Published in final edited form as:

Hum Genet. 2020 January ; 139(1): 95–102. doi:10.1007/s00439-019-02044-2.

Identifying causal variants and genes using functional genomics in specialized cell types and contexts

Boxiang Liu,

Department of Biology, Stanford University

Stephen B. Montgomery

Department of Pathology, Stanford University

Abstract

A central goal in human genetics is the identification of variants and genes that influence the risk of polygenic diseases. In the past decade, genome-wide association studies (GWAS) have identified tens of thousands of genetic loci associated with various diseases. Since the majority of such loci lie within non-coding regions and have many candidate variants in linkage disequilibrium, it has been challenging to accurately identify specific causal variants and genes. To aid in their discovery a variety of statistical and experimental approaches have been developed. These approaches often borrow information from functional genomics assays such as ATAC-seq, ChIP-seq and RNA-seq to annotate functional variants and identify regulatory relationships between variants and genes. While such approaches are powerful, given the diversity of cell types and environments, it is paramount to select disease-relevant contexts for follow-up analyses. In this review, we discuss the latest developments, challenges, and best practices for determining the causal mechanisms of polygenic disease risk variants with functional genomics data from specialized cell types.

Keywords

polygenic diseases; fine-mapping; functional genomics; specialized cell types

1 Introduction

Prior to the publication of the draft human genome and availability of high-throughput genotyping arrays, the identification of disease variants and genes was laborious and mostly limited to monogenic disorders. A notable example is cystic fibrosis, caused by abnormal regulation of ion transport in epithelial cells. The causal gene *CFTR* was discovered by cloning overlapping DNA segments over the disease-linked locus (Riordan et al., 1989). Unlike such successes in monogenic disorders, pinpointing variants and genes for complex, polygenic traits has remained difficult. Such traits may have tens to hundreds of risk variants where each variant explains only a small proportion of disease heritability. To accomplish this feat, genetic association studies on polygenic traits required a better resolved human

genome, a map of common genetic variants and sufficient sample sizes to gain statistical power. The completion of the draft of the human genome (International Human Genome Sequencing Consortium, 2001; Venter et al., 2001) provided an essential roadmap to the locations of human genes and set the foundation to genome-wide mapping of common risk variants. Soon after the publication of the human genome, the international HapMap consortium was established to produce the first multi-ethnic map of the human genome and common genetic variants. This effort led to a dense map of the genome covering more than 10 million common genetic variants across 1,184 reference individuals from 11 global populations to enable imputation on data from genotyping arrays and genome-wide association studies (GWAS) (The International HapMap 3 Consortium, 2010). Aided by the HapMap reference panel, the first large-scale GWAS sampled around 2000 cases for 7 common diseases and around 3000 shared normal controls (Wellcome Trust Case Control Consortium, 2007). This study marked the foundation of a decade-long acceleration in GWAS research to identify risk loci for diverse polygenic diseases.

A compelling finding across GWAS indicates that most complex, polygenic traits have tens to hundreds of independent risk loci, of which only a small fraction locate within coding regions. Despite the abundance of non-coding risk variants, their functional effects require increased scrutiny due to the absence of a clear link to protein sequences. Few studies, through considerable effort, have been successful in uncovering the functional mechanism of non-coding, disease-associated variants. An example is the *SORT1* loci for coronary artery disease, in which a non-coding variant rs12740374 creates a *CEBP* transcription factor binding site and alter the expression of *SORT1* in liver (Musunuru et al., 2010). For most traits however, the functional mechanisms have yet to be identified. Identification of the correct cell or tissue type remains a bottleneck and often times an essential first step in determining the causal mechanism(s). To address these challenges, multi-tissue reference datasets have been established (GTEx Consortium, 2017; Ward and Kellis, 2016; Dunham et al., 2012). Furthermore, advances in statistical and experimental approaches have facilitated more accurate mapping of causal non-coding risk variants, genes and mechanisms. In this review, we discuss advances in GWAS follow-up using functional genomics strategies, and highlight several studies that have pioneered these advances. We discuss existing large-scale, multi-tissue functional data sets with a focus on ENCODE, Roadmap Epigenomics Project, and GTEx, and present examples that necessitate the use of specialized cell types and contexts. We further describe approaches to combine new data with these reference datasets and highlight downstream statistical and experimental approaches to identify and interpret risk loci.

2 Existing datasets for follow-up of complex disease risk loci

Multi-tissue functional genomics datasets are powerful resources for the follow-up of polygenic disease-associated loci. To establish reference functional datasets, several consortia have conducted studies to delineate the genetic regulatory landscape across human tissues and cell types (GTEx Consortium, 2017; Dunham et al., 2012; Roadmap Epigenomics Consortium, 2015). Following the completion of the human genome project (International Human Genome Sequencing Consortium, 2001), the National Human Genome Research Institute launched the Encyclopedia of DNA element (ENCODE) project

to characterize all functional elements in the human genome (Dunham et al., 2012). Among the 3 billion basepairs in the human genome, only about 1.5% code for proteins. A key goal of ENCODE project is to identify the function of the remaining regions of the human genome. Using numerous techniques, including Hi-C, ATAC-seq, ChIP-seq, bisulfite sequencing and RNA-seq, the ENCODE project has profiled chromatin interaction and structure, transcription factor binding, histone modifications, methylation and transcription over hundreds of tissues, cell lines, primary cells and *in vitro* differentiated cells. At the time of writing, more than 10,000 assays have been made public (<https://www.encodeproject.org/>). A close sibling to the ENCODE project is the Roadmap Epigenomics project that produced a genome-wide regulatory map by jointly analyzing DNA accessibility, DNA methylation, histone marks, and RNA expression across 127 diverse cell and tissue types (Roadmap Epigenomics Consortium, 2015).

Complementing the ENCODE project, the Genotype-Tissue Expression (GTEx) project aims to identify regulatory variants by profiling gene expression and genotype information across a large number of individuals and tissues (GTEx Consortium, 2017). Collectively, the GTEx consortium has contributed RNA-seq and genotype data for over 600 individuals across 53 tissues and cell types. This reference dataset enables tissue-dependent mapping of expression and splicing quantitative trait loci (eQTL; sQTL) aiding in identifying the regulatory functions of non-coding variants. Specifically, GTEx has demonstrated that tissue-sharing of regulatory exhibits a bimodal pattern - an eQTL is likely to be either shared across all tissues or specific to a few. Furthermore, tissues with larger sample sizes showed greater numbers of tissue-specific eQTLs. However, despite the large number of tissues profiled (approximately 20 from each individual), many cell types remain to be individually discovered and profiled (Regev et al., 2017). This presents an ongoing challenge to interpret complex diseases with etiologies in deeply embedded cell types that are difficult to identify or obtain. To date, only a few studies have used specialized cell types to follow-up GWAS variants. We provide several examples in the next section.

3 Cell-type specific regulatory elements and mechanisms

A key to identifying pathologically-relevant genes is the selection of relevant tissues and cell types. A number of key cell types have been implicated in various polygenic diseases. Type-2 diabetes is a complex disease caused by dysregulation of insulin produced by beta cells in the islets of Langerhans (DeFronzo et al., 2015). To study the transcriptomic signature of pancreatic beta cells, Nica and colleagues purified beta cells from 11 individuals to compare against whole islets and beta-cell-depleted islets and identified expression and transcriptomic signatures specific to beta and non-beta cells (Nica et al., 2013). Mahajan and colleagues fine-mapped type 2 diabetes risk loci with T2D-relevant tissues including islets, liver, adipose and skeletal muscle, and discovered that 28 mapped to islet enhancers or promoter regions, 14 of which are islet specific (Mahajan et al., 2018). A second complex disease with significant metabolic component is coronary artery disease. Several cell types, including smooth muscle cells, endothelial cells, and immune cells, contributes to coronary artery disease (Khera and Kathiresan, 2017). We previously cultured primary human coronary artery smooth muscle cells (HCASMC) to follow-up coronary artery disease risk loci, and identified five candidate genes including growth factors (*SIPA1* and *PDGFRA*),

transcription factors (*TCF21*), and other genes (*FES* and *SMAD3*) (Liu et al., 2018). By comparison with large reference datasets (ENCODE and GTEx), we identified a colocalization between a *SIPA1* eQTL and a CAD GWAS signal is the strongest in HCASMC.

Many complex diseases lack clear annotations for key cell types. These diseases often require the identification of relevant cell types as a first step. Obesity is a polygenic disease with a multifactorial etiology that involves the central nervous system, energy metabolism, and adipose homeostasis. Claussnitzer and colleagues focused on the *FTO* locus, a well-replicated risk locus for obesity (Claussnitzer et al., 2015). In this locus, causal mechanisms involving the pancreas and the brain had further been suggested (Ragvin et al., 2010; Smemo et al., 2014a). To identify the relevant cell-type, they examined enhancer annotation across 127 tissue and cell-types from the ENCODE and Roadmap datasets and identified an unusually long enhancer in mesenchymal adipose progenitors. To confirm this region indeed carries enhancer function in adipose progenitors, they subsequently performed luciferase reporter assays in human adipocytes and four other cell types and determined that the enhancer activity is specific to adipocytes. Using a long-range chromatin interaction (Hi-C) map from human adipocyte progenitor cells, Claussnitzer and colleagues identified eight candidate genes in 3D contact with the *FTO* locus through chromatin looping. To reduce the search space, they profiled gene expression in mesenchymal adipocyte progenitors and found that the expression of *IRX3* and *IRX5* was influenced by variants in the *FTO* locus. Zooming in on these two genes, they demonstrated that overexpression of *IRX3* and *IRX5* leads to a cell-autonomous shift from beige (energy-dissipating) to white (energy-storing) adipocytes and reduces mitochondrial thermogenesis. Conversely, knockdown of *IRX3* and *IRX5* in primary adipocytes restored thermogenesis. Last but not least, they discovered that a risk variant rs1421085 lies in a *ARID5B* (a transcriptional repressor) motif. CRISPR-Cas9 genome editing from the risk to the protective allele restored *IRX3* and *IRX5* repression.

Identification of relevant cell types can also borrow information from cell-type specific eQTLs. eQTLs with small effect sizes, which are more likely to be discovered in large sample sizes, tend to have tissue-specific effects but are also less likely to replicate (GTEx Consortium, 2017). Small and colleagues identified a diabetes risk allele that reduces the expression of *KLF14* (a transcription factor) and modulates the expression of 385 target genes in adipose tissue (Small et al., 2018). They found no eQTL signal in skin, whole blood, or lymphoblastoid cell lines from the TwinsUK cohort (Moayyeri et al., 2013) or any tissue other than adipose from GTEx (GTEx Consortium, 2017). Focusing on the adipose tissue, they showed that reduced *KLF14* expression disrupts lipogenesis in human cell lines and demonstrated that *KLF14* knockout mouse demonstrated insulin resistance. Further, they showed that human carriers of the *KLF14* risk allele shift body fat from gynoid stores to abdominal stores, and their adipocyte cell size increases significantly.

For cell types with multiple states or environmental contexts (e.g. naive and stimulated), the cellular state can strongly influence follow-up results. Alasoo and colleagues studied the effect of IFN γ and Salmonella treatment on eQTLs in human macrophages (Alasoo et al., 2018). The systemic lupus erythematosus risk locus centered at rs11997338 colocalized with an eQTL signal from treated but not from naive macrophages. This implies that

interpretation of variants with eQTL requires careful examination of the cellular state. Similarly, we have found cellular state to be an important consideration in our study design. To understand the metabolic effect on genetic regulation in retinal pigment epithelium (RPE) cells, we treated cells with glucose and galactose growth media (Liu et al., 2019a). In total, we found 687 shared, 264 glucose-specific and 166 galactose-specific eQTLs, many of which have known implications in eye diseases. For example, a glucose-specific eQTL *ABCA1* has been associated with glaucoma (Chen et al., 2014) and a member of the WD repeat protein family, *WDR5*, demonstrated glucose-specific colocalization with GWAS for age-related macular degeneration.

4 Comparison between reference and cell-type specific datasets

To identify genetic elements specific to a novel cell type of interest, it can be useful to compare this cell type with established reference datasets. Since the two are likely to have been studied at different time point in separate facilities and processed independently, these comparisons can suffer from batch effects and other hidden technical confounders. One example of such a batch effect was demonstrated in the study of Nica and colleagues, where the pancreatic islet cells cluster away from all other cell types in a principal component analysis (Nica et al., 2013). Several steps can be taken to reduce batch effects. Batch effects from standard genomic analyses (e.g. alignment and quantification) can be mitigated by following the same pipeline used to process the reference datasets. Further, quantification metrics that control for library sizes, such as FPKM (Trapnell et al., 2010) and TPM (Wagner et al., 2012), will reduce batch effects from sequencing depth. In our experience, these aforementioned strategies can be sufficient in removing considerable batch effects present in clustering analysis (Liu et al., 2019a, 2018). Other analyses, such as differential expression and splicing analyses, are more sensitive to batch effects and require explicit modeling of hidden confounders. Packages such as surrogate variable analysis (sva) (Leek and Storey, 2007), probabilistic estimation of expression residuals (PEER) (Stegle et al., 2012) and remove unwanted variation (RUV) (Risso et al., 2014) are among the various methods to identify hidden confounders.

Specialized cell types often requires unique extraction and culturing protocols, which limits their availability and the number of samples that can be studied. Strategies leveraging external or internal information can be used to boost power to discover eQTLs. Studies have shown that joint modeling of shared regulatory effect across multiple tissues and cell types can boost discovery power (Sul et al., 2013; Flutre et al., 2013). MetaTissue (Sul et al., 2013) and mashR (Urbut et al., 2018) are two recent packages for jointly QTL modeling across tissues. Furthermore, integration of epigenomic or ATAC-seq data can enhance power when integrated into QTL mapping approaches using the TORUS software (Wen, 2016).

5 GWAS follow-up with functional data from specialized cell types

Since the majority of risk variants lie in non-coding regions, identifying the causal mechanism can be challenging. The pursuit of causal mechanisms broadly encompasses three interconnected tasks: GWAS fine-mapping, gene identification and experimental validation (Figure 1). Schaid and colleagues have written a comprehensive review on

statistical fine-mapping of GWAS (Schaid et al., 2018). Here, fine-mapping of causal variants with penalized regression or Bayesian variable selection techniques can be performed independent of cell or tissue type (Benner et al., 2016; Hormozdiari et al., 2014). However, the abundance of functional genomic data has prompted the development of methods that integrate functional annotation as prior information. Packages such as PAINTOR (Kichaev et al., 2014), fgwas (Pickrell, 2014), and CAVIARBF (Chen et al., 2016) leverage these annotations to improve fine-mapping accuracy.

Gene identification can be performed independently from fine-mapping. When molecular QTL data is available, packages such as eCAVIAR (Hormozdiari et al., 2016), coloc (Giambartolomei et al., 2014), and enloc (Wen et al., 2017) can estimate the posterior probability of colocalization, defined as one or more shared causal variants between the molecular QTL and GWAS interval. TWAS (Gusev et al., 2016) and PrediXcan (Gamazon et al., 2015) represent a second class of methods that use cis-eQTL data to perform association between GWAS and gene expression imputed from cis-eQTL effects. These methods do not provide posterior probability on the SNP-level. Instead, their main objective is identification of target genes. A closely related idea is summary-data-based Mendelian Randomization (SMR), which considers a regulatory variant to be a naturally occurring perturbation on gene expression and performs association between the perturbed gene expression and GWAS (Zhu et al., 2016). Despite the availability of multiple methods a main challenge lies in discrepant results across each. As an illustration, we previously used eCAVIAR (Hormozdiari et al., 2016) and SMR (Zhu et al., 2016) to identify coronary artery disease genes with HCASMC eQTLs (Liu et al., 2018). Among five candidate genes, only one was identified by both methods. As a complement to statistical methods, we have found it necessary to visually confirm the top-ranked candidate genes and colocalizations. A *bona fide* colocalization should have similar p-value distributions across the molecular QTL and GWAS datasets. One can verify such pattern with a pair of LocusZoom plots stacked on top of each other, one for each dataset. LocusCompare is another visualization technique that plots the GWAS against the eQTL p-values in a scatter plot (Liu et al., 2019b). In addition, due to the ambiguity in computational identification of target genes, experimental approaches that directly observe chromatin conformation are often used to complement computational findings. For example, the *FTO* locus has a strong association with obesity and is replicated across age groups and ancestries (Loos and Yeo, 2014). However, no significant association was found between *FTO* genotype and gene expression and the target genes of the *FTO* locus remained unclear. Smemo and colleagues used 4C in mouse embryos and brain to identify an interaction between the risk locus and *Irx3* promoter (Smemo et al., 2014b). Claussnitzer and colleagues later used Hi-C to identify *IRX5* as an additional target gene in human adipose progenitor cells (Claussnitzer et al., 2015). Gallagher and colleagues have written an excellent review on this general pipeline for identification and validation of target genes (Gallagher and Chen-Plotkin, 2018).

To confirm hypotheses generated by computational and high-throughput experimental approaches, follow-up experiments on the candidate gene or variants are necessary. One important reason that necessitates follow-up experiments is that the colocalization analyses are often conducted on cell types in their resting state, and a disease-relevant or environmentally-perturbed state may be required to verify the proper involvement of the

causal gene. Techniques such as CRISPR-Cas9 genome editing, gene knockout/knockdown experiments are among the numerous methods designed to test for regulatory effects (Visscher et al., 2017; Gaj et al., 2016; Tewhey et al., 2018). One challenge in testing regulatory effect in specialized cell types is the limited availability. Cultured primary cells often fail to grow sufficient quantity for genome editing and are prone to apoptosis after transfection. It may be necessary to use proxy cell types for experimental validation, such as those from model organisms, closely related cell lines, and iPSC-derived cell types (Edwards et al., 2013).

6 Challenges and Future Directions

Despite enormous progress in the past decade, several key challenges must be addressed to maximize the potential to use specialized cell types in GWAS follow-up analyses. Further, comparisons across different methods, such as colocalization, are still lacking, likely due to challenges in obtaining standardized test datasets and metrics. Although a single optimal method may not exist, we believe a comprehensive benchmark will provide insight into the strengths and weaknesses of each method.

A second challenge is the lack of comprehensive coverage on specialized cell types. In recent years, single-cell sequencing has shown potential to identify rare and novel cell types. For example, Plasschaert and colleagues used single-cell RNA-seq to discover pulmonary ionocytes, a new cell population in human airways that represent less than 1% of all cells. Downstream assays show that ionocytes are a major source of *CFTR* activity, and likely play a key role in cystic fibrosis (Plasschaert et al., 2018). Further, there is a lack in the understanding of developmental stages in specialized cells. For instance, new arteries can arise from vein during development and regeneration, but this process is poorly understood. Su and colleagues used single-cell analysis to show that vein cells of the developing heart undergo fate conversion to establish a pre-artery population that later differentiate into coronary artery (Su et al., 2018). Such single-cell profiling represents a new venue for discovering new cells in different developmental stages and under different contexts. Coupled with GWAS discoveries, single-cell sequencing represents a promising direction to uncover regulatory effect specific to rare cell types.

Despite enormous advances in genome-wide association studies, functional mechanisms of many risk loci remain to be discovered. Identification of causal mechanisms with new cell types, disease and environmental contexts have helped to explain several unknown risk loci, and will be a rapidly expanding part of future GWAS follow-up studies.

Acknowledgements

BL is supported by the Stanford Center for Evolution and Human Genomics fellowship and National Key R&D Program of China, 2016YFD0400800. SBM is supported by R33HL120757 (NHLBI), U01HG009431 (NHGRI; ENCODE4), R01MH101814 (NIH Common Fund; GTEx Program), R01HG008150 (NHGRI; Non-Coding Variants Program), R01HL142015 (NHLBI; TOPMED) and U01HG009080 (NHGRI; GSPAC).

References

- Alasoo K, Rodrigues J, Mukhopadhyay S, Knights AJ, Mann AL, Kundu K, Hale C, Dougan G, Gaffney DJ, Consortium H (2018) Shared genetic effects on chromatin and gene expression indicate a role for enhancer priming in immune response. *Nature Genetics* 50(3):424–, DOI 10.1038/s41588-018-0046-7, URL <GotoISI>://WOS:000427933400018 [PubMed: 29379200]
- Benner C, Spencer CCA, Havulinna AS, Salomaa V, Ripatti S, Pirinen M (2016) Finemap: efficient variable selection using summary data from genome-wide association studies. *Bioinformatics* 32(10):1493–1501, DOI 10.1093/bioinformatics/btw018, URL 10.1093/bioinformatics/btw018https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/btw018, URL https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/btw018 [PubMed: 26773131]
- Chen W, McDonnell SK, Thibodeau SN, Tillmans LS, Schaid DJ (2016) Incorporating functional annotations for fine-mapping causal variants in a bayesian framework using summary statistics. *Genetics* 204(3):933–958, DOI 10.1534/genetics.116.188953, URL 10.1534/genetics.116.188953https://www.ncbi.nlm.nih.gov/pubmed/27655946, URL https://www.ncbi.nlm.nih.gov/pubmed/27655946 [PubMed: 27655946]
- Chen Y, Lin Y, Vithana EN, Jia L, Zuo X, Wong TY, Chen LJ, Zhu X, Tam POS, Gong B, Qian S, Li Z, Liu X, Mani B, Luo Q, Guzman C, Leung CKS, Li X, Cao W, Yang Q, Tham CCY, Cheng Y, Zhang X, Wang N, Aung T, Khor CC, Pang CP, Sun X, Yang Z (2014) Common variants near *abca1* and in *pmm2* are associated with primary open-angle glaucoma. *Nature Genetics* 46(10):1115–1119, DOI 10.1038/ng.3078, URL 10.1038/ng.3078https://www-nature-com.laneproxy.stanford.edu/articles/ng.3078, URL https://www-nature-com.laneproxy.stanford.edu/articles/ng.3078 [PubMed: 25173107]
- Claussnitzer M, Dankel SN, Kim KH, Quon G, Meuleman W, Haugen C, Glunk V, Sousa IS, Beaudry JL, Puviondran V, Abdennur NA, Liu J, Svensson PA, Hsu YH, Drucker DJ, Mellgren G, Hui CC, Hauner H, Kellis M (2015) Fto obesity variant circuitry and adipocyte browning in humans. *The New England Journal of Medicine* 373(10):895–907, DOI 10.1056/NEJMoa1502214, URL 10.1056/NEJMoa1502214http://www.nejm.org/doi/abs/10.1056/NEJMoa1502214, URL http://www.nejm.org/doi/abs/10.1056/NEJMoa1502214 [PubMed: 26287746]
- DeFronzo RA, Ferrannini E, Groop L, Henry RR, Herman WH, Holst JJ, Hu FB, Kahn CR, Raz I, Shulman GI, Simonson DC, Testa MA, Weiss R (2015) Type 2 diabetes mellitus. *Nat Rev Dis Primers* 1:15019, DOI 10.1038/nrdp.2015.19, URL 10.1038/nrdp.2015.19https://www.ncbi.nlm.nih.gov/pubmed/27189025, URL https://www.ncbi.nlm.nih.gov/pubmed/27189025 [PubMed: 27189025]
- Dunham I, Kundaje A, Aldred SF, Collins PJ, Davis C, Doyle F, Epstein CB, Frietze S, Harrow J, Kaul R, Khatun J, Lajoie BR, Landt SG, Lee BK, Pauli F, Rosenbloom KR, Sabo P, Safi A, Sanyal A, Shores N, Simon JM, Song L, Trinklein ND, Altshuler RC, Birney E, Brown JB, Cheng C, Djebali S, Dong XJ, Dunham I, Ernst J, Furey TS, Gerstein M, Giardine B, Greven M, Hardison RC, Harris RS, Herrero J, Hoffman MM, Iyer S, Kellis M, Khatun J, Kheradpour P, Kundaje A, Lassmann T, Li QH, Lin X, Marinov GK, Merkel A, Mortazavi A, Parker SCJ, Reddy TE, Rozowsky J, Schlesinger F, Thurman RE, Wang J, Ward LD, Whitfield TW, Wilder SP, Wu W, Xi HLS, Yip KY, Zhuang JL, Bernstein BE, Birney E, Dunham I, Green ED, Gunter C, Snyder M, Pazin MJ, Lowdon RF, Dillon LAL, Adams LB, Kelly CJ, Zhang J, Wexler JR, Green ED, Good PJ, Feingold EA, Bernstein BE, Birney E, Crawford GE, Dekker J, Elnitski L, Farnham PJ, Gerstein M, Giddings MC, Gingeras TR, Green ED, Guigo R, Hardison RC, Hubbard TJ, Kellis M, Kent WJ, Lieb JD, Margulies EH, Myers RM, Snyder M, Stamatoyannopoulos JA, Tenenbaum SA, et al. (2012) An integrated encyclopedia of dna elements in the human genome. *Nature* 489(7414):57–74, DOI 10.1038/nature11247, URL <GotoISI>://WOS:000308347000039 [PubMed: 22955616]
- Edwards SL, Beesley J, French JD, Dunning AM (2013) Beyond gwas: Illuminating the dark road from association to function. *American Journal of Human Genetics* 93(5):779–797, DOI 10.1016/j.ajhg.2013.10.012, URL <GotoISI>://WOS:000326996600001 [PubMed: 24210251]
- Flutre T, Wen X, Pritchard J, Stephens M (2013) A statistical framework for joint eqtl analysis in multiple tissues. *PLoS genetics* 9(5):e1003486, DOI 10.1371/journal.pgen.1003486, URL 10.1371/journal.pgen.1003486http://dx.plos.org/10.1371/journal.pgen.1003486, URL http://dx.plos.org/10.1371/journal.pgen.1003486 [PubMed: 23671422]

- Gaj T, Sirk SJ, Shui SL, Liu J (2016) Genome-editing technologies: Principles and applications. *Cold Spring Harb Perspect Biol* 8(12), DOI 10.1101/cshperspect.a023754 , URL 10.1101/cshperspect.a023754<https://www.ncbi.nlm.nih.gov/pubmed/27908936>, URL <https://www.ncbi.nlm.nih.gov/pubmed/27908936>
- Gallagher MD, Chen-Plotkin AS (2018) The post-gwas era: From association to function. *Am J Hum Genet* 102(5):717–730, DOI 10.1016/j.ajhg.2018.04.002 , URL 10.1016/j.ajhg.2018.04.002<https://www.ncbi.nlm.nih.gov/pubmed/29727686>, URL <https://www.ncbi.nlm.nih.gov/pubmed/29727686> [PubMed: 29727686]
- Gamazon ER, Wheeler HE, Shah KP, Mozaffari SV, Aquino-Michaels K, Carroll RJ, Eyster AE, Denny JC, Consortium G, Nicolae DL, Cox NJ, Im HK (2015) A gene-based association method for mapping traits using reference transcriptome data. *Nature Genetics* 47(9):1091–1098, DOI 10.1038/ng.3367 , URL 10.1038/ng.3367<https://www-nature-com.laneproxy.stanford.edu/articles/ng.3367> [PubMed: 26258848]
- Giambartolomei C, Vukcevic D, Schadt EE, Franke L, Hingorani AD, Wallace C, Plagnol V (2014) Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS genetics* 10(5):e1004383, DOI 10.1371/journal.pgen.1004383 , URL 10.1371/journal.pgen.1004383<http://dx.plos.org/10.1371/journal.pgen.1004383>, URL <http://dx.plos.org/10.1371/journal.pgen.1004383> [PubMed: 24830394]
- GTEX Consortium (2017) Genetic effects on gene expression across human tissues. *Nature* 550(7675): 204–213, DOI 10.1038/nature24277 , URL 10.1038/nature24277<http://dx.doi.org/10.1038/nature24277>, URL <http://dx.doi.org/10.1038/nature24277> [PubMed: 29022597]
- Gusev A, Ko A, Shi H, Bhatia G, Chung W, Penninx BW, Jansen R, de Geus EJ, Boomsma DI, Wright FA, Sullivan PF, Nikkola E, Alvarez M, Civelek M, Lusi AJ, Lehtimäki T, Raitoharju E, Kahonen M, Seppälä I, Raitakari OT, Kuusisto J, Laakso M, Price AL, Pajukanta P, Pasaniuc B (2016) Integrative approaches for large-scale transcriptome-wide association studies. *Nat Genet* 48(3): 245–52, DOI 10.1038/ng.3506 , URL 10.1038/ng.3506<https://www.ncbi.nlm.nih.gov/pubmed/26854917>, URL <https://www.ncbi.nlm.nih.gov/pubmed/26854917> [PubMed: 26854917]
- Hormozdiari F, Kostem E, Kang EY, Pasaniuc B, Eskin E (2014) Identifying causal variants at loci with multiple signals of association. *Genetics* 198(2):genetics.114.167908–508, DOI 10.1534/genetics.114.167908 , URL 10.1534/genetics.114.167908<http://www.genetics.org/cgi/doi/10.1534/genetics.114.167908>, URL <http://www.genetics.org/cgi/doi/10.1534/genetics.114.167908>
- Hormozdiari F, van de Bunt M, Segr AV, Li X, Joo JW, Bilow M, Sul JH, Sankaranarayanan S, Pasaniuc B, Eskin E (2016) Colocalization of gwas and eqtl signals detects target genes. *The American Journal of Human Genetics* 99(6):1245–1260, DOI 10.1016/j.ajhg.2016.10.003 , URL 10.1016/j.ajhg.2016.10.003<http://www.cell.com/article/S0002929716304396/fulltext>, URL <http://www.cell.com/article/S0002929716304396/fulltext> [PubMed: 27866706]
- International Human Genome Sequencing Consortium (2001) Initial sequencing and analysis of the human genome. *Nature* 409(6822):860–921, DOI 10.1038/35057062 , URL 10.1038/35057062<https://www-nature-com.laneproxy.stanford.edu/articles/35057062>, URL <https://www-nature-com.laneproxy.stanford.edu/articles/35057062> [PubMed: 11237011]
- Khera AV, Kathiresan S (2017) Genetics of coronary artery disease: discovery, biology and clinical translation. *Nature Reviews Genetics* 18(6):331–344, DOI 10.1038/nrg.2016.160 , URL 10.1038/nrg.2016.160<https://www-nature-com.laneproxy.stanford.edu/articles/nrg.2016.160>, URL <https://www-nature-com.laneproxy.stanford.edu/articles/nrg.2016.160>
- Kichaev G, Yang WY, Lindstrom S, Hormozdiari F, Eskin E, Price AL, Kraft P, Pasaniuc B (2014) Integrating functional data to prioritize causal variants in statistical fine-mapping studies. *PLoS Genet* 10(10):e1004722, DOI 10.1371/journal.pgen.1004722 , URL 10.1371/journal.pgen.1004722<https://www.ncbi.nlm.nih.gov/pubmed/25357204>, URL <https://www.ncbi.nlm.nih.gov/pubmed/25357204> [PubMed: 25357204]
- Leek JT, Storey JD (2007) Capturing heterogeneity in gene expression studies by surrogate variable analysis. *PLoS genetics* 3(9):1724–1735, DOI 10.1371/journal.pgen.0030161 , URL 10.1371/journal.pgen.0030161<http://dx.plos.org/10.1371/journal.pgen.0030161>, URL <http://dx.plos.org/10.1371/journal.pgen.0030161> [PubMed: 17907809]

- Liu B, Pjanic M, Wang T, Nguyen T, Gloudemans M, Rao A, Castano VG, Nurnberg S, Rader DJ, Elwyn S, Ingelsson E, Montgomery SB, Miller CL, Quertermous T (2018) Genetic regulatory mechanisms of smooth muscle cells map to coronary artery disease risk loci. *The American Journal of Human Genetics* 103(3):377–388, DOI 10.1016/j.ajhg.2018.08.001 , URL 10.1016/j.ajhg.2018.08.001<https://www.ncbi.nlm.nih.gov/pubmed/30146127>, URL <https://www.ncbi.nlm.nih.gov/pubmed/30146127> [PubMed: 30146127]
- Liu B, Calton MA, Abell NS, Benchorin G, Gloudemans MJ, Chen M, Hu J, Li X, Balliu B, Bok D, Montgomery SB, Vollrath D (2019a) Genetic analyses of human fetal retinal pigment epithelium gene expression suggest ocular disease mechanisms. *Communications Biology*
- Liu B, Gloudemans MJ, Rao AS, Ingelsson E, Montgomery SB (2019b) Abundant associations with gene expression complicate gwas follow-up. *Nat Genet* 51(5):768–769, DOI 10.1038/s41588-019-0404-0 , URL 10.1038/s41588-019-0404-0<https://www.ncbi.nlm.nih.gov/pubmed/31043754> [PubMed: 31043754]
- Loos RJ, Yeo GS (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 , URL 10.1038/nrendo.2013.227<https://www.ncbi.nlm.nih.gov/pubmed/24247219>, URL <https://www.ncbi.nlm.nih.gov/pubmed/24247219> [PubMed: 24247219]
- Mahajan A, Taliun D, Thurner M, Robertson NR, Torres JM, Rayner NW, Payne AJ, Steinthorsdottir V, Scott RA, Grarup N, Cook JP, Schmidt EM, Wuttke M, Sarnowski C, Magi R, Nano J, Gieger C, Trompet S, Lecoecur C, Preuss MH, Prins BP, Guo X, Bielak LF, Below JE, Bowden DW, Chambers JC, Kim YJ, Ng MCY, Petty LE, Sim X, Zhang W, Bennett AJ, Bork-Jensen J, Brummett CM, Canouil M, Ec Kardt KU, Fischer K, Kardia SLR, Kronenberg F, Lall K, Liu CT, Locke AE, Luan J, Ntalla I, Nylander V, Schonherr S, Schurmann C, Yengo L, Bottinger EP, Brandslund I, Christensen C, Dedoussis G, Florez JC, Ford I, Franco OH, Frayling TM, Giedraitis V, Hackinger S, Hattersley AT, Herder C, Ikram MA, Ingelsson M, Jorgensen ME, Jorgensen T, Kriebel J, Kuusisto J, Ligthart S, Lindgren CM, Linneberg A, Lyssenko V, Mamakou V, Meitinger T, Mohlke KL, Morris AD, Nadkarni G, Pankow JS, Peters A, Sattar N, Stancakova A, Strauch K, Taylor KD, Thorand B, Thorleifsson G, Thorsteinsdottir U, Tuomilehto J, Witte DR, Dupuis J, Peyser PA, Zeggini E, Loos RJF, Froguel P, Ingelsson E, Lind L, Groop L, Laakso M, Collins FS, Jukema JW, Palmer CNA, Grallert H, Metspalu A, et al. (2018) Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. *Nat Genet* 50(11):1505–1513, DOI 10.1038/s41588-018-0241-6 , URL 10.1038/s41588-018-0241-6<https://www.ncbi.nlm.nih.gov/pubmed/30297969>, URL <https://www.ncbi.nlm.nih.gov/pubmed/30297969> [PubMed: 30297969]
- Moayyeri A, Hammond CJ, Hart DJ, Spector TD (2013) The uk adult twin registry (twinsuk resource). *Twin Research and Human Genetics* 16(1):144–9, DOI 10.1017/thg.2012.89 , URL 10.1017/thg.2012.89<https://www.ncbi.nlm.nih.gov/pubmed/23088889>, URL <https://www.ncbi.nlm.nih.gov/pubmed/23088889> [PubMed: 23088889]
- Musunuru K, Strong A, Frank-Kamenetsky M, Lee NE, Ahfeldt T, Sachs KV, Li X, Li H, Kuperwasser N, Ruda VM, Pirruccello JP, Muchmore B, Prokunina-Olsson L, Hall JL, Schadt EE, Morales CR, Lund-Katz S, Phillips MC, Wong J, Cantley W, Racie T, Ejebe KG, Orho-Melandier M, Melander O, Kotliansky V, Fitzgerald K, Krauss RM, Cowan CA, Kathiresan S, Rader DJ (2010) From noncoding variant to phenotype via sort1 at the 1p13 cholesterol locus. *Nature* 466(7307):714–719, DOI 10.1038/nature09266 , URL 10.1038/nature09266<http://www.nature.com/articles/nature09266> [PubMed: 20686566]
- Nica AC, Ongen H, Irminger JC, Bosco D, Berner T, Antonarakis SE, Halban PA, Dermitzakis ET (2013) Cell-type, allelic, and genetic signatures in the human pancreatic beta cell transcriptome. *Genome Research* 23(9):1554–1562, DOI 10.1101/gr.150706.112 , URL 10.1101/gr.150706.112<http://genome.cshlp.org/cgi/doi/10.1101/gr.150706.112>, URL <http://genome.cshlp.org/cgi/doi/10.1101/gr.150706.112> [PubMed: 23716500]
- Pickrell JK (2014) Joint analysis of functional genomic data and genome-wide association studies of 18 human traits (vol 94, pg 559, 2014). *American Journal of Human Genetics* 95(1):126–126, DOI 10.1016/j.ajhg.2014.06.001, URL <GotoISI>://WOS:000338904100012
- Plasschaert LW, Zilionis R, Choo-Wing R, Savova V, Knehr J, Roma G, Klein AM, Jaffe AB (2018) A single-cell atlas of the airway epithelium reveals the cftr-rich pulmonary ionocyte. *Nature*

560(7718):377–+, DOI 10.1038/s41586-018-0394-6, URL <GotoISI>://WOS:000441673400042 [PubMed: 30069046]

- Ragvin A, Moro E, Fredman D, Navratilova P, Drivenes O, Engstrom PG, Alonso ME, de la Calle Mustienes E, Gomez Skarmeta JL, Tavares MJ, Casares F, Manzanares M, van Heyningen V, Molven A, Njolstad PR, Argenton F, Lenhard B, Becker TS (2010) Long-range gene regulation links genomic type 2 diabetes and obesity risk regions to hhex, sox4, and irx3. *Proc Natl Acad Sci U S A* 107(2):775–80, DOI 10.1073/pnas.0911591107, URL 10.1073/pnas.0911591107https://www.ncbi.nlm.nih.gov/pubmed/20080751, URL https://www.ncbi.nlm.nih.gov/pubmed/20080751 [PubMed: 20080751]
- Regev A, Teichmann SA, Lander ES, Amit I, Benoist C, Birney E, Boden-miller B, Campbell P, Carninci P, Clatworthy M, Clevers H, Deplancke B, Dunham I, Eberwine J, Eils R, Enard W, Farmer A, Fugger L, Gottgens B, Hacohen N, Haniffa M, Hemberg M, Kim S, Klenerman P, Kriegstein A, Lein E, Linnarsson S, Lundberg E, Lundberg J, Majumder P, Marioni JC, Merad M, Mhlanga M, Nawijn M, Netea M, Nolan G, Pe'er D, Phillipakis A, Ponting CP, Quake S, Reik W, Rozenblatt-Rosen O, Sanes J, Satija R, Schumacher TN, Shalek A, Shapiro E, Sharma P, Shin JW, Stegle O, Stratton M, Stubbington MJT, Theis FJ, Uhlen M, van Oudenaarden A, Wagner A, Watt F, Weissman J, Wold B, Xavier R, Yosef N (2017) The Human Cell Atlas. *Elife* 6
- Riordan JR, Rommens JM, Kerem B, Alon N, Rozmahel R, Grzelczak Z, Zielenski J, Lok S, Plavsky N, Chou JL, et al (1989) Identification of the cystic fibrosis gene: cloning and characterization of complementary dna. *Science* 245(4922):1066–1073, DOI 10.1126/science.2475911, URL 10.1126/science.2475911http://www.sciencemag.org/cgi/doi/10.1126/science.2475911, URL http://www.sciencemag.org/cgi/doi/10.1126/science.2475911 [PubMed: 2475911]
- Risso D, Ngai J, Speed TP, Dudoit S (2014) Normalization of rna-seq data using factor analysis of control genes or samples. *Nature Biotechnology* 32(9):896–902, DOI 10.1038/nbt.2931, URL 10.1038/nbt.2931http://www.nature.com/doi/10.1038/nbt.2931, URL http://www.nature.com/doi/10.1038/nbt.2931
- Roadmap Epigenomics Consortium (2015) Integrative analysis of 111 reference human epigenomes. *Nature* 518(7539):317–330, DOI 10.1038/nature14248, URL 10.1038/nature14248http://www.nature.com/doi/10.1038/nature14248, URL http://www.nature.com/doi/10.1038/nature14248 [PubMed: 25693563]
- Schaid DJ, Chen W, Larson NB (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, URL 10.1038/s41576-018-0016-zhttps://www.ncbi.nlm.nih.gov/pubmed/29844615, URL https://www.ncbi.nlm.nih.gov/pubmed/29844615 [PubMed: 29844615]
- Small KS, Todorovic M, Civelek M, El-Sayed Moustafa JS, Wang X, Simon MM, Fernandez-Tajes J, Mahajan A, Horikoshi M, Hugill A, Glastonbury CA, Quaye L, Neville MJ, Sethi S, Yon M, Pan C, Che N, Vinuela A, Tsai PC, Nag A, Buil A, Thorleifsson G, Raghavan A, Ding Q, Morris AP, Bell JT, Thorsteinsdottir U, Stefansson K, Laakso M, Dahlman I, Arner P, Gloyn AL, Musunuru K, Lusis AJ, Cox RD, Karpe F, McCarthy MI (2018) Regulatory variants at klf14 influence type 2 diabetes risk via a female-specific effect on adipocyte size and body composition. *Nature Genetics* 50(4):572–580, DOI 10.1038/s41588-018-0088-x, URL 10.1038/s41588-018-0088-xhttps://www.ncbi.nlm.nih.gov/pubmed/29632379, URL https://www.ncbi.nlm.nih.gov/pubmed/29632379 [PubMed: 29632379]
- Smemo S, Tena JJ, Kim KH, Gamazon ER, Sakabe NJ, Gomez-Marin C, Aneas I, Credidio FL, Sobreira DR, Wasserman NF, Lee JH, Puviindran V, Tam D, Shen M, Son JE, Vakili NA, Sung HK, Naranjo S, Acemel RD, Manzanares M, Nagy A, Cox NJ, Hui CC, Gomez-Skarmeta JL, Nobrega MA (2014a) Obesity-associated variants within fto form long-range functional connections with irx3. *Nature* 507(7492):371–+, DOI 10.1038/nature13138, URL <GotoISI>://WOS:000333029000037 [PubMed: 24646999]
- Smemo S, Tena JJ, Kim KH, Gamazon ER, Sakabe NJ, Gomez-Marin C, Aneas I, Credidio FL, Sobreira DR, Wasserman NF, Lee JH, Puviindran V, Tam D, Shen M, Son JE, Vakili NA, Sung HK, Naranjo S, Acemel RD, Manzanares M, Nagy A, Cox NJ, Hui CC, Gomez-Skarmeta JL, Nobrega MA (2014b) Obesity-associated variants within fto form long-range functional connections with irx3. *Nature* 507(7492):371–5, DOI 10.1038/nature13138, URL 10.1038/nature13138https://www.ncbi.nlm.nih.gov/pubmed/24646999, URL https://www.ncbi.nlm.nih.gov/pubmed/24646999 [PubMed: 24646999]

- Stegle O, Parts L, Piipari M, Winn J, Durbin R (2012) Using probabilistic estimation of expression residuals (peer) to obtain increased power and interpretability of gene expression analyses. *Nature Protocols* 7(3):500–507, DOI 10.1038/nprot.2011.457 , URL 10.1038/nprot.2011.457http://www.nature.com/doifinder/10.1038/nprot.2011.457, URL http://www.nature.com/doifinder/10.1038/nprot.2011.457 [PubMed: 22343431]
- Su TY, Stanley G, Sinha R, D'Amato G, Das S, Rhee S, Chang AH, Poduri A, Raftrey B, Dinh TT, Roper WA, Li G, Quinn KE, Caron KM, Wu S, Miquerol L, Butcher EC, Weissman I, Quake S, Red-Horse K (2018) Single-cell analysis of early progenitor cells that build coronary arteries. *Nature* 559(7714):356–+, DOI 10.1038/s41586-018-0288-7, URL <GotoISI>://WOS:000439059800048 [PubMed: 29973725]
- Sul JH, Han B, Ye C, Choi T, Eskin E (2013) Effectively identifying eqtls from multiple tissues by combining mixed model and meta-analytic approaches. *Plos Genetics* 9(6), DOI ARTNe100349110.1371/journal.pgen.1003491, URL <GotoISI>://WOS:000321222600003
- Tewhey R, Kotliar D, Park DS, Liu B, Winnicki S, Reilly SK, Andersen KG, Mikkelsen TS, Lander ES, Schaffner SF, Sabeti PC (2018) Direct identification of hundreds of expression-modulating variants using a multiplexed reporter assay. *Cell* 172(5):1132–1134, DOI 10.1016/j.cell.2018.02.021 , URL 10.1016/j.cell.2018.02.021https://www.ncbi.nlm.nih.gov/pubmed/29474912, URL https://www.ncbi.nlm.nih.gov/pubmed/29474912 [PubMed: 29474912]
- The International HapMap 3 Consortium (2010) Integrating common and rare genetic variation in diverse human populations. *Nature* 467(7311):52–58, DOI 10.1038/nature09298 , URL 10.1038/nature09298https://www-nature-com.laneproxy.stanford.edu/articles/nature09298, URL https://www-nature-com.laneproxy.stanford.edu/articles/nature09298 [PubMed: 20811451]
- Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L (2010) Transcript assembly and quantification by rna-seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature Biotechnology* 28(5):511–U174, DOI 10.1038/nbt.1621, URL <GotoISI>://WOS:000277452700032
- Urbut SM, Wang G, Carbonetto P, Stephens M (2018) Flexible statistical methods for estimating and testing effects in genomic studies with multiple conditions. *bioRxiv*
- Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA, Gocayne JD, Amanatides P, Ballew RM, Huson DH, Wortman JR, Zhang Q, Kodira CD, Zheng XH, Chen L, Skupski M, Subramanian G, Thomas PD, Zhang J, Miklos GLG, Nelson C, Broder S, Clark AG, Nadeau J, McKusick VA, Zinder N, Levine AJ, Roberts RJ, Simon M, Slayman C, Hunkapiller M, Bolanos R, Delcher A, Dew I, Fasulo D, Flanigan M, Florea L, Halpern A, Hannenhalli S, Kravitz S, Levy S, Mobarry C, Reinert K, Remington K, Abu-Threideh J, Beasley E, Biddick K, Bonazzi V, Brandon R, Cargill M, Chandramouliswaran I, Charlab R, Chaturvedi K, Deng Z, Di Francesco V, Dunn P, Eilbeck K, Evangelista C, Gabrielian AE, Gan W, Ge W, Gong F, Gu Z, Guan P, Heiman TJ, Higgins ME, Ji RR, Ke Z, Ketchum KA, Lai Z, Lei Y, Li Z, Li J, Liang Y, Lin X, Lu F, Merkulov GV, Milshina N, Moore HM, Naik AK, Narayan VA, Neelam B, Nusskern D, Rusch DB, Salzberg S, Shao W, Shue B, Sun J, Wang ZY, Wang A, Wang X, Wang J, Wei MH, Wides R, Xiao C, Yan C, et al. (2001) The sequence of the human genome. *Science* 291(5507):1304–1351, DOI 10.1126/science.1058040 , URL 10.1126/science.1058040http://www.sciencemag.org/lookup/doi/10.1126/science.1058040, URL http://www.sciencemag.org/lookup/doi/10.1126/science.1058040 [PubMed: 11181995]
- Visscher PM, Wray NR, Zhang Q, Sklar P, McCarthy MI, Brown MA, Yang J (2017) 10 years of gwas discovery: Biology, function, and translation. *American Journal of Human Genetics* 101(1):5–22, DOI 10.1016/j.ajhg.2017.06.005, URL <GotoISI>://WOS:000404886800001 [PubMed: 28686856]
- Wagner GP, Kin K, Lynch VJ (2012) Measurement of mrna abundance using rna-seq data: Rpkms measure is inconsistent among samples. *Theory in Biosciences* 131(4):281–285, DOI 10.1007/s12064-012-0162-3, URL <GotoISI>://WOS:000310975800008 [PubMed: 22872506]
- Ward LD, Kellis M (2016) Haploreg v4: systematic mining of putative causal variants, cell types, regulators and target genes for human complex traits and disease. *Nucleic Acids Res* 44(D1):D877–81, DOI 10.1093/nar/gkv1340 , URL 10.1093/nar/gkv1340https://www.ncbi.nlm.nih.gov/pubmed/26657631, URL https://www.ncbi.nlm.nih.gov/pubmed/26657631 [PubMed: 26657631]

- Wellcome Trust Case Control Consortium (2007) Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 447(7145):661–678, DOI 10.1038/nature05911 , URL 10.1038/nature05911 <http://www.nature.com/doifinder/10.1038/nature05911>, URL <http://www.nature.com/doifinder/10.1038/nature05911> [PubMed: 17554300]
- Wen X (2016) Molecular qtl discovery incorporating genomic annotations using bayesian false discovery rate control. *Ann Appl Stat* 10(3):1619–1638, DOI 10.1214/16-AOAS952 , URL 10.1214/16-AOAS952 <https://doi.org/10.1214/16-AOAS952>, URL <https://doi.org/10.1214/16-AOAS952>
- Wen X, Pique-Regi R, Luca F (2017) Integrating molecular qtl data into genome-wide genetic association analysis: Probabilistic assessment of enrichment and colocalization. *PLoS genetics* 13(3):e1006646, DOI 10.1371/journal.pgen.1006646 , URL 10.1371/journal.pgen.1006646 <http://dx.plos.org/10.1371/journal.pgen.1006646>, URL <http://dx.plos.org/10.1371/journal.pgen.1006646> [PubMed: 28278150]
- Zhu Z, Zhang F, Hu H, Bakshi A, Robinson MR, Powell JE, Montgomery GW, Goddard ME, Wray NR, Visscher PM, Yang J (2016) Integration of summary data from gwas and eqtl studies predicts complex trait gene targets. *Nature Genetics* 48(5):481–487, DOI 10.1038/ng.3538 , URL 10.1038/ng.3538 <https://www-nature-com.laneproxy.stanford.edu/articles/ng.3538>, URL <https://www-nature-com.laneproxy.stanford.edu/articles/ng.3538> [PubMed: 27019110]

Specialized Cellular Context

Fig. 1. Using specialized cell types to improve GWAS follow-up analysis.

Functional genomic data from specialized cell types can facilitate GWAS follow-up (1), identification of target gene (2), and linking genes to diseases through experimental validation. Various experimental and computational approaches work based on the prior knowledge about the specialized cell type in disease-relevant states. Several fine-mapping methods such as penalized regression or multi-ethnic fine-mapping are cell-agnostic and are not included in this figure.