A NEW UNDERSTANDING OF THE HUMAN GENOME

JOHN MATTICK
j.mattick@imb.uq.edu.au

Institute for Molecular Bioscience
University of Queensland, AUSTRALIA

Abstract

It appears that the genetic programming of mammals and other complex organisms has been misunderstood for the past 50 years, because of the assumption – largely true in prokaryotes, but not in complex eukaryotes – that most genetic information is transacted by proteins. The numbers of protein-coding genes do not change appreciably across the metazoa, whereas the relative proportion of non-protein-coding sequences increases markedly. Moreover, while only a tiny fraction encodes proteins, it is now evident that the majority of the mammalian genome is transcribed in a developmentally regulated manner, and that most complex genetic phenomena in eukaryotes are RNA-directed. Evidence will be presented that (i) regulatory information scales quadratically with functional complexity and hence the majority of the genomes of the higher organisms comprises regulatory information; (ii) there are thousands of non-protein-coding transcripts in mammals that are dynamically expressed during differentiation and development, including in embryonal stem cell and neuronal cell differentiation, and T-cell and macrophage activation, among others, many of which show precise expression patterns and subcellular localization in the brain; (iii) many 3'UTRs are not only linked to but are also expressed in a regulated manner separately from their associated protein-coding sequences to transmit genetic information in trans; (iv) there are large numbers of small RNAs, including new classes, expressed from the human and mouse genomes, that may be discerned from bioinformatic analysis of genomic and deep sequencing transcriptomic datasets; and (v) much, if not most, of the mammalian genome may not be evolving neutrally, but rather is composed of different types of sequences (including transposon-derived sequences) that are evolving at different rates under different selection pressures and different structure-function constraints. There is also genome-wide evidence of editing of noncoding RNA sequences, especially in the brain and especially in humans (Alu elements), which may constitute a key part of the molecular basis of memory and cognition. Taken together, these and other observations suggest that the majority of the human genome is devoted to an
very sophisticated RNA regulatory system that directs developmental trajectories and mediates gene-environment interactions via the control of chromatin architecture and epigenetic memory, transcription, splicing, RNA modification and editing, mRNA translation and RNA stability.