



# Identifying Multi-Omics Causers and Causal Pathways for Complex Traits

Huaizhen Qin<sup>1,2\*</sup>, Tianhua Niu<sup>2,3</sup> and Jinying Zhao<sup>1</sup>

<sup>1</sup> Department of Epidemiology, College of Public Health and Health Professions and College of Medicine, University of Florida, Gainesville, FL, United States, <sup>2</sup> Department of Global Biostatistics and Data Science, Tulane University, New Orleans, LA, United States, <sup>3</sup> Department of Biochemistry and Molecular Biology, Tulane University School Medicine, New Orleans, LA, United States

## OPEN ACCESS

### Edited by:

Momiao Xiong,  
University of Texas Health Science  
Center, United States

### Reviewed by:

Christine W. Duarte,  
University of Alabama at Birmingham,  
United States  
Guimin Gao,  
University of Chicago, United States

### \*Correspondence:

Huaizhen Qin  
hqin@ufl.edu

### Specialty section:

This article was submitted to  
Statistical Genetics and Methodology,  
a section of the journal  
Frontiers in Genetics

**Received:** 17 April 2018

**Accepted:** 30 January 2019

**Published:** 21 February 2019

### Citation:

Qin H, Niu T and Zhao J (2019)  
Identifying Multi-Omics Causers and  
Causal Pathways for Complex Traits.  
*Front. Genet.* 10:110.  
doi: 10.3389/fgene.2019.00110

The central dogma of molecular biology delineates a unidirectional causal flow, i.e., DNA → RNA → protein → trait. Genome-wide association studies, next-generation sequencing association studies, and their meta-analyses have successfully identified ~12,000 susceptibility genetic variants that are associated with a broad array of human physiological traits. However, such conventional association studies ignore the mediate causers (i.e., RNA, protein) and the unidirectional causal pathway. Such studies may not be ideally powerful; and the genetic variants identified may not necessarily be genuine causal variants. In this article, we model the central dogma by a mediate causal model and analytically prove that the more remote an omics level is from a physiological trait, the smaller the magnitude of their correlation is. Under both random and extreme sampling schemes, we numerically demonstrate that the proteome-trait correlation test is more powerful than the transcriptome-trait correlation test, which in turn is more powerful than the genotype-trait association test. In conclusion, integrating RNA and protein expressions with DNA data and causal inference are necessary to gain a full understanding of how genetic causal variants contribute to phenotype variations.

**Keywords:** associations, causations, transcriptomics, proteomics, data integration, systems biology

## INTRODUCTION

The central dogma of molecular biology, as first proposed by Francis Crick (Crick, 1958, 1970), describes the transfer of sequence information during DNA replication, transcription into RNA and translation into amino-acid chains forming proteins. There are only ~23,500 predicted protein-coding genes in humans. Such genes constitute only ~2% of human DNA sequence. Genetic studies of thousands of single-gene disorders have revealed a large set of mutations in protein-coding regions, which appears to support the central dogma that the major output of the genome is protein (Plomin and Davis, 2009). Advanced multi-omics technologies have led to generation of genome-scale data sets at DNA, RNA, and protein levels. The multi-level data sources are illustrated in the context of the central dogma in **Figure 1**. At DNA level, genomic data uncover the information stored in the genomes of organisms. Variations at DNA level in populations include single nucleotide polymorphisms (SNPs), copy number variations (CNVs), and structural variations (SVs) (Koyutürk, 2010).

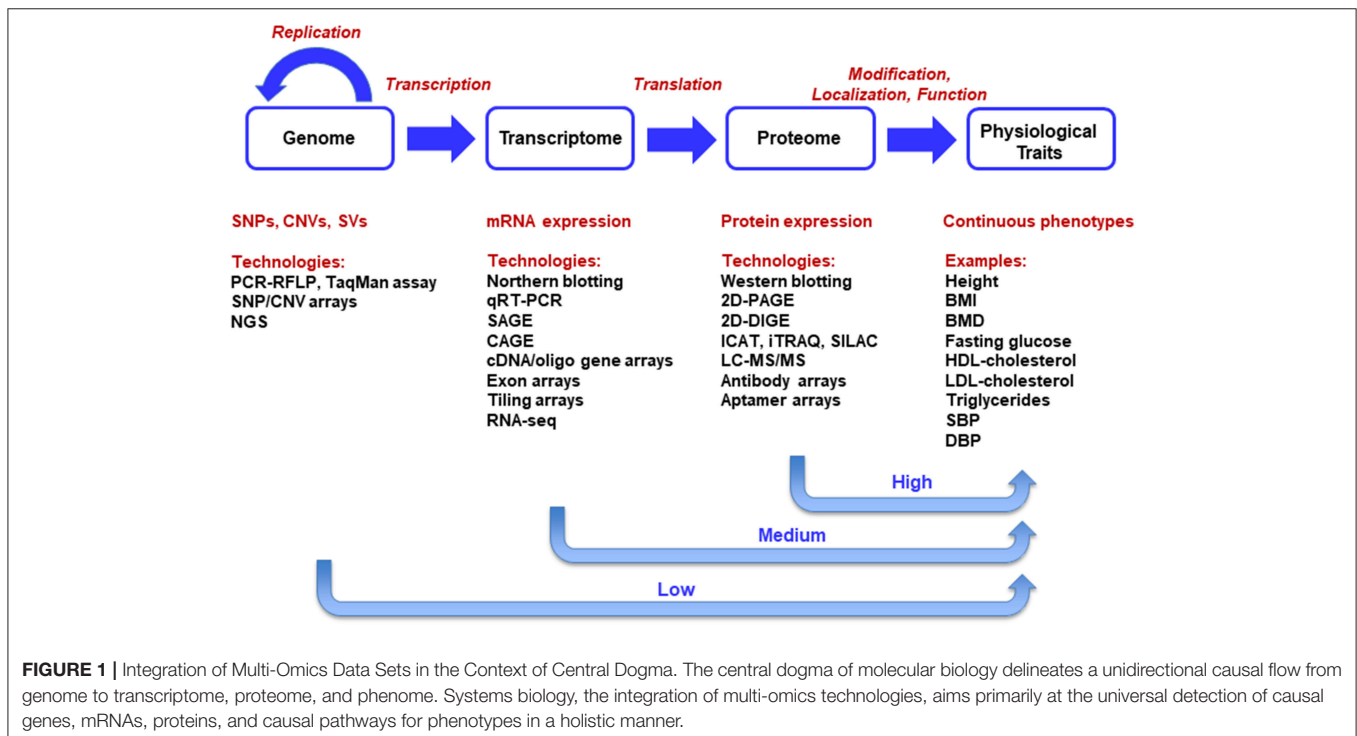
Identification of genetic causal variants for complex traits poses dramatically greater challenges than that for Mendelian traits, due to low penetrance, variable expressivity and pleiotropy, epistasis, and locus heterogeneity (Nadeau, 2001; Glazier et al., 2002). Klein et al. (2005) pioneered a genome-wide association (GWA) study and identified a SNP in *CFH* gene on age-related macular

degeneration. GWA studies have proven successful to identify susceptibility genetic variants for a range of complex traits (Christensen and Murray, 2007; Stranger et al., 2011; Evangelou and Ioannidis, 2013; Manolio, 2013). To date, at least 11,912 trait-associated SNPs from 1,751 curated publications have been reported at NHGRI Catalog of Published GWA studies (<http://www.genome.gov/gwastudies/>) (Welter et al., 2014). DNA variants for a plethora of physiological traits have been revealed by GWA studies, e.g., height (Weedon et al., 2008), body mass index (Monda et al., 2013), bone mineral density (Estrada et al., 2012), lipid levels (Willer et al., 2008), and hypertension (Kato et al., 2011). In past decade, researchers have also extended genetic association studies to rare genetic variants for a range of complex diseases. Typical examples include extreme lipoprotein cholesterol levels (Cohen et al., 2004; Haase et al., 2012), obesity (Ahituv et al., 2007; Meyre et al., 2009; Coassin et al., 2010), type 1 diabetes (Nejentsev et al., 2009), bone mineral density (Kung et al., 2010; Duncan et al., 2011). However, for a complex trait, the identified genetic variants only account for a small portion of phenotypic variation (Ruiz-Narváez, 2011).

A majority (88%) of trait-associated SNPs were found to reside in intergenic or intronic regions (Hindorf et al., 2009), suggesting that non-coding regions of the genome are responsible for most of the disease risk (Kung et al., 2010; Duncan et al., 2011) and some of this risk is likely to act through gene regulation (Bossé, 2013). Based on the central dogma of molecular biology, it is mechanistically very well understood how a gene get transcribed, how an mRNA get processed and sequentially translated into amino acid chains at the ribosome and subsequently fold into functional proteins. Recently, remarkable

advances of RNA-seq and mass spectrometry technologies have rapidly improved global identification, quantification, and analysis of transcriptome and proteome in same biological samples. Correlations between mRNA and protein abundances turn to be much stronger than that between genotype and trait. In bacteria and eukaryotes, they often show a squared Pearson correlation coefficient of  $\sim 0.40$ , which implies that  $\sim 40\%$  of the variation in protein abundance can be explained by knowing mRNA abundances. Emerging evidence shows that many regulatory mechanisms occur after mRNAs are made, and proteins exhibit a larger dynamic range of concentrations than do transcripts (Jacobs et al., 2005; Vogel et al., 2010; González-Porta et al., 2013; Schwanhäusser et al., 2013). A combination of post-transcriptional, translational and degradative regulation, acting through miRNAs (Mukherji et al., 2011) or other mechanisms to fine-tune protein abundances to their preferred levels. For example, miRNAs have been found to fine-regulate protein expression levels, rather than to cause large expression changes (Baek et al., 2008; Selbach et al., 2008). The combination of association studies on RNA and protein expressions allows systematic identifications of expression quantitative trait loci (eQTLs) and protein QTLs (pQTLs).

Single-platform studies, although popular, often neglect significant amount of genomic information. Under the central dogma, the genetic variant is most remote from the physiological trait; and the transcriptional and translational processes can dramatically attenuate the genetic effect of the genetic variant. Even if the sample size reaches several tens of thousands, the power is very limited to detect common genetic variants of modest effect sizes (Galvan et al., 2010). For example, Panagiotou

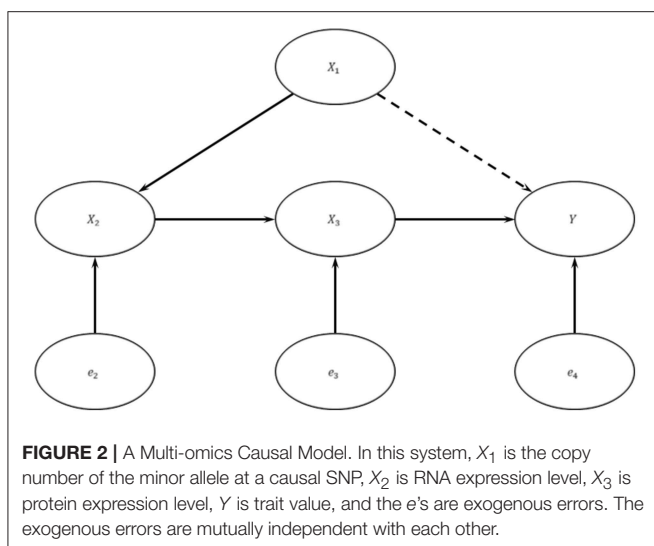


et al. (Panagiotou et al., 2013) depicted the relationship between the sample size and the number of trait-associated loci of genome-wide significance ( $P < 5 \times 10^{-8}$ ) in GWA studies of height, lipid levels, and blood pressure. Conventional genetic association studies ignore the mediate causers and the unidirectional causal pathways; and the genetic variants identified are not necessarily genuine causal variants, but are only in close physical proximities to them (Kingsley, 2011). Therefore, it is necessary to integrate multi-omics data and model a unidirectional causal graph. In this article, through extensive analytical explorations under the central dogma, we demonstrate that proteome data have the greatest potential to enhance our understanding of physiological traits, because protein is a direct causer for physiological trait and has a stronger correlation with trait than either genotype or transcriptome. We provide power and sample size analyses under extreme phenotype sampling (EPS) and random sampling schemes. EPS has been widely employed to detect genetic causal variants for complex diseases (Lander and Botstein, 1989; Abecasis et al., 2001; Xiong et al., 2002). Herein, we extend this sampling strategy to a multi-omics setting. The results would be helpful to design cost-effective multi-omics studies as well as to develop novel multi-stage causal association inference methods.

## METHODS

### Multi-Omics Causal Model (MCM)

A mediate causal graph may be suitable to model the unidirectional causal flow delineated by the central dogma of molecular biology. Assume the genuine data generating model is as depicted in **Figure 2**:  $\{Y = X_3\beta_3 + e_4, X_3 = X_2\beta_2 + e_3, X_2 = X_1\beta_1 + e_2\}$ , where  $X_1$  is the genotypic score (copy number of the minor allele) at a causal SNP,  $X_2$  is RNA expression,  $X_3$  is protein (PRT) expression, and  $Y$  is trait value of interest. Let exogenous errors  $e_2$ ,  $e_3$ , and  $e_4$  be independent.



Let  $\sigma^2$  denote the variance of a variable. Let  $r_1^2 = \beta_1^2 \sigma_{X_1}^2 / \sigma_{X_2}^2$ ,  $r_2^2 = \beta_2^2 \sigma_{X_2}^2 / \sigma_{X_3}^2$ , and  $r_3^2 = \beta_3^2 \sigma_{X_3}^2 / \sigma_Y^2$ . Then the heritability of the causal SNP is  $h^2 = r_1^2 r_2^2 r_3^2$ . In the numerical explorations, the minor allele frequency (MAF) of the causal SNP was fixed at  $p = 0.25$ ,  $r_1^2$  ranged from 0 to 15% (Schadt et al., 2003; Morley et al., 2004; Dimas et al., 2009; Webster et al., 2009; Bryois et al., 2014),  $r_2^2$  ranged from 0 to 50% (Guo et al., 2008; de Sousa Abreu et al., 2009; Lundberg et al., 2010; Vogel et al., 2010). Protein studies are based on extreme phenotype sampling (EPS) with certain truncation levels  $\alpha$  (e.g.,  $\alpha = 0.1 \sim 0.2$ ). In other words, an identical number of samples are drawn from top and bottom  $100\alpha\%$  tails of trait distribution, respectively. Reported fold changes between two tails of protein intensity ranged from 1.1 to 5 (Fields, 2001; Selbach et al., 2008; Cairns et al., 2009; Qiu et al., 2012). Under the MCM and EPS with a truncation level  $\alpha = 0.2$ , a 5-fold change in protein intensity is equivalent to  $r_3^2 = 34\%$  (**Appendix A**). Hence, we varied  $r_3^2$  from 0 to 34%, and varied the SNP heritability level  $h^2$  from 0.0% to 2.5% for numerical explorations.

### Working Models and Null Hypotheses

Under the MCM, the genotype-phenotype association test can be investigated under a simple additive genetic model (AGM), see **Appendix A**. To be specific, we rewrite the MCM as

$$\text{SNP: } Y = X_1\beta_1\beta_2\beta_3 + (\beta_2\beta_3e_2 + \beta_3e_3 + e_4). \quad (1)$$

Under this AGM working model, we test for  $H_{0,\text{SNP}}: \beta_1\beta_2\beta_3 = 0$ . Meanwhile, the tests for the associations between a mediate causal variable and the phenotype can be investigated under respective single mediate causal variable models (SMCVMs). To test for RNA-phenotype association, we rewrite the MCM as

$$\text{RNA: } \begin{cases} Y = X_2\beta_2\beta_3 + (\beta_3e_3 + e_4), \\ X_2 = X_1\beta_1 + e_2. \end{cases} \quad (2)$$

Under this SMCVM model, we test for  $H_{0,\text{RNA}}: \beta_2\beta_3 = 0$ . To test for protein-phenotype association, we rewrite the MCM as

$$\text{PRT: } \begin{cases} Y = X_3\beta_3 + e_4, \\ X_3 = X_1\beta_1\beta_2 + (e_2\beta_2 + e_3). \end{cases} \quad (3)$$

Under this working model, we test for  $H_{0,\text{PRT}}: \beta_3 = 0$ . The PRT and RNA models share one common indirect causal variable (SNP genotype  $X_1$ ) and have respective mediate causal variables. In the PRT model, protein expression  $X_3$  is taken as the mediate causal variable; and in the RNA model, RNA expression  $X_2$  is taken as the mediate causal variable.

### Sampling Designs and Association Tests

Under simple random sampling (SRS), the classical  $t$  test for slope is suitable for the association tests. Due to prohibitive experimental costs, however, extreme phenotype sampling (EPS) is often adopted for cost-effective multi-omics studies. In protein studies, for example, EPS is often utilized to reduce experimental costs and maintain statistical power of the two-sample  $t$ -tests. In EPS protein studies, fold changes in intensity of promising proteins are often reported. From major journals, however, we

have not identified protein studies that reported correlation coefficient between promising proteins and phenotypes. In **Appendix A**, we establish theoretical foundation to numerically evaluate powers of the SRS and EPS based  $t$ -tests for identifying the three kind of associations. In addition, we establish a method that converts a given fold change to corresponding correlation coefficient under the MCM.

## Software Package

Under the MCM, a freely available R package is developed for implementing numerical computations and graphical illustrations (<https://github.com/HuaizhenQin/MCM/>). It consists of R functions for multi-level power analyses and sample size evaluations under SRS and EPS. The function `MCVM.Power(.)` compute powers whereas the function `MCVM.n(.)` determines sample sizes to reach certain powers of level-specific tests under each sampling design. Both of the functions call corresponding functions for computing the non-centrality parameters. All the functions adopt fast and stable numerical algorithms without using any resampling techniques. The package is particularly useful for designing multi-omics studies and analyzing large-scale multi-omics data from such studies.

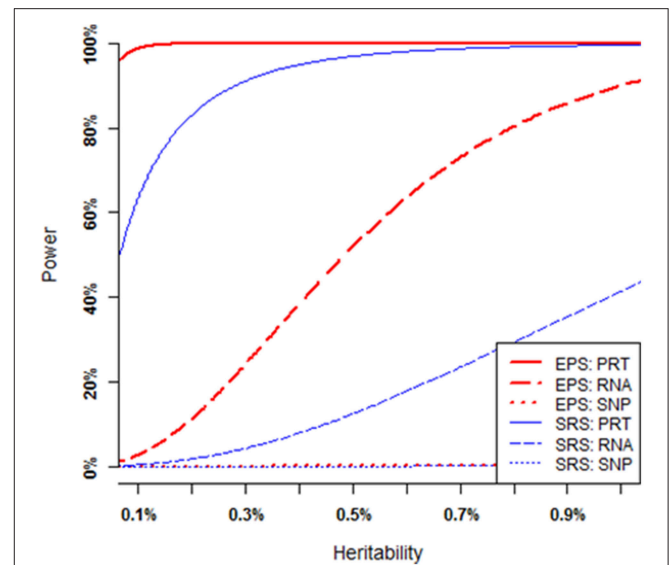
## RESULTS

In numerical power and sample size analyses, we use the nominal significance level of  $2.5e-6$  to depict relative power and sample size patterns of the widely used  $t$ -test under different sampling schemes. This nominal significance level is suggested in Goldstein et al. (2013) and often used for genome-wide gene-based association tests, because there are approximately 20,000 protein-coding genes in the human genome (Dunham, 2018).

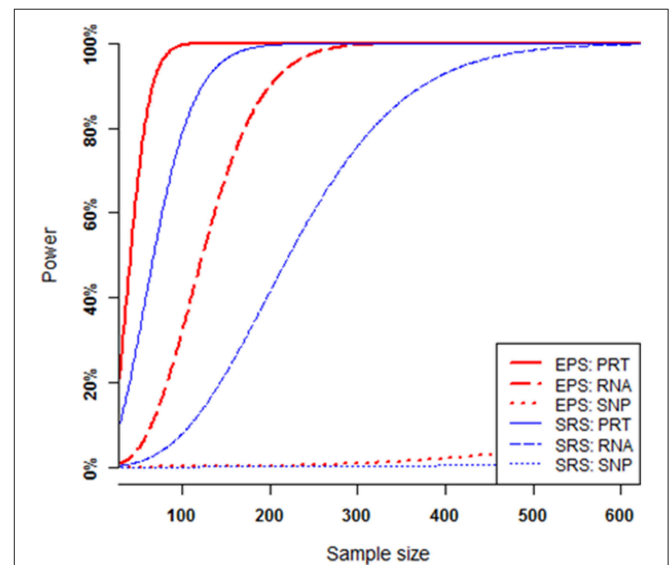
### Numerical Power Comparisons

As illustrated by **Figure 3**, the power of each strategy increases when the heritability of causal SNP increases (i.e., all the mediate correlations increase), provided that all other parameters (i.e., MAF of the causal SNP, nominal significance level, and sample size) are fixed (see **Appendix B** for a theoretical proof). Under SRS, the PRT test (blue solid curve) appears strikingly more powerful than the RNA test (blue dashed curve), and the RNA test appears strikingly more powerful than the SNP test (blue dotted line). The SRS-based single SNP test has little power to detect a genuine genotype-phenotype association over a large range of heritability levels. Under EPS, the PRT test (red solid curve) appears strikingly more powerful than the RNA test (red dashed curve), and the RNA test appears strikingly more powerful than the SNP test (red dotted curve). The EPS based single SNP test still has little power to detect the genuine genotype-phenotype association over the large range of heritability levels.

The power of each test increases when the sample size increases, provided that all other parameters (i.e., all the mediate correlations and hence the SNP's heritability, minor allele frequency at the causal SNP, and nominal significance level) are fixed (**Figure 4**). Under SRS, the PRT test appears strikingly more powerful than the RNA test, and RNA test appears strikingly

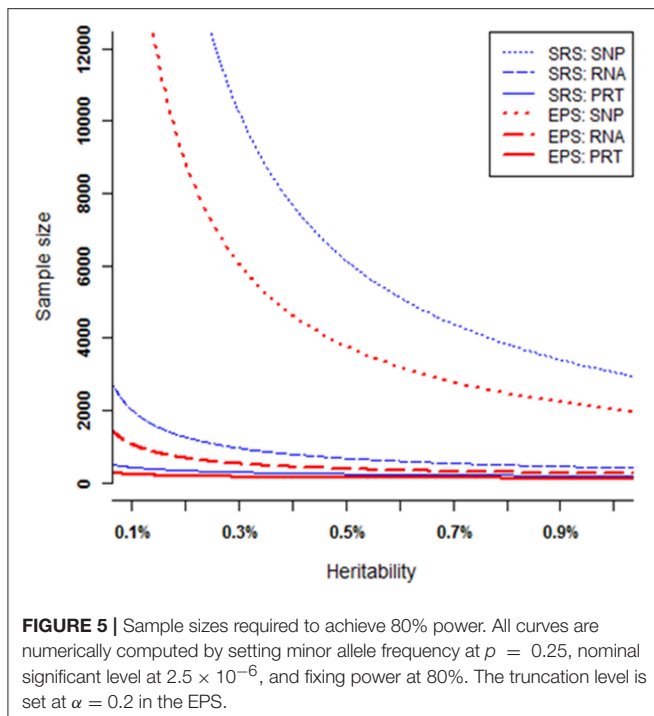


**FIGURE 3** | Power comparison at various heritability levels. Each overall heritability level of the causal SNP in the MCM is determined by a set of mediate effect sizes. All curves are numerically computed by setting minor allele frequency at  $p = 0.25$ , nominal significant level at  $2.5 \times 10^{-6}$ , and total sample size at  $n = 200$ . The truncation level is set at  $\alpha = 0.2$  in the EPS.



**FIGURE 4** | Power comparison at various sample sizes. All curves are numerically computed at nominal significant level  $2.5 \times 10^{-6}$  for a fixed SNP heritability level  $h^2 = 1.0\%$ , which is determined by a set of mediate effect sizes ( $\beta_1 = 0.5744$ ,  $\beta_2 = 0.7183$ ,  $\beta_3 = 0.4564$ ) and minor allele frequency  $p = 0.25$ . The truncation level is set at  $\alpha = 0.2$  in the EPS.

more powerful than the SNP test. The SRS-based single SNP test has little power to detect the genuine genotype-phenotype association over the range of sample sizes. Under EPS, the PRT test appears more powerful than the RNA test, and RNA test appears more powerful than the SNP test.



**TABLE 1** | Ranges in sample size to achieve 80% power.

Sampling	SRS			EPS		
	SNP	RNA	PRT	SNP	RNA	PRT
$h^2 = 0.1\%$	30,409	1,540	245	16,129	795	127
$h^2 = 1.0\%$	3,055	317	102	1,744	172	57

For a given set of non-zero effect sizes in the MCM (Figures 3, 4), the EPS may provide much stronger evidence than does the SRS to reveal genuine associations between study phenotype and PRT expression (red vs. blue solid curves), RNA expression (red vs. blue dashed curves) and SNP genotype (red vs. blue dotted curves).

### Sample Sizes Required to Achieve a Certain Power

At each heritability level  $h^2 > 0$ , to achieve 80% power, these six strategies require very different samples sizes (Figure 5). Under both EPS and SRS, the SNP test requires the largest sample sizes, the RNA test requires much smaller sample sizes, followed by the PRT test. For a given heritability level, the SRS-based SNP test requires strikingly larger sample sizes than the other five strategies; and the EPS-based PRT test requires the smallest sample sizes among all the six strategies. The smaller the heritability level, the larger the difference between sample sizes required by the SRS-based and the EPS-based SNP tests. This statement also applies to the RNA and PRT tests. Table 1 lists the ranges in sample sizes of all these six strategies when SNP heritability varies from 0.1% to 1.0%.

## DISCUSSION

“Omics” technologies have advanced rapidly in the past decade. Systems biology, the integration of multi-omics technologies, aims primarily at the universal detection of genes (genomics), mRNAs (transcriptomics), proteins (proteomics) and metabolites (metabolomics) in a holistic manner. Separate omics technologies and systems biology have generated and will continue to generate huge amounts of high dimensional multilevel data. The central dogma of molecular biology (Crick, 1958, 1970) delineates a unidirectional causal flow from DNA to mRNA, protein, and metabolite (physiological trait). Separate single-platform correlation tests are useful to understand causes of phenotype variation. We theoretically compared three single-platform tests under both random sampling and extreme phenotype sampling scenarios. The proteome-trait correlation test is more powerful than transcriptome-trait correlation test, which in turn is more powerful than genotype-trait association test. The sample size required to detect a causal gene is the smallest at proteomics level and the largest at the genomics level. A direct relationship between protein expression profiles and physiological traits implies that a smaller sample size can yield more meaningful insights than relating RNA expression profiles with traits.

RNA and protein expression levels can be mapped to chromosomal loci to identify functional DNA variants of a physiological trait. RNA and protein expression levels can be considered as intermediate phenotypes, as DNA variations contribute to the physiological trait by perturbing RNA and protein expressions (Civelek and Lusis, 2013). Gene expression levels are highly heritable (Morley et al., 2004; Lappalainen et al., 2013), and specific genetic variants that influence gene expressions are known as eQTL. Multiple studies have provided strong evidence that GWA signals are enriched with eQTLs in a tissue-specific manner (Dimas et al., 2009; Nicolae et al., 2010), highlighting their utility in understanding the mechanisms underlying GWA hits. Many resources, including online databases such as GeneVar (Yang et al., 2010), are now available for eQTL analyses. It is estimated that 50–90% of eQTLs are tissue-dependent (Dimas et al., 2009; Nica et al., 2011), and trait-associated variants tend to exert more tissue-specific effects (Fu et al., 2012; Brown et al., 2013). Most identified eQTLs are cis-acting, arbitrarily defined as regulation of genes within 1 Mb, given that their effect sizes are usually relatively large and can be detected with smaller sample sizes (Cheung and Spielman, 2009). However, genetic variants can also affect the expression of genes that reside further away or are on different chromosomes (trans-eQTL) (Westra et al., 2013), but the effect sizes of trans-eQTLs are generally small, and they require larger sample sizes to detect them (Cookson et al., 2009; Grundberg et al., 2012). As a result, the number of reported trans-eQTLs has remained small (Heinig et al., 2010; Consortium, 2011; Fehrmann et al., 2011; Innocenti et al., 2011; Fairfax et al., 2012). For example, the first GWA study of asthma was published in 2007 by the GABRIEL consortium (Moffatt et al., 2007), in which 317,000 SNPs in 994 patients with childhood-onset asthma and from 1243 nonasthmatics were genotyped using family and case-control

panels. This association was then replicated in 2320 individuals from a cohort of German children and in 3301 individuals from the British 1958-birth cohort. The 17q21 region is the most consistent locus associated with asthma. Further, the most compelling eQTL associated with GSDMA expression in the lung tissues was found to reside in the same locus, suggesting that the risk allele at this locus mediate its effect by modulating GSDMA expression. The strongest eQTL SNP on 17q was rs3859192 located in intron 6 of GSDMA, which is associated with asthma (Moffatt et al., 2007, 2010). This example demonstrates the value of eQTLs (lung-specific) to refine (i.e., to fine-map) previous GWA hits for asthma. Shared eQTLs across multiple cell types and tissues also have larger effect sizes and tend to cluster around the transcriptional start site (TSS) (Dimas et al., 2009; Grundberg et al., 2012). In contrast, cell- and tissue-specific eQTLs have smaller effects and are more widely distributed around the TSS. The directions of the allelic effects for eQTLs shared in different cell types are usually consistent (Dimas et al., 2009), but cell type-dependent and tissue-dependent direction effects were also observed (Fairfax et al., 2012; Fu et al., 2012).

For protein-coding genes, functionally important changes in mRNA expression are expected to be reflected by corresponding protein changes. However, a weak correlation between transcript and protein levels in yeast shows that various other mechanisms of post-transcriptional regulation can lead to changes in protein abundance in the absence of a corresponding transcript effect (Gygi et al., 1999). Variation in protein expression levels have recently been shown to be heritable (Wu et al., 2013; Parts et al., 2014). In humans, pQTL mapping has lagged behind eQTL mapping. To date, only a few studies have explored association between SNPs and protein abundance levels (Lourdusamy et al., 2012; Hause et al., 2014). Those pQTLs that overlap with SNPs associated with physiological traits support previously identified mechanistic relationships and provide testable hypotheses of functional relationships. pQTL analyses may be helpful for gaining additional mechanistic insights into molecular underpinnings of physiological traits that is separate from eQTLs. For example, rs3865444 on chromosome 19q13.3 is strongly associated with Alzheimer's disease (AD) in a meta-analysis of several case-control studies (OR = 0.91,  $P < 1.9 \times 10^{-9}$ ). The effect allele A of rs3865444 reduces the protein abundance of CD33 (beta = 20.45, FDR  $< 5.06 \times 10^{-9}$ ) indicating that this pQTL might influence AD susceptibility through a mechanism of altered protein abundance. CD33 is a member of sialic acid-binding immunoglobulin-like lectin (Siglec) family, which regulates functions of cell in the innate and adaptive immune systems (García-Domingo et al., 1999). Thus, discoveries of eQTLs and pQTLs across multiple populations, cell types, and tissues will facilitate the identification of regulatory variation in complex traits and diseases.

Pairwise association studies often neglect a significant amount of causality information. Under the central dogma, the DNA variant is most remote from the trait; and as we illustrated, the transcriptional and translational processes can dramatically attenuate the association between the DNA variant and the trait. In practical scenarios, inconsistent and/or non-replicable findings among separate GWA studies are quite common, which

posing the critical question as to how to properly interpret the incongruences of these results. Meta-analysis is a popular tool for combining multiple independent genetic association studies to identify associations with small genetic effect sizes. In the presence of genetic heterogeneity, interpreting the meta-analysis results is an important but often difficult task (Han and Eskin, 2012). Indeed, human disease is characterized by marked genetic heterogeneity, far greater than previously anticipated (McClellan and King, 2010). Such heterogeneities can greatly reduce the power of conventional association methods. Through extensive simulation studies, Pei et al. (2014) demonstrated that (i) in the presence of between-study heterogeneity, the true genetic effect might be diluted, and meta-analysis (even with the random-effects model) may particularly introduce elevated negative rates, and (ii) replicability between meta-analyses and independent individual studies is limited, and thus inconsistent findings are not unexpected. The presence of a substantial between-study heterogeneity could lead to a power loss in meta-analyses, implying that aggregating genetically heterogeneous samples into a meta-analysis may reduce power. Meta-analyses should not and cannot be used as a gold standard to evaluate the results of individual GWA studies (Liu et al., 2013). Conventional genetic association studies ignore the essential mediate causes (RNA, protein) and the unidirectional causal pathway. Thus, such studies are often underpowered, and may not necessarily discover genuine causal variants.

The central dogma of molecular biology indicates that the transcription of mRNA from DNA and subsequent translation of mRNA into protein transform genetic blueprints into cellular functions (Crick, 1958, 1970). However, epigenetic factors represent an additional layer of complexity to our understanding of gene regulation. Epigenetic changes, i.e., reversible, heritable changes in gene regulation that occur without a change in DNA sequence, include DNA methylation and histone modification (Baccarelli and Bollati, 2009), as well as microRNA and long non-coding RNA regulation (Gomes et al., 2013). Therefore, epigenetic regulation constitutes a key regulatory mechanism in the etiology of human complex diseases (Soejima, 2009). Further, high-order epistatic interactions of genes within-/across-pathways, environmental risk factors, and gene-environment interactions contribute to attenuations of genome-trait correlations. Systems biology, the integration of multi-omics techniques, aims at the universal detection of causes for diseases and understanding of newly emerging properties revealed by holistic analyses of high-dimensional multi-omics data. It relies on an interplay between hypothesis- and discovery-driven investigations, and offers significant promises in identifying intermediate causes and causal pathways for complex traits. Existing single-platform association studies, even if helpful and successful for some scenarios, are clearly incompetent to decipher the systems pathology of complex diseases.

In this article, we formulate the central dogma of molecular biology using a multi-omics model. Under this model, we inspect the power of combination of the extreme phenotype sampling scheme and the widely-used *t*-test, providing power and sample size analyses. These results would be helpful to design cost-effective studies as well as to develop novel

multi-stage causal association inference methods. As proven, detecting associations between the more proximal variables (protein and gene expression) and the trait is more powerful than detecting the genotype-trait association. Thus, it would be effective to unravel the unidirectional causal pathways from DNA to the endpoint trait using a multi-stage strategy. The first stage identifies candidate proteins by testing trait-protein associations. The second stage identifies candidate RNAs for each candidate protein identified at the first stage by testing protein-RNA associations. The third stage identifies candidate SNPs for each RNA identified at the second stage by testing RNA-SNP associations. Each stage would remove massive non-significant variables and thus essentially reduce multiple testing burden. This strategy would effectively identify candidate genetic instruments and vertical pleiotropy pathways for further causal inferences, i.e., Mendelian randomization inferences (Smith and Ebrahim, 2003; Davey Smith and Hemani, 2014). We acknowledge that reverse causations would exist although we do not model them herein. For example, overabundance of protein may trigger reduction in mRNA through signaling mechanism. It is instructive to examine reverse causal effects and inspect performance of existing modern multi-omics methods under extreme sampling. Systems biology offers promises in identifying intermediate causers as well as unidirectional and multidirectional causal pathways. Innovative

graphical inference methods and efficient computational toolkits are in crucial demands to holistically exploit high-dimensional multi-omics data.

## AUTHOR CONTRIBUTIONS

HQ conceived the project, prepared ground theoretical results and conducted numerical explorations. HQ and TN wrote the manuscript. JZ contributed constructive comments. All authors read and approved the final manuscript.

## FUNDING

This work was partially funded by the National Institutes of Health grants R01DK091369, R01MH097018, RF1AG052476, Carol Lavin Bernick Faculty Grant (632119), Tulane's Committee on Research fellowship (600890), and Tulane Interdisciplinary Innovative Programs Hub (632037).

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2019.00110/full#supplementary-material>

## REFERENCES

- Abecasis, G. R., Cookson, W. O., and Cardon, L. R. (2001). The power to detect linkage disequilibrium with quantitative traits in selected samples. *Am. J. Hum. Genet.* 68, 1463–1474. doi: 10.1086/320590
- Ahituv, N., Kavasar, N., Schackwitz, W., Ustaszewska, A., Martin, J., Hébert, S., et al. (2007). Medical sequencing at the extremes of human body mass. *Am. J. Hum. Genet.* 80, 779–791. doi: 10.1086/513471
- Baccarelli, A., and Bollati, V. (2009). Epigenetics and environmental chemicals. *Curr. Opin. Pediatr.* 21:243. doi: 10.1097/MOP.0b013e32832925cc
- Baek, D., Villén, J., Shin, C., Camargo, F. D., Gygi, S. P., and Bartel, D. P. (2008). The impact of microRNAs on protein output. *Nature* 455, 64–71. doi: 10.1038/nature07242
- Bossé, Y. (2013). Genome-wide expression quantitative trait loci analysis in asthma. *Curr. Opin. Allerg. Clin. Immunol.* 13, 487–494. doi: 10.1097/ACI.0b013e328364e951
- Brown, C. D., Mangravite, L. M., and Engelhardt, B. E. (2013). Integrative modeling of eQTLs and cis-regulatory elements suggests mechanisms underlying cell type specificity of eQTLs. *PLoS Genet.* 9:e1003649. doi: 10.1371/journal.pgen.1003649
- Bryois, J., Buil, A., Evans, D. M., Kemp, J. P., Montgomery, S. B., Conrad, D. F., et al. (2014). Cis and trans effects of human genomic variants on gene expression. *PLoS Genet.* 10:e1004461. doi: 10.1371/journal.pgen.1004461
- Cairns, D. A., Barrett, J. H., Billingham, L. J., Stanley, A. J., Xinarianos, G., Field, J. K., et al. (2009). Sample size determination in clinical proteomic profiling experiments using mass spectrometry for class comparison. *Proteomics* 9, 74–86. doi: 10.1002/pmic.200800417
- Cheung, V. G., and Spielman, R. S. (2009). Genetics of human gene expression: mapping DNA variants that influence gene expression. *Nat. Rev. Genet.* 10, 595–604. doi: 10.1038/nrg2630
- Christensen, K., and Murray, J. C. (2007). What genome-wide association studies can do for medicine. *N. Engl. J. Med.* 356, 1094–1097. doi: 10.1056/NEJMp068126
- Civelek, M., and Lusis, A. J. (2013). Systems genetics approaches to understand complex traits. *Nat. Rev. Genet.* 15, 34–48. doi: 10.1038/nrg3575
- Coassin, S., Schweiger, M., Kloss-Brandstätter, A., Lamina, C., Haun, M., Erhart, G., et al. (2010). Investigation and functional characterization of rare genetic variants in the adipose triglyceride lipase in a large healthy working population. *PLoS Genet.* 6:e1001239. doi: 10.1371/journal.pgen.1001239
- Cohen, J. C., Kiss, R. S., Pertsemlidis, A., Marcel, Y. L., McPherson, R., and Hobbs, H. H. (2004). Multiple rare alleles contribute to low plasma levels of HDL cholesterol. *Science* 305, 869–872. doi: 10.1126/science.1099870
- Consortium, M. (2011). Identification of an imprinted master trans regulator at the KLF14 locus related to multiple metabolic phenotypes. *Nat. Genet.* 43, 561–564. doi: 10.1038/ng.833
- Cookson, W., Liang, L., Abecasis, G., Moffatt, M., and Lathrop, M. (2009). Mapping complex disease traits with global gene expression. *Nat. Rev. Genet.* 10, 184–194. doi: 10.1038/nrg2537
- Crick, F. (1970). Central dogma of molecular biology. *Nature* 227, 561–563. doi: 10.1038/227561a0
- Crick, F. H. (1958). On protein synthesis. *Symp Soc Exp Biol.* 12, 138–163.
- Davey Smith, G., and Hemani, G. (2014). Mendelian randomization: genetic anchors for causal inference in epidemiological studies. *Hum. Mol. Genet.* 23, R89–R98. doi: 10.1093/hmg/ddu328
- de Sousa Abreu, R., Penalva, L. O., Marcotte, E. M., and Vogel, C. (2009). Global signatures of protein and mRNA expression levels. *Mol. Biosyst.* 5, 1512–1526. doi: 10.1039/b908315d
- Dimas, A. S., Deutsch, S., Stranger, B. E., Montgomery, S. B., Borel, C., Attar-Cohen, H., et al. (2009). Common regulatory variation impacts gene expression in a cell type-dependent manner. *Science* 325, 1246–1250. doi: 10.1126/science.1174148
- Duncan, E. L., Danoy, P., Kemp, J. P., Leo, P. J., McCloskey, E., Nicholson, G. C., et al. (2011). Genome-wide association study using extreme truncate selection identifies novel genes affecting bone mineral density and fracture risk. *PLoS Genet.* 7:e1001372. doi: 10.1371/journal.pgen.1001372
- Dunham, I. (2018). Human genes: time to follow the roads less traveled? *PLoS Biol.* 16:e3000034. doi: 10.1371/journal.pbio.3000034
- Estrada, K., Styrkarsdóttir, U., Evangelou, E., Hsu, Y. H., Duncan, E. L., Ntzani, E. E., et al. (2012). Genome-wide meta-analysis identifies 56 bone mineral density loci and reveals 14 loci associated with risk of fracture. *Nat. Genet.* 44, 491–501. doi: 10.1038/ng.2249
- Evangelou, E., and Ioannidis, J. P. (2013). Meta-analysis methods for genome-wide association studies and beyond. *Nat. Rev. Genet.* 14, 379–389. doi: 10.1038/nrg3472

- Fairfax, B. P., Makino, S., Radhakrishnan, J., Plant, K., Leslie, S., Dilthey, A., et al. (2012). Genetics of gene expression in primary immune cells identifies cell type-specific master regulators and roles of HLA alleles. *Nat. Genet.* 44, 502–510. doi: 10.1038/ng.2205
- Fehrmann, R. S., Jansen, R. C., Veldink, J. H., Westra, H. J., Arends, D., Bonder, M. J., et al. (2011). Trans-eQTLs reveal that independent genetic variants associated with a complex phenotype converge on intermediate genes, with a major role for the HLA. *PLoS Genet.* 7:e1002197. doi: 10.1371/journal.pgen.1002197
- Fields, P. A. (2001). Review: protein function at thermal extremes: balancing stability and flexibility. *Comparat. Biochem. Physiol. A Mol. Integr. Physiol.* 129, 417–431. doi: 10.1016/S1095-6433(00)00359-7
- Fu, J., Wolfs, M. G., Deelen, P., Westra, H. J., Fehrmann, R. S., Te Meerman, G. J., et al. (2012). Unraveling the regulatory mechanisms underlying tissue-dependent genetic variation of gene expression. *PLoS Genet.* 8:e1002431. doi: 10.1371/journal.pgen.1002431
- Galvan, A., Ioannidis, J. P., and Dragani, T. A. (2010). Beyond genome-wide association studies: genetic heterogeneity and individual predisposition to cancer. *Trends Genet.* 26, 132–141. doi: 10.1016/j.tig.2009.12.008
- García-Domingo, D., Leonardo, E., Grandien, A., Martínez, P., Albar, J. P., Izpisua-Belmonte, J. C., et al. (1999). DIO-1 is a gene involved in onset of apoptosis *in vitro*, whose misexpression disrupts limb development. *Proc. Natl. Acad. Sci. U.S.A.* 96, 7992–7997. doi: 10.1073/pnas.96.14.7992
- Glazier, A. M., Nadeau, J. H., and Aitman, T. J. (2002). Finding genes that underlie complex traits. *Science* 298, 2345–2349. doi: 10.1126/science.1076641
- Goldstein, D. B., Allen, A., Keebler, J., Margulies, E. H., Petrou, S., Petrovski, S., et al. (2013). Sequencing studies in human genetics: design and interpretation. *Nat. Rev. Genet.* 14:460. doi: 10.1038/nrg3455
- Gomes, A. Q., Nolasco, S., and Soares, H. (2013). Non-coding RNAs: multi-tasking molecules in the cell. *Int. J. Mol. Sci.* 14, 16010–16039. doi: 10.3390/ijms140816010
- González-Porta, M., Frankish, A., Rung, J., Harrow, J., and Brazma, A. (2013). Transcriptome analysis of human tissues and cell lines reveals one dominant transcript per gene. *Genome Biol.* 14:R70. doi: 10.1186/gb-2013-14-7-r70
- Grundberg, E., Small, K. S., Hedman, Å. K., Nica, A. C., Buil, A., Keildson, S., et al. (2012). Mapping cis- and trans-regulatory effects across multiple tissues in twins. *Nat. Genet.* 44, 1084–1089. doi: 10.1038/ng.2394
- Guo, Y., Xiao, P., Lei, S., Deng, F., Xiao, G. G., Liu, Y., et al. (2008). How is mRNA expression predictive for protein expression? A correlation study on human circulating monocytes. *Acta Biochim. Biophys. Sin.* 40, 426–436. doi: 10.1111/j.1745-7270.2008.00418.x
- Gygi, S. P., Rochon, Y., Franza, B. R., and Aebersold, R. (1999). Correlation between protein and mRNA abundance in yeast. *Mol. Cell. Biol.* 19, 1720–1730. doi: 10.1128/MCB.19.3.1720
- Haase, C. L., Frikke-Schmidt, R., Nordestgaard, B. G., and Tybjaerg-Hansen, A. (2012). Population-based resequencing of APOA1 in 10,330 individuals: spectrum of genetic variation, phenotype, and comparison with extreme phenotype approach. *PLoS Genet.* 8:e1003063. doi: 10.1371/journal.pgen.1003063
- Han, B., and Eskin, E. (2012). Interpreting meta-analyses of genome-wide association studies. *PLoS Genet.* 8:e1002555. doi: 10.1371/journal.pgen.1002555
- Hause, R. J., Stark, A. L., Antao, N. N., Gorsic, L. K., Chung, S. H., Brown, C. D., et al. (2014). Identification and Validation of Genetic Variants that Influence Transcription Factor and Cell Signaling Protein Levels. *Am. J. Hum. Genet.* 95, 194–208. doi: 10.1016/j.ajhg.2014.07.005
- Heinig, M., Petretto, E., Wallace, C., Bottolo, L., Rotival, M., Lu, H., et al. (2010). A trans-acting locus regulates an anti-viral expression network and type 1 diabetes risk. *Nature* 467, 460–464. doi: 10.1038/nature09386
- Hindorf, L. A., Sethupathy, P., Junkins, H. A., Ramos, E. M., Mehta, J. P., Collins, F. S., et al. (2009). Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc. Natl. Acad. Sci. U.S.A.* 106, 9362–9367. doi: 10.1073/pnas.0903103106
- Innocenti, F., Cooper, G. M., Stanaway, I. B., Gamazon, E. R., Smith, J. D., Mirkov, S., et al. (2011). Identification, replication, and functional fine-mapping of expression quantitative trait loci in primary human liver tissue. *PLoS Genet.* 7:e1002078. doi: 10.1371/journal.pgen.1002078
- Jacobs, J. M., Adkins, J. N., Qian, W. J., Liu, T., Shen, Y., Camp, D. G., et al. (2005). Utilizing human blood plasma for proteomic biomarker discovery. *J. Proteome Res.* 4, 1073–1085. doi: 10.1021/pr0500657
- Kato, N., Takeuchi, F., Tabara, Y., Kelly, T. N., Go, M. J., Sim, X., et al. (2011). Meta-analysis of genome-wide association studies identifies common variants associated with blood pressure variation in east Asians. *Nat. Genet.* 43, 531–538. doi: 10.1038/ng.834
- Kingsley, C. B. (2011). Identification of causal sequence variants of disease in the next generation sequencing era. *Methods Mol Biol.* 700, 37–46. doi: 10.1007/978-1-61737-954-3\_3
- Klein, R. J., Zeiss, C., Chew, E. Y., Tsai, J. Y., Sackler, R. S., Haynes, C., et al. (2005). Complement factor H polymorphism in age-related macular degeneration. *Science* 308, 385–389. doi: 10.1126/science.1109557
- Koyutürk, M. (2010). Algorithmic and analytical methods in network biology. *Wiley Interdiscipl. Rev.* 2, 277–292. doi: 10.1002/wsbm.61
- Kung, A. W., Xiao, S. M., Cherny, S., Li, G. H., Gao, Y., Tso, G., et al. (2010). Association of JAG1 with bone mineral density and osteoporotic fractures: a genome-wide association study and follow-up replication studies. *Am. J. Hum. Genet.* 86, 229–239. doi: 10.1016/j.ajhg.2009.12.014
- Lander, E. S., and Botstein, D. (1989). Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* 121, 185–199.
- Lappalainen, T., Sammeth, M., Friedländer, M. R., AC't Hoen, P., Monlong, J., Rivas, M. A., et al. (2013). Transcriptome and genome sequencing uncovers functional variation in humans. *Nature* 501, 506–511. doi: 10.1038/nature12531
- Liu, Y. J., Zhang, L., Pei, Y., Papanian, C. J., and Deng, H. W. (2013). On genome-wide association studies and their meta-analyses: lessons learned from osteoporosis studies. *J. Clin. Endocrinol. Metab.* 98, E1278–E1282. doi: 10.1210/jc.2013-1637
- Lourdusamy, A., Newhouse, S., Lunnon, K., Proitsi, P., Powell, J., Hodges, A., et al. (2012). Identification of cis-regulatory variation influencing protein abundance levels in human plasma. *Hum. Mol. Genet.* 21, 3719–3726. doi: 10.1093/hmg/dd186
- Lundberg, E., Fagerberg, L., Klevebring, D., Matic, I., Geiger, T., Cox, J., et al. Uhlen, M. (2010). Defining the transcriptome and proteome in three functionally different human cell lines. *Mol. Syst. Biol.* 6:450. doi: 10.1038/msb.2010.106
- Manolio, T. A. (2013). Bringing genome-wide association findings into clinical use. *Nat. Rev. Genet.* 14, 549–558. doi: 10.1038/nrg3523
- McClellan, J., and King, M. C. (2010). Genetic heterogeneity in human disease. *Cell* 141, 210–217. doi: 10.1016/j.cell.2010.03.032
- Meire, D., Delplanque, J., Chèvre, J. C., Lecoeur, C., Lobbens, S., Gallina, S., et al. (2009). Genome-wide association study for early-onset and morbid adult obesity identifies three new risk loci in European populations. *Nat. Genet.* 41, 157–159. doi: 10.1038/ng.301
- Moffatt, M. F., Gut, I. G., Demenis, F., Strachan, D. P., Bouzigon, E., Heath, S., et al. (2010). A large-scale, consortium-based genome-wide association study of asthma. *N. Engl. J. Med.* 363, 1211–1221. doi: 10.1056/NEJMoa0906312
- Moffatt, M. F., Kabisch, M., Liang, L., Dixon, A. L., Strachan, D., Heath, S., et al. (2007). Genetic variants regulating ORMDL3 expression contribute to the risk of childhood asthma. *Nature* 448, 470–473. doi: 10.1038/nature06014
- Monda, K. L., Chen, G. K., Taylor, K. C., Palmer, C., Edwards, T. L., Lange, L. A., et al. (2013). A meta-analysis identifies new loci associated with body mass index in individuals of African ancestry. *Nat. Genet.* 45, 690–696. doi: 10.1038/ng.2608
- Morley, M., Molony, C. M., Weber, T. M., Devlin, J. L., Ewens, K. G., Spielman, R. S., et al. (2004). Genetic analysis of genome-wide variation in human gene expression. *Nature* 430, 743–747. doi: 10.1038/nature02797
- Mukherji, S., Ebert, M. S., Zheng, G. X., Tsang, J. S., Sharp, P. A., and van Oudenaarden, A. (2011). MicroRNAs can generate thresholds in target gene expression. *Nat. Genet.* 43, 854–859. doi: 10.1038/ng.905
- Nadeau, J. H. (2001). Modifier genes in mice and humans. *Nat. Rev. Genet.* 2, 165–174. doi: 10.1038/35056009
- Nejentsev, S., Walker, N., Riches, D., Egholm, M., and Todd, J. A. (2009). Rare variants ofIFIH1, a gene implicated in antiviral responses, protect against type 1 diabetes. *Science* 324, 387–389. doi: 10.1126/science.1167728



- Nica, A. C., Parts, L., Glass, D., Nisbet, J., Barrett, A., Sekowska, M., et al. (2011). The architecture of gene regulatory variation across multiple human tissues: the MuTHER study. *PLoS Genet.* 7:e1002003. doi: 10.1371/journal.pgen.1002003
- Nicolae, D. L., Gamazon, E., Zhang, W., Duan, S., Dolan, M. E., and Cox, N. J. (2010). Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS. *PLoS Genet.* 6:e1000888. doi: 10.1371/journal.pgen.1000888
- Panagiotou, O. A., Willer, C. J., Hirschhorn, J. N., and Ioannidis, J. P. (2013). The Power of Meta-Analysis in Genome Wide Association Studies. *Annu. Rev. Genomics Hum. Genet.* 14:441. doi: 10.1146/annurev-genom-091212-153520
- Parts, L., Liu, Y. C., Tekkedil, M. M., Steinmetz, L. M., Caudy, A. A., Fraser, A. G., et al. (2014). Heritability and genetic basis of protein level variation in an outbred population. *Genome Res.* 24, 1363–1370. doi: 10.1101/gr.170506.113
- Pei, Y. F., Zhang, L., Pappasian, C. J., Wang, Y. P., and Deng, H. W. (2014). On individual genome-wide association studies and their meta-analysis. *Hum. Genet.* 133, 265–279. doi: 10.1007/s00439-013-1366-4
- Plomin, R., and Davis, O. S. (2009). The future of genetics in psychology and psychiatry: microarrays, genome-wide association, and non-coding RNA. *J. Child Psychol. Psychiatry* 50, 63–71. doi: 10.1111/j.1469-7610.2008.01978.x
- Qiu, L. Q., Stumpo, D. J., and Blackshear, P. J. (2012). Myeloid-specific tristetraprolin deficiency in mice results in extreme lipopolysaccharide sensitivity in an otherwise minimal phenotype. *J. Immunol.* 188, 5150–5159. doi: 10.4049/jimmunol.1103700
- Ruiz-Narváez, E. A. (2011). What is a functional locus? Understanding the genetic basis of complex phenotypic traits. *Med. Hypoth.* 76, 638–642. doi: 10.1016/j.mehy.2011.01.019
- Schadt, E. E., Monks, S. A., Drake, T. A., Lusk, A. J., Che, N., Colinayo, V., et al. (2003). Genetics of gene expression surveyed in maize, mouse and man. *Nature* 422, 297–302. doi: 10.1038/nature01434
- Schwahnäusser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., et al. (2013). Corrigendum: global quantification of mammalian gene expression control. *Nature* 495, 126–127. doi: 10.1038/nature11848
- Selbach, M., Schwahnäusser, B., Thierfelder, N., Fang, Z., Khanin, R., and Rajewsky, N. (2008). Widespread changes in protein synthesis induced by microRNAs. *Nature* 455, 58–63. doi: 10.1038/nature07228
- Smith, G. D., and Ebrahim, S. (2003). 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *Int. J. Epidemiol.* 32, 1–22. doi: 10.1093/ije/dyg070
- Soejima, H. (2009). Epigenetics-related diseases and analytic methods. *Rinsho Byori* 57, 769–778.
- Stranger, B. E., Stahl, E. A., and Raj, T. (2011). Progress and promise of genome-wide association studies for human complex trait genetics. *Genetics* 187, 367–383. doi: 10.1534/genetics.110.120907
- Vogel, C., de Sousa Abreu, R., Ko, D., Le, S.-Y., Shapiro, B. A., Burns, S. C., et al. (2010). Sequence signatures and mRNA concentration can explain two-thirds of protein abundance variation in a human cell line. *Mol. Syst. Biol.* 6:400. doi: 10.1038/msb.2010.59
- Webster, J. A., Gibbs, J. R., Clarke, J., Ray, M., Zhang, W., Holmans, P., et al. (2009). Genetic control of human brain transcript expression in Alzheimer disease. *Am. J. Hum. Genet.* 84, 445–458. doi: 10.1016/j.ajhg.2009.03.011
- Weedon, M. N., Lango, H., Lindgren, C. M., Wallace, C., Evans, D. M., Mangino, M., et al. (2008). Genome-wide association analysis identifies 20 loci that influence adult height. *Nat. Genet.* 40, 575–583. doi: 10.1038/ng.121
- Welter, D., MacArthur, J., Morales, J., Burdett, T., Hall, P., Junkins, H., et al. (2014). The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res.* 42, D1001–D1006. doi: 10.1093/nar/gkt1229
- Westra, H. J., Peters, M. J., Esko, T., Yaghootkar, H., Schurmann, C., Kettunen, J., et al. (2013). Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nat. Genet.* 45, 1238–1243. doi: 10.1038/ng.2756
- Willer, C. J., Sanna, S., Jackson, A. U., Scuteri, A., Bonnycastle, L. L., Clarke, R., et al. (2008). Newly identified loci that influence lipid concentrations and risk of coronary artery disease. *Nat. Genet.* 40, 161–169. doi: 10.1038/ng.76
- Wu, L., Candille, S. I., Choi, Y., Xie, D., Jiang, L., Li-Pook-Than, J., et al. (2013). Variation and genetic control of protein abundance in humans. *Nature* 499, 79–82. doi: 10.1038/nature12223
- Xiong, M., Fan, R., and Jin, L. (2002). Linkage disequilibrium mapping of quantitative trait loci under truncation selection. *Hum. Hered.* 53, 158–172. doi: 10.1159/000064978
- Yang, T. P., Beazley, C., Montgomery, S. B., Dimas, A. S., Gutierrez-Arcelus, M., Stranger, B. E., et al. (2010). Genevar: a database and Java application for the analysis and visualization of SNP-gene associations in eQTL studies. *Bioinformatics* 26, 2474–2476. doi: 10.1093/bioinformatics/btq452

**Conflict of Interest Statement:** 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.

Copyright © 2019 Qin, Niu and Zhao. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). 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.