Transcriptome meta-analysis reveals the human organs evolution

Katia de Paiva Lopes, Ricardo Assunção Vialle, José Miguel Ortega

Departamento de Bioquímica e Imunologia, Instituto de Ciências Biológicas (ICB), Universidade Federal de Minas Gerais (UFMG), Brasil.

RNA-Seq allows the measurement of transcripts expression levels in a manner far more precise and global than previous methods. Studies using this technology have already altered our view of the extend and complexity of eukaryotic transcriptomes. Actually, multiple efforts have been made to determine and analyze the gene expression patterns of many human cell types in different conditions. However, until recently, little has been reported about the perspective of gene evolution and gene expression. So, in these work a transcriptome meta-analysis was performed using 4 different databases. We examined the protein coding genes with at least 1 FPKM by tissue from Fantom5, GTEx, HPA (Human Protein Atlas) and IBM (Illumina Body Map). This corresponds to the following number of genes, respectively: Brain (10164, 12137, 12501, 12049), Colon (9412, 12205, 12175, 11063), Heart (8424, 10333, 10794, 10070), Kidney (8584, 11961, 12084, 12002), Lung (9023, 12474, 12610, 11873), Ovary (8587, 11687, 11597, 12272), Prostate (8779, 12577, 12544, 12233) and Testis (9978, 14307, 14033, 13832). In a global analysis of each database, we classified the genes in some different groups, such as: Tissue Specific (TS - Expressed in only one tissue), Tissue Enriched (TE - mRNAs levels in a particular tissue at least five times those in all other tissues), Elevated genes (EG - Total number of tissues from another 3 groups as tissue enriched, group enriched and tissue-enhanced) and Ubiquitous (UB - Expressed in all tissues). The number of genes in these categories are: Fantom5 (1448, 1847, 10452, 4083), GTEx (1177, 2163, 11227, 6188), HPA (1504, 2762, 11208, 5922) and IBM (1966, 2932, 9640, 6617). Accessing the group of orthologues and determining the Last Common Ancestor of each gene, we determined the fraction shared among the clades of the human lineage, from Cellular organisms to Homo sapiens. Our results showed that the origin of distinct organs was genetically told by examining the time of appearance of some specific genes or within distinct groups as elevated genes. Analyzing the cumulative curve of genes originated in each clade showed that the area underneath the curve of TS genes is much lower than the area of UB genes for all databases analyzed here: Fantom5 (18.85, 24.91), GTEx (16.74, 25.06), HPA (16.95, 25.09) and IBM (17.67, 24.94), indicating that the ubiquitous genes are more ancient than specific genes. Moreover, when we used the prominent genes (most expressed genes) we verified a large intersection of it with ubiquitous or housekeeping genes, showing the same list of genes for different tissues.

Supported by: Capes, FAPEMIG, CNPq.