لجنة Dentin 2020

Genetics

Writer : Zain Aljabali

doctor : Osama alsmadi

corrector: Tasneem Amer

Clinical variant interpretation I

Remember:

Human Genome

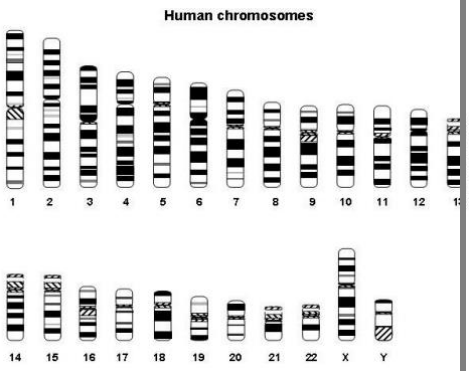
¬Consists of 23 pairs of chromosomes.

¬Chromosomes 1 through 22 are called autosomes.

¬The X & Y chromosomes are the sex chromosomes.

-Males are XY.

-Females are XX.



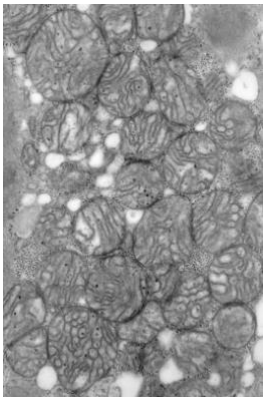
When we talk about Genome we mustn't forget another Genome in every cell in our body which is **Mitochondrial Genome (special)**:

¬Mitochondria have their own chromosome.

¬The human mitochondrial chromosome is a circular chromosome 16,569 base pairs long in every copy of the Genome .

¬Each mitochondrion contain several copies of its chromosome 3-10 or up to hundreds

¬Every cell contains many mitochondria for energy/ATPsynthesis , so different organs have different needs of energy and accordingly they differ in number of Mitochondria per cells .



Electron micrograph of mitochondria.

Back to the **human Genome** now ,

¬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).

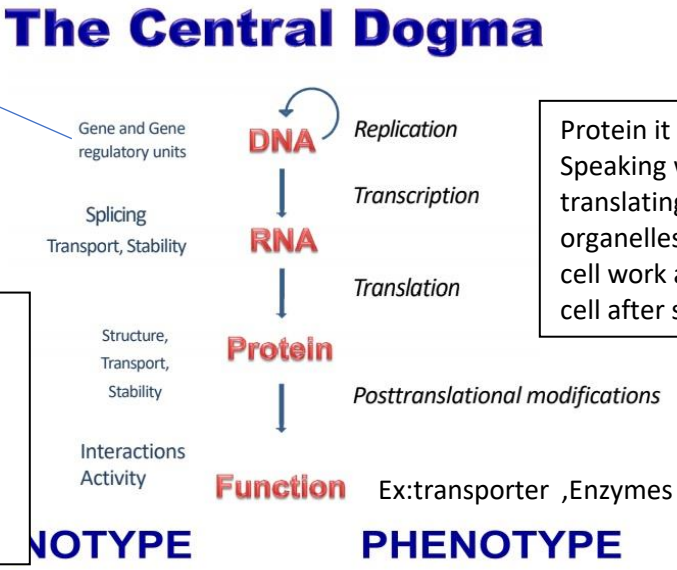
¬Humans are 99.9% identical at the DNA level.

¬Polymorphic variation (0.1%) between and within populations.

- This 0.1 percent make all diffrence in disease and health we Experiancing as indivdual in population

The Living Genome:

at the dna/ gene level they are subjected to variations that can affecting regulatory purposes or



Protein it a stable molecule generally Speaking which goes through translating from organelles to organelles and sometimes inside the cell work and also they ca exit the cell after synthesis

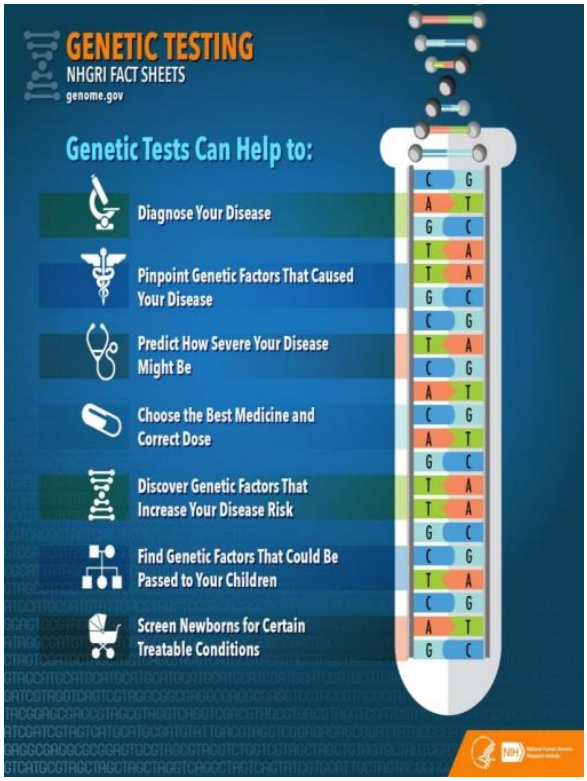
There's interaction between all of these aspects from the left side the genotypic similarities and differences and concerns on the right side as all of these processes they can eventually impact on any given phenotype of human being

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

◇Why would someone care to do genetic testing , for which purpose?
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 offspring .
Genetic testing identifies changes in levels of :

¬ DNA
¬ RNA
¬ Chromosomes
¬ Proteins
¬ Metabolites

These genetic testing are really crucial in number Of aspect such as diagnosis of a disease they can Be for example predict how severe a disease for Somebody and the information received from the Genetic testing can help in prescription of the Drugs for example what we call cell precision Medicine or smart medicine also for newborn Screening and can catch some risk early and lead To proper management



Genetic Testing Types:-

▪**Diagnostic testing** (symptomatic): to identify or rule out a specific condition, based on prior knowledge, we know what gene cause the mutation and simply look if it exists or not .

▪**Carrier testing:** to identify subjects heterozygous for a recessive condition. For example: cystic fibrosis found in certain population, we do the test to look for if the mutation existence or absent for someone at risk according to family history.

▪**Pre-symptomatic testing:** prior to symptoms appearance.

▪**Predictive testing** (susceptibility): identify mutations that increase risk of developing a disorder in a given individual.

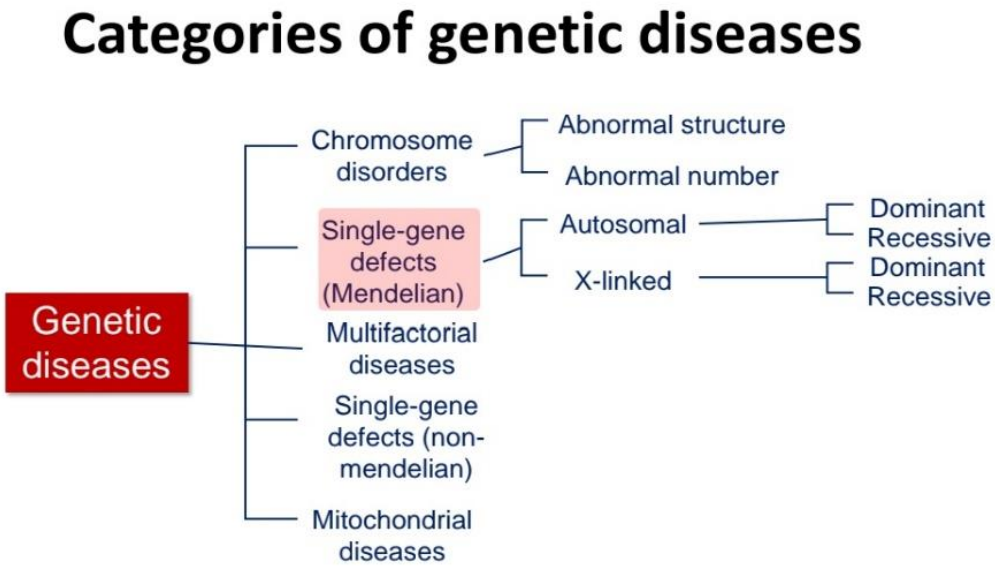
▪**Newborn screening**: just after birth, to identify genetic disorders that can be managed and treated early in life before reaching life-threatening situations .

▪**Prenatal testing:** to detect alterations in a fetus (in utero from amniotic fluid for example looking for genetic abnormalities).

▪**Preimplantation testing**: to detect genetic changes in embryos created in vitro fertilization by mother and father germ cells.

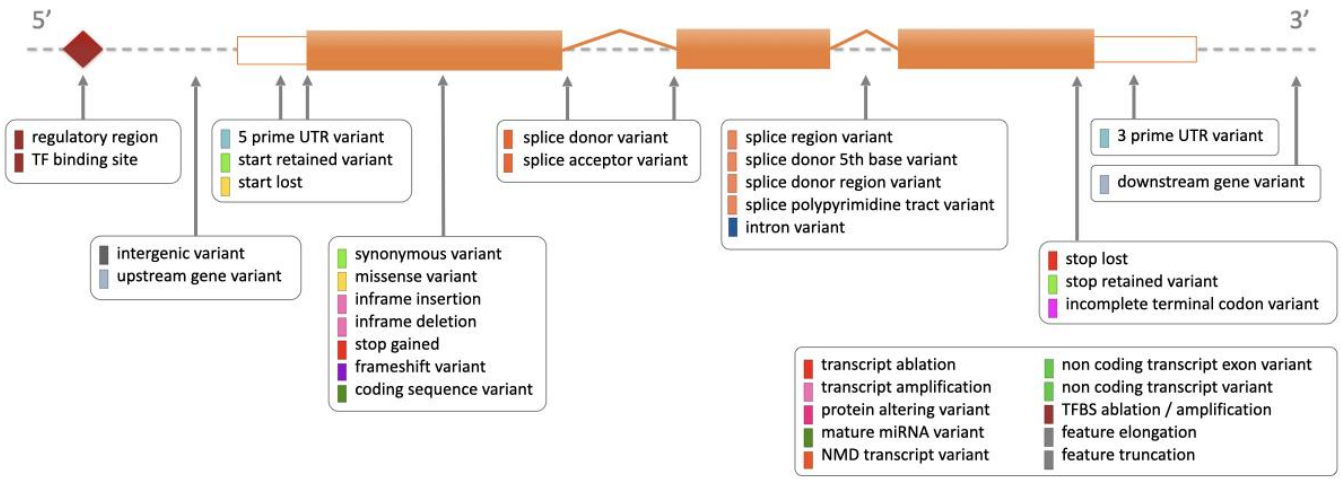
▪**Pharmacogenetic testing**: identification of the probable individual response (efficacy risk of adverse event) to drugs.

▪**Forensic testing:** to identify an individual for legal/criminal purposes as the paternal identity .



Genomic Variant Impact

The genetic testing all together will be looking at changes in sequencing of genome , here a slide showing the anatomy of way of a single gene for instance

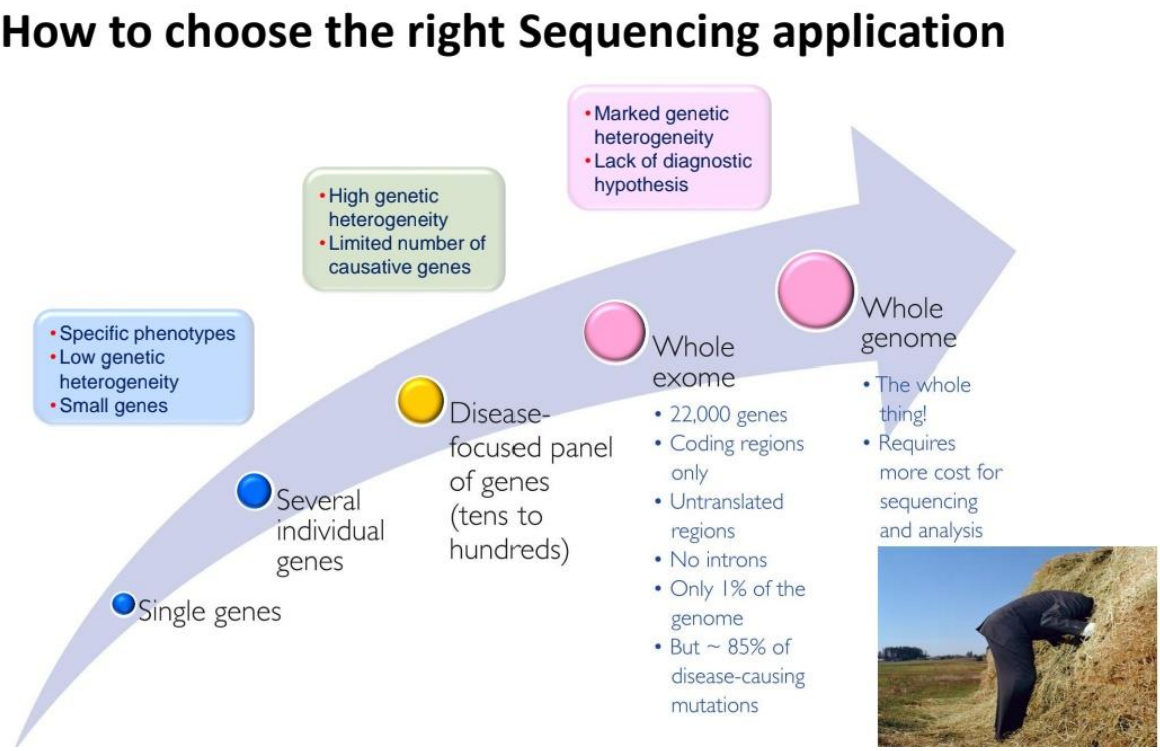


The orange part =exons , the dotting lines between =introns
Look from left to right (from 5' end to 3' end of the gene) ,we have different locations for example ; the untranslated region in 3' end and upstream from that there is a regulatory region for transcription factors binding site for initiation transcription.
And then we have 3 exons here ,and the introns in between .
The changes can happen anywhere within upstream or downstream the gene, where the mutation happen can carry impact or subsequence there will be,,
For example:
if it's a splice site, any changes on the donor or acceptor splice site "which is the first and last two base pairs of every intron ", there will be an interferences with the splicing mechanism and therefore this intron will not be excited out , meaning in RNA level this intron which supposed not to be present will being part of the transcript and will also impact later on the translation

If changes happen in the coding , they can be many variant types like we mentioned earlier synonymous meaning benign no changes, there is miss sence variance that can lead to amino acid change in frame and out of frame (insertion, deletions, stop gaining or loss frame shift changes.

In 3’ end is a critical part of the gene , it has a regulatory elements that are part of important in transcript modification and changes before generate a mature massage in the end .

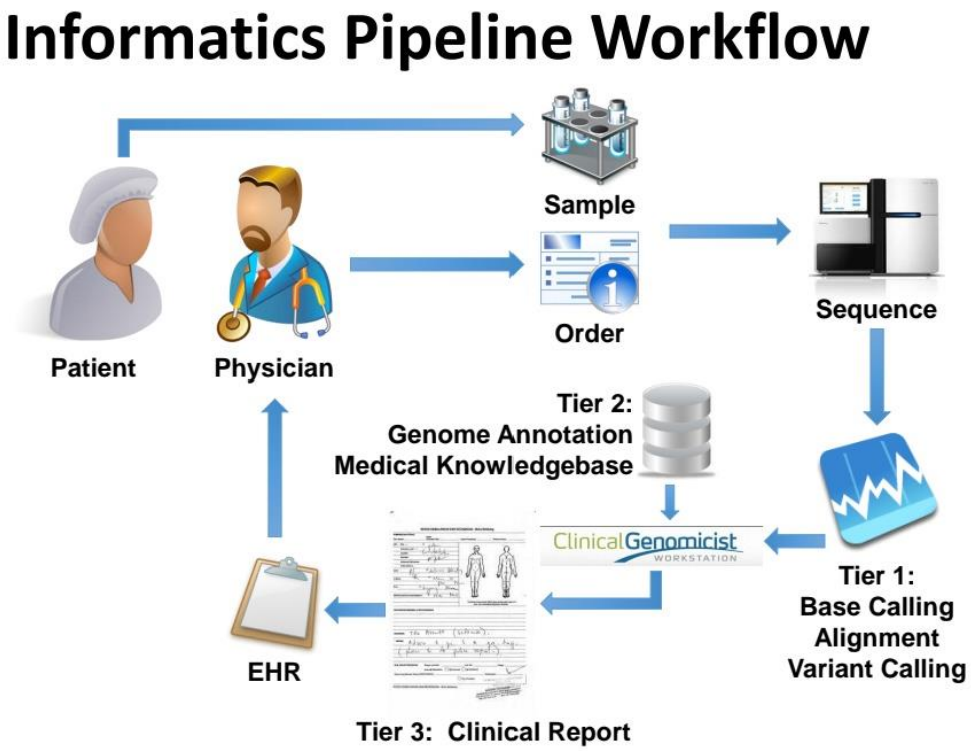
And changes anywhere can be critical and lead to loss of gene functions.



Left side:
You look at mendelian single gene disease, there will be a sequencing for example using the sanger sequencing, which looking at a small gene fragments that’s potentially the whole part of the mutation associated with the disease.
We also have a number of sets of genes, the fist two are associated with genetic diseases of known etiology (means we know the disease and the potential genes which can lead to it).

When we go higher in Complexity we have more number of genes panels , and we are now approaching more of multifactorial disorders such as Diabetes, hypertension , where there is need to look at tens of genes and sometimes at the whole exons or introns , so this will be associated with alot of Complexities and finding
When you look at the entire genome you looking 3 billions base pairs , زي الي بدور على ابرة بكومة قش

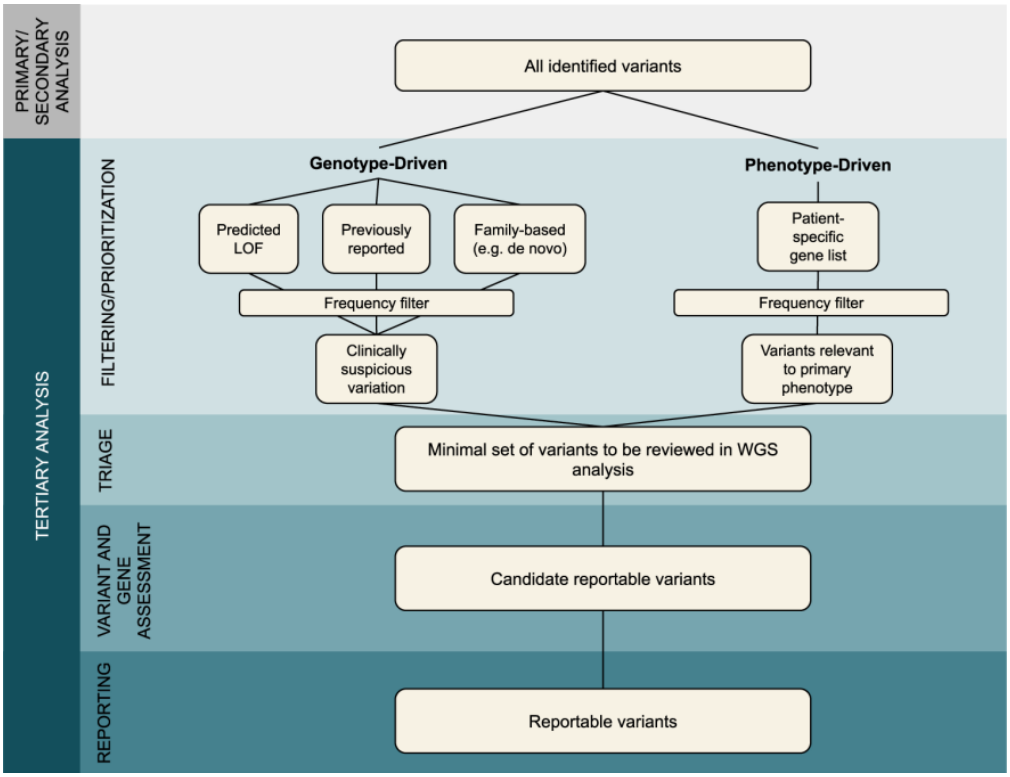
This type of studies not hypothesis based , but hypothesis generated meaning that you do a complete scan and sometimes you don't know what to expect, so the findings will guide you to generate hypothesis and explanation of gene relationship to a disease .
Remember in mendelian we had a prior knowledge about it .

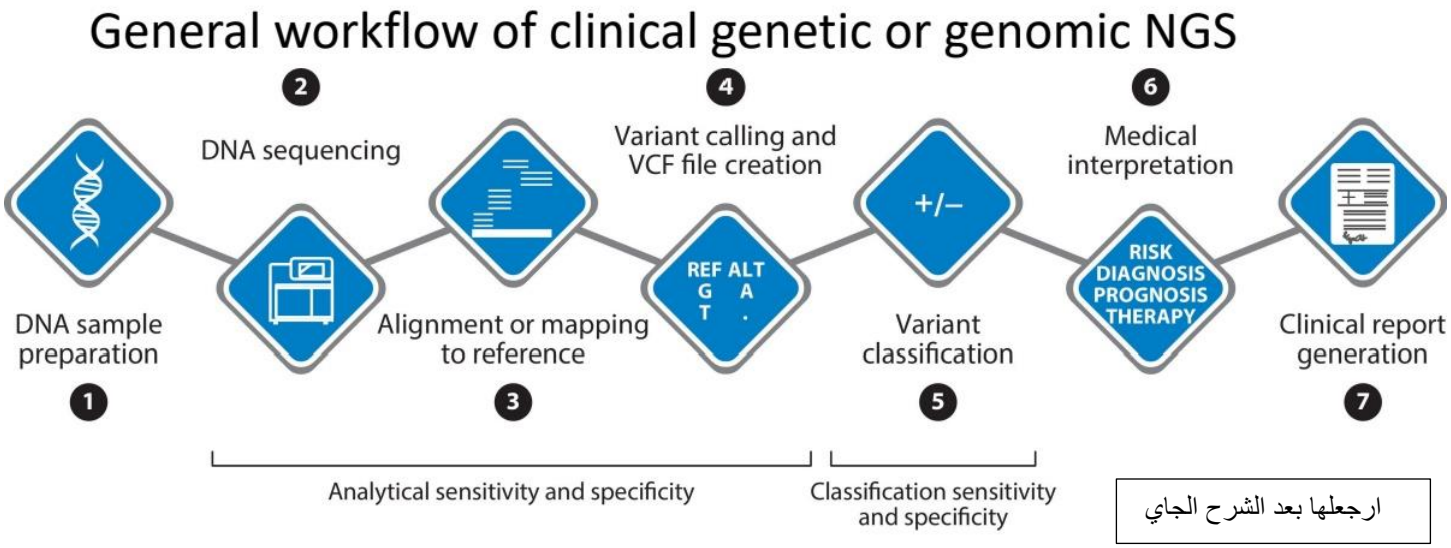


Sequencing the whole exon/genome depends on advanced technology and bioinformatics .
If you a clinician looking at a patient with a disease, there will be a decision to do a genetic testing (means collecting blood sample in order by the physician, the it go the lab for sequencing) and from that we get what called a **base calling alignment** of the reads that generated from the sequencing against a reference genome which they align to , and based on matches to the reference and differences from it there will be a variant calling , meaning the genetic variance will be determined , which called [**TIER 1 LEVEL**].
Then that information will subjected to further analysis and algorithms using automated software pipelines . and using multiple databases , this computing exercise will lead to generation of **[TIER 2]** which going from base calling to

Annotation ,meaning you have variant now you know the impact of the variant into functionality prescriptive

You put all this together to make a clinical report then goes back to the clinician to decide how to use that information to management of the patient.





ارجعلها بعد الشرح الجاي

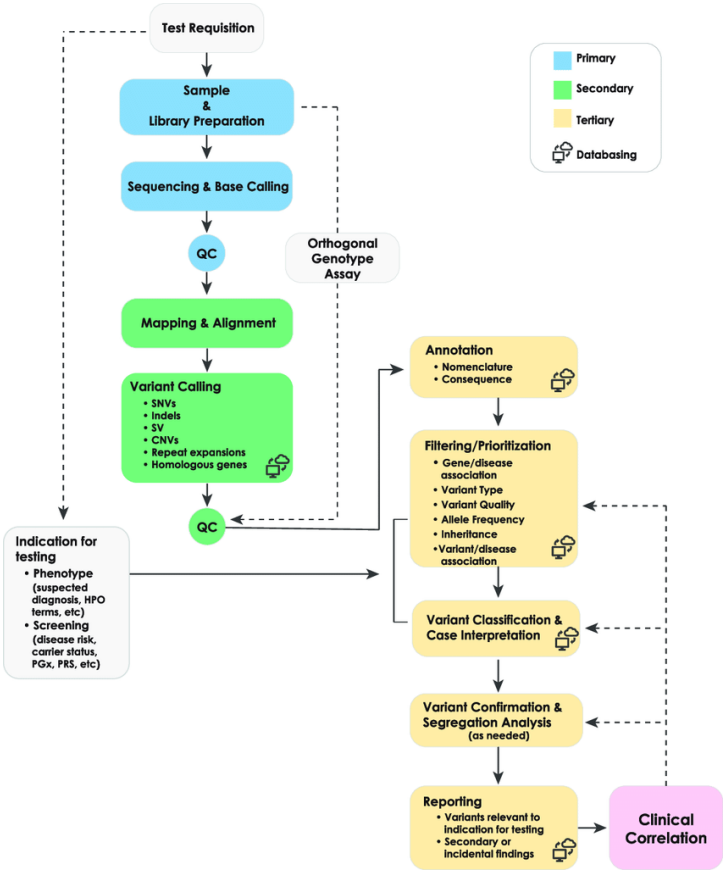
Giles HH, et al. 2021
Annu. Rev. Genom. Hum. Genet. 22:285–307

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

The data used is high volume data from which we will use genetic information connected to phenotypes and eventually we will have a list of variance that will be linked to patient information (clinical information) and phenotypic features , And all things put together.



And in the picture below; number Of databases that are one depends On, in this exercise of Annotation and variant calling .

Common framework and criteria for germline variant classification globally recognized and they are used eventually in every genomic labor clinical utility

AMP Association for Molecular Pathology

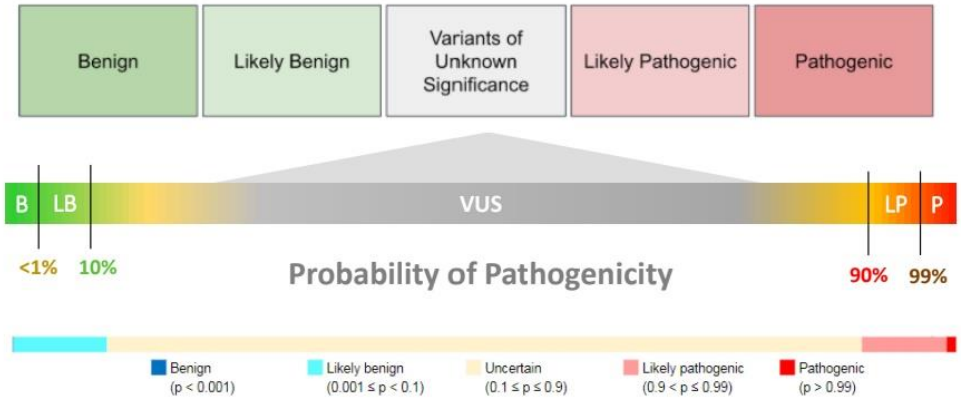
ACMG American College of Medical Genetics and Genomics *Translating Genes Into Health®*

cap

Most famous one

These sources have a vast amount of databases, diseases, references genomes, and millions of data sets from individuals collected which by those they have generated their databases which one can compare to it .

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.



these are sophisticated software that are developed by data engineer and biologist and programmers which they are able to take next generation sequencing which is a big data, millions of data sets that taken by software use algorithm built-in algorithm automatically and able to take all this data align the reads that come from a sequencers to a reference genome and compare , and eventually generate a list of variance , but in this process the software that are trained to compare the test or experimental data against reference genome data , will be able to group these genetic variations from benign to pathogenic (in between likely benign variant of non-significance Or known as the VUS or likely pathogenic) , all these classifications are based in algorithms and also associated with a score of pathogenicity ;

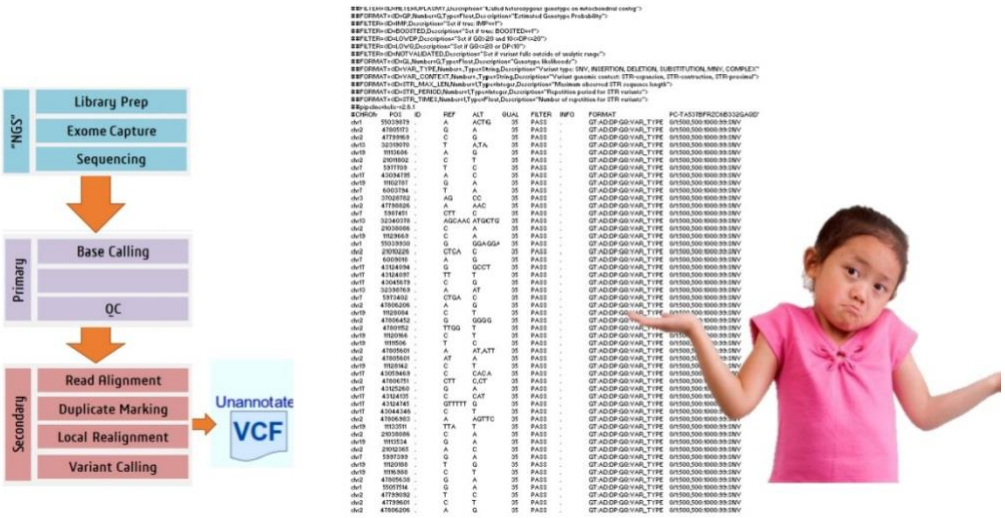
benign have a score under 1 percent ,

likely benign up to 10 percent,

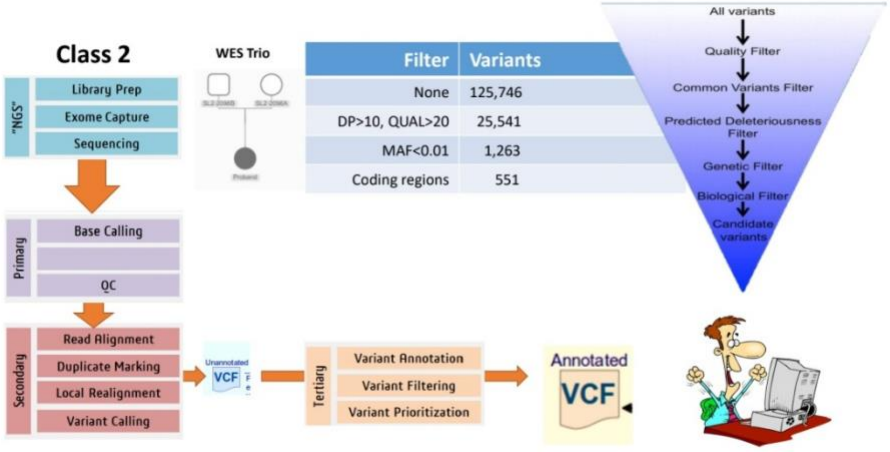
which is the vast majority of variance one comes across when they analyse the big data of sequencing they are of this class type ,

Between 10 -90 % are with pathogenic ,

However, if you approach the pathogenic and likely pathogenic this lesser of percent after this analysis.



if you work closely with someone with experience in next generation sequencing, and big volume data analysis for instance . the whole genome or exon that have analysis, you will go through making library (visiting laboratory) to prepare from the patient DNA , generating fragments that can go to the sequencers so they can sequence and read , also another steps such as **exome capture** followed by **sequencing** , based on that you will have a base call reads that can go into a **primary analysis** with quality checks that the algorithms applies and then will be a **secondary analysis** which means aligning these reads against the reference genome , up to this point you have a list of variance look at so many (above these are real results of somebody with some disease, impossible to make it manually, however thanks again to the technical advancements and algorithms and programs made by specialists, which can take that variant calling and apply it to tertiary analysis where there will be alot of filtering applies [you go from all variant to after quality checks you do some filtering =means it's an elimination process that you can go from hundreds or thousands of variants down to small sets of variants that are annotated and there will be connected to the disease of interest for some sort of link between phenotype and genotype, and of course with that knowledge you will be able to generate the clinical report

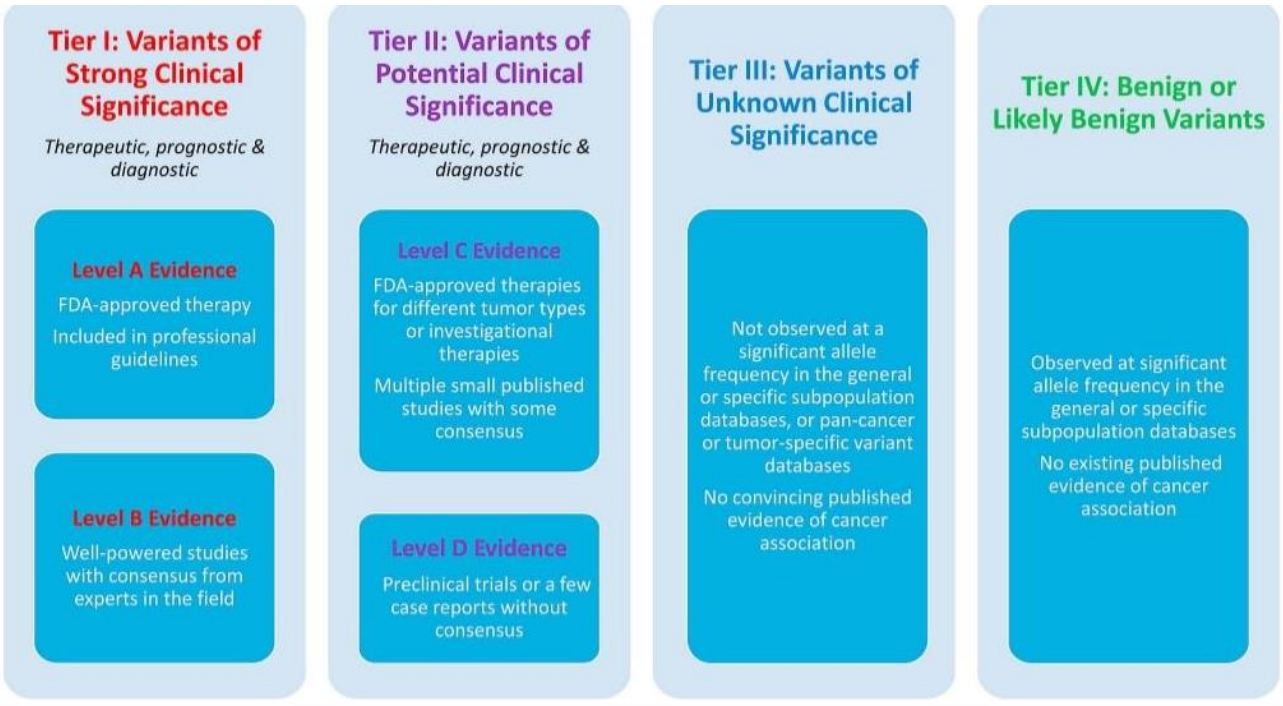


Variant Interpretation Framework Summary

(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 In-frame indel	PVS1 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

Databases using a sequential questions to be able eventually to annotate the variants and give them a functional meaning.
By going through it's automated process they apply these questions for example; at the allele frequency level , the question will be * common (= more likely non-pathogenic and exist at higher rate in the population)or rare?
Going through all these give a classification below ,



With all that filtering and using acmg and looking to the variants, the one is able to group that variance into 4 main groups; based on level of evidence

1.Variants of strong clinical significance
2.variants of potential clinical significance
3.variants of unknown clinical significance
4.bengin or likely benign variants

Variant Interpretation Framework Summary

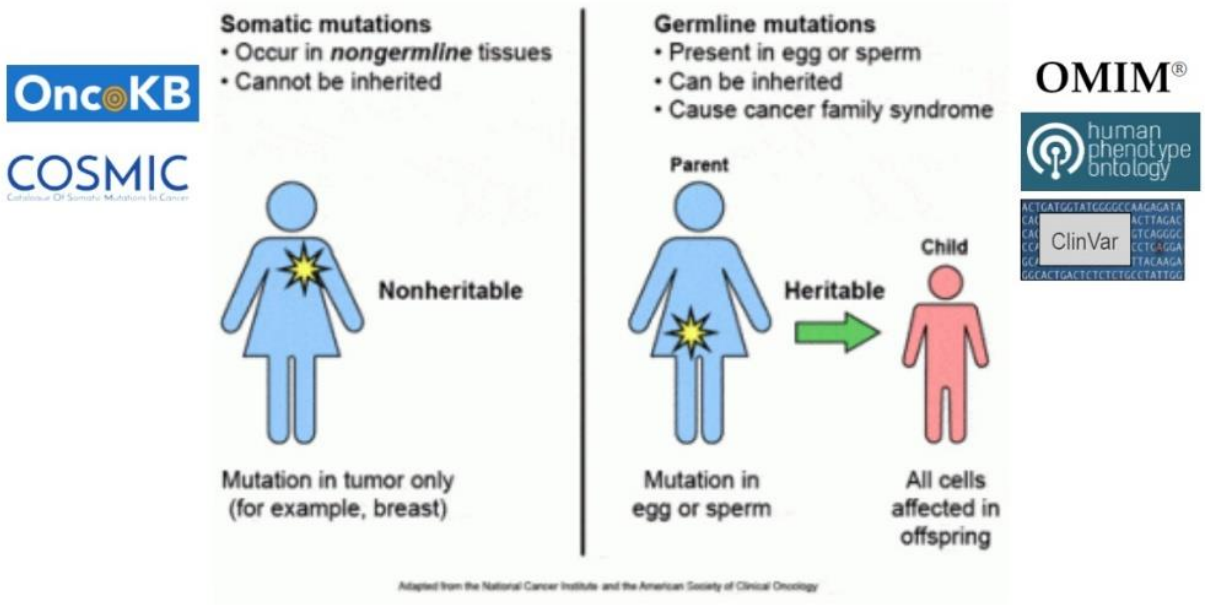
		BENIGN CRITERIA		PATHOGENIC CRITERIA			
Strength of evidence		Strong	Supporting	Supporting	Moderate	Strong	Very Strong
Odds of Pathogenicity*		–18.7	–2.08	2.08	4.33	18.7	350.0
Evidence Category and Corresponding ACMG/AMP Codes	Population Data	*BA1*+ BS1 BS2			PM2	PS4	
	Allelic Evidence & Cosegregation 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

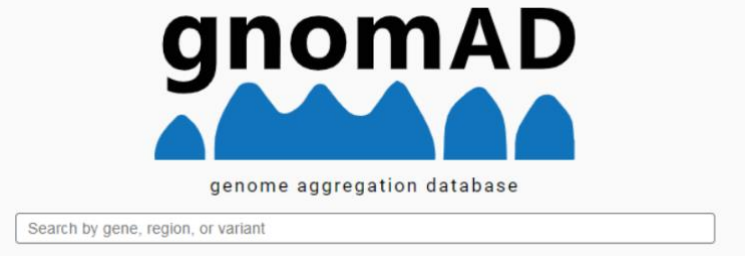
https://www.sciencedirect.com/science/article/pii/S1098360021017664

Example from the table ; PVS1=pathogenic, very strong.
Another reference to guide classifications known as clinical databases,

Germline and Somatic Classification and Catalogue Differences

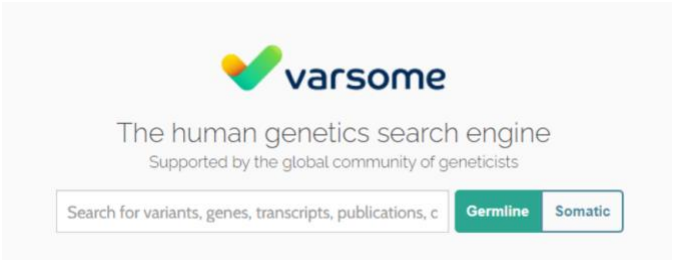


gnomAD browser
Exome Aggregation Consortium (ExAC)
125,748 exome sequences
15,708 whole-genome sequences
141,456 individuals



Another database you can use , that collect hundreds of thousands of genomic data points from individuals and will be reference to someone trying to make a knowledge or a judgment on a given variants what it means clinically and so on.

Another helpful website, also it's open Source meaning you don't have an Account but you can put Annotation, And a transcript name, nuclear number And changes in the search engine and the Website will tell you all knowledge of this variant in relation to pathogenicity, drug, and clinical trials and the published work that related to it , etc.



DONE