

Giới thiệu Giải trình tự DNA (DNA-seq)

Ngày 07 tháng 12 năm 2024

TS. Lưu Phúc Lợi

Email: luu.p.loi@googlemail.com

Zalo: 0901802182

Nội dung

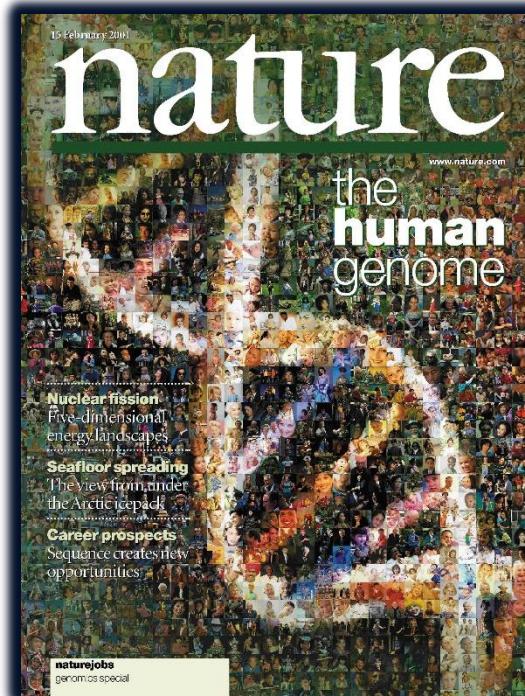
1. Giới thiệu về Dự án bộ gen người
2. Giới thiệu kỹ thuật giải trình tự thế hệ mới (NGS)
3. Quy trình **XÉT NGHIỆM** gen bằng phương pháp giải trình tự thế hệ mới
4. Năm ví dụ về **XÉT NGHIỆM** gen cho **BỆNH DI TRUYỀN** và **UNG THƯ**
5. Giải trình tự gen thế hệ mới trong nông nghiệp và môi trường

Giới thiệu về Dự án bộ gen người

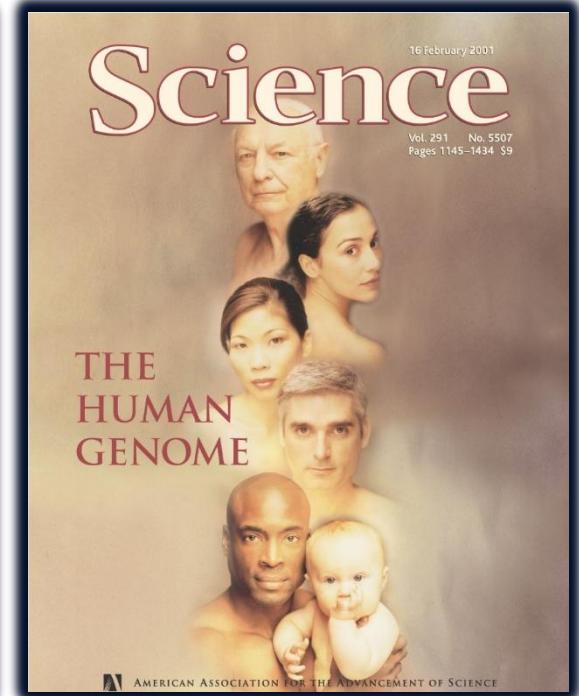
Dự án hệ gen người HGP (Oct 1990 - April 2003)

- 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

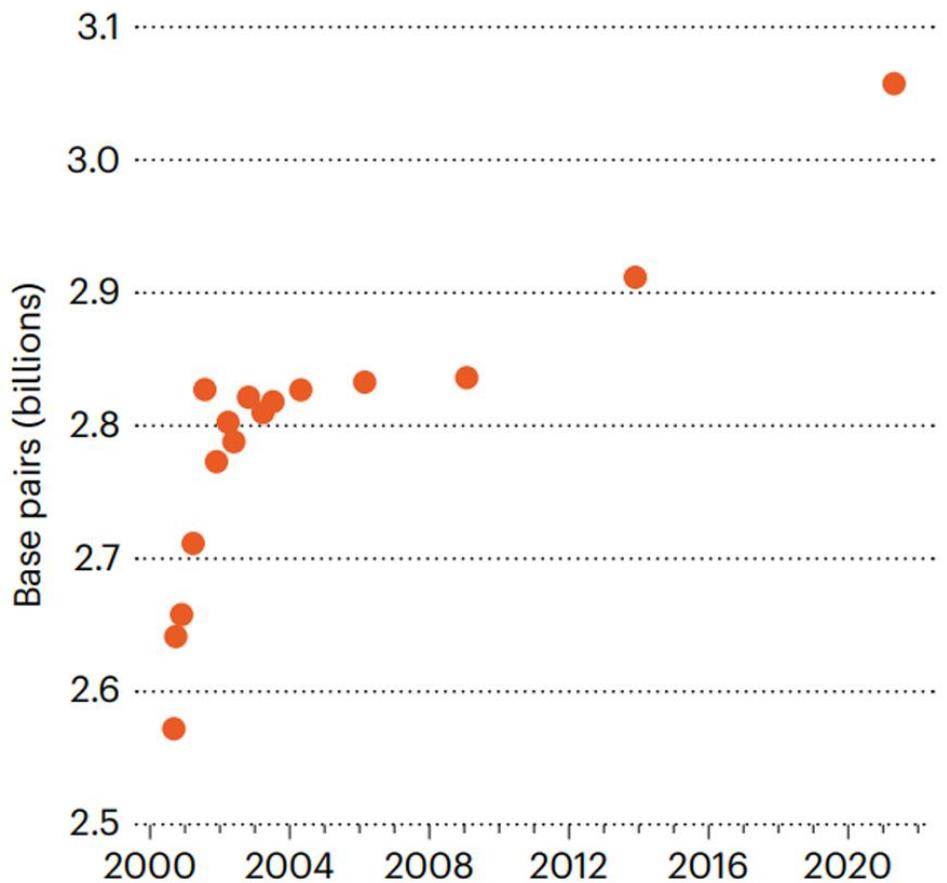
Human	hs1	Jan. 2022	T2T Consortium CHM13v2.0	Available
	hg38	Dec. 2013	Genome Reference Consortium GRCh38	Available
	hg19	Feb. 2009	Genome Reference Consortium GRCh37	Available
	hg18	Mar. 2006	NCBI Build 36.1	Available
	hg17	May 2004	NCBI Build 35	Available
	hg16	Jul. 2003	NCBI Build 34	Available
	hg15	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)

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

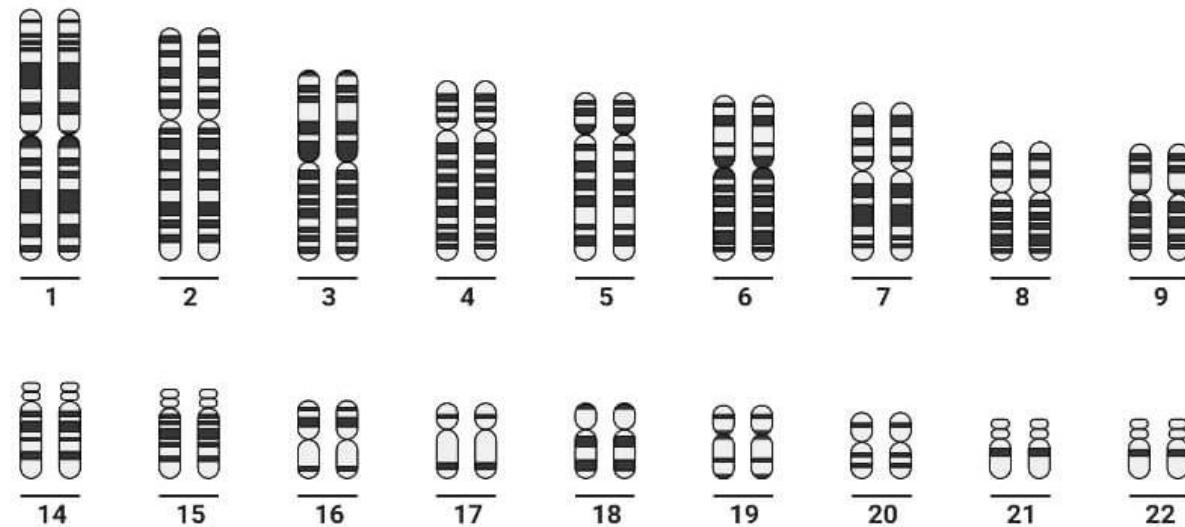
Researchers have been filling in incompletely sequenced parts of the human reference genome for 20 years, and have now almost finished it, with 3.05 billion DNA base pairs.



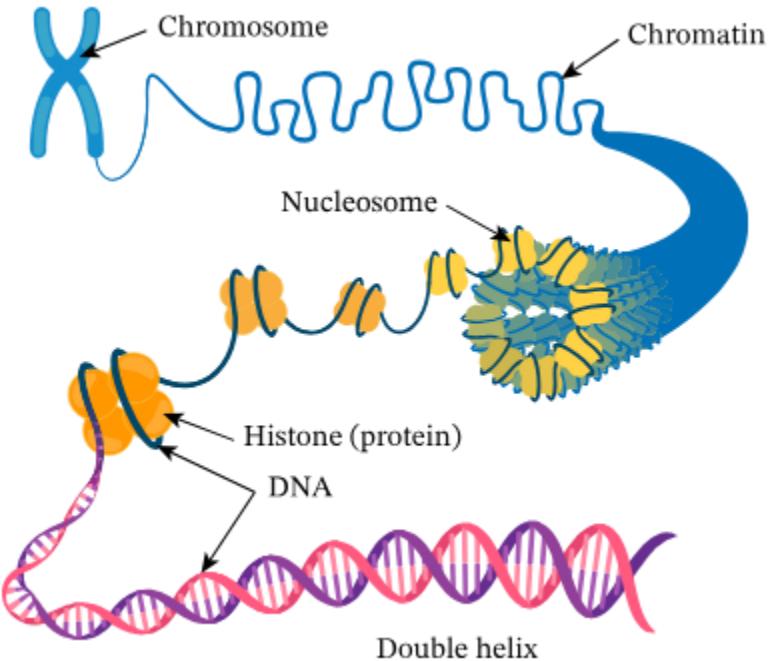
0.3% of sequence might still have errors. Includes X but not Y chromosome. Count excludes mitochondrial DNA.

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 ataaaaaatt 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 tgcgtcggct gggcaggcgg 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 aaacaatgca 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 tggggaaaaa gcaactccctt gcacatttggaa ggaaccgaga aagactgacc ccgaggacag
50401 cagccagcat ggccttcctt gggagccat gttggggat tccgtcgtca gccaaggctc
50461 agcccttgc gtcgcagggtg ctggccttgg cctttcccc tcccatgcag gggcaggg
50521 gagatggctt ctgaggaccc tttgcgttggg aatagattt ccaggaggt
50581 ttaaaggcgc tgagtgtgtc atccagctaa gcctggggaa ggagcttggc tcaggcctgt
50641 acagggtgtga cagggatggg gactgggaag taagagatga aaccctggct ggaggctgt
50701 agttccaca gccagcgtt gacaggagggt gtccagatat acccactgtt gccctcacca

Dữ liệu hệ gen cho quần thể



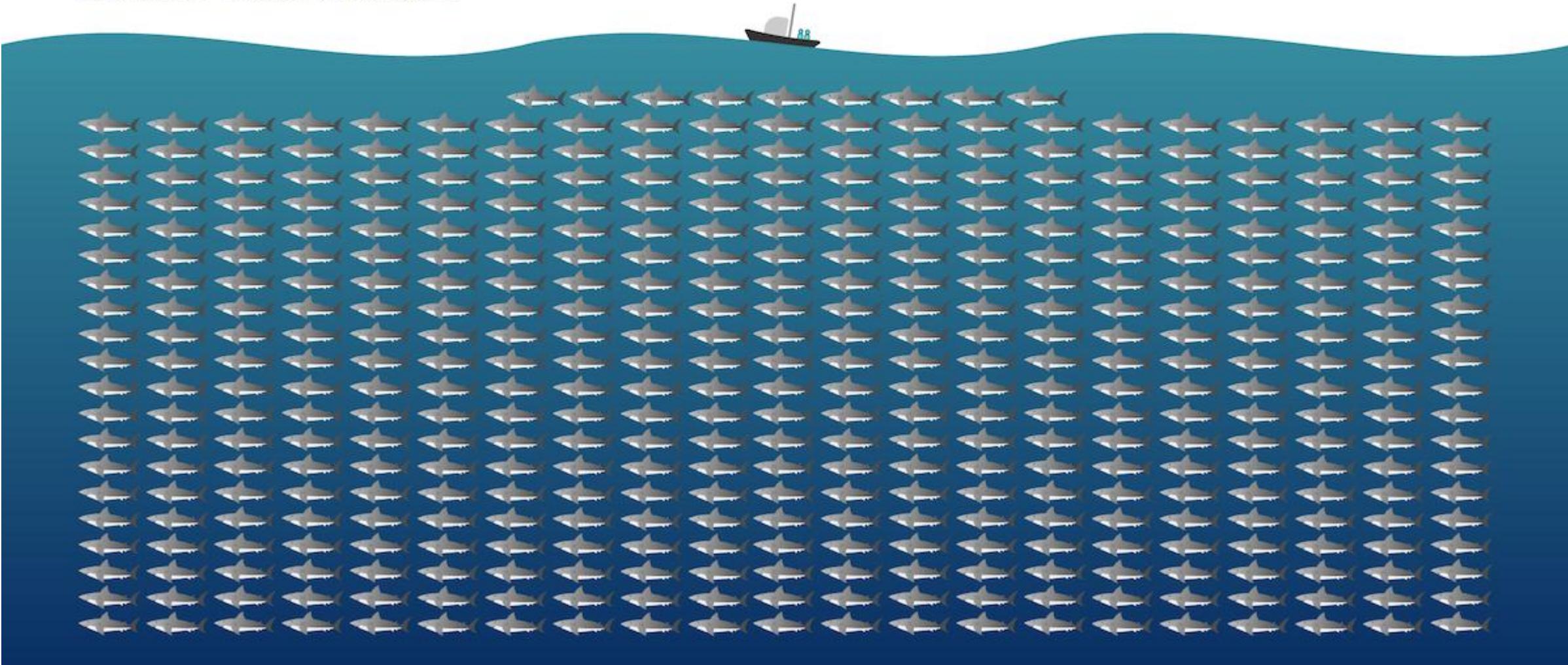
48541 agccctcaa agaaatgttc tcagcaggca tggagccag gacttgtcc cttgggttag
48601 agagccgggt tgaaggtgac tgaagtaaa tggagacta gaggcggggg ggtgtgttag
48661 ttctggagg tgggggtgtt gggacactgc tigtactga gatgcacccc tgccagttct
48721 gcttaaagat ttgaggcggg gggcaggggg gcggagtgaa gtcattttac ttgttaagtaa
48781 tttaaacct ttaatatta aagcaaacgt ggatatgtaa tgaataaaat tcattctgga
48841 atgaaaaatt cacgtatgt taaaaataa 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 ctgtcccggt ccagttctg aggacaagag ccaccaccc tagcaataaa agaccaggca
49141 acccttgac tcatcttgc gactctctgg aatcagagg tagccacatc gctggagaggt
49201 ggatgtaaagc actcgggtga aaaggatcaa ggaatcgagg gacaggatgt tggggacatc
49261 acatggacaa tgacagagaa gggggcaca gccgatgtgg gggaggggg cggcagttcc
49321 tacatccccct ggctgaaagc acgctccagg gcagaaggaa aaacactgtc tttgggtcc
49381 aagagacctg agttcaatt ctggctccac cactgaccac ctgttaacc ttgaaactgt
49441 gctgctgaa cctcagggtt ccctctaaa aatagaggag aaaaggatgtc atttccctt
49501 gcccctgtga gaacgaaatg gtcaagcac caaggagct cagcaaaagggt cggccctgcc
49561 cccgcctggc caaaccttc ctcttcaggaa ggccacggca accgtatgtt gacagaagag
49621 cagcaccttgc atttaatgtc tcccgatcg tgccttgcg caagtccact aacctctgt
49681 ggctgttcc tcattggaa aatatggctg ccagtaaaac ctggccctgc caccctctgg
49741 ggcacttggc aaacagacaa agatgtccaaat tgcgcggctt gggccaggcg cagtggctca
49801 tgcgtgtaat cccagcaatt taggaaggcc aggtggggcg atcaccctgat gtcaggaggt
49861 tgacccaggc ctggccaaact tggtaaaacc ttgtctctac aaaaataaca aaattagccg
49921 ggcatgtgg cggtgttgc taatcccaat tactcgggag gctggaggaa gagaatcgct
50081 tgaacccggc agggggagggt tgcatgtggc caagattgtt ccaactgcact ccagctgggg
50041 caacagagcg agactctgc taaaaaaaaaaa aaaaaaaaaaa aaacaatgca gagctggctg
50101 tgtaaaaaaaactgtttccact gcaggccca gtgtccacca ggctgggtt caggccatag
50161 gggggggggc ccagcatcg cctctcaggca gcccggggag gccccggcga tccctggccc
50221 ctgtgggtct ggatgtgttc tagcccaagt cctaggttac acctggcgcc gcttggccctc
50281 tcaggagagg cccagggtga ggaggagcat ggttaagggtt aagctgttgg ggaagtcgc
50341 tggggaaa gcaactctt gcacattggaa ggaaccgaga aagactgacc ccgaggacag
50401 cagccagcat ggccttccctt gggagccat gttggggat tccctgtgc gccaaggctc
50461 agccctgtgt gtcgcagggt ctggcttgc cctctccccc tcccatgtcg gacgcacagga
50521 gagatggctt ctgaggaccc ttggcaggctt tggccctggg aatagattt ccaggaggt
50581 ttaaaggcgc tgagtgttc atccagtaa gcctggggaa ggactggcc tcaaggctctg
50641 acaggtgtga cagggtggg gactggaaag taagatgtaa aaccctggctt ggaggctgt
50701 agttccaca gccaggcgctg gacaggggagg gtccagatgtt acccaactgtt 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

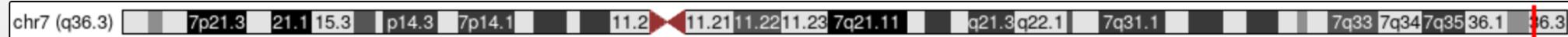


Gene Annotation

[Multi-region](#)

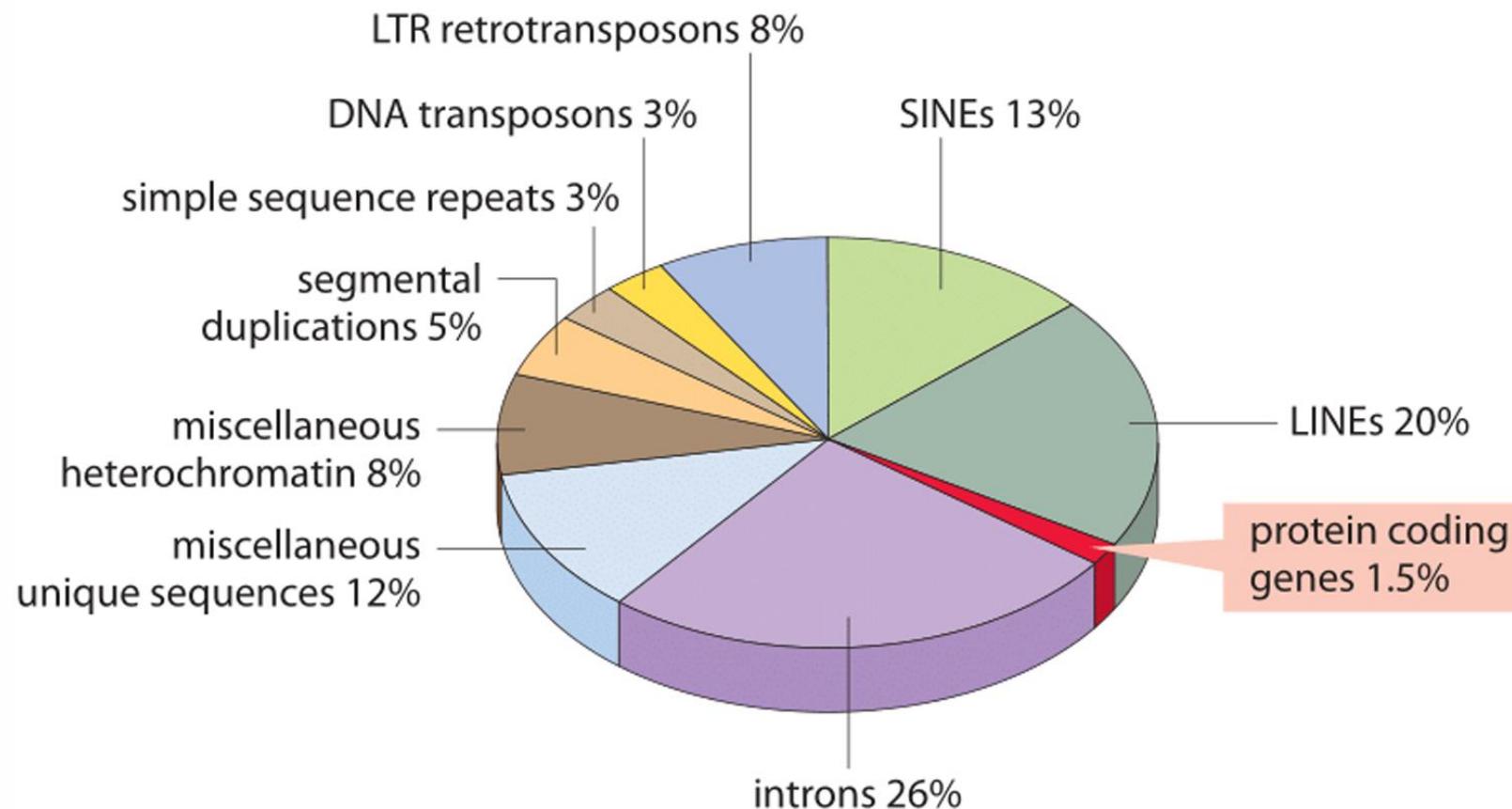
chr7:155,699,333-155,832,762 133,430 bp.

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](#)

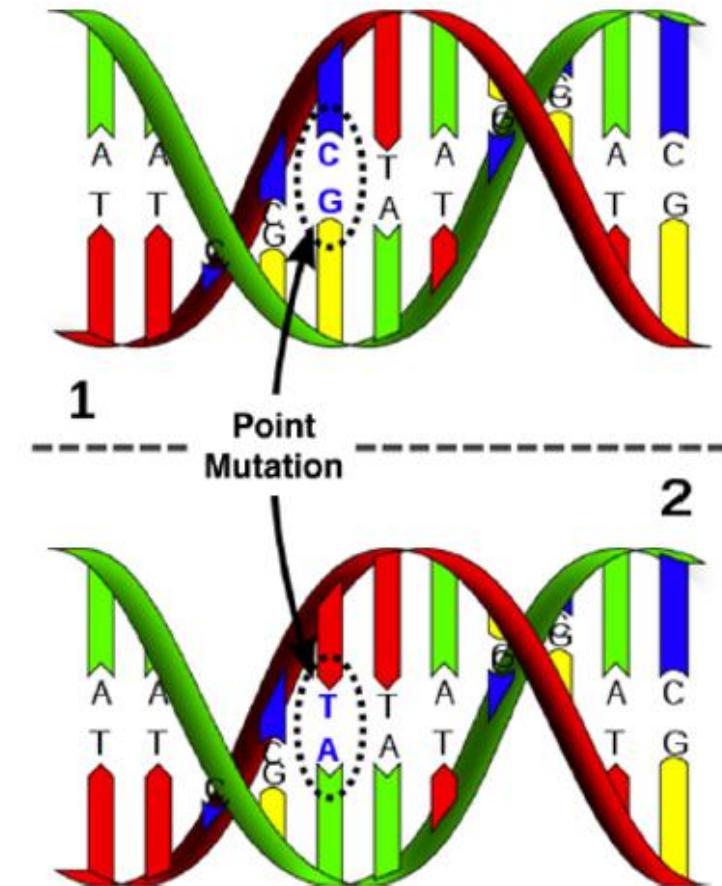
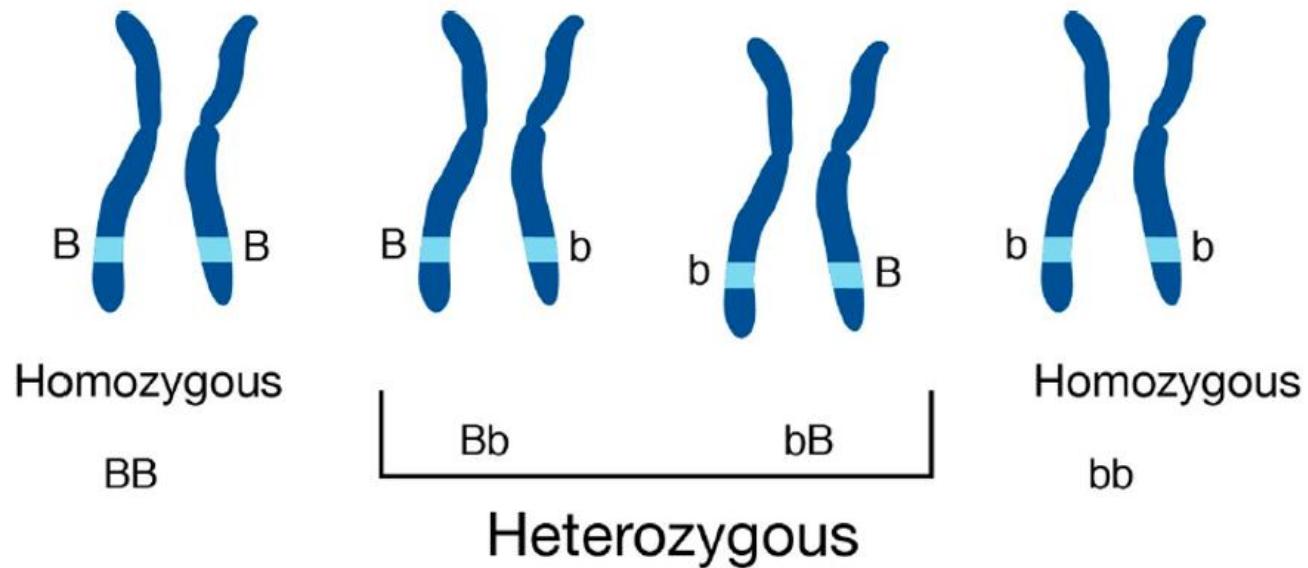
[More about GENCODE Human](#)[Current human data](#)[Release history](#)[Statistics](#)[Data format](#)[FTP site](#)

GTF / GFF3 files

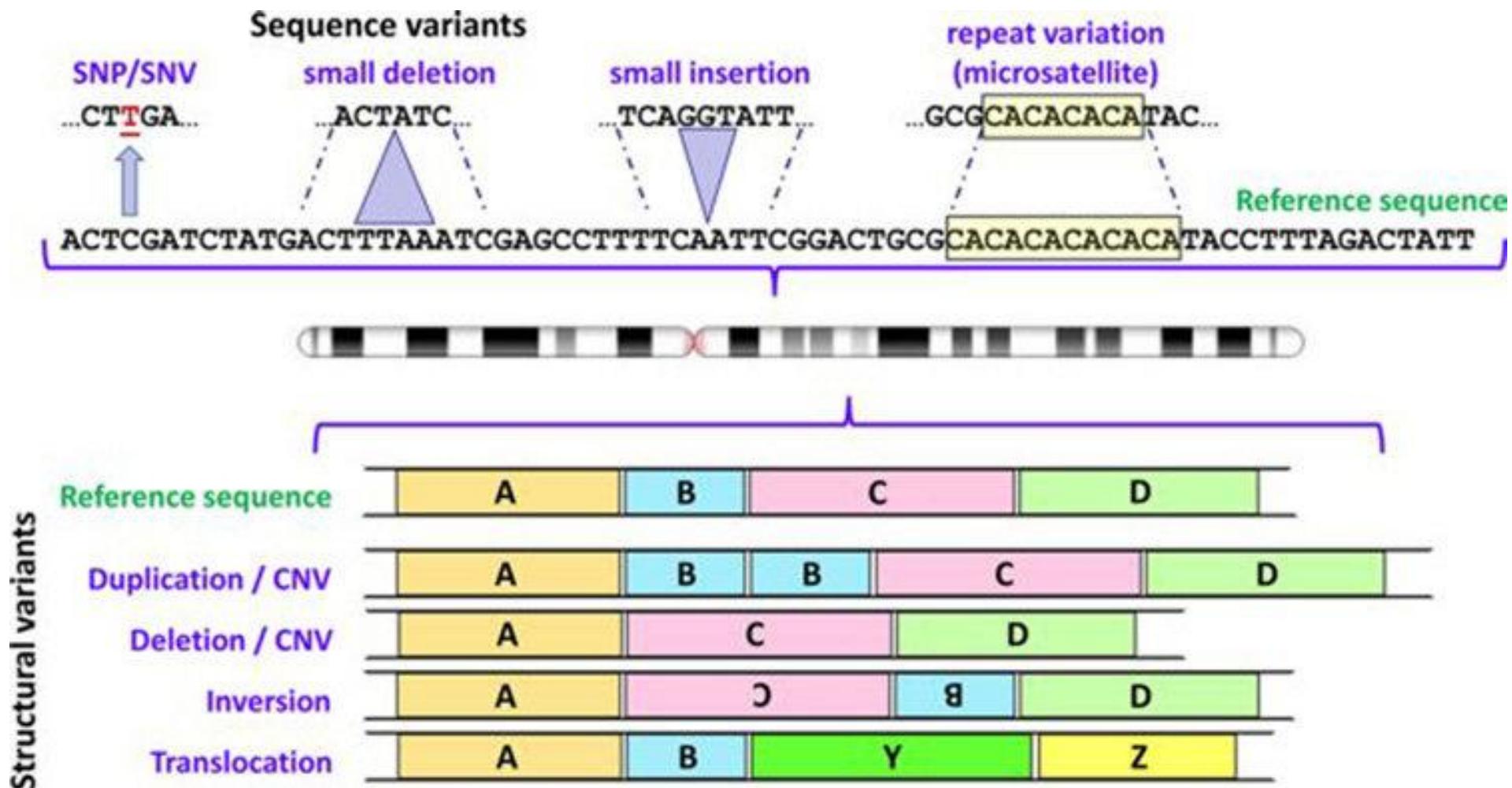
Content	Regions	Description	Download
Comprehensive gene annotation	CHR	<ul style="list-style-type: none">It contains the comprehensive gene annotation on the reference chromosomes only	GTF GFF3
Comprehensive gene annotation	ALL	<ul style="list-style-type: none">It contains the comprehensive gene annotation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

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



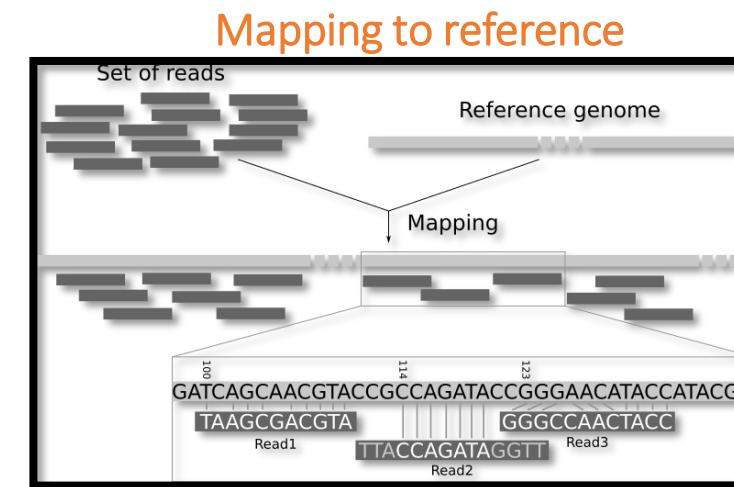
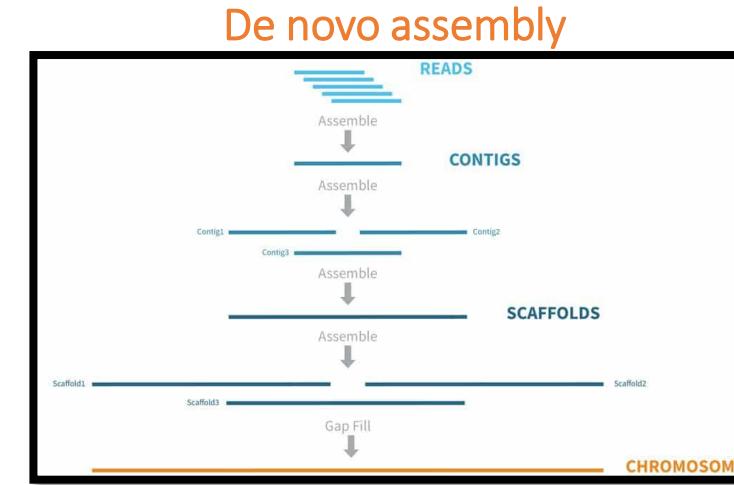
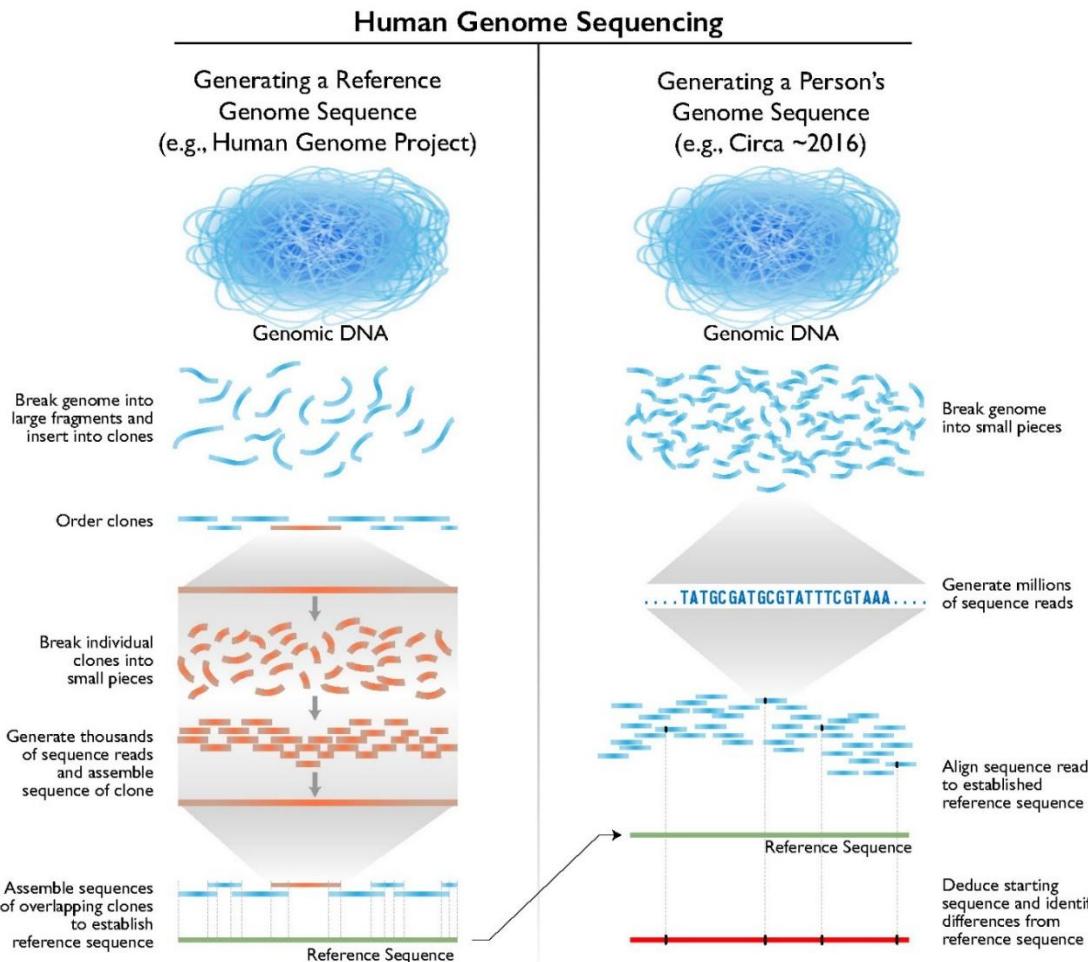
Các loại biến thể trên hệ gen



Giới thiệu về

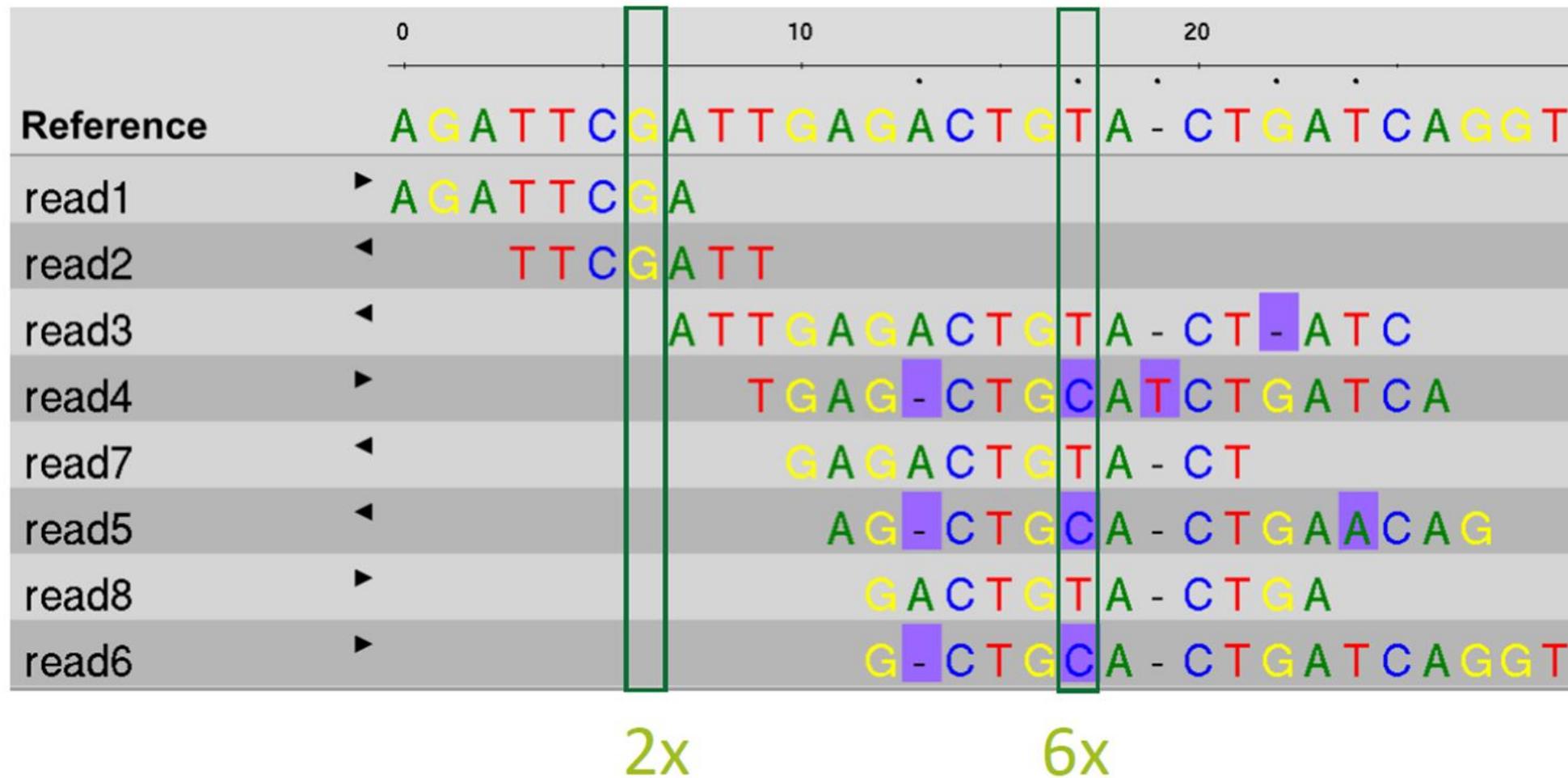
GIẢI TRÌNH TỰ GEN THẾ HỆ MỚI

Giải trình tự gen thế hệ mới (NGS)

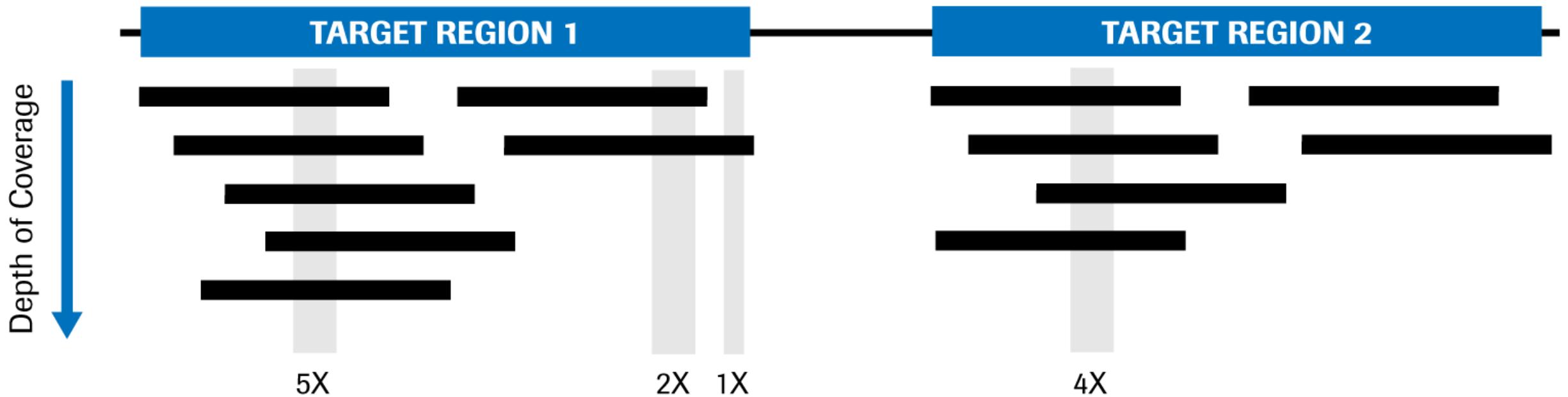


<https://www.genome.gov/about-genomics/fact-sheets/Sequencing-Human-Genome-cost>

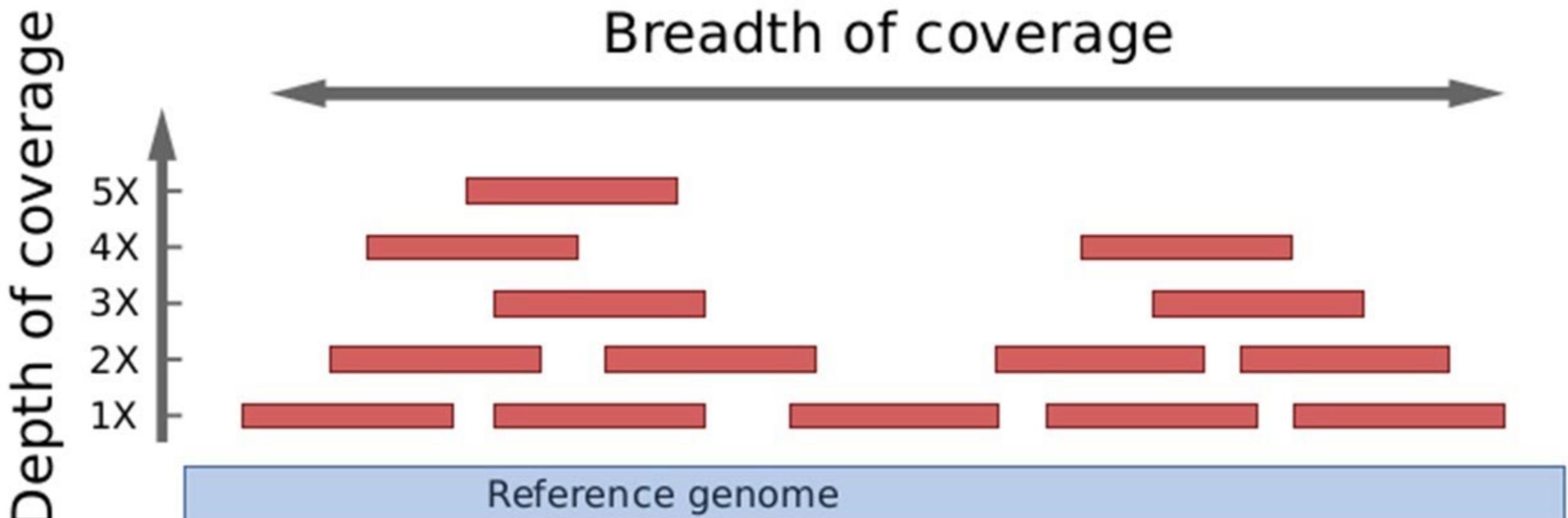
Kết quả của mapping



Depth of Coverage



Depth vs Breadth of Coverage



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

De novo assembly

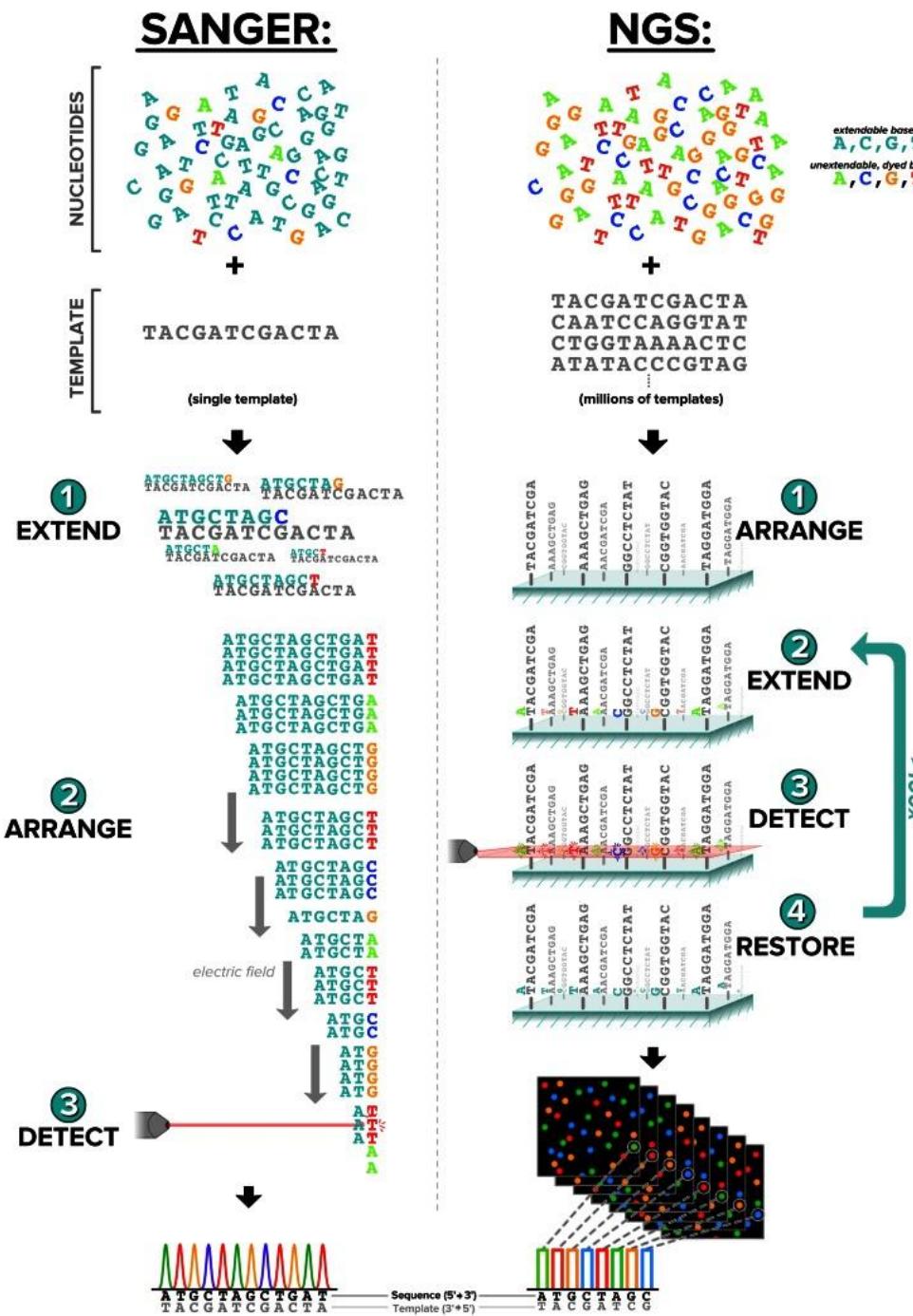


Mapping to reference



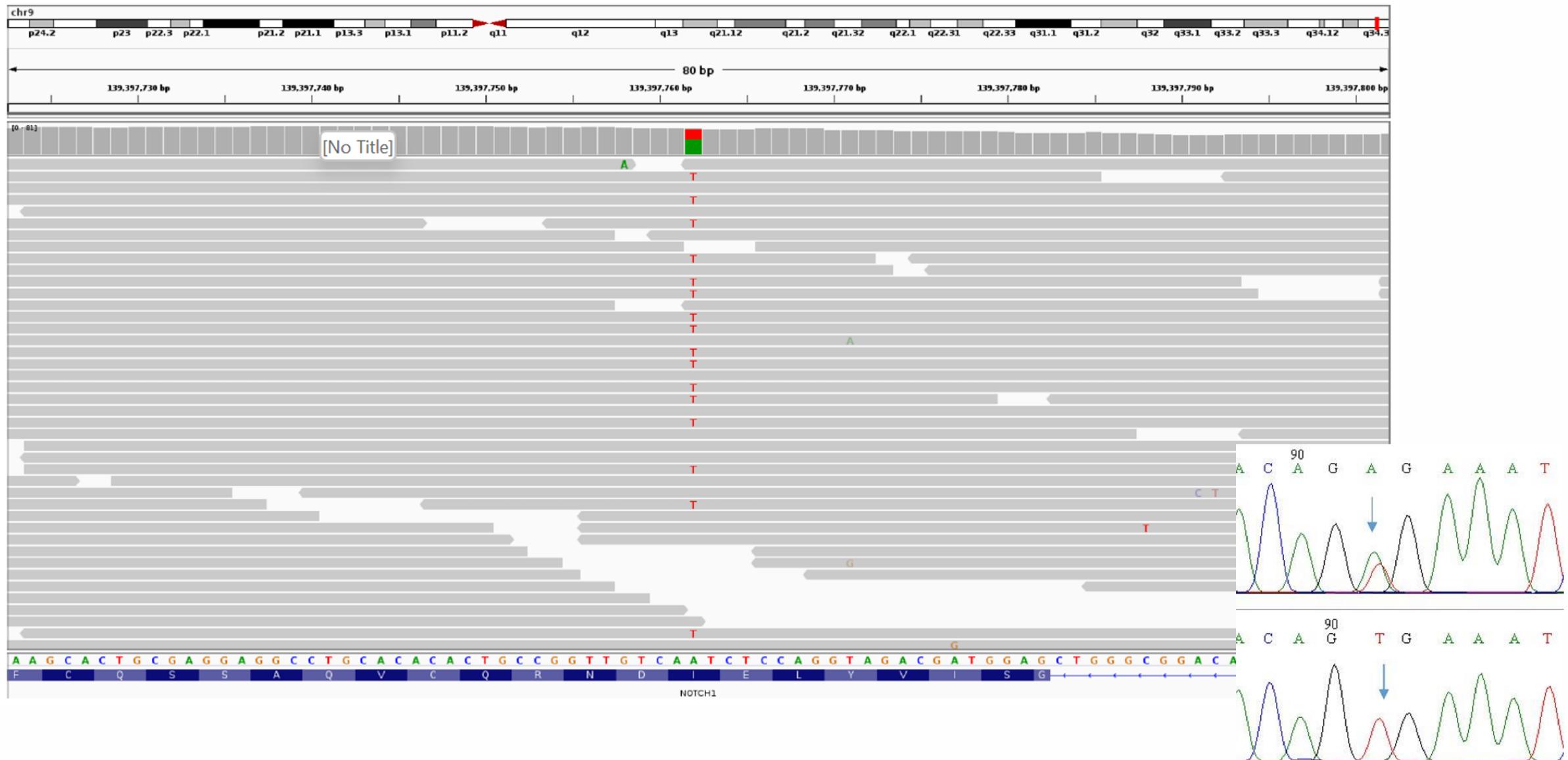
Giải trình tự gen thế hệ mới (NGS): giải trình tự song song

Read1: CTCGAATACG



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

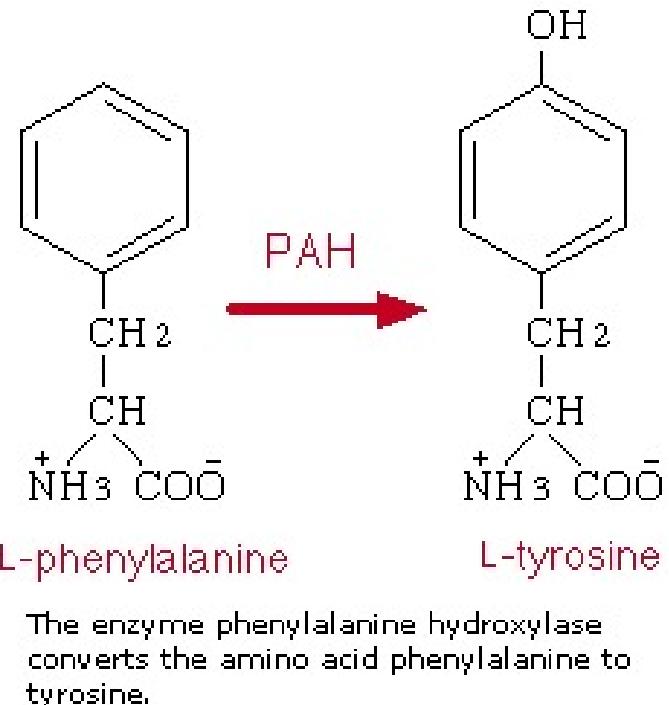
NGS vs Sanger



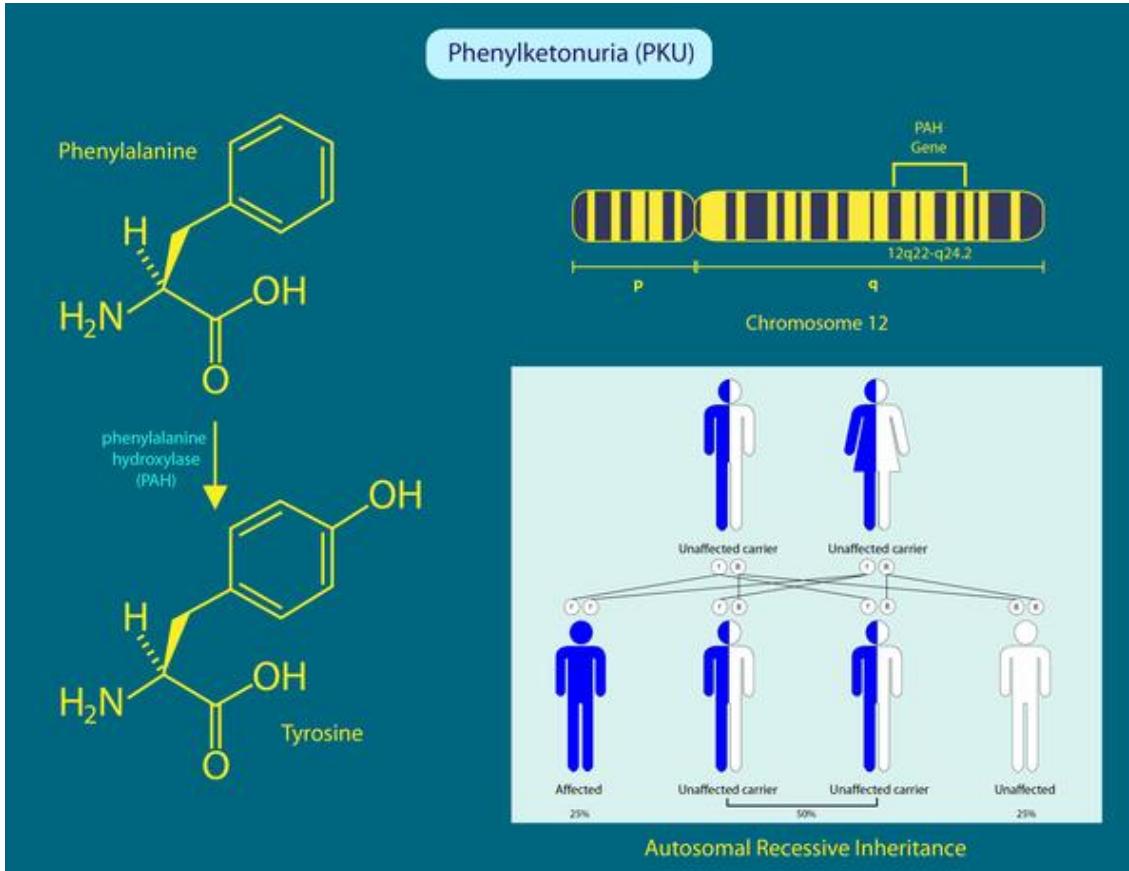
Ví dụ mối quan hệ của
BIẾN THỂ GEN và **BỆNH DI TRUYỀN**

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

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

Gene (Nucleotide)

NT seq	1359 nt NT seq atgtccactgcggctggaaaacccaggctggcaggaaactctctgactttggacag gaaacaagcttatattgaagacaactgaatcaaaatggccatatcactgatcttca ctcaaagaagaagtgggtcattggccaaagtattgcgttatttggaggagaatgtat aacctgaccacattgaatcttagacccctcgtttaaagaaaagatgagatgaattttc acccatttgataaaacgttagcctgcgtctgacaaaatcatcaagatcttgggcat gacattggtgcactgtccatgagcttcacgagataagaagaaaagacacagtggccctgg ttcccaagaaccattcaagagctggacagatggccaatcagattctcagctatggagcg gaactggatgctgaccaccctggtttaaagatccgtgtaccgtgcaagacggaaagcag tttgcgtacattgcctacaactaccgcattggcagccccatccctcgagttggaaatacatg gaggaagaaaaagaaaatggggcacagtgttcaagactctgaagtccctgtataaaaacc catgttgcgtatgagttacaatcacattttccacttctgaaaagtactgtggcccttccat gaagataacattccccagctggaaagacgttctcagttctgcagacttgcactggtttc cgccctccgacccctgtggctggcctgtccctcgggatccctgggatccctggcc cgagtcttccactgcacacagtacatcagatggatccaagccatgtataccccgaa cctgacatctgcatgagctgttggacatgtgcccttgcggatccctgtggc cagtttccaggaaattggcctgcctctgggtgcacctgtatgaaatacatgtaaaag ctcgccacaatttactgtggtagttggctctgcaaaacaaggagactccata aaggcatatggtgcggcctgtcatccttggtaattacagtactgttattcagag aagccaaagcttctcccccggagctggagaagacagccatccaaaattacactgtcag gagttccagcccttattacgtggcagagatgtttaatgtatgccaaggagaaaagtaagg aactttgctgccacaataccctcgcccttctcagttcgctacgaccatacacccaaagg attggaggcttggacaatacccagcagcttaagatggctgattccattaacagtgaa attggaaatcccttgcagtggccctccagaaaataaaagtaa
--------	--

Protein (Amino Acid)

AA seq	452 aa AA seq DB search MSTAVLENPGLGRKLSDFGQETSYIEDNCNQNGAISLIFSLKEEVGALAKVLRLFEENDV NLTHIESRPSRLKKDEYEFFTHLDKRSLPALTNIIKILRHDIGATVHELSRDKKKDTVPW FPTIQELDRFANQILSYGAELDADHPGFKDPMVYRARRKQFADIAYNYRHGQPIPRVEYM EEEKKTWGTVFKTLKSLYKTHACYEYNHIFPLLEKYCFGHEDNIPQLEDVSQFLQTCTGF RLRPVAGLLSSRDFLGLAFRVFHCTQYIRHGSKPMTPEPDICHELLGHVPLFSDRSFA QFSQEIGLASLGAPDEYIEKLATIYWFTVEFLCKQGDSIKAYGAGLLSSFGELQYCLSE KPKLLPLELEKTAIQNYTVTEFQPLYYVAESFNDAKEKVRNFAATIPRPF SVRYDPYTQR IEVLDNTQQLKILADSINSEIGILCSALQKIK
--------	---

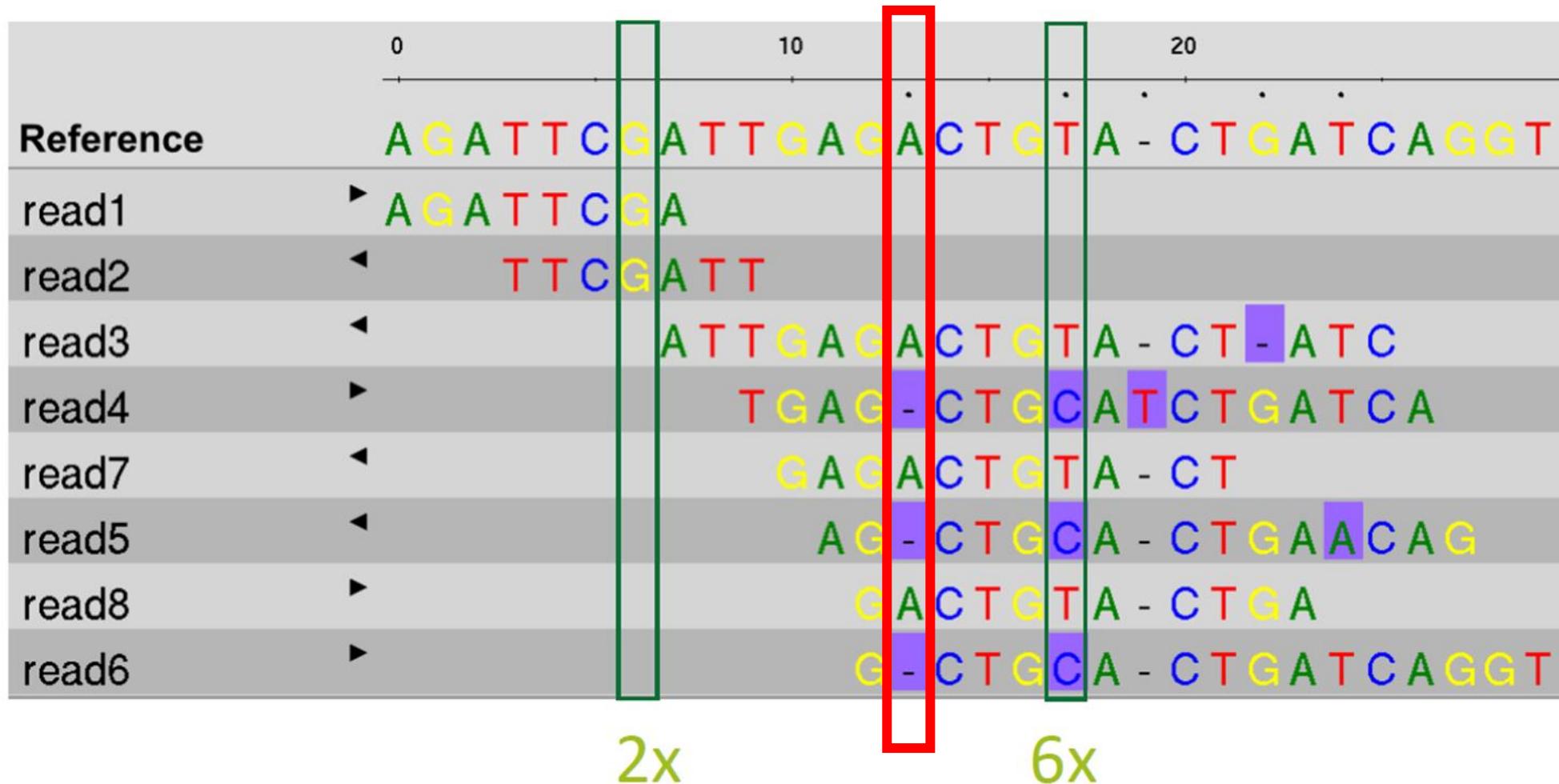
<https://www.genome.jp/entry/T01001:5053>

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

https://www.ncbi.nlm.nih.gov/clinvar?LinkName=gene_clinvar&from_uid=5053

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

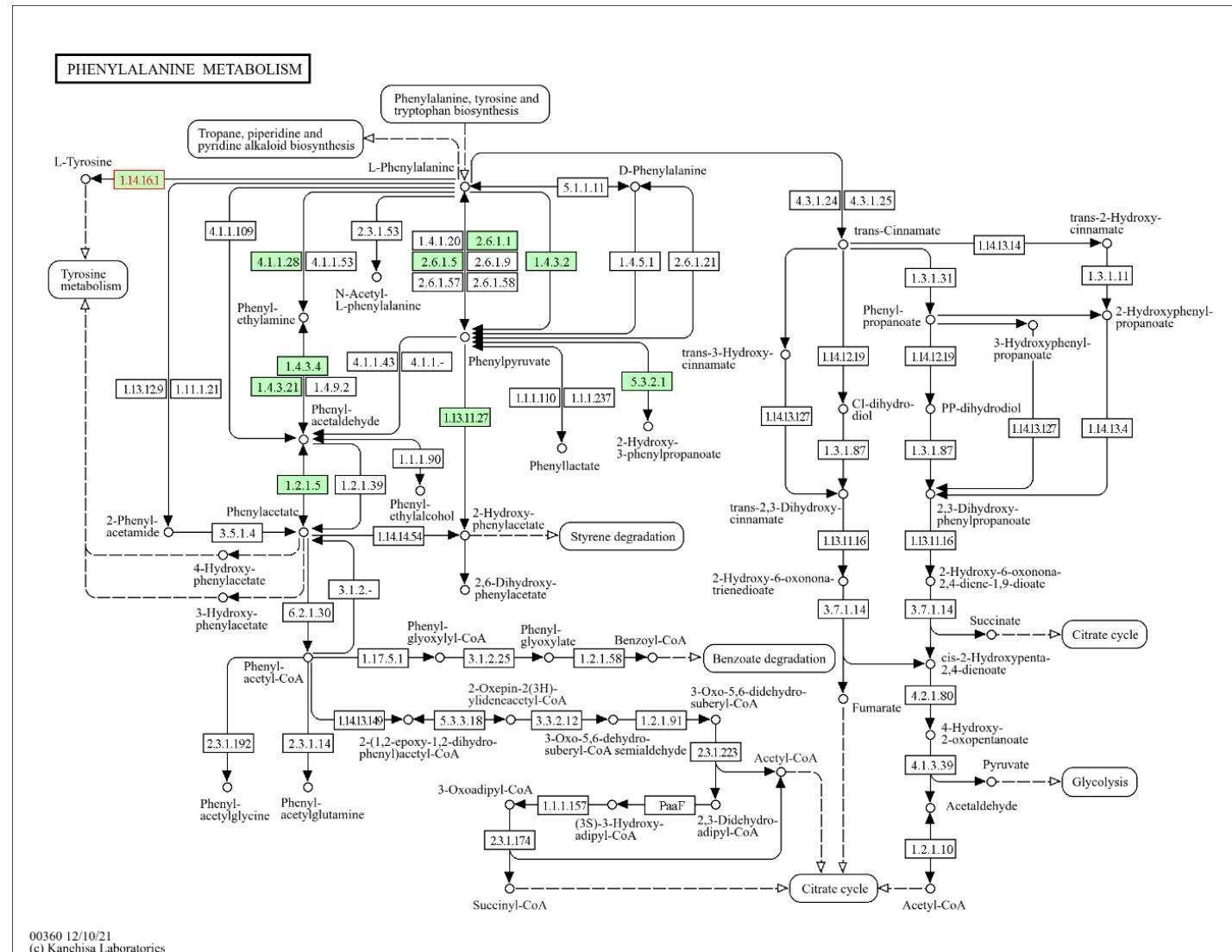


Show

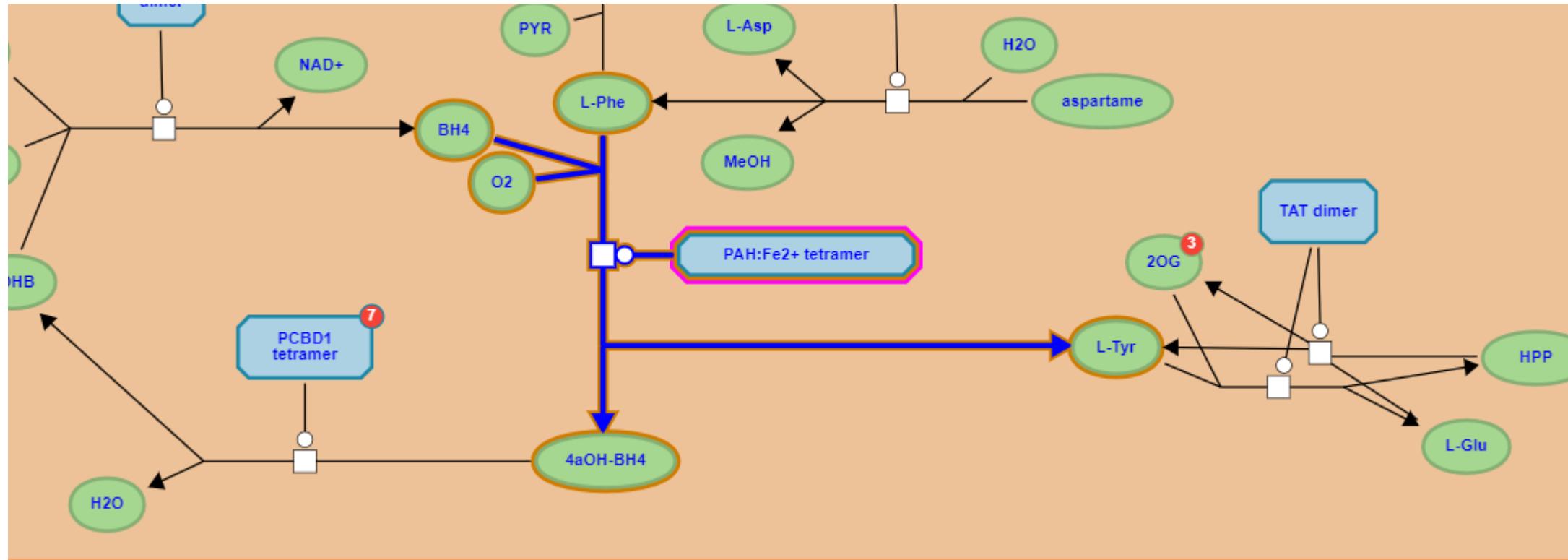
- All Residues
- Variants
- Pathogenic
- Likely Pathogenic
- Uncertain Significance
- Likely Benign
- Benign
- Current Variant

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

PAH: chuyển hóa Phenylalanine thành Tyrosine

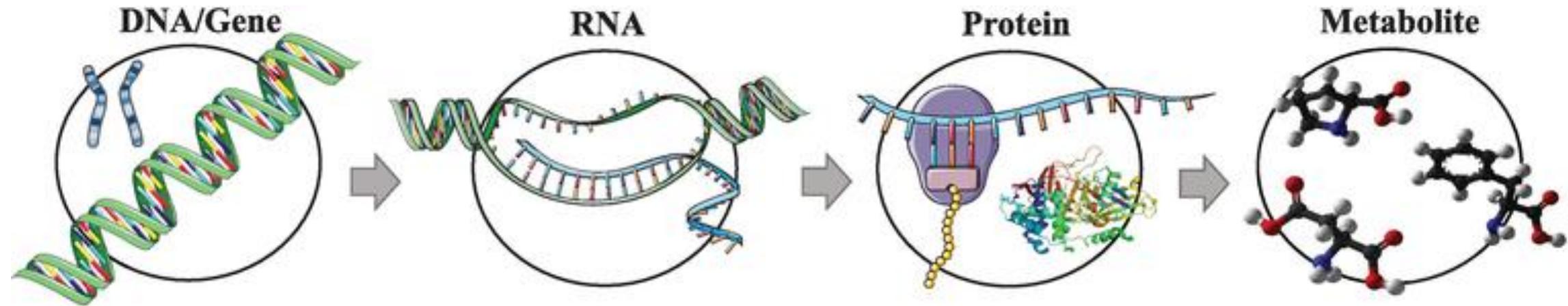


PAH: chuyển hóa Phenylalanine thành Tyrosine



<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

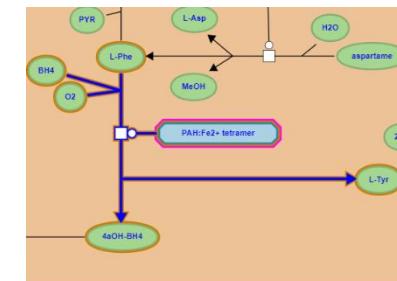
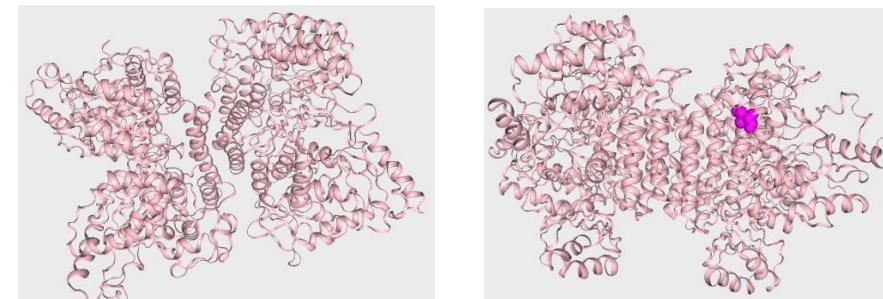
PAH protein
Ref ...Ile-Asp...
P1 ...Asn-Asp...

NM_000277.3(PAH):p.Ile324Asn

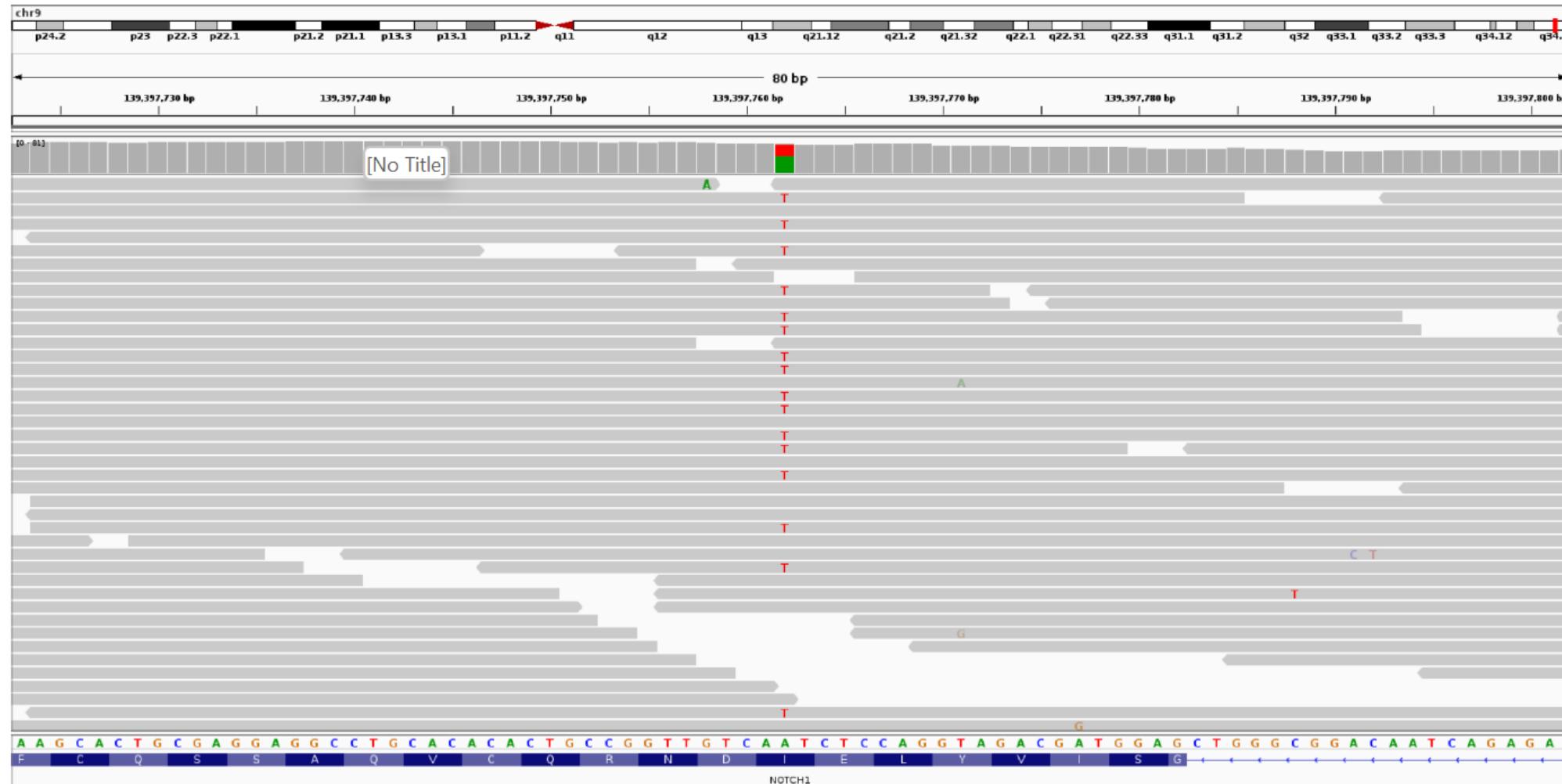
Metabolomics

PAH
Ref Phe → Tyr

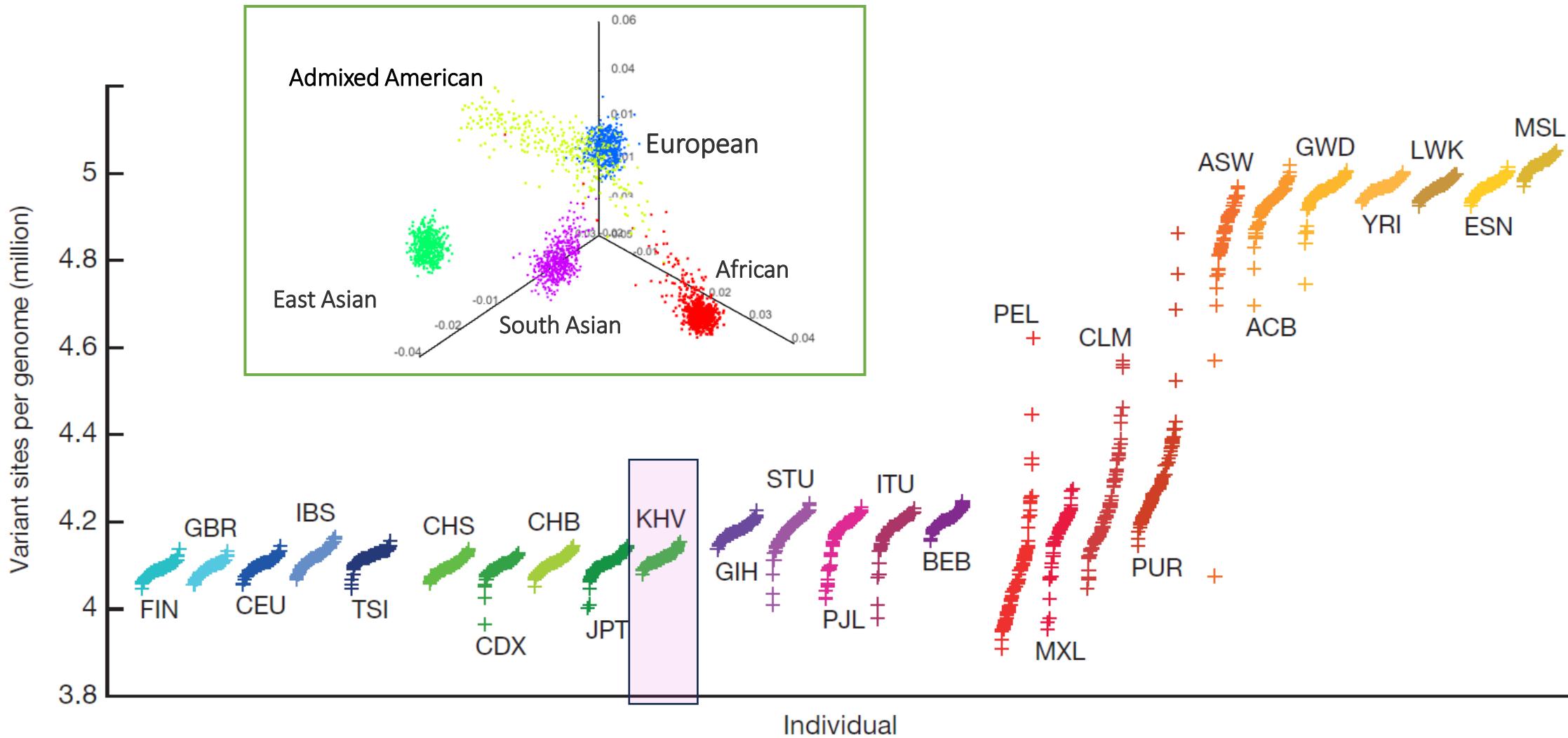
PAH
P1 Phe ~~→~~ Tyr



Alignment and variant viewers

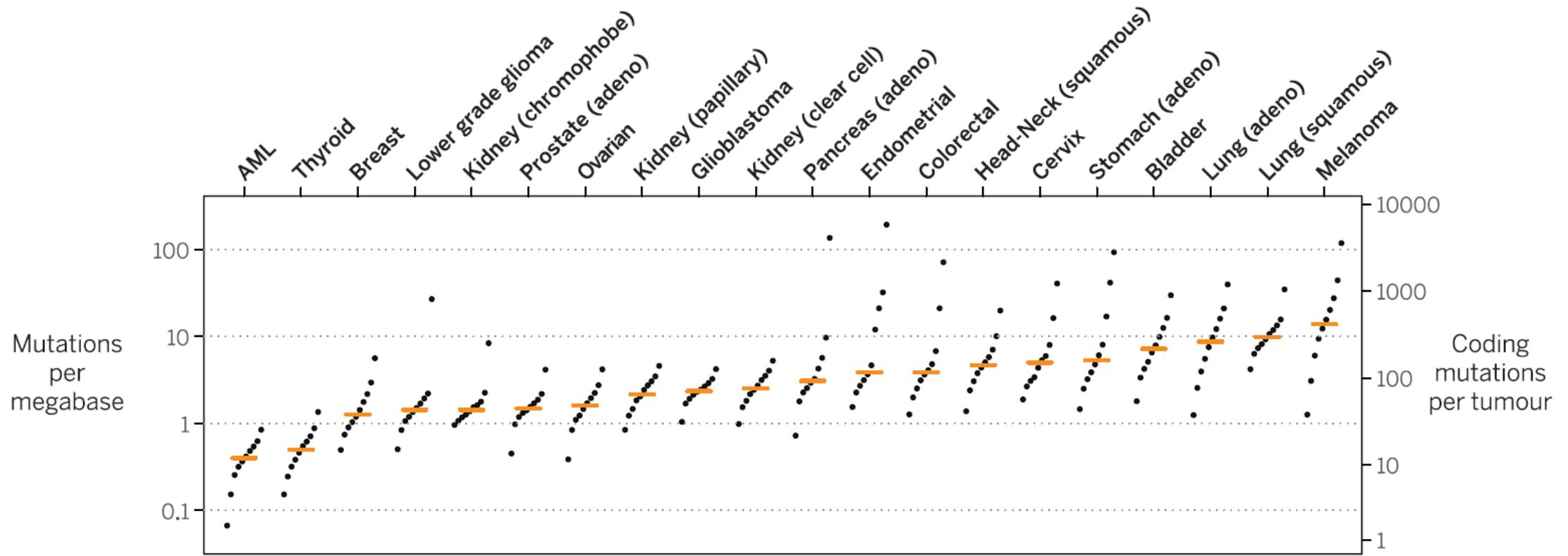


Human Genome Variation: 1000 Genomes Project



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

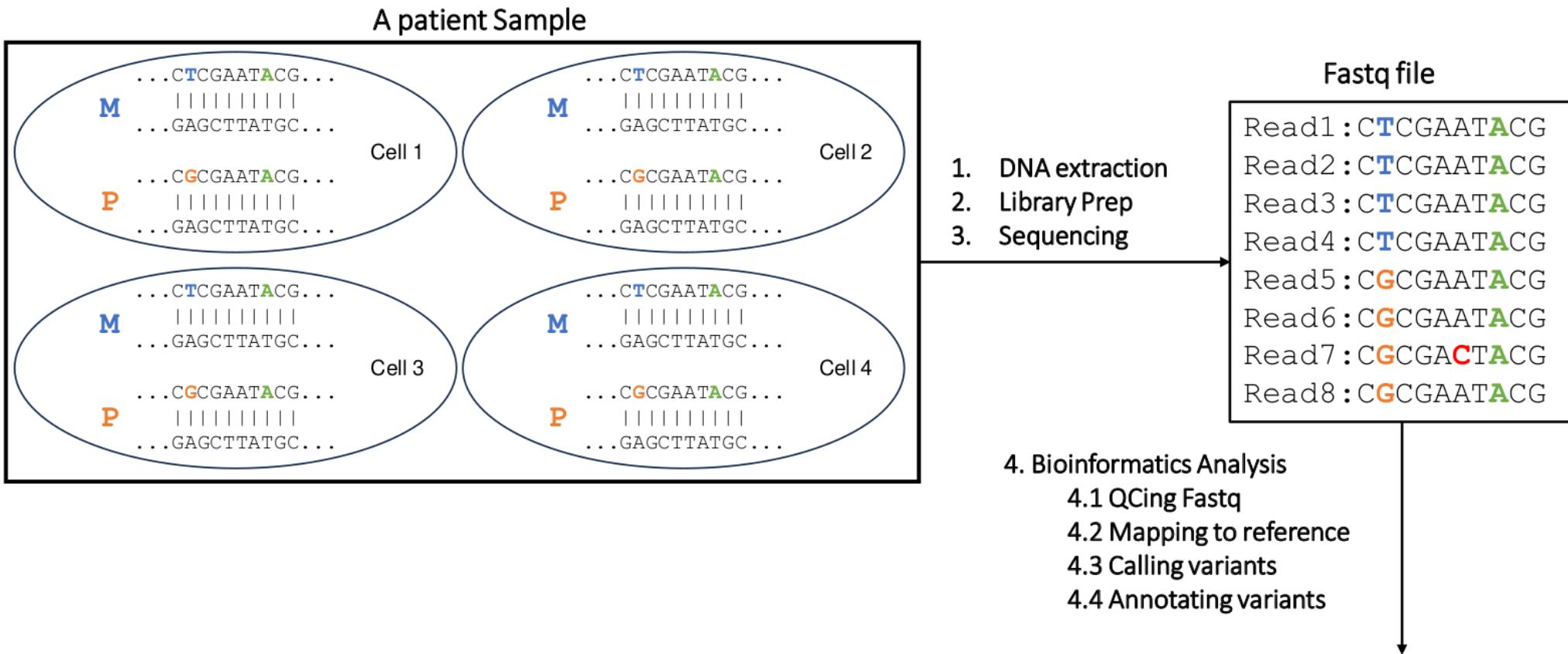
Cancer Genome Somatic Variation



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)

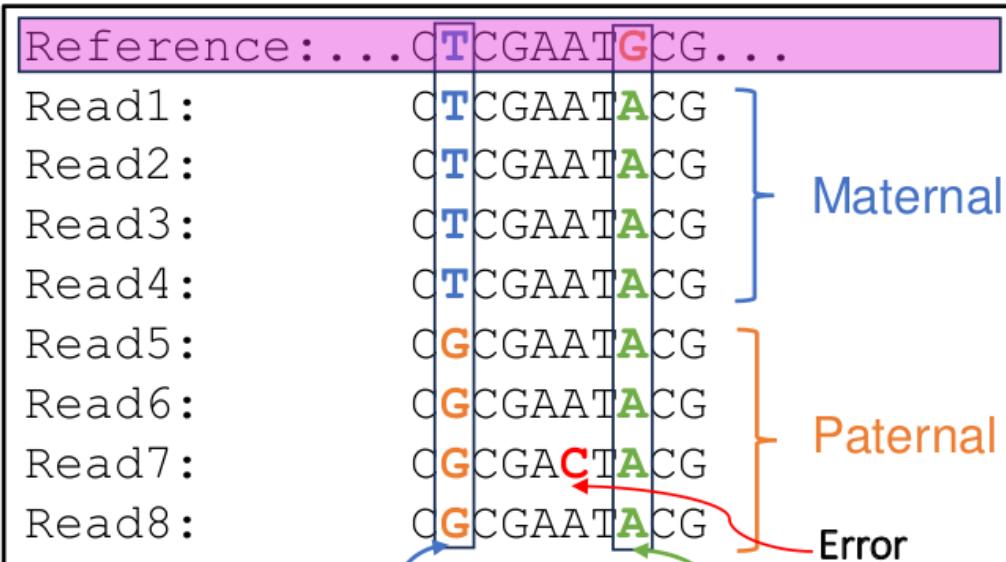
Quy trình XÉT NGHIỆM gen bằng phương pháp giải trình tự thế hệ mới (NGS)

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



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	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

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)

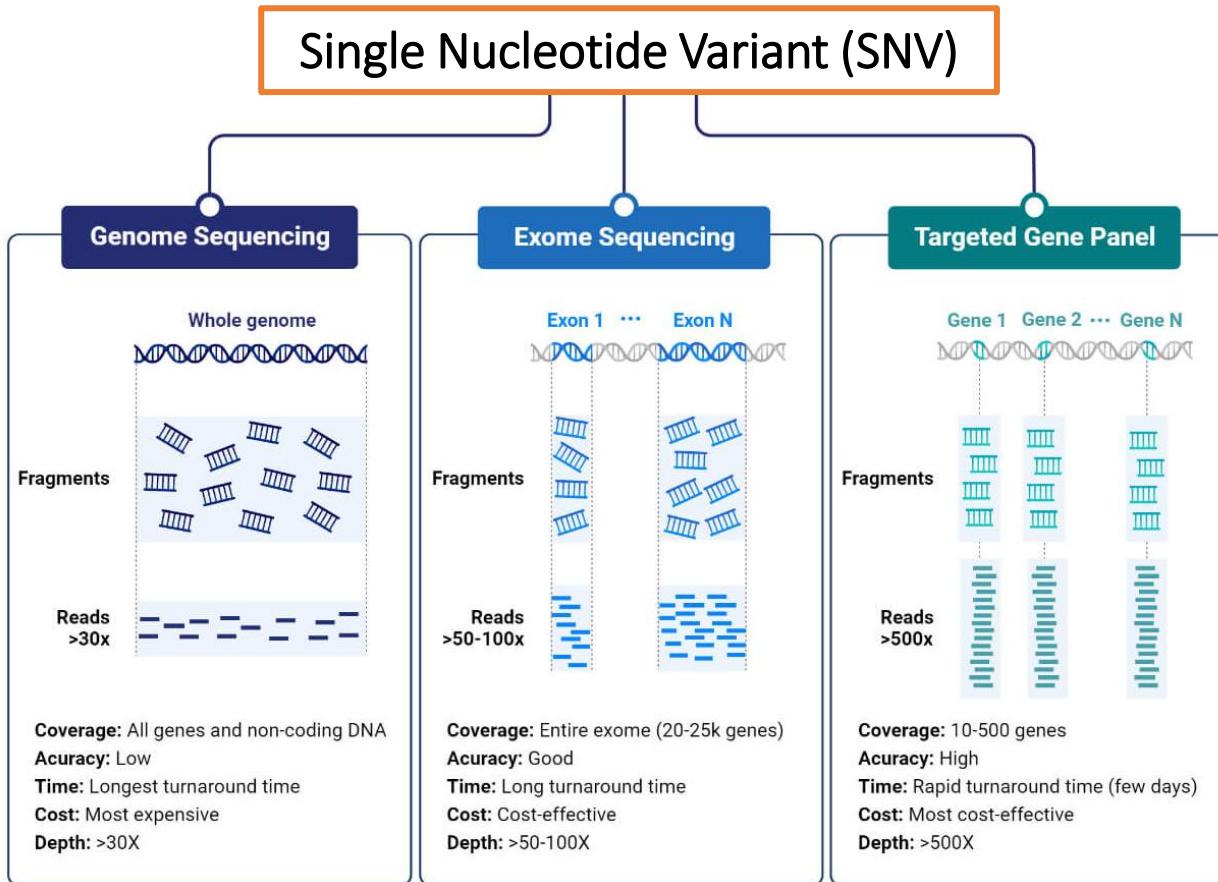
4.4 Annotating variants

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

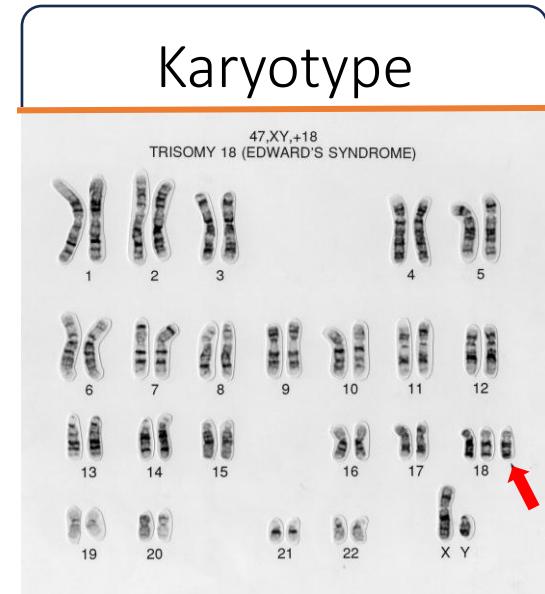
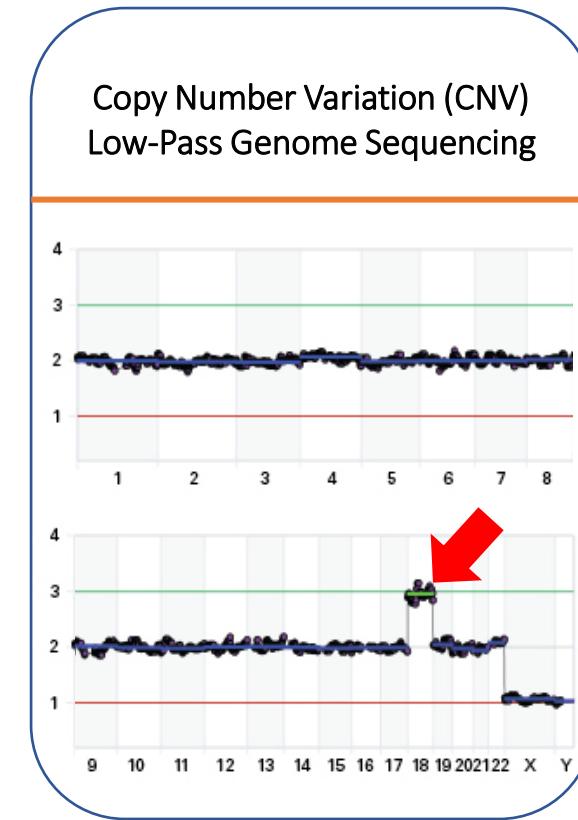


<https://ggb.a.swiss/en/the-first-medical-genomics-center-opens-in-geneva/>

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

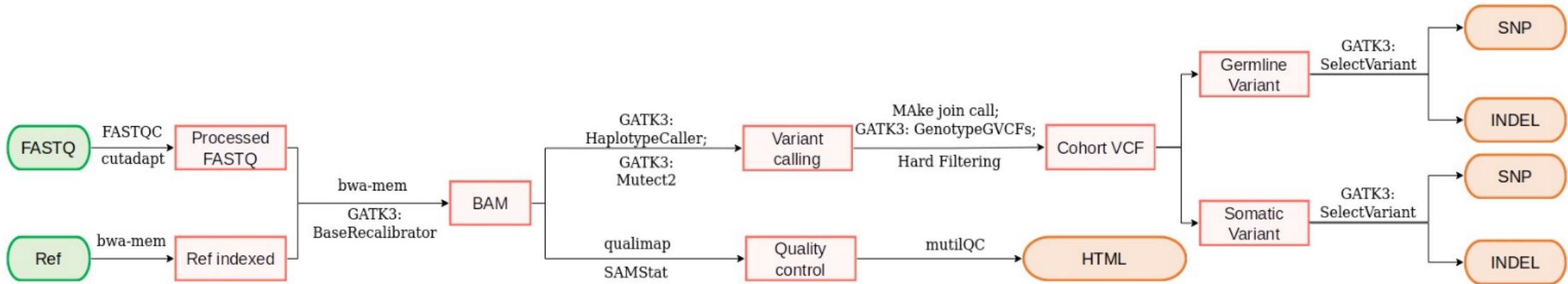


<https://microbenotes.com/next-generation-sequencing-ngs/>



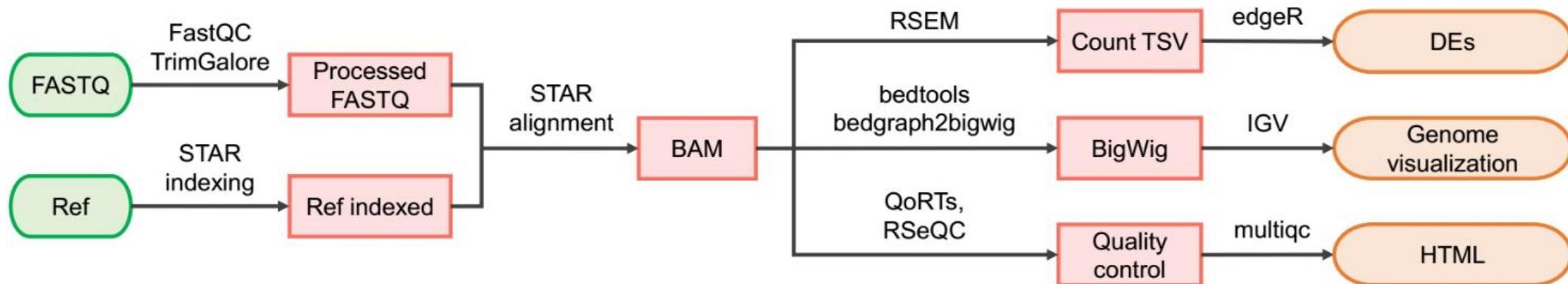
<https://wellcomecollection.org/search/images?query=eaahzt2u#>

Whole Genome Sequencing pipeline



Package	Version	Link
FastQC	0.11.9	https://www.bioinformatics.babraham.ac.uk/projects/fastqc/
cuadap	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/

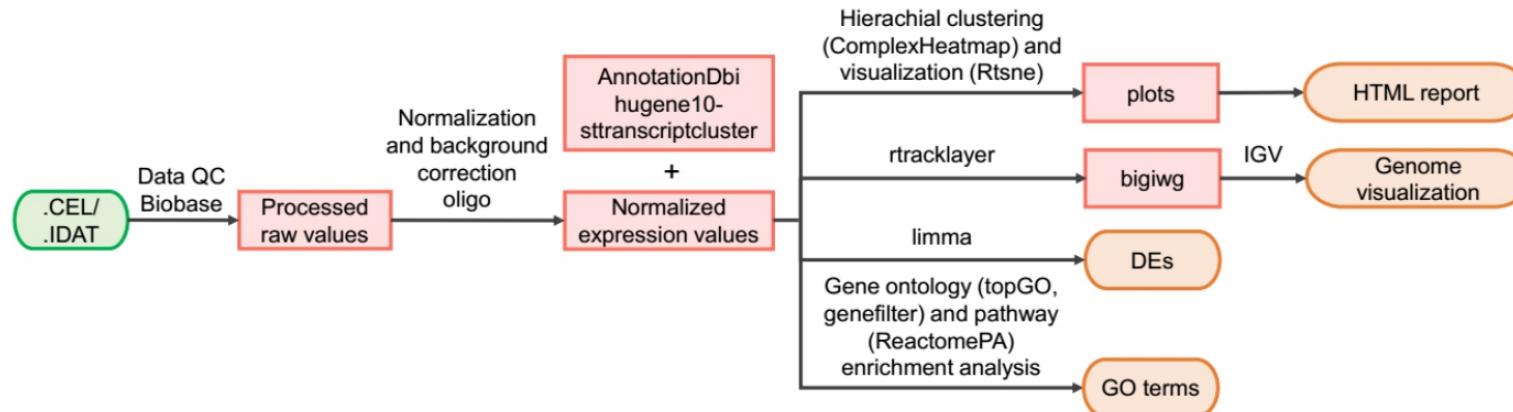
Bulk RNA-seq pipeline



Package	Version	Link
FastQC	0.11.9	https://www.bioinformatics.babraham.ac.uk/projects/fastqc/
TrimGalore	0.6.5	https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/
STAR	2.7.3a	https://github.com/alexdobin/STAR
RSEM	1.3.3	https://deweylab.github.io/RSEM/
edgeR	3.28.1	http://bioconductor.org/packages/release/bioc/html/edgeR.html
QoRTs	1.3.6	https://hartleys.github.io/QoRTs/
RSeQC	2.6.4	https://pythonhosted.org/RSeQC/
multiqc	1.8	https://multiqc.info/

Figure 6. The RNA-seq pipeline and software.

Gene Expression Microarray (Affymetrix) pipeline

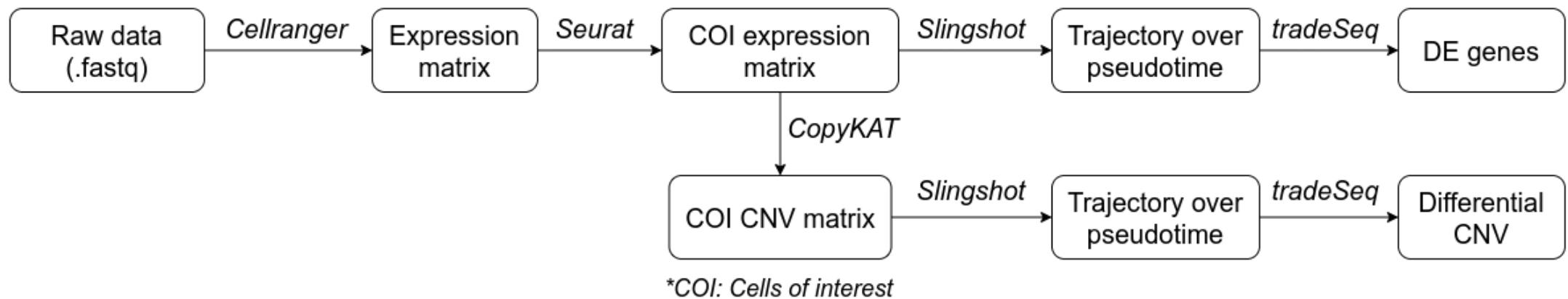


Package	Version	Link
Biobase	2.48.0	https://www.bioconductor.org/packages/release/bioc/html/Biobase.html
oligo	1.52.1	https://www.bioconductor.org/packages/release/bioc/html/oligo.html
AnnotationDbi	1.50.3	https://www.bioconductor.org/packages/release/bioc/html/AnnotationDbi.html
hugene10sttranscriptcluster	8.7.0	https://bioconductor.org/packages/release/data/annotation/html/hugene10sttranscriptcluster.db.html
Complex-Heatmap	2.4.3	https://www.bioconductor.org/packages/release/bioc/html/ComplexHeatmap.html
Rtsne	0.15	https://cran.r-project.org/web/packages/Rtsne/index.html
rtracklayer	1.48.0	https://bioconductor.org/packages/release/bioc/html/rtracklayer.html
limma	3.44.3	https://bioconductor.org/packages/release/bioc/html/limma.html
topGO	2.40.0	https://bioconductor.org/packages/release/bioc/html/topGO.html
genefilter	1.70.0	https://bioconductor.org/packages/release/bioc/html/genefilter.html
ReactomePA	1.32.0	http://bioconductor.org/packages/release/bioc/html/ReactomePA.html

Figure 7. The Expression array (Affymetrix) pipeline and software.

scRNA-seq

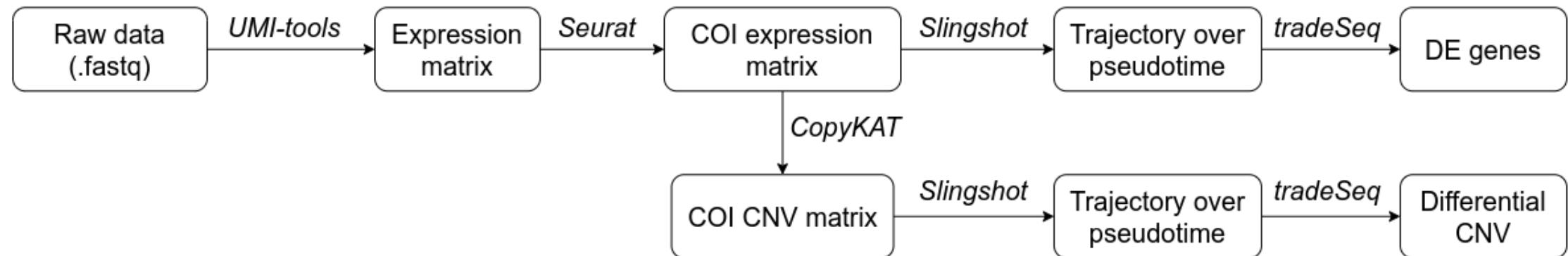
10X Chromium pipeline



Package	Version	Link
Cellranger	6.1.2	https://support.10xgenomics.com/single-cell-gene-expression/software/pipelines/latest/using/tutorials
Seurat	4.1.1	https://satijalab.org/seurat/articles/get_started.html
CopyKAT	1.0.5	https://github.com/navinlabcode/copykat
Slingshot	2.0.0	https://bioconductor.org/packages/devel/bioc/vignettes/slingshot/inst/doc/vignette.html
tradeSeq	1.6.0	https://bioconductor.org/packages/devel/bioc/vignettes/tradeSeq/inst/doc/tradeSeq.html

scRNA-seq

SMART-Seq2 pipeline



*COI: Cells of interest

Package	Version	Link
UMI-tools	1.1.2	https://umi-tools.readthedocs.io/en/latest/Single_cell_tutorial.html
Seurat	4.1.1	https://satijalab.org/seurat/articles/get_started.html
CopyKAT	1.0.5	https://github.com/navinlabcode/copykat
Slingshot	2.0.0	https://bioconductor.org/packages/devel/bioc/vignettes/slingshot/inst/doc/vignette.html
tradeSeq	1.6.0	https://bioconductor.org/packages/devel/bioc/vignettes/tradeSeq/inst/doc/tradeSeq.html

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

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

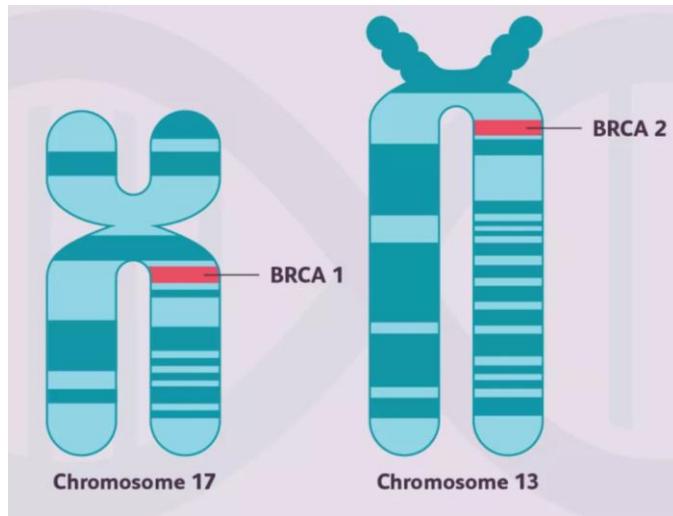
Ví dụ 2: Sàng lọc bệnh ung thư bằng xét nghiệm gen Gene Panel – 53 genes của MGI

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

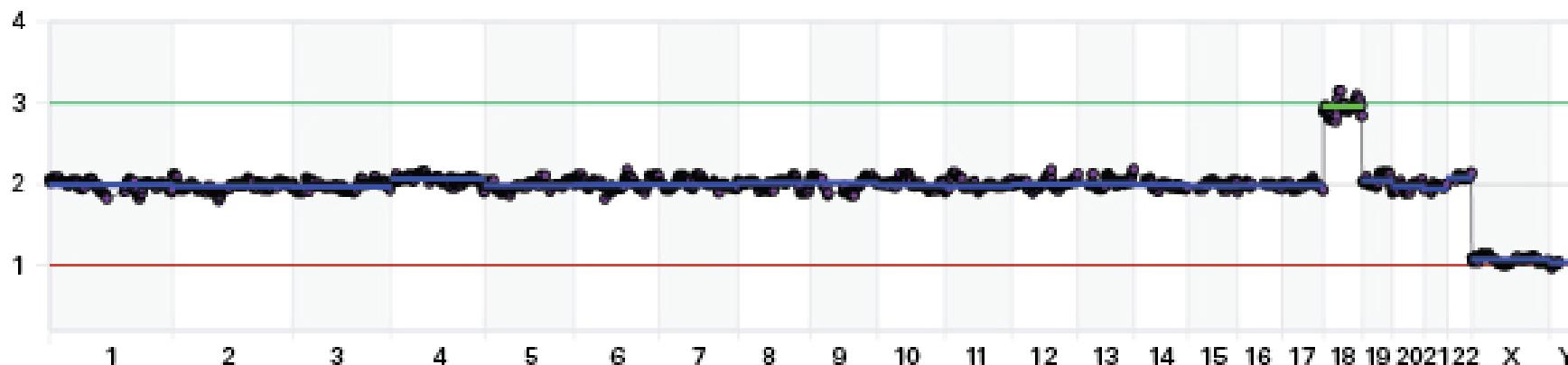
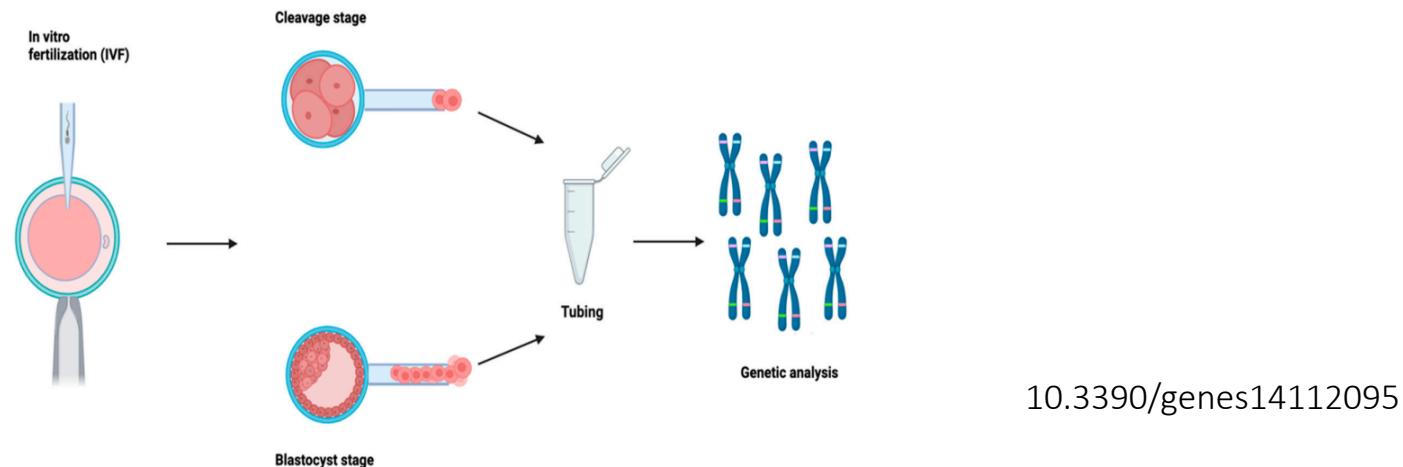
- Hệ thống xét nghiệm Illumina MiSeqDx (CE/US-IVD)
- Bộ xét nghiệm: NGeneBio BRCAaceuTest™Plus (CE-IVD)
- Phần mềm hỗ trợ phân tích kết quả: NGeneBio NGeneAnalySys™ (CE-IVD)

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

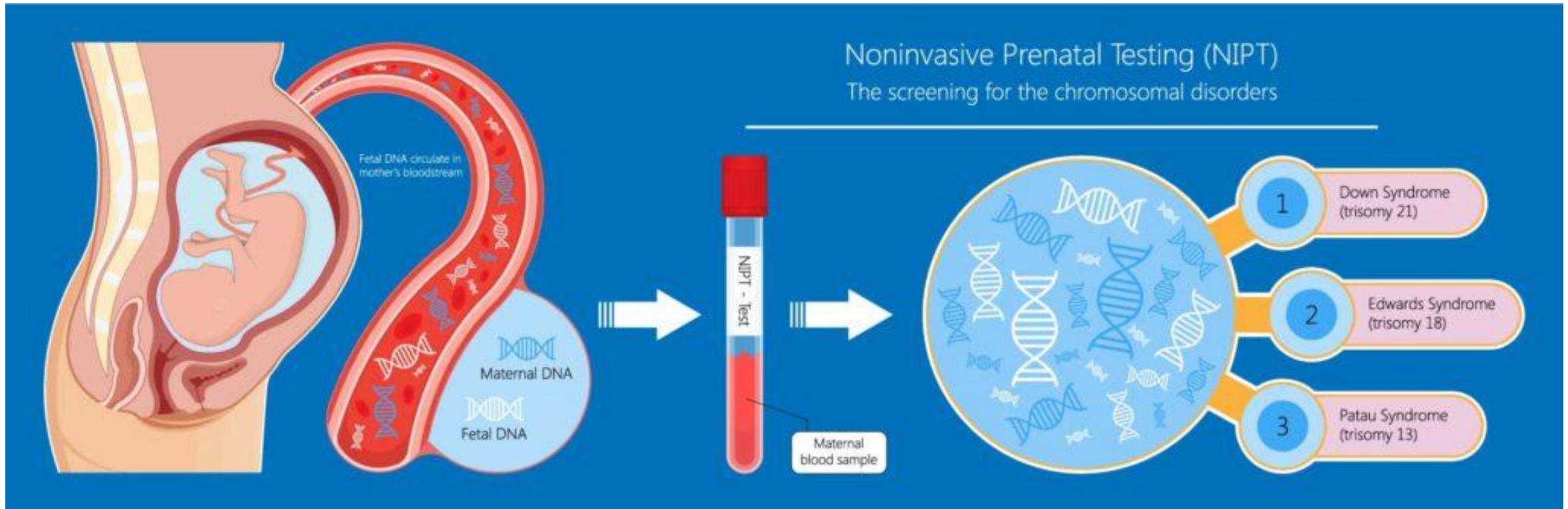
Ví dụ 4: Sàng lọc phôi trong hỗ trợ sinh sản IVF

Xét nghiệm tiền làm tổ PGT-A



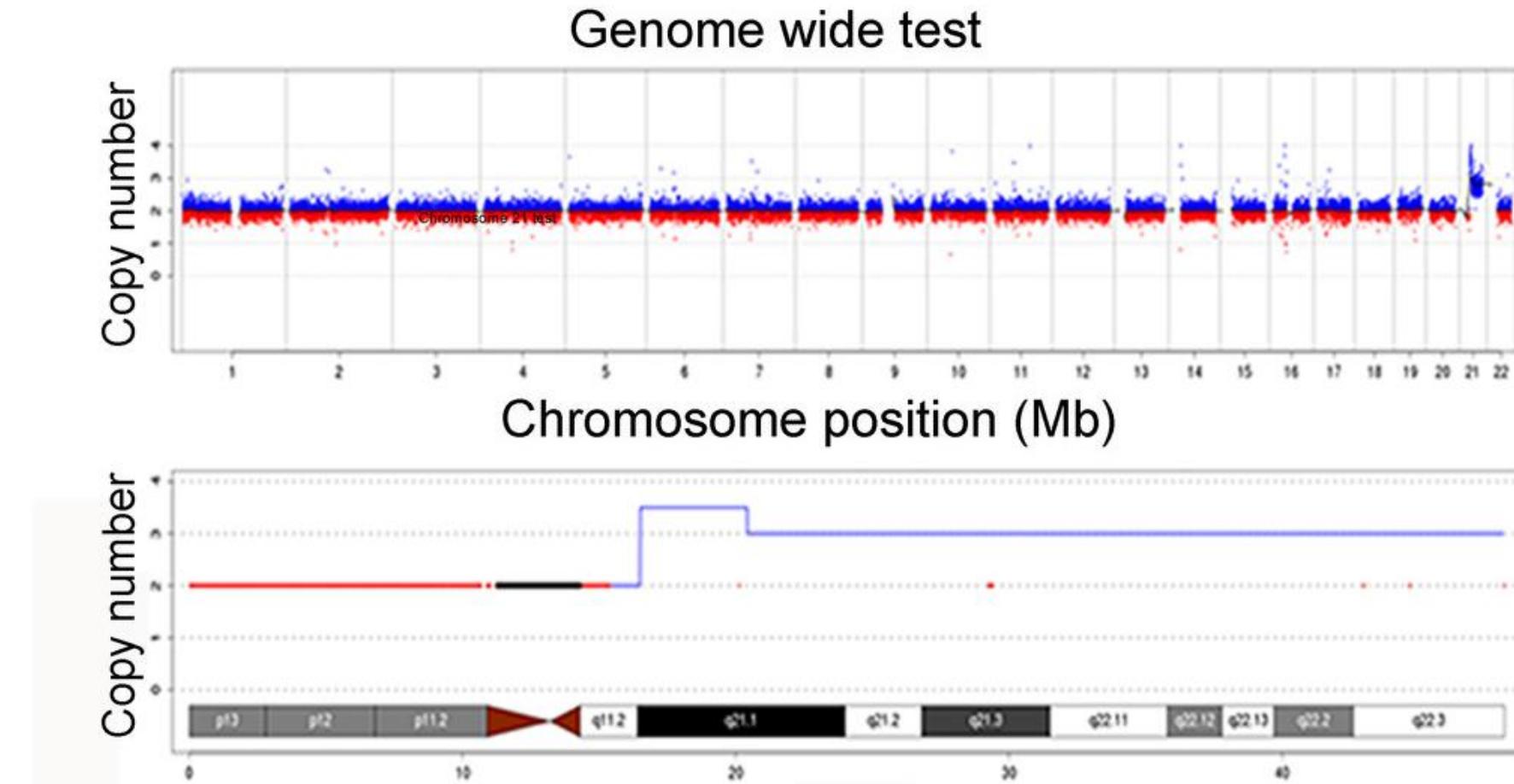
Mẫu phôi có 3 nhiễm sắc thể 18 (trisomy) trong DNA hệ gen

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 ^{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

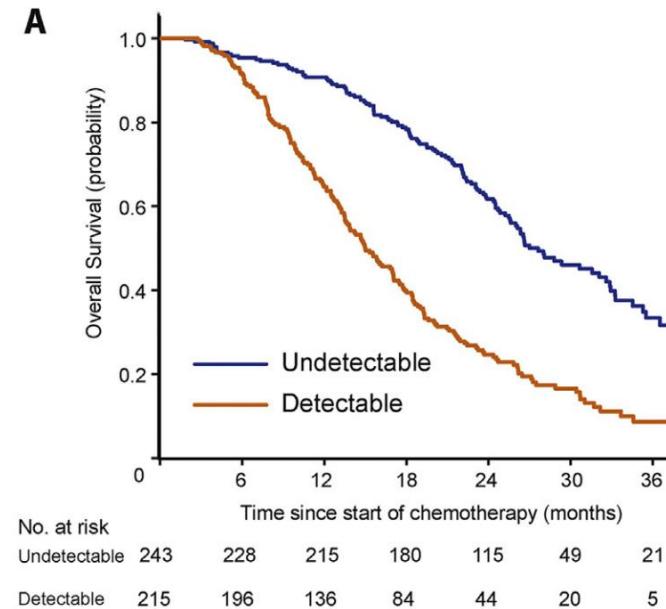
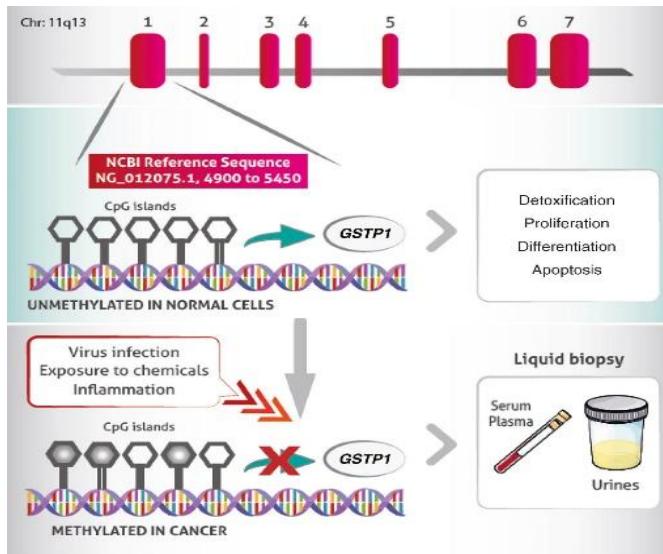
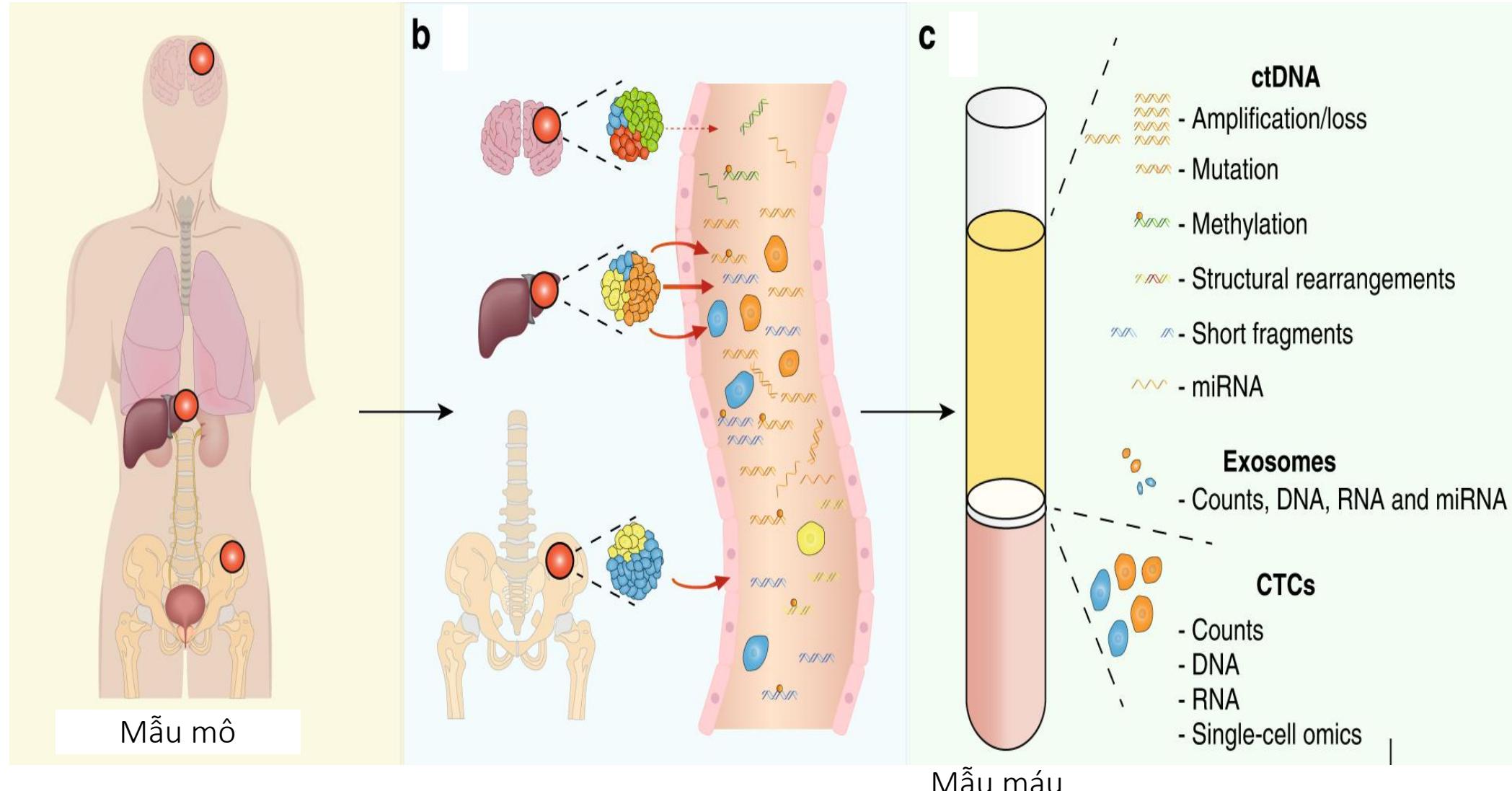
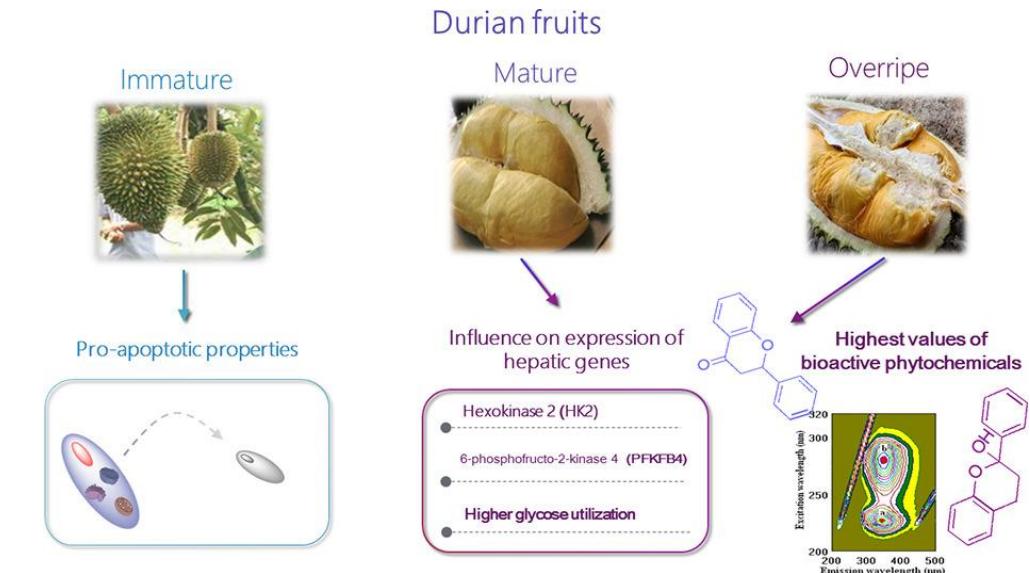
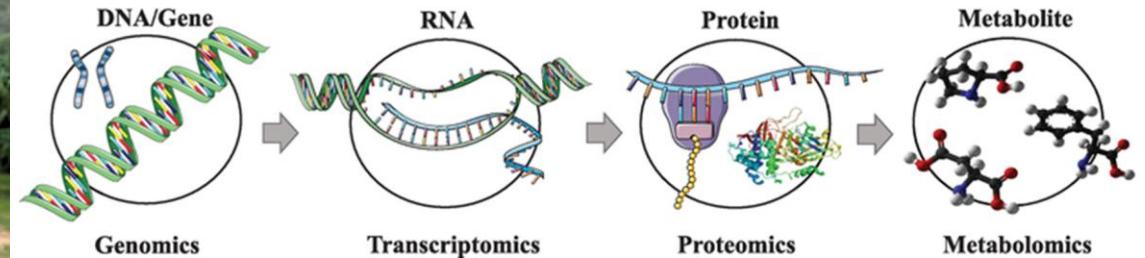


Fig. Kaplan-Meier estimates of survival in patients with a detectable mGSTP1 at baseline. (A) OS according to mGSTP1 detection after two cycles of docetaxel.

Mẫu cho XÉT NGHIỆM gen: mẫu mô và máu

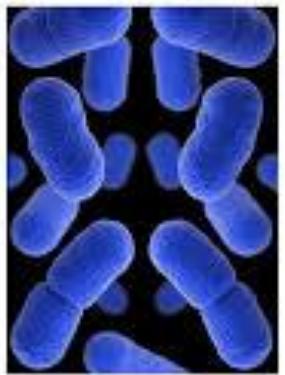


Ứng dụng kỹ thuật giải trình tự thê hệ mới trong nông nghiệp và thủy sản

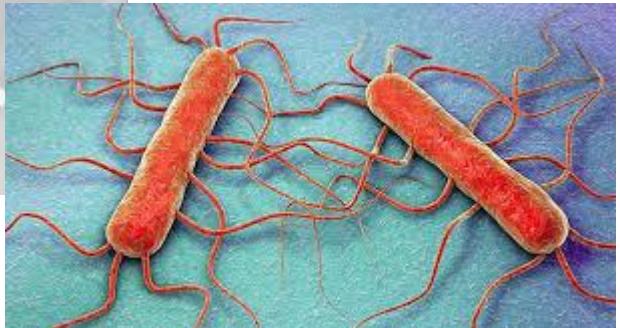


Listeria monocytogenes

microbeonline



Gram positive coccobacilli



Listeria monocytogenes EGD-e, complete genome

GenBank: AL591824.1

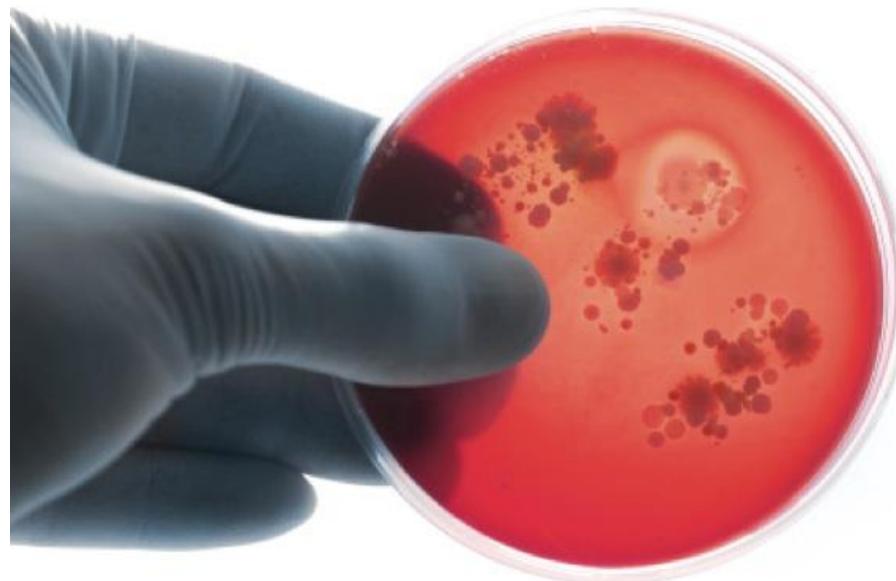
[GenBank](#) [Graphics](#)

>AL591824.1 Listeria monocytogenes EGD-e, complete genome

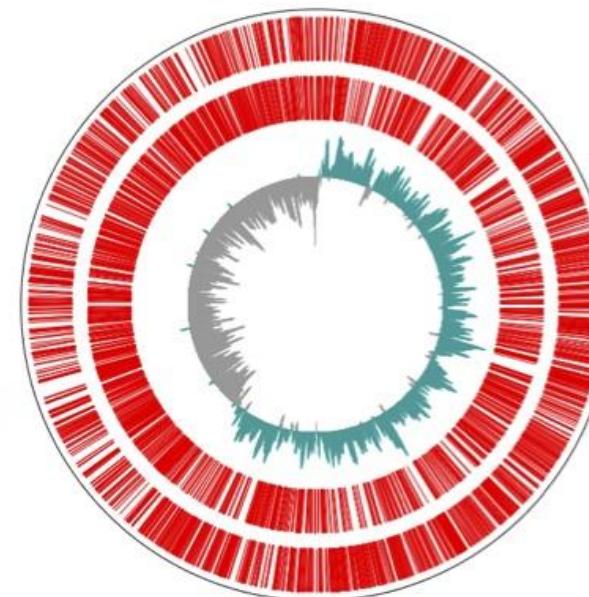
```
TAATTAATCTAACAAATTCTGTTACAGATTCTTACACACAAGTTACACACAAGTTAACGGCTGTGGA  
CAACCGTTTTCACATCTGGACAGTTTGTTGGATAGATTGGTAAGTCCTGCTATCAGAGTGTATTCT  
GATATTATAATTCCCTGTCGAATAGAAAATAGCTGGGAAAACTAAAGTTATCCACAATACATTTTACT  
TTGTGGATAATTTTAACAGTGTGGATAACCTTATCCATAGCTTTCTATCTGTGGATAACTTTAT  
AGCATCCATTACATTACATAAAAAGGGGGGTACTAGTGCAATCAATTGAAGACATCTGGCAGGAAACA  
CTGCAAATTGTTAAAAAAATATGAGTAAACCTAGTTACGATACATGGATGAAATCAACAAACCGCTCATT  
CACTTGAAGGTAACACGTTTATTATTCAGCGCCAAATAATTGTTGTCGCAATTGGTTAGAGAAAAGCTA  
CACTCAATTATCGCTAACATTGCAAGAAAATACTGGTCGTTATTGATGTCGCTTATTGATGGC  
GAGCAGGAAGAAAACTTGAATACACTGTGATTAAACCAATCCAGCATTAGATGAAGATGGCATTGAAA  
TTGGAAAACATATGCTTAATCCACGTTATGTTTGTACCTTGTCAATTGGTCAGGGAAACAGATTGC  
CCACGCAGCATCACTGCACTAGCCGAAGCACCAGCGAAAGCATATAATCCACTCTTCAATTGAGGAA  
GTTGGCCTCGGTTAAACACATTAAATGCAAGCAGTTGCCACTATGTTCAACAAACATAAAAGATAATGCGA  
AAGTAATGTACCTTCCAGCGAAAATTCCAATGAGTTATTAGCTTATTCGTGATAATAAAACCGA  
AGAATTGCAACAAATATCGGAATGTTGATGCTTACTTATTGATGATATTCAATTAGCCGGTAAA  
GAAGGAACACAAGAGGAATTTCACATTTAACACACTTTATGATGAACAAAAGCAAATTATT  
CCAGTGACCGGACCAACAAAAGAAATTCCACACTGGAAAGATCGACTGAGATCCGCTTGAATGGGCTT  
AATTACTGATATTAGCCACCAAGACTTAGAAACCCGGATGCCATTACGTAACAAAGCAGAC  
GGATTAGATATTCCAATGAAGTTATGCTTATATCGCAAACCAAATTGATTCGAATATTGCGAGCTAG  
AAGGGCTCTATCGAGTAGTTGCTTATTCTCCCTGTTAATAAGATATAACAGCTGGCTTGAGC  
AGAAGCACTAAAGATATTATCCCCTTCTAAATCACAAGTTACTATTAGTGGTATTCAAGAAGCA  
GTCGGTGAATATTCCACGTTGTTAGAAGATTAAAGCAAAAAACGGACGAAAGTATAGCATTCC  
CGCGCCAAATCGCCATGTATCTCAAGAGAGCTTACAGATGCCCTATTACCAAAATCGGTGATGAATT  
TGGTGGTCGAGATCATACAACAGTTATTGATGCACTGAAAAAAATATCGCAACTACTAAAAACCGATCAA  
GTGTTGAAAATGACCTGCCGAAATTGAAAAAAATTAAAGAAAAGCACAAAAATATGTTAATAGACCT  
GTGTACAATGTGGATAACTGAAACATACTTACCCACAAGTTACACATGTGGAAAACCTTATGCAGCAT  
GGCTTGTAAACCTACTTATCCACAAATCCACAGCGCTATTACTATTACTACGATTTTTATTAATTAAATT
```

<https://www.ncbi.nlm.nih.gov/nuccore/AL591824.1?report=fasta>

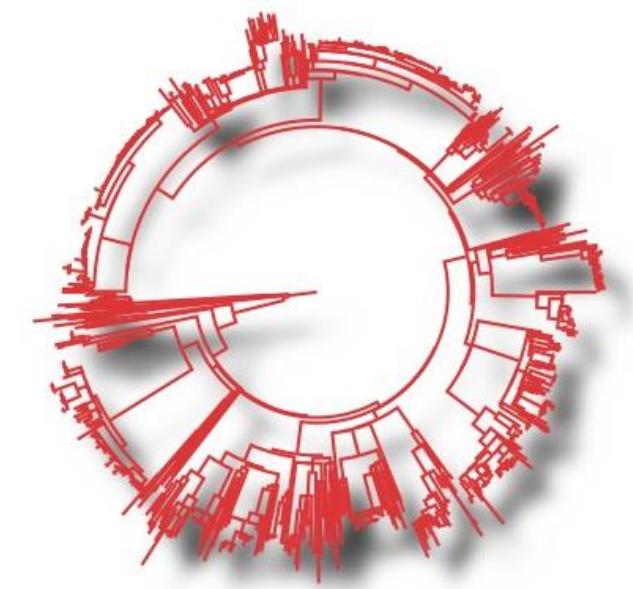
Microbial Genome Analysis: study one isolate at once



Culture microbe



Genome Sequencing



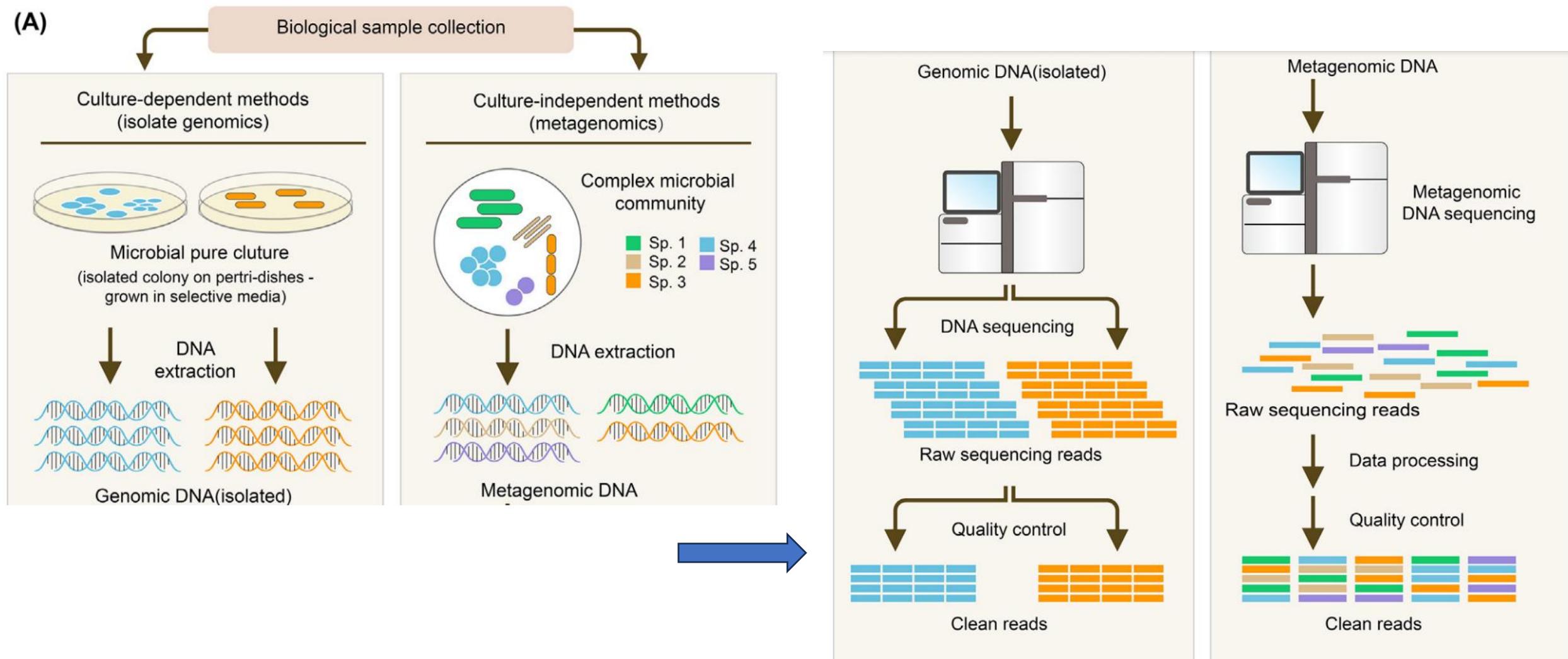
Serotyping/genotyping

Microbe = Microorganism

Prokaryotes = Bacteria & Archaea

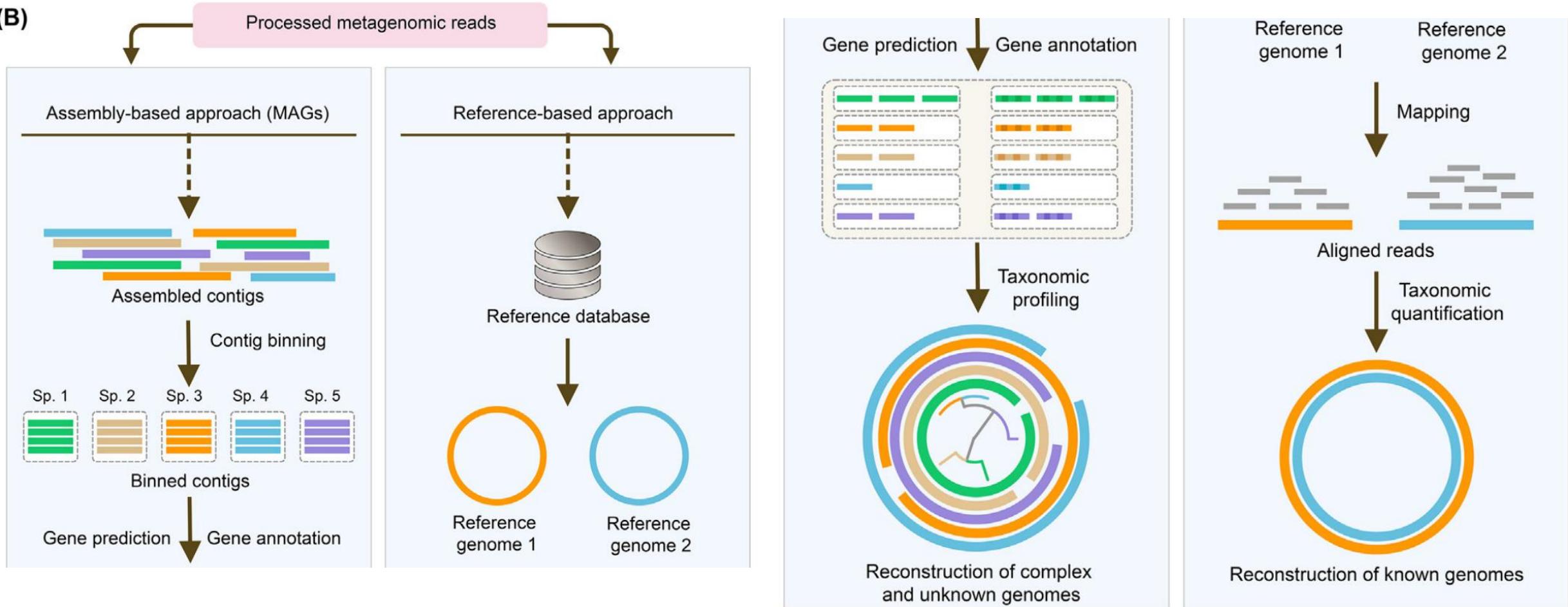
Microbial Genome vs Metagenome Analysis

(A)

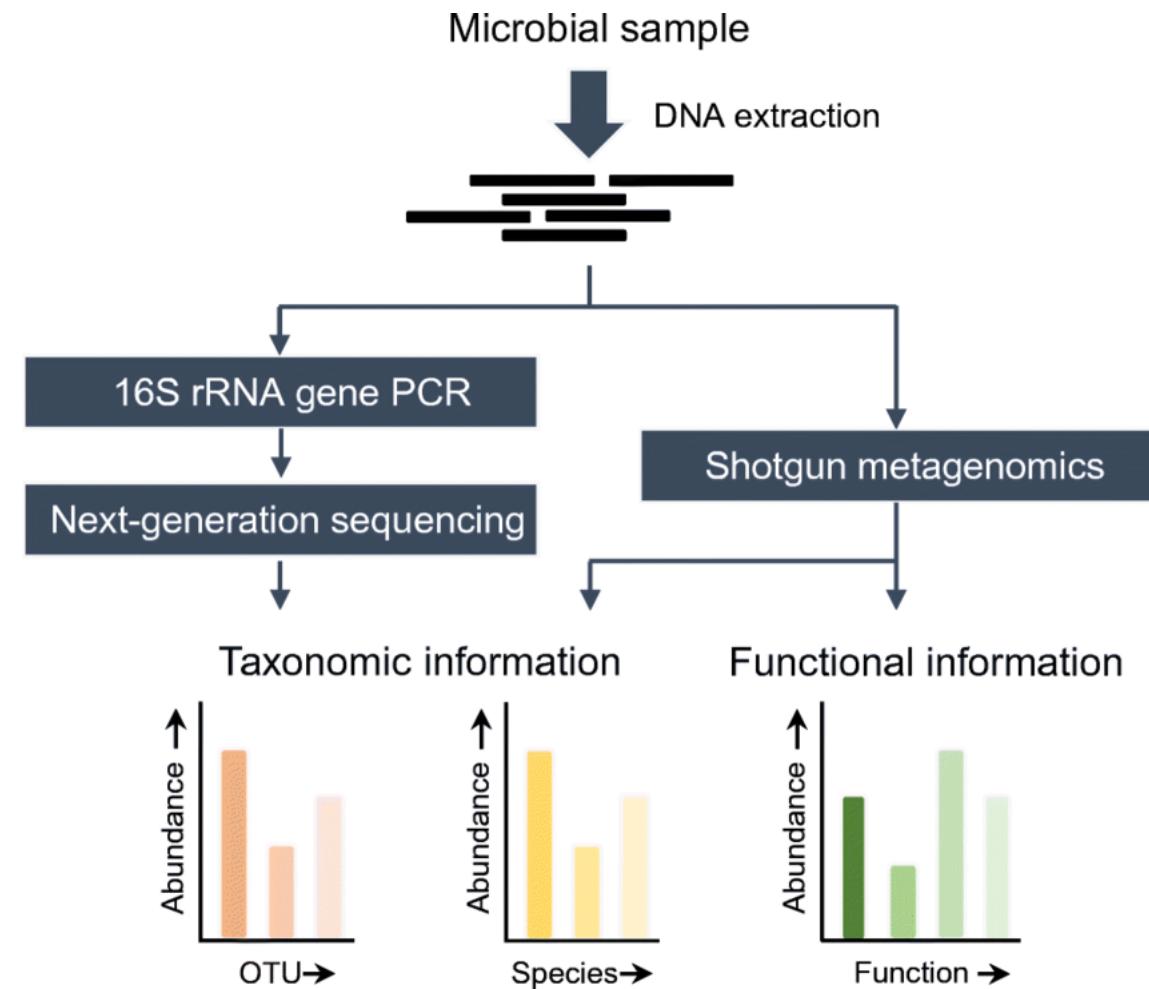
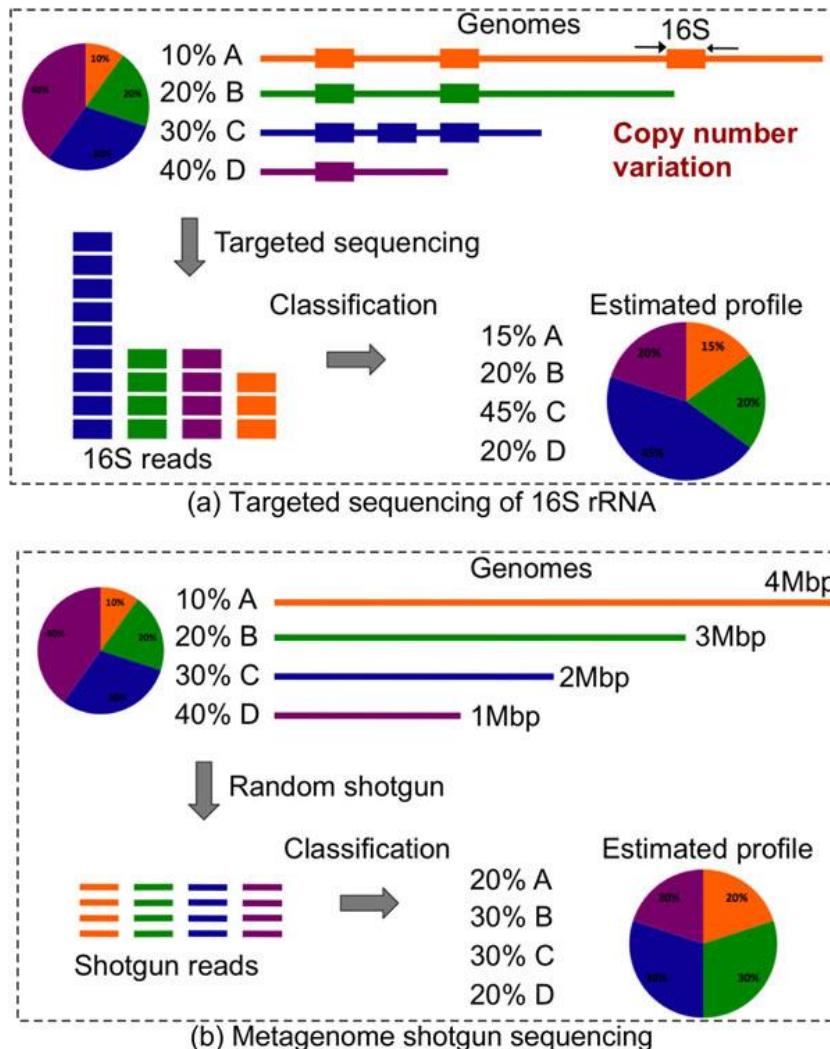


Shotgun Metagenome Analysis

(B)



Target vs Shotgun Microbiome Analysis



Precision microbiome testing: STI, HPV and AMR

Women's Health Test

Page 1 of 4

Patient Name:	nan	Provider:	Mony Sary, Yap Chew, Wendy Ullmer	Patient ID:	VS8
Gender:	nan	Provider NPI:	nan	Specimen ID:	KH20.45474
DOB:	nan	Order Date:	nan	Specimen Type:	nan
Age:	nan	Health Status Reported:	nan	Collection Date:	nan

Sexually Transmitted Infections

Name	Associated Condition	Result
<i>Neisseria gonorrhoeae</i>	Gonorrhea, urethritis, pelvic inflammatory disease, gonococcemia, gonococcal ophthalmia neonatorum	Detected
<i>Chlamydia trachomatis</i>	Chlamydia, cervicitis, urethritis, pelvic inflammatory disease	Not Detected
<i>Mycoplasma genitalium</i>	Urethritis, cervicitis, pelvic inflammatory disease	Not Detected
<i>Treponema pallidum</i>	Syphilis	Not Detected
<i>Haemophilus ducreyi</i>	Chancroid	Not Detected
<i>Trichomonas vaginalis</i>	Trichomoniasis	Not Detected
<i>Human papillomavirus</i>	Cervical and anogenital cancers, genital warts	Detected
<i>Herpes simplex virus</i>	Genital herpes, oral herpes	Not Detected

Viruses Detected

Name	Associated Condition
<i>Human papillomavirus 62 (HPV 62)</i>	Unknown risk for cervical cancer

Note: Human papillomavirus (HPV) 16, 18, 31, 33, 35, 39, 45, 51, 56, 58, 59, and 68 are considered high-risk or probable high-risk due to their association with cervical cancer. HPV 6, 11, 42, 43, and 44 are considered low-risk for cervical cancer but may cause genital warts. Other HPV genotypes found in the sample may have intermediate or unknown risk for cervical cancer.

Antimicrobial Resistance Genes Detected

AMR Gene Name	Function	Drug Class
<i>Neisseria.gonorrhoeae.folP</i>	Dihydropteroate synthase (mutated)	Sulfonamide

Women's Health Test

Page 1 of 4

Patient Name:	nan	Provider:	Yap Chew	Patient ID:	nan
Gender:	nan	Provider NPI:	nan	Specimen ID:	202122865
DOB:	nan	Order Date:	nan	Specimen Type:	nan
Age:	nan	Health Status Reported:	nan	Collection Date:	nan

Sexually Transmitted Infections

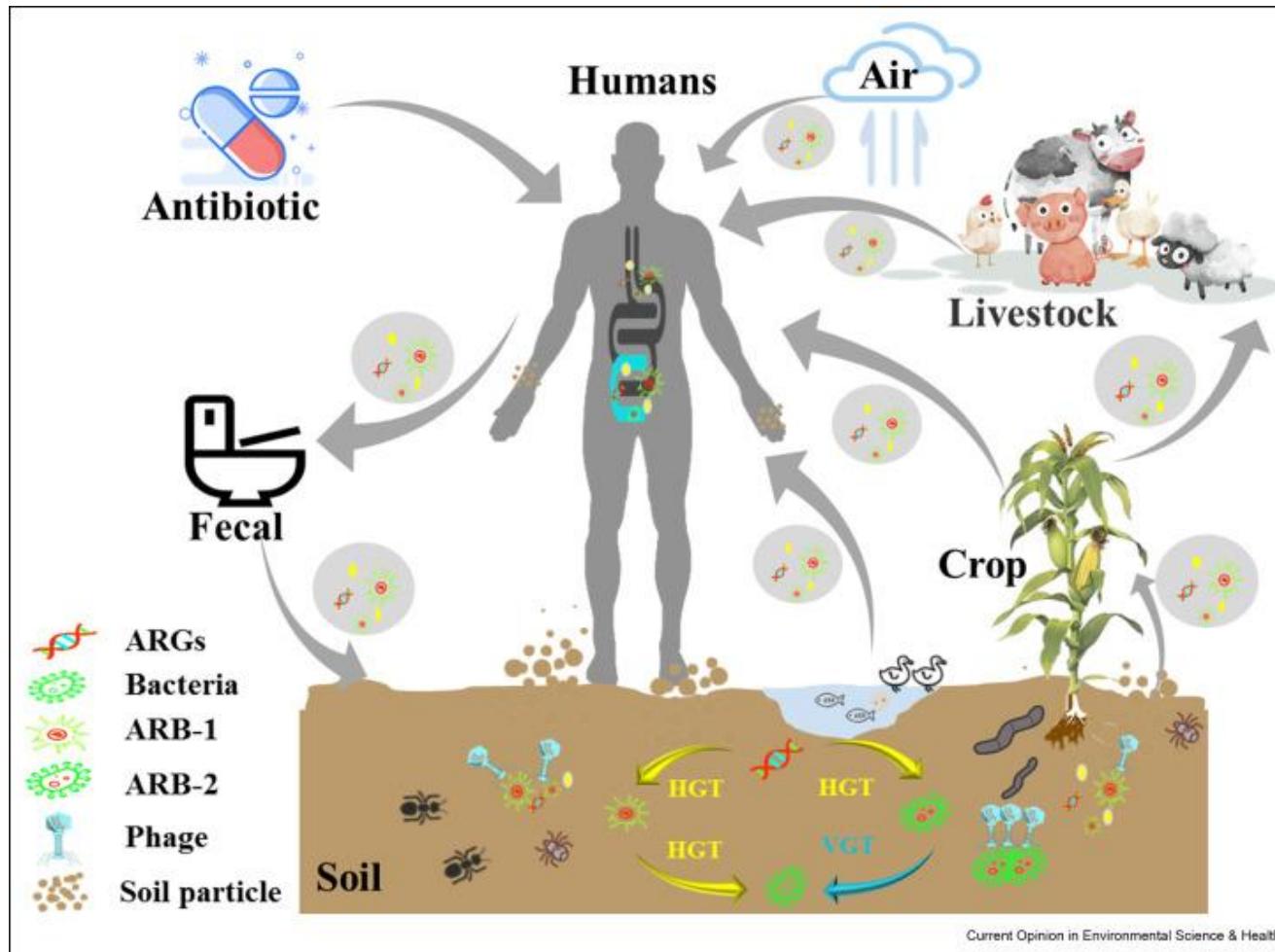
Name	Associated Condition	Result
<i>Neisseria gonorrhoeae</i>	Gonorrhea, urethritis, pelvic inflammatory disease, gonococcemia, gonococcal ophthalmia neonatorum	Not Detected
<i>Chlamydia trachomatis</i>	Chlamydia, cervicitis, urethritis, pelvic inflammatory disease	Not Detected
<i>Mycoplasma genitalium</i>	Urethritis, cervicitis, pelvic inflammatory disease	Not Detected
<i>Treponema pallidum</i>	Syphilis	Not Detected
<i>Haemophilus ducreyi</i>	Chancroid	Not Detected
<i>Trichomonas vaginalis</i>	Trichomoniasis	Not Detected
<i>Human papillomavirus</i>	Cervical and anogenital cancers, genital warts	Detected
<i>Herpes simplex virus</i>	Genital herpes, oral herpes	Not Detected

Viruses Detected

Name	Associated Condition
<i>Human papillomavirus 52 (HPV 52)</i>	High-risk for cervical cancer
<i>Human papillomavirus 68 (HPV 68)</i>	High-risk for cervical cancer

Note: Human papillomavirus (HPV) 16, 18, 31, 33, 35, 39, 45, 51, 56, 58, 59, and 68 are considered high-risk or probable high-risk due to their association with cervical cancer. HPV 6, 11, 42, 43, and 44 are considered low-risk for cervical cancer but may cause genital warts. Other HPV genotypes found in the sample may have intermediate or unknown risk for cervical cancer.

Transmission of antibiotic resistance genes in the soil ecosystem



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

Zalo: 0901802182