**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.ncifcrf.gov.

**Workflow**

Figure 1– Workflow for Interactive Analysis

**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, Enzyme Sequencing Project (ESP), ClinVar, COSMIC, Uniprot, NonP 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.
- **Annotation 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**: Using 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 Results**

Figure 2– AVIA Results pages showing Visualization options

**AVIA Prioritization Tab**

**AVIA Expanded Visualization Tabs**

- KEGG Pathways with Gene Summary overlays

**AVIA Variant Report Tab**

**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, Enzyme Sequencing Project (ESP), ClinVar, COSMIC, Uniprot, NonP 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: Using 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

**AVIA Variant Report Tab**

**References**


**Conclusion**

As the number of databases and tools become more abundant, it becomes increasingly difficult for users with limited bioinformatics knowledge to leverage all of the available tools and resources. Since very little is known about many genes and their role in the context of disease, our approach is to provide users with the most comprehensive, up-to-date data available.

Together, the extra functionalities detailed here help AVIA v2.0 mature as a hub for genomics, gene, and protein annotations by integrating several different types of databases and applications in a fast, comprehensive, and significant manner.