TREAT: a bioinformatics tool for variant annotations and visualizations in targeted and exome sequencing data

Yan W. Asmann, Surmit Middha, Asif Hossain, Saurabh Baheti, Ying Li, High-Seng Chai, Zhifu Sun, Patrick H. Duffy, Ahmed A. Hadad, Asha Nair, Xiaoyu Liu, Yuli Zhang, Eric W. Klee, Krishna R. Kalari and Jean-Pierre A. Kocher

Division of Biomedical Statistics and Informatics, Department of Health Sciences Research, Mayo College of Medicine, Rochester, MN, USA

ABSTRACT

Summary: TREAT (Targeted RE-sequencing Annotation Tool) is a tool for facile navigation and mining of the variants from both targeted resequencing and whole exome sequencing. It provides a rich integration of publicly available as well as in-house developed annotations and visualizations for variants, variant-hosting genes and host-gene pathways.

Availability and implementation: TREAT is freely available to non-commercial users as either a stand-alone annotation and visualization tool, or as a comprehensive workflow integrating sequencing alignment and variant calling. The executables, instructions and the Amazon Cloud Images of TREAT can be downloaded at the website: http://ndc.mayo.edu/mayo/research/biostat/stand-alone-packages.cfm

Contact: Hossain.Asif@mayo.edu; Kocher.JeanPierre@mayo.edu

Supplementary information: Supplementary data are provided at Bioinformatics online.
Viewer (IGV) (Robinson et al., 2011) for the visualization of read alignments and variant calling information at the variant position. The functions of the variant hosting genes are illustrated via hyperlinks to the KEGG pathways and Gene Ontology terms, and the tissue expression specificity graph.

2.3 Access
TREAT is deployed in two formats, a standalone annotation application and an integrated version for an end-to-end analysis of exome or targeted sequencing data. The standalone annotation tool takes the list of called variants as input files and allows users the flexibility of generating the variants using alignment and variant calling tools of their own choosing. The integrated version accepts either FASTQ or BAM files as input files and carries out sequence alignment using BWA (Li and Durbin, 2009) or Bowtie (Langmead et al., 2009), local sequence re-alignment (GATK; McKenna et al., 2010) and variant calling (GATK or SNVMix; Goya et al., 2010), which provides users with a convenient solution to their integrative needs. In addition, both versions of making TREAT available can be downloaded for local runs, or can be launched on the Amazon Elastic Compute Cloud (EC2) using Amazon Machine Images provided at our Website. The Machine Images are loaded with all the open-source tools and necessary annotation files for the direct execution of TREAT. The run time and cost estimate of TREAT Cloud version are provided in the Supplementary Data.

3 DISCUSSIONS
We have developed a bioinformatics tool, TREAT, which addresses the current challenges in analyzing and interpreting targeted and whole exome sequencing data. The annotations provided by TREAT have been carefully evaluated and selected from a pool of available open source tools and databases, and complimented by additional in-house developed annotations (details at the TREAT website). The variant reports in Excel format integrate the visualizations of the sequence alignment at variant positions, pathways and expression specificity of the variant hosting genes via clickable hyperlinks for each reported INDELs and SNVs. In addition, the summary of the targeted resequencing results is stored in a centralized HTML report with links to the TREAT website, the targeted region coverage report and the read QC report, the description of the TREAT workflow, and links to the website of the annotation tools and databases.

For maximum flexibility, two versions of TREAT were implemented: an annotation only version, and a version integrating read alignment, variant calling and annotations. Both versions can be downloaded as local installations or as Amazon Cloud images, which makes TREAT available for users with no access to local bioinformatics infrastructures. By targeting all user groups and enabling rapid integration of emerging analytic methods, we believe that TREAT provides a sustainable NGS analytic workflow with wide applicability to the research community.

We plan to continue adding new functionality and features to TREAT to make it a comprehensive tool for targeted and exome analysis. These include the development of an in-house variant database that collects all variants detected from hundreds of individuals with various types of diseases using exome and whole genome sequencing. This database will provide critical annotations whether the observed variants are truly ‘novel’ or disease specific. In addition, we are in the process of making TREAT applicable to whole genome sequencing data analysis, this would require adding annotation tracks for non-coding regions such as the conservations and regulatory domains.

In summary, the rich set of annotations provided by TREAT, the easy to use, centralized HTML summary report, and the Excel formatted variant reports with hyperlinked visualization utilities enable the filtering of detected variants based on their functional characteristics, and allow the researchers to navigate, filter and elucidate tens of thousands of variants to focus on potential disease-associated variant(s).

ACKNOWLEDGEMENTS
We want to thank Mrs Shannon McDonnell for her constructive suggestions and feedback during the implementation of this work flow. We are very grateful for the following investigators at Mayo Clinic who allowed us to use their exome sequencing data for the purpose of developing the work flow: Dr Stephen Thibodeau, Dr Ellen Goode and Dr Fergus Couch.

Funding: Development Fund from the Center for Individualized Medicine at Mayo Clinic Rochester MN, a generous gift from James and Donna Barksdale.

Conflict of Interest: none declared.

REFERENCES