

## Review



**Cite this article:** Nica AC, Dermitzakis ET.  
2013 Expression quantitative trait loci: present  
and future. *Phil Trans R Soc B* 368: 20120362.  
<http://dx.doi.org/10.1098/rstb.2012.0362>

One contribution of 12 to a Discussion Meeting  
Issue 'Regulation from a distance: long-range  
regulation of gene expression'.

### Subject Areas:

genetics, genomics, systems biology

### Keywords:

expression quantitative trait loci, regulation,  
genetics

### Author for correspondence:

Emmanouil T. Dermitzakis  
e-mail: [emmanouil.dermitzakis@unige.ch](mailto:emmanouil.dermitzakis@unige.ch)

# Expression quantitative trait loci: present and future

Alexandra C. Nica<sup>1,2,3</sup> and Emmanouil T. Dermitzakis<sup>1,2,3</sup>

<sup>1</sup>Department of Genetic Medicine and Development, University of Geneva Medical School,  
Geneva 1211, Switzerland

<sup>2</sup>Institute of Genetics and Genomics of Geneva (iGE3), 1211 Geneva, Switzerland

<sup>3</sup>Swiss Institute of Bioinformatics, 1 Rue Michel-Servet, 1211 Geneva, Switzerland

The last few years have seen the development of large efforts for the analysis of genome function, especially in the context of genome variation. One of the most prominent directions has been the extensive set of studies on expression quantitative trait loci (eQTLs), namely, the discovery of genetic variants that explain variation in gene expression levels. Such studies have offered promise not just for the characterization of functional sequence variation but also for the understanding of basic processes of gene regulation and interpretation of genome-wide association studies. In this review, we discuss some of the key directions of eQTL research and its implications.

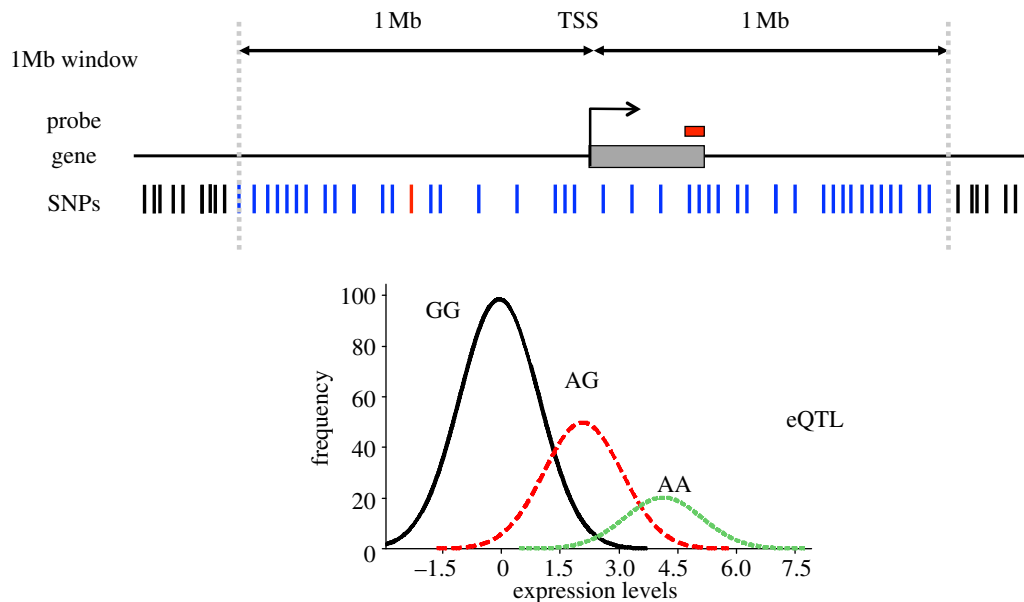
## 1. Introduction

Genome variability has been the focus of many studies in recent years due to its relevance to the differential disease risk among individuals. One of the fundamental needs for the interpretation of the effects of genome variants is the understanding of the specific biological effect such variants have in the cell, which provides a handle to the biology of the disease or organismal phenotype. Genome-wide association studies (GWAS) [1] have demonstrated that the majority of such variants are found in non-coding regions of the genome and are therefore likely to be involved in gene regulation. The analysis of such variants in the context of gene expression measured in cells or tissues has spawned a big field in human genetics studying expression quantitative trait loci (eQTLs).

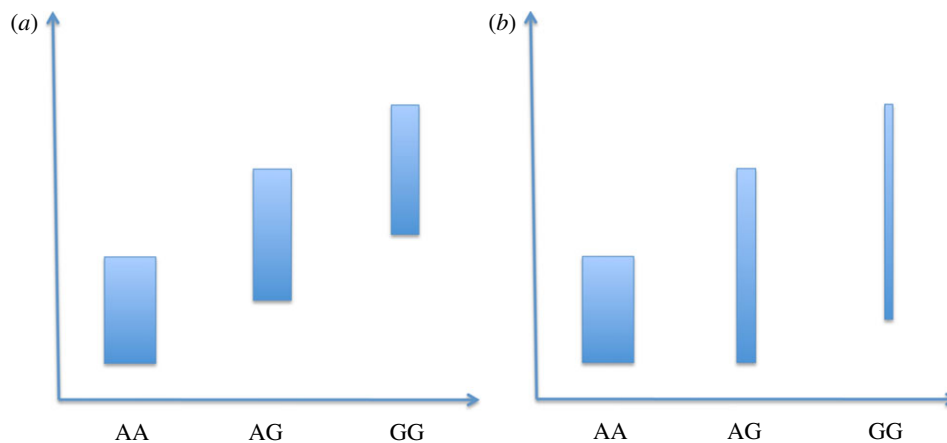
An eQTL is a locus that explains a fraction of the genetic variance of a gene expression phenotype. Standard eQTL analysis involves a direct association test between markers of genetic variation with gene expression levels typically measured in tens or hundreds of individuals. This association analysis can be performed proximally or distally to the gene. One of the major advantages of eQTL mapping using the GWAS approach is that it permits the identification of new functional loci without requiring any previous knowledge about specific *cis* or *trans* regulatory regions. Typically in the eQTL mapping literature, regulatory variants have been characterized as either *cis* or *trans* acting, reflecting the predicted nature of interactions and of course depending on the physical distance from the gene they regulate. In this review, conventionally, variants within 1 Mb (megabase) on either side of a gene's TSS were called *cis* (figure 1), while those at least 5 Mb downstream or upstream of the TSS or on a different chromosome were considered *trans* acting.

A number of studies have focused on eQTLs in model organisms, such as yeast [2,3], and this has provided a nice framework of basic knowledge that still informs studies in human populations.

Studies so far indicate that most of the regulatory control takes place locally, in the vicinity of genes [4–6]. Numerous genes were detected to have *cis* eQTLs (831 genes had a significant *cis* eQTL in a study performed by our laboratory on 270 lymphoblastoid cell lines derived from HapMap 2 individuals genotyped for 2.2. million common SNPs) [7]. As power increases with the availability of larger sample sizes, the number of genes detected to have eQTLs is also expected to increase. In fact, the recent use of transcriptome sequencing as well as the ability to correct for latent confounding factors has allowed a



**Figure 1.** A typical eQTL; many SNPs tested against levels of expression measured by a probe or by other means. The panel below illustrates the difference in distributions of expression values stratified by the SNP genotype of the most significant SNP. (Online version in colour.)



**Figure 2.** Correlation between genotype and expression levels in two populations as indicated by a boxplot. The same SNP is highly significantly correlated in (a) population 1 but not correlated in (b) population 2. This is mainly due the population minor allele frequency of the SNP. Box width is proportional to sample size for the corresponding genotypic category. (Online version in colour.)

substantial increase in power bringing the numbers to thousands of eQTLs even from about 100 individuals [8,9]. Finding *trans* eQTLs has been less successful so far, mainly because interrogating the whole-genome for potential regulatory effects is a daunting statistical and computational task. Whether the current enrichment of *cis* versus *trans* eQTLs reflects biological reality and is not just attributable to low power in *trans* is still under debate [10,11]. However, recent studies have shown that when a reasonable sample size is tested hundreds to thousands of replicated *trans* eQTLs could be found, and they tend to be very tissue specific [12,13].

## 2. Population differentiation of gene expression

Several studies have analysed expression data in populations of different ancestry and revealed substantial differences at many loci. A study on 16 individuals of European and African descent estimated that 17 per cent of genes were differentially expressed between populations [14].

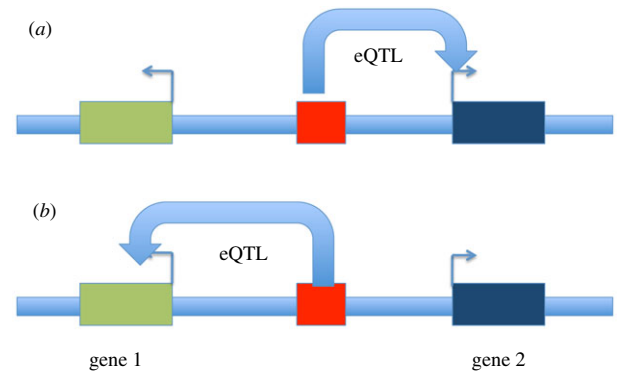
Differences were found also between European and Asian-derived populations for 1097 of 4197 genes tested [15]. Larger scale studies confirmed the initial estimates. The eQTL analysis on 270 individuals of the four worldwide HapMap II populations reported that 17–29% of loci have significant differences in mean expression levels between population pairs [7]. While some of these observations are due to environmental factors [16], genetics plays an important role in shaping the observed differences (figure 2). Price *et al.* [17] provide evidence for population differentiation due to genetic effects using cell lines derived from an admixed African American population. They estimated a mean value of 0.2 and a median of 0.12 in the proportion of gene expression variation attributable to population differences. A large proportion of the genetically determined variation in gene expression across populations has been explained by different allele frequencies [15], suggesting that regulatory mechanisms are probably not fundamentally different between populations. Finally, recent studies have shown that common regulatory effects are largely shared among populations [9].

### 3. Multiple-tissue studies

So far, the majority of human eQTL studies have been performed exclusively on blood-derived cells or cell lines. This relatively easily accessible cell-type has been very useful in understanding the genetics of gene expression and continues to be a great resource in other cohort studies. However, as gene expression signatures are cell-type specific [18], the question arises whether regulatory control of expression is also cell-type-dependent. Estimates vary depending on the tissues being compared and the eQTL methods used, but generally, a significant tissue-specific component of *cis* regulation has been systematically reported. In a study comparing adipose and blood expression patterns in two Icelandic cohorts, 50 per cent of the *cis* eQTLs detected were shared [19]. A comparison of cortical tissue and LCL regulatory overlap in a European population showed barely any overlap, albeit this difference is probably exacerbated by the different microarray platforms used in the two experiments [20]. Another study comparing eQTLs from autopsy-derived cortical tissue and peripheral blood mononucleated cells found less than 50 per cent sharing [21]. A study comparing the regulatory landscape in three tissues (fibroblasts, LCLs and primary T cells) derived from the same set of 75 European individuals reported that 69–80% of *cis* eQTLs are cell-type specific, augmenting thus the need to study multiple tissues to determine the full spectrum of regulatory variants [22]. Finally, recent studies with well-powered design, such as hundreds of individuals and twin structure have demonstrated that there is a rate of diminishing returns when sample size increases, arguing that there is significant tissue specificity but less than previously estimated and a tissue-dependent effect size that can be detected when the study is well powered [13,23].

### 4. Relevance of expression tissue for disease

Documenting cell-type specific regulatory variation is very important from the disease perspective. Integrating expression data with GWAS results can be informative for discovering genes and pathways whose disruption probably causes disease [24–26]. However, this is possible only when the tissue of expression is relevant to the interrogated complex trait [25]. eQTLs discovered in LCLs have helped explain GWAS associations with childhood asthma [27] and Crohn's disease [28], two autoimmune inflammatory disorders. The adipose and blood cohorts analysed by Emilsson *et al.* had been assessed for various phenotypes too, including obesity relevant traits. Notably, 50 per cent of the *cis* signals were estimated as overlapping between the two cohorts, but a marked correlation with obesity-related traits was only observed for gene expression measured in adipose tissue [19]. These observations certify the importance of integrating data from a relevant tissue when trying to interpret GWAS results using gene expression as an intermediate phenotype. An important caveat is that in several cases the same regulatory region and variant will be linked to one gene in one tissue and another gene in another tissue (figure 3). This raises the potential that limited tissue interrogation will give misleading biological interpretations about the gene mediating the regulatory effect to increase disease risk. Nevertheless, it is still unclear what the pattern of



**Figure 3.** The same regulatory regions and variant could be an eQTL for gene 2 in (a) tissue 1 and for gene 1 in (b) tissue 2, suggesting that limited interrogation of tissues would be misleading for the biological signal underlying disease. (Online version in colour.)

diminishing returns is across human tissues and what tissues could serve as highly informative in large cohorts. For example, LCLs have been useful in less-expected cases enabling candidate gene discovery for associations with autism [29] or bipolar disorder [30].

### 5. Promise of expression quantitative trait loci studies for disease genetics

Despite the impressive success of GWAS, there is a substantial gap between the susceptibility variants discovered and understanding how those respective loci contribute to disease. Frequently, such loci map to genomic regions of no apparent function (non-coding) or the genome's tight correlation structure (LD) does not permit firm conclusions about functional effects (i.e. which is the causal variant and the function of which gene does it affect). Under these circumstances, the need of incorporating additional information for interpreting GWAS results became apparent. The direct link between DNA polymorphisms (usually SNPs) and variable transcript levels along with the increasing role attributed to regulatory variation in shaping phenotypic differences nominated gene expression as an important mechanism underlying complex traits. Subsequently, we describe the main results obtained so far in support of this hypothesis.

### 6. Genome-wide association study snps can be strong expression quantitative trait loci

Comparing expression levels of individual genes between cases and controls may not be sufficiently powered to detect significant differences [31] and discriminating between causal and reactive expression changes would be a tough challenge. However, genetic markers simultaneously associated with disease status and eQTLs are very interesting: if one allele is more frequent in cases than controls and at the same time, it is causal for gene expression effects of a nearby gene, which is by itself important for the disease, then it is probable that causality can be established. Several recent studies have shown the value of this principle by incorporating eQTL analyses with GWAS results and thus proposing candidate disease genes. Moffatt *et al.* [27]

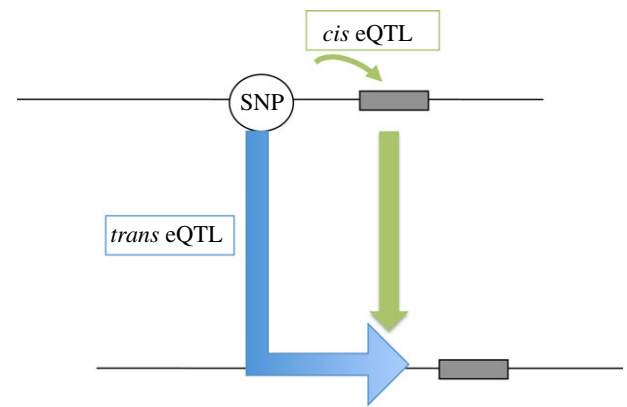
identified a series of strongly correlated SNPs in a 200 kb region of chromosome 17q23 associated with childhood asthma. The association region contained 19 genes, none of which had an evident disease role. Expression analysis on lymphoblastoid cell lines derived from the same families showed that the most significant GWAS SNPs also explained approx. 29.5 per cent of the variance in transcript levels of one of those 19 genes, *ORMDL3* (*ORM1*-like 3), now the best candidate for further functional studies.

Expression data have helped interpret some of the association signals for Crohn's disease as well. Initial findings of a recent GWAS included multiple susceptibility loci mapping to a 1.25 Mb gene desert region on chromosome 5 [32]. eQTL data showed that one or more of these loci act as long-range *cis* regulators of *PTGER4* (prostaglandin E receptor 4), a gene 270 kb away from the associated region whose homologue has been implicated in phenotypes similar to Crohn's disease in the mouse [28]. Similar other examples for height [33], systemic lupus erythematosus [34], type I diabetes [35] or bipolar disorder [36] support the use of eQTL data in aiding the interpretation of GWAS results.

However, not all cases are so straightforward, as shown by the association of the *SH2B1* (*SH2B* adaptor protein 1) locus to body mass index (BMI) [37]. In this case, a non-synonymous genome-wide significant SNP in *SH2B1* was associated also with differential expression of two other genes (*EIF3C*—eukaryotic translation initiation factor 3, subunit C and *TUFM* translation elongation factor, mitochondrial). Functional evidence from mice, where mutating a *SH2B1* homologue leads to extreme obesity, strengthens the hypothesis that the missense SNP is the actual functional variant, which is in high LD with a different causal regulatory variant affecting *EIF3C* and *TUFM* expression. This is a typical example of a coincidental overlap of GWAS and eQTL results, which must be carefully distinguished from causal cases where both the GWAS SNP and the eQTL tag the same functional variant. Given the ubiquitous nature of regulatory variants [7] and hence the high probability of such coincidental overlaps, integrative methods pinpointing true causal regulatory effects are desirable [25]. Finally, a recent study has shown the value of integrating *cis* and *trans* eQTLs together with GWAS analysis to detect previously unknown determinants of disease, such as the *KLF14* gene. Via eQTL analysis and by adding additional data from mice, it was inferred that *KLF14* is a likely signal from fat to other diabetes-related tissues to induce insulin resistance [38]. Nevertheless, since many traits manifest themselves only in certain tissues, such methods are only informative if expression measurements from disease-relevant cell-types are compared.

## 7. Gene regulatory networks

The large-scale disease studies performed so far have uncovered multiple variants of low-effect sizes affecting multiple genes. This suggests that common forms of disease are most probably not the result of single gene changes with a single outcome, but rather the outcome of perturbations of gene networks which are affected by complex genetic and environmental interactions [39]. A simple view of a network component can be viewed as a *cis* effect that transmits its signal in *trans* essentially making the *cis* SNP a *trans* SNP



**Figure 4.** Network relationships based on *cis* and *trans* eQTLs. Inference of network effects based on genetic effects allow for the determination of the direction of the effect through the network, which refines the network relationships much more finely. (Online version in colour.)

as well (figure 4). The numerous genetic factors involved in disease predisposition appear randomly distributed across the genome, but the expectation is that they are functionally linked and that these functional interactions are useful in prioritizing disease genes [40]. DNA sequencing of tumour samples from pancreatic and brain cancer, respectively, provided supporting evidence for this principle by identifying candidate genes belonging largely to core pathways involved in tumorigenesis or tumour progression [41,42]. Recently, analysis of gene regulatory networks has offered important insight into complex disease mechanisms. In a study integrating co-expression networks and genotypic data from an F2 intercross population, Chen *et al.* [24] identified a liver and adipose macrophage-enriched sub-network (MEMN) associated with metabolic syndrome relevant traits. Three genes in this network, lipoprotein lipase (*Lpl*), lactamase  $\beta$  (*Lactb*) and protein phosphatase 1-like (*Ppm1l*) were validated by gene knockouts as causal obesity genes, strengthening the association of MEMN to phenotypes characteristic to metabolic syndrome. A parallel study in humans identified a homologous transcriptional network constructed from adipose data, having substantial overlap with MEMN sub-modules and being enriched for genes involved in inflammatory and immune response [19]. Subsequent eQTL mapping identified *cis*-regulatory variants affecting specific genes in this network and the joint analysis of the strongest *cis* eQTLs revealed substantial enrichment for variants associated with obesity-related clinical traits. Classical genetic approaches would not be able to detect such variants with small individual effects. Identifying them as a group affecting gene networks which—when perturbed—result in a disease state, is in this case much better powered.

## 8. Conclusions and future directions

The emergence of eQTLs has transformed the field of human genetics by providing a comprehensive, easily accessible and interpretable molecular link between genetic variation and organismal phenotypes. Its use to bring better biological context in disease studies has motivated additional studies to expand the range of molecular phenotypes to be tested in large samples and multiple tissues



leading to a new field of genetics and genomics of cellular phenotypes. This new direction is likely to transform not only the way we learn biology, especially in non-model organisms such as humans where experiments are not possible in whole organisms, but also the way we implement biomedical information to the medical setting and practice. The future sees large projects interrogating large numbers of tissues in large numbers of individuals, such as the Genotype–Tissue Expression project which is a major NIH initiative to sample large numbers of autopsy tissues from

hundreds of individuals (<http://commonfund.nih.gov/GTEX/>). In addition, recent sample collections in cohorts almost always include biological material such as blood and other accessible tissues so that eQTL and other expression analyses can be performed. The development of reference datasets for eQTLs and other molecular phenotype variants will largely strengthen the interpretation of personalized genomes and will provide a valuable framework for the biological understanding of phenotypic variability and disease risk.

## References

- Manolio TA. 2010 Genomewide association studies and assessment of the risk of disease. *N. Engl. J. Med.* **363**, 166–176. (doi:10.1056/NEJMr0905980)
- Brem RB, Storey JD, Whittle J, Kruglyak L. 2005 Genetic interactions between polymorphisms that affect gene expression in yeast. *Nature* **436**, 701–703. (doi:10.1038/nature03865)
- Brem RB, Yvert G, Clinton R, Kruglyak L. 2002 Genetic dissection of transcriptional regulation in budding yeast. *Science* **296**, 752–755. (doi:10.1126/science.1069516)
- Goring HH *et al.* 2007 Discovery of expression QTLs using large-scale transcriptional profiling in human lymphocytes. *Nat. Genet.* **39**, 1208–1216. (doi:10.1038/ng2119)
- Schadt EE *et al.* 2008 Mapping the genetic architecture of gene expression in human liver. *PLoS Biol.* **6**, e107. (doi:10.1371/journal.pbio.0060107)
- Dixon AL *et al.* 2007 A genome-wide association study of global gene expression. *Nat. Genet.* **39**, 1202–1207. (doi:10.1038/ng2109)
- Stranger BE *et al.* 2007 Population genomics of human gene expression. *Nat. Genet.* **39**, 1217–1224. (doi:10.1038/ng2142)
- Montgomery SB, Sammeth M, Gutierrez-Arcelus M, Lach RP, Ingle C, Nisbett J, Guigo R, Dermizakis ET. 2010 Transcriptome genetics using second generation sequencing in a Caucasian population. *Nature* **464**, 773–777. (doi:10.1038/nature08903)
- Stranger BE *et al.* 2012 Patterns of cis regulatory variation in diverse human populations. *PLoS Genet.* **8**, e1002639. (doi:10.1371/journal.pgen.1002639)
- Wray GA. 2007 The evolutionary significance of cis-regulatory mutations. *Nat. Rev. Genet.* **8**, 206–216. (doi:10.1038/nrg2063)
- Wittkopp PJ, Haerum BK, Clark AG. 2008 Regulatory changes underlying expression differences within and between *Drosophila* species. *Nat. Genet.* **40**, 346–350. (doi:10.1038/ng.77)
- Franceschini N *et al.* 2012 Discovery and fine mapping of serum protein loci through transethnic meta-analysis. *Am. J. Hum. Genet.* **91**, 744–753. (doi:10.1016/j.ajhg.2012.08.021)
- Grundberg E *et al.* 2012 Mapping cis- and trans-regulatory effects across multiple tissues in twins. *Nat. Genet.* **44**, 1084–1089. (doi:10.1038/ng.2394)
- Storey JD, Madeoy J, Strout JL, Wurfel M, Ronald J, Akey JM. 2007 Gene-expression variation within and among human populations. *Am. J. Hum. Genet.* **80**, 502–509. (doi:10.1086/512017)
- Spielman RS, Bastone LA, Burdick JT, Morley M, Ewens WJ, Cheung VG. 2007 Common genetic variants account for differences in gene expression among ethnic groups. *Nat. Genet.* **39**, 226–231. (doi:10.1038/ng1955)
- Idaghdour Y, Storey JD, Jadallah SJ, Gibson G. 2008 A genome-wide gene expression signature of environmental geography in leukocytes of Moroccan Amazighs. *PLoS Genet.* **4**, e1000052. (doi:10.1371/journal.pgen.1000052)
- Price AL, Patterson N, Handks DC, Myers S, Reich D, Cheung VG, Spielman RS, Gibson G. 2008 Effects of cis and trans genetic ancestry on gene expression in African Americans. *PLoS Genet.* **4**, e1000294. (doi:10.1371/journal.pgen.1000294)
- Alberts B. 2002 *Molecular biology of the cell*, 4th edn. New York, NY: Garland Science.
- Emilsson V *et al.* 2008 Genetics of gene expression and its effect on disease. *Nature* **452**, 423–428. (doi:10.1038/nature06758)
- Myers AJ *et al.* 2007 A survey of genetic human cortical gene expression. *Nat. Genet.* **39**, 1494–1499. (doi:10.1038/ng.2007.16)
- Heinzen EL *et al.* 2008 Tissue-specific genetic control of splicing: implications for the study of complex traits. *PLoS Biol.* **6**, e1000001. (doi:10.1371/journal.pbio.1000001)
- Dimas AS *et al.* 2009 Common regulatory variation impacts gene expression in a cell type-dependent manner. *Science* **325**, 1246–1250. (doi:10.1126/science.1174148)
- Nica AC *et al.* 2011 The architecture of gene regulatory variation across multiple human tissues: the MuTHER study. *PLoS Genet.* **7**, e1002003. (doi:10.1371/journal.pgen.1002003)
- Chen Y *et al.* 2008 Variations in DNA elucidate molecular networks that cause disease. *Nature* **452**, 429–435. (doi:10.1038/nature06757)
- Nica AC, Dermizakis ET. 2008 Using gene expression to investigate the genetic basis of complex disorders. *Hum. Mol. Genet.* **17**, R129–134. (doi:10.1093/hmg/ddn285)
- Nica AC, Montgomery SB, Dimas AS, Stranger BE, Beazley C, Barroso I, Dermizakis ET, Gibson G. 2010 Candidate causal regulatory effects by integration of expression QTLs with complex trait genetic associations. *PLoS Genet.* **6**, e1000895. (doi:10.1371/journal.pgen.1000895)
- Moffatt MF *et al.* 2007 Genetic variants regulating ORM3L3 expression contribute to the risk of childhood asthma. *Nature* **448**, 470–473. (doi:10.1038/nature06014)
- Libioulle C *et al.* 2007 Novel Crohn disease locus identified by genome-wide association maps to a gene desert on 5p13.1 and modulates expression of PTGER4. *PLoS Genet.* **3**, e58. (doi:10.1371/journal.pgen.0030058)
- Nishimura Y *et al.* 2007 Genome-wide expression profiling of lymphoblastoid cell lines distinguishes different forms of autism and reveals shared pathways. *Hum. Mol. Genet.* **16**, 1682–1698. (doi:10.1093/hmg/ddm116)
- Iwamoto K, Bundo M, Washizuka S, Kakiuchi C, Kato T. 2004 Expression of HSPF1 and LIM in the lymphoblastoid cells derived from patients with bipolar disorder and schizophrenia. *J. Hum. Genet.* **49**, 227–231. (doi:10.1007/s10038-004-0136-5)
- Cookson W, Liang L, Abecasis G, Moffatt M, Lathrop M. 2009 Mapping complex disease traits with global gene expression. *Nat. Rev. Genet.* **10**, 184–194. (doi:10.1038/nrg2537)
- Barrett JC *et al.* 2008 Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease. *Nat. Genet.* **40**, 955–962. (doi:10.1038/ng.175)
- Gudbjartsson DF *et al.* 2008 Many sequence variants affecting diversity of adult human height. *Nat. Genet.* **40**, 609–615. (doi:10.1038/ng.122)
- Hom G *et al.* 2008 Association of systemic lupus erythematosus with C8orf13-BLK and ITGAM-ITGAX. *N. Engl. J. Med.* **358**, 900–909. (doi:10.1056/NEJMoa0707865)
- Hakonarson H *et al.* 2007 A genome-wide association study identifies KIAA0350 as a type 1 diabetes gene. *Nature* **448**, 591–594. (doi:10.1038/nature06010)

36. WTCCC. 2007 Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* **447**, 661–678. (doi:10.1038/nature05911)
37. Willer CJ *et al.* 2009 Six new loci associated with body mass index highlight a neuronal influence on body weight regulation. *Nat. Genet.* **41**, 25–34. (doi:10.1038/ng.287)
38. Small KS *et al.* 2011 Identification of an imprinted master trans regulator at the KLF14 locus related to multiple metabolic phenotypes. *Nat. Genet.* **43**, 561–564. (doi:10.1038/ng.833)
39. Schadt EE. 2009 Molecular networks as sensors and drivers of common human diseases. *Nature* **461**, 218–223. (doi:10.1038/nature08454)
40. Franke L, van Bakel H, Fokkens L, de Jong ED, Egmont-Petersen M, Wijmenga C. 2006 Reconstruction of a functional human gene network, with an application for prioritizing positional candidate genes. *Am. J. Hum. Genet.* **78**, 1011–1025. (doi:10.1086/504300)
41. Jones S *et al.* 2008 Core signaling pathways in human pancreatic cancers revealed by global genomic analyses. *Science* **321**, 1801–1806. (doi:10.1126/science.1164368)
42. Parsons DW *et al.* 2008 An integrated genomic analysis of human glioblastoma multiforme. *Science* **321**, 1807–1812. (doi:10.1126/science.1164382)