

ตอบโจทย์การวิจัยและธุรกิจจีโนมด้วย “computational genomics Platform” Research and business enabler as a national computational genomics Platform

ดร.ศิษเกศ ทองสีมา

ผู้อำนวยการธนาคารทรัพยากรชีวภาพแห่งชาติ
สำนักงานพัฒนาวิทยาศาสตร์และเทคโนโลยีแห่งชาติ



ประวัติการทำงาน:

2561-ปัจจุบัน ผู้อำนวยการธนาคารทรัพยากรชีวภาพแห่งชาติ
 สำนักงานพัฒนาวิทยาศาสตร์และเทคโนโลยีแห่งชาติ

2550-2561 Head of Biostatistics and Informatics Laboratory, Genome Institute, BIOTEC,
 Thailand

Computational Genomics Platform

Research and business enabler as a national computational genomics platform

Sissades Tongsimma, PhD

NAC2022
17th NSTDA Annual Conference
การประชุมวิชาการประจำปี สวทช. ครั้งที่ ๑๗



Genetic factors play “roles” in disorders



129.36
Meter

Paper stack of Human genome

> 3 billion nucleotides
or DNA “letters.”

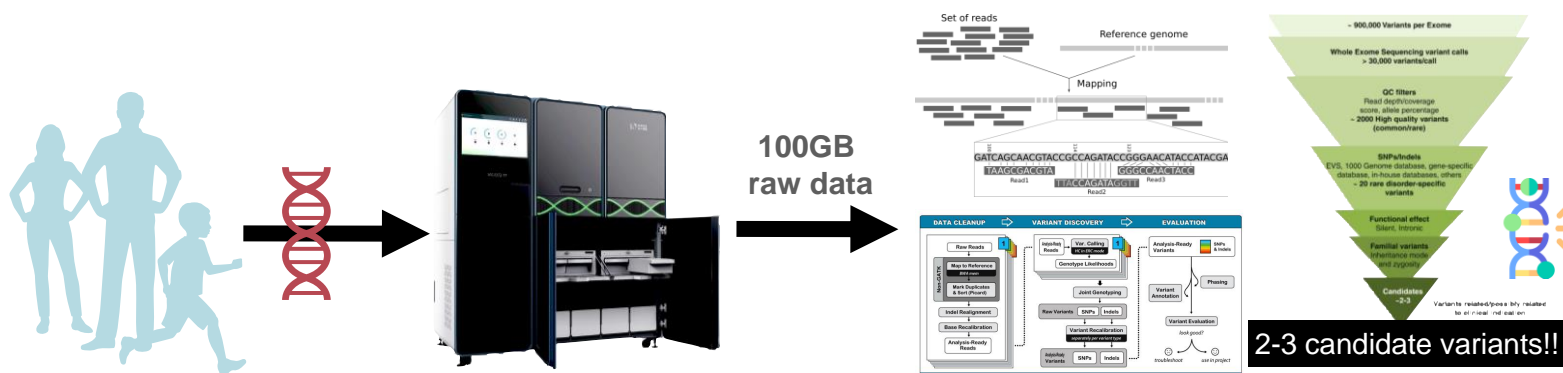


Rare diseases

Obesity
Diabetes
Cancers etc.

Infectious/
Accidents etc.

Population specific requirements



Thai reference exome database (T-REx)

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



ORIGINAL ARTICLE

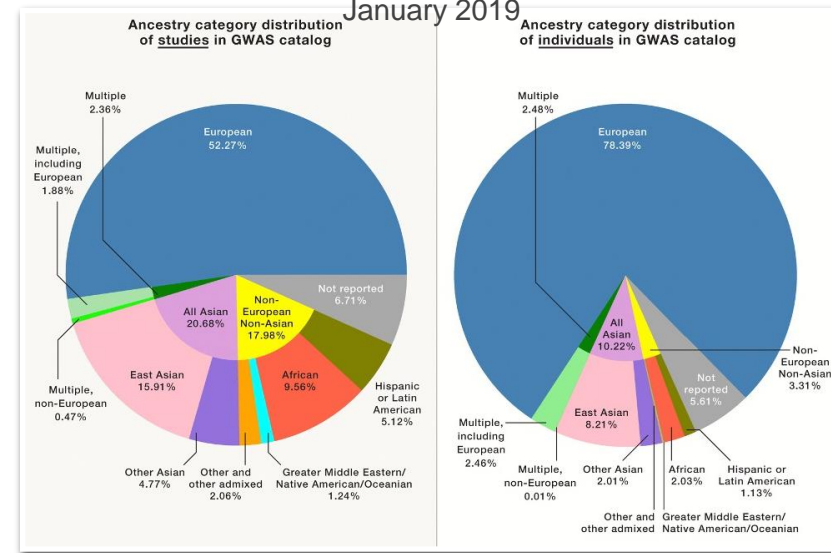
The Thai reference exome (T-REx) variant database

Vorasuk Shotelersuk, Duangdao Wichadakul, Chumpol Ngamphiw, Chalurmporn Srichomthong, Chureerat Phokaew, Alisa Wilantho, Sujiraporn Pakchuen, Vorthunju Nakhonsri ... See all authors

First published: 08 September 2021 | <https://doi.org/10.1111/cge.14060>

- ▶ **Caucasian bias reference genome**
- ▶ **Require specialized High Performance Computer (HPC)**
 - ▶ Time *****AND***** Storage ⇒ **critical demands**
 - ▶ Normal HPC ⇒ 30 hours VS customized HPC ⇒ ~1 hour
 - ▶ NBT HPC for genomic-related processing
- ▶ **Require advanced bioinformatic tools and services**
- ▶ **Require skillful bioinformaticians/programmers/system admin**

Summary of GWAS Studies in the GWAS Catalog through January 2019



GeTH 50K

1 HPC & Storage

- ▶ 2x Accelerated AI machines (DGX-A100)
- ▶ 3x FAT nodes (192 cores / 3TB)
- ▶ 8x Web & Database servers
- ▶ 3000 TB storage & 10000 TB buffered storage

2 Software Management & QC

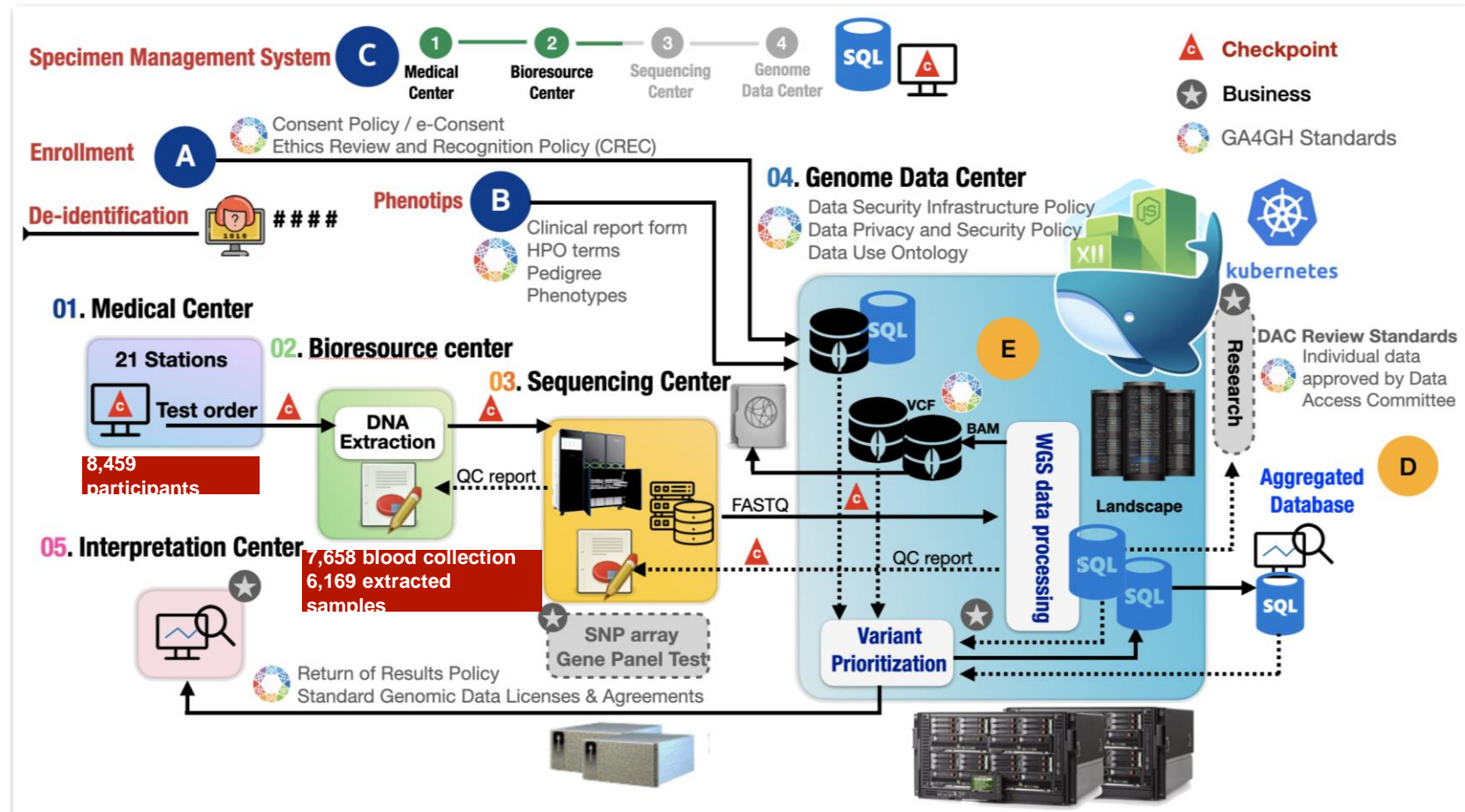
- Enrollment & e-consent
- PHENOTIPS[®]
- Specimen Mangement System (SMS)
- Whole Genome Sequence Data QC
- Accelerated WGS processing

3 Genomics dB & Prioritization platform

- Population genomics platform
- T-REx Database
- Thai Reference Genome dB
- Variant Annotation and Prioritization (VAPP)

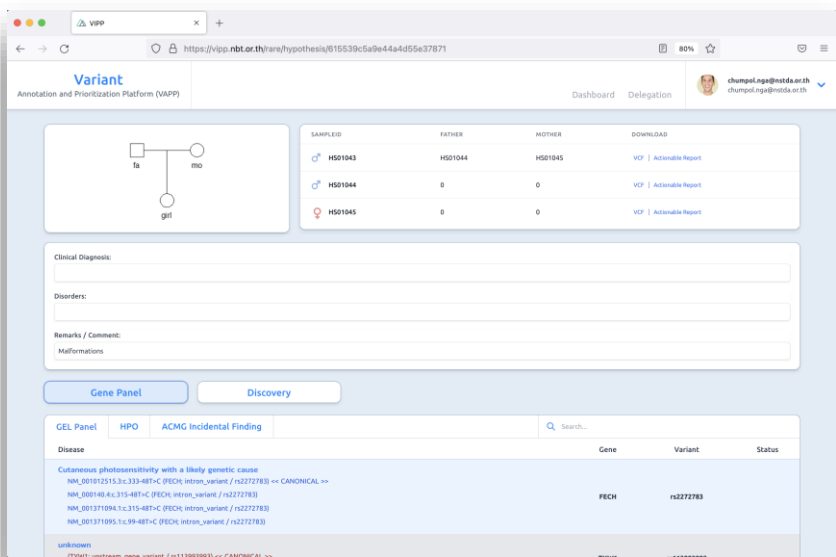
4 Data governance & Policy

- ▶ GeTH Data Governance policy
- ▶ Data Use Access policy
- ▶ Data sharing policy
- ▶ Intellectual Property policy

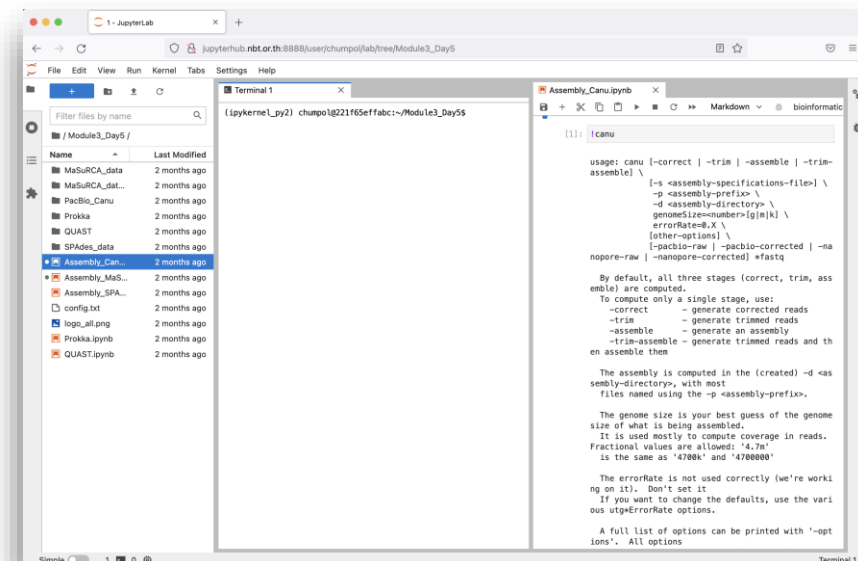


Computation services and Data access

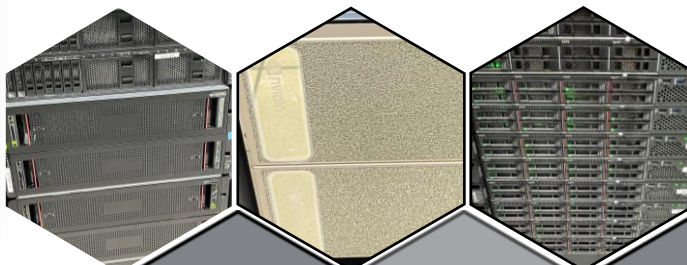
Web Application with preannotated results



Jupyterhub / Terminal / Virtual Env.
for Adv. proc on control data



Doctor - Researcher -
Private comp.



To transfer genome data
is Big problem !



HPC & Storage

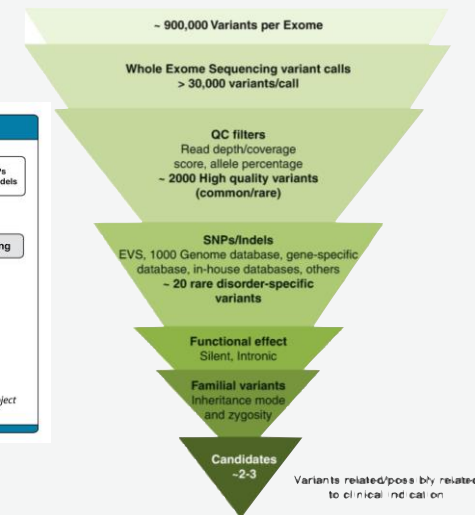
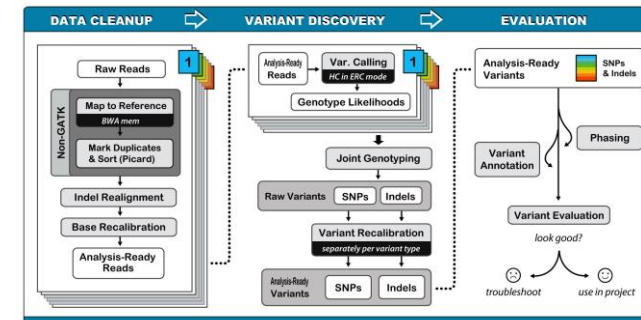
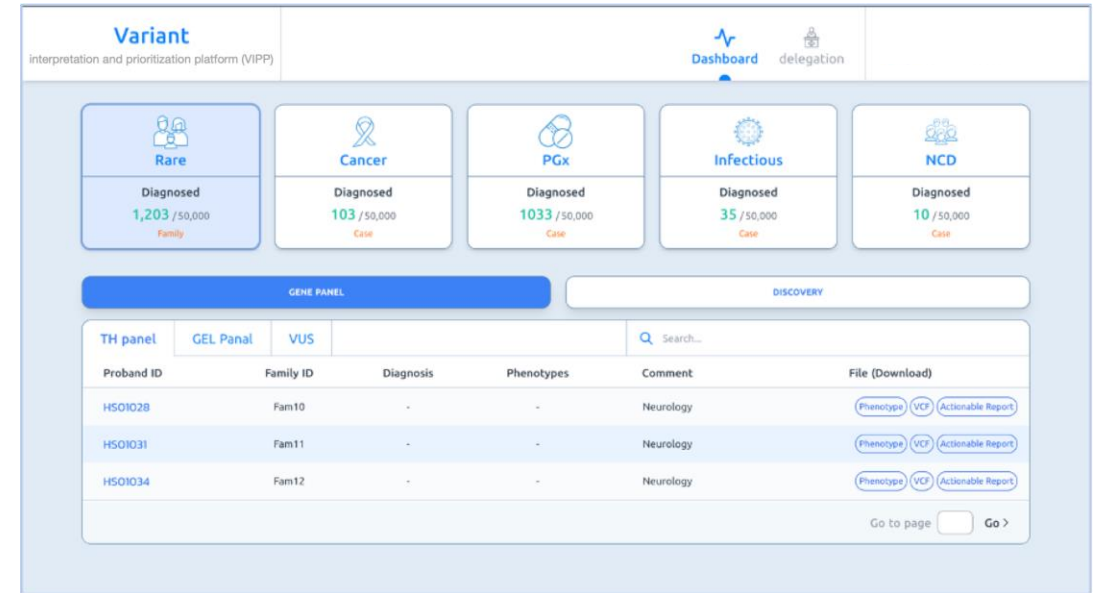


Internet bandwidth
Computing resources
Data leak !

Variant Annotation and Prioritization: VAPP

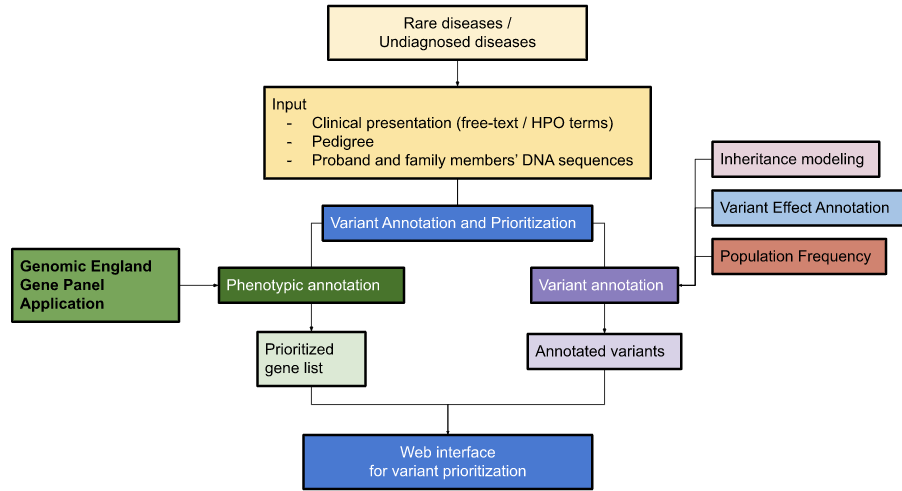
- Annotation & Prioritization (ANALYSIS)

- Rare
- Cancer
- NCD
- PGx
- Infectious
- GEL panel + HPO terms
- Large polygenic risk score (PGS) database
- Genomics sequence annotate for “supporting” clinical interpretation



VAPP

Rare disease + Canc



Variant
Annotation and Prioritization Platform (VAPP)

Dashboard Delegation

Rare

Diagnosed
1,203 / 50,000
Family

Cancer

Diagnosed
103 / 50,000
Case

PGx

Diagnosed
1033 / 50,000
Case

NCD

Diagnosed
10 / 50,000
Case

Infectious

Diagnosed
35 / 50,000
Case

Proband List		All Sample		Search...				
No	Proband ID	NBT ID	Family ID	Sex	Diagnosis	Phenotypes	Comment	File (Download)
1	HS01043	12345678	Fam15	♂		2	Malformations	Phenotype VCF Actionable Report
2	HS01007	12345678	Fam3	♂		2	Neurology	Phenotype VCF Actionable Report
3	HS01013	12345678	Fam5	♀		2	Neurology	Phenotype VCF Actionable Report

Gene Panel

+ New hypothesis

SNPs/INDEL
newdevtest
test detail
last modify
Mon Oct 25 23:33:48 2021

Discovery

SNPs/INDEL ★
demo2
jpdPqjgJf
test
last modify
Sun Oct 24 15:22:32 2021

SVs/CNVs ★
demo1
dev-test. detail:dev-test. detail
test
last modify
Sun Oct 24 15:22:32 2021

SVs/CNVs ★
dev-test
dev-test
last modify
Wed Oct 27 16:58:33 2021

SNPs/INDEL ★
ทดสอบภาษาไทย
ทดสอบภาษาไทย
last modify
Tue Oct 26 13:01:24 2021

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

Impact / Sequence ontology Filter 904 / 1,203

MODERATE

HIGH MODERATE LOW MODIFIER

Sequence ontology:

transcript_ablation splice_acceptor_variant splice_donor_variant stop_gained frameshift_variant
 stop_lost start_lost transcript_amplification inframe_insertion inframe_deletion
 missense_variant protein_altering_variant regulatory_region_ablation

5' 3'

regulatory region TF binding site 5 prime UTR variant start retained variant start lost splice donor variant splice acceptor variant splice region variant intron variant 3 prime UTR variant downstream gene variant

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

INTEGRATIVE GENOMICS VIEWER (IGV)

Genes: GRP42 CSH1

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

35,372,235 bp 35,372,240 bp 35,372,245 bp 35,372,250 bp

A G G A G G T T G T G T G G G C T C

Refseq Genes

Proband (HS02067)

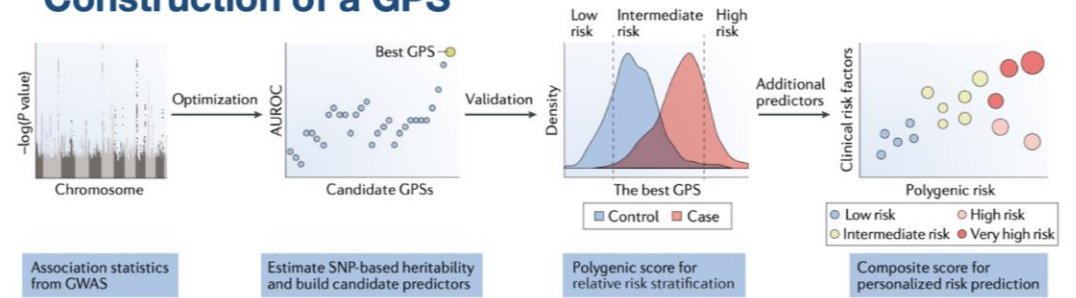
Father (HS02068)

Mother (HS02069)

VAPP NCDs



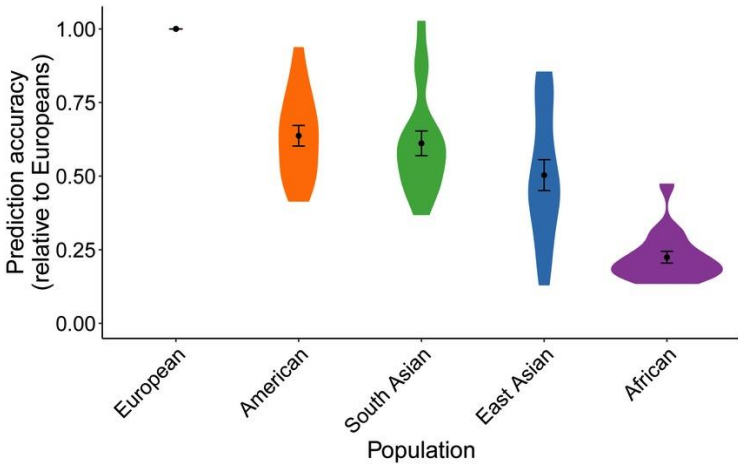
Construction of a GPS



PGS Catalog



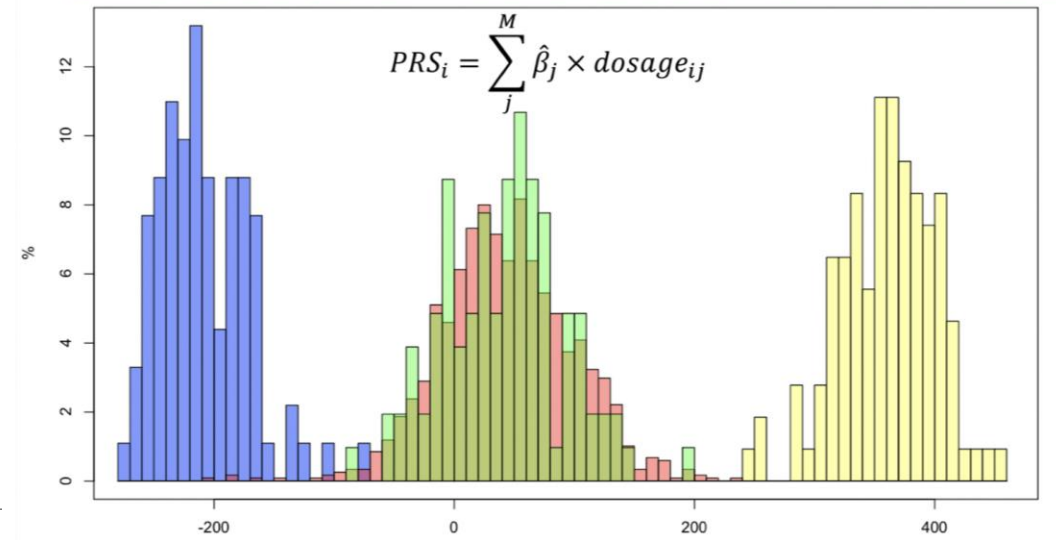
Polygenic Scores: 2,174
 Traits: 531
 Publications: 288



T breast cancer (MONDO_0007254)
 A primary or metastatic malignant neoplasm involving t C9335] < Show less
 Associated PGS: **110** >> Show PGS +

Which one should we use for Thai PGS?

Genetic Score Distribution of Skin Pigmentation in GBR(blue),Thai(red),CHB(green),and YRI(yellow) population



Martin et al., 2019. Clinical use of current polygenic risk scores may exacerbate health disparities. Nature Genetics, 51(4), pp.584-591.

VAPP

Pharmacogenomics: PGx



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

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

Video demo



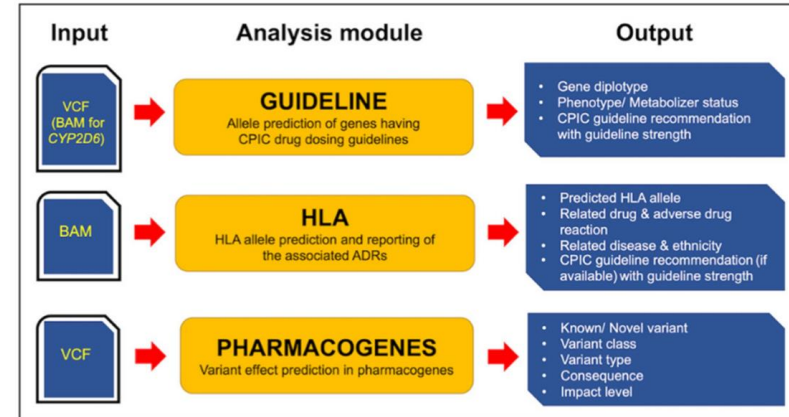
The screenshots show the following steps in the PharmVIP interface:

- STEP 1: Analysis options** - Select at least one option: Guideline, HLA, and PGx genes.
- STEP 2: Input files** - Upload VCF or BAM files for the selected options.
- STEP 3: Analysis parameters** - Configure filters and parameters for each selected option.
- STEP 4: Analysis details** - Review the selected options and parameters before starting the analysis.

Article

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

Jittima Piriyaongsa ^{1,*}, Chanathip Sukritha ¹, Pavita Kaewprommal ¹, Chalermpong Intarat ¹, Kwankom Tripam ¹, Krittin Phornsrichaenphant ¹, Chadapohn Chaosrikul ¹, Philip J. Shaw ², Wasun Chantratita ³, Surakameth Mahasirimongkol ⁴ and Sissades Tongsim ¹



The screenshot shows the output report for CYP2D6 (cytochrome P450 family 2 subfamily D member 6). It includes a mutation plot showing 78 mutations across the protein sequence (0-497 amino acids). A legend indicates the following mutation counts:

- Stop gained: 2 / 2
- Frameshift variant: 4 / 4
- Inframe deletion: 1 / 1
- Missense variant: 50 / 50
- Synonymous variant: 21 / 21

Below the plot is a table of variant entries:

ID	Gene	Location	Allele	Variant class	Transcript ID	Type	Consequence	cDNA position	CDS position	Protein position	Amino acid
1	CYP2D6	chr22_42126611_C/G	G	SNV	NP_000097.3	protein_coding	missense_variant	1547/1659	1457/1494	486/497	S/T
2	CYP2D6	chr22_42126658_A/G	G	SNV	NP_000097.3	protein_coding	synonymous_variant	1500/1659	1410/1494	470/497	T
3	CYP2D6	chr22_42126660_T/C	C	SNV	NP_000097.3	protein_coding	missense_variant	1498/1659	1408/1494	470/497	T/A

Other platform services

Infectious disease platform MDR-Tuberculosis

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

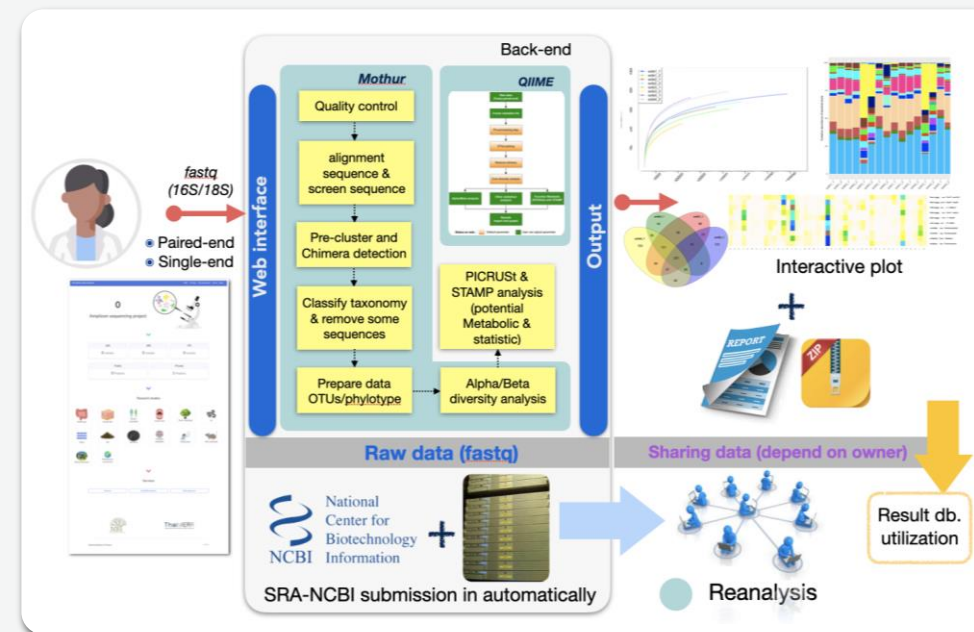
National Biobank of Thailand : TB Web Service

User Manual Contact us Login

Mycobacterium Tuberculosis (MTB) causes a major problem in public health. With the high prevalence of MTB in Thailand, the world health organization (WHO) assigned Thailand to the high tuberculosis (TB) burden group of 14 countries whose burdens encompass TB infection, infection with both TB and HIV (TB/HIV) and infection with multidrug-resistant TB (MDR-TB). More and more cases of drug resistant TB hampers the success of TB control program due to higher treatment failure rate. Quite often that new TB infected patients will be prescribed with multiple drugs because the standard TB drug sensitivity test usually takes up to two months. Such practices could promote the higher incidence of TB drug resistance. Thus, WHO recommends the use of TB whole genome sequencing (WGS) in the standard TB control program comprising diagnosis, prediction of TB drug resistance, and TB spreading management. WGS is faster and cheaper, soon TB WGS is hence becoming a standard practice in public health. However, WGS of TB entails generation of large sequencing data files. Such data require rather complex bioinformatic operations that could confuse TB interpretation personnel. With this challenge, we develop this platform to solicit support for development of a computational platform assisting TB drug resistant prediction with high accuracy using both TB's single nucleotide polymorphisms (SNPs) and structural variations (SV) from TB WGS data. The resulting computational prediction of drug-resistant TBs should promote the construction of an up-to-date national TB genomic portal that offer a prototype of TB drug program of TB in Thailand.

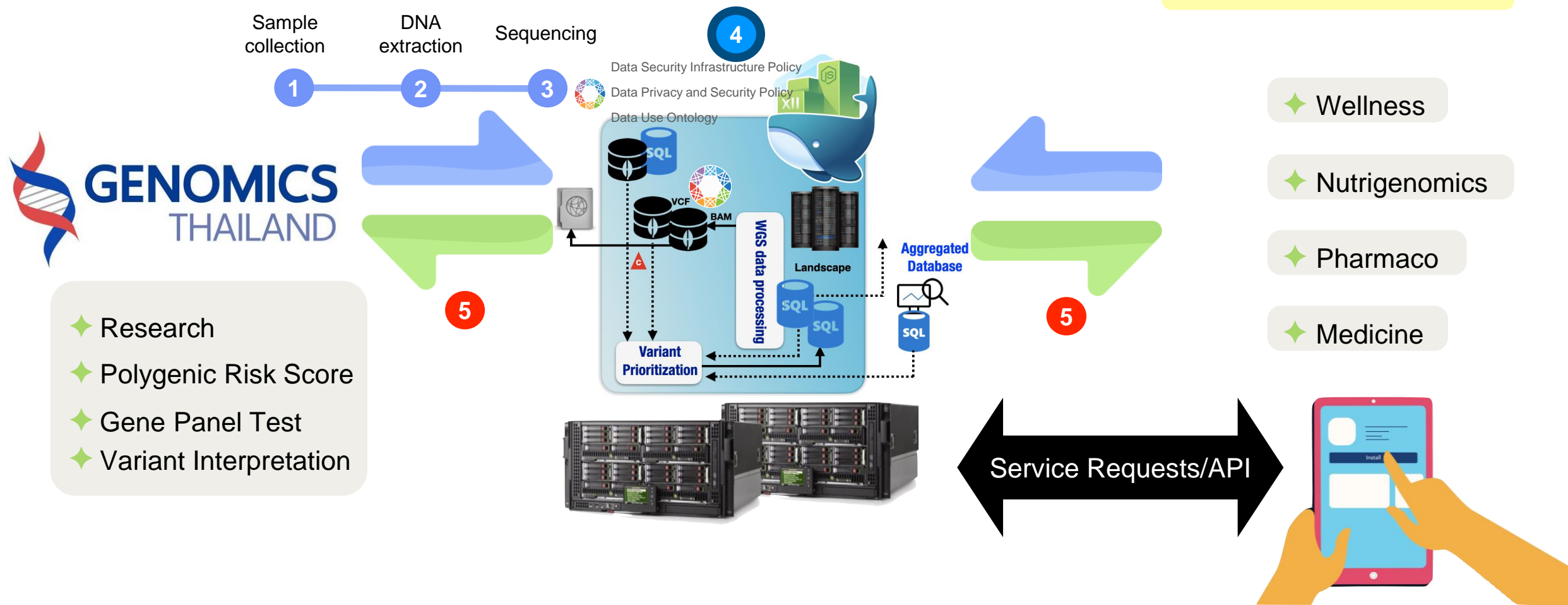
Metagenomics Amplicon Metagenomics Web Analysis (AmpMet)

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



Computational Genomic Platform

NBT/NSTDA as a ***service enabler***



GeTH Policy

IP Policy

Access Use data and restricted data

▶ Web-based interface

▶ Jupyter Notebook 

Data sharing

▶ Data Access Committee (DAC)

▶ Data stay local policy:

algorithms move to data analysis

▶ A reading library not a lending library:

can not remove the data

▶ Equal Data Governance and Securities

policy standard

Intellectual Property



Data Management policy

GeTH Data Governance policy

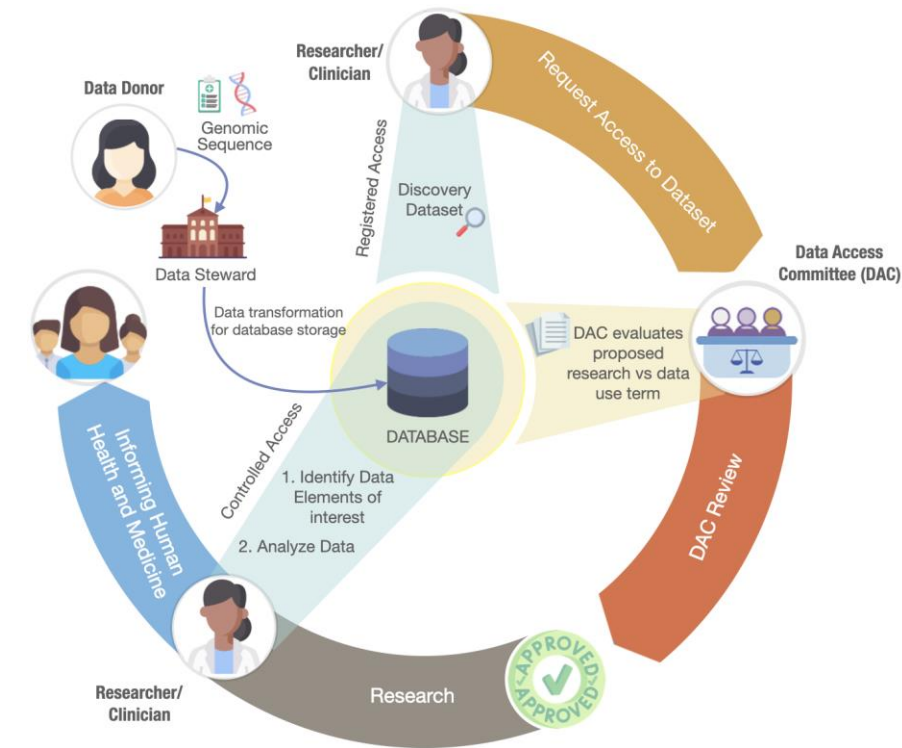
▶ Data Security Infrastructure

▶ Data Privacy and Security

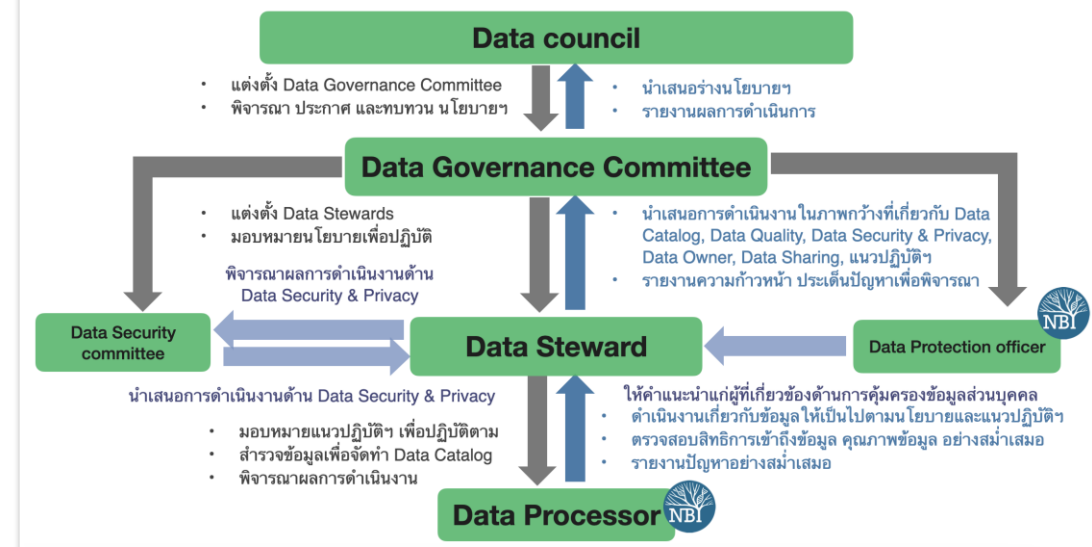
▶ Return of Results Policy



Global Alliance
for Genomics & Health
Collaborate. Innovate. Accelerate.



Structure Of Data Governance



Acknowledgement



EEC
เขตพัฒนาพิเศษภาคตะวันออก
เชื่อมโลก ไร้ไทยแลนด์



สวทช
NSTDA



กรมวิทยาศาสตร์การแพทย์
Department of Medical Sciences



คณะแพทยศาสตร์
FACULTY OF MEDICINE
Chulalongkorn University



Mahidol University
Faculty of Medicine
Siriraj Hospital

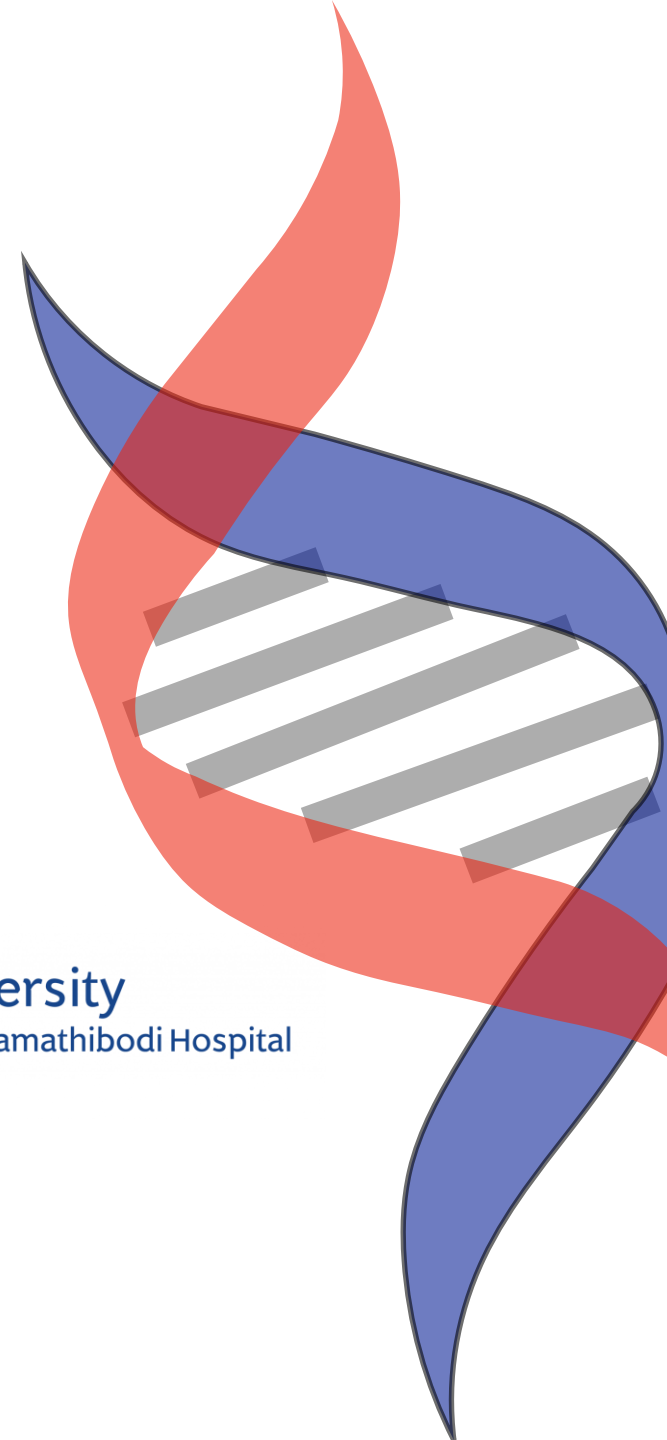


Mahidol University
Faculty of Medicine Ramathibodi Hospital



MDKKU

คณะแพทยศาสตร์ มหาวิทยาลัยขอนแก่น
FACULTY OF MEDICINE KHON KAEN UNIVERSITY





สวทช.
NSTDA

Bioeconomy
Circular economy
Green economy

ขอบคุณครับ

