Research Report

Functional consequences of genetic variation in primates on tyrosine hydroxylase (TH) expression in vitro

Lisa R. Warner\textsuperscript{a,b,*}, Courtney C. Babbitt\textsuperscript{a,b}, Alex E. Primus\textsuperscript{a}, Tonya F. Severson\textsuperscript{a}, Ralph Haygood\textsuperscript{b}, Gregory A. Wray\textsuperscript{a,b,c}

\textsuperscript{a}Institute for Genome Sciences & Policy, Duke University, Durham 27708, USA
\textsuperscript{b}Department of Biology, Duke University, Durham 27708, USA
\textsuperscript{c}Department of Evolutionary Anthropology, Duke University, Durham 27708, USA

ARTICLE INFO

Article history:
Accepted 25 June 2009
Available online 8 July 2009

Keywords:
Tyrosine hydroxylase
Regulation
Genetic variation
Expression
Primate

ABSTRACT

Tyrosine hydroxylase, the rate-limiting enzyme in catecholamine synthesis, is known to contain naturally occurring genetic variation in its promoter region that associates with a number of neuropsychological disorders. As such, examining non-coding regions is important for understanding tyrosine hydroxylase function in human health and disease. We examined \(\sim 2\) kb upstream of the translation start site within humans and non-human primates to obtain a fine resolution map of evolutionarily and functionally relevant cis-regulatory differences. Our study investigated \textit{Macaca mulatta}, \textit{Pan troglodytes}, \textit{Gorilla gorilla}, and \textit{Homo sapiens} haplotypes using transient dual-luciferase transfection in three neuroblastoma cell lines to assay the impact of naturally occurring sequence variation on expression level. In addition to trans effects between cell lines, there are several significant expression differences between primate species, but the most striking difference was seen between human haplotypes in one cell line. Underlying this variation are numerous sequence polymorphisms, two of which influence expression within humans in a non-additive and cell line-specific manner. This study highlights functional consequences of tyrosine hydroxylase genetic variation in primates. Additionally, the results emphasize the importance of examining more than one cell line, the existence of multiple functional variants in a given promoter region and the presence of non-additive cis-interactions.

© 2009 Elsevier B.V. All rights reserved.

1. Introduction

The presence of functional, non-coding genetic variation throughout the human genome has been well established (Pastinen et al., 2006; Rockman and Wray, 2002). Some of these cis-regulatory variants have functional consequences with respect to disease and fitness (De Gobbi et al., 2006; Enattah et al., 2008; Hamblin and Di Rienzo, 2000; Rockman et al., 2005; Tishkoff et al., 2007). One approach to assess the impact of non-coding sequence variation is to use unbiased scans to look for signatures of positive selection either between humans and other primates or between different human populations (Haygood et al., 2007; Voight et al., 2006). Another method is to identify regulatory regions for a candidate gene of interest experimentally. Here we assess the functional importance of genetic variation in the promoter region of \textit{tyrosine hydroxylase} (TH) using transfection assays.

Tyrosine hydroxylase is widely studied, as it is the first and rate-limiting enzyme in the synthesis of catecholamine neurotransmitters (Nagatsu et al., 1964; Nagatsu, 1989). Due
to the role TH plays in the synthesis of dopamine, norepi-

mphine and epinephrine, TH coding and non-coding variants

have been implicated in psychiatric disorders (Leboyer et al.,

1990; Lobos and Todd, 1997; Meloni et al., 1995; Serretti et al.,

1998), and neurological disease (Hoffmann et al., 2003;

Ludecke et al., 1995, 1996). For instance, there is a detectable

increase in tyrosine hydroxylase levels between individuals

that suffer from major depressive disorder compared to age

matched controls (Zhu et al., 1999) and the administration

of certain antidepressants decreases the transcription abundance

of tyrosine hydroxylase (Nestler et al., 1990). As promoter regions

have been associated with some of these disorders (Rao et al.,

2007; Ribases et al., 2007; Verbeek et al., 2007), understanding

natural genetic variation and its role in regulation could have

important clinical consequences.

Tyrosine hydroxylase is intrinsically regulated in vertebrates

(Kumer and Vrana, 1996). In humans, individual TH promoter

elements confer different levels of cellular specificity high-

lighting the importance of both cis and trans effects in TH

expression (Kim et al., 2006, 2003; Romano et al., 2005). Addi-

tionally, transcription factor binding sites have been experi-

mentally validated in the TH promoter (Kessler et al.,

2003; Kim et al., 2006). When the ∼11 kb region upstream of the

mouse, rat and human TH transcription start site are aligned

there are only five evolutionarily conserved regions (Kessler et al.,

2003). This lack of conservation indicates that the promoter

region may be changing relatively rapidly and phylogenetic comparisons closer than human–mouse are necessary to understand important changes on the human branch. Finally, within human populations, a recent study resequenced ∼1.2 kb 5′ of the TH translation start site in 80 ethnically diverse subjects (Rao et al., 2007). They found four common (minor allele frequency >10%) single nucleotide polymorphisms (SNPs) within this region, SNP-824 (rs10770141), SNP-801 (rs10840490), SNP-581 (rs10770140), and SNP-494 (rs11042962) (Rao et al., 2007). The effect of these common polymorphisms in a neural environment has not been characterized.

Given the importance of the promoter region to tyrosine hydroxylase expression and regulation, as well as its association with fitness and disease, we used transfection assays to assess the ∼2 kb region upstream of the translation start site, a region known to contain functional polymorphisms. Transfection assays in human cell lines have been successfully used to identify and verify cis-regulatory variants effecting expression between primate species (Chabot et al., 2007; Rockman et al., 2005). Our study is the first to investigate genetic variation in tyrosine hydroxylase using non-human primates to polarize human specific changes in expression. We tested the impact of interspecific and intraspecific variation as well as the functional consequence of cis-regulatory element interaction.

2. Results

2.1. Sequence analysis between non-human primates and humans

In order to assess naturally occurring primate sequence variation 5′ of the TH translation start site, we isolated ∼2 kb of this region from six haplotypes. The non-human primates investigated were macaque (AG07109), gorilla (AG05251), and chimpanzee (AG06939). The human variation examined comes from three HapMap individuals, representing three geographically distinct human populations: a Luhyan in Webuye, Kenya (GM19360), a Yoruba in Ibadan, Nigeria (GM18522), and a CEPH panel member who is a Utah resident with ancestry from northern or western Europe (GM12154). When all six haplotypes, which represent a single haplotype from the population, are aligned, there are over two hundred sequence differences (Fig. S1). We also obtained ∼2 kb of TH intronic regions from macaque, orangutan, chimpanzee, and human and compared divergence in the 5′-flanking region to that of the intronic sequences (excluding first introns, which often contain regulatory elements and intron ends, which generally contain splice sites), in order to test for signatures of positive selection between species (Haygood et al., 2007). We found no significant acceleration of substitutions in the TH promoter region relative to the introns. Assuming the intronic sequences examined are largely functionless and have not experienced appreciable positive selection, our finding suggests that the promoter region has experienced little or no positive selection along the human lineage (Table S1). Moreover, measures of sequence polymorphism and linkage disequilibrium show no evidence of positive selection within human in this region based on the HapMap analysis panels (Voight et al., 2006). Nevertheless, as tests for positive selection are generally underpowered and given the importance of tyrosine hydroxylase as the rate-limiting step in catecholamine synthesis, we wanted to functionally characterize the abundant variation in this regulatory region.

2.2. Expression of primate TH in multiple neuronal cell lines

To determine which aspects of these sequence differences have a functional impact, we performed in vitro dual-luciferase assays for each of the six haplotypes in three neuroblastoma cell lines SH-SYSY, SK-N-BE2, and IMR-32, each of which is commonly used as a proxy for neurons in cell culture. These transfection assays were repeated on three different days with eight wells of replication per construct. We fitted a mixed-model ANOVA to the background normalized transfection expression measurements, for each haplotype in each cell line using eight wells of replication per construct. We fitted a mixed-model ANOVA to the background normalized transfection expression measurements, for each haplotype in each cell line (Table S2). By this means, we sought to eliminate not only random but also systematic errors, such as those arising from changes in the condition of the cell cultures from one day of assays to another. The fitted fiducial expression levels served as inputs to subsequent analyses.

To investigate trans effects, the constructs were transiently transfected into three neuroblastoma cell lines. The rank order of expression when comparing the six constructs is different between the cell lines, indicating that different cellular environments affect TH gene expression (Fig. 1). For example, the gorilla haplotype has the highest TH expression in the IMR-32 cells, but the lowest expression in the SH-SYSY and SK-N-BE2 cells (Fig. 1).

When expression is compared between non-human pri-

mates and humans, there are differences between cell lines. The standard deviation of expression is greater among the
non-human primates for SH-SY5Y and SK-N-BE(2) as compared to among humans (Fig. 1 and Table S3). This is what one would expect, given there are many more sequence differences among the non-human primates than among the humans. However, the reverse is true in the IMR-32 background where there is a greater standard deviation of expression within humans as compared to non-human primates (Fig. 1 and Table S3).

Additionally, there are significant and repeatable differences between haplotypes within each cell line. The letters above the bar graph indicate significant differences for SH-SY5Y, SK-N-BE(2), and IMR-32 cell lines as tested by the Tukey-Kramer HSD test, which incorporates adjustment for multiple testing. Bars that do not have the same letter are significantly different from one another (SH-SY5Y p < 0.04; SK-N-BE(2) p < 0.001; and IMR-32 p < 0.05). When the constructs are analyzed in the SH-SY5Y cell line, there are differences between the non-human primates as the constructs for the macaque and chimpanzee are significantly different from the gorilla, but there are no intraspecific differences among humans (Fig. 1a). This is not the case when the TH 2 kb upstream region is assessed in the IMR-32 cell line. Strikingly, in this background there are significant intraspecific differences within humans. The GM19360 and GM18522 haplotypes have higher expression as compared to GM12154 individual (Fig. 1c).

Together, these results indicate that multiple sequence differences, both within humans and between humans and non-human primates affect TH transcription, and that these consequences are to some extent dependent on cell type. In all three cell lines, the range of functional variation within humans is similar to, but lower than, interspecies differences. Below, we focus on identifying specific functional variants within humans, and their possible interactions.

2.3. Examining the functional impact of TH promoter SNPs in IMR-32 cells

To assess the functional consequences of the previously identified four common polymorphisms (Rao et al., 2007) on TH expression in neuroblastoma cells, we used site-directed mutagenesis to alter these SNPs in the GM12154 background and tested these artificial haplotypes in IMR-32 cells. These transfection assays were repeated on three different days with eight wells of replication per construct. We again fitted a mixed-model ANOVA to the relevant transfection expression measurements, to obtain a fiducial expression level for each haplotype (Table S2). The letters above the bar graph indicate significant differences (p < 0.05) by the Tukey-Kramer HSD test (Fig. 2). The haplotype with all four major alleles has the highest expression (Fig. 2). Furthermore, the major allele haplotype is significantly different from the one containing the minor allele of SNP-801 or SNP-494 (Fig. 2).

In order to identify sequence features underlying the expression differences seen between the natural human haplotypes in the IMR-32 cell line, we examined the sequence features in the 2 kb region directly upstream of the translation start site. A total of eleven variants exist among the three human constructs, with seven specific to GM19360, two specific to GM18522, and two specific to GM12154 (Fig. 3). The only two sequence features that are common to the high expressing GM19360 and GM18522 constructs and different from the lower expressing GM12154 construct are SNP-1898 (rs7115640) and SNP-801 (rs10840490) (grey columns, Fig. 3). As such, these two features are the best candidates for explaining TH expression differences between humans in the IMR-32 cellular environment.

In order to assess the function of SNP-1898 and SNP-801 we mutated these two SNPs separately and together from their GM19360/GM18522 state to the GM12154 state in the GM12154 background and tested these in the IMR-32 cell line. These
transfection assays were repeated on three different days with eight wells of replication per construct. We again fitted a mixed-model ANOVA to the relevant transfection expression measurements, to obtain a fiducial expression level for each haplotype in IMR-32 cells (Table S2). As before, letters above the bar graph indicate significant differences (p < 0.001) by the Tukey–Kramer HSD test (Fig. 4). We would anticipate that if a particular SNP is functional, the expression of the mutated GM12154 construct would be increased to levels similar to the GM19360 and GM18522 constructs. We indeed saw this with both of the single mutations: −1898T>C and −801C>G each raise the expression to levels significantly different from the natural GM12154 haplotype and statistically indistinguishable from the GM19360 and GM18522 haplotypes (Fig. 4). Interestingly, a construct containing both of these SNPs (−1898T>C; −801C>G) drives expression at a level similar to the natural, lower expressing, GM12154 haplotype (Fig. 4). This result indicates that the interaction between these two SNPs is non-additive.

3. Discussion

This investigation set out to determine whether genetic differences in cis affect TH expression between primate species as well as within humans, using a method that has been successful in identifying and verifying cis-regulatory elements between primate species (Chabot et al., 2007; Rockman et al., 2005). In addition to confirming previous functional polymorphisms and identifying new variants in the 2 kb 5′ flanking region upstream of the TH translation start site (Fig. S1), this study revealed several interesting results.

Looking beyond sequence signals to expression, prior studies have revealed that the same TH haplotype transfected into different cell lines can result in different levels of expression (Kim et al., 2003). Our study, is the first to examine this in multiple primate species using three neuroblastoma cell lines, and demonstrates that variant-by-cell type
interactions are widespread at this locus. We recognize that the significant expression differences observed in vitro might not necessarily translate to biologically meaningful differences in vivo. However, differences in tyrosine hydroxylase transcript abundance have been associated with psychiatric (Zhu et al., 1999) and neurological phenotypes (Joyce et al., 1997).

By comparing humans to other non-human primates, we found that although there are clearly statistically significant differences between primate species and humans are not categorically unusual (Fig. 1). Some of the differences detected are attributed to trans effects, as the same construct can show different levels of expression depending on the cellular context. This is seen as the rank order of expression for the six haplotypes changes depending on the cell line (Fig. 1) and the standard deviations between non-human primates and humans are different depending on cellular context (Table S3). These findings highlight the importance of investigating expression in multiple cellular environments in order to reveal trans effects. The cell lines we used originate from different biopsies, and neuroblastomas are known to be somewhat representative of different environments (Ciccarello et al., 1989). If TH promoter activity were just studied in SH-SYSY or SK-N-BE(2) cells, we would have missed an interesting case of sequence polymorphisms driving expression variation within humans (Fig. 1c).

To extend the intraspecific analyses, we investigated the functional impact of standing common genetic variation in the TH promoter region. Previously, four common polymorphisms were discovered in the 1.2 kb region upstream of the translation start site (Rao et al., 2007). We isolated each of these polymorphisms in the GM12154 background and then examined their expression in IMR-32 cells. This approach is different from the previous study, which examined haplotypes in chromaffin cells and was concerned with associating SNPs with in vivo biochemical and physiological measurements (Rao et al., 2007). There are significant differences among these four common SNPs. Specifically, the minor allele of SNP-801 and SNP-494 has significantly different expression from the major alleles (Fig. 2). This result in conjunction with those presented in Fig. 4 demonstrates that multiple functional polymorphisms within a regulatory region can be segregating in human populations. To date, only a few other studies have searched for multiple promoter variants that segregate in human populations. To date, only a few other studies have searched for multiple promoter variants that segregate in human populations. To date, only a few other studies have searched for multiple promoter variants that segregate in human populations.

To further examine the natural intraspecific expression differences in IMR-32 cells, we investigated the sequence features associated with three human haplotypes (Fig. 3). In order to pinpoint functional variation, we focused on nucleotides where the two higher expressing haplotypes (GM19360 and GM18522) have the same state and the lower expressing haplotype (GM12154) has a different allele. The haplotypes we tested contain only two instances of this, at SNP-1898 and SNP-801, making these the most likely candidates for causing the expression differences (grey columns, Fig. 3). When either SNP was isolated individually in the GM12154 background, we saw a significant increase in expression from the natural GM12154 state (Fig. 4). Although these single nucleotide mutations were artificially created, we know that these combinations of SNPs segregate naturally in human populations based on HapMap genotypes (Rel_23A: The International HapMap Consortium, 2003).

Surprisingly, when these two SNPs were mutated together in the GM12154 background, the expression remained statistically unchanged (Fig. 4). This is highly suggestive of non-additive interactions between SNP-1898 and SNP-801 and further analysis in other genetic background could potentially provide additional insight. The current results indicate that testing a single sequence feature will not always reveal its functional consequences and that one needs to examine other variants that are segregating in the surrounding regions in order to assess the possibility of combinatorial effects. This example adds to a growing number of genes where non-additive interactions between cis acting SNPs have been found, including GH1 (Horan et al., 2003), KRT1 (Tao et al., 2006), and PDYN (Rockman et al., 2005). The number of examples is most likely due to the small number of studies that have carried out fine-scale functional studies rather than a reflection of biological reality. Indeed, the extent of non-additive interaction between regulatory variants on gene regulation is poorly understood in general.

Knowing that these two SNPs impact expression in a transfection assay, transcription factor binding could be modifying TH transcription as transfection experiments can separate the effects of transcription factors from other types of expression regulation. Using biochemical approaches, Kim et al. (2006) identified three NRSF/REST binding sites in the 5’ regulatory region of TH, one of which, NRSE-R, overlaps with SNP-1898 and is known to interact with the TH promoter in vitro. The GM19360 and GM18522 constructs have the same allele as the NRSE-R consensus sequence given for NRSF/REST binding at SNP-1898 (Kim et al., 2006), suggesting a possible molecular mechanism for the expression differences between haplotypes. Further work is needed to validate the impact of transcription factor binding events at SNP-1898 and SNP-801 on TH expression.

We identified many inter- and intraspecific variants in the 5’ flanking region of TH and examined their effects on expression in neuroblastoma cell lines. The expression differences we observed between human haplotypes in IMR-32 cells can be attributed partly to the functional variants SNP-1898 and SNP-801. Furthermore, we demonstrated the importance of using multiple cell lines due to trans effects, the presence of non-additive interactions between regulatory variants, and the existence of multiple functional polymorphisms within a promoter region. Finally, this study provides novel functional candidate variants for examining associations between tyrosine hydroxylase and neuropsychiatric disease.

4. Experimental procedures

4.1. Cloning and sequencing

 Constructs were made from non-human primate fibroblast purified DNA from macaque, chimpanzee and gorilla (respectively, Macaca mulatta — AG07109; Pan troglodytes — AG06939;
Gorilla gorilla — AG05251) and human lymphoblast purified DNA (Luhya from Webuye, Kenya — GM19360; Yoruba from Ibadan, Nigeria — GM18522; CEPH a Utah resident with ancestry from Northern or Western Europe — GM12154) received from Coriell Institute (Camden, New Jersey). The 2 kb TH cis-regulatory haplotypes were obtained through PCR-amplification (primers used: 5′ AGCGAAATCCCTCCAAGCC 3′ and 5′ GGCTCAGTGTGGAGGTC 3′) using the high-fidelity polymerase Phusion (Finzymes, Espoo, Finland). Individual PCR-amplification products were cloned into pGL4.1 vector using a KpnI restriction site on the 5′ end and a Xhol restriction site on the 3′ end. Constructs were then prepared using the Wizard midi-prep kit (Promega, Madison, WI, USA) and sequenced. These sequences have been deposited in GenBank under the following accession numbers GQ403011 (AG07109), GQ403010 (AG06939), GQ403009 (AG05251), GQ403014 (GM19360), GQ403013 (GM18522), and GQ403012 (GM12154).

4.2. Positive selection analysis

The 5′ flanking regions for chimpanzee and macaque came from AG06939 and AG07109. The 5′ flanking regions for human (ENST00000352909) and orangutan (Pongo pygmaeus, ENSPPT0000003515), as well as all of the intronic sequences (introns: 3, 4, 5, 6, 7, 8, 9, 10, 11, 12) came from Ensmbi (human — ENST00000352909; chimpanzee — ENSPTRT00000006067; orangutan — ENSPT00000003515; and macaque — ENSMUT00000005543). We fitted single nucleotide substitution models by likelihood maximization using HyPhy (http://www.hyphy.org), taking the best of 10 fits starting from random points to guard against local maxima of the likelihood function (Haygood et al., 2007). The models are equivalent to the HKY85 model (Hasegawa et al., 1985) modulated within the 5′-flanking region relative to the intronic sequences in the same way that preferred models are modulated at non-synonymous sites relative to synonymous sites (Zhang et al., 2009). The null model accommodates relaxation of negative selection on the so-called foreground lineage where the alternate model allows positive selection. (Hy Phy Batch Language code is available upon request to RH). We considered each terminal lineage (macaque, orangutan, chimpanzee, human) in turn as the foreground lineage. In each case, we tested the fit of the alternate model relative to that of the null model using a likelihood ratio test conservatively approximated as a chi-squared test with one degree of freedom (Zhang et al., 2005). In each case, we analyzed 100 bootstrap replicates over the intronic alignment to guard against small-sample stochasticity; each p-value in Table S1 is the median over the replicates.

4.3. TH RT-PCR

To verify that SH-SY5Y, SK-N-BE(2), and IMR-32 cells express TH, fragments from these cell lines were separately reverse transcribed and then PCR amplified. RNA was extracted from SH-SY5Y, SK-N-BE(2), and IMR-32 cells with an Aurum total RNA extraction kit (Bio-Rad, Hercules, CA, USA). The cDNA single strand synthesis used a High Capacity cDNA Archive kit (Applied Biosystems, Foster City, CA, USA). Primers were designed to amplify the four main TH alternative transcripts (Alternative Splicing Database at EMBL-EBI) (primers used: 5′ ACTGCTGCCAGGAGCTG 3′ and 5′ TGGACAGCTTCT-CAATTTCCT 3′). PCR was then completed with the Absolute qPCR SYBR kit (Abgene, Epsom, United Kingdom). In all three cell types, the expected amplicon was observed at 123 bp (Fig. S2). These results indicate that these cell lines are appropriate for assessing the expression of TH and that readings obtained are not artifacts from the experiment.

4.4. Cell culture

SH-SY5Y and SK-N-BE(2) cells were cultured in a 1:1 mixture of Ham’s F12K and MEME (ATCC, Manassas, VA, USA) supplemented with 10% FBS (HyClone, Logan, UT, USA). IMR-32 cells were cultured in MEME (Sigma-Aldrich Corp., St. Louis, MO, USA) supplemented with 1 mM sodium pyruvate, 0.1 mM non-essential amino acids (Gibco BRL), and 10% FBS (ATCC). These cell lines were acquired from ATCC or the Cell Culture Facility at Duke University and were maintained at 37 °C with 5% CO₂.

4.5. Site directed mutagenesis

The 2 kb TH cis-regulatory GM12154 haplotype in the pGL4.1 vector was methylated per GeneTailor Site-Directed Mutagenesis System (Invitrogen, Carlsbad, CA, USA) and was then used as the DNA template. The SNP location given is in reference to the human TH sequence from Ensmbi ENST00000352909. Six different single nucleotide changes were obtained through PCR-amplification using Platinum Taq DNA Polymerase high fidelity (Invitrogen, Carlsbad, CA, USA) using the following primer sets, −494G>A, (5′ AGACACACGGCTGAACTTCTTCTGAG 3′ and 5′ CCAAGCCGCTGTGTTCTTTAGCAGTTGT 3′), −581A>G (5′ GGCGGAGCTTTGGAAGCCGCTGCAAG 3′ and 5′ TTCCCAAGCTCAGGCTCGCTAAGGTT 3′), −801C>G (5′ CGTCTCTGTGCTCCTTGCTG 3′, and 5′ AGGAAAGCCAGAGCTAGCAGGCT 3′). These results indicate that these cell lines are appropriate for assessing the expression of TH and that readings obtained are not artifacts from the experiment.

4.6. Transfection and expression measurement

To control for variation each construct was transfected in eight wells each day and each experiment was repeated three times on separate days. The first set of experiments used the six natural haplotypes and all three cell types, SH-SY5Y, SK-N-BE(2), and IMR-32 (Fig. 1). The second set of experiments used the three human natural haplotypes and the six mutagenesis created GM12154 constructs, in IMR-32 cells (Figs. 2 and 4). Transfections were performed in 24-well plates and the cells were seeded at 8 × 10⁶ cells/ml in a volume of 500 µl of supplemented medium. Cells were transfected 24 h after
seeding using Lipofectamine 2000 (Invitrogen, Carlsbad, CA, USA). The transfection mixture used was 2 μl Lipofectamine 2000, 100 μl OPTI-MEM, 800 ng reporter construct and 200 ng of Renilla-TK as a co-reporter. Another control, 427 ng of empty pGL4.1, was added to the transfection mixture, which is the molar equivalent of the other constructs. Forty-eight hours after transfection the cells were lysed with 100 μl of Passive Lysis Buffer (Promega) per well for 30 min. Lysates were read in the automated 96-well Veritas Luminometer (Turner Biosystems, Sunnyvale, CA, USA) with auto-injection following procedures obtained from Promega, using a second following delay and 10 second read time.

4.7. In vitro expression analysis

In order to distinguish true biological signal from the background noise inherent in gene expression experiments, we fitted our measurements to a mixed-model ANOVA. We first calculated the mean of the promoterless pGL4 vector over eight replicates for each cell type on each day, subtracted this from each expression measurement for a TH cis-regulatory haplotype in that cell type on that day, and computed the base-2 logarithm of the difference (C. Babbitt, J. Silverman, R. Haygood, M. Rockman, J. Reininga-Craven, G. Wray, manuscript submitted). For each cell type, we then used restricted maximum likelihood (REML) to fit the “normalization model”:

\[ y_{ijk} = \mu + H_i + D_j + (HD)_{ij} + \epsilon_{ijk}, \]

where \( y_{ijk} \) is the logarithmically transformed, background-subtracted value of measured expression for haplotype \( i \) (1–12), day \( j \) (1–6), and well \( k \) (1–8), \( \mu \) is the overall mean, \( H_i \) is the fixed main effect of haplotype \( i \), \( D_j \) is the random main effect of day \( j \), \( (HD)_{ij} \) is the random interaction effect of haplotype \( i \) and day \( j \), and \( \epsilon_{ijk} \) is the residual. This model, which resembles models commonly used in the analysis of microarray data (Wolfinger et al., 2001), accounts for not only purely technical, well-to-well variation but also systematic effects of day arising from, for example, day-to-day variation in the cell cultures. The fits indicate that these effects are substantial, accounting for 35% to 70% of expression variance (Table S2). For each cell type, we used the Tukey–Kramer procedure to assess the significance of each pairwise difference between fitted \( H_i \). The fitted \( H_i \) constitute our best measures of the typical expression from the haplotypes, so we call them “fiducial”. We performed these computations using the R system for statistical analyses (Team R, 2005). Our R code is available on request.

All relevant measured data from the six days are included in the different analyses except for one data point from the macaque haplotype (AG07109) in the IMR-32 cell line. This point was not included as its expression level was similar to the empty vector control.

Acknowledgments

We thank Adam Pfefferle and Gregory Nachtrab for review of this manuscript and helpful suggestions. This work was supported by the Institute for Genome Sciences & Policy and the National Institutes of Health [T32-HD040372 to L.R.W.].

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.brainres.2009.06.086.

References


Kumer, S.C., Vrana, K.E., 1996. Intricate regulation of tyrosine


