

Using focused Gene Ontology curation to describe human telomere biology

Jessica Buxton, Nancy Campbell, Rachael Huntley and Ruth Lovering

Cardiovascular Gene Annotation group, Dept. of Cardiovascular Genetics, Institute of Cardiovascular Science, UCL



BACKGROUND

Coronary heart disease (CHD) and type 2 diabetes (T2D) are caused by a myriad of interacting genetic and environmental factors, but the risk of both is strongly influenced by increasing age.

Telomeres are protective nucleoprotein structures that cap the ends of linear chromosomes (see **Fig 1**), which shorten over time in most tissues. Recent genetic studies indicate a causal role for short telomeres in CHD¹. However, the mechanisms underlying the relationship between telomere shortening and risk of cardiometabolic disease have yet to be determined.

We undertook a manual annotation project focussed on gene products either known or suspected to be involved in telomere biology, using **Gene Ontology (GO)** terms². The GO resource provides dynamic, controlled vocabularies that allow consistent descriptions of the functional attributes and subcellular locations of all gene products (www.ucl.ac.uk/functional-gene-annotation).

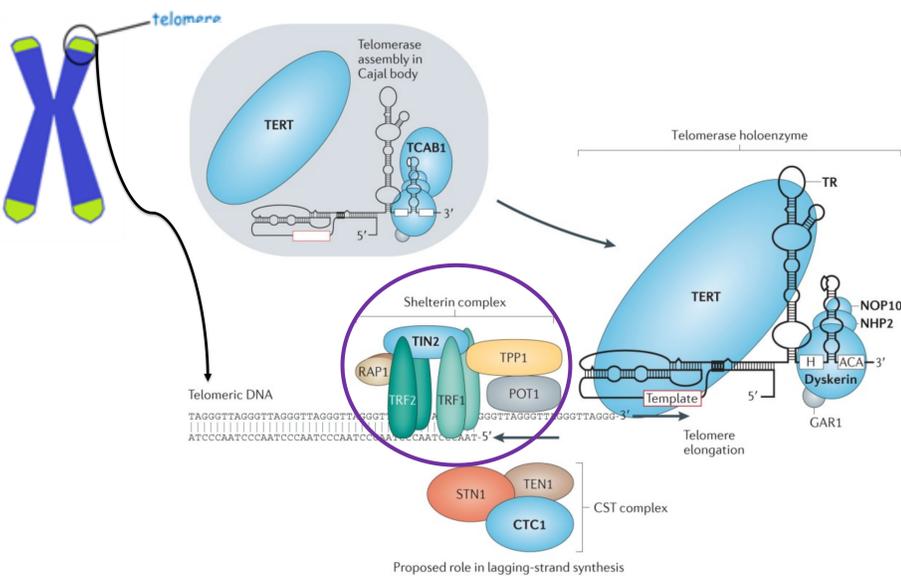


Fig. 1 Telomerase and telomere components³. In vertebrates, telomeres are composed of repeating units of (TTAGGG)_n, bound to a protective protein complex, shelterin (circled above)

METHODS

A list of human proteins, RNA molecules and macromolecular complexes known or suspected to be involved in telomere maintenance was drawn up. Using existing and newly created GO terms, annotations were made to these entities based on published evidence relating to their roles in telomere maintenance. Project progress was monitored by assessing the number of annotations present in the GO database with the GO terms (and their 'child' terms) listed in **Table 1**.

Parent GO term ID	Name	Annotations
GO:0000781	chromosome, telomeric region	192
GO:0000723	telomere maintenance	255
GO:0007004	telomere maintenance via telomerase	86
GO:0003720	telomerase activity	68
GO:0042162	telomeric DNA binding	61
GO:0005697	telomerase holoenzyme complex	39
GO:0090672	telomerase RNA localization	16
GO:0070034	telomerase RNA binding	28
GO:0000782	telomere cap complex	39
GO:0010521	telomerase inhibitor activity	12
GO:0098505	G-rich strand telomeric DNA binding	8
GO:0090661	box H/ACA telomerase RNP complex	8
GO:1904868	telomerase catalytic core complex assembly	2
GO:0061752	telomeric repeat-containing RNA binding	3
GO:0000333	telomerase catalytic core complex	8
GO:0090669	telomerase RNA stabilization	1
GO:0003721	telomerase RNA reverse transcriptase activity	6

Table 1 Terms relating to telomere maintenance in the GO database
The number of current annotations to each of these terms (or one or more child term) is given

RESULTS

A total of **91** scientific articles have so far been reviewed and used to create 1,151 manual annotations to proteins, including 503 annotations to the telomere maintenance-related GO terms (and relevant child terms) in **Table 1**. These include **45** new terms created during the course of this work.

This project has resulted in a **>200%** increase in the number of telomere-relevant annotations in the GO database. To achieve this, new and existing GO terms were associated to 147 human proteins, including 49 that had no previous telomere-related annotations (**Table 2**). As a result, there are currently **205** proteins with a total of **832** telomere-relevant annotations present in GO.

	Proteins	Telomere-related annotations
Before	98	329
After	205*	832*
% increase	209%	253%

Table 2: Human proteins with telomere maintenance-related annotations in GO
*Accessed from www.ebi.ac.uk/QuickGO on 8/1/16

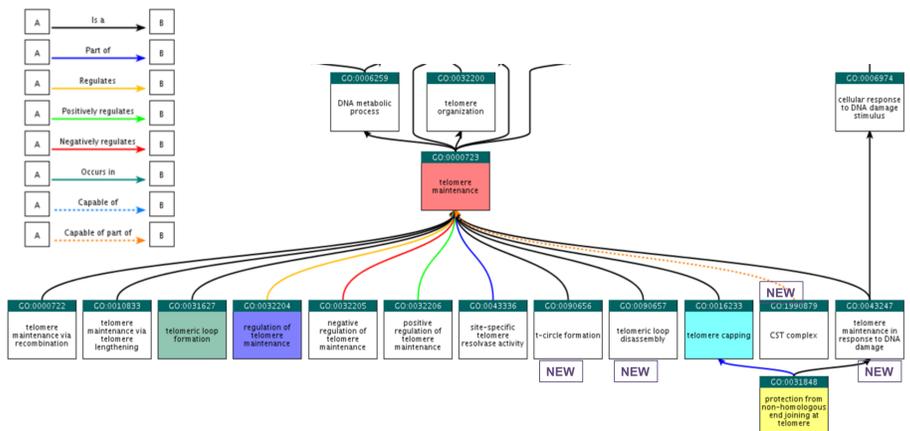


Fig. 2 Screenshot showing part of the GO ontology relating to telomere maintenance
The highlighted GO terms were used to annotate components of shelterin, the protective protein complex that binds to telomeric DNA (**Fig 1**) (www.ebi.ac.uk/QuickGO accessed 12/1/16). Colours of arrows joining GO terms indicate nature of relationships between them (see key). 4 of the 45 new terms created during this project are also shown, marked **NEW**

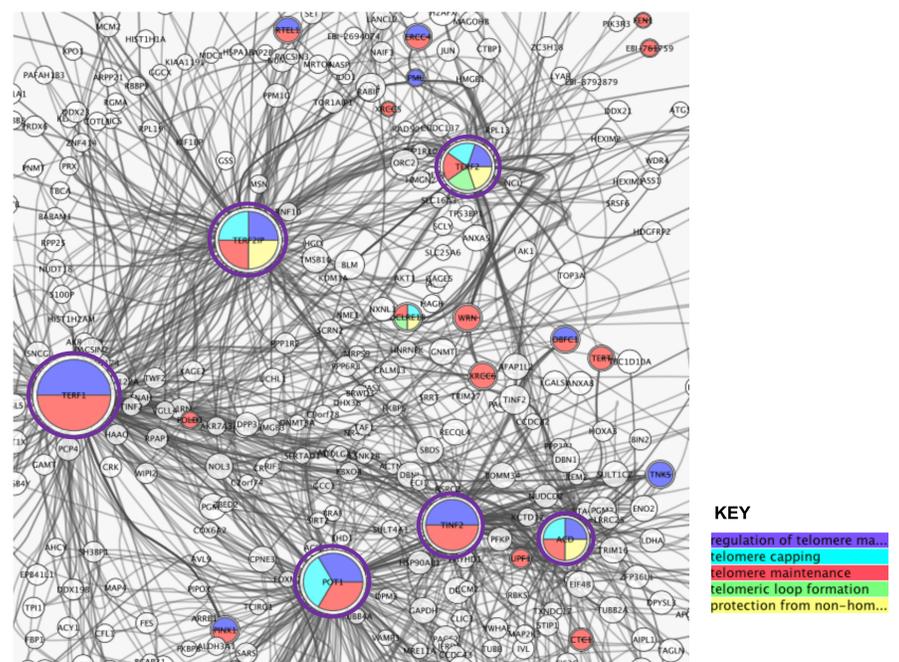


Fig. 3 Shelterin protein interaction network. The circled nodes are the 6 core shelterin proteins. Colours inside nodes indicate telomere maintenance GO terms associated with each protein see key (also match colours of terms in **Fig 2**) Created using Cytoscape v. 3.3.0⁴.

CONCLUSION

We have substantially enriched and improved the ontology relating to vertebrate telomere biology in the GO database. Accurate and comprehensive gene annotations are essential for optimising the performance of statistical analysis tools used to interpret genomic, transcriptomic and proteomic datasets. This work will support the rapid evaluation of new experimental data, and help elucidate the role of telomere shortening in disease.



References

- Codd V, Nelson CP, Albrecht E, Mangino M, Deelen J, Buxton JL et al. Identification of seven loci affecting telomere length and their association with disease (2013). *Nature Genetics*, 45: 422–427.
- Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene Ontology: tool for the unification of biology (2000). *Nature Genetics*, 25: 25–9.
- Armanios and Blackburn. The telomere syndromes (2012). *Nature Reviews Genetics* 13: 693–704.
- Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D et al. Cytoscape: a software environment for integrated models of biomolecular interaction network (2003). *Genome Res*. 13: 2498–504.