High Dimensional, Multi-Genomic Investigation into the Function and Evolution of Eukaryotic Single Exon Genes.

Jorquera R.\textsuperscript{1}, Ortiz R.\textsuperscript{2}, Pedroso A.\textsuperscript{3} and Holmes D.S.\textsuperscript{1}

\textsuperscript{1}Center for Bioinformatics and Genome Biology, Fundacion Ciencia \& Vida and Facultad de Ciencias Biologicas, Universidad Andres Bello, Santiago Chile; \textsuperscript{2}Facultad de Ciencias Biologicas, Pontificia Universidad Catolica de Chile, Santiago, Chile; \textsuperscript{3}Facultad de Ingenieria, Universidad de Talca, Chile.

Eukaryotic genes are generally interrupted by introns. However, some of them lack introns and are termed single exon genes or SEGs. Although some SEGs are thought to be non-functional pseudogenes, many have been demonstrated to be expressed, raising questions as to their origin, evolution and function.

A database of predicted SEGs was constructed from an analysis of 30 eukaryotic species ranging from yeast to human and including a plant. Gene information for each species was derived from the NCBI ftp website and parsed into multiexonic and single exonic genes using an in-house Pearl script. The database houses over 148,000 predicted SEGs. High dimensional analysis of the SEG information includes prediction of their function (KOG, PFAM and manually curated methods), their distribution between species, chromosome location, similarity to multi-exon orthologs and paralogs, relation to GC content, patterns of gene expression, association with diseases and others.

The frequency of occurrence of SEGs ranges from 99% of all genes in some protists and fungi to 2.6% in the worm \textit{C. elegans}. In higher metazoans such as mammals, SEGs are predicted to be enriched in functions encoding histones, translation functions and G protein-coupled cell surface receptors (GPCRs), whereas in lower metazoans (e.g. insects) and single celled Eukaryotes (e.g. fungi) SEGs appear to be enriched in metabolic functions. This contrast of functional categories suggests the operation of different selective strategies operating at different stages of evolution.

Cases of species specific SEGs were discovered, for example SEGs that were only present in the human lineage or SEGs found only in primates, providing a window of opportunity to evaluate the timing of SEG origin and potential insight into the mechanism(s) of SEG production. This generates useful information for the “introns early” versus “introns late” debate.