

AVIA: AN INTERACTIVE WEB-SERVER FOR ANNOTATION, VISUALIZATION AND IMPACT ANALYSIS OF GENOMIC VARIATIONS

Hue Vuong¹, Anney Che¹, Sarangan Ravichandran¹, Brian T. Luke¹, Jack R. Collins¹, Uma S. Mudunuri¹

Advanced Biomedical Computing Center, Advanced Technology Program, Leidos Biomedical Research Inc, Frederick National Laboratory for Cancer Research, Frederick, MD

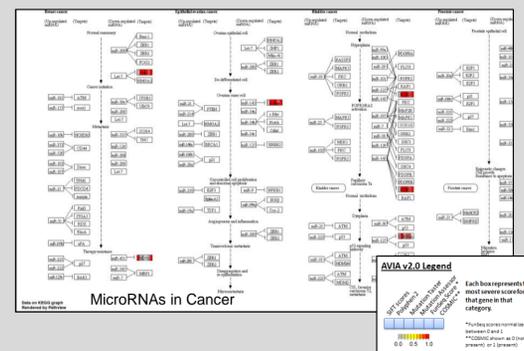


AVIA Functionality

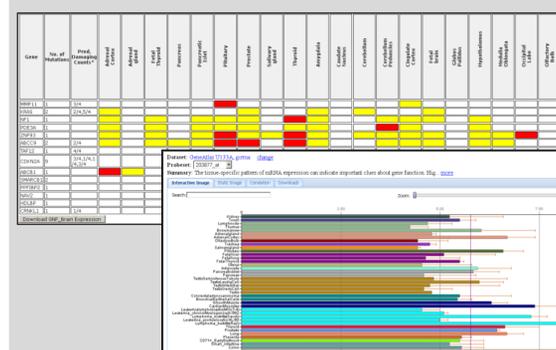
- Impact Analysis framework:** ANNOVAR (Wu *et al.* 2010) and bioDBnet (Mudunuri *et al.*, 2009), a conversion tool that allows easy retrieval of gene level annotations using any biological identifier
- Annotation Databases :** RefSeq, UCSC, SIFT, PolyPhen, dbSNP, 1000 Genomes, XOM Sequencing Project (ESP), ClinVar, COSMIC, Uniprot, NonB db, ENCODE, miRNA, miRNA targets, splicing
- User-driven annotation:** Ability to upload user-defined databases for on the fly annotation
- Customization:** Ability to customize reports based on interests.
- Clustering and Functional Annotations :** Using DAVID-API (Jiao, *et al.*, 2012), gene lists can be further interrogated by analyzing significance.
- Variety of input options :** Ability to use genomic and protein mutations, as well as gene lists, as input
- Pathview Integration :** Layering SNP information on top of KEGG Pathview maps (Luo, *et al.*, 2013) allows users to visualize SNPs in the context of pathways.
- Expression :** Ranked Reference expression by tissues and graphic visualization of all tissues.
- Gene and SNP Prioritization :** FunSeq (Khurana *et al.*, 2013) variant prioritization, as well as AVIA's variant summary. Gene prioritization by categories ranging from Disease causing to PTM.
- 3D protein structures :** Visualization of variant location in 3D protein structures modeled in-house using I-Tasser (Roy *et al.*, 2010).
- Circos Plots:** Visualization of variants in Circos

AVIA Expanded Visualization Tabs

KEGG Pathways with Gene Summary overlays



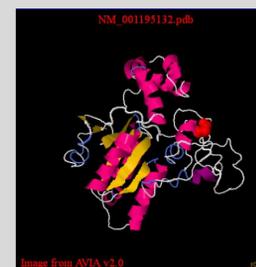
Reference Expression Tab



DAVID Clustering Analysis Tab

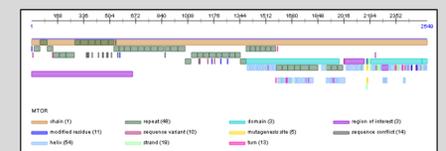
Gene	Count	%	Function	Category	Count	%	Function	Category
ABCC1	1	100.0	ATPase activity	Transporter	1	100.0	ATPase activity	Transporter
ABCC2	1	100.0	ATPase activity	Transporter	1	100.0	ATPase activity	Transporter
ABCC3	1	100.0	ATPase activity	Transporter	1	100.0	ATPase activity	Transporter
ABCC4	1	100.0	ATPase activity	Transporter	1	100.0	ATPase activity	Transporter
ABCC5	1	100.0	ATPase activity	Transporter	1	100.0	ATPase activity	Transporter
ABCC6	1	100.0	ATPase activity	Transporter	1	100.0	ATPase activity	Transporter
ABCC7	1	100.0	ATPase activity	Transporter	1	100.0	ATPase activity	Transporter
ABCC8	1	100.0	ATPase activity	Transporter	1	100.0	ATPase activity	Transporter
ABCC9	1	100.0	ATPase activity	Transporter	1	100.0	ATPase activity	Transporter
ABCC10	1	100.0	ATPase activity	Transporter	1	100.0	ATPase activity	Transporter

jsMol Visualization from in-house full length protein models



Future Work

- Our work includes tools that are developed in-house to support the scientists and help retrieve essential information to understand biology and aid in discovery of factors differentiating normal and disease states.
- REST API calls for programmatic annotation of variants, proteins, or genes
- Integration of comparative tools for comparing 2 different VCF/BAM files
- Continued work on modeling structures without any available PDB structure
- Release of an AVIA-lite package for easy installation and integration of custom annotations from other labs in the NCI community
- Include detailed position based protein annotations from UniProt and 2D visualization using BioJS



- Integrate pathway and disease association results from mining Medline and Pubmed Central using gene/protein names



References

Jiao, X., *et al.* (2012) DAVID-WS: a stateful web service to facilitate gene/protein list analysis, *Bioinformatics (Oxford, England)*, **28**, 1805-1806.

Khurana, E., *et al.* (2013) Integrative annotation of variants from 1092 humans: application to cancer genomics, *Science (New York, N.Y.)*, **342**, 1235-1238.

Luo, W. and Brouwer, C. (2013) Pathview: an R/Bioconductor package for pathway-based data integration and visualization, *Bioinformatics (Oxford, England)*, **29**, 1830-1831.

Mudunuri, U., *et al.* (2009) bioDBnet: the biological database network, *Bioinformatics (Oxford, England)*, **25**, 555-556.

Roy, A., Kucukural, A. and Zhang, Y. (2010) I-TASSER: a unified platform for automated protein structure and function prediction, *Nature protocols*, **5**, 725-738.

Vuong, H., Stephens, R.M. and Volfvsky, N. (2014) AVIA: an interactive web-server for annotation, visualization and impact analysis of genomic variations, *Bioinformatics (Oxford, England)*, **30**, 1013-1014.

Wang, K., Li, M. and Hakonarson, H. (2010) ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data, *Nucleic Acids Res*, **38**, e164.

AVIA Results

Figure 2- AVIA Results pages showing Visualization options



Figure 2. Overview of AVIA Results page. This tab shows the Visualization options available from AVIA.

AVIA Prioritization Tab

Gene	ABCC1	ABCC2	ABCC3	ABCC4	ABCC5	ABCC6	ABCC7	ABCC8	ABCC9	ABCC10
ABCC1	1	1	1	1	1	1	1	1	1	1
ABCC2	1	1	1	1	1	1	1	1	1	1
ABCC3	1	1	1	1	1	1	1	1	1	1
ABCC4	1	1	1	1	1	1	1	1	1	1
ABCC5	1	1	1	1	1	1	1	1	1	1
ABCC6	1	1	1	1	1	1	1	1	1	1
ABCC7	1	1	1	1	1	1	1	1	1	1
ABCC8	1	1	1	1	1	1	1	1	1	1
ABCC9	1	1	1	1	1	1	1	1	1	1
ABCC10	1	1	1	1	1	1	1	1	1	1

AVIA Variant Report Tab

Summary	Variant ID	ANNOVAR annot	Annot Feat	Gene	SIFT predictions and scores	PolyPhen 2 predictions and scores (Human Var)
C	737150018.07160518.A.C	exonic	non synonymous SNV/ABCC1_NM_009627	ABCC1	TOLERATED:0.2002793	Benign:0.005
C	15-22499228-22499228.T.C	UTR5	NA_ML009409204.NA_112559.n.1020	ZNF729	TOLERATED:0.0002703	Benign:0.002
C	18-28919118-28919118.T.C	exonic	non synonymous SNV/ZNF729-NM_081242	ZNF729	TOLERATED:0.7002403	Benign:0.001
C	7-21999158-21999158.T.C	exonic	non synonymous SNV/CARD11_NM_021441	CARD11	TOLERATED:0.4602393	Benign:0.003
C	11-27114008-27114008.T.C	exonic	non synonymous SNV/ABCC11_NM_020396	ABCC11	TOLERATED:0.4611303	Benign:0.047
C	22-2421568-2421568.T.C	exonic	non synonymous SNV/NMPL1-NM_005940	NMPL1	TOLERATED:0.4212303	Probably Damaging:0.6
C	9-21974606-21974606.T.C	exonic	non synonymous SNV/CANXA2L-NM_011919	CANXA2L	TOLERATED:0.3712433	Benign:0.421
C	30-20018196-20018196.T.C	exonic	non synonymous SNV/CRNK11L-NM_016665	CRNK11L	TOLERATED:0.2522893	Benign:0.009
C	11-20380008-20380008.T.C	exonic	non synonymous SNV/IT3-IT3-NC_004119.6	IT3	TOLERATED:0.1922303	Benign:0.001

Introduction

As sequencing becomes cheaper and more widely available, there is a greater need to quickly and effectively analyze large-scale genomics data. Although transitory, understanding the importance of epigenetic modifications in the landscape of cancer development requires comprehension of all changes affecting the cell. Many clinical panels focus on specific known genes associated with specific diseases or pathways. Accessing all available data about a given mutation can be difficult; from comprehending each database or algorithm to obtaining the computational power to annotate, filter, and prioritize lists of variants. Here, we present AVIA, an interactive Annotation, Visualization and Impact Analysis web server that enables users to explore their datasets on genomic variant, protein variant and gene-level, integrating resources from third party databases and software, as well as user-defined databases. AVIA includes several key annotation databases from the Encyclopedia of DNA elements (ENCODE) project, PhosphoSite and TargetScan, which provide insight on epigenetic markers that are involved with DNA methylation, posttranslational modifications, and microRNAs and their targets. Users can also explore data by many of the newly added features including visualization of genes in the context in KEGG pathway maps overlaying state data, comparison of gene expression in reference tissues, and prioritization of genes. The ability to upload a variety of input types, explore their datasets by mutations, genes and proteins, and easily navigate their data strengthens AVIA's applicability as a hub for genomics, epigenetic, gene, and protein annotations. AVIA is available at <http://avia.abcc.ncicrf.gov>.

Workflow

Figure 1- Workflow for Interactive Analysis

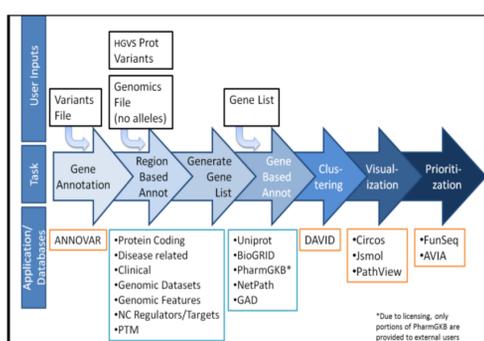


Figure 1. Workflow for AVIA annotations. Due to the modular framework, users can come in with a variety of different inputs to get annotations based on their input.

This work was supported with federal funds from the National Cancer Institute, National Institutes of Health, under contract HHSN26120080001E. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. government.



DEPARTMENT OF HEALTH AND HUMAN SERVICES • National Institutes of Health • National Cancer Institute

Frederick National Laboratory is a Federally Funded Research and Development Center operated by Leidos Biomedical Research, Inc., for the National Cancer Institute

Contract HHSN26120080001E - Funded by the National Cancer Institute