



## The impact of genomic neighborhood on the evolution of human and chimpanzee transcriptome

Subhajyoti De, Sarah A. Teichmann and M. Madan Babu

*Genome Res.* 2009 19: 785-794 originally published online February 19, 2009

Access the most recent version at doi:[10.1101/gr.086165.108](https://doi.org/10.1101/gr.086165.108)

---

**References** This article cites 52 articles, 14 of which can be accessed free at:  
<http://genome.cshlp.org/content/19/5/785.full.html#ref-list-1>

**Open Access** Freely available online through the *Genome Research* Open Access option.

**License** Freely available online through the Genome Research Open Access option.

**Email Alerting Service** Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).



---

To subscribe to *Genome Research* go to:  
<https://genome.cshlp.org/subscriptions>

---

Copyright © 2009 by Cold Spring Harbor Laboratory Press

# The impact of genomic neighborhood on the evolution of human and chimpanzee transcriptome

Subhajyoti De,<sup>1</sup> Sarah A. Teichmann, and M. Madan Babu<sup>1</sup>

MRC Laboratory of Molecular Biology, Cambridge CB2 0QH, United Kingdom

Divergence of gene expression can result in phenotypic variation, which contributes to the evolution of new species. Although the influence of *trans*- and *cis*-regulatory mutations is well known, the genome-wide impact of changes in genomic neighborhood of genes on expression divergence between species remains largely unexplored. Here, we compare the neighborhood of orthologous genes (within a window of 2 MB) in human and chimpanzee with the expression levels of their transcripts from several equivalent tissues and demonstrate that genes with altered neighborhood are more likely to undergo expression divergence than genes with conserved neighborhood. We observe the same trend when expression divergence data were analyzed from six different brain parts that are equivalent between human and chimpanzee. Additionally, we find enrichment for genes with altered neighborhood to be expressed in a tissue-specific manner in the human brain. These results suggest that expression divergence induced by this mechanism could have contributed to the phenotypic differences between human and chimpanzee. We propose that, in addition to other molecular mechanisms, change in genomic neighborhood is an important factor that drives transcriptome evolution.

[Supplemental material is available online at <http://www.genome.org>.]

One of the fundamental problems in the postgenomic era is to understand what mechanisms drive divergence in gene expression, and how this results in phenotypic changes, ultimately leading to the evolution of new species (King and Wilson 1975; White 2001; Carroll 2003; Wray et al. 2003; West-Eberhard 2005; Khaitovich et al. 2006). For example, changes in the expression level of *Bmp4* (Abzhanov et al. 2004; Wu et al. 2004) and calmodulin (Abzhanov et al. 2006) are linked to the evolutionary variation in beak morphology in finches, resulting in adaptation to different food sources and ultimately to the emergence of new species. The genetic material of eukaryotes is organized in a complex, hierarchical manner and is encoded in several linear chromosomes, providing ample opportunity for regulation of gene expression at several levels (Felsenfeld and Groudine 2003; Wray et al. 2003). Due to this intricate organization, mutations of different sizes can affect the expression pattern of one or several genes simultaneously. At one end of the spectrum are small-scale mutations (e.g., single nucleotide polymorphisms) that affect *cis*-regulatory elements, thereby changing the expression of a nearby gene. At the other end are large-scale mutations (e.g., translocation and change in ploidy) that involve large DNA segments or chromosome number, thereby affecting the expression of many genes. Though numerous studies have investigated the contribution of such mutations to expression divergence, the genome-wide impact of intermediate scale mutations, such as copy number variation and position effect, has only recently gained attention (Hurles et al. 2008).

Position effect is a phenomenon that results in an altered expression pattern of a gene as a consequence of a change in its genomic neighborhood, which may also lead to a phenotypic change (Fig. 1). Alteration of the genomic neighborhood can occur

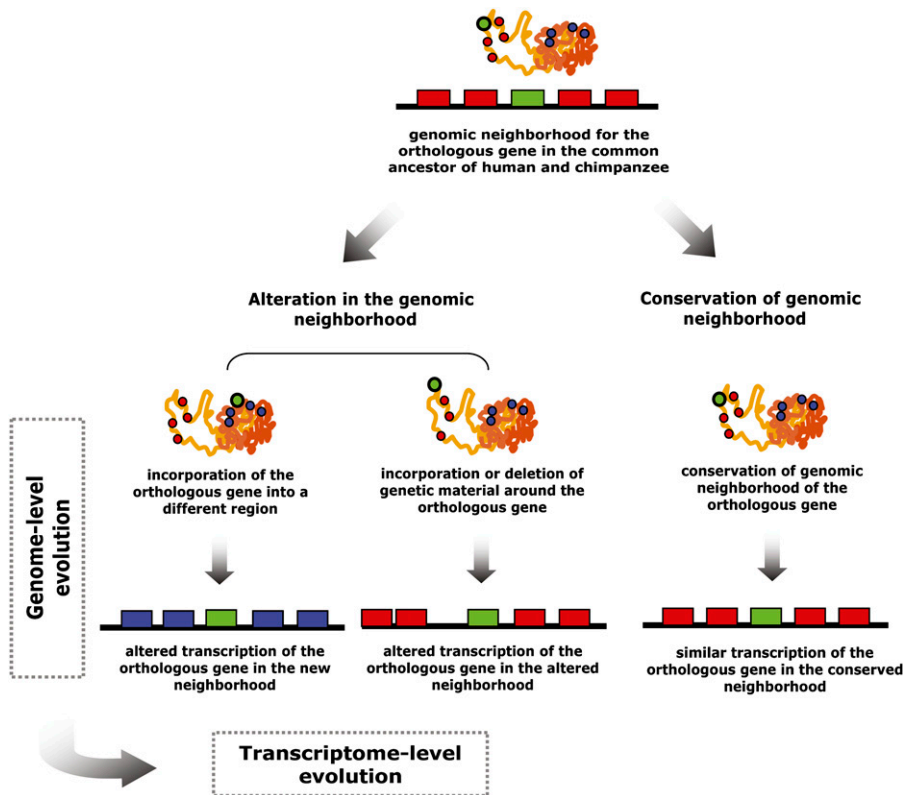
due to insertion or deletion of genetic material (e.g., L1 elements) around the gene, resulting in altered expression level (Han et al. 2004). Alternatively, a change in genomic neighborhood may be introduced due to a recombination or duplication event, resulting in the incorporation of the gene into a completely different region (e.g., euchromatic or heterochromatic region), thereby affecting its expression (Xiao et al. 2008). Often such changes in humans have been associated with genetic disorders (Kleinjan and van Heyningen 1998; Kleinjan and van Heyningen 2005) and have been described to be a result of (1) incorporation of a gene into a constitutive heterochromatic or euchromatic region, giving rise to position effect variegation (e.g., Saethre-Chotzen syndrome, which is characterized by facial asymmetry and short digits, is associated with a position effect mediated expression variation of the *TWIST1* [also known as *TWIST*] gene on chromosomes 7p21; Rose et al. 1997), (2) translocation of a gene next to another gene, thereby resulting in a competition for the same regulatory element or juxtaposition with an enhancer element (e.g., Burkitt's lymphoma, where the *MYC* gene is placed under the control of an immunoglobulin enhancer thereby resulting in a competition for the same regulatory element and misexpression of the genes; Joos et al. 1992), or (3) separation of the promoter from a distal tissue-specific regulatory element such as an enhancer, locus control region, or an insulator (e.g., several instances of Campomelic dysplasia, which result in skeletal malformations, involve misexpression of the *SOX9* gene in 17q24.3 due to a deletion around the gene; Velagaleti et al. 2005). Although a change in neighborhood leading to altered expression pattern can be deleterious for some genes, the effect in other instances may be neutral or may lead to variation in phenotypes (Jordan et al. 2005) upon which the beneficial changes can be selected during evolution (Fig. 1). For instance, a recent study (Xiao et al. 2008) has discovered that duplication and change in genomic neighborhood of *SUN* in tomato plants increased its expression relative to that of the ancestral copy. This event has resulted in an enormous variation in fruit shape that has been subsequently selected for in today's domesticated tomato plants.

## <sup>1</sup>Corresponding authors.

E-mail [madanm@mrc-lmb.cam.ac.uk](mailto:madanm@mrc-lmb.cam.ac.uk); fax 44-1223-213556.

E-mail [sde@mrc-lmb.cam.ac.uk](mailto:sde@mrc-lmb.cam.ac.uk); fax 44-1223-213556.

Article published online before print. Article and publication date are at <http://www.genome.org/cgi/doi/10.1101/gr.086165.108>. Freely available online through the *Genome Research* Open Access option.



**Figure 1.** A model describing how change in gene neighborhood might contribute to expression divergence. The genetic material is shown as a curly orange object and as a black line. The gene of interest is shown as a circle and box in green color. Neighbors of the gene, which are conserved between two species, are shown as red circles and boxes. The new neighbors of the gene upon incorporation into a new neighborhood are shown as blue circles and boxes. If the neighborhood of a gene is altered in one species relative to the other (and vice versa), the gene is likely to experience a different transcriptional environment. We postulate that an altered transcriptional environment is likely to lead to expression divergence between species. This may also result in a spectrum of phenotypic consequences, ranging from deleterious to neutral or advantageous upon which natural selection can operate.

We hypothesize that a change in the neighborhood of genes is likely to introduce variation at the genomic and transcriptomic level, potentially leading to phenotypic divergence (Fig. 1). In this work, we set out to investigate the prevalence and role of this mechanism in gene expression divergence during human and chimpanzee evolution. By using genomic and transcriptomic data from human and chimpanzee, we investigate the following questions: How many genes show an alteration in their genomic neighborhood in human or chimpanzee after the split from their common ancestor? Do genes with altered neighborhood display more expression divergence than other genes? In which parts of the body are such genes expressed?

## Results

Approximately 18% of the orthologous genes altered their genomic neighborhood either in human or chimpanzee

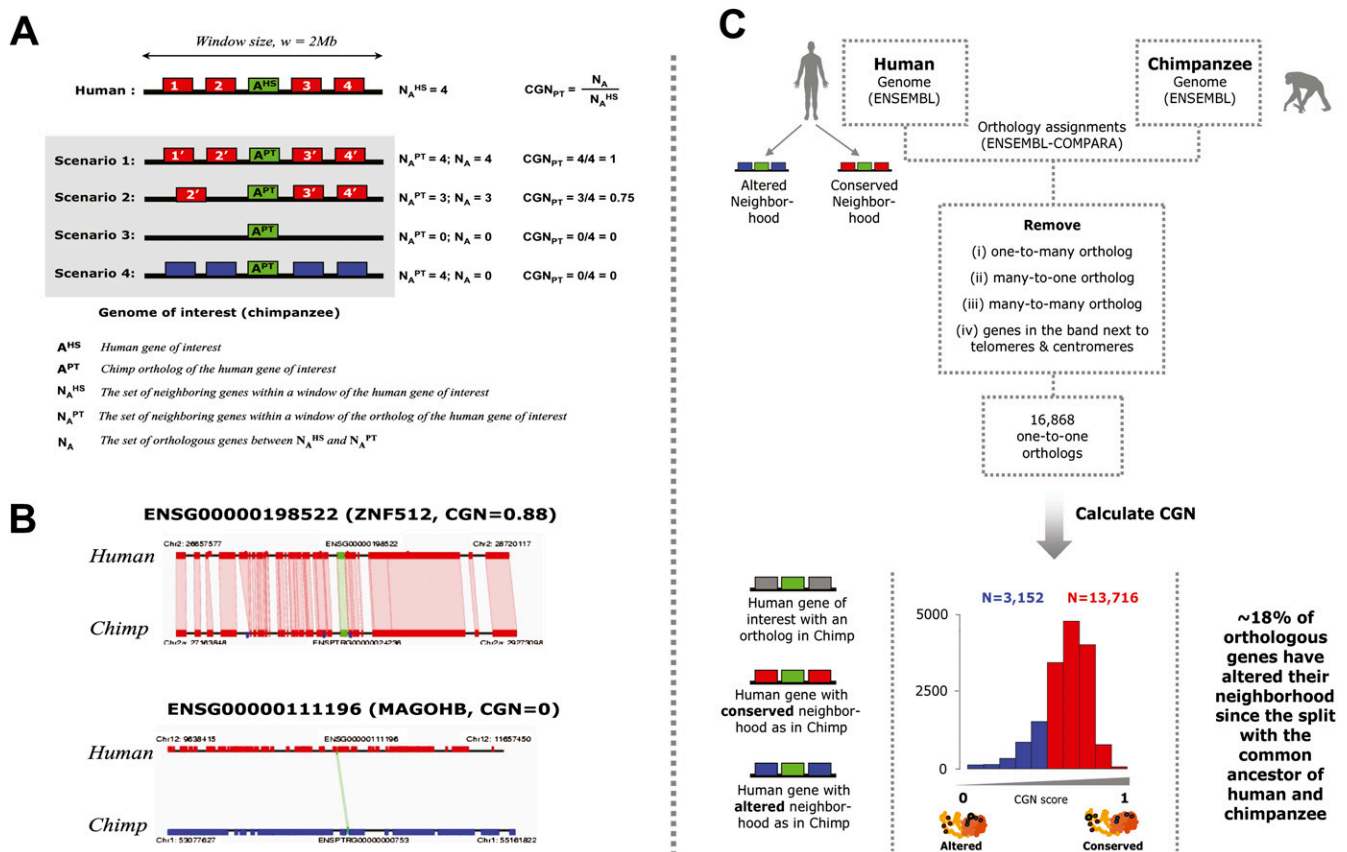
To identify genes that have alterations in their genomic neighborhood, we first defined a metric called conservation of genomic neighborhood (CGN) score for every gene. This metric estimates the extent of conservation of local chromosomal neighborhood of a gene between two species (Fig. 2A). The CGN score of a given gene

in human is simply the fraction of genes within a window (of size  $w$ ) surrounding it, which are orthologous and are also present in an equivalent window around the chimpanzee ortholog (see Methods section for details). Since single gene duplication events, insertion and deletion by recombination and transposable elements can alter the genomic neighborhood, and because such changes in primates generally affect less than 1 Mb of the genome (Bailey and Eichler 2006; Sharp et al. 2006), we chose a conservative window size of 2 Mb to calculate CGN scores to ensure that we account for such events. For example, if 27 of the 30 genes (within a window of 2 Mb centered on the gene of interest) are orthologous and are also present in an equivalent 2 Mb window around the chimpanzee ortholog, then the CGN score for that gene is 27/30, or 0.9. Unlike other measures, which assess overall sequence similarity of promoter regions alone, our proposed metric can account for large alterations, such as insertions and deletions in the neighborhood as well as change in chromosomal position. Therefore, the CGN score is a pragmatic measure of the magnitude of change to a gene's local environment.

### Calculation of CGN score

A high CGN score ( $\text{CGN} > 0.5$ ), where more than half the neighbors are conserved within a particular window ( $w = 2$  Mb) indicates conservation of its local chromosomal environment, whereas a low CGN score ( $\text{CGN} \leq 0.5$ ), where less than half the neighbors are conserved means that the gene might experience a different environment as compared to the reference species. The extreme values, CGN scores of 0 or 1, represent completely altered or absolutely conserved genomic context, respectively (Fig. 2B). It should be noted that the CGN scores were comparable when we considered different window sizes i.e.,  $w = 1$  Mb or 3 Mb or when we used a fixed number of neighboring genes around the gene of interest to calculate the CGN score i.e., a window of 30 or 40 genes around the gene of interest. These control calculations show that the definition of CGN score for a gene is robust to the choice of parameters used (see Supplemental Table S1). More importantly, the results we describe later in this work are not qualitatively affected by the cutoff used for identifying genes with altered or conserved neighborhood suggesting that the reported results are generally robust to the choice of parameters employed (see below).

Of the predicted human and chimpanzee genes, we obtained reliable orthologs for 19,256 genes from Ensembl-Compara, which assigns orthology mapping by building phylogenetic trees of homologs using a maximum likelihood approach (Hubbard et al. 2007). For each of the 19,256 human genes, we calculated the CGN score using chimpanzee as a reference species. To ensure that errors in genome assembly or orthology detection do not affect the



**Figure 2.** (A) Definition of conservation of genomic neighborhood (CGN). The gene of interest is shown in green and the neighbors of human genes are shown in red, labeled 1–4. The different evolutionary scenarios, resulting in a CGN score between 0 and 1, while comparing the neighborhood of orthologous chimpanzee and human genes are illustrated. Orthologs of the neighbors of the human gene in chimpanzee are shown in red and labeled 1'–4'. Neighbors of the chimpanzee gene that are not neighbors of the human ortholog are shown in blue. Because of the way in which the CGN score is calculated, the identification of genes with low CGN would include both the set of genes that have changed their neighborhood either in the human lineage or in the chimpanzee lineage after the split from their common ancestor. (B) An example of a gene with highly conserved genomic neighborhood (CGN = 0.88, *ZNF512*) and completely altered genomic neighborhood (CGN = 0, *MAGOHB*). (C) A CGN score was calculated for every one-to-one orthologous gene in human using chimpanzee as a reference genome and a histogram of the distribution is shown below. Genes with conserved neighborhood (CGN > 0.5) are shown in red while those with altered neighborhood (CGN ≤ 0.5) are shown in blue.

calculation of the CGN score, we performed control calculations and used several filters. First, we used the Macaque genome sequence as an out-group reference species to remove potentially spurious instances of genes with altered neighborhood (see Supplemental Table S2). We further removed all genes that are located in incompletely assembled genomic regions and the chromosomal band nearest to the telomeres and centromeres since such regions are often subject to sequencing errors and structural variations. We also excluded one-to-many, many-to-one, and many-to-many orthologs and focused only on one-to-one orthologs. This resulted in 16,868 genes with CGN scores (see Supplemental Table S3) of which ~18% (3152 genes) have altered their neighborhood either in human or in chimpanzee. It should be noted that because of the way in which the CGN score is calculated, the identification of genes with low CGN would include both the set of genes that have changed their neighborhood either in the human lineage or in the chimpanzee lineage after the split from their common ancestor. Of the 3152 genes, 98 (~0.6%) have a CGN score of 0 suggesting that they have completely changed their genomic neighborhood, possibly by being transferred into an entirely different region or due to an incorporation or deletion of genetic material around the gene after the split from the common ancestor.

### Chromosomal distribution

We observed that the genes with low CGN score map to all the human chromosomes and are scattered over the entire length of the chromosomes (see Supplemental Table S4). We then investigated the chromosomal distribution of genes with CGN scores to identify those chromosomes that showed an enrichment to contain genes with low or high CGN when compared to the genome-wide distribution using the Mann-Whitney two-tailed test. We found that chromosome (chr) 19 ( $n = 1016$  genes,  $P$ -value =  $<2.2 \times 10^{-16}$ ), chr 17 ( $n = 943$  genes,  $P$ -value =  $6.99 \times 10^{-12}$ ), and chr 11 ( $n = 1037$  genes,  $P$ -value =  $<2.2 \times 10^{-16}$ ) were the ones that were enriched in genes with high CGN score (i.e., genes with conserved neighborhood). On the other hand, apart from the sex chromosomes (chr X,  $n = 607$  genes,  $P$ -value =  $<2.2 \times 10^{-16}$  and chr Y,  $n = 28$  genes,  $P$ -value =  $3.63 \times 10^{-7}$ ), we find that chromosomes chr 18 ( $n = 248$  genes,  $P$ -value =  $1.23 \times 10^{-5}$ ), chr 7 ( $n = 780$  genes,  $P$ -value =  $2.2 \times 10^{-16}$ ), and chr 13 ( $n = 270$  genes,  $P$ -value =  $1.93 \times 10^{-13}$ ) show enrichment to contain genes with low CGN. While these observations about chromosomal distribution are intriguing, the biological and evolutionary implications of such patterns remain to be addressed in future studies. For

instance, it is well known that a significant part of the human chromosome 17 maps to a single large segment on chromosome 11 in mouse, highlighting the high conservation of genomic neighborhood of genes encoded in this chromosome during mammalian evolution (Zody et al. 2006).

### The impact of structural polymorphisms

Though the proportion of genes (~18% or 3152 genes) that have altered their neighborhood either in human or in chimpanzee appears to be substantial, it is not surprising since a comparable number of genes experience insertions, deletions and copy number variations in the current human population (Sharp et al. 2006) and after the split from the common ancestor of human and chimpanzee (Demuth et al. 2006). It should be noted that not all cases of alteration of genomic neighborhood reported in our study may be fixed in the human or chimpanzee lineage, as some of them might reside in structurally polymorphic regions in the current human and chimpanzee populations. To assess the impact of structural polymorphisms, we analyzed several recent data sets describing (1) copy number variations that are specific to human, specific to chimpanzee, and shared between the two species (Perry et al. 2008), (2) insertion and deletion polymorphisms in human populations (Iafate et al. 2004), (3) segmental duplications in human (She et al. 2004), and (4) inversion polymorphisms in human populations (Bansal et al. 2007) (see Supplemental Table S5). We found that ~350 genes are affected by segmental duplication, ~350 genes by CNV (insertion and deletion), and ~60 genes by inversion polymorphism. While the distribution of CGN scores of genes that reside in regions showing structural polymorphism tends to be lower when compared to the genome-wide distribution, removal of these genes did not qualitatively affect our conclusions discussed later. Our findings suggest incomplete fixation of these events and that such events may contribute to variation in gene expression within the current human and chimpanzee populations. We also note that the existing data on structural polymorphisms are not complete and therefore believe that high-resolution, population-level structural variation data for both species will be necessary to assess the extent of fixation and the true impact of structural polymorphisms on the reported observations.

### Genes with altered neighborhood tend to display high expression divergence

Given that a change in the genomic neighborhood can affect the expression of a gene (Kleinjan and van Heyningen 1998, 2005; Xiao et al. 2008), we investigated whether an alteration in genomic neighborhood is linked with gene expression divergence on a genome-wide scale.

### Expression divergence in different tissues

We compared the CGN scores of human–chimpanzee orthologous gene pairs with the extent of expression divergence of the orthologous genes in equivalent organs using the data set from Khaitovich et al. (2005). In short, Khaitovich and colleagues compared expression levels of ~21,000 orthologous genes from six humans and five chimpanzees using probes that are identical between the two organisms on the U133plus2 Affymetrix gene chip. The expression levels of these genes from five different tissues, i.e., heart, kidney, liver, brain, and testis were analyzed from both organisms. For 9248 genes, which were detected as being

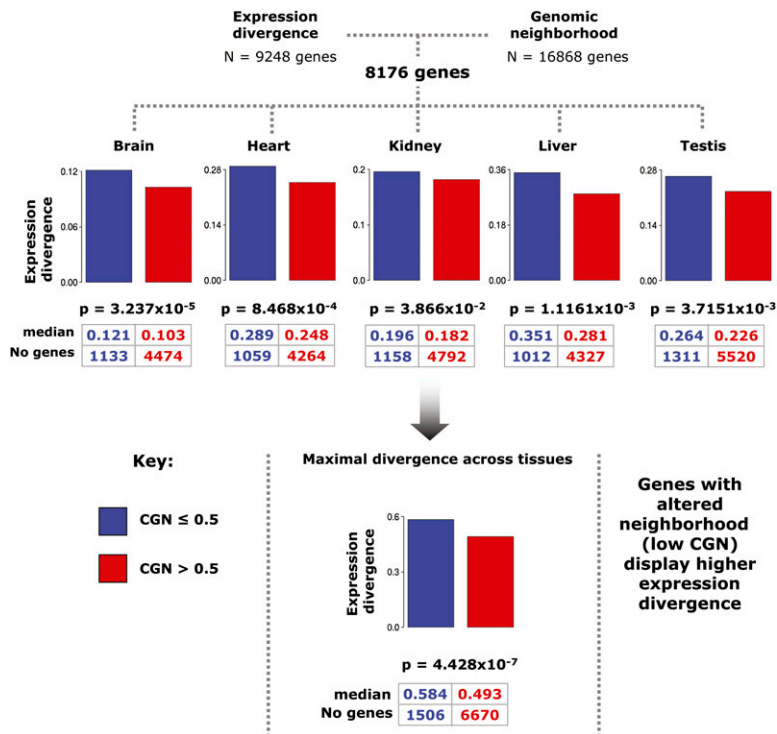
expressed in at least one of the tissues, Khaitovich et al. (2005) reported expression divergence values (see Methods).

Integration of the CGN score with expression divergence data of the genes for each tissue (Fig. 3) revealed the following trend: When the gene neighborhood is altered (i.e., low CGN), orthologous genes are more likely to undergo expression divergence (see Supplemental Table S6). Interestingly, the trend is most significant for the genes that are expressed in the brain ( $P$ -value =  $3.23 \times 10^{-5}$ , Fig. 3). More importantly, the pattern that we discovered is consistent and is statistically significant across the five different tissue types studied (Fig. 3; see Supplemental Table S6). In fact, the significance of the reported trend becomes higher when we compared the highest value of expression divergence among the five tissues against the CGN score for the genes ( $P$ -value =  $4.4 \times 10^{-7}$ , Fig. 3; see Supplemental Table S6). Conversely, an analysis of the data after grouping genes, which showed expression divergence between human and chimpanzee (i.e., controlling for expression level), revealed that if the expression levels are different, then the gene neighborhood is more likely to be altered compared to genes that had similar expression levels (see Supplemental Table S6). These results collectively suggest that a change in genomic neighborhood is likely to introduce alterations in expression pattern during human evolution. While the CGN score provides an estimate of the degree of change in the neighborhood of a gene, its relationship with the extent of expression divergence need not be linear for several reasons: (1) if the event (e.g., insertion, deletion, inversion, or translocation) that changed the neighborhood of the gene did not alter the *cis*-regulatory elements around the gene or (2) if it did not result in alterations to the higher order chromatin structures (e.g., euchromatin or heterochromatin) or nucleosome modification pattern (e.g., histone modification or DNA methylation), the effects are likely to be neutral. Therefore, while a large change in the neighborhood increases the chance of involving regulatory changes, the extent of gene expression divergence need not be proportional.

It should be noted that the results presented here are not affected by the cutoff used for identifying genes with altered ( $CGN \leq 0.5$ ) and conserved ( $CGN > 0.5$ ) neighborhood as we obtain similar results when we analyze the (1) the top 10% of the genes with low CGN ( $CGN < 0.416$ ) and high CGN ( $CGN > 0.775$ ) scores or (2) when we chose the median value ( $CGN = 0.637$ ) to group genes into the low CGN ( $CGN \leq 0.637$ ) and high CGN ( $CGN > 0.637$ ) group (see Supplemental Table S7). This suggests that the reported findings are robust to the definition used for identifying genes with altered or conserved neighborhood.

If we consider only those genes with conserved neighborhood, about half the genes display high expression divergence (see Supplemental Table S6). In addition, considering only those genes with high expression divergence, about three-quarters have conserved gene neighborhood (see Supplemental Table S6). These observations are not surprising and only support the fact that factors other than alterations to gene neighborhood, such as *trans*- and *cis*-regulatory mutations, contribute to expression divergence. In total, the observations described above are consistent with our understanding that gene expression divergence is not driven by a single factor, but by multiple mechanisms (Wray et al. 2003).

Taken together, our findings linking expression divergence and change in genomic neighborhood in human and chimpanzee implicate higher order effects, such as local chromosomal and chromatin environment in transcriptome evolution. More importantly, our findings suggest that such mechanisms act in combination with changes that affect *trans*- and *cis*-regulatory



**Figure 3.** Assessing the link between alteration in gene neighborhood and gene expression divergence. Distributions of the expression divergence value for genes with conserved (red) and altered (blue) neighborhood are compared using the Mann-Whitney test for five tissue types: brain, heart, kidney, liver, and testis. For visual clarity, median expression and sample size (i.e., number of genes) for genes with conserved and altered neighborhood categories are shown for each tissue. Y-axis is in log-scale. Maximal expression divergence values for a gene across all five tissues versus maximal expression divergence values for these two categories are compared using the Mann-Whitney test and the result is shown in the *bottom* panel. See Supplemental Table S7 for control calculations.

elements, in order to drive gene expression divergence. It is important to emphasize that the data on expression divergence were available for only five organs, which are in fact made up of multiple cell types. The pattern of expression divergence in other tissues remains unknown, and the estimates of expression divergence are likely to be dampened due to the presence of multiple cell types (i.e., higher noise-to-signal ratio compared to populations of pure cell types from equivalent tissues). Therefore, the findings reported here should be treated as a conservative estimate of the impact of alteration of genomic neighborhood on transcriptome evolution, which are nevertheless significant.

### Expression divergence in different brain parts

Having made the observation that the genes with altered neighborhood are more likely to undergo expression divergence and that this trend was most significant in the brain, we investigated if the individual parts within the brain also showed a similar trend. To investigate this question, we used the data on expression divergence between human and chimpanzee for the different brain parts from Khaitovich et al. (2004). In short, the authors measured gene expression levels using the U95AV2, U95B, U95C, U95D, and U95E gene chips and reported expression divergence in six equivalent brain parts (prefrontal cortex, primary visual cortex, anterior cingulate cortex, caudate nucleus, cerebellum, and Broca's area) from human and chimpanzee subjects. Expression divergence data in at least one of the six tissue types were available

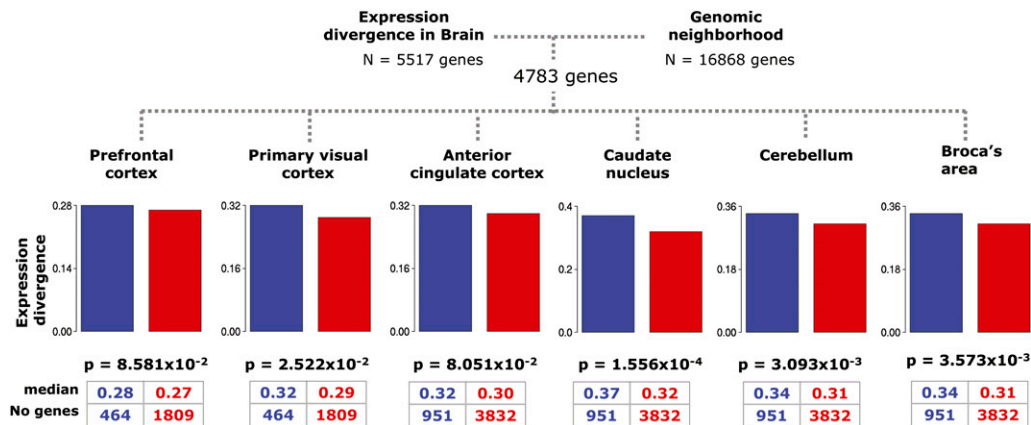
for 5517 genes (Ensembl v48), for which 4783 also had a CGN score available (see Methods).

Consistent with the results of the tissue-level analysis described in the previous section (Fig. 3), we found that when the genomic neighborhood is altered, orthologous genes expressed in equivalent brain parts are more likely to diverge in their expression level. More importantly, this pattern is consistent and statistically significant for all the six brain parts analyzed (Fig. 4). While the caudate nucleus, which plays an important role in learning and memory showed the highest statistical significance ( $P$ -value =  $1.55 \times 10^{-4}$ , Mann-Whitney test), the prefrontal cortex, which plays a role in planning complex cognitive behavior and in personality expression, showed marginal significance ( $P$ -value =  $9.2 \times 10^{-2}$ , Mann-Whitney test). The other parts still showed the same trend whose significance lie between these two values. The results described here were qualitatively similar when we used alternative CGN cutoffs to identify genes with altered or conserved neighborhood (see Supplemental Table S8). We did not analyze genes that are specifically expressed exclusively in one brain part as the number was too low to estimate statistical significance reliably.

While several factors (e.g., purity of the RNA samples, age of the human and chimpanzee subjects, sample preparation, storage, etc.) can increase the noise in expression data, it is difficult to control for all these possible factors. However, the consistency in the observed trends from the five different tissues and the six brain parts, taken from two independent experiments, imply that such factors are unlikely to have biased our conclusions. Thus, these observations suggest that alteration in genomic neighborhood is an important factor for expression divergence in the various tissues, including the different brain parts.

### Human brain shows an enrichment to express genes with low CGN in a tissue-specific manner

Having identified genes with altered neighborhood and established a link with expression divergence, we focused on where such genes are expressed in humans. In particular, we investigated if any tissue (1) showed an enrichment to express genes with low CGN and (2) had an enrichment to express genes with low CGN in a tissue-specific manner. To address these questions, we obtained the expression levels for 33,495 transcripts from 17,185 genes across 72 noncancerous human tissues from Su et al. (2004). CGN scores and expression levels were available for 13,963 genes of which 2574 genes had a low CGN. A transcript was considered to be expressed in a tissue if its expression level in that tissue was above the median value of all genes across all tissues. A transcript was considered to be expressed in a tissue-specific manner if it was expressed in less than 10% of the 72 tissue and cell types (i.e.,



**Figure 4.** Investigating the impact of alteration in gene neighborhood on gene expression divergence in different brain parts. Distributions of the expression divergence value for genes with conserved (red) and altered (blue) neighborhood are compared using the Mann-Whitney test for six different brain parts. Y-axis is in log-scale. For visual clarity, median expression and sample size (i.e., number of genes) for genes with conserved and altered neighborhood categories are shown for each brain part. See Supplemental Table S8 for control calculations.

expression breadth < 0.10). This resulted in the identification of 4009 transcripts that showed a strong tissue-specific expression pattern of which 683 map to genes with altered neighborhood (i.e., low CGN).

To investigate the first question, i.e., if any tissue shows an enrichment to express genes with low CGN, we first calculated the ratio of the number of transcripts from genes with low CGN that were expressed in a particular tissue to the total number of transcripts that were expressed in that tissue. An investigation of this ratio revealed, as expected, that on average ~18% of all transcripts that were expressed in a given tissue map to genes with low CGN, representing the average behavior of ~18% of genes with low CGN. This value was comparable across the different tissues suggesting that most tissues express genes with low CGN to a similar extent (Fig. 5A). To investigate the second question, i.e., whether any tissue showed an enrichment to express genes with low CGN in a tissue-specific manner, we calculated the proportion of transcripts of genes with low CGN among the total number of tissue specifically expressed transcripts in that tissue (Fig. 5B). One consistent theme, which stood out in this analysis, was that most (16 out of 17) brain tissues expressed an above-median proportion of tissue-specific transcripts from genes with low CGN when compared to all other tissues (Fig. 5; Supplemental Table S9) with the prefrontal cortex showing the highest enrichment (Supplemental Table S9). Expression data were not available for several of the other brain parts (e.g., caudate nucleus), which were analyzed in the previous section.

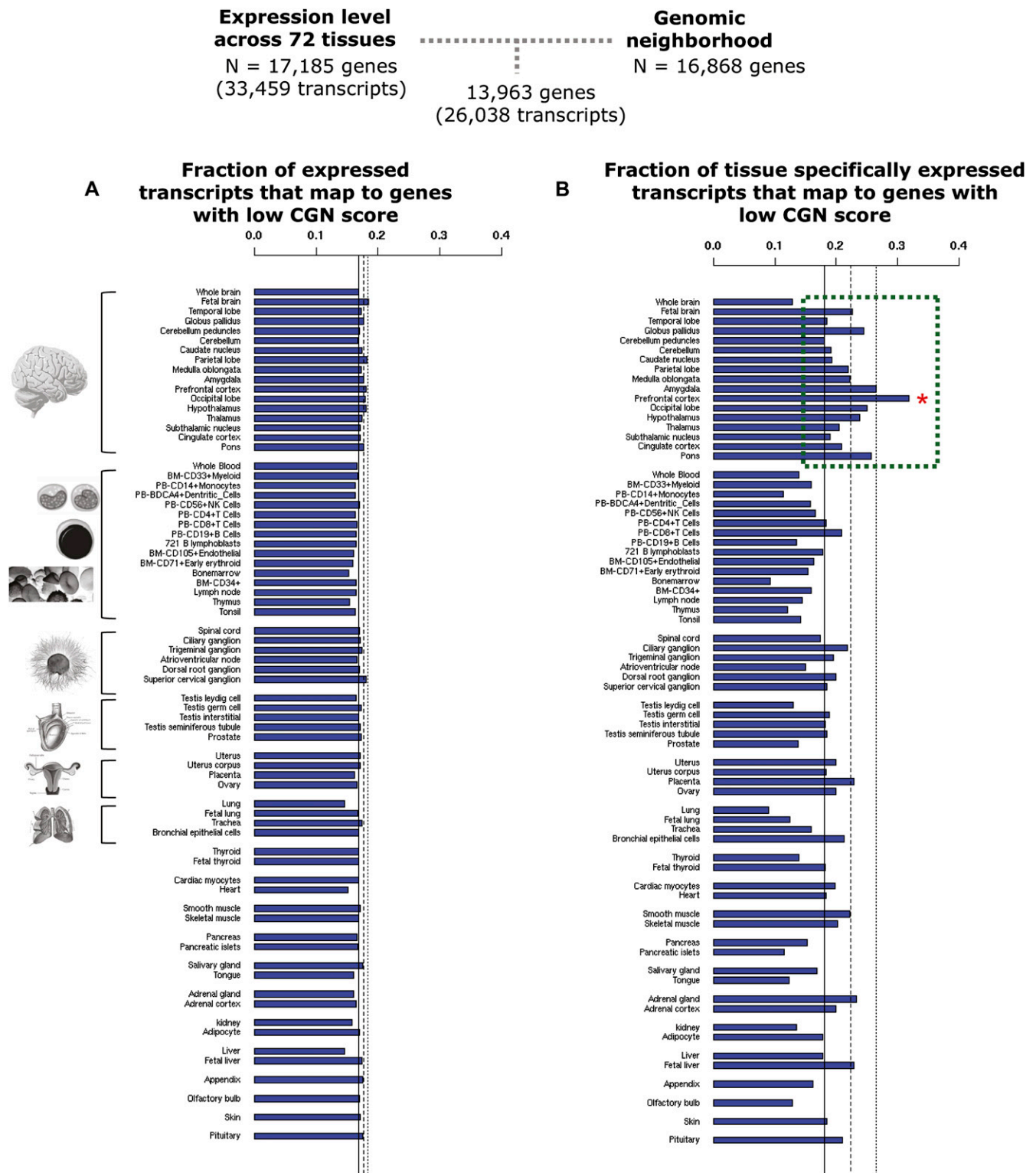
The results presented here are consistent when a different threshold was used to identify transcripts that were expressed in a tissue-specific manner, i.e., transcripts with expression breadth of <0.075 (genes expressed in five or fewer tissues) and <0.015 (genes expressed only in one tissue) (see Supplemental Table S9), suggesting that our observations are generally robust to the choice of parameters used. The list of genes with alterations in their genomic neighborhood and that are expressed in a tissue-specific manner are provided in Supplemental Table S9. It should be stressed that the genes with low CGN could have altered genomic neighborhood either in human or in chimpanzee lineage, thereby resulting in altered expression in that particular lineage. Since the data on gene expression in the different chimpanzee tissues are not available, it would not be currently possible to identify the

impact of alteration of genomic neighborhood on tissue-specific expression pattern in chimpanzee. However, we believe that the specific candidate genes identified in this study provide ideal starting points to independently investigate the evolutionary consequences of alteration of genomic neighborhood for future studies.

## Discussion

In summary, our work provides the first genome-wide view of the impact of change in genomic neighborhood on expression divergence during the evolution of human and chimpanzee. By integrating genomic and transcriptomic data for human and chimpanzee, we show that (1) a considerable fraction of genes have changed their genomic neighborhood in human and chimpanzee after the split from the common ancestor, (2) this is linked to divergence in gene expression levels of such genes in the different tissues and brain parts, and (3) the human brain shows enrichment to express such genes in a tissue-specific manner. These observations reveal the genome-wide impact of alteration of genomic neighborhood on expression divergence.

Though several factors (e.g., *cis*-regulatory elements, *trans*-factors and genomic context) contribute to expression divergence (Supplemental Table S10), it is difficult to quantify and tease apart the contribution of individual factors for several reasons. First, identification of true *cis*-regulatory elements in human promoters and linking their mutation patterns to expression divergence is not straightforward and only limited work has addressed issues related to this question (Haygood et al. 2007; Prud'homme et al. 2007; Wray 2007). Second, while studies by Wittkopp et al. (2004, 2008) have compared the relative contribution of *cis*- and *trans*-mutations in gene expression divergence for a set of genes in fruitfly, no such study at the genome scale has been carried out for humans. Third, changes in the neighborhood of a gene of interest (as measured by our CGN metric) will also be associated with potential alterations in non-coding regulatory elements. For instance, it was recently reported that single nucleotide mutation rate increases close to sites of insertion or deletion of genetic material in eukaryotes (Tian et al. 2008). Finally, gene expression divergence between human and chimpanzee has been studied only for a small number of tissues



**Figure 5.** Distribution of the fraction of transcripts that map to genes with low CGN which are expressed (A) and expressed in a tissue-specific manner (expression breadth < 0.1) (B) for the various tissues. 13,963 genes for which expression data and CGN score are available were investigated. (A) The ratio of transcripts that map to genes with low CGN to all the transcripts expressed in that tissue is shown for each of the 72 tissue types. (B) The ratio of tissue-specific transcripts from genes with low CGN to all transcripts that are expressed in a tissue-specific manner in that tissue is shown. The mean value across all tissues is shown by a continuous line and the first standard deviation represented by a broken line. Prefrontal cortex, which has the highest over-representation of tissue-specific expressed transcripts that map to genes with altered neighborhoods, is marked with a red asterisk. All brain tissues with above-mean proportion of tissue-specific transcripts from genes with altered neighborhoods are highlighted by a dashed green box.

(Khaltovich et al. 2004, 2005, 2006). Therefore, the currently available data and literature on expression divergence in humans and the role of these factors on transcriptome evolution on humans are insufficient to make a rigorous comparison of the different effects.

The findings presented in this study illustrate the evolutionary significance of an emerging concept: the local chromosomal and chromatin environment and spatial organization of the genome in the nucleus affect gene expression (Misteli 2004; Fraser and Bickmore 2007; Lanctot et al. 2007; Babu et al. 2008). For instance, a recent study demonstrated that the human genome is organized into discrete domains of transcriptional activity (Guelen et al. 2008). Another study (Gierman et al. 2007) reported that identical green fluorescent protein (GFP) reporter constructs integrated at several different chromosomal positions showed expression levels that correspond to the activity of the domain of integration. Given the existence of such domains of transcriptional activity, alterations in gene neighborhood could well lead to changes in gene expression pattern, thereby contributing to the transcriptome-level evolution. Interestingly, since most multicellular eukaryotes are diploid (or polyploid) a change in genomic neighborhood of one copy of a gene could contribute to allele-specific expression divergence between individuals in a population. Given the recent discovery that mono-allelic expression is a widespread phenomenon (Gimelbrant et al. 2007) and can have important consequences (e.g., issues with dosage balance; Birchler et al. 2005), our findings also suggest a possible mechanism that contributes to creating diversity in individual cell types.

Some previous studies have suggested that gene expression divergence is largely neutral (Khaltovich et al. 2006), while others have suggested that they may be selected during evolution (Haygood et al. 2007) and this has been a topic of intense debate in recent times (Pennisi 2008). Our observation that change in neighborhood is linked with gene expression divergence, together with our finding that the tissue-specific genes in the human brain tend to have altered genomic neighborhood, prompts us to speculate that expression divergence due to alterations in neighborhood could be one of the factors that contributed to the differences in cognitive abilities and other traits between human and chimpanzee. This hypothesis is particularly interesting in light of the fact that very little evidence is found for the differences in coding sequences between orthologous genes, which could explain the disparity in cognitive abilities of human and chimpanzee (The Chimpanzee Sequencing and Analysis Consortium 2005; Khaltovich et al. 2005). Although most of the observed gene expression divergence might be neutral, those which confer a fitness advantage would be selected during evolution. Even if the effects of gene expression divergence may be subtle, a minor change in expression of a regulator, pleiotropic, or developmental gene could have major phenotypic consequences (Abzhanov et al. 2004, 2006; Wu et al. 2004). This is also likely to be true for genes that function as evolutionary capacitors (Yeyati et al. 2007; Levy and Siegal 2008; Yeyati and van Heyningen 2008) and those that display expression-level dependent phenotypes (Dipple and McCabe 2000; Van Heyningen and Yeyati 2004).

To infer the real impact of alterations in genomic neighborhood of specific genes, carefully designed experiments are required as the effects are likely to be context dependent and subtle. In this regard, we believe that the candidate genes identified in our work will serve as a good starting point for future studies.

## Methods

### Data set

Chromosomal location of the genes and other relevant information for human (*Homo sapiens*, NCBI 36), chimpanzee (*Pan troglodytes*, PanTro 2.1), and macaque (*Macaca mulatta*, MMUL1.0) genomes were obtained from Ensembl Release 48 (Hubbard et al. 2007). Orthologs of human genes were obtained from Ensembl-Compara Release 48, which identifies orthologs by building phylogenetic trees of homologs using a maximum likelihood approach (Hubbard et al. 2007).

### Calculation of the conservation of genomic neighborhood score

We define a metric called conservation of genomic neighborhood (CGN) (see Fig. 2A) to estimate the extent to which the neighborhood of a gene is maintained between the species of interest (human) and the reference species (chimpanzee or macaque). For every human gene  $A^{HS}$  that has an ortholog  $A^{PT}$  in chimpanzee, we first identify all the neighbors ( $N_A^{HS}$ ) of  $A^{HS}$  within a window ( $w$ ) of 2 Mb centering on the gene. We then identify all the neighbors ( $N_A^{PT}$ ) of the chimpanzee ortholog  $A^{PT}$  within a window of 2 Mb and identify how many genes in  $N_A^{HS}$  have orthologs in  $N_A^{PT}$ , denoted as  $N_A$ . The CGN score for the gene  $A^{HS}$  is computed as  $N_A / N_A^{HS}$ . In other words, the CGN score represents the proportion of neighboring genes that is conserved between the two species within 2 Mb around a gene of interest. It should be noted that because of the way in which CGN score is calculated, the identification of genes with low CGN would include both the set of genes that have changed their neighborhood either in the human lineage or in the chimpanzee lineage after the split from their common ancestor. For a highly conserved neighborhood, the CGN score is closer to 1, and approaches 0 if the genomic neighborhood is completely altered. Since most of the primate duplication events, insertions or deletions via transposable elements are less than 1 Mb (Bailey and Eichler 2006) and because such events affect the genomic neighborhood, we chose a conservative window size of 2 Mb to calculate the CGN score of orthologous genes to capture such effects. Genes with CGN scores of  $\leq 0.5$  or  $> 0.5$  were defined as those that either altered or conserved their genomic neighborhood, respectively. See Supplemental Table S3 for the list of genes and their CGN scores.

### Quality control and control calculations for CGN score

CGN scores were computed for all the human–chimpanzee orthologs after removing (1) genes residing in regions of incomplete assembly, (2) one-to-many, many-to-one and many-to-many orthologs, and (3) genes residing in the band next to the telomeres and centromeres of all chromosomes (see Supplemental Table S2). To identify spurious cases of genes that appeared to have a low CGN score due to genome assembly errors, or arising from complications with ortholog identification, we computed the CGN values using macaque as a reference species (see Supplemental Table S2). By comparing the CGN scores obtained using chimpanzee and macaque as a reference species, we identified and removed such problematic genes (< 1% of all genes) from further analysis (see Supplemental Table S2). To ensure that the CGN score for a gene obtained using the above definition was not sensitive to the choice of the parameters used, we recalculated the scores using alternative methods: (1) by using different values of the window size (i.e.,  $w = 1$  Mb or 3 Mb) and (2) by using a variable window size but a fixed number of neighboring genes around the gene of interest (i.e.,  $N_A^{HS} = 30$  or 40 genes). The CGN score obtained using

these definitions were largely similar suggesting that the CGN scores are generally robust to the choice of the parameter used (see Supplemental Table S1).

### Structural variation analysis

Copy number variations specific to human and chimpanzee lineages and those shared between the two species were obtained from Perry et al. (2008). The positions of human segmental duplications were obtained from the Segmental Duplication Database (She et al. 2004). Positions for insertions and deletions in the human populations were obtained from the Database of Genomic Variants where InDel information was culled from several published studies (Iafate et al. 2004). Inversion polymorphism for the human HapMap population was obtained from Bansal et al. (2007).

### Expression divergence analysis for the five tissue types

Data on gene expression divergence were obtained from Khaitovich et al. (2005). The authors investigated expression divergence between human and chimpanzee using the U133plus2 Affymetrix gene chip containing probes for ~21,000 genes. Expression divergence values were available for 9248 orthologous genes in at least one of the five tissues (brain, heart, kidney, liver, and testis). Expression divergence was defined by the authors as an average squared difference in normalized gene expression across all probes with detectable gene expression between species. A gene was defined as differently expressed in both human and chimpanzee in a particular tissue if it was detected and had a consistent direction of change in all pairwise comparisons of the probes. Genes numbering 8176 had both CGN score and expression divergence data available (see Supplemental Table S6 for details of the calculation and an extended discussion). The results described here were similar when we used alternative cutoffs to identify genes with altered or conserved neighborhood (see Supplemental Table S7).

### Expression divergence analysis in six brain parts

Data on expression divergence of genes in six brain parts were obtained from Khaitovich et al. (2004). The authors investigated expression level of transcripts using the U95AV2, U95B, U95C, U95D, and U95E gene chips and reported expression divergence value for six equivalent brain parts (prefrontal cortex, primary visual cortex, anterior cingulate cortex, caudate nucleus, cerebellum, and Broca's area) between human and chimpanzee. Average signal log ratio between human and chimpanzee (AVE\_SLR) and standard deviation (SD) of measurement for each probe was available. We excluded the probes, which could not be mapped to Ensembl genes (v48) and those that mapped to multiple genes. If multiple probes map to the same genes, we observed that the extent and direction of expression divergence between such probes were not necessarily consistent. Therefore, we excluded such probes (and the corresponding genes) from further analysis. Expression divergence data in at least one of the six tissue types were available for 5517 genes, of which 4783 also had the CGN score calculated. Statistical analysis was performed on the genes for which the signal-to-noise ratio is not very low (AVE\_SLR/SD > 0.5). The results described here were similar when we used alternative cutoffs to identify genes with altered or conserved neighborhood (see Supplemental Table S8).

### Tissue-specific expression analysis

Mas5 condensed expression data (GNF1H) for 33,689 probes from 72 normal human tissues were obtained from Su et al. (2004).

Probes were mapped onto 33,495 transcripts from 17,185 human genes, of which 26,038 transcripts from 13,963 genes had CGN score available. A transcript was considered to show tissue-specific expression if it was expressed in at least one tissue and had above-median expression (>321 a.u.) in less than 10% (i.e., expression breadth < 0.1) of the tissue types considered (see Supplemental Table S9). The results presented here were similar when a different threshold was used to identify transcripts that were expressed in a tissue-specific manner, i.e., expression breadth of <0.075 (expressed in five or fewer tissues) and <0.015 (expressed in only one tissue) (see Supplemental Table S9).

### Acknowledgments

We thank MRC-LMB for support. S.D. is a recipient of the LMB Cambridge Scholarship and acknowledges King's College, S.A.T. acknowledges Trinity College, and M.M.B. acknowledges Darwin College and Schlumberger for generous support. We thank the anonymous referees, J. Gsponer, D. Hebenstreit, A. Wuster, A. Pombo, B. Lang, D. Rubinsztein, A.M. Arias, S. Janga, K. Weber, C. Chothia, P. Dear, and A. Travers for providing helpful comments on the manuscript.

### References

- Abzhanov, A., Protas, M., Grant, B.R., Grant, P.R., and Tabin, C.J. 2004. Bmp4 and morphological variation of beaks in Darwin's finches. *Science* **305**: 1462–1465.
- Abzhanov, A., Kuo, W.P., Hartmann, C., Grant, B.R., Grant, P.R., and Tabin, C.J. 2006. The calmodulin pathway and evolution of elongated beak morphology in Darwin's finches. *Nature* **442**: 563–567.
- Babu, M.M., Janga, S.C., de Santiago, I., and Pombo, A. 2008. Eukaryotic gene regulation in three dimensions and its impact on genome evolution. *Curr Opin Genet Dev* **18**: 571–582.
- Bailey, J.A. and Eichler, E.E. 2006. Primate segmental duplications: Crucibles of evolution, diversity and disease. *Nat. Rev. Genet.* **7**: 552–564.
- Bansal, V., Bashir, A., and Bafna, V. 2007. Evidence for large inversion polymorphisms in the human genome from HapMap data. *Genome Res.* **17**: 219–230.
- Birchler, J.A., Riddle, N.C., Auger, D.L., and Veitia, R.A. 2005. Dosage balance in gene regulation: Biological implications. *Trends Genet.* **21**: 219–226.
- Carroll, S.B. 2003. Genetics and the making of *Homo sapiens*. *Nature* **422**: 849–857.
- The Chimpanzee Sequencing and Analysis Consortium. 2005. Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature* **437**: 69–87.
- Demuth, J.P., De Bie, T., Stajich, J.E., Cristianini, N., and Hahn, M.W. 2006. The evolution of mammalian gene families. *PLoS One* **1**: e85. doi: 10.1371/journal.pone.0000085.
- Dipple, K.M. and McCabe, E.R. 2000. Phenotypes of patients with "simple" Mendelian disorders are complex traits: Thresholds, modifiers, and systems dynamics. *Am. J. Hum. Genet.* **66**: 1729–1735.
- Felsenfeld, G. and Groudine, M. 2003. Controlling the double helix. *Nature* **421**: 448–453.
- Fraser, P. and Bickmore, W. 2007. Nuclear organization of the genome and the potential for gene regulation. *Nature* **447**: 413–417.
- Gierman, H.J., Indemans, M.H., Koster, J., Goetze, S., Seppen, J., Geerts, D., van Driel, R., and Versteeg, R. 2007. Domain-wide regulation of gene expression in the human genome. *Genome Res.* **17**: 1286–1295.
- Gimelbrant, A., Hutchinson, J.N., Thompson, B.R., and Chess, A. 2007. Widespread monoallelic expression on human autosomes. *Science* **318**: 1136–1140.
- Guelen, L., Pagie, L., Brasset, E., Meuleman, W., Faza, M.B., Talhout, W., Eussen, B.H., de Klein, A., Wessels, L., de Laat, W., et al. 2008. Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature* **453**: 948–951.
- Han, J.S., Szak, S.T., and Boeke, J.D. 2004. Transcriptional disruption by the L1 retrotransposon and implications for mammalian transcriptomes. *Nature* **429**: 268–274.
- Haygood, R., Fedrigo, O., Hanson, B., Yokoyama, K.D., and Wray, G.A. 2007. Promoter regions of many neural- and nutrition-related genes have

- experienced positive selection during human evolution. *Nat. Genet.* **39**: 1140–1144.
- Hubbard, T.J., Aken, B.L., Beal, K., Ballester, B., Caccamo, M., Chen, Y., Clarke, L., Coates, G., Cunningham, F., Cutts, T., et al. 2007. Ensembl 2007. *Nucleic Acids Res.* **35**: D610–D617.
- Hurles, M.E., Dermitzakis, E.T., and Tyler-Smith, C. 2008. The functional impact of structural variation in humans. *Trends Genet.* **24**: 238–245.
- Iafraite, A.J., Feuk, L., Rivera, M.N., Listewnik, M.L., Donahoe, P.K., Qi, Y., Scherer, S.W., and Lee, C. 2004. Detection of large-scale variation in the human genome. *Nat. Genet.* **36**: 949–951.
- Joos, S., Falk, M.H., Lichter, P., Haluska, F.G., Henglein, B., Lenoir, G.M., and Bornkamm, G.W. 1992. Variable breakpoints in Burkitt lymphoma cells with chromosomal t(8;14) translocation separate c-myc and the IgH locus up to several hundred kb. *Hum. Mol. Genet.* **1**: 625–632.
- Jordan, I.K., Marino-Ramirez, L., and Koonin, E.V. 2005. Evolutionary significance of gene expression divergence. *Gene* **345**: 119–126.
- Khaitovich, P., Muetzel, B., She, X., Lachmann, M., Hellmann, I., Dietzsch, J., Steigle, S., Do, H.H., Weiss, G., Enard, W., et al. 2004. Regional patterns of gene expression in human and chimpanzee brains. *Genome Res.* **14**: 1462–1473.
- Khaitovich, P., Hellmann, I., Enard, W., Nowick, K., Leinweber, M., Franz, H., Weiss, G., Lachmann, M., and Paabo, S. 2005. Parallel patterns of evolution in the genomes and transcriptomes of humans and chimpanzees. *Science* **309**: 1850–1854.
- Khaitovich, P., Enard, W., Lachmann, M., and Paabo, S. 2006. Evolution of primate gene expression. *Nat. Rev. Genet.* **7**: 693–702.
- King, M.C. and Wilson, A.C. 1975. Evolution at two levels in humans and chimpanzees. *Science* **188**: 107–116.
- Kleinjan, D.J. and van Heyningen, V. 1998. Position effect in human genetic disease. *Hum. Mol. Genet.* **7**: 1611–1618.
- Kleinjan, D.A. and van Heyningen, V. 2005. Long-range control of gene expression: Emerging mechanisms and disruption in disease. *Am. J. Hum. Genet.* **76**: 8–32.
- Lancot, C., Cheutin, T., Cremer, M., Cavalli, G., and Cremer, T. 2007. Dynamic genome architecture in the nuclear space: Regulation of gene expression in three dimensions. *Nat. Rev. Genet.* **8**: 104–115.
- Levy, S.F. and Siegal, M.L. 2008. Network hubs buffer environmental variation in *Saccharomyces cerevisiae*. *PLoS Biol.* **6**: e264. doi: 10.1371/journal.pbio.0060264.
- Misteli, T. 2004. Spatial positioning: A new dimension in genome function. *Cell* **119**: 153–156.
- Pennisi, E. 2008. Evolutionary biology: Deciphering the genetics of evolution. *Science* **321**: 760–763.
- Perry, G.H., Yang, F., Marques-Bonet, T., Murphy, C., Fitzgerald, T., Lee, A.S., Hyland, C., Stone, A.C., Hurles, M.E., Tyler-Smith, C., et al. 2008. Copy number variation and evolution in humans and chimpanzees. *Genome Res.* **18**: 1698–1710.
- Prud'homme, B., Gompel, N., and Carroll, S.B. 2007. Emerging principles of regulatory evolution. *Proc. Natl. Acad. Sci.* **104** (Suppl 1): 8605–8612.
- Rose, C.S., Patel, P., Reardon, W., Malcolm, S., and Winter, R.M. 1997. The *TWIST* gene, although not disrupted in Saethre-Chotzen patients with apparently balanced translocations of 7p21, is mutated in familial and sporadic cases. *Hum. Mol. Genet.* **6**: 1369–1373.
- Sharp, A.J., Cheng, Z., and Eichler, E.E. 2006. Structural variation of the human genome. *Annu. Rev. Genomics Hum. Genet.* **7**: 407–442.
- She, X., Jiang, Z., Clark, R.A., Liu, G., Cheng, Z., Tuzun, E., Church, D.M., Sutton, G., Halpern, A.L., and Eichler, E.E. 2004. Shotgun sequence assembly and recent segmental duplications within the human genome. *Nature* **431**: 927–930.
- Su, A.I., Wiltshire, T., Batalov, S., Lapp, H., Ching, K.A., Block, D., Zhang, J., Soden, R., Hayakawa, M., Kreiman, G., et al. 2004. A gene atlas of the mouse and human protein-encoding transcriptomes. *Proc. Natl. Acad. Sci.* **101**: 6062–6067.
- Tian, D., Wang, Q., Zhang, P., Araki, H., Yang, S., Kreitman, M., Nagylaki, T., Hudson, R., Bergelson, J., and Chen, J.Q. 2008. Single-nucleotide mutation rate increases close to insertions/deletions in eukaryotes. *Nature* **455**: 105–108.
- Van Heyningen, V. and P.L. Yeyati. 2004. Mechanisms of non-Mendelian inheritance in genetic disease. *Hum Mol Genet* **13**: R225–R233.
- Velagaleti, G.V., Bien-Willner, G.A., Northup, J.K., Lockhart, L.H., Hawkins, J.C., Jalal, S.M., Withers, M., Lupski, J.R., and Stankiewicz, P. 2005. Position effects due to chromosome breakpoints that map approximately 900 kb upstream and approximately 1.3 Mb downstream of *SOX9* in two patients with campomelic dysplasia. *Am. J. Hum. Genet.* **76**: 652–662.
- West-Eberhard, M.J. 2005. Developmental plasticity and the origin of species differences. *Proc. Natl. Acad. Sci.* **102** (Suppl 1): 6543–6549.
- White, K.P. 2001. Functional genomics and the study of development, variation and evolution. *Nat. Rev. Genet.* **2**: 528–537.
- Wittkopp, P.J., Haerum, B.K., and Clark, A.G. 2004. Evolutionary changes in *cis* and *trans* gene regulation. *Nature* **430**: 85–88.
- Wittkopp, P.J., Haerum, B.K., and Clark, A.G. 2008. Regulatory changes underlying expression differences within and between *Drosophila* species. *Nat. Genet.* **40**: 346–350.
- Wray, G.A. 2007. The evolutionary significance of *cis*-regulatory mutations. *Nat. Rev. Genet.* **8**: 206–216.
- Wray, G.A., Hahn, M.W., Abouheif, E., Balhoff, J.P., Pizer, M., Rockman, M.V., and Romano, L.A. 2003. The evolution of transcriptional regulation in eukaryotes. *Mol. Biol. Evol.* **20**: 1377–1419.
- Wu, P., Jiang, T.X., Suksaweang, S., Widelitz, R.B., and Chuong, C.M. 2004. Molecular shaping of the beak. *Science* **305**: 1465–1466.
- Xiao, H., Jiang, N., Schaffner, E., Stockinger, E.J., and van der Knaap, E. 2008. A retrotransposon-mediated gene duplication underlies morphological variation of tomato fruit. *Science* **319**: 1527–1530.
- Yeyati, P.L. and van Heyningen, V. 2008. Inactivating the evolutionary capacitor: Hsp90 modulation of disease. *Curr. Opin. Genet. Dev.* **18**: 264–272.
- Yeyati, P.L., Banciewicz, R.M., Maule, J., and van Heyningen, V. 2007. Hsp90 selectively modulates phenotype in vertebrate development. *PLoS Genet.* **3**: e43. doi: 10.1371/journal.pgen.0030043.
- Zody, M.C., Garber, M., Adams, D.J., Sharpe, T., Harrow, J., Lupski, J.R., Nicholson, C., Searle, S.M., Wilming, L., Young, S.K., et al. 2006. DNA sequence of human chromosome 17 and analysis of rearrangement in the human lineage. *Nature* **440**: 1045–1049.

Received September 4, 2008; accepted in revised form January 14, 2009.