Contents lists available at ScienceDirect

Biochimica et Biophysica Acta

journal homepage: www.elsevier.com/locate/bbadis

Review

From genome to function by studying eQTLs☆

CrossMark

Harm-Jan Westra, Lude Franke *

University of Groningen, University Medical Center Groningen, Department of Genetics, Hanzeplein 1, 9700 RB Groningen, The Netherlands

ARTICLE INFO

ABSTRACT

Genome-wide association studies (GWASs) have shown a large number of genetic variants to be associated with complex diseases. The identification of the causal variant within an associated locus can sometimes be difficult because of the linkage disequilibrium between the associated variants and because most GWAS loci contain multiple genes, or no genes at all. Expression quantitative trait locus (eQTL) mapping is a method used to determine the effects of genetic variants on gene expression levels. eQTL mapping studies have enabled the prioritization of genetic variants within GWAS loci, and have shown that trait-associated single nucleotide polymorphisms (SNPs) often function in a tissue- or cell type-specific manner, sometimes having downstream effects on completely different chromosomes. Furthermore, recent RNA-sequencing (RNA-seq) studies have shown that a large repertoire of transcripts is available in cells, which are actively regulated by (trait-associated) variants. Future eQTL mapping studies will focus on broadening the range of available tissues and cell types, in order to determine the key tissues and cell types involved in complex traits. Finally, large meta-analyses will be able to pinpoint the causal variants within the trait-associated loci and determine their downstream effects in greater detail. This article is part of a Special Issue entitled: From Genome to Function.

© 2014 Elsevier B.V. All rights reserved.

1. Introduction

In the last few years, a large number of genome-wide association studies (GWASs) have been performed in attempts to uncover the genetic basis of many different complex diseases and traits. GWASs typically ascertain at least 300,000 common single nucleotide polymorphisms (SNPs) throughout the genome, and each of these variant associations with the disease is tested. For many traits, this approach has turned out to be highly successful; disease and trait associations for over 12,000 SNPs from over 1700 publications have now been reported (NHGRI Catalog of Published Genome-Wide Association studies) [1]. However, it soon became clear that the identified genetic variants typically explain only a very modest proportion of the total heritability of these traits.

One plausible explanation was that these GWASs had only investigated common SNPs (those with a minor allele frequency (MAF) above 5%). As such, many rare variants had not been ascertained, and it was therefore assumed that the common SNPs identified for a disease were actually tagging rarer variants (MAF < 5%) with a larger effect size. To test this hypothesis, fine-mapping studies were conducted, made possible with the availability of the next generation sequencing (NGS) methods: by sequencing candidate genes, whole exomes or genomes

it is possible to identify rare variants [2] and their association with disease became testable through the development of dedicated oligonucleotide arrays that specifically target these rare variants (e.g. the ImmunoChip and MetaboChip). Although this helped to fine-map loci for various diseases, few rare variants have so far been identified that have a large effect size.

These results, along with the observation that many smaller-effect loci became genome-wide significant upon increasing the sample sizes used in many GWASs, suggested that the genetic architecture for many traits could well be highly polygenic. This was further supported by the availability of polygenic models in 2009 [3,4]: these methods estimate the total proportion of variation that can be explained by all genotyped common SNPs, without requiring that any of the SNPs individually shows significant association (after correction for multiple testing). Initial results on adult height (which has an estimated heritability of 80% and is a phenotype that can be highly accurately quantified [4]) revealed that common genetic variants captured approximately 45% of the total variation in height, whereas the 180 genome-wide significant loci that had been found (when studying 180,000 samples) explained less than 10% of the variation in height. These results suggested that hundreds, or maybe even thousands, of genetic variants could well play a causal role in many traits.

These observations have proven highly problematic in trying to move from the discovery of these variants through GWAS to their biological interpretation for various reasons: given that many of the disease-causing variants are likely to be common, have small effect-sizes, and are often in near-perfect linkage disequilibrium (LD) with nearby SNPs. It is difficult to unequivocally identify the causal variant

for each locus through traditional fine-mapping methods. This also strongly impairs the ability to accurately pinpoint the causal gene(s) in each locus. Additionally, the mechanisms and function of each of these trait-associated variants are largely unknown, since many of the trait-associated SNPs are not actually changing the protein structure (i.e. are non-synonymous or nonsense mutations), but are often located in non-coding regions of the genome. This suggests that these variants have a regulatory function. A compounding problem is that often tens of disease-associated variants have now been identified for many diseases, making it infeasible to knock-down, knock-out or over-express each of the genes within these loci.

In order to identify which genes are regulated by genetic variation, Jansen and Nap introduced the concept of 'genetical genomics' [5] in 2001: by correlating the genetic variants with intermediate molecular quantitative traits (such as gene expression levels, protein levels or methylation levels), it is possible to identify quantitative trait loci (QTLs). The first product of the genome, mRNA levels, can be quantified easily for thousands of genes at once, by either using microarrays or by conducting RNA-sequencing. It soon became clear that gene expression levels are strongly heritable: for all human genes the average heritability was estimated to be around 0.25 [6–8]. Soon, expression QTL (eQTL) mapping was conducted in humans [9–11] (and model organisms such as *Arabidopsis thaliana* [12], *Caenorhabditis elegans* [13], mice and rats [14]), resulting in the identification of many genetic variants that affect gene expression levels.

2. eQTLs as a means to functionally annotate trait associated SNPs

eQTLs can be divided into those that have local effects (*cis*-eQTLs), where the genetic variant is located near the affected gene (e.g. within 1 megabase), and those with distant effects (*trans*-eQTLs), where the genetic variant is located further away from the affected gene (e.g. more than 5 megabases away, or on a completely different chromosome; Fig. 1).
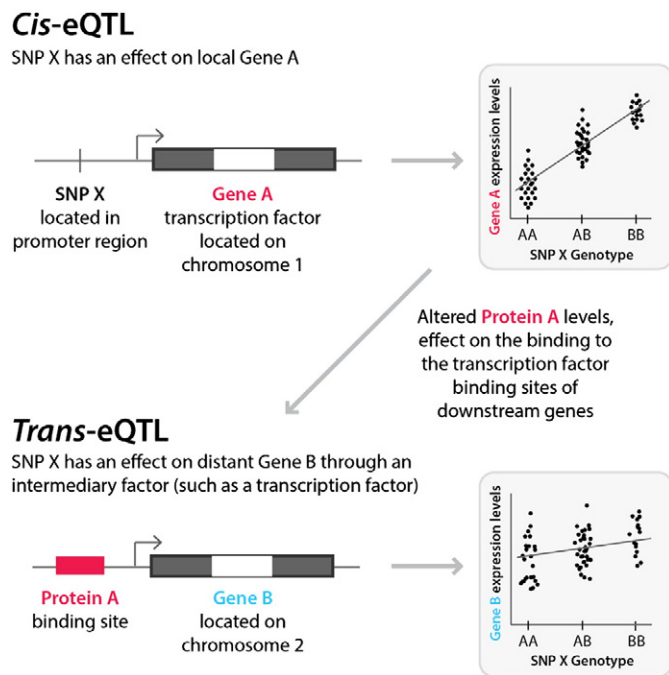
Cis-eQTL
SNP X has an effect on local Gene A

SNP X located in promoter region — Gene A transcription factor located on chromosome 1

Gene A expression levels — SNP X Genotype (AA, AB, BB)

Altered Protein A levels, effect on the binding to the transcription factor binding sites of downstream genes

Trans-eQTL
SNP X has an effect on distant Gene B through an intermediary factor (such as a transcription factor)

Protein A binding site — Gene B located on chromosome 2

Gene B expression levels — SNP X Genotype (AA, AB, BB)

Fig. 1. eQTLs can be either local effects (*cis*-eQTLs), or distant, indirect effects (*trans*-eQTLs).
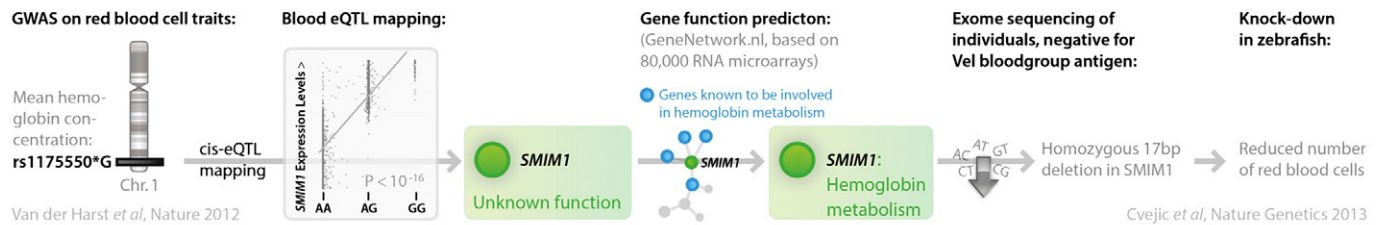
2.1. Cis-eQTLs

Since *cis*-eQTLs often have a large effect size [15], relatively modest sample sizes permit the detection of *cis*-eQTLs for thousands of genes [6,16–20]. *Cis*-eQTL effects appear to be mostly additive effects [21], and *cis*-eQTL SNPs are often located close to the transcription start site (TSS) of genes or within gene bodies [22–24]. As the distance between the eQTL SNP and the TSS decreases, the eQTL effect size generally increases. *Cis*-eQTL SNPs that are located close to the TSS may alter transcription factor binding sites or other *cis*-regulatory elements (CREs), which in turn may affect transcription. The observation that *cis*-eQTL SNPs tend to be overlapping with activating CREs, such as DNAse-I hypersensitive sites (DHSs) and transcription factor binding sites, and tend to be depleted for repressive CREs (such as CTCF binding sites) strengthened this hypothesis [25]. Finally, trait-associated SNPs have been shown to be enriched for *cis*-eQTL effects [20,26–28], which further indicates that trait-associated SNPs are often regulatory. *Cis*-eQTLs can aid in pinpointing the causal variant within a locus: after a GWAS on red blood cell traits [29], *cis*-eQTL mapping was performed in whole blood samples, which identified a *cis*-eQTL in the *SMIM1* locus on chromosome 1. Subsequent functional annotation using a gene expression co-regulation network suggested *SMIM1* was the causal gene within the locus. A follow-up exome sequencing study and knock down experiment in zebrafish revealed that *SMIM1* underlies the Vel blood group (Fig. 2A) [30].

Although *cis*-eQTLs, such as the *SMIM1* example, can provide valuable information about the likely causal gene for trait-associated SNPs, finding the causal gene underneath GWAS peaks is not always straightforward: LD might be so extensive that many candidate genes remain, or the regulatory regions that are influenced by the genetic variants may actually be located megabases away from the transcription start site of the causal gene. This has recently been observed for intronic variants within the *FTO* gene that have been found to be associated with type 2 diabetes and obesity [31,32]. Surprisingly, these variants do not show a *cis*-eQTL effect on *FTO*, but they do affect the gene expression levels of *IRX3*, which is located megabases away from the *FTO* locus [33]. Knocking-out *IRX3* in mice results in a 30% weight decrease in mice, confirming the importance of *IRX3* in regulating weight. These results illustrate that the genes that are located in very close proximity to the associated variant are not always the causal gene and also that variants associated with GWAS may have functional consequences on genes located megabases away, which raises the question whether such effects should be considered *trans*-eQTLs.
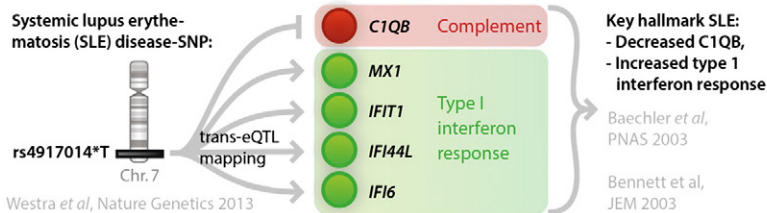
2.2. Trans-eQTLs

In contrast to *cis*-eQTL effects, the effect sizes of *trans*-eQTLs are generally small [9,34]. As a consequence, the sample sizes required to detect such effects are large, and as a result, the number of reported *trans*-eQTLs has remained small [9,17,19,35–37] in comparison to the number of reported *cis*-eQTLs. However, initial *trans*-eQTL studies have shown that *trans*-eQTL analysis provides valuable insight into disease pathogenesis. For example, multiple *trans*-eQTL genes were previously identified that are affected by a single SNP that is associated with type 2 diabetes and high-density lipoprotein levels. SNPs associated with these *trans*-genes also showed genetic association with various metabolic phenotypes [37], indicating that *trans*-eQTL mapping is able to identify coherent networks of genes that are likely to be causally involved in disease pathogenesis. Similarly, *trans*-eQTL genes were identified that are affected by a SNP in the *IRF7* locus, associated with the auto-immune disease type 1 diabetes. These downstream *trans*-genes showed an association with viral response [36]. To detect more *trans*-eQTL effects, sample-sizes were increased by performing meta-analyses [19,20]: a meta-analysis of 1469 whole blood samples showed that HLA SNPs were 10-fold enriched for showing *trans*-eQTL effects. For a few different complex traits it was also shown that SNPs
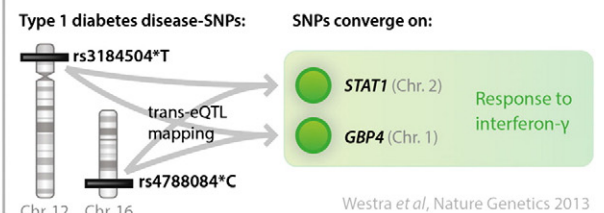
Fig. 2. Functional genomic studies translate GWAS findings into clear biological insight. (A) A recent GWAS conducted on red blood cell traits identified a locus on chromosome 1 associated with mean hemoglobin concentration. Through subsequent *cis*-eQTL mapping and gene function prediction (using a compendium of 80,000 microarrays), *SMIM1* was identified as the possibly causal gene in the locus on chromosome 1 that was predicted to be involved in hemoglobin metabolism. Subsequent exome-sequencing revealed this gene underlies the rare Vel blood group, and knock-down of Vel1 in zebrafish resulted in a reduced number of red blood cells. (B) Through *trans*-eQTL mapping in healthy individuals the downstream effects for the systemic lupus erythematosis (SLE) SNP rs4917014 were identified. These effects are identical to the key hallmarks of SLE: decreased complement 1q levels and an increased type 1 interferon response. (C) SNPs that increase risk for the same disease 'converge' on the same downstream genes: two unlinked type 1 diabetes SNPs affect exactly the same downstream genes in *trans* (*STAT1* and *GBP4*, both involved in the interferon-γ response).

independently associated with these traits affected the expression of exactly the same downstream genes in *trans*, creating functional converging pathways that are relevant for the traits associated with these SNPs [19]. A larger meta-analysis involving 5311 whole blood samples further increased the number of reported *trans*-eQTL genes to 430 and showed that *trans*-eQTLs can be informative of disease pathogenesis: in two previous cross-sectional studies, several interferon response genes had been identified that show strongly dysregulated expression in the blood of systemic lupus erythematosus (SLE) patients (Fig. 2B). The *trans*-eQTL study identified a single SNP, associated with SLE that affected exactly these genes, indicating that dysregulation of these interferon response genes is already detectable when a healthy individual is carrying SLE susceptibility alleles [20]. Similar to the meta-analysis in 1469 individuals, this larger meta-analysis provided information on the convergence of functional pathways, including converging effects originating from two type 1 diabetes associated variants, affecting the well-known type 1 diabetes gene *STAT1* (Fig. 2C).

2.3. Cell type and tissue specificity

Gene expression levels often vary considerably between different tissues and cell types [38]. As such, eQTL mapping studies have now been performed in various cell-types and tissues, such as fibroblasts, liver [17,39–41], lung [42], brain [16], muscle [41], adipose tissue [41], skin [43], various purified blood cell types (e.g. lymphoblastoid cell-lines (LCLs) [10,40,43–45], B-cells [35], monocytes [35], and T-cells [24]), and whole blood [19,20,46]. Early comparisons between cell types showed that the number of shared eQTL effects varies widely with the cell types or tissues under study. A comparison of skin and LCL eQTLs showed that 70% of *cis*-eQTLs found in LCLs could also be detected in skin [43], while a comparison of fibroblasts, T-cells and B-cells showed an overlap of up to 12% of the detected *cis*-eQTLs in any combination of two of these three cell types [24]. However, these studies had overestimated the cell-type specificity because of their small sample size and the statistical methods employed to make these comparisons. A more recent comparison of B-cells and monocytes in over 280 individuals showed a higher overlap: 21.8% of the detected *cis*-eQTLs and 7% of the detected *trans*-eQTLs were shared between both cell types [35],

which suggests that genetic regulation in *trans* is more cell-type-specific than *cis* regulation. Only a small proportion of the identified eQTL effects in this study could be replicated in whole blood (even though blood is partly comprised of monocytes and B-cells), indicating that eQTL mapping in a tissue that is composed of many cell types may reduce the power to find cell-type-specific eQTL effects.

Another recent study, comparing five tissues (subcutaneous and visceral adipose tissue, muscle, liver, and whole blood), described how 28.7% of the *cis*-eQTLs show differences across tissues [41]. Of these, 33% had eQTL effects unique to one of the tissues, 47.9% showed eQTLs originating from different SNPs in different tissues, and 4.4% unexpectedly showed a different direction of effect in one or more tissues, something that has recently been observed in other studies as well [35,47]. This study also showed *cis*-eQTL effects for 46.4% of the tested trait-associated SNPs, and indicated that these SNPs are enriched for tissue-dependent effects, compared to frequency matched SNPs. eQTLs that are shared across tissues and cell types have larger effect sizes, and their associated SNPs are located closer to the TSS [25,41] than tissue- and cell-type specific eQTLs. On average, 29% of the *cis*-eQTL loci also appear to have multiple independent SNPs affecting the same transcript [25]. Overall, these studies show that the genetic regulation of gene expression is complex and differs across cell-types and tissues, especially for disease-associated genetic variants.

2.4. Context specificity

Apart from differences in cell types, a large fraction of gene expression variation is due to the effect of environmental factors, begging the question whether some of these environmental factors might induce eQTLs. Several environmental factors have now been assessed in the context of eQTLs, which include response to radiation [48], geography [49], different treatments for disease [50,51], response to influenza vaccination [52], and infections with tuberculosis [53] and malaria [54]. However, the sample sizes for these studies have generally been rather small (up to 194 individuals), due to the difficulties and costs involved in acquiring samples that are relevant for the specific environmental factor. More powerful studies have been published as well: a study in monocytes from 1490 independent individuals showed 651 *cis*-eQTLs

that have interactions with either age, smoking status, gender, blood pressure and lipid traits [55]. However, the early stage of the studies of context-specific eQTLs has sometimes led to discrepant results being observed: for example, a gender-stratified analysis in 379 LCLs suggested that between 12% and 15% of the autosomal eQTLs function in a sex-biased manner [56], and a larger study in peripheral blood samples from 1240 individuals showed interactions for 623 eQTLs with age [57]. Although the large fraction of gender mediated effects in the LCLs may be caused by (epi-)genomic alterations caused by the Epstein Barr virus immortalization of these cell-lines [58], a subsequent, but much larger study of 5254 peripheral blood samples showed only 14 and 10 eQTLs interacting with gender and age, respectively [59]. One potential explanation for these discrepancies could be the statistical challenges, associated with performing large-scale gene environment interaction analysis: in order to get robust significance estimates of interaction effects, heteroscedasticity-consistent standard errors should be used [60] (e.g. available through the R package 'Sandwich').

Still, context-specific eQTL studies hold great promise. A recent study in monocytes, comparing the effect sizes of eQTLs before and after stimulation with interferon-γ and bacterial lipopolysaccharides (LPS; which was measured at two different time-points), reported that 51.4% of the eQTLs detected before stimulation were not detectable after stimulation, sometimes in a time-dependent manner [61]. Additionally, a study assessing the effect of the stimulation of dendritic cells with LPS, influenza and interferon-β, showed 121 eQTLs associated with changes in gene expression due to these stimuli (response-eQTLs; *cis*-reQTLs) [62]. Like cell-type specific effects, stimulus dependent eQTLs appeared to have a larger distance between the SNP and the transcript compared to effects shared with unstimulated cells [61], and can affect specific transcription factor binding sites [62]. Both studies showed that trait-associated SNPs can have stimulus dependent effects, which provides further insight in the downstream effects of disease associated SNPs [61,62].

2.5. RNA-sequencing

So far, most eQTL mapping studies have measured gene expression levels using microarray technology. With the advent of NGS, the sequencing of RNA molecules (RNA-seq) has also become feasible. RNA-seq has a much larger dynamic range than microarray based gene expression quantification, and as such, a smaller amount of RNA molecules is required to accurately quantify gene expression levels [63,64]. The initial eQTL mapping studies performed using RNA sequencing data on LCLs have shown that the gene expression measurements between micro-arrays and RNA sequencing data generally correlated well (with correlations ranging between 0.6 and 0.781) [65]. As such, *cis*-eQTLs detected using RNA-seq replicated well when using microarray data, with up to 70% of the *cis*-eQTLs detected on microarrays being replicated in the Nigerian HapMap population [65], and up to 81% being replicated in a Central European HapMap population [66]. RNA-seq allows for a higher resolution of gene expression quantification than microarrays, since RNA-seq is not limited to a predefined set of oligonucleotide probes. Consequently, the RNA-seq studies on LCLs showed that *cis*-eQTL effects are not limited to annotated genes: in the Nigerian HapMap population, for example, the expression of 4031 unannotated transcripts was reported [65]. The higher resolution of RNA-seq also allows for better estimation of the correlation structure between exons and can thus be used to impute missing gene expression data for exons or transcripts [66], and it allows for better mapping of *cis*-eQTLs within exons. Comparing RNA-seq with an earlier microarray-based study on the same samples, RNA-seq-based eQTL mapping was better able to detect exon *cis*-eQTLs, most of which were located in genes with a high level of transcription, which indicates that RNA-seq is less prone to saturation of the gene expression signal, and that splicing complexity is not properly picked up by microarray-based studies [66]. Apart from exon-based *cis*-eQTLs, the

relative ratios of different transcript isoforms can also be used as a quantitative trait in RNA-seq-based studies, in order to detect splicing-QTLs (sQTLs): 187 and 110 significant sQTLs were detected in the Nigerian and Central European HapMap populations, respectively [65,66], 639 genes were detected with significant sQTLs in a more recent LCL based study of 462 individuals [67], and 2851 sQTLs were detected in a whole blood study of 922 individuals [68], which indicates SNPs also regulate gene expression through altering different transcript isoforms. sQTLs appear to originate from different regulatory variants than eQTLs, since sQTL SNPs show less enrichment near the 5′ end of genes compared to *cis*-eQTLs [68], but more enrichment in splice sites, 3′ untranslated regions (3′ UTR) and promoters [67]. Additionally, 57% of the eQTL genes that also showed an sQTL had an independent effect when conditioning for sQTLs, further indicating the independence between eQTL and sQTL regulation [67]. Overall, these studies show that genetic variation has a smaller influence on splicing than on overall gene expression. Finally, different RNA sequencing strategies can be used to answer different biological questions. For example, DeepSAGE, a sequencing strategy that uses primer sequences that specifically target the 3′ ends of genes, is more suitable for detecting gene expression variation near the 3′ ends than conventional RNA-seq, which, due to its random hexamer library design, shows larger fragmentation near the ends of genes [69]. A study applying the DeepSAGE method showed 12 poly-adenylation QTLs that transcript more often and have an altered 3′UTR length, but also showed that different RNA-seq strategies can be successfully meta-analyzed [69].

2.6. Allele-specific expression

Because microarrays are only able to measure gene expression levels as an average over all alleles in diploid organisms, local eQTL effects were previously annotated as being *cis*-eQTL effects, while the *cis* annotation conventionally implies that the gene expression variation originates from the same allele as the genetic variant (i.e. allelic imbalance of transcription, or allele specific expression (ASE)). The traditional way to measure ASE was through RT-PCR, precluding genome-wide assessment of ASE. By assessing the number of reads in heterozygote individuals, and by inferring haplotypes from reference datasets, RNA-seq is now able to determine ASE on a large scale: out of the 929 eQTLs from the Nigerian HapMap population, 222 had informative SNPs in exons, of which 88% showed ASE [65], while a more recent RNA-seq study found that 73.8% of the genes with ASE also showed an eQTL, most often originating from a high frequency variant, although 21% of the detected ASE instances did not overlap with any local eQTL [68]. ASE without an overlapping eQTL signal suggests that the ASE variants are rare and can thus be applied to detect local eQTL effects originating from rare variants [67,68]. A hypothesis that was further strengthened by the observation of larger regions of homozygosity surrounding ASE SNPs, when the ASE signal is shared by fewer than four individuals [66]. When eQTLs and ASE overlap, the number of reads in the ASE signal is correlated with the effect size of the eQTL [66], with stronger effect sizes for lower frequency genetic variants [68]. The mechanisms behind ASE are still unclear, as one study suggested that ASE is mediated by CREs [66], while another study suggested that ASE is genetic rather than epigenetic, or that it may be mediated by transcript structure variation [67]. ASE can also be assessed as a quantitative trait, in order to map aseQTLs: 641 SNPs with an eQTL also showed an aseQTL, some of which were located more than 1 megabase away from the TSS [68].

3. Future perspectives

GWASs continue to identify ever smaller effects from common variants and larger effect sizes from rare variants, using sample sizes that now exceed 100,000 individuals for various traits. eQTL studies have been successful in identifying *cis*-eQTL effects for many of the trait-associated variants and, to a lesser extent, in identifying downstream

effects through *trans*-eQTL mapping. However, although eQTL study sizes are increasing as well, sample sizes are still limited to detect eQTL effects for rare variants, although ASE analysis can help in this respect. Additionally, because the range of available eQTL studies on different cell types and tissues is limited, the question about the effects of trait-associated SNPs in different tissues remains largely unanswered. Future eQTL studies should therefore focus on three different levels in order to find the downstream effects of trait-associated variants (Fig. 3): 1) they should focus on increasing the sample size by meta-analysis to identify small-effect *cis*- and *trans*-eQTLs in bulk tissues (such as blood), 2) they should then determine in which specific cell types these eQTLs show larger effects, and 3) should further zoom in using single-cell RNA sequencing to identify the specific context (e.g. when a blood cell has been activated by a viral trigger) in which these eQTLs show particularly strong effects.

3.1. Larger meta-analyses

Current eQTL studies have identified *cis*-eQTL effects for the majority of genes. However, it is likely that *trans*-eQTL effects are even more numerous, although their effects are likely to be very small: of the 430 *trans*-eQTLs detected in the largest eQTL meta-analysis to date, more than 70% explained less than 1% of the gene expression variation [20]. Thus, in order to find more *trans*-eQTLs with even smaller effect size, eQTL studies should be scaled up, similar to what was done for GWAS. Such meta-analyses will also permit us to more accurately fine-map existing eQTL loci, will provide a higher-resolution overview of the downstream effects of both common and rare SNPs, and will permit causal inference. Additionally, because the cumulative sample size for a single tissue or cell type may not approach an adequate meta-analysis sample size to find small-effect eQTLs, several methods have now been developed that allow meta-analysis over different tissues simultaneously [70,71]. These large-scale meta-analyses will likely generate important biological insights into the downstream effects of trait-associated variants.

An important issue that remains is multiple testing. When performing *trans*-eQTL analyses, billions of statistical tests need to be conducted. However, with ever increasing knowledge on the genes that are involved in specific pathways, it will also become possible to leverage external biological knowledge on these pathways to improve the statistical power. By averaging the expression levels of multiple genes that work in a specific pathway, we will improve signal-to-noise ratios and thereby
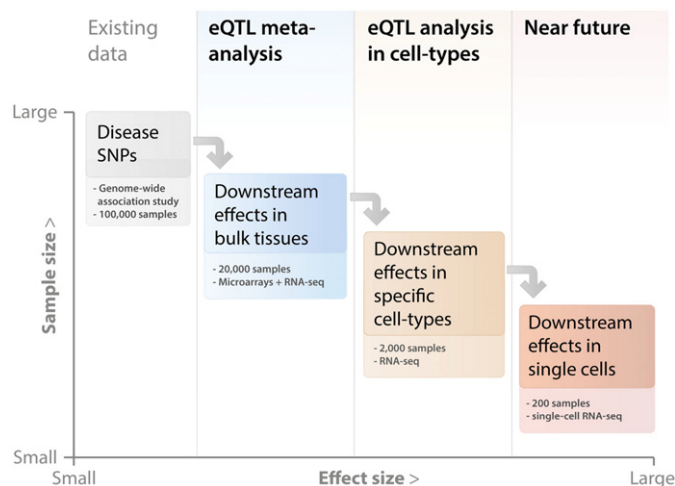
will need to perform fewer statistical tests. While such pathway-based eQTL methods have already been proposed [72], few studies have used them so far in a human setting [73].

Finally, another important issue is to have access to the large amounts of eQTL datasets that have been produced so far. For instance, human whole blood eQTL data is now available for over 20,000 samples. Joint re-analysis of this data will likely yield important biological insight into the downstream effects of many trait-associated variants. Additionally, it is possible to use such whole blood eQTL datasets to make inferences about the specific cell-types in which eQTLs manifest themselves (by using the abundance of such cell types as an interaction term and by performing meta-analysis across different datasets). However, while gene expression data is generally available (through databases such as the Gene Expression Omnibus, ArrayExpress, and more recently the European Nucleotide Archive), genotype data is available for only a limited number of eQTL datasets. This impedes progress on integrative approaches that can fully exploit such eQTL datasets to increase statistical power to identify smaller, but potentially very meaningful biological downstream eQTL effects. Initiatives such as dbGAP are therefore laudable, because they provide ways of sharing raw genotype data in a controlled and secure manner [74]. This might help to convince more researchers to make their data available to others (although researchers need to ensure that they have ethical approval and informed consent from their patients that raw genotype data can be exchanged). When (legal) hurdles preclude such sharing, alternative strategies, such as performing eQTL meta-analyses [20] (where no raw genotype data, but only summary statistics are being exchanged), might provide ways to share such data for gaining novel biological insight.

3.2. Larger tissue- and cell-type specific datasets

Although current studies have shown that numerous trait-associated variants act in a cell-type-specific way, it is likely that many cell-type-specific eQTLs have so far been missed. Additionally, current studies often lack the sample size that is necessary to determine the cell type specificity of *trans*-eQTL effects. As such, the question about what is the causal tissue or cell type for many diseases remains unanswered. To provide insight into this issue, large-scale studies are currently underway to assess many different tissues or cell types from the same individuals. The Genotype-Tissue Expression project (GTEx) [75], for example, aims to sample a range of tissues from a maximum of 900 samples. Gene expression is quantified using RNA-seq, which will enable the GTEx project to answer questions about the tissue specificity of *cis*- and *trans*-eQTLs, but will also provide insight into transcript isoform differences, ASE and differential exon usage between tissues, and their regulation by genetic variants. One of the aims of the GTEx project is to sample similar tissues to those used in the ENCODE project, which will add information about the tissue-specific epigenetic signals underlying the regulation of gene expression caused by genetic variants (e.g. DNASe-I hypersensitivity, various histone modifications, etc.) to the eQTL results. While the GTEx project focuses on tissues, the ImmVar project (http://www.immvar.org/) is focusing on a number of purified immunological cell types from approximately 600 individuals; it will be extended by an additional 28 cell types for approximately 60 individuals.

3.3. Single-cell eQTL analysis

So far, nearly every QTL study has been conducted while interrogating multiple cells per sample at once. Since many eQTL papers have moved from studying whole tissues to individual cell types, the next logical step would be to study individual cells. This is particularly interesting, because it will permit the identification of eQTLs that could well depend on the specific context in which a cell operates. Some effects of genetic variants on gene expression levels might manifest themselves only within those cells that have just been activated by a certain

Fig. 3. Future eQTL mapping studies will be focused on: 1) increasing the sample size through meta-analysis, in order to find more small-effect size eQTLs, 2) increasing the array of available tissues and cell types, in order to find cell-type-specific effects that are larger, and 3) single cell sequencing in order to identify context specific eQTLs that have large effect sizes.

external stimulus (e.g. viral, bacterial, or any other relevant trigger). If disease-associated SNPs only work in such a context, it could well be that such effects are not detectable when studying cells in bulk. The first paper to study this used single-cell RNA-seq and concentrated on 1440 single cells from 15 individuals [76]. They observed that many eQTLs are only detectable when studying single cell, and these would have been missed when the expression is averaged over multiple cells. Another attractive property of single-cell RNA-seq is that it might address a long-standing question in biology: how many different cell types are there for a given tissue (e.g. blood)? This can be addressed by performing single-cell RNA-seq and generating expression profiles on thousands of individual cells, with subsequent comparison of the expression profiles (e.g. through principal component analysis). Although many challenges still exist on how to generate and analyze single-cell RNA seq data reliably and robustly [77,78], this technology will likely mature quickly, leading to much lower prices and permitting many research groups to initiate single-cell eQTL studies in the near future.

3.4. Post-eQTL functional genomics

eQTL studies have now provided functional interpretation of many trait-associated SNPs and, with the various strategies outlined here and that are now being taken, future eQTL studies will likely yield substantial biological insights into many diseases. To gain an even better insight into the pathogenesis of these diseases, future integrative approaches that also ascertain the effect of trait-associated variants on different molecular levels (e.g. epigenetic levels, effects on protein levels, metabolite levels, or on the composition of the microbiome), and their possible interactions, will likely provide detailed mechanistic insights into the pathogenesis of many diseases.

References

[1] L.A. Hindorff, P. Sethupathy, H.A. Junkins, E.M. Ramos, J.P. Mehta, F.S. Collins, et al., Potential etiologic and functional implications of genome-wide association loci for human diseases and traits, Proc. Natl. Acad. Sci. U. S. A. 106 (2009) 9362–9367, http://dx.doi.org/10.1073/pnas.0903103106.

[2] M.J. Bamshad, S.B. Ng, A.W. Bigham, H.K. Tabor, M.J. Emond, D.A. Nickerson, et al., Exome sequencing as a tool for Mendelian disease gene discovery, Nat. Rev. Genet. 12 (2011) 745–755, http://dx.doi.org/10.1038/nrg3031.

[3] S.M. Purcell, N.R. Wray, J.L. Stone, P.M. Visscher, M.C. O'Donovan, P.F. Sullivan, et al., Common polygenic variation contributes to risk of schizophrenia and bipolar disorder, Nature 460 (2009) 748–752, http://dx.doi.org/10.1038/nature08185.

[4] J. Yang, B. Benyamin, B.P. McEvoy, S. Gordon, A.K. Henders, D.R. Nyholt, et al., Common SNPs explain a large proportion of the heritability for human height, Nat. Genet. 42 (2010) 565–569, http://dx.doi.org/10.1038/ng.608.

[5] R.C. Jansen, J.P. Nap, Genetical genomics: the added value from segregation, Trends Genet. 17 (2001) 388–391.

[6] B.E. Stranger, A.C. Nica, M.S. Forrest, A. Dimas, C.P. Bird, C. Beazley, et al., Population genomics of human gene expression, Nat. Genet. 39 (2007) 1217–1224, http://dx.doi.org/10.1038/ng2142.

[7] S.A. Monks, A. Leonardson, H. Zhu, P. Cundiff, P. Pietrusiak, S. Edwards, et al., Genetic inheritance of gene expression in human cell lines, Am. J. Hum. Genet. 75 (2004) 1094–1105, http://dx.doi.org/10.1086/426461.

[8] A.L. Dixon, L. Liang, M.F. Moffatt, W. Chen, S. Heath, K.C.C. Wong, et al., A genome-wide association study of global gene expression, Nat. Genet. 39 (2007) 1202–1207, http://dx.doi.org/10.1038/ng2109.

[9] E. Grundberg, K.S. Small, Å.K. Hedman, A.C. Nica, A. Buil, S. Keildson, et al., Mapping cis- and trans-regulatory effects across multiple tissues in twins, Nat. Genet. 44 (2012) 1084–1089, http://dx.doi.org/10.1038/ng.2394.

[10] A.C. Nica, L. Parts, D. Glass, J. Nisbet, A. Barrett, M. Sekowska, et al., The architecture of gene regulatory variation across multiple human tissues: the MuTHER study, PLoS Genet. 7 (2011) e1002003, http://dx.doi.org/10.1371/journal.pgen.1002003.

[11] G. Trynka, K.A. Hunt, N.A. Bockett, J. Romanos, V. Mistry, A. Szperl, et al., Dense genotyping identifies and localizes multiple common and rare variant association signals in celiac disease, Nat. Genet. 43 (2011) 1193–1201, http://dx.doi.org/10.1038/ng.998.

[12] J.J.B. Keurentjes, J. Fu, I.R. Terpstra, J.M. Garcia, G. van den Ackerveken, L.B. Snoek, et al., Regulatory network construction in Arabidopsis by using genome-wide gene expression quantitative trait loci, Proc. Natl. Acad. Sci. U. S. A. 104 (2007) 1708–1713, http://dx.doi.org/10.1073/pnas.0610429104.

[13] A. Viñuela, L.B. Snoek, J.A.G. Riksen, J.E. Kammenga, Aging uncouples heritability and expression-QTL in Caenorhabditis elegans, G3 (Bethesda) 2 (2012) 597–605, http://dx.doi.org/10.1534/g3.112.002212.

[14] B.M. Tesson, R.C. Jansen, eQTL analysis in mice and rats, Methods Mol. Biol. 573 (2009) 285–309, http://dx.doi.org/10.1007/978-1-60761-247-6_16.

[15] D.W. Huang, B.T. Sherman, R.A. Lempicki, Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources, Nat. Protoc. 4 (2009) 44–57, http://dx.doi.org/10.1038/nprot.2008.211.

[16] A.J. Myers, J.R. Gibbs, J.A. Webster, K. Rohrer, A. Zhao, L. Marlowe, et al., A survey of genetic human cortical gene expression, Nat. Genet. 39 (2007) 1494–1499, http://dx.doi.org/10.1038/ng.2007.16.

[17] F. Innocenti, G.M. Cooper, I.B. Stanaway, E.R. Gamazon, J.D. Smith, S. Mirkov, et al., Identification, replication, and functional fine-mapping of expression quantitative trait loci in primary human liver tissue, PLoS Genet. 7 (2011) e1002078, http://dx.doi.org/10.1371/journal.pgen.1002078.

[18] B.E. Stranger, M.S. Forrest, M. Dunning, C.E. Ingle, C. Beazley, N. Thorne, et al., Relative impact of nucleotide and copy number variation on gene expression phenotypes, Science 315 (2007) 848–853, http://dx.doi.org/10.1126/science.1136678.

[19] R.S.N. Fehrmann, R.C. Jansen, J.H. Veldink, H.-J. Westra, D. Arends, M.J. Bonder, et al., Trans-eQTLs reveal that independent genetic variants associated with a complex phenotype converge on intermediate genes, with a major role for the HLA, PLoS Genet. 7 (2011) 14, http://dx.doi.org/10.1371/journal.pgen.1002197.

[20] H.-J. Westra, M.J. Peters, T. Esko, H. Yaghootkar, C. Schurmann, J. Kettunen, et al., Systematic identification of trans eQTLs as putative drivers of known disease associations, Nat. Genet. 45 (2013) 1238–1243, http://dx.doi.org/10.1038/ng.2756.

[21] J.E. Powell, A.K. Henders, A.F. McRae, J. Kim, G. Hemani, N.G. Martin, et al., Congruence of additive and non-additive effects on gene expression estimated from pedigree and SNP data, PLoS Genet. 9 (2013) e1003502, http://dx.doi.org/10.1371/journal.pgen.1003502.

[22] B.E. Stranger, P.L. De Jager, Coordinating GWAS results with gene expression in a systems immunologic paradigm in autoimmunity, Curr. Opin. Immunol. 24 (2012) 544–551.

[23] J.-B. Veyrieras, S. Kudaravalli, S.Y. Kim, E.T. Dermitzakis, Y. Gilad, M. Stephens, et al., High-resolution mapping of expression-QTLs yields insight into human gene regulation, PLoS Genet. 4 (2008) e1000214, http://dx.doi.org/10.1371/journal.pgen.1000214.

[24] A.S. Dimas, S. Deutsch, B.E. Stranger, S.B. Montgomery, C. Borel, H. Attar-Cohen, et al., Common regulatory variation impacts gene expression in a cell type-dependent manner, Science 325 (2009) 1246–1250, http://dx.doi.org/10.1126/science.1174148.

[25] C.D. Brown, L.M. Mangravite, B.E. Engelhardt, Integrative modeling of eQTLs and cis-regulatory elements suggests mechanisms underlying cell type specificity of eQTLs, PLoS Genet. 9 (2013) e1003649, http://dx.doi.org/10.1371/journal.pgen.1003649.

[26] D.L. Nicolae, E. Gamazon, W. Zhang, S. Duan, M.E. Dolan, N.J. Cox, Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS, PLoS Genet. 6 (2010) e1000888, http://dx.doi.org/10.1371/journal.pgen.1000888.

[27] A.C. Nica, S.B. Montgomery, A.S. Dimas, B.E. Stranger, C. Beazley, I. Barroso, et al., Candidate causal regulatory effects by integration of expression QTLs with complex trait genetic associations, PLoS Genet. 6 (2010) e1000895, http://dx.doi.org/10.1371/journal.pgen.1000895.

[28] P.C.A. Dubois, G. Trynka, L. Franke, K.A. Hunt, J. Romanos, A. Curtotti, et al., Multiple common variants for celiac disease influencing immune gene expression, Nat. Genet. 42 (2010) 295–302, http://dx.doi.org/10.1038/ng.543.

[29] P. van der Harst, W. Zhang, I. Mateo Leach, A. Rendon, N. Verweij, J. Sehmi, et al., Seventy-five genetic loci influencing the human red blood cell, Nature 492 (2012) 369–375, http://dx.doi.org/10.1038/nature11677.

[30] A. Cvejic, L. Haer-Wigman, J.C. Stephens, M. Kostadima, P.A. Smethurst, M. Frontini, et al., SMIM1 underlies the Vel blood group and influences red blood cell traits, Nat. Genet. 45 (2013) 542–545.

[31] T.M. Frayling, N.J. Timpson, M.N. Weedon, E. Zeggini, R.M. Freathy, C.M. Lindgren, et al., A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity, Science 316 (2007) 889–894, http://dx.doi.org/10.1126/science.1141634.

[32] C. Dina, D. Meyre, S. Gallina, E. Durand, A. Körner, P. Jacobson, et al., Variation in FTO contributes to childhood obesity and severe adult obesity, Nat. Genet. 39 (2007) 724–726, http://dx.doi.org/10.1038/ng2048.

[33] S. Smemo, J.J. Tena, K.-H. Kim, E.R. Gamazon, N.J. Sakabe, C. Gómez-Marín, et al., Obesity-associated variants within FTO form long-range functional connections with IRX3, Nature 507 (2014) 371–375, http://dx.doi.org/10.1038/nature13138.

[34] W. Cookson, L. Liang, G. Abecasis, M. Moffatt, M. Lathrop, Mapping complex disease traits with global gene expression, Nat. Rev. Genet. 10 (2009) 184–194, http://dx.doi.org/10.1038/nrg2537.

[35] B.P. Fairfax, S. Makino, J. Radhakrishnan, K. Plant, S. Leslie, A. Dilthey, et al., Genetics of gene expression in primary immune cells identifies cell type-specific master regulators and roles of HLA alleles, Nat. Genet. 44 (2012) 502–510, http://dx.doi.org/10.1038/ng.2205.

[36] M. Heinig, E. Petretto, C. Wallace, L. Bottolo, M. Rotival, H. Lu, et al., A trans-acting locus regulates an anti-viral expression network and type 1 diabetes risk, Nature 467 (2010) 460–464, http://dx.doi.org/10.1038/nature09386.

[37] K.S. Small, A.K. Hedman, E. Grundberg, A.C. Nica, G. Thorleifsson, A. Kong, et al., Identification of an imprinted master trans regulator at the KLF14 locus related to multiple metabolic phenotypes, Nat. Genet. 43 (2011) 561–564, http://dx.doi.org/10.1038/ng.833.

[38] A.I. Su, M.P. Cooke, K.A. Ching, Y. Hakak, J.R. Walker, T. Wiltshire, et al., Large-scale analysis of the human and mouse transcriptomes, Proc. Natl. Acad. Sci. U. S. A. 99 (2002) 4465–4470, http://dx.doi.org/10.1073/pnas.012025199.

[39] E.E. Schadt, C. Molony, E. Chudin, K. Hao, X. Yang, P.Y. Lum, et al., Mapping the genetic architecture of gene expression in human liver, PLoS Biol. 6 (2008) e107, http://dx.doi.org/10.1371/journal.pbio.0060107.

[40] H. Zhong, J. Beaulaurier, P.Y. Lum, C. Molony, X. Yang, D.J. Macneil, et al., Liver and adipose expression associated SNPs are enriched for association to type 2 diabetes, PLoS Genet. 6 (2010) e1000932, http://dx.doi.org/10.1371/journal.pgen.1000932.

[41] J. Fu, M.G.M. Wolfs, P. Deelen, H.-J. Westra, R.S.N. Fehrmann, G.J. Te Meerman, et al., Unraveling the regulatory mechanisms underlying tissue-dependent genetic variation of gene expression, PLoS Genet. 8 (2012) e1002431, http://dx.doi.org/10.1371/journal.pgen.1002431.

[42] K. Hao, Y. Bossé, D.C. Nickle, P.D. Paré, D.S. Postma, M. Laviolette, et al., Lung eQTLs to help reveal the molecular underpinnings of asthma, PLoS Genet. 8 (2012) e1003029, http://dx.doi.org/10.1371/journal.pgen.1003029.

[43] J. Ding, J.E. Gudjonsson, L. Liang, P.E. Stuart, Y. Li, W. Chen, et al., Gene expression in skin and lymphoblastoid cells: refined statistical method reveals extensive overlap in cis-eQTL signals, Am. J. Hum. Genet. 87 (2010) 779–789, http://dx.doi.org/10.1016/j.ajhg.2010.10.024.

[44] B.E. Stranger, S.B. Montgomery, A.S. Dimas, L. Parts, O. Stegle, C.E. Ingle, et al., Patterns of cis regulatory variation in diverse human populations, PLoS Genet. 8 (2012) e1002639, http://dx.doi.org/10.1371/journal.pgen.1002639.

[45] E. Choy, R. Yelensky, S. Bonakdar, R.M. Plenge, R. Saxena, P.L. De Jager, et al., Genetic analysis of human traits in vitro: drug response and gene expression in lymphoblastoid cell lines, PLoS Genet. 4 (2008) e1000287, http://dx.doi.org/10.1371/journal.pgen.1000287.

[46] D. Mehta, K. Heim, C. Herder, M. Carstensen, G. Eckstein, C. Schurmann, et al., Impact of common regulatory single-nucleotide variants on gene expression profiles in whole blood, Eur. J. Hum. Genet. 21 (2013) 48–54, http://dx.doi.org/10.1038/ejhg.2012.106.

[47] M. Francesconi, B. Lehner, The effects of genetic variation on gene expression dynamics during development, Nature 505 (2014) 208–211, http://dx.doi.org/10.1038/nature12772.

[48] D.A. Smirnov, M. Morley, E. Shin, R.S. Spielman, V.G. Cheung, Genetic analysis of radiation-induced changes in human gene expression, Nature 459 (2009) 587–591, http://dx.doi.org/10.1038/nature07940.

[49] Y. Idaghdour, W. Czika, K.V. Shianna, S.H. Lee, P.M. Visscher, H.C. Martin, et al., Geographical genomics of human leukocyte gene expression variation in southern Morocco, Nat. Genet. 42 (2010) 62–67, http://dx.doi.org/10.1038/ng.495.

[50] C.E. Romanoski, S. Lee, M.J. Kim, L. Ingram-Drake, C.L. Plaisier, R. Yordanova, et al., Systems genetics analysis of gene-by-environment interactions in human cells, Am. J. Hum. Genet. 86 (2010) 399–410, http://dx.doi.org/10.1016/j.ajhg.2010.02.002.

[51] J.C. Maranville, F. Luca, A.L. Richards, X. Wen, D.B. Witonsky, S. Baxter, et al., Interactions between glucocorticoid treatment and cis-regulatory polymorphisms contribute to cellular response phenotypes, PLoS Genet. 7 (2011) e1002162, http://dx.doi.org/10.1371/journal.pgen.1002162.

[52] L.M. Franco, K.L. Bucasas, J.M. Wells, D. Niño, X. Wang, G.E. Zapata, et al., Integrative genomic analysis of the human immune response to influenza vaccination, elife 2 (2013) e00299, http://dx.doi.org/10.7554/eLife.00299.

[53] L.B. Barreiro, L. Tailleux, A.A. Pai, B. Gicquel, J.C. Marioni, Y. Gilad, Deciphering the genetic architecture of variation in the immune response to Mycobacterium tuberculosis infection, Proc. Natl. Acad. Sci. U. S. A. 109 (2012) 1204–1209, http://dx.doi.org/10.1073/pnas.1115761109.

[54] Y. Idaghdour, J. Quinlan, J.-P. Goulet, J. Berghout, E. Gbeha, V. Bruat, et al., Evidence for additive and interaction effects of host genotype and infection in malaria, Proc. Natl. Acad. Sci. U. S. A. 109 (2012) 16786–16793, http://dx.doi.org/10.1073/pnas.1204945109.

[55] T. Zeller, P. Wild, S. Szymczak, M. Rotival, A. Schillert, R. Castagne, et al., Genetics and beyond—the transcriptome of human monocytes and disease susceptibility, PLoS One 5 (2010) e10693, http://dx.doi.org/10.1371/journal.pone.0010693.

[56] A.S. Dimas, A.C. Nica, S.B. Montgomery, B.E. Stranger, T. Raj, A. Buil, et al., Sex-biased genetic effects on gene regulation in humans, Genome Res. 22 (2012) 2368–2375, http://dx.doi.org/10.1101/gr.134981.111.

[57] J.W. Kent, H.H.H. Göring, J.C. Charlesworth, E. Drigalenko, V.P. Diego, J.E. Curran, et al., , Genotype × age interaction in human transcriptional ageing, Mech. Ageing Dev. 133 (2012) 581–590, http://dx.doi.org/10.1016/j.mad.2012.07.005.

[58] K.D. Hansen, S. Sabunciyan, B. Langmead, N. Nagy, R. Curley, G. Klein, et al., Large-scale hypomethylated blocks associated with Epstein–Barr virus-induced B-cell immortalization, Genome Res. 24 (2014) 177–184, http://dx.doi.org/10.1101/gr.157743.113.

[59] C. Yao, R. Joehanes, A.D. Johnson, T. Huan, T. Esko, S. Ying, et al., Sex- and age-interacting eQTLs in human complex diseases, Hum. Mol. Genet. (2013) ddt582, http://dx.doi.org/10.1093/hmg/ddt582.

[60] H. White, A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity, Econometrica 48 (1980) 817–838.

[61] B.P. Fairfax, P. Humburg, S. Makino, V. Naranbhai, D. Wong, E. Lau, et al., Innate immune activity conditions the effect of regulatory variants upon monocyte gene expression, Science 343 (2014) 1246949, http://dx.doi.org/10.1126/science.1246949.

[62] M.N.H.M.N. Lee, C. Ye, A.-C.A.-C. Villani, T. Raj, W. Li, T.M. Eisenhaure, et al., Common genetic variants modulate pathogen-sensing responses in human dendritic cells, Science 343 (2014), http://dx.doi.org/10.1126/science.1246980 ((80-.),1246980–1246980).

[63] W. Sun, Y. Hu, eQTL mapping using RNA-seq data, Stat. Biosci. 5 (2013) 198–219, http://dx.doi.org/10.1007/s12561-012-9068-3.

[64] P.A.C. 't Hoen, M.R. Friedländer, J. Almlöf, M. Sammeth, I. Pulyakhina, S.Y. Anvar, et al., Reproducibility of high-throughput mRNA and small RNA sequencing across laboratories, Nat. Biotechnol. 31 (2013) 1015–1022, http://dx.doi.org/10.1038/nbt.2702.

[65] J.K. Pickrell, J.C. Marioni, A.A. Pai, J.F. Degner, B.E. Engelhardt, E. Nkadori, et al., Understanding mechanisms underlying human gene expression variation with RNA sequencing, Nature 464 (2010) 768–772, http://dx.doi.org/10.1038/nature08872.

[66] S.B. Montgomery, M. Sammeth, M. Gutierrez-Arcelus, R.P. Lach, C. Ingle, J. Nisbett, et al., Transcriptome genetics using second generation sequencing in a Caucasian population, Nature 464 (2010) 773–777, http://dx.doi.org/10.1038/nature08903.

[67] T. Lappalainen, M. Sammeth, M.R. Friedländer, P.A.C. 't Hoen, J. Monlong, M.A. Rivas, et al., Transcriptome and genome sequencing uncovers functional variation in humans, Nature 501 (2013) 506–511, http://dx.doi.org/10.1038/nature12531.

[68] A. Battle, S. Mostafavi, X. Zhu, J.B. Potash, M.M. Weissman, C. McCormick, et al., Characterizing the genetic basis of transcriptome diversity through RNA-sequencing of 922 individuals, Genome Res. 24 (2014) 14–24, http://dx.doi.org/10.1101/gr.155192.113.

[69] D.V. Zhernakova, E. de Klerk, H.-J. Westra, A. Mastrokolias, S. Amini, Y. Ariyurek, et al., DeepSAGE reveals genetic variants associated with alternative polyadenylation and expression of coding and non-coding transcripts, PLoS Genet. 9 (2013) e1003594, http://dx.doi.org/10.1371/journal.pgen.1003594.

[70] T. Flutre, X. Wen, J. Pritchard, M. Stephens, A statistical framework for joint eQTL analysis in multiple tissues, PLoS Genet. 9 (2013) e1003486, http://dx.doi.org/10.1371/journal.pgen.1003486.

[71] J.H. Sul, B. Han, C. Ye, T. Choi, E. Eskin, Effectively identifying eQTLs from multiple tissues by combining mixed model and meta-analytic approaches, PLoS Genet. 9 (2013) e1003491, http://dx.doi.org/10.1371/journal.pgen.1003491.

[72] S. Li, Q. Lu, Y. Cui, A systems biology approach for identifying novel pathway regulators in eQTL mapping, J. Biopharm. Stat. 20 (2010) 373–400, http://dx.doi.org/10.1080/10543400903572803.

[73] M. Rotival, T. Zeller, P.S. Wild, S. Maouche, S. Szymczak, A. Schillert, et al., Integrating genome-wide genetic variations and monocyte expression data reveals trans-regulated gene modules in humans, PLoS Genet. 7 (2011) e1002367, http://dx.doi.org/10.1371/journal.pgen.1002367.

[74] M. Gymrek, A.L. McGuire, D. Golan, E. Halperin, Y. Erlich, Identifying personal genomes by surname inference, Science 339 (2013) 321–324, http://dx.doi.org/10.1126/science.1229566.

[75] J. Lonsdale, J. Thomas, M. Salvatore, R. Phillips, E. Lo, S. Shad, et al., The Genotype-Tissue Expression (GTEx) project, Nat. Genet. 45 (2013) 580–585, http://dx.doi.org/10.1038/ng.2653.

[76] Q.F. Wills, K.J. Livak, A.J. Tipping, T. Enver, A.J. Goldson, D.W. Sexton, et al., Single-cell gene expression analysis reveals genetic associations masked in whole-tissue experiments, Nat. Biotechnol. 31 (2013) 748–752, http://dx.doi.org/10.1038/nbt.2642.

[77] L. Ning, G. Liu, G. Li, Y. Hou, Y. Tong, J. He, Current challenges in the bioinformatics of single cell genomics, Front. Oncol. 4 (2014) 7, http://dx.doi.org/10.3389/fonc.2014.00007.

[78] R. Sandberg, Entering the era of single-cell transcriptomics in biology and medicine, Nat. Methods 11 (2013) 22–24, http://dx.doi.org/10.1038/nmeth.2764