**Comparative Genomic Studies of ORFan genes in Mammalian Genomes**

|  |
| --- |
| Purpose |
| Find orphan genes of Human Chromosome 21 and compare results with the Clamp *at*. el study |

|  |
| --- |
| Method |
| **Step 1: Find all genes of the chromosome**  Ensembl BioMart[[1]](#footnote-1) online tool was used to retrieve all the sequences of the genes belongs to chromosome 21. Following parameters were used in the BioMart search:  *Dataset*: Homo sapiens genes (GRCh38.p7)  *Filters*: Chromosome: 21  *Attributes*: Peptide, Ensembl Gene ID, Ensembl Transcript ID  Sequence unavailable genes were removed from the output file.  **Step 2: Find orphan genes**  Sequences of all the genes were input into the ORFanFinder online tool[[2]](#footnote-2) to find orphan genes. Following parameters were used for the analysis:  *NCBI Taxonomy ID*: 9606  *Program*: blastp  *Use Premade Database*: nr  *E-value*: 1e-3  **Step 3: Compare results with Clamp *at*. el study**  The orphan genes reported from the ORFanFinder were compared against Clamp *at*. el study results. |

|  |
| --- |
| Results |
| 835 genes were found in the Human chromosome 21 and the sequences were saved as a FASTA file. After the analysis, ORFanFinder found 26 species-level orphan genes and other orphan genes in deferent taxonomy ranks as follows:  Screen%20Shot%202016-11-16%20at%203.39.41%20PM.png  It was found that, none of the genes that were excluded from the orphan gene list (Retroposons /Pseudogenes, Cross-species Orthologs, Cross-species Paralogs, Human-specific Paralogs, Pfam domains) by Clamp *at* el(2007) were matched with the results of OrphanFinder. However, the orphan gene list found by Clamp *at* el(2007) is **not** **accessible**(URLs are obsolete or not available) to view or download, therefore those genes cannot be compared with the genes found by ORFanFinder. |

|  |
| --- |
| Discussion |
| Clamp *at*. el (2007) stated that the human Open Reading Frames(ORFs) that lack cross-species counterparts, are termed as ‘‘orphan genes’’. However, they have only compared the human genome with dog and mouse species, which leads to a doubt that whether it is fair enough to compare the genome only with two species to claim for orphan genes (orphan genes are the genes that are belongs to only human, but not visible in other closely related species).  Further, ORFanFinder found only 26 species-specific (they call them “strict”) orphan genes for the chromosome 21. In contrast, Clamp *at*. el (2007) found 1,285 orphan genes for the entire human genome (excluding chr Y and MT). This could be due to several reasons such as:   1. ORFanFinder looking for only chromosome 21 genes, therefore search is get narrow down 2. Chromosome 21 is the shortest chromosome; therefore, it may contain less genes 3. Clamp *at*. el (2007) matches the human genome against only two species, but ORFanFinder matches the human genome against all the other species, therefore the chance to find a species-unique gene is dramatically reduced. 4. Current NCBI non-redundant database is highly rich with sequences compared to the comparative sequences (dog and mouse) in 2007. |

|  |
| --- |
| Conclusion |
| ORFanFinder finds more confident orphan genes due to broad search space (many species), but not specifically looking to issues like pseudogenes/transposons. Therefore, these results should be validated by comparing with other studies on human orphan genes and also by filtering pseudogenes/transposons. |

1. http://asia.ensembl.org/biomart/martview [↑](#footnote-ref-1)
2. http://cys.bios.niu.edu/orfanfinder/index.php [↑](#footnote-ref-2)