



# INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

## GenoART: Genome Annotation And Refinement Tool

**Kusumitha P S**  
**Junior Project Fellow**  
**Mangalore University**

### **Abstract:**

GenoART is a standalone application for detecting coding potential of transcripts complemented by experimentally derived proteomics (MS/MS) data. GenoART will make use of Python, PERL and JavaScript's for processing input genome and create a custom database for proteogenomic searches. In simple terms, the custom database generated by this application can be used as a reference protein database and search the unmatched or unassigned spectra in any search engines such as Mascot, SEQUEST, X! Tandem or Andromeda. GenoART can be queried to categorize and browse these resultant peptides at the click of a button.

The peptides will include the genomic coordinates which allows user to navigate to UCSC genome browser which can facilitate more details about the gene or transcript and can be compared with predicted genes, Expressed sequence Tag (EST), and perform comparative genomics for more evidence. A GTF (Gene Transfer Format) will be available to download upon querying, which can later be viewed in any genome visualizer such as UCSC or IGV to compare across different studies.

In the visualize module, we have incorporated Integrated Genome Viewer (IGV), developed in java script for aiding in visualization of the GTF files generated in GenoART and to aid in further understanding of the Genome and refinement or revising the existing gene or transcript models, Visualization part is a future work has to be done.

### **Overview:**

Genome concentrates on genetic information from each of the twenty-three pairs of chromosomes. Ridley has carefully selected the genes that drive the content of each uses them as springboards for discussing a variety of issues. Ridley describes some of the most well studied genes that caught the attention of researchers because they can cause disease when muted.

Genome annotation is the process of attaching biological information to sequences.

It consists of three main steps:

- Identifying portions of the genome that do not code for proteins
- Identifying elements on the genome, a process is called gene prediction
- Attaching biological information to these elements

**Statement of the problem:**

Genome data is so large that is difficult to view and analysis in genomic viewers or browser. The filtration process is complicated to reduce the size of the data and view it in the viewers.

**Motivation:**

Advancement in next generation sequencing technologies have enhanced several clinical applications in cancer biomarker discovery and aid in identifying pathogenic variants in inherited genetic disorders and identifying drug resistance of the emerging microbiome. however, most clinical studies rely on the existing knowledge of the gene models, and refinement of the genome is therefore essential for diagnostic relevance. We present GenoART a cross platform Graphical User Interface for the proteogenomics analysis.

**Conclusion:**

A cross-platform tool 'GenoART' for proteogenomics analysis is developed with simple web application wrapped around by the Electron framework. To our knowledge, this is the first of its kind and scientists with minimum skills in web applications and scripting languages can deploy their application for diverse users.

The aim of study was to provide a catalog of the proteins identified in the brain of *An. gambiae* mosquitoes based on experimental evidence and to additionally use the data to refine its genome annotation. We report here the identification of a set novel protein coding regions in *An. gambiae* genome obtained by high resolution tandem mass-spectrometry. With the proteogenomic approach, we were able to obtain a more comprehensive and accurate catalog of protein-coding genes.

**References:**

1. Datta KK, Patil AH, Patel K, et al. (2016). Proteogenomics of *Candida tropicalis*—An Opportunistic Pathogen with Importance for Global Health *OMICS* 20, 239-247.
2. Beausoleil SA, Villen J, Gerber SA, Rush J, and Gygi SP. (2006). A probability-based approach for high-throughput protein phosphorylation analysis and site localization. *Nat Biotechnol* 24, 1285-1292.