Variant caliing

Variant type

Explanation of Variant Types and Their Impact

- **Conservative Inframe Deletion:**

Count: 4
Frequency: 0.004%
Description: Deletion of a number of nucleotides that is a multiple of three, preserving the reading frame and resulting in the removal of one or more amino acids without drastic changes to protein structure or function.

- **Conservative Inframe Insertion:**

Count: 10
Frequency: 0.009%
Description: Insertion of a number of nucleotides that is a multiple of three, preserving the reading frame and resulting in the addition of one or more amino acids, generally with minimal impact on protein function.

- **Disruptive Inframe Deletion:**

Count: 4
Frequency: 0.004%
Description: Deletion within the coding sequence that does not alter the reading frame but may significantly affect protein function if critical regions are affected.

- **Disruptive Inframe Insertion:**

Count: 5
Frequency: 0.005%
Description: Insertion within the coding sequence that does not alter the reading frame but may significantly affect protein function if critical regions are affected.

- **Downstream Gene Variant:**

Count: 48,907
Frequency: 44.568%
Description: Variants located downstream of the coding region of a gene, typically having no direct impact on the gene's function.

- **Frameshift Variant:**

Count: 143
Frequency: 0.13%
Description: Insertions or deletions that alter the reading frame, potentially resulting in a completely altered protein sequence downstream of the mutation and often leading to loss of function.

- **Intergenic Region**:

Count: 880
Frequency: 0.802%
Description: Variants located between genes, generally having no direct impact on gene function unless regulatory elements are affected.

- **Intragenic Variant:**

Count: 33
Frequency: 0.03%
Description: Variants within a gene that do not affect the exons, possibly impacting introns, UTRs, or regulatory regions.

- **Missense Variant:**

Count: 7,235
Frequency: 6.593%
Description: Single nucleotide changes resulting in a different amino acid, potentially affecting protein function depending on the importance of the changed amino acid.

- **Splice Region Variant:**

Count: 36
Frequency: 0.033%
Description: Variants near exon-intron boundaries affecting splicing, potentially resulting in incorrect mRNA and altered protein products.

- **Start Lost:**

Count: 8
Frequency: 0.007%
Description: Variants that change the start codon, possibly resulting in no protein being produced or a truncated protein.

- **Stop Gained**:

Count: 43
Frequency: 0.039%
Description: Variants introducing a premature stop codon, leading to truncated, likely nonfunctional proteins.

- **Stop Lost:**

Count: 33
Frequency: 0.03%
Description: Variants that remove a stop codon, resulting in a longer protein that may be dysfunctional.

- **Stop Retained Variant:**

Count: 3
Frequency: 0.003%
Description: Variants that change a codon without altering the encoded stop signal, typically having no effect on protein function.

- **Synonymous Variant:**

Count: 2,465
Frequency: 2.246%
Description: Nucleotide changes that do not alter the amino acid sequence of the protein, generally considered neutral.

- **Upstream Gene Variant:**

Count: Not provided
Frequency: Not provided
Description: Variants located upstream of the coding region, potentially affecting gene regulation and

expression.

Summary Table of Provided Data

Variant Type	Count	Frequency
Conservative Inframe Deletion	4	0.004%
Conservative Inframe Insertion	10	0.009%
Disruptive Inframe Deletion	4	0.004%
Disruptive Inframe Insertion	5	0.005%
Downstream Gene Variant	48,907	44.568%
Frameshift Variant	143	0.13%
Intergenic Region	880	0.802%
Intragenic Variant	33	0.03%
Missense Variant	7,235	6.593%
Splice Region Variant	36	0.033%
Start Lost	8	0.007%
Stop Gained	43	0.039%
Stop Lost	33	0.03%
Stop Retained Variant	3	0.003%
Synonymous Variant	2,465	2.246%
Upstream Gene Variant	Not provided	Not provided

High Frequency of Downstream Gene Variants: These are the most common type, suggesting many variants are located outside coding regions, likely having minimal impact on gene function.
Significant Number of Missense Variants: These can potentially alter protein function and may be important for further functional studies.
Presence of Frameshift, Start Lost, Stop Gained, and Stop Lost Variants: These types of variants can have significant impacts on protein function, potentially leading to loss of function or gain of new, often deleterious, functions.
This summary helps in understanding the potential impact of detected variants on gene and protein function, guiding further biological and functional validation studies.

Let's visualize and describe the types of genetic variants with examples, including a diagram to illustrate the concepts.**

Example with Visualization
We'll use a simplified example of a gene sequence to demonstrate different types of genetic variants.

```
Gene Sequence:
Original DNA Sequence:

####5' - ATG GCC TGG CCG TAA - 3'
#Translated Protein Sequence:
#### Met - Ala - Trp - Pro - Stop
```

Here's a visual representation of the gene and its variants:

Descriptions of Variants:

Synonymous (Silent) Mutation:

Original Codon: GCC (Ala)
Mutated Codon: GCG (Ala)
Effect: No change in the amino acid sequence.

Missense Mutation:

Original Codon: GCC (Ala)
Mutated Codon: GTC (Val)
Effect: Alanine is replaced by Valine in the protein.
Nonsense Mutation:

Original Codon: CCG (Pro)
Mutated Codon: TAG (Stop)
Effect: Premature termination of the protein (Truncated protein).

Frameshift Mutation:

Original Sequence: TGG CCG (Trp - Pro)
Insertion: TGG A CCG (Trp - Thr - Pro)
Effect: Reading frame is shifted, resulting in different downstream amino acids.

In-frame Insertion:

Original Sequence: TGG CCG (Trp - Pro)
Insertion: TGG CAG CCG (Trp - Gln - Pro)
Effect: Insertion of one amino acid (Glutamine) without shifting the reading frame.

In-frame Deletion:

Original Sequence: TGG CCG (Trp - Pro)
Deletion: TGG G (Trp)
Effect: Deletion of one amino acid (Proline) without shifting the reading frame.

Splice Site Variant:

Original Sequence: Exon-Intron boundary
Mutation: Affecting the GT or AG at the boundary
Effect: Altered splicing, potentially leading to exon skipping or inclusion of intronic sequences.

Upstream Gene Variant:

Location: Before the start codon
Effect: Potentially affects transcriptional regulation and gene expression.

Downstream Gene Variant:

Location: After the stop codon
Effect: Typically minimal impact, but can affect regulatory elements.
Detailed Example for Missense Mutation:

Let's take a detailed look at a missense mutation example.

Original DNA and Protein Sequence:
@@@: 5' - ATG GCC TGG CCG TAA - 3'
@@@: Met - Ala - Trp - Pro - Stop
Missense Mutation Example:
Original Codon: GCC (Ala)
Mutated Codon: GTC (Val)

Original Sequence:

5' - ATG GCC TGG CCG TAA - 3'
Met - Ala - Trp - Pro - Stop
Mutated Sequence:
5' - ATG GTC TGG CCG TAA - 3'
Met - Val - Trp - Pro - Stop

Let's take a detailed look at a missense mutation example

Example of a Nonsense Variant:
Gene: BRCA1 (Breast Cancer 1)

Normal Sequence:

```
DNA: 5' - ATG GCT GAG TGA CTT CAG - 3'
mRNA: 5' - AUG GCU GAG UGA CUU CAG - 3'
Amino Acids: Met - Ala - Glu - Thr - Ser - Gln
```

Nonsense Mutation:

```
DNA: 5' - ATG GCT TAA TGA CTT CAG - 3'
mRNA: 5' - AUG GCU UAA UGA CUU CAG - 3'
```

```
Amino Acids: Met - Ala - Stop
```

Conclusion:

- Synonymous (Silent) Mutation: No change in protein sequence.
- Missense Mutation: One amino acid is replaced by another.
- Nonsense Mutation: Protein is prematurely terminated.
- Frameshift Mutation: Reading frame is shifted, altering downstream sequence.
- In-frame Insertion/Deletion: Addition or removal of amino acids without shifting the reading frame.
- Splice Site Variant: Affects mRNA splicing.
- Upstream/Downstream Variants: Can affect gene regulation.
- These examples and the visual representation help in understanding the different types of genetic variants and their potential impacts on gene and protein function.

A classification of genetic variants based on their predicted impact on gene function

Classification of Genetic Variants by Impact

- **HIGH Impact Variants:**

Count: 227
Frequency: 0.207%
Description: Variants that are very likely to have a severe impact on the gene or gene product. These include:
Nonsense mutations: Introduce a premature stop codon.
Frameshift mutations: Disrupt the reading frame of the gene.
Splice site mutations: Affect the splicing of the mRNA.
Start/stop codon mutations: Affect the initiation or termination of translation.

- **LOW Impact Variants:**

Count: 2,468
Frequency: 2.25%
Description: Variants that are expected to have a minimal or no effect on the gene or gene product. These include:
Synonymous mutations: Do not alter the amino acid sequence of the protein (silent mutations).

- **MODERATE Impact Variants:**

Count: 7,258
Frequency: 6.616%
Description: Variants that are likely to have a moderate impact on the gene or gene product. These include:

Missense mutations: Result in a single amino acid change in the protein.
In-frame insertions/deletions: Insert or delete amino acids without disrupting the reading frame.

- **MODIFIER Variants:**

Count: 99,278
Frequency: 90.866%
Description: Variants predicted to have a negligible or unknown impact on the gene or gene product.
These include:
Variants in non-coding regions: Intergenic regions, intronic regions (unless they affect splicing), and regulatory regions.

```
Summary Table
Impact  Level       Count        Frequency
HIGH                 227          0.207%
LOW                  2,468        2.25%
MODERATE             7,258        6.616%
MODIFIER             99,278       90.866%
```

Example Breakdown:

Impact	Example Nucleotide Change	Example Protein Change	Explanation
HIGH	5' - ATG GCT GAG **TAA** CTT CAG - 3'	Met - Ala - Glu - Stop	Nonsense mutation leading to premature stop codon and truncated protein.
LOW	5' - ATG GCT **GAA** TGA CTT CAG - 3'	Met - Ala - Glu - Thr - Leu - Gln	Synonymous mutation not changing the amino acid sequence.
MODERATE	5' - ATG GCT **GTG** TGA CTT CAG - 3'	Met - Ala - Val - Thr - Leu - Gln	Missense mutation changing one amino acid in the protein.
MODIFIER	5' - ATG GCT GAG TGA **CTA** CTT CAG - 3'	Met - Ala - Glu - Thr - Leu - Gln	Variant in an intergenic region with minimal or unknown impact.

Conclusion

- Understanding the distribution of these variants can help prioritize further studies, such as functional assays or clinical investigations, particularly focusing on HIGH and MODERATE impact variants due to their potential to significantly alter gene function.

link full Annotation explain:

https://pcingola.github.io/SnpEff/snpeff/inputoutput/