Comparative analysis of the complete transcriptome of the ENCODE regions in humans and chimpanzees

Philipp Khaitovich, Henriette Franz, Johann Visagie, Janet Kelso and Svante Pääbo

Max-Planck-Institute for Evolutionary Anthropology, Deutscher Platz 6, D-04103 Leipzig, Germany

RNA transcription is the first step in the transfer of functional information from genome sequences into phenotypic features. However, our knowledge of RNA transcription is mostly restricted to annotated gene regions. Studies employing recently introduced tiling array technology have shown that a significant proportion of RNA expression originates in intergenic regions. However, the functional significance of this expression remains obscure.

To address whether intergenic transcripts play a role in the evolution of human-specific phenotypic features, we used tiling arrays to measure RNA expression levels in the complete ENCODE region in brain, heart, liver and lymphoblast cell lines, from five humans and five chimpanzees. Our results show that 85% of RNA transcription occurs outside the exons of known genes in both species. Further, we find that approximately 40% of all RNA transcription occurs in intergenic regions.

Surprisingly, in all tissues, intergenic transcripts show comparable selective constraints on their expression divergence between species as observed for annotated exons. This indicates that intergenic transcripts are likely to have functional significance comparable to that of known genes. Co-regulation analysis indicates that a substantial proportion of intergenic transcripts is likely to represent extensions of known genes. Interestingly, this proportion is significantly larger in testis than in the other three tissues. The functional significance of remaining intergenic transcripts remains unknown. Nonetheless, our findings indicate that changes in intergenic RNA expression may have played a substantial role in the evolution of human-specific phenotypic features.