PROGRESS

From expression QTLs to personalized transcriptomics

Stephen B. Montgomery and Emmanouil T. Dermitzakis

Abstract | Approaches that combine expression quantitative trait loci (eQTLs) and genome-wide association (GWA) studies are offering new functional information about the aetiology of complex human traits and diseases. Improved study designs — which take into account technological advances in resolving the transcriptome, cell history and state, population of origin and diverse endophenotypes — are providing insights into the architecture of disease and the landscape of gene regulation in humans. Furthermore, these advances are helping to establish links between cellular effects and organismal traits.

Genome-wide association (GWA) studies have both the advantage and disadvantage that they are agnostic to the mechanism by which genetic variation influences the phenotype being tested. The disadvantage is particularly acute for complex phenotypes that are associated with multiple variants of small to intermediate effect sizes, which only explain a modest proportion of the phenotype's measured heritability[1]. In these cases there are two key challenges: first, to establish which gene functions are affected by the associated variants or markers and, second, to identify the causal variant, a task that is complicated by the linkage disequilibrium (LD) structure of the human genome. To address these challenges, it is necessary to obtain additional functional information that might explain the effect of an individual variant on specific gene(s). One way to enrich for variants that are likely to have causal effects is to account for the genetic influence of the variant on gene expression. Such effects are increasingly recognized as being of fundamental importance, as it has become apparent that in most GWA studies, common variants are found in non-coding regions[2]. Furthermore, recent studies have shown that the currently known GWA study signals are enriched for expression quantitative trait loci (eQTLs) in lymphoblastoid cell lines[3,4], and this pattern is likely to be repeated as we interrogate more tissues and cell types.

In this article, we explore how evolving designs of eQTL studies, facilitated by advances in genotyping and gene-expression-based technologies, are increasingly able to investigate the role of regulatory variation in different biological contexts. We highlight how such advances are further extending our understanding of the regulatory landscape by uncovering the degree to which cell-type and temporal specificity are represented in gene-regulatory architecture.

> “eQTL studies have illuminated the location and impact of regulatory variants, the tissue specificity of regulatory elements and the contribution of *cis*- versus *trans*-acting variation to the expression of any given gene.”

Furthermore, we demonstrate how this information is being used to uncover the molecular basis of complex traits at different layers of complexity (from individual transcripts to regulatory networks). Last, we propose how further integrated and large-scale measurements of cellular phenotypes, such as gene expression, will help to identify predictive biomarkers of disease.

eQTL studies: design and interpretation

The design and interpretation of eQTL studies is not straightforward. We describe four important parameters that are being incorporated into recent studies: increased sequence resolution of the genome and transcriptome, cellular history and state, population structure, and specific endophenotypes.

Increasing resolution through sequencing. Genome-wide mRNA sequencing (RNA-seq) is providing an enhanced framework for translating regulatory variants as studies can now obtain better resolution of the transcriptome when seeking eQTLs. In a recent example, RNA-seq was used to assess genetic-based differences in gene expression in a cohort of 60 Europeans and 69 Africans[5,6]. The technique provided a much improved ability to identify genes whose expression contains a genetic component as well as variants associated with differential abundance of alternatively spliced transcripts. Furthermore, it allowed genotypic association to be integrated with allelic association because, for each individual, we obtain both the overall abundance of the transcript and the relative abundance of the transcripts from the two alleles. Using RNA-seq reads that map to heterozygote positions in multiple individuals, rarer genetic effects were also observed. In many of these cases, these effects were due not only to the difference in overall expression but also to allele-specific differences in transcript termination and isoform expression. The increasing availability of population-scale sequencing will require new statistical methodologies to evaluate rare genetic variants and to pinpoint causal regulatory variants.

Accounting for cellular history and state. Many regulatory elements are tissue specific[7,8] and therefore sequence variants with functional consequences are expected, to some degree, to manifest their effects in a tissue-specific way. As such, complications in the design and interpretation of eQTL studies are largely due to the complexity of gene regulation, both spatially (cell-type and tissue specificity) and temporally (different developmental stages). Several

studies have performed eQTL analysis in different tissues[9–11], and a number of them have compared eQTL between tissue types[12–16]. These studies have suggested that eQTL analyses should aim to interrogate multiple well-defined cell types, especially those relevant to specific disease phenotypes. In order to sample the full range of biological effects of such regulatory variants, projects such as the Genotype-Tissue Expression (GTEx) project have embarked on the ambitious goal of interrogating many diverse tissues, with the aim of recording and comparing tissue-specific eQTLs to an unprecedented scale of tissue- and cell-type resolution.

Accounting for population structure. The resolution of any genetic study depends on the structure of genomic variation, which varies substantially between populations[17,18]. Studies of the genetics of gene expression have been very informative regarding the distribution and properties of functional variants when these are compared between populations[19–21]. In particular, if multiple population samples are studied, the differential pattern of LD among populations allows for the finer mapping of shared variants[22], in much the same way as does assessing clone overlaps in conventional physical mapping studies. Such studies will inform greatly the conclusions we can draw from the sharing of functional effects between different populations and about how transferrable are common variants for predicting disease risk[21]. However, as most of the eQTL studies have been largely restricted to reference samples (for example, the International HapMap Project), methodologies that integrate data across populations are becoming necessary[21].

Accounting for specific endophenotypes. If we are strictly interested in the genetic signal, without accounting for any lifestyle or environmental variants, we could argue that it is best to study eQTLs in reference samples and not in the sample of interest. Using the reference sample allows us to distinguish the causal effects leading to the phenotype from the reactive effects that emerged to the individuals after developing the phenotype[23]: most of the common disease variants simply confer risk but, at the cellular level, they may have the same effect in cases and controls. Therefore, a common eQTL variant is expected to have the same cellular effect in diseased and healthy tissue. This point was elegantly demonstrated by Moffat *et al.*, who showed that although a GWA study signal for asthma mapped to an eQTL for the gene *ORMDL3*, the mean expression of *ORMDL3* did not differ significantly between cases and controls[24]. However, as our understanding of complex disease improves, we would also like to account for environmental and epigenetic effects and, in this respect, eQTL studies directly performed on the subjects of the complex trait study may be informative. In addition, the ability to cluster the genetic effects in pathways (by identifying their molecular phenotypic attributes) may reveal a level of complexity in the manifestation of causality that was previously unknown. An increasing number of studies are interrogating gene expression in well-phenotyped samples, so this type of information should be available soon.

Architecture of gene regulation

eQTL studies provide an indirect interrogation of the functional landscape of gene regulation, providing insight into the location and modes of effect of functional sequences that influence gene expression. Specifically, eQTL studies have illuminated the location and impact of regulatory variants, the tissue specificity of regulatory elements and the contribution of *cis*- versus *trans*-acting variation to the expression of any given gene.

Location and effect sizes of eQTLs. Many eQTL studies have reported that the strongest eQTLs are most frequently located proximal (within 100 kb) to the transcription start site (TSS) of their respective genes[10,11,21,25]. Such observations imply that proximal transcription-factor binding sites involved in RNA polymerase II recruitment and subsequent transcription are key components of the regulatory architecture and suggest that distal-acting elements exert a weaker influence. However, the weak effects of distal elements are difficult to confirm, as most eQTL variants identified to date are presumed not to be the causal variant but to be in LD with the causal variant(s). Because of this, the underlying causal variants could be more or less distal and might target multiple tissue- or condition-dependent regulatory elements or the exon structure of the transcript.

The effect sizes and distribution of eQTLs operating through regulatory elements other than the basal promoter are beginning to be investigated by study designs that interrogate condition-specific eQTL discovery in the same population sample. An investigation of tissue-specificity of eQTLs in T cells, lymphoblastoid cell lines (LCLs) and fibroblasts demonstrated that eQTLs that were shared across tissues were more restricted to the TSS and had larger effect sizes[13]. By contrast, tissue-specific eQTLs were broadly distributed and had weaker effect sizes[13] (FIG. 1). This evidence suggests that even when proximal to the TSS, individual genetic effects on condition-specific regulatory sequences are less influential than those operating either on the basal promoter or on exons that are shared across isoforms.

Tissue specificity. Condition-specific differences in eQTL discovery are also providing insight into the degree of specificity that is required for transcription in different biological contexts. Canonical models of condition-specific gene regulation propose that there are discrete regulatory elements that modify basal transcription[26]. However, the degree of sharing of eQTLs across conditions allows us to estimate shared regulatory elements and networks among tissues. As many enhancers have been observed to be tissue-specific

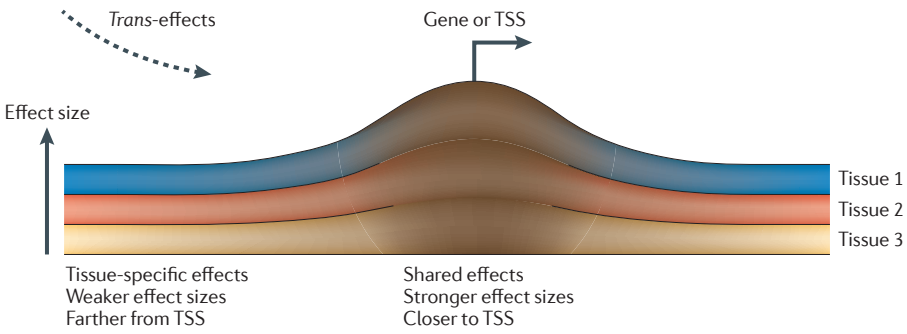


Figure 1 | **Gene regulatory architecture through expression quantitative trait locus studies.** Tissue-specific effects are typically weaker and are found more distal to the transcription start site (TSS). Shared effects are stronger and located closer to the TSS. Discrete effects are represented by discrete colours and shared effects are blended. *Trans*-effects, although regarded as weaker than *cis*-effects with respect to any particular locus, offer the potential to go beyond the dissection of local regulatory architecture to provide insight into gene-regulatory networks.

Table 1 | **Examples of disease risk loci correlated with eQTL genes**

Disease/trait study	Implicated eQTL genes	Expression source	Refs
Asthma	*ORMDL3*	EBV-transformed LCLs	24
Blood lipid levels	*SORT1*, *PPP1R3B*, *TTC39B*	Liver	53,54
Body mass index	*NEGR1*, *ZC3H4*, *TMEM160*, *MTCH2*, *NDUFS3*, *GTF3A*, *ADCY3*, *APOB48R*, *SH2B1*, *TUFM*, *GPRC5B*, *IQCK*, *SLC39A8*, *SULT1A1*, *SULT1A2*	Blood, brain, liver, lymphocytes, subcutaneous and visceral adipose tissue	48
Breast cancer	*RRP1B*	PyMT-induced primary tumours	55
Coeliac disease	*MMEL1*, *NSF*, *PARK7*, *PLEK*, *TAGAP*, *RRP1*, *UBE2L3*, *ZMIZ1*	Blood	47
Crohn's disease	*PTGER4*, *CARD9*, *ERAP2*, *TNFSF11*	EBV-transformed LCLs	56
Fat distribution	*GRB14*, *TBX15*, *PIGC*, *ZNRF3*, *STAB1*, AA553656	Blood, lymphocytes, omental fat, subcutaneous adipose tissue	57
Height	Multiple genes implicated	EBV-transformed LCLs, lymphocytes	58,59
Kidney-ageing	*MMP20*	Kidney	60
Migraine	*MTDH*	EBV-transformed LCLs	49
Multiple diseases	*CDKN2A* , *CDKN2B*, *CDKN2B-AS1*	Blood	61
Osteoporosis-related	*WLS*, *MEF2C*, *FOXC2*, *IBSP*, *TBC1D8*, *OSBPL1A*, *RAP1A*, *TNFRSF11B*	Liver, lymphocytes, primary osteoblasts	62,63
Parkinson's disease	*MAPT*, *LRRC37A*, *HLA-DRA*, *HLA-DQA2*, *HLA-DRB5*	EBV-transformed LCLs, frontal cortex	64,65
Psoriasis	*SDC4*, *SYS1*, *DBNDD2*, *PIGT*, *RPS26**	Lesional psoriatic skin	66
QRS duration and cardiac ventricular conduction	*TKT*, *CDKN1A*, *C6orf204*	Blood	67
Type 2 diabetes	*FADS1*, *FADS2*, *KLF14*, *CCNE2*, *IRS1*, *JAZF1*, *CAMK1D*	Blood, EBV-transformed LCLs, liver, subcutaneous adipose tissue	68,69

ADCY3, adenylate cyclase 3; *APOB48R*, apolipoprotein B48 receptor; *CAMK1D*, calcium/calmodulin-dependent protein kinase 1D; *CARD9*, caspase recruitment domain family, member 9; *CCNE2*, cyclin E2; *CDKN1A*, cyclin-dependent kinase inhibitor 1A; *CDKN2*, cyclin-dependent kinase inhibitor 2; *CDKN2B-AS1*, CDKN2B antisense RNA 1; *DBNDD2*, dysbindin (dystrobrevin binding protein 1) domain containing 2; EBV, Epstein–Barr virus; eQTL, expression quantitative trait locus; *ERAP2*, endoplasmic reticulum aminopeptidase 2; *FADS*, fatty acid desaturase; *FOXC2*, forkhead box C2; *GPRC5B*, G protein-coupled receptor, family C, group 5, member B; *GRB14*, growth factor receptor-bound protein 14; *GTF3A*, general transcription factor IIIA; *HLA*, human leukocyte antigen (major histocompatibility complex); *IBSP*, integrin-binding sialoprotein; *IQCK*, IQ motif containing K; *IRS1*, insulin receptor substrate 1; *JAZF1*, JAZF zinc finger 1; *KLF14*, Krüppel-like factor 14; LCLs, lymphoblastoid cell lines; *LRRC37A*, leucine rich repeat containing 37A; *MAPT*, microtubule-associated protein-τ; *MEF2C*, myocyte enhancer factor 2C; *MMEL1*, membrane metallo-endopeptidase-like 1; *MMP20*, matrix metallopeptidase 20; *MTCH2*, mitochondrial carrier homologue 2; *MTDH*, metadherin; *NDUFS3*, NADH dehydrogenase (ubiquinone) Fe-S protein 3, 30kDa (NADH-coenzyme Q reductase); *NEGR1*, neuronal growth regulator 1; *NSF*, *N*-ethylmaleimide-sensitive factor; *ORMDL3*, ORM1-like 3; *OSBPL1A*, oxysterol binding protein-like 1A; *PARK7*, Parkinson's disease (autosomal recessive, early onset) 7; *PIG*, phosphatidylinositol glycan anchor biosynthesis; *PLEK*, pleckstrin; *PPP1R3B*, protein phosphatase 1, regulatory (inhibitor) subunit 3B; *PTGER4*, prostaglandin E receptor 4 (subtype EP4); PyMT, Polyoma virus middle T antigen; QRS, electrocardiogram waveforms; *RAP1A*, RAP1A, member of RAS oncogene family; *RPS26*, ribosomal protein S26; *RRP1*, ribosomal RNA processing 1 homologue; *SDC4*, syndecan 4; *SH2B1*, SH2B adaptor protein 1; *SLC39A8*, solute carrier family 39 (zinc transporter), member 8; *SORT1*, sortilin 1; *STAB1*, stabilin 1; *SULT1A*, sulphotransferase family, cytosolic, 1A, phenol-preferring; *TAGAP*, T-cell activation RhoGTPase activating protein; *TBC1D8*, TBC1 domain family, member 8 (with GRAM domain); *TBX15*, T-box 15; *TKT*, transketolase; *TMEM160*, transmembrane protein 160; *TNFRSF11B*, tumour necrosis factor receptor superfamily, member 11b; *TNFSF11*, tumour necrosis factor (ligand) superfamily, member 11; *TTC39B*, tetratricopeptide repeat domain 39B; *TUFM*, Tu translation elongation factor, mitochondrial; *UBE2L3*, ubiquitin-conjugating enzyme E2L 3; *WLS*, wntless homologue; *ZC3H4*, zinc finger CCCH-type containing 4; *ZMIZ1*, zinc finger, MIZ-type containing 1; *ZNRF3*, zinc and ring finger 3. *eQTL genes for SNPs that did not achieve genome-wide significance but were of interest.

in the ENCyclopedia Of DNA Elements (ENCODE) project[7], it seems likely that cells with similar differentiation lineages will have increased eQTL sharing relative to developmentally distant tissues. However, in the aforementioned study of T cells, LCLs and fibroblasts — in which it was reported that 69–80% of *cis*-eQTLs are cell-type specific — it was found that even closely-related cell types (LCLs and T cells) only shared a minority of *cis*-eQTLs[13]. This suggests that high-frequency functional variation may be primarily present for late developmental processes. Unfortunately, it is difficult to directly test this hypothesis by comparing the complete regulatory element distribution to the elements for which eQTLs have been found, as our knowledge of developmental regulatory elements is incomplete.

Other studies have demonstrated that any estimate of sharing is likely to be highly context- and methodology-dependent. For example, a comparison of two distinct tissues — blood and adipose tissue — reported >50% sharing of *cis*-eQTLs[11]. Similarly, a study using linkage from mouse crosses found that >50% of the *cis*-eQTLs were shared across adipose, brain, liver and muscle tissues[27]. However, a separate study that investigated four differentiation states in haematopoiesis found only 365 shared eQTLs and 1,283 eQTLs that showed cell-type dependence[14]. Finally, a recent study looked beyond simple sharing of eQTLs to the level of expression fold change effect across tissues. This showed that even when an eQTL is shared at a statistically significant level, the fold difference in expression between tissues could be quite variable, representing another important level of tissue specificity[28].

All of the above methods have different study designs, power and analysis parameters, but the overall conclusion is that there is substantial tissue specificity of regulatory effects and the variants that influence these effects. By using patterns of eQTL sharing and specificity from studies such as these, we are increasingly able to understand gene regulatory pathways and locate context-specific regulatory sequences.

Trans *versus* cis *effects.* Understanding *cis*-regulatory architecture is complemented by eQTL studies designed to detect *trans*-regulatory effects. (In this Progress article, we adopt the distance-based convention

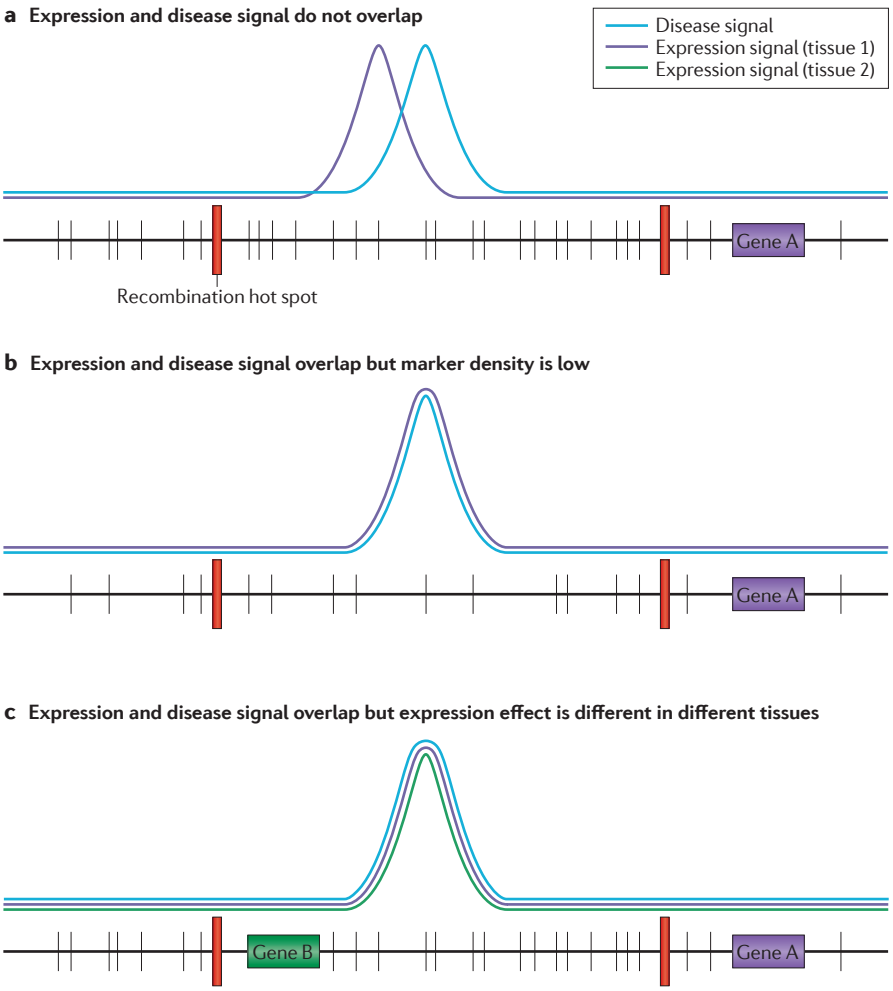
Box 1 | **Interpreting overlap between GWA and eQTL studies**

Methodologies that assess the degree of overlap of genome-wide association (GWA) studies with expression quantitative trait locus (eQTL) studies are important, but it is also crucial to understand the meaning of such overlap. There are three levels of caution one needs to consider.

First, the mere coincidence of an eQTL with a GWA study signal in the same genomic region does not necessarily suggest that they tag the same functional variant, even if they are not separated by a recombination hot spot and are historically correlated (see part **a** of the figure). The extent of correlation in the genome is tens to hundreds of kilobases, and it is very likely that multiple functional variants are found in each region of that size.

Second, even if GWA study and eQTL signal peak at the same marker, it does not mean that they tag the same functional variant, as the marker density may be variable and may not provide the resolution to disentangle nearby historically correlated functional variants (see part **b** of the figure). The above caveats could be addressed with the increasing availability of population-scale resequencing data and the increasing density of variants.

Third, the most challenging scenario is when the same functional variant is identified as causal for both the GWA study and the eQTL, but has been tested in one or a limited number of tissues (part **c** of the figure). Regulatory elements are known to be pleiotropic, so the same regulatory region may be regulating gene A in tissue 1 and gene B in tissue 2. If the GWA study signal is mediated by the effect of the functional variant in gene A in tissue 1 but we only have data for tissue 2, we will be misled into believing that gene B is relevant for the complex trait and we would incorrectly link gene B and its associated pathway to the phenotype. Connecting a causal variant to expression of a gene in the appropriate tissue is complicated further by the fact that it is often difficult to identify the 'relevant tissue' for a given phenotype. The tissue that is affected by a disease is not necessarily the tissue in which the causal genetic effect is manifested; for example, variants acting on the hypothalamus increase the risk for obesity even though adipose tissue is the affected tissue[52]. This last caveat strongly argues for extensive eQTL studies of multiple tissues in humans in order to dissect the circularity of cause and effect in genetic signals.



of *cis* versus *trans*[29].) An advantage of this approach is that by understanding how genetic differences that affect the expression and structure of distal genes influence the expression of a target gene, one can assign the importance of a particular protein product for that target gene's expression. Several studies have attempted to estimate the importance of these types of effects by estimating the proportion of a gene's expression that is influenced by *trans*-effects versus *cis*-effects[9–11,25,30–33]. A consistent result from most studies is that *trans*-eQTLs have weaker effects than *cis*-eQTLs. This view is contested by more recent studies that suggest that, despite lower effect-sizes, *trans*-eQTLs cumulatively explain more of the heritability of expression[30,31]. Despite doubts about the influence of *trans*-eQTLs, causal relationships underlying *trans*-eQTLs are being actively explored. Perhaps the most palatable hypothesis is that *trans*-eQTLs should be enriched outside transcription factor genes. However, such relationships have not been definitively observed[34,35]. There is growing evidence that this is because these *trans*-eQTL genes can mediate the effects of a transcription factor instead of directly encoding a transcription factor[31,36].

Given the complexity of potential *trans*-eQTLs, can regulatory pathways be reliably assembled from eQTL data? It appears so. One study identified and validated a new relationship between cyclin H (*CCNH*) and oxidative phosphorylation derived from eQTL data[37]. Another study identified a core regulatory network in human and mouse adipose tissue that was relevant to obesity-related traits[11]. A complementary study used gene co-expression to derive networks and observed enrichment for 'module-QTLs' with mouse weight[38]. On top of these studies, many systems-biology-based techniques have become available to categorize regulatory pathways from eQTL data[37,39–44]. Such techniques will increasingly illuminate the transcription factor complexes and pathways necessary for gene regulation.

Insights through sequencing. RNA-seq has provided further insight into the landscape of gene regulation and has highlighted patterns of transcription complexity that operate in *cis* and in *trans*. Such resolution allows the dissection of the specific functional components and genetic variants that influence gene expression. Two studies have reported that many eQTLs are driven by splicing variants and genetic differences in transcription termination[5,45]. We have used allele-specific

expression information derived from RNA-seq to identify patterns suggestive of haplotype-specific isoform expression. This work supports a previous observation, performed using gene expression microarrays, that the number of eQTLs for a gene is correlated with the number of known transcript isoforms[13]. Additionally, by integrating allele-specific expression from RNA-seq and linkage analysis, it has been shown that only 20% of *trans*-regulators have allelic imbalance, showing that *trans*-eQTLs can affect a target gene in various ways[31].

Other sequencing-based tools are contributing to our understanding of gene regulation through genetics. A family-based study of transcription factor binding using chromatin immunoprecipitation followed by DNA sequencing (ChIP–seq) is probably ushering in a new eQTL design that integrates not only expression-level differences but also the inheritance of different transcription factors and epigenetic marks[46]. With increasing resolution of gene regulation and subsequent expression, the functional role we can assign to genetic variants and their effect on physiological traits (particularly disease) will become more refined.

Integrating eQTLs with GWA studies

Ultimately, we would like to use eQTL data to interpret GWA study signals and elucidate the specific biological effects underlying them. This approach has been validated by several studies. For example, Dubois *et al.*[47] found that 20 of the 38 loci that had associated risk variants for coeliac disease are also correlated with variation in the expression of a nearby gene. In an analysis of body mass index in a quarter of a million people, multiple tissue eQTL analyses allowed 14 of 32 loci to be linked with specific gene candidates[48]. Finally, in an analysis of the genetics of migraine, genotypic correlation to expression of a candidate gene suggests a regulatory basis for this trait[49]. Increasingly, these types of meaningful links between disease-associated variants and eQTLs are providing the basis for selecting candidate genes and exploring the mechanisms underlying different diseases (TABLE 1). However, specific caveats in interpreting correlations between GWA and eQTL signals must be heeded (BOX 1).

Future directions

eQTL studies are increasingly informing the architecture of gene regulation and allowing us to connect genetic variation to important physiological traits. Many studies have already begun to link cellular traits (predominantly expression) to disease traits. There are certain challenges ahead, however.

From cellular to disease traits. First, given that measures of eQTL sharing vary across cell types and biological traits, context-specific discovery might affect our ability to relate genetic effects at the cellular level to disease-related traits. Future studies will need to be conducted in relevant tissues that underlie the aetiology of the trait of interest. This will require us to leverage advances in technologies such as induced pluripotent stem cells (iPSCs) to derive harder-to-acquire cell types[50]. An alternative solution could be to engineer cellular states that have components of the model system of interest in model organisms or through chemical-genetics-based approaches. These approaches might allow us to study networks in inaccessible cell types by activating relevant pathways in external environments or by using exogenous agents.

Personalized transcriptomics. A second major challenge is to elucidate the causal variants. Population-scale genome sequencing is providing the potential to achieve this with unprecedented resolution of the genetic variation that influences expression traits[51]. The diversity of functional-genomics-sequencing-based tools such as RNA-seq and ChIP–seq, in combination with full-genome sequencing, will allow the discovery both of specific causal variants and any associated functional information. This will allow the impact of novel variants to be inferred. The combination of this information will probably be more useful for determining disease risk and developing treatments than any one method alone: the expression data integrates the impact of multiple variants and the environment, and the genome data integrates the genetic data while being agnostic to cellular condition. Such information would, for instance, illuminate epistatic interactions between regulatory variants and protein-coding variants and inform about whether disease-predisposing variant effects are modulated, thus illuminating modes of differential penetrance.

eQTL studies are showing that future visits to the clinic will not be solely based on personalized genomics (that is, genome sequencing) but instead on personalized 'omics', which will combine in-depth analysis of DNA and functional genomics to tell us more about the medical condition of an individual.

Stephen B. Montgomery and Emmanouil T. Dermitzakis are at the Department of Genetic Medicine and Development, University of Geneva Medical School, 1 rue Michel-Servet, Geneva 1211, Switzerland.
e-mails: Emmanouil.Dermitzakis@unige.ch; Stephen.Montgomery@unige.ch

doi:10.1038/nrg2969
Published online 9 March 2011

1. Manolio, T. A. *et al.* Finding the missing heritability of complex diseases. *Nature* **461**, 747–753 (2009).
2. Manolio, T. A. Genomewide association studies and assessment of the risk of disease. *N. Engl. J. Med.* **363**, 166–176 (2010).
3. Nica, A. C. *et al.* Candidate causal regulatory effects by integration of expression QTLs with complex trait genetic associations. *PLoS Genet.* **6**, e1000895 (2010).
4. Nicolae, D. L. *et al.* Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS. *PLoS Genet.* **6**, e1000888 (2010).
5. Montgomery, S. B. *et al.* Transcriptome genetics using second generation sequencing in a Caucasian population. *Nature* **464**, 773–777 (2010).
6. Pickrell, A. M. & Moraes, C. T. What role does mitochondrial stress play in neurodegenerative diseases? *Methods Mol. Biol.* **648**, 63–78 (2010).
7. Birney, E. *et al.* Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* **447**, 799–816 (2007).
8. Heintzman, N. D. *et al.* Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature* **459**, 108–112 (2009).
9. Schadt, E. E. *et al.* Mapping the genetic architecture of gene expression in human liver. *PLoS Biol.* **6**, e107 (2008).
10. Myers, A. J. *et al.* A survey of genetic human cortical gene expression. *Nature Genet.* **39**, 1494–1499 (2007).
11. Emilsson, V. *et al.* Genetics of gene expression and its effect on disease. *Nature* **452**, 423–428 (2008).
12. Heinzen, E. L. *et al.* Tissue-specific genetic control of splicing: implications for the study of complex traits. *PLoS Biol.* **6**, e1 (2008).
13. Dimas, A. S. *et al.* Common regulatory variation impacts gene expression in a cell type-dependent manner. *Science* **325**, 1246–1250 (2009).
14. Gerrits, A. *et al.* Expression quantitative trait loci are highly sensitive to cellular differentiation state. *PLoS Genet.* **5**, e1000692 (2009).
15. Grundberg, E. *et al.* Population genomics in a disease targeted primary cell model. *Genome Res.* **19**, 1942–1952 (2009).
16. Kwan, T. *et al.* Tissue effect on genetic control of transcript isoform variation. *PLoS Genet.* **5**, e1000608 (2009).
17. Altshuler, D. M. *et al.* Integrating common and rare genetic variation in diverse human populations. *Nature* **467**, 52–58 (2010).
18. Frazer, K. A. *et al.* A second generation human haplotype map of over 3.1 million SNPs. *Nature* **449**, 851–861 (2007).
19. Spielman, R. S. *et al.* Common genetic variants account for differences in gene expression among ethnic groups. *Nature Genet.* **39**, 226–231 (2007).
20. Storey, J. D. *et al.* Gene-expression variation within and among human populations. *Am. J. Hum. Genet.* **80**, 502–509 (2007).
21. Stranger, B. E. *et al.* Population genomics of human gene expression. *Nature Genet.* **39**, 1217–1224 (2007).
22. Zaitlen, N., Pasaniuc, B., Gur, T., Ziv, E. & Halperin, E. Leveraging genetic variability across populations for the identification of causal variants. *Am. J. Hum. Genet.* **86**, 23–33 (2010).
23. Schadt, E. E. *et al.* An integrative genomics approach to infer causal associations between gene expression and disease. *Nature Genet.* **37**, 710–717 (2005).
24. Moffatt, M. F. *et al.* Genetic variants regulating ORMDL3 expression contribute to the risk of childhood asthma. *Nature* **448**, 470–473 (2007).
25. Dixon, A. L. *et al.* A genome-wide association study of global gene expression. *Nature Genet.* **39**, 1202–1207 (2007).
26. Arnosti, D. N. & Kulkarni, M. M. Transcriptional enhancers: Intelligent enhanceosomes or flexible billboards? *J. Cell Biochem.* **94**, 890–898 (2005).
27. van Nas, A. *et al.* Expression quantitative trait loci: replication, tissue- and sex-specificity in mice. *Genetics* **185**, 1059–1068 (2010).

28. Nica, A. E. A. The architecture of gene regulatory variation across multiple human tissues: the MuTHER Study. *PLoS Genet.* **7**, e1002003 (2011).
29. Gilad, Y., Rifkin, S. A. & Pritchard, J. K. Revealing the architecture of gene regulation: the promise of eQTL studies. *Trends Genet.* **24**, 408–415 (2008).
30. Price, A. L. *et al.* Effects of *cis* and *trans* genetic ancestry on gene expression in African Americans. *PLoS Genet.* **4**, e1000294 (2008).
31. Cheung, V. G. *et al.* Polymorphic *cis*- and *trans*-regulation of human gene expression. *PLoS Biol.* **8**, e1000480 (2010).
32. Goring, H. H. *et al.* Discovery of expression QTLs using large-scale transcriptional profiling in human lymphocytes. *Nature Genet.* **39**, 1208–1216 (2007).
33. Petretto, E. *et al.* New insights into the genetic control of gene expression using a Bayesian multi-tissue approach. *PLoS Comput. Biol.* **6**, e1000737 (2010).
34. Yvert, G. *et al.* Trans-acting regulatory variation in *Saccharomyces cerevisiae* and the role of transcription factors. *Nature Genet.* **35**, 57–64 (2003).
35. Breitling, R. *et al.* Genetical genomics: spotlight on QTL hotspots. *PLoS Genet.* **4**, e1000232 (2008).
36. Sun, W., Yu, T. & Li, K. C. Detection of eQTL modules mediated by activity levels of transcription factors. *Bioinformatics* **23**, 2290–2297 (2007).
37. Wu, C. *et al.* Gene set enrichment in eQTL data identifies novel annotations and pathway regulators. *PLoS Genet.* **4**, e1000070 (2008).
38. Ghazalpour, A. *et al.* Integrating genetic and network analysis to characterize genes related to mouse weight. *PLoS Genet.* **2**, e130 (2006).
39. Ren, X., Zhou, X., Wu, L. Y. & Zhang, X. S. An information-flow-based model with dissipation, saturation and direction for active pathway inference. *BMC Syst. Biol.* **4**, 72 (2010).
40. Li, S., Lu, Q. & Cui, Y. A systems biology approach for identifying novel pathway regulators in eQTL mapping. *J. Biopharm. Stat.* **20**, 373–400 (2010).
41. Rashid, I., McDermott, J. & Samudrala, R. Inferring molecular interactions pathways from eQTL data. *Methods Mol. Biol.* **541**, 211–223 (2009).
42. Wessel, J., Zapala, M. A. & Schork, N. J. Accommodating pathway information in expression quantitative trait locus analysis. *Genomics* **90**, 132–142 (2007).
43. Suthram, S., Beyer, A., Karp, R. M., Eldar, Y. & Ideker, T. eQED: an efficient method for interpreting eQTL associations using protein networks. *Mol. Syst. Biol.* **4**, 162 (2008).
44. Lee, E. & Bussemaker, H. J. Identifying the genetic determinants of transcription factor activity. *Mol. Syst. Biol.* **6**, 412 (2010).
45. Pickrell, J. K. *et al.* Understanding mechanisms underlying human gene expression variation with RNA sequencing. *Nature* **464**, 768–772 (2010).
46. McDaniell, R. *et al.* Heritable individual-specific and allele-specific chromatin signatures in humans. *Science* **328**, 235–239 (2010).
47. Dubois, P. C. *et al.* Multiple common variants for celiac disease influencing immune gene expression. *Nature Genet.* **42**, 295–302 (2010).
48. Speliotes, E. K. *et al.* Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nature Genet.* **42**, 937–948 (2010).
49. Anttila, V. *et al.* Genome-wide association study of migraine implicates a common susceptibility variant on 8q22.1. *Nature Genet.* **42**, 869–873 (2010).
50. Yamanaka, S. & Blau, H. M. Nuclear reprogramming to a pluripotent state by three approaches. *Nature* **465**, 704–712 (2010).
51. Durbin, R. M. *et al.* A map of human genome variation from population-scale sequencing. *Nature* **467**, 1061–1073 (2010).
52. Willer, C. J. *et al.* Six new loci associated with body mass index highlight a neuronal influence on body weight regulation. *Nature Genet.* **41**, 25–34 (2009).
53. Musunuru, K. *et al.* From noncoding variant to phenotype via SORT1 at the 1p13 cholesterol locus. *Nature* **466**, 714–719 (2010).
54. Teslovich, T. M. *et al.* Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* **466**, 707–713 (2010).
55. Crawford, N. P. *et al. Rrp1b*, a new candidate susceptibility gene for breast cancer progression and metastasis. *PLoS Genet.* **3**, e214 (2007).
56. Libioulle, C. *et al.* 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 (2007).
57. Heid, I. M. *et al.* Meta-analysis identifies 13 new loci associated with waist-hip ratio and reveals sexual dimorphism in the genetic basis of fat distribution. *Nature Genet.* **42**, 949–960 (2010).
58. Lango Allen, H. *et al.* Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature* **467**, 832–838 (2010).
59. Soranzo, N. *et al.* Meta-analysis of genome-wide scans for human adult stature identifies novel loci and associations with measures of skeletal frame size. *PLoS Genet.* **5**, e1000445 (2009).
60. Wheeler, H. E. *et al.* Sequential use of transcriptional profiling, expression quantitative trait mapping, and gene association implicates MMP20 in human kidney aging. *PLoS Genet.* **5**, e1000685 (2009).
61. Cunnington, M. S., Santibanez Koref, M., Mayosi, B. M., Burn, J. & Keavney, B. Chromosome 9p21 SNPs associated with multiple disease phenotypes correlate with ANRIL expression. *PLoS Genet.* **6**, e1000899 (2010).
62. Hsu, Y. H. *et al.* An integration of genome-wide association study and gene expression profiling to prioritize the discovery of novel susceptibility loci for osteoporosis-related traits. *PLoS Genet.* **6**, e1000977 (2010).
63. Rivadeneira, F. *et al.* Twenty bone-mineral-density loci identified by large-scale meta-analysis of genome-wide association studies. *Nature Genet.* **41**, 1199–1206 (2009).
64. Simon-Sanchez, J. *et al.* Genome-wide association study reveals genetic risk underlying Parkinson's disease. *Nature Genet.* **41**, 1308–1312 (2009).
65. Hamza, T. H. *et al.* Common genetic variation in the HLA region is associated with late-onset sporadic Parkinson's disease. *Nature Genet.* **42**, 781–785 (2010).
66. Stuart, P. E. *et al.* Genome-wide association analysis identifies three psoriasis susceptibility loci. *Nature Genet.* **42**, 1000–1004 (2010).
67. Sotoodehnia, N. *et al.* Common variants in 22 loci are associated with QRS duration and cardiac ventricular conduction. *Nature Genet.* **42**, 1068–1076 (2010).
68. Dupuis, J. *et al.* New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. *Nature Genet.* **42**, 105–116 (2010).
69. Voight, B. F. *et al.* Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. *Nature Genet.* **42**, 579–589 (2010).

Acknowledgements
We acknowledge funds from the Louis-Jeantet Foundation, the Swiss National Science Foundation and the European Commission and the help and comments of our Functional Population Genomics group in Geneva.

Competing interests statement
The authors declare no competing financial interests.

FURTHER INFORMATION
Author's homepage: http://www.medecine.unige.ch/recherche/groupes/b_donnees/sujet_892_4.html
Genotype-Tissue Expression (GTEx) Project: http://commonfund.nih.gov/GTEx
International HapMap Project: http://www.hapmap.org
ENCyclopedia Of DNA Elements (ENCODE) project: http://www.genome.gov/Pages/Research/ENCODE

ALL LINKS ARE ACTIVE IN THE ONLINE PDF