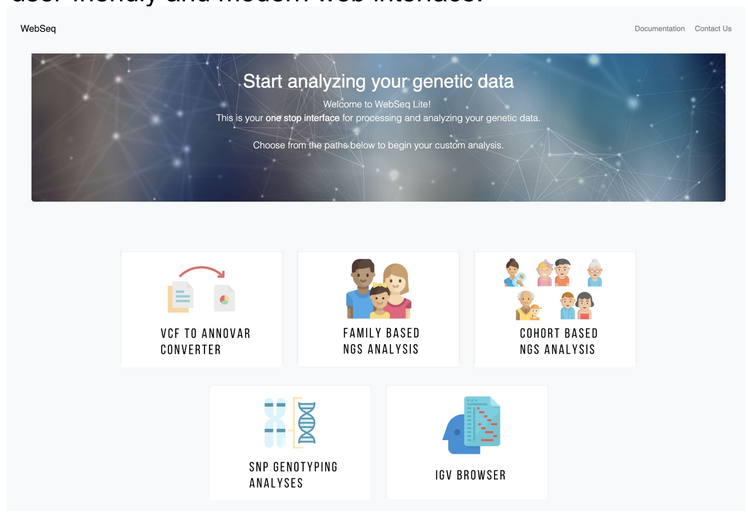
WebSeq: An Integrated Web Based Analytics Platform for Genomic Data

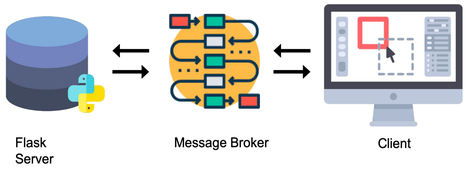
Milind Agarwal, Prof. Janet Markle
Computer Science, Molecular Microbiology and Immunology
Johns Hopkins University

Abstract

Next-generation sequencing technologies, including whole exome sequencing (WES) and whole genome sequencing (WGS), are commonly used to identify novel monogenic diseases. The application of these sequencing technologies is gaining popularity amongst clinicians and researchers as WES and WGS costs decline. However, this vast availability of WES and WGS data creates a need for a robust, flexible, scalable and easy-to-use analytics platforms to allow physicians and researchers to gain biological insight from this genomic data. We present WebSeq, a web application to facilitate intuitive genomic data analysis. WebSeq provides access to sophisticated tools and pipelines through a user-friendly and modern web interface.

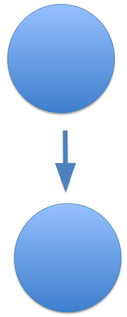


Application Architecture



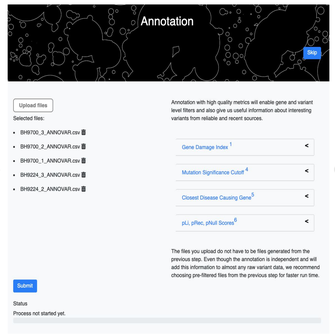
WebSeq utilizes the lightweight Flask micro framework as its backbone server. It employs microservices like Celery and Redis to monitor progress of long-running data-intensive tasks. All data processing takes places in a sandboxed environment within the user's system offering a secure way to conduct genomic data analysis. Nested services like KING[1], PLINK[2] Toolkit, and IGV[3] Reports are exposed through WebSeq's modern UI.

Results

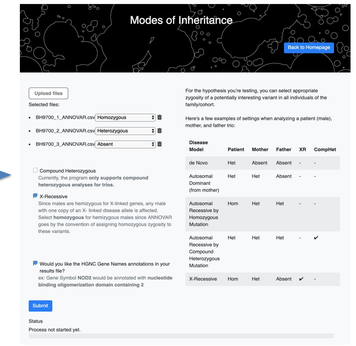


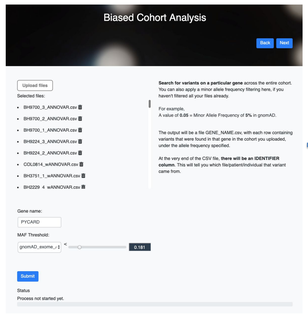
VCF Converter: WebSeq's converter abstracts away command line usage and removes file size restrictions to offer a robust, secure, in-house annotation of variant level data using ANNOVAR[4]

Cleaning: High-quality annotation enables further filtering while providing useful information about variants from reliable and recent sources. WebSeq currently annotates your data with GDI, MSC, CDG, and pLi annotations.

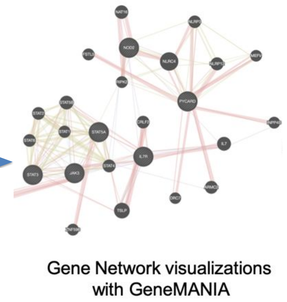


Filtering: High-quality annotation enables further filtering while providing useful information about variants from reliable and recent sources. WebSeq currently annotates your data with GDI, MSC, CDG, and pLi annotations.

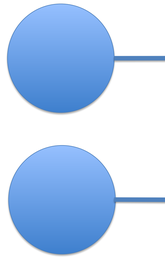




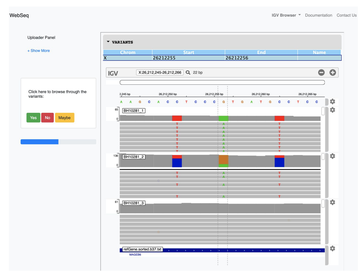
Degree of Homogeneity: annotation enables further filtering while providing useful information about variants from reliable and **recent** sources. WebSeq currently annotates your data with

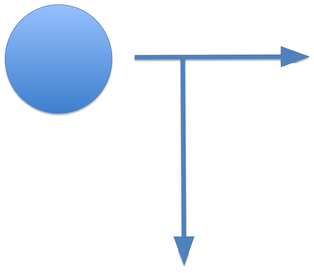


False Positive Removal: High-quality annotation enables further filtering while providing useful information about variants from reliable and recent sources. from reliable and recent sources. sources

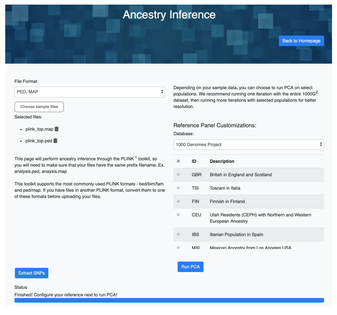


High-quality annotation enables further filtering while providing useful information about variants fromproviding useful information about variants from





Kinship and Sex Inference: High-quality annotation enables further filtering while providing useful information about variants from reliable and recent sources. WebSeq currently annotates your data with GDI, MSC, CDG, and pLi annotations.



References

[1] Manichaikul et al. Robust relationship inference in genome-wide association studies. Bioinformatics, 26(22). (2010)

[2] Purcell et al. PLINK: a tool set for whole-genome association association and population-based linkage analyses. American Journal of Human Genetics. (2007)

[3] Robinson et al. Integrated Genomics Viewer. Nature Biotechnology. (2011)

[4] Yang et al. Genomic variant annotation and prioritization using ANNOVAR and wANNOVAR. Nature Protocols. (Sep 2015)

Learn more about us!

www.webseq.org