Geographical, Environmental and Pathophysiological Influences on the Human Blood Transcriptome

Rubina Tabassum · Artika Nath ·
Marcela Preininger · Greg Gibson

Published online: 1 October 2013
© Springer Science + Business Media New York 2013

Abstract Gene expression variation provides a read-out of both genetic and environmental influences on gene activity. Geographical, genomic and sociogenomic studies have highlighted how life circumstances of an individual modify the expression of hundreds and in some cases thousands of genes in a coordinated manner. This review places such results in the context of a conserved set of 90 transcripts known as blood informative transcripts that capture the major conserved components of variation in the peripheral blood transcriptome. Pathophysiological states are also shown to associate with the perturbation of transcript abundance along the major axes. Discussion of false-negative rates leads us to argue that simple significance thresholds provide a biased perspective on assessment of differential expression that may cloud the interpretation of studies with small sample sizes.

Keywords Axes of variation · Sociogenomics · Geographical genomics · eQTL · Differential expression

Introduction

Modern humans have been exposed to diverse environments in different geographic regions and have recently experienced demographic changes. Novel selection pressures related to different cultures, food habits, lifestyles and exposures to infections and toxins may be leading to local adaptation along with divergence in genetic architectures [1, 2]. Local adaptation due to geographically varying natural selection has caused parallel divergence events such that the same genetic variants undergo repeated changes, which are likely to be associated with phenotypic diversity among human populations in diverse geographical regions [3]. Accordingly, human populations are characterized by population-specific traits, and there is enormous diversity in phenotypes, including disease susceptibility and immunological response to external challenges, both within and between populations. High divergence in genes related to skin pigmentation, immune response, dietary adaptation, neuronal development and behavior have been discussed in relation to population-specific traits [4]. Ethnic and racial differences in disease susceptibility are also well documented [5]. Though the majority of the risk variants identified thus far through genome-wide association studies are common among populations of European and Asian origin and most have similar effect sizes in European and Asian populations, transferability to African ancestry is quite low [6].

Though genetics is expected to play a major role in shaping phenotypic diversity and development of population-specific traits, extensive efforts toward quantification of the relative contributions of genetic and environmental factors have revealed quite low genetic diversity between populations. Early studies based on protein polymorphisms estimated ~15 % between-group diversity [7, 8]. Further

R. Tabassum · A. Nath · M. Preininger · G. Gibson (✉)
School of Biology, Georgia Institute of Technology, 310 Ferst
Drive, Atlanta, GA 30332, USA
e-mail: ggibson.gt@gmail.com;
greg.gibson@biology.gatech.edu

R. Tabassum
e-mail: rubina.tabassum@gmail.com

A. Nath
e-mail: natar210@gmail.com

M. Preininger
e-mail: mpreininger@gmail.com

studies investigating mitochondrial DNA [9], Y chromosome markers [10] and many autosomal polymorphisms [11–14] provided similar results. Many investigators have since investigated genetic divergence at a larger scale using genome-scale genetic variation in sub-populations defined by geography, language and culture [15–19]. These studies often suggest that the proportion of genetic differences between human populations only slightly exceeds that between unrelated individuals within a population. Around 93–95 % of human genetic diversity is attributed to within-population variation and only 3–5 % to between populations [18]. The findings of these studies may imply a major contribution of environment and/or gene–environment interaction in shaping phenotypic divergence.

Gene expression is the key mechanism through which genomic information and environmental influence are translated to developmental and physiological characteristics. Studies quantifying gene expression variation within and between natural populations of various species including yeast, *Drosophila*, fish and humans have demonstrated differences in gene expression patterns both within and between populations [20, 21], which reflect phenotypic diversity. Gene expression has also been shown to be an important target of evolutionary constraints and selection pressure. Stabilizing selection is a major source of gene expression variation within populations, whereas positive selection has been a driving force for between-population divergence [22, 23]. Perturbation of gene expression patterns has been implicated in numerous diseases [24], particularly complex disorders that have strong environmental components such as type 2 diabetes, hypertension and autoimmunity. Moreover, the dynamic transcriptome may provide a better understanding of the role of diverse environments and complex gene–environment interactions in phenotypic traits and/or diseases than static genetic variation alone. Thus, within and between-population variation in transcriptome profiles under different geographical, environment and physiological conditions could provide important insight into the basis of phenotypic diversity, evolutionary history and disease progression.

In this review, we first briefly discuss the genomic control of gene expression and how regulatory polymorphisms could lead to different patterns of gene expression. Then, we discuss potential factors leading to variability in gene expression including ethnicity, geography, lifestyle and various physiological conditions, as well as technical factors. We show that the methods used to normalize data and identify gene expression signatures in different conditions strongly impact inference. We re-analyze several studies to demonstrate how peripheral blood transcription profiles are dominated by major axes of variation, with the consequence that a relatively small set of genes can

robustly capture many of the reported findings. Finally, we discuss the potential impact of false-negative findings in comparative transcriptome analysis.

Genomic Control of Gene Expression

In 2003, Cheung et al. [25] provided evidence for heritability of gene expression and familial aggregation of expression phenotypes by comparing variation among unrelated individuals, siblings and monozygotic twins. Since then, numerous studies have quantified the direct impact of genetic polymorphisms that regulate gene expression at genome-wide significance levels across both tissues and populations [26]. Variable DNA sequences such as single nucleotide polymorphisms (SNPs) and copy number variations are known to regulate gene expression and are referred as expression quantitative trait loci (eQTL). Numerous so-called *cis* and *trans* eQTLs have been identified, though *cis* (or local) eQTLs are predominant and tend to be located near transcription start sites where they have relatively large effects on the abundance of the adjacent transcript. eQTLs are also known to exert tissue/cell type-specific regulatory control on gene expression [27, 28]. Studies across multiple tissues and cell types have demonstrated an abundance of eQTLs in all tissues, but the level of overlap across tissues or cell types is heavily debated: some studies suggest that 69–80 % of eQTLs operate in a cell-type specific manner [27, 28], whereas more highly powered studies suggest much more cross-tissue activity [29]. Also, housekeeping genes are more likely to have significant genetic correlation between different RNA sources than non-housekeeping genes [30].

Epigenetic modification of the genome is another mechanism of regulation of gene expression in eukaryotic genomes, which involves non-sequence based modifications in genomic DNA, for example by DNA methylation [31]. Epigenetic modifications are believed to mediate the effect of non-genetic environmental factors on gene expression [31] (some of which are listed above and include many unknown factors). The differences in DNA methylation across populations have not yet been explored extensively, but two studies investigating DNA methylation patterns in skin showed a very low degree of inter-individual differences in DNA methylation patterns between ethnic groups [32, 33].

Ethnic Differences in Gene Expression

The majority of studies involving ethnic differences in gene expression patterns have focused on the correlation of expression profiles with genetic variants, typically in

Epstein-Barr virus-transformed lymphoblastoid cell lines. These studies have mainly concerned individuals of Caucasian, African and Asian ethnicity, obtained from publicly available cell lines from the HapMap Project (<http://hapmap.ncbi.nlm.nih.gov/>), and have provided contradictory reports. An initial study showed that gene expression patterns vary significantly between Asians and Europeans (1,097 of 4,197 genes tested) [34]. However, this result seems to have been confounded by a batch effect [35], and other studies imply that most of the observed variation in gene expression is due to within- rather than between-population variation. Storey et al. [36] found instead that only ~20 % of the genes are differentially expressed between populations, while ~87 % of the transcripts vary among individuals within populations. This conclusion was corroborated by Li et al. [37], who reported that gene expressions in Caucasians and Africans have similar levels of within-population variability, suggesting similar constraints or regulatory mechanisms in both populations. Genes with higher within-population variation tend to be involved in human diseases. Interestingly, genes with population-specific expression variability often have SNPs in their untranslated regions and show pronounced difference in population heterogeneity, suggesting genetic control of translational regulation as well.

eQTL regulation of gene expression also varies by ethnicity, though there is also some debate over the extent due to confounding technical issues. Spielman et al. [34] reported differences in allele frequencies in *cis* eQTLs and found that specific genetic variants contribute to differences in gene expression among populations. Conversely, Stranger et al. [38] showed that there is extensive sharing of eQTLs across Asian, European and African populations with similar effect size and direction of effect, though with diversity in allele frequencies. On similar lines, Price et al. [39] also demonstrated that both global and local differences in ancestry account for the observed differences in gene expression between Caucasian and African populations and that the differences in gene expression are widespread, mirroring genotypic differentiation.

Role of Geography and Lifestyle

Geography, lifestyle and food habits are known contributors to the development of many complex diseases such as cancer, diabetes and cardiovascular diseases. Nutritional, demographic, epidemiological and socioeconomic transitions over a few decades have resulted in drastic changes in environmental factors and lifestyles that have contributed to increased prevalence of chronic disorders [40]. This altered pathophysiology is also evident in significantly higher prevalence and incidence of lifestyle-related

disorders in urban than in suburban or rural locations. Reasoning that lifestyle could play an important role in regulating the expression of genes that govern susceptibility to chronic diseases, Idaghdour et al. [41] estimated the extent of influence of lifestyle and geography on gene expression by examining the transcriptome of genetically homogeneous individuals from three diverse settings in Morocco: urban, rural and nomadic. The authors found that almost one-third of the transcriptome had differential expression among regions. Neither genetic differentiation nor differential methylation was indicated, suggesting a minimal role of these factors in the expression divergence.

The findings of this ‘geographical genomics’ study were confirmed with an expanded sample that included Arab and Berber populations in the same region of Morocco and also demonstrated an almost complete absence of genotype-by-environment interactions on eQTL effects [42]. Each of almost 400 genome-wide significant eQTLs had the same direction and magnitude of effect in the urban and rural settings despite strong lifestyle differences. They discuss how this situation could nevertheless lead to $G \times E$ interactions at the phenotype level. Subsequently, Nath et al. [43] made similar observations in Fiji, where migrant Indians and native Melanesian Fijians living in rural and urban settings were found to diverge in peripheral blood gene expression, with the most significant differences observed between rural Indians and the other three groups of residents, to some extent involving similar genes as observed in Morocco.

Signatures of Gene Expression

Transcriptome profiling generates a large amount of data about the basal expression of genes and differential expression of genes in different conditions, but interpretation and comparison across studies pose unique challenges. Instead of considering individual transcript profiles, signatures/modules of gene expression capturing transcriptional covariance in a sample, based on coordinated expression and clustering of transcript profiles, could provide a better understanding of the clinical implications. Different methods have been employed to quantify gene expression covariance. These include principal component methods as well as independent component analysis [44] and non-negative matrix factorization [45], which typically generate many components of variation that each explain a few percent of the total transcriptome variance. Hierarchical clustering methods including weighted gene co-expression network analysis [46] and modulated modularity clustering [47] are more directed toward identifying modules of coregulated genes. Chaussabel et al. [48] adopted a module construction strategy from coordinately expressed

genes in blood from patients ($n = 239$) suffering from one of several conditions including systemic juvenile idiopathic arthritis, systemic lupus erythematosus (SLE), type I diabetes, metastatic melanoma, acute infections (*Escherichia coli*, *Staphylococcus aureus*) and influenza A, or liver-transplant recipients undergoing immunosuppressive therapy. The authors identified 28 modules incorporating 4,742 transcripts, which provided a stable framework for functional interpretation of transcripts profiles in disease conditions.

From this foundation, we recently showed that nine axes of covariance are consistently conserved across diverse population-profiling data sets [49•]. These axes capture as much as half of the total transcriptome variance in each study, with several hundred to several thousand transcripts highly significantly associated with each axis. Furthermore, a conserved set of 90 transcripts known as blood informative transcripts (BIT), ten for each axis, can effectively capture much of this transcriptional covariance [42, 49•]. The first principal component of each set of BIT defines an axis score, and these show high correlations with genetic, environmental and phenotypic variables. These BITs can thus be used to classify individuals based on their blood and immune functions. Gene ontology enrichment indicates that the axes are enriched for specific immune functions (for example, B or T lymphocyte, neutrophil signaling, etc.) and responses to environmental stimuli such as antiviral response.

Axes of Variation Correspond to Reported Signatures of Environmental or Pathophysiological Differences

BIT axes could be very useful in the context of pre-clinical risk prediction for diseases, without whole transcriptome profiling, since just 90 transcripts can be assayed using relatively inexpensive qRT-PCR techniques. Here we evaluate the feasibility of using BIT axes as predictors of transcriptional changes under different environmental, geographical and physiological conditions. We used transcriptional profiling data from previous studies that investigated changes in blood gene expression under various pathophysiological conditions including infectious diseases, chronic non-communicable diseases, response to vaccine and mental health and behavior and re-analyzed these data sets to evaluate the correlation of gene expression changes in these diverse conditions with the nine axes of variation. We downloaded expression data from the public microarray repository Gene Expression Omnibus (<http://www.ncbi.nlm.nih.gov/geo/>) and generated BIT axes after normalization by taking the first principal component (PC1) of abundance levels of the set of ten BIT in each axis as described in Preiner et al. [49•]. Differences

in axis scores in different conditions were tested using either t tests or ANOVA, and their correspondence with described expression signatures was assessed by linear regression. The results are summarized in Table 1.

Since the BIT axes have the potential to track inflammatory and metabolic traits, we evaluated whether they could be useful in the context of pre-clinical risk prediction for diseases using transcriptome profiles of healthy controls and patients for any of the following diseases from the Chaussabel et al. [48] study: systemic juvenile idiopathic arthritis (sJIA), SLE, type I diabetes, metastatic melanoma, Laennec's cirrhosis, urinary tract infection and hepatitis C. These diseases include a spectrum of disorders (chronic, infectious and autoimmune disease) that mainly involve perturbations in immune function and inflammation. Hepatitis C, sJIA, Laennec's cirrhosis, melanoma type 1 diabetes and SLE were correlated with axis 5, which is related to Toll-like receptor (TLR) signaling, inflammatory response and immunity (Fig. 1a). This is consistent with pathogenesis of these diseases related to immunity and inflammation. Hepatitis C infection has been associated with increased levels of chemokines and cytokines in the liver [50]. The innate immunity pathway is triggered during the onset of sJIA through the TLR pathway, which upregulates expression of proinflammatory cytokines via NF- κ B activation, in turn initiating an inflammatory cascade [51]. TLRs and cytokines have also been implicated in autoimmune diseases [52]. In SLE, elevated levels of TLR9 have been found in B cells, which drive the increase in IFN α that promotes SLE disease progression [53, 54]. In addition, perturbation of serum proinflammatory cytokines (tumor necrosis factor and interleukin-6) has been reported in active SLE progression [55, 56]. Interestingly, axis 1 and axis 9 were specifically associated with type 1 diabetes and thus could distinguish type 1 diabetes from other diseases; axis 2 distinguishes sJIA and urinary tract infection from the other diseases (Table 1).

Socioeconomic status (SES) during early life plays an important role in determining health status in adult life. Low SES has been shown to be associated with an inflammatory state and development of many chronic diseases such as obesity, diabetes and cardiovascular diseases later in life [57]. Miller et al. [58] investigated differences in gene expression profiles of healthy adults who were raised in either low early life SES or high early life SES. The authors reported a significant increase in expression of genes bearing response elements for CREB/ATF transcription factors, decreased expression of genes with response elements of the glucocorticoid receptor and upregulation of proinflammatory transcripts among subjects with low early life SES. Our axis analysis showed positive correlation of axis 9 and negative correlation of axis 1 respectively. Axis 1 is involved in T cell function

Table 1 Effect of environmental and physiological conditions on the human transcriptome and their association with blood informative transcript (BIT) axis scores (Color table online)

Study	Condition	BIT Axes								
		1	2	3	4	5	6	7	8	9
Nakaya et al. 2011	Response to Influenza TIV									
	HAI response (low vs high at day3)	■	■	■	■	■	■	■	■	■
	Correlation with day28 titer	■	■	■	■	■	■	■	■	■
	Correlation with day0 titer	■	■	■	■	■	■	■	■	■
	PC1 of DEGs after TIV	■	■	■	■	■	■	■	■	■
	PC1 of DEGs after LAIV	■	■	■	■	■	■	■	■	■
Montano et al. 2006	HIV-1 infection and perinatal									
	HIV positive vs HIV negative	■	■	■	■	■	■	■	■	■
	HIV neg vs TR	■	■	■	■	■	■	■	■	■
	HIV neg vs NTR	■	■	■	■	■	■	■	■	■
	TR vs NTR	■	■	■	■	■	■	■	■	■
	Viral load	■	■	■	■	■	■	■	■	■
Chaussabel et al. 2008	Disease conditions									
	Type 1 diabetes	■	■	■	■	■	■	■	■	■
	SLE	■	■	■	■	■	■	■	■	■
	JIA Systemic Onset	■	■	■	■	■	■	■	■	■
	Melanoma	■	■	■	■	■	■	■	■	■
	Laennec's Cirrhosis	■	■	■	■	■	■	■	■	■
	Hepatitis C	■	■	■	■	■	■	■	■	■
	Urinary Tract Infection	■	■	■	■	■	■	■	■	■
Cole et al. 2007	Social loneliness									
	High loneliness vs low loneliness	■	■	■	■	■	■	■	■	■
	PC1 of over-expressed genes	■	■	■	■	■	■	■	■	■
	PC1 of under-expressed genes	■	■	■	■	■	■	■	■	■
Miller et al. 2009	Early life SES									
	LSSES vs HSES	■	■	■	■	■	■	■	■	■
	PC1 of DEG	■	■	■	■	■	■	■	■	■
Idaghdour et al. 2008	Lifestyle/Geography									
	Urban vs Rural	■	■	■	■	■	■	■	■	■
	Urban vs Nomad	■	■	■	■	■	■	■	■	■
	Rural vs Nomad	■	■	■	■	■	■	■	■	■

The increasing shades of red from light red to dark red shows association in positive direction at p value range 0.01–0.05, 0.01–0.001 and <0.001 respectively, while shades of blue represent association in negative direction at different p value ranges. p values > 0.05 are shown as gray cells and white cells where data was not available

DEG differentially expressed genes, *PC1* first principal component, *SES* socio-economic status, *LSSES* low SES, *HSES* high SES

while axis 9 corresponds to the major component of life-style divergence in Morocco. Moreover, Cole and colleagues [59•] have also surveyed sociogenomic effects related to loneliness, depression and bereavement and argue that a consistent pattern of altered regulation of proinflammatory genes and antiviral responses known as the conserved transcriptional response to adversity (CTRA) underlies many behavioral gene expression profiles. We find that axes 1, 3, 5 and 7 consistently correspond to the major CTRA components (unpublished data).

As discussed above, lifestyle and geography influence the expression of genes as demonstrated by Idaghdour et al.'s [41] study of transcriptome profiles of individuals

with different lifestyles in geographically diverse regions of Morocco. Axis 3, which is comprised of genes involved in B-cell activation, was found to be significantly higher in individuals from urban regions compared to individuals from rural and nomad regions, with a higher difference between urban and nomadic lifestyles. In the follow-up study where a complex interaction among gender, lifestyle and geography was observed, axis 9 was very strongly associated with the major signature of cultural differentiation, as reported in [49•], and implicated aspects of T-cell regulation in functional divergence among populations.

Montano et al. [60] in 2006 demonstrated differential patterns of gene expression during HIV-1C infection as

Fig. 1 Potential of BIT axes as predictors of various conditions. **a** Pathophysiological conditions: the plot shows the mean scores of axis 5 in children with the indicated diseases [48]. **b** HIV-1C infection [60]: regression plot of PC1 of differentially expressed genes in HIV load versus axis 9 scores. **c** Loneliness [61]: regression plot of PC1 of

overexpressed genes in high loneliness versus axis 4 scores. **d** Response to influenza vaccination [62]: regression plot of PC1 of differentially expressed genes in response to vaccination for influenza versus axis 8 scores

well as with perinatal transmission outcome; they showed association of HIV-1C infection with upregulation of multiple groups of innate response genes, namely TLR and interferon activated antiviral response. Perinatal transmission was attributed to RNA metabolism (processing and splicing) genes, whereas non-transmitter mothers displayed two clusters of gene expression correlating with low and high viral load, with the high viral load pattern being similar to the profiles of transmitters. Consistently, BIT axis analysis shows that axes 4, 7 and 9, which correspond to RNA metabolism, antiviral response and a subset of T-cell function, respectively, are associated with the status of HIV-1C infection (Table 1). Interestingly, axis 9 also clearly separates control and non-transmitter mothers with low viral load from transmitter mothers and non-transmitter mothers with high viral load (Fig. 1b).

To evaluate whether BIT axes could capture mental health status, we performed a comparison between

chronically high or low levels of social isolation from the Chicago Health Aging and Social Relations Study cohort [61]. We generated the first principal component of over- and under-expressed genes in high loneliness individuals, respectively. Genes upregulated in high loneliness were involved in immune activation, transcriptional regulation and cell division, while downregulated genes were part of mature B lymphocyte function and type I interferon response [61]. BIT axis analysis provided consistent results; PC1 of the overexpressed genes was highly correlated with BIT axis 4 (Fig. 1c), while that of underexpressed genes was correlated with axes 7 and 8.

The efficacy and immunogenicity of vaccines depend on many factors including the immunocompetence of vaccines and pre-existing level of antibody titers. Nakaya et al. [62] generated gene expression profiles of healthy individuals vaccinated with either inactivated influenza (TIV) or live attenuated influenza (LAIV) over three consecutive seasons

at days 0, 3 and 7 post-vaccination. They identified early molecular signatures that could predict later antibody titers after vaccination and showed that expression of genes involved in B cell and T cell function correlates with the antibody titer at day 28. BIT axis analyses revealed that HAI titer responses are very strongly correlated with axis 6 and 8. The first principal component of differentially expressed genes after TIV was negatively correlated with axis 6 and positively correlated with axis 8. Axis 8 comprises genes involved in lymphocyte activation and RNA processing. Thus, this analysis confirms that T cell function is involved in the response to vaccination for influenza and axis 8 genes could be useful in early detection of differences in response. There is no clear gene ontology enrichment for axis 6, but the implication is that the antibody titer is strongly correlated with as many as 1,000 coregulated genes, only a few of which reach experiment-wide significance in the traditional hypothesis-testing framework. In addition, LAIV vaccination induces differential expression of a different subset of genes including enrichment for T cells, monocytes and natural killer cells [62]. Consistently, axis 1, which comprises genes involved in T cell physiology, also showed a strong association with PC1 of differentially expressed genes after LAIV vaccination. Moreover, axis 1 was also able to differentiate high and low responders. This analysis suggests that expression levels of a few genes comprised in axes 1, 6 and 8 could help to predict the response to different vaccines for influenza and could help to identify responders and non-responders in a population (Fig. 1d).

Altogether, the analyses presented here confirm that BIT axes are robust in capturing patterns of gene expression in various conditions and could be helpful in predicting clinical outcomes and disease risks as the consequences of different physiological and environmental conditions. They also imply that many inferences of differential expression that are presented as specific to the conditions of the particular study actually reflect a common set of underlying physiological responses in well-coordinated coexpression networks.

The Impact of False Negative Findings in Inference from Gene Expression Profiling

A corollary of axis-based analysis of gene expression is that it highlights the strong potential for false negative findings. While genome-wide association studies typically adopt a stringent significance threshold for association, the standard in gene expression profiling is to adopt a false discovery rate criterion in which the ratio of false positives to false negatives is balanced [63]. In both cases, it is generally assumed that all of the tests are independent, but the strong covariance of gene expression in coordinated axes suggests an alternative interpretation, illustrated in Fig. 2. We took the Morocco data set contrasting urban and rural samples [41] and divided it into two halves. All of the genes upregulated in the first half of the data set at $p < 10^{-4}$ ($NLP > 4$) in the rural samples are colored red, and those downregulated are colored blue in the volcano

Fig. 2 An example illustrating the potential for false-negative findings in gene expression profiling. The two volcano plots depict differential gene expression between rural and urban regions in two randomly divided halves of the data set from the Morocco study [41]. Differences in the mean of each gene expression level between rural and urban (x axis) are plotted versus the negative logarithm of p values for statistical significance (y axis). Up- and downregulated

genes ($-\log p > 4$) in rural samples in the first half of the data set are colored *red* and *blue*, respectively, in both (a) and (b). For almost all of the transcripts, the direction of differential expression is the same, but fewer than one-third are significant in both halves. A similar result is seen when the significant genes in the second half are colored in the first half (not shown) (Color figure online)

plot in Fig. 2a. The corresponding genes are similarly shaded for the second half of the data in Fig. 2b, where it can be clearly seen that the direction of differential expression is consistent for all of the genes even though only a minority of them are significant at $p < 10^{-4}$ in both studies. In this case, many of the genes are associated with axis 3, so the reason why there is such consistency in directionality is that the entire axis is coregulated. However, in each of the subsamples only a fraction of the differential expression is strong enough to exceed the set threshold. Consequently, despite the power of transcriptome profiling to detect expression differences as small as 1.5-fold in samples as small as 20 or so individuals of each class, it is apparent that there is a high false-negative rate for the genes in axis 3. That threshold-based cutoffs fail to account for the coregulation of gene expression is almost certainly a general phenomenon; hence, it is imperative that archives of transcriptional covariance in different tissues be developed.

Conclusion

Peripheral blood gene expression profiling experiments consistently reveal differences among populations, which can be associated with pathophysiological, demographic and geographic phenomena. When traditional significance criteria are used to interpret the results, there is a tendency to focus attention on dozens to hundreds of genes that will tend to be specific to the particular study and hence to conclude that those genes provide a somewhat unique signature for the condition. However, recognition that gene expression involves conserved axes suggests that often the differential expression actually reflects perturbation of a small set of underlying regulatory processes where the most significant changes are better regarded as representatives of the affected axis of several thousand genes. Local (or *cis*-) eQTLs act on top of this coexpression network to regulate individual genes, in some cases leading to genetic differences between populations due to allele frequency shifts. Environmental and transregulatory influences are less well characterized, but are the predominant influence on gene expression variation.

Acknowledgments Research in the Gibson laboratory is supported by Project 3 of the NIH Program Project grant awarded to the University of Washington, 1P01GM099568, and we thank Bruce Weir for his leadership of the initiative. Artika Nath was supported by a Fulbright Fellowship while contributing to this review.

Conflict of Interest All authors declare no conflicts of interest.

Human and Animal Rights and Informed Consent This article does not contain any studies with human or animal subjects performed by any of the authors.

References

Papers of particular interest, published recently, have been highlighted as

- Of importance

1. Tang K, Thornton KR, Stoneking M. A new approach for using genome scans to detect recent positive selection in the human genome. *PLoS Biol.* 2007;5:e171.
2. Laland KN, Odling-Smee J, Myles S. How culture shaped the human genome: bringing genetics and the human sciences together. *Nat Rev Genet.* 2010;11:137–48.
3. Tennessen JA, Akey JM. Parallel adaptive divergence among geographically diverse human populations. *PLoS Genet.* 2011;7:e1002127.
4. de Magalhães JP, Matsuda A. Genome-wide patterns of genetic distances reveal candidate loci contributing to human population-specific traits. *Ann Hum Genet.* 2012;76:142–58.
5. Smith MW, O'Brien SJ. Mapping by admixture linkage disequilibrium: advances, limitations and guidelines. *Nat Rev Genet.* 2005;6:623–32.
6. Marigorta UM, Navarro A. High *trans*-ethnic replicability of GWAS results implies common causal variants. *PLoS Genet.* 2013;9:e1003566.
7. Lewontin R. The partitioning of human genetic diversity. *Evol Biol.* 1972;6:381–98.
8. Nei M, Livshits G. Evolutionary relationships of Europeans, Asians, and Africans at the molecular level. *Hum Hered.* 1989;39:276–81.
9. Merriwether DA, Clark AG, Ballinger SW, et al. The structure of human mitochondrial DNA variation. *J Mol Evol.* 1991;33:543–55.
10. Hammer MF, Spurdle AB, Karafet T, et al. The geographic distribution of Y chromosome variation. *Genetics.* 1997;145:787–805.
11. Bowcock AM, Kidd JR, Mountain JL, et al. Drift, admixture, and selection in human evolution: a study with DNA polymorphisms. *Proc Natl Acad Sci USA.* 1991;88:839–43.
12. Batzer MA, Stoneking M, Alegria-Hartman M, et al. African origin of human-specific polymorphic Alu insertions. *Proc Natl Acad Sci USA.* 1994;91:12288–92.
13. Deka R, Guangyun S, Smelser D, et al. Rate and directionality of mutations and effects of allele size constraints at anonymous, gene-associated, and disease-causing trinucleotide loci. *Mol Biol Evol.* 1999;16:1166–77.
14. Stoneking M. DNA and recent human evolution. *Evol Anthropol.* 1993;2:60–73.
15. Jorde LB, Watkins WS, Bamshad MJ, et al. The distribution of human genetic diversity: a comparison of mitochondrial, autosomal, and Y-chromosome data. *Am J Hum Genet.* 2000;66:979–88.
16. Brown RA, Armelagos GJ. Apportionment of racial diversity: a review. *Evol Anthropol.* 2001;10:34.
17. Romualdi C, Balding D, Nasidze IS, et al. Patterns of human diversity, within and among continents, inferred from biallelic DNA polymorphisms. *Genome Res.* 2002;12:602–12.
18. Rosenberg NA, Pritchard JK, Weber JL, et al. Genetic structure of human populations. *Science.* 2002;298:2381–5.
19. Wang C, Zöllner S, Rosenberg NA. A quantitative comparison of the similarity between genes and geography in worldwide human populations. *PLoS Genet.* 2012;8:e1002886.
20. Whitehead A, Crawford DL. Variation within and among species in gene expression: raw material for evolution. *Mol Ecol.* 2006;15:1197–211.
21. Gibson G. The environmental contribution to gene expression profiles. *Nat Rev Genet.* 2008;9:575–81.

22. Kudaravalli S, Veyrieras JB, Stranger BE, et al. Gene expression levels are a target of recent natural selection in the human genome. *Mol Biol Evol*. 2009;26:649–58.
23. Lemos B, Meiklejohn CD, Cáceres M, Hartl DL. Rates of divergence in gene expression profiles of primates, mice and flies: stabilizing selection and variability among functional categories. *Evolution*. 2005;59:126–37.
24. Zhang F, Xu X, Zhang Y, et al. Gene expression profile analysis of type 2 diabetic mouse liver. *PLoS One*. 2013;8:e57766.
25. Cheung VG, Conlin LK, Weber TM, et al. Natural variation in human gene expression assessed in lymphoblastoid cells. *Nat Genet*. 2003;33:422–5.
26. Zhang W, Duan S, Kistner EO, et al. Evaluation of genetic variation contributing to differences in gene expression between populations. *Am J Hum Genet*. 2008;82:631–40.
27. Dimas AS, Deutsch S, Stranger BE, et al. Common regulatory variation impacts gene expression in a cell type-dependent manner. *Science*. 2009;325:1246–50.
28. Nica AC, Parts L, Glass D, et al. The architecture of gene regulatory variation across multiple human tissues: the MuTHER study. *PLoS Genet*. 2011;7:e1002003.
29. Price AL, Helgason A, Thorleifsson G, et al. Single-tissue and cross-tissue heritability of gene expression via identity-by-descent in related or unrelated individuals. *PLoS Genet*. 2011;7:e1001317.
30. Powell JE, Henders AK, McRae AF, et al. Genetic control of gene expression in whole blood and lymphoblastoid cell lines is largely independent. *Genome Res*. 2012;22:456–66.
31. Jaenisch R, Bird A. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nat Genet*. 2003;33:245–54.
32. Gröniger E, Weber B, Heil O, et al. Aging and chronic sun exposure cause distinct epigenetic changes in human skin. *PLoS Genet*. 2010;6:e1000971.
33. Winnefeld M, Brueckner B, Gröniger E, et al. Stable ethnic variations in DNA methylation patterns of human skin. *J Invest Dermatol*. 2012;132:466–8.
34. Spielman RS, Bastone LA, Burdick JT, et al. Common genetic variants account for differences in gene expression among ethnic groups. *Nat Genet*. 2007;39:226–31.
35. Akey JM, Biswas S, Leek JT, Storey JD. On the design and analysis of gene expression studies in human populations. *Nat Genet*. 2007;39:807–8.
36. Storey JD, Madeoy J, Strout JL, et al. Gene-expression variation within and among human populations. *Am J Hum Genet*. 2007;80:502–9.
37. Li J, Liu Y, Kim T, et al. Gene expression variability within and between human populations and implications toward disease susceptibility. *PLoS Comput Biol*. 2010;6:e1000910.
38. Stranger BE, Montgomery SB, Dimas AS, et al. Patterns of cis regulatory variation in diverse human populations. *PLoS Genet*. 2012;8:e1002639.
39. Price AL, Patterson N, Hancks DC, et al. Effects of *cis* and *trans* genetic ancestry on gene expression in African Americans. *PLoS Genet*. 2008;4:e1000294.
40. Gibson G. Decanalization and the origin of complex disease. *Nat Rev Genet*. 2009;10:134–40.
41. Idaghdour Y, Storey JD, Jadhav SJ, Gibson G. A genome-wide gene expression signature of environmental geography in leukocytes of Moroccan Amazighs. *PLoS Genet*. 2008;4:e1000052.
42. Idaghdour Y, Czika W, Shianna KV, et al. Geographical genomics of human leukocyte gene expression variation in southern Morocco. *Nat Genet*. 2010;42:62–7.
43. Nath AP, Arafat D, Gibson G. Using blood informative transcripts in geographical genomics: impact of lifestyle on gene expression in fijians. *Front Genet*. 2012;3:243.
44. Biswas S, Storey JD, Akey JM. Mapping gene expression quantitative trait loci by singular value decomposition and independent component analysis. *BMC Bioinformatics*. 2008;9:244.
45. Fogel P, Young SS, Hawkins DM, Ledirac N. Inferential, robust non-negative matrix factorization analysis of microarray data. *Bioinformatics*. 2007;23:44–9.
46. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*. 2008;9:559.
47. Stone EA, Ayroles JF. Modulated modularity clustering as an exploratory tool for functional genomic inference. *PLoS Genet*. 2009;5:e1000479.
48. Chaussabel D, Quinn C, Shen J, et al. A modular analysis framework for blood genomics studies: application to systemic lupus erythematosus. *Immunity*. 2008;29:150–64.
49. • Preininger M, Arafat D, Kim J, et al. Blood-informative transcripts define nine common axes of peripheral blood gene expression. *PLoS Genet*. 2013;9:e1003362. *This study demonstrates that there are at least nine conserved axes of gene expression variation in peripheral blood and defines the concept of blood informative transcripts.*
50. Li K, Li NL, Wei D, et al. Activation of chemokine and inflammatory cytokine response in hepatitis C virus-infected hepatocytes depends on toll-like receptor 3 sensing of hepatitis C virus double-stranded RNA intermediates. *Hepatology*. 2012;55:666–75.
51. Mellins ED, Macaubas C, Grom AA. Pathogenesis of systemic juvenile idiopathic arthritis: some answers, more questions. *Nat Rev Rheumatol*. 2011;7:416–26.
52. Tadema H, Abdulahad WH, Stegeman CA, et al. Increased expression of toll-like receptors by monocytes and natural killer cells in ANCA-associated vasculitis. *PLoS One*. 2011;6:e24315.
53. Papadimitrakaki ED, Choulaki C, Koutala E, et al. Expansion of toll-like receptor 9-expressing B cells in active systemic lupus erythematosus: implications for the induction and maintenance of the autoimmune process. *Arthritis Rheum*. 2006;54:3601–11.
54. Anders HJ, Vielhauer V, Eis V, et al. Activation of toll-like receptor-9 induces progression of renal disease in MRL-Fas(lpr) mice. *FASEB J*. 2004;18:534–6.
55. Jacob N, Stohl W. Cytokine disturbances in systemic lupus erythematosus. *Arthritis Res Ther*. 2011;13:228.
56. Sabry A, Sheashaa H, El-Husseini A, et al. Proinflammatory cytokines (TNF- α and IL-6) in Egyptian patients with SLE: its correlation with disease activity. *Cytokine*. 2006;35:148–53.
57. Juonala M, Viikari JS, Raitakari OT. Main findings from the prospective Cardiovascular Risk in Young Finns Study. *Curr Opin Lipidol*. 2013;24:57–64.
58. Miller GE, Chen E, Fok AK, et al. Low early-life social class leaves a biological residue manifested by decreased glucocorticoid and increased proinflammatory signaling. *Proc Natl Acad Sci USA*. 2009;106:14716–21.
59. • Irwin MR, Cole SW. Reciprocal regulation of the neural and innate immune systems. *Nat Rev Immunol*. 2011;11:625–32. *A stimulating review of the evidence that a conserved regulatory network governs socio-genomic aspects of immune function.*
60. Montano M, Rarick M, Sebastiani P, et al. Gene-expression profiling of HIV-1 infection and perinatal transmission in Botswana. *Genes Immun*. 2006;7:298–309.
61. Cole SW, Hawkey LC, Arevalo JM, et al. Social regulation of gene expression in human leukocytes. *Genome Biol*. 2007;8:R189.
62. Nakaya HI, Wrammert J, Lee EK, et al. Systems biology of vaccination for seasonal influenza in humans. *Nat Immunol*. 2011;12:786–95.
63. Storey JD, Tibshirani R. Statistical significance for genome wide studies. *Proc Natl Acad Sci USA*. 2003;100:9440–5.