# Telomerase Holoenzyme Complex: The Impact of Focused Gene Ontology Curation

**Nancy Campbell*, Jessica Buxton, Rachael Huntley & Ruth Lovering**

Functional Gene Annotation Group, Genetic Epidemiology & Translational Cardiovascular Genomics, Institute of Cardiovascular Science

## BACKGROUND

Coronary heart disease (CHD) & type 2 diabetes (T2D) are caused by a myriad of interacting genetic & 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 **Figure 1**), which shorten over time in most tissues. Recent genetic studies indicate a causal role for short telomeres in CHD[1]. However, the mechanisms underlying the relationship between telomere shortening & risk of cardiometabolic disease have yet to be determined.

The GO resource provides dynamic, controlled vocabularies that allow consistent descriptions of the functional attributes & subcellular locations of all gene products[2]. Accurate & comprehensive gene annotations are essential for optimising the performance of statistical analysis tools used to interpret genomic, transcriptomic & proteomic datasets.

We undertook a manual annotation project focused on gene products either known or suspected to be involved in telomere biology, using **Gene Ontology (GO)** terms[2]. The **telomerase holoenzyme complex** is used here as a case study to demonstrate the impact of this annotation effort.

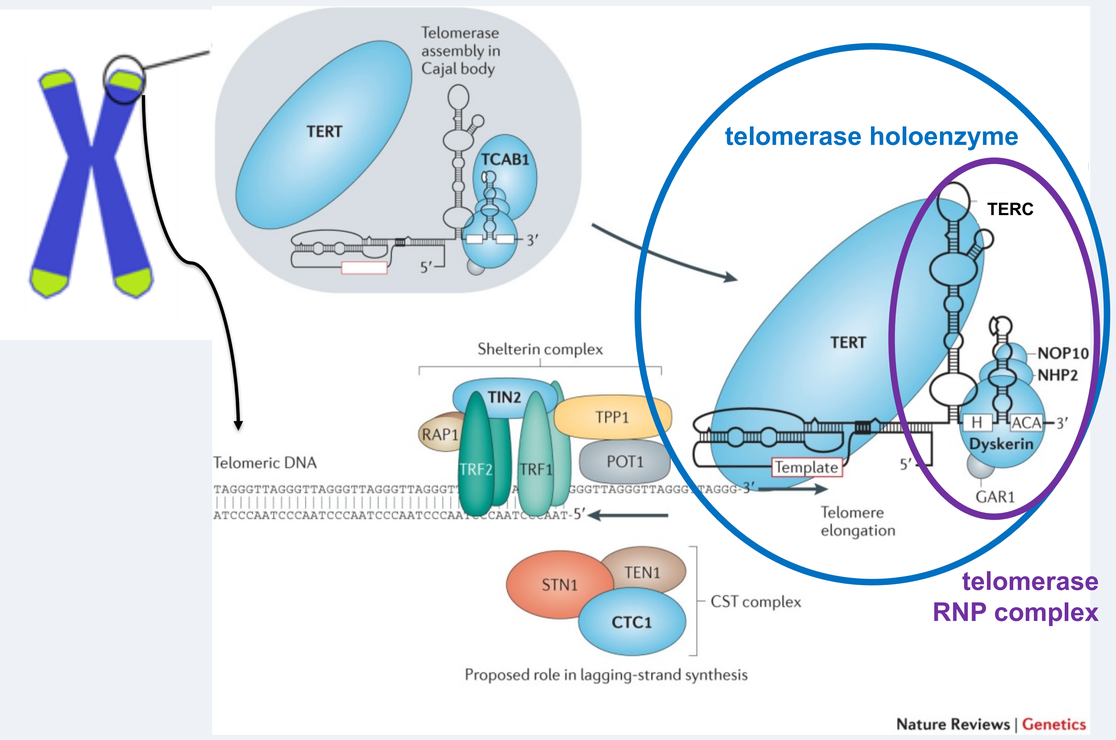


**Figure 1 Telomerase & telomere components[3].** In vertebrates, the telomerase holoenzyme complex (dark blue oval) is composed of telomerase (TERT) & the box H/ACA telomerase RNP complex (purple oval). The latter is formed of the telomerase RNA (TERC) & the proteins NOP10, NHP2, Dyskerin, GAR1.

## METHODS

**A Molecular Interactions Network** was created using Cytoscape[4] v3.5 by querying the web server PSICQUIC[5] for the interacting partners of six components of the telomerase holoenzyme complex: TERT, NOP10, NHP2, Dyskerin, GAR1 & TERC. Interactions from BHF-UCL, IntAct, MINT & EBI-GOA-nonIntAct databases were imported & merged.

**Enrichment Analysis & Colour-coding the Interactors (nodes)** with enriched functional GO terms were performed using BiNGO[6] v3.0.3 & Golorize[7] v1.0.0beta1.

**Gene association files** release 124 and 165 were downloaded from the QuickGO[8] website coinciding with release dates 22nd of July 2013 and 13th February 2017, respectively. The **Ontology file** releases/2017-02-14 format version:1.2 was downloaded from www.geneontology.org

**The Impact** was assessed by comparing the presence of four relevant pre-existing & new enriched terms **(Table 1)** prior to & following recording the appropriate annotations.

| TERM | GO ID | Aspect |
|---|---|---|
| telomerase RNA binding | GO:0070034 | Molecular Function |
| telomerase holoenzyme complex | GO:0005697 | Cellular Component |
| box H/ACA telomerase RNP complex* | GO:0090661 | Cellular Component |
| telomerase catalytic core complex | GO:0000333 | Cellular Component |

**Table 1 Terms relating to telomerase holoenzyme complex & its individual components in the GO database.** * New GO term created by this project.

## RESULTS

Enrichment of the four GO terms in **Table 1** is observed for the appropriate components of the telomerase holoenzyme complex and its interactors (see **Figure 2).** This enrichment reflects the relevant annotations whereby 62 of the 70 annotations made to 16 of the 23 gene products were the result of our focused effort.

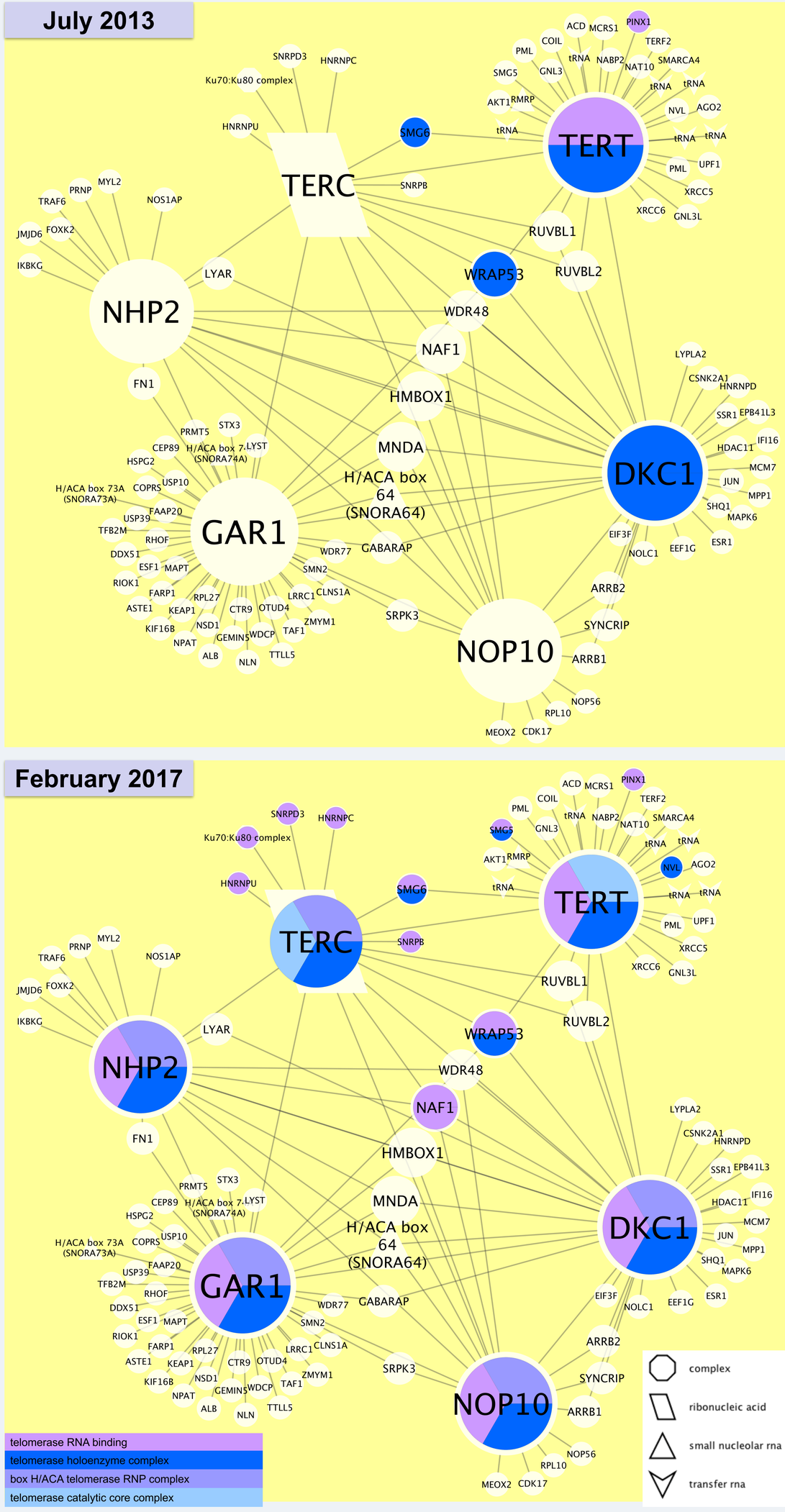


**Figure 2 The molecular interaction network for the components for the telomerase holoenzyme complex in human.** Colours inside nodes indicate enrichment with the relevant four GO terms associated with each protein, see key above.

## CONCLUSION

We have substantially enriched & improved the ontology & associated annotations relating to vertebrate telomere biology in the GO database. We have demonstrated the impact of our focused annotation approach by using the telomerase holoenzyme complex as a showcase. This work will support the rapid evaluation of new experimental data, & help elucidate the role of telomere shortening in disease.

Grant RG/13/5/30112

**References**

1. Codd, V. *et al.* Identification of seven loci affecting mean telomere length and their association with disease. *Nat. Genet.* **45,** 422–427, 427–2 (2013).
2. Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* **25,** 25–29 (2000).
3. Armanios, M. & Blackburn, E. H. The telomere syndromes. *Nat. Rev. Genet.* **13,** 693–704 (2012).
4. Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13,** 2498–2504 (2003).
5. del-Toro, N. *et al.* A new reference implementation of the PSICQUIC web service. *Nucleic Acids Res.* **41,** W601-606 (2013).
6. Maere, S., Heymans, K. & Kuiper, M. BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics* **21,** 3448–3449 (2005).
7. Garcia, O. *et al.* GOlorize: a Cytoscape plug-in for network visualization with Gene Ontology-based layout & coloring. *Bioinformatics* **23,** 394–396 (2007).
8. Binns D, Dimmer E, Huntley R, Barrell D, O'Donovan C, Apweiler R. QuickGO: a web-based tool for Gene Ontology searching. Bioinformatics. 2009 Nov 15;25(22):3045–6. 1.

*nancy.campbell@ucl.ac.uk

Visit our website, email us for more information or to receive our quarterly newsletter, or follow us on Twitter:

***www.ucl.ac.uk/functional-gene-annotation***

**goannotation@ucl.ac.uk**

**@UCLgene**