

The Convergence of Systems and Reductionist Approaches in Complex Trait Analysis

Evan G. Williams¹ and Johan Auwerx^{1,*}

¹Laboratory of Integrative and Systems Physiology, École Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland

*Correspondence: admin.auwerx@epfl.ch

<http://dx.doi.org/10.1016/j.cell.2015.06.024>

Research into the genetic and environmental factors behind complex trait variation has traditionally been segregated into distinct scientific camps. The reductionist approach aims to decrypt phenotypic variability bit by bit, founded on the underlying hypothesis that genome-to-phenome relations are largely constructed from the additive effects of their molecular players. In contrast, the systems approach aims to examine large-scale interactions of many components simultaneously, on the premise that interactions in gene networks can be both linear and non-linear. Both approaches are complementary, and they are becoming increasingly intertwined due to developments in gene editing tools, omics technologies, and population resources. Together, these strategies are beginning to drive the next era in complex trait research, paving the way to improve agriculture and toward more personalized medicine.

Over the past century, great strides have been made toward identifying and understanding the genetic and environmental factors affecting complex traits. This progress has come from divergent though complementary genetic techniques that address similar biological hypotheses from different angles: the study of humans and crops versus model organisms, forward versus reverse genetics, holistic versus additive models, and so forth. Results from each of these approaches include the identification of thousands of variants influencing complex traits (Kingsmore et al., 2007) and the mechanistic delineation of hundreds of molecular pathways (Kanehisa et al., 2014). This knowledge has led to numerous practical benefits, ranging from the development of vaccines for transmissible diseases to drug treatments, dietary, and lifestyle changes for metabolic diseases to the rational modification of crops and livestock for agricultural needs. However, our ability to predict when, where, and how genetic, environmental, or gene-by-environment interactions (G × E) will lead to specific phenotypic outcomes is still limited, and few rationally generated cures or preventative strategies are available for common diseases such as cancer and diabetes. In large part, this shortcoming stems from the challenge of identifying key genetic and environmental regulators that can be targeted by drugs, lifestyle changes, or genetic modifications. For decades, the identification of new targets and the implementation of treatments have been hampered by the conceptual complexity and diversity of genetic mechanisms. This obstacle is now being overcome by a combination of improvements in technical capability, new approaches in scientific thought, and increased resource sharing. With this in mind, how can we use these developments to find new medical and biological breakthroughs? To answer this, we must first consider the current state of research into complex traits and how it has evolved over time (Figure 1A).

Developments in Complex Trait Analysis

Traditionally, studies on complex traits could be separated into two distinct categories: those using forward or reverse genetics (Figure 1B). Both seek to answer fundamental and longstanding questions on how genes, the environment, and G × E factors influence complex traits. In forward genetics, a variable phenotype is measured and the upstream causal genetic variants are identified, whereas reverse genetics starts at the gene level and searches for the downstream phenotypic impact. Pioneering genetics studies relied on the forward-genetics paradigm: natural populations and later, randomly mutagenized stocks were screened for variant phenotypes and were then successively backcrossed to identify the causal locus. In the 1970s, research platforms began to shift toward reverse genetics as the first targeted mutagenesis techniques were developed, and by the 1990s, this concept had come to the forefront in the study of complex traits. This capability of performing gain- or loss-of-function (G/LOF) studies on target genes allowed the rational and mechanistic examination of genetic hypotheses gene by gene and potentially even the reverse engineering of complex traits. Consequently, many researchers began to advocate for—and generate—comprehensive genetic libraries of G/LOF tools in a variety of organisms, including yeast (Winzeler et al., 1999), *Arabidopsis* (Alonso et al., 2003), *Drosophila* (Ryder et al., 2007), *C. elegans* (Kamath et al., 2003; Rual et al., 2004), and mice (Auwerx et al., 2004; Skarnes et al., 2011). Unfortunately, the reality of using these resources to efficiently and comprehensively identify novel variants behind complex traits has been undermined by two major factors. First, we now know that much natural trait variation is driven by both the additive and non-additive interaction of dozens or more variants (Bogardus et al., 2002; Clark, 2000) (Figure 1C). Second, major genetic alterations, such as those typically induced in G/LOF studies, are a poor model for the common variants influencing

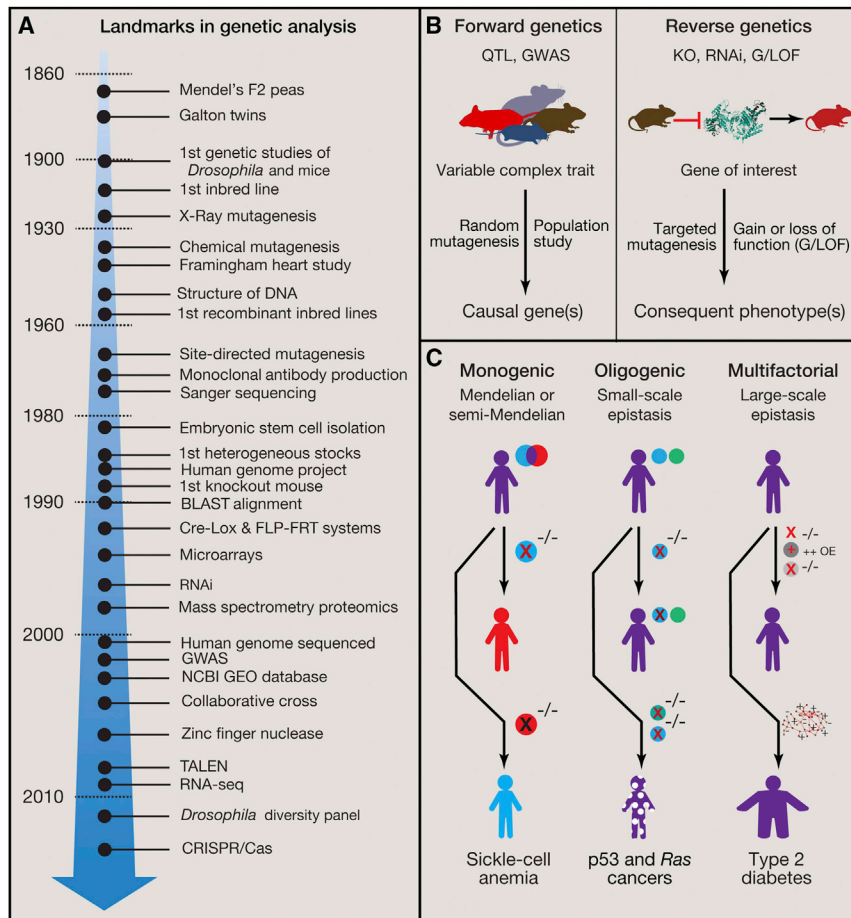


Figure 1. Broad Summary of Concepts and Developments in Genetic Analysis

(A) Relative timeline of complex trait analysis, listing some key landmarks and their approximate dates.

(B) (Left) The forward-genetics approach relies on identifying divergent phenotypes and then searching for the causative genetic factors, generally through QTL analysis or GWAS. (Right) The reverse-genetics approach relies on modifying a gene (or genes) of interest and then scanning for impact on downstream traits, often by G/LOF modifications.

(C) Single genes can fully regulate a phenotype (e.g., sickle-cell trait), or they can play a role in more complex phenotypes for part of the population (e.g., *PPAR* γ and metabolic disease; Deeb et al., 1998). However, complex trait variation also comes from epistasis (e.g., mutations in *p53* and other genes together are necessary to develop cancer; Soussi et al., 1994). To fully explain heritability and complex trait variation for an entire population (e.g., diabetes), it is likely that we will need the ability to model and test arbitrarily complex interaction of dozens or more variants simultaneously. In this cartoon, examples are given in which one, two, or many genes must be modified in order to change the observed phenotype (color, monogenic; spots, oligogenic; or size, polygenic).

scriptomic datasets. While the resulting and unprecedentedly thorough data sets aided both population and G/LOF research, they particularly boosted the population approach. In theory, omics coverage could provide the capacity to identify causal gene networks wholesale

trait variation in natural populations, which are generally more subtle (Chakravarti et al., 2013; MacArthur et al., 2012). Minor variants, gene \times gene, and G \times E interactions can be examined mechanistically using modern G/LOF tools, yet the exponential increase in the number of such possibilities as complexity expands necessitates the use of prior hypotheses instead of unbiased screens, particularly for vertebrate research. Finally, mechanisms that are uncovered in G/LOF models may not necessarily be generalizable to natural populations, whether in humans or in agriculture. These limitations of G/LOF models were recognized from the outset (Capecchi, 2005), but potential alternatives—particularly, population genetics—suffered from strong deficits as well.

In parallel to the developments in forward- and reverse-genetics techniques, progress continued steadily on molecular measurement technologies that expanded the scope and depth of genetic analysis. The debut of what has become the “omics revolution” began with massive investments in large-scale nucleotide sequencing (Smith and Hood, 1987), biological applications of mass spectrometry (Fenn et al., 1989; Wasinger et al., 1995), and array technology (Schena et al., 1995). By the late 1990s, the genomic and transcriptomic tools were sufficiently refined and affordable that small collaborative groups had the capability to generate and test hypotheses that required full pathway analysis by using comprehensive genomic and tran-

scriptomic datasets. While the resulting and unprecedentedly thorough data sets aided both population and G/LOF research, they particularly boosted the population approach. In theory, omics coverage could provide the capacity to identify causal gene networks wholesale through data-driven approaches—even directly in humans. Indeed, initial results using this approach to study common complex disorders were promising, as exemplified by the identification of variants in two genes, *PPAR* γ (Deeb et al., 1998) and *MC4R* (Yeo et al., 1998), causal for metabolic disease. However, human population studies examining such genome-to-phenome links (e.g., genome-wide association studies [GWAS]) ran into several major barriers, among them the issues of linkage disequilibrium, commonly detected SNPs having small effect sizes, poor long-term environmental control, and the perennial issue of “missing heritability” (Goldstein, 2009; Lander, 2011). Furthermore, while genotype information is fairly consistent across time and tissue, the ephemeral nature of transcripts, proteins, and metabolites hindered detailed mechanistic analyses in human populations due to the difficulty or impossibility—depending on tissue—of obtaining biopsies. In retrospect, we have now seen that human GWAS led only to a slow trickle of discoveries between novel gene variants and complex traits (McCarthy et al., 2008).

In principal, some of these issues could be bypassed by analyzing diverse populations of model organisms, the generation and application of which are relatively similar cross-species (Flint and Mackay, 2009) (Figure 2A). Such populations fall broadly into two groups: those of genetically unique individuals such as F2s and outbreds and those in specific and reproducible

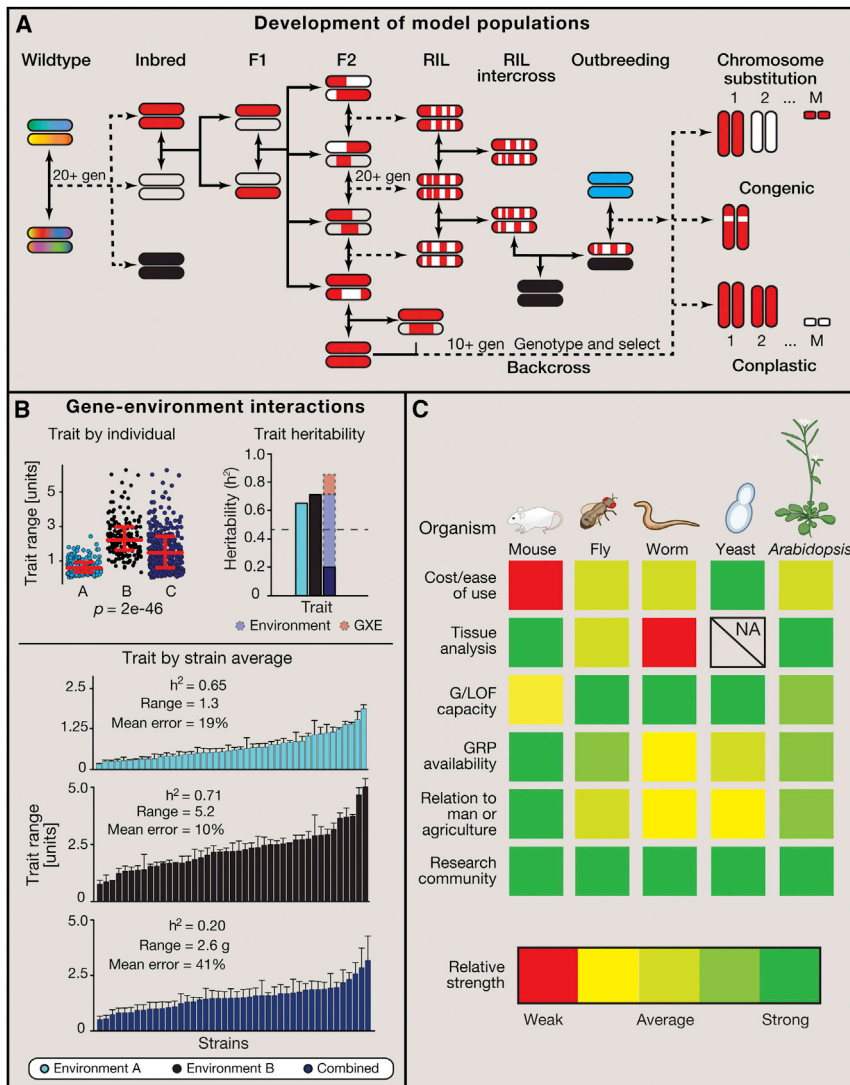


Figure 2. Model Population Development and Analysis

(A) A summary and basic breeding schema for major population types. F2s are generated ad hoc, which can be studied or crossed and inbred for 20+ generations to generate recombinant inbred lines (RIL). F2s can be backcrossed to generate congenic, consomic, or conplastic strains.

(B) Heritability should be considered as a function of gene-environment interactions, especially for complex traits influenced by environmental factors. (Upper-left) Across a diverse GRP, the observed trait is significantly elevated in "Environment B." (Lower-left) Despite the difference in trait expression across environments, the heritability is similar when groups are segregated, though it plummets when the groups are combined. If the environmental difference can be controlled and groups separated, as here, the environment and G \times E factors can be calculated by two-way ANOVA. (Right) When the same data are displayed as strain averages, the heritability drop can be visualized. Within either environment, variability due to genotype (y axis) is far lower than the error within a genotype (error bars). When these data are compressed without respect to environment, the cross-genotype variance compared to within-genotype variance decreases, hence the drop in observed heritability.

(C) A broad overview of relative species particularities and strengths for the most common model organisms in population genetics. The suitability of any model also depends substantially on the question to be addressed and specific experimental design.

has been developed across many model organisms that can simulate many aspects of the genetic complexity of natural populations (human, animal, or plant) but in controlled settings (Churchill et al., 2004; Kover et al., 2009; Mackay et al., 2012) (Table 1).

In addition to solving some of the key issues of human cohort studies, including

sample collection and long-term environmental control, model populations can also be used to address longstanding genetics questions. For instance, complex traits that are strongly influenced by G \times E factors may lead to incorrect calculations stating that these traits are not highly heritable, when in fact the heritability may be influenced by genes modulated by variable but unknown environmental factors (Figure 2B). When selecting which model organism (or model organisms) to examine, many trade-offs must also be considered, including: (1) speed, expense, efficiency, and technical concerns such as recombination events and self-fertilization or inbreeding depression; (2) ability to perform tissue and time-specific molecular phenotyping; (3) availability of gene-editing tools/libraries; (4) availability of population resources; (5) similarity to humans, agriculture, or livestock; and (6) the resource community surrounding the model (Figure 2C). Additionally, certain model organisms can provide unique capacity for study designs that would be infeasible or even impossible in other species. For example, one recent study exploited genetically diverse yeast to create sets of millions of individuals

genetic reference populations (GRPs). However, the implementation of these concepts during the early years of high-throughput sequencing and microarray transcriptomics was hindered by small cohort sizes and the limited selection of ready-made GRPs (Williams et al., 2001). The earliest GRPs, such as the BXD mice (Taylor et al., 1973), had been developed and utilized decades previously, primarily for forward-genetics analyses. However, such early studies often linked phenotypic variants to broad quantitative trait loci (QTLs) containing dozens to hundreds of candidate genes; thus, the specific causative genes were rarely identified (Flint et al., 2005). This shortcoming could be largely solved by analyzing larger populations and by improving recombination density (Darvasi and Soller, 1995), yet the development of GRPs is expensive and can take many years to generate, particularly in vertebrates. Thus, little development was done on expanding model organism populations—even in invertebrates—until technological advances in omics and the shortcomings (and benefits) of human population research had been well-recognized. Since then, a diverse collection of GRPs

Table 1. Landmark Dates and Papers in the Generation of Model Populations, along with Key Populations or First Populations in a Variety of Species

Date	Resource	Organism	Reference
	Wild-Type	(All)	
~1866	F2 intercross	Pea	G.J. Mendel
~1909	Inbred lines	Mouse	C.C. Little
1959	Recombinant inbred (CXB)	Mouse	(Bailey, 1981)
1971	Recombinant inbred (BXD)	Mouse	(Taylor et al., 1973)
1982	Recombinant inbred (BXH/HXB)	Rat	(Pravenec et al., 1989)
1984	Diversity panel (N/Nih)	Rat	(Li and Lumeng, 1984)
1988	Recombinant inbred (WSxM13)	Corn	(Burr et al., 1988)
1996	Recombinant inbred (WSxM13)	<i>Arabidopsis</i>	(Liu et al., 1996)
1997	Recombinant inbred (Rx2b)	<i>Drosophila</i>	(Nuzhdin et al., 1997)
2000	Chromosome substitution	Mice	(Nadeau et al., 2000)
2001	Recombinant inbred (BOxRC301)	<i>Caenorhabditis</i>	(Ayyadevara et al., 2001)
2002	Recombinant inbred	Soybean	(Yuan et al., 2002)
2003	Recombinant inbred	Barley	(Arru et al., 2003)
2004	Advanced recombinant inbred (BXD)	Mouse	(Peirce et al., 2004)
2004	Recombinant inbred (LXS)	Mouse	(Williams et al., 2004)
2004	Advanced recombinant inbred (collaborative cross)	Mouse	(Churchill et al., 2004)
2005	Diversity panel (heterogeneous stock)	<i>Caenorhabditis</i>	(Sivasundar and Hey, 2005)
2006	Diversity panel (heterogeneous stock)	Mouse	(Valdar et al., 2006)
2008	Diversity panel (1001 Genomes Project)	<i>Arabidopsis</i>	(Ossowski et al., 2008)
2009	Advanced recombinant inbred (MAGIC)	<i>Arabidopsis</i>	(Kover et al., 2009)
2009	Advanced recombinant inbred (N2xCB4856)	<i>Caenorhabditis</i>	(Rockman and Kruglyak, 2009)
2010	Chromosome substitution	Rice	(Xu et al., 2010)
2010	Diversity panel (hybrid mouse diversity panel)	Mouse	(Bennett et al., 2010)
2011	Recombinant inbred (four-way cross)	Yeast	(Cubillos et al., 2011)
2012	Diversity panel (DGRP)	<i>Drosophila</i>	(Mackay et al., 2012)
2012	Advanced recombinant inbred (DSPP)	<i>Drosophila</i>	(King et al., 2012)

While the technical ability to generate most of these systems has existed for decades, it is only within the last ~15 years that they have gained prominence, due primarily to developments in omics technologies. All dates are approximate, given that it can take many years between initial population conceptualization and their first publication.

that in turn display gene expression from very low (akin to “knockouts”) to average (akin to “wildtype”) to very high (akin to “transgenic overexpression”)—and everywhere in between (Albert et al., 2014). In such extensive populations, nearly all genes display very strong levels of variability. Moreover, the power to detect QTLs depends largely on the variance of the traits of interest. Thus, this approach (dubbed extreme QTL, or “X-QTL,” mapping) increases both the scale and scope of findings as compared using smaller cohorts. However, two reasons limit the generation of X-QTLs in more complex organisms. The first is due to scale, as analyzing millions of individuals remains logistically and economically prohibitive in most organisms, and the second due to technique, as the expression of any gene can be examined visually in yeast (and *C. elegans*) by tagging target genes with green fluorescent protein (GFP). As such, gene expression may be assessed across millions of individuals visually, groups may be separated into a few clusters based on expression, and omics analysis can be performed only on these subsets. For more complex organisms, such as

mice, omics analyses would be necessary for all millions of individuals to capture the full spectrum of variance.

Model populations can also address many of the shortcomings of studies using G/LOF models or single inbred lines. For instance, early drug trials in model organisms are predominantly performed as experiments with a genotypic $n = 1$. The reality of complex trait genetics, however, means that even a reproducible finding in one individual may not apply in another. For example, had researchers exclusively used the DBA/2J mouse strain in place of the prototypical C57BL/6J, they would have concluded that morphine is a non-addictive and inefficient painkiller (Elmer et al., 2010). Examples like this have led to a recognition that genetic diversity must be considered at all steps from target identification to final outcome, an approach that is beginning to be implemented medically (Barretina et al., 2012) and that is starting to assist in the design of compounds that are highly effective in particular, defined subsets of patients (e.g., *CFTR* variant-specific drug treatments for cystic fibrosis; Ledford, 2012). However, simplified models remain essential: early drug screens of dozens

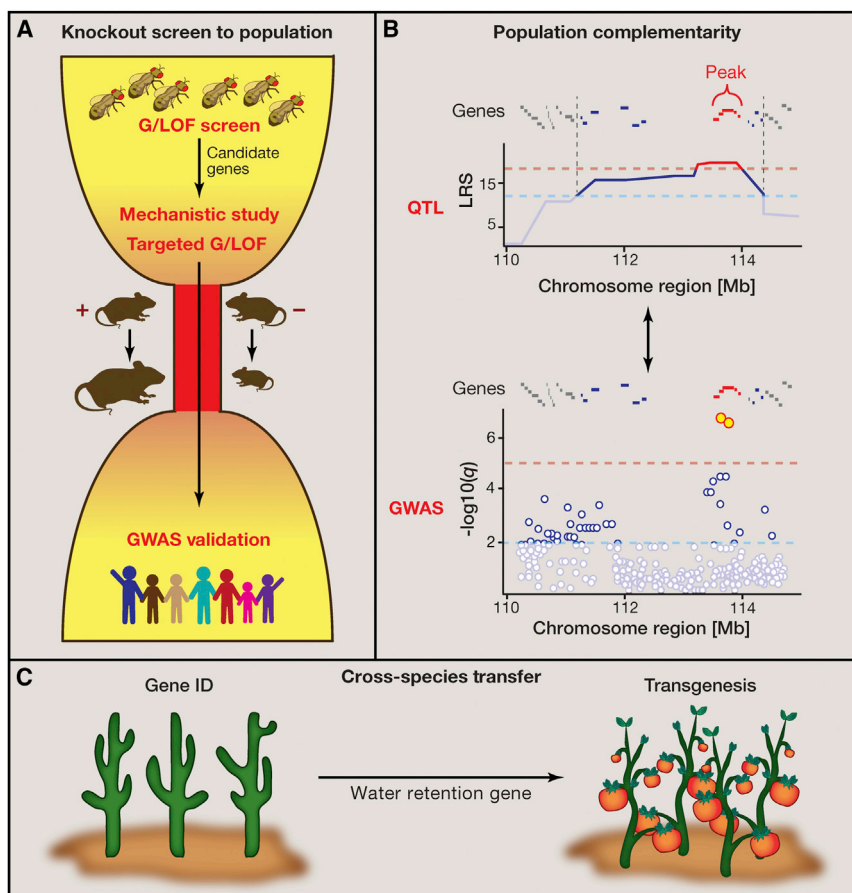


Figure 3. A Summary of a Few Study Designs Using Cross-Model and/or Cross-Species Approaches

(A) Conceptual schematic of G/LOF screenings performed in simpler model organisms (e.g., *Drosophila*), which may be used to generate targeted hypotheses for G/LOF studies in more complex (e.g., mammalian) models. These results may then provide targeted hypotheses for validation in human GWAS.

(B) Cross-model example of population studies benefitting from one-another. QTL results do not indicate a specific gene, and equivalent GWAS results, displayed as a Manhattan plot, would not approach significance on a genome-wide scan. Furthermore, candidate SNPs in GWAS may be in linkage disequilibrium. The complementarity of independent population studies can often be applied across species by comparing syntenic regions.

(C) Conceptual approach of gene transfer. Natural genetic pathways that allow certain plants to resist stresses such as drought or disease may potentially be translated into other species via transgenesis.

biased screens in invertebrates, thus providing specific hypotheses for G/LOF work in mammals—and eventually for the clinic (Figure 3A). Population studies as well may benefit from cross-species analysis—from GRP to GWAS or vice versa—such as in assisting the identification of causal genes under a QTL or within linkage disequilibrium (Figure 3B). Finally,

of novel compounds to treat an illness cannot be efficiently performed if they are each thoroughly tested in extensive GRPs. These tradeoffs between comprehensive genetic models (which may be excessively complex) and simplified models (which may be oversimplified) must be considered at each stage of the research process. This realization is beginning to lead to the complementary implementation of these approaches throughout the pipeline of complex trait analysis.

Complementary Approaches in Complex Trait Genetics

The parallel expansion of gene editing capabilities, omics technologies, and population resources has created the capability to implement experiments incorporating mechanistic *in vitro* experiments, comprehensive studies in G/LOF organisms and model populations, and research in natural populations. This complementarity is multidirectional. In one way, *in vitro* mechanistic studies can provide detailed hypotheses for *in vivo* verification using reductionist models, then tested in diverse genetic backgrounds for potential interaction effects of the variant, and finally examined outside of the laboratory setting. Conversely, population studies—often beset by false discovery—may use G/LOF models to distinguish several potential causal factors or to obtain detailed mechanistic understanding. In general, study design can also be streamlined by considering complementary cross-species techniques. For instance, the *de novo* identification of gene-phenotype links can be readily established using un-

it is not fundamentally necessary for the target genes, or even mechanisms, to be conserved across species to derive utility from a cross-species or cross-model approach. Discovery of unique aspects of an organism can be as informative as the commonalities, such as the key differences in atherosclerosis development between mice and humans (Jiang et al., 1992). Furthermore, particularly for agriculture, the identification of species-specific genes and mechanisms can also assist in transgenic development of crops, which are resistant to particular diseases or environmental conditions (Figure 3C).

The power of these combined approaches to complex trait analysis can be illustrated by a few recent studies. For instance, population research in any organism can be used to identify gene-phenotype relationships *en masse*. However, while only a few gene candidates may be practically assessed in vertebrate G/LOF studies, hundreds of genes can be analyzed in yeast, worms, or flies (related to Figure 3A). In one specific recent example, a study screened and phenotyped 11,594 transgenic lines of *Drosophila* for adiposity-related phenotypes (Pospisilik et al., 2010). For the ~500 positive hits, the authors performed mechanistic analyses to define the links between each candidate and adiposity and to inform upon the strongest candidates for G/LOF experiments in mice. Subsequently, fat-specific knockout mice were generated for the top candidate gene *Sufu*, which displayed robust changes in fat mass. Such large screens can also be used to inform directly upon human GWAS, as shown in a

separate recent example comparing results from a genetic loss-of-function screen in *Drosophila* to rare diseases in human exome data (Yamamoto et al., 2014). The complementarity of these cross-species exchanges can be quite surprising; for instance, genes linked to neurological diseases in humans can be effectively analyzed in reductive plant models (Xu and Møller, 2011).

In other cases, both hypothesis generation and validation may be performed using population studies, including for humans and mice (related to Figure 3B). For instance, a recent extensive metabolic study of the BXD mouse GRP identified more than a dozen novel and significant phenotype QTLs (Andreux et al., 2012). The QTL for systolic blood pressure, with five candidate genes, was prioritized for validation due to the small number of possibilities and, moreover, the ready availability of independent population studies that measured blood pressure (Koutnikova et al., 2009). In the subsequent analysis of three independent human GWAS, SNPs in a single gene, *Ubp1*, were associated consistently and significantly with elevated blood pressure. Although these SNPs were nominally significant in the original data, the correction for multiple testing across hundreds of thousands of SNPs across the whole genome drowned out the signal. Details about cellular mechanisms can also be confidently attributed solely using population data. In a separate BXD study, it was observed that diabetic mice have significantly lower levels of the metabolite 2-aminoadipate, independently from the effects of low- or high-fat diet feeding (Wu et al., 2014). In turn, the levels of this metabolite mapped significantly to the gene locus containing *Dhtkd1*, a known enzyme in 2-aminoadipate metabolism (Danhauser et al., 2012). Thus, a mechanistic link was identified between *Dhtkd1* and diabetes via the regulation of 2-aminoadipate. Furthermore, two independent populations (one mouse, one human) were examined using this hypothesis, and again the same links with diabetes were observed. As before, multiple testing correction had prevented the identification of this connection in the original data sets (Wu et al., 2014).

Finally, human medical application is not the only goal: diseases uniquely affecting animals (e.g., rinderpest, foot-and-mouth disease) and plants (e.g., coffee rust, black rot) are equally important. Additionally, genetic engineering in agriculture in response to environmental needs or changes (e.g., drought resistance, improved yield) is a critical research goal, as we must contend with expanding populations and general climatic variation. For instance, gene-phenotype links identified in *Arabidopsis*—itself not a crop—are often applied to agriculture. Mutagenesis screens in *Arabidopsis* have identified genetic variants influencing flowering promotion (e.g., *SFT*) and flowering repression (e.g., *SP*), which can be subsequently applied cross-species (related to Figure 3C). In tomato, the orthologs of these genes were rationally modified, leading to a striking 130% variation in tomato yield between the suppressed and promoted plants (Park et al., 2014). Similar cross-species plant studies have improved cold tolerance in tobacco (Zhao et al., 2009), drought resistance in rice (Datta et al., 2012), and salt resistance in barley (Schilling et al., 2014). Research for agricultural genetics is often parallel to medical genetics, yet this divergence is not fundamental, and the two can benefit equally from sharing technologies, methodologies, and findings.

Big Data

Moving across species and between or among populations and G/LOF models can provide a greater understanding of biological mechanisms underlying complex traits. However, the in vivo application of such resources still requires extensive experience and potentially time and expense. To this end, it is essential to consider and exploit the wealth of results available in public “big data” resources, which can allow hypothesis examination in silico. Over the last 10 years, the generation of full genomic and transcriptomic data sets has become commonplace. The full DNA sequence of more than 100 species on the UCSC Genome Browser (Kent et al., 2002), more than 1.2 million complete transcript data sets on the Gene Expression Omnibus (GEO) (Barrett et al., 2013), and dozens of other smaller repositories have sprung up as well, often organized per species or per topic (Bastian et al., 2008; Chesler et al., 2004; Harris et al., 2010). Conjointly, a wide array of software and web resources have been developed to aid in the analysis and interpretation of large systems data sets (Barrett et al., 2013; Subramanian et al., 2005; Wang et al., 2003). These resources can provide secondary study validation, study power, and founding hypotheses and can inform decisions on study design. These possibilities are rapidly expanding, as studies increasingly generate more data than can be analyzed under the scope of a single program. When made public, this “excess” information adds to a treasure trove for secondary analysis and independent validation. Consequently, primary research projects are increasingly taking advantage of independent research beyond the introduction and verbatim citations—whether for hypothesis generation, validation, or complete meta-analysis (Khan et al., 2014; Shin et al., 2014).

Historical data can also grow in value over time as additional, diverse data sets are collated. For example, early QTL studies often identified strong gene-phenotype links but were unable to establish the causal genes. These “unfinished stories” now serve as pre-made hypotheses, such as for detailed longevity data recorded in the BXD GRP well before the common implementation of omics technologies (De Haan and Van Zant, 1999; Gelman et al., 1988). These two early studies observed a common, significant QTL on chromosome 2 that contributed about 30% to the overall variation in lifespan, but the causal gene(s) were not defined, and the QTL went unproven. More than a decade later, using modern genetic maps, transcriptomic data, and information on sequence variants, specific candidate genes were found under the QTL. Subsequently, two of these genes (*mrps-5* and *nkcc-1*) were validated in *C. elegans* using G/LOF technologies that, again, did not exist at the time of the initial study—even if candidate genes had been identified (Houtkooper et al., 2013). Further mechanistic studies on *mrps-5* in *C. elegans* and in an independent set of BXDs were used to uncover that the longevity effects stemmed from the dysregulation of mitochondrial protein translation. The resulting stoichiometric imbalance between mitochondrial and nuclear-encoded proteins, dubbed mitonuclear imbalance, induces the mitochondrial unfolded protein response (known as UPR^{mi}), a known adaptive stress response protecting against aging (Jovaisaite et al., 2014). In other cases, rather than taking old data sets piecemeal, results may be taken together and collated. Decades

of detailed electronic medical records at hospitals now serve as the backbone for a recent approach called phenome-wide association studies (PheWAS)—essentially, the reverse-genetics analog of GWAS (Denny et al., 2013). In GWAS, a phenotype is queried against genotypes to determine the causative gene(s), whereas in PheWAS, a gene or SNP of interest is queried against phenotypes to determine associated traits. This is conceptually straightforward, yet to be effective, this approach requires extensive and comprehensive sets of phenotype data—a far more expensive and time-consuming task to generate than for genotype data. Some successes have already been reported with this approach, e.g., *FTO*—traditionally an obesity gene—was recently linked to fibrocystic breast disease by PheWAS (Cronin et al., 2014).

The increasing richness of omics data from both populations and G/LOF models also allows meta-analytical studies to detect and validate novel findings completely in silico (Horvath, 2013; Lee et al., 2012). In one recent high-profile example of meta-analysis, thousands of human methylation arrays spanning 51 tissues were cross-referenced with a single basic phenotype—the age of the patient (Horvath, 2013). A statistical model was then developed to accurately calculate the ages of the patients using only tissue samples and a set of DNA methylation sites. Using this model, it was observed that cancers have “older” methylation patterns than healthy cells; for instance, calculating the age of a 30-year-old breast cancer patient using a blood sample would yield an age of 27–33 years, while the same calculation using a breast tumor biopsy would yield an age of ~75 years. The confidence in the study results was made possible by the wealth of independently generated methylation data sets: half of the studies were used to generate hypotheses and the other half to test hypotheses, thus providing experimental validation without a need for “new” data. The crossover and “re-use” of data is not new, yet the viability of this approach has rapidly improved due to the remarkable increase in the quality and scope of omics data sets. However, the informatic difficulty in large-scale manipulation of this approach has thus far largely restricted its application to computational biology (Marx, 2013). Correspondingly, there has been a disconnect between bioinformatics groups working in silico and generating non-validated networks of profound complexity and traditional in vivo and in vitro biologists elucidating biochemical processes that may be only applicable in the specific circumstances of a single experiment. As with the converging applications of population genetics and G/LOF models, computational and wet lab biologists are now beginning to find common ground; bioinformatics platforms are becoming more widely established both institutionally and in individual wet labs.

Future Perspectives

The recent fundamental shifts we have discussed are beginning to drive the next era in complex trait analysis, yet it is only the next step of many in our push to understand genetics. We have mastered the ability to model and analyze monogenic-gene-to-phenotype interactions, both through their identification by forward genetics and by using precise reverse-genetics modifications to elucidate their mechanisms. Now, we are developing the capabilities to comprehensively identify and understand

more expansive oligogenic gene-to-phenotype and $G \times E$ -to-phenotype interactions (see Figure 1C). Despite these advances, a complete holistic understanding of systems biology is still beset by a number of challenges ahead: (1) there remains a need for higher-quality and more dynamic omics measurements; (2) epistatic gene modeling requires exponentially increasing sample sizes; (3) even the best statistical assurances of causality do not preclude a necessity for experimental validation, at least so far; and (4) there is no “ideal” experimental setup for the study of complex traits. For the first challenge, the next stage is already in sight, with single-cell dynamic profiling (Cai et al., 2006), quantitative and unbiased metabolite analysis (Fuhrer et al., 2011), larger-scale untargeted SWATH proteomics (Röst et al., 2014), and other such approaches breaking new ground. To the second point, model populations are becoming larger and more refined (Churchill et al., 2004), and G/LOF tools are becoming increasingly multiplexed (Chen et al., 2015; Sander and Joung, 2014)—together allowing identification and validation of moderate-scale epistatic interactions. The third obstacle may perhaps be addressed in time, but there remains little indication that accurate mathematical models are on the horizon, except for a few specific biomolecular processes such as protein folding. The last point is more fundamental: a new experiment will always be necessary for testing a new drug or treatment, regardless of the improvements in predictive models or genomics resources.

Finally, further changes in mindset are required. As hypotheses become larger and more complex, data and experimental methods must be more readily communicated within and outside of the standard publication cycle. In academic research, this need is now regularly accepted, yet a great deal of data and methods are not made available with publications, either due to privacy concerns (e.g., human data), a lack of a standard database (e.g., raw mass spectrometry data), or even active obfuscation (e.g., in patents, agricultural [Zamir, 2014], or pharmaceutical [Prayle et al., 2012] research). However, given the tremendous progress in resource sharing within the last decade, we expect many of these concerns to diminish in the coming years. There is a revolution in complex trait analysis well underway, and together the combined applications of systems biology and reductionist approaches are beginning to turn the trickle of genetic discoveries into a steady stream that will usher in a new era for personalized medicine and for environmentally safe genetically modified crops.

AUTHOR CONTRIBUTIONS

E.G.W. and J.A. outlined and designed the manuscript. E.G.W. drafted the manuscript and prepared figures. J.A. approved and edited the manuscript.

ACKNOWLEDGMENTS

We thank Dr. Xu Wang and Dr. Robert W. Williams for discussions on systems biology and cross-species approaches. J.A. is the Nestlé Chair in Energy Metabolism, and the J.A. laboratory is supported by grants from the École Polytechnique Fédérale de Lausanne, the Swiss National Science Foundation (31003A-140780), the Sinergia grant mechanism (CSRII3-136201), the AgingX program of the Swiss Initiative for Systems Biology (51RTPO-151019), and the NIH (R01AG043930).

REFERENCES

- Albert, F.W., Treusch, S., Shockley, A.H., Bloom, J.S., and Kruglyak, L. (2014). Genetics of single-cell protein abundance variation in large yeast populations. *Nature* 506, 494–497.
- Alonso, J.M., Stepanova, A.N., Leisse, T.J., Kim, C.J., Chen, H., Shinn, P., Stevenson, D.K., Zimmerman, J., Barajas, P., Cheuk, R., et al. (2003). Genome-wide insertional mutagenesis of *Arabidopsis thaliana*. *Science* 301, 653–657.
- Andreux, P.A., Williams, E.G., Koutnikova, H., Houtkooper, R.H., Champy, M.F., Henry, H., Schoonjans, K., Williams, R.W., and Auwerx, J. (2012). Systems genetics of metabolism: the use of the BXD murine reference panel for multiscalar integration of traits. *Cell* 150, 1287–1299.
- Arru, L., Francia, E., and Pecchioni, N. (2003). Isolate-specific QTLs of resistance to leaf stripe (*Pyrenophora graminea*) in the 'Steptoe' x 'Morex' spring barley cross. *Theor. Appl. Genet.* 106, 668–675.
- Auwerx, J., Avner, P., Baldock, R., Ballabio, A., Balling, R., Barbacid, M., Berns, A., Bradley, A., Brown, S., Carmeliet, P., et al. (2004). The European dimension for the mouse genome mutagenesis program. *Nat. Genet.* 36, 925–927.
- Ayyadevara, S., Ayyadevara, R., Hou, S., Thaden, J.J., and Shmookler Reis, R.J. (2001). Genetic mapping of quantitative trait loci governing longevity of *Caenorhabditis elegans* in recombinant-inbred progeny of a Bergerac-BO x RC301 interstrain cross. *Genetics* 157, 655–666.
- Bailey, D.W. (1981). Recombinant inbred strains and bilineal congenic strains (New York: Academic Press).
- Barretina, J., Caponigro, G., Stransky, N., Venkatesan, K., Margolin, A.A., Kim, S., Wilson, C.J., Lehár, J., Kryukov, G.V., Sonkin, D., et al. (2012). The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* 483, 603–607.
- Barrett, T., Wilhite, S.E., Ledoux, P., Evangelista, C., Kim, I.F., Tomashevsky, M., Marshall, K.A., Phillippy, K.H., Sherman, P.M., Holko, M., et al. (2013). NCBI GEO: archive for functional genomics data sets—update. *Nucleic Acids Res.* 41, D991–D995.
- Bastian, F., Parmentier, G., Roux, J., Moretti, S., Laudet, V., and Robinson-Rechavi, M. (2008). Bgee: Integrating and comparing heterogeneous transcriptome data among species. *Lect. N. Bioinform.* 5109, 124–131.
- Bennett, B.J., Farber, C.R., Orozco, L., Kang, H.M., Ghazalpour, A., Siemers, N., Neubauer, M., Neuhaus, I., Yordanova, R., Guan, B., et al. (2010). A high-resolution association mapping panel for the dissection of complex traits in mice. *Genome Res.* 20, 281–290.
- Bogardus, C., Baier, L., Permana, P., Prochazka, M., Wolford, J., and Hanson, R. (2002). Identification of susceptibility genes for complex metabolic diseases. *Ann. N Y Acad. Sci.* 967, 1–6.
- Burr, B., Burr, F.A., Thompson, K.H., Albertson, M.C., and Stuber, C.W. (1988). Gene mapping with recombinant inbreds in maize. *Genetics* 118, 519–526.
- Cai, L., Friedman, N., and Xie, X.S. (2006). Stochastic protein expression in individual cells at the single molecule level. *Nature* 440, 358–362.
- Capecchi, M.R. (2005). Gene targeting in mice: functional analysis of the mammalian genome for the twenty-first century. *Nat. Rev. Genet.* 6, 507–512.
- Chakravarti, A., Clark, A.G., and Mootha, V.K. (2013). Distilling pathophysiology from complex disease genetics. *Cell* 155, 21–26.
- Chen, S., Sanjana, N.E., Zheng, K., Shalem, O., Lee, K., Shi, X., Scott, D.A., Song, J., Pan, J.Q., Weissleder, R., et al. (2015). Genome-wide CRISPR screen in a mouse model of tumor growth and metastasis. *Cell* 160, 1246–1260.
- Chesler, E.J., Lu, L., Wang, J., Williams, R.W., and Manly, K.F. (2004). WebQTL: rapid exploratory analysis of gene expression and genetic networks for brain and behavior. *Nat. Neurosci.* 7, 485–486.
- Churchill, G.A., Airey, D.C., Allayee, H., Angel, J.M., Attie, A.D., Beatty, J., Beavis, W.D., Belknap, J.K., Bennett, B., Berrettini, W., et al.; Complex Trait Consortium (2004). The Collaborative Cross, a community resource for the genetic analysis of complex traits. *Nat. Genet.* 36, 1133–1137.
- Clark, A.G. (2000). Limits to prediction of phenotypes from knowledge of genotypes. *Evol. Biol.* 32, 205–224.
- Cronin, R.M., Field, J.R., Bradford, Y., Shaffer, C.M., Carroll, R.J., Mosley, J.D., Bastarache, L., Edwards, T.L., Hebring, S.J., Lin, S., et al. (2014). Phenome-wide association studies demonstrating pleiotropy of genetic variants within FTO with and without adjustment for body mass index. *Front. Genet.* 5, 250.
- Cubillos, F.A., Billi, E., Zörgö, E., Parts, L., Fargier, P., Omholt, S., Blomberg, A., Warringer, J., Louis, E.J., and Liti, G. (2011). Assessing the complex architecture of polygenic traits in diverged yeast populations. *Mol. Ecol.* 20, 1401–1413.
- Danhauser, K., Sauer, S.W., Haack, T.B., Wieland, T., Stauffer, C., Graf, E., Zschocke, J., Strom, T.M., Traub, T., Okun, J.G., et al. (2012). DHTKD1 mutations cause 2-aminoacidic and 2-oxoadipic aciduria. *Am. J. Hum. Genet.* 91, 1082–1087.
- Darvasi, A., and Soller, M. (1995). Advanced intercross lines, an experimental population for fine genetic mapping. *Genetics* 141, 1199–1207.
- Datta, K., Baisakh, N., Ganguly, M., Krishnan, S., Yamaguchi Shinozaki, K., and Datta, S.K. (2012). Overexpression of *Arabidopsis* and rice stress genes' inducible transcription factor confers drought and salinity tolerance to rice. *Plant Biotechnol. J.* 10, 579–586.
- De Haan, G., and Van Zant, G. (1999). Genetic analysis of hematopoietic cell cycling in mice suggests its involvement in organismal life span. *FASEB J* 13, 707–713.
- Deeb, S.S., Fajas, L., Nemoto, M., Pihlajamäki, J., Mykkänen, L., Kuusisto, J., Laakso, M., Fujimoto, W., and Auwerx, J. (1998). A Pro12Ala substitution in PPAR γ 2 associated with decreased receptor activity, lower body mass index and improved insulin sensitivity. *Nat. Genet.* 20, 284–287.
- Denny, J.C., Bastarache, L., Ritchie, M.D., Carroll, R.J., Zink, R., Mosley, J.D., Field, J.R., Pulley, J.M., Ramirez, A.H., Bowton, E., et al. (2013). Systematic comparison of phenome-wide association study of electronic medical record data and genome-wide association study data. *Nat. Biotechnol.* 31, 1102–1110.
- Elmer, G.I., Pieper, J.O., Hamilton, L.R., and Wise, R.A. (2010). Qualitative differences between C57BL/6J and DBA/2J mice in morphine potentiation of brain stimulation reward and intravenous self-administration. *Psychopharmacology (Berl.)* 208, 309–321.
- Fenn, J.B., Mann, M., Meng, C.K., Wong, S.F., and Whitehouse, C.M. (1989). Electrospray ionization for mass spectrometry of large biomolecules. *Science* 246, 64–71.
- Flint, J., and Mackay, T.F. (2009). Genetic architecture of quantitative traits in mice, flies, and humans. *Genome Res.* 19, 723–733.
- Flint, J., Valdar, W., Shifman, S., and Mott, R. (2005). Strategies for mapping and cloning quantitative trait genes in rodents. *Nat. Rev. Genet.* 6, 271–286.
- Fuhrer, T., Heer, D., Begemann, B., and Zamboni, N. (2011). High-throughput, accurate mass metabolome profiling of cellular extracts by flow injection-time-of-flight mass spectrometry. *Anal. Chem.* 83, 7074–7080.
- Gelman, R., Watson, A., Bronson, R., and Yunis, E. (1988). Murine chromosomal regions correlated with longevity. *Genetics* 118, 693–704.
- Goldstein, D.B. (2009). Common genetic variation and human traits. *N. Engl. J. Med.* 360, 1696–1698.
- Harris, T.W., Antoshechkin, I., Bieri, T., Blasiar, D., Chan, J., Chen, W.J., De La Cruz, N., Davis, P., Duesbury, M., Fang, R., et al. (2010). WormBase: a comprehensive resource for nematode research. *Nucleic Acids Res.* 38, D463–D467.
- Horvath, S. (2013). DNA methylation age of human tissues and cell types. *Genome Biol.* 14, R115.
- Houtkooper, R.H., Mouchiroud, L., Ryu, D., Moullan, N., Katsyuba, E., Knott, G., Williams, R.W., and Auwerx, J. (2013). Mitonuclear protein imbalance as a conserved longevity mechanism. *Nature* 497, 451–457.
- Jiang, X.C., Agellon, L.B., Walsh, A., Breslow, J.L., and Tall, A. (1992). Dietary cholesterol increases transcription of the human cholesteryl ester transfer protein gene in transgenic mice. Dependence on natural flanking sequences. *J. Clin. Invest.* 90, 1290–1295.

- Jovaisaite, V., Mouchiroud, L., and Auwerx, J. (2014). The mitochondrial unfolded protein response, a conserved stress response pathway with implications in health and disease. *J. Exp. Biol.* 217, 137–143.
- Kamath, R.S., Fraser, A.G., Dong, Y., Poulin, G., Durbin, R., Gotta, M., Kanapin, A., Le Bot, N., Moreno, S., Sohrmann, M., et al. (2003). Systematic functional analysis of the *Caenorhabditis elegans* genome using RNAi. *Nature* 421, 231–237.
- Kanehisa, M., Goto, S., Sato, Y., Kawashima, M., Furumichi, M., and Tanabe, M. (2014). Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic Acids Res.* 42, D199–D205.
- Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M., and Haussler, D. (2002). The human genome browser at UCSC. *Genome Res.* 12, 996–1006.
- Khan, D., Chan, A., Millar, J.L., Girard, I.J., and Belmonte, M.F. (2014). Predicting transcriptional circuitry underlying seed coat development. *Plant Sci* 223, 146–152.
- King, E.G., Macdonald, S.J., and Long, A.D. (2012). Properties and power of the *Drosophila* Synthetic Population Resource for the routine dissection of complex traits. *Genetics* 191, 935–949.
- Kingsmore, S.F., Lindquist, I.E., Mudge, J., and Beavis, W.D. (2007). Genome-wide association studies: progress in identifying genetic biomarkers in common, complex diseases. *Biomark. Insights* 2, 283–292.
- Koutnikova, H., Laakso, M., Lu, L., Combe, R., Paananen, J., Kuulasmaa, T., Kuusisto, J., Häring, H.U., Hansen, T., Pedersen, O., et al. (2009). Identification of the UBP1 locus as a critical blood pressure determinant using a combination of mouse and human genetics. *PLoS Genet.* 5, e1000591.
- Kover, P.X., Valdar, W., Trakalo, J., Scarcelli, N., Ehrenreich, I.M., Purugganan, M.D., Durrant, C., and Mott, R. (2009). A Multiparent Advanced Generation Inter-Cross to fine-map quantitative traits in *Arabidopsis thaliana*. *PLoS Genet.* 5, e1000551.
- Lander, E.S. (2011). Initial impact of the sequencing of the human genome. *Nature* 470, 187–197.
- Ledford, H. (2012). Drug bests cystic-fibrosis mutation. *Nature* 482, 145.
- Lee, P.N., Forey, B.A., and Coombs, K.J. (2012). Systematic review with meta-analysis of the epidemiological evidence in the 1900s relating smoking to lung cancer. *BMC Cancer* 12, 385.
- Li, T.K., and Lumeng, L. (1984). Alcohol preference and voluntary alcohol intakes of inbred rat strains and the National Institutes of Health heterogeneous stock of rats. *Alcohol. Clin. Exp. Res.* 8, 485–486.
- Liu, S.C., Kowalski, S.P., Lan, T.H., Feldmann, K.A., and Paterson, A.H. (1996). Genome-wide high-resolution mapping by recurrent intermating using *Arabidopsis thaliana* as a model. *Genetics* 142, 247–258.
- MacArthur, D.G., Balasubramanian, S., Frankish, A., Huang, N., Morris, J., Walter, K., Jostins, L., Habegger, L., Pickrell, J.K., Montgomery, S.B., et al.; 1000 Genomes Project Consortium (2012). A systematic survey of loss-of-function variants in human protein-coding genes. *Science* 335, 823–828.
- Mackay, T.F., Richards, S., Stone, E.A., Barbadilla, A., Ayroles, J.F., Zhu, D., Casillas, S., Han, Y., Magwire, M.M., Cridland, J.M., et al. (2012). The *Drosophila melanogaster* Genetic Reference Panel. *Nature* 482, 173–178.
- Marx, V. (2013). Biology: The big challenges of big data. *Nature* 498, 255–260.
- McCarthy, M.I., Abecasis, G.R., Cardon, L.R., Goldstein, D.B., Little, J., Ioannidis, J.P., and Hirschhorn, J.N. (2008). Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nat. Rev. Genet.* 9, 356–369.
- Nadeau, J.H., Singer, J.B., Matin, A., and Lander, E.S. (2000). Analysing complex genetic traits with chromosome substitution strains. *Nat. Genet.* 24, 221–225.
- Nuzhdin, S.V., Pasyukova, E.G., Dilda, C.L., Zeng, Z.B., and Mackay, T.F. (1997). Sex-specific quantitative trait loci affecting longevity in *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. USA* 94, 9734–9739.
- Ossowski, S., Schneeberger, K., Clark, R.M., Lanz, C., Warthmann, N., and Weigel, D. (2008). Sequencing of natural strains of *Arabidopsis thaliana* with short reads. *Genome Res.* 18, 2024–2033.
- Park, S.J., Jiang, K., Tal, L., Yichie, Y., Gar, O., Zamir, D., Eshed, Y., and Lippman, Z.B. (2014). Optimization of crop productivity in tomato using induced mutations in the florigen pathway. *Nat. Genet.* 46, 1337–1342.
- Peirce, J.L., Lu, L., Gu, J., Silver, L.M., and Williams, R.W. (2004). A new set of BXD recombinant inbred lines from advanced intercross populations in mice. *BMC Genet.* 5, 7.
- Pospisilik, J.A., Schramek, D., Schnidar, H., Cronin, S.J., Nehme, N.T., Zhang, X., Knauf, C., Cani, P.D., Aumayr, K., Todoric, J., et al. (2010). *Drosophila* genome-wide obesity screen reveals hedgehog as a determinant of brown versus white adipose cell fate. *Cell* 140, 148–160.
- Pravenec, M., Klír, P., Kren, V., Zicha, J., and Kunes, J. (1989). An analysis of spontaneous hypertension in spontaneously hypertensive rats by means of new recombinant inbred strains. *J. Hypertens.* 7, 217–221.
- Prayle, A.P., Hurley, M.N., and Smyth, A.R. (2012). Compliance with mandatory reporting of clinical trial results on ClinicalTrials.gov: cross sectional study. *BMJ* 344, d7373.
- Rockman, M.V., and Kruglyak, L. (2009). Recombinational landscape and population genomics of *Caenorhabditis elegans*. *PLoS Genet.* 5, e1000419.
- Röst, H.L., Rosenberger, G., Navarro, P., Gillet, L., Miladinović, S.M., Schubert, O.T., Wolski, W., Collins, B.C., Malmström, J., Malmström, L., and Aebersold, R. (2014). OpenSWATH enables automated, targeted analysis of data-independent acquisition MS data. *Nat. Biotechnol.* 32, 219–223.
- Rual, J.F., Ceron, J., Koreth, J., Hao, T., Nicot, A.S., Hirozane-Kishikawa, T., Vandenhaute, J., Orkin, S.H., Hill, D.E., van den Heuvel, S., and Vidal, M. (2004). Toward improving *Caenorhabditis elegans* phenome mapping with an ORFeome-based RNAi library. *Genome Res.* 14 (10B), 2162–2168.
- Ryder, E., Ashburner, M., Bautista-Llacer, R., Drummond, J., Webster, J., Johnson, G., Morley, T., Chan, Y.S., Blows, F., Coulson, D., et al. (2007). The DrosDel deletion collection: a *Drosophila* genomewide chromosomal deficiency resource. *Genetics* 177, 615–629.
- Sander, J.D., and Joung, J.K. (2014). CRISPR-Cas systems for editing, regulating and targeting genomes. *Nat. Biotechnol.* 32, 347–355.
- Schena, M., Shalon, D., Davis, R.W., and Brown, P.O. (1995). Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* 270, 467–470.
- Schilling, R.K., Marschner, P., Shavrukov, Y., Berger, B., Tester, M., Roy, S.J., and Plett, D.C. (2014). Expression of the *Arabidopsis vacuolar H⁺-pyrophosphatase* gene (AVP1) improves the shoot biomass of transgenic barley and increases grain yield in a saline field. *Plant Biotechnol. J.* 12, 378–386.
- Shin, S.Y., Fauman, E.B., Petersen, A.K., Krumsiek, J., Santos, R., Huang, J., Arnold, M., Erte, I., Forgetta, V., Yang, T.P., et al.; Multiple Tissue Human Expression Resource (MuTHER) Consortium (2014). An atlas of genetic influences on human blood metabolites. *Nat. Genet.* 46, 543–550.
- Sivasundar, A., and Hey, J. (2005). Sampling from natural populations with RNAi reveals high outcrossing and population structure in *Caenorhabditis elegans*. *Curr. Biol.* 15, 1598–1602.
- Skarnes, W.C., Rosen, B., West, A.P., Koutsourakis, M., Bushell, W., Iyer, V., Mujica, A.O., Thomas, M., Harrow, J., Cox, T., et al. (2011). A conditional knockout resource for the genome-wide study of mouse gene function. *Nature* 474, 337–342.
- Smith, L., and Hood, L. (1987). Mapping and sequencing the human genome: How to proceed. *Nat. Biotechnol.* 5, 933–939.
- Soussi, T., Legros, Y., Lubin, R., Ory, K., and Schlichtholz, B. (1994). Multifactorial analysis of p53 alteration in human cancer: a review. *Int. J. Cancer* 57, 1–9.
- Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., and Mesirov, J.P. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. USA* 102, 15545–15550.

- Taylor, B.A., Heiniger, H.J., and Meier, H. (1973). Genetic analysis of resistance to cadmium-induced testicular damage in mice. *Proc. Soc. Exp. Biol. Med.* *143*, 629–633.
- Valdar, W., Solberg, L.C., Gauguier, D., Burnett, S., Klenerman, P., Cookson, W.O., Taylor, M.S., Rawlins, J.N., Mott, R., and Flint, J. (2006). Genome-wide genetic association of complex traits in heterogeneous stock mice. *Nat. Genet.* *38*, 879–887.
- Wang, J., Williams, R.W., and Manly, K.F. (2003). WebQTL: web-based complex trait analysis. *Neuroinformatics* *1*, 299–308.
- Wasinger, V.C., Cordwell, S.J., Cerpa-Poljak, A., Yan, J.X., Gooley, A.A., Wilkins, M.R., Duncan, M.W., Harris, R., Williams, K.L., and Humphrey-Smith, I. (1995). Progress with gene-product mapping of the Mollicutes: *Mycoplasma genitalium*. *Electrophoresis* *16*, 1090–1094.
- Williams, R.W., Gu, J., Qi, S., and Lu, L. (2001). The genetic structure of recombinant inbred mice: high-resolution consensus maps for complex trait analysis. *Genome Biol* *2*, RESEARCH0046.
- Williams, R.W., Bennett, B., Lu, L., Gu, J., DeFries, J.C., Carosone-Link, P.J., Rikke, B.A., Belknap, J.K., and Johnson, T.E. (2004). Genetic structure of the LXS panel of recombinant inbred mouse strains: a powerful resource for complex trait analysis. *Mamm. Genome* *15*, 637–647.
- Winzeler, E.A., Shoemaker, D.D., Astromoff, A., Liang, H., Anderson, K., Andre, B., Bangham, R., Benito, R., Boeke, J.D., Bussey, H., et al. (1999). Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis. *Science* *285*, 901–906.
- Wu, Y., Williams, E.G., Dubuis, S., Mottis, A., Jovaisaite, V., Houten, S.M., Argmann, C.A., Faridi, P., Wolski, W., Kutalik, Z., et al. (2014). Multilayered genetics and omics dissection of mitochondrial activity in a mouse reference population. *Cell* *158*, 1415–1430.
- Xu, X.M., and Møller, S.G. (2011). The value of Arabidopsis research in understanding human disease states. *Curr. Opin. Biotechnol.* *22*, 300–307.
- Xu, J., Zhao, Q., Du, P., Xu, C., Wang, B., Feng, Q., Liu, Q., Tang, S., Gu, M., Han, B., and Liang, G. (2010). Developing high throughput genotyped chromosome segment substitution lines based on population whole-genome resequencing in rice (*Oryza sativa* L.). *BMC Genomics* *11*, 656.
- Yamamoto, S., Jaiswal, M., Chang, W.L., Gambin, T., Karaca, E., Mirzaa, G., Wiszniewski, W., Sandoval, H., Haelterman, N.A., Xiong, B., et al. (2014). A drosophila genetic resource of mutants to study mechanisms underlying human genetic diseases. *Cell* *159*, 200–214.
- Yeo, G.S., Farooqi, I.S., Aminian, S., Halsall, D.J., Stanhope, R.G., and O'Rahilly, S. (1998). A frameshift mutation in MC4R associated with dominantly inherited human obesity. *Nat. Genet.* *20*, 111–112.
- Yuan, J., Njiti, V.N., Meksem, K., Iqbal, M.J., Triwitayakorn, K., Kassem, M.A., Davis, G.T., Schmidt, M.E., and Lightfoot, D.A. (2002). Quantitative trait loci in Two Soybean Recombinant Inbred Line Populations Segregating for Yield and Disease Resistance. *Crop Sci.* *42*, 271–277.
- Zamir, D. (2014). Botany. A wake-up call with coffee. *Science* *345*, 1124.
- Zhao, L., Liu, F., Xu, W., Di, C., Zhou, S., Xue, Y., Yu, J., and Su, Z. (2009). Increased expression of OsSPX1 enhances cold/subfreezing tolerance in tobacco and *Arabidopsis thaliana*. *Plant Biotechnol. J.* *7*, 550–561.