

Cardiovascular Gene Ontology Annotation Initiative

Varsha Khodiyar¹, Daniel Barrell², Peter Scambler³, Mike Hubank⁴, Rolf Apweiler², Philippa Talmud¹ and Ruth Lovering¹

¹Centre for Cardiovascular Genetics, UCL Department of Medicine, Rayne Institute 5 University Street London WC1E 6JF.

²European Bioinformatics Institute, Hinxton, Cambridge, CB10 1SD.

³Molecular Medicine Unit, Institute of Child Health, 30 Guilford Street, London WC1N 1EH.

⁴Molecular Hematology and Cancer Biology Unit, Institute of Child Health, 30 Guilford Street, London WC1N 1EH.

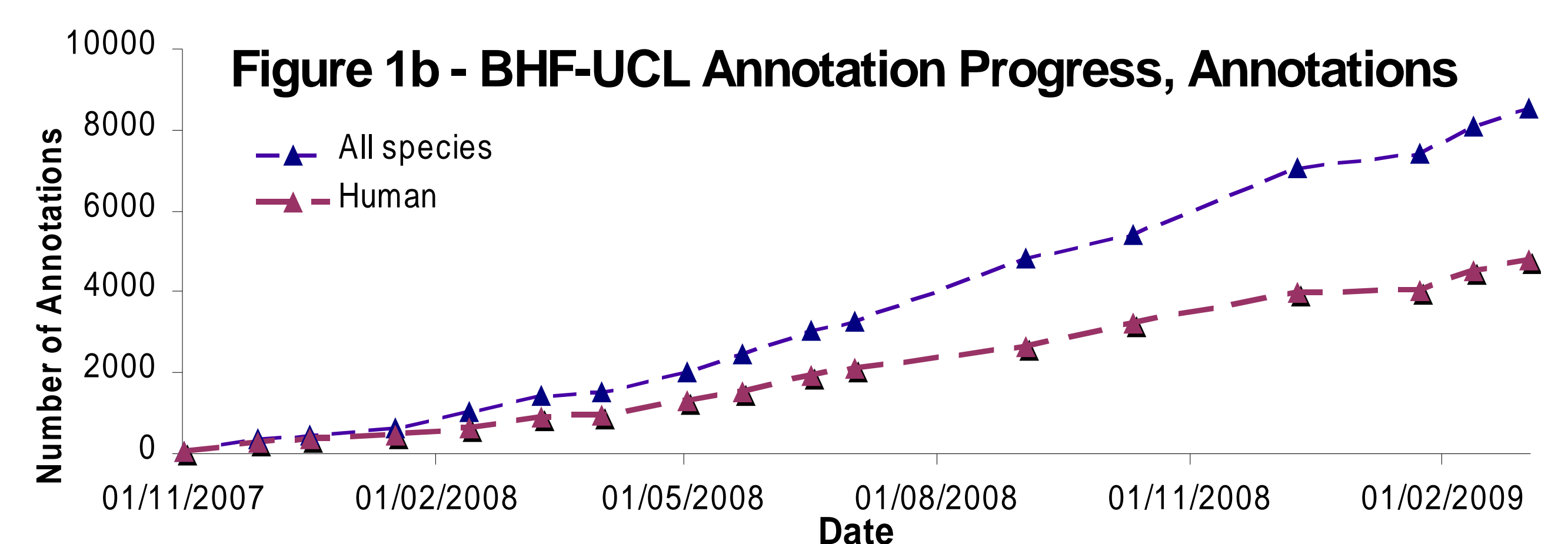
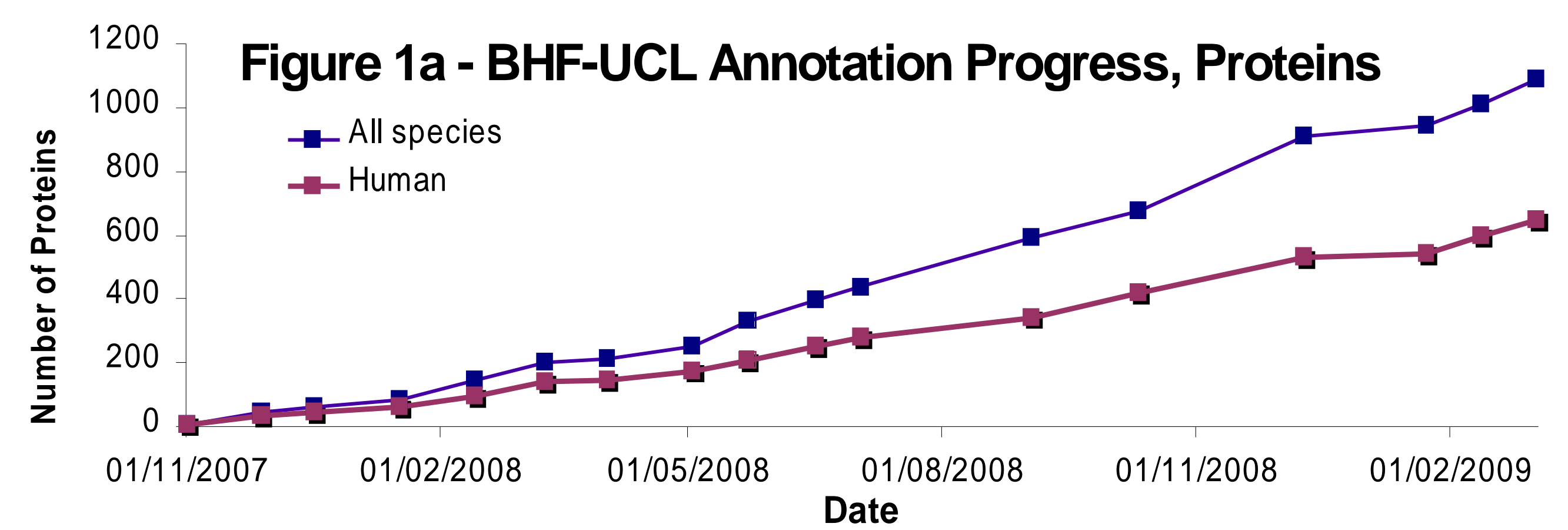


Gene Ontology (GO) provides a controlled vocabulary, which is used by several groups around the world to provide functional annotation to proteins across a wide range of species (www.geneontology.org). The Cardiovascular Gene Ontology Annotation Initiative is funded the British Heath Foundation to supply GO annotation specifically for human proteins involved in cardiovascular (CV) processes. This is the first time that a physiological process-centred approach has been used for human protein GO annotation¹. Experienced GO curators from the BHF-UCL team work alongside the bench scientists from the CV genetics group at University College London (UCL). By working in a CV dedicated environment the GO curators are developing their expertise in this field; this is leading to more detailed and accurate GO annotation of CV-relevant genes. Since the start of the project in November 2007 we have added over 4700 annotations to 649 human CV related proteins, as shown in figure 1.

Identification of CV related genes and proteins

Over 4000 CV-related genes were identified by combining:

- the ITMAT consortium list of approximately 2,200 genes²
- a list of 282 congenital heart disease candidate genes³
- a list of approximately 2,500 genes identified by the BHF-UCL team as associated with CV-relevant GO terms at the start of the project
- 170 genes provided by our advisory board



Process-centric annotation method

We followed a gene by gene annotation approach for the first year of the project. However, whilst reading the literature for a single gene, we often annotated several other genes within the same pathway as they would be discussed within the same papers. Thus for the second year we decided to trial a process-centric approach.

The benefit of annotating in this process-centric manner is a deeper understanding of the CV-related pathways by the BHF-UCL team. This increased understanding generally results in:

- increased granularity of GO annotation
- greater consistency of GO annotations across each process
- improved input to ontology development

For example, during the annotation of the process lipoprotein particle clearance, the CV genetics group at UCL provided considerable advice on the appropriate GO terms to use for each of the human apolipoproteins, identified relevant publications for the BHF-UCL team to annotate, and contributed to the development of the lipoprotein particle cellular component ontology as shown in figure 2.

Conclusions

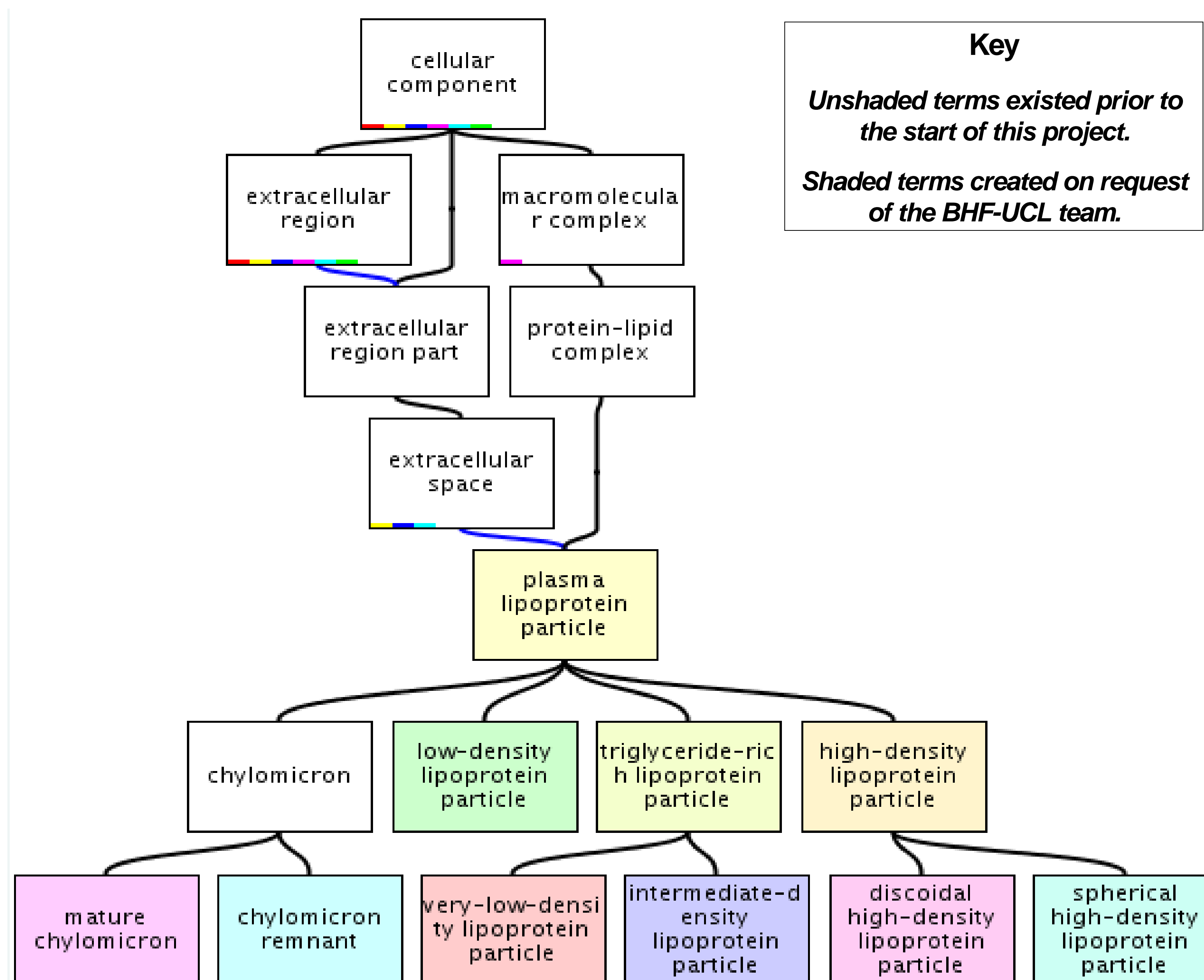
- A gene-by-gene annotation process promotes the complete annotation of a single gene.
- A process-centric approach allows the annotator to gain a deeper understanding of a specific pathway, and promotes the annotation of all the genes involved within that pathway.
- A major benefit of undertaking GO annotation in a process-centric approach is the development and refinement of the ontology, which occurs alongside the annotation process.

Contact us: Email: GOannotation@UCL.ac.uk
Wiki: wiki.geneontology.org/index.php/Cardiovascular
Website: www.cardiovasculargeneontology/feedback

References

- Lovering RC, Dimmer E, Khodiyar VK, Barrell DG, Scambler P, Hubank M, Apweiler R, Talmud PJ. Cardiovascular GO annotation initiative year 1 report: why cardiovascular GO? *Proteomics* 2008, 8:1950-1953
- Vascular Disease 50k SNP Array Consortia bmc.upenn.edu/cvdsn/index.php
- Bentham J, Bhattacharya S. Genetic mechanisms controlling cardiovascular development. *Ann N Y Acad Sci.* 2008, 1123:10-19
- Harris M.A. Developing an ontology. *Methods Mol Biol.* 2008, 452: 111-124

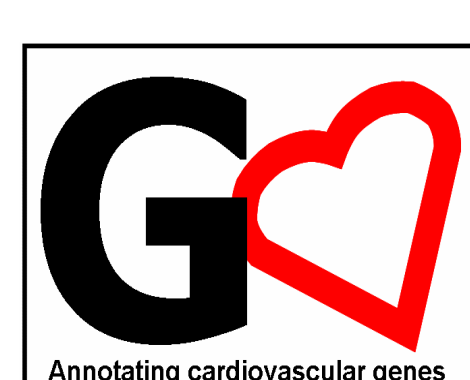
Figure 2 – How the GO was developed to reflect published knowledge on lipoprotein particle biology.



Ontology Development

The GO editorial office develops and refines the GO in small, medium and large scale projects⁴. The close association of the BHF-UCL annotation team in an active bench science CV research laboratory and the intensive reading required as part of the annotation process, means that the BHF-UCL team has gained considerable expertise in various aspects of CV system, processes and development. Therefore the BHF-UCL team is able to successfully instigate both small and medium scale GO development projects in CV related areas with the GO editorial office.

Figure 2 shows one example of an ontology development project that was undertaken by the BHF-UCL team. New terms were developed in consultation with lipoprotein researchers based at the UCL CV genetics laboratory.



Grant: SP/07/007/23671



BEATING HEART DISEASE TOGETHER

www.cardiovasculargeneontology.com