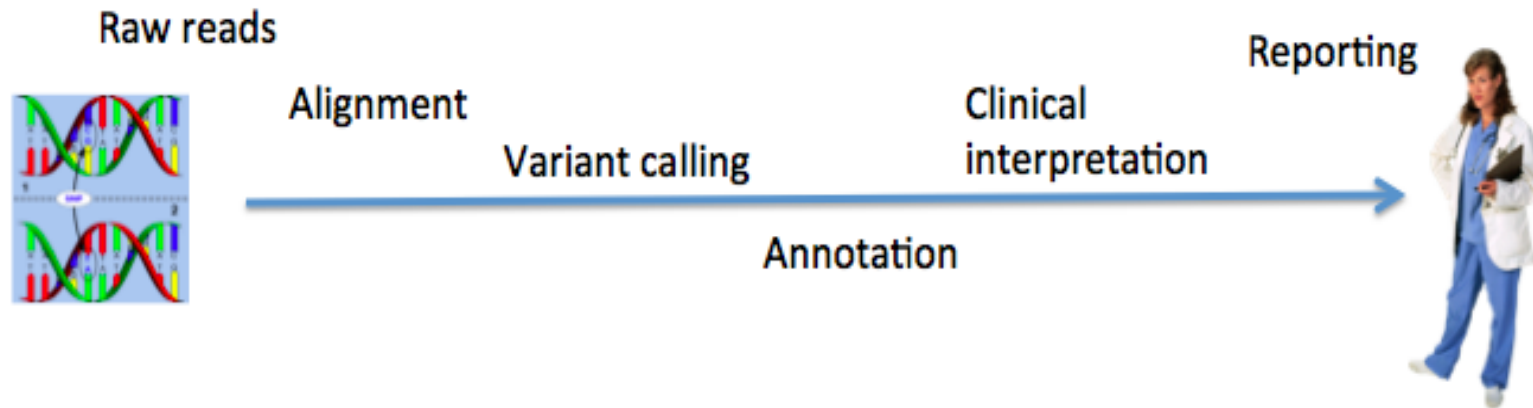


Module 2: Clinical interpretation of variants

Going From Sequence to Clinical Use



Typical Individual Differences from Reference

- ~5-10 million SNVs (varies by population)
 - 40-100,000 SNVs in coding exons
 - 10,000-12,000 synonymous (no amino acid change)
 - 8,000-11,000 non-synonymous, in 4,000-5,000 genes
- 200,000-500,000 indels (1-1000 bp) (varies by population)
 - ~150 in-frame indels in exons
 - ~200-250 shift the reading frame of an exon
- 500-1000 CNVs >1,000 bp

Approximate

Basic annotation of variants

- Gene name (if in a gene)
- Chromosome location of the change (position in reference genome)
- Location of the change within the mRNA/cDNA
- Location of the amino acid change in the protein
- Effect on protein (if in a gene)

<u>Gene</u>	<u>Chr.</u>	<u>Genomic</u>	<u>cDNA</u>	<u>Protein</u>	<u>Effect</u>
BRCA1	17	g.37038192G>T	c.199G>T	p.Gly67Trp	Non-synonymous
BRCA1	17	g.37042469_37042470delTG	c.231_232delTG	p.Cys77Ter	Stop-gained

Advanced annotation of variants

- Variant dependent methods
 - Allele frequency
 - Predicted effect of variant on protein
 - Evolutionary conservation, protein structure, amino acid properties
 - Functional characterization of variant (in vitro and/or in vivo)
- Disease-dependent methods
 - Mode of inheritance
 - Cosegregation with disease in families
 - Prior association of the gene with disease
 - Pathway analysis

Criteria used to evaluate variants

- Variant dependent methods
 - Allele frequency
 - Predicted effect of variant on protein
 - Evolutionary conservation, protein structure, amino acid properties
 - Functional characterization of variant (in vitro and/or in vivo)
- Disease-dependent methods
 - Mode of inheritance
 - Cosegregation with disease in families
 - Prior association of the gene with disease
 - Pathway analysis

Allele frequency in general population

- For a suspected Mendelian disease, a variant observed in the general, healthy population is assumed non-pathogenic
- 1000 Genomes (www.1000genomes.org)
- dbSNP
- And others.....

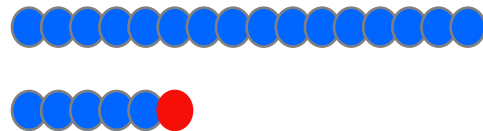
1000 Genomes

A Deep Catalog of Human Genetic Variation

Predicting the effect of a variant is CHALLENGING

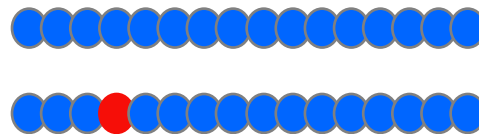
Probably Damaging

Stop-loss
Stop-gained
Frameshift
Splice disruptor



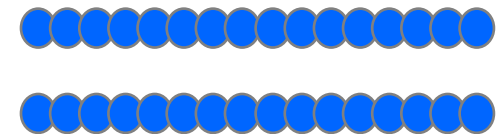
Possibly Damaging

Non-synonymous
In-frame In/Del



Likely not Damaging

5'/3' UTR
Synonymous
Intergenic
Intronic
Non-coding genes

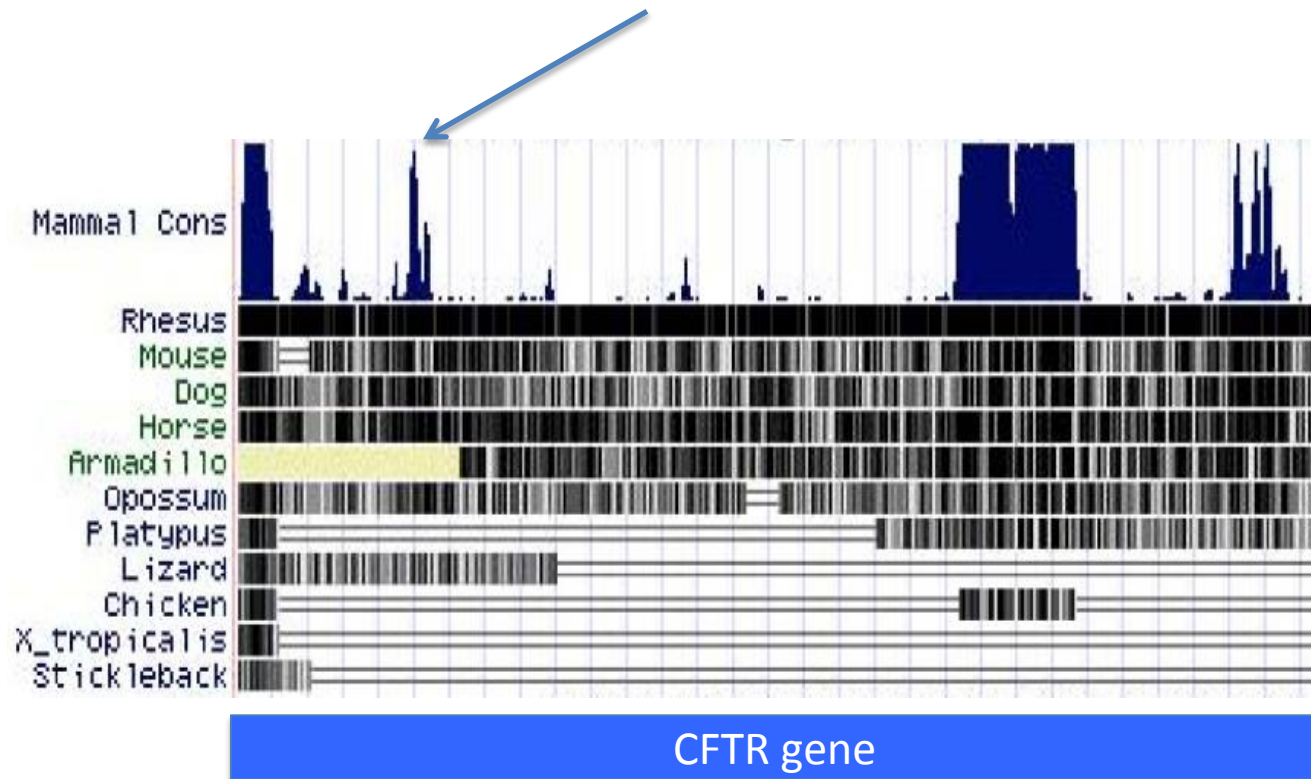


Predicting the effect of non-synonymous variants

- Evolutionary conservation
- Protein structure
- Amino acid properties
- These criteria are applied together by various computer algorithms to assess how damaging a change might be

Evolutionary conservation

Mutations in conserved positions more likely deleterious

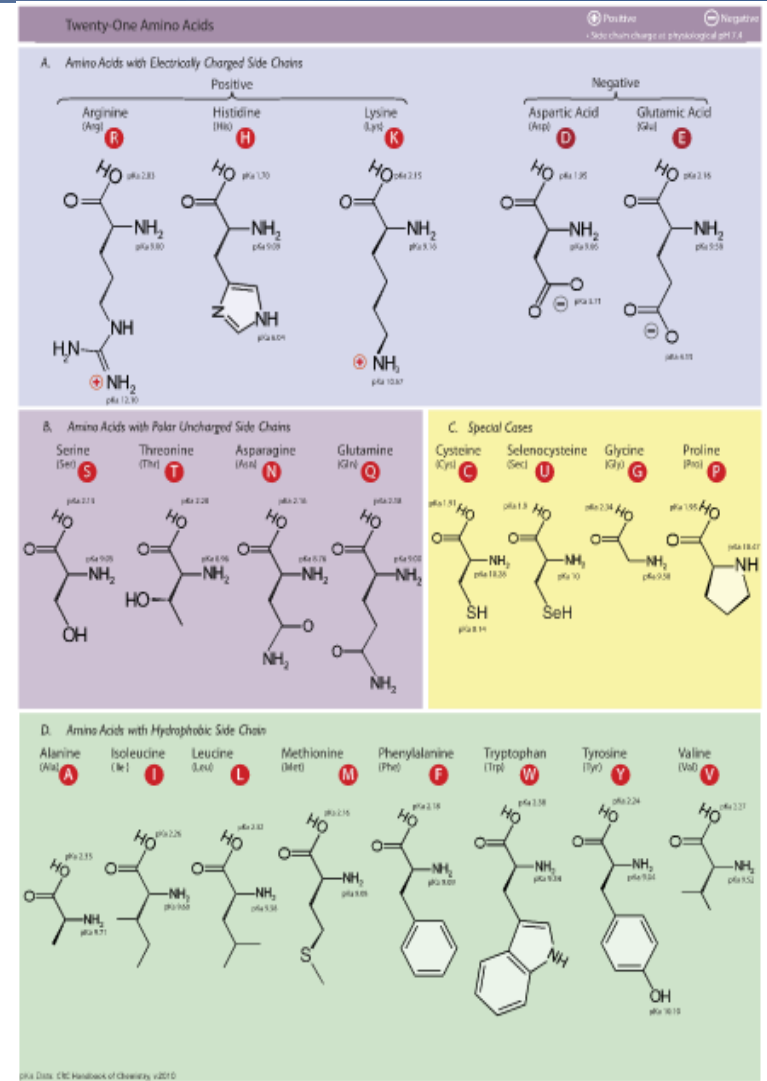


Amino acid properties and protein structure

Not all amino acid changes are equivalent

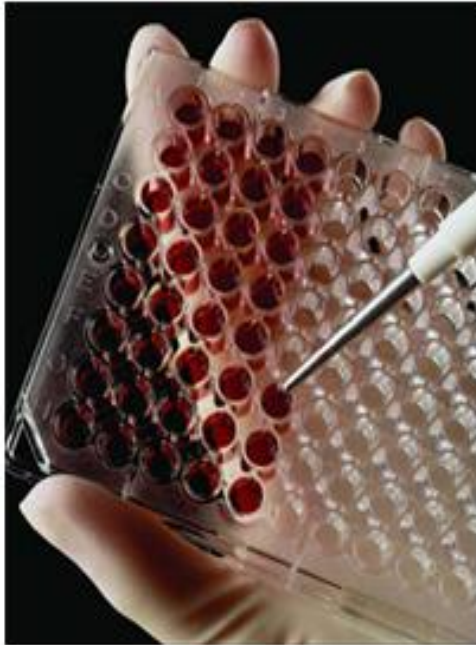
Conservative changes less likely to affect protein structure/function

Location of change within protein matters as well



In vitro/in vivo functional studies

In vitro (cell-based assays)



In vivo (animal models)



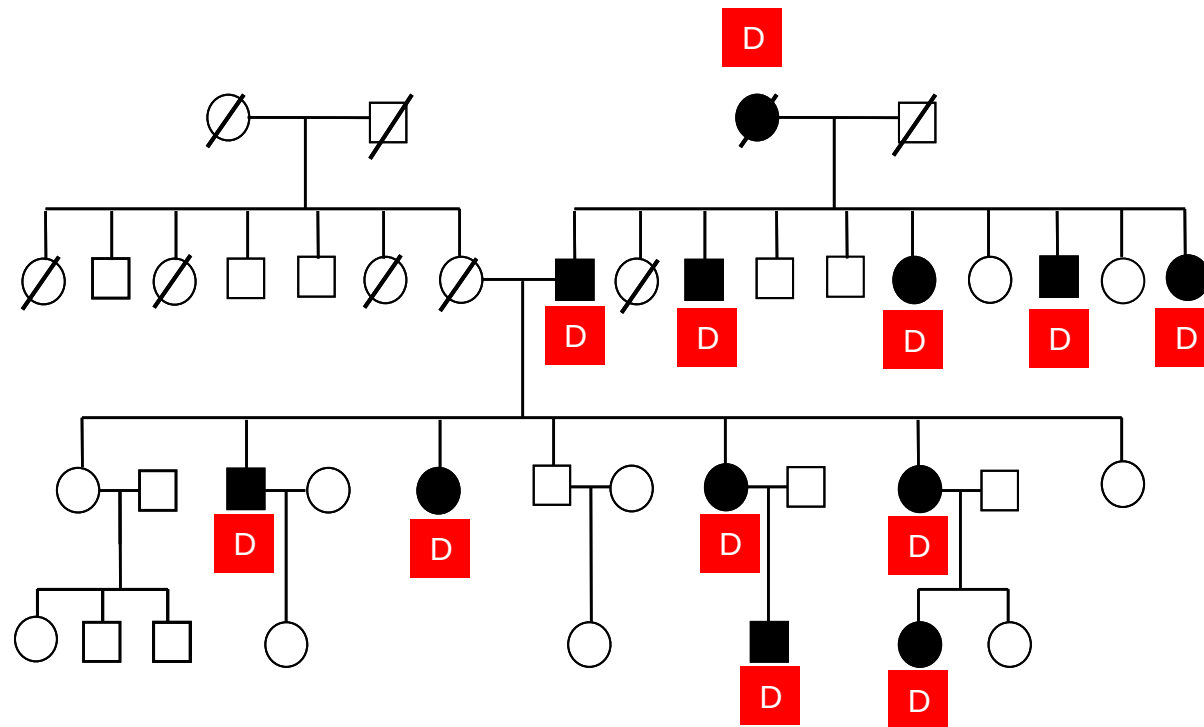
Criteria used to evaluate variants

- Variant dependent methods vary in their ability to predict the effect of a variant on gene or protein function. Some are highly predictive, others are, at best, suggestive or circumstantial.

Criteria used to evaluate variants

- Variant dependent methods
 - Allele frequency
 - Predicted effect of variant on protein
 - Evolutionary conservation, protein structure, amino acid properties
 - Functional characterization of variant (in vitro and/or in vivo)
- Disease-dependent methods
 - Cosegregation with disease in families
 - Prior association of the gene with disease (OMIM)
 - Pathway analysis

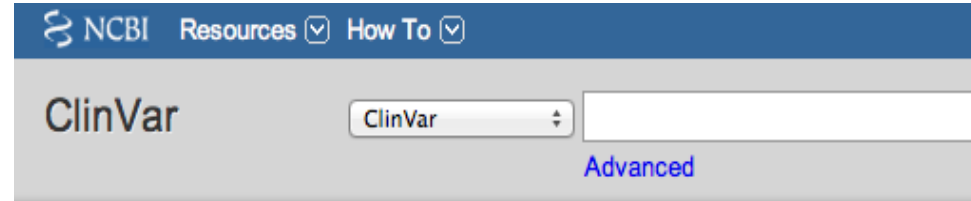
Cosegregation of variant with disease in families



Prior association of gene with disease



<http://www.ncbi.nlm.nih.gov/omim>



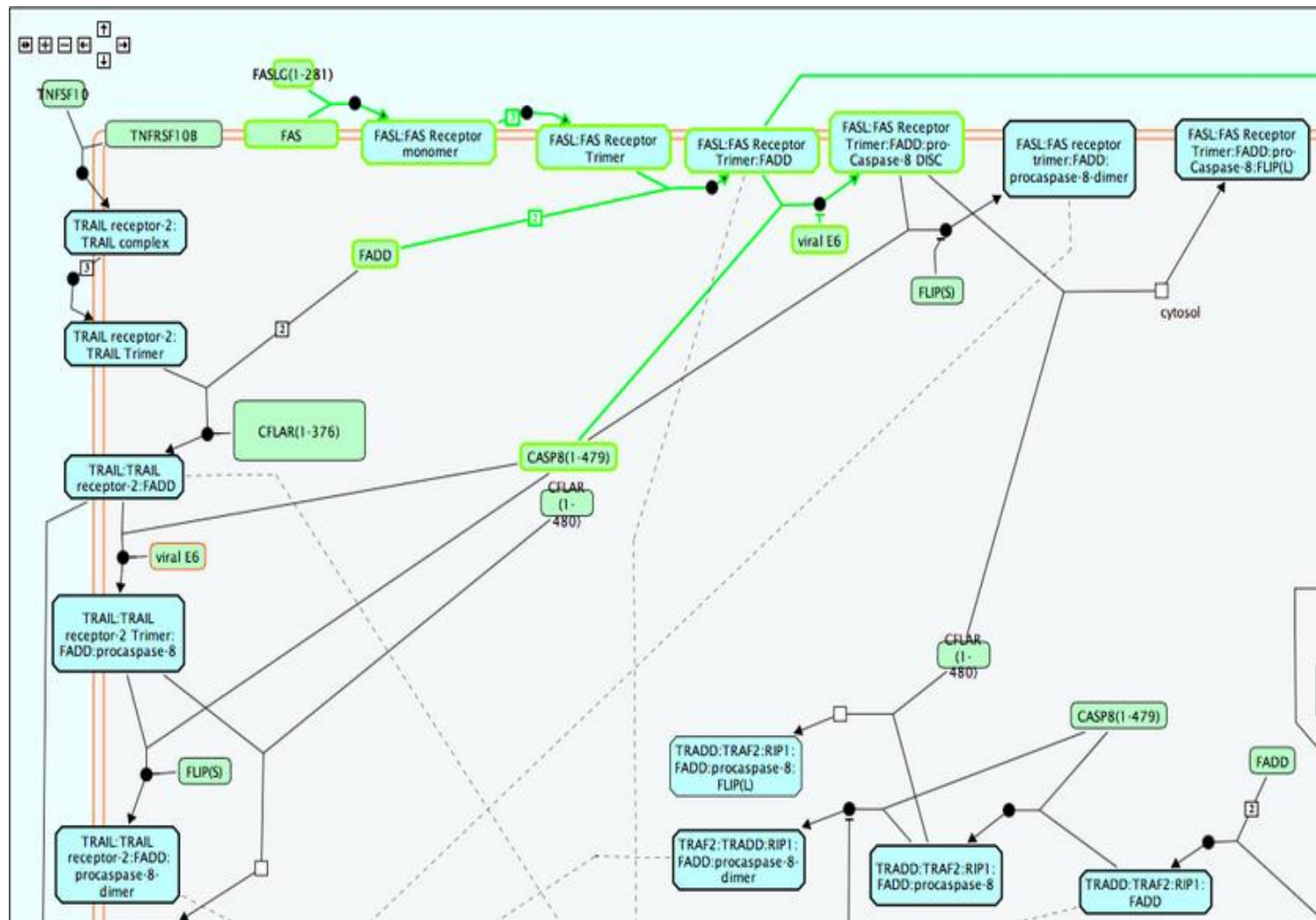
<https://www.ncbi.nlm.nih.gov/clinvar/>



Locus Specific Mutation Databases

<http://www.hgvs.org/dblist/dblist.html>

Gene in a disease pathway



Typical classification scheme

- Known pathogenic
- Likely pathogenic
- Variant of unknown significance (VOUS or VUS)
- Likely benign
- Benign

Classifying Variants

Deleterious Variant	Pathogenic Variant
Located in or near a coding exon and is..... • A frameshift or nonsense mutation that causes premature termination of translation	• Strongly associated with disease in affected individuals versus in unaffected individuals
• A non-synonymous amino acid change affecting a highly conserved residue through evolution	• Tracks with the disease in a family with multiple affected members
• A splice-site mutation in an intron that is highly likely to cause abnormal splicing	• Experimental evidence in animal models that the alteration causes disease
• NOT found as a common variant in a population of matching ethnic background	• <i>In vitro</i> experiments showing the variant changes function and is likely to cause disease

Classifying Variants

(Likely) Benign Variants
Common in an ethnic group without being associated with frequent disease
Synonymous change that does not change the amino acid encoded
Non-synonymous change of an amino acid residue that is not conserved between species

“Typical” Variant Classification Test Report

- “Mutations in this gene have been previously reported to cause this disease”
- “This variant affects a highly conserved cysteine residue in gene XYZ resulting in the substitution of a positively charged amino acid, arginine, for a sulfur-containing amino acid in an important functional domain of the enzyme encoded by gene XYZ”
- “Published biochemical studies have shown this substitution causes loss of activity of the enzyme”
- “This variant has been reported in 3 unrelated patients with this disease”
- “In one unrelated patient, the disease was co-inherited with the mutation in 5 other affected members of the family”