The genetics of quantitative traits: challenges and prospects

*Trudy F. C. Mackay**‡, Eric A. Stone*‡§ and Julien F. Ayroles**‡

Abstract | A major challenge in current biology is to understand the genetic basis of variation for quantitative traits. We review the principles of quantitative trait locus mapping and summarize insights about the genetic architecture of quantitative traits that have been obtained over the past decades. We are currently in the midst of a genomic revolution, which enables us to incorporate genetic variation in transcript abundance and other intermediate molecular phenotypes into a quantitative trait locus mapping framework. This systems genetics approach enables us to understand the biology inside the 'black box' that lies between genotype and phenotype in terms of causal networks of interacting genes.

Quantitative trait locus
A region of the genome containing one or more genes that affect variation in a quantitative trait, which is identified by its linkage to polymorphic marker loci.

*Department of Genetics, ‡W. M. Keck Center for Behavioral Biology and §Department of Statistics, North Carolina State University, Raleigh, North Carolina 27695, USA. Correspondence to T.F.C.M. e-mail: trudy_mackay@ncsu.edu
doi:10.1038/nrg2612
Published online 7 July 2009

Natural populations harbour a stunning diversity of phenotypic variation for morphology, physiology, behaviour and disease susceptibility. This phenotypic variation is typically due to underlying genetic complexity from multiple interacting loci, with allelic effects that are sensitive to the environmental conditions each individual experiences[1,2]. Understanding the relationship between DNA sequence variation and variation in phenotypes for these quantitative or complex traits will yield insights that are important for predicting disease risk and individual therapeutic treatments in human populations, for increasing the speed of selective breeding programmes in agriculturally important plants and animals and for predicting adaptive evolution.

The principles of mapping quantitative trait loci (QTLs) that affect the natural variation in complex traits by linkage to polymorphic marker loci with Mendelian segregation have been known since the early twentieth century[3]. Until the late 1980s, the lack of polymorphic markers limited the genetic dissection of complex traits to a few model organisms[4,5]. Since then, the discovery of abundant molecular markers, advances in rapid and cost-effective genotyping methods and the development of statistical methods for QTL mapping have revolutionized the field of mapping quantitative traits. The landmark paper by Lander and Botstein[6] launched an avalanche of studies mapping QTLs, which has culminated in recent large-scale genome-wide maps of QTLs that affect human quantitative traits and diseases[7–9].

Despite two decades of intensive effort, we have fallen short of our long-term goal of explaining genetic variation for quantitative traits in terms of the underlying genes, the effects of segregating alleles in different genetic backgrounds and in a range of ecologically relevant environments as well as on other traits, the molecular basis of functional allelic effects and the population frequency of causal variants[10]. The many studies mapping QTLs that affect human diseases and complex traits have uncovered new loci and provided unexpected insights into the biology of disease but, together, these loci account for only a small fraction of the total genetic variation in the population and they rarely map to individual genes[7–9]. The hurdle is not the intellectual foundation of QTL mapping methods but technological limitations. We are currently in the midst of another genomic revolution, with the development of economical, massively parallel technology for DNA and RNA sequencing and platforms for rapidly genotyping hundreds of thousands of polymorphic markers. Here, we review the principles of QTL mapping and summarize insights from previous studies about the genetic architecture of quantitative traits. We then indicate how new technologies can be applied to solve current challenges and describe how a systems genetics approach[11] for integrating genotype–phenotype relationships across multiple levels of biological organization can uncover genetic pathways that affect the variation of complex traits.

QTL mapping

The premise of QTL mapping is that QTLs can be localized through their genetic linkage to visible marker loci with genotypes that we can readily classify. If a QTL is linked to a marker locus, then individuals with different marker locus genotypes will have different mean values of the quantitative trait[3,6]. The most common molecular markers are SNPs, polymorphic insertions or deletions

(indels) and simple sequence repeats (microsatellites). QTLs can be mapped in families or the segregating progeny of crosses between genetically divergent strains (linkage mapping), or in unrelated individuals from the same population (association mapping) (BOX 1).

Mapping QTLs has two components: detection and localization. The power to detect QTLs depends on their effects and allele frequencies. By effect, we mean the average difference in the phenotype of the trait between marker allele genotypes (δ), scaled by the phenotypic standard deviation of the trait within marker genotype classes (σ_W). Homozygous effects refer to the difference in the mean of the trait between the two homozygous genotypes, and heterozygous effects refer to the difference between the mean of the trait in the heterozygous genotype from the average of the means of the trait in the two homozygous genotypes[1,2]. The number of individuals needed to map QTLs increases as δ/σ_W decreases and as allele frequencies depart from 0.5. The power to detect QTLs at an intermediate frequency is similar for both linkage and association mapping studies. However, allele frequencies can be more extreme with association mapping designs and this translates to an increase in the sample sizes that are required to detect QTLs (FIG. 1).

Localizing QTLs depends on the recombination frequency. In a linkage mapping context, recombination events need to occur in the mapping population. As the size of the interval in which we wish to localize the QTL decreases, the number of individuals required to detect at least one recombinant in the region of interest increases, as does the number of molecular markers necessary to detect recombination events. Association mapping uses historical recombination between QTLs and marker alleles in a random mating population and does not require as many individuals as linkage mapping for localizing QTLs (FIG. 1). The number of markers required in an association mapping study depends on the scale and pattern of linkage disequilibrium (LD). If a group of markers is in high LD, we only need to genotype one of them as a proxy for all of the other markers in the LD block. Thus, in species with large LD blocks, such as pure breeds of dogs, only a few markers might be required for QTL detection, but it will not be possible to localize QTLs precisely by within-breed association mapping[12]. By contrast, knowledge of all sequence variants is necessary for association mapping in species such as *Drosophila melanogaster*, in which LD can decline rapidly over short physical distances. However, in this scenario, QTL localization can be quite precise[13].

Because large numbers of individuals and genotypes per individual are necessary to detect and localize QTLs in a single mapping effort, QTL mapping is an iterative procedure, in which we first determine the general locations of QTLs and subsequently focus on high-resolution mapping of the regions containing the QTLs. The second phase requires generating or sampling more individuals to obtain the necessary recombinations and identifying molecular markers in the region of interest. These experiments are laborious and rarely result in positional cloning of QTLs, but instead delimit genomic regions that contain many positional candidate genes.

In organisms with well-annotated genomes, we can query which of the candidate genes in the QTL region are causal. High-resolution recombination mapping provides unambiguous proof of causality. Strategies to corroborate evidence of causality in the absence of recombination mapping include replication in independent studies, identifying potentially functional DNA polymorphisms between alternative alleles of one of the candidate genes, showing a difference in mRNA expression levels between genotypes, showing that mRNA or protein is expressed in tissues thought to be relevant to the trait and showing that mutations in candidate genes affect the trait or fail to complement QTL alleles. Formal proof that a specific allelic substitution affects the trait is provided by replacing the allele of a candidate gene in one strain with the allele in another strain without introducing any other changes in the genetic background, which is currently only possible in yeast[14].

Genetic architecture: lessons learned

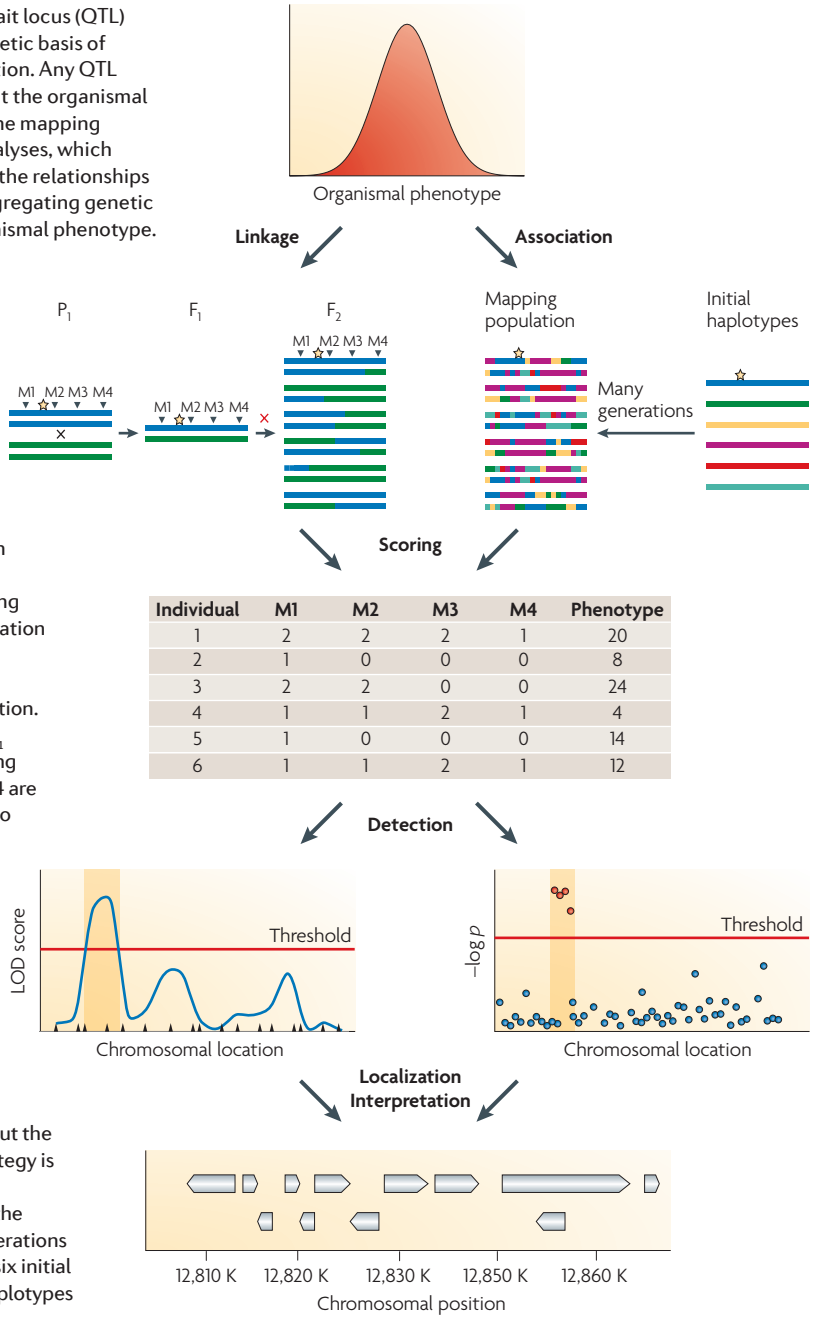
Many loci with small effects. Early QTL mapping studies were performed with sample sizes in the order of hundreds of individuals and approximately 100 molecular markers and, for most traits, consistently detected few QTLs with moderately large effects[10]. These results, combined with the successful positional cloning and identification of several QTLs with large effects[15–18], were encouraging and indicated that the genetic architecture of quantitative traits was moderately complex. This led to optimism that high-resolution mapping, one QTL at a time, would identify the genes that cause the natural variation in quantitative traits. A more pessimistic interpretation was that the experiments were underpowered and could not detect most of the QTLs with smaller effects that truly caused variation in the traits, and that the initial experiments were either lucky in mapping QTLs with smaller effects in the same direction that happened to cluster together or that the initial effects were overestimated[19]. If the effects were overestimated, one would expect that increasing the numbers of individuals and markers used would lead to estimates of larger numbers of QTLs with smaller effects, and that high-resolution mapping would identify multiple closely linked QTLs that underlie each linkage or association peak. These expectations have been confirmed. In both model organisms and humans, increasing sample sizes and marker densities increases the number of QTLs detected, with concomitant decreases in the average effect sizes[20].

High-resolution mapping typically shows that single QTLs fractionate into multiple closely linked QTLs, which often have opposite effects[20]. A striking example was provided by the detailed dissection of the effects on growth rate of a 210 kb region of the *Arabidopsis thaliana* genome that was not associated with growth rate in a QTL genome scan[21]. This random genomic region contained two tightly linked QTLs with clear — albeit small — effects on growth rate, which occurred in opposite directions in the two parental strains used to construct the mapping population. Furthermore, the

Linkage disequilibrium (LD). The correlation (non-random association) of alleles at two or more polymorphic loci. Alleles that are in LD co-occur in individuals more often than the random expectation from the product of their allele frequencies in the population.

Box 1 | **Quantitative trait locus mapping**

The purpose of quantitative trait locus (QTL) mapping is to uncover the genetic basis of quantitative phenotypic variation. Any QTL analysis therefore assumes that the organismal phenotype is variable within the mapping population. Linkage-based analyses, which focus on individuals for which the relationships are known, seek to identify segregating genetic markers that predict the organismal phenotype. Predictive markers are near (linked to) causal loci, and so the predictive markers and the causal loci tend to segregate together. This tendency is disrupted by recombination, and the probability of recombination increases with physical distance; the most predictive markers are therefore expected to reside in proximity of the causal locus.



Individual	M1	M2	M3	M4	Phenotype
1	2	2	2	1	20
2	1	0	0	0	8
3	2	2	0	0	24
4	1	1	2	1	4
5	1	0	0	0	14
6	1	1	2	1	12

The figure shows an F_2 mapping population. The parental generation (P_1) consists of two genetically divergent inbred lines that are crossed to create the F_1 generation. Crossing individuals from the F_1 generation yields the F_2 mapping population. M1, M2, M3 and M4 are markers that distinguish the two parental strains and are used to map the organismal phenotype. The yellow star indicates the position of a causal locus or QTL. Recombination in the F_2 population creates new haplotypes and can uncouple marker genotypes from the causal locus.

Association mapping is also based on recombination, but the recombination used in this strategy is historical. Thus, the association mapping population shown in the figure is removed by many generations from its progenitors. We show six initial founder haplotypes and the haplotypes in the population after many generations of random mating, indicating how recombination has effectively shuffled the initial haplotypes. The effect of this shuffling is to uncouple all but the most tightly linked markers from the causal locus; because only these tightly linked markers will predict the organismal phenotype, the causal locus can be localized with precision.

In either strategy, the purpose of the mapping population is to supply the genotypic variation through which variation in the organismal phenotype can be explained. As such, both approaches require that organismal phenotypes and marker genotypes are scored (0 and 2 indicate alternative homozygous genotypes and 1 indicates the heterozygous genotype at each biallelic marker). The marker and trait data are then assessed to determine whether there is a mean difference in the trait phenotypes between marker genotype classes. If there is, the marker is linked to the QTL. Linkage mapping typically uses interval mapping to estimate the map position and effect of each QTL[1,2,6,105,106], whereas markers are tested singly in association mapping designs. In both cases, the significance threshold needs to be adjusted for the number of independent tests performed. In linkage-based studies, the haplotype blocks in the mapping population might be large and, as a consequence, the causal locus might only be mapped to a large region. The haplotype blocks in an association mapping population tend to be much smaller, so it might be possible to localize the causal locus to a small genomic region. The QTL region might identify relevant genes for future study or suggest candidates for targeted sequencing or experimental perturbation.

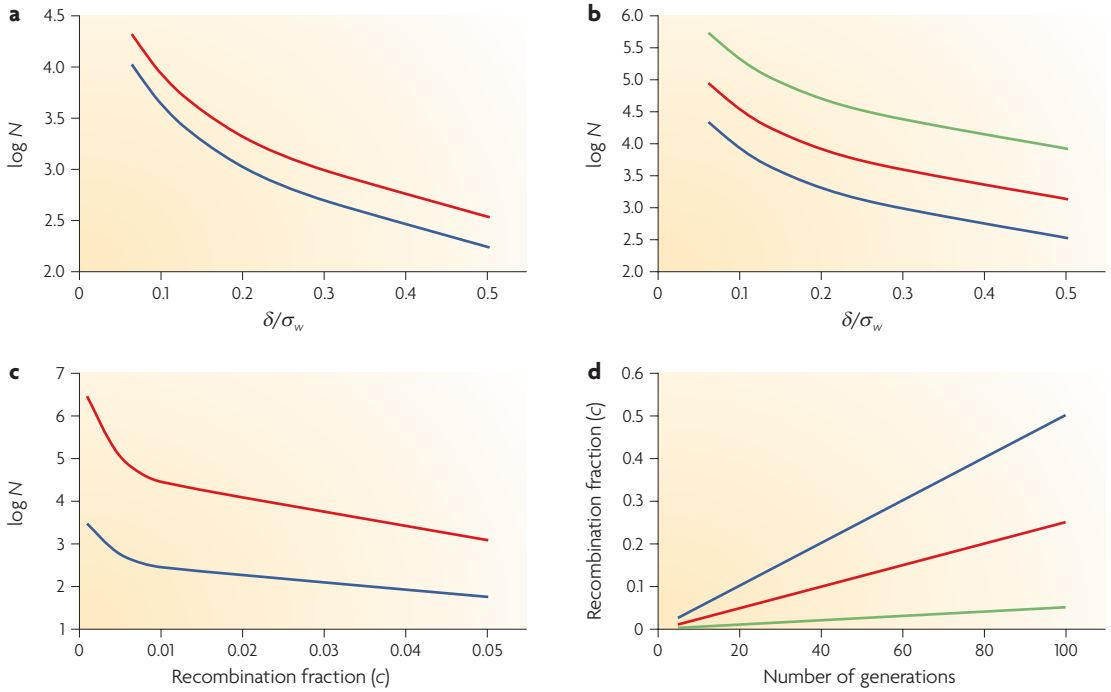


Figure 1 | **Power to localize and detect quantitative trait loci. a** | Numbers of individuals (log_{10} scale) required to detect quantitative trait loci (QTLs) for a range of effect sizes (δ/σ_w) in backcrossed (blue) and F_2 (red) linkage mapping populations. **b** | Numbers of individuals (log_{10} scale) required to detect QTLs for a range of effect sizes in association mapping populations in which the minor allele frequency is 0.5 (blue), 0.25 (red) and 0.1 (green). **c** | Log_{10} of the number of individuals required to detect at least one recombinant in an interval of size *c* (*c* = 100 centiMorgans; cM) (blue) and log_{10} of the number of marker genotypes needed to localize QTLs per 100 cM (red). **d** | The expected frequency of recombinants after *t* generations of recombination in a random mating population, for a per generation recombination fraction of *c* = 0.01 (blue), *c* = 0.005 (red) and *c* = 0.001 (green). δ, average difference in the trait phenotype between marker allele genotypes; σ_W, phenotypic standard deviation of the trait within marker genotype classes.

dissection of a 32 kb region containing a QTL with a large effect on the ability of yeast to grow at high temperatures identified 3 tightly linked QTLs with smaller effects as the true genetic basis of this trait[22]. The inescapable conclusions from the past two decades of studies are that QTL alleles with large effects are rare and that the bulk of genetic variation for quantitative traits is due to many loci with effects that were individually or in aggregate (owing to the tight linkage of QTLs with opposite effects) too small to detect because previous studies were underpowered.

Novel loci. Most knowledge on the genetic basis of complex traits comes from analysis of mutations in model organisms, which have been invaluable in identifying the genes and genetic networks required for producing the wild-type trait phenotype. Classical mutagenesis screens focus on null alleles with large phenotypic effects. It is possible that segregating variation might not be maintained in natural populations at loci that are required for wild-type expression of the trait, and that mutagenesis screens miss or possibly ignore loci at which mutations with subtle effects could affect quantitative traits. Mapping natural variants that affect quantitative trait phenotypes thus potentially complements mutagenesis; therefore, it is important to ask to what extent the genes detected by both methods overlap. The best evidence for some, but not extensive, overlap between the results from mutagenesis and QTL mapping comes from *D. melanogaster*, in which several quantitative traits have been studied by both approaches. For example, many genes in which mutations affect bristle and wing development map to the same regions as QTLs affecting sensory bristle number[23,24] and wing shape[25,26], and molecular polymorphisms at several of these loci are associated with quantitative genetic variation in these traits in natural populations[27–29]. However, many QTLs affecting sensory bristle number do not span genomic regions that contain obvious candidate genes[23,24,30]. QTLs for longevity, resistance to starvation stress, male mating behaviour, olfactory behaviour and locomotor behaviour rarely map to known genes affecting these traits, and conversely, complementation tests to mutations show that variation in unexpected and new loci potentially correspond to the QTLs[13,31–37]. Collectively, these results highlight how little we know about candidate genes that affect quantitative traits and that most of the genome is uncharted territory with respect to the phenotypic effects of naturally segregating alleles that affect even extensively studied phenotypes in a genetic model organism. More optimistically, it is clear that quantitative genetic analysis is an efficient method for functional genome annotation.

Context-dependent effects. If the effects of QTL alleles differ in their magnitude or direction in different genetic backgrounds, different environments or between males and females, they are said to be context dependent. Context-dependent effects are formally identified by appropriate statistical analyses as significant genotype-by-genotype interactions (epistasis), genotype-by-environment interactions (GEIs) and genotype-by-sex interactions (GSIs). Context-dependent effects are often considered a nuisance because estimates of allelic effects are relevant only to the sex, environment and genetic background in which the phenotypes were assessed, and inferences made under laboratory conditions might not be valid across a wide range of natural environments. Furthermore, marginal effects of alleles with highly context-dependent effects might not be detectable when they are averaged over multiple environments or genetic backgrounds. However, context-dependent effects are important. Epistatic interactions identify genetic networks that affect complex traits[38], and GEIs and GSIs are potential mechanisms that maintain the genetic variation of quantitative traits in natural populations[39,40].

In quantitative genetics, epistasis refers to the masking of genotypic effects at one locus by genotypes of another locus[38] and also to any statistical interaction between the genotypes at two or more loci[1,2]. Epistasis is common between mutations that affect the same quantitative trait, as shown by the extensive epistatic networks between *D. melanogaster* mutations that affect metabolic activity[41] and olfactory[42,43] and locomotor[44,45] behaviours. It is more difficult to detect epistatic interactions in QTL mapping studies because the significance threshold becomes low after adjusting for the large number of pairwise tests for marker–marker interactions, and large mapping populations are required to sample individuals in the rarer two-locus genotype classes. Nevertheless, epistasis is a common hallmark of the genetic architecture of quantitative traits in organisms in which controlled crosses reduce genetic heterogeneity (because the frequency of all segregating alleles is 0.5) and optimize the power to detect genotype-by-genotype interactions. Epistatic interactions have been documented in *D. melanogaster* between QTLs affecting numbers of sensory bristles, wing shape, longevity and locomotor behaviour[20,46]; in mice for a number of traits related to growth, body weight and morphometry[20]; and for the growth rate of chickens[47], *A. thaliana*[21] and yeast[22,48]. Epistatic effects can be as large as main QTL effects, and can occur in opposite directions between different pairs of interacting loci and between loci without significant main effects on the trait. Epistatic effects can also occur between closely linked QTLs[21,22,48] and even between polymorphisms at a single locus[49]. Given the difficulties in detecting epistasis in model organisms, it is not surprising that epistatic interactions have not been widely implicated in genome-wide association studies for human complex traits and diseases[20]. However, widespread epistasis might plausibly account for the small marginal effects of loci with significant associations in these studies.

QTL mapping studies are not usually performed in multiple environments, but when such studies are done, the effects often differ in magnitude and sometimes in direction, depending on the environmental circumstances. GEIs have been shown for most phenotypes for which they were assessed in mice[50–52] and *D. melanogaster*[20]. There is evidence that context-dependent effects are also important for human complex traits. A promoter variant of the monoamine oxidase A (*MAOA*) gene is associated with violent behaviour only if the individual was abused as a child[53], and a promoter variant in the solute carrier gene *SLC6A4* (also known as *5-HTT)* is associated with symptoms of depression, but only if the individuals have suffered stressful life events[54]. Finally, many QTLs have different effects in males and females, even if they are not X-linked and are not subject to dosage compensation[23,55–58].

Pleiotropy. In a broad sense, the term pleiotropy refers to the effect of a gene on more than one phenotype, and in a narrow sense, the term refers to the effect of a particular allele on more than one phenotype. Pleiotropy in the narrow sense is responsible for stable genetic correlations between quantitative traits if pleiotropic effects at multiple loci affecting the traits are in the same direction[1]. Positive genetic correlations can occur between traits that share a common biological process or are components of the same structure, and negative genetic correlations are often found between components of fitness[1]. Understanding the underlying pleiotropic connections between quantitative traits is thus important for predicting correlated responses to artificial selection[1] and assessing the contribution of new mutations to standing variation for quantitative traits[59,60] and understanding genetic constraints on the evolution of natural populations[61].

Evidence is accumulating that pleiotropy is even more pervasive than previously imagined and also occurs between traits that are not thought to be functionally related. Furthermore, the pleiotropic effects of different genes that affect pairs of traits are often not in the same direction and therefore do not result in significant genetic correlations between the traits. Widespread pleiotropy arises by necessity when large numbers of genes affect each trait. In linkage mapping studies, it is difficult to disentangle close linkage from pleiotropy, as the intervals to which QTLs map contain multiple genes. However, pleiotropy can be clearly shown by examining the effects of new mutations on multiple traits and by association mapping in instances in which there is little LD between adjacent genes. In *D. melanogaster*, homozygous viable transposable *P*-element insertional mutations in genes that are known to be essential for development or metabolism also affect adult quantitative traits, and several of the mutant alleles simultaneously affect multiple traits[20]. In mice, exhaustive phenotypic profiling of 250 knockout strains provides similar evidence for widespread and often unexpected pleiotropic effects (see Further information for a link to the Mouse Genome Informatics website). Pleiotropy is also implicated by the results of genome-wide association studies in humans. In several cases, the associations are between diseases that plausibly share a common aetiology and, in others, the shared associations show unexpected relationships between diseases and traits[20].

Epistasis
This occurs when the homozygous or heterozygous effects at one locus differ depending on the genotype of the interacting locus.

Genotype-by-environment interaction
This occurs when the homozygous and heterozygous effects of a locus change in magnitude or direction in different environments.

Pleiotropy at the level of individual genes is hardly surprising. The realization that most genes have multiple functions has motivated geneticists working on model organisms to develop methods for temporal and spatial control of mutant alleles and to create allelic series to parse gene function. The combinatorial possibilities of mutations in even a single gene are enormous and impractical to test systematically. However, analysis of the effects of genetic perturbations on quantitative traits using natural variants that have survived natural selection provides insights into the subfunctionalization of individual genes; that is, which regions of the gene are responsible for effects on multiple traits and which regions are specific for individual traits. Preliminary evidence for the specificity of individual natural polymorphisms comes from *D. melanogaster*, in which the scale of LD is sufficiently fine that polymorphisms in close physical proximity are not correlated, and it is possible to differentiate the effects of molecular polymorphisms in the same gene. In all cases in which polymorphisms in a single gene have been associated with more than one quantitative trait[20], different polymorphic sites were independently associated with the different traits. Thus, genes are pleiotropic, but individual variants are not. In genomic regions with little LD, pervasive pleiotropy does not necessarily impose evolutionary constraints in the form of strong genetic correlations between traits. Finally, pleiotropic effects can themselves be genetically variable when differences in epistatic interactions occur between loci that affect multiple traits[45].

Molecular basis of quantitative variation. Insights into the mechanisms of maintenance of quantitative genetic variation and the evolution of quantitative traits require that we understand the causal molecular variants (quantitative trait nucleotides; QTNs) affecting quantitative traits. The distribution of QTN allele frequencies can indicate the nature of the selective forces operating on the trait. Variation maintained by a balance between the input of new mutations and their removal by natural selection would lead to a frequency distribution that is skewed towards rare minor allele frequencies[62]. QTNs maintained by a balance of selective forces would tend to have a frequency distribution centred around intermediate frequencies[39,60], and the frequency distribution of selectively neutral alleles would span the entire frequency range[63]. Inference of QTN allele frequencies is restricted to association mapping designs in which all of the variants in a candidate gene or gene region have been identified. Otherwise, the polymorphism associated with a trait could be in LD with the true QTN with a different frequency. Evidence from *D. melanogaster*[13,64] and humans[65–69] indicates that rare variants and variants with minor allele frequencies less than 5% are frequently associated with variation in quantitative traits.

Most association mapping studies have used SNPs at intermediate frequencies, owing to the ease of multiplex genotyping of these SNPs and maximizing the power to detect QTLs. However, indels and larger-scale copy number variants might potentially have larger effects on complex traits and tend to be rare. In humans, rare copy number variants are more common in individuals with autism[70] and schizophrenia[71] than in unaffected individuals.

Determining the functional consequences of QTNs addresses the long-standing debate about the relative contribution of protein-coding changes versus regulatory changes in phenotypic evolution[72,73]. Associations of non-synonymous polymorphisms in coding regions with variation in quantitative traits are easy to understand mechanistically[17,18]. However, synonymous polymorphisms in coding regions could be associated with mRNA stability[64,74], and polymorphisms in promoters and introns could affect transcription factor binding and mRNA splicing, and affect the amount, timing or tissue-specific pattern of expression. For example, a QTL with a major effect on the difference in plant architecture between maize and its undomesticated ancestor, teosinte, maps to the *tb1* gene, but a polymorphism that causes the difference lies in a regulatory element 58–69 kb upstream of this gene[75]. Indeed, several replicated associations with human diseases lie in gene deserts far from any annotated gene[20].

Genetic architecture: challenges

The genetic dissection of quantitative traits faces two main challenges: the power to detect and localize QTLs and QTNs, and the biological context in which to place genotype–phenotype associations. Accurate phenotypes and high-density molecular genotypes are needed for many thousands of individuals to map QTLs with effect sizes of the magnitude we now expect, with the high resolution required to separate closely linked QTLs and with the power to detect interactions between QTLs. The implication of widespread pleiotropy is that we cannot accelerate the QTL-mapping end game by selecting likely candidate genes for functional validation, but need to perform unbiased scans for genes that correspond to the QTLs. Pervasive pleiotropy also highlights the fallacy that there are genes 'for' particular traits[76]. The challenge is to catalogue the full range of pleiotropic effects of each gene and to distinguish the QTNs affecting each trait. Superimposed on this challenge is the issue of environment-specific and sex-specific effects, which can only be estimated by repeating the mapping in a range of ecologically and medically relevant environments. Detecting epistatic interactions presents a statistical challenge given the large number of genes that are expected to be associated with any one trait and the expectation that epistasis can occur between QTLs without main effects.

The challenge of dissecting quantitative traits into individual genes and their causal QTNs should be met in the near future by applying new sequencing and genotyping technologies (TABLE 1, Supplementary information S1 (table)) to the problem, in combination with new community resources (BOX 2). However, a list of all genes and QTNs associated with quantitative traits is just that — a list, devoid of biological context. But QTNs do not affect traits directly; they do so through complex networks of transcriptional, protein, metabolic and other molecular endophenotypes. The new challenge is to understand the causative and correlative effects of genetic perturbations on these networks and their downstream effects on organismal phenotypes.

Quantitative trait nucleotide
A causal molecular variant (allele) that affects variation in a quantitative trait.

Endophenotype
An intermediate molecular phenotype associated with an organismal level quantitative trait. Variation in the endophenotype affects variation in the organismal trait.

Table 1 | **Technologies that allow systems genetics of quantitative traits**

Level of variation	Technology	Potential applications
High-throughput sequencing	• Deep sequencing • Sequence capture	• SNP identification and resequencing reference panels • Resequencing quantitative trait locus regions following detection to allow localization using a larger number of individuals; identification of rare alleles
Genotyping	• Genotyping arrays	• Multiplexed high-density markers for quantitative trait locus detection
Whole-genome transcriptional profiling	• cDNA microarrays • Tiling arrays and RNA-seq	• High-throughput, cost-effective examination of genome-wide mRNAs • High-throughput, unbiased examination of genome-wide mRNAs
Proteomics	• Tandem mass spectrophotometry	• Detecting quantitative and qualitative variation in proteins
Metabolomics	• Gas chromatography and high-performance liquid chromotagraphy mass spectrophotometry	• Detecting quantitative and qualitative variation in cellular metabolites
Organismal phenotypes	• Image or video analysis-based phenotyping	• Phenotyping large samples required for systems genetics analyses

Systems genetics: from QTL to biology

Associating DNA sequence variation with variation in organismal phenotypes omits all of the intermediate steps in the chain of causation from genetic perturbation to phenotypic variation. Intermediate molecular phenotypes such as transcript abundance also vary genetically in populations and are themselves quantitative traits[77,78]. 'Genetical genomics' (REF. 79) or systems genetics[11] approaches integrate DNA sequence variation, variation in transcript abundance and other molecular phenotypes and variation in organismal phenotypes in a linkage or association mapping population, and allow us to interpret quantitative genetic variation in terms of biologically meaningful causal networks of correlated transcripts. These approaches have been allowed by the development of massively parallel technologies for quantifying genome-wide levels of transcript abundance (TABLE 1, Supplementary information S1 (table)). The logic of systems genetics is outlined in FIG. 2. In addition to obtaining genotype and organism phenotype data, whole-genome transcript abundance for each individual in a linkage or association mapping population is quantified. As usual, marker–organismal trait associations are performed to map QTLs, but the same association tests are performed between the markers and gene expression traits to map expression quantitative trait loci (eQTLs), and the correlations between gene expression and organismal level phenotypes are determined to identify quantitative trait transcripts (QTTs).

eQTLs. Advances in high-throughput genotyping and transcriptional profiling have facilitated an increase in the number of eQTL studies[11,78,80–82]. Two features distinguish eQTL studies from their traditional predecessors: the number of traits, that is, transcript levels, tends to be much larger than the number of individuals in the study; and unlike organismal phenotypes, transcripts have a local genomic context. If the molecular variant is located within the gene region of the transcript under investigation, the regulation is called a *cis*, proximal or local eQTL, but if the polymorphism associated with variation in the transcript is in another gene, it is called a *trans* or distal eQTL[78]. Common features of most eQTL studies[11,78,80–82] are that large numbers of transcripts are genetically variable; *cis* eQTLs tend to have larger effects than *trans* eQTLs; there tend to be more *cis* than *trans*-acting polymorphisms; some genomic regions are associated with variation in the expression levels of many transcripts (eQTL hot spots); and the expression levels of many transcripts are highly correlated. A further commonality is that the sample sizes of most studies have been restricted to 30–100 individuals or strains, owing to the expense of whole-genome transcript profiling. Nevertheless, the same statistical considerations apply to eQTLs as to more traditional organismal level traits. The fact that these studies are underpowered to detect and localize eQTLs[77] could explain the observation that there are fewer *trans* than *cis* eQTLs, and raises the possibility that many *cis* eQTLs are *trans*-regulated by linked loci. eQTL studies are also plagued by the statistical challenge of the massive number of hypothesis tests required to associate a dense marker map with tens of thousands of transcripts. Few studies have attempted to control the false discovery rate, but instead have chosen to report the number of significant marker–eQTL associations at different significance levels. As the cost of genome-wide expression profiling declines (TABLE 1, Supplementary information S1 (table)), larger eQTL studies will be possible, which will alleviate many of these concerns.

Coexpression networks. Although many thousands of transcripts are genetically variable, they are not independent: the levels of expression of many transcripts co-vary between individuals in the mapping population[11,78,83] (FIGS 2,3). Genetically correlated transcripts might be coexpressed because they belong to a regulatory network, which could provide insights into the underlying biology. Several statistical methods[84,85] have been developed to group genetically correlated transcripts into modules, in which each module consists of a group of transcripts with higher correlations to each other than to the rest of the transcriptome (FIG. 3). The statistical information encoded in highly correlated transcripts is redundant; assembling such genes into modules reduces the number of hypothesis tests one must consider. The correlations between transcripts in a module can be visualized graphically as a network with nodes denoting transcripts and edges connecting nodes

Expression quantitative trait locus
A region of the genome containing one or more genes that affect variation in gene expression, which is identified by linkage to polymorphic marker loci.

Quantitative trait transcript
A transcript for which variation in its expression is correlated with variation in an organismal level quantitative trait phenotype.

Gene ontology
A widely used classification system of gene functions and other gene attributes that uses a controlled vocabulary.

KEGG pathway
The Kyoto Encyclopedia of Genes and Genomes (KEGG) is a database comprising a collection of graphical pathway maps for metabolism, regulatory processes and other biological processes.

that are genetically correlated (FIGS 2,3). In the absence of information on DNA polymorphisms, these networks represent indirect statistical relationships rather than direct interactions. Incorporating information about *cis* and *trans* eQTLs can be used to determine the direction of flow of information through the network and infer which relationships are caused directly by genetic perturbations and which ones are coregulated by genetic perturbations (FIG. 2). Coexpression networks are biologically plausible, as judged by enrichment of gene ontology categories, KEGG pathways, protein–protein interactions, tissue-specific expression patterns and transcription factor-binding sites[83,86,87]. Coexpression networks thus allow the prediction of the function of computationally predicted genes based on 'guilt by association' with well-annotated genes in the network[83,88].

QTTs. One insight into the molecules associated with variation in organismal phenotypes comes from identifying transcripts that are genetically correlated with an organismal quantitative trait[83,89,90]. Typically, several hundred quantitative trait transcripts (QTTs) are associated with any single organismal phenotype[83,89,90]. However, this observation is subject to the caveat described above regarding the low statistical power and high false-positive rate of studies with small sample sizes. QTTs associated with an organismal quantitative trait are also genetically correlated. Thus, one can construct networks relating transcriptional variation to organismal trait variation from modules of correlated QTTs, in the same manner as when networks are derived from transcript data alone (FIG. 3). Module-based approaches have been applied to various organisms to uncover pathways and processes associated with organismal quantitative traits, including Alzheimer's disease in humans[86], mouse models of type 2 diabetes[91] and sleep[90] and several ecologically relevant quantitative traits in *D. melanogaster*[83]. However, causal relationships cannot be inferred from modules of correlated QTTs alone, and also require information from DNA sequence variation.

Box 2 | From quantitative trait loci to quantitative trait nucleotides

The process of detecting and localizing quantitative trait nucleotides (QTNs) will be accelerated by full DNA sequences and comprehensive phenotypic descriptions of tens of thousands of individuals in linkage or association mapping populations. Next-generation sequencing technologies (TABLE 1, Supplementary information S1 (table)) are bringing this scenario closer to reality. Community projects are ongoing to determine the whole-genome sequence of 1,000 humans (the 1000 Genomes Project), 1,001 *Arabidopsis* strains[107] (the 1001 Genomes Project) and 192 *Drosophila* strains. The human project follows the HapMap model[108], with the goal of cataloguing SNPs and structural variants with frequencies of at least 0.5–1%, but with no phenotypic information. An immediate application of these data will be in studies testing for the functional effects of variants in regions that have previously been identified to affect disease risk. The *Arabidopsis* and *Drosophila* species projects are intended to allow genome-wide association studies.

Most linkage mapping studies are limited by the lack of dense polymorphic molecular marker maps. Next-generation sequencing methods empower individual laboratories to sequence parental strains of interest to rapidly identify marker loci and design custom genotyping arrays (TABLE 1, Supplementary information S1 (table)). Short-oligonucleotide arrays that represent the whole genome are efficient platforms for obtaining dense molecular marker maps[109] but the cost of genotyping thousands or more individuals is usually prohibitive. Strategies to overcome this limitation include genotyping individuals[110] or pools of individuals[111] in the extreme tails of the phenotypic distribution of the mapping population. Although the simultaneous detection and localization of QTLs may not become routine in the near future, the simultaneous dissection of multiple QTL regions identified by an initial genome scan in large populations is now possible. Methods for multiplexed capture of targeted sequences, before the application of next-generation sequencing technologies to only these targeted sequences[112–115], offer promise for rapidly identifying causal genes and variants associated with quantitative traits[116].

Understanding pleiotropy necessitates that tests of association to multiple phenotypes (including the same phenotype in multiple environments and both sexes[1]) are performed for the same genotypes. In model organisms, this is most conveniently achieved by community sharing of common inbred genetic reference panels, which only need to be genotyped once. First-generation mapping populations of recombinant inbred lines of mice[117], *A. thaliana*[118] and *D. melanogaster*[34–37] have shown the success of this approach. However, the small size of these reference populations gave low-resolution recombination maps and only two parental genomes were used to establish these maps. The second-generation projects address both deficiencies of the first-generation projects. The recombinant inbred line strategy has been adopted by the mouse Collaborative Cross Consortium[119] and the Maize Diversity Project[120]; sequencing the parental strains will yield a dense polymorphism map for genotyping. The *Arabidopsis* and *Drosophila* projects rely on association mapping and next-generation sequencing of all strains. Crosses between the inbred lines of each reference population extend the genotypic and phenotypic space that can be explored and directly address whether dominance and epistatic effects are important. In humans, large prospective cohort studies (for example, the UK Biobank and Framingham Heart Study) for which detailed phenotypic measurements and medical histories are obtained serve as reference panels. The challenge of genetic dissection in such 'natural' settings is to accurately measure and account for the direct and interaction effects of relevant environmental exposures on the traits[121].

Large-scale mapping of genotype to phenotype associations depends on accurate measurements of phenotype and strict attention to the principles of experimental design to avoid confounding of phenotypes and common environmental conditions[1]. In the future, the limitation of identifying genotype–phenotype associations will shift from assessing multilocus genotypes towards obtaining accurate, multidimensional phenotypes for large numbers of individuals. The development of high-throughput methods for automated phenotyping will be highly beneficial, especially in model organisms[122,123] (TABLE 1, Supplementary information S1 (table)).

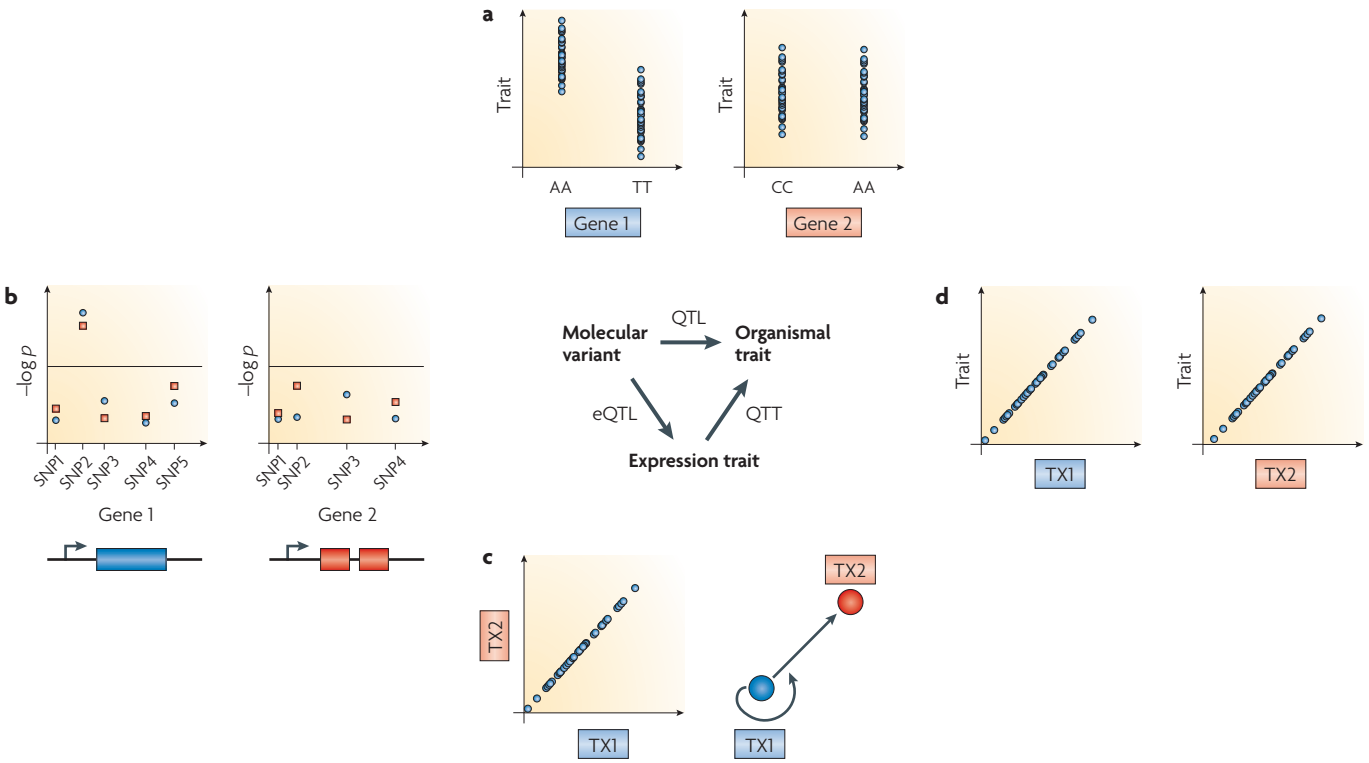


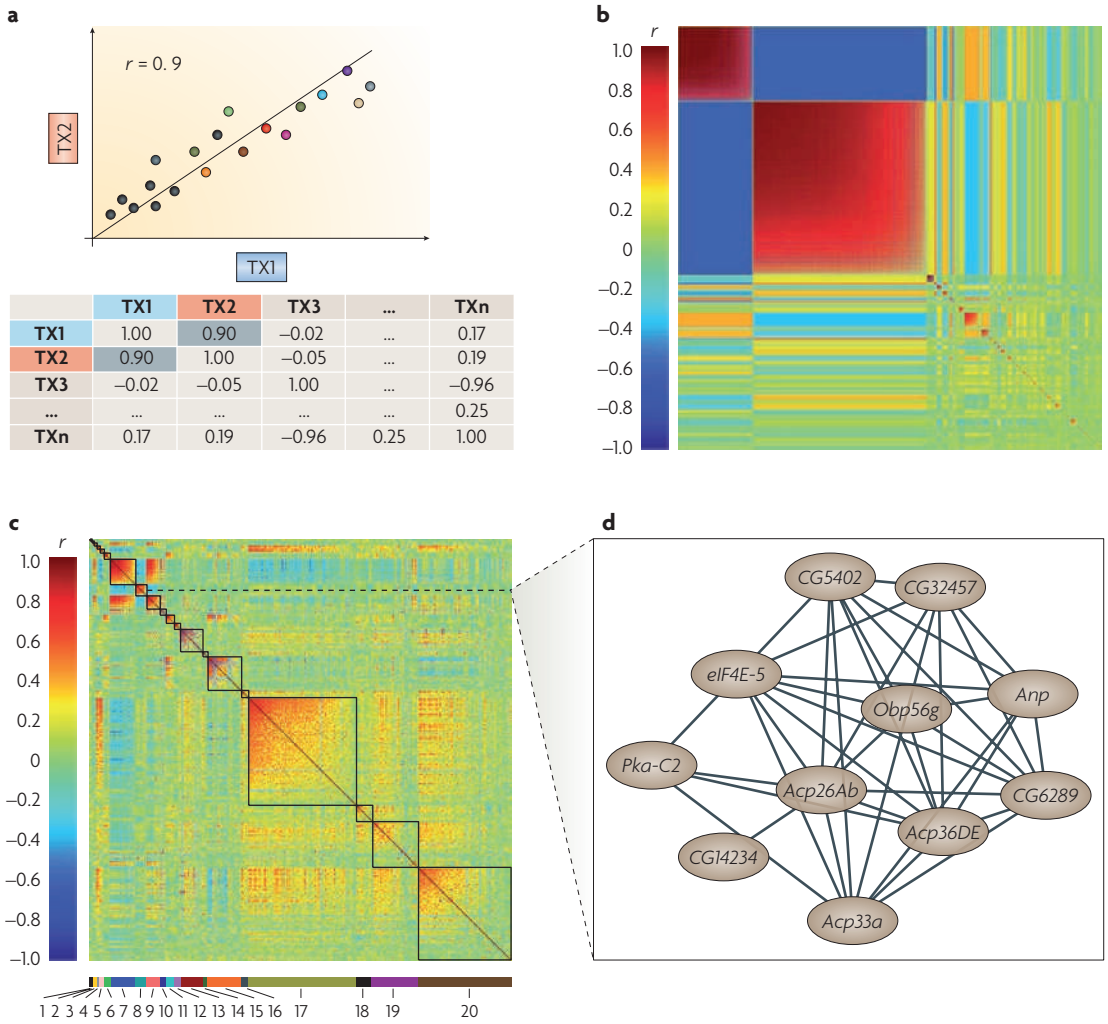
Figure 2 | **Systems genetics integrative framework.** The relationship between DNA sequence variation and quantitative variation for gene expression and an organismal phenotype is given for two genes. **a** | Values of the organismal phenotype are shown for several individuals that are homozygous at two SNPs (AA or TT in gene 1 and CC or AA in gene 2). The mean difference in the phenotype between the AA and TT genotypes identifies gene 1 as a quantitative trait locus (QTL) for the trait. **b** | An expression QTL (eQTL) mapping study in which SNPs in gene 1 and gene 2 are tested for association with the variation in transcript abundance of gene 1 (TX1) and gene 2 (TX2). The schematics below the x axes denote gene structure, with boxes indicating exons and lines non-coding regions. Circles indicate the result of an association test (scaled as –log *p*) and horizontal lines denote the significance threshold. SNP2 in the promoter of gene 1 is significantly associated with TX1 and TX2 abundance and there is no association with gene 2. SNP2 is a *cis*-regulatory polymorphism for TX1 and a *trans*-regulatory polymorphism for TX2. **c** | TX1 and TX2 are correlated, suggesting that the transcripts act as two nodes in a network. Combined with the information gained from the eQTL mapping, we can infer the direction in this network. SNP2 in gene 1 controls variation in the expression level of gene 2. **d** | Association between TX1 and TX2 abundance and the organismal phenotype. TX1 and TX2 are quantitative trait transcripts (QTTs).

Bayesian network
A graph with directed edges that connect nodes. The nodes represent assertions about relationships between the nodes; for example, a node A is related to a node B by an edge that represents that A is a cause of B with a certain probability. Bayesian networks with many interconnected nodes can be constructed.

Partial correlation analysis
This quantifies the association between a pair of variables after controlling for the effect of a set of potential confounders.

Empirical Bayes procedure
A hierarchical model in which the hyperparameter is not a random variable but is estimated by some other (often classical) means.

Systems genetics of complex traits. QTNs allow us to map phenotype to genotype in the absence of biological context. To gain this context, we need to describe the flow of information from DNA to the organismal phenotype through RNA intermediates, proteins, metabolites and other molecular endophenotypes. This could be achieved by dissecting a QTN into its constituent eQTLs and QTTs, but in practice this is not easy; one can expect a substantial coexpression network of relevant transcripts that associate both with the molecular variant and with the organismal phenotype[92]. Systems genetics promises to integrate these layers of information to produce directed biological networks that link molecular variants to organismal phenotypes. The systems genetics approach, in conjunction with community resources, will allow us to explore the space of possible genotypes with large sample sizes; nevertheless, sophisticated techniques are needed to construct meaningful networks from natural genetic perturbations. The sophistication that is required is based on subtlety: of the transcripts that associate both with sequence and trait, only a subset will be causal. Experimentally, a causal transcript is identified when directly perturbing it leads to variation in the organismal phenotype. By contrast, a consequence of observing natural genetic perturbations is that non-causal associations between a transcript and phenotype can be driven by upstream transcripts, which are the true causative transcripts. A fraction of QTTs will be consequential associations, the perturbation of which will not affect the organismal phenotype. Disentangling causal relationships from consequential relationships is the key to reconstructing biological networks, and the principal tool to study these relationships is the statistical concept of conditional dependence[93] (FIG. 4). Current techniques for causal inference include Bayesian networks[94], partial correlation analysis[95] and empirical Bayes procedures[96]; several of these are available as packages in the statistical programming language 'R'. The power of the systems genetics approach for understanding the biological basis of variation for quantitative traits and



	TX1	TX2	TX3	...	TXn
TX1	1.00	0.90	–0.02	...	0.17
TX2	0.90	1.00	–0.05	...	0.19
TX3	–0.02	–0.05	1.00	...	–0.96
...	...	...	...	...	0.25
TXn	0.17	0.19	–0.96	0.25	1.00

Figure 3 | **Coexpression gene networks. a** | The first step in building coexpression gene networks is to calculate a pairwise correlation matrix between all variable transcripts. Here we show a representation of such a matrix and the correlation between the abundances of two transcripts, TX1 and TX2. The scatterplot shows the mean expression of these two transcripts for different genotypes in a mapping population. **b** | An extensive coexpression network in 40 inbred *Drosophila melanogaster* lines[83]. The 10,096 genetically variable transcripts formed 241 modules and in each module the transcripts are more closely correlated with each other than to the other transcripts. **c** | There were 414 quantitative trait transcripts for competitive fitness in these lines, which formed 20 modules of correlated transcripts[83]. **d** | Network representation of fitness module 8. Each node in this network corresponds to a transcript, and an absolute correlation threshold of 0.6 was used for the edges connecting each node. *Acp*, accessory gland protein; *Anp*, andropin; *eIF4E-5*, eukaryotic initiation factor 4E-5; *Obp56g*, odorant-binding protein 56g; *Pka-C2*, cAMP-dependent protein kinase 2. Part **b** reproduced, with permission, from Nature Genetics REF. 83 © (2009) Macmillan Publishers Ltd. All rights reserved. Part **c** is modified from REF. 85.

the potential for evolutionary conservation of these networks is shown by two recent studies of body weight and obesity in mice[92] and humans[97]. In both studies, body weight was correlated with the same directed coexpression gene network.

Statistically defining the molecular interactions that govern phenotypic variation through natural genetic perturbations leads to tests of the models. One prediction that has been confirmed in several studies is that causal genes in the network will affect the trait when perturbed by an induced mutation[90,92,98,99]. A higher-level test is to determine whether the genomic effects of a new mutant allele are as predicted by the network; that is, whether transcripts downstream of the focal gene will be altered in the background of the mutant allele and whether transcripts unconnected with that gene will not be affected[90]. The biological networks provide a framework for targeted testing of epistatic interactions, with the prediction that genetically correlated transcripts that are also regulated by *cis* eQTLs in a module associated with the organismal trait fit the criteria for potential epistasis. Reducing the statistical penalty for testing all possible pairwise interactions should improve the power to detect two-way and higher-order epistasis. Finally,

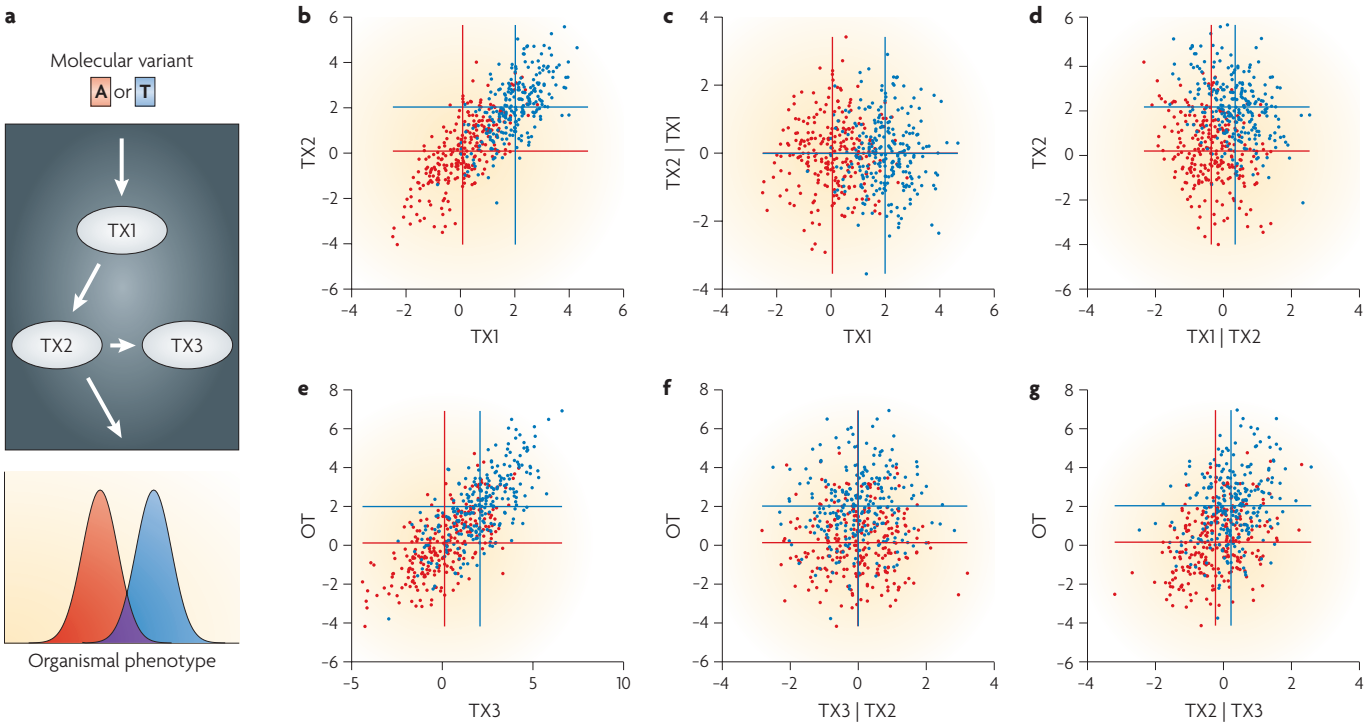


Figure 4 | **Inside the black box between genotype and phenotype.** A biological network encompassing a molecular variant, three transcripts (TX1–3) and an organismal phenotype (OT) is shown. **a** | A molecular variant upstream of the gene that encodes TX1 causes variation in TX1 abundance. TX1 increases the transcription of TX2, which in turn increases the transcription of TX3. Variation in TX2 is also the proximal cause of variation in the organismal phenotype. The molecular variant is a quantitative trait nucleotide with respect to the organismal phenotype and an expression quantitative trait locus for TX1 (*cis*), TX2 (*trans*) and TX3 (*trans*). The three transcripts form a coexpression network and each may manifest as a quantitative trait transcript; however, TX3 is a consequential rather than causal transcript. Data consistent with the relationships in panel **a** are shown in panels **b**–**g**. Each plot shows 500 individuals; both states of the molecular variant are equally represented. **b** | Linear relationship between the abundance of transcripts TX1 and TX2 (red shows molecular variant A and blue shows molecular variant T). The vertical and horizontal lines indicate the mean abundance of TX1 and TX2, respectively. In both cases, the T variant on average leads to higher values. **c** | Because the influence of the molecular variant on TX2 is realized though the variation it induces in TX1, after conditioning on TX1 (that is, statistically accounting for the effect of TX1), the variant-specific difference in TX2 abundance vanishes. **d** | Conversely, conditioning on TX2 does not eliminate the relationship between the molecular variant and TX1 abundance. Panels **b** – **d** thus support the order of TX1 upstream of TX2. **e** | Relationship between TX3 and the organismal phenotype. Despite the absence of a causal relationship, TX3 abundance is strongly correlated with the organismal phenotype. As the variables on both axes are downstream of the molecular variant, both show variant-specific differences. **f** | Conditioning on TX2 uncouples TX3 from both the molecular variant and from the organismal trait. **g** | Conversely, conditioning TX3 on TX2 does not eliminate the relationships between the molecular variant, TX2 abundance and the organismal phenotype. Data shown in panels **b**–**g** inspired by REF. 93.

assessing many different organismal level phenotypes for the same genotypes will provide insights into the molecular basis of pleiotropy[83].

The future of complex trait genetics

New technologies will allow us to perform genotype–phenotype mapping studies on the scale and with the density of molecular markers required to simultaneously identify many genes that affect variation for quantitative traits. We are beginning to interpret these associations in terms of genetic networks through the incorporation of information on whole-genome variation in transcript abundance in the mapping populations. This provides unprecedented insights into the biological underpinnings of complex traits and the pleiotropic connections between traits. As technologies advance and costs fall, we could perform systems genetics analyses on larger samples, more developmental time points and tissues, more environmental conditions and with an unbiased sample of the transcriptome. Not all functional molecular polymorphisms exert their effects on organismal traits through measurable alterations in gene expression. Adding information on qualitative and quantitative variation in proteins and metabolites, as well as epigenetic modifications, will give a more complete picture of the effects of genetic perturbations on the whole organism[100] (TABLE 1, Supplementary information S1 (table)). Profiting from the forthcoming avalanche of data will require interactive databases that can organize and integrate these heterogeneous

sources of information. New statistical methodology will be needed to exploit these databases and infer biological networks with confidence. As new technologies become increasingly affordable and widely available, all organisms will become model organisms, enabling us to understand the genetic basis of wide-ranging ecological specializations and adaptations. Finally, we now have the tools to address long-standing unanswered questions in evolutionary quantitative genetics, including, but not limited to, the mechanisms maintaining quantitative genetic variation[60,101], the polygenic mutation rate[101], the molecular basis of GEIs[102], the cause of limits to long-term selection[103] and phenotypic stability in the face of genetic and environmental variation[104].

1. Falconer, D. S. & Mackay, T. F. C. *Introduction to Quantitative Genetics* (Addison Wesley Longman, Harlow, 1996).
2. Lynch, M. & Walsh, B. *Genetics and Analysis of Quantitative Traits* (Sinauer Associates, Sunderland, Massachusetts, 1998).
3. Sax, K. The association of size differences with seed-coat pattern and pigmentation in *Phaseolus vulgaris*. *Genetics* **8**, 522–560 (1923).
4. Thoday, J. M. in *Quantitative Genetic Variation* (eds Thompson, J. N. Jr & Thoday, J. M.) 219–233 (Academic, New York, 1979).
5. Shrimpton, A. E. & Robertson, A. The isolation of polygenic factors controlling bristle score in *Drosophila melanogaster*. II. Distribution of third chromosome bristle effects within chromosome sections. *Genetics* **118**, 445–459 (1988).
6. Lander, E. S. & Botstein, D. Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* **121**, 185–199 (1989). **This seminal paper describes a statistical method for interval mapping of QTLs using molecular markers and provides the first power calculations for these studies.**
7. Altshuler, D., Daly, M. J. & Lander, E. S. Genetic mapping in human disease. *Science* **322**, 881–888 (2008).
8. Donnelly, P. Progress and challenges in genome-wide association studies in humans. *Nature* **456**, 728–731 (2008).
9. Frazer, K. A., Murray, S. S., Schork, N. J. & Topol, E. J. Human genetic variation and its contribution to complex traits. *Nature Rev. Genet.* **10**, 241–251 (2009).
10. Mackay, T. F. C. The genetic architecture of quantitative traits. *Annu. Rev. Genet.* **35**, 303–339 (2001).
11. Sieberts, S. K. & Schadt, E. E. Moving toward a system genetics view of disease. *Mamm. Genome* **18**, 389–401 (2007). **This is an excellent overview of systems genetics and the methods used to infer causal gene networks.**
12. Sutter, N. B. & Ostrander, E. A. Dog star rising: the canine genetic system. *Nature Rev. Genet.* **5**, 900–910 (2004).
13. Carbone, M. A. *et al.* Phenotypic variation and natural selection at *Catsup*, a pleiotropic quantitative trait gene in *Drosophila*. *Curr. Biol.* **16**, 912–919 (2006).
14. Steinmetz, L. M. & Davis, R. W. Maximizing the potential of functional genomics. *Nature Rev. Genet.* **5**, 190–201 (2004).
15. Colosimo, P. F. *et al.* The genetic architecture of parallel armor plate reduction in threespine sticklebacks. *PLoS Biol.* **2**, e109 (2004).
16. Shapiro, M. D. *et al.* Genetic and developmental basis of evolutionary pelvic reduction in threespine sticklebacks. *Nature* **428**, 717–723 (2004).
17. Hoekstra, H. E., Hirschmann, R. J., Bundey, R. A., Insel, P. A. & Crossland, J. P. A single amino acid mutation contributes to adaptive beach mouse color pattern. *Science* **313**, 101–104 (2006).
18. Mosher, D. S. *et al.* A mutation in the myostatin gene increases muscle mass and enhances racing performance in heterozygote dogs. *PLoS Genet.* **3**, e79 (2007).
19. Beavis, W. D. in *49th Annual Corn and Sorghum Research Conference* 252–268 (American Seed Trade Association, Washington DC, 1994).
20. Flint, J. & and Mackay, T. F. C. Genetic architecture of quantitative traits in flies, mice and humans. *Genome Res.* **19**, 723–733 (2009). **This is a recent review of the results of studies mapping QTLs in humans and two genetic model organisms.**
21. Kroymann, J. & Mitchell-Olds, T. Epistasis and balanced polymorphism influencing complex trait variation. *Nature* **435**, 95–98 (2005).
22. Steinmetz, L. M. *et al.* Dissecting the architecture of a quantitative trait locus in yeast. *Nature* **416**, 326–330 (2002).
23. Long, A. D. *et al.* High resolution mapping of genetic factors affecting abdominal bristle number in *Drosophila melanogaster*. *Genetics* **139**, 1273–1291 (1995).
24. Dilda, C. L. & Mackay, T. F. C. The genetic architecture of *Drosophila* sensory bristle number. *Genetics* **162**, 1655–1674 (2002).
25. Zimmerman, E., Palsson, A. & Gibson, G. Quantitative trait loci affecting components of wing shape in *Drosophila melanogaster*. *Genetics* **155**, 671–683 (2000).
26. Mezey, J. G., Houle, D. & Nuzhdin, S. V. Naturally segregating quantitative trait loci affecting wing shape of *Drosophila melanogaster*. *Genetics* **169**, 2101–2113 (2005).
27. Long, A. D., Lyman, R. F., Langley, C. H. & Mackay, T. F. C. Two sites in the *Delta* gene region contribute to naturally occurring variation in bristle number in *Drosophila melanogaster*. *Genetics* **149**, 999–1017 (1998).
28. Robin, C., Lyman, R. F., Long, A. D., Langley, C. H. & Mackay, T. F. C. *hairy*: a quantitative trait locus for *Drosophila* bristle number. *Genetics* **162**, 155–164 (2002).
29. Palsson, A. & Gibson, G. Association between nucleotide variation in *Egfr* and wing shape in *Drosophila melanogaster*. *Genetics* **167**, 1187–1198 (2004).
30. Macdonald, S. J. & Long, A. D. Joint estimates of quantitative trait locus effect and frequency using synthetic recombinant populations of *Drosophila melanogaster*. *Genetics* **176**, 1261–1281 (2007).
31. Pasyukova, E. G., Vieira, C. & Mackay, T. F. C. Deficiency mapping of quantitative trait loci affecting longevity in *Drosophila melanogaster*. *Genetics* **156**, 1129–1146 (2000).
32. De Luca, M. *et al.* Dopa decarboxylase (*Ddc*) affects variation in *Drosophila* longevity. *Nature Genet.* **34**, 429–433 (2003).
33. Mackay, T. F. C., Roshina, N. V., Leips, J. W. & Pasyukova, E. G. in *Handbook of the Biology of Aging* 6th edn (eds Masaro, E. J. & Austad, S. N.) 181–216 (Academic, New York, 2006).
34. Harbison, S. T., Yamamoto, A. H., Fanara, J. J., Norga, K. K. & Mackay, T. F. C. Quantitative trait loci affecting starvation resistance in *Drosophila melanogaster*. *Genetics* **166**, 1807–1823 (2004).
35. Moehring, A. J. & Mackay, T. F. C. The quantitative genetic basis of male mating behavior in *Drosophila melanogaster*. *Genetics* **167**, 1249–1263 (2004).
36. Fanara, J. J., Robinson, K. O., Rollmann, S., Anholt, R. R. H. & Mackay, T. F. C. *Vanaso* is a quantitative trait locus for *Drosophila* olfactory behavior. *Genetics* **162**, 1321–1328 (2002).
37. Jordan, K. W. & Mackay, T. F. C. Quantitative trait loci for locomotor behavior in *Drosophila melanogaster*. *Genetics* **174**, 271–284 (2006).
38. Phillips, P. C. Epistasis — the essential role of gene interactions in the structure and evolution of genetic systems. *Nature Rev. Genet.* **9**, 855–867 (2008). **This is a comprehensive review of epistasis and its role in determining genetic pathways, effects on complex traits and diseases, and evolution.**
39. Levene, H. Genetic equilibrium when more than one ecological niche is available. *Am. Nat.* **87**, 331–333 (1953).
40. Gillespie, J. H. & Turelli, M. Genotype–environment interactions and the maintenance of polygenic variation. *Genetics* **121**, 129–138 (1989).
41. Clark, A. G. & Wang, L. Epistasis in measured genotypes: *Drosophila P*-element insertions. *Genetics* **147**, 157–163 (1997).
42. Fedorowicz, G. M., Fry, J. D., Anholt, R. R. H. & Mackay, T. F. C. Epistatic interactions between *smell-impaired* loci in *Drosophila melanogaster*. *Genetics* **148**, 1885–1891 (1998).
43. Sambandan, D., Yamamoto, A., Fanara, J. J., Mackay, T. F. C. & Anholt, R. R. H. Dynamic genetic interactions determine odor-guided behavior in *Drosophila melanogaster*. *Genetics* **174**, 1349–1363 (2006).
44. van Swinderen, B. & Greenspan, R. J. Flexibility in a gene network affecting a simple behavior in *Drosophila melanogaster*. *Genetics* **169**, 2151–2163 (2005).
45. Yamamoto, A. *et al.* Neurogenetic networks for startle-induced locomotion in *Drosophila melanogaster*. *Proc. Natl Acad. Sci. USA* **105**, 12393–12398 (2008).
46. Mackay, T. F. C. & Anholt, R. R. H. Of flies and man: *Drosophila* as a model for human complex traits. *Ann. Rev. Genomics Hum. Genet.* **7**, 339–367 (2006).
47. Carlborg, O. *et al.* A global search reveals epistatic interaction between QTL for early growth in the chicken. *Genome Res.* **13**, 413–421 (2003).
48. Sinha, H. *et al.* Sequential elimination of major-effect contributors identifies additional quantitative trait loci conditioning high-temperature growth in yeast. *Genetics* **180**, 1661–1670 (2008).
49. Stam, L. F. & Laurie, C. C. Molecular dissection of a major gene effect on a quantitative trait: the level of alcohol dehydrogenase expression in *Drosophila melanogaster*. *Genetics* **144**, 1559–1564 (1996). **This classic paper uses *P*-element transformation to show that three separate molecular polymorphisms in alcohol dehydrogenase affect the level of gene expression, that these polymorphisms interact epistatically and that the polymorphisms have opposite effects on expression.**
50. Valdar, W. *et al.* Genetic and environmental effects on complex traits in mice. *Genetics* **174**, 959–984 (2006).
51. Crabbe, J. C. *et al.* Genetics of mouse behavior: interactions with laboratory environment. *Science* **284**, 1670–1672 (1999).
52. Wahlsten, D. *et al.* Stability of inbred mouse strain differences in behavior and brain size between laboratories and across decades. *Proc. Natl Acad. Sci. USA* **103**, 16364–16369 (2006).
53. Caspi, A. *et al.* Role of genotype in the cycle of violence in maltreated children. *Science* **297**, 851–854 (2002).
54. Caspi, A. *et al.* Influence of life stress on depression: moderation by a polymorphism in the 5-HTT gene. *Science* **301**, 386–389 (2003).
55. Nuzhdin, S. V., Pasyukova, E. G., Dilda, C. & Mackay, T. F. C. Sex-specific quantitative trait loci affecting longevity in *Drosophila melanogaster*. *Proc. Natl Acad. Sci. USA* **94**, 9734–9739 (1997).
56. Kenney-Hunt, J. P. *et al.* Pleiotropic patterns of quantitative trait loci for 70 murine skeletal traits. *Genetics* **178**, 2275–2288 (2008).
57. Farber, C. R. & Medrano, J. F. Fine mapping reveals sex bias in quantitative trait loci affecting growth, skeletal size and obesity-related traits on mouse chromosomes 2 and 11. *Genetics* **175**, 349–360 (2007).
58. Ober, C., Loisel, D. A. & Gilad, Y. Sex-specific genetic architecture of human disease. *Nature Rev. Genet.* **9**, 911–922 (2008).
59. Barton, N. H. & Keightley, P. D. Understanding quantitative genetic variation. *Nature Rev. Genet.* **3**, 11–21 (2002). **This is an excellent review of QTL mapping principles and the evolutionary importance of QTLs.**
60. Mitchell-Olds, T., Willis, J. H. & Goldstein, D. B. Which evolutionary processes influence natural genetic variation for phenotypic traits? *Nature Rev. Genet.* **8**, 845–856 (2007).

61. Roff, D. A. & Fairbairn, D. J. The evolution of trade-offs: where are we? *J. Evol. Biol.* **20**, 433–447 (2007).
62. Turelli, M. Heritable genetic variation via mutation–selection balance: Lerch's ζ meets the abdominal bristle. *Theor. Popul. Biol.* **25**, 138–193 (1984).
63. Watterson, G. A. On the number of segregating sites in genetical models without recombination. *Theor. Popul. Biol.* **7**, 256–276 (1975).
64. Wang, P., Lyman, R. F., Shabalina, S. A., Mackay, T. F. C. & Anholt, R. R. H. Functional evolution of odorant binding proteins in *Drosophila melanogaster*. *Genetics* **177**, 1655–1665 (2007).
65. Cohen, J. C. *et al.* Multiple rare alleles contribute to low plasma levels of HDL cholesterol. *Science* **305**, 869–872 (2004).
66. Cohen, J. C. *et al.* Low LDL cholesterol in individuals of African descent resulting from frequent nonsense mutations in *PCSK9*. *Nature Genet.* **37**, 161–165 (2005).
67. Kotowski, I. K. *et al.* A spectrum of *PCSK9* alleles contributes to plasma levels of low-density lipoprotein cholesterol. *Am. J. Hum. Genet.* **78**, 410–422 (2006).
68. Romeo, S. *et al.* Population-based resequencing of *ANGPTL4* uncovers variations that reduce triglycerides and increase HDL. *Nature Genet.* **39**, 513–516 (2007).
69. Ji, W. *et al.* Rare independent mutations in renal salt handling genes contribute to blood pressure variation. *Nature Genet.* **40**, 592–599 (2008).
70. Sebat, J. *et al.* Strong association of *de novo* copy number mutations with autism. *Science* **316**, 445–449 (2007).
71. International Schizophrenia Consortium. Rare chromosomal deletions and duplications increase risk of schizophrenia. *Nature* **455**, 237–241 (2008).
72. King, M. C. & Wilson, A. C. Evolution at two levels in humans and chimpanzees. *Science* **188**, 107–116 (1975).
73. Hoekstra, H. E. & Coyne, J. A. The locus of evolution: evo devo and the genetics of adaptation. *Evolution* **61**, 995–1016 (2007).
74. Nackley, A. G. *et al.* Human catechol-*O*-methyltransferase haplotypes modulate protein expression by altering mRNA secondary structure. *Science* **314**, 1930–1933 (2006).
75. Clark, R. M., Wagler, T. N., Quijada, P. & Doebley, J. A distant upstream enhancer at the maize domestication gene *tb1* has pleiotropic effects on plant and inflorescent architecture. *Nature Genet.* **38**, 594–597 (2006).
76. Buchanan, A. V., Sholtis, S., Richtsmeier, J. & Weiss, K. M. What are genes "for" or where are traits "from"? What is the question? *Bioessays* **31**, 198–208 (2009).
77. de Koning, D. J. & Haley, C. S. Genetical genomics in humans and model organisms. *Trends Genet.* **21**, 377–381 (2005).
78. Rockman, M. V. & Kruglyak, L. Genetics of global gene expression. *Nature Rev. Genet.* **7**, 862–872 (2006).
79. Jansen, R. C. & Nap, J. P. Genetical genomics: the added value from segregation. *Trends Genet.* **17**, 388–391 (2001).
This is the first paper to articulate the concept of eQTL mapping.
80. Hansen, B. G., Halkier, B. A. & Kliebenstein, D. J. Identifying the molecular basis of QTLs: eQTLs add a new dimension. *Trends Plant Sci.* **13**, 72–77 (2008).
81. Gilad, Y., Rifkin, S. A. & Pritchard, J. K. Revealing the architecture of gene regulation: the promise of eQTL studies. *Trends Genet.* **24**, 408–415 (2008).
82. Cookson, W., Liang, L., Abecasis, G., Moffatt, M. & Lathrop, M. Mapping complex disease traits with global gene expression. *Nature Rev. Genet.* **10**, 184–194 (2009).
83. Ayroles, J. F. *et al.* Systems genetics of complex traits in *Drosophila melanogaster*. *Nature Genet.* **41**, 299–307 (2009).
84. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559 (2008).
85. Stone, E. A. & Ayroles, J. F. Modulated modularity clustering as an exploratory tool for functional genomic inference. *PLoS Genet.* **5**, e1000479 (2009).
86. Miller, J. A., Oldham, M. C. & Geschwind, D. H. A systems level analysis of transcriptional changes in Alzheimer's disease and normal aging. *J. Neurosci.* **28**, 1410–1420 (2008).
87. Swindell, W. R. Genes regulated by caloric restriction have unique roles within transcriptional networks. *Mech. Ageing Dev.* **129**, 580–592 (2008).
88. Tian, W. *et al.* Combining guilt-by-association and guilt-by-profiling to predict *Saccharomyces cerevisiae* gene function. *Genome Biol.* **9**, S7 (2008).
89. 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).
90. Harbison, S. T. *et al.* Co-regulated transcriptional networks contribute to natural genetic variation in *Drosophila* sleep. *Nature Genet.* **41**, 371–375 (2009).
91. Keller, M. P. Choi, Y. & Wang, P. A gene expression network model of type 2 diabetes links cell cycle regulation in islets with diabetes susceptibility. *Genome Res.* **18**, 706–716 (2008).
92. Chen, Y. *et al.* Variations in DNA elucidate molecular networks that cause disease. *Nature* **452**, 429–435 (2008).
This paper uses systems genetics to identify gene networks in mice that are statistically causally related to disease traits associated with metabolic syndrome.
93. Rockman, M. V. Reverse engineering the genotype–phenotype map with natural genetic variation. *Nature* **456**, 738–744 (2008).
This is an excellent exposition of systems genetics and the methods for inferring causal transcriptional networks associated with organismal phenotypes.
94. Zhu, J. *et al.* An integrative genomics approach to the reconstruction of gene networks in segregating populations. *Cytogenet. Genome Res.* **105**, 363–374 (2004).
95. Aten, J. E., Fuller, T. F., Lusis, A. J. & Horvath, S. Using genetic markers to orient the edges in quantitative trait networks: the NEO software. *BMC Syst. Biol.* **2**, 34 (2008).
96. Chen, L. S., Emmert-Streib, F. & Storey, J. D. Harnessing naturally randomized transcription to infer regulatory relationships among genes. *Genome Biol.* **8**, R219 (2007).
97. Emilsson, V. *et al.* Genetics of gene expression and its effect on disease. *Nature* **452**, 423–428 (2008).
This paper uses systems genetics to identify a genetic network associated with human obesity; this network included a module enriched for genes involved in the inflammatory and immune response that was also found in the mouse study described in Reference 92.
98. Brem, R. B., Yvert, G., Clinton, R. & Kruglyak, L. Genetic dissection of transcriptional regulation in budding yeast. *Science* **296**, 752–755 (2002).
99. Yang, X. *et al.* Validation of candidate causal genes for obesity that affect shared metabolic pathways and networks. *Nature Genet.* **41**, 415–423 (2005).
100. Fu, J., *et. al.* System-wide molecular evidence for phenotypic buffering in *Arabidopsis*. *Nature Genet.* **41**, 166–167 (2009).
101. Barton, N. H. & Turelli, M. Evolutionary quantitative genetics: how little do we know? *Annu. Rev. Genet.* **23**, 337–370 (1989).
102. Via, S. *et al.* Adaptive phenotypic plasticity: consensus and controversy. *Trends Ecol. Evol.* **10**, 212–217 (1995).
103. Clayton, G. A. & Robertson, A. An experimental check on quantitative genetical theory. II. The long term effects of selection. *J. Genet.* **55**, 152–170 (1957).
104. Waddington, C. H. Canalization of development and genetic assimilation of acquired characters. *Nature* **183**, 1654–1655 (1959).
105. Zeng, Z. B., Kao, C. H. & Basten, C. J. Estimating the genetic architecture of quantitative traits. *Genet. Res.* **74**, 279–289 (1999).
106. Zou, W. & Zeng, Z. B. Statistical methods for mapping multiple QTL. *Int. J. Plant Genomics* **2008**, 286561 (2008).
107. Ossowski, S. *et al.* Sequencing of natural strains of *Arabidopsis thaliana* with short reads. *Genome Res.* **18**, 2024–2033 (2008).
108. The International HapMap Consortium. A second generation human haplotype map of over 3.1 million SNPs. *Nature* **449**, 851–861 (2007).
109. Gilad, Y. & Borevitz, J. Using DNA microarrays to study natural variation. *Curr. Opin. Genet. Dev.* **16**, 553–558 (2006).
110. Darvasi, A. & Soller, M. Selective genotyping for determination of linkage between a marker locus and a quantitative trait locus. *Theor. Appl. Genet.* **85**, 353–359 (1992).
111. Lai, C. Q. *et al.* Speed-mapping quantitative trait loci using microarrays. *Nature Methods* **10**, 839–841 (2007).
112. Albert, T. J. *et al.* Direct selection of human genomic loci by microarray hybridization. *Nature Methods* **4**, 903–905 (2007).
113. Porreca, G. J. *et al.* Multiplex amplification of large sets of human exons. *Nature Methods* **4**, 931–936 (2007).
114. Okou, D. T. *et al.* Microarray-based genomic selection for high-throughput resequencing. *Nature Methods* **4**, 907–909 (2007).
115. Turner, E. H., Lee, C., Ng, S. B., Nickerson, D. A. & Shendure, J. Massively parallel exon capture and library-free resequencing across 16 genomes. *Nature Methods* 6 Apr 2009 (doi: 10.1038/nmeth.f.248).
116. Nejentsev, S., Walker, N., Riches, D., Egholm, M. & Todd, J. A. Rare variants of *IFIH1*, a gene implicated in antiviral responses, protect against type 1 diabetes. *Science* **324**, 387–389 (2009).
117. Williams, R. W., Gu, J., Qi, S. &, Lu, L. The genetic structure of recombinant inbred mice: high-resolution consensus maps for complex trait analysis. *Genome Biol.* **2**, research0046.1–research0046.18 (2001).
118. Lister, C. & Dean, C. Recombinant inbred lines for mapping RFLP and phenotypic markers in *A. thaliana*. *Plant J.* **4**, 745–750 (1993).
119. Churchill, G. A. *et al.* The Collaborative Cross, a community resource for the genetic analysis of complex traits. *Nature Genet.* **36**, 1133–1137 (2004).
120. Canaran, P. *et al.* Panzea: an update on new content and features. *Nucleic Acids Res.* **36**, D1041–D1043 (2008).
121. Gibson, G. The environmental contribution to gene expression profiles. *Nature Rev. Genet.* **9**, 575–581 (2008).
122. Dankert, H., Wang, L., Hoopfer, E. D., Anderson, D. J. & Peron, P. Automated monitoring and analysis of social behavior in *Drosophila*. *Nature Methods* **6**, 297–303 (2009).
123. Lipkind, D. *et al.* New replicable anxiety-related measures of wall vs. center behavior of mice in the open field. *J. App. Physiol.* **97**, 347–352 (2004).

Acknowledgements
This work was supported by National Institutes of Health grants GM45146, GM076083 and AA016560.

DATABASES
Entrez Gene: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=gene
MAOA | *SLC6A4* | *tb1*

FURTHER INFORMATION
Mackay laboratory homepage: http://www.cals.ncsu.edu/genetics/mackay/mackay.html
1000 Genomes Project: http://www.1000genomes.org
1001 Genomes Project: http://www.1001genomes.org
Framingham Heart Study: http://www.framinghamheartstudy.org
Human Genome Sequencing Center at the Baylor College of Medicine: http://www.hgsc.bcm.tmc.edu
Mouse Genome Informatics at the Jackson Laboratories: http://www.informatics.jax.org/external/ko
UK Biobank: http://www.ukbiobank.ac.uk

SUPPLEMENTARY INFORMATION
See online article: S1 (table)

ALL LINKS ARE ACTIVE IN THE ONLINE PDF