Induced gene expression in human brain after the split from chimpanzee

Jianying Gu and Xun Gu

Dept of Zoology and Genetics, 332 Science II Hall, Iowa State University, Ames, IA 50011, USA

Despite only ~1% difference in genomic DNA sequence, humans and chimpanzees differ considerably in mental and linguistic capabilities, and in susceptibility to some diseases. A recent comparison of gene expression in human and great apes cast some light on the genetic basis of these differences, but more rigorous study is required. Our statistical reanalysis of these microarray data shows that there have indeed been dramatic alterations in the expression of genes in the human brain since the split from chimpanzees, mainly caused by a set of genes with increased (rather than decreased) expression in the human brain.

An unsolved mystery in human genetics and evolution is how humans and chimpanzees differ so considerably in many morphological, behavioral and cognitive aspects, given only ~1% difference in their genomic DNA sequences [1–4]. Using microarray technology to measure gene expression in human and chimpanzee, Enard *et al.* [5] concluded that since the divergence of the human and chimpanzee lineages, gene expression in human brain has altered more dramatically than that in the chimpanzee. Not surprisingly, this hypothesis has attracted a lot of attention because the result not only supports the long-term notion that the genetic basis of human–chimpanzee differences is the alteration in gene expression rather than in coding sequence [1], but it also has important implications for the evolution of the human brain. However, as will be shown later, statistical problems in the data analysis make it difficult to rule out the possibility of other explanations. Therefore, this data needs to be reexamined carefully.

The major result of the Enard *et al.* study [5] is as follows. After defining an *ad hoc* measure for expression distance (see reference 7 in [5]), the authors examined the expression of 12 600 genes (on an Affymetrix chip) in human, chimpanzee and orangutan, for both brain and liver. By comparing human, chimpanzee and orangutan, an 'overall expression distance' for the each lineage was obtained (Fig. 1). Enard *et al.* (2002) estimated that for the brain sample, the ratio of expression distance of human versus chimpanzee lineage is ~3.8. For the liver sample, this ratio is lower at ~1.7.

A similar pattern was obtained with another microarray technique (cDNA array). Because the expression distance was interpreted as a measure for the change of gene expression, Enard *et al.* concluded that it is the brain rather than the liver that has undergone a dramatic change in gene expression in the human lineage. Although this conclusion is within expectations, the authors did not provide any statistical analysis to test whether this estimated ratio (~3.8) is significantly larger than one, or to test the null hypothesis of no difference of expression distances between human and chimpanzee lineages. The lack of statistical testing of this critical result could cast some doubts about the significance of the work, particularly when the authors acknowledged that the ratio estimation is sensitive to the selection of distance measures and the biological interpretation of these measures is not always clear.

We have reanalyzed the Affymetrix microarray data from Enard *et al.* [5], including samples from three humans, three chimpanzees and one orangutan, for both brain and liver. The original Affymetrix data was processed through background adjustment, logarithm transformation and normalization [6]. Given a tissue sample (brain or liver), we adopted a statistical protocol to select genes that have significant expression differences between human and chimpanzee under a given significance level (α). To this end, we first assigned a P-value for each gene. Among various statistical techniques proposed for microarray data analysis (reviewed in [7,8]), we chose the t-test because it is simple, conservative and takes account of intraspecies variation [9–11], which is substantial in this case. For a given significance level (α), a gene was said to be expressed differently between human and chimpanzee brains (or livers) if $P < \alpha$. We selected these genes for further analysis. The false positive rate for selected genes under a given α can be computed [12]. For the brain sample, we found 1988, 1087, 670 and 131 genes at $\alpha = 0.05$, 0.02, 0.01 and 0.001, respectively. The corresponding false positive rates are estimated to be 21%, 15%, 12% and 5%, respectively.

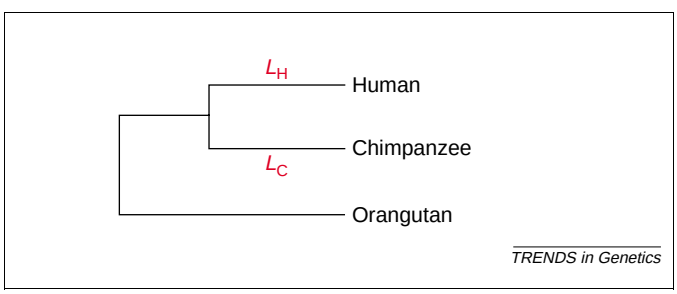

Fig. 1. A schematic tree for the human–chimpanzee–orangutan relationship, showing the human-lineage-specific expression change (L_H) and the chimpanzee-lineage-specific expression change (L_C), see Tables 1,2.

Corresponding author: Xun Gu (xgu@iastate.edu).

Table 1. Examples of genes expressed differently in human and chimpanzee brains[a]

	Diversified	Human specific (L_H)		Chimpanzee-specific (L_C)		Unclassified
		I	*R*	*I*	*R*	
Human	5.21	2.50	0.66	1.02	0.93	0.94
Chimpanzee	1.88	1.06	0.95	2.03	0.42	0.85
Orangutan	1	1	1	1	1	1

[a]Examples of genes that are expressed differentially in human and chimpanzee brains (or livers). The phylogenetic location of expression change is inferred by using orangutan as a reference (Fig. 1). Four types of expression change are shown, where L_H and L_C represent the human-lineage-specific and chimpanzee-lineage-specific expression change, respectively. The numbers represent the fold change of gene expression level in humans or chimpanzees, compared with that in orangutans. Genes in L_H or L_C group can be classified into two categories, induced (*I*) or repressed (*R*).

For each gene selected, we then used the orangutan as a reference to infer first, the phylogenetic location of the change of expression (i.e. did it occur in the human lineage or the chimpanzee lineage), and second, the trend of expression change (i.e. induced or repressed). At each significance level, we classified selected genes into four groups (Fig. 1; Tables 1,2):

(1) Diversified expression pattern: where the gene expression level in the orangutan (μ_O) is significantly different ($P < 0.05$, *t*-test) from the gene expression levels in both the chimpanzee and human (μ_C and μ_H, respectively).
(2) Chimpanzee-lineage (L_C) specific events: where μ_O is significantly different from μ_C but not from μ_H, suggesting the expression change occurred in the chimpanzee lineage after the human–chimpanzee split.
(3) Human-lineage (L_H) specific events: where μ_O is significantly different from μ_H but not μ_C, suggesting the expression change occurred in the human lineage after the human–chimpanzee split.
(4) Unclassified: where μ_O is not significantly different from both μ_C and μ_H.

In cases (1) and (4), one cannot use orangutan to infer the phylogenetic location of expression changes statistically.

Figure 2 shows clearly that for the brain, changes in gene expression in the human lineage are statistically more frequent than in the chimpanzee lineage, as measured by the ratio of changes in the human to chimpanzee lineage, L_H/L_C. However, this is not the case for the liver. This result holds true regardless of which significance level (α) is used for the selection of differential expressed gene between humans and chimpanzees (Fig. 2; Table 2). Indeed, the L_H/L_C ratio for brain-expressed genes ranges from 2.76 to 3.22 for $\alpha = 0.05$ to 0.001; in each case the null hypothesis of $L_H/L_C = 1$ is rejected at $P < 0.001$. By contrast, the L_H/L_C ratio for liver-expressed genes is virtually equal to one in any case. Hence, our result provides statistical support for the notion of dramatic gene expression changes in the brain of the human lineage [5].

For gene expression changes that are chimpanzee-lineage specific or human-lineage specific [i.e. case (2) or (3), above], we can infer the change in direction of the evolutionary event; that is, from low to high expression level (induction, denoted by *I*), or from high to low expression level (repression, *R*) (Table 1). In the human lineage, the induction/repression (*I*/*R*) ratio in brain ranges from 2.21 to 5.9; in each case it is greater than one with $P < 0.001$. By contrast, no evidence shows that the *I*/*R* ratio for liver-expressed genes in humans is significantly greater than one ($P > 0.05$) (Fig. 3a). Interestingly, in the chimpanzee lineage, the *I*/*R* ratio (~ 1.5) for both brain and liver expressed genes is not clear, and is sensitive to the significance level (Fig. 3b).

In summary, our analysis has provided statistical evidence to show that after the split of humans and chimpanzees, the change of expression pattern in the

Table 2. Classification of genes expressed differently between humans and chimpanzees[a]

Tissue	α	Number of genes			
		Diversified	L_H	L_C	Unclassified
Brain	0.05	238	572	183	995
	0.02	143	403	125	416
	0.01	89	268	93	220
	0.001	24	69	25	13
Liver	0.05	743	895	894	1079
	0.02	535	704	694	471
	0.01	411	559	550	205
	0.001	174	135	131	0

[a]Number of genes selected for each group under various significance levels (α) for genes to be expressed differentially between humans and chimpanzees (brain or liver). L_H and L_C represent the human-lineage-specific and chimpanzee-lineage-specific expression change, respectively. Instead of *t*-test, we also used several other approaches; for example, Bayesian frameworks approach [11], bootstrapping approach. Predictions from various methods mostly agree with each other and our main result holds (not shown).

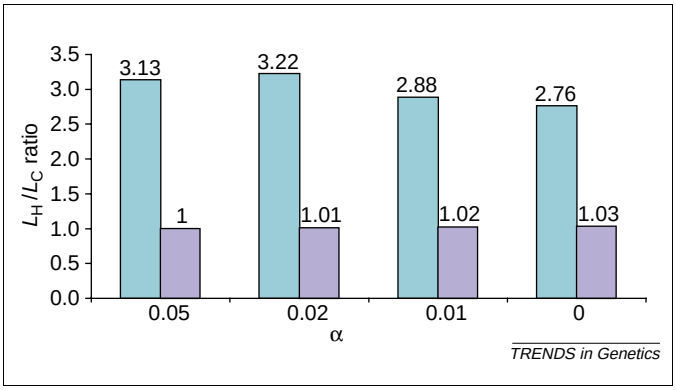

Fig. 2. Ratio of human-lineage specific expression changes to chimpanzee-lineage specific expression changes (L_H/L_C) in both brain (green) and liver (purple) under different significance levels ($\alpha = 0.05$, 0.02, 0.01 and 0.001). In all cases, the L_H/L_C ratio of brain expressed genes is significantly greater than one, whereas L_H/L_C ratio of liver expressed genes is not. Because oligonucleotides on the arrays are from humans, one might expect to see more genes with higher expression levels in humans than in apes. If this is the case, our inference of more genes with altered expressions in the human brain could be biased. However, because such pattern was not observed in the liver, it seems unlikely; otherwise one would have to assume that brain-expressed genes evolve much faster than live-expressed genes, which seems unreasonable.

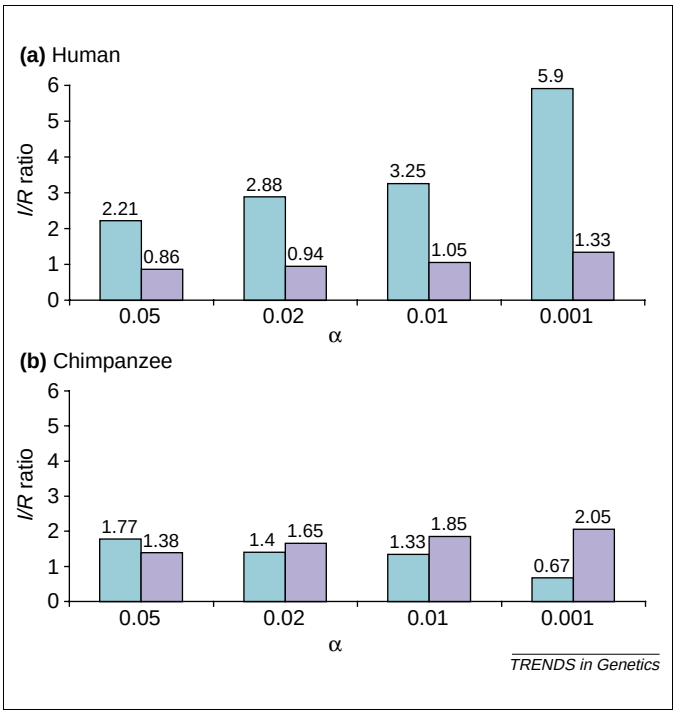

Fig. 3. Induction/repression (*I/R*) ratios for genes showing lineage-specific expression patterns. (a) In the human brain (green) and liver (purple). The *I/R* ratio of brain-expressed genes is statistically greater than one, whereas the *I/R* ratio of liver-expressed genes is not significant. (b) In the chimpanzee brain (green) and liver (purple). The *I/R* ratios for both brain- and liver-expressed genes are relatively similar, and sensitive to the significance level.

human brain was more dramatic than that in the chimpanzee [5]. Moreover, we have shown that these changes in the human brain involved induction (increased gene expression) much more frequently than repression. This pattern is not observed in chimpanzee brain, nor in the liver of humans or chimpanzees. The enhanced expression of genes in the human brain since the split from chimpanzees could be important in the emergence of human beings, and certainly deserves further investigation.

Acknowledgements

We are grateful to S. Paabo and P. Khaitovich for providing the Affymetrix dataset and comments, and to D. Nettleton, W. Nordstrom and T. Dunn for constructive discussions. This work was supported by an NIH grant.

References

1 King, M.C. and Wilson, A.C. (1975) Evolution at two levels in humans and chimpanzees. *Science* 188, 107–116
2 Goodman, M. *et al.* (1998) Toward a phylogenetic classification of primates based on DNA evidence complemented by fossil evidence. *Mol. Phylogenet. Evol.* 9, 585–598
3 Nickerson, E. and Nelson, D.L. (1998) Molecular definition of pericentric inversion breakpoints occurring during the evolution of humans and chimpanzees. *Genomics* 50, 368–372
4 Fujiyama, A. *et al.* (2002) Construction and analysis of a human–chimpanzee comparative clone map. *Science* 295, 131–134
5 Enard, W. *et al.* (2002) Intra- and interspecific variation in primate gene expression patterns. *Science* 296, 340–343
6 Irizarry, R.A. *et al.* (2002) Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* http://biosun01.biostat.jhsph.edu/~ririzarr/papers/affy1.pdf
7 Quackenbush, J. (2001) Computational analysis of microarray data. *Nat. Rev. Genet.* 2, 418–427
8 Nadon, R. and Shoemaker, J. (2002) Statistical issues with microarrays: processing and analysis. *Trends Genet.* 18, 265–271
9 Baldi, P. and Long, A.D. (2001) A Bayesian framework for the analysis of microarray expression data: regularized *t*-test and statistical inferences of gene changes. *Bioinformatics* 17, 509–519
10 Dudoit, S. *et al.* (2002) Statistical methods for identifying differentially expressed genes in replicated cDNA microarray experiments. *Statistica Sinica* 12, 111–140
11 Long, A.D. *et al.* (2001) Improved statistical inference from DNA microarray data using analysis of variance and a Bayesian statistical framework. *J. Biol. Chem.* 276, 19937–19944
12 Allison, D.B. *et al.* (2002) A mixture model approach for the analysis of microarray gene expression data. *ComPut. Stat. Data Anal.* 39, 1–20

PII: S0168-9525(02)00040-9

Genome Analysis

Gene conversion drives GC content evolution in mammalian histones

Nicolas Galtier

CNRS UMR 5000 – 'Génome, Populations, Interactions', Université Montpellier 2 – CC63, Place E. Bataillon, 34095 Montpellier, France

To examine the evolutionary influence of gene conversion on DNA base composition, I analysed an exhaustive dataset of histone paralogous genes from human and mouse. I show that those gene copies that belong to subfamilies of very similar sequences (presumably undergoing gene conversion) have a higher GC content than unique gene copies (presumably not undergoing gene conversion). Thus, it seems that gene conversion is a biased process that tends to increase the DNA GC content, a conclusion that has implications for the evolution of isochores in vertebrates.

Gene conversion is a mechanism by which two similar DNA fragments become identical: one genomic copy is 'copied and pasted' onto the other one. In diploid organisms, allelic gene conversion can occur at meiosis between

Corresponding author: Nicolas Galtier (galtier@univ-montp2.fr).