

# Giới thiệu Giải trình tự DNA (DNA-seq) và ứng dụng trong lâm sàng

Ngày 15 tháng 12 năm 2024

TS. Lưu Phúc Lợi

Email: [luu.p.loi@googlemail.com](mailto:luu.p.loi@googlemail.com)

Zalo: 0901802182

# Nội dung

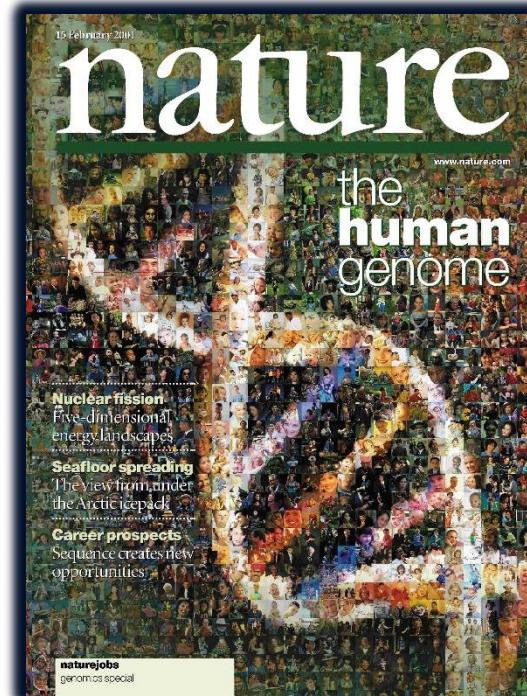
1. Giới thiệu về Dự án bộ gen người
2. Giới thiệu kỹ thuật giải trình tự thế hệ mới (NGS)
3. Quy trình **XÉT NGHIỆM** gen bằng phương pháp giải trình tự thế hệ mới
4. Năm ví dụ về **XÉT NGHIỆM** gen cho **BỆNH DI TRUYỀN** và **UNG THƯ**
5. Giải trình tự gen thế hệ mới trong nông nghiệp và môi trường

# Giới thiệu về Dự án bộ gen người

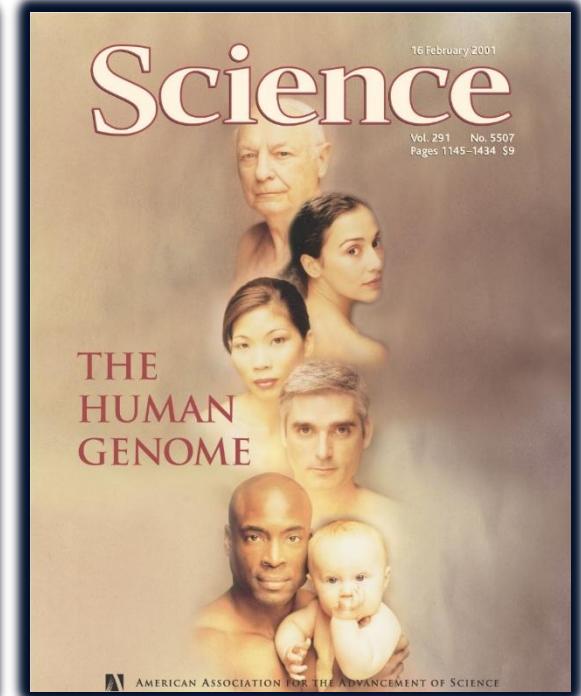
# Dự án hệ gen người HGP (Oct 1990 - April 2003)

- Năm 2003, dự án hệ gen người HGP đã tạo ra một chuỗi trình tự gần 3 tỉ nucleotide chiếm hơn 90% bộ gen người.
- Đây là chuỗi gen hoàn chỉnh nhất có thể đạt được với công nghệ giải trình tự DNA thời điểm đầu những năm 2000.

=> Sự phát triển của công nghệ giải trình tự thế hệ mới (NGS).



HGP Paper



Venter/Celera Paper

# Human Genome Build

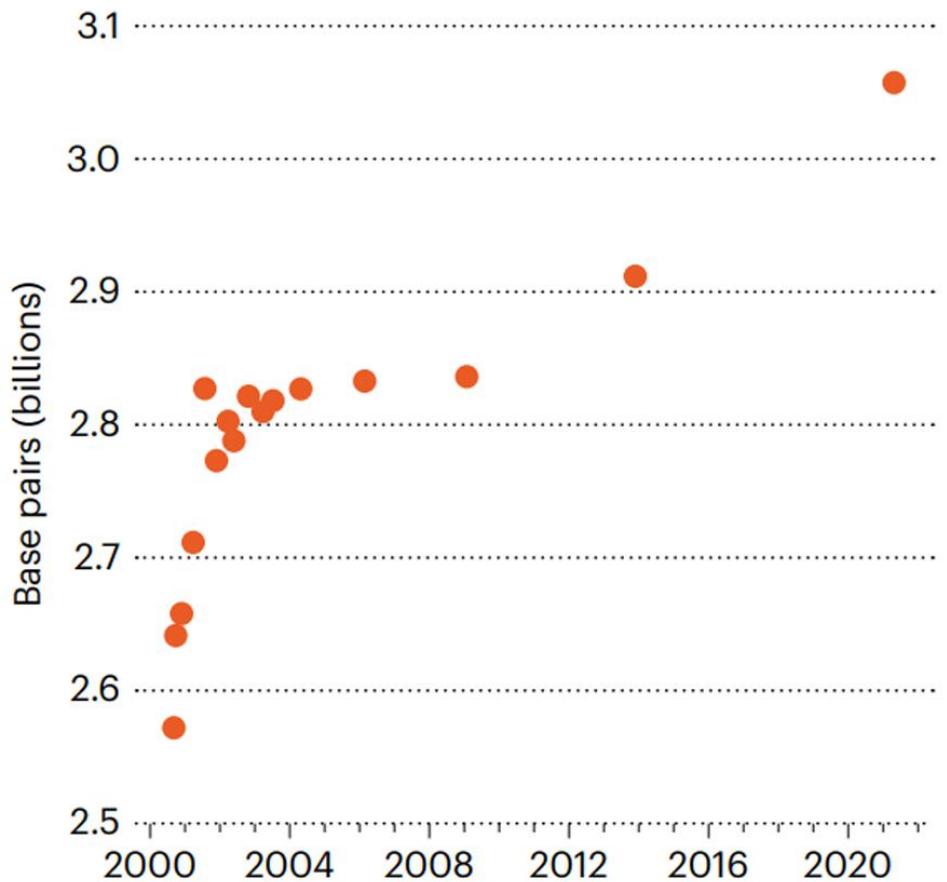
Human	hs1	Jan. 2022	T2T Consortium CHM13v2.0	Available
	hg38	Dec. 2013	Genome Reference Consortium GRCh38	Available
	hg19	Feb. 2009	Genome Reference Consortium GRCh37	Available
	hg18	Mar. 2006	NCBI Build 36.1	Available
	hg17	May 2004	NCBI Build 35	Available
	hg16	Jul. 2003	NCBI Build 34	Available
	hg15	Apr. 2003	NCBI Build 33	Archived
	hg13	Nov. 2002	NCBI Build 31	Archived
	hg12	Jun. 2002	NCBI Build 30	Archived
	hg11	Apr. 2002	NCBI Build 29	Archived (data only)
	hg10	Dec. 2001	NCBI Build 28	Archived (data only)
	hg8	Aug. 2001	UCSC-assembled	Archived (data only)
	hg7	Apr. 2001	UCSC-assembled	Archived (data only)
	hg6	Dec. 2000	UCSC-assembled	Archived (data only)
	hg5	Oct. 2000	UCSC-assembled	Archived (data only)
	hg4	Sep. 2000	UCSC-assembled	Archived (data only)
	hg3	Jul. 2000	UCSC-assembled	Archived (data only)
	hg2	Jun. 2000	UCSC-assembled	Archived (data only)
	hg1	May 2000	UCSC-assembled	Archived (data only)

# A COMPLETE HUMAN GENOME IS CLOSE: HOW THE GAPS WERE FILLED

Researchers added 200 million DNA base pairs and 115 genes – but they've yet to finish the Y chromosome.

## COMPLETING THE HUMAN GENOME

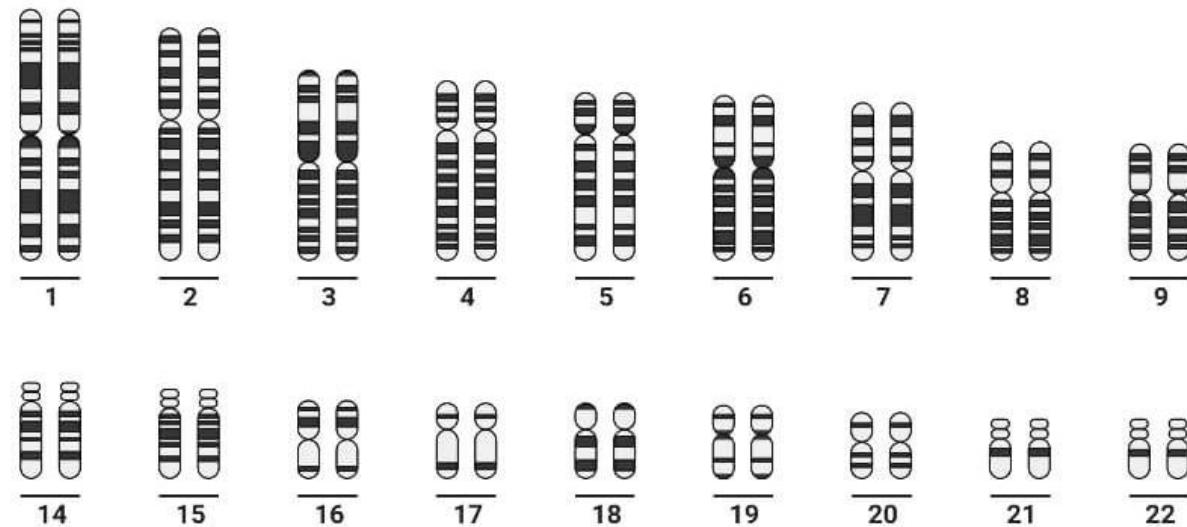
Researchers have been filling in incompletely sequenced parts of the human reference genome for 20 years, and have now almost finished it, with 3.05 billion DNA base pairs.



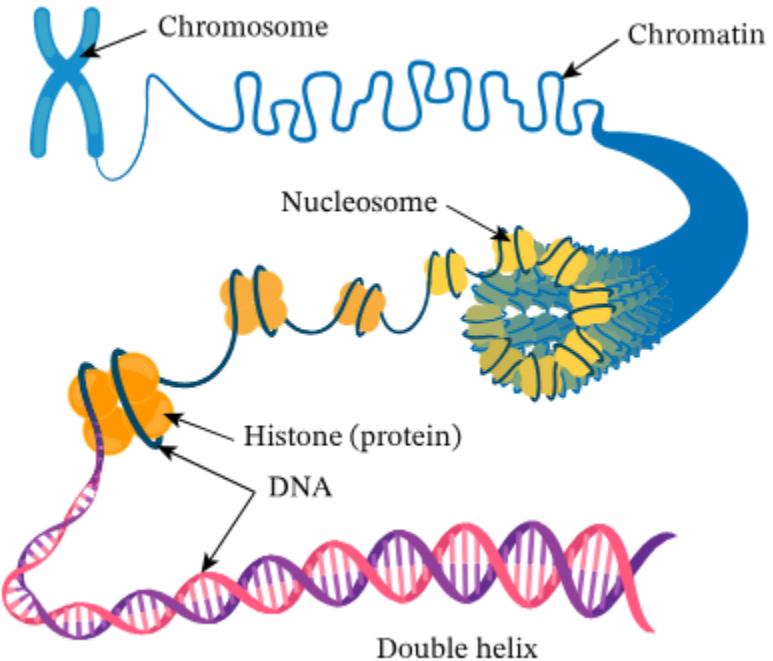
0.3% of sequence might still have errors. Includes X but not Y chromosome. Count excludes mitochondrial DNA.

SOURCE: ADAM PHILLIPY

# Human Karyotype



23  
X Y      or      X X



# Hệ Gen người



48541 agcccttcaa agaaatgttc tcagcaggca tggagccag gacttgctcc ctttgggtag  
48601 agagccgggt tgaaggtgac tgaagtaaa tgggacagta gaggccgggg gggtgttag  
48661 ttccctggagg tgggggtgt gggAACCTGC tgggtactga gatgcacccc tgccagttct  
48721 gcctgaagat ttgaggcggg gggcaggggg gcggagtgaa gtcattttac tggtaagtaa  
48781 ttttaaacct ttaaatatta aagcaaacgt ggatatgtaa tgaatgaaat tcattctgga  
48841 ataaaaaaatt cacgtgatgt taaaaataa cacggggctt cagagggac tttctggctg  
48901 gcagcagact ccagattccc agggccctg caccctcctc tgccacagg gcacctaatt  
48961 ggagaagggt tgggaggaga gccaggccgg agtcagagca cactggtgac tccacatttg  
49021 cagcgtgccc tgcctctc ctgaggctt gcaacgtca atatgtaag caaactcccc  
49081 ctgtccccgt ccagttctg aggacaagag ccaccacctg tagcaaataa agacccagca  
49141 acccttgcac tcatcttgcg gagttctgg aatcagaggg tagccacatc gctgagaggt  
49201 ggagtgaagc actcggtga aaaggtacaa ggaagtcaagg acaggagtg tggggacatc  
49261 acctagacaa tgacagagaa gaggggcaca gccgagttag gggagagggg cccgcagtcc  
49321 tacatccctt ggcctgaagc acgttccagg gcagaagaa aaacactgtc tttgggtcc  
49381 aagagacctg agttcaaatt ctggctccac cactgaccac ctgtgtaaacc ttgaactgtc  
49441 gctgcctgaa cctcaggttc ccctctaaa aatagaggag aaaaggatgc atttctcctt  
49501 gcccctgtga gaacaaaatg gtcaagcac caaggagct cagcaaagggt cgggcctgccc  
49561 cccgcctggc caaaccttcc ctcttcaggaa ggccacggca accgttagtt gacagaagag  
49621 cagcacccctt atttaatgtct tcccagcatg tgccttgag caagtccacct aacctctctg  
49681 ggctgcttcc tcattggaa aatatggctg ccagtaaaac ctgcctgtc cacctctgg  
49741 ggcacttggc aaacagcaaa agagtccaaa tgcgtcggctt gggcaggcgc cagtggctca  
49801 tgcctgtat cccagcaatt taggaagcca aggtggccgg atcaccttag gtcaggagtt  
49861 tggacccaggc ctggccaaca tggtaaacc ttgtctctac aaaaatacaa aaattagccg  
49921 ggcacatgtgg cgggtgcctg taatcccagt tactcggtgg gctgaggca gagaatcgct  
49981 tgaacccggaa agggaaagggt tgcgtcggc caagattgtg ccactgcact ccagcctggg  
50041 caacagacgc agactctgtc taaaaaaaaaaa aaaaaaaaaaa aaacaatgc gagctggctg  
50101 tggaaaaaac ctgttccact gcaggcccgc gtgtccacca ggctgggtg caggcctatg  
50161 gggggggggc ccagcatcag cctctcagca gcccctggag gggggccgc tcccgtgccc  
50221 ctctgtgtt ggtatgttcc tagcccaagt ccttagttac acctgcgtc gctggccctc  
50281 tcaggagagg cccagggtga ggaggagcat ggttaaagggt aagctgattt ggaagtgc  
50341 tggggaaaa gcaactccctt gcacatttggaa ggaaccgaga aagactgacc ccgaggacag  
50401 cagccagcat ggccttccctt gggagccat gttggggat tcctgtcgtca gccaaggctc  
50461 agcccttgc tgcgtcagggtg ctggccttgg cctttcccc tcccatgcag gggcaggg  
50521 gagatggctt ctgaggaccc tttgcgtcgg tggccctggg aatagattt ccagggagct  
50581 tttaaaggcgc tgagtgtgtc atccagctaa gcctggggaa ggagcttggc tcaggcctgt  
50641 acaggtgtga cagggatggg gactgggaag taagagatga aaccctggct ggaggctgt  
50701 agttccaca gccagcgtc gacaggagggt gtccagat acccactgtt gccctcacca

# Dữ liệu hệ gen cho quần thể



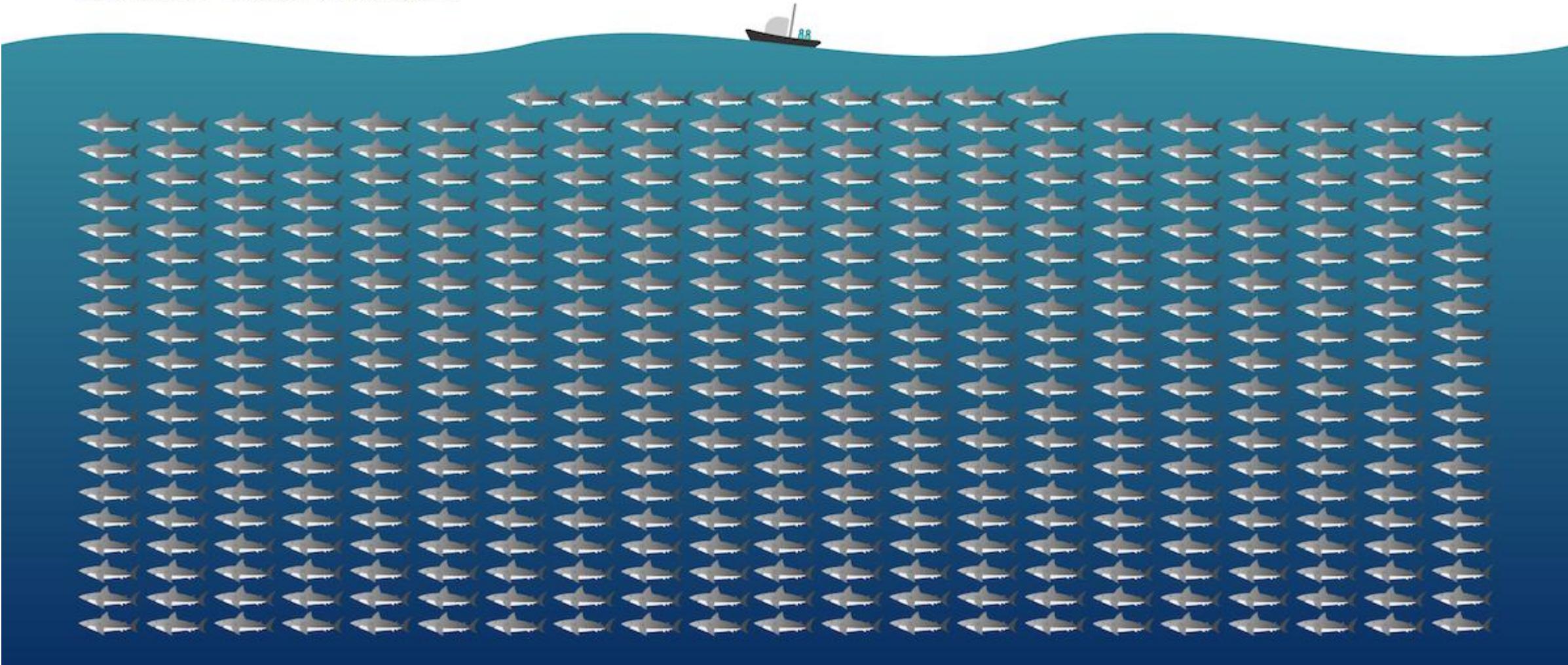
48541 agccctcaa agaaatgttc tcagcaggca tggagccag gacttgtcc cttgggttag  
48601 agagccgggt tgaaggtgac tgaagtaaa tggagacata gaggcggggg ggtgtgttag  
48661 ttctggagg tgggggtgtt gggacactgc tigtactga gatgcacccc tgccagttct  
48721 gcttaaagat ttgaggcggg gggcaggggg gcggagtgaa gtcattttac ttgttaagtaa  
48781 tttaaacct ttaatatta aagcaaacgt ggatatgtaa tgaataaaat tcattctgaa  
48841 atgaaaattt cacgtgtatg taaaaataaa cacggggctt cagagaggac tttctggctg  
48901 gcagcagact ccagattccc agggccctg caccctctc tgcccacagg gcacctaatt  
48961 ggagaaggta tgggaggaga gccaggccgg agtcagagca cactggtagc tccacatttg  
49021 cagcgtgccc tgccctctc ctgaggctt gcaacgtgca atatgtctaa caaactcccc  
49081 ctgtccctgt ccagttctg aggacaagag ccaccaccc tagcaataaa agaccaggca  
49141 acccttgac tcatctttgt gagtctctgg aatcagagg tagccacatc gctggagaggt  
49201 ggatgtaaagc actcgggtga aaaggatcaa ggaatcggg gacaggatgt tggggacatc  
49261 acatggaaa tgacagagaa gggggcaca gccggatgt gggaggggg cggcggatcc  
49321 tacatccccct ggctgaaagc acgctccagg gcagaaggaa aacactgtc tttgggtcc  
49381 aagagacctg agttcaattt ctggctccac cactgaccac ctgttaacc ttgaaactgt  
49441 gctgctgaa cctcagggtt cccttctaaa aatagaggag aaaaggatgtc atttccctt  
49501 gcccctgtga gaacgaaatg gtcaagcac caaggagct cagcaaaagggt cggccctgcc  
49561 cccgcctggc caaaccttc ctctttagga ggccacggca accgtatgtt gacagaagag  
49621 cagcacctt attaatgtc tcccgatcg tgcctttaga caagtccact aacctctgt  
49681 ggctgttcc tcattggaa aatatggctg ccagtaaaac ctggccctgc caccctctgg  
49741 ggcacttggc aaacagcaaa agatgtccaaat tgcgtggatc gggccaggcg cagtggctca  
49801 tgcgtgtaaat cccagcaatt taggaaggcc aggtggggcg atcaccctgat gtcggaggt  
49861 tgacccaggc ctggccaaaca tggtgaaacc ttgtctctac aaaaatacaa aaattagccg  
49921 ggcatgtgg cggtgtcc tcattccatc tactcgggg gctggaggaa gagaatcgct  
50081 tgaacccggc agggggatgt tgcatgtggc caagattgtt ccaactgact ccggccctgg  
50041 caacagagcg agactgttc taaaaaaaaaaa aaaaaaaaaaa aaacaatgca gagctggctg  
50101 tgtaaaaaaaat ctgttccact gcaggccca gtgtccacca ggctgggtt caggccatag  
50161 gggggggggc ccagcatcg cctctcgca gcccggggag gccccggca tccctggccc  
50221 ctgtgggtct ggatgtgttc tagcccaagt cctaggttac acctggcgcc gcttggccctc  
50281 tcaggagagg cccagggtga ggaggagcat ggttaagggtt aagctgttgg ggaagtcgc  
50341 tggggaaa gcaactcctt gcacattggaa ggaaccgaga aagactgacc ccgaggacag  
50401 cagccagcat ggccttcctt gggagccat gttggggat tccctgtgc gccaaggctc  
50461 agccctgtgt gtgcggatgt ctggctctgg cctctccccc tcccatcgac gacccacaggaa  
50521 gagatggctt ctgaggaccc ttggcggatgtt gggccctggg aatagattt ccaggaggt  
50581 ttaaaggcgc tgagtgttc atccagtaa ggcctggggaa ggacgtggc tcaaggctgt  
50641 acaggtgtga cagggtggg gactggggaa taagatgtaa aaccctggctt ggaggctgt  
50701 agttccaca gccaggcgatc gacagggggg gtccagatgtt accccactgtt gcccaccca



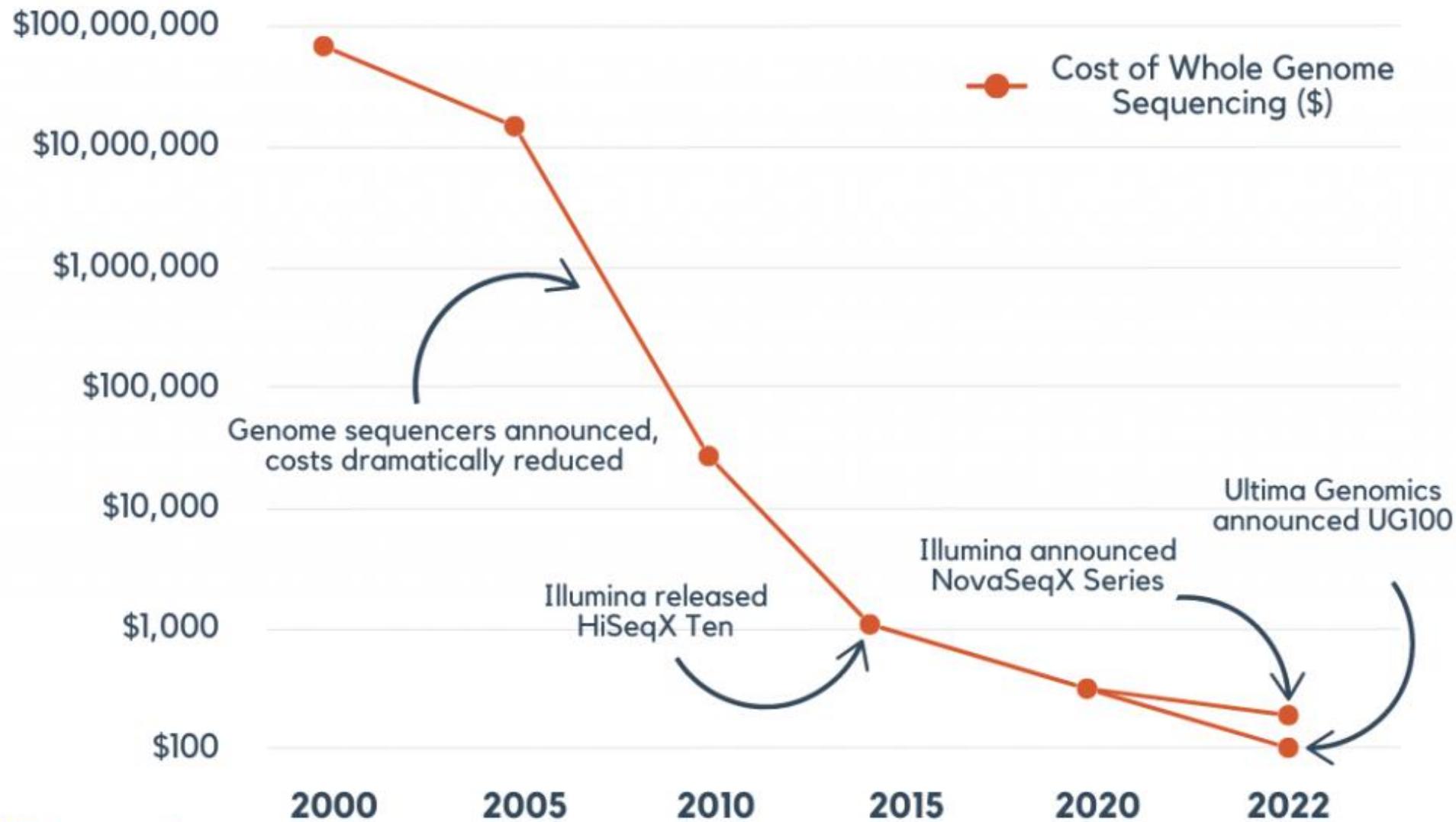
# How big is 40 exabytes?

Genomics projects will generate 40 exabytes of data in the next decade.

*Each shark = 100,000,000 GB of data*



# Decreasing Genome Sequencing Costs

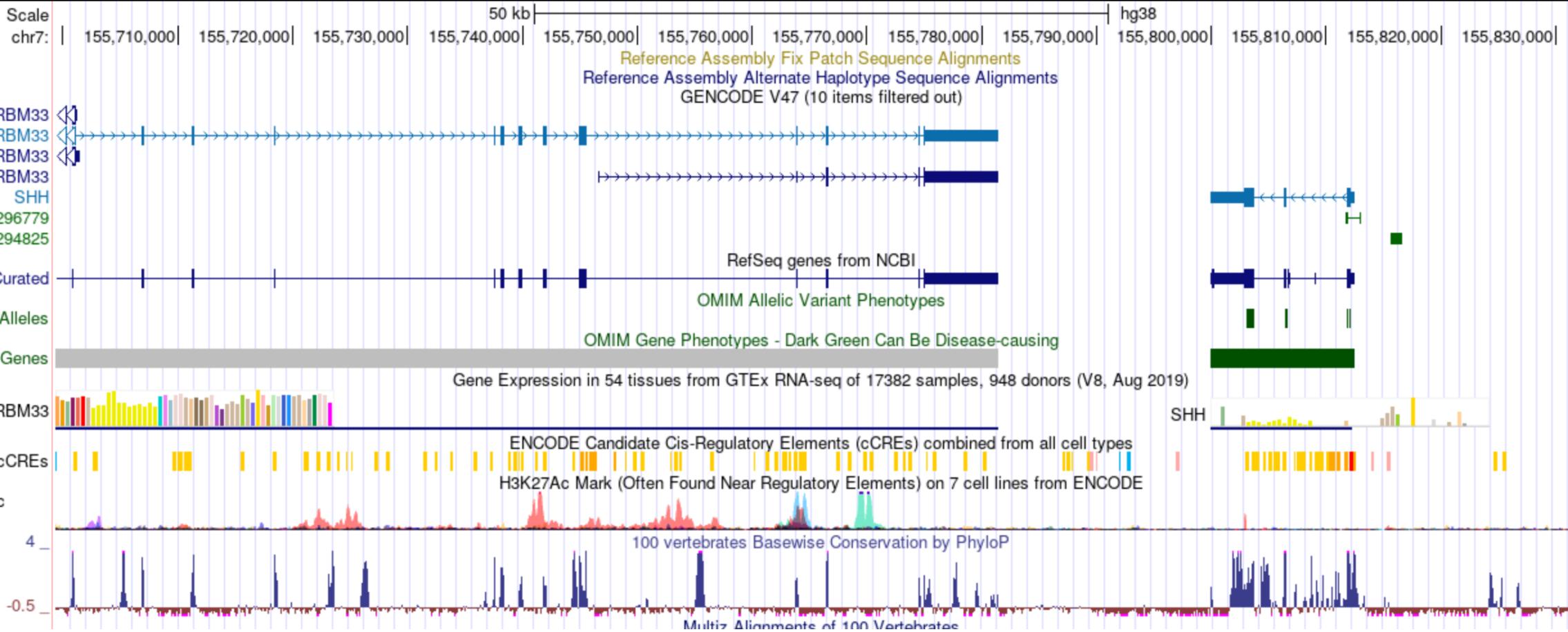
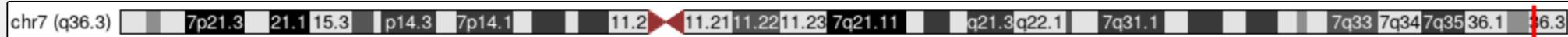


# Gene Annotation

[Multi-region](#)

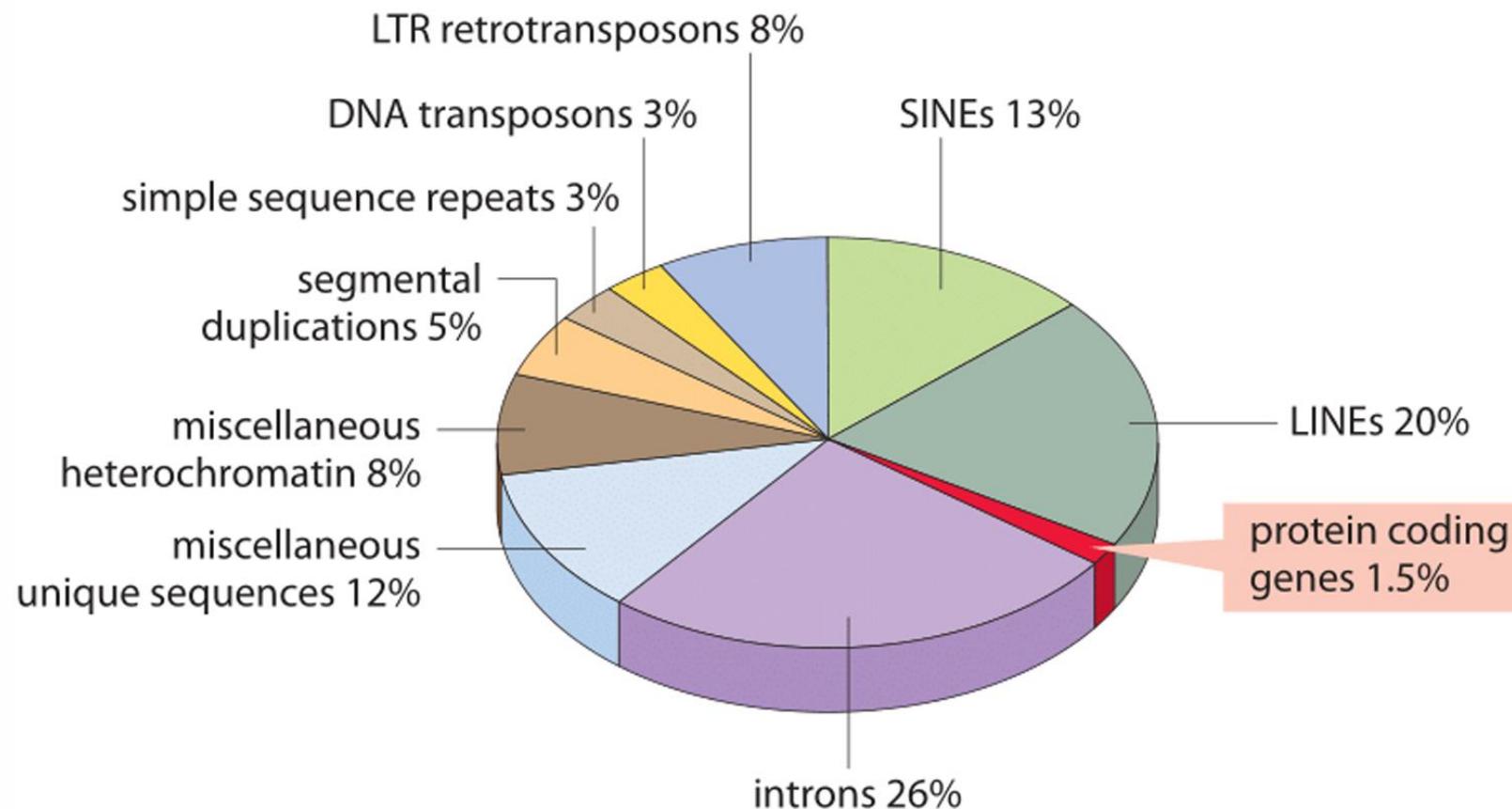
chr7:155,699,333-155,832,762 133,430 bp.

gene, chromosome range, search terms, help pages, see exam

[Search](#)[Examples](#)

# Gene Annotation

main components of the human genome



# Understanding Gene Annotation through GENCODE

(<https://www.gencodegenes.org/>)



Human

Mouse

How to access data

FAQ

Documentation

About us

## HUMAN

GENCODE 47 (October 2024)



## MOUSE

GENCODE M36 (October 2024)



The goal of the GENCODE project is to identify and classify all gene features in the human and mouse genomes with high accuracy based on biological evidence, and to release these annotations for the benefit of biomedical research and genome interpretation.

The GENCODE human and mouse lncRNA annotations are significantly expanding as we integrate models from our [Capture Long-read Sequencing project](#).

# Understanding Gene Annotation through GENCODE

[Human](#)[Mouse](#)[How to access data](#)[FAQ](#)[Documentation](#)[About us](#)**Human**

## Release 47 (GRCh38.p14)

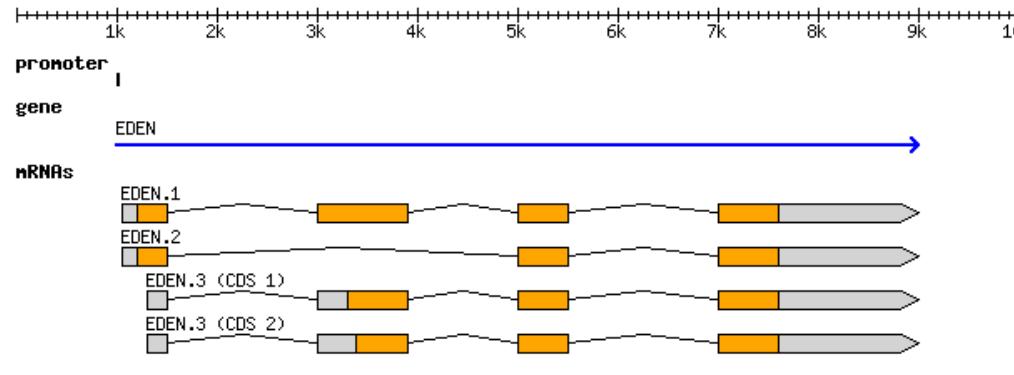
- [Statistics of this release](#)
- [More information about this assembly](#) (including patches, scaffolds and haplotypes)
- [Go to GRCh37 version of this release](#)

[More about GENCODE Human](#)[Current human data](#)[Release history](#)[Statistics](#)[Data format](#)[FTP site](#)

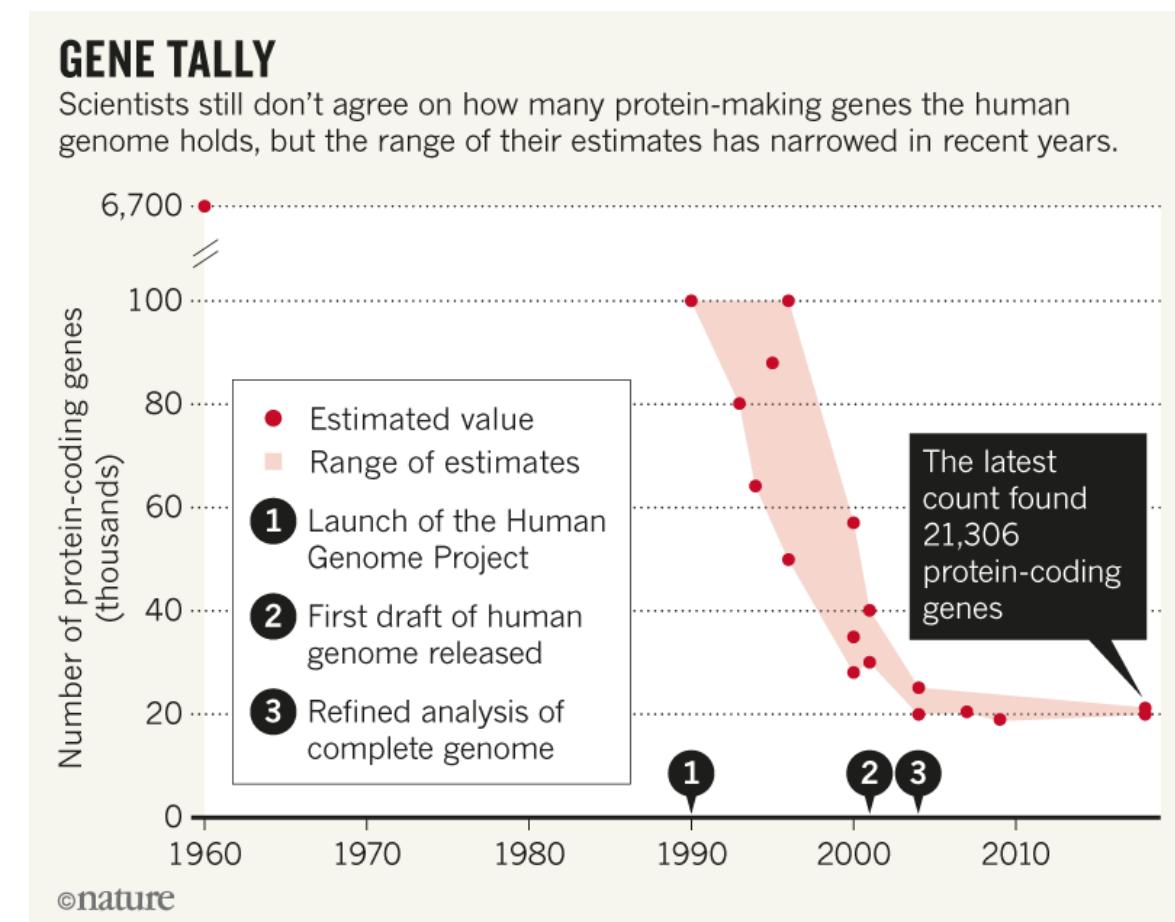
## GTF / GFF3 files

Content	Regions	Description	Download
Comprehensive gene annotation	CHR	<ul style="list-style-type: none"><li>It contains the comprehensive gene annotation on the reference chromosomes only</li></ul>	<a href="#">GTF</a> <a href="#">GFF3</a>
Comprehensive gene annotation	ALL	<ul style="list-style-type: none"><li>It contains the comprehensive gene annotation on the reference chromosomes, scaffolds, assembly patches and alternate loci (haplotypes)</li></ul>	<a href="#">GTF</a> <a href="#">GFF3</a>
Comprehensive gene annotation	PRI	<ul style="list-style-type: none"><li>It contains the comprehensive gene annotation on the primary assembly (chromosomes and scaffolds) sequence regions</li></ul>	<a href="#">GTF</a> <a href="#">GFF3</a>
Basic gene annotation	CHR	<ul style="list-style-type: none"><li>It contains the basic gene annotation on the reference chromosomes only</li><li>This is a <b>subset</b> of the corresponding comprehensive annotation, including</li></ul>	<a href="#">GTF</a> <a href="#">GFF3</a>

# GTF (Gene Transfer Format) and The number of Human Genes



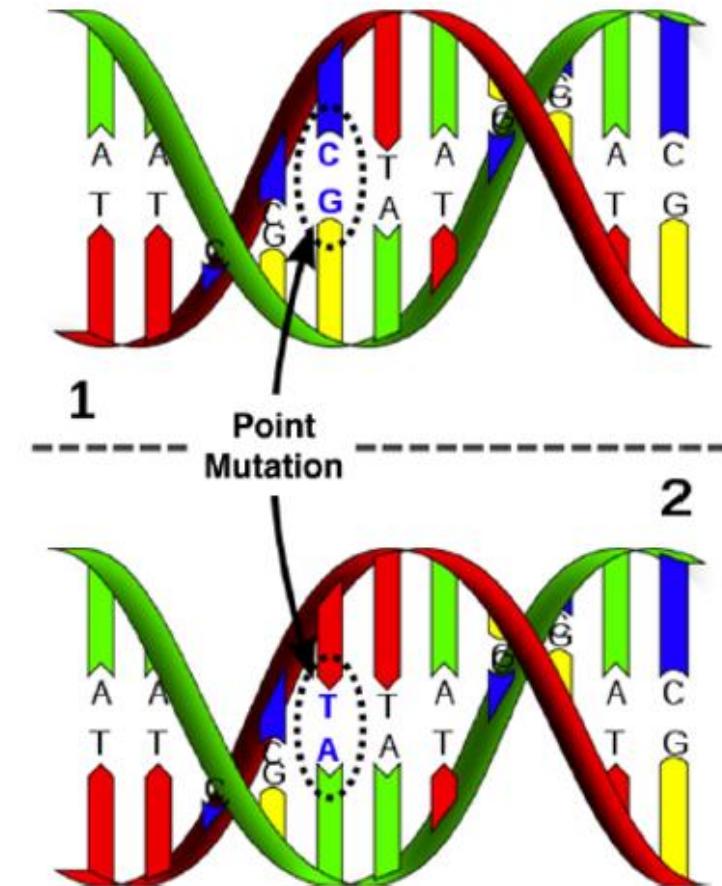
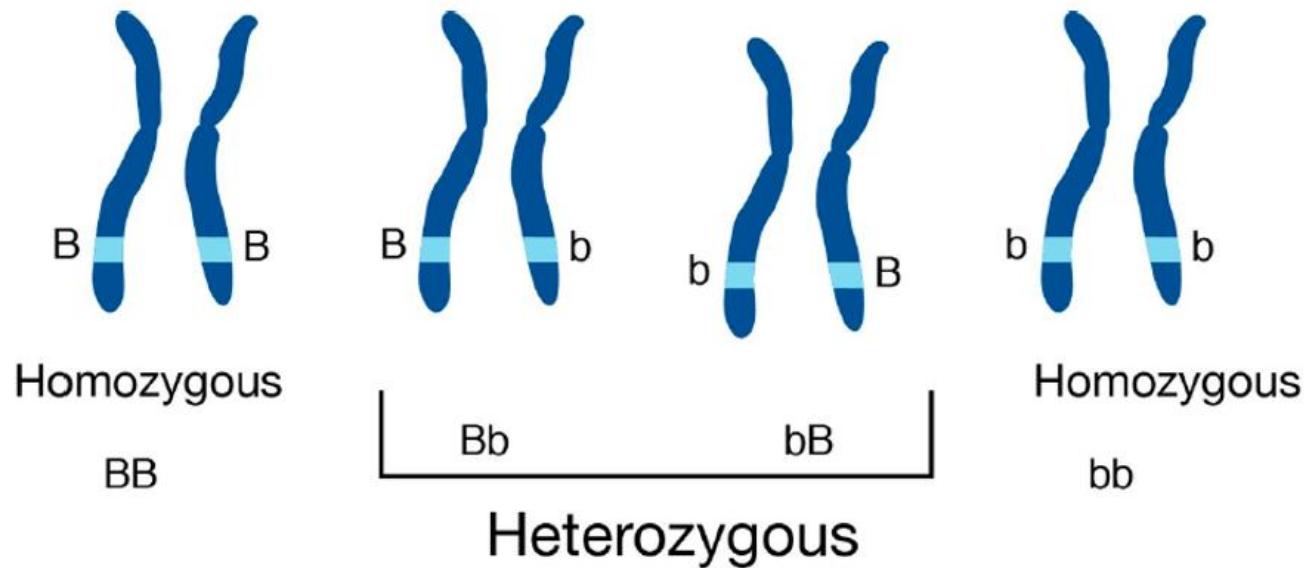
```
0 ##gff-version 3.2.1
1 ##sequence-region ctg123 1 1497228
2 ctg123 . gene      1000  9000 . + . ID=gene00001;Name=EDEN
3 ctg123 . TF_binding_site 1000 1012 . + . ID=tfbs00001;Parent=gene00001
4 ctg123 . mRNA      1050  9000 . + . ID=mRNA00001;Parent=gene00001;Name=EDEN.1
5 ctg123 . mRNA      1050  9000 . + . ID=mRNA00002;Parent=gene00001;Name=EDEN.2
6 ctg123 . mRNA      1300  9000 . + . ID=mRNA00003;Parent=gene00001;Name=EDEN.3
7 ctg123 . exon      1300  1500 . + . ID=exon00001;Parent=mRNA00003
8 ctg123 . exon      1050  1500 . + . ID=exon00002;Parent=mRNA00001,mRNA00002
9 ctg123 . exon      3000  3902 . + . ID=exon00003;Parent=mRNA00001,mRNA00003
10 ctg123 . exon     5000  5500 . + . ID=exon00004;Parent=mRNA00001,mRNA00002,mRNA00003
11 ctg123 . exon     7000  9000 . + . ID=exon00005;Parent=mRNA00001,mRNA00002,mRNA00003
12 ctg123 . CDS      1201  1500 . + 0 ID=cds00001;Parent=mRNA00001;Name=edenprotein.1
13 ctg123 . CDS      3000  3902 . + 0 ID=cds00001;Parent=mRNA00001;Name=edenprotein.1
14 ctg123 . CDS      5000  5500 . + 0 ID=cds00001;Parent=mRNA00001;Name=edenprotein.1
15 ctg123 . CDS      7000  7600 . + 0 ID=cds00001;Parent=mRNA00001;Name=edenprotein.1
16 ctg123 . CDS      1201  1500 . + 0 ID=cds00002;Parent=mRNA00002;Name=edenprotein.2
17 ctg123 . CDS      5000  5500 . + 0 ID=cds00002;Parent=mRNA00002;Name=edenprotein.2
18 ctg123 . CDS      7000  7600 . + 0 ID=cds00002;Parent=mRNA00002;Name=edenprotein.2
19 ctg123 . CDS      3301  3902 . + 0 ID=cds00003;Parent=mRNA00003;Name=edenprotein.3
20 ctg123 . CDS      5000  5500 . + 1 ID=cds00003;Parent=mRNA00003;Name=edenprotein.3
21 ctg123 . CDS      7000  7600 . + 1 ID=cds00003;Parent=mRNA00003;Name=edenprotein.3
22 ctg123 . CDS      3391  3902 . + 0 ID=cds00004;Parent=mRNA00003;Name=edenprotein.4
23 ctg123 . CDS      5000  5500 . + 1 ID=cds00004;Parent=mRNA00003;Name=edenprotein.4
24 ctg123 . CDS      7000  7600 . + 1 ID=cds00004;Parent=mRNA00003;Name=edenprotein.4
```



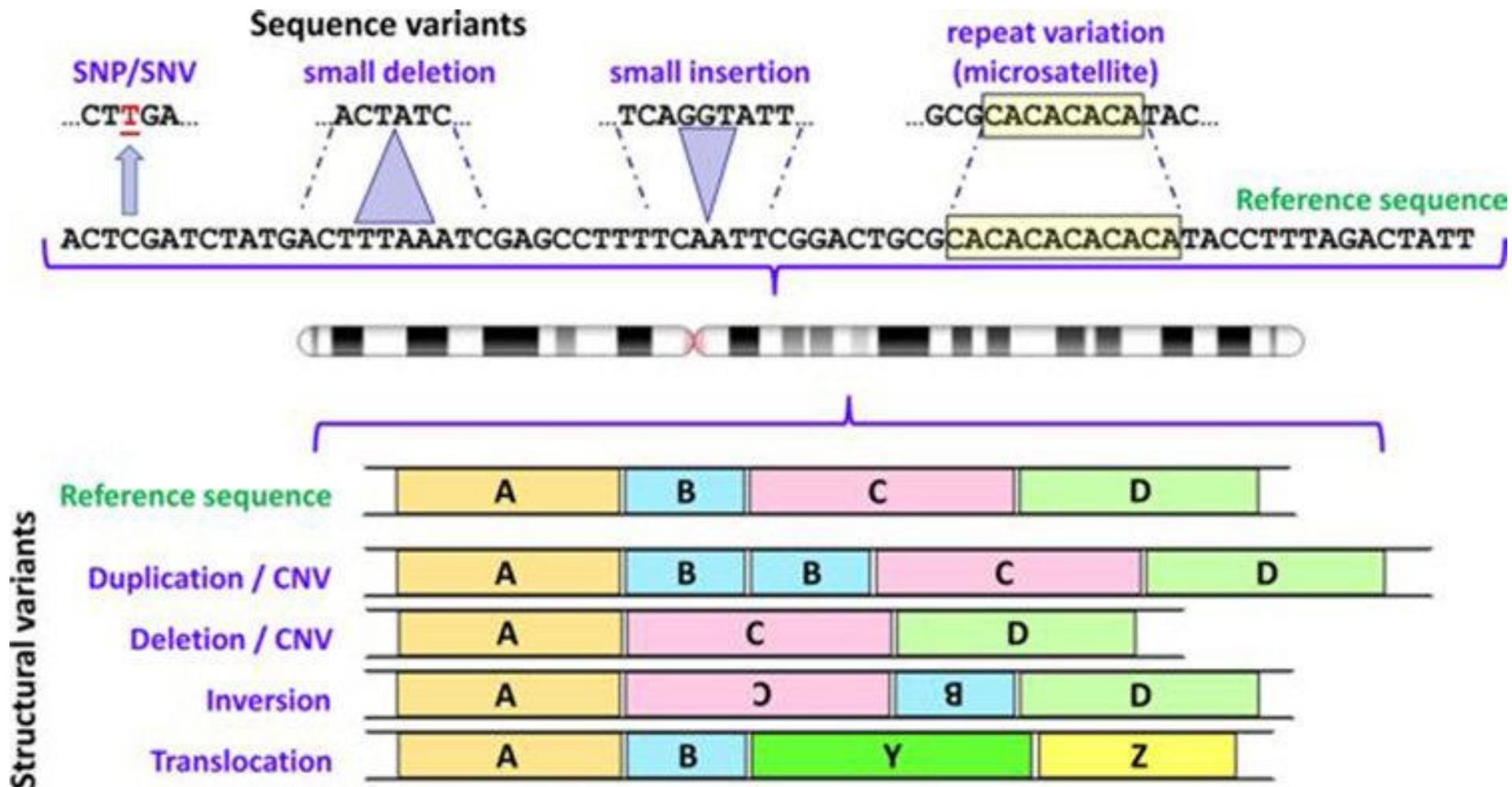
<https://www.nature.com/articles/d41586-018-05462-w>

# Các loại biến thể trên hệ gen

- Hệ gen giữa hai người giống nhau > 99%
- Mỗi người có khoảng 5 triệu biến thể, trong đó có 3 đến 4 triệu biến thể một nucleotide
- Hệ gen người là hệ lưỡng bội



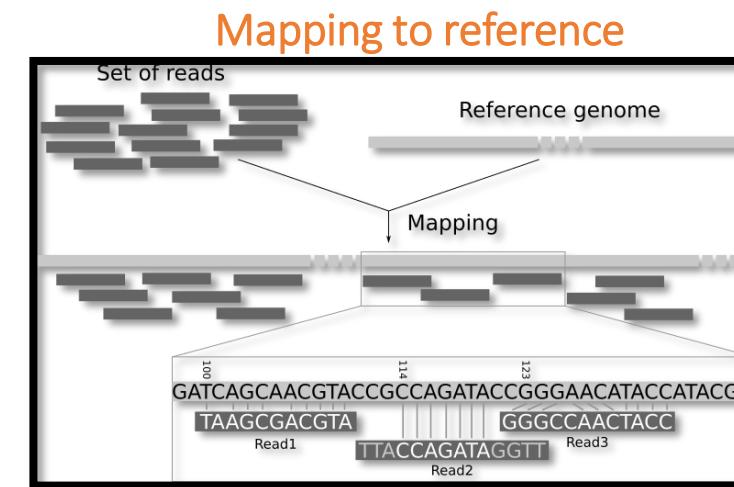
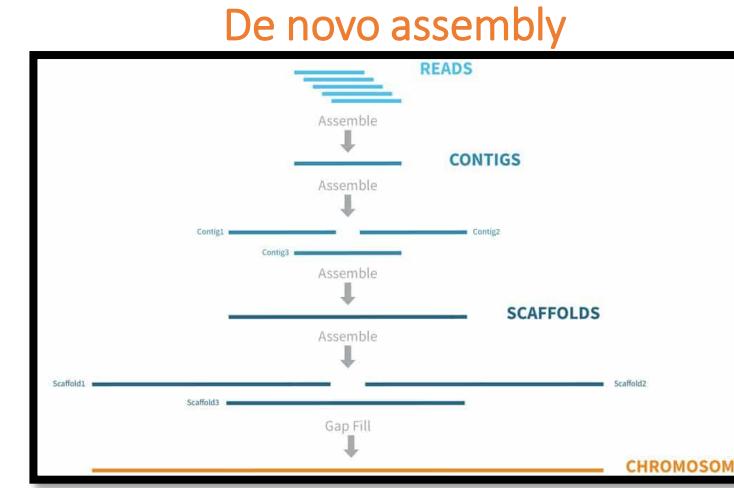
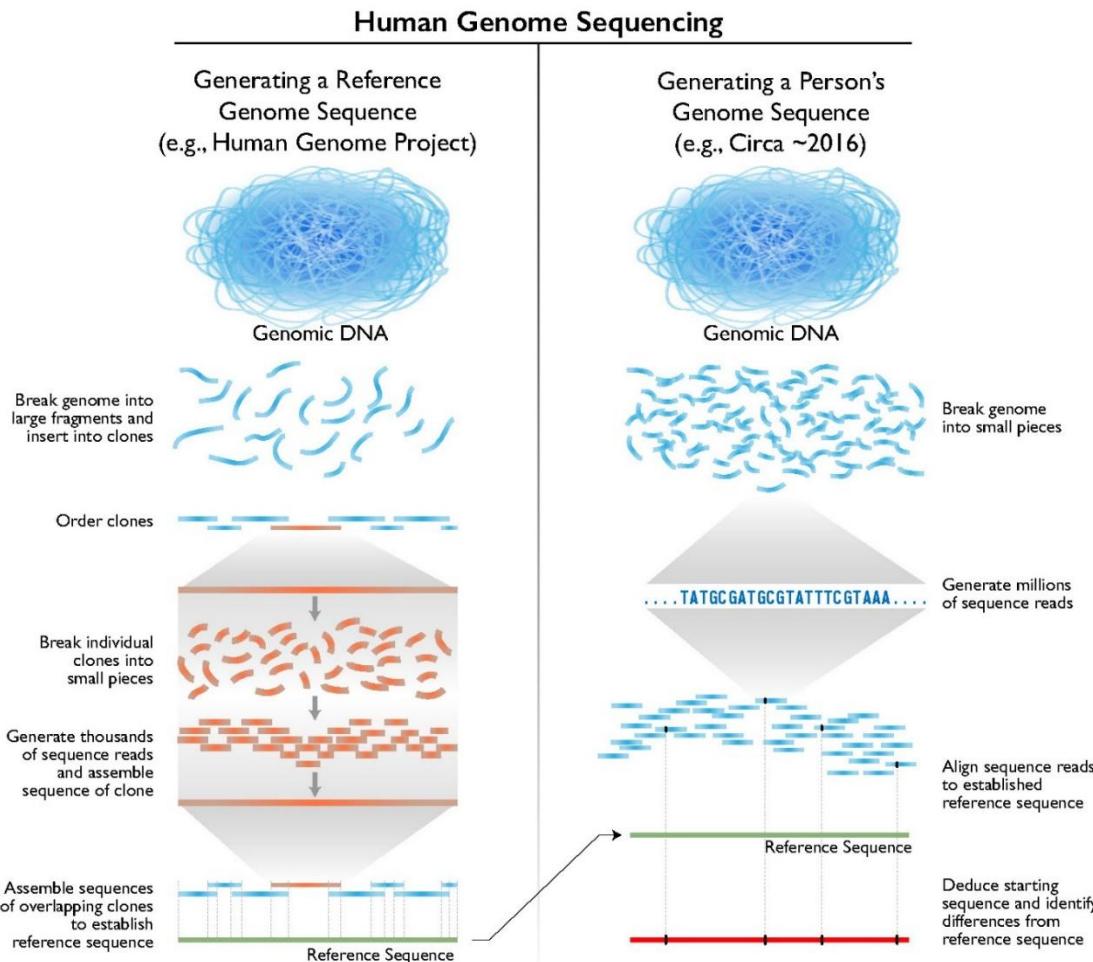
# Các loại biến thể trên hệ gen



Giới thiệu về

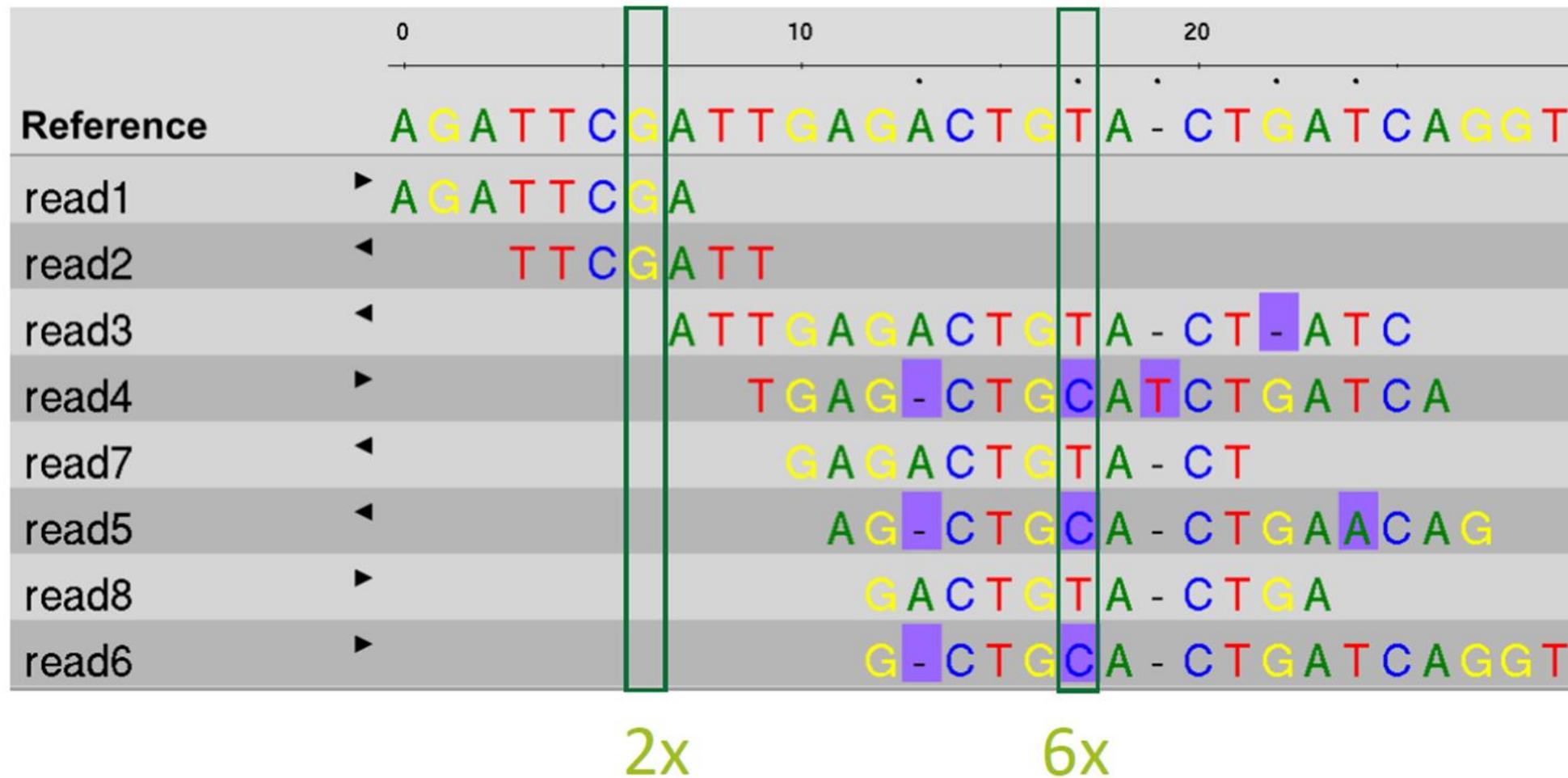
# GIẢI TRÌNH TỰ GEN THẾ HỆ MỚI

# Giải trình tự gen thế hệ mới (NGS)

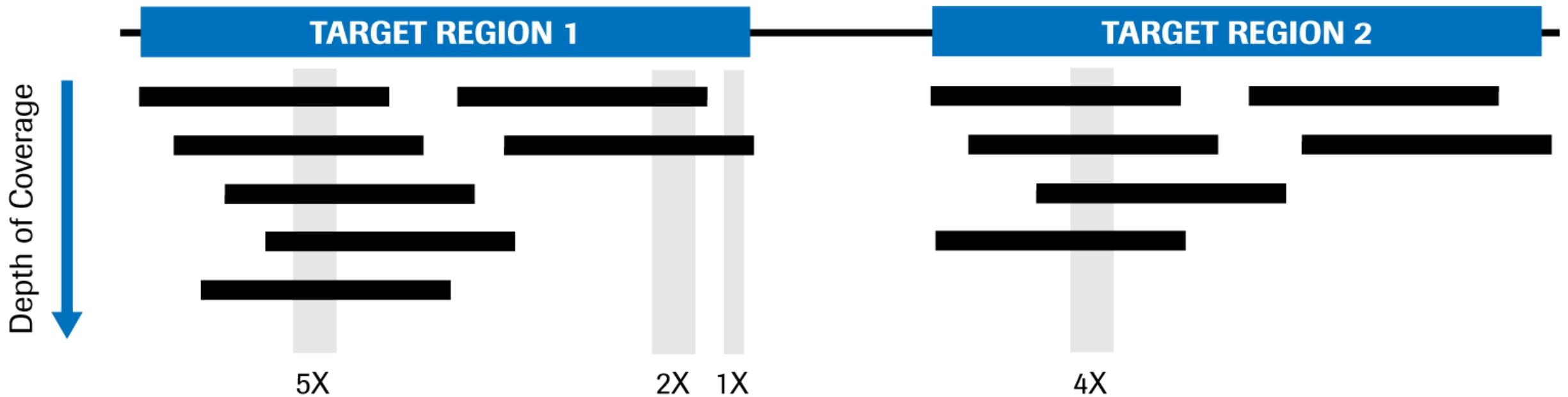


<https://www.genome.gov/about-genomics/fact-sheets/Sequencing-Human-Genome-cost>

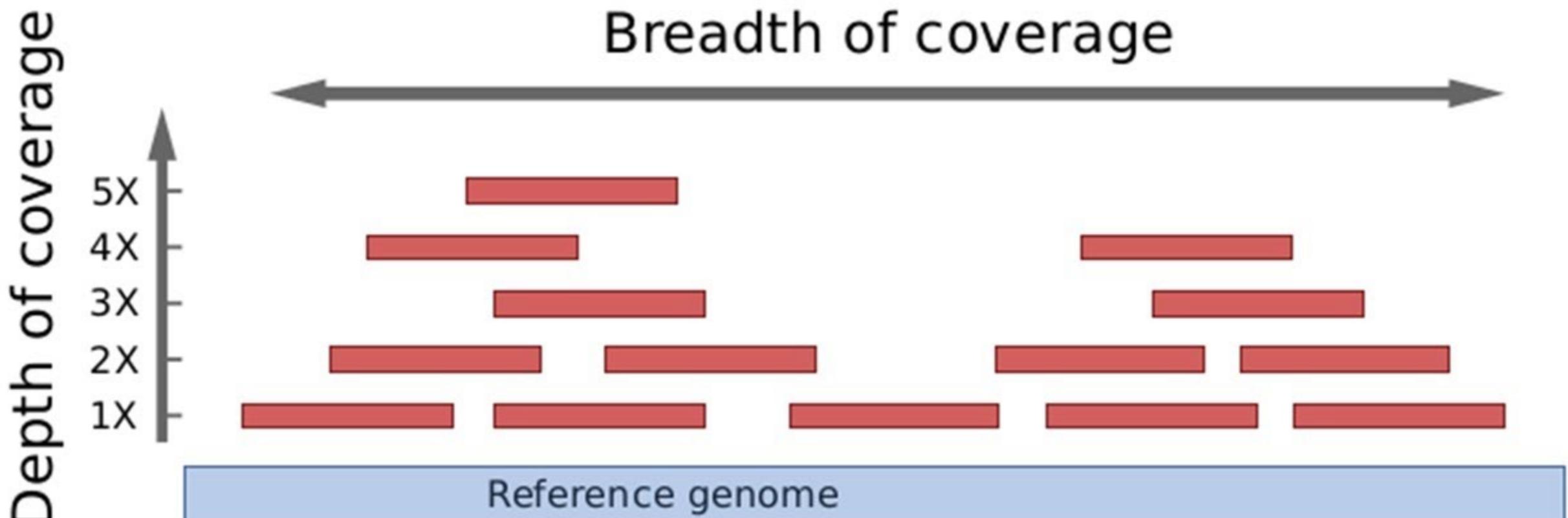
# Kết quả của mapping



# Depth of Coverage



# Depth vs Breadth of Coverage



# Giải trình tự gen thế hệ mới (NGS): Có hệ gen tham chiếu

De novo assembly

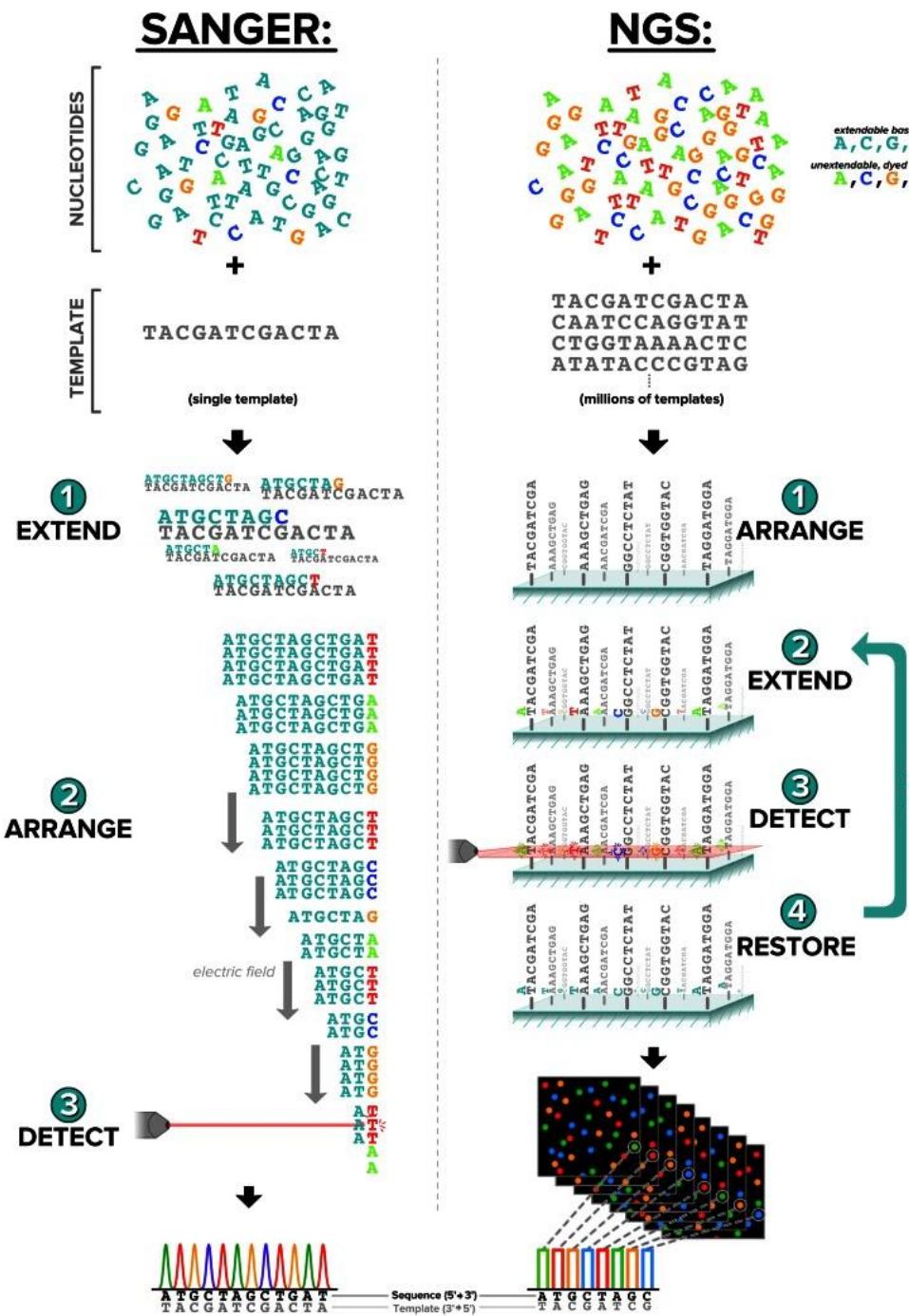


Mapping to reference



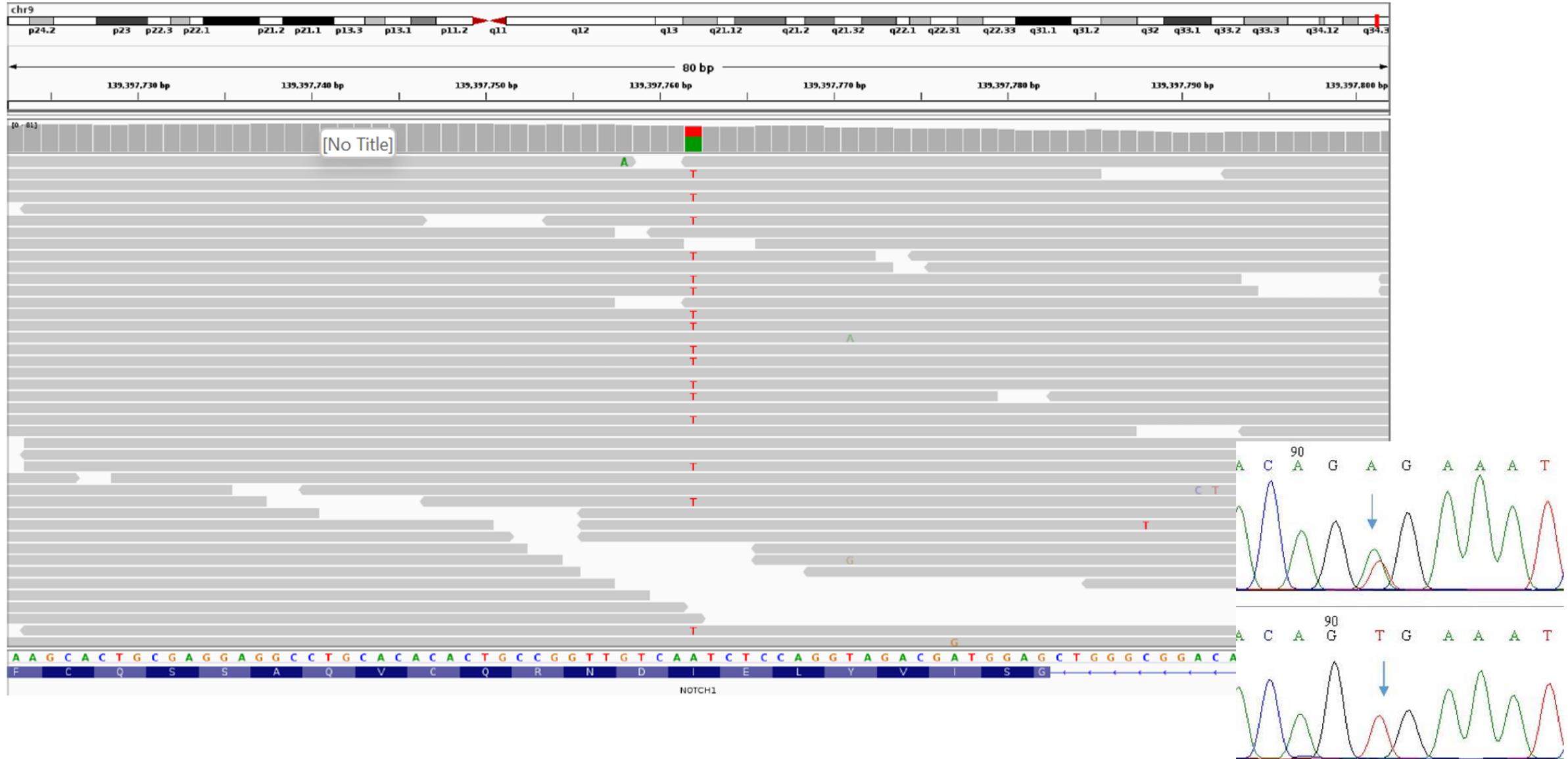
# Giải trình tự gen thế hệ mới (NGS): giải trình tự song song

Read1 : CTCGAATACG



Read1 : CTCGAATACG  
 Read2 : CTCGAATACG  
 Read3 : CTCGAATACG  
 Read4 : CTCGAATACG  
 Read5 : CGCGAATACG  
 Read6 : CGCGAATACG  
 Read7 : CGCGACTACG  
 Read8 : CGCGAATACG

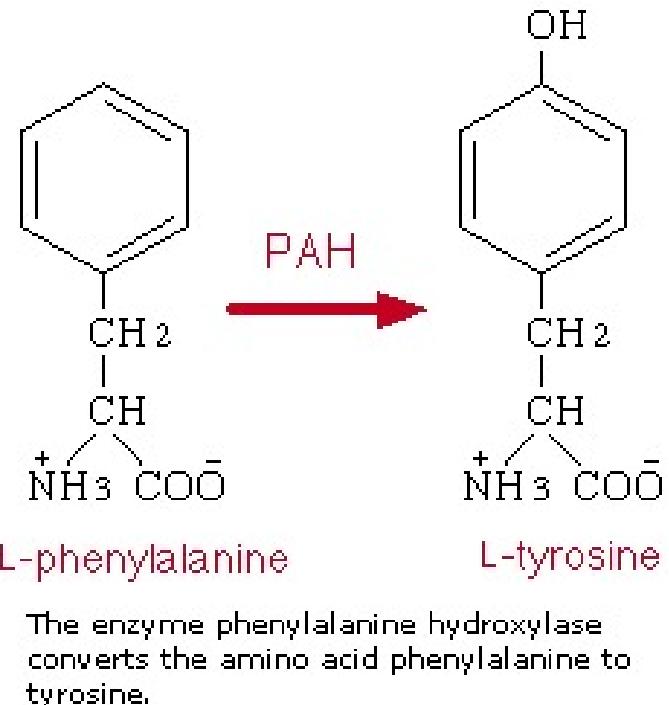
# NGS vs Sanger



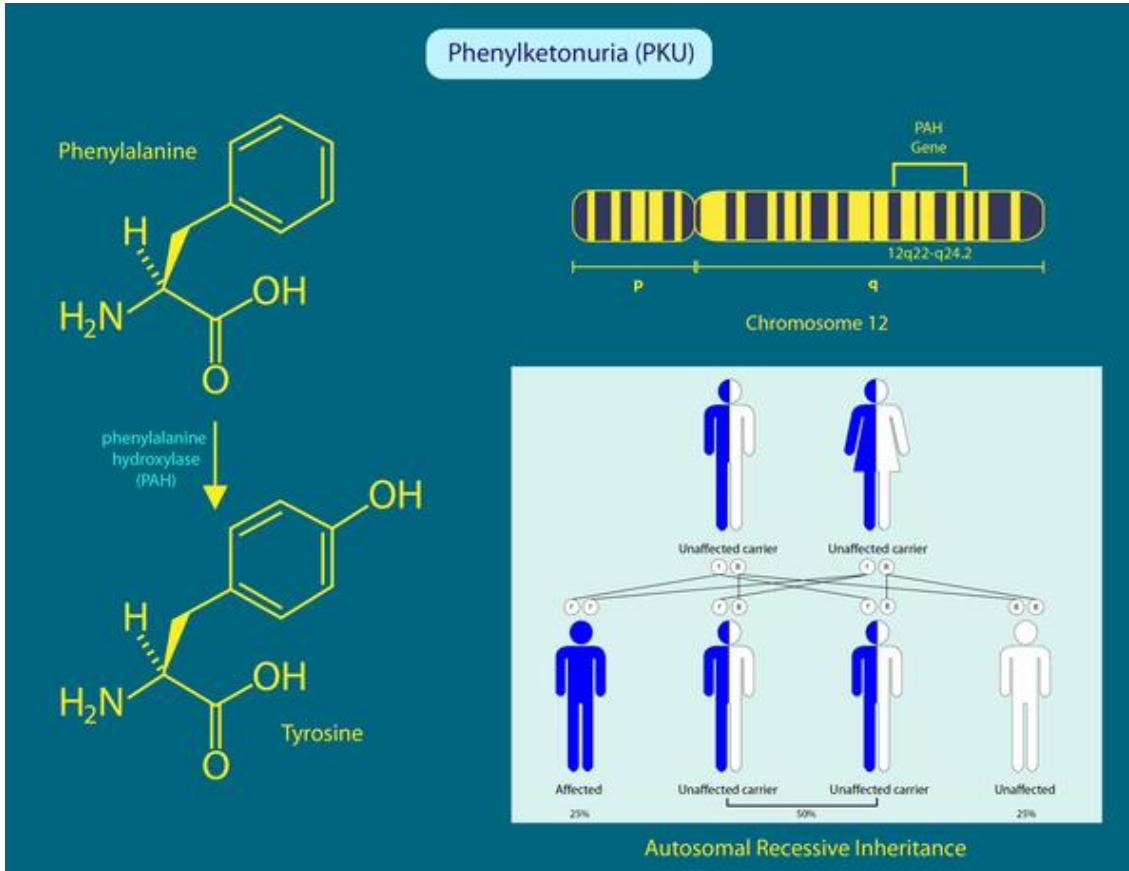
Ví dụ mối quan hệ của  
**BIẾN THỂ GEN** và **BỆNH DI TRUYỀN**

# Phenylketonuria (PKU): Rối loạn chuyển hóa di truyền

- Nguyên nhân do thiếu hụt enzyme phenylalanine hydroxylase.
  - Sự mất enzyme này dẫn đến suy giảm trí tuệ, tổn thương cơ quan, tư thế bất thường.
  - Tần suất xảy ra PKU khác nhau giữa các nhóm dân tộc và các vùng địa lý trên toàn thế giới. Ở Hoa Kỳ, PKU xảy ra ở 1 trong 25.000 trẻ sơ sinh.
  - Hầu hết các trường hợp PKU được phát hiện ngay sau sinh bằng sàng lọc sơ sinh và điều trị được bắt đầu ngay lập tức.



# Phenylketonuria (PKU): Rối loạn chuyển hóa di truyền



Tên khác của PKU

- Folling disease
- Folling's disease
- PAH deficiency
- Phenylalanine hydroxylase deficiency
- Phenylalanine hydroxylase deficiency disease

<https://medlineplus.gov/genetics/condition/phenylketonuria/>

# Trình tự gen PAH ở người - Homo sapiens (5053)

## Gene (Nucleotide)

NT seq	1359 nt <a href="#">NT seq</a> atgtccactgcggctggaaaacccaggctggcaggaaactctctgactttggacag gaaacaagcttatattgaagacaactgaatcaaaatggccatatcactgatcttca ctcaaagaagaagtgggtcattggccaaagtattgcgttatttggaggagaatgtat aacctgaccacattgaatcttagacccctcgtttaaagaaaagatgagatgaattttc acccatttgataaaacgttagcctgcgtctgacaaaatcatcaagatcttgggcat gacattggtgcactgtccatgagcttcacgagataagaagaaaagacacagtggccctgg ttcccaagaaccattcaagagctggacagatggccaatcagattctcagctatggagcg gaactggatgctgaccaccctggtttaaagatccgtgtaccgtgcaagacggaaagcag tttgcgtacattgcctacaactaccgcattggcagccccatccctcgagttggaaatacatg gaggaagaaaaagaaaatggggcacagtgttcaagactctgaagtccctgtataaaaacc catgttgcgtatgagttacaatcacattttccacttctgaaaagtactgtggcccttccat gaagataacattccccagctggaaagacgttctcagttctgcagacttgcactggtttc cgccctccgacacctgtggctggcctgtccctcgggatcttggggcccttc cgagtcttccactgcacacagtacatcagacatggatccaagccatgtataccccgaa cctgacatctgcatgagctgtggacatgtgcccttgcggatctggcggccatgttgc cagtttccaggaaattggccctgcctctgggtgcacctgtatgaaatacatgtaaaag ctcgccacaatttactgtggtagttggctctgcaaaacaaggagactccata aaggcatatggtgcggccctgtcatccttggtaattacagtactgtttatcagag aagccaaagcttctcccccggagctggagaagacagccatccaaaattacactgtcag gagttccagcccttcttacgtggcagagatgtttaatgtatgccaaggagaaaagtaagg aactttgctgccacaataccctcgcccttctcagttcgctacgacccatacacccaaagg attggaggcttggacaatacccagcagcttaagatggctgattccattaacagtgaa attggaaatcccttgcagtggccctccagaaaataaagtaa
--------	--

## Protein (Amino Acid)

AA seq	452 aa <a href="#">AA seq</a> <a href="#">DB search</a> MSTAVLENPGLGRKLSDFGQETSYIEDNCNQNGAISLIFSLKEEVGALAKVLRLFEENDV NLTHIESRPSRLKKDEYEFFTHLDKRSLPALTNIKILRHDIGATVHELSRDKKKDTVPW FPTIQELDRFANQILSYGAELDADHPGFKDPMVYRARRKQFADIAYNYRHGQPIPRVEYM EEEKKTWGTVFKTLKSLYKTHACYEYNHIFPLLEKYCFGHEDNIPQLEDVSQFLQTCTGF RLRPVAGLLSSRDFLGLAFRVFHCTQYIRHGSKPMTPEPDICHELLGHVPLFSDRSFA QFSQEIGLASLGAPDEYIEKLATIYWFTVEFLCKQGDSIKAYGAGLLSSFGELQYCLSE KPKLLPLELEKTAIQNYTVTEFQPLYYVAESFNDAKEKVRNFAATIPRPF SVRYDPYTQR IEVLDNTQQLKILADSINSEIGILCSALQKIK
--------	--

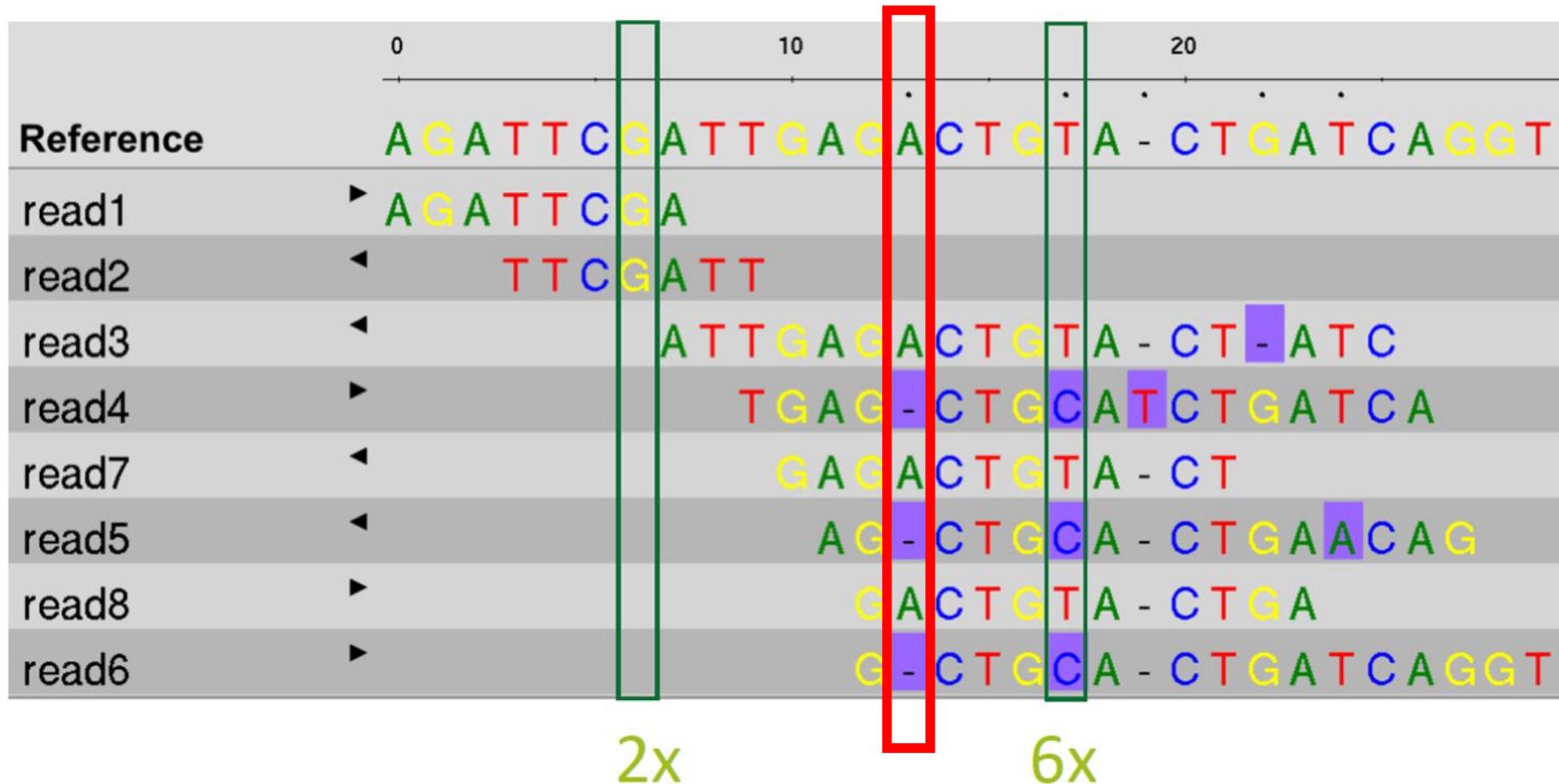
<https://www.genome.jp/entry/T01001:5053>

# Gen PAH



<https://www.ncbi.nlm.nih.gov/gene/5053>

# Variants in PAH



# Cơ sở dữ liệu ClinVar cho gen PAH

## Classification type

- Germline (208)
- Somatic (0)

## Germline classification

- Conflicting classifications (2)
- Benign (10)
- Likely benign (21)
- Uncertain significance (41)
- Likely pathogenic (40)
- Pathogenic (39)

## Types of conflicts

- P/LP vs LB/B (0)
- P/LP vs VUS (0)
- VUS vs LB/B (2)

## Molecular consequence

- Frameshift (20)
- Missense (96)
- Nonsense (7)
- Splice site (9)
- ncRNA (0)
- Near gene (0)
- UTR (27)

## Variation type

- Deletion (39)

## Links from Gene

[Display options](#) ▾ [Sort by Relevance](#) ▾ [Download](#) ▾

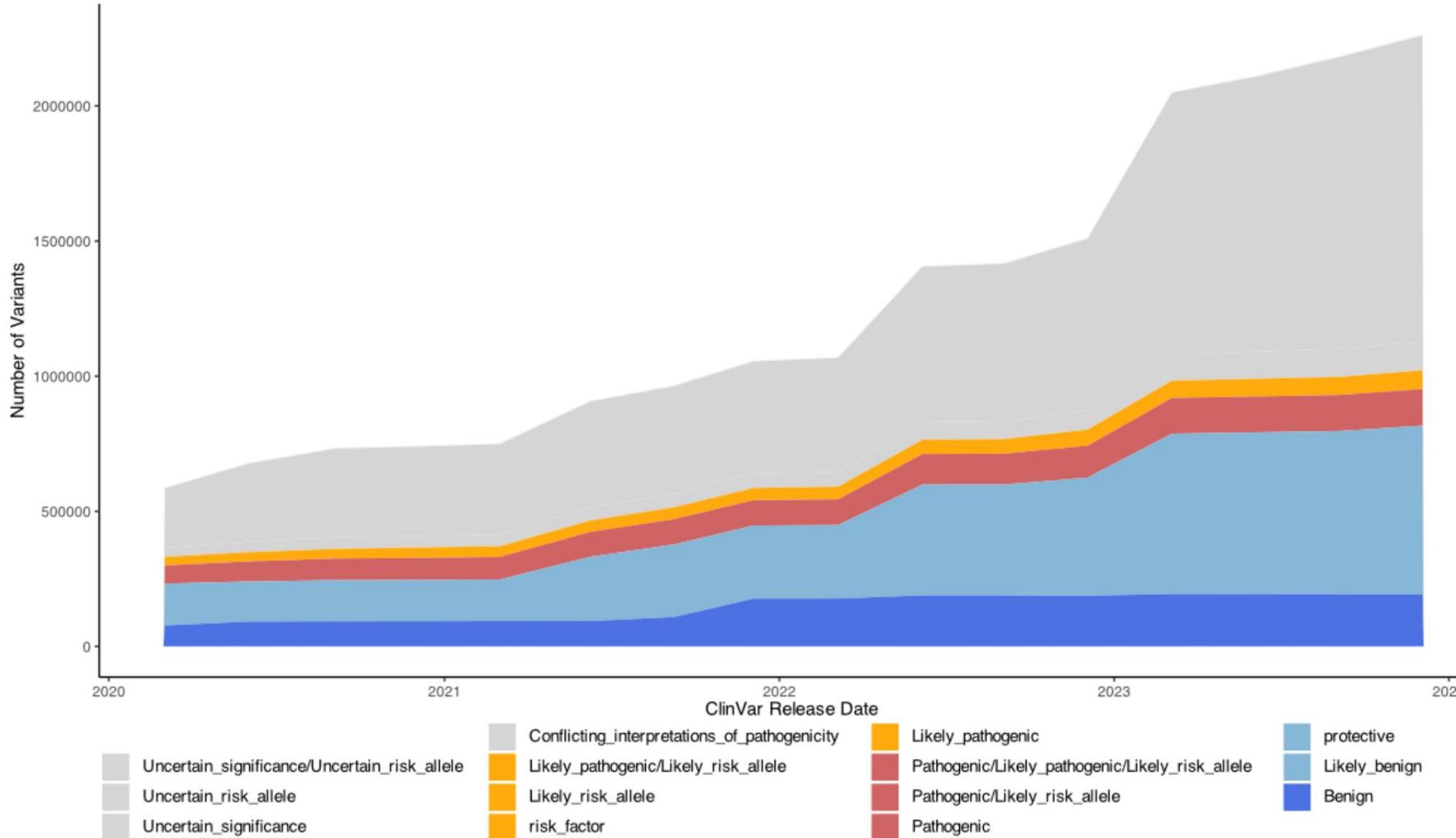
Items: 1 to 100 of 209

<< First < Prev Page  of 3 Next > Last >>

Variation	Gene (Protein Change)	Type (Consequence)	Condition	Classification, Review status
<input type="checkbox"/> <a href="#">NM_004316.4(ASCL1):c.51G&gt;T (p.Gln17His)</a>	ASCL1, PAH (Q17H)	Single nucleotide variant (missense variant +1 more)	not specified	 Uncertain significance ★
<input type="checkbox"/> <a href="#">NC_000012.11:g.(?_103232953)_(1_03240749_?)del</a>	PAH	Deletion	Phenylketonuria	 Pathogenic ★
<input type="checkbox"/> <a href="#">NC_000012.11:g.(?_103288493)_(1_03310908_?)del</a>	PAH	Deletion	Phenylketonuria	 Pathogenic ★
<input type="checkbox"/> <a href="#">NC_000012.11:g.(?_103248894)_(1_03249131_?)del</a>	PAH	Deletion	Phenylketonuria	 Pathogenic ★
<input type="checkbox"/> <a href="#">NC_000012.12:g.(?_102894715)_(102894938_?)del</a>	PAH	Deletion	Phenylketonuria	 Pathogenic ★
<input type="checkbox"/> <a href="#">NC_000012.11:g.(?_103306549)_(1_03306696_?)del</a>	PAH	Deletion	Phenylketonuria	 Pathogenic ★
<input type="checkbox"/> <a href="#">NM_000277.3(PAH):c.1179_1180del (p.Asn393fs)</a>	PAH (N393fs)	Deletion (frameshift variant)	Phenylketonuria	 Likely pathogenic ★

[https://www.ncbi.nlm.nih.gov/clinvar?LinkName=gene\\_clinvar&from\\_uid=5053](https://www.ncbi.nlm.nih.gov/clinvar?LinkName=gene_clinvar&from_uid=5053)

# Overview of ClinVar release trends over time



The classes have been divided into five colors to represent the degree of pathogenicity, from blue (benign) to red (pathogenic) with gray representing variants that are of uncertain clinical significance or have no clinical significance.

# Biến thể gây bệnh - Pathogenic variant in PAH

## NM\_000277.3(PAH):c.971T>A (p.Ile324Asn)

**ClinVar** Genomic variation as it relates to human health

Search by gene symbols, location, HGVS expressions, c-dot, p-dot, conditions, : [Search ClinVar](#) [?](#)

[Advanced search](#)

[About](#) [Access](#) [Submit](#) [Stats](#) [FTP](#) [Help](#) [Like](#) [Dislike](#)

**NM\_000277.3(PAH):c.971T>A (p.Ile324Asn)** [Cite](#) [Follow](#) [Print](#) [Download](#)

**We've updated the ClinVar website to better support classifications of somatic variants!**

Read more about changes to the website in our [web release notes](#); more information about somatic variants in ClinVar is available on [GitHub](#).

**Germline** Top reviewed classifications are shown here. Submission summary: **1 submission 1 submitter 1 condition**

Reviewed by expert panel **Pathogenic** for Phenylketonuria    
 Dec 2023 by [ClinGen PAH Va...](#) [FDA RECOGNIZED DATABASE](#)

**Somatic** No data submitted for somatic clinical impact **Somatic** No data submitted for oncogenicity

**On this page**

- [Classification Summary](#)
- [Variant Details](#)
- [Genes](#)
- [Germline](#)
- [Conditions](#)
- [Submissions](#)
- [Citations](#)
- [Text mined Citations](#)

**Feedback**

<https://www.ncbi.nlm.nih.gov/clinvar/variation/2682170/>

# Các biến thể gây bệnh trên gen PAH

Gene: PAH

[View on UniProt](#)

[View on SwissModel](#)

Transcript: ENST00000553106.6

Select protein structure

SwissModel:5den 20-450 (number o...)

X

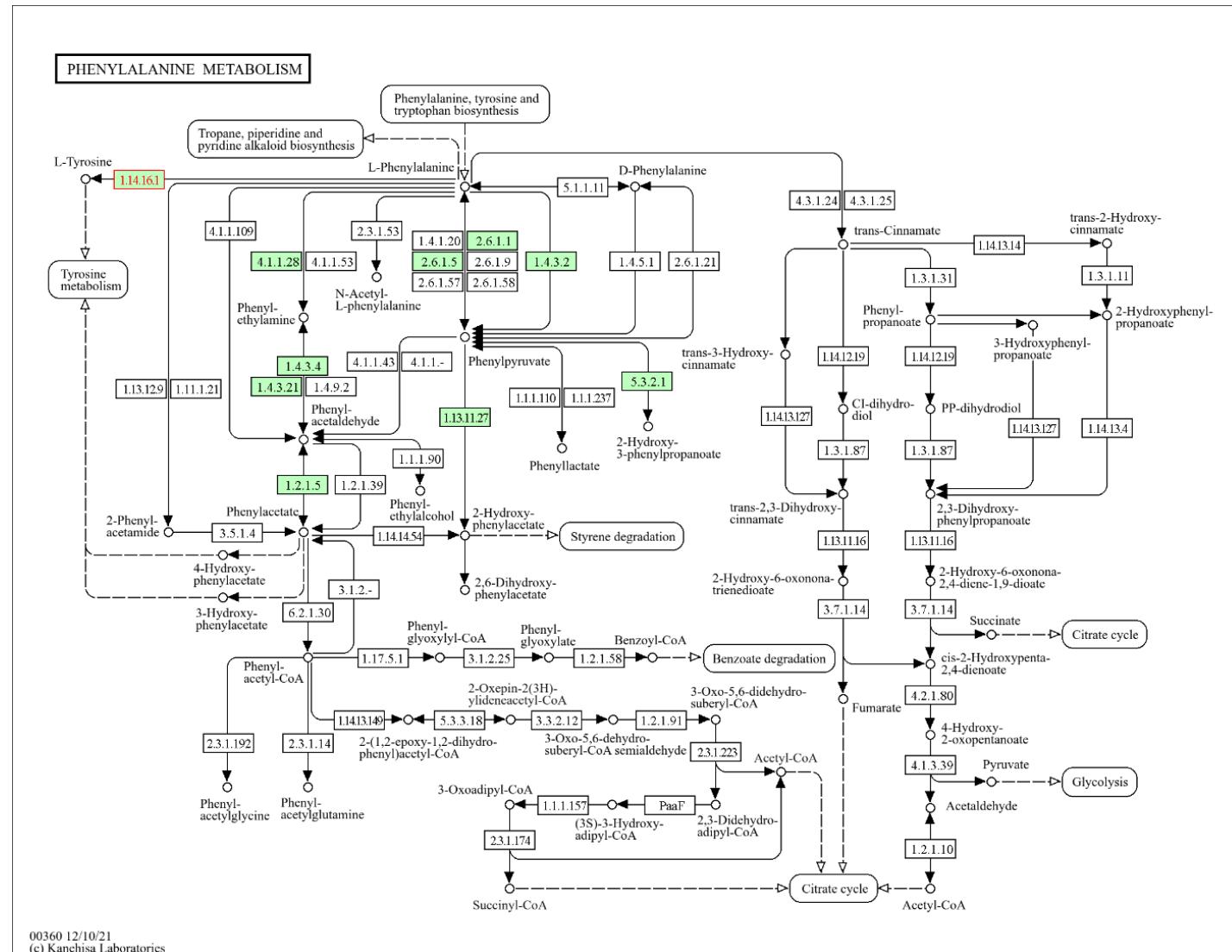


Show

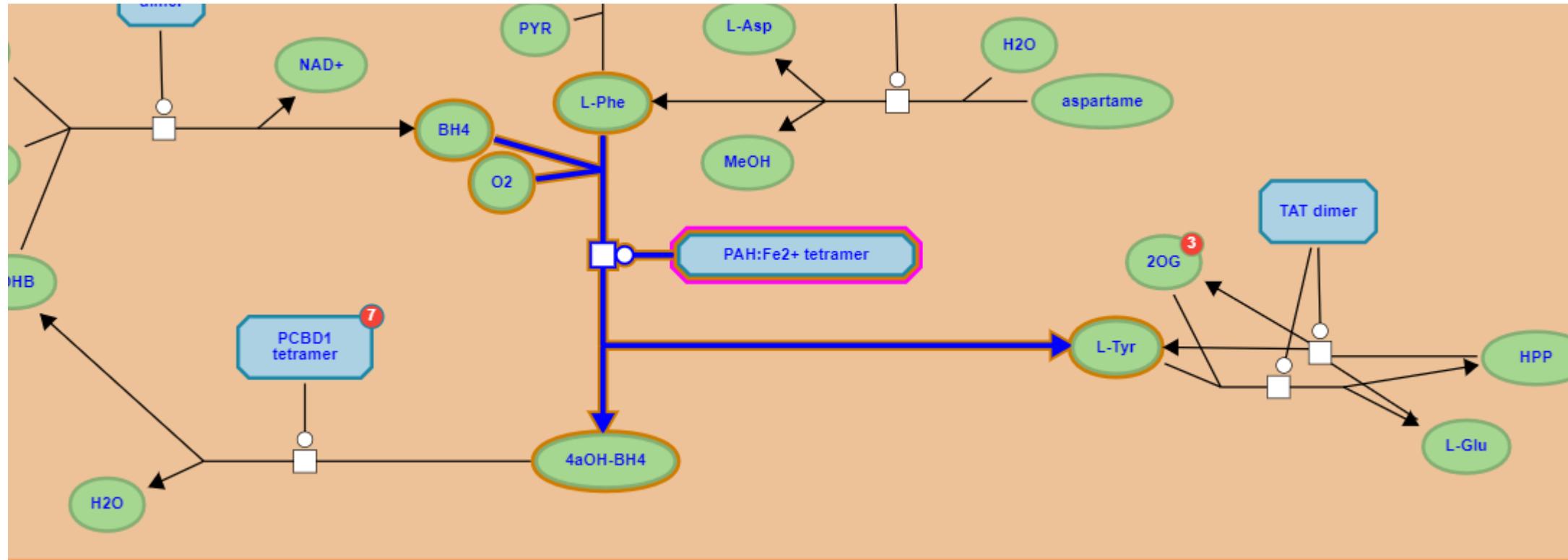
- All Residues
- Variants
- Pathogenic
- Likely Pathogenic
- Uncertain Significance
- Likely Benign
- Benign
- Current Variant

<https://varsome.com/variant/hg38/chr12%3A102844430%3AA%3AT?>

# PAH: chuyển hóa Phenylalanine thành Tyrosine

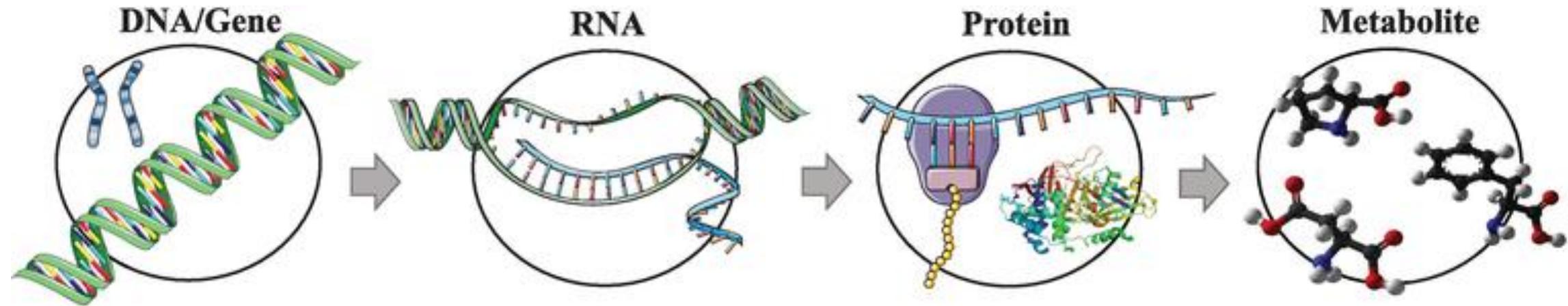


# PAH: chuyển hóa Phenylalanine thành Tyrosine



<https://reactome.org/PathwayBrowser/#/R-HSA-8963691&SEL=R-HSA-71118&PATH=R-HSA-1430728,R-HSA-71291&FLG=UniProt:P00439>

# Mối liên kết: Biến thể gen và bệnh di truyền



**Genomics**

PAH gene  
Ref ...ATCGAT...  
P1 ...AACGAT...  
  
NM\_000277.3(PAH):c.971T>A

**Transcriptomics**

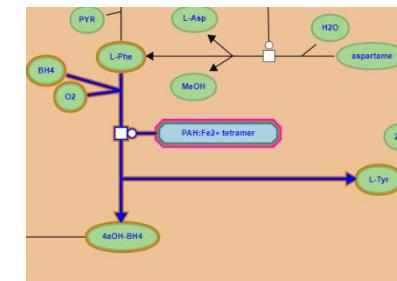
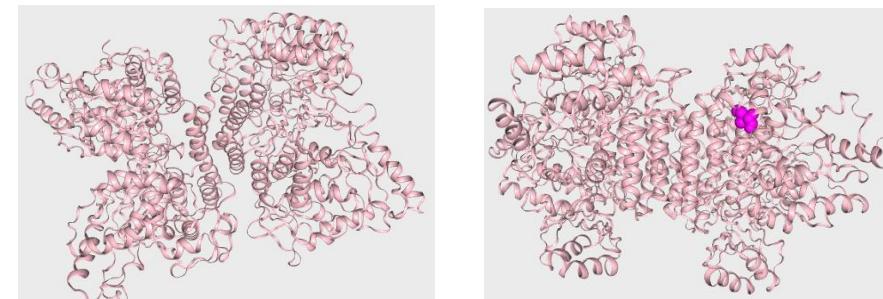
PAH mRNA  
Ref ...AUCGAU...  
P1 ...AACGAU...  
  
NM\_000277.3(PAH):c.971T>A

**Proteomics**

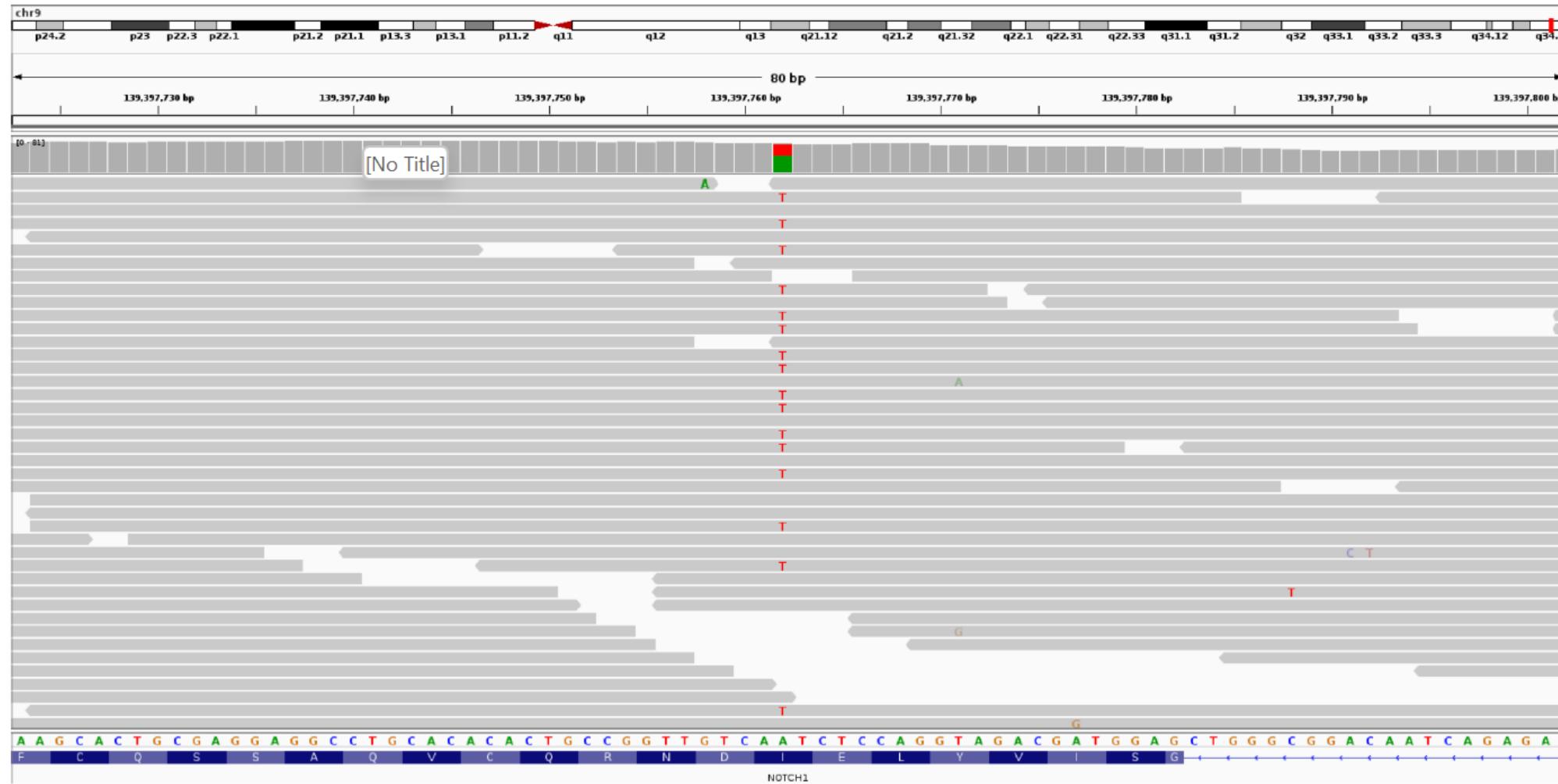
PAH protein  
Ref ...Ile-Asp...  
P1 ...Asn-Asp...  
  
NM\_000277.3(PAH):p.Ile324Asn

**Metabolomics**

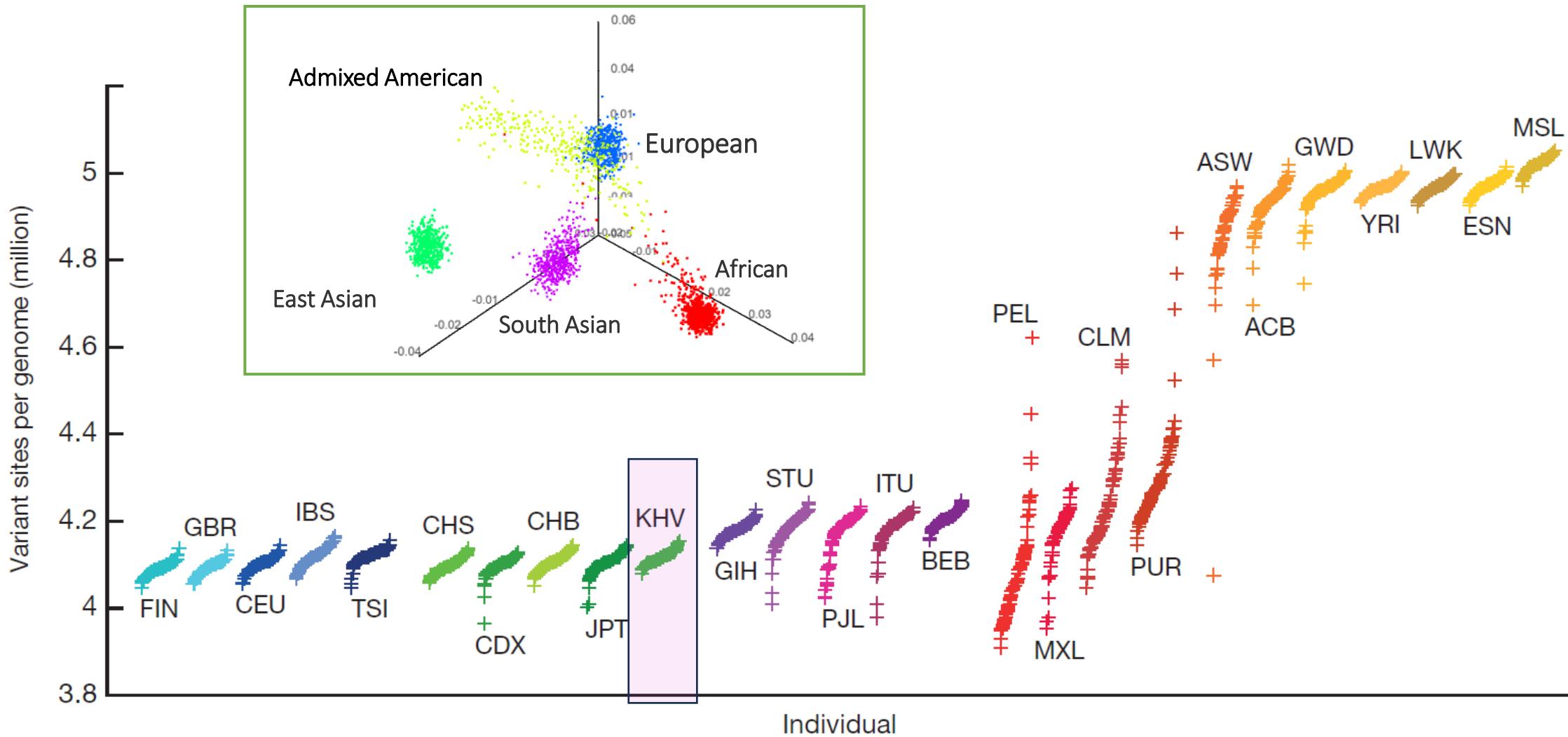
PAH  
Ref Phe → Tyr  
  
PAH  
P1 Phe ~~→~~ Tyr



# Alignment and variant viewers

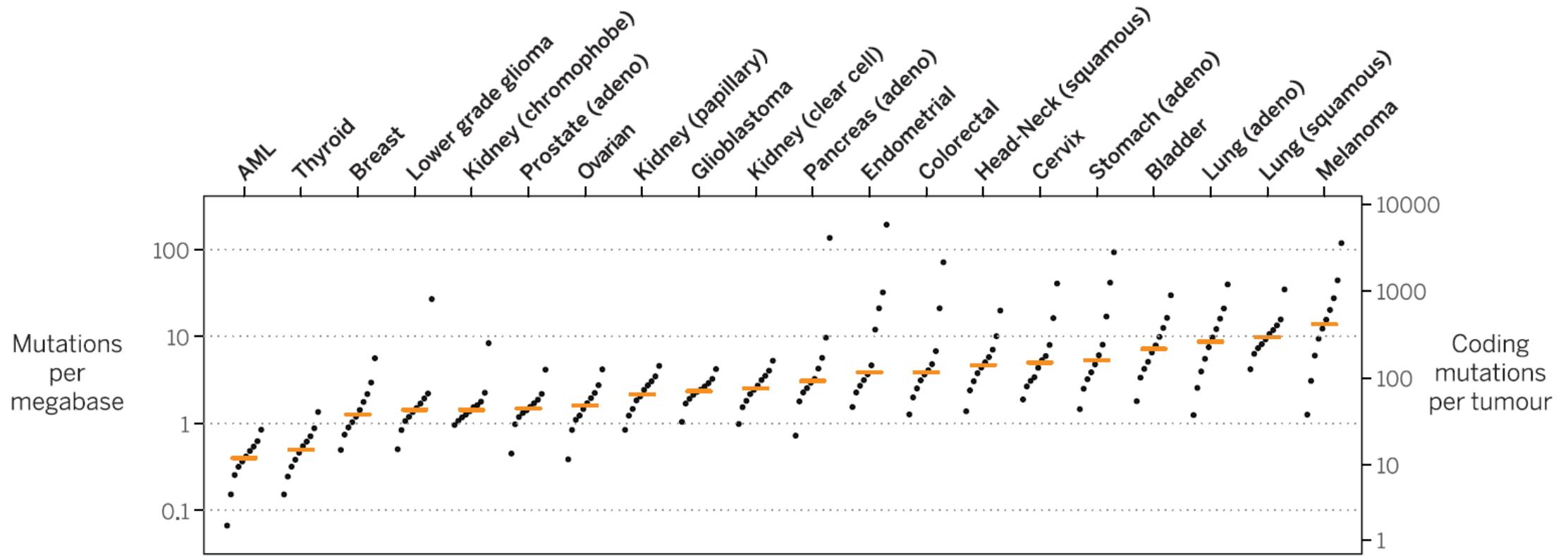


# Human Genome Variation: 1000 Genomes Project



The number of variant sites per genome of 1K human genomes project (2015)  
Kinh in Ho Chi Minh City, Vietnam (KHV)

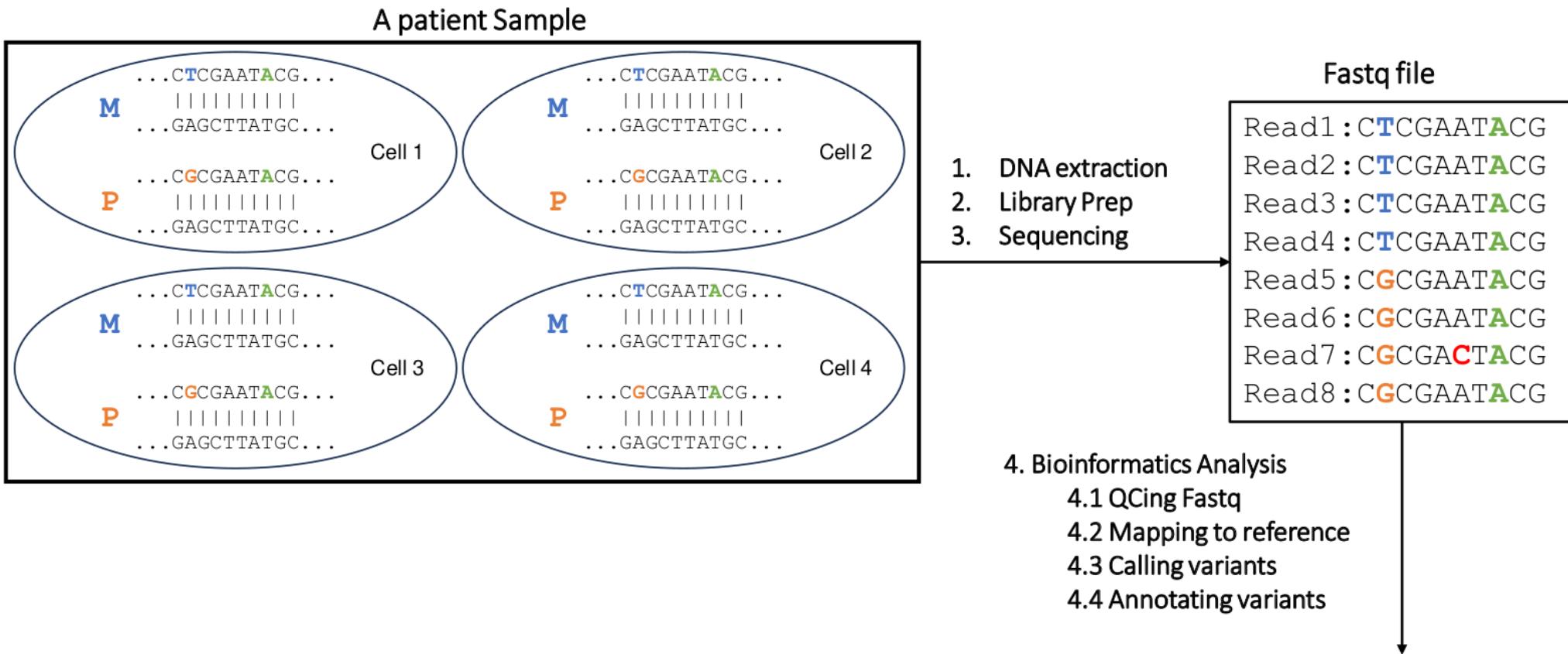
# Cancer Genome Somatic Variation



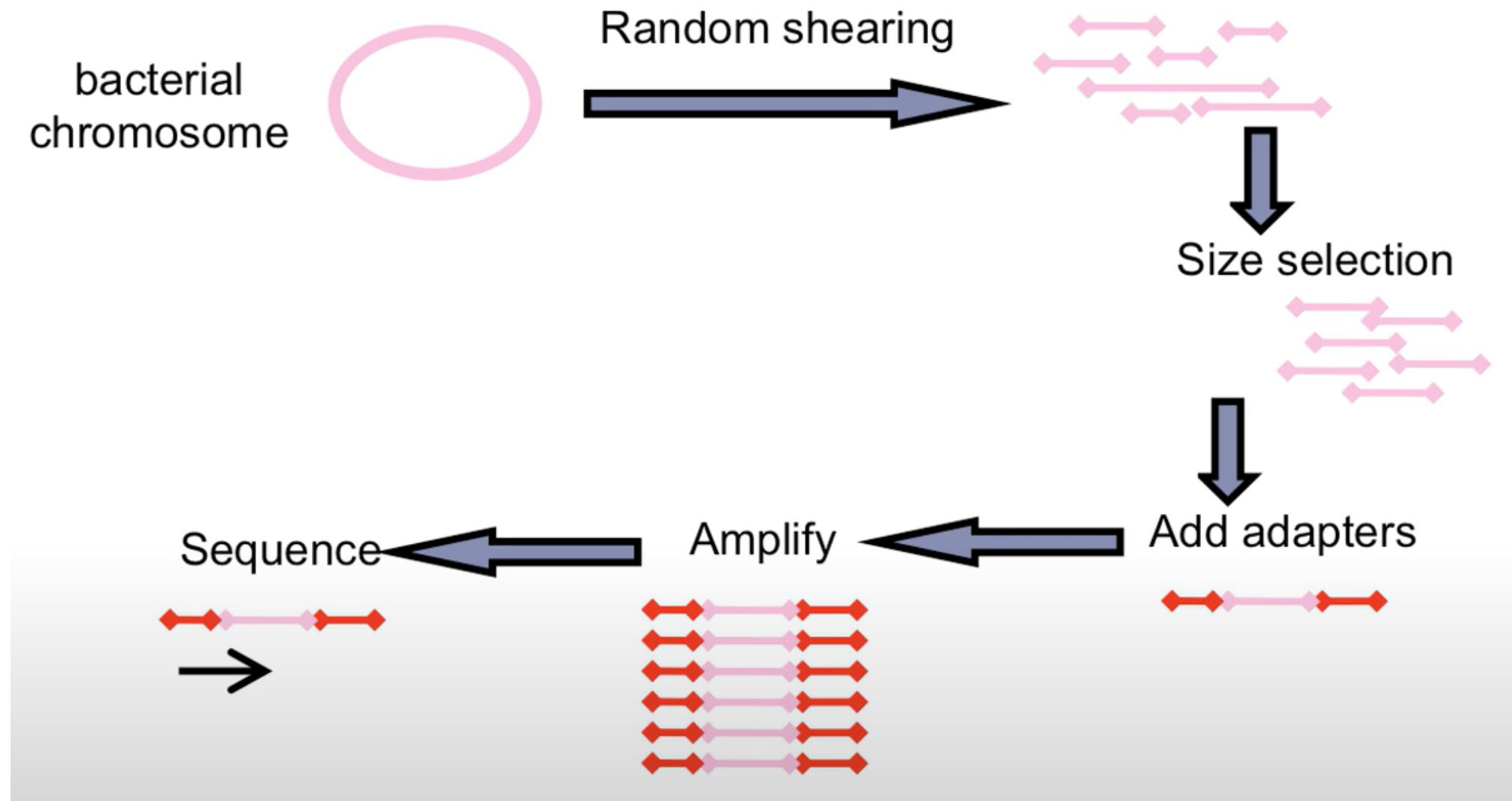
Mutation burden in 20 tumor types and relative contribution of different mutational processes.  
For each tumor type, samples were divided into deciles on the basis of their mutation burden. (2015)

# Quy trình XÉT NGHIỆM gen bằng phương pháp giải trình tự thế hệ mới (NGS)

# Các bước trong XÉT NGHIỆM gen bằng phương pháp giải trình tự thê hệ mới (1)



# Whole genome sequencing – Shotgun method



# Các bước trong XÉT NGHIỆM gen bằng phương pháp giải trình tự thế hệ mới (2)

## 4.2 Mapping reads to reference



Heterozygous

Homozygous

## 4.3 Calling variants

```
##fileformat=VCFv4.3
##FORMAT=<ID=GT,Number=1>Type=String>Description="Genotype">
##FORMAT=<ID=GQ,Number=1>Type=Integer>Description="Genotype Quality">
##FORMAT=<ID=DP,Number=1>Type=Integer>Description="Read Depth">
##FORMAT=<ID=AD,Number=2>Type=Integer>Description="Read depth for each allele">
```

#CHROM	POS	ID	REF	ALT	QUAL	FILTER	FORMAT	Sample1
20	14370	rs6054257	<b>T</b>	<b>G</b>	129	PASS	GT:GQ:DP:AD	0/1:48:8: <b>4,4</b>
20	17330	.	<b>G</b>	<b>A</b>	150	PASS	GT:GQ:DP:AD	1/1:49:8: <b>8,8</b>

ANN=G|stop\_gained|HIGH|OR4F5|ENSG00000186092|transcript|ENST0000641515.2|protein\_coding|3/3|c.822T>G|p.Trp274\*|882/2618|822/981|274/326||Pathogenic

ANN=A|frameshift\_variant|HIGH|ZSWIM2|ENSG00000163012|transcript|ENST00000295131.3|protein\_coding|9/9|c.1238G>A|p.Ile413|1293/2451|1238/1902|413/633||;LOF=(ZSWIM2|ENSG00000163012|1|1.00)

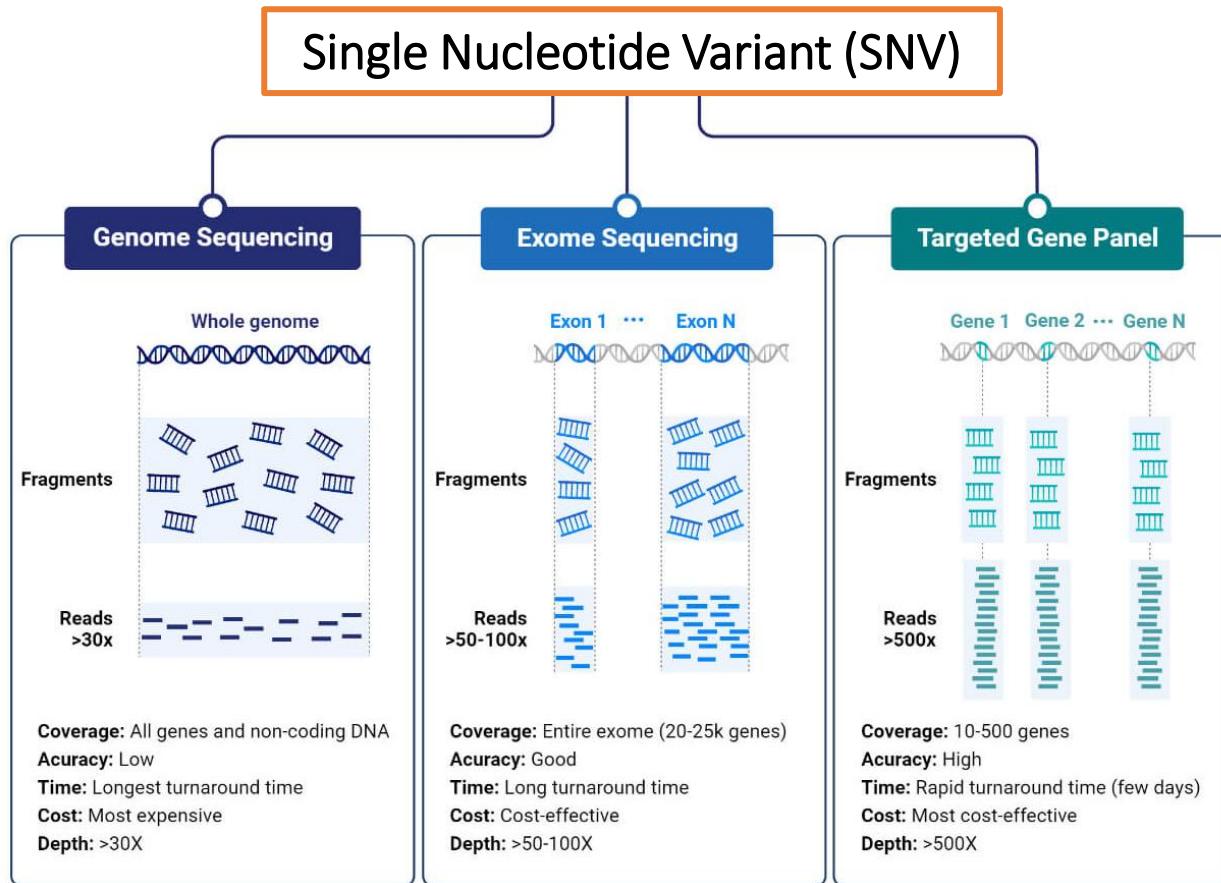
## 4.4 Annotating variants

# Hội đồng Hệ Gen: phiên giải và hội chẩn những biến thể trong báo cáo kết quả NGS

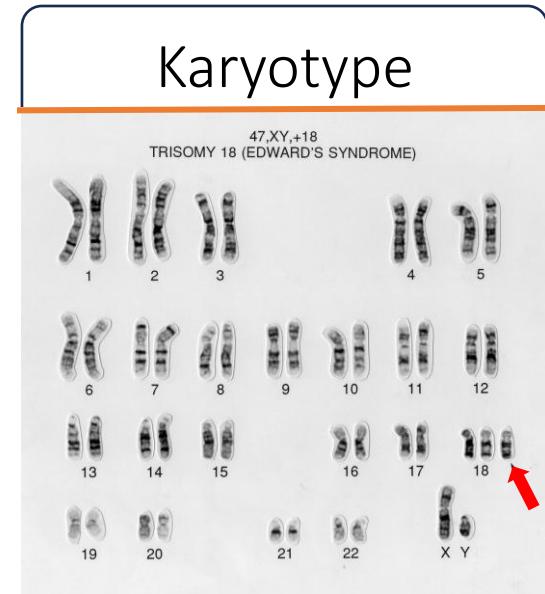
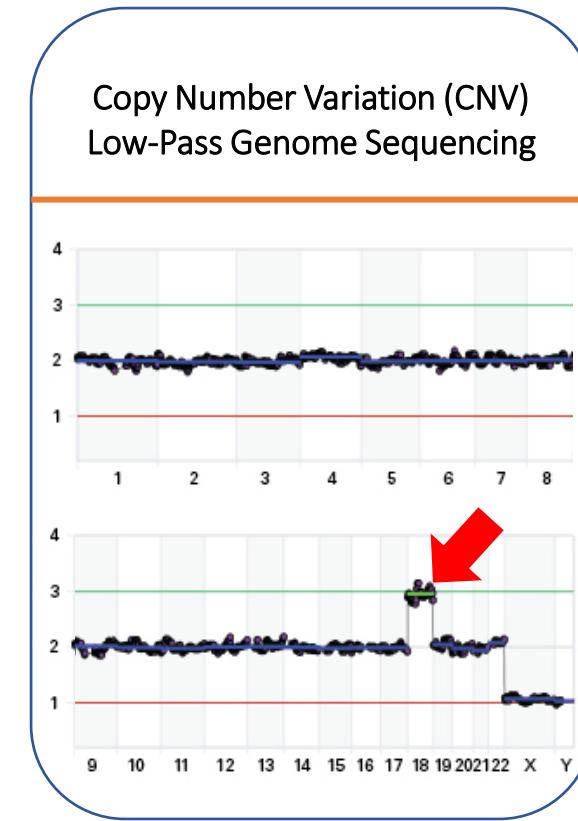


<https://ggba.swiss/en/the-first-medical-genomics-center-opens-in-geneva/>

# Ứng dụng giải trình tự gen thế hệ mới trong lâm sàng

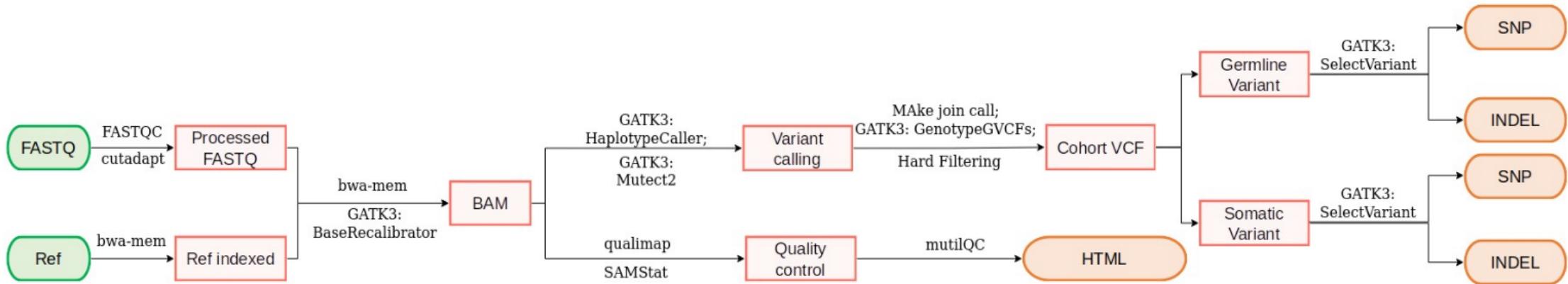


<https://microbenotes.com/next-generation-sequencing-ngs/>



<https://wellcomecollection.org/search/images?query=eaahzt2u#>

# Whole Genome Sequencing pipeline



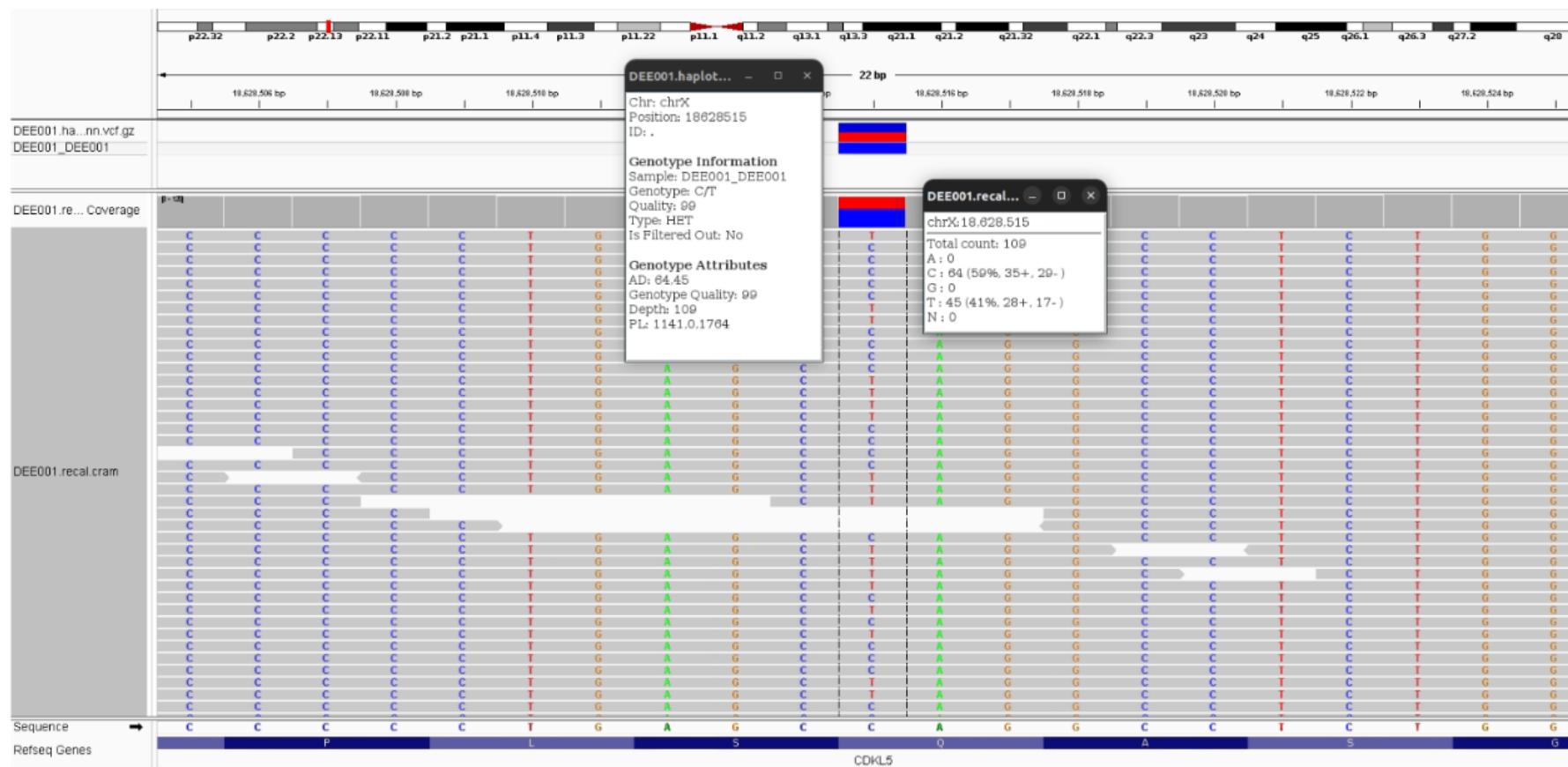
Package	Version	Link
FastQC	0.11.9	<a href="https://www.bioinformatics.babraham.ac.uk/projects/fastqc/">https://www.bioinformatics.babraham.ac.uk/projects/fastqc/</a>
cuadap	3.5	<a href="https://cutadapt.readthedocs.io/en/stable/">https://cutadapt.readthedocs.io/en/stable/</a>
bwa-mem	0.7.17	<a href="https://github.com/lh3/bwa">https://github.com/lh3/bwa</a>
qualimap	2.2.1	<a href="http://qualimap.bioinfo.cipf.es/">http://qualimap.bioinfo.cipf.es/</a>
SAMStat	1.08	<a href="http://samstat.sourceforge.net/">http://samstat.sourceforge.net/</a>
GATK	3.8	<a href="http://www.broadinstitute.org/gatk/">http://www.broadinstitute.org/gatk/</a>
multiqc	1.8	<a href="https://multiqc.info/">https://multiqc.info/</a>

# Năm ví dụ về XÉT NGHIỆM gen cho BỆNH DI TRUYỀN và UNG THƯ

# Ví dụ 1: Phát hiện sớm bệnh động kinh bằng xét nghiệm gen (1)

## Whole Exome Sequencing - WES

### IGV - DEE001 – CDKL5



Nguyen Le Duc Minh, MD

# Ví dụ 1: Phát hiện sớm bệnh động kinh bằng xét nghiệm gen (2)

## Whole Exome Sequencing - WES

DEE001							
Gene	Amino acid change	cDNA	Variant type	Allele frequency	Transcript	Variant effect	ClinVar significance
CPT2	p.Arg631Cys	c.1891C>T	SNP	0.5	ENST00000371486.4	Missense variant	Pathogenic
CDKL5	p.Gln881Ter	c.2641C>T	SNP	0.5	ENST00000623535.2	Stop gained (Nonsense)	Pathogenic
GALC		c.1162-4del	DEL (1bp)	1	ENST00000261304.7	Intron variant	Conflicting interpretations of pathogenicity
TUBB2B	p.Ala248Val	c.743C>T	SNP	0.5	ENST00000259818.8	Missense variant	Conflicting interpretations of pathogenicity

DEE001 – CDKL5 (Xp22.13)

Current Build 156  
Released September 21, 2022

**rs1057519541**

Organism	Homo sapiens	Clinical Significance	Reported in ClinVar
Position	chrX:18628515 (GRCh38.p14)	Gene : Consequence	CDKL5 : Stop Gained
Alleles	C>T	Publications	1 citation
Variation Type	SNV Single Nucleotide Variation	Genomic View	See rs on genome
Frequency	None		

**Clinical Significance**

Variant Details HGVS Submissions History Publications Flanks

Allele: T (allele ID: 362353 )

ClinVar Accession ▲ Disease Names ▷ Clinical Significance

RCV000416943.1 Focal epilepsy Pathogenic

[https://www.ncbi.nlm.nih.gov/snp/rs1057519541#clinical\\_significance](https://www.ncbi.nlm.nih.gov/snp/rs1057519541#clinical_significance)

NGUYEN Thuy-Minh-Thu, MD  
Nguyen Le Duc Minh, MD

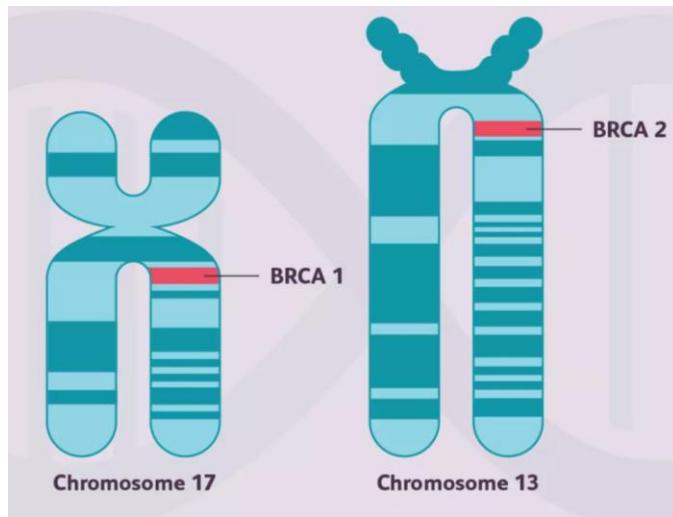
## Ví dụ 2: Sàng lọc bệnh ung thư bằng xét nghiệm gen Gene Panel – 53 genes của MGI

Gene	Amino Acid Change	Coding	Variant Type	Allele Frequency	Transcript	Variant effect	ClinVar Significance
CTNNB1	p.Ser33Tyr	c.98C>A	SNP	0.5	ENST00000349496	MISSENSE	Pathogenic/ Likely_pathogenic
PIK3CA	p.Gly914Arg	c.2740G>A	SNP	0.5	ENST00000263967	MISSENSE	Pathogenic
KRAS	p.Gly12Asp	c.35G>A	SNP	0.5	ENST00000256078	MISSENSE	Pathogenic
BRCA2	p.Ile2675AspfsTer6	c.8021dup	INS	0.5	ENST00000544455	FRAMESHIFT	Pathogenic

Nguyen Le Duc Minh, MD

# Ví dụ 3: Hỗ trợ điều trị bệnh ung thư vú bằng xét nghiệm gen Gene Panel BRCA1 và BRCA2

Olaparib (AstraZeneca) là một loại thuốc dùng để duy trì điều trị ung thư vú, buồng trứng, tuyến tiền liệt và tuyến tụy giai đoạn tiến triển có đột biến BRCA ở người lớn.



## KẾT QUẢ XÉT NGHIỆM BRCA1/2

Họ và tên : TRẦN THỊ X.	Tuổi : 1956	Giới tính : NỮ
Số hồ sơ:		
KHOA:	BS điều trị:	
Bệnh phẩm : Mô vúi nến	Số block: XXXX	
Yêu cầu: Xét nghiệm giải trình tự gen trên hệ thống MiSeq [02 gen BRCA1 và BRCA2]		
Ngày nhận chỉ định: 20/10/2022	Ngày thực hiện: 25/10/2022	

**Chẩn đoán lâm sàng:** Ung thư buồng trứng dịch trong grade cao/ Ung thư vú trái

**Chất lượng mẫu:** MẪU ĐẶT (kích thước 17mm x 16mm, thành phần bướu 70%)

**Phương pháp:** Giải trình tự gen bằng phương pháp NGS cho 02 gen BRCA1 và BRCA2

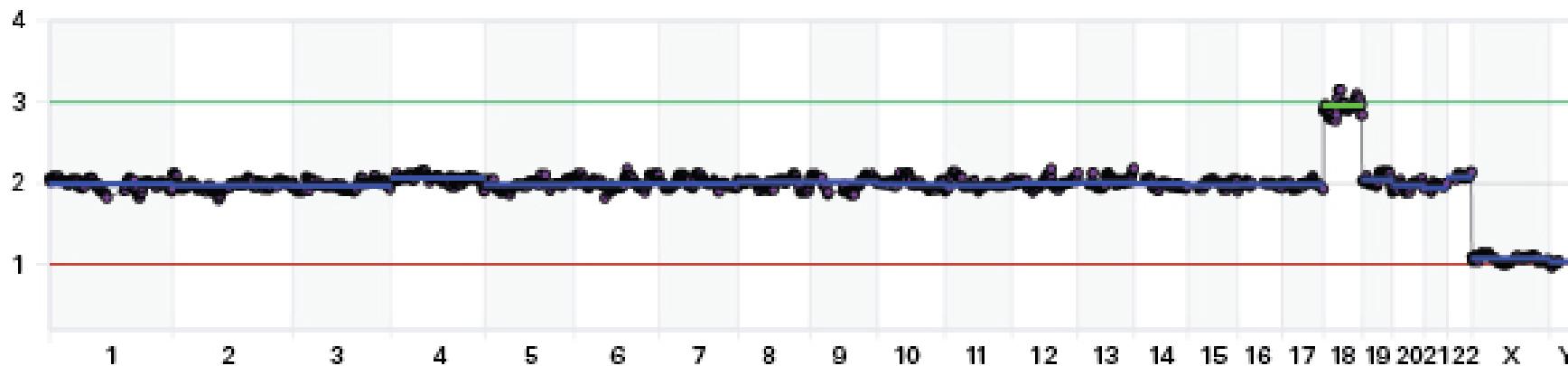
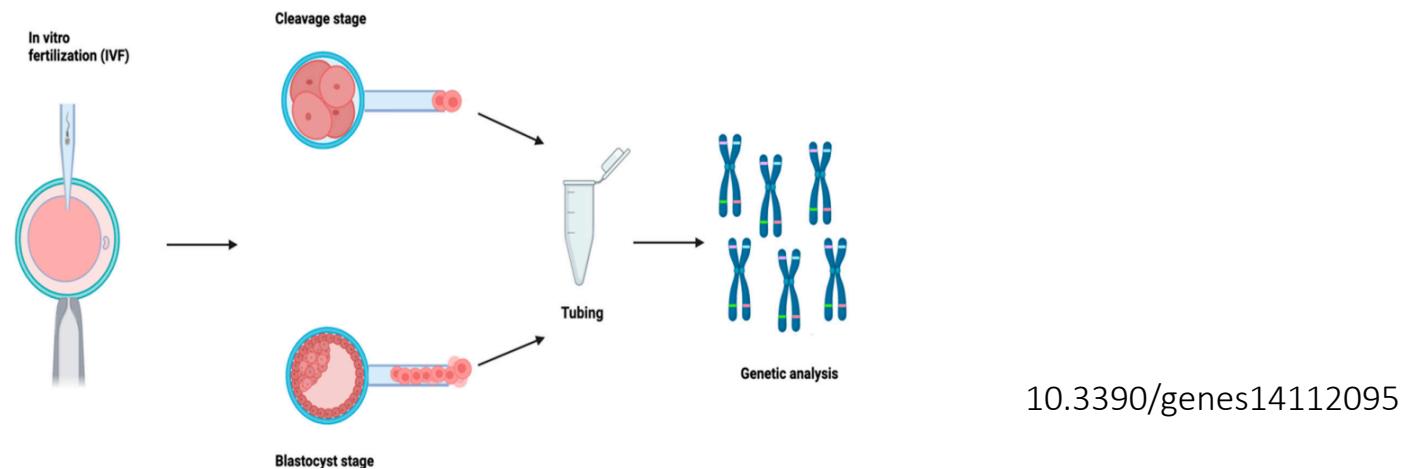
- Hệ thống xét nghiệm Illumina MiSeqDx (CE/US-IVD)
- Bộ xét nghiệm: NGeneBio BRCAaceuTest™Plus (CE-IVD)
- Phần mềm hỗ trợ phân tích kết quả: NGeneBio NGeneAnalySys™ (CE-IVD)

**Kết quả:** PHÁT HIỆN 1 BIÊN THẾ MẤT ĐOẠN NUCLEOTIDE NHỎ (DEL) GÂY BỆNH TRÊN GEN BRCA1

MÔ TẢ KẾT QUẢ			
Gen	Biến thể gây bệnh/có khả năng gây bệnh	Tỷ lệ	Phân loại
BRCA1	c.5335del (p.Gln1779AsnfsTer14)	82.03%	Gây bệnh (Pathogenic)
BRCA2	Không phát hiện	Không	Không

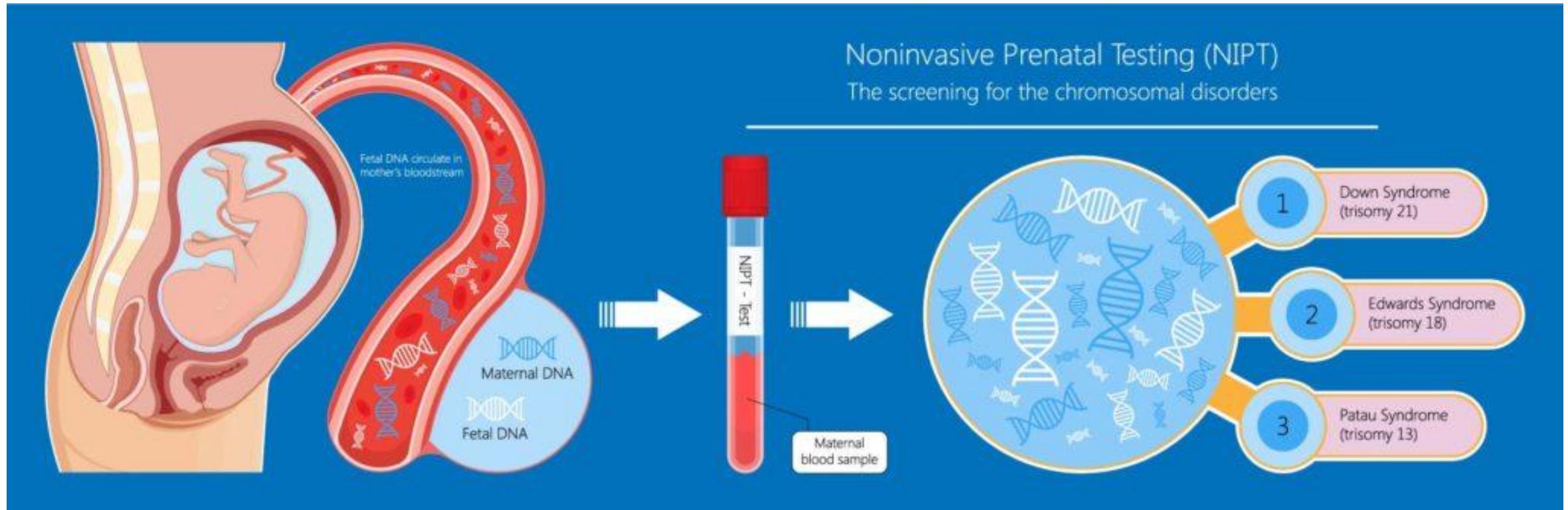
# Ví dụ 4: Sàng lọc phôi trong hỗ trợ sinh sản IVF

Xét nghiệm tiền làm tổ PGT-A



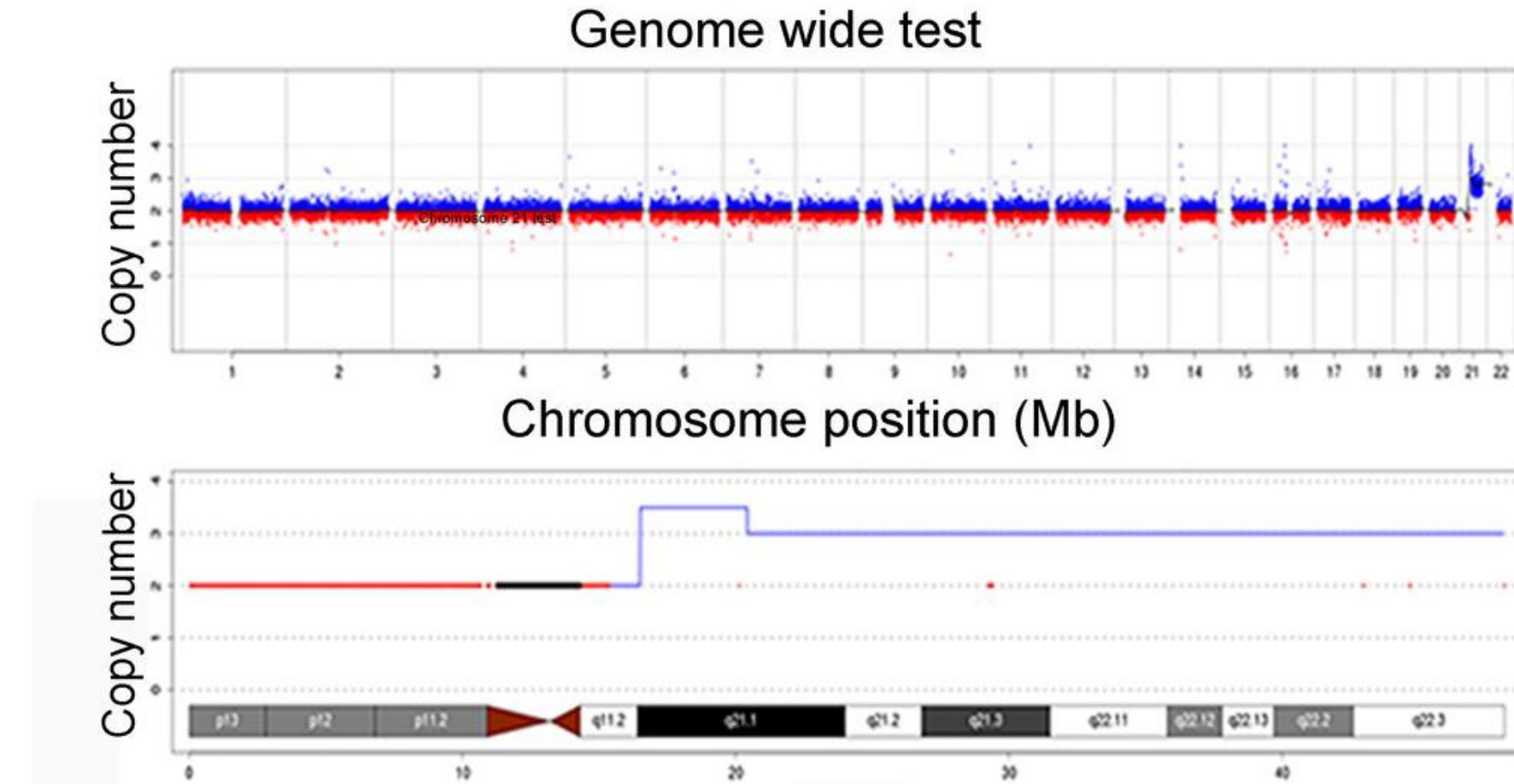
Mẫu phôi có 3 nhiễm sắc thể 18 (trisomy) trong DNA hệ gen

## Ví dụ 5: Sàng lọc bất thường nhiễm sắc thể ở thai nhi



<https://www.advancedwomensimaging.com.au/nipt/>

# Kết quả NIPT: Bất thường nhiễm sắc thể 21 ở thai nhi



# Ví dụ: Hỗ trợ điều trị ung thư tuyến tiền liệt kháng và di căn

## Serum Free Methylated Glutathione S-transferase 1 DNA Levels, Survival, and Response to Docetaxel in Metastatic, Castration-resistant Prostate Cancer: Post Hoc Analyses of Data from a Phase 3 Trial

Kate L. Mahon <sup>a,b,c,e,†</sup>, Wenjia Qu <sup>b,†</sup>, Hui-Ming Lin <sup>b,c</sup>, Calan Spielman <sup>b</sup>, Daniel Cain <sup>d</sup>, Cindy Jacobs <sup>d</sup>, Martin R. Stockler <sup>a,e,f</sup>, Celestia S. Higano <sup>g</sup>, Johann S. de Bono <sup>h</sup>, Kim N. Chi <sup>i</sup>, Susan J. Clark <sup>b,c,†</sup>, Lisa Glen Horvath <sup>a,b,c,e,†,\*</sup>

<sup>a</sup> Chris O'Brien Lifehouse, Sydney, Australia; <sup>b</sup> Garvan Institute of Medical Research, Sydney, Australia; <sup>c</sup> University of NSW, Sydney, Australia; <sup>d</sup> Oncogenex Pharmaceuticals Inc., Bothell, WA, USA; <sup>e</sup> University of Sydney, Sydney, Australia; <sup>f</sup> National Health and Medical Research Council Clinical Trials Centre, Sydney, Australia; <sup>g</sup> University of Washington, Fred Hutchinson Cancer Research Centre, Seattle, WA, USA; <sup>h</sup> Royal Marsden Hospital and Institute of Cancer Research, London, UK; <sup>i</sup> University of British Columbia, BC Cancer Agency, Vancouver Prostate Centre, Vancouver, BC, Canada

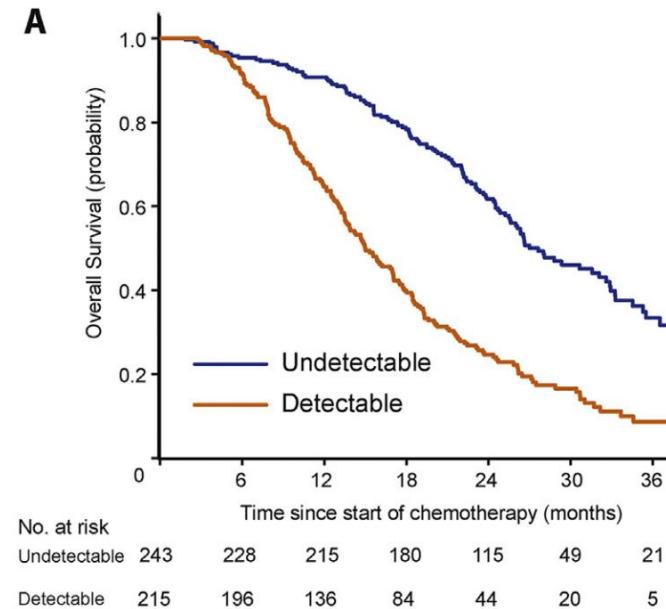
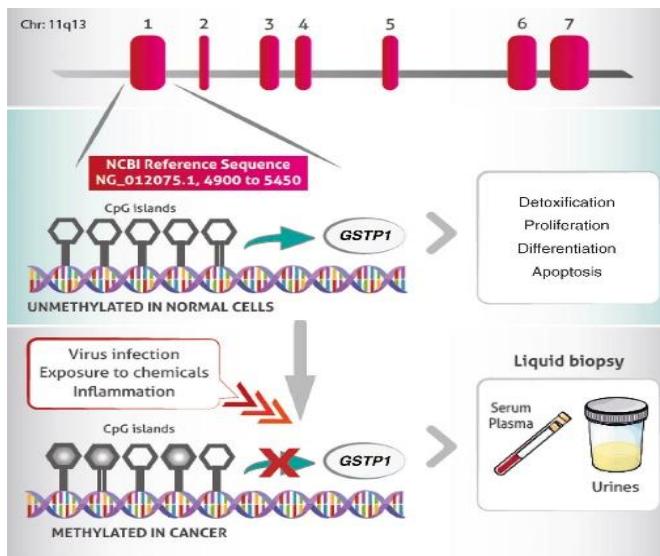
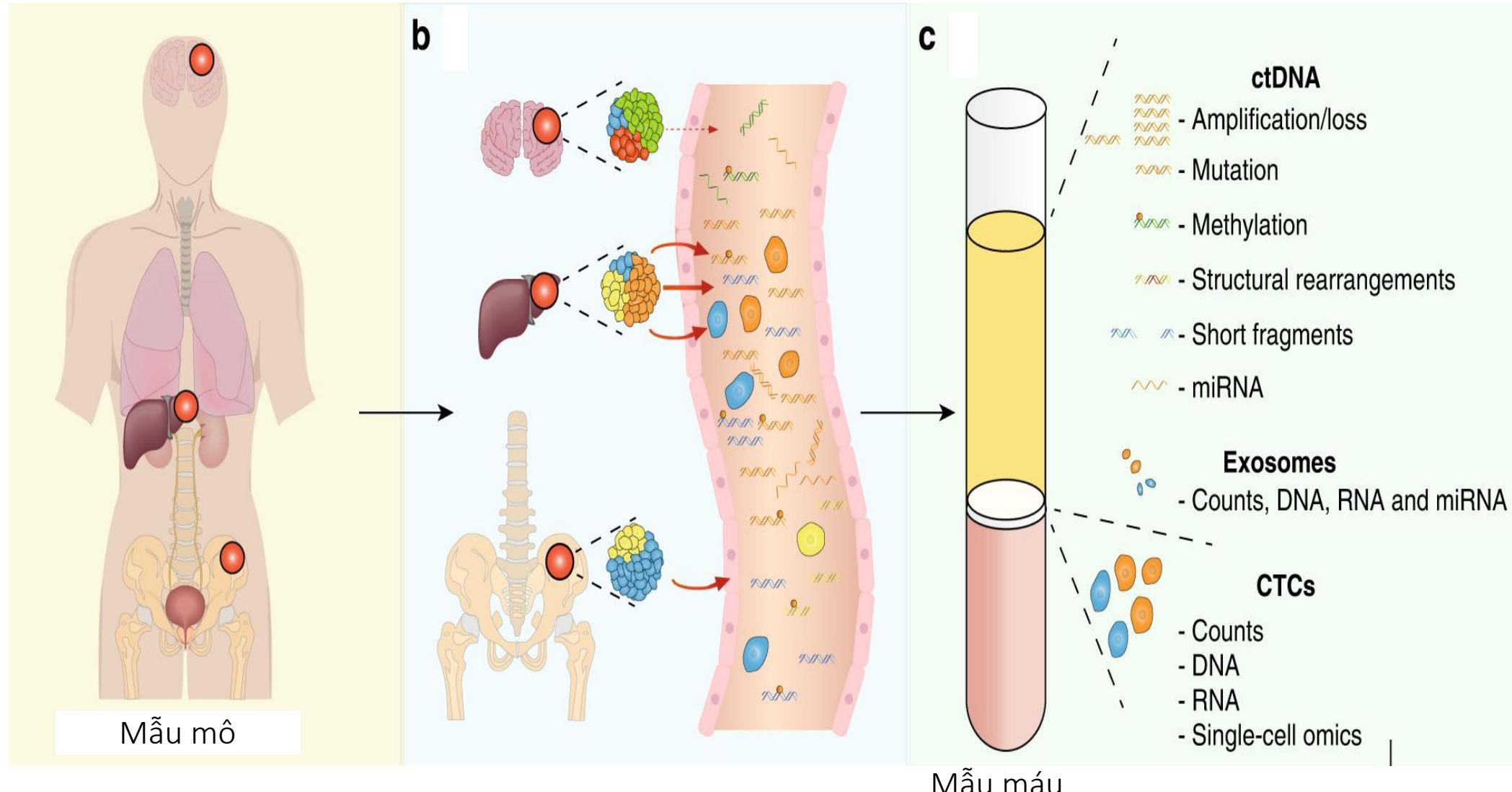
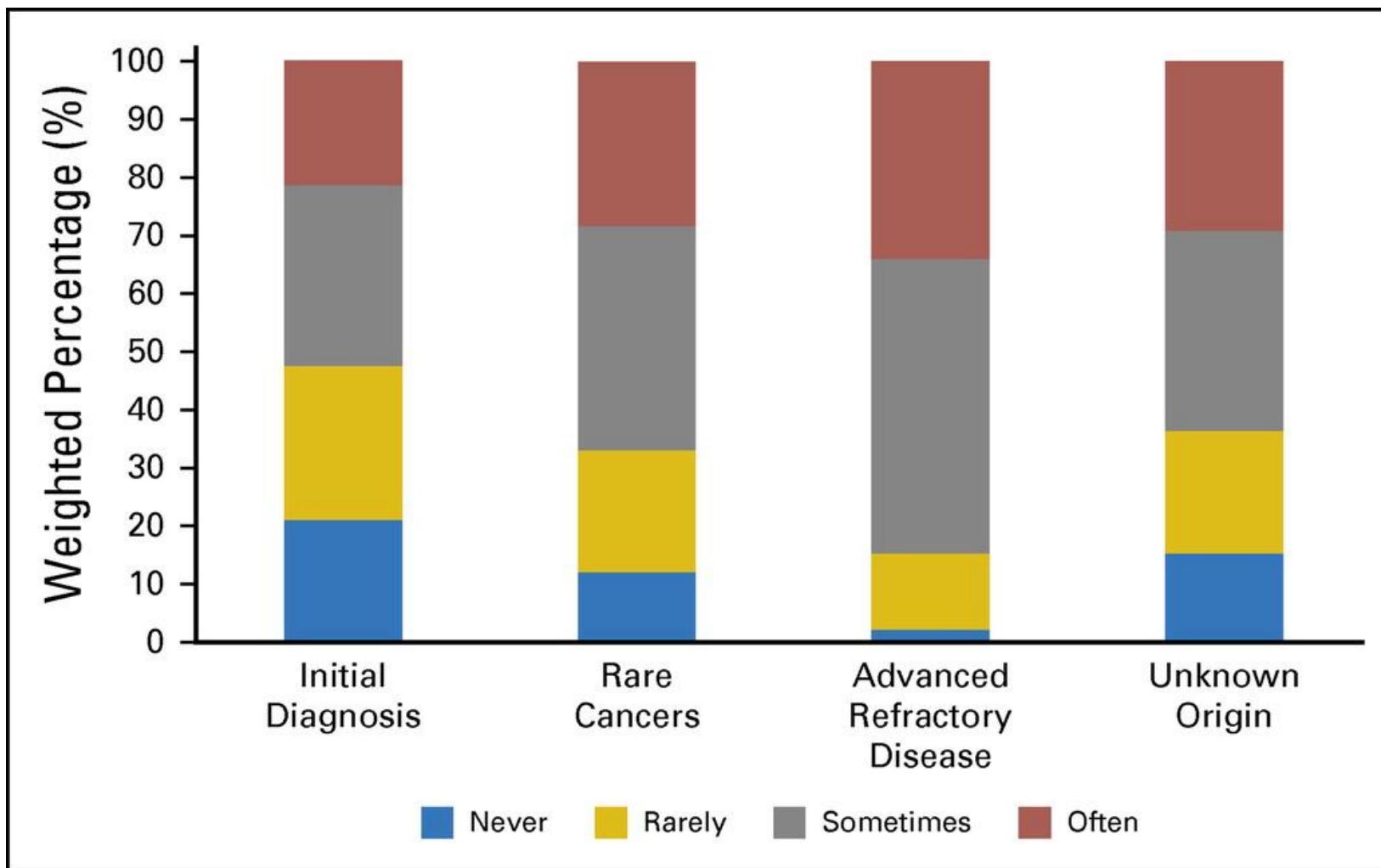


Fig. Kaplan-Meier estimates of survival in patients with a detectable mGSTP1 at baseline. (A) OS according to mGSTP1 detection after two cycles of docetaxel.

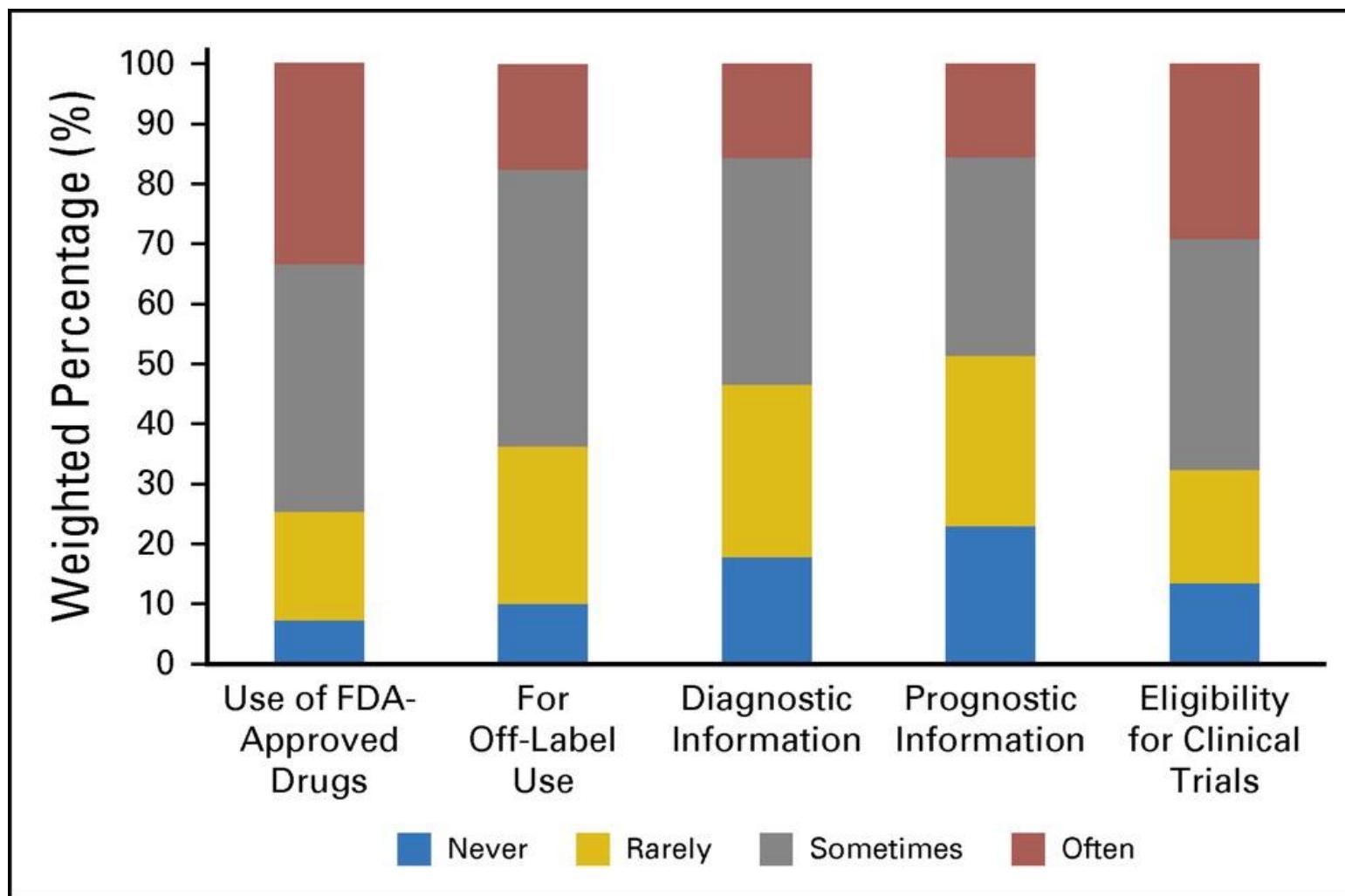
# Mẫu cho XÉT NGHIỆM gen: mẫu mô và máu



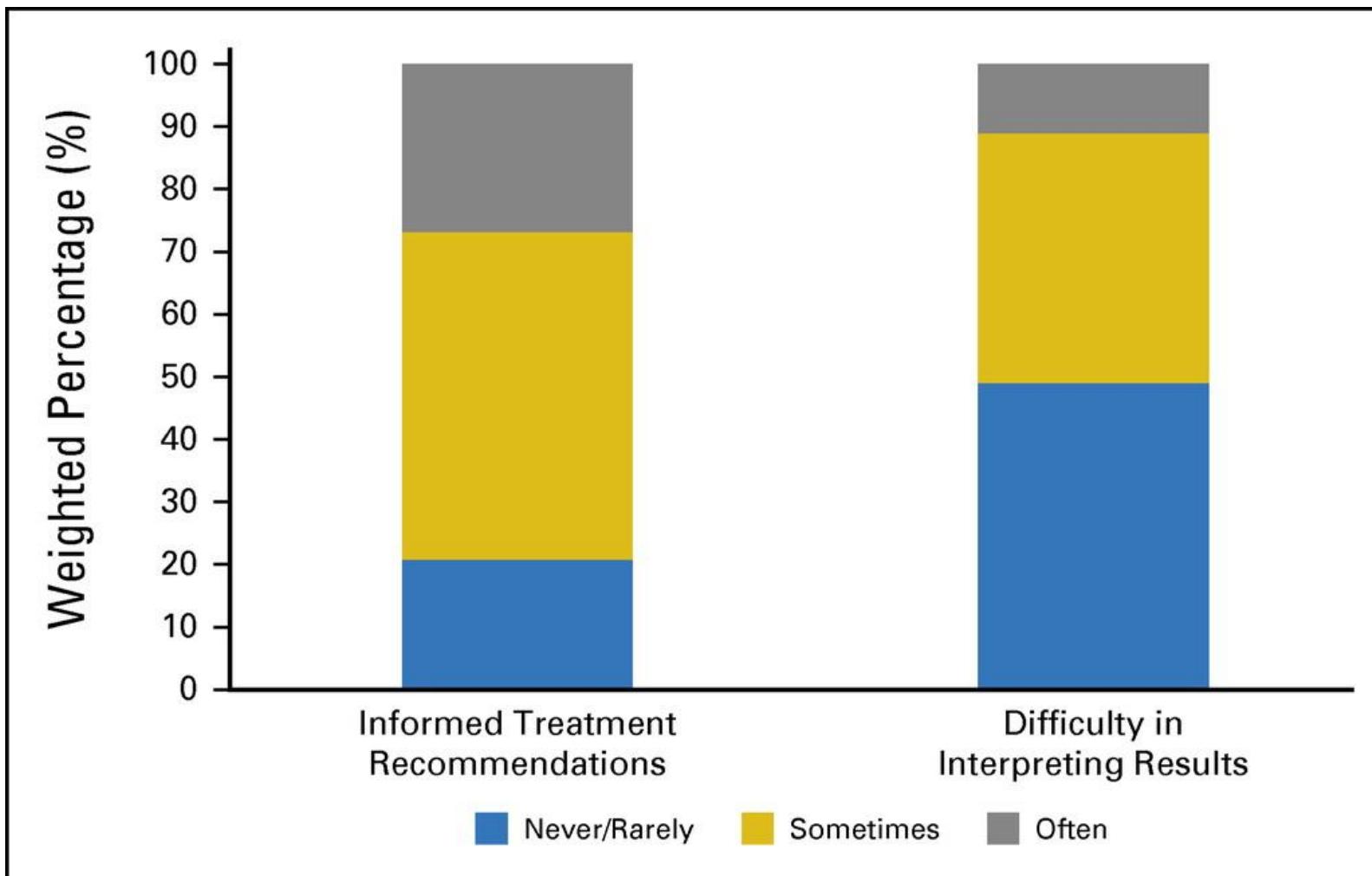
# NGS is used in clinical cancer care



# NGS is used in clinical cancer care



# NGS is used in clinical cancer care



# Xin gửi lời cảm ơn đến

## Nhóm nghiên cứu

- Bác sĩ Nguyễn Lê Đức Minh
- Thạc sỹ Đào Khương Duy
- Thạc sỹ Ngô Đại Phú
- Thạc sỹ Trần Thị Mỹ Qui
- Bác sĩ NGUYEN Thuý-Minh-Thu
- Thạc sỹ, bác sĩ Nguyễn Huy Thịnh
- Thạc sỹ Như
- KS Nguyễn Minh Hoàng

## Nhóm cộng tác nghiên cứu

- Trần Quang Hải
- TS Nguyễn Anh Thư
- Hoàng Kim
- Huy Hà
- Hoàng Sơn
- TS Bác sĩ Nguyễn Thị Kim Nhi
- Bác sĩ Hậu
- PGS TS Bác sĩ Phạm Lê An

# Xin chân thành cảm ơn!

Luu Phuc Loi, PhD

Email: [luu.p.loi@googlemail.com](mailto:luu.p.loi@googlemail.com)

Zalo: 0901802182