Leading Edge
Perspective

Cell

Common Disease Is More Complex Than Implied by the Core Gene Omnigenic Model

Naomi R. Wray,[1,2,*] Cisca Wijmenga,[3] Patrick F. Sullivan,[4,5,6] Jian Yang,[1,2] and Peter M. Visscher[1,2]
[1]Institute for Molecular Bioscience, The University of Queensland, Brisbane, Australia
[2]Queensland Brain Institute, The University of Queensland, Brisbane, Australia
[3]University of Groningen, University Medical Center Groningen, Department of Genetics, P.O. Box 30001, 9700 RB Groningen, Netherlands
[4]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden
[5]Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA
[6]Department of Psychiatry, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA
*Correspondence: naomi.wray@uq.edu.au
https://doi.org/10.1016/j.cell.2018.05.051

The evidence that most adult-onset common diseases have a polygenic genetic architecture fully consistent with robust biological systems supported by multiple back-up mechanisms is now overwhelming. In this context, we consider the recent "omnigenic" or "core genes" model. A key assumption of the model is that there is a relatively small number of core genes relevant to any disease. While intuitively appealing, this model may underestimate the biological complexity of common disease, and therefore, the goal to discover core genes should not guide experimental design. We consider other implications of polygenicity, concluding that a focus on patient stratification is needed to achieve the goals of precision medicine.

The Omnigenic Core versus Peripheral Gene Model

In a recent scholarly Perspective, Boyle et al. (2017b) undertook a series of elegant analyses of genome-wide association study (GWAS) summary statistics matched with genomic annotations. They hypothesize that GWAS results are "hinting at important new principles of biological organization, and how genetic perturbations percolate through cellular systems" (Boyle et al., 2017a). They introduce the term "omnigenic," (omni = "all") in acknowledgment of the very large number of genetic loci contributing to disease risk, which "is largely driven by genes with no direct relevance to disease and is propagated through regulatory networks to a much smaller number of core genes with direct effects." A key feature of the omnigenic model is the classification of genes as "peripheral" (which are generally regulatory in cellular networks and contribute to risk for many diseases and therefore to pleiotropy) or "core" (which are more disease specific with biologically interpretable roles). A defining feature of the omnigenic hypothesis is that only a modest number of genes or pathways have specific roles in the etiology of a specific disease, and these core genes, if mutated or deleted, have the strongest functional effects, since natural selection keeps risk alleles at these key genes at a low frequency in the population. The authors conclude by asking if the concept of core genes (and therefore, by implication, the assumption that only a few genes play key roles in a complex disease) is a useful one. Here, we explore this question and draw different conclusions for future experimental strategies based on our viewpoint on how to best propel genetic discoveries into outcomes for patients.

What's New in the Omnigenic Core versus Peripheral Gene Model?

Reconciliation of polygenic disease architecture with the epidemiological observation that disease is uncommon in the population is inherently consistent with robustness in biological systems (Masel and Siegal, 2009). The broad interpretation of the results of the last decade of genetic studies by Boyle et al. about the likely role of regulatory mechanisms in common disease is one with which most scholars in the field would concur (Cox, 2017; Franke, 2017). Fundamentals of the omnigenic model resemble those underpinning prior thinking, such as the cis-regulatory hypothesis (Stern and Orgogozo, 2008), discussions of pleiotropy based on results from the earliest days of GWASs (Sivakumaran et al., 2011), network medicine conceptualisations (Barabási et al., 2011; Civelek and Lusis, 2014), or the core gene regulatory networks hypothesis (Chakravarti and Turner, 2016). The key point of distinction of the omnigenic hypothesis is the emphasis on the importance of core genes, with the assumption that types of genes detected in rare variant studies—which can detect highly deleterious variants with large effect sizes—play more direct roles in complex disease than do genes identified from GWASs based on common variants. Thus, a consequence of the model is to focus experimental designs on discovery of rare variants. While intuitively appealing, this conclusion implies a simpler gene-disease biology than we have empirical evidence for. For example, one conclusion from sequencing genomes from healthy individuals was the high level of redundancy/robustness in the human genome, since most apparently normal humans have ~100 loss-of-function mutations (MacArthur et al., 2012). An alternative view is to fully

embrace the concept of multiple biological back-up systems that are implied by polygenicity. From a genetic architecture perspective, the core/peripheral properties align closely with those of older conceptualizations considering the relative importance of rare/common variants (Pritchard, 2001; Pritchard and Cox, 2002), implying that the omnigenic model is partly a reframing of older ideas while trying to accommodate the empirical evidence that confirms polygenicity and a role of risk variants from across the allelic spectrum. Thus, a closer look at the definition of the core gene is warranted.

Evidence to Support the Key Assumption of a Core Gene?

Mendelian disease clearly fulfills the core gene definition, as disease only occurs in the context of a given mutation. Huntington's disease, for example, only presents in those with an expanded trinucleotide repeat in the gene *HTT*. Yet important advances for identification of potential therapeutic targets have been made through GWASs of age of onset (GeM-HD, 2015) and rate of disease progression (Hensman Moss et al., 2017). Hence, a definition of Mendelian disease might be one in which the complex compensatory network response mechanisms to the presence of a mutation in a core gene have failed. For common disease, the core gene hypothesis acknowledges that a single core gene is not causal for disease but makes the strong assumption that there is a small number of genes that can be identified against a background of contributions from peripheral genes that have non-disease-specific roles in modulating disease risk. We first assume that the goal to detect core genes for common disease is valid and examine the overlap in genes detected through GWASs and other studies. Next, we question if the definition of core genes is useful when studying GWAS results alone. Third, we consider whether association effect size is the key driver of determining biological relevance in complex disease. These are key questions that impact on priorities for experimental design.

First, the classical polygenic model incorporates contributions from both common and rare variants to disease risk, with many more rare variants than common (Crow and Kimura, 1970; Reich and Lander, 2001). While the vast majority of rare variants have small effect sizes, under a selection model, some rare variants may have larger effect than variants that are more common in the population (Hansen et al., 2006). Disease-risk genes are expected to harbor both common and rare risk variants, and empirical data for height (Kemper et al., 2012; Marouli et al., 2017), type 2 diabetes (Fuchsberger et al., 2016), inflammatory bowel disease (IBD) (Luo et al., 2017), and high-density lipoprotein (HDL) cholesterol (Rosenson et al., 2018) show common-variant associations variants in genes responsible for related monogenic disorders.

So, if we assume that the concept of core genes is useful and that identification of these genes will further our understanding of the underlying biology, the question is then of the best experimental design to identify such genes. Boyle et al. advocate whole-exome sequencing (WES) studies for identification of such genes, but they make no mention of the required sample sizes. WES *de novo* mutation studies of epilepsies, developmental delay, and congenital heart disease have shown that hundreds of genes may each be a core gene for different children presenting with severe childhood syndromes, and mutations in the same core gene are associated with extensive phenotypic heterogeneity (Jin et al., 2017; Short et al., 2018; Thomas and Berkovic, 2014). But for common disorders, the largest WES studies conducted to date have not been sufficiently powered to detect the effect sizes that exist in nature. For example, in schizophrenia (4,877 cases), WES did not detect any individual gene with excess rare variants compared to controls (Genovese et al., 2016). For IBD, WES (4,280 cases) identified a single rare variant (in a previously known locus) and an excess of very rare, damaging missense variants in known Crohn's disease risk genes (including those identified through GWAS) (Luo et al., 2017). For type 2 diabetes, the conclusion from analysis of WES (7,380 cases) was "large-scale sequencing does not support the idea that lower-frequency coding variants have a major role in predisposition" (Fuchsberger et al., 2016). A plausible interpretation of these results is that when rare large-effect variants occur, the resulting clinical phenotypes differ importantly from the common-variant-associated trait due to, for example, lethality, broader somatic impact, greater severity, or earlier onset. This conclusion is consistent with a key finding from genetic studies in psychiatric disorders that report an enrichment of common variants in genes associated with intellectual disability, developmental delay, and epilepsy (Gandal et al., 2016; Sullivan et al., 2012).

Therefore, the rare monogenic disorders help identify core genes (Antonarakis and Beckmann, 2006), but for common disease, larger samples than studied to date are needed to identify rare variants of the effect size that exist in nature and survive natural selection (Kiezun et al., 2012). Large GWAS samples imputed to the ever-increasing sequenced referenced samples mean that disease-risk rare variants can be identified by this paradigm. For example, a GWAS meta-analysis (74,124 type 2 diabetes cases) identified 24 imputed genome-wide significant SNPs with minor allele frequency of less than 0.05% (Mahajan et al., 2018), which point to both known and novel genes. While most of their 243 genome-wide significant association signals mapped to regulatory regions, 18 genes were highlighted as human-validated therapeutic targets (Mahajan et al., 2018). The power of association studies is dependent on the variance explained by a locus; thus, for a disease that affects 1% of the population, we have the same power to detect a risk locus of 50% frequency and odds ratio of 1.1 as we do for a risk locus of 0.1% frequency and odds ratio of 2.9. Thus, sample size—not genotyping technology—is the limiting factor in associated/causal variant discovery.

Second, in proposing the omnigenic core genes model, four examples where GWAS results have helped identify important core genes and highlight specific molecular processes (role of autophagy in Crohn's disease, role of adipocyte thermogenesis, central nervous system genes in obesity, and role of *C4A* on synaptic pruning in development that contributes to risk of schizophrenia) were reported (Boyle et al., 2017b). The fact that these examples have led to better understanding of underlying biology does not preclude similar future understanding for other genes identified through common variants, since we remain ignorant of the full complexity of the biology in which

most genes function. Thus, more genes of key biological relevance have likely been identified from GWASs; it is just that we are not yet able to annotate them as such (Cox, 2017).

Third, Boyle et al. conclude that the GWAS enrichment signal in relevant genes (based on current knowledge) is surprisingly weak. These observations are not necessarily contradictory and are indeed expected. Current annotation of gene function is very imperfect, and reports of novel and surprising functions of genes are common. As a specific example, typical annotations of synaptic genes have little resemblance to what is known in synaptic biology (Lips et al., 2012). Weak effect size is not unexpected if the effect size of a SNP on disease is via regulation of the expression level of a gene. If the GWAS SNP is an eQTL (i.e., its alleles correlate with levels of gene expression, with a stronger eQTL effect than SNP-disease association effect), then even if these eQTLs are tissue specific, it does not follow that all will have tissue-specific effects on gene expression large enough to generate detectable differences in mean expression level between tissues (Qi et al., 2018). Moreover, it is well recognized that association effect size is not well correlated with clinical relevance, as many FDA-approved medications have drug targets linked to common risk variants identified in GWASs (Faraone, 2017; Gandal et al., 2016; Nelson et al., 2015). In other words, while the effect size for a SNP associated with disease risk is small, pharmacological intervention targeted at the associated gene or gene product can be effective for those with and without specific risk variants, because the biological pathway targeted is relevant to their disease etiology.

To summarize this section, the broader literature suggests that once identified, risk genes likely harbor both rare and common variants, and rare variants of very large effect are inconsistent with most diagnoses of common diseases, since when such variants are consistent with life, they attract more severe and/or early-onset diagnoses. Given the effect sizes now expected for rare variants associated with common disease, sample size is the primary limiting factor in their detection. Importantly, effect size has little association with biological or clinical relevance. Taken together, in the context of common disease, a primary focus on the types of genes detected in rare-variant studies seems misplaced.

A More Complete Picture of Polygenicity

Although some in the human genetics community have been surprised by the empirical evidence for polygenicity of complex disease (Boyle et al., 2017b), for many researchers, this was expected and had been long hypothesized (Gottesman and Shields, 1967; Penrose, 1953), using models in which liability to complex disease has the properties of a quantitative trait. For example, genetic architecture modeling that aimed to match simulated data to early GWAS results included genetic architectures in which all independent genetic sites harbored risk loci (International Schizophrenia Consortium et al., 2009). Such polygenic disease hypotheses were firmly grounded in the knowledge gained from artificial selection studies set up to confirm quantitative genetic theory, many derived prior to the discovery of DNA (Dunnington et al., 2013; Hill, 2010). Response to artificial selection results can only be explained by a massive number of combinations of DNA variants being able to generate the same phenotype, which in turn reflects the massive amount of variation hidden in the genome, such that half the genetic variance in a population is available through the segregation variance between children from any pair of parents (Lynch and Walsh, 1998). With this in mind, a conclusion from the omnigenic model, "that many of the more dramatic phenotypic differences seen between species are also driven by an accumulation of tiny effects and that larger-effect differences are likely to be exceptions to the rule" (Boyle et al., 2017b), is consistent with quantitative-genetics thinking since Darwin—that evolutionary adaptations are mostly from substitutions of many gene variants of small effects (Denny et al., 2010; Orr and Coyne, 1992). Much quantitative genetic theory has used the infinitesimal model, which assumes that all genetic loci influence a trait and is a useful approximation of polygenic traits. From our point of view, the term omnigenic is describing the same genetic architecture as the infinitesimal model. The term polygenic covers any genetic architecture from few to all contributing variants and thus captures many architectures that exist both between and within disease classifications. Therefore, we believe the introduction of the term omnigenic over the existing terms of polygenic and infinitesimal is not a useful addition to our vocabulary.

Other key properties of polygenicity are sometimes underrecognized. These properties are emphasized here because they contribute to our perspective for prioritisation of current and future experimental strategies. While polygenicity describes the genetic architecture in a population, it also has important implications for individuals. All humans carry many risk alleles for all common diseases, and each individual affected with disease likely carries a higher burden and unique portfolio of risk variants (see Figure 1). Paradoxically, while the description of a polygenic model at the population level is very simply defined, it generates considerable genetic heterogeneity between individuals, which in turn is consistent with characteristics of common complex disease, such as heterogeneity in clinical presentation and variation in response to treatments (although other factors certainly may contribute to these observations).

Understanding the consequences of polygenicity for individuals also links into an understanding of epistasis, the interacting effects of risk loci. As discussed by others (Paixão and Barton, 2016; Phillips, 2008), expectations for the role of epistasis in complex genetic disease are confusing and confused. Molecular biology studies provide unequivocal evidence that gene-gene interactions are common and impart a strong desire to undertake studies to detect epistatic associations, yet quantitative genetic theory suggests that contributions from non-additive effects to phenotypic variation in the population and differences between people are small (Hill et al., 2008; Mäki-Tanila and Hill, 2014; Polderman et al., 2015; Zhu et al., 2015). These differing viewpoints are accommodated under a polygenic genetic architecture, as the only way to reconcile disease that impacts only a small fraction of the population with a genetic architecture of many risk loci is to have a highly non-linear relationship between probability of disease and burden of risk alleles (Slatkin, 2008). In quantitative genetics language, polygenic disease is non-additive on the disease scale. Under polygenicity, this means that

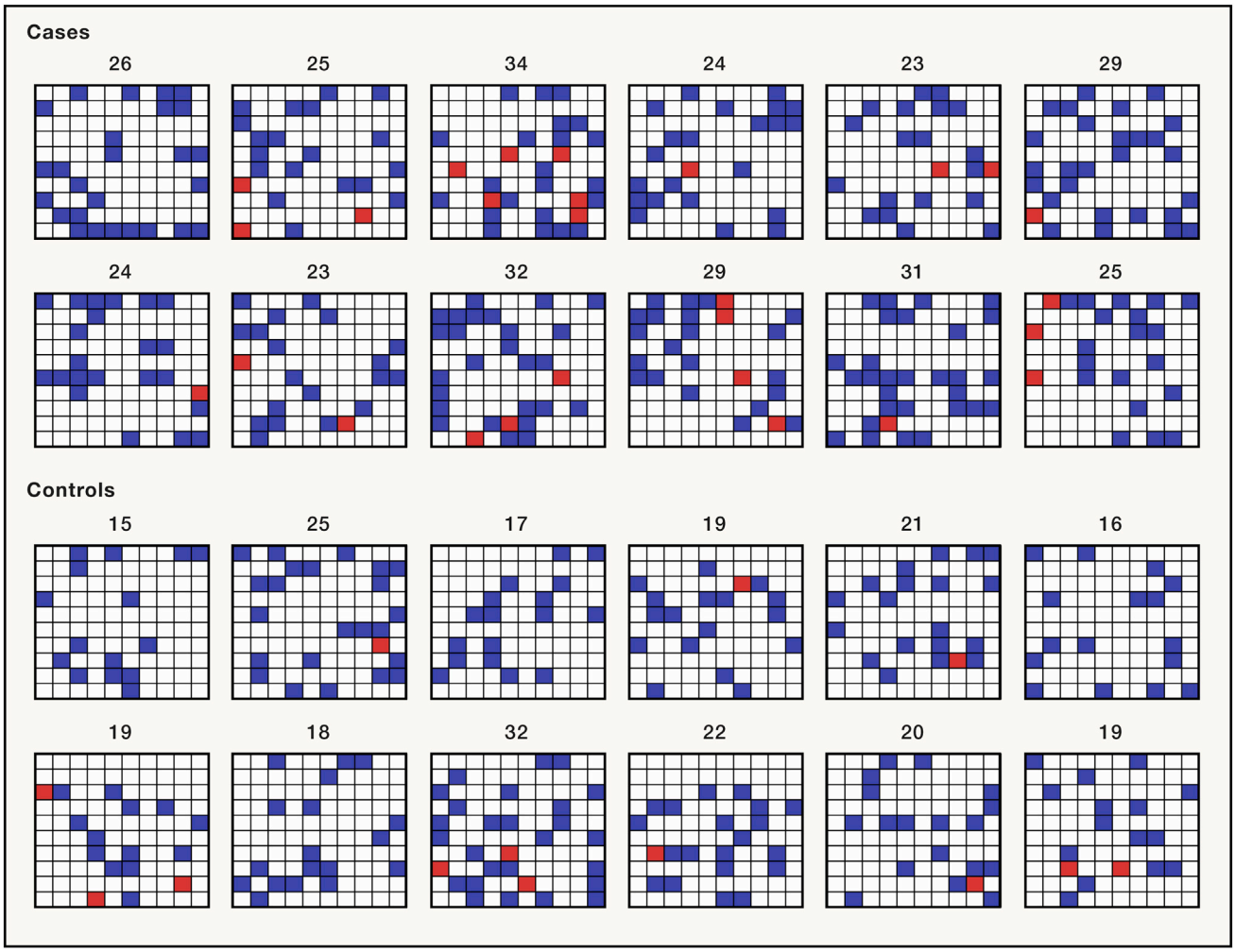

Figure 1. Between Individual Genetic Heterogeneity under a Polygenic Model

We use a toy example to show how a polygenic genetic architecture generates considerable genetic heterogeneity between individuals. Each box represents the genomic risk profile of an individual, assuming 100 independent risk loci contribute to disease (polygenic common diseases are underpinned by 10- to 1,000-fold more loci). Each box is divided into 100 squares, each representing the genotype of a locus. The squares are colored white if the person carries two non-risk alleles at a locus, blue if they carry one risk allele, and red if they carry two risk alleles. Each risk allele has frequency of 0.1 and contributes equal risk to disease. Heritability for the disease is 0.5, and disease risk is 0.01. The total number of risk alleles in the risk profile is listed at the top of each box. An average person in the population carries 20 risk alleles; thus, controls all carry risk alleles, but usually not as many as cases. Precision medicine may depend on matching genomic profiles to drug treatments. If the first row represents 10% of variants in "core" genes, there is also considerable heterogeneity in risk profiles between people. Heterogeneity increases with the number of risk loci contributing to disease. Varying risk allele frequencies and effect sizes have little impact on the message of the visualization.

each individual with a disease carries a high burden of risk alleles that together make a biological system that is vulnerable to non-genetic risk factors and cumulatively result in disease. However, each individual carries a (near-)unique combination of the many risk variants and environmental exposures. Since there is no specific combination of risk alleles that cause this non-additive genetic variation, this sort of epistasis is removed by transformation to liability or log (odds ratio) scales of disease. So while the statistical genetics community prefers to say that complex disease is underpinned by genetic effects working additively in liability to risk, for a molecular geneticist studying samples from diseased and healthy individuals, interactions between genetic effects are indeed implied, but on a scale that is challenging to study, both in terms of number of contributing genes and uniqueness of individuals. We can hypothesize about epistasis on the liability scale (Zuk et al., 2012), but unachievably massive sample sizes would be needed to distinguish epistasis from additivity empirically.

In summary, the additive polygenic disease architecture provides a homogeneous representation of disease for a population that is entirely consistent with a high degree of non-additivity at the biological level and between-individual heterogeneity that is challenging to differentiate based on clinical presentation. These properties of common disease must be acknowledged when formulating priorities for experimental design if we are to accelerate clinical outcomes from genetic research.

Implications for Future Experimental Design for Gene Discovery

Boyle et al. conclude their Perspective by asking if the conceptual distinction between core genes and peripheral genes is useful for understanding disease. Addressing their question matters because it potentially informs experimental design for gene mapping and functional studies. If it is not a useful distinction, then the implications for experimental designs that follow from the core gene model, including the most effective strategies for gene mapping and functional studies, are invalid. It seems to us to be a strong assumption that only a few genes have a core role in a common disease. Given the extent of biological robustness, we cannot exclude an etiology of many core genes, which in turn may become indistinguishable from a model of no core genes. Boyle et al. suggest that in the short term, exome sequencing should be a priority in common disease research to identify additional core genes to identify larger-effect variants that are more likely to affect protein-coding sequences. The core gene hypothesis has interpreted by some (Callaway, 2017) to imply that continuing to build GWAS cohorts for common disease will not be cost effective as part of the research portfolio to advance our understanding of common disease. However, this viewpoint assumes a simplicity that may not apply to some common diseases. In our opinion, whether the goal is discovery of rare variants or common variants, sample sizes are a key limiting factor for furthering our understanding of polygenic diseases, and increasing sample size remains a research priority needed to further the genetic discoveries that will ultimately make impact for those affected.

To date, follow-up analyses have been limited by a lack of depth of phenotypic information in GWAS cohorts. A pragmatic choice is to focus on collection of large sample sizes today (with relatively detailed but cost-effectively collected, phenotypes) and use of cheap SNP chip arrays (now less than $40 USD/sample) to generate genome-wide genotype data from which variants of frequency >0.5% can be imputed with high accuracy from fully sequenced reference samples (McCarthy et al., 2016; Yang et al., 2015). As these reference cohorts have become larger and better sequenced, disease associations with variants of frequency as low as 0.25% have been reported (Mahajan et al., 2018). For many diseases, it will take time before accumulated sample sizes are powered to detect associations of risk alleles that are less common in the population than alleles accurately assessed through imputation.

By the time such samples have been accumulated, sequencing technology is likely to have improved (to allow better detection in some of the most structurally challenging but perhaps very interesting genomic regions (Frith and Khan, 2018; Sekar et al., 2016; Treangen and Salzberg, 2011) and become cheaper, allowing for experimental designs less prone to confounders (Leek et al., 2010). Moreover, enhancement of reference datasets (whole-genome sequence, gene expression, etc.) will lead to improved annotation of GWAS-imputed data—for example, to include loss-of-function variants (Havrilla et al., 2017; Samocha et al., 2017), the key motivator of WES studies. To bias experimental design toward a hypothesis based on a critical assumption that only a few genes play key roles in complex disease would be putting all eggs in one basket. Of course, sequencing is the appropriate technology in many experimental settings (e.g., discovery of *de novo* mutations in severe childhood syndromes, Deciphering Developmental Disorders, 2017; rare Mendelian disorders; extreme-phenotype families, Chakravarti and Turner, 2016; cancer tumor versus normal cells; single-cell gene expression). Even for common complex diseases and disorders, a transition to sequencing is inevitable as prices decline; however, we should focus first on building sample size and GWASs. In the long term, we believe this approach will be both more cost effective and more productive compared to turning immediately to underpowered WES. Our approach will deliver not only in discovery of less common associated variants, but also in advancing disease-risk prediction and patient stratification. For example, risk prediction has already achieved clinical utility for prostate cancer (Grönberg et al., 2015) and for stratification in adult-onset diabetes (Oram et al., 2016).

In our opinion—for some flagship disorders, at least—we need to increase sample sizes until the discovery of common associated variants starts to plateau, which has not happened yet. In psychiatry, for example, obvious choices are schizophrenia, as it is the most intensively studied disorder (Pardiñas et al., 2016; Schizophrenia Working Group of the Psychiatric Genomics, 2014) and major depression, which as a very common disease can benefit from GWAS meta-analyses utilizing international biobank projects (Levinson et al., 2014). Together, these disorders span the extremes of several key epidemiological dimensions, including frequency of disease, sex ratios, known non-genetic risk factors, and heritability. There is a rapidly increasing number of ways to combine GWAS summary statistics with biologically informed datasets linked through associated SNPs or genes (and enhancements of such reference datasets are also a priority) for further understanding of complex diseases (Pasaniuc and Price, 2017).

To access the large samples needed to extend and complete risk-variant discovery of common disease, new strategies may be needed to enable large-scale sample collections through clinical facilities, which may require clarification and reform on consenting policy (Caulfield and Murdoch, 2017). With very large sample sizes genotyped on cheap SNP chips, followed by imputation, we will gain a very good understanding and more complete picture of the gene regulatory networks that contribute to complex disease. Such samples allow for studies that associate phenotypes to risk variants previously identified as associated with a disease or trait (so-called PheWAS analyses) (Denny et al., 2010). Notably, in a very short period of time, the UK Biobank dataset (Sudlow et al., 2015) of 500,000 deeply phenotyped and SNP-genotyped individuals has proven to be a phenomenal resource of discovery of new genetic associations for quantitative traits, as well as very common diseases such as type 2 diabetes and depression—exactly in line with *a priori* predictions. Therefore, new gene discoveries for common disease will follow once powered in the same way as the UK Biobank, which as a community sample has at most 5,000 cases of a disease with lifetime risk of 1%. Accumulation of larger samples is also needed for patient stratification, which needs to progress in parallel with biological research.

Patient Stratification Is a Limiting Factor for Precision Medicine

The ultimate goal of research into common complex genetic diseases is to improve outcomes for those individuals affected through prevention, diagnosis, and treatment. Many research directions are needed to achieve this goal, and the Boyle et al. omnigenic core gene Perspective has been widely interpreted as a call for a research focus on cell-specific gene regulatory networks. We concur, and indeed, this work is already underway in many disease areas.

However, while the results of gene discovery from human samples feeds into research to identify treatments, it is prudent to prepare, in parallel, for the patient stratification that is likely needed for new (or repositioned) drugs to pass the hurdles for approved use. The goal of precision medicine is to tailor treatments to individuals (Collins and Varmus, 2015). In cancer, progress toward this goal has been made based on genomic profiles of tumors (Schram et al., 2017). For other complex genetic diseases, it is not clear what criteria will be used to determine allocation of patient groups to treatment options. Heterogeneity in clinical presentations and variation between individuals in response to treatment options imply that we need better ways to map the unique portfolio of risk alleles harbored by individuals to phenotypic presentations. The realities of data collection (inherent even in model species; Lucanic et al., 2017; Mott, 2015) can make it difficult to generate data not confounded by site of collection on the scale needed for interpretation. Biomarkers are proposed as a route to achieve biologically based phenotyping, particularly for disorders of the brain, where the tissue of most interest is hard to study. However, despite many research dollars spent on hypothesis-driven candidates, reproducibility and utility of such biomarkers seems elusive (Venkatasubramanian and Keshavan, 2016). New technologies may provide opportunities for assessment of genome-wide, hypothesis-free biomarkers.

In their Perspective, Boyle et al. called for an emphasis on developing cell-based models to enable study of key aspects of complex traits. We agree that there are deep challenges to fully understanding the impact of very small effects in organismal systems, so there is a need to develop cell-based model systems that can recapitulate aspects of complex traits. Our concurrence is both from the viewpoint of developing a better understanding of the biological complexity underpinning disease and from the viewpoint of penetrating the heterogeneity between individuals predicted by polygenicity. We need high-throughput scale platforms to characterize vulnerable human genomes (i.e., ascertained as cases) for a battery of measures. Technological advances in cellular reprogramming now approach the cost-achievable possibility of integrating multiple disease-specific cell types in complex synthetic human tissues (Fantuzzo et al., 2017; Junaid et al., 2017; Wevers et al., 2016). These models can mimic disease pathology by integration with physiological or pathophysiological stressors, inflammatory cytokines and inflammatory cells, bacterial or viral challenges, or a wide range of experimental perturbations (chemical screens, CRISPR mutations, etc.). The quantitative genetics community should contribute to experimental design as soon as these technologies can be applied on a high-sample-size scale, since an understanding of genetic variation will be critical to progressing applications in complex disease. The resulting models need to generate technically reproducible cellular phenotypes, but with genetically controlled variation between individuals (i.e., high heritability) (Hoffman and Brennand, 2018; Schwartzentruber et al., 2018). High-throughput cellular phenotyping may generate the data needed to allow stratification of patients who will then need full phenotyping to help interpretation of that stratification at the clinical level. At least, and perhaps more realistically, such data will generate the next set of questions that we do not yet know we need to ask.

Conclusion

In conclusion, Boyle et al. are congratulated for their synthesis of current data and for articulation of a biological framework that has prompted extensive constructive discussion. We agree that understanding the cell-specific role of disease-associated variants is a crucial step for advancing knowledge of common disease. However, whereas those authors extrapolate results of analyses of GWAS summary statistics to make fundamental assumptions that rare variants of large effect in a small number of genes play the most critical roles in clinical conditions that attract a common disease diagnosis, we believe it would be a major disservice to the field to allow these assumptions to guide the next steps of research. To assume that a limited number of core genes are key to our understanding of common disease may underestimate the true biological complexity, which is better represented by systems genetics and network approaches (Baliga et al., 2017; Parikshak et al., 2015). While Boyle et al. advocate for WES studies, they did not discuss the sample sizes needed for such discovery. We believe that in the short term, large samples recorded for key measures of phenotypic heterogeneity and genome-wide SNP data are the best next steps for research using human DNA samples in moving forward our understanding of complex genetic diseases. Large numbers of samples, biobanked for cellular reprogramming, will position us well for the next generation of sequencing and other new technologies. High-throughput phenotyping to characterize cellular properties associated with disease-associated genomes may be the key to penetrate the polygenic complexity of common disease and provide the data needed for patient stratification, as well as to progress toward the goal of new drug treatments. These are research paths that need to advance in parallel to advance the promise of precision medicine.

ACKNOWLEDGMENTS

We thank members of The University of Queensland Program in Complex Trait Genomics group, Bill Hill, Nick Martin, Georgia Chenevix-Trench, and Alexander Arguello for commenting on a draft. We acknowledge funding from the Australian National Health and Medical Research Council (1078901, 1113400, 1087889, 1078037; N.R.W., J.Y., P.M.V.), Charles & Sylvia Viertel Foundation (J.Y.), ERC advanced grant (2012-322698; C.W.) and Spinoza grant (NWO SPI 92-266; C.W.). We thank Yan Holtz (www.r-graph-gallery.com/all-graphs/) for making the figure.

REFERENCES

Antonarakis, S.E., and Beckmann, J.S. (2006). Mendelian disorders deserve more attention. Nat. Rev. Genet. *7*, 277–282.

Baliga, N.S., Björkegren, J.L., Boeke, J.D., Boutros, M., Crawford, N.P., Dudley, A.M., Farber, C.R., Jones, A., Levey, A.I., Lusis, A.J., et al. (2017). The State of Systems Genetics in 2017. Cell Syst. *4*, 7–15.

Barabási, A.L., Gulbahce, N., and Loscalzo, J. (2011). Network medicine: a network-based approach to human disease. Nat. Rev. Genet. *12*, 56–68.

Boyle, E.A., Li, Y., and Pritchard, J.K. (2017a). The Omnigenic Model: Response from the Authors. J. Psychiatry Brain Sci. *2*, S8.

Boyle, E.A., Li, Y.I., and Pritchard, J.K. (2017b). An Expanded View of Complex Traits: From Polygenic to Omnigenic. Cell *169*, 1177–1186.

Callaway, E. (2017). Genome studies attract criticism. Nature *546*, 463.

Caulfield, T., and Murdoch, B. (2017). Genes, cells, and biobanks: Yes, there's still a consent problem. PLoS Biol. *15*, e2002654.

Chakravarti, A., and Turner, T.N. (2016). Revealing rate-limiting steps in complex disease biology: The crucial importance of studying rare, extreme-phenotype families. BioEssays *38*, 578–586.

Civelek, M., and Lusis, A.J. (2014). Systems genetics approaches to understand complex traits. Nat. Rev. Genet. *15*, 34–48.

Collins, F.S., and Varmus, H. (2015). A new initiative on precision medicine. N. Engl. J. Med. *372*, 793–795.

Cox, N.J. (2017). Comments on Pritchard Paper. J. Psychiatry Brain Sci. *2*, S5.

Crow, J.F., and Kimura, M. (1970). An introduction to population genetics theory (New York: Harper & Row).

Deciphering Developmental Disorders, S.; Deciphering Developmental Disorders Study (2017). Prevalence and architecture of de novo mutations in developmental disorders. Nature *542*, 433–438.

Denny, J.C., Ritchie, M.D., Basford, M.A., Pulley, J.M., Bastarache, L., Brown-Gentry, K., Wang, D., Masys, D.R., Roden, D.M., and Crawford, D.C. (2010). PheWAS: demonstrating the feasibility of a phenome-wide scan to discover gene-disease associations. Bioinformatics *26*, 1205–1210.

Dunnington, E.A., Honaker, C.F., McGilliard, M.L., and Siegel, P.B. (2013). Phenotypic responses of chickens to long-term, bidirectional selection for juvenile body weight–historical perspective. Poult. Sci. *92*, 1724–1734.

Fantuzzo, J.A., De Filippis, L., McGowan, H., Yang, N., Ng, Y.H., Halikere, A., Liu, J.J., Hart, R.P., Wernig, M., Zahn, J.D., and Pang, Z.P. (2017). μNeurocircuitry: Establishing *in vitro* models of neurocircuits with human neurons. Technology (Singap World Sci) *5*, 87–97.

Faraone, S.V. (2017). The Omnigenic Model: Implications for Psychiatric Genetics. J. Psychiatry Brain Sci. *2*, S1.

Franke, B. (2017). What's in a Name: The "Omnigenic" Model as a Point of Departure from Polgenic Psychiatric Disorders. J. Psychiatry Brain Sci. *2*, S7.

Frith, M.C., and Khan, S. (2018). A survey of localized sequence rearrangements in human DNA. Nucleic Acids Res. *46*, 1661–1673.

Fuchsberger, C., Flannick, J., Teslovich, T.M., Mahajan, A., Agarwala, V., Gaulton, K.J., Ma, C., Fontanillas, P., Moutsianas, L., McCarthy, D.J., et al. (2016). The genetic architecture of type 2 diabetes. Nature *536*, 41–47.

Gandal, M.J., Leppa, V., Won, H., Parikshak, N.N., and Geschwind, D.H. (2016). The road to precision psychiatry: translating genetics into disease mechanisms. Nat. Neurosci. *19*, 1397–1407.

Genetic Modifiers of Huntington's Disease (GeM-HD) Consortium (2015). Identification of Genetic Factors that Modify Clinical Onset of Huntington's Disease. Cell *162*, 516–526.

Genovese, G., Fromer, M., Stahl, E.A., Ruderfer, D.M., Chambert, K., Landén, M., Moran, J.L., Purcell, S.M., Sklar, P., Sullivan, P.F., et al. (2016). Increased burden of ultra-rare protein-altering variants among 4,877 individuals with schizophrenia. Nat. Neurosci. *19*, 1433–1441.

Gottesman, I.I., and Shields, J. (1967). A polygenic theory of schizophrenia. Proc. Natl. Acad. Sci. USA *58*, 199–205.

Grönberg, H., Adolfsson, J., Aly, M., Nordström, T., Wiklund, P., Brandberg, Y., Thompson, J., Wiklund, F., Lindberg, J., Clements, M., et al. (2015). Prostate cancer screening in men aged 50-69 years (STHLM3): a prospective population-based diagnostic study. Lancet Oncol. *16*, 1667–1676.

Hansen, T.F., Alvarez-Castro, J.M., Carter, A.J., Hermisson, J., and Wagner, G.P. (2006). Evolution of genetic architecture under directional selection. Evolution *60*, 1523–1536.

Havrilla, J.M., Pedersen, B.S., Layer, R.M., and Quinlan, A.R. (2017). A map of constrained coding regions in the human genome. bioRxiv. https://doi.org/10.1101/220814.

Hensman Moss, D.J., Pardiñas, A.F., Langbehn, D., Lo, K., Leavitt, B.R., Roos, R., Durr, A., Mead, S., Holmans, P., Jones, L., and Tabrizi, S.J.; TRACK-HD investigators; REGISTRY investigators (2017). Identification of genetic variants associated with Huntington's disease progression: a genome-wide association study. Lancet Neurol. *16*, 701–711.

Hill, W.G. (2010). Understanding and using quantitative genetic variation. Philos. Trans. R. Soc. Lond. B Biol. Sci. *365*, 73–85.

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.

Hoffman, G.E., and Brennand, K.J. (2018). Mapping regulatory variants in hiPSC models. Nat. Genet. *50*, 1–2.

Jin, S.C., Homsy, J., Zaidi, S., Lu, Q., Morton, S., DePalma, S.R., Zeng, X., Qi, H., Chang, W., Sierant, M.C., et al. (2017). Contribution of rare inherited and de novo variants in 2,871 congenital heart disease probands. Nat. Genet. *49*, 1593–1601.

Junaid, A., Mashaghi, A., Hankemeier, T., and Vulto, P. (2017). An end-user perspective on Organ-on-a-Chip: Assays and usability aspects. Curr. Opin. Biomed. Eng. *1*, 15–22.

Kemper, K.E., Visscher, P.M., and Goddard, M.E. (2012). Genetic architecture of body size in mammals. Genome Biol. *13*, 244.

Kiezun, A., Garimella, K., Do, R., Stitziel, N.O., Neale, B.M., McLaren, P.J., Gupta, N., Sklar, P., Sullivan, P.F., Moran, J.L., et al. (2012). Exome sequencing and the genetic basis of complex traits. Nat. Genet. *44*, 623–630.

Leek, J.T., Scharpf, R.B., Bravo, H.C., Simcha, D., Langmead, B., Johnson, W.E., Geman, D., Baggerly, K., and Irizarry, R.A. (2010). Tackling the widespread and critical impact of batch effects in high-throughput data. Nat. Rev. Genet. *11*, 733–739.

Levinson, D.F., Mostafavi, S., Milaneschi, Y., Rivera, M., Ripke, S., Wray, N.R., and Sullivan, P.F. (2014). Genetic studies of major depressive disorder: why are there no genome-wide association study findings and what can we do about it? Biol. Psychiatry *76*, 510–512.

Lips, E.S., Cornelisse, L.N., Toonen, R.F., Min, J.L., Hultman, C.M., Holmans, P.A., O'Donovan, M.C., Purcell, S.M., Smit, A.B., Verhage, M., et al.; International Schizophrenia Consortium (2012). Functional gene group analysis identifies synaptic gene groups as risk factor for schizophrenia. Mol. Psychiatry *17*, 996–1006.

Lucanic, M., Plummer, W.T., Chen, E., Harke, J., Foulger, A.C., Onken, B., Coleman-Hulbert, A.L., Dumas, K.J., Guo, S., Johnson, E., et al. (2017). Impact of genetic background and experimental reproducibility on identifying chemical compounds with robust longevity effects. Nat. Commun. *8*, 14256.

Luo, Y., de Lange, K.M., Jostins, L., Moutsianas, L., Randall, J., Kennedy, N.A., Lamb, C.A., McCarthy, S., Ahmad, T., Edwards, C., et al. (2017). Exploring the genetic architecture of inflammatory bowel disease by whole-genome sequencing identifies association at ADCY7. Nat. Genet. *49*, 186–192.

Lynch, M., and Walsh, B. (1998). Genetics and Analysis of Quantitative Traits (Sunderland, Massachusetts, USA: Sinauer Associates).

MacArthur, D.G., Balasubramanian, S., Frankish, A., Huang, N., Morris, J., Walter, K., Jostins, L., Habegger, L., Pickrell, J.K., Montgomery, S.B., et al.; 1000 Genomes Project Consortium (2012). A systematic survey of loss-of-function variants in human protein-coding genes. Science *335*, 823–828.

Mahajan, A., Taliun, D., Thurner, M., Robertson, N.R., Torres, J.M., Rayner, N.W., Steinthorsdottir, V., Scott, R.A., Grarup, N., Cook, J.P., et al. (2018). Fine-mapping of an expanded set of type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. bioRxiv. https://doi.org/10.1101/245506.

Mäki-Tanila, A., and Hill, W.G. (2014). Influence of gene interaction on complex trait variation with multilocus models. Genetics *198*, 355–367.

Marouli, E., Graff, M., Medina-Gomez, C., Lo, K.S., Wood, A.R., Kjaer, T.R., Fine, R.S., Lu, Y., Schurmann, C., Highland, H.M., et al.; EPIC-InterAct Consortium; CHD Exome+ Consortium; ExomeBP Consortium; T2D-Genes Consortium; GoT2D Genes Consortium; Global Lipids Genetics Consortium; ReproGen Consortium; MAGIC Investigators (2017). Rare and low-frequency coding variants alter human adult height. Nature *542*, 186–190.

Masel, J., and Siegal, M.L. (2009). Robustness: mechanisms and consequences. Trends Genet. *25*, 395–403.

McCarthy, S., Das, S., Kretzschmar, W., Delaneau, O., Wood, A.R., Teumer, A., Kang, H.M., Fuchsberger, C., Danecek, P., Sharp, K., et al.; Haplotype Reference Consortium (2016). A reference panel of 64,976 haplotypes for genotype imputation. Nat. Genet. *48*, 1279–1283.

Mott, R. (2015). Genetic differential calculus. Nat. Genet. *47*, 965–966.

Nelson, M.R., Tipney, H., Painter, J.L., Shen, J., Nicoletti, P., Shen, Y., Floratos, A., Sham, P.C., Li, M.J., Wang, J., et al. (2015). The support of human genetic evidence for approved drug indications. Nat. Genet. *47*, 856–860.

Oram, R.A., Patel, K., Hill, A., Shields, B., McDonald, T.J., Jones, A., Hattersley, A.T., and Weedon, M.N. (2016). A Type 1 Diabetes Genetic Risk Score Can Aid Discrimination Between Type 1 and Type 2 Diabetes in Young Adults. Diabetes Care *39*, 337–344.

Orr, H.A., and Coyne, J.A. (1992). The genetics of adaptation: a reassessment. Am. Nat. *140*, 725–742.

Paixão, T., and Barton, N.H. (2016). The effect of gene interactions on the long-term response to selection. Proc. Natl. Acad. Sci. USA *113*, 4422–4427.

Pardiñas, A.F., Holmans, P., Pocklington, A.J., Escott-Price, V., Ripke, S., Carrera, N., Legge, S.E., Bishop, S., Cameron, D., Hamshere, M.L., et al. (2018). Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. Nat. Genet. *50*, 381–389.

Parikshak, N.N., Gandal, M.J., and Geschwind, D.H. (2015). Systems biology and gene networks in neurodevelopmental and neurodegenerative disorders. Nat. Rev. Genet. *16*, 441–458.

Pasaniuc, B., and Price, A.L. (2017). Dissecting the genetics of complex traits using summary association statistics. Nat. Rev. Genet. *18*, 117–127.

Penrose, L.S. (1953). The genetical background of common diseases. Acta Genet. Stat. Med. *4*, 257–265.

Phillips, P.C. (2008). Epistasis–the essential role of gene interactions in the structure and evolution of genetic systems. Nat. Rev. Genet. *9*, 855–867.

Polderman, T.J., Benyamin, B., de Leeuw, C.A., Sullivan, P.F., van Bochoven, A., Visscher, P.M., and Posthuma, D. (2015). Meta-analysis of the heritability of human traits based on fifty years of twin studies. Nat. Genet. *47*, 702–709.

Pritchard, J.K. (2001). Are rare variants responsible for susceptibility to complex diseases? Am. J. Hum. Genet. *69*, 124–137.

Pritchard, J.K., and Cox, N.J. (2002). The allelic architecture of human disease genes: common disease-common variant...or not? Hum. Mol. Genet. *11*, 2417–2423.

International Schizophrenia Consortium, Purcell, S.M., Wray, N.R., Stone, J.L., Visscher, P.M., O'Donovan, M.C., Sullivan, P.F., and Sklar, P. (2009). Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. Nature *460*, 748–752.

Qi, T., Wu, Y., Zeng, J., Zhang, F., Xue, A., Jiang, L., Zhu, Z., Kemper, K., Yengo, L., Zheng, Z., et al. (2018). Identifying gene targets for brain-related traits using transcriptomic and methylomic data from blood. bioRxiv. https://doi.org/10.1101/274472.

Reich, D.E., and Lander, E.S. (2001). On the allelic spectrum of human disease. Trends Genet. *17*, 502–510.

Rosenson, R.S., Brewer, H.B., Jr., Barter, P.J., Björkegren, J.L.M., Chapman, M.J., Gaudet, D., Kim, D.S., Niesor, E., Rye, K.A., Sacks, F.M., et al. (2018). HDL and atherosclerotic cardiovascular disease: genetic insights into complex biology. Nat. Rev. Cardiol. *15*, 9–19.

Samocha, K.E., Kosmicki, J.A., Karczewski, K.J., O'Donnell-Luria, A.H., Pierce-Hoffman, E., MacArthur, D.G., Neale, B.M., and Daly, M.J. (2017). Regional missense constraint improves variant deleteriousness prediction. bioRxiv. https://doi.org/10.1101/148353.

Schizophrenia Working Group of the Psychiatric Genomics Consortium (2014). Biological insights from 108 schizophrenia-associated genetic loci. Nature *511*, 421–427.

Schram, A.M., Berger, M.F., and Hyman, D.M. (2017). Precision oncology: Charting a path forward to broader deployment of genomic profiling. PLoS Med. *14*, e1002242.

Schwartzentruber, J., Foskolou, S., Kilpinen, H., Rodrigues, J., Alasoo, K., Knights, A.J., Patel, M., Goncalves, A., Ferreira, R., Benn, C.L., et al.; HIPSCI Consortium (2018). Molecular and functional variation in iPSC-derived sensory neurons. Nat. Genet. *50*, 54–61.

Sekar, A., Bialas, A.R., de Rivera, H., Davis, A., Hammond, T.R., Kamitaki, N., Tooley, K., Presumey, J., Baum, M., Van Doren, V., et al.; Schizophrenia Working Group of the Psychiatric Genomics Consortium (2016). Schizophrenia risk from complex variation of complement component 4. Nature *530*, 177–183.

Short, P.J., McRae, J.F., Gallone, G., Sifrim, A., Won, H., Geschwind, D.H., Wright, C.F., Firth, H.V., FitzPatrick, D.R., Barrett, J.C., and Hurles, M.E. (2018). De novo mutations in regulatory elements in neurodevelopmental disorders. Nature *555*, 611–616.

Sivakumaran, S., Agakov, F., Theodoratou, E., Prendergast, J.G., Zgaga, L., Manolio, T., Rudan, I., McKeigue, P., Wilson, J.F., and Campbell, H. (2011). Abundant pleiotropy in human complex diseases and traits. Am. J. Hum. Genet. *89*, 607–618.

Slatkin, M. (2008). Exchangeable models of complex inherited diseases. Genetics *179*, 2253–2261.

Stern, D.L., and Orgogozo, V. (2008). The loci of evolution: how predictable is genetic evolution? Evolution *62*, 2155–2177.

Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., Danesh, J., Downey, P., Elliott, P., Green, J., Landray, M., et al. (2015). UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. PLoS Med. *12*, e1001779.

Sullivan, P.F., Daly, M.J., and O'Donovan, M. (2012). Genetic architectures of psychiatric disorders: the emerging picture and its implications. Nat. Rev. Genet. *13*, 537–551.

Thomas, R.H., and Berkovic, S.F. (2014). The hidden genetics of epilepsy-a clinically important new paradigm. Nat. Rev. Neurol. *10*, 283–292.

Treangen, T.J., and Salzberg, S.L. (2011). Repetitive DNA and next-generation sequencing: computational challenges and solutions. Nat. Rev. Genet. *13*, 36–46.

Venkatasubramanian, G., and Keshavan, M.S. (2016). Biomarkers in Psychiatry - A Critique. Ann. Neurosci. *23*, 3–5.

Wevers, N.R., van Vught, R., Wilschut, K.J., Nicolas, A., Chiang, C., Lanz, H.L., Trietsch, S.J., Joore, J., and Vulto, P. (2016). High-throughput compound evaluation on 3D networks of neurons and glia in a microfluidic platform. Sci. Rep. *6*, 38856.

Yang, J., Bakshi, A., Zhu, Z., Hemani, G., Vinkhuyzen, A.A., Lee, S.H., Robinson, M.R., Perry, J.R., Nolte, I.M., van Vliet-Ostaptchouk, J.V., et al.; LifeLines Cohort Study (2015). Genetic variance estimation with imputed variants finds negligible missing heritability for human height and body mass index. Nat. Genet. *47*, 1114–1120.

Zhu, Z., Bakshi, A., Vinkhuyzen, A.A., Hemani, G., Lee, S.H., Nolte, I.M., van Vliet-Ostaptchouk, J.V., Snieder, H., Esko, T., Milani, L., et al.; LifeLines Cohort Study (2015). Dominance genetic variation contributes little to the missing heritability for human complex traits. Am. J. Hum. Genet. *96*, 377–385.

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. USA *109*, 1193–1198.