Investigating the Role of Pseudogenes as the Source of Conserved Non-coding Elements in the Human Genome

Liu, Joyce

More than 98% of the human genome is defined as non-coding DNA. Many of these noncoding elements are simply remnants of ancient genetic material, but studies indicate that 5% of the genome is composed of non-coding sequences that have been highly conserved. The strong evolutionary pressure on these conserved non-coding elements (CNEs) imply that these regions likely perform some significant function in the genome. Recent studies also suggest that pseudogenes, once-functioning protein-coding genes that have lost their gene expression or their ability to code proteins, may account half of the CNEs in the human genome. In this paper, we investigate the relationship between pseudogenes and CNEs in order to identify the function, purpose, and properties of conserved non-coding elements in the human genome. Using two sets of human CNEs of varying degrees of robustness, we analyzed the overlap between pseudogenes and CNEs. After determining the relationship between pseudogenes and CNEs we analyzed the proximity of each CNE to known genes and DNase hypersensitivity peaks. Results showed that CNEs of a more nuanced statistical model of significant conservation, demonstrated a minuscule percentage of overlap with pseudogenes. This suggests that real CNEs, showing conservation over a broad evolutionary distribution, seem not to be the result of pseudogene activity. Furthermore, CNEs demonstrated a strong correlation to DNase hypersensitivity peaks, suggesting that CNEs are highly accessible. Some CNEs may also be closely associated with known protein coding regions, as many CNEs displayed close proximity to known genes.

Awards Won: