

49. Boyle, A. P. *et al.* High-resolution mapping and characterization of open chromatin across the genome. *Cell* **132**, 311–322 (2008).
50. Jin, C. *et al.* H3.3/H2A.Z double variant-containing nucleosomes mark ‘nucleosome-free regions’ of active promoters and other regulatory regions. *Nature Genet.* **41**, 941–945 (2009).
51. Fu, Y., Sinha, M., Peterson, C. L. & Weng, Z. The insulator binding protein CTCF positions 20 nucleosomes around its binding sites across the human genome. *PLoS Genet.* **4**, e1000138 (2008).
52. Li, M., Belozero, V. E. & Cai, H. N. Modulation of chromatin boundary activities by nucleosome-remodeling activities in *Drosophila melanogaster*. *Mol. Cell Biol.* **30**, 1067–1076 (2009).
53. Gerasimova, T. I., Byrd, K. & Corces, V. G. A chromatin insulator determines the nuclear localization of DNA. *Mol. Cell* **6**, 1025–1035 (2000).
54. Capelson, M. & Corces, V. G. The ubiquitin ligase dTopors directs the nuclear organization of a chromatin insulator. *Mol. Cell* **20**, 105–116 (2005).
55. Golovnin, A. *et al.* ‘Insulator bodies’ are aggregates of proteins but not of insulators. *EMBO Rep.* **9**, 440–445 (2008).
56. Byrd, K. & Corces, V. G. Visualization of chromatin domains created by the *gypsy* insulator of *Drosophila*. *J. Cell Biol.* **162**, 565–574 (2003).
57. Maeda, R. K. & Karch, F. Making connections: boundaries and insulators in *Drosophila*. *Curr. Opin. Genet. Dev.* **17**, 394–399 (2007).
58. Kuhn, E. J. & Geyer, P. K. Genomic insulators: connecting properties to mechanism. *Curr. Opin. Cell Biol.* **15**, 259–265 (2003).
59. Nolis, I. K. *et al.* Transcription factors mediate long-range enhancer–promoter interactions. *Proc. Natl Acad. Sci. USA* **106**, 20222–20227 (2009).
60. Bartolomei, M. S. Genomic imprinting: employing and avoiding epigenetic processes. *Genes Dev.* **23**, 2124–2133 (2009).
61. Mishiro, T. *et al.* Architectural roles of multiple chromatin insulators at the human apolipoprotein gene cluster. *EMBO J.* **28**, 1234–1245 (2009).
62. Dorsett, D. Cohesin, gene expression and development: lessons from *Drosophila*. *Chromosome Res.* **17**, 185–200 (2009).
63. Petrykowska, H. M., Vockley, C. M. & Elnitski, L. Detection and characterization of silencers and enhancer-blockers in the greater *CFTR* locus. *Genome Res.* **18**, 1238–1246 (2008).

Acknowledgements

We would like to thank B. Cairns, O. Rando, M. Bulger, D. Clark, G. Hartzog, M. Okí, K. Noma, S. Henikoff, P. Geyer and G. Felsenfeld for comments on an early draft of this

Review. We would also like to thank members of the Ro laboratory for comments and criticisms. We apologize for the selective citations, which are a consequence of space limitations. This work was supported by grants from the US National Institutes of Health to R.T.K. (GM078068) and J.R.R. (T32-GM008646), and by grant 2008-16 from the GREAT Training Program of the University of California System-wide Biotechnology Research and Education Program to J.R.R.

Competing interests statement

The authors declare no competing financial interests.

DATABASES

Entrez Gene: <http://www.ncbi.nlm.nih.gov/gene>
 FOLR1 | IGF2 | H19
 FlyBase: <http://flybase.org>
 achaete-scute | scs | scs' | yellow
 UniProtKB: <http://www.uniprot.org>
 BEAF | CP190 | CTCF | GAF | MOD(MDG4) | SU(HW) | USF1 | VEZF1

FURTHER INFORMATION

Rohinton T. Kamakaka's homepage:
<http://research.pbsci.ucsc.edu/mcdb/chromatin>
 ALL LINKS ARE ACTIVE IN THE ONLINE PDF

VIEWPOINT

Missing heritability and strategies for finding the underlying causes of complex disease

Evan E. Eichler, Jonathan Flint, Greg Gibson, Augustine Kong, Suzanne M. Leal, Jason H. Moore and Joseph H. Nadeau

Abstract | Although recent genome-wide studies have provided valuable insights into the genetic basis of human disease, they have explained relatively little of the heritability of most complex traits, and the variants identified through these studies have small effect sizes. This has led to the important and hotly debated issue of where the ‘missing heritability’ of complex diseases might be found. Here, seven leading geneticists offer their opinion about where this heritability is likely to lie, what this could tell us about the underlying genetic architecture of common diseases and how this could inform research strategies for uncovering genetic risk factors.

Q How should we solve the problem of ‘missing heritability’ in complex diseases?

Evan E. Eichler. Understanding the heritability of genetic diseases requires a more comprehensive assessment of human genetic variation. Human genomes are rich in structural diversity, but the discovery and genotyping of this type of variation has lagged far behind those of the SNP^{1,2}. Although there has been a tremendous push to close this gap over the past 4 years^{3–5}, two aspects remain understudied. The first is the exploration of the landscape and impact of large variants (deletions, duplications and inversions) that

are individually rare but collectively common in the human population^{6,7}. An estimated 8% of the general population carry a large (>500 kb) deletion or duplication that occurs at an allele frequency of <0.05%⁷. The available data suggest that these variants are under strong selection, affect transcription⁸ and contribute to a variety of different diseases⁹. These genomic imbalances represent a special class of rare variants that can potentially affect many genes and pathways in a single individual. Not only are large numbers of cases and controls required to assess the clinical significance of particular events, but the modelling of other forms of genetic variation in this

sensitized background of localized haploidy or triploidy remains largely unexplored. The second aspect involves the several hundred genes that map to regions of copy-number polymorphic (CNP) duplications. Available data suggest that these genes are highly variable among individuals, are enriched in genes associated with drug detoxification, immunity and environmental interaction¹⁰, and have been subject to bursts of rapid, and sometimes adaptive, evolution in humans and our ape relatives. However, because of their repetitive and multicopy nature, these genes are considered inaccessible by most existing genotyping and sequencing technologies. There is a pressing need to characterize not only copy number but also the sequence content and structural arrangement in these diverse regions of our genome. Such regions are more likely to be subject to recurrent mutations and be inadequately assayed by a correlated neighbouring SNP. It is therefore premature to conclude that CNPs have limited impact in terms of common disease until these more complex regions are tested¹¹. Excluding the most variable and diverse regions of human genetic variation because they are difficult to study is an unacceptable loss in the pursuit of genotype–phenotype correlations.

Jonathan Flint. I find it hard to imagine that there will be a single answer to the question of where to find missing heritability, but I have a suggestion as to what might help find it. Even in crosses between inbred mouse strains, in which the genetics is simplified to a comparison between two genomes (and related genomes at that), there is variation in genetic architecture among phenotypes. For example, susceptibility to infectious disease has often

The contributors*

Eván E. Eichler is a professor and Howard Hughes Medical Institute (HHMI) Investigator in the Department of Genome Sciences, University of Washington School of Medicine, Seattle, Washington, USA. He received his Ph.D. in 1995 from the Department of Molecular and Human Genetics at Baylor College of Medicine, Houston, Texas, USA. He joined the faculty of Case Western Reserve University, Cleveland, Ohio, USA, in 1997 and later the Department of Genome Sciences in 2004. He was appointed as an HHMI Investigator in 2005 and an American Association for the Advancement of Science Fellow in 2006, and was awarded the American Society of Human Genetics Curt Stern Award in 2008. He is an editor of the journal *Genome Research* and a participant in numerous genome sequencing projects. His research group develops experimental and computational methods for studying complex regions of genome structural variation, including deletions, duplications and inversions. The long-term goal of his research is to understand the evolution and mechanisms of primate gene duplication and their relationship to copy-number variation and human disease.

Jonathan Flint is a psychiatrist working at the University of Oxford, UK. After working on small chromosomal deletions (or copy-number variants) that cause mental retardation, he turned to investigating the genetic basis of anxiety and depression in animal models and in humans. He developed strategies for fine-mapping genetic loci underpinning complex traits using heterogeneous stocks for genetic mapping. These strategies apply to all phenotypes, not just behaviour.

Greg Gibson is a professor in the School of Biology at the Georgia Institute of Technology in Atlanta, Georgia, USA. He trained as a developmental molecular geneticist at the University of Basel, Switzerland, and at Stanford University, California, USA, and then worked on genomic approaches to quantitative genetic analysis of complex traits in *Drosophila melanogaster* for 12 years, mainly at North Carolina State University, Raleigh, USA. Recently, he has shifted his research programme to human genetics, with a particular interest in the causes of environment and culture-dependent increases in the incidence of common complex diseases. His 'human transition project' uses genome-wide genetics of gene expression and metabolomics studies in mixed populations in developing countries, such as Morocco and Fiji, as well as in diverse urban settings.

Augustine Kong is Vice President of Statistics at deCODE genetics in Iceland. He received his Ph.D. from Harvard University, Cambridge, Massachusetts, USA, and was a professor at the University of Chicago, Illinois, USA, from 1987 to 2000. He started working in Iceland in 1996, and was the architect of the deCODE genetic map constructed in 2002. He has extensive experience with large-scale and population-based genetic studies covering a wide range of human traits. Recently, by combining principles from family and genome-wide association studies, he and his team derived methods that can, reliably and systematically, phase markers over a long range and determine the parental origins of alleles.

Suzanne M. Leal is a professor in the Department of Molecular and Human Genetics at Baylor College of Medicine, Houston, Texas, USA. She received an M.S. in biostatistics and a Ph.D. in epidemiology from Columbia University, New York, USA. She is interested in understanding the genetic etiology of complex and Mendelian traits in relation to common and rare variants, structural variation and gene–gene and gene–environment interactions. She is involved in the study of several traits, including non-syndromic hearing impairment, aneurysm and dissection, opiate addiction, pain perception and platelet reactivity. She is also involved in developing methods for analysing complex traits, and her latest work focuses on the analysis of next-generation sequence data and detecting associations with rare variants. This work led to the development of the combined multivariate and collapsing (CMC) method, which was one of the first approaches designed specifically to detect associations with rare variants.

Jason H. Moore is the Frank Lane Research Scholar in Computational Genetics, Professor of Genetics, Professor of Community and Family Medicine and Director of Bioinformatics at Dartmouth Medical School, Lebanon, New Hampshire, USA. He holds adjunct appointments in Computer Science at the University of New Hampshire, Durham, USA, and the University of Vermont, Burlington, USA. He also holds an adjunct appointment in Psychiatry and Human Behavior at Brown University, Providence, Rhode Island, USA. He serves as a founding Editor-in-Chief of the journal *BioData Mining* and is an editor of the Cambridge University Press book series *Systems Genetics*. He has published extensively on the detection, characterization and interpretation of epistasis and other phenomena that account for the complex mapping relationship between genotype and phenotype in the context of human health.

Joseph H. Nadeau is Chair of the Department of Genetics, James H. Jewell Professor of Genetics and Director of the Division of Bioinformatics in the Center for Proteomics and Bioinformatics at Case Western Reserve University School of Medicine, Cleveland, Ohio, USA. He has joint appointments in the Case Comprehensive Cancer Center and the Department of Electrical Engineering and Computer Science in the Case School of Engineering. He is an Elected Fellow of the American Association for the Advancement of Science, founding director of the Mouse Genome Database and Informatics Program, and co-founding editor of *Mammalian Genome* and *WIREs Systems Biology and Medicine*, which recently won the R. R. Hawkins Award for outstanding scholarly publication from the American Publishers Association. His work focuses on the genetics and systems properties of complex traits and common diseases in mouse models of human disease.

*Listed in alphabetical order.

turned out to be due to variants of large effect^{12,13}; however, apparently equally complex phenotypes (such as cell counts of red and white blood cells, variation in high and low density lipoproteins and obesity) have a much more complex genetic architecture due to the joint action of very many loci of small effect¹⁴. Differences also exist in the extent to which epistasis shapes a phenotype: pervasive epistatic effects have been documented in autoimmune conditions¹⁵, morphology¹⁶ and susceptibility to cancer¹⁷, but the genetic architecture underlying fear-related phenotypes consists almost entirely of multiple small additive effects^{14,18}.

Differences in genetic architecture reflect the complex, often opposing effects of selection, population history, migration and mutation rates. Is it possible to be more specific, to make predictions about genetic architecture? Interactions between selection and the size and structure of populations contribute to allele frequencies in predictable ways¹⁹, and theoretical models and data have already been used to argue that additive genetic effects are likely to be common in complex phenotypes²⁰. But this conclusion applies broadly, averaging across all phenotypes in different populations. A more fine-grained analysis, examining individual

phenotypes and taking into account the characteristics of individual populations, has yet to be undertaken. We would not, for example, expect the genetic architecture of schizophrenia and autism, both conditions that considerably lower reproductive fitness, to be the same as that of intelligence, height or weight. Modelling per locus effect sizes has been used to constrain the genetic architecture of schizophrenia²¹. Presumably, modelling fitness would constrain possibilities yet further. So, understanding why genetic architecture differs for different traits could help when choosing the correct tools to find the underlying genes and deciding whether

to look for common or rare variants, and studying the genetic architecture might even tell us what type of variant to expect (for example, a SNP or a copy-number variant). Population and theoretical genetics approaches may hold the key to finding the missing heritability.

Greg Gibson. In a nutshell, I think the missing heritability problem is overblown, and the focus on hits that are significant genome-wide is distracting attention from more general concerns over the ability of genome-wide association (GWA) studies to fully describe the architecture of phenotypic variation. A lot of the confusion may arise because heritability seems often to be equated with genetic contributions, but it actually refers to the ratio of the genetic to the total phenotypic variance in a population²². The population is assumed to share a common environment, but if there is hidden environmental structure that interacts with genes, then this effect shunts genetic variance to the denominator and reduces the (narrow sense) heritability estimate. So the heritability gives a lower bound on estimates of how much total variance the genetic component should explain. Exceptions will occur if heritability is estimated in a biased way with respect to environmental exposures or from pedigrees that have a high risk.

More fundamentally then, there is a missing genetic variance problem, which really relates to misplaced preconceptions. It would have been nice if GWA studies typically uncovered a dozen associations each explaining 5–10% of the variance, but the fact that they do not suggests only that the allelic effects are smaller or the causal alleles are too rare. With so many tests, there is a high false-negative rate, as true associations are hidden in the fog of random associations. I am convinced that myriad common variants of small effect do explain the vast majority of genetic effects predicted by heritability estimates, even though we cannot detect them individually²³. Rare variants of large effect, sometimes in synthetic association with common variants, will also contribute²⁴.

To my mind, the really interesting questions concern differences in the architecture of diseases, and the rising prevalence of chronic disease. Gene–environment (G×E) interactions may be important in both of these situations, but they will not appear in the heritability estimates. For example, life-style changes may alter the distribution of genetic effects in some people, and genetic buffering may be disrupted in some pedigrees. There are many reasons why such G×E interactions will not be revealed by GWA

studies. Foremost is lack of statistical power, especially if a small fraction of individuals experience the adverse exposure. Population stratification can also induce false negatives in which the allelic effect is in the opposite direction to the population effect — and I suspect this is very common. And of course we are largely unable to identify what the relevant environments are. For these reasons, stratifying GWA studies by environments is unlikely to reveal any but the largest interaction effects, and testing for G×E effects will not obviously explain much more of the variance. G×E interactions may be strongest for rare alleles, but this will be very hard to detect on a case-by-case basis. But we certainly need to learn a great deal more about how the environment does modify allelic effects, because after all it is not genotype–phenotype so much as genotype effect–phenotype associations that really matter. A good place to start is environmental influences on the transcriptome and metabolome.

Augustine Kong. Recently, some susceptibility variants for cancer and type 2 diabetes (T2D) were shown to confer risk only when inherited from a specific parent, and a variant was discovered that can either confer or reduce risk of T2D depending on the parent of origin²⁵. Such variants contribute to missing heritability in two ways: first, they are more difficult to discover and second, even if discovered, their contribution to heritability would be underestimated when evaluated under models that do not take parental origin into account. In another example, variants that increase the recombination rate for fathers reduce the recombination rate for mothers²⁶. As recombination rates of parents affect transmissions and recombinations are sometimes associated with mutations, there could be sex-specific association between variants in parents and risks in offspring. More generally, for diseases in which prenatal conditions have a role, interactions between the genetic variants present in the parent and offspring are possible. In T2D, we estimate that about 13 to 14% of the heritability accounted for by the known variants can be attributed to parent-of-origin effects. Parent-of-origin effects have also been reported for type 1 diabetes²⁷. Considering that the power to detect such variants is low and these effects could be more prevalent with rare variants, such effects should not be overlooked.

More important to consider, however, is the fact that complex inheritance can take on numerous forms. Epigenetic effects beyond imprinting that are sequence-independent

and that might be environmentally induced but can be transmitted for one or more generations²⁸ could contribute to missing heritability. Phase-dependent interactions between variants that are not in linkage disequilibrium and that are difficult to detect without long-range haplotypes²⁹ are another possibility. Predicting the amount of missing heritability explained by each of these would be speculation, but it is reasonable to assume that complex inheritance as a whole could account for a substantial fraction of heritability. Everyone is looking forward to full-genome sequencing of large samples. Simply being able to identify and type variants, rare or common, old or new, that are not tagged by the current common SNPs will no doubt lead to many important discoveries. However, various other types of information can add value to sequence data, such as knowledge about the epigenome³⁰ and information on families and parental origins (including information from long-range phasing, which among other usages²⁹ would assist in determining the age of the nearest common ancestor when two individuals share a region by descent). The possibility that some heritability comes from entirely unforeseen sources is actually something to look forward to in the future.

Suzanne M. Leal. Although GWA studies have been successful in identifying common variants involved in complex trait aetiology, for the majority of complex traits, <10% of genetic variance is explained by common variants³¹. Genetic variance may also be explained by gene interactions and structural variation, and there is strong evidence that rare variants have an important role. These variants, although individually rare, are collectively frequent, and even though their effect sizes are greater than those observed for common variants, they are not large enough to produce familial aggregation³². For a variety of complex traits, exome data are currently being generated, and whole-genome sequencing will follow for the next wave of GWA studies. Detecting associations with rare variants is the first step towards a better understanding of the extent of their role in complex trait aetiology. Common variant analyses can be used for direct mapping of rare variants; however, these methods are underpowered owing to low allele frequencies and allelic heterogeneity^{33,34}. Association methods developed specifically for rare variants jointly analyse variants in a locus or gene instead of individually testing each variant. These methods include the comparison

of rare variants found exclusively in cases to those found only in controls (RVE)³⁵, the combined multivariate and collapsing (CMC)⁴ method and the weighted sum statistic (WSS)³⁶, with the CMC and WSS methods having a power advantage over the RVE method. Some methods, such as the CMC method, can easily be adapted for the analysis of quantitative traits³⁷, the control of confounding factors, and the inclusion of variant-predicted functionality and accuracy of genotype calls. Because sequencing uncovers both causal and non-functional variants, methods for analysing rare variants must be robust to misclassification. Even if detected gene associations are replicated, it is not possible within this analysis framework to tease apart causal from neutral variation. Due to the current cost of exome sequencing, study designs that maximize power for a given number of sequenced individuals are beneficial; these designs include extreme quantitative trait sampling and the use of common controls. When analysing rare variants, it is important to adequately control for population substructure and admixture; rare variants tend to have occurred more recently and therefore have greater population diversity than common variants. Analysis methods that evaluate effect sizes are necessary for the estimation of the amount of explained genetic variance. Computationally efficient methods are also important, in particular for the analysis of exome and whole-genome sequence data. Sequencing and the analysis of phenotyped samples will elucidate over the next few years whether or not the majority of the missing heritability for complex traits is due to rare variants.

Jason H. Moore. The case of the missing heritability for common human diseases should not be a mystery to anyone given the inherent complexity of the relationship between genotype and phenotype. Consider, for example, the discovery of non-coding microRNAs that provide a new mechanism of translational regulation. Genetic variation influencing the expression of non-coding RNAs has the potential to add a previously unexpected layer of complexity to the genetic architecture of biological and clinical traits. For example, Nicoloso *et al.*³⁸ recently showed that SNPs associated with breast cancer susceptibility can alter microRNA gene regulation. This is one of many layers of complexity that need to be considered before we will truly be able to identify the missing heritability that has not been accounted for by agnostic or unbiased GWA studies that focus on one genetic variation at a time.

Given this expected complexity, it is likely that synergistic interactions among variants in the sequences that regulate both microRNA and mRNA expression have a big impact on protein expression. Similarly, it is reasonable to presume that coding sequence variations, for example, could synergistically act through protein–protein interactions and protein–DNA interactions in transcriptional networks and biochemical systems. Such biomolecular interactions that depend on multiple genetic variations can substantially complicate the relationship between genotype and phenotype, making it impossible to explain phenotypic variation simply by adding together independent genetic effects. This hypothesis is completely consistent with the current results of GWA studies. Indeed, this kind of complexity is routinely observed in simple organisms, such as yeast³⁹. If this is true for yeast, why would we expect humans to be markedly simpler?

The idea that multiple genetic variations can and do interact through different layers of genomic complexity is not new. In fact, Bateson⁴⁰ coined the term epistasis to describe one gene standing upon or modifying the effects of another gene. This early definition of epistasis has given way to more modern definitions that recognize the complexity of gene networks and biochemical systems^{41–43}. This logically leads to the idea that a significant proportion of the missing heritability is not due to single common variants, nor single rare variants, but rather to rare combinations of common variants. As such, solving part of the missing heritability problem will require the application of statistical and computational methods that detect patterns of epistasis across the genome implemented in a systems biology framework that accounts for the highly interconnected nature of biomolecular networks^{44,45}. High-throughput technology alone will not solve this problem. The time is now to philosophically and analytically retool for a complex genetic architecture or we will continue to underdeliver on the promises of human genetics. Indeed, life, and thus genetics, is complicated⁴⁶ and some will soon ask, as seismologists have⁴⁷, whether we are trying to predict the unpredictable.

Joseph H. Nadeau. Testing associations between genotype and phenotype is central to many genetic studies of inherited traits and disease. Historically, Mendel's laws and Morgan's chromosome theory of inheritance have dominated research such as GWA studies, and alternative modes of inheritance have been largely and appropriately

dismissed for lack of evidence. Now we discover that most genetic variants that account for the heritable component of phenotypic variation elude discovery⁴⁸. Perhaps heritability is overestimated. Alternatively, perhaps the missing variants reside in largely unexplored regions of the genome, or in largely untested classes of genetic variation. Another possibility is that genetic variants are missed because they are rare and their effects are small. Or perhaps genetic complexity is greater than imagined, with a very large number of closely linked genes that show context-dependent and non-additive effects⁴⁹. But will these possibilities reveal the 'whole truth' about 'missing heritability'?

Recent studies in mice provide striking evidence for transgenerational genetic effects in which phenotypic variation in the present generation results from genetic variants in previous generations²⁸. Remarkably, these studies show that transgenerational effects persist across several if not many generations, and that these effects are common and usually as strong as conventional inheritance. The original discoveries involve pigmentation⁵⁰, germ cell^{51,52} and heart development⁵³, embryogenesis⁵⁴ and growth⁵⁴, and ongoing work provides examples involving metabolism, behaviour and many other traits⁶¹ (J.H.N., unpublished observations). Some transgenerational effects are reminiscent of paramutations^{50,53,54}, whereas others involve interacting genes in different generations^{51,52}. The obvious molecular mechanisms for epigenetic inheritance are DNA methylation and histone modifications⁵⁵. However, evidence from plants, flies and more recently from mice suggests another possibility²⁸ — a combination of small RNAs^{50,53,54}, RNA-binding proteins that are involved in both RNA editing⁵⁶ and microRNA access to their target mRNAs⁵⁷, and DNA methylation mediated by RNA-editing enzymes^{58,59} controls translation in RNA granules that are abundant in the gametes of both males and females^{50,60}.

How are these discoveries relevant to missing heritability? Because transgenerational effects loosen conventional genotype–phenotype associations, even complete genome surveys will fail to reveal the full repertoire of genetic variants. Under these conditions, traits are heritable, with family members being more similar to each other than unrelated individuals. Remarkably in these cases, the genotype of individuals in previous generations is a better predictor of phenotype than the individual's own genotype. Two key questions emerge: first, do transgenerational effects occur in humans?

This test could simply involve examining associations between genotypes and phenotypes across generations. Second, what is the molecular basis for epigenetic inheritance? Although other possibilities exist, perhaps the most provocative evidence suggests that small RNAs and related protein functions are responsible for the epigenetic persistence of genetic memory.

Evan E. Eichler is at the Department of Genome Sciences, University of Washington, 3413C Foege Building, 1705 NE Pacific St., Seattle, Washington 98195-5065, USA. e-mail: eee@gs.washington.edu

Jonathan Flint is at the Wellcome Trust Centre for Human Genetics, Roosevelt Drive, Oxford OX3 7BN, UK. e-mail: jf@well.ox.ac.uk

Greg Gibson is at the School of Biology, Georgia Institute of Technology, 310 Ferst Drive, Atlanta, Georgia 30332, USA. e-mail: greg.gibson@biology.gatech.edu

Augustine Kong is at deCODE genetics, Sturlugata 8, IS-101 Reykjavik, Iceland. e-mail: kong@decode.is

Suzanne M. Leal is at the Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas 77025, USA. e-mail: sleal@bcm.edu

Jason H. Moore is at the Computational Genetics Laboratory, Dartmouth Medical School, One Medical Center Drive, Lebanon, New Hampshire 03756, USA. e-mail: jason.h.moore@dartmouth.edu

Joseph H. Nadeau is at the Department of Genetics, Case Western Reserve University School of Medicine, Biomedical Research Building 731, 2109 Adelbert Road, Cleveland, Ohio 44106-4955, USA. e-mail: joseph.nadeau@case.edu

doi:10.1038/nrg2809

- Cooper, G. M., Zerr, T., Kidd, J. M., Eichler, E. E. & Nickerson, D. A. Systematic assessment of copy number variant detection via genome-wide SNP genotyping. *Nature Genet.* **40**, 1199–1203 (2008).
- McCarroll, S. A. et al. Integrated detection and population-genetic analysis of SNPs and copy number variation. *Nature Genet.* **40**, 1166–1174 (2008).
- Redon, R. et al. Global variation in copy number in the human genome. *Nature* **444**, 444–454 (2006).
- Kidd, J. M. et al. Mapping and sequencing of structural variation from eight human genomes. *Nature* **453**, 56–64 (2008).
- Conrad, D. F. et al. Origins and functional impact of copy number variation in the human genome. *Nature* **464**, 704–712 (2010).
- Sebat, J. et al. Large-scale copy number polymorphism in the human genome. *Science* **305**, 525–528 (2004).
- Itsara, A. et al. Population analysis of large copy number variants and hotspots of human genetic disease. *Am. J. Hum. Genet.* **84**, 148–161 (2009).
- Henrichsen, C. N. et al. Segmental copy number variation shapes tissue transcriptomes. *Nature Genet.* **41**, 424–429 (2009).
- Mefford, H. C. & Eichler, E. E. Duplication hotspots, rare genomic disorders, and common disease. *Curr. Opin. Genet. Dev.* **19**, 196–204 (2009).
- Alkan, C. et al. Personalized copy number and segmental duplication maps using next-generation sequencing. *Nature Genet.* **41**, 1061–1067 (2009).
- The Wellcome Trust Case Control Consortium. Genome-wide association study of CNVs in 16,000 cases of eight common diseases and 3,000 shared controls. *Nature* **464**, 713–720 (2010).
- Min-Oo, G. et al. Pyruvate kinase deficiency in mice protects against malaria. *Nature Genet.* **35**, 357–362 (2003).
- Diez, E. et al. *Birc1e* is the gene within the *Lgn1* locus associated with resistance to *Legionella pneumophila*. *Nature Genet.* **33**, 55–60 (2003).
- Valdar, W. et al. Genome-wide genetic association of complex traits in heterogeneous stock mice. *Nature Genet.* **38**, 879–887 (2006).
- Wanstrat, A. & Wakeland, E. The genetics of complex autoimmune diseases: non-MHC susceptibility genes. *Nature Immunol.* **2**, 802–809 (2001).
- Leamy, L. J., Routman, E. J. & Cheverud, J. M. An epistatic genetic basis for fluctuating asymmetry of mandible size in mice. *Evolution* **56**, 642–653 (2002).
- Fijneman, R. J., de Vries, S. S., Jansen, R. C. & Demant, P. Complex interactions of new quantitative trait loci, *Sluc1*, *Sluc2*, *Sluc3*, and *Sluc4*, that influence the susceptibility to lung cancer in the mouse. *Nature Genet.* **14**, 465–467 (1996).
- Flint, J., De Fries, J. C. & Henderson, N. D. Little epistasis for anxiety-related measures in the DeFries strains of laboratory mice. *Mamm. Genome* **15**, 77–82 (2004).
- Wright, S. Evolution in Mendelian populations. *Genetics* **16**, 97–159 (1931).
- Hill, W. G., Goddard, M. E. & Visscher, P. M. Data and theory point to mainly additive genetic variance for complex traits. *PLoS Genet.* **4**, e1000008 (2008).
- Wray, N. R. & Visscher, P. M. Narrowing the boundaries of the genetic architecture of schizophrenia. *Schizophr. Bull.* **36**, 14–23 (2009).
- Visscher, P. M., Hill, W. G. & Wray, N. R. Heritability in the genomics era — concepts and misconceptions. *Nature Rev. Genet.* **9**, 255–266 (2008).
- International Schizophrenia Consortium et al. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* **460**, 748–752 (2009).
- Dickson, S. P., Wang, K., Krantz, I., Hakonarson, H. & Goldstein, D. B. Rare variants create synthetic genome-wide associations. *PLoS Biol.* **8**, e1000294 (2010).
- Kong, A. et al. Parental origin of sequence variants associated with complex diseases. *Nature* **462**, 868–874 (2009).
- Kong, A. et al. Sequence variants in the *RNF212* gene associate with genome-wide recombination rate. *Science* **319**, 1398–1401 (2008).
- Wallace, C. et al. The imprinted *DLK1-MEG3* gene region on chromosome 14q32.2 alters susceptibility to type 1 diabetes. *Nature Genet.* **42**, 68–71 (2010).
- Nadeau, J. H. Transgenerational genetic effects on phenotypic variation and disease risk. *Hum. Mol. Genet.* **18**, R202–R210 (2009).
- Kong, A. et al. Detection of sharing by descent, long-range phasing and haplotype imputation. *Nature Genet.* **40**, 1068–1075 (2008).
- Feinberg, A. P. Genome-scale approaches to the epigenetics of common human disease. *Virchows Arch.* **456**, 13–21 (2010).
- Frazer, K. A., Murray, S. S., Schork, N. J. & Topol, E. J. Human genetic variation and its contribution to complex traits. *Nature Rev. Genet.* **10**, 241–251 (2009).
- Bodmer, W. & Bonilla, C. Common and rare variants in multifactorial susceptibility to common diseases. *Nature Genet.* **40**, 695–701 (2008).
- Gorlov, I. P., Gorlova, O. Y., Sunyaev, S. R., Spitz, M. R. & Amos, C. I. Shifting paradigm of association studies: value of rare single-nucleotide polymorphisms. *Am. J. Hum. Genet.* **82**, 100–112 (2008).
- Li, B. & Leal, S. M. Methods for detecting associations with rare variants for common diseases: application to analysis of sequence data. *Am. J. Hum. Genet.* **83**, 311–321 (2008).
- Cohen, J. C. et al. Multiple rare alleles contribute to low plasma levels of HDL cholesterol. *Science* **305**, 869–872 (2004).
- Madsen, B. E. & Browning, S. R. A groupwise association test for rare mutations using a weighted sum statistic. *PLoS Genet.* **5**, e1000384 (2009).
- Morris, A. P. & Zeggini, E. An evaluation of statistical approaches to rare variant analysis in genetic association studies. *Genet. Epidemiol.* **34**, 188–193 (2010).
- Nicoloso, M. S. et al. Single-nucleotide polymorphisms inside microRNA target sites influence tumor susceptibility. *Cancer Res.* **70**, 2789–2798 (2010).
- Gao, H., Granka, J. M. & Feldman, M. W. On the classification of epistatic interactions. *Genetics* **184**, 827–837 (2010).
- Bateson, W. *Mendel's Principles of Heredity* (Cambridge Univ. Press, 1909).
- Moore, J. H. & Williams, S. M. Epistasis and its implications for personal genetics. *Am. J. Hum. Genet.* **85**, 309–320 (2009).
- Phillips, P. C. Epistasis — the essential role of gene interactions in the structure and evolution of genetic systems. *Nature Rev. Genet.* **9**, 855–867 (2008).
- Tyler, A. L., Asselbergs, F. W., Williams, S. M. & Moore, J. H. Shadows of complexity: what biological networks reveal about epistasis and pleiotropy. *Bioessays* **31**, 220–227 (2009).
- Cordell, H. J. Detecting gene–gene interactions that underlie human diseases. *Nature Rev. Genet.* **10**, 392–404 (2009).
- Moore, J. H., Asselbergs, F. W. & Williams, S. M. Bioinformatics challenges for genome-wide association studies. *Bioinformatics* **26**, 445–455 (2010).
- Hayden, E. C. Life is complicated. *Nature* **464**, 664–667 (2010).
- Hough, S. *Predicting the Unpredictable: The Tumultuous Science of Earthquake Prediction* (Princeton Univ. Press, 2010).
- Manolio, T. A. et al. Finding the missing heritability of complex diseases. *Nature* **461**, 747–753 (2009).
- Shao, H. et al. Genetic architecture of complex traits: large phenotypic effects and pervasive epistasis. *Proc. Natl Acad. Sci. USA* **105**, 19910–19914 (2008).
- Rassoulzadegan, M. et al. RNA-mediated non-mendelian inheritance of an epigenetic change in the mouse. *Nature* **441**, 469–474 (2006).
- Lam, M.-Y. J., Heaney, J. D., Youngren, K. K., Kawasoe, J. H. & Nadeau, J. H. Trans-generational epistasis between *Dnd1*^{ter} and other modifier genes controls susceptibility to testicular germ cell tumors. *Hum. Mol. Genet.* **16**, 2233–2240 (2007).
- Heaney, J. D., Lam, M.-Y. J., Michelson, M. V. & Nadeau, J. H. Loss of the transmembrane but not the soluble kit ligand isoform increases testicular germ cell tumor susceptibility in mice. *Cancer Res.* **68**, 5193–5197 (2008).
- Wagner, K. D. et al. RNA induction and inheritance of epigenetic cardiac hypertrophy in the mouse. *Dev. Cell* **14**, 962–969 (2008).
- Grandjean, V. et al. The *miR-124–Sox9* paramutation: RNA-mediated epigenetic control of embryonic and adult growth. *Development* **136**, 3647–3655 (2009).
- Hammoud, S. S. et al. Distinctive chromatin in human sperm packages genes for embryo development. *Nature* **460**, 473–478 (2009).
- Youngren, K. K. et al. The *Ter* mutation in the dead-end gene causes germ cell loss and testicular germ cell cancer. *Nature* **435**, 360–365 (2005).
- Kedde, M. et al. RNA-binding protein Dnd1 inhibits microRNAs access to target mRNA. *Cell* **131**, 1273–1286 (2007).
- Morgan, H. D., Dean, W., Coker, H. A., Reik, W. & Petersen-Mahrt, S. K. Activation-induced cytidine deaminase deaminates 5-methylcytosine in DNA and is expressed in pluripotent tissues. *J. Biol. Chem.* **279**, 52353–52360 (2004).
- Popp, C. et al. Genome-wide erasure of DNA methylation in mouse primordial germ cell is affected by AID deficiency. *Nature* **463**, 1101–1106 (2010).
- Kedersha, N. & Anderson, P. Chapter 4 Regulation of translation by stress granules and processing bodies. *Prog. Mol. Biol. Transl. Sci.* **90C**, 155–185 (2009).
- Nelson, V. R. & Nadeau, J. H. Transgenerational genetic effects of the paternal chromosome on daughter phenotypes. *Epigenomics* (in the press).

Acknowledgements
E.E.E. thanks G. Cooper for helpful suggestions. J.F. is supported by the Wellcome Trust. J.H.M. is supported by US National Institutes of Health R01s LM009012, LM010098 and AI59694.

Competing interests statement
E.E.E. and A.K. declare **competing financial interests**; see Web version for details.

FURTHER INFORMATION
Evan E. Eichler's homepage: <http://eichlerlab.gs.washington.edu>
Jonathan Flint's homepage: <http://www.well.ox.ac.uk/flint>
Greg Gibson's homepage: <http://www.gibsongroup.biology.gatech.edu>
Suzanne M. Leal's homepage: <http://www.bcm.edu/genetics/leal>
Jason H. Moore's homepage: <http://www.epistasis.org>
ALL LINKS ARE ACTIVE IN THE ONLINE PDF