## Genes of the Quarter: Excision-Repair Proteins

In order to maintain the integrity of DNA, various repair mechanisms exist in cells repair damage caused by to endogenous and exogenous factors. One such mechanism is nucleotide excision repair (NER). This versatile repair mechanism senses and responds to lesions and distortions in DNA, such UV-induced dimers and bulky as chemical adducts. When this mechanism fails, genetic disorders develop; in humans, these include xeroderma pigmentosum and Cockayne syndrome, both characterized by photosensitivity and



photosensitivity and neurological *Part of the annotation graph for ERCC4* dysfunction.

A conserved set of seven NER proteins has been annotated in nine species (*H. sapiens, M. musculus, R. norvegicus, C. elegans, S. pombe, D. discoideum, D. melanogaster* and *S. cerevisiae*). All proteins in the set are associated with the biological process term "nucleotide-excision repair" or one of its child terms. Each protein has a specific activity within the repair process, and this is reflected in the GO molecular function annotations: ERCC3 and ERCC2 are helicases that open the denaturation bubble of the DNA lesion, and are annotated to "3'-5' DNA helicase activity" and "5'-3' DNA helicase activity" respectively; ERCC5 and the heterodimer ERCC4/ERCC1 cut the damaged strand around the lesion, and are both annotated to "endodeoxyribonuclease activity". Cellular component annotations are to "nucleotide-excision repair complex" or "holo TFIIH complex", a general transcription factor complex involved in DNA repair. These annotations are very similar across the nine species curated, which suggests a highly conserved mechanism of DNA repair that may have arisen early in evolution.

Annotations and the full version of the ERCC4 graphic (above), as well as those for other genes, can be accessed at the GO website.

Graphs: http://www.geneontology.org/images/RefGenomeGraphs/ Annotations: http://www.geneontology.org/GO.current.annotations.shtml

## Announcements

## Dual taxon annotations now in the GO association files

For several years, GO has had the ability to capture the taxon IDs of both the organism encoding the annotated gene product, and another organism with which the gene product interacts. This is needed to fully annotate gene products involved in symbiotic (including pathogenic) interactions. Gene association files containing dual taxon information are now available in the repository; see *P. syringae*, *M. grisea* and *Oomycetes*. For more information on how dual taxon annotations should be created and used, please see

http://www.geneontology.org/GO.annotation.conventions.shtml#interactions.

#### Process-function relationships postponed

The introduction of relationships between the biological process and molecular function ontologies has been delayed until 9 February 2009. More information is available on the GO wiki at <a href="http://wiki.geneontology.org/index.php/FP-regulates">http://wiki.geneontology.org/index.php/FP-regulates</a>.

# **Resource List: ID Mapping**

Translating between different gene and protein identifiers can be a problem when looking at annotations from different databases. There are a number of web-based resources available to help.

UniProtKB (choose 'ID mapping' tab) http://www.uniprot.org/ EBI Protein Identifier Cross-Reference Service (PICR) http://www.ebi.ac.uk/Tools/picr/ Synergizer http://llama.med.harvard.edu/cgi/synergizer/translate Onto-Translate (part of Onto-Tools) http://vortex.cs.wayne.edu/projects.htm g:Convert (part of the g:Profiler toolkit) http://biit.cs.ut.ee/qprofiler/qconvert.cqi Rosetta IDconverter (Babelomics) http://www.babelomics.org MatchMiner (part of the Miner suite of tools) http://discover.nci.nih.gov/matchminer/ SOURCE http://source.stanford.edu Gene Name Converter (part of GeneMerge) http://genemerge.bioteam.net/convertgenenames

Draghici *et al.* compare the performance of some of these tools in a 2006 paper, available as a PDF at <u>http://vortex.cs.wayne.edu/papers/Babel.tower.pdf</u>.

Mappings files between UniProtKB and RefSeq accessions for equivalent proteins are available from <u>ftp://ftp.ncbi.nih.gov/gene/DATA/</u>.

# **Upcoming Events**

- 1-5 Sept 2008: MGED 11
  Riva del Garda, Trentino, Italy
  <u>http://www.mgedmeeting.org/</u>
- 10-14 Sept 2008: Genome Informatics
  Wellcome Trust Genome Campus, Hinxton, UK
  <u>http://meetings.cshl.edu/meetings/infouk08.shtml</u>
- 27-30 Sept 2008: HUGO Human Genome Meeting Hyderabad, India http://hgm2008.hugo-international.org/
- 1-3 Oct 2008: Eukaryotic Genome Annotation and Analysis Course
- 14-16 Oct 2008: Prokaryotic Genome Annotation and Analysis Course
   J. Craig Venter Institute, Rockville, Maryland http://www.jcvi.org/EukaryoticClass
   http://www.jcvi.org/ProkaryoticClass
- 9-10 Sept, 5-6 Nov 2008: IGS Genome Annotation Workshops
   Institute for Genome Sciences, Baltimore, MD http://ae.igs.umaryland.edu/workshop.html
- 16-19 Apr 2009: Third International Biocurator Meeting Badin, Cormony, Mars information to follow:

Berlin, Germany. More information to follow.

View Expanded Newsletter Online: <u>http://www.geneontology.org/newsletter/archive/200808.shtml</u> To Receive Future Newsletters: Subscribe to the GO Friends mailing list (gofriends@geneontology.org) Contact the Gene Ontology Consortium: Please send comments or questions to <u>gohelp@geneontology.org</u>