



Giới thiệu môn học chuyên biệt

Phân tích dữ liệu Tin Y Sinh

Ngày 18 tháng 11 năm 2024

TS. Lưu Phúc Lợi

Email: luu.p.loi@googlemail.com

Nội dung

1. Giới thiệu về dữ liệu y sinh
2. Ví dụ dự án DeepHE, DeepECG và VECV
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
6. Nội dung môn học “Phân tích dữ liệu tin y sinh”

PHUC–LOI LUU, PhD

Current Affiliation	Head of Data Science Division Tam Anh Research Institute (TamRI)
Webpage	https://scholar.google.com.au/citations?user=KPIqpJsAAAAJ&hl=en
Address	592/17 Lac Long Quan, P5, Q11, TpHCM
Home	5 Rose st, Sefton, Sydney 2162 Australia
Telephone	+84 (0)901802182
	E-mail luu.p.loi@googlemail.com

Education and Research interest

Undergraduate at University of Science - Ho Chi Minh City National University (2000 – 2005)

- Major: Bioinformatics

- Research Topics:

- Genotyping HCV, HBV, HPV
- Anti-biotics resistance in bacteria

PENICILLIN BINDING PROTEINS (PBPs) IN *STREPTOCOCCUS PNEUMONIAE*: DATABASE SETTING UP AND APPLICATION

Luu Phuc Loi, Thai Ke Quan, Nguyen Hoang Chuong, Thai Thien Minh, Ho Huynh Thuy Duong

Department Genetics, Faculty of Biology, University of Natural Sciences, Viet Nam National University - Ho Chi Minh City

ABSTRACT

β -Lactam antibiotics inhibit the growth of *Streptococcus pneumoniae* by inactivation of cell-wall-synthesizing enzymes, the penicillin-binding proteins (PBPs). Resistance to β -lactam antibiotics of *S. pneumoniae* is essentially due to altered PBPs with decreased affinities to the antibiotic. We investigated relationship between point mutations on ppp genes – pbp1a, 2b, 2x - and penicillin resistance in *S. pneumoniae* clinical isolates. With sequences determined from local isolates combining to data retrieved from GenBank, we established a database named PGD (Penicillin Binding Protein Gene Database) which integrated information about nucleotide sequence and point mutations of 472 strains. Using the set up database, we applied a PCR-based protocol to detect penicillin resistant *S. pneumoniae* strains. The protocol detects 18/20 resistant strains and needs further improvement.

Hepatitis C virus isolate MB14 5' UTR

GenBank: AY690659.1

[FASTA](#) [Graphics](#) [PopSet](#)

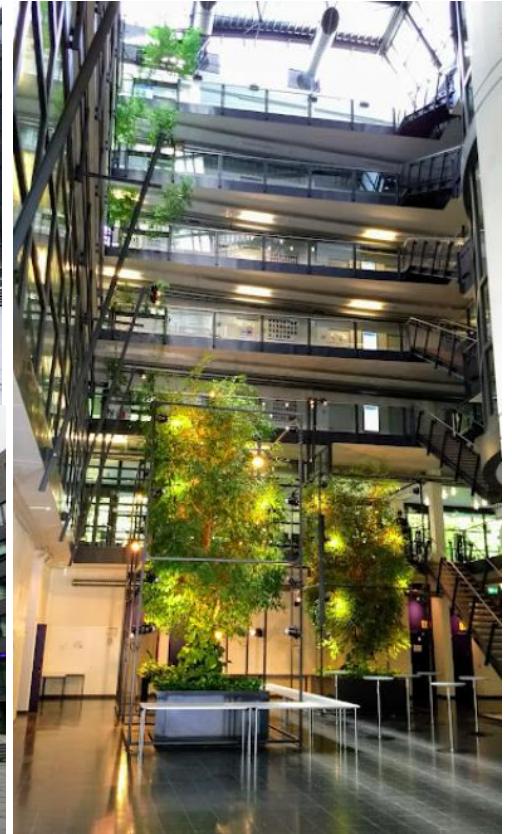
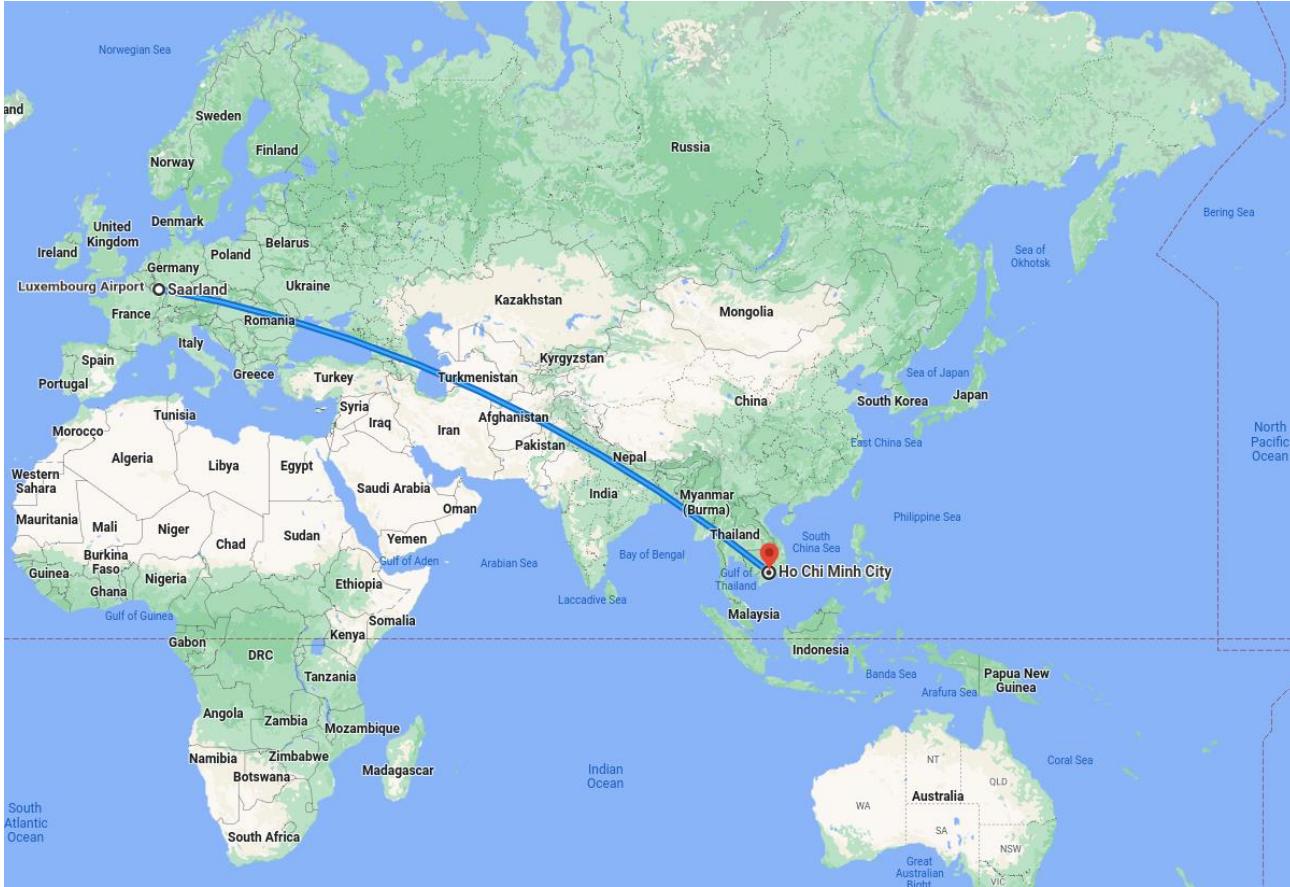
[Go to:](#)

LOCUS AY690659 94 bp RNA linear VRL 22-AUG-2004
DEFINITION Hepatitis C virus isolate MB14 5' UTR.
ACCESSION AY690659
VERSION AY690659.1
KEYWORDS .
SOURCE Hepacivirus C
ORGANISM [Hepacivirus C](#)
Viruses; Riboviria; Orthornavirae; Kitrinoviricota; Flasuviricetes; Amarillovirales; Flaviviridae; Hepacivirus.
REFERENCE 1 (bases 1 to 94)
AUTHORS Nguyen,H.C., Ho,T.T.T., Luu,P.L. and Ho,H.T.D.
TITLE Genotyping of Hepatitis C virus by real-time RT-PCR
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 94)
AUTHORS Nguyen,H.C., Ho,T.T.T., Luu,P.L. and Ho,H.T.D.
TITLE Direct Submission
JOURNAL Submitted (20-JUL-2004) Genetics, University of Natural Sciences Viet Nam National University - Ho Chi Minh City, 227 Nguyen Van Cu Street Dist. 5, Ho Chi Minh City 0848, Viet Nam
FEATURES Location/Qualifiers
source 1..94
/organism="Hepacivirus C"
/mol_type="genomic RNA"
/isolate="MB14"
/db_xref="taxon:[11103](#)"
/country="Viet Nam"
5' UTR <1..>94
ORIGIN //
1 ctcgcttccct ggagctggcc ggatccgagg atgatttcat gtgcgtatgag ctgcctgacg
61 gccaggaacg accgggtcct ttcttgatc accc

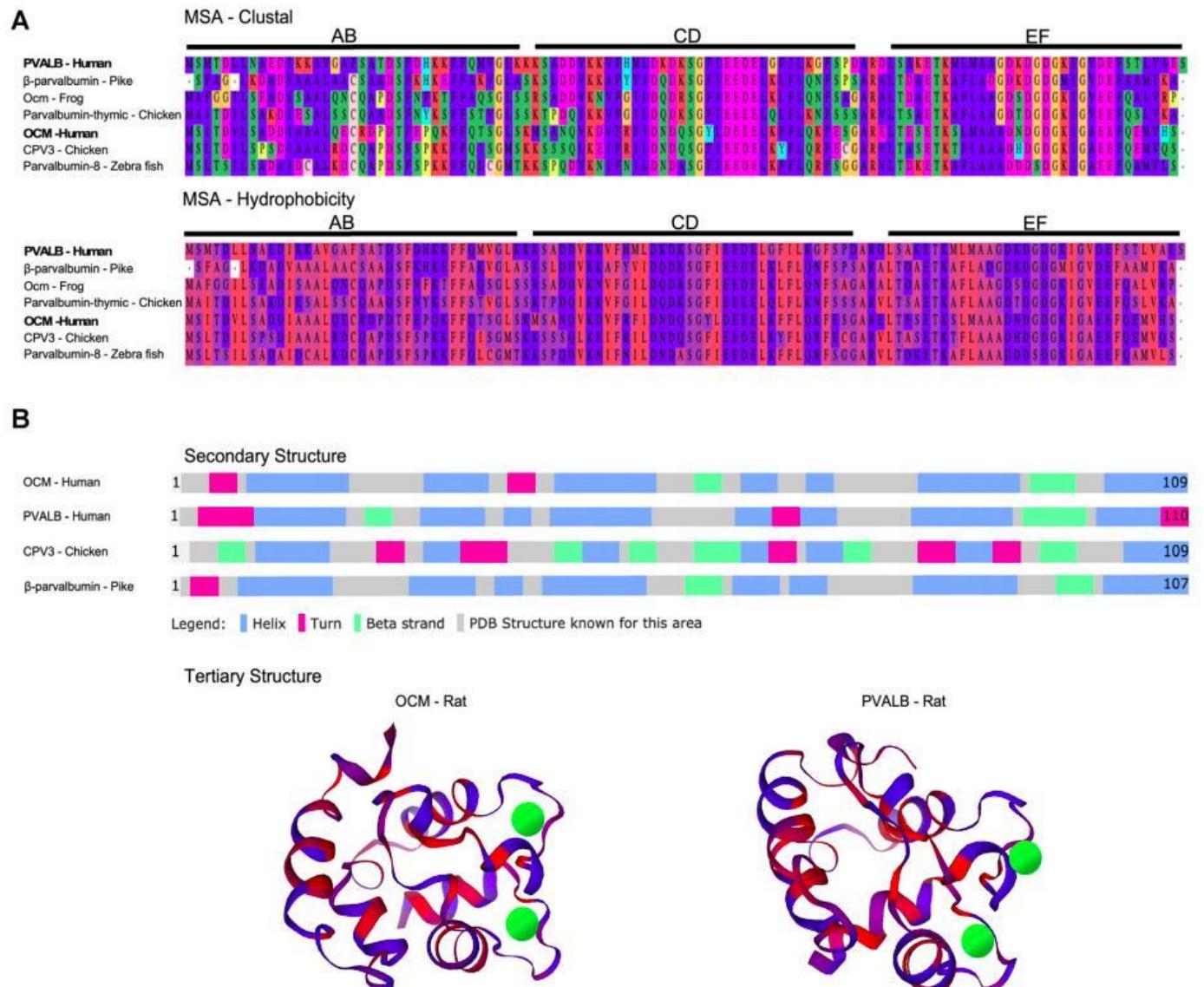
The American journal of tropical medicine and hygiene 73(6):55-55 (2005).

Master degree in Bioinformatics at Max Planck Institute for Informatics

University of Saarland in Germany (2007 - 2010)



Identifying functional discriminative motifs in protein families



- Protein structures and functions
- Bigdata mining and Machine Learning



PhD at Max Planck Institute for Molecular Biomedicine and University of Muenster in Germany (2011 - 2014)



PhD in Bioinformatics and iPS technology

- Research:

- The dynamics of stem cell and induced pluripotent stem cell (iPS) epigenomes using genome-wide Next Generation Sequencing (NGS)
- Somatic memory in iPS technology
- Bioinformatics pipeline development



HOME | ABOUT | ARCHIVE | SUBMIT | SUBSCRIBE | ADVERTISE | AUTHOR INFO | C



Disclosing the crosstalk among DNA methylation, transcription factors, and histone marks in human pluripotent cells through discovery of DNA methylation motifs

Phuc-Loi Luu¹, Hans R. Schöler^{2,3} and Marcos J. Araúzo-Bravo^{1,4}

Author Affiliations

¹Computational Biology and Bioinformatics Group, Max Planck Institute for Molecular Biomedicine, 48149 Münster, Germany;

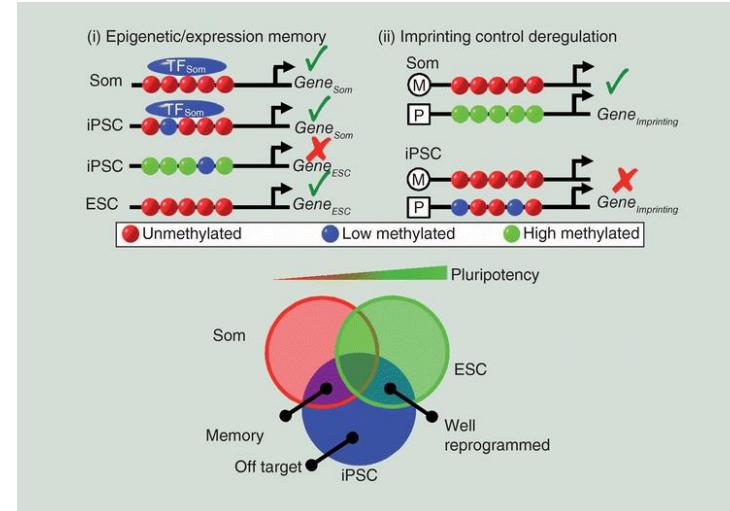
²Department of Cell and Developmental Biology, Max Planck Institute for Molecular Biomedicine, 48149 Münster, Germany;

³University of Münster, Medical Faculty, 48149 Münster, Germany

EPIGENOMICS, VOL. 10, NO. 2 | RESEARCH ARTICLE
Rules governing the mechanism of epigenetic reprogramming memory

Phuc-Loi Luu[†], Daniela Gerovska[†], Hans R Schöler & Marcos J Araúzo-Bravo

Published Online: 16 Jan 2018 | <https://doi.org/10.2217/epi-2017-0098>



P3BSseq: parallel processing pipeline software for automatic analysis of bisulfite sequencing data

Phuc-Loi Luu, Daniela Gerovska, Mikel Arrospide-Elgarresta, Sugoi Retegi-Carrión, Hans R Schöler, Marcos J Araúzo-Bravo

Bioinformatics, Volume 33, Issue 3, 1 February 2017, Pages 428–431, <https://doi.org/10.1093/bioinformatics/btw633>

Published: 06 October 2016 Article history ▾

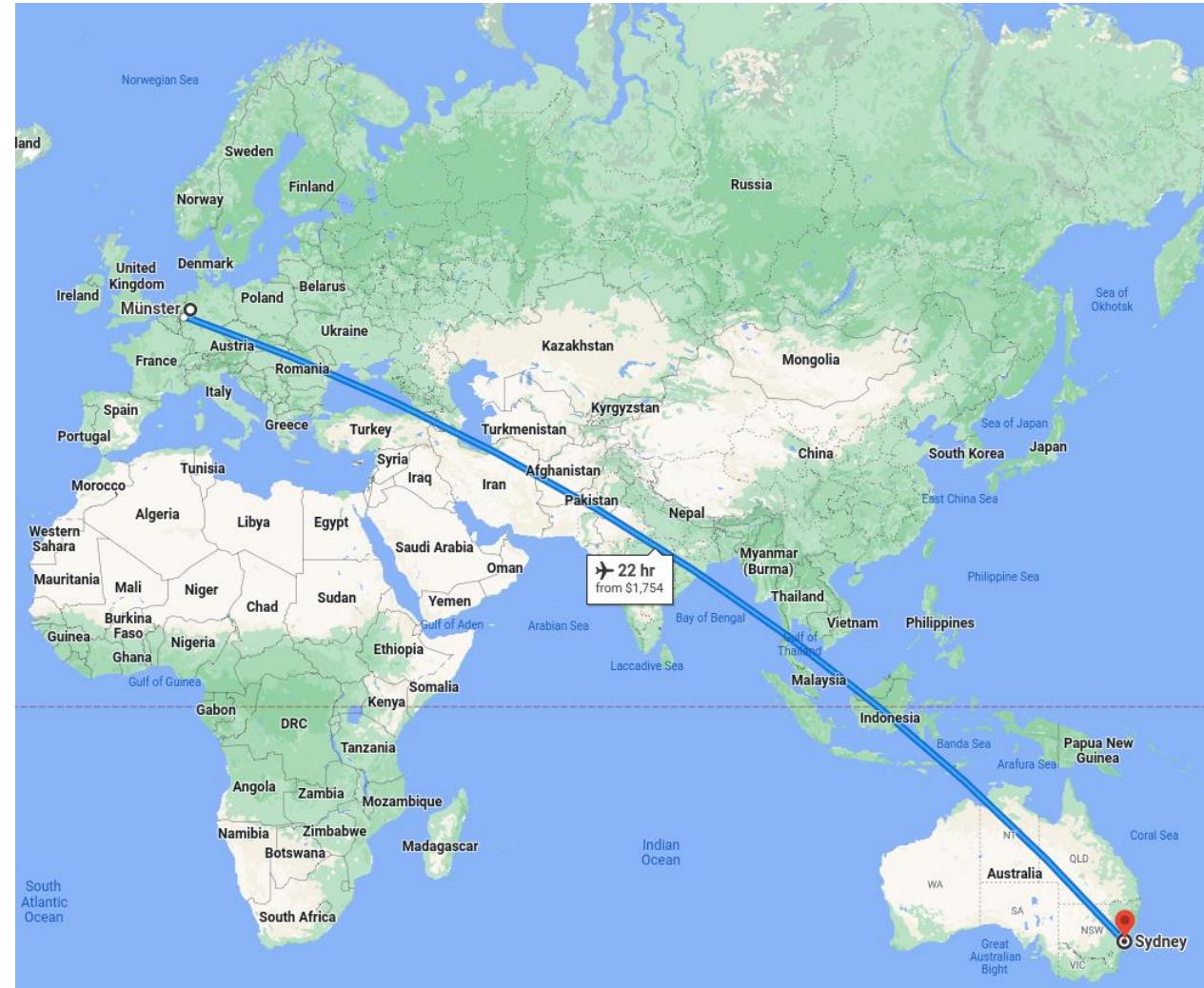
PDF Split View Cite Permissions Share ▾

Abstract

Motivation

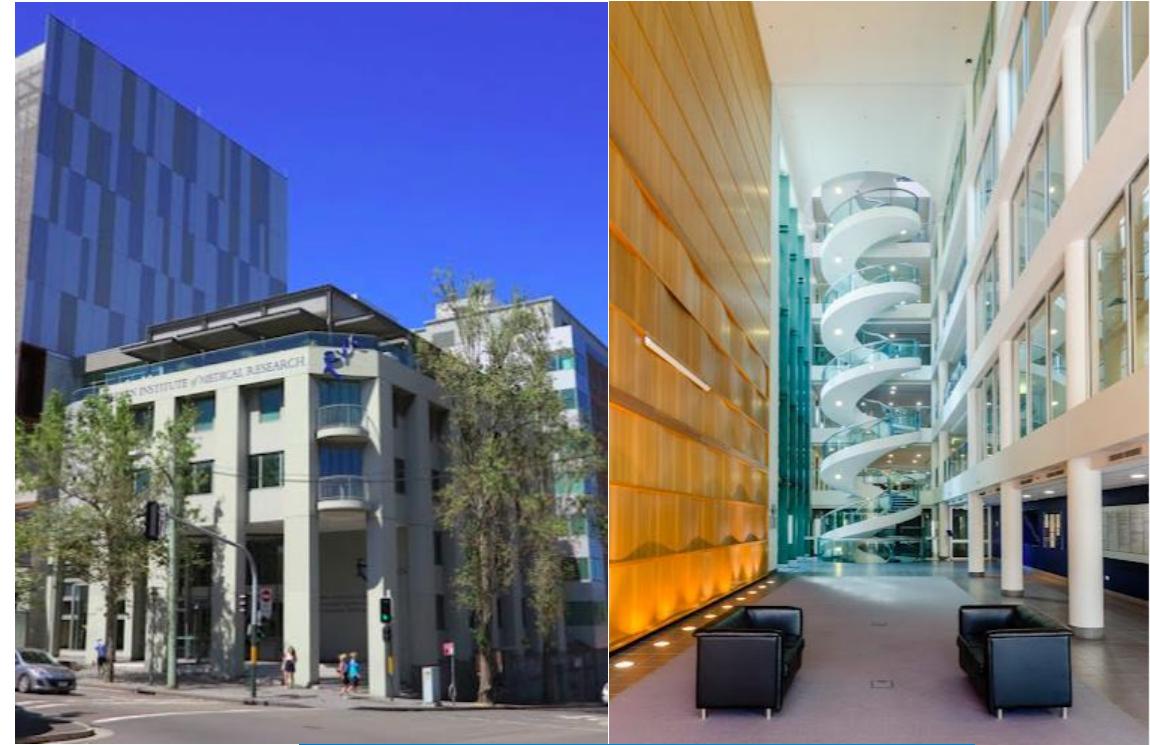
Bisulfite sequencing (BSseq) processing is among the most cumbersome next generation sequencing (NGS) applications. Though some BSseq processing tools are available, they are scattered, require puzzling parameters and are running-time and memory-usage demanding.

Seven years of research at Garvan Institute of Medical Research in Australia (2014 - 2022)



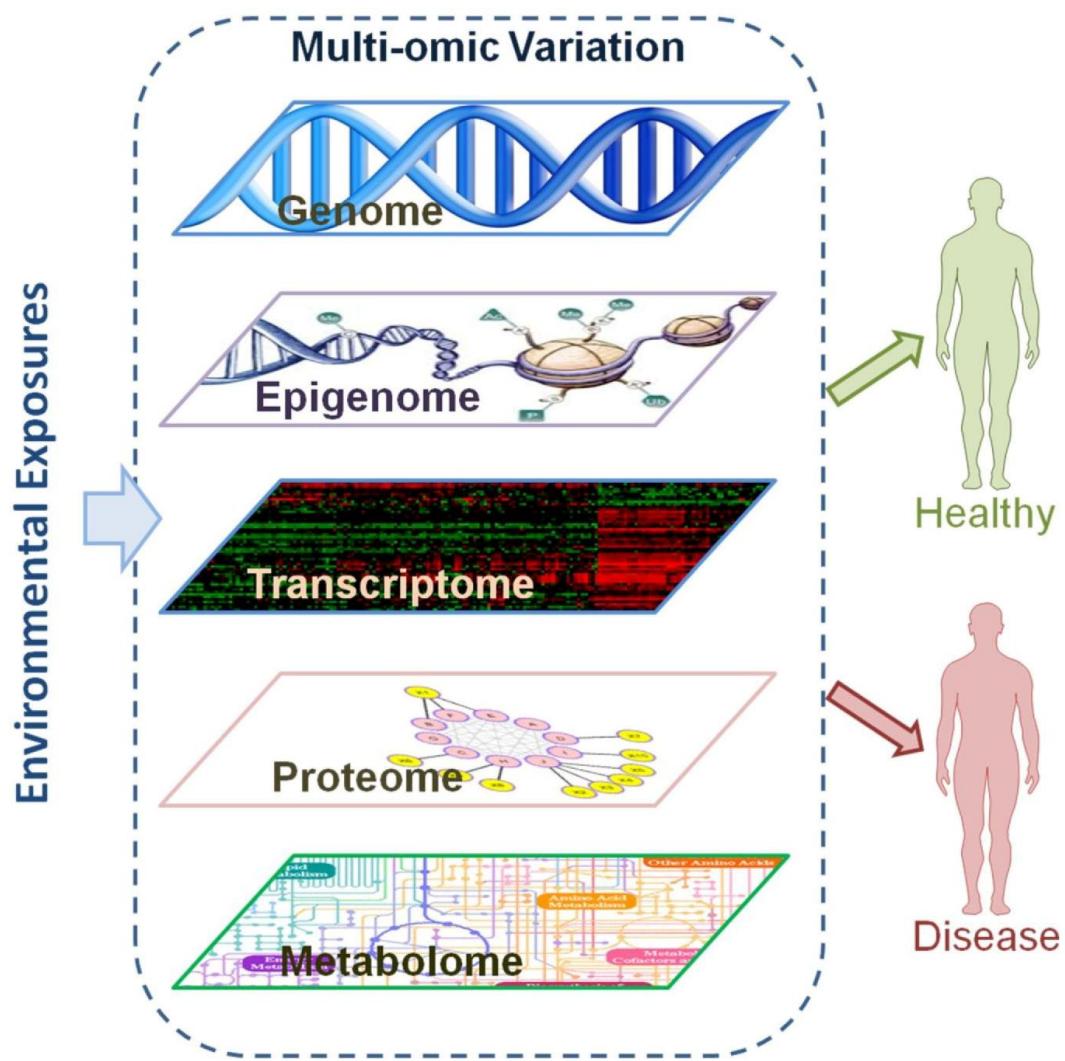
Seven years of research at Garvan Institute of Medical Research in Australia (2014-2022)

- Major: Bioinformatics and Machine Learning
- Research:
 - Cancer Genomics and Epigenomics
 - Single cell and Spatial transcriptomics technology and other diagnosis methods
 - Bioinformatics pipeline development



More than one year at Zymo Research (Sep 2022-Nov 2023)

- Microbiomics
- Epigenomics
- Transcriptomics
- Bioinformatics



THE NOBEL PRIZE IN CHEMISTRY 2024



**David
Baker**

"for computational
protein design"

**Demis
Hassabis**

"for protein structure prediction"

**John M.
Jumper**

THE ROYAL SWEDISH ACADEMY OF SCIENCES

THE NOBEL PRIZE IN PHYSICS 2024



John J. Hopfield

"for foundational discoveries and inventions
that enable machine learning
with artificial neural networks"

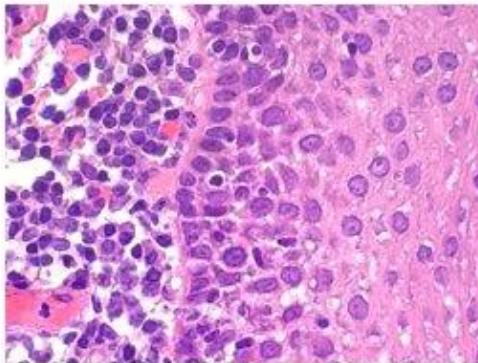
Geoffrey E. Hinton

THE ROYAL SWEDISH ACADEMY OF SCIENCES

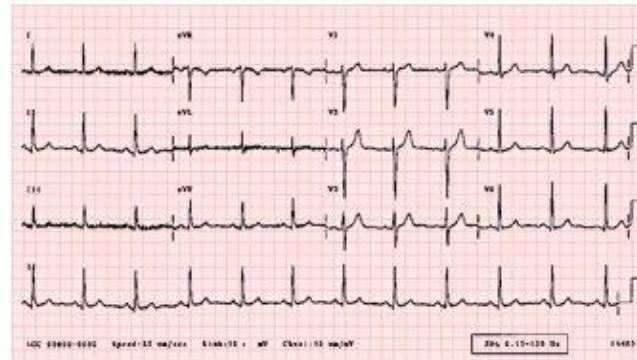
Giới thiệu về dữ liệu y sinh

Dữ liệu ảnh giải phẫu bệnh và chẩn đoán hình ảnh

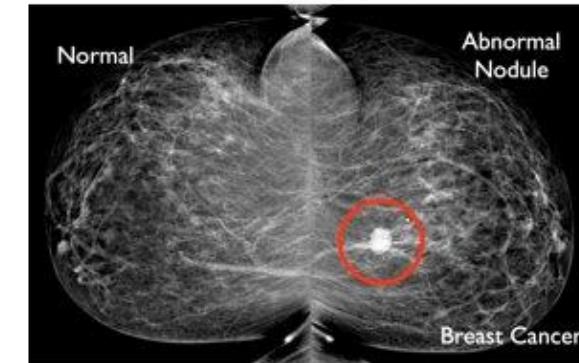
1. HE



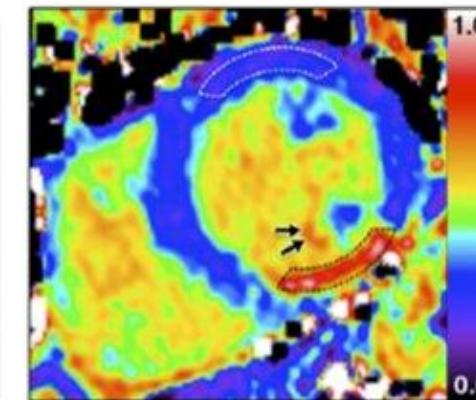
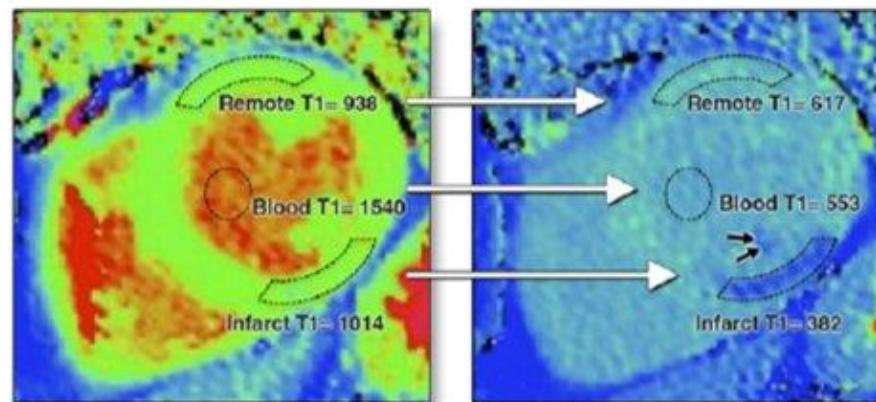
2. ECG



3. Mammography



4. Heart MRI



Dữ liệu hệ gen



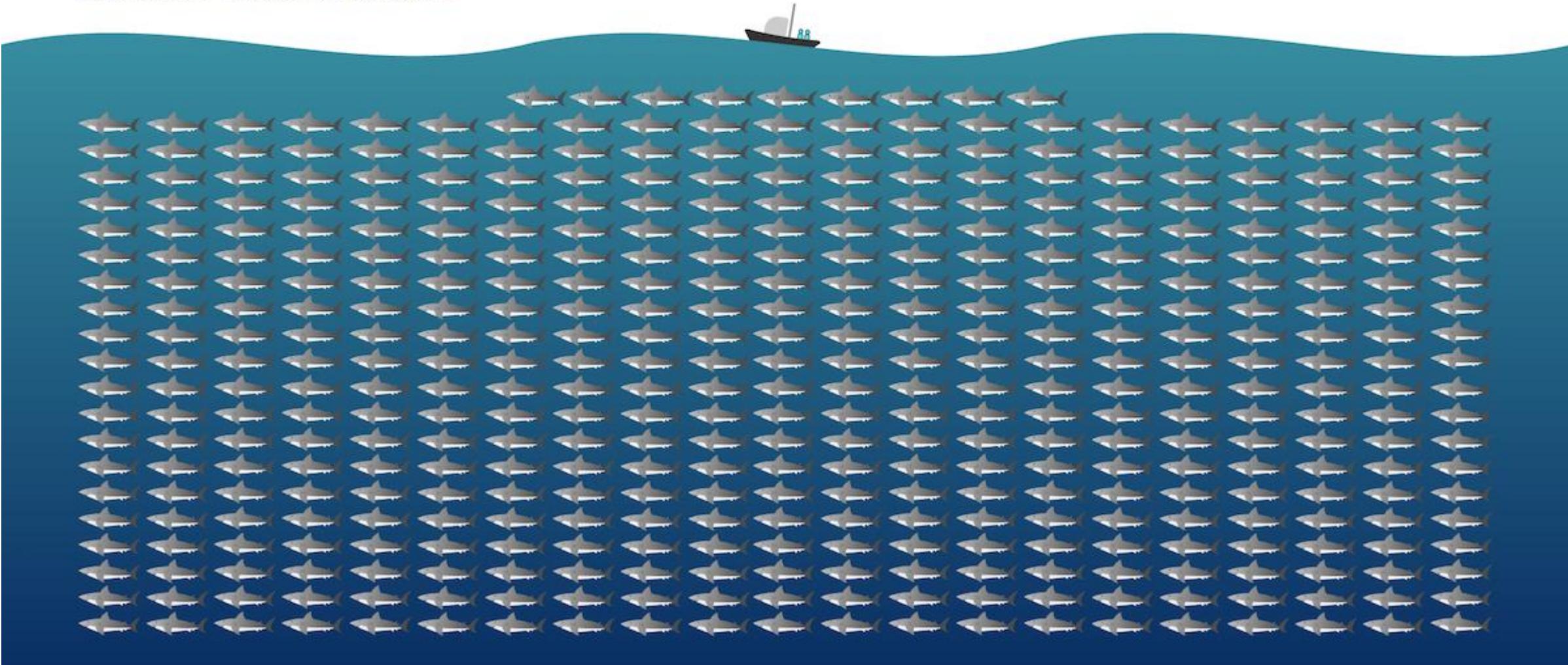
48541 agccctcaa agaaatgttc tcagcaggca tggagccag gacttgtcc cttgggtag
48601 agagccgggt tgaaggtgac tgaagtaaa tggagacata gaggcgaaaa ggttggtag
48661 ttccggagg tgggggtgt gggacactgc tggactcta gatgcacccc tgccagttct
48721 gcttaaagat ttggccggg gggcgggggg gcggaggtaa gtcattttac tggtaagtaa
48781 tttaaacct tttatattaa aagcaaacgt ggatgtttaa tgaatgaaat tcattctgaa
48841 atgaaaattt cacgtgtatg tggaaataa caccgggtt cagagggac tttctggctg
48901 gcacgacact ccagattccc agggccctgt caccctctc tgcccacagg gcacccaaat
48961 ggagaaggta tgggaggaga gccaggccgg agtcagagca cactggtgac tccacatttg
49021 cagcgtcccc tgcctctc ctgaggctt gcaacgttca atatgtctaa caaaactcccc
49081 ctgtcccggt ccagttctg aggacaagag ccaccaccc tagcaataa agaccaggca
49141 accctttgac tcatctttgt ggttctctgg aatcagagg tagccacatc gctggaggg
49201 ggatgttggc actcggttga aaaggatcaa ggaatggggg gacaggatgt tggggacatc
49261 acatggaaa tgacagagaa gggggcaca gccggatgg ggggggggg ccggcgttcc
49321 tacatccccct ggcttgaagc acgctccagg gcagaaggaa aacactgtc tttgggtcc
49381 aagagacctg agttcaattt ctggctccac cactgaccac ctgttaacc ttgaaactgt
49441 gctgctgaa cttcagggtt cccttctaaa aatagaggaa aaaaggatgtc atttccctt
49501 gcccgttga gaacgaaatg gtcaaggac caaggagct cagcaaaatg cggccgttcc
49561 cccgcgttgc caaaccttc ctctttagga ggccacggca accgtatgtt gacagaagag
49621 cagcacctt attaatgtc tcccttgcgtat tggtctttagga caagtccatc aaccccttgc
49681 ggctgttcc tcattggaa aatatggctg ccagtaaaac ctggcttgc cacccttgg
49741 ggcacttggc aaacagcaaa agatgtccaaat tggcggatgg gggccaggcc cagtggctca
49801 tgccgttgcattt cccaggcaatt taggaaggcc aggtgggggg atcaccctgg gtcaggatgt
49861 tgacccatggc ctggccaaata tggcttgcgtat aaaaataaaaaaa aatattagccg
49921 ggcacatggc tgggttgcgtat taaatccatg tactggggag gtcggatgg gagaatcgct
50041 tgaaccggca agggggaaatg tgcatgttgc caagattgtt ccaactgcact ccagcttgg
50101 caacagagcg agactgttgc taaaaaaaaaaaaaaa aaaaaaaa aaaaaaaa aaaaaaaa
50161 tgtaaaaaaaatctttccactt gcaaggccca gtttccacca ggctgggtt caggccatgg
50221 gggggggggcc ccaagcatcg cctcttgcgtat gcccggggag gccccggca tcccttgg
50281 tcaggagagg cccagggttgc gggggggggat ggttggggat tcccttgcgtat gccaaggcc
50341 tggggggaaa gcaacttgcgtt gcaatggggaa ggaaccggaa aagactgacc cggaggac
50401 cagccatggc tgggttgcgtat gggggggggat ggttggggat tcccttgcgtat gccaaggcc
50461 agcccttgcgtt gtcgttgcgtat tggggggggat ggttggggat tcccttgcgtat gccaaggcc
50521 gagatggccctt ctggggatggc tggggggggat ggttggggat tcccttgcgtat gccaaggcc
50581 ttaaaggcgc tggtgttgcgtat tggggggggat ggttggggat tcccttgcgtat gccaaggcc
50641 acaggtgtga cagggtgggg gactggggaa taagatggat aaccctggctt gggggggggat
50701 agttccaca gccaggcgatgtt gacaggggggat ggttggggat tcccttgcgtat gccaaggcc



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



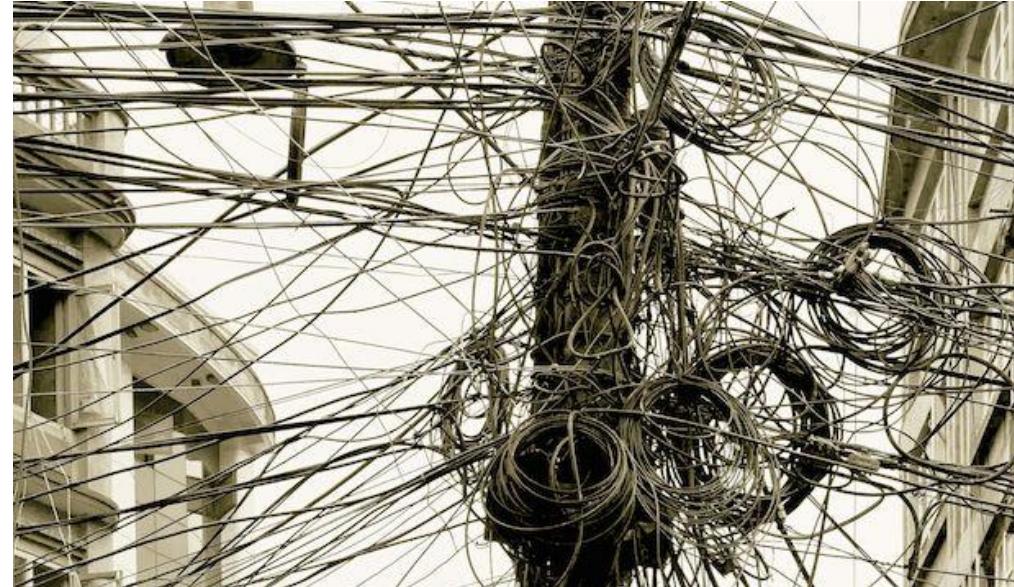
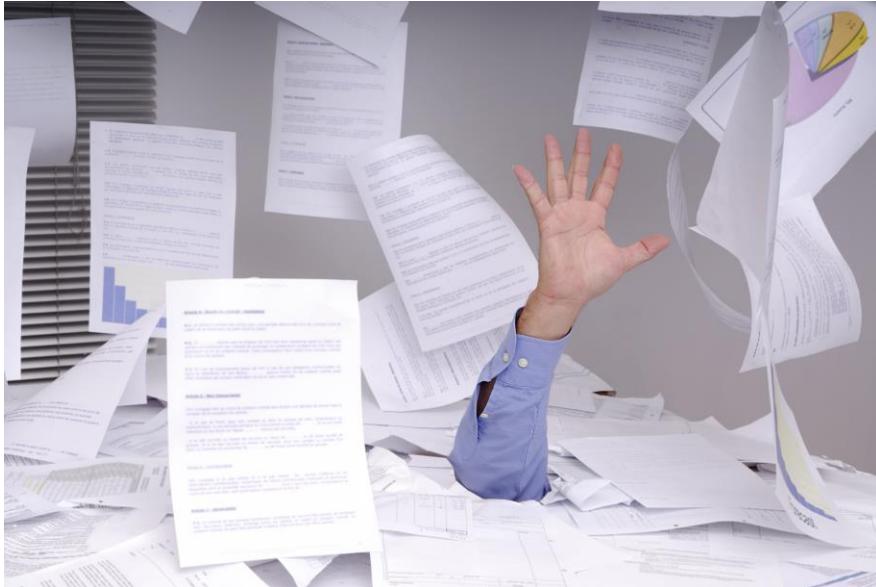
Người ngoài nghề nghĩ ...

CHỈ SỐ HUYẾT ÁP CHUẨN (Tiêu chuẩn của Tổ Chức y tế thế giới - WHO)

Độ tuổi	NAM		NỮ	
	Huyết áp tối đa	Huyết áp tối thiểu	Huyết áp tối đa	Huyết áp tối thiểu
15-19	120	70	111	67
20-29	124	75	114	69
30-39	126	79	118	73
40-49	130	83	126	78
50-59	137	85	134	81
60-69	143	84	139	81
70 trở lên	145	82	146	79

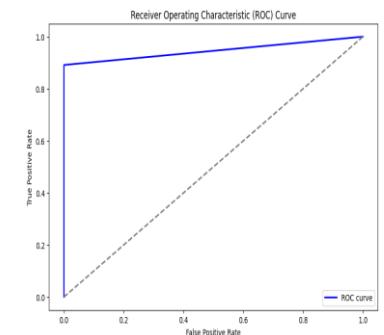
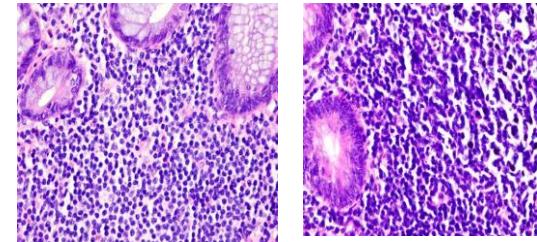
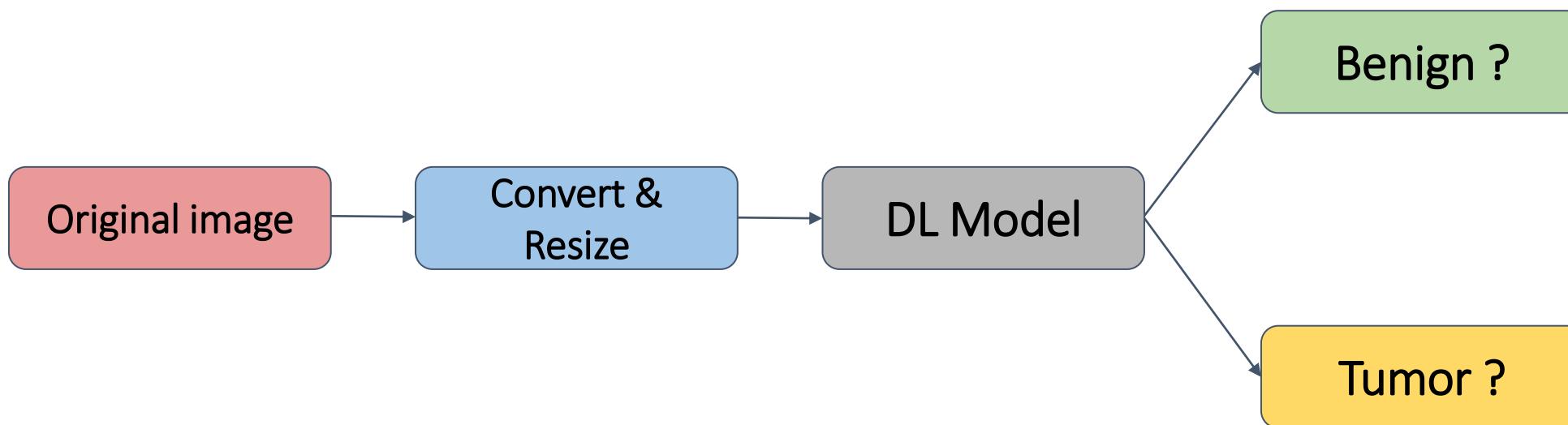
Microlife - «Hãy theo dõi huyết áp thường xuyên. Vì tăng huyết áp là kẻ giết người thầm lặng!»

Thật sự khi mình thật sự bắt tay vào ...

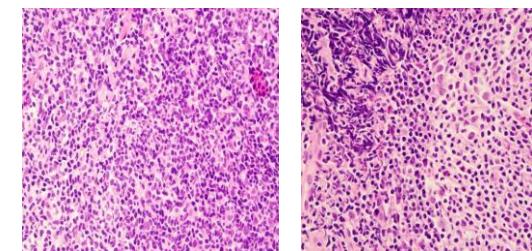


- ➔ Thu thập, xây dựng, lưu trữ và tối ưu hóa các hệ thống dữ liệu (**Data engineer**)
 - Kiến thức chuyên môn
 - Kiến thức về cơ sở dữ liệu (SQL and Non-SQL)
 - Kỹ năng lập trình
 - Hadoop

DeepHE: Lymphôm đường tiêu hóa

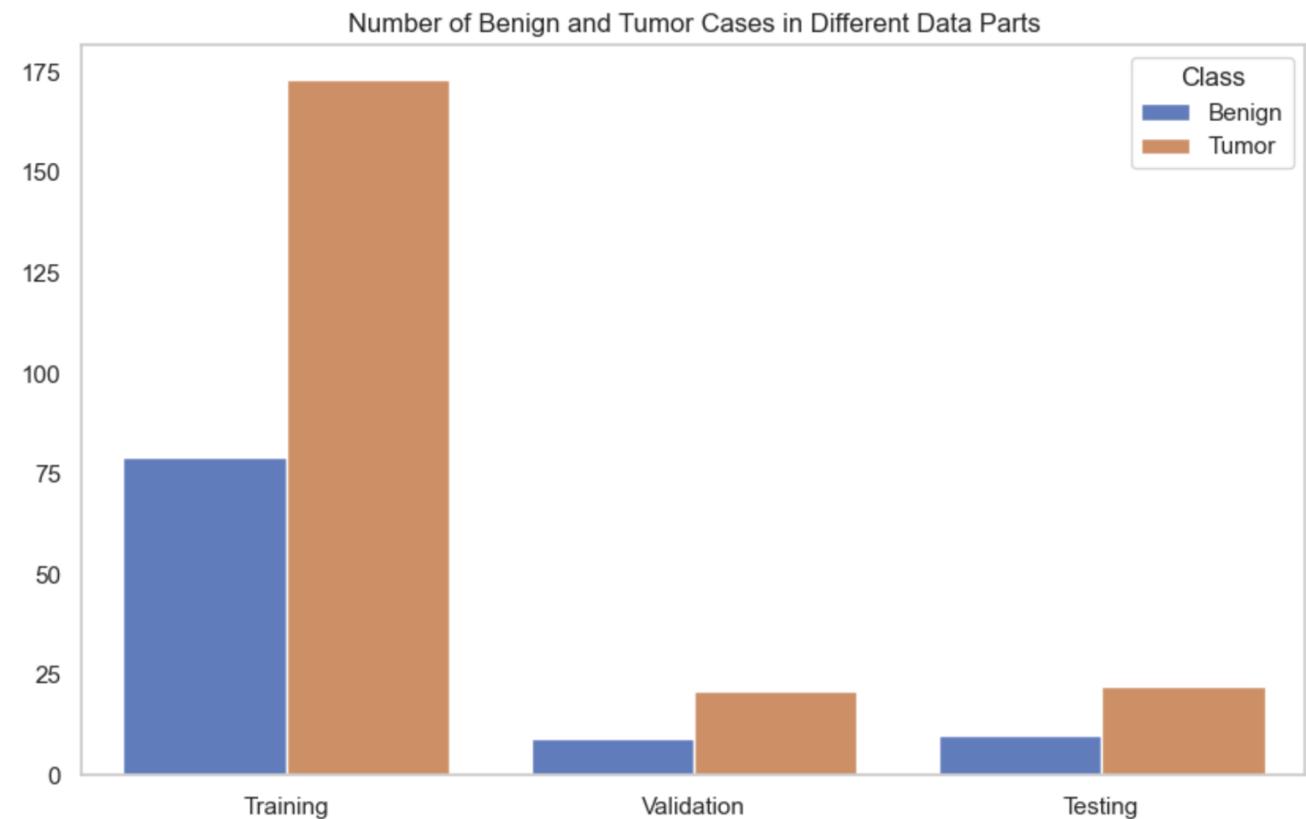


- **Accuracy:** 0.9398
- **Sensitivity:** 0.8913
- **Specificity:** 1.000
- **AUC:** 0.9413

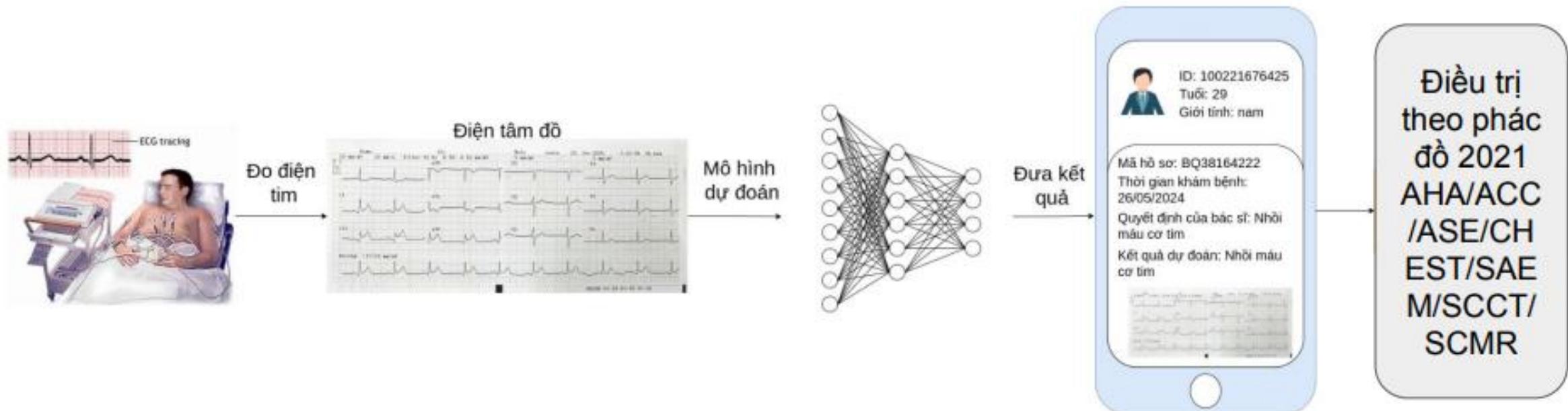


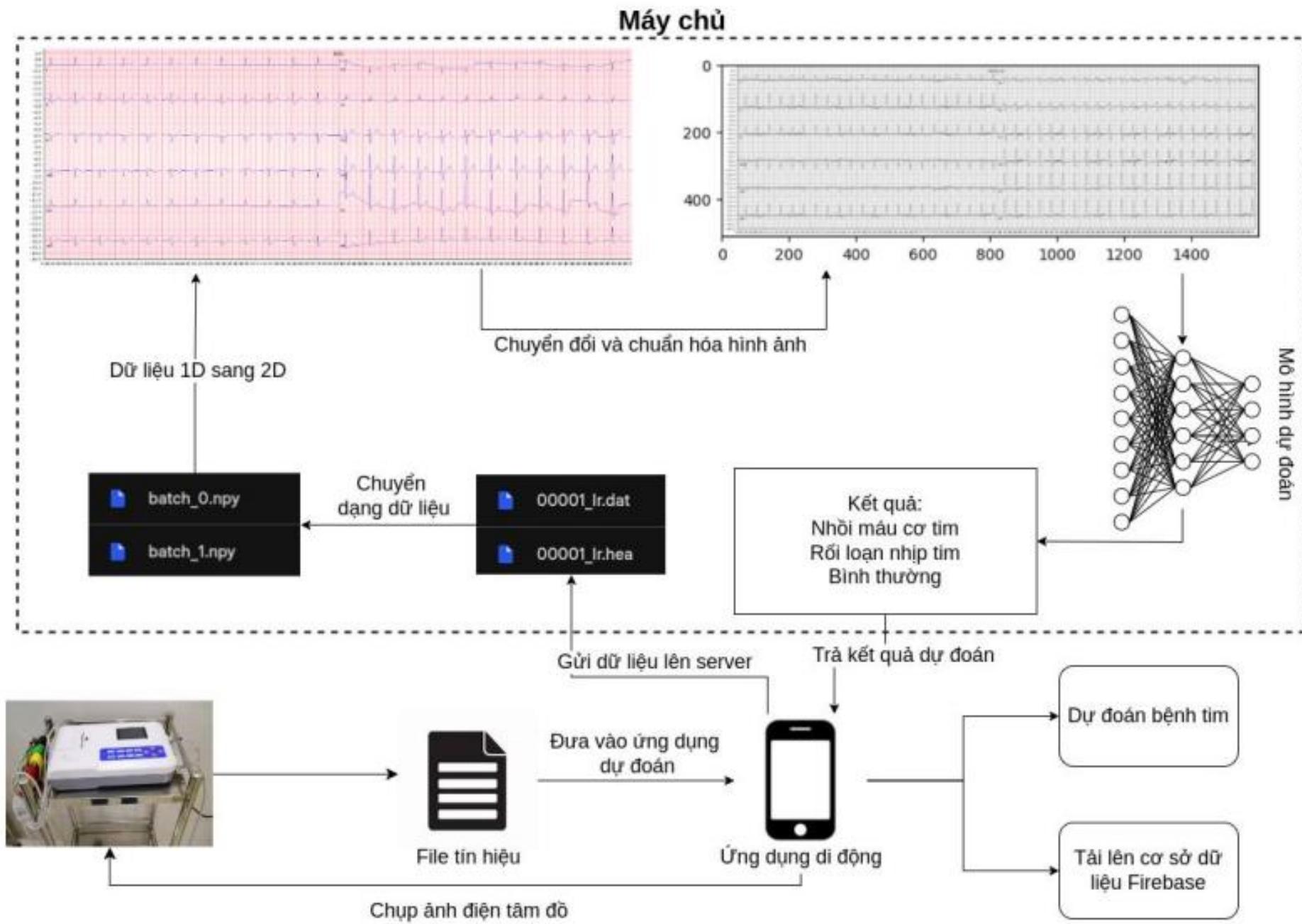
Dataset for DeepHE

- **Total:** 314 images
 - **Training:** 252 images
 - **Validation:** 30 images
 - **Testing:** 32 images

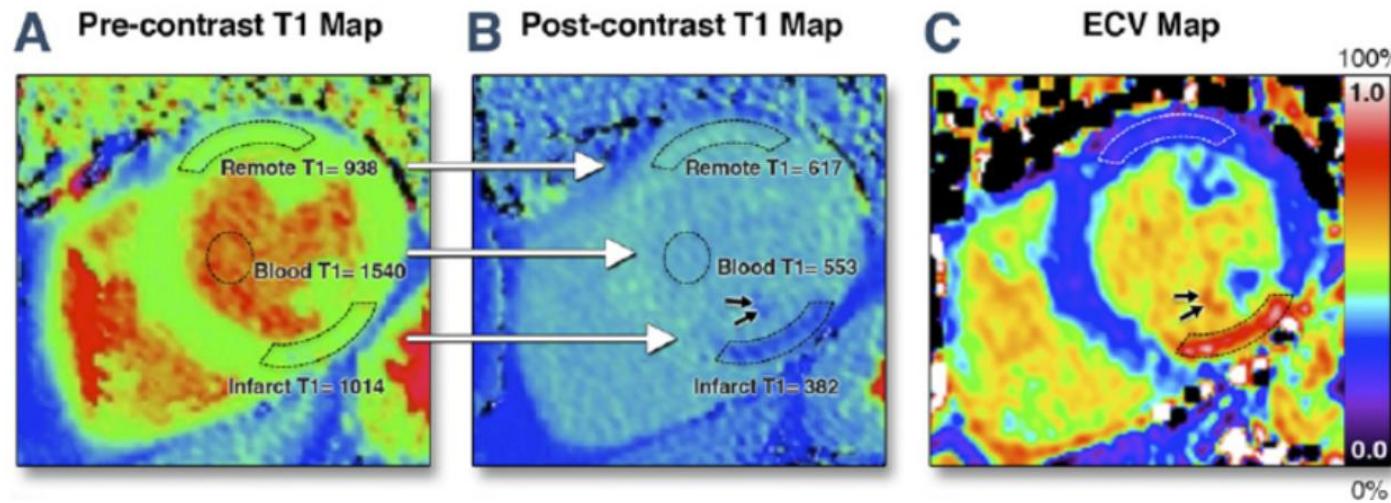


DeepECG: các bệnh về tim mạch

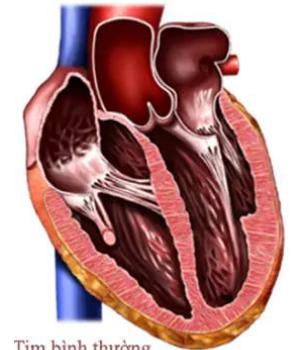




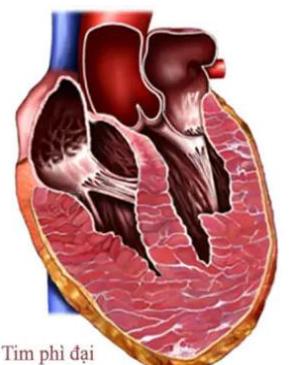
VECV: bệnh cơ tim phì đại



DL Model



Tim bình thường



Tim phì đại

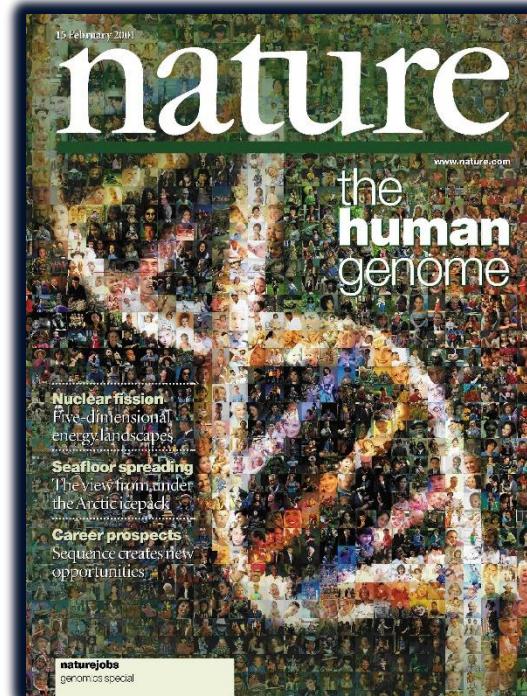
Giới thiệu về

GIẢI TRÌNH TỰ GEN THẾ HỆ MỚI

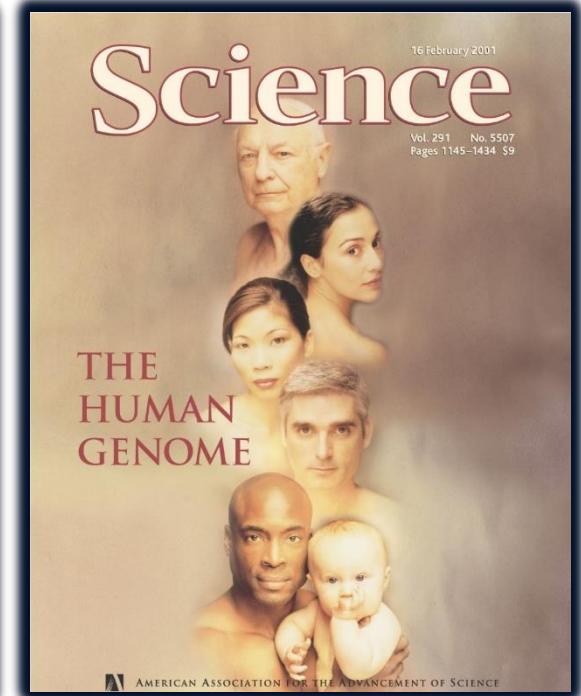
Dự án hệ gen người HGP (Oct 1990 - April 2003)

- 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).

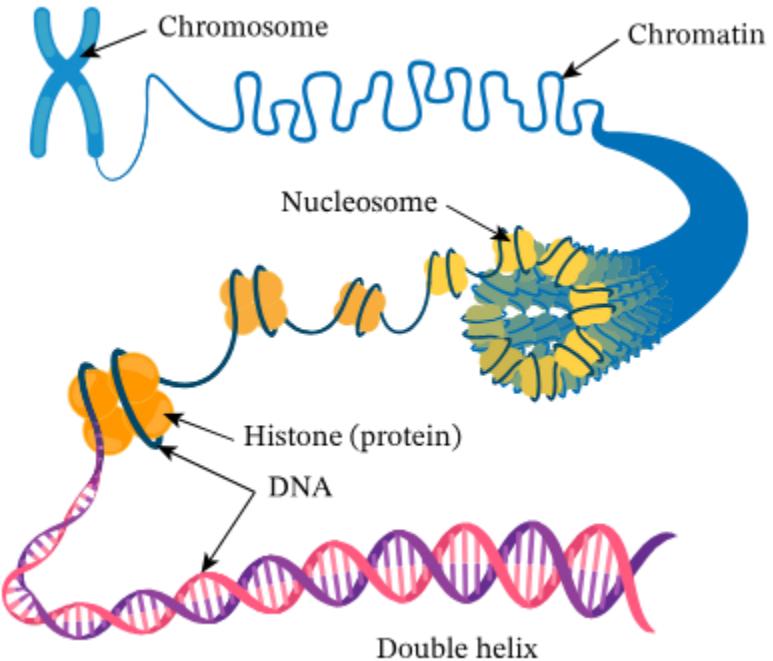
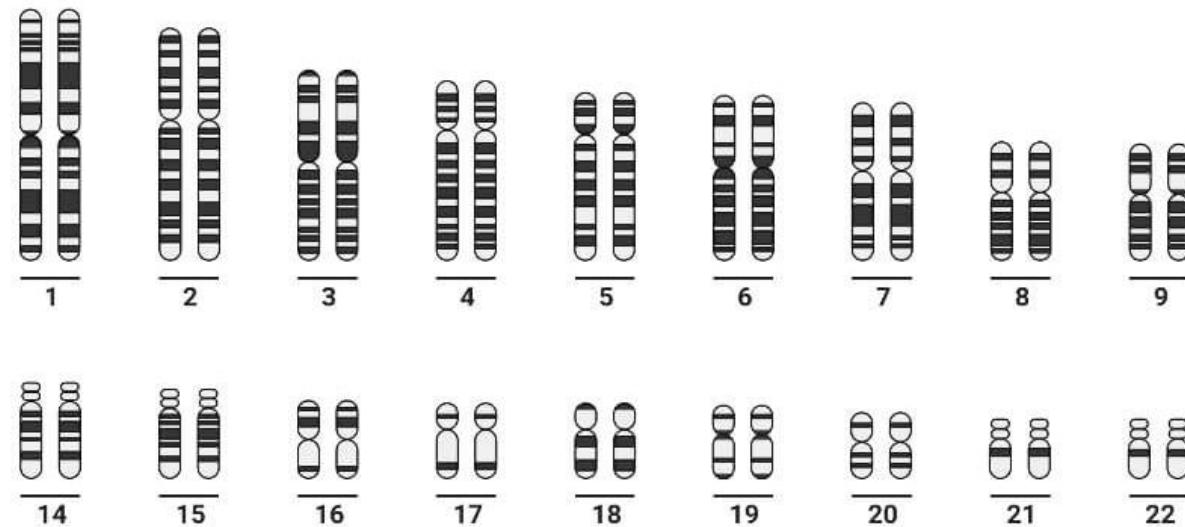


HGP Paper



Venter/Celera Paper

Human Karyotype



Hệ Gen người

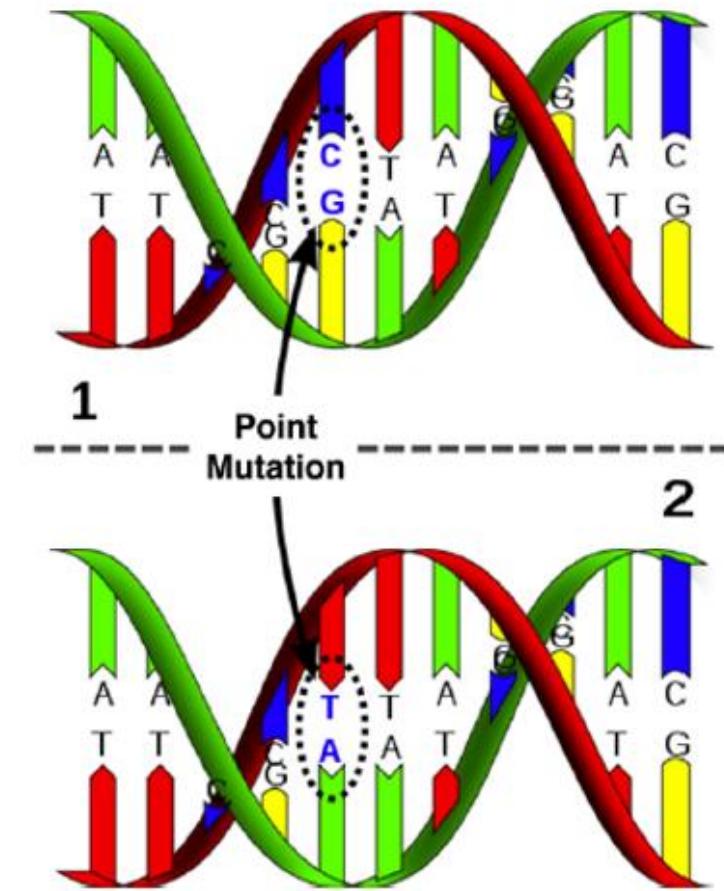
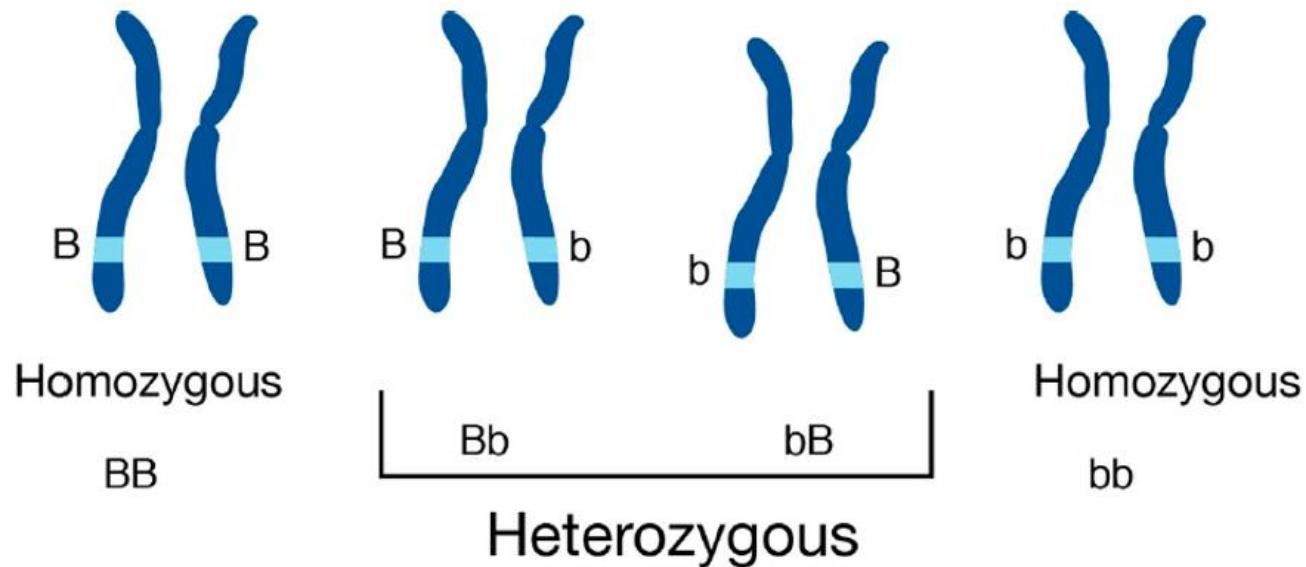


<https://www.ncbi.nlm.nih.gov/nuccore/806904736>

48541 agcccttcaa agaaatgttc tcagcaggca tggagcccag gacttgctcc cttttggtag
48601 agagccgggt tgaaggtgac tgaagtaaa tgggacagta gaggcgaaaa ggttgttag
48661 ttccctggagg tgggggtgt gggAACCTGC tgtgtactga gatgcacccc tgccaggct
48721 gcccgtaaat ttgaggcccc gggcaggggg gcccggatgaa gtcattttac tggtaagtaa
48781 ttttaaacct ttaatattaa aagcaaacgt ggatatgtaa tgaatgaaat tcattctgg
48841 atgaaaaatt cacgtatgt taaaaataa cacggggctt cagagaggac ttctggct
48901 gcagcagact ccagattccc agggccctg caccctcctc tgcccacagg gcacccaat
48961 ggagaagggtg tgggaggaga gccaggccgg agtcagagca cactggtgac tccacattt
49021 cagcgtgccc tgcctcttc ctgaggctt gcaacgtgca atatgctaag caaactcccc
49081 ctgtccccgt ccagttctg aggacaagag ccaccacctg tagcaaataa agaccagg
49141 accctttgac tcatcttggt gagtctctgg aatcagagg tagccacatc gctgagagg
49201 ggagtgaagc actcgggtga aaaggtaaaa ggaagtcaag gacaggatgt tggggacatc
49261 acctagacaa tgacagagaa gaggggcaca gccgagttag gggagagggg ccggcagtcc
49321 tacatccccct ggcctgaagc acgcctccagg gcagaaggaa aaacactgtc ttgggggtcc
49381 aagagacctg agttcaatt ctggctccac cactgaccac ctgtgttaacc ttgaactgt
49441 gtcgcctgaa ctcagggtt cccttctaaa aatagaggag aaaaggatgc attttcttcc
49501 gcccctgtga gaacgaaatg gtgcaggcac caaggagccct cagcaaagggt cggccctgcc
49561 cccgcctggc caaaccttc ctcttcagg ggcacccggca accgtatgtt gacagaagag
49621 cagcacccctg atttaatgtc tcccaagcatg tgccttgag caagtacct aacccctctg
49681 ggctgccttc tcattggaa aatatggctg ccagtaaaaac ctgcctgtc caccctctgg
49741 ggcacttggc aaacagcaaa agagtccaaa tgcaggctt gggccaggcg cagtggctca
49801 tgcctgtaat cccagcaatt taggaagcca aggtggggcg atcacctgag gtcaggatgt
49861 tgagaccagc ctggccaaca tggtaaaacc ttgtctctac aaaaatacaa aaattagccg
49921 ggcatgtatgg cgggtgcctg taatccctgt tactcgggag gctgaggca gagaatcgt
49981 tgaacccggc aggggaaggt tgcagtgc caagattgtg ccactgcact ccaggctggg
50041 caacagagcg agactctgtc tcaaaaaaaaa aaaaaaaaaaa aaacaatgcg gagctggct
50101 tgtaaaaaaaact ctgttccact gcaggccccca gtgtccacca ggctgggtg caggctatg
50161 ggggtgggggc ccagcatcg cctctcagca gccctggag gcggggcgca tccctgtcccc
50221 ctcgtggctc ggatgtttc tagcccaagt cctagggttac acctggcgtc gcctggctc
50281 tcaggagagg cccagggtga ggaggagcat ggtaaagggtg aagctgattt ggaagtgc
50341 tggggggaaa gcaactccct gcacattggg ggaaccggaga aagactgacc ccgaggac
50401 cagccagcat ggccttctt gggagcccat gttgggggat tcctgtcga gccaaggctc
50461 agcccttgtc gtcgcagggtg ctggctctgg cctcttcccc tcccatgcag gagcacagg
50521 gagatggctt ctgaggacct gttgcagctg tggccctggg aatagattt ccaggaggt
50581 ttaaaggcgc tgagtgttc atccagctaa gcctggggaa ggagcttgc tcaggtctgg
50641 acagggtgtc cagggatgg gactggaaag taagagatga aaccctggct ggaggctgt
50701 agttccacca gccagcgtc gacaaggagg gtccagatatt acccactagt gcccctcacca

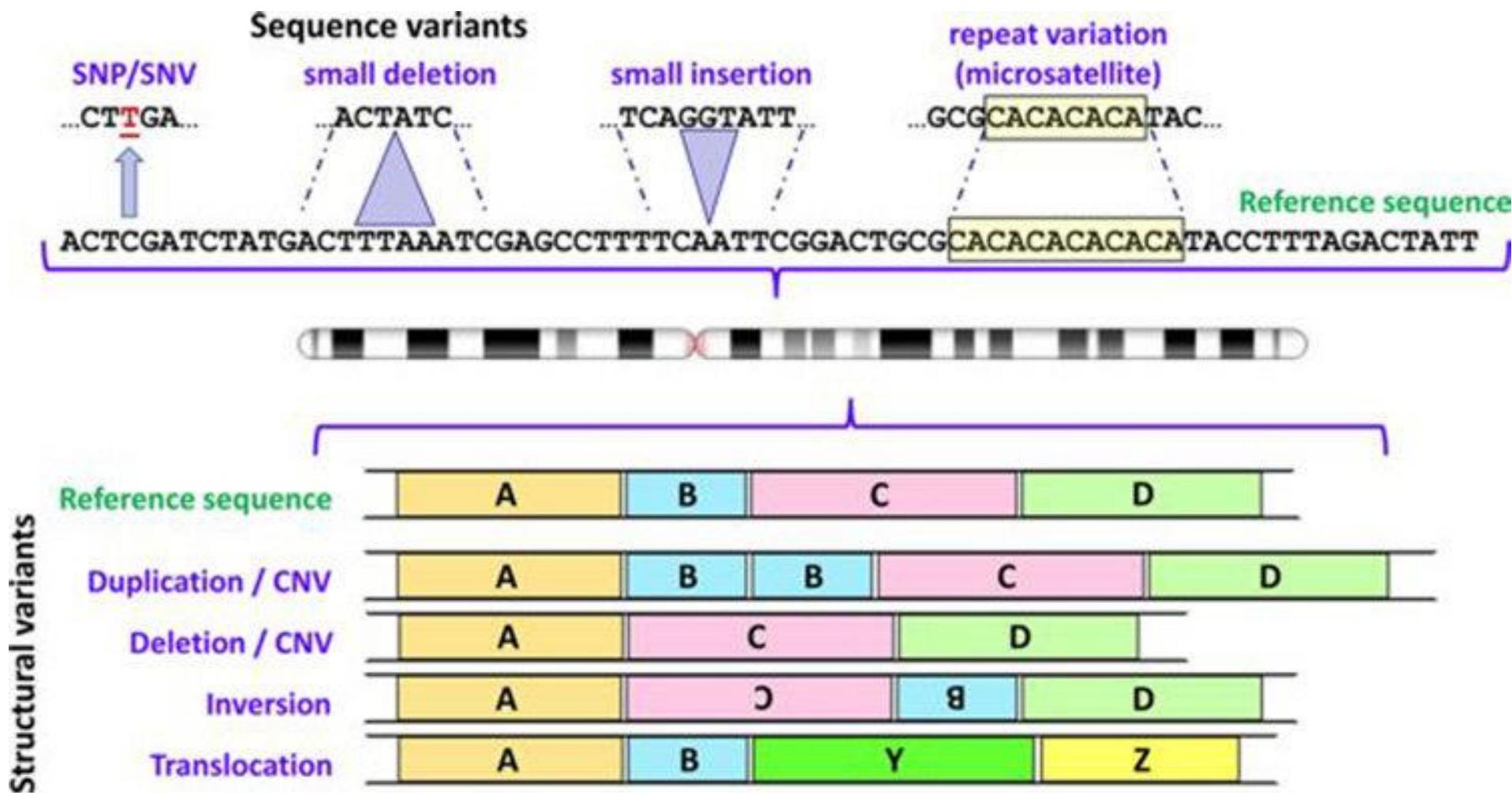
Các loại biến thể trên hệ gen

- 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

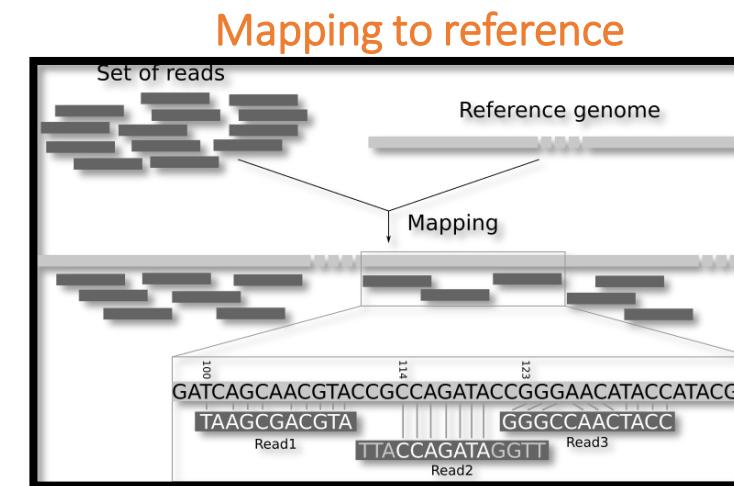
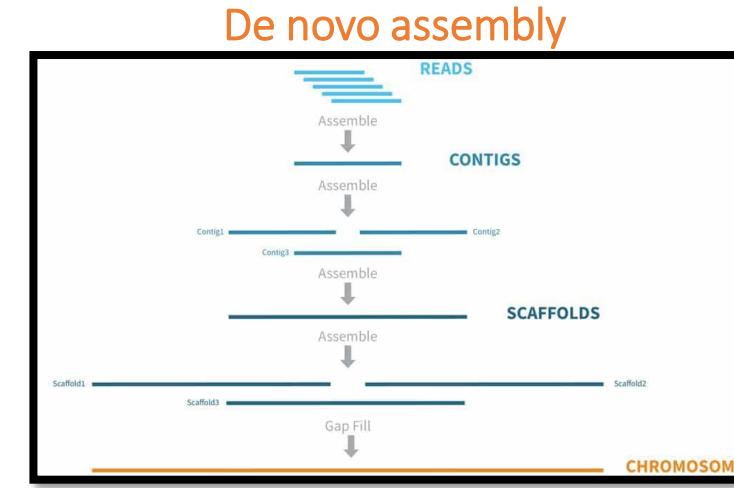
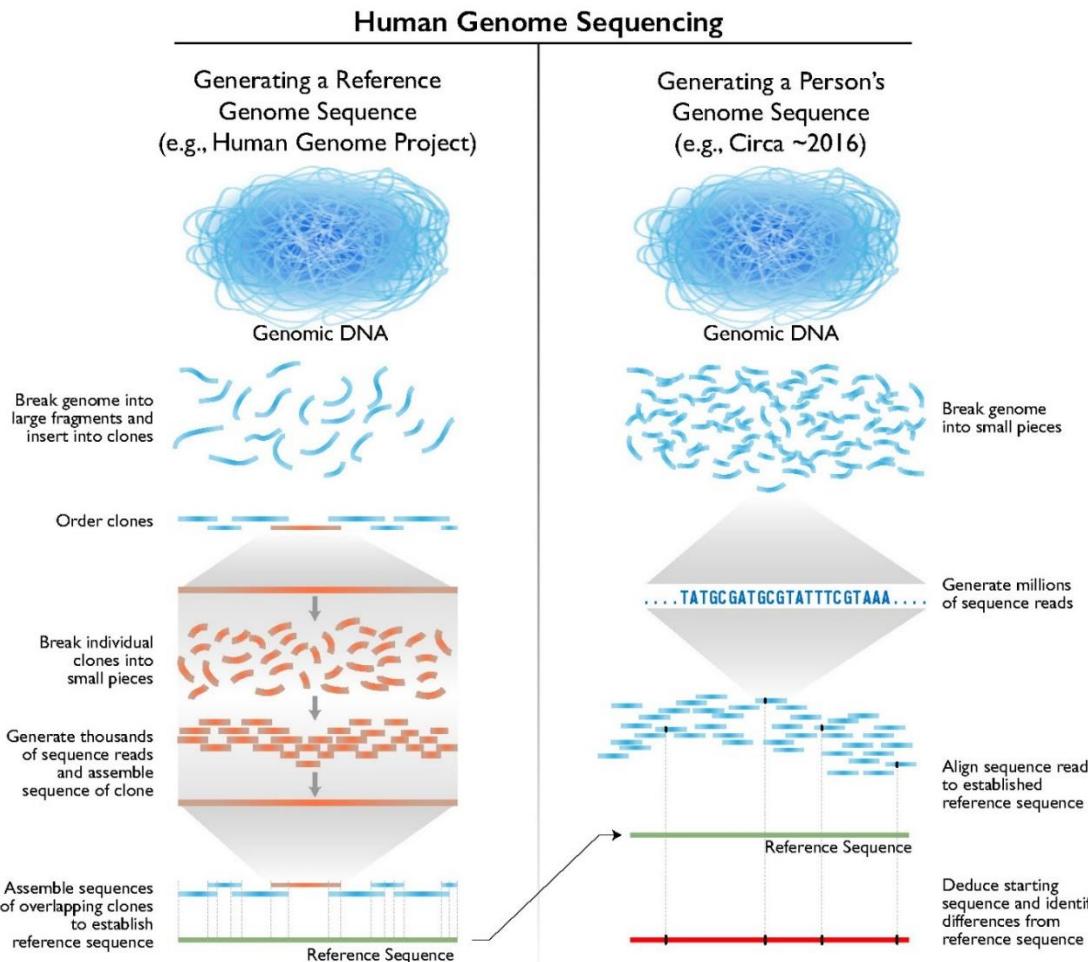


https://rosalind.info/media/point_mutation.png, <https://en.wikipedia.org/wiki/Zygosity>

Các loại biến thể trên hệ gen



Giải trình tự gen thế hệ mới (NGS)



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

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

De novo assembly

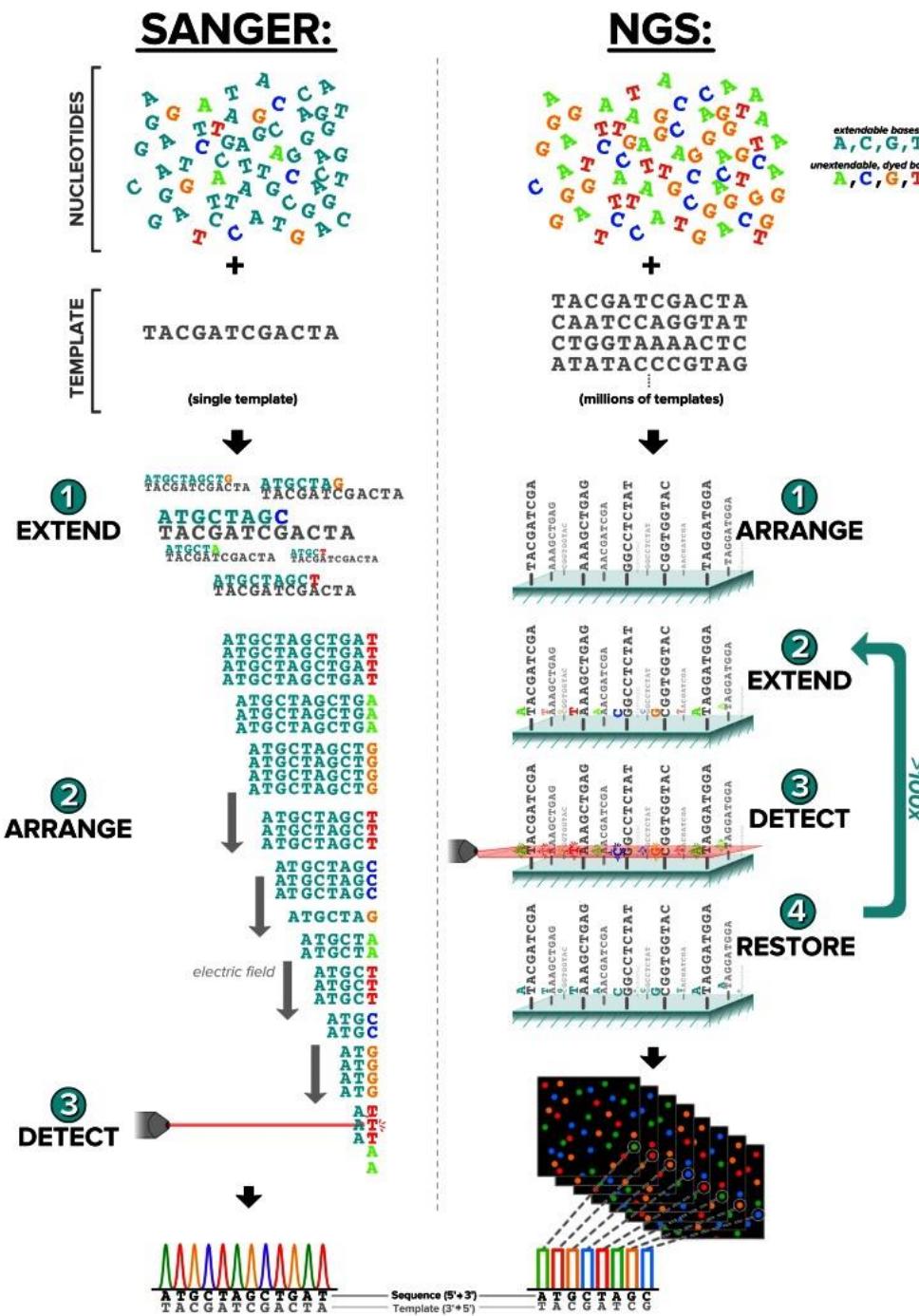


Mapping to reference



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

Read1 : CTCGAATACG

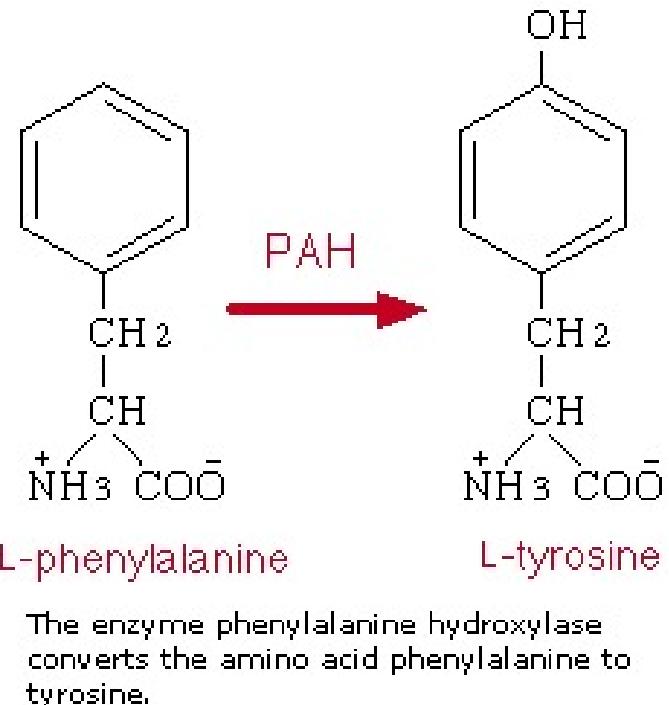


Read1 : CTCGAATACG
 Read2 : CTCGAATACG
 Read3 : CTCGAATACG
 Read4 : CTCGAATACG
 Read5 : CGCGAATACG
 Read6 : CGCGAATACG
 Read7 : CGCGACTACG
 Read8 : CGCGAATACG

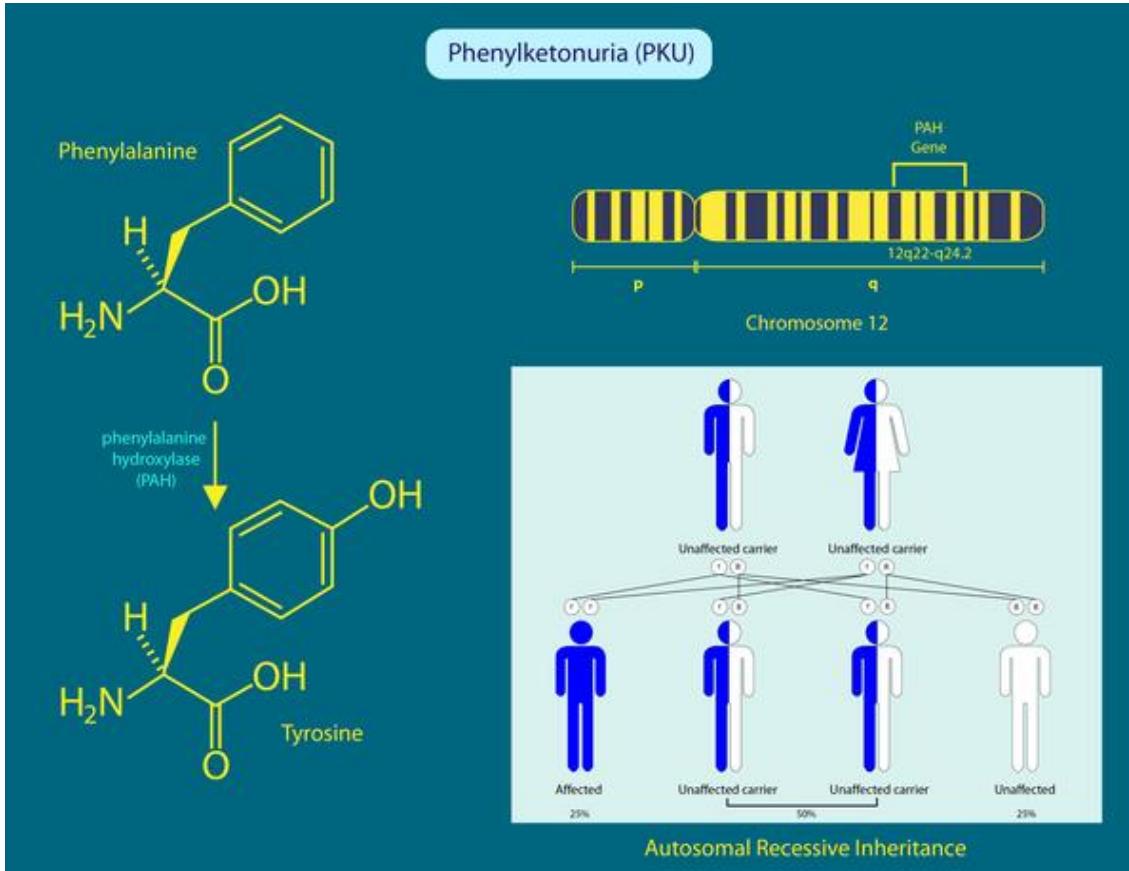
Ví dụ mối quan hệ của
BIẾN THỂ GEN và **BỆNH DI TRUYỀN**

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 ctcaaagaagaagtgggcattggccaaagtattgcgttatttggaggagaatgtat aacctgaccacattgaatcttagacccctcgtttaaagaaaagatgagatgaattttc acccatttgataaaacgttagcctgcgtctgacaaaatcatcaagatcttgggcat gacattggtgcactgtccatgagcttcacgagataagaagaaaagacagtcggctgg ttcccaagaaccattcaagagctggacagatggccaatcagattctcgtatggagcg gaactggatgctgaccaccctggtttaaagatccgtgtaccgtgcaagacggaaagcg tttgcgtacattgcctacaactaccgcattggcagccccatccctcgagtggaaatacatg gaggaagaaaaagaaaatggggcacagtgttcaagactctgaagtccctgtataaaaacc catgttgcgtatgagtacaatcacattttccacttctgaaaagtactgtggcttccat gaagataacattccccagctggaaagacgttctcgttccctgcagacttgcactggtttc cgccctccgacacctgtggctggcctgtccctcgggatcttggggcccttc cgagtctccactgcacacagtacatcagacatggatccaagccatgtataccccgaa cctgacatctgcatgagctgtggacatgtgcccttgcggatctggcggccatgt cagtttccaggaaattggccctgcctctgggtgcacctgtatgaaatacatgtaaaag ctcgccacaatttactgtggtagttggctctgcaaaacaaggagactccata aaggcatatggtgcggccctgtcatccttggtaattacagtactgttattcagag aagccaaagcttctcccccggagctggagaagacagccatccaaaattacactgtcag gagttccagcccttattacgtggcagagatgtttaatgtatgccaaggagaaaagtaagg aactttgctgccacaataccctcgcccttctcgttgcgtacgacccatcacccaaagg attggaggcttggacaatacccagcagcttaagatggctgattccattaacagtgaa attggaaatcccttgcagtggccctccagaaaataaagtaa
--------	--

Protein (Amino Acid)

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

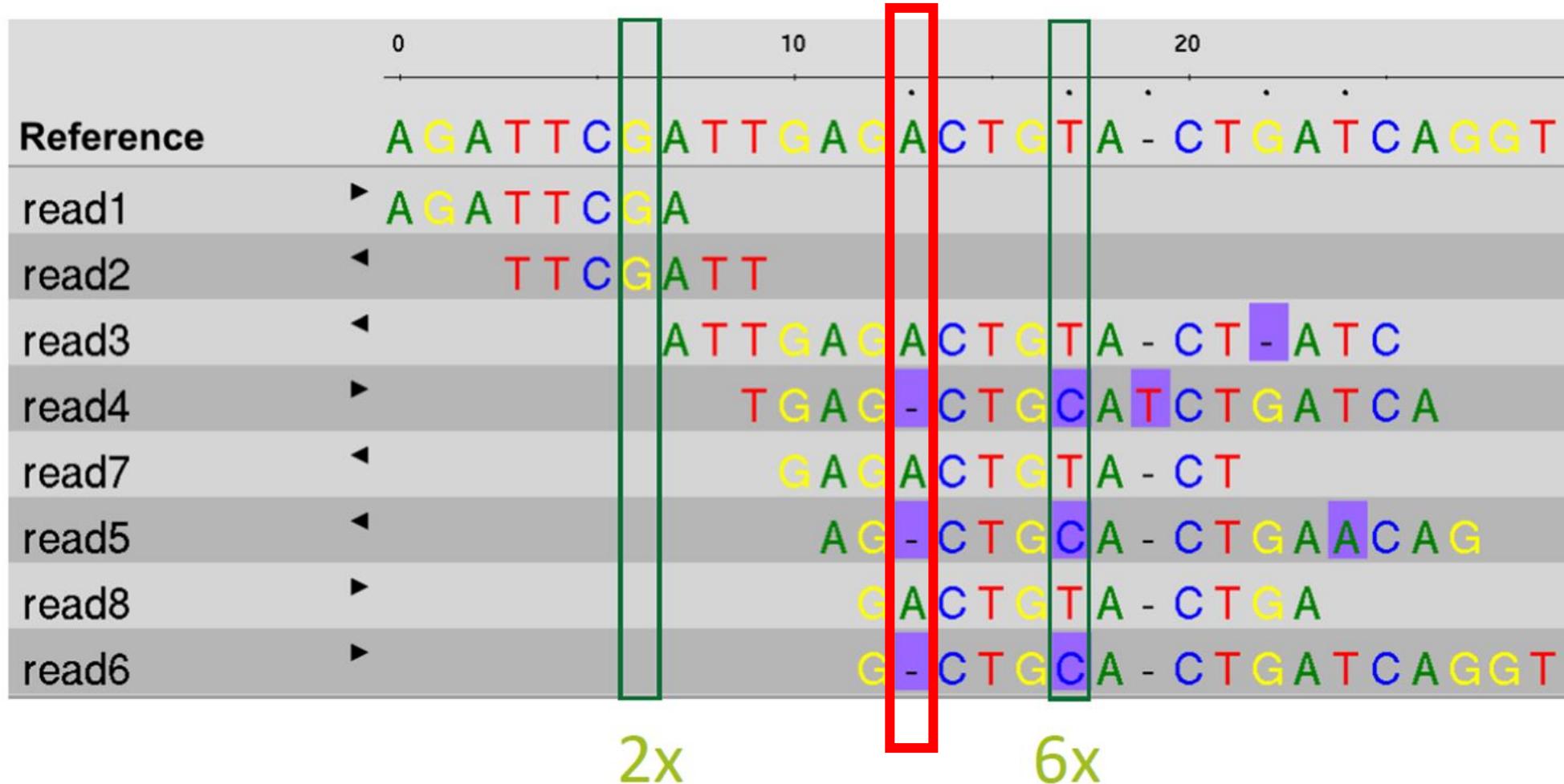
<https://www.genome.jp/entry/T01001:5053>

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** ?

About Access Submit Stats FTP Help Advanced search 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](#)

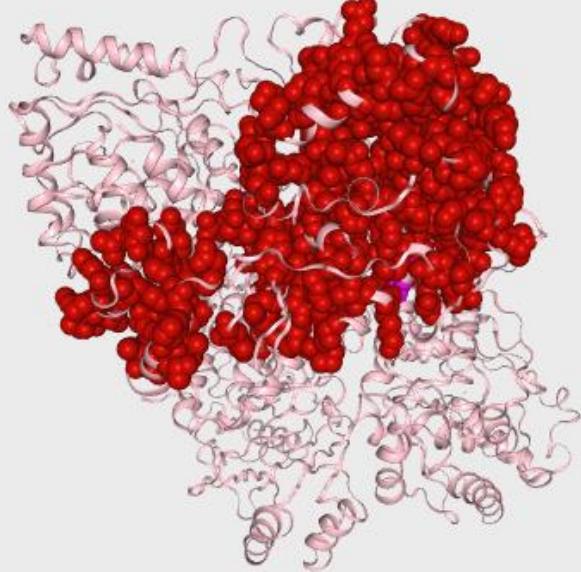
[View on SwissModel](#)

Transcript: ENST00000553106.6

Select protein structure

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

X

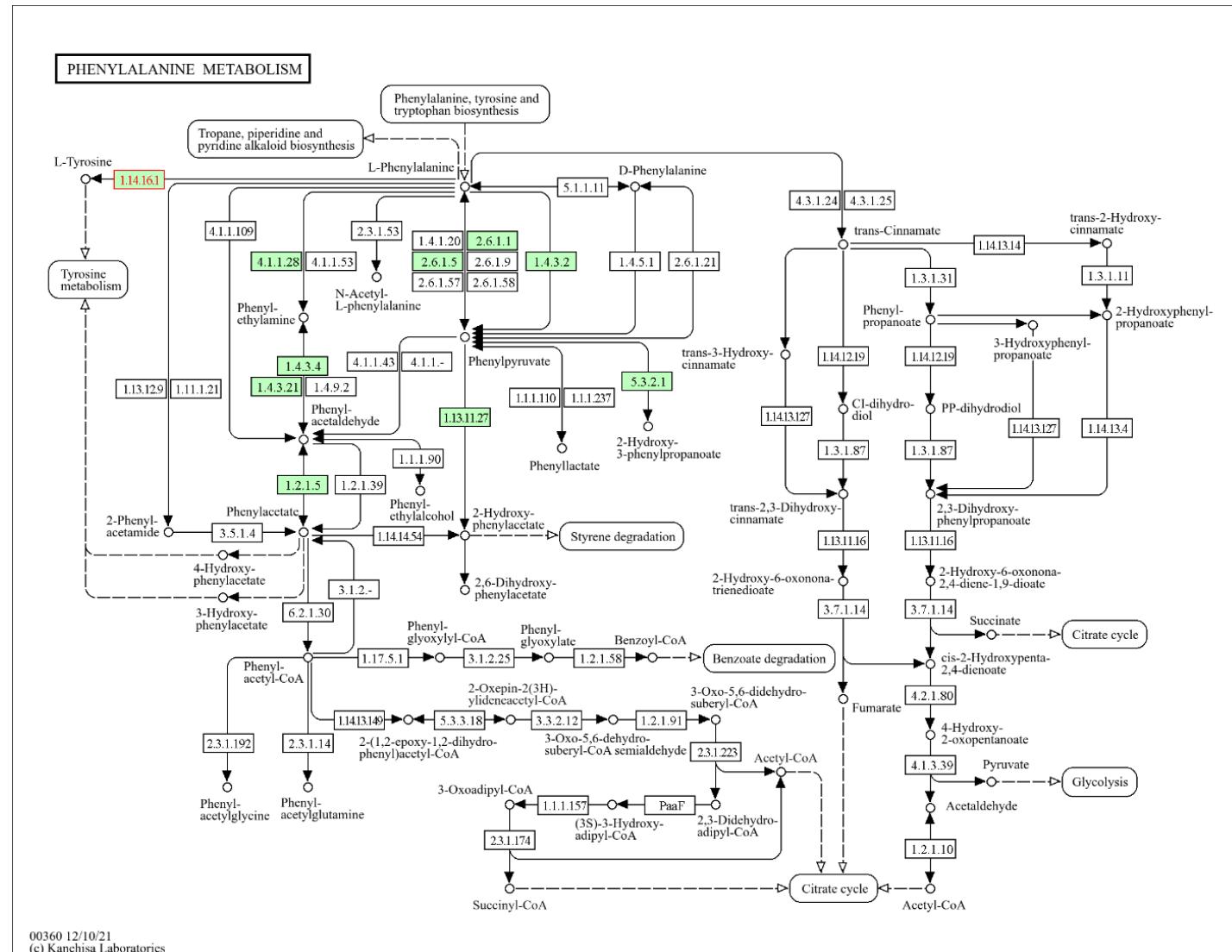


Show

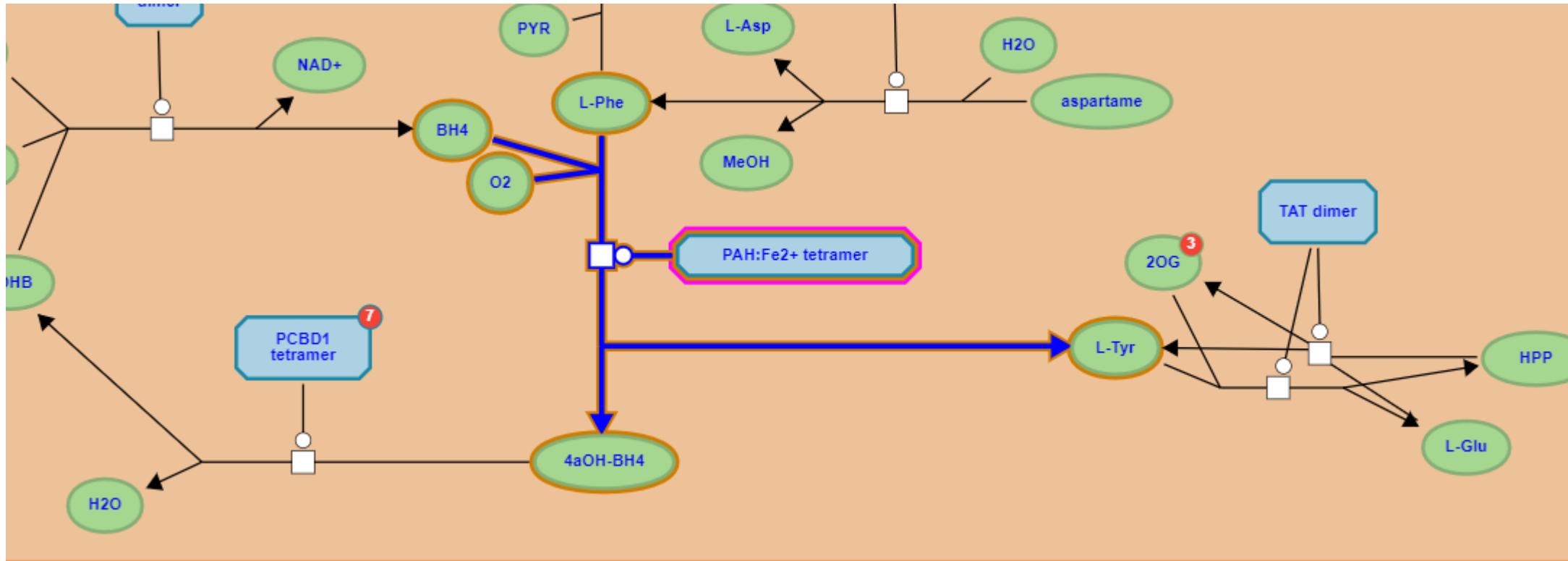
- All Residues
- Variants
- Pathogenic
- Likely Pathogenic
- Uncertain Significance
- Likely Benign
- Benign
- Current Variant

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

PAH: chuyển hóa Phenylalanine thành Tyrosine

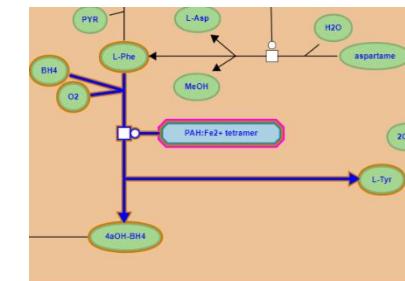
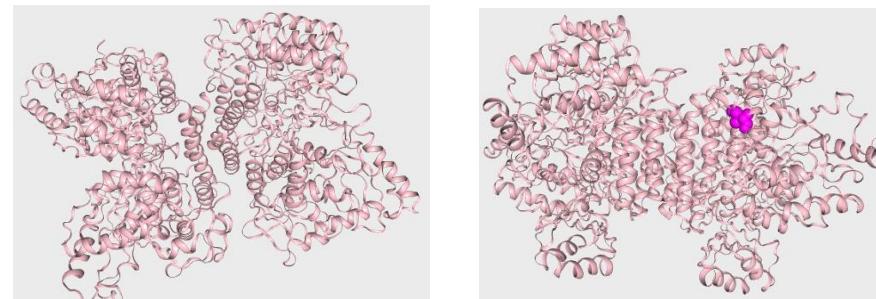
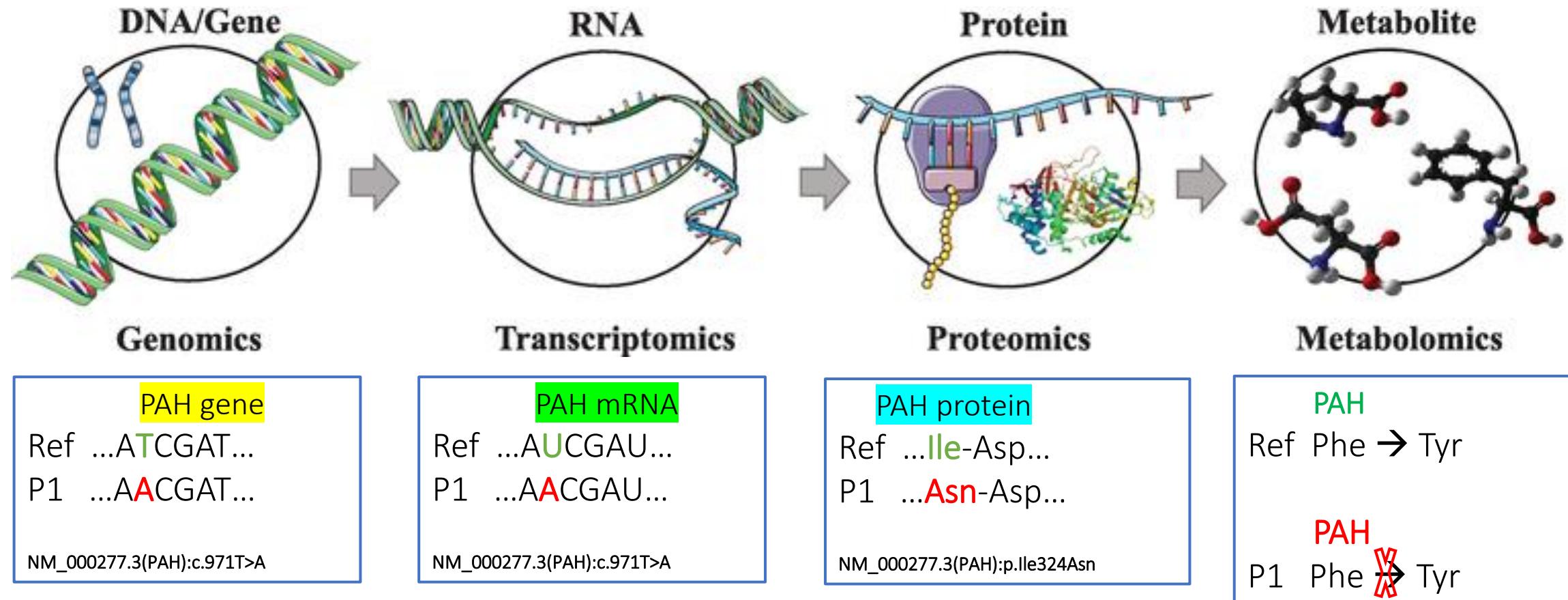


PAH: chuyển hóa Phenylalanine thành Tyrosine

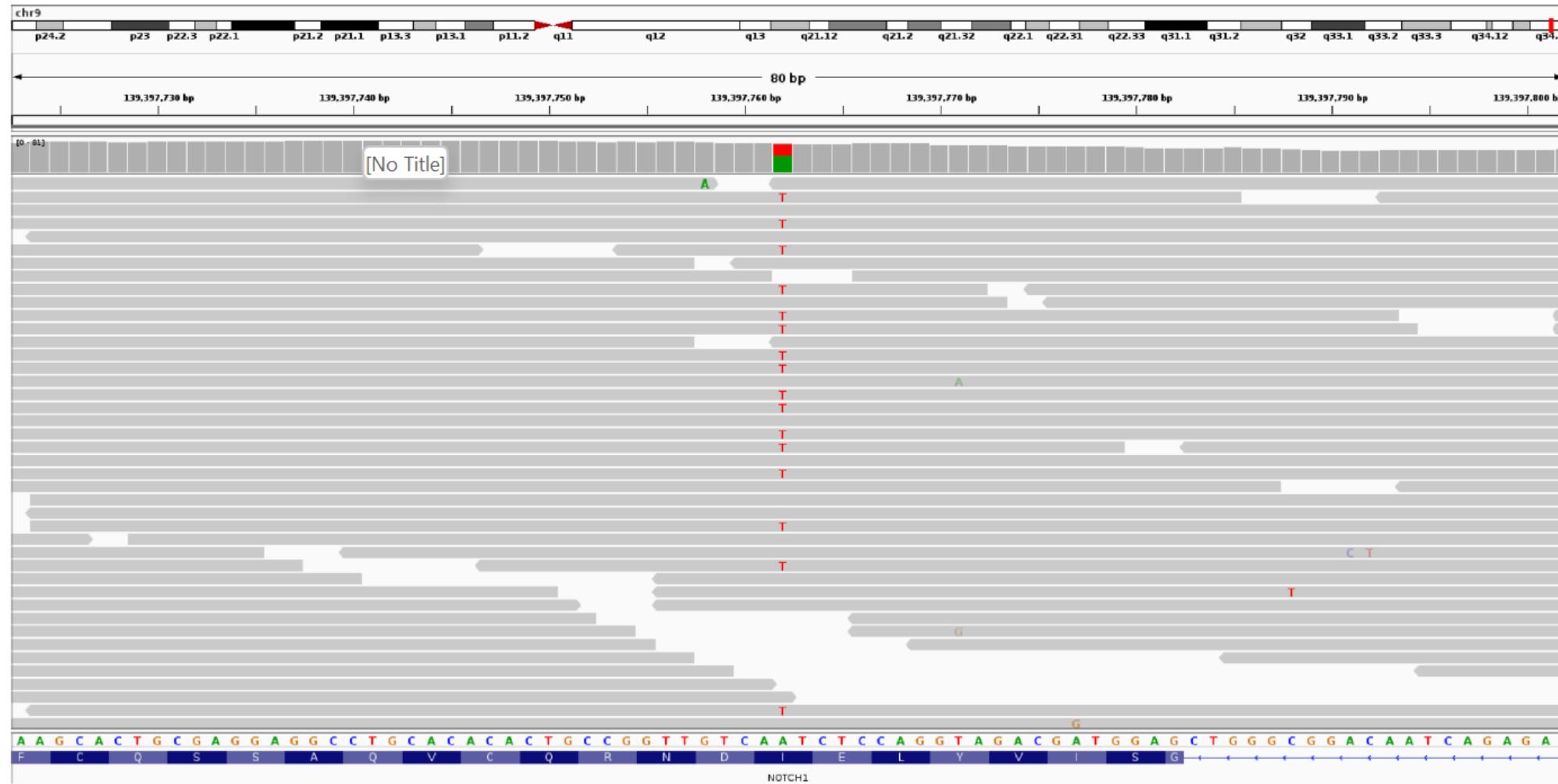


<https://reactome.org/PathwayBrowser/#/R-HSA-8963691&SEL=R-HSA-71118&PATH=R-HSA-1430728,R-HSA-71291&FLG=UniProt:P00439>

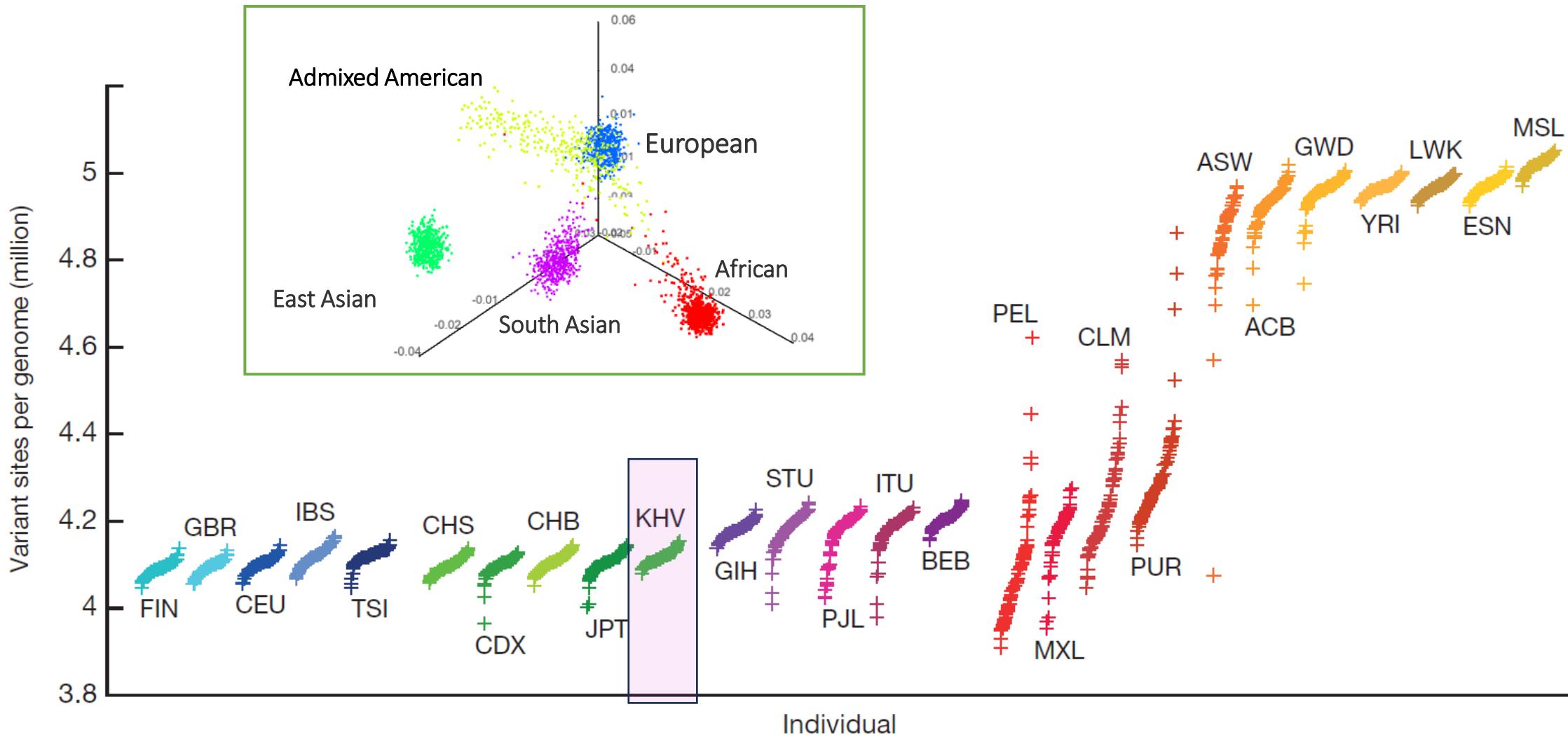
Mối liên kết: Biến thể gen và bệnh di truyền



Alignment and variant viewers

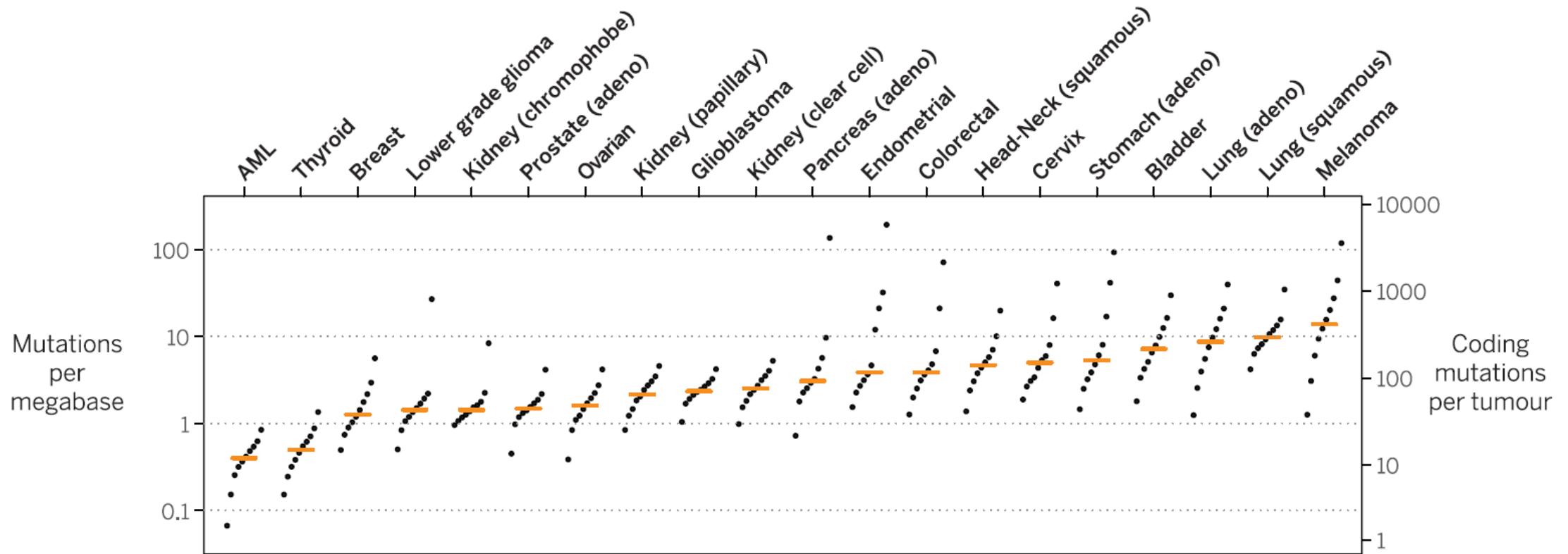


Human Genome Variation: 1000 Genomes Project



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

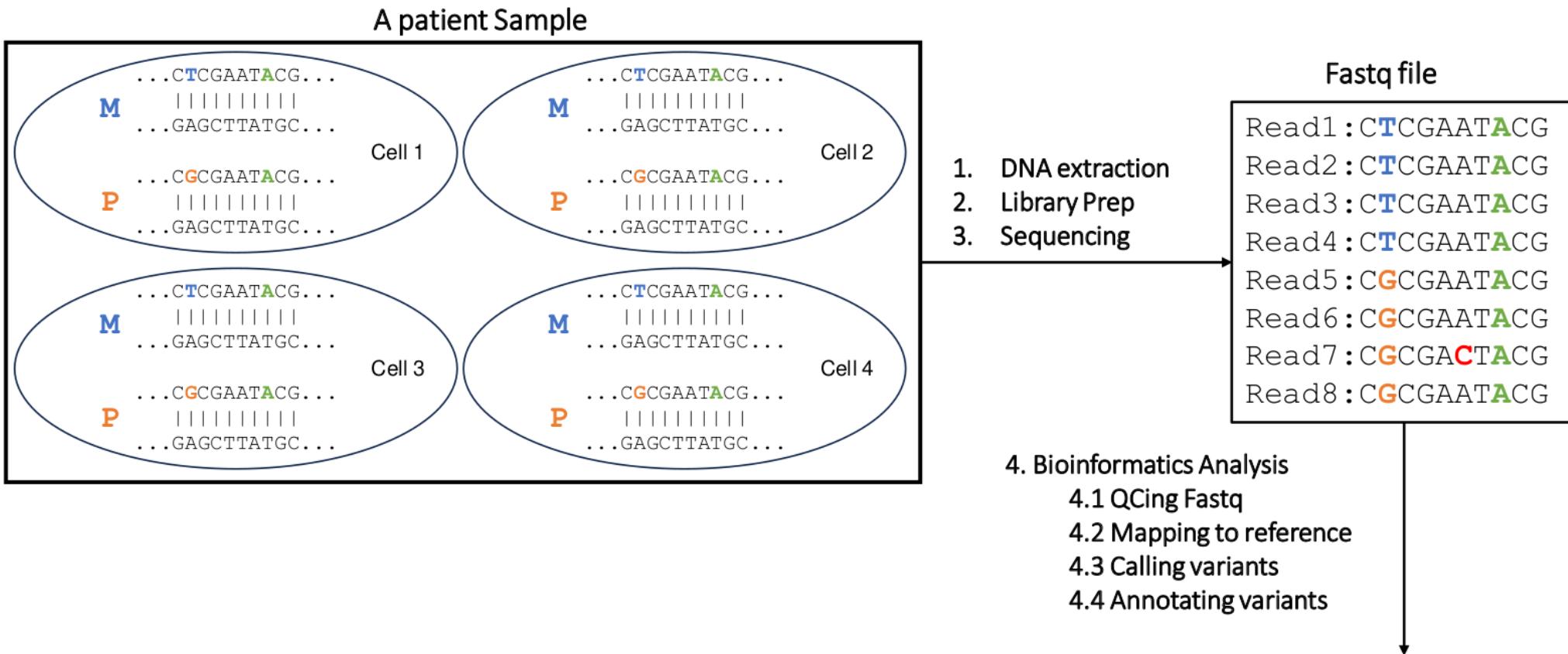
Cancer Genome Somatic Variation



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)

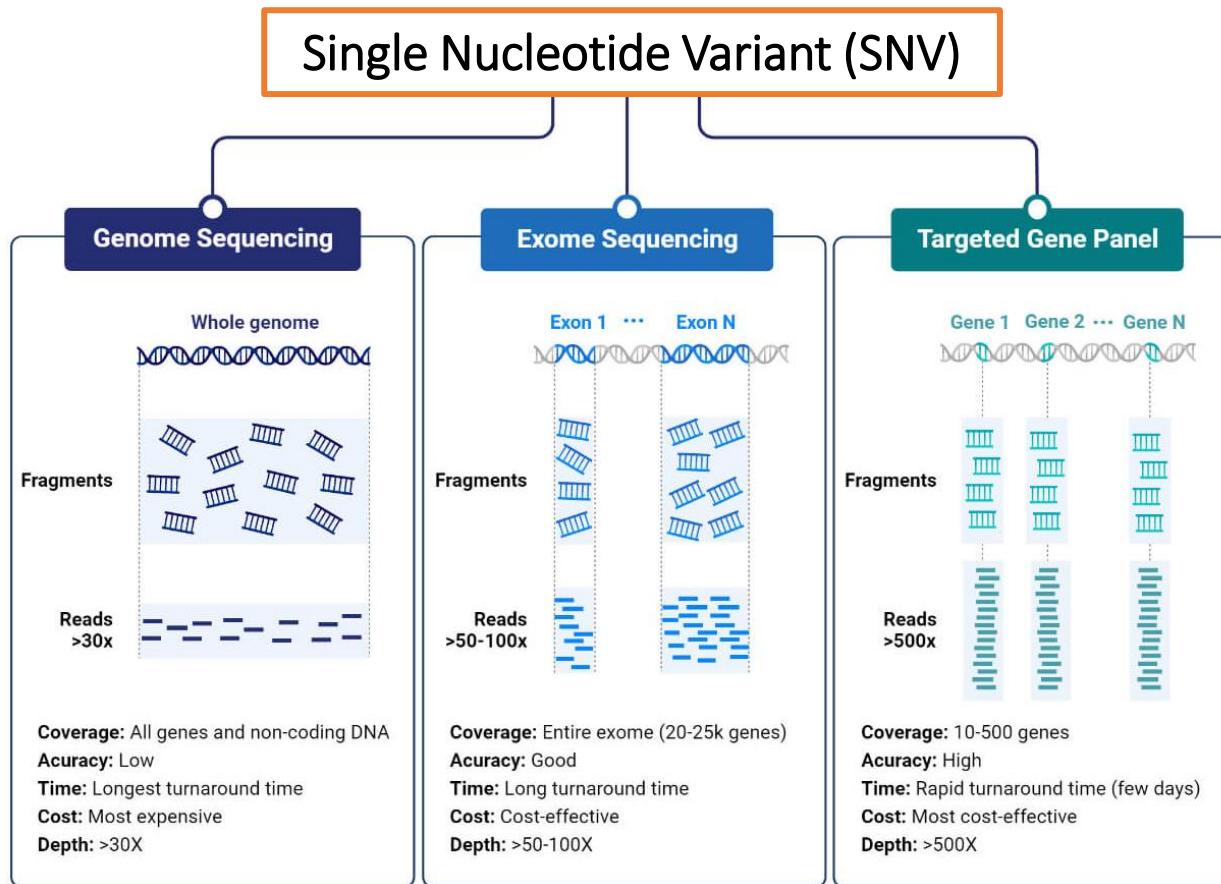
4.4 Annotating variants

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

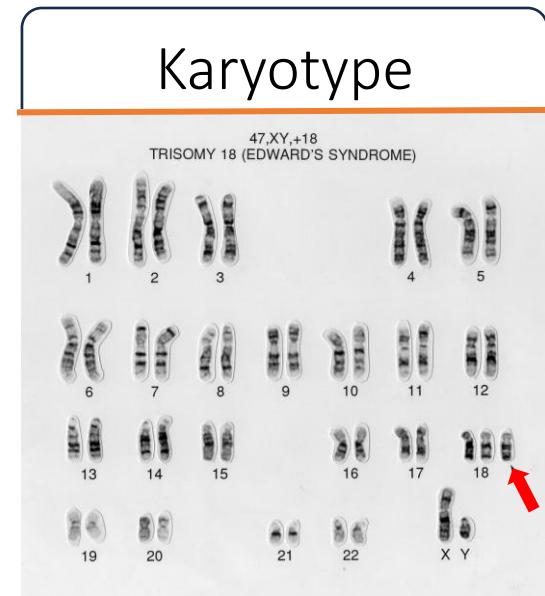
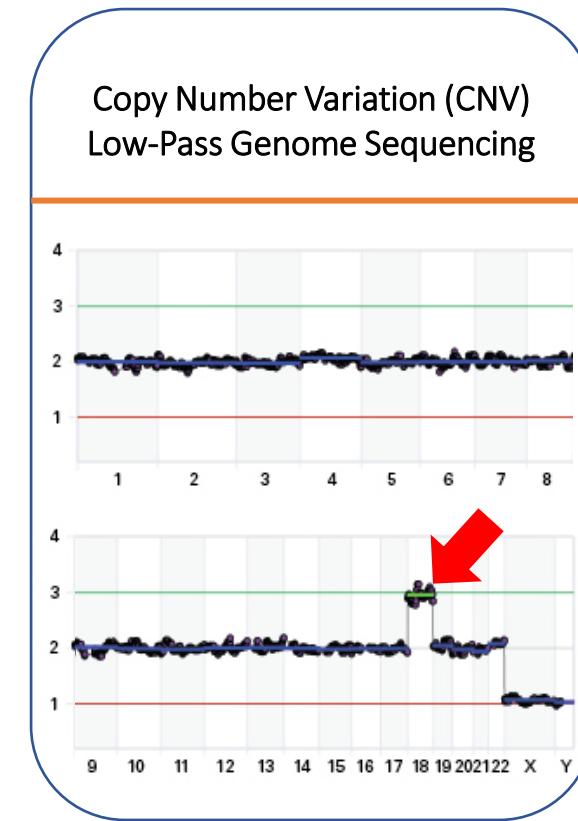


<https://ggb.a.swiss/en/the-first-medical-genomics-center-opens-in-geneva/>

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#>

Data storage

WGS_Level_1

```
|── FR07885321_1 (sample)
|   └── H3T75CCXX_1_150416_FR07885321_Human__R_150219 (lane)
|       ├── H3T75CCXX_1_150416_FR07885321_Human__R_150219_R1.fastq.gz.md5
|       └── H3T75CCXX_1_150416_FR07885321_Human__R_150219_R1.fastq.gz
```

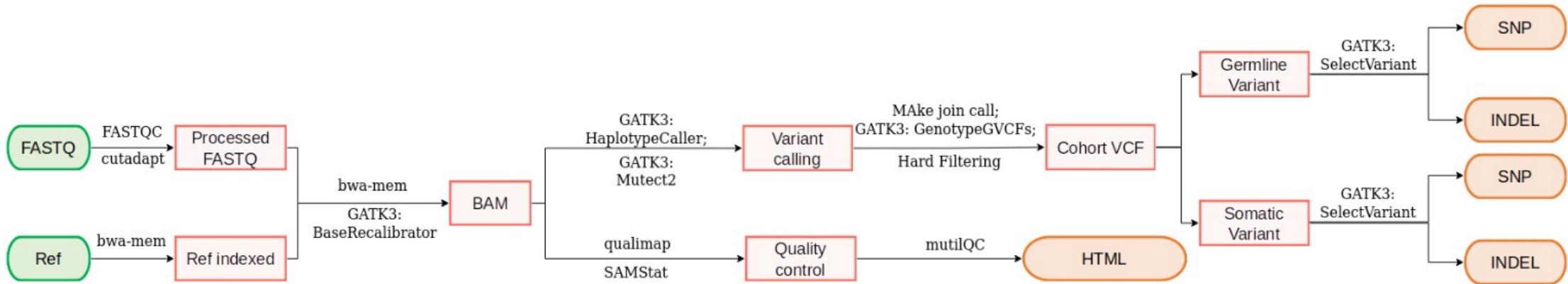
WGS_Level_2

```
|── FR07885321_1 (sample)
|   └── H3T75CCXX_1_150416_FR07885321_Human__R_150219 (lane)
|       ├── H3T75CCXX_1_150416_FR07885321_Human__R_150219.bam.md5
|       └── H3T75CCXX_1_150416_FR07885321_Human__R_150219.bam
```

WGS_Level_3

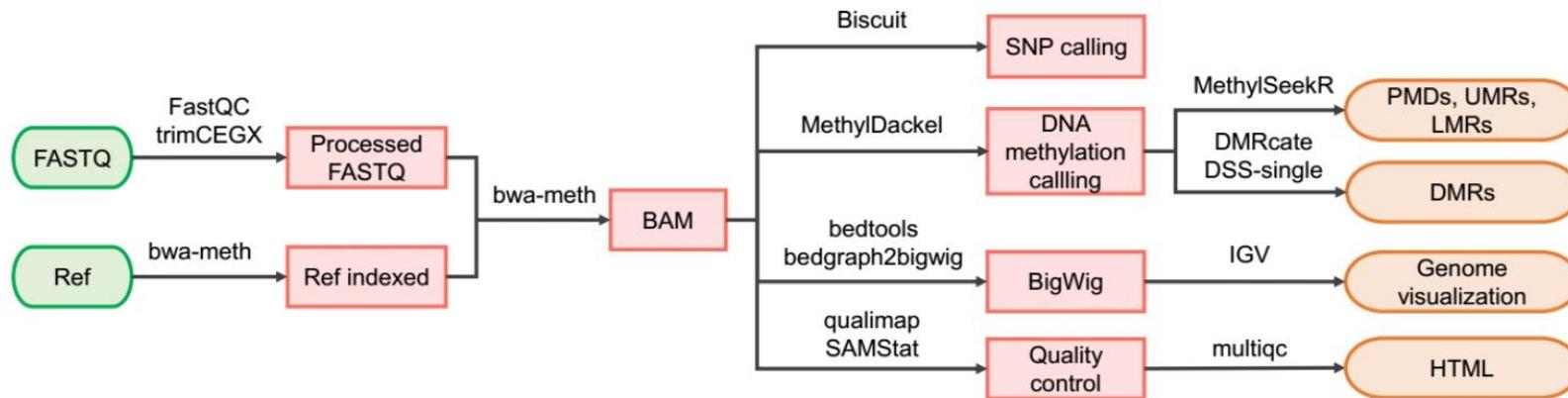
```
|── FR07885321_1 (sample)
|   └── FR07885321_1.bam
|   └── FR07885321_1.bam.md5
|   └── FR07885321_1.GenotypeGVCFs.vcf.gz
```

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/

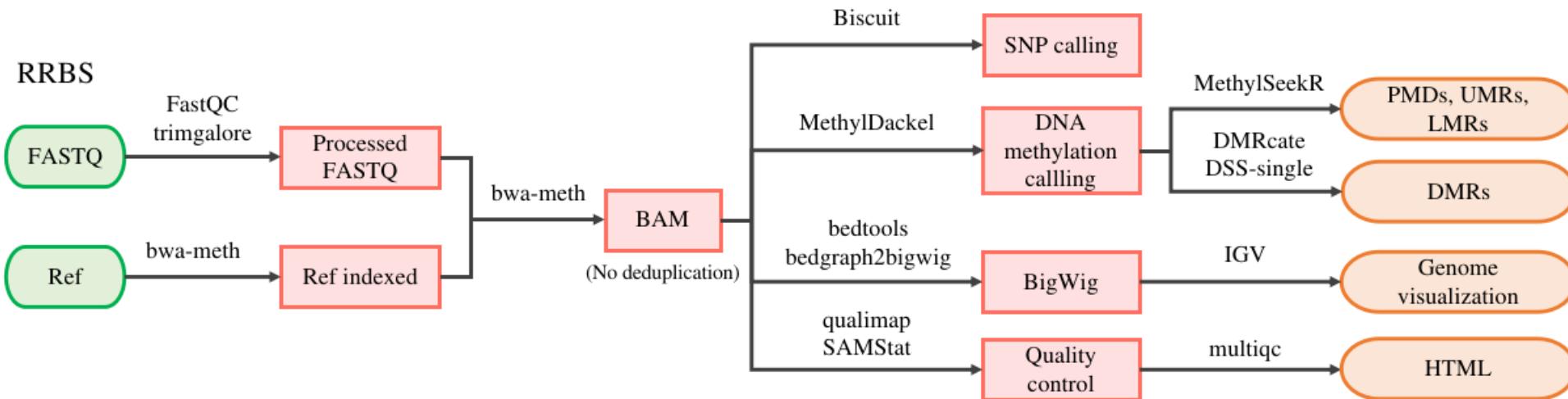
Whole Genome Bisulphite Sequencing pipeline



Package	Version	Link
FastQC	0.11.9	https://www.bioinformatics.babraham.ac.uk/projects/fastqc/
trimCEGX	1.0	https://github.com/luuloi/trim.paired.read
bwa-meth	0.10	https://github.com/brentp/bwa-meth
Biscuit	0.2.0	https://github.com/zwdzwd/biscuit
MethylDackel	0.2.0	https://github.com/dpryan79/MethylDackel
qualimap	2.2.1	http://qualimap.bioinfo.cipf.es/
SAMStat	1.08	http://samstat.sourceforge.net/
MethylSeekR	1.0	https://github.com/Bioconductor-mirror/MethylSeekR
DMRcate	2.2.3	https://bioconductor.org/packages/release/bioc/html/DMRcate.html
DSS-single	2.36.0	https://bioconductor.org/packages/release/bioc/html/DSS.html
multiqc	1.8	https://multiqc.info/

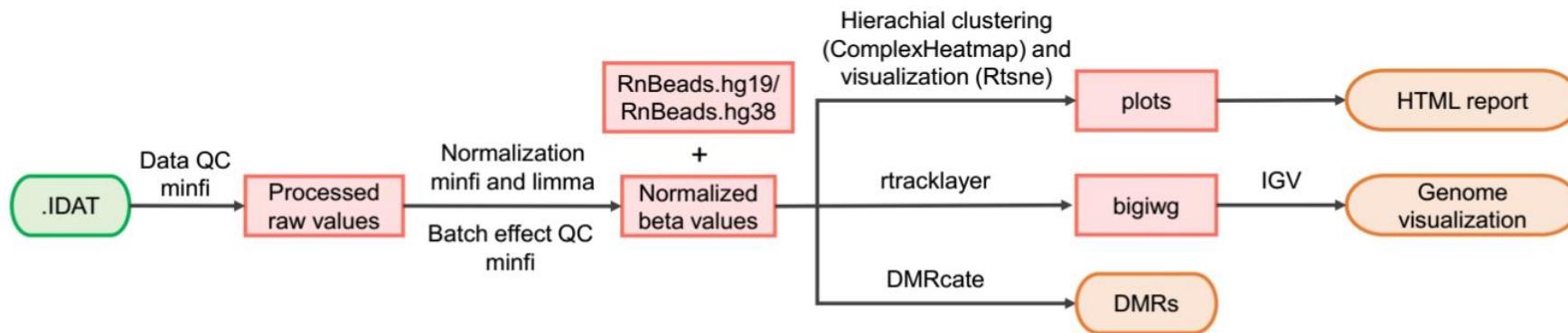
Figure 4. The WGBS pipeline and software.

Reduced-representation bisulfite sequencing (RRBS-Seq) pipeline



Package	Version	Link
FastQC	0.11.9	https://www.bioinformatics.babraham.ac.uk/projects/fastqc/
trimgalore	0.6.7	https://github.com/FelixKrueger/TrimGalore
bwa-meth	0.10	https://github.com/brentp/bwa-meth
Biscuit	0.2.0	https://github.com/zwdzwd/biscuit
MethylDackel	0.2.0	https://github.com/dpryan79/MethylDackel
qualimap	2.2.1	http://qualimap.bioinfo.cipf.es/
SAMStat	1.08	http://samstat.sourceforge.net/
MethylSeekR	1.0	https://github.com/Bioconductor-mirror/MethylSeekR
DMRcate	2.2.3	https://bioconductor.org/packages/release/bioc/html/DMRcate.html
DSS-single	2.36.0	https://bioconductor.org/packages/release/bioc/html/DSS.html
multiqc	1.8	https://multiqc.info/

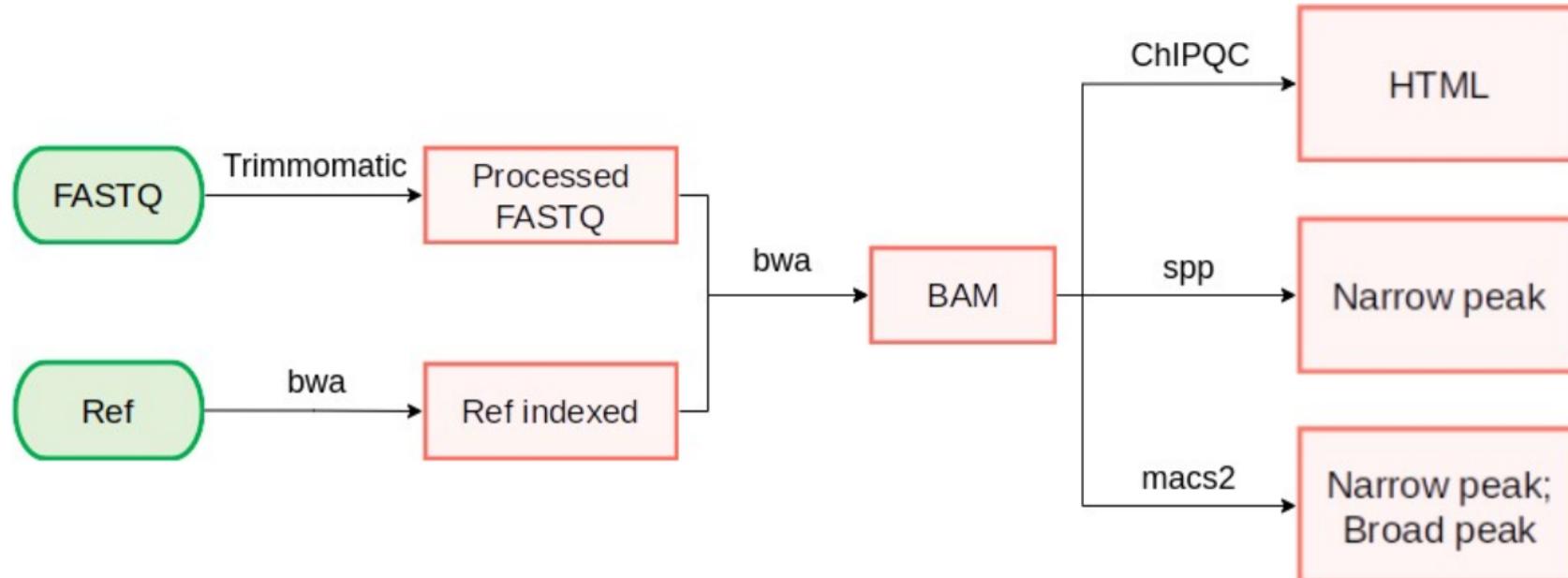
DNA Methylation Array (450K and EPIC/850K) pipeline



Package	Version	Link
minfi	1.34.0	http://bioconductor.org/packages/release/bioc/html/minfi.html
limma	3.44.3	https://bioconductor.org/packages/release/bioc/html/limma.html
RnBeads.hg19	1.20.0	https://bioconductor.org/packages/release/data/experiment/html/RnBeads.hg19.html
RnBeads.hg38	1.20.0	https://bioconductor.org/packages/release/data/experiment/html/RnBeads.hg38.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
DMRcate	2.2.3	https://bioconductor.org/packages/release/bioc/html/DMRcate.html

Figure 5. The DNA methylation array (Illumina Methylation 450K and 850K) pipeline and software.

ChIP-seq pipeline



Package	Version	Link
Trimmomatic	0.39	https://bioweb.pasteur.fr/packages/pack@Trimmomatic@0.39
bwa	0.7.17	https://github.com/lh3/bwa
ChIPQC	3.14	https://bioconductor.org/packages/release/bioc/html/ChIPQC.html
spp	2.0.1	http://xinwang2hms.github.io/SPP/
macs2	2.2.7.1	https://doi.org/10.1186/gb-2008-9-9-r137

HiC-seq pipeline: HiC-Pro

Software | Open Access | Published: 01 December 2015

HiC-Pro: an optimized and flexible pipeline for Hi-C data processing

Nicolas Servant Nelle Varoquaux, Bryan R. Lajoie, Eric Viara, Chong-Jian Chen, Jean-Philippe Vert, Edith Heard, Job Dekker & Emmanuel Barillot

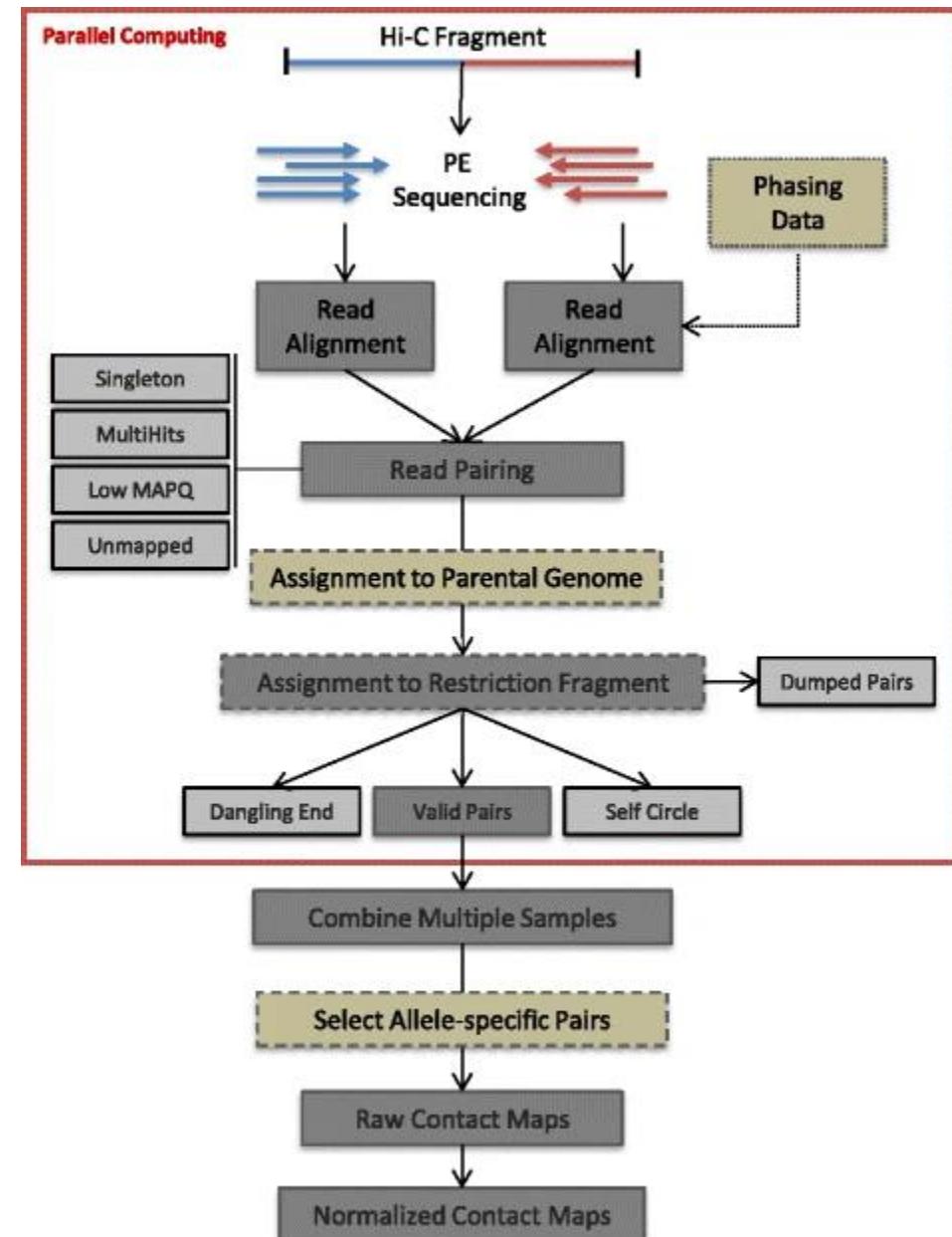
Genome Biology 16, Article number: 259 (2015) | [Cite this article](#)

46k Accesses | 687 Citations | 13 Altmetric | [Metrics](#)

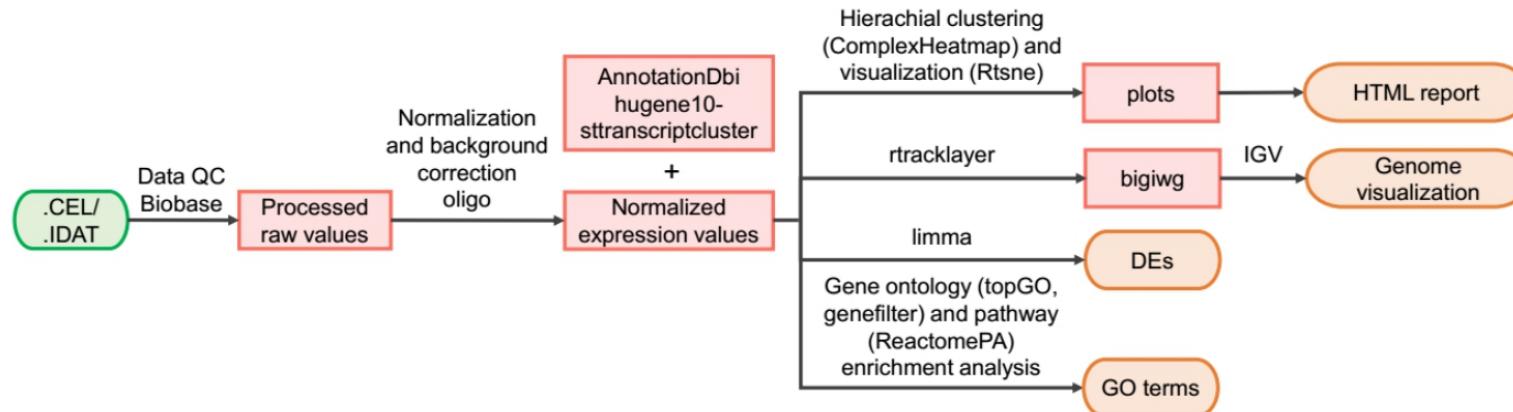
Abstract

HiC-Pro is an optimized and flexible pipeline for processing Hi-C data from raw reads to normalized contact maps. HiC-Pro maps reads, detects valid ligation products, performs quality controls and generates intra- and inter-chromosomal contact maps. It includes a fast implementation of the iterative correction method and is based on a memory-efficient data format for Hi-C contact maps. In addition, HiC-Pro can use phased genotype data to build allele-specific contact maps. We applied HiC-Pro to different Hi-C datasets, demonstrating its ability to easily process large data in a reasonable time.

Source code and documentation are available at <http://github.com/nservant/HiC-Pro>.



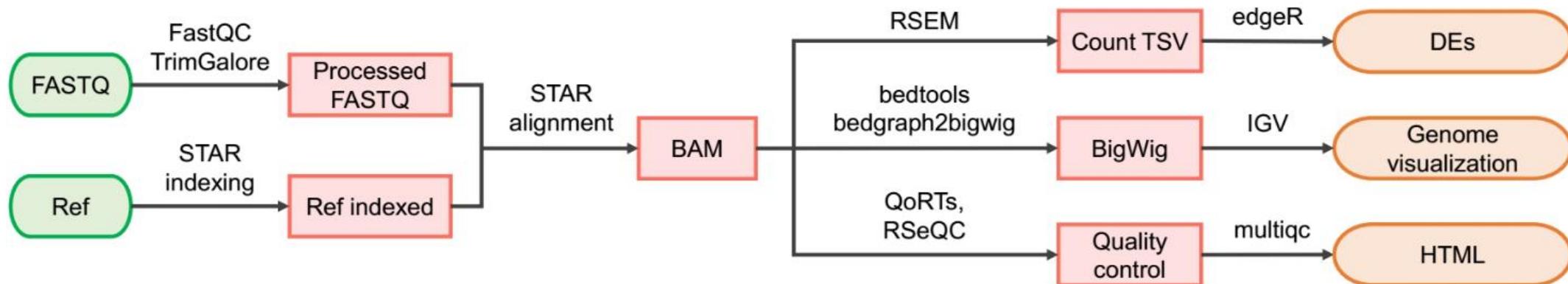
Gene Expression Microarray (Affymetrix) pipeline



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.

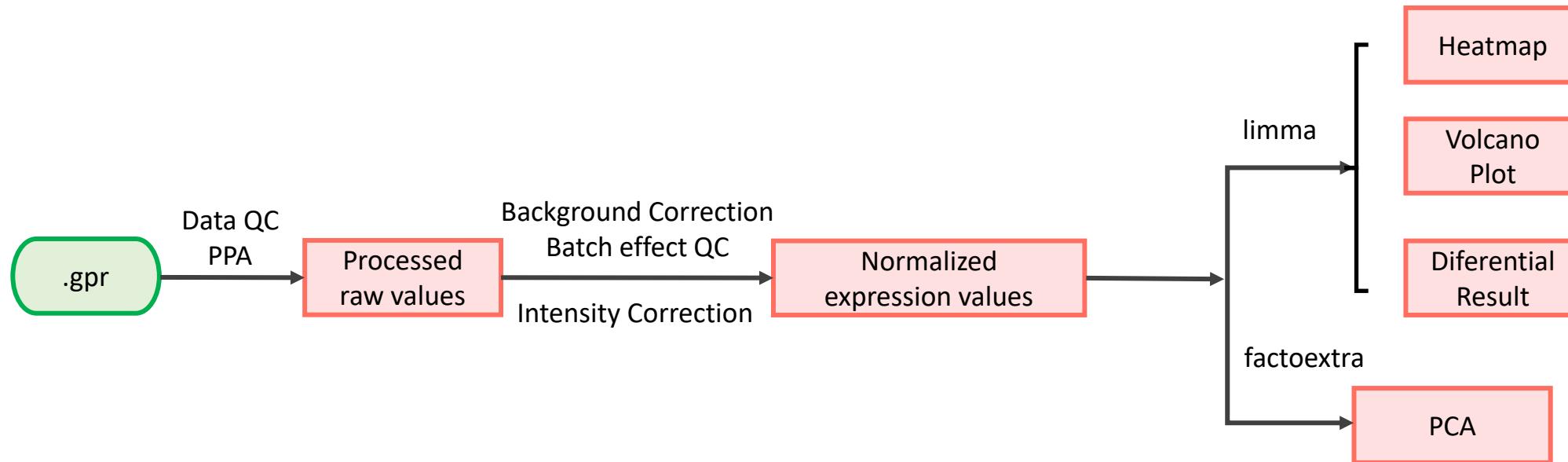
Bulk RNA-seq pipeline



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.

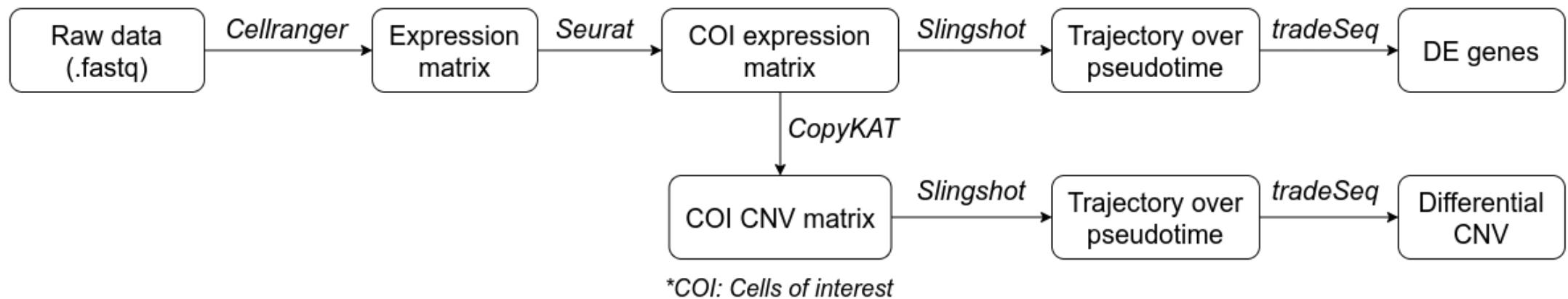
Protein Microarray Expression (ProtoArray) pipeline



Package	Version	Link
PPA	1.7.1	https://bioconductor.org/packages/release/bioc/vignettes/PAA/inst/doc/PAA_1.7.1.pdf
limma	3.44.3	https://bioconductor.org/packages/release/bioc/html/limma.html
Complex-Heatmap	2.4.3	https://www.bioconductor.org/packages/release/bioc/html/ComplexHeatmap.html
factoextra	1.0.7	https://cran.r-project.org/web/packages/factoextra/index.html

scRNA-seq

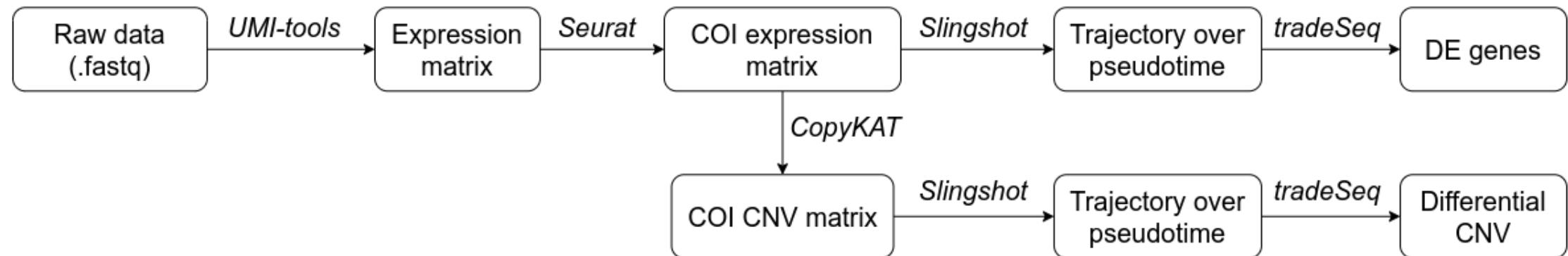
10X Chromium pipeline



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



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

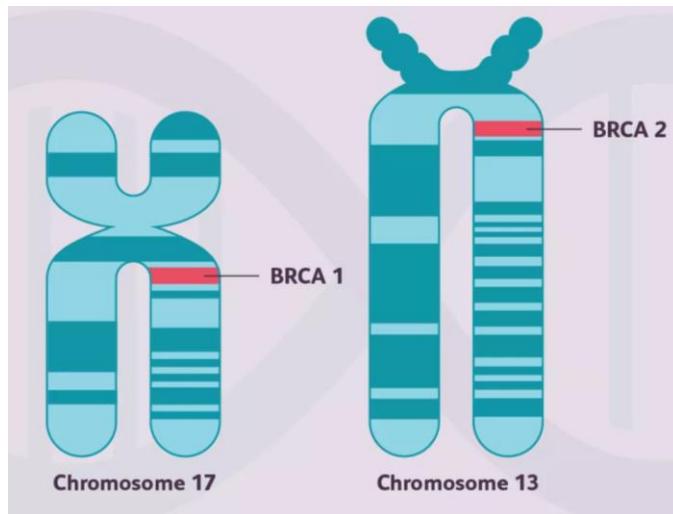
Ví dụ 2: Sàng lọc bệnh ung thư bằng xét nghiệm gen Gene Panel – 53 genes của MGI

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

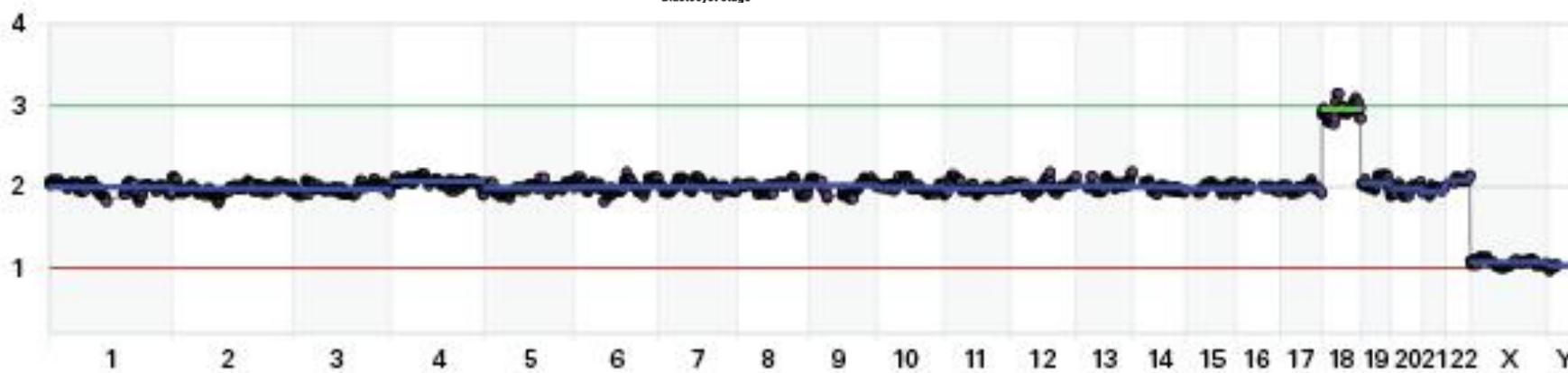
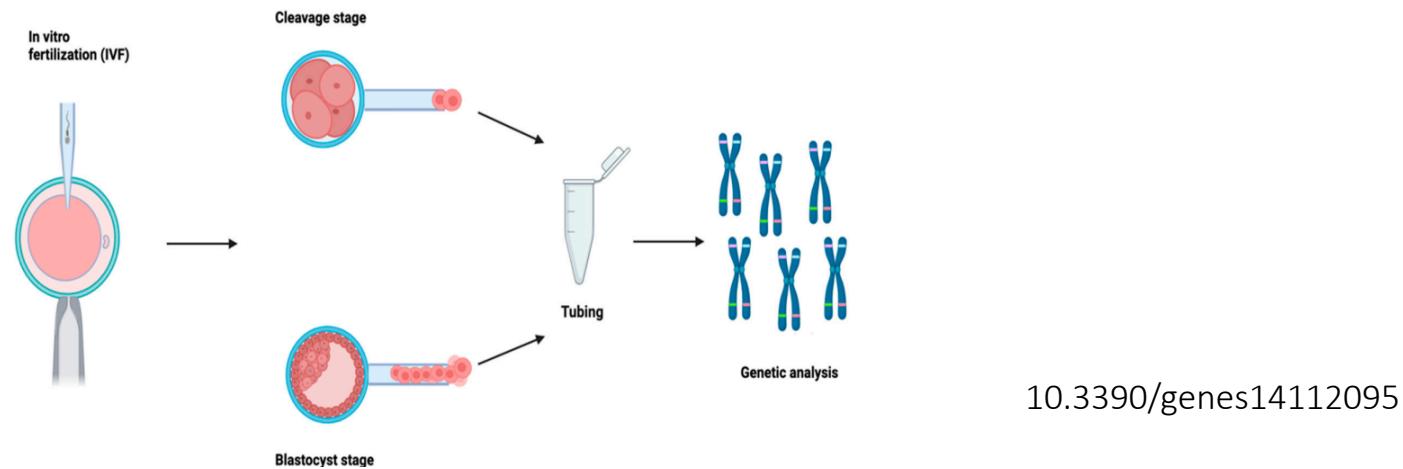
- Hệ thống xét nghiệm Illumina MiSeqDx (CE/US-IVD)
- Bộ xét nghiệm: NGeneBio BRCAaceuTest™Plus (CE-IVD)
- Phần mềm hỗ trợ phân tích kết quả: NGeneBio NGeneAnalySys™ (CE-IVD)

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

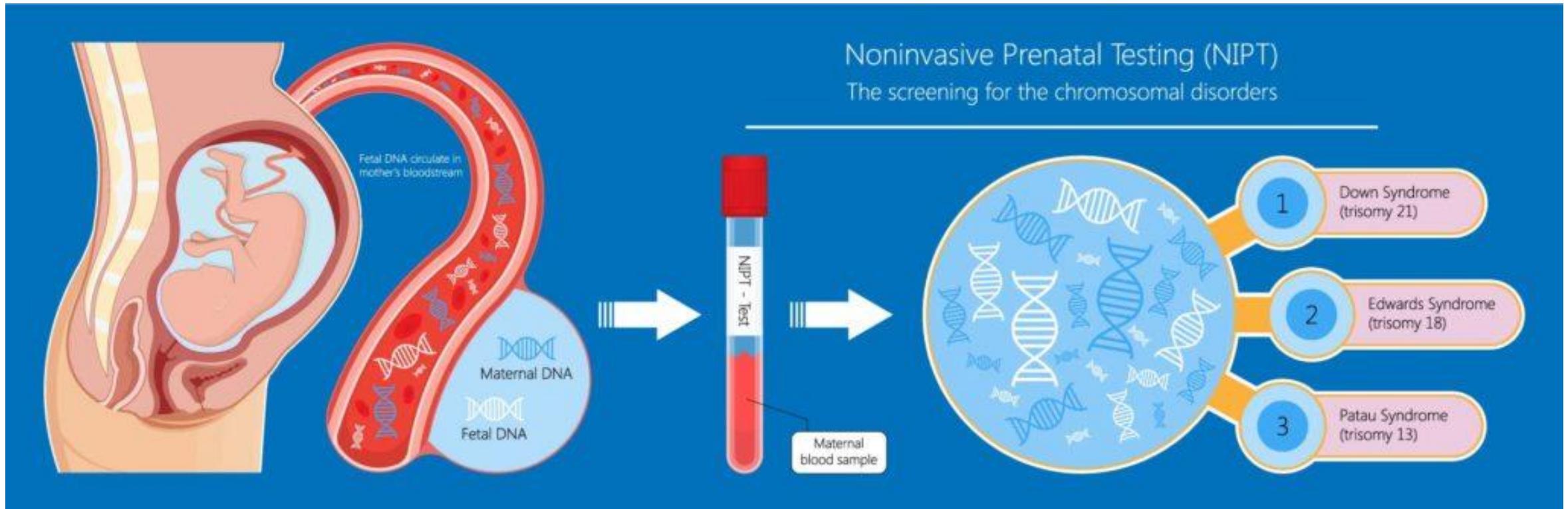
Ví dụ 4: Sàng lọc phôi trong hỗ trợ sinh sản IVF

Xét nghiệm tiền làm tổ PGT-A



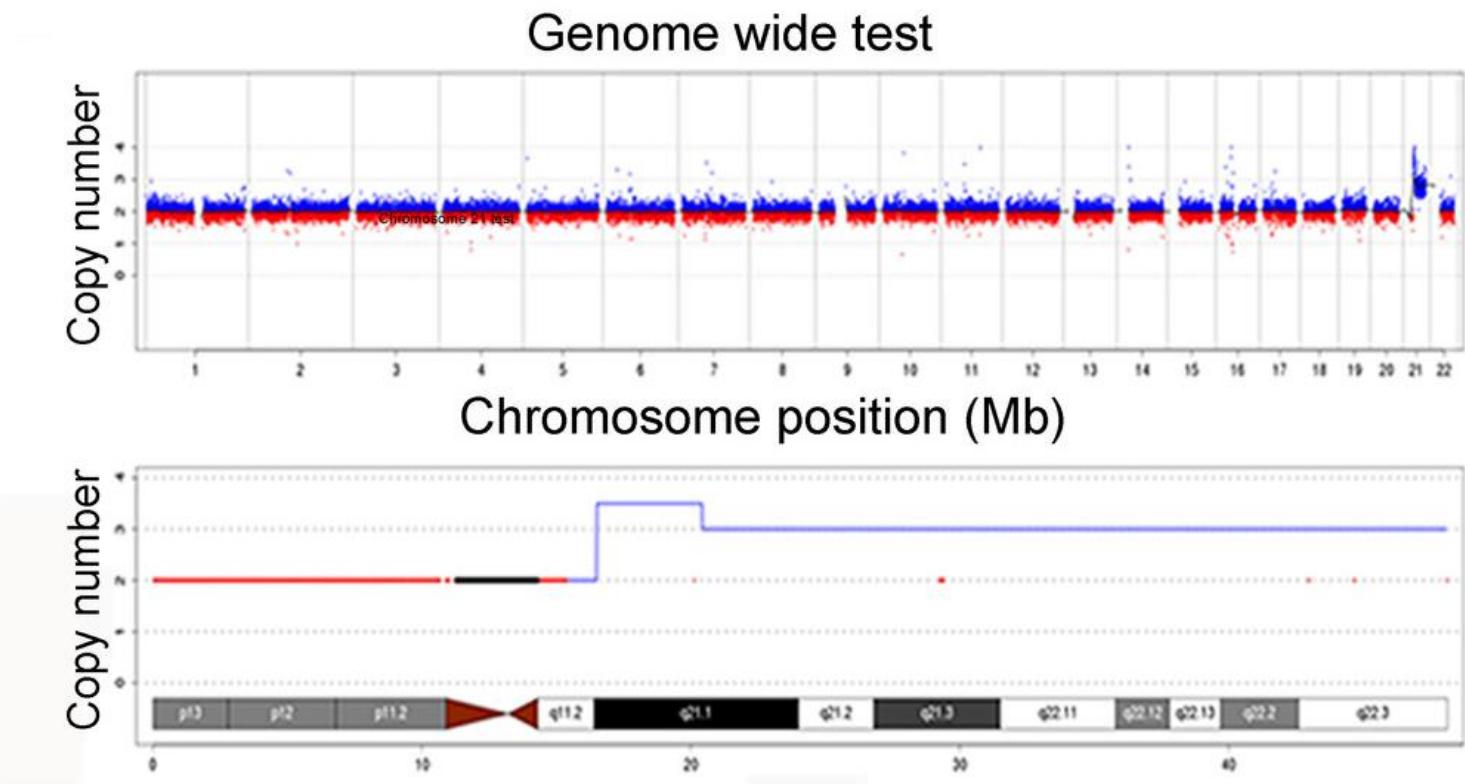
Mẫu phôi có 3 nhiễm sắc thể 18 (trisomy) trong DNA hệ gen

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



10.7717/peerj.14400

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

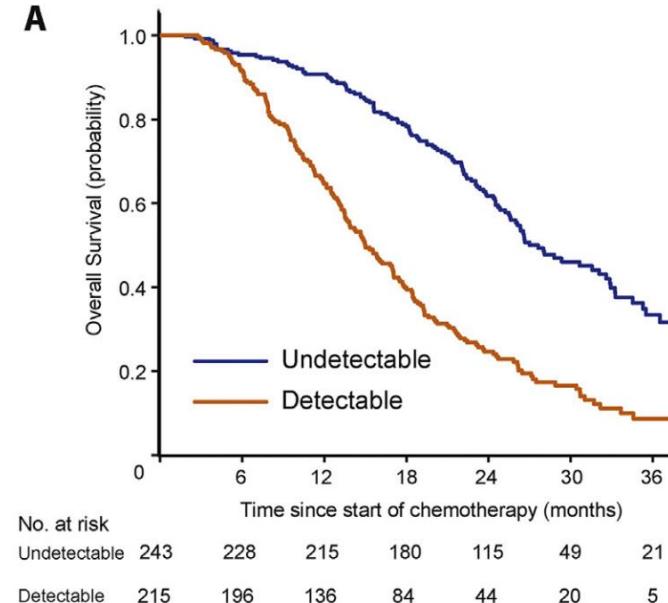
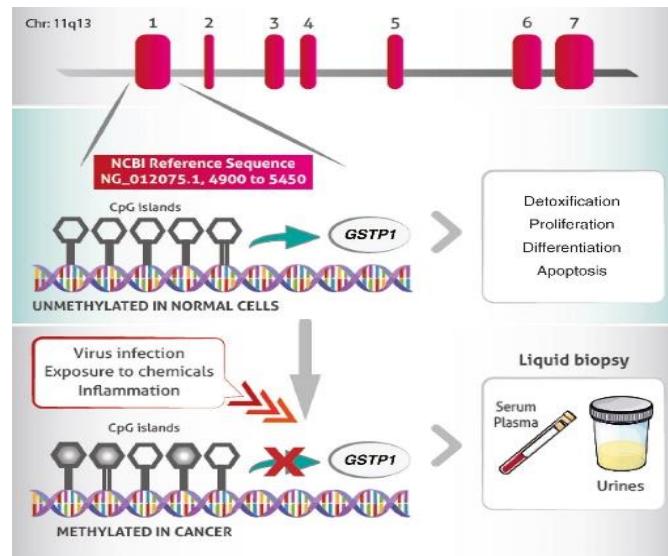
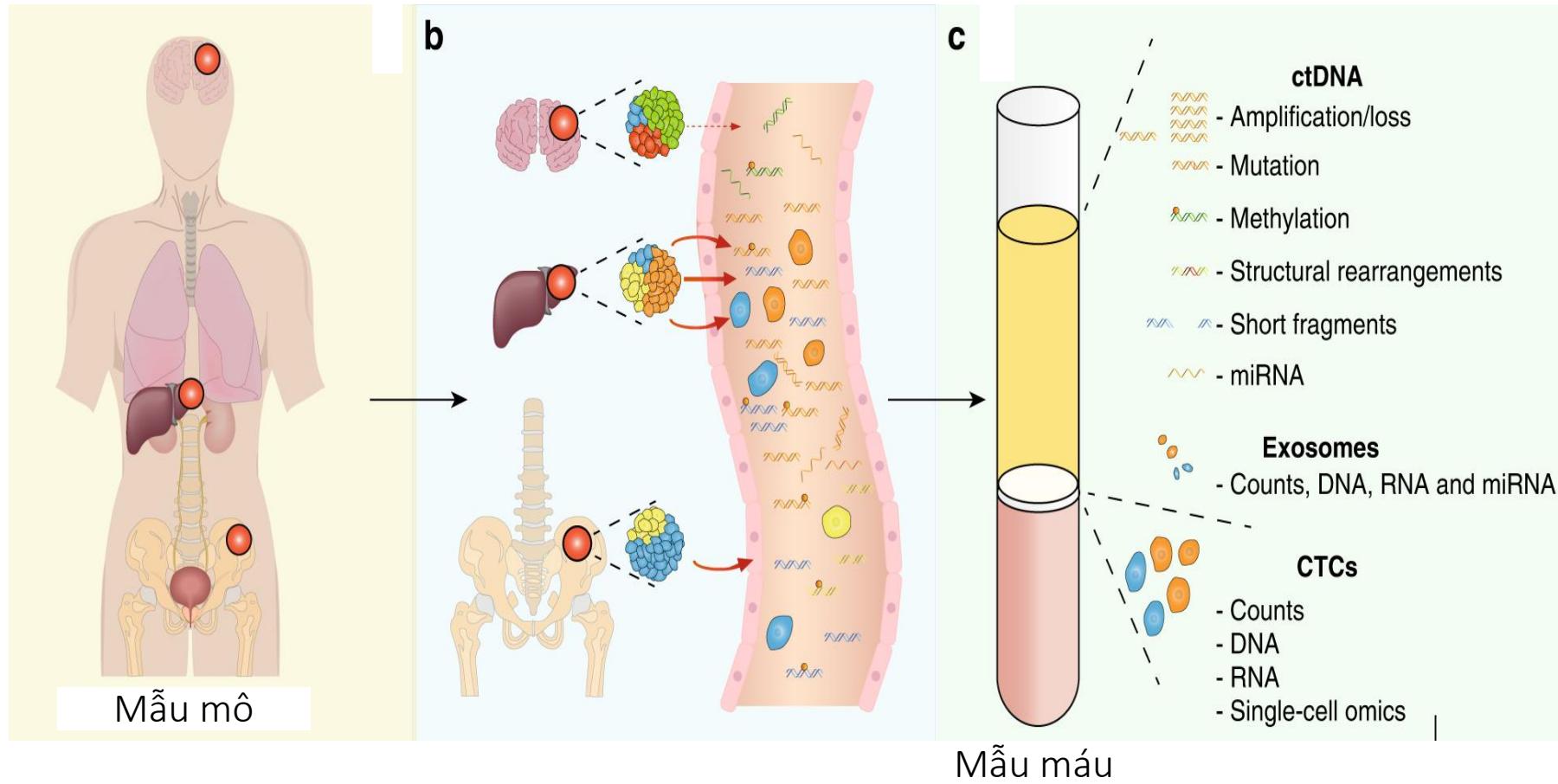


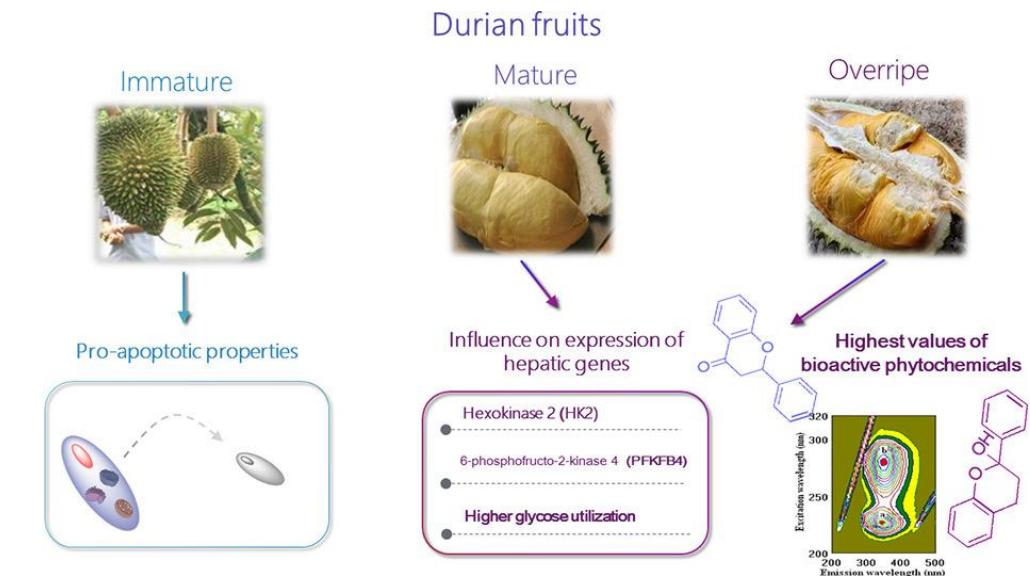
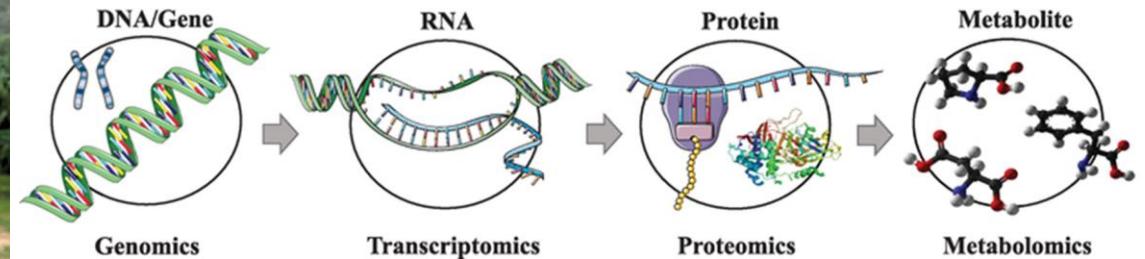
Fig. Kaplan-Meier estimates of survival in patients with a detectable mGSTP1 at baseline. (A) OS according to mGSTP1 detection after two cycles of docetaxel.

Mẫu cho XÉT NGHIỆM gen: mẫu mô và máu



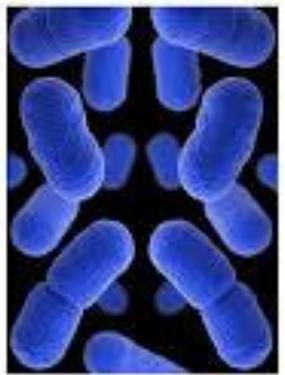
10.1038/s41416-022-01881-9

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

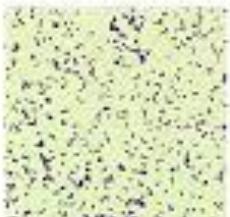


Listeria monocytogenes

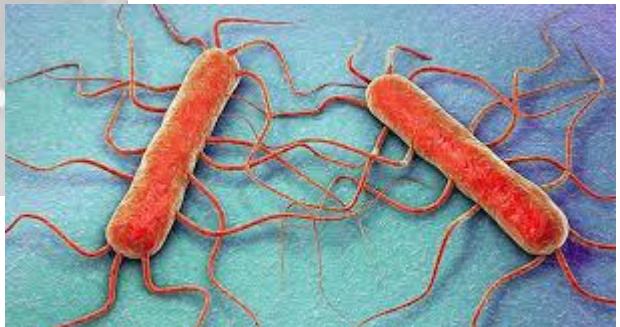
microbeonline



Beta-hemolytic colonies



Gram positive coccobacilli



Listeria monocytogenes EGD-e, complete genome

GenBank: AL591824.1

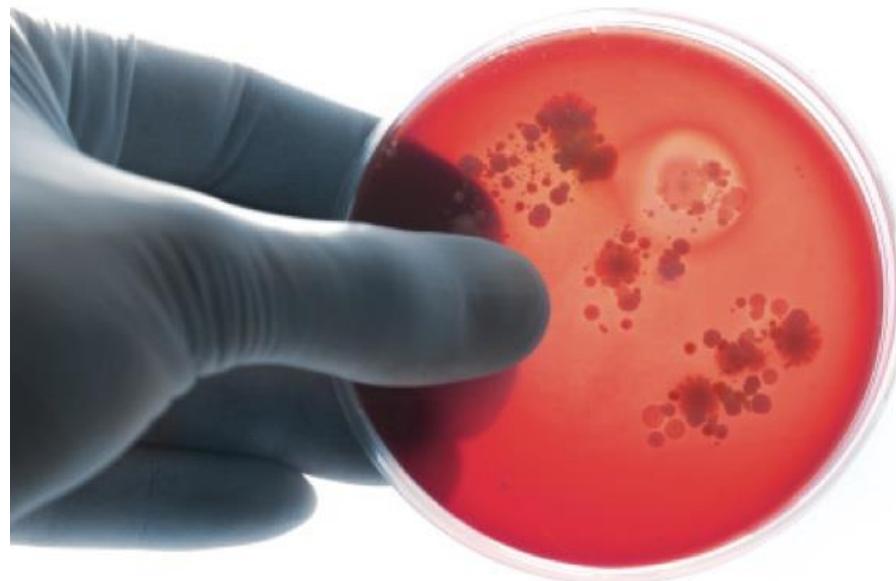
[GenBank](#) [Graphics](#)

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

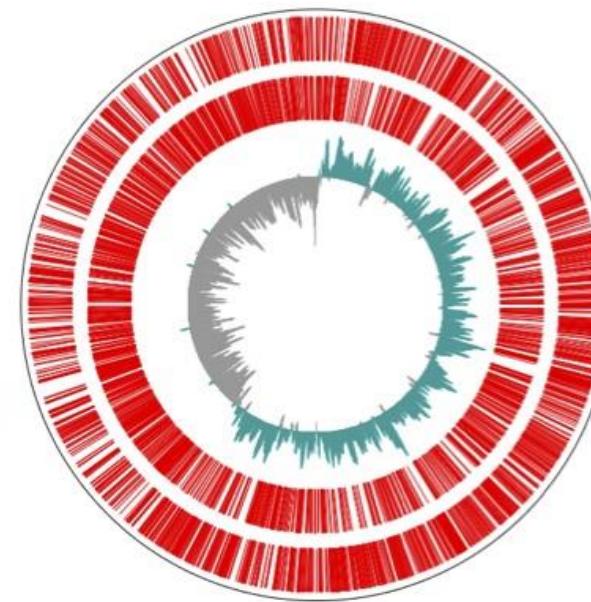
```
TAATTAATCTAACAAATTCTGTTACAGATTCTTACACACAAGTTACACACAAGTTAACGGCTGTGGA  
CAACCGTTTTCACATCTGGACAGTTTGTTGGATAGATTGGTAAGTCCTGCTATCAGAGTGAATTCT  
GATATTATAATTCCCTGCGAATAGAAAATAGCTGGGAAAACTAAAGTTATCCACAATACATTTTACT  
TTGTGGATAATTTTAACAGTGTGGATAACCTTATCCATAGCTTTCTATCTGTGGATAACTTTAT  
AGCATCCATTACATTACATAAAAAGGGGGGTACTAGTGCAATCAATTGAAGACATCTGGCAGGAAACA  
CTGCAAATTGTTAAAAAAATATGAGTAAACCTAGTTACGATACATGGATGAAATCAACAAACCGCTCATT  
CACTTGAAGGTAACACGTTTATTATTCAGCGCCAAATAATTGTTGCGATTGGTAGAGAAAAGCTA  
CACTCAATTATCGCTAACATTGCAAGAAAATACTGGTCGTTATTGATGTCGCTTATTGATGGC  
GAGCAGGAAGAAAACTTGAATACACTGTGATTAAACCAATCCAGCATTAGATGAAGATGGCATTGAAA  
TTGGAAAACATATGCTTATCCACGTTATGTTTGATACTTGTCAATTGGTCAGGGAACAGATTGC  
CCACGCAGCATCACTGCACTAGCCGAAGCACCAGCGAAAGCATAATCCACTCTCATTATGGAGGA  
GTTGGCCTCGGTTAAACACATTAAATGCAAGCAGTTGCCACTATGTTCAACAAACATAAAAGATAATGCGA  
AAGTAATGTACCTTCCAGCGAAAATTCCAATGAGTTATTAGCTTATTCGTGATAATAAAACCGA  
AGAATTGCAACAAATATCGGAATGTTGATGCTTACTTATTGATGATATTCAATTAGCCGGTAAA  
GAAGGAACACAAGAGGAATTTCACATTTAACACACTTTATGATGAACAAAAGCAAATTATT  
CCAGTGACCGGACCAACAAAAGAAATTCCACACTGGAAAGATCGACTGAGATCCGCTTGAATGGGCTT  
AATTACTGATATTAGCCACCAAGACTTAGAAACCCGGATGCCATTACGTAAAAAAGCAAAGCAGAC  
GGATTAGATATTCCAATGAAGTTATGCTTATATCGCAAACCAAATTGATTCGAATATTGCGAGCTAG  
AAGGGCTCTATCGAGTAGTTGCTTATTCTCCCTGTTAATAAGATATAACAGCTGGCTTGAGC  
AGAAGCACTAAAGATATTATCCCCTTTCAAATCACAAGTTACTATTAGTGGTATTCAAGAAGCA  
GTCGGTGAATATTCCACGTTGTTAGAAGATTAAAGCAAAAAACGGACGAAAGTATAGCATTCC  
CGCGCCAAATCGCCATGTATCTCAAGAGAGCTTACAGATGCCCTATTACCAAAATCGGTGATGAATT  
TGGTGGTCGAGATCATACAACAGTTATTGATGACATGAAAAAAATATCGCAACTACTAAAAACCGATCAA  
GTGTTGAAAATGACCTGCCGAAATTGAAAAAAATTAAAGAAAAGCACAAAAATATGTTAATAGACCT  
GTGTACAATGTGGATAACTGAAACATACTTACCCACAAGTTACACATGTGGAAAACTTATGCAGCAT  
GGCTTGTAAACCTACTTATCCACAAATCCACAGCGCTATTACTATTACTACGATTTTTATTAATTAAATT
```

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

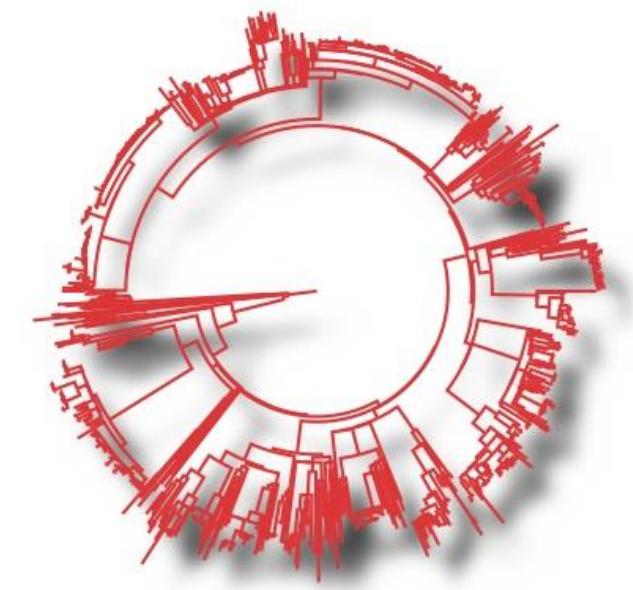
Microbial Genome Analysis: study one isolate at once



Culture microbe



Genome Sequencing



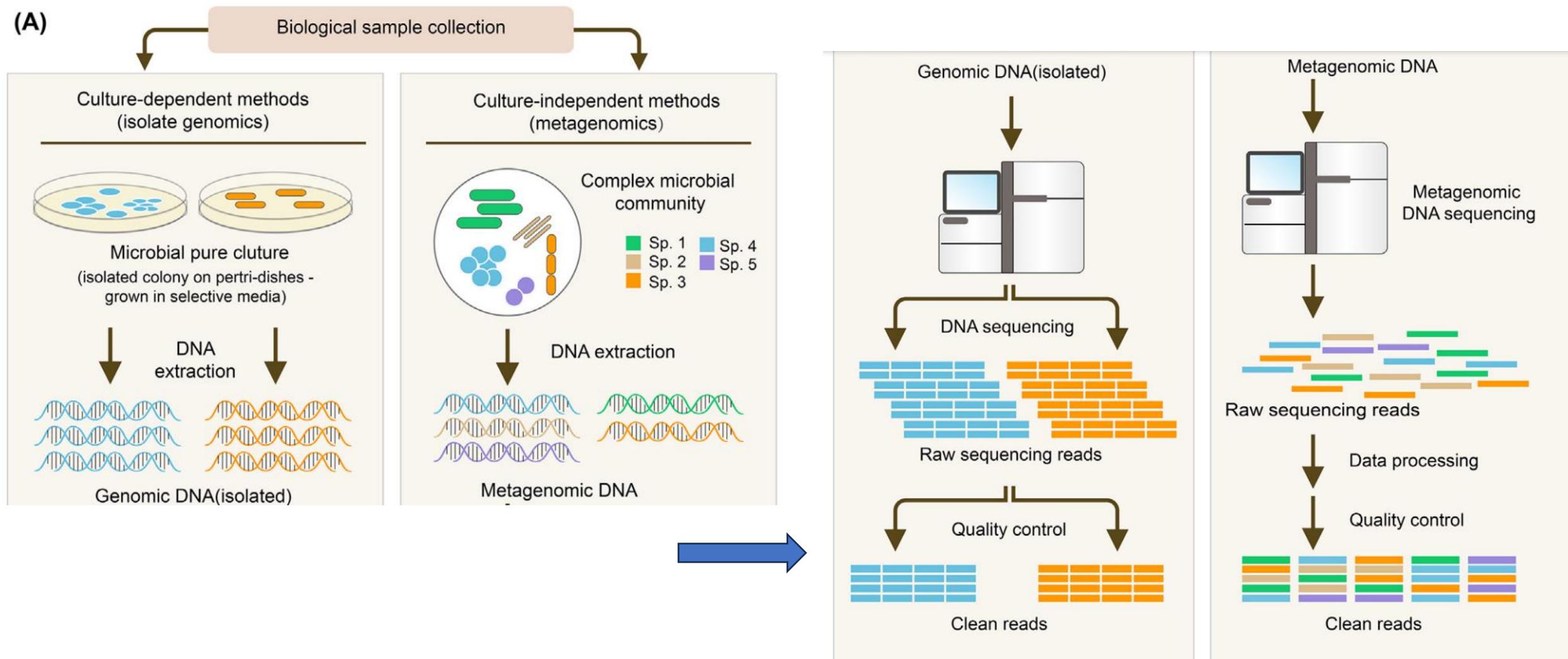
Serotyping/genotyping

Microbe = Microorganism

Prokaryotes = Bacteria & Archaea

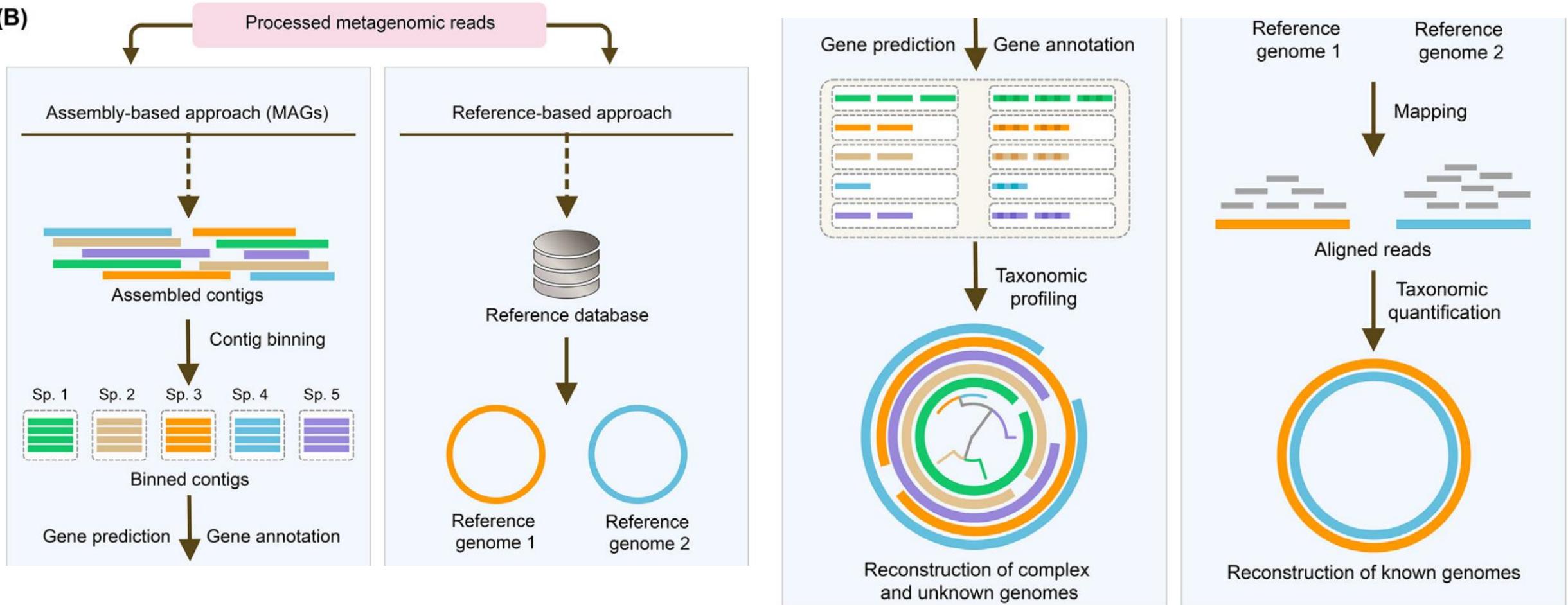
Microbial Genome vs Metagenome Analysis

(A)

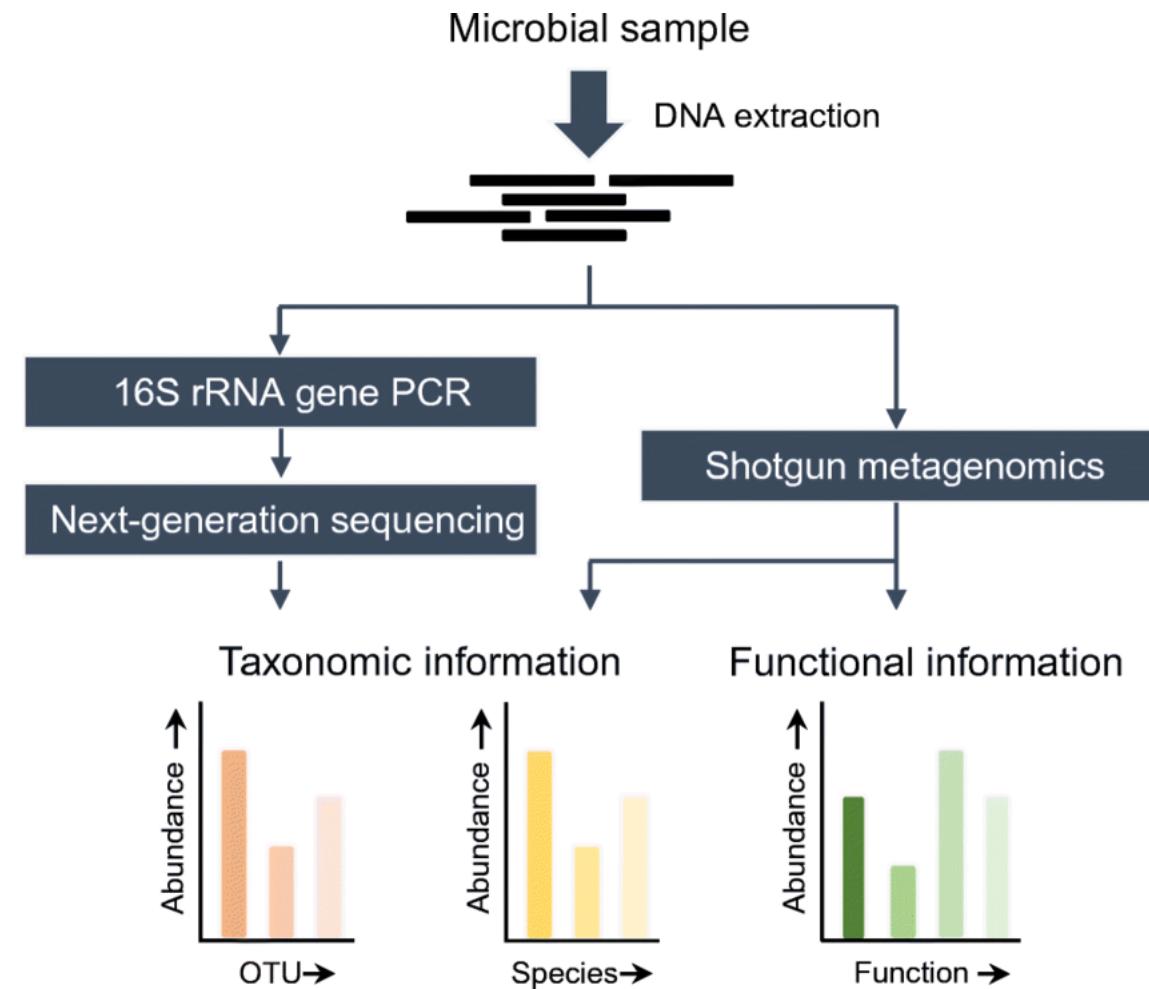
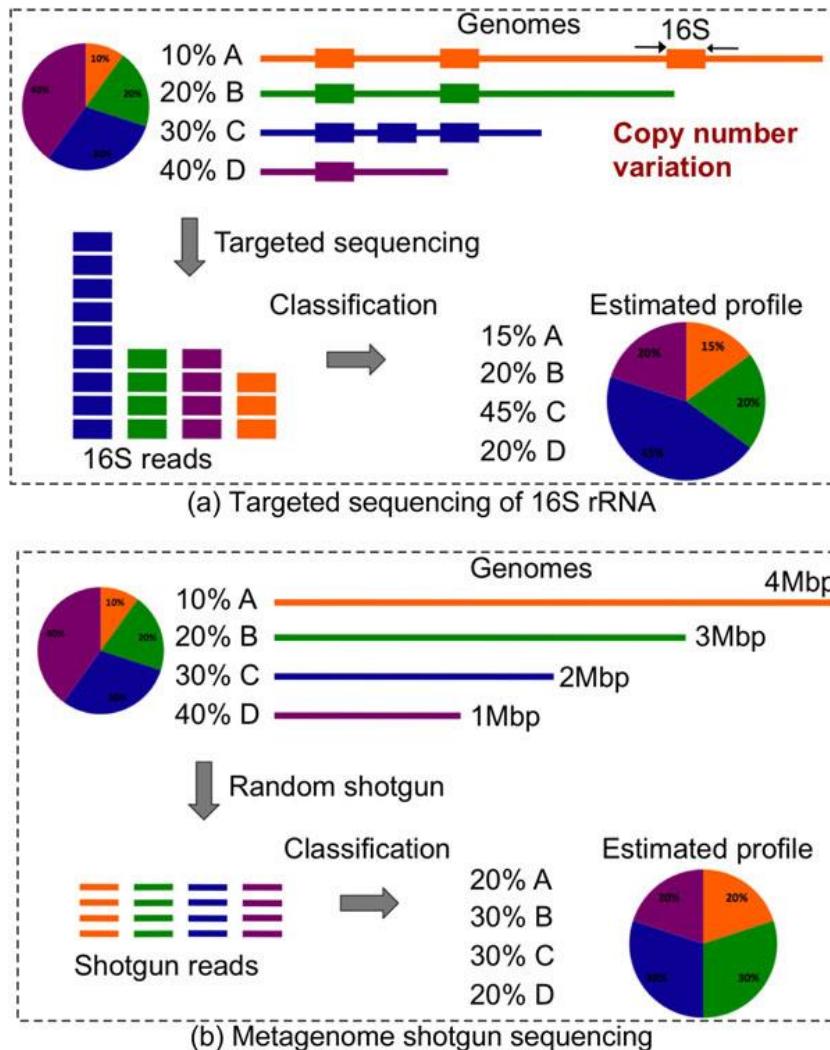


Shotgun Metagenome Analysis

(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

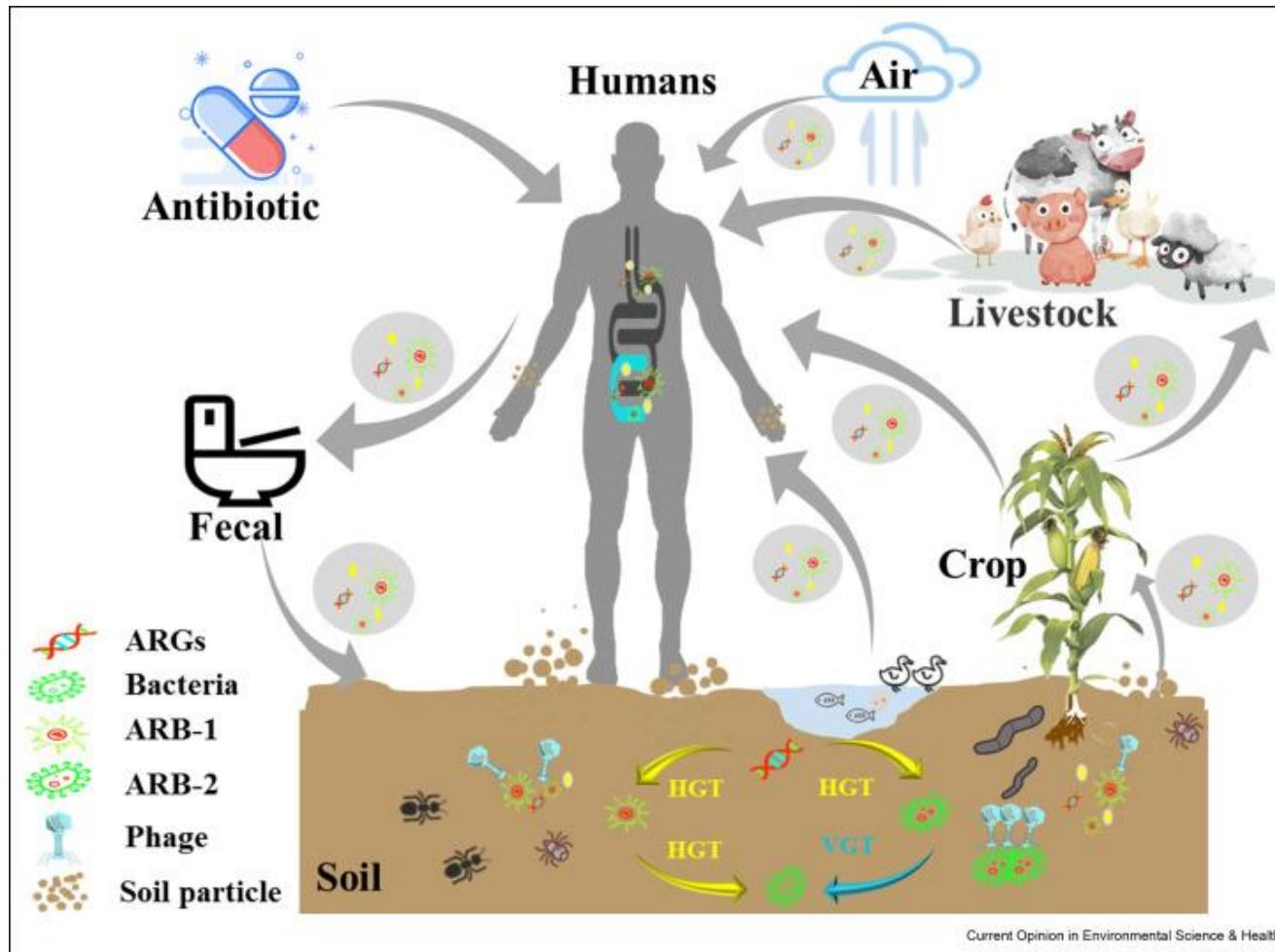
Name	Associated Condition	Result
<i>Neisseria gonorrhoeae</i>	Gonorrhea, urethritis, pelvic inflammatory disease, gonococcemia, gonococcal ophthalmia neonatorum	Not Detected
<i>Chlamydia trachomatis</i>	Chlamydia, cervicitis, urethritis, pelvic inflammatory disease	Not Detected
<i>Mycoplasma genitalium</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

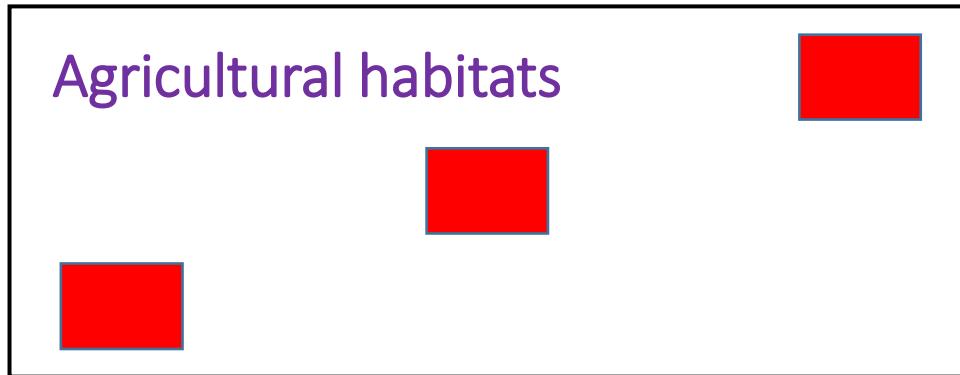
Note: Human papillomavirus (HPV) 16, 18, 31, 33, 35, 39, 45, 51, 56, 58, 59, and 68 are considered high-risk or probable high-risk due to their association with cervical cancer. HPV 6, 11, 42, 43, and 44 are considered low-risk for cervical cancer but may cause genital warts. Other HPV genotypes found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

Transmission of antibiotic resistance genes in the soil ecosystem



EXPERIMENTAL SETUP

Agricultural habitats



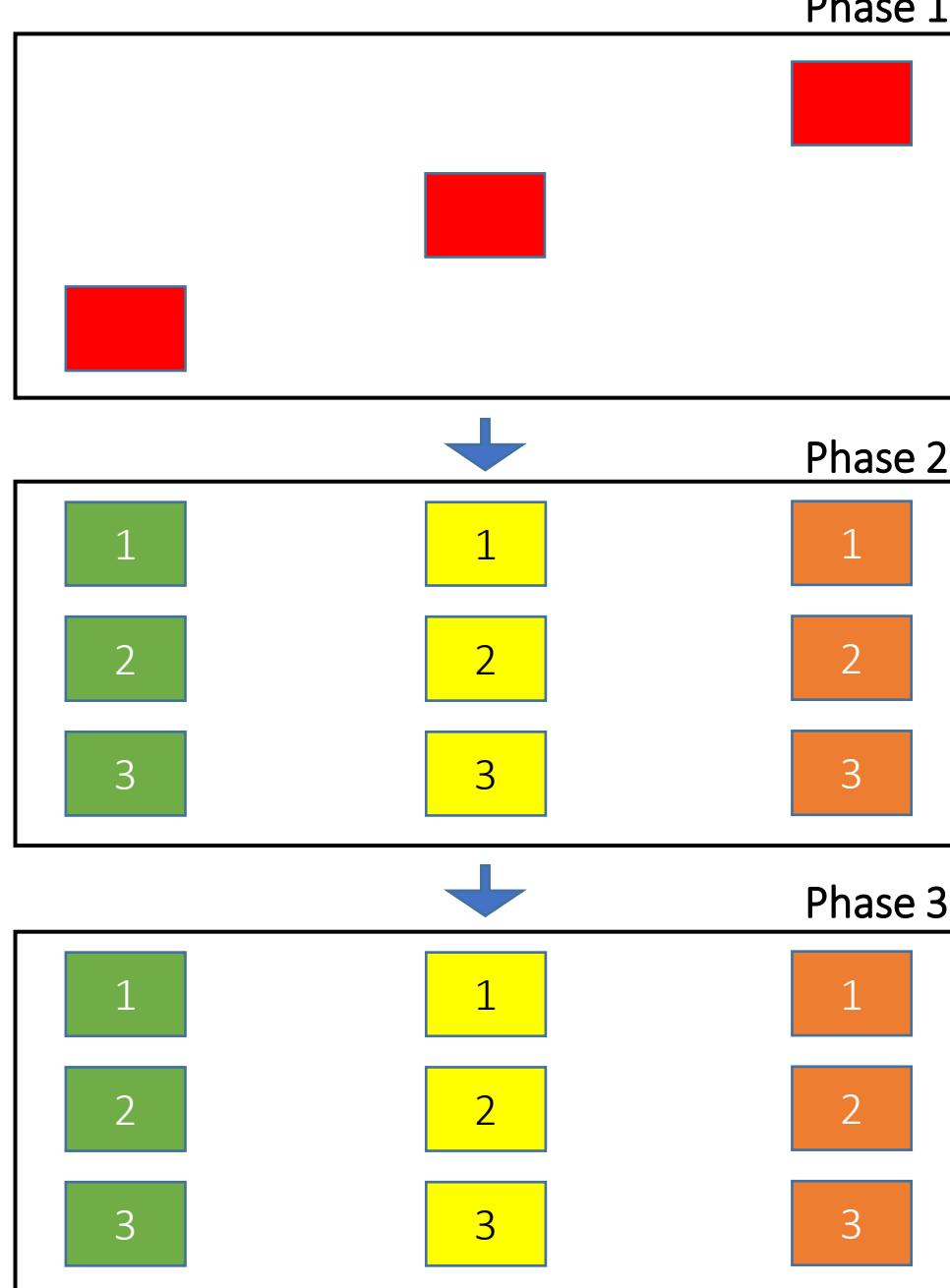
Vegetation: *Brassica integrifolia*

Phase 1: Non-fertilization and non-pesticide



Shotgun Metagenomics

EXPERIMENTAL SETUP



Phase 1: Non-fertilization and Non-pesticide

Phase 2: Apply Fertilization + Non-pesticide

Phase 3: Apply Fertilization + Pesticide

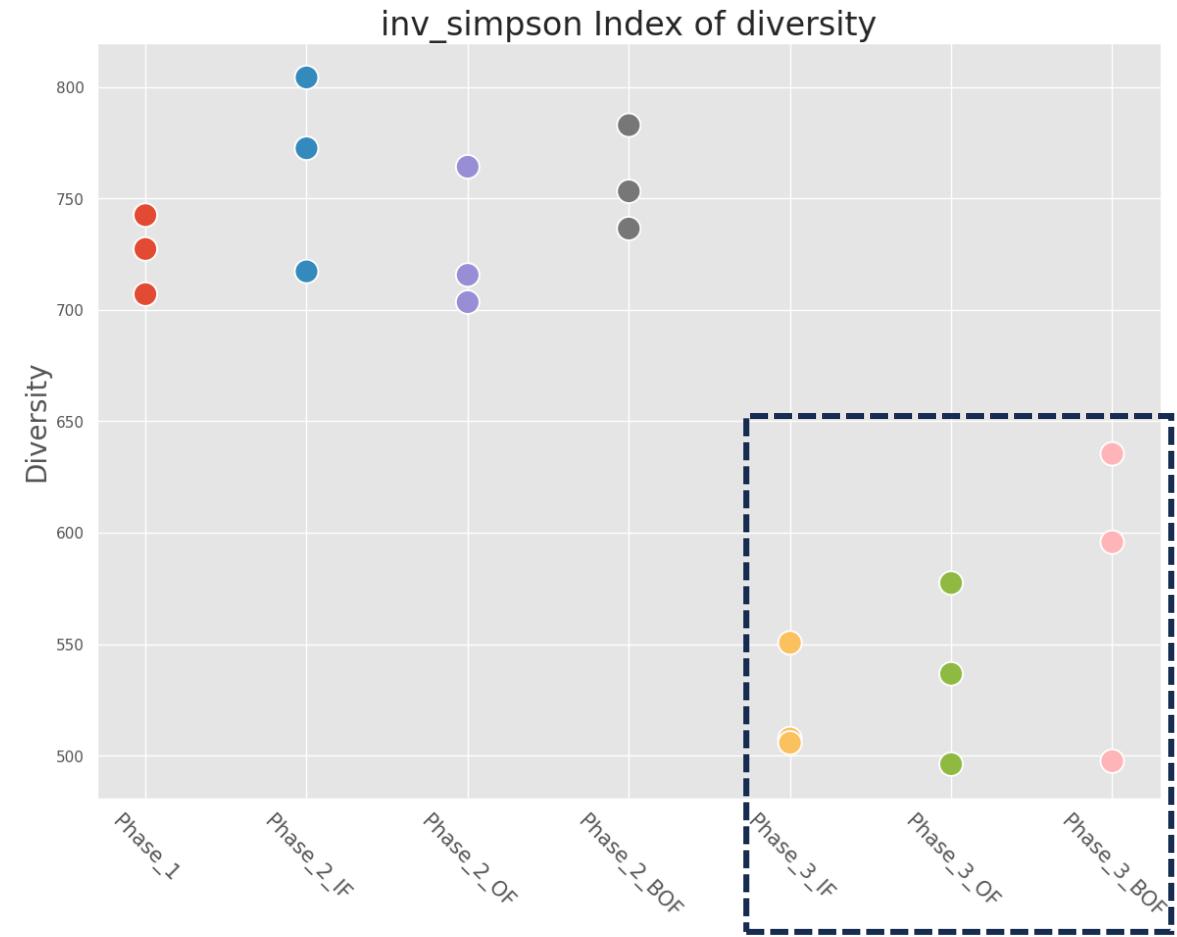
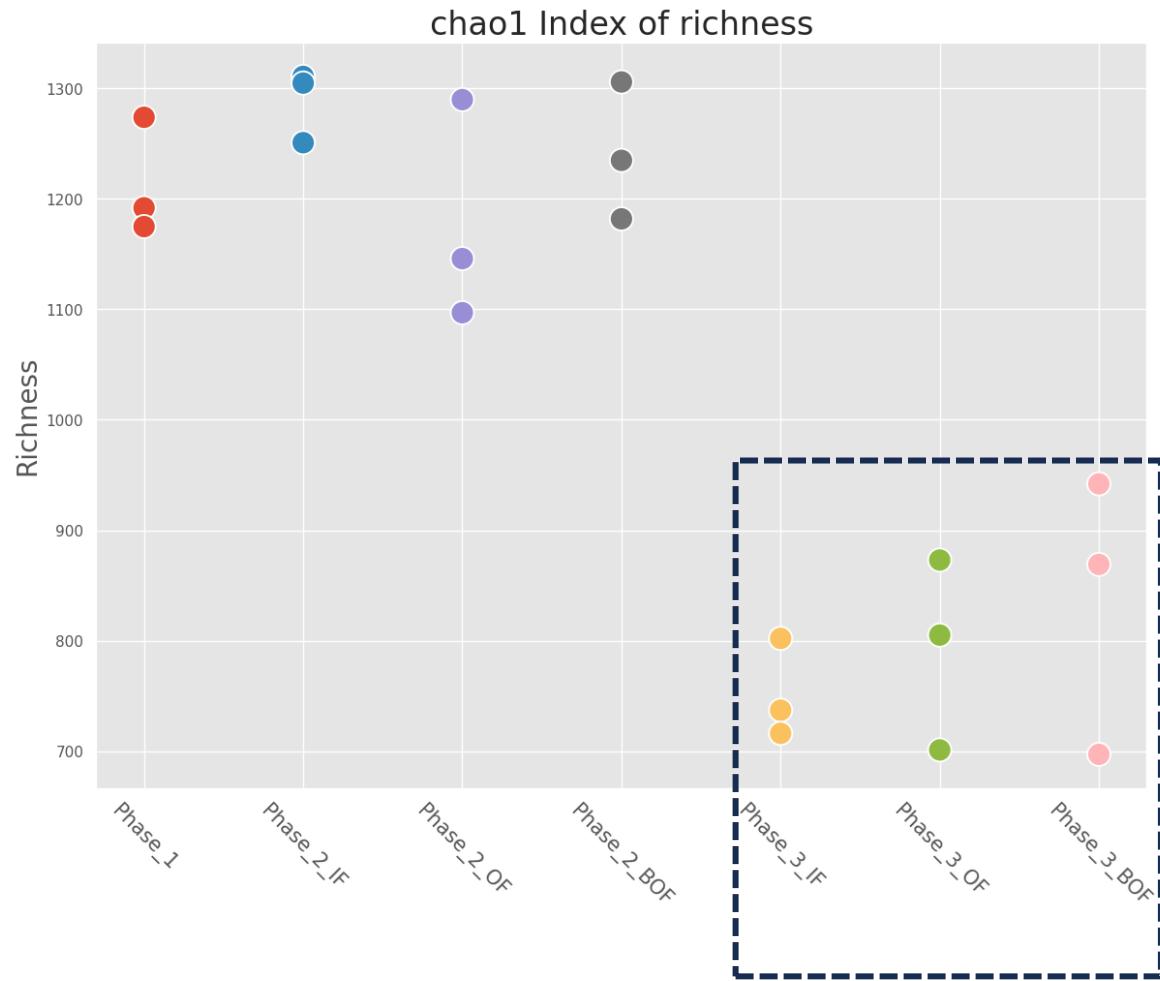
IF: No organic fertilizer, apply 100% inorganic fertilizer

BOF: Apply bio-organic fertilizer + 100% inorganic fertilizer

OF: Apply organic fertilizer + 100% inorganic fertilizer

21 Shotgun Metagenomics

Application of fertilizers and pesticides significantly impacts ARGs



Nội dung môn học “Phân tích dữ liệu tin y sinh”

- Processing medical and genomic raw data
- Explore data
- Feature selection
- Dimension reduction
- Clustering data
- Building machine learning model

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