Review TRENDS in Genetics Vol.21 No.11 November 2005

Full text provided by www.sciencedirect.com
SCIENCE DIRECT
ELSEVIER

The quantitative genetics of transcription

Greg Gibson[1,2] and Bruce Weir[2]

[1]Department of Genetics, Gardner Hall, North Carolina State University, Raleigh, NC 37695-7614, USA
[2]Bioinformatics Research Center, North Carolina State University, Centennial Campus Partners II Building, Raleigh, NC 27695-7566, USA

Quantitative geneticists have become interested in the heritability of transcription and detection of expression quantitative trait loci (eQTLs). Linkage mapping methods have identified major-effect eQTLs for some transcripts and have shown that regulatory polymorphisms *in cis* and *in trans* affect expression. It is also clear that these mapping strategies have little power to detect polygenic factors, and some new statistical approaches are emerging that paint a more complex picture of transcriptional heritability. Several studies imply pervasive non-additivity of transcription, transgressive segregation and epistasis, and future studies will soon document the extent of genotype–environment interaction and population structure at the transcriptional level. The implications of these findings for genotype–phenotype mapping and modeling the evolution of transcription are discussed.

Introduction

Microarray-based gene expression profiling has provided a key with which quantitative geneticists can begin to open up the 'black box' that lies between genotype and phenotype. For almost a century, mathematical models of inheritance have assumed a fairly simple mapping of genetic onto phenotypic variation: allelic effects are modeled as if they have small and similar effects and act in a predominantly additive manner [1]. Although mechanistic intuition and an increasingly large body of empirical data [2] contradict these assumptions, there have not been any genome-scale methods to test for the prevalence of phenomena such as epistasis and genotype–environment interaction. An emerging approach is to ask whether the parameters of gene activity at the level of transcription regarding additivity, heritability and complexity parallel those of classical phenotypic traits.

The objectives of this research program are to apply quantitative and population genetic methodologies to the dissection of the genetic basis of disease and complex traits, and to generate sophisticated models of the evolution of transcription (and hence of morphological, physiological and behavioral evolution). To date, much of the literature, recently reviewed by Stamatoyannopoulos [3], has involved descriptive studies that apply standard statistical procedures to new transcriptomic data sets. In this review, we address three fundamental questions that have arisen naturally: how heritable is transcription, is it possible to identify genes that affect transcriptional variation, and how is transcriptional variation structured within populations? We then conclude with some observations on the potential impact of gene expression profiling on disease mapping and understanding the roles that drift and selection have in biological evolution.

The heritability of transcription

Estimates of genetic variance and heritability

Heritability refers to the proportion of the phenotypic variance among individuals in a population that can be attributed to genotypic (as opposed to environmental and random) differences. It is most commonly measured either by assessment of the resemblance among relatives, or by comparison of the variance among individuals having different genotypes with that observed among individuals having the same genotype [1]. Soon after the development of microarrays, studies of yeast [4,5], mice [6,7], flies [8–10], fish [11,12] and humans [13,14] all suggested a strong component of differential expression among genotypes, but because of experimental shortfalls most of these did not provide a direct estimate of heritability. Furthermore, any statement about the proportion of genes that vary because of genotype is necessarily a function of the statistical power of the experiment [7], in particular the number of replicates of each class of individual. There is a common misconception that the resolution of expression profiling is limited by microarray technology – and that verification techniques such as quantitative reverse transcription-polymerase chain reaction (RT-PCR) are more accurate – but actually the ability to detect small differences in transcript abundance is a statistical issue. Moderate replication can generally uncover differences as small as 1.5-fold, but if the resources permit – say, ten replicate measurements of several genotypes – there is no reason why differences as small as 1.2-fold or less cannot be attributed to the genotype. Nevertheless, there is now a reasonable expectation that for any tissue from any organism sampled under a particular set of environmental conditions, 10%–50% of the transcripts will be found to vary as a result of heritable differences [3].

Corresponding authors: Gibson, G. (ggibson@ncsu.edu), Weir, B. (weir@ncsu.edu).

Available online 8 September 2005

www.sciencedirect.com 0168-9525/$ - see front matter © 2005 Elsevier Ltd. All rights reserved. doi:10.1016/j.tig.2005.08.010

Several recent studies have turned to pedigree analysis for the measurement of heritability, notably using the Centre d' Etude du Polymorphisme Humain (CEPH) panel of transformed lymphoblasts [15]. Two different groups, using Affymetrix short-oligonucleotide [16] (http://www.affymetrix.com/index.affx) and Agilent long-oligonucleotide [17] arrays (http://www.agilent.com), independently concluded that a large proportion of the human lymphoblastoid transcriptome varies between individuals. These experiments used three-generation pedigrees consisting of four grandparents, two parents and up to ten children. Heritability of transcription for each gene can be measured from parent–offspring regression between each pair of generations, or using the complete data set in a restricted maximum likelihood framework. Ideally, heritability estimates computed independently for the same trait in a similar sample of individuals should be very similar. However, re-analysis of the CEPH data suggests that this is not the case: heritability estimates for all differentially expressed genes, inferred from grandparent to parent, parent to children, or across all three generations, are not correlated (D. Nielsen, personal communication). This means that sample sizes of the order of 100 individuals are too small to support robust estimates of heritability [18], or that there are unknown sources of experimental artifact (perhaps relating to sample processing in batches) that produce false positive measures of genetic variance components, or that the genetic components affecting transcription vary themselves from generation to generation as a result of the complexity of the architecture of the variation. We suspect that all three factors are operating, and although cautioning against too literal an interpretation of heritability (and hence expression QTL) measurements, note that these are early days and that much more robust estimates of quantitative genetic parameters of transcription are both needed and in progress.

Mutational variance

A complementary approach to demonstrating heritability is to monitor the divergence among replicate lines derived from an isogenic strain as they accumulate mutations over the course of several hundred generations [19]. This approach has been taken recently in studies of flies (S. Rifkin, personal communication) and nematodes [20], both of which arrived at two robust conclusions. First, genes are affected by mutations that alter transcript abundance considerably more rapidly than they accumulate changes in the DNA sequence. This observation is simply attributed to the likelihood that mutations in numerous genes can affect the transcription of any one locus, and that single transcription factors regulate numerous target genes. It implies that a new mutation is likely to be predominantly *trans* acting, a fact that is particularly intriguing in light of the finding (using a clever pyrosequencing assay) that transcriptional differentiation between species is at least as likely to be caused by *cis*- as *trans*-acting polymorphisms [21]. Second, mutation accumulation lines diverge at least an order of magnitude more rapidly than would be expected on the basis of observed levels of divergence among natural isolates. This is strong evidence that stabilizing selection is a potent force constraining divergence at the transcriptional level. Similar observations have been made regarding isolated morphological phenotypes, but this is a striking result applied to tens of thousands of expression phenotypes.

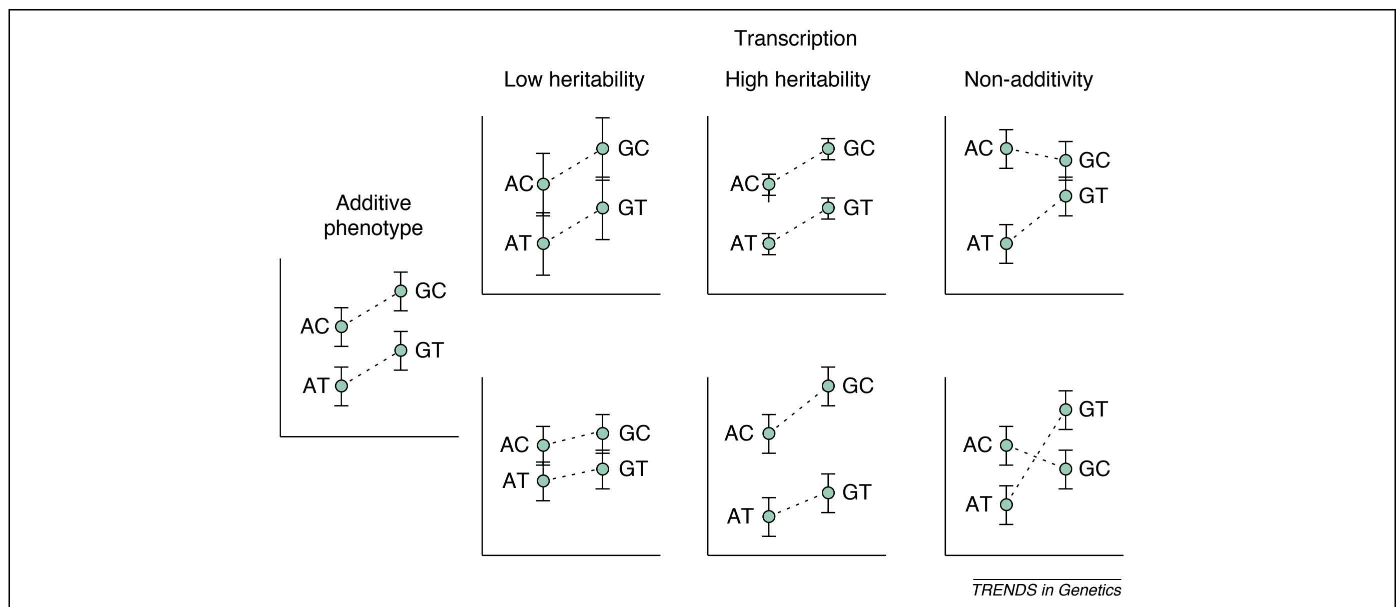

Figure 1. Genetic variance for transcription. The relationship between phenotypic and transcriptional heritability. Assume that two biallelic loci affect a trait with purely additive effects, then the mean value of the trait can be plotted on the *y*-axis for each of the four genotypes. At one locus, substitution of G for A (left to right in the panels) increases the trait value and at the other substitution of C for T (top to bottom in most panels) increases the trait value. Heritability is a function of the magnitude of the allelic effects and the variance among individuals. If transcription also varies among the four genotypes, we can draw similar plots where the trait on the *y*-axis is transcript abundance. Low heritability can arise either from high variance (top left panel) or small allelic effects (bottom left panel); and high heritability from low variance (top middle panel) or large allelic effects (lower middle panel). Non-additive effects result when one locus affects the allelic effects at the other one, possibly resulting in so-called 'crossing of line means' (lower right panel). Note that estimation of the variance of each genotype class is also affected by technical error, the relative magnitude of which between phenotypic and expression studies is unknown and probably dependent on the technology and degree of replication.

Clearly, mutation is not a limiting factor with regards to the production of variation for transcript abundance.

Non-additivity of transcription

The next question we can ask is whether transcriptional variation is largely additive or not. That is, if expression differs between two lines, and these are crossed, do the progeny show an intermediate level of mRNA? The first transcriptome-wide assessment of this question was conducted by Wayne, Nuzhdin and colleagues in a study of the reciprocal F1 of two isogenic strains of *Drosophila* [22]. Remarkably, twice as many genes differentiated either F1 from the parents as differentiated the two parental lines, and more detailed analysis of the modes of inheritance implied pervasive non-additivity. Numerous examples were observed of over- and under-dominance (F1 with higher or lower expression respectively than either parent), parent-of-origin, maternal and reciprocal F1 effects, indicating an unexpected complexity to the mapping of genotype onto transcriptional phenotype. Similar results have been observed in studies targeting specific candidate genes in maize [23] and wheat [24], in a massively parallel signature sequencing (MPSS) analysis of hybrid oysters [25] (D. Hedgecock, personal communication), and appear in another species of *Drosophila* [26] as well as in re-analysis of the aforementioned CEPH pedigrees (G. Gibson, unpublished data).

It is tempting to relate such non-additivity of transcription to phenomena such as heterosis and hybrid inviability, but it is not clear that this is appropriate. Maize and oyster breeding both rely on hybrid vigor to increase yield and so there is great interest in detecting genes that are differentially expressed in hybrids, but there is no evidence that this expression is responsible for phenotypic differentiation – particularly in regard to economically important traits. In fact, traditional models of heterosis assume either masking of deleterious recessive alleles in heterozygotes, or summation of the effects of a series of dominant alleles, neither of which need be reflected at the transcriptional level. There is, however, a pressing need for studies that explicitly relate transcriptional variation among genotypes to phenotypic variation [27].

Relationship to phenotypic heritability

A basic question in this regard is whether the heritability of transcription for genes that regulate a trait is greater or less than the heritability of the trait itself. Several possible scenarios are diagrammed in Figure 1; unfortunately no data sets that we are aware of address this fundamental problem. Given two regulatory polymorphisms that affect both transcription and a visible phenotype, it should be possible to measure both the mean and the variance of the phenotypes of each two-locus genotype. Transcriptional heritability might be relatively low either because small transcriptional differences are amplified into large phenotypic differences through later developmental processes, or because transcription is inherently 'noisy'. Conversely, transcriptional heritability might be relatively high either because later development tends to buffer large differences in transcript abundance, or because control of transcription is tight and additional noise is introduced post-transcriptionally.

The precision of the mapping of genotype onto phenotype will be strongly affected by the nature of the relative magnitudes of transcriptional and phenotypic heritability. Furthermore, it cannot be assumed that additive effects at the phenotypic level reflect additive contributions at the transcriptional level. A radical prospect, at least in light of classical quantitative genetic theory, is that epistasis for transcription is rampant, and that relatively chaotic gene expression is buffered at the level of translation, protein function and cellular interaction, producing an apparently more orderly phenotypic output. On this view, genetic additivity is more

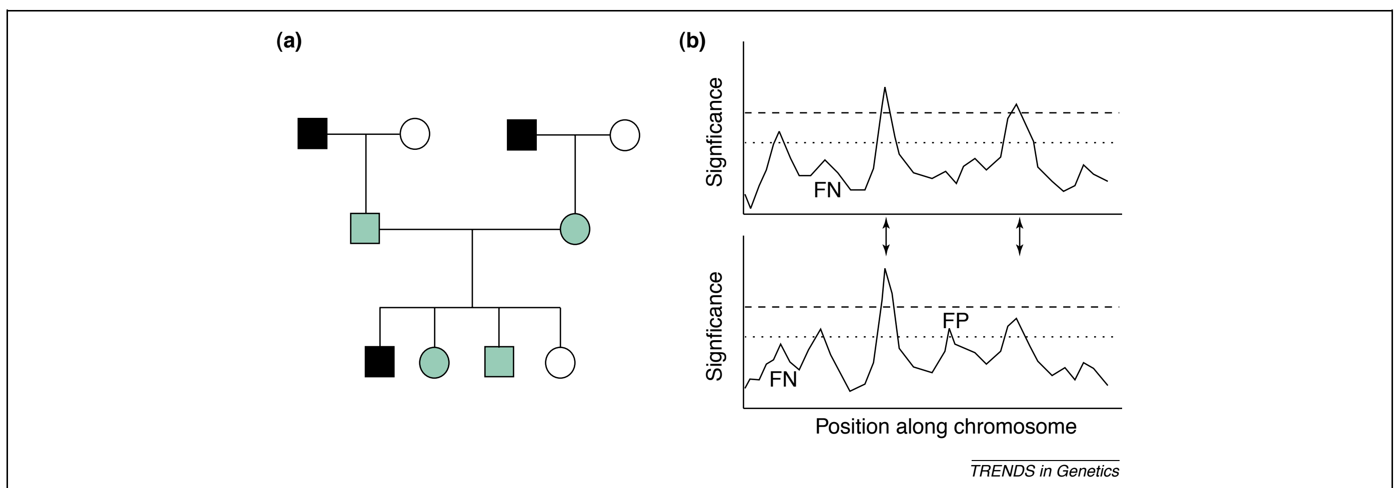

Figure 2. The principles of QTL mapping and genetical genomics. **(a)** QTL mapping is performed either in pedigrees or defined crosses. If the original lines are inbred, as implied here for male (black squares) and female (white circles) grandparents, their children will be genetically similar heterozygotes. Mapping is performed by scanning for markers that are associated with the phenotypes that segregate in the next generation. **(b)** Hypothetical profiles of two replicates of the same mapping experiment illustrate some of the issues involved in QTL detection. QTLs are identified as peaks of statistical significance walking along the chromosome: these must exceed an appropriate threshold indicated by the horizontal lines. Adopting a stringent threshold (upper broken lines) identifies only the strongest QTLs, indicated by the double arrowheads, but there is no guarantee that even these will appear in all replicates. Relaxing the stringency uncovers more peaks, but elevates the risk that some of these are false positives (FP) that appear by chance, whereas many QTLs will remain below the threshold as false negatives (FN). These problems are particularly acute when just 100 or fewer segregants are screened, as is usually the case in eQTL profiling, because of the expense of the experiments.

appropriately viewed as an averaging over thousands of possible transcriptional states than a reflection of precise mapping of genotype onto transcript abundance and in turn onto phenotype. Assessment of the quantitative relationship between these levels of variation is thus fundamental. It will also be fascinating to ask how often the variability of transcription measured across individuals is increased under perturbed circumstances, as it appears to be in a mouse model of Down syndrome [28], implying decanalization at the level of gene expression [29].

The genetic architecture of transcription

Expression QTL designs

Expression quantitative trait loci, or eQTLs, are peaks of statistical significance in a genome-wide scan for linkage between genetic markers and transcript abundance [30,31]. The basic experimental design of eQTL studies (Figure 2) is identical to that of classical F2 or recombinant inbred line linkage mapping for visible quantitative traits, except that tens of thousands of phenotypes (individual gene expression levels) are evaluated simultaneously. Studies have been reported to date for yeast [32,33], eucalyptus [34], mice [35–37], rats [38], maize [35] and humans [16,17,35]. In the rodent examples, eQTLs have been evaluated jointly with QTLs for obesity or other physiological traits, leading to novel insights that will not be discussed here. An example of a practical outcome has been the identification of a major-effect gene for lignin biosynthesis in forest trees [34].

The most notable conclusion from all of these studies is the ease with which at least one QTL can be identified for a majority of all transcripts that show heritable variation. Whereas visible trait variation is often described by several QTLs that collectively account for up to half of the genetic variance and individually rarely $>20\%$ of it, the summary in Table 1 shows that eQTLs accounting for 25%–50% of transcriptional variation are prevalent. These measures are likely to be overestimates of the magnitude of effect as a result of sampling biases in exploratory studies (Beavis effects [39]), and should not be extrapolated to the conclusion that single genes explain this proportion of the variation in a population. However, it is clear that major-effect QTLs are more prevalent than many investigators would have expected.

Cis- and trans-acting eQTLs

A second major conclusion is that up to one-third of eQTLs are *cis* acting, as summarized in Table 1. This means that the eQTL maps to the same genomic location as the gene whose variation it explains, at least within the limits of mapping resolution. In most of these cases, intuition suggests that the eQTL effect is likely to be caused by polymorphism in the regulatory region of the gene, namely sequence variation in the binding sites for transcription factors. Proof that this is the case requires either transgenic assays that demonstrate differential gene expression from constructs containing two different regulatory alleles, or demonstration of an association between a regulatory polymorphism and transcript abundance in a separate sample of individuals. Examples of the latter were provided by Morley *et al.* [16], who demonstrated that for several *cis*-acting eQTLs identified by linkage mapping between parents and offspring in the CEPH pedigrees, promoter polymorphisms showed association with transcript abundance in the grandparents.

Remarkably, several of these instances suggest that the regulatory polymorphism accounts for $>50\%$ of the transcriptional variance, resulting in P values for the test statistic $<10^{-10}$, despite sample sizes of <50 individuals. This result might not be surprising to experimentalists who are familiar with cellular transfection studies, which often demonstrate that a regulatory polymorphism has a measurable impact on transcription. From a quantitative genetic standpoint, though, magnitudes of effect this large can rarely be attributed to single QTLs associated with directly observable phenotypes. Data such as this support the view that transcriptional heritability is likely to be higher than phenotypic heritability, at least in some circumstances.

Equally impressively, most of these studies also detect eQTL ‘hotspots’ that explain variation for multiple transcripts. The straightforward interpretation of these cases is that the eQTL identifies a regulatory gene that coregulates as many as 25 downstream targets (but see Refs [18,40] for a more critical interpretation).

Table 1. Summary of eQTL studies

Tissue	*n*	Platform	Genes on array	Differentially expressed genes	Markers[a]	eQTL (threshold)[b]	*cis* eQTLs[c]	Refs
Mouse liver	111	Agilent	23 574	7861	13 cM	4339 (gwp 0.05)	34%	[35]
Maize	76	Agilent	24 473	18 805	12 cM	7372 (LOD 3.0)	Not applicable	[35]
Yeast	86	Custom open reading frames	6200	2294	3114	1011 ($P<0.000\,04$)	25%	[33]
Eucalyptus	91	cDNA	2608	12[d]	10 cM	9 (gwp 0.05)	22%	[34]
Human CEPH	112	Affymetrix	8500	3554	1 cM	142 (gwp 0.001)	19%	[16]
Human CEPH	167	Agilent	23 499	2499	4 cM	132 ($P<0.0005$)	19%	[16]
Rat kidney	30	Affymetrix	15 923	1553	1011	2490 (gwp 0.05)	30%	[38]
Rat fat body	30	Affymetrix	15 923	2046	1011	2118 (gwp 0.05)	35%	[38]
Mouse HSC	30	Affymetrix	12 422	Not applicable	779	352 (gwp 0.01)	33%	[36]
Mouse brain	35	Affymetrix	12 422	608	779	88 (fdr 0.10)	92%	[37]

[a]Some studies report number of markers, whereas others report approximate marker density in cM.
[b]Reported significance threshold: gwp, genome-wide *P* value; fdr, false discovery rate; LOD, likelihood cutoff.
[c]The proportion of eQTLs that map to the same interval as the transcript (this proportion is also affected by statistical thresholds).
[d]These authors only reported eQTL analysis for a subset of enzymes involved in lignin biosynthesis.

Permutation methods are used to define the maximum number of transcripts expected to map to one genomic interval given the number of observed linkage peaks. By relaxing stringency to adjust the balance between false positive and false negative detection of eQTLs, more hotspots can be identified, in the hope that key candidate genes can be located for more detailed evaluation. Although many of the pleiotropic eQTLs probably encode transcription factors, the analysis of the yeast data set that has the highest density of genotype markers [33] led to the confident conclusion that many of them also encode a range of different types of regulatory gene that could have indirect effects on transcription of the target gene.

It should also be noted that the proportion of eQTLs that are found to be *trans* acting tends to drop as the stringency for eQTL detection is increased [41]. Several of the studies suggest that *cis*-acting eQTLs typically have larger effects on transcription, and so tend to be associated with smaller *P* values. Indeed, the last study reported in Table 1 adopted a very stringent statistical cutoff, and perhaps as a consequence found that the majority of eQTLs were *cis* acting. A further concern relates to the possible impact of sequence polymorphism on estimation of transcript abundance in microarray studies, which will also tend to elevate the number of *cis*-acting eQTLs detected. Some of these could be false positives as a result of linkage disequilibrium between regulatory and coding single nucleotide polymorphisms (SNPs) [41].

Transcriptional variation is probably highly polygenic

It is important to recognize that even in the cases where a major-effect eQTL explains half of the genetic variance for transcript abundance, the other half remains to be accounted for, and in most cases will be caused by undetected loci. Because conservative thresholds of detection are required to adjust for the extraordinarily large number of comparisons involved in a genome-wide linkage scan for several thousand transcripts (the so-called 'multiple comparison problem'), most true eQTLs remain undetected. Only relatively few of the transcripts in the aforementioned studies are associated with multiple eQTLs, generally because the main effects of secondary eQTLs (those that explain <20% of the variance) are below the detection threshold. Again, this is an experimental design and/or statistical problem, not a technical issue related to the use of microarrays. A rigorous method for detecting two-locus effects and interactions has just been introduced [42], and several groups are working on new strategies employing multiple interval mapping methods [43], or comprehensive model comparisons, in an attempt to gradually increase the resolution of eQTL mapping. Adoption of false discovery rates [44] will probably be crucial to this endeavor, but for the time being it is unlikely that epistatic interactions involving loci other than those that contribute major effects will be detected routinely. Impressive as they have been, eQTL approaches are also undoubtedly giving a biased view tending towards an oversimplified view of transcriptional regulation.

For this reason, Brem and Kruglyak recently turned the problem on its head [45], and instead of focusing on identifying eQTLs asked whether increasingly polygenic models explained more of the variance for transcription in the yeast data set. Their major conclusion is that transcription is more often likely to be highly polygenic than monogenic: only 3% of highly heritable transcripts are consistent with single locus inheritance, 18% suggest control by two loci, and >50% require at least five loci under an additive model (Figure 3). They also argue that more than half of all transcripts show transgressive segregation (transcript abundance in F2 progeny falls outside the range of both grandparents) and that >15% are better explained by models that include epistatic interaction. Clearly, the landscape of gene expression in yeast is genetically complex, and it is difficult to imagine that it will be anything but more complex in higher eukaryotes.

The population structure of transcription: transcriptional frequency classes and cliques

QTL mapping necessarily provides only a partial picture of the architecture of genetic variation, because it is restricted to analysis of the variation segregating in a cross between two strains. The most important question it leaves unanswered concerns the frequency distribution of QTL effects in natural populations: are the frequencies of the two eQTL alleles more often equally common, skewed towards one major allele, or are eQTL effects more often the result of rare alleles? Unfortunately, resolution of this issue requires dissection of the eQTL to the nucleotide level, and there are no high-throughput methods yet

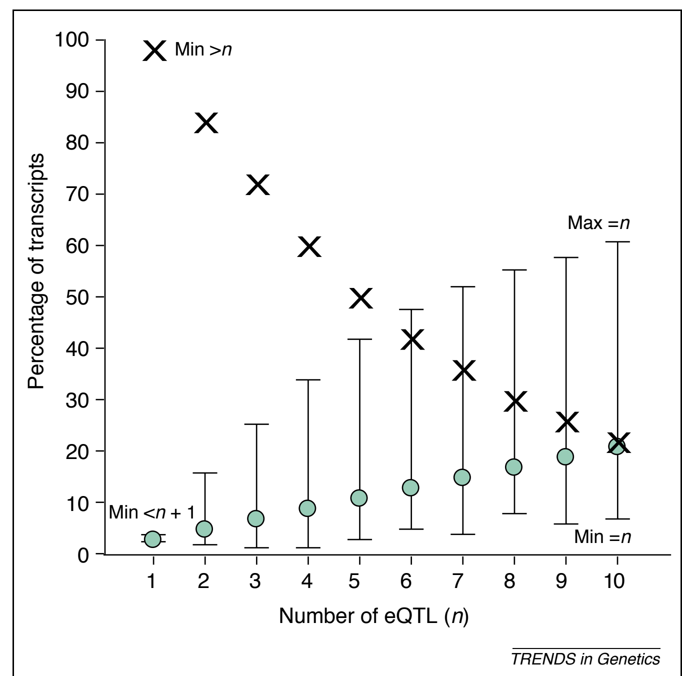

Figure 3. Inference of polygenic regulation from eQTL analysis. A plot of the data from Ref. [45] showing the range of complexity of transcriptional regulation in yeast inferred by their likelihood analysis. The error bars indicate the range of the percentage of differentially expressed transcripts in the F2 generation that are predicted to be regulated by *n* genes indicated on the *x*-axis. The large Xs indicate the minimum number of transcripts regulated by more than *n* eQTLs: for example, at least 20% of transcripts are predicted to have more than ten eQTLs. The green circles place a lower limit on the number of transcripts regulated by up to *n* eQTLs: for example, at least 10% of transcripts are regulated by four or fewer eQTLs.

available for this. With respect to transcription, though, we also need to address the distribution of transcript abundance across the transcriptome, as well as the covariance of transcript classes. Although a few studies have sampled dozens of individuals, none has yet addressed the statistical issues involved in defining the frequency distribution of transcript abundance. The core issues are schematized in Figure 4, which introduces the concept of 'transcriptional frequency classes' (TFCs), namely classes of transcript abundance among individuals in a population that can be distinguished from one another.

If major-effect eQTLs are indeed common, then it should be expected that gene expression will often depart from a normal frequency distribution. A dominant *cis*-acting regulatory polymorphism segregating at an allele frequency of 0.3, for example, might give rise to a bimodal transcript distribution with approximately equal proportions in each class. Large additive effects might even give rise to trimodal distributions. Unlike genotype frequencies, for which a large body of theory predicts expected distributions with the majority of minor allele frequencies $<5\%$, there is no theory predicting the distribution of transcriptional frequencies, because these will generally be influenced by both *cis*- and *trans*-acting loci. The detection of TFCs is complicated by the fact that departures from normality caused by skew and/or a small fraction of outliers are not easy to support statistically, and technical measurement error is likely to be high relative to genotypic differences when just one or two microarrays are performed per sample. Nevertheless, we have fitted mixture models [46] to data from 58 lines of *Drosophila* and using a Bayesian information criterion found evidence for a slight excess of multimodal transcript distributions suggesting that at least 5% of the transcriptome could show distinguishable frequency classes of transcription (W. Hsieh *et al*., unpublished data).

If major-effect eQTLs are *trans* acting and regulate several target transcripts in a similar manner, it might be further predicted that suites of transcripts are coregulated so that the frequency classes fall into clusters. We call these clusters 'transcriptional cliques', and note that the clustering is not the result of linkage and so might only weakly carry across generations. If a high degree of transcriptional clique structure is prevalent within populations, it would have profound implications for understanding the molecular basis of phenotypic covariance, evolutionary constraint, disease mapping and modeling transcriptional divergence. Measurement of the population structure of transcription is treated in more detail in Box 1.

Concluding remarks: towards models of the evolution of transcription

Quantitative characterization of transcript abundance within and among populations has several practical applications in biomedical and agricultural research, as well as evolutionary biology. These include scanning for transcripts whose abundance correlates with a quantitative phenotype, searching for functional associations among genes that fall into transcriptional cliques (genes that are coregulated often function together), and controlling for population structure in the search for transcripts that are associated with disease. Just as differences in allele and disease frequency between populations can lead to false positive attribution of genetic susceptibility to

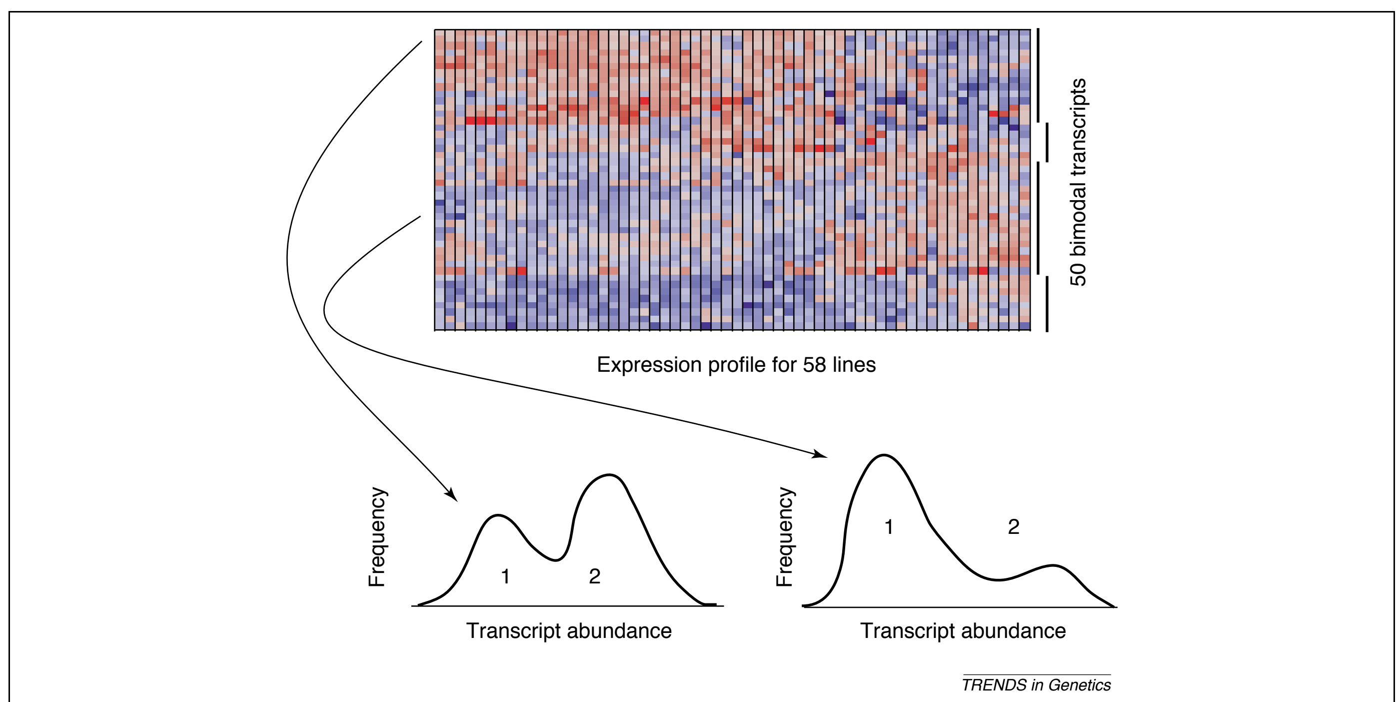

Figure 4. Hierarchical clustering of transcription in populations. A hypothetical heatmap (upper panel) showing ~50 genes (each row is a gene) that are bimodally expressed within a population sample (columns represent individuals). Red indicates relatively high transcript abundance; blue indicates relatively low transcript abundance. The genes cluster into 'transcriptional cliques' indicated by the bars to the right, which correspond to genes that are coregulated across individuals. Frequency distributions of transcript abundance (lower panel) for two representative genes show bimodality corresponding to TFCs at different frequencies in the sample.

Box 1. Measures of population structure

As argued in the main text, it is likely that there is considerable population structure to the distribution of transcript abundance. Aside from a handful of studies documenting differential expression between populations or species, theoretical and statistical treatment of this topic is lacking. We envisage two complementary approaches to robust definition of population structure, employing modified F_{ST} and Q_{ST} statistics [48,49]. Assuming that bimodal or even multi-modal transcript frequency classes can be detected, as suggested in Figure 4, expression in each individual can be assigned to a particular TFC (transcript frequency class), and the frequency distribution of these compared across populations. A quantitative analog of this approach known as Q_{ST} analysis seeks to determine to what degree continuous frequency distributions vary among populations.

Comparison of transcriptional Q_{ST} (tQ_{ST}) and genotypic F_{ST} statistics is a promising approach to detection of selection on transcript abundance, as first suggested by Lewontin [50], discussed in Refs [51–53], recently reviewed in Ref. [54] and implemented in relation to human disease in Refs [55,56]. The logic of the approach rests on the realization that F_{ST} serves both as a measure of population differentiation and as a measure of relationship within populations. For allele frequencies, the between-population variance component is proportional to F_{ST} statistics, whereas for a quantitative trait F_{ST} determines the covariance of trait values for pairs of individuals in the same population. For a haploid organism and a trait with additive genetic variance σ_A^2 and non-genetic variance σ_e^2, the trait value variance components within and between populations are $v_w = \left[(1-F_{ST})\sigma_A^2 + \sigma_e^2\right]$ and $v_b = \left(F_{ST}\sigma_A^2 + \sigma_e^2\right)$ respectively. The quantity defined as Q_{ST} for quantitative traits, $v_b/(v_b+v_w)$, is then algebraically equivalent to h^2F_{ST}, the product of heritability and F_{ST}. Consequently, the degree of divergence among populations in transcript abundance measured by tQ_{ST} is expected to be proportional to the narrow-sense heritability of transcription, and the inference of selection based on transcriptional divergence requires detailed knowledge of the genetic variance components underlying gene expression.

disease, the use of transcript abundance to define disease must be considered in the context of population variation [47].

The linkage between population genetic variation and divergence among species requires precise study of mutational, genetic and environmental variance components, all of which are now achievable at the gene expression level. However, technical and experimental design issues need to be addressed to handle the large data sets that are being generated, and new statistical tools are still being evaluated. Furthermore, novel population genetic theory will be required to deal with the covariance structure and potential non-additivity of transcript abundance. New data sets are needed that explicitly address the degree of family and population structure for gene expression and, where possible, evaluate gene expression simultaneously with visible phenotypic trait variation that is expected to be regulated by the transcripts being measured. It will be exciting to track the impact of this new transcriptional perspective on our basic understanding of biological evolution.

Acknowledgement
We thank colleagues too numerous to mention for discussions, but in particular Dahlia Nielsen, Dennis Hedgecock and Scott Rifkin for communicating unpublished results, and Glib Savych for an early version of Table 1. Work in our groups is supported in part by the NIH Program Project P01-GM45344.

References

1 Lynch, M. and Walsh, B. (1997) *Genetics and Analysis of Quantitative Traits*, Sinauer
2 Mackay, T.F. (2001) The genetic architecture of quantitative traits. *Annu. Rev. Genet.* 35, 303–339
3 Stamatoyannopoulos, J.A. (2004) The genomics of gene expression. *Genomics* 84, 449–457
4 Cavalieri, D. *et al*. (2000) Manifold anomalies in gene expression in a vineyard isolate of *Saccharomyces cerevisiae* revealed by DNA microarray analysis. *Proc. Natl. Acad. Sci. U. S. A.* 97, 12369–12374
5 Townsend, J.P. *et al*. (2003) Population genetic variation in genome-wide gene expression. *Mol. Biol. Evol.* 20, 955–963
6 Sandberg, R. *et al*. (2000) Regional and strain-specific gene expression mapping in the adult mouse brain. *Proc. Natl. Acad. Sci. U. S. A.* 97, 11038–11043
7 Pavlidis, P. and Noble, W.S. (2001) Analysis of strain and regional variation in gene expression in mouse brain. *Genome Biol.* 2. doi: 10.1186/gb-2001-2-10-research0042 (http://genomebiology.com/2001/2/10/research/0042)
8 Jin, W. *et al*. (2001) The contributions of sex, genotype and age to transcriptional variance in *Drosophila melanogaster*. *Nat. Genet.* 29, 389–395
9 Rifkin, S.A. *et al*. (2003) Evolution of gene expression in the *Drosophila melanogaster* subgroup. *Nat. Genet.* 33, 138–144
10 Meiklejohn, C.D. *et al*. (2003) Rapid evolution of male-biased gene expression in *Drosophila*. *Proc. Natl. Acad. Sci. U. S. A.* 100, 9894–9899
11 Oleksiak, M.F. *et al*. (2002) Variation in gene expression within and among natural populations. *Nat. Genet.* 32, 261–266
12 Oleksiak, M.F. *et al*. (2005) Natural variation in cardiac metabolism and gene expression in *Fundulus heteroclitus*. *Nat. Genet.* 37, 67–72
13 Whitney, A.R. *et al*. (2003) Individuality and variation in gene expression patterns in human blood. *Proc. Natl. Acad. Sci. U. S. A.* 100, 1896–1901
14 Radich, J.P. *et al*. (2004) Individual-specific variation of gene expression in peripheral blood leukocytes. *Genomics* 83, 980–988
15 Cheung, V.G. *et al*. (2003) Natural variation in human gene expression assessed in lymphoblastoid cells. *Nat. Genet.* 33, 422–425
16 Morley, M. *et al*. (2004) Genetic analysis of genome-wide variation in human gene expression. *Nature* 430, 743–747
17 Monks, S.A. *et al*. (2004) Genetic inheritance of gene expression in human cell lines. *Am. J. Hum. Genet.* 75, 1094–1105
18 de Koning, D.J. and Haley, C.S. (2005) Genetical genomics in humans and model organisms. *Trends Genet.* 21, 377–381
19 Mukai, T. (1964) The genetic structure of natural populations of *Drosophila melanogaster*. I. Spontaneous mutation rate of polygenes controlling viability. *Genetics* 50, 1–19
20 Denver, D.R. *et al*. (2005) The transcriptional consequences of mutation and natural selection in *Caenorhabditis elegans*. *Nat. Genet.* 37, 544–584
21 Wittkopp, P.J. *et al*. (2004) Evolutionary changes in *cis* and *trans* gene regulation. *Nature* 430, 85–88
22 Gibson, G. *et al*. (2004) Extensive sex-specific nonadditivity of gene expression in *Drosophila melanogaster*. *Genetics* 167, 1791–1799
23 Auger, D.L. *et al*. (2005) Nonadditive gene expression in diploid and triploid hybrids of maize. *Genetics* 169, 389–397
24 Sun, Q.X. *et al*. (2004) Differential gene expression patterns in leaves between hybrids and their parental inbreds are correlated with heterosis in a diallelic cross. *Plant Sci.* 166, 651–657
25 Hedgecock, D. *et al*. (2002) Analysis of gene expression in hybrid Pacific oysters by massively parallel signature sequencing. *Plant & Animal Genome X Conference Abstract* (http://www.intl-pag.org/pag/10/abstracts/PAGX_W15.html)
26 Wayne, M.L. *et al*. (2004) Additivity and trans-acting effects on gene expression in male *Drosophila simulans*. *Genetics* 168, 1413–1420
27 Birchler, J.A. *et al*. (2005) Dosage balance in gene regulation: biological implications. *Trends Genet.* 21, 219–226
28 Saran, N.G. *et al*. (2003) Global disruption of the cerebellar transcriptome in a Down syndrome mouse model. *Hum. Mol. Genet.* 12, 2013–2019
29 Gibson, G. and Dworkin, I. (2004) Uncovering cryptic genetic variation. *Nat. Rev. Genet.* 5, 681–690

30 Jansen, R.C. and Nap, J.P. (2001) Genetical genomics: the added value from segregation. *Trends Genet.* 17, 388–391
31 Farrall, M. (2004) Quantitative genetic variation: a post-modern view. *Hum. Mol. Genet.* 13, R1–R7
32 Brem, R.B. *et al*. (2002) Genetic dissection of transcriptional regulation in budding yeast. *Science* 296, 752–755
33 Yvert, G. *et al*. (2003) Trans-acting regulatory variation in *Saccharomyces cerevisiae* and the role of transcription factors. *Nat. Genet.* 35, 57–64
34 Kirst, M. *et al*. (2004) Coordinated genetic regulation of growth and lignin revealed by quantitative trait locus analysis of cDNA microarray data in an interspecific backcross of eucalyptus. *Plant Physiol.* 135, 2368–2378
35 Schadt, E.E. *et al*. (2003) Genetics of gene expression surveyed in maize, mouse and man. *Nature* 422, 297–302
36 Bystrykh, L. *et al*. (2005) Uncovering regulatory pathways that affect hematopoietic stem cell function using 'genetical genomics'. *Nat. Genet.* 37, 225–232
37 Chesler, E.J. *et al*. (2005) Complex trait analysis of gene expression uncovers polygenic and pleiotropic networks that modulate nervous system function. *Nat. Genet.* 37, 233–242
38 Hubner, N. *et al*. (2005) Integrated transcriptional profiling and linkage analysis for identification of genes underlying disease. *Nat. Genet.* 37, 243–253
39 Beavis, W.D. (1994) The power and deceit of QTL experiments: lessons from comparative QTL studies. In: *49th Annual Corn and Sorghum Research Conference*, pp. 252–268, American Seed Trade Association
40 Pérez-Enciso, M. (2004) *In silico* study of transcriptome genetic variation in outbred populations. *Genetics* 166, 547–554
41 Doss, S. *et al*. (2005) Cis-acting expression quantitative trait loci in mice. *Genome Res.* 15, 681–691
42 Storey, J.D. *et al*. (2005) Multiple locus linkage analysis of genomewide expression in yeast. *PLoS Biol.* e267
43 Kao, C.H. *et al*. (1999) Multiple interval mapping for quantitative trait loci. *Genetics* 152, 1203–1216
44 Storey, J.D. and Tibshirani, R. (2003) Statistical significance for genomewide studies. *Proc. Natl. Acad. Sci. U. S. A.* 100, 9440–9445
45 Brem, R.B. and Kruglyak, L. (2005) The landscape of genetic complexity across 5,700 gene expression traits in yeast. *Proc. Natl. Acad. Sci. U. S. A.* 102, 1572–1577
46 Yeung, K.Y. *et al*. (2001) Model-based clustering and data transformations for gene expression data. *Bioinformatics* 17, 977–987
47 Gibson, G. (2003) Population genomics: celebrating individual expression. *Heredity* 90, 1–2
48 Weir, B.S. and Hill, W.G. (2002) Estimating F-statistics. *Annu. Rev. Genet.* 36, 721–750
49 Ritland, K. (2000) Marker-inferred relatedness as a tool for detecting heritability in nature. *Mol. Ecol.* 9, 1195–1204
50 Lewontin, R.C. (1984) Detecting population differences in quantitative characters as opposed to gene frequencies. *Am. Nat.* 123, 115–124
51 Rogers, A.R. (1986) Population differences in quantitative characters as opposed to gene frequencies. *Am. Nat.* 127, 729–730
52 Felsenstein, J. (1986) Population differences in quantitative characters and gene frequencies: a comment on papers by Lewontin and Rogers. *Am. Nat.* 127, 731–732
53 Lewontin, R.C. (1986) A comment on the comments of Rogers and Felsenstein. *Am. Nat.* 127, 733–734
54 Merilä, J. and Crnokrak, P. (2001) Comparison of genetic differentiation at marker loci and quantitative traits. *J. Evol. Biol.* 14, 892–903
55 Rockman, M.V. *et al*. (2003) Positive selection on a human-specific transcription factor binding site regulating IL4 expression. *Curr. Biol.* 13, 2118–2123
56 Rockman, M.V. *et al*. (2004) Positive selection on MMP3 regulation has shaped heart disease risk. *Curr. Biol.* 14, 1531–1539

Elsevier celebrates two anniversaries with gift to university libraries in the developing world

In 1580, the Elzevir family began their printing and bookselling business in the Netherlands, publishing works by scholars such as John Locke, Galileo Galilei and Hugo Grotius. On 4 March 1880, Jacobus George Robbers founded the modern Elsevier company intending, just like the original Elzevir family, to reproduce fine editions of literary classics for the edification of others who shared his passion, other 'Elzevirians'. Robbers co-opted the Elzevir family's old printer's mark, visually stamping the new Elsevier products with a classic old symbol of the symbiotic relationship between publisher and scholar. Elsevier has since become a leader in the dissemination of scientific, technical and medical (STM) information, building a reputation for excellence in publishing, new product innovation and commitment to its STM communities.

In celebration of the House of Elzevir's 425th anniversary and the 125th anniversary of the modern Elsevier company, Elsevier will donate books to 10 university libraries in the developing world. Entitled 'A Book in Your Name', each of the 6 700 Elsevier employees worldwide has been invited to select one of the chosen libraries to receive a book donated by Elsevier. The core gift collection contains the company's most important and widely used STM publications including *Gray's Anatomy, Dorland's Illustrated Medical Dictionary, Essential Medical Physiology, Cecil Essentials of Medicine, Mosby's Medical, Nursing and Allied Health Dictionary, The Vaccine Book, Fundamentals of Neuroscience, and Myles Textbook for Midwives.*

The 10 beneficiary libraries are located in Africa, South America and Asia. They include the Library of the Sciences of the University of Sierra Leone; the library of the Muhimbili University College of Health Sciences of the University of Dar es Salaam, Tanzania; the library of the College of Medicine of the University of Malawi; and the libraries of the University of Zambia, Universite du Mali, Universidade Eduardo Mondlane, Mozambique; Makerere University, Uganda; Universidad San Francisco de Quito, Ecuador; Universidad Francisco Marroquin, Guatemala; and the National Centre for Scientific and Technological Information (NACESTI), Vietnam.

Through 'A Book in Your Name', the 10 libraries will receive approximately 700 books at a retail value of approximately 1 million US dollars.

For more information, visit www.elsevier.com