Differential Allelic Expression in the Human Genome: A Robust Approach To Identify Genetic and Epigenetic Cis-Acting Mechanisms Regulating Gene Expression

David Serre1,2*, Scott Gurd1, Bing Ge1,2, Robert Sladek1, Donna Sinnett1,2, Eef Harmsen1,2, Marina Bibikova3, Eugene Chudin1,2b, David L. Barker3, Todd Dickinson3, Jian-Bing Fan3, Thomas J. Hudson1,2c

1 McGill University, Montreal, Quebec, Canada, 2 Genome Quebec Innovation Centre, Montreal, Quebec, Canada, 3 Illumina Inc., San Diego, California, United States of America

Abstract

The recent development of whole genome association studies has lead to the robust identification of several loci involved in different common human diseases. Interestingly, some of the strongest signals of association observed in these studies arise from non-coding regions located in very large introns or far away from any annotated genes, raising the possibility that these regions are involved in the etiology of the disease through some unidentified regulatory mechanisms. These findings highlight the importance of better understanding the mechanisms leading to inter-individual differences in gene expression in humans. Most of the existing approaches developed to identify common regulatory polymorphisms are based on linkage/association mapping of gene expression to genotypes. However, these methods have some limitations, notably their cost and the requirement of extensive genotyping information from all the individuals studied which limits their applications to a specific cohort or tissue. Here we describe a robust and high-throughput method to directly measure differences in allelic expression for a large number of genes using the Illumina Allele-Specific Expression BeadArray platform and quantitative sequencing of RT-PCR products. We show that this approach allows reliable identification of differences in the relative expression of the two alleles larger than 1.5-fold (i.e., deviations of the allelic ratio larger than 60:40) and offers several advantages over the mapping of total gene expression, particularly for studying humans or outbred populations. Our analysis of more than 80 individuals for 2,968 SNPs located in 1,380 genes confirms that differential allelic expression is a widespread phenomenon affecting the expression of 20% of human genes and shows that our method successfully captures expression differences resulting from both genetic and epigenetic cis-acting mechanisms.


Editor: Emmanouil T. Dermietzelis, The Wellcome Trust Sanger Institute, United Kingdom

Received January 10, 2007; Accepted January 15, 2008; Published February 29, 2008

Copyright: © 2008 Serre et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work is supported by Genome Canada and Genome Quebec. TJH is the recipient of a Clinician-Scientist Award in Translational Research by the Burroughs Wellcome Fund and an Investigator Award from CIHR. Illumina Inc. provided reagents/materials/analysis tools necessary for the conduct of the ASE experiments.

Competing Interests: MB, DLB, TD, and J-BF are employed by the company that makes the array chips analyzed. EC is a former employee.

* E-mail: serredd@ccf.org

† Current address: Genomic Medicine Institute, Cleveland Clinic Foundation, Cleveland, Ohio, United States of America

‡ Current address: Merck & Co. Inc., Seattle, Washington, United States of America

§ Current address: Ontario Institute for Cancer Research, Toronto, Ontario, Canada

Introduction

Understanding the genetic causes of phenotypic variation in humans still remains a major challenge for human genetics. In hundreds of cases, a single DNA sequence polymorphism affecting a protein coding sequence has been linked to a clear simple Mendelian phenotype (see e.g. [1]) and, for a much smaller but increasing number of cases, to more complex phenotypes [2–4]. Recent developments in high-density genotyping technologies have led to the completion of several whole genome association studies that test hundreds of thousands of markers for a specific disease. While earlier studies essentially focused on variants in coding sequences and regions immediately surrounding candidate genes, whole genome scans interrogate, in an unbiased way, most of the human genome including large regions of non-coding DNA that had not been studied previously. Interestingly, some of the strongest signals observed in these association studies are located in non-coding regions, either in large introns (e.g. [5–7]) or far away from any annotated loci (e.g. [8] and references therein). The mechanisms connecting these polymorphisms to the etiology of the diseases are still unclear but regulation of gene expression remains an obvious candidate. It is thus becoming particularly important to have a powerful and reliable method to easily test the influence of DNA polymorphisms on gene expression. One of the approaches commonly used to identify regulatory polymorphisms is to look for statistical associations between variation in gene expression and individual genotypes [9,10]. This method offers the advantage of simultaneously analyzing thousands of genes using gene expression arrays and has yielded fascinating results in yeast [11,12] and mouse [13–16]. Its application in humans [17–24] suffers from relatively low statistical power due to potential inter-individual differences in a large number of causal variants involved in the regulation of a specific gene [25], their modest effects and the burden of the multiple testing correction necessary to take into
account the large number of independent tests performed. In addition, since this approach requires extensive genotype information for all individuals, it is costly to apply to new samples. An alternative approach is to compare the relative expression of the two alleles in one individual: the effect of a polymorphism affecting in cis the regulation of a particular transcript can be detected by measuring the relative expression of the two alleles in heterozygous individuals using a transcribed SNP as a marker [26–32]. Several studies have used this approach in humans but have been criticized for their low throughput or the apparent high variability. Here we describe a novel array-based method that allows high-throughput assessment of differential allelic expression. We used a modified version of the Illumina GoldenGate genotyping platform, the Allele-Specific Expression (ASE) assay, to assess the extent of differential allelic expression for over 1300 genes in more than 80 human lymphoblastoid cell lines (LCLs). Our analyses include 352 genes located in ENCODE regions and chromosome 21 that have been previously screened for cis-regulatory polymorphisms using total gene expression [19]. This allows us to directly compare the advantages and drawbacks of the two approaches in terms of range, sensitivity and robustness. We specifically address the issue of experimental noise and reproducibility of the findings and show that biology, not experimental variability, is responsible for the patterns observed. We discuss the relevance of our results for the identification of the molecular mechanisms regulating gene expression, as well as their implications for future genetic studies.

Results
Assessment of Differential Allelic Expression Using Illumina ASE Assay

We first assessed the extent of differential allelic expression at 1,432 exonic SNPs using 81 individual LCLs with the Illumina ASE technology (Figure 1 and Table S1 for the composition of the Illumina ASE Cancer Panel). This technology uses primer extension assays with fluorescence-labeled allele-specific primers to measure the proportion of each allele separately at the genomic and transcriptomic levels (Figure 2). Five hundred and twelve SNPs (in 345 genes) displayed an expression level significantly higher than background in at least three heterozygous individuals and were further investigated (see Materials and Methods for details). The extent of differential allelic expression at each SNP was obtained by comparing the relative amount of each allele in RNA to the ratio observed in DNA.

As a first effort to determine if the assay could reliably be used to assess differential expression we generated spike mixes using varying proportions of total RNA extracts from two individuals. For 20 exonic SNPs located in expressed transcript, the two individuals are homozygous for the different alleles (i.e. respectively AA and BB), while for 192 SNPs one individual is

Author Summary

We describe a new methodology to identify individual differences in the expression of the two copies of one gene. This is achieved by comparing the mRNA level of the two alleles using a heterozygous polymorphism in the transcript as marker. We show that this approach allows an exhaustive survey of cis-acting regulation in the genome; we can identify allelic expression differences due to epigenetic mechanisms of gene regulation (e.g. imprinting or X-inactivation) as well as differences due to the presence of polymorphisms in regulatory elements. The direct comparison of the expression of both alleles nullifies possible trans-acting regulatory effects (that influence equally both alleles) and thus complements the findings from gene expression association studies. Our approach can be easily applied to any cohort of interest for a wide range of studies. It notably allows following up association signals and testing whether a gene sitting on a particular haplotype is over- or under-expressed, or can be used for screening cancer tissues for aberrant gene expression due to newly arisen mutations or alteration of the methylation patterns.

Figure 1. Experiment design and results obtained for the two panels used in the study. The Overall column corresponds to the combination of the two panels. The detailed composition of each panel is presented on Supplemental Table S1.

doi:10.1371/journal.pgen.1000006.g001
heterozygous and the other homozygous (i.e. either AB and AA, or AB and BB). Since the expression of each gene may differ between the two individuals, one does not expect to observe an exact translation of the proportions of total RNA mixed to the “allelic” expression level. However, the allelic expression differences estimated for the different spike mixes should be the proportional to each other. For all homozygous/homozygous mixes (20 out of 20 SNPs) and 83% of the heterozygous/homozygous mixes (159 out of 192), we observed a significant linear correlation (p < 0.05) between the proportion of mixed total RNAs and the “allele-specific expression” estimated by the assay (Figure 3 and Figure S1).

We then tried to assess the threshold above which differential allelic expression would be genuine: even if the two alleles are equally expressed in one individual, we expect the ratio of allelic expression measured at a given marker to deviate stochastically from 50:50 due to experimental variability. In order to differentiate technical noise from biological signal (i.e. the differences in allelic expression due to differential cis-acting regulation), we evaluated the extent of experimental variability in the assay by comparing independent estimates of allelic imbalance for duplicates of individual RNA. We used duplicated measurements from 81 individuals at all SNPs expressed to determine a robust estimation of the experimental variability (N = 31,503 duplicates). After averaging duplicate differences for each SNP over all individuals, we observed than less than 3% of the SNPs show a population average variability greater than 10% (see Materials and Methods for more details and Figure S2). This level of experimental variability corresponds to a ratio of allelic expression of 60:40 (i.e. 1.5-fold difference). Thus, population-average allelic expression ratio at any SNP lower than 60:40 can be explained by experimental noise, while a SNP displaying a population-average differential allelic expression greater than this threshold most likely reflects a biological process affecting cis-

Figure 2. Overview of the Illumina Allele-Specific Expression assay. Genomic DNA and total RNA are separately converted into biotinylated DNA and amplified using fluorescence-labeled universal primers following extension and ligation of allele-specific assay oligo-nucleotides. PCR products are captured by locus-specific beads and the fluorescence of each dye (i.e. allele) at each locus is measured by quantitative fluorescence imaging (see Materials and Methods for details).

doi:10.1371/journal.pgen.1000006.g002

Figure 3. Result of the spike mixes experiment for four genes. The graph shows the logarithm of the dye ratio (y-axis) for four mixes of total RNA extract (x-axis) of NA12155 (at 0 on the x-axis) and NA07000 (at 1 on the x-axis). The original RNA extracts mixed and the four mixes for FRK (rs580396) are shown in blue, for GNAS (rs7121) in pink, for ITPR3 (rs2229634) in yellow and for RAD52 (rs1051672) in green.

doi:10.1371/journal.pgen.1000006.g003
regulation. We are interested in the present study in identifying loci with common allelic expression differences and we thus focused on population-average differential allelic expression: the average over all heterozygous individuals of the extent of allelic expression differences, regardless of which allele is over expressed (this is addressed later). The identification of a single individual with dramatic allelic expression difference is also possible using the same approach [but a different detection cut-off] but is beyond the scope of this paper. Among the 345 genes expressed in this first panel (512 SNPs), 72 (87 SNPs) displayed an average level of allelic imbalance larger than this 40:60 cut-off and were thus considered to display significant differences in allelic expression (Figure 4).

These analyses rely on the observation of the three genotypes in the population (i.e. AA, AB and BB). To also include SNPs with a lower minor allele frequency for each it was not possible to observe homozygotes for the minor allele in our small sample, we designed a second analysis method using solely the heterozygous individuals. If the alleles are differentially regulated we expect to observe in some cases a very large variance in the ratio of the two alleles in the population. We used this approach to determine SNPs for which the heterozygous individuals harbor a variance of the allelic ratio higher than expected using a Maximum expectation algorithm (see Materials and Methods for details). This approach does not allow us to quantify the overall extent of differential allelic expression but identifies 8 genes with differential allelic expression that were not identified by the previous method.

When one considers the estimates of allelic expression obtained using different SNPs in a same transcript showing significant differences in allele expression (i.e. with a population-average ratio greater than 60:40), we note that 36 out of 44 correlations between individuals estimates are significant (for an average $r^2$ of 0.83). Individuals showing a large allelic expression difference at one SNP display similar patterns at all heterozygous positions of the transcript (an example is shown on Figure 5). This observation supports our findings that the experimental variability is low in the Illumina ASE assay and that this assay allows quantitative assessment of differential allelic expression. Consequently, the population-average estimates of allelic imbalance obtained with different markers in the same transcript tend to be similar (Table S2) but can vary since different individuals will be included in the average (depending on whether they are heterozygotes at this marker).

Validation of Allelic Imbalance Estimates Using Quantitative Sequencing

To further assess the validity of our results, we randomly selected 25 genes tested on the Illumina ASE platform and used quantitative sequencing of RT-PCR products [33] to measure allelic imbalance for the same SNP in the same individuals (Figure 6).
The selected genes consisted of eight autosomal genes with significant allelic imbalance and 17 genes for which the level of differential allelic expression did not reach our significance threshold. We analyzed the same 81 individual LCLs using RNA from the same extract as for the Illumina assay. Overall, we observed a strong correlation between the estimates of allelic imbalance obtained for each individual using the two methods for the genes with a ratio of allelic expression larger than our 40:60 cut-off ($r^2 > 0.8$ for 6 out of 8 genes, see Figure 6A as an example). The correlations were not statistically significant for the genes for which the average difference in allelic imbalance did not reach our significance threshold (16 out of 17 genes, Figure 6B); minor deviations observed in the allelic ratio for these genes likely correspond to random variations and are therefore not expected to be reproducible. The strength of the correlation (measured by Pearson’s $r^2$) for all 25 genes is shown on Figure 4. One SNP in CD44 (rs8193) displayed a low but significant correlation between our estimates of allelic imbalance obtained from the Illumina assay and those using quantitative sequencing ($p < 10^{-7}, r^2 = 0.4075$), even though the average level of allelic imbalance was below our significance cut-off on the Illumina platform. Allelic imbalance at CD44 has been previously reported [30] and it is likely that the signal observed at that gene is real but corresponds to a low level of differential expression. Two genes (ABL2, XRCC1) showed significant allelic imbalance in the Illumina ASE assay (with a mean allelic ratio of, respectively, 70:30 and 65:35) but were not validated by quantitative sequencing. Manual inspection of the Illumina results for these genes revealed that the allelic expression ratios were estimated using a small number of homozygotes for the minor allele (respectively, 1 and 2 individuals) which led to an incorrect estimation of the expected dye ratio for heterozygotes and to a general over-estimation of allelic imbalance. For further analyses, we manually curated the list of all genes with significant differential allelic expression to remove potential false positives due to low number of homozygous individuals.

Reproducibility of Allelic Imbalance Measured in Cell Lines

Our study uses lymphoblastoid cell lines (LCLs) and it remains controversial whether culture conditions could artefactually generate differential allelic expression. We therefore tested whether allelic imbalance is influenced by harvesting the cells after different numbers of passages. This allowed us to control the effect of changes in the culture environment including pH, nutrient concentration and cell density at the time of harvest. We compared our estimation of allelic imbalance for three genes with significant population-average allelic imbalance in 47 individuals.
using recently thawed LCLs harvested after the 2nd, 4th and 6th successive passages (respectively, “growths” 1, 2 and 3). The correlations between the allelic imbalance estimations are displayed in Figure S4 for the comparison of growths 2 and 3. The estimations of allelic imbalance after different passages were very similar to each other (r²>0.9), supporting the idea that differential allelic expression is little influenced by variations in culture environment.

**Comparison of Differential Allelic Expression with Total Gene Expression Association**

An alternative approach to identifying cis-regulatory polymorphisms is to test for statistical association in a population between total gene expression measurements and the genotypes at markers in or surrounding the transcript. Interestingly, one of the genes showing the most marked difference in allele expression in our analysis is one of the 14 genes identified by Cheung and colleagues in a previous genome-wide study [17]. One comprehensive analysis was recently conducted for 512 ReSeq genes in ENCODE regions and chromosome 21 using LCLs from 60 unrelated individuals genotyped by the HapMap project [19]. In order to compare the respective strengths and weaknesses of total gene expression mapping and differential allelic expression, we designed a second panel that includes SNPs in the same genomic regions to analyze the same individual LCLs (Figure 1 and Table S1 for details). Using the information from the HapMap phase 1 (release 16) to select common exonic SNPs, we were able to include 228 and 124 genes from, respectively, ENCODE regions and chromosome 21, while Stranger and colleagues selected 321 and 191 genes (after screening for genes with high variance in their expression among individuals, see Materials and Methods for details). From the regions analyzed by Stranger and al., two-hundreds and ninety SNPs (in 170 genes) showed an expression level significantly higher than the background in three or more heterozygous individuals and were further investigated. Forty-nine out of 170 genes show significant level of differential allelic expression including 6 out of the 21 genes identified by Stranger and colleagues and present in our panel. Additionally, TTC3 which shows significant association between total gene expression and genotypes in the study by Stranger et al. shows patterns of allelic expression consistent with differential allelic expression (including a very high correlation between the extent of differential allelic expression estimated using different SNPs) on the Illumina ASE assay, even though it did not pass the significance cutoff. Overall in this second panel, 497 SNPs in 317 genes were expressed in three or more heterozygous individuals (out of 1536 SNPs in 674 genes) and 78 SNPs in 63 genes showed a significant level of differential allelic expression (Figure 1).

**Intronic SNPs Can Be Used To Assess Differential Allelic Expression**

To test whether intronic SNPs could be used instead of exonic SNP, we included for each gene on the second panel one intronic SNP. In general, intronic SNPs were less successfully analyzed and SNP, we included for each gene on the second panel one intronic SNP of a gene was detected in the RNA extract, it typically yields estimates of differential allelic expression very similar with those obtained using exonic SNPs.

**Differential Allelic Expression in the Human Genome**

Overall, 177 out of 1,009 expressed SNPs (in 140 out of 643 genes expressed, 22%) display population-average ratios of allelic expression larger than 40:60 or an higher than expected variance in allelic expression among heterozygous individuals and are thus unlikely to result solely from stochastic variation in the experiment (Figure S6). Table 1 shows the 133 SNPs (100 genes) with significant allelic imbalance after manual curation to remove possible false positives due to a low number of individual homozygous for the minor allele (this list is likely over-conservative and the complete data is presented in Table S2).

Many of the genes with the highest extent of allelic imbalance in LCLs are located on the X-chromosome. While it is known that one allele at most X-linked genes is silenced in females by inactivation of one entire chromosome [34,35], we would expect that a polyclonal cell population (in which half of the cells inactivate one X chromosome and the other 50% inactivate the alternate X chromosome) would give a similar level of expression for both alleles. However, all X-linked genes on our two SNP panels (22 SNPs in 12 genes) were among the top 5% of genes with most dramatic allelic imbalance patterns. The extent of allelic imbalance at a given gene varies among individual LCLs but interestingly, the patterns of allelic imbalance are very consistent across genes for a given individual (Figure S7). Additionally, the inheritance of the expressed allele (determined, when possible, using the pedigree information for the two families included in this study) appeared random. It has been previously proposed that the extent of clonality of a cell line could explain the patterns of allelic imbalance at genes with random mono-allelic expression [30]: clonal cells will all have the same X chromosome inactivated and thus display very high ratios of allelic imbalance. In contrast, cell-lines composed of a polyclonal population of lymphoblasts will have one or the other of their X chromosomes inactivated in different cells and thus an apparent expression of both alleles (i.e., a low extent of differential allelic expression). Our observations at X-linked genes are consistent with this hypothesis and the biased clonality of these LCLs, which were created over 20 years ago and passed numerous times (see also [30]).

The two autosomal genes displaying the most dramatic allelic imbalance patterns have previously been shown to be imprinted in humans: PEG10 [36] and SNRPN [37]. In addition, KCNQ1, MEST and ZNF215 which are imprinted in humans [38–40] also show significant differences in allelic expression (Table 1). The mode of inheritance of the expressed allele also corresponds, in each case, to what has been described for the expression of these genes: for PEG10 and SNRPN, heterozygous individuals express the paternally-inherited allele (i.e. maternally imprinted) while for KCNQ1 the maternally-inherited allele is expressed. Our limited pedigree information is not conclusive for MEST and ZNF215.

The only other known imprinted gene analyzable in our panel, PLAGL1 [41,42] did not pass the significance threshold (i.e. an allelic ratio greater than 60:40) but shows a population average allelic imbalance larger than 55:45 and a high correlation between the two SNPs analyzable in the panel (r²=0.9), supporting the idea that differential allelic expression is little influenced by variations in culture environment.

**Intronic SNPs Can Be Used To Assess Differential Allelic Expression**

To test whether intronic SNPs could be used instead of exonic SNP, we included for each gene on the second panel one intronic SNP. In general, intronic SNPs were less successfully analyzed and passed our expression threshold only for genes highly expressed in each case, to what has been described for the expression of these genes: for PEG10 and SNRPN, heterozygous individuals express the paternally-inherited allele (i.e. maternally imprinted) while for KCNQ1 the maternally-inherited allele is expressed. Our limited pedigree information is not conclusive for MEST and ZNF215.

The only other known imprinted gene analyzable in our panel, PLAGL1 [41,42] did not pass the significance threshold (i.e. an allelic ratio greater than 60:40) but shows a population average allelic imbalance larger than 55:45 and a high correlation between the two SNPs analyzable in the panel (r²=0.9), therefore it likely represents a significant difference in allelic expression.

The 83 remaining genes (103 SNPs) with significant population-average allelic imbalance included several genes for which allelic imbalance had been shown in previous studies (e.g. IL1A or IGF1 described in [30]). For some genes (e.g., CHISL2), one allele/haplotype is clearly expressed more than the other in heterozygotes and the inheritance pattern in families supports a genetic cause for allelic imbalance. For other genes, neither the direction of allelic imbalance nor the pedigree analysis allowed us to easily differentiate the genetic/epigenetic cause of the differential allelic expression (Table 1). For 56 genes with significant differences in allelic expression we tested whether differential allelic expression...
Table 1. Genes with significant allelic expression difference.

<table>
<thead>
<tr>
<th>rs</th>
<th>Gene</th>
<th>Panel</th>
<th>Chr</th>
<th># Hets</th>
<th>Variance &amp; Mean</th>
<th>Average Al</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs2292305</td>
<td>THBS1</td>
<td>Cancer</td>
<td>15</td>
<td>13</td>
<td>Mean shifted</td>
<td>NA</td>
</tr>
<tr>
<td>rs2288539</td>
<td>NR2F6</td>
<td>Cancer</td>
<td>19</td>
<td>18</td>
<td>Mean shifted</td>
<td>NA</td>
</tr>
<tr>
<td>rs2459216</td>
<td>OAT</td>
<td>Cancer</td>
<td>10</td>
<td>6</td>
<td>High variance</td>
<td>NA</td>
</tr>
<tr>
<td>rs3750105</td>
<td>PEG10</td>
<td>Cancer</td>
<td>7</td>
<td>8</td>
<td>High variance</td>
<td>NA</td>
</tr>
<tr>
<td>rs3745410</td>
<td>LILRB3</td>
<td>Encode</td>
<td>19</td>
<td>5</td>
<td>High variance</td>
<td>NA</td>
</tr>
<tr>
<td>rs2839500</td>
<td>TMRP553</td>
<td>Encode</td>
<td>21</td>
<td>13</td>
<td>High variance</td>
<td>NA</td>
</tr>
<tr>
<td>rs546782</td>
<td>FGF9</td>
<td>Cancer</td>
<td>13</td>
<td>10</td>
<td>High variance</td>
<td>NA</td>
</tr>
<tr>
<td>rs17756426</td>
<td>DDX43</td>
<td>Encode</td>
<td>6</td>
<td>6</td>
<td>High variance</td>
<td>NA</td>
</tr>
<tr>
<td>rs1893963</td>
<td>DSC2</td>
<td>Cancer</td>
<td>18</td>
<td>14</td>
<td>High variance</td>
<td>NA</td>
</tr>
<tr>
<td>rs9978281</td>
<td>C21orf7</td>
<td>Encode</td>
<td>21</td>
<td>12</td>
<td>High variance</td>
<td>NA</td>
</tr>
<tr>
<td>rs2834601</td>
<td>CUC6</td>
<td>Encode</td>
<td>21</td>
<td>8</td>
<td>High variance</td>
<td>NA</td>
</tr>
<tr>
<td>rs2070807</td>
<td>DNASE1L1</td>
<td>Encode</td>
<td>X</td>
<td>5</td>
<td>High variance</td>
<td>NA</td>
</tr>
<tr>
<td>rs705</td>
<td>SNRPN</td>
<td>Cancer</td>
<td>15</td>
<td>42</td>
<td>High variance</td>
<td>95:05</td>
</tr>
<tr>
<td>rs7810469</td>
<td>PEG10</td>
<td>Cancer</td>
<td>7</td>
<td>12</td>
<td>High variance</td>
<td>95:05</td>
</tr>
<tr>
<td>rs13073</td>
<td>PEG10</td>
<td>Cancer</td>
<td>11</td>
<td>39</td>
<td>High variance</td>
<td>95:05</td>
</tr>
<tr>
<td>rs2834601</td>
<td>CUC6</td>
<td>Encode</td>
<td>21</td>
<td>8</td>
<td>High variance</td>
<td>95:05</td>
</tr>
<tr>
<td>rs9978281</td>
<td>C21orf7</td>
<td>Encode</td>
<td>21</td>
<td>12</td>
<td>High variance</td>
<td>95:05</td>
</tr>
<tr>
<td>rs2834601</td>
<td>CUC6</td>
<td>Encode</td>
<td>21</td>
<td>8</td>
<td>High variance</td>
<td>95:05</td>
</tr>
<tr>
<td>rs2070807</td>
<td>DNASE1L1</td>
<td>Encode</td>
<td>X</td>
<td>5</td>
<td>High variance</td>
<td>NA</td>
</tr>
</tbody>
</table>

Table 1. cont.

<table>
<thead>
<tr>
<th>rs</th>
<th>Gene</th>
<th>Panel</th>
<th>Chr</th>
<th># Hets</th>
<th>Variance &amp; Mean</th>
<th>Average Al</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs1571858</td>
<td>GSTM3</td>
<td>Encode</td>
<td>1</td>
<td>17</td>
<td>High variance</td>
<td>75:25</td>
</tr>
<tr>
<td>rs3817405</td>
<td>PLXDC2</td>
<td>Cancer</td>
<td>10</td>
<td>12</td>
<td>High variance</td>
<td>75:25</td>
</tr>
<tr>
<td>rs10863</td>
<td>MEST</td>
<td>Cancer</td>
<td>7</td>
<td>29</td>
<td>High variance</td>
<td>70:30</td>
</tr>
<tr>
<td>rs10336</td>
<td>CCL5</td>
<td>Cancer</td>
<td>4</td>
<td>11</td>
<td>High variance</td>
<td>70:30</td>
</tr>
<tr>
<td>rs5351</td>
<td>EDNRB</td>
<td>Cancer</td>
<td>13</td>
<td>22</td>
<td>High variance</td>
<td>70:30</td>
</tr>
<tr>
<td>rs1022477</td>
<td>RIBC2</td>
<td>Encode</td>
<td>22</td>
<td>26</td>
<td>High variance</td>
<td>70:30</td>
</tr>
<tr>
<td>rs6007897</td>
<td>CELSR1</td>
<td>Encode</td>
<td>22</td>
<td>10</td>
<td>High variance</td>
<td>70:30</td>
</tr>
<tr>
<td>rs11264793</td>
<td>FCR1L3</td>
<td>Encode</td>
<td>1</td>
<td>24</td>
<td>High variance</td>
<td>70:30</td>
</tr>
<tr>
<td>rs4445669</td>
<td>IGF5</td>
<td>Cancer</td>
<td>11</td>
<td>40</td>
<td>High variance</td>
<td>70:30</td>
</tr>
<tr>
<td>rs4767884</td>
<td>PDXN</td>
<td>Cancer</td>
<td>12</td>
<td>26</td>
<td>High variance</td>
<td>70:30</td>
</tr>
<tr>
<td>rs1041985</td>
<td>CDH2</td>
<td>Encode</td>
<td>18</td>
<td>38</td>
<td>High variance</td>
<td>70:30</td>
</tr>
<tr>
<td>rs140519</td>
<td>KLHC7B</td>
<td>Encode</td>
<td>22</td>
<td>30</td>
<td>High variance</td>
<td>70:30</td>
</tr>
<tr>
<td>rs17197</td>
<td>PTGER2</td>
<td>Encode</td>
<td>14</td>
<td>10</td>
<td>High variance</td>
<td>70:30</td>
</tr>
<tr>
<td>rs1803965</td>
<td>MGMT</td>
<td>Cancer</td>
<td>10</td>
<td>5</td>
<td>High variance</td>
<td>70:30</td>
</tr>
<tr>
<td>rs2837029</td>
<td>C21orf13</td>
<td>Encode</td>
<td>21</td>
<td>15</td>
<td>Mean shifted</td>
<td>70:30</td>
</tr>
<tr>
<td>rs724558</td>
<td>SERPINB10</td>
<td>Encode</td>
<td>18</td>
<td>23</td>
<td>High variance</td>
<td>70:30</td>
</tr>
<tr>
<td>rs6007594</td>
<td>C22orf8</td>
<td>Encode</td>
<td>22</td>
<td>28</td>
<td>High variance</td>
<td>70:30</td>
</tr>
<tr>
<td>rs7561</td>
<td>LAMB1</td>
<td>Cancer</td>
<td>7</td>
<td>21</td>
<td>High variance</td>
<td>70:30</td>
</tr>
<tr>
<td>rs165602</td>
<td>NEFH</td>
<td>Encode</td>
<td>22</td>
<td>6</td>
<td>High variance</td>
<td>70:30</td>
</tr>
<tr>
<td>rs10593</td>
<td>ITGB1BP1</td>
<td>Cancer</td>
<td>2</td>
<td>23</td>
<td>High variance</td>
<td>65:35</td>
</tr>
<tr>
<td>rs1042531</td>
<td>PCK1</td>
<td>Encode</td>
<td>20</td>
<td>9</td>
<td>High variance</td>
<td>65:35</td>
</tr>
<tr>
<td>rs1025689</td>
<td>IL17RB</td>
<td>Cancer</td>
<td>3</td>
<td>29</td>
<td>High variance</td>
<td>65:35</td>
</tr>
<tr>
<td>rs225334</td>
<td>TFF2</td>
<td>Encode</td>
<td>21</td>
<td>17</td>
<td>High variance</td>
<td>65:35</td>
</tr>
<tr>
<td>rs17207369</td>
<td>LILRB2</td>
<td>Encode</td>
<td>19</td>
<td>18</td>
<td>High variance</td>
<td>65:35</td>
</tr>
<tr>
<td>rs3856806</td>
<td>PPARG</td>
<td>Encode</td>
<td>3</td>
<td>5</td>
<td>High variance</td>
<td>65:35</td>
</tr>
<tr>
<td>rs300239</td>
<td>ENC1</td>
<td>Cancer</td>
<td>5</td>
<td>27</td>
<td>High variance</td>
<td>65:35</td>
</tr>
<tr>
<td>rs677688</td>
<td>IMPACT</td>
<td>Cancer</td>
<td>18</td>
<td>7</td>
<td>High variance</td>
<td>65:35</td>
</tr>
<tr>
<td>rs6104</td>
<td>SERPINB2</td>
<td>Cancer</td>
<td>18</td>
<td>19</td>
<td>High variance</td>
<td>65:35</td>
</tr>
<tr>
<td>rs8097425</td>
<td>SERPINB10</td>
<td>Encode</td>
<td>18</td>
<td>24</td>
<td>High variance</td>
<td>65:35</td>
</tr>
<tr>
<td>rs1071676</td>
<td>IL1B</td>
<td>Cancer</td>
<td>2</td>
<td>29</td>
<td>High variance</td>
<td>65:35</td>
</tr>
<tr>
<td>rs9612234</td>
<td>GNAZ</td>
<td>Encode</td>
<td>22</td>
<td>16</td>
<td>High variance</td>
<td>65:35</td>
</tr>
<tr>
<td>rs2042333</td>
<td>WNT2</td>
<td>Encode</td>
<td>7</td>
<td>24</td>
<td>High variance</td>
<td>65:35</td>
</tr>
<tr>
<td>rs7927012</td>
<td>TRIM6</td>
<td>Encode</td>
<td>11</td>
<td>30</td>
<td>High variance</td>
<td>65:35</td>
</tr>
<tr>
<td>rs2042333</td>
<td>WNT2</td>
<td>Encode</td>
<td>7</td>
<td>15</td>
<td>High variance</td>
<td>65:35</td>
</tr>
<tr>
<td>rs162549</td>
<td>CYP1B1</td>
<td>Cancer</td>
<td>2</td>
<td>22</td>
<td>High variance</td>
<td>65:35</td>
</tr>
<tr>
<td>rs2075760</td>
<td>PLSCR3</td>
<td>Encode</td>
<td>17</td>
<td>19</td>
<td>High variance</td>
<td>65:35</td>
</tr>
<tr>
<td>rs2832236</td>
<td>C21orf7</td>
<td>Encode</td>
<td>21</td>
<td>40</td>
<td>High variance</td>
<td>65:35</td>
</tr>
<tr>
<td>rs15017</td>
<td>MOXD1</td>
<td>Encode</td>
<td>6</td>
<td>6</td>
<td>High variance</td>
<td>65:35</td>
</tr>
<tr>
<td>rs1035474</td>
<td>IMPACT</td>
<td>Cancer</td>
<td>11</td>
<td>31</td>
<td>High variance</td>
<td>60:40</td>
</tr>
<tr>
<td>rs12593359</td>
<td>RAD51</td>
<td>Cancer</td>
<td>15</td>
<td>44</td>
<td>High variance</td>
<td>60:40</td>
</tr>
<tr>
<td>rs1368439</td>
<td>IL12B</td>
<td>Encode</td>
<td>5</td>
<td>23</td>
<td>High variance</td>
<td>60:40</td>
</tr>
<tr>
<td>rs1103229</td>
<td>PPIA2</td>
<td>Encode</td>
<td>22</td>
<td>26</td>
<td>High variance</td>
<td>60:40</td>
</tr>
<tr>
<td>rs2839600</td>
<td>NDUFV3</td>
<td>Encode</td>
<td>21</td>
<td>17</td>
<td>High variance</td>
<td>60:40</td>
</tr>
<tr>
<td>rs3734744</td>
<td>MOXD1</td>
<td>Encode</td>
<td>6</td>
<td>11</td>
<td>High variance</td>
<td>60:40</td>
</tr>
<tr>
<td>rs18807</td>
<td>HLA</td>
<td>Cancer</td>
<td>6</td>
<td>15</td>
<td>High variance</td>
<td>60:40</td>
</tr>
<tr>
<td>rs743616</td>
<td>ARSA</td>
<td>Encode</td>
<td>22</td>
<td>37</td>
<td>High variance</td>
<td>60:40</td>
</tr>
<tr>
<td>rs6214</td>
<td>IGFI</td>
<td>Cancer</td>
<td>12</td>
<td>31</td>
<td>High variance</td>
<td>60:40</td>
</tr>
</tbody>
</table>
could be statistically associated to one of the SNP in the vicinity of the gene genotyped by the HapMap project (see Materials & Methods for details). The results of these tests for SERPINB10 and ABCG1 are shown on Figure 7 and the strongest nominal association for each gene is displayed on Table 2. Twenty-three genes still display statistical significant associations after Bonferroni correction for multiple testing (highlighted in green on Table 2) showing a clear enrichment relative to the 2–3 associations expected by chance. Our power to detect a significant association between a HapMap SNP and the under-/over-expressing chromosome in this setting is low due to our reduced sample size (only the heterozygous individuals are taken into account in this analysis) and the number of regulatory haplotypes identified is thus likely underestimated. Additionally, many SNPs are tested for each gene and it is thus possible that some of the regulatory haplotypes result from spurious associations (i.e. they are false positives). One argument against a very high rate of false positive in our analysis is that imprinted genes such as MEST or PEG10 do not show any signal of association (Figure S8) consistent with the fact that the cis-regulatory mechanism at these genes is not encoded in the DNA sequence. To further investigate the validity of our association, we attempted to independently confirm these regulatory haplotypes by testing for the statistical association between one SNP in the regulatory haplotype and gene expression level. We used gene expression measurements performed at the Wellcome Trust Sanger Institute (kindly provided by M. Dermitza-kis) on the same individual cell lines assayed by Illumina gene expression arrays. For each gene, we tested whether the homozygotes for the regulatory haplotype associated with low allelic expression in heterozygotes show a significantly lower gene expression level than the homozygous individuals for the regulatory haplotype associated with high allelic expression. We also performed locus-specific RT-PCR and quantified the level of gene expression using SYBR-Green for eleven genes for which differential allelic expression was significantly associated with allelic expression but for which expression data were not available (2 genes) or genes with strong association with a regulatory haplotype but that were not validated using the Sanger dataset (9 genes). Overall, out of the 47 genes with a significant association between a SNP (or several, defining the regulatory haplotype) and differential allelic expression at the nominal cut-off, 10 were confirmed using gene expression measurements while 5 other genes showed a trend but did not reach statistical significance (Table 2).

### Discussion

**Biological, Not Experimental Noise, Is Responsible for Differential Allelic Expression**

We analyzed differences in relative allelic expression (or allelic imbalance) at 1,380 human genes using 2,968 SNPs and more than 80 lymphoblastoid cell lines from individuals with European ancestry. Using quantitative sequencing we validated our results for a subset of genes and showed that the experimental variability in both settings is low and that the Illumina ASE assay and quantitative sequencing of RT-PCR products yield reproducible estimates of allelic imbalance consistent with each other. Overall, the experimental noise is much lower than the difference in relative allelic expression observed at many loci and therefore cannot be responsible for it. Additionally, the high concordance of the results obtained using different SNPs in the same transcript supports our findings that allelic imbalance, as we estimated it, is not an experimental artefact but reflects inherent biological differences in the relative expression of both alleles in heterozygous individuals. We also showed that lymphoblastoid cell lines, despite being simplified biological materials, are suitable resources to investigate mechanisms of gene regulation. Here, we demonstrated that our estimation of allelic imbalance is little affected by growth conditions and that LCLs harvested from different passages yield very similar results. Finally, the results efficiently recapitulate the consequences of the epigenetic mechanisms established in the individuals from which the cells have been derived (see also [43]). We are therefore confident that, overall, the patterns of allelic imbalance we observed are neither experimental artifacts, nor specific to the material studied, but represent a common biological phenomenon affecting human gene expression.

### Table 1. cont.

<table>
<thead>
<tr>
<th>rs</th>
<th>Gene</th>
<th>Panel Chr</th>
<th># Hets</th>
<th>Variance &amp; Mean</th>
<th>Average Ai</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs2822445 RMB11</td>
<td>Encode 21</td>
<td>38</td>
<td>60:40</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs4947963 EGFR</td>
<td>Encode 7</td>
<td>19</td>
<td>60:40</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs5275 PTGS2</td>
<td>Cancer 1</td>
<td>41</td>
<td>60:40</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs2257905 MGC33648</td>
<td>Encode 5</td>
<td>29</td>
<td>60:40</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs1029365 FLJ21062</td>
<td>Encode 7</td>
<td>30</td>
<td>60:40</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs2839536 TSGA2</td>
<td>Encode 21</td>
<td>19</td>
<td>60:40</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs6518322 LOC284837</td>
<td>Encode 21</td>
<td>29</td>
<td>60:40</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs2258119 C2orf91</td>
<td>Encode 21</td>
<td>25</td>
<td>60:40</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs2229730 CSK</td>
<td>Cancer 15</td>
<td>4</td>
<td>60:40</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs2206593 PTGS2</td>
<td>Cancer 1</td>
<td>10</td>
<td>High variance</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs2829777 JAM2</td>
<td>Encode 21</td>
<td>17</td>
<td>60:40</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs1053395 TUBB4</td>
<td>Cancer 19</td>
<td>32</td>
<td>60:40</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs1044104 BMP6</td>
<td>Cancer 6</td>
<td>22</td>
<td>60:40</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs1801719 F2R</td>
<td>Cancer 5</td>
<td>31</td>
<td>60:40</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs235768 BMP2</td>
<td>Cancer 20</td>
<td>7</td>
<td>High variance</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs2239730 ZNF215</td>
<td>Cancer 11</td>
<td>41</td>
<td>60:40</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs2270121 GAT7</td>
<td>Cancer 17</td>
<td>38</td>
<td>60:40</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs963075 SERPINB10</td>
<td>Encode 18</td>
<td>32</td>
<td>High variance</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs4820268 TMRPS56</td>
<td>Encode 22</td>
<td>38</td>
<td>60:40</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs4798 ITGB1B1P</td>
<td>Cancer 2</td>
<td>34</td>
<td>60:40</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs9782 ASC1L</td>
<td>Cancer 12</td>
<td>28</td>
<td>60:40</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs180817 BCR</td>
<td>Cancer 22</td>
<td>15</td>
<td>60:40</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs406271 TFRC</td>
<td>Cancer 3</td>
<td>33</td>
<td>60:40</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs1746217 FGF2</td>
<td>Cancer 4</td>
<td>43</td>
<td>High variance</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs2239731 ZNF215</td>
<td>Cancer 11</td>
<td>44</td>
<td>60:40</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs10916 CYP1B1</td>
<td>Cancer 2</td>
<td>13</td>
<td>High variance</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs2330033 KCNJ5</td>
<td>Encode 21</td>
<td>33</td>
<td>60:40</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs3088440 CDKN2A</td>
<td>Cancer 9</td>
<td>31</td>
<td>60:40</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs2855658 CYP1B1</td>
<td>Cancer 2</td>
<td>37</td>
<td>High variance</td>
<td>55:45</td>
<td></td>
</tr>
<tr>
<td>rs14983 MMP7</td>
<td>Cancer 11</td>
<td>22</td>
<td>High variance</td>
<td>55:45</td>
<td></td>
</tr>
<tr>
<td>rs3747676 FGF2</td>
<td>Cancer 4</td>
<td>42</td>
<td>High variance</td>
<td>55:45</td>
<td></td>
</tr>
<tr>
<td>rs10502001 MMP7</td>
<td>Cancer 11</td>
<td>19</td>
<td>High variance</td>
<td>55:45</td>
<td></td>
</tr>
<tr>
<td>rs2066575 DLEU1</td>
<td>Cancer 13</td>
<td>27</td>
<td>High variance</td>
<td>55:45</td>
<td></td>
</tr>
</tbody>
</table>

Hets: number of heterozygous individuals which express the transcript. Variance & Mean: indicates whether the analyses of variance/mean allelic expression detected significant deviation of the expression of both alleles (see Materials and Methods for details). Average Ai: population-average difference in allelic expression using all individuals heterozygous at this position (rounded down). These values correspond to values reported on the y-axis on Figure 4 and Figure S6. doi:10.1371/journal.pgen.1000006.t001
Differential Allelic Expression Identifies Consequences of Epigenetic Mechanisms of Gene Regulation

We showed that LCLs derived from female individuals still harbor the consequences of X-inactivation at all X-linked genes investigated, with one allele being transcriptionally silenced [34]. The extent of allelic imbalance detected at X-linked genes can vary among LCLs due to the various degrees of clonality of these cells but clonal LCLs consistently show complete silencing of one allele at all X-linked genes investigated (Figure S7). In addition, imprinting, established in the germ lines of the parents of the individuals from which the cells are derived [44], is also maintained in LCLs. In our experiments, PEG10, SNPRN, MEST and KCNQ1 show reduced or absent expression of one...
Table 2. Results of the association mapping and validation assays.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Chr #</th>
<th>Hets</th>
<th>Association</th>
<th>Sanger</th>
<th>RT-PCR</th>
<th>Mapped to</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABCG1</td>
<td>21</td>
<td>18</td>
<td>2.20E-10</td>
<td>6.81E-01</td>
<td>5.78E-01</td>
<td>Intron</td>
</tr>
<tr>
<td>ARSA</td>
<td>22</td>
<td>17</td>
<td>2.26E-05</td>
<td>NA</td>
<td></td>
<td>Gene</td>
</tr>
<tr>
<td>ASCL1</td>
<td>12</td>
<td>13</td>
<td>1.69E-02</td>
<td>1.37E-02</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BCR</td>
<td>22</td>
<td>6</td>
<td>6.06E-02</td>
<td>9.83E-01</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BMP6</td>
<td>6</td>
<td>5</td>
<td>4.76E-02</td>
<td>8.62E-01</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C21orf13</td>
<td>21</td>
<td>9</td>
<td>4.11E-05</td>
<td>No data</td>
<td>3.78E-02</td>
<td>Gene-5'</td>
</tr>
<tr>
<td>C21orf7</td>
<td>21</td>
<td>12</td>
<td>3.71E-01</td>
<td>7.06E-01</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C21orf91</td>
<td>21</td>
<td>17</td>
<td>1.54E-08</td>
<td>8.47E-04</td>
<td></td>
<td>Gene</td>
</tr>
<tr>
<td>C22orf8</td>
<td>22</td>
<td>8</td>
<td>1.55E-04</td>
<td>No data</td>
<td></td>
<td>Gene</td>
</tr>
<tr>
<td>CDH2</td>
<td>18</td>
<td>22</td>
<td>5.32E-05</td>
<td>7.00E-01</td>
<td></td>
<td>Gene-3'</td>
</tr>
<tr>
<td>CKC5L9</td>
<td>4</td>
<td>7</td>
<td>4.66E-03</td>
<td>3.43E-01</td>
<td>Whole Region</td>
<td></td>
</tr>
<tr>
<td>CYP1B1</td>
<td>2</td>
<td>17</td>
<td>2.11E-06</td>
<td>1.43E-02</td>
<td>Gene-5'</td>
<td></td>
</tr>
<tr>
<td>DDX43</td>
<td>6</td>
<td>9</td>
<td>3.47E-01</td>
<td>5.56E-01</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EDNRB</td>
<td>13</td>
<td>14</td>
<td>7.03E-03</td>
<td>6.17E-01</td>
<td>Whole Region</td>
<td></td>
</tr>
<tr>
<td>EGFR</td>
<td>7</td>
<td>6</td>
<td>4.55E-01</td>
<td>6.21E-01</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F2R</td>
<td>5</td>
<td>10</td>
<td>1.08E-05</td>
<td>7.32E-01</td>
<td>9.65E-02</td>
<td>3'UTR</td>
</tr>
<tr>
<td>FCR3L3</td>
<td>1</td>
<td>14</td>
<td>6.88E-05</td>
<td>No data</td>
<td>8.04E-03</td>
<td>Gene-3'</td>
</tr>
<tr>
<td>FGFR2</td>
<td>4</td>
<td>22</td>
<td>6.00E-03</td>
<td>7.63E-01</td>
<td>Gene-3'</td>
<td></td>
</tr>
<tr>
<td>FLJ35801</td>
<td>22</td>
<td>13</td>
<td>4.72E-02</td>
<td>6.23E-01</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GAS7</td>
<td>17</td>
<td>12</td>
<td>7.40E-07</td>
<td>2.96E-01</td>
<td>1.45E-01</td>
<td>3'UTR</td>
</tr>
<tr>
<td>GNAZ</td>
<td>22</td>
<td>5</td>
<td>7.94E-03</td>
<td>6.42E-01</td>
<td>1.16E-01</td>
<td>Gene-3'</td>
</tr>
<tr>
<td>IG1</td>
<td>12</td>
<td>13</td>
<td>2.38E-01</td>
<td>9.17E-01</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IG5F4</td>
<td>11</td>
<td>15</td>
<td>1.70E-03</td>
<td>2.91E-01</td>
<td>2.59E-01</td>
<td>Gene-3'</td>
</tr>
<tr>
<td>IL12B</td>
<td>5</td>
<td>13</td>
<td>1.92E-07</td>
<td>4.51E-02</td>
<td>Whole Region</td>
<td></td>
</tr>
<tr>
<td>IL17RB</td>
<td>3</td>
<td>14</td>
<td>5.98E-06</td>
<td>6.81E-01</td>
<td>7.33E-01</td>
<td>5'</td>
</tr>
<tr>
<td>IL1A</td>
<td>2</td>
<td>9</td>
<td>1.49E-01</td>
<td>8.74E-01</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IMPACT</td>
<td>18</td>
<td>11</td>
<td>7.52E-03</td>
<td>6.63E-01</td>
<td>5'distal ?</td>
<td></td>
</tr>
<tr>
<td>ITGB1BP1</td>
<td>2</td>
<td>18</td>
<td>1.58E-03</td>
<td>No data</td>
<td>Gene</td>
<td></td>
</tr>
<tr>
<td>JAM2</td>
<td>21</td>
<td>9</td>
<td>2.26E-03</td>
<td>NA</td>
<td>5'</td>
<td></td>
</tr>
<tr>
<td>KCNJ15</td>
<td>21</td>
<td>13</td>
<td>1.69E-02</td>
<td>3.25E-01</td>
<td>Gene</td>
<td></td>
</tr>
<tr>
<td>KCNQ1</td>
<td>11</td>
<td>21</td>
<td>2.07E-02</td>
<td>1.59E-01</td>
<td></td>
<td></td>
</tr>
<tr>
<td>KLHDC7B</td>
<td>22</td>
<td>21</td>
<td>3.72E-12</td>
<td>No data</td>
<td>Gene-3'</td>
<td></td>
</tr>
<tr>
<td>LILRP2</td>
<td>19</td>
<td>10</td>
<td>2.30E-02</td>
<td>No data</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOC284837</td>
<td>21</td>
<td>16</td>
<td>3.73E-02</td>
<td>No data</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MAPK10</td>
<td>4</td>
<td>11</td>
<td>2.84E-06</td>
<td>5.25E-01</td>
<td>Gene</td>
<td></td>
</tr>
<tr>
<td>MCAM</td>
<td>11</td>
<td>12</td>
<td>6.44E-04</td>
<td>5.21E-01</td>
<td>Gene/3'</td>
<td></td>
</tr>
<tr>
<td>MEST</td>
<td>7</td>
<td>9</td>
<td>1.53E-01</td>
<td>5.50E-01</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MGC33648</td>
<td>5</td>
<td>21</td>
<td>4.40E-06</td>
<td>3.59E-05</td>
<td>5'</td>
<td></td>
</tr>
<tr>
<td>MMP7</td>
<td>11</td>
<td>8</td>
<td>4.66E-03</td>
<td>7.46E-01</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MBD2</td>
<td>6</td>
<td>8</td>
<td>6.99E-03</td>
<td>1.12E-01</td>
<td>5'</td>
<td></td>
</tr>
<tr>
<td>NDURF3</td>
<td>21</td>
<td>11</td>
<td>2.84E-06</td>
<td>3.73E-05</td>
<td>Gene-3'</td>
<td></td>
</tr>
<tr>
<td>PEG10</td>
<td>7</td>
<td>22</td>
<td>4.59E-02</td>
<td>NA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PLSCR3</td>
<td>17</td>
<td>10</td>
<td>1.98E-02</td>
<td>2.22E-01</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PLXDC2</td>
<td>10</td>
<td>8</td>
<td>7.69E-02</td>
<td>3.40E-02</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PTG52</td>
<td>1</td>
<td>5</td>
<td>2.06E-01</td>
<td>NA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PXN</td>
<td>12</td>
<td>14</td>
<td>4.23E-04</td>
<td>9.24E-01</td>
<td>Gene-3'</td>
<td></td>
</tr>
<tr>
<td>RBM11</td>
<td>21</td>
<td>10</td>
<td>1.08E-05</td>
<td>4.59E-01</td>
<td>1.71E-01</td>
<td>3'</td>
</tr>
<tr>
<td>RIB2</td>
<td>22</td>
<td>12</td>
<td>7.40E-07</td>
<td>No data</td>
<td>Gene</td>
<td></td>
</tr>
<tr>
<td>SERPINB10</td>
<td>18</td>
<td>22</td>
<td>1.84E-05</td>
<td>5.98E-11</td>
<td>5'</td>
<td></td>
</tr>
<tr>
<td>SERPINB2</td>
<td>18</td>
<td>10</td>
<td>1.19E-04</td>
<td>2.98E-01</td>
<td>Gene-5'</td>
<td></td>
</tr>
</tbody>
</table>
allele and, when the mode of inheritance can be determined, it corresponds to the imprinting mechanism described in the literature (i.e., PEG10 and SNPRN are maternally imprinted, KCNQ1 is paternally imprinted). We thus observe extensive differential allelic expression (i.e., allelic ratio larger than 70:30) for all genes whose expression is known to be epigenetically regulated. This clearly shows that analysis of differential allelic expression is a suitable method for identifying the consequences of epigenetic mechanisms of gene regulation. The Illumina ASE assay would thus provide an efficient method to screen tumor tissues and identify patterns of differential allelic expression resulting from aberrant methylation or loss of imprinting that are known to be statistically associated with differences in allele expression that could indicate the existence of a regulatory haplotype (i.e., a region of one chromosome likely containing the polymorphism(s) causing the differential cis-regulation). These regions are often tens of kb long, consistent with previous descriptions of the linkage disequilibrium patterns in humans [50]. Although this approach does not identify the actual polymorphism(s) responsible for the differential cis-regulation, examination of these regulatory haplotypes provides some valuable insights on the mechanisms leading to differential expression and can guide future investigations. For example, the regulatory haplotype for GAS7 is almost exclusively restricted to the 3'UTR of the gene and may indicate that the patterns of allelic imbalance observed are due to differential mRNA processing, stability or the presence of a 3' enhancer. In contrast, the regulatory haplotype identified for MGC33648 is located in the 5' region and does not seem to overlap with the gene itself. This might be indicative of alternative promoter usage or differential transcription efficiency (e.g. due to differential transcription factor binding site affinity).

### Allelic Imbalance Is Complementary of Total Gene Expression Association

Several recent studies have used large-scale associations between gene expression and extensive genotype information to investigate gene regulation in humans, some of them using cell lines included in our study. In particular, Stranger and colleagues analyzed 630 genes located in ENCODE regions, on chromosome 21 and in one portion of chromosome 20. They found evidence of cis-acting regulation for 63 genes [19]. 2005). We were able to analyze 21 of these genes in our experiment. Six of them also showed evidence of cis-acting regulation (e.g. SERPINB10 or TSGA2) in our study while a seventh gene (TTC3) showed patterns consistent with differential allelic expression but did not reach our significance threshold. The remaining 14 genes did not show evidence of differential allelic expression in our analysis. Alternatively, we identified 10 new genes located in ENCODE region or chromosome 21 that showed significant level of differential allelic expression but were not detected in the Stranger study. Several non-exclusive reasons could explain the discrepancies between the results of the two approaches. First, it is worth noting that, even if the same individuals are analyzed by allelic-specific expression and gene expression association, the power to detect cis-acting effect differs depending on the allele frequency of the marker used: in gene expression association analysis all individuals are analyzed but the power in the regression analysis

<table>
<thead>
<tr>
<th>Gene</th>
<th>Chr</th>
<th># Hets</th>
<th>Association</th>
<th>Sanger</th>
<th>RT-PCR</th>
<th>Mapped to</th>
</tr>
</thead>
<tbody>
<tr>
<td>SNPRN</td>
<td>15</td>
<td>27</td>
<td>3.94E-02</td>
<td>3.78E-01</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TMPRSS6</td>
<td>22</td>
<td>12</td>
<td>7.40E-07</td>
<td>8.42E-01</td>
<td>3'UTR</td>
<td></td>
</tr>
<tr>
<td>TRIM6</td>
<td>11</td>
<td>8</td>
<td>2.56E-02</td>
<td>9.84E-01</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TSGA2</td>
<td>21</td>
<td>8</td>
<td>1.55E-04</td>
<td>9.81E-01</td>
<td></td>
<td>Gene</td>
</tr>
<tr>
<td>WNT2</td>
<td>7</td>
<td>9</td>
<td>9.05E-03</td>
<td>3.60E-01</td>
<td>6.45E-01</td>
<td>5' Gene</td>
</tr>
<tr>
<td>ZNF215</td>
<td>11</td>
<td>20</td>
<td>3.73E-02</td>
<td>1.45E-03</td>
<td>1.82E-02</td>
<td>Gene-5'</td>
</tr>
</tbody>
</table>

#Hets: number of heterozygous individuals for the gene considered among the HapMap individuals analyzed.

Association: lowest nominal p-value obtained by Fischer exact test of the association of the alleles of all SNPs with over-/under-expressing chromosomes. Associations remaining significant after Bonferroni correction for multiple testing are highlighted in green.

Sanger: p-value obtained in the linear regression of the Sanger gene expression measurements with the alleles of the SNP most strongly associated with allelic expression.

RT-PCR: p-value obtained in the linear regression of the locus-specific RT-PCR gene expression measurements with the alleles of the SNP most strongly associated with allelic expression.

Mapped to: broad localization of the regulatory haplotype with regards to the gene with significant difference in allele expression.

doi:10.1371/journal.pgen.1000006.t002
depends on their genotypes (e.g. the genotypes AA, AB and BB are encoded in the linear regression as 0, 1 and 2) while in allelic expression analysis only the individuals heterozygotes at the marker considered are analyzed. This can become particularly problematic to study differential allelic expression at some genes since it requires a relatively common exonic SNP to detect allelic imbalance. In this context, it is worth noting that intrinsic SNPs can successfully be used for genes that are highly expressed (see also [30]). Second, associations of gene expression to genotypes depends greatly on the linkage disequilibrium (LD) patterns and requires extensive genotype information from all the individuals in order to include one marker in LD with the regulatory polymorphism. Allelic expression, on the other hand, directly investigate cis effect directly at the gene level and thus only requires physical link between the gene and the regulatory polymorphism affecting it (i.e. they need to be on the same chromosome). Finally, the differences between allelic expression and gene expression mapping might indicate that some genes are also regulated by trans-acting mechanisms that differ among individuals: differential allelic expression is influenced only by cis-acting mechanisms of gene regulation while gene expression is influenced by cis- and trans-acting gene regulation. It is thus not unlikely that individual differences in trans-acting regulation swamp the signal from cis-acting polymorphisms. In this context, it is noteworthy that total gene expression mapping has been much more successful in mice and yeast for which the genetic heterogeneity is much lower and can be controlled (reviewed in [9,10,51]). In humans, or in any other outbred population, genetic heterogeneity greatly limits the identification of cis-acting mechanisms using gene expression data while measurements of differential allelic expression are unaffected.

We showed here that allelic expression assays are complementary from gene expression mapping and that the Illumina ASE assay overcomes two of the major limitations and criticisms of the former methodologies used to assess differential allelic expression: it allows a robust and high-throughput estimation of allelic imbalance: it is now possible to reliably screen hundreds of RNAs for several hundreds of genes in a couple of days. Additionally, when several SNPs can be used to assess differential allelic expression, the assay becomes very robust since each marker provides an independent estimation and one can test the correlation among estimates obtained at different positions. It is worth noting here that since this assay relies on the comparison of allelic ratio in DNA and RNA of each individual, it internally controls for the existence of polymorphisms in the primer sites or copy number variation encompassing the gene studied (that will affect equally DNA and RNA). Likely, the greatest advantage of the analysis of differential allelic expression over total gene expression is its flexibility. To identify differential regulation of gene expression using total gene expression, one needs extensive genotype information to test whether, at any polymorphic position, the gene expression differences among individuals segregate according to their genotype. This precludes a quick assessment of the expression of one locus in one cohort of particular interest or using a specific tissue. In contrary, differential allelic expression offers the advantage that any one gene can be quickly assessed in any cohort or tissue by simply comparing the expression of the two alleles in each individual (the amount of genetic information recently made available by the HapMap project allows a quick and easy selection of markers likely to be polymorphic for a given gene). The determination of regulatory haplotypes would still require extensive information concerning surrounding polymorphisms but the initial screening to determine whether one transcript is differentially cis-regulated can be done very efficiently with a handful of markers.

Conclusion
We showed that differential allelic expression is a robust approach to identify cis-acting mechanism of gene regulation. It complements gene expression association studies and offers additional perspectives, notably on epigenetic mechanisms of gene regulation. It could thus be particularly interesting to apply this assay to tumors to detect mis-regulated genes due to aberrant methylation patterns or loss of imprinting. In addition, our approach is applicable to any new cohort or tissue since it is self-sufficient to identify differential cis-regulation and does not require additional genotyping. It can be easily used to follow-up interesting non-coding regions associated to a particular disease and test if they are involved in the etiology of the disease through some regulatory effects on neighboring genes.

Materials and Methods
Sample Preparation
83 lymphoblastoid cell lines (LCL) derived from blood samples from the CEPH collection were selected for this project. They included 60 unrelated individuals obtained from Utah residents with ancestry from western and northern Europe for which DNA was genotyped for millions of SNPs covering the entire genome by the International HapMap Project. Additionally, 21 LCLs from CEPH pedigrees 1420 and 1444 were included to provide complete information on two three-generation CEPH families. Cells were grown at 37°C and 5% CO2 in RPMI 1640 medium (Invitrogen, Burlington, Canada) supplemented with 15% heat-inactivated fetal bovine serum (Sigma-Aldrich, Oakville, Canada), 2 mM L-glutamine (Invitrogen, Burlington, Canada) and penicillin/streptomycin (Invitrogen, Burlington, Canada). The cell growth was monitored with a hemocytometer and the cells were harvested when the density reached 0.8–1.1 x 10⁶ cells/mL. Cells were then resuspended and lysed in TRIzol reagent (Invitrogen, Burlington, Canada). For all LCLs, three successive growths were performed (corresponding to the 2nd, 3rd and 6th passages) after thawing frozen cell aliquots.

Illumina Allele-Specific Expression (ASE) Assay
We estimated allelic imbalance at 1,380 genes (two panels of ~1,500 SNPs, Figure 1) using the Illumina ASE assay (Figure 2). The experiment is similar to the one used for large-scale SNP genotyping [52] and gene expression profiling [53] except that DNA and RNA are independently assessed and compared to each other. RNA was first converted into biotinylated cDNA [53] while DNA was treated according to the usual GoldenGate assay protocol [52]. Biotinylated DNA (derived from genomic DNA or mRNA) was immobilized on paramagnetic beads and pooled SNP-specific oligonucleotides were annealed on the DNA. Hybridized oligonucleotides were then extended and ligated to generate DNA templates, which were amplified using universal fluorescently-labeled primers. Finally, single-stranded PCR products were hybridized to a Sentrix Array Matrix [52], and the arrays were imaged using the BeadArray Reader Scanner [54]. 96 samples (DNA or RNA) were analyzed per Sentrix Array for ~1,500 SNPs. All RNA measurements were performed in duplicates.

Analyses of ASE Results
To estimate the extent of allelic imbalance in heterozygote individuals at each SNP of the Illumina ASE panel, we developed algorithms using two different approaches: i) we used information from individuals of all three genotypes (AA, AB and BB), and/or ii) we used only the heterozygote individuals.
We first determined whether a given gene was expressed above a determined background in a given individual. To do so, we made use of the fact that the genotypes were known (from the DNA analysis) and developed a locus-specific expression background cut-off: homozygote individuals (i.e., AA or BB) can only express the corresponding allele, respectively A or B, at the RNA level (if at all). We thus determined a background fluorescence level (i.e., corresponding to random noise) for each allele (i.e., A and B) by measuring the emission in the corresponding dye (respectively, Cy3 and Cy5) in individuals homozygous for the other allele (respectively BB and AA). This is represented schematically on Figure S9. To avoid false positive results due to the inclusion of transcripts not expressed in the cell lines considered, we used a conservative approach and arbitrarily fixed the background emission cutoff to the maximum emission of the absent allele of all homozygotes, plus the mean emission of the absent allele divided by the number of homozygotes (to weight the uncertainty in the determination of the “maximum noise” by the numbers of individuals used to determine it). This procedure allowed us to independently estimate the background emission of each allele/dye specifically for each SNP, which is particularly important because the fluorescence emission can differ drastically between the dyes and among loci (data not shown). We then proceeded to the detection call using the background cut-offs: individuals with genotypes AA were considered to express a given transcript if the emission was larger than twice the cutoff background emission of A, individuals with genotypes AB if the fluorescence was larger than the sum of the background emission of A and the background emission of B, and individuals with genotypes BB if the emission is larger than twice the background emission of B. Since the inclusion in the analyses of transcripts expressed at low level (or not expressed at all) is very problematic, we excluded from our analyses all loci for which less than 75% of the individuals had discordant replicate expression (i.e., one replicate above expression background, the other under the cut-off value).

The first method used to determine whether some heterozygote individuals expressed significantly differently the two alleles is locus-specific but requires having at least one individual expressed from each homozygote genotype (AA and BB). In this case, we determined the median log ratio of the two dyes for each homozygote clusters at the DNA and RNA level \( \left( D_{AA}, R_{AA}, D_{BB}, R_{BB} \right) \) as well as the median absolute deviations (MAD). We used medians and MADs instead of means and standard deviations, to down weight the influence of possible outliers. We then determined a range of “expected” (i.e., non significant) variation of allelic expression for the heterozygote individuals. We calculated the equation of the lines joining the median values plus/minus two MAD of AA and BB and estimated the range, for the log ratio of the dyes at the RNA level, between the lines at the value corresponding to the median of DNA in heterozygote individuals (Figure S10). If the observed log ratio of dyes for a given heterozygote individual fell outside the expected range of variation in absence of AI (Figure S6), we scored each heterozygote individual separately to obtain a quantitative estimation of allelic imbalance using the ratio:

\[
\frac{(R_{AB} - E(R_{AB}))}{E(R_{AB})}, \text{ if } R_{AB} > E(R_{AB})
\]

or

\[
\frac{(R_{AB} - E(R_{AB}))}{E(R_{AB}) - R_{BB}}, \text{ if } R_{AB} < E(R_{AB}).
\]

This simple estimate indicates both the magnitude of the allelic imbalance (i.e. the fold difference) and its direction (i.e. which allele is more expressed than its counterpart).

In order to assess allelic imbalance for SNPs with low minor allele frequencies (for which homozygote individuals with the minor allele may not be present in a small sample size panel), we developed a second method based solely on the heterozygote individuals. If a given transcript is affected by allelic imbalance we expect that either the variance of the log ratio of dyes for heterozygote RNAs to be greatly increased relative to the variance of homozygote RNAs, or, if one allele is systematically more expressed than the other, the mean value of these log ratios to be drastically shifted from its expected intermediate position (between the mean for AA and the mean for BB homozygote RNAs). For all SNPs with at least five individuals with the same genotype expressed, we estimated the standard deviation of the log ratio of dyes for DNA and RNA. The distribution of the log ratio of the standard deviations (i.e., log \( \sigma_{DNA} / \sigma_{RNA} \)) over all loci for heterozygous individuals differed from those observed using homozygous individuals and did not seem to fit a normal distribution (Figure S11). Based on the assumption that this distribution may include some loci in allelic imbalance (and thus with a higher than expected RNA variance), we fitted a mixture of two Gaussians on our dataset (i.e., one corresponding to the loci with allelic imbalance, the second including all other loci) using a Maximum Expectation algorithm implemented in R (mixdist package). For our data, the best fit was obtained with a minor distribution (including ~3% of the loci) corresponding to the most extremely negative log ratios of variances (i.e., that the RNA standard deviation was larger than expected). For each locus, we then used the probability of belonging to the “higher-than-expected RNA variance” distribution as an indication of allelic imbalance.

Quantitative Sequencing of RT-PCR Products

We assessed the extent of allelic imbalance by quantitative sequencing following the method described in Ge et al. [33]. Briefly, we isolated RNA using TRIzol reagent following the manufacturer’s instructions. We assessed RNA quality with an Agilent 2100 Bioanalyzer (Agilent, Palo Alto, USA) before synthesizing first strand cDNA using random hexamers (Invitrogen, Burlington, Canada) and Superscript II reverse transcriptase (Invitrogen, Burlington, Canada). For each locus, we designed locus-specific primers, in the exon/UTR containing the SNP studied. 5 ng of genomic DNA and 10 ng of total cDNA were then amplified by PCR using Hot Start Taq Polymerase (Qiagen, Mississauga, Canada) with an activation step (95 °C for 15 minutes) followed by 40 cycles (95 °C for 30 s, 55 °C for 30 s and 72 °C for 45 s) and a final extension step (72 °C for 6 minutes). PCR products were purified using Exonuclease I and Shrimp Alkaline Phosphatase (USB, Cleveland, USA) and sequenced using either one of the former primers or a nested primer, on an Applied Biosystems 3730xl DNA analyzer. We used PeakPeaker v.2.0 [33] with the default settings to quantify the relative amount of the two alleles measured from the chromatogram after peak intensity normalization. To estimate the experimental variability of the entire experimental setup we used a hierarchical strategy for two genes (cf. Figure S12): for two/three individual cell lines, we extracted independently RNAs three times and performed, on each extract, three independent RT-PCRs. All cDNA obtained were then split into three aliquots, each amplified independently by locus-specific PCR. These PCR products were finally sequenced each three times (i.e. three independent sequencing reactions). To estimate the variability at each experimental stage we calculated the mean standard variation normalized to the mean using the independent
triplicates. To calculate the variance in the higher hierarchical levels (PCR, RT-PCR), we averaged the values from the lower level (e.g., to estimate the variability at the PCR level, we compared the means of the three sequencing values performed on each of the three PCR's: [s1, s2, s3] vs [s4, s5, s6] vs [s7, s8, s9]). The results are presented in Table 2.

**Association Mapping of Differential Allelic Expression**

We attempted to map allelic imbalance to regulatory haplotypes for all genes with significant differences in allelic expression that fulfilled these criteria: i) they are mapped on the build 34 of the human genome, ii) the SNP used in the Illumina ASE assay has also been genotyped by the HapMap [53] and iii) there are more than four HapMap individuals heterozygous at the marker SNP. For each gene, we retrieved the haplotype information from the phased chromosomes of each of the 57 HapMap CEPH individuals for 100,000 bp upstream and downstream of the SNP used to assess allelic imbalance. When a transcript contains more than one SNP or if two SNPs used to assess allelic imbalance at two transcripts are separated by less than 200,000 bp, the region obtained spans from the most upstream marker plus 100,000 bp to the most downstream marker minus 100,000 bp. For each individual LCL, the over expressed and under expressed haplotype/chromosome were identified and each SNP was tested for segregation of the alleles in under- and over-expressed chromosomes using a Fischer's exact test. Between 47 and 592 SNPs were tested for each gene (mean = 229) and the associations remaining significant after Bonferroni correction for multiple testing are shown in green in Table 2.

**Validation of Regulatory Haplotypes**

Illumina total gene expression data were obtained from the Wellcome Trust Sanger Institute for the 60 unrelated CEPH individuals genotyped by the HapMap project and included in our assay. We also determined the total expression for 10 genes using Real-Time PCR and SYBR Green labeling on an ABI 7900HT (Applied Biosystems, Foster City, CA) instrument. 8–10 ng of first strand cDNA were amplified using 0.32 μM of gene specific primers and Power SYBR Green PCR master mix (Applied Biosystems) according to the manufacturer’s instructions. The amplifications started by 95°C for 10 min followed by 40 cycles at 95°C for 20 s, 58°C for 30 s and 72°C for 45 s. We performed the Real-Time PCR assays for the 60 individuals LCLs genotyped by the HapMap projects and analyzed 6 replicates per each sample. A standard curve was established using a dilution series of total cDNA of known concentration. The Cg for each replicate was transformed to a relative concentration using the estimated standard curve function (SDS 2.1, Applied Biosystems) and normalized based on 18S rRNA Taqman (Applied Biosystems) expression data obtained for each sample to account for well to well variability.

**Software**

All analysis scripts are available upon request. PeakPicker v.2.0 is available at [http://www.genomequebec.mcgill.ca/EST-HapMap/](http://www.genomequebec.mcgill.ca/EST-HapMap/).

**Supporting Information**

**Text S1** Experimental variability using quantitative sequencing of RT-PCR products.

Found at: doi:10.1371/journal.pgen.1000006.s001 (0.03 MB DOC)
Table S1 List of the 2,968 SNPs analyzed using the Illumina ASE assay. Origins, Displays if the gene is located in a ENCODE region, on chromosome 21 or 22 and whether the genes was included for its potential involvement in disease etiology. Intron/exon. SNPs in 3′UTR are shown as “exon”.

Table S2 All SNPs expressed in at least three heterozygous individuals

References


