IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 4, May 2025
Impact Factor: 7.67

# Functional Annotation of Cancer Variants: Tools And Techniques

**Alka Manisha [1] Jyoti Prakash[2], Akanksha Pandey[2], Rachna Chaturvedi[1*]**
Amity Institute of Biotechnology, Amity University, Lucknow, UP, India[1]
Queen Mary Hospital, KGMU Lucknow, UP, India[2]
[*]Corresponding Author: rchaturvedi1@lko.amity.edu
Orchid ID https://orcid.org/0000-0003-2875-4384

**Abstract**: *Cancer is a complex genetic disease driven by somatic and germline mutations that alter critical cellular functions. With the advancement of Next-Generation Sequencing (NGS), researchers can now identify these mutations at high throughput; however, interpreting their biological and clinical relevance requires robust functional annotation. This study investigates the mutational landscape of prostate, breast, and pancreatic cancers using public RNA-seq datasets and an open-source bioinformatics pipeline. Key oncogenic drivers and tumor suppressor genes, including BRCA1, BRCA2, TP53, and KRAS, were analyzed using tools such as FastQC, Bowtie2, Samtools, FreeBayes, and SnpEff. Our findings highlight the predominance of missense mutations and frequent transition substitutions such as C→T and G→A, with silent and nonsense mutations also contributing to disease mechanisms. These insights emphasize the utility of integrative computational tools in variant annotation and their potential to enhance cancer diagnostics, prognostics, and targeted therapy selection in precision oncology.*

**Keywords**: Functional annotation, BRCA1, BRCA2, Prostate cancer, Breast cancer, Pancreatic cancer, NGS, Precision oncology

## I. INTRODUCTION

Cancer is a genetically complex disease driven by both somatic and germline mutations, which disrupt normal cellular mechanisms, leading to tumorigenesis. Advances in Next-Generation Sequencing (NGS) have transformed cancer research by enabling rapid identification of genetic variants associated with various cancers, such as breast, prostate, and pancreatic. These cancers commonly involve mutations in genes like BRCA1, BRCA2, TP53, KRAS, and PIK3CA, which play critical roles in DNA repair, cell cycle regulation, and signaling pathways.

In breast cancer, BRCA1/2 mutations increase hereditary risk and influence treatment decisions, such as the use of PARP inhibitors. Prostate cancer often features mutations in BRCA2, ATM, and TP53, which are linked to aggressive disease. In pancreatic cancer, frequent mutations in KRAS, CDKN2A, and PALB2 contribute to its high lethality. Functional annotation of these variants is vital for interpreting their impact, guiding personalized therapies, and enhancing clinical outcomes in precision oncology.

## II. MATERIALS AND METHODS

This study used RNA-based NGS data from the European Nucleotide Archive (ENA) to identify and annotate genetic variants in breast, prostate, and pancreatic cancers. The workflow integrates open-source bioinformatics tools available through the Galaxy platform to ensure reproducibility and accessibility.

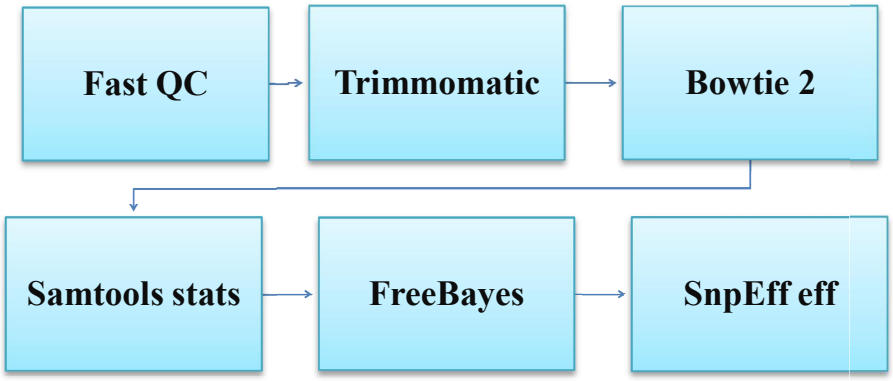
**Table 2: Provided information on the data functional annotation is performed on from European Nucleotide Archive Database.**

| S.No. | Project | Sample | Run Accession | No. of Datasets | Cancer type |
|---|---|---|---|---|---|
| 01 | PRJNA874314 | SAMN30541240 | SRR21275094 | 1 | PANCREATIC |

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 4, May 2025
Impact Factor: 7.67

| | | | | | |
|---|---|---|---|---|---|
| 02 | PRJNA1078399 | SAMN40004198 | SRR28020223 | 1 | PROSTATE |
| 03 | PRJNA1078399 | SAMN40004204 | SRR28020229 | 1 | PROSTATE |
| 04 | PRJNA451206 | SAMN08964539 | SRR7050672 | 2 | BREAST |

(Archive, 2025)




**FASTQ Files**:

Raw sequencing data is stored in FASTQ format, containing nucleotide sequences and quality scores. Since sequencing errors can lead to false-positive variant calls, quality control is critical. Evaluations include GC content, sequencing biases, and base-call errors (Bolger, 2014).

**FastQC**:

This tool evaluates per-base and per-sequence quality scores, GC content, overrepresented sequences, and potential contaminants. It highlights sequencing artifacts that may impact downstream analysis (Andrews, 2010; Zhou, 2023).

**Trimmomatic:**

Used to clean sequencing data by trimming adapters and low-quality bases. It ensures reliable alignment and variant calling by applying sliding window trimming, leading/trailing trimming, and length filtering (Ritchie, 2016; Tuteja, 2022; Bolger, 2014).

**Bowtie2:**

Efficiently aligns reads to a reference genome with support for gapped alignment. It produces SAM/BAM files for further analysis, and its accuracy is crucial in cancer variant detection (Langmead, 2012; Zhou, 2023).

**Samtools:**

Handles SAM/BAM files, enabling filtering, indexing, and conversion. It also computes mapping statistics and depth coverage, facilitating structural variant discovery and variant calling when paired with samtools (Li, 2009; Tuteja, 2022).

**FreeBayes:**

A haplotype-based variant caller that identifies SNPs, indels, MNPs, and complex variants. It integrates into Galaxy workflows and outputs results in VCF format for downstream annotation (Lee, 2015).

**SnpEff:**

Annotates variants by predicting their functional impact, including synonymous, nonsynonymous, and frameshift mutations. It integrates cancer-specific databases to identify pathogenic variants and assess drug response relevance (Cingolani, 2012).

**Dataset overview**

Samples included datasets from prostate (PRJNA1078399), pancreatic (PRJNA874314), and breast cancer (PRJNA451206) projects. The combination of computational pipelines and publicly available data enables efficient detection of clinically relevant mutations across cancer types.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 4, May 2025
Impact Factor: 7.67

## III. RESULTS

**Table :-3 Number of variants by type, Number of effects by functional class of dataset SAMN40004204 of project id PRJNA1078399**

| Type | Total | Type | Count | Percent |
|---|---|---|---|---|
| **SNP** | 102,741 | **MISSENSE** | 1,907 | 54.486% |
| **MNP** | 2,753 | **NONSENSE** | 68 | 1.943% |
| **INS** | 946 | **SILENT** | 1,525 | 43.571% |
| **DEL** | 1,002 | | | |
| **MIXED** | 209 | | | |
| **Total** | **107,651** | | | |

**Table 4: Frequency of Base Substitutions Due to SNPs of dataset SAMN40004204 of project id PRJNA1078399**

| | A | C | G | T |
|---|---|---|---|---|
| A | 0 | 5,196 | 14,360 | 5,167 |
| C | 5,761 | 0 | 6,179 | 14,849 |
| G | 14,535 | 6,186 | 0 | 5,683 |
| T | 5,007 | 14,618 | 5,200 | 0 |

**Table:-5 Number of variants by type and Number of effects by functional class of dataset SAMN08964539 of project id PRJNA1078399**

| Type | Total | Type | Count | Percent |
|---|---|---|---|---|
| **SNP** | 137,646 | MISSENSE | 2,399 | 52.958% |
| **MNP** | 3,069 | NONSENSE | 68 | 1.501% |
| **INS** | 1,565 | SILENT | 2,063 | 45.541% |
| **DEL** | 1,657 | | | |
| **MIXED** | 210 | | | |
| **Total** | 144,147 | | | |

**Table 6: Frequency of Base Substitutions Due to SNPs of dataset SAMN08964539 of project id PRJNA1078399**

| | **A** | **C** | **G** | **T** |
|---|---|---|---|---|
| **A** | 0 | 6,488 | 20,829 | 5,608 |
| **C** | 6,876 | 0 | 8,045 | 20,782 |
| **G** | 20,951 | 7,874 | 0 | 6,667 |
| **T** | 5,835 | 21,224 | 6,467 | 0 |

The dataset provides insight into genetic alterations in prostate cancer, focusing on missense, nonsense, and silent mutations, which help identify functional changes linked to disease progression. Tables 3 and 5 summarize the types of variants and their functional effects.

In Table 3 (dataset SAMN40004204, project PRJNA1078399), a total of 107,651 variants were identified. SNPs were the most common (95.43%), followed by MNPs (2.56%), insertions (0.88%), deletions (0.93%), and mixed variants (0.19%). Functionally, missense mutations were the most frequent (54.49%), followed by silent mutations (43.57%) and nonsense mutations (1.94%). Table 5 (dataset SAMN08964539, project PRJNA1078399) shows similar trends across 144,147 variants, with SNPs again dominating (95.48%), followed by MNPs (2.13%), insertions (1.09%), deletions (1.15%), and mixed variants (0.15%). The functional distribution included 52.96% missense, 45.54% silent, and 1.50% nonsense mutations.


DOI: 10.48175/IJARSCT-26429

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 4, May 2025
Impact Factor: 7.67

These results reflect common genomic patterns, with SNPs being the most frequent, and most functional coding changes being missense or silent.

Tables 4 and 6 represent the SNP substitution chart, showing the frequency of base substitutions between the four nucleotides (A, C, G, and T). The rows indicate the original (reference) base, while the columns represent the substituted (alternate) base. The diagonal values are zero, indicating no substitution. In Table 3 (dataset SAMN40004204, project PRJNA1078399), the most frequent substitutions occur between C → T (14,849) and G → A (14,535), followed by T → C (14,618) and A → G (14,360). These are typical transition mutations (purine ↔ purine or pyrimidine ↔ pyrimidine), which are more common than transversions (purine ↔ pyrimidine). Table 6 of dataset SAMN08964539 of project id PRJNA1078399 reveals that the most frequent substitutions occur between G → A (20,951) and C → T (21,224), followed by A → G (20,829) and T → C (21,224). These substitutions are primarily transition mutations (purine ↔ purine or pyrimidine ↔ pyrimidine), which are typically more common than transversions (purine ↔ pyrimidine). The chart highlights common mutational patterns and potential hotspots, which could be influenced by biological processes such as deamination in methylated CpG regions.

**Table 7: Number variants by type and Number of effects by functional class of dataset SAMN30541240 for project ID PRJNA874314**

| Type | Total | Type | Count | Percent |
|---|---|---|---|---|
| **SNP** | 347,691 | **MISSENSE** | 111,490 | 46.972% |
| **MNP** | 68,118 | **NONSENSE** | 4,260 | 1.795% |
| **INS** | 24,934 | **SILENT** | 121,603 | 51.233% |
| **DEL** | 15,588 | | | |
| **MIXED** | 76,781 | | | |
| **Total** | **533,112** | | | |

**Table:- 8 Frequency of Base Substitutions Due to SNPs of dataset SAMN30541240 for project ID PRJNA874314**

| | A | C | G | T |
|---|---|---|---|---|
| **A** | 0 | 28,199 | 49,955 | 19,750 |
| **C** | 23,763 | 0 | 16,769 | 35,295 |
| **G** | 35,843 | 18,264 | 0 | 23,822 |
| **T** | 19,740 | 48,902 | 27,389 | 0 |

Table 7 provides a summary of the genetic variants identified in a dataset, classified by variant type and their functional effects. A total of 533,112 variants were observed. The majority are Single Nucleotide Polymorphisms (SNPs), with 347,691 variants (65.23%), followed by Multi-Nucleotide Polymorphisms (MNPs) at 68,118 (12.78%), insertions (INS) at 24,934 (4.68%), deletions (DEL) at 15,588 (2.92%), and mixed variants at 76,781 (14.41%). In terms of functional classification, silent mutations, which do not alter the protein sequence, were the most common, accounting for 121,603 variants (51.23%). Missense mutations, which change amino acid sequences and can affect protein function, made up 111,490 variants (46.97%). Nonsense mutations, which create premature stop codons and can disrupt protein production, were relatively rare, with 4,260 variants (1.80%). (Reva, 2011)

Table 8 presents a SNP substitution profile, illustrating the frequency of base changes among the four DNA nucleotides: A, C, G, and T. Each row represents the original (reference) base, while each column shows the substituted (alternate) base. The diagonal entries are zero, indicating no mutation. The most frequent substitutions observed are A → G (49,955), T → C (48,902), C → T (35,295), and G → A (35,843). These are predominantly transition mutations, involving purine-to-purine or pyrimidine-to-pyrimidine changes, which are biologically more common than transversions. The distribution of these substitutions reflects typical mutational patterns in genomic data and may be influenced by processes such as deamination or replication errors, particularly in methylated regions of the genome.

These findings provide valuable insights into the variant landscape of pancreatic cancer, contributing to a deeper understanding of its genomic alterations.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 4, May 2025
Impact Factor: 7.67

**Table 9: Number variants by type and Number of effects by functional class of dataset 1 SAMN08964539 for project id PRJNA451206**

| Type | Total | Type | Count | Percent |
|---|---|---|---|---|
| **SNP** | 329,222 | **MISSENSE** | 30,361 | 25.133% |
| **MNP** | 181,460 | **NONSENSE** | 419 | 0.347% |
| **INS** | 19,186 | **SILENT** | 90,019 | 74.52% |
| **DEL** | 10,822 | | | |
| **MIXED** | 48,427 | | | |
| **Total** | **589,117** | | | |

**Table:- 10 Frequency of Base Substitutions Due to SNPs of dataset 1 SAMN08964539 for project id PRJNA451206**

| | A | C | G | T |
|---|---|---|---|---|
| **A** | 0 | 15,630 | 63,775 | 15,133 |
| **C** | 15,305 | 0 | 13,001 | 41,244 |
| **G** | 42,692 | 14,357 | 0 | 16,875 |
| **T** | 14,272 | 62,058 | 14,880 | 0 |

| Type | Total | Type | Count | Percent |
|---|---|---|---|---|
| **SNP** | 336,448 | **MISSENSE** | 30,367 | 24.491% |
| **MNP** | 54,663 | **NONSENSE** | 306 | 0.247% |
| **INS** | 21,398 | **SILENT** | 93,318 | 75.262% |
| **DEL** | 10,125 | | | |
| **MIXED** | 54,470 | | | |
| **Total** | **477,104** | | | |

**Table:-11 Number variants by type and Number of effects by functional class of dataset 2 SAMN08964539 for project id PRJNA451206**

**Table:-12 Frequency of Base Substitutions Due to SNPs of dataset 2 SAMN08964539 for project id PRJNA451206**

| | A | C | G | T |
|---|---|---|---|---|
| **A** | 0 | 18,636 | 62,825 | 16,210 |
| **C** | 15,055 | 0 | 13,988 | 41,317 |
| **G** | 42,751 | 16,250 | 0 | 16,826 |
| **T** | 14,867 | 60,724 | 16,999 | 0 |

Tables 9 and 11 present the distribution of genetic variants by type and functional impact in two breast cancer datasets (Dataset 1 and Dataset 2) from the project PRJNA451206. In Dataset 1 (SAMN08964539), a total of 589,117 variants were detected. The majority were Single Nucleotide Polymorphisms (SNPs) (329,222; 55.89%), followed by Multi-Nucleotide Polymorphisms (MNPs) (181,460; 30.79%), insertions (19,186; 3.26%), deletions (10,822; 1.84%), and mixed variants (48,427; 8.22%). Functionally, silent mutations accounted for the largest proportion (90,019; 74.52%), indicating changes that do not affect the amino acid sequence. Missense mutations, which may alter protein function, were also significant (30,361; 25.13%), while nonsense mutations (419; 0.35%) were relatively rare.

In comparison, Dataset 2 contains 477,104 variants, including SNPs (336,448; 70.5%), MNPs (54,663; 11.46%), insertions (21,398; 4.48%), deletions (10,125; 2.12%), and mixed variants (54,470; 11.42%). The functional breakdown shows a similar trend: silent mutations are most common (93,318; 75.26%), followed by missense (30,367; 24.49%) and nonsense mutations (306; 0.25%). The analysis across all datasets highlights key differences in the mutational landscape of prostate, breast, and pancreatic cancers, emphasizing the varying roles of missense, silent, and nonsense mutations in each cancer type.


DOI: 10.48175/IJARSCT-26429

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 4, May 2025
Impact Factor: 7.67

Tables 10 and 12 show the frequency of base substitutions due to SNPs in two datasets (Dataset 1 and Dataset 2) from the breast cancer project PRJNA451206. Both tables present the original (reference) nucleotide in the rows and the substituted (alternate) nucleotide in the columns. In Dataset 1 (Table 10), the most frequent substitutions are A → G (63,775), T → C (62,058), G → A (42,692), and C → T (41,244). In Dataset 2 (Table 12), the most common substitutions are A → G (62,825), T → C (60,724), G → A (42,751), and C → T (41,317).

Both datasets show a similar pattern, with transition mutations (purine ↔ purine or pyrimidine ↔ pyrimidine) being the most common.

Missense mutations are the most prevalent across all three cancers, underscoring their impact on protein function. These mutations, which alter amino acid sequences, can disrupt crucial signaling pathways and contribute to tumor progression. In **prostate cancer**, missense mutations frequently affect androgen receptor signaling, a key driver of disease progression. In **breast cancer**, they often impact genes involved in hormone regulation and DNA repair, while in **pancreatic cancer**, they frequently occur in oncogenes such as **KRAS**, a major driver of tumorigenesis. (Miguel A. Molina-Vila1, 2020)

Silent mutations, though not altering protein sequences, play an important regulatory role in gene expression, splicing, and mRNA stability. Their frequency across all three cancer types suggests they may contribute to tumor development through mechanisms beyond direct protein alteration. In **breast and pancreatic cancers**, silent mutations may influence alternative splicing of key tumor suppressor genes, while in **prostate cancer**, they could impact regulatory elements involved in androgen signaling.

Nonsense mutations, though less frequent, have a significant impact by introducing premature stop codons, leading to truncated, nonfunctional proteins. Their effects vary across cancer types: in **prostate cancer**, they may disrupt tumor suppressor genes like **TP53** and **RB1**; in **breast cancer**, they often affect genes such as **BRCA1/2**, critical for DNA repair; and in **pancreatic cancer**, they are frequently found in tumor suppressors like **CDKN2A** and **SMAD4**, contributing to aggressive tumor behavior. (Cingolani, 2012)

In addition to point mutations, the prevalence of structural changes such as SNPs, MNPs, insertions (INS), and deletions (DEL) differs amongst tumors. SNPs are the most common variant type in all three malignancies, demonstrating that single-base substitutions are a significant source of genetic variation. MNPs, while being less prevalent, introduce numerous nucleotide alterations that may have a greater influence on protein function than SNPs do. Insertions and deletions that add or delete nucleotide sequences can trigger frameshift mutations, resulting in loss-of-function effects in important tumor suppressor genes. The increased occurrence of insertions and deletions in pancreatic cancer reflects a greater degree of genomic instability than in prostate and breast cancer.

These findings underscore the diverse and complex genetic landscapes of prostate, breast, and pancreatic cancers, emphasizing the significance of both small-scale mutations, such as single-nucleotide polymorphisms (SNPs) and multiple-nucleotide polymorphisms (MNPs), as well as larger structural variations, including insertions (INS) and deletions (DEL). These mutations contribute to tumor heterogeneity, influencing cancer progression, treatment response, and PROGNOSIS. (Kumar, 2018)

## V. CONCLUSION

This study underscores the critical role of functional annotation in comprehensively understanding the mutational landscape of prostate, breast, and pancreatic cancers. By utilizing Next-Generation Sequencing (NGS) and bioinformatics tools, we systematically identified and classified key genetic variants, emphasizing the significance of single-nucleotide polymorphisms (SNPs) and their functional consequences. Among these, missense mutations emerged as the most prevalent, playing a crucial role in altering protein structure and function, thereby driving tumor progression. While silent mutations were traditionally considered neutral, our findings suggest their potential regulatory impact on gene expression and splicing. Additionally, although nonsense mutations occur less frequently, their ability to generate truncated proteins often leads to the disruption of essential cellular processes, contributing to tumor development. (Cingolani, 2012)

The variation in mutation prevalence across different cancer types reflects their distinct genetic landscapes. In prostate cancer, mutations in genes such as AR, TP53, and BRCA2 influence tumor growth and therapeutic response. Breast

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 4, May 2025
Impact Factor: 7.67

cancer is primarily driven by mutations in BRCA1/2, TP53, and PIK3CA, affecting DNA repair and hormone signaling pathways. Pancreatic cancer is characterized by a high frequency of KRAS, TP53, and CDKN2A mutations, contributing to its aggressive nature. Notably, missense mutations in oncogenes and tumor suppressors within these cancers can lead to protein dysfunction, loss of regulation, and increased tumorigenicity. The absence of large-scale structural variations (such as inversions, duplications, and breakends) in the analyzed datasets suggests that small-scale mutations, including SNPs and indels, play a more prominent role in these cancers. (Mary-Claire King, 2003)

These findings highlight the necessity of integrating functional annotation with variant analysis to distinguish pathogenic mutations from benign polymorphisms, ultimately improving our understanding of cancer genetics. SNPs, particularly those resulting in missense mutations, are of high clinical significance, as they can alter protein function and influence treatment response. Future advancements in computational tools, multi-omics approaches, and machine learning models will be crucial in refining mutation interpretation, improving biomarker discovery, and advancing precision oncology. Additionally, expanding functional annotation frameworks to include non-coding mutations and conducting experimental validation will provide deeper insights into their roles in cancer pathogenesis. By bridging the gap between genomic data and clinical application, these efforts will contribute to more effective targeted therapies and personalized treatment strategies for cancer patients.

## REFERENCES

[1]. Andrews, S. (2010). FastQC: A quality control tool for high-throughput sequence data. Babraham Bioinformatics.
[2]. Archive, E. N. (2025). cancer. European Nucleotide Archive.
[3]. Bolger, A. M. (2014). . Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics, 2114–2120.
[4]. Cingolani, P. P. (2012). A program for annotating and predicting the effects of single-nucleotide polymorphisms.
[5]. Galaxy. (2025). Galaxy Project.
[6]. Halim-Fikri H, S.-H. S.-J. (2023). Central resources of variant discovery and annotation and their role in precision medicine. Asian Biomed, 285-298.
[7]. Kumar, S. e. (2018). Systematic Functional Annotation of Somatic Mutations in Cancer. Cancer cells, 291.
[8]. Langmead, B. &. (2012). Fast gapped-read alignment with Bowtie 2. Nature Methods, 357–359.
[9]. Lee, L. A. (2015). Annotation of Sequence Variants in Cancer Samples: Processes and Pitfalls for Routine Assays in the Clinical Laboratory. The Journal of Molecular Diagnostics, 339-351.
[10]. Li, H. H. (2009). 1he Sequence Alignment/Map format and SAMtools. Bioinformatics, 2078–2079.
[11]. Mary-Claire King, J. H. (2003). Breast and Ovarian Cancer Risks Due to Inherited Mutations in BRCA1 and BRCA2. Science, 643-646.
[12]. Miguel A. Molina-Vila1, C. M.-d.-l.-C.-I.-A.-P.-B. (2020). Annotating the next-generation sequencing report. Precision Cancer Medicine, 936-939.
[13]. Ng, S. B. (2009). Targeted capture and massively parallel sequencing of 12 human exomes. nature, 461(7261), 272-276.
[14]. Reva, B. A. (2011). Predicting the functional impact of protein mutations: Application to cancer genomics. Nucleic Acids Research, 118.
[15]. Ritchie, G. E. (2016). In Silico Functional Annotation of Genomic Variation. Current Protocols in Bioinformatics, 375-379.
[16]. Schaafsma, G. C. (s.d.). ariSNP, a benchmark database for SNP effect predictions. . Human Mutation, 805-811.
[17]. Tuteja, S. (2022). A performance evaluation study: Variant annotation tools - the enigma of clinical next generation sequencing (NGS) based genetic testing. The Journal of Pathology Informatics, 100130.
[18] Zhou, J. Z. (2023). Automated bioinformatics. autoba, 2309.03242