NEAR EAST UNIVERSITY

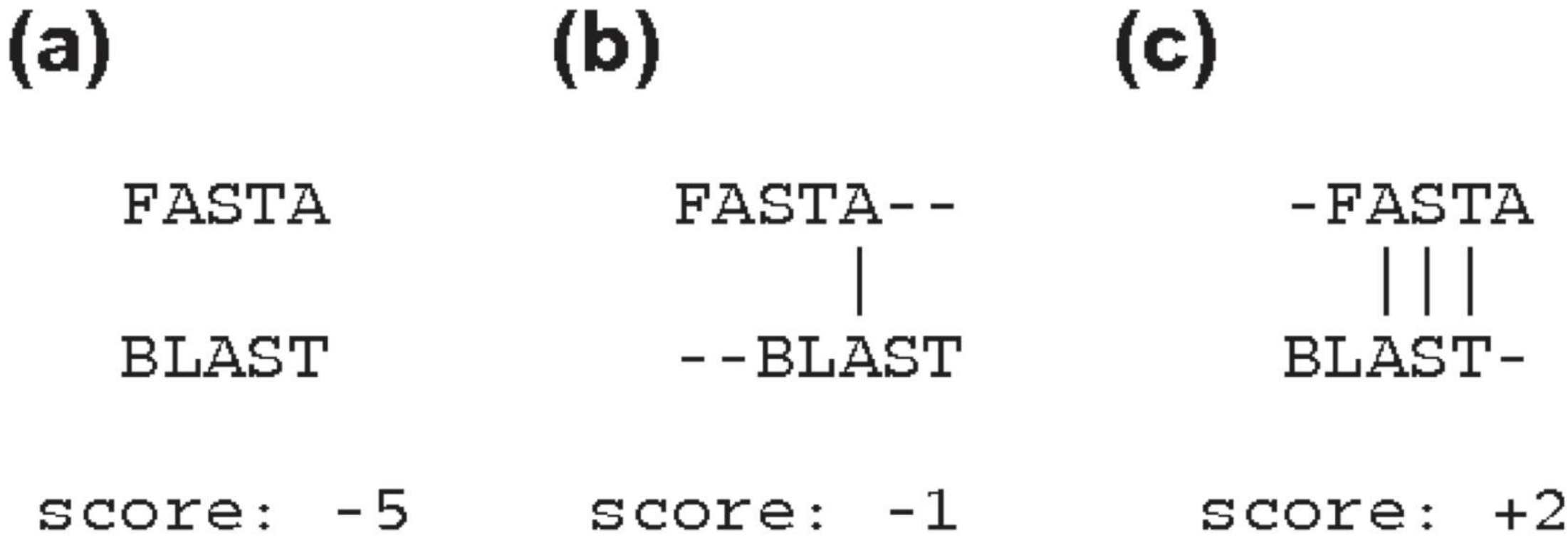
37 Years in Education

NEAR EAST UNIVERSITY
EXPERIMENTAL HEALTH
SCIENCES RESEARCH CENTER

Bioinformatics Fall School:
Applications in Molecular Basic and Clinical Sciences

A DNA Variant Scoring for Pathogenicity Assesments in Mendelian and Rare Disorders: Using Clinical Cases

Mahmut Çerkez Ergören


(a)
FASTA
BLAST
score: -5
(b)
FASTA--
|
--BLAST
score: -1
(c)
-FASTA
|||
BLAST-
score: +2

The Greatest Puzzle

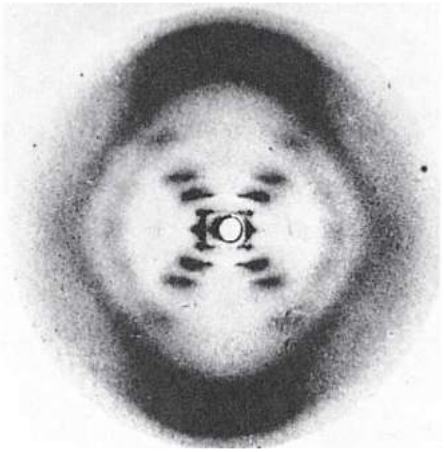
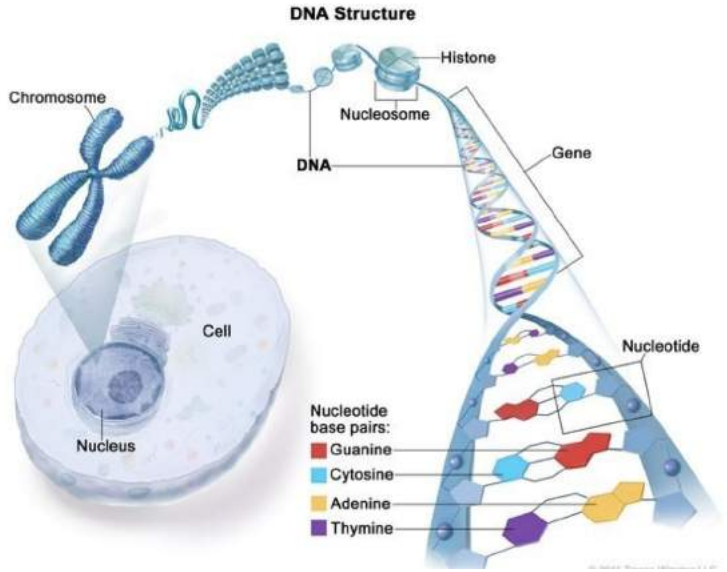
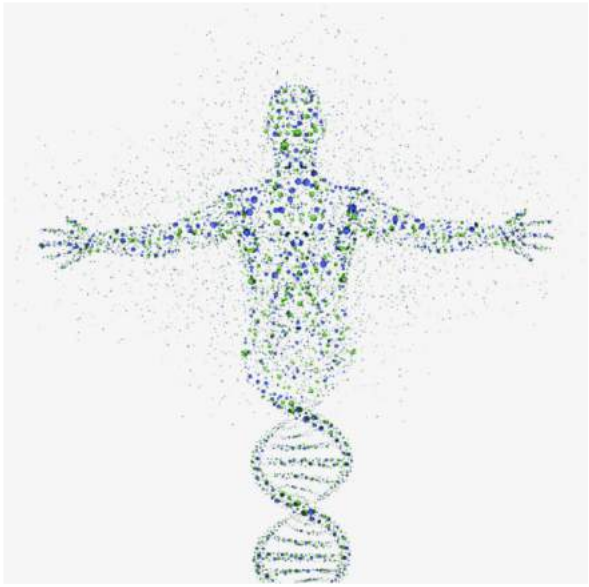


Photo 51
(1952)

Rosalind Franklin
(1920-1958)

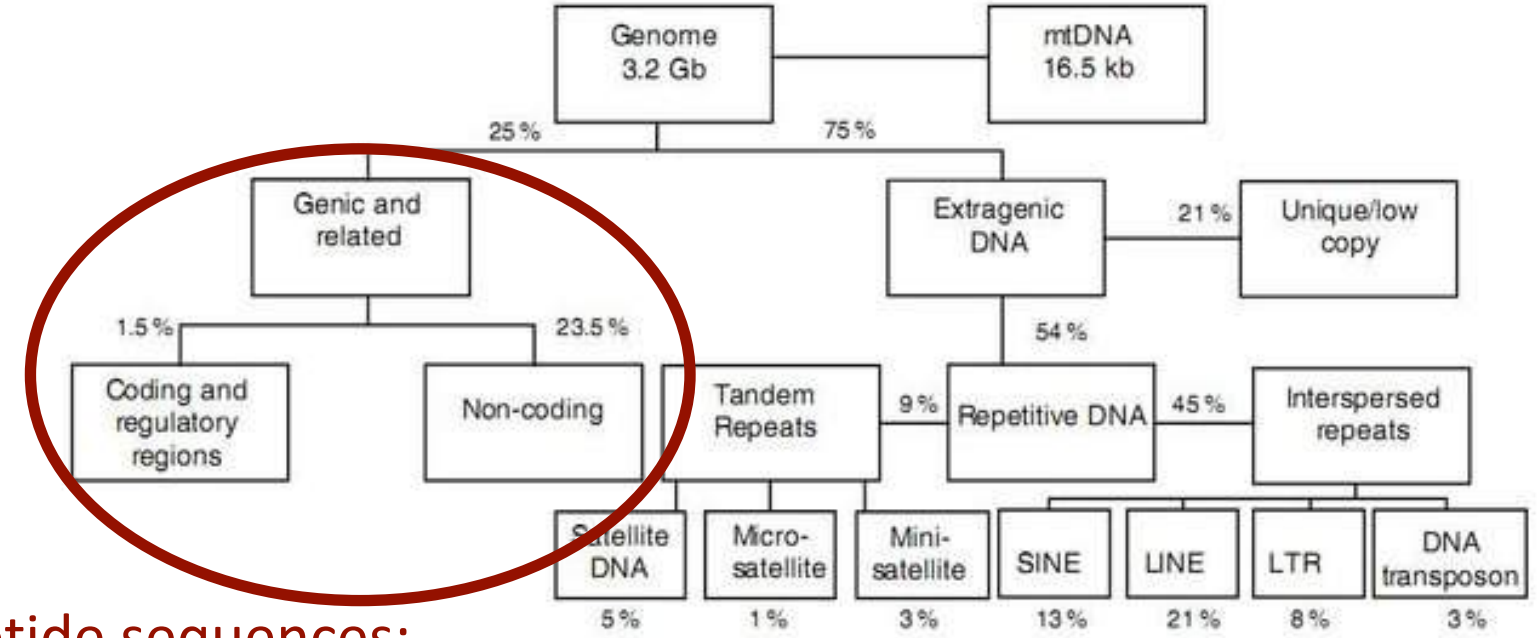




23 pairs of chromosome
3 billion bp
~20.000 genes

IT WAS ONLY A BEGINNING!

Organization of the Human Genome



Changes in nucleotide sequences:

Disorders/ Diseases

Diversity

1990-2003

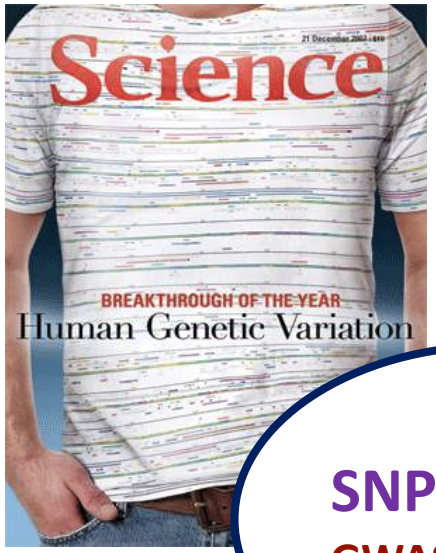
Banana
60.0%

Chimp.
96.5%

Homo sapiens
99.9%

Homo neanderthalis
99.7%

2004-2007

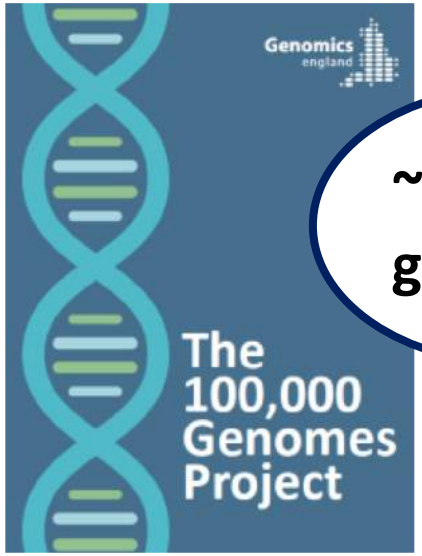


SNP
GWAS was born

2008-2015

26 population
2504 individuals
88 million variants

2015-present



~92.000 genomes

What numbers tell us?

3 million SNPs among people

20.000 SNP in genome coding sequence

1000 stop mutations

THANKS to diploid nature!

Human *vs* Human

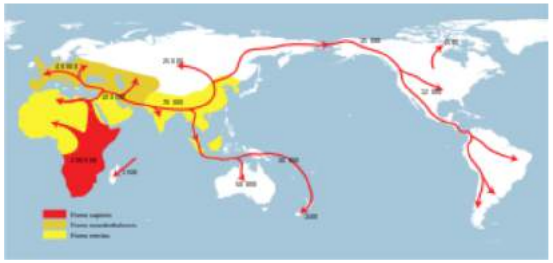
99.9%

Why we care about genetic variations?

1. Genetic variations underlie **phenotypic differences** among different individuals

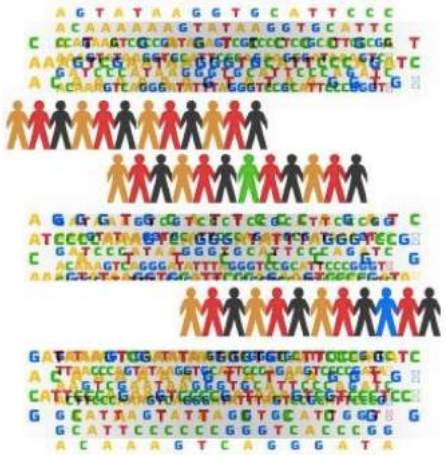
2. Genetic variations determine our **predisposition to complex diseases** and **responses to drugs** and **environmental factors.**

3. Genetic variations reveals clues of **ancestral human migration history.**



Types of variations between human genomes

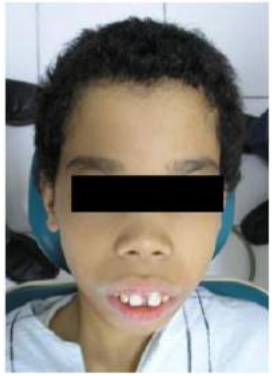
Most of many differences between individuals human genomes seem to have **no effect**.



Other differences do affect the **phenotype**, **producing the normal range of genetically determined variants in body build, pigmentation, metabolism** and so on, that make each us individual.

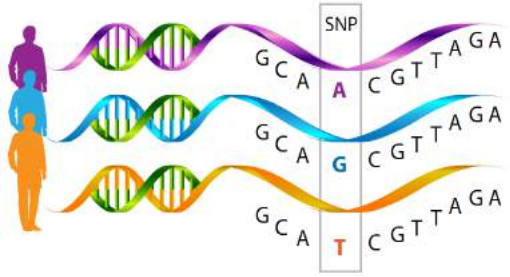
Some variants are **pathogenic**:

They either **cause** disease or make their bearer **susceptible** to a disease



Main types of genetic variations

1. Single Nucleotide Polymorphisms (SNPs)



- 90% of human genetic variations
- Majority of SNPs do NOT directly or significantly contribute to any phenotype

2. Insertion or deletion of one or more nucleotide(s)

Tandem repeat polymorphisms

- Are genomic regions consisting of variable lenght of sequence motifs repeating in tandem with variable copy number.
- Used as genetic markers for DNA fingerprinting (forensic & parental testing)
- Many cause genetic diseases

Microsatellites (Short tandem repeats) – repeat unit 1-6 bases long

Minisatellites – repeat unit 11-100 bases long

Insertion/Deletion polymorphisms

Often resulted from localized rearragement between homologous tandem repeats.

3. Gross chromosomal aberrations

- Deletions, inversions or translocation of large DNA fragments
- Often causing serious genetic diseases

ttggaaaagcaacccctgccttgaagccaggatgatggtatctgcagcagttgccaacacaagagaaggatccatagttcatcatttaaaaaagaaaacaaaatagaaaaaggaaaacta
tttctgagcataagaagttgtagggtaagtctttaagaaggtgacaatttctgccaatcaggatttcaaagctcttgctttgacaattttggtctttcagaatactataaatataacctatattata
atttcataaagtctgtgcattttctttgacccaggatatttgcaaaagacatattcaaacttccgcagaacactttatttcacatatacatgcctcttatatcagggatgtgaaacagggtcttgaa
aactgtctaaatctaaaacaatgctaatgcaggtttaaatttaataaaataaaatccaaaatctaacagccaagtcaaatctgcatgttttaacatttaaaatattttaaagacgtcttttccca
ggattcaacatgtgaaatctttctcagggatacacgtgtgcctagatcctcattgctttagtttttacagaggaatgaatataaaaagaaaatacttaaattttatccctcttacctctataatca
tacataggcataatttttaacctaggctccagatagccatagaagaaccaaacactttctgcgtgtgtgagaataatcagagtgagatttttcacaagtacctgatgagggttgagacaggt
agaaaaagtgagagatctctatttatttagcaataatagagaaagcatttaagagaataaagcaatggaaataagaaatttgtaaatttccttctgataactagaaatagaggatccagtttct
tttggttaacctaaattttatttcattttattgttttatttttatttttatttttatttttgtgtaatcgtagtttcagagtgttagagctgaaaggaagaagtaggagaaacatgcaaagtaaaagtat
aacactttccttactaaaccgacatgggtttccaggtaggggcaggattcaggatgactgacagggcccttagggaacactgagaccctacgctgacctcataaatgcttgctacctttgctgt
tttaattacatcttttaatagcaggaagcagaactctgcacttcaaaagtttttcctcacctgaggagttaatttagtacaaggggaaaaagtacaggggggatgggagaaaggcgatcacgtt
gggaagctatagagaaagaagagtaaattttagtaaaggaggtttaaacaaacaaaatataaagagaaataggaacttgaatcaaggaaatgattttaaaacgcagtattcttagtggact
agaggaaaaaaataatctgagccaagtagaagacctttcccctcctacccctactttctaagtcacagaggcttttgttcccccagacactcttgcagattagtccaggcagaaacagttag
atgtccccagttaacctcctatttgacaccactgattaccccattgatagtcacactttgggttgtaagtgacttttatttatttgtatttttgactgcattaagaggtctctagtttttatctcttgttt
cccaaaacctaataagtaactaatgcacagagcacattgatttgtatttattctatttttagacataatttattagcatgcatgagcaaattaagaaaaacaacaacaaatgaatgcatatatat
gtatatgtatgtgtgtatatatacacacatatatatatatttttttctttttcttaccagaaggttttaatccaaataaggagaagatatgcttagaaccgaggtagagttttcatccattctgtcctg
taagtattttgcatattctggagacgcaggaagagatccatctacatatcccaaagctgaattatggtagacaaaactcttccacttttagtgcatcaacttcttatttgtgtaataagaaaattg
ggaaaacgatcttcaatatgcttaccaagctgtgattccaaatattacgtaaatacacttgcaaaggaggatgtttttagtagcaatttgtactgatggtatggggccaagagatatatcttaga
gggagggctgagggtttgaagtccaactcctaagccagtgccagaagagccaaggacaggtacggctgtcatcacttagacctcaccctgtggagccacaccctagggttggccaatctact
cccaggagcagggagggcaggagccagggctgggcataaaagtcagggcagagccatctattgcttacatttgcttctgacacaactgtgttcactagcaacctcaaacagacaccatggt
gcatctgactcctgaggagaagtctgccgttactgccctgtggggcaaggtgaacgtggatgaagttggtggtgaggccctgggcaggttggtatcaaggttacaagacaggtttaaggaga
ccaatagaaactgggcatgtggagacagagaagactcttgggtttctgataggcactgactctctctgcctattggtctattttcccacccttaggctgctggtggtctacccttggacccagag
gttctttgagtcctttggggatctgtccactcctgatgctgttatgggcaaccctaaggtgaaggctcatggcaagaaagtgctcggtgcctttagtgatggcctggctcacctggacaacctca
agggcacctttgccacactgagtgagctgcactgtgacaagctgcacgtggatcctgagaacttcagggtgagtctatgggacgcttgatgttttctttcccccttcttttctatggttaagttcatg
tcataggaaggggataagtaacagggtacagtttagaatgggaaacagacgaatgattgcatcagtgtggaagtctcaggatcgttttagtttcttttatttgctgttcataacaattgttttcttt
tgtttaattcttgctttctttttttttcttctccgcaattttactattatacttaatgccttaacattgtgtataacaaaaggaaatatctctgagatacattaagtaacttaaaaaaaaactttacac
agtctgcctagtacattactatttggaatatatgtgtgcttatttgcatattcataatctccctactttatttcttttatttttaattgatacataatcattatacatatttatgggttaaagtgtaatgtt
ttaatatgtgtacacatattgaccaaatcagggtaattttgcatttgtaattttaaaaaatgctttcttcttttaatatactttttgtttatcttatttctaatactttccctaatctctttctttcagggc
aataatgatacaatgtatcatgcctctttgcaccattctaaagaataacagtgataatttctgggttaaggcaatagcaatatctctgcatataaatatttctgcatataaattgtaactgatgta

Codon 17 (A>T)

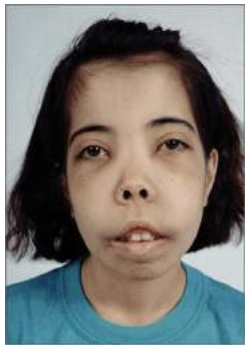
Why exploring gene variations are important?

Allowing researchers to **trace** the pattern of **disease-causing** DNA sequence variants in the population.

Nearly **half** of the genes causing 7,000 **rare monogenic disorders** were identified.

More than 50% of individuals with rare genetic disorders are **yet** to be **diagnosed** and **treated** in order to **improve** their **life quality**.

Clinical features can be used to distinguish one condition from another; however,

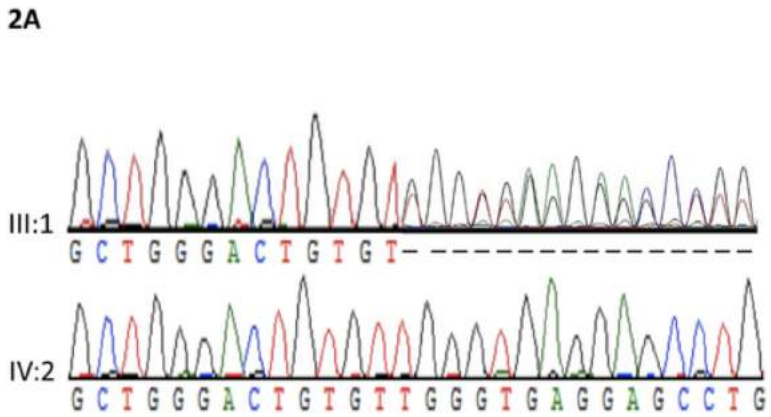


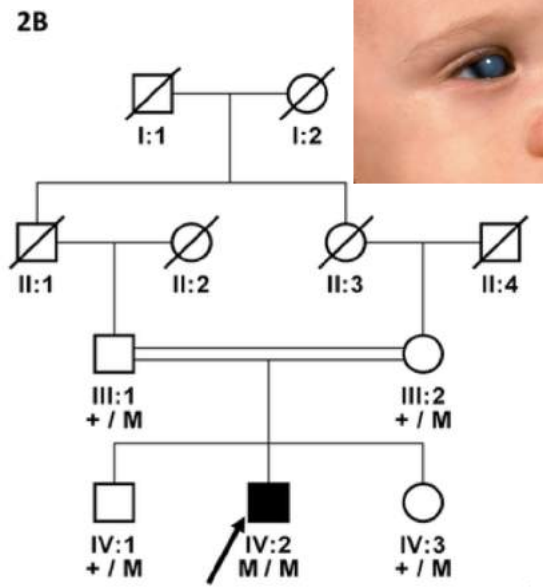
1. **Some phenotypes are associated with a single gene,**

2. **Many are associated with multiple genes,**

3. **Some clinical features overlap with several other genetic conditions,**

4. **The pathogenicity of detected variants are not known.**





I detected a novel variant
Is every variant pathogenic?

high-throughput next generation sequencing evolution!

Genotyping

Single gene to Gene panels

WES

WGS

Transcriptome

Epigenome

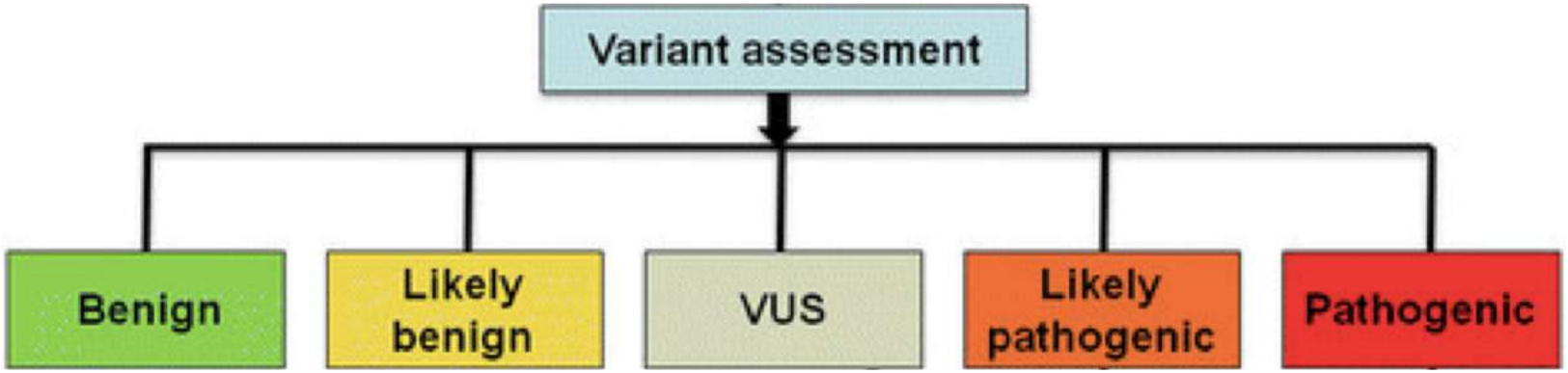
Clinical labs are performing increased catalogue of genetics testing for genetic disorders

Mutation or Polymorphism Tripe

A **mutation** is defined as a **permanent** change in the nucleotide sequence.

A **polymorphism** is defined as a **variant** with a frequency **above 1%.**

Medical genetics: Leading to confusion! **Pathogenic? Bening?**



These modifiers may not address all human phenotypes (Mendelian disease)

Nomenclature

A standard gene variant nomenclature versioned by the HGVS
(http://www.hgvs.org/mutnomen)

Tools to provide correct HGVS nomenclature for describing variants (https://mutalyzer.nl)

The reference sequence should be complete and derived from
a. the NCBI RefSeq database with the version number (http://www.ncbi.nlm.nih.gov/RefSeq/),
b. the Locus Reference Genomic (LRG) database (http://www.lrg-sequence.org).

Clinical reports should include sequence reference(s) to ensure unambiguous naming of the variant:
“g.” for genomic sequence,
“c.” for coding DNA sequence,
“p.” for protein,
“m.” for mitochondria
“A” of the ATG translation initiation codon as position number 1
“X”, “*”, “Ter” Nonsense variant

When describing coding variants:

A reference transcript for each gene should be provided in the report.

The transcript should either represent the longest known transcript and/or most clinically relevant transcript.

Community-supported reference transcripts:

LRG10, CCDS Database11,

Human Gene Mutation Database (http://www.hgmd.cf.ac.uk),

ClinVar (http://www.ncbi.nlm.nih.gov/clinvar)

A locus-specific database.

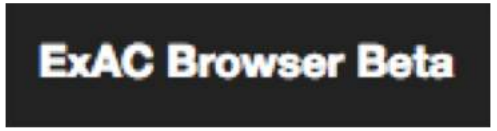
Additional exons or extended untranslated regions when there are known variants in these regions that are clinically interpretable should be evaluated.

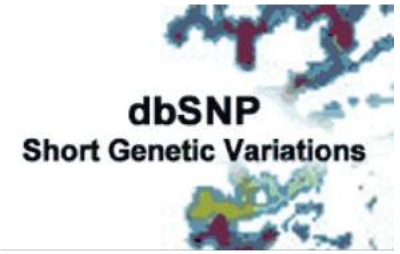
Literature and Data Use

A large number of databases contain a growing number of variants that are being discovered in the human genome.

Valuable Information can be find:

- Databases
- Published Literature





When using databases, clinical laboratories should:

(1) determine how **frequently** the database is updated (what methods were used);

(2) confirm the use of **HGVS nomenclature**;

(3) determine the degree to which **data** is validated for analytical accuracy

(*e.g.* next generation sequencing *vs* Sanger validated variants)

(4) requiring to read **associated publications**;

(5) genotype-phenotype correlation **(ClinVar)**

Population Databases	
Exome Aggregation Consortium http://exac.broadinstitute.org/	Database of variants found during exome sequencing of 61,486 unrelated individuals sequenced as part of various disease-specific and population genetic studies. Pediatric disease subjects as well as related individuals were excluded.
Exome Variant Server http://evs.gs.washington.edu/EVS	Database of variants found during exome sequencing of several large cohorts of individuals of European and African American ancestry. Includes coverage data to inform the absence of variation.
1000 Genomes http://browser.1000genomes.org	Database of variants found during low-coverage and high-coverage genomic and targeted sequencing from 26 populations. Provides more diversity compared to EVS but also contains lower quality data and some cohorts contain related individuals.
dbSNP http://www.ncbi.nlm.nih.gov/snp	Database of short genetic variations (typically 50 bp or less) submitted from many sources. May lack details of originating study and may contain pathogenic variants.
dbVar http://www.ncbi.nlm.nih.gov/dbvar	Database of structural variation (typically greater than 50 bp) submitted from many sources.
Disease Databases	
ClinVar http://www.ncbi.nlm.nih.gov/clinvar	Database of assertions about the clinical significance and phenotype relationship of human variation.
OMIM http://www.omim.org	Database of human genes and genetic conditions that also contains a representative sampling of disease-associated genetic variants.
Human Gene Mutation Database http://www.hgmd.org	Database of variant annotations published in the literature. Requires fee-based subscription for much of the content.
Locus/Disease/Ethnic/Other-Specific Databases http://www.hgvs.org/dblist/dblist.html http://www.lovd.nl	The HGVS site developed a list of thousands of different databases that provide variant annotations on specific subsets of human variation. A large percentage of databases are built in the LOVD system.
DECIPHER http://decipher.sanger.ac.uk	A molecular cytogenetic database for clinicians and researchers linking genomic microarray data with phenotype using the Ensembl genome browser.
Sequence Databases	
NCBI Genome http://www.ncbi.nlm.nih.gov/genome	Source of full human genome reference sequences.
RefSeqGene http://www.ncbi.nlm.nih.gov/refseq/rsg and Locus Reference Genomic (LRG) http://www.lrg-sequence.org	Medically relevant gene reference sequence resource
MitoMap http://www.mitomap.org/MITOMAP/HumanMitoSeq	Revised Cambridge reference sequence (rCRS) for the Human Mitochondrial DNA

Missense prediction	Name	Website	Basis
	ConSurf	http://consurftest.tau.ac.il	Evolutionary conservation
	FATHMM	http://fathmm.biocompute.org.uk	Evolutionary conservation
	MutationAssessor	http://mutationassessor.org	Evolutionary conservation
	PANTHER	http://www.pantherdb.org/tools/csnpScoreForm.jsp	Evolutionary conservation
	PhD-SNP	http://snps.biofold.org/phd-snp/phd-snp.html	Evolutionary conservation
	SIFT	http://sift.jcvi.org	Evolutionary conservation
	SNPs&GO	http://snps-and-go.biocomp.unibo.it/snps-and-go	Protein structure/function
	Align GVGD	http://agvgd.iarc.fr/agvgd_input.php	Protein structure/function and evolutionary conservation
	MAPP	http://mendel.stanford.edu/SidowLab/downloads/MAPP/index.html	Protein structure/function and evolutionary conservation
	MutationTaster	http://www.mutationtaster.org	Protein structure/function and evolutionary conservation
	MutPred	http://mutpred.mutdb.org	Protein structure/function and evolutionary conservation
	PolyPhen-2	http://genetics.bwh.harvard.edu/pph2	Protein structure/function and evolutionary conservation
	PROVEAN	http://provean.jcvi.org/index.php	Alignment and measurement of similarity between variant sequence and protein sequence homolog
	nsSNPAnalyzer	http://snpanalyzer.uthsc.edu	Multiple sequence alignment and protein structure analysis
	Condel	http://bg.upf.edu/condel/home	Combines SIFT, PolyPhen-2 and MutationAssessor
	CADD	http://cadd.gs.washington.edu	Contrasts annotations of fixed/nearly fixed derived alleles in humans with simulated variants
Splice site prediction			
	GeneSplicer	http://www.cbcb.umd.edu/software/GeneSplicer/gene_spl.shtml	Markov models
	Human Splicing Finder	http://www.umd.be/HSF	Position-dependent logic

60-80% Accurancy in prediction

sensitivity (~90–100%) relative to specificity (~60–80%)

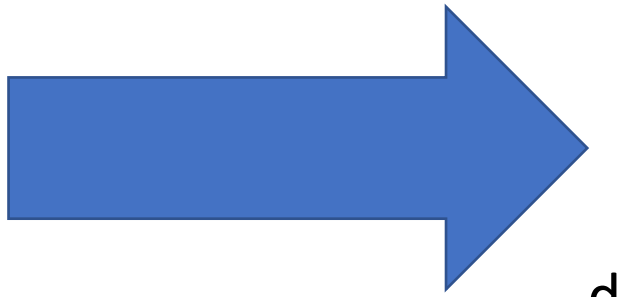
Proposed Criteria for Interpretation of Sequence Variants

Inheritance patern?

- Mendelian
- Somatic variation
- Pharmacogenomic variants
- Variants in genes associated with multigenic non-Mendelian complex disorders

Which analysis used to identified the variant?

- A single gene,
- Gene panel,
- Exome,
- Genome,
- Transcriptome,



pathogenic for a

One Case is NOT enough!

disruptive/d aging to the protein, but not implicated in a disease.

Standards and Guidelines for the Interpretation of Sequence Variants: A Joint Consensus Recommendation of the **American College of Medical Genetics and Genomics** and **the Association for Molecular Pathology**

Criteria for Classifying Pathogenic Variants

Very strong evidence of pathogenicity

PVS1 Null variant (nonsense, frameshift, canonical +/–1 or 2 splice sites, initiation codon, single or multi-exon deletion) in a gene where loss of function (LOF) is a known mechanism of disease

Caveats:

- Beware of genes where LOF is not a known disease mechanism (e.g. *GFAP*, *MYH7*)
- Use caution interpreting LOF variants at the extreme 3' end of a gene
- Use caution with splice variants that are predicted to lead to exon skipping but leave the remainder of the protein intact
- Use caution in the presence of multiple transcripts

Strong evidence of pathogenicity

PS1 Same amino acid change as a previously established pathogenic variant regardless of nucleotide change

Example: Val->Leu caused by either G>C or G>T in the same codon

Caveat: Beware of changes that impact splicing rather than at the amino acid/protein level

PS2 *De novo* (both maternity and paternity confirmed) in a patient with the disease and no family history

Note: Confirmation of paternity only is insufficient. Egg donation, surrogate motherhood, errors in embryo transfer, *etc*. can contribute to non-maternity

PS3 Well-established *in vitro* or *in vivo* functional studies supportive of a damaging effect on the gene or gene product

Note: Functional studies that have been validated and shown to be reproducible and robust in a clinical diagnostic laboratory setting are considered the most well-established

PS4 The prevalence of the variant in affected individuals is significantly increased compared to the prevalence in controls

Note 1: Relative risk (RR) or odds ratio (OR), as obtained from case-control studies, is >5.0 and the confidence interval around the estimate of RR or OR does not include 1.0. See manuscript for detailed guidance.

Note 2: In instances of very rare variants where case-control studies may not reach statistical significance, the prior observation of the variant in multiple unrelated patients with the same phenotype, and its absence in controls, may be used as moderate level of evidence.

Moderate evidence of pathogenicity

PM1 Located in a mutational hot spot and/or critical and well-established functional domain (*e.g*. active site of an enzyme) without benign variation

PM2 Absent from controls (or at extremely low frequency if recessive) (see Table 6) in Exome Sequencing Project, 1000 Genomes or ExAC

Caveat: Population data for indels may be poorly called by next generation sequencing

PM3 For recessive disorders, detected in *trans* with a pathogenic variant

Note: This requires testing of parents (or offspring) to determine phase

PM4 Protein length changes due to in-frame deletions/insertions in a non-repeat region or stop-loss variants

PM5 Novel missense change at an amino acid residue where a different missense change determined to be pathogenic has been seen before

Example: Arg156His is pathogenic; now you observe Arg156Cys

Caveat: Beware of changes that impact splicing rather than at the amino acid/protein level

PM6 Assumed *de novo*, but without confirmation of paternity and maternity

Supporting evidence of pathogenicity

PP1 Co-segregation with disease in multiple affected family members in a gene definitively known to cause the disease

Note: May be used as stronger evidence with increasing segregation data

PP2 Missense variant in a gene that has a low rate of benign missense variation and where missense variants are a common mechanism of disease

PP3 Multiple lines of computational evidence support a deleterious effect on the gene or gene product (conservation, evolutionary, splicing impact, etc)

Caveat: As many *in silico* algorithms use the same or very similar input for their predictions, each algorithm should not be counted as an independent criterion. PP3 can be used only once in any evaluation of a variant.

PP4 Patient's phenotype or family history is highly specific for a disease with a single genetic etiology

PP5 Reputable source recently reports variant as pathogenic but the evidence is not available to the laboratory to perform an independent evaluation

Criteria for Classifying Benign Variants

Stand-Alone evidence of benign impact

BA1 Allele frequency is above 5% in Exome Sequencing Project, 1000 Genomes, or ExAC

Strong evidence of benign impact

BS1 Allele frequency is greater than expected for disorder (see table 6)

BS2 Observed in a healthy adult individual for a recessive (homozygous), dominant (heterozygous), or X-linked (hemizygous) disorder with full penetrance expected at an early age

BS3 Well-established *in vitro* or *in vivo* functional studies shows no damaging effect on protein function or splicing

BS4 Lack of segregation in affected members of a family

Caveat: The presence of phenocopies for common phenotypes (*i.e.* cancer, epilepsy) can mimic lack of segregation among affected individuals. Also, families may have more than one pathogenic variant contributing to an autosomal dominant disorder, further confounding an apparent lack of segregation.

Supporting evidence of benign impact

BP1 Missense variant in a gene for which primarily truncating variants are known to cause disease

BP2 Observed in *trans* with a pathogenic variant for a fully penetrant dominant gene/disorder; or observed in *cis* with a pathogenic variant in any inheritance pattern

BP3 In-frame deletions/insertions in a repetitive region without a known function

BP4 Multiple lines of computational evidence suggest no impact on gene or gene product (conservation, evolutionary, splicing impact, etc)

Caveat: As many *in silico* algorithms use the same or very similar input for their predictions, each algorithm cannot be counted as an independent criterion. BP4 can be used only once in any evaluation of a variant.

BP5 Variant found in a case with an alternate molecular basis for disease

BP6 Reputable source recently reports variant as benign but the evidence is not available to the laboratory to perform an independent evaluation

BP7 A synonymous (silent) variant for which splicing prediction algorithms predict no impact to the splice consensus sequence nor the creation of a new splice site AND the nucleotide is not highly conserved

	Benign		Pathogenic			
	Strong	**Supporting**	**Supporting**	**Moderate**	**Strong**	**Very Strong**
Population Data	MAF is too high for disorder *BA1/BS1* **OR** observation in controls inconsistent with disease penetrance *BS2*			Absent in population databases *PM2*	Prevalence in affecteds statistically increased over controls *PS4*	
Computational And Predictive Data		Multiple lines of computational evidence suggest no impact on gene /gene product *BP4* Missense in gene where only truncating cause disease *BP1* Silent variant with non predicted splice impact *BP7*	Multiple lines of computational evidence support a deleterious effect on the gene /gene product *PP3*	Novel missense change at an amino acid residue where a different pathogenic missense change has been seen before *PM5* Protein length changing variant *PM4*	Same amino acid change as an established pathogenic variant *PS1*	Predicted null variant in a gene where LOF is a known mechanism of disease *PVS1*
Functional Data	Well-established functional studies show no deleterious effect *BS3*		Missense in gene with low rate of benign missense variants and path. missenses common *PP2*	Mutational hot spot or well-studied functional domain without benign variation *PM1*	Well-established functional studies show a deleterious effect *PS3*	
Segregation Data	Non-segregation with disease *BS4*		Co-segregation with disease in multiple affected family members *PP1*	Increased segregation data →		
De novo Data				*De novo* (without paternity & maternity confirmed) *PM6*	*De novo* (paternity & maternity confirmed) *PS2*	
Allelic Data		Observed in *trans* with a dominant variant *BP2* Observed in *cis* with a pathogenic variant *BP2*		For recessive disorders, detected in *trans* with a pathogenic variant *PM3*		
Other Database		Reputable source w/out shared data = benign *BP6*	Reputable source = pathogenic *PP5*			
Other Data		Found in case with an alternate cause *BP5*	Patient's phenotype or FH highly specific for gene *PP4*			

Pathogenic

1. 1 Very Strong (PVS1) *AND*
   - a. ≥1 Strong (PS1–PS4) *OR*
   - b. ≥2 Moderate (PM1–PM6) *OR*
   - c. 1 Moderate (PM1–PM6) and 1 Supporting (PP1–PP5) *OR*
   - d. ≥2 Supporting (PP1–PP5)
2. ≥2 Strong (PS1–PS4) *OR*
3. 1 Strong (PS1–PS4) *AND*
   - a. ≥3 Moderate (PM1–PM6) *OR*
   - b. 2 Moderate (PM1–PM6) *AND* ≥2 Supporting (PP1–PP5) *OR*
   - c. 1 Moderate (PM1–PM6) *AND* ≥4 Supporting (PP1–PP5)

Likely Pathogenic

1. 1 Very Strong (PVS1) *AND* 1 Moderate (PM1–PM6) *OR*
2. 1 Strong (PS1–PS4) *AND* 1–2 Moderate (PM1–PM6) *OR*
3. 1 Strong (PS1–PS4) *AND* ≥2 Supporting (PP1–PP5) *OR*
4. ≥3 Moderate (PM1–PM6) *OR*
5. 2 Moderate (PM1–PM6) *AND* ≥2 Supporting (PP1–PP5) *OR*
6. 1 Moderate (PM1–PM6) *AND* ≥4 Supporting (PP1–PP5)

Benign

1. 1 Stand-Alone (BA1) *OR*
2. ≥2 Strong (BS1–BS4)

Likely Benign

1. 1 Strong (BS1–BS4) and 1 Supporting (BP1–BP7) *OR*
2. ≥2 Supporting (BP1–BP7)

Special Considerations

1. Evaluating and Reporting Variants in Genes of Uncertain Significance (GUS) based on the Indication for Testing

a gene has never been associated with any patient phenotype,

the gene has been associated with a different phenotype.

de novo observation is no longer strong evidence for pathogenicity,

thousands of variants in genome could segregate with a significant LOD score,

insufficient evidence for a causative rol → “Uncertain Significance”

2. Evaluating Variants in Healthy Individuals or as Incidental Findings

Negative results in performing disease-targeted testing

Predicted penetrance of pathogenic variants found in the absence of a phenotype

Family history

3. Mitochondrial variants

%Heteroplasmy/%Homoplasmy?

275 mtDNA variants relating to disease have been recorded

(http://mitomap.org/bin/view.pl/MITOMAP/WebHome)

(http://www.mtdb.igp.uu.se/),

secondary structures, sequences, and alignment of mitochondrial tRNAs (http://mamit-trna.u-strasbg.fr/),

mitochondrial haplogroups (http://www.phylotree.org/)

(http://www.mtdnacommunity.org/default.aspx).

Muscle, liver or urine aslo are specimen types useful for clinical evaluation

To test nuclear genes associated with mitochondrial disorders

4. Pharmacogenomics (PGx)

Challenge: Identification the effects of variants in drug metabolism

Reason: a phenotype is only apparent upon exposure to a drug

drug efficacy and risk for adverse increasingly used in clinical care

The traditional nomenclature of PGx alleles uses star (*) alleles, which often represent haplotypes, or a combination of variants on the same allele.

Pharmacogenomics Knowledge Base (http://www.pharmgkb.org/),

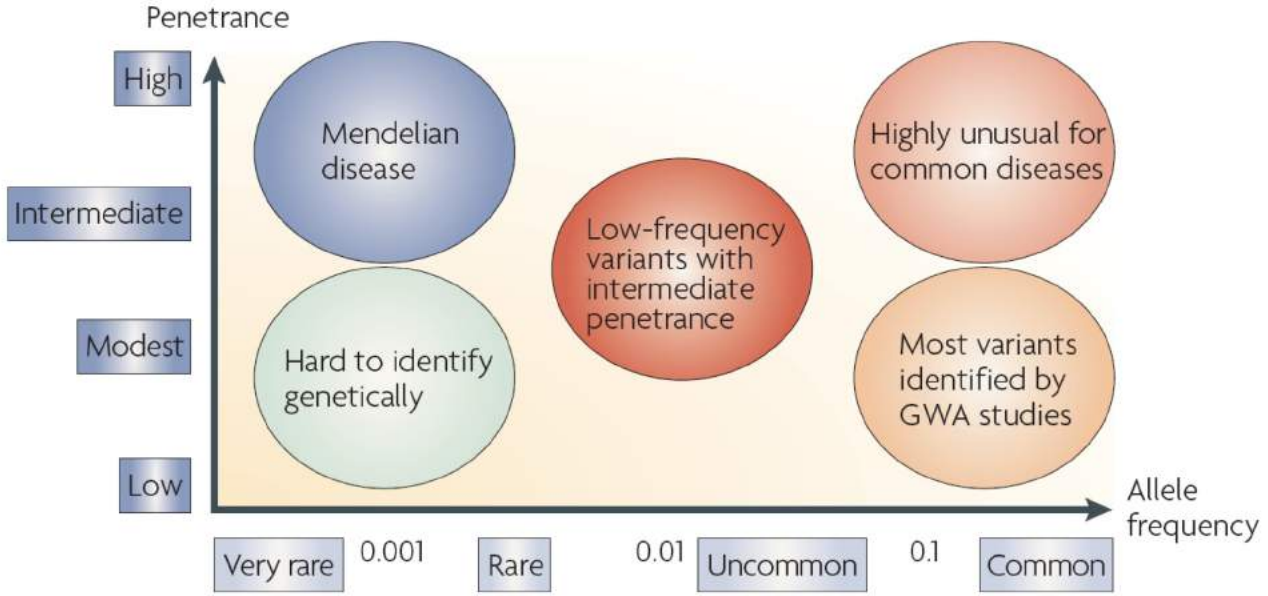
48 Alleles for the cytochrome P450 gene family http://www.cypalleles.ki.se/.49

There are 18 mammalian cytochrome P450 (*CYP*) families, which encode 57 genes in the human genome.

To metabolize drugs and other foreign chemicals

5. Common Complex Disorders

Unlike Mendelian diseases, the identification of CCD genes, have relied on population-based approaches (GWAS) rather than family-based studies.



GWAS reported the cataloguing of over 1200 risk alleles for common, complex diseases and traits.

Most of them not directly causitive;
If thought to be causitive = Linkage Disequlibrium with causal variant?

Reporting/ Cataloguing the risk in the case-control/GWAS studies:

“established risk allele”,
“likely risk allele”,
“uncertain risk allele”,

6. Somatic Variation

Cancer cells

Problem: the allele ratios are highly variable and tumor heterogeneity

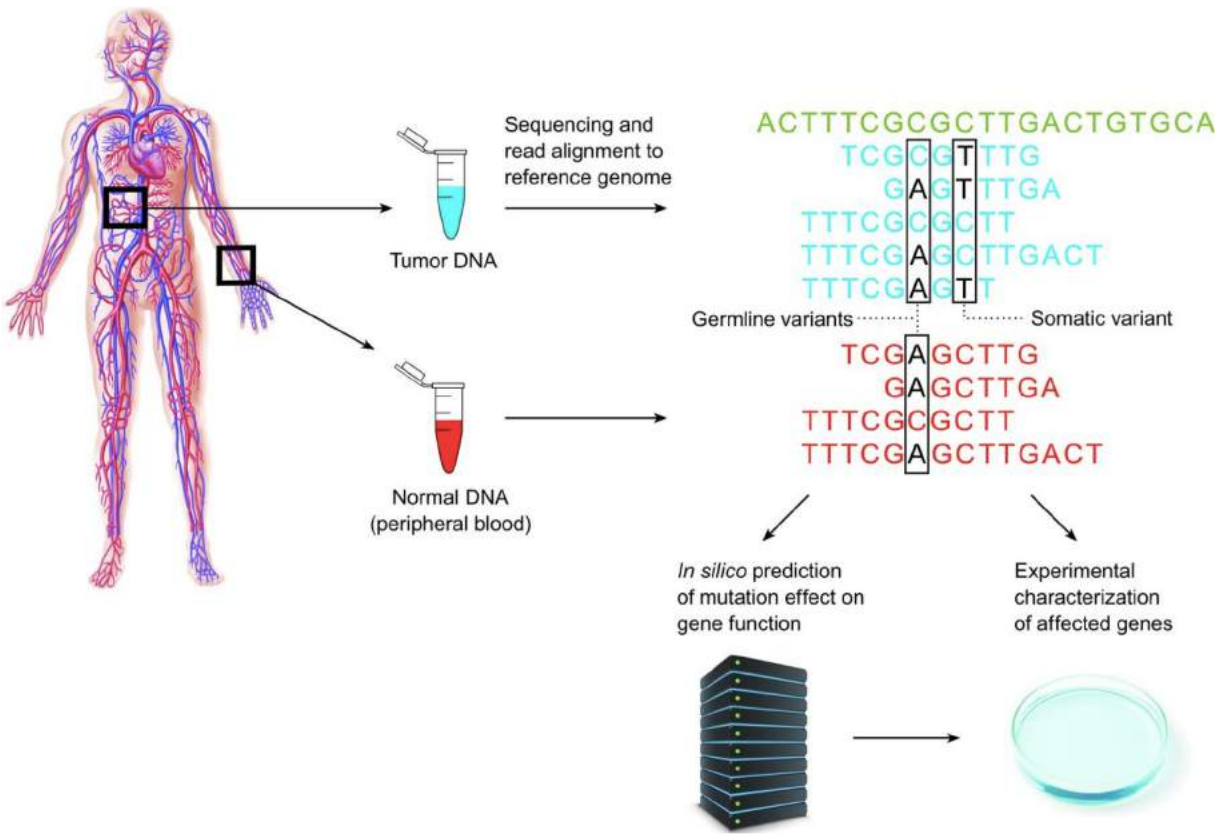
Variant classification category:

“responsive”,

“resistant”,

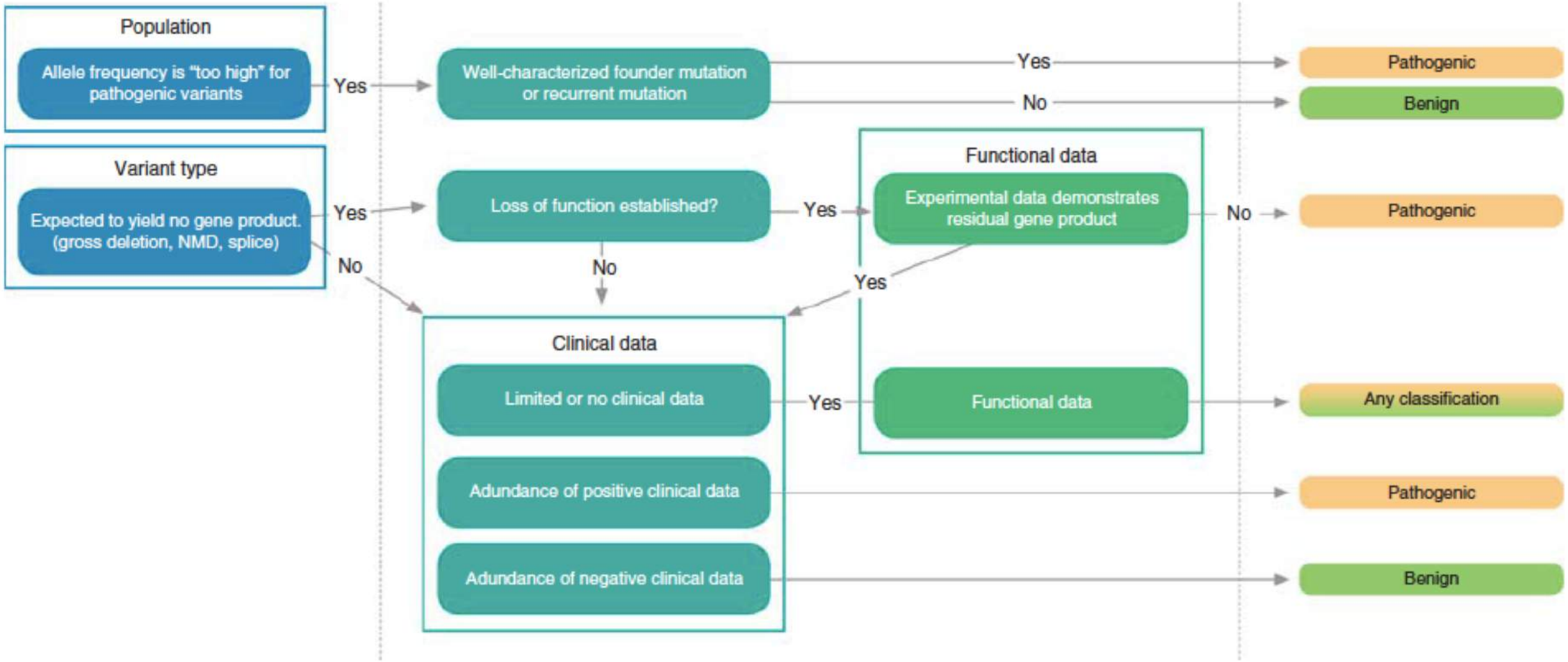
“driver”,

“passenger”

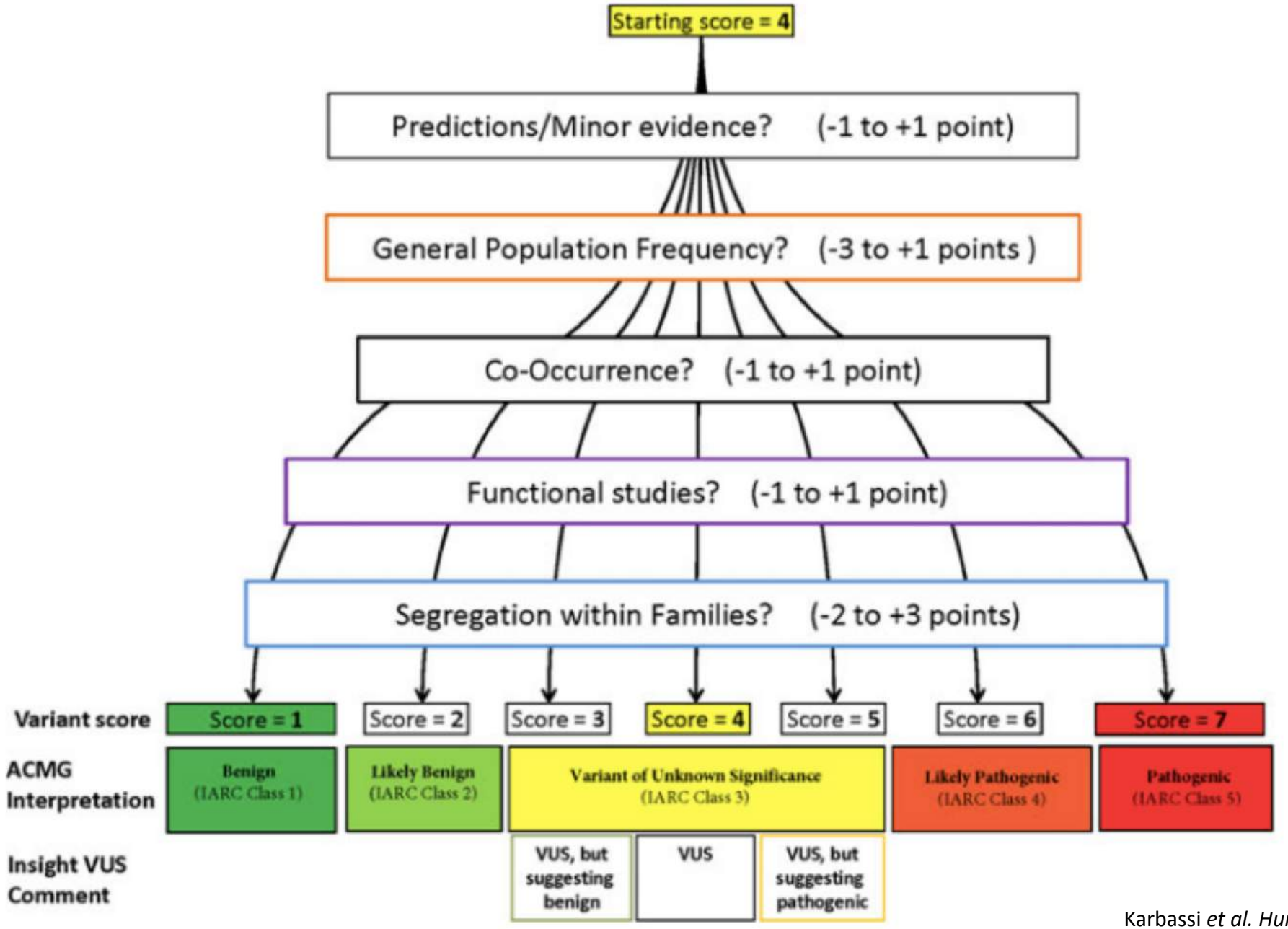


Samuel and Hudson, Clin Chem. 2013

Hierarchical approach to effect variant research:

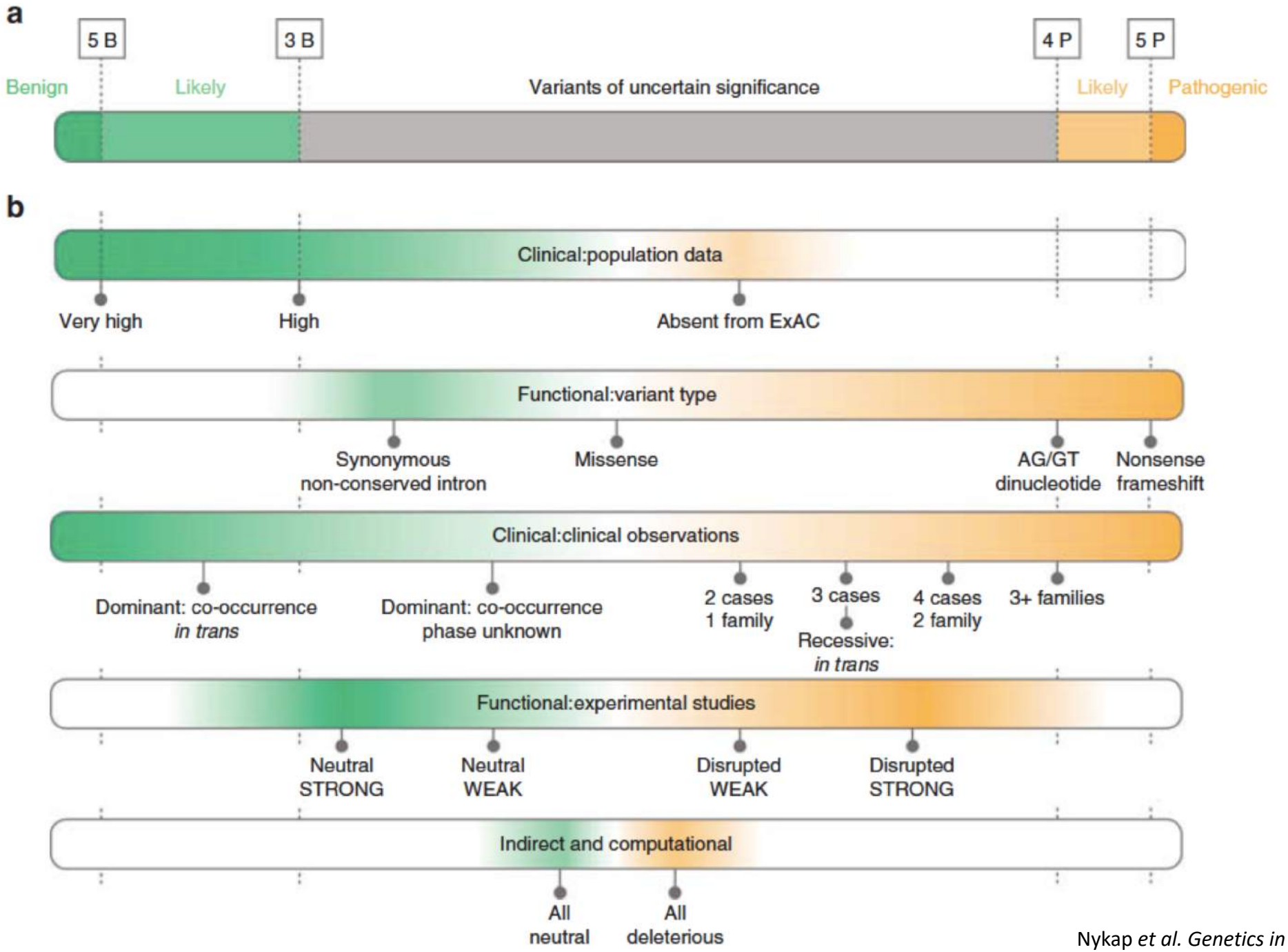


A Standardized DNA Variant Scoring System for Pathogenicity Assessments in Mendelian Disorders



Karbassi *et al. Human Mutation*. 2015

Sherloc: a comprehensive refinement of the ACMG–AMP variant classification criteria



Nykap *et al. Genetics in Medicine*. 2017