Systems genetics approaches to understand complex traits

Mete Civelek[1–3] *and Aldons J. Lusis*[1–3]

Abstract | Systems genetics is an approach to understand the flow of biological information that underlies complex traits. It uses a range of experimental and statistical methods to quantitate and integrate intermediate phenotypes, such as transcript, protein or metabolite levels, in populations that vary for traits of interest. Systems genetics studies have provided the first global view of the molecular architecture of complex traits and are useful for the identification of genes, pathways and networks that underlie common human diseases. Given the urgent need to understand how the thousands of loci that have been identified in genome-wide association studies contribute to disease susceptibility, systems genetics is likely to become an increasingly important approach to understanding both biology and disease.

Systems genetics
A global analysis of the molecular factors that underlie variability in physiological or clinical phenotypes across individuals in a population. It considers not only the underlying genetic variation but also intermediate phenotypes such as gene expression, protein levels and metabolite levels, in addition to gene-by-gene and gene-by-environment interactions.

[1]*Department of Microbiology, Immunology, and Molecular Genetics, University of California, Los Angeles.*
[2]*Department of Human Genetics, University of California, Los Angeles.*
[3]*Department of Medicine, A2-237 Center for Health Sciences, University of California, Los Angeles, California 90095-1679, USA.*
Correspondence to A.J.L.
e-mail: jlusis@mednet.ucla.edu
doi:10.1038/nrg3575
Published online
3 December 2013

Genome-wide association studies (GWASs) have identified thousands of genetic loci that contribute to common diseases in humans. However, aside from their typically modest value for predicting future disease occurrence, this information will provide little mechanistic insight until the loci are translated into genes and pathways. Beyond that, it will be important to understand how the alleles interact with each other or with environmental factors. Although the genes at each locus can be individually tested either in cell cultures or in animal models and their mechanistic effect on the clinical trait defined using classical molecular biology approaches[1], this strategy will clearly be challenging. The vast majority of loci that have been identified for common diseases show modest effects that may be hard to reproduce experimentally. It is clear from studies using experimental organisms that natural phenotypic variation usually results from many interacting alleles that show context-dependent and environmentally sensitive effects, and there is a reason to believe that common diseases will be similarly complex[2–6].

An alternative, or complementary, method to studying one locus at a time is to carry out global analyses of biological molecules in populations that show inter-individual variability for the clinical traits. Recent technological developments have made it possible to quantitatively survey hundreds or thousands of biological molecules, from DNA sequence variations to epigenetic marks to levels of transcripts, proteins and metabolites (FIG. 1). For example, it is reasonably straightforward to globally quantify transcript levels in tissues, assuming that the relevant tissues are available, using hybridization or sequencing technologies. The transcript levels can then be either tested for correlation with the clinical trait or mapped to chromosomal loci to identify functional variants that may contribute to the clinical trait. The transcript levels can, in a sense, be considered intermediate phenotypes, as DNA variation contributes to the clinical trait by perturbing gene expression, proteins and metabolites — of course, such molecular traits are dynamic and can also be reactive to the phenotype. The advantage of this systems genetics approach is that it allows an analysis of molecular interactions in a context that is the most relevant to the clinical trait, namely, multiple genetic perturbations (as in a natural population) rather than an individual genetic perturbation (as in a transgenic mouse). This point is central to a 'systems genetics perspective': inferences about biological phenomena are rarely separable from the genetic system in which they are embedded; thus, to generalize results across genetic backgrounds, experiments must be carried out across multiple genetic backgrounds (W. Valdar, personal communication). Besides common diseases, systems genetics provides a useful window into the general architecture of complex traits and into the flow of biological information. Indeed, systems genetics studies in the past decade have addressed some classic questions about the underlying molecular genetic architecture of complex traits: How common is functional variation in natural populations? How does information flow from DNA to phenotype? And what is the nature of gene-by-environment (G×E) interactions? Systems genetics is, of course, limited by the extent of natural genetic variation and, currently,

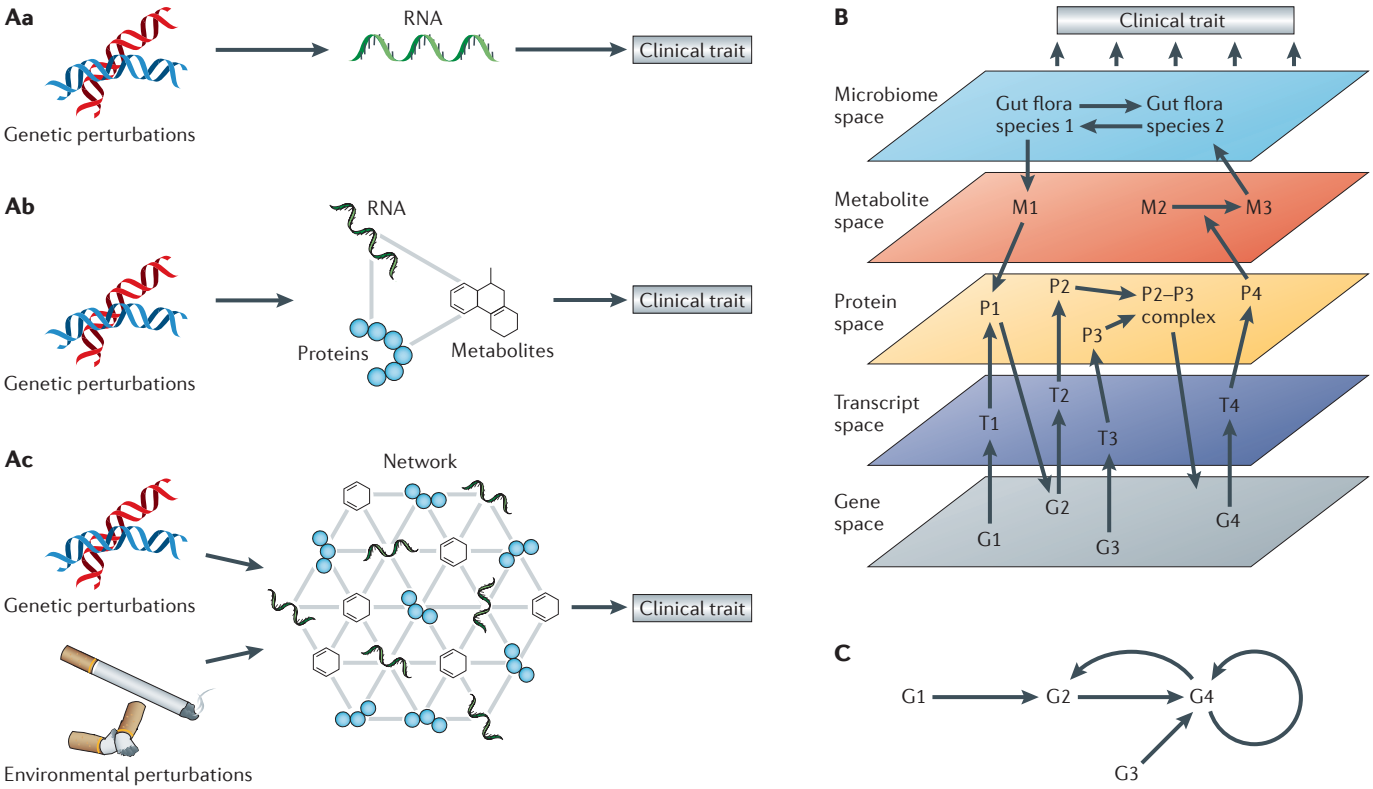


Figure 1 | **Systems genetics strategies.** The left panel shows various designs of systems genetics studies. **Aa** | In the simplest scenario, an intermediate phenotype, such as transcript levels, is quantitated in a population and integrated with a clinical trait on the basis of correlation and mapping. **Ab** | In the second scenario, multiple intermediate phenotypes are studied, which allows interactions across biological scales to be examined. **Ac** | In the third scenario, data across multiple scales are used to model a biological network. **B** | Interactions (shown as arrows) of molecular phenotypes across multiple biological scales — including genes (G), transcripts (T), proteins (P), metabolites (M) and microbiome — can be used to create a map on the basis of natural variation. **C** | Based on correlations of the traits that occur across individuals in a population, one can model a biological network. For example, based on natural variations of genes 1–4 (G1–4), a directional expression network can be modelled. Part **A** is modified, with permission, from REF. 123

Natural populations
Human populations, or animal populations in wild environments, that are experiencing normal selective pressures. By contrast, laboratory animal populations, such as inbred strains, can show natural genetic variation, but they have been subjected to nonrandom breeding and artificial selection.

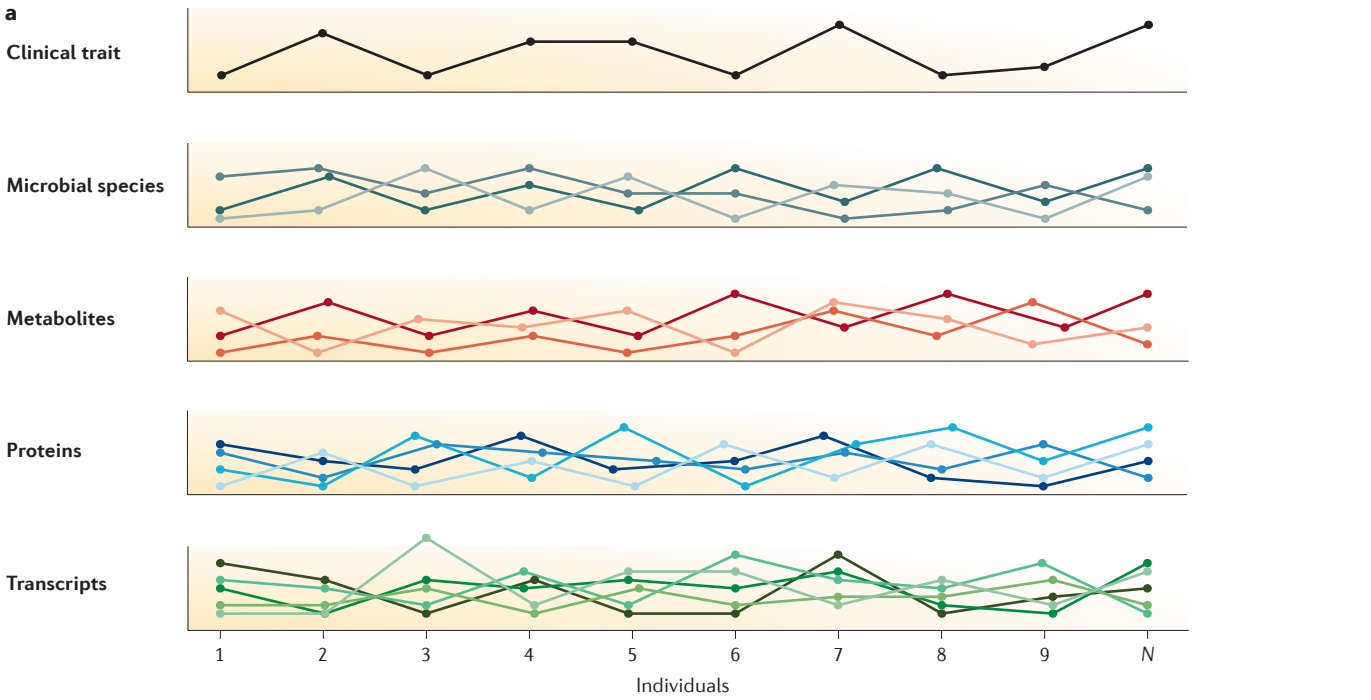
Natural genetic variation
Genetic variation that is present in all populations as a result of mutations that occur in the germline; the frequencies of such mutations in populations are affected by selection and by random drift. This is in contrast with experimental variation that is introduced by techniques such as gene targeting and chemical mutagenesis.

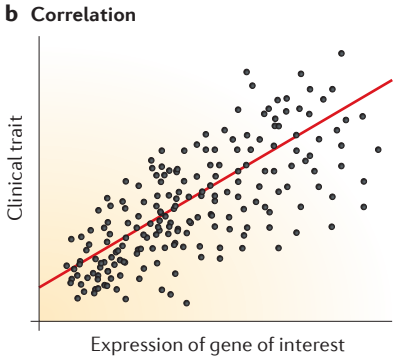
it is mostly hypothesis generating. However, it can be combined with complementary experiments and data sets, such as transgenic mice, global small interfering RNA (siRNA) knockdown and predictive modelling, as discussed below, to provide detailed genotype–phenotype maps that are similar to those envisioned by Waddington and others[7,8] (FIG. 1). In this Review, we describe systems genetics approaches and discuss how they have provided insights into both the molecular underpinnings of complex traits and the understanding of common, complex diseases.

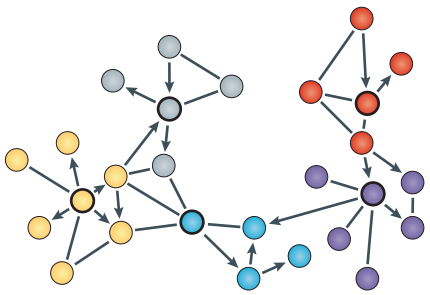
Overview of systems genetics studies

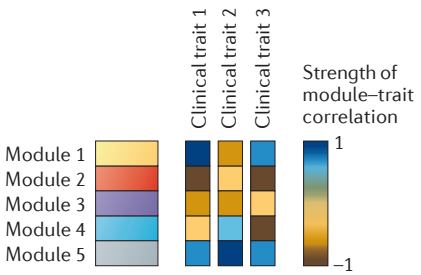
Systems genetics shares with systems biology a holistic, global perspective. The typical strategy in systems biology is to perturb a system, monitor the responses, integrate the data and formulate mathematical models that describe the system. Systems genetics is a particular type of systems biology, in which genetic variation within a population is used to perturb the system. Ultimately, the goal of systems genetics is to understand the broad molecular underpinnings, such as genetic architecture and intermediate physiological phenotypes, of complex traits, including diseases.

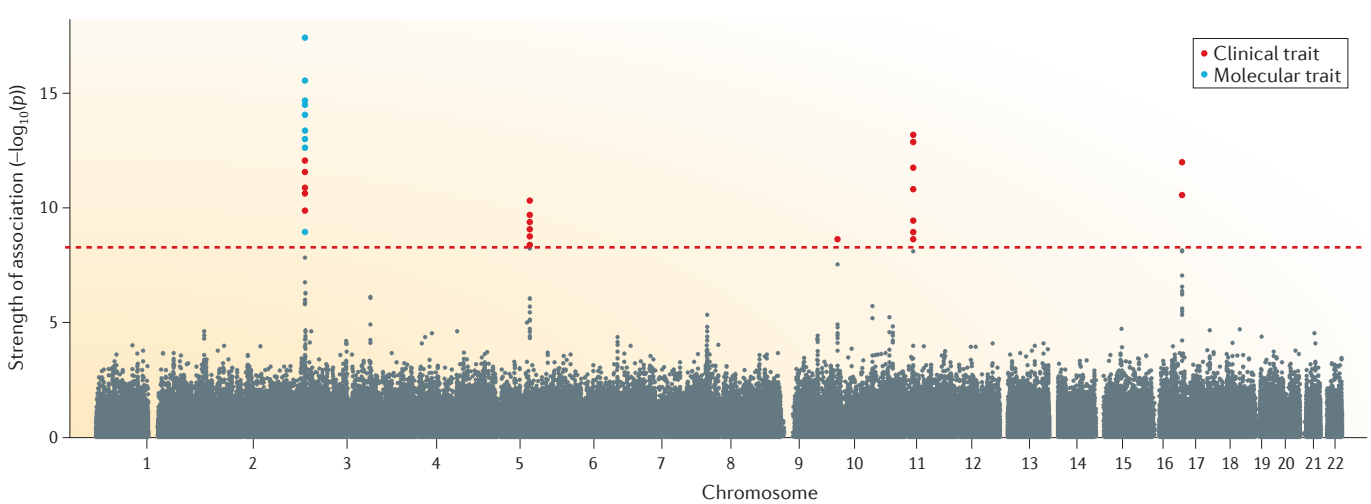
In a hypothetical systems genetics study, numerous individuals are interrogated for a clinical trait, for transcript, protein and metabolite levels in a relevant tissue, and for microbiota composition in the gut (FIG. 2a). The variations in molecular phenotypes can be related to each other and to clinical traits in three ways. First, a simple correlation provides evidence of a possible relationship between two traits[9–11] (FIG. 2b). In this example, one might postulate three probable explanations[12]: the molecular trait influences the clinical trait (that is, the molecular trait is causal for the clinical trait); the clinical trait influences the molecular trait (that is, the molecular trait is reactive to the clinical trait); or both are affected by a confounding factor (that is, the molecular and the clinical traits are independent). Second, genetic mapping can provide evidence of a relationship through co-mapping (FIG. 2c). Thus, if a molecular trait and a clinical trait map to the same genomic region, then one may be the cause of the other. Third, statistical modelling approaches can be used to integrate the











◀ Figure 2 | **Collection and analysis of systems genetics data.** An overview of the steps of a systems genetics study is shown. **a** | A population of individuals who differ in traits of interest is identified. The population could be either a group of unrelated individuals or a segregating population (that is, a family). These individuals are then examined for clinical traits of interest, and one or more intermediate phenotypes from tissues of interest are quantified using high-throughput technologies. Each intermediate phenotype is shown by different shades of the same colour in the graphs. **b** | The relationships between these traits can be analysed by examining pairwise correlations. A correlation could result from causal, reactive or independent relationships. **c** | Loci that contribute to these traits can be mapped either by association or by linkage. In this example, single-nucleotide polymorphisms (SNPs) that are genotyped using a high-density genotyping microarray are tested for association with the traits using linear regression. The negative logarithm of the *p*-values for each SNP are plotted against the position of the SNP across the genome. Coincident mapping of multiple traits (the peak on the left) indicates the possibility of a causal relationship. The red dashed line represents the *p*-value threshold. **d** | Higher-order interactions among molecular phenotypes can be modelled using both statistical and network-based approaches. This example shows part of a genetic interaction network, in which highly correlated transcripts are clustered to form modules of co-regulated genes. Relationships among genes can be either directional (arrows) or non-directional (lines). Here, genes are grouped into modules that are denoted by different colours, and the outlined circles represent the hub genes, which have the most connections in their respective modules. **e** | The relationship between such modules and clinical traits can then be examined by correlating either the average gene expression levels or principal components in a module with the trait.

data. For example, various network approaches, such as co-expression, can identify groups of molecular traits that share characteristics (FIG. 2d). These groups, which are termed modules, can then be tested for relationships to the clinical trait (FIG. 2e). Examples of each of these operations are described below.

In designing a systems genetics study, there are various considerations. First, it is important to examine a sufficient number of individuals such that there is adequate power for genetic mapping and for other analyses. Mapping resolution is also an important issue because the ultimate goal is to relate specific genetic elements to changes in molecular and clinical phenotypes. If the locus is large and contains many genes, as is the case for quantitative trait locus (QTL) analyses in mice and rats, then it is difficult to identify the underlying causal genetic variation. Additionally, the structure of the population may have considerable effects on the architecture of complex traits. For example, a study of a collection of inbred strains of *Drosophila melanogaster* yielded a different set of QTLs for certain quantitative traits compared with a study of an outbred population that was derived from the same strains, although there was a significant overlap of the networks that underlie those traits[4]. Systems genetics studies are becoming increasingly powerful as additional phenotypes, multiple biological scales, environmental conditions and changes over time are examined (FIG. 1). It is not possible to collect such extensive data on unique individuals, and renewable reference populations — such as collections of haploid yeast segregants and inbred strains of worms, flies, mice or rats — are therefore proving particularly useful (TABLE 1). Some novel experimental designs, such as 'bulk segregant analysis' that involves examining individuals with extreme phenotypes from large populations of phenotyped individuals, have been used to address specific questions[13].

Principal components
Dominant patterns in multivariate data, as extracted by the principal component analysis data reduction method.

Modules
In the context of network modelling, groups of components that are tightly connected or correlated across a set of conditions, perturbations or genetic backgrounds.

Inbred strains
Strains in which a set of naturally occurring genetic variations have been fixed by many generations of inbreeding.

Biological scales
Various levels in the flow of information from DNA to proteins to metabolites to cell structures to cell interactions.

The flow of biological information

Gene expression. The analysis of transcript levels is fairly straightforward and is routinely done using either microarray-based or high-throughput RNA sequencing (RNA-seq) methods. Transcript levels, similarly to other quantitative traits (such as height), can be mapped to genomic loci that underlie the variation using either linkage analyses in segregating populations or association analyses in population-wide surveys. Although mapping of expression levels of individual genes has been carried out since the 1980s, the first genome-wide mapping of expression levels that used microarrays to measure transcript levels was carried out in 2002 in a linkage analysis of a cross between two strains of yeast[14]. The results showed widespread regulation of hundreds of transcripts. Subsequent studies in bacteria, yeast, plants, worms, fish, flies and various cell types and tissues of mice, rats and humans increased the understanding of the genetic regulation of gene expression and have shown that genetic variations that affect gene expression are widespread in natural populations[14–18].

The genomic regions that are associated with transcript levels are termed expression quantitative trait loci (eQTLs) (FIG. 3a). When an eQTL is near the location of the gene that encodes the transcript (commonly within ≤1 Mb), it is termed a local eQTL. Such proximity indicates that the variation is likely to act on gene expression in *cis* (that is, only in the chromosome copy in which it resides), and local eQTLs are therefore often referred to as *cis*-eQTLs. Of course, it is possible that some local eQTLs act in *trans* — for example, a gene could show feedback and could regulate the gene on the homologous chromosome as well — and it is thus more accurate to use the term local. When a locus that affects the expression level of a transcript maps distally to the gene that encodes the transcript (such as to a gene on another chromosome), it is termed a distal eQTL or a *trans*-eQTL. In contrast to *cis*-acting loci, *trans*-acting loci would be expected to affect both alleles of the target gene equally. In some cases loci have been found to affect the expression of hundreds of genes, and such eQTLs are termed *trans* bands or hot spots[19] (FIG. 3b).

Studies using mice, rats, and human cells and tissues have revealed that the expression of a high percentage of genes (≥30%) is substantially influenced by eQTLs[18,20,21]. Most peak single-nucleotide polymorphisms (SNPs) for GWASs map outside protein-coding regions, and >75% of GWAS SNPs map to functional regulatory elements that have been identified in the Encyclopedia of DNA Elements (ENCODE) project[22]. These results suggest that genetic variants that alter gene expression, rather than variants that alter protein sequences, form the primary basis of natural variation in complex traits. This is in contrast to Mendelian disorders, for which variation in protein-coding regions predominates. eQTL studies of non-coding transcripts, such as microRNAs[23] and large intergenic non-coding RNAs[24], have started to emerge. It seems that the degree of genetic variation that underlies levels of these transcripts is less than that observed for protein-coding RNAs.

The examination of gene expression using RNA-seq provides information about both RNA splicing and transcripts that are not detected by commercial gene expression microarrays. Comparisons of RNA-seq results with expression array results indicate fairly high concordance for transcript abundance[25,26]. Next-generation sequencing has also made it possible to carry out global *cis*-eQTL analyses using allele-specific expression (ASE) analyses, which complement eQTL analyses[27]. ASE analyses identify sequence differences in the transcripts that are derived from the two chromosome copies in a diploid organism and use this information to quantitate these transcripts separately in order to determine whether they are expressed at different levels. A disadvantage of ASE analyses compared with eQTL analyses is that ASE analyses are restricted to the identification of *cis*-eQTLs (as *trans*-regulated genes would not be expected to show ASE) and to genes for which the alleles give rise to transcripts with sequence differences. An obvious advantage of ASE analyses is the requirement for a much smaller number of individuals to identify a comparable number of *cis*-acting loci[28]. Currently, only a few ASE studies have been carried out. In two separate studies using crosses between strains of mice, the overlaps of *cis*-eQTLs that were identified by linkage and by ASE were modest — 40% and 60% overlap was found, respectively, when technical artefacts were eliminated[27,29]. The explanation for the discrepancies is unclear.

Table 1 | **Systems genetics reference populations**

Species	Description
Saccharomyces cerevisiae	• A cross between a laboratory strain and a wine strain of *S. cerevisiae* has been used to investigate interactions among multiple cellular and molecular traits[15,57,62,124] • The haploid segregants from the cross constitute a permanent resource that can be grown in large quantities and studied under a range of environmental conditions
Arabidopsis thaliana	A set of 191 recombinant inbred strains from *A. thaliana* were phenotyped for many quantitative traits, gene expression and metabolites[125]
Caenorhabditis elegans	• A cross between Bristol and Hawaiian isolates of *C. elegans* has been used to generate >200 recombinant inbred advanced intercross lines • These lines have been used to investigate epistatic loci that regulate complex phenotypes[126] and to carry out eQTL studies[127]
Drosophila melanogaster from the Genome Reference Panel	Extensive phenotyping, including physiology, disease resistance, gene expression, behaviour and morphology, was carried out[128]
D. melanogaster from the *Drosophila* Synthetic Population Resource	• A panel of >1,700 recombinant inbred lines were derived from two highly recombined synthetic populations, each created by intercrossing a different set of eight inbred founder lines (with one founder line that was common to both populations) • The strategy that was used to create these recombinant inbred strains is similar to that of the mouse Collaborative Cross[129]
Mice from the Collaborative Cross	• A large set of highly diverse recombinant inbred strains that are derived from an eight-way cross is currently under construction[130] • A unique aspect is the extensive genetic diversity (for example, ~17 million SNPs), as the cross includes several 'wild' parental strains • A Diversity Outbred Population with the same founders has been constructed and is available from The Jackson Laboratory[131]
Mice from the Hybrid Mouse Diversity Panel	• A collection of ~100 commercially available inbred strains of mice that consists of ~30 classic inbred strains and 70 recombinant inbred strains[16] • Association mapping with correction for population structure allows fine mapping (~1 Mb resolution) with fair power for complex traits • This panel has been typed for tissue transcript levels, protein levels, and numerous clinical and physiological traits; published data are also available in the Systems Genetics Resource[108]
Mouse and rat CCSs	• The CSSs consist of an inbred genetic background onto which one chromosome at a time has been substituted from a separate strain • The mouse CSS panel consists of the background strain C57BL/6J onto which individual chromosomes from strain A/J have been bred [51]; this panel has been extensively phenotyped for metabolic and other traits and has proved especially useful for studying genetic interactions
HxB/BxH recombinant inbred rat strains	• A set of 30 recombinant inbred strains that were derived from parental Spontaneously Hypertensive and Brown Norway strains have been systematically studied for a variety of metabolic, cardiovascular and behavioural traits[96,132,133] • A comprehensive inventory of genomic and transcriptomic differences has been generated[134]
Human cohort from the Metabolic Syndrome in Men study	• One of the largest single-site population-based prospective cohorts that comprises ~10,000 participants who have been subjected to extensive clinical examinations, including oral glucose tolerance tests with measurements of glucose, insulin, proinsulin and free fatty acids levels at 0, 30 and 120 minutes; body composition analysis through bioelectrical impedance; and measurement of plasma biomarkers such as cytokines, hormones, lipids, lipoprotein subtypes and metabolites using NMR • The population is examined for clinical and molecular traits every five years • Third examination of the participants is currently ongoing[135]

CCSs, chromosome substitution strains; eQTL, expression quantitative trait locus; SNP, single-nucleotide polymorphism.

Recent studies show that a substantial proportion of eQTLs is in open chromatin regions[30], methylated regions[31] or transcription factor-binding sites[32]. Sequence-specific transcription factors form a network that enables integration of multiple internal and external cues to produce a specific epigenetic state and gene expression output. Such networks have been modelled on the basis of diverse cell types[33], and specific interactions have been studied using classic single-gene perturbations. As individuals in a population probably differ at hundreds or thousands of transcription factor-binding sites, systems genetics can provide a useful window into the network interactions. Recently, the genome-wide effects of sequence variation on transcription factor binding and on transcriptional outcomes were examined in primary macrophages of two different inbred strains of mice. Many SNPs that affect the binding of different transcription factors were identified using chromatin immunoprecipitation followed by sequencing (ChIP–seq), and the results provided convincing evidence that lineage-specific transcription factors select enhancer-like regions in a collaborative manner[34]. Similar studies were carried out in humans using lymphoblastoid cell lines (LCLs) and showed extensive effects of genetic variants on epigenetic marks and on transcriptional activation and repression[35–37].

Proteins. As proteins constitute the primary 'machines' of biology, any comprehensive genotype–phenotype maps will require detailed analyses of protein levels and their modifications. It is possible to map protein QTLs (pQTLs; that is, loci that control protein levels)[25,38–42] but only a tiny fraction of all proteins, which are generally the most abundant ones in a sample, can currently be quantitated using high-throughput proteomic approaches, such as mass spectrometry and immunoassays[25,38,41]. If transcript levels were closely correlated with protein levels, they could act as surrogates. However, systems genetics studies in both yeast and mice suggest that transcript levels explain a small proportion of the overall variation in protein levels among individuals in a population[25,38]. For example, the correlation between the levels of >500 proteins and their corresponding transcripts in a population of 100 strains of mice was only 27%[25]. Undoubtedly, this low level of correspondence is partly due to technical issues in proteomic analyses. Heterogeneity is an important issue, as splicing results in many more polypeptide chains than the number of genes, and >200 post-translational modifications have been described in proteins[43]. Given that signalling is mostly mediated through protein modifications, such as acetylation and phosphorylation, methods to quantitate such modifications would be particularly informative for understanding the flow of biological information.

Metabolites. The large-scale analysis of metabolites (that is, metabolomics) using techniques such as mass spectrometry or NMR is a developing field that has already provided important insights into diseases and metabolic processes[44]. For example, metabolite profiling revealed that a phosphatidylcholine metabolite couples diurnal lipid synthesis to energy use in the muscle[45]. Efforts are being made to catalogue the thousands of metabolites that are present in the human body, and examples of such efforts include the Human Metabolome Project and the LIPID Metabolites and Pathways Strategy (LIPID MAPS). In principle, metabolite profiling offers a particularly attractive approach for the integration of genetic and environmental factors that contribute to complex traits. Several systems genetics studies of metabolites in human plasma have been reported and, in these studies, levels of many metabolites showed high heritability. The levels of these metabolites could be mapped to specific loci, and some of these loci co-mapped with GWAS-identified loci for diseases[46–48].

Complexity of interactions

Gene-by-gene interactions. Most studies of quantitative traits in animal models suggest that epistatic (that is, non-additive) interactions between loci are widespread[4,49–51], and various examples of gene-by-gene (G×G) interactions for human complex traits have been identified[6]. The highly varied pathological phenotypes that occur among individuals with specific Mendelian disorder (for example, sickle cell anaemia) represent a form of epistasis. Such variation is also commonly observed in studies using gene-targeted mice; in the extreme case, a null mutation is lethal on one genetic background and has little or no phenotype on another[52]. In natural populations, the varieties of common variations are constrained by selection, such that most combinations of alleles must be compatible with adequate functioning (that is, 'good enough solutions') but are still sufficient for adaptation to changing environments. Common diseases are likely to result from the inheritance of particular combinations of common and rare alleles that are 'poor solutions' given the age-related effects and environmental conditions[53].

At the molecular level, epistasis can take many forms. One common mechanism concerns the dependence of the steady-state levels of a molecule on its rates of production and degradation. For example, in a recent human GWAS among individuals who consume alcohol, the incidence of oesophageal cancer involved a strong genetic interaction between loci that contribute to the production of acetaldehyde (which is a carcinogen) from alcohol and those that contribute to the degradation of acetaldehyde[6] (FIG. 3c). Another example involves the formation of molecular complexes, in which the final levels of the complex are limited by the least abundant component. For example, a systems genetics study of transcript levels and protein levels in mouse liver across a panel of mouse strains observed that transcript–protein correlations were weaker for multisubunit proteins than for homopolymers[25]. In particular, the correlation between the levels of ribosomal proteins and those of their transcripts was essentially zero. This presumably results from the fact that any excess proteins that are not assembled into ribosomes are rapidly degraded. This phenomenon could partly explain the 'phenotypic buffering' that is observed in *Arabidopsis thaliana*, as discussed below.

An advantage of systems genetics for examining the prevalence of G×G interactions is that hundreds

Chromatin immunoprecipitation followed by sequencing (ChIP–seq). A method that is used to analyse protein–DNA interactions by combining chromatin immunoprecipitation with next-generation sequencing to identify binding sites of DNA-associated proteins.

Epistasis A statistical interaction between two or more genetic loci, such that their effects are non-additive.

or thousands of molecular phenotypes (such as transcript levels) can be examined in a single population (FIG. 3b). However, in such global mapping studies, the identification of interactions at multiple loci is difficult because of the low significance threshold that results from multiple comparisons[54]. This has usually been addressed by restricting the analysis to QTLs that have significant main effects, which may miss the vast majority of epistatic interactions. Although methods that overcome the computational burden of identifying G×G interactions have been proposed[55], it remains difficult to determine the importance of G×G interactions in high-dimensional data sets. In addition to the problem of multiple comparisons, G×G interactions are difficult to detect in outbred populations for alleles that have small effect sizes or low frequencies[56].

A recent study using a large cross between two yeast strains provided an estimate of the importance of epistatic interactions for 46 highly heritable complex traits[57]. The authors implemented a high-throughput endpoint colony size assay, which was carried out under a variety of conditions (such as different pH, nutrient types and temperatures) to develop the traits. The identified QTLs explained nearly the entire additive contribution to the heritable variation of the traits. The authors quantitated epistasis from the difference between broad-sense heritability (which is estimated from the reproducibility of trait measures) and narrow-sense heritability (which is estimated from phenotypic similarity for different degrees of relatedness). They observed that the traits showed epistasis that ranged from almost zero to ~50%[57]. The results have implications for the problem of missing heritability, as epistatic interactions would affect heritability estimates[54]. The level of interactions in that study may be somewhat limited compared with that in a mammalian population, as the yeast were haploid and physiological interactions may be more complex in multicellular organisms.

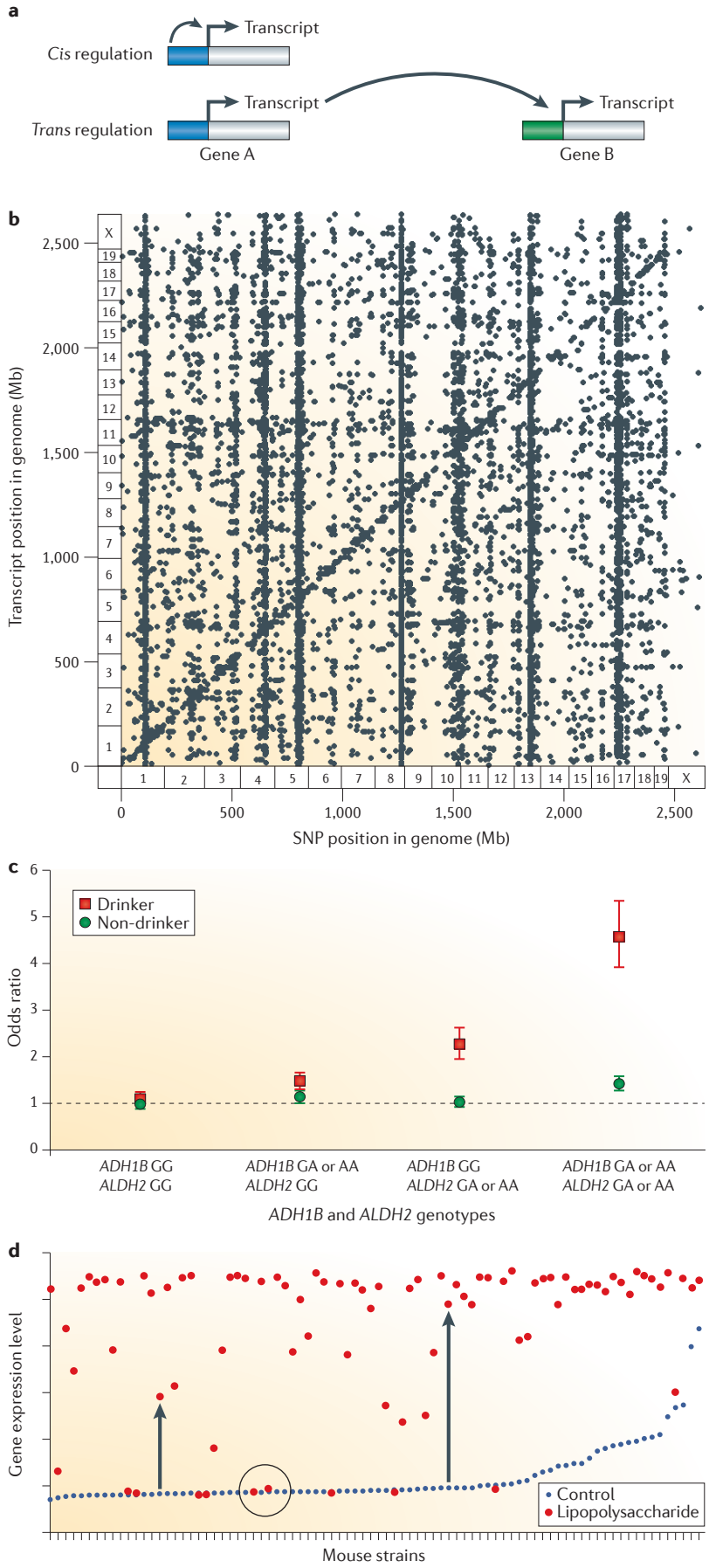
Gene-by-environment interactions. Most complex traits have a substantial environmental component, and G×E interactions seem to be pervasive[49,58]. In fact, almost all common diseases result from a combination of genetic and environmental factors. The frequency and the nature of G×E interactions can be conveniently studied using global gene expression traits. Studies in yeast[59], human LCLs[60] and endothelial cells[21], and mouse macrophages[20] all observed a high frequency of G×E interactions. Environmental changes are much more likely to modulate the effect of a distal eQTL than that of a local eQTL[20]. For example, one study examined gene expression in peritoneal macrophages from 100 strains of mice either in culture medium alone or in culture medium that contained the inflammatory mediator bacterial lipopolysaccharide[20] (FIG. 3d). Of 2,802 significant eQTLs that were detected, 2,607 (93%) showed significant evidence of G×E interactions. A particularly striking example of a G×E interaction in this study was the finding of hot spots that regulated hundreds of genes in macrophages only when stimulated with lipopolysaccharide[20] (FIG. 3b).

Missing heritability
The phenomenon whereby the fraction of the heritability of a trait that is explained by a genome-wide association study is modest.

Figure 3 | **Genetics of gene expression and genetic interactions.** Common genetic variations that affect transcript levels can be examined globally using either gene expression arrays or high-throughput RNA sequencing (RNA-seq). **a** | *Cis* and *trans* effects of gene expression are shown. Genomic loci that regulate the expression of a gene in the same locus are termed *cis*-expression quantitative trait loci (*cis*-eQTLs), whereas loci that regulate the expression of genes that are distant (which are often on another chromosome) are termed *trans*-eQTLs. **b** | A global view of the genetic architecture of gene expression is shown. The *x* and *y* axes show the genomic location of the single-nucleotide polymorphism (SNP) variants and the transcripts, respectively. Each dot shows a significant association. In this example, dots along the diagonal represent *cis*-eQTLs, and the rest show *trans*-eQTLs. There are also several hot spots that regulate hundreds of genes in *trans*, which are shown by dots along the vertical lines. **c** | Gene-by-gene interaction in oesophageal squamous cell carcinoma is shown[6]. The effect size of each allele (A and G) of the *ADH1B* (alcohol dehydrogenase 1B (class I), β-polypeptide) and the *ALDH2* (aldehyde dehydrogenase 2 family (mitochondrial)) genes, as indicated by the odds ratio for the incidence of oesophageal cancer, is not additive and is also influenced by alcohol consumption. **d** | In one example of a gene-by-environment interaction, the expression of a gene is examined in macrophages from 100 strains of mice that are cultured in either the presence or the absence of bacterial endotoxin[20]. Mice with certain genetic backgrounds do not respond to the treatment (some examples of which are circled), whereas others respond to the treatment to different degrees (some examples of which are indicated by arrows). Part **a** is modified, with permission, from REF. 122 © (2011) Macmillan Publishers Ltd. All rights reserved. Part **b** is modified, with permission, from REF. 20 © (2012) Elsevier Science. Part **c** is modified, with permission, from REF. 6 © (2012) Macmillan Publishers Ltd. All rights reserved.

The fact that G×G and G×E interactions seem to be common among complex traits has implications for the mystery of missing heritability in human GWASs because these interactions may lead to inflated estimates of heritability[54]. As human studies are poorly powered to identify such interactions, experimental models will be essential for any comprehensive understanding.

Phenotypic buffering. The integration of data for transcript, protein and metabolite levels using inbred lines of *A. thaliana* resulted in surprising findings[61]. Although eQTLs were associated with the expression levels of several thousand transcripts, these eQTLs corresponded poorly to protein or metabolite levels, which suggests a buffering system for phenotypic variation. A few genomic hot spots were also found to regulate the majority of the molecular phenotypes, which is consistent with the concept of hub genes as inherent components of biological networks. Multiple biological scales have also been examined in several reference populations, including yeast[62] and mice[16,58].



Network modelling

Systems biology approaches that mathematically model higher-order relationships between molecular phenotypes have been developed to understand how they interact with each other and influence complex traits[63–65]. Network approaches, in particular, have proven informative. Networks are graphical representations of the interactions between the molecular traits. Molecules are represented as nodes and the relationships among them as edges. They can be broadly divided into three categories: those that are based on curated knowledge, such as metabolic pathways; those that are derived from experimental data on the basis of physical interactions, such as protein–protein interactions from yeast two-hybrid experiments; and those that are inferred from high-throughput data. Although the curated pathways can capture experimentally supported knowledge, they are typically not comprehensive. To uncover novel relationships and regulatory interactions, data-driven network construction methods use various algorithms. These algorithms can be based on regression models, information theory, linear and nonlinear correlations, entropy maximization, graphical Gaussian modelling, Bayesian networks or combinations of these[66]. The assessment of the insights that are provided by these methods remains challenging because there is no experimental approach to create the 'true' network structure and to compare it with computational predictions. The Dialogue on Reverse Engineering Assessment and Methods (DREAM) project has attempted an objective assessment of reverse engineering methods for biological networks using both simulated and experimental data[66]. They evaluated the performance of >30 network inference method implementations using gene expression microarray data from *Escherichia coli*, *Staphylococcus aureus* and *Saccharomyces cerevisiae* cells. The results showed that no single network method outperformed the others in all data sets. Different network connectivity patterns were predicted by various approaches to various levels of success. For example, although linear cascades of regulation were more accurately predicted by regression and by Bayesian network methods, feed-forward loops — such as those that occur when two mutually dependent transcription factors regulate the expression level of a gene — were more accurately predicted by mutual-information and correlation-based methods. Inference of the eukaryotic regulatory network was less successful than that of the bacterial regulatory network, which suggests the requirement for additional data, such as time course series and transcription factor-binding data, for accurate predictions in complex systems. The conclusion was that a consensus network created on the basis of the inferences from multiple approaches showed the most robust performances across diverse data sets.

The integration of genetic information with network modelling approaches has been used to refine inferences, to highlight pathways that contribute to clinical traits and to identify genes that are likely to be 'key drivers' in biological processes[67,68]. In these implementations,

Box 1 | **Systems genetics in rats and humans for the identification of novel disease genes**

One study[69] identified a *trans*-regulated gene co-expression network that confers risk to type 1 diabetes (T1D) using a cross-species systems genetics approach. First, the authors examined expression quantitative trait loci (eQTLs) from 7 tissues of 30 rat recombinant inbred strains (TABLE 1) and focused on the eQTLs that altered the expression levels of transcription factors[69]. They identified 147 transcription factors with these eQTLs, most of which were regulated in *trans*. They then identified genes that were likely to be controlled by those transcription factors on the basis of co-mapping to the same eQTL, enrichment of the transcription factor-binding sites in their promoters and chromatin immunoprecipitation results. Combining these results with genome-wide co-expression network analysis, they identified a co-expression module with 305 genes that was enriched for inflammatory genes and was partly regulated by the transcription factor interferon regulatory factor 7 (IRF7). Bayesian regression models revealed the rat 15q25 locus as a hot spot that regulates the expression of the members of the inflammatory gene network. Therefore, they postulated that this locus regulates the expression of IRF7 and its target genes. Second, they tested whether a similar network might occur in humans. For this, they used monocytes isolated from humans who were part of the Gutenberg Heart and Cardiogenics Study cohorts. Indeed, they observed evidence of a conserved IRF7-regulated inflammatory network in the monocytes and went on to show that single-nucleotide polymorphisms (SNPs) in the human orthologous locus regulates *IRF7* and the inflammatory network gene expression in *trans*. This network contains interferon-induced with helicase C domain 1 (*IFIH1*), which is a well-characterized T1D susceptibility gene. Knowing the role of macrophages in the immunopathology of T1D, the authors tested for association of T1D susceptibility with SNPs that are located near the inflammatory network genes and observed that these SNPs were more likely to associate with T1D than those that are located near non-network genes. Furthermore, SNPs in the human locus that is orthologous to the rat 15q25 hot spot were associated with T1D risk in a genome-wide association meta-analysis of ~7,500 cases and 9,000 controls. In summary, this study was able to identify a conserved inflammatory network in two species and connect it to T1D in humans.

eQTLs are regarded as 'causal anchors' and are added to the network construction process as prior information. For example, in Bayesian networks, genes that show evidence of *cis* regulation can be modelled as 'parents' of genes that are not *cis*-regulated but not vice versa. Thus, the direction of regulation can be established for certain gene–gene pairs using systems genetics data (FIG. 2d). This integrative approach has recently been used to identify molecular interactions that are disrupted in the brains of patients with Alzheimer's disease[68]. In this study, the authors first constructed co-expression networks using gene expression data from three different brain regions of >500 affected and healthy individuals. Comparison of these networks showed that the connectivity of genes in several modules was reconfigured in affected individuals. The *cis*-eQTL results that were identified using brain gene expression data were used as causal anchors to construct Bayesian networks to predict the key drivers of the differential connectivity that results from the disease. The TYRO protein tyrosine kinase-binding protein gene (*TYROBP*) was identified as a key driver of a gene module that was enriched for genes expressed in the immune system and the microglia; this module was reconfigured in the disease state and was significantly correlated with disease progression. Forced overexpression of TYROBP in microglial cells led to expression changes that validated many of the network predictions. Another systems genetics study also integrated co-expression networks from rats and humans, and identified the mechanism of action of a transcription factor in a locus that is associated with type 1 diabetes[69] (BOX 1).

Causal interactions among molecular and clinical traits can be predicted from systems genetics data using various algorithms[12,70,71]. For example, in a population that is studied for global transcript levels and clinical traits, one can ask, given sufficient data, whether the relationship between levels of a transcript and a clinical trait is causal, reactive or independent. As natural genetic variation in a population is randomly distributed among individuals, these causality algorithms take into account the effect of the multifactorial genetic perturbations on several phenotypic outcomes. The causal predictions are depicted as directed edges on the graphical models that represent the relationships among molecular traits (FIG. 2d). However, these methods have some shortcomings, such as the use of linear models to infer relationships that are not necessarily linear[72,73]. Moreover, networks that are constructed from static systems genetics data cannot predict feedback loops unless time course data are available. Despite these shortcomings, integrative approaches have been used to infer causal relationships among phenotypes[12,74,75].

Additional approaches to data integration

Integration with computational predictions. Incorporating diverse data sets from multiple organisms may substantially add to the predictive power of systems genetics data, although the extent to which inferences in one population can be generalized to another is unclear. Reasonable prediction models of the effects of protein variants, such as missense variants, have been developed, and large-scale protein interaction maps have been constructed. Systematic maps of transcription factor-binding sites and chromatin modifications are being generated. Two useful databases — HaploReg[76] and RegulomeDB[77] — have automated this process using results from the ENCODE project. Early efforts to integrate such modelling data with systems genetics data in yeast have been promising[62,78].

Integration with experimental perturbations. An alternative approach to globally link genes to molecular or even clinical phenotypes is to use high-throughput methods such as genome-wide RNA interference screens in flies and worms, gene deletion collections in bacteria and yeast, and siRNA knockdown

in mammalian tissue culture cells[78]. Such 'systematic genetics' screens differ from systems genetics in that they examine the effects of single gene perturbations on a single genetic background. Although they have the advantage that causality is easier to establish, they do not address the gene–gene interactions that are central to complex traits and are generally not applicable to organismal traits in mammals. Nevertheless, the two approaches should complement each other. The integration of systematic genetics with systems genetics has so far been used only in unicellular organisms[62,79,80], but it should also be feasible in mammals using cultured cells.

Applications for common human diseases

As discussed above, a major goal of systems genetics is to dissect common, complex disease. In particular, eQTL studies have proven useful for prioritizing genes at GWAS loci (FIG. 4) and, in some cases, for identifying potential *trans*-acting interactions (BOX 2). Metabolomics and global analyses of microbiota have also revealed novel interactions with common diseases. However, human studies are complicated by the difficulty of obtaining samples from relevant tissues and by the lack of reliable environmental data, and transformed cell lines can show biological noise that reduces power to observe genetic effects[81]. Thus, as discussed below, experimental organisms such as mice and rats have been widely used to model human disease.

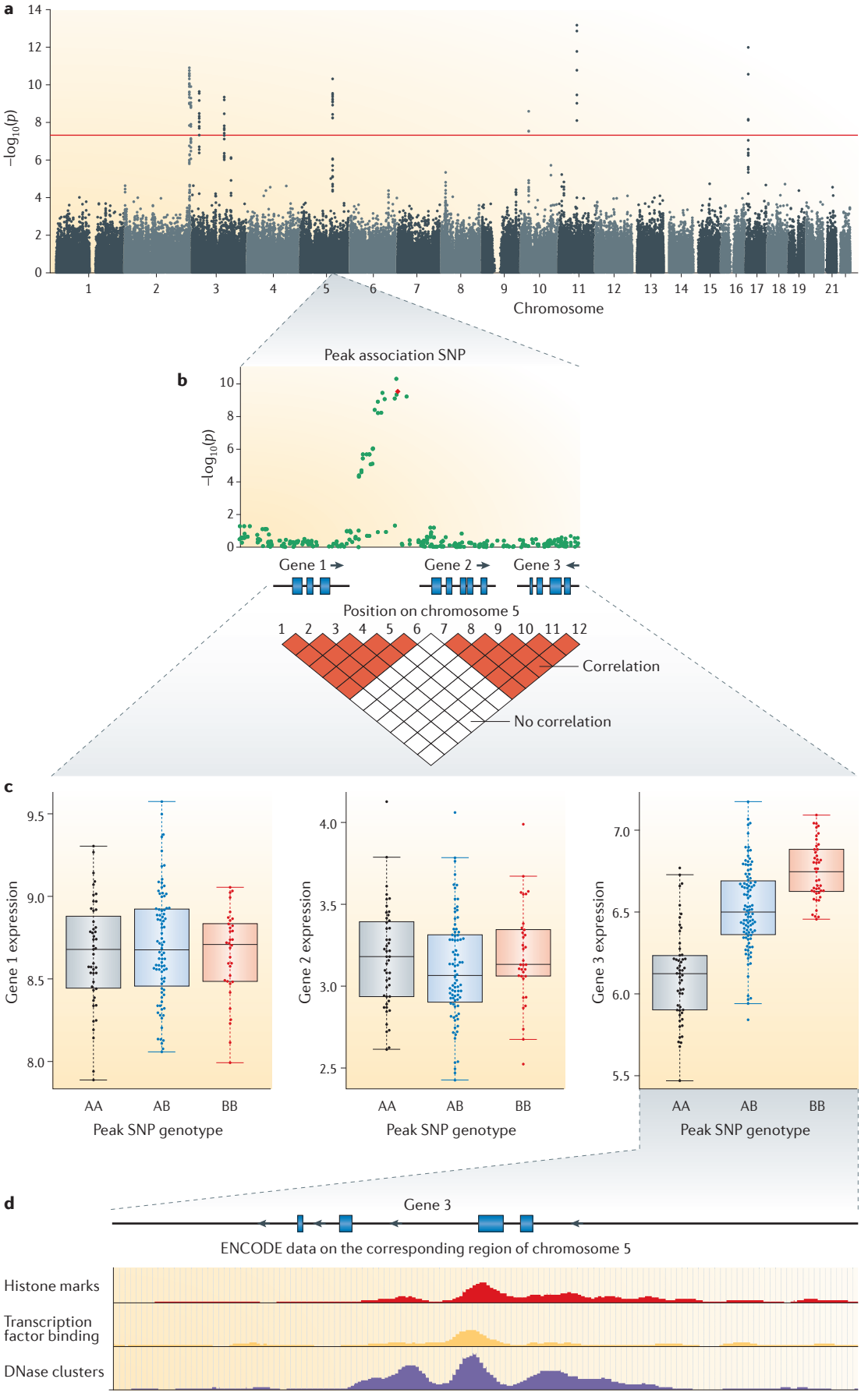
Integration of gene expression with human GWASs. Systems genetics approaches that determine the associations of identified risk variants with the expression levels of transcripts in risk loci in various human tissue samples provide a powerful way to assign priority to the candidate genes. For example, one of the first such studies involved a locus on chromosome 1p13 that is associated with plasma cholesterol levels and myocardial infarction[1]. This locus contains three genes: *CELSR2* (cadherin, EGF LAG seven-pass G-type receptor 2), proline/serine-rich coiled-coil 1 (*PSRC1*) and sortilin 1 (*SORT1*). The GWAS SNPs were found to be associated with transcript levels of *SORT1* in the liver but not in the adipose tissue, and analysis of multiple haplotypes further suggested that the causal SNP affected the binding of CCAAT-enhancer-binding protein (CEB/P) transcription factors. The role of the SNP in *SORT1* regulation was confirmed using cell transfection studies, and the causal role of *SORT1* was verified using overexpression and knockdown studies in mice. In another study, the maternally imprinted Krüppel-like factor 14 gene (*KLF14*) was identified as a regulator of multiple metabolic phenotypes using *cis* and *trans* associations of gene expression[82]. This study shows how eQTL results have been useful in identifying the molecular architecture of complex diseases (BOX 2). A considerable problem with human eQTL studies is obtaining samples of the relevant tissues. In this regard, it is noteworthy that *cis*-eQTLs are often conserved among various tissues, whereas *trans*-eQTLs are usually tissue specific[18].

Haplotypes
Combinations of alleles at genetic loci that are inherited together.

Metabolomics and the microbiome. Broad metabolite screens of human populations have identified novel associations between specific metabolites and diseases such as cancer[83], atherosclerosis[84] and diabetes[85]. The levels of many plasma metabolites are influenced by diet, and studies suggest that the composition of the gut microbiota can also have a major effect. For example, a recent report showed that the metabolite trimethylamine-*N*-oxide (TMAO) is strongly associated with atherosclerosis. TMAO is derived entirely from trimethylamine, which is produced by the gut microbiota during the catabolism of phospholipids[84]. A subsequent study identified the enzyme that can generate trimethylamine and showed that it occurs in only certain genera of bacteria[86]. These and other studies[87,88] highlight a complex set of interactions that involve host genetic background, diet and gut microbiota composition, all of which are likely to contribute to metabolic disorders. In humans, the role of host genetics in setting the composition of the gut microbiota is still unclear — whereas family members have more similar microbiota compositions than unrelated individuals, monozygotic and dizygotic twins have equally similar compositions[89]. However, in mice, in which age, diet and other environmental factors can be more carefully controlled, it seems that the composition of the gut microbiota is heritable[58]. Thus, large studies of humans that emphasize a genetic perspective and more detailed studies in experimental organisms are of considerable interest.

Cancer. One of the first important applications of systems genetics was to improve the categorization of cancers using global gene expression analyses. Thus, expression patterns from histologically similar cancers were found to cluster into distinct groups, which were then found to associate with different clinical outcomes[90]. Recent sequencing studies have revealed clear patterns of genetic variations that are associated with cancers. For example, a recent study from The Cancer Genome Atlas project revealed similarities between some endometrial cancers and subtypes of ovarian and breast cancers[91]. Given that relevant cancer tissues are available, the integration of transcriptomic, proteomic and metabolomic data using systems genetics approaches should provide a useful strategy for unravelling the flow of biological information from these mutations to pathways that control cell growth and other aspects of tumour formation[92,93].

Animal models. Given the difficulties of carrying out systems genetics studies directly in humans, animal models have proven invaluable for studies of complex traits, including atherosclerosis[94,95], heart failure[10,96,97], diabetes[69,98,99], obesity[58,100,101], osteoporosis[102,103], cancer[92] and behavioural disorders[104]. These organisms have been frequently studied in response to environmental stressors or to sensitizing genetic mutations[58,94]. The assumption in such studies is that even if inferences do not directly translate between animal models and humans, the pathways that contribute to



◀ Figure 4 | **Predicting causal genes in GWAS loci.** In this hypothetical example, the association of a clinical trait with multiple genomic loci is discovered through a genome-wide association study (GWAS; part **a**). The red line represents the *p*-value threshold. In order to understand the causal gene (or genes) in the chromosome 5 region, a detailed regional association plot is generated. Although the peak single-nucleotide polymorphism (SNP; shown in red) is in a linkage disequilibrium (LD) block with Gene 1, the neighbouring LD block contains Genes 2 and 3 in close proximity. The matrix below the association plot shows loci that are co-inherited without recombination (red diamonds) and hence form an LD block (part **b**). Expression quantitative trait locus (eQTL) mapping of transcript abundance of the three genes in various tissues can help to predict the causal gene. In this example, Gene 3 has a significant association with the peak GWAS SNP and is therefore the probable causal candidate gene in this locus. Note that it resides in a different LD block from where the peak SNP is located (part **c**). Overlaying the Encyclopedia of DNA Elements (ENCODE) data available for the genomic region that contains Gene 3 helps to generate specific hypotheses for the mechanism of how the GWAS SNP, or the SNPs that are in high LD based on the 1,000 Genomes data, regulates the expression of this gene (part **d**).

the pathologies will be shared. Although there will certainly be important differences, animal models have generally led the way to an understanding of both basic biology and pathology. In this regard, a comprehensive study observed conservation of *trans*-regulated genes that contribute to hypertension between rats and humans[105]. Some recent studies indicate that there is even a substantial overlap between rodents and humans in the genes that contribute to complex traits[58,106]. Conversely, a large mapping study of 122 diverse phenotypes in outbred rats concluded that orthologous genes rarely contribute to the same phenotype in rats and mice[107].

Various reference populations of rats and mice are proving useful for systems genetics studies (TABLE 1). Renewable populations that can be interrogated repeatedly — including inbred strains, recombinant inbred strains and congenic strains — have been particularly useful. These have made it possible to develop large data sets that contain detailed molecular, physiological and pathological phenotypic data under a range of environmental conditions[108]. An important weakness of genetic studies of complex traits in rats and mice has been the poor mapping resolution of linkage analyses using genetic crosses (regions identified are tens of megabases and encode hundreds of genes). This has now been overcome to a large extent using GWASs of either outbred stocks[107] or panels of inbred and recombinant inbred strains (in which most loci show linkage disequilibrium blocks that are <1Mb in size)[16].

New therapies. Systems genetics provides a potentially powerful approach for drug development. One can construct networks, identify modules that are associated with disease traits and target one or more genes in that module[109]. Such an approach takes a broad view of the causes of a disease and can incorporate nonlinear G×G and G×E interactions. Systems-level approaches are proving particularly useful for the development of anticancer therapies[110]. For example, a recent study[111] observed that the inhibition of epidermal growth factor receptors in a subset of breast cancers markedly sensitized these cells to DNA damage, provided that certain drugs were given sequentially but not simultaneously. The authors concluded that the enhanced efficacy results from a dynamic network 'rewiring' by one drug which, in turn, unmasks sensitivity to another[111].

Systems genetics has also been used to screen existing promising compounds against a reference population to identify additional targets or to reveal harmful gene-by-drug interactions. Human LCLs, such as CEPH cell lines, have been widely investigated for such pharmacogenetic studies, which could be considered as a sub-category of systems genetics[112]. Typically, such studies treat the LCLs with various concentrations of drugs, test for differences in toxicity or responsiveness and carry out global gene expression analyses. For example, a recent study of gene expression and statin responsiveness using LCLs from 480 participants of a clinical trial identified the RAS homologue family member A gene (*RHOA*) as a modulator of the cholesterol-lowering effects of a statin[113] and the glycine amidinotransferase gene (*GATM*) as a causal gene for statin-induced myopathy[114].

Large-scale studies. It has been argued that high-throughput, high-dimensional phenotyping will be crucial for the understanding of both genotype–phenotype maps and environmental interactions[115]. Given the complexity of natural variation, particularly that of common disease, and considering that modest genetic variations can have considerable biological consequences, it will be important to assemble large data sets to maximize power through the formation of consortia and meta-analyses. The Genotype–Tissue Expression Program — an ambitious collaborative project sponsored by the US National Institutes of Health — is a step in the right direction[116]. The goal of this project is to collect gene expression data in ~30 different tissues from 1,000 people for systems genetics analyses.

Conclusions and future prospects

The concept of linking population genetic variation with biochemical variation dates back to at least 1970 (REF. 7). The current interest in systems genetics has been driven by recent technological and computational advances, which continue to progress rapidly and are paving the way for large-scale applications of systems genetics in both model organisms and humans. The approach should also be enhanced by the incorporation of additional data modalities, such as high-resolution clinical imaging data, single-cell gene expression data, pathway readout data and time course data. In addition to quantitating the steady-state levels of macromolecules in populations, it may be possible to determine their rates of synthesis and degradation, as well as their localization[117].

Despite striking differences between the sexes in the prevalence of common diseases, the underlying mechanisms have been fairly neglected. Clearly, these differences could be 'mined' to identify factors that confer resistance to disease[118]. Systems genetics analyses of

Recombinant inbred strains
A set of inbred strains that is generally produced by crossing two parental inbred strains and then inbreeding random intercross progeny; they provide a permanent resource for examining the segregation of traits that differ between the parental strains.

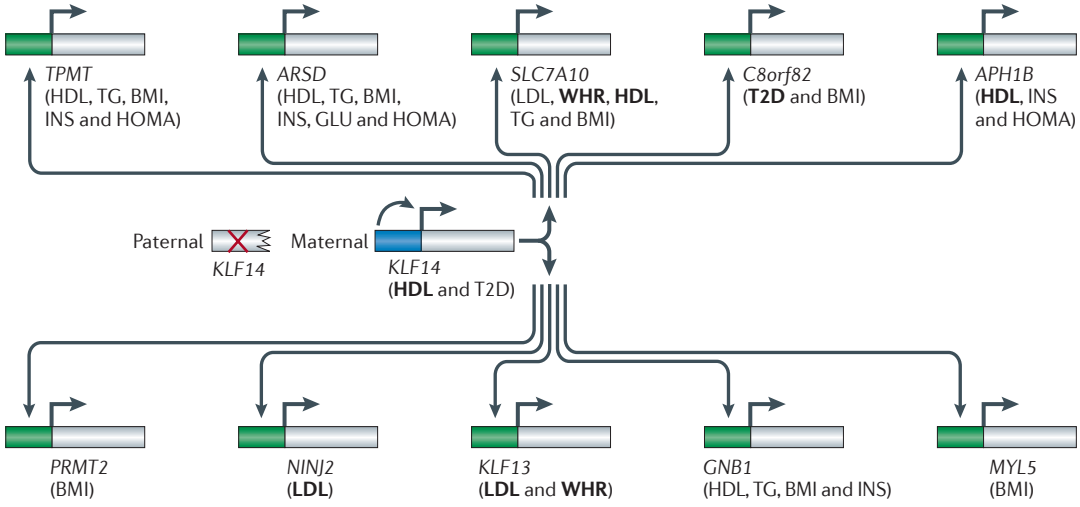
Congenic strains
Strains in which a small region of the genome from one strain has been placed, by repeated crossing, onto the genetic background of a second strain.

Linkage disequilibrium blocks
Regions of high correlation across genetic markers, which results from their linkage in *cis* on a chromosome and thus infrequent recombination during meiosis. LD blocks are often demarcated by recombination hot spots.

CEPH cell lines
A large set of lymphoblastoid cell lines from European pedigrees that serves as a reference collection for studies of allele frequencies, linkage mapping and the genetics of gene expression.

Box 2 | **KLF14 pathway mediates metabolic syndrome traits**

Variants that are located near the maternally expressed imprinted Krüppel-like factor 14 gene (*KLF14*) have been associated with type 2 diabetes (T2D)[120] and with high-density lipoprotein (HDL) cholesterol levels[121] in genome-wide association studies (GWASs) (see the figure). The Multiple Tissue Human Expression Resource (MuTHER) consortium started by asking whether these variants affected the expression of the nearby *KLF14* gene in *cis* in metabolically active adipose tissue that was isolated from >700 women[82]. The significant association between the variants and *KLF14* transcript levels strongly suggested that this gene is the causal gene in the GWAS locus. As KLF14 is a transcription factor, and transcription factors regulate the expression of their target genes in *trans*, members of the consortium reasoned that the transcript abundance of *KLF14*-target genes should be influenced by the GWAS variants that are located near *KLF14*. Forty-six genes with *trans*-eQTLs that mapped to the *KLF14* locus were enriched for KLF-binding sites in their promoter regions, which provided *in silico* evidence for the binding of KLF14 regulating the expression of these genes in *trans*. Ten of these genes (see the figure) had strong *trans* associations that reached the typical genome-wide *p*-value cutoff of 5×10^{-8}. Furthermore, variants near the 5′ end of these genes showed significant associations with metabolic traits. Traits in bold show genome-wide significant associations with local single-nucleotide polymorphisms at the corresponding *trans*-regulated genes; other traits are associated with expression of the corresponding *trans*-regulated genes. Collectively, the *cis*- and *trans*-eQTL networks identified KLF14 as the master regulator of the expression of multiple genes that, in turn, mediate the effects of this transcription factor on metabolic disease. This study showed that *cis*- and *trans*-eQTL networks can be successfully interrogated to predict the causal gene in a GWAS locus and how the gene functions to affect disease susceptibility.



APH1B, APH1B γ-secretase subunit; *ARSD*, arylsulfatase D; BMI, body mass index; *C8orf82*, chromosome 8 open reading frame 82; GLU, glucose levels; *GNB1*, guanine nucleotide-binding protein, β-polypeptide 1; HOMA, index of insulin sensitivity; INS, insulin levels; LDL, low-density lipoprotein levels; *MYL5*, myosin, light chain 5, regulatory; *NINJ2*, ninjurin 2; *PRMT2*, protein arginine methyltransferase 2; *SLC7A10*, solute carrier family 7 member 10; TG, triglyceride levels; *TPMT*, thiopurine *S*-methyltransferase; WHR, waist:hip ratio. The figure is modified, with permission, from REF. 122 © (2011) Macmillan Publishers Ltd. All rights reserved.

experimental organisms, as well as of human populations, have revealed marked differences between the sexes in both gene expression and metabolic activities[119]. Network modelling of transcriptomes suggest that the higher-order interactions differ between the sexes[119], and it would be of interest to understand the relationships of such interactions to disease states or to drug responses.

As discussed above, it is now clear that the gut microbiome contributes, importantly, to multiple common disorders, including atherosclerosis, cancer, colitis, diabetes, depression and fatty liver. Systems genetics should provide a useful approach for understanding the interactions between diet, microbiota composition, plasma metabolites and genetic background.

An increased focus of systems genetics studies on biological processes that are involved in homeostasis — such as cell cycling, oxidative stress, mitochondrial function, and synthesis and turnover of macromolecules — might be productive. Such 'process traits' are likely to be less genetically complex than any of the disease traits to which they contribute, and will hence require smaller samples sizes for mapping. Nevertheless, these traits may integrate considerable genetic variation that is relevant to multiple disease traits.

An important challenge will be to make systems genetics data broadly available to biologists. Currently, much high-throughput data are fairly inaccessible to everyday bench scientists. As educating all biologists in basic computational and statistical skills will be difficult to achieve, it will be desirable to display association, linkage and models on browsers that can be quickly scanned for potential links and insights.

1. Musunuru, K. *et al.* From noncoding variant to phenotype via *SORT1* at the 1p13 cholesterol locus. *Nature* **466**, 714–719 (2010).
2. Ayroles, J. F. *et al.* Systems genetics of complex traits in *Drosophila melanogaster*. *Nature Genet.* **41**, 299–307 (2009).
3. Falconer, D. S. & Mackay, T. F. C. *Introduction to Quantitative Genetics* 4th edn (Longman, 1996).
4. Huang, W. *et al.* Epistasis dominates the genetic architecture of *Drosophila* quantitative traits. *Proc. Natl Acad. Sci. USA* **109**, 15553–15559 (2012).
5. Lynch, M. & Walsh, J. B. *Genetics and Analysis of Quantitative Traits* (Sinauer Associates, 1998).
6. Wu, C. *et al.* Genome-wide association analyses of esophageal squamous cell carcinoma in Chinese identify multiple susceptibility loci and gene–environment interactions. *Nature Genet.* **44**, 1090–1097 (2012).
7. Burns, J. in *Towards a Theoretical Biology* Vol. 3 (ed. Waddington, C. H.) 47–51 (Edinburgh Univ. Press, 1970).
8. Waddington, C. H. *The Strategy of the Genes* 262 (Allen & Unwin, 1957).
9. Passador-Gurgel, G., Hsieh, W. P., Hunt, P., Deighton, N. & Gibson, G. Quantitative trait transcripts for nicotine resistance in *Drosophila melanogaster*. *Nature Genet.* **39**, 264–268 (2007).
10. Petretto, E. *et al.* Integrated genomic approaches implicate osteoglycin (*Ogn*) in the regulation of left ventricular mass. *Nature Genet.* **40**, 546–552 (2008).
11. Aitman, T. J. *et al.* Identification of *Cd36* (*Fat*) as an insulin-resistance gene causing defective fatty acid and glucose metabolism in hypertensive rats. *Nature Genet.* **21**, 76–83 (1999).
12. Schadt, E. E. *et al.* An integrative genomics approach to infer causal associations between gene expression and disease. *Nature Genet.* **37**, 710–717 (2005).
13. Ehrenreich, I. M. *et al.* Genetic architecture of highly complex chemical resistance traits across four yeast strains. *PLoS Genet.* **8**, e1002570 (2012).
14. Brem, R. B., Yvert, G., Clinton, R. & Kruglyak, L. Genetic dissection of transcriptional regulation in budding yeast. *Science* **296**, 752–755 (2002).
This is the first study to carry out a linkage analysis of global gene expression in a cross between a laboratory strain and wild strain of *S. cerevisiae*, which shows widespread *cis* and *trans* regulation of gene expression.
15. Brem, R. B. & Kruglyak, L. The landscape of genetic complexity across 5,700 gene expression traits in yeast. *Proc. Natl Acad. Sci. USA* **102**, 1572–1577 (2005).
16. Bennett, B. J. *et al.* A high-resolution association mapping panel for the dissection of complex traits in mice. *Genome Res.* **20**, 281–290 (2010).
17. Lappalainen, T. *et al.* Transcriptome and genome sequencing uncovers functional variation in humans. *Nature* **501**, 506–511 (2013).
18. van Nas, A. *et al.* Expression quantitative trait loci: replication, tissue- and sex-specificity in mice. *Genetics* **185**, 1059–1068 (2010).
19. Breitling, R. *et al.* Genetical genomics: spotlight on QTL hotspots. *PLoS Genet.* **4**, e1000232 (2008).
20. Orozco, L. D. *et al.* Unraveling inflammatory responses using systems genetics and gene–environment interactions in macrophages. *Cell* **151**, 658–670 (2012).
21. Romanoski, C. E. *et al.* Systems genetics analysis of gene-by-environment interactions in human cells. *Am. J. Hum. Genet.* **86**, 399–410 (2010).
22. Schaub, M. A., Boyle, A. P., Kundaje, A., Batzoglou, S. & Snyder, M. Linking disease associations with regulatory information in the human genome. *Genome Res.* **22**, 1748–1759 (2012).
This study investigates the overlap between the disease-associated SNPs that were identified in GWASs and multiple types of ENCODE data; it shows that up to 80% of the disease-associated variants lie in functional regions of the genome.
23. Civelek, M. *et al.* Genetic regulation of human adipose microRNA expression and its consequences for metabolic traits. *Hum. Mol. Genet.* **22**, 3023–3037 (2013).
24. Kumar, V. *et al.* Human disease-associated genetic variation impacts large intergenic non-coding RNA expression. *PLoS Genet.* **9**, e1003201 (2013).
25. Ghazalpour, A. *et al.* Comparative analysis of proteome and transcriptome variation in mouse. *PLoS Genet.* **7**, e1001393 (2011).
26. Marioni, J. C., Mason, C. E., Mane, S. M., Stephens, M. & Gilad, Y. RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. *Genome Res.* **18**, 1509–1517 (2008).
27. Babak, T. *et al.* Genetic validation of whole-transcriptome sequencing for mapping expression affected by *cis*-regulatory variation. *BMC Genomics* **11**, 473 (2010).
28. Almlof, J. C. *et al.* Powerful identification of *cis*-regulatory SNPs in human primary monocytes using allele-specific gene expression. *PLoS ONE* **7**, e52260 (2012).
29. Lagarrigue, S. *et al.* Analysis of allele specific expression in mouse liver by RNA-seq: a comparison with *cis*-eQTL identified using genetic linkage. *Genetics* **195**,1157–1166 (2013).
30. Degner, J. F. *et al.* DNase I sensitivity QTLs are a major determinant of human expression variation. *Nature* **482**, 390–394 (2012).
31. Bell, J. T. *et al.* DNA methylation patterns associate with genetic and gene expression variation in HapMap cell lines. *Genome Biol.* **12**, R10 (2011).
32. Gaffney, D. J. *et al.* Dissecting the regulatory architecture of gene expression QTLs. *Genome Biol.* **13**, R7 (2012).
This study combines eQTL results from LCLs and regulatory information from the ENCODE project to annotate the putative function of variants that affect gene expression.
33. Neph, S. *et al.* Circuitry and dynamics of human transcription factor regulatory networks. *Cell* **150**, 1274–1286 (2012).
34. Heinz, S. *et al.* Effect of natural genetic variation on enhancer selection and function. *Nature* http://dx.doi.org/10.1038/nature12615 (2013).
This study compares the binding of lineage-determining and specific transcription factors in primary macrophages of two different strains of mice.
35. Kasowski, M. *et al.* Extensive variation in chromatin states across humans. *Science* **342**, 750–752 (2013).
36. Kilpinen, H. *et al.* Coordinated effects of sequence variation on DNA binding, chromatin structure, and transcription. *Science* **342**, 744–747 (2013).
37. McVicker, G. *et al.* Identification of genetic variants that affect histone modifications in human cells. *Science* **342**, 747–749 (2013).
38. Foss, E. J. *et al.* Genetic basis of proteome variation in yeast. *Nature Genet.* **39**, 1369–1375 (2007).
39. Holdt, L. M. *et al.* Quantitative trait loci mapping of the mouse plasma proteome (pQTL). *Genetics* **193**, 601–608 (2013).
40. Lourdusamy, A. *et al.* Identification of *cis*-regulatory variation influencing protein abundance levels in human plasma. *Hum. Mol. Genet.* **21**, 3719–3726 (2012).
41. Melzer, D. *et al.* A genome-wide association study identifies protein quantitative trait loci (pQTLs). *PLoS Genet.* **4**, e1000072 (2008).
42. Wu, L. *et al.* Variation and genetic control of protein abundance in humans. *Nature* **499**, 79–82 (2013).
43. Krishna, R. G. & Wold, F. Post-translational modification of proteins. *Adv. Enzymol. Relat. Areas Mol. Biol.* **67**, 265–298 (1993).
44. Patti, G. J., Yanes, O. & Siuzdak, G. Innovation: Metabolomics: the apogee of the omics trilogy. *Nature Rev. Mol. Cell Biol.* **13**, 263–269 (2012).
45. Liu, S. *et al.* A diurnal serum lipid integrates hepatic lipogenesis and peripheral fatty acid use. *Nature* **502**, 550–554 (2013).
46. Gieger, C. *et al.* Genetics meets metabolomics: a genome-wide association study of metabolite profiles in human serum. *PLoS Genet.* **4**, e1000282 (2008).
47. Kettunen, J. *et al.* Genome-wide association study identifies multiple loci influencing human serum metabolite levels. *Nature Genet.* **44**, 269–276 (2012).
48. Suhre, K. *et al.* Human metabolic individuality in biomedical and pharmaceutical research. *Nature* **477**, 54–60 (2011).
This study profiles >250 metabolites that represent >60 biochemical pathways in ~3,000 people. It shows that many GWAS loci are associated with serum metabolite levels and that the effect sizes for metabolites are much larger than those for clinical traits.
49. Flint, J. & Mackay, T. F. Genetic architecture of quantitative traits in mice, flies, and humans. *Genome Res.* **19**, 723–733 (2009).
50. Jarvis, J. P. & Cheverud, J. M. Mapping the epistatic network underlying murine reproductive fatpad variation. *Genetics* **187**, 597–610 (2011).
51. Shao, H. *et al.* Genetic architecture of complex traits: large phenotypic effects and pervasive epistasis. *Proc. Natl Acad. Sci. USA* **105**, 19910–19914 (2008).
52. Naya, F. J. *et al.* Mitochondrial deficiency and cardiac sudden death in mice lacking the MEF2A transcription factor. *Nature Med.* **8**, 1303–1309 (2002).
53. Weiss, J. N. *et al.* "Good enough solutions" and the genetics of complex diseases. *Circ. Res.* **111**, 493–504 (2012).
54. Zuk, O., Hechter, E., Sunyaev, S. R. & Lander, E. S. The mystery of missing heritability: genetic interactions create phantom heritability. *Proc. Natl Acad. Sci. USA* **109**, 1193–1198 (2012).
This paper discusses that the estimates of missing heritability may be misleading owing to the assumptions of no epistasis when calculating heritability from population data.
55. Prabhu, S. & Pe'er, I. Ultrafast genome-wide scan for SNP–SNP interactions in common complex disease. *Genome Res.* **22**, 2230–2240 (2012).
56. 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).
57. Bloom, J. S., Ehrenreich, I. M., Loo, W. T., Lite, T. L. & Kruglyak, L. Finding the sources of missing heritability in a yeast cross. *Nature* **494**, 234–237 (2013).
This paper uses a cross in yeast to identify the additive and epistatic contributions to heritability of 46 different traits and shows that contribution of gene–gene interactions varies among traits, from near zero to ~50%.
58. Parks, B. W. *et al.* Genetic control of obesity and gut microbiota composition in response to high-fat, high-sucrose diet in mice. *Cell. Metab.* **17**, 141–152 (2013).
This study uses the Hybrid Mouse Diversity Panel to identify the genetic loci that regulate body fat gain and gut microbiota composition in response to a high fat diet. It shows that the estimated heritability of body fat changes can be as high as 85%.
59. Smith, E. N. & Kruglyak, L. Gene–environment interaction in yeast gene expression. *PLoS Biol.* **6**, e83 (2008).
60. Smirnov, D. A., Morley, M., Shin, E., Spielman, R. S. & Cheung, V. G. Genetic analysis of radiation-induced changes in human gene expression. *Nature* **459**, 587–591 (2009).
61. Fu, J. *et al.* System-wide molecular evidence for phenotypic buffering in *Arabidopsis*. *Nature Genet.* **41**, 166–167 (2009).
62. Zhu, J. *et al.* Integrating large-scale functional genomic data to dissect the complexity of yeast regulatory networks. *Nature Genet.* **40**, 854–861 (2008).
63. Pearl, J. *Causality* (Cambridge Univ. Press, 2009).
64. Schwartz, S. M., Schwartz, H. T., Horvath, S., Schadt, E. & Lee, S. I. A systematic approach to multifactorial cardiovascular disease: causal analysis. *Arterioscler Thromb. Vasc. Biol.* **32**, 2821–2835 (2013).
65. Shipley, B. *Cause and Correlation in Biology: A User's Guide to Path Analysis, Structural Equations, and Causal Inference* (Cambridge Univ. Press, 2002).
66. Marbach, D. *et al.* Wisdom of crowds for robust gene network inference. *Nature Methods* **9**, 796–804 (2012).
This study compares >30 methods that aim to reconstruct regulatory networks from high-throughput data and concludes that a consensus network that is constructed by integrating the predictions of different methods has the best performance to infer regulatory interactions.
67. Huan, T. *et al.* A systems biology framework identifies molecular underpinnings of coronary heart disease. *Arterioscler Thromb. Vasc. Biol.* (2013).
68. Zhang, B. *et al.* Integrated systems approach identifies genetic nodes and networks in late-onset Alzheimer's disease. *Cell* **153**, 707–720 (2013).
69. Heinig, M. *et al.* A *trans*-acting locus regulates an anti-viral expression network and type 1 diabetes risk. *Nature* **467**, 460–464 (2010).
70. Hageman, R. S., Leduc, M. S., Korstanje, R., Paigen, B. & Churchill, G. A. A Bayesian framework for inference of the genotype–phenotype map for segregating populations. *Genetics* **187**, 1163–1170 (2011).
71. Neto, E. C. *et al.* Modeling causality for pairs of phenotypes in system genetics. *Genetics* **193**, 1003–1013 (2013).
72. Blair, R. H., Kliebenstein, D. J. & Churchill, G. A. What can causal networks tell us about metabolic pathways? *PLoS Comput. Biol.* **8**, e1002458 (2012).

73. Li, Y., Tesson, B. M., Churchill, G. A. & Jansen, R. C. Critical reasoning on causal inference in genome-wide linkage and association studies. *Trends Genet.* **26**, 493–498 (2010).
74. Chaibub Neto, E., Ferrara, C. T., Attie, A. D. & Yandell, B. S. Inferring causal phenotype networks from segregating populations. *Genetics* **179**, 1089–1100 (2008).
75. Li, R. *et al.* Structural model analysis of multiple quantitative traits. *PLoS Genet.* **2**, e114 (2006).
76. Ward, L. D. & Kellis, M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res.* **40**, D930–D934 (2012).
77. Boyle, A. P. *et al.* Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res.* **22**, 1790–1797 (2012).
78. Lehner, B. Genotype to phenotype: lessons from model organisms for human genetics. *Nature Rev. Genet.* **14**, 168–178 (2013).
79. Costanzo, M. *et al.* The genetic landscape of a cell. *Science* **327**, 425–431 (2010).
80. Dixon, S. J., Costanzo, M., Baryshnikova, A., Andrews, B. & Boone, C. Systematic mapping of genetic interaction networks. *Annu. Rev. Genet.* **43**, 601–625 (2009).
81. Choy, E. *et al.* Genetic analysis of human traits *in vitro*: drug response and gene expression in lymphoblastoid cell lines. *PLoS Genet.* **4**, e1000287 (2008).
82. Small, K. S. *et al.* Identification of an imprinted master *trans* regulator at the *KLF14* locus related to multiple metabolic phenotypes. *Nature Genet.* **43**, 561–564 (2011).
This study identifies *KLF14* as the causal gene in a GWAS locus that is associated with both diabetes and lipoprotein levels and dissects its role as a master regulator of gene expression in human fat tissues.
83. Sreekumar, A. *et al.* Metabolomic profiles delineate potential role for sarcosine in prostate cancer progression. *Nature* **457**, 910–914 (2009).
84. Wang, Z. *et al.* Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease. *Nature* **472**, 57–63 (2011).
Using an unbiased metabolomics approach this study reports the identification of a serum metabolite that is derived from dietary choline produced by the gut microbiota as a novel risk factor for cardiovascular disease.
85. Wang, T. J. *et al.* Metabolite profiles and the risk of developing diabetes. *Nature Med.* **17**, 448–453 (2011).
86. Craciun, S. & Balskus, E. P. Microbial conversion of choline to trimethylamine requires a glycyl radical enzyme. *Proc. Natl Acad. Sci. USA* **109**, 21307–21312 (2012).
87. Koeth, R. A. *et al.* Intestinal microbiota metabolism of L-carnitine, a nutrient in red meat, promotes atherosclerosis. *Nature Med.* **19**, 576–585 (2013).
88. Qin, J. *et al.* A metagenome-wide association study of gut microbiota in type 2 diabetes. *Nature* **490**, 55–60 (2012).
89. Lozupone, C. A., Stombaugh, J. I., Gordon, J. I., Jansson, J. K. & Knight, R. Diversity, stability and resilience of the human gut microbiota. *Nature* **489**, 220–230 (2012).
90. Perou, C. M. *et al.* Molecular portraits of human breast tumours. *Nature* **406**, 747–752 (2000).
91. Kandoth, C. *et al.* Integrated genomic characterization of endometrial carcinoma. *Nature* **497**, 67–73 (2013).
92. Quigley, D. & Balmain, A. Systems genetics analysis of cancer susceptibility: from mouse models to humans. *Nature Rev. Genet.* **10**, 651–657 (2009).
93. Fendler, B. & Atwal, G. Systematic deciphering of cancer genome networks. *Yale J. Biol. Med.* **85**, 339–345 (2012).
94. Wang, S. S. *et al.* Identification of pathways for atherosclerosis in mice: integration of quantitative trait locus analysis and global gene expression data. *Circ. Res.* **101**, e11–e30 (2007).
95. Yang, X. *et al.* Identification and validation of genes affecting aortic lesions in mice. *J. Clin. Invest.* **120**, 2414–2422 (2010).
96. Hubner, N. *et al.* Integrated transcriptional profiling and linkage analysis for identification of genes underlying disease. *Nature Genet.* **37**, 243–253 (2005).
97. McDermott-Roe, C. *et al.* Endonuclease G is a novel determinant of cardiac hypertrophy and mitochondrial function. *Nature* **478**, 114–118 (2011).
References 10 and 97 use various systems genetics approaches to identify both endonuclease G and osteoglycin as causal genes in loci that underlie left ventricular heart mass in rats.
98. Hodgin, J. B. *et al.* Identification of cross-species shared transcriptional networks of diabetic nephropathy in human and mouse glomeruli. *Diabetes* **62**, 299–308 (2012).
This study shows the conservation of glomerular gene expression networks of humans and of different mouse models of diabetic nephropathy.
99. Keller, M. P. & Attie, A. D. Physiological insights gained from gene expression analysis in obesity and diabetes. *Annu. Rev. Nutr.* **30**, 341–364 (2010).
100. Wang, S. *et al.* Genetic and genomic analysis of a fat mass trait with complex inheritance reveals marked sex specificity. *PLoS Genet.* **2**, e15 (2006).
101. Yang, X. *et al.* Validation of candidate causal genes for obesity that affect shared metabolic pathways and networks. *Nature Genet.* **41**, 415–423 (2009).
102. Calabrese, G. *et al.* Systems genetic analysis of osteoblast-lineage cells. *PLoS Genet.* **8**, e1003150 (2012).
103. Farber, C. R. *et al.* Mouse genome-wide association and systems genetics identify *Asxl2* as a regulator of bone mineral density and osteoclastogenesis. *PLoS Genet.* **7**, e1002038 (2011).
104. Park, C. C. *et al.* Gene networks associated with conditional fear in mice identified using a systems genetics approach. *BMC Syst. Biol.* **5**, 43 (2011).
105. Langley, S. R. *et al.* Systems-level approaches reveal conservation of *trans*-regulated genes in the rat and genetic determinants of blood pressure in humans. *Cardiovasc. Res.* **97**, 653–665 (2013).
106. Davis, R. C. *et al.* Genome-wide association mapping of blood cell traits in mice. *Mamm. Genome* **24**, 105–118 (2013).
107. Baud, A. *et al.* Combined sequence-based and genetic mapping analysis of complex traits in outbred rats. *Nature Genet.* **45**, 767–775 (2013).
108. van Nas, A. *et al.* The systems genetics resource (SGR): a web application to mine global data for complex disease traits. *Front. Genet.* **4**, 84 (2013).
109. Schadt, E. E., Friend, S. H. & Shaywitz, D. A. A network view of disease and compound screening. *Nature Rev. Drug Discov.* **8**, 286–295 (2009).
110. Erler, J. T. & Linding, R. Network medicine strikes a blow against breast cancer. *Cell* **149**, 731–733 (2012).
111. Lee, M. J. *et al.* Sequential application of anticancer drugs enhances cell death by rewiring apoptotic signaling networks. *Cell* **149**, 780–794 (2012).
112. Min, J. L. *et al.* The use of genome-wide eQTL associations in lymphoblastoid cell lines to identify novel genetic pathways involved in complex traits. *PLoS ONE* **6**, e22070 (2011).
113. Medina, M. W. *et al. RHOA* is a modulator of the cholesterol-lowering effects of statin. *PLoS Genet.* **8**, e1003058 (2012).
114. Mangravite, L. M. *et al.* A statin-dependent QTL for *GATM* expression is associated with statin-induced myopathy. *Nature* **502**, 377–380 (2013).
115. Houle, D., Govindaraju, D. R. & Omholt, S. Phenomics: the next challenge. *Nature Rev. Genet.* **11**, 855–866 (2010).
116. GTEx Consortium. The Genotype-Tissue Expression (GTEx) project. *Nature Genet.* **45**, 580–585 (2013).
117. Pai, A. A. *et al.* The contribution of RNA decay quantitative trait loci to inter-individual variation in steady-state gene expression levels. *PLoS Genet.* **8**, e1003000 (2012).
118. Arnold, A. P. & Lusis, A. J. Understanding the sexome: measuring and reporting sex differences in gene systems. *Endocrinology* **153**, 2551–2555 (2012).
119. van Nas, A. *et al.* Elucidating the role of gonadal hormones in sexually dimorphic gene coexpression networks. *Endocrinology* **150**, 1235–1249 (2009).
120. Voight, B. F. *et al.* Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. *Nature Genet.* **42**, 579–589 (2010).
121. Teslovich, T. M. *et al.* Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* **466**, 707–713 (2010).
122. Civelek, M. & Lusis, A. J. Conducting the metabolic syndrome orchestra. *Nature Genet.* **43**, 506–508 (2011).
123. Schadt, E. E. Molecular networks as sensors and drivers of common human diseases. *Nature* **461**, 218–223 (2009).
124. Zhu, J. *et al.* Stitching together multiple data dimensions reveals interacting metabolomic and transcriptomic networks that modulate cell regulation. *PLoS Biol.* **10**, e1001301 (2012).
125. Atwell, S. *et al.* Genome-wide association study of 107 phenotypes in *Arabidopsis thaliana* inbred lines. *Nature* **465**, 627–631 (2010).
126. Gaertner, B. E., Parmenter, M. D., Rockman, M. V., Kruglyak, L. & Phillips, P. C. More than the sum of its parts: a complex epistatic network underlies natural variation in thermal preference behavior in *Caenorhabditis elegans*. *Genetics* **192**, 1533–1542 (2012).
127. Rockman, M. V., Skrovanek, S. S. & Kruglyak, L. Selection at linked sites shapes heritable phenotypic variation in *C. elegans*. *Science* **330**, 372–376 (2010).
128. Jumbo-Lucioni, P. *et al.* Systems genetics analysis of body weight and energy metabolism traits in *Drosophila melanogaster*. *BMC Genomics* **11**, 297 (2010).
129. King, E. G. *et al.* Genetic dissection of a model complex trait using the *Drosophila* Synthetic Population Resource. *Genome Res.* **22**, 1558–1566 (2012).
130. Philip, V. M. *et al.* Genetic analysis in the Collaborative Cross breeding population. *Genome Res.* **21**, 1223–1238 (2011).
131. Churchill, G. A., Gatti, D. M., Munger, S. C. & Svenson, K. L. The Diversity Outbred mouse population. *Mamm. Genome* **23**, 713–718 (2012).
132. Aitman, T. J. *et al.* Progress and prospects in rat genetics: a community view. *Nature Genet.* **40**, 516–522 (2008).
133. Printz, M. P., Jirout, M., Jaworski, R., Alemayehu, A. & Kren, V. Genetic models in applied physiology. HXB/BXH rat recombinant inbred strain platform: a newly enhanced tool for cardiovascular, behavioral, and developmental genetics and genomics. *J. Appl. Physiol.* **94**, 2510–2522 (2003).
134. Simonis, M. *et al.* Genetic basis of transcriptome differences between the founder strains of the rat HXB/BXH recombinant inbred panel. *Genome Biol.* **13**, R31 (2012).
135. Stancakova, A. *et al.* Hyperglycemia and a common variant of *GCKR* are associated with the levels of eight amino acids in 9,369 Finnish men. *Diabetes* **61**, 1895–1902 (2012).

Acknowledgements
The authors thank R. Chen for assistance in the preparation of this paper. M.C. is supported by Ruth L. Kirschstein National Research Service Award T32HL69766; A.J.L. is supported by the US National Institutes of Health grants HL30568, HL28481, HL094322, HL110667 and DP3D094311, and Transatlantic Networks of Excellence Award from Foundation Leducq. They are also grateful to the detailed and critical reviewers.

Competing interests statement
The authors declare no competing interests.

FURTHER INFORMATION
***Drosophila* Synthetic Population Resource:** http://wfitch.bio.uci.edu/~dspr/index.html
Genome Reference Panel: http://www.dpgp.org/
Human Metabolome Project: www.metabolomics.ca/
LIPID MAPS: www.lipidmaps.org/
System Genetics Resource: http://systems.genetics.ucla.edu

ALL LINKS ARE ACTIVE IN THE ONLINE PDF