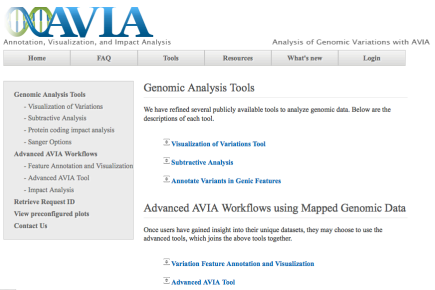
AVIA: AN INTERACTIVE WEBSERVER FOR ANNOTATION, VISUALIZATION, AND IMPACT ANALYSIS OF GENOMIC VARIATIONS

Hue Vuong, Robert Stephens, and Natalia Volfovsky

Advanced Biomedical Computing Center, SAIC-Frederick, Inc., Frederick National Laboratory for Cancer Research, Frederick, MD, USA.

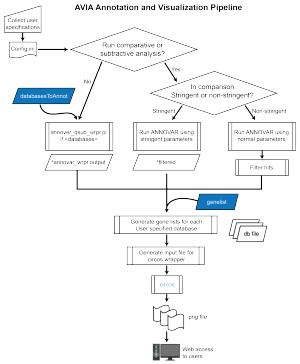


INTRODUCTION

The plethora of information that emerges from large-scale genome characterization studies has triggered the development of computational frameworks and tools for efficient analysis, interpretation, and visualization of genomic data. However, functional annotation of genomic variations and the ability to visualize the data in the context of the whole genome and/or multiple genomes have remained a challenging task. We have developed an interactive web-based tool, Annotation, Visualization, and Impact Analysis, or AVIA (avia.abcc.ncifcrf.gov), to explore and interpret large sets of genomic variations (single nucleotide variations [SNVs] and insertion/deletions) in order to help explore and summarize genomic experiments.

AVIA WORKFLOW

Our tool is based on coupling a comprehensive annotation pipeline with a flexible visualization method.

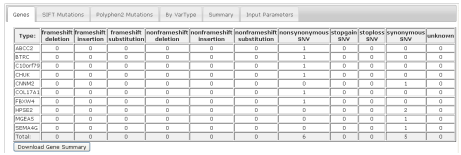


AVIA FUNCTIONALITY:

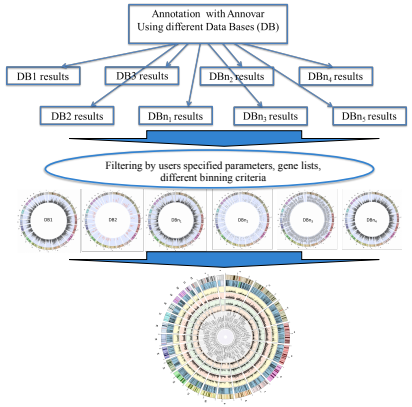
- **Impact analysis framework**: extended ANNOVAR with the capablity to add new modules with analysis tools (current tools: multiple genome analysis, tumor vs. normal analysis. MiR_SNP)
- **Annotation databases**: RefSeq, UCSC, SIFT, PolyPhen, dbSNP, 1000 Genomes, CGI publicly available genomes; option to add new databases, e.g., NonB db, PolyBrowse (ABCC annotation DBs)
- **Visualization**: Summary tables and genome-wide Circos plots with gene information
- **User-driven annotation**: Ability to upload user annotation files and combine them with a variety of pre-computed data sets, including whole-genome distributions of different genomic features (genes, exons, repeats), as well as variation analysis tracks for the 69 CGI public genomes for reference

VISUALIZATION OPTIONS

Summary Tables: Statistics of the analysis of the submitted variation by the type of variation

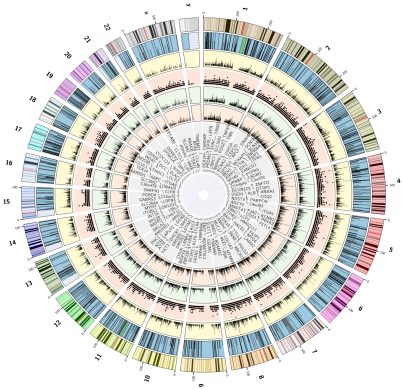


Visualization with Circos plots:



AVIA SIGNATURE Circos PLOT

- Using the ANNOVAR-AVIA workflow, we identified the functional impact of genomic variations in the liver cancer data set from RIKEN (http://dcc.icgc.org).
- Using a subtraction (filtering) option of AVIA, we removed variations found in normal populations (JPT, CGI genomes: NA18940, NA18942, NA18947, NA18956).
- Using the Circos plots option, we selected tracks corresponding to 6 annotation databases and built a whole-genome signature of the data shown below.



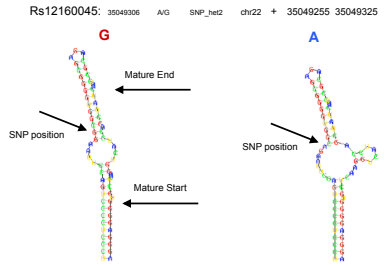
CONCLUSIONS

This version of AVIA is focused on gene-related impact assessment (coding). Tracks showing the distribution of genes with variations that have specific functional effects are shown here. Users can screen all RefSeq genes or other publicly available data sets already loaded, or provide a user-selected list of genes. Another option of the interactive analysis with AVIA is the ability to compare two input genomes, whereby users can generate views of shared or distinct variations. This allows for the comparison of tumor/normal data sets to illustrate tumor-specific mutations. During exploratory work with AVIA, users can browse different tracks with their data and then re-generate signature plots to summarize the project.

FUTURE PLANS

- Addition of new annotation databases: protein domain annotation, protein–protein interaction, drug–target interactions, comprehensive pathway information
- Extension of visualization options: cancer vs. tumor analysis, structural variation comparison, phasing and haplotype information
- Addition of new analytical modules to assess the functional impact of unknown variations, including deeper coding analysis and adding non-coding assessments
- Using haplotype and phasing information to estimate the impact of variations; modules under development: miR_SNP and NonB_SNP

Example of SNP effect on the structure of microRNAs (RNAfold):



REFERENCES

- Wang K, Le M, Hakonarson H: **ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data.** *Nucleic Acids Res* 2010, 38(16):e164
- Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA: **Circos: an information aesthetic for comparative genomics.** *Genome Res* 2009, 19(9):1639-45.

ACKNOWLEDGMENT

The authors wish to thank members of Bioinformatics Support and Scientific Web Programming Groups at ABCC and ABCC's In Silico Research Center of Excellence (ISRCE yellow task #09-260 to Frederick National Lab). This work was supported with federal funds from the National Cancer Institute, National Institutes of Health, under contract HHSN261200800001E.