



Reflections on the Field of Human Genetics: A Call for Increased Disease Genetics Theory

Steven J. Schrodj^{1,2*}

¹ Marshfield Clinic Research Foundation, Center for Human Genetics, Marshfield, WI, USA, ² Computation and Informatics in Biology and Medicine, University of Wisconsin-Madison, Madison, WI, USA

OPEN ACCESS

Edited by:

Eli Stahl,
Icahn School of Medicine at
Mount Sinai, USA

Reviewed by:

Jian Li,
Tulane University, USA
Guo-Bo Chen,
Evergreen Landscape & Architecture
Studio, China

*Correspondence:

Steven J. Schrodj
schrodj.steven@mcrf.mfldclin.edu

Specialty section:

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

Received: 26 February 2016

Accepted: 25 May 2016

Published: 08 June 2016

Citation:

Schrodj SJ (2016) Reflections on the
Field of Human Genetics: A Call for
Increased Disease Genetics Theory.
Front. Genet. 7:106.
doi: 10.3389/fgene.2016.00106

Development of human genetics theoretical models and the integration of those models with experiment and statistical evaluation are critical for scientific progress. This perspective argues that increased effort in disease genetics theory, complementing experimental, and statistical efforts, will escalate the unraveling of molecular etiologies of complex diseases. In particular, the development of new, realistic disease genetics models will help elucidate complex disease pathogenesis, and the predicted patterns in genetic data made by these models will enable the concurrent, more comprehensive statistical testing of multiple aspects of disease genetics predictions, thereby better identifying disease loci. By theoretical human genetics, I intend to encompass all investigations devoted to modeling the heritable architecture underlying disease traits and studies of the resulting principles and dynamics of such models. Hence, the scope of theoretical disease genetics work includes construction and analysis of models describing how disease-predisposing alleles (1) arise, (2) are transmitted across families and populations, and (3) interact with other risk and protective alleles across both the genome and environmental factors to produce disease states. Theoretical work improves insight into viable genetic models of diseases consistent with empirical results from linkage, transmission, and association studies as well as population genetics. Furthermore, understanding the patterns of genetic data expected under realistic disease models will enable more powerful approaches to discover disease-predisposing alleles and additional heritable factors important in common diseases. In spite of the pivotal role of disease genetics theory, such investigation is not particularly vibrant.

Keywords: disease genetics, theoretical model, human genetics, GWAS (genome-wide association study), complex diseases, statistical genetics and genomics

INTRODUCTION

Development of human genetics theoretical models and the integration of those models with experiment and statistical evaluation are critical for scientific progress. This perspective argues that increased effort in disease genetics theory, complementing experimental, and statistical efforts, will escalate the unraveling of molecular etiologies of complex diseases. In particular, the development of new, realistic disease genetics models will help elucidate complex disease pathogenesis, and the predicted patterns in genetic data made by these models will enable the concurrent, more comprehensive statistical testing of multiple aspects of disease genetics predictions, thereby better

identifying disease loci. By theoretical human genetics, I intend to encompass all investigations devoted to modeling the heritable architecture underlying disease traits and studies of the resulting principles and dynamics of such models. Hence, the scope of theoretical disease genetics work includes construction and analysis of models describing how disease-predisposing alleles (1) arise, (2) are transmitted across families and populations, and (3) interact with other risk and protective alleles across both the genome and environmental factors to produce disease states. Theoretical work improves insight into viable genetic models of diseases consistent with empirical results from linkage, transmission, and association studies as well as population genetics. Furthermore, understanding the patterns of genetic data expected under realistic disease models will enable more powerful approaches to discover disease-predisposing alleles and additional heritable factors important in common diseases. In spite of the pivotal role of disease genetics theory, such investigation is not particularly vibrant. Currently, activities in human disease genetics are primarily centered upon large-scale empirical studies and, to a lesser extent, statistical methods, with limited contribution to theory.

BACKGROUND AND FRAMEWORK

Broadly speaking, scientific progress is predicated on a robust interplay between three activities: (1) empirical experimentation and observation, (2) the development of theoretical models and extraction of predicted patterns thereof, and (3) the statistical evaluation of the probabilistic correspondence between the predicted patterns and empirical data. Highly impactful discoveries can certainly occur in the absence of formalization of these activities, but these three aspects are nonetheless critical. To exemplify, consider the relatively recent remarkable finding of complex, low-level admixture between modern humans and archaic humans (Green et al., 2010; Reich et al., 2010; Gronau et al., 2011; Li and Durbin, 2011; Sankararaman et al., 2016). This discovery was made through heroic efforts to isolate, sequence, assemble, and align archaic DNA from Neanderthal and Denisovan remains. In parallel, predictions of genetic architecture, divergence patterns, and shared chromosomal regions from admixture models were developed using both molecular phylogenetics and population genetics theory involving mutation, genetic drift, migration, and demographics. Lastly, correspondence between the observed genetic data and theoretical predictions were accomplished through a variety of likelihood-based, Bayesian, and Fisherian approaches. It is not overreaching to claim that this advance of our scientific knowledge hinged on careful empirical observations/experiments, the development of population genetics theory, and the formal evaluation of rich theoretical predictions against observed data through rigorous statistical methods. The most casual of observers will note legions of additional examples of this paradigm from a diverse set of scientific fields such as particle physics (Glashow, 1961; Higgs, 1964; Weinberg, 1967), mechanics (Einstein, 1916; Schrodinger, 1926), enzyme kinetics (Michaelis and Menten,

1913), semiconductors (Hall, 1879; Wilson, 1931; Mott, 1938; Schottky, 1938), atomic chemistry (Hund, 1926; Mulliken, 1932; Huckel, 1934), classical genetics (Mendel, 1866; Fisher, 1918), heredity and evolution (Fisher, 1930; Wright, 1932; Price, 1970), population genetics (Hardy, 1908; Weinberg, 1908; Hudson, 1982, 1983; Kingman, 1982a; Gillespie, 1993, 2000), and predator-prey ecology (Lotka, 1925). Across these and many other fields, theoretical work is a dynamic component of the scientific process: unexplained empirical phenomena motivate new theory, often in a relatively seamless manner, and, conversely, predicted patterns stemming from mechanistic models are digested by experimenters and promptly tested, leading to an expeditious and efficient expansion in our understanding of these phenomena.

Comparatively, disease gene mapping has grown a relatively barren landscape of theory. Largely motivated by the desire to impact clinical practice, human disease genetics has historically been a highly pragmatic field where technological advancement in genotyping and sequencing has spurred large-scale studies and the bulk of quantitative work has focused on statistical methods of analysis, rather than a more equitable partition of statistics and theory. This has continued despite instances where profound shifts in approaches have been driven by insights from theory. For example, an extended stagnation in mapping common, complex diseases was fractured by theoretical developments in the mid- to late-1990s showing that high density genotyping using population-based samples would have dramatically increased power to detect high frequency disease-predisposing alleles of moderate effect sizes, motivating the GWAS paradigm from an overly-simplistic disease genetics model (Kaplan et al., 1995; Risch and Merikangas, 1996; Long et al., 1997; Xiong and Guo, 1998; Kruglyak, 1999; Long and Langley, 1999).

Although not commonplace, there are other historical examples of disease genetics theory driving accelerated progress in human genetics, including the heterozygote selective advantage theory of malaria and sickle-cell disease (Allison, 1954), that complex diseases do have a heritable component (Steinberg et al., 1951; Pickering, 1978; Debray et al., 1979; Kendler and Diehl, 1993; DeBraekeleer, 1991; Lynn et al., 1995; Stein et al., 2005), that large multiplex families were ideal for linkage studies of diseases under Mendelian disease models (Thompson, 1978; Botstein et al., 1980), and the more diffuse impact of theoretical ideas from population genetics such as allele frequency spectra being highly skewed toward very rare alleles (Ewens, 1972; Watterson, 1975; Slatkin and Rannala, 1997; Long and Langley, 1999; Eyre-Walker, 2010; Hudson, 2015)—accentuated in rapidly expanding populations, haplotypes exhibiting block-like structure in LD patterns (Hill and Robertson, 1968; Hudson and Kaplan, 1985; Nothnagel et al., 2002; Wiuf and Posada, 2003), population bottlenecks followed by expansion accentuates allelic dominance effects on fitness (Balick et al., 2015), and alleles with deleterious effects on fitness originating more recently than those neutral with respect to fitness given the same allele frequency (Maruyama, 1974; Zieun et al., 2013). Theoretical work has been done on applying the highly polygenic, additive model to common diseases, offering some testable predictions (Yang et al., 2010, 2011a; Vinkhuyzen

et al., 2013; Loh et al., 2015). Additional, useful efforts have focused on widespread epistatic interactions (Hodge, 1981; Neuman and Rice, 1992; Majewski et al., 2001; Zuk et al., 2012). However, competing theoretical models of common disease genetics are sparse even though both history and reason argue for a more vigorous theoretician community and heightened interaction between theory, experiment and statistical methods.

WHAT CONSTITUTES A USEFUL THEORY OF DISEASE GENETICS?

Although many areas of investigation rightly fall under this rubric, the general focus should be the development of testable models that describe the set of heritable factors and interactions that generate disease states. This includes allele and genotype frequencies, numbers of susceptibility loci, numbers and types of susceptibility alleles, penetrances, epistatic interactions, properties of familial transmission, and effect modification with environmental variables. It is important to draw a distinction between the genetics that predispose an individual to a disease and the genetic repertoire that underlies the predisposition to a disease across a population of affected individuals, for we do not know the extent in which each individual's disease etiology is unique for complex disease phenotypes. That is, although it is well established from population-level analyses that complex diseases are polygenic, we currently have little evidence that definitively speaks to the level of allelic and locus heterogeneity in any complex disease. What set of genotypes at a set of loci are sufficient to generate disease in an individual? What is the variation in these disease-predisposing sets of genotypes and loci across diseased individuals? Do the alleles co-segregate with disease states across relatedness structures? Coherent, useful theories of disease genetics must address these questions.

Theoretical population genetics, with numerous practitioners using coalescent theory and similar tools to study the maintenance of alleles in populations and elucidate the evolutionary forces responsible for genetic variation, is relatively advanced (Kaplan et al., 1988; Hudson, 1991; Charlesworth et al., 1997; Calafell et al., 2001). As population genetics is concerned with the dynamics and distributions of alleles in populations (Ewens, 1972; Moran, 1975; Watterson, 1975; Kingman, 1982a,b; Charlesworth and Jain, 2014; Greenbaum, 2015), the relevance of population genetics theory to disease gene mapping—particularly for case-control association studies, fine-scale mapping, and population stratification—is undeniably clear, with several important advances demonstrating applicability (Pritchard et al., 2000; Morris et al., 2002; Molitor et al., 2003; Burkett et al., 2014). However, population genetics theory, in and of itself, is inadequate to serve as a complete theory for modeling the disease genetics: (1) coalescent theory is largely concerned with samples of random chromosomes from a population, rather than from disease-affected individuals; (2) there is limited focus on the treatment of related individuals; (3) whereas population genetics aims to delineate the relative impact of natural selection, genetic drift, mutation and demographic effects, diseases have a complex, enigmatic relationship to fitness—some diseases may

result from mutation-selection balance, other diseases may carry susceptibility genes that are neutral with respect to selection, while some disease genes may be subjected to directional selection, and many diseases, such as type 2 diabetes (Hu, 2011), may result from a shift to a modern environment; (4) theoretical population genetics concentrates on the dynamics of individual loci in isolation; and (5) somatic mutations and heritable epigenetic factors, which play important roles in at least some common diseases, are often not the subject of mainstream population genetics theory.

Similarly, the theoretical models from quantitative genetics are also problematic in their direct applicability to investigations of disease genetics architecture. These models are almost exclusively direct derivatives of the infinitely polygenic, miniscule additive effects model (IPMAE model) (Falconer and MacKay, 1996; Frank, 2011). Often, when applied to a dichotomous outcome, a threshold (Wright, 1934) or liability function (Falconer, 1965; Curnow and Smith, 1972) is overlaid on the IPMAE model. Historically, work on the IPMAE model was designed for the study of quantitative traits, such as livestock lean body weight and crop yield, in agriculturally important organisms and specifically-designed pedigrees to assess measures such as breeding values (Falconer and MacKay, 1996; Lynch and Walsh, 1998). Although this model carries utility for analysis of quantitative traits in general populations, and the application to human disease is strongly argued by some (Hill et al., 2008; Plomin et al., 2009), whether or not the coupling of a liability function with the IPMAE model is indeed the appropriate model of allelic architecture for any dichotomous complex disease is currently unknown. Many, if not most disease physiologies are fundamentally different than naturally-occurring phenotype variation investigated by quantitative geneticists, and it is reasonable to assume that their underlying allelic architecture also differs. Recently, several have strongly argued against the continued use of the IPMAE model for the purpose of dissecting complex diseases (Nelson et al., 2013; Génin and Clerget-Darpoux, 2015). Although I personally favor models other than the IPMAE model for complex diseases, I do not think that either theoretical nor empirical evidence is currently sufficient to completely dismiss the IPMAE model. It is certainly possible, *a priori*, that tens or hundreds of thousands loci across the genome harbor alleles of very small effect sizes, all marginally contributing to additively increase disease risk. Moreover, many types of models may appear to have additive and nearly independent effects as those effect sizes become small. If complex diseases are a conglomeration of distinct physiological entities with their own genetic etiologies, erroneously aggregated by physicians, it appears possible that the IPMAE model may be reasonable, at least for interpreting data from population-based studies. If molecular networks are highly redundant and numerous pathogenic changes are necessary to compromise the function of these networks, then the IPMAE model might be appropriate. So, rather than disbanding the IPMAE model entirely, a prudent direction would be encouraging the development of alternative theoretical models. A competitive marketplace of disease genetics models is a critically important cog in the unraveling the genetic architecture of all diseases. Certainly, the correspondence

between IPMAE predictions and experimental data will be the ultimate arbitrator. Empirically, the jury is mixed with some studies offering moderate evidence of consistency between genetic association data and the IPMAE model (Yang et al., 2010, 2011a; Vinkhuyzen et al., 2013; Bulik-Sullivan et al., 2015; Loh et al., 2015), while others do not (Ritchie et al., 2001; Kirino et al., 2013; Ridge et al., 2013; Fritsche et al., 2014), and familial data has yet to definitively support or refute the model. Of note, a useful global measure of the magnitude of polygenic inheritance has been discussed by Yang et al. (2011b). Interestingly, testing polygenetic architecture models on GWAS data for four complex diseases—rheumatoid arthritis, celiac disease, myocardial infarction/coronary artery disease, and type 2 diabetes—using Bayesian Approximate Computation, Stahl and colleagues estimated the joint density of the number of independent disease-predisposing SNPs and the liability-scale variance explained showing consistency with models using roughly 2000 SNPs (Stahl et al., 2012).

Within human genetics, the majority of models of common disease genetic architecture used in practice fall into two overly-simplistic camps: (1) monogenic and two-locus models with a biallelic markers and typically one of four classical modes of inheritance (fully dominant, fully recessive, additive, or multiplicative), and (2) direct derivatives of the IPMAE model. One only has to go as far as to look at commonly-used power calculators for genetic association or linkage studies to observe this rather ubiquitous, long-standing, yet fairly impotent state of affairs. Parametric linkage studies using monogenic models produced spurious results for complex diseases (Génin and Clerget-Darpoux, 2015). Much of their use has resulted from convenience—both the monogenic/two-locus models and the IPMAE model are mathematically tractable and other, more realistic models may necessitate complex mathematical treatment or computational approaches. Not only do these two classes of models represent the ends of a wide spectrum of models, but this limited number of disease genetics models is symptomatic of an anemic theoretical effort. That said, what we have learned about the properties and dynamics of the IPMAE (Blangero et al., 2013; Zhou et al., 2013) and monogenic/two-locus models (Li and Reich, 2000; Zaykin et al., 2006; Schrodi et al., 2007; Zaykin and Shibata, 2008) will serve us well for the development of the next generation of theoretical disease genetics models. For example, the finite, additive polygenic model relaxes from the extremely large number of disease loci assumption of the IPMAE (Cannings et al., 1978; Lange, 1997). Further, new statistical approaches, explicitly harnessing theoretical models of polygenic inheritance to better understand genetic variation of complex traits are starting to be developed (Zhou et al., 2013). In my view, finite rare allele models of moderately high effect sizes, high allelic, and high locus heterogeneity with effect modification by genetic background deserve attention. Importantly, very recent results from simulations appear to favor incomplete recessivity models for complex trait etiologies, demonstrating consistency with both realistic population genetic models, heritability data, and GWAS findings (Sanjak et al., 2016). Such work suggests prioritizing tests of recessive modes

of inheritance and compound heterozygosity testing for common disease mapping.

While it is undeniable that substantial biological insights and clinical utility have resulted from identifying alleles truly associated/linked with complex diseases (Sabbagh and Darlu, 2006; Roychowdhury and Chinnaiyan, 2013; Bottini and Peterson, 2014; Kavanaugh et al., 2014; Everett et al., 2015; Lueck et al., 2015), we are currently in the infancy of understanding disease genetics where prediction of any common, complex disease is not yet clinically practicable (Schrodi et al., 2014), and efficacious, highly targeted therapies are sparse. One bright point for disease prediction, borrowed from quantitative genetics and work on highly polygenic additive models, is the use of best linear unbiased prediction (BLUP) (Speed and Balding, 2014; Vilhjalmsón et al., 2015). That said, the overall lack of realistic theoretical models dramatically hinders our progress, for powerful experimental designs and analysis techniques could be optimized to suit the predictions of such models. Yet the tools are available to make significant theoretical inroads. Data processing approaches (Fan et al., 2014) and machine learning has become incorporated into development of genetic models and their evaluation (Libbrecht and Noble, 2015). Graphical modeling programming software are well-developed (Hall et al., 2009). In addition, the investigation and use of causal models may also advance human genetics theory (Pearl, 2000; Madsen et al., 2011a,b). Fast Markov-chain-Monte-Carlo algorithms to screen complicated, vast parameter spaces are accessible. And, most importantly, the accumulated results from multiplex linkage studies, affected sibling pair studies, studies assessing disease concordance between relative pairs of varying relatedness, twin studies, family-based transmission/disequilibrium studies, GWAS, and familial and population-based sequencing studies are available. Moreover, high-throughput genotyping and sequencing have painted a detailed picture of the raw materials from which disease genetics are sampled: the allele frequency spectrum and LD patterns. Disease genetics models must be consistent with these results. Ideally, an abundant assortment of viable theoretical disease genetics models will be developed, generating informative, distinguishing predictions. These predictions can then be tested against the accumulated patterns of genetic data, producing posterior probabilities, or likelihoods for each model. As the empirical data accumulates, the posterior probability density across the parameter space of the models will indicate those models with reasonably high posterior probabilities, with many models being ruled out. Not only would such work illuminate plausible etiological models of complex diseases, but it would suggest highly-powered experimental designs and statistical methods. In particular, with the determination of likely disease genetics models, one could harness a variety of predicted patterns to improve the discovery and assessment of casual loci.

A more detailed example may provide additional weight and clarity to this argument. Consider a standard common disease case/control GWAS study. With notable exceptions of issues such as clustering of subjects using dimensional reduction methods (Price et al., 2006), most such studies are designed and

analyzed to solely test the simple hypothesis of independence between disease status and genotype frequencies at single sites. However, formal disease genetics models could offer a wealth of predictions concerning genetic architecture patterns in the data: (1) Diseases with early onset and probable ancestral effects on fitness predict selection against disease-predisposing alleles which would generate departures from neutrality as measured by metrics such as Tajima's (1989). (2) Departures from Hardy-Weinberg Equilibrium differ between cases and controls under several disease models (Nielsen et al., 1999). (3) The linkage disequilibrium patterns within cases at the susceptibility locus are expected to differ from those patterns observed in controls (Zaykin et al., 2006; Schrodi et al., 2007; Pan, 2010). (4) The decay of disease association with declining linkage disequilibrium between a causal site and closely-linked markers follows a particular form (Lai et al., 1994; Pritchard and Przeworski, 2001; Garcia et al., 2008; Schrodi et al., 2009; Maadooliat et al., 2016). (5) Cases are expected to exhibit increased sharing of chromosomal segments compared to controls (Houwen et al., 1994; Te Meerman et al., 1995; Browning and Thompson, 2012). (6) Models generating allelic heterogeneity such as the rare allele/large effect (RALE) model suggest investigating multiple predisposing sequence variants segregating at each gene/functional motif (Personal communication with Ray White, 2000-2010; Terwilliger and Göring, 2000; Pritchard, 2001; Thornton et al., 2013) and perhaps testing for linkage. Little imagination is necessary to presume that additional, highly useful predicted genetic patterns exist under disease genetics models, hereto underutilized. So long as the disease genetics model is sufficiently accurate or the predictions are robust across models, concurrently testing the rich panoply of theoretical predictions extracts increased information, enabling more refined, credible, and localized discovery of pathogenic alleles. Notably, Agarwala and colleagues have conducted excellent work in this area for type 2 diabetes (Agarwala et al., 2013). They have used a combination of simulation results and results from affected sibling linkage studies, GWAS, a polygenic score logistic regression, and sequencing studies to reduce the model space of possible architecture models. They foresee further reduction in the space of possible models being dependent on the findings from very large-scale sequencing studies. I applaud this considerable effort and hope that further work in this area is strongly supported.

There are similar implications for such theoretical disease genetics when applied to family-based studies. Predictions from realistic disease genetics models enable the coherent exploration of critical questions such as (1) Is the distribution of chromosomal regions shared by affected individuals indicative of disease loci? (2) Are the observed phenotypic variance within families and familial aggregation patterns consistent with specific disease genetics models? (3) Are the transmission patterns within families consistent with disease genetics models? And (4) Given a specific disease genetics model, what is the optimal size of family structure for finding chromosomal regions linked and associated with complex diseases (i.e., siblings, multiplex families, founder populations, or general populations)? Just as with population-based studies, the development of theoretical

models of disease genetics illuminates the path to jointly testing numerous observed genetic patterns within familial structures. Wray and Goddard have started to explore some of these issues and have shown that three disease models are roughly consistent with data on disease risk in relatives (Wray and Goddard, 2010).

RELATED AREAS OF THEORY DEVELOPMENT

One area of active research that would profit from the advancement of disease genetics theory is the development of fine mapping methods to identify causal variants within a disease-associated region. Discovering causal variants is critically important for several reasons, most notably that expensive, time-consuming, follow-up laboratory experiments are predicated on which gene or functional motif are indicted by the genetic evidence. This problem of identifying disease-causing variants in regions of often complex linkage disequilibrium patterns and allelic heterogeneity, although dramatically understudied in the past, has now been increasingly recognized as being vital in the human genetics toolbox. One example is the fine-mapping method of Maller and colleagues which uses a ranked set of Bayes factors (one for each polymorphism in an associated region) (Wellcome Trust Case Control Consortium et al., 2012). Other approaches include Bim-Bam (Servin and Stephens, 2007), CAVIAR (Hormozdiari et al., 2014), CAVIARBF (Chen et al., 2015), coalescent-based methods (Graham, 1998; Morris et al., 2002; Zollner and Pritchard, 2005), and PAINTOR (Kichaev et al., 2014), which incorporates functional information probabilistically. While I applaud these excellent, thoughtful methods, as a generalization these approaches are statistically appropriate and easily interpretable, but use overly simplistic disease genetics models. If we held a more complete understanding of the theoretical properties of alleles that underlie complex diseases and their correlation patterns with linked variants, one could incorporate this information into more powerful fine mapping methods.

Aside from the need for the development and analysis of DNA-based models of disease predisposition, similar models of other heritable factors involved in pathogenesis such as inherited RNA pools, histone acetylation, and DNA methylation effects are essential for the rapid advancement of disease genetics. It is becoming increasingly clear that epigenetic factors play a role in heritable diseases (Uddin et al., 2010; Williams et al., 2010; Allum et al., 2015; Montano et al., 2016). However, just as clear is the near absence of theoretical models describing epigenetics as an etiological factor in diseases. Several simple questions need investigation: What are the probabilistic laws that govern the transmission of these epigenetic factors? That is, what is the distribution of probabilities that a given epigenetic state is transmitted to a subsequent generation? How do these probabilities attenuate across multiple generations? What are the frequencies of various epigenetic changes and the corresponding effects on disease risk? What is the fraction of an individual's disease risk that is generated by epigenetic changes? And how is this fraction distributed across a population? Development

of these theoretical models will allow for the calculation of testable predictions and aid in the construction of more powerful experimental designs.

CONCLUSION

To be balanced, there certainly are efforts in human genetics theory, several of which have been discussed, enabling statistical methods that harness informative theoretical predictions (Reich and Lander, 2001; Zhu et al., 2015). The crux of the argument made here, however, is one of degree: Additional training of human geneticists in theoretical models, complementing, and motivating statistical methods, would be beneficial. Additional construction of disease genetics models to evaluate against empirical data is vitally needed. Additional work on identifying the patterns of genetic data expected under theoretical models is essential. Additional evaluation of empirical data from large multiplex families, affected sibling pairs, isolated populations, founder populations, transmission-based tests, GWAS, whole genome/exome sequencing studies in families, and population-based sequencing studies to determine which disease models are supported and which can be excluded based on these results would be highly productive. And additional interplay between theory, experiment, and analysis is critical. My central thesis is that funding and effort must be balanced in a way that produces complementarily functioning triad of theory, experiment, and statistics, so that the entire field of human genetics moves forward unabated. Currently, experimental studies and statistical methods are clearly active and highly-functioning subfields, whereas, the field has a dearth of theoretical disease genetics models, impeding the entire disease gene mapping enterprise. The timing is ideal for the institution of these changes. We have amassed vast amounts of genetic data for many hundreds of common diseases and rarer, related conditions and yet the heritable causes of each common disease remain poorly understood. Colossal, expensive shifts in focus have been historically driven by simplistic, undeveloped theoretical models, e.g., common disease/common variant hypothesis (Reich and

Lander, 2001), so it may be more fruitful for our field to further develop resources and capabilities that generate more fully developed theoretical models of disease genetics. Perhaps it is time for a new field of theoretical disease genetics.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and approved it for publication.

FUNDING

This work was supported by generous donors to the Marshfield Clinic Research Foundation, a pilot grant award from the NIH-NCATS/University of Wisconsin-Madison Institute for Clinical and Translational Research (UL1TR000427) and NIMH RO1MH097464. The content is solely the responsibility of the author and does not necessarily represent the official views of the National Institutes of Health.

ACKNOWLEDGMENTS

This manuscript benefited from conversations with Louis Ptacek, Andreas Ziegler, Murray Brilliant, Scott Hebring, Harold Ye, Sarah Murray, Mark Leppert, Nori Matsunami, and Ingrid Borecki, and comments from reviewers. The editorial comments were especially insightful and played an instrumental role in greatly improving the manuscript. I would like to particularly thank Tony Long and Ray White for sharing their keen observations and highly refined insights into genetic architectures of traits over many years. This work was supported by generous donors to the Marshfield Clinic Research Foundation, a pilot grant award from the NIH-NCATS/University of Wisconsin-Madison Institute for Clinical and Translational Research (UL1TR000427) and NIMH RO1MH097464. The content is solely the responsibility of the author and does not necessarily represent the official views of the National Institutes of Health.

REFERENCES

- Agarwala, V., Flannick, J., Sunyaev, S., GoT2D Consortium, and Altshuler, D. (2013). Evaluating empirical bounds on complex disease genetic architecture. *Nat. Genet.* 45, 1418–1427. doi: 10.1038/ng.2804
- Allison, A. C. (1954). The distribution of sickle cell trait in East Africa and elsewhere and its apparent relationship to the incidence of subtertian malaria. *Trans. R. Soc. Trop. Med. Hyg.* 48, 312–318. doi: 10.1016/0035-9203(54)90101-7
- Allum, F., Shao, X., Guenard, F., Simon, M. M., Busche, S., Caron, M. et al. (2015). Characterization of functional methylomes by next-generation capture sequencing identifies novel disease-associated variants. *Nat. Commun.* 6, 7211. doi: 10.1038/ncomms9016
- Balick, D. J., Do, R., Cassa, C. A., Reich, D., and Sunyaev, S. R. (2015). Dominance of deleterious controls the response to a population bottleneck. *PLoS Genet.* 11:e1005436. doi: 10.1371/journal.pgen.1005436
- Blangero, J., Diego, V. P., Dyer, T. D., Almeida, M., Peralta, J., Kent, J. W. Jr., et al. (2013). A kernel of truth: Statistical advances in polygenic variance component models for complex human pedigrees. *Adv. Genet.* 81, 1–31. doi: 10.1016/B978-0-12-407677-8.00001-4
- Botstein, D., White, R. L., Skolnick, M., and Davis, R. W. (1980). Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am. J. Hum. Genet.* 32, 314–331.
- Bottini, N., and Peterson, E. J. (2014). Tyrosine phosphatase PTPN22: multifunctional regulator of immune signaling, development, and disease. *Annu. Rev. Immunol.* 32, 83–119. doi: 10.1146/annurev-immunol-032713-120249
- Browning, S. R., and Thompson, E. A. (2012). Detecting rare variant associations by identity-by-descent mapping in case-control studies. *Genetics* 190, 1521–1531. doi: 10.1534/genetics.111.136937
- Bulik-Sullivan, B. K., Loh, P. R., Finucane, H. K., Ripke, S., Yang, J., Schizophrenia Working Group of the Psychiatric Genomics Consortium, et al. (2015). LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* 47, 291–295. doi: 10.1038/ng.3211
- Burkett, K. M., McNeney, B., Graham, J., and Greenwood, C. M. T. (2014). Using gene genealogies to detect rare variants associated with complex traits. *Hum. Hered.* 78, 117–130. doi: 10.1159/000363443

- Calafell, F., Grigorenko, E. L., Chikarian, A. A., and Kidd, K. K. (2001). Haplotype evolution and linkage disequilibrium: a simulation study. *Hum. Hered.* 51, 85–96. doi: 10.1159/000022963
- Cannings, C., Thompson, E. A., and Skolnick, M. H. (1978). Probability functions on complex pedigrees. *Adv. Appl. Prob.* 10, 26–61. doi: 10.2307/1426718
- Charlesworth, B., and Jain, K. (2014). Purifying selection, drift, and reversible mutation with arbitrarily high mutation rates. *Genetics* 198, 1587–1602. doi: 10.1534/genetics.114.167973
- Charlesworth, B., Nordborg, M., and Charlesworth, D. (1997). The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided population. *Genet. Res.* 70, 155–174. doi: 10.1017/S0016672397002954
- Chen, W., Larrabee, B. R., Ovsyannikova, I. G., Kennedy, R. B., Haralambieva, I. H., Poland, G. A., et al. (2015). Fine mapping causal variants with an approximate Bayesian method using marginal test statistics. *Genetics* 200, 719–736. doi: 10.1534/genetics.115.176107
- Curnow, R. N., and Smith, C. (1972). Multifactorial models for familial diseases in man. *J. R. Stat. Soc. Ser. A* 138, 131–169. doi: 10.2307/2984646
- DeBraekeleer, M. (1991). Hereditary disorders in Saguenay-Lac-St-Jean. (Quebec, Canada). *Hum. Hered.* 41, 141–146. doi: 10.1159/000153992
- Debray, Q., Caillard, V., and Stewart, J. (1979). Schizophrenia: a study of genetic models. *Hum. Hered.* 29, 27–36. doi: 10.1159/000153012
- Einstein, A. (1916). Die grundlage der allgemeinen relativitatstheorie. *Ann. Phys.* 49, 769–822. doi: 10.1002/andp.19163540702
- Everett, B. M., Smith, R. J., and Hiatt, W. R. (2015). Reducing LDL with PCSK9 inhibitors—The clinical benefit of lipid drugs. *N. Engl. J. Med.* 373, 1588–1591. doi: 10.1056/NEJMp1508120
- Ewens, W. J. (1972). The sampling theory of selectively neutral alleles. *Theor. Pop. Biol.* 3, 87–112. doi: 10.1016/0040-5809(72)90035-4
- Eyre-Walker, A. (2010). Genetic architecture of a complex trait and its implications for fitness and genome-wide association studies. *Proc. Natl. Acad. Sci. U.S.A.* 107(Suppl. 1), 1752–1756. doi: 10.1073/pnas.0906182107
- Falconer, D. S. (1965). The inheritance of liability to certain diseases estimated from the incidence in relatives. *Ann. Hum. Genet.* 29, 51–76. doi: 10.1111/j.1469-1809.1965.tb00500.x
- Falconer, D. S., and MacKay, T. F. C. (1996). *Introduction to Quantitative Genetics, 4th Edn.* Harlow: Longmans Green.
- Fan, J., Han, F., and Liu, H. (2014). Challenges of Big Data analysis. *Natl. Sci. Rev.* 1, 293–314. doi: 10.1093/nsr/nwt032
- Fisher, R. A. (1918). The correlation between relatives on the supposition of Mendelian inheritance. *Trans. R. Soc. Edinb.* 52, 399–433. doi: 10.1017/S0080456800012163
- Fisher, R. A. (1930). *The Genetical Theory of Natural Selection.* Oxford: Clarendon.
- Frank, S. A. (2011). “Wright’s adaptive landscape versus Fisher’s fundamental theorem,” in *The Adaptive Landscape in Evolutionary Biology*, eds E. I. Svensson and R. Calsbeek This model is often attributed to Fisher, but this view may not be entirely correct. (Oxford: Oxford University Press), 41–57.
- Fritsche, L. G., Fariss, R. N., Stambolian, D., Abecasis, G. R., Curcio, C. A., and Swaroop, A. (2014). Age-related macular degeneration: genetics and biology coming together. *Ann. Rev. Genomics Hum. Genet.* 15, 151–171. doi: 10.1146/annurev-genom-090413-025610
- Garcia, V. E., Chang, M., Brandon, R., Li, Y., Matsunami, N., Callis-Duffin, K. P., et al. (2008). Detailed genetic characterization of the interleukin-23 receptor in psoriasis. *Genes Immun.* 9, 546–555. doi: 10.1038/gene.2008.55
- Génin, E., and Clerget-Darpoux, F. (2015). The missing heritability paradigm: A dramatic resurgence of the GIGO syndrome in genetics. *Hum. Hered.* 79, 1–4. doi: 10.1159/000370327
- Gillespie, J. H. (1993). Substitution processes in molecular evolution. I. Uniform and clustered substitutions in a haploid model. *Genetics* 134, 971–981.
- Gillespie, J. H. (2000). Genetic drift in an infinite population: The pseudohitchhiking model. *Genetics* 155, 909–919.
- Glashow, S. L. (1961). Partial-symmetries of weak interactions. *Nucl. Phys.* 22, 579–588. doi: 10.1016/0029-5582(61)90469-2
- Graham, J. (1998). *Disequilibrium Fine-Mapping of a Rare Allele via Coalescent Models of Gene Ancestry.* Ann Arbor: UMI Dissertation Services.
- Green, R. E., Krause, J., Briggs, A. W., Maricic, T., Stenzel, U., Kircher, M., et al. (2010). A draft sequence of the Neandertal genome. *Science* 328, 710–722. doi: 10.1126/science.1188021
- Greenbaum, G. (2015). Revisiting the time until fixation of a neutral mutant in a finite population – A coalescent theory approach. *J. Theor. Biol.* 380, 98–102. doi: 10.1016/j.jtbi.2015.05.019
- Gronau, I., Hubisz, M. J., Gulko, B., Danko, C. G., and Siepel, A. (2011). Bayesian inference of ancient human demography from individual genome sequences. *Nat. Genet.* 43, 1031–1034. doi: 10.1038/ng.937
- Hall, E. H. (1879). On a new action of the magnet on electric currents. *Am. J. Math.* 2, 287–292. doi: 10.2307/2369245
- Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I. H., et al. (2009). The WEKA data mining software: an update; SIGKDD explorations. 11, 10–18. doi: 10.1145/1656274.1656278
- Hardy, G. H. (1908). Mendelian proportions in a mixed population. *Science* 28, 49–50. doi: 10.1126/science.28.706.49
- Higgs, P. W. (1964). Broken symmetries and the masses of gauge bosons. *Phys. Rev. Lett.* 13, 508–509. doi: 10.1103/PhysRevLett.13.508
- Hill, W. G., Goddard, M. E., and Visscher, P. M. (2008). Data and theory point to mainly additive genetic variance for complex traits. *PLoS Genet.* 4:e1000008. doi: 10.1371/journal.pgen.1000008
- Hill, W. G., and Robertson, A. (1968). Linkage disequilibrium in finite populations. *Theor. Appl. Genet.* 38, 226–231. doi: 10.1007/BF01245622
- Hodge, S. E. (1981). Some epistatic two-locus models of disease. I. Relative risks and identity-by-descent distributions in affected sib pairs. *Am. J. Hum. Genet.* 33, 381–395.
- Hormozdiari, F., Kostem, E., Kang, E. Y., Pasianic, B., and Eskin, E. (2014). Identifying causal variants at loci with multiple signals of association. *Genetics* 198, 497–508. doi: 10.1534/genetics.114.167908
- Houwen, R. H. J., Baharloo, S., Blankenship, K., Raeymaekers, P., Juyn, J., Sandkuijl, L. A., et al. (1994). Genome screening by searching for shared segments: mapping a gene for benign recurrent intrahepatic cholestasis. *Nat. Genet.* 8, 380–386. doi: 10.1038/ng1294-380
- Hu, F. B. (2011). Globalization of diabetes: the role of diet, lifestyle, and genes. *Diabetes Care* 34, 1249–1257. doi: 10.2337/dc11-0442
- Huckel, E. (1934). Theory of free radicals of organic chemistry. *Trans. Faraday Soc.* 30, 40–52. doi: 10.1039/TF9343000040
- Hudson, R. R. (1982). Estimating genetic variability with restriction endonucleases. *Genetics* (New York, NY: Oxford University Press) 100, 711–719.
- Hudson, R. R. (1983). Properties of a neutral allele model with intragenic recombination. *Theor. Popul. Biol.* 23, 183–201. doi: 10.1016/0040-5809(83)90013-8
- Hudson, R. R. (1991). “Gene genealogies and the coalescent process,” in *Oxford Surveys in Evolutionary Biology*, Vol. 7, eds D. Futuyma and J. Antonovics (New York, NY: Oxford University Press), 1–44.
- Hudson, R. R. (2015). A new proof of the expected frequency spectrum under the standard neutral model. *PLoS ONE* 10:e0118087. doi: 10.1371/journal.pone.0118087
- Hudson, R. R., and Kaplan, N. (1985). Statistical properties of the number of recombination events in the history of a sample of DNA sequences. *Genetics* 111, 147–164.
- Hund, F. (1926). Zur deutung einiger erscheinungen in den molekelspektren. *Zeitschrift Physik* 36, 657–674. doi: 10.1007/BF01400155
- Kaplan, N. L., Darden, T., and Hudson, R. R. (1988). The coalescent process in models with selection. *Genetics* 120, 819–829.
- Kaplan, N. L., Hill, W. G., and Weir, B. S. (1995). Likelihood methods for locating disease genes in nonequilibrium populations. *Am. J. Hum. Genet.* 56, 18–32.
- Kavanaugh, A., Ritchlin, C., Rahman, P., Puig, L., Gottlieb, A. B., Li, S., et al. (2014). Ustekinumab, an anti-IL-23/23 p40 monoclonal antibody, inhibits radiographic progression in patients with active psoriatic arthritis: results of an integrated analysis of radiographic data from the phase 3, multicenter, randomized, double-blind, placebo-controlled P SUMMIT-1 and P SUMMIT-2 trials. *Ann. Rheum. Dis.* 73, 1000–1006. doi: 10.1136/annrheumdis-2013-204741
- Kendler, K. S., and Diehl, S. R. (1993). The genetics of schizophrenia: a current, genetic-epidemiologic perspective. *Schizophr. Bull.* 19, 261–285. doi: 10.1093/schbul/19.2.261
- Kichaev, G., Yang, W.-Y., Lindstrom, S., Hormozdiari, F., Eskin, E., Price, A. L., et al. (2014). Integrating functional data to prioritize causal variants in statistical fine-mapping studies. *PLoS Gene* 10:e1004722. doi: 10.1371/journal.pgen.1004722

- Kingman, J. F. C. (1982a). On the genealogy of large populations. *J. Appl. Prob.* 184, 27–43. doi: 10.2307/3213548
- Kingman, J. F. C. (1982b). The coalescent. *Stochastic Process. Appl.* 13, 235–248.
- Kirino, Y., Bertias, G., Ishigatsubo, Y., Mizuki, N., Tugal-Tutkun, I., Seyahi, E., et al. (2013). Genome-wide association analysis identifies new susceptibility loci for Behcet's disease and epistasis between HLA-B*51 and ERAP1. *Nat. Genet.* 45, 202–207. doi: 10.1038/ng.2520
- Kruglyak, L. (1999). Prospects for whole-genome linkage disequilibrium mapping of common disease genes. *Nat. Genet.* 22, 139–144. doi: 10.1038/9642
- Lai, C., Lyman, R. F., Long, A. D., Langley, C. H., and Mackay, T. F. (1994). Naturally occurring variation in bristle number and DNA polymorphisms at the scabrous locus of *Drosophila melanogaster*. *Science* 266, 1697–1702. doi: 10.1126/science.7992053
- Lange, K. (1997). An approximate model of polygenic inheritance. *Genetics* 147, 1423–1430.
- Li, H., and Durbin, R. (2011). Inference of human population history from individual whole-genome sequences. *Nature* 475, 493–496. doi: 10.1038/nature10231
- Li, W., and Reich, J. (2000). A complete enumeration and classification of two-locus disease models. *Hum. Hered.* 50, 334–349. doi: 10.1159/000022939
- Libbrecht, M. W., and Noble, W. S. (2015). Machine learning applications in genetics and genomics. *Nat. Rev. Genet.* 16, 321–332. doi: 10.1038/nrg3920
- Loh, P. R., Bhatia, G., Gusev, A., Finucane, H. K., Bulik-Sullivan, B. K., Pollack, S. J., et al. (2015). Contrasting genetic architectures of schizophrenia and other complex disease using fast variance-components analysis. *Nat. Genet.* 47, 1385–1392. doi: 10.1038/ng.3431
- Long, A. D., Grote, M. N., and Langley, C. H. (1997). Genetic analysis of complex diseases. *Science* 275, 1328.
- Long, A. D., and Langley, C. H. (1999). The power of association studies to detect the contribution of candidate genetic loci to variation in complex traits. *Genome Res.* 9, 720–731.
- Lotka, A. J. (1925). *Elements of Physical Biology*. Baltimore, MD: Williams and Wilkins.
- Lueck, K., Busch, M., Moss, S. E., Greenwood, J., Kasper, M., Lommatzsch, A., et al. (2015). Complement stimulates retinal pigment epithelial cells to undergo pro-inflammatory changes. *Ophthalmic Res.* 54, 195–203. doi: 10.1159/000439596
- Lynch, M., and Walsh, B. (1998). *Genetics and Analysis of Quantitative Traits*. Sunderland, MA: Sinauer Associates.
- Lynn, A. H., Kwoh, C. K., Venglish, C. M., Aston, C. E., and Chakravarti, A. (1995). Genetic epidemiology of rheumatoid arthritis. *Am. J. Hum. Genet.* 57, 150–159.
- Maadooliat, M., Bansal, N. K., Upadhyay, J., Farazi, M., Ye, Z., Li, X., et al. (2016). The decay of disease association with declining linkage disequilibrium: a fine mapping theorem. *bioRxiv*. doi: 10.1101/052381
- Madsen, A. M., Hodge, S. E., and Ottman, R. (2011a). Causal models for investigating complex disease: I. A primer. *Hum. Hered.* 72, 54–62. doi: 10.1159/000330779
- Madsen, A. M., Ottman, R., and Hodge, S. E. (2011b). Causal models for investigating complex genetic disease: II. What causal models can tell us about penetrance for additive, heterogeneity, and multiplicative two-locus models. *Hum. Hered.* 72, 63–72. doi: 10.1159/000330780
- Majewski, J., Li, H., and Ott, J. (2001). The Ising model in physics and statistical genetics. *Am. J. Hum. Genet.* 69, 853–862. doi: 10.1086/323419
- Maruyama, T. (1974). The age of a rare mutant gene in a large population. *Am. J. Hum. Genet.* 26, 669–673.
- Mendel, G. (1866). “Versuche über pflanzenhybriden,” in *Verhandlungen des Naturforschenden Vereines* (Brunn), 3–47. IV für das Jahr 1865, abhandlungen.
- Michaelis, L., and Menten, M. L. (1913). Die kinetic der invertinwirkung. *Biochem. Z.* 49, 333–369.
- Molitor, J., Marjoram, P., and Thomas, D. (2003). Fine-scale mapping of disease genes with multiple mutations via spatial clustering techniques. *Am. J. Hum. Genet.* 73, 1368–1384. doi: 10.1086/380415
- Montano, C., Taub, M. A., Jaffe, A., Briem, E., Feinberg, J. I., Trygvadottir, R., et al. (2016). Association of DNA methylation differences with schizophrenia in an epigenome-wide association study. *JAMA Psychol.* 73, 506–514. doi: 10.1001/jamapsychiatry.2016.0144
- Moran, P. A. P. (1975). Wandering distributions and the electrophoretic profile. *Theor. Pop. Biol.* 8, 318–330. doi: 10.1016/0040-5809(75)90049-0
- Morris, A. P., Whittaker, J. C., and Balding, D. J. (2002). Fine-scale mapping of disease loci via shattered coalescent modeling of genealogies. *Am. J. Hum. Genet.* 70, 686–707. doi: 10.1086/339271
- Mott, N. F. (1938). Note on the contact between a metal and an insulator or semiconductor. *Proc. Camb. Philol. Soc.* 34, 568–572. doi: 10.1017/S0305004100020570
- Mulliken, R. S. (1932). Electronic structures of polyatomic molecules and valence. II. General considerations. *Phys. Rev.* 41, 49–71. doi: 10.1103/PhysRev.41.49
- Nelson, R. M., Pettersson, M. E., and Carlborg, O. (2013). A century after Fisher: time for a new paradigm in quantitative genetics. *Trends Genet.* 29, 669–676. doi: 10.1016/j.tig.2013.09.006
- Neuman, R. J., and Rice, J. P. (1992). Two-locus models of disease. *Genet. Epidemiol.* 9, 347–365. doi: 10.1002/gepi.1370090506
- Nielsen, D. M., Ehm, M. G., and Weir, B. S. (1999). Detecting marker disease association by testing for Hardy-Weinberg disequilibrium at a marker locus. *Am. J. Hum. Genet.* 63, 1531–1540. doi: 10.1086/302114
- Nothnagel, M., Furst, R., and Rohde, K. (2002). Entropy as a measure for linkage disequilibrium over multilocus haplotype blocks. *Hum. Hered.* 54, 186–198. doi: 10.1159/000070664
- Pan, W. (2010). A unified framework for detecting genetic association with multiple SNPs in a candidate gene or region: contrasting genotype scores and LD patterns between cases and controls. *Hum. Hered.* 69, 1–13. doi: 10.1159/000243149
- Pearl, J. (2000). *Causality: Models, Reasoning, and Inference*. Cambridge, UK: Cambridge University Press.
- Personal communication with Ray White (2000–2010). Ray White was a strong advocate of rare variants/allelic heterogeneity models driving common diseases since the early 1990s. He often promoted thinking of the “genetics of genes” to underscore this idea.
- Pickering, G. (1978). Normotension and hypertension: the mysterious viability of the false. *Am. J. Med.* 65, 561–563. doi: 10.1016/0002-9343(78)90839-2
- Plomin, R., Haworth, C. M. A., and Davis, O. S. P. (2009). Common disorders are quantitative traits. *Nat. Rev. Genet.* 10, 872–878. doi: 10.1038/nrg2670
- Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. A., and Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* 38, 904–909. doi: 10.1038/ng1847
- Price, G. R. (1970). Selection and covariance. *Nature* 227, 520–521. doi: 10.1038/227520a0
- Pritchard, J. K. (2001). Are rare variants responsible for susceptibility to complex diseases? *Am. J. Hum. Genet.* 69, 124–137.
- Pritchard, J. K., and Przeworski, M. (2001). Linkage disequilibrium in humans: models and data. *Am. J. Hum. Genet.* 69, 1–14. doi: 10.1086/321275
- Pritchard, J. K., Stephens, M., Rosenberg, N. A., and Donnelly, P. (2000). Association mapping in structured populations. *Am. J. Hum. Genet.* 67, 170–181. doi: 10.1086/302959
- Reich, D., Green, R. E., Kircher, M., Krause, J., Patterson, N., Durand, E. Y., et al. (2010). Genetic history of an archaic hominin group from Denisova cave in Siberia. *Nature* 468, 1053–1060. doi: 10.1038/nature09710
- Reich, D. E., and Lander, E. S. (2001). On the allelic spectrum of human disease. *Trends Genet.* 17, 502–510. doi: 10.1016/S0168-9525(01)02410-6
- Ridge, P. G., Mukherjee, S., Crane, P. K., Kauwe, J. S. K., and Alzheimer's Disease Genetics Consortium (2013). Alzheimer's disease: Analyzing the missing heritability. *PLoS ONE* 8:e79771. doi: 10.1371/journal.pone.0079771
- Risch, N., and Merikangas, K. (1996). The future of genetic studies of complex human diseases. *Science* 273, 1516–1517. doi: 10.1126/science.273.5281.1516
- Ritchie, M. D., Hahn, L. W., Roodi, N., Bailey, L. R., Dupont, W. D., Parl, F. F., et al. (2001). Multifactor-dimensionality reduction reveals high-order interactions among estrogen-metabolism genes in sporadic breast cancer. *Am. J. Hum. Genet.* 69, 138–147. doi: 10.1086/321276
- Roychowdhury, S., and Chinnaiyan, A. M. (2013). Advancing precision medicine for prostate cancer through genomics. *J. Clin. Oncol.* 31, 1866–1873. doi: 10.1200/JCO.2012.45.3662
- Sabbagh, A., and Darlu, P. (2006). Data-mining methods as useful tools for predicting individual drug response: application to CYP2D6 data. *Hum. Hered.* 62, 119–134. doi: 10.1159/000096416
- Sanjak, J. S., Long, A. D., and Thornton, K. R. (2016). The genetic architecture of a complex trait is more sensitive to genetic model than population growth. *bioRxiv*. doi: 10.1101/048819

- Sankararaman, S., Mallick, S., Patterson, N., and Reich, D. (2016). The combined landscape of Denisovan and Neanderthal ancestry in present-day humans. *Curr. Biol.* 26, 1241–1247. doi: 10.1016/j.cub.2016.03.037
- Schottky, W. (1938). Halbleitertheorie der sperrschicht. *Naturwissenschaften* 26, 843. doi: 10.1007/BF01774216
- Schrodi, S. J., Garcia, V. E., and Rowland, C. M. (2009). “A fine mapping theorem to refine results from association genetic studies,” in *ASHG Abstract*. Honolulu, HI.
- Schrodi, S. J., Garcia, V. E., Rowland, C., and Jones, H. B. (2007). Pairwise linkage disequilibrium under disease models. *Eur. J. Hum. Genet.* 15, 212–220. doi: 10.1038/sj.ejhg.5201731
- Schrodi, S. J., Mukherjee, S., Shan, Y., Tromp, G., Sninsky, J. J., Callear, A. P., et al. (2014). Genetic-based prediction of disease traits: prediction is very difficult, especially about the future. *Front. Genet.* 5:162. doi: 10.3389/fgene.2014.00162
- Schrodinger, E. (1926). Quantisierung als eigenwertproblem. *Annalen Phys.* 384, 273–376. doi: 10.1002/andp.19263840404
- Servin, B., and Stephens, M. (2007). Imputation-based analysis of association studies: candidate regions and quantitative traits. *PLoS Genet.* 3:e114. doi: 10.1371/journal.pgen.0030114
- Slatkin, M., and Rannala, B. (1997). The sampling distribution of disease-associated alleles. *Genetics* 147, 1855–1861.
- Speed, D., and Balding, D. J. (2014). MultiBLUP: improved SNP-based prediction for complex traits. *Genome Res.* 24, 1550–1557. doi: 10.1101/gr.169375.113
- Stahl, E. A., Wegmann, D., Trynka, G., Gutierrez-Achury, J., Do, R., Voight, B. F., et al. (2012). Bayesian inference analyses of the polygenic architecture of rheumatoid arthritis. *Nat. Genet.* 44, 483–489. doi: 10.1038/ng.2232
- Stein, C. M., Nshuti, L., Chiunda, A. B., Boom, W. H., Elston, R. C., Mugerwa, R. D., et al. (2005). Evidence for a major gene influence on tumor necrosis factor- α expression in tuberculosis: path and segregation analysis. *Hum. Hered.* 60, 109–118. doi: 10.1159/000088913
- Steinberg, A. G., Becker, S., Fitzpatrick, T. B., and Kierland, R. R. (1951). A genetic and statistical study of psoriasis. *Am. J. Hum. Genet.* 3, 267–281.
- Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123, 585–595.
- Te Meerman, G. J., Van Der Meulen, M. A., and Sandkuijl, L. A. (1995). Perspectives of identity by descent (IBD). mapping in founder populations. *Clin. Exp. Allergy* 25, 97–102. doi: 10.1111/j.1365-2222.1995.tb00433.x
- Terwilliger, J. D., and Göring, H. H. H. (2000). Gene mapping in the 20th and 21st centuries: Statistical methods, data analysis, and experimental design. *Hum. Biol.* 72, 63–132. doi: 10.3378/027.081.0615
- Thompson, E. A. (1978). “Linkage and the power of a pedigree structure,” in *Genetic Epidemiology*, eds N. E. Morton and C. S. Chung (New York, NY: Academic), 247–253.
- Thornton, K. R., Foran, A. J., and Long, A. D. (2013). Properties and modeling of GWAS when complex disease risk is due to non-complementing, deleterious mutations in genes of large effect. *PLoS Genet.* 9:e1003258. doi: 10.1371/journal.pgen.1003258
- Uddin, M., Aiello, A. E., Wildman, D. E., Koenen, K. C., Pawelec, G., de los Santos, R., et al. (2010). Epigenetic and immune function profiles associated with posttraumatic stress disorder. *Proc. Natl. Acad. Sci. U.S.A.* 107, 9470–9475. doi: 10.1073/pnas.0910794107
- Vilhjalmsón, B. J., Yang, J., Finucane, H. K., Gusev, A., Lindström, S., Ripke, S., et al. (2015). Modeling linkage disequilibrium increases accuracy of polygenic risk scores. *Am. J. Hum. Genet.* 97, 576–592. doi: 10.1016/j.ajhg.2015.09.001
- Vinkhuyzen, A. A., Wray, N. R., Yang, J., Goddard, M. E., and Visscher, P. M. (2013). Estimation and partition of heritability in human populations using whole genome analysis methods. *Annu. Rev. Genet.* 47, 75–95. doi: 10.1146/annurev-genet-111212-133258
- Watterson, G. A. (1975). On the number of segregating sites in genetical models without recombination. *Theor. Pop. Biol.* 7, 256–276. doi: 10.1016/0040-5809(75)90020-9
- Weinberg, S. (1967). A model of leptons. *Phys. Rev. Lett.* 19, 1264–1266. doi: 10.1103/PhysRevLett.19.1264
- Weinberg, W. (1908). Über den nachweis der verebung beim menschen. *Jahreshefte des vereins für vaterländische naturkunde in Württemberg.* 64, 368–382.
- Wellcome Trust Case Control Consortium, Maller, J. B., McVean, G., Byrnes, J., Vukcevic, D., Palin, K., et al. (2012). Bayesian refinement of association signals for 14 loci in 3 common diseases. *Nat. Genet.* 44, 1294–1301. doi: 10.1038/ng.2435
- Williams, S. R., Aldred, M. A., Der Kaloustian, V. M., Halal, F., Gowans, G., McLeod, D. R., et al. (2010). Haploinsufficiency of HDAC4 causes brachydactyly mental retardation syndrome, with brachydactyly type E, developmental delays, and behavioral problems. *Am. J. Hum. Genet.* 87, 219–228. doi: 10.1016/j.ajhg.2010.07.011
- Wilson, A. H. (1931). The theory of electronic semi-conductors. *Proc. R. Soc. Lond. A* 133, 458–491. doi: 10.1098/rspa.1931.0162
- Wiuf, C., and Posada, D. (2003). A coalescent model of recombination hotspots. *Genetics* 164, 407–417.
- Wray, N. R., and Goddard, M. E. (2010). Multi-locus models of genetic risk of disease. *Genome Med.* 2, 10. doi: 10.1186/gm131
- Wright, S. (1932). “The roles of mutation, inbreeding, crossbreeding and selection in evolution,” in *Proceedings of the 6th International Congress of Genetics*, Austin, TX. Vol. 1, 356–366.
- Wright, S. (1934). An analysis of variability in number of digits in an inbred strain of guinea pigs. *Genetics* 19, 506–536.
- Xiong, M., and Guo, S. W. (1998). The power of linkage detection by the transmission/disequilibrium tests. *Hum. Hered.* 48, 295–312. doi: 10.1159/000022821
- Yang, J., Benyamin, B., McEvoy, B. P., Gordon, S., Henders, A. K., Nyholt, D. R., et al. (2010). Common SNPs explain a large proportion of the heritability for human height. *Nat. Genet.* 42, 565–569. doi: 10.1038/ng.608
- Yang, J., Manolio, T. A., Pasquale, L. R., Boerwinkle, E., Caporaso, N., Cunningham, J. M., et al. (2011a). Genome partitioning of genetic variation for complex traits using common SNPs. *Nat. Genet.* 43, 519–525. doi: 10.1038/ng.823
- Yang, J., Weedon, M. N., Purcell, S., Lettre, G., Estrada, K., Willer, C. J., et al. (2011b). Genomic inflation factors under polygenic inheritance. *Eur. J. Hum. Genet.* 19, 807–812. doi: 10.1038/ejhg.2011.39
- Zaykin, D. V., and Shibata, K. (2008). Genetic flip-flop without an accompanying change in linkage disequilibrium. *Am. J. Hum. Genet.* 82, 794–796. doi: 10.1016/j.ajhg.2008.02.001
- Zaykin, D. V., Meng, Z., and Ehm, M. G. (2006). Contrasting linkage-disequilibrium patterns between cases and controls as a novel association-mapping method. *Am. J. Hum. Genet.* 78, 737–746. doi: 10.1086/503710
- Zhou, X., Carbonetto, P., and Stephens, M. (2013). Polygenic modeling with Bayesian sparse linear mixed models. *PLoS Genet.* 9:e1003264. doi: 10.1371/journal.pgen.1003264
- Zhu, Z., Baksji, A., Vinkhuyzen, A. A., Hemani, G., Lee, S. H., Nolte, I. M., et al. (2015). Dominance genetic variation contributes little to the missing heritability for human complex traits. *Am. J. Hum. Genet.* 96, 377–385. doi: 10.1016/j.ajhg.2015.01.001
- Ziezun, A., Pulit, S. L., Francioli, L. C., van Dijk, F., Swertz, M., Boomsma, D. I., et al. (2013). Deleterious alleles in the human genome are on average younger than neutral alleles of the same frequency. *PLoS Genet.* 9:e1003301. doi: 10.1371/journal.pgen.1003301
- Zollner, S., and Pritchard, J. K. (2005). Coalescent-based association mapping and fine mapping of complex trait loci. *Genetics* 169, 1071–1092. doi: 10.1534/genetics.104.031799
- Zuk, O., Hechter, E., Sunyaev, S. R., and Lander, E. S. (2012). The mystery of missing heritability: Genetic interactions create phantom heritability. *Proc. Natl. Acad. Sci. U.S.A.* 109, 1193–1198. doi: 10.1073/pnas.1119675109

Conflict of Interest Statement: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Schrodi. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.