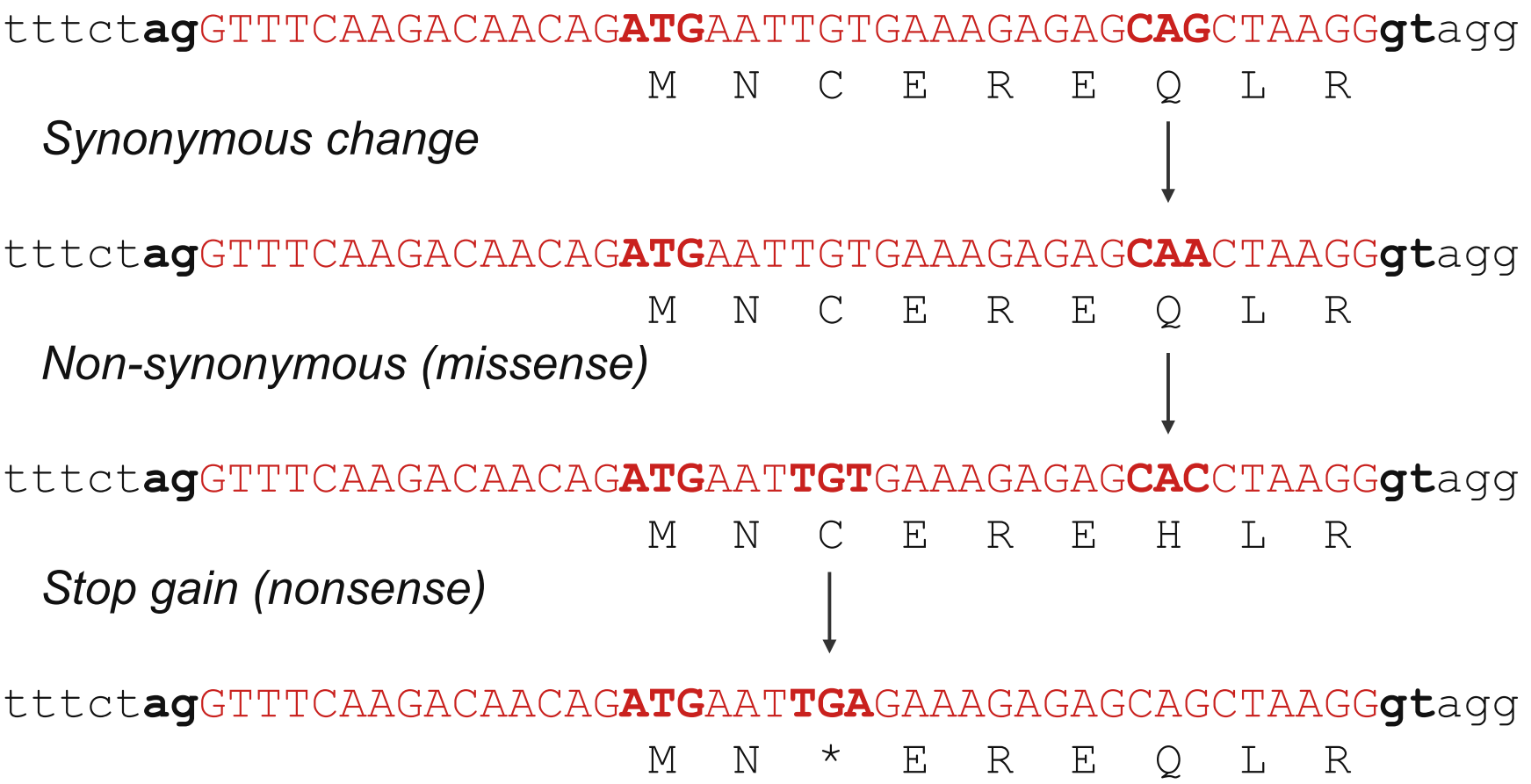
Examples of coding changes in *RBFOX1*

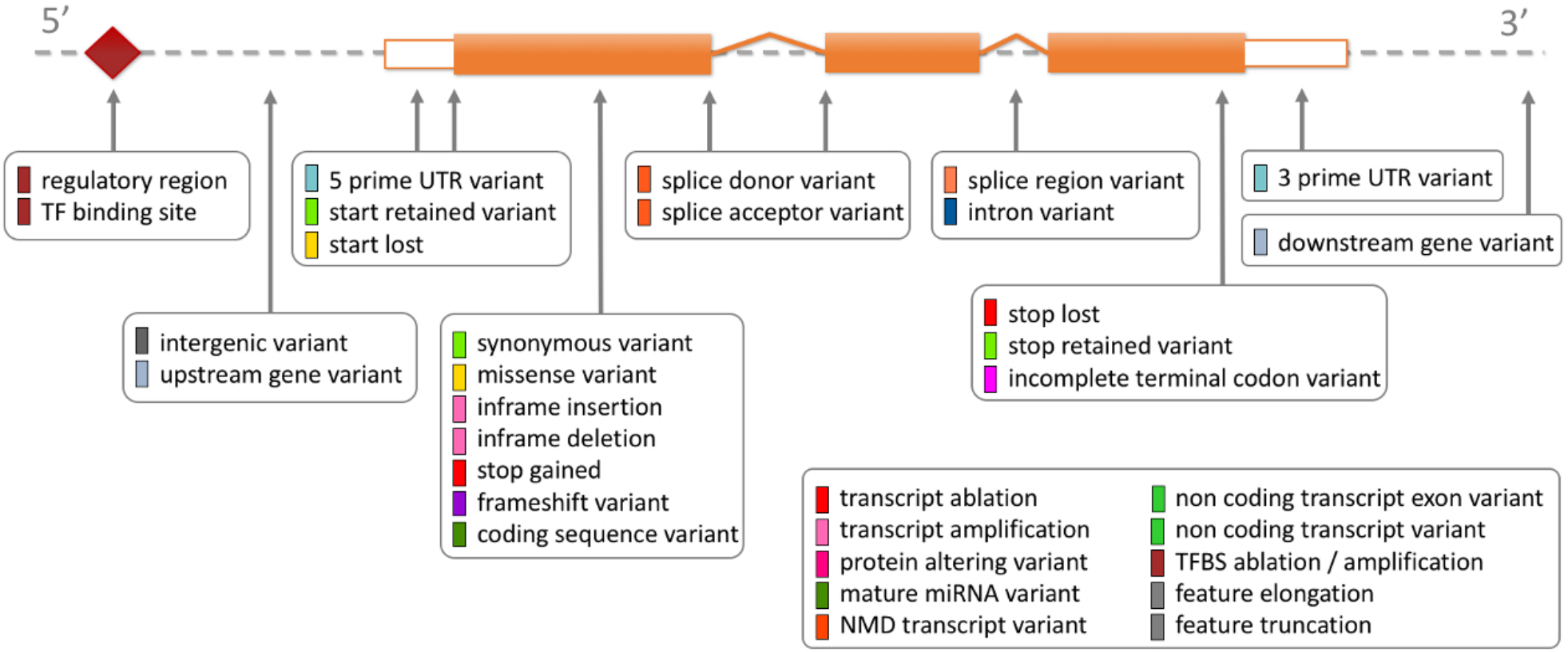


Examples of coding changes in *RBFOX1*

ENSEMBL Variant Effect Predictor

Variation consequences

Promoter ♦ 5'-UTR ♦ Start (ATG) ♦ Donor(GT) ♦ Acceptor(AG) ♦ … ♦ Stop(TAA,…) ♦ 3'-UTR



https://www.ensembl.org/info/genome/variation/prediction/predicted_data.html#consequences

ENSEMBL Variant Effect Predictor

Variation consequences and impact

*	SO term	SO description	SO accession	Display term	IMPACT
	transcript_ablation	A feature ablation whereby the deleted region includes a transcript feature	SO:0001893	Transcript ablation	HIGH
	splice_acceptor_variant	A splice variant that changes the 2 base region at the 3' end of an intron	SO:0001574	Splice acceptor variant	HIGH
	splice_donor_variant	A splice variant that changes the 2 base region at the 5' end of an intron	SO:0001575	Splice donor variant	HIGH
	stop_gained	A sequence variant whereby at least one base of a codon is changed, resulting in a premature stop codon, leading to a shortened transcript	SO:0001587	Stop gained	HIGH
	frameshift_variant	A sequence variant which causes a disruption of the translational reading frame, because the number of nucleotides inserted or deleted is not a multiple of three	SO:0001589	Frameshift variant	HIGH
	stop_lost	A sequence variant where at least one base of the terminator codon (stop) is changed, resulting in an elongated transcript	SO:0001578	Stop lost	HIGH
	start_lost	A codon variant that changes at least one base of the canonical start codon	SO:0002012	Start lost	HIGH
	transcript_amplification	A feature amplification of a region containing a transcript	SO:0001889	Transcript amplification	HIGH
	inframe_insertion	An inframe non synonymous variant that inserts bases into in the coding sequence	SO:0001821	Inframe insertion	MODERATE
	inframe_deletion	An inframe non synonymous variant that deletes bases from the coding sequence	SO:0001822	Inframe deletion	MODERATE
	missense_variant	A sequence variant, that changes one or more bases, resulting in a different amino acid sequence but where the length is preserved	SO:0001583	Missense variant	MODERATE
	protein_altering_variant	A sequence_variant which is predicted to change the protein encoded in the coding sequence	SO:0001818	Protein altering variant	MODERATE
	splice_region_variant	A sequence variant in which a change has occurred within the region of the splice site, either within 1-3 bases of the exon or 3-8 bases of the intron	SO:0001630	Splice region variant	LOW
	incomplete_terminal_codon_variant	A sequence variant where at least one base of the final codon of an incompletely annotated transcript is changed	SO:0001626	Incomplete terminal codon variant	LOW
	start_retained_variant	A sequence variant where at least one base in the start codon is changed, but the start remains	SO:0002019	Start retained variant	LOW
	stop_retained_variant	A sequence variant where at least one base in the terminator codon is changed, but the terminator remains	SO:0001567	Stop retained variant	LOW
	synonymous_variant	A sequence variant where there is no resulting change to the encoded	SO:0001819	Synonymous variant	LOW

https://www.ensembl.org/info/genome/variation/prediction/predicted_data.html#consequences

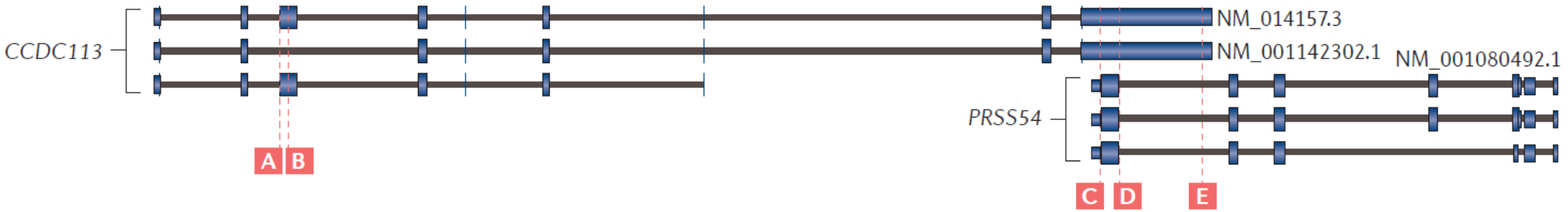
ENSEMBL Variant Effect Predictor

Variation consequences and impact

IMPACT	*Consequence examples*	*Description*
HIGH	splice_acceptor_variant, splice_donor_variant, stop_gained, stop_lost, start_lost	The variant is assumed to have high (disruptive) impact in the protein, probably causing protein truncation, loss of function or triggering nonsense mediated decay
MODERATE	inframe_insertion, inframe_deletion, missense_variant	A non-disruptive variant that might change protein effectiveness
LOW	splice_region_variant, synonymous_variant	A variant that is assumed to be mostly harmless or unlikely to change protein behaviour
MODIFIER	5_prime_UTR_variant, 3_prime_UTR_variant, intron_variant, TFBS_ablation	Usually non-coding variants or variants affecting non-coding genes, where predictions are difficult or there is no evidence of impact

https://www.ensembl.org/info/genome/variation/prediction/predicted_data.html#consequences

Complexity of variant annotation

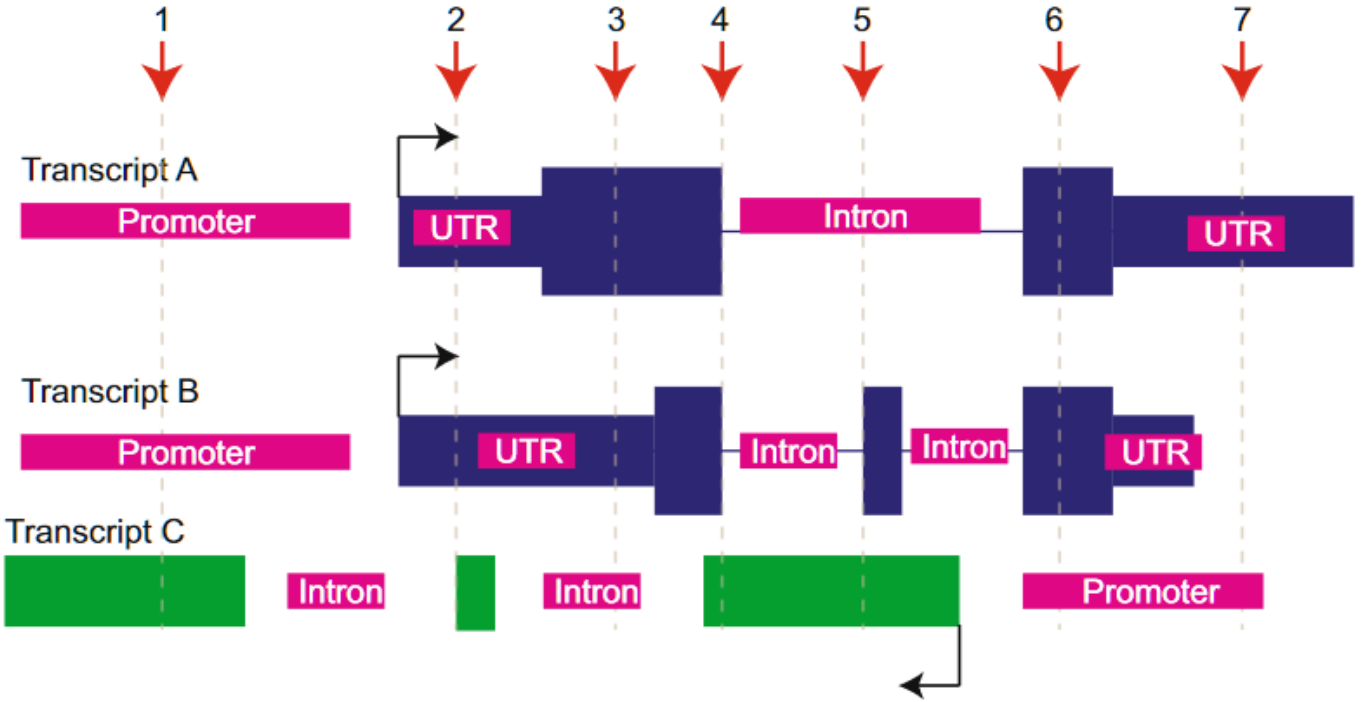


	Variant allele	Gene	Transcript change	RefSeq	Protein change	Molecular consequence
A rs765957496	G	*CCDC113*	c.228+1143A>G	NM_001142302.1	—	Intron variant
	G	*CCDC113*	c.229•2A>G	NM_014157.3	—	Splice acceptor variant
B rs775877153	A	*CCDC113*	c.228+1182T>A	NM_001142302.1	—	Intron variant
	A	*CCDC113*	c.266T>A	NM_014157.3	Met89Lys	Missense variant
C rs780162055	T	*PRSS54*	c.1135G>A	NM_001080492.1	Glu379Lys	Missense variant
	T	*CCDC113*	c.*500C>T	NM_001142302.1	—	3' UTR variant
D rs776101237	A	*PRSS54*	c.655-2A>T	NM_001080492.1	—	Splice acceptor variant
	A	*CCDC113*	c.*962T>A	NM_001142302.1	—	3' UTR variant
E rs745863465	C	*PRSS54*	c.655-18T>G	NM_001080492.1	—	Intron variant
	C	*CCDC113*	c.*996A>C	NM_001142302.1	—	3' UTR variant

A demonstration of the multiple possible effects of a single variant across transcripts and genes. The complexity of genomic annotation adds to the complexity of variant annotation. In this example, two genes, coiled-coil domain-containing 113 (*CCDC113*) and protease serine 54 (*PRSS54*) overlap on different strands of the genome, and both have multiple observed transcripts. Variants intersecting this extent of the genome show different effects depending on the gene and the transcript inspected.

Eilbeck (2017) *Nat Rev Genet*

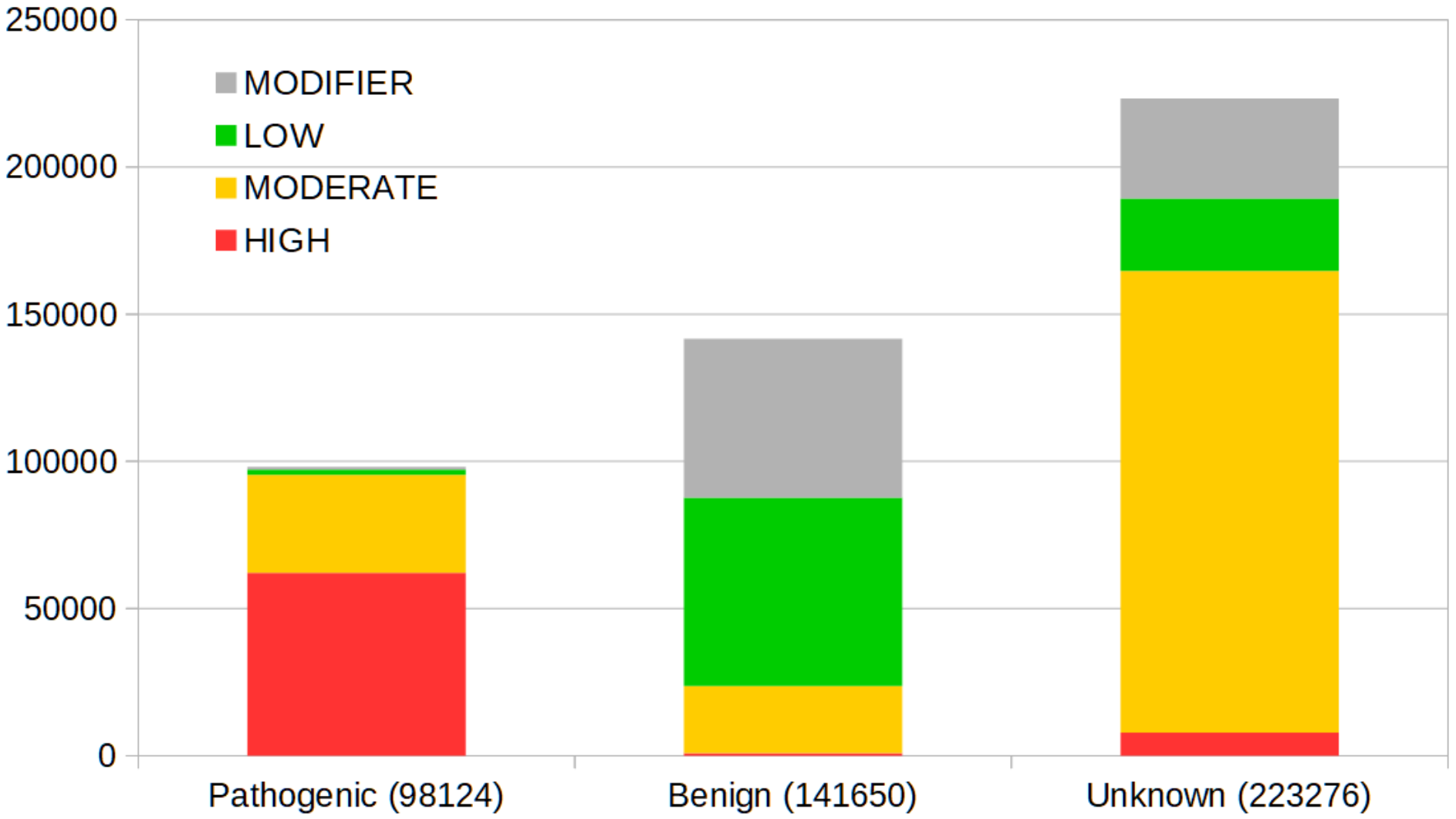
Complexity of variant annotation



Variant	Transcript A	Transcript B	Transcript C
1	Promoter	Promoter	Exon
2	Non Coding Exon	Non Coding Exon	Non Coding Splice
3	Coding Exon	Non Coding Exon	Intron
4	Coding Splice	Coding Splice	Non Coding Exon
5	Intron	Coding Splice	Non Coding Exon
6	Coding Exon	Coding Exon	Promoter
7	Non Coding Exon	Downstream	Prompter

Gloss (2018) *Exp Mol Med*

Ensembl VEP impact for ClinVar variants



ClinVar (Oct. 2019), 498,742 variants annotated with Ensembl VEP

Pathogenic variants in ClinVar (Oct. 2019)

Gene	Frameshift	Stop gain or loss	Splice site	Missense	Inframe	Synonymous	UTR	Intronic	Upstream	Start codon	Phenotype
HBB	30	14	21	35	3	1	7	12	7	4	Beta thalassemia
LDLR	387	171	51	77	9	3	7	6	0	2	Familial hypercholesterolemia
CFTR	123	111	70	105	5	3	0	20	0	4	Cystic fibrosis
GALT	21	15	11	100	1	2	0	4	1	1	Deficiency of UDPglucose-hexose-1-phosphate uridylyltransferase
KCNQ2	61	20	20	102	7	2	0	1	1	1	Benign familial neonatal seizures; Early infantile epileptic encephalopathy
MECP2	268	60	12	27	12	2	0	1	0	3	Mental retardation; Rett syndrome
MLH1	316	132	76	69	4	6	1	11	0	10	Hereditary nonpolyposis colon cancer; Lynch syndrome
OTC	22	32	39	203	5	2	0	7	0	4	Ornithine carbamoyltransferase deficiency

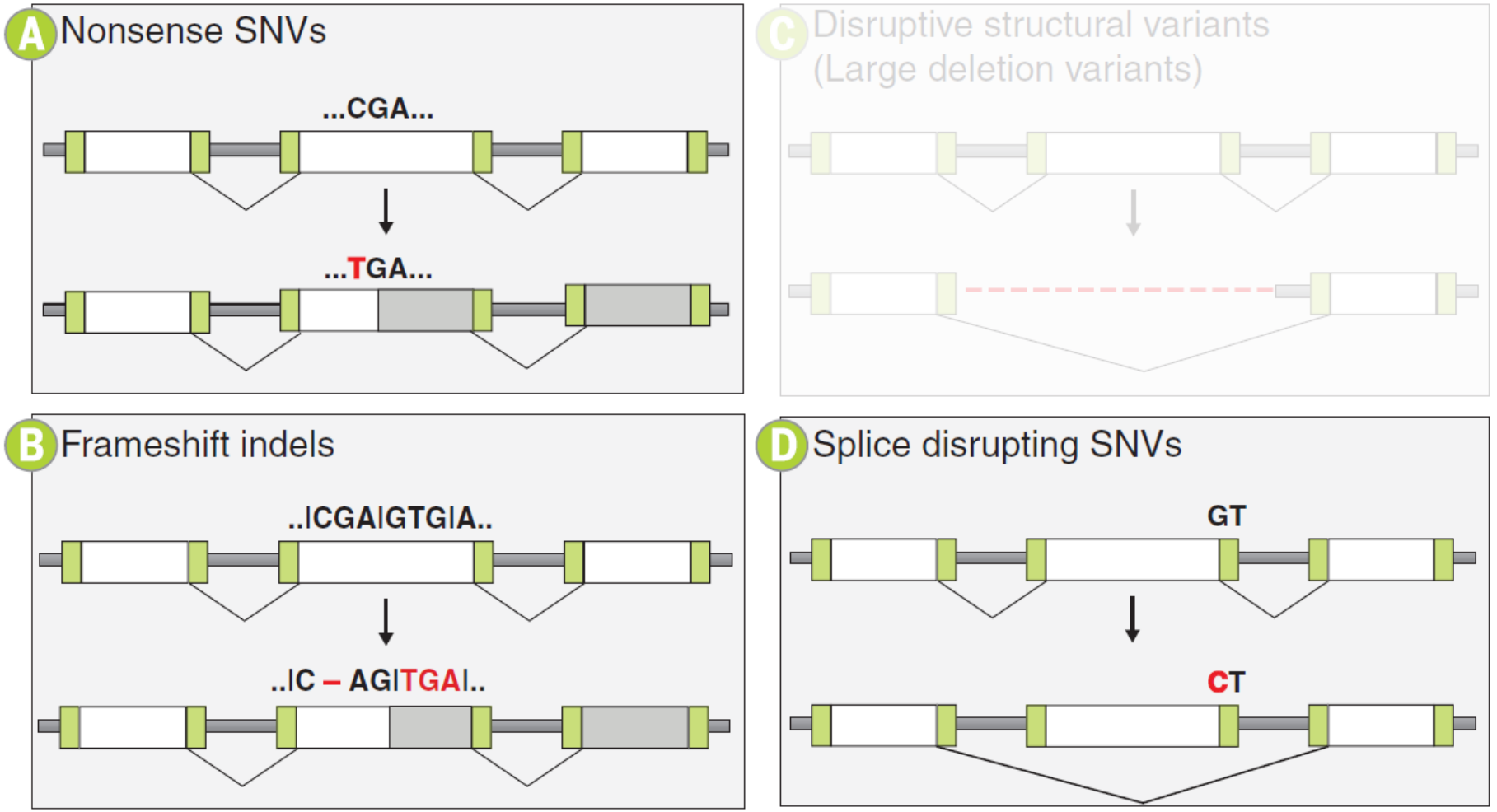
Exercise

Use ClinVar (OMIM, SNPedia) to find and save one example of disease-associated pathogenic mutation for *each* annotation type:

- stop-gain
- synonymous
- missense
- splice-site
- frameshift indel

PTVs and LoF variants

Protein-truncating variants: stop-gain, splice site, frameshift indels.
VEP impact: HIGH.



Rivas (2015) *Science*

PTVs and LoF variants

Protein-truncating variants: stop-gain, splice site, frameshift indels.
VEP impact: HIGH. *However, not all PTVs are loss-of-function*

LOFTEE tool (K.Karczewski et al): filters and flags to predict pLoF (putative LoF) from candidate PTVs. https://github.com/konradjk/loftee

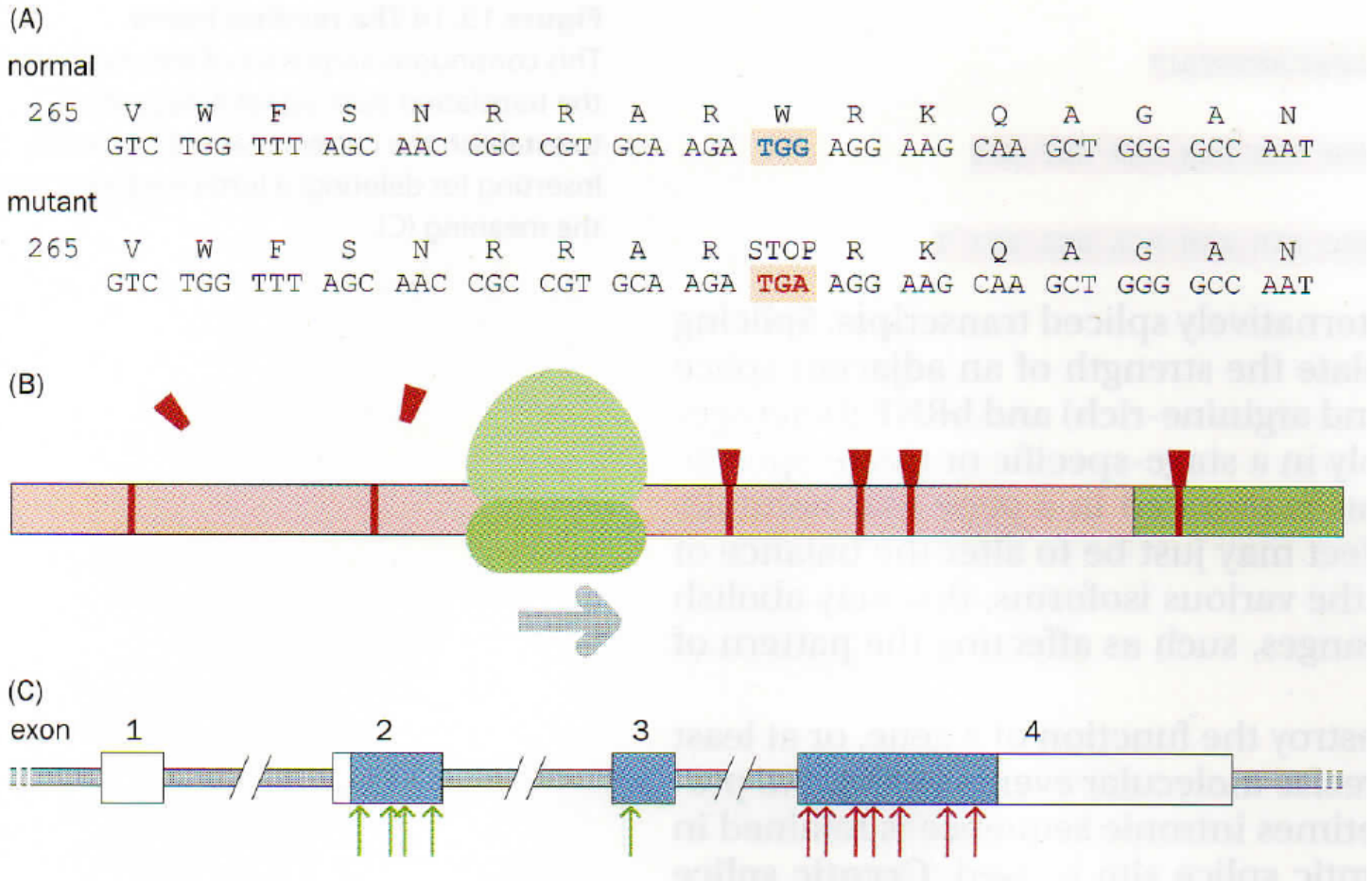
PTVs not predicted as pLoF, examples:

- Stop-gain and frameshift variants near the end of the transcript, based on the 50 bp rule
- Variants in an exon with non-canonical splice sites (GT, AG) around it
- Splice site variants rescued by nearby, in-frame splice site
- Variants in small introns

Flagged PTVs, examples:

- Variants in NAGNAG sites (acceptor sites rescued by in-frame acceptor site)
- Variants that fall in an intron with a non-canonical splice site

PTVs and nonsense-mediated decay (NMD)



(A) G>A change in exon 6 of the *PAX3* gene (B) Nonsense-mediated decay (NMD). Splice junctions (red bars) retain proteins of the exon junction complex (EJC, red triangles). Ribosome moves along the mRN A and displaces the EJC proteins. If it encounters a premature stop codon and detaches before displacing all EJCs, the mRNA is targeted for degradation. **Stop codons in the last exon or less than 50 nucleotides upstream of the last splice junction (the green zone) do not trigger NMD.** (C) Depending on whether or not a premature stop codon triggers NMD, the consequences of a nonsense mutation can be very different.

Strachan, Read – *Human Molecular Genetics*

PTVs and nonsense-mediated decay (NMD)

Ideally: **PTV → NMD → Transcript level → Protein level → Cellular functions**

However, variation in mRNA and protein expression levels are often uncorrelated: the reduction in RNA levels may not reduce the protein level, and vice versa

Battle, A., Khan, Z., Wang, S.H., Mitrano, A., Ford, M.J., Pritchard, J.K., and Gilad, Y. (2015). Impact of Regulatory Variation from RNA to Protein. Science 347, 664–667.

Narasimhan VM, Xue Y, Tyler-Smith C. Human Knockout Carriers: Dead, Diseased, Healthy, or Improved? Trends in Molecular Medicine. 2016;22(4):341-351. doi:10.1016/j.molmed.2016.02.006.

Examples of PTV impact

Protein-truncating variants: stop-gain, splice site, frameshift indels.
VEP impact: HIGH.

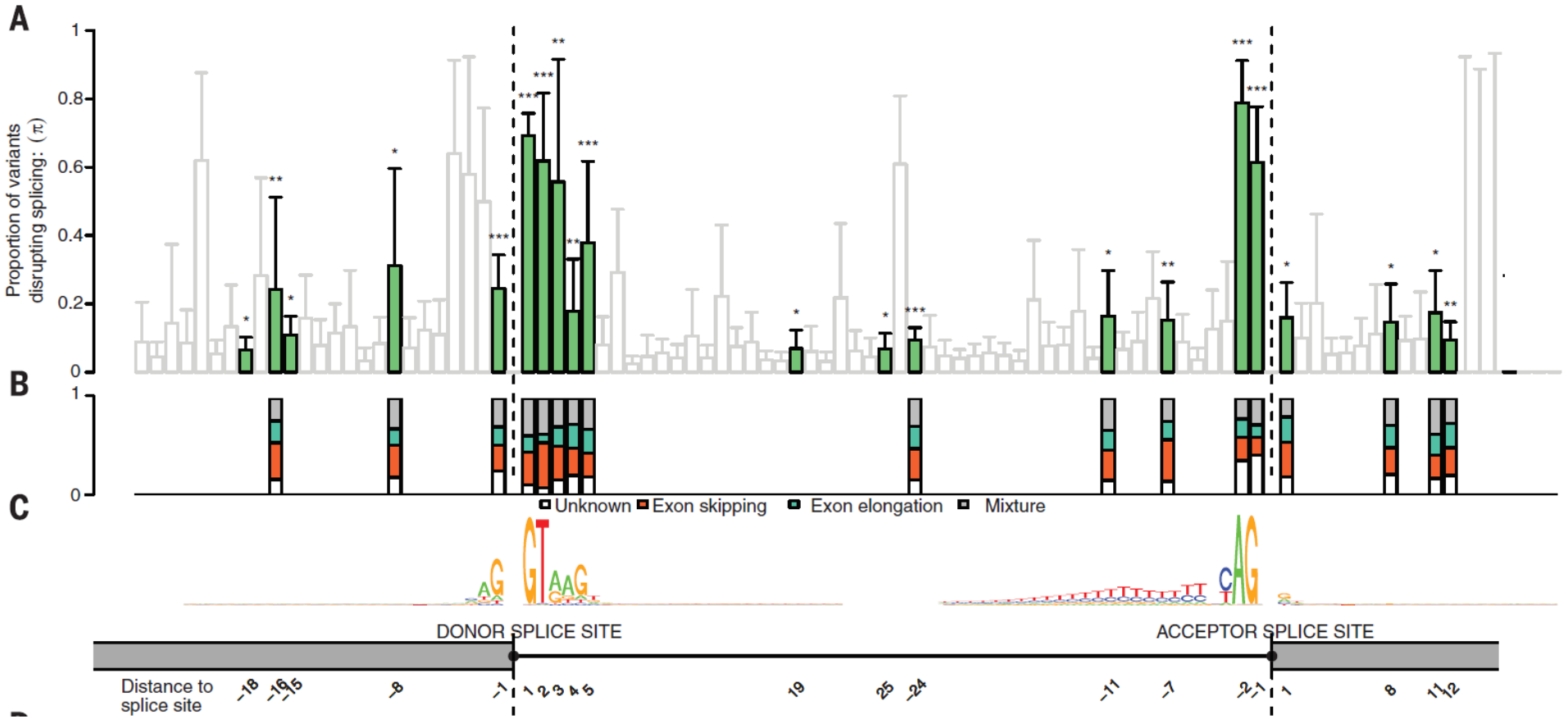
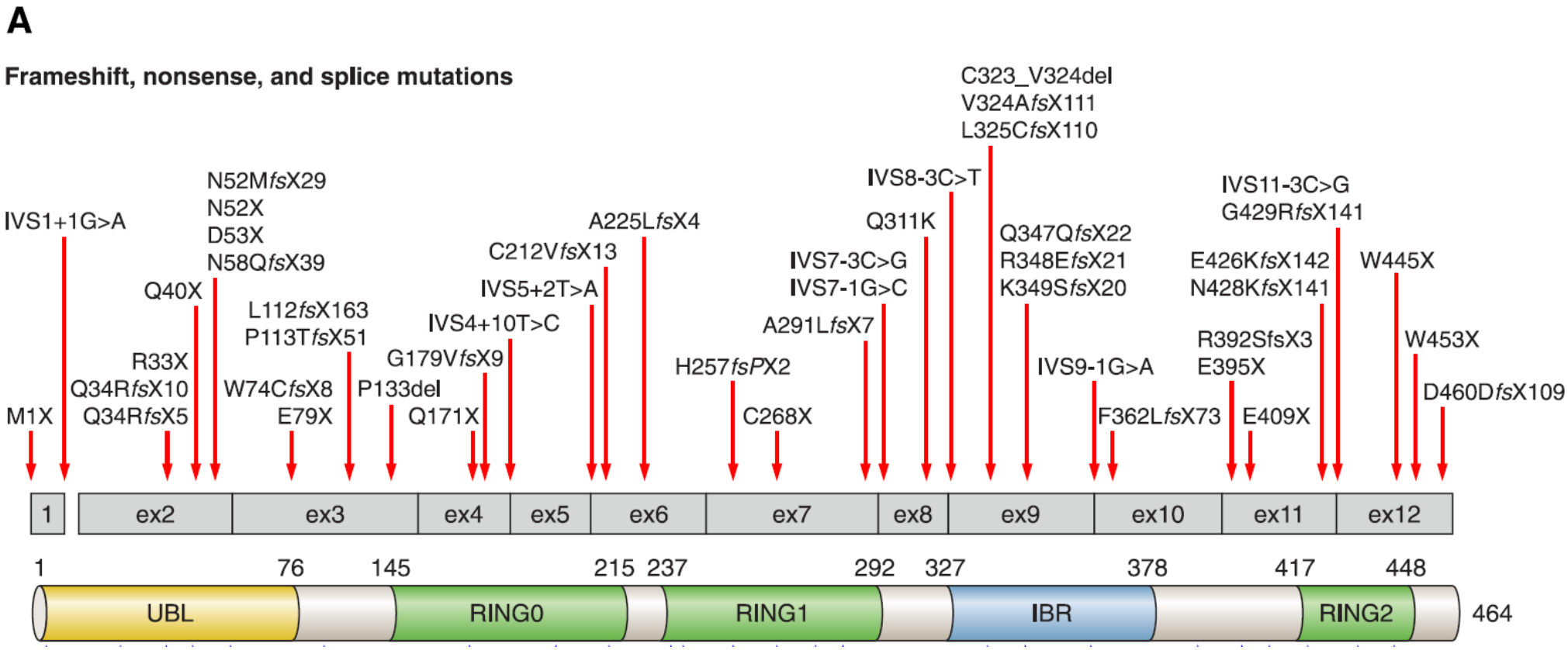


Fig. 3. Splicing disruption. (A) Proportion of variants disrupting splicing at each distance +/-25 bp from donor and acceptor site (B) Classification of splice disruption events: exon skipping, exon elongation and mixture (C) Diagram of donor and acceptor splice junctions and sequence logo of represented sequences.

Rivas (2015) *Science*

Examples of PTV impact



Mutations in the Parkin RBR E3 Ubiquitin Protein Ligase *PARK2* are the most frequent known cause of early-onset (40–50 yr) Parkinson's disease. PD is the second most common neurodegenerative disorder, after Alzheimer's disease, with prevalence in industrialized countries ~0.3%.

Corti (2011) *Physiol Rev*

Examples of PTV impact

1. Narasimhan VM, Xue Y, Tyler-Smith C. (2016) Human Knockout Carriers: Dead, Diseased, Healthy, or Improved? *Trends Mol Med* 22:341-351.

- A knockout of the immune gene *IRF7* was shown to confer **susceptibility to flu viruses**, leading to life-threatening influenza in an otherwise healthy child (Ciancanelli 2015 *Science*)
- Instances where a naturally-occurring **LoF variant proves beneficial to health**. These discoveries have stimulated drug development:
  - lowering LDL levels: *PCSK9*
  - decreasing susceptibility to HIV: *CCR5*
  - increasing endurance: *ACTN3*
  - increasing sepsis resistance: *CASP12*
  - reduced triglyceride levels in humans: *APOC3*

2. DeBoever, C., Tanigawa, Y., Lindholm, M.E., et al. (2018). Medical relevance of protein-truncating variants across 337,205 individuals in the UK Biobank study. *Nat Commun* 9, 1–10.

- 18,228 PTVs × 135 phenotypes; find **27 associations between medical phenotypes and PTVs** in genes outside the MHC

Examples of PTV impact

1. The stop-gain variant in *GNAS* (MIM:139320) is present in the highly variable **first exon** of the gene and is likely to result in nonsense-mediated RNA decay; in contrast, pathogenic *GNAS* variants that cause Albright hereditary osteodystrophy (MIM:103580) are located in **later**, highly constrained exons.

2. Similarly, the stop-gain variant in *TGIF1* (MIM:602630) is located in the **first exon**, where multiple PTVs in gnomAD are also located, but *TGIF1* pathogenic variants causing holoprosencephaly are located in the **final exons**, where they affect DNA binding affinity.

3. Finally, a frameshift deletion in *HIST1H1E* (MIM:142220) is located near **the start** of the single exon of this gene; however, pathogenic *HIST1H1E* frameshift deletions that cause child overgrowth and intellectual disability are located near **the end** of the exon, where they result in a truncated histone protein with lower net charge that is less effective at binding DNA.

We believe that these three rare PTVs are benign because of their locations, despite the fact that they occur in genes that cause dominant DD via haploinsufficiency.

Wright (2019) *Am J Hum Genet*

Are PTVs actually LoFs?

Lek (2016) *Nature*, ExAC paper, ~60,000 individuals:
– 13.2 expected pLoF variants per gene, 62.8% of genes have >10 pLoF variants on the canonical transcript
– Each individual harbors ~85 heterozygous and ~34 homozygous PTVs

Sulem (2015) *Nat Genet*, ~101,000 Icelanders: // founder population
– 7.7% individuals have 1 gene completely knocked out by loss-of-function variants with a MAF under 2%
– 553 were predicted to have >1 gene completely knocked out
– 1,171 of the 19,135 RefSeq genes (6.1%) were completely knocked out

Saleheen (2017) Nature, ~10,000 Pakistanis // consanguineous
– 1,317 distinct genes were predicted to be inactivated owing to homozygous pLoF mutations
– 17.5% participants had at least one gene knocked out by a homozygous pLoF mutation, ~18% of them >1 gene knocked out

Missense variant, classic example

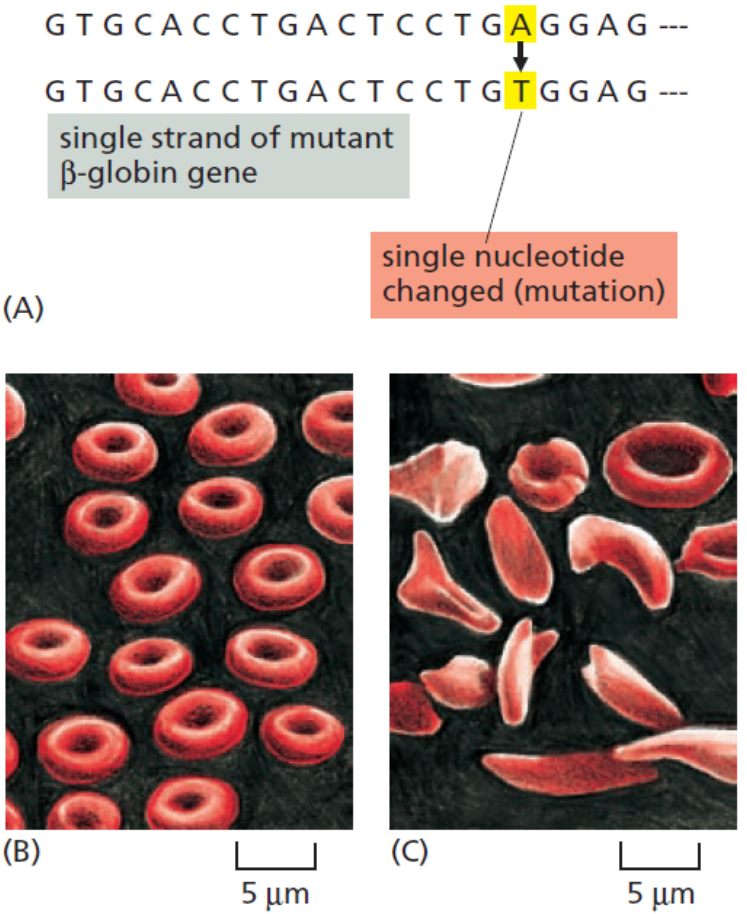
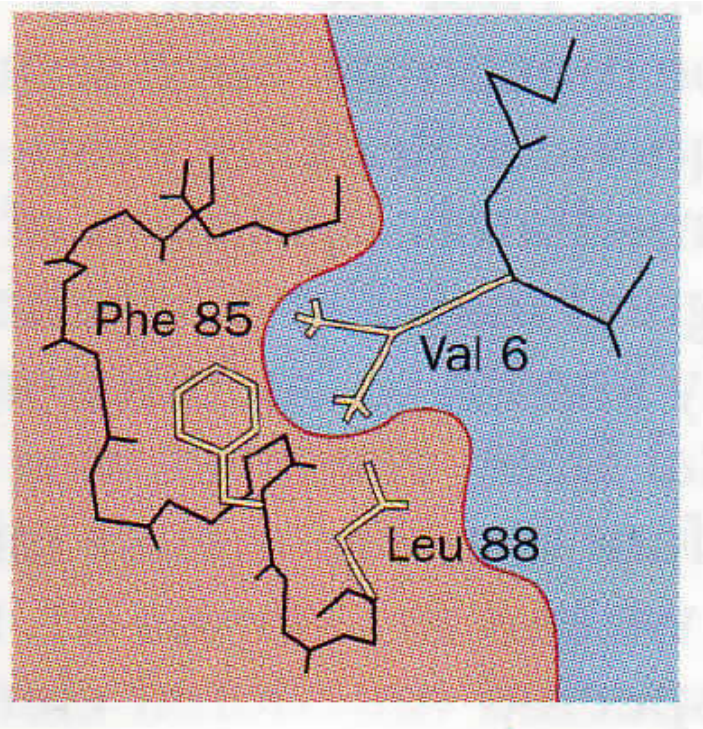


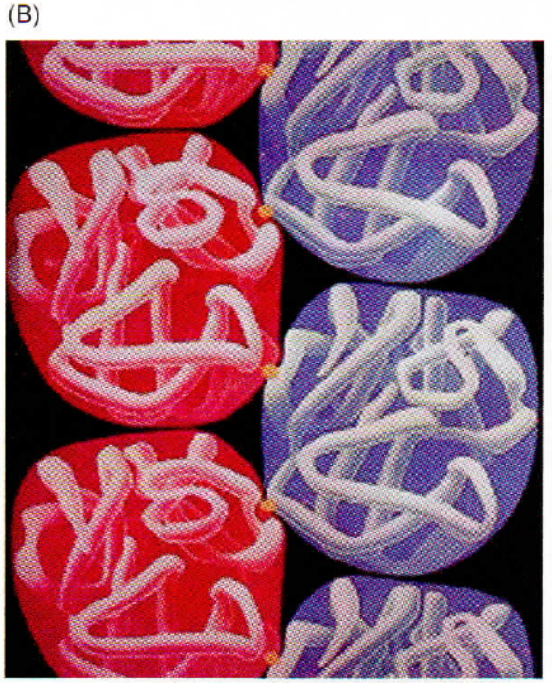
Figure 6–19 **A single nucleotide change causes the disease sickle-cell anemia.** (A) β-globin is one of the two types of subunit that form hemoglobin (see Figure 4–20). A single nucleotide change (mutation) in the β-globin gene produces a β-globin subunit that differs from normal β-globin only by a change from glutamic acid to valine at the sixth amino acid position. (Only a small portion of the gene is shown here; the β-globin subunit contains a total of 146 amino acids.) Humans carry two copies of each gene (one inherited from each parent); a sickle-cell mutation in one of the two β-globin genes generally causes no harm to the individual, as it is compensated for by the normal gene. However, an individual who inherits two copies of the mutant β-globin gene displays the symptoms of sickle-cell anemia. Normal red blood cells are shown in (B), and those from an individual suffering from sickle-cell anemia in (C). Although sickle-cell anemia can be a life-threatening disease, the mutation responsible can also be beneficial. People with the disease, or those who carry one normal gene and one sickle-cell gene, are more resistant to malaria than unaffected individuals, because the parasite that causes malaria grows poorly in red blood cells that contain the sickle-cell form of hemoglobin.

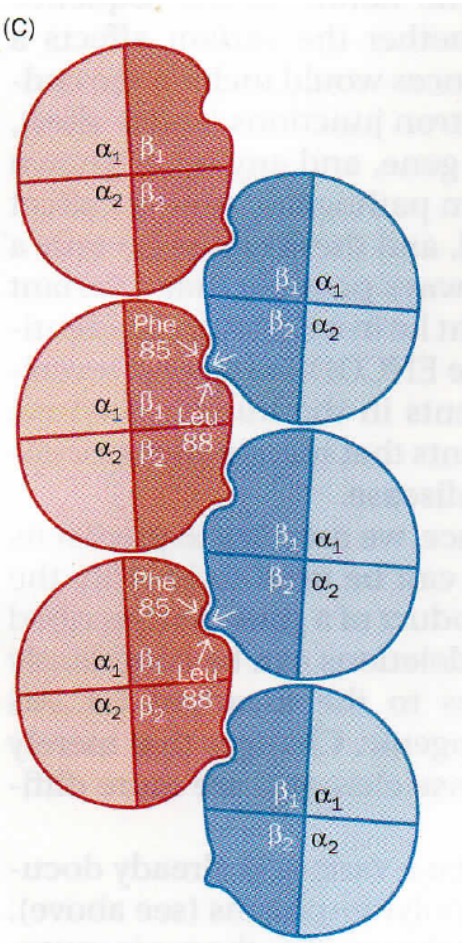
HBB.Glu7Val **Sickle cell anemia** [MIM:603903]: Characterized by abnormally shaped red cells resulting in **chronic anemia and periodic episodes of pain, serious infections and damage to vital organs**. Normal red blood cells are round and flexible and flow easily through blood vessels, but in sickle cell anemia, the abnormal hemoglobin (called Hb S) causes red blood cells to become stiff. They are C-shaped and resembles a sickle. These stiffer red blood cells can led to microvascular occlusion thus cutting off the blood supply to nearby tissues // www.genecards.org

Alberts - *Essential Cell Biology*

Missense variant, classic example







The sickle cell mutation. An A>T mutation in the β-globin *(HBB)* gene causes an amino acid change in the β-globin protein. The mutation replaces glutamic acid, a hydrophilic charged amino acid, with valine, a hydrophobic nonpolar amino acid. This change on the surface of the globin protein allows adhesive interactions between hemoglobin molecules.

s held in the PDB are shown here at a magnification of about
ented as a small sphere. The enormous range of molecular sizes is
le (H_2O) with only three atoms (shown at the left) to the ribosomal
toms.

(Complex III) 1bgy
33. Succinate Dehydrogenase (Complex II) 1nek
34. NADH-Quinone Oxidoreductase (Complex I) 3m9s, 3rko
35. ATP Synthase 1e79, 1c17, 1l2p, 2a7u
36. Myoglobin 1mbd
37. Hemoglobin 4hhb

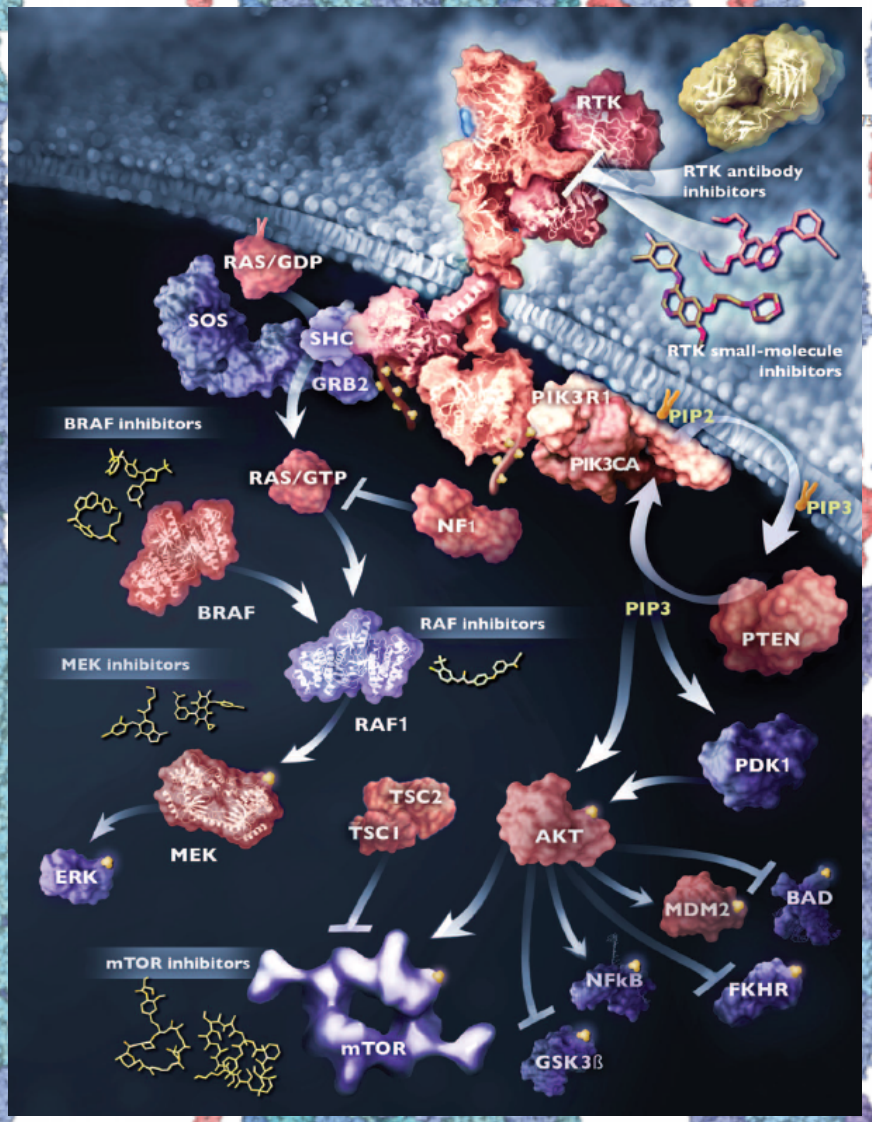
Storage: containing nutrients for future consumption

38. Ferritin 1hrs

Enzymes: cutting and joining the molecules of life

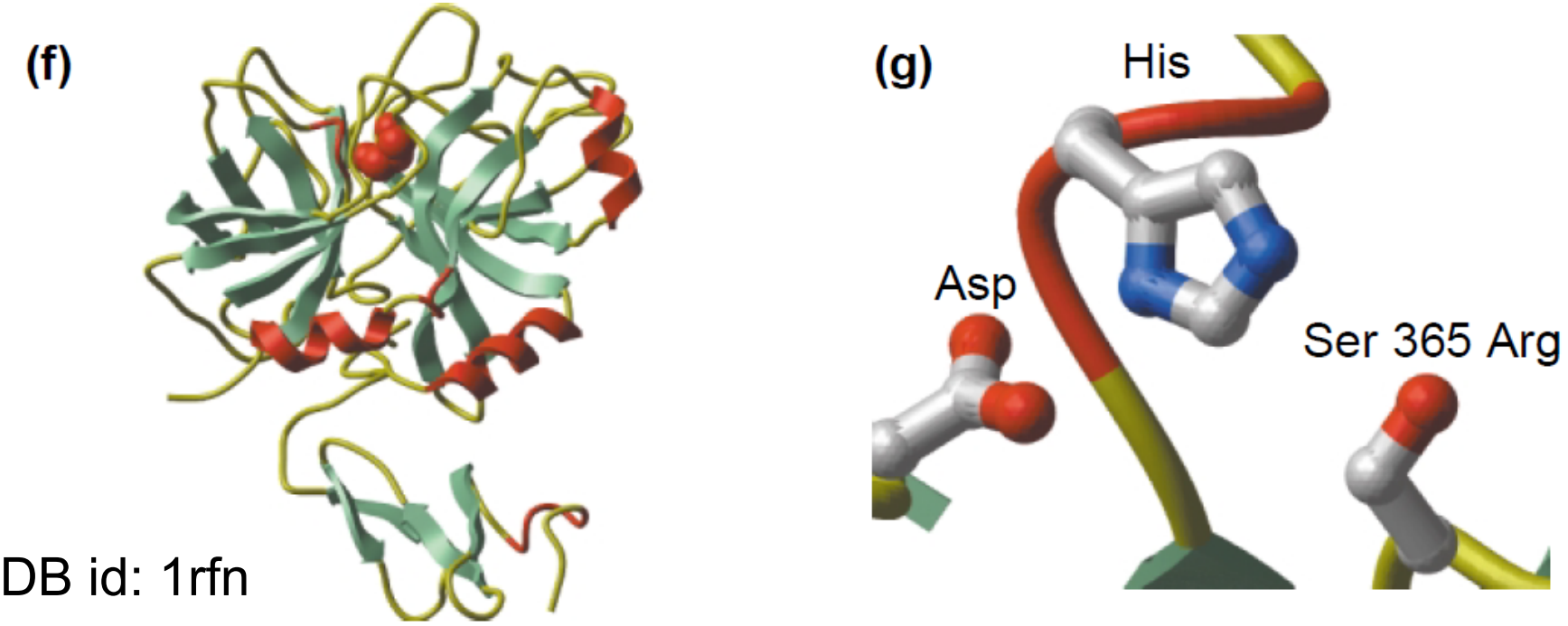
39. Fatty Acid Synthase 2uvb, 2uvc
40. RuBisCo: Ribulose Bisphosphate Carboxylase/Oxygenase 1rcx
41. Green Fluorescent Protein 1gfl
42. Luciferase 2d1s
43. Glutamine Synthetase 2gls
44. Alcohol Dehydrogenase 2ohx
45. Dihydrofolate Reductase 1dhf
46. Nitrogenase 1n2c
47. Leucine Aminopeptidase 1lap
48. beta-Lactamase 4blm
49. Catalase 1qqw
50. Thymidylate Synthase 2tsc
51. Tryptophan Synthase 1wsy
52. Aspartate Carbamoyltransferase 4at1
53. Hexokinase 1dgk
54. Phosphoglucose Isomerase 1hox
55. Phosphofructokinase 4pfk
56. Aldolase 4ald
57. Triosephosphate Isomerase 2ypi
58. Glyceraldehyde-3-phosphate Dehydrogenase 3gpd
59. Phosphoglycerate Kinase 3pgk
60. Phospoglycerate Mutase 3pgm
61. Enolase 5enl
62. Pyruvate Kinase 1a3w

78. Prefoldin 1fxk
79. Chaperonin GroEL/ES 1aon
80. Proline cis/trans Isomerase 2cpl
81. Heat Shock Protein Hsp90 2cg9
82. Proteasome 4b4t
83. Ubiquitin 1ubq



Vogelstein (2013) *Science*
Protein Data Bank rcsb.org

Examples of missense disease variants



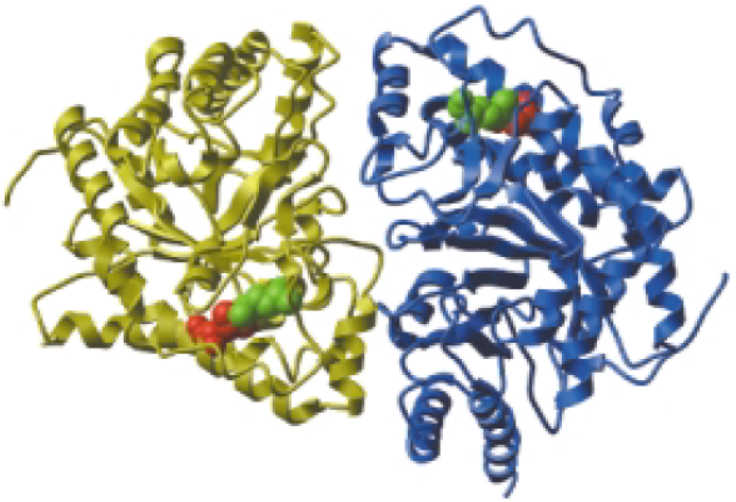
Factor IX *F9* is a serine protease with Ser-His-Asp catalytic triade that participates in the intrinsic pathway of blood coagulation by converting factor X to its active form Xa. Disease mutations in *F9* are associated with the X-linked recessive bleeding disorder haemophilia B (OMIM:306900).

Disruption of catalytic residues. Mutations of the catalytic serine residue to an arginine results in the loss of enzyme activity and a severe haemophilia phenotype.

Steward (2003) *Trends Genet*

Examples of missense disease variants

Introduction of buried charged residues:
Met165Arg ⟹ arginine sidechain cannot be accommodated in a hydrophobic pocket ⟹ no soluble protein.



PDB id: 1uro

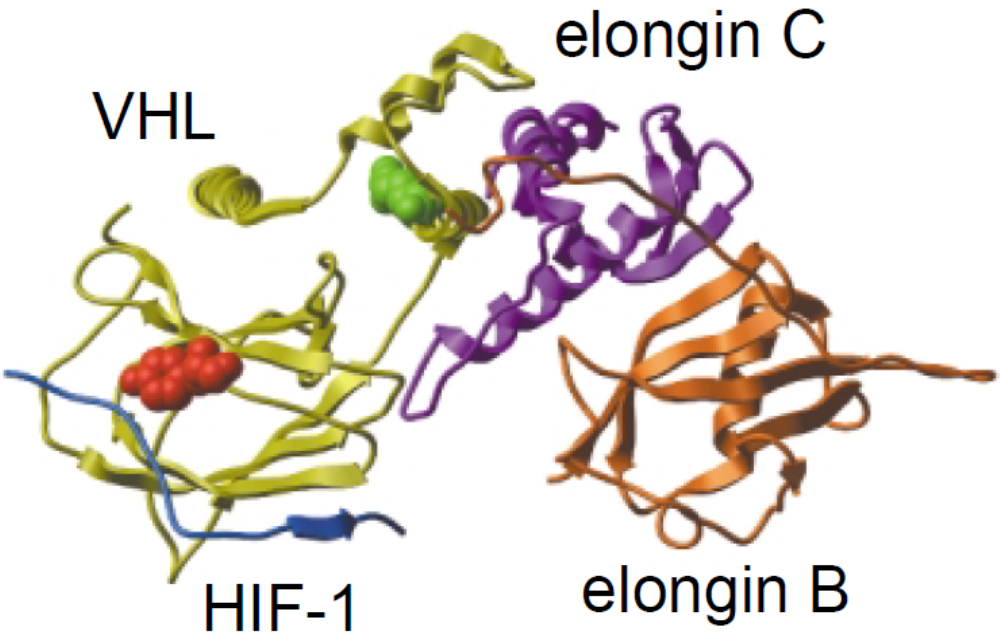
Size changes in the hydrophobic core:
Leu195Phe ⟹ rearrangement of surrounding side-chains ⟹ 30% of the wild-type activity.

Mutations in the uroporphyrinogen decarboxylase *UROD* are associated with Porphyria cutanea tarda (OMIM:176100), accumulation of uroporphyrins in the liver and plasma, leading to skin fragility and photosensitive dermatitis.

Steward (2003) *Trends Genet*

Examples of missense disease variants

Disruption of protein–protein interactions:
Tyr98His destroys binding between HIF and VHL ⟹ HIF not degraded ⟹ over-expression of angiogenic growth factors ⟹ local proliferation of blood vessels.



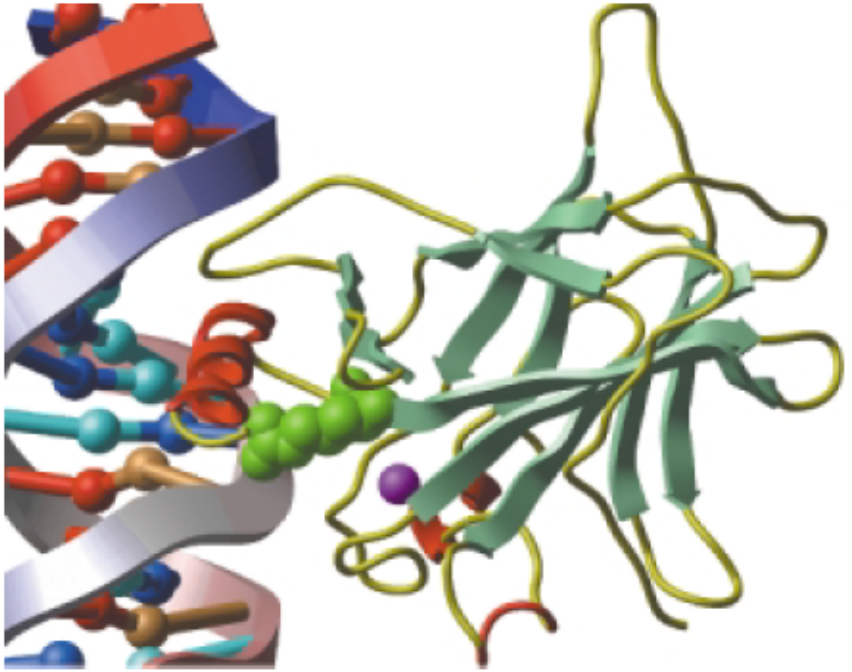
PDB id: 1lm8

Von Hippel-Lindau syndrome (OMIM:193300) is an inherited predisposition to a variety of cancers. Von Hippel-Lindau disease tumor suppressor *VHL* codes for a protein with two structural domains. The β-domain of VHL binds to hypoxia-inducible transcription factor HIF, ultimately leading to HIF degradation.

Steward (2003) *Trends Genet*

Examples of missense disease variants

Disruption of DNA binding
Arg273 contacts the DNA phosphate backbone with its charged side-chain. Arg273His is associated with low p53 DNA-binding and Li-Fraumeni syndrome.



PDB id: 1tsr

Li-Fraumeni syndrome (OMIM 191170), a predisposition to a broad spectrum of cancers at an early age. Cellular tumor antigen p53 (*TP53)* is a tumor suppressor in many tumor types, induces growth arrest or apoptosis. Three functional domains: an N-terminal transcription factor domain, a DNA-binding core domain, and a C-terminal homooligomerization domain.

Steward (2003) *Trends Genet*

Examples of missense disease variants

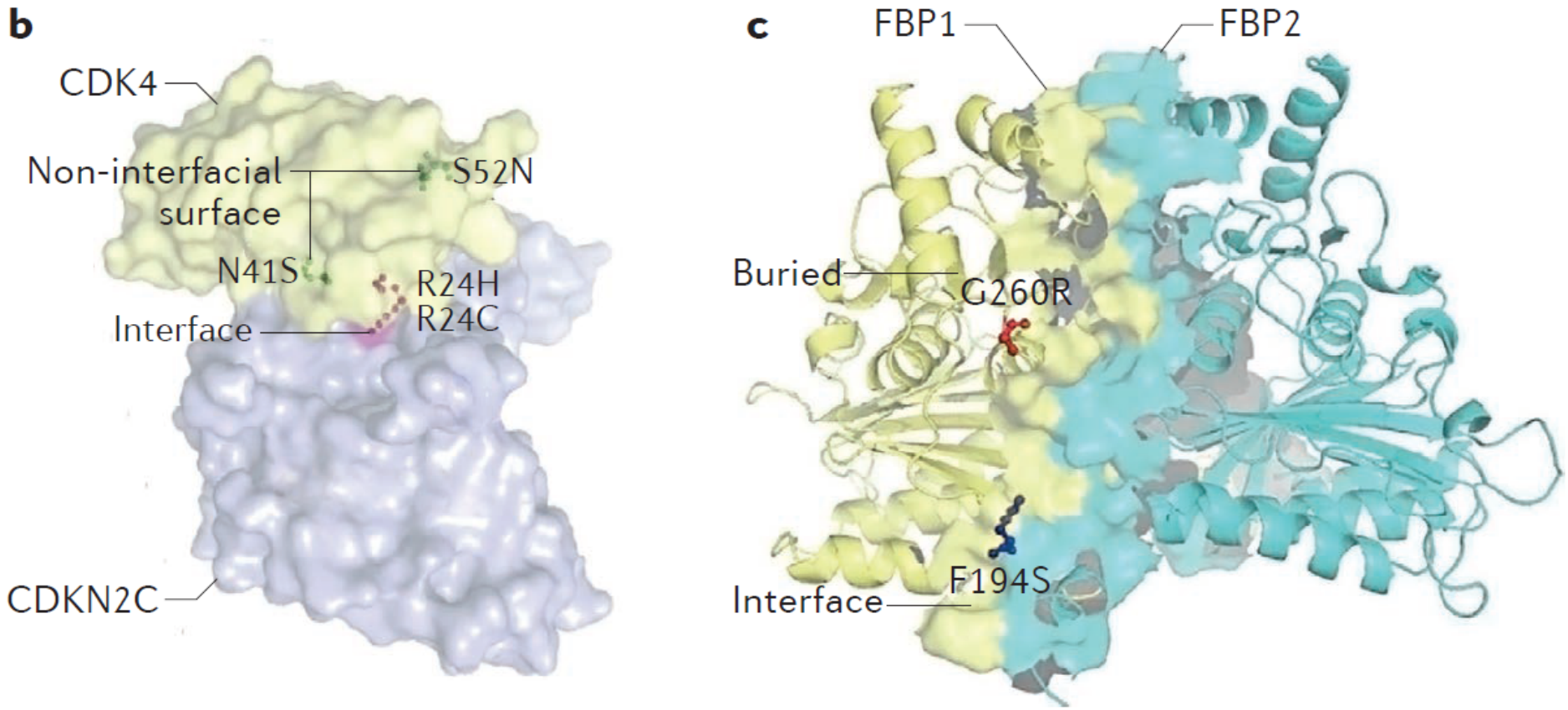
Fig. 5 Protein structural variation. Organization of the descriptive VariO terms, which facilitate very detailed annotation of observed effects

VariO: Variant effect on protein…

- Dynamics
- Quaternary structure
- Amino acid size
- Folding rate
- Interactions
- Post-translational modification
- Secondary structural element
- Fold
- Epigenetic modification
- Abundance
- Accessibility
- Activity
- Charge
- Degradation
- Solubility
- Stability
- Subcellular localization
- …

www.variationontology.org Vihinen (2015) *Human Genet*

Missense disease mutations: stability or PPI?



b | Locations of residues affected by mutations are highlighted on the cyclin-dependent kinase 4 (CDK4) structure based on homology modelling (PDB: 1bi7). CDKN2C, CDK inhibitor 2C. **c** | Locations of residues affected by mutations are highlighted on the fructose bisphosphatase 1 (FBP1) structure (PDB: 1fpi).

Yi (2017) *Nat Rev Genet*

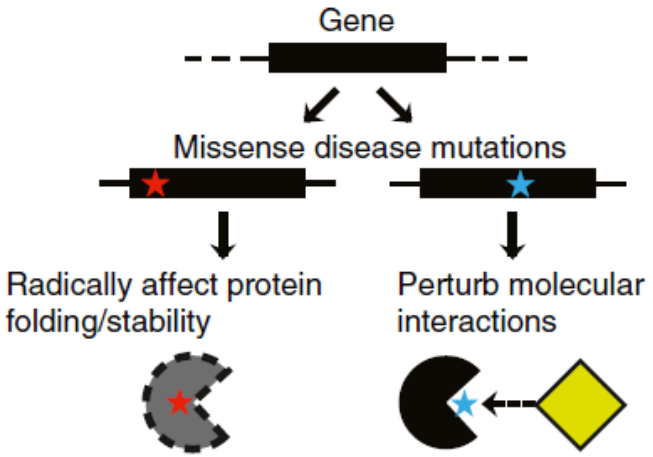
Missense disease mutations: stability or PPI?

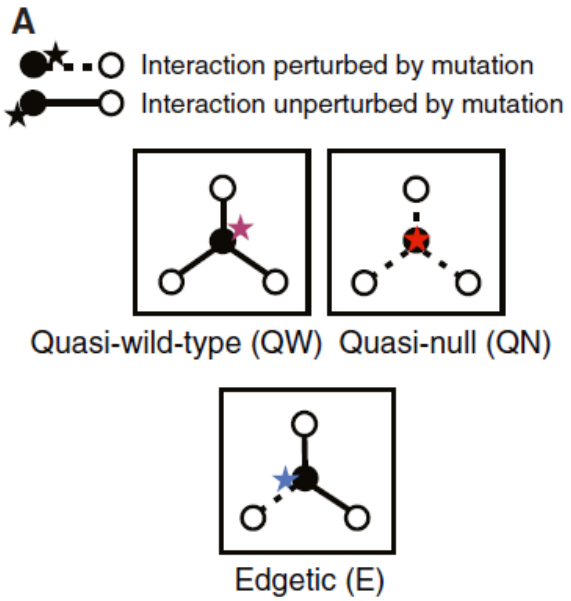
Table 1 | **Human diseases caused by defects in protein folding, stability and aggregation**

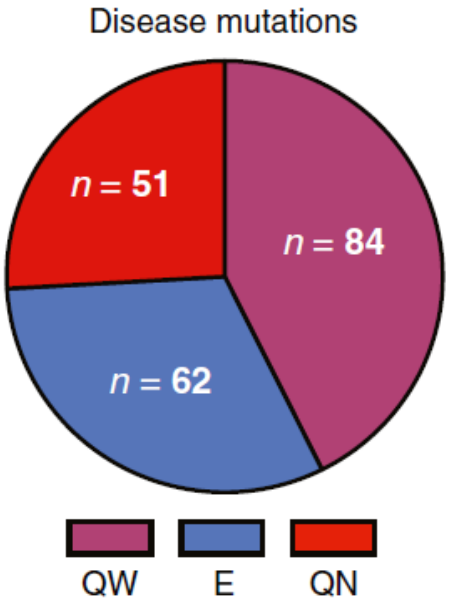
Disease	Protein affected	Description	References
Cystic fibrosis	Cystic fibrosis transmembrane conductance regulator (CFTR)	The ΔPhe508 mutant has wild-type activity, but impaired folding in the endoplasmic reticulum leads to degradation.	97
α1 Antitrypsin deficiency	α1 Antitrypsin (also known as SERPINA1)	80% of Glu342Lys mutants misfold and are degraded. Pathology is due to aggregation in patients with a reduced degradation rate.	97
SCAD deficiency	Short-chain acyl-CoA dehydrogenase (SCAD)	Impaired folding of Arg22Trp mutants leads to rapid degradation.	98
Alzheimer disease	Presenilin, γ-secretase	Mutations cause incorrect cleavage by the γ-secretase protease to produce the amyloid β-peptide; this aggregates into extracellular amyloid plaques.	99,100
Parkinson disease	α-Synuclein	Oxidative damage causes misfolding and aggregation. Hereditary forms are linked to deficiency in ubiquitin-mediated degradation.	101
Huntington disease	Huntingtin	CAG expansions in the Huntingtin gene lead to an abundance of polyglutamine fragments that aggregate and associate non-specifically with other cellular proteins.	101,102
Sickle cell anaemia	Haemoglobin	The Glu6Val mutation leads to aggregation in red blood cells.	103

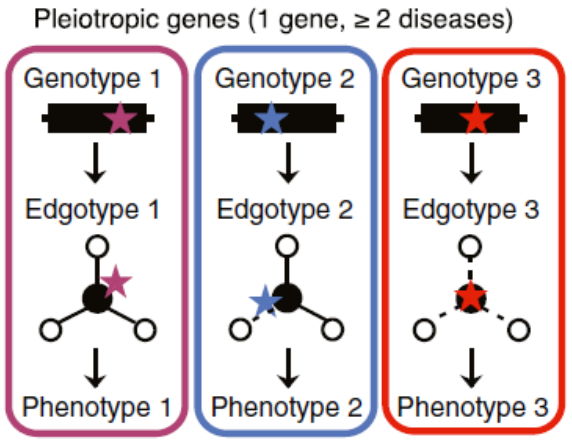
De Pristo (2005) *Nat Rev Genet*

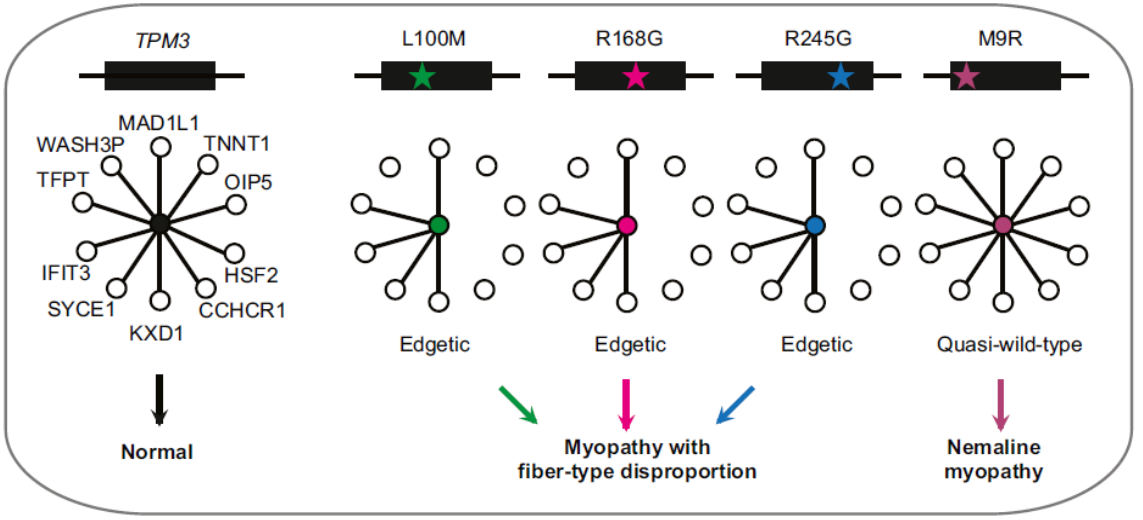
Missense disease mutations: stability or PPI?











The effects of missense disease mutations on molecular interactions could range from no apparent detectable change in interactions (**quasi-WT**), to specific loss of some interactions (**edgetic**), to an apparent complete loss of interactions (**quasinull**)

Sahni (2015) *Cell*