Human Molecular Genetics, 2007, Vol. 16, No. 16 ***1931–1939***
doi:10.1093/hmg/ddm140
Advance Access published on June 12, 2007

In vitro assays fail to predict *in vivo* effects of regulatory polymorphisms

Elizabeth T. Cirulli and David B. Goldstein*

Institute for Genome Sciences and Policy, Duke University, Durham, NC 27710, USA

Received May 8, 2007; Revised May 8, 2007; Accepted May 16, 2007

A typical paradigm in the investigation of complex human disease is to assess the effects of *cis*-regulatory polymorphisms implicated in association studies on transcription in cellular expression systems. Evidence from *in vitro* transfection studies is often assumed to be sufficient evidence for the *in vivo* functional importance of a polymorphism in the context of human disease, even though many confounding effects (e.g. temporal regulation, tissue specificity, genetic background) are not considered. In this study, we evaluate this assumption directly by examining the translation of *in vitro* results on allele-specific expression to an *in vivo* system using four genes that have been well documented through reporter assays to have promoter polymorphisms affecting transcription level: monoamine oxidase A (*MAOA*), neuropeptide Y (*NPY*), endothelial nitric oxide synthase (*NOS3*), and prodynorphin (*PDYN*). In our study, *MAOA* was found to have large allelic imbalances, which indicates that there is *in vivo* variation in the expression of this gene. However, the imbalances observed were not correlated with genotype at the putatively functional polymorphism. *PDYN, NOS3* and *NPY* did not have large allelic imbalances. Overall, there was no statistically significant effect of these polymorphisms on expression level as measured by imbalance ratios in any of these genes. These results suggest that the functional effects of a polymorphism on gene expression may be more complicated and context dependent than is often assumed and also imply that the use of cell-based expression studies to support the role of such polymorphisms in disease etiology should be treated with caution.

INTRODUCTION

There is a growing consensus that gene expression changes are likely to play an important role in human phenotypic variation and, in particular, in disease predisposition (1,2). Consistent with this, experimental evaluation of the effects of genetic variation on gene transcription plays an important role in the effort to dissect the genetic bases of human complex disease. However, there is limited information on how human genetic variation affects expression levels and how that translates into phenotypic variation.

In the investigation of complex disease, it is common to assess the effects of polymorphisms that are implicated in association studies on transcription in cellular expression systems (2). Most frequently, expression differences are assayed by transfecting each allele separately into a cell line. It is unknown, however, whether differential expression *in vitro* typically translates into differential expression *in vivo*. Cellular systems are also not useful for ascertaining temporal (i.e. developmentally regulated) and tissue-specific effects of polymorphisms. Despite these limitations, evidence from transfection studies is often accepted as sufficient evidence for the functional importance of a polymorphism in the context of human disease (2–4).

Allele-specific expression offers a method to assess these limitations by accurately comparing the expression levels of two putatively functional alleles in native tissue. Allele-specific expression has recently been used in a number of studies, providing evidence for *cis*-acting sites causing large expression differences in native tissue (5–9). In many cases, the variant that causes this expression difference has been validated, either through disease association studies (7) or by transfecting cell lines (6,8,9). However, little effort has been made to test for *in vivo* effects in the large number of genes that have shown evidence of an imbalance *in vitro*.

In this study, we use allele-specific expression to assess the applicability of genetic variants with experimentally determined *in vitro* expression effects in human tissue. Using

*To whom correspondence should be addressed at: Institute for Genome Sciences and Policy, Duke University, PO Box 3471, Durham, NC 27710, USA. Tel: +1 9196840896; Fax: +1 9196686787; Email: d.goldstein@duke.edu

human brain tissue, we assayed four genes [monoamine oxidase A (*MAOA*), neuropeptide Y (*NPY*), endothelial nitric oxide synthase (*NOS3* or *eNOS*), and prodynorphin (*PDYN*)] that have been determined *in vitro* to harbor common promoter variants that affect expression (10–20). We conclude that there is no correlation between the effects of polymorphisms on expression in cellular constructs and the effects on expression in native tissue, a point previously demonstrated in the gene *SLC6A4* by Wang and Sadée (8).

RESULTS

Study design

Our basic study design was to identify samples that were heterozygous for one or more polymorphisms that were present in the mRNA (referred to here as 'reporter polymorphisms') in order to identify samples for each gene in which allele-specific expression assays were possible. All such samples were then genotyped for the upstream variant that putatively affects expression level. The effect of this variant (here called 'functional variant') in the tissue studied was then inferred by comparing the allelic imbalances for samples with different genotypes at the functional variant. In particular, in the simplest case where the variant in question was the only determinant of expression, we would expect to see that the two homozygotes have no imbalance and that the heterozygotes (for the functional variant) would either have one or two imbalance levels, depending on the phase arrangement between the reporter and the functional variant in the sampled samples. More generally, an effect of the functional variant could be inferred by comparing the variability in imbalance scores amongst the samples of each genotypic class (Fig. 1).

Accuracy of methods

PDYN and *NPY* were subjected to pyrosequencing assays, while *NOS3* was analyzed using real-time PCR. To compare methods, *MAOA* was analyzed using both pyrosequencing and real-time PCR. For *MAOA*, there was a high correlation between pyrosequencing results and real-time PCR results ($P < 0.0001$, $r^2 = 0.9032$), although real-time PCR did tend to return higher ratios than pyrosequencing for this gene. Because of this inconsistency, only the pyrosequencing data were used to calculate the final ratios for *MAOA*. Both methods produced fairly consistent results: the average standard deviation for a sample for replicates across all runs was 0.07 for *NPY*, 0.10 for *PDYN*, 0.15 for *NOS3*, and 0.11 (pyrosequencing) and 0.12 (real-time PCR) for *MAOA*. In contrast, the total standard deviation of all composite allelic ratios across all tissues for all samples was 0.14 for *NPY*, 0.19 for *PDYN*, 0.17 for *NOS3*, and 1.05 (pyrosequencing) and 0.82 (real-time PCR) for *MAOA*. This demonstrates that while observed allelic ratios for a sample on a single run could vary by up to 30%, in the case of *MAOA* the variation between samples was much greater than the variation found within replicates of a sample. This makes it possible to distinguish between samples that truly have an allelic imbalance and those that do not.

An analysis of the apportionment of variance within runs and between runs, where each run involves a separate initial PCR mix, was also performed for each gene. It is often the case in studies of allele-specific expression that only one run is performed per sample, and it is assumed that the standard deviation for replicates within this run accurately reports the amount of variation one can expect for that sample. In contrast, we found that when comparing the amount of variation within runs to the amount of variation between runs, on average the amount of variation within runs only accounts for 70% of the total variation for *NPY,* 60% for *PDYN*, 58% for *NOS3* and 61% for *MAOA*. This demonstrates that there is nearly as much variation between runs as there is within runs, and multiple runs should be performed in the future to account for this extra source of variation.

For *MAOA* and *PDYN*, more than one reporter single nucleotide polymorphism (SNP) was used. A fundamental assumption of allele-specific expression analysis is that since allelic ratios are normalized by the gDNA ratio obtained in each sample, the imbalance ratio obtained in the cDNA should not be influenced by the identity of the reporter SNP. Indeed, this was the case in our sample, as the correlation between allelic ratios obtained using different reporter SNPs was high ($P = 0.0020$, $r^2 = 0.8180$). This correlation was calculated using a regression model with data from both genes.

Allelic ratios were obtained from cDNA from four different brain regions for each sample. Regression analysis was performed to compare the ratios obtained for each sample in each tissue (Supplementary material, Table S1). When there was a significant correlation between multiple tissues for a gene, only one tissue is presented. For *MAOA* and *NOS3*, only frontal cortex is presented, as allelic ratios from this tissue were highly correlated with those from all other tissues. For *NPY* and *PDYN*, allelic ratios from temporal cortex were highly correlated with those from frontal cortex and occipital cortex, but allelic ratios from cerebellum were very different. For this reason, both temporal cortex and cerebellum are presented for these genes.

MAOA

There are four alleles of the functional *MAOA-uVNTR* polymorphism: 3, 3.5, 4 and 5 (10,11). Alleles 3.5 and 4 show higher expression than allele 3 in transfected cell lines (10,11,14), while allele 5 has been reported to have either high (10) or low (11) expression. The high linkage disequilibrium between the functional polymorphism and our reporter SNPs (*rs6323* and *rs1800464*, both T/G) allows us to infer that 97–99% of the time 3/4 heterozygotes should have a T/G ratio between 2 and 10 (Fig. 1). Instead, ratios seem to be a continuous trait that extends from 0.26 to 5.10 (Fig. 2). Many 3/4 heterozygotes have ratios very close to one. There are some samples that show large allelic imbalances, however most of these have ratios of far less than one. Overall, the results presented are inconsistent with results from studies of transfected cell lines, where the 4 allele was more highly expressed than the 3 allele (10,11). There was also no significant difference between ratios found in functional homozygotes and those found in heterozygotes according to the regression analysis (Supplementary Material, Table S2).

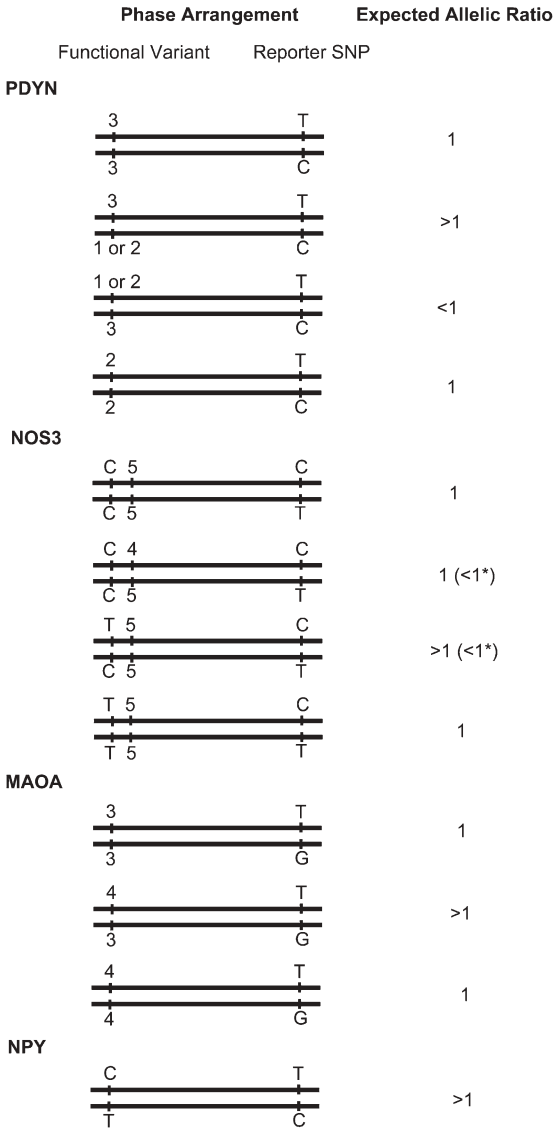

Figure 1. Phase arrangements and allelic imbalance expectations. *PDYN*: the reporter SNP was used to calculate the allelic imbalance as a ratio T to C. At the functional locus, the 3 allele is expected to show higher expression than the 1 or 2 allele. When the sample is homozygous for the functional variant (a 3/3 or a 2/2), no imbalance is expected. For samples that are heterozygous at the functional variant, in the case where the T allele is paired with the 3 allele, a ratio larger than one is expected. When the T allele is paired with the 1 or 2 allele, a ratio of less than one is expected. Due to low linkage disequilibrium, phase was not able to be determined for this gene. However, even in the case of phase uncertainty, an effect of the functional variant should still be discernable due to the greater variability in imbalance in heterozygotes than in homozygotes. *NOS3*: the reporter SNP was used to calculate the imbalance as a ratio of C to T. For the promoter variant, the T allele is expected to show higher expression than the C allele. As in *PDYN*, when the sample is homozygous for the promoter variant, no imbalance is expected. High linkage disequilibrium in this gene allowed us to infer that 96.5% of double heterozygotes would have the phase promoter T with reporter C, promoter C with reporter T, resulting in an expected ratio greater than 1. *If one takes the intronic *VNTR* into account, there are now two types of double heterozygote (phase assigned with >96% certainty): one has the phase promoter C with *VNTR* 4 and reporter C, promoter C with *VNTR* 5 and reporter T, while the other has the phase promoter T with *VNTR* 5 and reporter C, promoter C with *VNTR* 5 and repoter T. In this case, both heterozygotes are expected to have a ratio of less than 1. *MAOA*: the reporter SNP was used to calculate the allelic imbalance as a ratio T to G. At the functional locus, the 4 allele is expected to show higher expression than the 3 allele. As in *PDYN*, when the sample is homozygous for the functional variant (3/3 or 4/4), no imbalance is expected. High linkage disequilibrium in this gene allowed us to infer that 97–99% of double heterozygotes would have the phase 4 with T, 3 with G. Therefore all 3/4 heterozygotes assayed in this study are expected to have an allelic ratio greater than 1. *NPY*: the reporter SNP was used to calculate the allelic imbalance as a ratio T to C. At the functional locus, the C allele is expected to show higher expression than the T allele. Due to high linkage disequilibrium, all samples in our study that were heterozygous for the reporter SNP were also heterozygous for the functional SNP, and all were in the same phase. Therefore, all samples studied are expected to have an allelic ratio greater than one.

Balciuniene *et al*. (21) found that in autopsied cortical brain tissue, *MAOA* expression level was correlated not with the *MAOA-uVNTR* but rather with a haplotype formed by this *VNTR* and the genotype at a SNP in the neighboring gene *MAOB* (*rs1799836*). Haplotype 3G was found to have lower expression than other haplotypes in their study. We also genotyped this SNP to see if there was an association between the 3G haplotype and lower expression. We found no such association (data not shown).

There was one 3/5 heterozygote in our sample set, which exhibited a large expression imbalance between the two alleles (T/G ratio of 0.29 in the frontal cortex). Although the phase of this heterozygote is uncertain, the large imbalance is contrary to the results found by Sabol *et al*. (11), which report no expression difference between alleles 3 and 5.

Our results indicate that some samples that are heterozygous for alleles 3 and 4 at the *MAOA-uVNTR* have an expression imbalance; however, this imbalance is a continuous trait and overall shows no statistical dependency on the *VNTR* polymorphism (Supplementary Material, Table S2).

NPY

We examined 22 samples that were heterozygous for the functional SNP, *rs16147* (12,15). Because of the high linkage disequilibrium between this functional SNP and our pyrosequencing reporter SNP (*rs9785023*), all of the double heterozygotes examined are expected to exhibit an allelic ratio between 2.5 and 5 (Fig. 1). Instead, the ratios observed were very close to 1 and formed a continuous distribution in temporal cortex from 0.74 to 1.28, with no bias towards higher ratios observed (Fig. 3). Two samples, C673 and C591, did seem to exhibit a true imbalance, with ratios of 0.32 and 0.54. This was true only in the cerebellum, however, which had very low amplification and was therefore more prone to variable and questionable results.

Since we only analyzed samples that were heterozygous for the functional variant, it was not possible to use a regression model to interpret its effect. However, the pattern of imbalance ratios exhibited in our study is inconsistent with results from transfection studies.

NOS3

The T allele of the promoter SNP *rs2070744* has been shown in reporter assays to increase transcription levels 2-fold over

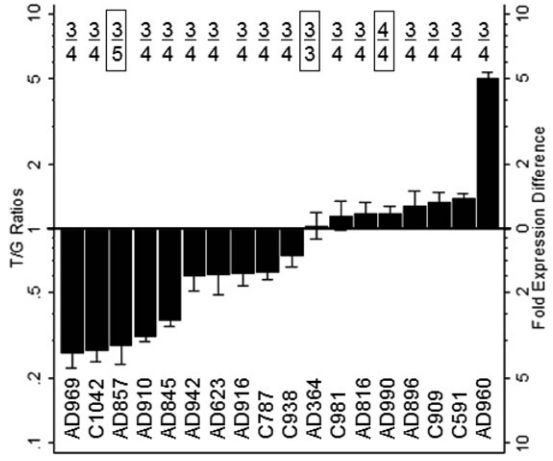

Figure 2. Allelic ratios for *MAOA* in frontal cortex. There was a high correlation between results obtained from each tissue, so only frontal cortex is shown. All samples are 3/4 heterozygotes except those marked otherwise. The homozygotes for the *VNTR* are not expected to have allelic imbalances dependent on this variant, and indeed do not show a significant allelic imbalance. All 3/4 heterozygotes are expected to have ratios well above one. Instead, allelic ratios for these heterozygotes present as a continuous trait ranging from 0.25 to 5.1. Phase is not known for the 3/5 heterozygote; however, it is clear that one of the alleles is being expressed at a much higher level than the other.

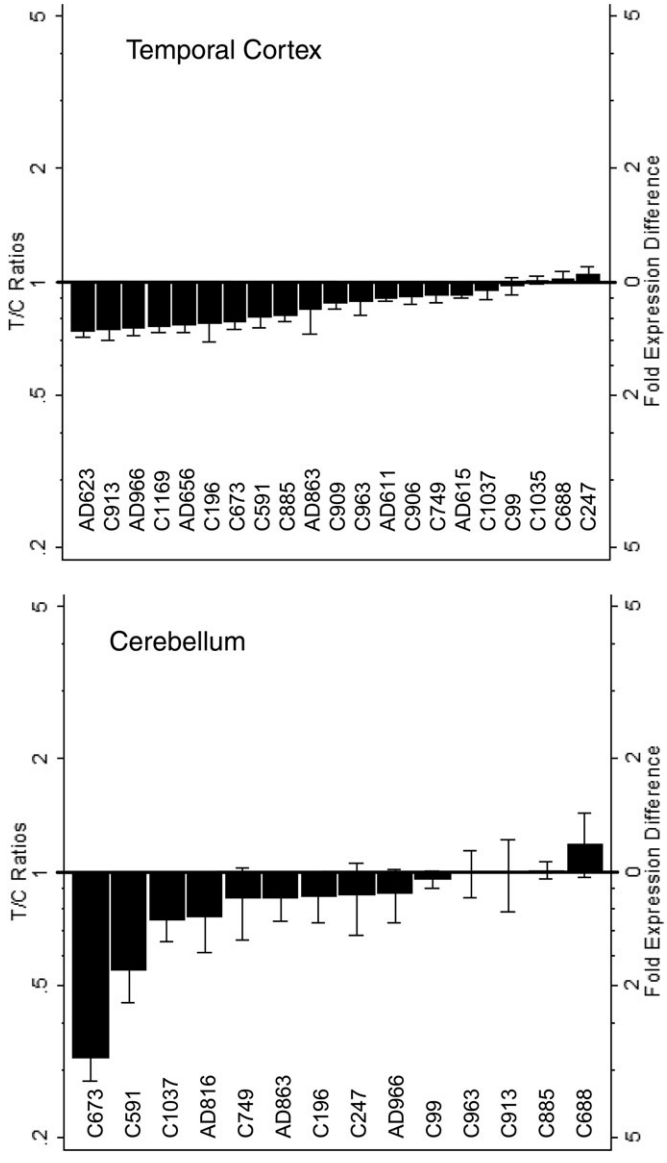

Figure 3. Allelic ratios for *NPY* in temporal cortex and cerebellum. Ratios obtained in temporal cortex were highly correlated with those obtained from frontal and occipital cortex, while cerebellum was not. All samples are heterozygous for both the functional and reporter SNPs. Allelic ratios present as a continuous trait.

the C allele (16–19). We utilized a T/C reporter SNP (*rs1549758*) to assess allelic imbalances. The high linkage disequilibrium between the two SNPs allowed us to infer that 96.5% of the time a sample heterozygous for the functional variant should have a C/T ratio from the reporter SNP of about 2 (Fig. 1). Instead, most heterozygotes have ratios of less than 1, and allelic imbalance in general presents as a continuous trait without regard to genotype of the promoter SNP (Fig. 4).

Wang *et al*. (20) use reporter assays to demonstrate that transcription level is not controlled solely by the promoter SNP, but also by a *VNTR* in intron 4. They show that the combination of the C allele for the promoter SNP and five repeats for the *VNTR* results in the highest level of expression, whereas the combination of the T allele for the promoter SNP and five repeats for the *VNTR* results in the lowest level of expression. The combination of either the T or C allele with four repeats results in intermediate levels of expression. We sampled two types of heterozygotes in our study: those homozygous for the C allele of the promoter variant but heterozygous for the *VNTR*, and those homozygous for the five repeat allele of the *VNTR* but heterozygous for the promoter variant. High linkage disequilibrium allowed us to infer that all of our C4/C5 samples should have reporter C/T ratios of about 0.7, and 96.5% of our T5/C5 samples should have reporter C/T ratios of about 0.25 (Fig. 1). Both sets of heterozygotes do tend to have ratios below 1 in our dataset; however, most of our samples that are homozygous for both variants also have ratios below 1 (Fig. 4). Regression analysis shows no effect of either variant, taken separately or considered together, on allelic imbalance (Supplementary Material, Table S2).

There were two samples with unique haplotypes that could not be analyzed in the regression. The phase arrangements for these samples were C4/T5 (C750) and T4/T6 (C591). A ratio greater than 1 is expected for C750 according to *in vitro* evidence from Wang *et al*. (20), while there is no prior expectation of imbalance level for C591 since the six-repeat allele of the *VNTR* has not been previously studied. Both samples show ratios lower than 1, although convincingly so only in C591.

The non-synonymous coding SNP *rs1799983*, which is often reported to be associated with disease (reviewed in 22,23), was also examined, both by itself and as part of a haplotype with the promoter SNP and *VNTR*. It was not found to affect imbalance levels (data not shown).

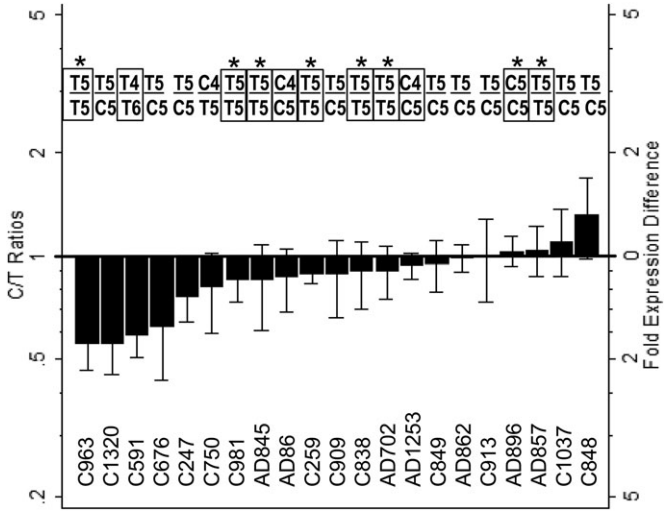

Figure 4. Allelic ratios for *NOS3* in frontal cortex. Ratios obtained in frontal cortex were highly correlated with those obtained from all other tissues. Shown above are the genotypes for the promoter SNP and intronic *VNTR*. All samples homozygous for the promoter SNP are boxed: based on *in vitro* evidence for this promoter SNP, no imbalances are expected in these samples. Instead, homozygous samples show a wide range of imbalance ratios. Wang *et al*. (20) showed that *in vitro* transcription for this gene is influenced by both the promoter SNP and a *VNTR* in intron 4. Those samples homozygous for both variants are marked with asterisks. Again, based on the *in vitro* evidence no imbalances are expected for these samples, yet a range of imbalances occur.

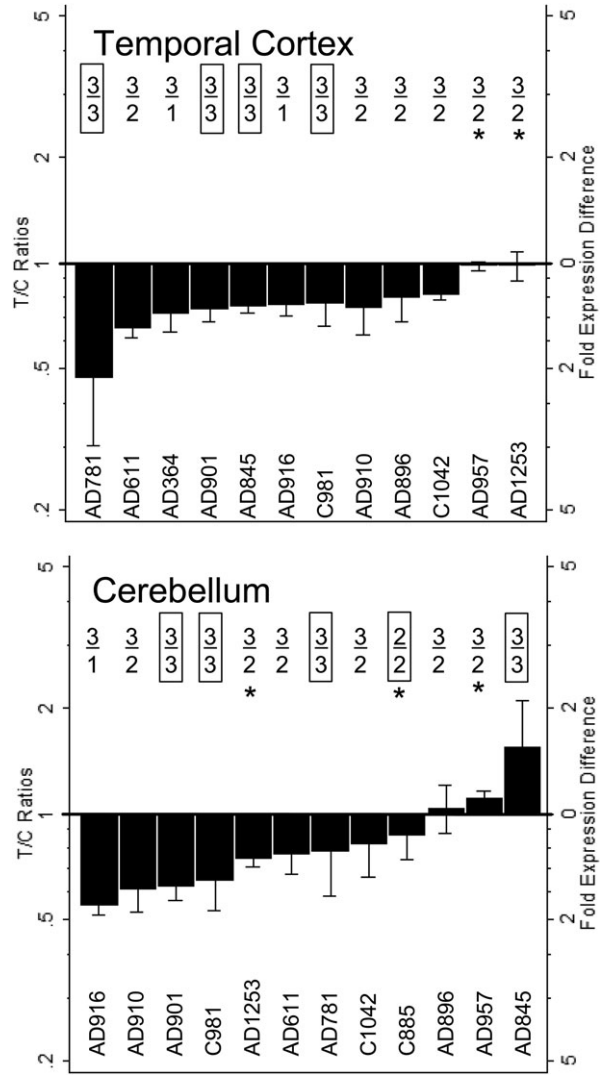

Figure 5. Allelic ratios for *PDYN* in temporal cortex and cerebellum. Ratios obtained in temporal cortex were highly correlated with those obtained from frontal and occipital cortex, while cerebellum was not. While most samples have an imbalance in favor of the C allele, this pattern is independent of genotype at the reportedly functional *VNTR*. Those samples marked with an asterisk are homozygous for *rs6045733*. Heterozygotes for this SNP had larger imbalances than homozygotes for all tissues except cerebellum.

PDYN

There is a functional promoter *VNTR* for *PDYN* which has four alleles: 1,2,3, and 4, the last two demonstrating 150% the expression of the first two in reporter assays (13). We utilized two reporter SNPs *(rs910080*, *rs6045819*, both T/C) to analyze allele-specific expression. There was low linkage disequilibrium between these SNPs and the *VNTR* variants, so phase was not able to be assigned. T/C ratios were therefore expected to be either 1.5 (T with 3) or 0.67 (C with 3) (Fig. 1). In our results (Fig. 5), most of the samples had an allelic ratio near 0.67. However, this imbalance was present in samples that were both homozygous and heterozygous for the *VNTR*, and there was no statistically significant correlation between imbalance ratio and genotype at the functional variant according to the regression analysis (Supplementary Material, Table S2).

Genetic determinants of allelic imbalance

Our analyses suggest that in each case studied above the putative *cis*-acting regulatory variant has little or no impact on gene expression in the studied tissues. This still however leaves open the possibility that other *cis*-acting sites may be affecting imbalance levels. To check for whether common polymorphisms present in our samples may affect imbalance we used a tagging strategy. The studied samples were each genotyped using the Illumina HumanHap550 Genotyping BeadChip. Since we were looking for *cis*-acting sites, we restricted attention in each case to only those SNPs that are in or within 10 kb of each of the genes studied. SNPs that met these criteria were examined for association with imbalance level (Supplementary Material, Table S4). No SNPs were suggestive of influencing imbalance in *MAOA, NPY* or *NOS3*. For *PDYN*, one SNP, *rs6045733*, seemed to be correlated with allelic imbalance (permuted $P = 0.03530$). Heterozygotes for this SNP had consistently lower ratios than homozygotes in temporal cortex, with similar trends in frontal and occipital cortex (Fig. 5). The trend was not seen in cerebellum. However, the significance of this SNP does not stand up to correction for multiple testing, and there were only three homozygotes in our study.

DISCUSSION

Expression analysis is an increasing focus in studies of both normal variation and genetic contribution to disease.

There has been little empirical evaluation, however, of how well cell-based expression assays reflect expression patterns in human tissue relevant to the disease or phenotype under study. At one extreme, the effect of regulatory variants may be general, applying to most tissues, most genetic backgrounds and most environments. At the other extreme, the effects of variants may be highly dependent on developmental stage, tissue type, genetic background and the precise environmental context. Where reality sits in this spectrum is currently almost entirely unknown.

In this study, we examined four variants that had been previously shown to affect expression levels in transfected cells. Three of these variants are extremely well documented as affecting transcription *in vitro*, the results having been replicated in multiple cell line types by multiple groups (10–12,14–20). We found varying levels of allelic imbalances for these genes; however, not one of the genes showed an imbalance pattern consistent with previous findings. There is the possibility that the presence of multiple *cis*-acting sites masked the effect of the variants we studied; however, it is clear that the functional variants studied are not the major contributors to allelic imbalance in the brain regions examined. In fact, for all genes analyzed, there was no consistent or significant effect of the 'functional variant' on expression patterns at all (Table 1). The imbalance results we present are not artifacts, however, since imbalance levels consistently replicate within our sample set and it is well known that true allelic imbalances can be accurately detected when they exist (24,25). Our results therefore imply that it is important to perform allele-specific expression in native tissue when measuring the effect of a variant on gene expression level. If the expression difference that is found in transfected cell lines is not found in the tissue of interest, or is not found in a manner that coincides with the hypothesis of how this variant affects the phenotype, then the evidence from the transfected cell lines is close to meaningless. It is vital to keep context in mind when interpreting expression data. For *NPY, NOS3* and *PDYN*, there were indications of modest imbalances for some samples; however, there was no evidence for any effect of the functional variant on imbalance ratio.

In the case of *MAOA*, there does appear to be a *cis*-acting site that causes a large difference in expression between the two alleles of some samples. However, this is true only for some samples heterozygous for the promoter *VNTR*, and the direction of the effect is inconsistent. Most of our samples that displayed a large imbalance displayed it in a direction contrary to that expected from transfection studies, while in general the expression differences presented as a continuous trait. This is consistent with other studies of allele-specific expression in native tissue, which have also found gene expression differences to present as a continuous trait (6–9). This suggests that *cis*-acting sites may be very sensitive to modifiers, with multiple environmental or *trans*-acting sites contributing to the continuous distribution of imbalances.

Pinsonneault *et al.* (27) also looked for allelic imbalances in *MAOA* in human brain tissue. They too found that imbalances occurred but were independent of genotype at the promoter *VNTR*. They genotyped several SNPs in *MAOA* and found an association between allelic imbalance and *rs6323, rs2205718, rs979606, rs979605, rs1801291, rs3027407* and *rs909525*. *Rs6323* was our reporter SNP and showed a wide range of imbalances, so it is not the causal variant. We genotyped *rs1801291, rs3027407* and *rs909525*, and also *rs2235186*, which has an r^2 of 1 with *rs2205718, rs979606* and *rs979605*. None of these SNPS were correlated with allelic imbalance in our sample. Although there is weighty evidence for a *cis*-acting site affecting expression in *MAOA*, it appears that the causal variant has yet to be identified.

The potentially confounding effect of X-inactivation must be considered when analyzing the imbalances observed in MAOA. Since allele-specific expression of *MAOA* can only be studied in females, random X-inactivation could create an additional source of uncertainty and variability. However, it has previously been shown that MAOA is one of the many X-chromosome genes that escapes inactivation (26). It has also been shown that allelic imbalances in MAOA are not correlated with which X is inactivated (27).

Although we genotyped tagging SNPs in each of the four genes to attempt to locate the variant causing imbalance, no convincing variant was found. There was no correlation between any of the SNPs genotyped and allelic imbalance for *MAOA, NPY* or *NOS3*. For *PDYN*, one SNP, *rs6045733*, was found to be associated with imbalance level in cortical tissue but not cerebellum. It is difficult to tell if this SNP is truly influencing transcription, however, since there were only three homozygous samples in our study. It is also difficult to postulate how this SNP would affect transcription levels, as it is positioned 8.5 kb after the gene ends and is not in the vicinity of any alternative transcripts. Finally, we should emphasize that this analysis was based on a tagging panel of SNPs and would not be expected to well represent rare variants present in our sample. It is therefore likely that there are variants present that affect imbalance patterns that have not been detected.

Taken together these results imply that regulatory polymorphisms are generally highly complex and dependent on unknown modifiers. Transfection studies can be useful in adding further evidence to an imbalance already observed *in vivo*, but a transfection study alone should not be considered evidence for a functional polymorphism in reference to a

Table 1. Expectations and results for each control gene

Gene	Expected imbalance (from previous studies)	Evidence of imbalance?	Consistent with expectations?	Any effect of the functional variant?
MAOA	2–10× (10,11,14)	Strong	No	No
NPY	2.5–5× (12,15)	Moderate	No	No
NOS3	2× (16–20)	Moderate	No	No
PDYN	1.5× (13)	Moderate	No	No

given phenotype, since there seems to be no correlation between expression level changes observed in cell lines and those present in native tissue. This implies that if we are to understand how polymorphisms affect phenotypes of interest, then a framework will be needed for assessing the affects of polymorphisms in native tissue, with attention given to the possibility of tissue specificity of effects as well as modification by the genetic and environmental background. This argues for the development of large-scale mapping efforts that use allele-specific expression and other assays to map the effects of polymorphisms *in vivo* and to compare their effects across tissue types and developmental stages. After variants that affect allelic imbalance *in vivo* have been found, *in vitro* approaches can be utilized to understand the causes of the imbalance. Some promising starts have been made, as some recent studies have begun to use allele-specific expression to identify functional variants (6–9). This is an important step in making the connection between phenotype and genotype. Instead of using transfection experiments as a follow-up to association results, studies of allele-specific expression in multiple tissues should be extended to these genome wide searches. This will be a difficult challenge, but it must be seen as a necessary component to association studies on end phenotypes.

MATERIALS AND METHODS

Samples

Post-mortem tissue samples were collected from 32 patients that had been diagnosed with Alzheimers Disease (AD) and 31 controls with no diagnosis of dementia. All patients were recruited to a study run by the ADRC and consented to the use of their brain tissue for research (IRB No. 0182-03-5R3). Samples were obtained from four brain regions for each individual: temporal cortex, frontal cortex, occipital cortex and cerebellum. Brain tissue was obtained through the Kathleen Price Bryan Alzheimer's Disease Brain Bank at Duke University. Control subjects ($n = 31$) were between the ages of 56 and 90 (mean 85.5 yrs) and consisted of 14 males and 17 females. AD subjects ($n = 32$) were between the ages of 79 and 90 (mean 83.5 years) and consisted of 8 males and 24 females. The samples were not categorized as cases or controls for this study, however, as the genes we studied were not thought to affect or be affected by AD. Brain tissue was flash frozen in liquid nitrogen and stored at -80°C until extraction. DNA was extracted using the Puregene Kit® (Gentra) according to the manufacturer's instructions. Total RNA was extracted using the RNeasy® Lipid Tissue Purification Kit (Qiagen) according to the manufacturer's instructions. Concentrations were checked using a Molecular Devices M2 spectrophotometer. cDNA was produced from 5 μg total RNA in a 100 μl reaction using the High Capacity cDNA Synthesis Kit (Applied Biosystems) according to the manufacturer's instructions.

Genotyping of samples

For the *MAOA-uVNTR* (variable number of tandem repeats), the *NOS3 VNTR*, the *NOS3* promoter SNP *rs2070744* and the *PDYN VNTR*, PCR was performed in 16 μl reactions using 30 ng of gDNA. The reaction conditions for *PDYN* were as stated in Zimprich *et al*. (13) with the addition of 0.5 μl DMSO per reaction. The reaction for the *MAOA-uVNTR* and the *NOS3 VNTR* and SNP consisted of 1× PCR Buffer II (Roche), deoxynucleoside triphosphates each at 0.2 mM, 0.025 units/μl of Amplitaq Gold® (Roche), 2 mM $MgCl_2$, 5% DMSO, and 0.5 μM forward and reverse primer. The following touchdown PCR was used: 5 min at 94ºC, then 8 × (30 s at 94ºC, 30 s at 64ºC–56ºC (decreasing by one degree each cycle), 30 s at 72ºC), then 25 × (30 s at 94ºC, 30 s at 56ºC, 30 s at 72ºC). The PCR products for all *VNTRs* were visualized on 4% agarose gels. The PCR product for *rs2070744* was visualized using a 3730 DNA Analyzer (Applied Biosystems). For *NPY-rs16147, NOS3-rs1799983* (a common nonsynonymous coding SNP) and *MAOB-rs1799836* (used to determine haplotype in *MAOA*), PCR was performed in 5 μl reactions using 1× TaqMan® (Applied Biosystems) universal PCR master mix, 1× assay and 30 ng of DNA, using the following conditions: 10 min at 95°C followed by 40 × (15 s at 92°C and 1 min at 60°C). Results were then obtained using a 7900 HT Fast real-time PCR System.

SNPs with high heterozygote frequencies located in the mRNA of *MAOA*, *NPY, NOS3*, and *PDYN* were chosen for use as reporter SNPs. The SNPs utilized were *rs6323* (*MAOA*, G/T, exon 8), *rs1800464*(*MAOA*, G/T, exon 4), *rs9785023*(*NPY*, C/T, exon 2), *rs1549758*(*NOS3*, C/T, exon 6), *rs910080* (*PDYN*, C/T, 3′-UTR) and *rs6045819* (*PDYN*, C/T, exon 4). For genotyping of the SNPs in *MAOA, NPY* and *PDYN*, 2 μl of DNA were used in a 50 μl PCR. The conditions for this PCR were as described above for genotyping of the *MAOA-uVNTR*, except that the amplification time used was 20 s instead of 30. The product was then analyzed using the pyrosequencer, a PSQ 96 MA. For genotyping of the SNP in *NOS3*, PCR was performed using a TaqMan® (Applied Biosystems) assay, as described above. Samples that were heterozygous for the reporter SNP were then analyzed using allele-specific expression. The genotypes at the functional loci for these samples are shown in Figures 2–5. Primer sequences are available in Supplementary Material, Table S3.

Allele-specific expression using pyrosequencing

Allele-specific expression was performed for *MAOA, PDYN* and *NPY* with the same procedures listed for genotyping reporter variants: PCR was run in a 50 μl reaction and the results, measured as peak heights, were obtained from a pyrosequencer. For each sample, a run consisted of three to five replicates (separate reactions, same PCR mix) of gDNA and three to five replicates (separate reactions, same PCR mix) of cDNA from each tissue. Each replicate was analyzed separately, and for each sample, between one and seven such runs (different PCR mix) were performed. Breaking replicates up in this manner allowed us to determine how sensitive pyrosequencing was to different sources of variation.

Allele-specific expression using real-time PCR

For *MAOA-rs6323* and *NOS3- rs1549758*, allele-specific expression was performed using a TaqMan®-based quantitative real-time PCR assay. Real-time PCR was performed in a 10 μl reaction containing 0.4 μl cDNA or gDNA, 1× universal primer mix and 1× assay. For each sample, two replicates of gDNA and two replicates of cDNA from each tissue were performed for each run. For each sample, two runs were performed. The results from the real-time PCR, measured in Ct values, were obtained using a 7900 HT Fast Real-Time PCR System. A standard curve was calculated from the relationship between $\log_{10}$(ng input DNA) and resultant Ct values, with separate curves for Vic and Fam. The amount of each allele present in each sample was calculated using the equation $10^{((\text{Ct-intercept})/\text{slope})}$, where intercept and slope are from the standard curve.

Data processing

The raw data obtained from pyrosequencing and real-time PCR were controlled for quality before being analyzed. For pyrosequencing, if the peak heights for both alleles of the SNP were less than 4, the data were not used, except when all of the peak heights for a sample were low (this was the minority of cases), in which case the data were analyzed by hand to determine validity. For real-time PCR, if both alleles had a Ct value outside the range of the standard curve, the data were not used.

To determine allelic ratios, the score (peak height for pyrosequencing, nanograms of cDNA or gDNA for real-time PCR) of one allele of the SNP was divided by the score of the other allele. For each run, the average allelic ratio of the gDNA for a sample was then determined. All of this sample's cDNA allelic ratios from this run were divided by the sample's average gDNA ratio, to normalize them. Normalizing the cDNA ratios by the gDNA ratios negates the effect of any biased amplification. The average standard deviation for replicates within each run was 0.07 (relative to the expected ratio of 1 when there is no imbalance) for *MAOA* and *NPY*, 0.08 for *PDYN* and 0.10 for *NOS3*, so as a measure of quality control, runs with a standard deviation greater than 0.5 were discarded. The average of all normalized cDNA ratios from all replicates across all valid runs were averaged to create one composite ratio per tissue for each sample.

Phase calculation

The probability of potential phase arrangements between the functional and reporter variant for each sample was calculated using the program PHASE, utilizing the genotypes of all 63 of our samples at all loci (28,29). For *MAOA* and *PDYN*, a second reporter SNP was also used, as a further check for sources of variation. Phase probabilities from PHASE were used to assure that the ratios obtained from these second reporter SNPs would be in the same orientation as those ratios obtained from the original reporter SNPs.

Statistical analysis

We developed a regression model to test whether the functional variant had any impact on imbalance ratios for each gene. The ratios obtained for each sample were assessed by the regression model $\ln(\text{ratio}) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$, where x_1 was the probability that phase was in configuration 1 and x_2 was the probability that phase was in configuration 2 (see Fig. 1 for examples). Samples that were homozygous for the functional SNP had a zero entered for both x_1 and x_2. The null hypothesis for this model was that the imbalance ratio obtained in homozygotes was equal to the ratios obtained in the two possible phases of heterozygotes, indicating no effect of the functional variant. The alternative hypotheses were that either one phase or the other of heterozygotes had an allelic ratio that was not equal to that found in the homozygotes. This regression model was analyzed for *MAOA*, *PDYN* and *NOS3*. We could not use this regression model for *NPY*, as all samples were heterozygous for the functional allele. For *NOS3*, a second regression model was also utilized, one that took genotype at both the promoter variant and the intronic *VNTR* into account. For this regression, there were three configurations of heterozygotes possible, resulting in the regression model $\ln(\text{ratio}) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$. Because of our relatively small sample size, the model was analyzed using 10 000 permutations in a multiple linear regression program written by Pierre Legendre [(30), available at http://www.bio.umontreal.ca/Casgrain/en/labo/regression.html].

All other regression analyses and ANOVAs were performed using STATA 9.2 (31) using the natural log of the allelic ratio. Regressions were run to look for correlations between allelic ratios obtained from different tissues, correlations between different reporter SNPs within a gene and correlations between pyrosequencing and real-time PCR results. To determine apportionment of variance within runs and between runs, an ANOVA was first performed using STATA, with all replicates of a sample categorized by which run they were from. Variance components were calculated as in Sokal and Rohlf (32). Each tissue of each sample received its own estimate for the percent of total variation that could be attributed to variation within runs and variation between runs. All percentages were then used to obtain the average percent of total variation that was accounted for by variation within runs for each gene.

Finding genetic determinants of allelic imbalance

Each sample was genotyped using the Illumina HumanHap550 BeadChip according to the manufacturer's instructions. Genotypes for all SNPs within 10 kb of each gene studied were extracted. Each SNP was analyzed using STATA 9.2 (31). A *t*-test was used to determine if there was a difference between the mean ratio of the homozygotes and the mean ratio of the heterozygotes. In case phase was switched for some heterozygotes, causing large ratios in one type and small ratios in the other, a test of equal variance was also performed. Any SNPs that looked suggestive by either of these two tests were then assigned phases using PHASE (28,29) and tested by regression, as described above.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at HMG Online.

ACKNOWLEDGEMENTS

We thank E. Heinzen for assistance with brain samples and comments on the manuscript, L. Goering for assistance with pyrosequencing, G. Gibson for the use of lab space and equipment, A. Need for assistance in gene selection, N. Walley for many very helpful comments on the manuscript, M. Weale and V. Kondragunta for assistance with statistical analyses, and IGSP for funding.

Conflict of Interest statement. None declared.

REFERENCES

1. Khaitovich, P., Enard, W., Lachmann, M. and Paabo, S. (2006) Evolution of primate gene expression. *Nat. Rev. Genet.*, **7**, 693–702.
2. Pastinen, T. and Hudson, T.J. (2004) *Cis*-acting regulatory variation in the human genome. *Science*, **306**, 647–650.
3. Chen, J.M., Ferec, C. and Cooper, D.N. (2006) A systematic analysis of disease-associated variants in the 3′ regulatory regions of human protein-coding genes II: the importance of mRNA secondary structure in assessing the functionality of 3′-UTR variants. *Hum. Genet.*, **120**, 301–333.
4. D'Souza, U.M. and Craig, I.W. (2006) Functional polymorphisms in dopamine and serotonin pathway genes. *Hum. Mutat.*, **27**, 1–13.
5. Bray, N.J., Buckland, P.R., Owen, M.J. and O'Donovan, M.C. (2003) *Cis*-acting variation in the expression of a high proportion of genes in human brain. *Hum. Genet.*, **113**, 149–153.
6. Zhang, Y., Wang, D., Johnson, A.D., Papp, A.C. and Sadee, W. (2005) Allelic expression imbalance of human mu opioid receptor (OPRM1) caused by variant A118G. *J. Biol. Chem.*, **280**, 32618–32624.
7. Li, Y., Grupe, A., Rowland, C., Nowotny, P., Kauwe, J.S., Smemo, S., Hinrichs, A., Tacey, K., Toombs, T.A., Kwok, S. *et al.* (2006) DAPK1 variants are associated with Alzheimer's disease and allele-specific expression. *Hum. Mol. Genet.*, **15**, 2560–2568.
8. Wang, D. and Sadée, W. (2006) Searching for polymorphisms that affect gene expression and mRNA processing: example ABCB1 (MDR1). *AAPS J.*, **8**, 515–520.
9. Tao, H., Cox, D.R. and Frazer, K.A. (2006) Allele-specific KRT1 expression is a complex trait. *PLoS Genet.*, **2**, e93.
10. Deckert, J., Catalano, M., Syagailo, Y.V., Bosi, M., Okladnova, O., Di Bella, D., Nothen, M.M., Maffei, P., Franke, P., Fritze, J. *et al.* (1999) Excess of high activity monoamine oxidase A gene promoter alleles in female patients with panic disorder. *Hum. Mol. Genet.*, **8**, 621–624.
11. Sabol, S.Z., Hu, S. and Hamer, D. (1998) A functional polymorphism in the monoamine oxidase A gene promoter. *Hum. Genet.*, **103**, 273–279.
12. Buckland, P.R., Hoogendoorn, B., Guy, C.A., Coleman, S.L., Smith, S.K., Buxbaum, J.D., Haroutunian, V. and O'Donovan, M.C. (2004) A high proportion of polymorphisms in the promoters of brain expressed genes influences transcriptional activity. *Biochim. Biophys. Acta*, **1690**, 238–249.
13. Zimprich, A., Kraus, J., Woltje, M., Mayer, P., Rauch, E. and Hollt, V. (2000) An allelic variation in the human prodynorphin gene promoter alters stimulus-induced expression. *J. Neurochem.*, **74**, 472–477.
14. Inoue-Murayama, M., Mishima, N., Hayasaka, I., Ito, S. and Murayama, Y. (2006) Divergence of ape and human monoamine oxidase A gene promoters: comparative analysis of polymorphisms, tandem repeat structures and transcriptional activities on reporter gene expression. *Neurosci. Lett.*, **405**, 207–211.
15. Itokawa, M., Arai, M., Kato, S., Ogata, Y., Furukawa, A., Haga, S., Ujike, H., Sora, I., Ikeda, K. and Yoshikawa, T. (2003) Association between a novel polymorphism in the promoter region of the neuropeptide Y gene and schizophrenia in humans. *Neurosci. Lett.*, **347**, 202–204.
16. Abe, K., Nakayama, M., Yoshimura, M., Nakamura, S., Ito, T., Yamamuro, M., Sakamoto, T., Miyamoto, Y., Yoshimasa, Y., Saito, Y. *et al.* (2005) Increase in the transcriptional activity of the endothelial nitric oxide synthase gene with fluvastatin: a relation with the -786T>C polymorphism. *Pharmacogenet. Genomics*, **15**, 329–336.
17. Cattaruzza, M., Guzik, T.J., Slodowski, W., Pelvan, A., Becker, J., Halle, M., Buchwald, A.B., Channon, K.M. and Hecker, M. (2004) Shear stress insensitivity of endothelial nitric oxide synthase expression as a genetic risk factor for coronary heart disease. *Circ. Res.*, **95**, 841–847.
18. Miyamoto, Y., Saito, Y., Nakayama, M., Shimasaki, Y., Yoshimura, T., Yoshimura, M., Harada, M., Kajiyama, N., Kishimoto, I., Kuwahara, K. *et al.* (2000) Replication protein A1 reduces transcription of the endothelial nitric oxide synthase gene containing a -786T−>C mutation associated with coronary spastic angina. *Hum. Mol. Genet.*, **9**, 2629–2637.
19. Nakayama, M., Yasue, H., Yoshimura, M., Shimasaki, Y., Kugiyama, K., Ogawa, H., Motoyama, T., Saito, Y., Ogawa, Y., Miyamoto, Y. *et al.* (1999) T-786−>C mutation in the 5′-flanking region of the endothelial nitric oxide synthase gene is associated with coronary spasm. *Circulation*, **99**, 2864–2870.
20. Wang, J., Dudley, D. and Wang, X.L. (2002) Haplotype-specific effects on endothelial NO synthase promoter efficiency: modifiable by cigarette smoking. *Arterioscler Thromb. Vasc. Biol.*, **22**, e1–e4.
21. Balciuniene, J., Emilsson, L., Oreland, L., Pettersson, U. and Jazin, E. (2002) Investigation of the functional effect of monoamine oxidase polymorphisms in human brain. *Hum. Genet.*, **110**, 1–7.
22. Casas, J.P., Cavalleri, G.L., Bautista, L.E., Smeeth, L., Humphries, S.E. and Hingorani, A.D. (2006) Endothelial nitric oxide synthase gene polymorphisms and cardiovascular disease: a HuGE review. *Am. J. Epidemiol.*, **164**, 921–935.
23. Wang, X.L. and Wang, J. (2000) Endothelial nitric oxide synthase gene sequence variations and vascular disease. *Mol. Genet. Metab.*, **70**, 241–251.
24. de Meaux, J., Goebel, U., Pop, A. and Mitchell-Olds, T. (2005) Allele-specific assay reveals functional variation in the chalcone synthase promoter of Arabidopsis thaliana that is compatible with neutral evolution. *Plant Cell*, **17**, 676–690.
25. Sun, A., Ge, J., Siffert, W. and Frey, U.H. (2005) Quantification of allele-specific G-protein beta3 subunit mRNA transcripts in different human cells and tissues by Pyrosequencing. *Eur. J. Hum. Genet.*, **13**, 361–369.
26. Carrel, L. and Willard, H.F. (2005) X-inactivation profile reveals extensive variability in X-linked gene expression in females. *Nature*, **434**, 400–404.
27. Pinsonneault, J.K., Papp, A.C. and Sadee, W. (2006) Allelic mRNA expression of X-linked monoamine oxidase a (MAOA) in human brain: dissection of epigenetic and genetic factors. *Hum. Mol. Genet.*, **15**, 2636–2649.
28. Stephens, M. and Donnelly, P. (2003) A comparison of bayesian methods for haplotype reconstruction from population genotype data. *Am. J. Hum. Genet.*, **73**, 1162–1169.
29. Stephens, M., Smith, N.J. and Donnelly, P. (2001) A new statistical method for haplotype reconstruction from population data. *Am. J. Hum. Genet.*, **68**, 978–989.
30. Legendre, P. (1999) Program for multiple linear regression with permutation test. 32-bit DOS version ed. Département de sciences biologiques. Universitéde Montréal, Montreal, Quebec, Canada.
31. StatCorp. (2006) Intercooled Stata. *Intercooled Stata 9.2 for Windows.*, 9.2 ed. College Station, USA, pp. TX 77845.
32. Sokal, R.R. and Rohlf, F.J. (1995) Estimation of variance components in a single-classification ANOVA with unequal sample sizes. *Biometry: the principles and practice of statistics in biological research*, 3rd edn. W. H. Freeman and Company, New York, NY, pp. 214.