

Annotation Tools

Variant Annotation Purpose

- Predict functional effect of variants
- Add gene names and transcript information
- Include population frequency data
- Clinical significance and disease associations
- Conservation scores and pathogenicity predictions

VEP

Ensembl

Variant Effect Predictor

ANNOVAR

Comprehensive

Multiple databases

SnEff

Fast

Genomic annotations

Annotation Databases

Population Databases

- gnomAD (global frequencies)
- 1000 Genomes
- ExAC, dbSNP

Clinical Databases

- ClinVar (pathogenicity)
- OMIM (disease-gene)
- COSMIC (cancer)

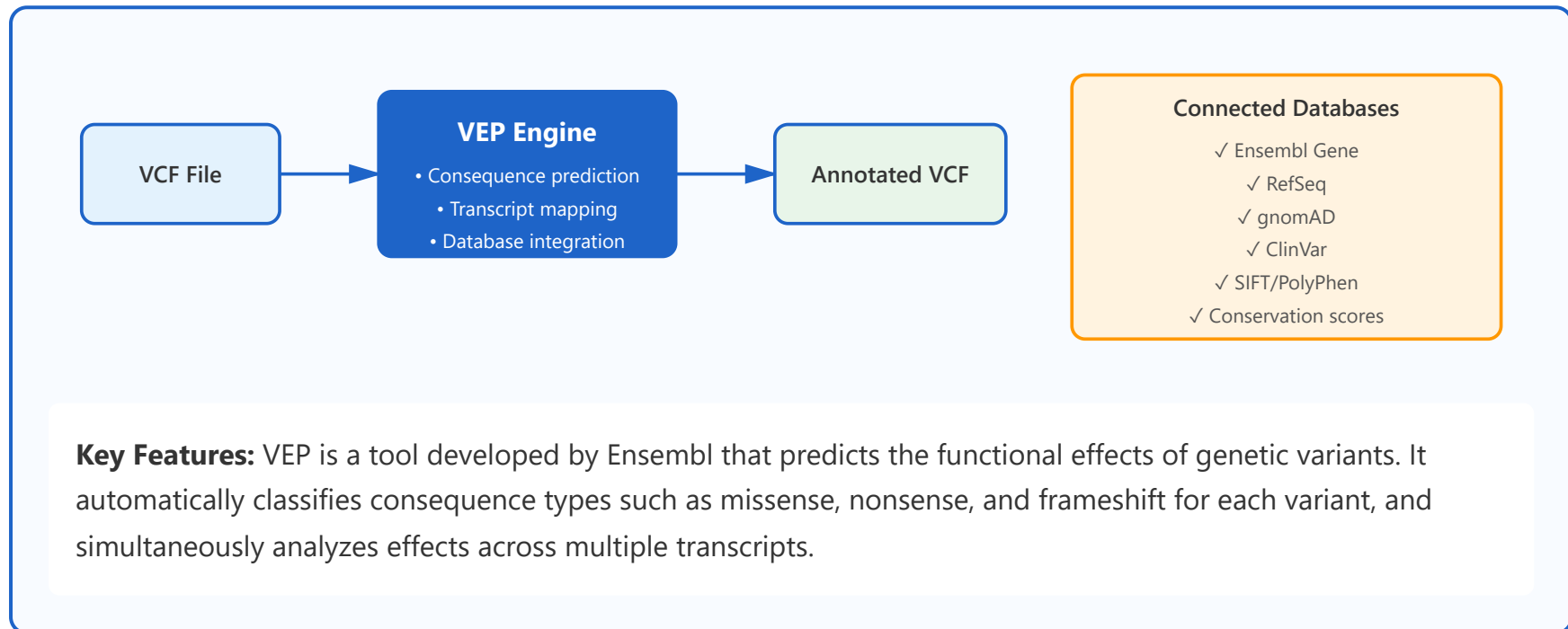
Prediction Tools

- SIFT (deleteriousness)
- PolyPhen-2
- CADD scores

Conservation

- PhyloP
- GERP++
- PhastCons

VEP (Variant Effect Predictor) Details



ANNOVAR Workflow



Features: ANNOVAR has a modular architecture that allows sequential application of various databases. Gene-based annotation identifies gene functions, while filter-based annotation adds frequency and clinical information. Command-line based, it is efficient for large-scale data processing.

SnpEff Annotation Process



Advantages: SnpEff is Java-based and boasts very fast processing speeds, automatically generating variant effect statistics in HTML format along with annotation results. It is particularly efficient for large-scale WGS data or population studies, storing all information in a structured format in the ANN field.

Population Database Usage Example

gnomAD (Genome Aggregation Database)

Variant: chr17-43044295-G-A (BRCA1)

Global AF: 0.00015 (15/100,000)

East Asian AF: 0.00008 (8/100,000)

European AF: 0.00022 (22/100,000)



Interpretation: Very low frequency → Consider pathogenicity possibility as rare variant

Clinical Database Example

ClinVar - Clinical Significance

Pathogenic ★★★★★

Review Status: 4 stars
Condition: Breast cancer
Submitters: 12 labs

OMIM - Disease Association

Gene: BRCA1

MIM: 113705

Associated:

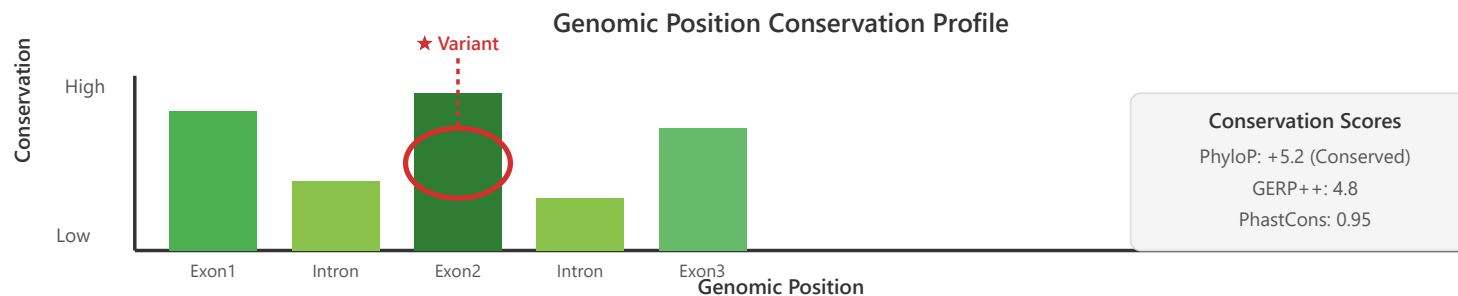
- Breast-ovarian cancer
- Fanconi anemia

Pathogenicity Prediction Comparison

Tool	Score/Result	Interpretation	Method
SIFT	0.01 (Deleterious)	Score < 0.05	Sequence homology
PolyPhen-2	0.98 (Damaging)	Score > 0.85	Structure + conservation
CADD	28.5 (Pathogenic)	Phred > 20	Machine learning

✓ **Overall Assessment:** All three tools predict pathogenicity with concordant results. Concordance across multiple prediction tools increases confidence in the functional impact of the variant.

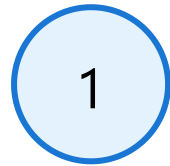
Conservation Score Visualization



💡 **Interpretation Guide:** High conservation scores indicate that the position is evolutionarily conserved, and variants at these positions are likely to have significant effects on protein function. In particular, high

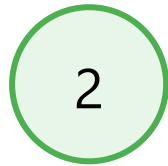
conservation in exon regions suggests functional importance.

Practical Annotation Workflow Example



VCF Preparation

Quality filtering complete



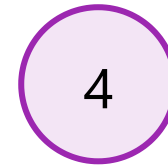
Functional Annotation

VEP/ANNOVAR



Frequency/Clinical Info

gnomAD, ClinVar



Prioritization

Filtering & Interpretation



Recommended Filtering Criteria

- ✓ Population AF < 0.01 (1%)
- ✓ CADD score > 15
- ✓ Loss-of-function variants (frameshift, nonsense, splice site)
- ✓ ClinVar: Pathogenic/Likely pathogenic
- ✓ Gene: Disease-related gene panel