Genome Korea

10,000 Korean Genomes

KOGIC

Clinomics

울산광역시
ULSAN METROPOLITAN CITY

울산과학기술원
ULSAN NATIONAL INSTITUTE OF SCIENCE AND TECHNOLOGY

KOGIC
KOrean Genomics Center
게 놈 산 업 기 술 센 터

KOGIC
KOrean Genomics Industrialization and Commercialization Center
게 놈 산 업 기 술 센 터

게놈산업기술센터는

게놈기반 바이오 헬스케어 연구, 기술개발, 상업화를 지원하고, 실용적인 게놈기반 원천기술 상용화에 집중하기 위해 센터, 기업, 지역이 연합하여 수익성 창출과 글로벌 시장 개척에 초점을 맞춘 핵심 인프라 역할을 수행합니다.

대형 게놈 사업을 성공적으로 추진하고
미래 게놈 산업을 확대하는 것을 목표로 하고 있습니다.

대형 게놈 사업기반 빅데이터 프로젝트 추진

한국인 표준게놈
첨단게놈기술 확보 사업

- 게놈코리아/울산 게놈 프로젝트
- 게놈 빅데이터 지능적 처리기술
- 게놈기반 맞춤의료 인프라 구축

게놈기반 바이오메디컬 기업과의 산학밀착형 연구개발

게놈.바이오메디컬 기업과의
산학밀착형 연구개발

- 인프라기술지원(중소기업을 위한 센터)
- 게놈원천 기술개발
- 진단예측 기기/기술 개발

게놈 융합 사업의 핵심 인프라 역할

게놈 융합 사업의
핵심 인프라 역할

- ICT-바이오지능형(AI) 융합 산업의 허브 역할
- 부울경 동남권게놈산업 발전의 촉진제 역할
- 게놈기반 바이오메디컬 산업 핵심 인프라 역할

게놈산업기술센터 장비 현황

게놈분석 슈퍼컴퓨팅 장비 2021.03 기준

CPU : 4,450core
GPU : 9,000core
Memory : 40,324GB (39.38TB)
Storage : 14,170TB (13.84PB)

게놈분석 실험장비

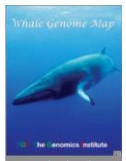
게놈분석 실험장비 : 25종 37여대 구축

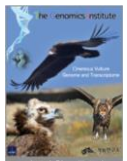
(나노포어 해독기, 시료보관용 냉동고, 세포조직배양기, 단일 세포 분류기, 위상차현미경, 중합효소연쇄반응기, 핵산 분석/정량장치, 역상형광 현미경이미지 시스템, 유전자증폭시스템, 분자병리진단 현미경 등)

KOGIC | 게놈산업기술센터 협력기관

울산광역시 · 울산대학교
울산대학교병원 · 연세대학교 의과대학 송당암연구센터
부산대학교 · 제주대학교 제주국제동물연구센터
DGIST 대구경북과학기술원 융합연구센터 · 아주대학교의료원 분산형 바이오헬스 빅데이터사업단
농촌진흥청 포스트게놈다부처유전체사업단 · 한국식품연구원
한국한의학연구원 미병연구단 · 안전성평가연구소
(주)테라젠이텍스 · (주)클리노믹스
(주)힐릭스코 · (주)데이타커맨드


최초의 호랑이 표준게놈지도 완성


최초의 고래 표준게놈지도 완성


최초의 한국 독수리 게놈 해독


최초의 한국 표범 표준게놈지도 완성


최초의 황금박쥐 게놈 해독


노무라 입깃 해파리 표준게놈 지도 완성

2013 – 2014 – 2015 – 2016 – 2017 – 2018 – 2019 – 2020

한국인 41명의 게놈 정보를 통합해 한국인 고유 특징을 분석한 **한국인 표준게놈지도(KOREF) 공개**


최초로 신석기 동아시아인 고대 게놈 해독 한국인의 유전적 뿌리 규명


KoVariome: 한국인 변이체


한국인 1천명 게놈 (Korea1K) 분석

Genome Korea Project

Aims:

- To characterize the non-pathological ethnicity-relevant genetic variation and **diversity present in the Korean population**
- To predict candidate loci, inheritance patterns, and genetic risk for certain diseases
- To improve the **consensus Korean Reference** genome standard
- To identify Korean population-specific **pathogenic rare variants**

Milestones

2014. 7. UNIST Genome Center

2014. 12. Stressomics

2015. 11. Ulsan 10,000 Genomes

2016. 2. 10k Korean genome Started

2016. 9. Welfare Genome Project

2016. 11. Korean Reference (KOREF)

2017. 6. UNIST KOGIC Established

2018. 4. Korean Variome

2018. 8. Genome Expo 2018 개최

2019. 8. Genome Expo 2019

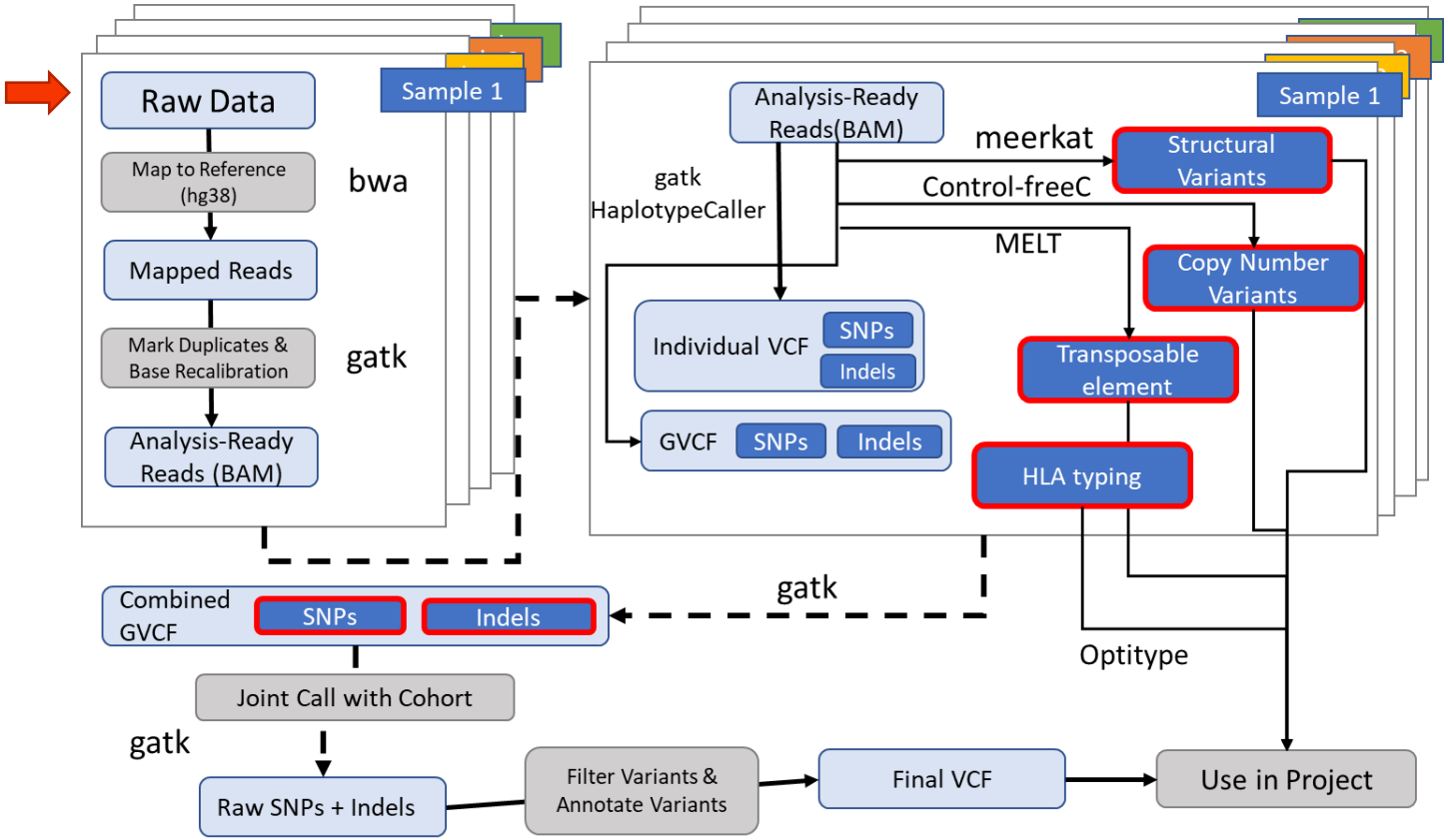
2020. 5. Korea1K *Sci. Adv.*

2020. 7. Biodatafarm project

2021. 04. 26. Korean **10,000명** Whole Genome Sequencing Completed

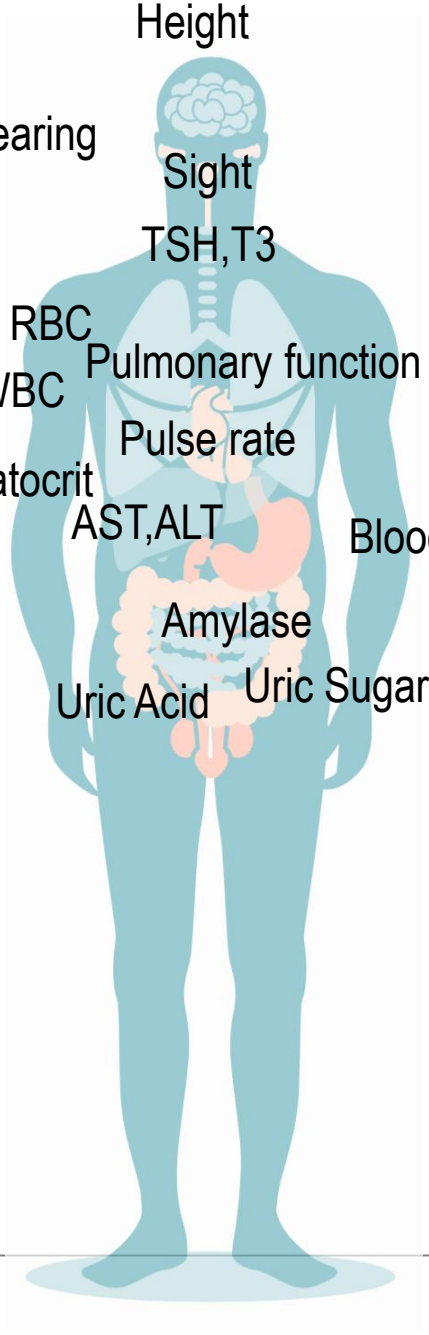
Genome Korea in Ulsan [2016~2021.04]

Whole Genome Processing Pipelines



Whole Genomes: SNP, Indel ,HLA type, Transposable element, Copy number variants, Structural variants

Clinical information

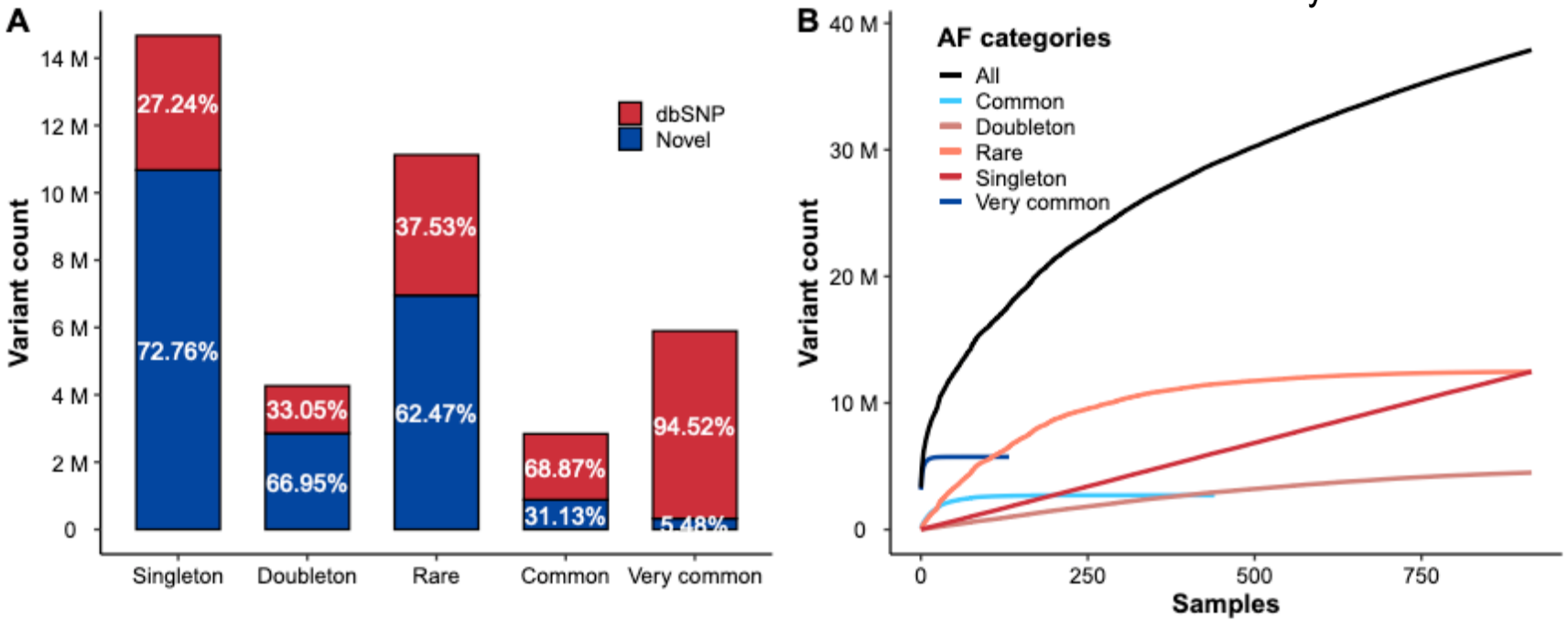


Variant Statistics

Rare:≥AC=3,<1%

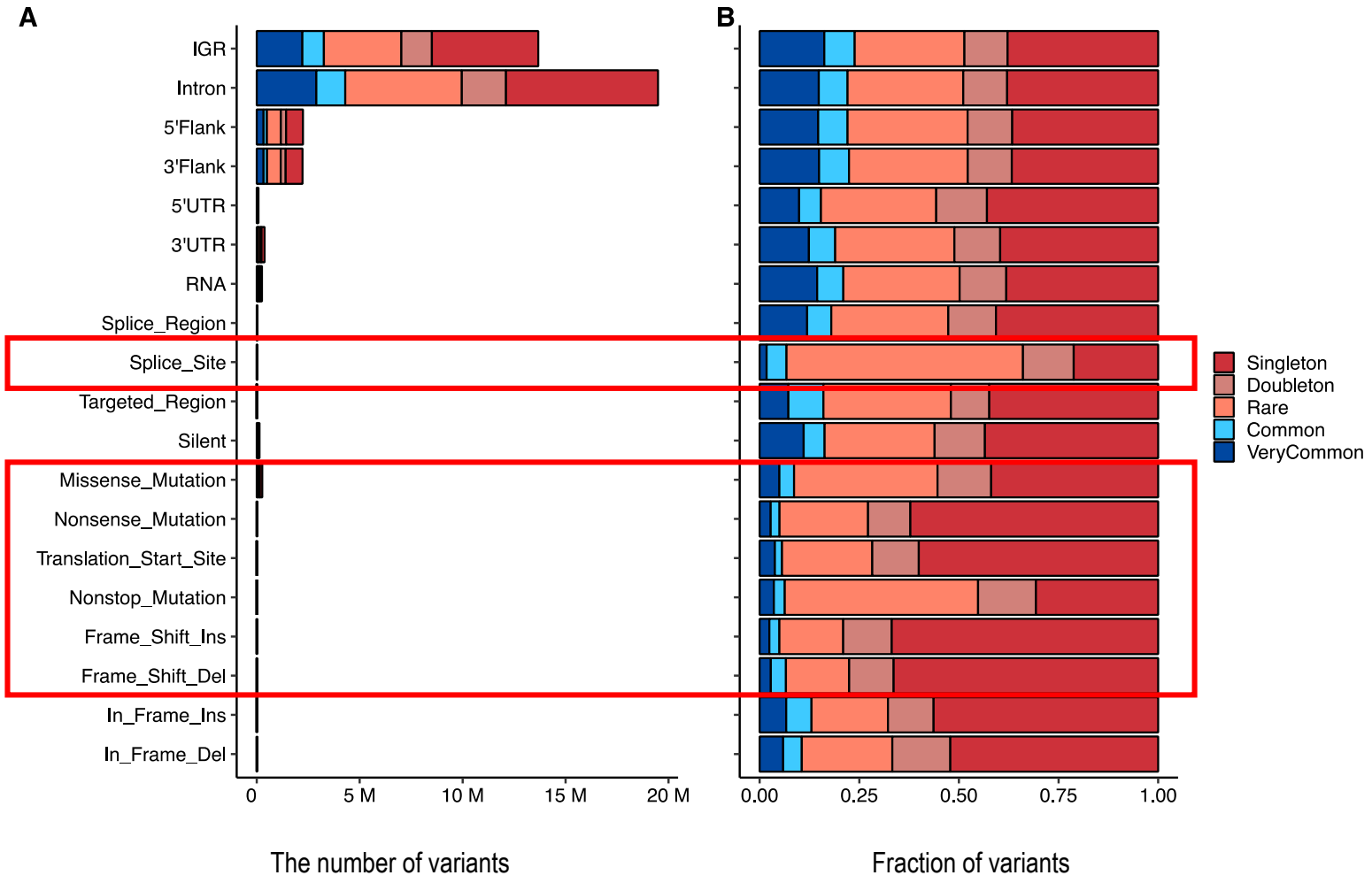
Common:≥1%,<5%

Very Common:>5%



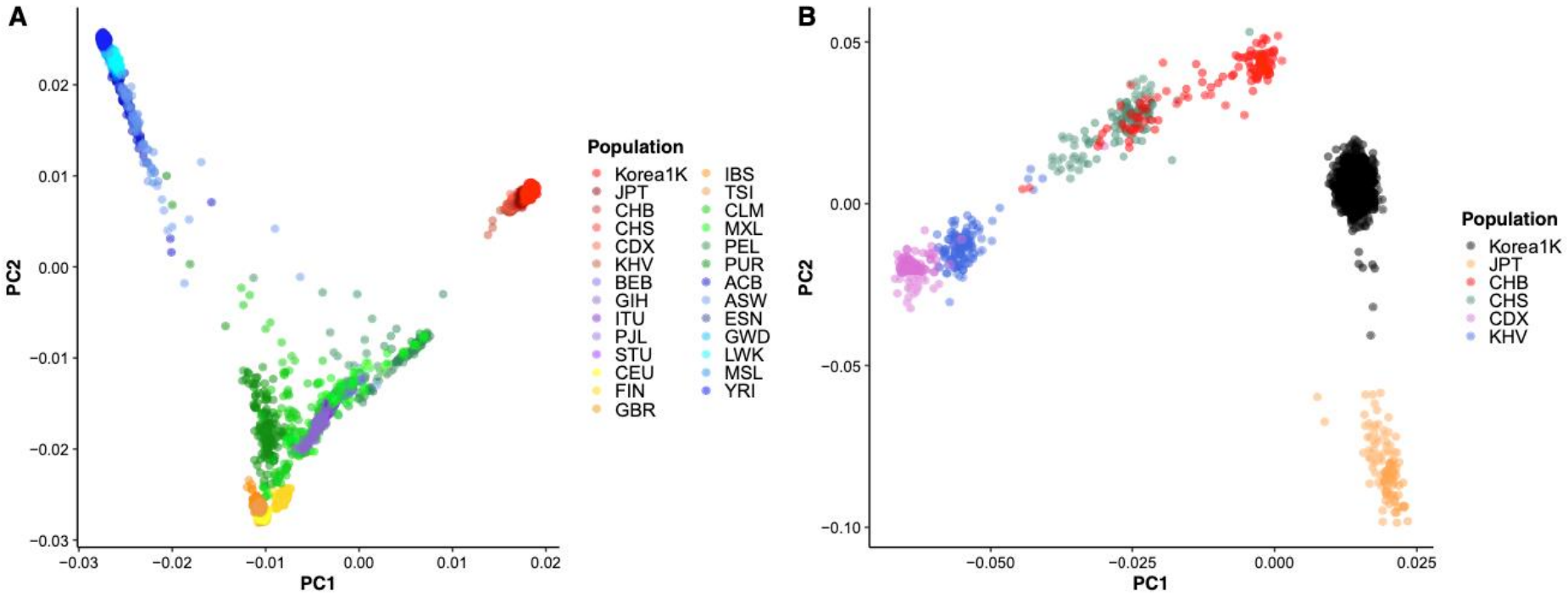
- 39,025,362 variants (13,462,495 singleton) from 1,094 Koreans and most of them are rare variants, although Korean population is known to be homogenous
- Korea1K covered most of very common variants, while more samples are needed to cover very rare variants.

Variant Distribution by Variant Class



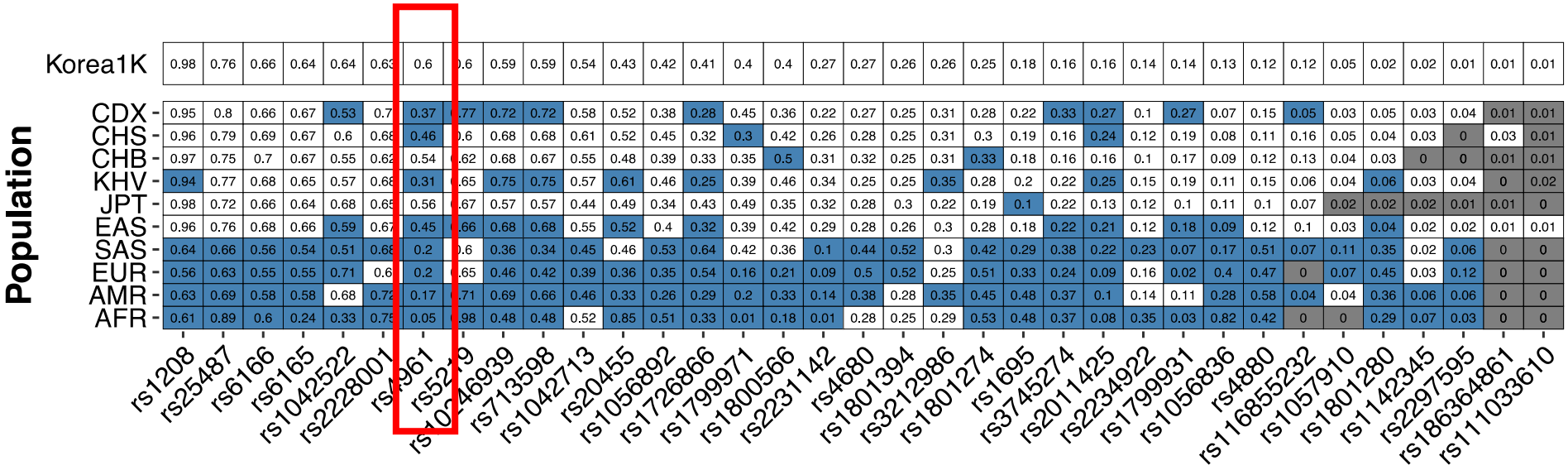
Possible deleterious mutations have higher proportion of rare variants (purifying selection)

Principle Component Analysis with 1KGP data



Koreans can be distinguished from the Chinese and Japanese.
Koreans are a fairly homogenous population.a

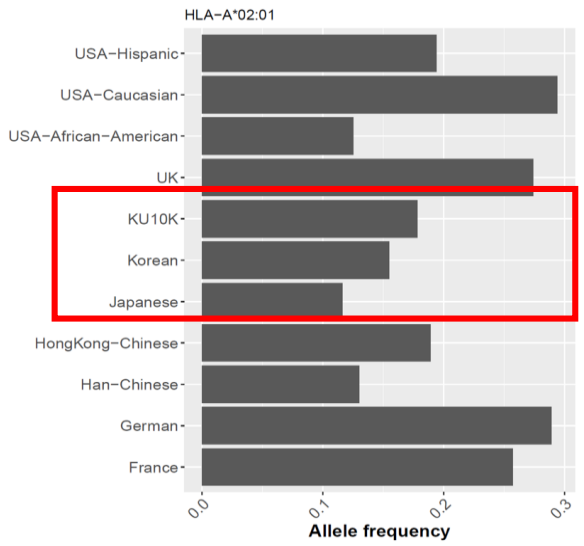
Drug response variants found in Korea1Ka

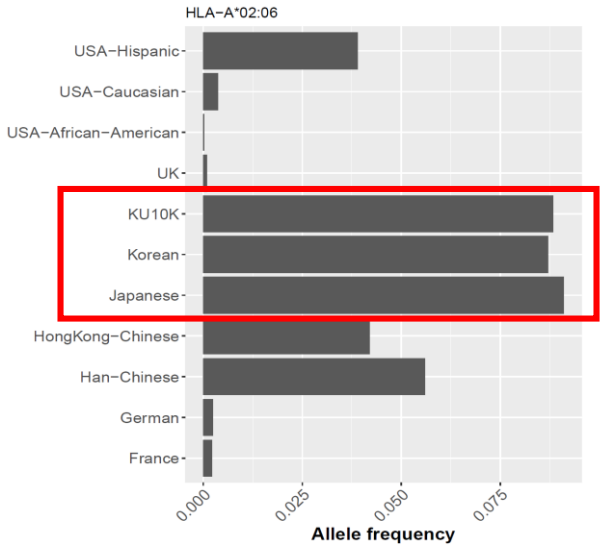


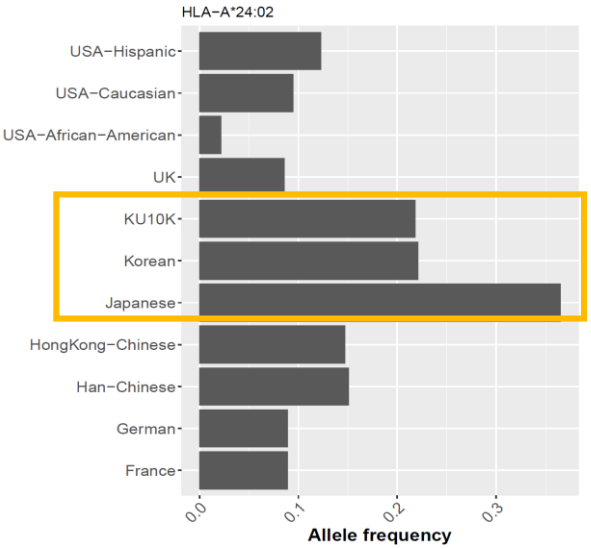
We also found 35 drug- response variants annotated in ClinVar (fig. S16), and ***11 of them displayed significantly different allele frequencies from those of the Chinese or Japanese individuals in the 1KGP set,*** *highlighting* ***the importance of population-specific datasets when interpreting pathogenic or drug-response variants****.*
For example, the variant rs4961 in ADD1 had the highest frequency in the Korea1K compared to other populations and is associated with hypertension and responsiveness to furosemide and spironolactone as shown in a European study.

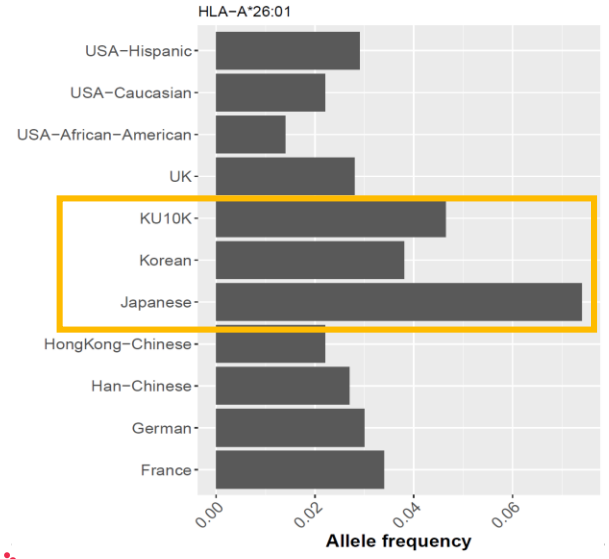
HLA typing by WGS dataa

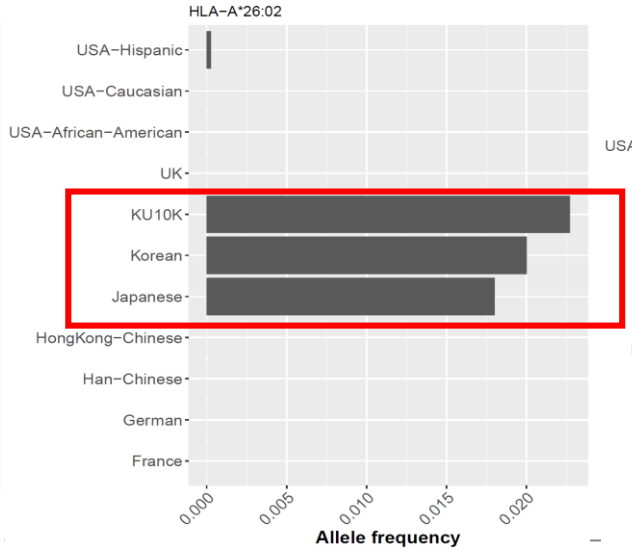
Korean population have different HLA allele frequency pattern with the Japanese, although they are known to be genetically very close

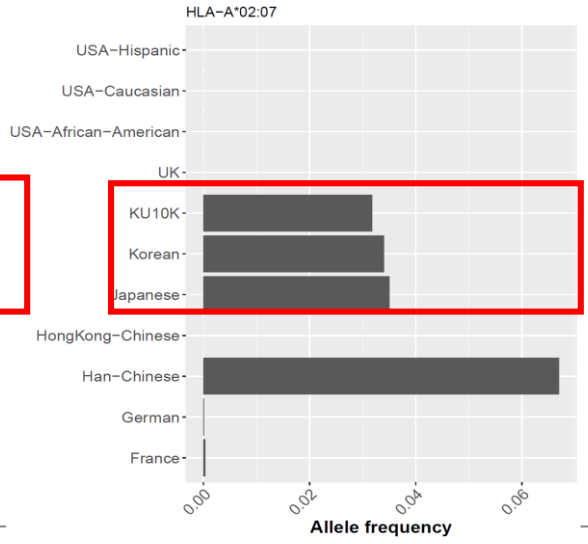










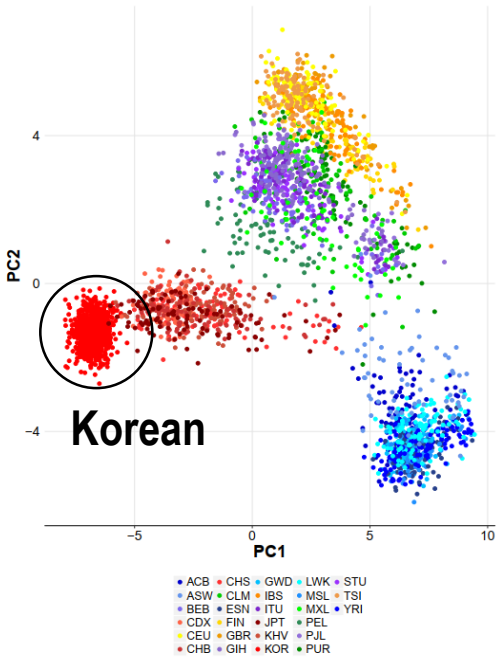


KOGIC
KOrean Genomics Center

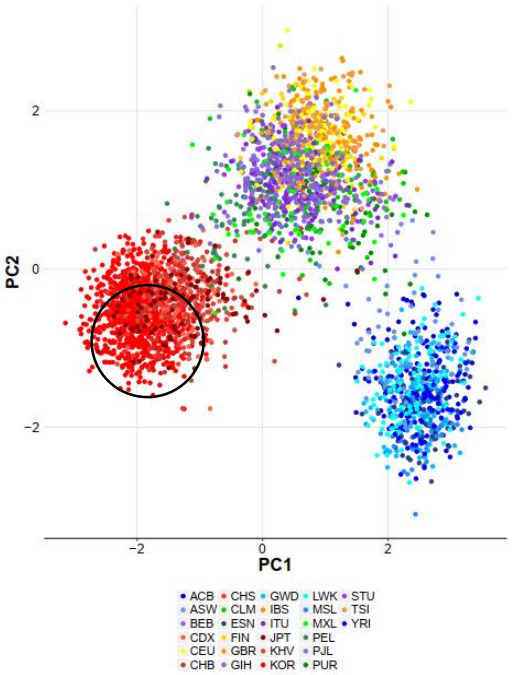
Other ethnic groups: http://www.allelefrequencies.net/gold.aspx

Population stratification confirmed by PCA with TEs in Korea1Ka

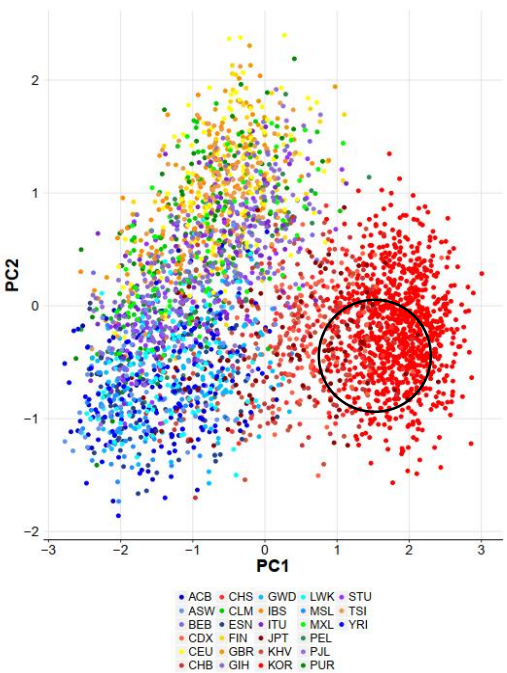
ALU

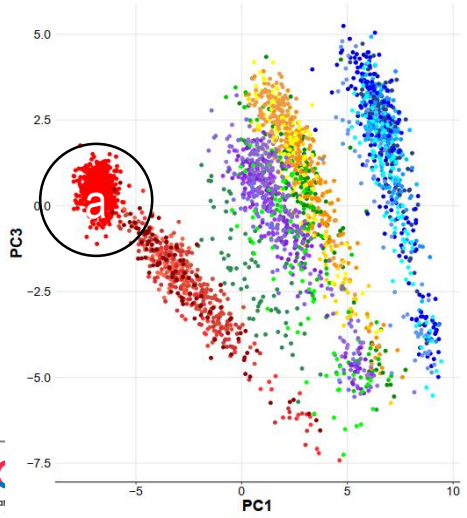


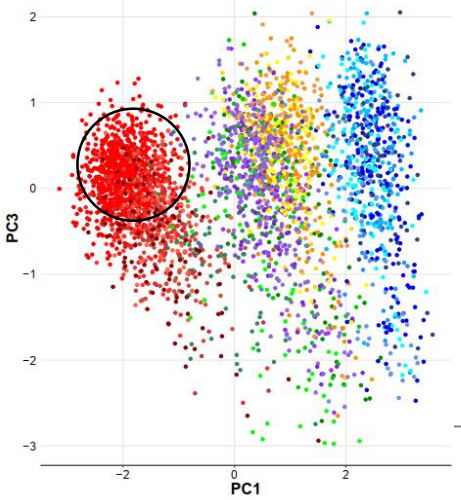
LINE1

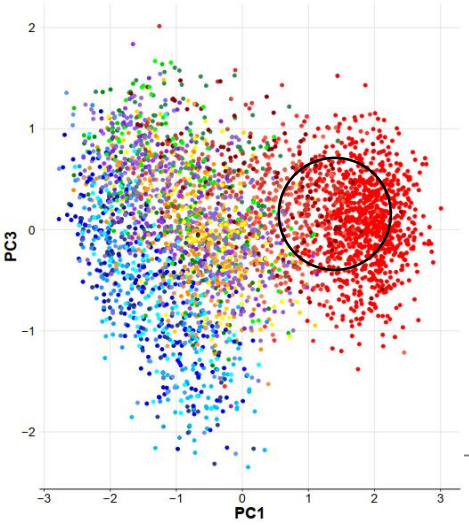


SVA





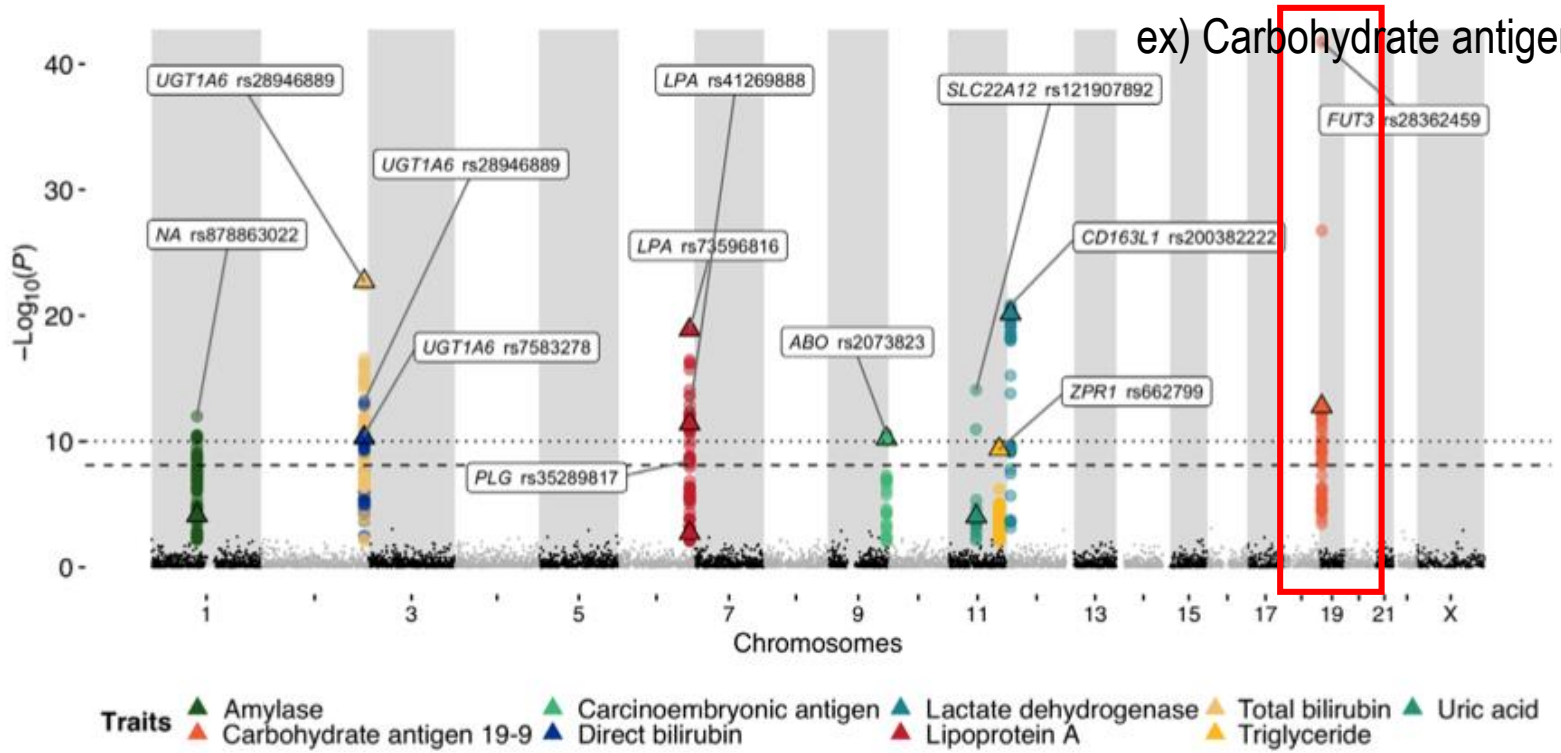




Whole-genome sequencing based GWAS identified more significant GWAS marker than chip-based

*Among the 11 loci of reported variants, 9 contained variants reported in the GWAS catalog, but **their index variants were newly identified in this study**.*

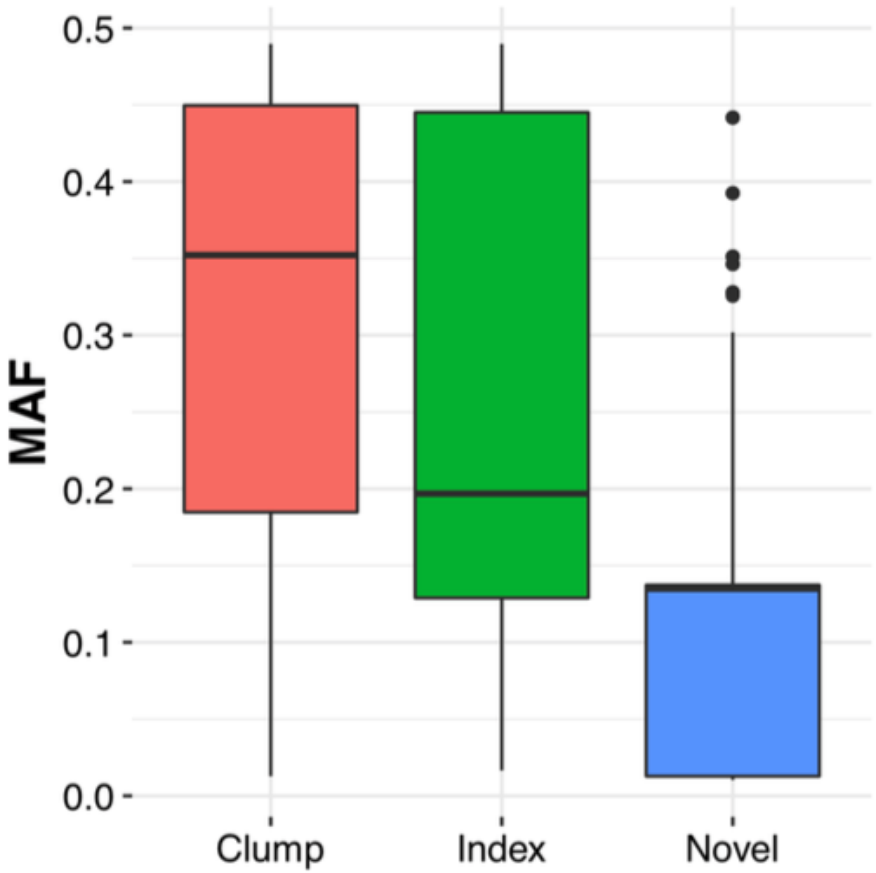
ex) Carbohydrate antigen 19-9



GWAS variants

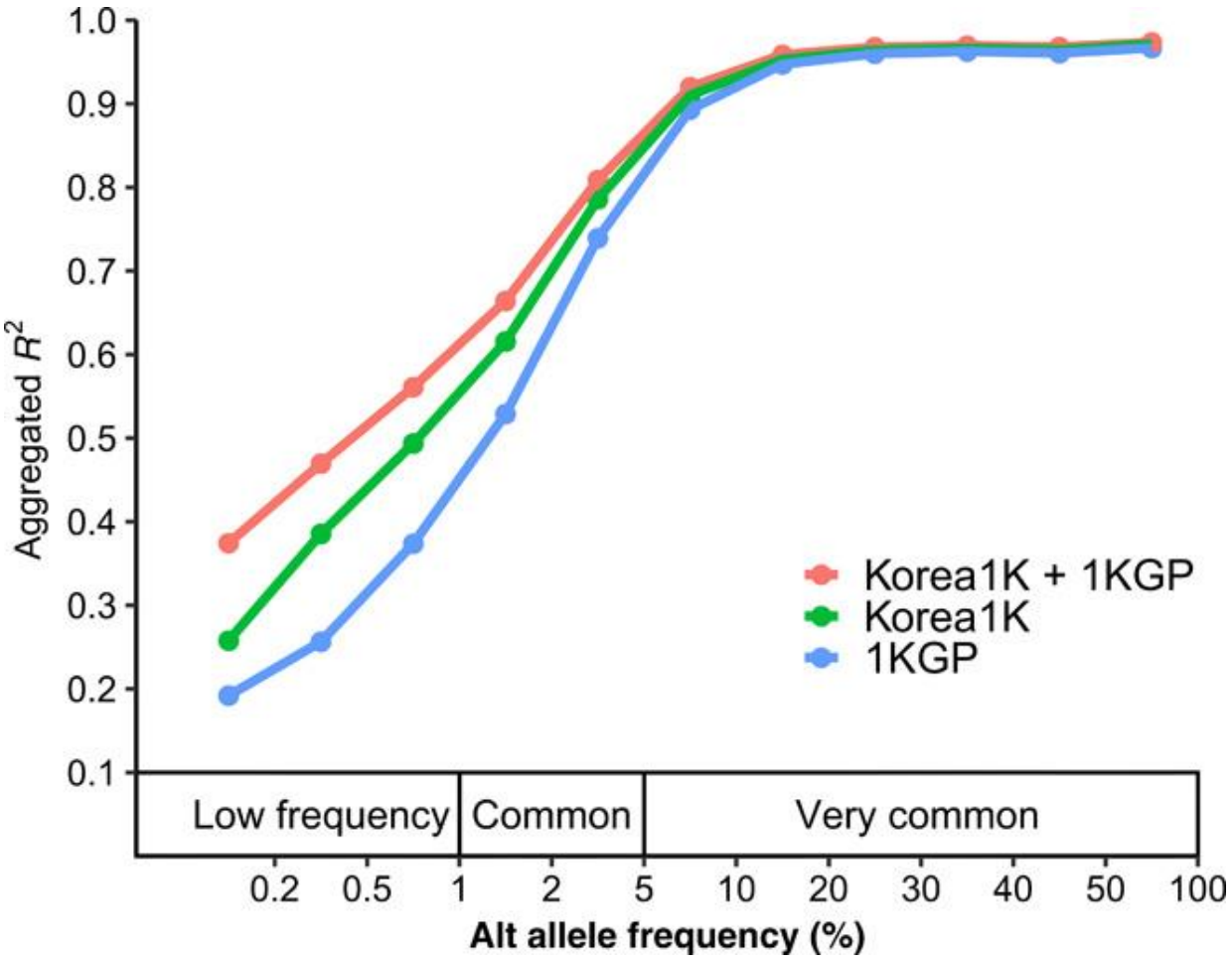
The most significant variants within previously reported markers

WGS data help identify low-frequency alleles and unreported GWAS loci



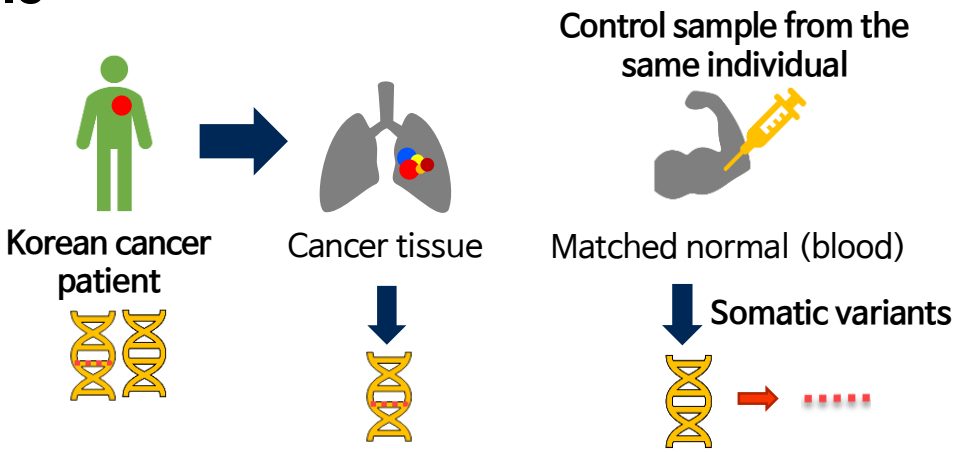
The MAFs in the previously unidentified loci were markedly lower than those previously reported when we compared the MAFs of GWAS variants in these previously reported and the unreported loci (fig. S29). This means ***that large-scale variomes from WGS data help identify low-frequency alleles and unreported loci via whole genome–based GWA studies****.*

Korea1K imputation panel



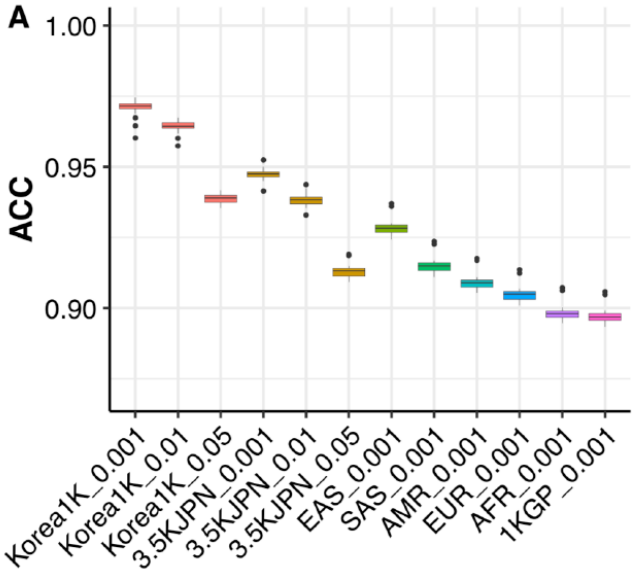
"The accuracy of imputation is improved when a population-specific reference panel is used."

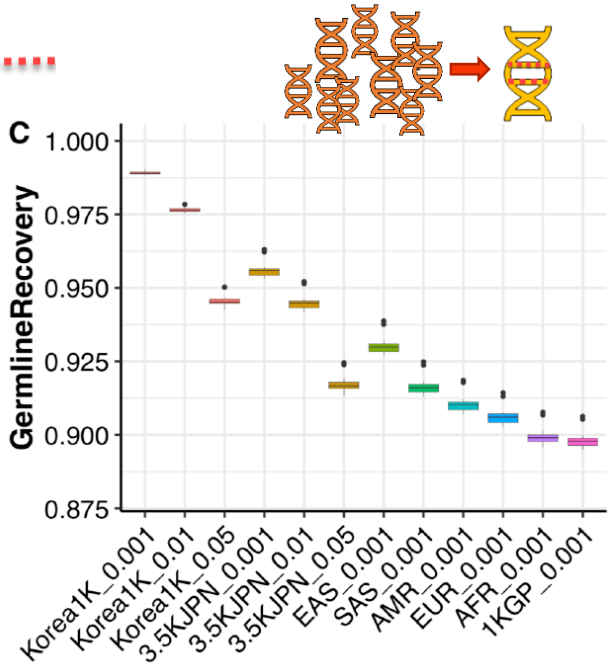
Performance of the variant classification using different panels of normals



Variome DB (Panel of normals)

Somatic variants is usually rare in the population





Although the 3.5KJPN set contained the largest number of variants, the Korea1K dataset had the highest accuracy of prediction of germline and somatic variants.

Publication of Korea1K

SCIENCE ADVANCES | RESEARCH ARTICLE

HUMAN GENETICS

Korean Genome Project: 1094 Korean personal genomes with clinical information

Sungwon Jeon[1,2]*, Youngjune Bhak[1,2,3]*, Yeonsong Choi[1,2]*, Yeonsu Jeon[1,2], Seunghoon Kim[1,2], Jaeyoung Jang[1], Jinho Jang[1,2], Asta Blazyte[1], Changjae Kim[1,3], Yeonkyung Kim[1], Jungae Shim[1], Nayeong Kim[1], Yeo Jin Kim[1], Seung Gu Park[1], Jungeun Kim[4], Yun Sung Cho[3], Yeshin Park[3], Hak-Min Kim[1,2,3], Byoung-Chul Kim[3], Neung-Hwa Park[5,6], Eun-Seok Shin[7], Byung Chul Kim[3], Dan Bolser[3], Andrea Manica[8], Jeremy S. Edwards[9], George Church[10†], Semin Lee[1,2†], Jong Bhak[1,2,3,4†]

We present the initial phase of the Korean Genome Project (Korea1K), including 1094 whole genomes (sequenced at an average depth of 31×), along with data of 79 quantitative clinical traits. We identified 39 million single-nucleotide variants and indels of which half were singleton or doubleton and detected Korean-specific patterns based on several types of genomic variations. A genome-wide association study illustrated the power of whole-genome sequences for analyzing clinical traits, identifying nine more significant candidate alleles than previously reported from the same linkage disequilibrium blocks. Also, Korea1K, as a reference, showed better imputation accuracy for Koreans than the 1KGP panel. As proof of utility, germline variants in cancer samples could be filtered out more effectively when the Korea1K variome was used as a panel of normals compared to non-Korean variome sets. Overall, this study shows that Korea1K can be a useful genotypic and phenotypic resource for clinical and ethnogenetic studies.

https://advances.sciencemag.org/content/6/22/eaaz7835

http://1000genomes.kr/

Korea1K Variome

This webpage is for data sharing of 1K Korean human genomes.

For allele frequency data

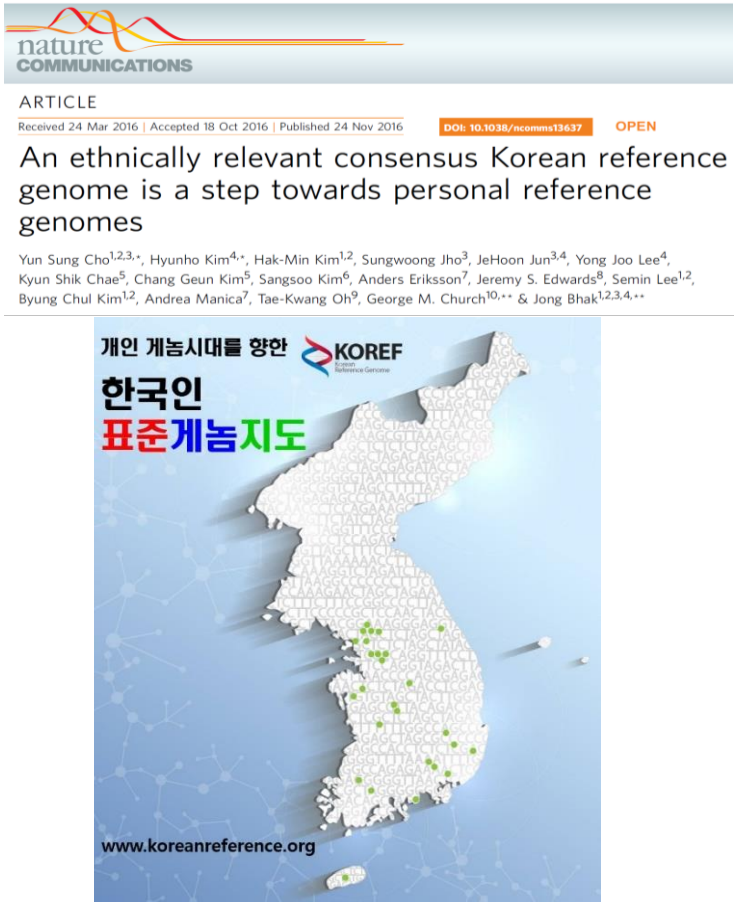
You can download the allele frequency of SNVs and Indels, Copy Number variants (CNV), Transposable element (TE) insertions of Korea1K genomes for academic purposes. For the method about calling variants, please refer to the publication (S. Jeon, Y. Bhak, Y. Choi, Y. Jeon, S. Kim, J. Jang, J. Jang, A. Blazyte, C. Kim, Y. Kim, J. Shim, N. Kim, Y. J. Kim, S. G. Park, J. Kim, Y. S. Cho, Y. Park, H.-M. Kim, B.-C. Kim, N.-H. Park, E.-S. Shin, B. C. Kim, D. Bolser, A. Manica, J. S. Edwards, G. Church, S. Lee, J. Bhak, Korean Genome Project: 1094 Korean personal genomes with clinical information. Sci. Adv. 6, eaaz7835 (2020).).

☐ I agree with the terms and conditions

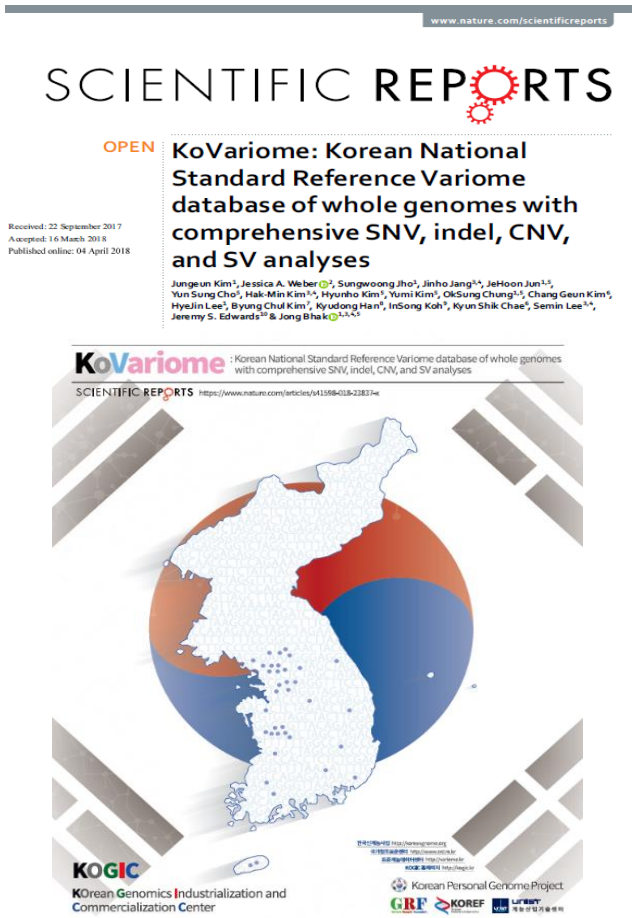
Korea1K SNP and Indel Allele frequency data
Korea1K Copy Number Variants Allele frequency data
Korea1K TE insertion Allele frequency data

한국인 참조표준 게놈(KOREF), 표준 변이체 (KoVariome)

❖ 고해상도의 한국인 표준 게놈지도(KOREF) 확립

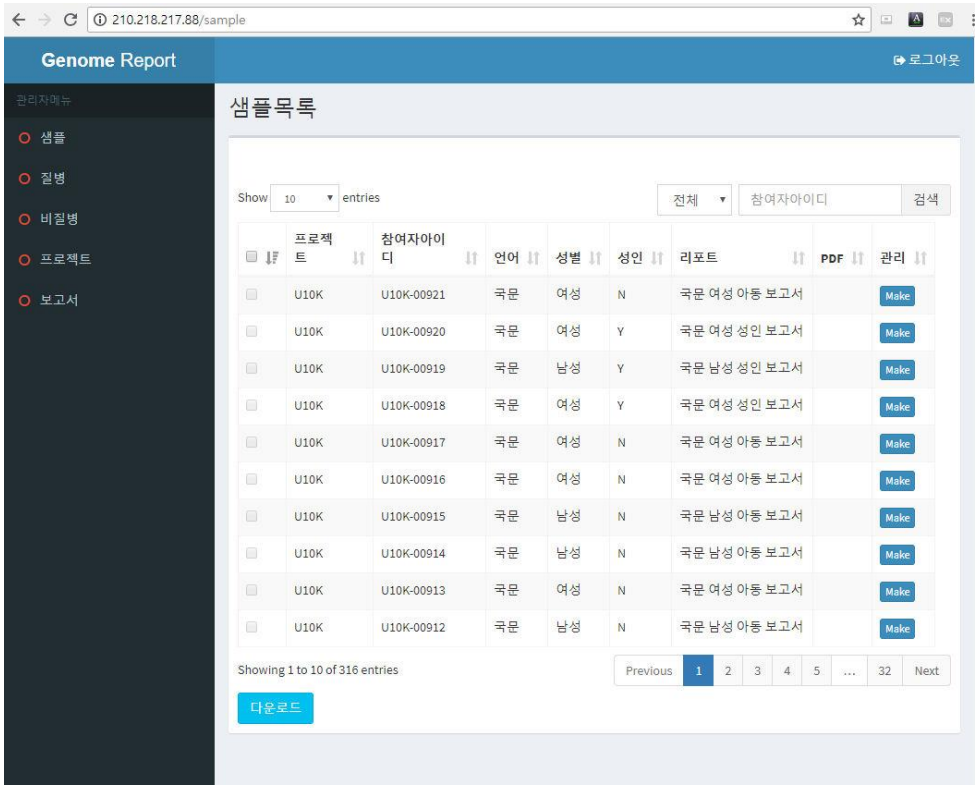


❖ 한국인 참조표준 변이체(KoVariome) 고도화



게놈분석 연구리포트 개발 및 리포팅 자동화 시스템 구축

수백명 단위의 대규모 연구 참여자들에게 **게놈 연구 분석 결과 정보인 게놈 분석 리포트**를 혜택으로 제공하기 위해 **대량의 결과 데이터를 효율적으로 리포팅 할 수 있는 자동화 시스템 구축**



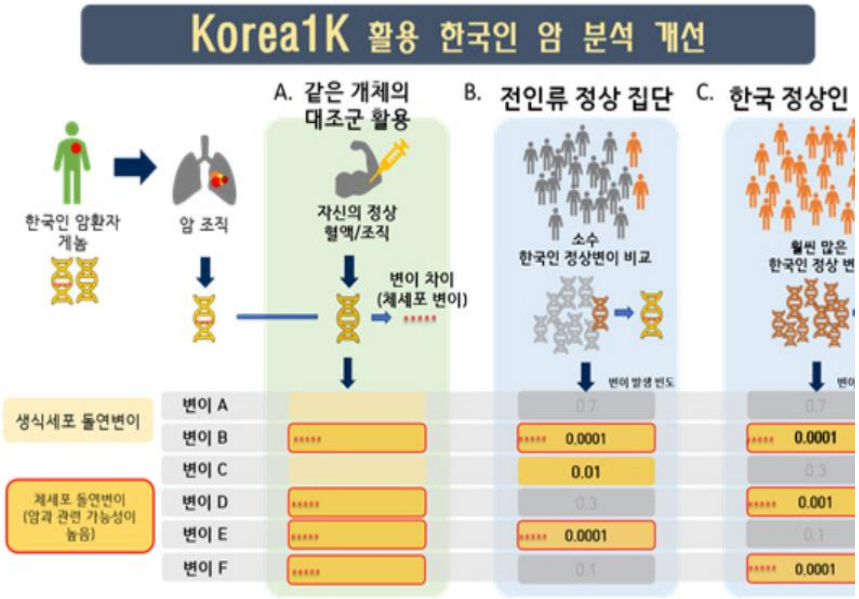
게놈 데이터의 공유

❖ 20여건의 전장게놈 데이터 공유 진행(Korea1K)

요청 기관	제공 날짜	요청 기관	제공 날짜
양산부산대병원	2020.05.13.	게놈연구재단	2020.09.07.
울산대 의대 약리학교실	2020.05.27.	성균관대 삼성융합의과학원/미 NIH	2020.10.04.
Fudan University	2020.06.09.	KAIST 의과학대학원	2020.11.25.
울산대병원	2020.06.11.	한양대학교	2020.12.03.
울산대 의대 생화학분자 생물학교실	2020.06.12.	UC Berkeley	2020.12.08.
연세대 의대 약리학교실	2020.06.12.	양산부산대병원	2020.12.08.
게놈연구재단	2020.06.15.	연세대 의대	분양절차 진행중
연세대 의대	2020.06.21.	UNIST	분양절차 진행중
MRC-LMB	2020.08.13.	Lieber Institute for Brain Development, Johns Hopkins 의대	분양절차 진행중
Department of Psychiatry, UCLA	2020.08.19.	연세대학교 의과학과	분양절차 진행중
서울대 보건대학원 보건통계연구실	2020.08.24.	서울대학교병원	분양절차 진행중

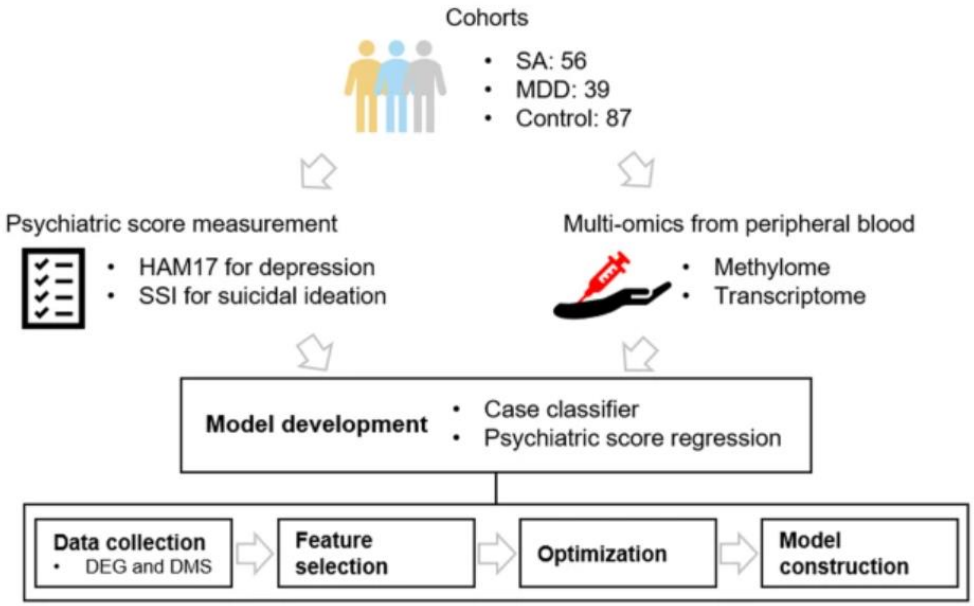
게놈 및 다중오믹스 연구 논문 16편 게재

Korean Genome Project: 1,094 Korean personal genomes with clinical information



대량(1천명)의 한국인 게놈데이터가 한국인의 **암과 각종 질병 및 표현형 분석의 정밀도**를 높일 수 있음

Depression and suicide risk prediction models using blood-derived multi-omics data

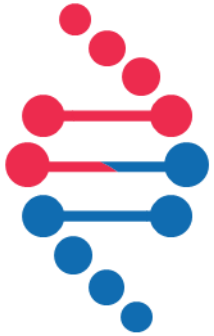


다중오믹스 분석을 통해 우울증과 자살을 예측하는 **기계학습 모델을 개발**하여 **치료에 활용**할 수 있음

Korea10K Data ➔ BioDataFarm

❖ Genome Based Biomedical

Korea10K Data ➔ Personalized Medicine



KOGIC

KOrean Genomics Center

게 놈 산 업 기 술 센 터

FIRST IN CHANGE