Published in final edited form as:

Curr Opin Genet Dev. 2013 December ; 23(6): 678–683. doi:10.1016/j.gde.2013.10.008.

Genetic architecture of quantitative traits and complex diseases

Wenqing Fu¹, Timothy D. O'Connor¹, Joshua M. Akey¹

¹Department of Genome Sciences, University of Washington, Seattle, Washington 98195-5065, USA

Abstract

More than 150 years after Mendel discovered the laws of heredity, the genetic architecture of phenotypic variation remains elusive. Here, we discuss recent progress in deciphering how genotypes map onto phenotypes, sources of genetic complexity, and how model organisms are illuminating general principles about the relationship between genetic and phenotypic variation. Moreover, we highlight insights gleaned from large-scale sequencing studies in humans, and how this knowledge informs outstanding questions about the genetic architecture of quantitative traits and complex diseases. Finally, we articulate how the confluence of technologies enabling whole-genome sequencing, comprehensive phenotyping, and high-throughput functional assays of polymorphisms will facilitate a more principled and mechanistic understanding of the genetic architecture of phenotypic variation

Introduction

One of the most ubiquitous characteristics of life is the incredible phenotypic diversity that exists within and between species. Delineating the heritable basis of phenotypic variation remains a fundamental goal of applied, basic, and biomedical research [1]. Although our understanding of how genetic diversity maps onto phenotypic diversity is becoming increasingly sophisticated, formidable gaps in knowledge remain. Of particular interest, and complexity, is elucidating the genetic architecture of quantitative traits and complex diseases. To be clear, the genetic architecture of a trait refers to a comprehensive description of how genes and the environment conspire to produce phenotypes (Figure 1), including the number of quantitative trait loci (QTL) that contribute to variability of the trait between individuals, their effect sizes, whether alleles for causal polymorphisms are additive, dominant, or recessive and their frequency in the population, patterns of gene-gene and gene-environment effects, and levels of pleiotropy [2]. There is an extensive literature on the genetic architecture of quantitative traits, and many excellent recent reviews exist [2–5]. Here, we specifically focus on recent discoveries that illuminate general principles of genetic complexity and its implications for understanding the genetic architecture of quantitative

Corresponding Author: Joshua M. Akey, Department of Genome Sciences, University of Washington, Foege Building, S303, Seattle, WA. 98195-7730, Phone: (206) 543-7254, Fax: (206) 685-7301, akeyj@u.washington.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

traits and complex diseases. Although we focus primarily on humans, important insights into the genetic architecture of phenotypic diversity gleaned from model organisms will also be discussed.

Architectural diversity across traits

There is enormous architectural diversity, or genetic complexity, across traits. For example, human diseases span the range of “simple” Mendelian traits, such as Cystic Fibrosis that result from mutations of large effect size in the *CFTR* gene [6], to exceedingly complex traits such as height that are likely influenced by thousands of variants and environmental factors [7]. Note, even simple Mendelian traits can exhibit striking levels of allelic and locus heterogeneity [8]. For instance, nearly 2,000 mutations have been identified in the *CFTR* gene that lead to Cystic Fibrosis [9], and variation at additional loci can modulate the severity of symptoms [10].

Furthermore, considerable architectural diversity is also observed even among closely related phenotypes. For example, numerous studies have been performed to elucidate the genetics of gene expression, where transcript abundance is treated as a quantitative trait and expression QTL (eQTLs) are mapped by linkage or association methods [11,12]. A powerful feature of this study design is that thousands of phenotypes are studied, allowing insights into the distribution of architectures across traits. The first eQTL study was performed in the budding yeast *Saccharomyces cerevisiae* [13] by crossing two divergent parental strains (BY and RM) and mapping eQTLs among the progeny. The median heritability of gene expression traits was 84% [13]. The majority of eQTLs detected, however, had modest effects and explained a median variance of 27%. Subsequent modeling showed that only 3% of highly heritable transcripts are consistent with single-locus inheritance, 17–18% are consistent with control by one or two loci, 50% of all transcripts have at least five additive QTLs, and 20% have at least 10 additive QTLs [14]. Thus, even in a simple eukaryotic organism grown in a controlled environment, the genetic architecture of gene expression levels encompasses a range of complexity.

Finally, it is important to note that formally defining a measure of architectural diversity that quantifies genetic complexity remains challenging. The number of loci that govern variability of a trait is often used as an implicit measure of genetic complexity. However, this is a rather crude summary that fails to account for potentially important architectural features such as gene-gene and gene-environment interactions. Recently, Thompson and Galitsky [15] proposed a more formal measure of genetic complexity, which they defined as the excess of genotypic diversity over phenotypic diversity, and used this framework to investigate the level and determinants of genetic complexity in Boolean networks. Interestingly, higher levels of genetic complexity were associated with topological features of the network and the number of periodic attractors (such as observed in cell cycle networks) [15]. Although this is an important step towards more rigorously defining and quantifying genetic complexity, additional theoretical work in this area is needed.

GWAs – so many associations, so little variation explained

With the advent of high-density SNP arrays, genome-wide association studies (GWAs) have been performed at a frenetic rate in humans. Indeed, a catalog of published GWAs [16] (<http://www.genome.gov/gwastudies/>) includes 1,659 publications and 10,986 associated SNPs as of July 24, 2013. As expected from this study design, most of the reported disease/trait associated SNPs are common with small effect sizes (i.e., odds ratio <1.5). Notably, the considerable number of significant associations fails to account for a substantial fraction of heritability for most traits studied to date, an observation that has been dubbed the “missing heritability” problem [5,17]. For example, 54 validated variants have been associated with human height, a classic complex trait with an estimated narrow sense heritability (the proportion of phenotypic variance that is due to additive genetic variance) of 80%; yet these 54 QTLs account for only ~5% of the variation in height observed among the tens of thousands of individuals studied [18]. Similarly, in a study of more than ten thousand individuals to identify loci that contribute to the risk of Crohn’s disease, the 32 identified loci explain ~20% of disease heritability [19]. Note, an association that explains only a small proportion of heritability may nonetheless provide valuable biological insights that could prove useful in designing drug targets [17,20]. However, the missing heritability revealed in many well-powered GWAs clearly highlights that many important architectural details of traits remain to be discovered. Below, we discuss features of genetic architecture that likely contribute to why GWAs fail to account for substantial amounts of heritability.

Detecting the undetectable – many loci of small effect

The infinitesimal model, proposed by Fisher nearly a century ago [21], has been a key conceptual foundation in the analysis of quantitative traits [22]. It posits that quantitative trait variation (or disease liability) is dictated by a large number of loci that have very small effect sizes. Thus, one potential explanation for why GWAs with even tens of thousands of individuals fail to reveal substantial amounts of heritability is simply because of inadequate statistical power. Consistent with this hypothesis, Yang et al. [23] examined variation in height in an Australian population and found that 294,000 SNPs examined could collectively explain 45% of the phenotypic variance, the vast majority of which fail to meet stringent significance thresholds because of their small effect sizes. Obviously, larger sample sizes can increase the statistical power to detect variants with weak effects, but there are practical limitations to detection, particularly for rare variants.

Model organism studies also clearly demonstrate that the architecture of some traits is governed by many loci of small effect. Specifically, a recent study in yeast measured growth rates in different environments and found that the 46 traits studied had an estimated narrow sense heritability of 52% [24]. A powerful genetic mapping approach (applicable only in model organisms) was used to identify 591 QTLs (ranging from 5 and 29 per trait). Strikingly, these QTL explained 72–100% of the measured heritability with a median of 88% [24]. In addition, the contribution to heritability of gene-gene interactions varies among traits, from near zero to approximately 50% [24]. In short, a large number of loci with weak phenotypic effects are clearly a prominent feature in the architecture of some quantitative and complex traits. However, it remains unclear whether this is the predominant cause for

the difficulty in mapping QTL and disease susceptibility or whether other architectural features are more important.

A Rare Portrait of Human Genetic Variation

Next-generation sequencing technology has dramatically increased the ability to assess patterns of genetic variation in thousands of individuals. To date, the largest surveys of human variation have primarily focused on subsets of the genome, such as the exome or set of all protein-coding regions [25]. Large-scale targeted gene or exome sequencing has revealed an unexpected abundance of rare variants [26,27], typically defined as a minor allele frequency less than 0.5% [27]. Strikingly, >80% of protein-coding variants are rare [27] largely due to the result of recent accelerated population growth over the last 5–10 thousand years [26–28]. Moreover, ~86% of single nucleotide variants (SNVs) predicted to be deleterious arose during this period [29]. Although these exact numbers will likely change as increasingly large samples are analyzed, they clearly reveal an abundance of rare, recently arisen mutations in humans, have led to the speculation that rare variants of modest to large effect size may be a source of missing heritability. Indeed, there have been notable examples of rare variants that influence quantitative traits and complex human disease [30–33], and recent findings suggest that *de novo* mutations play a prominent role in rare and common forms of neurodevelopmental diseases [34–39]. However, the relative contribution of rare and common variants to the burden of human disease remains unknown [4,40,41].

In considering the issue of rare versus common variation, it is important to note that while most protein-coding variation is rare at the population level, the majority of SNVs found in an individual are common [27]. For example, Figure 2 shows the proportion of protein-coding variants found in individuals as a function of minor allele frequency in the population using data from Fu et al. [29]. Among individuals, 17% of variants are found at a frequency of <0.5% in the population, while 65% are >5% on average. This paradoxical observation is simply due to the fact that while the absolute number of common variants is small relative to rare variants, they are shared among individuals. Thus, it seems likely that both rare and common variants play an important role in phenotypic variation. If the distribution of effect sizes for rare variants is commensurate with that of common variants, however, their detection will be much more difficult by standard population based association studies. Therefore, continued methodological development and innovative study designs will be critical to more comprehensively identifying rare variants that influence traits of interest.

Gene-Gene and Gene-Environment Interactions

The contribution of gene-gene and gene-environment interactions to human quantitative traits and complex disease remains unclear, but insights from model organisms unambiguously show that context dependent effects are pervasive. In yeast, comprehensive genetic interaction maps have been developed allowing global insights into the frequency and nature of interactions among loci [42]. Typically, large numbers of strains carrying pairwise mutations (encompassing all genes or defined subsets of genes) are constructed and interactions are identified by comparing the observed phenotype to that predicted from single locus mutants. In theory, such an approach could be used to interrogate the landscape

of genetic interactions for any phenotype of interest, but in practice most interaction maps have been constructed using growth rates or lethality as a phenotype [43–45]. The most salient observation that has emerged from the construction and analysis of yeast interaction networks is how extensive they are. For example, a yeast interaction network of synthetic lethality revealed ~200,000 pairwise deletions that resulted in lethality, whereas only 1,000 single-gene perturbations were lethal. Thus, ~200-fold more digenic mutant combinations can result in a lethal phenotype compared to a single gene alone. Such context dependent effects, where the phenotypic consequences of a variant are conditional on the genetic background significantly complicates the task of correlating genotypes with phenotypes in natural populations, including humans. Furthermore, context dependent effects in yeast are also a prominent feature in the genetic architecture of transcript abundance. For example, 57% of gene expression traits exhibit significant evidence of genetic interaction among eQTL [46,47] and 47% of transcripts have eQTL whose effects show gene-environment interactions [48].

Genetic interactions have also been systematically cataloged in multicellular model organisms. For example, in *Drosophila melanogaster* a map of pairwise gene-gene interactions was constructed by RNAi for 93 signal transduction genes [49] revealing 634 interactions that influenced a range of quantitative cellular phenotypes. Interestingly, only ~20% of interactions were shared among traits suggesting that different phenotypes reveal different functional relationships among loci. Furthermore, a recent study in flies found pervasive gene interactions for three complex life history [50] measured in the sequenced inbred lines of the *D. melanogaster* Genetic Reference Panel (DGRP) and an outbred population derived from 40 DGRP lines (referred to as Flyland). Surprisingly, none of the SNPs associated with traits in Flyland replicated in the DGRP and vice versa. Consistent with this observation, widespread interactions were found between SNPs in both Flyland and the DGRP, and thus their phenotypic effects were context dependent. Although genetic networks were constructed from the interactions separately in Flyland and DGRP, the resulting networks were found to be highly interconnected, suggesting commonalities between the genetic perturbations and biological pathways governing phenotypic variation in both mapping populations.

Although the evidence for gene and environmental interactions is compelling in model organisms, the lack of statistical power has greatly hampered efforts to robustly explore the landscape of non-additive gene action for human quantitative traits and complex diseases. Nonetheless, it seems likely that such interactions are an important component in the architecture of many human traits, and may contribute to the missing heritability phenomenon. For example, a recent study showed that interactions among loci can inflate estimates of total heritability, causing in turn estimates of narrow-sense heritability to be overestimated [51]. To illustrate the potentially large contribution interactions can make to estimates of variance explained, Zuk et al. [51] showed the amount of heritability accounted for in 71 Crohn's disease associated variants increased from 22% to 84% in models that considered interactions. Several notable examples of gene-environment interactions have also been discovered in humans, including a recent study of over 70,000 individuals that showed risk variants of several breast cancer susceptibility genes exhibit a strong dependency on environment variable such as the number of births (*LSP1*) and alcohol

consumption (*CASP8*) [52]. Despite these encouraging examples, detecting interactions in outbred populations and uncontrolled experimental settings remains a daunting challenge.

Conclusions and future directions

The rapid maturation of new technology to assay genetic variation and high-dimensional molecular phenotypes has led to many important insights into how genotypes map onto phenotypes. A key finding is that architectural features are highly heterogeneous across phenotypes, and that complex traits are complex for potentially different reasons such as being governed by many loci with small effect sizes or the product of non-additive interactions among loci and environments. Model organisms have led the way for much of our understanding about the complexities arising in relationship between genotype and phenotype, and we anticipate they will continue to be at the forefront of novel discoveries that help inform the architectural details of human phenotypic variation.

Despite the tangible progress made, the potentially small and context dependent effects a variant has on a particular phenotype poses considerable problems for interpreting human genetic variation and fulfilling the promise of personal genomics [53]. Although increasingly large sample sizes will improve power to detect variants of weak effect, there are diminishing returns on the information provided by such study designs, and opportunities to replicate findings in independent samples decreases. Furthermore, an important challenge to address moving forward is to translate statistical correlations between particular loci and traits into molecular, mechanistic, and biological insight. New technology and approaches are needed for large-scale functional characterization of genetic variants [54,55]. In this vein, systematic and large-scale phenotyping, referred to as phenomics [56], superimposed with whole-genome sequence data may be a particularly powerful framework for revealing general principles of genotype-phenotype relationships [57]. Ultimately, a deeper, mechanistic, and more principled understanding of the genetic architecture of quantitative and complex traits will have profound implications for basic and biomedical science.

Acknowledgements

This work was supported in part by NIH grants GM094810 and GM098360 to JMA.

References

1. Botstein D, Risch N: Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease. *Nat Genet* 2003, 33 Suppl:228–237. [PubMed: 12610532]
2. Mackay TF: The genetic architecture of quantitative traits. *Annu Rev Genet* 2001, 35:303–339. [PubMed: 11700286]
3. Mackay TF, Stone EA, Ayroles JF: The genetics of quantitative traits: challenges and prospects. *Nat Rev Genet* 2009, 10:565–577. [PubMed: 19584810]
4. Gibson G: Rare and common variants: twenty arguments. *Nat Rev Genet* 2011, 13:135–145.
5. Eichler EE, Flint J, Gibson G, Kong A, Leal SM, Moore JH, Nadeau JH: Missing heritability and strategies for finding the underlying causes of complex disease. *Nat Rev Genet* 2010, 11:446–450. [PubMed: 20479774]

6. Riordan JR, Rommens JM, Kerem B, Alon N, Rozmahel R, Grzelczak Z, Zielenski J, Lok S, Plavsic N, Chou JL, et al.: Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA. *Science* 1989, 245:1066–1073. [PubMed: 2475911]
7. Sinclair DC: *Human growth after birth* edn 5th Oxford; New York;: Oxford University Press; 1989.
8. Gilissen C, Hoischen A, Brunner HG, Veltman JA: Unlocking Mendelian disease using exome sequencing. *Genome Biol* 2011, 12:228. [PubMed: 21920049]
9. Drumm ML, Ziady AG, Davis PB: Genetic variation and clinical heterogeneity in cystic fibrosis. *Annu Rev Pathol* 2012, 7:267–282. [PubMed: 22017581]
10. Emond MJ, Louie T, Emerson J, Zhao W, Mathias RA, Knowles MR, Wright FA, Rieder MJ, Tabor HK, Nickerson DA, et al.: Exome sequencing of extreme phenotypes identifies DCTN4 as a modifier of chronic *Pseudomonas aeruginosa* infection in cystic fibrosis. *Nat Genet* 2012, 44:886–889. [PubMed: 22772370]
11. Skelly DA, Ronald J, Akey JM: Inherited variation in gene expression. *Annu Rev Genomics Hum Genet* 2009, 10:313–332. [PubMed: 19630563]
12. Rockman MV, Kruglyak L: Genetics of global gene expression. *Nat Rev Genet* 2006, 7:862–872. [PubMed: 17047685]
13. Brem RB, Yvert G, Clinton R, Kruglyak L: Genetic dissection of transcriptional regulation in budding yeast. *Science* 2002, 296:752–755. [PubMed: 11923494] ** The first study to systematically explore the genetic architecture of gene expression levels.
14. Brem RB, Kruglyak L: The landscape of genetic complexity across 5,700 gene expression traits in yeast. *Proc Natl Acad Sci U S A* 2005, 102:1572–1577. [PubMed: 15659551]
15. Thompson EG, Galitski T: Quantifying and analyzing the network basis of genetic complexity. *PLoS Comput Biol* 2012, 8:e1002583. [PubMed: 22792055]
16. Hindorff LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS, Manolio TA: Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci U S A* 2009, 106:9362–9367. [PubMed: 19474294]
17. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, McCarthy MI, Ramos EM, Cardon LR, Chakravarti A, et al.: Finding the missing heritability of complex diseases. *Nature* 2009, 461:747–753. [PubMed: 19812666]
18. Visscher PM: Sizing up human height variation. *Nat Genet* 2008, 40:489–490. [PubMed: 18443579]
19. Barrett JC, Hansoul S, Nicolae DL, Cho JH, Duerr RH, Rioux JD, Brant SR, Silverberg MS, Taylor KD, Barmada MM, et al.: Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease. *Nat Genet* 2008, 40:955–962. [PubMed: 18587394]
20. Sanseau P, Agarwal P, Barnes MR, Pastinen T, Richards JB, Cardon LR, Mooser V: Use of genome-wide association studies for drug repositioning. *Nat Biotechnol* 2012, 30:317–320. [PubMed: 22491277]
21. Fisher RA: The correlation between relatives on the supposition of Mendelian inheritance. *Transactions of the Royal Society of Edinburgh* 1918, 52:399–433.
22. Falconer DS, Mackay TFC: *Introduction to quantitative genetics*. New York, NY: Longman 1996.
23. Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, Nyholt DR, Madden PA, Heath AC, Martin NG, Montgomery GW, et al.: Common SNPs explain a large proportion of the heritability for human height. *Nat Genet* 2010, 42:565–569. [PubMed: 20562875]
24. Bloom JS, Ehrenreich IM, Loo WT, Lite TL, Kruglyak L: Finding the sources of missing heritability in a yeast cross. *Nature* 2013, 494:234–237. [PubMed: 23376951] * This study used a powerful mapping method to show that variants of weak effect can be a component of missing heritability.
25. Ng SB, Turner EH, Robertson PD, Flygare SD, Bigham AW, Lee C, Shaffer T, Wong M, Bhattacharjee A, Eichler EE, et al.: Targeted capture and massively parallel sequencing of 12 human exomes. *Nature* 2009, 461:272–276. [PubMed: 19684571]
26. Nelson MR, Wegmann D, Ehm MG, Kessner D, St Jean P, Verzilli C, Shen J, Tang Z, Bacanu SA, Fraser D, et al.: An Abundance of Rare Functional Variants in 202 Drug Target Genes Sequenced in 14,002 People. *Science* 2012, 337:100–104. [PubMed: 22604722]

27. Tennessen JA, Bigham AW, O'Connor TD, Fu W, Kenny EE, Gravel S, McGee S, Do R, Liu X, Jun G, et al.: Evolution and Functional Impact of Rare Coding Variation from Deep Sequencing of Human Exomes. *Science* 2012, 337:64–69. [PubMed: 22604720]
28. Keinan A, Clark AG: Recent explosive human population growth has resulted in an excess of rare genetic variants. *Science* 2012, 336:740–743. [PubMed: 22582263]
29. Fu W, O'Connor TD, Jun G, Kang HM, Abecasis G, Leal SM, Gabriel S, Rieder MJ, Altshuler D, Shendure J, et al.: Analysis of 6,515 exomes reveals the recent origin of most human protein-coding variants. *Nature* 2013, 493:216–220. [PubMed: 23201682] * This study explored patterns of protein-coding variation in over 6,000 individuals and showed most putatively deleterious variants are rare and arose very recently.
30. Cohen JC, Kiss RS, Pertsemlidis A, Marcel YL, McPherson R, Hobbs HH: Multiple rare alleles contribute to low plasma levels of HDL cholesterol. *Science* 2004, 305:869–872. [PubMed: 15297675] * One of the first examples showing that rare variants of moderate effect influence quantitative traits in humans.
31. Cohen J, Pertsemlidis A, Kotowski IK, Graham R, Garcia CK, Hobbs HH: Low LDL cholesterol in individuals of African descent resulting from frequent nonsense mutations in PCSK9. *Nat Genet* 2005, 37:161–165. [PubMed: 15654334]
32. Nejentsev S, Walker N, Riches D, Egholm M, Todd JA: Rare variants of IFIH1, a gene implicated in antiviral responses, protect against type 1 diabetes. *Science* 2009, 324:387–389. [PubMed: 19264985]
33. Walsh T, King MC: Ten genes for inherited breast cancer. *Cancer Cell* 2007, 11:103–105. [PubMed: 17292821]
34. Girard SL, Gauthier J, Noreau A, Xiong L, Zhou S, Jouan L, Dionne-Laporte A, Spiegelman D, Henrion E, Diallo O, et al.: Increased exonic de novo mutation rate in individuals with schizophrenia. *Nat Genet* 2011, 43:860–863. [PubMed: 21743468]
35. Neale BM, Kou Y, Liu L, Ma'ayan A, Samocha KE, Sabo A, Lin CF, Stevens C, Wang LS, Makarov V, et al.: Patterns and rates of exonic de novo mutations in autism spectrum disorders. *Nature* 2012, 485:242–245. [PubMed: 22495311]
36. O'Roak BJ, Deriziotis P, Lee C, Vives L, Schwartz JJ, Girirajan S, Karakoc E, Mackenzie AP, Ng SB, Baker C, et al.: Exome sequencing in sporadic autism spectrum disorders identifies severe de novo mutations. *Nat Genet* 2011, 43:585–589. [PubMed: 21572417]
37. Sanders SJ, Murtha MT, Gupta AR, Murdoch JD, Raubeson MJ, Willsey AJ, Ercan-Sencicek AG, DiLullo NM, Parikshak NN, Stein JL, et al.: De novo mutations revealed by whole-exome sequencing are strongly associated with autism. *Nature* 2012, 485:237–241. [PubMed: 22495306]
38. Vissers LE, de Ligt J, Gilissen C, Janssen I, Steehouwer M, de Vries P, van Lier B, Arts P, Wieskamp N, del Rosario M, et al.: A de novo paradigm for mental retardation. *Nat Genet* 2010, 42:1109–1112. [PubMed: 21076407]
39. Xu B, Roos JL, Dexheimer P, Boone B, Plummer B, Levy S, Gogos JA, Karayiorgou M: Exome sequencing supports a de novo mutational paradigm for schizophrenia. *Nat Genet* 2011, 43:864–868. [PubMed: 21822266]
40. McClellan J, King MC: Genetic heterogeneity in human disease. *Cell* 2010, 141:210–217. [PubMed: 20403315]
41. Lupski JR, Belmont JW, Boerwinkle E, Gibbs RA: Clan genomics and the complex architecture of human disease. *Cell* 2011, 147:32–43. [PubMed: 21962505]
42. Dixon SJ, Costanzo M, Baryshnikova A, Andrews B, Boone C: Systematic mapping of genetic interaction networks. *Annu Rev Genet* 2009, 43:601–625. [PubMed: 19712041]
43. Davierwala AP, Haynes J, Li Z, Brost RL, Robinson MD, Yu L, Mnaimneh S, Ding H, Zhu H, Chen Y, et al.: The synthetic genetic interaction spectrum of essential genes. *Nat Genet* 2005, 37:1147–1152. [PubMed: 16155567]
44. Tong AH, Evangelista M, Parsons AB, Xu H, Bader GD, Page N, Robinson M, Raghibizadeh S, Hogue CW, Bussey H, et al.: Systematic genetic analysis with ordered arrays of yeast deletion mutants. *Science* 2001, 294:2364–2368. [PubMed: 11743205]

45. Tong AH, Lesage G, Bader GD, Ding H, Xu H, Xin X, Young J, Berriz GF, Brost RL, Chang M, et al.: Global mapping of the yeast genetic interaction network. *Science* 2004, 303:808–813. [PubMed: 14764870]
46. Brem RB, Storey JD, Whittle J, Kruglyak L: Genetic interactions between polymorphisms that affect gene expression in yeast. *Nature* 2005, 436:701–703. [PubMed: 16079846]
47. Storey JD, Akey JM, Kruglyak L: Multiple locus linkage analysis of genomewide expression in yeast. *PLoS Biol* 2005, 3:e267. [PubMed: 16035920]
48. Smith EN, Kruglyak L: Gene-environment interaction in yeast gene expression. *PLoS Biol* 2008, 6:e83. [PubMed: 18416601]
49. Horn T, Sandmann T, Fischer B, Axelsson E, Huber W, Boutros M: Mapping of signaling networks through synthetic genetic interaction analysis by RNAi. *Nat Methods* 2011, 8:341–346. [PubMed: 21378980]
50. Huang W, Richards S, Carbone MA, Zhu D, Anholt RR, Ayroles JF, Duncan L, Jordan KW, Lawrence F, Magwire MM, et al.: Epistasis dominates the genetic architecture of *Drosophila* quantitative traits. *Proc Natl Acad Sci U S A* 2012, 109:15553–15559. [PubMed: 22949659]
51. Zuk O, Hechter E, Sunyaev SR, Lander ES: The mystery of missing heritability: Genetic interactions create phantom heritability. *Proc Natl Acad Sci U S A* 2012, 109:1193–1198. [PubMed: 22223662]
52. Nickels S, Truong T, Hein R, Stevens K, Buck K, Behrens S, Eilber U, Schmidt M, Haberle L, Vrieling A, et al.: Evidence of gene-environment interactions between common breast cancer susceptibility loci and established environmental risk factors. *PLoS Genet* 2013, 9:e1003284. [PubMed: 23544014] * An extensive study of the landscape of gene-environment interactions for breast cancer susceptibility genes in humans.
53. Gonzaga-Jauregui C, Lupski JR, Gibbs RA: Human genome sequencing in health and disease. *Annu Rev Med* 2012, 63:35–61. [PubMed: 22248320]
54. Bernstein BE, Birney E, Dunham I, Green ED, Gunter C, Snyder M: An integrated encyclopedia of DNA elements in the human genome. *Nature* 2012, 489:57–74. [PubMed: 22955616]
55. Patwardhan RP, Lee C, Litvin O, Young DL, Pe'er D, Shendure J: High-resolution analysis of DNA regulatory elements by synthetic saturation mutagenesis. *Nat Biotechnol* 2009, 27:1173–1175. [PubMed: 19915551] ** This study describes a clever approach for systematically studying the functional consequences of regulatory variation.
56. Freimer N, Sabatti C: The human phenome project. *Nat Genet* 2003, 34:15–21. [PubMed: 12721547]
57. Skelly DA, Merrihew GE, Riffle M, Connelly CF, Kerr EO, Johansson M, Jaschob D, Graczyk B, Shulman NJ, Wakefield J, et al.: Integrative phenomics reveals insight into the structure of phenotypic diversity in budding yeast. *Genome Res* 2013.

Figure 1. Schematic illustration of the mapping of genotypes onto phenotypes.

The top two rectangles represent the set of all possible genotypic and environmental states, respectively. Genetic architecture refers to the rules that govern how a set of multilocus genotypes maps onto phenotypes, and how the environment influences this mapping. Genotype space can be extremely large with 3^n possible multilocus genotypes, where n denotes the number of trait influencing variants. Lines represent the mapping of particular multilocus genotypes (red circles) in particular environments to trait values. For simplicity, phenotypic space is only shown for two traits (the height of the bivariate distribution indicates population frequency). Note, identical genotypes in different environments can culminate in different phenotypic values (representing gene-environment interactions) and

distinct genotypes either in the same or different environments can result in the same phenotypic values (representing robustness).

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 2. Most protein-coding variants found in individuals are common at the population level. The plot shows the cumulative proportion of protein-coding variants found in individuals as a function of minor allele frequency. The data used in this analysis is based on exome sequencing data from 4,298 European Americans and 2,217 African Americans [29]. Light pink lines show each of the 6,515 individuals and the dark red line is the median across all individuals.