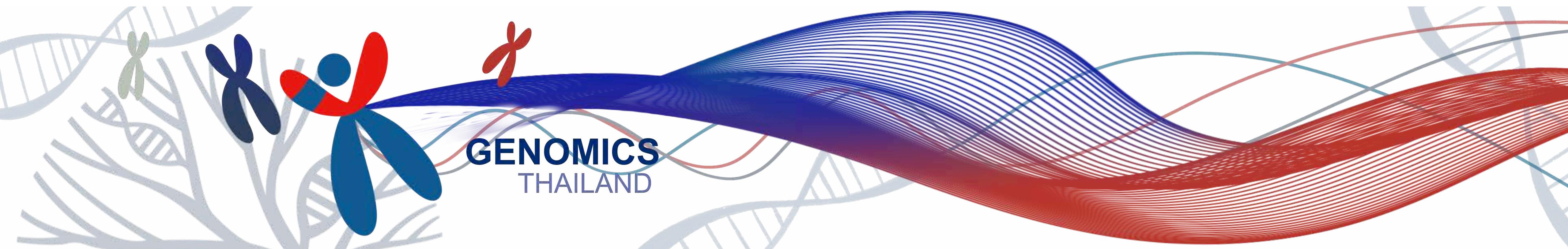


Toward Precision Medicine: Genomics Thailand Project

Worawich Phornsiricharoenphant

PhD student, Computational Oncogenomics Group

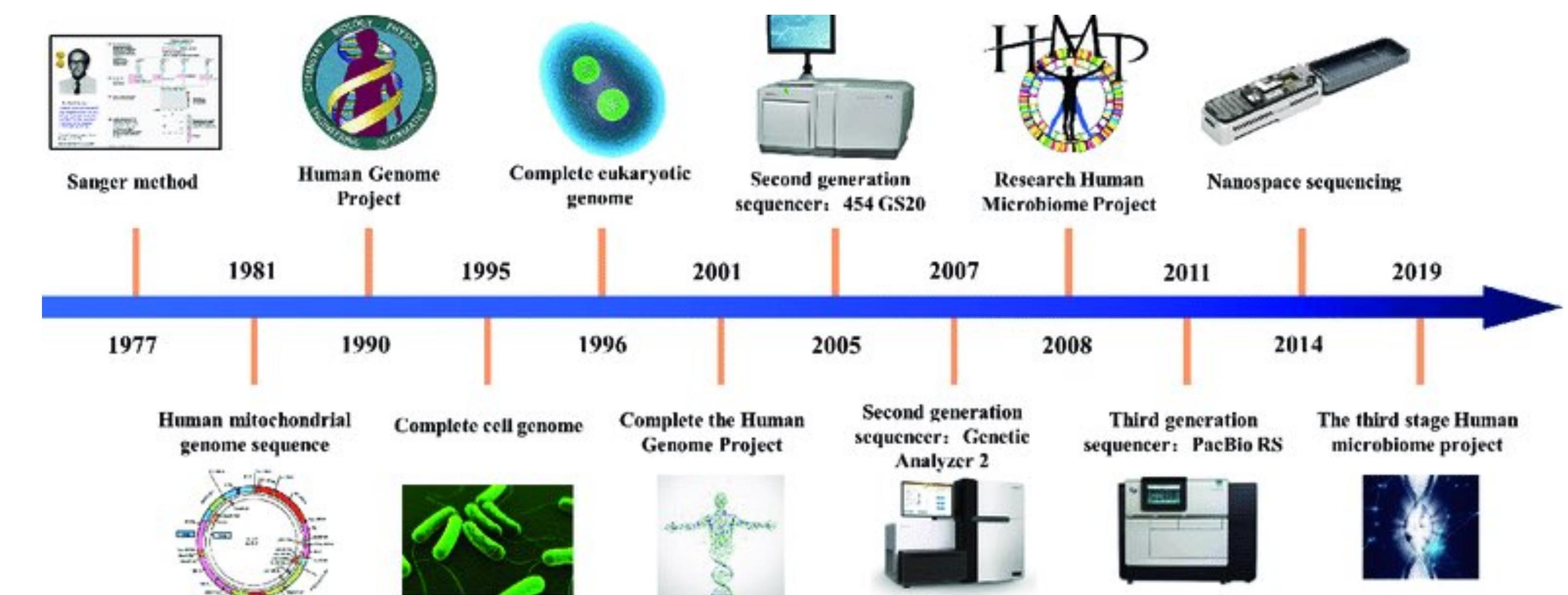
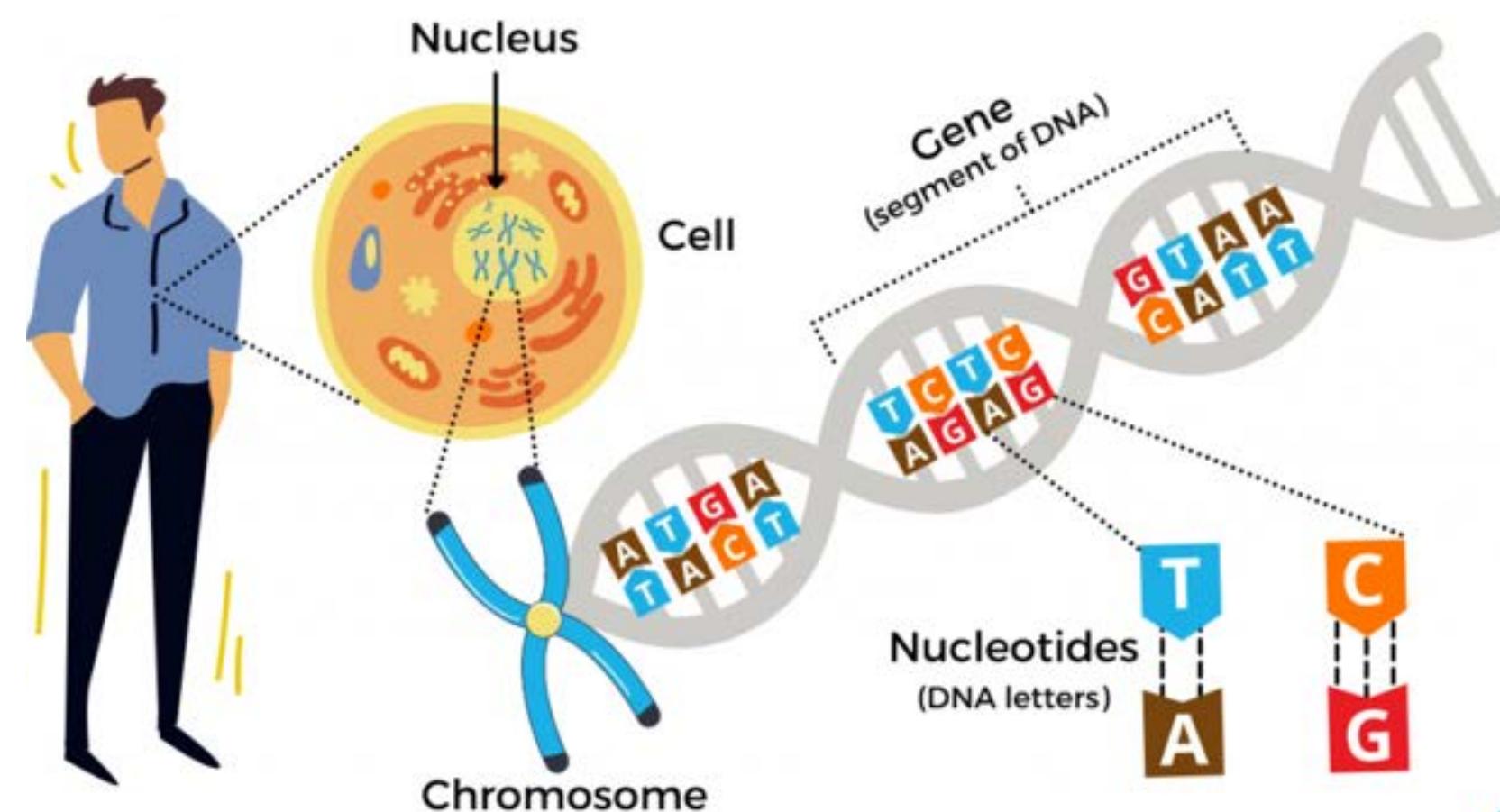
University of Zurich



View from “genomics”

Genomic medicine

- **Prediction:** ↗ gather variants ↗ evaluate genetic profile
- **Diagnosis:** ↗ gather variants ↗ prioritize them ↗ report the mutation(s)
- **Treatment:** ↗ accurate diagnosis ↗ PharmGx ↗ prescription



Genomics Thailand Phase I

Here you go!

Thailand large-scale population genomic initiative

- Aims to construct a national ref variation database of 50,000 Thais and focus on **RUD, ONC, PGx, NCD and EID**
- Starting off as a **consortium** of 6 universities and 6 governmental organizations (March 2018)
- **Health Systems Research Institute (HSRI)** allocates budget for research and infrastructure



Genomics Thailand

GeTH 50K (2022-2025)



- ▶ ₩4.47 billion project (26 March 2019) (112 million CHF)
- ▶ 50K **population genomics**: 5yr-project
- ▶ 30-40x NGS using DNBSeq T7 @ Burapha U.
- ▶ ₩9700/sample : ~120 GB/file (243 CHF/sample)
- ▶ 5 disease-groups:

Rare and undiagnosed
Oncogenomics
Pharmacogenomics

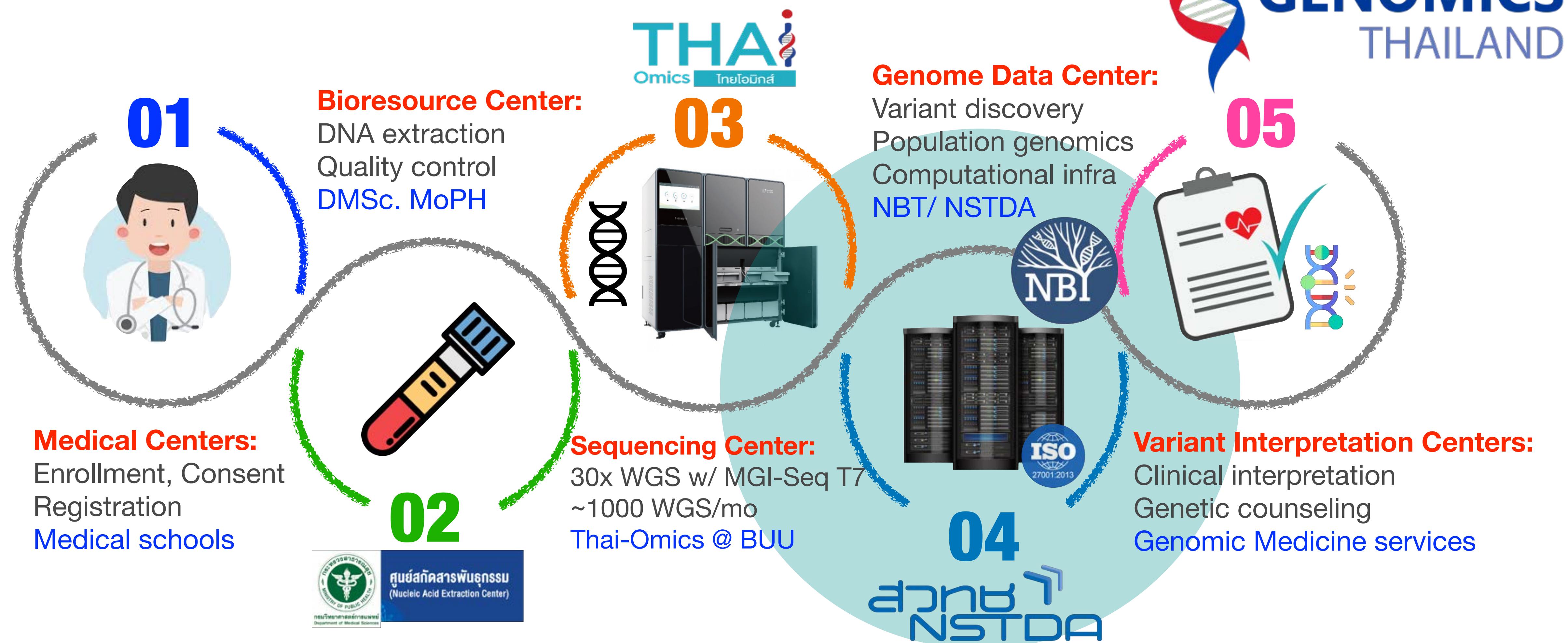
Non-communicable
Infectious



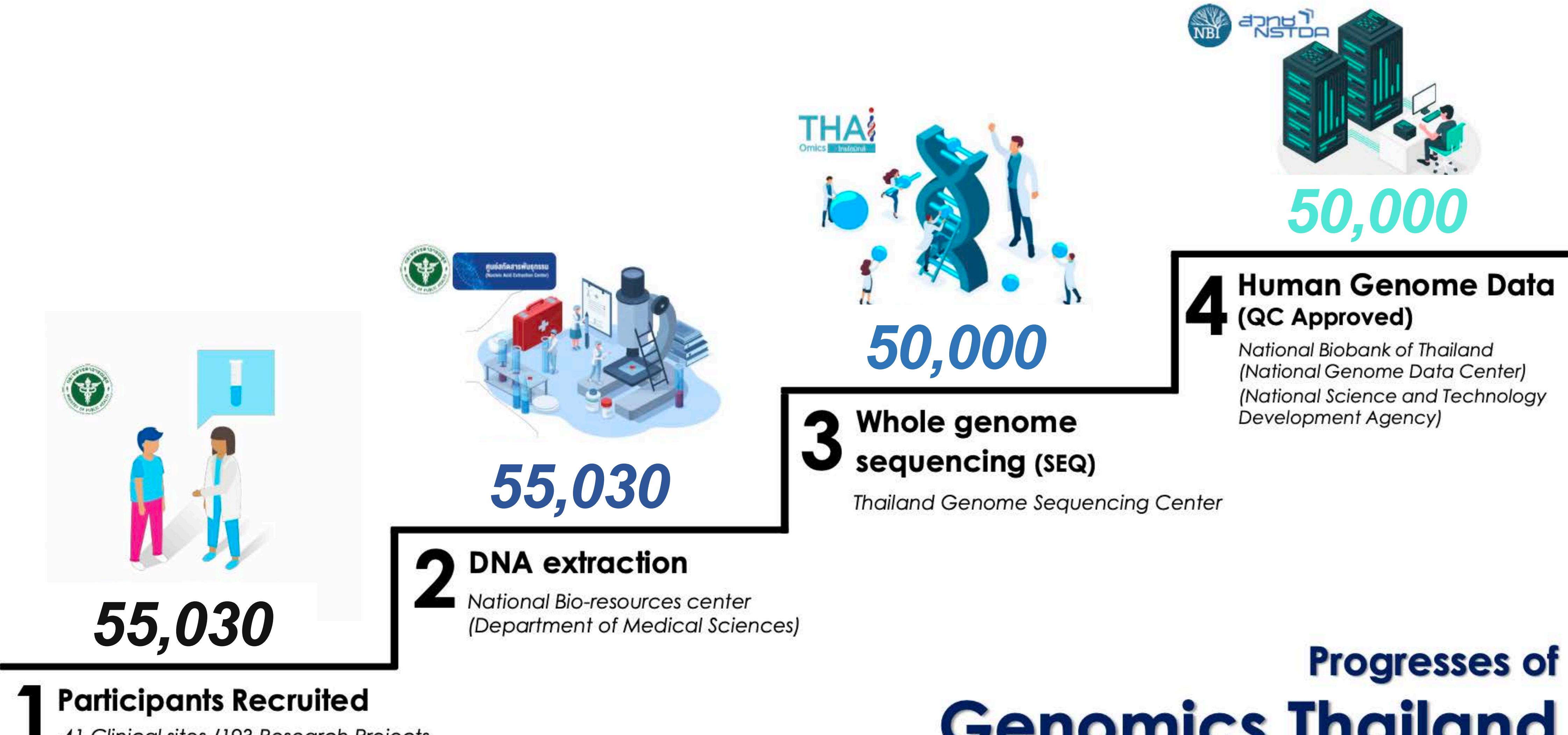
กองวิทยาศาสตร์การแพทย์
Department of Medical Sciences

Large-scale WGS

50,000 whole genome sequencing



How far we've come



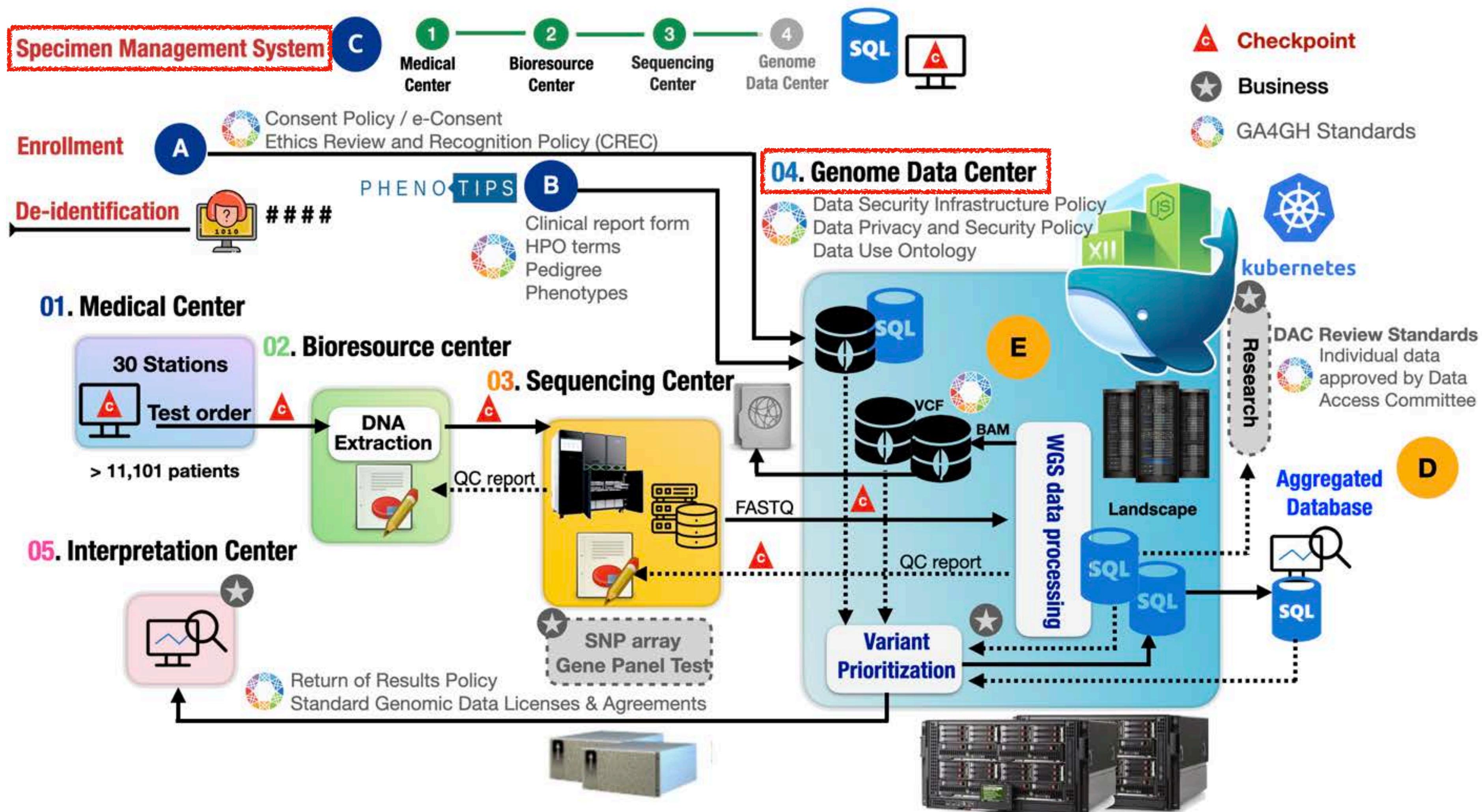
Progresses of Genomics Thailand

(As of 31 July 2025)

Behind the scenes

The grand scheme

Computational Genomic Infrastructure



Management Software

- Enrollment & e-consent
- Clinical database
- PHENOTIPS**
- Specimen Mangement System (SMS)

Data Processing & Analytic Platform

- Whole Genome Sequence Data QC
- Accelerated WGS processing
- Variant Annotation and Prioritization (V@PP)
- PharmVIP
- Population genomics platform
- T-Rex Database
- Thai Reference Genome dB
- MDR Tuberculosis platform
- Amplicon Metagenomic platform

Very big data indeed

Three billion basepairs



A vertical stack of icons. At the top is a black stopwatch icon. Below it is a black DNA double helix icon. To the right of the DNA icon is a black silhouette of a person. At the bottom left is a large red 'X' mark, and to its right is the number '40' in a large, dark gray font.

48 indv. / 36 hours

\sim 120 GE



A black icon depicting a family unit. It features three stylized human figures: two adults standing on either side of a smaller child figure. The figures are represented by simple black outlines and filled shapes.

50,000 indv.

> 6 PB



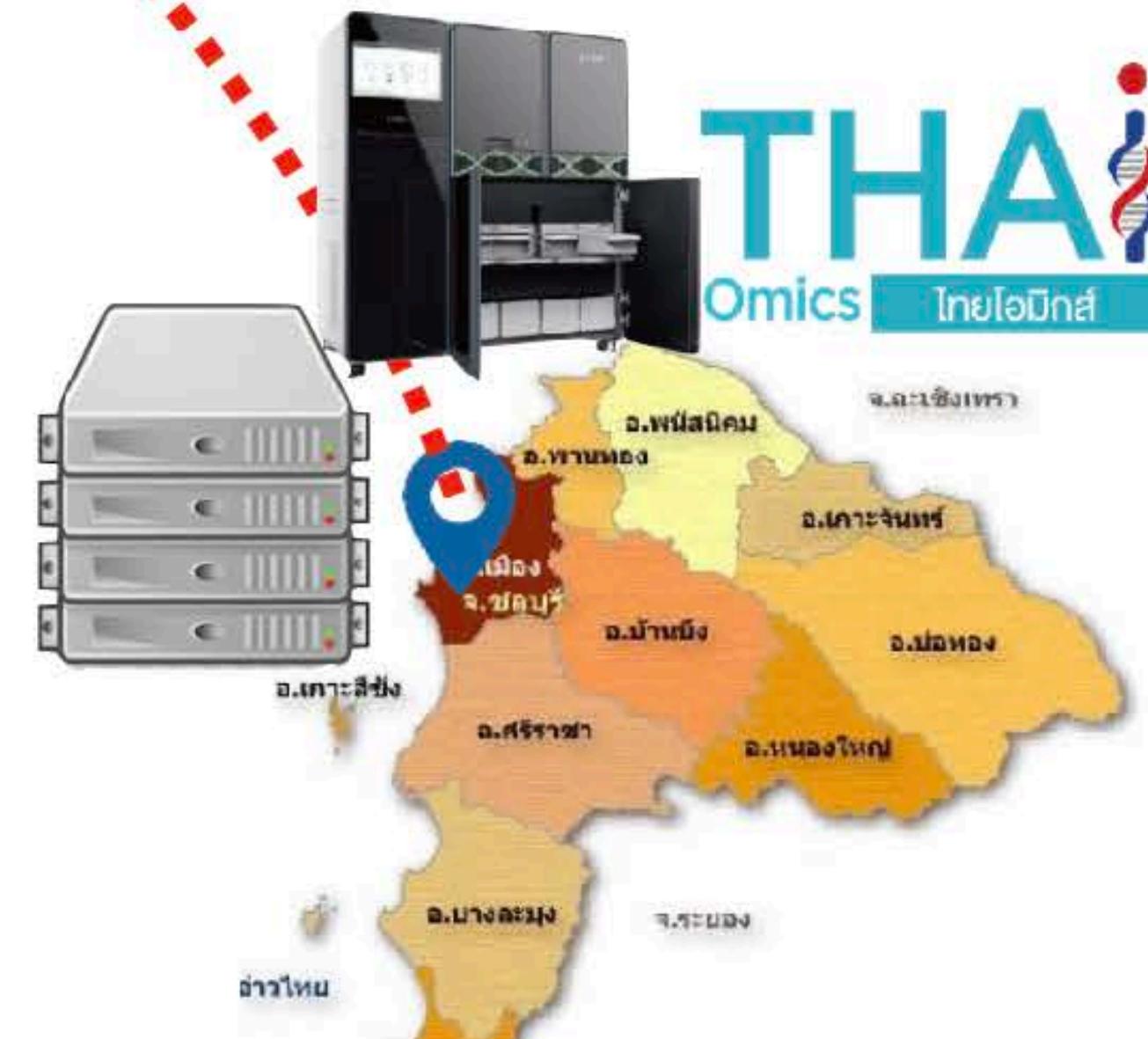
129 m

The journey of WGS

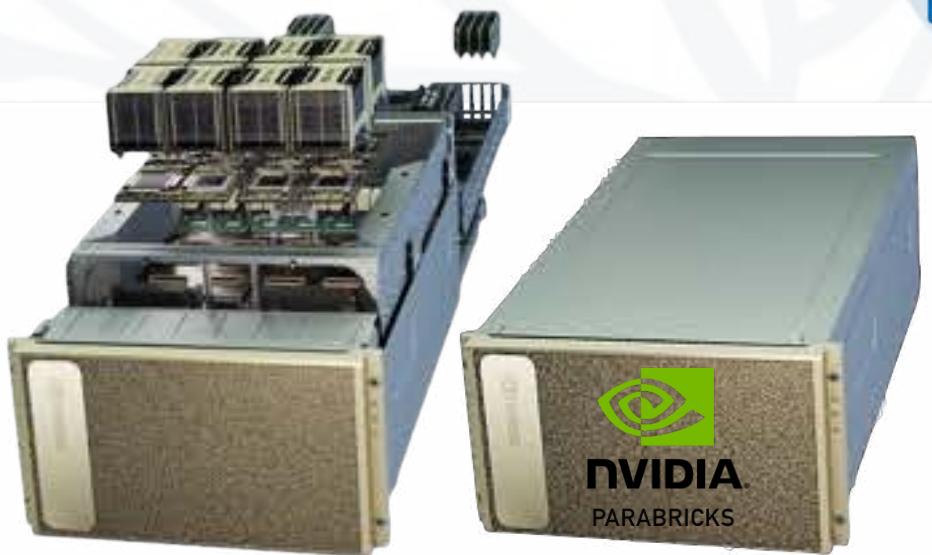
Chonburi to Pathum Thani



UniNet
48 WGS (5TB) ~ 2.15 h
134 Km.



HPC : Hardware Infrastructure



2x DGX A100 + Parabricks



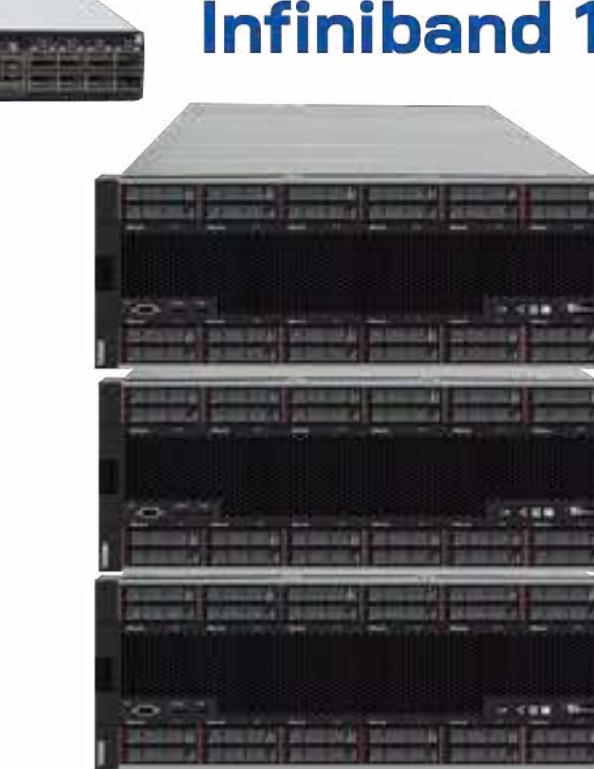
**18x Web/Database servers
3x Frontend/Transfer servers**



NSTDA Data Center : ISO/IEC 27001:2013



**Storage 7500 TB
for huge genomics data**



9x Fat nodes

- 2x 192 cores/3TB Mem
- 7x 224 cores/3TB Mem/
12 NVMe



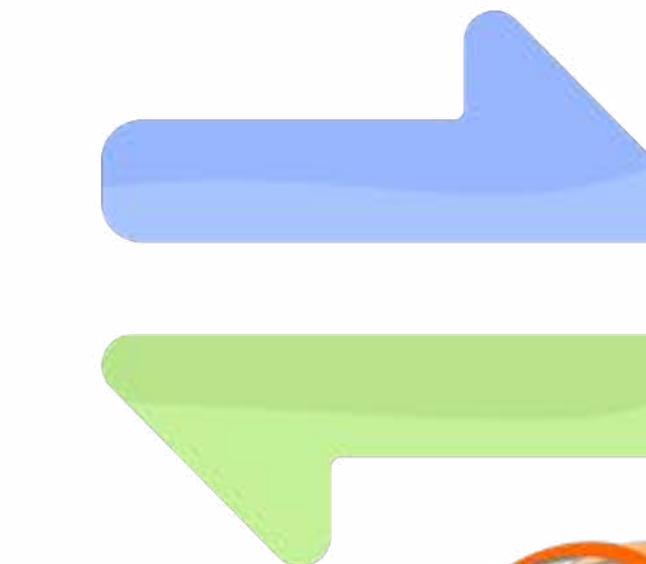
180x LTO9 (18 TB)

**Expand
system capacity &
computational**

Software IBM Aspera 500 Mbps



rsync with multi-sessions



10 Gbps direct link



**2x Fat node
(224 cores/3 TB Mem/
12 TB NVMe)**



**2x Buffer Storage
500 TB**

Specimen Management System

<https://sms.genomicsthailand.com>

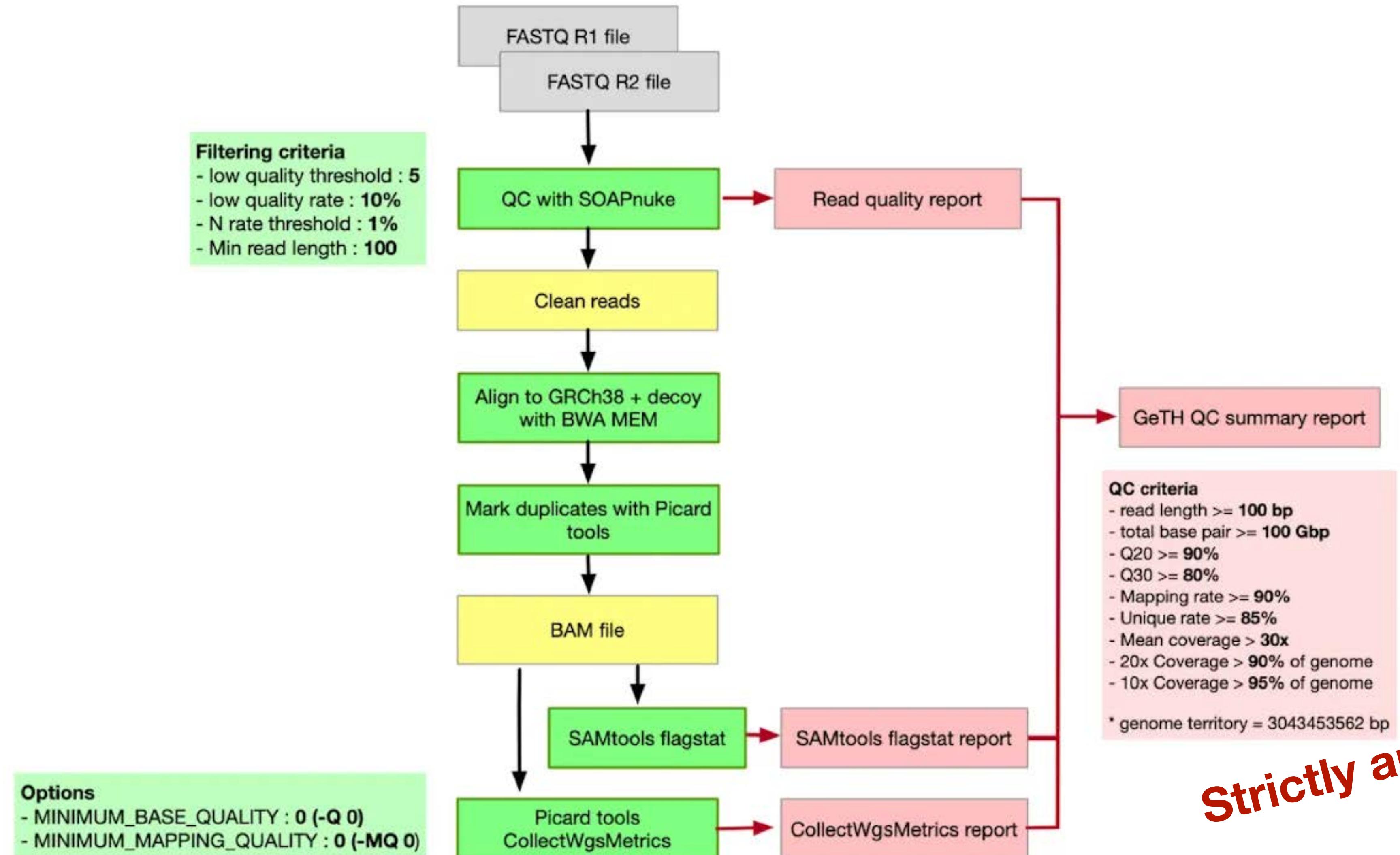
The dashboard features a top navigation bar with links for Specimen Management System, Genome Databank (4), Dashboard, Activities, Samples, Members, Medical Centers, and a SuperAdmin dropdown. Below the navigation is a summary section for Today (18 January, 2024) showing 0 samples input and 0 output. To the right is a Disease Statistics chart with categories: Cancer (61%), Rare (114%), PGx (20%), NCD (54%), Infectious (17%), and Popgen (0%). A Sample Statistics chart shows daily sample counts from Jan 3 to Jan 19. The main area displays a Sample List table with columns: No, NBT ID, Project ID, Sampling Date, Medical Center, DNA Status, Sample Status, and Progress (with five green circles).

No	NBT ID	Project ID	Sampling Date	Medical Center	DNA Status	Sample Status	Progress
1	T0000001	00-000	24:00 2/23/2022	กรุงเทพมหานคร	Completed	Completed	1 2 3 4 5
2	T0000002	00-000	24:00 2/23/2022	กรุงเทพมหานคร	Completed	Completed	1 2 3 4 5
3	T0000003	00-000	24:00 2/23/2022	กรุงเทพมหานคร	Completed	Completed	1 2 3 4 5
4	T0000004	00-000	24:00 2/23/2022	กรุงเทพมหานคร	Completed	Completed	1 2 3 4 5
5	T0000005	00-000	24:00 2/23/2022	กรุงเทพมหานคร	Completed	Completed	1 2 3 4 5
6	T0000006	00-000	24:00 2/23/2022	กรุงเทพมหานคร	Completed	Completed	1 2 3 4 5
7	T0000007	00-000	24:00 2/23/2022	กรุงเทพมหานคร	Completed	Completed	1 2 3 4 5
8	T0000008	00-000	24:00 2/23/2022	กรุงเทพมหานคร	Completed	Completed	1 2 3 4 5

- Handshake between stations.
- Progress tracking.
- Sample record database.

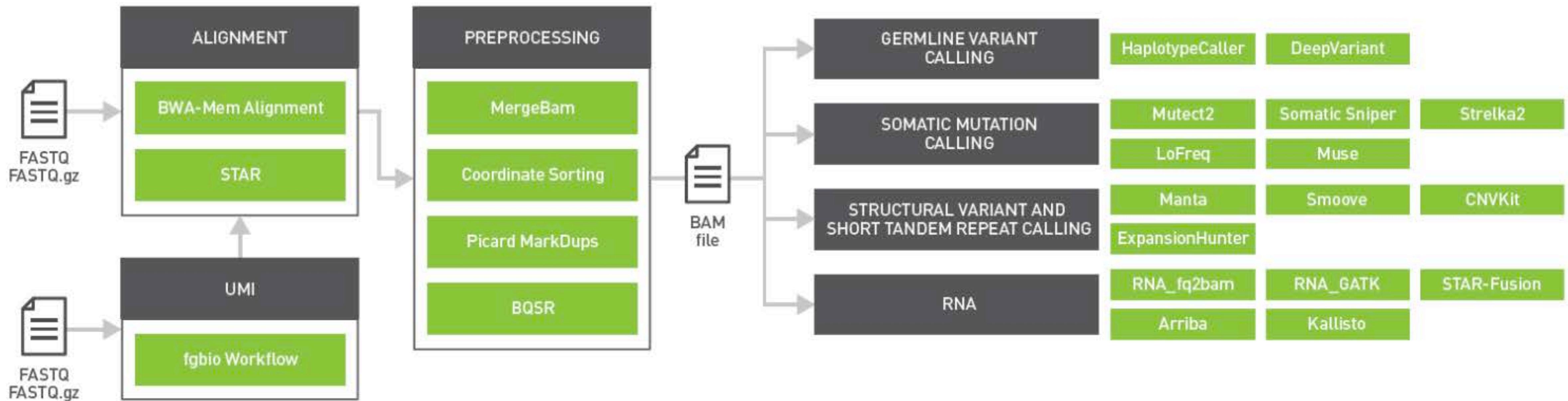
WGS Quality Control

Ensure high quality sequencing data



Strictly applied criteria

Variant Calling

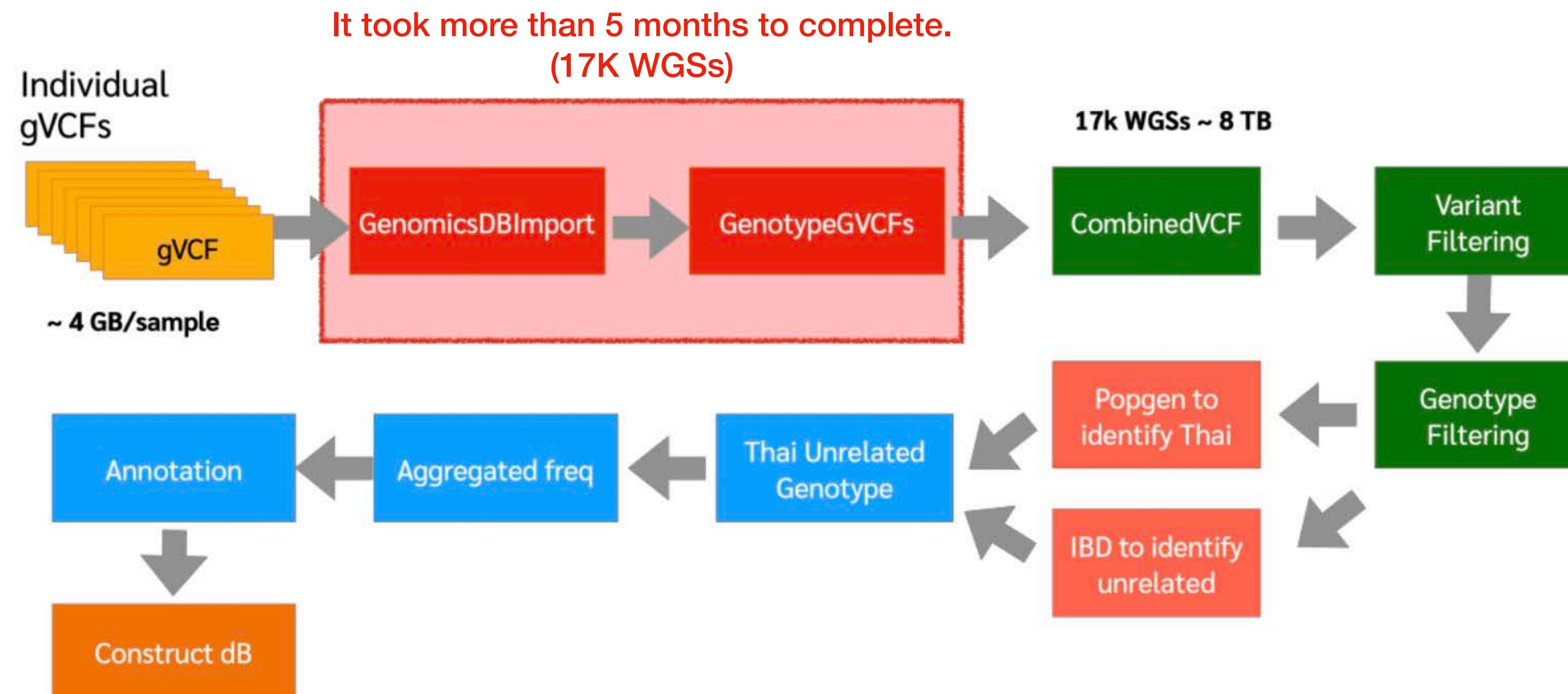


GPU accelerated WGS analysis
(Alignment + Variant calling)

- ~ 1 hour per sample (4 GPU)
- ~ 48 sample per day per DGX A100



Variant Aggregation



Thai Variant Annotation and Prioritization (v@pp)

The screenshot displays the Thailand Variant Annotation and Prioritization Platform (v@pp) interface. At the top, a blue header bar includes icons for Data Management, Dashboard, Delegation, Activities, Statistics, Management, and a gear icon. Below the header, a navigation bar offers links to GeTHlyzer, Discovery, Selected, Predicted, GEL, GeneSet, ACMG, and a back button.

The main content area shows two variant entries:

- Variant 1:** chr7_94426514_G_A, COL1A2, Heterozygous. ACMG Classification: PM1, PM2, PMS, PP3, PPS. ACMG Prediction: likely pathogenic. Probability of Pathogenicity: high. Feature: NM_000089.4; Variant: c.3089G>A; Protein: p.Gly1030Asp; Impact: MODERATE; Consequence: missense_variant; Maneuver: YES.
- Variant 2:** chr6_146027829_G_T, GRM1, Heterozygous. ACMG Classification: PVS1. ACMG Prediction: vus. Probability of Pathogenicity: low. Feature: NM_000089.4; Variant: c.3089G>A; Protein: p.Gly1030Asp; Impact: MODERATE; Consequence: missense_variant; Maneuver: YES.

To the right, a dashboard provides summary statistics:

Total	Rare	Cancer	PGx	NCD	Infectious	Popgen
55,109	19,158	9,726	3,388	15,823	4,419	2595

It also shows resource utilization for the Medical Center, National Bioresource Center, Sequencing Lab, and Genome Databank.

At the bottom, an IGV track displays genomic data for chromosome 7, highlighting a specific variant at position 94,426,573. A callout box provides detailed sequencing statistics for this position.

<https://vapp.genomicsthailand.com>

Return WGS results

The screenshot shows the Data Management interface. At the top, there are navigation links: Dashboard, Delegation, Activities, Statistics, Management, and a user icon. Below the header, a search bar shows '12,753 Samples' and a search input field. A table lists samples with columns: No, NBT ID, Disease, Project ID, Create Date, and a 'View BAM' button. The 'View BAM' button for sample 1 (NBT ID 25136151) is highlighted with a blue circle and a blue arrow pointing to a BAM viewer window. The BAM viewer window displays a genomic track for chromosome 124,479,127-124,479,29. The main interface also features a 'File (Download)' section with a list of files (GVCF, VCF, Excel) and a 'Share' section with a list of users.

No	NBT ID	Disease	Project ID	Create Date	
1	25136151	Rare	63-080	06/05/2022	View BAM
2	43959867	Rare	63-096	06/05/2022	
3	75177387	Rare	63-096	06/05/2022	
4	76049212	Rare	00-001	06/05/2022	

Data management

- PI/Doctor can view/download their own samples
- BAM must view online
- GVCF, Filtering VCF, Annotated Excel can download **5** samples per time
- Use OTP (8 hrs / OTP)
- Owner can share permission to the others
- The granted users can access with time limit

Rare : Ranked associated/pathogenic variants

The screenshot shows a clinical informatics interface. At the top, there are three main categories: 'Rare' (with 7 samples), 'Cancer' (with 7 samples), and 'PGx' (with 7 samples). Below these are tabs for 'Sample List' and 'Disable Family'. The 'Sample List' table includes columns for NBT ID, Family ID, Family, Gender, Life Status, Project ID, and Disorders. A row for sample '7' is highlighted with a red box. To the right, a detailed view of sample '7' is shown, including phenotypic data: Disorders (MIM:259420, Label: OSTEOGENESIS IMPERFECTA, TYPE III), Genes (ENSG00000164692, Gene: COL1A2, Strategy: sequencing, Status: candidate), and HPO terms (HP:0000592, Label: Blue sclerae; HP:0000938, Label: Osteopenia; HP:0004322, Label: Short stature).

Results from TAPES
ACMG Classification

The screenshot shows the 'GeTHlyzer' tool interface. A pink arrow points from the highlighted sample '7' in the previous screenshot to the 'Ranked variants' section here. The 'Ranked variants' section displays three variants (1, 2, 3) with their ACMG classification, OMIM, and ClinVar details. Variant 1 is highlighted with a yellow box.

Sample	Gene	Variant	ACMG Classification	OMIM	ClinVar Submitter(s)
1	COL1A2	c.3089G>A	PMS1, PMS2, PMD, PP3, PPD	OMIM: 259420	Pathogenic (1), Benign (0)
2	GRHL1	chr6_146027629_G_T	PVS1	gnomADg, ThAdg	
3	CASR	chr3_122261891_C_T	PMS1, PMS2, PMD, PP3	OMIM: 259420	Pathogenic (0), Benign (0)

Variants prioritization in Cancer

This screenshot shows a software interface for variant prioritization. On the left, there's a 'Sample List' table with columns for NBT ID, Family ID, Family, Gender, Life Status, and Project. A red arrow points from the 'Family' column towards the right panel. On the right, there are two main sections: 'Rare' and 'Cancer'. The 'Cancer' section is highlighted with a red box and contains a sub-section for 'Pending / Ready / Analyze'. Below these sections is a 'Genes List (182)' table.

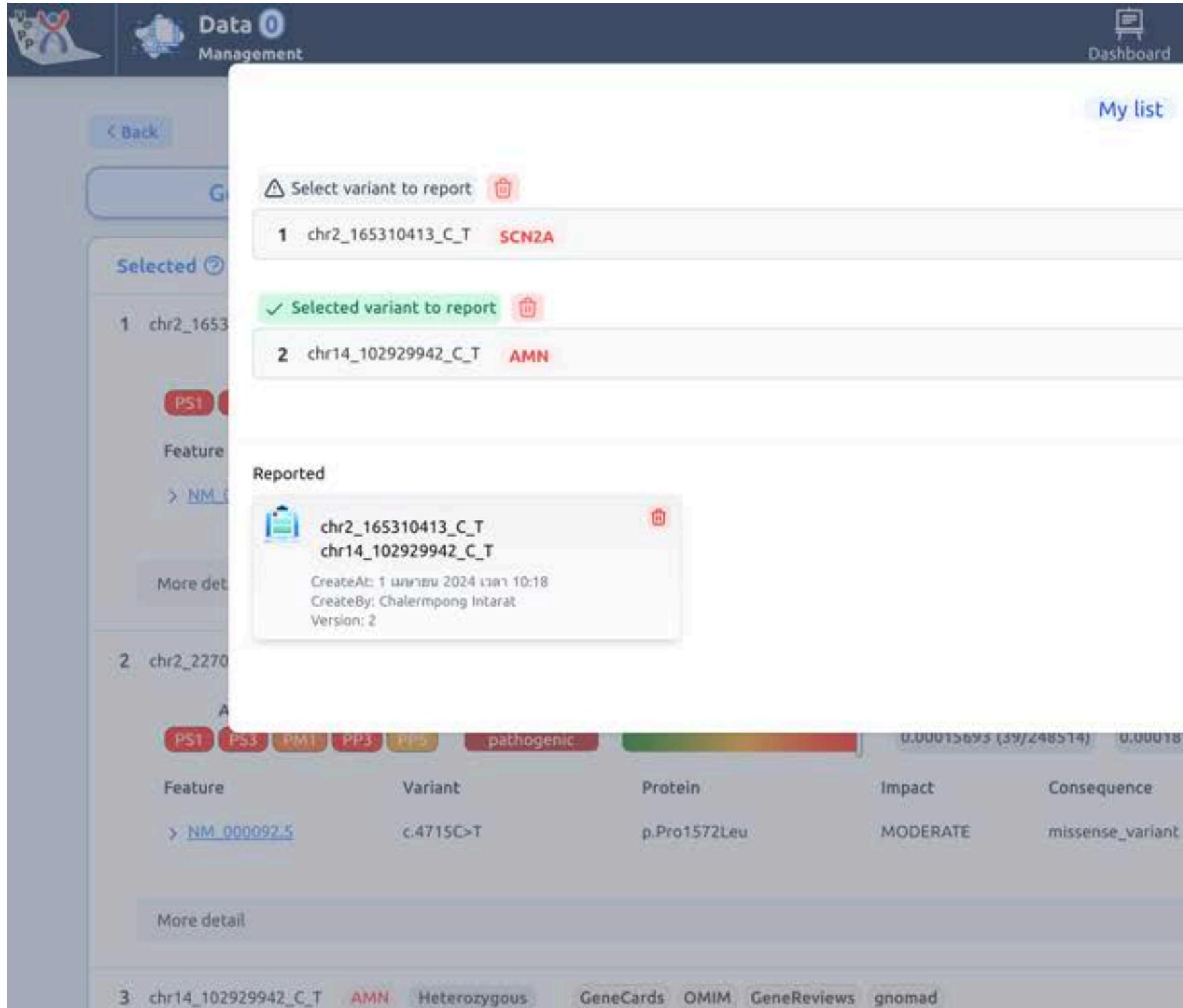
No	NBT ID	Family ID	Family	Gender	Life Status	Project
1	89103524	-	-	♀	-	00
2	01389206	-	-	♀	-	63
3	01821919	-	-	♀	-	00
4	02684126	-	-	♀	-	00
5	03947815	-	-	♂	-	63
6	04779110	-	-	♀	-	63
7	04898509	-	-	♀	-	00
8	09430262	-	-	♀	-	00
9	11708705	-	-	♀	-	63
10	10253208	-	-	♀	-	63
11	13212611	-	-	♀	-	63
12	09450947	-	-	♂	-	63
13	13540646	-	-	♀	-	63
14	13971244	-	-	♂	-	00
15	14189818	-	-	♂	-	63
16	16129198	-	-	♀	-	00
17	17720466	-	-	♂	-	63-115
18	18120417	-	-	♀	-	63-099
19	17997251	-	-	♂	-	00-000
20	18240149	-	-	♀	-	63-142
21	19770540	-	-	♀	-	63-000

This screenshot shows the 'GeTHlyzer' tool interface. At the top, it displays 'NBTID: 89103524' and tabs for 'GEL', 'GeneSet', and 'ACMG'. The 'GEL' tab is selected. The interface is divided into three levels: Level 2 (0), Level 3 (0), and Level 4 (17). The Level 3 section is highlighted with a red box and contains the text '182 Genes from GEL Cancer Panel'. Below this is a 'Genes List (182)' table. The bottom section shows detailed variant analysis for three specific variants:

- Variant 1:** chr2_127271354_C_T, ERCC3, Heterozygous. ACMG Classification: PM1, PM2, PP3. ACMG Prediction: VUS. Probability of Pathogenicity: 0.00000398 (1/251442). Feature: NM_000122.2. Variant: c.1927G>A. Protein: p.Val643Met. Impact: MODERATE. Consequence: missense_variant. Mane Select: YES. Predicted Deleterious: 12 (Pathogenic) / 5 (Benign). ClinVar Submitter(s): 0.
- Variant 2:** chr5_34937582_C_T, DNAJC21, Heterozygous. ACMG Classification: PM1, PM2. ACMG Prediction: VUS. Probability of Pathogenicity: 0.00004786 (12/250736). Feature: NM_001012339.3. Variant: c.695C>T. Protein: p.Ala232Val. Impact: MODERATE. Consequence: missense_variant. Mane Select: YES. Predicted Deleterious: 1 (Pathogenic) / 12 (Benign). ClinVar Submitter(s): 0.
- Variant 3:** chr3_10149931_A_G, VHL, Heterozygous. ACMG Classification: PM1, PM2, PP3. ACMG Prediction: VUS. Probability of Pathogenicity: 0.00000683 (1/143320). Feature: NM_000337.1. Variant: c.100G>A. Protein: p.Thr33Asn. Impact: MODERATE. Consequence: missense_variant. Mane Select: YES. Predicted Deleterious: 1 (Pathogenic) / 12 (Benign). ClinVar Submitter(s): 0.

The interface also includes buttons for 'varsome', 'IGV Compound', 'IGV', 'Add', and 'Hide'.

Report Variant



The screenshot shows the 'Data Management' interface with a 'Selected' list containing two variants: 'chr2_165310413_C_T SCN2A' and 'chr14_102929942_C_T AMN'. A modal window titled 'Report Version!' is open, prompting the user to 'Please select report version'. Below the modal, three report templates are displayed:

- version1 (limit 2 variant)**: Shows a form for reporting two variants, with the first variant filled in.
- version2 (unlimit variant) default**: Shows a more detailed form designed for reporting multiple variants.
- version3 (Chula Report)**: Shows a simplified report template with sections for patient information, clinical diagnosis, and results summary.

Cancel **Confirm**



Pharmacogenomics: PGx



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

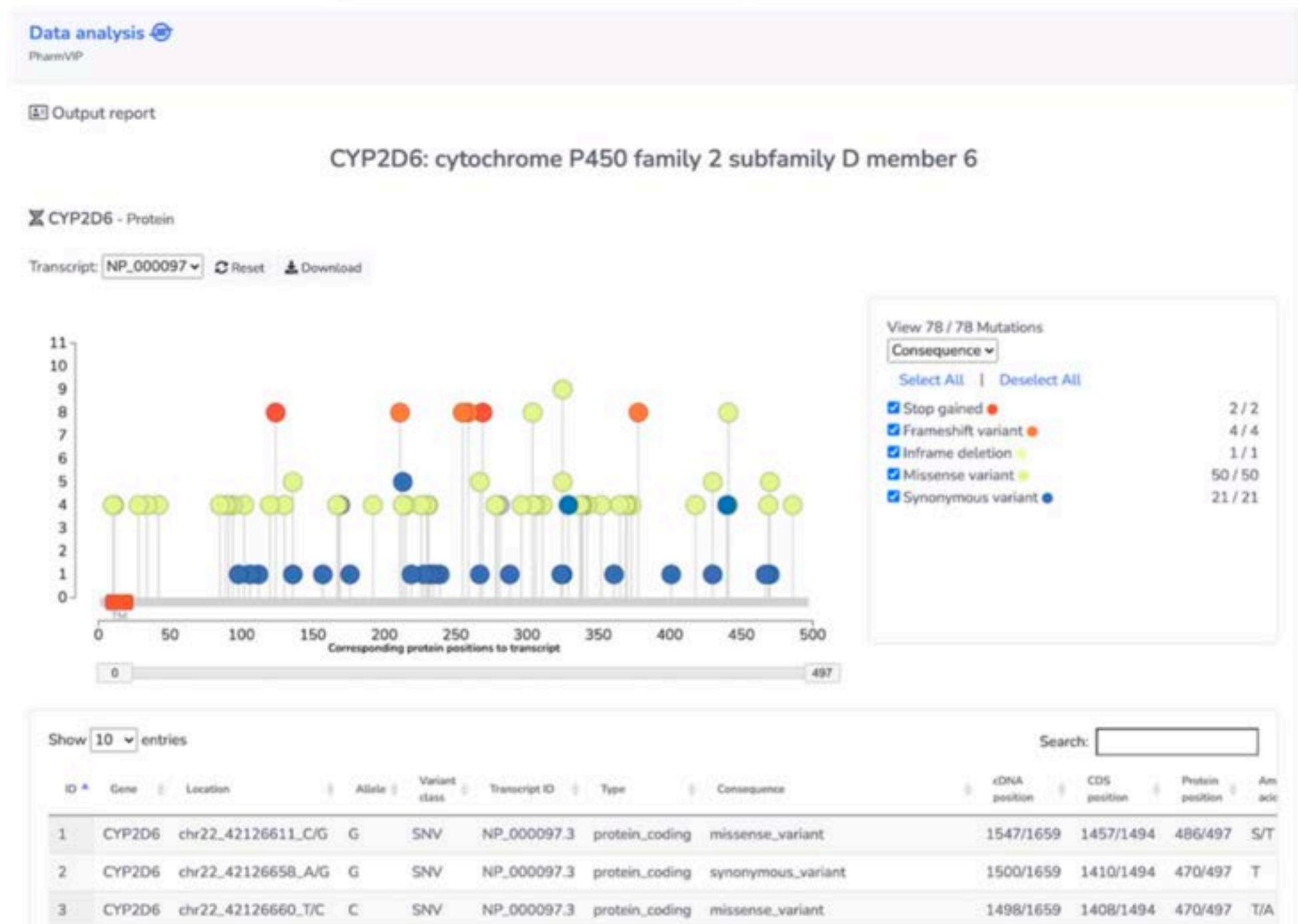
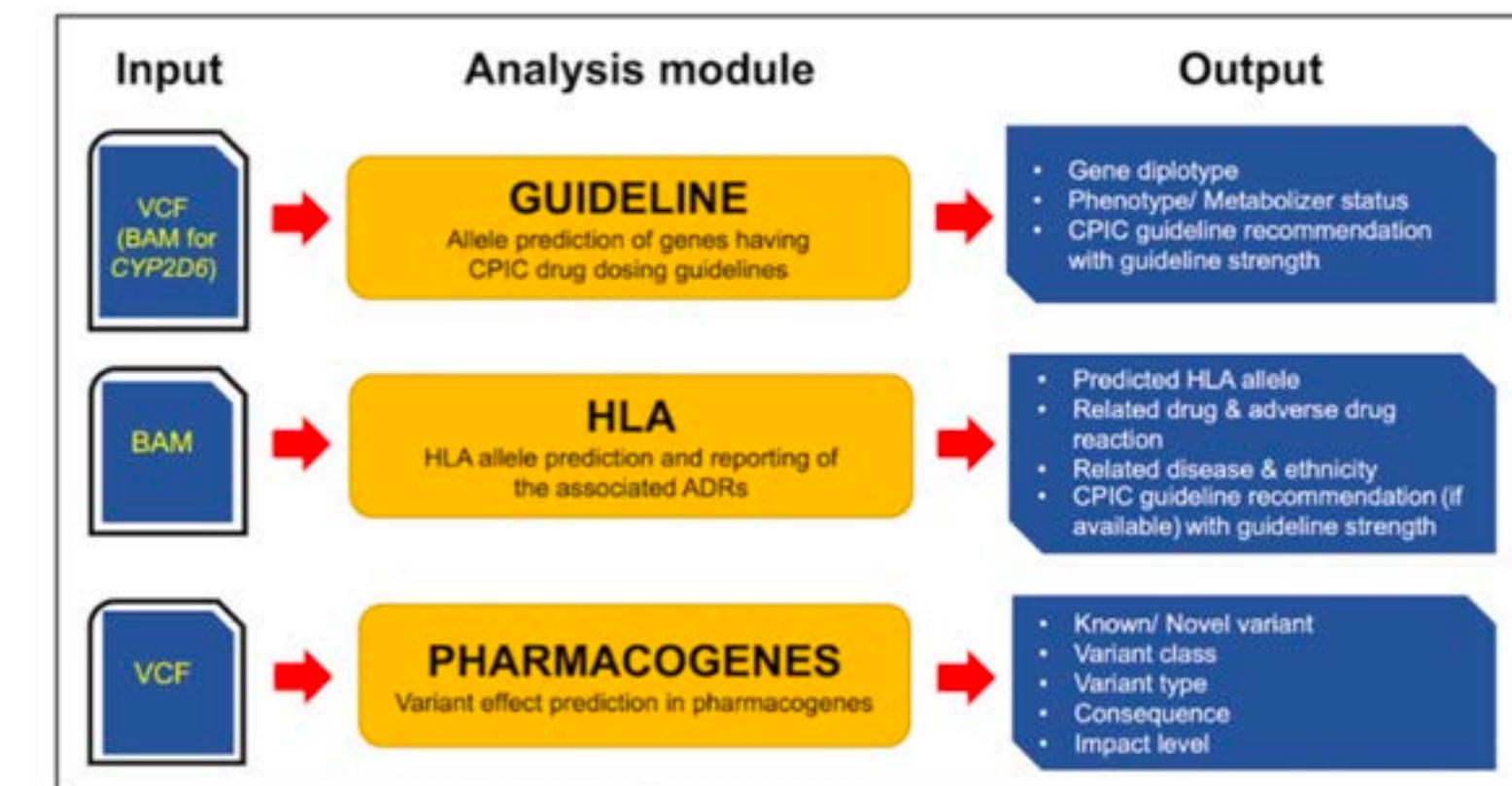
- ◎ Pharmacokinetics (Cytochrome P450)
- ◎ Pharmacodynamics
- ◎ Immune-mediated adverse drug reaction; IM-ADR

Video demo

Open Access Article

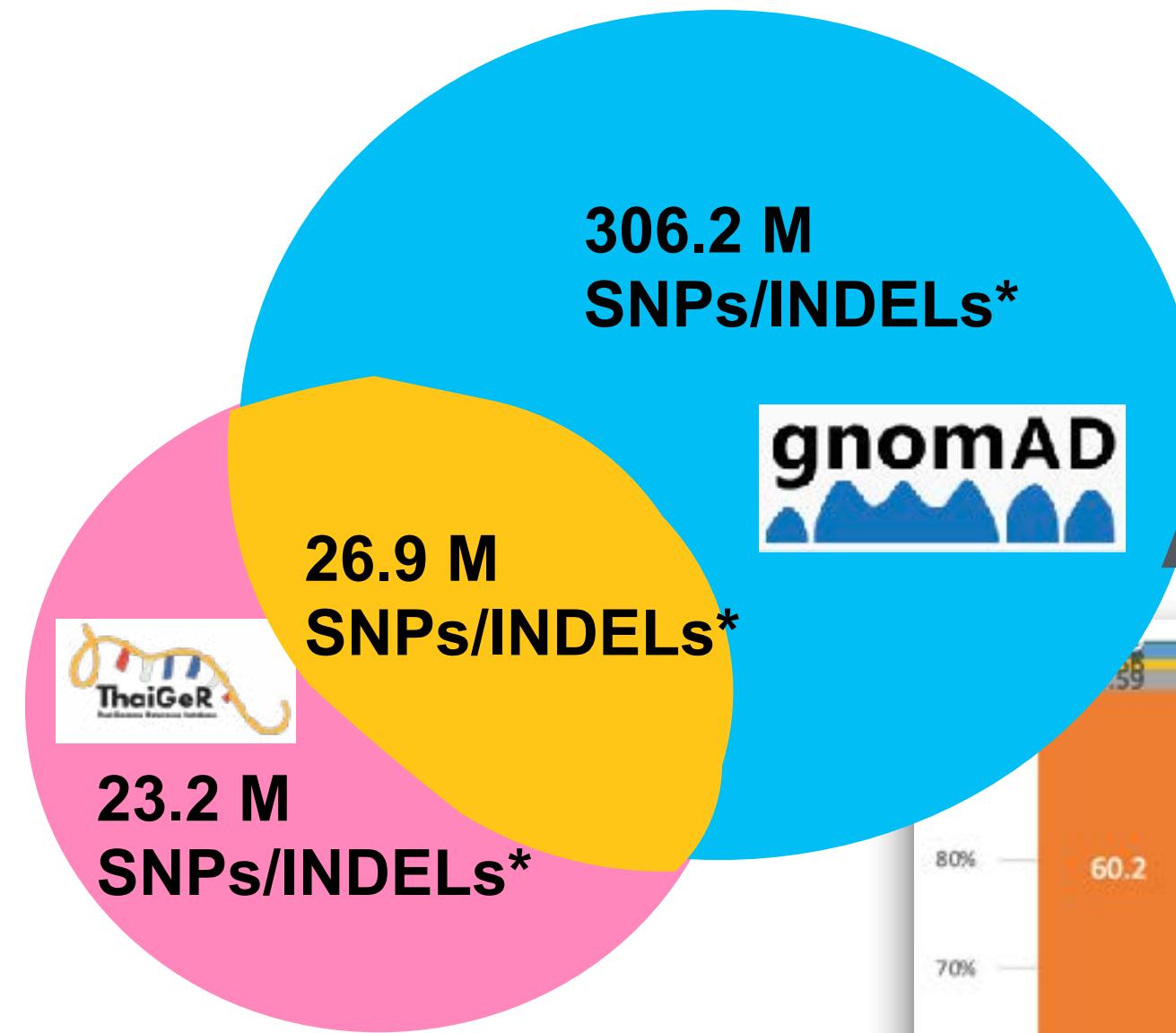
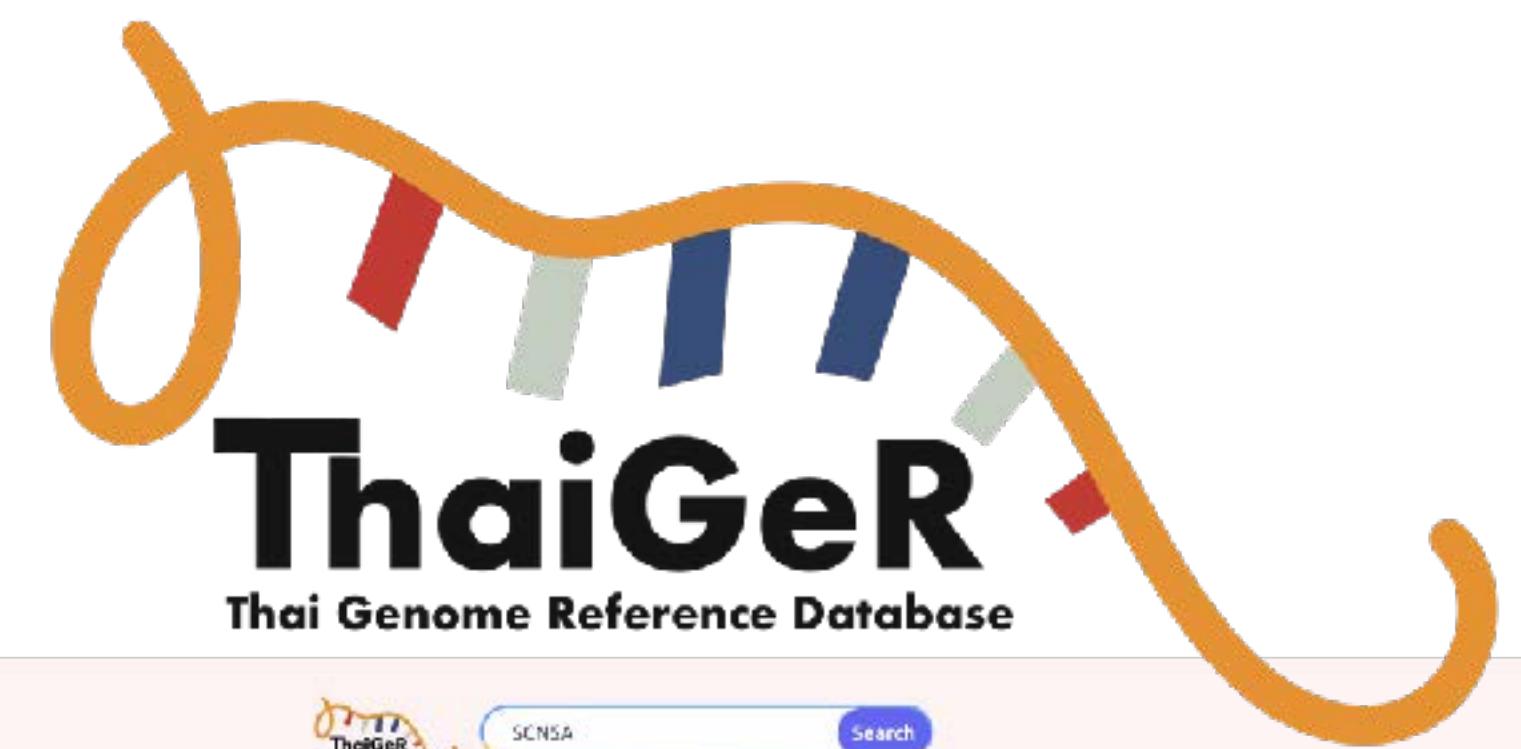
PharmVIP: A Web-Based Tool for Pharmacogenomic Variant Analysis and Interpretation

by Jittima Piriyapongsa 1,* , Chanathip Sukritha 1 , Pavita Kaewprommal 1 , Chalermpong Intarat 1 , Kwankom Triparn 1 , Krittin Phornsiricharoenphant 1 , Chadapohn Chaosrikul 1 , Philip J. Shaw 2 , Wasun Chanratita 3 , Surakameth Mahasirimongkol 4 and Sissades Tongsim 1



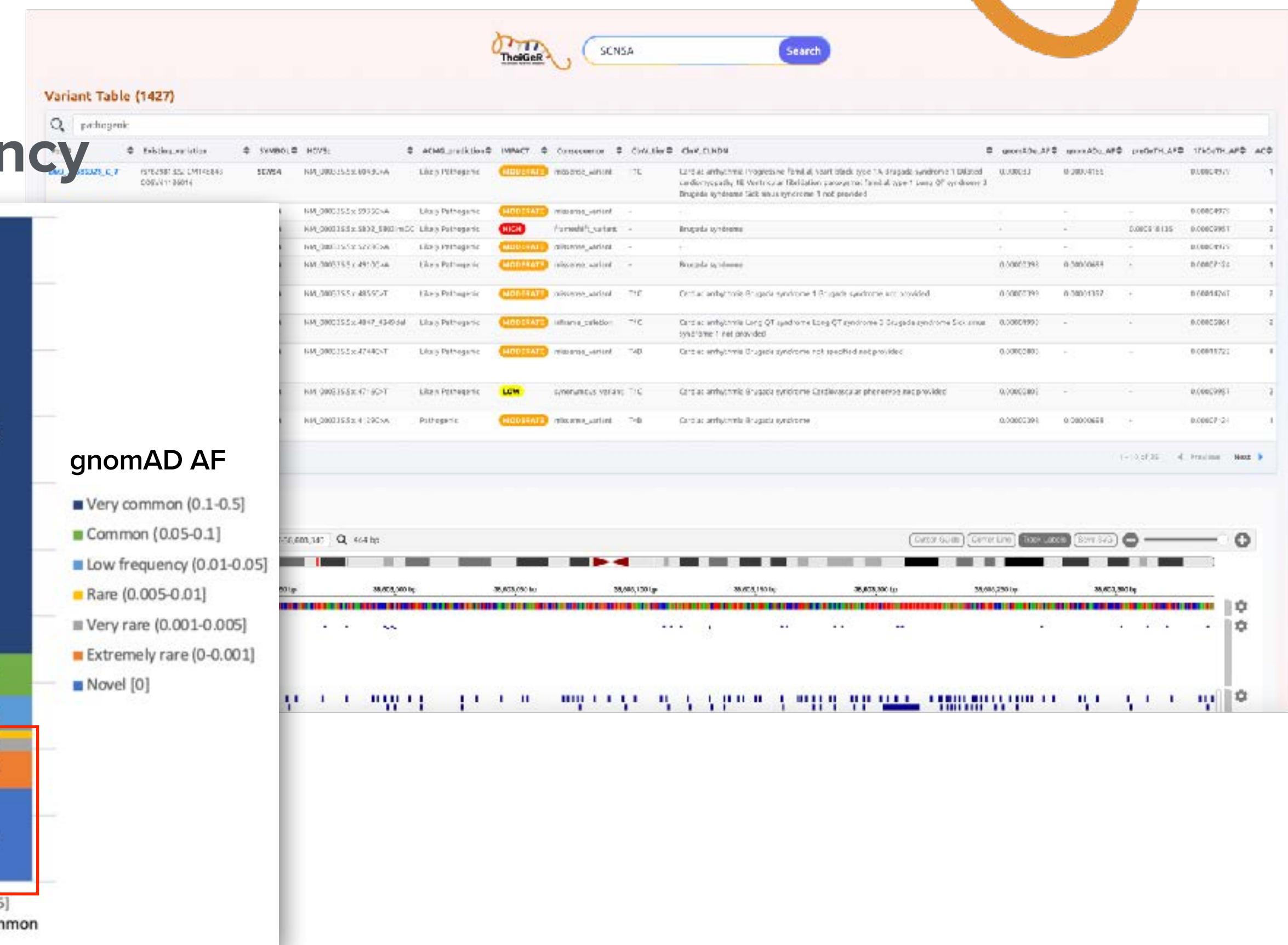
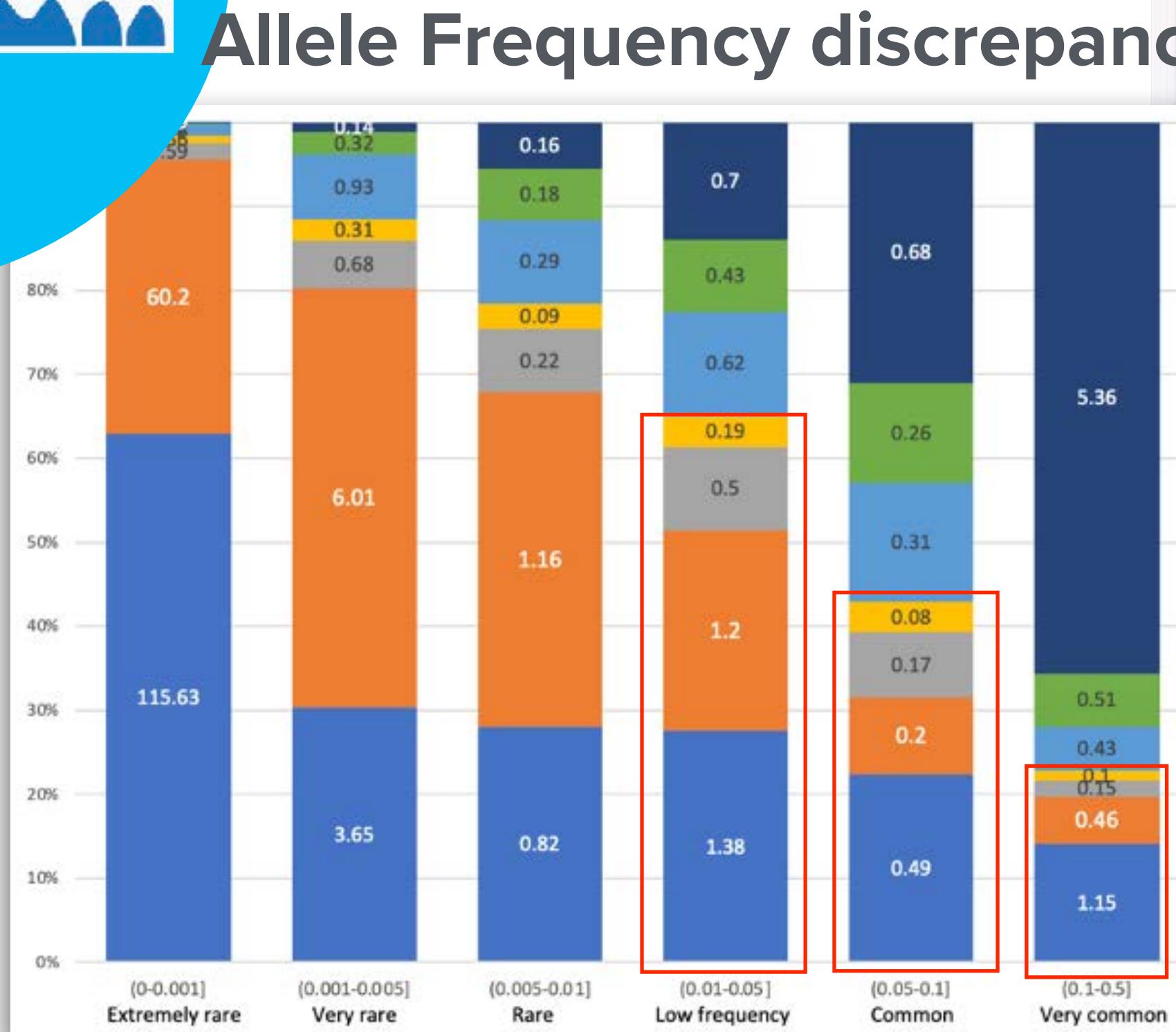
Thai Genome Reference Database

<https://thaiger.genomicsthailand.com>



*Non-private variants (AC>1)

14K Thai WGSs + gnomADv41
Total : 203,105,650 variants
186,739,173 SNPs
18,547,701 INDELs
Novel : 176,197,559 variants
(compared to dbSNP v138)



Infectious Genomics Projects

Mycobacterium tuberculosis (MTB) Platform

The platform includes:

- Account Summary
- Phylogenetic tree
- Structural variant analysis
- Lineage analysis
- Drug resistance prediction
- Outbreak analysis
- Genome browser
- MTB Report
- Mycobacterium Whole Genome Sequencing Report

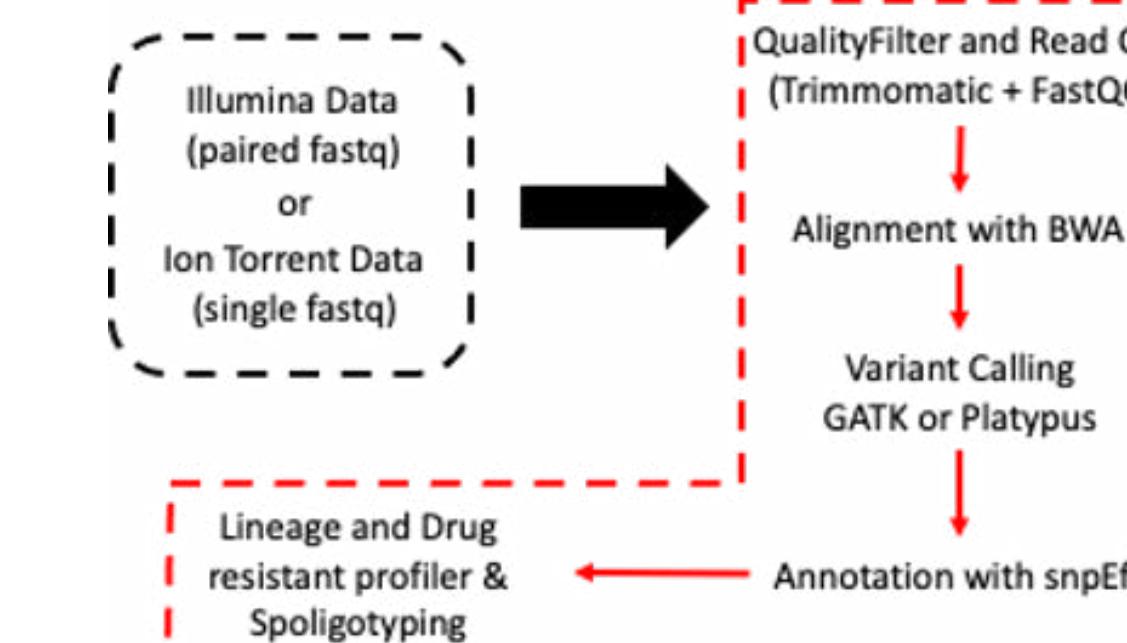
Features

- Drug resistance prediction
- Lineage classification
- Outbreak investigation

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

Sequencing read processing & annotation

Short reads

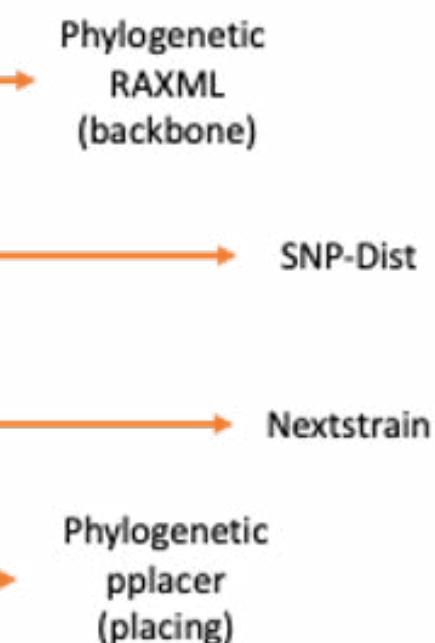
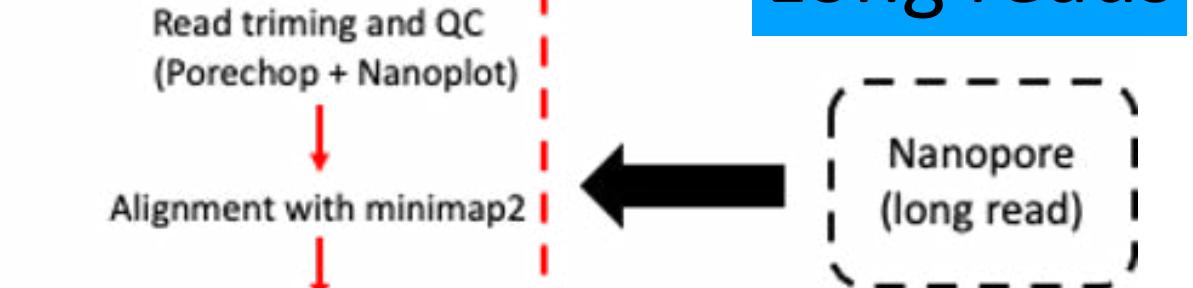


Flexible and Customizable Databases

- Post process for phylogenetic
 - Select only SNP
 - Filter homo variant only
 - Filter biallelic variant only
 - Filter out region PE/PPE and drug resistant

Phylogenetic analyses

Long reads



BIO-TEC
a member of **NSTDA**



Mahidol University
Faculty of Medicine
Siriraj Hospital

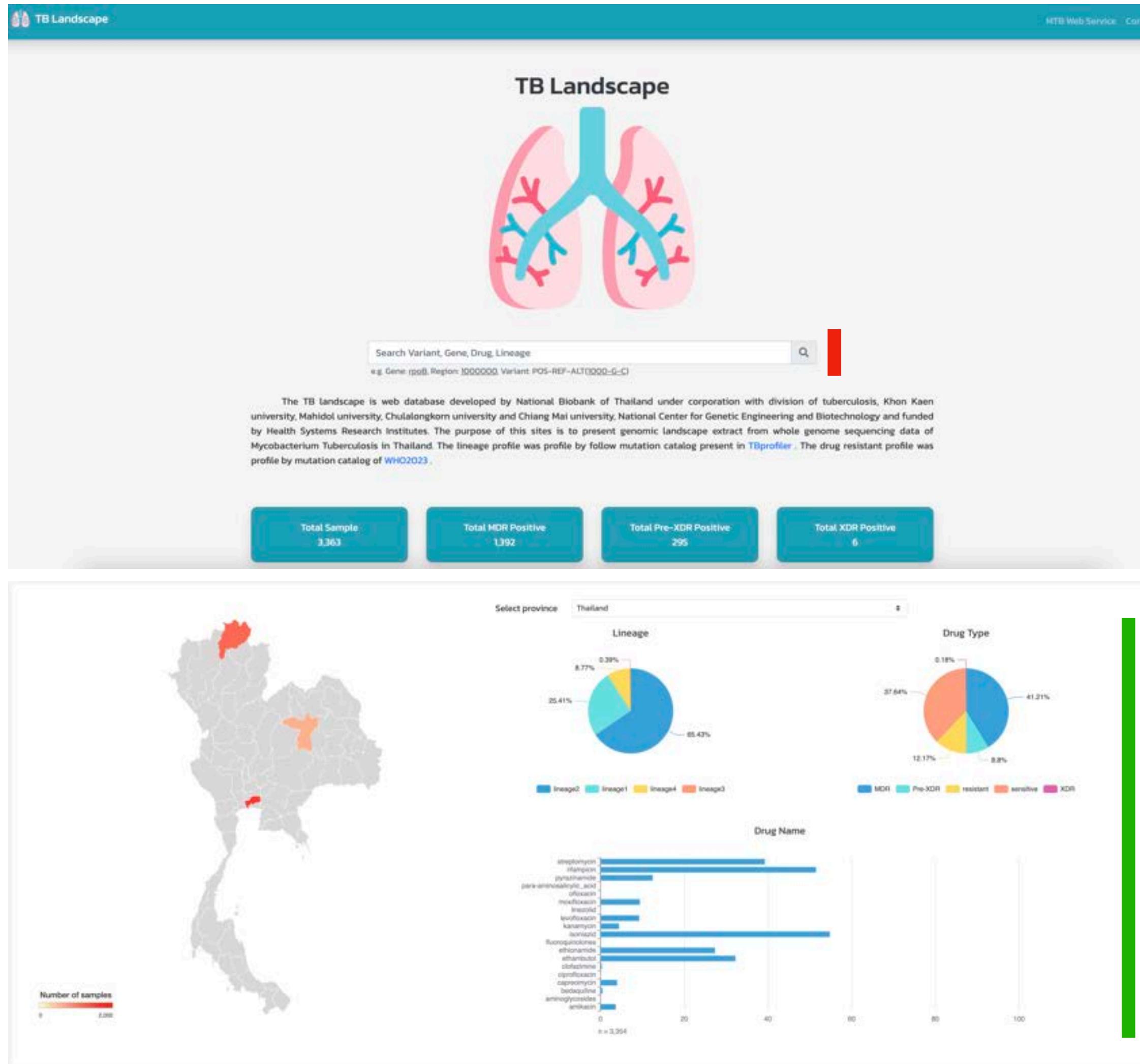


กองวัณโรค
Division of Tuberculosis



Mycobacterium tuberculosis Landscape

<https://mtbview.genomicsthailand.com/>



“Currently store **3,353** samples of TB genome present in Thailand”

Features

Variant querying

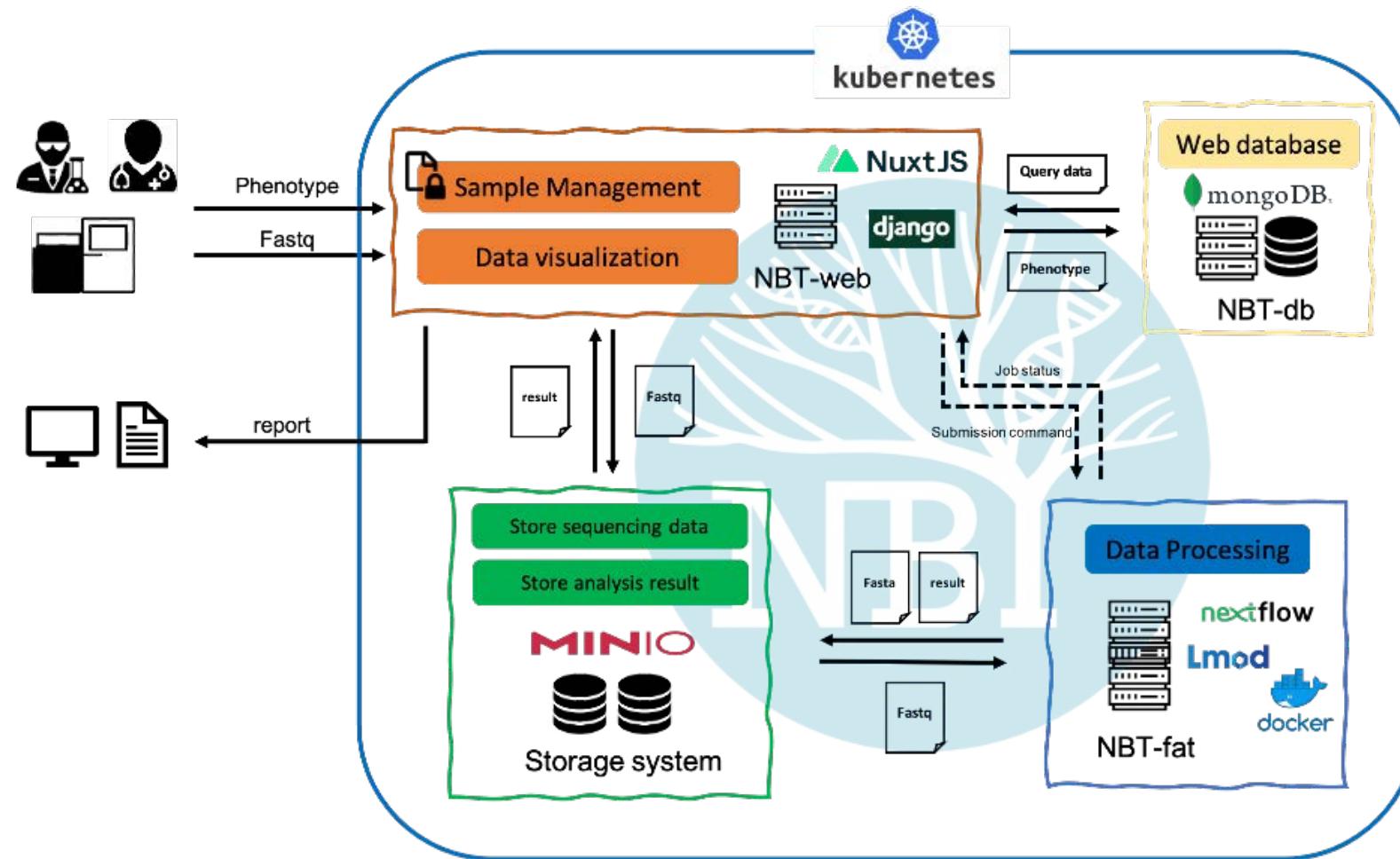
Variant frequency relate to Lineage

Variant frequency relate to Drug-resistant

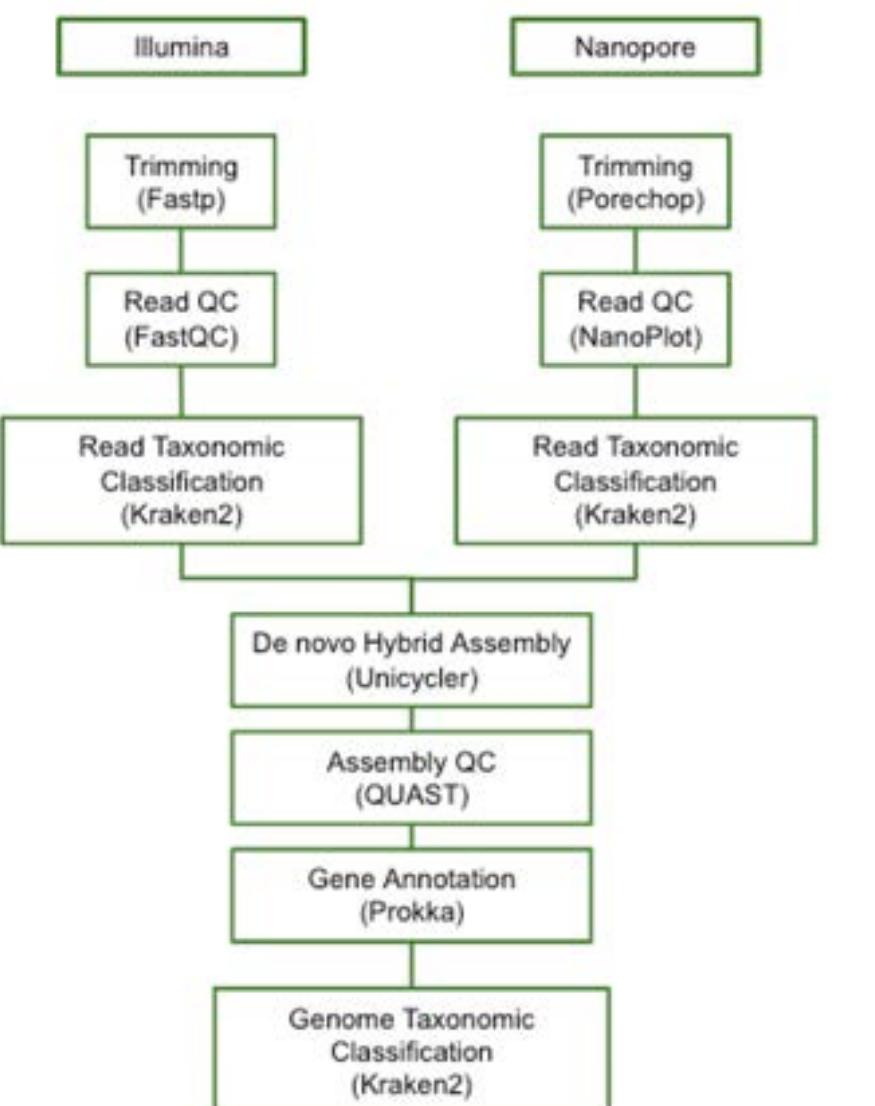
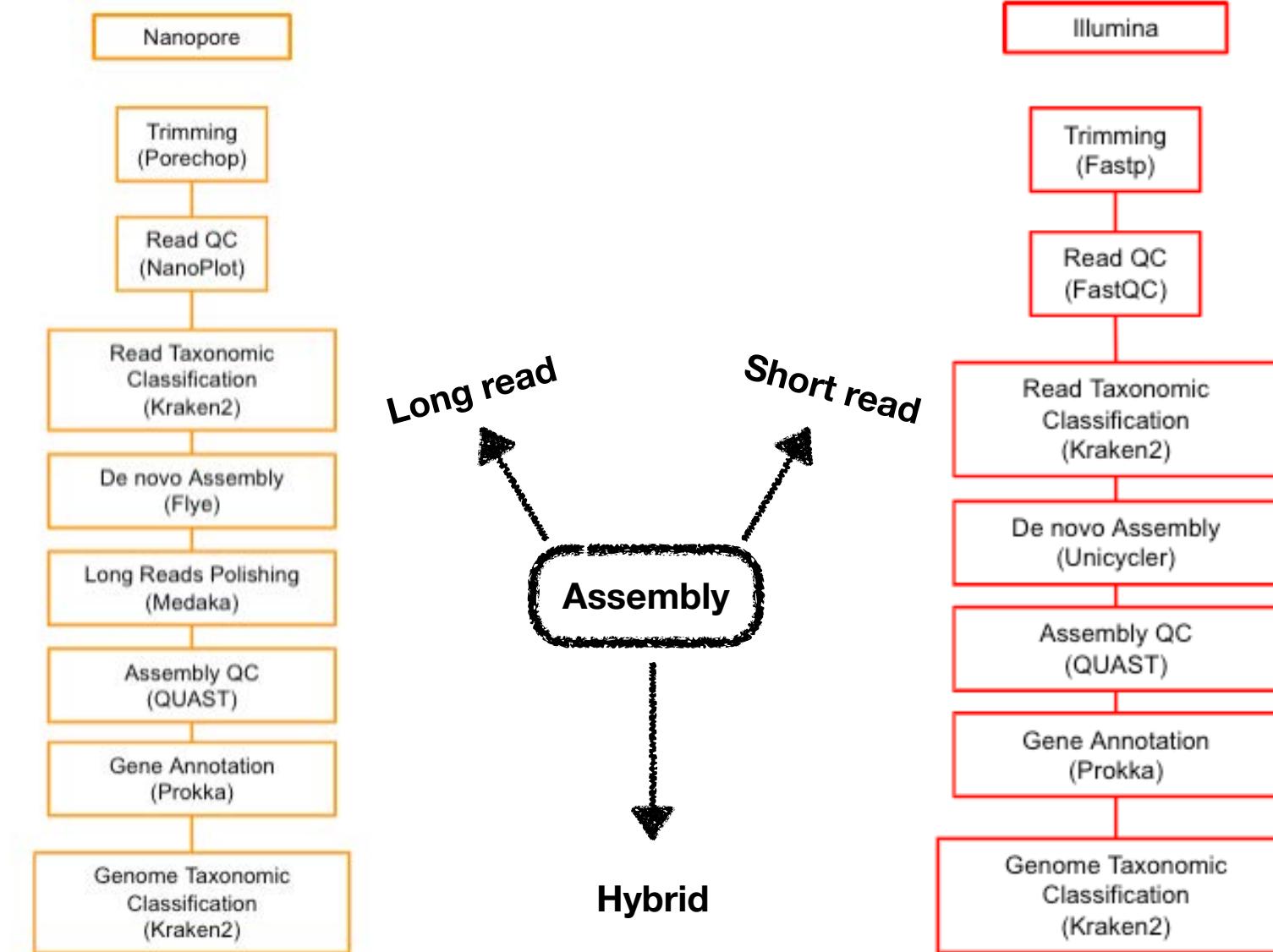
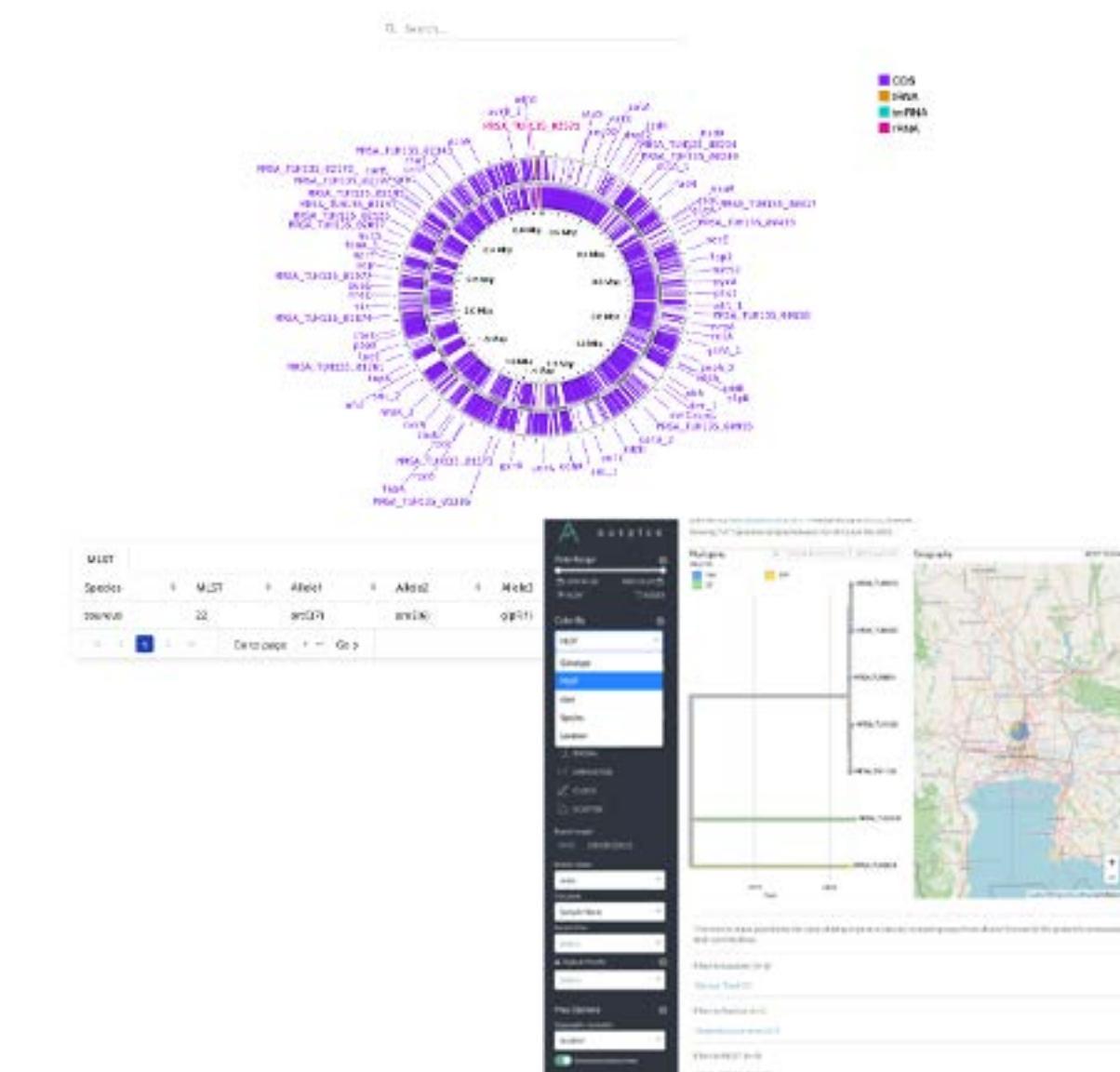
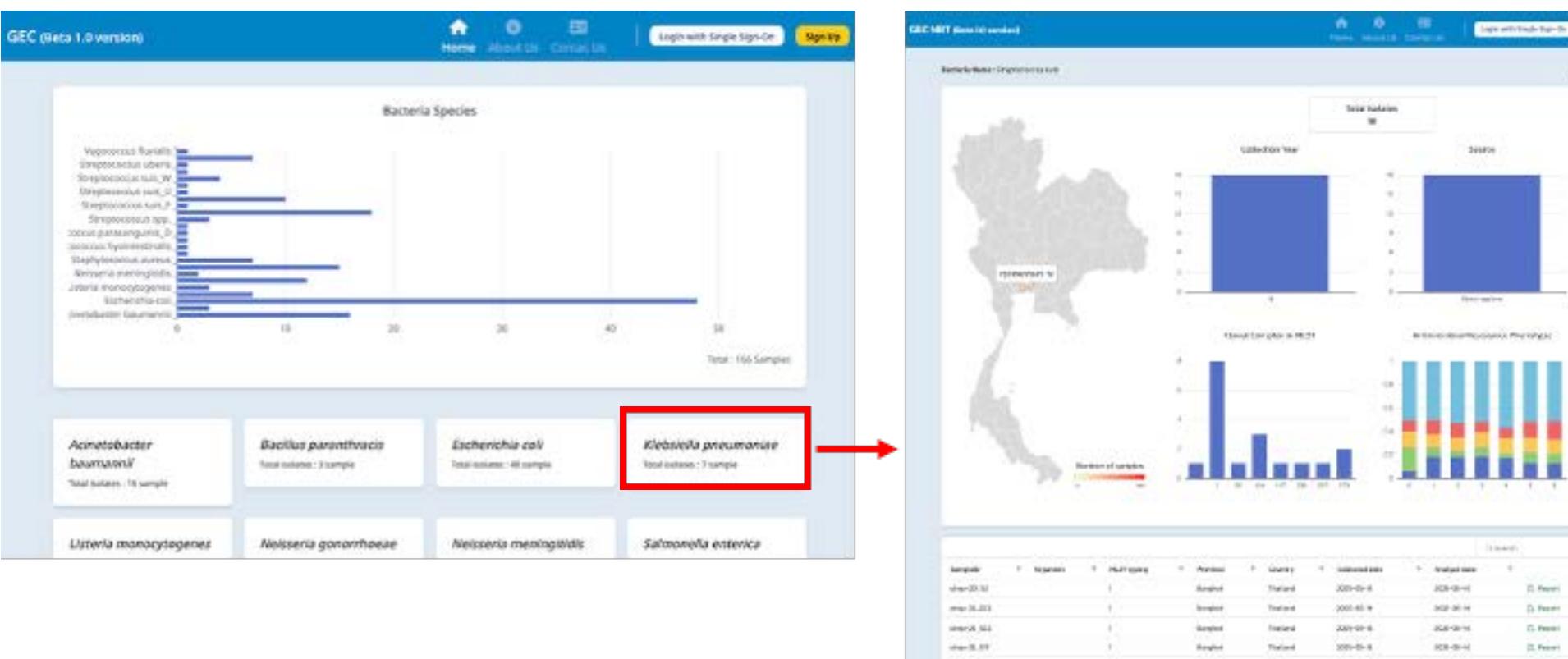
Geological distribution view

Antimicrobial resistance (AMR)

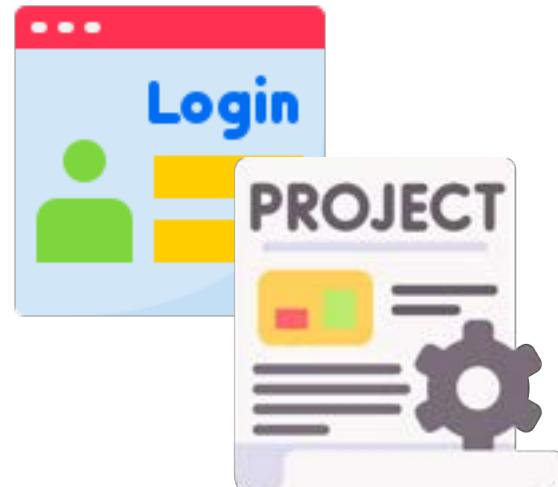
<https://gec.nbt.or.th/>



- ◆ Genome Assembly
- ◆ Taxonomic Classification
- ◆ Resistant Gene Identification
- ◆ Pangenome Analysis
- ◆ Plasmid and Mobile genetic elements Identification
- ◆ Genotypic classification (MLST, Serotyping)



Pathogen Genomic Surveillance (PathGS) Platform



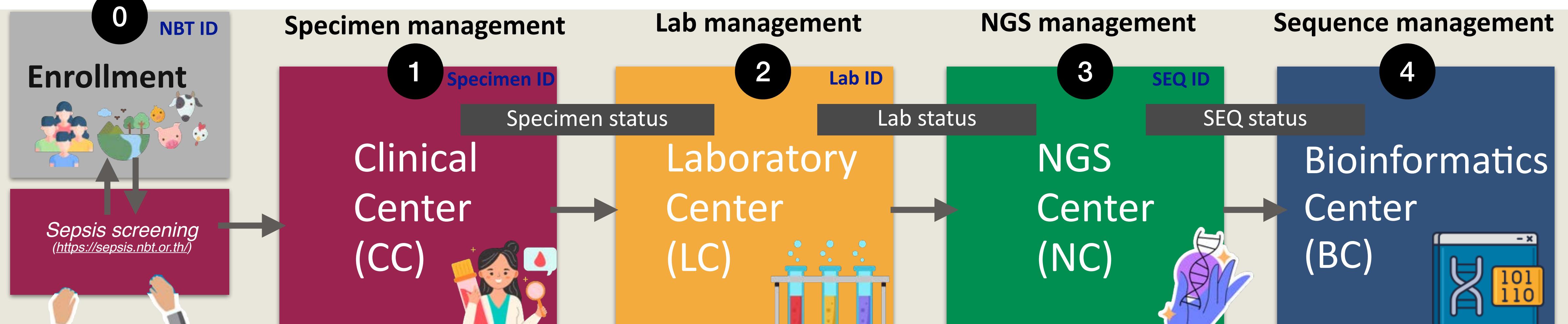
Sample store

- Login & registration
- Project management: create project, project approval, launch project, enroll project
- Sample store: sample dashboard, enroll samples
- Specimen management: SMS dashboard, samples



กรมวิทยาศาสตร์การแพทย์
Department of Medical Sciences

Specimen Management



- Enroll data & samples
- Return result

- Enroll specimen
- Collect specimen

- Cultivation & Isolation
- Genomic extraction
- Molecular identification
- Genomic quality control

- Genomic quality control
- Next-generation sequencing

- NGS data analysis

Acknowledgment



Sissades Tongsima Ph.D.
Director of Medical Molecular
Biotechnology Research Group



Chumpol Ngamphiw Ph.D.
Head of Integrative Genomics for
Precision Health Research Team

Bioinformatician / Data scientists

Jittima Piriyapongsa Ph.D.
Pongsakorn Wangkumhang Ph.D.
Vorthunjul Nakhonsri Ph.D.
Pimonwan Phokhaphan Ph.D.
Alisa Wilantho
Pavita Tipsombatboon

Programmer / Web developer

Watcharapot Janpoung
Peerapat Khunkham



Thank you

