

Minireview

Measuring *cis*-acting regulatory variants genome-wide: new insights into expression genetics and disease susceptibility

Wolfgang Sadee

Address: Program in Pharmacogenomics, College of Medicine, The Ohio State University, Columbus, OH 43210, USA.
Email: wolfgang.sadee@osumc.edu

Abstract

Regulatory polymorphisms have emerged as a prevalent source of phenotypic variability, capable of driving rapid evolution. mRNA profiling combined with genome-wide genotyping of polymorphisms has revealed pervasive genetic influences on gene expression, acting both *in cis* and *in trans*. Measuring allelic ratios of RNA transcripts makes it possible to focus on *cis*-acting factors separately from *trans*-acting processes. Using large-scale allelic expression analysis, a recent study by Ge and colleagues demonstrates a high incidence of *cis*-acting regulatory variants, promising insights into the ‘missing heritability’ component of complex disorders. Here, I evaluate their results and discuss the limitations of the current approach and avenues for exploring disease risk, guiding successful therapy, early intervention, and prevention.

Introduction

Advances in large-scale genotyping and DNA sequencing have yielded unprecedented insights into human genomic diversity, and yet a large proportion of genetic risk factors for complex human diseases remains unknown. How can we shed light on the ‘missing heritability’ [1]? Whereas genetics has traditionally focused on nonsynonymous polymorphisms that alter the encoded amino acid sequence (coding single nucleotide polymorphisms (SNPs); the term ‘SNP’ is used here for all variants), the focus has now shifted to regulatory variants (rSNPs), which are likely to be more prevalent than coding SNPs. Suspected as being a primary driver of evolution [2-4], rSNPs can undergo positive selection, potentially reaching high frequency. Intense exploration of regulatory variants has been accelerated by new genomic technologies. Here, I discuss the findings of a recent genome-wide analysis of regulatory variation [5], which is among the largest of such studies conducted so far. In a broader context, I further assess new avenues that could lead to a better understanding of human health and disease.

Measuring *cis*- and *trans*-acting factors in mRNA expression

Several studies have used expression arrays to measure mRNA levels and coupled this with genome-wide SNP

analyses, mostly in transformed lymphocytes. mRNA levels can then serve as quantitative phenotypes, and associations can be found with genomic regions (expression quantitative trait loci or eQTLs) that act either *in cis* or *in trans*, depending on whether the eQTL maps to the same gene as the measured mRNA or to another genomic region [6-10] (Figure 1). This approach reveals that mRNA expression is subject to pervasive genetic factors, which are mostly located *in cis*. On the other hand, if one measures allelic mRNA expression, any differences between expression from one allele compared with the other reveals the presence of *cis*-acting regulatory factors, and not *trans*-acting influences (Figure 1) [5,11-13].

Ge *et al.* [5] measured genome-wide allelic expression (AE) differences on Illumina Human1M BeadChips in lymphoblastoid cells; they then compared these with allelic genomic DNA ratios to detect AE imbalance (AEI). Using multiple filters, they detected AE ratios of ± 0.05 deviation from unity, confirming pervasive *cis* regulation. The loci with AEI involved 30% of the measured RefSeq transcripts and extended to unannotated transcripts. Varying estimates of AEI prevalence are a result of different cutoff values for AE ratios, methodology, and numbers of individuals studied [11-13]. The simultaneous availability of genome-wide SNP analysis enabled further fine mapping of the *cis*-eQTLs, which showed that common SNPs accounted for 45% of the loci with AEI (when sequences up to 250 kb upstream and downstream were included) [5]. The authors demonstrated the utility of their results for finding disease-associated variants using the example of a region associated with systemic lupus erythematosus (SLE). Ge *et al.* [5] further compared the *cis*-eQTL loci detected using AE analysis with eQTLs obtained from mRNA expression arrays, and found a partial overlap. Differences between these two approaches are attributable to strong *trans*-acting factors (which can mask weaker *cis* effects), epigenetic events, and limitations of the AE analysis at individual SNPs (see below).

The authors [5] concluded that *cis*-acting regulatory variants are frequent and could be used to clarify the

AE, allelic expression; AEI, allelic expression imbalance; eQTL, expression quantitative trait locus; rSNP, regulatory SNP; srSNP, structural RNA SNP.

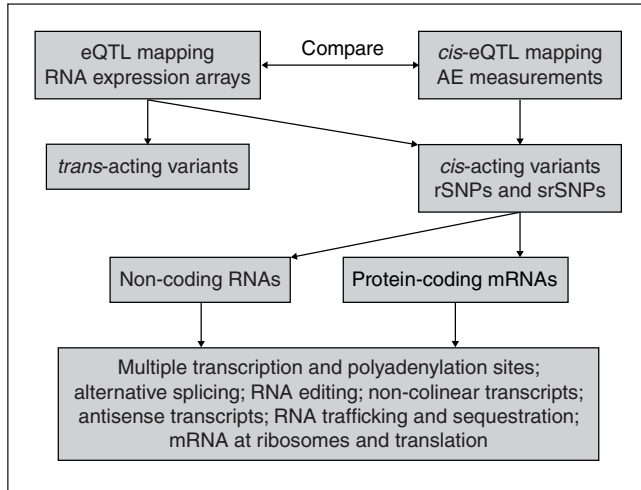


Figure 1

Schematic representation of the detection of *cis*- and *trans*-regulatory variants and the type of polymorphisms involved in gene expression. eQTL mapping and expression arrays give information about *cis*- and *trans*-acting variants, and this can be compared with information from *cis*-eQTL mapping and AE measurements to determine which variants are *cis*-acting. These variants come in various forms, as shown at the bottom. To simplify, 'SNP' is taken here as representing all sequence variations; rSNPs affect transcription, and srSNP (structural RNA SNPs) affect RNA processing and translation.

genetic risk of complex disorders. To evaluate the potential of 'expression genetics', we must account for the complexity of transcription, mRNA processing, and translation; and we must ask what we can learn from AE assays at individual SNPs and what the limitations of this approach are.

Regulatory variants and the complexity of RNA transcripts

An allelic RNA expression imbalance measured at an individual SNP indicates the presence of a *cis*-regulatory process [14]. Epigenetic effects can account for AEI, for example through imprinting or the random mono-allelic silencing that is observed for numerous genes in lymphoblastic cells [15], which are often highly clonal [16]; however, Ge *et al.* [5] suggest that epigenetic silencing occurs less frequently than previously thought in transformed B lymphocytes. Moreover, this phenomenon may be less prevalent in other (non-transformed) tissues [13]. Rather, AEI seems to arise mainly from *cis*-regulatory variants. However, the AE ratio measurements provide only a crude picture of a highly dynamic process from transcription to translation [14]. First, many genes have multiple transcription initiation sites, so that SNPs in the transcripts typically represent multiple species of RNA, each subject to distinct regulation. Second, docking sites for proteins and RNAs (such as microRNAs) can be affected,

leading to altered (m)RNA processing, splicing, editing, polyadenylation, cellular trafficking, and the formation of non-colinear transcripts [17] or antisense RNAs [18]. Given that alternative splicing is a near universal phenomenon in human genes [19], AE analysis without separating the main RNA species at any given locus cannot provide a clear answer. Ge *et al.* [5] have addressed alternative splicing by analyzing windows of multiple SNPs across a gene locus, offering a broad, if incomplete, glimpse of alternative splicing genetics. However, this approach fails if a splice variant has similar turnover but distinct functions, or the spliced exon does not carry a polymorphism. AE analysis must be performed specifically for each splice variant, as demonstrated for the short and long mRNA isoforms of dopamine receptor D2 [20]. Two intronic SNPs were found to alter splicing and brain activity *in vivo* during cognitive processing in humans [20].

SNPs residing in transcribed RNAs have extensive potential to affect function, because the RNA transcript consists of a single-stranded nucleic acid, which folds onto itself to yield an assembly of structures that determine the RNA's biology. Over 90% of all SNPs alter RNA folding - a fact exploited in single-stranded conformational polymorphism (SSCP) SNP analysis - and thus have the potential to affect function [14]. We have named polymorphisms occurring in the RNA transcript 'structural RNA SNPs' (srSNPs) (Figure 1); this type of variant might be at least as prevalent as rSNPs [13]. Furthermore, synonymous SNPs located in protein-coding regions have been neglected as carriers of functional information; however, they can alter mRNA turnover, splicing, translation, and are particularly adapted towards RNA folding structures that may have a role in evolution [21]. Increasing knowledge of transcript complexity has led to reassessment of the role of RNA variation in evolution and disease etiology.

Tissue selectivity of *cis*-regulatory variants

Ge *et al.* [5] found considerable overlap in AEI between lymphoblasts and a few tested primary cell lines of mesenchymal origin, whereas Dimas *et al.* [22] found from testing various blood cell types that 69 to 80% of *cis*-regulatory variants operate in a cell-type-specific manner. Tissue-specific enhancers determine selective expression for most genes [23] and, moreover, a large proportion of the machinery regulating transcription, mRNA processing, and translation differs from one tissue to the next. For example, a promoter SNP in *VKORC1* (encoding vitamin K epoxide reductase complex subunit 1, the target of warfarin) affects expression only in the liver but not in the heart or lymphocytes [24]. Studying the *TPH2* gene (encoding tryptophan hydroxylase 2, which is involved in serotonin biosynthesis) requires pontine tissues, in which the gene is actively transcribed before the protein is distributed throughout the brain [25]. Therefore, AE analysis must focus on relevant target tissues, whereas

blood lymphocytes can serve as a surrogate only for a limited subset of genes.

The role of regulatory variants in evolution

Regulation of gene expression is now considered a primary driver of evolution [2-4]. The potential to alter gene expression only in specific target tissues imposes less constraint for developing new selectable traits. We must assume that positive selection to allele frequencies beyond those expected in a neutral model implies strong phenotypic penetrance associated with fitness, either of the individual or, more controversially, a group of individuals. When applied to humans, the concept of selection on a group includes cultural influence on human evolution and may involve 'balanced evolution', that is, the accumulation of high- and low-activity variants for key genes. Because such regulatory variants are linked to fitness rather than disease, it is not surprising that genome-wide association studies have failed to detect them. However, fitness genes can be a two-edged sword: for example, the activity of a gene product may be optimal for long life but not reproductive success. Similarly, fitness genes could conceivably contribute to disease risk if several interrelated genes have variants that cause a change in the same direction in any given individual. A disease association would become apparent only if interactions between several genes are considered. Knowing the functional variants is essential to tackle these complex interactions.

The way forward: how do we identify regulatory variants germane to fitness and disease

The results of Ge *et al.* [5] significantly advance our understanding of *cis*-regulatory factors, and their possible role in heritability of complex disorders. We can now propose steps that are required to shed light on this hidden area.

First, AE should be measured for each transcript isoform, rather than at single marker SNPs that represent the mean of all isoform transcripts. Next generation sequencing has the potential to provide this level of detail [9,10]. Second, equal attention must be given to rSNPs and srSNPs; the latter affect mRNA processing and translation. Moreover, noncoding RNAs should be considered, as many hits from genome-wide association studies are in intergenic regions.

Because of the tissue selectivity of gene expression, the third step is that AE must be determined in relevant target tissues. Numerous tissue banks are available that provide human autopsy tissues from diseased subjects and controls that are suitable for AE analysis. Also, SNP scanning and subsequent molecular genetics studies are needed to identify the polymorphisms responsible for AEI. Knowing the main functional variants for a candidate gene greatly facilitates subsequent clinical association studies with accessible DNA samples. Furthermore, we should focus on genes that show positive selection in the human lineage,

which indicates phenotypic penetrance. If multiple genes in a given pathway have frequent regulatory variants, appropriate multifactorial models should be tested for combined effects on fitness and disease.

Finally, drug targets presumably reside at critical intersections of protein networks, thereby altering the disease process. These targets should be revisited in order to check whether *cis*-regulatory factors have been overlooked. Polymorphisms in drug target genes often have a large effect on disease risk or treatment outcomes, which are the focus of pharmacogenomic studies.

Given the rapid advances in genomic technologies, these goals are achievable and promise breakthroughs in resolving complex disease risks, prevention strategies, and therapy outcomes.

Competing interests

The author declares that he has no competing interests.

References

1. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorf LA, Hunter DJ, McCarthy MI, Ramos EM, Cardon LR, Chakravarti A, Cho JH, Guttmacher AE, Kong A, Kruglyak L, Mardis E, Rotimi CN, Slatkin M, Valle D, Whittemore AS, Boehnke M, Clark AG, Eichler EE, Gibson G, Haines JL, Mackay TF, McCarrroll SA, Visscher PM: **Finding the missing heritability of complex diseases.** *Nature* 2009, **461**:747-753.
2. Britton RJ, Davidson EH: **Gene regulation for higher cells: a theory.** *Science* 1969, **165**:349.
3. Hawks J, Wang ET, Cochran GM, Harpending HC, Moyzis RK: **Recent acceleration of human adaptive evolution.** *Proc Natl Acad Sci USA* 2007, **104**:20753-20758.
4. Wray GA: **The evolutionary significance of cis-regulatory mutations.** *Nat Rev Genet* 2007, **8**:206-216.
5. Ge B, Pokholok DK, Kwan T, Grundberg E, Morcos L, Verlaan DJ, Le J, Koka V, Lam KC, Gagné V, Dias J, Hoberman R, Montpetit A, Joly MM, Harvey EJ, Sinnett D, Beaulieu P, Hamon R, Graziani A, Dewar K, Harmsen E, Majewski J, Göring HH, Naumova AK, Blanchette M, Gunderson KL, Pastinen T: **Global patterns of cis variation in human cells revealed by high-density allelic expression analysis.** *Nat Genet* 2009, **41**:1216-1222.
6. Stranger BE, Nica AC, Forrest MS, Dimas A, Bird CP, Beazley C, Ingle CE, Dunning M, Flicek P, Koller D, Montgomery S, Tavaré S, Deloukas P, Dermitzakis ET: **Population genomics of human gene expression.** *Nat Genet* 2007, **39**:1217-1224.
7. Spielman RS, Bastone LA, Burdick JT, Morley M, Ewens WJ, Cheung VG: **Common genetic variants account for differences in gene expression among ethnic groups.** *Nat Genet* 2007, **39**:226-231.
8. Göring HH, Curran JE, Johnson MP, Dyer TD, Charlesworth J, Cole SA, Jowett JB, Abraham LJ, Rainwater DL, Comuzzie AG, Mahaney MC, Almasy L, MacCluer JW, Kissebah AH, Collier GR, Moses EK, Blangero J: **Discovery of expression QTLs using large-scale transcriptional profiling in human lymphocytes.** *Nat Genet* 2007, **39**:1208-1216.
9. Zhang K, Li JB, Gao Y, Egli D, Xie B, Deng J, Li Z, Lee JH, Aach J, Leproust EM, Eggan K, Church GM: **Digital RNA allelotyping reveals tissue-specific and allele-specific gene expression in human.** *Nat Methods* 2009, **6**:613-618.
10. Heap GA, Yang JH, Downes K, Healy BC, Hunt KA, Bockett N, Franke L, Dubois PC, Mein CA, Dobson RJ, Albert TJ, Rodesch MJ, Clayton DG, Todd JA, van Heel DA, Plagnol V: **Genome-wide analysis of allelic expression imbalance in human**

- primary cells by high throughput transcriptome resequencing.** *Hum Mol Gen* 2009, doi:10.1093/hmg/ddp473.
11. Campino S, Forton J, Raj S, Mohr B, Auburn S, Fry A, Mangano VD, Vandiedonck C, Richardson A, Rockett K, Clark TG, Kwiatkowski DP: **Validating discovered cis-acting regulatory genetic variants: application of an allele specific expression approach to HapMap populations.** *PLoS One* 2008, **3**:e4105.
 12. Serre D, Gurd S, Ge B, Sladek R, Sinnett D, Harmsen E, Bibikova M, Chudin E, Barker DL, Dickinson T, Fan JB, Hudson TJ: **Differential allelic expression in the human genome: a robust approach to identify genetic and epigenetic cis-acting mechanisms regulating gene expression.** *PLoS Genet* 2008, **4**:e1000006.
 13. Johnson AD, Zhang Y, Papp AC, Pinsonneault JK, Lim JE, Saffen D, Dai Z, Wang D, Sadee W: **Polymorphisms affecting gene transcription and mRNA processing in pharmacogenetic candidate genes: detection through allelic expression imbalance in human target tissues.** *Pharmacogenet Genomics* 2008, **18**:781-791.
 14. Johnson AD, Wang D, Sadee W: **Polymorphisms affecting gene regulation and mRNA processing: broad implications for pharmacogenetics.** *Pharmacol Ther* 2005, **106**:19-38.
 15. Gimelbrant A, Hutchinson JN, Thompson BR, Chess A: **Wide-spread monoallelic expression on human autosomes.** *Science* 2007, **318**:1136-1140.
 16. Plagnol V, Uz E, Wallace C, Stevens H, Clayton D, Ozcelik T, Todd JA: **Extreme clonality in lymphoblastoid cell lines with implications for allele specific expression analyses.** *PLoS One* 2008, **3**:e2966.
 17. Gingeras TR: **Implications of non-co-linear transcripts.** *Nature* 2009, **461**:206-211.
 18. He Y, Vogelstein B, Velculescu VE, Papadopoulos N, Kinzler KW: **The antisense transcriptomes of human cells.** *Science* 2008, **322**:1855-1857.
 19. Wang ET, Sandberg R, Luo S, Khrebtkova I, Zhang L, Mayr C, Kingsmore SF, Schroth GP, Burge CB: **Alternative isoform regulation in human tissue transcriptomes.** *Nature* 2008, **456**:470-476.
 20. Zhang Y, Bertolino A, Fazio L, Blasi G, Rampino A, Romano R, Lee ML, Xiao T, Papp A, Wang D, Sadee W: **Polymorphisms in human dopamine D2 receptor gene affect gene expression, splicing, and neuronal activity during working memory.** *Proc Natl Acad Sci USA* 2007, **104**:20552-20557.
 21. Biro JC: **Correlation between nucleotide composition and folding energy of coding sequences with special attention to wobble bases.** *Theor Biol Med Model* 2008, **5**:14.
 22. Dimas AS, Deutsch S, Stranger BE, Montgomery SB, Borel C, Attar-Cohen H, Ingle C, Beazley C, Gutierrez Arcelus M, Sekowska M, Gagnebin M, Nisbett J, Deloukas P, Dermitzakis ET, Antonarakis SE: **Common regulatory variation impacts gene expression in a cell type-dependent manner.** *Science* 2009, **325**:1246-1250.
 23. Heintzman ND, Hon GC, Hawkins RD, Kheradpour P, Stark A, Harp LF, Ye Z, Lee LK, Stuart RK, Ching CW, Ching KA, Antosiewicz-Bourget JE, Liu H, Zhang X, Green RD, Lobanenkov VV, Stewart R, Thomson JA, Crawford GE, Kellis M, Ren B: **Histone modifications at human enhancers reflect global cell-type-specific gene expression.** *Nature* 2009, **459**:108-112.
 24. Wang D, Chen H, Momary KM, Cavallari LH, Johnson JA, Sadee W: **Regulatory polymorphism in vitamin K epoxide reductase complex subunit 1 (VKORC1) affects gene expression and warfarin dose requirement.** *Blood* 2008, **112**:1013-1021.
 25. Lim JE, Pinsonneault J, Sadee W, Saffen D: **Tryptophan hydroxylase 2 (TPH2) haplotypes predict levels of TPH2 mRNA expression in human pons.** *Mol Psychiatry* 2007, **12**:491-501.
-
- Published: 22 November 2009
doi:10.1186/gm116
© 2009 BioMed Central Ltd