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

Ngày 15 tháng 02 năm 2025

TS. Lưu Phúc Lợi

Email: luu.p.loi@googlemail.com

Zalo: 0901802182

Nội dung

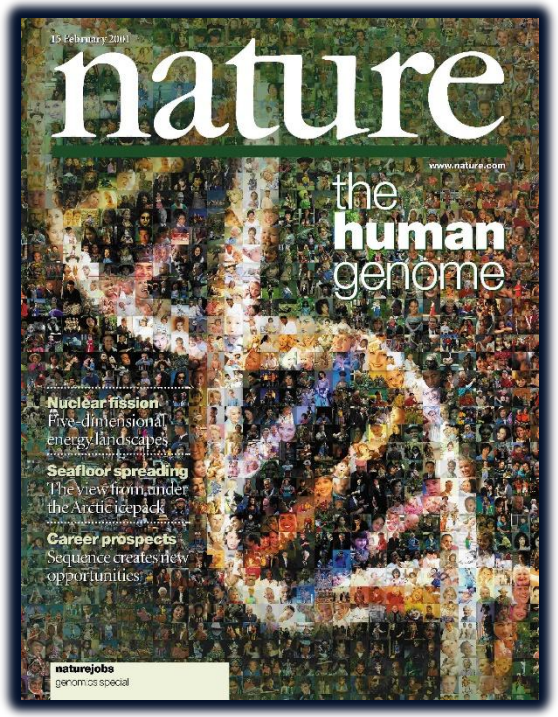
1. Giới thiệu về Dự án bộ gen người
2. Giới thiệu kỹ thuật giải trình tự thế hệ mới (NGS)
3. Quy trình XÉT NGHIỆM gen bằng phương pháp giải trình tự thế hệ mới
4. Năm ví dụ về XÉT NGHIỆM gen cho BỆNH DI TRUYỀN và UNG THƯ
5. Giải trình tự gen thế hệ mới trong nông nghiệp và môi trường

Giới thiệu về
Dự án bộ gen người

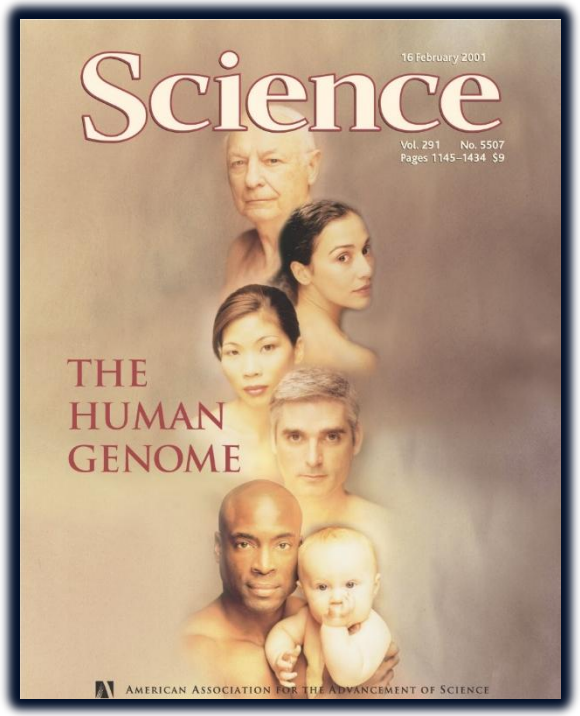
Dự án hệ gen người HGP (Oct 1990 - April 2003)

- 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)

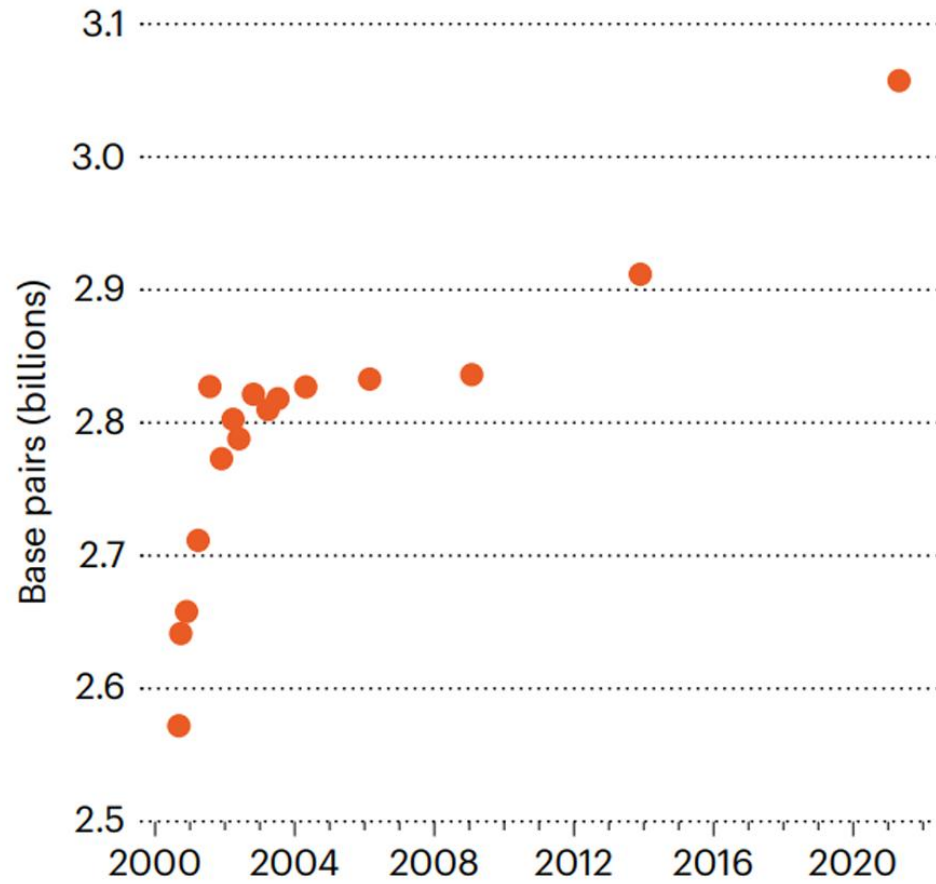
https://genome.ucsc.edu/FAQ/FAQreleases.html

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

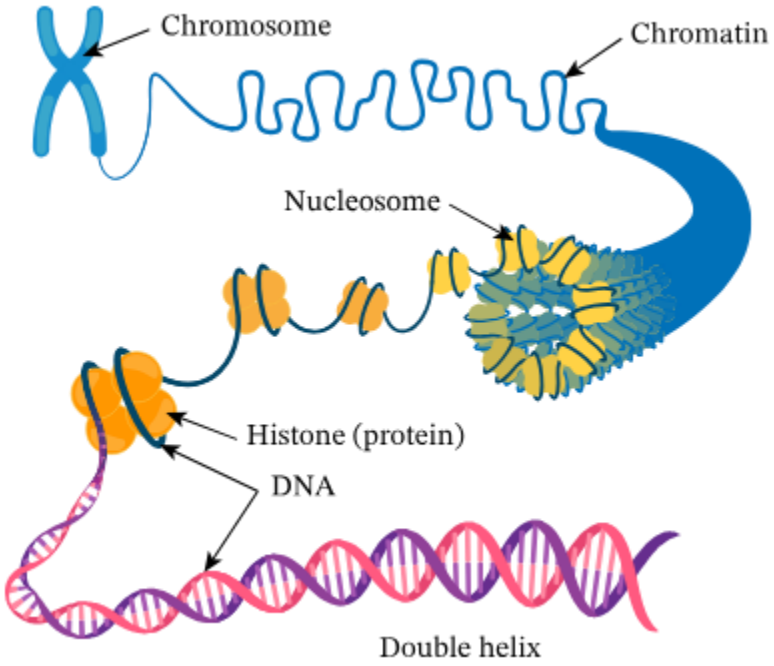
Researchers have been filling in incompletely sequenced parts of the human reference genome for 20 years, and have now almost finished it, with 3.05 billion DNA base pairs.

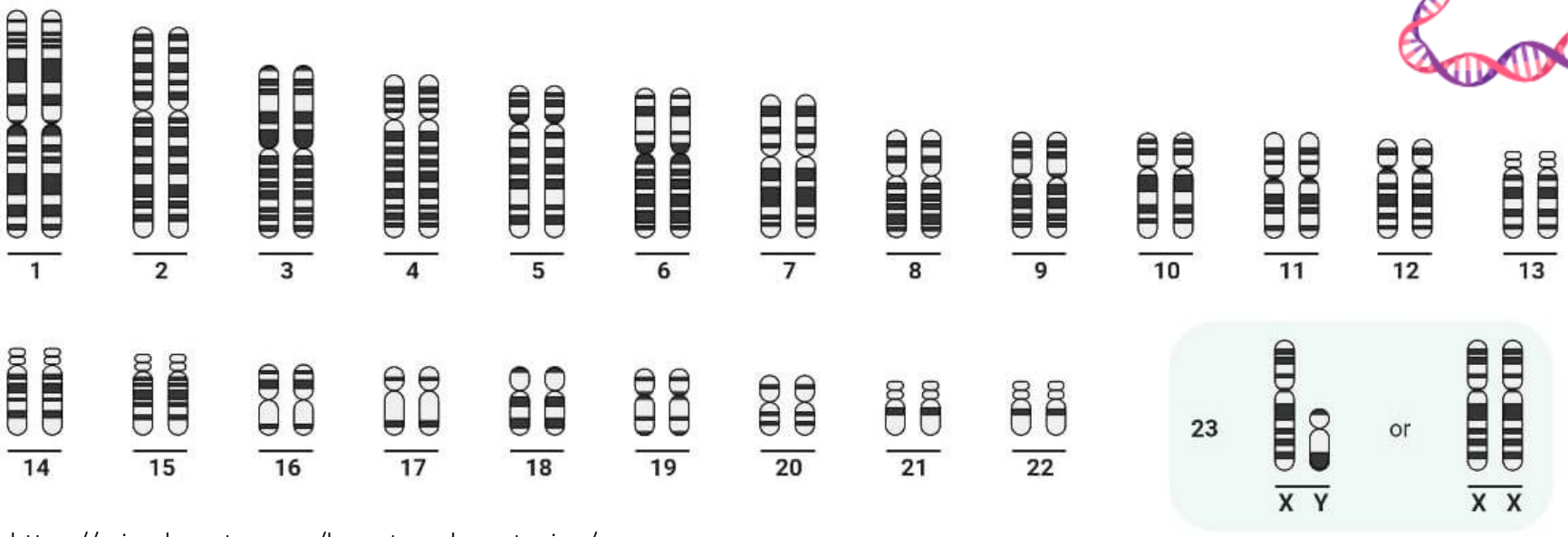


0.3% of sequence might still have errors. Includes X but not Y chromosome. Count excludes mitochondrial DNA.

SOURCE: ADAM PHILLIPPY

Human Karyotype





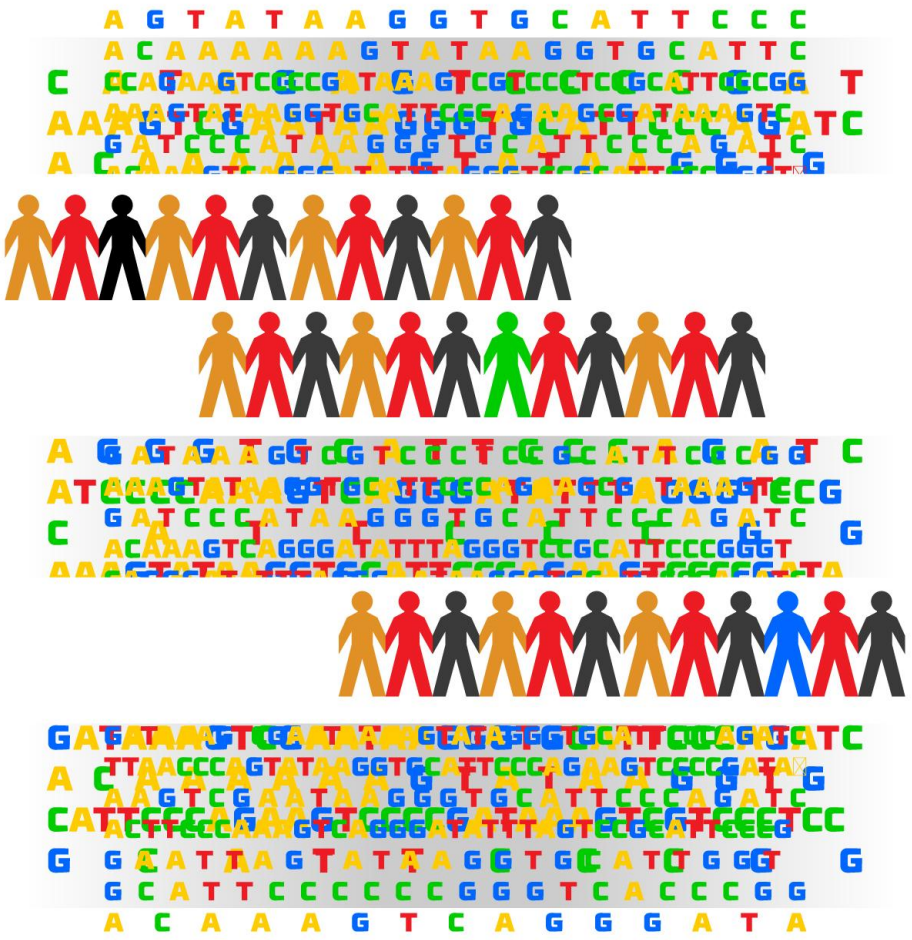
https://microbenotes.com/karyotype-karyotyping/

Hệ Gen người

```
48541 agcccttcaa agaaatgttc tcagcaggca tggagcccag gacttgctcc ctttgggtag
48601 agagccgggt tgaaggtgac tgaagtgaaa tgggacagta gaggcggggg gggtggtgag
48661 ttcctggagg tgggggtgt gggaacctgc tgtgtactga gatgcacccc tgccagttct
48721 gcctgaagat ttgaggcggg gggcaggggg gcggagtgaa gtcattttac tggtaagtaa
48781 ttttaaacct tttaatatta aagcaaacgt ggatatgtaa tgaatgaaat tcattctgga
48841 atgaaaaatt cacgtgatgt tgaaaaataa cacggggctt cagagaggac tttctggctg
48901 gcagcagact ccagattccc agggcccctg caccctcctc tgcccacagg gcaccttaat
48961 ggagaaggtg tgggaggaga gccaggccgg agtcagagca cactggtgac tccacatttg
49021 cagcgtgccc tgcctctctc ctgaggcttg gcaacgtgca atatgctaag caaactcccc
49081 ctgtccccgt ccagtttctg aggacaagag ccaccacctg tagcaaataa agacccagca
49141 acccttttgac tcatctttgt gagtctctgg aatcagaggg tagccacatc gctgagaggt
49201 ggagtgaagc actcgggtga aaaggtacaa ggaagtcagg gacaggagtg tggggacatc
49261 acctagacaa tgacagagaa gaggggcaca gccgagtgag gggagagggg ccggcagtcc
49321 tacatcccct ggcctgaagc acgctccagg gcagaaggaa aaacactgtc tttggggtcc
49381 aagagacctg agttcaaatt ctggctccac cactgaccac ctgtgtaacc ttgaactgct
49441 gctgcctgaa cctcaggttc cccttctaaa aatagaggag aaaaggatgc atttctcctt
49501 gcccctgtga gaacgaaatg gtgcaagcac caaggagcct cagcaaaggt cgggcctgcc
49561 cccgcctggc caaacctttc ctcttcagga ggccacggca accgtagttt gacagaagag
49621 cagcaccttg atttaatgct tcccagcatg tgtccttgag caagtcacct aacctctctg
49681 ggctgcttcc tcattgggaa aatatggctg ccagtaaaac ctgccctgtc cacctcctgg
49741 ggcacttggc aaacagcaaa agagtccaaa tgtgcaggct gggccaggcg cagtggctca
49801 tgcctgtaat cccagcaatt taggaagcca aggtgggcgg atcacctgag gtcaggagtt
49861 tgagaccagc ctggccaaca tggtgaaacc ttgtctctac aaaaatacaa aaattagccg
49921 ggcatgatgg cgggtgcctg taatcccagt tactcgggag gctgaggcaa gagaatcgct
49981 tgaacccgga aggggaaggt tgcagtgagc caagattgtg ccactgcact ccagcctggg
50041 caacagagcg agactctgtc tcaaaaaaaa aaaaaaaaaa aaacaatgca gagctggctg
50101 tgtaaaaaac ctgttccact gcagggccca gtgtccacca ggctggggtg caggcctatg
50161 gggtgggggc ccagcatcag cctctcagca gccctgggag gcggggcgca tcccgtgccc
50221 ctcgtggtct ggatgtgttc tagcccaagt cctaggttac acctgccgtc gcctggcctc
50281 tcaggagagg cccagggtga ggaggagcat ggtaaaggtg aagctgattg ggaagtcagc
50341 tgttgggaaa gcaactcctt gcacattgga ggaaccgaga aagactgacc ccgaggacag
50401 cagccagcat ggcccttcct gggagcccat gttgggggat tcctgctgca gccaaggctc
50461 agcccttgtg gtcgcaggtg ctggtcctgg cctcttcccc tcccatgcag gagcacagga
50521 gagatggctt ctgaggacct gttgcagctg tggccctggg aatagatttg ccagggagct
50581 ttaaagcagc tgagtgtgtc atccagctaa gcctggggaa ggagcttggc tcaggtcctg
50641 acaggtgtga cagggatggg gactgggaag taagagatga aaccctggct ggaggctgtg
50701 agcttccaca gccagcgctg gacagggagg gtccagatat acccactagt gccctcacca
```

https://www.ncbi.nlm.nih.gov/nuccore/806904736

Dữ liệu hệ gen cho quần thể



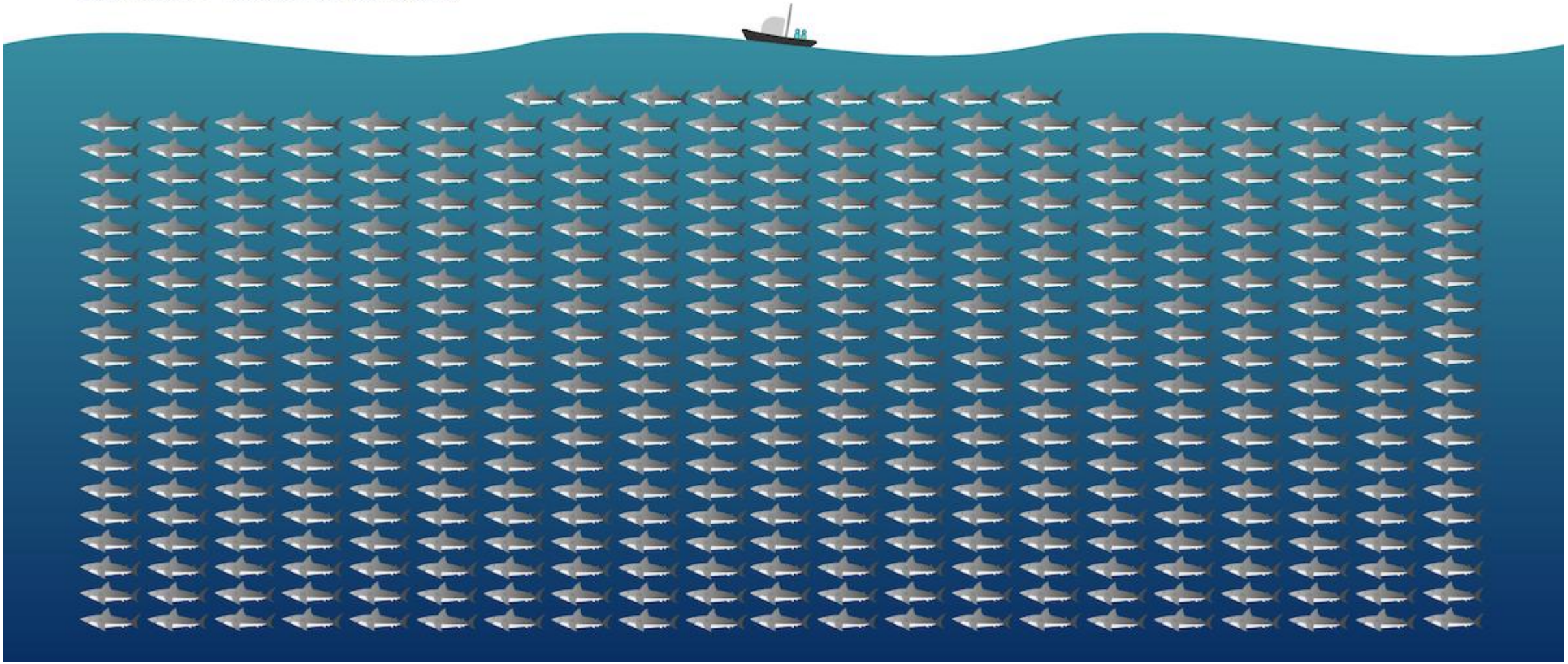
```
48541 agcccttcaa agaaatgttc tcagcaggca tggagcccag gacttgctcc ctttgggtag
48601 agagccgggt tgaaggtgac tgaagtgaaa tgggacagta gaggcggggg gggtggtgag
48661 ttcctggagg tgggggtgt gggaacctgc tgtgtactga gatgcacccc tgccagttct
48721 gcctgaagat ttgaggcggg gggcaggggg gcggagtgaa gtcattttac tggtaagtaa
48781 ttttaaacct tttaatatta aagcaaacgt ggatatgtaa tgaatgaaat tcattctgga
48841 atgaaaaatt cacgtgatgt tgaaaaataa cacggggctt cagagaggac tttctggctg
48901 gcagcagact ccagattccc agggcccctg caccctcctc tgcccacagg gcaccttaat
48961 ggagaaggtg tgggaggaga gccaggccgg agtcagagca cactggtgac tccacatttg
49021 cagcgtgccc tgcctctctc ctgaggcttg gcaacgtgca atatgctaag caaactcccc
49081 ctgtccccgt ccagtttctg aggacaagag ccaccacctg tagcaaataa agacccagca
49141 acccttttgac tcatctttgt gagtctctgg aatcagaggg tagccacatc gctgagaggt
49201 ggagtgaagc actcgggtga aaaggtacaa ggaagtcagg gacaggagtg tggggacatc
49261 acctagacaa tgacagagaa gaggggcaca gccgagtgag gggagagggg ccggcagtcc
49321 tacatcccct ggcctgaagc acgctccagg gcagaaggaa aaacactgtc tttggggtcc
49381 aagagacctg agttcaaatt ctggctccac cactgaccac ctgtgtaacc ttgaactgct
49441 gctgcctgaa cctcaggttc cccttctaaa aatagaggag aaaaggatgc atttctcctt
49501 gcccctgtga gaacgaaatg gtgcaagcac caaggagcct cagcaaaggt cgggcctgcc
49561 cccgcctggc caaacctttc ctcttcagga ggccacggca accgtagttt gacagaagag
49621 cagcaccttg atttaatgct tcccagcatg tgtccttgag caagtcacct aacctctctg
49681 ggctgcttcc tcattgggaa aatatggctg ccagtaaaac ctgccctgtc cacctcctgg
49741 ggcacttggc aaacagcaaa agagtccaaa tgtgcaggct gggccaggcg cagtggctca
49801 tgcctgtaat cccagcaatt taggaagcca aggtgggcgg atcacctgag gtcaggagtt
49861 tgagaccagc ctggccaaca tggtgaaacc ttgtctctac aaaaatacaa aaattagccg
49921 ggcatgatgg cgggtgcctg taatcccagt tactcgggag gctgaggcaa gagaatcgct
49981 tgaacccgga aggggaaggt tgcagtgagc caagattgtg ccactgcact ccagcctggg
50041 caacagagcg agactctgtc tcaaaaaaaa aaaaaaaaaa aaacaatgca gagctggctg
50101 tgtaaaaaac ctgttccact gcagggccca gtgtccacca ggctggggtg caggcctatg
50161 gggtgggggc ccagcatcag cctctcagca gccctgggag gcggggcgca tcccgtgccc
50221 ctcgtggtct ggatgtgttc tagcccaagt cctaggttac acctgccgtc gcctggcctc
50281 tcaggagagg cccagggtga ggaggagcat ggtaaaggtg aagctgattg ggaagtcagc
50341 tgttgggaaa gcaactcctt gcacattgga ggaaccgaga aagactgacc ccgaggacag
50401 cagccagcat ggcccttcct gggagcccat gttgggggat tcctgctgca gccaaggctc
50461 agcccttgtg gtcgcaggtg ctggtcctgg cctcttcccc tcccatgcag gagcacagga
50521 gagatggctt ctgaggacct gttgcagctg tggccctggg aatagatttg ccagggagct
50581 ttaaagcagc tgagtgtgtc atccagctaa gcctggggaa ggagcttggc tcaggtcctg
50641 acaggtgtga cagggatggg gactgggaag taagagatga aaccctggct ggaggctgtg
50701 agcttccaca gccagcgctg gacagggagg gtccagatat acccactagt gccctcacca
```

https://www.youtube.com/watch?v=mDi-crPgjFw&t=3s

How big is 40 exabytes?

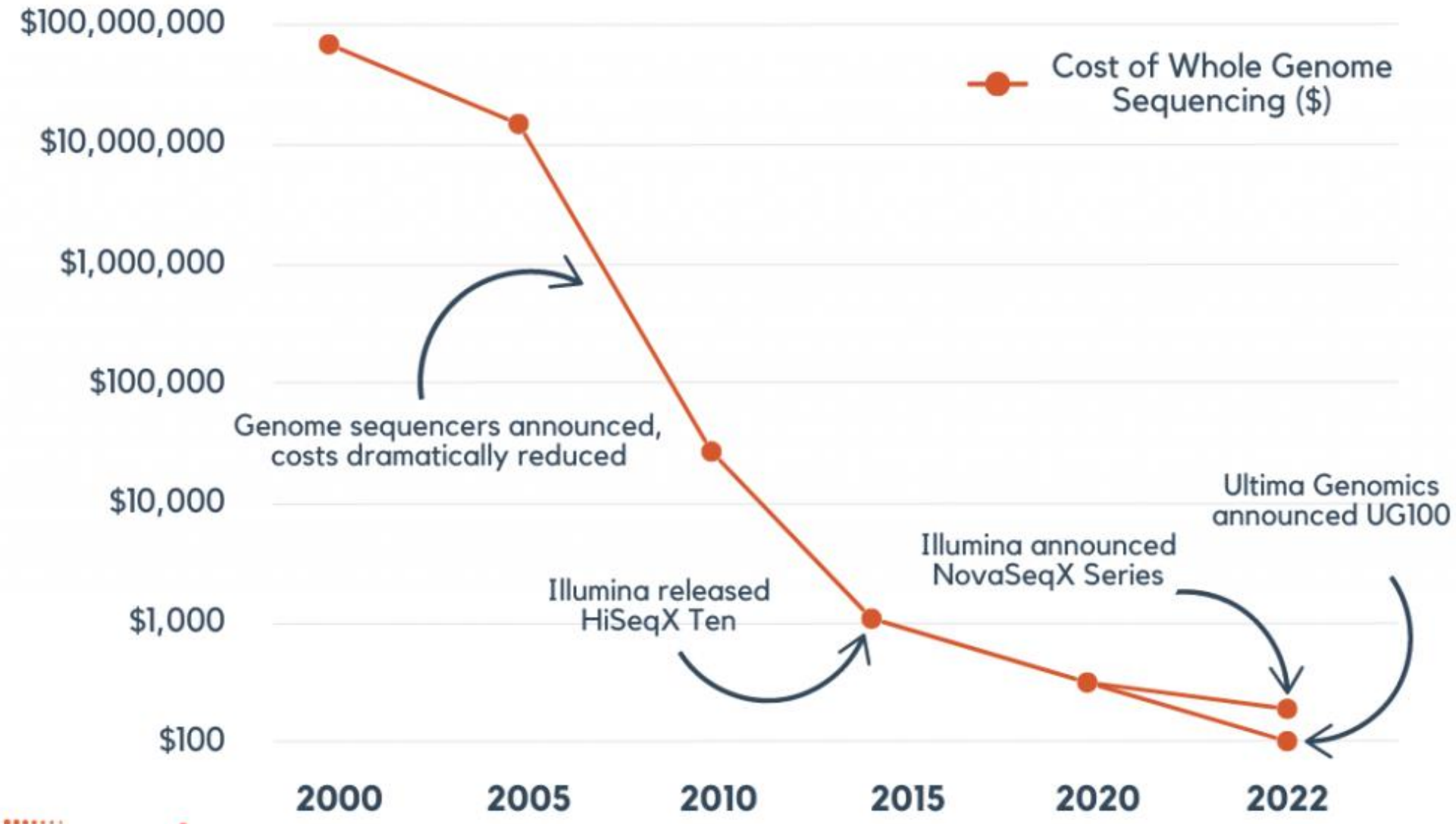
Genomics projects will generate 40 exabytes of data in the next decade.

Each shark = 100,000,000 GB of data



https://www.genome.gov/about-genomics/fact-sheets/Genomic-Data-Science

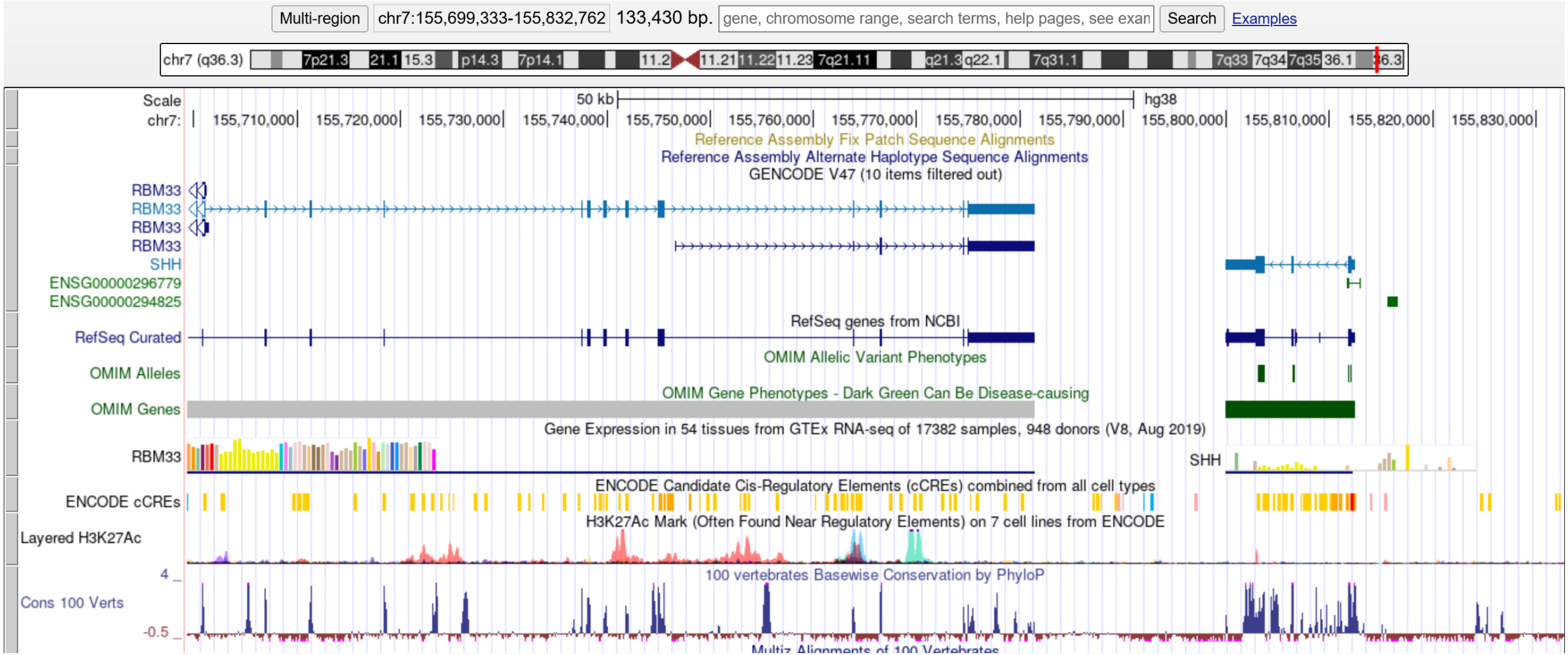
Decreasing Genome Sequencing Costs



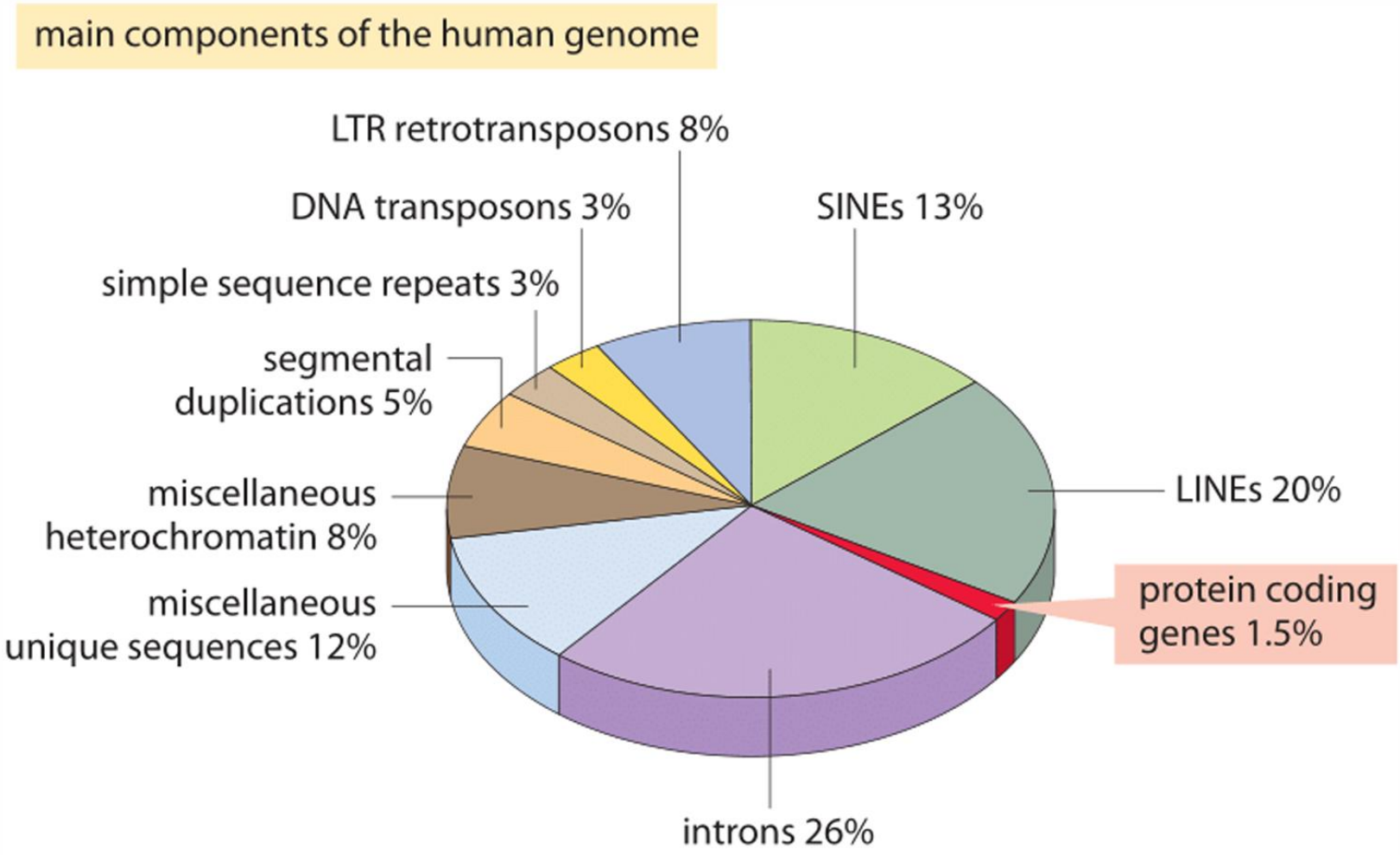
genomize

Source: National Human Genome Research Institute

Gene Annotation



Gene Annotation



» How many genes are in a genome? (bionumbers.org)

Understanding Gene Annotation through GENCODE (https://www.gencodegenes.org/)

GENCODE

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

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

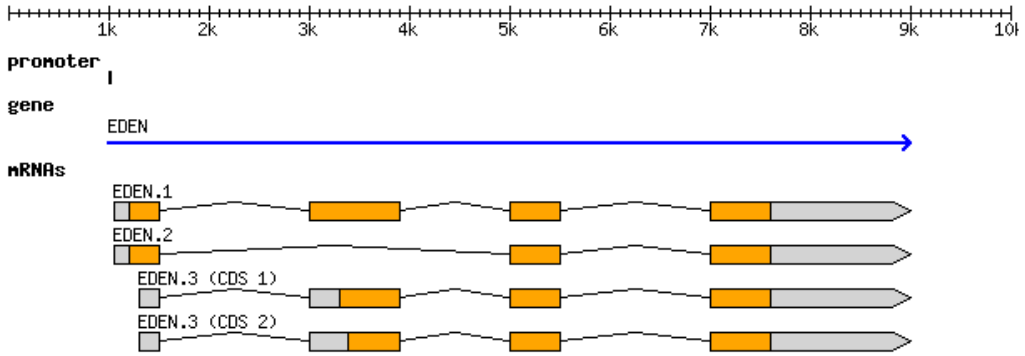
GTF / GFF3 files

Content	Regions	Description	Download
Comprehensive gene annotation	CHR	• It contains the comprehensive gene annotation on the reference chromosomes only	GTF GFF3
Comprehensive gene annotation	ALL	• It contains the comprehensive gene annotation on the reference chromosomes, scaffolds, assembly patches and alternate loci (haplotypes)	GTF GFF3
Comprehensive gene annotation	PRI	• It contains the comprehensive gene annotation on the primary assembly (chromosomes and scaffolds) sequence regions	GTF GFF3
Basic gene annotation	CHR	• It contains the basic gene annotation on the reference chromosomes only • This is a **subset** of the corresponding comprehensive annotation, including	GTF GFF3

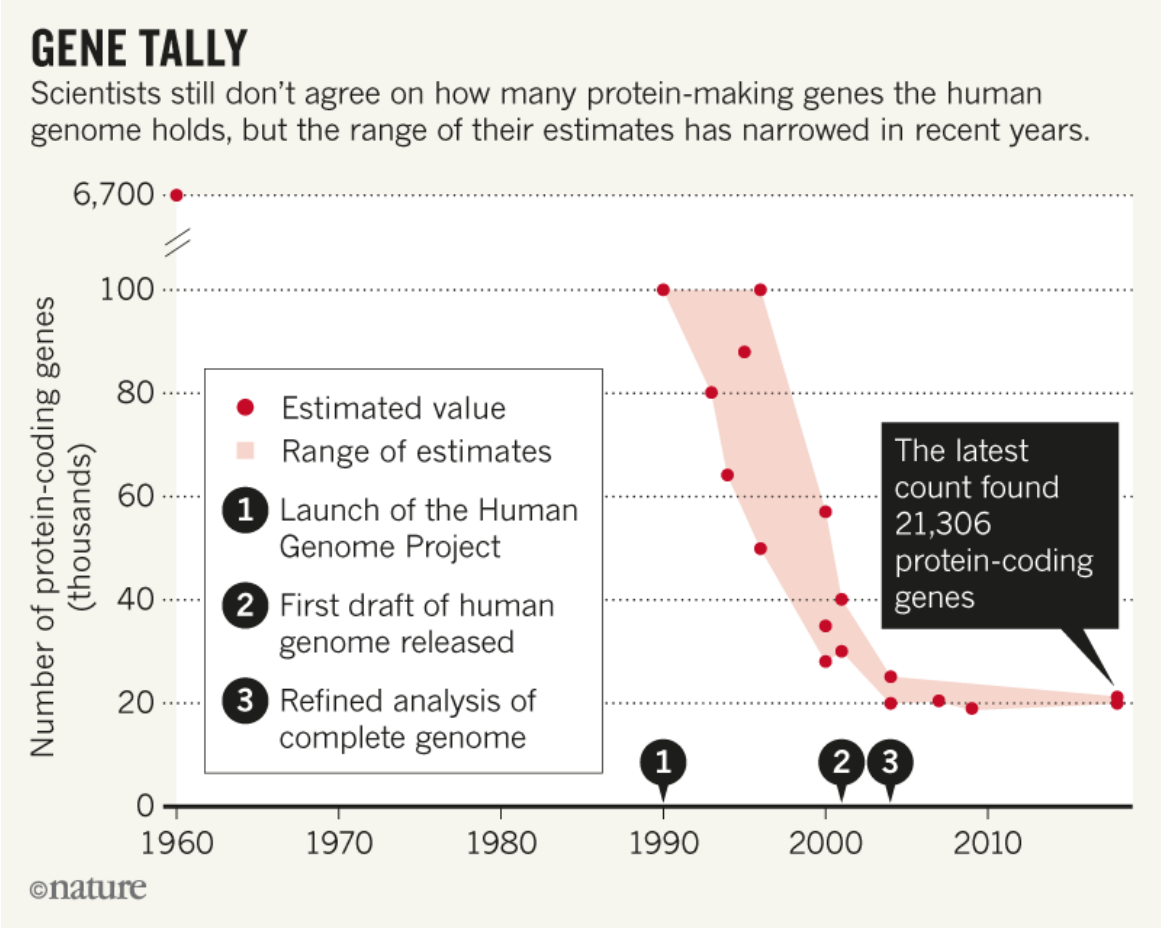
More about GENCODE Human

- Current human data
- Release history
- Statistics
- Data format
- FTP site

GTF (Gene Tranfer Format) and The number of Human Genes



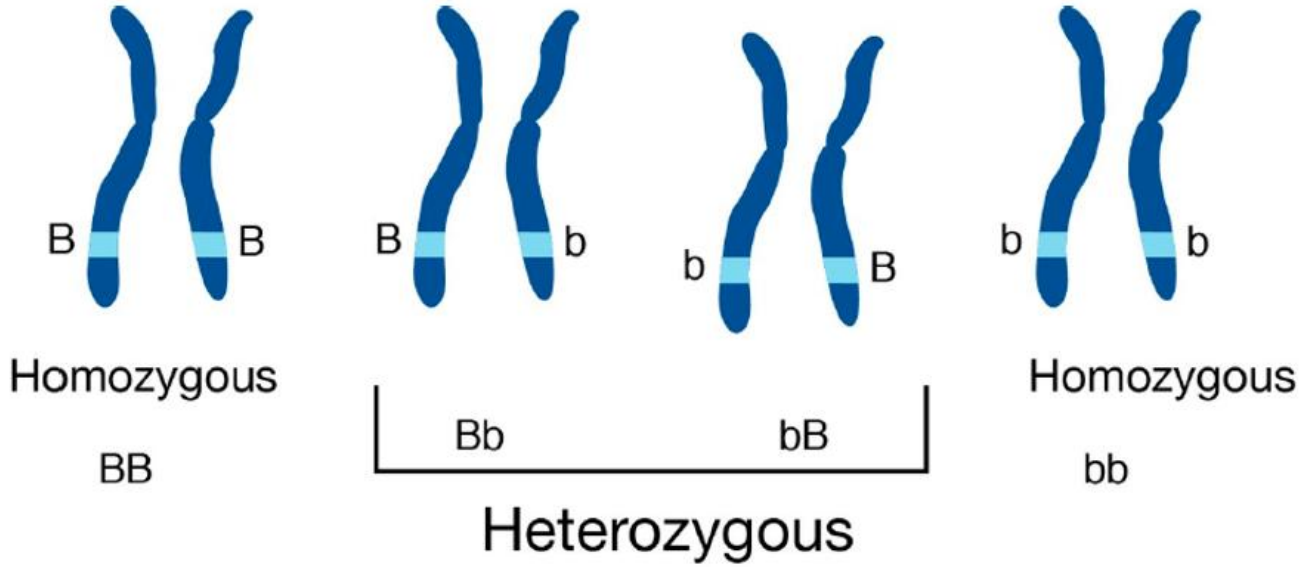

```
##gff-version 3.2.1
##sequence-region ctg123 1 1497228
ctg123 . gene            1000  9000  .  +  .  ID=gene00001;Name=EDEN
ctg123 . TF_binding_site 1000  1012  .  +  .  ID=tfbs00001;Parent=gene00001
ctg123 . mRNA            1050  9000  .  +  .  ID=mRNA00001;Parent=gene00001;Name=EDEN.1
ctg123 . mRNA            1050  9000  .  +  .  ID=mRNA00002;Parent=gene00001;Name=EDEN.2
ctg123 . mRNA            1300  9000  .  +  .  ID=mRNA00003;Parent=gene00001;Name=EDEN.3
ctg123 . exon            1300  1500  .  +  .  ID=exon00001;Parent=mRNA00003
ctg123 . exon            1050  1500  .  +  .  ID=exon00002;Parent=mRNA00001,mRNA00002
ctg123 . exon            3000  3902  .  +  .  ID=exon00003;Parent=mRNA00001,mRNA00003
ctg123 . exon            5000  5500  .  +  .  ID=exon00004;Parent=mRNA00001,mRNA00002,mRNA00003
ctg123 . exon            7000  9000  .  +  .  ID=exon00005;Parent=mRNA00001,mRNA00002,mRNA00003
ctg123 . CDS             1201  1500  .  +  0  ID=cds00001;Parent=mRNA00001;Name=edenprotein.1
ctg123 . CDS             3000  3902  .  +  0  ID=cds00001;Parent=mRNA00001;Name=edenprotein.1
ctg123 . CDS             5000  5500  .  +  0  ID=cds00001;Parent=mRNA00001;Name=edenprotein.1
ctg123 . CDS             7000  7600  .  +  0  ID=cds00001;Parent=mRNA00001;Name=edenprotein.1
ctg123 . CDS             1201  1500  .  +  0  ID=cds00002;Parent=mRNA00002;Name=edenprotein.2
ctg123 . CDS             5000  5500  .  +  0  ID=cds00002;Parent=mRNA00002;Name=edenprotein.2
ctg123 . CDS             7000  7600  .  +  0  ID=cds00002;Parent=mRNA00002;Name=edenprotein.2
ctg123 . CDS             3301  3902  .  +  0  ID=cds00003;Parent=mRNA00003;Name=edenprotein.3
ctg123 . CDS             5000  5500  .  +  1  ID=cds00003;Parent=mRNA00003;Name=edenprotein.3
ctg123 . CDS             7000  7600  .  +  1  ID=cds00003;Parent=mRNA00003;Name=edenprotein.3
ctg123 . CDS             3391  3902  .  +  0  ID=cds00004;Parent=mRNA00003;Name=edenprotein.4
ctg123 . CDS             5000  5500  .  +  1  ID=cds00004;Parent=mRNA00003;Name=edenprotein.4
ctg123 . CDS             7000  7600  .  +  1  ID=cds00004;Parent=mRNA00003;Name=edenprotein.4
```

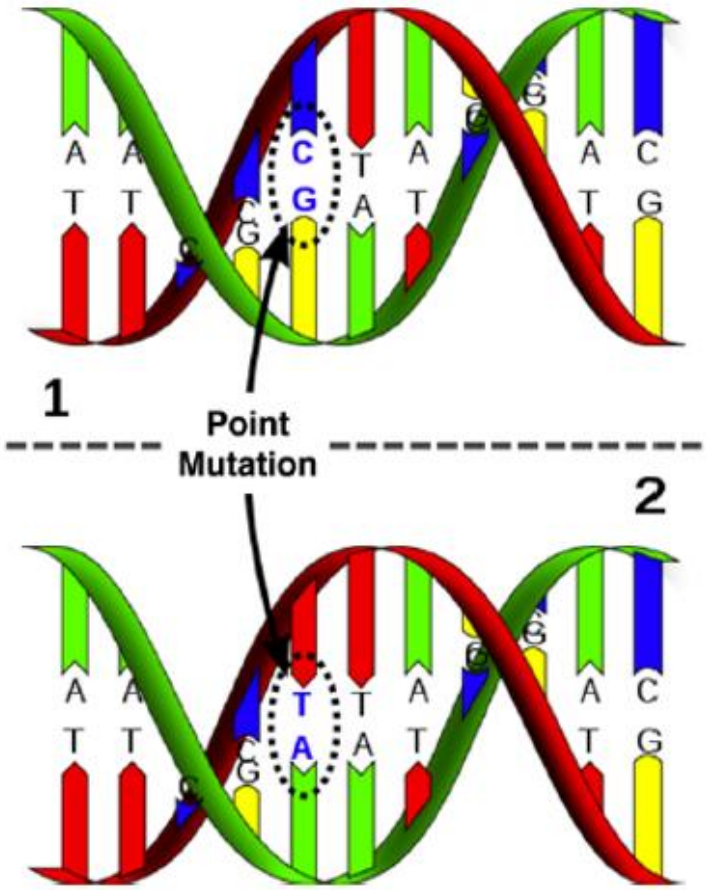


https://www.nature.com/articles/d41586-018-05462-w

Các loại biến thể trên hệ gen

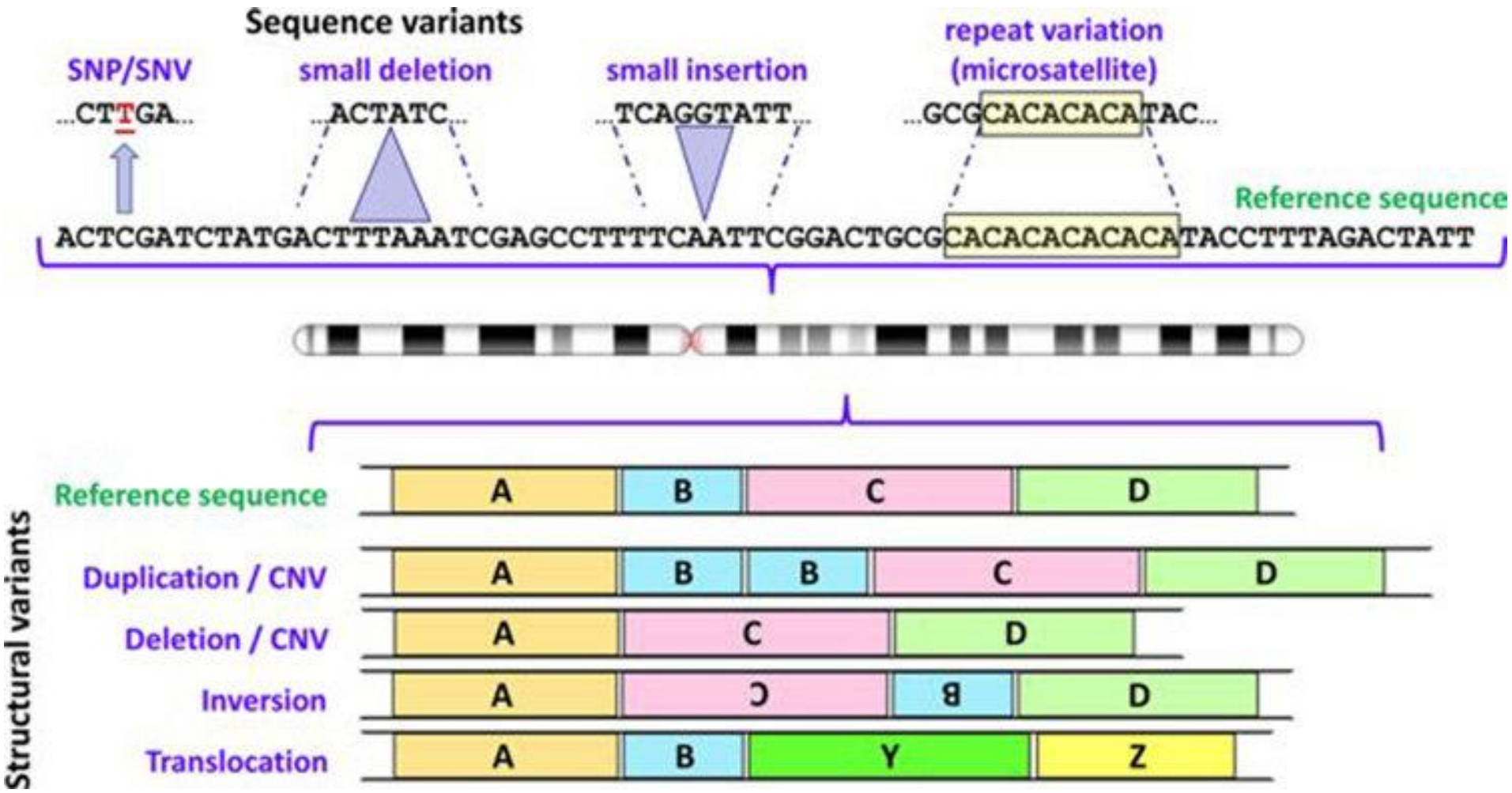
- Hệ gen giữa hai người giống nhau > 99%
- Mỗi người có khoảng 5 triệu biến thể, trong đó có 3 đến 4 triệu biến thể một nucleotide
- Hệ gen người là hệ lưỡng bội





https://rosalind.info/media/point_mutation.png, https://en.wikipedia.org/wiki/Zygosity

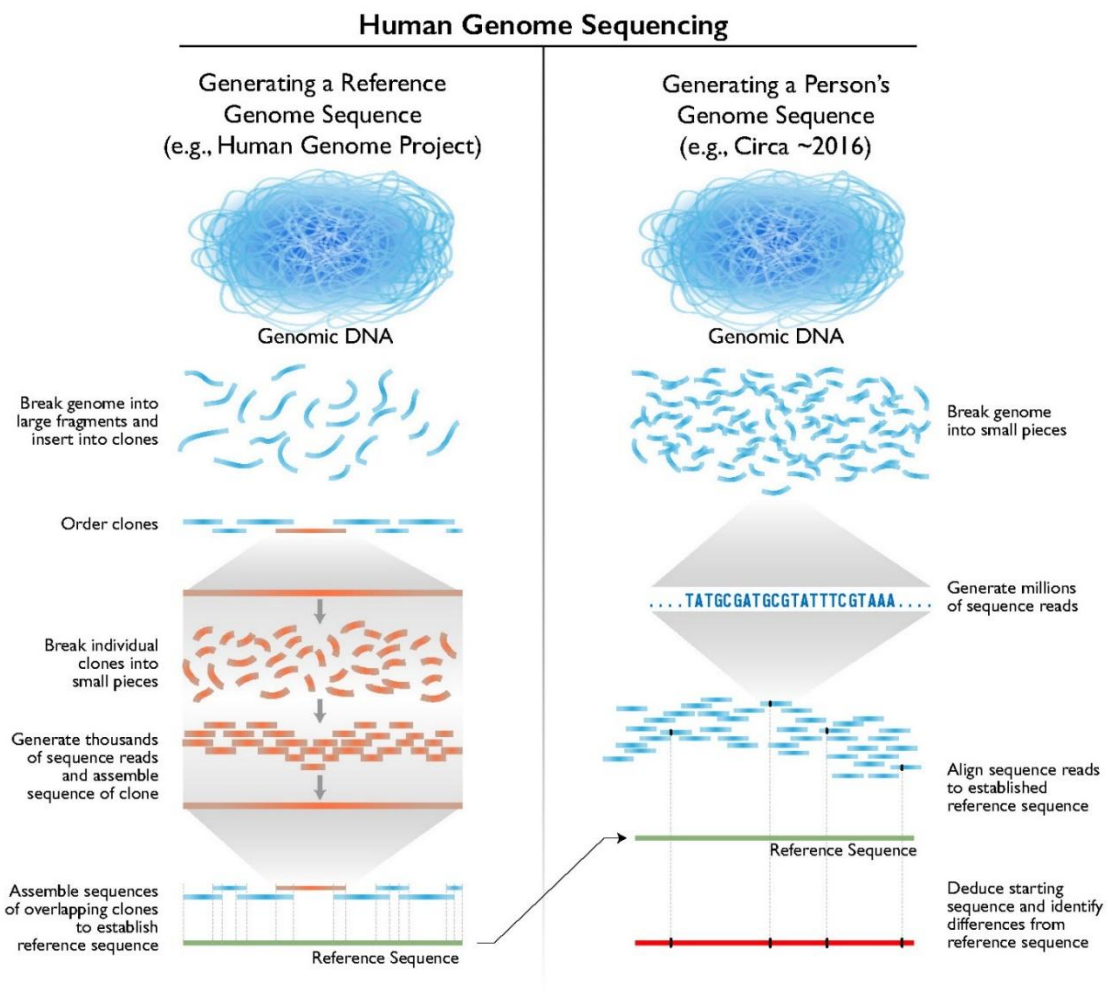
Các loại biến thể trên hệ gen



10.1042/EBC20170053

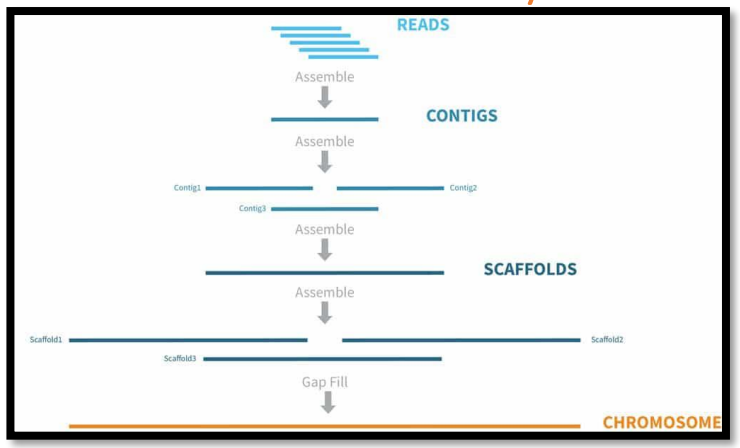
Giới thiệu về
GIẢI TRÌNH TỰ GEN THẾ HỆ MỚI

Giải trình tự gen thế hệ mới (NGS)

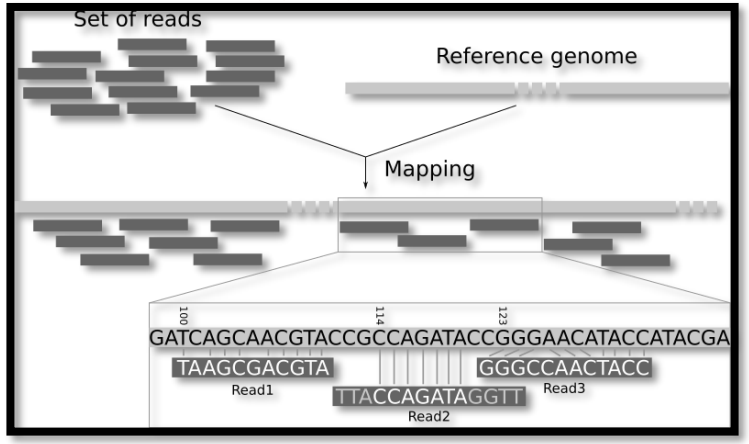


https://www.genome.gov/about-genomics/fact-sheets/Sequencing-Human-Genome-cost

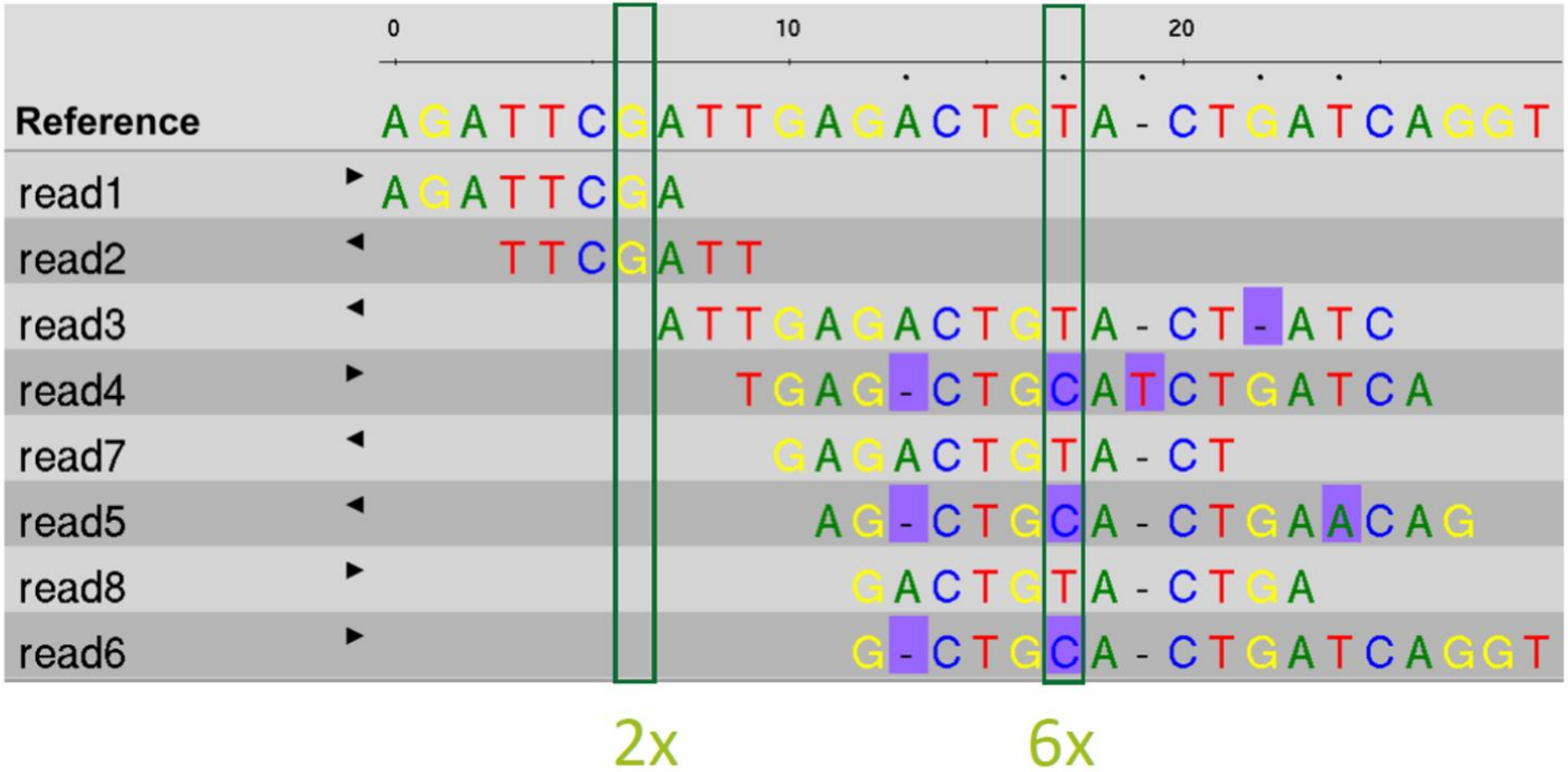
De novo assembly



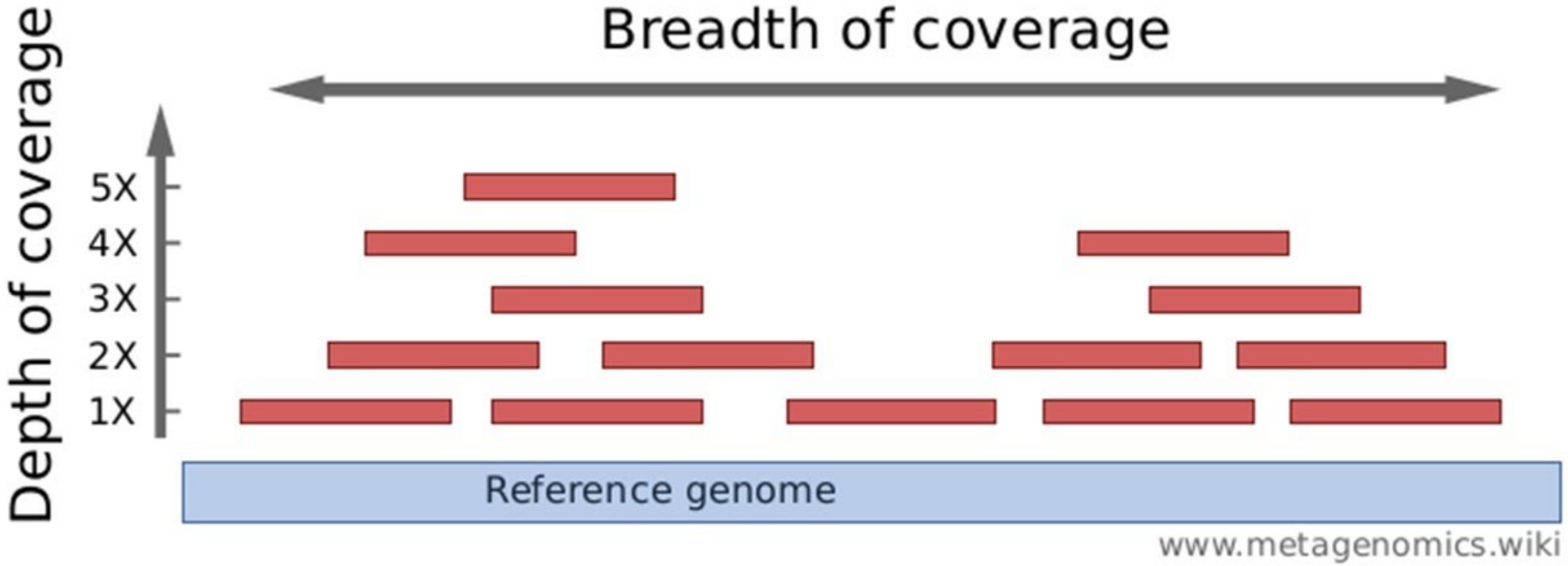
Mapping to reference



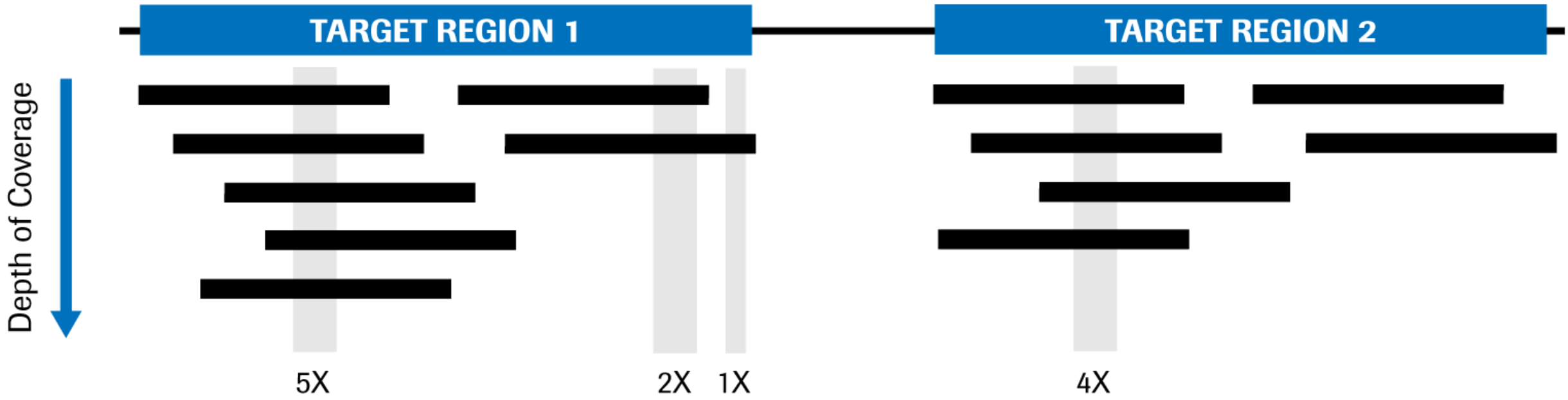
Kết quả của mapping



Depth vs Breadth of Coverage



Depth of Coverage



https://sequencing.roche.com/us/en/article-listing/metrics-for-target-enriched-ngs.html

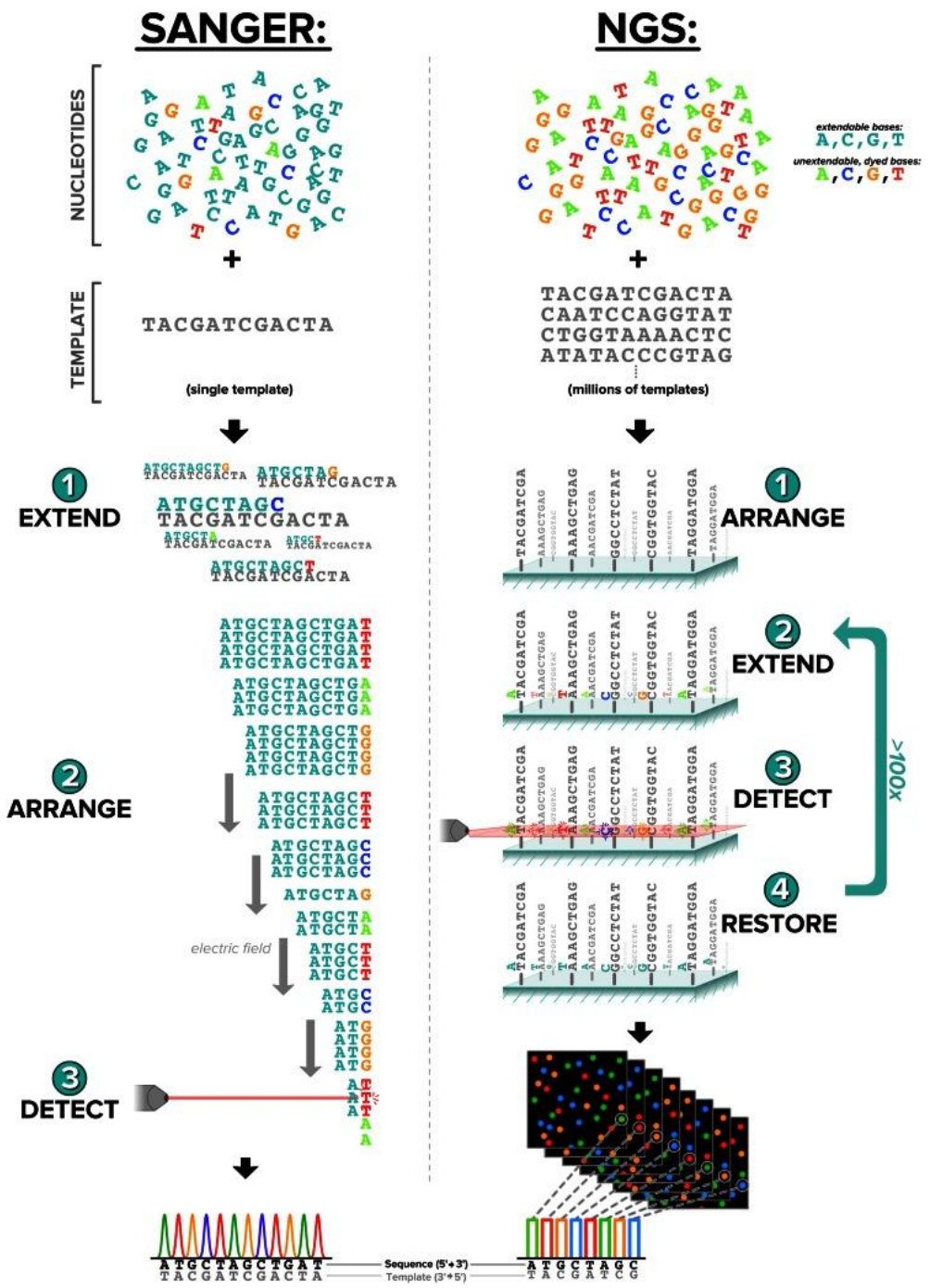
Giải trình tự gen thế hệ mới (NGS): Có hệ gen tham chiếu

De novo assembly

Mapping to reference

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

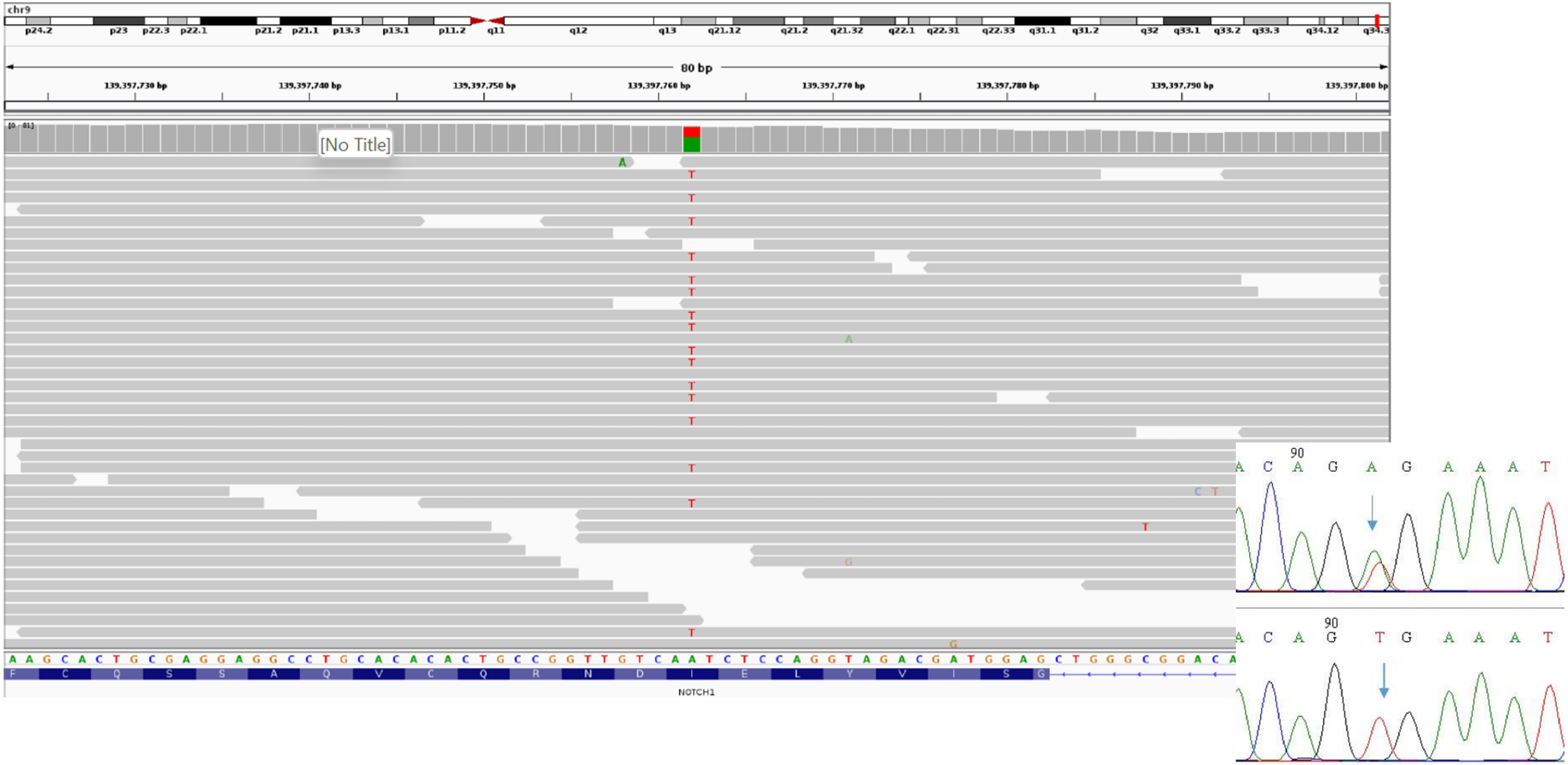
Read1:CTCGAATACG



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

10.1007/s40142-015-0076-8

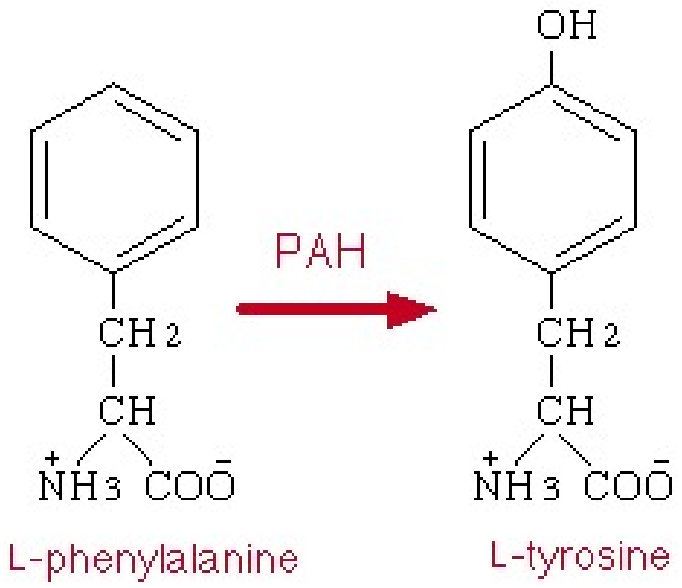
NGS vs Sanger



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

Phenylketonuria (PKU): Rối loạn chuyển hóa di truyền

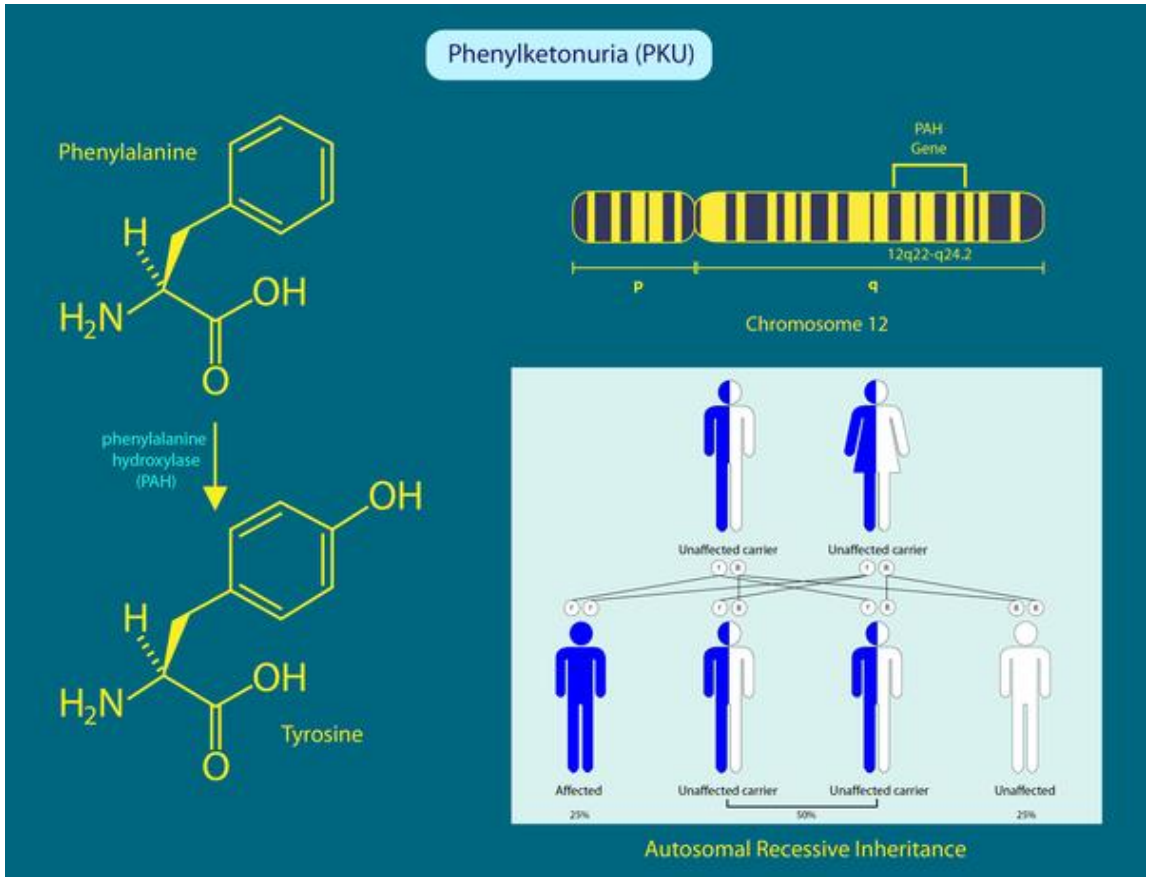
- Nguyên nhân do thiếu hụt enzyme phenylalanine hydroxylase.
- Sự mất enzyme này dẫn đến suy giảm trí tuệ, tổn thương cơ quan, tư thế bất thường.
- Tần suất xảy ra PKU khác nhau giữa các nhóm dân tộc và các vùng địa lý trên toàn thế giới. Ở Hoa Kỳ, PKU xảy ra ở 1 trong 25.000 trẻ sơ sinh.
- Hầu hết các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



The enzyme phenylalanine hydroxylase converts the amino acid phenylalanine to tyrosine.

https://www.ncbi.nlm.nih.gov/books/NBK22253/

Phenylketonuria (PKU): Rối loạn chuyển hóa di truyền



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

Tên khác của PKU

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

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

Gene (Nucleotide)

NT seq — 1359 nt [NT seq]

```
atgtccactgcggtcctggaaaacccaggcttgggcaggaaactctctgactttggacag
gaaacaagctatattgaagacaactgcaatcaaaatggtgccatatcactgatcttctca
ctcaaagaagaagttggtgcattggccaaagtattgcgcttatttgaggagaatgatgta
aacctgacccacattgaatctagaccttctcgtttaaagaaagatgagtatgaatttttc
acccatttggataaacgtagcctgcctgctctgacaaacatcatcaagatcttgaggcat
gacattggtgccactgtccatgagctttcacgagataagaagaaagacacagtgccctgg
ttcccaagaaccattcaagagctggacagatttgccaatcagattctcagctatggagcg
gaactggatgctgaccaccctggttttaaagatcctgtgtaccgtgcaagacggaagcag
tttgctgacattgcctacaactaccgccatgggcagcccatccctcgagtggaatacatg
gaggaagaaaagaaaacatggggcacagtgttcaagactctgaagtccttgtataaaacc
catgcttgctatgagtacaatcacatttttccacttcttgaaaagtactgtggcttccat
gaagataacattccccagctggaagacgtttctcagttcctgcagacttgcactggtttc
cgcctccgacctgtggctggcctgctttcctctcgggatttcttgggtggcctggccttc
cgagtcttccactgcacacagtacatcagacatggatccaagcccatgtatacccccgaa
cctgacatctgccatgagctgttgggacatgtgcccttgttttcagatcgcagctttgcc
cagttttcccaggaaattggccttgcctctctgggtgcacctgatgaatacattgaaaag
ctcgccacaatttactggtttactgtggagtttgggctctgcaaacaaggagactccata
aaggcatatggtgctgggctcctgtcatcctttggtgaattacagtactgcttatcagag
aagccaaagcttctccccctggagctggagaagacagccatccaaaattacactgtcacg
gagttccagcccctctattacgtggcagagagttttaatgatgccaaggagaaagtaagg
aactttgctgccacaatacctcggcccttctcagttcgctacgacccatacacccaaagg
attgaggtcttggacaatacccagcagcttaagattttggctgattccattaacagtgaa
attggaatcctttgcagtgccctccagaaaataaagtaa
```

Protein (Amino Acid)

AA seq — 452 aa [AA seq] [DB search]

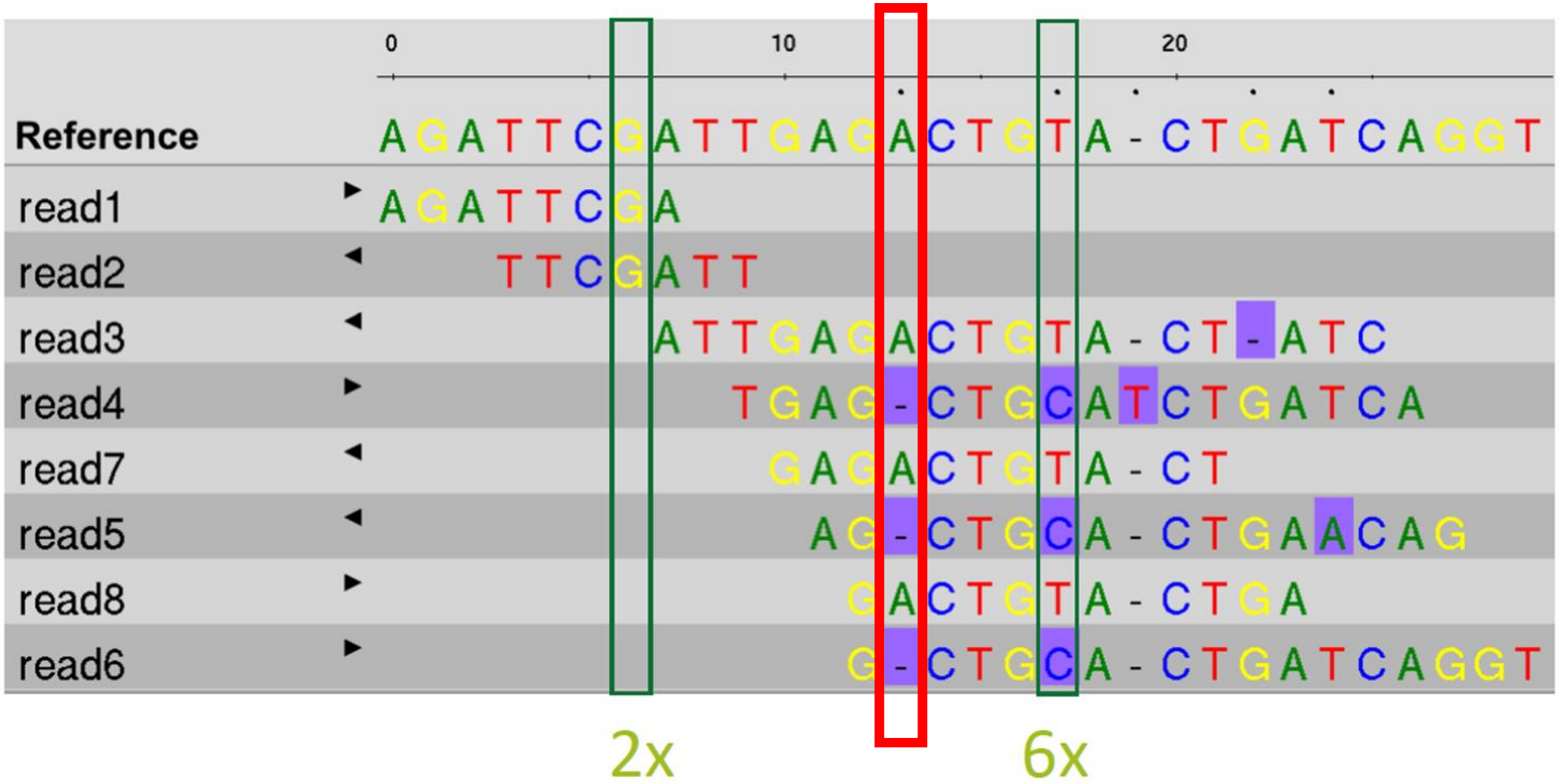
```
MSTAVLENPGLGRKLSDFGQETSYIEDNCNQNGAISLIFSLKEEVGALAKVLRLFEENDV
NLTHIESRPSRLKKDEYEFFTHLDKRSLPALTNIIKILRHDIGATVHELSRDKKKDTVPW
FPRTIQELDRFANQILSYGAELDADHPGFKDPVYRARRKQFADIAYNYRHGQPIPRVEYM
EEEKKTWGTVFKTLKSLYKTHACYEYNHIFPLLEKYCGFHEDNIPQLEDVSQFLQTCTGF
RLRPVAGLLSSRDFLGGLAFRVFHCTQYIRHGSKPMYTPEPDICHELLGHVPLFSDRSFA
QFSQEIGLASLGAPDEYIEKLATIYWFTVEFGLCKQGDSIKAYGAGLLSSFGELQYCLSE
KPKLLPLELEKTAIQNYTVTEFQPLYYVAESFNDAKEKVRNFAATIPRPFSVRYDPYTQR
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)

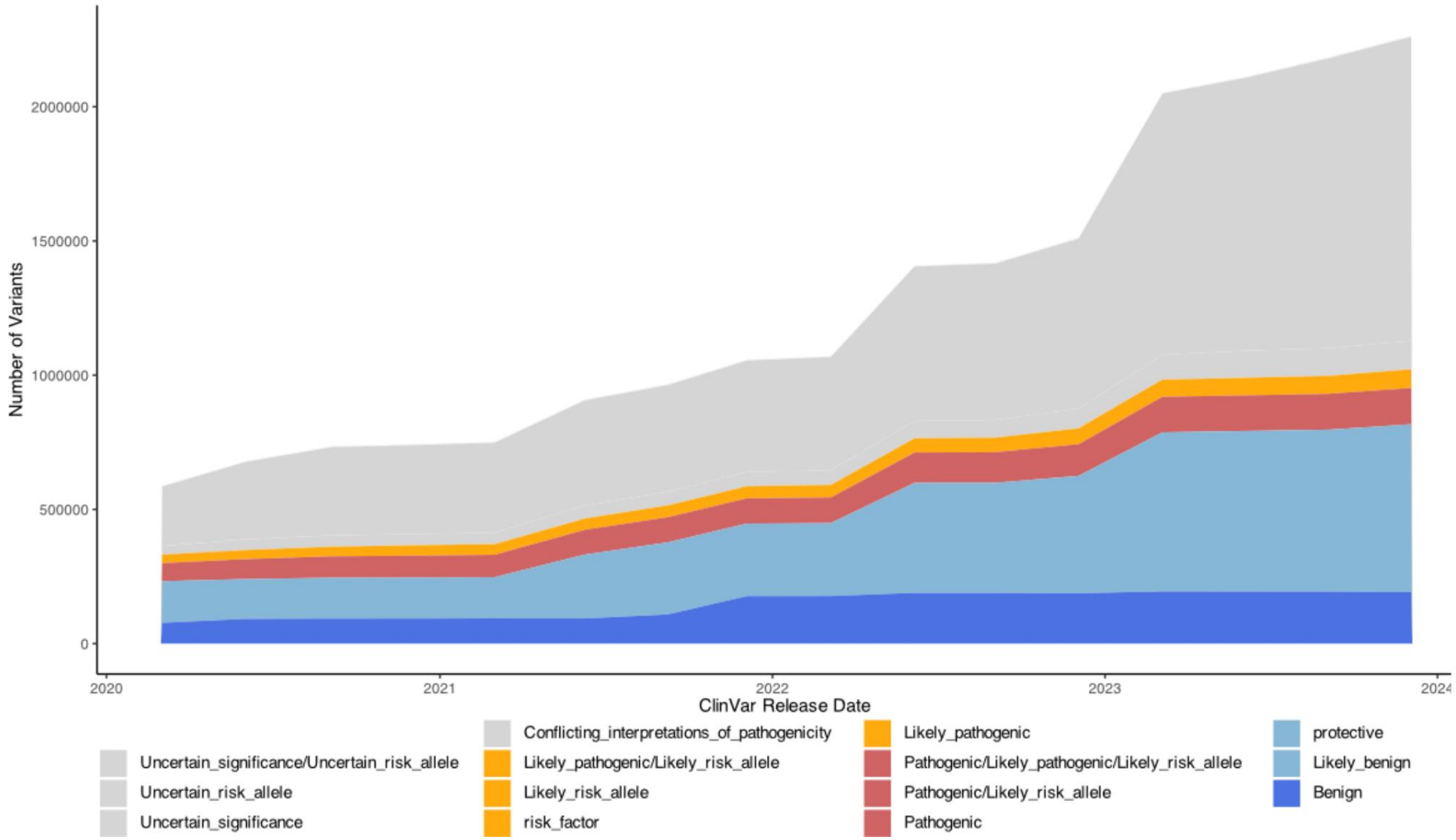
Links from Gene

Display options ▾ Sort by Relevance ▾ Download ▾ Items: 1 to 100 of 209 << First < Prev Page 1 of 3 Next > Last >>

Variation	Gene (Protein Change)	Type (Consequence)	Condition	Classification, Review status
NM_004316.4(ASCL1):c.51G>T (p.Gln17His)	ASCL1, PAH (Q17H)	Single nucleotide variant (missense variant +1 more)	not specified	G Uncertain significance ★
NC_000012.11:g.(?_103232953)_(103240749_?)del	PAH	Deletion	Phenylketonuria	G Pathogenic ★
NC_000012.11:g.(?_103288493)_(103310908_?)del	PAH	Deletion	Phenylketonuria	G Pathogenic ★
NC_000012.11:g.(?_103248894)_(103249131_?)del	PAH	Deletion	Phenylketonuria	G Pathogenic ★
NC_000012.12:g.(?_102894715)_(102894938_?)del	PAH	Deletion	Phenylketonuria	G Pathogenic ★
NC_000012.11:g.(?_103306549)_(103306696_?)del	PAH	Deletion	Phenylketonuria	G Pathogenic ★
NM_000277.3(PAH):c.1179_1180delI(p.Asn393fs)	PAH (N393fs)	Deletion (frameshift variant)	Phenylketonuria	G Likely pathogenic ★

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.

10.1186/s12967-024-05508-w

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

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 Dec 2023 by ClinGen PAH Va...

FDA RECOGNIZED DATABASE

for Phenylketonuria

Somatic

No data submitted for somatic clinical impact

Somatic

No data submitted for oncogenicity

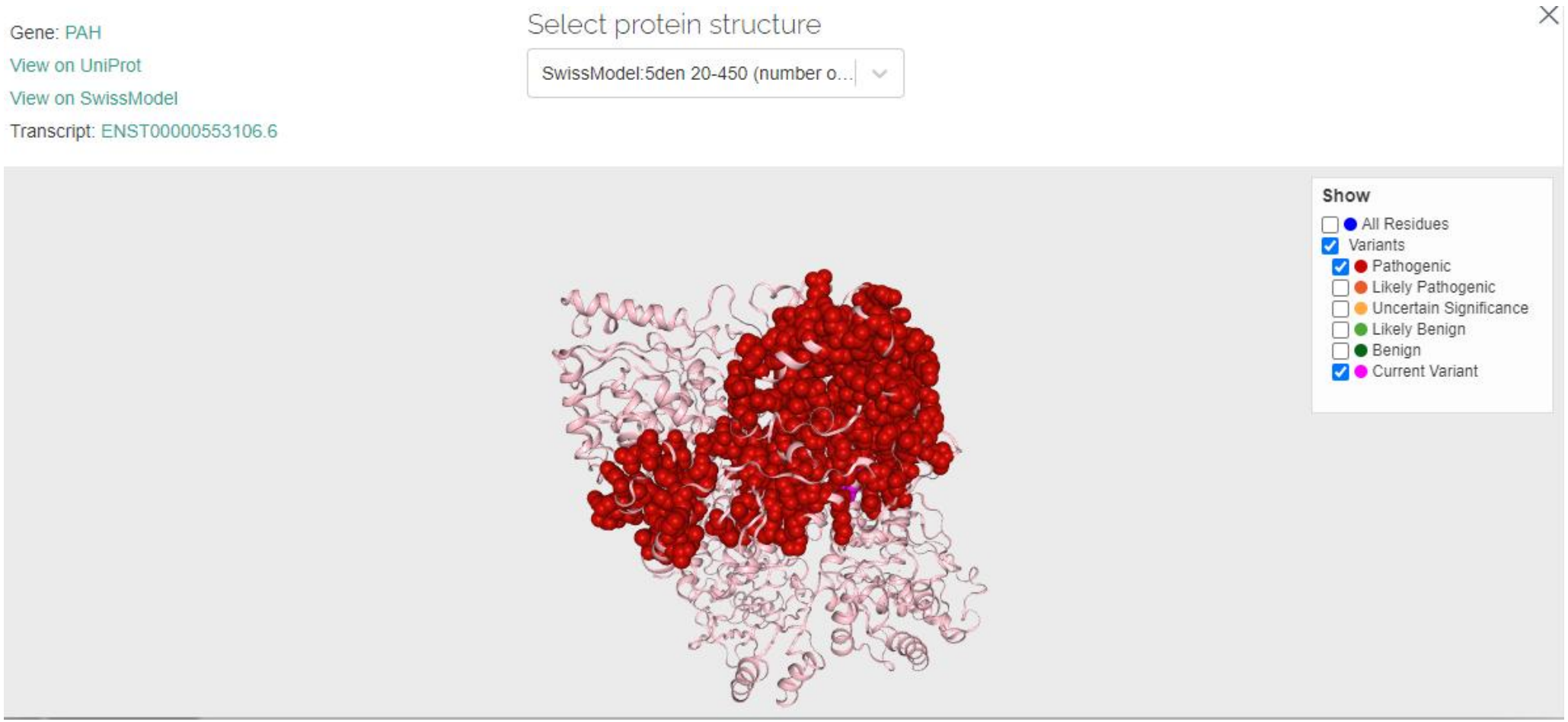
On this page

- Classification Summary
- Variant Details
- Genes
- Germline
  - Conditions
  - Submissions
  - Citations
- Text mined Citations

Feedback

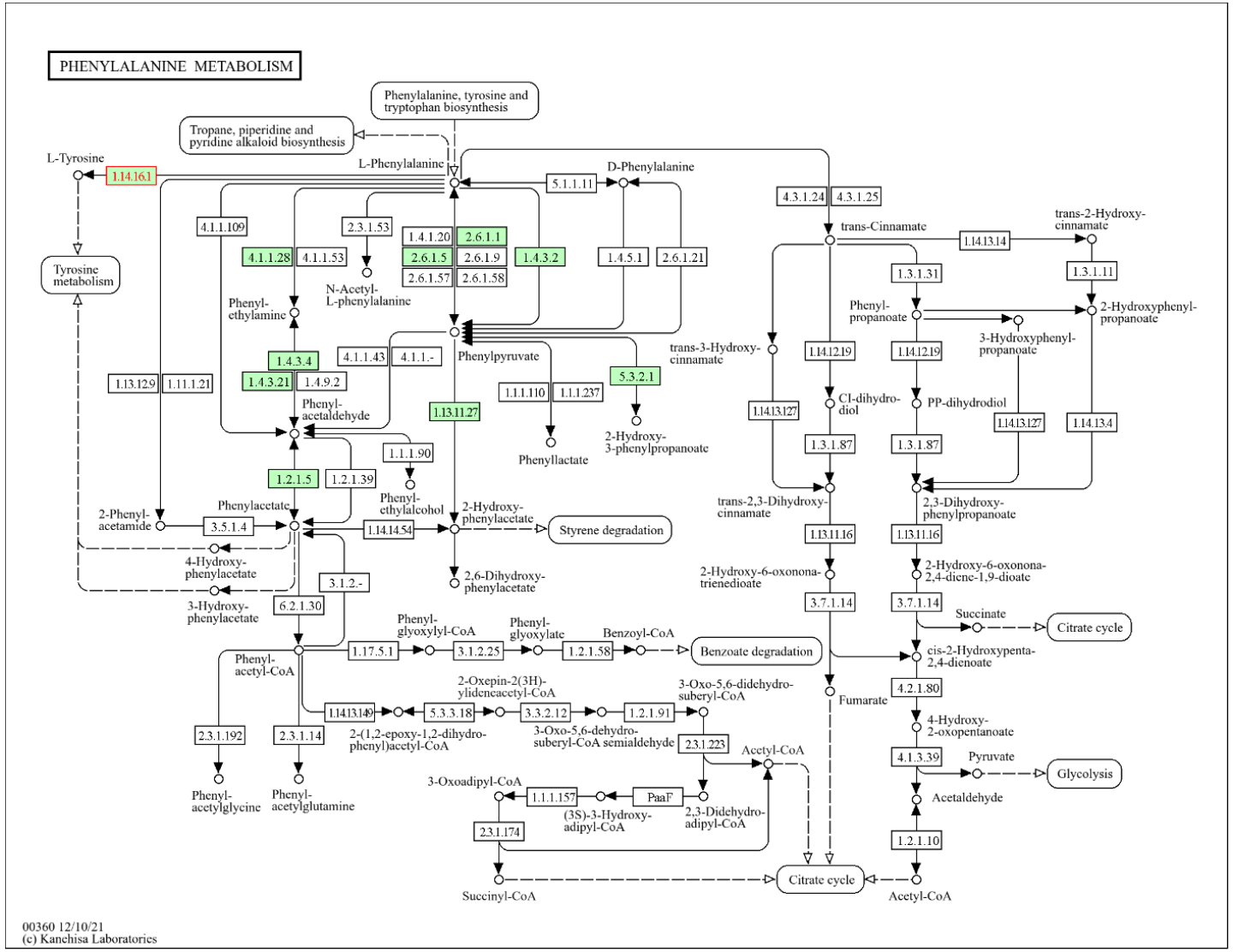
https://www.ncbi.nlm.nih.gov/clinvar/variation/2682170/

Các biến thể gây bệnh trên gen PAH



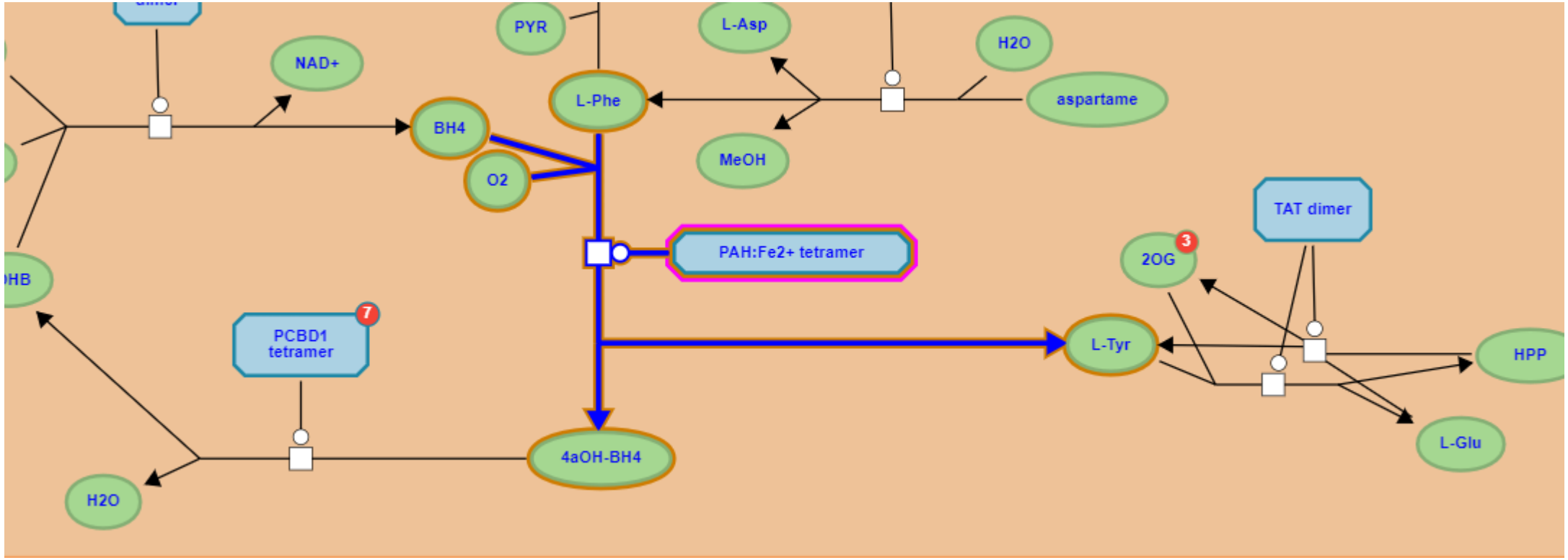
https://varsome.com/variant/hg38/chr12%3A102844430%3AA%3AT?

PAH: chuyển hóa Phenylalanine thành Tyrosine



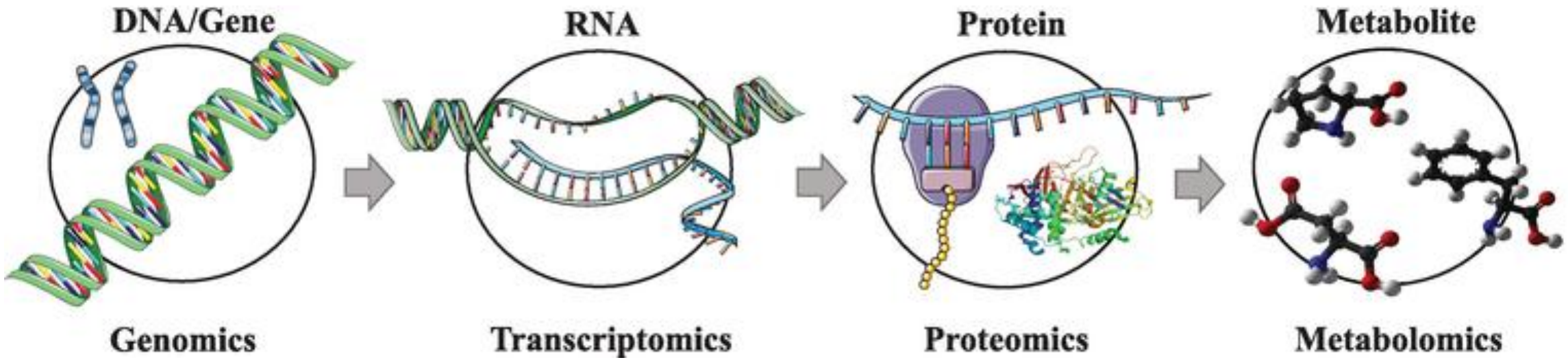
https://www.genome.jp/pathway/hsa00360+5053

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



PAH gene

Ref ...ATCGAT...

P1 ...AACGAT...

NM_000277.3(PAH):c.971T>A

PAH mRNA

Ref ...AUCGAU...

P1 ...AACGAU...

NM_000277.3(PAH):c.971T>A

PAH protein

Ref ...Ile-Asp...

P1 ...Asn-Asp...

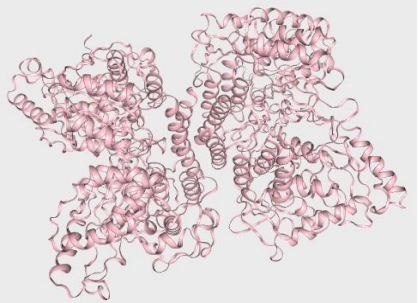
NM_000277.3(PAH):p.Ile324Asn

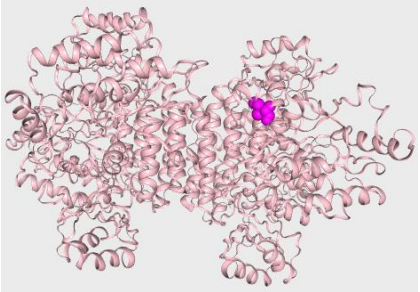
PAH

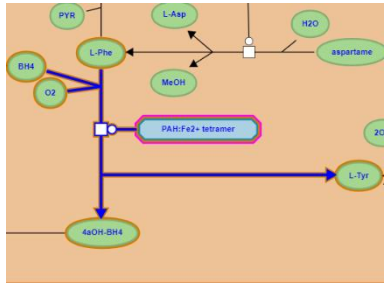
Ref Phe → Tyr

PAH

P1 Phe ~~→~~ Tyr

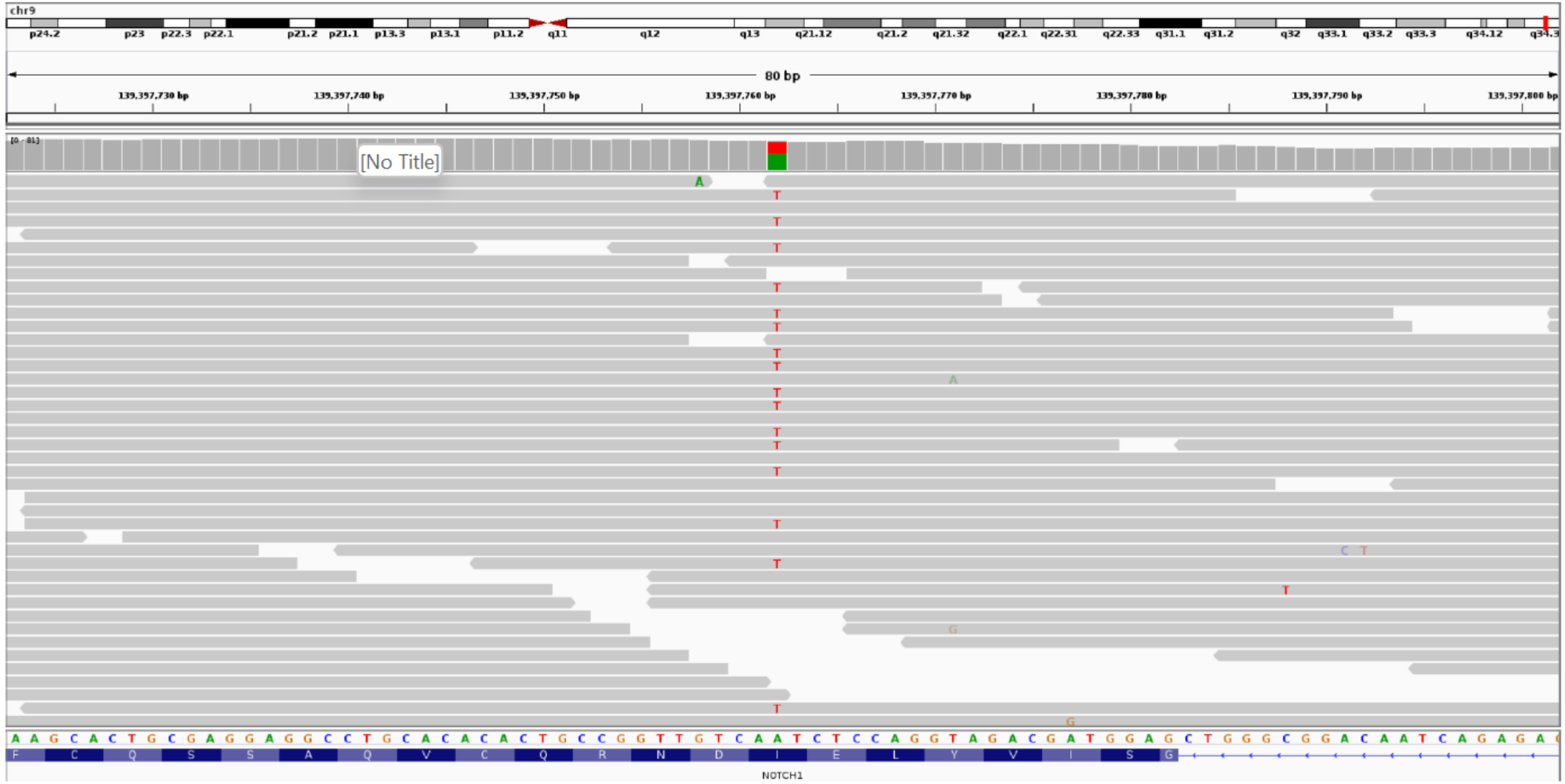




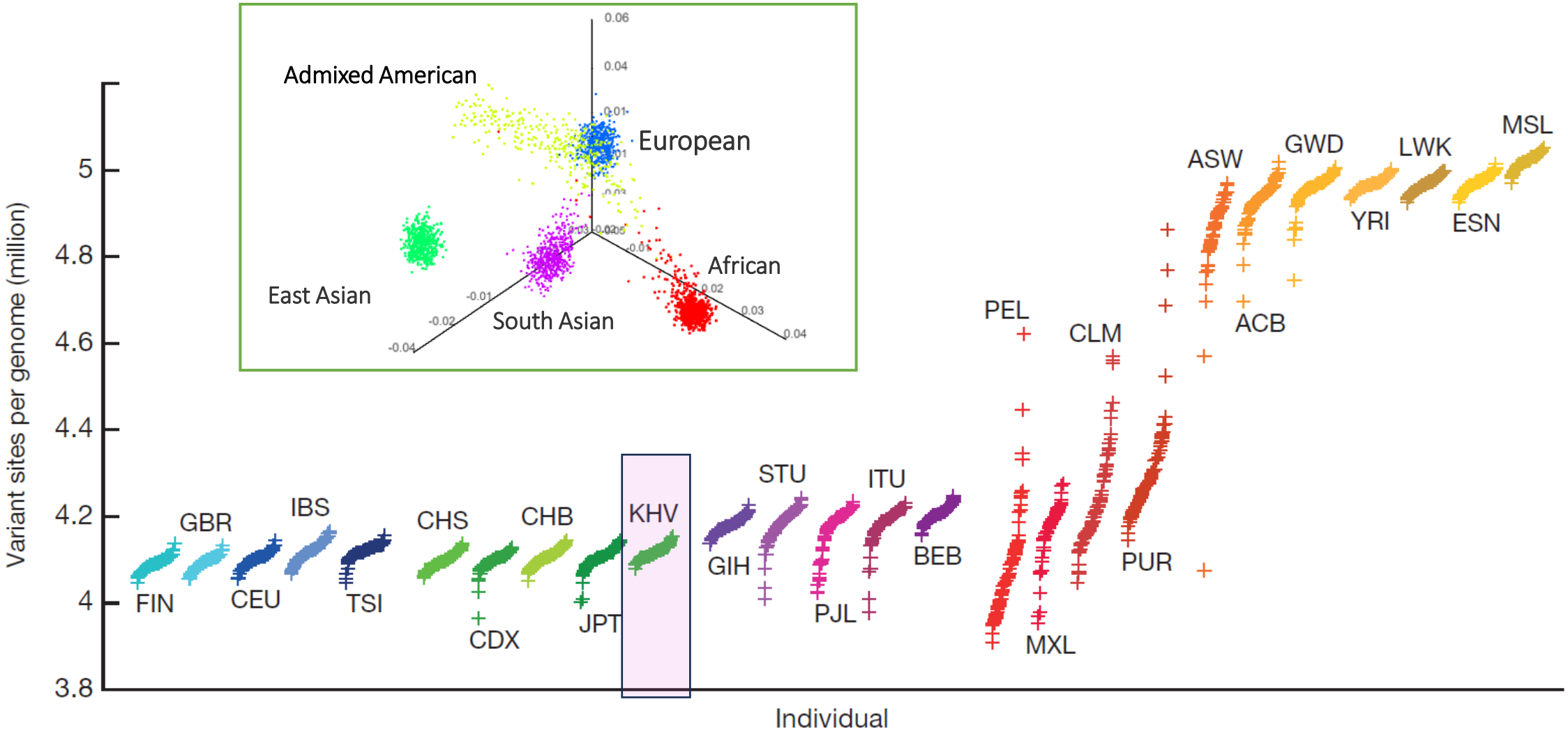


10.5772/intechopen.71769

Alignment and variant viewers



Human Genome Variation: 1000 Genomes Project

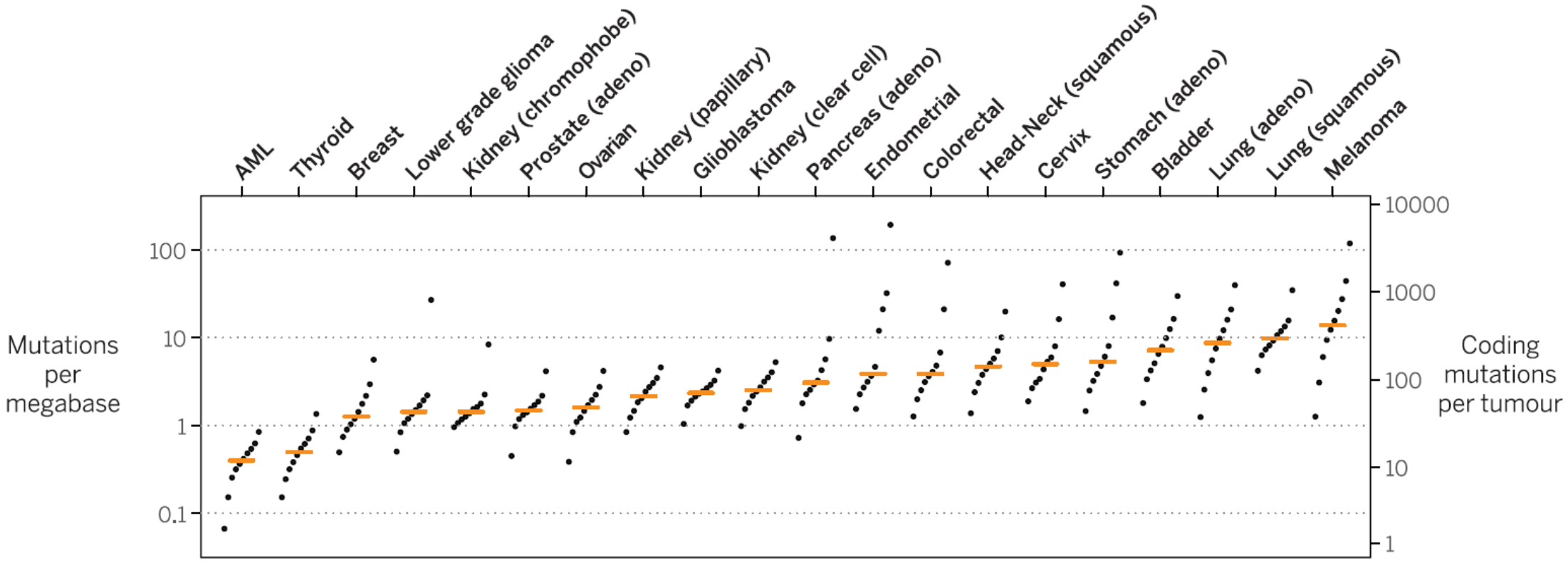


The number of variant sites per genome of 1K human genomes project (2015)
Kinh in Ho Chi Minh City, Vietnam (KHV)

doi:10.1038/nature15393

https://bwlewis.github.io/1000_genomes_examples/PCA_overview.html

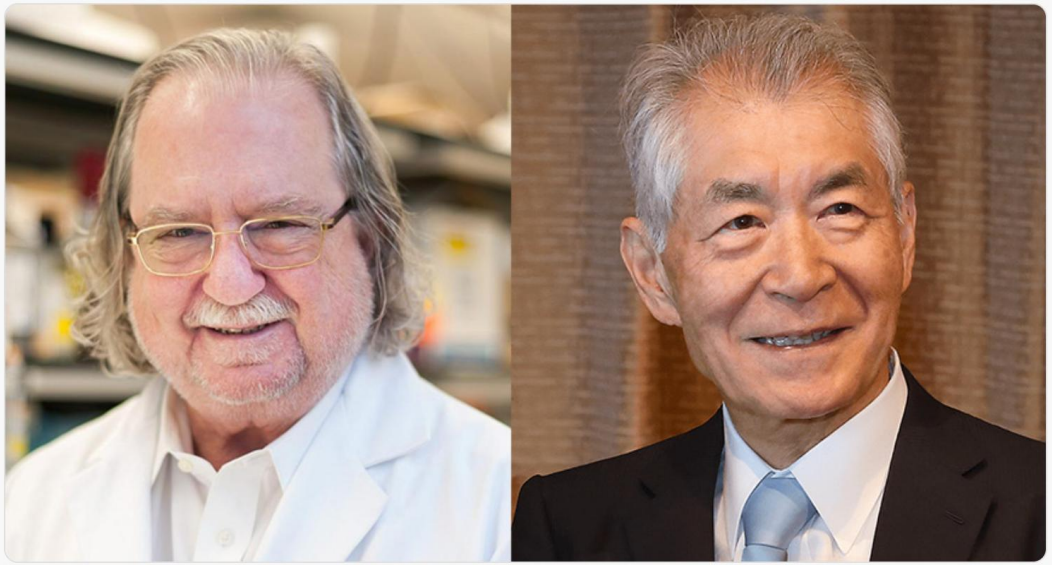
Cancer Genome Somatic Variation



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)

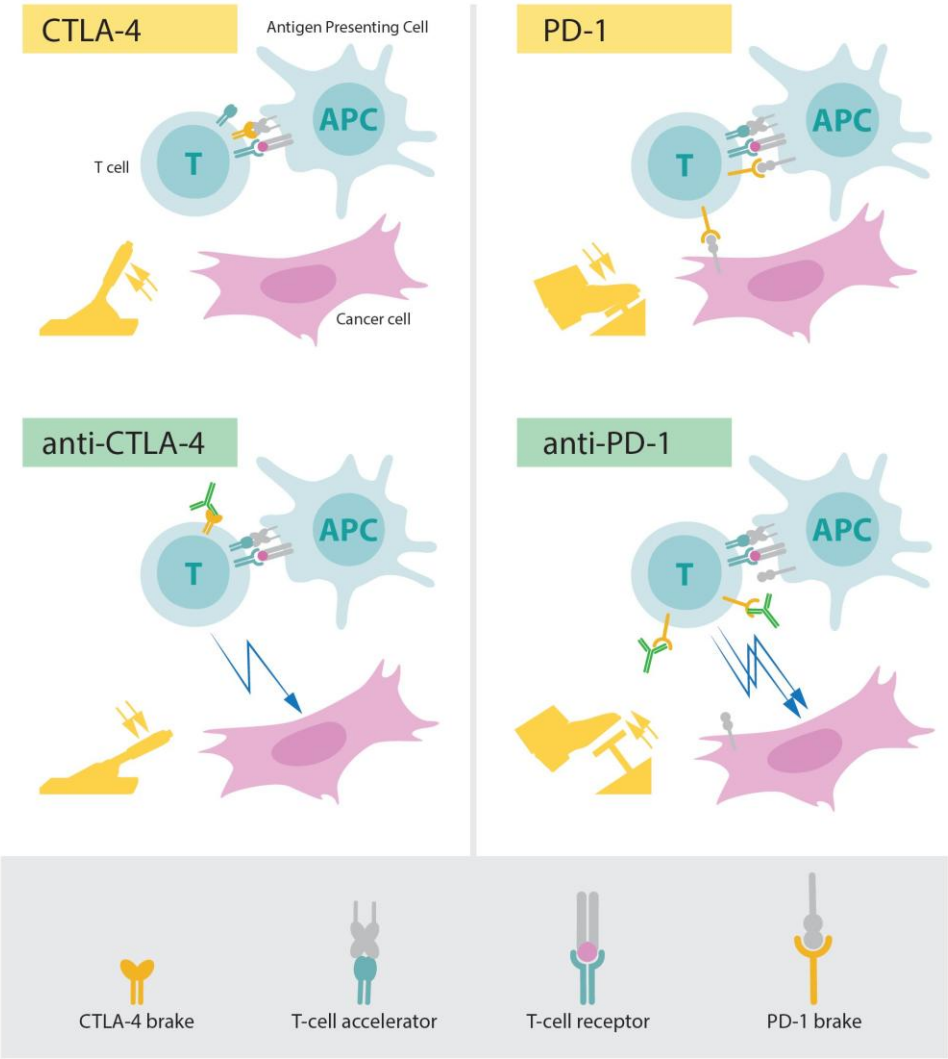
DOI: 10.1126/science.aab4082

Nobel Prize 2018 in Physiology or Medicine: **Cancer immunotherapy**



As a testament to the significance of immunotherapy for the treatment of cancer, the 2018 Nobel Prize in Physiology or Medicine was awarded to Drs. James P. Allison and Tasuku Honjo "for their discovery of cancer therapy by inhibition of negative immune regulation". Dr. Allison discovered that a molecule called CTLA-4 acts as a brake on T cells and Dr. Honjo discovered another T cell brake called PD-1.

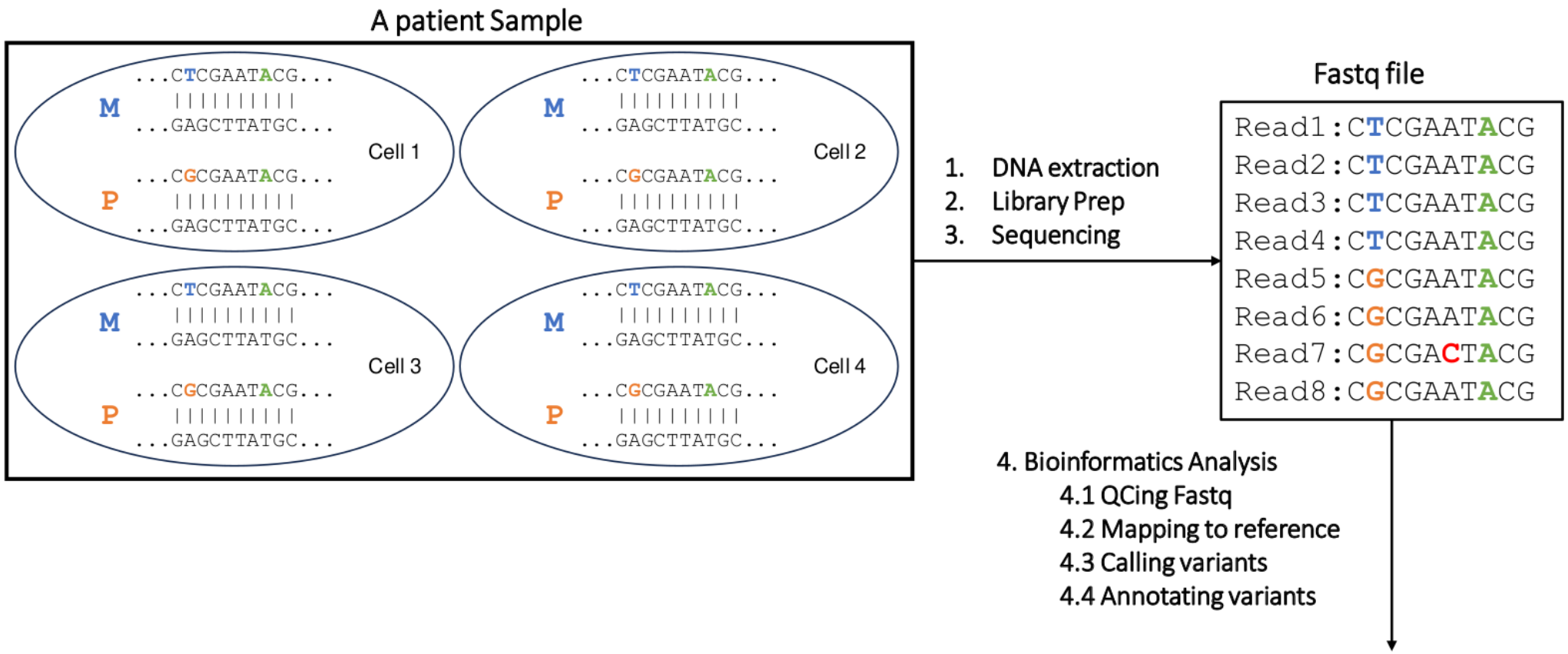
https://www.ucir.org/articles/2019-12-26-news_article_nobel_prize_allison



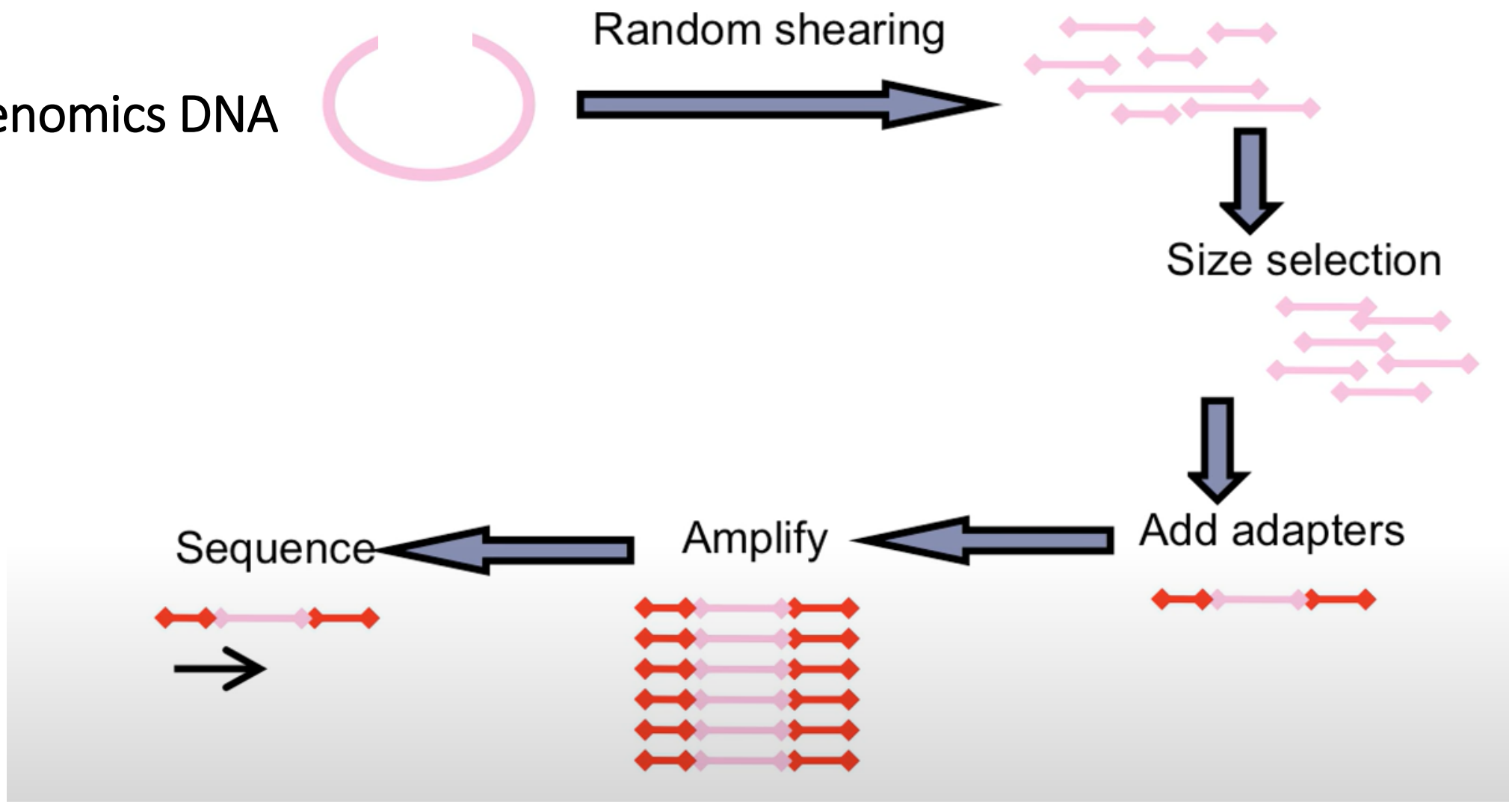
https://www.nobelprize.org/prizes/medicine/2018/press-release/

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 thế hệ mới (1)



Whole genome sequencing – Shotgun Library Prep



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

4.2 Mapping reads to reference

```
Reference:...CTCGAATGCG...
Read1:       CTCGAATACG
Read2:       CTCGAATACG
Read3:       CTCGAATACG
Read4:       CTCGAATACG
Read5:       CGCGAATACG
Read6:       CGCGAATACG
Read7:       CGCGACTACG
Read8:       CGCGAATACG
```

Maternal

Paternal

Error

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   T    G    129   PASS    GT:GQ:DP:AD  0/1:48:8:4,4
20      17330   .           G    A    150   PASS    GT:GQ:DP:AD  1/1:49:8:8,8
```

4.4 Annotating variants

ANN=G|stop_gained|HIGH|OR4F5|ENSG00000186092|transcript|ENST00000641515.2|protein_coding|3/3|c.822T>G|p.Trp274*|882/2618|822/981|274/326||Pathogenic

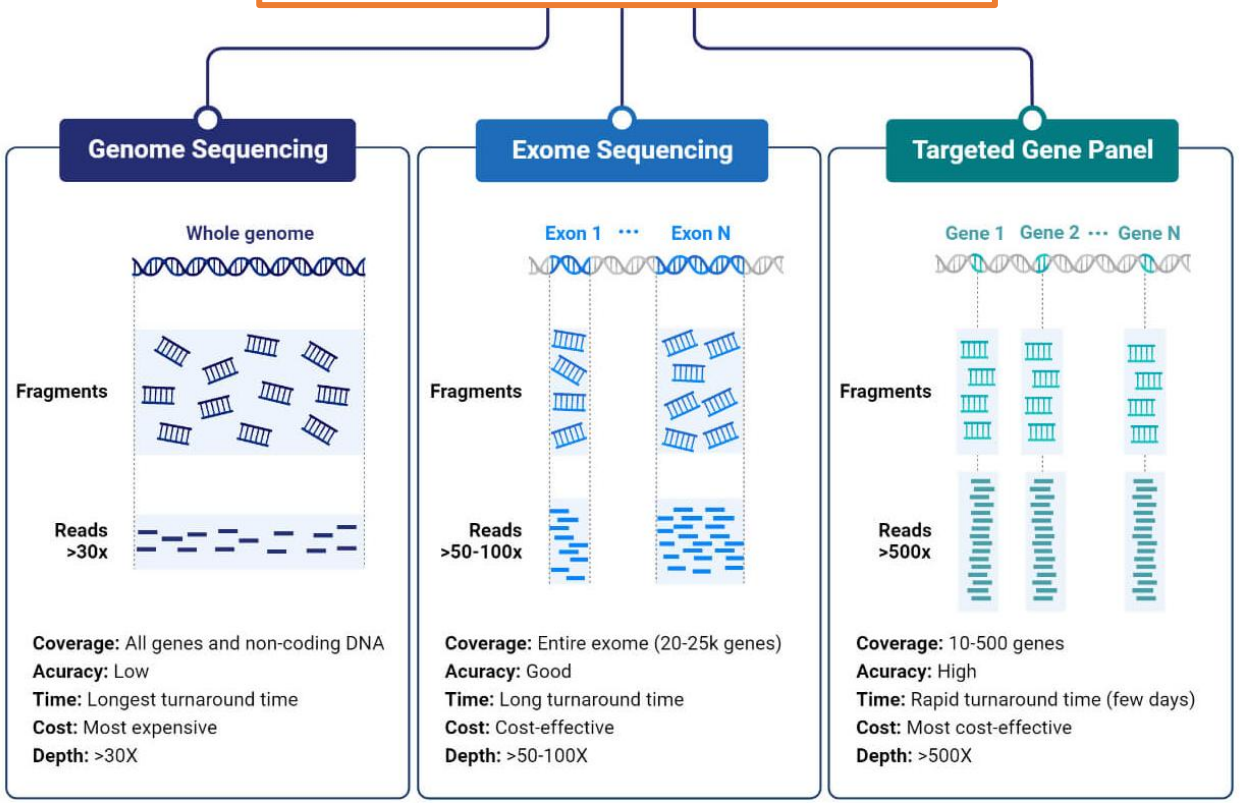
ANN=A|frameshift_variant|HIGH|ZSWIM2|ENSG00000163012|transcript|ENST00000295131.3|protein_coding|9/9|c.1238G>A|p.Ile413|1293/2451|1238/1902|413/633||;LOF=(ZSWIM2|ENSG00000163012|1|1.00)

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/

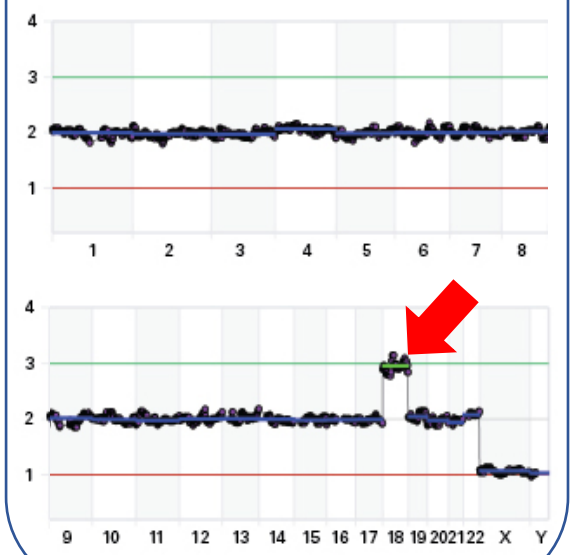
Ứng dụng giải trình tự gen thế hệ mới trong lâm sàng

Single Nucleotide Variant (SNV)

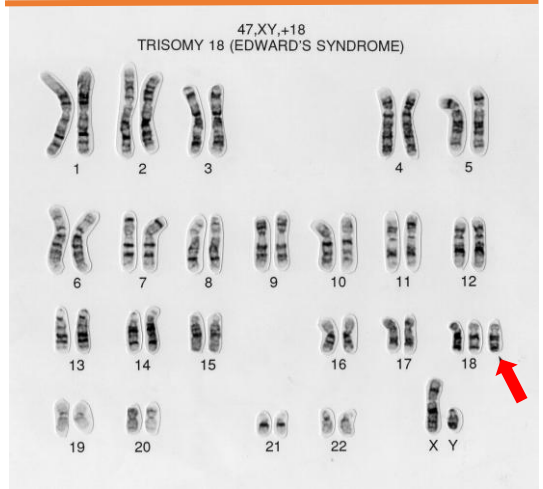


https://microbenotes.com/next-generation-sequencing-ngs/

Copy Number Variation (CNV)
Low-Pass Genome Sequencing

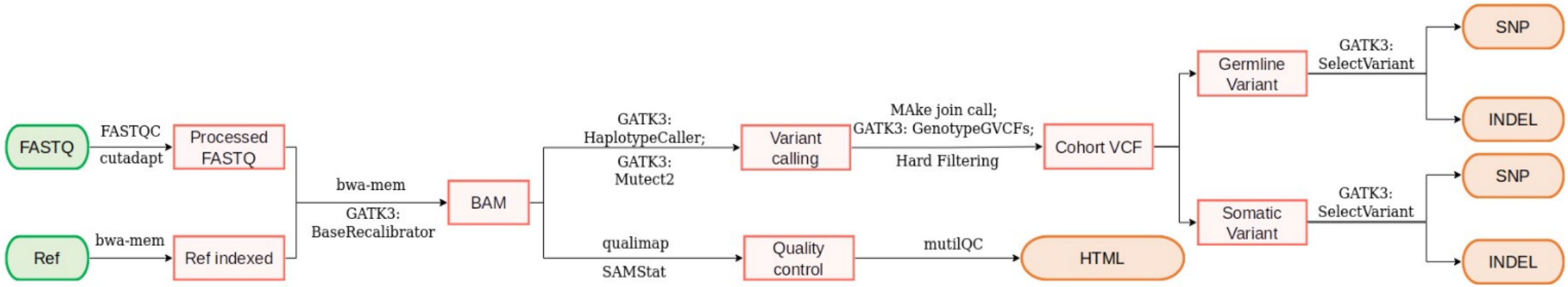


Karyotype



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

Whole Genome Sequencing pipeline



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

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)

rs1057519541

Current Build 156
Released September 21, 2022

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		

Variant Details | **Clinical Significance** | 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

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

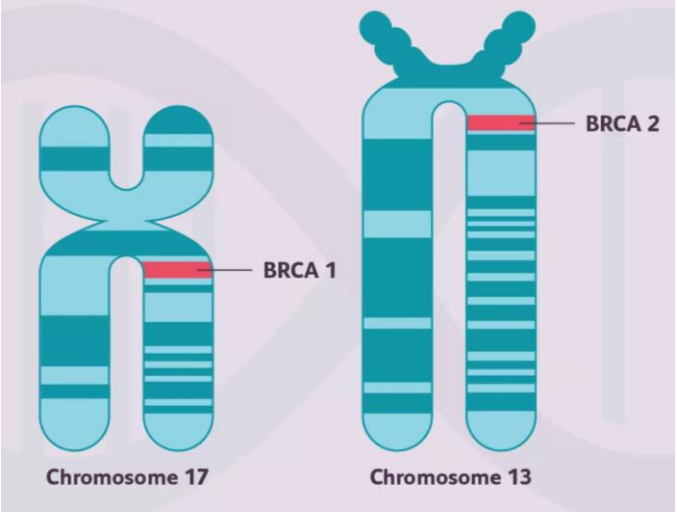
Ví dụ 2: Sàng lọc bệnh ung thư bằng xét nghiệm gen Gene Panel – 53 genes của MGI

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 vùi 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 ĐẠT (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 BRCAaccuTest™Plus (CE-IVD)**
- Phần mềm hỗ trợ phân tích kết quả: **NGeneBio NGeneAnalySys™ (CE-IVD)**

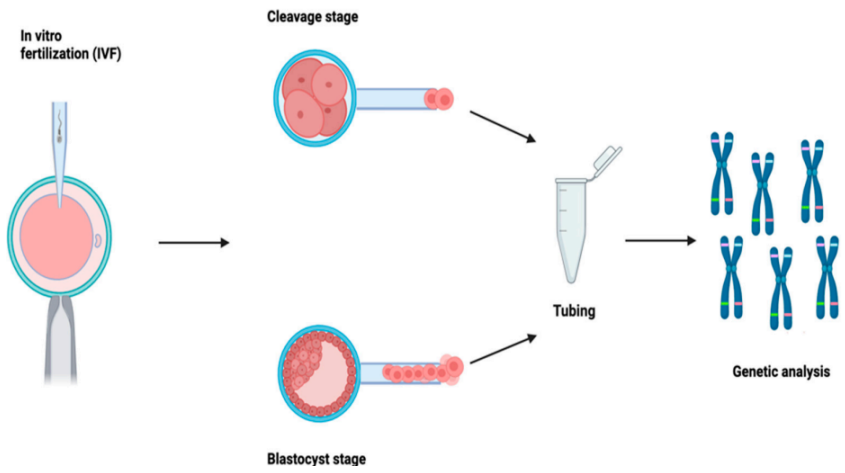
Kết quả: **PHÁT HIỆN 1 BIẾN THỂ MẤT ĐOẠN NUCLEOTIDE NHỎ (DEL) GÂY BỆNH TRÊN GEN *BRCA1***

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**

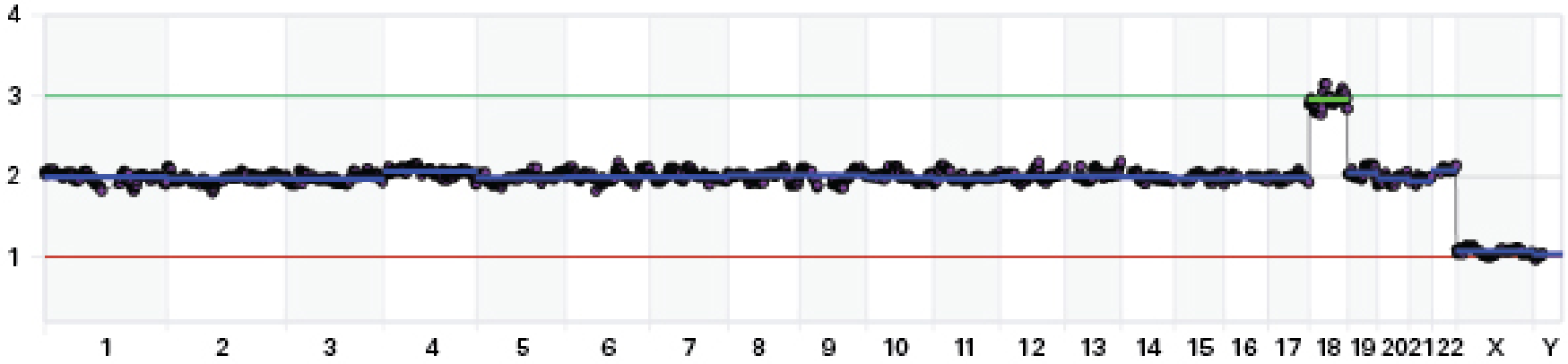
Nguyễn Huy Thịnh, MD
Nguyễn Hoài Thu, MD
Nguyen Le Duc Minh, MD

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

Xét nghiệm tiền làm tổ PGT-A



10.3390/genes14112095

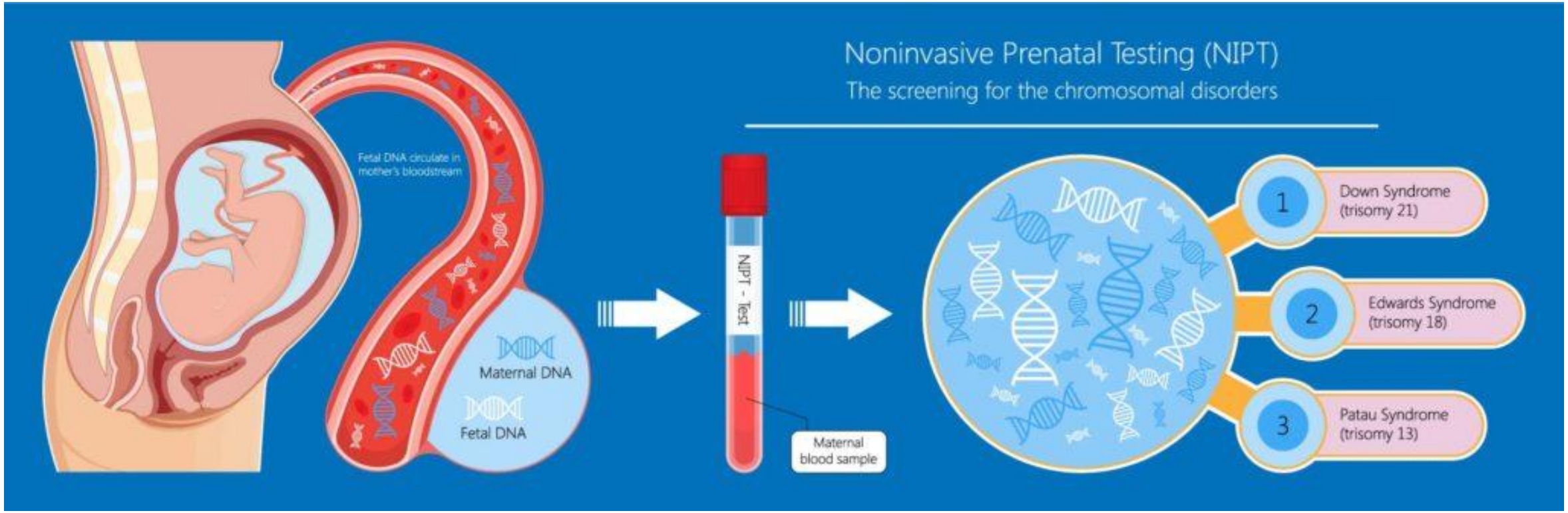


Mẫu phôi có 3 nhiễm sắc thể 18 (trisomy) trong DNA hệ gen

BS. Nguyen Le Duc Minh
CN. Dao Khuong Duy
ThS. Ngô Đại Phú

https://www.takarabio.com/learning-centers/next-generation-sequencing/technical-notes/reproductive-health-technologies/embgenix-pgt-a

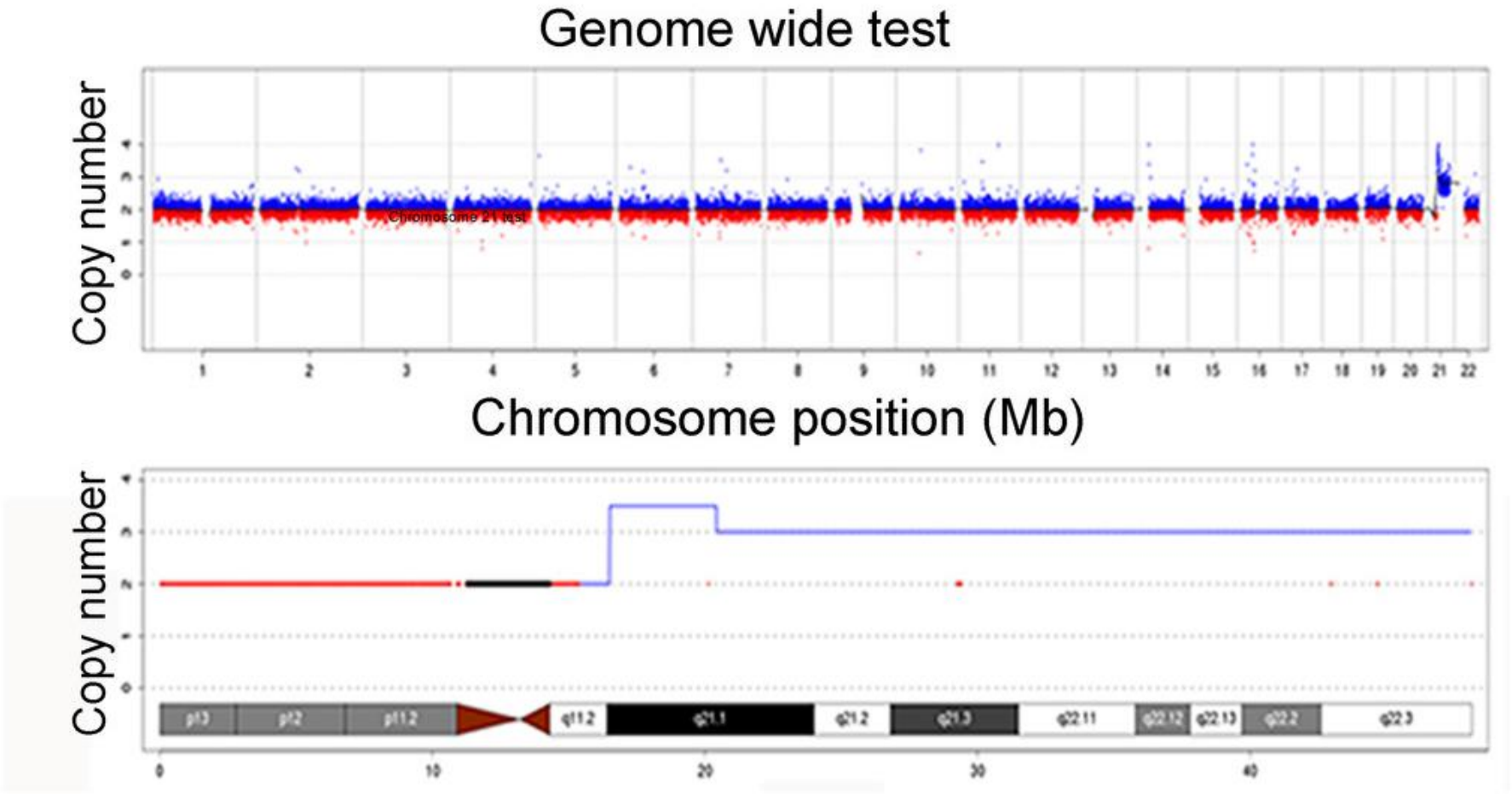
Ví dụ 5: Sàng lọc bất thường nhiễm sắc thể ở thai nhi



CN. Dao Khuong Duy
BS. Nguyen Le Duc Minh
ThS. Ngô Đại Phú

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

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



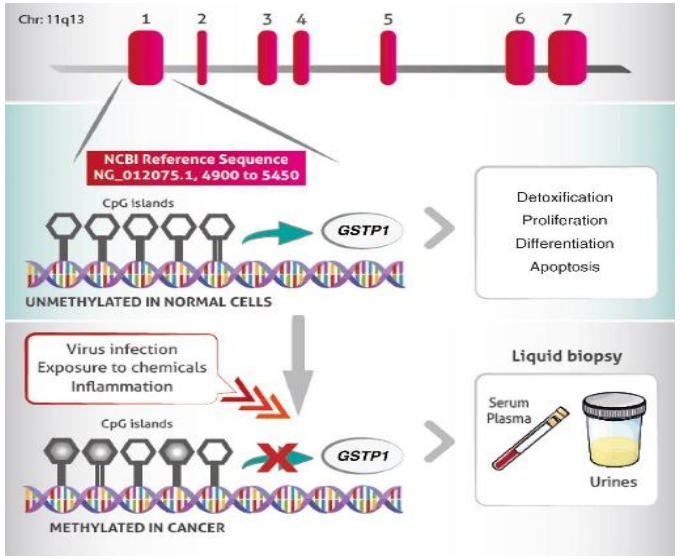
10.7717/peerj.14400

Ví dụ: Hỗ trợ điều trị ung thư tuyến tiền liệt kháng hay 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 [a,b,c,e,†], Wenjia Qu [b,†], Hui-Ming Lin [b,c], Calan Spielman [b], Daniel Cain [d], Cindy Jacobs [d], Martin R. Stockler [a,e,f], Celestia S. Higano [g], Johann S. de Bono [h], Kim N. Chi [i], Susan J. Clark [b,c,†], Lisa Glen Horvath [a,b,c,e,†,]*

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



European Association of Urology

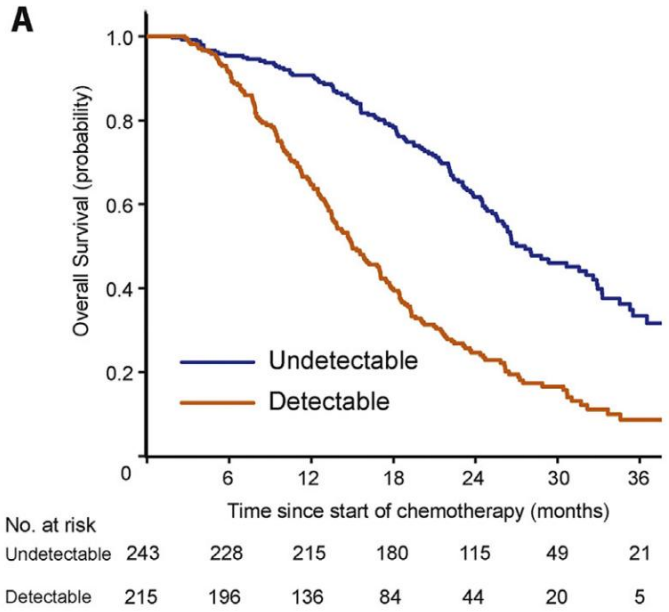
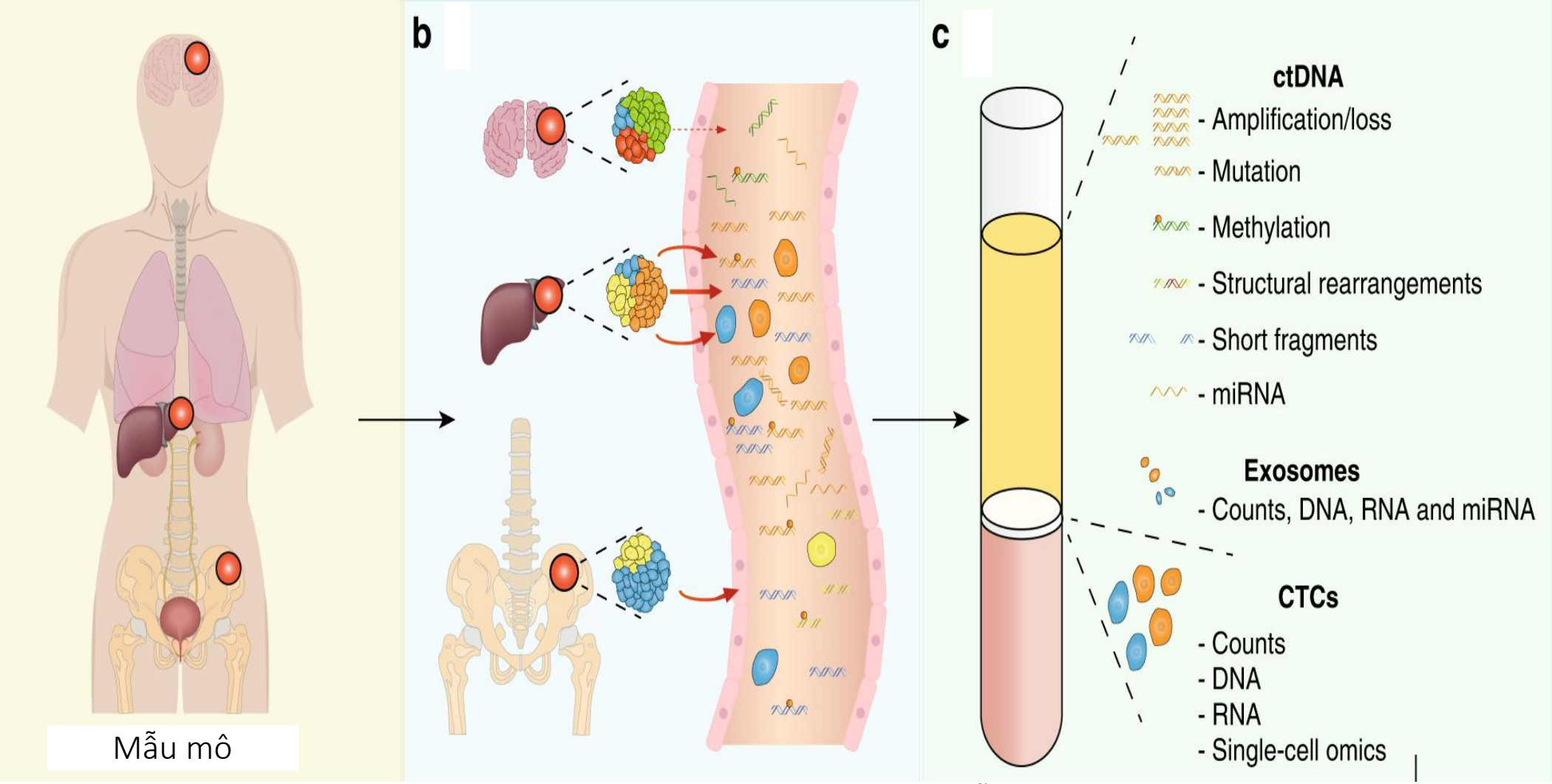


Fig. Kaplan-Meier estimates of survival in patients with a detectable mGSTP1 at baseline. (A) OS according to mGSTP1 detection after two cycles of docetaxel.

10.1515/cclm-2017-0703

Mẫu cho XÉT NGHIỆM gen: mẫu mô và máu

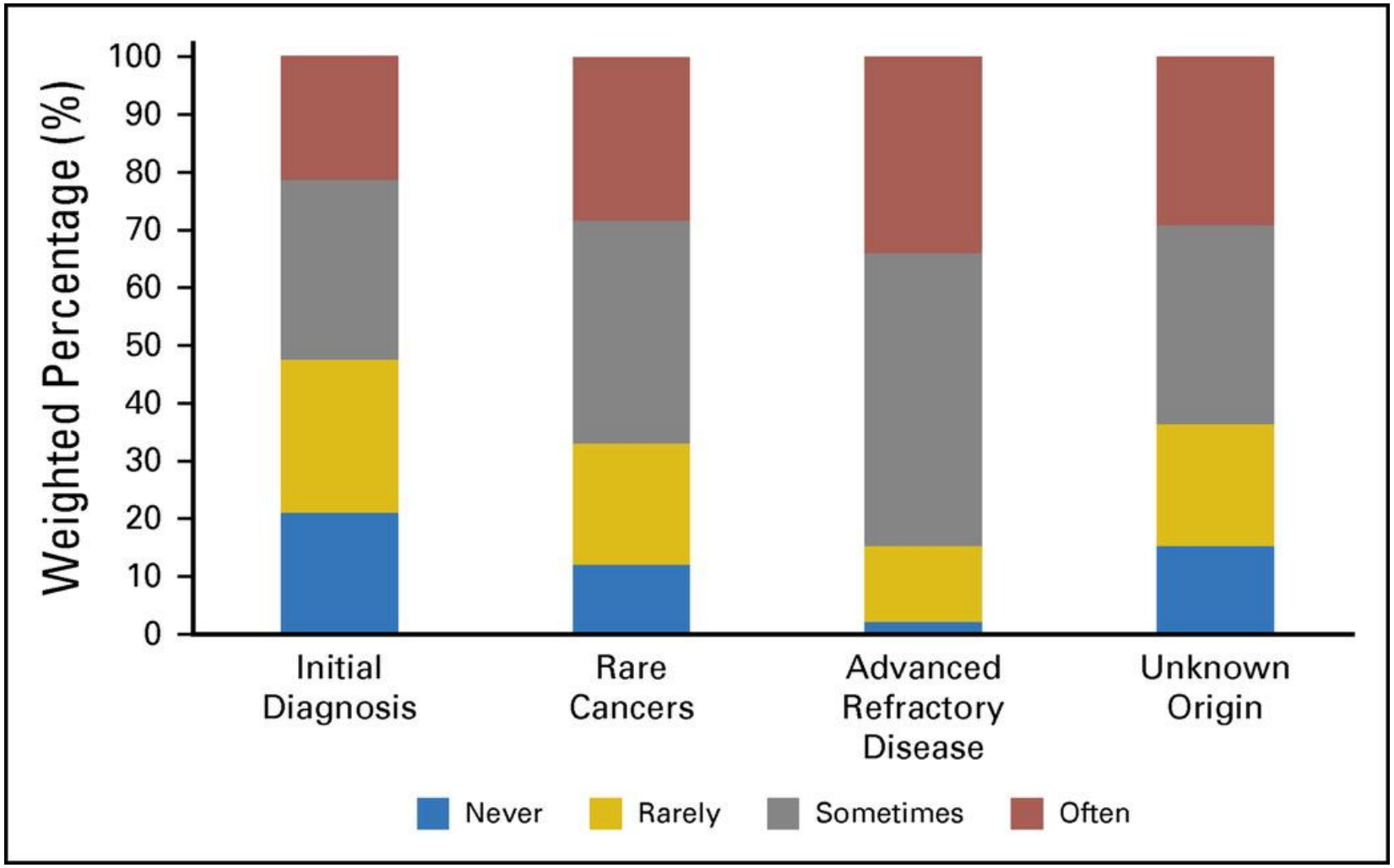


Mẫu mô

Mẫu máu

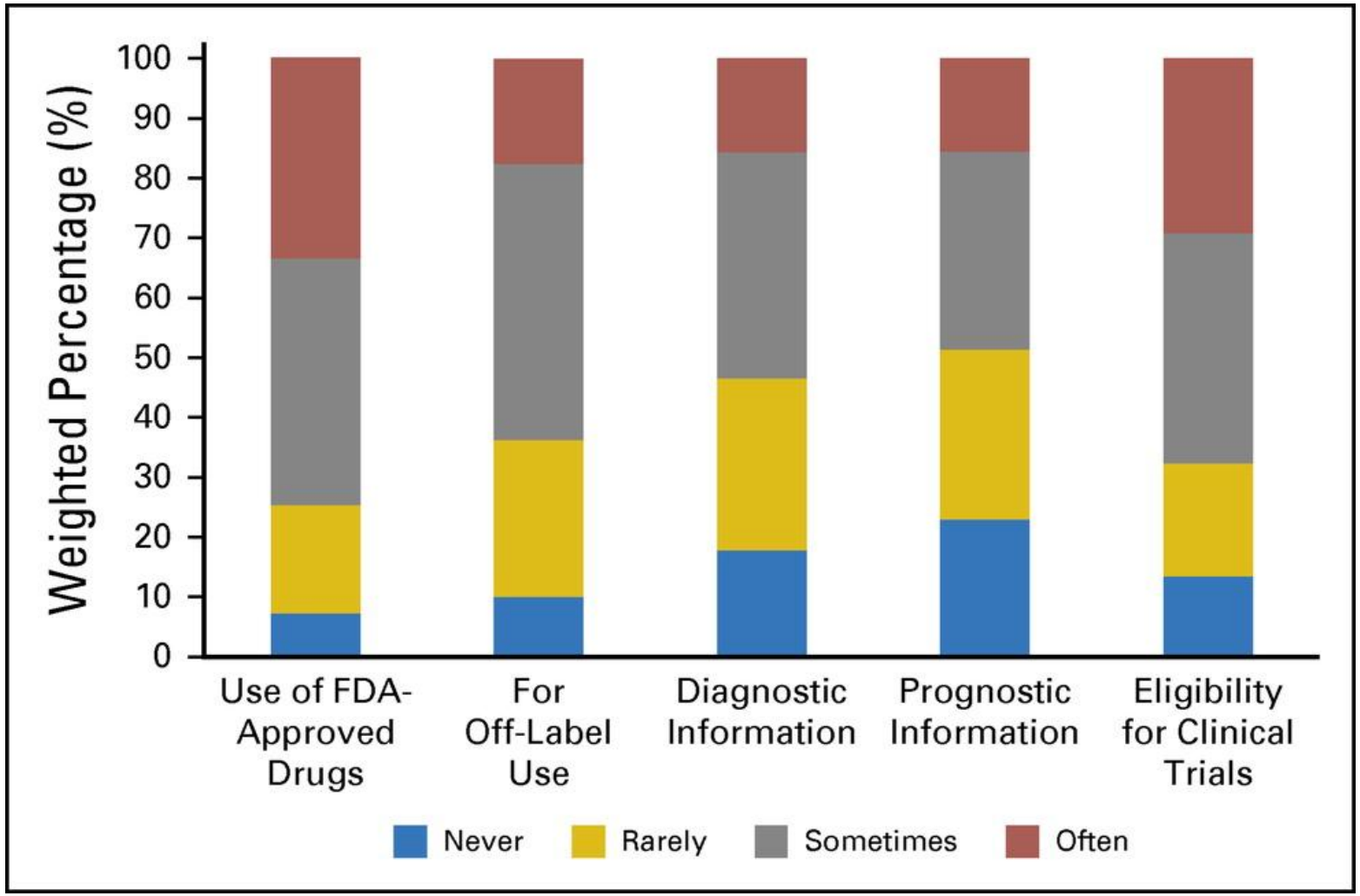
10.1038/s41416-022-01881-9

NGS is used in clinical cancer care



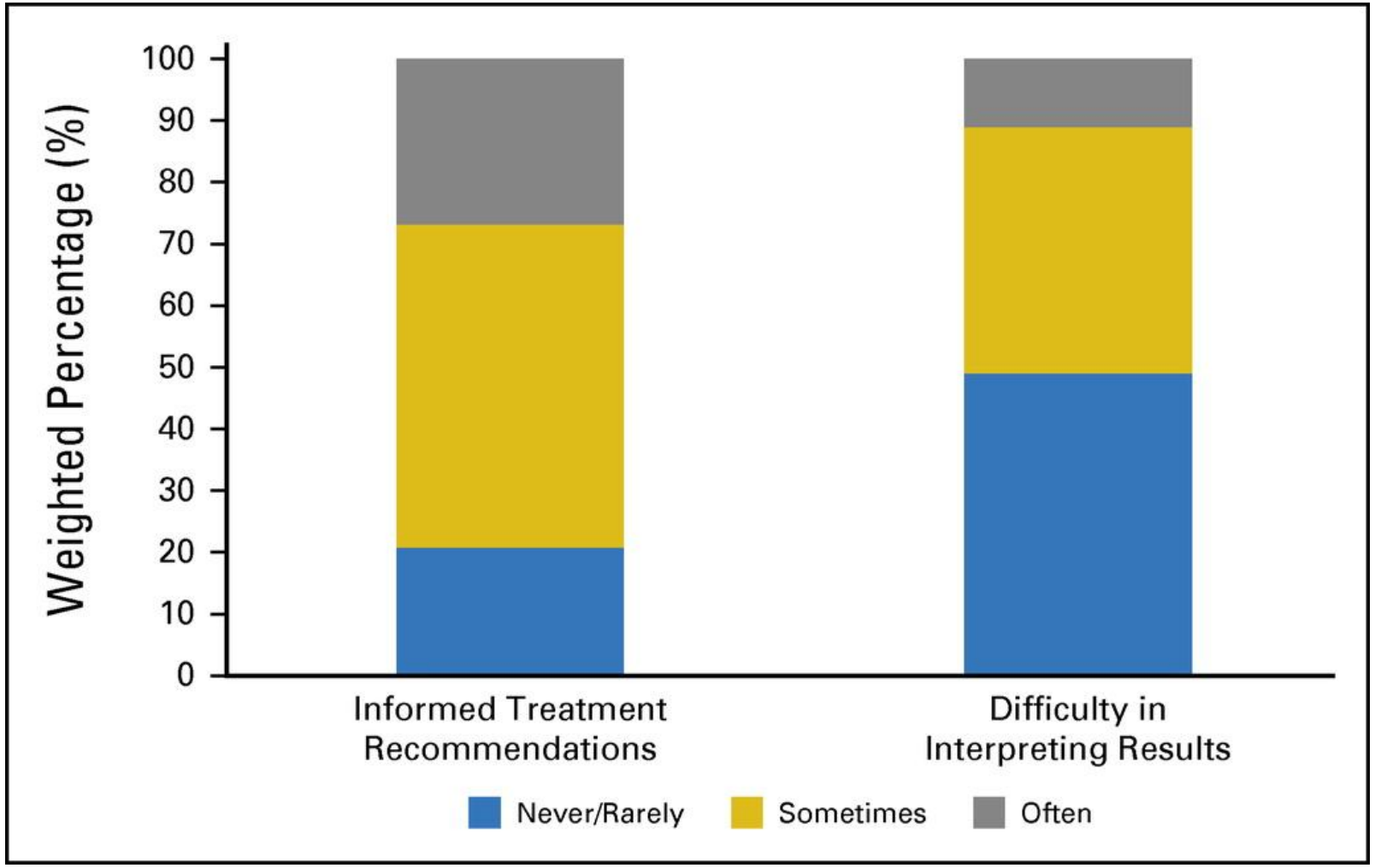
https://ascopubs.org/doi/10.1200/PO.18.00169

NGS is used in clinical cancer care



https://ascopubs.org/doi/10.1200/PO.18.00169

NGS is used in clinical cancer care



https://ascopubs.org/doi/10.1200/PO.18.00169

Xin gởi 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 Thuy-Minh-Thu
- Thạc sỹ, bác sĩ Nguyễn Huy Thịnh
- Thạc sỹ Như
- KS Nguyễn Minh Hoàng

Nhóm cộng tác nghiên cứu

- Trần Quang Khải
- 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

Zalo: 0901802182