

## **1. Entrez Gene ID species conversion tool**

### *1.1. Description*

Since ConceptGen uses human Entrez Gene IDs, the mouse IDs from our experiment needed to be converted to human IDs. The ConceptGen species conversion tool only examines homologous genes using the NCBI HomoloGene database and was not used. Instead, a custom script was written that performed a three-step process to convert a gene ID from mouse to human:

1. Any orthologous genes were identified using the Mouse Genome Informatics (MGI) Orthology database. If the mouse gene only had a single human orthology, this was used as the converted gene.
2. If there was more than one orthology, or none at all, the homologous genes were identified using the National Center for Biotechnology Information (NCBI) HomoloGene database. If the mouse gene only had a single human homology, this was used as the converted gene.
3. If the gene also had more than one homology, or no homologies, then the script searched for a human gene with the exact same official symbol in NCBI Entrez Gene database. If an exact match was found, then that gene was used. Otherwise, the gene was marked as not having a human gene ID.

### *1.2. Comparison to ConceptGen conversion tool*

Both conversion tools were tested on all 16,331 on the microarray annotated by the custom CDF. The custom conversion tool successfully converted 14,924 genes (91.38%), and the ConceptGen conversion tool successfully converted 14,366 genes (87.97%). The vast majority of the input mouse genes (14,131) were converted to the same human gene by both tools. However, the custom script was able to convert an additional 558 genes, and so it was selected for use over the ConceptGen conversion tool.