

Big Data ฐานข้อมูลพันธุกรรมคนไทย ใครบ้างจะได้ประโยชน์

การประชุมวิชาการ 30 ปี สวรส.

ศิษย์เก่า ทองสีมา/ 7 ก.ค. 2565



พันธุกรรม vs. สิ่งแวดล้อม

Prevalence curve

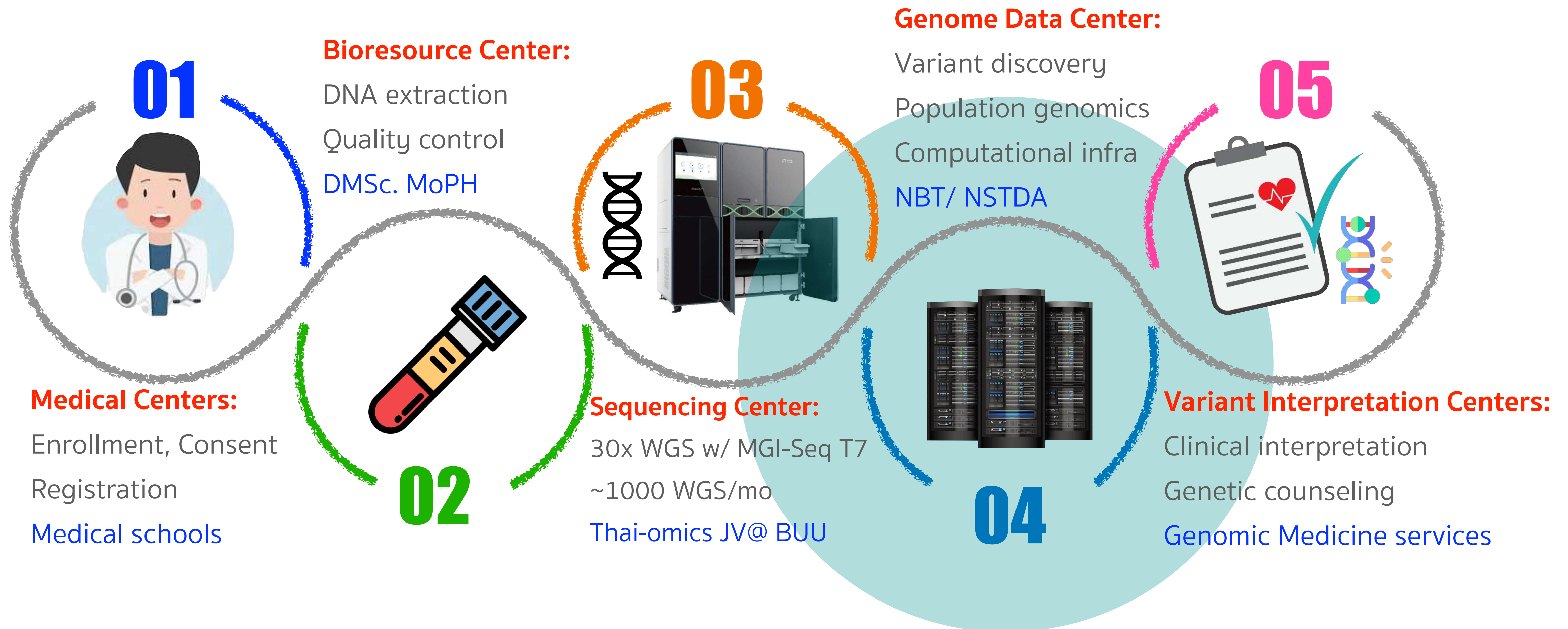


Rare diseases

Obesity
Diabetes
Cancers etc.

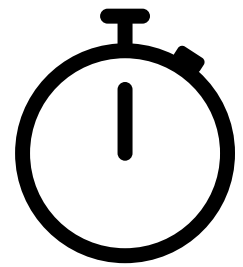
Infectious/
Accidents etc.

โครงการถอดรหัสพันธุกรรม 50,000 ราย

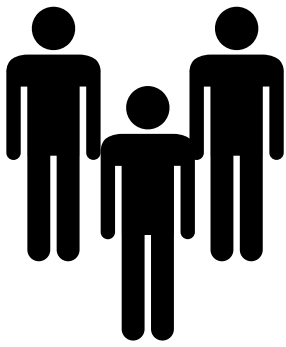


ข้อมูล ดีเอ็นเอ ของเราใหญ่แค่ไหน

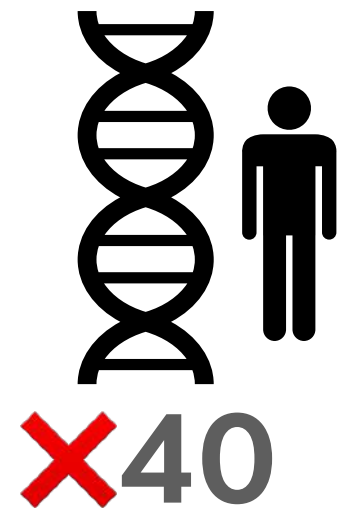
3 พันล้านเบส



48 คน/36 ชั่วโมง



50,000 คน

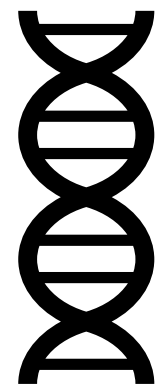


X40

120 ทิกะไบท์



1,000,000 แผ่น



129 เมตร

```
(qiime2-2018.2) [yuuuhee@node01 0410_qiime2]$ head QY1-48.new.fastq
@QY38_1 M01056:42:000000000-A4R2M:1:1101:16586:1574
TACGGAGGGTCAAGCGTTAATCGGAATTACTGGGCGTAAAGCGCATGCAGGTGGTTTGTAAAGTCAGATGTGAAAGCCCTGGGCTCAACCTAGGAATCGCATTTGAAACTGACAAGCTAGAGTACTGTAGAGG
GTAGAATTTTCAGGTGTAGCGGTGAAATGCGTAGAGATCTGAAGGAATACCGGTGGCGAAGGCGGCCCCCTGGACAGATACTGACACTCAGATGCGAAAGCGTGGGGAGCAAACAGG
+
HHHHGGGGGGGGHHHHGGGGGHGGGGHHHHHHHHGGGGHHHHHHGGGGHHHHHHGGGGHGHHHHHHHHHHHHHHHHHHHHGHGGGGHGHGGGGHHHHHHHHHHHHHHHHHHHHHHGGGGHHHHHHHHHHHHHHHHHHHHHH
HHGHFHHHFGHHHGGGGHGHGGGGFFGFHGHGHGHHGHHGHHGGGGGGGHHGGGGGGGGGHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
@QY12_1 M01056:42:000000000-A4R2M:1:1101:17149:1607
TACGTATGCCCCGAGCGTTATCCGGATTTATTGGGCGTAAAGCGAGCGCAGACGGTTGATTAAGTCTGATGTGAAAGCCCGGAGTTCAACTCCGGAATGGCATTGAAACTGGTTAACTTGAGTGTTGTAGAGG
GTGAACTCCATGTGTAGCGGTGGAATGCGTAGATATATGGAAGAACCAGTGGCGAAGGCGGCTTACTGGACAACAACTGACGTTGAGGCTCGAAAGTGTGGGTAGCAAACAGG
+
GGGGHHHHHHHHHHGGGGGGEGHHGGG?EEHHHHHHHFFGGGHHHHGGGGGGGFFHCEFFFFGHDFFHHGGGHHGGGGHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
HHHHHHHHHHHHHHHHHHGGDDHFEHGGGGHHHHHFGHGHGHHHHHHGGGGFFHHGGGFE1HGGGECGGF5BFG35HFFHHGGB5?GHGHFGGEEAAA5HHHHGGGGFFHHGD3HHHHHHF
@QY38_2 M01056:42:000000000-A4R2M:1:1101:16704:1609
TACGGAGGGTCAAGCGTTAATCGGAATTACTGGGCGTAAAGCGCATGCAGGTGGTTTGTAAAGTCAGATGTGAAAGCCCTGGGCTCAACCTAGGAATCGCATTTGAAACTGACAAGCTAGAGTACTGTAGAGG
GTAGAATTTTCAGGTGTAGCGGTGAAATGCGTAGAGATCTGAAGGAATACCGGTGGCGAAGGCGGCCCCCTGGACAGATACTGACACTCAGATGCGAAAGCGTGGGGAGCAAACAGG
```


จากชลบุรี-สู่-ปทุมธานี

การเดินทางของข้อมูลดีเอ็นเอ

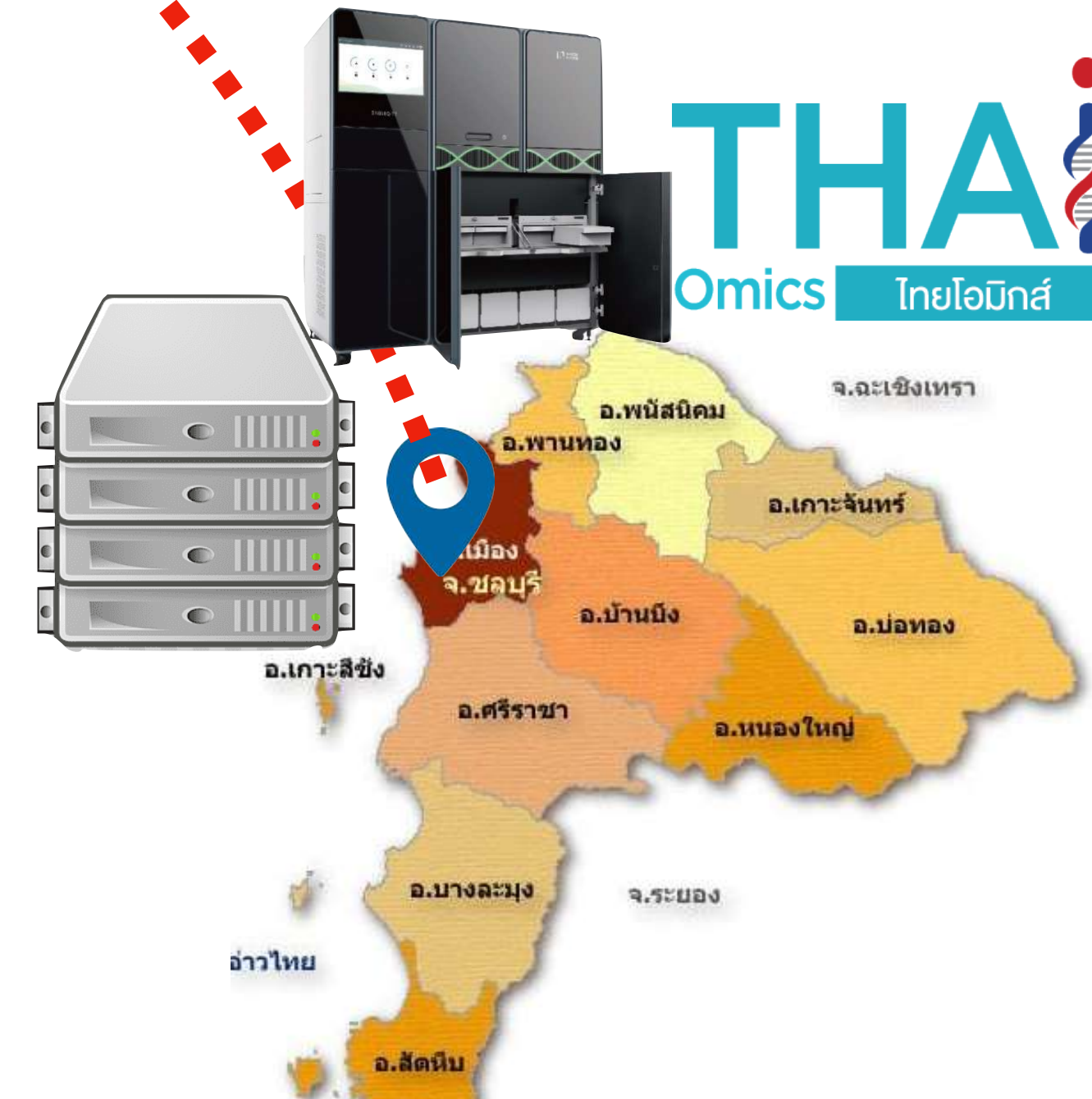
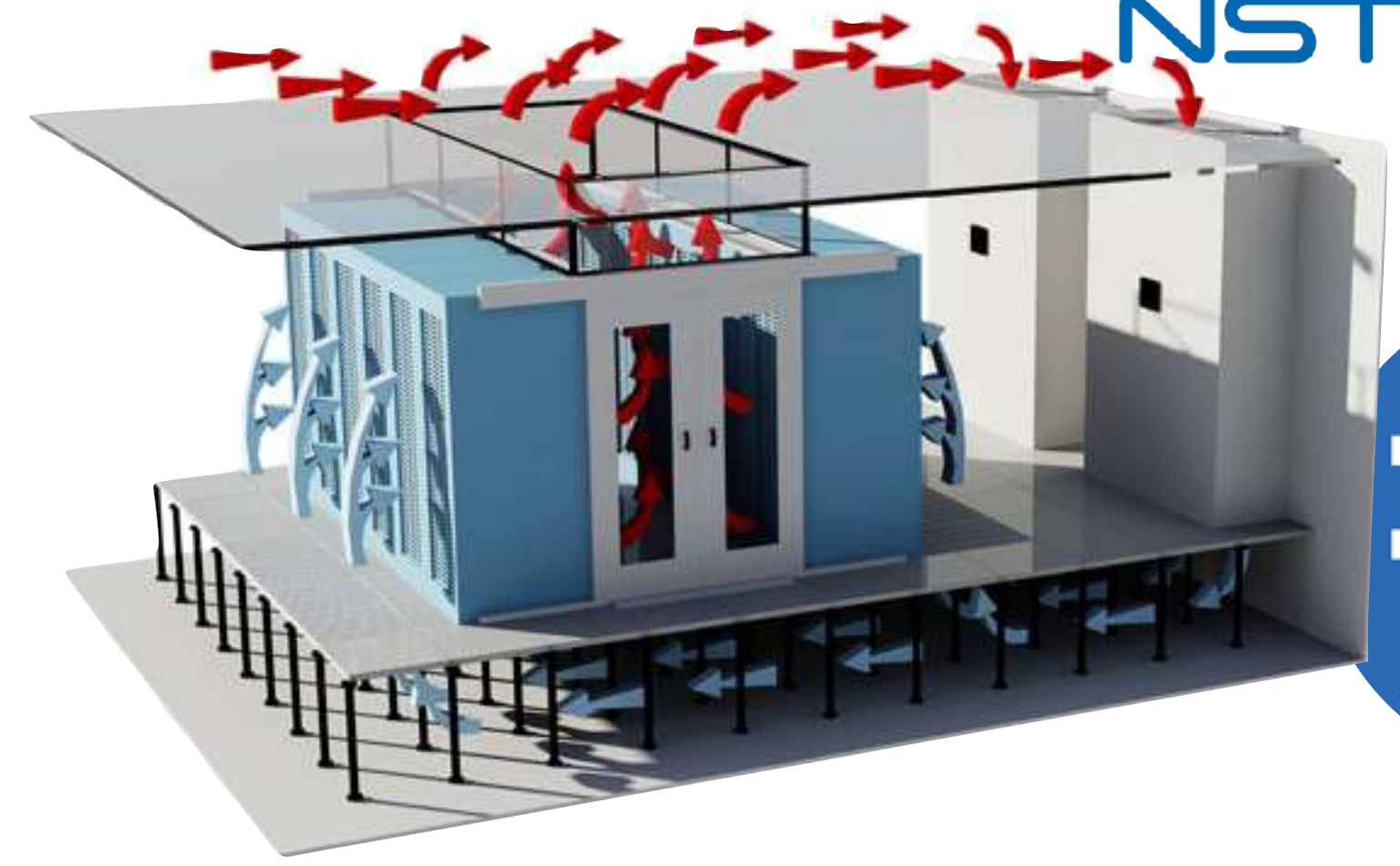


ข้อมูล 48 คน (5TB) ~ 2.15 ชม

134 กิโลเมตร



สาขา NSTDA



THA
Omics ไทยโอมิกส์

ข้อมูล ดีเอ็นเอ ไปทำอะไร

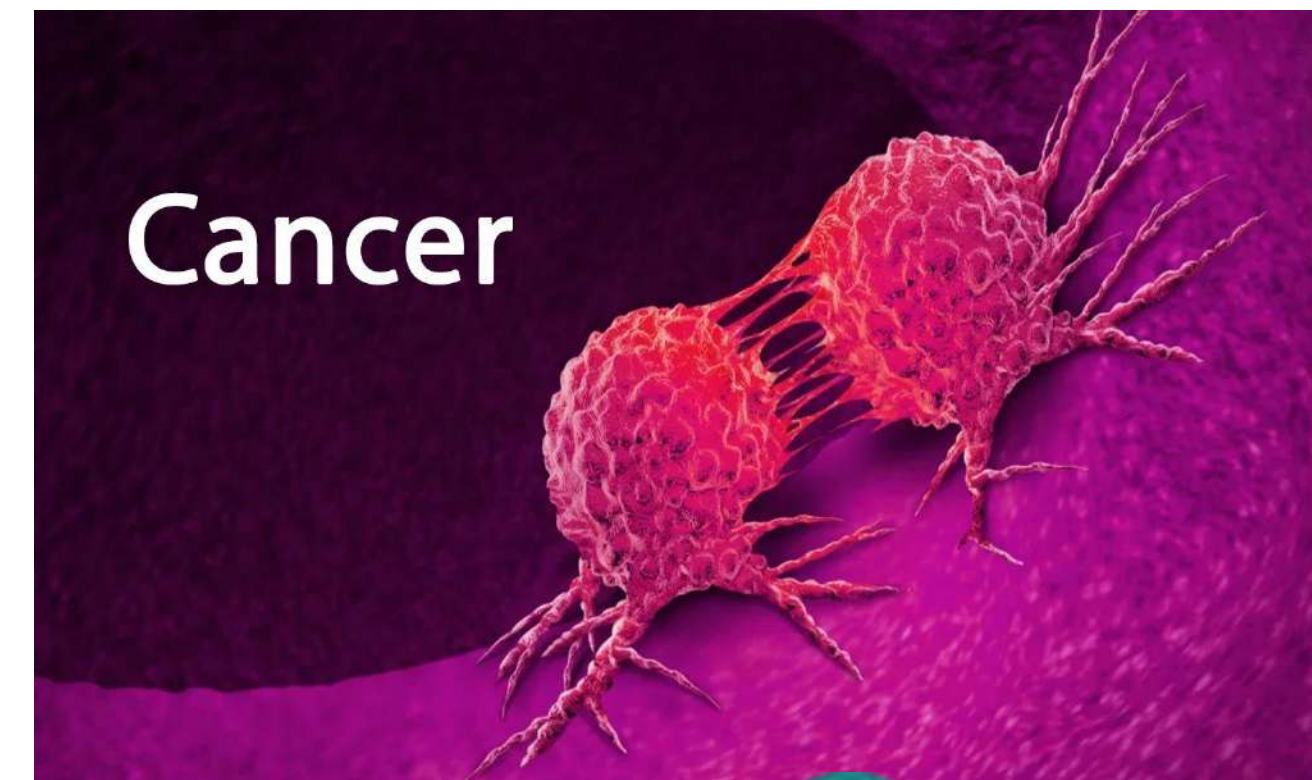
วินิจฉัย และเลือกการรักษาที่เหมาะสม/ พยากรณ์โรค

- การอ่าน = ถอดรหัสพันธุกรรม
- การวินิจฉัย = ระบุตำแหน่งกลายพันธุ์
- ผล = เลือกการรักษา/ความเสี่ยง
- การเขียน = เข้าไปแก้ไขสารพันธุกรรม



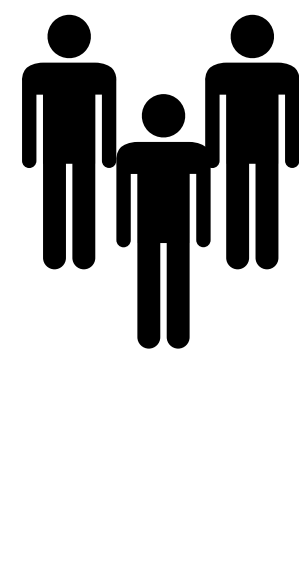
ผลการวิเคราะห์ใช้ทำอะไร

- **โรคหายาก:** > 7000 วินิจฉัยเร็วช่วยรักษา
- **โรคมะเร็ง:** วินิจฉัยเร็ว วางแผนในการดูแล
- **โรคซับซ้อนไม่ติดต่อ (NCD):** พยากรณ์ความเสี่ยง หลีกเลี่ยงความเสี่ยง
- **เภสัชพันธุศาสตร์:** รับยาเหมาะสม หลีกเลี่ยงอาการรุนแรง
- **โรคติดต่อ:** หลีกเลี่ยงความเสี่ยงการติดเชื้อ



ไล่ล่าหาคนทำผิด

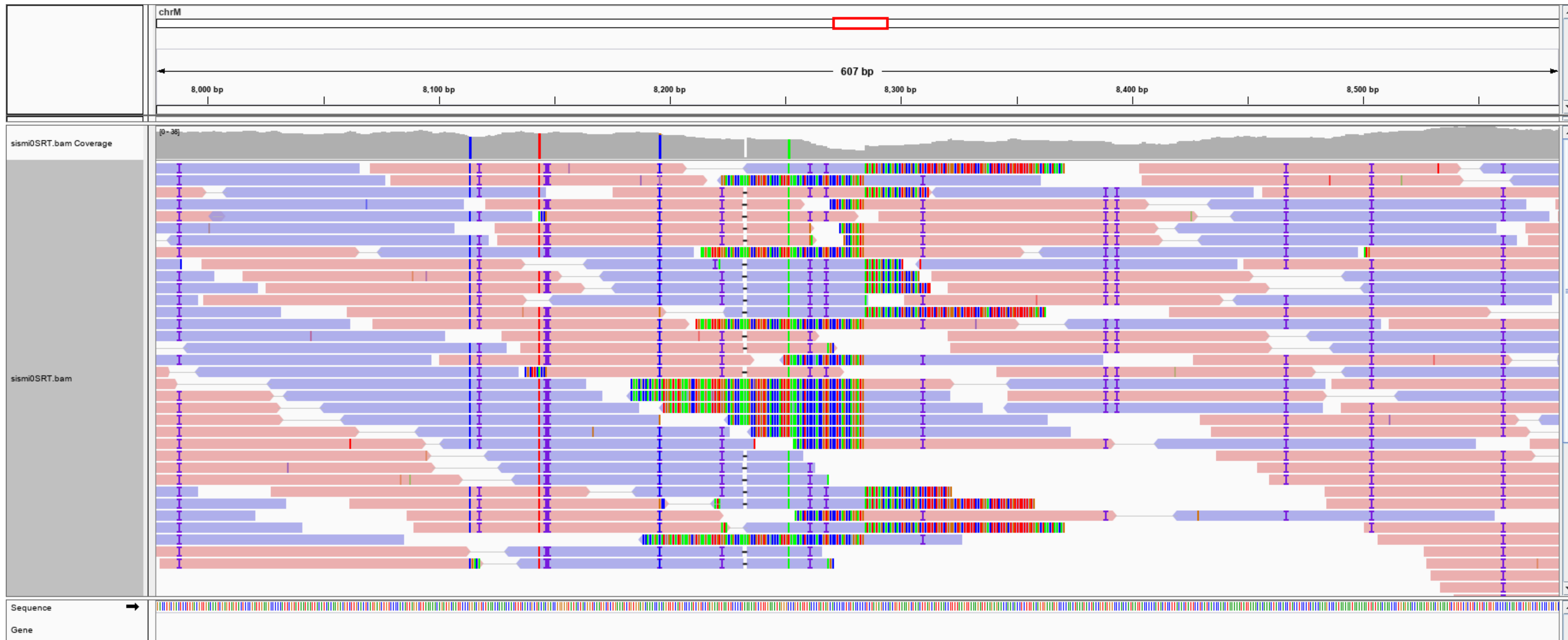
Hunting for the culprit(s)



50,000 คน = 5 ล้าน กิกะไบต์ (เพตาไบต์)

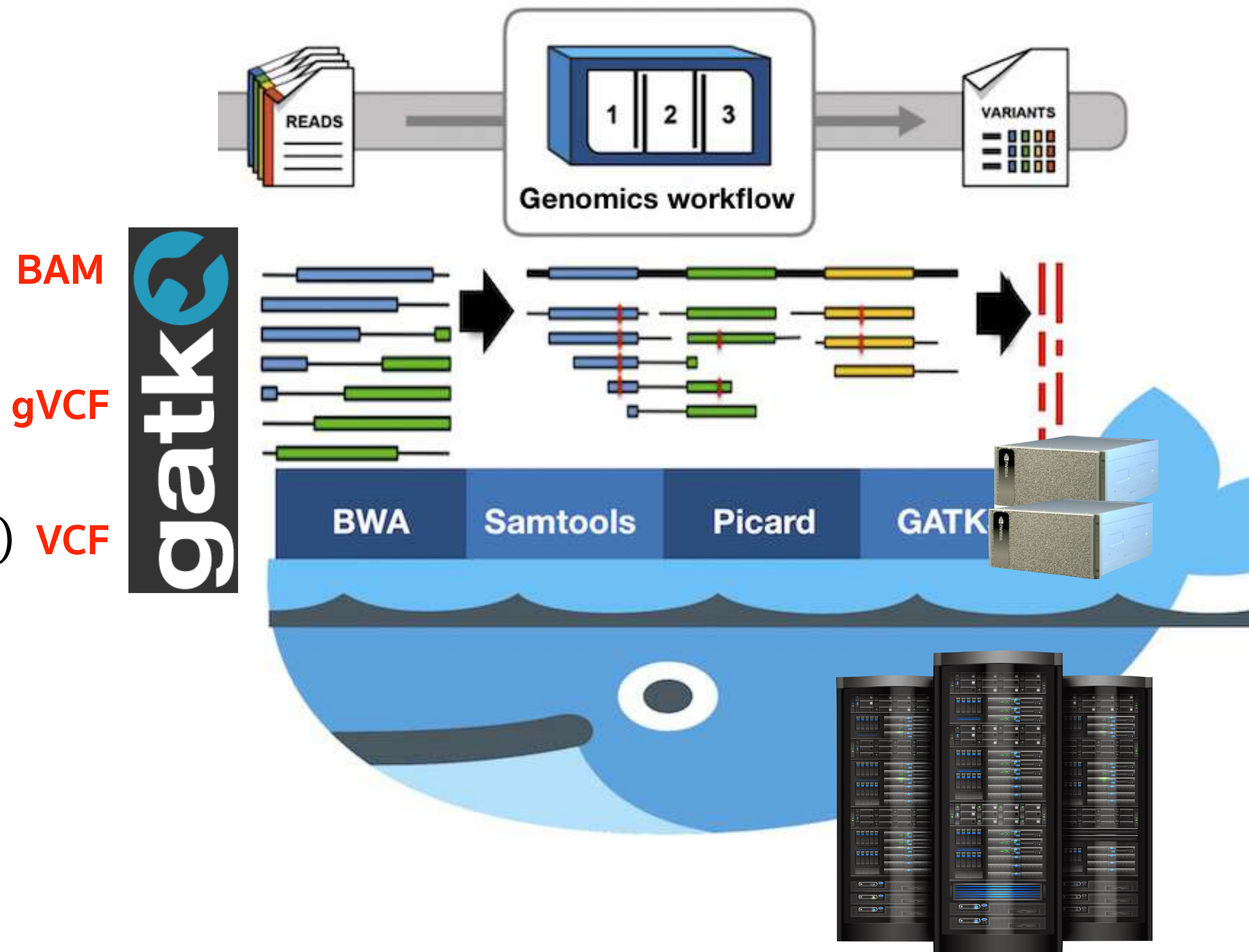


1 คน = 120 กิกะไบต์



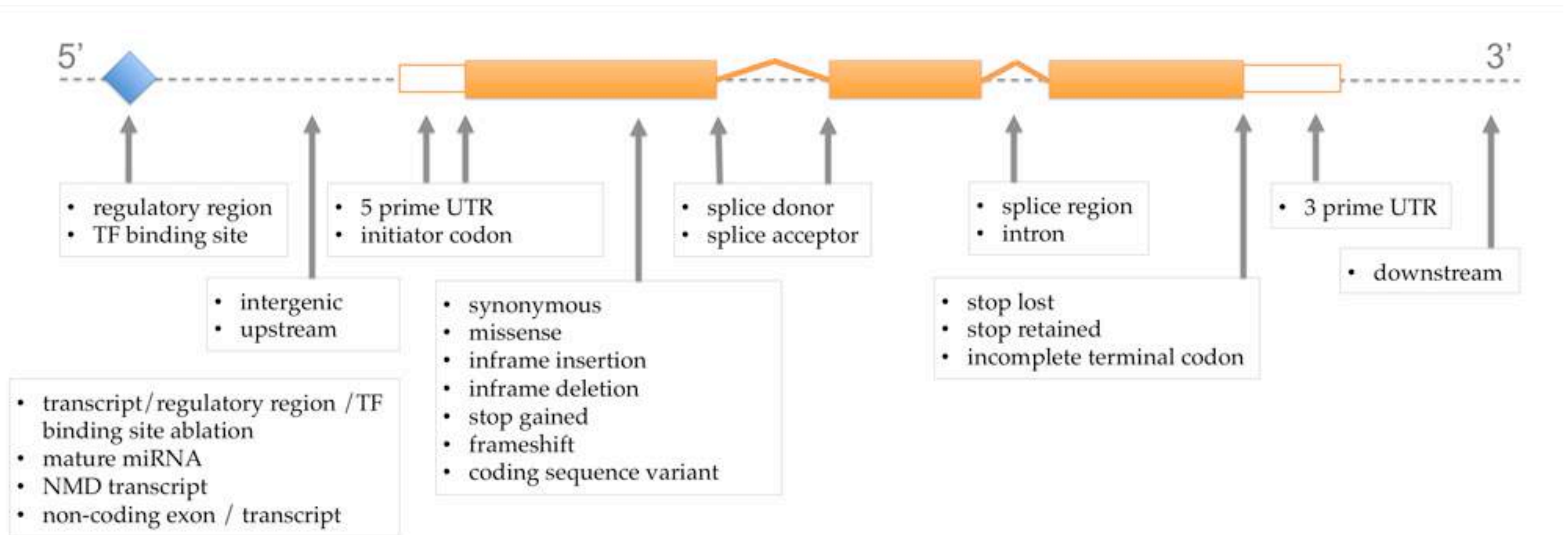
ขั้นตอนการประมวลผล

- 1 Base-calling (raw sequencing data)
- 2 Alignment to GRCh38 (RefSeq)
- 3 Variant calling (GATK4)
- 4 GenomicsDBImport (CombineGVCF) **VCF**
- 5 Variant annotation (ClinVarDB)
- 6 Variant effect prediction



ทำนายผลของการกลายพันธุ์

Variant effect predictor

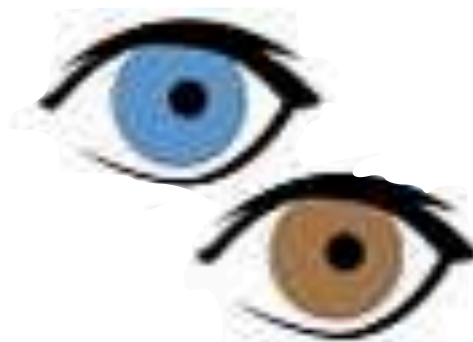


ฐานข้อมูลการกลายพันธุ์มีไว้ทำไม

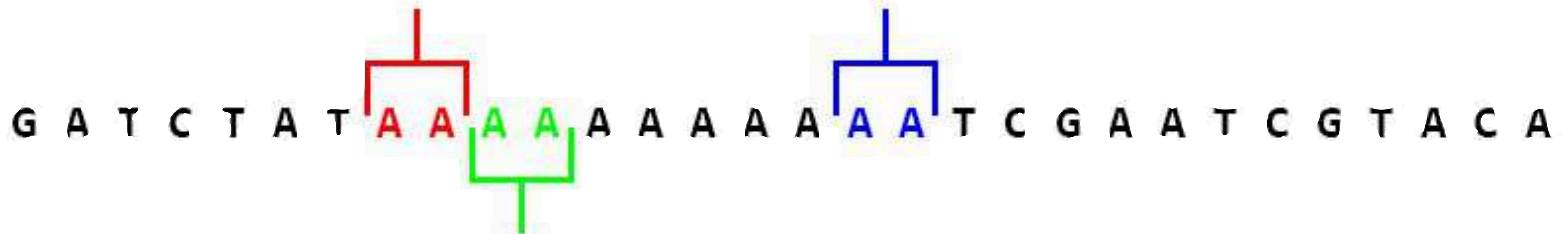
Variant annotation



Population frequency of deletion from gnomAD in VCF



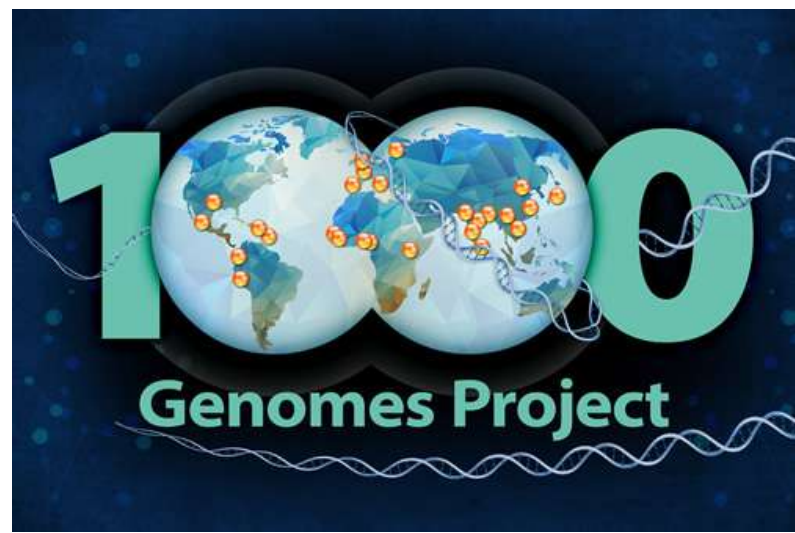
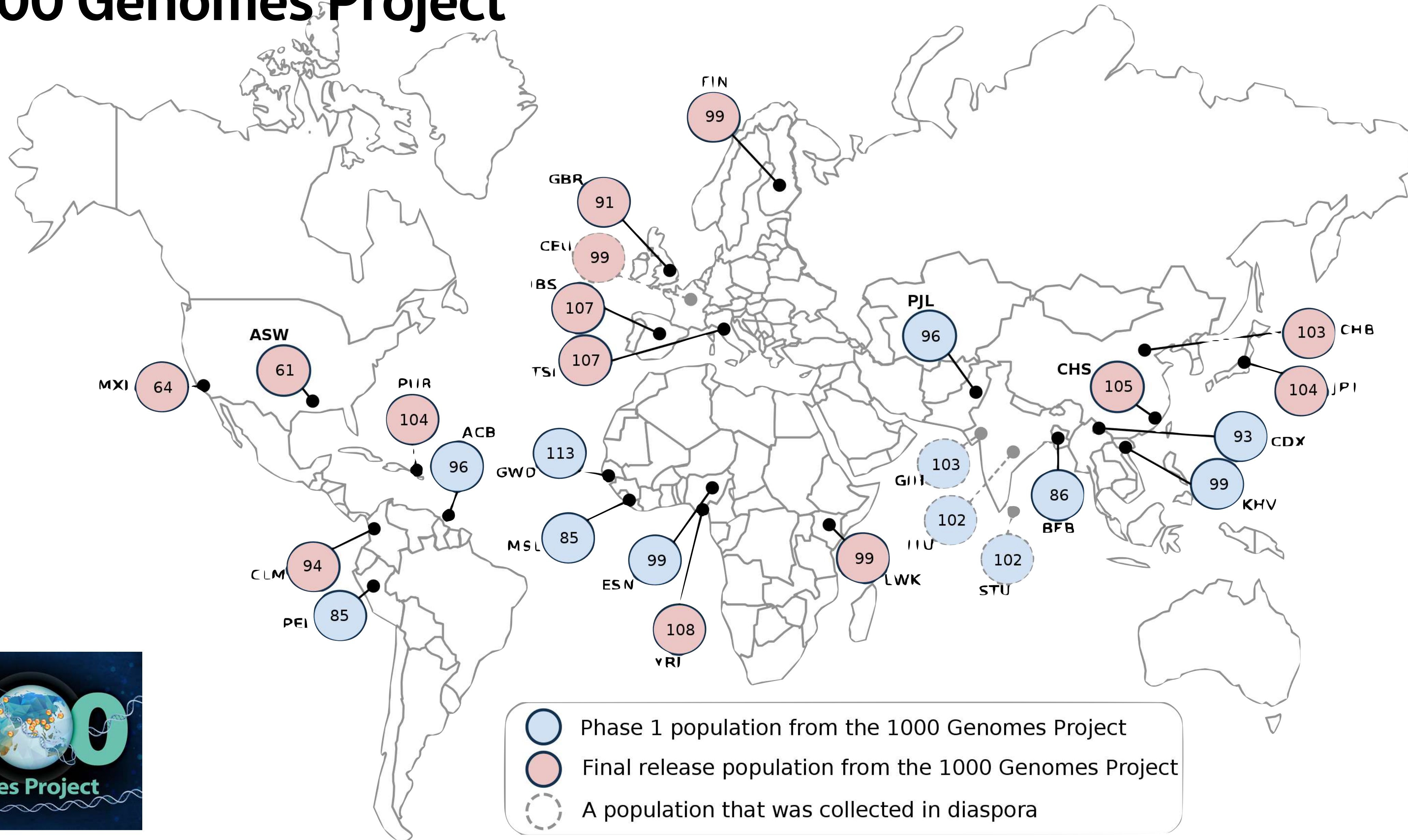
Clinical report on deletion in HGVS



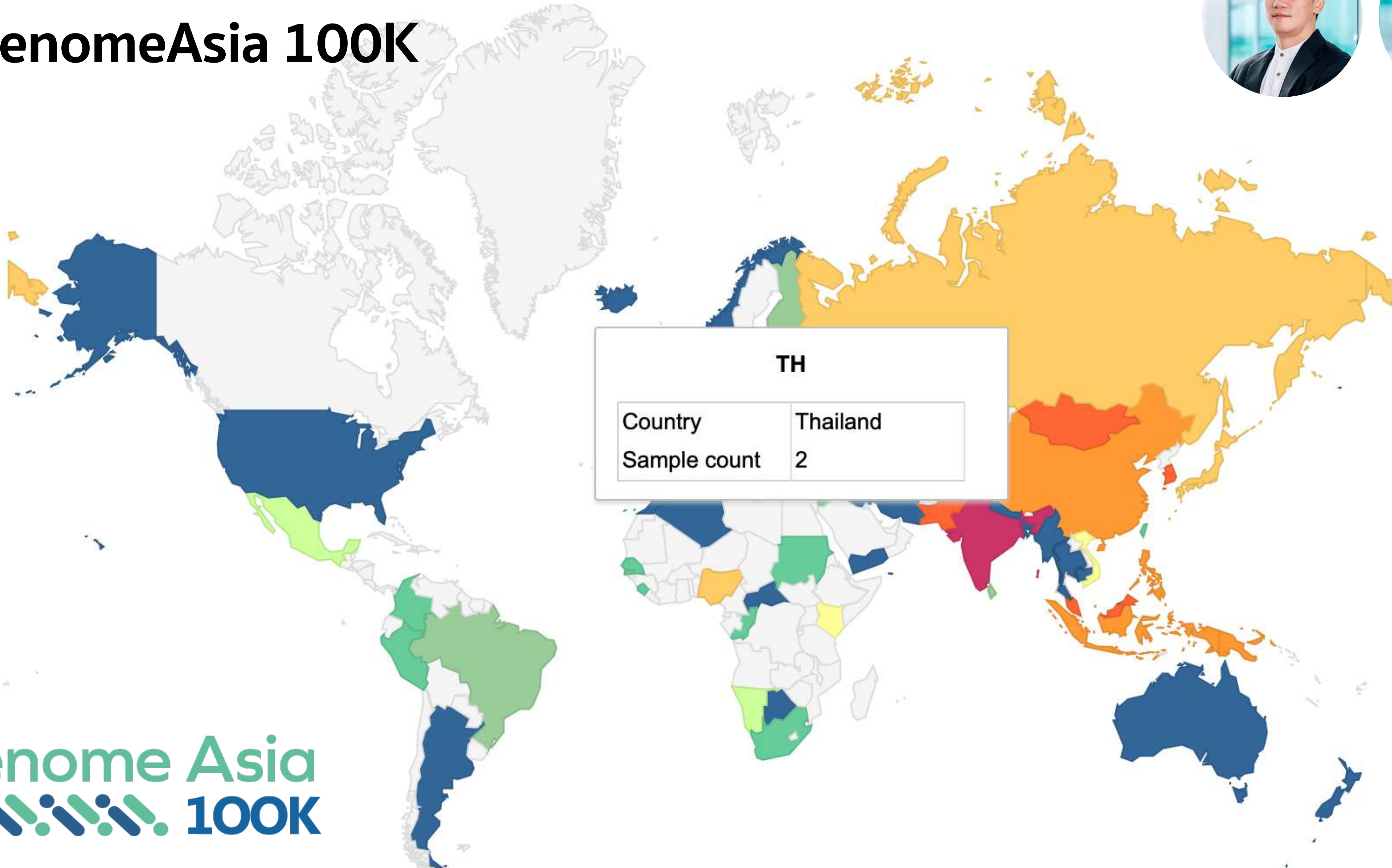
Deletion described in paper



1000 Genomes Project

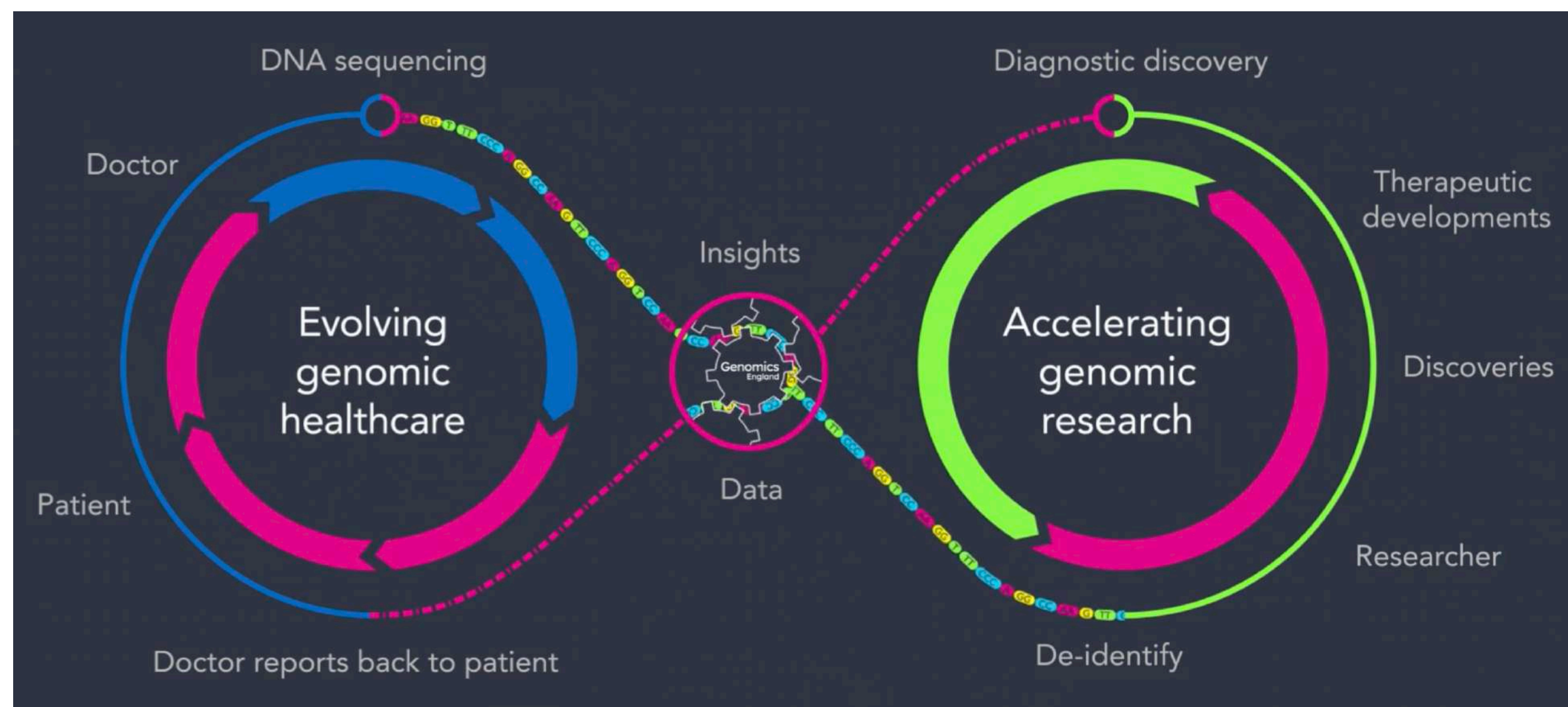


GenomeAsia 100K



โครงการคล้ายกับเรา

All-of-US และ Genomics England



Participants at a Glance

506,000+

Participants

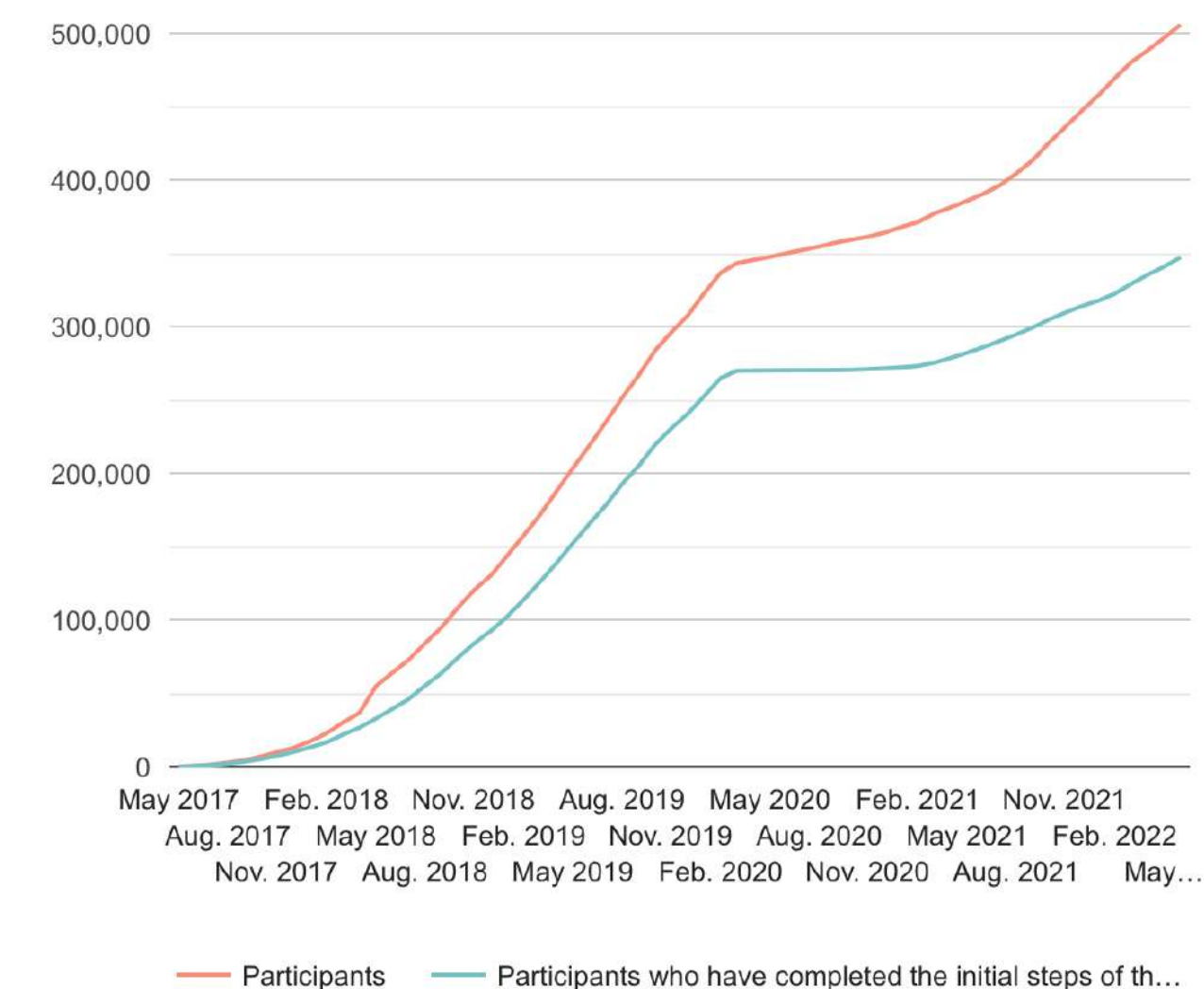
347,000+

Participants who have completed initial steps of the program

Enrollment Numbers

This graph represents participants who have consented to join the program and those who have completed all initial steps of the program. The initial steps are consenting, agreeing to share electronic health records, completing the first three surveys, providing physical measurements, and donating at least one biospecimen to be stored at the biobank.

The following numbers are approximated to protect participants' privacy. Numbers are updated as of July 5, 2022.



All of Us | The Future of Health Begins With You
RESEARCH PROGRAM

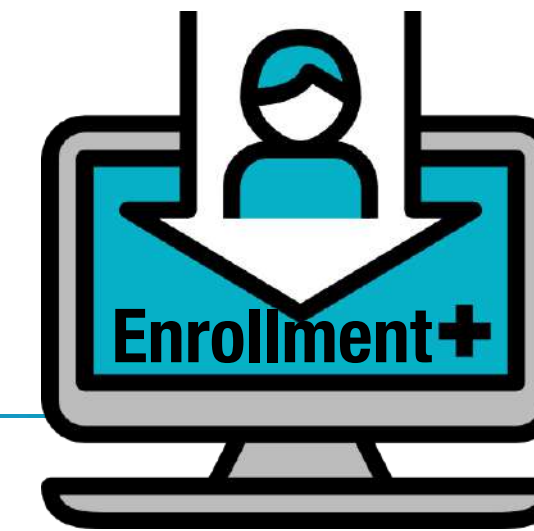
บทบาท สวทช. กับจีโนมิกส์ประเทศไทย

บริหารจัดการข้อมูล ประมวลผล และพัฒนาซอฟต์แวร์

- ออกแบบจัดซื้อระบบ HPC:
 - เร่งการประมวลผลข้อมูลพันธุกรรม
 - ดูแลข้อมูลรหัสพันธุกรรมและข้อมูลส่วนบุคคล
- พัฒนานโยบายและธรรมาภิบาลข้อมูล
 - การแบ่งปันข้อมูล
 - การแบ่งปันทรัพยากรสินทางปัญญา



ซอฟต์แวร์ด้านการบริหารจัดการข้อมูล



» data » P0000061

P0000061
Reported by Worrachet Intachai on 2020/06/29 15:45 · Last modified on 2020/06/29 16:24

Patient information

PhenoTips identifier:	61
Identifier:	2859
Exam Date:	2020-06-29
Life status:	Alive
Date of birth:	2013-06-07
Sex:	Female

Family history and pedigree
FAM0000015

Family members:
• 2859 (2013-06-07)

This is the proband

Clinical symptoms and physical findings

CRANIOFACIAL
Mesiodens

Specimen Management System

Executive Projects Executive

Medical Center

Week

0 / 225

0 Resampling

Month

705 / 900

0 Resampling

Year

9,080 / 10,800

0 Resampling

Disease Year Statistics

- Rare 3,211/10,800
- Cancer 2,702/10,800
- Infectious 9/10,800
- PGx 396/10,800
- NCD 2,762/10,800

Medical Center Weekly Report [See all](#)

Medical Center	Samples	Resampling	%
1. โรงพยาบาลตำรวจ	0 / 2	0	0%
2. จุฬาลงกรณ์มหาวิทยาลัย	0 / 30	0	0%
3. โรงพยาบาลชลบุรี	0 / 2	0	0%
4. โรงพยาบาลศูนย์มະເຣິງ ສລຸຣີ	0 / 5	0	0%
5. กรมวิทยาศาสตร์การแพทย์	0 / 0	0	0%

National Bioresource Center

Week

Input Output

0 / 1

Month **705** / 400 Year **9,080** / 808

Sequencing Lab

Week

Input Output

1 / 144

Month **400** / 240 Year **808** / 440

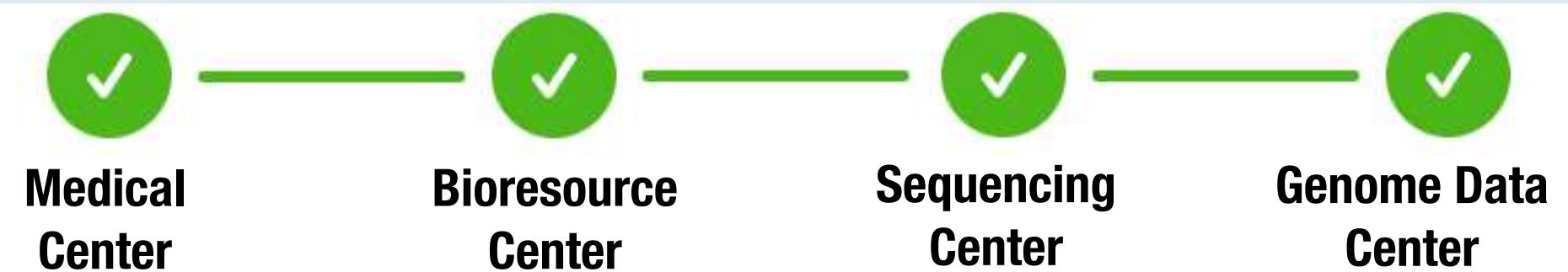
Genome Databank

Week

Input Output

144 / 0

Month **240** / 0 Year **440** / 0



ซอฟต์แวร์ช่วยระบุการกลายพันธุ์

Variant Annotation & Prioritization Platform (v@pp)



Variant Annotation and Prioritization Platform (V@PP) | Total Patient 100

Dashboard | Delegation

Rare: 70 / 40 / 20 (Pending / Ready / Analyze)

Cancer: 70 / 40 / 20 (Pending / Ready / Analyze)

PGx: 70 / 40 / 20 (Pending / Ready / Analyze)

NCD: - / - / - (Pending / Ready / Analyze)

Infectious: - / - / - (Pending / Ready / Analyze)

Proband List | All Sample | Search...

No	NBT ID	Family ID	Sex	Status	File (Download)
1	NBT-HSO1043	Fam15	♂	Pending	Phenotype VCF Actionable Report
2	NBT-HSO1007	Fam3	♂	Pending	Phenotype VCF Actionable Report
3	NBT-HSO1013	Fam5	♀	Pending	Phenotype VCF Actionable Report
4	NBT-HSO1046	Fam16	♂	Pending	Phenotype VCF Actionable Report
5	NBT-HSO1052	Fam18	♀	Pending	Phenotype VCF Actionable Report
6	NBT-HSO1049	Fam17	♀	Pending	Phenotype VCF Actionable Report
7	NBT-HSO1055	Fam19	♀	Pending	Phenotype VCF Actionable Report
8	NBT-HSO1001	Fam1	♂	Pending	Phenotype VCF Actionable Report
9	NBT-HSO1010	Fam4	♂	Pending	Phenotype VCF Actionable Report
10	NBT-HSO1040	Fam14	♂	Pending	Phenotype VCF Actionable Report
11	NBT-HSO1004	Fam2	♂	Pending	Phenotype VCF Actionable Report

Impact / Sequence ontology | Filter 904 / 1,203

Sequence ontology:

- frameshift variant
- stop lost
- inframe insertion
- splice acceptor variant
- splice donor variant
- missense variant
- stop gained
- transcript ablation
- protein altering variant
- start lost
- transcript amplification
- inframe deletion

Mode of Inheritance and QC | Filter 1,203 / 500,000

- AD
- AD_dn
- AR_comp
- AR_comp_dn
- AR_hom
- AR_hom_dn
- XD
- XD_dn
- XR
- XR_dn

Alternate Allele Frequency | Filter 835 / 904

1000Gp3_AF: 0.05 | gnomAD_exomes_AF: 0.05 | gnomAD_genomes_AF: 0.05 | TREx: 0.05 | GeTH_AF: 0.05

Variants table | Explore 2 Variants

#CHROMOM	POSCHROM	ID#CHROM	REFCHROM	ALTCHROM	QUALHROM	FILTEROM	INFOHROM	Consequence	IMPACTOM
chr10	3138915	.	G	A	184439.66	PASS	0.123000003397	missense_variant	MODERATE
chr10	3138915	.	G	A	184439.66	PASS	0.123000003397	missense_variant	MODERATE

INTEGRATIVE GENOMICS VIEWER (IGV)

Genes: GRP42 CSH1

IGV hg38 chr19 chr19:35,372,233-35,372,297 65 bp

Refseq Genes

Proband (HS02067)

Father (HS02068)

Mother (HS02069)

ซอฟต์แวร์ด้านเภสัชพันธุศาสตร์



Variant
Annotation and Prioritization
Platform (V@PP)



Total
Patient 100

Dashboard Delegation

Rare

70 / 40 / 20
Pending / Ready / Analyze

Cancer

70 / 40 / 20
Pending / Ready / Analyze

PGx

70 / 40 / 20
Pending / Ready / Analyze

NCD

- / - / -
Pending / Ready / Analyze

Infectious

- / - / -
Pending / Ready / Analyze

No	NBT ID	Sex	Status	Date analysed	File (Download)
1	NBT-HS01005	♂	Completed	05/01/2022	Actionable Guidelines HLA Variant Effect Clinical Annotation SV/CNV
2	NBT-HS01003	♀	Completed	05/01/2022	
3	NBT-HS01042	♀	Completed	05/01/2022	
4	NBT-HS01041	♂	Completed	05/01/2022	
5	NBT-HS01043	♂	Completed	05/01/2022	
6	NBT-HS01044	♂	Completed	05/01/2022	
7	NBT-HS01007	♂	Completed	05/01/2022	
8	NBT-HS01013	♀	Completed	05/01/2022	
9	NBT-HS01046	♂	Completed	05/01/2022	
10	NBT-HS01047	♂	Completed	05/01/2022	
11	NBT-HS01045	♀	Completed	05/01/2022	

Variant
Annotation and Prioritization
Platform (V@PP)

Total
Patient 100

Dashboard Delegation

Actionable Guidelines
Research

Sample ID: HS01005 | NBT ID: 12345678

HLA Variant effect Clinical annotation SV/CNV

Filtering options:

Label drugs related to diagnosed disease

HLA genes:

Drug causing ADRs:

Cohort ethnicities:

Patient diseases:

ADRs:



Variant
Annotation and Prioritization
Platform (V@PP)



Total
Patient 100

Dashboard Delegation

Sample ID: HS01005 | NBT ID: 12345678

Actionable Guidelines Research

All guidelines NLEM panel Anticancer panel

Filtering options:

Report only best candidate allele Label drugs related to diagnosed disease

Phenotypes:

Intermediate Metabolizer Normal Metabolizer Indeterminate Poor Metabolizer

Rapid Metabolizer Ultra Rapid Metabolizer

Guideline Strengths:

Strong Moderate Optional No recommendation

Gene list:

Drug list:

Output

Exported report:

Summary Detail

Selected Report view:

Gene-view Drug-view

Gene	Phased status	# Missing positions / # Total...	# Matched diplotypes	Matched diplotypes
CYP2C9,CYP4F2,VKORC1	Unphased	2/53,0/2,0/1	1	*1/*1,*1/*1,-1639A/-1639G
CYP2B6	Unphased	0/35	1	*6/*6
CYP2C19	Unphased	0/34	1	*3/*3
CYP2C19	Unphased	0/34	1	*3/*3
CYP2C19	Unphased	0/34	1	*3/*3
CYP2C19	Unphased	0/34	1	*3/*3
CYP2C19,CYP2D6	Unphased	0/34,0/0	1	*3/*3,*10/*2
CFTR	Unphased	5/54	1	No CPIC variants found
CYP2C19,CYP2D6	Unphased	0/34,0/0	1	*3/*3,*10/*2
CYP2C19,CYP2D6	Unphased	0/34,0/0	1	*3/*3,*10/*2

Gene	Drug name	Diplotype	Phenotype	Recommendation	Guideline strength
CYP2C9,CYP4F2,VKORC1	warfarin	*1/*1,*1/*1,-1639A/-1639G	Poor Metabolizer	<text>See dosing guideline ...	N/A
CYP2B6	efavirenz	*6/*6	Poor Metabolizer	<text>Consider initiating ef...	Moderate
CYP2C19	citalopram, escitalopram	*3/*3	Poor Metabolizer	<text>Consider a 50% reduc...	Moderate
CYP2C19	clopidogrel	*3/*3	Poor Metabolizer	<text>Alternative antiplatelet...	Strong
CYP2C19	sertraline	*3/*3	Poor Metabolizer	<text>Consider a 50% reduc...	Optional
CYP2C19	voriconazole	*3/*3	Poor Metabolizer	<text>For pediatric or adult...	Moderate
CYP2C19,CYP2D6	doxepin	*3/*3,*10/*2	Poor Metabolizer,Normal M...	<text>Avoid amitriptyline us...	Optional
CFTR	ivacaftor	No CPIC variants found		<text>This guideline does n...	N/A
CYP2C19,CYP2D6	clomipramine	*3/*3,*10/*2	Poor Metabolizer,Normal M...	<text>Avoid amitriptyline us...	Optional
CYP2C19,CYP2D6	amitriptyline	*3/*3,*10/*2	Poor Metabolizer,Normal M...	<text>Avoid amitriptyline us...	Moderate

ตรวจการดื้อยาของเชื้อวัณโรค

<https://mtb.nbt.or.th>



The image shows a screenshot of the Tuberculosis NBT system. On the left is a blue sidebar with the title 'TUBERCULOSIS' and navigation options: Home, Account Summary, and Dashboard. The main area is titled 'All Project' and displays a grid of project cards. Each card shows project details (e.g., 'Project : Test', 'Total Sample : 3'), NGS Analysis status (with progress bars), Phylogenetic status, and DST results (with color-coded bars for sensitive, resistant, and MDR). A vertical blue navigation bar is overlaid on the project cards.

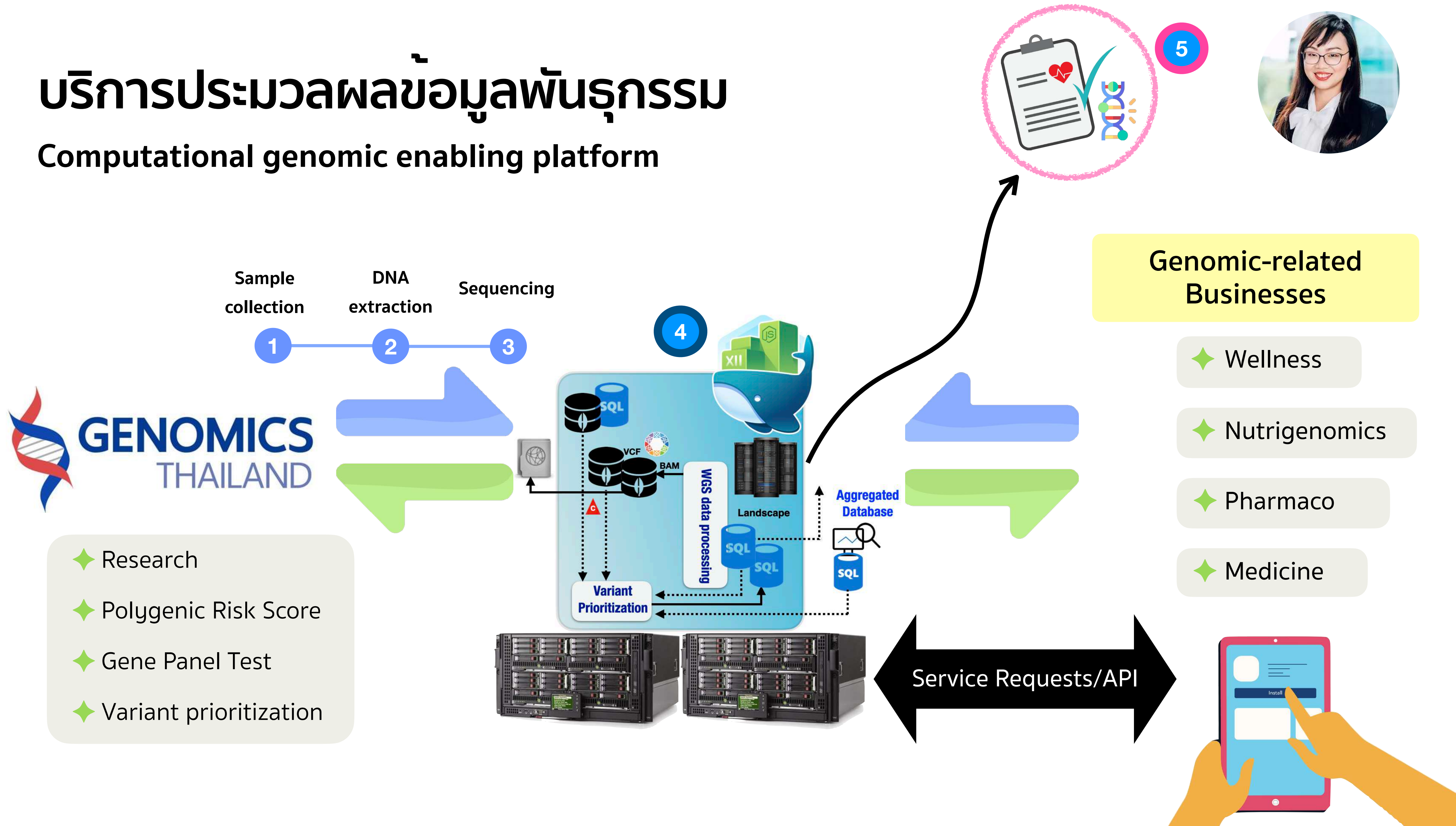
On the right, a 'Result' page is shown for a 'Mycobacterium Whole Genome Sequencing Report'. It includes the following information:

- Specimen ID : ERR718196
- Report Date : 08 September 2021
- Platform : Illumina
- Summary of Finding : The specified from xxxx is positive for Mycobacterium tuberculosis (lineage1, lineage1.1.1.5). It is predicted to be resistant to 0 antibiotics ()
- Diagnosis : The specified was speciated as Mycobacterium tuberculosis.
- Treatment Details : The specified was considered to be susceptible
- Drug Resistance database : TBprofiler
- Lineage database : Thai SNP DB

Drug	Prediction	Mutation	Resistant confident	Resistant Level
Isoniazid	No mutation detected			
Rifampicin	No mutation detected			
Ethambutol	No mutation detected			
Pyrazinamide	No mutation detected			
Streptomycin	No mutation detected			
Fluoroquinolones				
Ofloxacin	No mutation detected			
Levofloxacin	No mutation detected			
Moxifloxacin	No mutation detected			
Aminoglycosides				
Amikacin	No mutation detected			
Kanamycin	No mutation detected			
Capreomycin	No mutation detected			

บริการประมวลผลข้อมูลพันธุกรรม

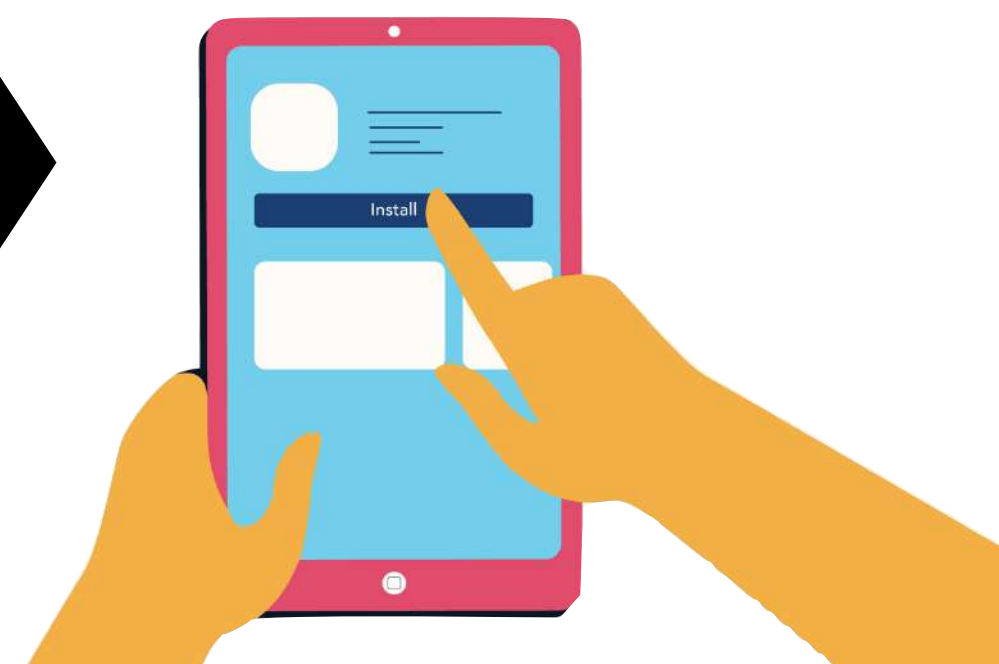
Computational genomic enabling platform



Genomic-related Businesses

- Wellness
- Nutrigenomics
- Pharmaco
- Medicine

- Research
- Polygenic Risk Score
- Gene Panel Test
- Variant prioritization



กิตติกรรมประกาศ



กรมวิทยาศาสตร์การแพทย์
Department of Medical Sciences



คณะแพทยศาสตร์
FACULTY OF MEDICINE
Chulalongkorn University



Mahidol University
Faculty of Medicine
Siriraj Hospital



Mahidol University
Faculty of Medicine Ramathibodi Hospital



MDKKU
คณะแพทยศาสตร์ มหาวิทยาลัยขอนแก่น
FACULTY OF MEDICINE KHON KAEN UNIVERSITY

