Single-Tissue and Cross-Tissue Heritability of Gene Expression Via Identity-by-Descent in Related or Unrelated Individuals

Alkes L. Price1,2,3, Agnar Helgason4,5, Gudmar Thorleifsson4, Steven A. McCarroll3,6, Augustine Kong4, Kari Stefansson4,5

1 Department of Epidemiology, Harvard School of Public Health, Boston, Massachusetts, United States of America, 2 Department of Biostatistics, Harvard School of Public Health, Boston, Massachusetts, United States of America, 3 Broad Institute of Massachusetts Institute of Technology and Harvard, Cambridge, Massachusetts, United States of America, 4 deCODE Genetics, Reykjavik, Iceland, 5 University of Iceland, Reykjavik, Iceland, 6 Department of Genetics, Harvard Medical School, Boston, Massachusetts, United States of America

Abstract

Family studies of individual tissues have shown that gene expression traits are genetically heritable. Here, we investigate cis and trans components of heritability both within and across tissues by applying variance-components methods to 722 Icelanders from family cohorts, using identity-by-descent (IBD) estimates from long-range phased genome-wide SNP data and gene expression measurements for ~19,000 genes in blood and adipose tissue. We estimate the proportion of gene expression heritability attributable to cis regulation as 37% in blood and 24% in adipose tissue. Our results indicate that the correlation in gene expression measurements across these tissues is primarily due to heritability at cis loci, whereas there is little sharing of trans regulation across tissues. One implication of this finding is that heritability in tissues composed of heterogeneous cell types is expected to be more dominated by cis regulation than in tissues composed of more homogeneous cell types, consistent with our blood versus adipose results as well as results of previous studies in lymphoblastoid cell lines. Finally, we obtained similar estimates of the cis components of heritability using IBD between unrelated individuals, indicating that transgenerational epigenetic inheritance does not contribute substantially to the “missing heritability” of gene expression in these tissue types.

Introduction

The genome contains a complex set of instructions for the assembly and maintenance of an organism. A fundamental goal in biology is to understand the relationship between genotype and phenotype. This goal can be achieved in part by studying the genetic basis of gene expression, as many genotype-phenotype correlations are a consequence of genetically driven variation in gene expression [1]. A number of studies have mapped individual cis and trans regulatory variants in humans, and recent work has suggested that the majority of regulators act in trans [2-5]; regulation of gene expression has also been widely studied in animal models [6-9]. However, the bulk of variability in gene expression remains unexplained. Heritability analyses can shed light on the genetic basis of gene expression. Several previous studies have demonstrated substantial overall heritability of gene expression in family data sets, and heritability approaches have also been broadly applied to other phenotypes [10-14].

In this study, we used gene expression measurements [11] and genome-wide single nucleotide polymorphism (SNP) data [13] from 722 Icelanders from family cohorts to examine the heritability of gene expression in blood and adipose tissue. By studying more than one tissue type, we were able to analyze the regulation of gene expression both within and across tissues. Our goal was to answer three key questions about gene expression heritability. First, can heritability be partitioned into cis and trans components using local and genome-wide IBD between pairs of individuals? Second, to what extent are heritable components of variance shared across tissues? Third, to what extent does heritability extend to distantly related individuals inheriting IBD segments from distant ancestors?

We sought to partition the heritability of gene expression into cis versus trans components by comparing the effects of IBD at the genome-wide level (trans) to those of IBD at the local level (cis), defined as the number of chromosomes (0, 1 or 2) shared IBD at the genomic location containing the expressed gene. Our results show a substantially higher proportion of heritability due to cis regulation, 37% in blood and 24% in adipose tissue, than the 12% reported in a previous ancestry-based study of lymphoblastoid cell lines (LCL) in African Americans [16]. One possible explanation for this discrepancy is transgenerational epigenetic inheritance, which is one of the explanations proposed to account for the “missing heritability” in genetic studies of human traits [17-23]. Epigenetic inheritance would regulate gene expression at the cis locus, and would be expected to contribute to cis heritability in...
Author Summary

An important goal in biology is to understand how genotype affects gene expression. Because gene expression varies across tissues, the relationship between genotype and gene expression may be tissue-specific. In this study, we used heritability approaches to study the regulation of gene expression in two tissue types, blood and adipose tissue, as well as the regulation of gene expression that is shared across these tissues. Heritability can be partitioned into cis and trans effects by assessing identity-by-descent (IBD) at the genomic location close to the expressed gene or genome-wide, respectively, and applying variance-components methods to partition the heritability of each gene. We estimated the proportion of gene expression heritability explained by cis regulation as 37% in blood and 24% in adipose tissue. Notably, the heritability shared across tissue types was primarily due to cis regulation. Thus, the relative contribution of cis versus trans regulation is expected to increase with the number of cell types present in the tissue being assayed, just as observed in our study and in a comparison to previous work on lymphoblastoid cell lines (LCL). We specifically ruled out a substantial contribution of transgenerational epigenetic inheritance to heritability of gene expression in these cohorts by repeating our heritability analyses using segments shared IBD in distantly related Icelanders.

Methods

Ethics statement

This research was approved by the Data Protection Commission of Iceland and the National Bioethics Committee of Iceland. The appropriate informed consent was obtained for all sample donors.

Icelandic Family Blood cohort

Relative abundances of 23,720 transcripts were obtained for blood samples from each of 1,001 individuals from the IFB cohort, as described previously [11] (see Web Resources). Values were adjusted for sex and age. We removed 4,985 transcripts that either had >5% missing data, did not map to an autosomal chromosome, or mapped to more than one genomic location. We removed 16 individuals with >5% missing data and 269 individuals for which long-range phased SNP data were not available. This left 10,735 transcripts and 716 individuals. Most of our analyses focused on 1,700 related pairs (individuals from the same family pedigree with genome-wide IBD >0.05) spanning a subset of 531 individuals.

Icelandic Family Adipose cohort

Relative abundances of 23,720 transcripts were obtained for adipose tissue samples from each of 673 individuals from the IFA cohort, as described previously [11] (see Web Resources). Values were adjusted for sex, age and body mass index (BMI), restricting to 638 individuals with BMI data. We removed 4,621 transcripts that either had >5% missing data, did not map to an autosomal chromosome, or mapped to more than one genomic location. We removed 2 individuals with >3% missing data and 67 individuals for which long-range phased SNP data were not available. This left 19,099 transcripts and 569 individuals. Most of our analyses focused on 1,700 related pairs (individuals from the same family pedigree with genome-wide IBD >0.05) spanning a subset of 531 individuals.

Local IBD estimates

Individuals were genotyped using the Illumina 300K chip. Owing to the sensitive nature of genotype data, access to these data can only be granted at the headquarters of deCODE Genetics in Iceland. Given long-range phased Illumina 300K data [15] for a pair of individuals, we partitioned the genome into 2cM blocks and for each block performed 2 × 2 = 4 comparisons between haplotypes from the two individuals. We declared two haplotypes to be IBD if they matched at >95% of alleles in the block, non-IBD if they matched at <85% of alleles, and unknown-IBD otherwise. We excluded SNPs with missing data in one or both individuals, so that lack of a match implies a mismatch, and set IBD status to unknown for pairs of haplotypes with >5% of SNPs excluded. We defined local IBD as the total number of comparisons producing a match. We verified that this approach infers 0:1:2 copies IBD between parent-child pairs with probabilities 0.2%:99.3%:0.4% and 0:1:2: copies IBD between sibling pairs with probabilities 24.9%:50.1%:24.9%, excluding from this computation the 7% of pairs and blocks for which inferred IBD was unknown. These numbers are a function of the thresholds we used to define IBD and non-IBD; the thresholds were largely chosen for specificity rather than sensitivity since for our application it does not matter that inferred IBD is sometimes unknown. The numbers are very close to the expected theoretical probabilities for parent-child pairs, 2 copies IBD is expected to occasionally occur due to IBD in “unrelated” parents. This validates our use of long-range phased SNP genotypes to compute local IBD estimates. We computed genome-wide IBD estimates as the average of local IBD estimates across all 2cM blocks.

Heritability estimates using genome-wide IBD only

We applied variance-components methods to estimate narrow-sense heritability [14,24]. The source code used in our heritability analyses is available for download (see Web Resources). Let $\epsilon_{gs}$ denote the gene expression for gene $g$ and individual $s$, normalized to have mean 0 and variance 1 across individuals. Let $\theta_{gs}$ denote the genome-wide IBD between individuals $s$ and $t$ ($0 \leq \theta_{gs} \leq 1$) and $\Theta = (\theta_{gs})$ be the $N \times N$ matrix of genome-wide IBD, where $N$ is the number of individuals. Let $V_g$ denote the covariance matrix of normalized gene expression for gene $g$. We consider the model $V_g = h^2 g \Theta + (1 - h^2 g) I$ and fit $h^2_g$, the heritability of gene $g$, to the observed normalized gene expression values $\epsilon_{gs}$ by maximizing the likelihood $L(\epsilon_{gs}|V_g) \propto \frac{1}{\sqrt{\det(V_g)}} \exp \left( -\frac{1}{2} \epsilon_{gs}^T V_g^{-1} \epsilon_{gs} \right)$, where $\epsilon_{gs} = (\epsilon_{gs})$. Values of $\epsilon_{gs}$, $h^2_g$, and $V_g$ vary with tissue type, but we view tissue type as an implicit index rather than an explicit index for simplicity of notation. For both blood and adipose tissue, the estimated values of $h^2_g$ were ~80% correlated to values that were computed previously using similar methods [11], despite the fact that the current analysis was restricted to a subset of individuals for which long-range phased SNP data was available for local IBD.
Heritability estimates using both cis and trans IBD

We extended the variance-components approach to cis and trans heritability via the model $Y_i = h_{g,cis}r_i^2 + h_{g,trans}r_i^2 + \varepsilon_i$, where $r_i$ is the $N \times N$ matrix of local (cis) IBD between individuals $i$ and $j$ at the genomic location proximal to gene $g$. We used the midpoint of the gene expression probe to define genomic location, but the value of $r_i$ is not sensitive to this choice as local IBD segments between related individuals span many megabases. We scale $r_i$ to have value 0.0, 0.5, or 1.0 (for 0, 1, or 2 copies shared). We fit the cis heritability $h_{g,cis}^2$ and trans heritability $h_{g,trans}^2$ by maximizing the usual likelihood, fitting $h_{g,cis}^2 = h_{g,trans}^2 + h_{g,cis}^2$ and $h_{g,trans}^2$ in turn. We average across genes $g$ to estimate $h_{g,cis}^2$ and $h_{g,trans}^2$. We define the proportion of heritable gene expression variation that is due to cis regulation as $\pi_c = h_{g,cis}^2/(h_{g,cis}^2 + h_{g,trans}^2)$. As above, all values vary with tissue type, which we view as an implicit index.

Cross-tissue analysis

The cross-tissue correlation $\rho$ was computed as the correlation between normalized expression levels in blood and adipose tissue across genes and individuals. Due to the normalization, this is equal to the average of gene-specific correlations $\rho_{ij}$. We computed standard errors of both gene-specific and average cross-tissue correlations via jackknife, repeating the computation with each individual removed in turn and estimating the standard error as $\sqrt{N}$ times the standard deviation of the $\rho$ estimates. We now describe our estimation of cross-tissue heritability. Let $e_{bg} = (e_{bgj})$ and $e_{ag} = (e_{ajg})$ denote normalized expression levels for gene $g$ and individual $j$ in blood and adipose tissue, respectively. Let $W_{g}$ denote the covariance matrix of the vector $(e_{bgj},e_{ajg})$ of length $2N$. Here are the relevant equations.

\[
X_{g,cis} = \left( \begin{array}{c} h_{g,cis}^2 \varepsilon_{g,cis}^2 + h_{ag,cis}^2 \\ \varepsilon_{g,cis}^2 - h_{ag,cis}^2 \end{array} \right), \quad
X_{g,trans} = \left( \begin{array}{c} h_{g,trans}^2 \varepsilon_{g,trans}^2 + h_{ag,trans}^2 \\ \varepsilon_{g,trans}^2 - h_{ag,trans}^2 \end{array} \right),
\]

\[
X_{g,env} = \left( \begin{array}{c} 1 - h_{g,cis}^2 - h_{g,trans}^2 \\ \rho_{g} - \varepsilon_{g,cis}^2 - \varepsilon_{g,trans}^2 \\ \rho_{g} - \varepsilon_{g,cis}^2 - \varepsilon_{g,trans}^2 \\ 1 - h_{g,cis}^2 - h_{g,trans}^2 \end{array} \right),
\]

\[
W_{g} = X_{g,cis} \otimes \Gamma_g + X_{g,trans} \otimes \Theta + X_{g,env} \otimes I,
\]

where $\varepsilon^2$ denotes cross-tissue heritability, $\rho$ denotes cross-tissue correlation, and $\otimes$ denotes the tensor product of a $2 \times 2$ matrix with an $N \times N$ matrix to form a $2N \times 2N$ matrix. For example, the first term of $W_g$ has entries $h_{g,cis}^2/\varepsilon_{g,cis}^2$ in the upper left $N \times N$ block, $\varepsilon_{g,cis}^2/\varepsilon_{g,trans}^2$ in the upper right $N \times N$ block, and so on. This generalization of the variance-components approach to cross-phenotype analyses has been previously described (for the case of genome-wide IBD) in an analysis of two height phenotypes, self-reported height and clinically measured height [25]. The likelihood is defined in the usual way, replacing $V_g$ with $W_g$ and $e_j$ with $(e_{bgj},e_{ajg})$. We fit $h_{g,cis}^2 = h_{g,cis}^2 + h_{g,trans}^2$, $h_{g,cis}^2 = h_{g,cis}^2 + h_{g,trans}^2$, $h_{g,trans}^2 = h_{g,trans}^2 + h_{g,trans}^2$, $h_{g,trans}^2$, $\rho_{g} - \varepsilon_{g,cis}^2 - \varepsilon_{g,trans}^2 = \varepsilon_{g,cis}^2 - \varepsilon_{g,trans}^2$, and $\varepsilon_{g,cis}^2$ in turn. For each of the parameters estimated, we compute average values by averaging across genes $g$.

Web Resources

- http://www.ncbi.nlm.nih.gov/geo/ (Gene Expression Omnibus). Gene expression data sets have been deposited into the GEO database under accession numbers GSE7965 and GPL3991, as described previously [11].

- http://www.hpsb.harvard.edu/faculty/alkes-price/software/.

The source code used in our heritability analyses is available for download, along with the results presented in Table S1.

Results

Overall heritability of gene expression

For the analysis of gene expression in blood, we analyzed normalized intensity values for 18,735 mRNA transcripts. Analysis was restricted to 667 individuals from the IFB cohort for whom long-range phased SNP data were available (see Methods). For each pair of individuals, we used the long-range phased SNP data to compute the number of chromosomes shared IBD at each location in the genome, and computed the genome-wide IBD as an average of these values (Figure 1; see Methods). Our initial analyses focused on 2,233 related pairs with genome-wide IBD $>0.05$. For the analysis of gene expression in adipose tissue, we similarly analyzed 19,099 mRNA transcripts of 531 individuals from the IFA cohort, focusing on 1,700 related pairs with genome-wide IBD $>0.05$ (see Methods). The IFA cohort largely overlaps the IFB cohort, with 496 of the 722 individuals analyzed appearing in both cohorts.

We estimated the overall heritability $h^2$ for each gene $g$ using variance-component methods [14] (see Methods). Although estimates for each gene $g$ are statistically noisy at these sample sizes, histograms show a clear positive bias for both IFB and IFA cohorts (Figure S1 and Table S1), and $h^2>0$ was nominally significant ($P = 0.05$; see Methods) for an excess of genes: 42% for IFB and 63% for IFA. We computed the average $h^2$ as the average of $h^2$ across genes $g$. A relevant question is whether or not to allow negative values of $h^2$ when computing this average [26]. Such values have no biological interpretation (except in the case of negative correlation among siblings in traits that depend on birth order). However, because values close to zero may be either increased or decreased by statistical noise—leading to negative estimates of $h^2$ for 3,031 of 18,735 genes for IFB and 1,038 of 19,099 genes for IFA—we elected to allow negative values in our main computations so as to produce an unbiased estimate of average $h^2$. We obtained estimates of $h^2 = 0.130$ for blood and $h^2 = 0.234$ for adipose tissue. We obtained similar results when using a regression-based approach to estimate average $h^2$ (Text S1), which more readily lends itself to visualization (Figure 2A and 2B). (When clipping negative $h^2$ values to zero, we obtained $h^2 = 0.159$ for blood and $h^2 = 0.237$ for adipose tissue.) Our results are consistent with previous analyses reporting that expression levels of a substantial fraction of genes are significantly heritable at the level of $h^2 = 0.3$ or higher [10-15,26].

Cis versus trans heritability of gene expression

While estimates of overall heritability are based on genome-wide IBD, it is possible to estimate cis versus trans heritability by extending variance components to consider both local (cis) IBD at the genomic
location close to the expressed gene, and genome-wide (trans) IBD (see Methods). As before, analyses were restricted to 2,233 and 1,700 related pairs from the IFB and IFA cohorts, respectively. Histograms of \( h_{\text{cis}}^2 \) and \( h_{\text{trans}}^2 \) estimates for each gene \( g \) show a clear positive bias for both IFB and IFA cohorts (Figure S2 and Table S1), with an excess of nominally significant (\( P < 0.05 \)) genes for IFB (\( h_{\text{cis}}^2 > 0: 16\%; h_{\text{trans}}^2 > 0: 19\% \)) and IFA (\( h_{\text{cis}}^2 > 0: 16\%; h_{\text{trans}}^2 > 0: 30\% \)). For IFB, we obtained average \( h_{\text{cis}}^2 = 0.055 \) and \( h_{\text{trans}}^2 = 0.095 \), which sum to \( h^2 = 0.150 \). This leads to the conclusion that the proportion of heritability of expression due to \( \text{cis} \) variants in blood is \( \pi_{\text{cis}} = 37\% \). For IFA, we obtained estimates of \( h_{\text{cis}}^2 = 0.057 \) and \( h_{\text{trans}}^2 = 0.177 \), which sum to \( h^2 = 0.234 \). This yields an estimate of \( \pi_{\text{cis}} = 24\% \) in adipose tissue. The values of \( h^2 \) and \( h_{\text{trans}}^2 \) in adipose tissue are significantly higher than for blood, but \( h_{\text{cis}}^2 \) is similar, leading to a lower value of \( \pi_{\text{cis}} \). We obtained similar results when using a regression-based approach to estimate average \( h_{\text{cis}}^2 \) and \( h_{\text{trans}}^2 \) (Text S1; Figure 2C and 2D). We note that there is considerably less statistical uncertainty in estimates of \( h_{\text{cis}}^2 \) (Figure 2C and 2D) than in estimates of \( h^2 \) (Figure 2A and 2B). Indeed, we obtained standard errors of \( h^2 = 0.150 \pm 0.011 \), \( h_{\text{cis}}^2 = 0.055 \pm 0.001 \), and \( h_{\text{trans}}^2 = 0.095 \pm 0.010 \) for blood and \( h^2 = 0.234 \pm 0.011 \), \( h_{\text{cis}}^2 = 0.057 \pm 0.002 \) and \( h_{\text{trans}}^2 = 0.177 \pm 0.010 \) for adipose tissue (see Methods). These standard errors are 7–100 times lower than standard errors for single-gene heritability estimates, which are inadequate for estimating \( \pi_{\text{cis}} \) (see Text S1). The much lower standard errors for \( h_{\text{cis}}^2 \) are a consequence of variation in \( \text{cis} \) IBD across the genome that decouples the estimation of this parameter from the systematic noise covariance structure across all pairs of individuals (see Text S1). Based on these standard errors for \( h_{\text{cis}}^2 \) and \( h_{\text{trans}}^2 \), \( \pi_{\text{cis}} \) has little statistical uncertainty, although results may be affected by modeling uncertainty.

Our heritability model does not account for the possibility of phenotypic similarity in related individuals due to shared environment, which can confound estimates of heritability [14]. We note that such effects would inflate estimates of \( h^2 \) and \( h_{\text{trans}}^2 \), but have a negligible impact on \( h_{\text{cis}}^2 \), since the extent of shared environment would be related to genome-wide (trans) rather than local (cis) IBD. To investigate the possibility of confounding due to shared environment, we computed the average correlation in gene expression between spouses, who are genetically unrelated but have a shared environment. We observed average correlations of \( 0.074 \pm 0.042 \) in 33 IFB spouse pairs and \( 0.076 \pm 0.033 \) in 28 IFA spouse pairs, which are similar in magnitude to correlations between sib-sib or parent-child pairs that correspond to the average heritabilities reported above (see Text S1 and Table S2). Thus, there is strong evidence that shared environment can lead to similarity in gene expression phenotypes. We further investigated whether the gene by gene signature of correlations in spouse pairs matches the signature of correlations in sib-sib or parent-child pairs or estimates of \( h_{\text{cis}}^2 \), but found that it does not (see Text S1 and Table S3). Thus, we hypothesize that the correlations in spouse
pairs are due to very recent shared environment (e.g. diet) arising from sharing the same household, whereas the correlations in sibling and parent-child pairs in this study (who are unlikely to share the same household, since only adult individuals were sampled) are due to genetic heritability. However, we cannot rule out a small amount of inflation in $h^2$ and $h_{trans}^2$ estimates due to shared environment in related individuals.

**Assessing the impact of epigenetic inheritance on cis heritability**

Our family-based estimates of $\pi_{cis}$ in blood and adipose tissue are considerably greater than a previous estimate of $12\pm3\%$ obtained using lymphoblastoid cell lines (LCL) from African-Americans, in which local versus genome-wide European ancestry was used to infer the relative contribution of cis versus trans heritability [16]. An analogous ancestry-based analysis of LCL gene expression data [27] from admixed HapMap 3 Mexican-Americans has produced a similarly low value of $\pi_{cis}=13\pm9\%$. One possible explanation for the lower values as compared to family-based estimates could be the epigenetic inheritance of cis-acting factors other than DNA sequence that are transmitted from parent to offspring. Given the relatively short time scale of epigenetic inheritance, this would be expected to have a much greater impact on family-based estimates of $\pi_{cis}$ than those based on ancestry [22-23].

To further explore the epigenetic hypothesis, we repeated the cis versus trans analysis using subsets of unrelated or distantly related individuals (genome-wide IBD <0.01) from the IFB and IFA cohorts. The mean genome-wide IBD for all such pairs of individuals was 0.0044, with a standard deviation of 0.0018, consistent with the known properties of distant relatedness between “unrelated” individuals from Iceland as well as other world populations [29-31]. We independently generated five random subsets of IFB individuals (85, 87, 92, 93, 91 individuals) and five random subsets of IFA individuals (127, 85, 92, 95, 89 individuals) with genome-wide IBD <0.01 between each pair of individuals in each subset, such that each subset was maximal subject to this constraint. The resulting estimates of $h_{cis}^2$ were $0.057\pm0.008$ for blood and $0.067\pm0.005$ for adipose tissue (mean $\pm$ standard deviation across five subsets). These estimates of $h_{cis}^2$ were close to our previous estimates based on closely related pairs, thereby ruling out a substantial contribution of epigenetic inheritance to cis heritability (see Discussion). However, we did not obtain meaningful estimates of $h_{trans}^2$ using distantly related individuals, due to the systematic noise covariance structure (see Text S1), and therefore $\pi_{cis}$ could not be estimated. We note that

---

**Figure 2. Family heritability in the IFB and IFA cohorts.** (A) Gene expression covariance (average value of product of normalized gene expression measurements) between related individuals in the IFB cohort varies with genome-wide IBD. Each point represents one pair of related individuals. The slope of this plot corresponds to the regression-based estimate of $h^2$. (B) Same as (A), for IFA cohort. (C) Gene expression covariance between siblings for genes with 0, 1 or 2 copies IBD at the cis locus, minus total covariance as displayed above. The slope of this plot corresponds to the regression-based estimate of $h_{cis}^2$. The signal to noise ratio is higher in this plot due to reduced effects of systematic noise covariance. (D) Same as (C), for IFA cohort.

doi:10.1371/journal.pgen.1001317.g002

---
similar results for distantly related pairs were obtained using different IBD estimation algorithms (see Text S1).

Cross-tissue analysis

We conducted a cross-tissue analysis of expression heritability in blood and adipose tissue in 496 individuals who overlapped between the IFB and IFA cohorts. We determined that an individual’s blood expression for a particular gene is slightly but significantly correlated to the same individual’s adipose expression for the same gene, with an average correlation of $\rho = 0.041 \pm 0.005$ (mean $\pm$ standard error) (see Methods). Although estimates for each gene $g$ are statistically noisy at these sample sizes, histograms show a clear positive bias in $\rho_g$ (Figure S3), and $\rho_g > 0$ was nominally significant ($P = 0.05$) for 20% of genes, a significant excess.

We next investigated the relationship between an individual’s blood expression and a related individual’s adipose expression, using variance-components methods (see Methods). This revealed that cross-tissue similarity varies with the level of family relatedness, with an average cross-tissue heritability estimate of $\hat{c}^2 = 0.030 \pm 0.006$. Analogous to the analyses for single tissues, we partitioned the cross-tissue heritability into cis and trans components, yielding values of $\hat{c}^2_{cis} = 0.031 \pm 0.001$ and $\hat{c}^2_{trans} = -0.001 \pm 0.006$. We obtained similar results using regression-based approaches (Text S1; Figure 3A and 3B). Histograms of cross-heritability estimates for each gene $g$ show a positive bias for $\hat{c}^2_{cis}$ and $\hat{c}^2_{trans}$, but not $\hat{c}^2_{trans,adip}^2$, for which the histogram is symmetric about zero (Figure S4). While our estimate of $\hat{c}^2_{trans}^2$ is not significantly different from zero, $\hat{c}^2_{cis}$ is highly significant and explains the bulk of our estimate of $\rho$. This implies that the extent to which gene expression in blood and adipose tissue is similar across genes and individuals is dominated by heritable effects at the cis locus.

Averaging across cell types with shared cis effects increases the value of $\pi_{cis}$

Our finding that cross-tissue similarities are dominated by heritable cis effects leads to the mathematical result that $\pi_{cis}$ is expected to increase with tissue heterogeneity: as the number of cell types represented in a tissue increases, the strongly correlated cis effects will add linearly but the uncorrelated trans effects will be diluted. In detail, let $x$ and $y$ denote cell types and suppose that $\text{Cov}(g_{x,y}) = \text{Cov}(g_{x,y}) = h_{cis,x}^2 + h_{cis,y}^2 \theta_{cis}$ for all genes $g$ and individuals $x \neq y$, and that all cis effects (but no trans or non-genetic effects) are shared across cell types. Thus, $	ext{Cov}(g_{x,y}) = h_{cis,x}^2 \theta_{cis}$.

Now consider a tissue $z$ containing cell types $x$ and $y$. Up to a normalization constant, $\text{Cov}(g_{x,y}) = \text{Cov}(g_{x,y}) = h_{cis,x}^2 \theta_{cis} + 0.5 h_{trans} \theta_{trans}$, so that $\pi_{cis,z} = h_{cis,x}^2 / (h_{cis,x}^2 + 0.5 h_{trans} \theta_{trans})$ is larger than $\pi_{cis,x} = \pi_{cis,y} = h_{cis,x}^2 / (h_{cis,x}^2 + h_{trans} \theta_{trans})$.

We verified this theoretical result empirically by defining $g_{x,y} = e_{cis} + e_{adip}$, as the average of normalized gene expression in blood and adipose tissue, normalized to mean 0 and variance 1. For synthetic tissue $z$, we obtained the value $\pi_{cis,x} = 0.41$, which is larger than the value of $\pi_{cis}$ for either blood or adipose tissue, and similar to the predicted value of $0.055/(0.055+0.250+0.095+0.177) = 0.45$ based on $h_{cis}$ and $h_{trans}$ ($\pi_{cis} < 0.45$ is actually expected since not all cis effects are shared). Thus, the variability in $\pi_{cis}$ across tissue types ($0.12$ for LCL, $0.24$ for adipose, $0.37$ for blood) is consistent with the fact that LCL represent a single cell type, whereas adipose tissue and blood contain many cell types: adipose tissue contains smooth muscle cells, fibroblasts, adipocytes, mast-cells and endothelial cells, while blood contains erythrocytes, thrombocyte, neutrophils, lymphocytes, monocytes, eosinophils and basophils in proportions that vary across individuals [32-34]. This also explains why studies of individual cell types have been more successful in identifying trans eQTLs than studies of whole tissues, and why most replications across tissue types occur at cis eQTLs [11,34-37].
between closely related individuals but not between distantly related individuals, given that this mode of inheritance persists over a relatively short time scale [22-23]. Our failure to observe any such discordance suggests that transgenerational epigenetic inheritance is unlikely to play a major role in the missing heritability of gene expression and other traits, although it does not rule out a very small aggregate effect across all genes or large effects at certain metatelic epialleles [40-41], nor does it shed light on the importance of mitotically conserved epigenetic effects that are not transmitted from parent to offspring.

Our results highlight the utility of using IBD in distantly related individuals to make inferences about heritability. This approach will be particularly valuable as sample sizes increase, since the number of pairs of individuals increases quadratically with sample size. Indeed, IBD in distantly related individuals has already proven useful for mapping specific loci [42], and heritability-related analyses using identity-by-state (IBS) instead of IBD have also yielded important insights [43-45]. By using IBD segments shorter than those analyzed here to consider IBD sharing at different distances from genes, it may even be possible to draw conclusions about the distribution of genomic distances at which cis regulation contributes to heritability.

Supporting Information

Figure S1  Histograms of heritability estimates for each gene. We plot histograms of (a) $\hat{h}_g^2$ estimates for IFB and (b) $\hat{h}_g^2$ estimates for IFA, across genes $g$. Found at: doi:10.1371/journal.pgen.1001317.s001 (0.23 MB TIF)

Figure S2  Histograms of cis and trans heritability estimates for each gene. We plot histograms of (a) $\hat{h}_{g,cis}^2$ estimates for IFB, (b) $\hat{h}_{g,trans}^2$ estimates for IFB, (c) $\hat{h}_{g,cis}^2$ estimates for IFA and (d) $\hat{h}_{g,trans}^2$ estimates for IFA, across genes $g$. Found at: doi:10.1371/journal.pgen.1001317.s002 (0.19 MB TIF)

Figure S3  Histograms of cross-tissue correlations for each gene. We plot a histogram of observed gene-specific cross-tissue correlations $\rho_{g,t}$.

Found at: doi:10.1371/journal.pgen.1001317.s003 (0.14 MB TIF)

Figure S4  Histograms of cross-tissue heritability estimates for each gene. We plot histograms of (a) $\hat{h}_{g,t}^2$ estimates, (b) $\hat{h}_{g,t}^2$ estimates, across genes.

Found at: doi:10.1371/journal.pgen.1001317.s004 (0.04 MB DOC)

Table S1  Heritability results for each gene.

Found at: doi:10.1371/journal.pgen.1001317.s005 (1.82 MB TXT)

Table S2  Average correlations between spouse-spouse, sib-sib, and parent-child pairs. We list the average correlation for each pair type and cohort, averaging across correlations for each gene $g$. We also list standard errors, computed via jackknife.

Found at: doi:10.1371/journal.pgen.1001317.s006 (0.03 MB DOC)

Table S3  Concordance of gene-by-gene signatures of correlations in each pair type. We list values of $R_{sib-sib,parent-child}$ and $R_{parent-parent}$ for each cohort (see text), along with the number of pairs of each type used to compute those values. For comparison purposes, we also list (in italics) values of $R_{sib-sib,parent-child}$ computed using smaller subsets of pairs to match the number of pairs used to compute $R_{parent-parent}$, as a smaller number of pairs leads to lower values of $R$.

Found at: doi:10.1371/journal.pgen.1001317.s007 (0.03 MB DOC)

Text S1  Supplementary Note.

Found at: doi:10.1371/journal.pgen.1001317.s008 (0.04 MB DOC)

Acknowledgments

We are grateful to David Reich and Adiï Hazra for valuable discussions. We thank Barbara Stranger and Manolis Dermitzakis for sharing LCL gene expression data from HapMap 3 samples; the International HapMap 3 Consortium for providing genotype data; and Benjamin Raby, Laiming Liang, and Noah Zaitlen for helpful comments. Finally, we thank the Icelanders who participated in this study.

Author Contributions

Conceived and designed the experiments: ALP AH GT SAM AK KS. Analyzed the data: ALP AH GT. Wrote the paper: ALP AH GT SAM AK KS.

References

1. Cookson W, Liang L, Abecasis G, Moffatt M, Lathrop M (2009) Mapping complex disease traits with global gene expression. Nat Rev Genet 10: 184–194.
2. Schadt EE, Monks SA, Drake TA, Lussi AJ, Che N, et al. (2003) Genetics of gene expression surveyed in maize, mouse and man. Nature 422: 297–302.
3. Mohamed M, Molony CM, Weber TM, Devlin JL, Ewens KG, et al. (2007) A genome-wide association study of global gene expression. PLoS Biol 5: e1000480. doi:10.1371/journal.pbio.1000480.
4. Gilad Y, Rifkin SA, Pritchard JK (2008) Revealing the architecture of gene regulation: the promise of eQTL studies. Trends Genet 24: 408–415.
5. Dixon AL, Liang L, Moffatt MF, Chen W, Heath S, et al. (2007) A genome-wide association study of global gene expression. Nature 430: 743–747.
6. Metspalu A, Metspalu M, Kivisild T (2008) The origin and prehistory of fully modern humans in Oceania. Nature 453: 549–552.
7. Williams RB, Cotsapas CJ, Cowley MJ, Chan E, Nott DJ, et al. (2006) Heritability of gene expression uncovers polygenic and pleiotropic networks that modulate nervous system function. Nat Genet 37: 233–242.
8. Pettretto E, Manjour JD, Dickens NJ, Cook SA, Kumaran MK, et al. (2006) Heritability and tissue specificity of expression quantitative trait loci. PLoS Genet 2: e172. doi:10.1371/journal.pgen.0020172.
9. Pettretto E, Bottolo L, Langlely SR, Hennig M, McDermott-Roe C, et al. (2010) New insights into the genetic control of gene expression using a Bayesian multivariate approach. PLoS Comput Biol 6: e1000737. doi:10.1371/journal.pcbi.1000737.
10. Monks SA, Leonardson A, Zhu H, Cundiff P, Pietrusik P, et al. (2004) Genetic inheritance of gene expression in human cell lines. Am J Hum Genet 75: 1094–1105.
11. Emilsson V, Thorleifsson G, Zhang B, Leonardson AS, Zink F, et al. (2008) Genetics of gene expression and its effect on disease. Nature 452: 423–428.
12. Stranger BE, Nica AC, Forrest MS, Dimas A, Bird CP, et al. (2007) Population genomics of human gene expression. Nat Genet 39: 1217–1224.
13. Dixon AL, Liang L, Moffatt MF, Chen W, Heath S, et al. (2007) A genome-wide association study of global gene expression. Nat Genet 39: 1202–1207.
14. Visscher PM, Hill WG, Wray NR (2008) Heritability in the genetics era: concepts and misconceptions. Nat Rev Genet 9: 253–266.
heritability of gene expression across tissues

Kong A, Masson G, Frigge ML, Glynnson A, Zuurmanovich P, et al. (2008) Detection of sharing by descent, long-range phasing and haplotype imputation. Nat Genet 40: 1068–1075.

Price AL, Patterson N, Plenge RM, Weinblatt ME, Shriver MD, et al. (2006) Principal components for population structure. Nat Genet 38: 968–974.

The International HapMap Consortium (2007) A second generation human haplotype map of 3.1 million SNPs. Nature 449: 851–860.

McRae AF, Matigian NA, Vadlamudi L, Mulley JC, Mowry B, et al. (2007) Replicated effects of sex and genotype on gene expression in human lymphoblastoid cell lines. Hum Mol Genet 16: 364–373.

Youngson NA, Whitelaw E (2008) Transgenerational epigenetic effects. Annu Rev Genomics Hum Genet 9: 233–257.

Jablonka E, Raz G (2009) Transgenerational epigenetic inheritance: prevalence, mechanisms, and implications for the study of heredity and evolution. Q Rev Biol 84: 131–176.

Amos CI (1994) Robust variance-components approach for assessing genetic linkage in pedigrees. Am J Hum Genet 54: 535–543.

Macgregor S, Cornes BK, Martin NG, Vischer PM (2006) Bias, precision and heritability of self-reported and clinically measured height in Australian twins. Hum Genet 120: 571–580.

McRae AF, Matigian NA, Yaslamodi L, Mulley JC, Mowry B, et al. (2007) Replicated effects of sex and genotype on gene expression in human lymphoblastoid cell lines. Hum Mol Genet 16: 364–373.

Stranger BE, Montgomery SB, Dimas A, Parts L, Ingle CE, et al. Patterns of cis regulatory variation in diverse human populations. Submitted.

The International HapMap 3 Consortium (2010) An integrated haplotype map of rare and common genetic variation in diverse human populations. Nature 467: 1028–1035.

Helgason A, Palsson S, Gudbjartsson DF,Kristjansson T, Stefansson K (2008) An association between the kinship and fertility of human couples. Science 319: 813–816.

The International HapMap Consortium (2007) A second generation human haplotype map of over 3.1 million SNPs. Nature 449: 851–860.

Gusev A, Low JK, Stoffel M, Daly MJ, Abshuler D, et al. (2009) Whole population, genome-wide mapping of hidden relatedness. Genome Res 19: 316–326.

Astor G, Vignati F, Bardelli S, Tubio M, Gola M, et al. (2007) "In vitro" and multicolor phenotypic characterization of cell subpopulations identified in fresh human adipose tissue stromal vascular fraction and in the derived mesenchymal stem cells. J Transl Med 5: 55.

Josephson B, Dahlberg G (1952) Variations in the cell-content and chemical composition of the human blood due to age, sex and season. Scand J Clin Lab Invest 4: 216–236.

Cheung VG, Spielman RS (2009) Genetics of human gene expression: mapping DNA variants that influence gene expression. Nat Rev Genet 10: 595–604.

Schadt EE, Molony C, Hudson E, Zhao H, Wang X, et al. (2008) Mapping the genetic architecture of gene expression in human liver. PLoS Biol 6: e107. doi:10.1371/journal.pbio.0060107.

Dinnas AS, Deutsch S, Stranger BE, Montgomery SB, Borel C, et al. (2009) Common regulatory variation impacts gene expression in a cell type-dependent manner. Science 325: 1246–1250.

Idaghdour Y, Ceika W, Shianma KV, Lee SH, Vischer PM, et al. (2010) Geographical genomics of human leukocyte gene expression variation in southern Morocco. Nat Genet 42: 62–67.

Choy E, Yelenosky R, Bonakdar S, Plenge RM, Saxena R, et al. (2008) Genetic analysis of human traits in vitro: drug response and gene expression in lymphoblastoid cell lines. PLoS Genet 4: e1000287. doi:10.1371/journal.pgen.1000287.

Shen-Orr SS, Tishkarian R, Khatri P, Bodian DL, Staedtler F, et al. (2010) Cell type-specific gene expression differences in complex tissues. Nat Methods 7: 287–289.

Morgan HD, Sutherland HG, Martin DI, Whitelaw E (1999) Epigenetic inheritance at the agouti locus in the mouse. Nat Genet 23: 314–318.

Rahyan VK, Chong S, Champ ME, Guthbert PC, Morgan HD, et al. (2003) Transgenerational inheritance of epigenetic states at the murine Axin/Fu allele occurs after maternal and paternal transmission. Proc Natl Acad Sci U S A 100: 2538–2543.

Kenny EE, Gusev A, Krieg K, Lutjohann D, Lowe JK, et al. (2009) Systematic haplotype analysis resolves a complex plasma plant sterol locus on the Micronesian Island of Kosrae. Proc Natl Acad Sci U S A 106: 13896–13891.

Kang HM, Sul JH, Service SK, Zaitlen NA, Kong SY, et al. (2010) Variance component model to account for sample structure in genome-wide association studies. Nat Genet 42: 348–354.

Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, et al. (2010) Common SNPs explain a large proportion of the heritability for human height. Nat Genet 42: 565–569.

Powell JE, Vischer PM, Godillard ME (2010) Reconciling the analysis of IBD and IBS in complex trait studies. Nat Rev Genet.

Vischer PM, Medland SE, Ferreira MA, Morley KI, Zhu G, et al. (2006) Assumption-free estimation of heritability from genome-wide identity-by-descent sharing between full siblings. PLoS Genet 2: e41. doi:10.1371/journal.pgen.0020041.