Clinical variant interpretation I

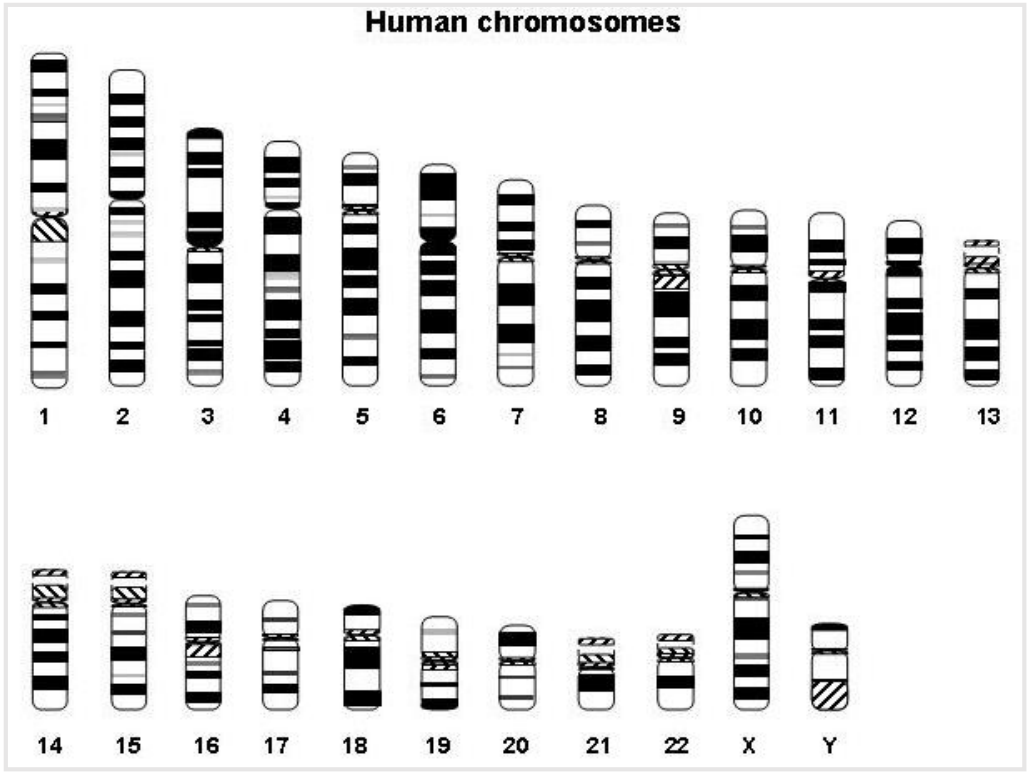
Genetics in Medicine - 0504321

2022-2023 Second Semester

Dr. Osama Alsmadi

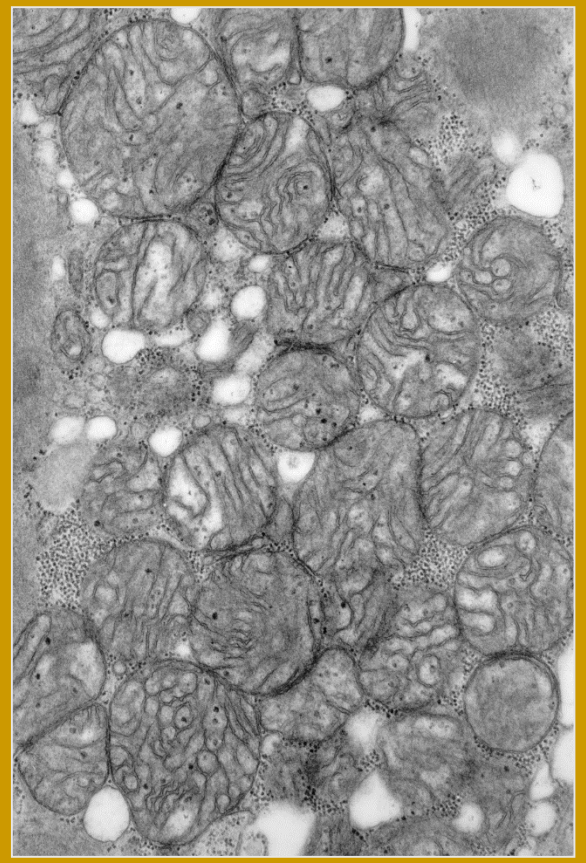
The Human Genome

- Consists of 23 pairs of chromosomes.
- Chromosomes 1 through 22 are called autosomes.
- The X & Y chromosomes are the sex chromosomes.
  - Males are XY.
  - Females are XX.



Mitochondria

- Mitochondria have their own chromosome.
  - The human mitochondrial chromosome is a circular chromosome 16,569 base pairs long.
  - Each mitochondrion contain several copies of its chromosome.
  - Every cell contains many mitochondria



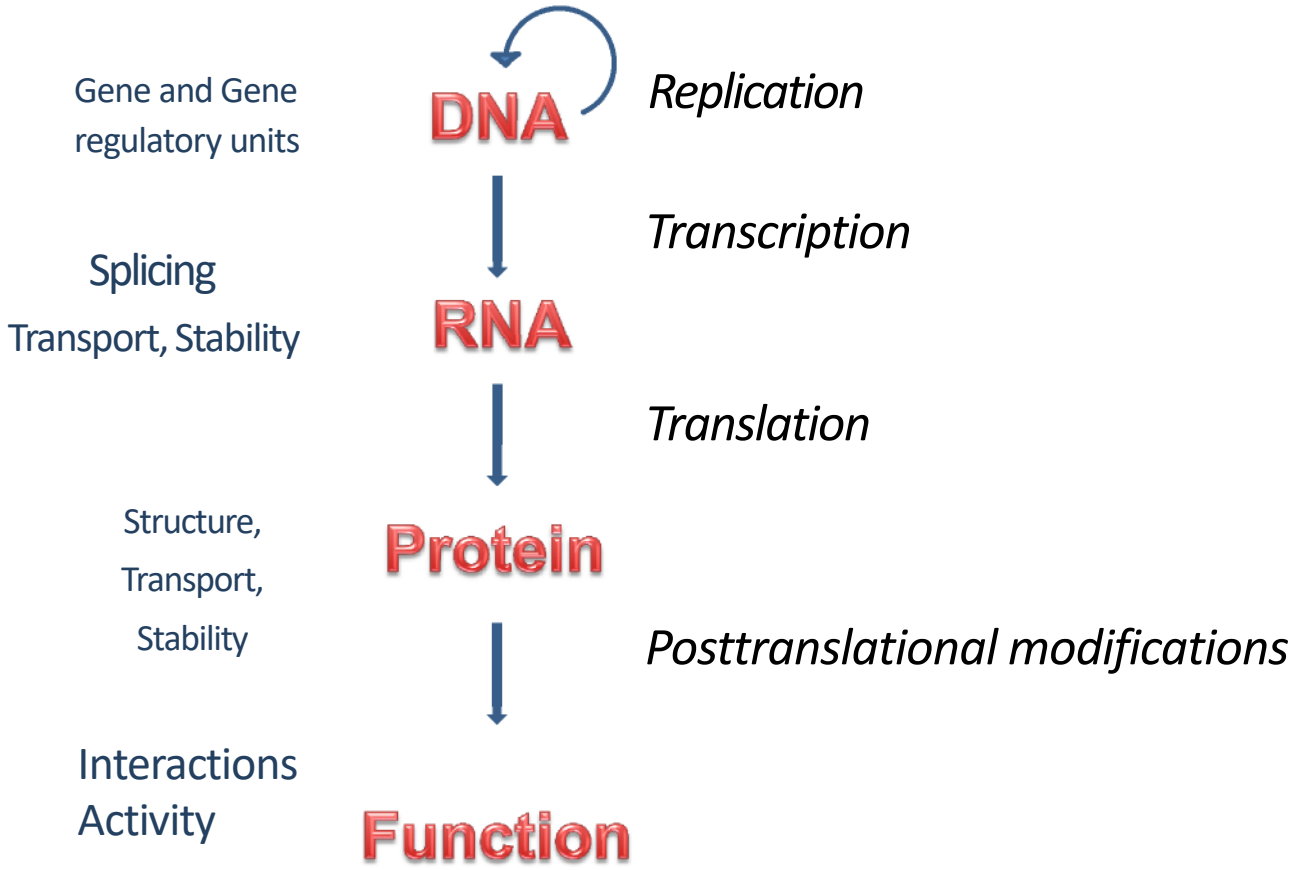
Electron micrograph of mitochondria.
Courtesy of
M. John Hicks, MD, PhD, DDS

The Human Genome

- The **Haploid** human genome contains 3 billion base pairs of DNA.
- Only **1 %** of the human genome codes for proteins.
- Humans have about 22,000 genes (**Controversy**).
- Humans are 99.9% identical at the DNA level.
- Polymorphic variation (0.1%) between and within populations

The Living Genome: Reading the Book of Life, Houston Museum of Natural Science

The Central Dogma



GENOTYPE **PHENOTYPE**

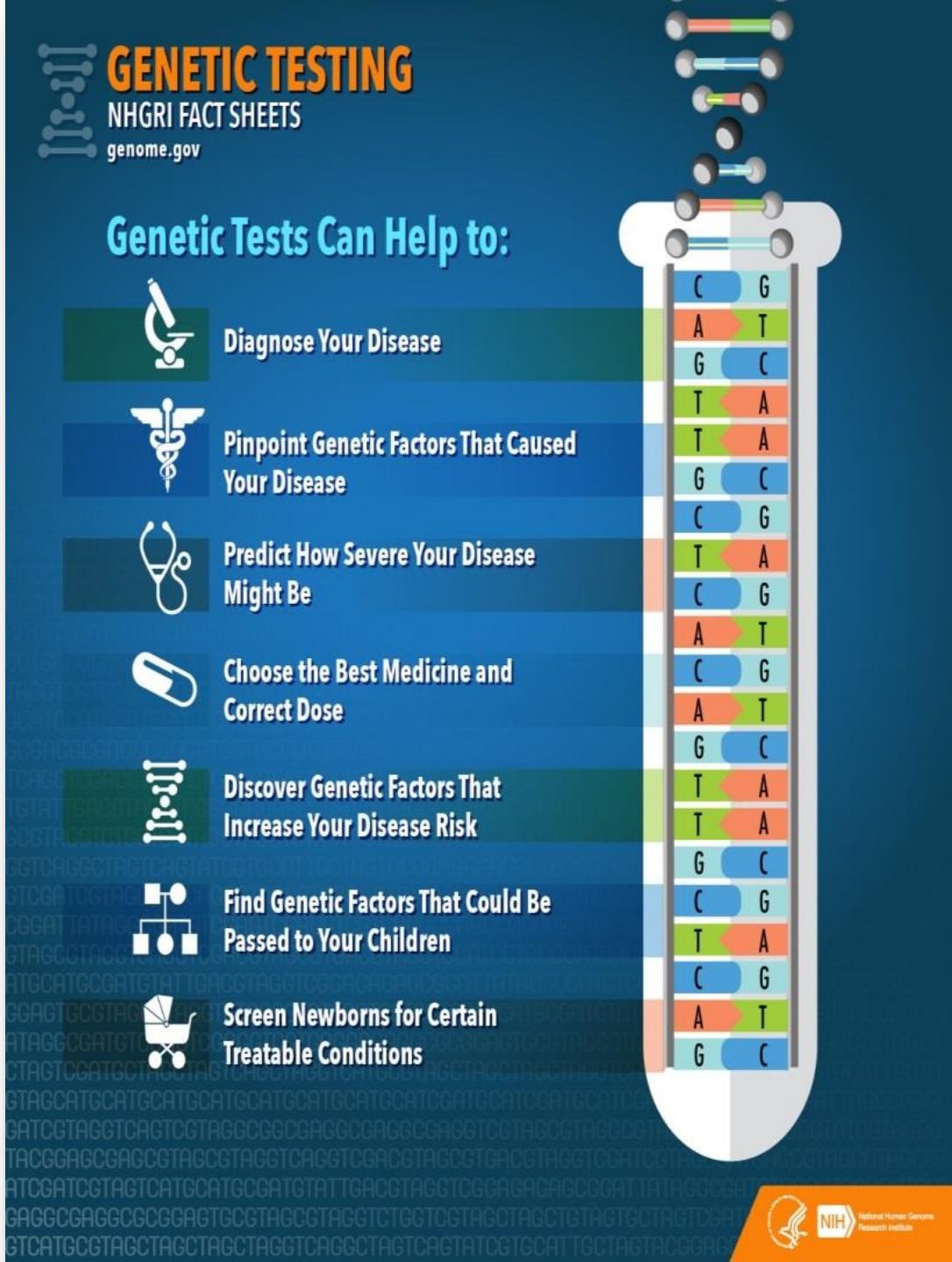
https://www.dnalc.org/resources/animations/

What is genetic testing?

https://ghr.nlm.nih.gov/primer/testing/genetictesting

- Genetic testing identifies changes in:
  - DNA
  - RNA
  - Chromosomes
  - Proteins
  - Metabolites

- Results of a genetic test can **rule in/out** a **suspected genetic condition** or help **determine** the risk of **developing** or passing on a **genetic disorder to the offsprings**

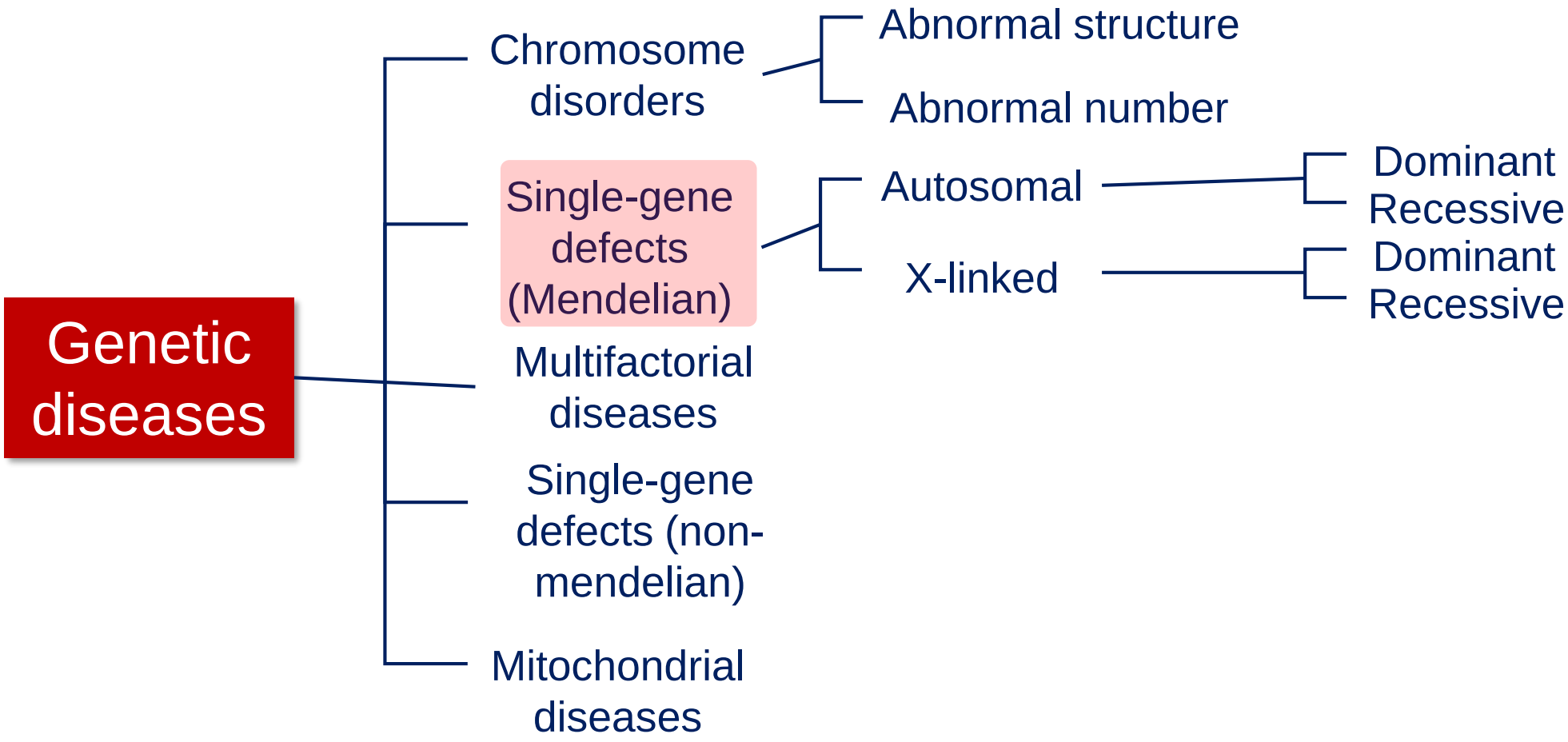


Genetic Testing Types

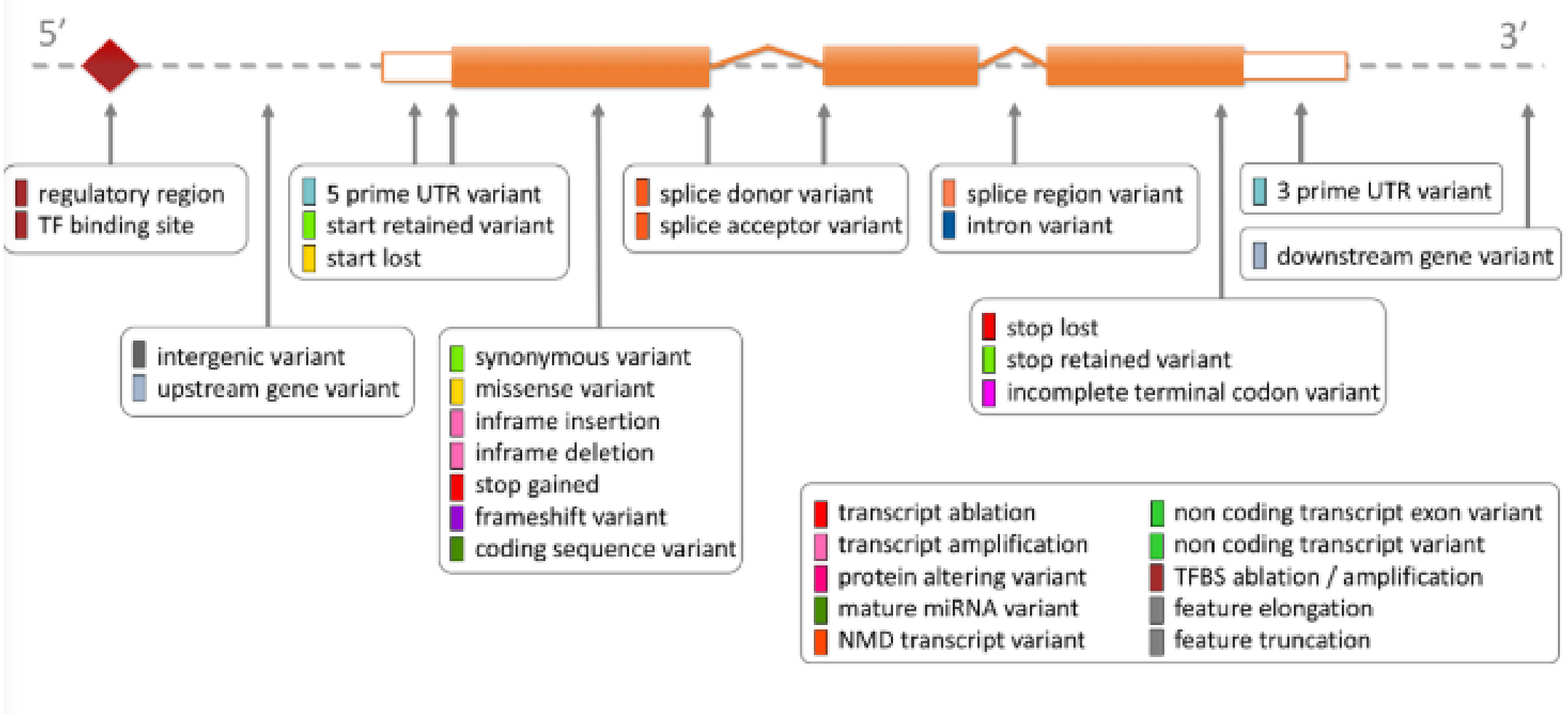
- ✓ **Diagnostic testing** (symptomatic): to identify or rule out a specific condition
- ✓ **Carrier testing**: to identify subjects heterozygous for a recessive condition
- ✓ **Pre-symptomatic testing**: prior to symptoms appearence
- ✓ **Predictive testing** (susceptibility): identify mutations that increase risk of developing a disorder
- ✓ **Newborn screening**: just after birth, to identify genetic disorders that can be treated early in life
- ✓ **Prenatal testing**: to detect alterations in a fetus (in utero)
- ✓ **Preimplantation testing**: to detect genetic changes in embryos created *in vitro* fertilization
- ✓ **Pharmacogenetic testing**: identification of the probable individual response (efficacy-risk of adverse event) to drugs
- ✓ **Forensic testing**: to identify an individual for legal/criminal purposes.

https://ghr.nlm.nih.gov/primer/testing/genetictesting

Categories of genetic diseases



Genomic Variant Impact



How to choose the right Sequencing application

- Specific phenotypes
- Low genetic heterogeneity
- Small genes

Single genes

Several individual genes

- High genetic heterogeneity
- Limited number of causative genes

Disease-focused panel of genes (tens to hundreds)

- Marked genetic heterogeneity
- Lack of diagnostic hypothesis

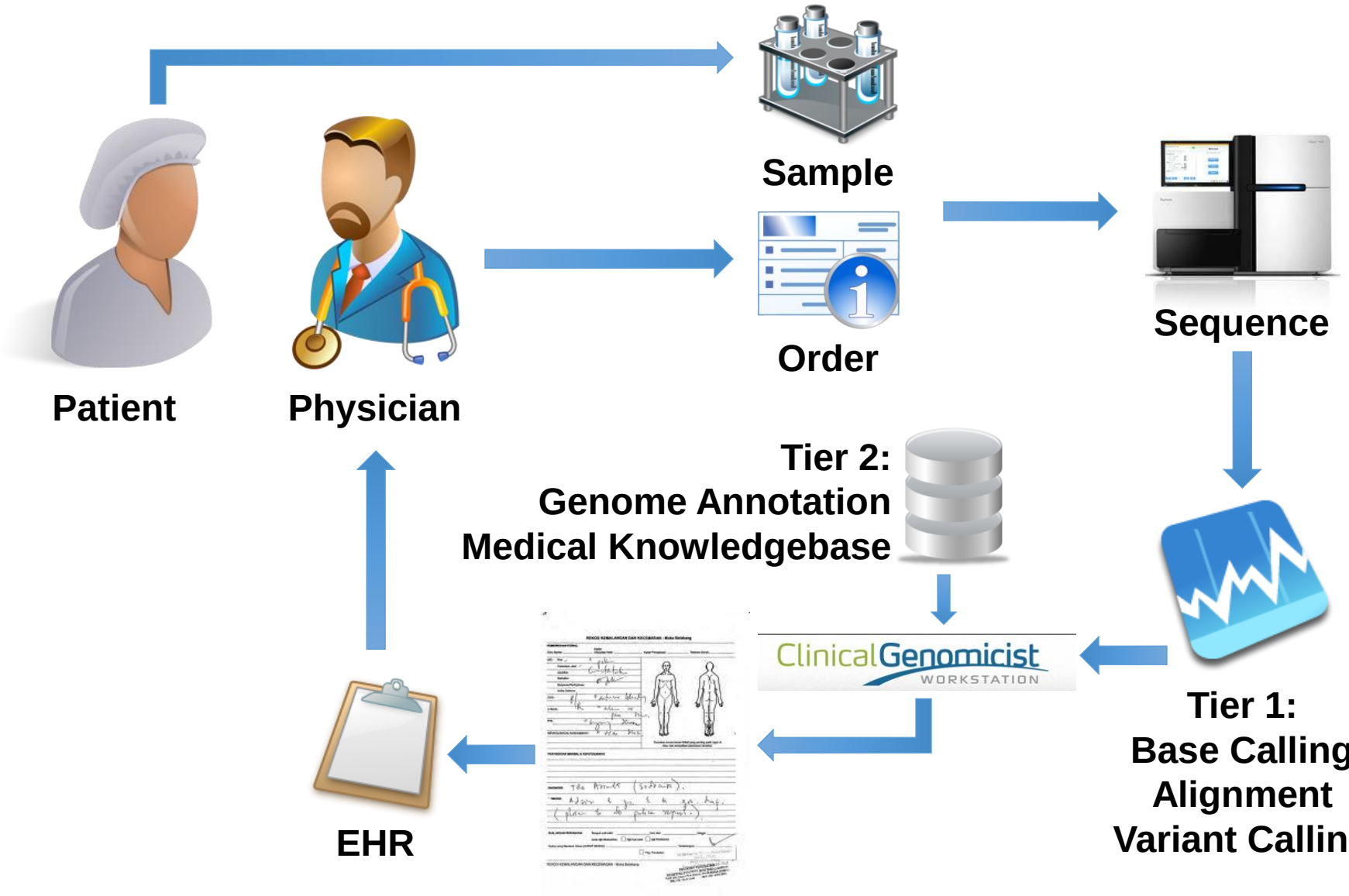
Whole exome

- 22,000 genes
- Coding regions only
- Untranslated regions
- No introns
- Only 1% of the genome
- But ~ 85% of disease-causing mutations

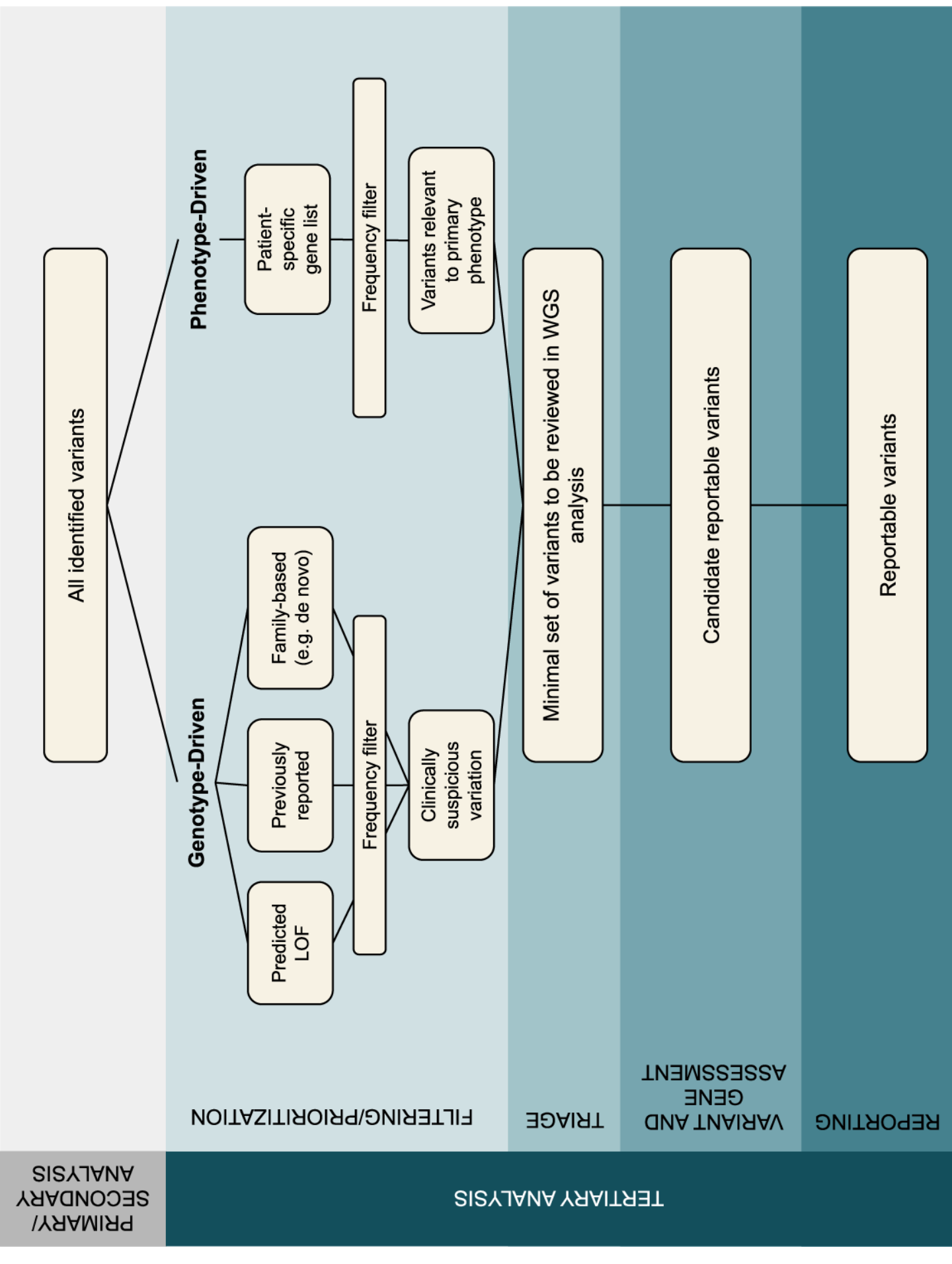
Whole genome

- The whole thing!
- Requires more cost for sequencing and analysis

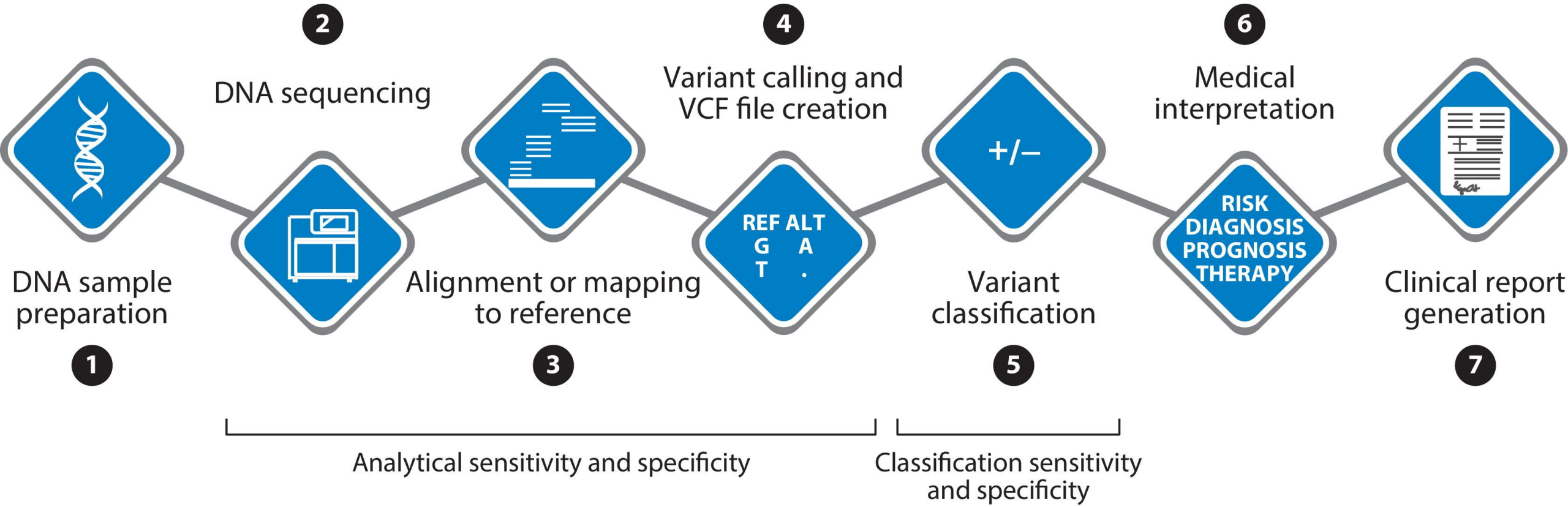
Informatics Pipeline Workflow



https://www.nature.com/articles/s41525-022-00295-z/figures/2



General workflow of clinical genetic or genomic NGS

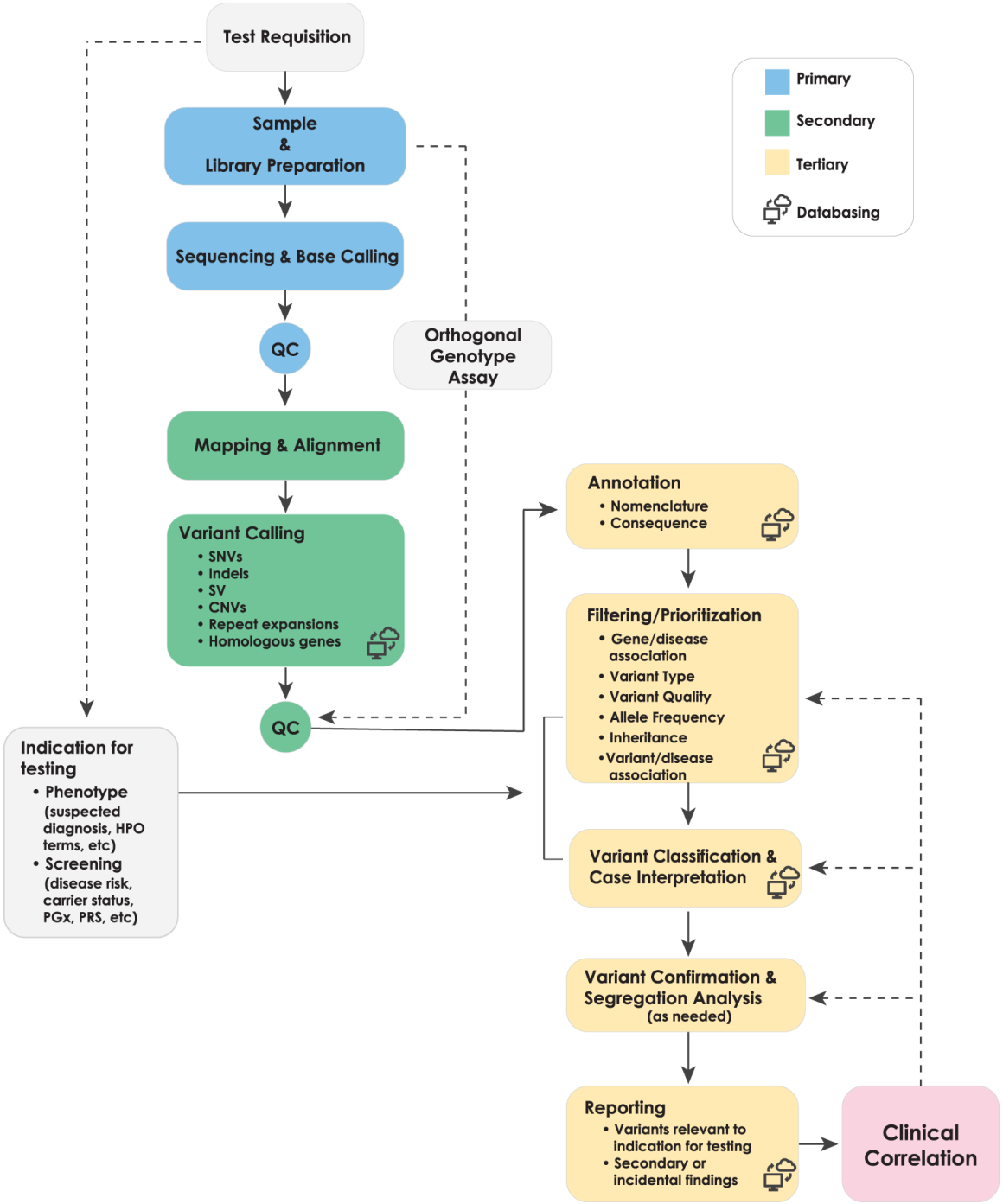


Giles HH, et al. 2021
Annu. Rev. Genom. Hum. Genet. 22:285–307

https://www.broadinstitute.org/videos/variant-classification-using-acmgamp-interpreting-sequence-guidelines

https://www.annualreviews.org/doi/10.1146/annurev-genom-121620-082709

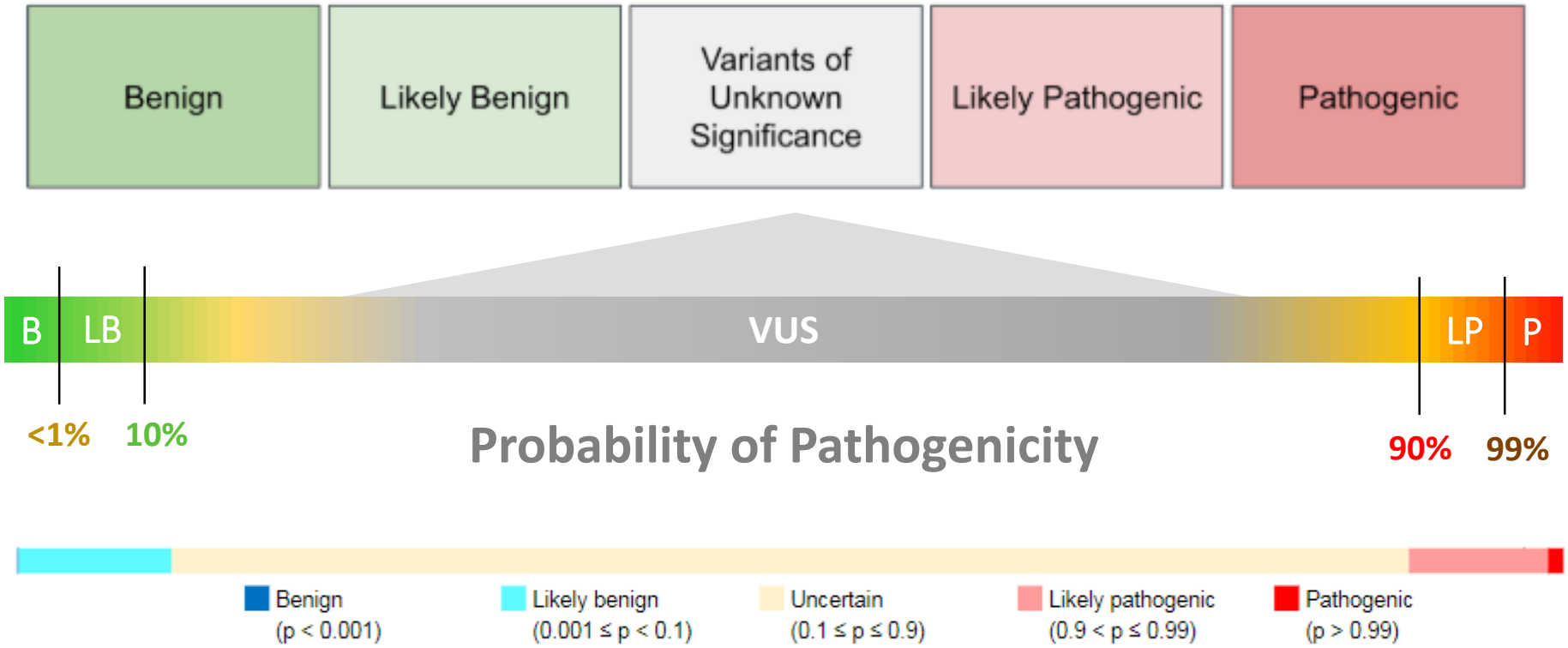
Best practices for the interpretation and reporting of clinical whole genome sequencing

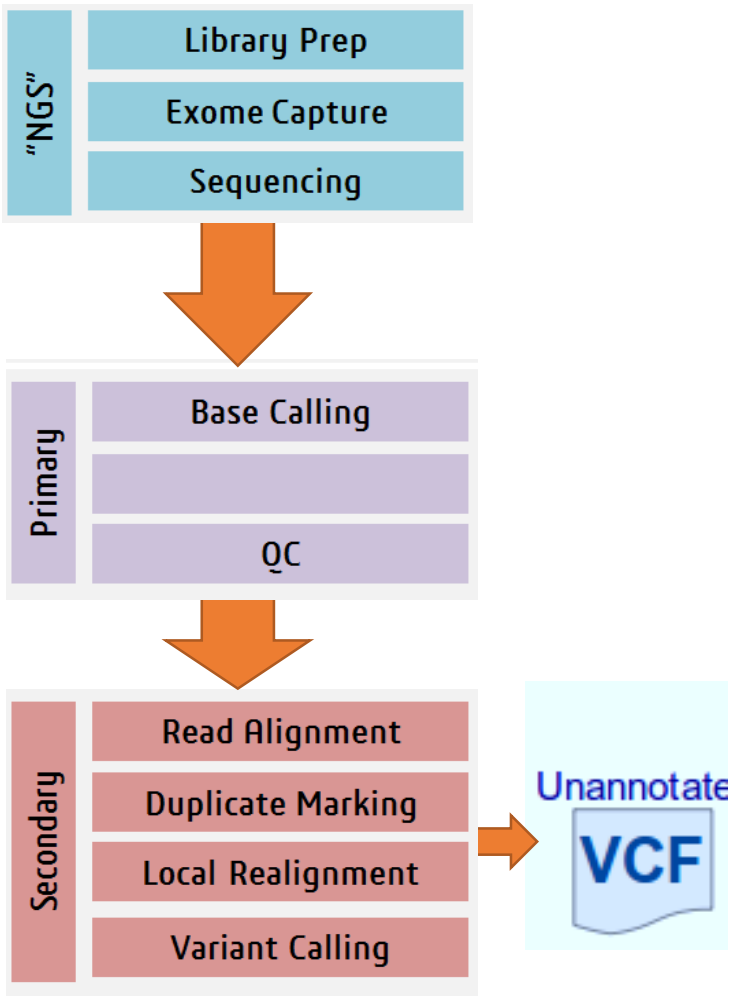


https://www.nature.com/articles/s41525-022-00295-z/figures/2

Common framework and criteria for germline variant classification

The rules proposed to classify sequence variants follows is a heuristic system for variant classification that is compatible with a formal, quantitative, naive Bayesian classifier.






```
##FILTER=<ID=HETEROPLASMY,Description="Called heterozygous genotype on mitochondrial contig">
##FORMAT=<ID=GP,Number=G,Type=Float,Description="Estimated Genotype Probability">
##FILTER=<ID=IMP,Description="Set if true: IMP==1">
##FILTER=<ID=BOOSTED,Description="Set if true: BOOSTED==1">
##FILTER=<ID=LOWDP,Description="Set if GQ>20 and 10<=DP<=20">
##FILTER=<ID=LOWQ,Description="Set if GQ<=20 or DP<10">
##FILTER=<ID=NOTVALIDATED,Description="Set if variant falls outside of analytic range">
##FORMAT=<ID=GL,Number=G,Type=Float,Description="Genotype likelihoods">
##FORMAT=<ID=VAR_TYPE,Number=.,Type=String,Description="Variant type: SNV, INSERTION, DELETION, SUBSTITUTION, MNV, COMPLEX"
##FORMAT=<ID=VAR_CONTEXT,Number=.,Type=String,Description="Variant genomic context: STR-expansion, STR-contraction, STR-proximal">
##FORMAT=<ID=STR_MAX_LEN,Number=1,Type=Integer,Description="Maximum observed STR sequence length">
##FORMAT=<ID=STR_PERIOD,Number=1,Type=Integer,Description="Repetition period for STR variants">
##FORMAT=<ID=STR_TIMES,Number=1,Type=Float,Description="Number of repetition for STR variants">
##pipeline=helix-v2.6.1
#CHROM  POS       ID  REF      ALT      QUAL  FILTER  INFO  FORMAT                PC-TA537BFRZC6B332GAQD'
chr1    55033879  .   A        ACTG     35    PASS    .     GT:AD:DP:GQ:VAR_TYPE  0/1:500,500:1000:99:SNV
chr2    47805173  .   G        A        35    PASS    .     GT:AD:DP:GQ:VAR_TYPE  0/1:500,500:1000:99:SNV
chr2    47793163  .   C        G        35    PASS    .     GT:AD:DP:GQ:VAR_TYPE  0/1:500,500:1000:99:SNV
chr13   32319070  .   T        A,TA     35    PASS    .     GT:AD:DP:GQ:VAR_TYPE  0/1:500,500:1000:99:SNV
chr19   11113686  .   A        G        35    PASS    .     GT:AD:DP:GQ:VAR_TYPE  0/1:500,500:1000:99:SNV
chr2    21011802  .   C        T        35    PASS    .     GT:AD:DP:GQ:VAR_TYPE  0/1:500,500:1000:99:SNV
chr7    5977709   .   T        C        35    PASS    .     GT:AD:DP:GQ:VAR_TYPE  0/1:500,500:1000:99:SNV
chr17   43094795  .   A        C        35    PASS    .     GT:AD:DP:GQ:VAR_TYPE  0/1:500,500:1000:99:SNV
chr19   11102787  .   G        A        35    PASS    .     GT:AD:DP:GQ:VAR_TYPE  0/1:500,500:1000:99:SNV
chr7    6003794   .   T        A        35    PASS    .     GT:AD:DP:GQ:VAR_TYPE  0/1:500,500:1000:99:SNV
chr3    37028782  .   AG       CC       35    PASS    .     GT:AD:DP:GQ:VAR_TYPE  0/1:500,500:1000:99:SNV
chr2    47798826  .   A        AAC      35    PASS    .     GT:AD:DP:GQ:VAR_TYPE  0/1:500,500:1000:99:SNV
chr7    5987451   .   CTT      C        35    PASS    .     GT:AD:DP:GQ:VAR_TYPE  0/1:500,500:1000:99:SNV
chr13   32340378  .   AGCAAG   ATGCTG   35    PASS    .     GT:AD:DP:GQ:VAR_TYPE  0/1:500,500:1000:99:SNV
chr2    21038086  .   C        A        35    PASS    .     GT:AD:DP:GQ:VAR_TYPE  0/1:500,500:1000:99:SNV
chr19   11123663  .   C        A        35    PASS    .     GT:AD:DP:GQ:VAR_TYPE  0/1:500,500:1000:99:SNV
chr1    55033930  .   G        GGAGGA   35    PASS    .     GT:AD:DP:GQ:VAR_TYPE  0/1:500,500:1000:99:SNV
chr2    21010226  .   CTCA     C        35    PASS    .     GT:AD:DP:GQ:VAR_TYPE  0/1:500,500:1000:99:SNV
chr7    6003018   .   A        G        35    PASS    .     GT:AD:DP:GQ:VAR_TYPE  0/1:500,500:1000:99:SNV
chr17   43124034  .   G        GCCT     35    PASS    .     GT:AD:DP:GQ:VAR_TYPE  0/1:500,500:1000:99:SNV
chr17   43124037  .   TT       T        35    PASS    .     GT:AD:DP:GQ:VAR_TYPE  0/1:500,500:1000:99:SNV
chr17   43045679  .   C        G        35    PASS    .     GT:AD:DP:GQ:VAR_TYPE  0/1:500,500:1000:99:SNV
chr13   32338769  .   A        AT       35    PASS    .     GT:AD:DP:GQ:VAR_TYPE  0/1:500,500:1000:99:SNV
chr7    5973402   .   CTGA     C        35    PASS    .     GT:AD:DP:GQ:VAR_TYPE  0/1:500,500:1000:99:SNV
chr2    47806206  .   A        G        35    PASS    .     GT:AD:DP:GQ:VAR_TYPE  0/1:500,500:1000:99:SNV
chr19   11128084  .   C        T        35    PASS    .     GT:AD:DP:GQ:VAR_TYPE  0/1:500,500:1000:99:SNV
chr2    47806452  .   G        GGGG     35    PASS    .     GT:AD:DP:GQ:VAR_TYPE  0/1:500,500:1000:99:SNV
chr2    47801152  .   TTGG     T        35    PASS    .     GT:AD:DP:GQ:VAR_TYPE  0/1:500,500:1000:99:SNV
chr19   11120166  .   C        T        35    PASS    .     GT:AD:DP:GQ:VAR_TYPE  0/1:500,500:1000:99:SNV
chr19   11111506  .   T        C        35    PASS    .     GT:AD:DP:GQ:VAR_TYPE  0/1:500,500:1000:99:SNV
chr2    47805601  .   A        AT,ATT   35    PASS    .     GT:AD:DP:GQ:VAR_TYPE  0/1:500,500:1000:99:SNV
chr2    47805601  .   AT       A        35    PASS    .     GT:AD:DP:GQ:VAR_TYPE  0/1:500,500:1000:99:SNV
chr19   11128142  .   C        T        35    PASS    .     GT:AD:DP:GQ:VAR_TYPE  0/1:500,500:1000:99:SNV
chr17   43059469  .   C        CACA     35    PASS    .     GT:AD:DP:GQ:VAR_TYPE  0/1:500,500:1000:99:SNV
chr2    47806751  .   CTT      C,CT     35    PASS    .     GT:AD:DP:GQ:VAR_TYPE  0/1:500,500:1000:99:SNV
chr17   43125260  .   G        A        35    PASS    .     GT:AD:DP:GQ:VAR_TYPE  0/1:500,500:1000:99:SNV
chr17   43124135  .   C        CAT      35    PASS    .     GT:AD:DP:GQ:VAR_TYPE  0/1:500,500:1000:99:SNV
chr17   43124745  .   GTTTTT   G        35    PASS    .     GT:AD:DP:GQ:VAR_TYPE  0/1:500,500:1000:99:SNV
chr17   43044346  .   C        T        35    PASS    .     GT:AD:DP:GQ:VAR_TYPE  0/1:500,500:1000:99:SNV
chr2    47806983  .   A        AGTTC    35    PASS    .     GT:AD:DP:GQ:VAR_TYPE  0/1:500,500:1000:99:SNV
chr19   11133511  .   TTA      T        35    PASS    .     GT:AD:DP:GQ:VAR_TYPE  0/1:500,500:1000:99:SNV
chr2    21038086  .   C        A        35    PASS    .     GT:AD:DP:GQ:VAR_TYPE  0/1:500,500:1000:99:SNV
chr19   11113534  .   G        A        35    PASS    .     GT:AD:DP:GQ:VAR_TYPE  0/1:500,500:1000:99:SNV
chr2    21012365  .   A        C        35    PASS    .     GT:AD:DP:GQ:VAR_TYPE  0/1:500,500:1000:99:SNV
chr7    5997399   .   G        A        35    PASS    .     GT:AD:DP:GQ:VAR_TYPE  0/1:500,500:1000:99:SNV
chr19   11120188  .   T        G        35    PASS    .     GT:AD:DP:GQ:VAR_TYPE  0/1:500,500:1000:99:SNV
chr19   11116388  .   C        T        35    PASS    .     GT:AD:DP:GQ:VAR_TYPE  0/1:500,500:1000:99:SNV
chr2    47805638  .   G        A        35    PASS    .     GT:AD:DP:GQ:VAR_TYPE  0/1:500,500:1000:99:SNV
chr1    55057514  .   G        A        35    PASS    .     GT:AD:DP:GQ:VAR_TYPE  0/1:500,500:1000:99:SNV
chr2    47799092  .   T        C        35    PASS    .     GT:AD:DP:GQ:VAR_TYPE  0/1:500,500:1000:99:SNV
chr2    47799601  .   C        T        35    PASS    .     GT:AD:DP:GQ:VAR_TYPE  0/1:500,500:1000:99:SNV
chr2    47806206  .   A        G        35    PASS    .     GT:AD:DP:GQ:VAR_TYPE  0/1:500,500:1000:99:SNV
```

Class 2

"NGS"
- Library Prep
- Exome Capture
- Sequencing

Primary
- Base Calling
- QC

Secondary
- Read Alignment
- Duplicate Marking
- Local Realignment
- Variant Calling

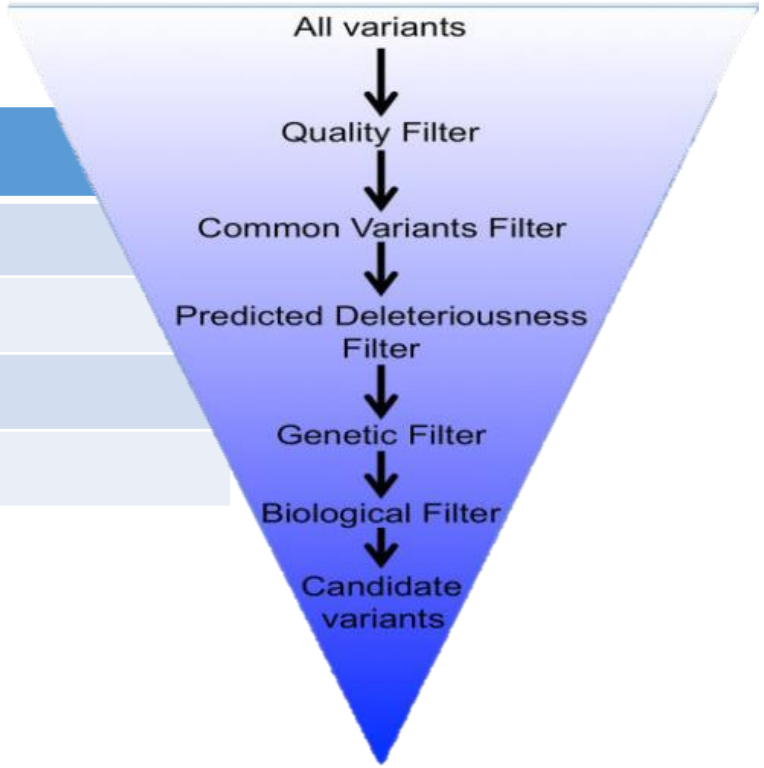
Unannotated VCF

Tertiary
- Variant Annotation
- Variant Filtering
- Variant Prioritization

Annotated VCF

WES Trio

Filter	Variants
None	125,746
DP>10, QUAL>20	25,541
MAF<0.01	1,263
Coding regions	551



Variant Interpretation Framework Summary

(11 questions to always ask from a variant)

Concept	Questions		ACMG Criteria
Allele Frequency	(1) Common or rare?		BA1, BS1, PM2
	(2) Variant Impact/Type	Loss of function	PVS1
		In-frame indel	PM4, BP3
Computational & Predictive Data	(3) In-silico predictions? Potential splicing impact?		PP3, BP4 BP7
	(4) Constraint metrics Gene/regional level		PP2, BP1
Functional Knowledge	(5) Residue/Domain? Hotspot?		PM1
	(6) Variant effect functionally studied?		PS3, BS3
Clinical Knowledge (published, or case/sample specific)	(7) Interpretation Databases - ClinVar		~~PP5 ,~~ PM5, PS1
	(8) Previously reported cases?		PS4, BS2, BP5
	(9) Phenotype specificity		PP4
	(10) Segregation? De novo?		PP1, BS4. PS2, PM6

Tier I: Variants of Strong Clinical Significance

Therapeutic, prognostic & diagnostic

Level A Evidence

FDA-approved therapy

Included in professional guidelines

Level B Evidence

Well-powered studies with consensus from experts in the field

Tier II: Variants of Potential Clinical Significance

Therapeutic, prognostic & diagnostic

Level C Evidence

FDA-approved therapies for different tumor types or investigational therapies

Multiple small published studies with some consensus

Level D Evidence

Preclinical trials or a few case reports without consensus

Tier III: Variants of Unknown Clinical Significance

Not observed at a significant allele frequency in the general or specific subpopulation databases, or pan-cancer or tumor-specific variant databases

No convincing published evidence of cancer association

Tier IV: Benign or Likely Benign Variants

Observed at significant allele frequency in the general or specific subpopulation databases

No existing published evidence of cancer association

Variant Interpretation Framework Summary

		BENIGN CRITERIA		PATHOGENIC CRITERIA			
Strength of evidence		Strong	Supporting	Supporting	Moderate	Strong	Very Strong
Odds of Pathogenicity*		–18.7	–2.08	2.08	4.33	18.7	350.0
Evidence Category and Corresponding ACMG/AMP Codes	Population Data	*BA1*+ BS1 BS2			PM2	PS4	
	Allelic Evidence & Cosegregation Data	BS4	BP2 BP5	PP1 →	PM3 PM6	PS2	
	Computation & Predictive Data		BP1 BP3 BP4 BP7	PP2 PP3	PM1 PM4 PM5	PS1	PVS1
	Functional Data	BS3				PS3	
	Other		BP6	PP4 PP5			

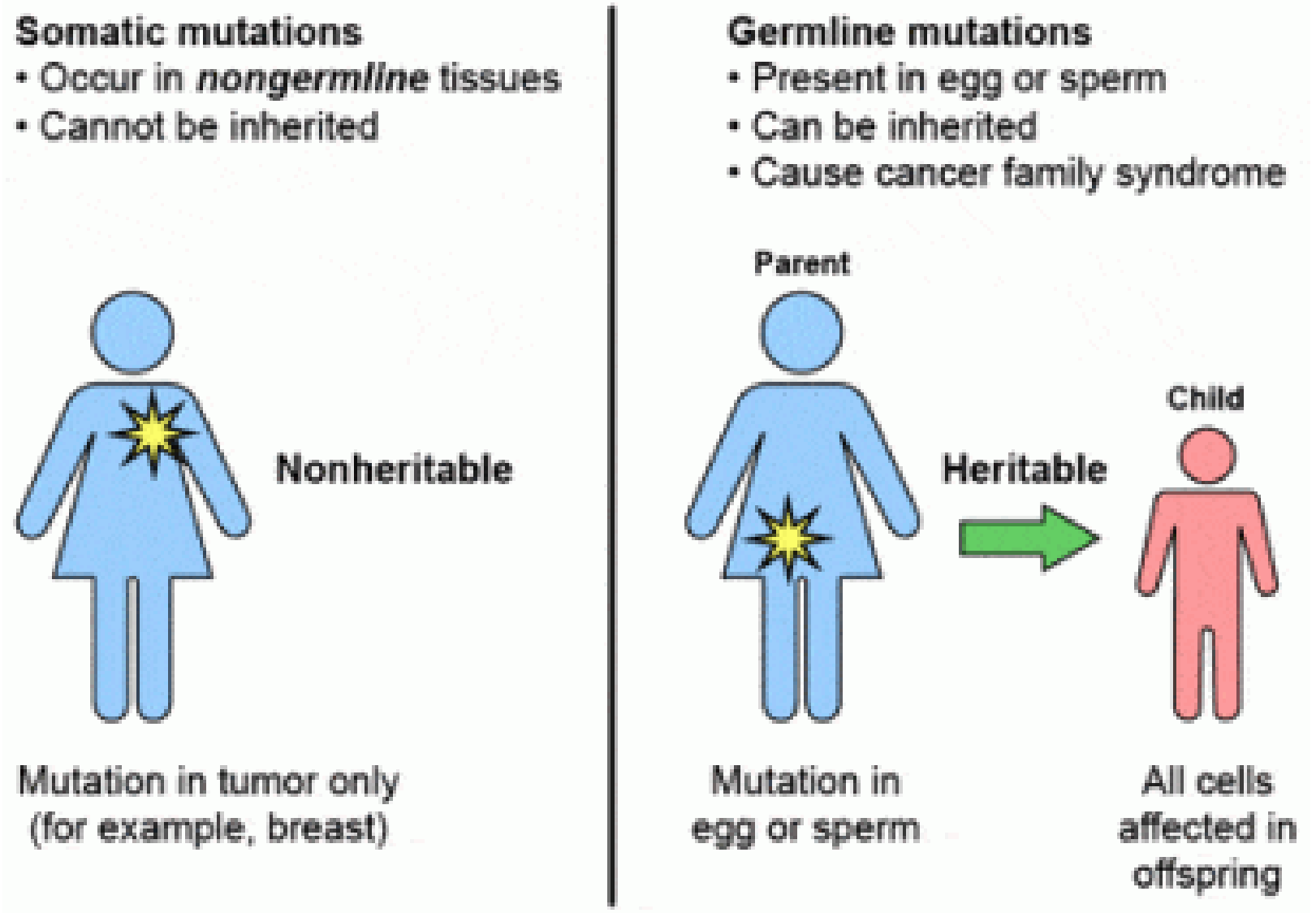
benign (B) or pathogenic (P) classification (first letter of code); Relative strength (second letter of code): VS very strong, S strong, M moderate, P supporting

https://www.sciencedirect.com/science/article/pii/S1098360021017664

Germline and Somatic Classification and Catalogue Differences

OncoKB

COSMIC
Catalogue Of Somatic Mutations In Cancer



OMIM®

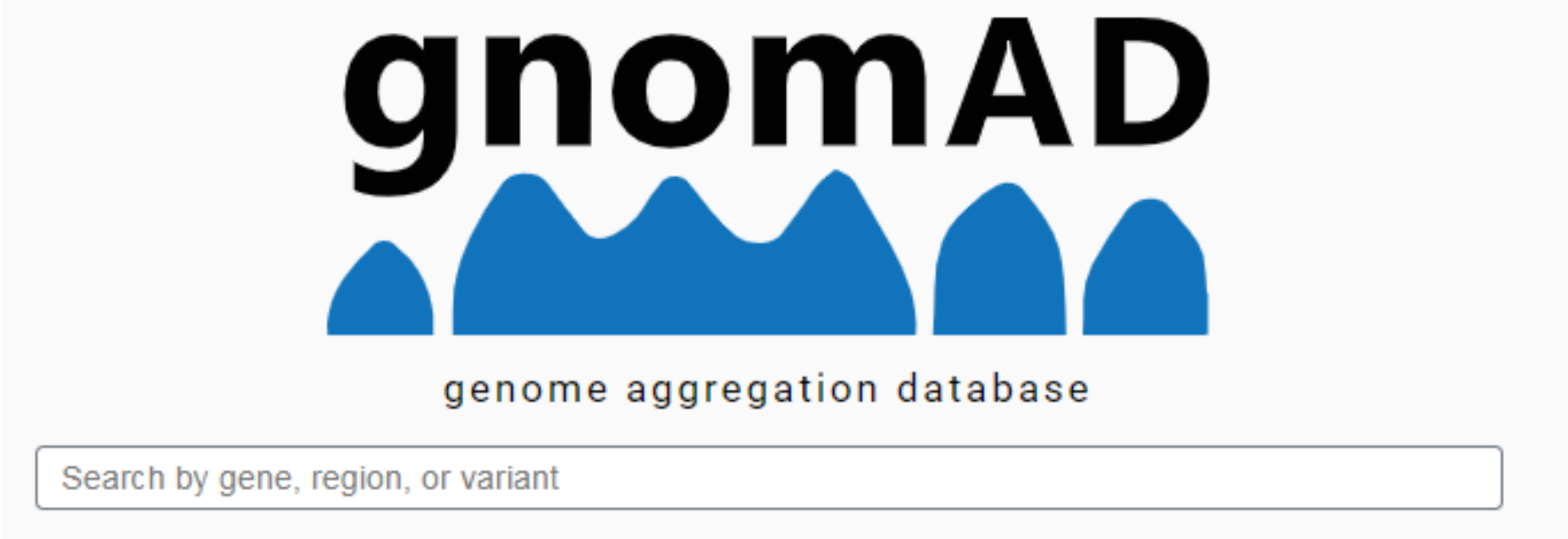
human phenotype ontology

ClinVar

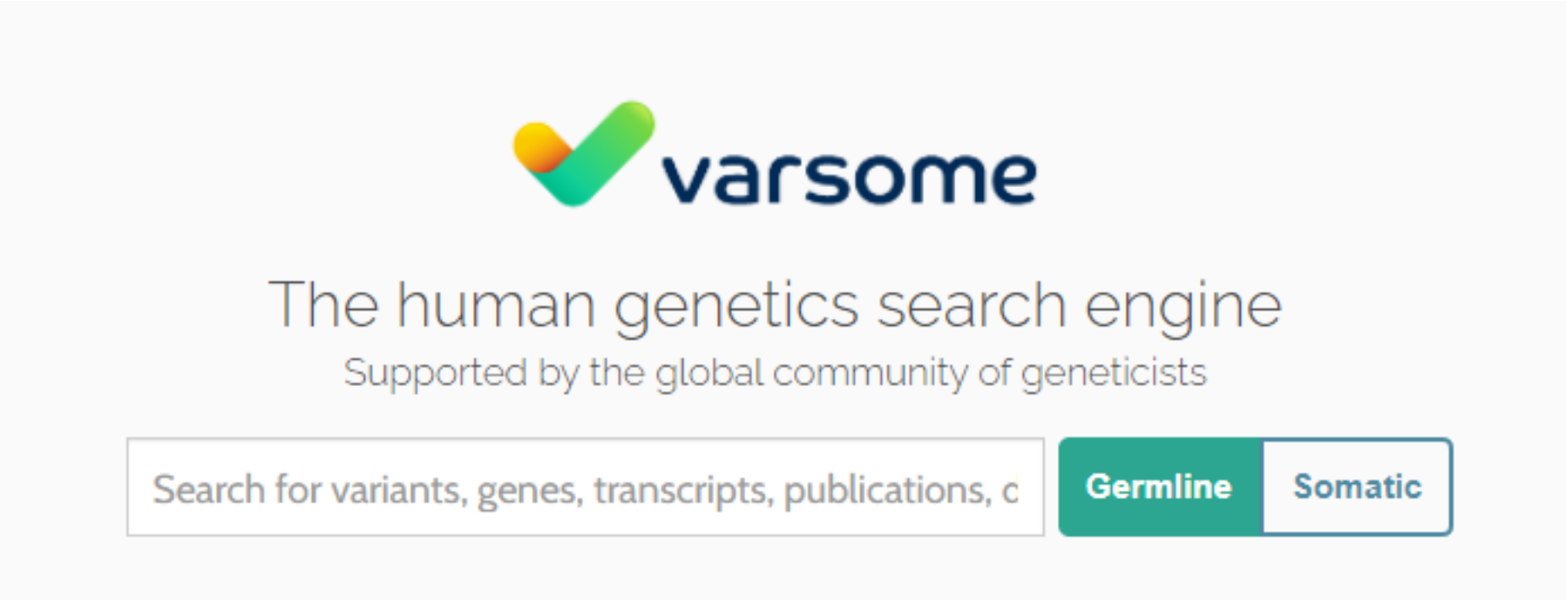
gnomAD browser

Exome Aggregation Consortium (ExAC)

125,748 exome sequences
15,708 whole-genome sequences
141,456 individuals



Please sign in to **Varsome** for the lab section



Background Reading

Multifactorial contributions to disease etiology are interrogated by statistical means (e.g., genome-wide association studies), but the clinical utility of this information is currently limited by its poor predictive power.

In contrast, Mendelian or monogenic disorders are characterized by rare variants in a single gene with a high impact on disease risk

assertion of variant pathogenicity for a Mendelian disorder implies causality, although this does not always correlate with disease manifestations in a given individual, due to incomplete penetrance and variable expressivity.

The American College of Medical Genetics and Genomics (ACMG), together with the Association for Molecular Pathology (AMP), established guidelines for reporting and interpreting sequence variation in an effort to standardize clinical evaluation of genomic information.

The original 2000 guidelines established five categories of classifications but provided little guidance on evidence selection and weighting.[3] Revision of these guidelines in 2007 incorporated a sixth category of variants: those associated with clinical symptoms, but that are unexpected or unknown to cause disease (e.g., risk variants)

The 2007 version also recommended that clinical laboratories utilize standardized variant nomenclature and include testing limitations in reports. These initial iterations were qualitative and assigned variants to categories based on certain features (e.g., reported in the literature or observed in other affected individuals).

Variability in the interpretation of variants between laboratories, in the setting of massively increased data generation due to next-generation sequencing, necessitated more thorough guidance and ultimately led to the development of a more structured approach for variant interpretation in 2015

updated guidelines establish a 5-tier classification system (pathogenic, likely pathogenic, uncertain significance, likely benign, or benign) and specify lines of evidence necessary for clinical interpretation. Critically, the guidelines include a relative measure of strength (stand-alone, very strong, strong, moderate, or supporting) for each piece of evidence for or against pathogenicity

Additionally, the 2015 guidelines present rules for combining evidence to make a given assertion, defaulting to variant of uncertain significance (VUS) when these rules are not met or conflicting evidence exists.

The National Institutes of Health (NIH) funded endeavors, such as the Clinical Genome Resource (ClinGen)[11] and ClinVar,[12] are developing curation tools that incorporate the ACMG/AMP guidelines to assist in variant interpretation and data sharing.

Population-level minor allele frequency (MAF) data is critical because disease-causing alleles for most Mendelian disorders are expected to be rare, and five ACMG/AMP criteria (BA1, BS1, BS2, PM2, and PS4) utilize these data. It is important to consider disease prevalence, penetrance, and genetic (locus and allelic) heterogeneity when applying any of these criteria, although such information is often unknown or inaccurate due to ascertainment bias. A MAF >5% in any global population is considered a “stand-alone” benign classification (BA1) for the vast majority of Mendelian disorders, with the exception of well-known founder alleles.[5] The most frequently applied ACMG/AMP criteria across 99 variants assessed by nine laboratories were PM2 (absent from control populations) and BS1 (MAF higher than expected for disorder).[7]

To date, Exome Aggregation Consortium (ExAC) and Genome Aggregation Database (**gnomAD**) are the largest and most ethnically comprehensive datasets of variant allele frequency. However, even these datasets do not represent all populations and it is important to consider population size and error in MAF estimates when assessing thresholds for BS1. Furthermore, inadequate population representation and lack of phenotypic details may cause difficulties in applying BS1 or BS2 (variant for highly penetrant condition seen in healthy individual) criteria. In addition, the data quality of any population allele frequency database should be evaluated for sufficient depth of coverage to ensure accurate MAF estimates.

Allelic evidence and cosegregation

Due to the Mendelian patterns of inheritance seen in most monogenic disorders, evidence of segregation in family members (or lack thereof) can inform variant interpretation. The ACMG/AMP guidelines include a number of criteria (PS2, PM6, PP1, and BS4) that apply to segregation evidence. The de novo occurrence of a variant is considered strong evidence of pathogenicity (PS2) when (1) maternity and paternity are confirmed, (2) the variant is in a gene associated with a condition consistent with the patient's phenotype, and (3) there is no past family history of disease (i.e., unaffected parents). When the first criterion is not met, the evidence is considered moderate strength (PM6). The second and third criteria apply for both PS2 and PM6, because all humans are expected to have approximately 44–82 de novo single-nucleotide variants (SNVs) of which 1–2 are expected to be exonic.[18]

Computational and predictive criteria

A large number of the ACMG/AMP variant interpretation criteria (PVS1, PS1, PM4, PM5, PP3, BP1, BP3, BP4, and BP7) are categorized as predictive or computational evidence.[5] These criteria relate to the type of variant in question and its predicted impact on the protein product based on knowledge about the protein's function, structure, and evolutionary conservation. Notably, PP2 (missense variant in a gene in which missense changes are rare) and PM1 (mutational hotspot/functional domain), which are categorized as "functional data," could also be regarded as "predictive" evidence (see Fig. 1). Together, these 11 criteria provide a way to predict variant pathogenicity by extrapolating from what we know about the functional and clinical impact of similar variants, with respect to both variant type and location within a protein

Functional criteria

Data from well-established functional assays showing a deleterious effect (PS3) or no effect (BS3) are considered strong evidence in the ACMG/AMP variant interpretation framework.[5] As with in silico tools, functional assays are poised to tackle the reclassification of many rare missense VUSs because they are not dependent on clinical data.[22,26] However, lack of guidance on what constitutes a well-established assay has resulted in subjective application of these criteria, and interlaboratory interpretation differences are not always resolved through data sharing.[8] Often, assays are performed in research laboratories, which may not meet clinical laboratory standards. For sufficient predictive power, functional assays require extensive reproducibility and experimental rigor, including benchmarking against multiple variants with definitive clinical interpretations as determined by genetic or other evidence.

While the 2017 ACMG/AMP sequence variant interpretation guidelines establish a foundation for uniform and transparent variant analysis, there is still room for improvement and refinement. The guidelines are expected to “evolve as technology and knowledge improve,” and many groups have published their experiences and their critiques and/or modifications to the guidelines