GOing forward with the cardiac conduction system using Gene Ontology

Kan Yan Chloe Li[1, 2], Andrew C. Cook[2], Ruth C. Lovering[1]

[1]Functional Gene Annotation, Preclinical and Fundamental Science, UCL Institute of Cardiovascular Science, London, UK
[2]Centre for Cardiac Morphology, UCL Institute of Cardiovascular Science, London, UK

Correspondence: kan.li.14@ucl.ac.uk

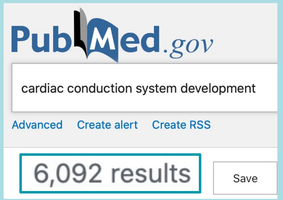
1. Background

What is Gene Ontology (GO)?

The Gene Ontology Consortium (GOC) provides a structured, controlled vocabulary known as the Gene Ontology (GO; http://geneontology.org/). GO describes the characteristics of genes and gene products in any organism and does this in 3 levels: molecular function, biological process and cellular component. GO and GO annotations are freely available and frequently updated.

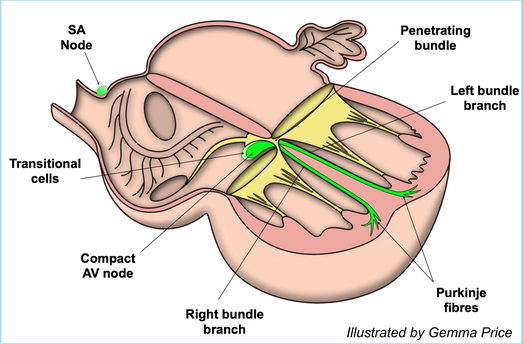
Why do we need ontologies?

- To cope with large volumes of biological data from many sources
- To organise and analyse data and share it universally
- English can be an imprecise language



Cardiac conduction system (CCS):

The electrical CCS consists of the sinoatrial node (SAN), atrioventricular node (AVN) and the ventricular conduction system (VCS). The VCS comprises the bundle of His (atrioventricular bundle, AVB) which bifurcates into the left and right bundle branches (BB).



The components of the CCS are essential for the initiation, propagation, and coordination of each action potential that travels across the heart. Aberrant CCS development can cause conduction abnormalities, such as sick sinus syndrome and atrioventricular and bundle branch blocks.

Most GO annotations have focussed on heart development as opposed to CCS development.

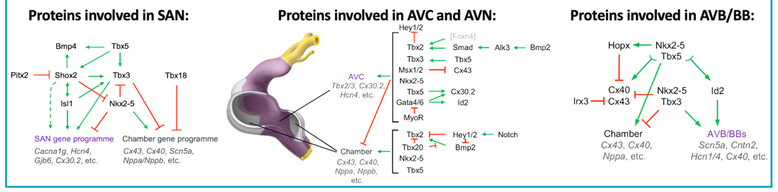
2. Aims

The main priority of this project was to use GO terms to capture the contribution of 35 essential proteins (listed in [1]) to CCS development and submit these to the GO annotation database.

3. Methods

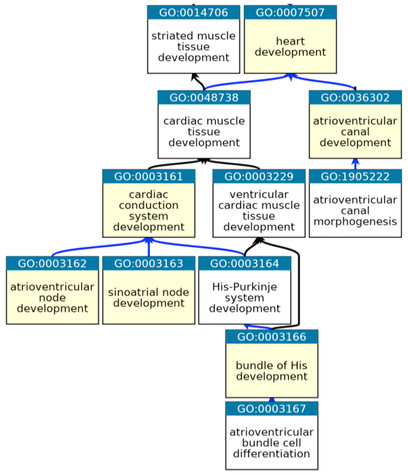
A list of 35 proteins reported to be required for CCS development was made based on [1]

17 proteins prioritised based on the proposed pathways mentioned in [1]



14 of the 17 prioritised proteins were eligible for curation

BMP2, BMPRA1, GJA5, GJD3, IRX3, ISL1, NKX2-5, PITX2, SHOX2, TBX2, TBX3, TBX5, TBX18, TBX20



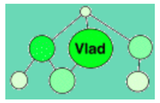
Curate proteins using appropriate annotation extension statements

Protein2GO X

Investigate annotations using 6 GO terms used as GO slims

Quick GO

Functional enrichment analysis of GO annotations (October 2020 vs March 2021)



4. Results

152 descriptive GO annotations were created, including those describing SAN and AVN development. A full list of GO terms associated with the 35 prioritised proteins can be found using https://tinyurl.com/2h8wutmy.

In October 2020, the total number of CCS development annotations associated with the 35 priority proteins was 3. Our project has increased this number to 57 annotations. An additional 157 annotations related to heart development were associated with these proteins in October 2020 and this number was increased to 350 (March 2021) by this project.

Impact of cardiac conduction-focussed annotation project

GO term	Number of annotations	
	October 2020	March 2021
Cardiac conduction system development	3	22
Atrioventricular node development	0	9
Sinoatrial node development	0	6
Bundle of His development	0	10
Total CCS development	**3**	**57**
Heart development	177	350
Atrioventricular canal development	3	17
Total number of GO annotations	**180**	**407**

5. Discussion

This project has expanded the available GO description of CCS development and heart development by creating 152 new annotations for human proteins of which 54 of the annotations described CCS development. Other development terms arose from this project even though the key priority of this study was to curate proteins involved in CCS development.

Our contribution to the GO database may elucidate the key mechanisms involved in CCS disorders such as Wolf-Parkinson-White Syndrome and sick sinus syndrome.

References

[1] van Weerd and Christoffels (2016). The formation and function of the cardiac conduction system. *Development*, 143; 197-210.