VERSITA

Biologia **67**/3: 435—446, 2012
Section Cellular and Molecular Biology
DOI: 10.2478/s11756-012-0026-9

Review

Systems genetics: challenges and developing strategies

Hong Li[1] & Ping Zhang[2]

[1] *Chongqing Technology and Business University, Department of Biological Engineering, Key Laboratory of Catalysis Science and Technology of Chongqing Education Commission, Chongqing Nanan Xuefu Avenue 22#, 400067, People's Republic of China; e-mail: lhctbu@163.com*

[2] *Chongqing Technology and Business University, Library, Chongqing Nanan Xuefu Avenue 22#, 400067, People's Republic of China*

Abstract: Systems genetics is a new discipline based on the transcription mapping, which is also called "genetical genomics". In recent years, systems genetics has become more practical because of advances in science and technology. Analysis of expression quantitative trait loci (eQTLs) is an emerging technique in which individuals are genotyped across a panel of genetic markers and, simultaneously, phenotyped using DNA microarrays. Depending on eQTL mapping, one can infer the underlying regulatory network responsible for complex diseases or quantitative trait phenotypes. Systems genetics 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 made possible due to the development of massively parallel technologies for quantifying genome-wide levels of transcript abundance. The predictive power of the networks could be enhanced by more systematically integrating protein-protein interactions, protein-DNA interactions, protein-RNA interactions, RNA-RNA interactions, protein state information, methylation state, and interactions with metabolites. Systems genetics research will change the traditional approaches based on reductionism, and allows us to reconsider the living phenomenon and complex disease mechanism. Systems genetics benefits from varied "omics" researches (such as transcriptomics, metabolomics, and phenomics) and the development of bioinformatics tools and mathematical modeling, and will become mature in the near future like many other branches of genetics. Systems genetics is leading researchers to understand genetics systems from holism's viewpoint, and will open a wide field of vision for genetics researchers in systems biology era.

Key words: complex trait; eQTL mapping; systems genetics; regulatory network.

Abbreviations: CNVs, copy-number variations; ENCODE, Encyclopedia of DNA Elements; eQED, eQTL electrical diagrams; eQTL, expression quantitative trait locus; mQTL, metabolite quantitative trait locus; meQTL, methyl quantitative trait locus; GWAS, genome-wide association study; QTL, quantitative trait loci; QTN, quantitative trait nucleotides; QTTs, quantitative trait transcripts; SNP, single nucleotide polymorphism.

Introduction

Genetics is at the dawn of a new era with maturing technologies that enable low-cost, high-throughput genotyping of hundreds of thousands of DNA markers, which in turn can be tested for association to complex traits of interest like disease and drug response. A number of studies have already leveraged the availability of such technologies to identify polymorphisms in genes that associate with human complex diseases like age-related macular degeneration (Edwards et al. 2005; Haines et al. 2005; Klein et al. 2005), diabetes (Grant et al. 2006; Sladek et al. 2007), obesity (Herbert et al. 2006), hypertension (Yang et al. 2009), osteoporosis (Deng et al. 2002; Huang et al. 2003; Xiong et al. 2003), and age at menarche (Liu et al. 2009), to name just a few. In addition, there are scores of similar genome-wide association studies (GWAS) that promise to deliver scores of genes that harbor variations that associate with diseases. A new paradigm for bridging the gap between our knowledge of the physiology and predisposition of diseases is the combination of quantitative trait loci (QTL) mapping with large-scale gene expression analysis. Transcriptome mapping, also called "genetical genomics" (Jansen & Nap 2001; Jansen 2003), treats gene expression levels of any particular gene measured across different individuals as an expression-level polymorphism that in principle reflects the underlying genetic variation (Jansen & Nap 2001).

Many of the genetic variations that underlie disease susceptibility and morphology are complex and governed by loci that have quantitative effects on the phenotype. Genetic variation that affects gene regulation plays an important role in the genetics of human complex diseases and adaptive evolution (Kleinjan & van Heyningen 2005; Wray 2007). Determining the genetic architecture of complex traits poses a challenge because most phenotypic variations are caused by

numerous interactions between multitudes of environmentally sensitive genes (Swami et al. 2009). Systems genetics seeks to understand this complexity by integrating the questions and methods of systems biology with those of genetics to solve the fundamental problem of interrelating genotype and phenotype in complex traits and diseases (Nadeau & Dudley 2011). Systems genetics promises to integrate biological information to produce directed genetic networks that link molecular variants to organismal phenotypes, which enables us to understand the genetic basis inside the "black box" that lies between genotype and phenotype (Mackay et al. 2009). Recently, systems genetics approach for integrating a genetic polymorphism with genome-wide gene expression levels has successfully been applied to complex diseases and metabolism (Plaisier et al. 2009; Fleet et al. 2011). On 1–2 October 2009, the 1st Symposium of Systems Genetics was successfully held in the University of Groningen. Systems genetics has thus become the hotspot of biological and medical fields, and will recruit more genetists to understand the genetic basis underlying phenotypic variations of complex traits by combining bioinformatics tools and genome-scale analysis.

eQTL mapping and genetical genomics

A QTL, where mRNA expression is responsible for the variation of quantitative trait phenotype of interest, is generally referred to as an expression QTL (eQTL) (Rockman & Kruglyak 2006). eQTL mapping is based on the fact that the transcription efficiency of any gene is determined by the interaction of gene regulatory regions with transcription factors and RNA polymerase. Genome-wide mapping of eQTLs can provide a great insight into the genetic architecture of gene expression variations and elucidate the genetic regulation of entire transcriptomes, and are beginning to build biochemical pathways of interacting genes or causal networks on the basis of transcript level variations and molecular marker information (Narain 2010). Such eQTL studies are similar to traditional multi-trait QTL mapping but with thousands of phenotypes, which are useful for elucidating the molecular mechanisms of human complex diseases (Schadt & Lum 2006; Sieberts & Schadt 2007; Chen et al. 2008; Michaelson et al. 2009). By using high-throughput microarrays analysis technologies, eQTL can be measured for many genes in the genome, rendering eQTL data information rich and potentially very powerful. Recent studies on a variety of organisms, such as yeast, *Drosophila*, mouse, rat, human, and plants (Brem et al. 2005; Petretto et al. 2006; Druka et al. 2008; Veyrieras et al. 2008; Ruden et al. 2009; Lu et al. 2011), have shown that levels of gene expression are often highly heritable (Dixon et al. 2007; Göring et al. 2007; Emilsson et al. 2008). Generally, it is possible for many genes to map *cis*- and *trans*-acting factors using linkage (Göring et al. 2007; Emilsson et al. 2008) or association mapping (Dixon et al. 2007; Stranger et al. 2007a). As one gene regulates the level of expression of another (*trans*-acting eQTL), novel upstream or downstream components in gene regulation pathways can be identified (Holloway & Li 2010).

Functional genomics has become more practical because of advances in science and technology. Some examples are DNA-capturing techniques to facilitate gene hunting efforts in highly repetitive genomes, motif-directed profiling to specifically target genetic variation in functional parts of the genome, gene expression profiling to identify eQTL, and RNA interference to silence individual genes without altering the genome structure (Li & Deng 2010). Although eQTL studies have achieved considerable advances, there are still open questions about the biology and applications of eQTL mapping. First, there are important technical questions about the extent to which eQTLs are replicated across independent samples and independent platforms for measuring gene expression. Second, most of the human eQTL studies to date have analyzed transformed lymphoblast cell lines or lymphocyte samples (Dixon et al. 2007; Göring et al. 2007; Stranger et al. 2007a,b; Kwan et al. 2008). Indeed, three recent studies have started moving beyond cell types in blood by characterizing eQTLs in cortical (Myers et al. 2007), adipose (Emilsson et al. 2008) and liver (Schadt et al. 2008) tissues. Discovery of similarities (or differences) across tissues will add power to experimental findings, providing validation especially for *cis*-acting QTL, and revealing important underlying biology for the quantitative traits of interest. Further studies of eQTL mapping need to be applied to different development stages for human, such as embryogenesis and the comparison of different tissues. The research about the comparison of the differential expressions of genes in varied tissues, different time and environment conditions will enhance our recognition to the mechanism of complex diseases.

Unlike protein-coding sequences, however, we still know little about how to identify the DNA sequence elements that regulate the expression level of a gene of interest. It is still difficult for us to predict with any confidence which single nucleotide polymorphisms (SNPs) in genome are likely to affect gene expression, without performing targeted experimental assays. To address this gap, recent experimental and computational approaches have made progress on identifying DNA sequence elements that may be functional, for example through experimental methods that identify transcription factor binding sites (ENCODE Project Consortium 2007; Kim et al. 2007), by *in vivo* testing of possible enhancers (Pennacchio et al. 2006), and by computational analysis of sequence data (Blanchette et al. 2006; Xie et al. 2007). However, our understanding of the importance of different types of functional elements in gene regulation remains rudimentary. As a complementary approach, genome-wide studies of gene expression are now starting to provide information on genetic variation that affects gene expression levels (Gilad et al. 2008). Recently, the locations of SNPs within eQTL genes in experimental crosses in mice have been used to provide further information about the

identity of functional elements (GuhaThakurta et al. 2006). In studies of human lymphoblastoid cells, it has been reported that most strong signals of association lie within 100 kb of the transcribed region (Dixon et al. 2007), and that eQTLs cluster roughly symmetrically around the transcription start site (Stranger et al. 2007b). In addition, structural variations including short insertions/deletions (indels) and other more complex ones, such as duplications and translocations, have been demonstrated as the genetic basis of some complex diseases in the human population (Zhang et al. 2011). There are some tools or methods devised for the detection of genomic structural variation, such as BreakDancer (Chen et al. 2009), Indelign (Kim & Sinha 2007), MAQ (Li et al. 2008), VariationHunter (Hormozdiari et al. 2009), MoDIL (Lee et al. 2009), PEMer (Korbel et al. 2009), GASV (Sindi et al. 2009), VarScan (Kobolt et al. 2009), Pindel (Ye et al. 2009), CNV-seq (Xie et al. 2009), etc. Some databases containing clinical findings associated with submicroscopic chromosomal imbalance (including deletions, duplications, insertions, translocations, and inversions) are also developed, such as DECIPHER (DatabasE of Chromosomal Imbalance and Phenotype in Humans using Ensembl Resources), dbVar (database of Variation), CHOP (the Copy Number Variation project at the Children's Hospital of Philadelphia), and DGV (Database of Genomic Variants). Next generation sequencing has a perspective of the important applications in human genetics, including the detection of large and small insertions/deletions, of inversions, and of homozygous and heterozygous SNPs.

Results from eQTL studies have been used for identifying hotspots (Bystrykh et al. 2005; Lan et al. 2006), constructing gene networks (Li et al. 2005; Schadt et al. 2005), elucidating subclasses of clinical phenotypes (Bystrykh et al. 2005), and narrowing down lists of candidate genes (Bystrykh et al. 2005). Each of these tasks relies largely on the ability to generate a list of mapping transcripts and the genomic locations to which these transcripts map. At each genome region, the total number of mapping transcripts is tallied. The identification of hotspots provides lists of co-mapping transcripts and often leads to the inspection of putative candidates controlling the collection. Jansen & Nap (2001) were perhaps the first to formally recognize how hotspot lists could be used to construct networks. Although some of these might be artifacts of microarray normalization, regions with a high number of distal eQTLs might harbor "master regulators" that affect the expression levels of many genes. For example, Mehrabian et al. (2005) dissected an eQTL hotspot in mice and identified a locus that affected the regulation of several metabolic traits associated with obesity and bone density.

Computation and biology have been converging ever more closely for the past two decades, but with a vision of computing as a resource for biology that has propelled bioinformatics. To facilitate the interpretation of the data and the design of follow-up experimental validations, some databases and new algorithms for genomic data analysis and gene network inference are developed at a high rate. Gamazon et al. (2010) developed a database called SCAN (SNP and copy number annotation) that enables the sensible prioritization of these variants by combining several approaches, involving not only publicly available physical and functional annotations but also multilocus linkage disequilibrium annotations as well as annotations of eQTLs. A number of these data sets, particularly those for mouse and rat, are combined with sophisticated statistical and computational tools for the genetic dissection and synthesis of single traits or entire systems of traits, and have been integrated into the GeneNetwork (http://www.genenetwork.org/). Building on generic standard model FuGE (Functional Genomics Experiment) (Jones et al. 2007) for describing the experimental metadata on samples, protocols and experimental variables of functional genomics experiments, Swertz et al. (2010) developed the XGAP (eXtensible Genotype And Phenotype platform) object model to uniformly capture the wide variety of genotype and phenotype data. Some softwares, such as SysGenSIM, have been used for evaluating data analysis of systems genetics and studying gene expression dynamics (http://www.bioinformatica.crs4.it/node/156). This tool has recently been used to create a benchmark for the DREAM5 Systems Genetics challenges and is used to generate data for the National Institute of Health grant "Highly multivariate methods for quantitative trait loci mapping in systems genetics".

Systems genetics: combining genetics and systems biology

Over the last fifty years, numerous genetic resources have been devised and developed for specific purposes using a variety of inbred strains as progenitors. The major genetic resources that are widely used currently include recombinant inbred lines (Broman 2005), recombinant congenic strains (Demant & Hart 1986), genome-tagged or congenic lines (Iakoubova et al. 2001), chromosome substitution strains (Nadeau et al. 2000), heterogeneous stocks (Hitzemann et al. 1994), laboratory strain diversity panels drawn from the mouse phenome project (Paigen & Eppig 2000) for association studies, and, more recently, the collaborative cross of the Oak Ridge National Laboratory constructed for high-dimensional studies and high precision mapping (Chesler et al. 2008).

"Genetical genomics" or systems genetics (Sieberts & Schadt 2007) 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 made possible by the development of massively parallel technologies for quantifying genome-wide levels of transcript abundance. Associating DNA sequence

variation with variation in organismal phenotypes dispenses with 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 (Rockman & Kruglyak 2006).

As a statistical technique, eQTL mapping has become a powerful tool in systems biology (Gilad et al. 2008). For example, it is well-known that individuals differ in their sensitivity to radiation, and the role of genetics in radiosensitivity is most evident in humans with radiosensitivity syndromes, such as ataxia telangiectasia, Nijmegen breakage syndrome, xeroderma pigmentosum, and others (Gatti 2001). The systems genetics research of low dosage radiation has discovered the related genes of radiosensitivity, e.g. p53, Rb, xeroderma pigmentosum genes, etc. Radiation can damage DNA, alter gene expression, activate the signal pathway of p53, and modify intracellular oxidative status, etc., which can, in turn, result in changes at the systems level. The expression of some p53-regulated genes, such as *Cdkn1a* (cyclin-dependent kinase inhibitor 1), *Gadd45a* (growth arrest and DNA-damage-inducible protein), and *Mdm2* (murine double minute 2), appear to be up-regulated in response. The genes which responded to low dosage radiation are involved in functions relating to cell-to-cell signaling, signal transduction and DNA damage repair. In contrast, those which responded to high dosage radiation are involved in apoptosis and cell proliferation (Ding et al. 2005). By combining QTL mapping with gene expression data, researchers uncovered two particularly compelling candidate genes – *Ptprk* (protein tyrosine phosphatase receptor type K) and *Acp1* (acid phosphatase 1) – within the two significant QTLs associated with CD4:CD8 (Lynch 2010).

Applying systems genetics approaches to quantitative genetic variation is enhancing our ability to dissect the risk factor of human diseases by both reducing the number of candidate genes and directly identifying putative causal polymorphisms. One common approach now for examining complex traits in humans is the GWAS, which compares variation across the genomes of two groups of individuals and looks for markers that are associated more with one group than the other. After the variants linked to a trait have been identified by such approaches, the effects of these variants on the molecular networks in a cell or organism can be further investigated (Gunter 2008). GWAS takes advantage of the knowledge of linkage disequilibrium patterns in humans and the rapid development of high-throughput SNP genotyping platforms, and enables genetic variants at specific loci to be associated with particular diseases (Donnelly 2008). With high SNP density that makes possible the detection of culprit DNA changes within a narrow genomic region, the GWAS approach has demonstrated its great power to identify novel genes for human complex diseases/traits (Duerr et al. 2006; Frayling et al. 2007; Hunter et al. 2007). These advantages have been well illustrated over the past several years in the context of human GWAS (Altshuler et al. 2008; Emilsson et al. 2008; Schadt et al. 2008), for example, GWAS is a promising strategy to facilitate the identification of age at menarche genes (Liu et al. 2009). Moreover, the remarkable success in mapping genes linked to a number of disease traits using GWAS in human cohorts has renewed interest in applying this same technique in model organisms, such as inbred laboratory mice. Unlike humans, however, the limited genetic diversity in the ancestry of laboratory mice combined with selection pressure over the past decades has yielded an intricate population genetic structure that can complicate the results obtained from association studies. This problem is further exacerbated by the small number of strains typically used in such studies where multiple spurious associations arise as a result of random chance (Su et al. 2010).

At present, the reconstruction of biological networks that can be used to define how single genes increase disease risk is an intractable problem, partly because today's methods and data are inadequate for representing biological systems in a comprehensive fashion. Genetic networks provide a convenient framework for exploring the context within which single genes operate (Sieberts & Schadt 2007). Genomics networks interacting with transcriptional networks that in turn interact with protein, metabolic, and signaling networks, present challenges to elucidating disease (Schadt & Lum 2006). Until highly accurate genomics networks can be constructed from more complete data, inferences drawn from biological networks should be considered as hypothesis that need to be further tested with experimental method, where the results can then illuminate the representation of the biological system. The predictive power of the genomics networks could be enhanced by more systematically integrating interactions of informational molecules, such as protein-protein interactions, protein-DNA interactions, protein-RNA interactions, RNA-RNA interactions, protein state information, methylation state, and interactions with metabolites, as these types of data have become available (Schadt et al. 2009). A combination of functional, phenotypic, and genetic definitions of cell subpopulations will be an invaluable resource in providing evidence as to whether distinct phenotypic assays are in fact measuring the same cells, and potentially reveal evidence of relationships between apparently dissimilar phenotypes (Peirce et al. 2006). Causal inference approaches in systems genetics exploit QTL genotypes to infer causal relationships among phenotypes. QTLnet is a QTL-driven phenotype network method which jointly infers a causal phenotype network and associated genetic architecture for sets of correlated phenotypes (Neto et al. 2010).

The advent of high-throughput methods for measuring changes in gene expression has facilitated the study of biological systems as a whole (i.e. gene interaction networks), rather than on the discovery and characterization of its individual components (i.e. genes)

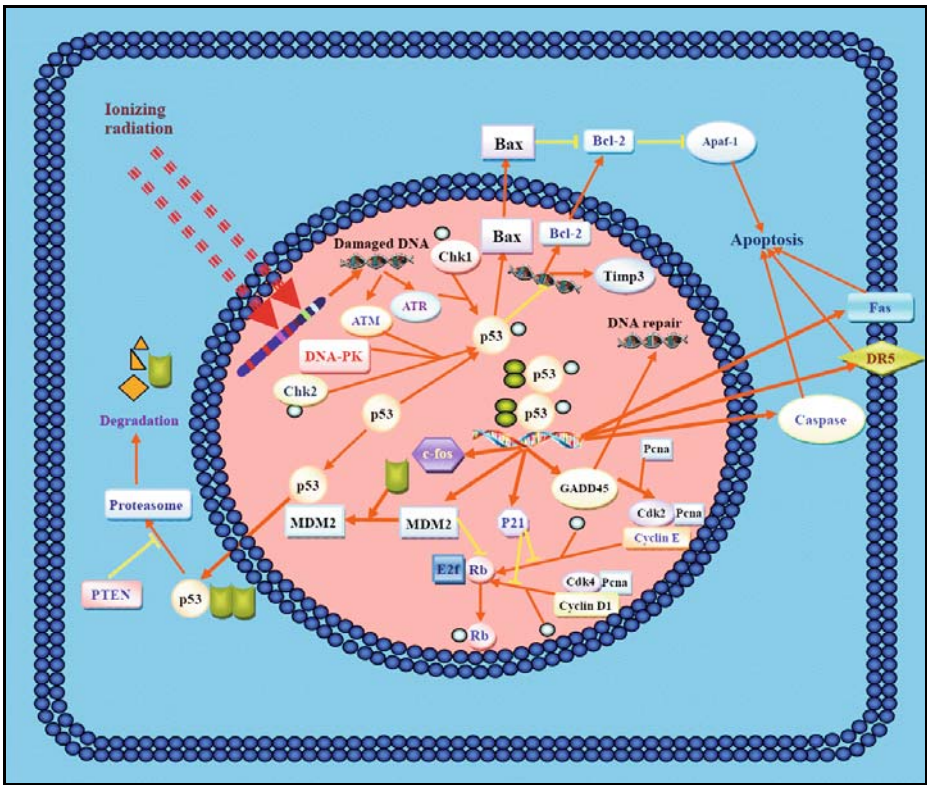

Fig. 1. Activation of the p53 signaling pathway by radiation. p53 is a transcription factor whose activity is regulation by phosphorylation. Damage to DNA can lead to apoptosis via *Bcl2* and *Bax* transcription. Alternately, *Cdkn1a*(*p21*) transcription can halt the cell cycle until DNA is repaired.

(Hood et al. 2004). The combination of hybridization and sequencing-based technologies such as DNA microarrays and RNA-seq with high-throughput methods for studying protein-DNA and protein-protein binding has enabled us to glean insights into global networks of interactions (Harbison et al. 2004; Workman et al. 2006; Simonis et al. 2009; The FANTOM Consortium & Riken Omics Science Center 2009). However, deep understanding of a complex dynamic system requires an examination of the dynamics of individual components during normal function or following perturbations (Kitano 2002). The existing methods of measuring transcription suffer from two major drawbacks. First, quantifying transcription dynamics using microarrays at multiple time-points is expensive when long processes are under study. Second, despite improvements in assay sensitivity, these approaches typically involve pooling mRNA from thousands of cells. Recent developments in cell-based assays combined with advances in reporter technology allow us to address these limitations, since expression levels can be repeatedly assayed in single cells. The Living Microarray platform is a step towards creating a more comprehensive platform for furthering our understanding of dynamic cellular processes at the systems level, as it provides us the ability to make parallel high-throughput measurements of transcriptional changes in single cells.

Epigenetic changes include modifications in the genome that do not affect the DNA sequence, such as DNA methylation, histone modification, and RNA silencing. DNA methylation plays a key role in regulating eukaryotic gene expression and has shown to be involved in the development of diseases, such as cancer, multiple sclerosis, diabetes, and schizophrenia. The studies of Gibbs et al. (2010) in brain illustrated that the majority of eQTLs or methyl QTLs (meQTLs) with strong effect sizes were consistent across tissues. For example, a large effect eQTL was found for *churc1*, which encodes a protein supposed to be involved in transcriptional regulation in all tissues. However, there were also rare, but observable, events where a large effect QTL was detected within a single tissue and was completely absent in the other three tissues. For example, the *cis*-eQTL for *ppapdc1a*, encoding a phosphatidic acid phosphatase that displays hydrolase and phosphotase activity at the membrane, has a large effect that appears to be restricted to the cerebellum, despite reliable detection of the transcript in all four brain regions. Several methods have been developed to map DNA methylation (meQTL) on a genomic scale, such as methylated DNA immunoprecipitation sequencing, methylated DNA capture by affinity purification, reduced representation bisulfite sequencing and the Infinium HumanMethylation27 BeadChip assay (Bock et al. 2010). Instead of gene expression being the quantitative trait, protein, and metabolite levels can be used as quantitative traits. In this manner, protein QTL and metabolite-QTL (mQTL) can be identified in both *cis*- and *trans*-genes that regulate the levels of molecules. In addition, data mining of expression sequence tag and microarray studies showed that the human brain carries more tissue-specific alternative splice forms than any

other tissue in the body. Alternative splicing may be enriched in certain systemic functions, including those of the immune and nervous systems.

Challenges and developing strategies of systems genetics

The eQTL analysis using high-throughput microarray technology is a potentially powerful way to detect transcriptional regulatory relationships on the genomic scale. Although this approach has led to several important discoveries, one persistent challenge in eQTL studies is the selection of loci and genes that should receive further biological investigation (Gatti et al. 2009). This requires us to improve the precision of eQTL mapping and to understand eQTL associations based on mechanistic explanation (Rockman & Kruglyak 2006; Schadt & Lum 2006). Firstly, we must break the major bottleneck for fine mapping caused by the limited number of recombinant inbred strains providing for eQTL analysis. Secondly, developing robust method to map the relationship between the expression trait and genotype is a promising strategy to handle the underexploited problem of large scale eQTL datasets in legacy QTL approaches. These data sets are so massive that standard publishing procedures cannot accommodate them, which limits the application of these methods in the analysis of complex traits, particularly in the case of epistasis (Michaelson et al. 2010). In addition, the extension of the paradigm of transcriptome mapping to the proteome and metabolome would be unique. The question of whether microarrays are valid indicators of actual protein levels is still of major importance, and such results would test and extend this question in regard to comparison of the underlying polygenic control at each step of the central dogma of biology. Phenotyping methods, including gene expression arrays, or phenotyping based on non-invasive imaging will have to be integrated into the complex trait studies. This shift to microphenotypes requires new statistical tools because of multiple-testing issues (Rudi 2007). Recent achievements in multiple testing issues, such as SLIDE (Sliding-window approach for Locally Inter-correlated markers with asymptotic Distribution Errors corrected) and SLIP (Sliding-window approach for Locally Inter-correlated markers for Power estimation) method of Han et al. (2009) and MCFDR algorithm for false discovery rate modulated sequential Monte Carlo multiple hypothesis testing of Sandve et al. (2011) will shed light on the application of microphenotypes. In addition, "quantitative trait nucleotide" (QTN) analyses have been also developed to circumvent this problem by associating individual nucleotide polymorphisms in large populations with a quantitative trait. QTNs allow us to map phenotype to genotype in the absence of the knowledge about the flow of information from DNA to the organismal phenotype through RNA intermediates, proteins, metabolites and other molecular endo-phenotypes. This could be achieved by dissecting a QTN into its constituent eQTLs and QTTs (quantitative trait transcripts), but in practice this is not easy; one can expect a substantial co-expression network of relevant transcripts that associate both with the molecular variant and with the organismal phenotype (Chen et al. 2008). The QTN approach is more clinically relevant and is thus more readily applicable to humans, where larger sample sizes are almost always available (Ruden 2007). A recent proposed method called "eQTL electrical diagrams" (eQED) sheds light on the problem for prioritizing candidate genes at a locus, which models the flow of information as electric currents through the protein network from a locus to target genes and provides a means of predicting the direction of signal transmission (Silpa et al. 2008).

High-throughput microarray analysis technology is central to systems genetics, which can simultaneously examine tissues from as many as 100 related strains and integrate the transcriptomic data with the underlying genotypic data. Systems genetics makes it possible to define networks and clusters of co-regulated transcripts; to correlate phenotype with genotype; and to map genes which affect the expression levels of other genes (Morahan et al. 2008). Measurements of gene expression with high-throughput analysis technologies provide us with an effective approach for learning which types of SNPs in genome are most likely to affect gene regulation, in a way that complements other experimental approaches, such as allele-specific expression studies (ENCODE Project Consortium 2007) or reporter gene assays (Chabot et al. 2007). eQTL mapping can provide important information for dissecting the genetics background underlying human complex disease. Not a lot of SNP functional data have been produced from ChIP-on-chip or ChIP-Seq for some samples, such as brain, so eQTL mapping is the primary mode of discovery for functional genetic variants. In its simplest form, the identification of eQTLs can provide a tool for connecting SNPs that are significant in GWAS of disease to a molecular mechanism (Moffatt et al. 2007; Schadt et al. 2008). eQTLs may be also useful for linking genes and individual variants to cellular phenotypes, such as cell line sensitivity to chemotherapeutic agents (Huang et al. 2007). More ambitiously, one might be able to use patterns of gene expression and eQTL mapping in people with and without disease to identify networks of genes that are differentially regulated in the two groups (Chen et al. 2008). Any eQTL that up- or down-regulates such a network is a natural candidate for affecting complex disease phenotype itself and would be of particular interest in association studies of complex diseases (Gilad et al. 2008).

Since GWAS tests simultaneously millions of SNPs, omitting multiple-testing correction with high threshold will result in a large number of false positives. It means that very large sample sizes are needed to achieve sufficient statistical power to detect risk alleles with weak effects. Based on previously identified candidate genes or pathways, and the functional effects of SNPs, the power of GWAS can be increased through limiting tests to the most likely relevant variants or functional

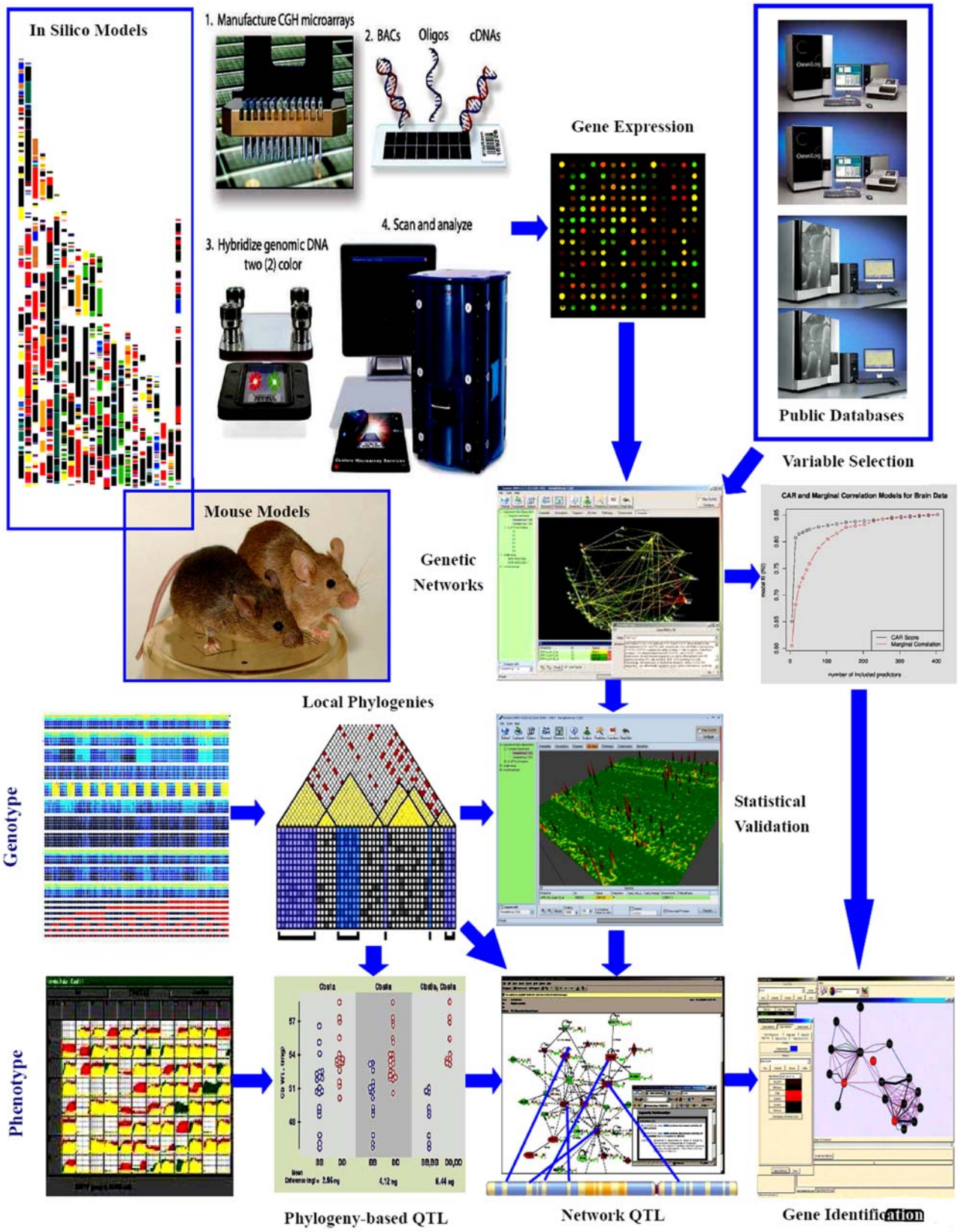

Fig. 2. The experimental approach of systems genetics.

variants, which will allow us to detect weak effects in a relatively small but sufficiently-powered study and can generate a new priori that can be tested in other studies. In large-scale epidemiological studies ($n > 1{,}000$), association between metabolic phenotypes and disease phenotypes gave rise to metabolome-wide association studies. QTL mapping of metabolomic traits (mQTL), and metabolome-wide genome-wide association studies

are robust and accurate strategies for the integration of genome-wide genotyping and metabolome-wide profiling by ^{1}H-NMR and mass spectroscopy, identifying candidate biomarkers and susceptibility genes. Recent developments such as metabolite-set enrichment analysis or integrated metabolome interactome mapping allow a better mechanistic understanding of metabolic phenotypes.

Advances in genotyping and sequencing technologies have revolutionized the genetics of complex diseases by locating rare and common variants that influence an individual's risk for diseases, such as diabetes, cancers, and psychiatric disorders. However, to capitalize on these data for prevention and therapies requires the identification of causal alleles and a mechanistic understanding of how these variants contribute to the disease (Raychaudhuri 2011). Recent studies suggest that all the risk alleles identified so far explain only a small proportion of total disease susceptibility variance (about 3–20%); the remaining, unexplained variation (referred to as the "missing heritability") may be caused by gene-gene interaction, heritable epigenetic variants and rare variants (Purcell et al. 2009). In the past, conventional genetic studies for inherited diseases, such as cystic fibrosis identified rare, mutated genes that have a high penetrance, meaning that the gene has an effect in almost everyone who carries it. However, it quickly became apparent that high-penetrance variants would not underlie most common diseases because evolution largely keeps them in check (Brendan 2008). Some researchers are now homing in on copy-number variations (CNVs), stretches of DNA tens or hundreds of base pairs long that are deleted or duplicated between individuals. Variations in these features could begin to explain missing heritability in disorders, such as schizophrenia and autism, for which GWAS have turned up almost nothing. Two recent studies looked at hundreds of CNVs in normal people and in those with schizophrenia, and found strong associations between the disease and several CNVs (Stefansson et al. 2008; The International Schizophrenia Consortium 2008). They commonly arise in an individual without any family history of the mutation. These structural variants might account for a lot of the genetic variability from person to person and could account for some of those rare "out-of-sight" mutations with moderate penetrance that GWAS can not pick up. Many CNVs go undetected because they do not alter SNP sequences. Duplicated regions can also be difficult to sequence. Reduced fecundity, associated with severe mental disorders, places negative selection pressure on risk alleles and may explain, in part, why common variants have not been found that confer risk of disorders, such as autism, schizophrenia and mental retardation. Thus, rare variants may account for a larger fraction of overall genetic risks than previously assumed. In contrast to rare single nucleotide mutations, rare CNVs can be detected using genome-wide SNP arrays. This has led to the identification of CNVs associated with mental retardation and autism (Stefansson et al. 2008). In GWAS, common variants are typically identified though individual testing, whereas rare variants, each with incommensurable effects on phenotypic traits, are difficult to identify using the traditional methods. If various rare variants in a group influence phenotype of complex traits, focusing on the group rather than on an individual, variant helps enrich the association signals, reduce the number of degrees of freedom in tests, and subsequently increase statistical power (Mccarthy et al. 2008). Recently, Zhu et al. (2011) used re-sequenced candidate gene fragment data to showcase the Resequenced-based Genome-Wide Association Studies (CR-GWAS) analysis and expected similar analysis strategies to be applied to exome sequencing and whole-genome re-sequencing studies. On the other hand, data generated through array-based genotyping approaches could also be analyzed in a similar framework if the ultrahigh-density genotyping chip containing rare SNPs provides adequate context sequence polymorphisms for function prediction. In addition, it would be challenging but highly desirable incorporating various analytical methods into a common platform developed for population stratification correction, testing of common variants and rare variants (with flexible weight assignment), threshold determination, and computational load reduction (Pritchard et al. 2000; Devlin et al. 2004; Price et al. 2006; Yu et al. 2006; Aulchenko et al. 2007; Kang et al. 2010; Zhang et al. 2010). Based on the above, we propose several view-points for the systems genetics research as follows: (1) combining GWAS studies and RNA-sequencing for explaining the mechanism of complex traits or diseases; (2) reconstructing regulatory networks based both on common variants and rare variants; (3) developing more robust statistical method of eQTL mapping and software tools; (4) incorporating various "-omics" data into a common platform developed for systems genetics study; and (5) introducing novel analysis technologies into systems genetics.

Conclusions

Systems genetics as a new discipline is still in its infancy. Maturation of the field will proceed as many challenges that it faces are successfully solved. In one word, there are still many challenges confronting systems genetics in its development process, including the design of operable experiment programs and the collection of reliable data, the development of new instrumentation for making rapid, highly parallel, inexpensive and accurate measurements, the development and refinement of network theory and the effective engineering of simulation tools, the combination of different disciplines, the cooperation of research teams, the difficulties in obtaining funding, and the like. Any of the above will construct an obstacle to systems genetics research if these problems cannot be successfully solved. Although there are many challenges in the avenue of systems genetics, we believe that systems genetics will achieve rapid development and become a mature discipline in the near

future as many branches of genetics through benefiting from the integration of varied "omics" and bioinformatics tools. Systems genetics is leading researchers to understand genetics systems from holism's viewpoint, and will open a wide field of vision for genetics researchers in systems biology era.

References

Altshuler D., Daly M.J. & Lander E.S. 2008. Genetic mapping in human disease. Science **322:** 881–888. doi: 10.1126/science.1156409

Aulchenko Y.S., de Koning D.J. & Haley C. 2007. Genome-wide rapid association using mixed model and regression: a fast and simple method for genomewide pedigree-based quantitative trait loci association analysis. Genetics **177:** 577–585. doi: 10.1534/genetics.107.075614

Bailey D.W. 1971. Recombinant-inbred strains. An aid to finding identity, linkage, and function of histocompatibility and other genes. Transplantation **11:** 325–327.

Blanchette M., Bataille A., Chen X., Poitras C., Laganiere J., Lefebvre C., Deblois G., Giguere V., Ferretti V., Bergeron D., Coulombe B. & Robert F. 2006. Genome-wide computational prediction of transcriptional regulatory modules reveals new insights into human gene expression. Genome Res. **16:** 656–668. doi: 10.1101/gr.4866006

Bock C., Tomazou E.M., Brinkman A.B., Müller F., Simmer F., Gu H.C., Jäger N., Gnirke A., Stunnenberg H.G. & Meissner A. 2010. Quantitative comparison of genome-wide DNA methylation mapping technologies. Nat. Biotechnol. **28:** 1106–1114. doi: 10.1038/nbt.1681

Brem R.B., Storey J.D., Whittle J. & Kruglyak L. 2005. Genetic interactions between polymorphisms that affect gene expression in yeast. Nature **436:** 701–703. doi: 10.1038/nature03865

Brendan M. 2008. Personal genomes: the case of the missing heritability. Nature **456:** 18–21. doi: 10.1038/456018a

Broman K.W. 2005. The genomes of recombinant inbred lines. Genetics **169:** 1133–1146. doi: 10.1534/genetics.104.035212

Bystrykh L., Weersing E., Dontje B., Sutton S., Pletcher M.T., Wiltshire T., Su A.I., Vellenga E., Wang J., Manly K.F., Lu L., Chesler E.J., Alberts R., Jansen R.C., Williams R.W., Cooke M.P. & de Haan G. 2005. Uncovering regulatory pathways that affect hematopoietic stem cell function using "genetical genomics". Nat. Genet. **37:** 225–232. doi: 10.1038/ng1497

Chabot A., Shrit R.A., Blekhman R. & Gilad Y. 2007. Using reporter gene assays to identify *cis*-regulatory differences between humans and chimpanzees. Genetics **176:** 2069–2076. doi: 10.1534/genetics.107.073429

Chen K., Wallis J.W., McLellan M.D., Larson D.E., Kalicki J.M., Pohl C.S., McGrath S.D., Wendl M.C., Zhang Q., Locke D.P., Shi X., Fulton R.S., Ley T.J., Wilson R.K., Ding L. & Mardis E.R. 2009. BreakDancer: an algorithm for high-resolution mapping of genomic structural variation. Nat. Meth. **6:** 677–681. doi: 10.1038/nmeth.1363

Chen Y., Zhu J., Lum P.Y., Yang X., Pinto S., MacNeil D.J., Zhang C., Lamb J., Edwards S., Sieberts S.K., Leonardson A., Castellini L.W., Wang S., Champy M.F., Zhang B., Emilsson V., Doss S., Ghazalpour A., Horvath S., Drake T.A., Lusis A.J. & Schadt E.E. 2008. Variations in DNA elucidate molecular networks that cause disease. Nature **452:** 429–435. doi: 10.1038/nature06757

Chesler E.J., Miller D.R., Branstetter L.R., Galloway L.D., Jackson B.L., Philip V.M., Voy B.H., Culiat C.T., Threadgill D.W., Williams R.W., Churchill G.A., Johnson D.K. & Manly K.F. 2008. The Collaborative Cross at Oak Ridge National Laboratory: developing a powerful resource for systems genetics. Mamm. Genome **19:** 382–389. doi: 10.1007/s00335-008-9135-8

Demant P. & Hart A.A. 1986. Recombinant congenic strains – a new tool for analyzing genetic traits determined by more than one gene. Immunogenetics **24:** 416–422. doi: 10.1007/BF00377961

Deng H.W., Xu F.H., Huang Q.Y., Shen H., Deng H., Conway T., Liu Y.J., Liu Y.Z., Li J.L., Zhang H.T., Davies K.M. & Recker R.R. 2002. A whole-genome linkage scan suggests several genomic regions potentially containing quantitative trait loci for osteoporosis. J. Clin. Endocrinol Metab. **87:** 5151–5159. doi: 10.1210/jc.2002-020474

Devlin B., Bacanu S. A. & Roeder K. 2004. Genomic control to the extreme. Nat. Genet. **36:** 1129–1130. doi: 10.1038/ng1104-1129

Ding L.H., Shingyoji M., Chen F., Hwang J.J., Burma S., Lee C., Cheng J.F. & Chen D.J. 2005. Gene expression profiles of normal human fibroblasts after exposure to ionizing radiation: a comparative study of low and high doses. Radiat. Res. **164:** 17–26. doi: 10.1667/RR3354

Dixon A., Liang L., Moffatt M., Chen W., Heath S., Wong K., Taylor J., Burnett E., Gut I., Farrall M., Lathrop G.M., Abecasis G.R. & Cookson W.O.C. 2007. A genome-wide association study of global gene expression. Nat. Genet. **39:** 1202–1207. doi: 10.1038/ng2109

Donnelly P. 2008. Progress and challenges in genome-wide association studies in humans. Nature **456:** 728–731. doi: 10.1038/nature07631

Druka A., Potokina E., Luo Z., Bonar N., Druka I., Zhang L., Marshall D.F., Steffenson B.J., Close T.J., Wise R.P., Kleinhofs A., Williams R.W., Kearsey M.J. & Waugh R. 2008. Exploiting regulatory variation to identify genes underlying quantitative resistance to the wheat stem rust pathogen *Puccinia graminis* f. sp. *tritici* in barley. Theor. Appl. Genet. **117:** 261–272. doi: 10.1007/s00122-008-0771-x

Duerr R.H., Taylor K.D., Brant S.R., Rioux J.D., Silverberg M.S., Daly M.J., Steinhart A.H., Abraham C., Regueiro M., Griffiths A., Dassopoulos T., Bitton A., Yang H., Targan S., Datta L.W., Kistner E.O., Schumm L.P., Lee A.T., Gregersen P.K., Barmada M.M., Rotter J.I., Nicolae D.L. & Cho J.H. 2006. A genome-wide association study identifies IL23R as an inflammatory bowel disease gene. Science **314:** 1461–1463. doi: 10.1126/science.1135245

Edwards A.O., Ritter III R., Abel K.J., Manning A., Panhuysen C. & Farrer L.A. 2005. Complement factor H polymorphism and age-related macular degeneration. Science **308:** 421–424. doi: 10.1126/science.1110189

Eichler E.E., Flint J., Gibson G., Kong A., Leal S.M., Moore J.H. & Nadeau J.H. 2010. Missing heritability and strategies for finding the underlying causes of complex disease. Nat. Rev. Genet. **11:** 446–450. doi: 10.1038/nrg2809

Emilsson V., Thorleifsson G., Zhang B., Leonardson A., Zink F., Zhu J., Carlson S., Helgason A., Walters G.B., Gunnarsdottir S., Mouy M., Steinthorsdottir V., Eiriksdottir G.H., Bjornsdottir G., Reynisdottir I., Gudbjartsson D., Helgadottir A., Jonasdottir A., Jonasdottir A., Styrkarsdottir U., Gretarsdottir S., Magnusson K.P., Stefansson H., Fossdal R., Kristjansson K., Gislason H.G., Stefansson T., Leifsson B.G., Thorsteinsdottir U., Lamb J.R., Gulcher J.R., Reitman M.L., Kong A., Schadt E.E. & Stefansson K. 2008. Genetics of gene expression and its effect on disease. Nature **452:** 423–428. doi: 10.1038/nature06758

ENCODE Project Consortium. 2007. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature **447:** 799–816. doi: 10.1038/nature05874

Fleet J.C., Replogle R. & Salt D.E. 2011. Systems genetics of mineral metabolism. J. Nutrit. **141:** 520–525. doi: 10.3945/jn.110.128736

Frayling T.M., Timpson N.J., Weedon M.N., Zeggini E., Freathy R.M., Lindgren C.M., Perry J.R.B., Elliott K.S., Lango H., Rayner N.W., Shields B., Harries L.W., Barrett J.C., Ellard S., Groves C.J., Knight B., Patch A.M., Ness A.R., Ebrahim S., Lawlor D.A., Ring S.M., Ben-Shlomo Y., Jarvelin M.R., Sovio U., Bennett A.J., Melzer D., Ferrucci L., Loos R.J.F., Barroso I., Wareham N.J., Karpe F., Owen K.R., Cardon L.R., Walker M., Hitman G.A., Palmer C.N.A., Doney A.S.F., Morris A.D., Smith G.D., The Wellcome Trust Case Control Consortium, Hattersley A.T. & McCarthy M.I. 2007. A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. Science **316:** 889–894. doi: 10.1126/science.1141634

Gatti D.M., Harrill A.H., Wright F.A., Threadgill D.W. & Rusyn I. 2009. Replication and narrowing of gene expression quantitative trait loci using inbred mice. Mamm. Genome **20:** 437–446. doi: 10.1007/s00335-009-9199-0

Gatti R.A. 2001. The inherited basis of human radiosensitivity. Acta Oncol. **40:** 702–711.

Gibbs J.R., van der Brug M.P., Hernandez D.G., Traynor B.J., Nalls M.A., Lai S.L., Arepalli S., Dillman A., Rafferty Ian P., Troncoso J., Johnson R., Zielke H.R., Ferrucci L., Longo D.L., Cookson M.R. & Singleton A.B. 2010. Abundant quantitative trait loci exist for DNA methylation and gene expression in human brain. PLoS Genet. **6:** e1000952. doi: 10.1371/journal.pgen.1000952

Gilad Y., Rifkin S. & Pritchard J. 2008. Revealing the architecture of gene regulation: the promise of eQTL studies. Trends Genet. **24:** 408–415. doi: 10.1016/j.tig.2008.06.001

Göring H.H.H., Curran J.E., Johnson M.P., Dyer T.D., Charlesworth J., Cole S.A., Jowett J.B.M., Abraham L.J., Rainwater D.L., Comuzzie A.G., Mahaney M.C., Almasy L., MacCluer J.W., Kissebah A.H., Collier G.R., Moses E.K. & Blangero J. 2007. Discovery of expression QTLs using large-scale transcriptional profiling in human lymphocytes. Nat. Genet. **39:** 1208–1216. doi: 10.1038/ng2119

Grant S.F., Thorleifsson G., Reynisdottir I., Benediktsson R., Manolescu A., Sainz J., Helgason A., Stefansson H., Emilsson V., Helgadottir A., Styrkarsdottir U., Magnusson K.P., Walters G.B., Palsdottir E., Jonsdottir T., Gudmundsdottir T., Gylfason A., Saemundsdottir J., Wilensky R.L., Reilly M.P., Rader D.J., Bagger Y., Christiansen C., Gudnason V., Sigurdsson G., Thorsteinsdottir U., Gulcher J.R., Kong A. & Stefansson K. 2006. Variant of transcription factor 7–like 2 (TCF7L2) gene confers risk of type 2 diabetes. Nat. Genet. **38:** 320–323. doi: 10.1038/ng1732

GuhaThakurta D., Xie T., Anand M., Edwards S.W., Li G., Wang S.S. & Schadt E.E. 2006. *Cis*-regulatory variations: A study of SNPs around genes showing *cis*-linkage in segregating mouse populations. BMC Genomics **7:** 235. doi: 10.1186/1471-2164-7-235

Gunter C. 2008. Quantitative genetics. Nature **456:** 719. doi: 10.1038/456719a

Haines J.L., Hauser M.A., Schnidt S., Scott W.K., Olson L.M., Gallins P., Spencer K.L., Kwan S.Y., Noureddine M., Gilbert J.R., Schnetz-Boutaud N., Agarwal A., Postel E.A. & Pericak-Vance M.A. 2005. Complement factor H variant increases the risk of age-related macular degeneration. Science **308:** 419–421. doi: 10.1126/science.1110359

Han B., Kang H.M. & Eskin E. 2009. Rapid and accurate multiple testing correction and power estimation for millions of correlated markers. PLoS Genet. **5:** e1000456. doi: 10.1371/journal.pgen.1000456

Harbison C.T., Gordon D.B., Lee T.I., Rinaldi N.J., Macisaac K.D., Danford T.W., Hannett N.M., Tagne J., Reynolds D.B., Yoo J., Jennings E.G., Zeitlinger J., Pokholok D.K., Kellis M., Rolfe P.A., Takusagawa K.T., Lander E.S., Gifford D.K., Fraenkel E. & Young R.A. 2004. Transcriptional regulatory code of a eukaryotic genome. Nature **431:** 99–104. doi: 10.1038/nature02800

Herbert A., Gerry N.P., McQueen M.B., Heid I.M., Pfeufer A., Illig T., Wichmann H.E., Meitinger T., Hunter D., Hu F.B., Colditz G., Hinney A., Hebebrand J., Koberwitz K., Zhu X., Cooper R., Ardlie K., Lyon H., Hirschhorn J.N., Laird N.M., Lenburg M.E., Lange C. & Christman M.F. 2006. A common genetic variant is associated with adult and childhood obesity. Science **312:** 279–283. doi: 10.1126/science.1124779

Hitzemann B., Dains K., Kanes S. & Hitzemann R. 1994. Further studies on the relationship between dopamine cell density and haloperidol-induced catalepsy. J. Pharmacol. Exp. Ther. **271:** 969–976.

Holloway B. & Li B. 2010. Expression QTLs: applications for crop improvement. Mol. Breeding **26:** 381–391. doi: 10.1007/s11032-010-9396-2

Hood L., Heath J.R., Phelps M.E. & Lin B. 2004. Systems biology and new technologies enable predictive and preventative medicine. Science **306:** 640–643. doi: 10.1126/science.1104635

Hormozdiari F., Alkan C., Eichler E.E. & Sahinalp S.C. 2009. Combinatorial algorithms for structural variation detection in high-throughput sequenced genomes. Genome Res. **19:** 1270–1278. doi: 10.1101/gr.088633.108

Huang Q.Y., Recker R.R. & Deng H.W. 2003. Searching for the osteoporosis gene(s) in the post-genome era: progress and challenge. Osteoporos. Int. **14:** 701–715. doi: 10.1007/s00198-003-1445-9

Huang R.S., Duan S., Bleibel W.K., Kistner E.O., Zhang W., Clark T.A., Chen T.X., Schweitzer A.C., Blume J.E., Cox N.J. & Dolan M.E. 2007. A genome-wide approach to identify genetic variants that contribute to etoposide-induced cytotoxicity. Proc. Natl. Acad. Sci. USA **104:** 9758–9763. doi: 10.1073/pnas.0703736104

Hunter D.J., Kraft P., Jacobs K.B., Cox D.G., Yeager M., Hankinson S.E., Wacholder S., Wang Z., Welch R., Hutchinson A., Wang J., Yu K., Chatterjee N., Orr N., Willett W.C., Colditz G.A., Ziegler R.G., Berg C.D., Buys S.S., McCarty C.A., Feigelson H.S., Calle E.E., Thun M.J., Hayes R.B., Tucker M., Gerhard D.S., Fraumeni J.F. Jr, Hoover R.N., Thomas G. & Chanock S.J. 2007. A genome-wide association study identifies alleles in FGFR2 associated with risk of sporadic postmenopausal breast cancer. Nat. Genet. **39:** 870–874. doi: 10.1038/ng2075

Iakoubova O.A., Olsson C.L., Dains K.M., Ross D.A., Andalibi A., Lau K., Choi J., Kalcheva I., Cunanan M., Louie J., Nimon V., Machrus M., Bentley L.G., Beauheim C., Silvey S., Cavalcoli J., Lusis A.J. & West D.B. 2001. Genome-tagged mice (GTM): two sets of genome-wide congenic strains. Genomics **74:** 89–104. doi: 10.1006/geno.2000.6497

Jansen R.C. 2003. Studying complex biological systems using multifactorial perturbation. Nat. Rev. Genet. **4:** 145–151. doi: 10.1038/nrg996

Jansen R.C. & Nap J.P. 2001. Genetical genomics: the added value from segregation. Trends Genet. **17:** 388–391.doi: 10.1016/S0168-9525(01)02310-1

Kang H.M., Sul J.H., Service S.K., Zaitlen N.A., Kong S.Y., Freimer N.B., Sabatti C. & Eskin E. 2010. Variance component model to account for sample structure in genome-wide association studies. Nat. Genet. **42:** 348–354. doi: 10.1038/ng.548

Kim I. & Sinha S. 2007. Indelign: a probabilistic framework for annotation of insertions and deletions in a multiple alignment. Bioinformatics **23:** 289–297. doi: 10.1093/bioinformatics/btl578

Kim T., Abdullaev Z., Smith A., Ching K., Loukinov D., Green R., Zhang M., Lobanenkov V. & Ren B. 2007. Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome. Cell **128:** 1231–1245. doi: 10.1016/j.cell.2006.12.048

Kitano H. 2002. Systems biology: a brief overview. Science **295:** 1662–1664. doi: 10.1126/science.1069492

Kleinjan D. & van Heyningen V. 2005. Long-range control of gene expression: emerging mechanisms and disruption in disease. Amer. J. Hum. Genet. **76:** 8–32. doi: 10.1086/426833

Klein R.J., Zeis C., Chew E.Y., Tsai J.Y., Sackler R.S., Haynes C., Henning A.K., SanGiovanni J.P., Mane S.M., Mayne S.T., Bracken M.B., Ferris F.L., Ott J., Barnstable C. & Hoh J. 2005. Complement factor H polymorphism in age-related macular degeneration. Science **308:** 385–389. doi: 10.1126/science.1109557

Kobolt D.C., Chen K., Wylie T., Larson D.E., McLellan M.D., Mardis E.R., Weinstock G.M., Wilson R.K. & Ding L. 2009. VarScan: variant detection in massively parallel sequencing of individual and pooled samples. Bioinformatics **25:** 2283–2285. doi: 10.1093/bioinformatics/btp373

Korbel J.O., Abyzov A., Mu X.J., Carriero N., Cayting P., Zhang Z., Snyder M. & Gerstein M.B. 2009. PEMer: a computational framework with simulation-based error models for inferring genomic structural variants from massive paired-end sequencing data. Genom. Biol. **10:** R23. doi: 10.1186/gb-2009-10-2-r23

Kwan T., Benovoy D., Dias C., Gurd S., Provencher C., Beaulieu P., Hudson T.J., Sladek R. & Majewski J. 2008. Genome-wide analysis of transcript isoform variation in humans. Nat. Genet. **40:** 225–231. doi: 10.1038/ng.2007.57

Lan H., Chen M., Flowers J.B., Yandell B.S., Stapleton D.S., Mata C.M., Mui E.T., Flowers M.T., Schueler K.L., Manly

K.F., Williams R.W., Kendziorski C. & Attie A.D. 2006. Combined expression trait correlations and expression quantitative trait locus mapping. PLoS Genet. **2:** e6. doi: 10.1371/journal.pgen.0020006

Lee S., Cheran E. & Brudno M. 2008. A robust framework for detecting structural variations in a genome. Bioinformatics **24:** i59–i67. doi: 10.1093/bioinformatics/btn176

Lee S., Hormozdiari F., Alkan C. & Brudno M. 2009. MoDIL: detecting small indels from clone-end sequencing with mixtures of distributions. Nat. Meth. **6:** 473–474. doi: 10.1038/nmeth.f.256

Li H. & Deng H. 2010. Systems genetics, bioinformatics and eQTL mapping. Genetica **138:** 915–924. doi: 10.1007/s10709-010-9480-x

Li H., Lu L., Manly K.F., Chesler E.J., Bao L., Wang J., Zhou M., Williams R.W. & Cui Y. 2005. Inferring gene transcriptional modulatory relations: a genetical genomics approach. Hum. Mol. Genet. **14:** 1119–1125. doi: 10.1093/hmg/ddi124

Li H., Ruan J. & Durbin R. 2008. Mapping short DNA sequencing reads and calling variants using mapping quality scores. Genome Res. **18:** 1851–1858. doi: 10.1101/gr.0788212.108

Liu Y.Z., Guo Y.F., Wang L., Tan L.J., Liu X.G., Pei Y.F., Yan H., Xiong D.H., Deng F.Y., Yu N., Zhang Y.P., Zhang L., Lei S.F., Chen X.D., Liu H.B., Zhu X.Z., Levy S., Papasian C.J., Drees B.M., Hamilton J.J., Recker R.R. & Deng H.W. 2009. Genome-wide association analyses identify *SPOCK* as a key novel gene underlying age at menarche. PLoS Genet. **5:** e1000420. doi: 10.1371/journal.pgen.1000420

Lu H., Wang X., Pullen M., Guan H., Chen H., Sahu S., Zhang B., Chen H., Williams R.W., Geisert E.E., Lu L. & Jablonski MM. 2011. Genetic dissection of the Gpnmb network in the eye. Invest. Ophthalmol. Vis. Sci. **52:** 4132–42. doi: 10.1167/iovs.10-6493

Lynch R.M. 2010. A systems genetics approach to the characterization of differential low dose radiation responses in BXD recombinant inbred mice. PhD Thesis, University of Tennessee.

Mackay T.F.C., Stone E.A. & Ayroles J.F. 2009. The genetics of quantitative traits: challenges and prospects. Nat. Rev. Genet. **10:** 565–577. doi: 10.1038/nrg2612

McCarthy M.I., Abecasis G.R., Cardon L.R., Goldstein D.B., Little J., Ioannidis J.P.A. & Hirschhorn J.N. 2008. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. Nat. Rev. Genet. **9:** 356–369. doi: 10.1038/nrg2344

Mehrabian M. Allayee H., Stockton J., Lum P.Y., Drake T.A., Castellani L.W., Suh M., Armour C., Edwards S., Lamb J., Lusis A.J. & Schadt E.E. 2005. Integrating genotypic and expression data in a segregating mouse population to identify 5–lipoxygenase as a susceptibility gene for obesity and bone traits. Nat. Genet. **37:** 1224–1233. doi: 10.1038/ng1619

Michaelson J.J., Alberts R., Schughart K. & Beyer A. 2010. Data-driven assessment of eQTL mapping methods. BMC Genomics **11:** 502. doi: 10.1186/1471-2164-11-502

Michaelson J.J., Loguercio S. & Beyer A. 2009. Detection and interpretation of expression quantitative trait loci (eQTL). Methods **48:** 265–276. doi: 10.1016/j.ymeth.2009.03.004

Moffatt M.F., Kabesch M., Liang L., Dixon A.L., Strachan D., Heath S., Depner M., von Berg A., Bufe A., Rietschel E., Heinzmann A., Simma B., Frischer T., Willis-Owen S.A.G., Wong K.C.C., Illig T., Vogelberg C., Weiland S.K., von Mutius E., Abecasis G.R., Farrall M., Gut I.G., Lathrop G.M. & Cookson W.O.C. 2007. Genetic variants regulating ORMDL3 expression contribute to the risk of childhood asthma. Nature **448:** 470–473. doi: 10.1038/nature06014

Morahan G., Peeva V., Mehta M. & Williams R. 2008. Systems genetics can provide new insights in to immune regulation and autoimmunity. Journal of Autoimmunity **31:** 233–236. doi: 10.1016/j.jaut.2008.04.011

Myers A.J., Gibbs J.R., Webster J.A., Rohrer K., Zhao A., Marlowe L., Kaleem M., Leung D., Bryden L., Nath P., Zismann V.L., Joshipura K., Huentelman M.J., Hu-Lince D., Coon K.D., Craig D.W., Pearson J.V., Holmans P., Heward C.B., Reiman E.M., Stephan D. & Hardy J. 2007. A survey of genetic human cortical gene expression. Nat. Genet. **39:** 1494–1499. doi: 10.1038/ng.2007.16

Nadeau J.H. & Dudley A.M. 2011. Systems genetics. Science **331:** 1015–1016. doi: 10.1126/science.1203869

Nadeau J.H., Singer J.B., Matin A. & Lander E.S. 2000. Analysing complex genetic traits with chromosome substitution strains. Nat. Genet. **24:** 221–225. doi: 10.1038/73427

Narain P. 2010. Quantitative genetics: past and present. Mol. Breeding **26:** 135–143. doi: 10.1007/s11032-010-9406-4

Neto E.C., Keller M.P., Attie A.D. & Yandell B.S. 2010. Causal graphical models in systems genetics: a unified framework for joint inference of causal network and genetic architecture for correlated phenotypes. Ann. Appl. Stat. **4:** 320–339. doi: 10.1214/09-AOAS288

Paigen K. & Eppig J.T. 2000. A mouse phenome project. Mamm. Genome **11:** 715–717. doi: 10.1007/s003350010152

Peirce J.L., Li H., Wang J., Manly K.F., Hitzemann R.J., Belknap J.K., Rosen G.D., Goodwin S., Sutter T.R., Williams R.W. & Lu L. 2006. How replicable are mRNA expression QTL? Mamm. Genome **17:** 643–656. doi: 10.1007/s00335-005-0187-8

Pennacchio L., Ahituv N., Moses A., Prabhakar S., Nobrega M., Shoukry M., Minovitsky S., Dubchak I., Holt A., Lewis K.D., Plajzer-Frick I., Akiyama J., Val S.D., Afzal V., Black B.L., Couronne O., Eisen M.B., Visel A. & Rubin E.M. 2006. *In vivo* enhancer analysis of human conserved non-coding sequences. Nature **444:** 499–502. doi: 10.1038/nature05295

Petretto E., Mangion J., Dickens N.J., Cook S.A., Kumaran M.K., Lu H., Fischer J., Maatz H., Kren V., Pravenec M., Hubner N. & Aitman T.J. 2006. Heritability and tissue specificity of expression quantitative trait loci. PLoS Genet. **2:** e172. doi: 10.1371/journal.pgen.0020172

Plaisier C.L., Horvath S., Huertas-Vazquez A., Cruz-Bautista I., Herrera M.F., Tusie-Luna T., Aguilar-Salinas C. & Pajukanta P. 2009. A systems genetics approach implicates USF1, FADS3, and other causal candidate genes for familial combined hyperlipidemia. PLoS Genet. **5:** e1000642. doi: 10.1371/journal.pgen.1000642

Price A.L., Patterson N.J., Plenge R.M., Weinblatt M.E., Shadick N.A. & Reich D. 2006. Principal components analysis corrects for stratification in genome-wide association studies. Nat. Genet. **38:** 904–909. doi: 10.1038/ng1847

Pritchard J.K., Stephens M. & Donnelly P. 2000. Inference of population structure using multilocus genotype data. Genetics **155:** 945–959.

Purcell S.M., Wray N.R., Stone J.L., Visscher P.M., O'Donovan M.C., Sullivan P.F. & Sklar P. 2009. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. Nature **460:** 748–752. doi: 10.1038/nature08185

Raychaudhuri S. 2011. Mapping rare and common causal alleles for complex human diseases. Cell **147:** 57–69. doi: 10.1016/j.cell.2011.09.011

Rockman M.V. & Kruglyak L. 2006. Genetics of global gene expression. Nat. Rev. Genet. **7:** 862–872. doi: 10.1038/nrg1964

Ruden D.M. 2007. Personalized medicine and quantitative trait transcripts. Nat. Genet. **39:** 144–145. doi: 10.1038/ng0207-144

Ruden D.M., Chen L., Possidente D., Possidente B., Rasouli P., Wang L., Lu X., Garfinkel M.D., Hirsch H.V.B. & Page G.P. 2009. Genetical toxicogenomics in *Drosophila* identifies master modulatory loci that are regulated by developmental exposure to lead. Neurotoxicology **30:** 898–914. doi: 10.1016/j.neuro.2009.08.011

Rudi B. 2007. From mouse genetics to systems biology. Mamm. Genome **18:** 383–388. doi: 10.1007/s00335-007-9044-2

Sandve G.K., Ferkingstad E., & Nygard S. 2011. Sequential Monte Carlo multiple testing. Bioinformatics **27**: 3235–3241. doi: 10.1093/bioinformatics/btr568

Schadt E.E., Lamb J., Yang X., Zhu J., Edwards S., GuhaThakurta D., Sieberts S.K., Monks S., Reitman M., Zhang C., Lum P.Y., Leonardson A., Thieringer R., Metzger J.M., Yang L., Castle J., Zhu H., Kash S.F., Drake T.A., Sachs A. & Lusis A.J. 2005. An integrative genomics approach to infer causal associations between gene expression and disease. Nat. Genet. **37:** 710–717. doi: 10.1038/ng1589

Schadt E.E. & Lum P.Y. 2006. Reverse engineering gene networks to identify key drivers of complex disease phenotypes. J. Lipid Res. **47:** 2601–2613. doi: 10.1194/jlr.R600026-JLR200

Schadt E.E., Molony C., Chudin E., Hao K., Yang X., Lum P.Y., Kasarskis A., Zhang B., Wang S., Suver C., Zhu J., Millstein J., Sieberts S., Lamb J., GuhaThakurta D., Derry J., Storey J.D., Avila-Campillo I., Kruger M.J., Johnson J.M., Rohl C.A., van Nas A., Mehrabian M., Drake T.A., Lusis A.J., Smith R.C., Guengerich F.P., Strom S.C., Schuetz E., Rushmore T.H. & Ulrich R. 2008. Mapping the genetic architecture of gene expression in human liver. PLoS Biol. **6:** e107. doi: 10.1371/journal.pbio.0060107

Schadt E.E., Zhang B. & Zhu J. 2009. Advances in systems biology are enhancing our understanding of disease and moving us closer to novel disease treatments. Genetica **136:** 259–269. doi: 10.1007/s10709-009-9359-x

Sieberts S.K. & Schadt E.E. 2007. Moving toward a system genetics view of disease. Mamm. Genome **18:** 389–401. doi: 10.1007/s00335-007-9040-6

Silpa S., Andreas B., Richard M.K., Yonina E. & Trey I. 2008. eQED: an efficient method for interpreting eQTL associations using protein networks. Mol. Sys. Biol. **4:** 162. doi: 10.1038/msb.2008.4

Simonis N., Rual J., Carvunis A., Tasan M., Lemmens I., Hirozane-Kishikawa T., Hao T., Sahalie J.M., Venkatesan K., Gebreab F., Cevik S., Klitgord N., Fan C., Braun P., Li N., Ayivi-Guedehoussou N., Dann E., Bertin N., Szeto D., Dricot A., Yildirim M.A., Lin C., de Smet A.S., Kao H.L., Simon C., Smolyar A., Ahn J.S., Tewari M., Boxem M., Milstein S., Yu H., Dreze M., Vandenhaute J., Gunsalus K.C., Cusick M.E., Hill D.E., Tavernier J., Roth F.P. & Vidal M. 2009. Empirically controlled mapping of the *Caenorhabditis elegans* protein-protein interactome network. Nat. Meth. **6:** 47–54. doi: 10.1038/nmeth.1279

Sindi S., Helman E., Bashir A. & Raphael B.J. 2009. A geometric approach for classification and comparison of structural variants. Bioinformatics **25:** i222–i230. doi: 10.1093/bioinformatics/btp208

Sladek R., Rocheleau G., Rung J., Dina C., Shen L., Serre D., Boutin P., Vincent D., Belisle A., Hadjadj S., Balkau B., Heude B., Charpentier G., Hudson T.J., Montpetit A., Pshezhetsky A.V., Prentki M., Posner B.I., Balding D.J., Meyre D., Polychronakos C. & Froguel P. 2007. A genome-wide association study identifies novel risk loci for type 2 diabetes. Nature **445:** 881–885. doi: 10.1038/ nature05616

Stefansson H., Rujescu D., Cichon S., Pietiläinen O.P.H., Ingason A., Steinberg S., Fossdal R., Sigurdsson E., Sigmundsson T., Buizer-Voskamp J.E., Hansen T., Jakobsen K.D., Muglia P., Francks C., Matthews P.M., Gylfason A., Halldorsson B.V., Gudbjartsson D., Thorgeirsson T.E., Sigurdsson A., Jonasdottir A., Jonasdottir A., Bjornsson A., Mattiasdottir S., Blondal T., Haraldsson M., Magnusdottir B.B., Giegling I., Möller H.J., Hartmann A., Shianna K.V., Ge D., Need A.C., Crombie C., Fraser G., Walker N., Lonnqvist J., Suvisaari J., Tuulio-Henriksson A., Paunio T., Toulopoulou T., Bramon E., Forti M.D., Murray R., Ruggeri M., Vassos E., Tosato S., Walshe M., Li T., Vasilescu C., Mühleisen T.W., Wang A.G., Ullum H., Djurovic S., Melle I., Olesen J., Kiemeney L.A., Franke B., GROUP, Sabatti C., Freimer N.B., Gulcher J.R., Thorsteinsdottir U., Kong A., Andreassen O.A., Ophoff R.A., Georgi A., Rietschel M., Werge T., Petursson H., Goldstein D.B., Nöthen M.M., Peltonen L., Collier D.A., St Clair D. & Stefansson K. 2008. Large recurrent microdeletions associated with schizophrenia. Nature **455:** 232–237. doi: 10.1038/nature07229

Stranger B.E., Forrest M.S., Dunning M., Ingle C.E., Beazley C., Thorne N., Redon R., Bird C.P., Grassi A., Lee C., Tyler-Smith C., Carter N., Scherer S.W., Tavaré S., Deloukas P., Hurles M.E. & Dermitzakis E.T. 2007a. Relative impact of nucleotide and copy number variation on gene expression phenotypes. Science **315:** 848–853. doi: 10.1126/science.1136678

Stranger B., Nica A., Forrest M., Dimas A., Bird C., Beazley C., Ingle C.E., Dunning M., Flicek P., Koller D., Montgomery S., Tavaré S., Deloukas P. & Dermitzakis E.T. 2007b. Population genomics of human gene expression. Nat. Genet. **39:** 1217–1224. doi: 10.1038/ng2142

Su W.L., Sieberts S.K., Kleinhanz R.R., Lux K., Millstein J., Molony C. & Schadt E.E. 2010. Assessing the prospects of genome-wide association studies performed in inbred mice. Mamm. Genome **21:** 143–152. doi: 10.1007/s00335-010-9249-7

Swami W. 2009. Systems genetics: networking complex traits. Nat. Rev. Genet. **10:** 219. doi: 10.1038/nrg2566

The FANTOM Consortium & Riken Omics Science Center. 2009. The transcriptional network that controls growth arrest and differentiation in a human myeloid leukemia cell line. Nat. Genet. **41:** 553–562. doi: 10.1038/ng.375

The International Schizophrenia Consortium. 2008. Rare chromosomal deletions and duplications increase risk of schizophrenia. Nature **455:** 237–241. doi: 10.1038/nature07239

Veyrieras J.B., Kudaravalli S., Kim S.Y., Dermitzakis E.T., Gilad Y., Stephens M. & Pritchard J.K. 2008. High-resolution mapping of expression-QTLs yields insight into human gene regulation. PLoS Genet. **4:** e1000214. doi: 10.1371/journal.pgen.1000214

Vineis P. & Pearce N. 2010. Missing heritability in genome-wide association study research. Nat. Rev. Genet. **11:** 589. doi: 10.1038/nrg2809-c2

Williams R.W. 2006. Expression genetics and the phenotype revolution. Mamm. Genome **17:** 496–502. doi: 10.1007/s00335-006-0006-x

Workman C.T., Mak H.C., McCuine S., Tagne J., Agarwal M., Ozier O., Begley T.J., Samson L.D. & Ideker T. 2006. A systems approach to mapping DNA damage response pathways. Science **312:** 1054–1059. doi: 10.1126/science.1122088

Wray G. 2007. The evolutionary significance of *cis*-regulatory mutations. Nat. Rev. Genet. **8:** 206–216. doi: 10.1038/nrg2063

Xie C. & Tammi M. 2009. CNV-seq, a new method to detect copy number variation using high-throughput sequencing. BMC Bioinformatics **10:** 80. doi: 10.1186/1471-2105-10-80

Xie X., Mikkelsen T., Gnirke A., Lindblad-Toh K., Kellis M. & Lander E.S. 2007. Systematic discovery of regulatory motifs in conserved regions of the human genome, including thousands of CTCF insulator sites. Proc. Natl. Acad. Sci. USA **104:** 7145–7150. doi:10.1073/pnas.0701811104

Xiong D.H., Long J.R., Recker R.R. & Deng H.W. 2003. Pharmacogenomic approaches to osteoporosis. Pharmacogenom. J. **3:** 261–263. doi:10.1038/sj.tpj. 6500199

Yang H.C., Liang Y.J., Wu Y.L., Chung C.M., Chiang K.M., Ho H.Y., Ting C.T., Lin T.H., Sheu S.H., Tsai W.C., Chen J.H., Leu H.B., Yin W.H., Chiu T.Y., Chen C.I., Fann C.S.J., Wu J.Y., Lin T.N., Lin S.J., Chen Y.T., Chen J.W. & Pan W.H. 2009. Genome-wide association study of young-onset hypertension in the Han Chinese population of Taiwan. PLoS ONE **4:** e5459. doi:10.1371/journal.pone.0005459

Ye K., Schulz M.H., Long Q., Apweiler R. & Ning Z. 2009. Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. Bioinformatics **25:** 2865–2871. doi: 10.1093/bioinformatics/btp394

Yu J., Pressoir G., Briggs W.H., Vroh B.I., Yamasaki M., Doebley J.F., McMullen M.D., Gaut B.S., Nielsen D.M., Holland J.B., Kresovich S. & Buckler E.S. 2005. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat. Genet. **38:** 203–208. doi: 10.1038/ng1702

Zhang Z., Ersoz E., Lai C.Q., Todhunter R.J., Tiwari H.K., Gore M.A., Bradbury P.J., Yu J., Arnett D.K., Ordovas J.M. & Buckler E.S. 2010. Mixed linear model approach adapted for genome- wide association studies. Nat. Genet. **42:** 355–360. doi: 10.1038/ng.546

Zhang Z.D., Du J., Lam H., Abyzov A., Urban A.E., Snyder M. & Gerstein M. 2011. Identification of genomic indels and structural variations using split reads. BMC Genomics **12:** 375. doi: 10.1186/1471-2164-12-375

Zhu C., Li X. & Yu J. 2011. Integrating rare-variant testing, function prediction, and gene network in composite resequencing-based genome-wide association studies (CR-GWAS). G3 Genes Genomes Genetics **1:** 233–243. doi: 10.1534/g3.111.000364

Received June 12, 2011
Accepted February 14, 2012