

A report on recent activities at Korea Bioinformation Center

Seungwoo Hwang

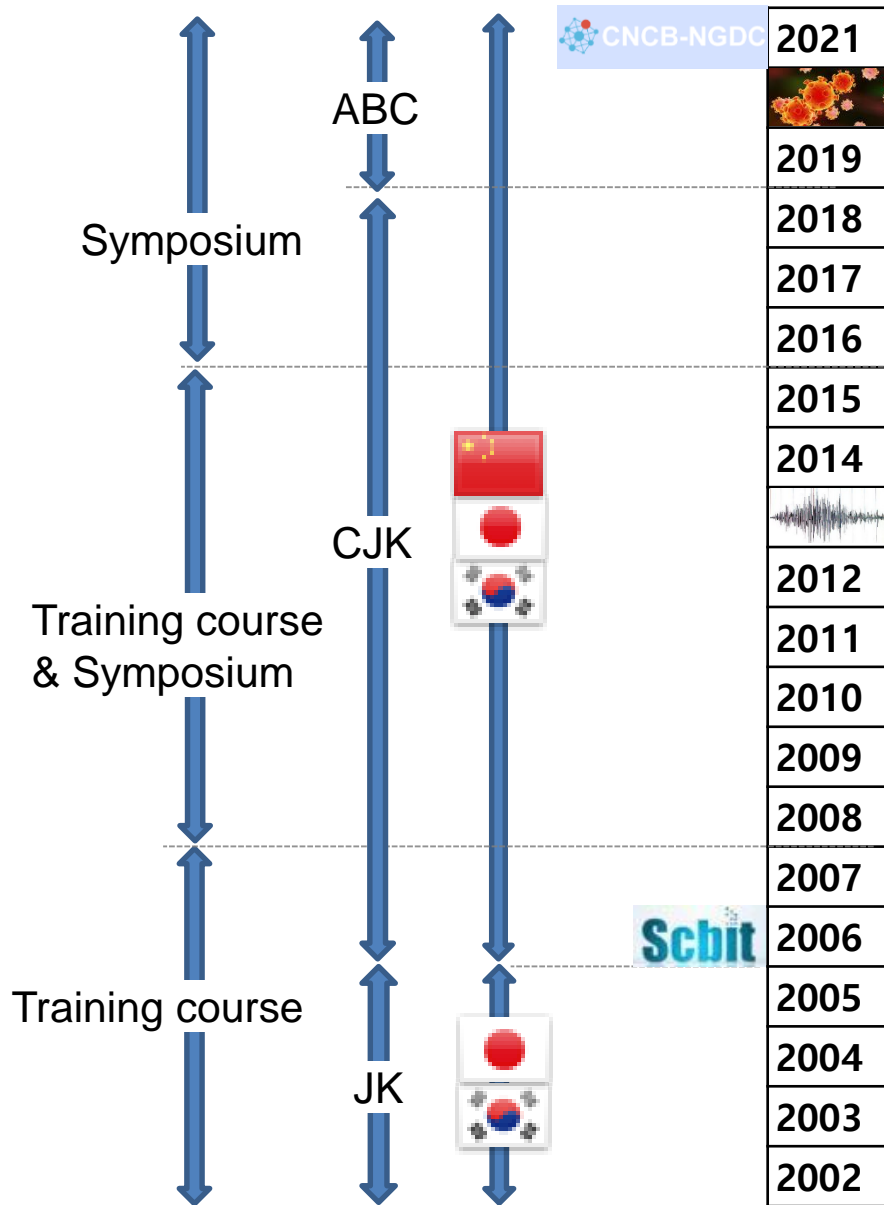
Korea Bioinformation Center



Korea Research Institute of Bioscience and Biotechnology



The history of ABC and KOBIC



kobic Korea Bioinformation Center

NGIC

6. Seon-Young Kim



5. Yong-Kyung Choi



4. Ryan W. Kim



3. Kiejung Park



2. Sanghyuk Lee



1. Jong Bhak



Sangsoo Kim



KOBIC has been skyrocketing since 2019

New building!!!



New server room!!!



More people!!!

(2019. 2.)
27

➔

(Now)
57

* Excluding interns

More funding!!!

(2019. 2.)
~4,800 million KRW
(~40 M USD)

➔

(Now)
~27,700 M KRW
(~320 M USD)

* Including WGS costs

New big projects!!!

1. BioData Station Project
2. National Genome Project

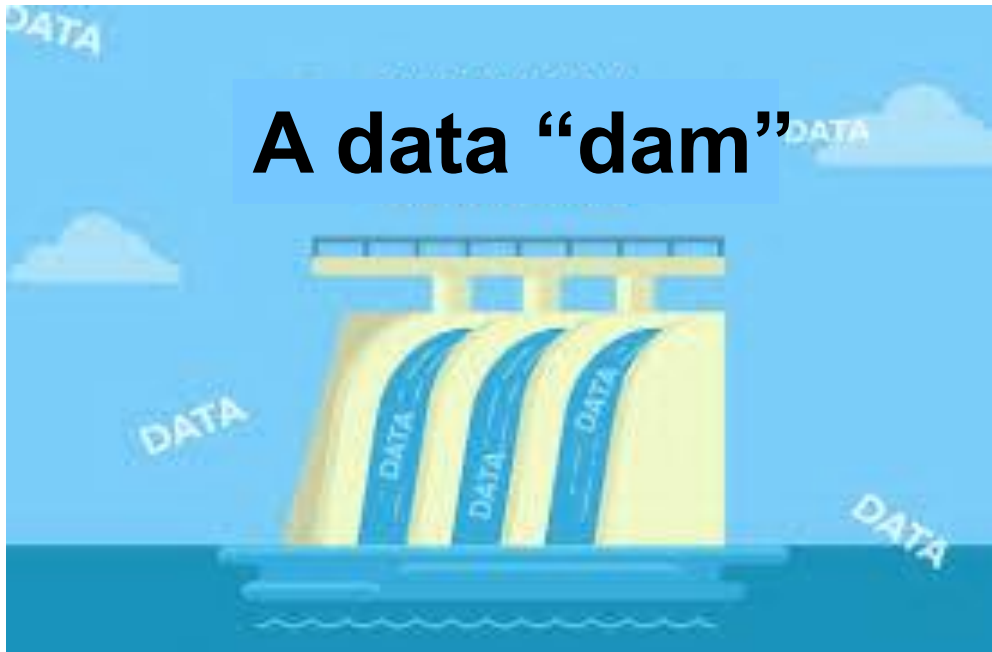
BioData Station Project (2020~)

Goal of BioData Station Project

Long-term goal:

To collect **all types** of biological data (i.e., not just omics) that were produced from **all grants** from government funding

A data “dam”



Why can't just we leave data sharing to journal publishers?

FAQ:

To publish a paper, I already deposit data somewhere (e.g., established international data repositories) after all. So, why do you bother to make another repository in Korea?

Data deposit requirements from journals are surprisingly low

Absolutely required
(not accepted without deposit)

Required

Recommended

Only indirectly mentioned

Mentioning only about omics data

No mentioning at all

- 1 Required as condition of publication, barring exceptions
- 2 Required but no explicit statement regarding effect on publication/editorial decisions
- 3 Explicitly encouraged/addressed, but not required.
- 4 Mentioned indirectly
- 5 Only protein, proteomic, and/or genomic data sharing are addressed.
- 6 No mention

11.9%

9.1%

23.3%

9.1%

14.8%

31.8%

Vasilevsky et al (2017) PeerJ PMID:28462024

Reproducible and reusable research: are journal data sharing policies meeting the mark?

To make data sharing more active, funding agencies need to act too.

For more data sharing, funding agencies need to act too

While journals are international, funding agencies are national.

[Home](#) > [BMIC Home](#) > [NIH Data Sharing Repositories](#)

Open Domain-Specific Data Sharing Repositories

This table lists NIH-supported domain-specific data repositories that make data accessible for reuse and are open for both submitting and accessing data. Submission is typically limited to data of a certain type or related to a certain discipline. The table provides links to information about submitting data to and accessing data from the listed repositories. Repositories in this list have current NIH funding, sustained support, open data submission and access, and open time frame for data deposit, based on information [provided by the repository](#) about funding and data availability. This non-exhaustive list is also available in a [downloadable Excel version](#).

NIH does not list non-US repositories.

Need to establish a policy on data management

Final NIH Policy for Data Management and Sharing

Notice Number:

NOT-OD-21-013

Key Dates

Release Date:

October 29, 2020

Effective Date:

January 25, 2023

To get an approval for NIH-funded grants (of any amount), researchers need to file a Data Management Plan (DMP), and prove that they deposited the data according to the plan.

→ Korea is preparing such a policy.

Need a proper format for data and metadata deposit

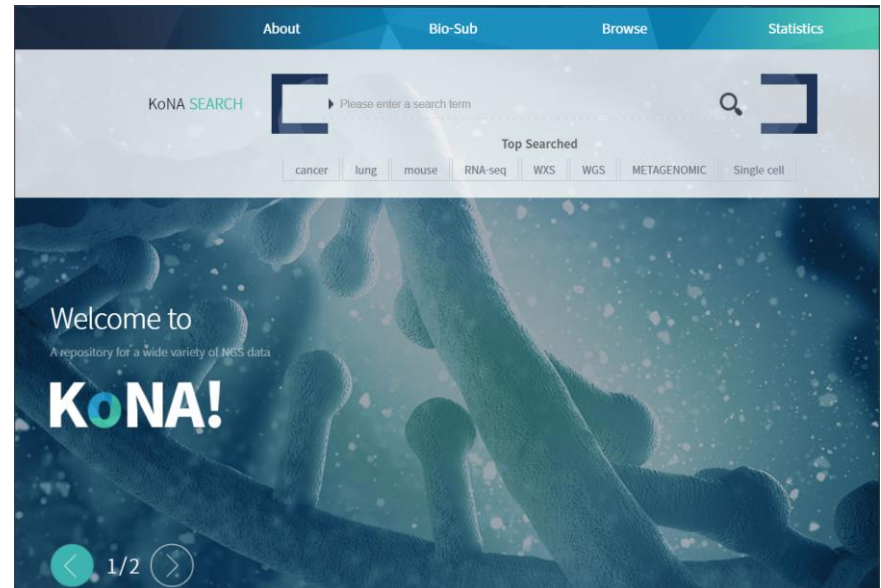
- **To maximize information content of deposited data**
- **Generally following existing formats**
 - **NGS:** INSDC
 - **Microarray:** NCBI GEO
 - **Nucleotide sequence:** NCBI GenBank
 - **Proteomics:** EBI PRIDE
 - **Metabolomics:** EBI MetaboLights
 - **Chemical activity:** NCBI PubChem Bioassay
 - **Microscopy image:** IDR, EBI EMPIAR, EBI BioImage
 - **Medical imaging:** DICOM, BIDS, ADNI
 - **Medical instrument R&D:** ClinicalTrials
 - **Public health:** NCD-RisC, WHO STEPS
 - **Pre-clinical:** SEND
 - **Plant metadata:** UPOV
 - **Other unstructured data:** EBI BioStudies, Dryad, FigShare, Zenodo
 - ...

A pilot system for data deposit is now open

KBDS: Korea BioData Station



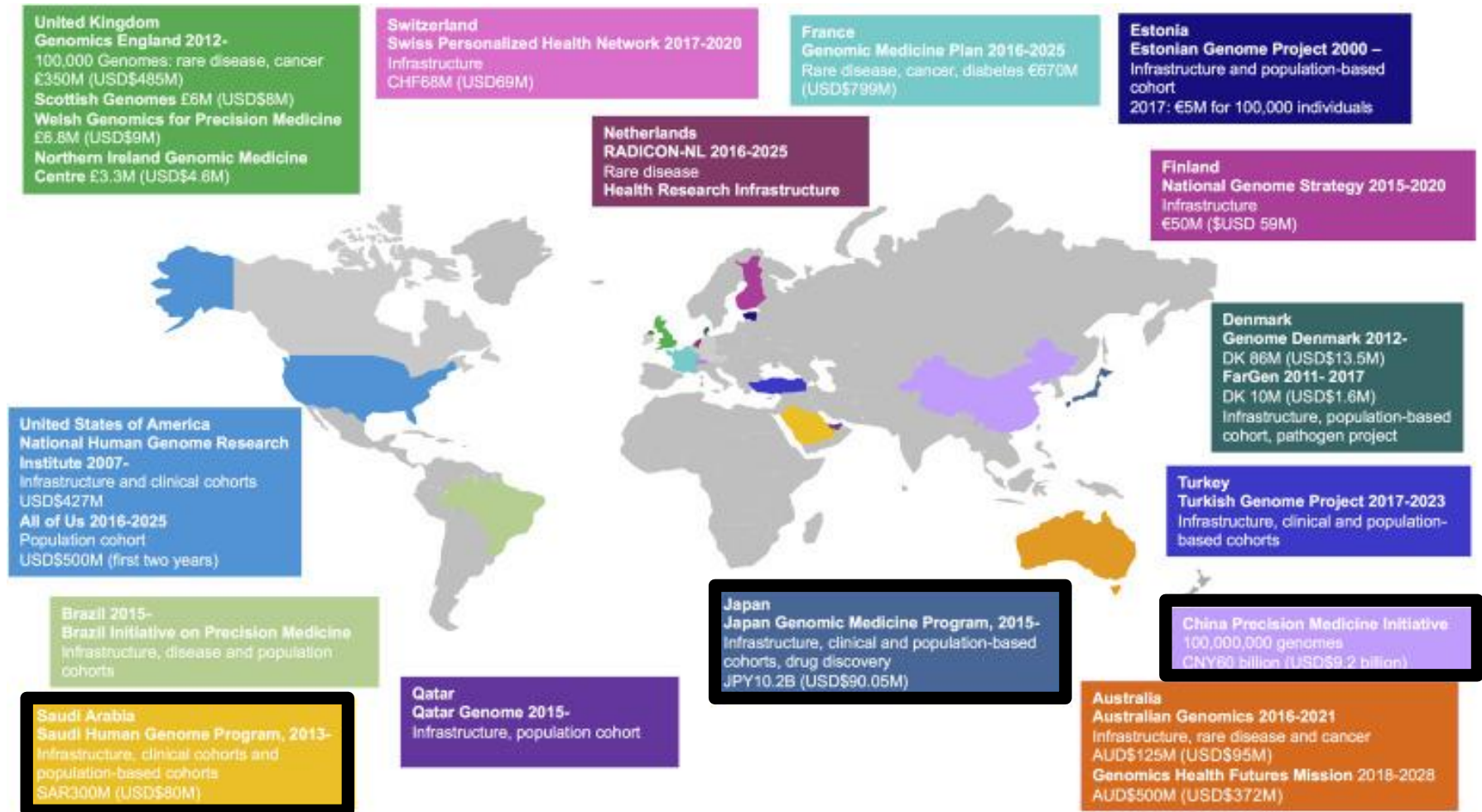
KoNA: Korean Nucleotide Archive



National Genome Project (2020~)

Nation-wide genome projects across the world

14 countries with nation-wide genome projects (as of 2019)



Stark *et al.*, *AJHG* (2019)

National Genome Project (NGP) in Korea

Phase 1. Pilot project (2020~2022)	25,000 genomes
Phase 2. Main project (2023~2028)	1 million genomes

Mission Statement:

To provide large-scale genomic and clinical data
to scientific and industrial community
for the purpose of studying precision medicine,
while protecting the participants' privacy

Overall scheme of the pilot phase of NGP

Collection of
Samples & Data

Data Processing
& Quality Control

Data Analysis
& Sharing

Clinical report & assessment

2 cohorts (newly sequenced by NGP)

Rare Disease (RD)
15,000 (mostly trio)



Normal Participants
2,500 (Korean Genome
Epidemiological Study)

Sequencing data

Clinical info

NGP
consortium

kobic



Sequencing data

Clinical info



Research
Environment
Platform



Computing
resources



Cloud services



Data sharing



Integrative data analysis
in the secure platform

3 cohorts (sequenced prior to NGP)

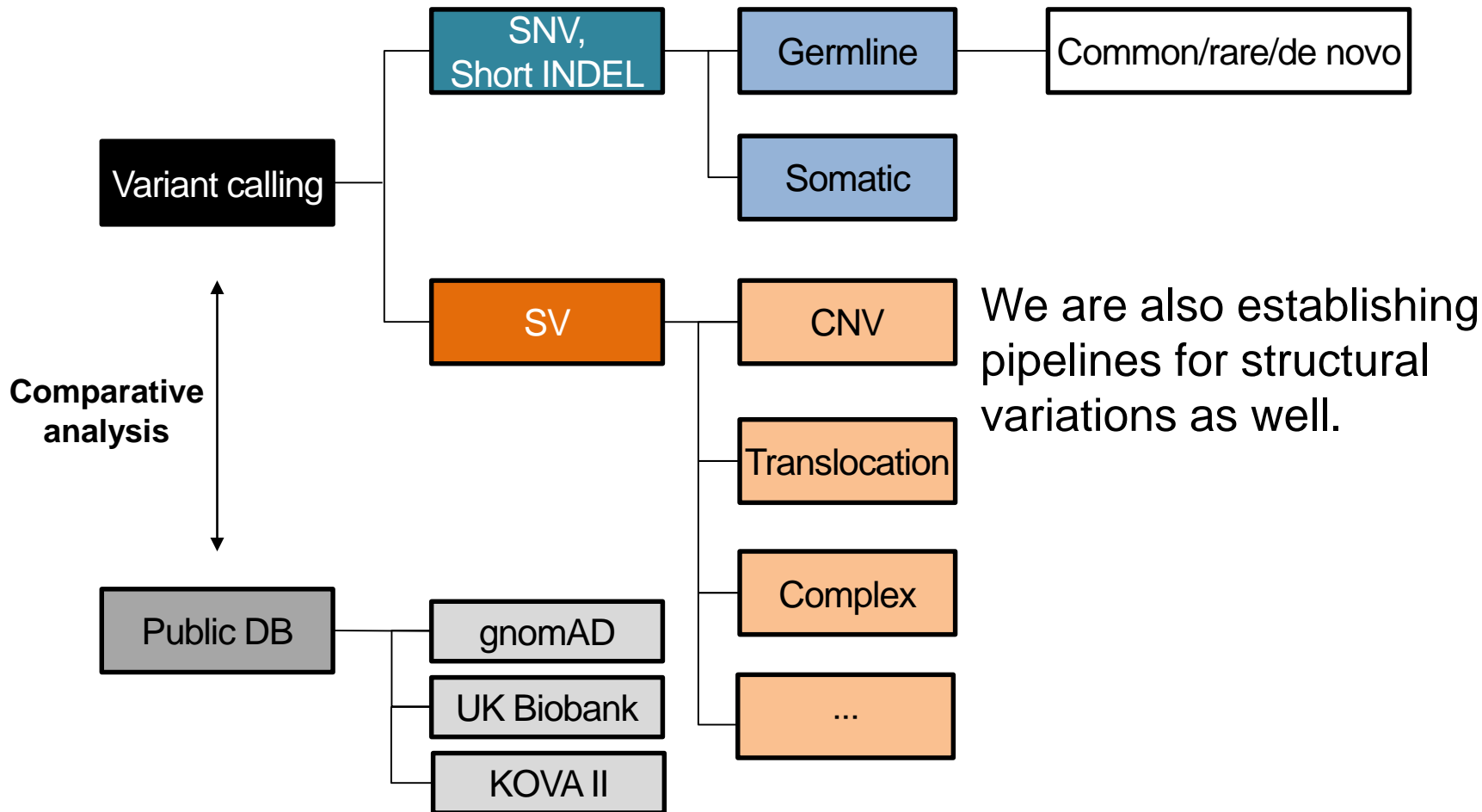
Normal Participants
1,600 (KGP)(Jong Bhak)

**Autism Spectrum
Disorder** (500, familial)

Colorectal Cancer
(200, tumor and
adjacent normal)

Data processing and QC

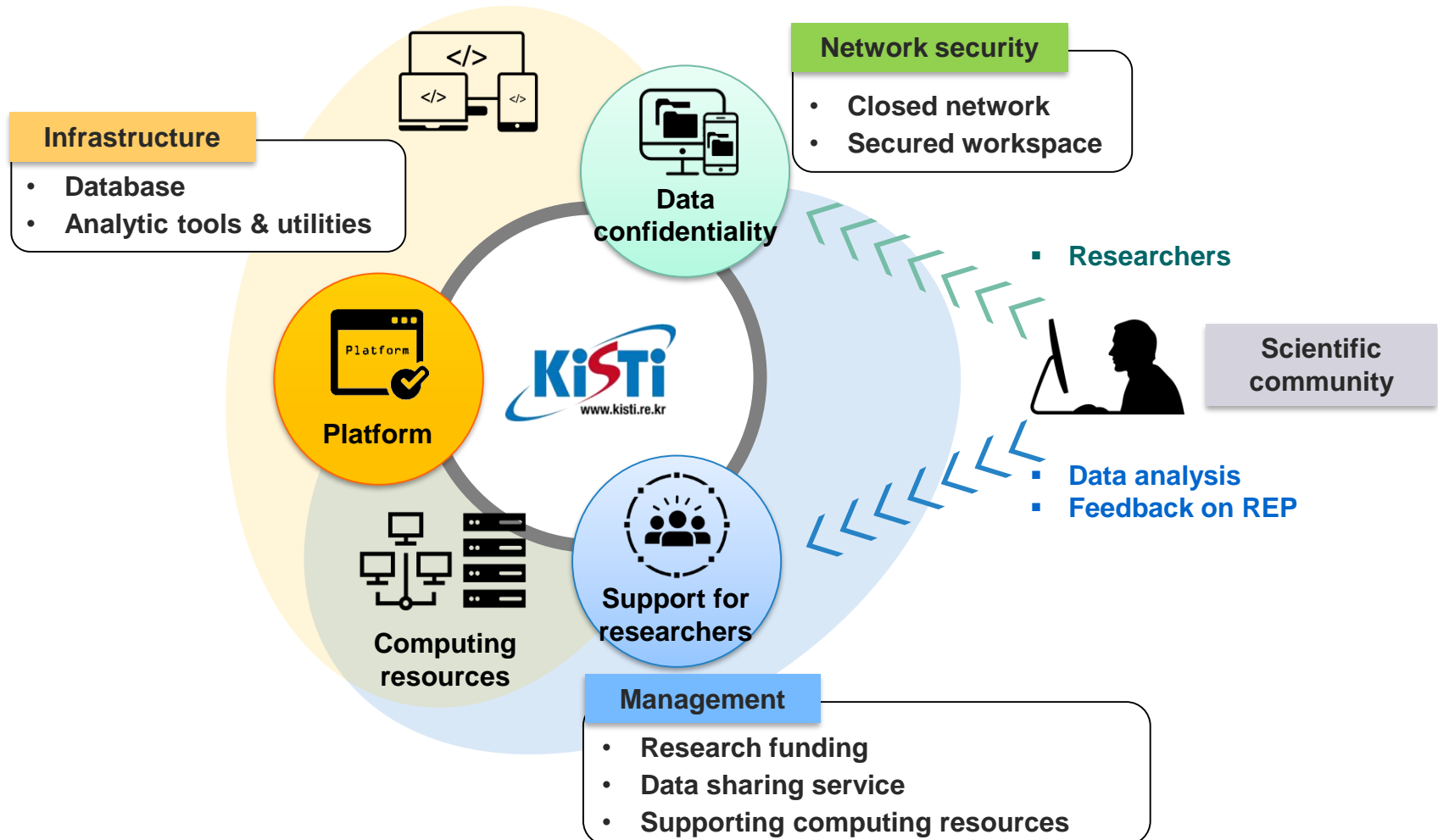
We established data processing and QC pipelines to identify germline or somatic variants with high consistency and confidence.



We are also plan on performing a comparative analysis between our results and other public databases.

Research Environment Platform

Researchers can access and analyze the NGP data (processed WGS data and de-identified clinical information) at the Research Environment Platform, a secure workspace.



Next phase (2023~2028) of NGP

- The next phase of NGP aims to collect samples and data from **1 million participants**:
 - 0.3 million from existing cohorts
 - 0.7 million from newly recruited participants
- The 1 million participants will consist of
 - Rare disease patients
 - Cancer patients
 - Severe incurable disease patients
 - Normal/Healthy participants
- The data collection will include not only WGS, but also multi-omics data as well as longitudinal clinical information and personal health records.
- All these datasets will be available in the expanded Research Environment Platform with dynamic consent system. So the researchers can use the data for a wide range of purposes without any ethical and legal issues.

National Genome Project

Korea BioData Station

May our collaboration continues!

Since 2002

ABC (Asian Bioinformatics Consortium) Symposium

Home

Home

About ABC Symposium

2021

2019

History in Brief

2018

2017

2016

2015

2014

2013

2011

2010

2009

2008

Main Organizers

2007

2006

2005

2004

2003

2002






In 2001, the 4th Korea-Japan Science and Technology Forum was held in Seoul. One of its agreements was that both countries should collaborate to hold a bioinformatics training course in order to educate students in this field from the two countries. On the basis of this agreement, the first training course was held in Japan as "The 1st Japan-Korea Bioinformatics Training Course" in 2002. This annual course continued between Japan and Korea until the 4th course in 2005.

From the 5th course in 2006, it expanded to include China as well, and has continued ever since. In 2015, its nature has been changed from training course to symposium, and was thus named as "China-Japan-Korea Bioinformatics Symposium". Since the three countries take turns hosting the annual course, the alphabetical order of the C, J, and K is rearranged each year, in the order of hosting country of the current year, that of the next year, and that of the year after next.

From the 17th symposium in 2019, it was renamed as ABC (Asian Bioinformatics Consortium) Symposium.

From the 18th symposium in 2021, the organizing institute of China was changed from SCBIT (Shanghai Center for Bioinformation Technology) to CNCB-NGDC (National Genomics Data Center, part of China National Center for Bioinformation).

The ABC Symposium is organized by the following institutions.

China	Japan	Korea
CNCB-NGDC (National Genomics Data Center, part of China National Center for Bioinformation)	DDBJ (DNA Data Bank of Japan) at NIG (National Institute of Genetics)	KOBIC (Korea Bioinformation Center) at KRIBB (Korea Research Institute of Bioscience and Biotechnology)
	 DNA Data Bank of Japan	 Korean Bioinformation Center
	 National Institute of Genetics Research Organization of Information and Systems	 Korea Research Institute of Bioscience & Biotechnology

<http://cjk-bioinfo.org>