

CLINVAR-BASED- VARIANT-IMPACT- STUDY

INTERPRETING PATHOGENICITY OF
VARIANTS IN BRCA1, CFTR, AND
GJB2 USING PUBLIC DATABASES
AND PREDICTIVE TOOLS

WHAT THIS PROJECT DOES

Goal:

Build a mini diagnostic pipeline to evaluate and classify clinically significant genetic variants using publicly available tools and databases.

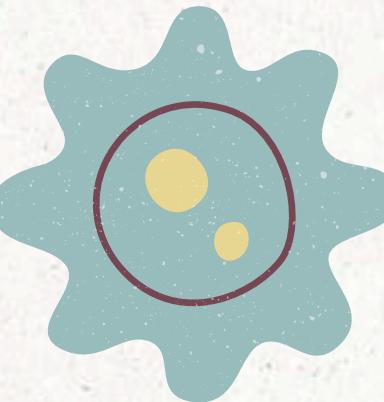
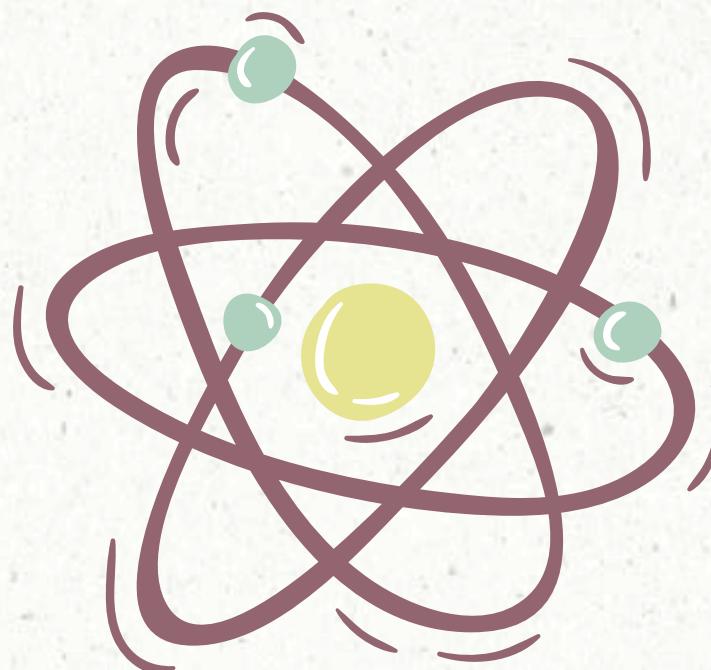
Scope:

- Selected Genes: BRCA1, CFTR, GJB2
- Source: ClinVar, dbSNP, Ensembl
- Tools Used: VEP, SIFT, PolyPhen-2, MutationTaster, CADD
- Output: Variant classification + interpretation



WHAT THIS PROJECT DOES

- Simulates a **clinical diagnostic workflow** for pathogenic variant interpretation
- Focused on **3 high-impact genes** (BRCA1, CFTR, GJB2)
- Uses ClinVar, dbSNP, Ensembl VEP, SIFT, PolyPhen-2, MutationTaster, and CADD
- Interprets **20 real variants** using a unified pipeline
- Produces annotated tables, variant summaries, and structured analysis

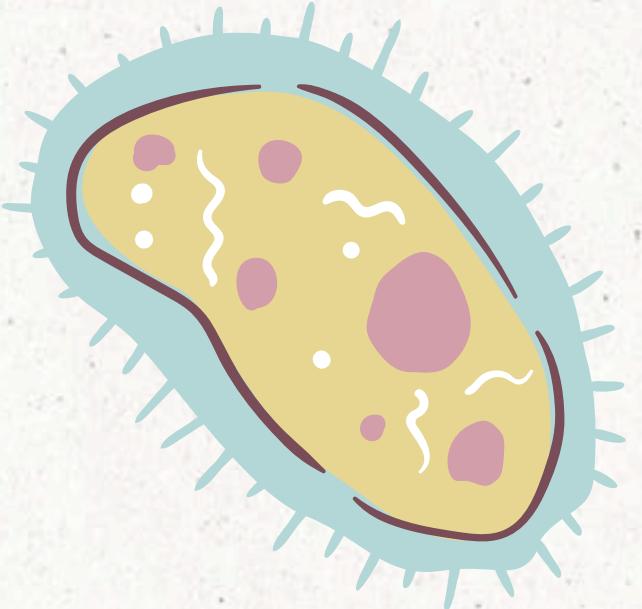


GENE SELECTION RATIONALE

WHY BRCA1, CFTR, AND GJB2?

BRCA1

TUMOR SUPPRESSOR GENE LINKED TO HEREDITARY BREAST AND OVARIAN CANCER.
WELL-STUDIED WITH NUMEROUS CLINICALLY ANNOTATED VARIANTS.
CRITICAL IN CANCER RISK ASSESSMENT AND EARLY INTERVENTION.



CFTR

GENE RESPONSIBLE FOR CYSTIC FIBROSIS, A COMMON AUTOSOMAL RECESSIVE DISORDER.
RICH DATABASE OF PATHOGENIC VARIANTS WITH WELL-DEFINED CLINICAL CONSEQUENCES.
WIDELY USED IN NEWBORN SCREENING AND CARRIER TESTING PANELS.

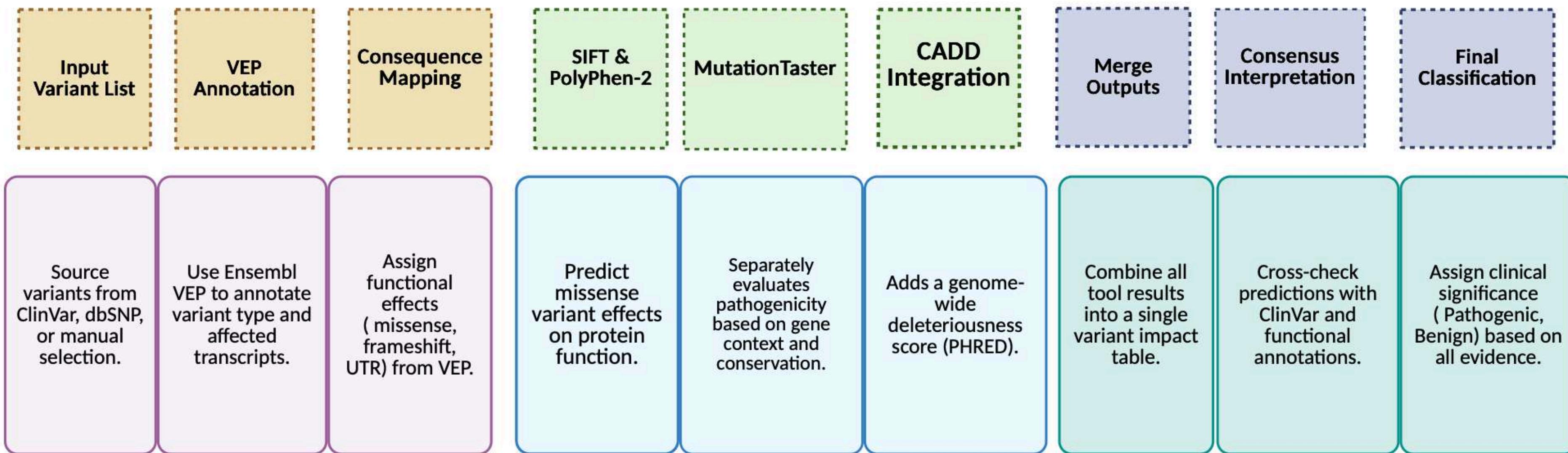
GJB2

ENCODES CONNEXIN 26: MUTATIONS LEAD TO CONGENITAL SENSORINEURAL HEARING LOSS.
ONE OF THE MOST FREQUENT CAUSES OF INHERITED DEAFNESS WORLDWIDE.
REPRESENTS THE AUDITORY GENETICS DOMAIN IN DIAGNOSTICS.



VARIANT INTERPRETATION PIPELINE

Predictive Pipeline for Interpreting Genomic Variants



Phase 1: Variant Collection & Annotation

Phase 2: Computational Scoring

Phase 3: Interpretation & Classification

TOOLS FOR VARIANT INTERPRETATION

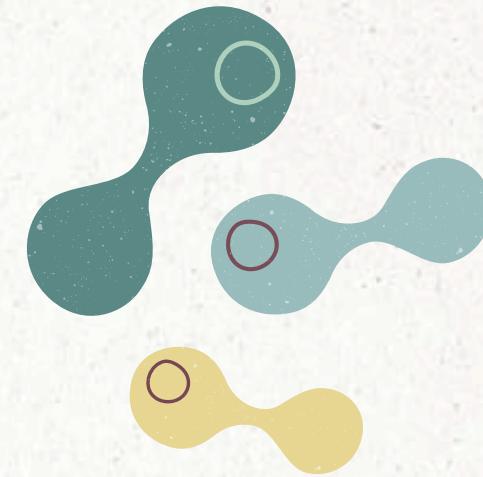
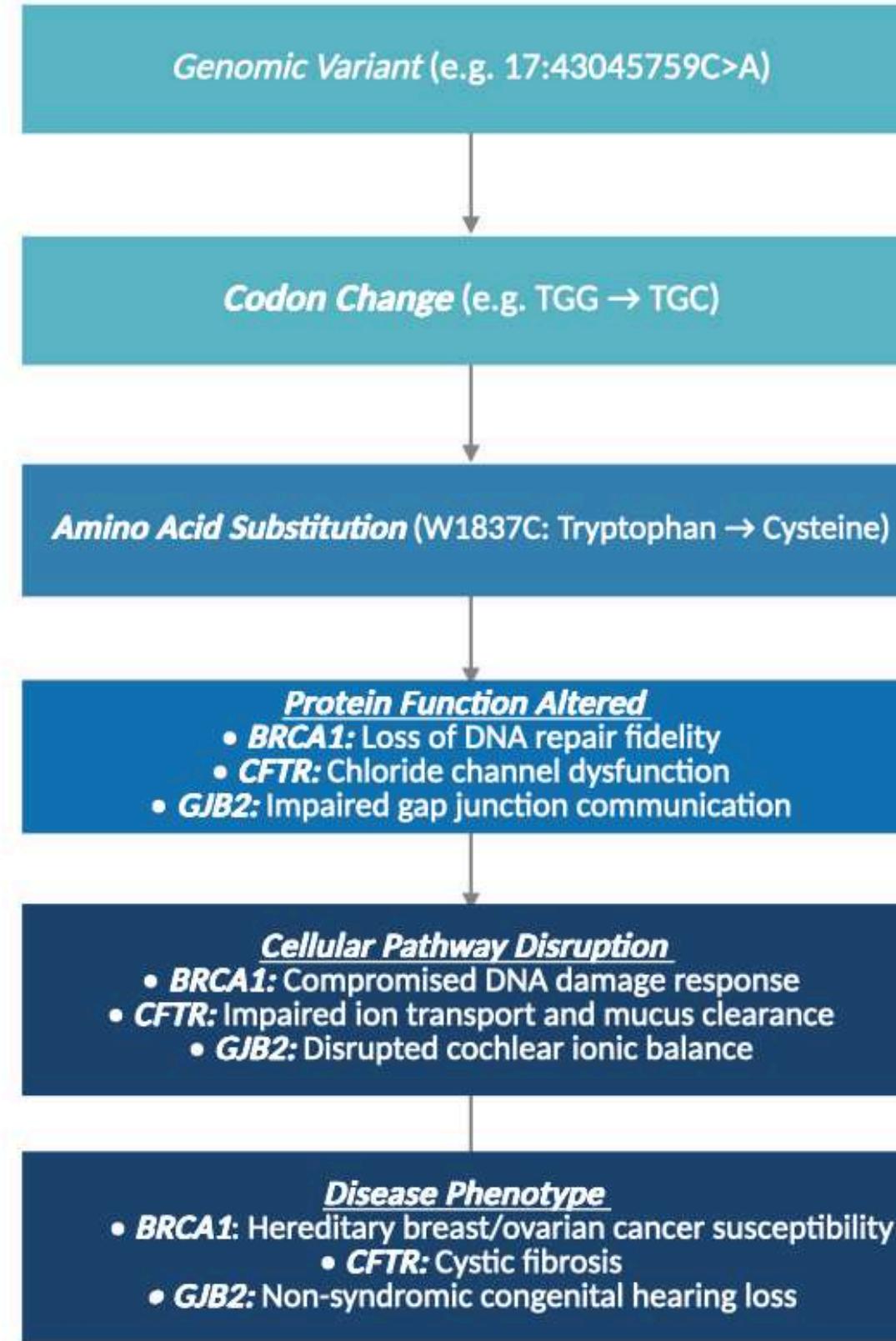
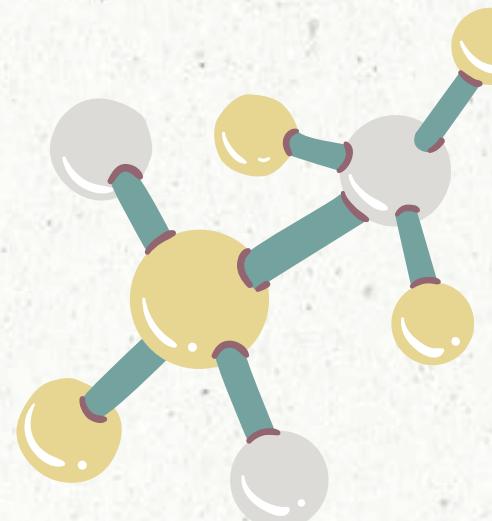
- Ensembl VEP – Annotates functional consequences of variants (e.g., missense, stop-gain) using Ensembl gene models.
- SIFT – Predicts whether an amino acid substitution affects protein function based on sequence homology and physical properties.
- PolyPhen-2 – Estimates the potential impact of missense mutations using sequence-based and structure-based predictive features.
- MutationTaster – Predicts the disease-causing potential of sequence variants by evaluating conservation, protein features, and splicing.
- CADD (Combined Annotation Dependent Depletion) – Integrates multiple annotations into a single score representing variant deleteriousness across the genome.
- ClinVar – Public database of clinically annotated variants used here as the ground truth for benchmarking predictions.

COMPARISON OF TOOLS USED

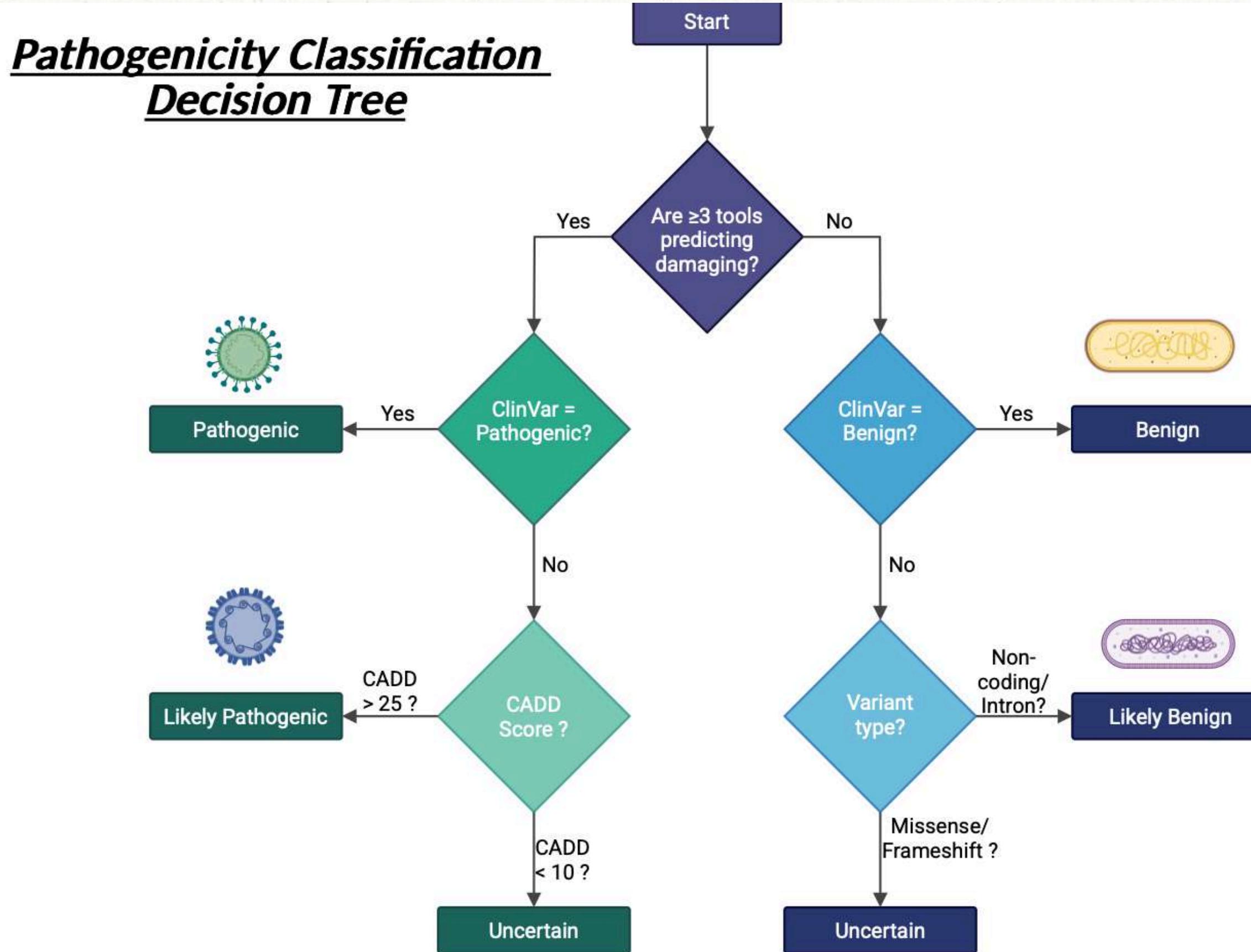
Tool	Input Format	Output	Score Interpretation
SIFT	VCF / protein FASTA	Tolerated / Damaging	< 0.05 = damaging
PolyPhen-2	Protein sequence / VCF	Benign / Possibly or Probably Damaging	> 0.85 = probably damaging
MutationTaster	Genomic coordinates (VCF or HGVS)	Disease-causing / Polymorphism	Qualitative + confidence score (max 215)
CADD	VCF / BED	Raw score + PHRED-like score	PHRED > 20 = top 1% most deleterious variants

MECHANISTIC PATHWAY OF DISEASE-CAUSING VARIANTS

Mechanistic Pathway of Disease-Causing Variants

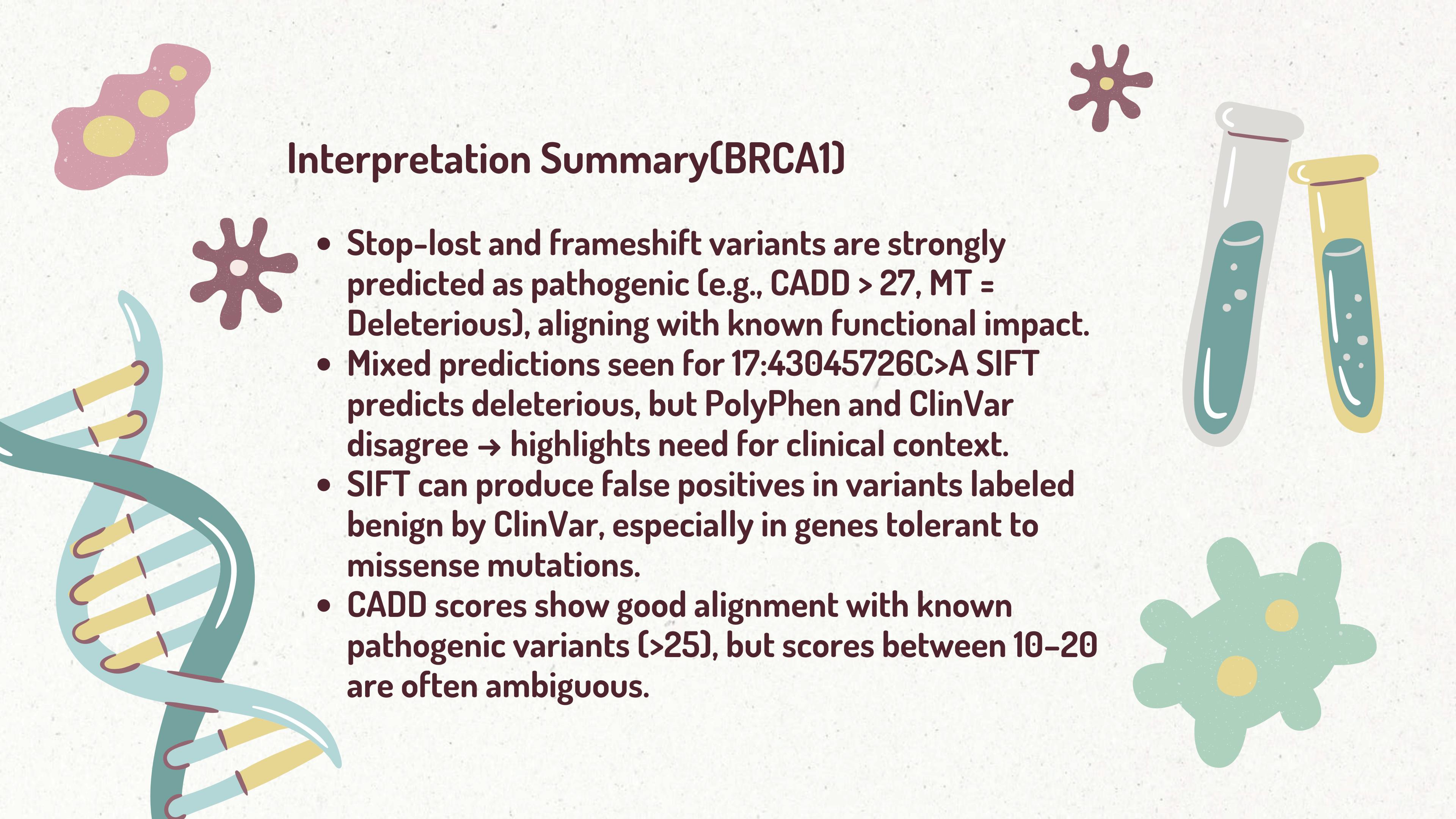


PATHOGENICITY CLASSIFICATION DECISION TREE



BRCA1 MASTER VARIANT IMPACT TABLE

Variant	MutationTaster Result	AA Change	MutationTaster Score	ClinVar Significance	CADD PHRED	Variant Type	SIFT	PolyPhen	Notes	Molecular consequence
17:43045685T>A	Deleterious	H1862L	99	Benign/Likely benign	10.58	single nucleotide variant	deleterious_low_confidence(0.01)	benign(0.013)	choosed based on MANE	missense variant 3 prime UTR variant non-coding transcript variant
17:43045726C>A	Deleterious	Q1848H	24	Benign/Likely benign	22.6	single nucleotide variant	deleterious(0)	benign(0.263)	choosed based on MANE	synonymous variant 3 prime UTR variant
17:43045752A>AG	Deleterious	Stop codon lost + prolonged protein	stop codon lost	Pathogenic/Likely pathog	27.1	Insertion	N/A	N/A	choosed based on MANE	frameshift variant 3 prime UTR variant non-coding transcript variant
17:43045759C>A	Deleterious	W1837C	215	Pathogenic/Likely pathog	27.6	single nucleotide variant	deleterious(0)	probably_damaging(0.991)	choosed based on MANE	missense variant 3 prime UTR variant non-coding transcript variant
17:43045784T>TA	Deleterious	Stop codon lost + prolonged protein	stop codon lost	Pathogenic	36	Duplication	N/A	N/A	choosed based on MANE	nonsense frameshift variant stop lost
17:43047654T>A	Benign	N1819I	149	Benign/Likely benign	7.686	single nucleotide variant	deleterious(0)	benign(0.015)	choosed based on MANE	missense variant synonymous variant

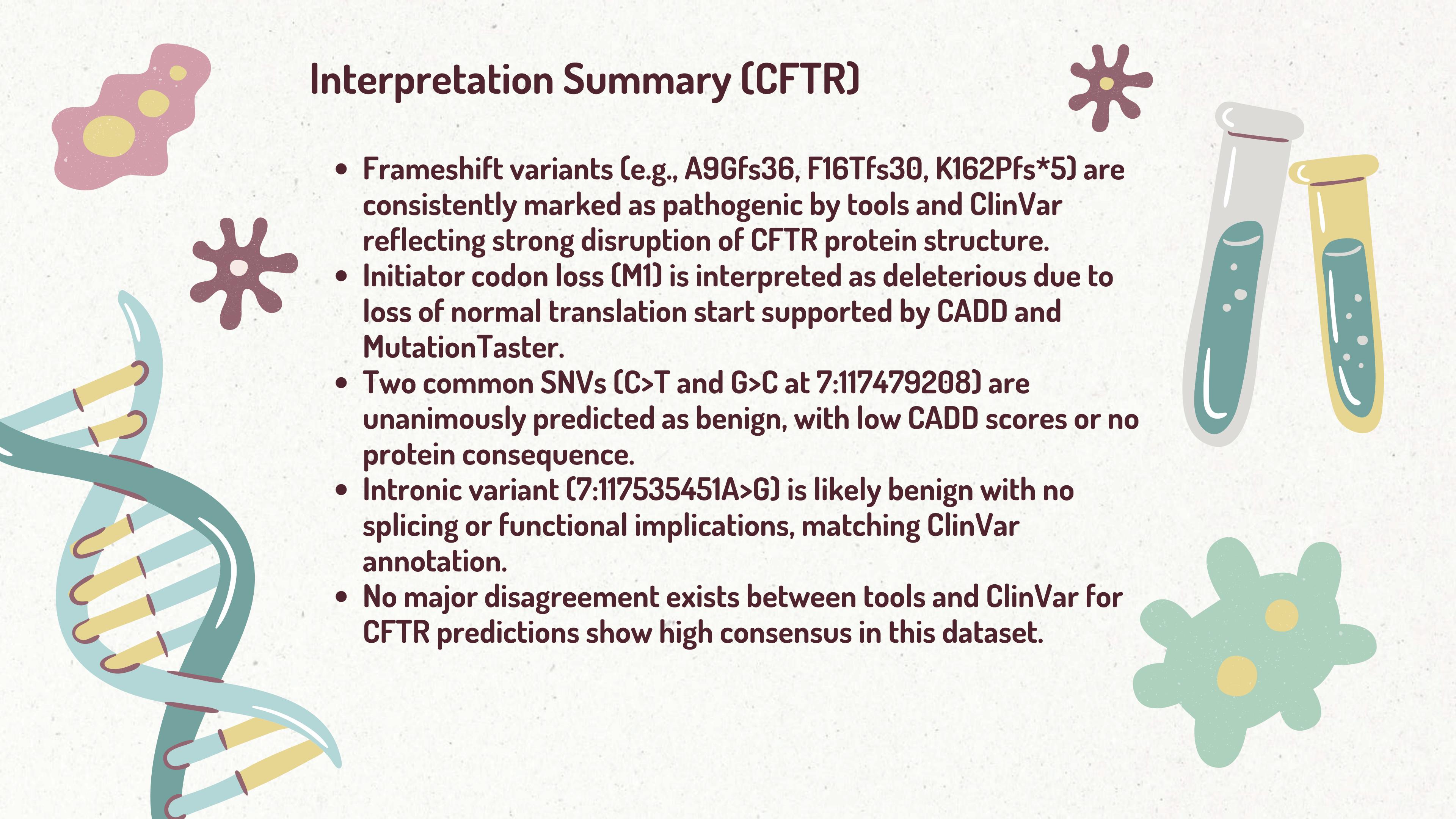


Interpretation Summary(BRCA1)

- Stop-lost and frameshift variants are strongly predicted as pathogenic (e.g., CADD > 27, MT = Deleterious), aligning with known functional impact.
- Mixed predictions seen for 17:43045726C>A SIFT predicts deleterious, but PolyPhen and ClinVar disagree → highlights need for clinical context.
- SIFT can produce false positives in variants labeled benign by ClinVar, especially in genes tolerant to missense mutations.
- CADD scores show good alignment with known pathogenic variants (>25), but scores between 10–20 are often ambiguous.

CFTR MASTER VARIANT IMPACT TABLE

Variant	MutationTaster Result	AA Change	MutationTaster Score	ClinVar Significance	CADD PHRED	Variant type	SIFT	PolyPhen	Molecular consequence
7:117479208C>T	Benign	N/A	100	Benign/Likely benign	6.987	single nucleotide variant	N/A	N/A	N/A
7:117479208G>C	Benign	N/A	N/A	Benign/Likely benign	N/A	single nucleotide variant	N/A	N/A	N/A
7:117480097G>A	Deleterious	MI? Loss of initiating Methionine, new frame start	N/A	Pathogenic/Likely pathogenic	24.9	single nucleotide variant	deleterious_low_confidence(0)	benign(0.362)	missense variant initiator_codon variant
7:117480117insG	Deleterious	A9Gfs*36	N/A	Pathogenic/Likely pathogenic	21.8	Duplication	N/A	N/A	frameshift variant
7:117480134delAACT	Deleterious	F16Tfs*30	N/A	Pathogenic	38	Deletion	N/A	N/A	frameshift variant
7:117531108insCC	Deleterious	K162Pfs*5	N/A	Pathogenic/Likely pathogenic	N/A	Insertion	N/A	N/A	frameshift variant
7:117535451A>G	Benign	N/A	N/A	Benign	N/A	single nucleotide variant	N/A	N/A	intron variant



Interpretation Summary (CFTR)

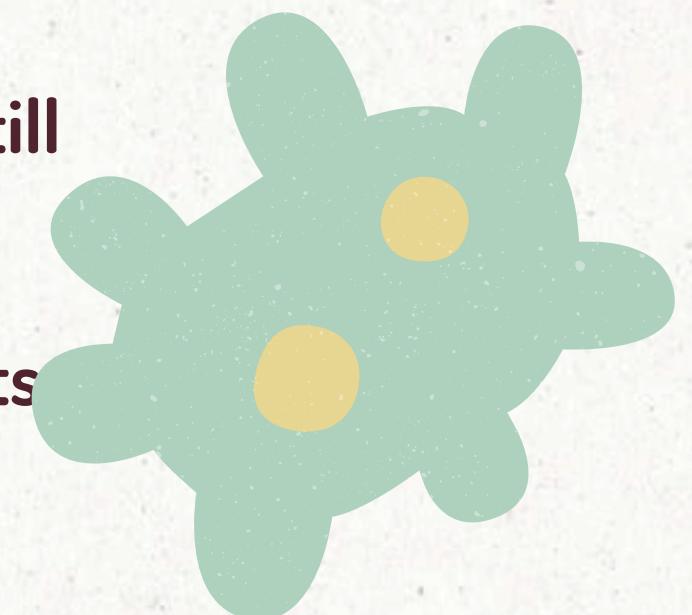
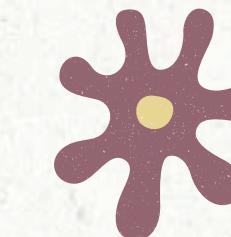
- **Frameshift variants** (e.g., A9Gfs36, F16Tfs30, K162Pfs*5) are consistently marked as pathogenic by tools and ClinVar reflecting strong disruption of CFTR protein structure.
- **Initiator codon loss (M1)** is interpreted as deleterious due to loss of normal translation start supported by CADD and MutationTaster.
- Two common SNVs (C>T and G>C at 7:117479208) are unanimously predicted as benign, with low CADD scores or no protein consequence.
- Intronic variant (7:117535451A>G) is likely benign with no splicing or functional implications, matching ClinVar annotation.
- No major disagreement exists between tools and ClinVar for CFTR predictions show high consensus in this dataset.

GJB2 MASTER VARIANT IMPACT TABLE 2

Variant	MutationTaster Result	CADD PHRED	AA change	Variant type	SIFT	Polyphen	ClinVar label	Molecular consequence
13:20188790G>A	Benign	4.39		single nucleotide variant	N/A	N/A	Benign	3 prime UTR variant
13:20188817A>C	Benign	15.06		single nucleotide variant	N/A	N/A	Benign	3 prime UTR variant
13:20188976insAAATTCC AGACACTGCAATCATGA ACACTGTGAAGACAGTC TTCTC	Deleterious		C202Wfs*10	Duplication	N/A	N/A	Pathogenic/Likely pathogenic	nonsense
13:20189006delTG	Deleterious		V193Cfs*3	Microsatellite	N/A	N/A	Pathogenic/Likely pathogenic	frameshift variant
13:20189017delCT				Deletion	N/A	N/A	Pathogenic/Likely pathogenic	frameshift variant
13:20189125C>A	Benign	20.4	AAE: V153F Score: 50	single nucleotide variant	tolerated_low_confidence(0.19)	N/A	Benign/Likely benign	missense variant
13:20189303C>G	Deleterious	25.1	AAE: M93I Score: 10	single nucleotide variant	deleterious_low_confidence(0)	N/A	Pathogenic/Likely pathogenic	missense variant



Interpretation Summary (GJB2)



- Stop-lost and frameshift variants (13:20188976insAAA.., 13:20189006delTG, 13:20189017delCT) are consistently predicted as deleterious by MutationTaster and labeled Pathogenic/Likely pathogenic in ClinVar, supporting strong concordance between computational tools and clinical annotations.
- Missense variant 13:20189303C>G (M93I) is flagged as deleterious by MutationTaster and SIFT and also marked Pathogenic by ClinVar, demonstrating high confidence in functional impact.
- Benign predictions (13:20188790G>A, 13:20188817A>C) lie in the 3' UTR, which is non-coding and often tolerates variation. Their low CADD scores and benign ClinVar status confirm low likelihood of disease relevance.
- Mixed predictions for 13:20189125C>A (V153F): CADD is moderately high (20.4), but SIFT gives low-confidence tolerance, and ClinVar still classifies it as Benign/Likely benign, suggesting it may lie in a non-critical or tolerated protein region.
- Overall, coding-region variants with truncating or frameshift effects align well across all tools and ClinVar, whereas non-coding UTR variants tend to show consistent benign predictions.

RESOURCES & LINKS

INCLUDE DIRECT LINKS TO:

1. [CLINVAR](#)
2. [DBSNP](#)
3. [ENSEMBL VEP](#)
4. [SIFT](#)
5. [POLYPHEN-2](#)
6. [MUTATIONASTER](#)
7. [CADD](#)

THANK YOU!

