BRAINIACS
BRAIN IMAGING AND COMPUTING SCIENCES

2023 Volume 4
Issue 2 Edoc GFA4E8812

Fallacies and Pitfalls in Genome-Wide Association Studies*

Julian Hecker, Adam Craig, Andrew Hughes, Julie Neidich, Carl Taswell, Nan Laird†

Abstract

Since the first genome-wide association study (GWAS) identifying variants associated with myocardial infarction was published over 20 years ago, GWASs have emerged as a powerful tool for exploring the genetic basis of complex traits. To date, hundreds of thousands of statistically significant associations have been reported across thousands of human phenotypes. Nevertheless, the design, implementation, and analysis of GWASs remain complex, and the results are easily misinterpreted. Common mistakes include 1) assuming that variants with the strongest statistical associations are causal instead of correlative, 2) believing that associated loci act through nearby genes, and 3) overemphasizing the contribution of individual loci to the total variability of particular traits. Clinical assays have been designed using the results of GWAS that rely on the contribution of such erroneous data interpretations to predict clinical phenotypes, reactions to medications or foods, and/or propensity to develop diseases. The failure to recognize these errors due to fallacies in logical reasoning and statistical inference presents problems for both the scientific community when the wrong targets may be prioritized in future research studies, as well as for communication with the general public when our understanding of the genetic basis of important traits may be misrepresented and overstated. Here, we review statistical data quality, analysis, and meta-analysis, of GWAS results with an emphasis on accurate and reliable interpretation. Placed in the appropriate context, GWASs enable genome-wide discovery of loci associated with diverse traits, but they constitute only a first step towards understanding the biological mechanism(s) underlying the observed associations. Scientific elucidation of these biological mechanisms must be required to establish causality with biochemical and pathophysiological explanations for any putative statistical correlations.

Keywords

Genome-wide association studies (GWAS), correlation-causation fallacy, meta-analysis, random effects model, fixed effects model, population stratification, family-based association studies (FBAS).

Contents

Introduction

Genome-wide association studies (GWASs) aim to identify associations of genetic variants with phenotypes (Visscher et al. 2017). Most commonly, so-called single-nucleotide polymorphisms (SNPs) are considered in GWASs, and each available SNP is tested for association separately. After more than 15 years of GWASs, thousands of genetic associations were reported and partially replicated (Abdellaoui et al. 2023). Examples include identification of a female-specific association between SNPs at the *PAX1* enhancer locus and idiopathic scoliosis (Sharma et al. 2015), linking of *TAF3* to control of corpuscular hemoglobin concentration (Pistis et al. 2013), and discovery of the role of introns of the *FTO* gene in obesity (Smemo et al. 2014; Claussnitzer et al. 2016). Many GWASs focus on so-called complex traits and diseases that are described by a polygenic architecture (Visscher et al. 2017). A trait with a polygenic architecture is influenced by thousands of causal genetic variants with rather small effect sizes (Tam et al. 2019). Examples include asthma, schizophrenia, body mass index, and human height (Vicente et al. 2017; Tam et al. 2019; Yengo et al. 2022).

Consequently, massive efforts by the research community to collect genetic and phenotypic data in large databases, such as the UK Biobank (Bycroft et al. 2018), led sample sizes in GWASs to grow rapidly over the years, enabling the identification of an increasing number of genetic risk loci (Visscher et al. 2017). According to one projection, use of GWASs to inform selection of drug targets and indications could double the number of drug candidates that successfully pass from phase I clinical trials to approval (Nelson et al. 2015). Nowadays, GWASs are considered to be a success story that identified several important genetic factors

*Presented 2023-10-09 with JH slides, NL slides, and JH+NL video at Guardians 2023

†JH and NL affiliated with Harvard University; AH and JN with Washington University St Louis; AC and CT with Brain Health Alliance; CT and NL contributed as senior co-authors; correspondence to J Hecker at Harvard.

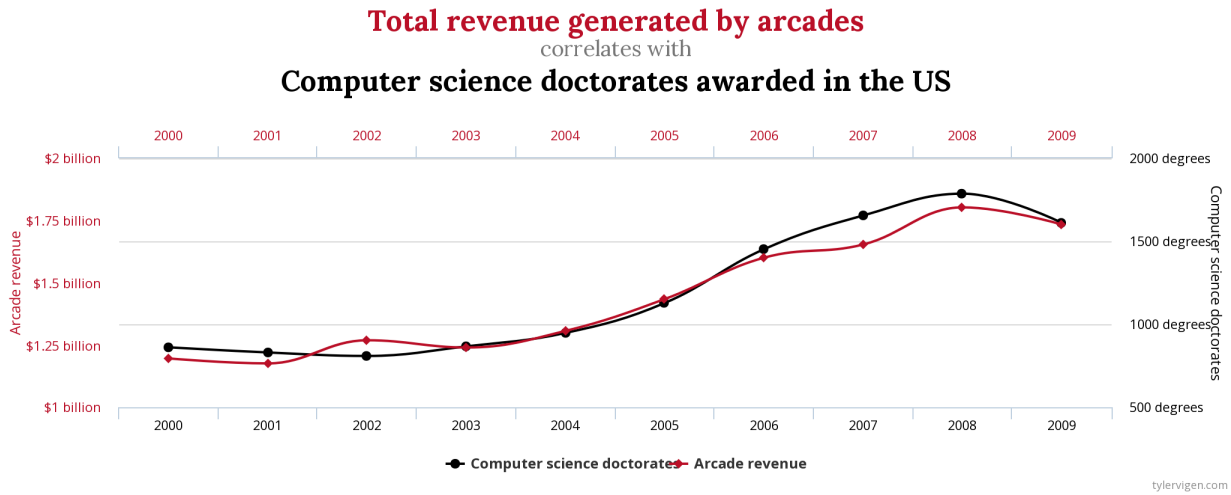

Figure 1: Correlation does not prove causation; see Vigen (2023) for this and other examples of the correlation-causation fallacy.

of complex diseases, but GWASs also face challenges, pitfalls, and limitations that we would like to discuss and review here.

Multiple Testing

As described above, each available SNP is tested for association with the phenotype of consideration. The density of available SNPs in a GWAS depends on the underlying platform used. Over time, the number of genes tested in a single GWAS has grown, starting from microarrays of a few thousand followed by genetic imputations (predicting genetic information based on reference panels), and expanding until, at present, whole-exome and whole-genome sequencing datasets are common (Tam et al. 2019). This implies that a typical GWAS incorporates more than a million common SNPs (DerSimonian and Laird 2015), leading to a substantial multiple testing burden. The established significance level for so-called genome-wide significance in a GWAS is p=5e-08 (Tam et al. 2019). This significance level corresponds to a nominal level of 0.05 corrected for 1 million independent statistical tests by the Bonferroni correction (Tam et al. 2019). More specific gene-based tests, such as the versatile gene-based association study (VEGAS) methodology, can arrive at appropriately corrected *p*-values by using statistics on common genetic variation and the corresponding linkage disequilibrium (LD) structures in reference panels such as the HapMap project (Hecker et al. 2017). LD describes nonrandom association of alleles at different loci resulting from complex interactions between recombination, mutation, selection, and genetic drift (Slatkin 2008). It provides the key patterns of information on which statistical methods of fine-scale gene mapping rely (Slatkin 2008). A pitfall in replication attempts is the winner's curse (Zhong and Prentice 2010). The winner's curse describes the phenomenon that the effect sizes of genetic variants that just passed the genome-wide significance level tend to be overestimated (Zhong and Prentice 2010). This in turn, leads to overestimated power calculations in replication GWASs and was one of the driving factors for the lack of replication in early GWASs (Zhong and Prentice 2010). A wide variety of statistical correction methods can partially account for the winner's curse effect, with empirical Bayesian, FDR inverse quantile transformation, and bootstrap resampling methods outperforming commonly used conditonal likelihood methods (Forde et al. 2023).

Linkage Disequilibrium

Moreover, since nearby genetic variants are often in LD, the genetic information for local SNPs is usually correlated (Tam et al. 2019; Lappalainen and MacArthur 2021). Consequently, the presence of a causal genetic variant resulting in a significant association with the phenotype also leads to a substantial number of significant associations for nearby SNPs (Lappalainen and MacArthur 2021). Therefore, GWASs typically report associated genetic risk loci that contain these multiple associations. A common pitfall is that a genetic variant with a genome-wide significant association p-value is interpreted to be causal, although it potentially only tags a causal variant through LD (Visscher et al. 2017). This is an example of the causation-correlation fallacy.

Fine-mapping is an approach to tackle this problem (Schaid et al. 2018). Fine-mapping prioritizes a set of genetic variants that most likely contains the causal variant at this genetic risk locus, often assuming the presence of at most one causal SNP (Schaid et al. 2018). This procedure incorporates LD information and the individual SNP association statistics (Schaid et al. 2018). Interestingly, there are potential scenarios with multiple genetic effects within the loci in which the most significant SNP is not causal, especially when the statistical power of the study is low (Schaid et al. 2018). Recent fine-mapping approaches are based on Bayesian computations (Schaid et al. 2018; Tam et al. 2019). This purely statistical fine-mapping can be improved by incorporating external biological information such as functional annotations. This external information can be integrated into the prior distributions of the Bayesian models and therefore guides these analyses.

Study Design

Another challenge in GWAS concerns the selected study design. The design of a GWAS with regard to how the subjects are selected for participation can impact the results (Heid, Huth, et al. 2009). Two time-honored epidemiological designs for studying causality in the absence of randomization are the case-control study and the cohort study. The case-control design has been very successful in GWAS (Clarke et al. 2011). Likewise, many GWAS have taken advantage of existing cohorts, such as the Nurse's Health Study, provided that the phenotypes of interest can be obtained. An alternative approach chooses subjects as part of a random sample from a population. Random samples from

selected populations are generally difficult and expensive to obtain, but may be available in some countries as a part of ongoing research programs, for example, the Framingham Heart Study in the United States (Dawber 1980) and the KORA Study in Germany (Heid, Vollmert, et al. 2005).

Another paradigm developed for GWAS involves the use of genomic repositories or biobanks. The idea is to provide access to very large data sets. Genotypes are recorded in a central repository without regard to the phenotypic status of subjects. Selection bias may make it difficult to interpret the results of such studies. Even more importantly, allele frequencies vary with genetic ancestries (Derks et al. 2022). If phenotypic differences correlate with genetic ancestry in the study population, often because of specific participant sampling procedures, genetic association testing based on this phenotype can lead to false positive findings if not appropriately controlled for genetic ancestry (Derks et al. 2022). Furthermore, predictive models of quantitative traits based on data from a single ancestry group can generalize poorly to other populations, as with a predictive model of height found to account for 45% of variation in European populations but only 14-24% in others (Yengo et al. 2022).

The most established approach to adjust for genetic ancestry and therefore reducing the likelihood for false positive associations is to include principal components of genetic ancestry derived from genome-wide data as covariates in the statistical association tests (Price et al. 2006). However, this approach is not guaranteed to fully adjust for ancestry-induced signals, as even large data sets, such as that of the 1000 Genomes Project, may not have sufficient coverage of genetic diversity in some populations (D. Lu and Xu 2013), and the urge to increase sample sizes in recent GWAS can amplify this issue.

An approach to address the issue of population stratification is a family-based study design (Rabinowitz and Laird 2000; Tam et al. 2019). Here cases, or affected individuals, and their family members (ideally their parents) are chosen to be study participants. The family members serve as the controls. Family-based study designs allow genetic association tests for SNPs that are robust to population stratification (Derks et al. 2022). They are particularly useful for observing segregation of rare variants with very large effect sizes when those variants segregate within a family (Visscher et al. 2017).

The classical example is the transmission disequilibrium test (TDT) (Schaid 1998). The TDT considers affected offspring trios and tests the observed allele transmissions against Mendelian expectations (Schaid 1998). Since the test statistic conditions on parental genotypes, the test does not require any assumptions about the underlying allele frequencies and distributions (Ewens and Spielman 1995). This concept was extended to general pedigrees, general phenotypes, and groups of genetic variants in the Family-Based Association Test framework by Laird and Lange (2006).

Meta-Analyses

To achieve the desired large sample sizes, researchers combine their association results in meta-analyses across cohorts and studies (Abdellaoui et al. 2023; Mikolajewicz and Komarova 2019; Steel et al. 2021). Since meta-analyses can achieve the same results by combining summary statistics as with individual-level data (DerSimonian and Laird 2015), they also have the advantage that data resources can be combined without sharing protected individual genetic information across institutions and scientific groups. Several consortia of researchers and institutions have formed to pool data and set standards for studies to be included in metanalyses relevant to particular areas of health and wellness, including the Psychiatric Genomics Consortium, the Genetic Investigation of Anthropometric Traits (GIANT) Consortium, and the Global Lipids Genetics Consortium (O'Donovan 2015; H. Park et al. 2016; Klarin et al. 2018). The approaches to meta-analyses include fixed effects and random effects models (Steel et al. 2021). The latter explicitly allows for heterogeneity in the data (DerSimonian and Laird 2015; Steel et al. 2021). The combination of association results across studies with varying genetic ancestry has the advantage that the differences in the LD structure can lead to an improved resolution in the fine-mapping step of genetic risk loci since the LD effects dilute (DerSimonian and Laird 2015). However, careful interpretation is required. Since most GWAS so far were based on participants of European genetic ancestry, the analysis of other genetic ancestries and ethnicities has great potential to reveal a refined picture of genetic associations and generalizability across populations (Derks et al. 2022). Such meta-analyses rely on accurate and detailed metadata to ensure that results across different studies are comparable (Mikolajewicz and Komarova 2019; Steel et al. 2021). The NHGRI-EBI GWAS Catalog represents one attempt to compile such data in an online repository on a large scale (Sollis et al. 2022).

From GWAS to Biology

Even though GWAS publications have reported thousands of genetic associations with a plethora of complex diseases and traits, and recent advances in fine-mapping in combination with large sample sizes have pinpointed genetic variants with potential causal associations, the underlying biological mechanisms of these associations remain largely unknown. This is because the exact regulatory function of most GWAS hits, which are often located in non-coding regions of the genome, is poorly understood (Abdellaoui et al. 2023; Aguet et al. 2023). Therefore, the role of a SNP and its downstream effects on other genes and pathways is often unknown. Projects such as the Encyclopedia of DNA Elements (ENCODE) are working to fill this knowledge gap by compiling an extensive repository of millions of human and mouse functional elements, including protein-coding genes, regulatory RNA-coding genes, and non-coding regions with known mechanistic functions, such as promoters and enhancers (Moore et al. 2020). Similarly, the GENCODE project publishes extensive annotations of the human and mouse genomes, including protein-coding genes, pseudogenes, and long non-coding RNA genes (Frankish et al. 2020). Using such gene annotations enables new approaches to weighting the significance of association scores based on this prior knowledge in addition to LD and Bonferroni correction (Kichaev et al. 2019). The candidate causal gene is commonly inferred based on the smallest physical distance, but recent investigations showed that this might be misleading. One possibility to gain further insights into the identified genetic associations is to study molecular quantitative trait loci (QTLs) (Lappalainen and MacArthur 2021; Aguet et al. 2023). These SNPs are associated with molecular phenotypes such as RNA expression, DNA methylation, or metabolite levels (Aguet et al. 2023).

Colocalization analyses attempt to test if GWAS findings colocalize with both molecular and expression QTLs (i.e., the same genetic variant is implicated), and such successful colocalizations provide the basis for mediation hypotheses (Rheenen et al. 2021). A systematic version of this concept, a post-GWAS transcriptome-wide association study (TWAS), tests for associations between traits and gene expression levels imputed from eQTLs across the entire genome, while a proteome-wide

Figure 2: Mixed measures? When are meta-analyses reproducible and valid? (New Cuyama sign image by Gogulski 2007.)

association study (PWAS) tests for associations with protein abundance as predicted from population-level protein QTL (pQTL) data (Gedik et al. 2023). By studying the downstream effects of genetic variants, these approaches can identify genes that affect health via differences in quantitatively measured expression traits better correlated with phenotypes even when the direct association between the genotypes and phenotypes otherwise would be weak (Gedik et al. 2023).

One approach that has built on this idea further is the use of colocalization in conjunction with similarity of annotations from single-cell gene expression, protein-protein interaction, and pathway participation features to compute a polygenic priority score to identify associations between non-coding loci and protein-coding genes that are likely to be causal (Weeks et al. 2023). As noted in (Weeks et al. 2023), combining such similarity-based methods with complementary locus-based methods can achieve better results than either one can alone. Taking that reasoning even further, (Gazal et al. 2022) propose a framework for arriving at combinations SNP-to-gene strategies and apply it to select seven such strategies that together achieve higher recall than attainable with any one strategy alone. Ultimately, GWAS alone cannot determine the causal mechanisms behind human health and diversity, which requires taking the next step of analyzing the GWAS-identified candidate genes through both statistical and bench-based functional testing (Gallagher and Chen-Plotkin 2018).

Direct-to-Consumer Testing

Some privately held laboratories, especially those offering direct-to-consumer testing (US Food & Drug Administration 2019; Malgorzata et al. 2021), have used GWAS data for the interpretation of genomic tests for a variety of diverse indications including fear of heights to cat allergy to anxiety to dietary advice. These kinds of indications often have vague symptoms, little evidence of heritability, or are common disorders that may have a multifactorial pattern of inheritance without a specific genotype-phenotype association. The commercial labs often describe the tests as available for personal amusement and not for diagnosis of any specific condition. For the FDA-approved assays that also run at these labs, the reported results are not based on GWAS data. For example, some offer FDA-approved tests for pathogenic variants in the genes associated with increased risk of the development of breast or other cancers (National Human Genome Research Institute 2023). Studies have identified ethical and legal concerns with the DTC testing modality (Martins et al. 2022; Panacer 2023), and specifically with the use of polygenic risk scores that are based solely on GWAS data (J. K.

Park and C. Y. Lu 2023).

Conclusion

Genome-wide association studies identified thousands of genetic associations with a wide range of phenotypes. As a consequence of the polygenic architecture of complex traits and diseases, recent GWASs reached sample sizes of 1 million samples to identify novel genetic risk loci. However, the interpretation of GWAS results requires careful consideration. Technical artefacts such as population stratification can introduce false positive findings in GWAS and identified genetic associations should always be replicated in independent studies. A significant GWAS signal does not imply causality and the identification of causal genetic variants within a genetic risk locus remains a challenge. Furthermore, most GWAS hits are in non-coding regions of the genome and mapping genetic associations to candidate genes for functional follow-up analyses is non-trivial and of limited success so far. Overall, the underlying mechanisms of genetic associations remain poorly understood and there is a risk of overinterpreting their individual relevance in clinical risk prediction and other complex traits such as educational attainment (Okbay et al. 2022; Cesarini and Visscher 2017). While there is great potential in utilizing the findings from GWAS to support the development of new drugs and approach the reality of personalized medicine based on individual risk evaluation, the application of GWAS as a research tool comes with ethical and social responsibility.

Citation

Brainiacs 2023 Volume 4 Issue 2 Edoc GFA4E8812
Title: "Fallacies and Pitfalls in Genome-Wide Association Studies"
Authors: Julian Hecker, Adam Craig, Andrew Hughes, Julie Neidich, Carl Taswell, Nan Laird
Dates: created 2023-10-01, received 2023-10-03, presented 2023-10-09, updated 2023-12-21, published 2023-12-21, endorsed 2023-12-30

Contact: J Hecker at Harvard
URL: BrainiacsJournal.org/arc/pub/Hecker2023FPGWAS
PDP: /Nexus/Brainiacs/Hecker2023FPGWAS
DOI: /10.48085/GFA4E8812

References

[1] A. Abdellaoui, L. Yengo, K. J. Verweij, and P. M. Visscher. "15 years of GWAS discovery: Realizing the promise." *The American Journal of Human Genetics* 110.2 (Feb. 2023), pp. 179–194. ISSN: 0002-9297. DOI: 10.1016/j.ajhg.2022.12.011 (cited pp. 1, 3).

[2] F. Aguet, K. Alasoo, Y. I. Li, A. Battle, H. K. Im, S. B. Montgomery, and T. Lappalainen. "Molecular quantitative trait loci." *Nature Reviews Methods Primers* 3.1 (Jan. 2023). ISSN: 2662-8449. DOI: 10.1038/s43586-022-00188-6 (cited p. 3).

[3] C. Bycroft, C. Freeman, D. Petkova, G. Band, et al. "The UK Biobank resource with deep phenotyping and genomic data." *Nature* 562.7726 (Oct. 2018), pp. 203–209. ISSN: 1476-4687. DOI: 10.1038/s41586-018-0579-z (cited p. 1).

[4] D. Cesarini and P. M. Visscher. "Genetics and educational attainment." *npj Science of Learning* 2.1 (Feb. 2017). ISSN: 2056-7936. DOI: 10.1038/s41539-017-0005-6 (cited p. 5).

[5] G. M. Clarke, C. A. Anderson, F. H. Pettersson, L. R. Cardon, A. P. Morris, and K. T. Zondervan. "Basic statistical analysis in genetic case-control studies." *Nature Protocols* 6.2 (Feb. 2011), pp. 121–133. ISSN: 1750-2799. DOI: 10.1038/nprot.2010.182 (cited p. 2).

[6] M. Claussnitzer, S. N. Dankel, K.-H. Kim, G. Quon, W. Meuleman, et al. "FTO Obesity Variant Circuitry and Adipocyte Browning in Humans." *New England Journal of Medicine* 374.2 (Jan. 2016), pp. 190–193. ISSN: 1533-4406. DOI: 10.1056/nejmc1513316 (cited p. 1).

[7] T. R. Dawber. *The Framingham Study. The Epidemiology of Atherosclerotic Disease.* Harvard University Press, 1980. ISBN: 9780674492080. DOI: 10.4159/harvard.9780674492097 (cited p. 3).

[8] E. M. Derks, J. G. Thorp, and Z. F. Gerring. "Ten challenges for clinical translation in psychiatric genetics." *Nature Genetics* 54.10 (Sept. 2022), pp. 1457–1465. ISSN: 1546-1718. DOI: 10.1038/s41588-022-01174-0 (cited p. 3).

[9] R. DerSimonian and N. Laird. "Meta-analysis in clinical trials revisited." *Contemporary Clinical Trials* 45 (Nov. 2015), pp. 139–145. ISSN: 1551-7144. DOI: 10.1016/j.cct.2015.09.002 (cited pp. 2, 3).

[10] W. J. Ewens and R. S. Spielman. "The transmission/disequilibrium test: history, subdivision, and admixture." *American journal of human genetics* 57.2 (1995), p. 455 (cited p. 3).

[11] A. Forde, G. Hemani, and J. Ferguson. "Review and further developments in statistical corrections for Winner's Curse in genetic association studies." *PLoS Genetics* 19.9 (2023), e1010546 (cited p. 2).

[12] A. Frankish, M. Diekhans, I. Jungreis, J. Lagarde, et al. "GENCODE 2021." *Nucleic Acids Research* 49.D1 (Dec. 2020), pp. D916–D923. ISSN: 1362-4962. DOI: 10.1093/nar/gkaa1087 (cited p. 3).

[13] M. D. Gallagher and A. S. Chen-Plotkin. "The Post-GWAS Era: From Association to Function." *The American Journal of Human Genetics* 102.5 (May 2018), pp. 717–730. ISSN: 0002-9297. DOI: 10.1016/j.ajhg.2018.04.002 (cited p. 4).

[14] S. Gazal, O. Weissbrod, F. Hormozdiari, K. K. Dey, et al. "Combining SNP-to-gene linking strategies to identify disease genes and assess disease omnigenicity." *Nature Genetics* 54.6 (June 2022), pp. 827–836. ISSN: 1546-1718. DOI: 10.1038/s41588-022-01087-y (cited p. 4).

[15] H. Gedik, R. E. Peterson, B. P. Riley, V. I. Vladimirov, and S.-A. Bacanu. "Integrative Post-Genome-Wide Association Study Analyses Relevant to Psychiatric Disorders: Imputing Transcriptome and Proteome Signals." *Complex Psychiatry* 9.1-4 (2023), pp. 130–144 (cited p. 4).

[16] M. Gogulski. *Image of sign at 4923 Primero St, New Cuyama, CA 93254*. Ed. by Wikipedia. Aug. 4, 2007. URL: https://en.wikipedia.org/wiki/New_Cuyama,_California#/media/File:US-CA,_New_cuyama.jpg (cited p. 4).

[17] J. Hecker, A. Maaser, D. Prokopenko, H. L. Fier, and C. Lange. "Reporting Correct p Values in VEGAS Analyses." *Twin Research and Human Genetics* 20.3 (Mar. 2017), pp. 257–259. ISSN: 1839-2628. DOI: 10.1017/thg.2017.16 (cited p. 2).

[18] I. M. Heid, C. Vollmert, A. Hinney, A. Döring, et al. "Association of the 103I MC4R allele with decreased body mass in 7937 participants of two population based surveys." *Journal of Medical Genetics* 42.4 (Apr. 2005), e21. ISSN: 1468-6244. DOI: 10.1136/jmg.2004.027011 (cited p. 3).

[19] I. M. Heid, C. Huth, R. J. F. Loos, F. Kronenberg, et al. "Meta-Analysis of the INSIG2 Association with Obesity Including 74,345 Individuals: Does Heterogeneity of Estimates Relate to Study Design?" *PLoS Genetics* 5.10 (Oct. 2009). Ed. by D. B. Allison, e1000694. ISSN: 1553-7404. DOI: 10.1371/journal.pgen.1000694 (cited p. 2).

[20] G. Kichaev, G. Bhatia, P.-R. Loh, S. Gazal, et al. "Leveraging Polygenic Functional Enrichment to Improve GWAS Power." *The American Journal of Human Genetics* 104.1 (Jan. 2019), pp. 65–75. ISSN: 0002-9297. DOI: 10.1016/j.ajhg.2018.11.008 (cited p. 3).

[21] D. Klarin, S. M. Damrauer, K. Cho, Y. V. Sun, et al. "Genetics of blood lipids among~ 300,000 multi-ethnic participants of the Million Veteran Program." *Nature genetics* 50.11 (2018), pp. 1514–1523 (cited p. 3).

[22] N. Laird and C. Lange. "Family-based designs in the age of large-scale gene-association studies." *Nature Reviews Genetics* 7.5 (May 2006), pp. 385–394. ISSN: 1471-0064. DOI: 10.1038/nrg1839 (cited p. 3).

[23] T. Lappalainen and D. G. MacArthur. "From variant to function in human disease genetics." *Science* 373.6562 (Sept. 2021), pp. 1464–1468. ISSN: 1095-9203. DOI: 10.1126/science.abi8207 (cited pp. 2, 3).

[24] D. Lu and S. Xu. "Principal component analysis reveals the 1000 Genomes Project does not sufficiently cover the human genetic diversity in Asia." *Frontiers in Genetics* 4 (2013). ISSN: 1664-8021. DOI: 10.3389/fgene.2013.00127 (cited p. 3).

[25] M. Malgorzata, S. Maria, and W. Michał. "Genetic testing—whether to allow complete freedom? Direct to consumer tests versus genetic tests for medical purposes." *Journal of Applied Genetics* 63.1 (Nov. 2021), pp. 119–126. ISSN: 2190-3883. DOI: 10.1007/s13353-021-00670-z (cited p. 4).

[26] M. F. Martins, L. T. Murry, L. Telford, and F. Moriarty. "Direct-to-consumer genetic testing: an updated systematic review of healthcare professionals' knowledge and views, and ethical and legal concerns." *European Journal of Human Genetics* 30.12 (Oct. 2022), pp. 1331–1343. ISSN: 1476-5438. DOI: 10.1038/s41431-022-01205-8 (cited p. 4).

[27] N. Mikolajewicz and S. V. Komarova. "Meta-Analytic Methodology for Basic Research: A Practical Guide." *Frontiers in Physiology* 10 (Mar. 2019). ISSN: 1664-042X. DOI: 10.3389/fphys.2019.00203 (cited p. 3).

[28] J. E. Moore, M. J. Purcaro, H. E. Pratt, C. B. Epstein, et al. "Expanded encyclopaedias of DNA elements in the human and mouse genomes." *Nature* 583.7818 (2020), pp. 699–710 (cited p. 3).

[29] National Human Genome Research Institute. *Direct-to-Consumer Genetic Testing FAQ for Healthcare Professionals*. June 14, 2023. URL: https://www.genome.gov/For-Health-Professionals/Provider-Genomics-Education-Resources/Healthcare-Provider-Direct-to-Consumer-Genetic-Testing-FAQ (cited p. 4).

[30] M. R. Nelson, H. Tipney, J. L. Painter, J. Shen, et al. "The support of human genetic evidence for approved drug indications." *Nature Genetics* 47.8 (June 2015), pp. 856–860. ISSN: 1546-1718. DOI: 10.1038/ng.3314 (cited p. 1).

[31] M. C. O'Donovan. "What have we learned from the Psychiatric Genomics Consortium." *World Psychiatry* 14.3 (2015), p. 291 (cited p. 3).

[32] A. Okbay, Y. Wu, N. Wang, H. Jayashankar, et al. "Polygenic prediction of educational attainment within and between families from genome-wide association analyses in 3 million individuals." *Nature Genetics* 54.4 (Mar. 2022), pp. 437–449. ISSN: 1546-1718. DOI: 10.1038/s41588-022-01016-z (cited p. 5).

[33] K. S. Panacer. "Ethical Issues Associated With Direct-to-Consumer Genetic Testing." *Cureus* (June 2023). ISSN: 2168-8184. DOI: 10.7759/cureus.39918 (cited p. 4).

[34] H. Park, X. Li, Y. E. Song, K. Y. He, and X. Zhu. "Multivariate analysis of anthropometric traits using summary statistics of genome-wide association studies from GIANT Consortium." *PloS one* 11.10 (2016), e0163912 (cited p. 3).

[35] J. K. Park and C. Y. Lu. "Polygenic Scores in the Direct-to-Consumer Setting: Challenges and Opportunities for a New Era in Consumer Genetic Testing." *Journal of Personalized Medicine* 13.4 (Mar. 2023), p. 573. ISSN: 2075-4426. DOI: 10.3390/jpm13040573 (cited p. 4).

[36] G. Pistis, S. U. Okonkwo, M. Traglia, C. Sala, et al. "Genome Wide Association Analysis of a Founder Population Identified TAF3 as a Gene for MCHC in Humans." *PLoS ONE* 8.7 (July 2013). Ed. by F. Di Cunto, e69206. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0069206 (cited p. 1).

[37] A. L. Price, N. J. Patterson, R. M. Plenge, M. E. Weinblatt, N. A. Shadick, and D. Reich. "Principal components analysis corrects for stratification in genome-wide association studies." *Nature genetics* 38.8 (2006), pp. 904–909 (cited p. 3).

[38] D. Rabinowitz and N. Laird. "A Unified Approach to Adjusting Association Tests for Population Admixture with Arbitrary Pedigree Structure and Arbitrary Missing Marker Information." *Human Heredity* 50.4 (2000), pp. 211–223. ISSN: 1423-0062. DOI: 10.1159/000022918 (cited p. 3).

[39] W. van Rheenen, R. A. A. van der Spek, M. K. Bakker, J. J. F. A. van Vugt, et al. "Common and rare variant association analyses in amyotrophic lateral sclerosis identify 15 risk loci with distinct genetic architectures and neuron-specific biology." *Nature Genetics* 53.12 (Dec. 2021), pp. 1636–1648. ISSN: 1546-1718. DOI: 10.1038/s41588-021-00973-1 (cited p. 3).

[40] D. J. Schaid. "Transmission disequilibrium, family controls, and great expectations." *The American Journal of Human Genetics* 63.4 (1998), pp. 935–941 (cited p. 3).

[41] D. J. Schaid, W. Chen, and N. B. Larson. "From genome-wide associations to candidate causal variants by statistical fine-mapping." *Nature Reviews Genetics* 19.8 (May 2018), pp. 491–504. ISSN: 1471-0064. DOI: 10.1038/s41576-018-0016-z (cited p. 2).

[42] S. Sharma, D. Londono, W. L. Eckalbar, X. Gao, et al. "A PAX1 enhancer locus is associated with susceptibility to idiopathic scoliosis in females." *Nature Communications* 6.1 (Mar. 2015). ISSN: 2041-1723. DOI: 10.1038/ncomms7452 (cited p. 1).

[43] M. Slatkin. "Linkage disequilibrium — understanding the evolutionary past and mapping the medical future." *Nature Reviews Genetics* 9.6 (June 2008), pp. 477–485. ISSN: 1471-0064. DOI: 10.1038/nrg2361 (cited p. 2).

[44] S. Smemo, J. J. Tena, K.-H. Kim, E. R. Gamazon, et al. "Obesity-associated variants within FTO form long-range functional connections with IRX3." *Nature* 507.7492 (Mar. 2014), pp. 371–375. ISSN: 1476-4687. DOI: 10.1038/nature13138 (cited p. 1).

[45] E. Sollis, A. Mosaku, A. Abid, A. Buniello, et al. "The NHGRI-EBI GWAS Catalog: knowledgebase and deposition resource." *Nucleic Acids Research* 51.D1 (Nov. 2022), pp. D977–D985. ISSN: 1362-4962. DOI: 10.1093/nar/gkac1010 (cited p. 3).

[46] P. Steel, S. Beugelsdijk, and H. Aguinis. "The anatomy of an award-winning meta-analysis: Recommendations for authors, reviewers, and readers of meta-analytic reviews." *Journal of International Business Studies* 52.1 (Jan. 2021), pp. 23–44. ISSN: 1478-6990. DOI: 10.1057/s41267-020-00385-z (cited p. 3).

[47] V. Tam, N. Patel, M. Turcotte, Y. Bossé, G. Paré, and D. Meyre. "Benefits and limitations of genome-wide association studies." *Nature Reviews Genetics* 20.8 (May 2019), pp. 467–484. ISSN: 1471-0064. DOI: 10.1038/s41576-019-0127-1 (cited pp. 1–3).

[48] US Food & Drug Administration. *Direct-to-Consumer Tests*. Dec. 20, 2019. URL: https://www.fda.gov/medical-devices/in-vitro-diagnostics/direct-consumer-tests (cited p. 4).

[49] C. T. Vicente, J. A. Revez, and M. A. Ferreira. "Lessons from ten years of genome-wide association studies of asthma." *Clinical & translational immunology* 6.12 (2017), e165 (cited p. 1).

[50] T. Vigen. *Spurious Correlations*. 2023. URL: https://tylervigen.com/spurious-correlations (cited p. 2).

[51] P. M. Visscher, N. R. Wray, Q. Zhang, P. Sklar, M. I. McCarthy, M. A. Brown, and J. Yang. "10 Years of GWAS Discovery: Biology, Function, and Translation." *The American Journal of Human Genetics* 101.1 (July 2017), pp. 5–22. ISSN: 0002-9297. DOI: 10.1016/j.ajhg.2017.06.005 (cited pp. 1–3).

[52] E. M. Weeks, J. C. Ulirsch, N. Y. Cheng, B. L. Trippe, et al. "Leveraging polygenic enrichments of gene features to predict genes underlying complex traits and diseases." *Nature Genetics* 55.8 (July 2023), pp. 1267–1276. ISSN: 1546-1718. DOI: 10.1038/s41588-023-01443-6 (cited p. 4).

[53] L. Yengo, S. Vedantam, E. Marouli, J. Sidorenko, et al. "A saturated map of common genetic variants associated with human height." *Nature* 610.7933 (Oct. 2022), pp. 704–712. ISSN: 1476-4687. DOI: 10.1038/s41586-022-05275-y (cited pp. 1, 3).

[54] H. Zhong and R. L. Prentice. "Correcting "winner's curse" in odds ratios from genomewide association findings for major complex human diseases." *Genetic Epidemiology: The Official Publication of the International Genetic Epidemiology Society* 34.1 (2010), pp. 78–91 (cited p. 2).

