COMBAT TB Explorer,
a TB data exploration workbench.

Peter van Heusden (pvh@sanbi.ac.za), Ziphozake Mashologu, Alan Christoffels (alan@sanbi.ac.za)
South African National Bioinformatics Institute, University of the Western Cape

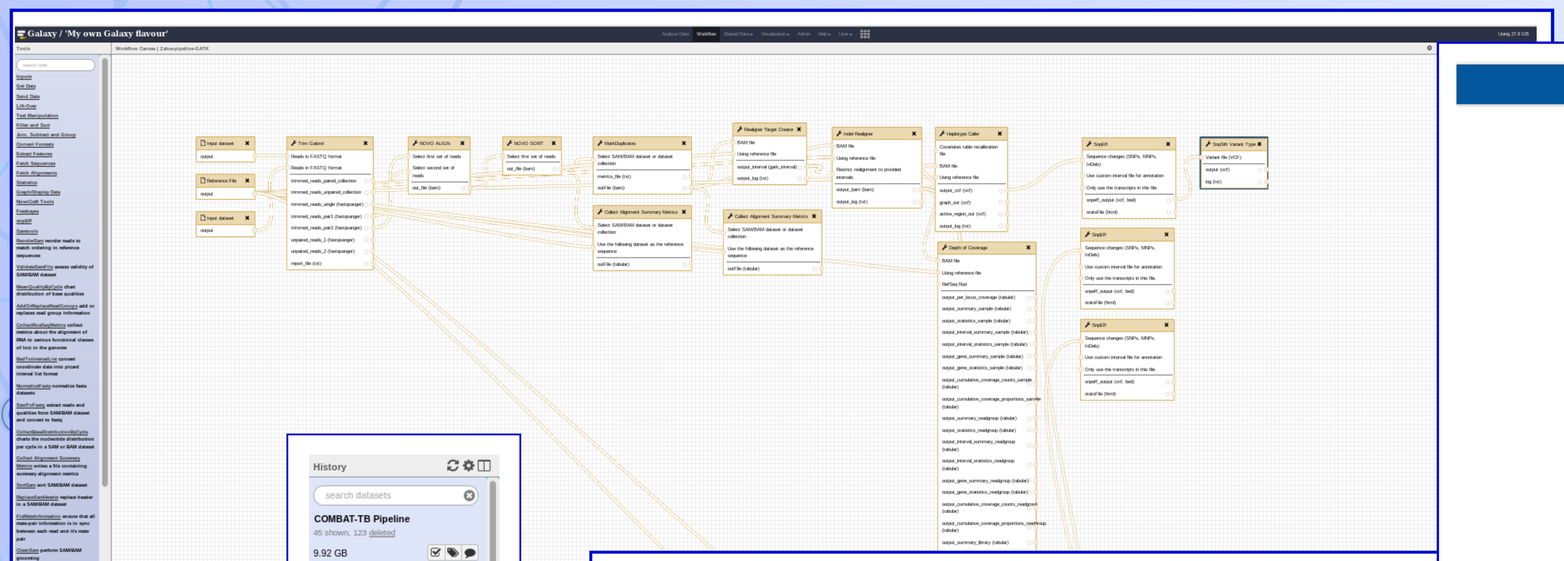
Introduction

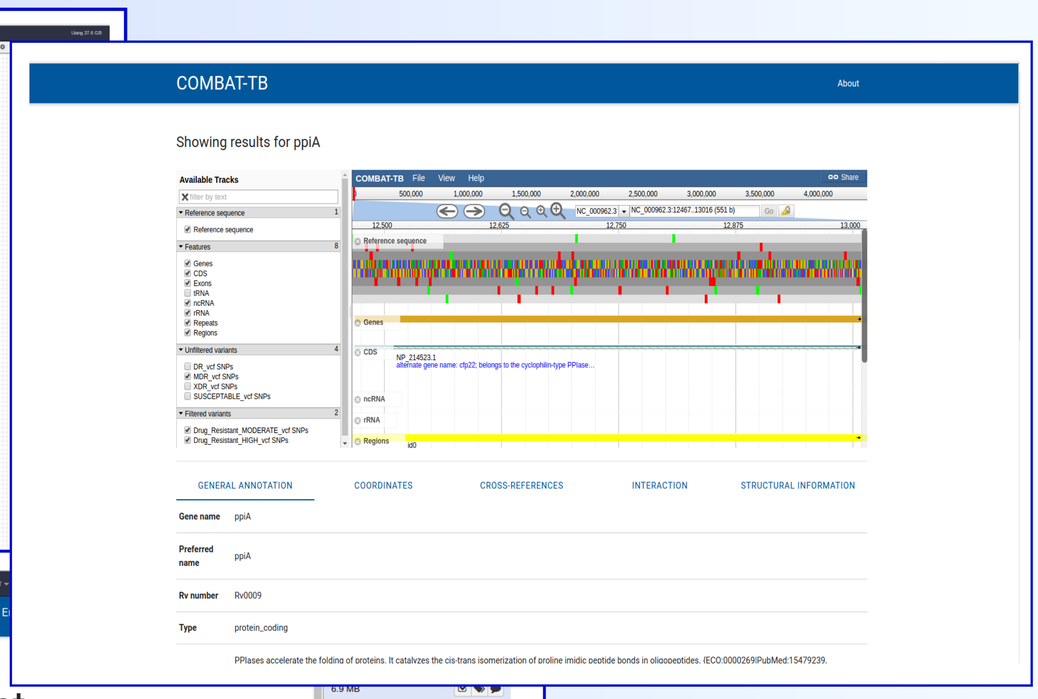
Tuberculosis (TB), an infectious disease caused by the *Mycobacterium tuberculosis*, ranks as one of the leading causes of death worldwide, with WHO recording 9.6 million people falling ill with 1.5 million deaths in 2014. This disease burden has arguably been matched with continued increase in genomic, transcriptomic and proteomic data for *M. tuberculosis* as a result of NGS technologies. This continued expansion of data is exemplified by the growth of data repositories such as the tuberculosis database (TBDB) and the pathosystems resource integrated center (PATRICBRC). Unfortunately, these resources only present pre-computed data and do not provide the computational toolkit for biomedical researchers to analyze their own data. We have created the COMBAT TB Explorer, a Galaxy-based environment for annotating and exploring *M. tuberculosis* sequence data.

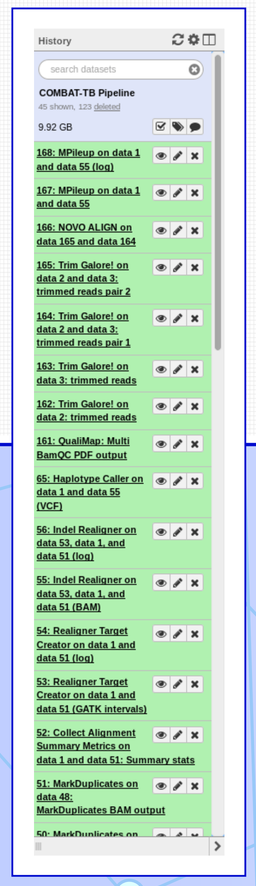
Methods and Results

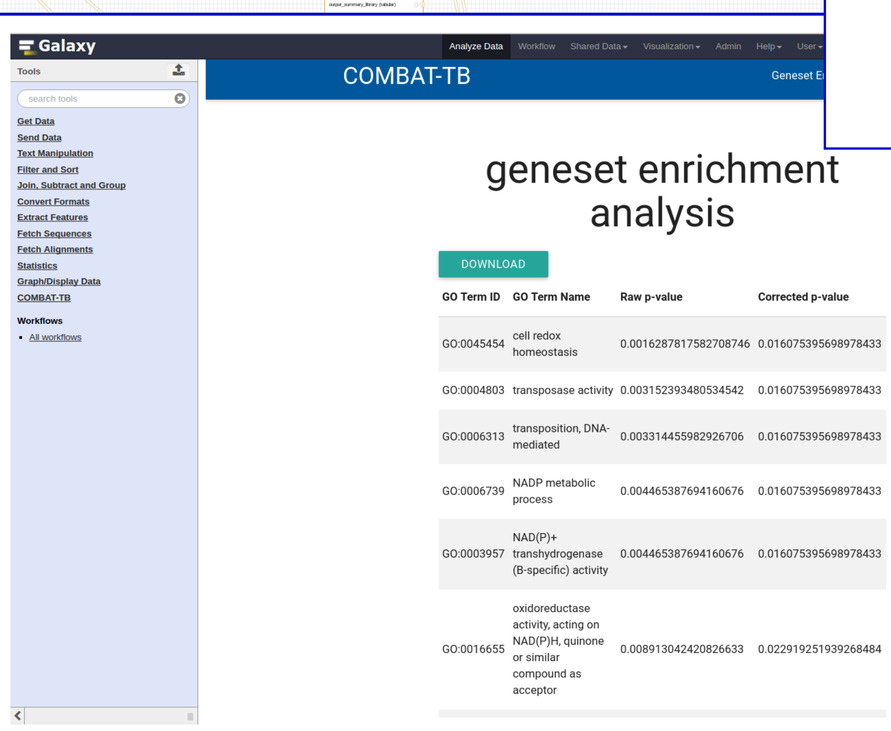
We implemented a bacterial genomic variant calling workflow in Galaxy based on previous SANBI work on analysing mutations in drug resistance strains of *M. tuberculosis* and a downstream tool that combines variant calls with a reference database of *M. tuberculosis* genome annotation and a JBrowse based genome browser to produce a new Galaxy datatype (ctb_report). To visualise the resulting datasets we implemented COMBAT TB Explorer, a Galaxy Interactive Environment. The result is that variants gleaned from novel sequence data can be explored in the context of known annotation, both from genome-centric and annotation-centric perspectives. The COMBAT TB Explorer also contains a module for geneset enrichment analysis (GSEA).

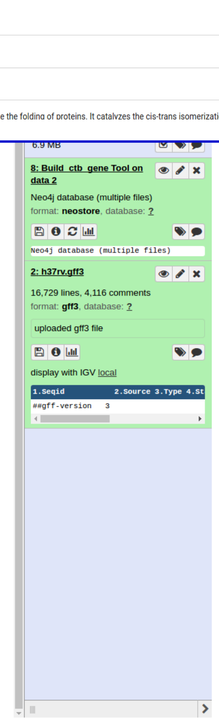
Trimming | Novo Align | Novo Sort | Mark Duplicates | Realigner Target Creator | Indel Realigner | Depth of Coverage | SnpEff











Conclusions and acknowledgements

- Integrating novel sequencing results with existing annotation will be key to researching M. tuberculosis and ultimately fighting tuberculosis.
- Our data exploration environment allows biomedical researchers to analyse, visualise and explore their own data in the context of existing annotation.
- We intend to build on the foundation we have created by extending the COMBAT TB Explorer with additional analysis modules for interrogation molecular pathway relationships, human – pathogen and drug interaction data.

The COMBAT TB project is funded by the South African Medical Research Council (MRC). We would like to acknowledge the support of Zahra Jalali (SANBI) the Galaxy Team and community.