

Goliath, a NGS web-based platform.

David Berl, Thiago Luiz Araujo Miller, Daniel T. Ohara, Pedro Alexandre Favoretto Galante

USP

Abstract

The development of Next-generation sequencing platforms (NGS) in the past decade created an accurate and cost-effective methodology with application in many areas ranging from basic research to individual patient care. Nowadays, it is trouble-free, fast and inexpensive to generate NGS data. However, this ongoing revolution is placing a significant demand for expertise in processing these large sequencing datasets produced by NGS platforms. It is not rare to see researchers with NGS data on hand and stuck in the step of processing these data. Several initiatives have been produced in order to simplify the NGS data processing, but most of them have pitfalls, such as incompleteness in terms of pipelines and/or difficulty of usage. We present Goliath, a web-accessible platform for NGS processing. At its core functionality, Goliath will provide support for both transcriptomics and genomics analysis. In its launch version, Goliath processes FASTQ format data from human RNA sequencing (RNA-seq) and produces both gene expression patterns and differential expression. With a receptive interface, multiple samples and their replicates can be uploaded and combined to create an assortment of comparisons. This is achieved using the latest reference transcriptome available (e.g., GENCODE, for humans) and some of the most commonly used algorithms to quantify gene expression (e.g., Kallisto, which uses an alignment-free approach), and DESeq2 for obtaining the set of differentially expressed genes. Goliath also uses the R environment to produce clever and customizable graphs. In essence, Goliath aims to aid researchers that can produce NGS data for elucidating relevant biological questions but have a tough time processing these NGS data by their own.

Funding: CNPq