Acta Scientific MEDICAL SCIENCES (ISSN: 2582-0931)

Volume 10 Issue 2 February 2026

Mini Review

Integrating Bioinformatic Tools in Genomic Applications

Anusha Sunder[1]*, Kirtikaa Chezhian[2], Varshini D[2] and Vishalsai BJ[2]

[1]Doctorate in Life Science/Human Nutrition, Lead Scientist and Nutrigenetic Expert, Xcode Life Sciences, Pvt. Ltd. Chennai, India

[2]R&D Intern, Xcode Life Sciences, Pvt. Ltd. Chennai, India

***Corresponding Author:** Anusha Sunder, Doctorate in Life Science/Human Nutrition, Lead Scientist and Nutrigenetic Expert, Xcode Life Sciences, Pvt. Ltd. Chennai, India.

Received: December 22, 2025

Published: January 31, 2026

Science is an organized way of studying things and solving problems through skills. An interdisciplinary approach of integrating one field of science with another enhances the discovery of findings in a time-efficient, and validated manner. For instance, medical technology and clinical biochemistry find a role in identifying a health condition, for which the underlying cause can be traced through cell/molecular biology and genetics. While biotechnology helps to locate genetic changes through methods like polymerase chain reaction/PCR, the science of bioinformatics with its databases and computational software helps in a sequential analysis of a health-related outcomes. In this paper we have detailed on a few aspects where bioinformatics' tools have certainly enhanced genomic applications like gene panel creation.

Three practical examples of exploring bioinformatics tools have been listed below for the following genomic applications

- Comprehensive listing of genes using bioinformatics tools and databases while creating an assessment panel
- Ascertaining the two prominent alleles for a gene marker that is not biallelic
- Automating gene marker mismatches across health-based gene panels

Comprehensive listing of genes using bioinformatics tools and databases while creating an assessment panel

The traditional review research methods (manual literature search) are time- consuming and may miss relevant genes. Hence, exploring Bioinformatic tools and databases may enable a quicker search of all relevant genes and their markers/single nucleotide polymorphisms (SNPs) relevant to a health panel. Bioinformatic tools and databases such as Cytoscape, MetaCyc, and KEGG help to identify and validate genes that are specific to health traits. Such tools and databases provide a detailed network and pathway insights for better understanding of gene interactions. And their complementary use enables rapid, systematic, and reliable identification of genes associated with health and metabolic pathways. The gene list obtained from these tools and databases can subsequently be consolidated to identify gene markers/SNPs using databases such as GeneCards, GWAS Catalog, SNPedia. This bioinformatics-driven, multi- database strategy provides an efficient framework for gene panel creation. The following table (Table 1) lists the notable features of each bioinformatics tool.

Tools	Notable features
Cytoscape	A network visualization and analysis tool focuses on gene-gene, protein-protein, pathway interaction networks.
MetaCyc	A curated metabolic pathway database that contains metabolic reactions, enzymes and genes across multiple organisms.
KEGG	A comprehensive biological database that provides standardized metabolic pathway maps,gene ontologies and their disease associations.

Table 1: Notable features of each bioinformatics tool/database.

The process of exploring such tools and databases for a gene panel creation is explained with an example of vision health (vitamin A-based).

Cytoscape

Cytoscape is an open-source bioinformatics tool designed for the visualization and analysis of molecular interaction networks. It integrates gene expression with functional annotation data. We are presenting two different approaches for the Cytoscape tool. These approaches are different owing to the installation of plugins and the subsequent additional information provided. For instance, Approach 1 details the list of genes derived from the metabolic pathway, whereas Approach 2 gives additional information such as gene descriptions and its associated publications.

Approach 1: Pathway-centric network analysis

In the first approach, Cytoscape is used to identify genes associated with vision health through pathway-based network exploration. Within the Cytoscape interface, keywords such as vitamin A and vision health is given as a query under the network search bar. This search yields a set of associated biological and metabolic networks corresponding to a relevant pathway. From these results, an appropriate metabolic pathway such as vitamin A and carotenoid metabolism is selected. The genes associated with the selected pathway are extracted and compiled.

Approach 2: Plugin-enhanced network analysis

In the second approach, Cytoscape functionality is extended through the installation of the STRING plugin through Cytoscape App store. This plugin enables us with easier access to various queries across the databases. Within the network search bar, the search mode is switched to STRING Pubmed query and keywords such as vitamin A and vision health is entered resulting in a gene interaction network. Functional enrichment analysis is accessed from the right-hand panel, where a list of genes along with their functional descriptions is displayed. The enriched publications option in the same panel is then used to retrieve the corresponding literature references (PMIDs).

MetaCyc

To further refine and validate the pathway-specific gene list, the MetaCyc database can be employed as a curated resource of experimentally validated metabolic pathways. MetaCyc provides comprehensive information on metabolic reactions, enzymes, genes across multiple organisms, including humans. Within the MetaCyc interface, the “Search in Current Database” option allows targeted identification of metabolic pathways relevant to vision health and vitamin A metabolism by using keywords such as ‘retinol biosynthesis’. This approach enables direct retrieval of well-defined metabolic pathways along with the enzymes and genes involved in each reaction step.

KEGG

Alongside MetaCyc, KEGG (Kyoto Encyclopedia of Genes and Genomes) is a widely used and comprehensive database that connects genomic data with higher level biological functions especially metabolic pathways and disease mechanisms.

Within the KEGG interface, the “Search for” option enables the identification of metabolic pathways associated with vision health and vitamin A metabolism by entering keywords such as ‘retinol metabolism’. This search retrieves map00830 which represents the retinol metabolism pathway. Clicking on this pathway displays an interactive map showing the network of genes and enzymes involved in vitamin A metabolism. KEGG also offers additional

contextual information such as Gene Ontology(GO) terms, literature references (PMIDs) and standardized pathway annotations, which helps in understanding the functional roles of the identified genes.

Ascertaining the two prominent alleles for a gene marker that is not biallelic

When considering a gene marker, there is an effect allele and the other allele based on the research annotation for a specific health trait. For certain gene markers, the effect allele gets listed in the research paper, while finding the other allele becomes challenging when the marker is not biallelic. In such cases, the following approach can be taken:

- Gene markers can initially be obtained along with their reported chromosomal base pair positions from the study
- If the reported base pair position doesn't match the variant details, map their position through the ClinGen registry using GnomAD browser data from dbSNP to ascertain the other allele, with justified exclusion of rare allele.

Methodology

- Consider a gene marker which is mentioned in the scientific study that has the map position (base pair).
- The study gives the effect allele based on the OR value and significant risk contribution
- The other allele should be found in dbSNP which gives two alternative alleles and the map position (base pair) did not match the variant details in dbSNP.
- Enter the rsid in query in the ClinGen registry and select the type of search as "dbSNP ID".
- The next page displays the CA IDs for all the alleles.
- Click on the CA ID which has a gnomAD reference. The base pair position from the study was mapped in ClinGen registry.
- Click on the gnomAD reference link from the ClinGen registry. In which the other allele is given, duly mapped as the base pair position was mapped

For example

Consider the gene marker rs4792181, which relates with high myopia risk in a Singapore Chinese population-specific study [1]. The effect allele as given in the study is 'T'. The Other allele needs to be traced based on the process given above. The following are the steps –

- DbSNP gives two alleles against the effect allele (T>C / T>G), which one to consider?
- The Map position (base pair) for the rs4792181 from the study is 11653216.
- In ClinGen registry for rs4792181, the CA ID was CA288030351. This maps with the base pair position 11653216 and it also has GnomAD references hence the other allele was takes as G.

Automating gene marker mismatches across health-based gene panels

When considering two gene panels, where the common traits need to be aligned on gene markers or SNPs (single nucleotide polymorphisms), the following approach is efficient to resolve mismatches. It also serves as a final step in validation, where the program is re-run to check whether both panels still contain unique entries.

- Use the Directional Approach to locate the unique SNPs (single nucleotide polymorphism from each panel and obtain both the files separately
- Once unique SNPs are identified, align panels by resolving unmatched markers
- Retain shared SNPs (a unique SNP of one panel becomes shared when it is
- entered in the other also)

The following process flow is suggested -

Identify the mismatch between the panels

Each panel might have unique as well as gene markers (SNPs)

Example:

- Panel A has the following SNPs: rs101, rs202, rs303, rs404
- Panel B has the following SNPs: rs202, rs303, rs505, rs606

Observation:

- rs101 and rs404 appear only in panel A
- rs505 and rs606 appear only in panel B

These unique SNPs cause inconsistency between panels.

What to do?

Start with a panel and proceed directionally with the other.

To precisely find the inconsistencies, the comparison must be done in two directions:

A to B comparison

List panel A first and then panel B, as we are finding SNPs that are present in panel A but not in Panel B.

B to A comparison

List panel B first and then panel A, as we are finding SNPs that are present in panel B but not in Panel A.

Generate Two Outputs.

Obtaining unique SNPs

A mathematical approach is the logic, which gets coded in the system using Python programming:

Logic-Mathematical approach

(A Δ B)

Unique SNPs = AΔB= (A−B) ∪ (B−A)

A logical approach is to convert each of the panels into a set of elements (a text file containing SNPs of panel A and panel B separately) and apply the difference:

Unique to A = A-B

Unique to B = B-A

The symmetric difference is broken down into simple set-wise differences here to obtain the unique SNPs from each panel. Instead of using it entirely in a single step, this ensures that unique SNPs from each panel are obtained separately.

Symbols and their denotation with explanation

Δ -Symmetric Difference

∪-Union.

The symmetric difference of two sets is always used to identify entries that are unique to each set by excluding what is in their intersection.

- Coding-Python program
- Using Python, a basic Script can be written to solve this problem without manual effort.

PROGRAM 1 — Find SNPs in Panel A NOT in Panel B

Logic of the program:

Input part

Input file 1 – the first panel as a text file containing SNPs

(Note: copy the SNPs separately from both files and name them as per the panel.)

Input file 2 – The second panel as a text file containing SNPs

Output file – this would contain the unique SNPs from panel A.

Generating the output file

Running the program will produce a text file with unique SNPs for the directional approach, both for A to B and B to A.

Align the panels

- With unique SNPs being obtained, use them to solve the following:
- No unique entries in any panel
- Removing or resolving unmatched markers
- Ensuring only shared, validated SNPs are kept for comparison
- Re-running the unique-check programs until both output files become empty, meaning both panels now contain the same SNP list.

Bibliography

1. Li Y J., *et al*. "Genome-wide association studies reveal genetic variants in CTNND2 for high myopia in Singapore Chinese". *Ophthalmology* 118.2 (2011): 368-375.