

Review

Another Round of “Clue” to Uncover the Mystery of Complex Traits

Shefali Setia Verma^{1,2} and Marylyn D. Ritchie^{1,2,*}

¹ The Huck Institute of Life Sciences, The Pennsylvania State University, University Park, PA 16802, USA; shefali.setiaverma@pennmedicine.upenn.edu

² Department of Genetics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA

* Correspondence: marylyn@upenn.edu

Received: 1 November 2017; Accepted: 15 January 2018; Published: 25 January 2018

Abstract: A plethora of genetic association analyses have identified several genetic risk loci. Technological and statistical advancements have now led to the identification of not only common genetic variants, but also low-frequency variants, structural variants, and environmental factors, as well as multi-omics variations that affect the phenotypic variance of complex traits in a population, thus referred to as complex trait architecture. The concept of heritability, or the proportion of phenotypic variance due to genetic inheritance, has been studied for several decades, but its application is mainly in addressing the narrow sense heritability (or additive genetic component) from Genome-Wide Association Studies (GWAS). In this commentary, we reflect on our perspective on the complexity of understanding heritability for human traits in comparison to model organisms, highlighting another round of clues beyond GWAS and an alternative approach, investigating these clues comprehensively to help in elucidating the genetic architecture of complex traits.

Keywords: complex traits; meta-dimensional analysis; multi-omics datasets; heritability; game of “Clue”

1. Complex Diseases and the Concept of Heritability

Elucidating the genetic underpinnings of diseases that occur frequently in the population, such as obesity, type 2 diabetes, hypertension, and cancer, are essential research foci among researchers in the human genetics community. The underlying genetic etiologies behind these common human traits are inherently complex due to the effects of multiple genes on the phenotype in comparison to diseases such as cystic fibrosis which follow a Mendelian pattern of inheritance [1]. Remarkably, genetic association studies have been shown to be successful over the past two decades in identifying genetic variants associated with common, complex traits. For example, the first successful Genome-Wide Association Studies (GWAS) identified moderate effect size variants such as *CFH* for age-related macular degeneration (odds ratio 1.25–20.28) [2–4], while rare variant association studies have identified multiple rare variants in *ANGPTL4* associated with Coronary Artery Disease (odds ratio 0.32–0.81) [5]. Discovery of many such associations has also impacted the field of pharmacogenomics where variants from these association studies have resulted in drug development and repositioning strategies [6,7].

GWAS are mainly performed using retrospective study designs from sample populations collected from either academic medical centers/healthcare provider organizations (examples include eMERGE Network, MyCode Community Health Initiative, and BioVU among others) [8–10] or population-based, epidemiological study designs (NHANES, GIANT, CHARGE, etc.) [11–13].

Despite the success of association analyses among heritable complex diseases, a meager proportion of phenotypic variance has been explained [14–16]. Many factors contribute to the unexplained

proportion of variance. In most studies, environmental factors and longitudinal effects on the population remain underutilized merely due to unavailability resulting from the strenuous task of collecting these measures [17,18]. Alongside, contributions of additional factors such as structural variations, epistasis, and environmental factors have been proposed as alternative hypotheses for understanding the genetic architecture of complex traits. Comprehensive conventional approaches for validating these hypotheses in model organisms have also shown great success [19–23]. A myriad of approaches are applied for testing the effects of genetic associations in model organisms such as yeast, flies, and mice [24–28]. These model organisms possess human orthologous genes as well as phenotypes that can be directly correlated with human phenotypes [29,30]. Testing of associations in model organisms has helped us in many ways, but the gap between validation of all possible genetic associations and the limited number that have been achieved lies in the potential genetic and phenotypic overlap between humans and model systems. Phenotypic stability among model organisms results in straightforward phenotypic changes due to the low effects of external factors, such as environment, which occasionally makes the concept of missing heritability in humans seem delusional [31]. Validation of associations in model organisms are usually evaluated as quantitative traits but that is not true for all human phenotypes and likewise not all human genes have model organism gene orthologs. Another difference between humans and model organisms lies in the complexity and heterogeneity on both the phenotype and genotype side in humans, whereas model organisms are much simpler. The human genome has much more complex linkage disequilibrium and population diversity than model organisms. In addition, in model organisms, the environment and phenotypes are well controlled in laboratory testing. These differences are depicted in Figure 1.

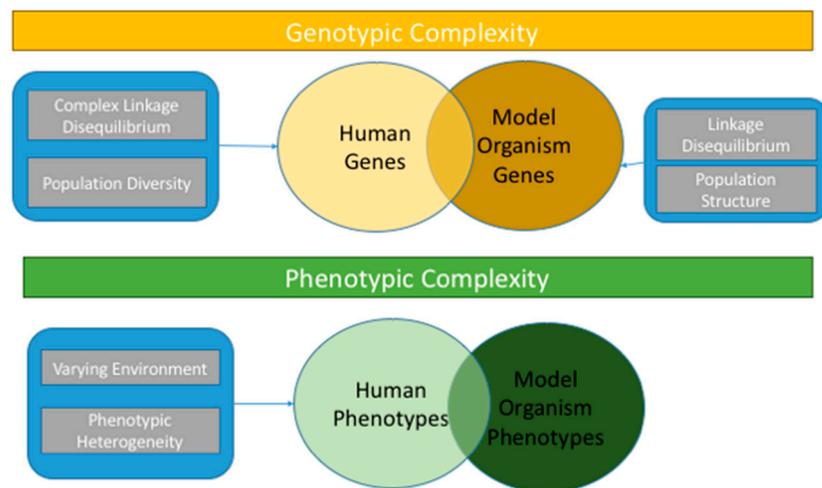


Figure 1. Differences among genotypic and phenotypic complexity in humans and model organisms. The intersection represents orthologous genes (yellow section) and phenotypes (green section).

Heritability of a disease trait refers to the proportion of variance that can be explained by genetic factors. Estimation of heritability is usually done by observing patterns of inheritance among samples either in family based studies or in population based studies. In family based studies, patterns among a pedigree of family members or among monozygotic and dizygotic twins are estimated. In these studies, environmental factors are assumed to be constant. Whereas in population based studies, patterns of inheritance among the population are observed in non-stationary environmental conditions. In this commentary, we highlight studies aimed towards understanding the heritability of complex traits in the realm of sifting through the stack of clues to uncover the mystery of complex trait genetic architecture. We will also emphasize the validation of association analyses from model organisms where available and challenges in interpreting these validated associations.

2. Clues to Elucidating the Underlying Genetic Architecture of Complex Traits

Genetic association studies are founded on the bedrock of the concept of heritability. Significant challenges for association studies arise in the realm of answering the following fundamental questions:

1. How much closer do we get to explaining heritability as estimated by family and twin studies by exploiting genetic variations in population based studies?
2. How much of the phenotypic variance is additive (i.e., the combinatorial effect of all variations)? This is referred to as narrow sense heritability.
3. How much of variance cannot be explained merely by adding all variations in a single model?

Answers to these questions vary from one phenotype to another phenotype. For example, a trait such as height is highly heritable ($>90\%$ h^2_{FAM} (family based estimate of heritability) [32]; offspring resemble their parents) and association studies have shown that the majority of height heritability can be explained by the additive variance components of common variations h^2_{SNP} (SNP-based estimate of heritability), whereas complex traits such as obesity, type 2 diabetes, and neurological disorders among others are also estimated to be highly heritable (h^2_{FAM} 47–90%, 20–80%, and 37–81%, respectively) [33,34]. However, association studies using common variants have only explained 27%, 10%, and 21–28% proportion of variance for these traits, respectively [33,35,36]. Considering the predominant explanation for height is through additive genetic variance, GWAS studies have also explained a substantial amount of the genomic heritability for some other traits such as body mass index (BMI), where over 50% of the heritability estimated by family and twin studies has been explained. The same does not stand true for many other binary phenotypes, such as type 2 diabetes, Crohn's disease, Alzheimer's disease, etc. There are many clues that may contribute towards the unexplained heritability of these types of complex traits [37–39] which include effects due to: (1) rare variations; (2) structural variations; and (3) gene–gene or gene–environment interactions effects among others. Exploring multi-omics data in many studies has also gotten us a bit closer to uncovering the mystery of complex trait architecture [40]. That said, it is still proven difficult to predict which genomic components/elements will be essential for each different complex phenotypic trait. How can we know which elements are responsible for the underlying genetic architecture of each unique trait? To uncover the heritability of common, complex traits, it seems as though most or all these components are playing a different “Game of Clue” in explaining the proportion of phenotypic variance as attributable to genetic/genomic variations. In the following section, we propose our model for understanding this mystery of complex trait architecture by putting the genomic elements contributing to the different effects in the context of “suspects”, types of statistical analysis methods in the context of “weapons”, and which tissue these elements are being tested/evaluated in the context of “rooms” to compare it to the Hasbro board game “Clue: The Classic Mystery Game” (Hasbro, Pawtucket, RI, USA). In this classic board game, the goal is to identify who committed a crime, with what weapon the crime was committed, and where the crime took place. A similar analogy has been used in the area of genome sequencing to solve the medical mystery of a drug resistant bacterial spread [41]. This analogy has also been used to describe the mystery of missing heritability [42]; we expand upon this idea in this review. We describe this analogy further in Figure 2. We propose that each common, complex trait should be studied in the same manner that a game of “Clue” would be played. Here, the “crime” is the complex trait risk. In each new round, there is no a-priori hypothesis about who the suspect would be. There is no preconceived notion of the weapon or the room. Similarly, in studies of complex traits, we can expect different genomic and environmental elements to be responsible for disease risk, based on different underlying models, functioning in different tissues. We should not assume that all complex traits will follow a polygenic, additive model simply because height demonstrates this type of model. Other traits may be due to different types of effects, in alternative models, and in different tissues. We discuss all of these components in detail in the following sections.

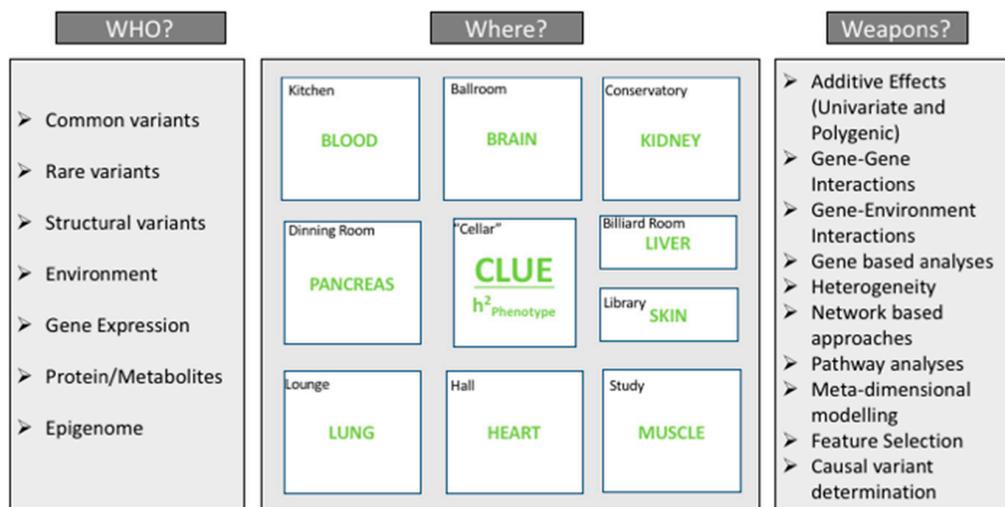


Figure 2. Unriddling undercover heritability. A depiction of the mystery of heritability in the context of the “Game of Clue.” Here, tools and methods to understand heritability are shown as weapons, suspects are genomic elements contributing to heritability, and tissues that are impacted are represented as rooms on the “Clue” game board. The size of rooms does not correspond to importance. This figure is adapted from [43].

3. Suspects/Who Did It? Considering Different Types of Omics and Environmental Variability as the Suspects in the Crime (Influencing Disease Risk)

3.1. Common Variants

Over the past decade, studies to correlate common genetic variations with phenotypic effects have become a standard part of genetic association studies. Among the most popular tools/weapons for investigating common variants are GWAS where single nucleotide polymorphisms (SNPs) obtained via genotyping chips and/or imputed data are tested for association with a phenotype of interest. GWAS have identified over 50 K SNP-trait associations [44–46]. GWAS studies are based on patterns of linkage disequilibrium (LD) and assume that tag SNPs in LD with causal SNPs could help in increasing our understanding of complex traits. Analyses of common variants (SNPs) in common diseases follows the common-variant-common-disease (CDCV) hypothesis, which suggests that a combination of common loci could be responsible for common diseases [47,48]. GWAS have been successful in identifying tag SNPs for complex disorders such as Crohn’s disease [49], obesity [50], type 2 diabetes (T2D) [51], multiple sclerosis [52], age-related macular degeneration (AMD) [4] and breast cancer [53] among many others. Many associations identified by GWAS have been validated in model organisms and have also shown pharmacological implications. For example, GWAS have identified variations in the *FTO* gene to be associated with obesity and studies in mouse models also suggest a lean body type when these genetic variations are present [54]. Orthologous phenotypes for metabolic traits such as obesity and T2D are easily captured in model organisms whereas other complex human diseases such as AMD, which is comprised of many factors, are difficult to recapitulate exactly in model organisms [55]. Despite this difficulty, mouse models have suggested an important role of GWAS-identified variants in *CFH* with AMD progression and pharmacological progress has been made regarding the design of a drug targeting *CFH* for AMD patients [6].

Biological processes are complex, and they arise due to the integration of many genes that interact in a pathway to complete a cellular process. Thus, exploiting the interactive effects of common variants is a natural extension in elucidating the complexity of common diseases. Analyses of the non-linear genetic effects of common variants are referred to as epistasis [19,38,56,57]. Non-additive genetic effects can be obtained via statistical methods that may not necessarily correspond to biological epistasis [58].

While additive effects can only explain narrow-sense heritability, exploring dominance and interacting effects could potentially get us closer to the estimated heritability, thus getting one step closer to uncovering the underlying heritability. Significant evidence of epistasis has been demonstrated in model organisms such as *Drosophila melanogaster* and *Saccharomyces cerevisiae* [59–62], but the role and validation of epistasis for most (but not all) complex human diseases are still hidden due to numerous challenges in discovery and replication of epistatic effects. These challenges are out of the scope of this review but have been discussed previously by many [14,37,63]. We hypothesize that for some complex traits, common variants will act through additive models; where other common variants will act through dominant and epistatic models. We should allow the common variant data to be explored under each of these scenarios to maximize our ability to identify common variants associated with complex traits.

3.2. Rare Variants

Genetic variants that are less frequent in the population could have potentially large effects on complex diseases, as illustrated in common-disease-rare-variant (CDRV) hypothesis. The CDRV model is not captured by common variant GWAS designs [48]. Analyzing the single effects of each rare variant is not statistically powerful due to the number of samples sequenced in rare variant association studies (RVAS) [64,65] and the rarity of the genetic variants. Thus, collapsing variants into a gene, pathway, or other variant-set to test the association of the contributions of multiple rare variants in collapsed regions on the phenotype of interest is the most popular method. Many analysis tools have been developed for these analyses such as BioBin [66], Sequence Kernel Association Test (SKAT) [67], Variant Association Tools [68], and RVTESTS [69]. Several genes containing rare variants have been identified with moderate to high effects on complex traits such as the association of low-frequency variants in *IFIH1* with type 2 diabetes [70] and the association of gain and loss of function (LOF) rare variants in *PCSK9* with low density lipoproteins (LDL) levels [7,71]. The role of *PCSK9* in designing lipid-lowering medications has been implicated in mouse model testing [72,73]. Large-scale studies of multiple rare variants are still in their infancy. However, we expect to see an emergence of rare variant association analyses in the coming years. As these studies emerge, we will develop a better understanding of the role of rare variants in the architecture of common, complex traits.

3.3. Structural Variations

Single nucleotide variants (SNVs) as explained in the previous two sections are of great importance in understanding the link between genotype and phenotypes, but large structural variations such as insertions, deletions, translocations, and inversions [74] also play a major role in affecting complex diseases and traits. It has been shown that the human genome consists of multiple recurrent sections of insertions and deletions [75,76]. These variations are rather rare in the population occurring at a frequency of less than 0.05% [77]. Structural variations are commonly referred to as copy number variations (CNV) and these large variations span across one or multiple genes. Given the inherent complexity of genotype and phenotype, it is easy to believe that these large variations which span through many genes could influence a large variety of pathways underlying complex diseases [78]. Since CNV frequency and distribution varies to a large extent in even closely related samples, in population-based studies, a measure of the burden of CNV is sometimes used to associate with diseases and traits. Evidence of association of CNV burden has been shown for neurological and behavioral traits such as Autism Spectrum Disorders [79]. We expect that the evidence for the role of structural variation in complex traits will continue to emerge as the resolution and sensitivity of our molecular technologies continue to improve for structural variant detection.

3.4. Environmental Factors

The broad sense heritability (H^2) of complex diseases can be estimated from both genetic and environmental factors. Heterogeneity in association results could be due to the effects of phenotype interactions with exposure or the interaction of genotypes with exposure. Thus, studying the effect of environment is crucial in understanding biological pathways and mechanisms behind complex traits. Studying the effects of environmental exposures on phenotypes is referred to as environment-wide association study (EWAS) [80,81] and the study of the effects of genetic variants in the context of the environment is performed through gene–environment interaction studies [82–84]. Environmental exposures vary across a population-based study to a large extent and thus pose several challenges to gene–environment studies. A comprehensive collection of complex heritable measures is required. Several studies by Patel et al. [80,85] have identified disease-associated exposure factors such as pesticides with type 2 diabetes and effects of heavy metals with serum lipids among others. The effect sizes of the exposures on phenotypes are quite high (odds ratios range from 2–4). Thus, many efforts to measure the exposome are in place to obtain standard global environmental variables on study populations [86]. We hypothesize that gene–environment interactions will explain a great deal of undercover heritability.

3.5. Gene Expression

Specific regions of the genome consist of genetic variations that lead to highly heritable variability of gene expression [87]. These regions are known as expression quantitative trait loci (eQTL). The effects of eQTLs on the expression of genes are highly tissue dependent. Thus, testing the effect of eQTLs on diseases or traits is also important because these directly affect the expression of genes. Heritability of gene expression is mediated due to the presence of specific variants (such as eQTLs as explained above). Studies by Moffat et al. [88] and Zhu et al. [89] have identified eQTL signals associated with several complex traits. Even though gene expression as measured by next generation sequencing technologies might not directly explain heritability, it is an important suspect to investigate in elucidating the genetic architecture of complex traits to observe how gene expression influences disease risk (hence the crime in investigation). Various methods to test the effects of gene expression on diseases exist, such as Summary based Mendelian Randomization (SMR), PrediXcan, MetaXcan, and CAVIAR (Causal Variant Identification in Associated Regions) [40,89–91].

Gene expression data have also been utilized in multi-omics integrative approaches. For example, LaCriox et al. [92] proposed a pyramid approach to test for the effect of SNPs and gene expression on the disease. Kim et al. [93] utilized gene expression data from The Cancer Genome Atlas (TCGA) to predict integrative non-linear effects of SNPs, gene expression, and methylation data on complex cancer phenotypes via an artificial neural network based approach (ATHENA) [94].

Another popular approach to leverage gene expression data to identify how gene expression mediates the effect of genetic variants on diseases and trait is Transcriptome Wide Association Study (TWAS). Gusev et al. [95] showed the efficiency of TWAS in identifying genes associated with anthropometric traits. TWAS methods are still in their early stages. Two types of methods have been suggested in this category: Summary Based Mendelian Randomization (TWAS-SMR) and imputation of the cis-genetic component using SNP information TWAS multi-SNP prediction (TWAS-MP) (e.g., PrediXcan). It is not surprising that gene expression affects disease, but TWAS methods do require more thorough testing to understand the suitability to use these methods to understand various genetic architectures and how these analyses can be validated in model organisms.

3.6. Protein/Metabolites

The genetic underpinning of disease phenotypes can be better understood better if we can pinpoint the agitations in normal cellular functions that lead to the disease process. The approaches as mentioned above to identify how DNA variations predispose individuals to disease are important, but they do not get us very close to identifying the underlying mechanisms affecting the phenotype unless the specific causal variants are tested. Amino acid changes can lead to the disruptions of proteins in biological pathways. Testing for the effect of protein variability on human diseases could have potential implications for identifying drug targets, susceptibility to diseases, and also in developing preventive care measures. The effect of proteins can be tested in many different ways: analyzing the effect of protein–protein interactions, the impact of protein complexes, and the effect of metabolites. The field of study to link proteins and their functions to diseases is referred to as proteomics. Biomarkers identified by proteomics are used in developing diagnostic measures for early detection of various types of cancers such as prostate cancer [96,97], as well as monitor progression of ovarian cancer [98,99]. Information on proteins such as metabolites can also be used as filters for pinpointing the underlying etiologies for complex traits. For example, Lee et al. [100] hypothesized that testing the effect of metabolites on metabolic disorders would yield important insights. They utilized a network-based approach to identify connections among multiple metabolic disorders based on circulating metabolites. The protein–protein interactions database is a resource consisting of information on biological interactions among proteins. Sun et al. [101] filtered CNVs to identify epistatic effects among CNVs that mapped to proteins based on the protein–protein interactions database and evaluated the impact of CNV–CNV interactions on the expression of genes.

3.7. Epigenome

Germline genetic variations in DNA of an individual do not change significantly during the life course, but the chemical changes in DNA such as methylation and histone modification do change [102,103]. Many factors influence the change in the conformation of DNA structure that could affect underlying cellular mechanisms. Thus, it is also impactful to understand how epigenetic factors influence the disease risk by investigating differences in epigenomes of carriers and non-carriers of diseases. Analysis of the epigenome is mostly studied in the field of cancer genomics to see how these epigenetic factors differ in the cancerous vs. normal tissue [104,105]. Epigenetic changes observed in model organisms can attribute similar biological behaviors in humans because of similar genes in conserved pathways [106]. Thus, the epigenetic analysis is relevant and likely fruitful in model organisms. Model organisms such as yeast and fruit flies, among others, are used to study epigenetic changes such as chromatin structure, DNA methylation, RNA interference (RNAi) pathways, histone modifications [107–111].

4. What Is the Weapon of Choice? Which Type of Tools Can Help Elucidate the Significant Risk Factors for Complex Diseases?

In the previous section, we introduced the lineup suspects (or “WHO?”) in the still uncovered mystery of complex trait architecture. These suspects are responsible for influencing disease risk. Another important aspect is to identify which weapons (or “HOW?” the independent variables are modeled) can be used for studying the behaviors of suspects discussed above. Weapons here refer to the tools that can be used by researchers in identifying underlying biological mechanisms. Multiple analysis tools exist in the literature to focus on one or more factors that can contribute towards the susceptibility of complex traits. These weapons or modeling tools are listed in Table 1. Some of these tools have already been described in previous sections as a means of understanding the methods used to test or evaluate each suspect’s behavior.

Table 1. A brief list of “weapons” (i.e., models/tools) available to identify genome-phenome associations to uncover heritability. Many of the tools are compiled from Omic Tools resource [112].

Weapon	Suspects	Tool Name	Reference
Additive Model	Common variations	PLINK	[113]
	Common variations	PLATO	[114]
	Common variations	QCTool	[115]
	Common variations	GenAbeL	[116]
	Common and Rare Variations	BOLT-LMM	[117]
	Common and Rare Variations	FAST-LMM	[118]
	Structural Variations	CNVTools	[119]
	Structural Variations	PennCNV	[120]
	Structural Variations	CKAT	[121]
	Structural Variations	ParseCNV	[122]
	SNPs and Structural Variations	CNVassoc	[123]
	Common and Rare Variations	RVTests	[124]
	Common and Rare Variations	PLINK/SEQ	[125]
	Rare Variations	EPACTS	[126]
Common variations	MAGMA	[127]	
Rare Variations	EMMAX	[128]	
Gene–Gene Interactions Model	Common variations	MDR	[129]
	Common variations	AntEpiSeeker	[130]
	Common variations	MultiSURF	[131]
	Common variations	BOOST	[132]
	Common variations	PLATO	[114]
	SNPs and Structural Variations	CNVassoc	[123]
	Common variations	SNPTEST	[133]
	Common variations	TS-GSIS	[134]
	Common variations	SNPAssociation	[135]
	Common variations	PLINK	[113]
Common Variants and Phenotypes	CAPE	[136]	
Gene-Environment Interactions Model	Common variations and Environment	PLATO	[114]
Detecting Heterogeneity	Genetic variations and Phenotypes	JBASE	[137]
	Gene Expression and Phenotype	SMR	[138]
	Gene Expression and Phenotype	FAST-LMM-EWASher	[139]
	Phenotype	LiChE	[140]
	Genetic and phenotypic	BUHMBOX	[141]
	Genetic Heterogeneity	ForestPMPlot	[142]
	Genetic variations and Phenotypes	NetDx	[143]
	Genetic Heterogeneity	BioGranat-IG	[144]
Network based approaches	SNPs, Phenotypes and Gene Expression	NETAM	[145]
	Common variations	EINVis	[146]
	Gene Expression and Phenotype	NetDecoder	[147]
	Common variations	ViSEN	[148]
	All genetic variations	Cytoscape	[149]
Pathway analyses	Common variations	PARIS	[150]
	Genes	SNPSea	[151]
	Genes	GSEA	[152]
	Common variations	VEGAS2Pathway	[153]
	Common variations	MAGENTA	[154]
Meta-dimensional modelling	Multi-Omic Datasets	ATHENA	[155]
	Multi-Omic Datasets	NetDX	[143]
	Multi-Omic Datasets	iCluster	[156]
Gene-based analyses	All genetic variations	Biofilter	[157]
	Common and Rare Variations	SKAT	[158]
	Rare Variations	BioBin	[159]
	Rare Variations	Variant Association Tools	[68]
	Rare Variations	EPACTS	[126]
Feature Selection/Prioritization	All genetic variations	Biofilter	[157]
	Common variations	GLM (LASSO and Elastic-Net)	[160]
	Common variations	RANGER	[161]
	Common variations	Gradient Boosting	[162]
Causal Variant Determination	Common variations	TATES	[163]
	Common variation and eQTL	CAVIAR	[164]
	Common variation and eQTL	PrediXcan	[165]

A key conclusion from analyzing all of these weapons is that there is likely no one single weapon that can investigate all variations (“omics”) together simultaneously. Meta-dimensional approaches aim to explore not only genomic features but also proteomic and epigenetic features by integrating the effects of all variations in ingenious ways [94,166]. However, even those approaches cannot be used in isolation, and other methods should be explored to improve our understanding of complex trait architecture. There are obvious strengths and caveats to using each of these approaches which are out of scope of this review to highlight. However, we recommend that, when exploring a new complex disease or trait, multiple suspects are considered along with multiple weapons (or analysis tools) to evaluate the trait of interest thoroughly. It is crucial to evaluate multiple possible analysis strategies for the study of each unique disease or trait.

5. Where? Which Tissue(s) Are Important for the Evaluation of Omics Associations?

Every suspect (“variations” as described in previous sections) may alter different tissue types in various biological processes. DNA variations are primarily obtained from blood or serum plasma samples, but it is essential to know which pathway, tissue, or cell type that the risk locus affects. The GTEx database is a powerful resource where the link between gene and the tissue that is affected is investigated [167]. Similarly, for many ocular traits, an ocular tissues database has been utilized [168]. SNPsea is a tool developed at the Broad Institute (Cambridge, MA, USA) to help prioritize the genetic associations to identify tissues and pathways that are affected by the expression of specific genes [169]. Liu et al. utilized SNPsea to identify tissues that are affected by GWAS associations for Systemic Lupus Erythematosus [170] and Hu et al. [171] utilized the same approach to identify tissues affecting the risk loci for autoimmune diseases. Another example where tissue was explored is in an attempt to identify the location of the causes of missing heritability in Type 2 Diabetes, which is one of the most well-studied diseases to identify genetic associations. To identify causal variants and tissues affected by the risk loci, researchers have found that risk associated loci for diabetes and many other metabolic disorders affect adipose tissue as well as observed effects in islets of the pancreas [172–174].

In the search for undercover heritability, many suggestions have been made. Genotyping and sequencing technologies rely on the collection of blood samples from study population. Thus, the tissue investigated remains constant in majority of the studies. The common variant common diseases hypothesis (CDCV) suggests that many common variants are the suspects, an additive genetic model is the weapon, and the disease risk manifests in the blood (where the SNPs are measured). Likewise, the rare variant common disease (RVCD) hypothesis suggests that rare variants are the suspects, either burden or distribution of multiple rare variants are the weapons, and blood is the location where the variation is measured. All of these suggestions are tested in one study or the another, but in reality, it seems that we should consider both rare and common variants in multiple possible genetic models for complex diseases. In addition, in elucidating the etiologies of complex traits, the “murder” (or undercover heritability) can take place any location, where location refers to the tissue that is affected by the risk loci. Hence, the secret of this game to identify undercover heritability is to be more open-minded and evaluate each new study with the broad set of data types, tools, and tissues where possible.

6. Estimating Heritability (Making a Suggestion in the Game of “Clue”)

The measure of heritability that explains the degree to which a trait is inherited from one generation to another refers to how the heritability values are quantified. The process in which heritability is studied can take place in several different designs as explained below:

- Family and Twin studies: In family and twin studies, a set of related individuals and their phenotypic traits are analyzed to identify how heritable the phenotypic trait is in families and in sets of identical twins respectively. Family study estimates are usually lower than twin study estimates. For example, family studies for BMI estimate that BMI is 24–81% heritable whereas twin studies estimate BMI to be 47–90% heritable [34]. The estimates from studies of related

individuals take the effects of the environment into consideration and, thus, generally broad sense heritability is estimated by these methods.

- **Population based studies:** Genomic heritability mainly refers to the proportion of trait variance that can be attributable to genetic factors such as common variants and low frequency or rare variations. Many methods and tools exist in the literature to measure heritability among a set of unrelated individuals. These include mixed model approaches (GCTA, REACTA, PLINK, etc.) [175,176], Bayesian approaches (example BGLR) [177], LD based weighted methods in mixed linear model approaches (LDAK) [178], and machine learning approaches (HERRA and MEGHA) [179,180]. All of these methods are focused towards explaining the additive variance component (i.e., narrow sense heritability). Narrow sense heritability estimates from GWAS studies and for either all variants on the genotyping chip assayed or only a subset of statistically significant variants. Locke et al. [181] determined the variance components for BMI based on statistically significant GWAS variants from their study and showed that 97 genome-wide significant loci can explain only 2.7% of the variance, whereas the overall SNP heritability (i.e., heritability from all available genotyped and imputed SNPs) is 75%, as shown by Robinson et al. [182]. Recent studies have also looked at the proportion of phenotypic variance explained by partitioning the genome. Speed et al. [183] showed how the proportion of variance explained for 19 different traits varies across the genome by chromosome and by minor allele frequency ranges. Finucane et al. [184] proposed LD score regression method using summary statistics from GWAS to partition heritability across the genome based on functional annotations.

Methods described above refer to the narrow sense heritability as estimated from GWAS studies which reflects only the additive variance component. With the advent of sequencing technologies, new methods analyzing the effects of rare variants and genetic and gene–environment interactions are also necessary. Heritability estimates for rare variants can be obtained from single variants as well as collapsed regions containing information on the burden of rare variants. To determine unbiased estimates of heritability for gene-based rare variants, Liu et al. [185] proposed a mixed model approach. Next, an estimation of unbiased estimates of interacting variance components remains to be a challenge. Ronnegard et al. [186] estimated genetic marker interactive effects using a ridge regression-based approach. This analysis was conducted in 74 samples from *Arabidopsis thaliana*, but its application on big, large-scale human datasets is still to be explored.

7. The Focus of Future Studies, What to Expect?

Addressing a phenotype of interest and analyzing an accurate representation of the phenotype is of great importance. Association analyses are often conducted as retrospective studies, for example by utilizing electronic health record datasets [8], epidemiological datasets [187] and clinical trials datasets [188], among others. Testing of multiple phenotypes simultaneously in PheWAS (Phenome-Wide Association Studies) [189] is also becoming immensely popular, especially with the use of electronic health record (EHR) data and availability of multiple phenotypes. PheWAS provides a significant advantage to test for phenotype relationships (comorbidities and potential pleiotropy) which a single phenotype association study lacks. Occurrences of phenotypes are dependent on many demographic factors such as ancestry, sex, age, etc. Association studies account for these factors in some ways by considering them as confounders and adjusting for their effects. However, it is also crucial to investigate these factors in the context of sub-phenotypes. Verma et al. approached this strategy to address the context of phenotypes in an Aids Clinical Trial Group (ACTG) dataset [190] and identified variants that showed effects on each phenotype of interest in the presence of a drug and in a specific context, where a similar phenotype was not observed in other contexts. This leads to the concept of phenotypic heterogeneity. It is essential to recognize the differences in a phenotype. Association studies aim at understanding a complex phenotype such as type 2 diabetes, cataract, glaucoma, chronic obstructive pulmonary disorder, etc. The analyses of these phenotypes include a single dependent variable which is a phenotype labeled as presence or absence of the disease status for

each sample. If we look closer, we observe that these complex phenotypes are a collection of many sub-phenotypes. For example, Li et al. analyzed EHR data to identify patterns among patients and were able to cluster type 2 diabetes patients into three distinct subgroups [191]. Similarly, another study by Verma et al. [192] has shown the importance of incorporating more than one data type from an EHR to identify robust associations. EHRs are highly resourceful as they provide longitudinal information on patients, but utilizing this information is essential and challenging. Longitudinal data tell a patient's history, and careful assessment of this history can be very fruitful in designing preventive and diagnostic measures. Machine learning methods have been proposed to use longitudinal data from EHR. For example, methods by Zhao et al. [193] and Singh et al. [194] address the issues of temporality from an EHR. Clinical laboratory measures that are part of routine care in health systems are also utilized for association analyses. Historically, one set of values (mean, median, or most recent measurement) values are used. However, clinical measures can provide inferences of a disease diagnosis precisely in the situation where one value for each patient might not tell the complete story. Thus, exploring these longitudinal measures in many different ways such as interquartile ranges (IQR), variance (referring to variability in measures), time-series analysis, and extremely low and high values, etc. can be very beneficial in elucidating the etiology of complex traits.

In this era of rapid molecular technological advancements, we expect to see future studies not looking at the relevance of one omics dataset at a time, but instead considering the combinatorial effects (including additive and interactive) of all possible omics datasets. The field is moving in the direction of integrative modeling approaches (such as meta-dimensional analyses) [166], and, thus, along with detection of risk factors, we should also expect the calculations of variance components attributable by additive, interactive, and environmental effects to be included in the same comprehensive models elucidating undercover heritability. Utilizing a-priori biological information for interpretation of results and identifying interconnected networks affecting disease etiology and what tissues are affected also seems necessary [195]. Many association studies of phenotypes such as lipid traits, cancer, Alzheimer's disease, and autism to name a few, have been conducted to identify the impact of different genomic variations (suspects) utilizing several analytic methodologies (weapons) to represent their effect in tissues (where) (see Table 2 for example studies in lipid traits).

Table 2. Examples for use of different methods to exploit genetic architecture of lipid traits.

Analysis Type	References
Common variants	Rotroff et al. [196], Ligthart et al. [197]
Rare Variants	Liu et al. [185], Surakka et al. [198]
Gene–Gene Interactions	Ma et al. [199], De et al. [200], Holzinger et al. [201]
Gene–Environment Interactions	Ordovas [202], Shungin et al. [203]
Gene Expression analysis	Wen et al. [204]
Proteomics	Luczak et al. [205]
Meta-dimensional analysis	Holzinger et al. [206]
Phenotype Heterogeneity	Morabia et al. [207]

However, future studies will likely involve the development of methods to integrate all types of omics variations in identifying models affecting disease traits. Methods should also focus on dealing with the sparsity of data because it is highly unlikely to have all of the different data types variations collected on large enough sample sizes. Still, our challenge will be to initiate each new investigation as we would a new round of the board game “Clue”. We allow for some variations to classic game of “Clue” for one or team of the suspects, weapons, and locations to be the accurate solutions to the crime (disease risk) in question. Another extension to consider in this game is that, in studying the underlying genetic underpinnings of complex traits, we should consider investigating series of crimes and not a single crime. This refers to the concept of phenotypic heterogeneity where many sub-phenotypes (multiple crimes) could be leading to a complex phenotype in investigation. Our ability to explore the different types of omics variation, using different underlying genetic architectures and modeling

methods, will be critical to achieving our maximal understanding of the genomics of common, complex diseases and traits.

Acknowledgments: We would like to thank current and previous members of Ritchie Lab, Sarah Pendergrass, Anurag Verma, Molly Hall, Anastasia Lucas and Alex Frase in discussions about the concept of application of Game of Clue towards investigating complexity of disease traits. This work was supported in part by the following NIH grants: AI116794 and AI077505.

Author Contributions: S.S.V. drafted the manuscript; S.S.V. and M.D.R. revised and edited the manuscript. S.S.V. and M.D.R. created the concept of the manuscript collaboratively.

Conflicts of Interest: The authors declare no conflict of interest. The founding sponsors had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, and in the decision to publish the results.

References

1. Cutting, G.R. Modifier genes in Mendelian disorders: The example of cystic fibrosis. *Ann. N. Y. Acad. Sci.* **2010**, *1214*, 57–69. [[CrossRef](#)] [[PubMed](#)]
2. Afshari, N.A.; Igo, R.P.; Morris, N.J.; Stambolian, D.; Sharma, S.; Pulagam, V.L.; Dunn, S.; Stamler, J.F.; Truitt, B.J.; Rimmler, J.; et al. Genome-wide association study identifies three novel loci in Fuchs endothelial corneal dystrophy. *Nat. Commun.* **2017**, *8*, 14898. [[CrossRef](#)] [[PubMed](#)]
3. Klein, I.; Danzi, S. Thyroid Disease and the Heart. *Circulation* **2007**, *116*, 1725–1735. [[CrossRef](#)] [[PubMed](#)]
4. Fritsche, L.G.; Igl, W.; Bailey, J.N.C.; Grassmann, F.; Sengupta, S.; Bragg-Gresham, J.L.; Burdon, K.P.; Hebbbring, S.J.; Wen, C.; Gorski, M.; et al. A large genome-wide association study of age-related macular degeneration highlights contributions of rare and common variants. *Nat. Genet.* **2016**, *48*, 134–143. [[CrossRef](#)] [[PubMed](#)]
5. Dewey, F.E.; Gusarova, V.; O’Dushlaine, C.; Gottesman, O.; Trejos, J.; Hunt, C.; Van Hout, C.V.; Habegger, L.; Buckler, D.; Lai, K.-M.V.; et al. Inactivating Variants in ANGPTL4 and Risk of Coronary Artery Disease. *N. Engl. J. Med.* **2016**, *374*, 1123–1133. [[CrossRef](#)] [[PubMed](#)]
6. Chu, X.K.; Tuo, J.; Chan, C.-C. Genetics of age-related macular degeneration: Application to drug design. *Future Med. Chem.* **2013**, *5*, 13–15. [[CrossRef](#)] [[PubMed](#)]
7. Everett, B.M.; Smith, R.J.; Hiatt, W.R. Reducing LDL with PCSK9 Inhibitors—The Clinical Benefit of Lipid Drugs. *N. Engl. J. Med.* **2015**, *373*, 1588–1591. [[CrossRef](#)] [[PubMed](#)]
8. McCarty, C.A.; Chisholm, R.L.; Chute, C.G.; Kullo, I.J.; Jarvik, G.P.; Larson, E.B.; Li, R.; Masys, D.R.; Ritchie, M.D.; Roden, D.M.; et al. The eMERGE Network: A consortium of biorepositories linked to electronic medical records data for conducting genomic studies. *BMC Med. Genom.* **2011**, *4*, 13. [[CrossRef](#)] [[PubMed](#)]
9. Carey, D.J.; Fetterolf, S.N.; Davis, F.D.; Faucett, W.A.; Kirchner, H.L.; Mirshahi, U.; Murray, M.F.; Smelser, D.T.; Gerhard, G.S.; Ledbetter, D.H. The Geisinger MyCode community health initiative: An electronic health record-linked biobank for precision medicine research. *Genet. Med.* **2016**, *18*, 906–913. [[CrossRef](#)] [[PubMed](#)]
10. Roden, D.M.; Pulley, J.M.; Basford, M.A.; Bernard, G.R.; Clayton, E.W.; Balser, J.R.; Masys, D.R. Development of a large-scale de-identified DNA biobank to enable personalized medicine. *Clin. Pharmacol. Ther.* **2008**, *84*, 362–369. [[CrossRef](#)] [[PubMed](#)]
11. McQuillan, G.M.; Pan, Q.; Porter, K.S. Consent for genetic research in a general population: An update on the National Health and Nutrition Examination Survey experience. *Genet. Med.* **2006**, *8*, 354–360. [[CrossRef](#)] [[PubMed](#)]
12. Speliotes, E.K.; Willer, C.J.; Berndt, S.I.; Monda, K.L.; Thorleifsson, G.; Jackson, A.U.; Lango Allen, H.; Lindgren, C.M.; Luan, J.; Mägi, R.; et al. Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nat. Genet.* **2010**, *42*, 937–948. [[CrossRef](#)] [[PubMed](#)]
13. Patel, C.J.; Pho, N.; McDuffie, M.; Easton-Marks, J.; Kothari, C.; Kohane, I.S.; Avillach, P. A database of human exposomes and phenomes from the US National Health and Nutrition Examination Survey. *Sci. Data* **2016**, *3*, 160096. [[CrossRef](#)] [[PubMed](#)]
14. Zuk, O.; Hechter, E.; Sunyaev, S.R.; Lander, E.S. The mystery of missing heritability: Genetic interactions create phantom heritability. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 1193–1198. [[CrossRef](#)] [[PubMed](#)]
15. Gibson, G. Hints of hidden heritability in GWAS. *Nat. Genet.* **2010**, *42*, 558–560. [[CrossRef](#)] [[PubMed](#)]

16. Lee, S.H.; Wray, N.R.; Goddard, M.E.; Visscher, P.M. Estimating Missing Heritability for Disease from Genome-wide Association Studies. *Am. J. Hum. Genet.* **2011**, *88*, 294–305. [[CrossRef](#)] [[PubMed](#)]
17. Kerner, B.; North, K.E.; Fallin, M.D. Use of Longitudinal Data in Genetic Studies in the Genome-wide Association Studies Era: Summary of Group 14. *Genet. Epidemiol.* **2009**, *33*, S93–S98. [[CrossRef](#)] [[PubMed](#)]
18. Ioannidis, J.P.A.; Loy, E.Y.; Poulton, R.; Chia, K.S. Researching genetic versus nongenetic determinants of disease: A comparison and proposed unification. *Sci. Transl. Med.* **2009**, *1*, 7ps8. [[CrossRef](#)] [[PubMed](#)]
19. Cheverud, J.M.; Routman, E.J. Epistasis and its contribution to genetic variance components. *Genetics* **1995**, *139*, 1455–1461. [[PubMed](#)]
20. Verma, S.S.; Cooke Bailey, J.N.; Lucas, A.; Bradford, Y.; Linneman, J.G.; Hauser, M.A.; Pasquale, L.R.; Peissig, P.L.; Brilliant, M.H.; McCarty, C.A.; et al. Epistatic Gene-Based Interaction Analyses for Glaucoma in eMERGE and NEIGHBOR Consortium. *PLoS Genet.* **2016**, *12*, e1006186. [[CrossRef](#)] [[PubMed](#)]
21. Gatz, M.; Reynolds, C.A.; Fratiglioni, L.; Johansson, B.; Mortimer, J.A.; Berg, S.; Fiske, A.; Pedersen, N.L. Role of genes and environments for explaining Alzheimer disease. *Arch. Gen. Psychiatry* **2006**, *63*, 168–174. [[CrossRef](#)] [[PubMed](#)]
22. Lord, J.; Lu, A.J.; Cruchaga, C. Identification of rare variants in Alzheimer’s disease. *Front. Genet.* **2014**, *5*. [[CrossRef](#)] [[PubMed](#)]
23. Wu, X.; Gu, J. Heritability of prostate cancer: A tale of rare variants and common single nucleotide polymorphisms. *Ann. Transl. Med.* **2016**, *4*. [[CrossRef](#)] [[PubMed](#)]
24. Connelly, C.F.; Akey, J.M. On the Prospects of Whole-Genome Association Mapping in *Saccharomyces cerevisiae*. *Genetics* **2012**, *191*, 1345–1353. [[CrossRef](#)] [[PubMed](#)]
25. Ivanov, D.K.; Escott-Price, V.; Ziehm, M.; Magwire, M.M.; Mackay, T.F.C.; Partridge, L.; Thornton, J.M. Longevity GWAS Using the Drosophila Genetic Reference Panel. *J. Gerontol. A Biol. Sci. Med. Sci.* **2015**, *70*, 1470–1478. [[CrossRef](#)] [[PubMed](#)]
26. Wangler, M.F.; Hu, Y.; Shulman, J.M. Drosophila and genome-wide association studies: A review and resource for the functional dissection of human complex traits. *Dis. Models Mech.* **2017**, *10*, 77–88. [[CrossRef](#)] [[PubMed](#)]
27. Flint, J.; Mackay, T.F.C. Genetic architecture of quantitative traits in mice, flies, and humans. *Genome Res.* **2009**, *19*, 723–733. [[CrossRef](#)] [[PubMed](#)]
28. Cox, R.D.; Church, C.D. Mouse models and the interpretation of human GWAS in type 2 diabetes and obesity. *Dis. Models Mech.* **2011**, *4*, 155–164. [[CrossRef](#)] [[PubMed](#)]
29. Blake, J.A.; Richardson, J.E.; Bult, C.J.; Kadin, J.A.; Eppig, J.T. MGD: The Mouse Genome Database. *Nucleic Acids Res.* **2003**, *31*, 193–195. [[CrossRef](#)] [[PubMed](#)]
30. Smith, C.L.; Eppig, J.T. The Mammalian Phenotype Ontology: Enabling robust annotation and comparative analysis. *Wiley Interdiscip. Rev. Syst. Biol. Med.* **2009**, *1*, 390–399. [[CrossRef](#)] [[PubMed](#)]
31. Queitsch, C.; Carlson, K.D.; Girirajan, S. Lessons from Model Organisms: Phenotypic Robustness and Missing Heritability in Complex Disease. *PLoS Genet.* **2012**, *8*, e1003041. [[CrossRef](#)] [[PubMed](#)]
32. Silventoinen, K.; Magnusson, P.K.E.; Tynelius, P.; Kaprio, J.; Rasmussen, F. Heritability of body size and muscle strength in young adulthood: A study of one million Swedish men. *Genet. Epidemiol.* **2008**, *32*, 341–349. [[CrossRef](#)] [[PubMed](#)]
33. Cross-Disorder Group of the Psychiatric Genomics Consortium; International Inflammatory Bowel Disease Genetics Consortium (IIBDGC); Lee, S.H.; Ripke, S.; Neale, B.M.; Faraone, S.V.; Purcell, S.M.; Perlis, R.H.; Mowry, B.J.; Thapar, A.; et al. Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nat. Genet.* **2013**, *45*, 984–994. [[CrossRef](#)] [[PubMed](#)]
34. Elks, C.E.; den Hoed, M.; Zhao, J.H.; Sharp, S.J.; Wareham, N.J.; Loos, R.J.F.; Ong, K.K. Variability in the Heritability of Body Mass Index: A Systematic Review and Meta-Regression. *Front. Endocrinol. (Lausanne)* **2012**, *3*. [[CrossRef](#)] [[PubMed](#)]
35. Yang, J.; Bakshi, A.; Zhu, Z.; Hemani, G.; Vinkhuyzen, A.A.E.; Lee, S.H.; Robinson, M.R.; Perry, J.R.B.; Nolte, I.M.; van Vliet-Ostaptchouk, J.V.; et al. Genetic variance estimation with imputed variants finds negligible missing heritability for human height and body mass index. *Nat. Genet.* **2015**, *47*, 1114–1120. [[CrossRef](#)] [[PubMed](#)]
36. Ali, O. Genetics of type 2 diabetes. *World J. Diabetes* **2013**, *4*, 114–123. [[CrossRef](#)] [[PubMed](#)]
37. Cordell, H.J. Detecting gene-gene interactions that underlie human diseases. *Nat. Rev. Genet.* **2009**, *10*, 392–404. [[CrossRef](#)] [[PubMed](#)]

38. Cordell, H.J. Epistasis: What it means, what it doesn't mean, and statistical methods to detect it in humans. *Hum. Mol. Genet.* **2002**, *11*, 2463–2468. [CrossRef] [PubMed]
39. Manolio, T.A.; Collins, F.S.; Cox, N.J.; Goldstein, D.B.; Hindorff, L.A.; Hunter, D.J.; McCarthy, M.I.; Ramos, E.M.; Cardon, L.R.; Chakravarti, A.; et al. Finding the missing heritability of complex diseases. *Nature* **2009**, *461*, 747–753. [CrossRef] [PubMed]
40. Gamazon, E.R.; Huang, R.S.; Dolan, M.E.; Cox, N.J.; Im, H.K. Integrative Genomics: Quantifying Significance of Phenotype-Genotype Relationships from Multiple Sources of High-Throughput Data. *Front. Genet.* **2013**, *3*. [CrossRef] [PubMed]
41. Lewis, R.; Lewis, R. Like a Game of Clue, Genomics Tracks Outbreak, Revealing Evolution in Action. Available online: <https://blogs.scientificamerican.com/guest-blog/like-a-game-of-clue-genomics-tracks-outbreak-revealing-evolution-in-action/> (accessed on 1 November 2017).
42. Clue Emerges in Case of the Missing Heritability. Available online: <https://www.genengnews.com/gen-news-highlights/clue-emerges-in-case-of-the-missing-heritability/81249819> (accessed on 1 November 2017).
43. Cluedo. Available online: <https://en.wikipedia.org/wiki/Cluedo> (accessed on 1 November 2017).
44. MacArthur, J.; Bowler, E.; Cerezo, M.; Gil, L.; Hall, P.; Hastings, E.; Junkins, H.; McMahon, A.; Milano, A.; Morales, J.; et al. The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucleic Acids Res.* **2017**, *45*, D896–D901. [CrossRef] [PubMed]
45. Welter, D.; MacArthur, J.; Morales, J.; Burdett, T.; Hall, P.; Junkins, H.; Klemm, A.; Flicek, P.; Manolio, T.; Hindorff, L.; et al. The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res.* **2014**, *42*, D1001–D1006. [CrossRef] [PubMed]
46. Hindorff, L.A.; Sethupathy, P.; Junkins, H.A.; Ramos, E.M.; Mehta, J.P.; Collins, F.S.; Manolio, T.A. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 9362–9367. [CrossRef] [PubMed]
47. Hirschhorn, J.N. Genomewide Association Studies—Illuminating Biologic Pathways. *N. Engl. J. Med.* **2009**, *360*, 1699–1701. [CrossRef] [PubMed]
48. Schork, N.J.; Murray, S.S.; Frazer, K.A.; Topol, E.J. Common vs. rare allele hypotheses for complex diseases. *Curr. Opin. Genet. Dev.* **2009**, *19*, 212–219. [CrossRef] [PubMed]
49. Franke, A.; McGovern, D.P.B.; Barrett, J.C.; Wang, K.; Radford-Smith, G.L.; Ahmad, T.; Lees, C.W.; Balschun, T.; Lee, J.; Roberts, R.; et al. Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci. *Nat. Genet.* **2010**, *42*, 1118–1125. [CrossRef] [PubMed]
50. Fall, T.; Ingelsson, E. Genome-wide association studies of obesity and metabolic syndrome. *Mol. Cell. Endocrinol.* **2014**, *382*, 740–757. [CrossRef] [PubMed]
51. Billings, L.K.; Florez, J.C. The genetics of type 2 diabetes: What have we learned from GWAS? *Ann. N. Y. Acad. Sci.* **2010**, *1212*, 59–77. [CrossRef] [PubMed]
52. Bashinskaya, V.V.; Kulakova, O.G.; Boyko, A.N.; Favorov, A.V.; Favorova, O.O. A review of genome-wide association studies for multiple sclerosis: Classical and hypothesis-driven approaches. *Hum. Genet.* **2015**, *134*, 1143–1162. [CrossRef] [PubMed]
53. Michailidou, K.; Beesley, J.; Lindstrom, S.; Canisius, S.; Dennis, J.; Lush, M.J.; Maranian, M.J.; Bolla, M.K.; Wang, Q.; Shah, M.; et al. Genome-wide association analysis of more than 120,000 individuals identifies 15 new susceptibility loci for breast cancer. *Nat. Genet.* **2015**, *47*, 373–380. [CrossRef] [PubMed]
54. Church, C.; Lee, S.; Bagg, E.A.L.; McTaggart, J.S.; Deacon, R.; Gerken, T.; Lee, A.; Moir, L.; Mecinović, J.; Quwailid, M.M.; et al. A mouse model for the metabolic effects of the human fat mass and obesity associated FTO gene. *PLoS Genet.* **2009**, *5*, e1000599. [CrossRef] [PubMed]
55. Pennesi, M.E.; Neuringer, M.; Courtney, R.J. Animal models of age related macular degeneration. *Mol. Asp. Med.* **2012**, *33*, 487–509. [CrossRef] [PubMed]
56. Culverhouse, R.; Suarez, B.K.; Lin, J.; Reich, T. A perspective on epistasis: Limits of models displaying no main effect. *Am. J. Hum. Genet.* **2002**, *70*, 461–471. [CrossRef] [PubMed]
57. Moore, J.H. A global view of epistasis. *Nat. Genet.* **2005**, *37*, 13–14. [CrossRef] [PubMed]
58. Moore, J.H.; Williams, S.M. Traversing the conceptual divide between biological and statistical epistasis: Systems biology and a more modern synthesis. *Bioessays* **2005**, *27*, 637–646. [CrossRef] [PubMed]
59. Mackay, T.F.C. Epistasis for quantitative traits in *Drosophila*. *Methods Mol. Biol.* **2015**, *1253*, 47–70. [CrossRef] [PubMed]

60. Mackay, T.F.C. Epistasis and Quantitative Traits: Using Model Organisms to Study Gene-Gene Interactions. *Nat. Rev. Genet.* **2014**, *15*, 22–33. [[CrossRef](#)] [[PubMed](#)]
61. Reedy, J.J.; Cavalier, F.P. Epistasis in eye colors of *Drosophila melanogaster*. *J. Hered.* **1971**, *62*, 131–134. [[CrossRef](#)] [[PubMed](#)]
62. Huang, W.; Richards, S.; Carbone, M.A.; Zhu, D.; Anholt, R.R.H.; Ayroles, J.F.; Duncan, L.; Jordan, K.W.; Lawrence, F.; Magwire, M.M.; et al. Epistasis dominates the genetic architecture of *Drosophila* quantitative traits. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 15553–15559. [[CrossRef](#)] [[PubMed](#)]
63. Sun, X.; Lu, Q.; Mukherjee, S.; Mukheerjee, S.; Crane, P.K.; Elston, R.; Ritchie, M.D. Analysis pipeline for the epistasis search—statistical versus biological filtering. *Front. Genet.* **2014**, *5*, 106. [[CrossRef](#)] [[PubMed](#)]
64. Lee, S.; Abecasis, G.R.; Boehnke, M.; Lin, X. Rare-Variant Association Analysis: Study Designs and Statistical Tests. *Am. J. Hum. Genet.* **2014**, *95*, 5–23. [[CrossRef](#)] [[PubMed](#)]
65. Moutsianas, L.; Agarwala, V.; Fuchsberger, C.; Flannick, J.; Rivas, M.A.; Gaulton, K.J.; Albers, P.K.; Consortium, G.; McVean, G.; Boehnke, M.; et al. The Power of Gene-Based Rare Variant Methods to Detect Disease-Associated Variation and Test Hypotheses About Complex Disease. *PLoS Genet.* **2015**, *11*, e1005165. [[CrossRef](#)] [[PubMed](#)]
66. Moore, C.B.; Wallace, J.R.; Frase, A.T.; Pendergrass, S.A.; Ritchie, M.D. BioBin: A bioinformatics tool for automating the binning of rare variants using publicly available biological knowledge. *BMC Med. Genom.* **2013**, *6* (Suppl. 2). [[CrossRef](#)]
67. Zhu, C.S.; Pinsky, P.F.; Cramer, D.W.; Ransohoff, D.F.; Hartge, P.; Pfeiffer, R.M.; Urban, N.; Mor, G.; Bast, R.C.; Moore, L.E.; et al. A framework for evaluating biomarkers for early detection: Validation of biomarker panels for ovarian cancer. *Cancer Prev. Res. (Phila.)* **2011**, *4*, 375–383. [[CrossRef](#)] [[PubMed](#)]
68. Variant Association Tools. Available online: <http://varianttools.sourceforge.net> (accessed on 10 November 2017).
69. Zhan, X.; Hu, Y.; Li, B.; Abecasis, G.R.; Liu, D.J. RVTESTS: An efficient and comprehensive tool for rare variant association analysis using sequence data: Table 1. *Bioinformatics* **2016**, *32*, 1423–1426. [[CrossRef](#)] [[PubMed](#)]
70. Nejentsev, S.; Walker, N.; Riches, D.; Egholm, M.; Todd, J.A. Rare variants of IFIH1, a gene implicated in antiviral responses, protect against type 1 diabetes. *Science* **2009**, *324*, 387–389. [[CrossRef](#)] [[PubMed](#)]
71. Peterson, A.S.; Fong, L.G.; Young, S.G. PCSK9 function and physiology. *J. Lipid Res.* **2008**, *49*, 1152–1156. [[CrossRef](#)] [[PubMed](#)]
72. Steinberg, D.; Witztum, J.L. Inhibition of PCSK9: A powerful weapon for achieving ideal LDL cholesterol levels. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 9546–9547. [[CrossRef](#)] [[PubMed](#)]
73. Lopez, D. Inhibition of PCSK9 as a novel strategy for the treatment of hypercholesterolemia. *Drug News Perspect.* **2008**, *21*, 323–330. [[CrossRef](#)] [[PubMed](#)]
74. Pettenati, M.J.; Rao, P.N.; Phelan, M.C.; Grass, F.; Rao, K.W.; Cosper, P.; Carroll, A.J.; Elder, F.; Smith, J.L.; Higgins, M.D. Paracentric inversions in humans: A review of 446 paracentric inversions with presentation of 120 new cases. *Am. J. Med. Genet.* **1995**, *55*, 171–187. [[CrossRef](#)] [[PubMed](#)]
75. Mullaney, J.M.; Mills, R.E.; Pittard, W.S.; Devine, S.E. Small insertions and deletions (INDELs) in human genomes. *Hum. Mol. Genet.* **2010**, *19*, R131–R136. [[CrossRef](#)] [[PubMed](#)]
76. Fan, Y.; Wang, W.; Ma, G.; Liang, L.; Shi, Q.; Tao, S. Patterns of Insertion and Deletion in Mammalian Genomes. *Curr. Genom.* **2007**, *8*, 370–378. [[CrossRef](#)]
77. Itsara, A.; Cooper, G.M.; Baker, C.; Girirajan, S.; Li, J.; Absher, D.; Krauss, R.M.; Myers, R.M.; Ridker, P.M.; Chasman, D.I.; et al. Population analysis of large copy number variants and hotspots of human genetic disease. *Am. J. Hum. Genet.* **2009**, *84*, 148–161. [[CrossRef](#)] [[PubMed](#)]
78. Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3000 shared controls. *Nature* **2007**, *447*, 661–678. [[CrossRef](#)]
79. Chung, B.H.-Y.; Tao, V.Q.; Tso, W.W.-Y. Copy number variation and autism: New insights and clinical implications. *J. Formos. Med. Assoc.* **2014**, *113*, 400–408. [[CrossRef](#)] [[PubMed](#)]
80. Patel, C.J.; Bhattacharya, J.; Butte, A.J. An Environment-Wide Association Study (EWAS) on Type 2 Diabetes Mellitus. *PLoS ONE* **2010**, *5*, e10746. [[CrossRef](#)] [[PubMed](#)]
81. Hall, M.A.; Dudek, S.M.; Goodloe, R.; Crawford, D.C.; Pendergrass, S.A.; Peissig, P.; Brilliant, M.; Mccarty, C.A.; Ritchie, M.D. Environment-wide association study (EWAS) for type 2 diabetes in the Marshfield Personalized Medicine Research Project Biobank. In Proceedings of the Pacific Symposium on Biocomputing, Kohala Coast, HI, USA, 3–7 January 2014; pp. 200–211.

82. Ottman, R. Gene–Environment Interaction: Definitions and Study Designs. *Prev. Med.* **1996**, *25*, 764–770. [[CrossRef](#)] [[PubMed](#)]
83. Khoury, M.J. Editorial: Emergence of Gene-Environment Interaction Analysis in Epidemiologic Research. *Am. J. Epidemiol.* **2017**, *186*, 751–752. [[CrossRef](#)] [[PubMed](#)]
84. Ritchie, M.D.; Davis, J.R.; Aschard, H.; Battle, A.; Conti, D.; Du, M.; Eskin, E.; Fallin, M.D.; Hsu, L.; Kraft, P.; et al. Incorporation of Biological Knowledge Into the Study of Gene-Environment Interactions. *Am. J. Epidemiol.* **2017**, *186*, 771–777. [[CrossRef](#)] [[PubMed](#)]
85. Patel, C.J.; Cullen, M.R.; Ioannidis, J.P.; Butte, A.J. Systematic evaluation of environmental factors: Persistent pollutants and nutrients correlated with serum lipid levels. *Int. J. Epidemiol.* **2012**, *41*, 828–843. [[CrossRef](#)] [[PubMed](#)]
86. Thomas, D. Gene–environment-wide association studies: Emerging approaches. *Nat. Rev. Genet.* **2010**, *11*, 259–272. [[CrossRef](#)] [[PubMed](#)]
87. Wright, F.A.; Sullivan, P.F.; Brooks, A.I.; Zou, F.; Sun, W.; Xia, K.; Madar, V.; Jansen, R.; Chung, W.; Zhou, Y.-H.; et al. Heritability and Genomics of Gene Expression in Peripheral Blood. *Nat. Genet.* **2014**, *46*, 430–437. [[CrossRef](#)] [[PubMed](#)]
88. Moffatt, M.F.; Kabesch, M.; Liang, L.; Dixon, A.L.; Strachan, D.; Heath, S.; Depner, M.; von Berg, A.; Bufe, A.; Rietschel, E.; et al. Genetic variants regulating ORMDL3 expression contribute to the risk of childhood asthma. *Nature* **2007**, *448*, 470–473. [[CrossRef](#)] [[PubMed](#)]
89. Zhu, Z.; Zhang, F.; Hu, H.; Bakshi, A.; Robinson, M.R.; Powell, J.E.; Montgomery, G.W.; Goddard, M.E.; Wray, N.R.; Visscher, P.M.; et al. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat. Genet.* **2016**, *48*, 481–487. [[CrossRef](#)] [[PubMed](#)]
90. Barbeira, A.N.; Dickinson, S.P.; Torres, J.M.; Bonazzola, R.; Zheng, J.; Torstenson, E.S.; Wheeler, H.E.; Shah, K.P.; Edwards, T.; Garcia, T.; et al. Exploring the phenotypic consequences of tissue specific gene expression variation inferred from GWAS summary statistics. *bioRxiv* **2017**, 045260. [[CrossRef](#)]
91. Hormozdiari, F.; van de Bunt, M.; Segrè, A.V.; Li, X.; Joo, J.W.J.; Bilow, M.; Sul, J.H.; Sankararaman, S.; Pasaniuc, B.; Eskin, E. Colocalization of GWAS and eQTL Signals Detects Target Genes. *Am. J. Hum. Genet.* **2016**, *99*, 1245–1260. [[CrossRef](#)] [[PubMed](#)]
92. LaCroix, B.; Gamazon, E.R.; Lenkala, D.; Im, H.K.; Geeleher, P.; Ziliak, D.; Cox, N.J.; Huang, R.S. Integrative analyses of genetic variation, epigenetic regulation, and the transcriptome to elucidate the biology of platinum sensitivity. *BMC Genom.* **2014**, *15*, 292. [[CrossRef](#)] [[PubMed](#)]
93. Kim, D.; Li, R.; Dudek, S.M.; Ritchie, M.D. ATHENA: Identifying interactions between different levels of genomic data associated with cancer clinical outcomes using grammatical evolution neural network. *BioData Min.* **2013**, *6*, 23. [[CrossRef](#)] [[PubMed](#)]
94. Holzinger, E.R.; Dudek, S.M.; Frase, A.T.; Pendergrass, S.A.; Ritchie, M.D. ATHENA: The analysis tool for heritable and environmental network associations. *Bioinformatics* **2014**, *30*, 698–705. [[CrossRef](#)] [[PubMed](#)]
95. Gusev, A.; Lee, S.H.; Trynka, G.; Finucane, H.; Vilhjálmsson, B.J.; Xu, H.; Zang, C.; Ripke, S.; Bulik-Sullivan, B.; Stahl, E.; et al. Partitioning Heritability of Regulatory and Cell-Type-Specific Variants across 11 Common Diseases. *Am. J. Hum. Genet.* **2014**, *95*, 535–552. [[CrossRef](#)] [[PubMed](#)]
96. Madu, C.O.; Lu, Y. Novel diagnostic biomarkers for prostate cancer. *J. Cancer* **2010**, *1*, 150–177. [[CrossRef](#)] [[PubMed](#)]
97. Li, D.; Chan, D.W. Proteomic cancer biomarkers from discovery to approval: It’s worth the effort. *Expert Rev. Proteom.* **2014**, *11*, 135–136. [[CrossRef](#)] [[PubMed](#)]
98. Coticchia, C.M.; Yang, J.; Moses, M.A. Ovarian Cancer Biomarkers: Current Options and Future Promise. *J. Natl. Compr. Cancer Netw.* **2008**, *6*, 795–802. [[CrossRef](#)]
99. Mai, P.L.; Wentzensen, N.; Greene, M.H. Challenges related to developing serum-based biomarkers for early ovarian cancer detection. *Cancer Prev. Res. (Phila.)* **2011**, *4*, 303–306. [[CrossRef](#)] [[PubMed](#)]
100. Lee, D.-S.; Park, J.; Kay, K.A.; Christakis, N.A.; Oltvai, Z.N.; Barabási, A.-L. The implications of human metabolic network topology for disease comorbidity. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 9880–9885. [[CrossRef](#)] [[PubMed](#)]
101. Sun, Y.V.; Kardia, S.L.R. Identification of epistatic effects using a protein–protein interaction database. *Hum. Mol. Genet.* **2010**, *19*, 4345–4352. [[CrossRef](#)] [[PubMed](#)]
102. Egger, G.; Liang, G.; Aparicio, A.; Jones, P.A. Epigenetics in human disease and prospects for epigenetic therapy. *Nature* **2004**, *429*, 457–463. [[CrossRef](#)] [[PubMed](#)]

103. Jones, P.A.; Baylin, S.B. The fundamental role of epigenetic events in cancer. *Nat. Rev. Genet.* **2002**, *3*, 415–428. [[CrossRef](#)] [[PubMed](#)]
104. Ballestar, E.; Esteller, M. The epigenetic breakdown of cancer cells: From DNA methylation to histone modifications. *Prog. Mol. Subcell. Biol.* **2005**, *38*, 169–181. [[PubMed](#)]
105. Novak, K. Epigenetics Changes in Cancer Cells. *MedGenMed* **2004**, *6*, 17. [[PubMed](#)]
106. Bonasio, R. The expanding epigenetic landscape of non-model organisms. *J. Exp. Biol.* **2015**, *218*, 114–122. [[CrossRef](#)] [[PubMed](#)]
107. Grunstein, M.; Gasser, S.M. Epigenetics in *Saccharomyces cerevisiae*. *Cold Spring Harb. Perspect. Biol.* **2013**, *5*. [[CrossRef](#)] [[PubMed](#)]
108. Lyko, F.; Beisel, C.; Marhold, J.; Paro, R. Epigenetic regulation in *Drosophila*. *Curr. Top. Microbiol. Immunol.* **2006**, *310*, 23–44. [[PubMed](#)]
109. Weissmann, F.; Muyrers-Chen, I.; Musch, T.; Stach, D.; Wiessler, M.; Paro, R.; Lyko, F. DNA Hypermethylation in *Drosophila melanogaster* Causes Irregular Chromosome Condensation and Dysregulation of Epigenetic Histone Modifications. *Mol. Cell. Biol.* **2003**, *23*, 2577–2586. [[CrossRef](#)] [[PubMed](#)]
110. Zhou, S.; Mackay, T.F.; Anholt, R.R. Transcriptional and epigenetic responses to mating and aging in *Drosophila melanogaster*. *BMC Genom.* **2014**, *15*. [[CrossRef](#)] [[PubMed](#)]
111. Mano, Y.; Kobayashi, T.J.; Nakayama, J.; Uchida, H.; Oki, M. Single Cell Visualization of Yeast Gene Expression Shows Correlation of Epigenetic Switching between Multiple Heterochromatic Regions through Multiple Generations. *PLoS Biol.* **2013**, *11*, e1001601. [[CrossRef](#)] [[PubMed](#)]
112. Omic Tools. Available online: <https://omictools.com> (accessed on 22 January 2018).
113. PLINK2. Available online: <https://www.cog-genomics.org/plink2> (accessed on 22 January 2018).
114. PLATO. Available online: <https://ritchielab.psu.edu/software/plato-download> (accessed on 22 January 2018).
115. QCTOOL. Available online: <http://www.well.ox.ac.uk/~gav/qctool/#overview> (accessed on 22 January 2018).
116. Genabel. Available online: <http://www.genabel.org> (accessed on 22 January 2018).
117. BOLT-LMM. Available online: <https://data.broadinstitute.org/alkesgroup/BOLT-LMM/> (accessed on 22 January 2018).
118. FAST-LMM. Available online: <https://github.com/MicrosoftGenomics/FaST-LMM> (accessed on 22 January 2018).
119. CNVTools. Available online: <http://www.bioconductor.org/packages/release/bioc/html/CNVtools.html> (accessed on 22 January 2018).
120. PennCNV. Available online: <http://penncnv.openbioinformatics.org/en/latest/> (accessed on 22 January 2018).
121. CKAT. Available online: https://works.bepress.com/debashis_ghosh/75/ (accessed on 22 January 2018).
122. ParseCNV. Available online: <http://parsecnv.sourceforge.net> (accessed on 22 January 2018).
123. CNVassoc. Available online: <https://cran.r-project.org/web/packages/CNVassoc/index.html> (accessed on 22 January 2018).
124. RVtests. Available online: <https://genome.sph.umich.edu/wiki/Rvtests> (accessed on 22 January 2018).
125. PLINK/SEQ. Available online: <https://atgu.mgh.harvard.edu/plinkseq/> (accessed on 22 January 2018).
126. EPACTS-Genome Analysis Wiki. Available online: <https://genome.sph.umich.edu/wiki/EPACTS> (accessed on 29 October 2017).
127. MAGMA. Available online: <https://ctg.cncr.nl/software/magma> (accessed on 22 January 2018).
128. EMMAX. Available online: <http://varianttools.sourceforge.net> (accessed on 22 January 2018).
129. MDR. Available online: <https://sourceforge.net/projects/mdr/> (accessed on 22 January 2018).
130. AntEpiSeeker. Available online: <http://nce.ads.uga.edu/~romdhane/AntEpiSeeker/index.html> (accessed on 22 January 2018).
131. MultiSurf. Available online: <https://github.com/EpistasisLab/scikit-rebate/blob/master/skrebate/multisurf.py> (accessed on 22 January 2018).
132. BOOST. Available online: <http://bioinformatics.ust.hk/BOOST.html> (accessed on 22 January 2018).
133. SNPTest. Available online: https://mathgen.stats.ox.ac.uk/genetics_software/snptest/snptest.html (accessed on 22 January 2018).
134. TS-GSIS. Available online: <https://cran.r-project.org/web/packages/TSGSIS/index.html> (accessed on 22 January 2018).
135. SNPAssociation. Available online: <https://cran.r-project.org/web/packages/SNPAssoc/index.html> (accessed on 22 January 2018).

136. CAPE. Available online: <https://cran.r-project.org/web/packages/cape/cape.pdf> (accessed on 22 January 2018).
137. Colak, R.; Kim, T.; Kazan, H.; Oh, Y.; Cruz, M.; Valladares-Salgado, A.; Peralta, J.; Escobedo, J.; Parra, E.J.; Kim, P.M.; et al. JBASE: Joint Bayesian Analysis of Subphenotypes and Epistasis. *Bioinformatics* **2016**, *32*, 203–210. [CrossRef] [PubMed]
138. SMR. Available online: <http://cnsgenomics.com/software/smr/> (accessed on 29 October 2017).
139. EWASher. Available online: <https://www.microsoft.com/en-us/download/details.aspx?id=52501> (accessed on 22 January 2018).
140. LiCHe. Available online: <http://viq854.github.io/lichee/> (accessed on 22 January 2018).
141. BUHMBBOX. Available online: <http://software.broadinstitute.org/mpg/buhmbox/> (accessed on 22 January 2018).
142. ForestPMPlot. Available online: http://genetics.cs.ucla.edu/meta_jemdoc/ (accessed on 22 January 2018).
143. NetDx. Available online: <https://github.com/nrn/GoogleSummerOfCode/issues/70> (accessed on 22 January 2018).
144. BioGranat-IG. Available online: <http://www.thomas-schlitt.net/Biogramat.html> (accessed on 22 January 2018).
145. NETAM. Available online: <http://www.sailing.cs.cmu.edu/main/> (accessed on 22 January 2018).
146. EINVis. Available online: <http://www.robwu.net/einvis/> (accessed on 22 January 2018).
147. NetDecoder. Available online: <http://netdecoder.hms.harvard.edu> (accessed on 22 January 2018).
148. ViSEN. Available online: <https://sourceforge.net/projects/visen/> (accessed on 22 January 2018).
149. Cytoscape. Available online: <http://www.cytoscape.org> (accessed on 22 January 2018).
150. PARIS. Available online: <https://ritchielab.psu.edu/software/paris-download> (accessed on 22 January 2018).
151. SNPsea. Available online: <http://pubs.broadinstitute.org/mpg/snpsea/> (accessed on 22 January 2018).
152. GSEA. Available online: <http://software.broadinstitute.org/gsea/index.jsp> (accessed on 22 January 2018).
153. Vegas2Pathway. Available online: <https://vegas2.qimrberghofer.edu.au> (accessed on 22 January 2018).
154. MAGENTA. Available online: <https://software.broadinstitute.org/mpg/magenta/> (accessed on 22 January 2018).
155. ATHENA. Available online: <https://ritchielab.psu.edu/software/athena-downloads> (accessed on 22 January 2018).
156. iCluster. Available online: <https://www.mskcc.org/departments/epidemiology-biostatistics/biostatistics/icluster> (accessed on 22 January 2018).
157. Biofilter. Available online: <https://ritchielab.psu.edu/software/biofilter-download-1> (accessed on 22 January 2018).
158. SKAT. Available online: <https://www.hsph.harvard.edu/skat/> (accessed on 22 January 2018).
159. Biobin. Available online: <https://ritchielab.psu.edu/software/biobin-download> (accessed on 22 January 2018).
160. GLM. Available online: <https://cran.r-project.org/web/packages/glmnet/index.html> (accessed on 22 January 2018).
161. RANGER. Available online: <https://cran.r-project.org/web/packages/ranger/ranger.pdf> (accessed on 22 January 2018).
162. Gradient Boosting. Available online: <https://cran.r-project.org/web/packages/gbm/index.html> (accessed on 22 January 2018).
163. TATES. Available online: <https://ctg.cncr.nl/software/tates> (accessed on 22 January 2018).
164. CAVIAR. Available online: <http://genetics.cs.ucla.edu/caviar/> (accessed on 22 January 2018).
165. PrediXcan. Available online: <https://github.com/hakyimlab/PrediXcan> (accessed on 22 January 2018).
166. Ritchie, M.D.; Holzinger, E.R.; Li, R.; Pendergrass, S.A.; Kim, D. Methods of integrating data to uncover genotype-phenotype interactions. *Nat. Rev. Genet.* **2015**, *16*, 85–97. [CrossRef] [PubMed]
167. Lonsdale, J.; Thomas, J.; Salvatore, M.; Phillips, R.; Lo, E.; Shad, S.; Hasz, R.; Walters, G.; Garcia, F.; Young, N.; et al. The Genotype-Tissue Expression (GTEx) project. *Nat. Genet.* **2013**, *45*, 580–585. [CrossRef] [PubMed]
168. Braun, T.; Wagner, A.; DeLuca, A.; Casavant, T.; Scheetz, T.; Clark, A.; Mullins, R.; Stone, E. The Ocular Tissue Database. *Investig. Ophthalmol. Vis. Sci.* **2013**, *54*, 3383.
169. Slowikowski, K.; Hu, X.; Raychaudhuri, S. SNPsea: An algorithm to identify cell types, tissues and pathways affected by risk loci. *Bioinformatics* **2014**, *30*, 2496–2497. [CrossRef] [PubMed]

170. Liu, L.; Yin, X.; Wen, L.; Yang, C.; Sheng, Y.; Lin, Y.; Zhu, Z.; Shen, C.; Shi, Y.; Zheng, Y.; et al. Several Critical Cell Types, Tissues, and Pathways Are Implicated in Genome-Wide Association Studies for Systemic Lupus Erythematosus. *G3 (Bethesda)* **2016**, *6*, 1503–1511. [[CrossRef](#)] [[PubMed](#)]
171. Hu, X.; Kim, H.; Stahl, E.; Plenge, R.; Daly, M.; Raychaudhuri, S. Integrating autoimmune risk loci with gene-expression data identifies specific pathogenic immune cell subsets. *Am. J. Hum. Genet.* **2011**, *89*, 496–506. [[CrossRef](#)] [[PubMed](#)]
172. Fadista, J.; Vikman, P.; Laakso, E.O.; Mollet, I.G.; Esguerra, J.L.; Taneera, J.; Storm, P.; Osmark, P.; Ladenvall, C.; Prasad, R.B.; et al. Global genomic and transcriptomic analysis of human pancreatic islets reveals novel genes influencing glucose metabolism. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 13924–13929. [[CrossRef](#)] [[PubMed](#)]
173. Small, K.S.; Hedman, A.K.; Grundberg, E.; Nica, A.C.; Thorleifsson, G.; Kong, A.; Thorsteindottir, U.; Shin, S.-Y.; Richards, H.B.; GIANT Consortium; et al. Identification of an imprinted master trans regulator at the KLF14 locus related to multiple metabolic phenotypes. *Nat. Genet.* **2011**, *43*, 561–564. [[CrossRef](#)] [[PubMed](#)]
174. Grotz, A.K.; Gloyn, A.L.; Thomsen, S.K. Prioritising Causal Genes at Type 2 Diabetes Risk Loci. *Curr. Diabetes Rep.* **2017**, *17*. [[CrossRef](#)] [[PubMed](#)]
175. Yang, J.; Lee, S.H.; Goddard, M.E.; Visscher, P.M. GCTA: A Tool for Genome-wide Complex Trait Analysis. *Am. J. Hum. Genet.* **2011**, *88*, 76–82. [[CrossRef](#)] [[PubMed](#)]
176. Purcell, S.; Neale, B.; Todd-Brown, K.; Thomas, L.; Ferreira, M.A.R.; Bender, D.; Maller, J.; Sklar, P.; de Bakker, P.I.W.; Daly, M.J.; et al. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *Am. J. Hum. Genet.* **2007**, *81*, 559–575. [[CrossRef](#)] [[PubMed](#)]
177. Pérez, P.; de los Campos, G. Genome-Wide Regression and Prediction with the BGLR Statistical Package. *Genetics* **2014**, *198*, 483–495. [[CrossRef](#)] [[PubMed](#)]
178. Speed, D.; Hemani, G.; Johnson, M.R.; Balding, D.J. Improved heritability estimation from genome-wide SNPs. *Am. J. Hum. Genet.* **2012**, *91*, 1011–1021. [[CrossRef](#)] [[PubMed](#)]
179. Gorfine, M.; Berndt, S.I.; Chang-Claude, J.; Hoffmeister, M.; Marchand, L.L.; Potter, J.; Slattery, M.L.; Keret, N.; Peters, U.; Hsu, L. Heritability Estimation using a Regularized Regression Approach (HERRA): Applicable to continuous, dichotomous or age-at-onset outcome. *PLoS ONE* **2017**, *12*, e0181269. [[CrossRef](#)] [[PubMed](#)]
180. Ge, T.; Nichols, T.E.; Lee, P.H.; Holmes, A.J.; Roffman, J.L.; Buckner, R.L.; Sabuncu, M.R.; Smoller, J.W. Massively expedited genome-wide heritability analysis (MEGHA). *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 2479–2484. [[CrossRef](#)] [[PubMed](#)]
181. Locke, A.E.; Kahali, B.; Berndt, S.I.; Justice, A.E.; Pers, T.H.; Day, F.R.; Powell, C.; Vedantam, S.; Buchkovich, M.L.; Yang, J.; et al. Genetic studies of body mass index yield new insights for obesity biology. *Nature* **2015**, *518*, 197–206. [[CrossRef](#)] [[PubMed](#)]
182. Robinson, M.R.; English, G.; Moser, G.; Lloyd-Jones, L.R.; Triplett, M.A.; Zhu, Z.; Nolte, I.M.; van Vliet-Ostapchouk, J.V.; Snieder, H.; The LifeLines Cohort Study; et al. Genotype-covariate interaction effects and the heritability of adult body mass index. *Nat. Genet.* **2017**, *49*, 1174–1181. [[CrossRef](#)] [[PubMed](#)]
183. Speed, D.; Cai, N.; UCLEB Consortium; Johnson, M.R.; Nejentsev, S.; Balding, D.J. Reevaluation of SNP heritability in complex human traits. *Nat. Genet.* **2017**, *49*, 986–992. [[CrossRef](#)] [[PubMed](#)]
184. Finucane, H.K.; Bulik-Sullivan, B.; Gusev, A.; Trynka, G.; Reshef, Y.; Loh, P.-R.; Anttila, V.; Xu, H.; Zang, C.; Farh, K.; et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet.* **2015**, *47*, 1228–1235. [[CrossRef](#)] [[PubMed](#)]
185. Liu, D.J.; Leal, S.M. Estimating Genetic Effects and Quantifying Missing Heritability Explained by Identified Rare-Variant Associations. *Am. J. Hum. Genet.* **2012**, *91*, 585–596. [[CrossRef](#)] [[PubMed](#)]
186. Ronnegard, L.; Shen, X. Genomic prediction and estimation of marker interaction effects. *bioRxiv* **2016**, 038935. [[CrossRef](#)]
187. Pendergrass, S.A.; Brown-Gentry, K.; Dudek, S.; Frase, A.; Torstenson, E.S.; Goodloe, R.; Ambite, J.L.; Avery, C.L.; Buyske, S.; Bůžková, P.; et al. Phenome-wide association study (PheWAS) for detection of pleiotropy within the Population Architecture using Genomics and Epidemiology (PAGE) Network. *PLoS Genet.* **2013**, *9*, e1003087. [[CrossRef](#)] [[PubMed](#)]
188. Daar, E.S.; Tierney, C.; Fischl, M.A.; Sax, P.E.; Mollan, K.; Budhathoki, C.; Godfrey, C.; Jahed, N.C.; Myers, L.; Katzenstein, D.; et al. Atazanavir plus ritonavir or efavirenz as part of a 3-drug regimen for initial treatment of HIV-1. *Ann. Intern. Med.* **2011**, *154*, 445–456. [[CrossRef](#)] [[PubMed](#)]

189. Denny, J.C.; Ritchie, M.D.; Basford, M.A.; Pulley, J.M.; Bastarache, L.; Brown-Gentry, K.; Wang, D.; Masys, D.R.; Roden, D.M.; Crawford, D.C. PheWAS: Demonstrating the feasibility of a phenome-wide scan to discover gene-disease associations. *Bioinformatics* **2010**, *26*, 1205–1210. [[CrossRef](#)] [[PubMed](#)]
190. Verma, A.; Bradford, Y.; Verma, S.S.; Pendergrass, S.A.; Daar, E.S.; Venuto, C.; Morse, G.D.; Ritchie, M.D.; Haas, D.W. Multiphenotype association study of patients randomized to initiate antiretroviral regimens in AIDS Clinical Trials Group protocol A5202. *Pharmacogenet. Genom.* **2017**, *27*, 101–111. [[CrossRef](#)] [[PubMed](#)]
191. Li, L.; Cheng, W.-Y.; Glicksberg, B.S.; Gottesman, O.; Tamler, R.; Chen, R.; Bottinger, E.P.; Dudley, J.T. Identification of type 2 diabetes subgroups through topological analysis of patient similarity. *Sci. Transl. Med.* **2015**, *7*, 311ra174. [[CrossRef](#)] [[PubMed](#)]
192. Verma, A.; Leader, J.B.; Verma, S.S.; Frase, A.; Wallace, J.; Dudek, S.; Lavage, D.R.; Van Hout, C.V.; Dewey, F.E.; Penn, J.; et al. Integrating clinical laboratory measures and icd-9 code diagnoses in phenome-wide association studies. In Proceedings of the Pacific Symposium on Biocomputing, Kohala Coast, HI, USA, 4–8 January 2016; Volume 21, pp. 168–179.
193. Zhao, J.; Papapetrou, P.; Asker, L.; Boström, H. Learning from heterogeneous temporal data in electronic health records. *J. Biomed. Inform.* **2017**, *65*, 105–119. [[CrossRef](#)] [[PubMed](#)]
194. Singh, A.; Nadkarni, G.; Gottesman, O.; Ellis, S.B.; Bottinger, E.P.; Guttag, J.V. Incorporating temporal EHR data in predictive models for risk stratification of renal function deterioration. *J. Biomed. Inform.* **2015**, *53*, 220–228. [[CrossRef](#)] [[PubMed](#)]
195. Magger, O.; Waldman, Y.Y.; Ruppim, E.; Sharan, R. Enhancing the Prioritization of Disease-Causing Genes through Tissue Specific Protein Interaction Networks. *PLoS Comput. Biol.* **2012**, *8*, e1002690. [[CrossRef](#)] [[PubMed](#)]
196. Rotroff, D.M.; Pijut, S.S.; Marvel, S.W.; Jack, J.R.; Havener, T.M.; Pujol, A.; Schluter, A.; Graf, G.A.; Ginsberg, H.N.; Shah, H.S.; et al. Genetic variants in HSD17B3, SMAD3, and IPO11 impact circulating lipids in response to fenofibrate in individuals with type 2 diabetes. *Clin. Pharmacol. Ther.* **2017**. [[CrossRef](#)] [[PubMed](#)]
197. Ligthart, S.; Vaez, A.; Hsu, Y.-H.; Inflammation Working Group of the CHARGE Consortium; PMI-WG-XCP; LifeLines Cohort Study; Stolk, R.; Uitterlinden, A.G.; Hofman, A.; Alizadeh, B.Z.; et al. Bivariate genome-wide association study identifies novel pleiotropic loci for lipids and inflammation. *BMC Genom.* **2016**, *17*, 443. [[CrossRef](#)]
198. Surakka, I.; Horikoshi, M.; Mägi, R.; Sarin, A.-P.; Mahajan, A.; Lagou, V.; Marullo, L.; Ferreira, T.; Miraglio, B.; Timonen, S.; et al. The impact of low-frequency and rare variants on lipid levels. *Nat. Genet.* **2015**, *47*, 589–597. [[CrossRef](#)] [[PubMed](#)]
199. Ma, L.; Brautbar, A.; Boerwinkle, E.; Sing, C.F.; Clark, A.G.; Keinan, A. Knowledge-Driven Analysis Identifies a Gene–Gene Interaction Affecting High-Density Lipoprotein Cholesterol Levels in Multi-Ethnic Populations. *PLoS Genet.* **2012**, *8*, e1002714. [[CrossRef](#)] [[PubMed](#)]
200. De, R.; Verma, S.S.; Drenos, F.; Holzinger, E.R.; Holmes, M.V.; Hall, M.A.; Crosslin, D.R.; Carrell, D.S.; Hakonarson, H.; Jarvik, G.; et al. Identifying gene-gene interactions that are highly associated with Body Mass Index using Quantitative Multifactor Dimensionality Reduction (QMDR). *BioData Min.* **2015**, *8*, 41. [[CrossRef](#)] [[PubMed](#)]
201. Holzinger, E.R.; Verma, S.S.; Moore, C.B.; Hall, M.; De, R.; Gilbert-Diamond, D.; Lanktree, M.B.; Pankratz, N.; Amuzu, A.; Burt, A.; et al. Discovery and replication of SNP-SNP interactions for quantitative lipid traits in over 60,000 individuals. *BioData Min.* **2017**, *10*, 25. [[CrossRef](#)] [[PubMed](#)]
202. Ordovas, J.M. Gene-diet interaction and plasma lipid responses to dietary intervention. *Biochem. Soc. Trans.* **2002**, *30*, 68–73. [[CrossRef](#)] [[PubMed](#)]
203. Shungin, D.; Deng, W.Q.; Varga, T.V.; Luan, J.; Mihailov, E.; Metspalu, A.; Morris, A.P.; Forouhi, N.G.; Lindgren, C.; Magnusson, P.K.E.; et al. Ranking and characterization of established BMI and lipid associated loci as candidates for gene-environment interactions. *PLoS Genet.* **2017**, *13*. [[CrossRef](#)] [[PubMed](#)]
204. Wen, X.; Pique-Regi, R.; Luca, F. Integrating molecular QTL data into genome-wide genetic association analysis: Probabilistic assessment of enrichment and colocalization. *PLoS Genet.* **2017**, *13*, e1006646. [[CrossRef](#)] [[PubMed](#)]
205. Luczak, M.; Formanowicz, D.; Marczak, Ł.; Suszyńska-Zajczyk, J.; Pawliczak, E.; Wanic-Kossowska, M.; Stobiecki, M. iTRAQ-based proteomic analysis of plasma reveals abnormalities in lipid metabolism proteins in chronic kidney disease-related atherosclerosis. *Sci. Rep.* **2016**, *6*, 32511. [[CrossRef](#)] [[PubMed](#)]

206. Holzinger, E.R.; Dudek, S.M.; Frase, A.T.; Krauss, R.M.; Medina, M.W.; Ritchie, M.D. ATHENA: A tool for meta-dimensional analysis applied to genotypes and gene expression data to predict HDL cholesterol levels. In Proceedings of the Pacific Symposium on Biocomputing, Kohala Coast, HI, USA, 3–7 January 2013; pp. 385–396.
207. Morabia, A.; Cayanis, E.; Costanza, M.C.; Ross, B.M.; Flaherty, M.S.; Alvin, G.B.; Das, K.; Gilliam, T.C. Association of extreme blood lipid profile phenotypic variation with 11 reverse cholesterol transport genes and 10 non-genetic cardiovascular disease risk factors. *Hum. Mol. Genet.* **2003**, *12*, 2733–2743. [[CrossRef](#)] [[PubMed](#)]



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).