Rapporteurs: Melpomeni Kasapi and Taylorlyn Stephan, NHGRI

Missing Heritability, Ten Years on
May 1-2, 2018 - Silver Spring, MD
Meeting Summary

Welcome / Objectives

The purpose of this meeting was to review the scientific progress in the area of missing heritability and identify: 1) what we have learned since the first missing heritability workshop in 2008, 2) what has been and/or will be the value of identifying the sources of missing heritability, and 3) what research can and/or should be pursued to determine these sources.

Impact of missing heritability publication in *Nature* (Orli Bahcall)

Orli Bahcall addressed the impact of missing heritability by pointing out that the concept arose as genome-wide association studies (GWAS) were increasingly published after 2005; reviews of the usefulness of GWAS were not always favorable. Since then, however, there has been a lot of discovery in gene-disease associations and the GWAS method has been the driver.

The review produced by the first Missing Heritability workshop has been cited more than 4,000 times and the field-weighted citation impact is comparable to other highly-cited reviews published by *Nature*. This has sparked the generation of more reviews addressing missing heritability and bringing attention to the issue for others in the field. We need to start thinking of how this missing heritability problem has influenced research directions and whether it has changed the public perception of human genetics.

Missing heritability circa 2009 (Teri Manolio)

During the first Missing Heritability workshop in 2008, the GWAS method was about 3 years old and had identified hundreds of associated variants. Most of those conferred small increments in risk and explained only a small portion of heritability (h^2). As an example, 40 loci for height explained 5% of the phenotypic variance, but the estimated h^2 was about 80%, raising the question of how to identify sources of this missing heritability. Proposed explanations included that there could be a much larger number of variants of smaller effect, rarer variants (with possibly larger effects), structural variants (SVs) poorly detected by available arrays, potential gene-gene (gxg) and gene-environment (gxe) interactions and correlations, inadequate accounting for shared environment, and over-estimation of h^2. Since then there has been much progress in exploring these explanations. We need best approaches for combining functional and statistical evidence, using common SNPs to predict and control for differences in rare SNPs, and pooling variants by classes and minor allele frequencies (MAF). Accounted-for heritability has increased since 2008 with increasing numbers of alleles discovered to influence a trait. Methods have improved, including more accurate algorithms for SV detection. The NHGRI-EBI GWAS catalog (https://www.ebi.ac.uk/gwas/) shows that in 2008, there weren't any known associated rare alleles ($MAF < 0.005$); 12 high-effect common variants ($MAF > 0.05$) were known, and only 4 low frequency ($0.005 < MAF \leq 0.05$) and moderately high odds ratio ($1.5 < OR \leq 3.0$) variants were identified. By April 2018 these numbers were much higher with about 24 rare alleles identified, 142 low-frequency moderate effect variants, 111 high-effect ($OR > 3.0$) common variants, and almost 5,000 common low effect (< 1.5) variants. Most rare variants were discovered in sequencing studies and not through GWAS.

Quantifying the genetic architecture and heritability of complex traits: estimation and prediction (Peter Visscher)

Yang et al. in 2010 *Nature Genetics* showed the substantial amount of variance captured by low effect-size loci. Most of the h^2 is not missing but has not been detected because the individual effects were too small to pass stringent significance tests. About 45% of variance in height, for example, can be explained if all genotyped SNPs are considered simultaneously, irrespective of statistical significance. In addition, as GWAS sample sizes increase, more genome-wide significant (GWS) loci are discovered. New data generated since 2009 include GWAS summary statistics, larger GWAS, transcriptional and epigenetic resources, fully sequenced reference panels with imputation accuracy down to MAF = 0.005, and large single-cohort studies. New methods include GREML (estimating genetic variation without hypothesis testing), linkage disequilibrium (LD) score regression, and prediction methods.

Height has been a well-studied trait for assessing explained variance. Mendelian forms of "tallness" and "shortness" exist, but most variation is polygenic. In partitioning the variance of height, heritability based on twin or family studies is 80%, within-family estimates are 70%, SNP heritability from imputation to sequenced reference is 60%, SNP-heritability (variance explained by all genotyped SNPs on an array) is 45%, and variance explained by GWS SNPs is 25%. Hidden h^2, if the SNP-h^2 estimates are correct, is generally due to lack of study power. Overestimation, untagged variants and disease heterogeneity are other possible explanations for the missing h^2. The key experiment (analysis) that is needed to better understand the gap between SNP-heritability and estimates of heritability from family data is to estimate heritability from a sufficiently large sample of individuals with WGS data. If such an analysis recovers the full pedigree heritability from WGS data then there is no more missing heritability and the problem can be considered as 'solved'.

Further issues include the difference between within-family and population estimates of SNP effects. Population samples suffer from population stratification, gene-environment (G-E) correlation, and assortative mating. Prediction from DNA sequence (or imputed SNP arrays) is limited by how much phenotypic variance is captured by all variants and how well the effects of all variants are estimated. GWAS data show little evidence for non-additive genetic variance when investigated at genome-wide-significant (GWS) loci or when estimating dominance variance from all SNPs simultaneously. However, the loss of information due to imperfect LD is larger for non-additive effects than it is for additive effects, and this can contribute to a lack of power to detect gxg interactions from population data.

Model organisms could help infer environmental impact on missing heritability. Genes shape behavior and behavior profoundly affects environmental factors, therefore environmental tracking studies should be integrated with genetic studies and not separated. The UK Biobank is a good example of an integrative study that collected exercise and other lifestyle information. The *All of Us* cohort project also plans on using wearable devices and biomarkers to improve measurement of environmental factors.

Discussion

If all the SNPs in a genome are implicated, unless all effects are infinitesimal, then they are diluting the impact of the causative SNPs. Heritability estimation should be considered in the context of prediction;

there isn't a good way of ensuring that the identified rare variants of large effect are actually causative. Whole genome sequencing (WGS) can be used to estimate genetic variance in addition to discovering variants. In addition, although GWAS is now over 12 years old, there's still very little information from non-European populations. Expanding GWAS to diverse populations will allow for novel discoveries and strengthen the imputation.

SNP-based heritability is based on some measure of how phenotypically similar pairs of individuals in a population are, even without a known pedigree relationship. It might be useful to compare findings across the whole range, from studies with unrelated individuals to traditional studies with twins and families, and look at the trends.

Relatively little work has been done on the genetics of disease progression, though there have been some efforts in diabetes and renal disease. One of the main challenges in this area is getting phenotypic information on patients and generating phenotypic algorithms to predict disease progression. For these algorithms to show accurate results we need to factor in the environment, transcriptome, genetic variance, and genetic effects modified by the onset of the disease. The genetic factors contributing to variation between people for disease progression and for response to treatment may be different from the genetic factors contributing to variation between people in disease onset.

In the early days of GWAS there were efforts to study heritability in affected siblings, but if parents are genotyped analyses are much more powerful. To compare estimates from siblings with SNP/pedigree estimates, larger sample sizes are needed. UK Biobank currently includes about 80,000 sib pairs, which for many traits is sufficient, however for any particular trait there is need for more samples.

Quantifying (missing) heritability for common disease from GWAS data (Naomi Wray)

In psychiatric disorders the difference between SNP-$h^2$and pedigree-h^2 is much larger than the difference seen in quantitative traits. Subtle confounding also has a greater impact on discrete than continuous traits. If the missing h^2 is greater for binary disease traits than for quantitative traits despite the same technology, then the assumptions of the methodology are not upheld, there are technical artifacts, and/or assumptions about genetic architecture are incorrect. A study of amyotrophic lateral sclerosis (ALS) in 171 twin pairs estimated h^2 at 61%, for example, but a similar diagnosis in a second twin is more likely once their co-twin is diagnosed. Schizophrenia has been studied more extensively with twin-based h^2 estimates up to 81%. In contrast, a study of common genetic determinants of schizophrenia using 9 million Swedish national records estimated h^2 at 64-67%. The LD score regression method likely underestimates h^2 in disease traits but is helpful as a quick benchmark.

Key challenges in binary disease datasets are restricted sample sizes of disease cases vs. controls and greater variability in their ascertainment. Additional challenges include overestimation of h^2 in pedigrees and a tendency to infer a greater level of accuracy than the data deserve. Residual population stratification may be confounding case-control comparisons, and disease-specific architectures increase the complexity of defining h^2, as do unknown genetic heterogeneity and polygenicity.

For common diseases and disorders the difference between SNP-$h^2$and pedigree-h^2 is larger than the difference seen in quantitative traits, despite the same genotyping and imputation strategies. Possible explanations for this observation include technical artifacts, the assumptions of the methodology are not upheld, and/or the genetic architectures of common disease are different to those of binary traits. Technical artifacts are more likely in disease traits than continuous traits as subtle confounding between the binary values of the trait and genotyping experimental design or population stratification may be difficult to resolve fully through statistical analysis. Assumptions of methodology that could contribute to the difference between SNP-$h^2$and pedigree-h^2 may reflect that estimates of heritability for disease are likely less reliable than those for quantitative traits. For example, the estimate of heritability usually quoted for ALS of 61% is estimated from 171 twin pairs. Schizophrenia has been studied more extensively and usually heritability is quoted as 81% from a meta-analysis of mostly twin-based estimates. However, a study of common genetic determinants of schizophrenia using 9 million Swedish national records estimated h^2 at 64-67%, with a similar estimate made from Danish national records. Heritability estimates (both family-based and SNP-heritability) are made on the liability scale, which includes methodological assumptions that may not be upheld. The LD score regression method likely underestimates h^2 in disease traits but is useful as a quick benchmark. Haseman-Elston methodology is the most robust method for estimation of SNP-h^2 when there is very extreme ascertainment.

The greater difference between pedigree h^2 and SNP-h^2 estimates for disease compared to quantitative traits may also reflect true differences in genetic architecture. One explanation could be a greater contribution from rare variants, but whole exome sequencing studies (which have been highly successful in identifying rare variants of large effect in severe childhood syndromes) have been less successful in similarly powered studies for common disorders (psychiatric disorders, type 2 diabetes, inflammatory bowel diseases), implying that rare variant effect sizes will also be small. Rare variants of large effect likely lead to more severe childhood diagnoses than to common disorders with onset typically later in life. Another explanation for the greater difference between pedigree- and SNP-h^2 estimates for disease compared to quantitative traits may be the heterogeneity in clinical diagnoses, in which multiple (but likely correlated) biological routes could lead to clinical presentations that attract the same diagnosis. This is perhaps the most important explanation to investigate in future studies and it may have downstream consequences for precision medicine, i.e., the stratification of patients to drug treatments.

Whatever the explanation of the greater difference between pedigree- and SNP-h^2 estimates for common diseases vs quantitative traits, larger samples with more detailed phenotyping and with consistent phenotyping across cohorts will help to resolve them. Continuing to focus on increasing sample size is a priority for common diseases.

Rare variants (David Goldstein)

Clinical diagnostic sequencing can be used as a paradigm for studying rare variants in Mendelian diseases. The diagnostic analysis framework includes identifying rare and functional variation in genes having known associations with disease, looking for extremely rare sequence variants of high technical quality that preferably are previously reported pathogenic or at the same or adjacent genomic sites, or that show loss of function (LoF) in genes known to be depleted for LoF variants. The methodology is

remarkably effective, identifying causative variants in roughly a quarter of patients even when their conditions are considered genetically complex. Often causative variants can be identified from extreme rarity in population databases without needing to sequence the parents and infer their *de novo* origin. In general, association signals are much stronger in individuals with known positive family histories than those without, which isn't surprising, but that it's so dramatically less in "sporadic" cases is surprising as many of these are probably familial and just don't have extensive family information. Signal comes entirely from the rarest variants that appear to have been kept out of the population, with no continuum between rare and common variation. This implies they're all very recent and restricted to small families, which has significant implications for precision medicine.

High effect alleles depend on the disease. Some diseases have an early-onset and thus more evolutionary pressure due to which risk variants have increased in prevalence. Sequencing in presumed common, complex diseases such as chronic kidney disease can identify a small but significant proportion with monogenic etiologies such as Alport's syndrome, which has profound implications for treatment and screening of family members. Similarly (though anecdotally), a patient with presumed non-alcoholic fatty liver disease was sequenced and found to have Wilson's disease, a highly treatable condition that had been missed clinically. The group agreed that Mendelian diseases are associated with a wide range of presentations and often the line between Mendelian and complex disease is hard to distinguish.

Monogenic contributions to complex traits: scaling pleiotropy (Judy Cho)
Health-system based biobanks have several advantages over the simplistic case vs. control model. Health systems have enormous "phenotype" data including labs, radiology and pathology data. Development of disease is time-dependent, and through these records we can look at progression and have data on many endpoints. Age-dependent prevalence of hypertension is an example of a disease that can be studied in these cohorts. Monogenic forms of hypertension are estimated to be 3-5% of all cases and a genetics-first approach is needed to identify the early cases.

Primary immunodeficiencies are under-diagnosed diseases with viral and bacterial infections that are very common. There is a likely continuum of genetic differences in the capacity to fight infections due to major evolutionary selection. Unusual phenotypic characteristics such as two major infections before age 50 in a non-alcoholic can help identify patients with primary immunodeficiencies. This group of diseases is defined by a crucial time element with decreased capacities at extremes of age. Pleiotropy may be expected or unexpected; *IL23R* demonstrates "expected" pleiotropy as an IBD (Crohn's disease (CD) and ulcerative colitis (UC)) gene. The majority of inflammatory bowel disease (IBD) loci show similar trends between CD vs. UC, but there is also unexpected pleiotropy with a strong protective association against tongue-tie. In cystic fibrosis (CF), a multi-organ recessive Mendelian disease where the chloride channel is defective, heterozygous *CFTR* carriers present with recurrent pancreatitis phenotypically distinct from CF patient manifestations, showing the need for systematic evaluation of these carriers.

Ashkenazi Jews have been studied for a variety of diseases. They show a 3-fold higher prevalence of IBD vs. non-Jewish Europeans due to a much higher functional effect size of *NOX1* variants in Jews, even though variants are present in both European populations. An example of unexpected pleiotropy is in

Jewish predominant mutations of the *LRRK2* gene. Protein-altering variants in *LRRK2* include distinct risk and common protective variants between Parkinson's and Crohn's.

Next steps for studying monogenic genes in complex traits should include a gene-centric view of disease pathogenesis, systematic analysis of biobank-based genetic data to provide specific medical context to phenotypic variability of monogenic or high-effect genetic variants, exploration of population differences, and domain-based sequence annotation. Increased appreciation of the ubiquity of pleiotropy should lead to investigation of modifying factors such as age and time, recognizing that phenome coverage is more limited than genome coverage.

Discussion

Pure exome rare variant studies have paved the way to some informative findings, but the issue is still very complex. 80-85% of GWAS signals are not in the coding region. Non-coding variants of large effect are rare due to the very extensive buffering capacity of cells/organisms. Modulation of gene expression is very contextual, complex, and difficult to interpret in an experiment. Tools for analyzing regulatory variants are really poor right now, the actual mutation target is very small, the mutational target space is very large, and there are very few bases that when mutated create a strong phenotype. The solution to this could come with perfecting and applying cell-based analysis.

The nature of the connection between genes and phenotype is still very hard to define. We might have to think about gene networks influencing disease rather than direct gene to phenotype pathways. Unexpected pleiotropy is to be expected; in many experimental designs, the more phenotypes you study the more pleiotropy you find.

Functional follow-up of findings is an unmet need. It can be accomplished by leveraging the functional domains. There is an undervaluation of domains in the field and deeper functional annotation of variants would be productive. Moreover, combining functional and statistical evidence has been anecdotal to date and the more we look the more we find. Systematic mutagenesis will allow for scaling and this will be facilitated by sequencing phenotype-linked biobanks.

Structural and multi-allelic variation (Steve McCarroll)

Structural variation presents in a variety of ways and has variable effects on heritability across the genome. But at individual loci, SVs explain 2-4x more of the variation than "lead SNPs". Large copy number variants are reasonably easy to detect and impute, while structurally unstable loci are subject to high evolutionary pressures, making them more challenging to analyze. Loci with recurring structural mutations can have many functionally distinct alleles. In schizophrenia, *C4* (complement component 4) *A* and *B* genes in the major histocompatibility (MHC) locus have recurring structural mutations and include an ancient retroviral insertion that acts as a brain-specific enhancer. The more *C4A* RNA expression an allele generates, the greater the risk of schizophrenia. As a second example, recurring exon deletions in haptoglobin (*HP*) alter the multimerization of HP and act to reduce blood cholesterol, particularly in synergy with a nearby SNP that regulates *HP.*

Difficult to copy parts of the genome will always have a higher proportion of SVs, but they may appear as microsatellites over timescales so long that they do not impact human health. Across the genome SVs do not appear to explain much of the missing h^2 because there appear to be at most 1 SV (with current knowledge) affecting any given disease. At individual loci, however, SVs may explain 2-4 times more variance than the "lead SNP."

An interesting new line of investigation is examining acquired somatic mutations as a source of missing h^2 because mosaic mutations cluster in genomic hotspots and it may be the tendency to somatic mutation that is inherited, making "fragile sites" much more fragile. eQTLs can be used in genome-wide searches to impute polygenic risk scores and sub-classify mutation mechanisms. When somatic mutations are added to stem cells in culture, the cells have a selected advantage of up to 2-fold in pluripotent stem cells. The dichotomy between inheritance and acquired mutations may thus not be as firm as thought. Despite the availability of ever-more-complex and expensive technologies for characterizing the genome, these are often too easy an excuse to abandon painstaking and consistent application of established forms of genetic analysis. Most of what has been learned to date has come from large, widely available SNP datasets because SNP data are available for so many people.

Omnigenic architecture of human complex traits (Jonathan Pritchard)

Key questions are why lead GWAS hits for a given trait contribute so little to h^2, and why so much of the genome appears to contribute to h^2. For example, schizophrenia has 108 genome-wide significant loci but they only explain ~10% of variation. Significant loci for low-density lipoprotein cholesterol (LDL), only explain ~20% of its h^2, but all known LDL loci cumulatively explain ~80%. Genes with trait-relevant functions typically contribute only small fractions of total disease risk, while low frequency/large effect variants often have clearer enrichment in trait- and disease-specific genes. Contributing variants are highly concentrated in regions of active chromatin in relevant tissues, suggest that most effects of low frequency/large effect variants are mediated through gene regulation.

The omnigenic model describes three types of genes:

- Tier 1: core genes that have direct roles in disease
- Tier 2: peripheral genes, or all other expressed genes that can trans-regulate core genes
- Tier 3: genes not expressed in cell types that do not contribute to heritability

Most phenotypic variance is due to regulatory variation in peripheral genes. It is hypothesized that peripheral genes outnumber core genes by 100:1, so they dominate h^2 by having effects in gene networks. About 70% of h^2 is expressed in *trans*-peripheral genes, while ~30% of mRNA h^2 is expressed in *cis*-core genes. Because *trans* eQTLs have small effect sizes compared to *cis* eQTLs, a typical gene must have many weak *trans*-regulators that contribute to its h^2. *Cis* effects are independent for each core gene, while *trans* effects are often shared across core genes. This means that most of the h^2 is transferred to peripheral genes. If core genes are highly correlated in the same network, when one increases its expression so do the others so peripheral effects end up dominating. *Trans* effects shared across co-regulated networks can thus act as amplifiers for peripheral variation. GWAS are thus telling

us something fundamental about how genetic variation affects phenotypes, and the paradigm of direct links from variant to phenotype is actually quite restricted.

Epigenetic effects and gene expression (Alexis Battle)

Given that the majority of trait-associated variation occurs in non-coding regions, and presumably functions by altering gene expression, gene expression and epigenetic data can be used to inform missing h^2. *Cis* eQTLs have been identified in nearly every human gene, and thus can be leveraged for h^2 studies, but *trans* eQTLs have thus far been poorly replicated and validated. They are believed to be underpowered for use in h^2 studies even though they contribute more to gene expression h^2. Most SNPs look like eQTLs in some tissue but most just tag functional variants, while ~50% of genetic variants implicated in human disease co-localize (appear to share the same causal variant) with an eQTL. *Trans*-eQTLs appear to be more tissue-specific than *cis*-eQTLs.

Disease-relevant states occur during different developmental stages, in response to variable environmental exposures, and in different cell types. eQTL data are needed from more diverse cell types, developmental stages, and environmental perturbations. Rare variants have been shown to drive extreme expression levels in individuals, but it remains uncertain what fraction of missing h^2 they explain. Analyzing gene expression polygenicity is a reasonable next step to calculate missing h^2, but will require large samples and meta-analyses of all available expression data, or other experimental approaches, particularly to find *trans* effects. Improving gene expression analysis will help determine how much more we need to invest in WGS and epigenetic data and will also power existing studies.

Discussion

Lack of power in studies does affect the *trans*-eQTL tissue-specificity. Power also affects our ability to discover *cis* and *trans* eQTLs. However, trans-eQTLs still display much greater tissue-specificity than cis-eQTLs with matched MAF and effect sizes. Some may also affect tissues and developmental timepoints that have not yet been assayed. These eQTLs specific to conditions such as immune response and stem-cell development are potentially missed due to not focusing on specific time intervals. We expected that when we found SNPs in pathways they'd ripple across cancer, but now 90% of SNPs or even loci in cancer haven't been seen in another cancer. Could this relate to developmental order, considering cancer as a process of unraveling development, and they're missed because we're not looking at development? There is an emerging eQTL analysis effort in tumor tissues that will become more available soon for the community to utilize.

The potential for confounding between *cis*- and *trans*-eQTLs due to the haplotype structure of the genome also needs consideration—could it be that *cis*-eQTLs are also *trans*-acting? There are no data to assess *trans* effects of *cis*-eQTLs. Most *cis*-eQTLs (only things cross-chromosomal are unambiguously *trans*) have allelic effects but the ability to call them is based on sequencing depth. If the omnigenic model were taken to the extreme, then every *cis*-eQTL would be *trans* to every other eQTL. In general, we do see more complex effects including feedback loops, target signaling, etc. After defining these effects, analyzing polygenicity of gene expression will be a good next direction. We should also start testing strawman models of regulation, particularly for traits with a relatively well-defined set of core

genes such as lipids. Core genes might be considered and evaluated as potential bottlenecks, one for which loss of function can't be compensated; would this be consistent with polygenicity?

How sex-specific, environment-specific and genetic background-specific effects generate missing heritability (Trudy Mackay)

Effects of variants affecting human complex traits may be small due to genotype-by-sex interaction (genetic variation in sexual dimorphism), genotype-by-environment interaction (genetic variation in environmental plasticity) and genotype-by-genetic background interaction (epistasis). *Drosophila* as a model organism fulfills many criteria needed to study these genetic interactions. The *D. melanogaster* Genetic Reference Panel (DGRP) includes 205 sequenced lines derived from a single natural population for genome wide association mapping in a scenario where all variants are known. DGRP analyses show that the genetic architecture of the *Drosophila* lifespan is dominated by sex- and environment-specific variants as well as epistasis. These variants have small effects on lifespan, averaged over both sexes and all environments, and may account in part for missing h^2 when context is not accounted for. With epistasis, effects of variants will not replicate between populations with different allele frequencies. Context-dependent variants may 'hide' from natural selection in natural populations experiencing heterogeneous environments, leading to maintenance of variation for lifespan in natural populations.

Gene-environment interaction (David Hunter)

Differences in rates of most diseases between countries (and over time within countries) are due to differences in environmental and "lifestyle" risk factors, not genetic differences. Differences in individual risk of most diseases within countries are due to differences in both genetic and environmental and "lifestyle" risk factors; therefore, there is need to measure both and see how they interact. With some exceptions (e.g., drug idiosyncrasies), genetic, environmental, and "lifestyle" risk factors are independent and the risks multiply. Despite interest in gene-environment interactions, there are few agreed-upon instances where the effect of exposure differs across genotypes (and vice versa). Reasons for these few true interactions could include poor measurement of genes, low power of studies, poor measurement of environment, and the possibility that there might not be many interactions to discover.

Environment-to-environment interactions are also important. These include: smoking/alcohol in esophageal cancer, BMI/menopausal status in breast cancer, postmenopausal hormones/BMI in breast cancer, aflatoxin/HBV in liver cancer, radiation/smoking in lung cancer, skin type/UV and skin cancer. Interactions that depart from the multiplicative model are the exception, not the rule.

In ten more years we have discovered few examples of synergy between genes and environment. Gene variants that dramatically alter drug metabolism can dramatically alter drug response and efficacy. Most genetic and environmental risk factors conform to the multiplicative model which makes risk prediction algorithms more stable. The multiplicative model implies that environmental risk reduction in those at high genetic risk prevents more cases and conforms to our new understanding of highly polygenic risk and complex environmental causation.

Selection effects on complex trait architecture (Guy Sella)

Missing heritability largely reflects the limited statistical power of current GWAS to identify loci which together account for the bulk of genetic variance in complex traits. The power of GWAS can be well approximated in terms of two thresholds, where loci that exceed both thresholds are identified: i) a threshold contribution to genetic variance, which is a simple function of minor allele frequency and effect size, and ii) a threshold minor allele frequency (in studies based on genotyping and imputation).

Because many quantitative complex traits are subject to stabilizing selection and because genetic variation affecting one trait often affects many others, the genetic architecture of a focal trait that arises under stabilizing selection can be modeled in a multidimensional trait space. When the degree of pleiotropy, or the effective number of associated traits, is sufficiently high, and when selection is sufficiently strong, the distribution of genetic variances among loci is insensitive to the specific degree of pleiotropy or strength of selection. The distribution depends on a single parameter: the expected contribution of a strongly selected locus to genetic variance. Weakly selected loci contribute much less to genetic variance than do strongly selected ones. As GWAS sample sizes increase and thus the threshold variance for discovery decreases, strongly selected large-effect loci will be identified first. Only when study sizes become very large will weakly selected loci begin to be picked up. It therefore seems likely that at the moment we are seeing only loci that are strongly selected. Indeed, our theoretical predictions for the distribution of variance among strongly selected loci fit GWAS data for height and body mass index. By extrapolating the fitted distributions to lower thresholds, we can predict the explained heritability and number of loci expected in larger studies.

While our results are insensitive to many variations on modeling assumptions, they nonetheless are sensitive to historical changes in population size that substantially affect allele frequencies in the population under consideration. Another implication of recent demographic history is that relying on genotyping rather than resequencing in GWAS has little effect on explained h^2, because strongly selected loci are expected to contribute much less to genetic variance than loci under intermediate selection. Some complex diseases may be primarily subject to polygenic mutation selection balance rather than to stabilizing selection.

Understanding how evolution shapes genetic architecture helps explain missing h^2. Such understanding should enable inferences about the mode of selection and about the parameters that shape variation in specific complex traits. In turn, these inferences should enable prediction of explained h^2 under different study designs, such as genotyping and resequencing, and thus inform future mapping efforts.

Discussion

If allele frequency and effect size are considered separately, one wonders if there are populations where allele frequency has drifted upwards. Inbreeding will increase the frequency of rare genotypes; could that affect estimates of epistasis? This could be because large effects would give greater power to detect effect modifiers. Newly arising mutations may have large epistatic effects and would be selected against. It does seem curious that strong gxe and epistatic interactions are seen in the animal and plant literature but not in humans. Evidence is increasing of environmental interactions with polygenic risk scores, as in the interaction effects seen in UK Biobank at upper levels of BMI.

Family studies (Lynn Jorde)

Families provide opportunities for long-term, longitudinal studies and return of clinically significant results. Multigenerational pedigrees allow detection of shared genomic segments containing rare disease-causing variants and detection of causal *de novo* mutations, such as SVs in autism and schizophrenia and SNVs in intellectual disability and congenital heart disease. They enable detection of parent-of-origin effects, germline/somatic mosaicism, and Mendelian subsets with complex disease phenotypes (*BRCA1, BRCA2, APC*, etc.). The Utah Population Database (UPDB), initiated in the 1970s, consists of 10 million people in large, multigenerational pedigrees that are linked to more than 25 million medically relevant records. The database has now expanded to 100 million people and includes geocode information. Utah pedigrees offer excellent power for detecting and validating *de novo* mutations (DNM) by transmission and identifying unlikely DNMs (likely false positives) if they aren't observed in any of the offspring of a putative DNM carrier. Multigenerational pedigrees allow estimation of the false negative rate of DNMs, which appears to be < 4% for 30x WGS data. Validating DNM in the F1 generation by transmission produces an estimated average of 71 DNMs per individual, and about one in 5-7 have a *de novo* structural variant. The effect of paternal age on DNMs is greater than maternal age, with ~1.3 additional DNM per year of paternal age and 0.3 DNM per year of maternal age. Large F2 generations in Utah pedigrees allow for longitudinal evaluation of DNM rate and assessment of the paternal age effect within individual pedigrees. 339 germline mosaic events (DNMs in multiple F2s) were identified across Utah families. Families were re-contacted in 1990s and 180 phenotype variables were collected. Fourth-generation data collection will be initiated with re-contact of families and follow-up.

An additional study done in this large cohort included finding new shared genomic segments (SGS). SGS analysis was applied to multiple myeloma (MM) for which GWAS accounts only for about 20% of h^2. 11 high-risk MM pedigrees were identified from UPDB. One genome-wide significant SGS region (1.8 Mb) containing nine genes was found, of which one gene, *USP45* (a DNA-repair gene), contained pathogenic variants. More regions that play prominent roles in cancer somatic mutations were identified by using multiple pedigrees. SGS analysis is being applied to other diseases such as young onset atrial fibrillation, preterm birth, and autism in the large Utah pedigrees.

Better phenotyping and use of biomarkers (Dave Valle)

For even the strongest variant, nearly all phenotypes exhibit variation in expressivity, while for nearly all phenotypes, heterogeneity of etiology is the rule. Informed, rigorous, iterative phenotyping yields the best data. The Centers for Mendelian Genomics (CMGs) aim to identify all genes with high penetrance variants to produce a map of phenotype relationships for all genes in the genome as "integrative, whole organism phenotypes." The CMGs describe ~300 new phenotypes per year.

An example of locus heterogeneity from the CMG findings is shown in Robinow syndrome and the Wnt-PCP pathway. Robinow genes include *DVL1* and *DVL3* that have very similar alterations: 6 total pathogenic variants are tightly clustered within 100 nucleotides of each other. Contrary to *DVL1*, the variants in *DVL3* are located in the final exon and include two splice acceptor mutations. But they are all still -1 frameshifting and escape from nonsense-mediated decay.

"Multi-Mendels," the occurrence of dual molecular diagnoses resulting in a blended phenotype, was first reported in 2017. Multi-Mendels are an important cause of heterogeneity and have implications for clinical care, as these dual diagnoses may not be ascertained clinically, and without WES may never be properly diagnosed. Recognition of dual diagnosis is important as it informs recurrence risk counseling and may also inform management and surveillance recommendations.

Epigenome patterns in Kabuki syndrome identified important pathogenic histone alterations. Disease is caused by alterations in patterns of DNA methylation secondary to defects in genes encoding histone modifying enzymes. This suggests "cross talk" between epigenome changes in histones and DNA methylation. Patterns similar in KS secondary to variants in *KMT2D, KDM6A* and *KMT2A* (Wiedermann-Steiner) suggest that a similar phenotype derives from a similar epigenetic pattern. Results point to a set of downstream genes that may be important for pathogenesis.

Two-locus models are another source of heterogeneity, as seen in craniosynostosis. WES was performed in a cohort of 191 cases yielding 13 (7%) rare damaging *de novo* or transmitted variants in *SMAD6* that showed about 60% incomplete penetrance. Previous GWAS had identified one common variant (with allele frequency 0.35) 345 kb downstream of *BMP2*. *SMAD6* encodes a BMP-induced osteoblast differentiation. The *BMP2* downstream variant increases the risk of disease when the base is cytosine, and acted as a protective variant when the base is thymine.

Discussion

One can liken the omnigenic model to the Kerplunk game—marbles are core genes, which are supported by sticks (peripheral genes). Peripheral gene effects are only seen in select people; similar to how some Kerplunk marbles only fall when certain sticks are removed. Core genes are not necessarily hub genes (defined as such by systems biology framework). The omnigenic model can also be additive, and thus it is unnecessary to superimpose epistatic interactions to understand the model.

eQTL studies can be used to find variant effects with bigger impacts than GWAS alone, due to differences in penetrance. A 1mm effect on height averaged across all participants may actually be a 10cm effect in 1% of the participants. eQTL mapping analyses also allow variant effects to be interpreted in clusters. Pedigree studies may someday be able to estimate the penetrance of common and rare variants. When analyzing variant effect, impact, and penetrance, one must also consider selection. Selection can be intense and intensify over generations, which gives rise to polygenicity.

Mind the (diversity) gap: contributions of diverse populations to common disease studies (Lucia Hindorff)

Genetic studies over-represent European ancestry populations. The Population Architecture Using Genomics and Epidemiology (PAGE) program comprises has 50,000 individuals of non-European ancestry. PAGE developed a multi-ethnic genotyping array (MEGA) which is useful for fine-mapping and finding secondary alleles in diverse populations. PAGE recommends that population studies combine rather than stratify the data sets by ethnicity, as the former has more statistical power. PAGE has

developed the GENESIS and SUGEN analytic tools for computing on multiple ancestry populations. Missing heritability is due to a lack of scientific information, and lack of information differs among racial and ethnic minorities. To address the information disparity, studies need to include diverse populations and ensure analyses are state-of-the-art.

Missing heritability: contributions from genomic studies in African ancestry populations (Charles Rotimi)

Environmental and social data (e.g. educational attainment, health insurance, tax policies, water quality, housing opportunities) are frequently missing from genetic studies, but they impact health outcomes and health disparities. Hypertension is assumed to have the highest prevalence in African Americans, but prevalence in rural African populations is actually quite low, suggesting an important role of environmental and social factors. Gxe interactions are infrequent but can help to identify associated SNPs; ethnic-specific SNPs are more common and can only be found with samples of diverse ancestry.

In a study of phenotypic variance stratified by local ancestry in admixed African Americans, most additive genetic variation was explained by genetic markers undifferentiated by ancestry. Results suggested the proportion of health disparities due to genetic risk factors, and adjusting for global ancestry did not control local ancestry effects.

Contributions of diverse populations and expanded catalogues of human variation to our understanding of low frequency and rare variants (Eimear Kenny)

Human genetic history is complex and has changed dramatically in recent evolutionary history. The 1000 Genomes Project showed that common variants are shared globally while rare variants are geospatially restricted. Diverse reference populations facilitate adjudicating more variants clinically. Local ancestry and recent demography are important for rare variant mapping studies, and diverse reference and comparative sequences are required to improve rare/low frequency mapping. Knowing genetic ancestry can also reduce false positives and find founder populations thought to be missing.

Discussion

Genotypes can vary in prevalence in different populations due to the population's propensity to exposure and subsequent selection. Gene-by-environment correlations should therefore be assessed and included in heritability studies.

Common variant studies have more power when multiple ancestry populations are combined. Still, stratifying analyses by ancestry may be more fruitful for rare variants. Some rare variants are missing in entire populations, and thus a combined analysis will dilute those variants and under- or overestimate their prevalence. Using chromosomal segments to adjust for local ancestry would be ideal because chromosomal positions are important in admixed ancestry populations. However, the research community should be wary of pigeonholing genes as population-specific. Studies focusing on non-European ancestry populations have found more variants and elucidated more genetic architectures than predominantly European studies. The diversity in regional African populations alone presents a good opportunity to understand genetic variation, gene flow, and environmental impact on h^2.

Environmental effects on specific populations, and population prevalence data, have not been leveraged to their full potential. The tuberculosis bacterium has shaped the genome toward increased tendency to IBD; more explicit modeling of infectious agents might permit identification of more risk variants such as those in *APOL1*, We are not in a post-GWAS era, and disease and population architecture are far from completely understood. Failing to correct for environmental artifacts will skew a variant's actual effect size and penetrance. It is estimated that 75-80% of effect sizes will be shared across all populations.

Missing epistasis (Andy Clark)

Fisher's infinitesimal model supports the idea of a very large number of unlinked loci, each with very small effect. Results of the model show that the variance of offspring does not depend on trait values of the parents; selection produces negligible change in allele frequency (or variance); the model can accommodate epistasis; and consequences of stabilizing selection, inbreeding, and assortative mating are easily derived. The infinitesimal model of epistasis supposes each pairwise interaction is small and only very few are genome-wide significant. A substantial portion of variation caused by epistatic interaction ends up in the additive variance and thus contributes to h^2. Epistasis matters in evolution as genes exist in networks and epistasis at the molecular level is pervasive. If selection is weak, drift dominates and variance components are unchanged. If selection is strong, allele frequencies change, and the genotype-phenotype map matters more than variance components. Reasons for difficulties in detecting epistasis include that markers are in imperfect LD with causal variants, rapid population growth leads to more rare alleles, multi-dimensionality and small effect size reduce power, and effects are embedded in higher dimension gxg and gxe interactions.

Impact of indirect genetic effects on effect estimates, heritability estimates, and missing heritability (Augie Kong)

Sib-regression for traits in Iceland suggest that twin estimates could overestimate h^2 in the general population (at least in Iceland). For two probands, Relatedness Disequilibrium Regression (RDR) uses the identity-by-descent relatedness between the two pairs of parents as baseline/control for the IBD relatedness between the probands. If these RDR estimates are to be believed, this is evidence that Scandinavian twin estimates tend to be too high when applied to the general population of Iceland. Non-transmitted (NT) alleles only have nurturing effects and transmitted (T) alleles have both direct and nurturing effects. Thus, basic GWAS effect estimates would tend to be overestimates of the direct effects when there is genetic nurturing. Existence of genetic nurture can profoundly affect how various h^2 estimates should be interpreted; GREML estimates, for example, would unavoidably also capture the genetic nurturing effects. Twin estimates and sib-regression are not affected by parental/ancestral genetic nurture, but can be biased due to genetic nurture from siblings.

H^2 estimates based on twins, for whatever reason, appear to be too high for the general population. Sib-regression has its appeal but requires very large sample sizes. The RDR method might work well for probands with parents who are also genotyped. RDR and sib-regression can complement each other. Genetic nurture can lead to positive bias of both effect estimates and h^2 estimates from GREML. If 'explained heritability' only counts GWS markers and GREML h^2 estimates are used, this could inflate

missing h^2 for many health-related traits. The genetic components of educational attainment and BMI are estimated to have a correlation of -0.13 (Bulik-Sullivan *et al*., NG 2015). A part of that could be shared genetic nurturing components.

Discussion

Adoption studies would be a good model design to study h^2 estimates, as genetic nurturing modifications would not be a factor because the parents would be unrelated to the child. In principle, the estimates are not biased but one has to reinterpret what they are estimating. Claiming something to be unbiased is meaningless unless you define what you are estimating. Lyndon Eves studied this in the 1980s before the measured genotype era. There is a lot of literature between adoption and twin studies that might be relevant to understand this, but it has not been solved.

The rapid expansion of the human population has led to an excess of rare alleles that contribute to epistatic interactions. This is also shown in studies of model organisms and it is attributed to population growth. However, comparing these rare variants to each other will not give epistatic variants. In addition, there is no distinction made in epistasis between quantitative and disease (binary) traits.

On the underlying evolutionary forces, we know that genetic nurture affects fitness but have not defined how allelic effects evolve due to these forces. In the literature this is described in maternal effects. The challenge is to define the exceptions due to the lack of power in our studies. There is not enough power to estimate the phenomenon well with individual parents. On the example of educational attainment, even in countries such as Iceland where society favors a very equal environment across all study participants, we still see discrepancies. If the genetic nurture model is removed from educational attainment and applied to diet, for example, we could recognize an impact on disease because of inheriting risk alleles and behaviors from parents. There are data on genetic nurture that show effects on BMI, HDL, smoking, etc. and it is highly likely that genetic nurturing affects all health traits.

Summary and recommendations (Teri Manolio and Peter Visscher)

A number of lessons learned in the areas of quantifying missing h^2; missing h^2 in the clinic; contributions of rare variants, structural variants, and gene expression; environmental effects; diverse populations; and genetic architecture of complex traits were summarized and are listed in the Executive Summary. Future directions for new or enhanced analyses or studies are also detailed there.

Enormous progress has been made in the past decade in contrast with decades before. H^2 appears to have been over-estimated using traditional methods, leading to more apparent missing h^2 than is probably the case. GWAS as an experimental design is no longer questioned and has been highly successful in explaining h^2 and transitioning from classic Mendelian to omnigenic or infinitesimal models. Powerful data resources are fueling discovery, including GWAS summary statistics, the GWAS catalogue, GTEx, the Epigenetic Roadmap, ENCODE, and UK Biobank. Substantial proportions of h^2 are now captured from known variants, and nearly all traits appear to be polygenic. It remains to be determined how polygenicity works biologically, and how natural selection shapes genetic architecture.