorporate publicly and shared available data. Important to understand local legal guidelines, especially when analyzing human genomic datasets on a web-based platform. https://usegalaxy.org
Open Targets Genetics (Ghousaini et al., 2021; Mountjoy et al., 2021)	Web-based platform that integrates GWAS, various QTL, Hi-C, and DNase Hypersensitivity Sites to aid in causal variant and gene target prioritization. https://genetics.opentargets.org/
Genotype-Tissue Expression project (GTEx) (The Genotype-Tissue Expression GTEx project, 2013)	Both RNA-seq and genotype data is used to provide information on human gene expression by tissue. https://www.gtexportal.org/home/

and indeed TADs can be defined with chromosomal capture techniques (Zufferey et al., 2022).

Inclusion of chromosomal conformation capture techniques to determine gene regulation by GWAS loci has been successful in a multitude of studies involving complex diseases and traits. One study identified 38 candidate genes potentially involved in obesity by incorporating promoter Capture-Hi-C techniques within primary adipocytes (Pan et al., 2018). Following promoter Capture-Hi-C, the study implicated interaction with GWAS SNPs and three additional genes (rs8076131 to *ORMDL3*, rs1017546 and rs3784671 with *LACTB*, and rs10774569 with *ACADS*) (Pan et al., 2018). Use of related techniques led to the discovery of inflammatory bowel syndrome (IBD) target genes that were regulated by 92 regulatory elements previously associated with IBD (Meddens et al., 2016). Our study on systemic lupus erythematosus (SLE) evaluated putative target genes of GWAS signals by combining follicular helper T cells (TFH) (cells with SLE involvement), open chromatin sites, and three-dimensional genomic architecture as defined by Capture C (63). Incorporation of Capture C implicated genes *BCL6* and *CXCR5*, which were previously identified as TFH regulators (Su et al., 2020). The study further assessed the putative target genes via CRISPR/CAS9 genome editing techniques, which revealed key genes important in regulating crucial cytokine involvement in B cell antibody production (Su et al., 2020). Such examples show the power of integrating chromosomal conformation techniques to determine targeted genes by non-coding variants and can lead to a more comprehensive understanding of the genomics driving specific complex traits or diseases.

Studies involving chromatin 3D interactions and genome organization have found regions with increased frequency of contact referred to as topologically associated domains (TADs) (McArthur and Capra, 2021). TADs provide a spatial framework for the genome while facilitating proper gene regulation (McArthur and Capra, 2021), and has been used by investigators to define the search space for defining an underlying effector gene at a given GWAS locus. Variants within noncoding regulatory elements, such as enhancers, localize within three dimensional TADs that are readily assessed through various chromosomal capture methods that assess chromatin 3D interactions (Dixon et al., 2012; McArthur and Capra, 2021). An overview of popular and useful resources for chromatin accessibility data integration can be found in Table 1.

Chromosomal capture techniques including Capture-C (65), ChIA-Pet (Han et al., 2018), Hi-C (59,65), chromosome conformation capture carbon copy (5C) (Dixon et al., 2012; Han et al., 2018; Dostie et al., 2006), chromosome conformation capture-on-chip (4C) (Dixon et al., 2012; Han et al., 2018; Simonis et al., 2006), and other chromosome conformation capture (3C) (Han et al., 2018) techniques are all methods that are available for genomic interaction assessment purposes (Luo et al., 2020; Karličić et al., 2010). Although there are some differences among chromosomal capture technique methods, each method includes four main steps (crosslinking, fragmenting, ligating, and sequencing steps). Each approach requires an initial step for crosslinking chromatin. The genome is subsequently fragmented with endonucleases and then ligated. The ligation step is used to join the interacting genetic loci to each other, which is reverse crosslinked in preparation for

sequencing or quantitative polymerase chain reaction (PCR) to identify genomic interactions (Dostie et al., 2006; Simonis et al., 2006; Naumova et al., 2012; Lieberman-Aiden et al., 2009; Fullwood et al., 2009; Li et al., 2010; Zhao et al., 2019a; Fullwood and Ruan, 2009).

Quantitative trait loci (QTL)

QTL analysis is another method that can aid in prioritizing variants that are involved in a disease or trait. A QTL can be used to focus on a genomic region associated with a phenotype or trait, which is beneficial when attempting to implicate a gene for GWAS V2F follow-up (Powder, 2020). QTLs will typically incorporate marker and quantitative data to connect observed trait variation to genetic variations within a given population (Powder, 2020). Various types of QTL analysis exist and rely on various combinations of genotype, RNA-seq, chromatin accessibility, and/or methylation data (Powder, 2020). QTL-based data is based on observable, quantitative traits motivated by answering the question of how genetic variation impacts said trait. QTLs can further be used to assess the influence of genetic variation diseases and response to various treatments. Some of the more commonly used QTL analyses include expression quantitative trait loci (eQTLs) (Porcu et al., 2019), protein quantitative trait loci (pQTLs) (Xu et al., 2023), and chromatin accessibility quantitative trait loci (caQTLs) (Khetan et al., 2021; Kumasaka et al., 2019).

The more well-known type of QTL is the eQTL, which is used to determine how a specific genome region impacts gene expression variation. In order to conduct an eQTL study, both genotype and RNA expression data are required. Combining such data for eQTLs can aid in the GWAS V2F follow-up process to determine which genes are influenced by genetic differences in complex diseases or traits. Studies that have overlapped eQTL and GWAS data have been able to address the link between GWAS signals and potential target genes through 'co-localization'. One example was seen in the context of an AD study, where combined eQTL and mQTL data was used to prioritize SNPs and then to connect them to putative target effector genes (Zhao et al., 2019b). Findings from the study indicated an association between 653 SNPs and 25 genes, with 93 of the SNPs being significant for both eQTL and mQTL data (Zhao et al., 2019b). Furthermore, 10 out of the 25 genes found were previously identified in the literature as already being involved in the genetic etiology of AD (79). Another study used eQTL to assess immune related diseases and nominate 127 candidate disease genes following colocalization of eQTL and GWAS associated SNPs across 11 different diseases (Soskic et al., 2022). The eQTL data was generated from 119 human derived isolated and activated naive and memory CD4⁺ T-cell. Each cell was profiled at resting, 16 h, 40 h, and 5 days (Soskic et al., 2022). The eQTL and GWAS SNP results implicated genes in a broad range of immune related diseases, which included Crohn's disease, multiple sclerosis and SLE (80). More details on various types of QTL/GWAS colocalization and methods can be found elsewhere (Hormozdiari et al., 2014; Kang et al., 2023; Zuber et al., 2022; Cano-Gamez and Trynka, 2020; Suhre et al., 2021; Zhang et al., 2024; Abood and Farber, 2021; Fabo and Khavari, 2023).

Similar to eQTLs, pQTL and caQTL have major applications. Although these approaches serve slightly different purposes, they incorporate genotype data with protein levels for pQTLs, and

chromatin accessibility for caQTLs. A previous study focusing on serum protein successfully overlapped pQTLs with lead GWAS variants for multiple different phenotypes (Gudjonsson et al., 2022). In the study, two body mass index (BMI) associated loci were found to overlap with protein serum levels of Agouti signaling protein (ASIP) (Gudjonsson et al., 2022). The study also found waist-to-hip ratio GWAS signals within the *LRRC36* gene overlapping with levels of Agouti-related protein (Gudjonsson et al., 2022). One caQTL study had success when assessing GWAS signals for T2D (77). Chromatin accessibility sites were assessed in pancreatic islet cells, resulting in the nomination of causal variants at 13 GWAS loci (Khetan et al., 2021). These candidate causal loci were then functionally assessed *in vitro* via luciferase assay in MIN6 cells (Khetan et al., 2021). Out of the 13 loci, more than half were identified as having differential allelic regulatory roles (Khetan et al., 2021).

Additional QTL approaches that can be applied to GWAS follow-up are outlined in Table 2.

Transcriptome-wide association study (TWAS)

Similar to QTL, Transcriptome-Wide Association Study (TWAS) (and the newer causal TWAS that allows for confounding adjustments within the model (Zhao et al., 2024)) can aid in gene prioritization for GWAS follow-up. TWAS leverages genomic and transcriptomic data to discover how genetic differences might impact gene expression across different tissues. This approach can further explain how genetic variants found through GWAS not only affect gene expression, but also influence disease risk. A study conducted by Gusev et al. applied TWAS by using expression data across multiple tissues in conjunction with GWAS summary stats focused on traits such as height, body mass index (BMI), and lipids (Gusev et al., 2016). By leveraging this type of data, the study revealed 69 novel gene-trait associations (Gusev et al., 2016).

While many TWAS methods tend to be univariate (such as PMR-Egger (Yuan et al., 2020), PrediXcan (Gamazon et al., 2015), and FUSION(90)), more recent TWAS methods have tried to expand this type of model with the rationale of potential pleiotropic effects (Liu et al., 2021; Feng et al., 2021). One such study that developed a TWAS method referred to as moPMR-Egger which accommodates the analysis of multiple traits at a time (as opposed to one trait), lead to 13.5% increased gene associations findings when applied to United Kingdom biobank traits (Liu et al., 2021). While has different models available, it can still be a great genomic-based tool that offers a potential resolution to determining how GWAS identified variants might function within a biological setting by suggesting the genetic target(s) of GWAS variants. More in depth information of the various TWAS approaches along with valuable resources can be found elsewhere (Feng et al., 2021; Li and Ritchie, 2021; Zhu and Zhou, 2021; Mashhour et al., 2024; Xie et al., 2021).

While outside the scope of this review, it is worth mentioning the use of Phenome-wide association studies (PheWAS) for its ability to also connect gene-trait associations. A general overview on PheWAS and some resources can be found elsewhere (Liu and Crawford, 2022; Bastarache et al., 2022).

TABLE 2 List of various QTL resources available with additional information on the data required to perform each type of QTL and what question each type of QTL can answer.

Gene expression stage	QTL type	Abbreviations	Utility for incorporation into study design
Epigenetic Regulation	Methylation QTL	meQTL (Villicaña and Bell, 2021) mQTL (Lyu et al., 2021) methQTL (Scherer et al., 2021)	Genome-wide genotype in addition to epigenetic/histone marker data (methylation and/or acetylation) can be used to distinguish variant vs. epigenetic impact on trait of interest. Such data is also incorporated to identify associations between variant(s) and chromatin state (open or closed)
	Histone QTL	hQTL (Zheng et al., 2020)	
	Chromatin Accessibility QTL	caQTL (Zheng et al., 2020)	
	Transcription factor binding QTL	tfQTL (Watt et al., 2021) bQTL (Tehranchi et al., 2016)	Determining potential genomic regions involved in gene regulation in a cell specific manner
Transcriptional Regulation	Expression QTL	eQTL (Zhang and Zhao, 2023)	Genome wide genotype data and gene expression data (RNA-seq) to determine how variants impact gene expression of gene(s)
Post-Transcriptional Regulation	RNA Editing QTL	reQTL (Zheng et al., 2020) edQTL (Park et al., 2021)	When assessing the impact of genetic variation on protein levels. This can be due to underlying involvement in regulation involving pre-mRNA regulation by variation in capping, splicing, or polyadenylation
	microRNA QTL	miRQTL (Huan et al., 2015)	
	Splicing QTL	sQTL (Zheng et al., 2020)	
	Competing Endogenous RNA QTL	cerQTL (Zheng et al., 2020)	
	Alternative Polyadenylation QTL	apaQTL (Li et al., 2023b)	
Translational Regulation	Ribosome Occupancy QTL	riboQTL (Zheng et al., 2020) rQTL (Battle et al., 2015) roQTL (Cenik et al., 2015)	Ribosomal occupancy assessment can aid in determining potential underlying translational efficiency impact from putative variants (Ozadam et al., 2023)
	Protein Expression QTL	pQTL (Zheng et al., 2020)	GWAS/WES/WGS data combined with protein-based data to link variants to protein quantity (Xu et al., 2023)
Post-Translational Regulation	Metabolic QTL	metaQTL (Zheng et al., 2020)	Determine whether allelic differences explain differences in metabolite levels (Carreno-Quintero et al., 2012)

Pathway and network analysis

Successful incorporation of pathway and network-based analyses can inform V2F by implicating potential hubs influencing a phenotype of interest. One AD study identified 32 additional candidate genes by including a network analysis in their study (Schwartzentruber et al., 2021). They first generated an overall gene interaction list based on information retrieved from the STRING, IntAct, and BioGRID databases (Schwartzentruber et al., 2021). Through propagation methods utilizing genes found in the initial retrieval process, the study was able to highlight previously established genes associated with AD (Schwartzentruber et al., 2021). Interestingly, the study included high ranked genes for AD GWAS loci with the lowest *P*-values, indicating larger GWAS may be required to fully evaluate all putative loci (Schwartzentruber et al., 2021). Another recent study focused on kidney renal clear cell carcinoma (KIRC) utilized a Protein-Protein Interaction (PPI) network method and the authors elected to further evaluate the top four genes (referred to as “hub” genes) in the associated list (Ali et al., 2023). Through further assessment of their implicated hub genes, the study reported that two out of the four genes were upregulated while the other two were downregulated in KIRC patients (Ali et al., 2023).

Although there is debate regarding the utility and reproducibility of pathway and network based analyses (Tomczak et al., 2018), a wide-range of studies have included such results as a means to further understand potential underlying gene interactions on a broader scale (Yu et al., 2019; Jin et al., 2022). Pathway analysis can be accomplished through either a non-topology or a topology-driven method. A non-topology-driven method is traditionally known as an overrepresentation analysis (ORA) or functional enrichment analysis. An enrichment analysis considers a list of significantly differentially expressed genes (DEG) from a larger given data set. The list is then used to determine the percent of DEG present within a pathway. When more than 10% of DEGs are present in a given pathway, that specific network is considered “enriched” and worth further investigation. ORA tools include FuncAssociate (Berriz et al., 2009), GeneMerge (Castillo-Davis and Hartl, 2003), EASE (Hosack et al., 2003), g:Profiler (Kolberg et al., 2023), DAVID ((Huang et al., 2009a), (Huang et al., 2009b)), WebGestalt (Liao et al., 2019), AmiGO 2 (Carbon et al., 2009), GeneWeaver (Baker et al., 2012), BiNGO (Maere et al., 2005), GoMiner (Zeeberg et al., 2003), ontologizer (Bauer et al., 2008), etc. Functional class scoring (FCS) is another type of ORA similar to functional enrichment analysis but utilizes the entire gene set data

instead of only the DEG. Common FCS approaches include GSEA (Subramanian et al., 2005; Mootha et al., 2003), GSA (Mooney and Wilmot, 2015), GlobalTest (Hulsegge et al., 2009), PADOG (Tarca et al., 2012), SAM-GS (Dinu et al., 2007), FunCluster (Henegar et al., 2006), etc.

Gene Ontology (GO) analysis is another ORA-based enrichment method that allows users to assign functional annotations based on various categories. GO analysis can help assign genes or gene products based on molecular, biological, and cellular functions (Zhao et al., 2020). GO analysis is an example of a pathway analysis, which identifies biological pathways involved in a phenotype by using gene expression data and available pathway databases. This approach is often used to help elucidate how genes and variants impact specific cellular pathways. Pathway-based databases can help discern involved gene and protein networks, which helps further understand the underlying interactions within a given complex trait or disease. There are multiple online resources available to incorporate GO analysis. Most notable databases include Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa and Goto, 2000), PANTHER (Thomas et al., 2022), Reactome (Gillespie et al., 2022), Pathway Commons (Rodchenkov et al., 2020), Wiki-Pathways (Martens et al., 2021), and PathBank (Wishart et al., 2020).

Although non-topology ORA analyses are helpful and can yield biological insights, such analyses do not take into account interactions of genes, which often play a role in contributing to complex disease and traits. For this reason, topology-based approaches can be a better alternative to non-topology ORAs. Pathway topology (PT) considers the independent gene role, position, magnitude, and interactions. Fold change of a gene is propagated onto the gene directly downstream in a pathway. iPathwayGuide (Ahsan and Drăghici, 2017) provides a web-based system to conduct a PT approach to gene expression analysis tools to better rank pathways involved in a given phenotype. More information on pathway analysis and additional resources can be found elsewhere (Maleki et al., 2020).

Pathway analyses provide information regarding gene and gene products within a set pathway impacting a phenotype of interest, but not necessarily how proteins across different pathways potentially interact. Such information can yield additional insight into the underlying biological network leading to a complex disease or trait. For this purpose, PPI networks can help highlight how proteins within various pathways intersect and interact at a given point to serve an overall biological process. A valuable tool for a PPI network includes the STRING (Szklarczyk et al., 2021) database, which can also be used for enrichment and PPI network analysis.

Although it aids understanding of the underlying biological network, the main goal is for functional and mechanistic significance. Insight from additional approaches like phenotype annotation and comparative genomics can further aid in understanding how genetic variants lead to a specific functional outcome (i.e., disease, drug response, biological process, etc.). By incorporating tools like Ensembl (Martin et al., 2023), linking genetic changes to phenotype can help determine functional consequences. Furthermore, using evolutionally conservation data can also help understand functional significance through comparative genomics. Comparing genomic regions across species for GWAS signal interpretation may reveal a conserved gene regulatory site. Using such knowledge can further aid in understanding whether a specific locus is in a region potentially

involved in regulating an essential gene in a given biological pathway. Available tools for a comparative genomic approach include VISTA (Frazer et al., 2004; Dubchak et al., 2000), CoGE (Lyons and Freeling, 2008), PipMakers (Schwartz et al., 2000), etc.

Functional validation methods

Individual reporter-based method

Regulatory abilities of putative non-coding regulatory elements are traditionally assessed through individual reporter-based assays. Although reporter assays are not used to explicitly nominate genes being regulated by a putative enhancer region, they can represent an initial step in determining which nominated variants drive expression changes. The luciferase reporter assay system is an example of an individual reporter-based assay that is very useful when investigating the regulation on gene expression. Incorporating reporter assays has shown success in assessing variant regulatory impact of across different complex disease and traits (Zhang et al., 2010; Ustiugova et al., 2019; Rivas et al., 2011; Ramachandran et al., 2022). While investigating opioid addiction and the role of rs569356 in the gene *OPRD1* gene promoter, luciferase reporter plasmids were constructed with the major (A) and minor (G) alleles (Zhang et al., 2010). These constructs were transfected into HEK293 cells, and results indicated the G allele led to an increased expression in the reporter assay (Zhang et al., 2010). The differential allelic response was used to suggest a potential mechanism in regulation of *OPRD1* leading to opioid addiction (Zhang et al., 2010). Another study focused on GWAS associated loci in autoimmune diseases, incorporating luciferase reporter assay into their study design to determine allele specific functionality (Ustiugova et al., 2019). Both risk and protective alleles from six associated loci (rs12946510, rs2313430, rs4795397, rs12709365, rs13380815, rs8067378) were included in the reporter assay and assessed in six different cell lines (Nalm6, MP1, Jurkat, MT-2, U-937, and activated U-937) (Ustiugova et al., 2019). The results indicated cell specific differential allelic activity for at least four of the six loci, with the strongest effect seen for rs12946510 across three cell types (Nalm6, MP1, and activated U-937) (Ustiugova et al., 2019).

Although simplistic in design, luciferase assays are still useful in determining whether a specific genomic region harboring GWAS loci potentially impacts gene expression. Luciferase assays use plasmids containing a luciferase reporter gene located downstream of a regulatory element of interest and a minimal promoter (Fan and Wood, 2007). The final reporter construct is transfected into a specific cell type (animal, plant, or bacteria). Since the reporter gene is essentially fused to the regulatory element, detecting transcription changes in the reporter gene is directly correlated to the relative regulatory activity of the regulatory element (cis-acting) (Fan and Wood, 2007). The luciferase gene encodes for a specific enzyme that produces fluorescence, which is quantified by measuring the light intensity (Fan and Wood, 2007).

Although individual reporter assays help determine functional non-coding variants nominated from GWAS, they are very time-consuming. The time constraints associated with individual reporter assays require stringent prioritization of nominated GWAS variants due to their inability to incorporate multiple variants in a single experiment. Current genomic technologies have led to an expanded and comprehensive form of individual reporter-based assays,

referred to as high-throughput reporter assays. Such high-throughput reporter assays allows for the evaluation of thousands of putative non-coding regulatory variants simultaneously (Melnikov et al., 2012; Kheradpour et al., 2013; Kwasniewski et al., 2012; White et al., 2013; Inoue and Ahituv, 2015; Arnold et al., 2013). High-throughput reporter assays result in a more streamlined and comprehensive follow-up approach for GWAS-nominated variants.

Most high-throughput reporter assays, like Massively Parallel Reporter Assay (MPRA) and Self-Transcribing Active Regulatory Region Sequencing (STARR-seq) (both discussed below), use a similar plasmid-based concept as utilized in individual reporter assays when determining the relative regulatory ability of specific non-coding regions (Gallego Romero and Lea, 2023; Das et al., 2023). The main difference between individual reporter assays and high-throughput assays is how the expression is quantified. In individual reporter assays, the relative enhancer activity is qualitatively determined using light emission created by substrate-enzymatic reactions (Fan and Wood, 2007; Melnikov et al., 2012; Kheradpour et al., 2013; Tewhey et al., 2016). High-throughput reporter-based assays are far more quantitative by incorporating high throughput sequencing to count generated sequences caused by a specific regulatory region (Gallego Romero and Lea, 2023; Das et al., 2023).

High-throughput methods (MPRA)

MPRA is a highly reproducible and sensitive high-throughput reporter assay that simultaneously evaluates thousands of putative regulatory sequences while determining allele-specific activity (Melnikov et al., 2012; Kheradpour et al., 2013; Ulirsch et al., 2016; Lu et al., 2021). Inclusion of the MPRA design has been utilized in a wide range of contexts from answering evolutionary questions (Du et al., 2022) to determining previously unexplored allele specific activity of complex diseases/traits. In fact, multiple studies have shown great success in identifying functional GWAS loci that were previously unexplored through incorporation of MPRA (Matoba et al., 2020; Long et al., 2022; Mouri et al., 2022). One study, in particular, focused on autoimmune diseases where T cell involvement was known. Integrating variants associated with inflammatory bowel syndrome (IBD), multiple sclerosis (MS), type 1 diabetes (T1D), psoriasis and rheumatoid arthritis (RA) were included in the MPRA design and subsequently assessed for their functionality within T-cells (Mouri et al., 2022). With over 18,000 loci assessed, the study found 313 variants with differential expression when comparing the reference and alternative alleles (Mouri et al., 2022). Other studies have achieved success by leveraging data from MPRA and eQTLs combined, allowing further biological perturbations from previously overlooked loci (Choi et al., 2020). This study included over 800 loci associated with melanoma, and assessed their regulatory influence within a melanoma cell line (Choi et al., 2020). Overlapping the MPRA data with local eQTL data, the study could prioritize nine variants for which differential gene expression data appeared to corroborate potential variant endogenous gene regulation (Choi et al., 2020).

Incorporating MPRA to evaluate GWAS-associated variants include constructs with sentinel single nucleotide polymorphisms (SNPs) and variants in LD. Including variants in LD aids in

determining whether the SNP directly genotyped for the GWAS is influencing the biological system or the SNP in high LD with the SNP directly genotyped. Once a list of noncoding variants is determined, various manufacturers (such as Agilent Technologies, Dynegene, and Twist Bioscience (Agilent, 2024; Dynegene, 2024; Bioscience, 2024)) can create large-scale oligonucleotide (oligo) library pools. The entire MPRA oligo library is generated or synthesized by microarray technologies and is currently limited to 230 base pairs (bp) (Melnikov et al., 2014). Although traditional MPRA are highly reproducible and sensitive enough to determine allelic variation, the MPRA reporter plasmid design cannot guarantee biological relevance. The construct may not represent an actual regulatory element since MPRA oligos are limited in sequence length. Regulatory elements often span much larger regions than the length-restricted regions found in MPRA oligos, which may lead to type I and II errors (Blackwood and Kadonaga, 1998). Since regulatory regions can span large regions, it is essential to validate MPRA findings to include genomic regions that span larger bp regions centered on the putative variant that can be transformed into individual reporter plasmids (Blackwood and Kadonaga, 1998). The ability to further evaluate larger regions of any high-confidence regulatory element found in MPRA assays with individual reporter assays is one way to mitigate the negative consequences of the limited sequence lengths in MPRA oligo library pools. Another alternative approach to the limited length of current MPRA designs is the Tiling MPRA method (Ernst et al., 2016). Tiling MPRA allows one to extend the length of testable regulatory regions by using multiple 175bp constructs for one variant loci represented at the center of the construct and varying bp lengths to the left and right from the center of the constructs (Ernst et al., 2016).

In addition to size restrictions, MPRA plasmid pool designs do not consider the endogenous biological context. Although using the same minimal promoter and reporter gene for the entire MPRA library pool is beneficial for direct comparison across individual regulatory elements within the MPRA pool, it lacks biological context and relevance by excluding endogenous promoters and genes. MPRA experiments cannot measure the putative regulatory element's activity in its endogenous chromatin environment. Additionally, GWAS can nominate noncoding variants that may be involved in gene regulation but do not indicate putative tissue or cell types involved. Lack of cell-specific involvement in a disease or trait becomes another limitation for MPRA when incorporating library pools for a research design. MPRA libraries are created based on all top GWAS hits, which means false positives and negatives can occur when utilizing this technique within cell types that are not actively involved in the complex trait or disease of interest. Including chromatin accessibility data for specific cell types is one way to avoid such errors when determining which cell type best suits the MPRA design. Alternatively, the high-throughput reporter assay called STARR-seq can supplement the MPRA design to analyze active enhancer regions within a given temporal context (Arnold et al., 2013).

High-throughput methods (STARR-Seq)

Similar to MPRA, STARR-seq can simultaneously determine the regulatory activities of thousands of regulatory elements. Unlike MPRA, STARR-seq method can determine active regulatory regions

of the genome in a state/temporal dependent manner while directly assessing cell-specific regulatory activity (Arnold et al., 2013). Studies incorporating STAR-seq design have had success in determining regulatory elements for cellular response to various pharmacological drugs, such as dexamethasone and other glucocorticoids (GC) (Johnson et al., 2018; Penner-Goeke et al., 2023). Such studies placed cells under the influence of steroid-based drugs to induce a GC- response. Regulatory regions that were responsive to the pharmacological insult would reside in open chromatin, allowing for fragmentation and placement into the reporter plasmid for incorporation into the STARR-seq design to detect novel GC-responsive regulatory loci within the genome (Johnson et al., 2018; Penner-Goeke et al., 2023). Incorporating GC-responsive elements into the STARR-seq technique has previously allowed researchers to identify an overlap between functionally validated drug responsive elements that were enriched with identified GWAS variants associated with psychiatric traits (Penner-Goeke et al., 2023). Under the driving idea that a “stress” induced response would result in various psychiatric outcomes, the STARR-seq reporter plasmid was introduced into osteosarcoma and brain glioblastoma cell lines (Penner-Goeke et al., 2023). Furthermore, the identified GC-responsive elements were found to regulate transcripts enriched in genes shown to be differentially expressed in the cerebral cortex of psychiatric disorders, such as schizophrenia (SCZ), major depressive disorder (MDD) and autism spectrum disorder (ASD) (Penner-Goeke et al., 2023).

Importantly, STARR-seq can only include open chromatin accessibility sites in the reporter plasmid pool (Arnold et al., 2013). In some situations, the direct cell-specific assessment makes STARR-seq cost effective compared to MPRA as a plasmid pool is unnecessary for STARR-seq (Das et al., 2023). Instead, any cell type or tissue needed for a project can be used to extract cellular DNA and subsequently sheared to obtain smaller fragments of DNA (typically 300-500bp fragments) (Arnold et al., 2013). The fragmented DNA is transformed into the reporter plasmid. Open chromatin regions can be obtained by targeting regions by incorporating chromatin immunoprecipitation (CHIP) techniques to select for specific transcription factor binding sites or histone modifications associated with open chromatin sites and active enhancers (H3K27ac and H3K4me1) which can be advantageous when only wanting to assess GWAS associated SNP regions (Arnold et al., 2013; Heintzman et al., 2007). The STARR-seq reporter plasmid is then transfected into the cell type from which the regulatory elements were initially derived. Following transfection, the cells are lysed and submitted for high-throughput sequencing (Arnold et al., 2013).

Another difference seen between MPRA and STAR-seq methods is in the subsequent analysis. Variation in the analysis technique is due to the plasmid organization. MPRA plasmids place the regulatory regions upstream of a minimal promoter and reporter gene. The only way to determine regulatory activity in an MPRA experiment is to sequence the unique identifier downstream of the reporter gene. Unlike the MPRA plasmid organization, the STARR-seq method places the regulatory element downstream of the core promoter and reporter gene (Arnold et al., 2013). The basic concept driving the STARR-seq method assumes active enhancers can act independently of their position (Arnold et al., 2013). Placement of the regulatory elements in the STARR-seq plasmid will allow active enhancer regions to transcribe themselves, essentially being their

own unique identifier. Analysis for STARR-seq experiments differs from MPRA analysis primarily due to the need for alignment, which is due to the incorporation of a fragmentation step in creating the STAR-seq reporter plasmid pool (Arnold et al., 2013).

Genome editing based methods (CRISPR)

Incorporating functional annotation and validation methods can collectively implicate a link between GWAS-nominated variants and target gene expression; however, these techniques still lack the ability to physically confirm whether GWAS loci are causal within an endogenous setting. Recent techniques like CRISPR become useful in fully elucidating how variants impact phenotypes within an endogenous setting. This powerful genome editing technique has shown great utility in a wide range of diseases (such as Parkinson’s (Soldner et al., 2016), schizophrenia (Forrest et al., 2017) and nasopharyngeal carcinoma (Wang et al., 2023)). Furthermore, CRISPR has been used in pursuit of GWAS loci for complex traits such as bone mineral density (Pippin et al., 2021). Our study focused on bone mineral density connected GWAS associated variants to a target gene by incorporating promoter capture C (172). After obtaining a putative gene target (*EPDR1*), CRISPR was used to knockdown the gene in a cell model equivalent to mesenchymal stem-cell derived osteoblasts and we showed the gene impacted osteoblast differentiation (Pippin et al., 2021).

CRISPR aids in such investigations by allowing direct quantification of gene expression based on endogenous single nucleotide changes (Wang et al., 2016; Irion et al., 2014). Editing an individual nucleotide in a regulatory sequence allows direct assessment of variant function on putative target gene expression. This was accomplished in a study interesting in bone mineral density and hyperglycemic phenotypes (Sinnott-Armstrong et al., 2021). Chromosomal capture techniques were used to assess *ADCY5* as the potential target of the rs56371916 pleiotropic locus (Sinnott-Armstrong et al., 2021). CRISPR was implemented to edit the locus in adipose-derived mesenchymal stem cells (AMSCs) to be either homozygous for CC or TT (Sinnott-Armstrong et al., 2021). Following osteoblast or adipocyte induction, *ADCY5* expression was assessed (Sinnott-Armstrong et al., 2021). Findings indicated *ADCY5* expression was increased in osteoblast induction for the TT homozygous genotype, while conversely *ADCY5* expression was decreased in the adipocyte induction for the CC homozygous genotype (Sinnott-Armstrong et al., 2021). Overall, the change in *ADCY5* expression confirmed a pleiotropic effect of rs56371916 seen through lipolysis regulation in adipocytes and lipid-oxidation differentiation of osteoblasts (Sinnott-Armstrong et al., 2021).

Using CRISPR to edit single nucleotide changes requires the Cas9 protein, gRNA, and a repair template strand (Wang et al., 2016; Irion et al., 2014). The repair template strand is used to repair the genomic break caused by the Cas9 protein through the homology-directed repair (HDR) and nonhomologous end joining (NHEJ) mechanisms (Li et al., 2023a). Following nucleotide editing validation, the impact of downstream gene expression can be determined with traditional qPCR techniques. Unfortunately, off-target effects while using CRISPR and designing functional guide RNAs (gRNAs) are a valid concern. Exploiting the high-fidelity use of HDR for single nucleotide editing has the unfortunate drawback of having low efficiency with previous rate estimates of approximately <5% (Wang et al., 2016; Irion et al., 2014; Komor

et al., 2017). However, this rate has recently been drastically improved to >50% (Li et al., 2023a).

CRISPR screens can be implemented as a pooled (gRNAs in bulk cells) and arrayed (gRNAs into CRISPR screens) for a large-scale GWAS follow-up (Bock et al., 2022) or at a single-cell resolution (Faial, 2023). While pooled screens are great for discovery-based research, array-based screens are beneficial for follow-up-based studies. In the context of GWAS signals, array-based CRISPR screens can make implementing single-base pair changes in the genome possible, and measurable phenotypes can be determined based on the perturbation. With the continued use of CRISPR screen applications for GWAS functional follow-up, focus on single-cell CRISPR-based (scCRISPR) applications has increased. Some notable methods that incorporate the scCRISPR approach include CRISP-seq (Jaitin et al., 2016), CROP-seq (Datlinger et al., 2017), Mosaic-seq (Xie et al., 2017), Perturb-seq (Adamson et al., 2016; Dixit et al., 2016), and transcript-informed single-cell CRISPR sequencing (TISCC-seq) (Kim et al., 2024a). Additional details on the progress of CRISPR-based screen approaches has been recently explored by Kim et al. (2024b), Cooper et al. (2024).

Artificial intelligence (AI) methods

AI is a general term used to describe technology that utilizes computer algorithms like machine learning or deep learning (Sarker, 2022). The benefits of incorporating AI programs can be seen in their ability to handle large amounts of data, which can significantly decrease the overall time of determining putative gene targets of GWAS variants (or variants in LD with GWAS variants). AI tools have been trained on different datasets for the overall purpose of pattern recognition. Having the ability to assess patterns quickly, AI has become an efficient way to predict which variants might disrupt regulatory motifs and gene function, leading to a more streamline method to prioritize the most relevant variant and putative target gene for functional validation.

Researchers have already benefited from AI integration in tools that can help suggest the potential functional impact of GWAS variants such as the online tools GeneMANIA (Warde-Farley et al., 2010), STRING (Szklarczyk et al., 2021; Martin et al., 2023), PhenoScanner V2 (Kamat et al., 2019), and Variant Effect Predictor (VEP) (Martin et al., 2023; McLaren et al., 2016). Tools like these, capitalize on pattern recognition from intersecting AI and multi-omics datasets (such as transcriptomics, functional annotations, etc.). Other tools, like DeepSEA (Zhou and Troyanskaya, 2015) and ExCAPE (Sturm et al., 2020), use a convolutional neural network (Gu et al., 2018) to determine the potential impact of regulatory elements and gene expression and how this could translate to a pharmaceutical approach, respectively (Zhou and Troyanskaya, 2015; Sturm et al., 2020).

Genetic variants effect predicting tools have further been inhibited by the inclusion of deep learning tool. Some notable examples include ESM (Lin et al., 2023) (iterations include ESM-1, ESM-2, and recently ESM-3), AlphaMissense (Cheng et al., 2023; Tordai et al., 2024), AlphaFold2 (Yang et al., 2023b), and DeepVariant (Chen et al., 2023). Various iterations of ESM focus on model protein structures. ESM assesses evolutionary sequence data and genetic mutations into its model to suggest how variants affect protein folding, stability, and/or interactions (Lin et al., 2023; Callaway, 2024). ESM-3 has extended this protein-based application to predicting new proteins that could be created to enhance current

GFP and CRISPR proteins to enhance current techniques used in functional assays (Callaway, 2024). Similarly, AlphaMissense AI tool predicts protein activity caused by variants that impact amino acid sequences by using deep learning and structural bioinformatics (Tordai et al., 2024). Apart from protein prediction effect, other advancements in AI have provided newer tools like DeepVariant that have improved accuracy of mapping variants to genes and/or regulatory elements they might affect (Chen et al., 2023).

Once a target gene or regulatory element is prioritized with the available AI algorithms, other forms of AI tools can further aid in developing the assays used for functional assays. Already mentioned was the latest ESM-3 tool that is actively attempting to create new protein structures based on known sequences that could provide higher efficient CRISPR proteins for genome editing (Callaway, 2024). ESM-3 is currently too new to fully know the degree of efficiency this may provide to current CRISPR techniques used for functional follow-up, but a important to mention in the context of how AI tools may further advance functional assessments in the future (Callaway, 2024). AI-powered design tools like CRISPR-DO provides predictive feedback on designing guide RNAs used for genome editing, which is useful for designing the most efficient gRNA for successful application of CRISPR technique (Ma et al., 2016). Additional AI powered tools and method have been explored elsewhere (Sigala et al., 2023; Nicholls et al., 2020).

Concluding remarks

Functional work of GWAS signals continue to be an arduous endeavor. Most of the complexity of GWAS follow-up stems from the large percent of predicted variants reside within the noncoding genome. Such nominated variants likely involved in gene regulation, which requires the need to determine target gene(s) and how this may vary across different tissues and under varying stimuli. As technology and computation methods continue to advance, such challenges will steadily decrease (or at least minimize the time to find causal variants and their target gene(s)). Continued advancements will further help overcome the additional challenge of answering further questions regarding how GWAS variants might impact gene function within different cells/tissues, cellular state (such as development state), and environmental stimuli (drug exposure, altitude, temperature, etc.).

The future of GWAS signal follow-up is beginning to be positively impacted by the continued interest in artificial intelligence (AI) and machine learning integration. As GWAS data is often too large to follow-up efficiently, technological advancements involving AI promise to quickly assess the most probable GWAS signals responsible for complex diseases and traits. Moreover, an AI-based framework can be incorporated to identify underlying networks of regulation that are currently time-consuming and difficult to explain fully in complex phenotypes. In this manner, AI has the potential to overcome a large portion of the time and financial burden often associated with follow-up studies of GWAS results and further facilitate quicker personalized therapy and medicine.

As genomics and computational approaches continue to advance, most notably with the recent advances in V2F based algorithms such as activity-by-contact (ABC) (Nasser et al., 2021), polygenic priority score (PoPS) (Weeks et al., 2023) and the Effector Index (Liang et al., 2023), GWAS V2F follow-up will

become increasingly capable of uncovering the underlying gene-gene and gene-environment interactions that collectively contribute to disease and trait etiologies. Currently, combining multiple techniques for variant prioritization and functional validation methods is an established process for determining causal variants and pathways contributing to complex phenotypes of interest. Most of the methods and techniques lead to large multi-omics and pathway-based datasets that require computational models to dissect the genetic basis of a complex trait comprehensively. Ongoing improvements with high-throughput techniques and computational methods will allow for easier identification of causal genes and pathways than previous capabilities. With the advent and continued advancements of computational algorithms and AI, the inclusion of multi-omics and “big data” will undoubtedly experience a further reduction in the current time restraints required for completing GWAS V2F follow-up.

Author contributions

WB: Writing—original draft, Writing—review and editing. SG: Writing—original draft, Writing—review and editing.

References

- Abood, A., and Farber, C. R. (2021). Using “-omics” data to inform genome-wide association studies (GWAS) in the osteoporosis field. *Curr. Osteoporos. Rep.* 19 (4), 369–380. doi:10.1007/s11914-021-00684-w
- Adamson, B., Norman, T. M., Jost, M., Cho, M. Y., Nuñez, J. K., Chen, Y., et al. (2016). A multiplexed single-cell CRISPR screening platform enables systematic dissection of the unfolded protein response. *Cell.* 167 (7), 1867–1882. doi:10.1016/j.cell.2016.11.048
- Agilent (2024). Chemical analysis, life sciences, and diagnostics | agilent. Available at: <https://www.agilent.com/> (Accessed October 23, 2023).
- Ahsan, S., and Drághici, S. (2017). Identifying significantly impacted pathways and putative mechanisms with iPathwayGuide. *Curr. Protoc. Bioinforma.* 57 (1), 7.15.1–7. doi:10.1002/cpbi.24
- Akiyama, M., Okada, Y., Kanai, M., Takahashi, A., Momozawa, Y., Ikeda, M., et al. (2017). Genome-wide association study identifies 112 new loci for body mass index in the Japanese population. *Nat. Genet.* 49 (10), 1458–1467. doi:10.1038/ng.3951
- Ali, L., Raza, A. A., Zaheer, A. B., Alhomrani, M., Alamri, A. S., Alghamdi, S. A., et al. (2023). *In vitro* analysis of PI3K pathway activation genes for exploring novel biomarkers and therapeutic targets in clear cell renal carcinoma. *Am. J. Transl. Res.* 15 (7), 4851–4872.
- Anttila, V., Winsvold, B. S., Gormley, P., Kurth, T., Bettella, F., McMahon, G., et al. (2013). Genome-wide meta-analysis identifies new susceptibility loci for migraine. *Nat. Genet.* 45 (8), 912–917. doi:10.1038/ng.2676
- Arnold, C. D., Gerlach, D., Stelzer, C., Boryń, L. M., Rath, M., and Stark, A. (2013). Genome-wide quantitative enhancer activity maps identified by STARR-seq. *Science* 339 (6123), 1074–1077. doi:10.1126/science.1232542
- Baker, E. J., Jay, J. J., Bubier, J. A., Langston, M. A., and Chesler, E. J. (2012). GeneWeaver: a web-based system for integrative functional genomics. *Nucleic Acids Res.* 40 (Database issue), D1067–D1076. doi:10.1093/nar/gkr968
- Banerjee, A., Chitnis, U. B., Jadhav, S. L., Bhawalkar, J. S., and Chaudhury, S. (2009). Hypothesis testing, type I and type II errors. *Ind. Psychiatry J.* 18 (2), 127–131. doi:10.4103/0972-6748.62274
- Barski, A., Cuddapah, S., Cui, K., Roh, T. Y., Schones, D. E., Wang, Z., et al. (2007). High-resolution profiling of histone methylations in the human genome. *Cell.* 129 (4), 823–837. doi:10.1016/j.cell.2007.05.009
- Bastarache, L., Denny, J. C., and Roden, D. M. (2022). Phenome-wide association studies. *JAMA* 327 (1), 75–76. doi:10.1001/jama.2021.20356
- Battle, A., Khan, Z., Wang, S. H., Mitrano, A., Ford, M. J., Pritchard, J. K., et al. (2015). Genomic variation. Impact of regulatory variation from RNA to protein. *Science* 347 (6222), 664–667. doi:10.1126/science.1260793
- Bauer, S., Grossmann, S., Vingron, M., and Robinson, P. N. (2008). Ontologizer 2.0—a multifunctional tool for GO term enrichment analysis and data exploration. *Bioinforma. Oxf Engl.* 24 (14), 1650–1651. doi:10.1093/bioinformatics/btn250
- Berriz, G. F., Beaver, J. E., Cenik, C., Tasan, M., and Roth, F. P. (2009). Next generation software for functional trend analysis. *Bioinformatics* 25 (22), 3043–3044. doi:10.1093/bioinformatics/btp498
- Bioscience (2024). Oligo pools for high throughput screens - twist bioscience. Available at: <https://www.twistbioscience.com/products/oligopools?tab%27overview=&adgroup=114820678743&creative=491174669992&device=c&matchtype=e&location=9007378> (Accessed October 23, 2023).
- Blackwood, E. M., and Kadonaga, J. T. (1998). Going the distance: a current view of enhancer action. *Science* 281 (5373), 60–63. doi:10.1126/science.281.5373.60
- Bock, C., Datlinger, P., Chardon, F., Coelho, M. A., Dong, M. B., Lawson, K. A., et al. (2022). High-content CRISPR screening. *Nat. Rev. Methods Primer* 2 (1), 8–23. doi:10.1038/s43586-021-00093-4
- Bradfield, J. P., Taal, H. R., Timpson, N. J., Scherag, A., Lecoeur, C., Warrington, N. M., et al. (2012). A genome-wide association meta-analysis identifies new childhood obesity loci. *Nat. Genet.* 44, 526–531. doi:10.1038/ng.2247
- Bradfield, J. P., Voegelzang, S., Felix, J. F., Chesi, A., Helgeland, Ø., Horikoshi, M., et al. (2019). A trans-ancestral meta-analysis of genome-wide association studies reveals loci associated with childhood obesity. *Hum. Mol. Genet.* 28, 3327–3338. doi:10.1093/hmg/ddz161
- Buenrostro, J. D., Wu, B., Chang, H. Y., and Greenleaf, W. J. (2015). ATAC-seq: a method for assaying chromatin accessibility genome-wide. *Curr. Protoc. Mol. Biol.* 109, 21.29.1–21. doi:10.1002/0471142727.mb2129s109
- Bysani, M., Agren, R., Davegårdh, C., Volkov, P., Rönn, T., Unneberg, P., et al. (2019). ATAC-seq reveals alterations in open chromatin in pancreatic islets from subjects with type 2 diabetes. *Sci. Rep.* 9 (1), 7785. doi:10.1038/s41598-019-44076-8
- Cai, Y., Zhang, Y., Loh, Y. P., Tng, J. Q., Lim, M. C., Cao, Z., et al. (2021). H3K27me3-rich genomic regions can function as silencers to repress gene expression via chromatin interactions. *Nat. Commun.* 12 (1), 719. doi:10.1038/s41467-021-20940-y
- Callaway, E. (2024). ‘It will change everything’: DeepMind’s AI makes gigantic leap in solving protein structures. *Nature* 588, 203–204. doi:10.1038/d41586-020-03348-4
- Cano-Gamez, E., and Trynka, G. (2020). From GWAS to function: using functional genomics to identify the mechanisms underlying complex diseases. *Front. Genet.* 11, 424. doi:10.3389/fgene.2020.00424
- Carbon, S., Ireland, A., Mungall, C. J., Shu, S., Marshall, B., Lewis, S., et al. (2009). AmiGO: online access to ontology and annotation data. *Bioinformatics* 25 (2), 288–289. doi:10.1093/bioinformatics/btn615
- Carrero-Quintero, N., Acharjee, A., Maliepaard, C., Bachem, C. W. B., Mumm, R., Bouwmeester, H., et al. (2012). Untargeted metabolic quantitative trait loci analyses reveal a relationship between primary metabolism and potato tuber quality. *Plant Physiol.* 158 (3), 1306–1318. doi:10.1104/pp.111.188441
- Castillo-Davis, C. I., and Hartl, D. L. (2003). GeneMerge—post-genomic analysis, data mining, and hypothesis testing. *Bioinforma. Oxf Engl.* 19 (7), 891–892. doi:10.1093/bioinformatics/btg114

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher’s note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

- Cenik, C., Cenik, E. S., Byeon, G. W., Grubert, F., Candille, S. I., Spacek, D., et al. (2015). Integrative analysis of RNA, translation, and protein levels reveals distinct regulatory variation across humans. *Genome Res.* 25 (11), 1610–1621. doi:10.1101/gr.193342.115
- Chen, N. C., Kolesnikov, A., Goel, S., Yun, T., Chang, P. C., and Carroll, A. (2023). Improving variant calling using population data and deep learning. *BMC Bioinforma.* 24 (1), 197. doi:10.1186/s12859-023-05294-0
- Cheng, J., Novati, G., Pan, J., Bycroft, C., Žemgulytė, A., Applebaum, T., et al. (2023). Accurate proteome-wide missense variant effect prediction with AlphaMissense. *Science* 381 (6664), eadg7492. doi:10.1126/science.adg7492
- Choi, J., Zhang, T., Vu, A., Ablain, J., Makowski, M. M., Colli, L. M., et al. (2020). Massively parallel reporter assays of melanoma risk variants identify MX2 as a gene promoting melanoma. *Nat. Commun.* 11, 2718. doi:10.1038/s41467-020-16590-1
- Clausnitzer, M., Dankel, S. N., Kim, K. H., Quon, G., Meuleman, W., Haugen, C., et al. (2015). FTO obesity variant circuitry and adipocyte browning in humans. *N. Engl. J. Med.* 373 (10), 895–907. doi:10.1056/NEJMoa1502214
- Cooper, S., Obolenski, S., Waters, A. J., Bassett, A. R., and Coelho, M. A. (2024). Analyzing the functional effects of DNA variants with gene editing. *Cell. Rep. Methods* 4 (5), 100776. doi:10.1016/j.crmeth.2024.100776
- Creyghton, M. P., Cheng, A. W., Welstead, G. G., Kooistra, T., Carey, B. W., Steine, E. J., et al. (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc. Natl. Acad. Sci. U. S. A.* 107 (50), 21931–21936. doi:10.1073/pnas.1016071107
- Das, M., Hossain, A., Banerjee, D., Praul, C. A., and Girirajan, S. (2023). Challenges and considerations for reproducibility of STARR-seq assays. *Genome Res.* 33 (4), 479–495. doi:10.1101/gr.27204.122
- Datlinger, P., Rendeiro, A. F., Schmid, C., Krausgruber, T., Traxler, P., Klughammer, J., et al. (2017). Pooled CRISPR screening with single-cell transcriptome readout. *Nat. Methods* 14 (3), 297–301. doi:10.1038/nmeth.4177
- de los Campos, G., Grueneberg, A., Funkhouser, S., Pérez-Rodríguez, P., and Samadpour, A. (2023). Fine mapping and accurate prediction of complex traits using Bayesian Variable Selection models applied to biobank-size data. *Eur. J. Hum. Genet.* 31 (3), 313–320. doi:10.1038/s41431-022-01135-5
- Deng, Y., Bartosovic, M., Ma, S., Zhang, D., Kukanja, P., Xiao, Y., et al. (2022). Spatial profiling of chromatin accessibility in mouse and human tissues. *Nature* 609 (7926), 375–383. doi:10.1038/s41586-022-05094-1
- Dinu, I., Potter, J. D., Mueller, T., Liu, Q., Adewale, A. J., Jhangri, G. S., et al. (2007). Improving gene set analysis of microarray data by SAM-GS. *BMC Bioinforma.* 8 (1), 242. doi:10.1186/1471-2105-8-242
- Dixit, A., Parnas, O., Li, B., Chen, J., Fulco, C. P., Jerby-Arnon, L., et al. (2016). Perturb-seq: dissecting molecular circuits with scalable single-cell RNA profiling of pooled genetic screens. *Cell.* 167 (7), 1853–1866. doi:10.1016/j.cell.2016.11.038
- Dixon, J. R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., et al. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* 485 (7398), 376–380. doi:10.1038/nature11082
- Dostie, J., Richmond, T. A., Arnaout, R. A., Selzer, R. R., Lee, W. L., Honan, T. A., et al. (2006). Chromosome Conformation Capture Carbon Copy (5C): a massively parallel solution for mapping interactions between genomic elements. *Genome Res.* 16 (10), 1299–1309. doi:10.1101/gr.5571506
- Du, A. Y., Zhuo, X., Sundaram, V., Jensen, N. O., Chaudhari, H. G., Saccone, N. L., et al. (2022). Functional characterization of enhancer activity during a long terminal repeat's evolution. *Genome Res.* 32 (10), 1840–1851. doi:10.1101/gr.276863.122
- Dubchak, I., Brudno, M., Loots, G. G., Pachter, L., Mayor, C., Rubin, E. M., et al. (2000). Active conservation of noncoding sequences revealed by three-way species comparisons. *Genome Res.* 10 (9), 1304–1306. doi:10.1101/gr.142200
- Dynegene (2024). Dynegene technologies official website. Available at: <https://www.dynegene.com/en/> (Accessed October 23, 2023).
- Easton, D. F., Pooley, K. A., Dunning, A. M., Pharoah, P. D. P., Thompson, D., Ballinger, D. G., et al. (2007). Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature* 447 (7148), 1087–1093. doi:10.1038/nature05887
- Elks, C. E., Loos, R. J. F., Sharp, S. J., Langenberg, C., Ring, S. M., Timpson, N. J., et al. (2010). Genetic markers of adult obesity risk are associated with greater early infancy weight gain and growth. *PLoS Med.* 7 (5), e1000284. doi:10.1371/journal.pmed.1000284
- Elks, C. E., Loos, R. J. F., Hardy, R., Wills, A. K., Wong, A., Wareham, N. J., et al. (2012). Adult obesity susceptibility variants are associated with greater childhood weight gain and a faster tempo of growth: the 1946 British Birth Cohort study. *Am. J. Clin. Nutr.* 95 (5), 1150–1156. doi:10.3945/ajcn.111.027870
- Ernst, J., Melnikov, A., Zhang, X., Wang, L., Rogov, P., Mikkelsen, T. S., et al. (2016). Genome-scale high-resolution mapping of activating and repressive nucleotides in regulatory regions. *Nat. Biotechnol.* 34 (11), 1180–1190. doi:10.1038/nbt.3678
- Fabo, T., and Khavari, P. (2023). Functional characterization of human genomic variation linked to polygenic diseases. *Trends Genet. TIG* 39 (6), 462–490. doi:10.1016/j.tig.2023.02.014
- Faial, T. (2023). Single-cell CRISPR screen for GWAS loci. *Nat. Genet.* 55 (6), 904. doi:10.1038/s41588-023-01432-9
- Fan, F., and Wood, K. V. (2007). Bioluminescent assays for high-throughput screening. *Assay. Drug Dev. Technol.* 5 (1), 127–136. doi:10.1089/adt.2006.053
- Felix, J. F., Bradfield, J. P., Monnereau, C., van der Valk, R. J. P., Stergiakouli, E., Chesi, A., et al. (2016). Genome-wide association analysis identifies three new susceptibility loci for childhood body mass index. *Hum. Mol. Genet.* 25 (2), 389–403. doi:10.1093/hmg/ddv472
- Feng, H., Mancuso, N., Pasaniuc, B., and Kraft, P. (2021). Multitrait transcriptome-wide association study (TWAS) tests. *Genet. Epidemiol.* 45 (6), 563–576. doi:10.1002/gepi.22391
- Forrest, M. P., Zhang, H., Moy, W., McGowan, H., Leites, C., Dionisio, L. E., et al. (2017). Open chromatin profiling in hiPSC-derived neurons prioritizes functional noncoding psychiatric risk variants and highlights neurodevelopmental loci. *Cell Stem Cell.* 21 (3), 305–318. doi:10.1016/j.stem.2017.07.008
- Frayling, T. M., Timpson, N. J., Weedon, M. N., Zeggini, E., Freathy, R. M., Lindgren, C. M., et al. (2007). A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. *Science* 316 (5826), 889–894. doi:10.1126/science.1141634
- Frazer, K. A., Pachter, L., Paliakov, A., Rubin, E. M., and Dubchak, I. (2004). VISTA: computational tools for comparative genomics. *Nucleic Acids Res.* 32 (Web Server issue), W273–W279. doi:10.1093/nar/gkh458
- Fullwood, M. J., and Ruan, Y. (2009). ChIP-based methods for the identification of long-range chromatin interactions. *J. Cell. Biochem.* 107 (1), 30–39. doi:10.1002/jcb.22116
- Fullwood, M. J., Liu, M. H., Pan, Y. F., Liu, J., Xu, H., Mohamed, Y. B., et al. (2009). An oestrogen-receptor-alpha-bound human chromatin interactome. *Nature* 462 (7269), 58–64. doi:10.1038/nature08497
- Galaxy Community (2022). The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2022 update. *Nucleic Acids Res.* 50 (W1), W345–W351. doi:10.1093/nar/gkac247
- Gallego Romero, I., and Lea, A. J. (2023). Leveraging massively parallel reporter assays for evolutionary questions. *Genome Biol.* 24 (1), 26. doi:10.1186/s13059-023-02856-6
- Gamazon, E. R., Wheeler, H. E., Shah, K. P., Mozaffari, S. V., Aquino-Michaels, K., Carroll, R. J., et al. (2015). A gene-based association method for mapping traits using reference transcriptome data. *Nat. Genet.* 47 (9), 1091–1098. doi:10.1038/ng.3367
- Genome-wide association study identifies new (2009). Genome-wide association study identifies new multiple sclerosis susceptibility loci on chromosomes 12 and 20. *Nat. Genet.* 41, 824–828. doi:10.1038/ng.396
- Ghousaini, M., Mountjoy, E., Carmona, M., Peat, G., Schmidt, E. M., Hercules, A., et al. (2021). Open Targets Genetics: systematic identification of trait-associated genes using large-scale genetics and functional genomics. *Nucleic Acids Res.* 49 (D1), D1311–D1320. doi:10.1093/nar/gkaa840
- Gibson, G. (2012). Rare and common variants: twenty arguments. *Nat. Rev. Genet.* 13 (2), 135–145. doi:10.1038/nrg3118
- Gillespie, M., Jassal, B., Stephan, R., Milacic, M., Rothfels, K., Senff-Ribeiro, A., et al. (2022). The reactome pathway knowledgebase 2022. *Nucleic Acids Res.* 50 (D1), D687–D692. doi:10.1093/nar/gkab1028
- Giresi, P. G., Kim, J., McDaniell, R. M., Iyer, V. R., and Lieb, J. D. (2007). FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) isolates active regulatory elements from human chromatin. *Genome Res.* 17 (6), 877–885. doi:10.1101/gr.5533506
- Grandi, F. C., Modi, H., Kampman, L., and Corces, M. R. (2022). Chromatin accessibility profiling by ATAC-seq. *Nat. Protoc.* 17 (6), 1518–1552. doi:10.1038/s41596-022-00692-9
- Grant, S. F. A., Li, M., Bradfield, J. P., Kim, C. E., Annaiah, K., Santa, E., et al. (2008). Association analysis of the FTO gene with obesity in children of Caucasian and African ancestry reveals a common tagging SNP. *PLoS One* 3 (3), e1746. doi:10.1371/journal.pone.0001746
- Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B., et al. (2018). Recent advances in convolutional neural networks. *Pattern Recognit.* 77, 354–377. doi:10.1016/j.patcog.2017.10.013
- Gudjonsson, A., Gudmundsdottir, V., Axelsson, G. T., Gudmundsson, E. F., Jonsson, B. G., Launer, L. J., et al. (2022). A genome-wide association study of serum proteins reveals shared loci with common diseases. *Nat. Commun.* 13 (1), 480. doi:10.1038/s41467-021-27850-z
- Gusev, A., Ko, A., Shi, H., Bhatia, G., Chung, W., Penninx, B. W. J. H., et al. (2016). Integrative approaches for large-scale transcriptome-wide association studies. *Nat. Genet.* 48 (3), 245–252. doi:10.1038/ng.3506
- Han, J., Zhang, Z., and Wang, K. (2018). 3C and 3C-based techniques: the powerful tools for spatial genome organization deciphering. *Mol. Cytogenet.* 11, 21. doi:10.1186/s13039-018-0368-2
- Hassanein, M. T., Lyon, H. N., Nguyen, T. T., Akylbekova, E. L., Waters, K., Lettre, G., et al. (2010). Fine mapping of the association with obesity at the FTO locus in African-derived populations. *Hum. Mol. Genet.* 19 (14), 2907–2916. doi:10.1093/hmg/ddq178
- Heintzman, N. D., Stuart, R. K., Hon, G., Fu, Y., Ching, C. W., Hawkins, R. D., et al. (2007). Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat. Genet.* 39 (3), 311–318. doi:10.1038/ng1966

- Henegar, C., Canello, R., Rome, S., Vidal, H., Clément, K., and Zucker, J. D. (2006). Clustering biological annotations and gene expression data to identify putatively co-regulated biological processes. *J. Bioinform. Comput. Biol.* 4 (4), 833–852. doi:10.1142/s0219720006002181
- Hormozdiari, F., Kostem, E., Kang, E. Y., Pasiñic, B., and Eskin, E. (2014). Identifying causal variants at loci with multiple signals of association. *Genetics* 198 (2), 497–508. doi:10.1534/genetics.114.167908
- Hosack, D. A., Dennis, G., Sherman, B. T., Lane, H. C., and Lempicki, R. A. (2003). Identifying biological themes within lists of genes with EASE. *Genome Biol.* 4 (10), R70. doi:10.1186/gb-2003-4-10-r70
- Huan, T., Rong, J., Liu, C., Zhang, X., Tanriverdi, K., Joehanes, R., et al. (2015). Genome-wide identification of microRNA expression quantitative trait loci. *Nat. Commun.* 6 (1), 6601. doi:10.1038/ncomms7601
- Huang, D. W., Sherman, B. T., and Lempicki, R. A. (2009a). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* 4 (1), 44–57. doi:10.1038/nprot.2008.211
- Huang, D. W., Sherman, B. T., and Lempicki, R. A. (2009b). Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* 37 (1), 1–13. doi:10.1093/nar/gkn923
- Hulsegge, I., Kommadath, A., and Smits, M. A. (2009). Globaltest and GOEAST: two different approaches for Gene Ontology analysis. *BMC Proc.* 3 (Suppl. 4), S10. doi:10.1186/1753-6561-3-S4-S10
- Inoue, F., and Ahituv, N. (2015). Decoding enhancers using massively parallel reporter assays. *Genomics* 106 (3), 159–164. doi:10.1016/j.ygeno.2015.06.005
- Irion, U., Krauss, J., and Nüsslein-Volhard, C. (2014). Precise and efficient genome editing in zebrafish using the CRISPR/Cas9 system. *Dev. Camb. Engl.* 141 (24), 4827–4830. doi:10.1242/dev.115584
- Jaitin, D. A., Weiner, A., Yofe, I., Lara-Astiaso, D., Keren-Shaul, H., David, E., et al. (2016). Dissecting immune circuits by linking CRISPR-pooled screens with single-cell RNA-seq. *Cell* 167 (7), 1883–1896. doi:10.1016/j.cell.2016.11.039
- Jin, X. X., Xie, X. L., Niu, F., Yin, K. G., Ji, C. G., Cui, J. F., et al. (2022). A single-center follow-up study of low-grade gastric intraepithelial neoplasia and the screening of key genes of precancerous lesions. *Front. Oncol.* 12, 899055. doi:10.3389/fonc.2022.899055
- Johnson, G. D., Barrera, A., McDowell, I. C., D'Ippolito, A. M., Majoros, W. H., Vockley, C. M., et al. (2018). Human genome-wide measurement of drug-responsive regulatory activity. *Nat. Commun.* 9, 5317. doi:10.1038/s41467-018-07607-x
- Kamat, M. A., Blackshaw, J. A., Young, R., Surendran, P., Burgess, S., Danesh, J., et al. (2019). PhenoScanner V2: an expanded tool for searching human genotype–phenotype associations. *Bioinformatics* 35 (22), 4851–4853. doi:10.1093/bioinformatics/btz469
- Kanehisa, M., and Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 28 (1), 27–30. doi:10.1093/nar/28.1.27
- Kang, J. B., Raveane, A., Nathan, A., Soranzo, N., and Raychaudhuri, S. (2023). Methods and insights from single-cell expression quantitative trait loci. *Annu. Rev. Genomics Hum. Genet.* 24, 277–303. doi:10.1146/annurev-genom-101422-100437
- Karlič, R., Chung, H. R., Lasserre, J., Vlahovick, K., and Vingron, M. (2010). Histone modification levels are predictive for gene expression. *Proc. Natl. Acad. Sci. U. S. A.* 107 (7), 2926–2931. doi:10.1073/pnas.0909344107
- Kent, W. J., Sugnet, C. W., Furey, T. S., Roskin, K. M., Pringle, T. H., Zahler, A. M., et al. (2002). The human genome browser at UCSC. *Genome Res.* 12 (6), 996–1006. doi:10.1101/gr.229102
- Kheradpour, P., Ernst, J., Melnikov, A., Rogov, P., Wang, L., Zhang, X., et al. (2013). Systematic dissection of regulatory motifs in 2000 predicted human enhancers using a massively parallel reporter assay. *Genome Res.* 23 (5), 800–811. doi:10.1101/gr.144899.112
- Khetan, S., Kales, S., Kursawe, R., Jillette, A., Ulirsch, J. C., Reilly, S. K., et al. (2021). Functional characterization of T2D-associated SNP effects on baseline and ER stress-responsive β cell transcriptional activation. *Nat. Commun.* 12 (1), 5242. doi:10.1038/s41467-021-25514-6
- Kichaev, G., Yang, W. Y., Lindstrom, S., Hormozdiari, F., Eskin, E., Price, A. L., et al. (2014). Integrating functional data to prioritize causal variants in statistical fine-mapping studies. *PLoS Genet.* 10 (10), e1004722. doi:10.1371/journal.pgen.1004722
- Kim, H. S., Grimes, S. M., Chen, T., Sathe, A., Lau, B. T., Hwang, G. H., et al. (2024a). Direct measurement of engineered cancer mutations and their transcriptional phenotypes in single cells. *Nat. Biotechnol.* 42 (8), 1254–1262. doi:10.1038/s41587-023-01949-8
- Kim, H. S., Kwon, J., and Kim, Y. (2024b). Recent advances in CRISPR-based functional genomics for the study of disease-associated genetic variants. *Exp. Mol. Med.* 56 (4), 861–869. doi:10.1038/s12276-024-01212-3
- Kocarnik, J. M., Richard, M., Graff, M., Haessler, J., Bien, S., Carlson, C., et al. (2018). Discovery, fine-mapping, and conditional analyses of genetic variants associated with C-reactive protein in multiethnic populations using the MetaboChip in the Population Architecture using Genomics and Epidemiology (PAGE) study. *Hum. Mol. Genet.* 27 (16), 2940–2953. doi:10.1093/hmg/ddy211
- Kolberg, L., Raudvere, U., Kuzmin, I., Adler, P., Vilo, J., and Peterson, H. (2023). g:Profiler—interoperable web service for functional enrichment analysis and gene identifier mapping (2023 update). *Nucleic Acids Res.* 51 (W1), W207–W212. doi:10.1093/nar/gkad347
- Komor, A. C., Badran, A. H., and Liu, D. R. (2017). CRISPR-based technologies for the manipulation of eukaryotic genomes. *Cell* 168 (1–2), 20–36. doi:10.1016/j.cell.2016.10.044
- Krzywinski, M., and Altman, N. (2013). Power and sample size. *Nat. Methods* 10 (12), 1139–1140. doi:10.1038/nmeth.2738
- Kumasaka, N., Knights, A., and Gaffney, D. (2019). High resolution genetic mapping of putative causal interactions between regions of open chromatin. *Nat. Genet.* 51 (1), 128–137. doi:10.1038/s41588-018-0278-6
- Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., Heravi-Moussavi, A., et al. (2015). Integrative analysis of 111 reference human epigenomes. *Nature* 518 (7539), 317–330. doi:10.1038/nature14248
- Kwasniewski, J. C., Mogno, I., Myers, C. A., Corbo, J. C., and Cohen, B. A. (2012). Complex effects of nucleotide variants in a mammalian cis-regulatory element. *Proc. Natl. Acad. Sci. U. S. A.* 109 (47), 19498–19503. doi:10.1073/pnas.1210678109
- Li, B., and Ritchie, M. D. (2021). From GWAS to gene: transcriptome-wide association studies and other methods to functionally understand GWAS discoveries. *Front. Genet.* 12, 713230. doi:10.3389/fgene.2021.713230
- Li, G., Fullwood, M. J., Xu, H., Mulawadi, F. H., Velkov, S., Vega, V., et al. (2010). ChIA-PET tool for comprehensive chromatin interaction analysis with paired-end tag sequencing. *Genome Biol.* 11 (2), R22. doi:10.1186/gb-2010-11-2-r22
- Li, T., Yang, Y., Qi, H., Cui, W., Zhang, L., Fu, X., et al. (2023a). CRISPR/Cas9 therapeutics: progress and prospects. *Signal Transduct. Target Ther.* 8 (1), 36–23. doi:10.1038/s41392-023-01309-7
- Li, L., Ma, X., Cui, Y., Rotival, M., Chen, W., Zou, X., et al. (2023b). Immune-response 3'UTR alternative polyadenylation quantitative trait loci contribute to variation in human complex traits and diseases. *Nat. Commun.* 14 (1), 8347. doi:10.1038/s41467-023-44191-1
- Liang, K. Y. H., Farjoun, Y., Forgetta, V., Chen, Y., Yoshiji, S., Lu, T., et al. (2023). Predicting ExWAS findings from GWAS data: a shorter path to causal genes. *Hum. Genet.* 142 (6), 749–758. doi:10.1007/s00439-023-02548-y
- Liao, Y., Wang, J., Jaehnig, E. J., Shi, Z., and Zhang, B. (2019). WebGestalt 2019: gene set analysis toolkit with revamped UIs and APIs. *Nucleic Acids Res.* 47 (W1), W199–W205–W205. doi:10.1093/nar/gkz401
- Lieberman-Aiden, E., van Berkum, N. L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., et al. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* 326 (5950), 289–293. doi:10.1126/science.1181369
- Lin, Z., Akin, H., Rao, R., Hie, B., Zhu, Z., Lu, W., et al. (2023). Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science* 379 (6637), 1123–1130. doi:10.1126/science.ade2574
- Liu, S., and Crawford, D. C. (2022). Maturation and application of phenome-wide association studies. *Trends Genet.* 38 (4), 353–363. doi:10.1016/j.tig.2021.12.002
- Liu, L., Zeng, P., Xue, F., Yuan, Z., and Zhou, X. (2021). Multi-trait transcriptome-wide association studies with probabilistic Mendelian randomization. *Am. J. Hum. Genet.* 108 (2), 240–256. doi:10.1016/j.ajhg.2020.12.006
- Lizio, M., Harshbarger, J., Shimoji, H., Severin, J., Kasukawa, T., Sahin, S., et al. (2015). Gateways to the FANTOM5 promoter level mammalian expression atlas. *Genome Biol.* 16 (1), 22. doi:10.1186/s13059-014-0560-6
- Lizio, M., Abugessaisa, I., Noguchi, S., Kondo, A., Hasegawa, A., Hon, C. C., et al. (2019). Update of the FANTOM web resource: expansion to provide additional transcriptome atlases. *Nucleic Acids Res.* 47 (D1), D752–D758–8. doi:10.1093/nar/gky1099
- Long, E., Yin, J., Funderburk, K. M., Xu, M., Feng, J., Kane, A., et al. (2022). Massively parallel reporter assays and variant scoring identified functional variants and target genes for melanoma loci and highlighted cell-type specificity. *Am. J. Hum. Genet.* 109 (12), 2210–2229. doi:10.1016/j.ajhg.2022.11.006
- Loos, R. J. F., and Yeo, G. S. H. (2014). The bigger picture of FTO: the first GWAS-identified obesity gene. *Nat. Rev. Endocrinol.* 10 (1), 51–61. doi:10.1038/nrendo.2013.227
- Lu, X., Chen, X., Forney, C., Donmez, O., Miller, D., Parameswaran, S., et al. (2021). Global discovery of lupus genetic risk variant allelic enhancer activity. *Nat. Commun.* 12 (1), 1611. doi:10.1038/s41467-021-21854-5
- Luo, Y., Hitz, B. C., Gabdank, I., Hilton, J. A., Kagda, M. S., Lam, B., et al. (2020). New developments on the Encyclopedia of DNA Elements (ENCODE) data portal. *Nucleic Acids Res.* 48 (D1), D882–D889–9. doi:10.1093/nar/gkz1062
- Lyons, E., and Freeling, M. (2008). How to usefully compare homologous plant genes and chromosomes as DNA sequences. *Plant J.* 53 (4), 661–673. doi:10.1111/j.1365-313X.2007.03326.x
- Lyu, C., Huang, M., Liu, N., Chen, Z., Lupo, P. J., Tycko, B., et al. (2021). Detecting methylation quantitative trait loci using a methylation random field method. *Brief. Bioinform.* 22 (6), bbab323. doi:10.1093/bib/bbab323
- Ma, J., Köster, J., Qin, Q., Hu, S., Li, W., Chen, C., et al. (2016). CRISPR-DO for genome-wide CRISPR design and optimization. *Bioinform. Oxf. Engl.* 32 (21), 3336–3338. doi:10.1093/bioinformatics/btw476

- Maere, S., Heymans, K., and Kuiper, M. (2005). BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinforma. Oxf Engl.* 21 (16), 3448–3449. doi:10.1093/bioinformatics/bti551
- Maleki, F., Owens, K., Hogan, D. J., and Kusalik, A. J. (2020). Gene set analysis: challenges, opportunities, and future research. *Front. Genet.* 11, 654. doi:10.3389/fgene.2020.00654
- Manolio, T. A., Collins, F. S., Cox, N. J., Goldstein, D. B., Hindorf, L. A., Hunter, D. J., et al. (2009). Finding the missing heritability of complex diseases. *Nature* 461 (7265), 747–753. doi:10.1038/nature08494
- Mansisor, A. R., and Risca, V. I. (2022). Chromatin accessibility: methods, mechanisms, and biological insights. *Nucleus* 13 (1), 236–276. doi:10.1080/19491034.2022.2143106
- Martens, M., Ammar, A., Riutta, A., Waagmeester, A., Slenker, D. N., Hanspers, K., et al. (2021). WikiPathways: connecting communities. *Nucleic Acids Res.* 49 (D1), D613–D621. doi:10.1093/nar/gkaa1024
- Martin, F. J., Amode, M. R., Aneja, A., Austine-Orimoloye, O., Azov, A. G., Barnes, I., et al. (2023). Ensembl 2023. *Nucleic Acids Res.* 51 (D1), D933–D941. doi:10.1093/nar/gkac958
- Mashhour, M. A., Kandil, A. H., AbdElwahed, M., and Mabrouk, M. S. (2024). Harmony in transcripts: a systematic literature review of transcriptome-wide association studies. *J. Eng. Appl. Sci.* 71 (1), 167. doi:10.1186/s44147-024-00499-3
- Matoba, N., Liang, D., Sun, H., Aygün, N., McAfee, J. C., Davis, J. E., et al. (2020). Common genetic risk variants identified in the SPARK cohort support DDHD2 as a candidate risk gene for autism. *Transl. Psychiatry* 10 (1), 265–314. doi:10.1038/s41398-020-00953-9
- Matthews, L. J., and Turkheimer, E. (2022). Three legs of the missing heritability problem. *Stud. Hist. Philos. Sci.* 93, 183–191. doi:10.1016/j.shpsa.2022.04.004
- McArthur, E., and Capra, J. A. (2021). Topologically associating domain boundaries that are stable across diverse cell types are evolutionarily constrained and enriched for heritability. *Am. J. Hum. Genet.* 108 (2), 269–283. doi:10.1016/j.ajhg.2021.01.001
- McCarthy, M. I., Abecasis, G. R., Cardon, L. R., Goldstein, D. B., Little, J., Ioannidis, J. P. A., et al. (2008). Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nat. Rev. Genet.* 9 (5), 356–369. doi:10.1038/nrg2344
- McLaren, W., Gil, L., Hunt, S. E., Riat, H. S., Ritchie, G. R. S., Thormann, A., et al. (2016). The Ensembl variant effect predictor. *Genome Biol.* 17 (1), 122. doi:10.1186/s13059-016-0974-4
- Meddens, C. A., Harakalova, M., van den Dungen, N. A. M., Foroughi Asl, H., Hijma, H. J., Cuppen, EPJG, et al. (2016). Systematic analysis of chromatin interactions at disease associated loci links novel candidate genes to inflammatory bowel disease. *Genome Biol.* 17, 247. doi:10.1186/s13059-016-1100-3
- Melnikov, A., Murugan, A., Zhang, X., Tesileanu, T., Wang, L., Rogov, P., et al. (2012). Systematic dissection and optimization of inducible enhancers in human cells using a massively parallel reporter assay. *Nat. Biotechnol.* 30 (3), 271–277. doi:10.1038/nbt.2137
- Melnikov, A., Zhang, X., Rogov, P., Wang, L., and Mikkelsen, T. S. (2014). Massively parallel reporter assays in cultured mammalian cells. *J. Vis. Exp. JoVe* (90), 51719. doi:10.3791/51719
- Momozawa, Y., and Mizukami, K. (2021). Unique roles of rare variants in the genetics of complex diseases in humans. *J. Hum. Genet.* 66 (1), 11–23. doi:10.1038/s10038-020-00845-2
- Mooney, M. A., and Wilmot, B. (2015). Gene set analysis: a step-by-step guide. *Am. J. Med. Genet. Part B Neuropsychiatr. Genet. Off. Publ. Int. Soc. Psychiatr. Genet.* 168 (7), 517–527. doi:10.1002/ajmg.b.32328
- Mootha, V. K., Lindgren, C. M., Eriksson, K. F., Subramanian, A., Sihag, S., Lehar, J., et al. (2003). PGC-1 α -responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat. Genet.* 34 (3), 267–273. doi:10.1038/ng1180
- Mountjoy, E., Schmidt, E. M., Carmona, M., Schwartzentruber, J., Peat, G., Miranda, A., et al. (2021). An open approach to systematically prioritize causal variants and genes at all published human GWAS trait-associated loci. *Nat. Genet.* 53 (11), 1527–1533. doi:10.1038/s41588-021-00945-5
- Mouri, K., Guo, M. H., de Boer, C. G., Lissner, M. M., Harten, I. A., Newby, G. A., et al. (2022). Prioritization of autoimmune disease-associated genetic variants that perturb regulatory element activity in T cells. *Nat. Genet.* 54 (5), 603–612. doi:10.1038/s41588-022-01056-5
- Nasser, J., Bergman, D. T., Fulco, C. P., Guckelberger, P., Doughty, B. R., Patwardhan, T. A., et al. (2021). Genome-wide enhancer maps link risk variants to disease genes. *Nature* 593 (7858), 238–243. doi:10.1038/s41586-021-03446-x
- Naumova, N., Smith, E. M., Zhan, Y., and Dekker, J. (2012). Analysis of long-range chromatin interactions using Chromosome Conformation Capture. *Methods San. Diego Calif.* 58 (3), 192–203. doi:10.1016/j.ymeth.2012.07.022
- Nicholls, H. L., John, C. R., Watson, D. S., Munroe, P. B., Barnes, M. R., and Cabrera, C. P. (2020). Reaching the end-game for GWAS: machine learning approaches for the prioritization of complex disease loci. *Front. Genet.* 11, 350. doi:10.3389/fgene.2020.00350
- Noguchi, S., Arakawa, T., Fukuda, S., Furuno, M., Hasegawa, A., Hori, F., et al. (2017). FANTOM5 CAGE profiles of human and mouse samples. *Sci. Data* 4 (1), 170112. doi:10.1038/sdata.2017.112
- Okada, Y., Kubo, M., Ohmiya, H., Takahashi, A., Kumasaka, N., Hosono, N., et al. (2012). Common variants at CDKAL1 and KLF9 are associated with body mass index in east Asian populations. *Nat. Genet.* 44 (3), 302–306. doi:10.1038/ng.1086
- Ozadam, H., Tonn, T., Han, C. M., Segura, A., Hoskins, I., Rao, S., et al. (2023). Single-cell quantification of ribosome occupancy in early mouse development. *Nature* 618 (7967), 1057–1064. doi:10.1038/s41586-023-06228-9
- Palermo, J., Chesi, A., Zimmerman, A., Sonti, S., Pahl, M. C., Lasconi, C., et al. (2023). Variant-to-gene mapping followed by cross-species genetic screening identifies GPI-anchored biosynthesis as a regulator of sleep. *Sci. Adv.* 9 (1), eabq0844. doi:10.1126/sciadv.abq0844
- Pan, D. Z., Garske, K. M., Alvarez, M., Bhagat, Y. V., Boocock, J., Nikkola, E., et al. (2018). Integration of human adipocyte chromosomal interactions with adipose gene expression prioritizes obesity-related genes from GWAS. *Nat. Commun.* 9 (1), 1512. doi:10.1038/s41467-018-03554-9
- Papaemmanuil, E., Hosking, F. J., Vijayakrishnan, J., Price, A., Olver, B., Sheridan, E., et al. (2009). Loci on 7p12.2, 10q21.2 and 14q11.2 are associated with risk of childhood acute lymphoblastic leukemia. *Nat. Genet.* 41 (9), 1006–1010. doi:10.1038/ng.430
- Park, E., Jiang, Y., Hao, L., Hui, J., and King, Y. (2021). Genetic variation and microRNA targeting of A-to-I RNA editing fine tune human tissue transcriptomes. *Genome Biol.* 22 (1), 77. doi:10.1186/s13059-021-02287-1
- Penner-Goeke, S., Bothe, M., Rek, N., Kreitmaier, P., Pöhlchen, D., Kühnel, A., et al. (2023). High-throughput screening of glucocorticoid-induced enhancer activity reveals mechanisms of stress-related psychiatric disorders. *Proc. Natl. Acad. Sci. U. S. A.* 120 (49), e2305773120. doi:10.1073/pnas.2305773120
- Peterson, R. E., Kuchenbaecker, K., Walters, R. K., Chen, C. Y., Popejoy, A. B., Periyasamy, S., et al. (2019). Genome-wide association studies in ancestrally diverse populations: opportunities, methods, pitfalls, and recommendations. *Cell.* 179 (3), 589–603. doi:10.1016/j.cell.2019.08.051
- Pippin, J. A., Chesi, A., Wagley, Y., Su, C., Pahl, M. C., Hodge, K. M., et al. (2021). CRISPR-Cas9-Mediated genome editing confirms EPDR1 as an effector gene in the BMD GWAS-implicated 'STARD3NL' locus. *JBMR Plus* 5 (9), e10531. doi:10.1002/jbmr.4.10531
- Porcu, E., Rieger, S., Lepik, K., Santoni, F. A., Reymond, A., Kutalik, Z., et al. (2019). Mendelian randomization integrating GWAS and eQTL data reveals genetic determinants of complex and clinical traits. *Nat. Commun.* 10 (1), 3300. doi:10.1038/s41467-019-10936-0
- Powder, K. E. (2020). Quantitative trait loci (QTL) mapping. *Methods Mol. Biol. Clifton N. J.* 2082, 211–229. doi:10.1007/978-1-0716-0026-9_15
- Ramachandran, D., Dennis, J., Fachal, L., Schürmann, P., Bousset, K., Hülse, F., et al. (2022). Genome-wide association study and functional follow-up identify 14q12 as a candidate risk locus for cervical cancer. *Hum. Mol. Genet.* 31 (15), 2483–2497. doi:10.1093/hmg/ddac031
- Ramilowski, J. A., Yip, C. W., Agrawal, S., Chang, J. C., Ciani, Y., Kulakovskiy, I. V., et al. (2020). Functional annotation of human long noncoding RNAs via molecular phenotyping. *Genome Res.* 30 (7), 1060–1072. doi:10.1101/gr.254219.119
- Raychaudhuri, S. (2011). Mapping rare and common causal alleles for complex human diseases. *Cell.* 147 (1), 57–69. doi:10.1016/j.cell.2011.09.011
- Rivas, M. A., Beaudoin, M., Gardet, A., Stevens, C., Sharma, Y., Zhang, C. K., et al. (2011). Deep resequencing of GWAS loci identifies independent rare variants associated with inflammatory bowel disease. *Nat. Genet.* 43 (11), 1066–1073. doi:10.1038/ng.952
- Rodchenkov, I., Babur, O., Luna, A., Aksoy, B. A., Wong, J. V., Fong, D., et al. (2020). Pathway Commons 2019 Update: integration, analysis and exploration of pathway data. *Nucleic Acids Res.* 48 (D1), D489–D497–97. doi:10.1093/nar/gkz946
- Sarker, I. H. (2022). AI-based modeling: techniques, applications and research issues towards automation, intelligent and smart systems. *Sn Comput. Sci.* 3 (2), 158. doi:10.1007/s42979-022-01043-x
- Schaid, D. J., Chen, W., and Larson, N. B. (2018). From genome-wide associations to candidate causal variants by statistical fine-mapping. *Nat. Rev. Genet.* 19 (8), 491–504. doi:10.1038/s41576-018-0016-z
- Scherer, M., Gasparoni, G., Rahmouni, S., Shashkova, T., Arnoux, M., Louis, E., et al. (2021). Identification of tissue-specific and common methylation quantitative trait loci in healthy individuals using Magar. *Epigenetics Chromatin* 14 (1), 44. doi:10.1186/s13072-021-00415-6
- Schipper, M., and Posthuma, D. (2022). Demystifying non-coding GWAS variants: an overview of computational tools and methods. *Hum. Mol. Genet.* 31 (R1), R73–R83. doi:10.1093/hmg/ddac198
- Schwartz, S., Zhang, Z., Frazer, K. A., Smit, A., Riemer, C., Bouck, J., et al. (2000). PipMaker—a web server for aligning two genomic DNA sequences. *Genome Res.* 10 (4), 577–586. doi:10.1101/gr.10.4.577
- Schwartzentruber, J., Cooper, S., Liu, J. Z., Barrio-Hernandez, I., Bello, E., Kumasaka, N., et al. (2021). Genome-wide meta-analysis, fine-mapping, and integrative prioritization implicate new Alzheimer's disease risk genes. *Nat. Genet.* 53 (3), 392–402. doi:10.1038/s41588-020-00776-w
- Serdar, C. C., Cihan, M., Yücel, D., and Serdar, M. A. (2021). Sample size, power and effect size revisited: simplified and practical approaches in pre-clinical, clinical and laboratory studies. *Biochem. Medica* 31, 010502. doi:10.11613/BM.2021.010502

- Sigala, R. E., Lagou, V., Shmeliov, A., Atito, S., Kouchaki, S., Awais, M., et al. (2023). Machine learning to advance human genome-wide association studies. *Genes*. 15 (1), 34. doi:10.3390/genes15010034
- Simonis, M., Klous, P., Splinter, E., Moshkin, Y., Willemsen, R., de Wit, E., et al. (2006). Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). *Nat. Genet.* 38 (11), 1348–1354. doi:10.1038/ng1896
- Sinnott-Armstrong, N., Sousa, I. S., Laber, S., Rendina-Ruedy, E., Nitter Dankel, S. E., Ferreira, T., et al. (2021). A regulatory variant at 3q21.1 confers an increased pleiotropic risk for hyperglycemia and altered bone mineral density. *Cell. Metab.* 33 (3), 615–628.e13. doi:10.1016/j.cmet.2021.01.001
- Slager, S. L., Achenbach, S. J., Asmann, Y. W., Camp, N. J., Rabe, K. G., Goldin, L. R., et al. (2013). Mapping of the IRF8 gene identifies a 3' UTR variant associated with risk of chronic lymphocytic leukemia but not other common non-Hodgkin lymphoma subtypes. *Cancer Epidemiol. Biomark. Prev. Publ. Am. Assoc. Cancer Res. Cosponsored Am. Soc. Prev. Oncol.* 22 (3), 461–466. doi:10.1158/1055-9965.EPI-12-1217
- Smemo, S., Tena, J. J., Kim, K. H., Gamazon, E. R., Sakabe, N. J., Gómez-Marín, C., et al. (2014). Obesity-associated variants within FTO form long-range functional connections with IRX3. *Nature* 507 (7492), 371–375. doi:10.1038/nature13138
- Soldner, F., Stelzer, Y., Shivalila, C. S., Abraham, B. J., Latourelle, J. C., Barrasa, M. I., et al. (2016). Parkinson-associated risk variant in distal enhancer of α -synuclein modulates target gene expression. *Nature* 533 (7601), 95–99. doi:10.1038/nature17939
- Soleimanpour, S. A., Gupta, A., Bakay, M., Ferrari, A. M., Groff, D. N., Fadista, J., et al. (2014). The diabetes susceptibility gene Clec16a regulates mitophagy. *Cell*. 157 (7), 1577–1590. doi:10.1016/j.cell.2014.05.016
- Song, L., and Crawford, G. E. (2010). DNase-seq: a high-resolution technique for mapping active gene regulatory elements across the genome from mammalian cells. *Cold Spring Harb. Protoc.* 2010 (2), prot5384. doi:10.1101/pdb.prot5384
- Soskic, B., Cano-Gamez, E., Smyth, D. J., Ambridge, K., Ke, Z., Matte, J. C., et al. (2022). Immune disease risk variants regulate gene expression dynamics during CD4+ T cell activation. *Nat. Genet.* 54 (6), 817–826. doi:10.1038/s41588-022-01066-3
- Srivastava, R., Rolyan, H., Xie, Y., Li, N., Bhat, N., Hong, L., et al. (2019). TCF7L2 (transcription factor 7-like 2) regulation of GATA6 (GATA-Binding protein 6)-dependent and -independent vascular smooth muscle cell plasticity and intimal hyperplasia. *Arterioscler. Thromb. Vasc. Biol.* 39 (2), 250–262. doi:10.1161/ATVBAHA.118.311830
- Stunnenberg, H. G., and Hirst, M. (2016). The international human epigenome Consortium: a blueprint for scientific collaboration and discovery. *Cell*. 167 (5), 1145–1149. doi:10.1016/j.cell.2016.11.007
- Sturm, N., Mayr, A., Le Van, T., Chupakhin, V., Ceulemans, H., Wegner, J., et al. (2020). Industry-scale application and evaluation of deep learning for drug target prediction. *J. Cheminformatics* 12 (1), 26. doi:10.1186/s13321-020-00428-5
- Su, C., Johnson, M. E., Torres, A., Thomas, R. M., Manduchi, E., Sharma, P., et al. (2020). Mapping effector genes at lupus GWAS loci using promoter Capture-C in follicular helper T cells. *Nat. Commun.* 11 (1), 3294. doi:10.1038/s41467-020-17089-5
- Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., et al. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci.* 102 (43), 15545–15550. doi:10.1073/pnas.0506580102
- Suhre, K., McCarthy, M. I., and Schwenk, J. M. (2021). Genetics meets proteomics: perspectives for large population-based studies. *Nat. Rev. Genet.* 22 (1), 19–37. doi:10.1038/s41576-020-0268-2
- Szkarczyk, D., Gable, A. L., Nastou, K. C., Lyon, D., Kirsch, R., Pyysalo, S., et al. (2021). The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res.* 49 (D1), D605–D612. doi:10.1093/nar/gkaa1074
- Tarca, A. L., Draghici, S., Bhatti, G., and Romero, R. (2012). Down-weighting overlapping genes improves gene set analysis. *BMC Bioinforma.* 13, 136. doi:10.1186/1471-2105-13-136
- Tehranchi, A. K., Myrthil, M., Martin, T., Hie, B. L., Golan, D., and Fraser, H. B. (2016). Pooled ChIP-seq links variation in transcription factor binding to complex disease risk. *Cell*. 165 (3), 730–741. doi:10.1016/j.cell.2016.03.041
- Tewhey, R., Kotliar, D., Park, D. S., Liu, B., Winnicki, S., Reilly, S. K., et al. (2016). Direct identification of hundreds of expression-modulating variants using a multiplexed reporter assay. *Cell*. 165 (6), 1519–1529. doi:10.1016/j.cell.2016.04.027
- The Genotype-Tissue Expression (GTEx) project (2013). The genotype-tissue expression (GTEx) project. *Nat. Genet.* 45 (6), 580–585. doi:10.1038/ng.2653
- Thomas, P. D., Ebert, D., Muruganujan, A., Mushayama, T., Albou, L. P., and Mi, H. (2022). PANTHER: making genome-scale phylogenetics accessible to all. *Protein Sci.* 31 (1), 8–22. doi:10.1002/pro.4218
- Tomczak, A., Mortensen, J. M., Winnenburger, R., Liu, C., Alessi, D. T., Swamy, V., et al. (2018). Interpretation of biological experiments changes with evolution of the Gene Ontology and its annotations. *Sci. Rep.* 8 (1), 5115. doi:10.1038/s41598-018-23395-2
- Tordai, H., Torres, O., Csepi, M., Padányi, R., Lukács, G. L., and Hegedűs, T. (2024). Analysis of AlphaMissense data in different protein groups and structural context. *Sci. Data* 11 (1), 495. doi:10.1038/s41597-024-03327-8
- Uffelmann, E., Huang, Q. Q., Munung, N. S., de Vries, J., Okada, Y., Martin, A. R., et al. (2021). Genome-wide association studies. *Nat. Rev. Methods Primer* 26, 59. doi:10.1038/s43586-021-00056-9
- Ulirsch, J. C., Nandakumar, S. K., Wang, L., Giani, F. C., Zhang, X., Rogov, P., et al. (2016). Systematic functional dissection of common genetic variation affecting red blood cell traits. *Cell*. 165 (6), 1530–1545. doi:10.1016/j.cell.2016.04.048
- Ustugova, A. S., Korneev, K. V., Kuprash, D. V., and Afanasyeva, M. A. (2019). Functional SNPs in the human autoimmunity-associated locus 17q12-21. *Genes*. 10 (2), 77. doi:10.3390/genes10020077
- Villicaña, S., and Bell, J. T. (2021). Genetic impacts on DNA methylation: research findings and future perspectives. *Genome Biol.* 22 (1), 127. doi:10.1186/s13059-021-02347-6
- Wainschtein, P., Jain, D., Zheng, Z., Cupples, L. A., Shadyab, A. H., McKnight, B., et al. (2022). Assessing the contribution of rare variants to complex trait heritability from whole-genome sequence data. *Nat. Genet.* 54 (3), 263–273. doi:10.1038/s41588-021-00997-7
- Walker, J. T., Saunders, D. C., Rai, V., Chen, H. H., Orchard, P., Dai, C., et al. (2023). Genetic risk converges on regulatory networks mediating early type 2 diabetes. *Nature* 624 (7992), 621–629. doi:10.1038/s41586-023-06693-2
- Wang, Y., Zhang, Z. T., Seo, S. O., Lynn, P., Lu, T., Jin, Y. S., et al. (2016). Bacterial genome editing with CRISPR-cas9: deletion, integration, single nucleotide modification, and desirable “clean” mutant selection in *Clostridium beijerinckii* as an example. *ACS Synth. Biol.* 5 (7), 721–732. doi:10.1021/acssynbio.6b00060
- Wang, G., Sarkar, A., Carbonetto, P., and Stephens, M. (2020). A simple new approach to variable selection in regression, with application to genetic fine mapping. *J. R. Stat. Soc. Ser. B Stat. Methodol.* 82 (5), 1273–1300. doi:10.1111/rssb.12388
- Wang, T. M., Xiao, R. W., He, Y. Q., Zhang, W. L., Diao, H., Tang, M., et al. (2023). High-throughput identification of regulatory elements and functional assays to uncover susceptibility genes for nasopharyngeal carcinoma. *Am. J. Hum. Genet.* 110 (7), 1162–1176. doi:10.1016/j.ajhg.2023.06.003
- Warde-Farley, D., Donaldson, S. L., Comes, O., Zuberi, K., Badrawi, R., Chao, P., et al. (2010). The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Res.* 38 (Suppl. 1_2), W214–W220. doi:10.1093/nar/gkq537
- Watt, S., Vasquez, L., Walter, K., Mann, A. L., Kundu, K., Chen, L., et al. (2021). Genetic perturbation of PU.1 binding and chromatin looping at neutrophil enhancers associates with autoimmune disease. *Nat. Commun.* 12 (1), 2298. doi:10.1038/s41467-021-22548-8
- Weeks, E. M., Ulirsch, J. C., Cheng, N. Y., Trippe, B. L., Fine, R. S., Miao, J., et al. (2023). Leveraging polygenic enrichments of gene features to predict genes underlying complex traits and diseases. *Nat. Genet.* 55 (8), 1267–1276. doi:10.1038/s41588-023-01443-6
- Wen, W., Cho, Y. S., Zheng, W., Dorajoo, R., Kato, N., Qi, L., et al. (2012). Meta-analysis identifies common variants associated with body mass index in east Asians. *Nat. Genet.* 44 (3), 307–311. doi:10.1038/ng.1087
- White, M. A., Myers, C. A., Corbo, J. C., and Cohen, B. A. (2013). Massively parallel *in vivo* enhancer assay reveals that highly local features determine the cis-regulatory function of ChIP-seq peaks. *Proc. Natl. Acad. Sci. U. S. A.* 110 (29), 11952–11957. doi:10.1073/pnas.1307449110
- Wishart, D. S., Li, C., Marcu, A., Badran, H., Pon, A., Budinski, Z., et al. (2020). PathBank: a comprehensive pathway database for model organisms. *Nucleic Acids Res.* 48 (D1), D470–D478–8. doi:10.1093/nar/gkz861
- Wong, Y. Y., Harbison, J. E., Hope, C. M., Gundambuu, B., Brown, K. A., Wong, S. W., et al. (2023). Parallel recovery of chromatin accessibility and gene expression dynamics from frozen human regulatory T cells. *Sci. Rep.* 13 (1), 5506. doi:10.1038/s41598-023-32256-6
- Xie, S., Duan, J., Li, B., Zhou, P., and Hon, G. C. (2017). Multiplexed engineering and analysis of combinatorial enhancer activity in single cells. *Mol. Cell*. 66 (2), 285–299. doi:10.1016/j.molcel.2017.03.007
- Xie, Y., Shan, N., Zhao, H., and Hou, L. (2021). Transcriptome wide association studies: general framework and methods. *Quant. Biol.* 9 (2), 141–150. doi:10.15302/j-qb-020-0228
- Xu, F., Yu, E. Y. W., Cai, X., Yue, L., Jing, L., peng, Liang, X., et al. (2023). Genome-wide genotype-serum proteome mapping provides insights into the cross-ancestry differences in cardiometabolic disease susceptibility. *Nat. Commun.* 14 (1), 896. doi:10.1038/s41467-023-36491-3
- Xue, A., Wu, Y., Zhu, Z., Zhang, F., Kemper, K. E., Zheng, Z., et al. (2018). Genome-wide association analyses identify 143 risk variants and putative regulatory mechanisms for type 2 diabetes. *Nat. Commun.* 9 (1), 2941. doi:10.1038/s41467-018-04951-w
- Yang, Z., Wang, C., Liu, L., Khan, A., Lee, A., Vardarajan, B., et al. (2023a). CARMA is a new Bayesian model for fine-mapping in genome-wide association meta-analyses. *Nat. Genet.* 55 (6), 1057–1065. doi:10.1038/s41588-023-01392-0

- Yang, Z., Zeng, X., Zhao, Y., and Chen, R. (2023b). AlphaFold2 and its applications in the fields of biology and medicine. *Signal Transduct. Target Ther.* 8 (1), 115–214. doi:10.1038/s41392-023-01381-z
- Yu, M., Georges, A., Tucker, N. R., Kyryachenko, S., Toomer, K., Schott, J. J., et al. (2019). Genome-wide association study-driven gene-set analyses, genetic, and functional follow-up suggest GLIS1 as a susceptibility gene for mitral valve prolapse. *Circ. Genomic Precis. Med.* 12 (5), e002497. doi:10.1161/CIRCGEN.119.002497
- Yuan, Z., Zhu, H., Zeng, P., Yang, S., Sun, S., Yang, C., et al. (2020). Testing and controlling for horizontal pleiotropy with probabilistic Mendelian randomization in transcriptome-wide association studies. *Nat. Commun.* 11 (1), 3861. doi:10.1038/s41467-020-17668-6
- Zeeberg, B. R., Feng, W., Wang, G., Wang, M. D., Fojo, A. T., Sunshine, M., et al. (2003). GoMiner: a resource for biological interpretation of genomic and proteomic data. *Genome Biol.* 4 (4), R28. doi:10.1186/gb-2003-4-4-r28
- Zhang, J., and Zhao, H. (2023). eQTL studies: from bulk tissues to single cells. *J. Genet. Genomics Yi Chuan Xue Bao.* 18 (23), S1673–S8527.
- Zhang, H., Gelernter, J., Gruen, J. R., Kranzler, H. R., Herman, A. I., and Simen, A. A. (2010). Functional impact of a single nucleotide polymorphism in the OPRD1 promoter region. *J. Hum. Genet.* 55 (5), 278–284. doi:10.1038/jhg.2010.22
- Zhang, Y., Wang, M., Li, Z., Yang, X., Li, K., Xie, A., et al. (2024). An overview of detecting gene-trait associations by integrating GWAS summary statistics and eQTLs. *Sci. China Life Sci.* 67 (6), 1133–1154. doi:10.1007/s11427-023-2522-8
- Zhao, L., Wang, S., Cao, Z., Ouyang, W., Zhang, Q., Xie, L., et al. (2019a). Chromatin loops associated with active genes and heterochromatin shape rice genome architecture for transcriptional regulation. *Nat. Commun.* 10 (1), 3640. doi:10.1038/s41467-019-11535-9
- Zhao, T., Hu, Y., Zang, T., and Wang, Y. (2019b). Integrate GWAS, eQTL, and mQTL data to identify Alzheimer's disease-related genes. *Front. Genet.* 10, 1021. doi:10.3389/fgene.2019.01021
- Zhao, Y., Wang, J., Chen, J., Zhang, X., Guo, M., and Yu, G. (2020). A literature review of gene function prediction by modeling gene ontology. *Front. Genet.* 11, 400. doi:10.3389/fgene.2020.00400
- Zhao, S., Crouse, W., Qian, S., Luo, K., Stephens, M., and He, X. (2024). Adjusting for genetic confounders in transcriptome-wide association studies improves discovery of risk genes of complex traits. *Nat. Genet.* 56 (2), 336–347. doi:10.1038/s41588-023-01648-9
- Zheng, Z., Huang, D., Wang, J., Zhao, K., Zhou, Y., Guo, Z., et al. (2020). QTLbase: an integrative resource for quantitative trait loci across multiple human molecular phenotypes. *Nucleic Acids Res.* 48 (D1), D983–D991–91. doi:10.1093/nar/gkz888
- Zhou, J., and Troyanskaya, O. G. (2015). Predicting effects of noncoding variants with deep learning-based sequence model. *Nat. Methods* 12 (10), 931–934. doi:10.1038/nmeth.3547
- Zhu, H., and Zhou, X. (2021). Transcriptome-wide association studies: a view from Mendelian randomization. *Quant. Biol. Beijing China* 9 (2), 107–121. doi:10.1007/s40484-020-0207-4
- Zuber, V., Grinberg, N. F., Gill, D., Manipur, I., Slob, E. A. W., Patel, A., et al. (2022). Combining evidence from Mendelian randomization and colocalization: review and comparison of approaches. *Am. J. Hum. Genet.* 109 (5), 767–782. doi:10.1016/j.ajhg.2022.04.001
- Zufferey, M., Tavernari, D., and Ciriello, G. (2022). Methods for the analysis of topologically associating domains (TADs). *Methods Mol. Biol. Clifton NJ.* 2301, 39–59. doi:10.1007/978-1-0716-1390-0_3