REVIEW OF SOFTWARE TOOLS FOR GENETIC MUTATION ANALYSIS

Augustine Chidibere Onuora
Department of Computer Science,
Akanu Ibiam Federal Polytechnic Unwana, (AIFPU)
Ebonyi State, Nigeria

Nonye Emmanuel Maidoh
Department of Computer Science,
Akanu Ibiam Federal Polytechnic Unwana, (AIFPU)
Ebonyi State, Nigeria

Godwin Oko Epia
Department of Computer Science,
Akanu Ibiam Federal Polytechnic Unwana, (AIFPU)
Ebonyi State, Nigeria

Prince Ana
Department of Computer Science,
Cross-River University of Technology (CRUTECH),
Cross-River State, Nigeria

Abstract— **Genes are basic physical and functional unit of heredity. Genes are made up of DNA. Some genes act as instructions to make molecules called proteins. However, many genes do not code for proteins. Gene Mutation refers to modification of gene's DNA sequence. Mutation of genes can vary within a genome and between genomes. Mutations range in size; they can affect anywhere from a single DNA building block (base pair) to a large segment of a chromosome that includes multiple genes. Because mutation is one of the fundamental forces of evolution, this work looked into the various computer software tools that can be used for genetic mutation analysis. These applications were reviewed stating the OS platform it depends on, the nature and format of data it accept, the input and output from these software tools and urls were they can be found were also stated.**

Keywords— **gene, genome, mutation, software, tool, analysis, DNA, chromosome**

I. INTRODUCTION

Mutations are changes in the genetic sequence, and they are a main cause of diversity among organisms. These changes occur at many different levels, and they can have widely differing consequences. In biological systems that are capable of reproduction, we must first focus on whether they are heritable; specifically, some mutations affect only the individual that carries them, while others affect all of the carrier organism's offspring, and further descendants. For mutations to affect an organism's descendants, they must:

1) occur in cells that produce the next generation, and
2) affect the hereditary material. Ultimately, the interplay between inherited mutations and environmental pressures generates diversity among species [1].

Although various types of molecular changes exist, the word "mutation" typically refers to a change that affects the nucleic acids. In cellular organisms, these nucleic acids are the building blocks of DNA, and in viruses they are the building blocks of either DNA or RNA. One way to think of DNA and RNA is that they are substances that carry the long-term memory of the information required for an organism's reproduction. Recall that the DNA sequence found within a gene controls protein synthesis. If the DNA sequence is altered, this can alter the amino acid sequence within a protein [2].

If mutations occur in non-germline cells, then these changes can be categorized as somatic mutations. The word somatic comes from the Greek word soma which means "body", and somatic mutations only affect the present organism's body. From an evolutionary perspective, somatic mutations are uninteresting, unless they occur systematically and change some fundamental property of an individual--such as the capacity for survival. For example, cancer is a potent somatic mutation that will affect a single organism's survival. As a different focus, evolutionary theory is mostly interested in DNA changes in the cells that produce the next generation [2].

II. CLASSIFICATION OF GENE MUTATION

According to [3], Gene mutations can be classified in two major ways:

A. **Heriditary Mutation**

Hereditary mutations are inherited from a parent and are present throughout a person's life in virtually every cell in the body. These mutations are also called germline mutations because they are present in the parent's egg or sperm cells, which are also called germ cells. When an egg and a sperm cell unite, the resulting fertilized egg cell receives DNA from both parents. If this DNA has a mutation, the child that grows from the fertilized egg will have the mutation in each of his or her cells.

B. Acquired Mutation

Acquired (or somatic) mutations occur at some time during a person's life and are present only in certain cells, not in every cell in the body. These changes can be caused by environmental factors such as ultraviolet radiation from the sun, or can occur if a mistake is made as DNA copies itself during cell division. Acquired mutations in somatic cells (cells other than sperm and egg cells) cannot be passed on to the next generation.
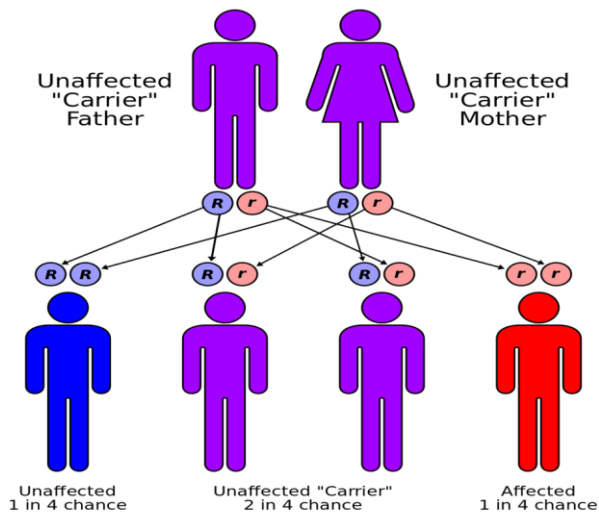

Figure 1: The red individual has inherited two mutated alleles of a gene from their parents. This is an example of a hereditary mutation [3].

Genetic changes that are described as de novo (new) mutations can be either hereditary or somatic. In some cases, the mutation occurs in a person's egg or sperm cell but is not present in any of the person's other cells. In other cases, the mutation occurs in the fertilized egg shortly after the egg and sperm cells unite. It is often impossible to tell exactly when a de novo mutation happened. As the fertilized egg divides, each resulting cell in the growing embryo will have the mutation. De novo mutations may explain genetic disorders in which an affected child has a mutation in every cell in the body but the parents do not, and there is no family history of the disorder [4].

Somatic mutations that happen in a single cell early in embryonic development can lead to a situation called mosaicism. These genetic changes are not present in a parent's egg or sperm cells, or in the fertilized egg, but happen a bit later when the embryo includes several cells. As all the cells divide during growth and development, cells that arise from the cell with the altered gene will have the mutation, while other cells will not. Depending on the mutation and how many cells are affected, mosaicism may or may not cause health problems [4].

Most disease-causing gene mutations are uncommon in the general population. However, other genetic changes occur more frequently. Genetic alterations that occur in more than one percent of the population are called polymorphisms. They are common enough to be considered a normal variation in the DNA. Polymorphisms are responsible for many of the normal differences between people such as eye color, hair color, and blood type. Although many polymorphisms have no negative effects on a person's health, some of these variations may influence the risk of developing certain disorders [5].

III. TYPES OF MUTATIONS

The DNA sequence of a gene can be altered in a number of ways. Gene mutations have varying effects on health, depending on where they occur and whether they alter the function of essential proteins [6]. The types of mutations include:

1. Missense mutation: This type of mutation is a change in one DNA base pair that results in the substitution of one amino acid for another in the protein made by a gene.

2. Nonsense mutation: A nonsense mutation is also a change in one DNA base pair. Instead of substituting one amino acid for another, however, the altered DNA sequence prematurely signals the cell to stop building a protein. This type of mutation results in a shortened protein that may function improperly or not at all.

3. Insertion: An insertion changes the number of DNA bases in a gene by adding a piece of DNA. As a result, the protein made by the gene may not function properly.

4. Deletion: A deletion changes the number of DNA bases by removing a piece of DNA. Small deletions may remove one or a few base pairs within a gene, while larger deletions can remove an entire gene or several neighbouring genes. The deleted DNA may alter the function of the resulting protein(s).

5. Duplication: A duplication consists of a piece of DNA that is abnormally copied one or more times. This type of mutation may alter the function of the resulting protein.

6. Frameshift mutation: This type of mutation occurs when the addition or loss of DNA bases changes a gene's reading frame. A reading frame consists of groups of 3 bases that each code for one amino acid. A frameshift mutation shifts the grouping of these bases and changes the code for amino acids. The resulting protein is usually nonfunctional. Insertions, deletions, and duplications can all be frameshift mutations.

7. Repeat expansion: Nucleotide repeats are short DNA sequences that are repeated a number of times in a row. For example, a trinucleotide repeat is made up of 3-base-pair sequences, and a tetranucleotide repeat is made up of 4-base-pair sequences. A repeat expansion is a mutation

that increases the number of times that the short DNA sequence is repeated. This type of mutation can cause the resulting protein to function improperly.

IV. EFFECTS OF MUTATIONS

Reference [7] elaborated on the effects of mutation. A single mutation can have a large effect, but in many cases, evolutionary change is based on the accumulation of many mutations with small effects. Mutational effects can be beneficial, harmful, or neutral, depending on their context or location. Most non-neutral mutations are deleterious. In general, the more base pairs that are affected by a mutation, the larger the effect of the mutation, and the larger the mutation's probability of being deleterious.

To better understand the impact of mutations, researchers have started to estimate distributions of mutational effects (DMEs) that quantify how many mutations occur with what effect on a given property of a biological system. In evolutionary studies, the property of interest is fitness, but in molecular systems biology, other emerging properties might also be of interest. It is extraordinarily difficult to obtain reliable information about DMEs, because the corresponding effects span many orders of magnitude, from lethal to neutral to advantageous; in addition, many confounding factors usually complicate these analyses. To make things even more difficult, many mutations also interact with each other to alter their effects; this phenomenon is referred to as epistasis. However, despite all these uncertainties, recent work has repeatedly indicated that the overwhelming majority of mutations have very small effects [8].

V. SOFTWARE TOOLS

Technological advancement has indeed contributed its quota in bioinformatics. Analysis of gene's constituent for mutation can be done using software tools. The accuracy of these tools differs from one to the other. Here are some these tools considered in the review:

1. Biomedical Mutation Analysis (BMA)
2. MutaNET
3. VariantMaster
4. Mutation Surveyor
5. Mutation taster
6. Polyphen2
7. SIFT
8. KGGSeq
9. Galaxy
10. GaTK
11. AlaMut
12. Imutant
13. SNPEffect
14. HOPE
15. PredictSNP
16. Meta-SNp
17. PON-P
18. CONDEL
19. iSTABLE
20. PROVEAN
21. Sibyl
22. PhD-SNP
23. PMut
24. HANSA
25. SNPs&GO
26. MutPred

VI. REVIEW OF SOME SOFTWARE TOOLS FOR ANALYSIS OF GENE MUTATION

There are lot of software available in this field and each of them can be used depending on what the user need. Some of these softwares listed above will be reviewed based on data available at OmicX. OmicX have developed a unique dynamic analytics ecosystem that capitalizes on collective scientific intelligence to generate fast, accurate and actionable meta-insights, unleashing the full value of big biodata and allows researchers to achieve new scientific breakthroughs [9] [10].

1. **Biomedical Mutation Analysis (BMA):** was designed for an accessible analysis of mutations online. BMA is a user-friendly application by which the user compare aligned sequences with a reference sequence. The output given will be the place where the changes are and to what it changed. Assists in computing changes in nucleotide and amino acid sequences. BMA is an online application that stores all information related to the mutation analyses. This tool performs thanks to an analysis algorithm able to evaluate multiple patients, where each one can include multiple sequences. Moreover, the algorithm can analyze desired positions that the analyst can define.

Website identifier	http://bma.itiud.org/
Name	BMA
Alternative name	Biomedical Mutation Analysis

Interface	Web user interface
Restrictions to use	None
Output format	PDF
Database management system	MySQL
License	GNU General Public License version 3.0
Computer skills	Basic
Stability	Stable
Maintained	Yes

system	
Programming languages	Python
License	GNU General Public License version 3.0
Computer skills	Medium
Version	1.0.0
Stability	Stable
Maintained	Yes

2. **MutaNET:** Mutations in genomic key elements can influence gene expression and function in various ways, and hence greatly contribute to the phenotype. MutaNET performs statistical analyses of mutations in different genomic regions. The tool also incorporates the mutations in a provided gene regulatory network to estimate their global impact. The integration of a next-generation sequencing pipeline enables calling mutations prior to the analyses. An application example analysed the impact of mutations in antibiotic resistance (AR) genes and their potential effect on AR of bacterial strains.

Website identifier	https://sourceforge.net/projects/mutanet/
Name	MutaNET
Software type	Application/Script
Interface	Graphical user interface
Restrictions to use	None
Input data	Gene annotations, protein domains, transcription factors and their binding sites, or antibiotic resistance for multiple bacterial strains.
Input format	FASTA
Output data	Images and tables.
Operating	Unix/Linux, Mac OS, Windows

3. **VariantMaster:** Extracts causative variants in familial and sporadic genetic diseases. VariantMaster implements a methodology to evaluate the status (presence or absence) of a variant in familial or case-control contexts. The software allows users to identify causative variants in familial, sporadic germline, and somatic genetic disorders, including cancers. It also allows for the search of causative variants in one or more recurrently mutated genes in a pool of unrelated individuals sharing the same phenotype.

Website identifier	https://sourceforge.net/projects/variantmaster/
Name	VariantMaster
Software type	Package/Module
Interface	Command line interface
Restrictions to use	None
Input format	BAM, VCF
Operating system	Unix/Linux
Programming languages	Python
Computer skills	Advanced

Version	1.02
Stability	Stable
Maintained	Yes

4. **PROVEAN:** (Protein Variation Effect Analyzer) is a software tool which predicts whether an amino acid substitution or indel has an impact on the biological function of a protein. PROVEAN is useful for filtering sequence variants to identify nonsynonymous or indel variants that are predicted to be functionally important. The performance of PROVEAN is comparable to popular tools such as SIFT or PolyPhen-2.

Website identifier	http://provean.jcvi.org/index.php
Name	PROVEAN
Alternative name	Protein Variation Effect Analyzer
Software type	Package/Module
Interface	Command line interface
Restrictions to use	None
Operating system	Unix/Linux
Programming languages	C++
License	GNU General Public License version 2.0
Computer skills	Advanced
Stability	Stable
Maintained	Yes

5. **Mutation Surveyor:** is a powerful and accurate DNA Sequencing analysis tool for Sanger Sequencing files generated by Applied Biosystems Genetic Analyzers, MegaBACE, and Beckman CEQ electrophoresis systems. Capable of performing variant analysis of up to 2000 Sanger sequencing files (ab1, RSD, .ESD & .scf) in 15 minutes, Mutation Surveyor delivers excellent accuracy, sensitivity, and low false positive and false negative rates in the analysis of DNA variants, including: single

nucleotide polymorphisms (SNPs), insertions and deletions (INDELS), and somatic mutations in direct sequencing, PCR sequencing, Mitochondrial DNA sequencing, and resequencing projects.

Website identifier	https://softgenetics.com/mutationSurveyor.php
Name	Mutation Surveyor
Software type	Package/Module
Interface	Command line interface
Restrictions to use	License purchase required
Operating system	Windows
License	Commercial
Computer skills	Advanced
Version	5.0
Stability	Stable
Free trial	Yes
Maintained	Yes

6. **Mutation Taster:** Assists researchers to perform evaluation of the pathogenic potential of DNA sequence alterations. MutationTaster is an online application that aims to determine the functional consequences of amino acid substitutions, short insertion and/or deletion (indel) mutations, variants spanning intron-exon borders, intronic and synonymous alterations. Moreover, this tool is able to categorize confirmed polymorphisms and known disease mutations.

Website identifier	http://www.mutationtaster.org/
Name	MutationTaster
Alternative name	MutationTaster2

Interface	Web user interface
Restrictions to use	None
Input data	Some coding sequence (ORF), gene (genomic sequence), and transcript (cDNA sequence).
Computer skills	Basic
Version	2.0
Stability	Stable
Maintained	Yes
Wikipedia	https://en.wikipedia.org/wiki/MutationTaster

7. **SIFT:** SIFT predicts whether an amino acid substitution affects protein function based on sequence homology and the physical properties of amino acids. SIFT can be applied to naturally occurring nonsynonymous polymorphisms and laboratory-induced missense mutations.

Website identifier	https://sift.bii.a-star.edu.sg/
Name	SIFT
Alternative names	Sorting Intolerant From Tolerant, SIFT for genomes, SIFT 4G
Software type	Application/Script
Interface	Command line interface
Restrictions to use	None
Input data	Some sequence conservation and amino acid properties.
Input format	VCF,FASTA,GTF,GFF
Output format	VCF,XLS

Operating system	Unix/Linux
License	GNU General Public License version 3.0
Computer skills	Advanced
Version	2.0.0
Stability	Stable
Requirements	g++, GNU make, nvcc
Maintained	Yes

8. **Imutant:** A support vector machine (SVM)-based tool for the automatic prediction of protein stability changes upon single point mutations. I-Mutant predictions are performed starting either from the protein structure or, more importantly, from the protein sequence.

Website identifier	http://gpcr2.biocomp.unibo.it/cgi/predictors/I-Mutant3.0/I-Mutant3.0.cgi
Name	I-Mutant
Interface	Web user interface
Restrictions to use	None
Computer skills	Basic
Version	3.0
Stability	Stable
Maintained	Yes

9. **HOPE:** Serves for automatic mutant analysis. HOPE is an online tool that furnishes information about disease related phenotype caused by mutations in human proteins. For performing, this tool collects data from sources such as the protein's 3D structure and the UniProt database of well-annotated protein sequences. This program works in three steps: users (1) enter an input sequence; (2) select a residue to mutate; and (3) select mutation.

Website identifier	http://www.cmbi.ru.nl/hope/input/
Name	HOPE
Alternative name	Have yOur Protein Explained
Interface	Web user interface
Restrictions to use	None
Input data	A protein sequence.
Programming languages	Javascript
Database management system	PostgreSQL
Computer skills	Basic
Version	1.1.1
Stability	Stable
Maintained	Yes

Input data	PredictSNP1: protein sequence; PredictSNP2: list of variants
Input format	PredictSNP1: FASTA; PredictSNP2: VCF, GVF, HGVS
Computer skills	Basic
Version	2.0
Stability	Stable
Maintained	Yes

10. **PredictSNP:** Consensus classifiers for prediction of disease-related mutations. PredictSNP1 offers its users a consensus score based on the output of six different amino acid-based predictors. Because of the nature of the tools whose results are combined to generate its consensus, PredictSNP1 can only be used to analyze substitutions in an amino acid sequence. PredictSNP2 complements PredictSNP1 by evaluating the effects of nucleotide variants located in any region of the genome. PredictSNP2 represents the first unified platform for nucleotide-based predictions of deleterious variants.

Website identifier	https://loschmidt.chemi.muni.cz/predictsnp/
Name	PredictSNP
Interface	Web user interface
Restrictions to use	Academic or non-commercial use

VII. CONCLUSION

Bioinformatics researchers especially those who want to run mutation prediction analysis cannot do without the part this technological software tools is playing currently. This software review will enable researcher know which tool to use and how they can get them because every software tool reviewed has a website URL where you can run your analysis online is they are web-based or downloaded to your computing device if a desktop software.

Another software tool worth mentioning even though it is not tied to biological computation like the one mentioned above is the MATHLAB software. This software has relevance in almost every statistical and computational related field [11]. Researchers can also explore its capabilities in genetic analysis.

VIII. REFERENCE

[1] Loewe, L., Genetic Mutation | Learn Science at Scitable. 2008,https://www.nature.com/scitable/ topicpage/genetic-mutation-1127?error=cookies_not_supported&code= f4b6c104-de21-4785-987b-84d03ed7b087

[2] Brennan, J., How Can a Mutation in DNA Affect Protein Synthesis? 2018, https://sciencing.com/can-mutation-dna-affect-protein-synthesis-2028.html

[3] Khan S and Vihinen M., Performance of protein stability predictors. - PubMed - NCBI. 2019 https://www.ncbi.nlm.nih.gov/pubmed/20232415

[4] Genetics Home Reference, What is a gene mutation and how do mutations occur? 2019, https://ghr.nlm.nih.gov/primer/mutationsanddisorders/gen emutation

[5] Bartee, L., How Gene Mutations Occur – Mt Hood Community College Biology 102. 2016, https://openoregon.pressbooks.pub/mhccbiology102/chapt

er/7-1-what-is-a-gene-mutation-and-how-do-mutations-occur/

[6] Genetics Home Reference, What kinds of gene mutations are possible? May 9, 2019, https://ghr.nlm.nih.gov/primer/mutationsanddisorders/possiblemutations

[7] Brennan, J., How Can a Mutation in DNA Affect Protein Synthesis? 2018, https://sciencing.com/can-mutation-dna-affect-protein-synthesis-2028.html

[8] Eyre-Walker, A., Keightley, P., The distribution of fitness effects of new mutations. Nat Rev Genet 8, 2007, 610–618. https://doi.org/10.1038/nrg2146

[9] omicX. (2019). Bioinformatics tools - omicX. May 9, 2019, from https://omictools.com/

[10] SoftGenetics. (2019). SoftGenetics - Software PowerTools for Genetic Analysis. May 9, 2019, https://softgenetics.com/

[11] Potapov , V., Cohen , M., & Schreiber, G., Assessing computational methods for predicting protein stability upon mutation: good on average but not in the details. 2009, https://academic.oup.com/peds/article/22/9/553/1515777

[12] Christophe Béroud, Thierry Soussi, "p53 gene mutation: Software and database", Nucleic Acids Research, Volume 26, Issue 1, 1 January 1998, Pages 200–204, https://doi.org/10.1093/nar/26.1.200

[13] Stenson, P.D., Mort, M., Ball, E.V. et al. The Human Gene Mutation Database: towards a comprehensive repository of inherited mutation data for medical research, genetic diagnosis and next-generation sequencing studies. Hum Genet 136, 665–677 (2017). https://doi.org/10.1007/s00439-017-1779-6

[14] Stenson PD, Mort M, Ball EV, Shaw K, Phillips A, Cooper DN (2014) The Human Gene Mutation Database: building a comprehensive mutation repository for clinical and molecular genetics, diagnostic testing and personalized genomic medicine. Hum Genet 133:1–9

[15] Schwarz JM, Cooper DN, Schuelke M, Seelow D (2014) MutationTaster2: mutation prediction for the deep-sequencing age. Nat Methods 11:361–362