Generating a global mRNA segregation pattern for a comparative analysis of the Human and Chimpanzee transcriptomes

William Brown

Systems Biology

ABSTRACT

As a fundamental question we endeavor to understand our origins and our developmental history which has led to our emergence. This inquiry is so profound because of the indisputable importance and indelible desire to know thyself. In looking to our origins we have identified a species that, from an evolutionary perspective, we have shared a common ancestor with, the chimpanzee. Because chimpanzees are the most closely related species to humans a comparative analysis of their genetic and transcriptomic constitutions can potentially illuminate the molecular evolution that engendered the divergence of the two species. This will help to provide an understanding for how genetic perturbations can produce the phenotypic differences between humans and chimpanzees, and indeed this is currently one of the most intensely studied subjects in genomics. The Human and Chimpanzee genomes have been compared and the differences and similarities between the two are highly defined. However the transcriptomes have not been thoroughly compared and it is from this analysis that insight into the apparent phenotypic differences is hoped to be gained because the divergence is hypothesized to be due primarily to differential regulation of the transcriptome. So while there are around 7.5×10^7 nucleotide differences within the genomes it is not thought that this is a sufficient divergence of sequence to produce speciation as it only accounts for a difference of 1.22 – 1.23% of the genetic code. Because the 3' untranslated region (3'UTR) of mRNA transcripts is the target of post-

transcriptional regulatory machinery structural differences of the 3'UTR of human and chimpanzee transcripts in the same tissue will show that orthologous genes are under disparate regulation in the two species. Therefore a transcriptome-wide profiling analysis will be performed by analyzing the different sizes of the mRNA transcripts contained within the neuron cells of a human and a chimpanzee for comparative analysis. It is expected that conserved gene orthologs will be producing mRNA transcripts with different 3'UTR structures. This will show that while the genomes of chimpanzees and humans are highly conserved there are differences in the regulation of the expressed mRNAs which would support the hypothesis that this is the driving factor of phenotypic divergence between the two species.

INTRODUCTION

This is a proposal for the examination of the post-transcriptional regulatory system of the chimpanzee, which will be analyzed and compared to that of a human. The following outline details a method to produce a global representation of the varying lengths of the 3'UTRs from a human and chimpanzee transcriptome for comparative analysis. Such comparative analyses have been performed in great detail for the genome and much valuable insight has been gained (Chen *et al*, 2001, The Chimpanzee Sequencing and Analysis Consortium, 2005). However it has long been suggested that a more encompassing understanding of the divergence of our lineage will only come from an examination of the differential expression of our genomes, given the high level of conservation between the two (King and Wilson, 1975). This points the way to the regulatory regions of the genome and the transcriptomic regulatory systems.

In an investigation of the divergence of the human and chimpanzee transcriptome by mapping full length cDNAs (FLcDNAs) to the human genome it was found that 3'UTRs had the largest amount of structural differences, occurring primarily from indels (Sakate *et al*, 2007). The indels were caused by transposons (*Alu* elements), which provides a possible explanation for divergence because 3'UTR sequence motifs involved in processing and regulation could be inserted or deleted by the retroposons. Indeed in the study performed by Sakate *et al* it was found that an *Alu* insertion added an exon onto the 3' end of the chimpanzee genome. Therefore the addition of microRNA (miRNA) recognition sites and alternative cleavage and polyadenylation signals (APA) perturbed the regulation of the transcriptome.

The comparative analysis drawn between humans and chimpanzees can indicate the types of selective forces that caused divergence between the species. For example the 5'UTRs appear

to be under positive selection in humans and chimpanzees and selective constraints appear to act on the 3'UTR because it exhibits 29% less divergence than intergenic regions, but at the same time the 3'UTR has diverged almost twice as much as coding sequence (CDS) regions (Hellman *et al*, 2003). This seems to indicate the importance of the untranslated regions of mRNA, because they are under selective pressure unlike the intergenic regions, and therefore they are important to the survivability of the organism, and yet they have experienced more divergence than the coding sequences, they are therefore the most likely candidates for where the alterations to the genome have occurred that drove speciation. This is the reason why the 3'UTR is the area of scrutiny proposed for this investigation. Selection seems to have retained changes in the genome that affected the expression of the genes, and not changes in the genes themselves. Even in genes where novel exons have been identified, functional analysis indicates that they are primarily involved in the regulatory pathways and perturb gene expression as such (Huang *et al*, 2008).

Experimental Design and Methodology

A single tissue type will be selected to be used in obtaining transcriptome-wide mRNAs from a chimpanzee. The tissue type of choice is the cerebellar cortex because the tissue must be presumed to be under differential regulation. One can imagine that epidermal cells will have minimal differences in expression or regulation when compared between humans and chimpanzees. The brain however is the organ with the most obvious functional differences between the two species. Indeed Enard *et al* found that the human brain had the most pronounced changes in protein and gene expression after analyzing interspecific variation in primate gene expression patterns (Enard *et al*, 2002). This functional difference is most likely attributable to differences in structure and organization of the brain and not to the production of

novel proteins (proteins arising from mutation that is, not proteins arising from alternative processing events such as alternative exon splicing). The organization of the brain is very much under the control of post-transcriptional regulatory machinery as this directs the patterning of the nervous system via manipulation of the neuronal synapses (Landthaler *et al*, 2008). Only a single tissue type is to be selected and utilized because it has been shown that the variation in mRNA isoforms is approximately threefold greater between tissues than it is between individuals (Wang *et al*, 2008). Therefore including numerous tissue types would greatly obfuscate any intelligible interpretations for the effect of differential regulation on the divergence of species and in fact this would be a study of merit on its own. As many tissue samples as can be obtained will be analyzed. It is hoped that 5-10 samples will be obtained.

It would be expected that the difference in mRNA regulation will be greater than that observed between two individuals of the same species but less than that observed between two different tissues of the same individual. This is because while there should be enough differential regulation to produce unique functionality of the same tissues between species, such that in neurons the cumulative effects of differences in neuronal synaptic structure produces different patterning of the chimpanzee nervous system as compared to a humans, it is not so great that the brain cell of a chimpanzee can be considered as divergent from a humans to the same degree as a brain cell would be compared to the epithelial cell of the same organism.

Intriguingly it was found in the study performed by Enard *et al* that the mRNA expression pattern is more different between two individual humans than it was between a human and a chimpanzee (Enard *et al*, 2002). This could be because the mRNA transcripts were hybridizing to human oligonucleotides on the microarray chips being used and therefore this lowered the resolution for the other primate species and introduced bias. However they also

found that there was a 5.5-fold acceleration in the rate of change of gene expression in the lineage leading to humans for brain cortex, making it more divergent from chimpanzees than chimpanzee brain cortex gene expression are from macaques. Nevertheless the large difference in gene expression levels observed in humans could be attributable to a greater variation in the level of transcriptional regulation and will potentially provide enhanced resolution to our study of differential regulation.

Total RNA extracts of the selected tissues will be obtained using RNeasy kits and the full length of the mRNA transcripts will be amplified using 3' Rapid Amplification of cDNA Ends and PCR (3'RACE-PCR). The PCR cycles will be performed with a fluorescently labeled primer so that the amplicons are fluorescently active. After the 3'RACE-PCR the products will be ran through an ABI 3700 capillary electrophoresis machine allowing them to be separated by size. This will produce a transcriptome segregation pattern from the varying sizes of the mRNA transcripts which can be compared between human and chimpanzee (Figure 1). The length of the 3'UTR can be greater than 3kb making it a major contributor to the size of the mRNA transcript. But more importantly differences in size between the human and chimpanzee mRNA transcripts will rarely be attributable to changes in the length of coding sequences but instead to variations in the length of the untranslated regions. For example between humans and chimpanzees the number of differences occurring in the CDSs is 0.45% while in the 3'UTR it is 0.86% (Hellman *et al*, 2003). This makes the transcriptome segregation pattern method a portrayal of the varying lengths of the 3'UTRs and therefore an indication of the difference in regulation.

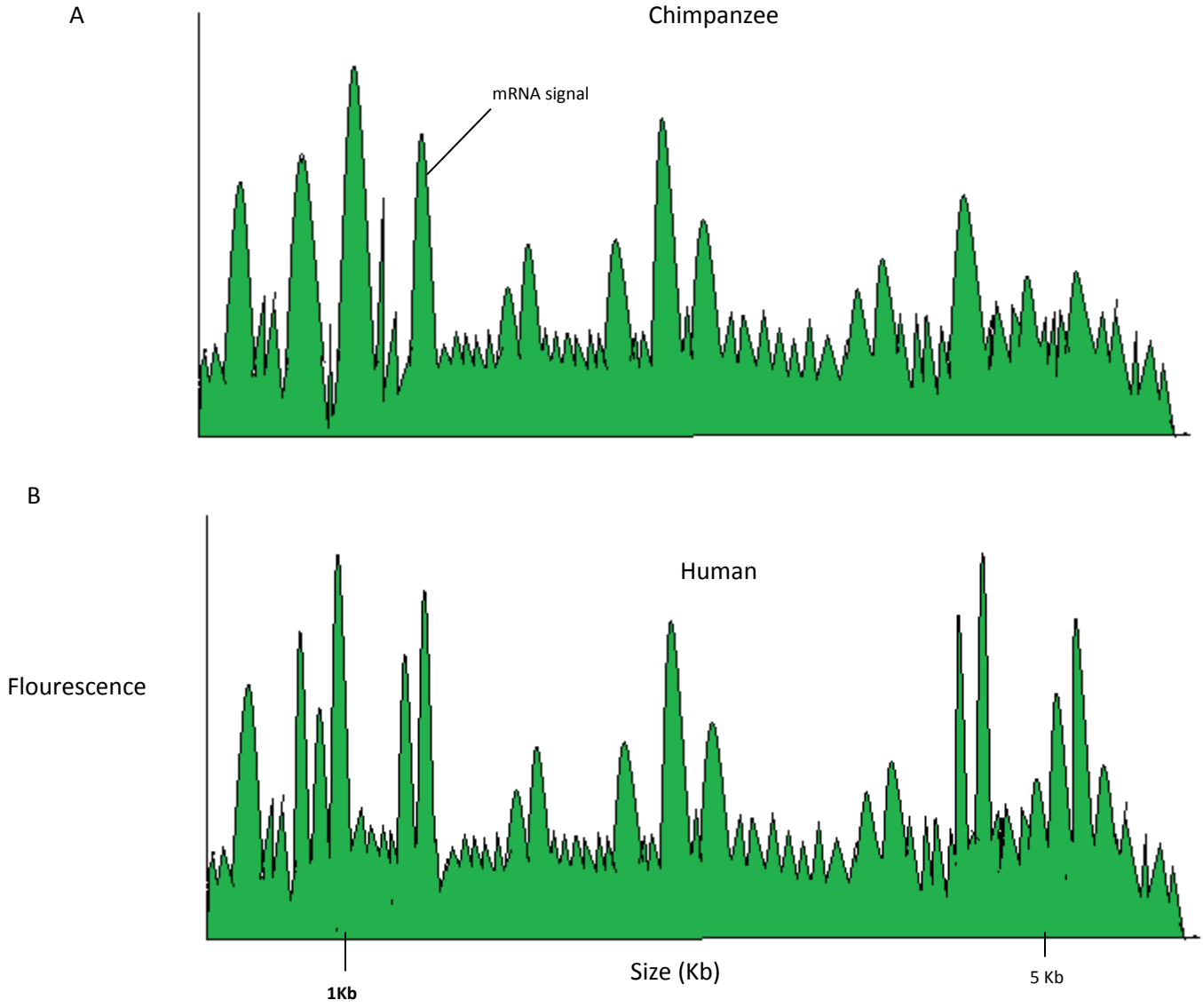

Figure 1. Diagram showing the expected transcriptome "fingerprint" generated from the variable sizes of the mRNA 3'UTRs. The peaks are 3'RACE-PCR amplicons from the mRNA transcripts. The amplicons have been segregated by size. Flourescent instensity from the labeled primer is on the y-axis and size in kilobases is on the x-axis. (A) The chimpanzee transcriptome "fingerprint" and (B) the Human transcriptome "fingerprint".

In order to quantify the transcriptome "fingerprints" a size ratio (ω) will be generated by selecting a midpoint on the x-axis and adding all of the signals to the right (N_R) and dividing them over the sum of the signals to the left (N_L) of the midpoint: $\omega = (N_R / N_L)$

The size ratio will be a quantification of the amount of regulation as $\omega > 1$ will occur in systems under higher levels of regulation and $\omega < 1$ will occur in systems under lower levels of regulation. The values will also be used to compare the human and the chimpanzee transcriptome "fingerprints" quantitatively.

Although the transcriptome segregation pattern is a global representation of 3'UTR lengths, a number of mRNAs from the neuron cells will be selected and the exact sequence of their 3'UTRs will be ascertained to directly show the differences in lengths. Since the mRNA molecules have already been amplified by 3'RACE-PCR the amplicons will be isolated via a Northern blot and conventional sequencing can be performed on a number of transcripts in order to compare them to that of a human. In probing the 3'UTR structures of the mRNA transcripts from a chimpanzee and comparing them to those of a human it is predicted that there will be differences in the length of the 3'UTR (Figure 2). The length of the 3'UTR affects the stability, localization, and translation of the mRNA transcript, with shorter 3'UTRs making the mRNA transcripts more robust and consequently having higher rates of translation (Mayr & Bartel, 2009). This is because the 3'UTR contains the complementary seed sequences for miRNA repressor molecules and RNA binding proteins. When the 3'UTR is alternatively cleaved the number of miRNA recognition elements (miREs) changes. Shorter 3'UTRs have fewer miREs and are therefore under less targeted repression. Therefore differences in the length of the 3'UTR between chimpanzees and humans will be indicative of differing levels of regulation being imposed on the transcripts.

In order to control for differences in environmental condition the history of the chimpanzees and humans from which samples are derived must be thoroughly evaluated. Environmental conditions such as malnutrition, isolation, and captivity have a pronounced

influence on genetic expression and hence will perturb the transcriptome as compared to that of a healthy environment. Nowhere will this be more evident than in the brain as neurotransmitters and receptors are suppressed because of depression or malnutrition. It must be concluded that the human and chimpanzee subjects that are being compared were both living under healthy conditions that would not cause aberrant brain functioning.

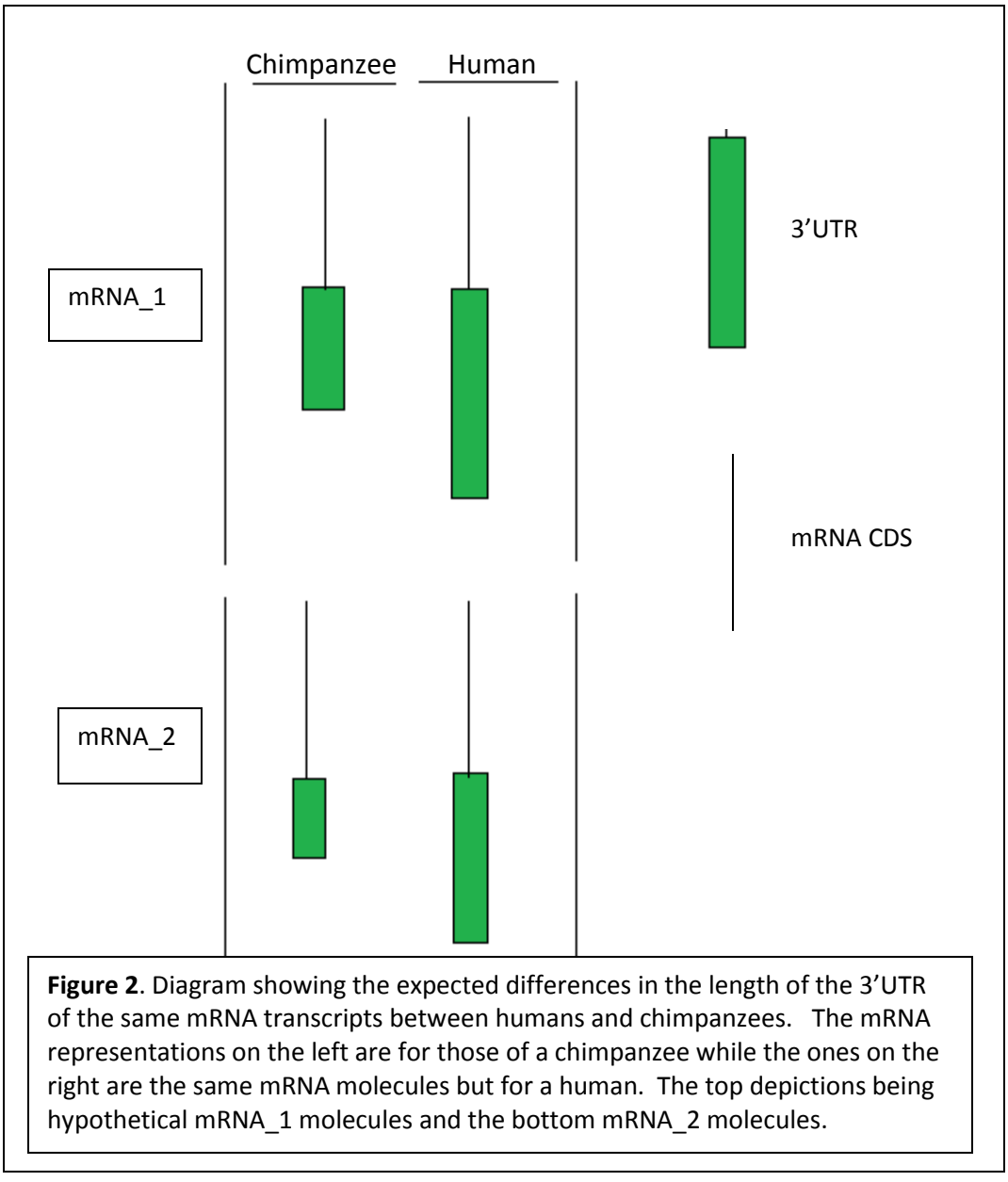

Figure 2. Diagram showing the expected differences in the length of the 3'UTR of the same mRNA transcripts between humans and chimpanzees. The mRNA representations on the left are for those of a chimpanzee while the ones on the right are the same mRNA molecules but for a human. The top depictions being hypothetical mRNA_1 molecules and the bottom mRNA_2 molecules.

Projected Duration and Timeline:

- Two to three months to obtain suitable human and chimpanzee cadavers for cerebral cortex biopsies.
- Three months of optimization to obtain clear and interpretable signals from the capillary electrophoresis.
- A year to achieve the desired number of tissue samples (5-10). Because it is not expected that deceased chimpanzees will be readily available for biopsy.
- One month of analysis for interpretations of the data and a full conclusion of the study.

Future Directions

If the data obtained from this experiment supports the propounded hypothesis then it will be the first transcriptome-wide comparison of a component of the chimpanzee and human post-transcriptional regulatory systems and will show that they are under differential regulation in the two species. This will advance the paradigm from suppositions about the causes of divergence of chimpanzees and humans and show that differences in gene expression, and not gene products, was the main factor affected by the impetus driving speciation. The 3730 ABI capillary electrophoresis machines are designed to handle multiple injections in a single run, with a capacity to analyze an entire 96 well plate. Therefore if the method proves to be informative then it can be scaled up to analyze over a hundred samples in a couple of injections. If epidermal or blood cells were chosen for analysis of the transcriptome segregation pattern then entire human and chimpanzee populations could be studied and compared. The knowledge gained from an investigation such as this would most likely prove very informative.

LITERATURE CITED

F.-C. Chen, E. J. Vallender, H. Wang, C.-S. Tzeng, and W.-H. Li. Genomic Divergence Between Human and Chimpanzee Estimated from Large-Scale of Genomic Sequences 2001 The American Genetic Association 92:481–489.

The Chimpanzee Sequencing and Analysis Consortium. Initial sequence of the chimpanzee genome and comparison with the human genome. 2005. Nature. Vol 437.

Wolfgang Enard, *et al*. Intra- and Interspecific Variation in Primate Gene Expression Patterns. 2002. Science 296, 340.

Ines Hellmann, Sebastian Zöllner, Wolfgang Enard, et al. Selection on Human Genes as Revealed by Comparisons to Chimpanzee cDNA. 2003. Genome Research 13: 831-837.

Yao-Ting Huang, Feng-Chi Chen, Chiuan-Jung Chen, et al. *Genome Res.* Identification and analysis of ancestral hominoid transcriptome inferred from cross-species transcript and processed pseudogene comparisons. 2008 18: 1163-1170.

Mary-Claire King and A. C. Wilson. Evolution at Two Levels in Humans and Chimpanzees. 11 April 1975, Science, Volume 188, Number 4184.

Markus Landthaler, Dimos Gaidatzis, Andrea Rothballer, et al. Molecular characterization of human Argonaute-containing ribonucleoprotein complexes and their bound target mRNAs. 2008. RNA 14: 2580-2596.

Christine Mayr and David P.Bartel. Widespread Shortening of 3'UTRs by alternative cleavage and polyadenylation activates oncogenes in cancer cells. 2009. Cell.

Ryuichi Sakate, Yumiko Suto, Tadashi Imanishi, Tetsuya Tanoue, Munetomo Hida, Ikuo Hayasaka, Jun Kusuda, Takashi Gojobori, Katsuyuki Hashimoto, Momoki Hirai. Mapping of Chimpanzee full-length cDNAs onto the human genome unveils large potential of the transcriptome. 2007. Gene 399, 1–10

Y. Sakaki*, H. Watanabe, A. Fujiyama, M. Hattori, A. Toyoda, T.D. Taylor. Chimpanzee genome project for understanding ourselves. 2002. International Congress Series 1246 183– 187.

Eric T. Wang, Rickard Sandberg, Shujun Luo4, Irina Khrebtukova, Lu Zhang, Christine Mayr, Stephen F. Kingsmore6, Gary P. Schroth4 & Christopher B. Burge. Alternative isoform regulation in human tissue transcriptomes. 27 November 2008, Nature. Vol 456.