The Mammalian Ortholog and Annotation Database: enhanced ortholog information with an accurate identifier mapping between NCBI and Ensembl accessible from a Galaxy-server

Jochen Bick¹, Susanne E. Ulbrich¹ and Stefan Bauersachs¹,²

¹ETH Zurich, Animal Physiology, Institute of Agricultural Sciences, Zurich, Switzerland
²University of Zurich, Genetics and Functional Genomics, Clinic of Reproductive Medicine, Zurich, Switzerland

Introduction

With the availability of more mammalian and other genomes, gene expression studies, particularly RNA sequencing became feasible for many non-model organisms. Although, the quality of the results for mapping and counting of RNA-seq reads can be comparable to the classical model organisms, incomplete gene annotation and assignment to functional annotation databases can lead to a substantial loss of information in downstream data analysis. To overcome this, we developed a database tool, called Mammalian Ortholog and Annotation Database (MOAdb, www.moadb.ethz.ch), to conveniently provide ortholog/homolog information for selected mammalian species. The information is obtained in three steps, namely the match of NCBI Entrez gene symbols, the use of the ortholog information contained in EnsemblCompara, and the results of pairwise BLAST comparisons of all transcripts. Ortholog gene lists derived from the MOAdb can be used for functional annotation with tools such as DAVID.

Identifier pairs per source

A genomic position based identifier mapper for NCBI and Ensembl Gene IDs.

Number of ortholog genes within the MOAdb

Number of annotation clusters per database

Conclusions: Our novel database MOAdb presents a useful tool for researchers working on non-classical as well as the classical mammalian model organisms who are interested in the improvement of functional annotation of lists of DEGs in across-species comparisons of transcriptomics datasets. Furthermore, it is possible to translate between NCBI and Ensembl gene IDs, and also adding annotation information from lists of genes or transcripts (from RefSeq).

References: