



# Giới thiệu môn học Phân tích dữ liệu chuyên biệt Phân tích dữ liệu Tin Y Sinh

Review at [https://www.youtube.com/watch?v=Qac1Bd8V8\\_I](https://www.youtube.com/watch?v=Qac1Bd8V8_I)

Ngày 05 tháng 11 năm 2025

TS. Lưu Phúc Lợi

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

Zalo: 0901802182

# 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



### **Head of Scientific Research Office**

**Current Affiliation** Institute for Applied Research in Health Sciences and Aging (ARiHA) - Thong Nhat Hospital  
01 Ly Thuong Kiet street, Tan Son Nhat ward, HCM city, Vietnam

---

**Webpage** <https://scholar.google.com.au/citations?user=KPIqpJsAAAAJ&hl=en>

---

**ORCID** 0000-0001-8045-718X

---

**Address** 592/17 Lac Long Quan, Hoa Binh ward, HCM city, Vietnam  
**Home** 5 Rose st, Sefton, Sydney 2162 Australia

---

**Telephone** +84 (0)901802182      **E-mail** [luu.p.loi@googlemail.com](mailto:luu.p.loi@googlemail.com)  
[loilp@bvtn.org.vn](mailto:loilp@bvtn.org.vn)

# 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

$\beta$ -Lactam antibiotics inhibit the growth of *Streptococcus pneumoniae* by inactivation of cell-wall-synthesizing enzymes, the penicillin-binding proteins (PBPs). Resistance to  $\beta$ -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

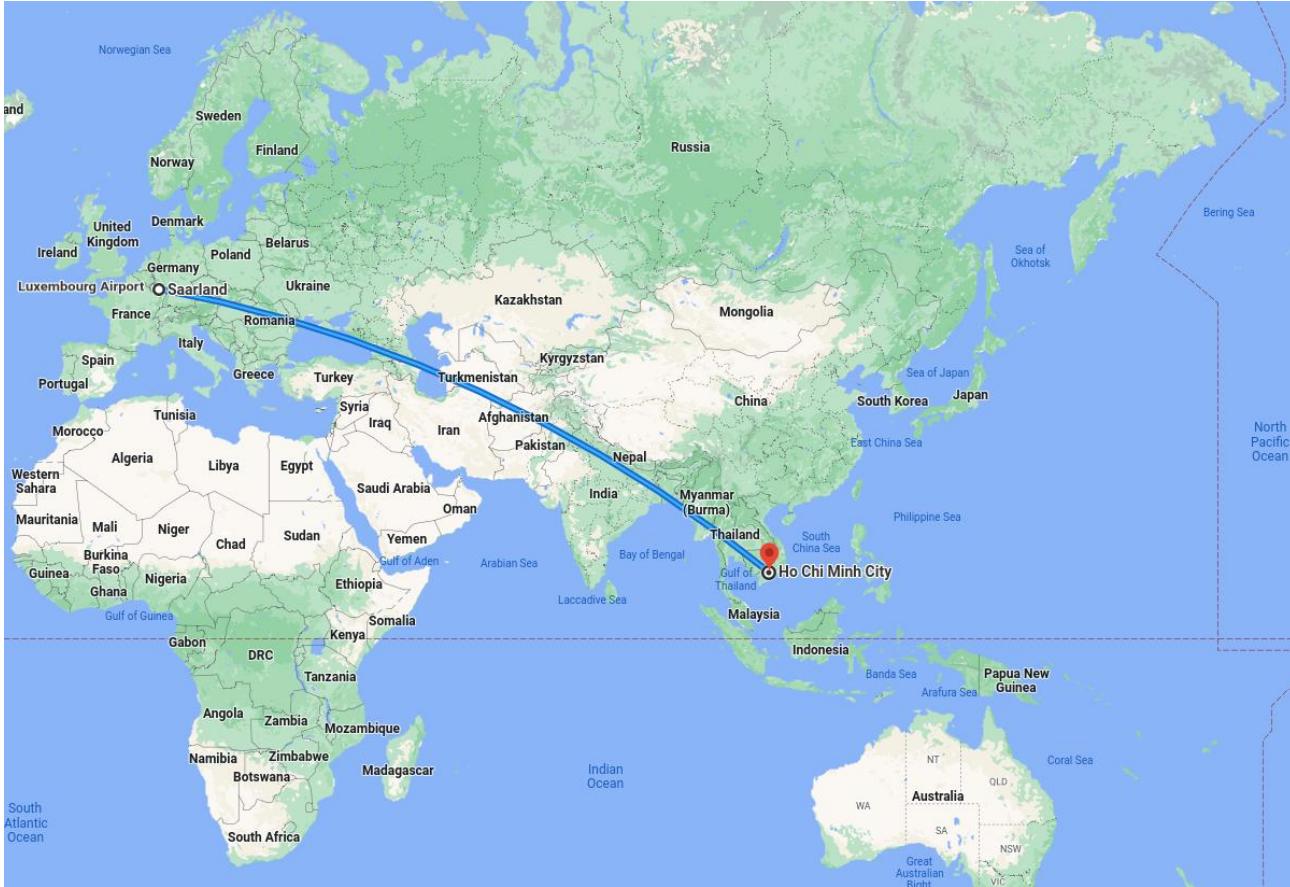
[FASTA](#) [Graphics](#) [PopSet](#)

Go to:

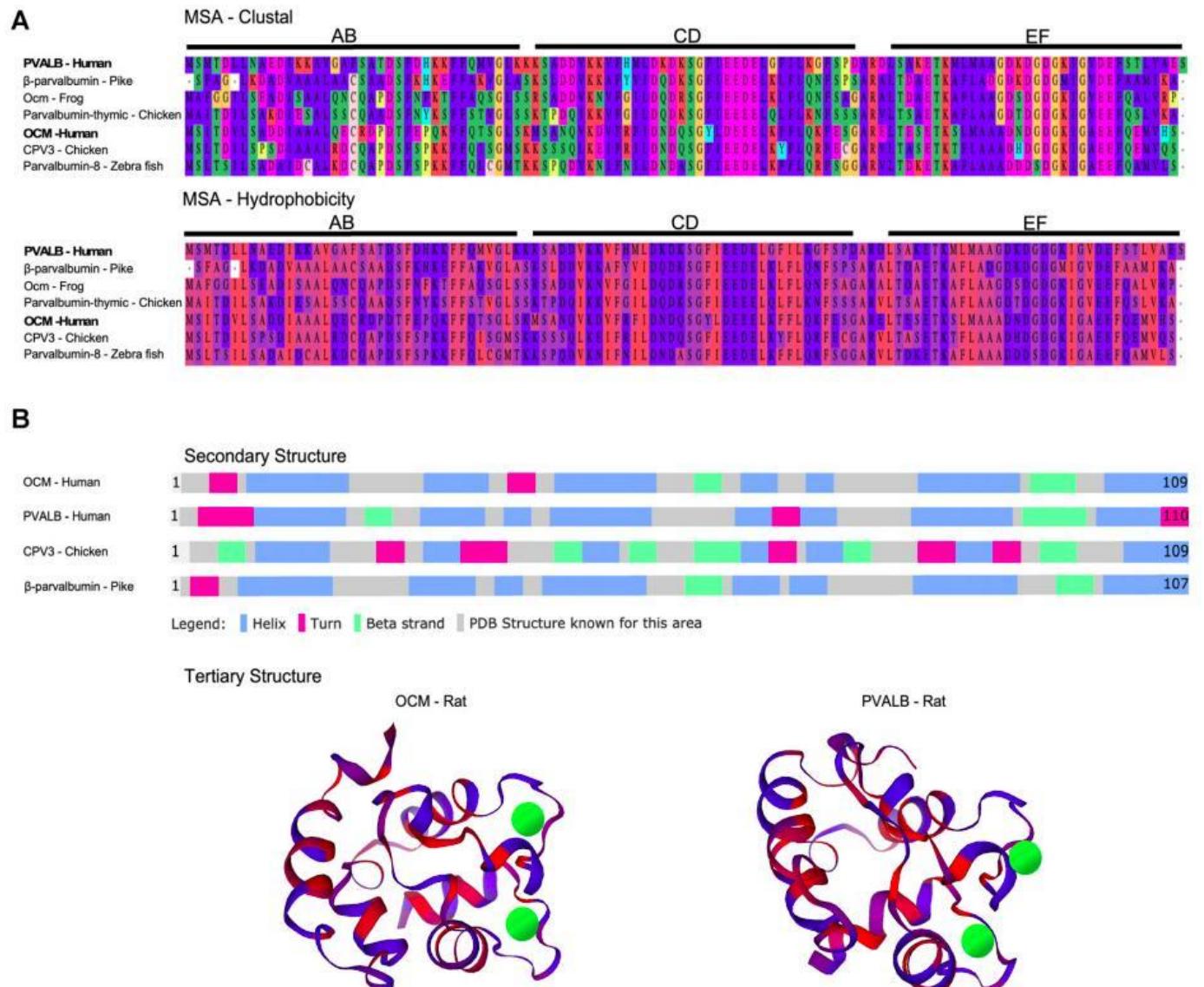
LOCUS AY690659 94 bp RNA linear VRL 22-AUG-2004  
DEFINITION Hepatitis C virus isolate MB14 5' UTR.  
ACCESSION AY690659  
VERSION AY690659.1  
KEYWORDS .  
SOURCE Hepacivirus C  
ORGANISM [Hepacivirus C](#)  
Viruses; Riboviria; Orthornavirae; Kitrinoviricota; Flasuviricetes; Amarillovirales; Flaviviridae; 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 Luu<sup>1</sup>, Hans R. Schöler<sup>2,3</sup> and Marcos J. Araúzo-Bravo<sup>1,4</sup>

Author Affiliations

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

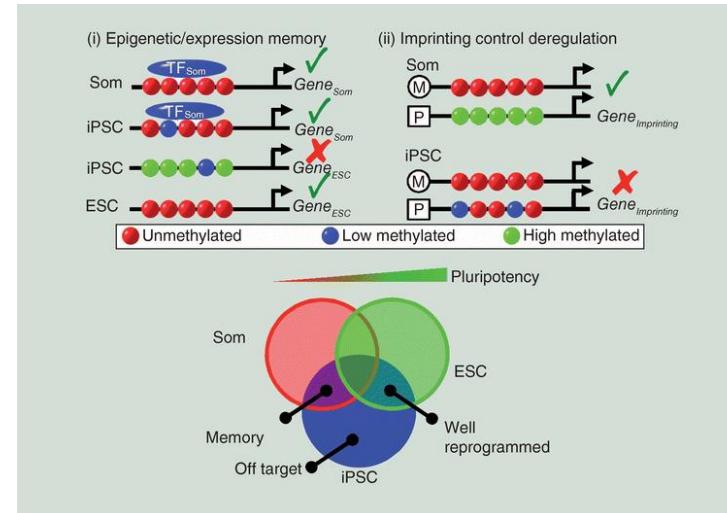
<sup>2</sup>Department of Cell and Developmental Biology, Max Planck Institute for Molecular Biomedicine, 48149 Münster, Germany;

<sup>3</sup>University 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<sup>†</sup>, Daniela Gerovska<sup>†</sup>, 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 ▾

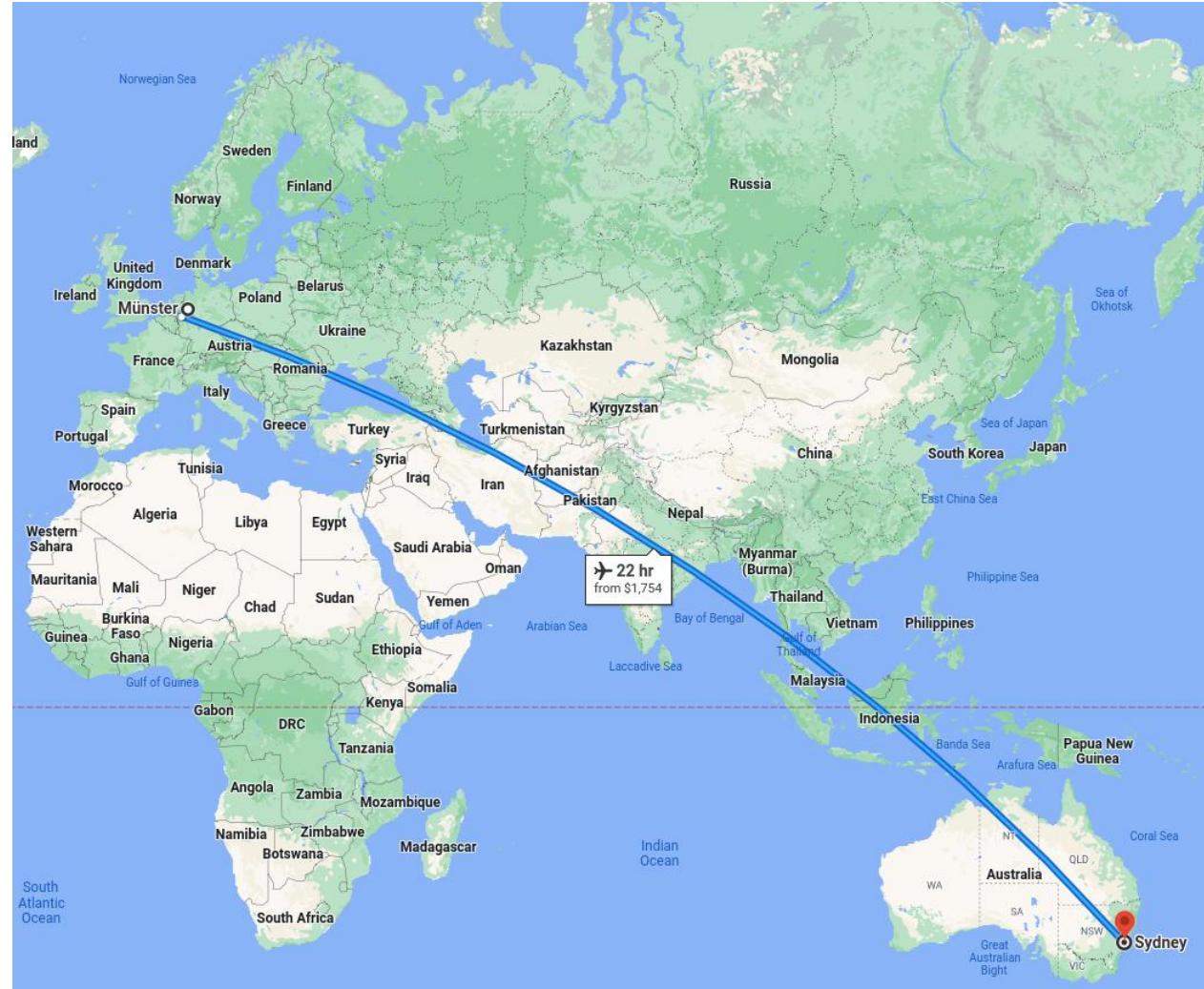
PDF Split View Cite Permissions Share ▾

## 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)



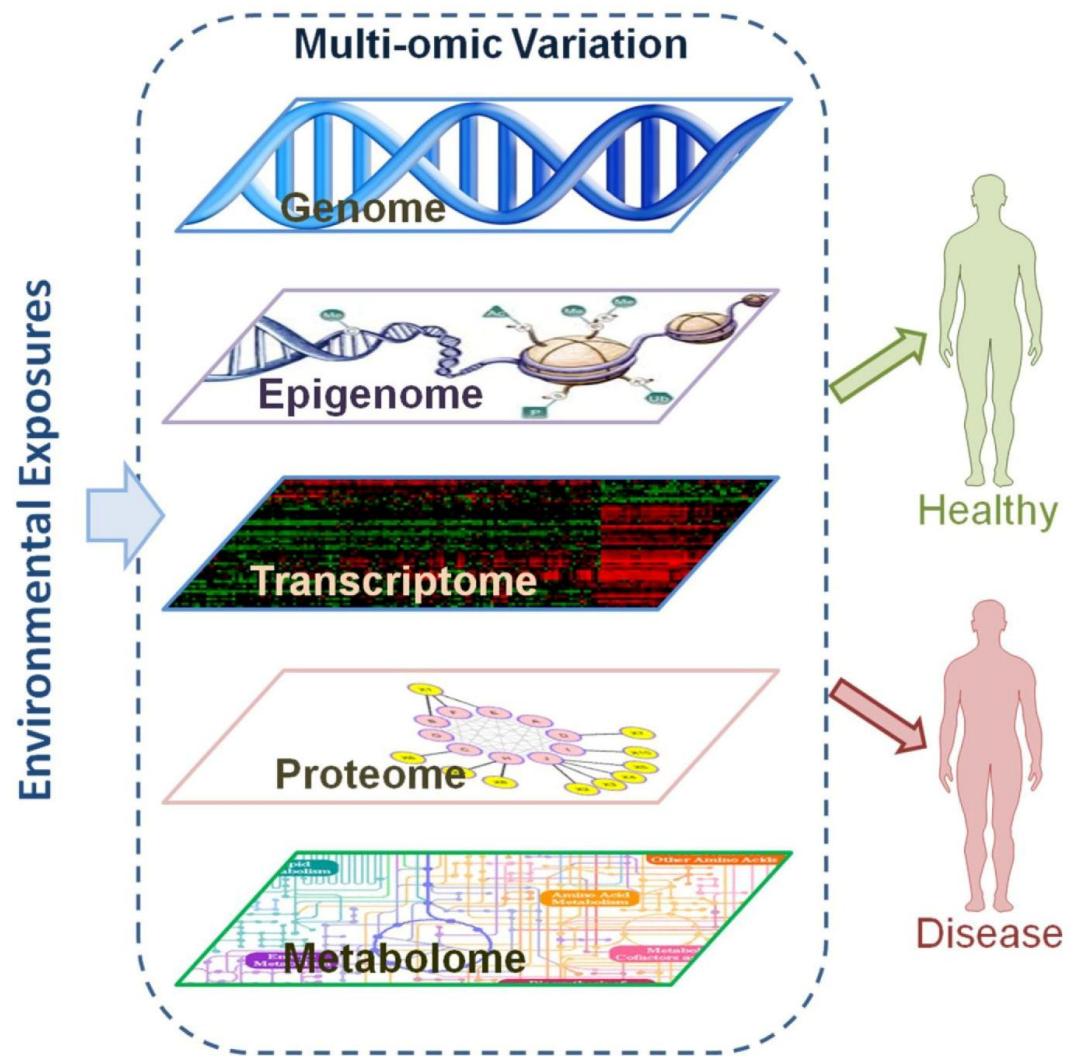
# Seven years of research at Garvan Institute of Medical Research in Australia (2014-2022)

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

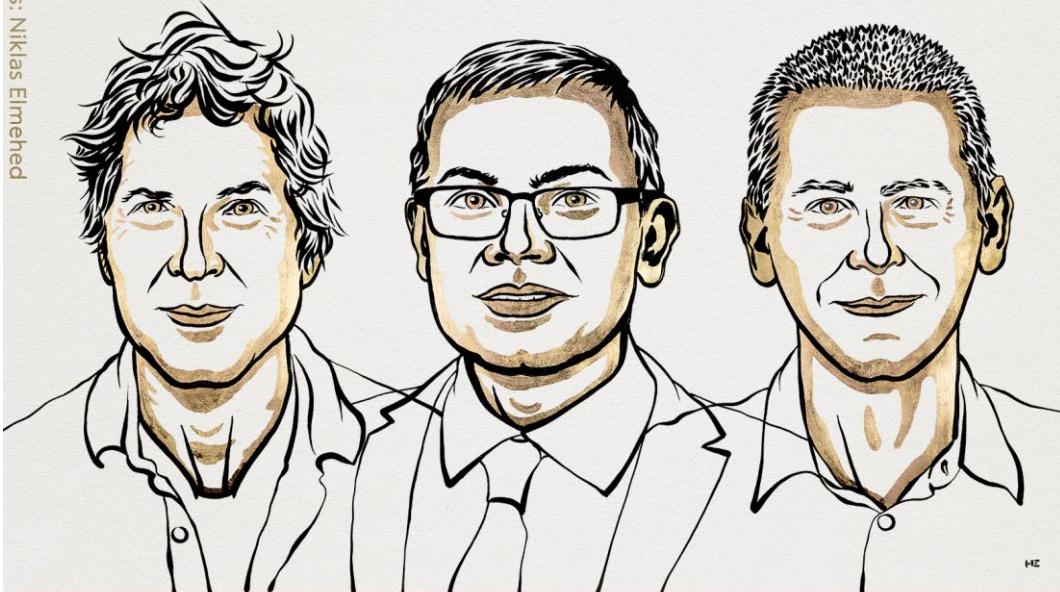


# More than one year at Zymo Research (Sep 2022-Nov 2023)

- Microbiomics
- Epigenomics
- Transcriptomics
- Proteomics
- Metabolomics
- 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**

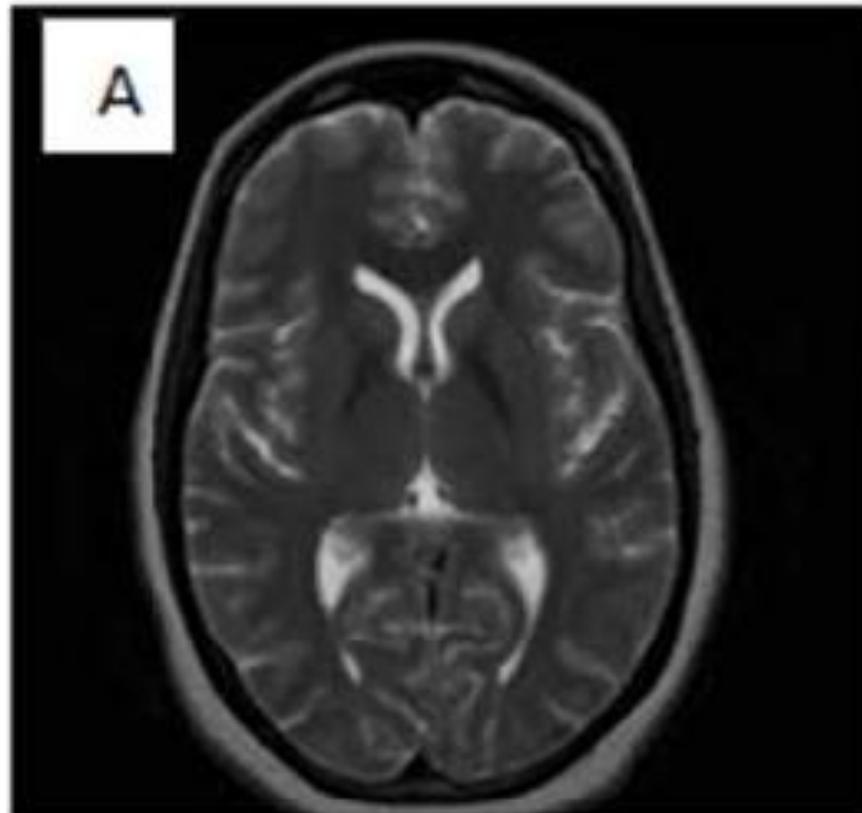
THE ROYAL SWEDISH ACADEMY OF SCIENCES

# Giới thiệu về dữ liệu y sinh

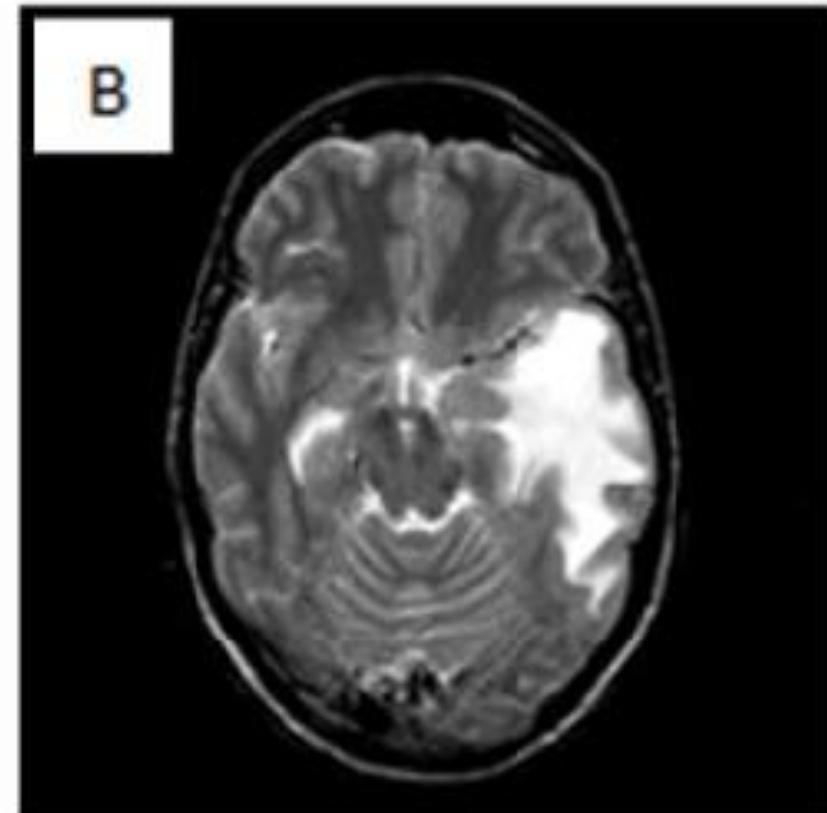
# Medical Data

- Text (Patient Medical Record or Health Record)
- Image (Hematoxylin Eosin (HE), Immunohistochemistry (IHC) CT nd MRI)
- Video (Medical Ultrasonography and Endoscopy)
- Signal (Electrocardiogram (ECG) and Electroencephalogram (EEG))

# MRI



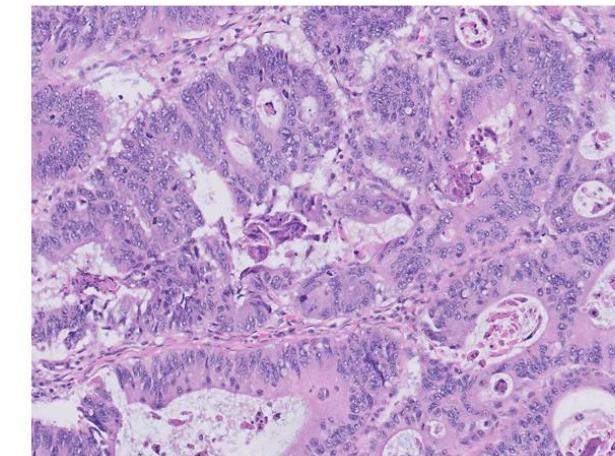
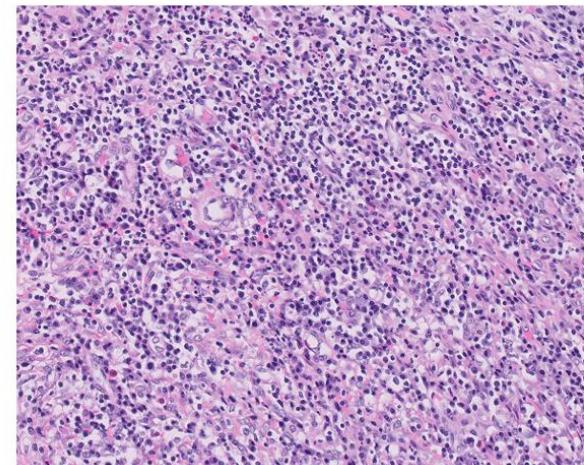
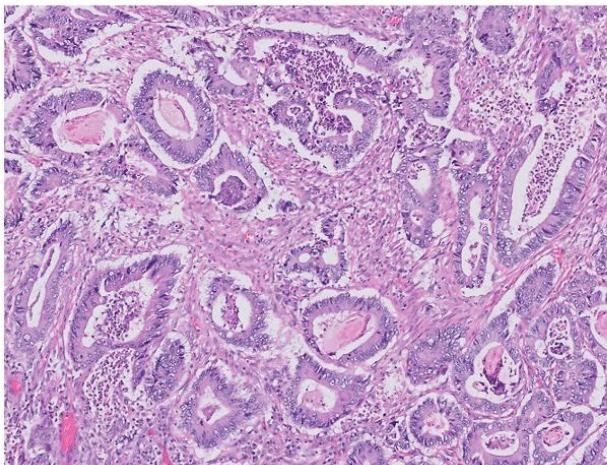
(A)-Normal brain



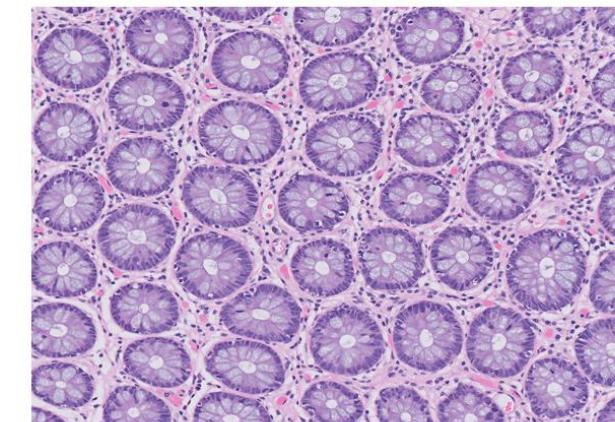
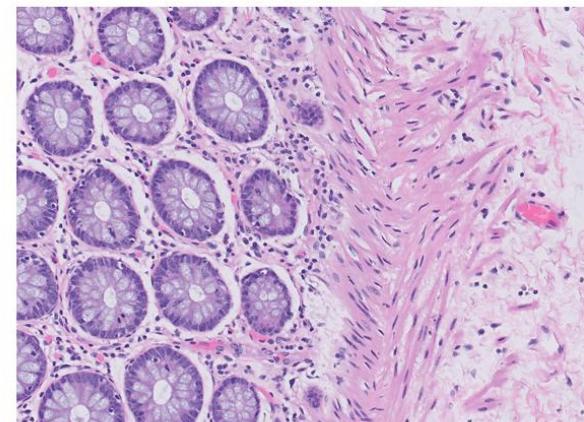
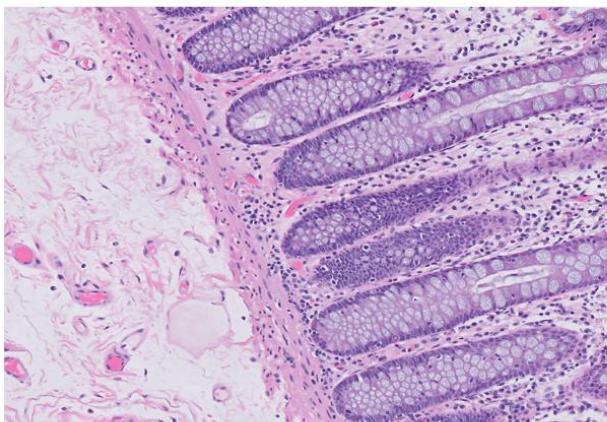
(B)-Tumor Brain

# Hematoxylin Eosin

**Normal  
(10X)**



**Tumor  
(10X)**

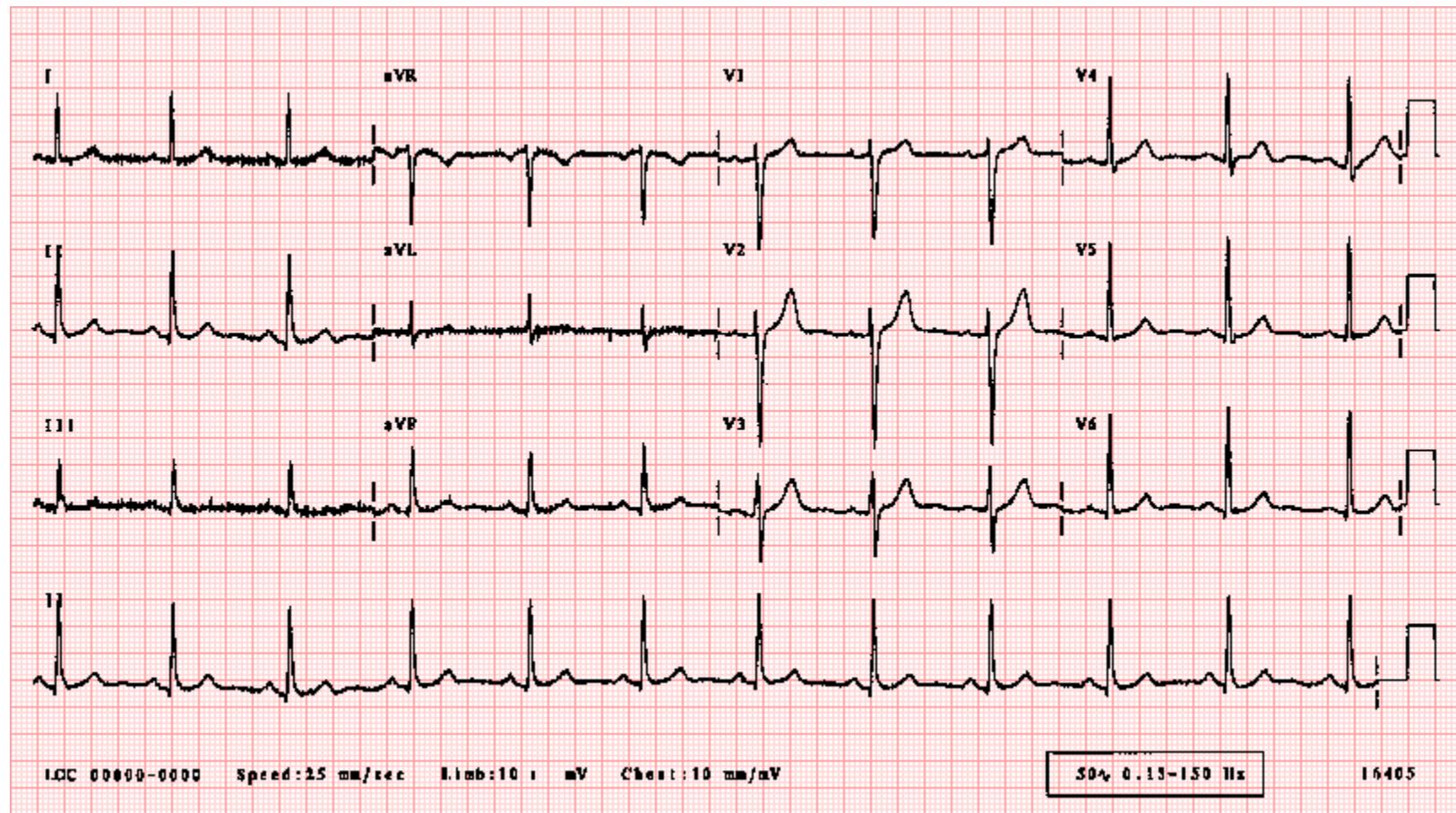


# Medical Ultrasonography



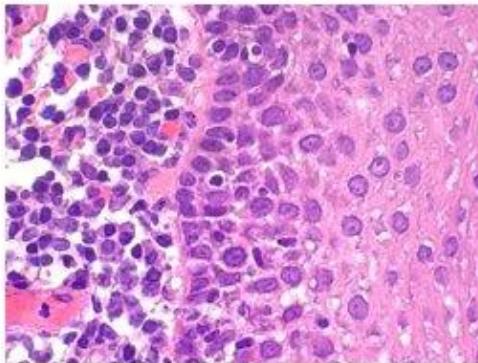
[https://commons.wikimedia.org/wiki/File:Ultrasound\\_Scan\\_ND\\_0105143736\\_1439320.gif](https://commons.wikimedia.org/wiki/File:Ultrasound_Scan_ND_0105143736_1439320.gif)

# ECG

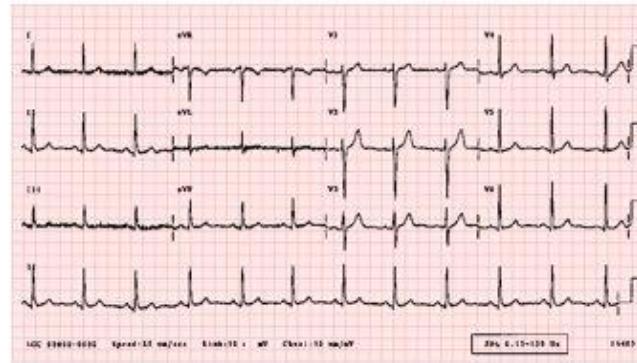


# Dữ liệu ảnh giải phẫu bệnh và chẩn đoán hình ảnh

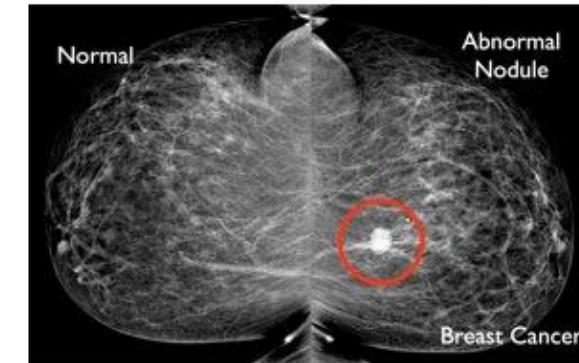
1. HE



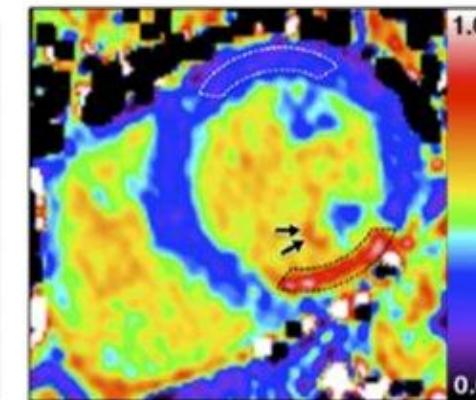
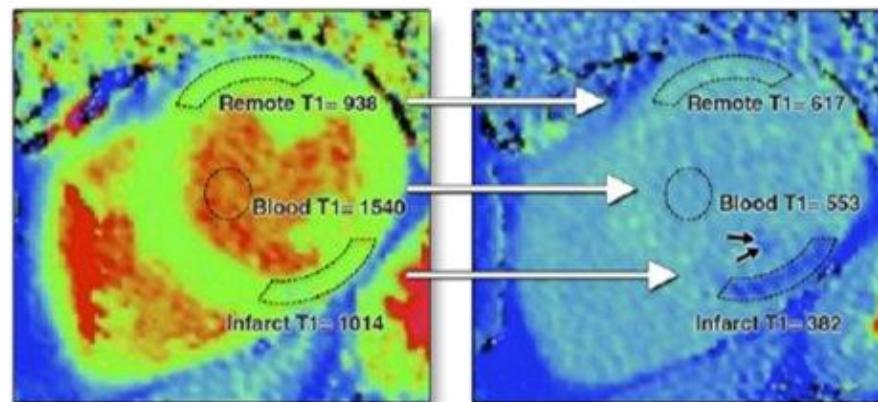
2. ECG



3. Mammography



4. Heart MRI



# Dữ liệu hệ gen



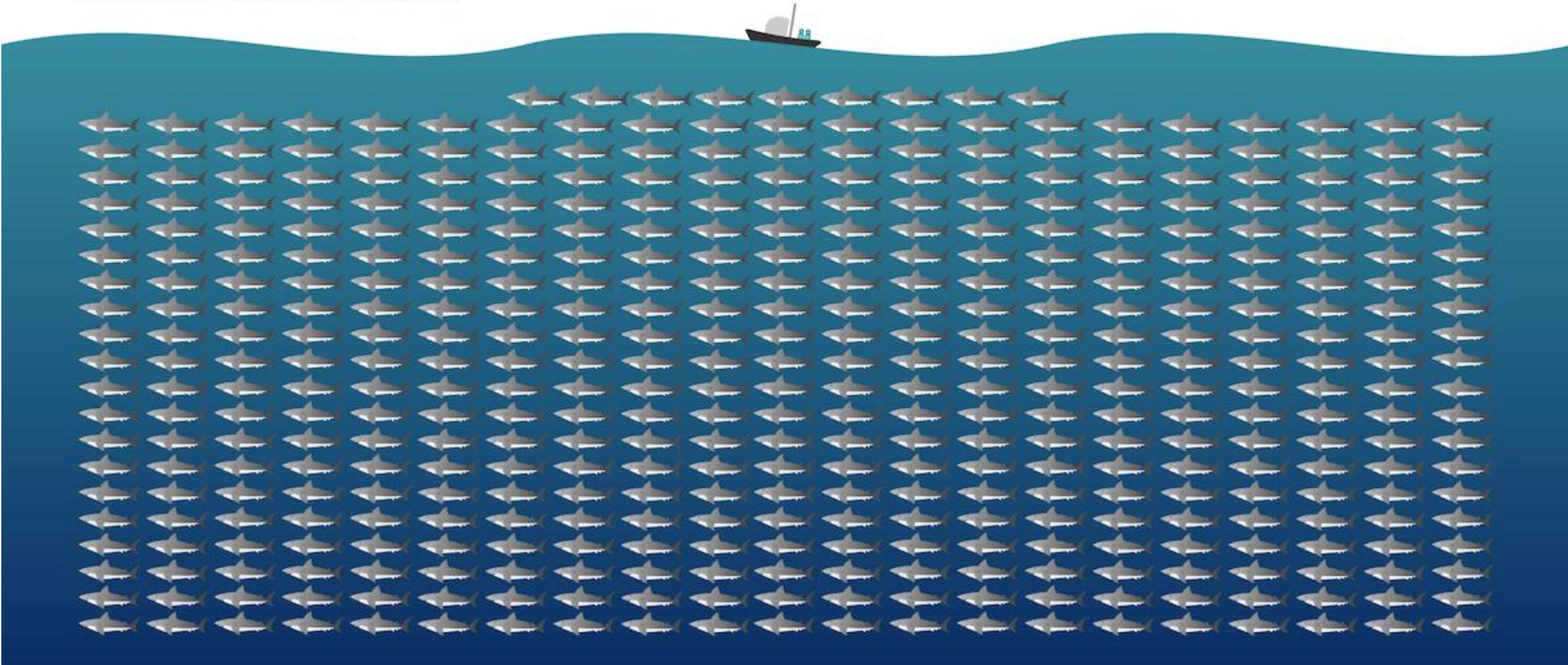
48541 agccctcaa agaaatgttc tcagcaggca tggagccag gacttgtcc cttgggttag  
48601 agagccgggt tgaaggtgac tgaagtaaa tggagacata gaggcggggg ggtgttgag  
48661 ttctggagg tgggggtgtt gggacactgc tigtactga gatgcacccc tgccagttct  
48721 gcttaagat ttgaggccgg gggcaggggg gcggaggtaa gtcattttac ttgttaagtaa  
48781 tttaaacct ttaatatta aagcaaacgt ggatgtttaa tgaatgaaat tcattctgga  
48841 atgaaaattt cacgtgtatg taaaaataa cacggggctt cagagaggac tttctggctg  
48901 gcagcagact ccagattccc agggccctg caccctctc tgcccacagg gcacctaatt  
48961 ggagaaggta tgggaggaga gccaggccgg agtcagagca cactggtagc tccacatttg  
49021 cagcgtcccc tgcctctc ctgaggctt gcaacgtca atatgtcaaa caaactcccc  
49081 ctgtcccggt ccagttctg aggacaagag ccaccaccc tagcaataa agaccaggca  
49141 accctttgac tcatctttgtt ggtctctgg aatcagagg tagccacatc gctggaggg  
49201 ggatgtaaaggc actcggtgtaa aaggtaaaaa ggaatcgagg gacaggatgt tggggacatc  
49261 acatggaaaa tggacagaaaa gggggcaca gccgatgtt ggggggggg cggcagttcc  
49321 tacatccccctt ggcctgaagc acgctccagg gcagaaggaa aaacactgtc tttgggtcc  
49381 aagagacctg agttcaattt ctggctccac cactgaccac ctgttaacc ttgaaactgt  
49441 gctgctgaa cctcagggtt cccttctaaa aatagaggaa aaaaggatgtc atttccctt  
49501 gcccctgtga gaacgaaatg gtcaagcac caaggagct cagcaaaagggt cggccctgcc  
49561 cccgcctggc caaaccttc ctcttcaggaa ggccacggca accgtatgtt gacagaagag  
49621 cagcacctt attaatgtc tcccgatcg tgcctttagt caagtccactt aacctctgt  
49681 ggctgttcc tcattggaa aatatggctg ccagtaaaac ctggccctgc caccctctgg  
49741 ggcacttggc aaacagacaa agatgtccaaat tgcgtcggctt gggccaggcg cagtggctca  
49801 tgcgtgttaat cccagcaatt taggaaggcc aggtggggcg atcacccttagt gtcaggatgt  
49861 tgacccaggc ctggccaaaca tggtaaaacc ttgtctctac aaaaataaca aaattatggc  
49921 ggcatgtgg cggtgtcc tcattcccgat tactcggtt gctggaggaa gagaatcgct  
50041 tgaaccggca agggggaggat tgcgtgagc caagattgtt ccaactgactt ccagctgggg  
50101 caacagagcg agactgttc taaaaaaaaaaa aaaaaaaaaaa aaacaatgtca gagctggctg  
50161 tgtaaaaaaaacttgcgttccacttgcaggccca gttgtccacca ggctgggtt caggccatag  
50161 gggggggggccagcatcg cctctcaggca gcccggggag gccccggca tccctggccc  
50221 ctgtgggtctt ggtgtgttccacttgcgttccacttgcaggccca gttgtccacca ggctgggtt caggccatag  
50281 tcaggagagg cccagggtgttgcgttccacttgcgttccacttgcaggccca gttgtccacca ggctgggtt caggccatag  
50341 tggggaaa gcaactccttgcgttccacttgcaggccca gttgtccacca ggctgggtt caggccatag  
50401 cagccagcat gggcccttccat tggggggat tgcgttccacttgcaggccca gttgtccacca ggctgggtt caggccatag  
50461 agccctgtgttgcgttccacttgcaggccca gttgtccacca ggctgggtt caggccatag  
50521 gagatggctt ctgaggacccat tggggggat tgcgttccacttgcaggccca gttgtccacca ggctgggtt caggccatag  
50581 ttaaaggcgc tgagtgttgcgttccacttgcaggccca gttgtccacca ggctgggtt caggccatag  
50641 acaggtgtga cagggtgggg gactggggaa taagatgtt aaccctggctt ggaggctgt  
50701 agcttccaca gccagcgctg gacaggggagg gtccagatgtt aaccctggctt ggaggctgt



# 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*



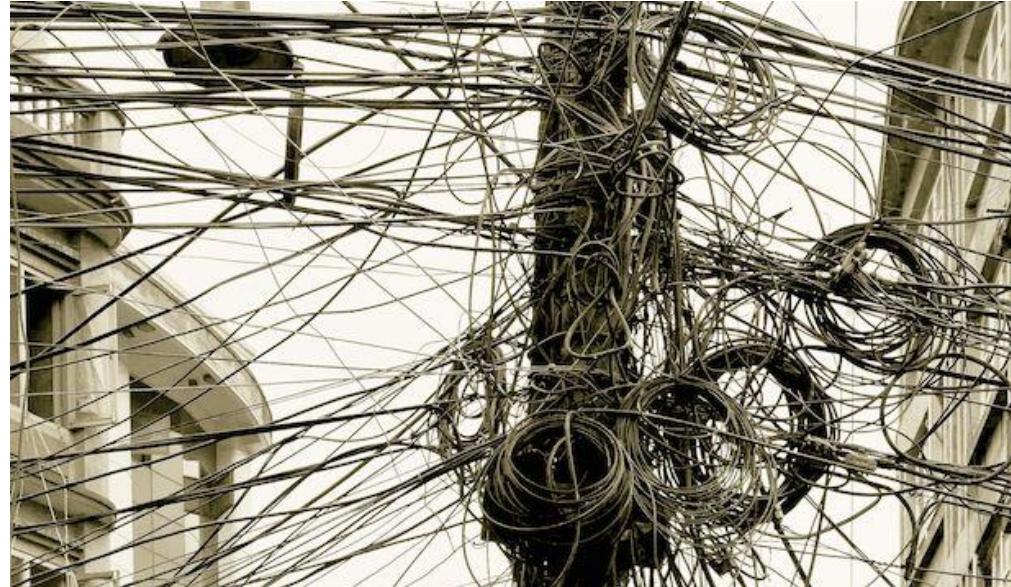
# Người ngoài nghề nghĩ ...

## 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!»

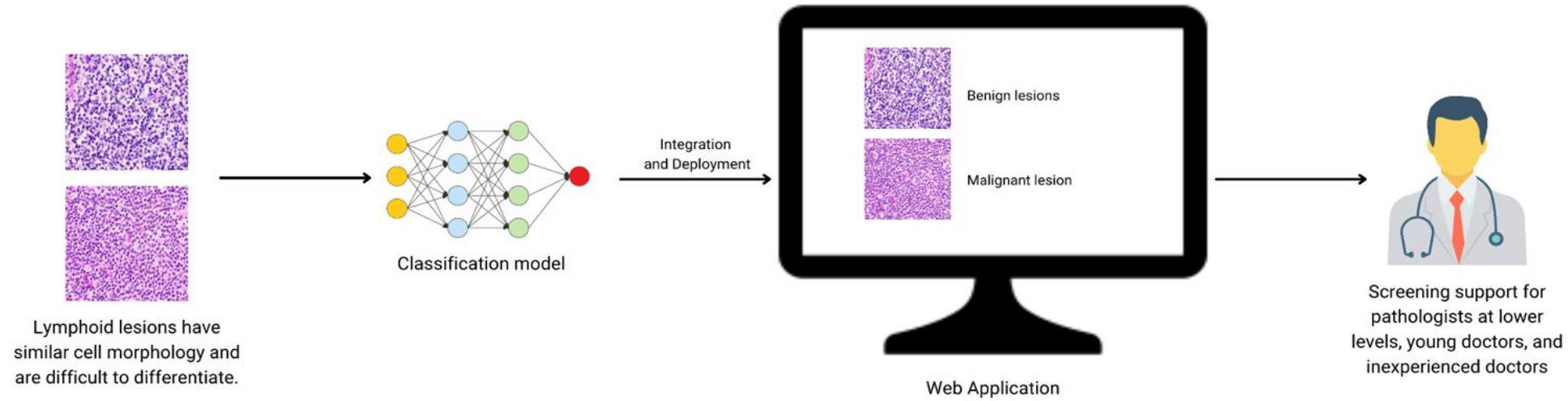
# Thật sự khi mình thật sự bắt tay vào ...



- ➔ 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

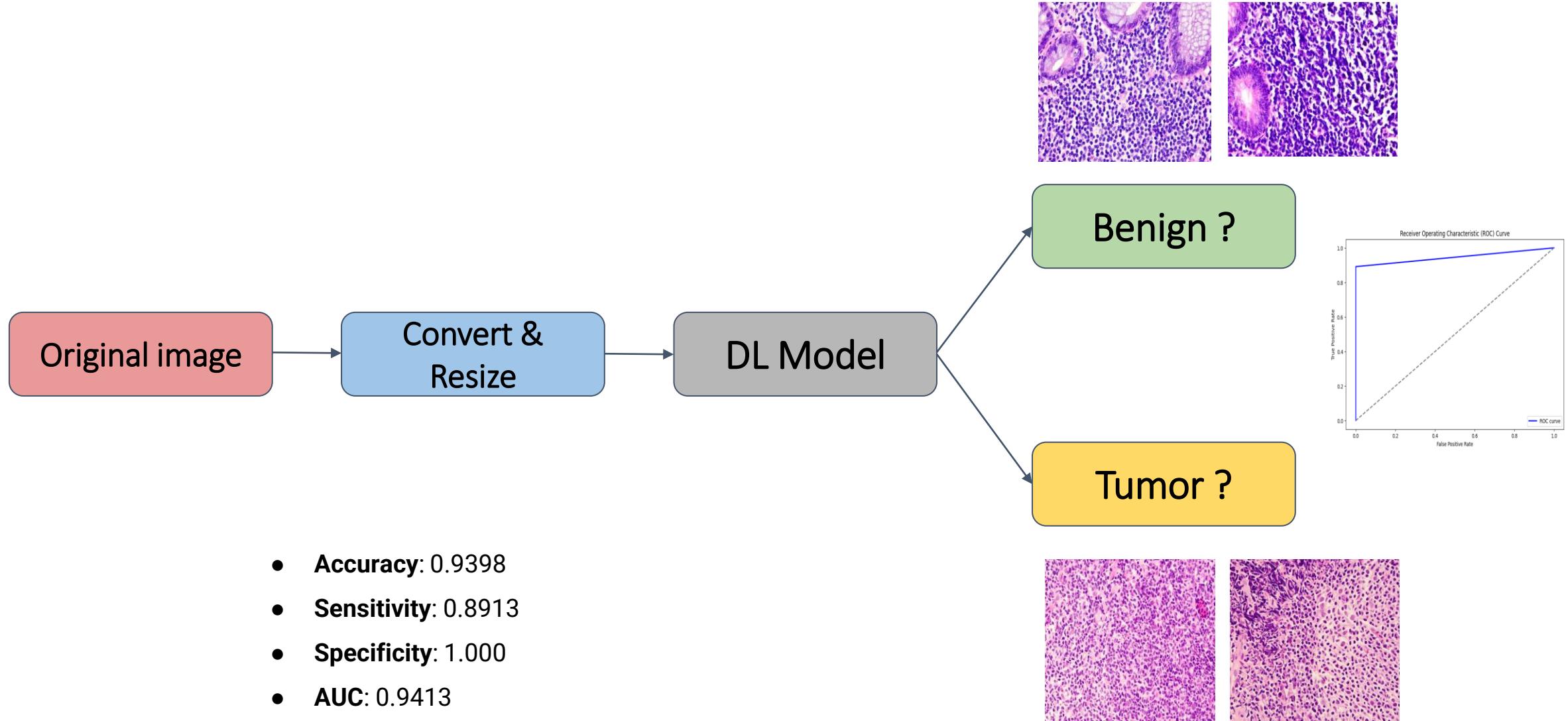
# DeepHE: Lymphoma đường tiêu hóa

1. Develop a deep learning model, capable of accurately classifying images of H&E-stained biopsy specimens of follicular lymphoma cells in the digestive tract.

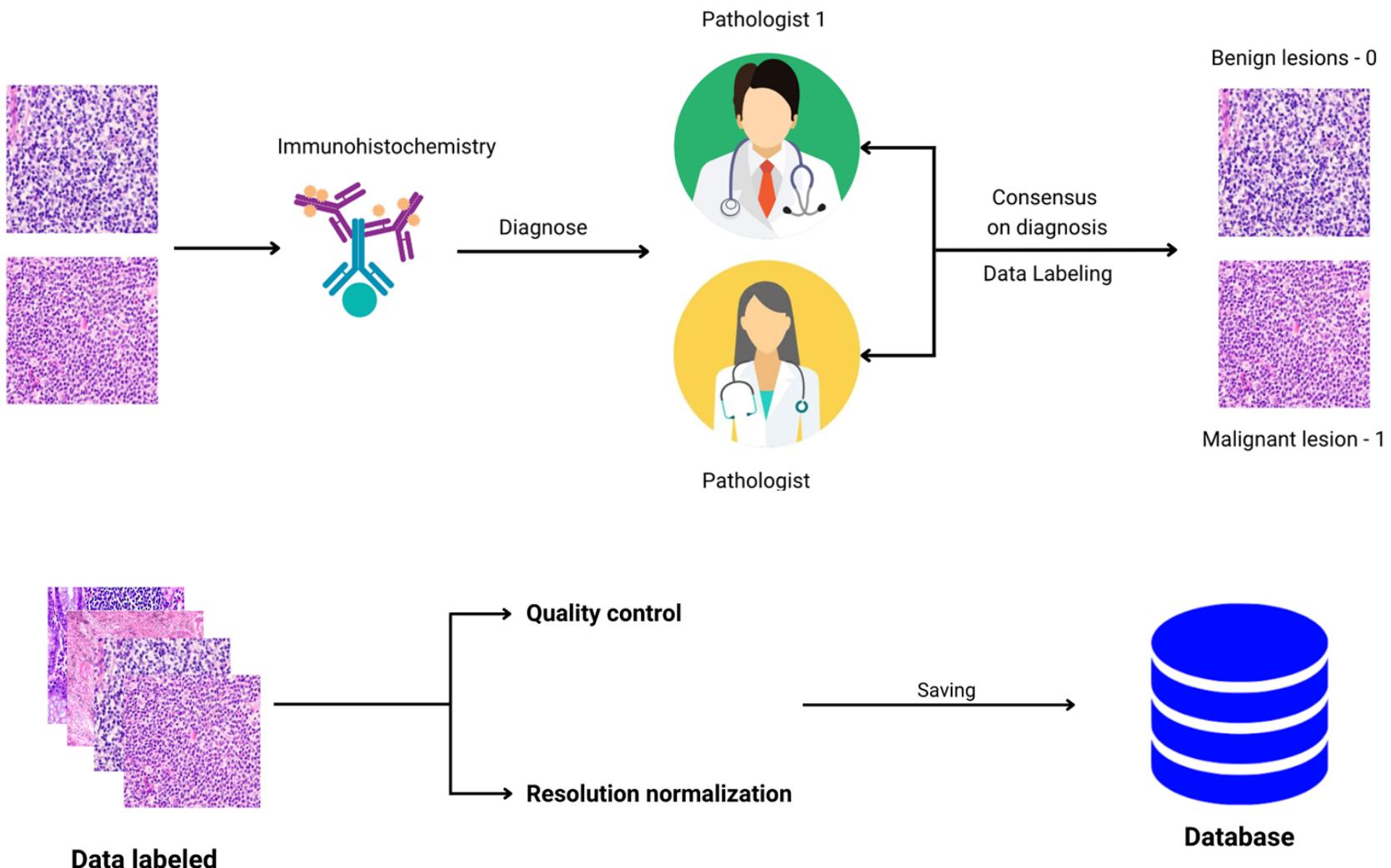


2. Build a simple web application, allowing the deployment and operation of the H&E stained biopsy specimen image classification model.

# DeepHE: Lymphôma đường tiêu hóa



## Methods | Data Collection



# Model Performance | Comparison

**Our method**

Model	Accuracy	Precision	Recall	F1-Score	AUC	Sensitivity	Specificity
Fusion by Case	0.96296	0.96629	0.94505	0.95556	0.99288	0.94505	0.976
Fusion by Image	0.98131	0.97753	0.97753	0.97753	0.99613	0.97753	0.984

**Miyoshi et al, 2020**  
*Laboratory Investigation*

	Category	Recall	Precision	F1 Score
Classifier	DLBCL	1.000	0.969	0.984
	FL	0.941	1.000	0.970
	RL	0.905	0.950	0.927

**Li et al, 2020**  
*Nature Communications*

CNNs	Diagnostic accuracy (%)		
	Model A	Model B	Model C
AlexNet	92.08	93.57	95.12
GoogleNet	95.05	90.68	95.12
Vgg16	95.05	94.53	99.50
ResNet18	92.08	88.42	95.12
SqueezeNet	92.08	89.39	92.68
MobileNetv2	90.10	88.42	92.68
Inceptionv3	90.10	93.89	87.80
DenseNet201	90.10	84.57	95.12
Xception	98.02	91.32	85.37
Vgg19	87.13	93.25	92.68
Places365GoogleNet	96.04	92.93	95.12
InceptionResNetv2	94.06	96.14	96.02
ResNet50	86.14	90.68	87.80
ResNet101	89.11	91.96	97.56
NASNetMobile	95.05	85.21	90.24
NASNetLarge	95.05	91.96	92.50
ShuffleNet	87.13	88.42	85.37
GOTDP-MP-CNNs (with combined 17 CNNs)	100.00	99.71	100.00

## Data Collection | To 25 Jul 2025

- Number of IDs in Benign class: **165**
- Number of IDs in Tumor class: **45**

**Image-level**

	<b>Training</b>	<b>Evaluation</b>	
<b>Benign</b>	626	157	<b>783</b>
<b>Tumor</b>	502	126	<b>628</b>
<b>Total</b>	<b>1128</b>	<b>283</b>	<b>1411</b>

**Case-level**

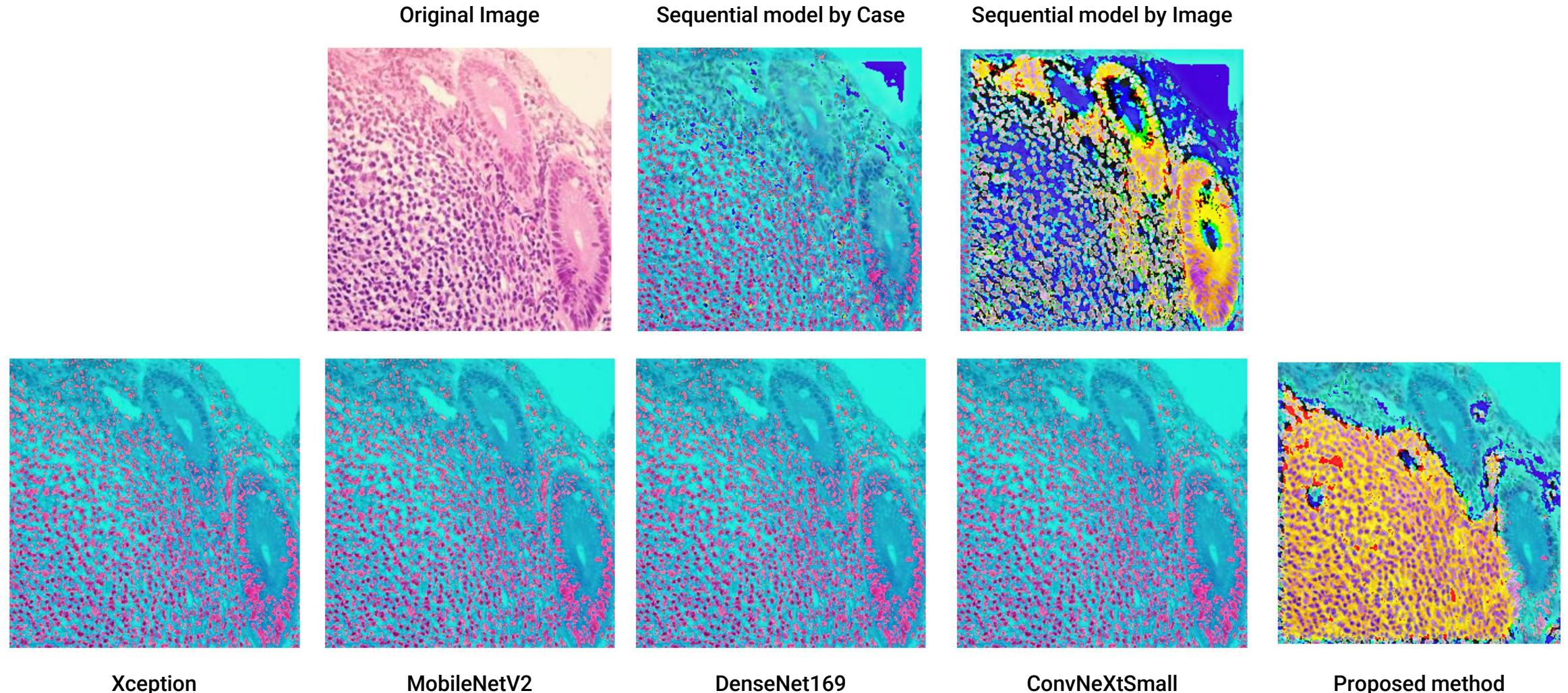
	<b>Training</b>		<b>Evaluation</b>	
	<b>Patient ID</b>	<b>Samples</b>	<b>Patient ID</b>	<b>Samples</b>
<b>Benign</b>	130	626	35	157
<b>Tumor</b>	37	497	8	131
<b>Total</b>	167	<b>1123</b>	43	<b>288</b>

## Model Performance | To 25 Jul 2025

Strategy	Resolutions	Trainable params	Model's size (GB)	Performance on testing set				
				Accuracy	Sensitivity	Specificity	F1-score	AUC
Case-level	256 x 256	133,982,985	0.51	0.9792	0.9847	0.9745	0.9773	0.9994
Image-level	256 x 256	133,982,985	0.51	0.9965	0.9921	1.000	0.9960	0.9999

# Model Explanation

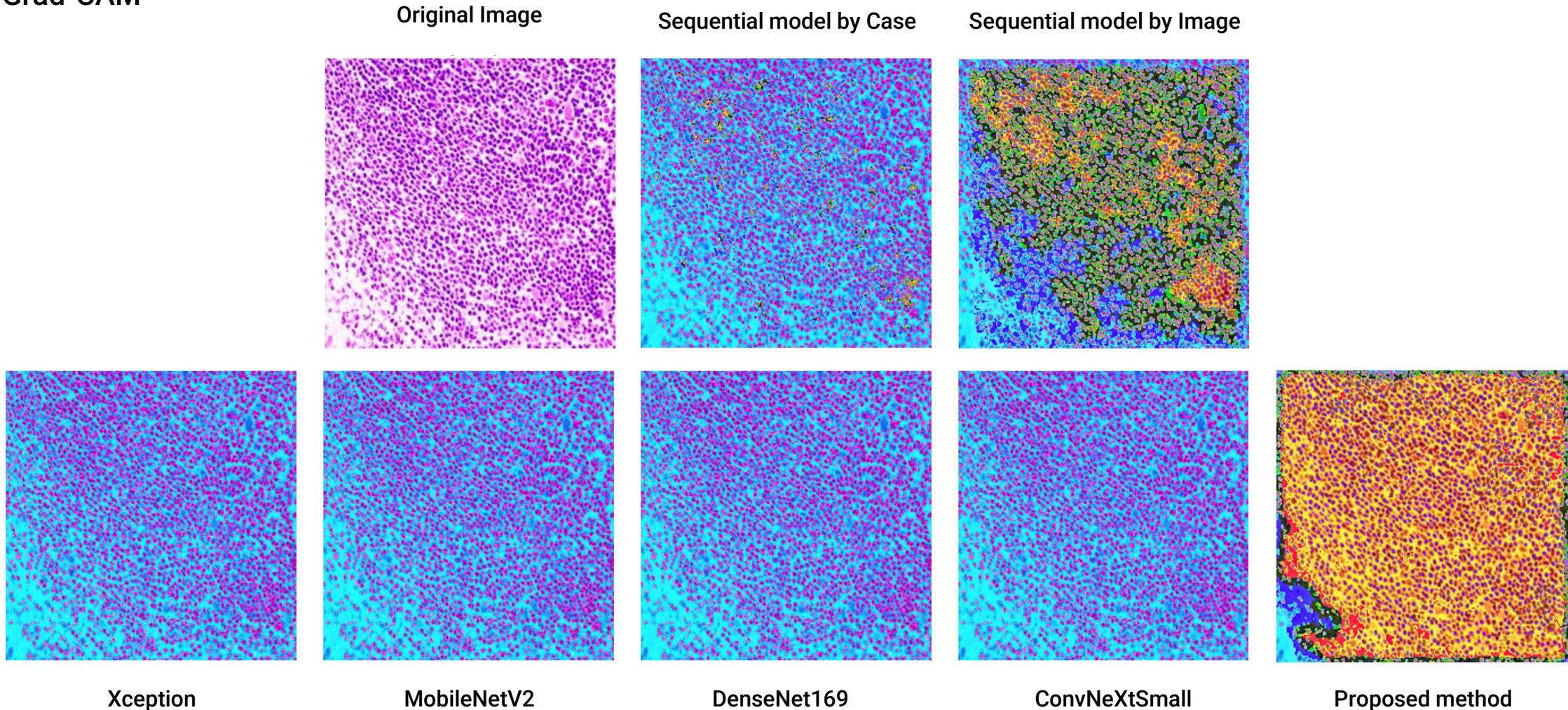
## Grad-CAM



Comprehensive “behavior” of propose method

# Model Explanation

## Grad-CAM

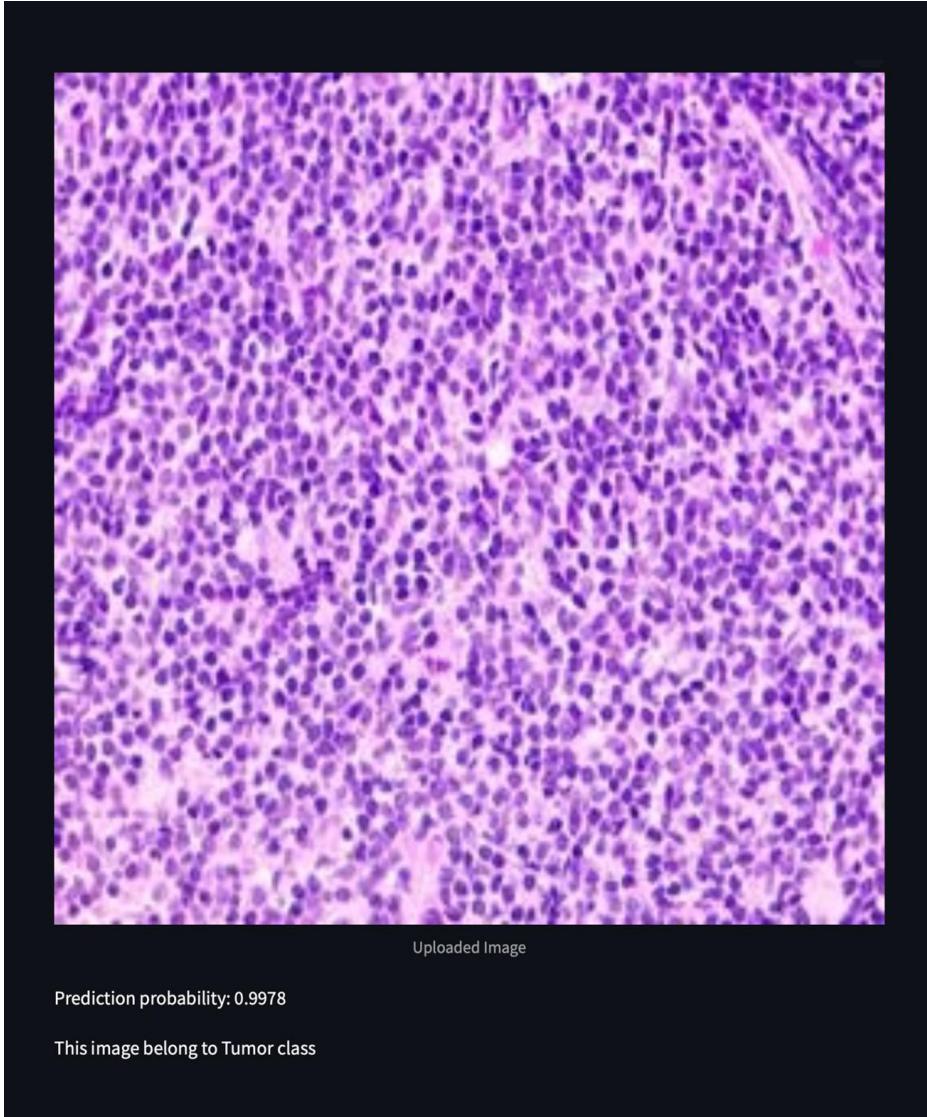


Comprehensive “behavior” of propose method

# Model Deployment | Machine learning operation platform

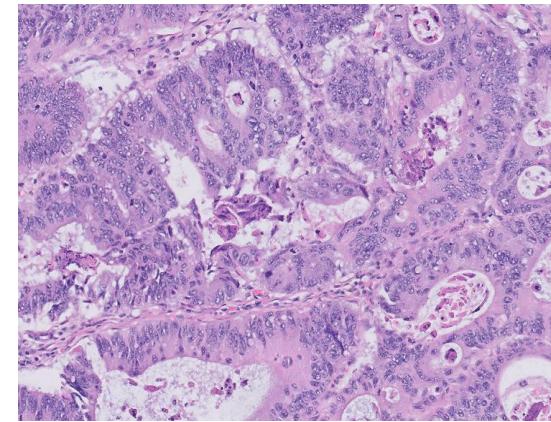
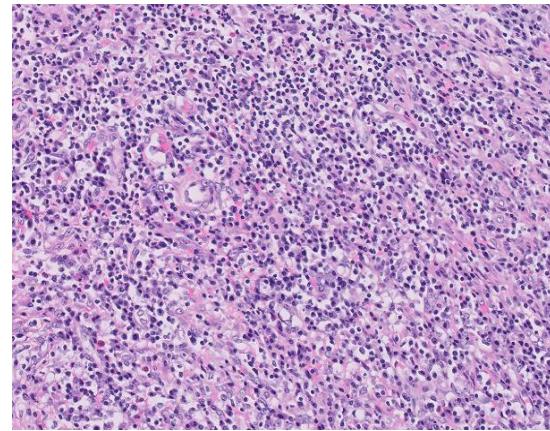
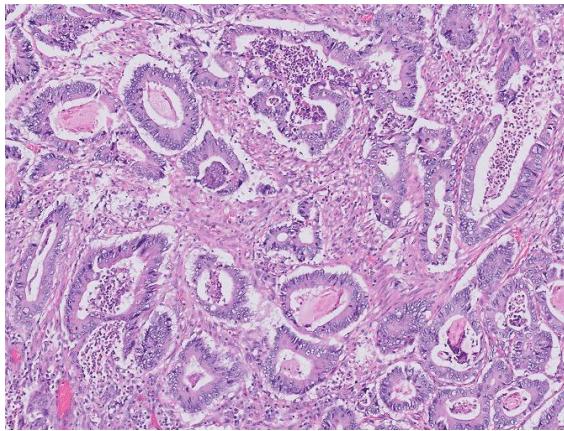
User Interface by Streamlit

The screenshot shows a Streamlit application window titled "CLASSIFYING LYMPHOID LESIONS IN THE GASTROINTESTINAL TRACT". Below the title, a subtitle reads "For follicular lymphoma involving small to medium-sized cells." A photograph of a light-colored microscope is displayed. A text instruction "Upload an image to get a class prediction." is present, followed by a file upload input field with the placeholder "Drag and drop file here" and a "Browse files" button. A file preview section shows an image file named "tm\_fl\_tu\_tho\_14-SO-3419\_10X-1\_40X\_(11)\_mlpc.jpeg" with a size of "27.1KB". A "Submit" button is at the bottom. The top of the window shows browser controls for localhost.

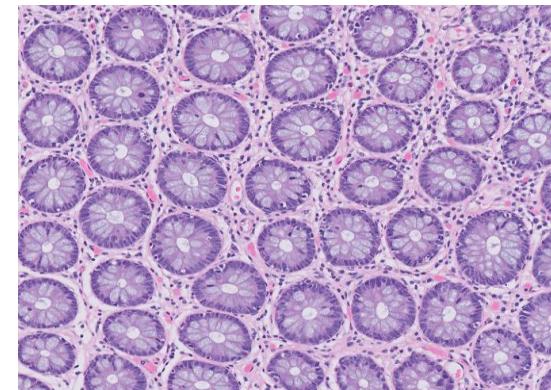
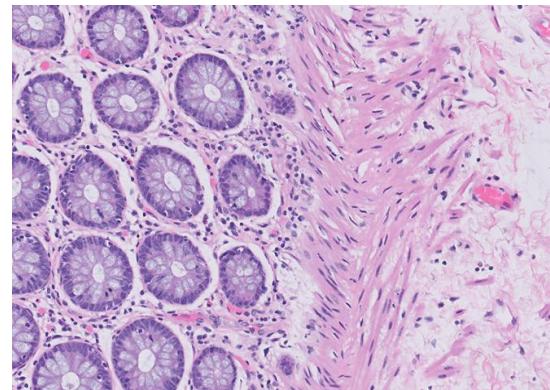


# Xây dựng ứng dụng AI cho phân loại ảnh nhuộm Hematoxylin và Eosin (HE) trong sàng lọc ung thư đại thực tràng

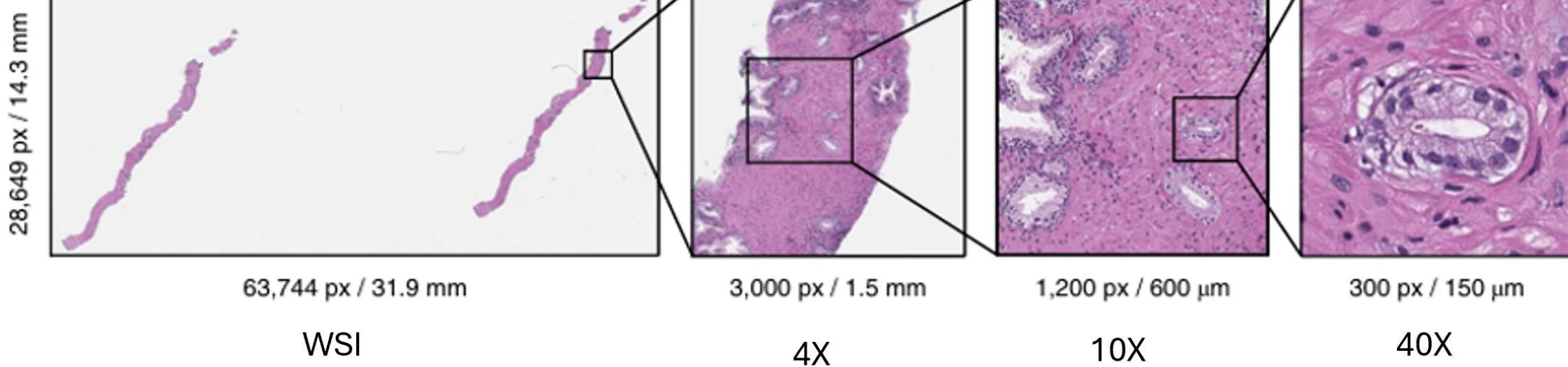
**Normal  
(10X)**



**Tumor  
(10X)**



# XÂY DỰNG MÔ HÌNH HỌC SÂU TRONG CHẨN ĐOÁN UNG THƯ TUYẾN TIỀN LIỆT TRÊN ẢNH CHỤP TIÊU BẢN SINH THIẾT KIM NHỎ NHUỘM HEMATOXYLIN-EOSIN (HE)

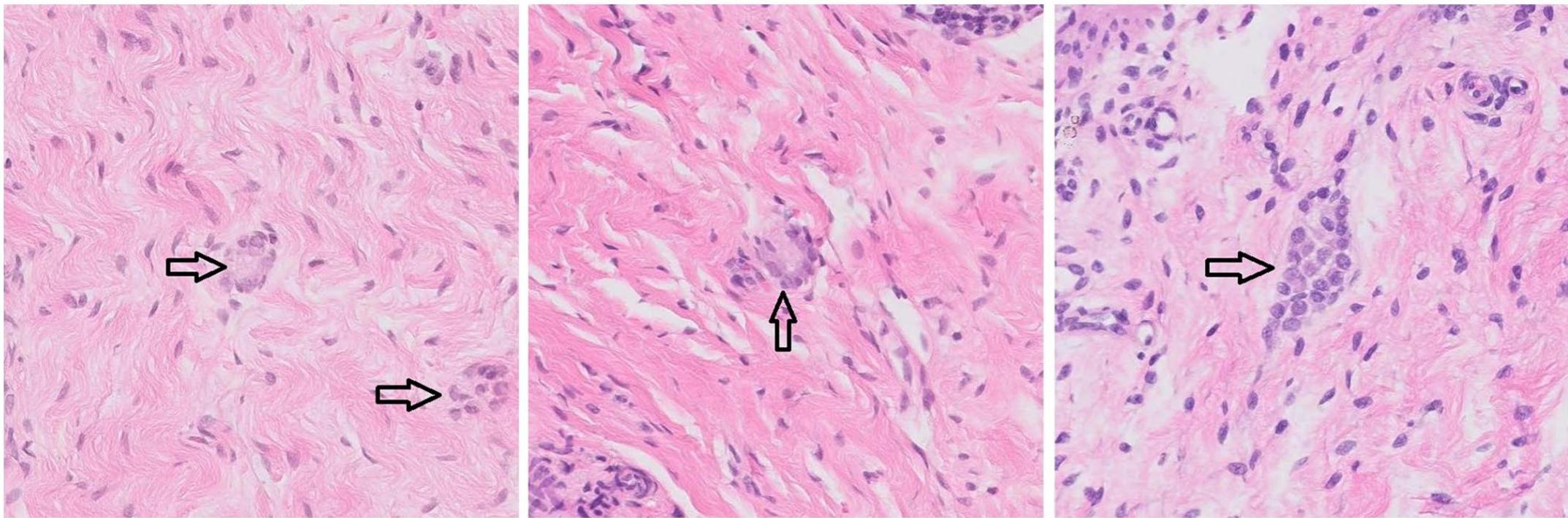


BÁC SĨ GPB NHÌN  
TỔNG THỂ

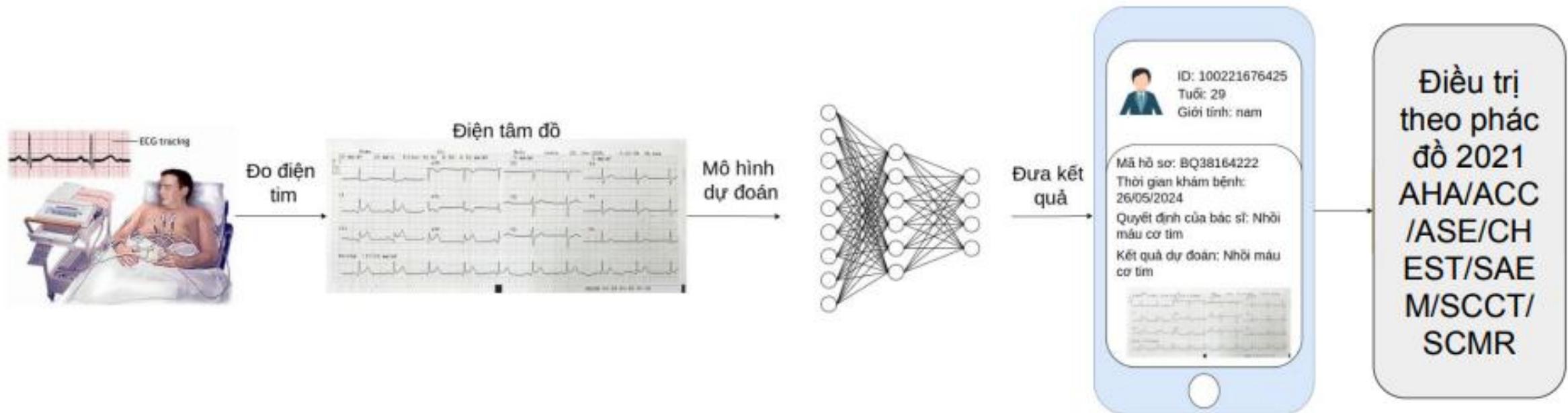
VÙNG NGHI NGỜ

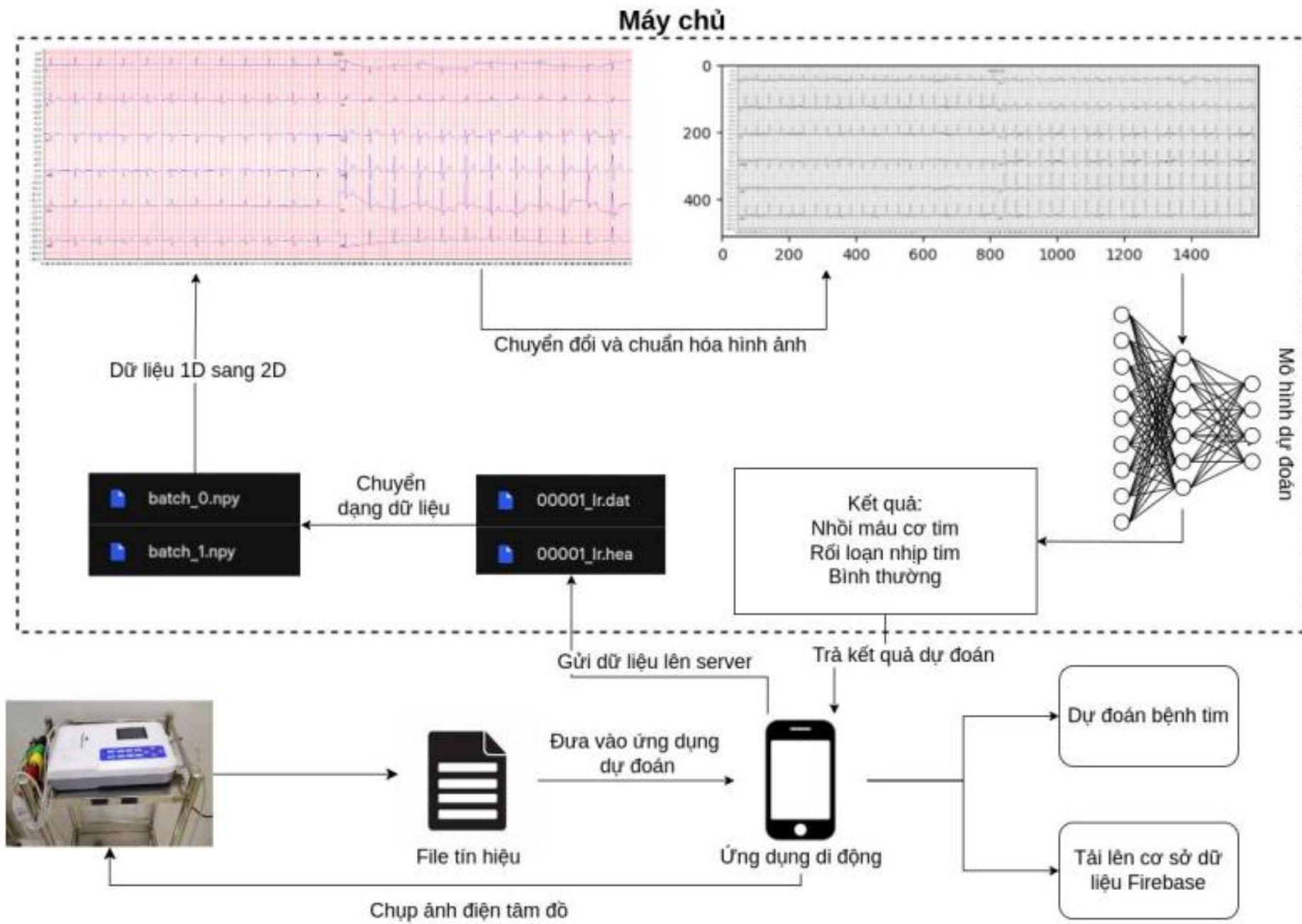
CHI TIẾT

# Xây dựng mô hình học sâu trên ảnh nhuộm Hematoxylin và Eosin (HE) cho chẩn đoán chính xác bệnh hirschprung



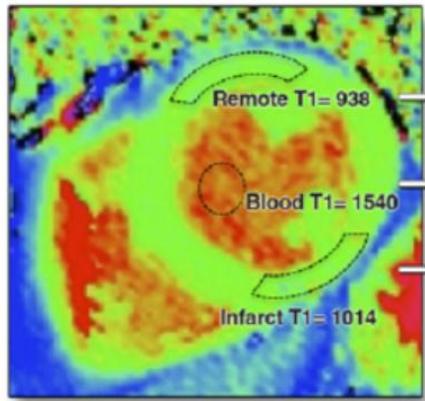
# DeepECG: các bệnh về tim mạch



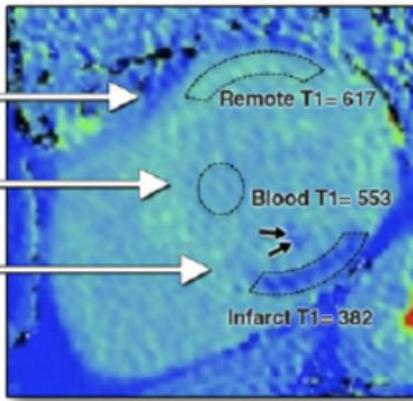


# VECV: bệnh cơ tim phì đại

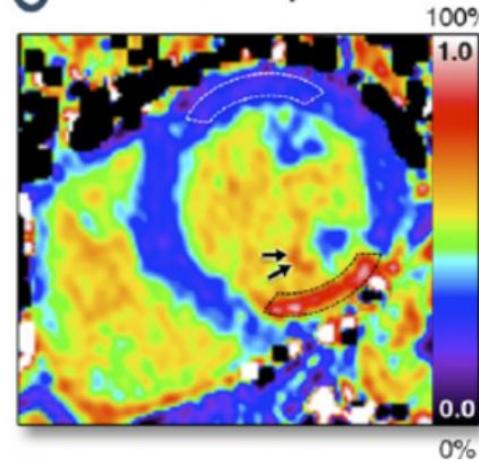
A Pre-contrast T1 Map



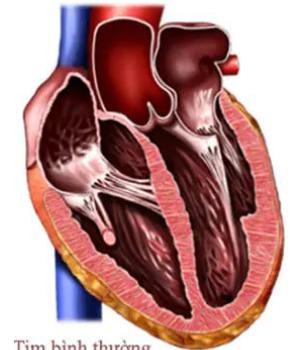
B Post-contrast T1 Map



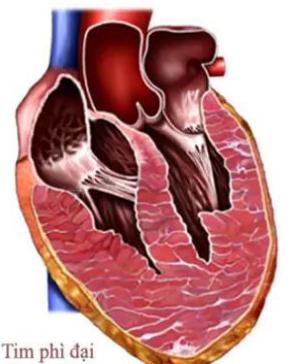
C ECV Map



DL Model



Tim bình thường



Tim phì đại

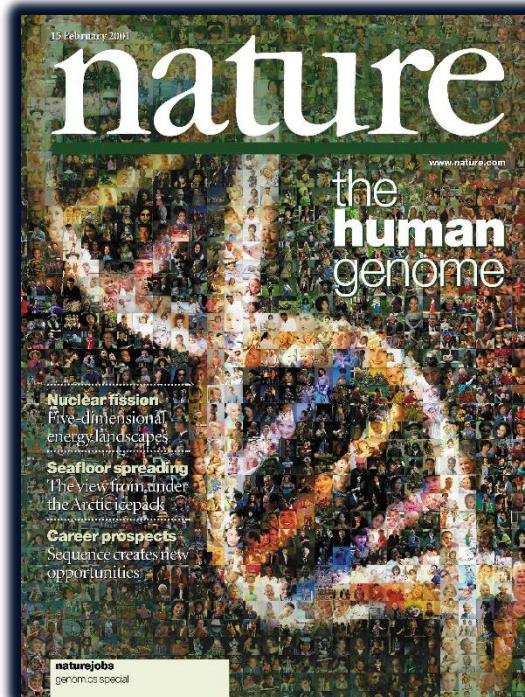
Giới thiệu về

# GIẢI TRÌNH TỰ GEN THẾ HỆ MỚI

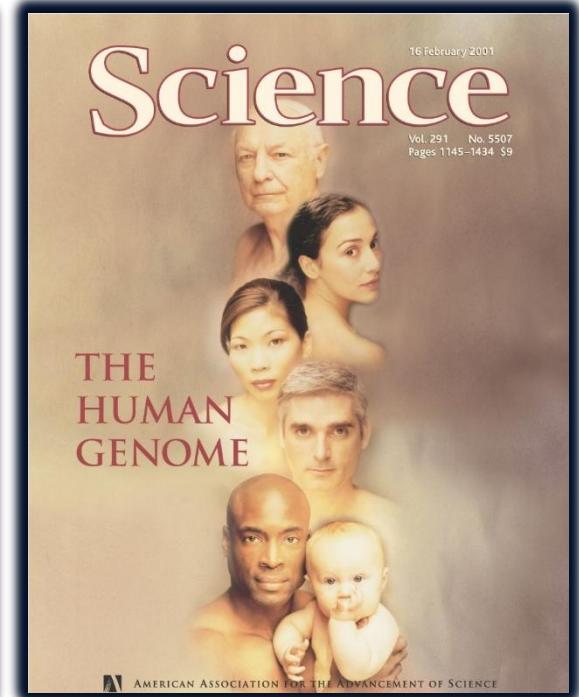
# Dự án hệ gen người HGP (Oct 1990 - April 2003)

- 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).

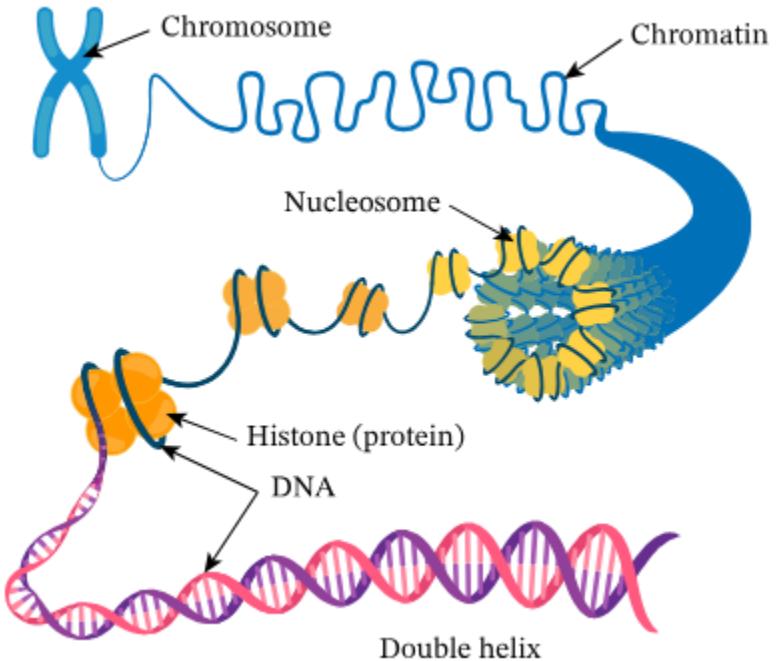
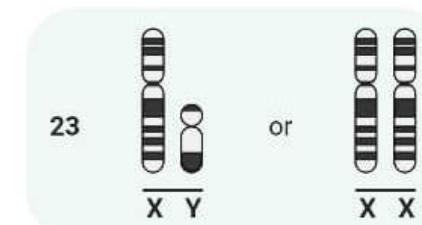
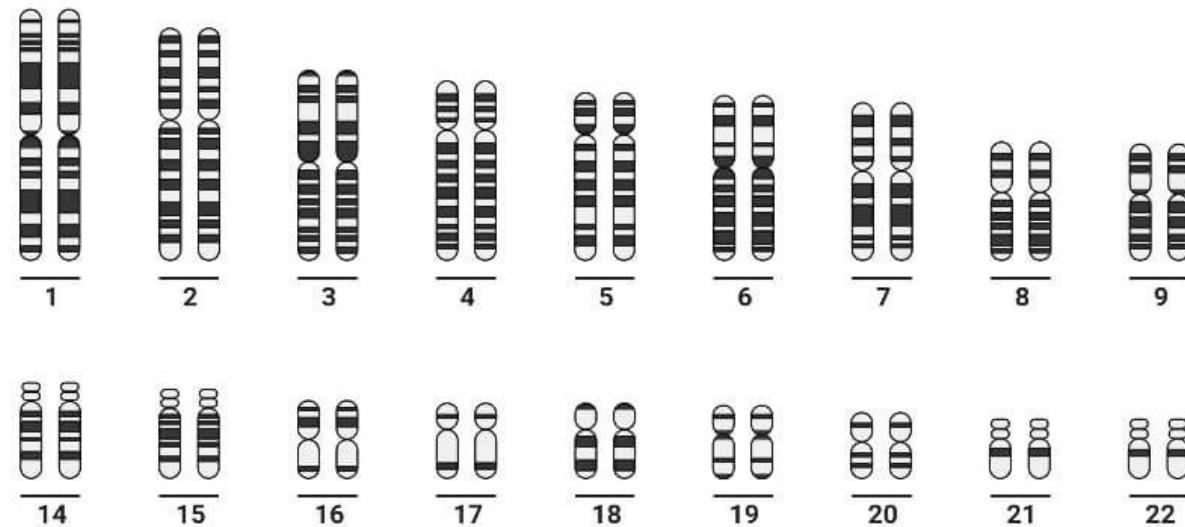


HGP Paper



Venter/Celera Paper

# Human Karyotype



# Hệ Gen người

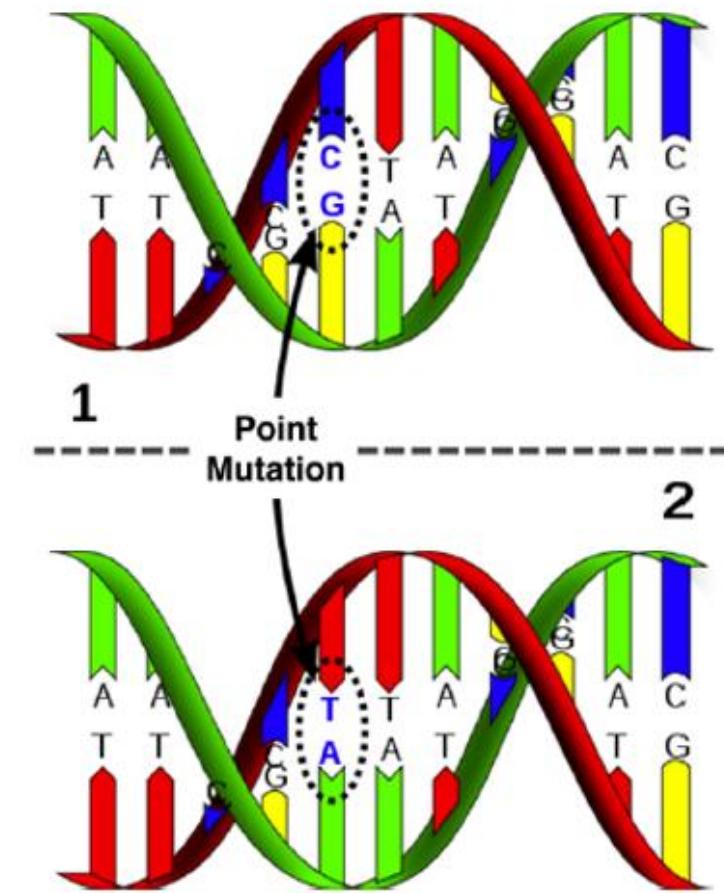
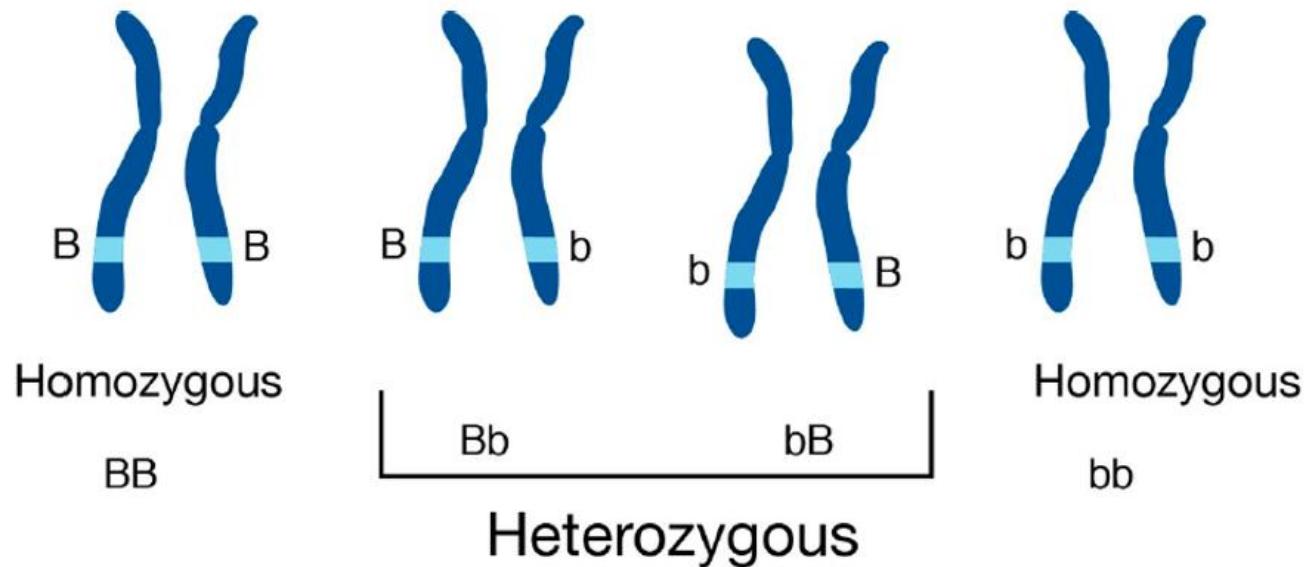


<https://www.ncbi.nlm.nih.gov/nuccore/806904736>

48541 agcccttcaa agaaatgttc tcagcaggca tggagccca gacttgctcc ctttggtag  
48601 agagccgggt tgaagggtac tgaagtaaa tgggacgta gaggcggggg ggggtgtg  
48661 ttccctggagg tgggggggtg gggAACCTGC tttgtactga gatgcacccc tgccagg  
48721 gcctgaagat ttgaggcggg gggcaggggg gcggagtcaa gtcattttac tggtaagtaa  
48781 ttttaaacct tttaatatta aagcaaacgt ggatgttaa tgaataaaat tcattctga  
48841 ataaaaattt cacgtgatgt tgaaaaataa cacggggctt cagagaggac ttctggctg  
48901 gcagcagact ccagattccc agggccccctg caccctcctc tgcccacagg gcaccta  
48961 ggagaagggtg tgggaggaga gccaggccgg agtcagagca cactggtagc tccacattt  
49021 cagcgtgccc tgccctctc ctgaggctt gcaacgtgca atatgctaag caaactcccc  
49081 ctgtcccggt ccagttctg aggacaagag ccaccacatg tagcaaataa agaccagca  
49141 accctttgac tcatctttgt gagttctgg aatcagagg tagccacatc gctgagaggt  
49201 ggagtgaagc actcgggtga aaaggtacaa ggaagtccgg gacaggagtg tggggacatc  
49261 acctagacaa tgacagagaa gaggggcaca gccgagttag gggagagggg cggcagtc  
49321 tacatccctt ggcctgaagc acgtccagg gcagaaggaa aaacactgtc tttgggggtcc  
49381 aagagacctg agttcaattt ctggctccac cactgaccac ctgtgttaacc ttgaactgt  
49441 gctgcctgaa ctcagggtt cccttctaaa aatagaggag aaaaggatgc atttctc  
49501 gcccctgtga gaacaaatg gtcaagcac caaggaccc cagcaaagggt cgggctg  
49561 cccgcctggc caaacctttc ctcttcggg ggcacggca accgttagttt gacagaagag  
49621 cagcacccctt atttaatgtt tcccagcatg ttttttttttttttttttttttttttttttt  
49681 ggctgcttcc tcatt  
49741 ggcacttggc aaacagcaaa agagtccaaa ttgtcaggct gggccaggcg cagttggct  
49801 tgcctgtat cccagcaatt taggaagcca aggtggccgg atcaccttag gtcaggagtt  
49861 tgagaccagc ctggccaaaca ttgtgaaacc ttgtctctac aaaaatacaa aaattagcc  
49921 ggcatgtatgg cgggtgcctg taatccact tactcggag gctgaggca gagaatcg  
49981 tgaacccgga agggaaagggt tgcaatgtgc caagttgtt ccactgcact ccagctgg  
50041 caacagagcg agactctgtc taaaaaaaaaaa aaaaaaaaaaa aaacaatgca gagctggct  
50101 tgaaaaaaaaac ctgtttccact gcagggccca gtgtccacc ggctgggtg caggctatg  
50161 ggggtggggc ccagcatcag cctctcagca gcccgggag gcggggcgca tccctgt  
50221 ctcgtggctt ggatgtttc tagcccaagt cctagttac acctgcccgc gcctggct  
50281 tcaggagagg cccagggtga ggaggagcat gttaaagggtt aagctgattt ggaagtc  
50341 ttgtggggaaa gcaactcctt gcacattggta ggaaccgaga aagactgacc ccgaggac  
50401 cagccagcat gggcccttcc gggagcccat gttggggat tccctgtgc gccaaggct  
50461 agcccttgtt gtcgcagggtt ctggctctgg cctttcccc tcccatgtcag gacacag  
50521 gagatggctt ctgaggaccc ttgtcagctg tggccctggg aatagattt ccagggagct  
50581 ttaaaggcgc tgagtgttc atccagctaa gcctggggaa ggagcttggc tcaggct  
50641 acagggtgtga caggatggg gactggaaag taagatgtaa aaccctggct ggaggct  
50701 agttccaca gccagcgtt gacaggaggat acccactgtt gcccacca

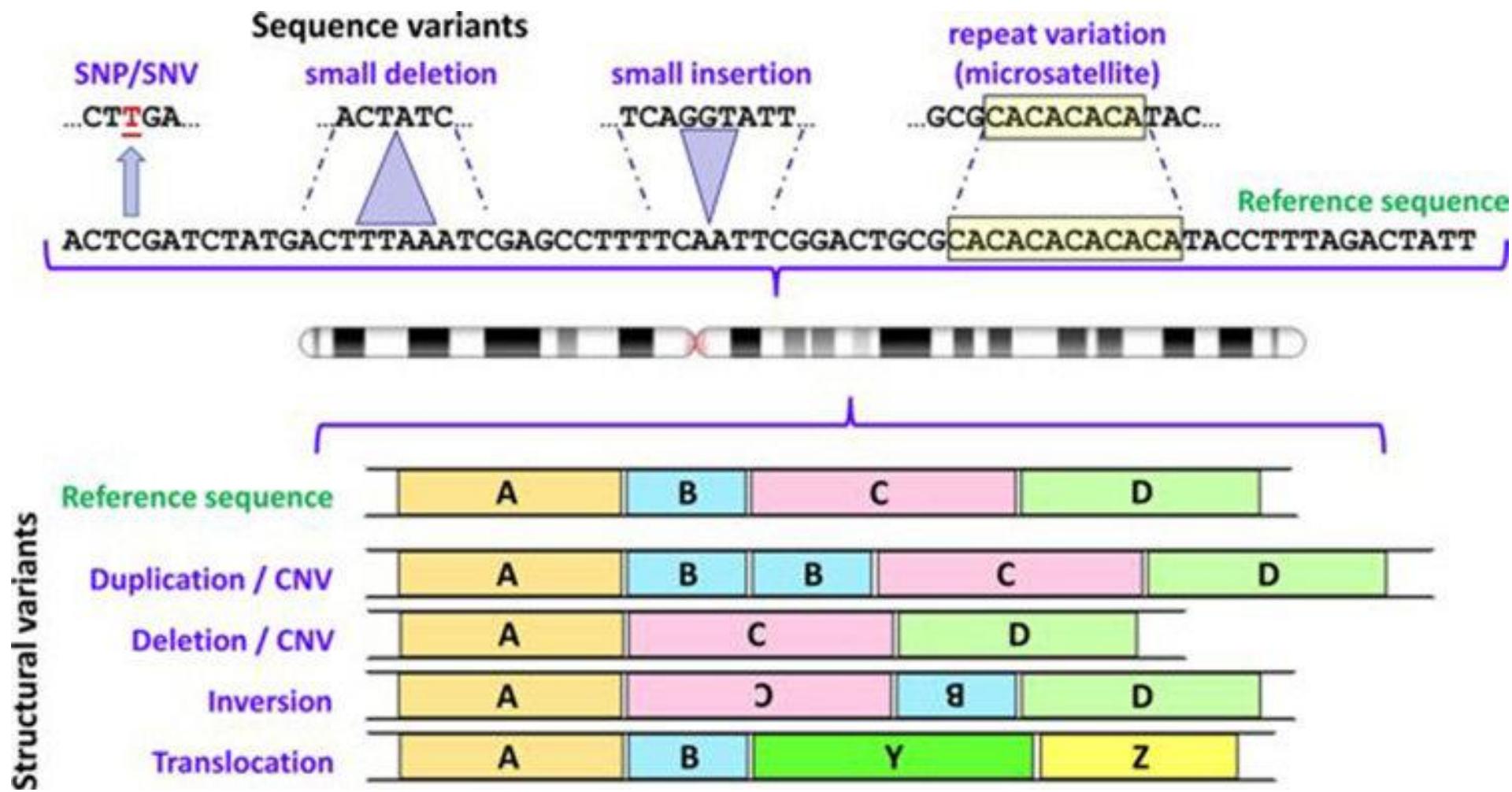
# Các loại biến thể trên hệ gen

- 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

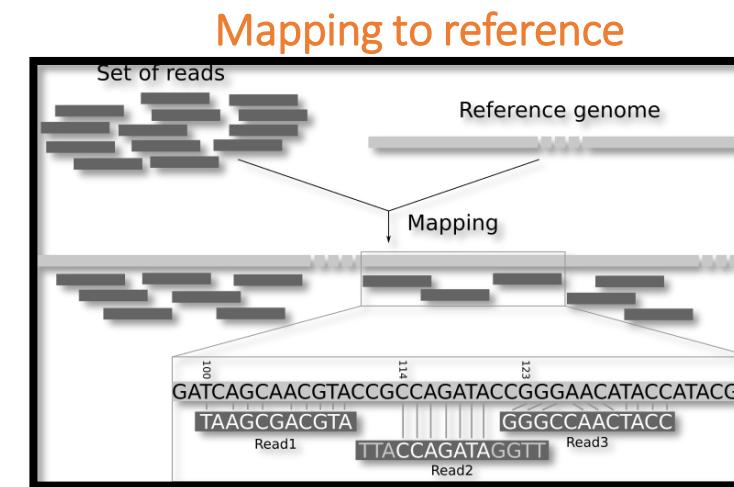
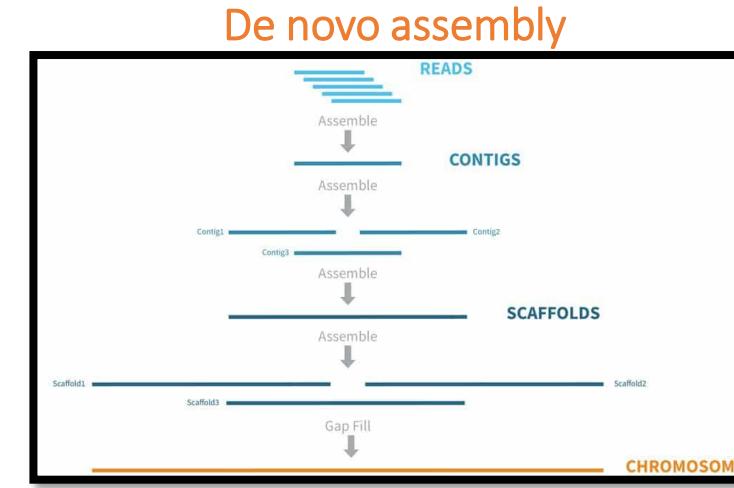
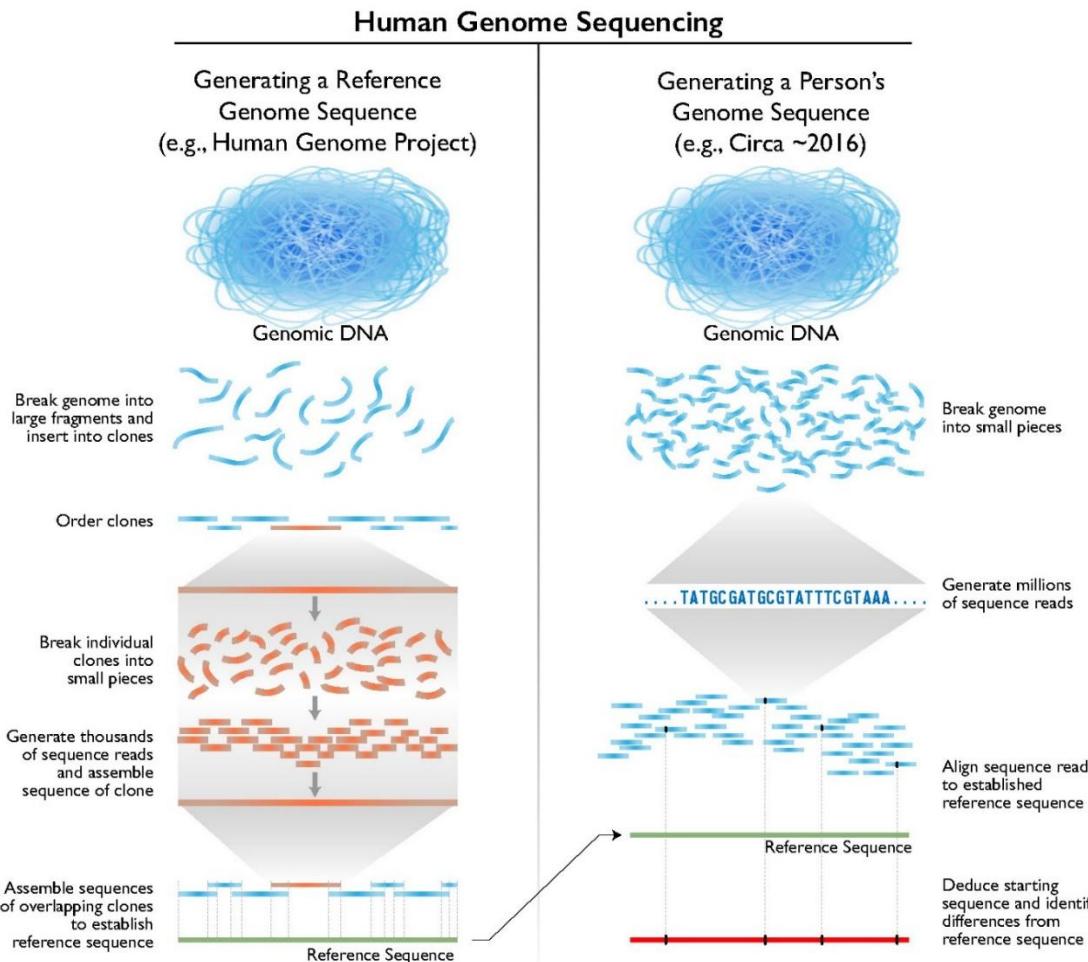


[https://rosalind.info/media/point\\_mutation.png](https://rosalind.info/media/point_mutation.png), <https://en.wikipedia.org/wiki/Zygosity>

# 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

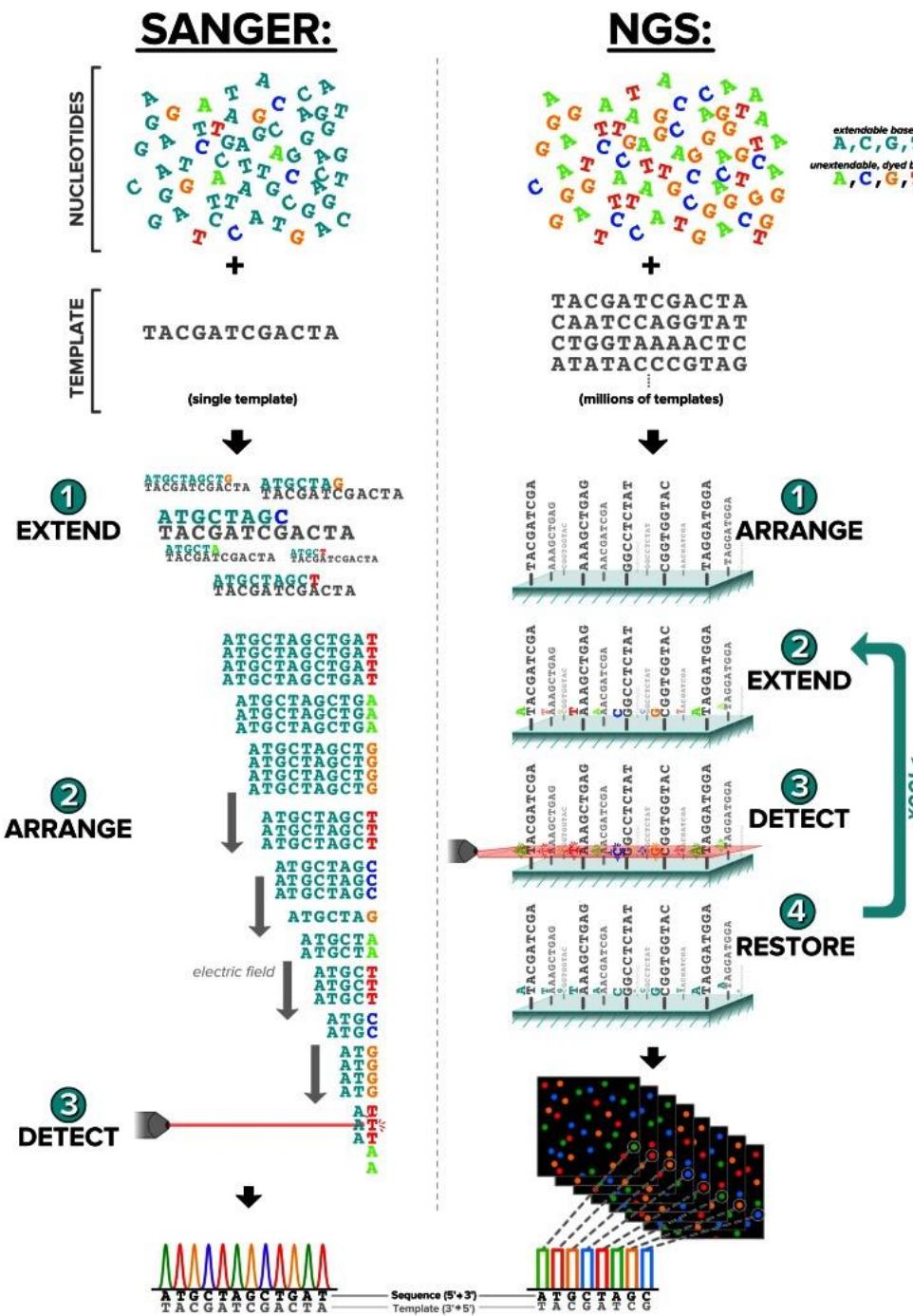


Mapping to reference



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

Read1 : CTCGAATACG

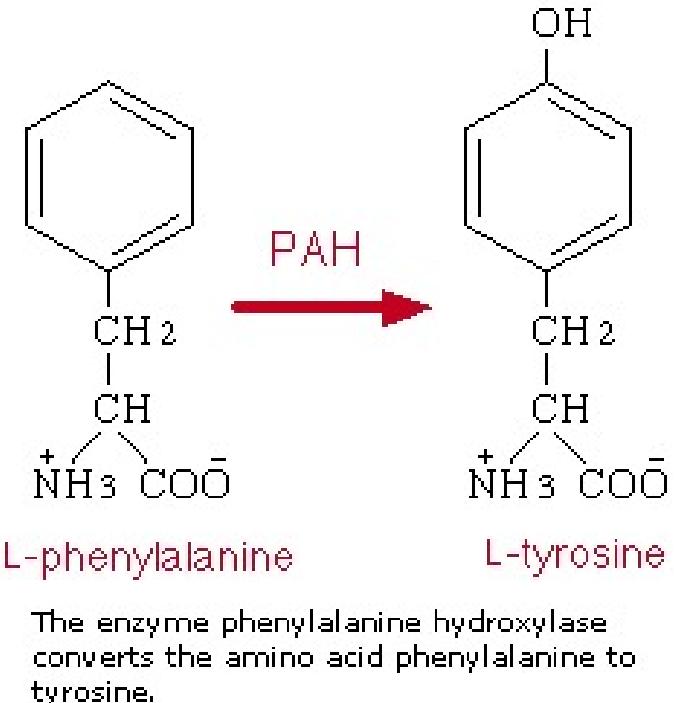


Read1 : CTCGAATACG  
 Read2 : CTCGAATACG  
 Read3 : CTCGAATACG  
 Read4 : CTCGAATACG  
 Read5 : CGCGAATACG  
 Read6 : CGCGAATACG  
 Read7 : CGCGACTACG  
 Read8 : CGCGAATACG

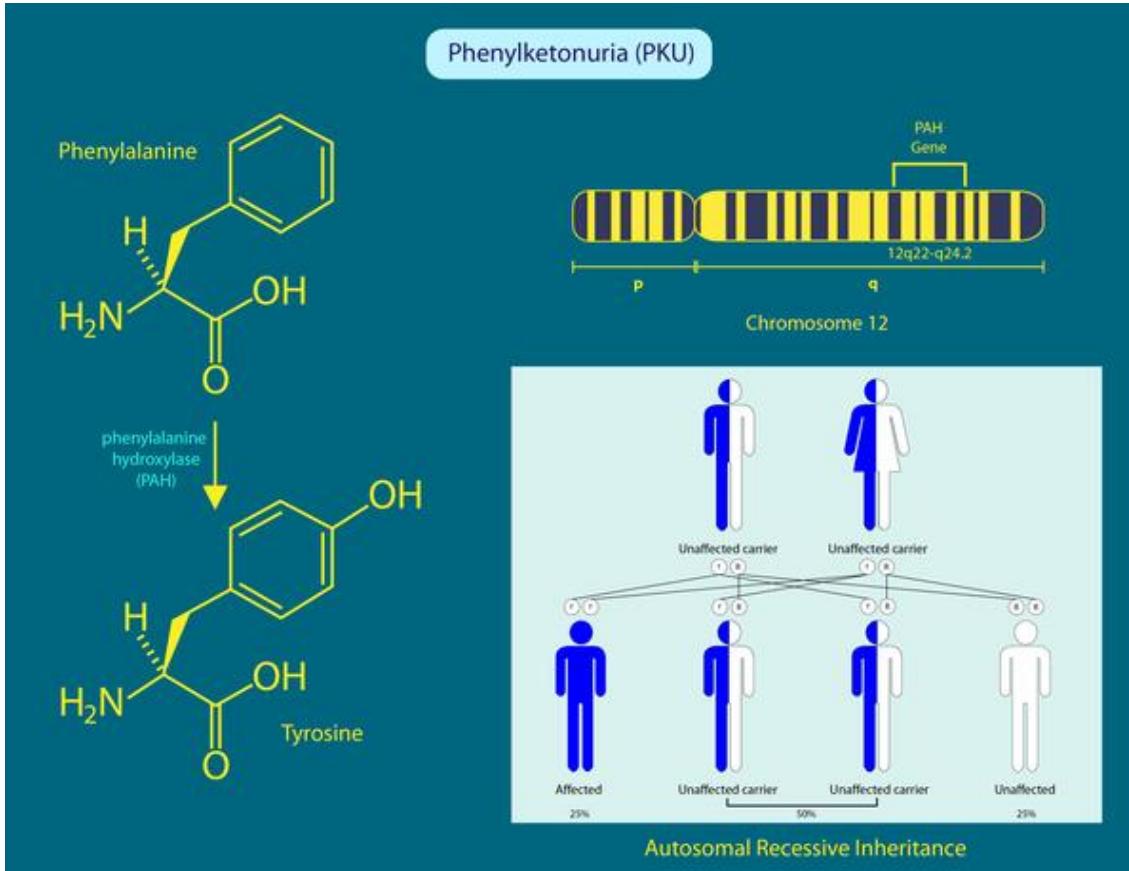
Ví dụ mối quan hệ của  
**BIẾN THỂ GEN** và **BỆNH DI TRUYỀN**

# 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

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

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

## Gene (Nucleotide)

NT seq	1359 nt <a href="#">NT seq</a> atgtccactgcggctggaaaacccaggctggcaggaaactctctgactttggacag gaaacaagcttatattgaagacaactgaatcaaaatggccatatcactgatcttca ctcaaagaagaagtgggtcattggccaaagtattgcgttatttggaggagaatgtat aacctgaccacattgaatcttagacccctcgtttaaagaaaagatgagatgaattttc acccatttgataaaacgttagcctgcgtctgacaaaatcatcaagatcttgggcat gacattggtgcactgtccatgagcttcacgagataagaagaaaagacacagtggccctgg ttcccaagaaccattcaagagctggacagatggccaatcagattctcagctatggagcg gaactggatgctgaccaccctggtttaaagatccgtgtaccgtgcaagacggaaagcag tttgcgtacattgcctacaactaccgcattggcagccccatccctcgagttggaaatacatg gaggaagaaaaagaaaatggggcacagtgttcaagactctgaagtccctgtataaaaacc catgttgcgtatgagttacaatcacattttccacttctgaaaagtactgtggcccttccat gaagataacattccccagctggaaagacgttctcagttctgcagacttgcactggtttc cgccctccgacacctgtggctggcctgtccctctcgggatttctgggtggcccttc cgagtctccactgcacacagtacatcagacatggatccaagccatgtataccccgaa cctgacatctgcatgagctgtggacatgtgcccttgcggatgtggcactgtatggcc cagtttccaggaaattggcctgcctctgggtgcacctgtatgaaatacatgtaaaag ctcgccacaatttactgtggtagttggctctgcaaaacaaggagactccata aaggcatatggtgcggcctgtcatccttggtaattacagtactgtttatcagag aagccaaagcttctcccccggagctggagaagacagccatccaaaattacactgtcag gagttccagcccttctattacgtggcagagatgtttatgatgcacaggagaaagtaagg aactttgctgccacaataccctcgcccttctcagttcgctacgacccatacacccaaagg attggaggcttggacaatacccagcagcttaagatggctgattccattaacagtgaa attggaaatccttgcagtggccctccagaaaataaagtaa
--------	--

## Protein (Amino Acid)

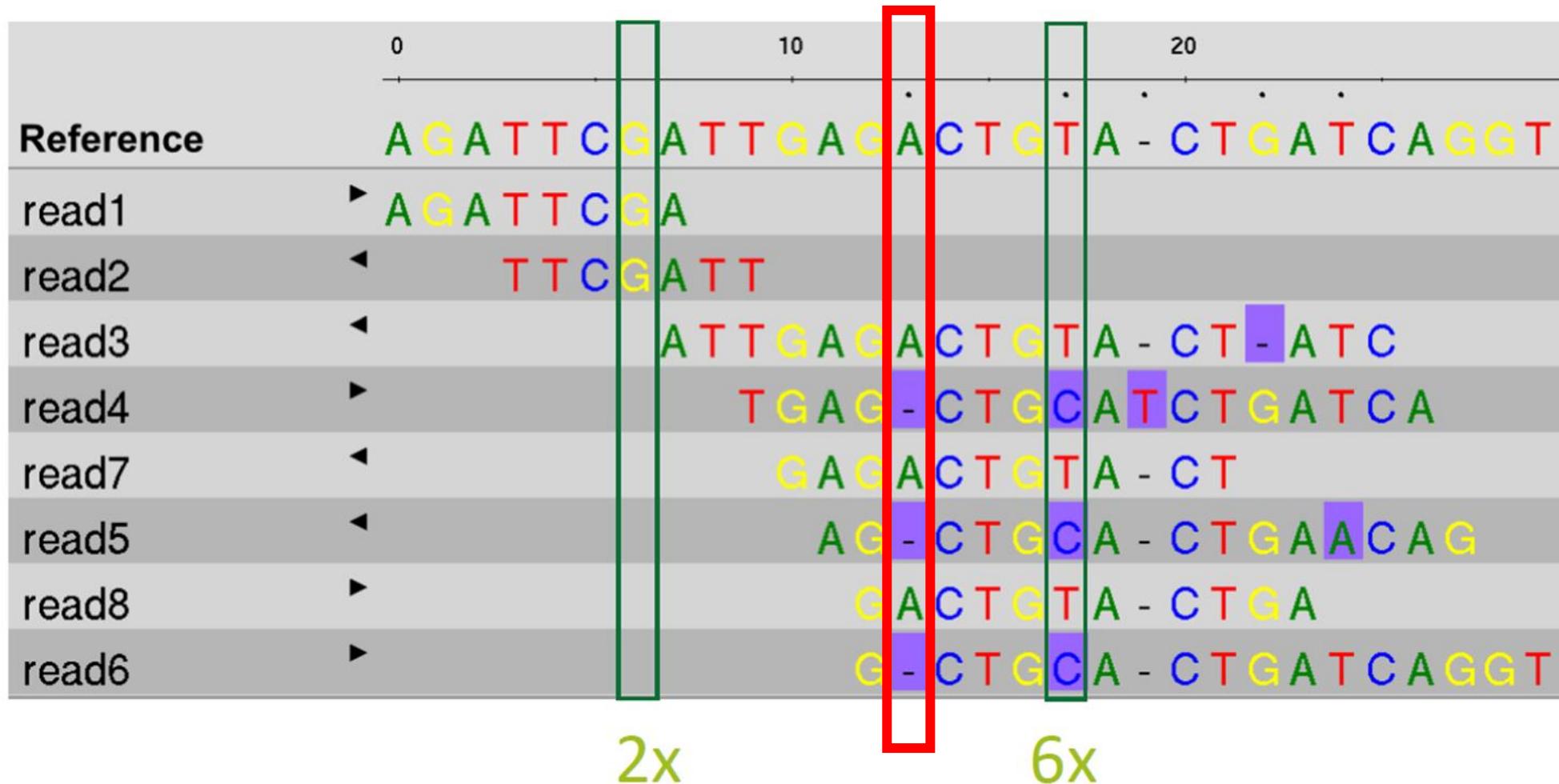
AA seq	452 aa <a href="#">AA seq</a> <a href="#">DB search</a> MSTAVLENPGLGRKLSDFGQETSYIEDNCNQNGAISLIFSLKEEVGALAKVLRLFEENDV NLTHIESRPSRLKKDEYEFFTHLDKRSLPALTNIKILRHDIGATVHELSRDKKKDTVPW FPTIQELDRFANQILSYGAELDADHPGFKDPMVYRARRKQFADIAYNYRHGQPIPRVEYM EEEKKTWGTVFKTLKSLYKTHACYEYNHIFPLLEKYCFGHEDNIPQLEDVSQFLQTCTGF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

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

# Gen PAH



# Variants in PAH



# Cơ sở dữ liệu ClinVar cho gen PAH

## Classification type

- Germline (208)
- Somatic (0)

## Germline classification

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

## Types of conflicts

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

## Molecular consequence

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

## Variation type

- Deletion (39)

## Links from Gene

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

Items: 1 to 100 of 209

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

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

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

# 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

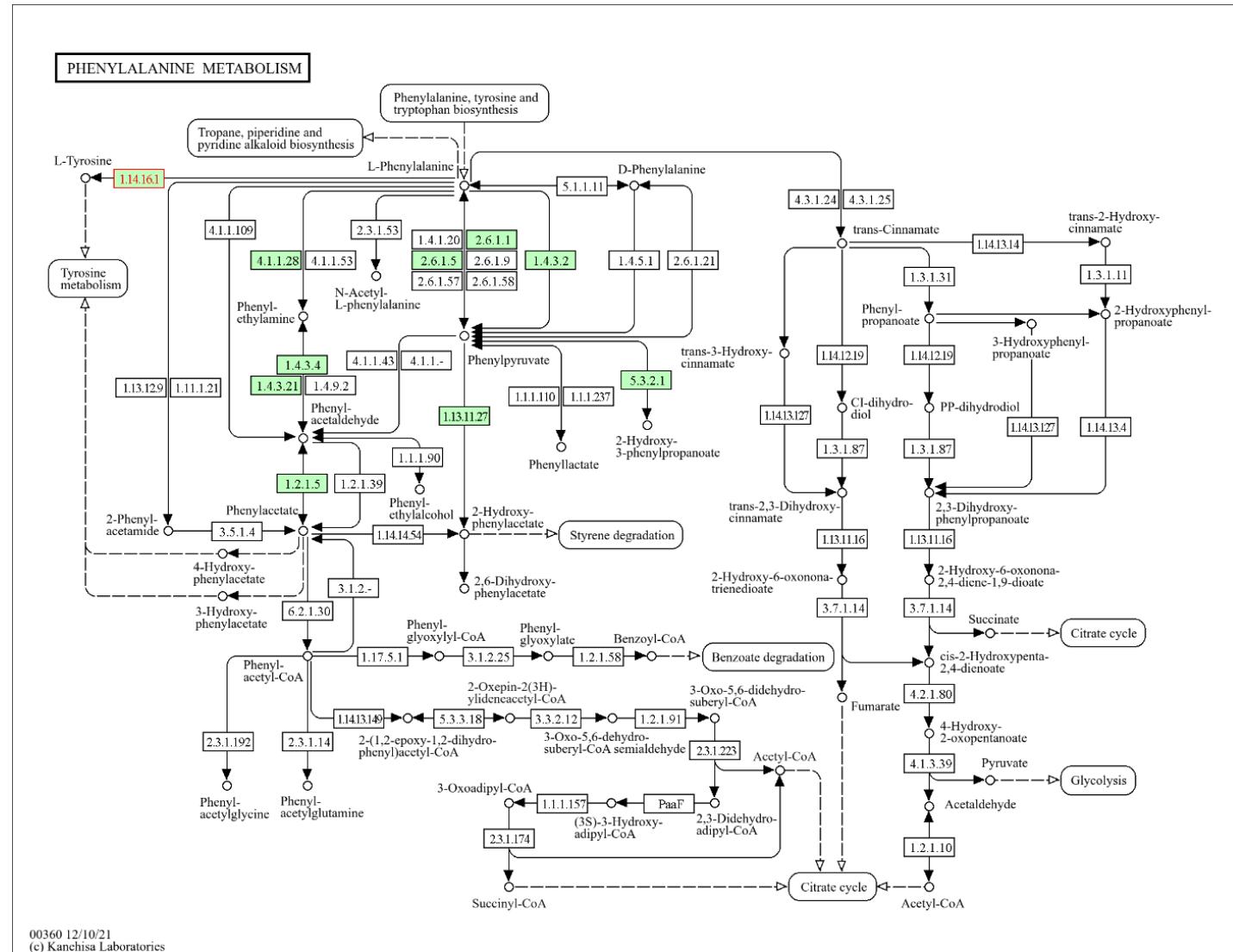


Show

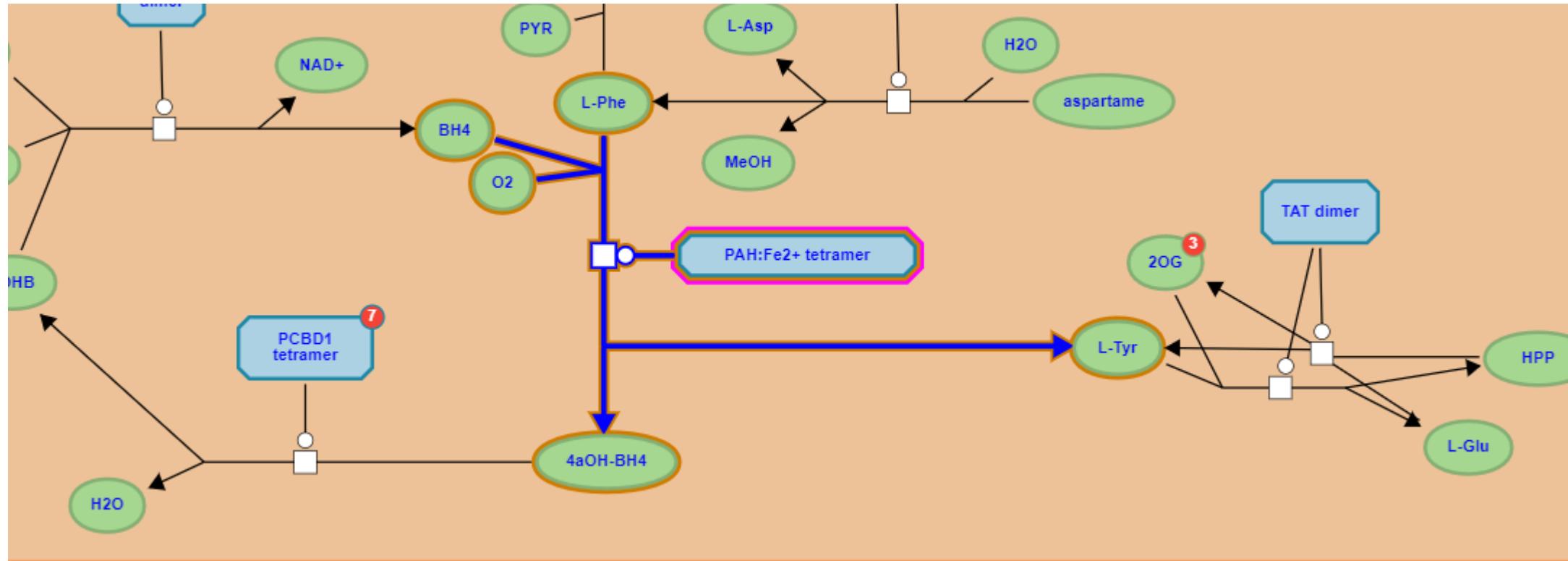
- All Residues
- Variants
- Pathogenic
- Likely Pathogenic
- Uncertain Significance
- Likely Benign
- Benign
- Current Variant

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

# PAH: chuyển hóa Phenylalanine thành Tyrosine

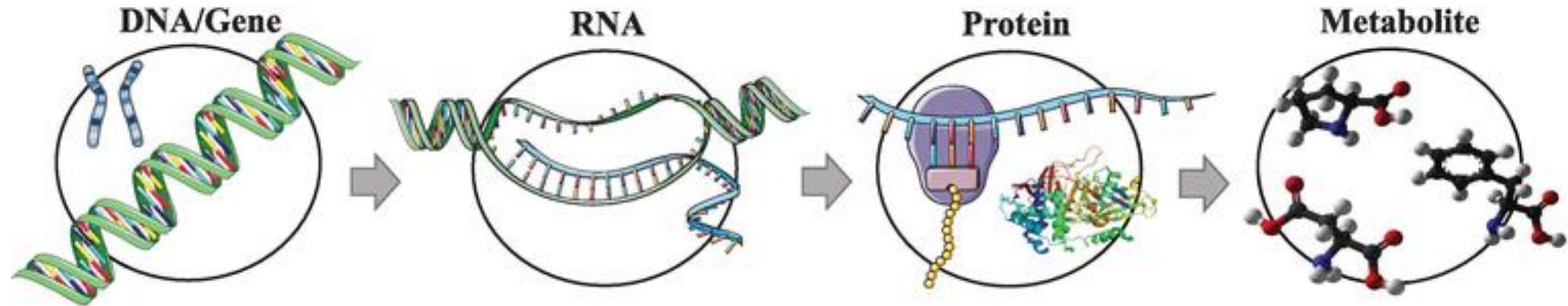


# PAH: chuyển hóa Phenylalanine thành Tyrosine



<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



**Genomics**

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

**Transcriptomics**

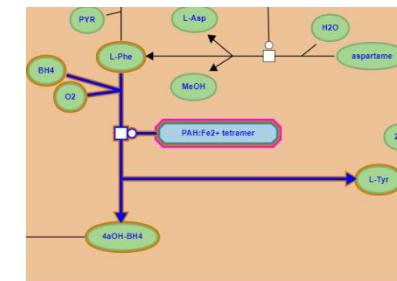
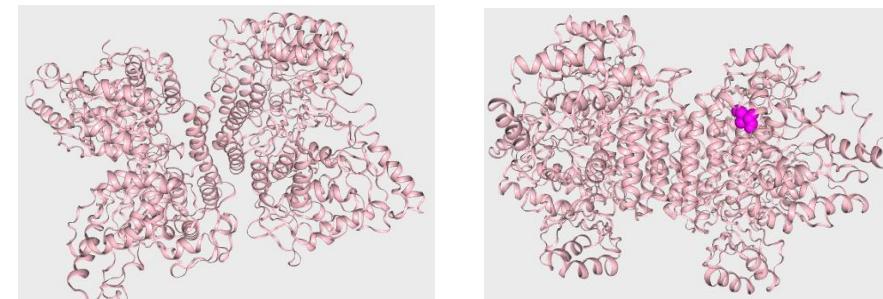
**PAH mRNA**  
Ref ...AUCGAU...  
P1 ...AACGAU...  
  
NM\_000277.3(PAH):c.971T>A

**Proteomics**

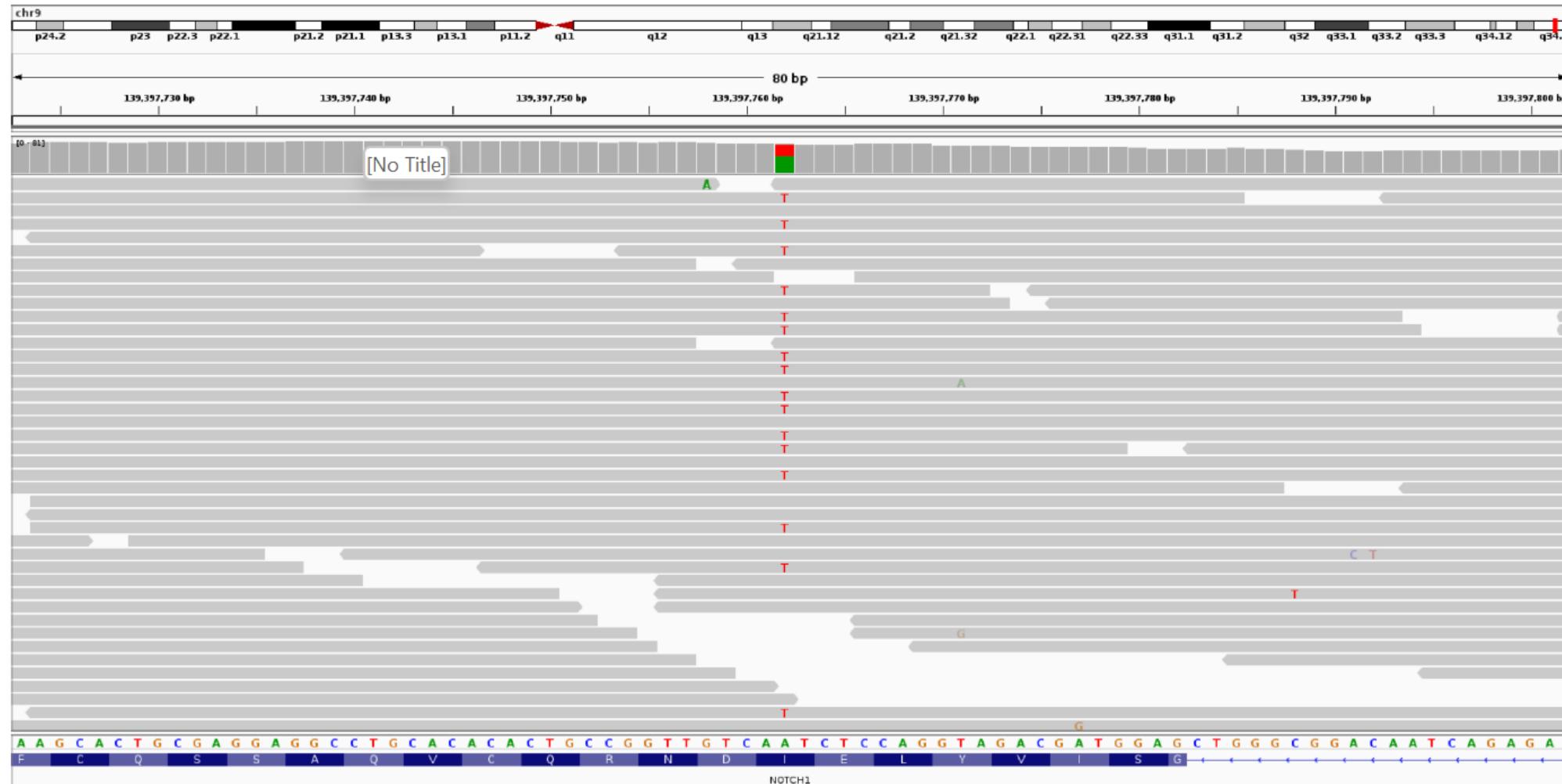
**PAH protein**  
Ref ...Ile-Asp...  
P1 ...Asn-Asp...  
  
NM\_000277.3(PAH):p.Ile324Asn

**Metabolomics**

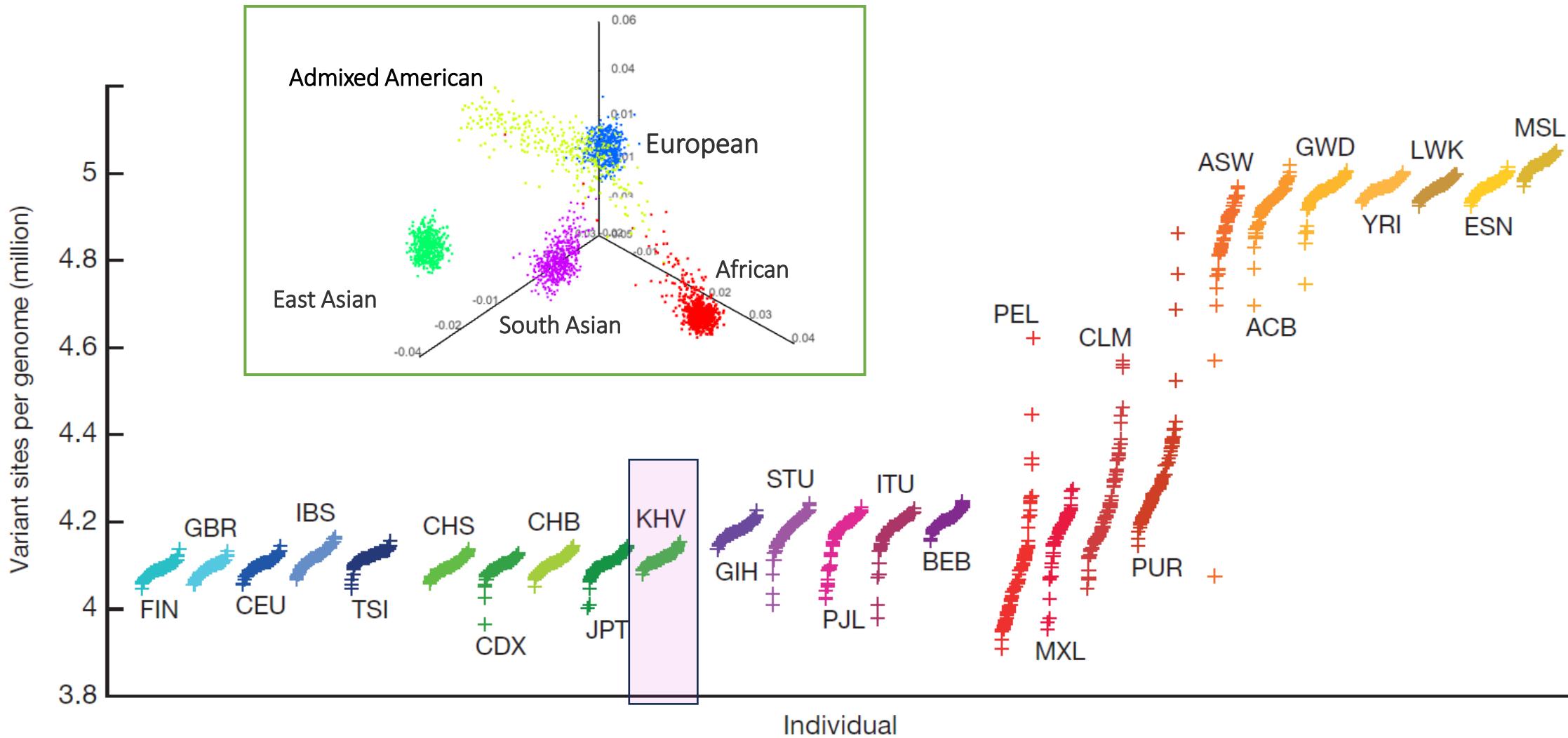
**PAH**  
Ref Phe → Tyr  
  
**PAH**  
P1 Phe ~~→~~ Tyr



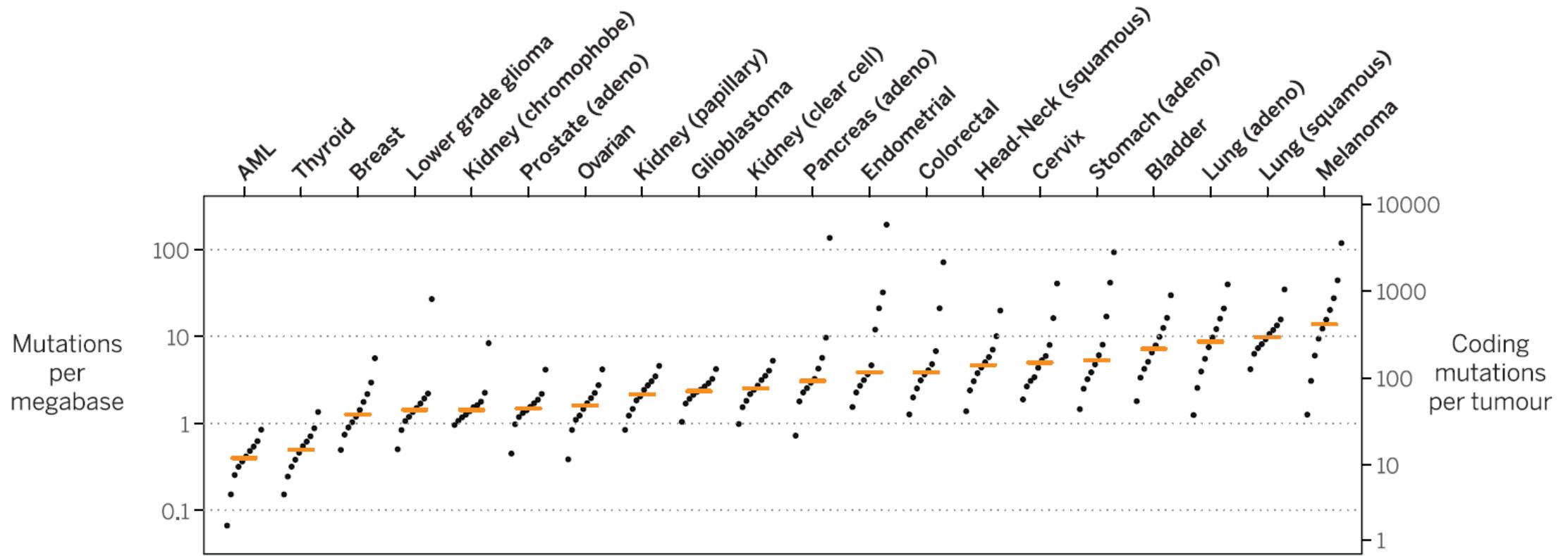
# Alignment and variant viewers



# Human Genome Variation: 1000 Genomes Project



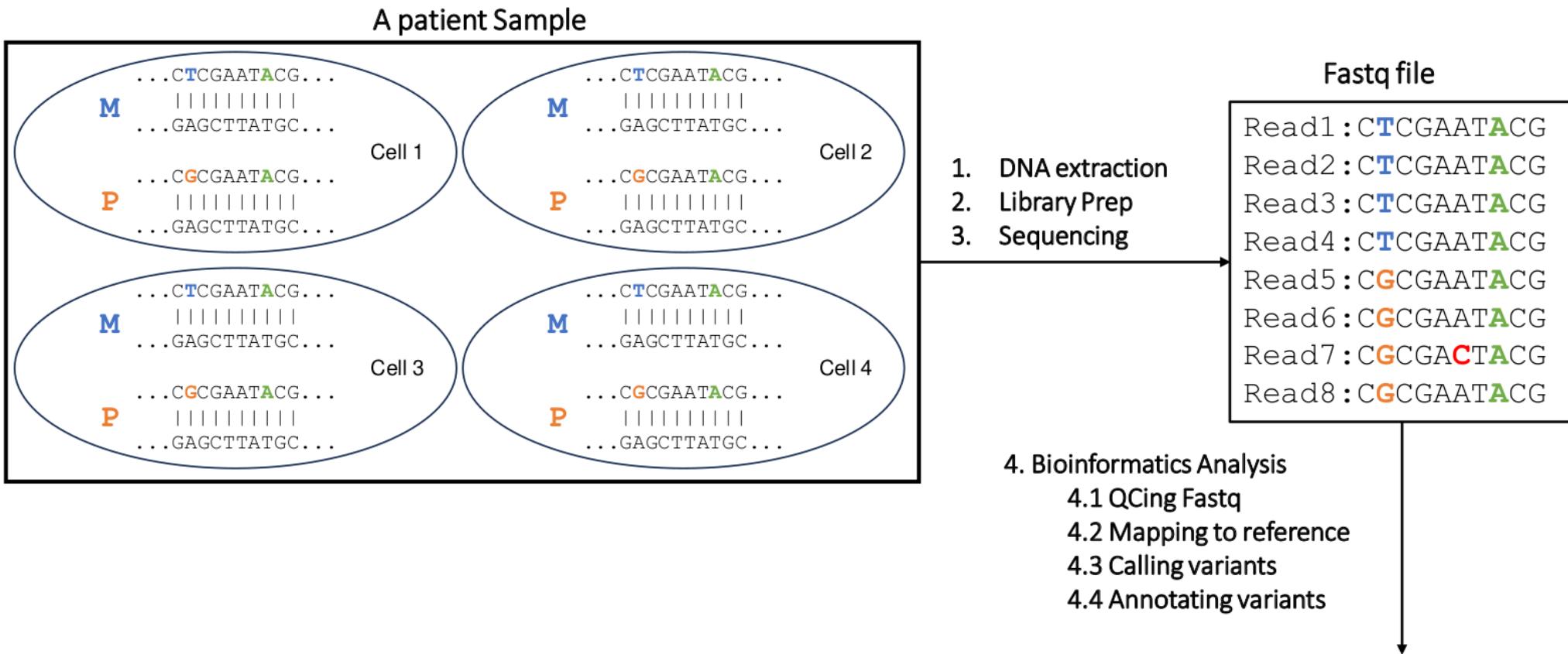
# Cancer Genome Somatic Variation



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)

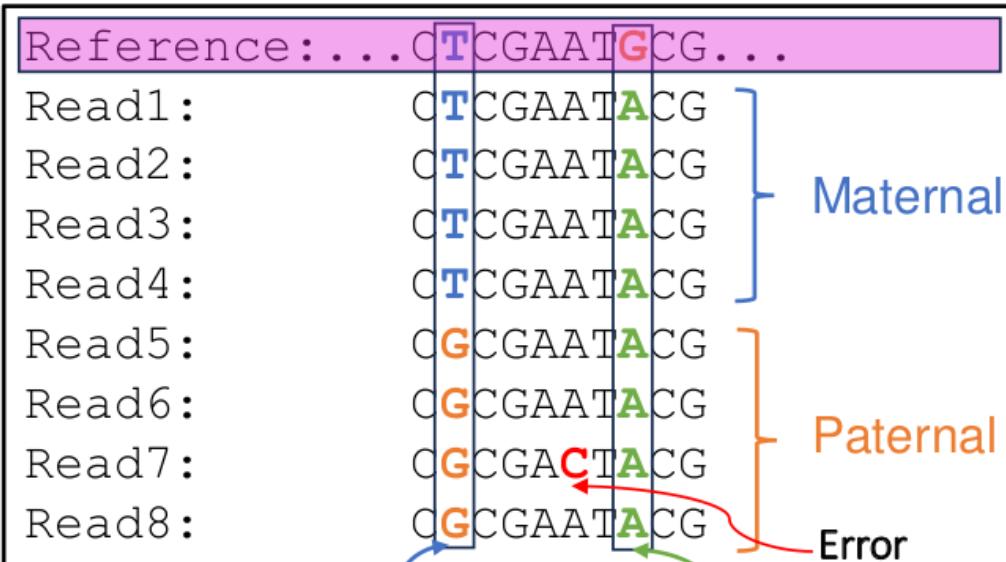
# Quy trình XÉT NGHIỆM gen bằng phương pháp giải trình tự thế hệ mới (NGS)

# 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: <b>4,4</b>
20	17330	.	G	A	150	PASS	GT:GQ:DP:AD	1/1:49:8: <b>8,8</b>

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

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

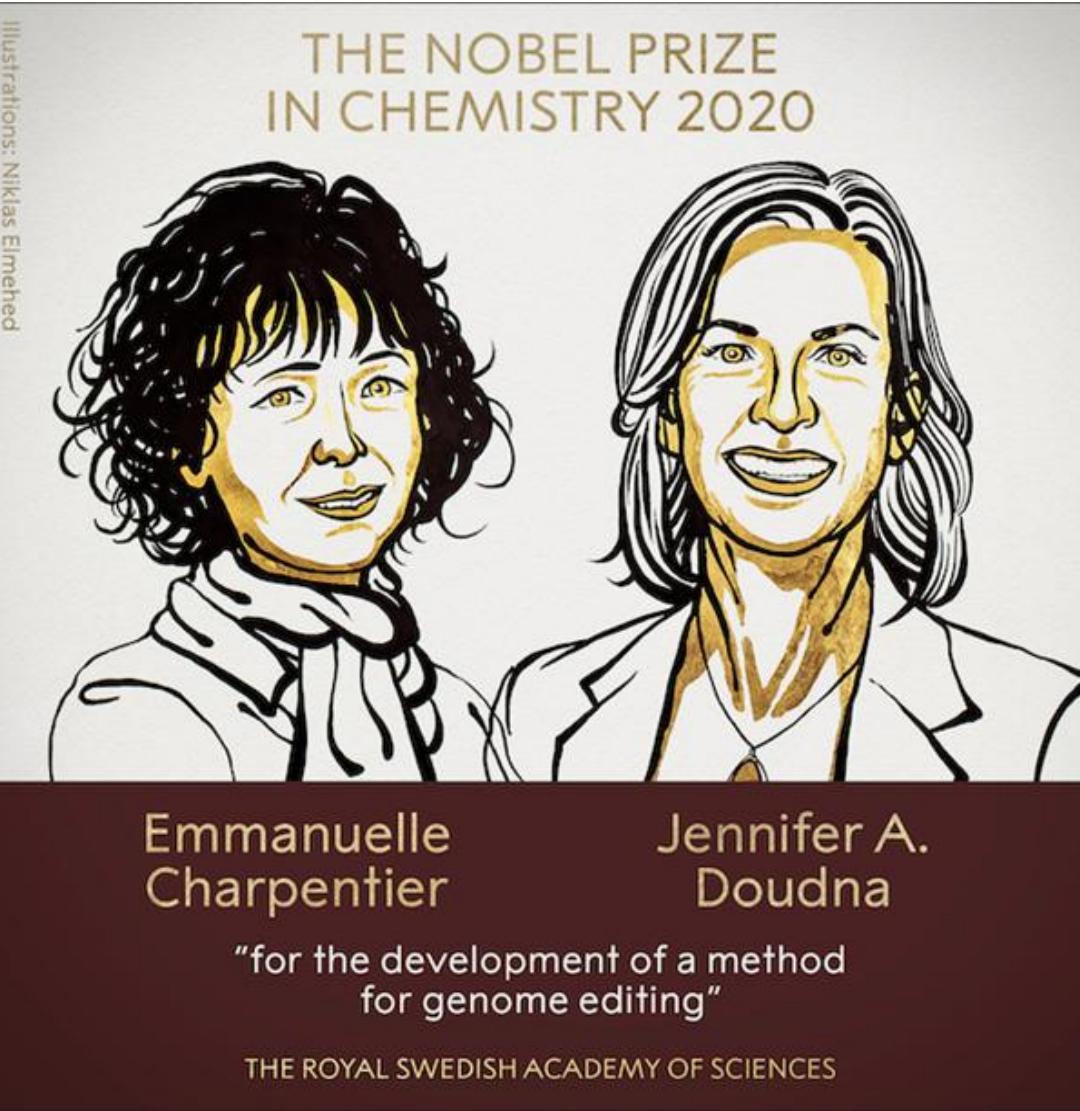
## 4.4 Annotating variants

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



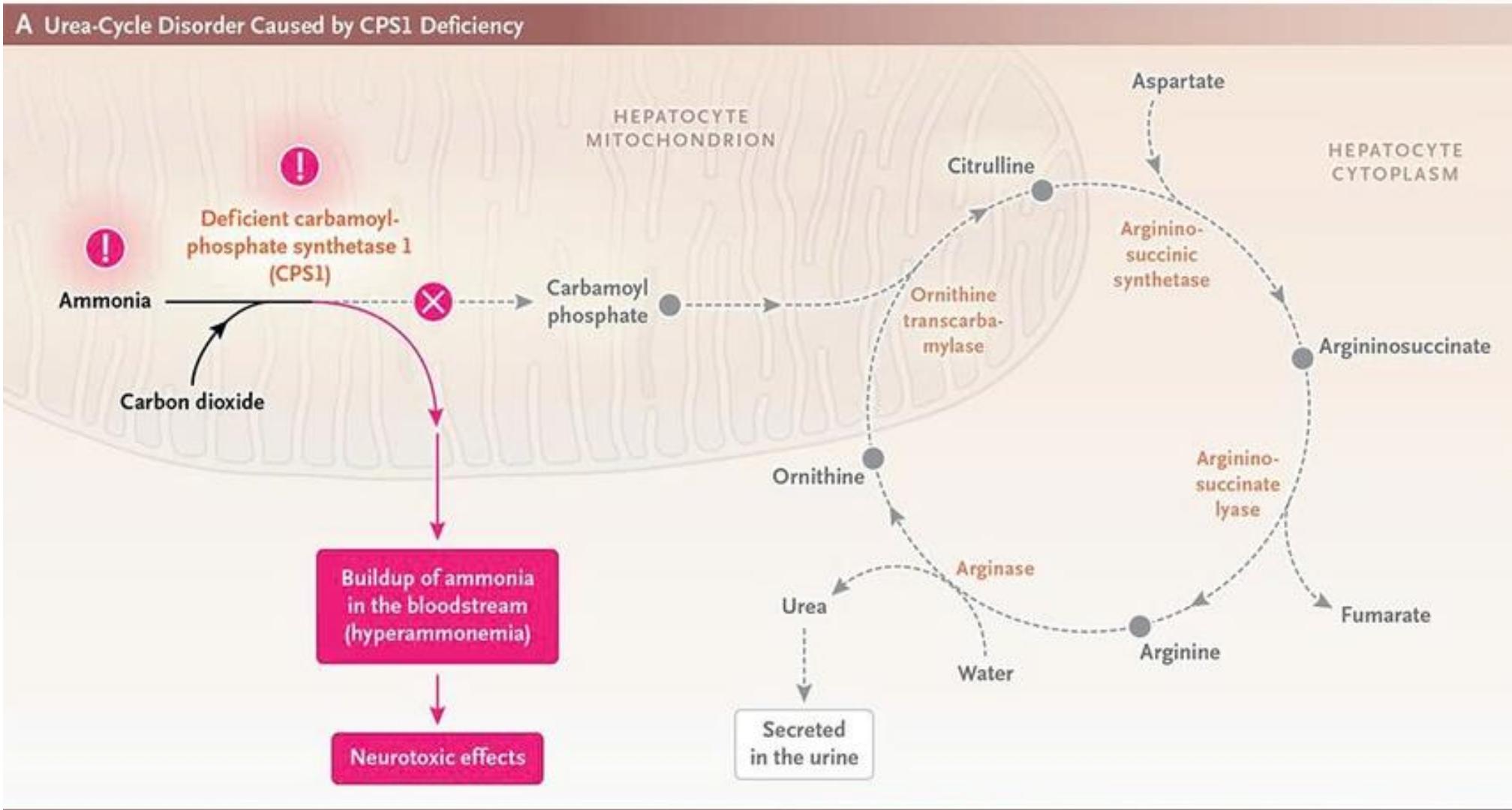
<https://ggb.a.swiss/en/the-first-medical-genomics-center-opens-in-geneva/>

LIỆU CÓ GIÚP  
ÍCH?

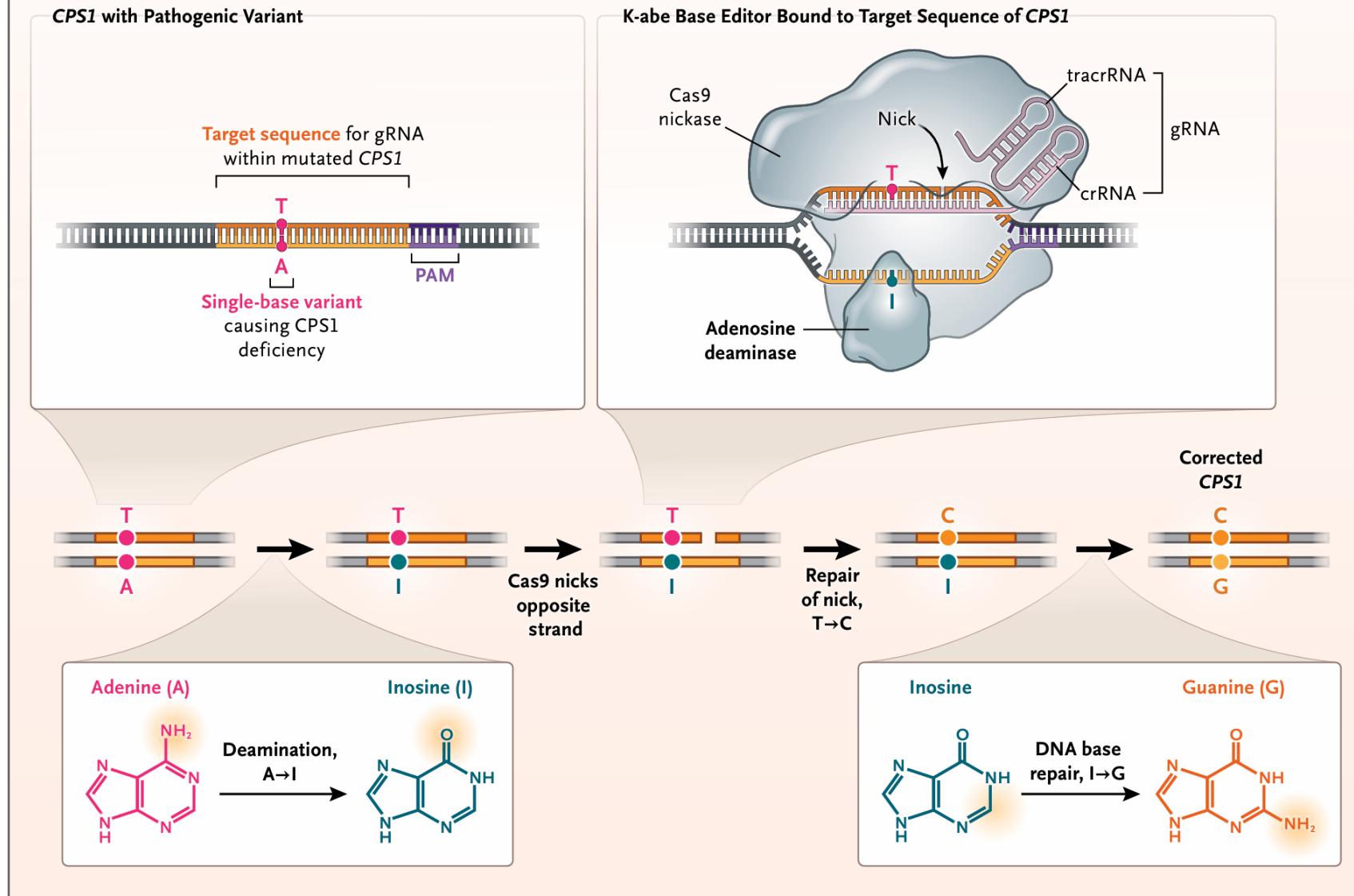


# CRISPR-CAS9

# New England Journal of Medicine (NEJM), May 2025



### B Adenine Base Editor-Mediated Repair of *CPS1* Variant



# Gene-editing therapy made in just 6 months helps baby with life-threatening disease

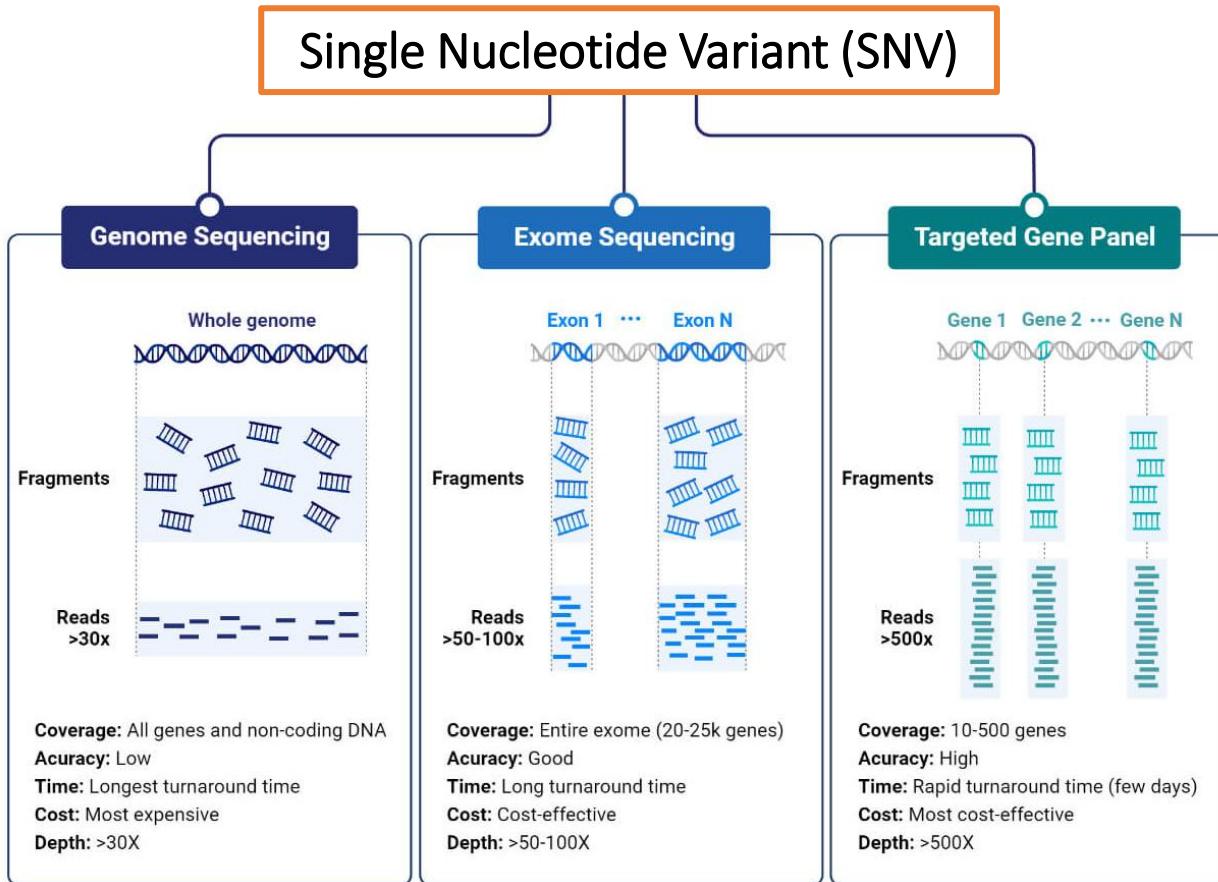
Custom CRISPR paves the way for treating genetic disorders in tailor-made ways

13 MAY 2023 · 1:00 PM ET · BY JOCELYN KAISER

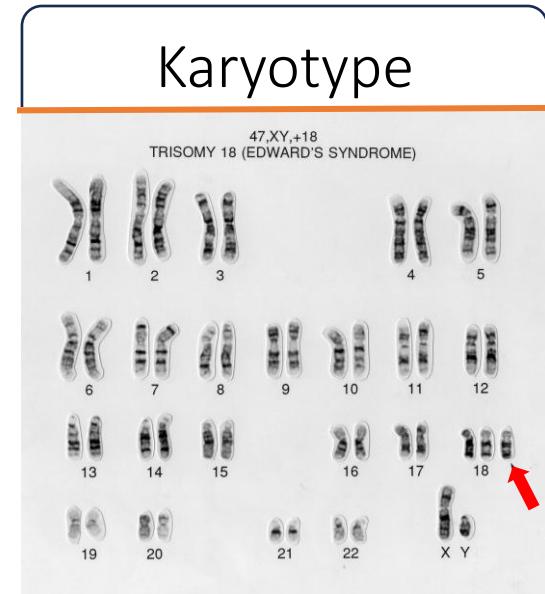
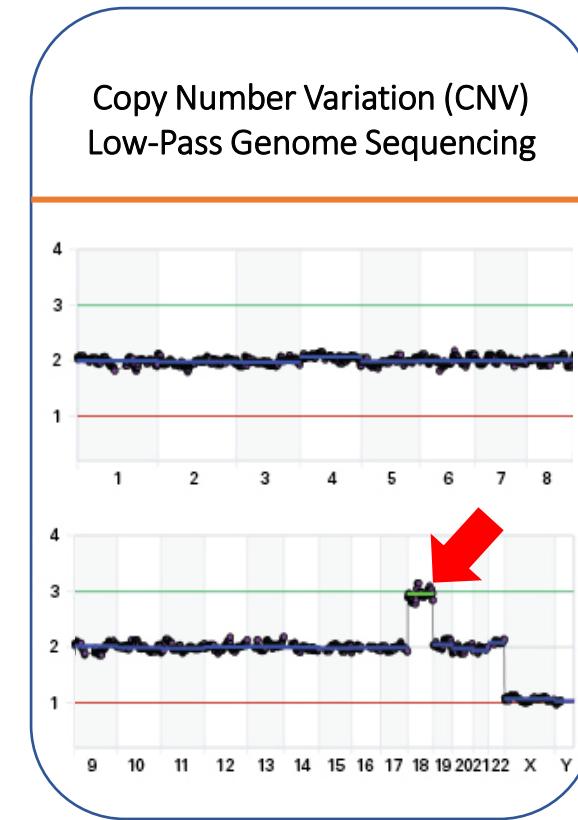


An infant who received a customized gene-editing treatment, shown here with researchers Kiran Musunuru (left) and Rebecca Ahrens-Nicklas, now needs less medicine to defuse a blood buildup of ammonia. From "Gene-editing therapy made in just 6 months helps baby with life-threatening disease" by Jocelyn Kaiser, 2023. <https://www.science.org/content/article/gene-editing-therapy-made-just-6-months-helps-baby-life-threatening-disease>

# 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
```

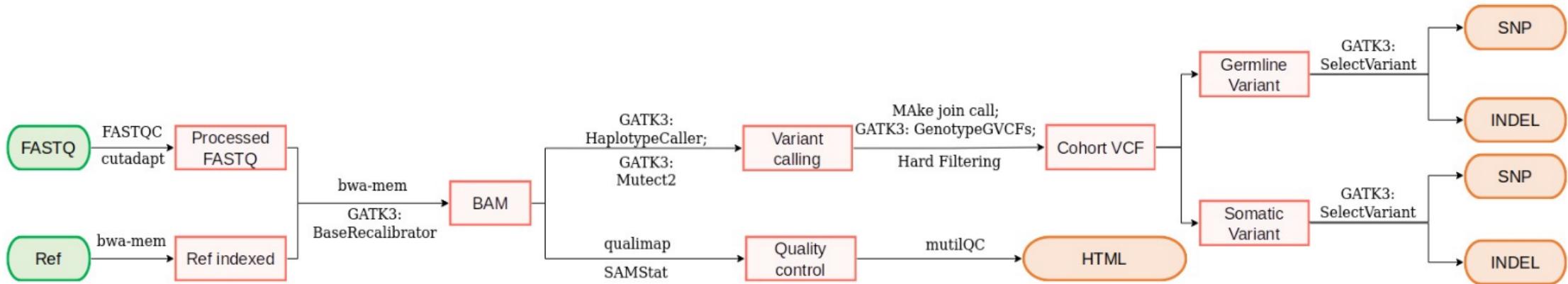
## WGS\_Level\_2

```
|── 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

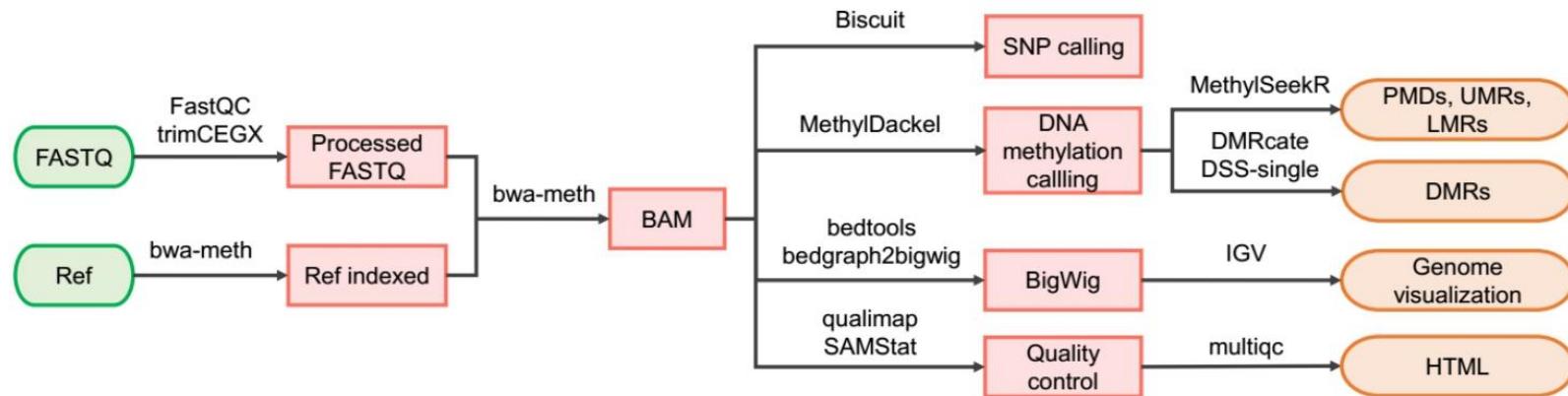
```
|── FR07885321_1 (sample)
|   └── FR07885321_1.bam
|   └── FR07885321_1.bam.md5
|   └── FR07885321_1.GenotypeGVCFs.vcf.gz
```

# Whole Genome Sequencing pipeline



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

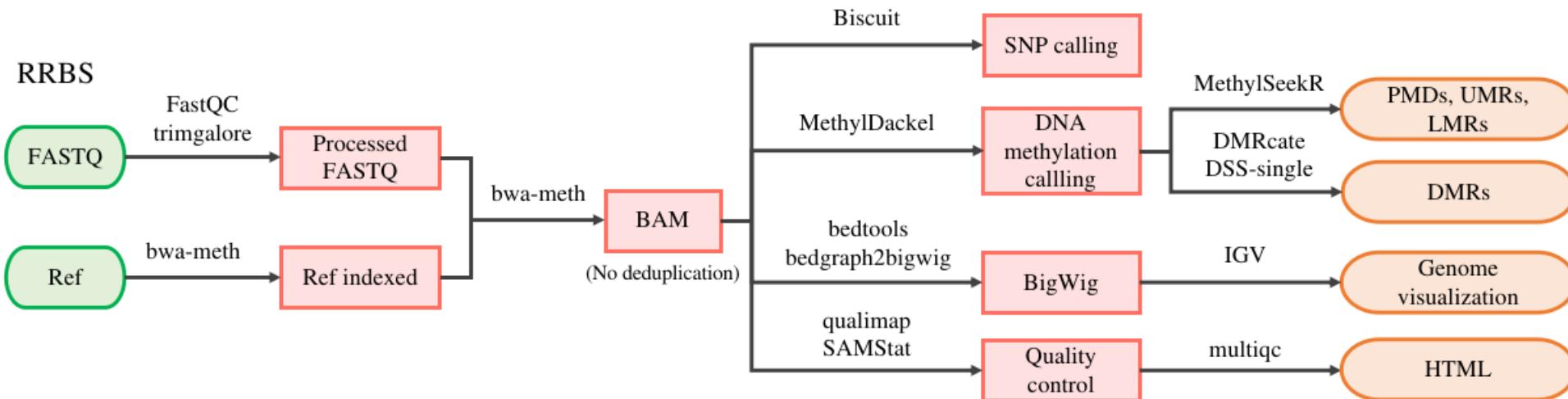
# Whole Genome Bisulphite Sequencing pipeline



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

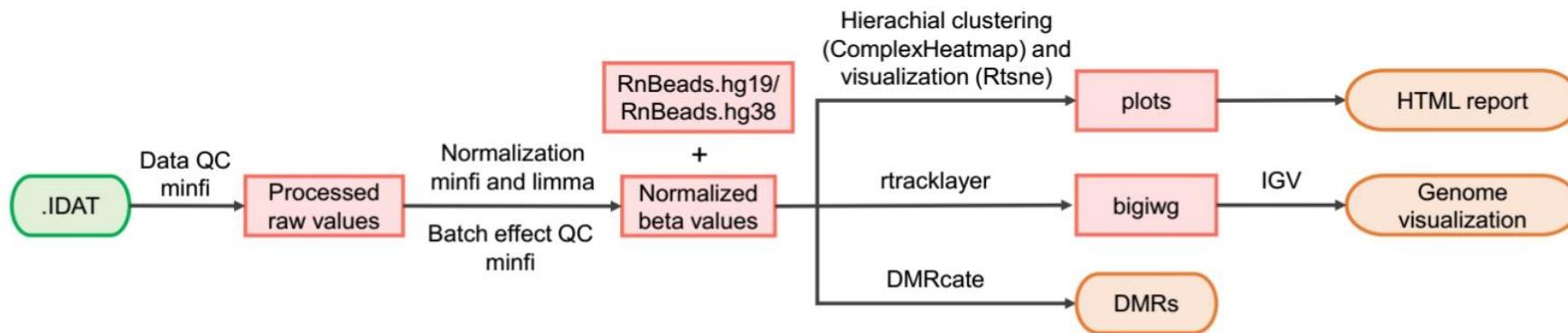
Figure 4. The WGBS pipeline and software.

# Reduced-representation bisulfite sequencing (RRBS-Seq) pipeline



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

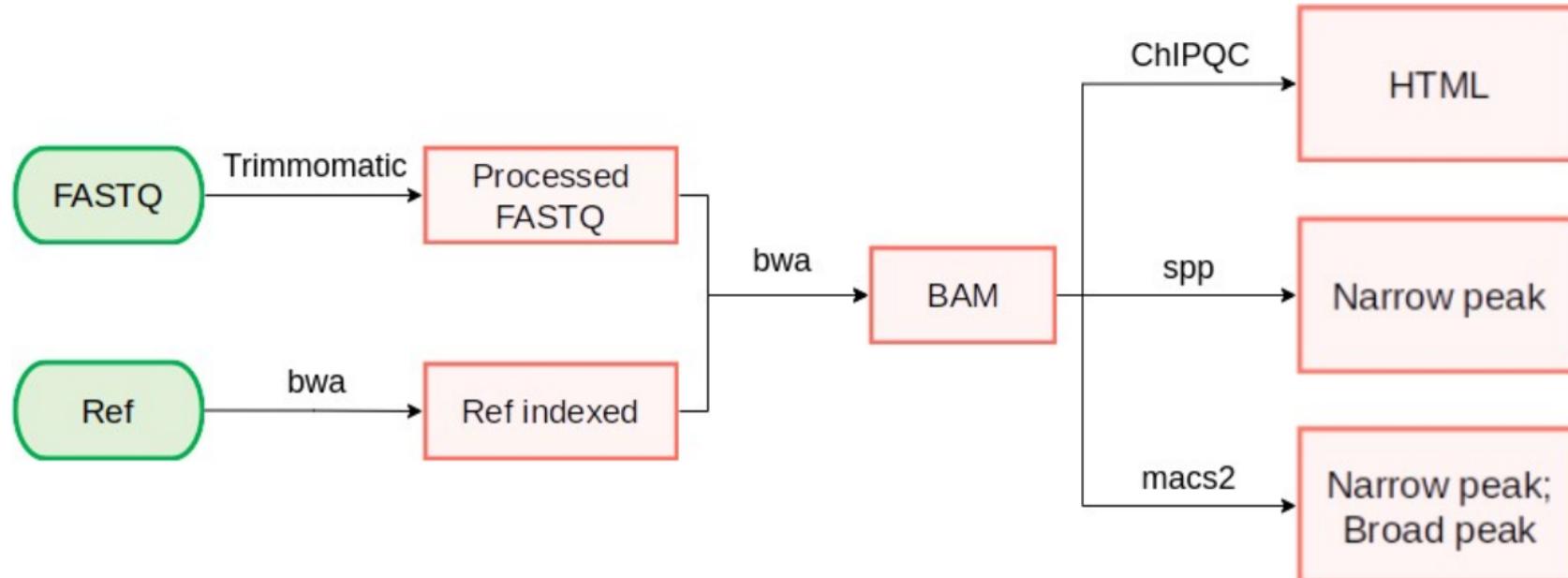
# DNA Methylation Array (450K and EPIC/850K) pipeline



Package	Version	Link
<b>minfi</b>	1.34.0	<a href="http://bioconductor.org/packages/release/bioc/html/minfi.html">http://bioconductor.org/packages/release/bioc/html/minfi.html</a>
<b>limma</b>	3.44.3	<a href="https://bioconductor.org/packages/release/bioc/html/limma.html">https://bioconductor.org/packages/release/bioc/html/limma.html</a>
<b>RnBeads.hg19</b>	1.20.0	<a href="https://bioconductor.org/packages/release/data/experiment/html/RnBeads.hg19.html">https://bioconductor.org/packages/release/data/experiment/html/RnBeads.hg19.html</a>
<b>RnBeads.hg38</b>	1.20.0	<a href="https://bioconductor.org/packages/release/data/experiment/html/RnBeads.hg38.html">https://bioconductor.org/packages/release/data/experiment/html/RnBeads.hg38.html</a>
<b>Complex-Heatmap</b>	2.4.3	<a href="https://www.bioconductor.org/packages/release/bioc/html/ComplexHeatmap.html">https://www.bioconductor.org/packages/release/bioc/html/ComplexHeatmap.html</a>
<b>Rtsne</b>	0.15	<a href="https://cran.r-project.org/web/packages/Rtsne/index.html">https://cran.r-project.org/web/packages/Rtsne/index.html</a>
<b>rtracklayer</b>	1.48.0	<a href="https://bioconductor.org/packages/release/bioc/html/rtracklayer.html">https://bioconductor.org/packages/release/bioc/html/rtracklayer.html</a>
<b>DMRcate</b>	2.2.3	<a href="https://bioconductor.org/packages/release/bioc/html/DMRcate.html">https://bioconductor.org/packages/release/bioc/html/DMRcate.html</a>

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

# ChIP-seq pipeline



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

# 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

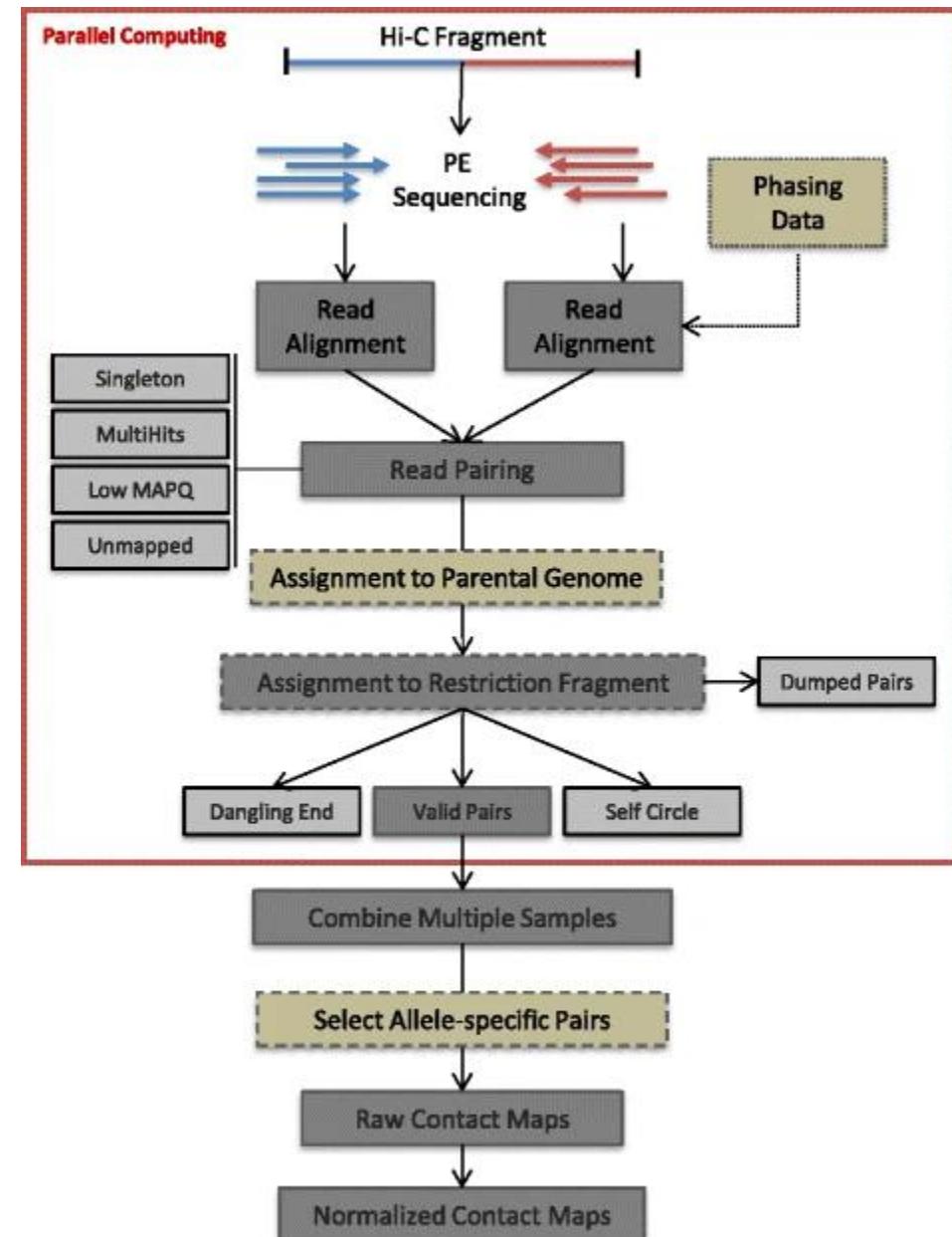
*Genome Biology* 16, Article number: 259 (2015) | [Cite this article](#)

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

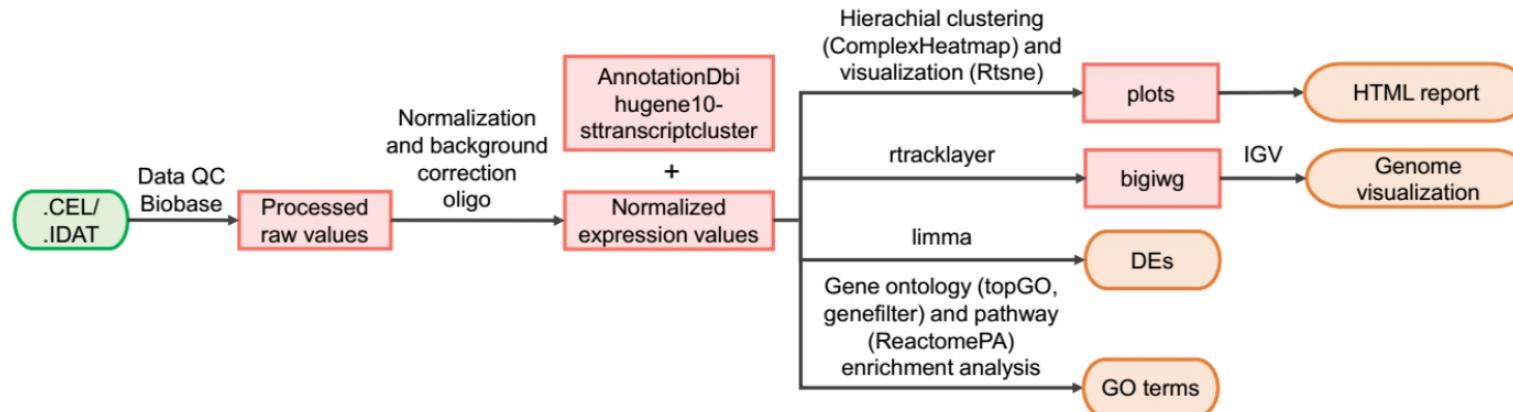
### Abstract

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>.



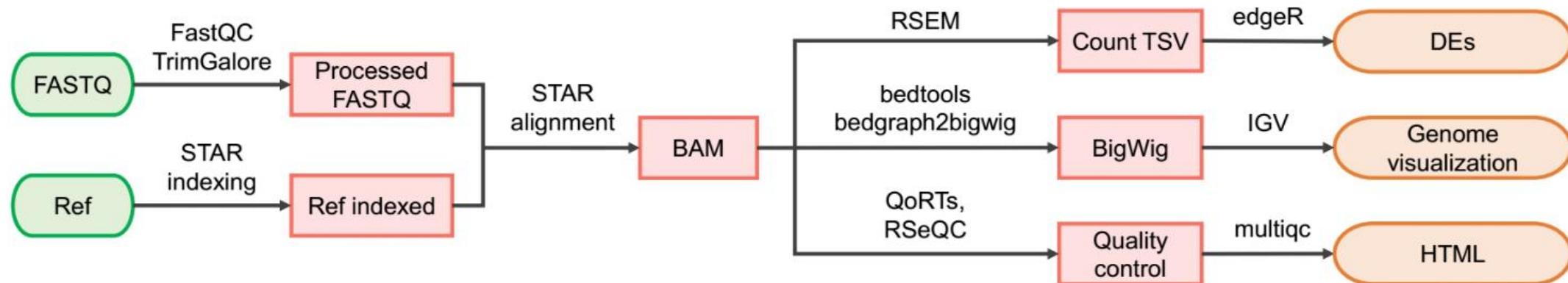
# Gene Expression Microarray (Affymetrix) pipeline



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

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

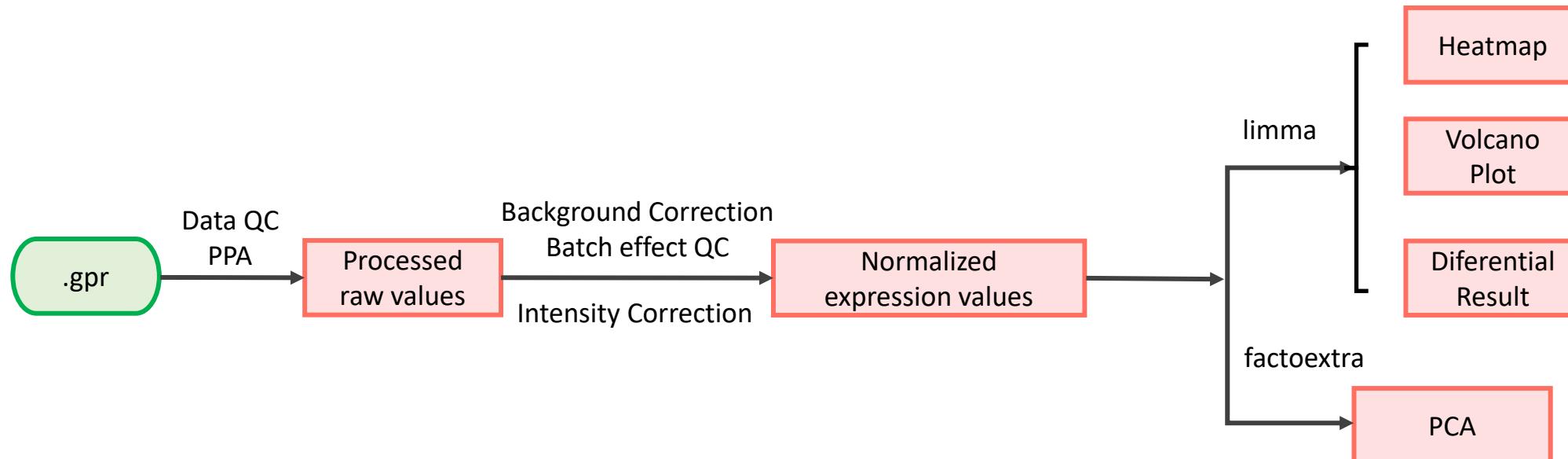
# Bulk RNA-seq pipeline



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

Figure 6. The RNA-seq pipeline and software.

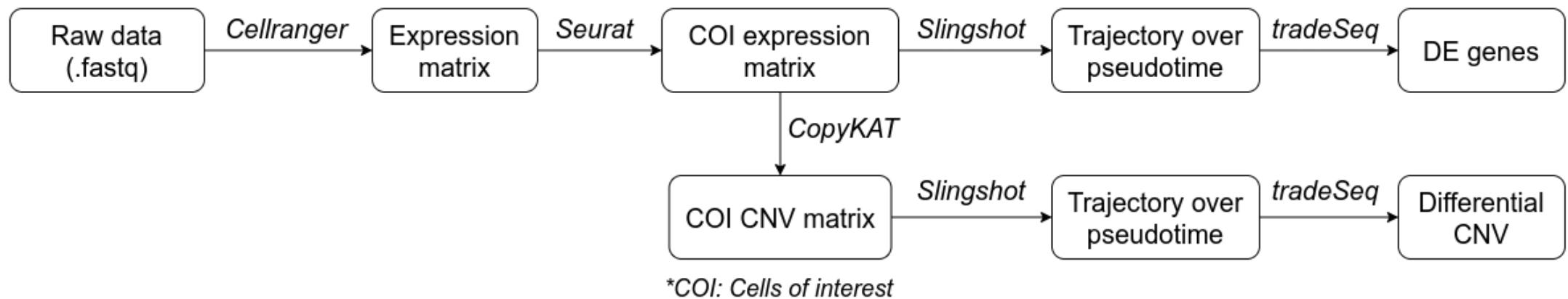
# Protein Microarray Expression (ProtoArray) pipeline



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

# scRNA-seq

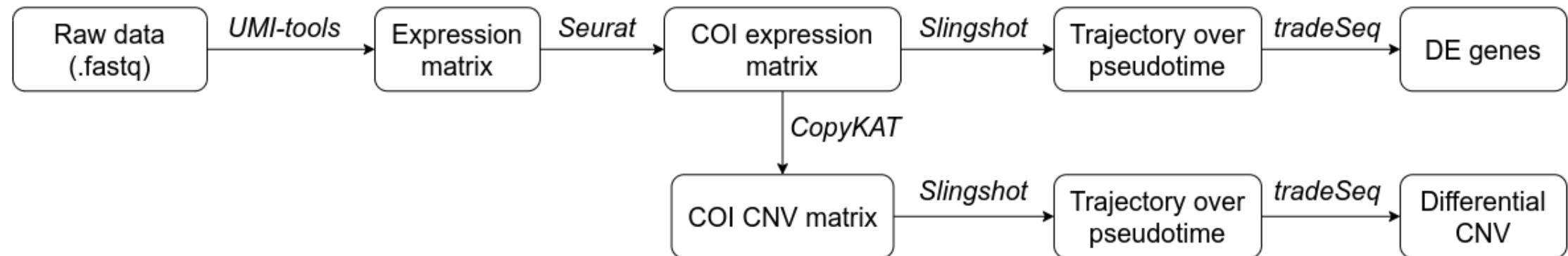
## 10X Chromium pipeline



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

# scRNA-seq

## SMART-Seq2 pipeline



\*COI: Cells of interest

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

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

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

## Whole Exome Sequencing - WES

### IGV - DEE001 – CDKL5



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

## Whole Exome Sequencing - WES

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

DEE001 – CDKL5 (Xp22.13)

Current Build 156  
Released September 21, 2022

**rs1057519541**

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

**Clinical Significance**

Variant Details HGVS Submissions History Publications Flanks

Allele: T (allele ID: 362353 )

ClinVar Accession ▲ Disease Names ▷ Clinical Significance

RCV000416943.1 Focal epilepsy Pathogenic

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

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

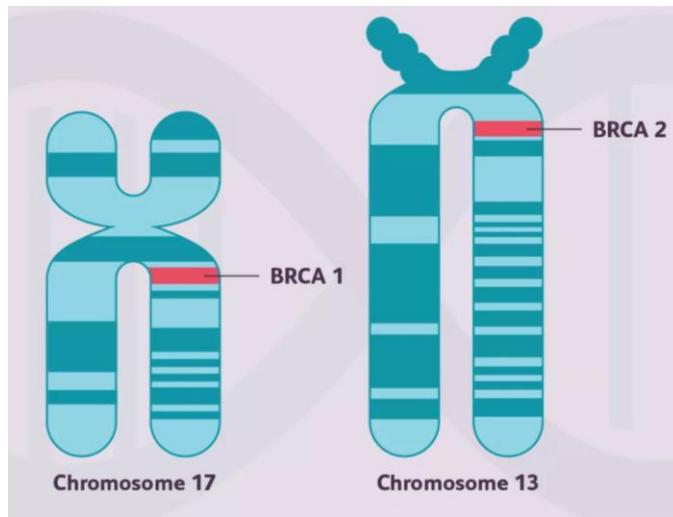
## Ví dụ 2: Sàng lọc bệnh ung thư bằng xét nghiệm gen Gene Panel – 53 genes của MGI

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

Nguyen Le Duc Minh, MD

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

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



## KẾT QUẢ XÉT NGHIỆM BRCA1/2

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

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

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

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

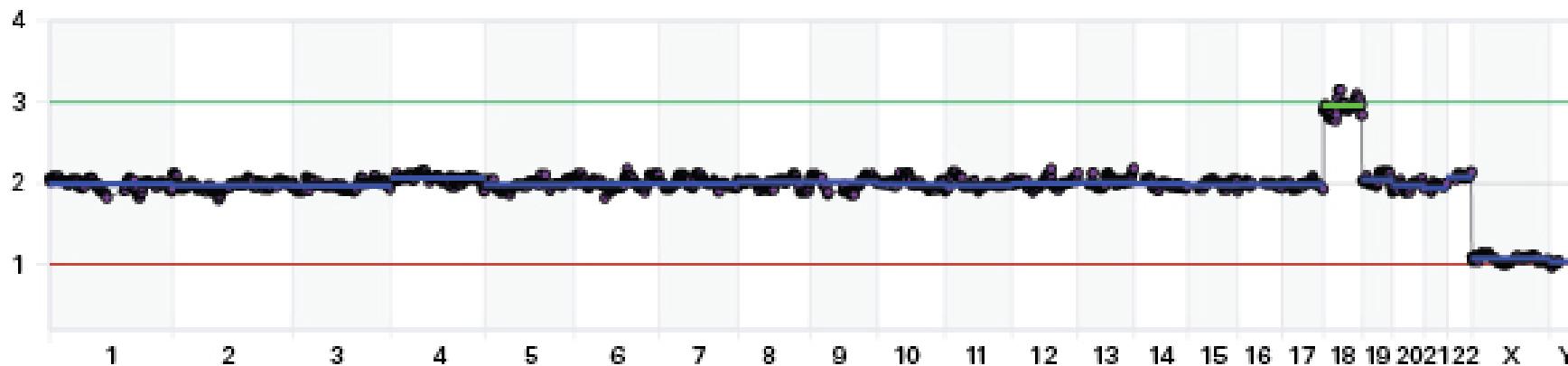
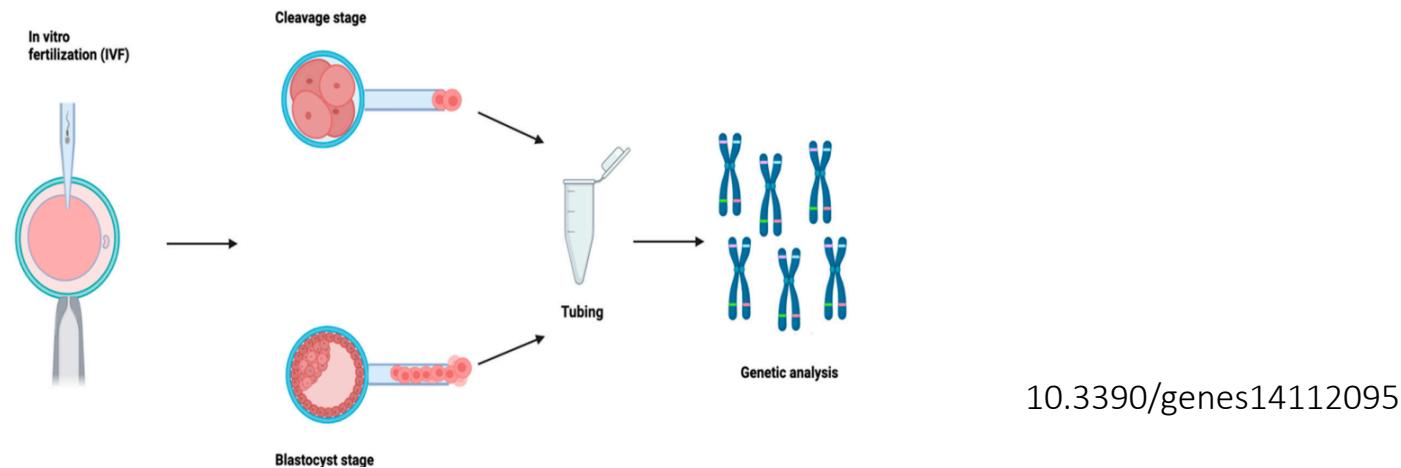
- Hệ thống xét nghiệm Illumina MiSeqDx (CE/US-IVD)
- Bộ xét nghiệm: NGeneBio BRCAaceuTest™Plus (CE-IVD)
- Phần mềm hỗ trợ phân tích kết quả: NGeneBio NGeneAnalySys™ (CE-IVD)

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

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

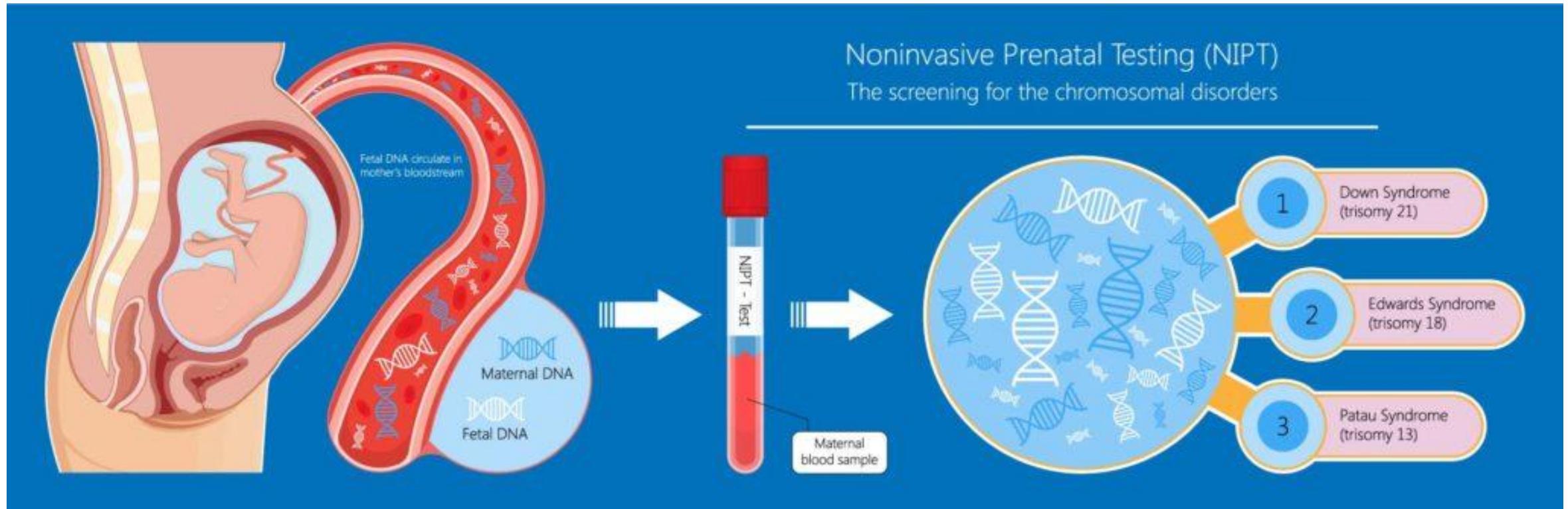
# Ví dụ 4: Sàng lọc phôi trong hỗ trợ sinh sản IVF

Xét nghiệm tiền làm tổ PGT-A



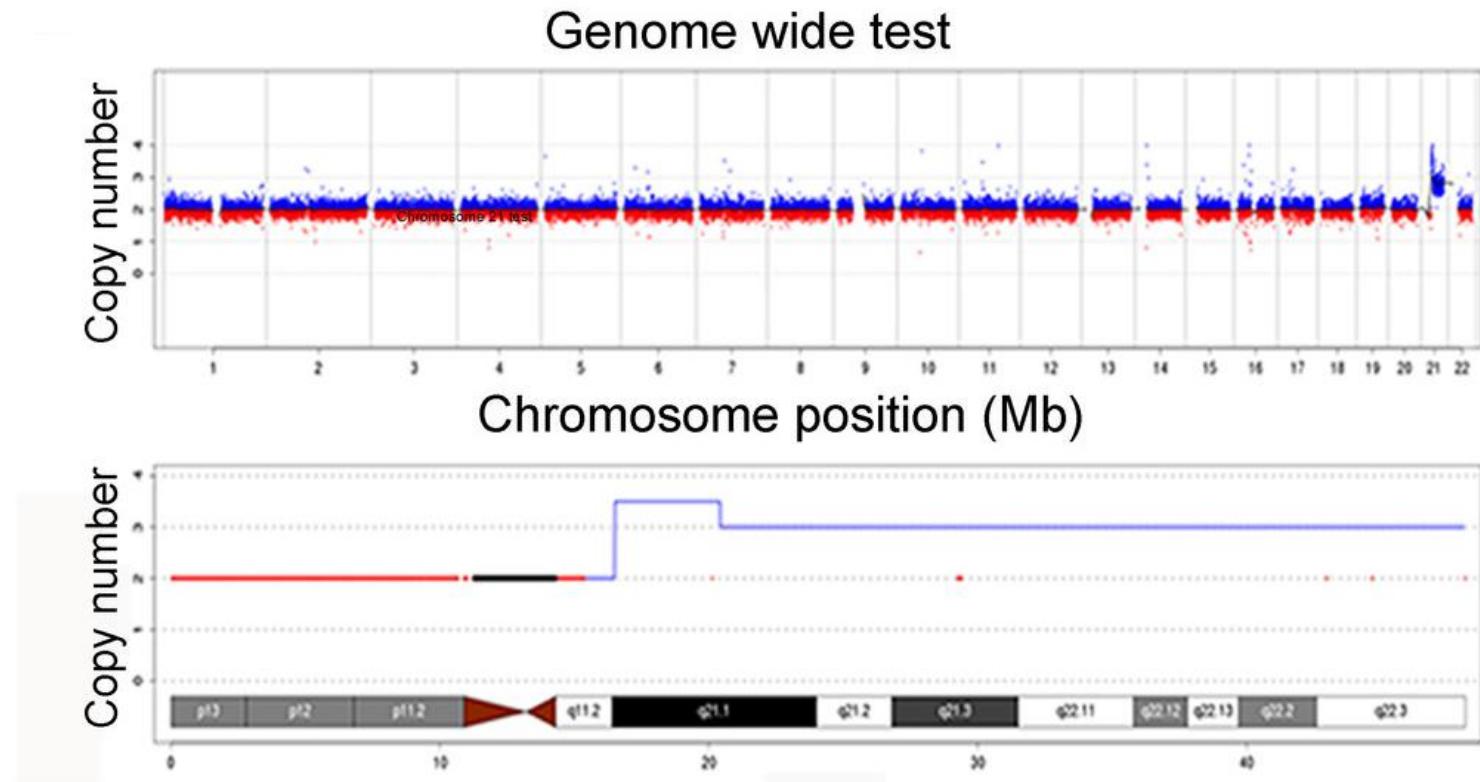
Mẫu phôi có 3 nhiễm sắc thể 18 (trisomy) trong DNA hệ gen

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



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

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



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

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

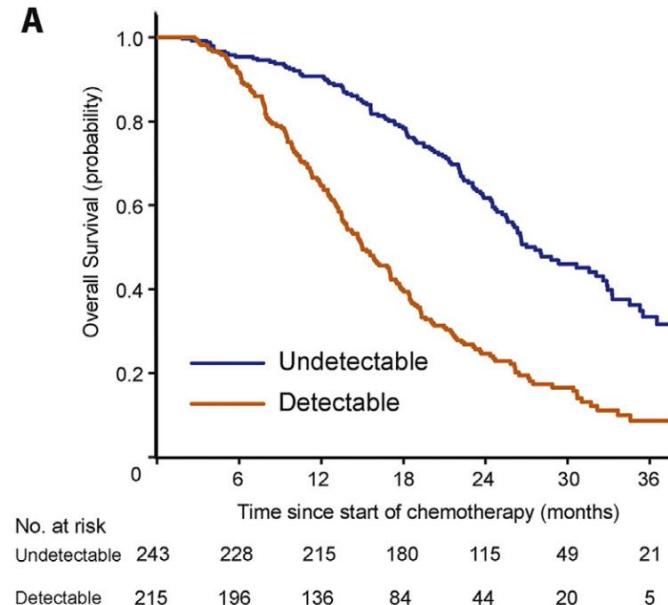
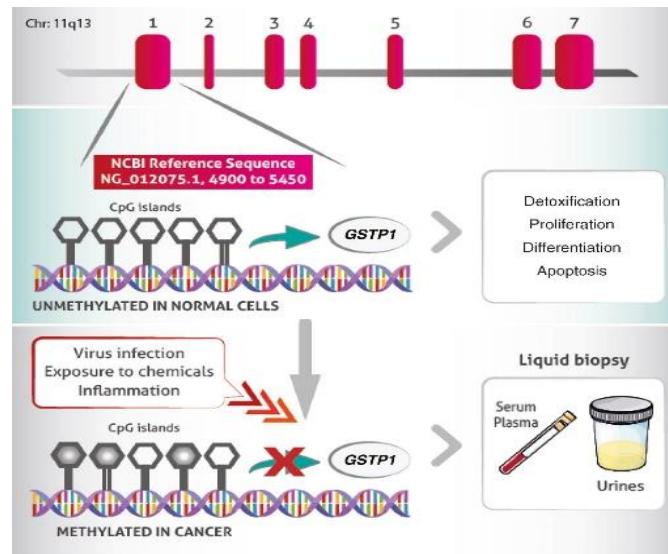
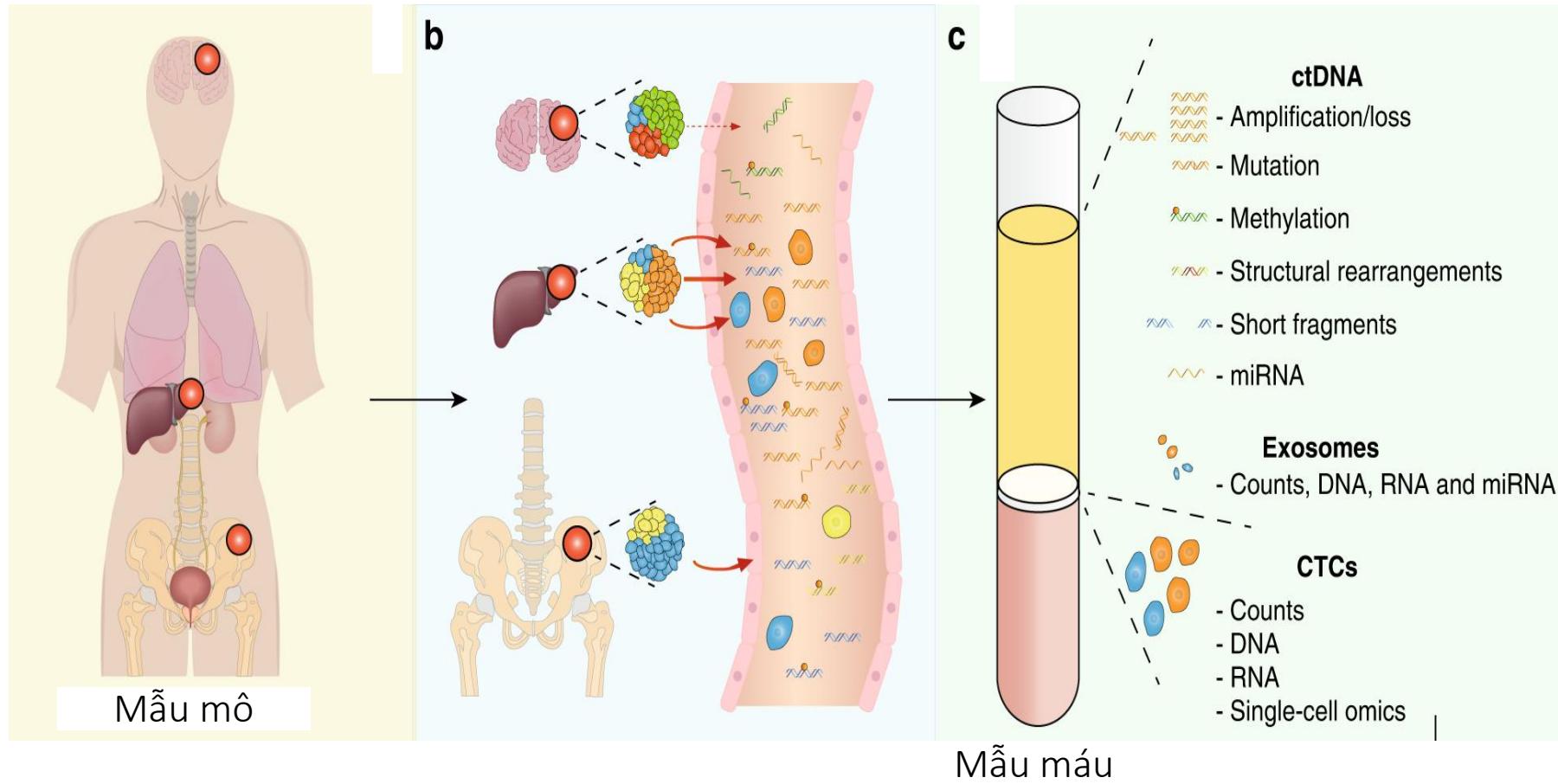


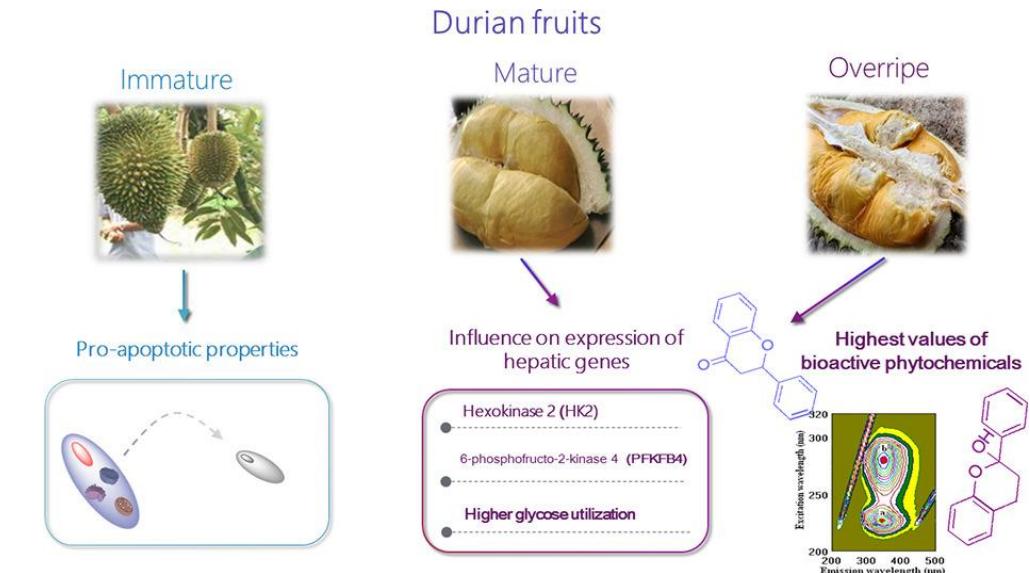
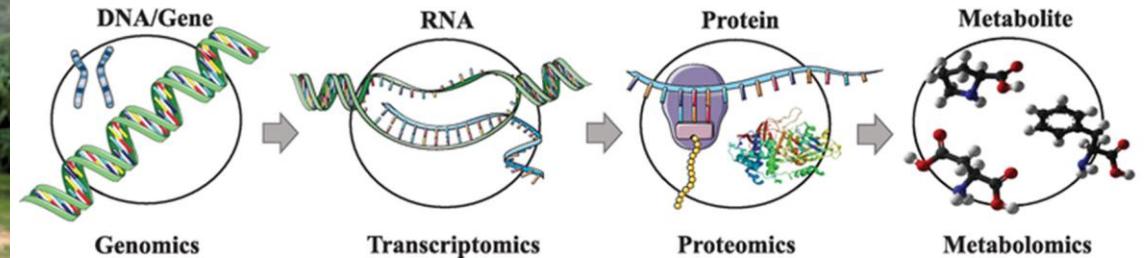
Fig. Kaplan-Meier estimates of survival in patients with a detectable mGSTP1 at baseline. (A) OS according to mGSTP1 detection after two cycles of docetaxel.

# Mẫu cho XÉT NGHIỆM gen: mẫu mô và máu



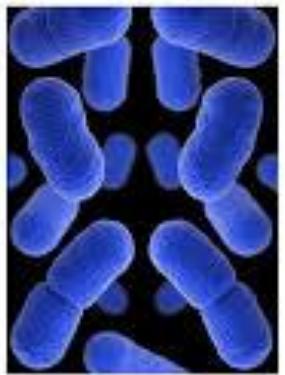
10.1038/s41416-022-01881-9

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



## Listeria monocytogenes

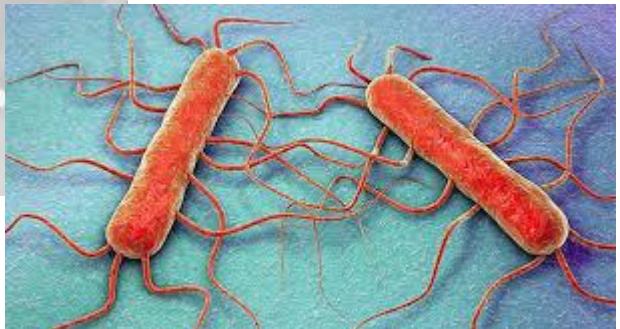
microbeonline



Beta-hemolytic colonies



Gram positive coccobacilli



## Listeria monocytogenes EGD-e, complete genome

GenBank: AL591824.1

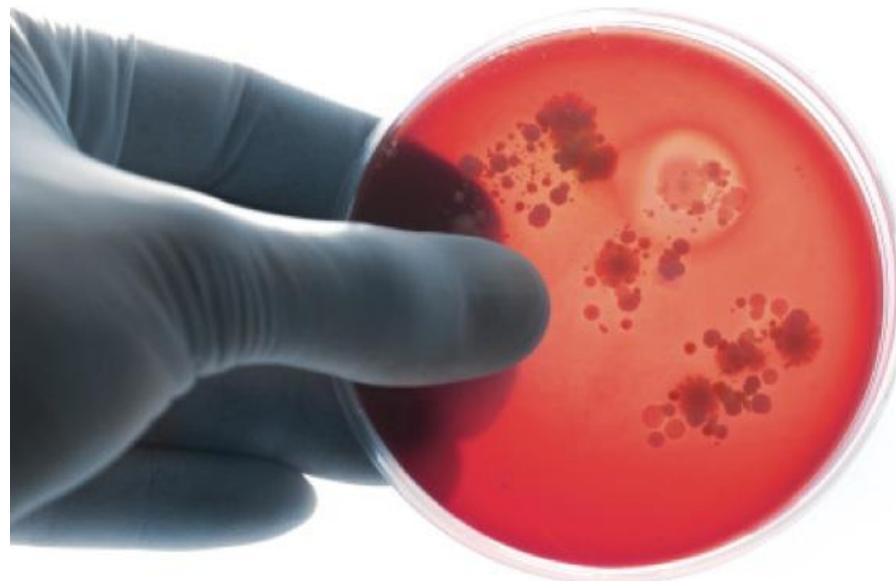
[GenBank](#) [Graphics](#)

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

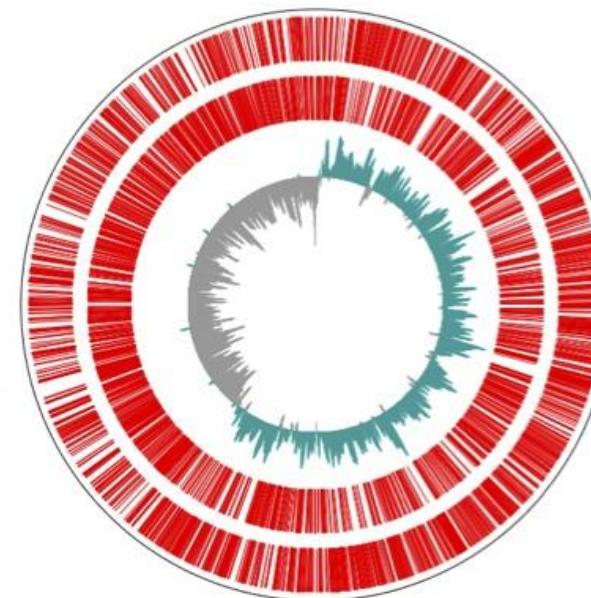
```
TAATTAATCTAACAAATTCTGTTACAGATTCTTACACACAAGTTACACACAAGTTAACGGCTGTGGA  
CAACCGTTTTCACATCTGGACAGTTTGTTGGATAGATTGGTAAGTCCTGCTATCAGAGTGTATTCT  
GATATTATAATTCCCTGCGAATAGAAAATAGCTGGGAAAACTAAAGTTATCCACAATACATTTTACT  
TTGTGGATAATTTTAACAGTGTGGATAACCTTATCCATAGCTTTCTATCTGTGGATAACTTTAT  
AGCATCCATTACATTACATAAAAAGGGGGGTACTAGTGCAATCAATTGAAGACATCTGGCAGGAAACA  
CTGCAAATTGTTAAAAAAATATGAGTAAACCTAGTTACGATACATGGATGAAATCAACAAACCGCTCATT  
CACTTGAAGGTAACACGTTTATTATTCAGCGCCAAATAATTGTTGCGATTGGTAGAGAAAAGCTA  
CACTCAATTATCGCTAACATTGCAAGAAAATACTGGTCGTTATTGATGTCGCTTATTGATGGC  
GAGCAGGAAGAAAACTTGAATACACTGTGATTAAACCAATCCAGCATTAGATGAAGATGGCATTGAAA  
TTGGAAAACATATGCTTAACTCACGTTATGTTTGATACTTGTCAATTGGTCAGGGAACAGATTGC  
CCACGCAGCATCACTGCACTAGCCGAAGCACCAGCGAAAGCATAATCCACTCTCATTATGGAGGA  
GTTGGCCTCGGTTAAACACATTTAATGCAAGCAGTTGCCACTATGTTCAACAAACATAAAAGATAATGCGA  
AAGTAATGTACCTTCCAGCGAAAATTCCAATGAGTTATTAGCTTATTCGTGATAATAAAACCGA  
AGAATTGCGACAAAATATCGGAATGTTGATGCTTACTTATTGATGATATTCAATTAGCCGGTAAA  
GAAGGAACACAAGAGGAATTTCACATTTAACACACTTTATGATGAACAAAAGCAAATTATT  
CCAGTGACCGGACCAACAAAAGAAATTCCACTGGAAGATCGACTGAGATCCGCTTGAATGGGCTT  
AATTACTGATATTAGCCACCAAGACTTAGAAACCCGGATGCCATTACGTAAAAAAGCAAAGCAGAC  
GGATTAGATATTCCAATGAAGTTATGCTTATATCGCAAACCAAATTGATTCGAATATTGCGAGCTAG  
AAGGGCTCTATCGAGTAGTTGCTTATTCTCCCTGTTAATAAGATATAACAGCTGGCTTGAGC  
AGAAGCACTAAAAGATATTATCCCCTTTCAAATCACAAGTTACTTATTAGTGGTATTCAAGAAGCA  
GTCGGTGAATATTCCACGTTGTTAGAAGATTAAAGCAAAAAACGGACGAAAAGTATAGCATTCC  
CGCGCCAAATCGCCATGTATCTCAAGAGAGCTTACAGATGCCCTATTACCAAAAAATCGGTGATGAATT  
TGGTGGTCGAGATCATACAACAGTTATTGATGCACTGAAAAAAATATCGCAACTACTAAAAACCGATCAA  
GTGTTGAAAATGACCTGCCGAAATTGAAAAAAATTAAAGAAAAGCACAAAATATGTTAATAGACCT  
GTGTACAATGTGGATAACTGAAACATACTTACCCACAAGTTACACATGTGGAAAACCTTATGCAGCAT  
GGCTTGTAAACCTACTTATCCACAAATCCACAGCGCTATTACTTACTACGATTTTTATTAATTAAATT
```

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

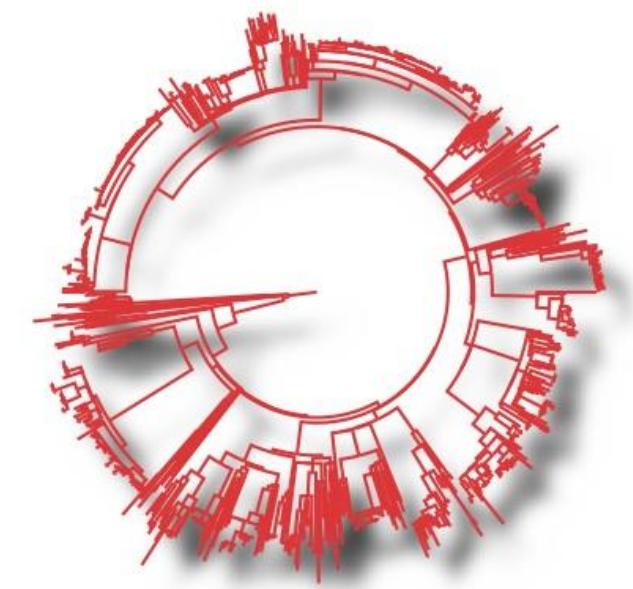
# Microbial Genome Analysis: study one isolate at once



Culture microbe



Genome Sequencing



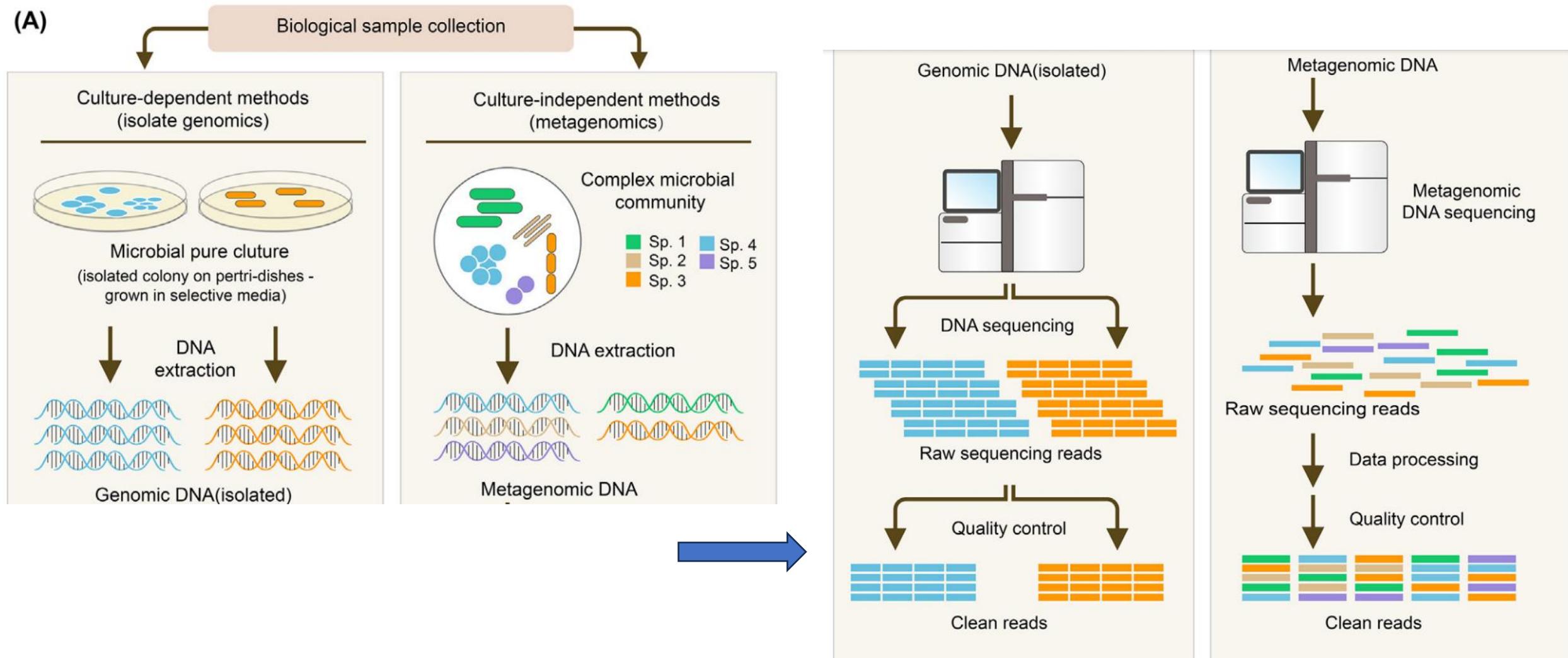
Serotyping/genotyping

Microbe = Microorganism

Prokaryotes = Bacteria & Archaea

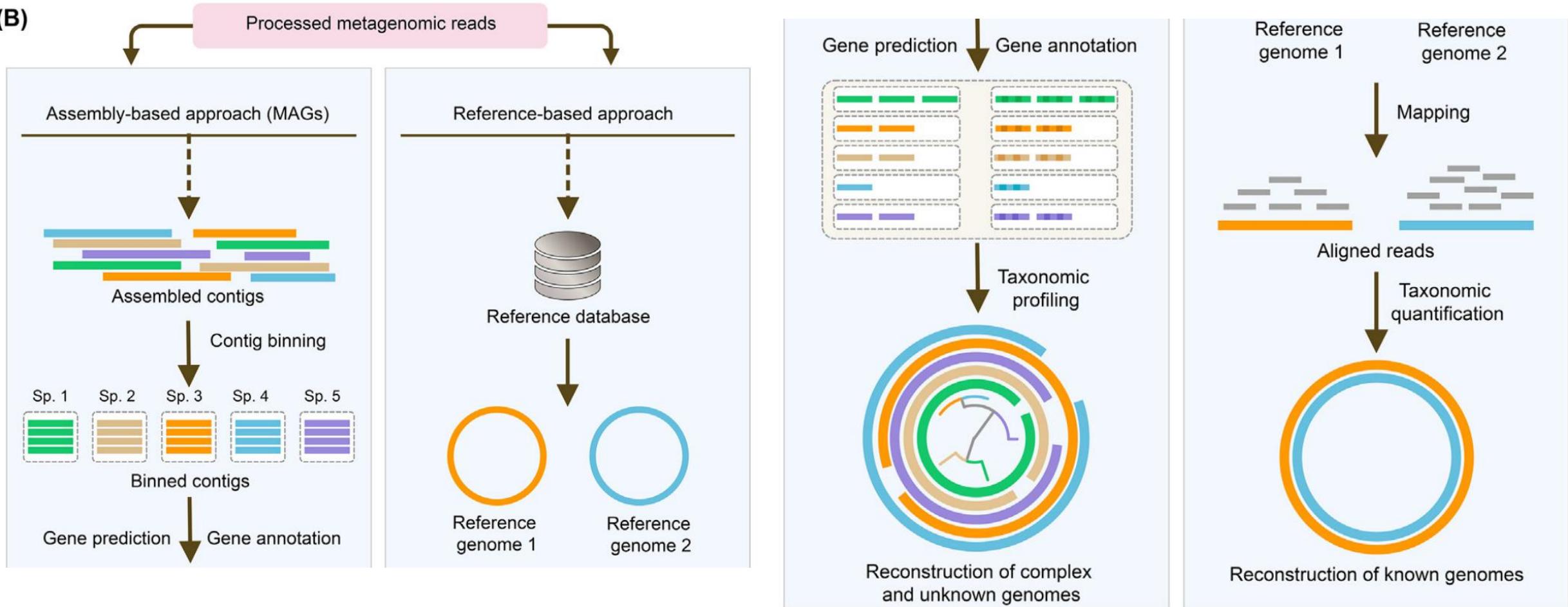
# Microbial Genome vs Metagenome Analysis

(A)

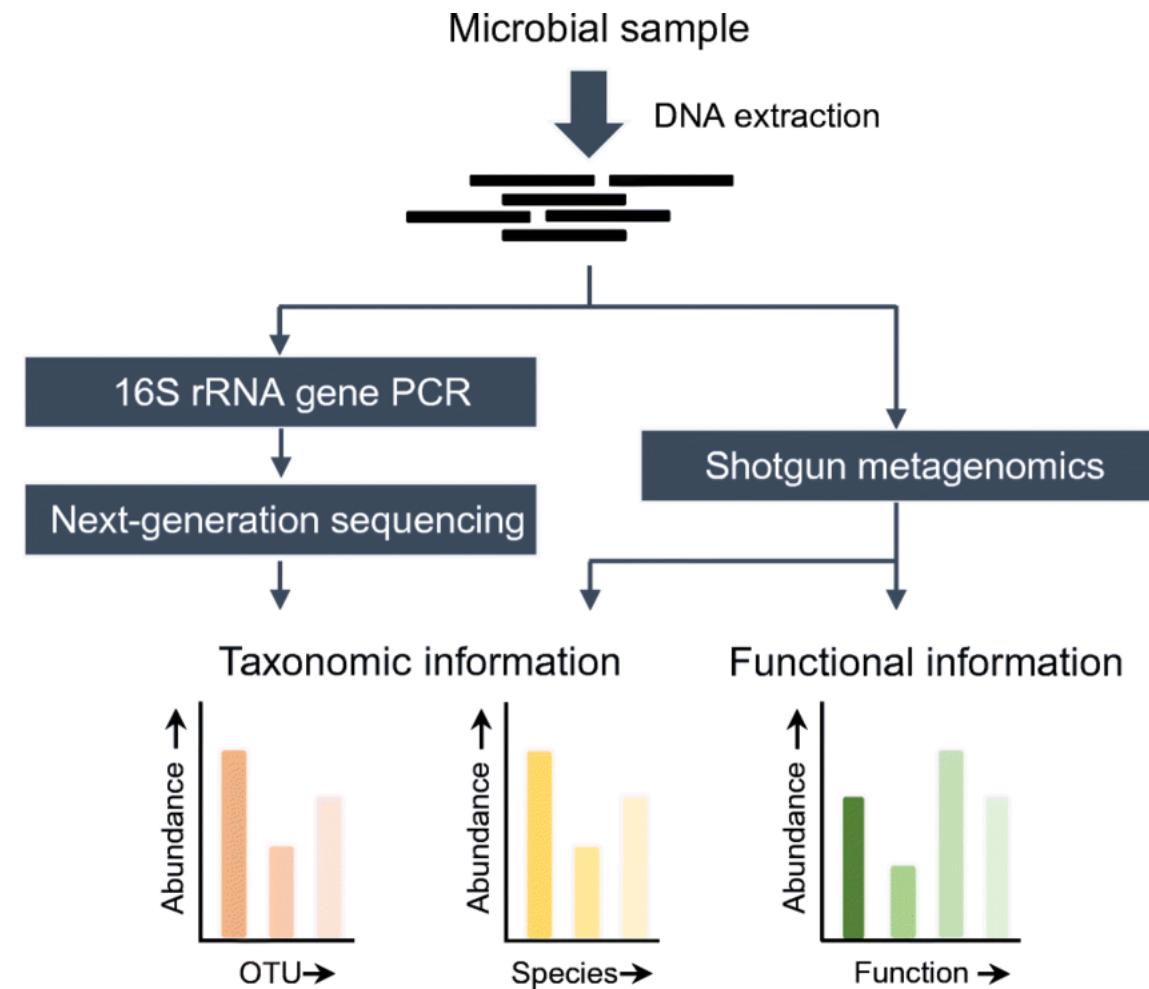
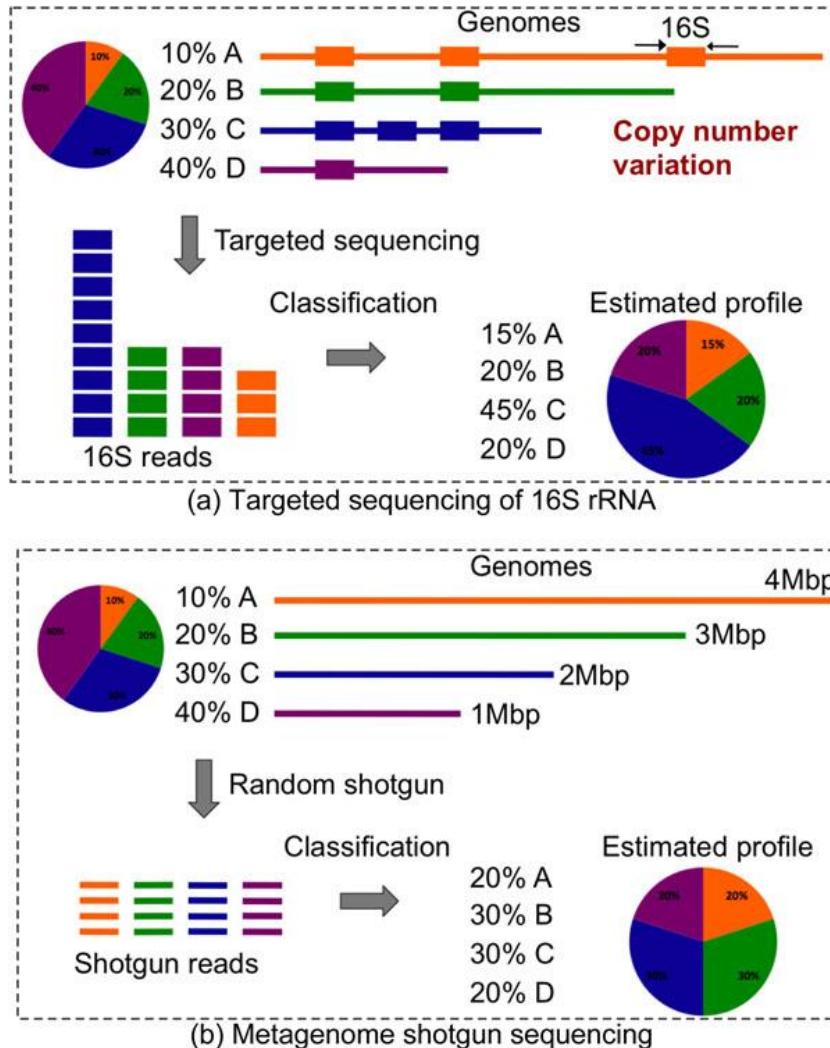


# Shotgun Metagenome Analysis

(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

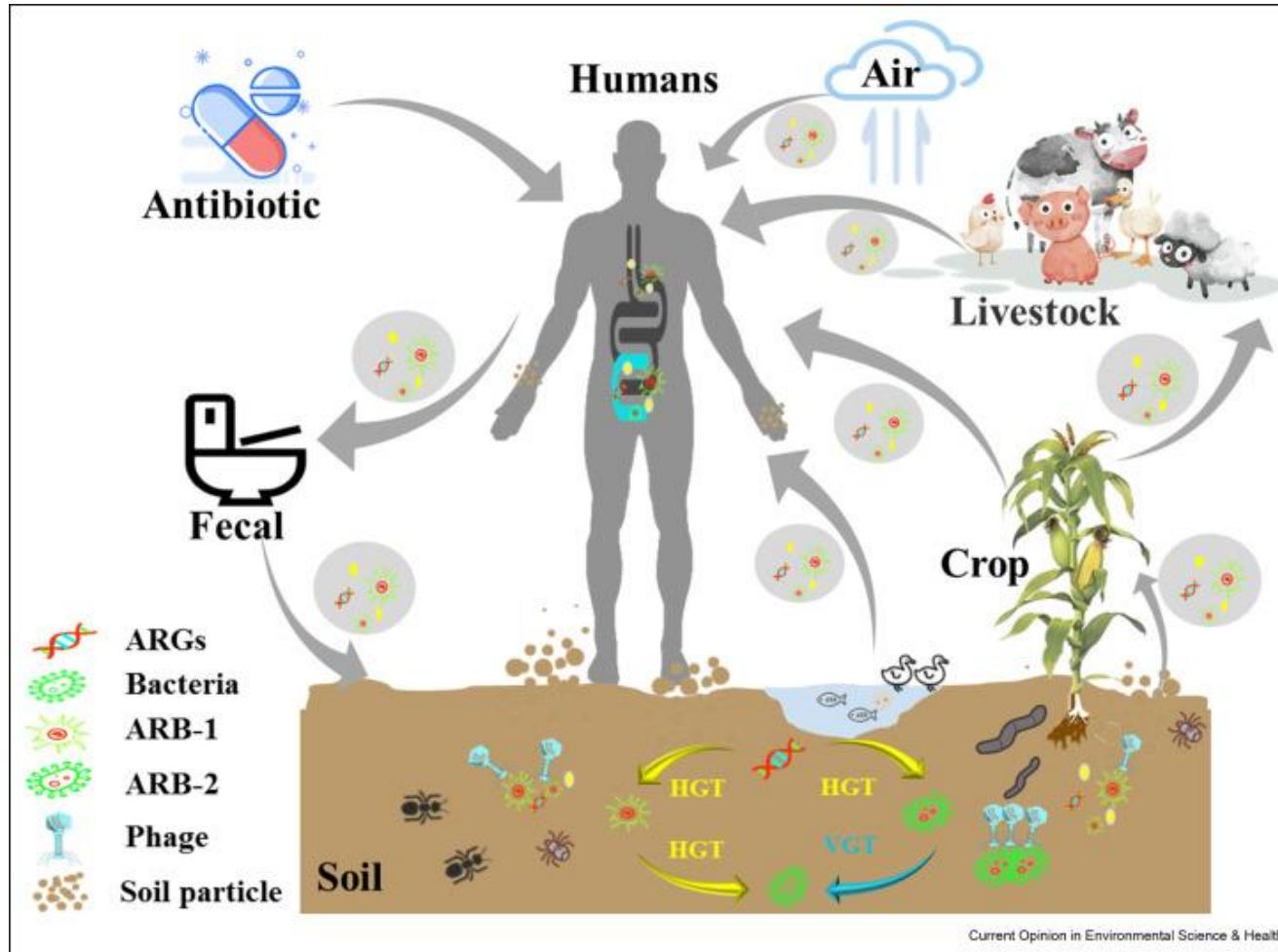
Name	Associated Condition	Result
<i>Neisseria gonorrhoeae</i>	Gonorrhea, urethritis, pelvic inflammatory disease, gonococcemia, gonococcal ophthalmia neonatorum	Not Detected
<i>Chlamydia trachomatis</i>	Chlamydia, cervicitis, urethritis, pelvic inflammatory disease	Not Detected
<i>Mycoplasma genitalium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

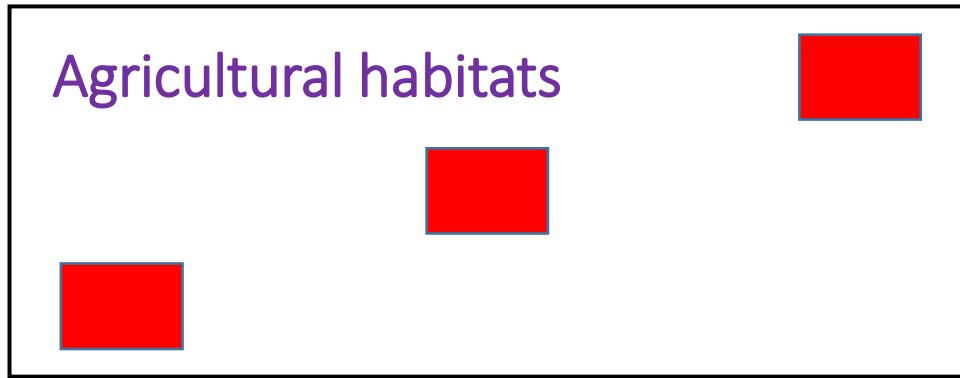
Note: Human papillomavirus (HPV) 16, 18, 31, 33, 35, 39, 45, 51, 56, 58, 59, and 68 are considered high-risk or probable high-risk due to their association with cervical cancer. HPV 6, 11, 42, 43, and 44 are considered low-risk for cervical cancer but may cause genital warts. Other HPV genotypes found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



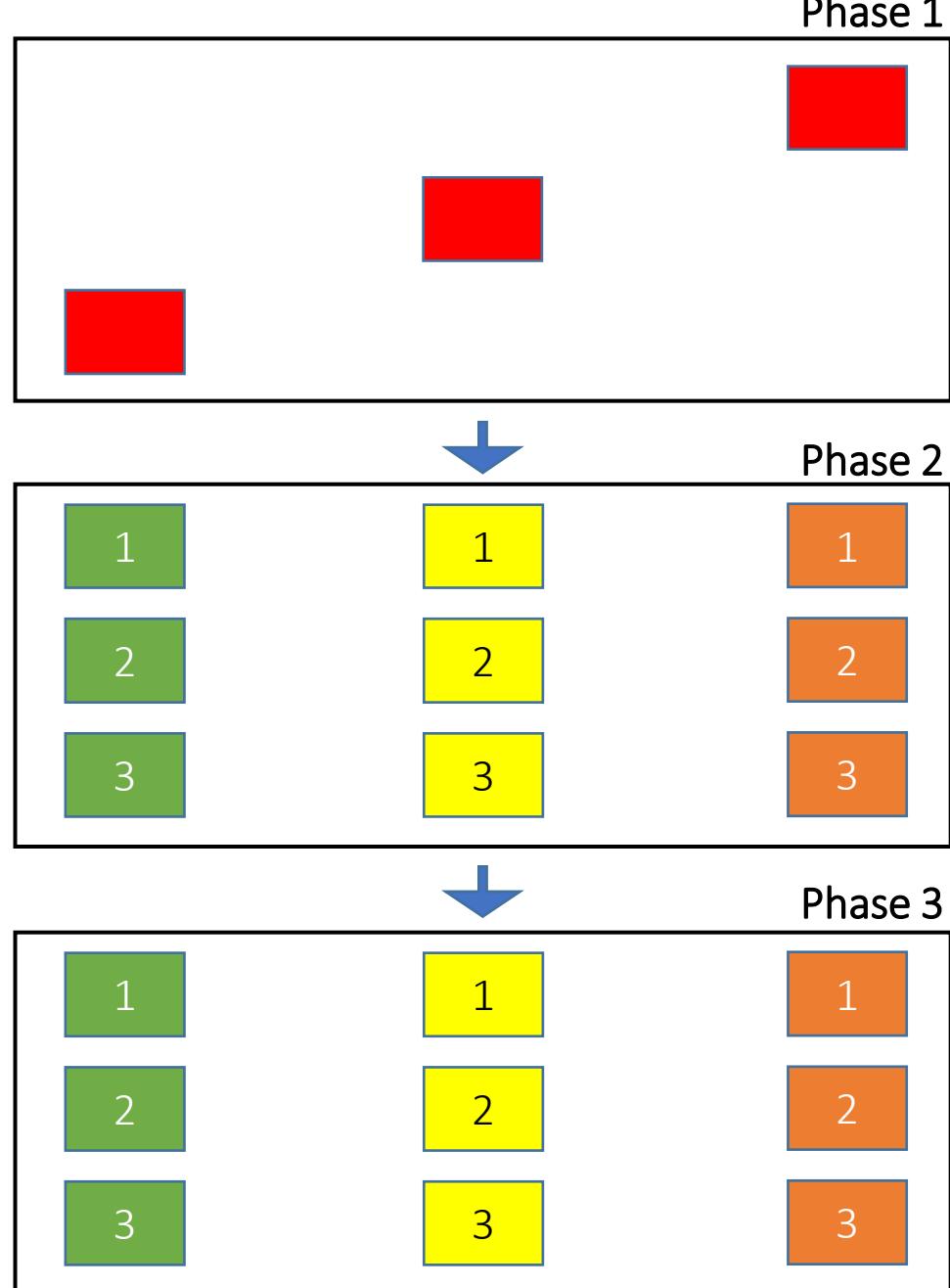
Vegetation: *Brassica integrifolia*

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

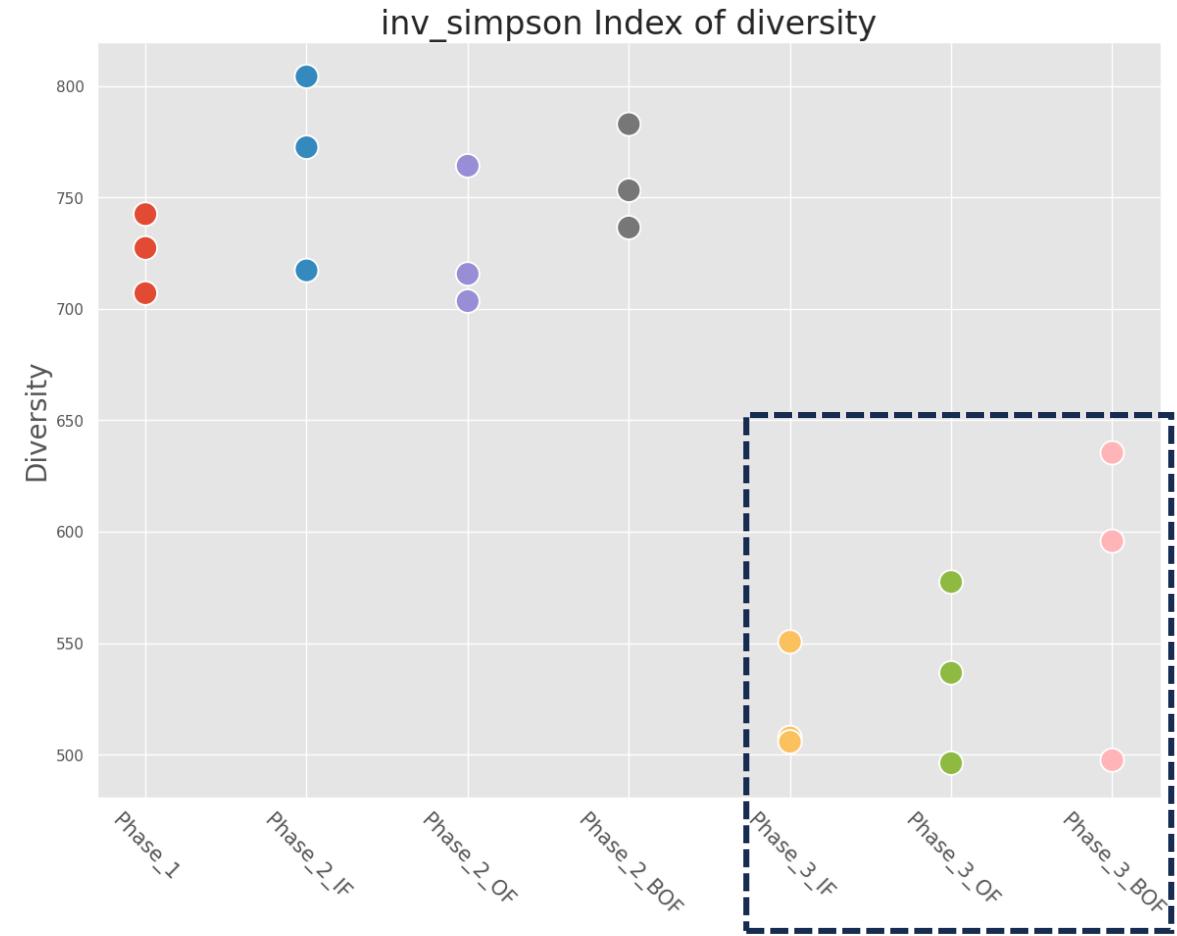
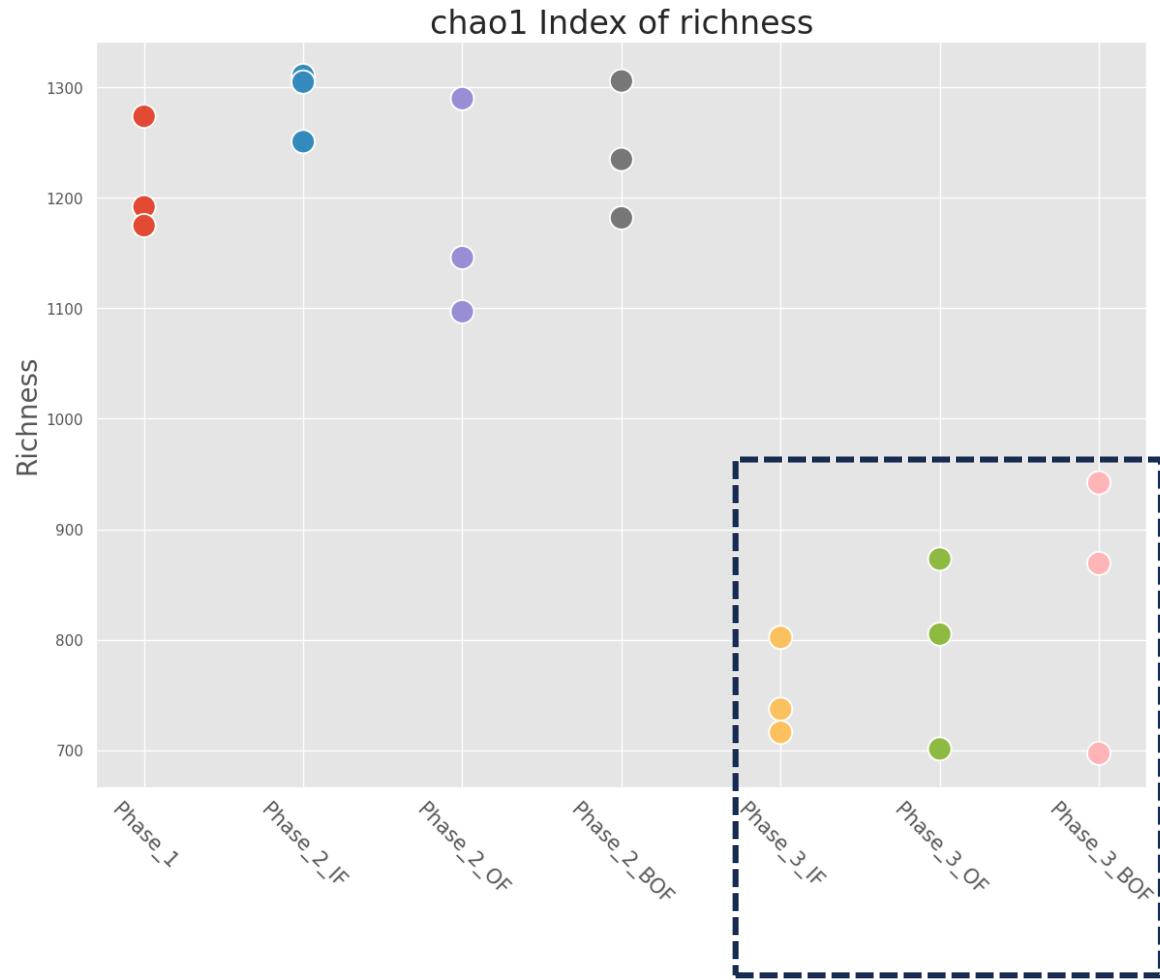
IF: No organic fertilizer, apply 100% inorganic fertilizer

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



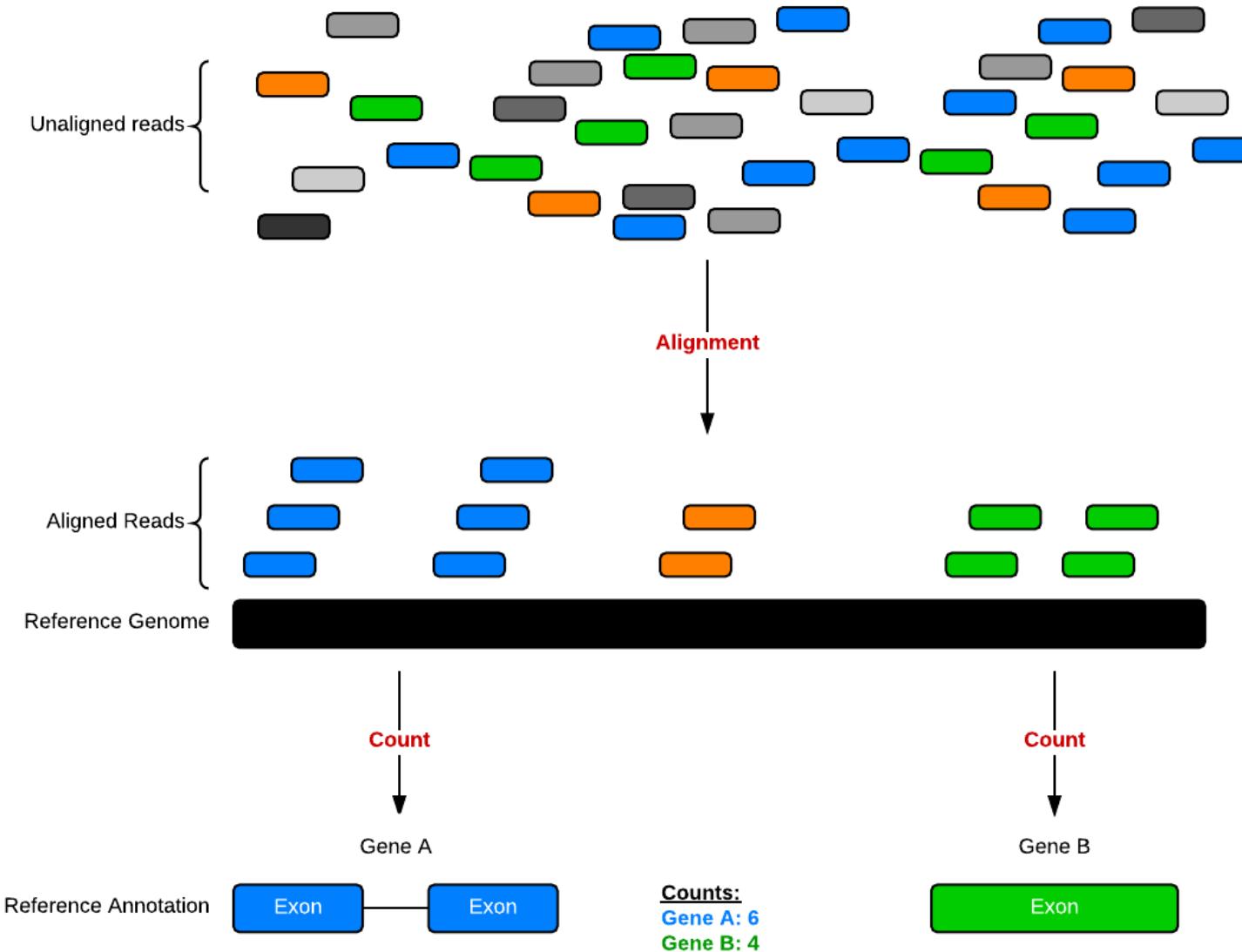
# Nội dung môn học “Phân tích dữ liệu tin y sinh”

- Processing medical and genomic raw data
- Explore data
- Feature selection
- Dimension reduction
- Clustering data
- Building machine learning and deep learning model
- Explainable AI (XAI)

# Genomic Data

- **Genomics** (SNP microarray, CNV microarray and long or short read DNA-seq/WGS)
- **Transcriptomics** (microarray, bulk RNA-seq, single-cell RNA-seq and spatial transcriptomics)
- **Proteomics** (protein microarrays and mass spectrometry)
- **Metabolomics** (mass spectrometry and Nuclear Magnetic Resonance Spectroscopy)
- **Epigenomics**
- **Methylomics** (methylation microarray, WGBS, EM-seq and long-read sequencing)
- **Metagenomics** (long or short read DNA-seq)

# How to generate genomic data: RNA-seq



# RNA-seq count table

**countData**

gene	ctrl_1	ctrl_2	exp_1	exp_1
geneA	10	11	56	45
geneB	0	0	128	54
geneC	42	41	59	41
geneD	103	122	1	23
geneE	10	23	14	56
geneF	0	1	2	0
...	...	...	...	...
...	...	...	...	...
...	...	...	...	...

**colData**

id	treatment	sex
ctrl_1	control	male
ctrl_2	control	female
exp_1	treatment	male
exp_2	treatment	female

Sample names:

**ctrl\_1, ctrl\_2, exp\_1, exp\_2**

**countData** is the count matrix  
(number of reads mapping to each gene for each sample)

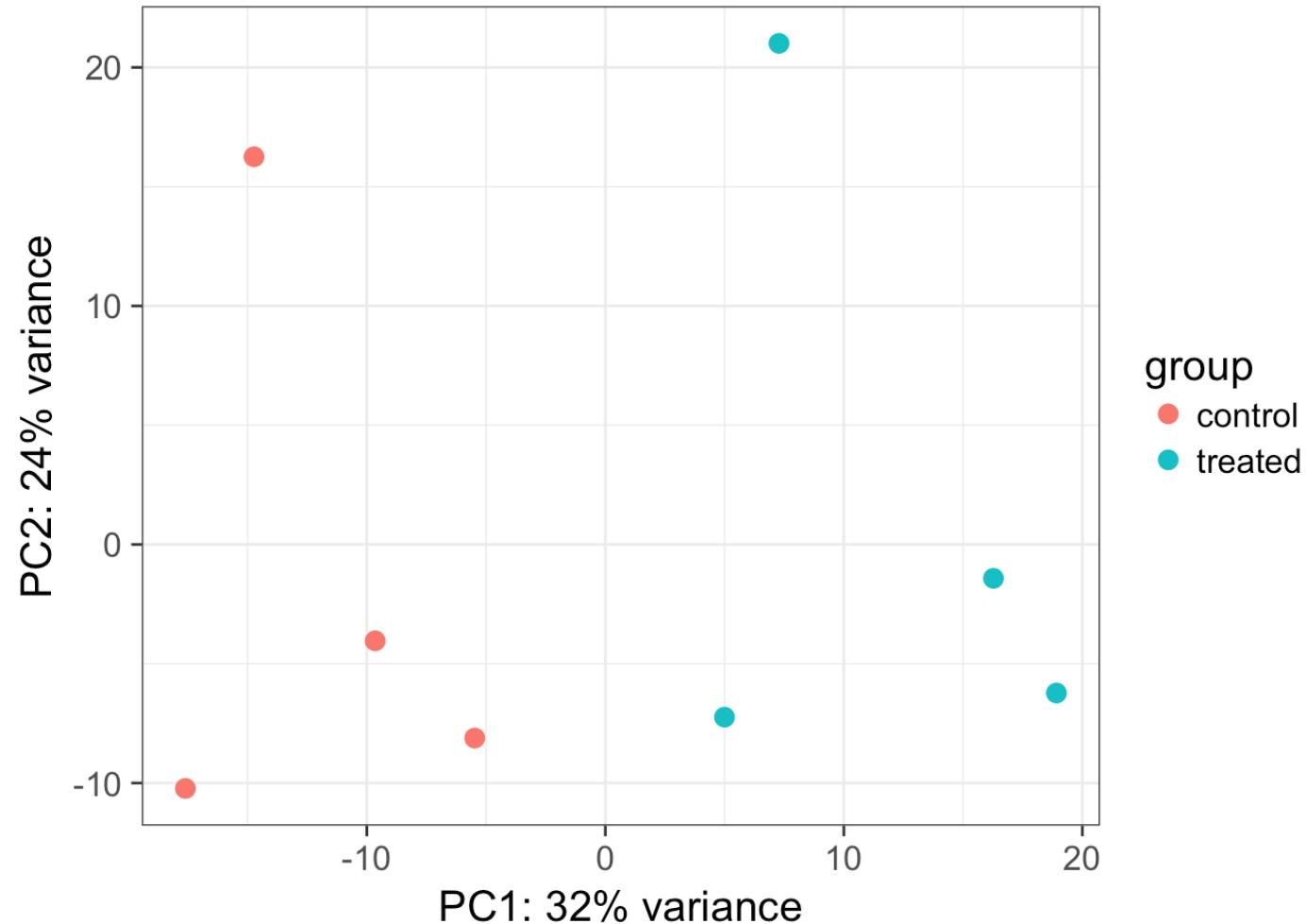
**colData** describes metadata about the *columns* of countData

**First column of colData must match column names of countData (-1st)**

# RNA-seq count table

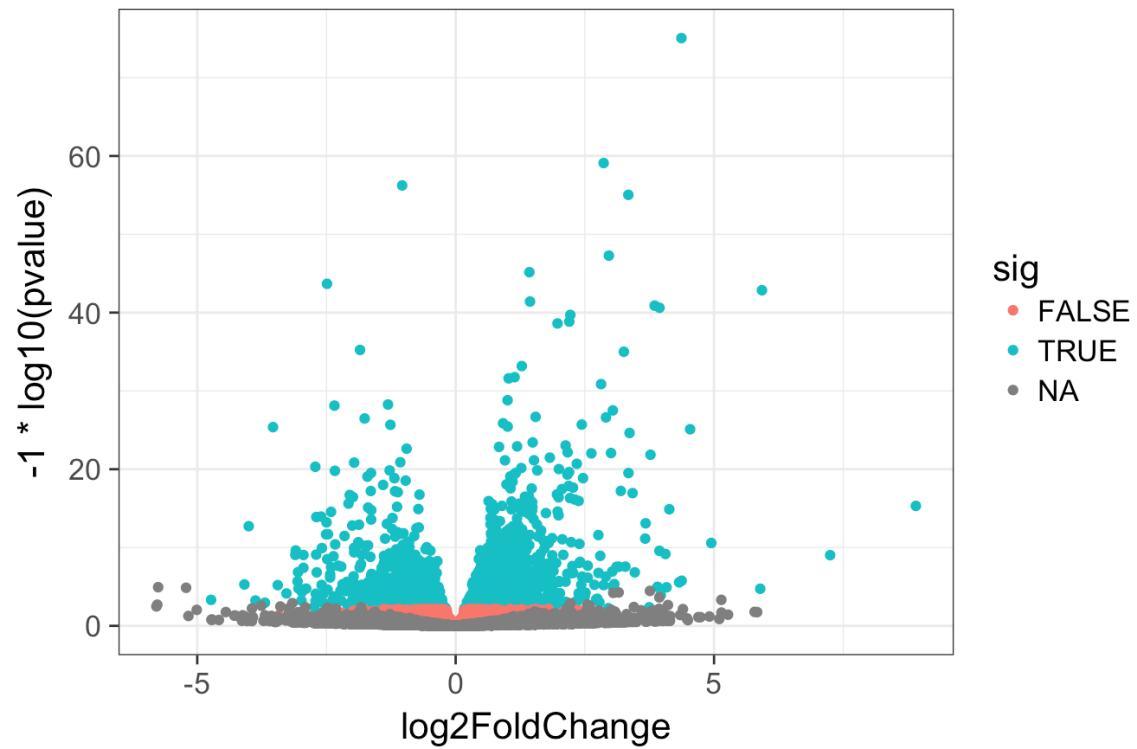
```
## # A tibble: 38,694 x 9
##       ensgene SRR1039508 SRR1039509 SRR1039512 SRR1039513 SRR1039516
##   <chr>     <dbl>     <dbl>     <dbl>     <dbl>     <dbl>
## 1 ENSG00000000003     723      486      904      445     1170
## 2 ENSG00000000005      0        0        0        0        0
## 3 ENSG00000000419     467      523      616      371      582
## 4 ENSG00000000457     347      258      364      237      318
## 5 ENSG00000000460      96       81       73       66      118
## 6 ENSG00000000938      0        0        1        0        2
## 7 ENSG00000000971    3413     3916     6000     4308     6424
## 8 ENSG00000001036    2328     1714     2640     1381     2165
## 9 ENSG00000001084    670      372      692      448      917
## 10 ENSG00000001167    426      295      531      178      740
## # ... with 38,684 more rows, and 3 more variables: SRR1039517 <dbl>,
## #   SRR1039520 <dbl>, SRR1039521 <dbl>
```

# RNA-seq Downstream Analysis



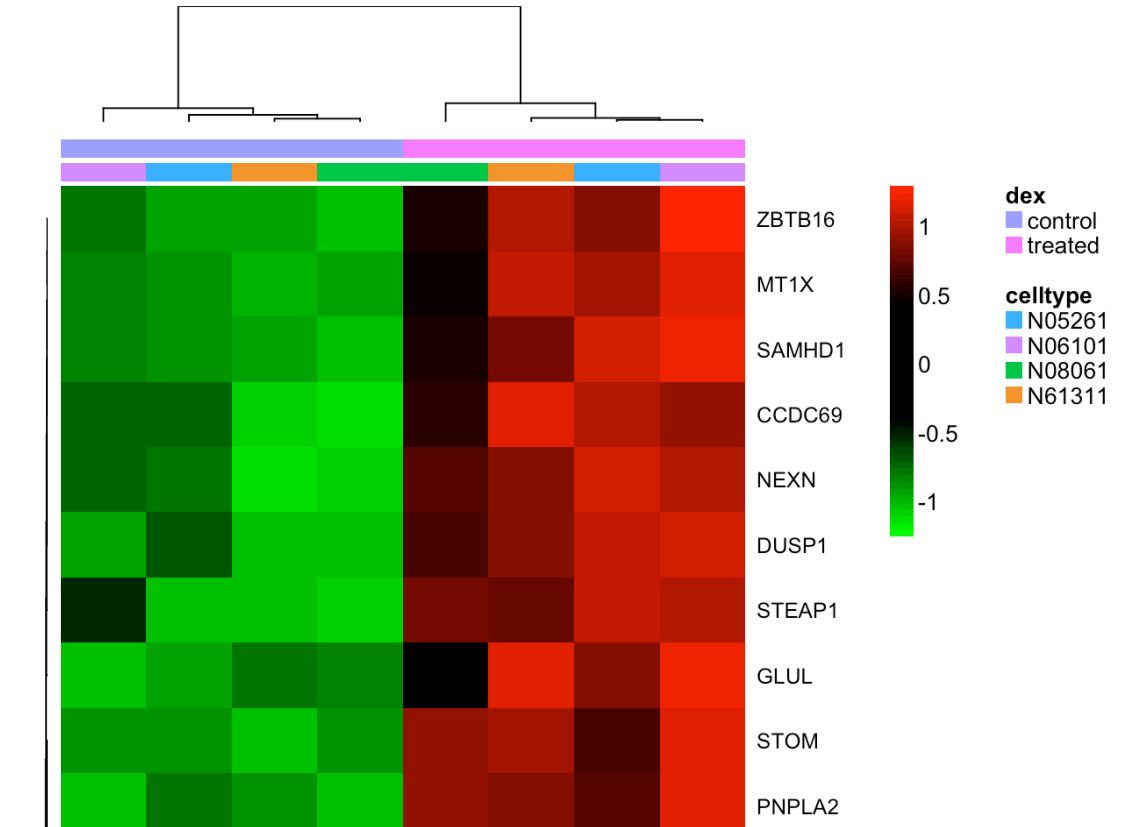
# RNA-seq Downstream Analysis

Volcano plot

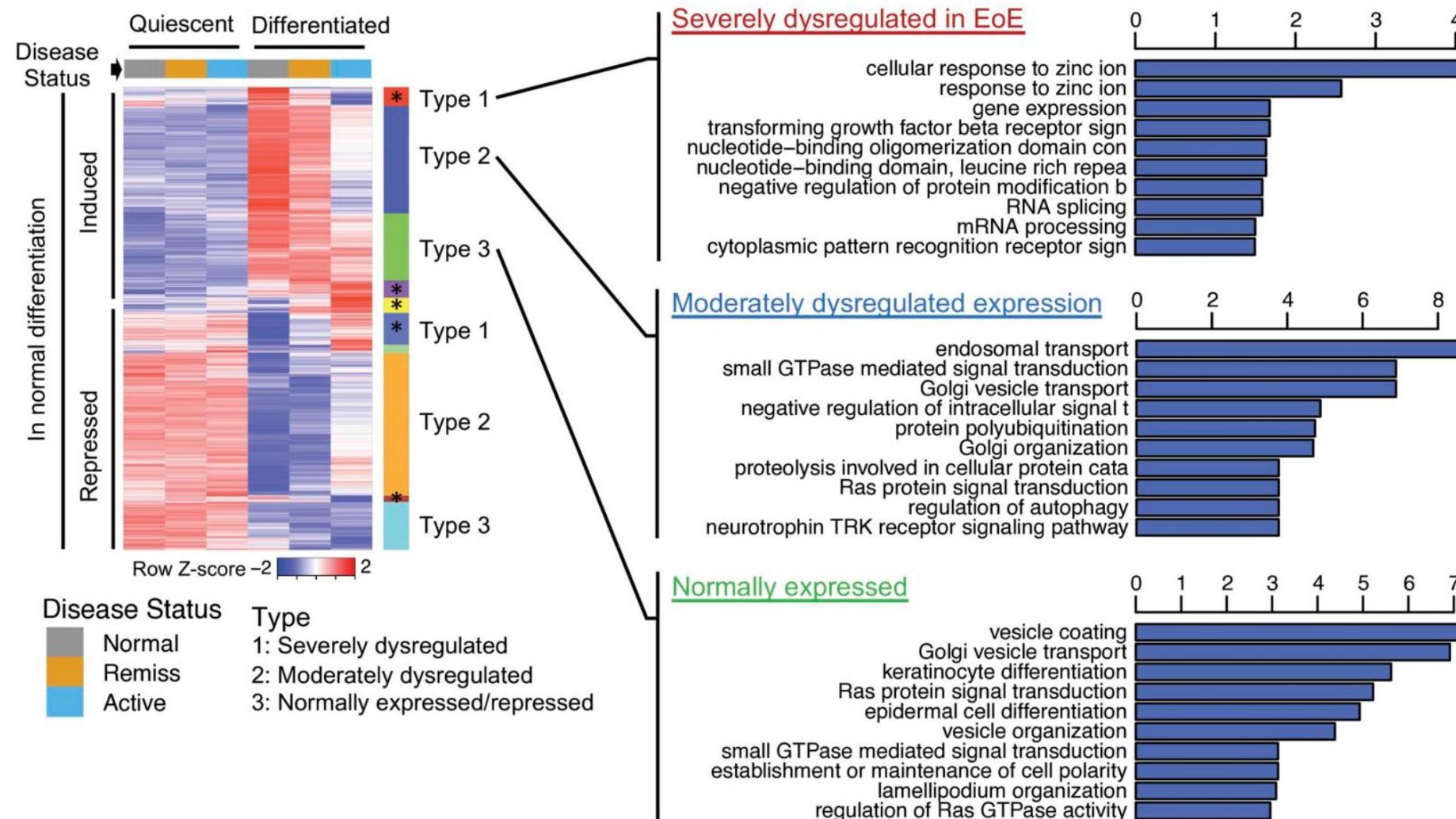


sig

- FALSE
- TRUE
- NA



# RNA-seq Downstream Analysis



# 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 Quỳnh Như
- KS. Nguyễn Minh Hoàng
- ThS. Ông Xuân Hồng
- DS. Nguyễn Thanh Trà

## Nhóm cộng tác nghiên cứu

- Trần Quang Hải
- TS. Nguyễn Anh Thư
- KS. Hoàng Kim
- KS. Huy Hà
- Bác sĩ Hoàng Sơn
- TS. Bác sĩ Nguyễn Thị Kim Nhi
- ThS. Bác sĩ Ngô Viết Hậu
- GS. TS. Bác sĩ Phạm Lê An
- TS. Bác sĩ Lương Thị Mỹ Hạnh
- TS. Trịnh Vạn Ngữ

# Xin chân thành cảm ơn!

Luu Phuc Loi, PhD

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

Zalo: 0901802182