

Cross-species Annotation of Cardiovascular Processes

Varsha Khodiyar, Ruth Lovering, Reference Genome Group of the Gene Ontology Consortium* and Philippa Talmud

Centre for Cardiovascular Genetics, UCL Department of Medicine, Rayne Institute 5 University Street London WC1E 6JF.



What is Gene Ontology (GO)?

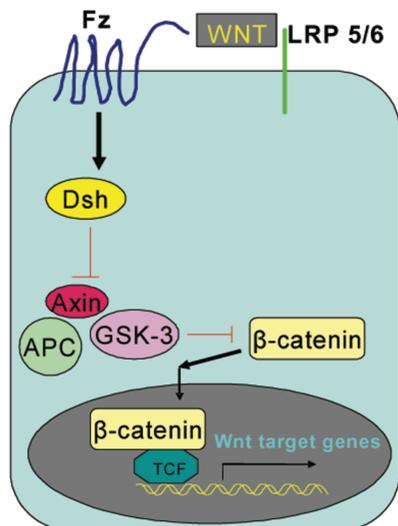
GO is a freely available public resource established by the Gene Ontology consortium (GOC) for the description of gene products (proteins and RNA) across all species¹ (www.geneontology.org). The UCL-based GO curation team is part of the GOC, specializing in GO annotation human cardiovascular proteins². See poster BG1 for more information on the Cardiovascular GO Annotation initiative.

Why annotate non-human species?

Much cardiovascular research is carried out in model organisms and so it is important that relevant information gathered from these species is extrapolated to orthologous human proteins in a consistent and conserved manner. The Reference Genomes project³ has been established by the GOC to maximize and optimize the GO annotation of a large and representative set of key genomes, known as 'reference genomes'. The goal of this project is to completely annotate twelve reference genomes so that those annotations may be used to effectively seed the automatic annotation efforts of other genomes.

Wnt annotation project

The Wnt signaling pathway has a number of distinct roles to play during embryogenesis, including cardiovascular development⁴. Additionally, Wnt signaling occurs in a number of distinct species from fly and frog to human and mouse⁵. Thus it was an ideal choice for annotation with the Reference Genomes' cross-species approach.



From Liu *et al.* *Breast Cancer Research* 2005 7:86

Figure 1 - The canonical Wnt signaling pathway

Establishing orthology groups

The PANTHER (Protein ANalysis THrough Evolutionary Relationships) Classification System⁶ classifies genes by their functions, using published scientific experimental evidence and evolutionary relationships to predict function even in the absence of direct experimental evidence. We used PANTHER to generate Wnt signaling pathway orthologous family groups across all twelve reference genome species.

It was decided to initially focus on the so-called canonical Wnt pathway (figure 1). Each protein in the pathway has multiple paralogs or family members, and all members of each family were annotated in a specified time period by each MOD. The first focus was on the Wnt family (WNT1, WNT2, etc), the second focus was the catenins and so on

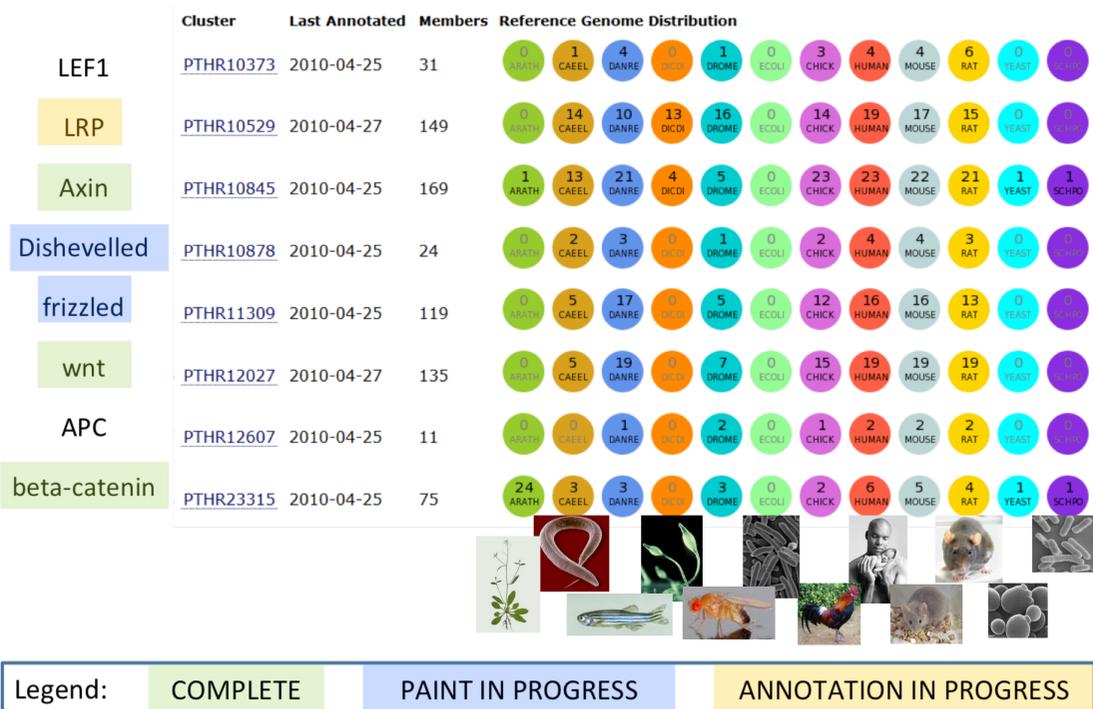


Figure 2 - Progress of reference genomes WNT annotation project

Annotation across phylogenetic trees

Comparative genomics works on the premise that homologous genes descended from a common ancestor often have related functions. The Phylogenetic Annotation INference Tool (PAINT) allows the transfer of experimentally derived GO annotations to the ancestral gene⁷. The ancestral annotations are then propagated to other descendants for which there may be little or no annotatable data.

Figure 2 shows the progress made on the Wnt project to date. The annotation phase is due for completion by the end of September 2010, and the PAINT phase in mid-October. This will be followed by a phase of each MOD reviewing the PAINT derived annotations for consistency and accuracy.

Future plans

This is the first time that the Reference Genome cross-species approach has been applied to a single signaling pathway. Using the procedures established by this project, in particular the defined timeframes, we intend to annotate several more pathways in the future. In particular the UCL-based curation team will bid for Reference Genome project resources to be focused on cardiovascular specific projects. If you have any ideas of specific CV areas which would benefit from this approach we would like to hear from you.

References

- Ashburner, M., Ball, C.A., Blake, J.A., *et al.* *Gene ontology: tool for the unification of biology.* *The Gene Ontology Consortium. Nat Genet*, 2000; 25: 25-29.
- Lovering, R.C., Dimmer, E., Khodiyar, V.K., *et al.* *Cardiovascular GO annotation initiative year 1 report: why cardiovascular GO?* *Proteomics*, 2008; 8(10): 1950-1953.
- Reference Genome Group of the Gene Ontology Consortium. *The Gene Ontology's Reference Genome Project: a unified framework for functional annotation across species.* *PLoS Comput Biol*, 2009; 5(7): e1000431.
- Gessert S., Kuhl M. *The Multiple Phases and Faces of Wnt Signaling During Cardiac Differentiation and Development.* *Circ Res*, 2010; 107(2): 186-199.
- van Amerongen R., Nusse R. *Towards an integrated view of Wnt signaling in development.* *Development*, 2009; 136(19): 3205-3214.
- Thomas P.D., Campbell M.J., Kejariwal A., *et al.* *PANTHER: a library of protein families and subfamilies indexed by function.* *Genome Res*, 2003; 13: 2129-2141.
- <http://wiki.geneontology.org/index.php/PAINT>

*Reference Genome Group

Annotators: Pascale Gaudet, Petra Fey, Rex Chisholm (dictyBase), Tanya Berardini, Donghui Li (TAIR), Emily Dimmer, Rachael Huntley (GOA), Ruth C. Lovering, Varsha K. Khodiyar (UCL), Julie Park, Karen Christie, Stacia R. Engel (SGD), David P. Hill, Li Ni (MGI), Doug Howe (ZFIN), Jim Hu, Deborah Siegel (E.coliWiki), Kimberly Van Auken, Ranjana Kishore (WormBase), Fiona McCarthy (AgBase), Victoria Petri, Stan Laulederkind (RGD), Susan Tweedie (FlyBase), Valerie Wood (PomBase).

Computational staff representatives: Siddhartha Basu (dictyBase), Seth Carbon (BBOP), Mary Dolan (MGI), Chris Mungall (BBOP), Sven Heinicke (PPOD); Protein families generated by Paul Thomas (PANTHER). Trees annotated by Kara Dolinski (PPOD), Michael S. Livstone (PPOD) and Pascale Gaudet (dictyBase).

PIs of the GO Consortium: Michael Ashburner (FlyBase), Judith Blake (MGI), Mike Cherry (SGD), Suzi Lewis (BBOP) and Paul Thomas (PANTHER).



The UCL-based curation team is supported by the BHF, grant SP/07/007/23671



For more information about contributing to the annotation of cardiovascular genes and proteins email GOannotation@UCL.ac.uk or see our website www.cardiovasculargeneontology.com