Check for updates

# eQTLs as causal instruments for the reconstruction of hormone linked gene networks

Sean Bankier* and Tom Michoel

Computational Biology Unit, Department of Informatics, University of Bergen, Bergen, Norway

Hormones act within in highly dynamic systems and much of the phenotypic response to variation in hormone levels is mediated by changes in gene expression. The increase in the number and power of large genetic association studies has led to the identification of hormone linked genetic variants. However, the biological mechanisms underpinning the majority of these loci are poorly understood. The advent of affordable, high throughput next generation sequencing and readily available transcriptomic databases has shown that many of these genetic variants also associate with variation in gene expression levels as expression Quantitative Trait Loci (eQTLs). In addition to further dissecting complex genetic variation, eQTLs have been applied as tools for causal inference. Many hormone networks are driven by transcription factors, and many of these genes can be linked to eQTLs. In this mini-review, we demonstrate how causal inference and gene networks can be used to describe the impact of hormone linked genetic variation upon the transcriptome within an endocrinology context.

KEYWORDS

causal inference, hormones, genetics, eQTL, networks

## Introduction

Since the inception of Genome Wide Association Studies (GWAS), nearly two decades ago (1), there has been a steady expansion in the number of studies conducted as well as increases in sample size, yielding a wealth of new genetic associations with complex traits and disease. This approach has offered many new opportunities in endocrinology (2), where hormonal networks are well understood and hence lend themselves to informed mapping approaches. However, loci identified by GWAS alone are insufficient to elucidate the mechanisms by which these traits emerge (3) and efforts to understand the biology underpinning these associations has proved to be a significant challenge.

Much of the genetic research related to hormones has focused on monogenic endocrine disorders with scope for clinical intervention through genetic testing

schemes. Examples include Autoimmune Polyglandular Syndrome Type 1 (4) and IPEX syndrome (5) involving germline mutations within the *AIRE* and *FOXP3* genes respectively. However many genetic variants that contribute to different endocrine disorders and risk factors arise from common variants identified from GWAS (2). Hormone measurements have been exploited as GWAS traits to identify regions of the genome associated with hormone levels (Table 1), but the downstream consequences of these variants are still poorly understood. However, strides have also been made to link these GWAS hits to transcriptomic variation as expression Quantitative Trait Loci (eQTLs) (6).

This mini-review aims to describe how advances in population genetics have helped to identify regions of the genome linked to hormone variation, while highlighting how causal network inference and multi-omic integration can enhance these findings with added biological relevance and context. We first describe how GWAS have been used to identify genetic variants that are associated with complex traits, as well as discussing the limitations of association based approaches. The role of in systems genetics is discussed, including how these have been used to model the impact of genetic variation upon molecular phenotypes. Finally, we describe how causal inference methods have been used to overcome some of the limitations of GWAS and how these can be integrated for the reconstruction of causal molecular networks by using eQTLs as genetic instruments.

## Moving beyond GWAS

GWAS have exploded in popularity over the course of the last decade. In 2022, the NHGRI-EBI GWAS Catalog lists 5690 studies for more than 372,752 genetic associations (7). However most of the genome-wide significant loci identified are of low to moderate penetrance, exacerbating the issue of missing heritability that has been predicted for complex traits (8–11). This includes traits such as type II diabetes where only 10% of heritability is explained by the GWAS variants that have currently been identified (12), although twin and population studies estimate heritability to be between 20-80% (13). Missing heritability, combined with the uniform distribution of GWAS hits across the genome, has even led to speculation of an "omnigenic" model of inheritance in which all genes in trait-related cells play a functional role in the resultant phenotype (8).

The genetic drivers behind many common disease phenotypes present through complex multi-factorial models of inheritance (14), as is the case in instances of obesity (15), cardiovascular disease (16) and type II diabetes (12). The identification of causal Single Nucleotide Polymorphisms (SNPs) is further complicated by the presence of pleiotropy, the phenomenon whereby genetic variation can be seen to influence multiple phenotypic traits (17). Pleiotropy has been shown to be highly prevalent across the human genome (18), with studies showing that up to 90% of trait associated loci are associated with multiple traits (19).

TABLE 1   Top 10 GWAS as determined by discovery sample size, obtained from GWAS catalog (7) under EFO term "hormone measurment" (EFO_0004730).

| Reported Trait | Discovery sample number | Publication date | First author |
|---|---|---|---|
| Nonatopic asthma or fasting insulin levels | 502,660 European | 2019-10-24 | Zhu Z (37) |
| IGF-1 measurement | 435,516 European | 2021-07-05 | Barton AR (38) |
| Total testosterone levels | 425,097 European | 2020-02-10 | Ruth KS (39) |
| Testosterone levels x smoking behaviour interaction | 414,294 European | 2021-01-06 | Liang X (40) |
| IGF 1 (Gene-based burden) | 409,926 European | 2021-10-18 | Backman JD (41) |
| Sex hormone-binding globulin levels | 389,354 European | 2021-05-12 | Martin S (42) |
| Body mass index and fasting insulin (pairwise) | 374,012 European, 16,962 African American or Afro-Caribbean, East Asian, Hispanic or Latin American, South East Asian | 2021-02-22 | Huang LO (43) |
| Insulin-like growth factor 1 levels | 340,567 European, 5974 African unspecified, 7283 South Asian | 2021-01-18 | Sinnott-Armstrong N (44) |
| Circulating leptin levels or HOMA-IR | 254,263 Asian unspecified, European, Hispanic or Latin American, NR, South Asian | 2020-04-01 | Wang X (45) |
| Fasting insulin | 213,645 African American or Afro- Caribbean, East Asian, European, Hispanic or Latin American, South Asian | 2021-05-31 | Chen J (46) |

These limitations have encouraged the wider integration of multi-omic data to provide functional context for GWAS results, linking SNPs to intermediate molecular phenotypes using systems genetics approaches that consider the global response to genetic variation (20) (Figures 1A, B). This is particularly relevant in the case of hormone associated genetic variation, as many hormones mediate signalling across tissues (21) through transcriptional changes and can be modelled using these systems based approaches.

## Retrospective of eQTL studies

As transcriptomic data has become more readily available from highly powered studies, there has been a drive to link SNPs to variation in gene expression. High throughput sequencing technologies such as RNA-sequencing (RNA-seq) have facilitated the analysis of gene expression on a genome wide scale, replacing SNP microarrays as the leading method for gene expression analysis (22). In conjunction with the emergence of large deeply genotyped cohorts, this has allowed for the mapping of eQTLs on an unprecedented scale.

RNA-seq also provides the advantage of allowing for the measurement of Allele-Specific Expression (AES), where it is possible to measure the relative contributions of the maternal and paternal allele, something that is not possible in microarray based methods (23). RNA-seq estimates the expression of different genes through the total read count method, which measures the number of mapped sequence reads (24). This allows for the use of traditional eQTL mapping methods such as linear regression, but also facilitates the direct modelling of

gene expression of total read count using discrete distributions (22).

Linking GWAS loci to gene expression provides some indication of a functional relationship, and indeed data have demonstrated that trait-associated SNPs are more likely than non-trait associated SNPs to also be associated with changes in gene expression (eSNPs) (25). eSNPs describe the association between a single SNP with changes in gene expression whereas eQTLs are reflective of the association between a genetic locus and gene expression (26). Although association with changes in gene expression is not a direct proxy for function, this does help to better characterise systemic changes that are elicited in response to trait-associated genetic variation.

eQTLs are categorised on the basis of their proximity to the gene locus with which they are associated (Figure 1B). This distinction is important as it provides insight into the mechanisms by which an eQTL mediates an effect on gene expression. Cis-eQTLs, located close to their associated gene, are more likely to be acting locally than those located further away. Typically this distance is defined as being within 1 Mb of the associated transcription start site (27, 28). Outside of this threshold, eQTLs are said to be acting distally with associated genes in trans. Trans-eQTLs can be associated with genes located several megabases away including those on other chromosomes (29).

If an eQTL is cis-acting, this is likely to suggests a physical interaction between the eQTL and the associated gene. For example, a cis-eQTL sitting in an enhancer region may facilitate either an increased or decreased affinity for binding with a transcription factor (30, 31). Trans-eQTLs, on the other hand, associated with a distal gene, may influence transcription
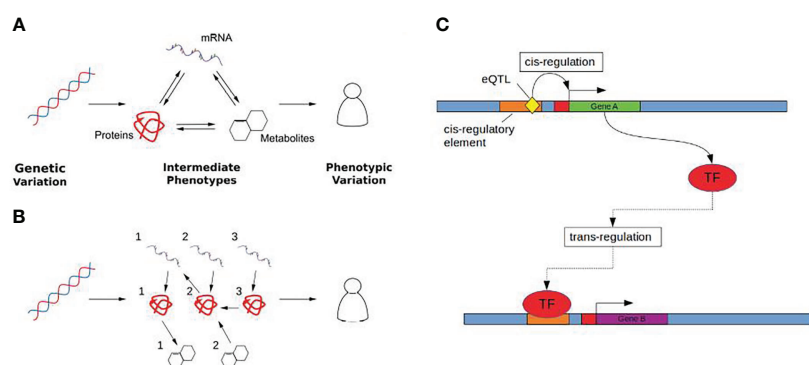


FIGURE 1
**(A)** Genetic variation (Left) influences complex traits (Right) through quantitative changes in intermediate phenotypes (Middle). Molecular interactions are shown as arrows, where the direction of the arrow indicates the direction of the flow of biological information. **(B)** Intermediate phenotypes can be modelled as biological networks using causal inference to uncover directed relationships between the molecular determinants that mediate the effect of genetic changes on complex traits. **(C)** Cis and trans gene regulation. Gene A (green) encodes a transcription factor (TF) which regulates the expression of gene B (purple). The eQTL (yellow), acts as a cis-eQTL for gene A by causing a change in the sequence of gene A's cis-regulatory element (orange) which may either increase or decrease the binding affinity of any corresponding TFs. The same eQTL is a trans-eQTL for Gene B as by changing the expression of the TF encoded by gene A, this in turn influences the expression of gene B.

indirectly through an intermediary gene product or working in conjunction with local cis-eQTLs (32). Most gene regulation takes place in cis, within regulatory regions and this is reflected by cis signals appearing more strongly than trans effects from eQTL mapping studies (33).

Genomes are consistent between cell types, but the way in which these genes are expressed varies drastically between tissues, and much of the regulation that mediates this disparity takes place at the transcriptome. Tissue gene expression profiles as measured by bulk RNA-seq are composed of distinct cell types, however cellular deconvolution methods can estimate the relative contributions of different cells types within the bulk sample. Methods that use single cell RNA-seq data perform most effectively as described in a recent benchmarking study (34).

Developments and reduced costs for RNA-seq methods has led to the establishment of large multi-tissue eQTL catalogues (Table 2). One of the most comprehensive projects to produce an atlas of transcriptome wide genetic effects has been conducted by the GTEx consortium, who have generated tissue specific eQTL data from an impressive number of post-mortem samples. In the latest release of data from GTEx v8 (35), the authors present both trans and cis associations from 49 different tissues. AES methods have also been applied using GTEx with the aim of improving the power of eQTL mapping studies. Zhabotynsky and colleagues show that AES both improves the power to detect eQTLs in GTEx tissues and the quantification of individual-specific genetic effects, while observing a similar levels of enrichment of GWAS hits within eQTL sets as seen using linear models (36).

## Causal inference in genetic epidemiology

Mendelian randomisation (MR) is a causal inference based approach that has been applied extensively in relation to SNP associated phenotypes (57). During meiosis, alleles are randomly seg- regated within chromosomes during gamete production. This independent assortment ensures that alleles are randomly distributed across a given population, much in the same way that treatments are allocated during randomised controlled trials, hence the "randomisation" in MR refers to the way in which alleles randomly segregate from parent to offspring (58).

MR uses an Instrumental Variable (IV) analysis framework to obtain causal relationships between biological traits (Figure 2A). For this methodology, the IV is used to infer a causal relationship between an exposure and an outcome variable. IV analysis requires the following assumptions: 1) The IV should be robustly associated with the exposure. 2) The IV should only be causal for the outcome through the exposure (58). 3) The IV should be independent from any confounding factors that are causal for the exposure or the outcome (59, 60). Given these assumptions, the IV acts as a proxy for the exposure to infer a directed relationship between the exposure and outcome, where the detection of a causal relationship between the instrument and outcome can be inferred as a causal relationship between exposure and outcome, due to the elimination of an alternative causal path. Given the IV assumptions and continuous traits, the average causal effect of the exposure on the outcome can be estimated by the ratio of the covariances of the IV and the exposure and outcome respectively (60).

The use of genetic variants as instruments in MR has become an important method for establishing causal relationships in biological systems where gene expression acts as an exposure. eQTLs have been shown to satisfy the IV assumptions through a robust association with gene expression, given the same eQTL is not also directly associated with the outcome (61). This also overcomes issues related to confounding as genetic variation is fixed at conception and is therefore highly unlikely to be confounded by the same causal factors influencing downstream phenotypes (62) outside of systemic variation in population structure. Issues can arise when an eQTL is also directly associated with the outcome, however this can be overcome through careful instrument selection and testing for pleiotropy (63).

It has been challenging to identify causal variants from GWAS results alone as a result of Linkage Disequilibrium

TABLE 2  Top 10 RNA-seq based eQTL studies as determined by sample size, obtained from eQTL catalogue (47).

| Study name | Cell type or tissues | Number of samples | Number of donors |
| --- | --- | --- | --- |
| GTEx (v8) (35) | 49 tissues | 15,178 | 838 |
| TwinsUK (48) | adipose, LCLs1, skin, blood | 1364 | 433 |
| Schmiedel (49) | 15 immune cell types | 1331 | 91 |
| Quach (50) | monocytes | 969 | 200 |
| CommonMind (51) | brain (DLPFC3) | 590 | 590 |
| ROSMAP (52) | brain (DLPFC3) | 576 | 576 |
| GENCORD (53) | LCLs1, fibroblasts, T cells | 560 | 195 |
| FUSION (54) | adipose, muscle | 559 | 302 |
| BLUEPRINT (55) | monocytes, neutrophils, CD4+ T cells | 554 | 197 |
| Nedelec (2016) (56) | macrophages | 493 | 168 |
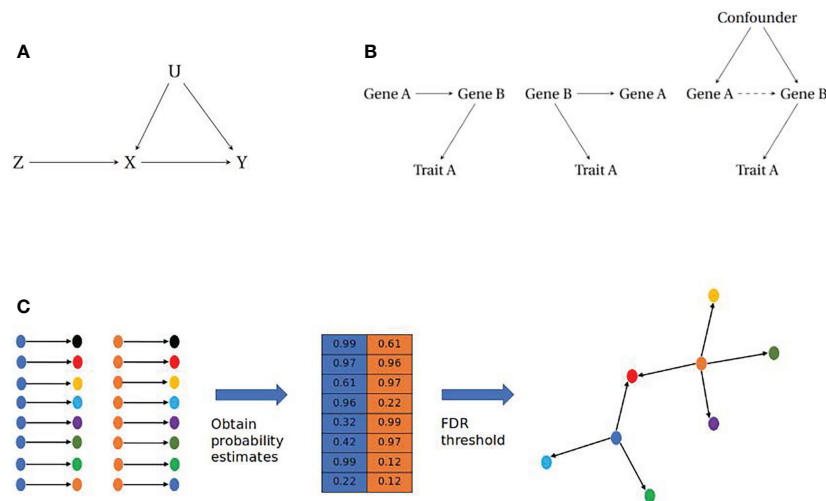
**FIGURE 2**
**(A)** Instrumental Variable paradigm. The instrumental variable (Z) is causally associated with the exposure (X) which in turn is causally associated with the outcome (Y). The IV will account for any confounding (U) that affects the exposure or outcome, assuming independence of U. **(B)** Causal modelling of pairwise gene-gene relationships. (Left) Simple causal model where Trait A is influenced by Gene A, through Gene B (Middle) Reactive model where Gene B influences both Gene A and Trait A, therefore any association between Gene A and Trait A is a non-causal relationship. (Right) Association between Genes A and B is a result of unobserved confounding, therefore there is no causal relationship between Gene A and Trait A. **(C)** Reconstructing gene networks from pairwise relationships. (Left) Prospective pairwise relationships between genes with a robust eQTL (blue and orange) and other genes within a dataset. (Middle) Causal inference approaches are employed to obtain a probability matrix for the likelihood of a causal relationship between gene pairs. (Right) A filtering step is imposed e.g. a False Discovery Rate (FDR) cut-off, which will return relationships that cross this threshold to be assembled as directed networks.

(LD), resulting in the observation of the non-random inheritance of alleles at a given loci with SNPs that are in LD (64). Therefore, if a true causal SNP for a trait is present and detectable at a given locus, the causal SNP and all other SNPs in LD will be identified as being associated with the trait in question, leading to an increase in type I error rate (65). In the instance of using eQTLs to obtain causal estimates between traits, as the eQTL is being used as an instrument it is not necessary that it is causal for the trait in question and eQTLs in LD will also be suitable.

# Reconstruction of causal gene networks

Jansen and Nap first proposed the integration of genomic information to identify changes in continuous molecular traits associated with the segregation of genotypes within a population in 2001 (66). What the authors originally describe as "genetical genomics", outlines a strategy to link genetic variation within a population to gene expression data, at the time obtained from microarray assays, and to other sources of expression data relating to proteins and metabolites. This has provided the foundation for modern day systems genetics, which allows for the integration of genetic and quantitative data with the ultimate aim of generating biological networks that can be linked to complex traits (20).

Most network based approaches to date have focused on correlation, through the development of co-expression networks using transcriptomic data (67). Co-expression networks were first proposed in the 1990s (68) and have been used to identify novel pathways in complex traits and disease as wide ranging as depression (69), muscular disease (70), and cardiovascular disease (71). They have also been used to identify clusters of genes that are linked to different phenotypic characteristics for conditions such as endometriosis (72). These methods are capable of reconstructing edges (connections between nodes) between co-expressed genes but are limited due to their inability to distinguish between different causal models (Figure 2B).

Pairwise gene-gene relationships are capable of providing a foundation for gene network reconstruction using sufficiently large transcriptomic datasets (73–77). MR based methods can be used to obtain probability estimates for causal relationships between genes when provided with robust genetic instruments. A method that facilities this approach is the tool Findr, which incorporates eQTLs within an MR framework to obtain the Bayesian posterior probability of a causal relationship between a pair of genes, using a combination of likelihood ratio tests to account for any unobserved confounding (77).

Bayesian networks are acyclic graphs that have been used for modelling gene networks as they allow for the incorporation of prior knowledge and are capable of resolving issues of conditional independence in data (73, 78–81). Bayesian networks are developed using frequency tables from discrete data, however in cases of continuous data such as transcriptomic datasets, posterior probabilities can be calculated from density functions (82). By obtaining posterior probabilities for pairwise relationships between genes with tools such as Findr, it is possible to reconstruct networks of genes (nodes) that are connected by posterior probabilities (edges) at a given thresh-old (83) (Figure 2C).

An issue encountered within Bayesian network analysis, is that as the number of network nodes increases so does the number of potential network edges. Given the high dimensional datasets commonly generated from next generation sequencing, standard Bayesian network methods are often computationally prohibitive (84). Novel methods to overcome the computational burden include the use of eQTL and transcriptomic data within a node ordering approach which prioritises given relationships, reducing the number of possible networks (83).

There is a wealth of data relating to the role of gene regulation, including available cis-regulatory elements (85) and transcription factor binding sites (86). The incorporation of these data allows for the construction of robust priors for Bayesian causal inference (83). eQTLs are also particularly well suited to filling this role and have been used to identify genes driving cardiovascular disease (71), type II diabetes (79) and Acute Myeloid Leukaemia (87) when combined with gene expression data.

A combination of different approaches can be used in the dissection of GWAS hits for complex disease, as demonstrated by Small and colleagues who were able to reconstruct networks of genes associated with SNPs linked to type II diabetes and mediated through the gene *KLF14* (88).

The researchers were able to show that cis-eQTLs for *KLF14* regulated a larger adipose specific gene network that was significantly enriched for metabolic pathways. This highlights the role of a network approach when combined with traditional genetic association and linkage studies. More recently, a 2022 study demonstrated how it was possible to use eQTLs to identify tissue specific clusters driven by "key driver genes" (89). Some of these key drivers were then validated using similar MR based methods as described in this mini-review.

## Conclusion

Over the course of this mini-review we have described how GWAS have been used to link genetic variants to different endocrine traits, including changes in hormone levels. Much of this variation is reflected at the transcriptome, indicated by the presence of cell type specific eQTLs. There have been many

thorough studies that have used network based approaches to understand the impact of genetic variation upon phenotypes, although without causal inference, it is challenging to identify the causal drivers of these networks. The use of eQTLs in MR has lead to identification of causal relationships between different traits, including molecular phenotypes. Through an extension of this pairwise approach to causal inference, we propose an systems genetics framework through which the reconstruction of causal gene networks is possible, with particular relevance to endocrinology.

## Author contributions

SB and TM both contributed to the conception and design of this manuscript. The manuscript was written by SB with TM providing supervision, oversight and review. All authors contributed to the article and approved the submitted version.

## Funding

## Acknowledgments

## 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.

# References

1. Ozaki K, Ohnishi Y, Iida A, Sekine A, Yamada R, Tsunoda T, et al. "Functional SNPs in the lymphotoxin-alpha gene that are associated with susceptibility to myocardial infarction". *Nat Genet* (2002) 32.4:650–4. doi: 10.1038/ng1047

2. Goodarzi MO. Genetics of common endocrine disease: The present and the future. *J Clin Endocrinol Metab* (2016) 101.3:787–94. doi: 10.1210/jc.2015-3640

3. Hunter DJ, Kraft P. Drinking from the fire hose–statistical issues in genomewide association studies. *New Engl J Med* (2007) 357.5:436–9. doi: 10.1056/NEJMp078120

4. Husebye ES, Perheentupa J, Rautemaa R, Kämpe O. Clinical manifestations and manage- ment of patients with autoimmune polyendocrine syndrome type I. *J Internal Med* (2009) 265.5:514–29. doi: 10.1111/j.1365-2796.2009.02090.x

5. d'Hennezel E, Dhuban KB, Torgerson T, Piccirillo C. The immunogenetics of immune dysregulation, polyendocrinopathy, enteropathy, X linked (IPEX) syndrome. *J Med Genet* (2012) 49.5:291–302. doi: 10.1136/jmedgenet-2012-100759

6. Porcu E, Rüeger S, Lepik KeQTLGen Consortium and BIOS Consortium, , Santoni FA, Reymond A. Mendelian randomization integrating GWAS and eQTL data reveals genetic determinants of complex and clinical traits. *Nat Commun* (2019) 10.1:1–12. doi: 10.1038/s41467-019-10936-0

7. Buniello A, MacArthur JAL, Cerezo M, Harris LW, Hayhurst J, Malangone C, et al. The NHGRI-EBI GWAS catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res* (2019) 47.D1:D1005–D1012. doi: 10.1093/nar/gky1120

8. Boyle EA, Li YI, Pritchard JK. An expanded view of complex traits: From polygenic to omnigenic. *Cell* (2017) 169.7:1177–86. doi: 10.1016/j.cell.2017.05.038

9. Bogardus C. Missing heritability and GWAS utility. *Obesity* (2009) 17.2:209–10. doi: 10.1038/oby.2008.613

10. Clarke AJ, Cooper DN. GWAS: heritability missing in action? *Eur J Hum Genet* (2010) 18.8:859–61.

11. Makowsky R, Pajewski NM, Klimentidis YC, Vazquez AI, Duarte CW, Allison DB, et al. Beyond missing heritability: Prediction of complex traits. *PloS Genet* (2011) 7.4:e1002051. doi: 10.1371/journal.pgen.1002051

12. Xue A, Wu Y, Zhu Z, Zhang F, Kemper KE, Zheng Z, et al. Genome-wide association analyses identify 143 risk variants and putative regulatory mechanisms for type 2 diabetes. *Nat Commun* (2018) 9.1:2941. doi: 10.1038/s41467-018-04951-w

13. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ. Finding the missing heritability of complex diseases. *Nature* (2009) 461.7265:747–53. doi: 10.1038/nature08494

14. Schadt EE. Molecular networks as sensors and drivers of common human diseases. *Nature* (2009) 461.7261:218–23. doi: 10.1038/nature08454

15. Nagrani R, Foraita R, Gianfagna F, Iacoviello L, Marild S, Michels N, et al. Common genetic variation in obesity, lipid transfer genes and risk of metabolic syndrome: Results from IDEFICS/I.Family study and meta-analysis. *Sci Rep* (2020) 10.1:7189. doi: 10.1038/s41598-020-64031-2

16. Humphries SE. Common variants for cardiovascular disease. *Circulation* (2017) 135.22:2102–5. doi: 10.1161/CIRCULATIONAHA.117.027798

17. olovieff N, Cotsapas C, Lee PH, Purcell SM, Smoller JW. Pleiotropy in complex traits: challenges and strategies. *Nat Rev Genet* (2013) 14.7:483–95. doi: 10.1038/nrg3461

18. Chesmore K, Bartlett J, Williams SM. The ubiquity of pleiotropy in human disease. *Hum Genet* (2018) 137.1:39–44. doi: 10.1007/s00439-017-1854-z

19. Watanabe K, Stringer S, Frei O, Umićević Mirkov M, de Leeuw C, Polderman TJC. A global overview of pleiotropy and genetic architecture in complex traits. *Nat Genet* (2019) 51.9:1339–48. doi: 10.1038/s41588-019-0481-0

20. Civelek M, Lusis AJ. Systems genetics approaches to understand complex traits. *Nat Rev Genet* (2014) 15.1:34–48. doi: 10.1038/nrg3575

21. Oakley RH, Cidlowski JA. The biology of the glucocorticoid receptor: New sig- naling mechanisms in health and disease. *J Allergy Clin Immunol* (2013) 132.5:1033–44. doi: 10.1016/j.jaci.2013.09.007

22. Sun W. A statistical framework for eQTL mapping using RNA-seq data. *Biometrics* (2012) 68.1:1–11. doi: 10.1111/j.1541-0420.2011.01654.x

23. Sun W, Hu Y. eQTL mapping using RNA-seq data. *Stat Biosci* (2013) 5.1:198–219. doi: 10.1007/s12561-012-9068-3

24. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet* (2009) 10.1:57–63. doi: 10.1038/nrg2484

25. Nicolae DL, Gamazon E, Zhang W, Duan S, Dolan ME, Cox NJ. Trait-associated SNPs are more likely to be eQTLs: Annotation to enhance discovery from GWAS. *PloS Genet* (2010) 6.4:e1000888. doi: 10.1371/journal.pgen.1000888

26. Cuomo ASE, Alvari G, Azodi CB, McCarthy DJ, Bonder MJ. Optimizing expression quantitative trait locus mapping workflows for single-cell studies. *Genome Biol* (2021) 22.1:188. doi: 10.1186/s13059-021-02407-x

27. Kirsten H, Al-Hasani H, Holdt L, Gross A, Beutner F, Krohn K. Dissecting the genetics of the human transcriptome identifies novel trait- related trans-eQTLs and corroborates the regulatory relevance of non-protein coding loci. *Hum Mol Genet* (2015) 24:4746–63:16. doi: 10.1093/hmg/ddv194

28. Fauman EB, Hyde C. An optimal variant to gene distance window derived from an empirical definition of cis and trans protein QTLs. *BMC Bioinf* (2022) 23.1:169. doi: 10.1186/s12859-022-04706-x

29. Joehanes R, Zhang X, Huan T, Yao C, Ying S-X, Tri Nguyen Q, et al. Integrated genome-wide analysis of expression quantitative trait loci aids interpretation of genomic association studies. *Genome Biol* (2017) 18.1:16. doi: 10.1186/s13059-016-1142-6

30. Cesar ASM, Regitano LCA, Reecy JM, Poleti MD, Oliveira PSN, de Oliveira GB. Identification of putative regulatory regions and transcription factors associated with intramuscular fat content traits. *BMC Genomics* (2018) 19:499. doi: 10.1186/s12864-018-4871-y

31. Brown CD, Mangravite LM, Engelhardt BE. Integrative modeling of eQTLs and cis-regulatory elements suggests mechanisms underlying cell type specificity of eQTLs. *PloS Genet* (2013) 9.8:e1003649. doi: 10.1371/journal.pgen.1003649

32. Albert FW, Bloom JS, Siegel J, Day L, Kruglyak L. Genetics of trans-regulatory variation in gene expression. *eLife* (2018) 7:e35471. doi: 10.7554/eLife.35471

33. Dixon AL, Liang L, Moffatt MF, Chen W, Heath S, Wong KCC. A genome-wide association study of global gene expression. *Nat Genet* (2007) 39.10:1202–7. doi: 10.1038/ng2109

34. Avila Cobos F, Alquicira-Hernandez JoséVerifytat, Powell JE, Mestdagh P, De Preter K. Benchmarking of cell type deconvolution pipelines for transcriptomics data. *Nat Com- munications* (2020) 11.1:5650. doi: 10.1038/s41467-020-19015-1

35. Consortium. The GTEx consortium atlas of genetic regulatory effects across human tissues. *Science* (2020) 369.6509:1318–30. doi: 10.1126/science.aaz1776

36. Zhabotynsky V, Huang L, Little P, Hu Y-J, Pardo-Manuel de Villena F, Zou F. eQTL mapping using allele-specific count data is computationally feasible, powerful, and provides individual-specific estimates of genetic effects. *PloS Genet* (2022) 18.3:e1010076. doi: 10.1371/journal.pgen.1010076

37. Zhu Z, Guo Y, Shi H, Liu C-L, Panganiban RA, Chung W. Shared genetic and experimental links between obesity-related traits and asthma subtypes in UK biobank. *J Allergy Clin Immunol* (2020) 145.2:537–49. doi: 10.1016/j.jaci.2019.09.035

38. Barton AR, Sherman MA, Mukamel RE, Loh P-R. Whole-exome imputation within UK biobank powers rare coding variant association and fine-mapping analyses. *Nat Genet* (2021) 53.8:1260–9. doi: 10.1038/s41588-021-00892-1

39. Ruth KS, Day FR, Tyrrell J, Thompson DJ, Wood AR, Mahajan A, et al. Using human genetics to understand the disease impacts of testosterone in men and women. *Nat Med* (2020) 26.2:252–8. doi: 10.1038/s41591-020-0751-5

40. Liang X, Cheng SQ, Ye J, Chu XM, Wen Y, Liu L. Evaluating the genetic effects of sex hormone traits on the development of mental traits: a polygenic score analysis and gene-environment-wide interaction study in UK biobank cohort. *Mol Brain* (2021) 14.1:3. doi: 10.1186/s13041-020-00718-x

41. Backman JD, Li AH, Marcketta A, Sun D, Mbatchou J, Kessler MD. Exome sequencing and analysis of 454,787 UK biobank participants. *Nature* (2021) 599.7886:628–34. doi: 10.1038/s41586-021-04103-z

42. Martin S, Cule M, Basty N, Tyrrell J, Beaumont RN, Wood AR. Genetic evidence for different adiposity phenotypes and their opposing influences on ectopic fat and risk of cardiometabolic disease. *Diabetes* (2021) 70.8:1843–56. doi: 10.2337/db21-0129

43. Huang LO, Rauch A, Mazzaferro E, Preuss M, Carobbio S, Bayrak CS, et al. Genome-wide discovery of genetic loci that uncouple excess adiposity from its comorbidities. *Nat Metab* (2021) 3.2:228–43. doi: 10.1038/s42255-021-00346-2

44. Sinnott-Armstrong N, Tanigawa Y, Amar D, Mars N, Benner C, Aguirre M. Genetics of 35 blood and urine biomarkers in the UK biobank. *Nat Genet* (2021) 53.2:185–94. doi: 10.1038/s41588-020-00757-z

45. Wang X, Jia J, Huang T. Shared genetic architecture and casual relationship between leptin levels and type 2 diabetes: large-scale cross-trait meta-analysis and mendelian randomization analysis. *BMJ Open Diabetes Res Care* (2020) 8.1: e001140. doi: 10.1136/bmjdrc-2019-001140

46. Chen J, Spracklen CN, Marenne G, Varshney A, Corbin LJ, Luan J. The trans-ancestral genomic architecture of glycemic traits. *Nat Genet* (2021) 53.6:840–60. doi: 10.1038/s41588-021-00852-9

47. Kerimov N, Hayhurst JD, Peikova K, Manning JR, Walter P, Kolberg L. A compendium of uniformly processed human gene expression and splicing quantitative trait loci. *Nat Genet* (2021) 53:1290–9. doi: 10.1038/s41588-021-00924-w

48. Buil A, Brown AA, Lappalainen T, Viñuela A, Davies MN, Zheng H-F. Gene-gene and gene-environment interactions detected by transcriptome sequence analysis in twins. *Nat Genet* (2015) 47.1:88–91. doi: 10.1038/ng.3162

49. Schmiedel BJ, Singh D, Madrigal A, Valdovino-Gonzalez AG, White BM, Zapardiel-Gonzalo J. Impact of genetic polymorphisms on human immune cell gene expression. *Cell* (2018) 175.6:1701–15. doi: 10.1016/j.cell.2018.10.022

50. Quach H, Rotival M, Pothlichet J, Loh Y-HE, Dannemann M, Zidane N. Genetic adaptation and neandertal admixture shaped the immune system of human populations. *Cell* (2016) 167.3:643–56. doi: 10.1016/j.cell.2016.09.024

51. Hoffman GE, Bendl J, Voloudakis G, Montgomery KS, Sloofman L, Wang Y-C, et al. CommonMind consortium provides transcriptomic and epigenomic data for schizophrenia and bipolar disorder. *Sci Data* (2019) 6.1:180. doi: 10.1038/s41597-019-0183-6

52. Ng B, White CC, Klein H-U, Sieberts SK, McCabe C, Patrick E. An xQTL map integrates the genetic architecture of the human brain's transcriptome and epigenome. *Nat Neurosci* (2017) 20.10:1418–26. doi: 10.1038/nn.4632

53. Gutierrez-Arcelus M, Lappalainen T, Montgomery SB, Buil A, Ongen H, Yurovsky A. Passive and active DNA methylation and the interplay with genetic variation in gene regulation. *eLife* (2013) 2:e00523. doi: 10.7554/eLife.00523

54. Taylor DL, Jackson AU, Narisu N, Hemani G, Erdos MR, Chines PS, et al. Integrative analysis of gene expression, DNA methylation, physiological traits, and genetic variation in human skeletal muscle. *Proc Natl Acad Sci* (2019) 116.22:10883–8. doi: 10.1073/pnas.1814263116

55. Chen L, Ge B, Casale FP, Vasquez L, Kwan T, Garrido-Martín D. Genetic drivers of epigenetic and transcriptional variation in human immune cells. *Cell* (2016) 167.5:1398–414. doi: 10.1016/j.cell.2016.10.026

56. Nédélec Y, Sanz J, Baharian G, Szpiech ZA, Pacis A, Dumaine A. Genetic ancestry and natural selection drive population differences in immune responses to pathogens. *Cell* (2016) 167.3:657–69. doi: 10.1016/j.cell.2016.09.025

57. Davey Smith G, Hemani G. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. *Hum Mol Genet* (2014) 23.R1:R89–98. doi: 10.1093/hmg/ddu328

58. Hinke Sv, Davey Smith G, Lawlor DA, Propper C, Windmeijer F. Genetic markers as instrumental variables. *J Health Economics* (2016) 45:131–48. doi: 10.1016/j.jhealeco.2015.10.007

59. Glymour MM, Tchetgen Tchetgen EJ, Robins JM. Credible mendelian randomization studies: Approaches for evaluating the instrumental variable assumptions. *Am J Epidemiol* (2012) 175.4:332–9. doi: 10.1093/aje/kwr323

60. Lousdal ML. An introduction to instrumental variable assumptions, validation and estimation. *Emerging Themes Epidemiol* (2018) 15. doi: 10.1186/s12982-018-0069-7

61. Richardson TG, Hemani G, Gaunt TR, Relton CL, Davey Smith G. A transcriptome-wide mendelian randomization study to uncover tissue-dependent regulatory mechanisms across the human phenome. *Nat Commun* 11.1. doi: 10.1038/s41467-019-13921-9

62. Neto EC, Keller MP, Attie AD, Yandell BS. CAUSAL GRAPHICAL MODELS IN SYSTEMS GENETICS: A UNIFIED FRAMEWORK FOR JOINT INFERENCE OF CAUSAL NETWORK AND GENETIC ARCHITECTURE FOR CORRELATED PHENOTYPES. *Ann Appl Stat* (2010) 4.1:320–39. doi: 10.1214/09-aoas288

63. Neumeyer S, Hemani G, Zeggini E. Strengthening causal inference for complex disease using molecular quantitative trait loci. *Trends Mol Med* (2019) 26.2:232–41. doi: 10.1016/j.molmed.2019.10.004

64. Slatkin M. Linkage disequilibrium — understanding the evolutionary past and mapping the medical future. *Nat Rev Genet* (2008) 9.6:477–85. doi: 10.1038/nrg2361

65. Joiret M, Mahachie John JM, Gusareva ES, Van Steen K. Confounding of linkage disequilibrium patterns in large scale DNA based gene-gene interaction studies. *Bio-Data Min* (2019) 12.1:11. doi: 10.1186/s13040-019-0199-7

66. Jansen RC, Nap J-P. Genetical genomics: the added value from segregation. *Trends Genet* (2001) 17.7:388–91. doi: 10.1016/S0168-9525(01)02310-1

67. Dam Sv, Võsa U, Graaf Avd, Franke L, Magalhães JoãoPde. Gene co-expression analysis for functional classification and gene–disease predictions. *Briefings Bioinf* (2018) 19.4:575–92. doi: 10.1093/bib/bbw139

68. Butte AJ, Kohane IS. Unsupervised knowledge discovery in medical databases using relevance networks. *Proc AMIA Symposium* (1999) 1999:711–5.

69. Gerring ZF, Gamazon ER, Derks EMConsortium, for the Major Depressive Disorder Working Group of the Psychiatric Genomics. A gene co-expression network-based analysis of multiple brain tissues reveals novel genes and molecular pathways underlying major depression. *PloS Genet* (2019) 15.7:e1008245. doi: 10.1371/journal.pgen.1008245

70. Mukund K, Subramaniam S. Co-Expression network approach reveals functional similarities among diseases affecting human skeletal muscle. *Front Physiol* (2017) 8. doi: 10.3389/fphys.2017.00980

71. Talukdar HA, Asl HF, Jain RK, Ermel R, Ruusalepp A, Franzén O, et al. Cross-tissue regulatory gene networks in coronary artery disease. *Cell Syst* (2016) 2.3:196–208. doi: 10.1016/j.cels.2016.02.002

72. Bakhtiarizadeh MR, Hosseinpour B, Shahhoseini M, Korte A, Gifani P. Weighted gene Co-expression network analysis of endometriosis and identification of functional modules associated with its main hallmarks. *Front Genet* (2018) 9. doi: 10.3389/fgene.2018.00453

73. Schadt EE, Lamb J, Yang X, Zhu J, Edwards S, GuhaThakurta D, et al. An integrative genomics approach to infer causal associations between gene expression and disease. *Nat Genet* (2005) 37:7. doi: 10.1038/ng1589

74. Chen LS, Emmert-Streib F, Storey JD. Harnessing naturally randomized transcription to infer regulatory relationships among genes. *Genome Biol* (2007) 8.10:R219. doi: 10.1186/gb-2007-8-10-r219

75. Millstein J, Zhang B, Zhu J, Schadt EE. Disentangling molecular relationships with a causal inference test. *BMC Genet* (2009) 10.1:23. doi: 10.1186/1471-2156-10-23

76. Meinshausen N, Hauser A, Mooij JM, Peters J, Versteeg P, Bühlmann P. Methods for causal inference from gene perturbation experiments and validation. *Proc Natl Acad Sci* (2016) 113.27:7361–8. doi: 10.1073/pnas.1510493113

77. Wang L, Michoel T. Efficient and accurate causal inference with hidden confounders from genome-transcriptome variation data. *PloS Comput Biol* (2017) 13.8:e1005703. doi: 10.1371/journal.pcbi.1005703

78. Friedman N, Linial M, Nachman I, Pe'er D. Using Bayesian networks to analyze expression data. *J Comput Biol* (2000) 7.3-4:601–20. doi: 10.1089/106652700750050961

79. Schadt EE, Molony C, Chudin E, Hao K, Yang X, Lum PY, et al. Mapping the genetic architecture of gene expression in human liver. *PloS Biol* (2008) 6.5:e107. doi: 10.1371/journal.pbio.0060107

80. Zhang B, Gaiteri C, Bodea L-G, Wang Z, McElwee J, Podtelezhnikov AA, et al. Integrated systems approach identifies genetic nodes and networks in late-onset alzheimer's disease. *Cell* (2013) 153.3:707–20. doi: 10.1016/j.cell.2013.03.030

81. Beckmann ND, Lin W-Y, Wang M, Cohain AT, Charney AW, Wang P. Multiscale causal networks identify VGF as a key regulator of alzheimer's disease. *Nat Commun* (2020) 11.1:3942. doi: 10.1038/s41467-020-17405-z

82. Pearl J. *Causality*. Cambridge: Cambridge University Press (2009).

83. Wang L, Audenaert P, Michoel T. High-dimensional Bayesian network inference from systems genetics data using genetic node ordering. *Front Genet* (2019) 10. doi: 10.3389/fgene.2019.01196

84. Wilkinson DJ. Bayesian Methods in bioinformatics and computational systems biology. *Briefings Bioinf* (2007) 8.2:109–16. doi: 10.1093/bib/bbm007

85. Wittkopp PJ, Kalay G. Cis -regulatory elements: molecular mechanisms and evolutionary processes underlying divergence. *Nat Rev Genet* (2012) 13.1:59–69. doi: 10.1038/nrg3095

86. Prestridge DS. Predicting pol II promoter sequences using transcription factor binding sites. *J Mol Biol* (1995) 249.5:923–32. doi: 10.1006/jmbi.1995.0349

87. Agrahari R, Foroushani A, Docking TR, Chang L, Duns G, Hudoba M, et al. Applications of Bayesian network models in predicting types of hematological malignancies. *Sci Rep* (2018) 8.1:6951. doi: 10.1038/s41598-018-24758-5

88. Small KS, Todorčević M, Civelek M, El-Sayed Moustafa JS, Wang X, Simon MM, et al. Regulatory variants at KLF14 influence type 2 diabetes risk *via* a female-specific effect on adipocyte size and body composition. *Nat Genet* (2018) 50.4:572–80. doi: 10.1038/s41588-018-0088-x

89. Koplev S, Seldin M, Sukhavasi K, Ermel R, Pang S, Zeng L. A mechanistic framework for cardiometabolic and coronary artery diseases. *Nat Cardiovasc Res* (2022) 1.1:85–100. doi: 10.1038/s44161-021-00009-1