

Genome Korea

10,000 Korean Genomes

KOGIC

Clinomics

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 맞춤형 의료 인프라 구축



게놈기반 바이오메디컬
기업과의 산학밀착형 연구개발

게놈 바이오메디컬 기업과의 산학밀착형 연구개발

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

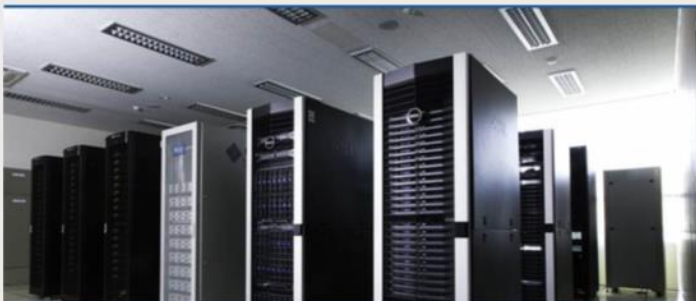


게놈 융합 사업의
핵심 인프라 역할

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, 세포조작배양기, 단일 세포 분류기, 위상차현미경, 중합효소연쇄반응기, 핵산 분석/정량장치, 역상형광 현미경이미지 시스템, 유전자증폭시스템, 분자병리진단 현미경 등)



최초의 호랑이
표준게놈지도 완성



최초의 고래
표준게놈지도 완성



최초의 한국 독수리
게놈 해독



최초의 한국 표범
표준게놈지도 완성



최초의 황금박쥐 게놈
해독



노무라 입깃 해마리
표준게놈 지도 완성

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



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



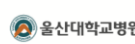
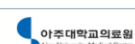
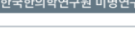
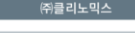
최초로 산성기 동아시아인 고대
게놈 해독 한국인의 유전적
뿌리 규명



KoVariome: 한국인
변이체



한국인 1천명 게놈
(Korea1K) 분석

 울산광역시	 울산대학교
 울산대학교병원	 연세대학교 의과대학 연세대학교 의과대학 송당암연구소
 부산대학교	 제주대학교 제주국제동물연구소
 DGIST 슈퍼컴퓨팅 빅데이터 융합연구소	 아주대학교의료원 아주대학교의료원 분당백병원 바이오메디칼 센터
 농촌진흥청 농촌진흥청 포스트게놈데이터유전체사업단	 KFRRI 한국식품연구원 한국식품연구원
 한국화학연구원 한국화학연구원 미생연구단	 KITT 안전성평가연구소 안전성평가연구소
 TheraGen ETEX Bio Institute (주)테라젠이텍스	 Clinomics (주)클리노믹스
 Helixco (주)헬릭스코	 DATA COMMAND (주)데이터커맨드

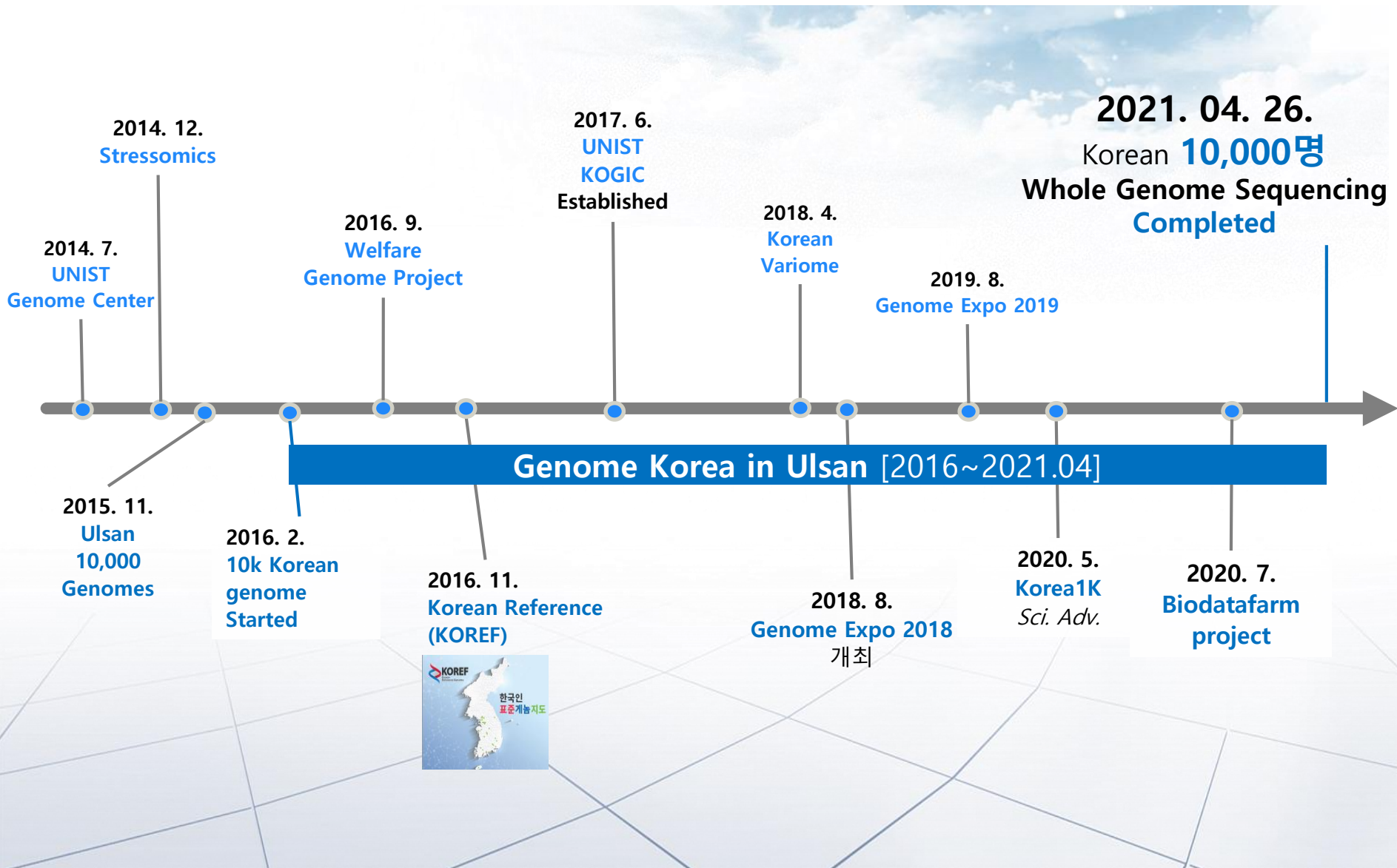
Genome Korea Project

Aims:

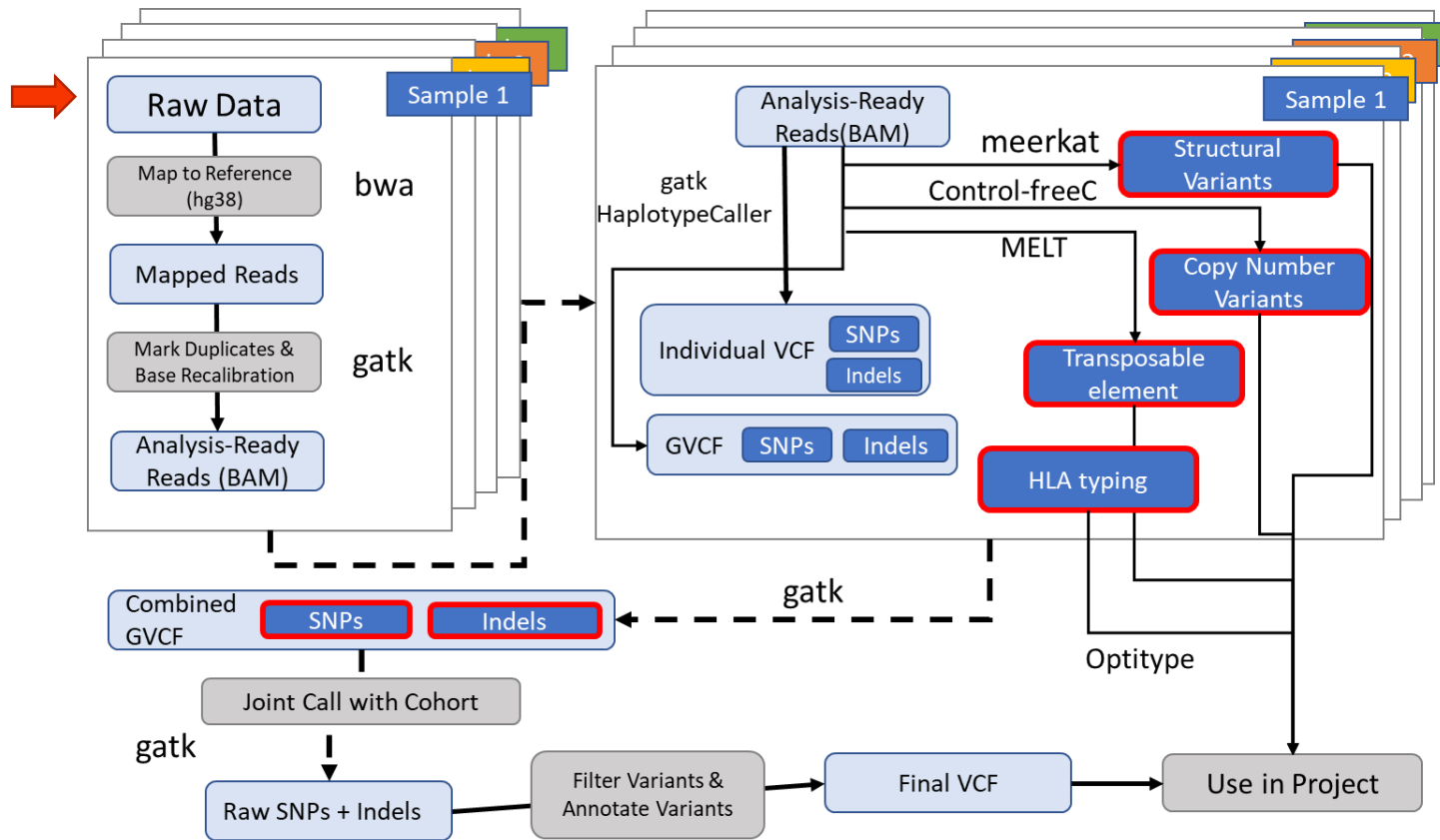
- To characterize the non-pathological ethnicity-relevant genetic variation and **diversity present in the Korean population**
- To predict candidate loci, inheritance patterns, and genetic risk for certain diseases
- To improve the **consensus Korean Reference** genome standard
- To identify Korean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



Milestones

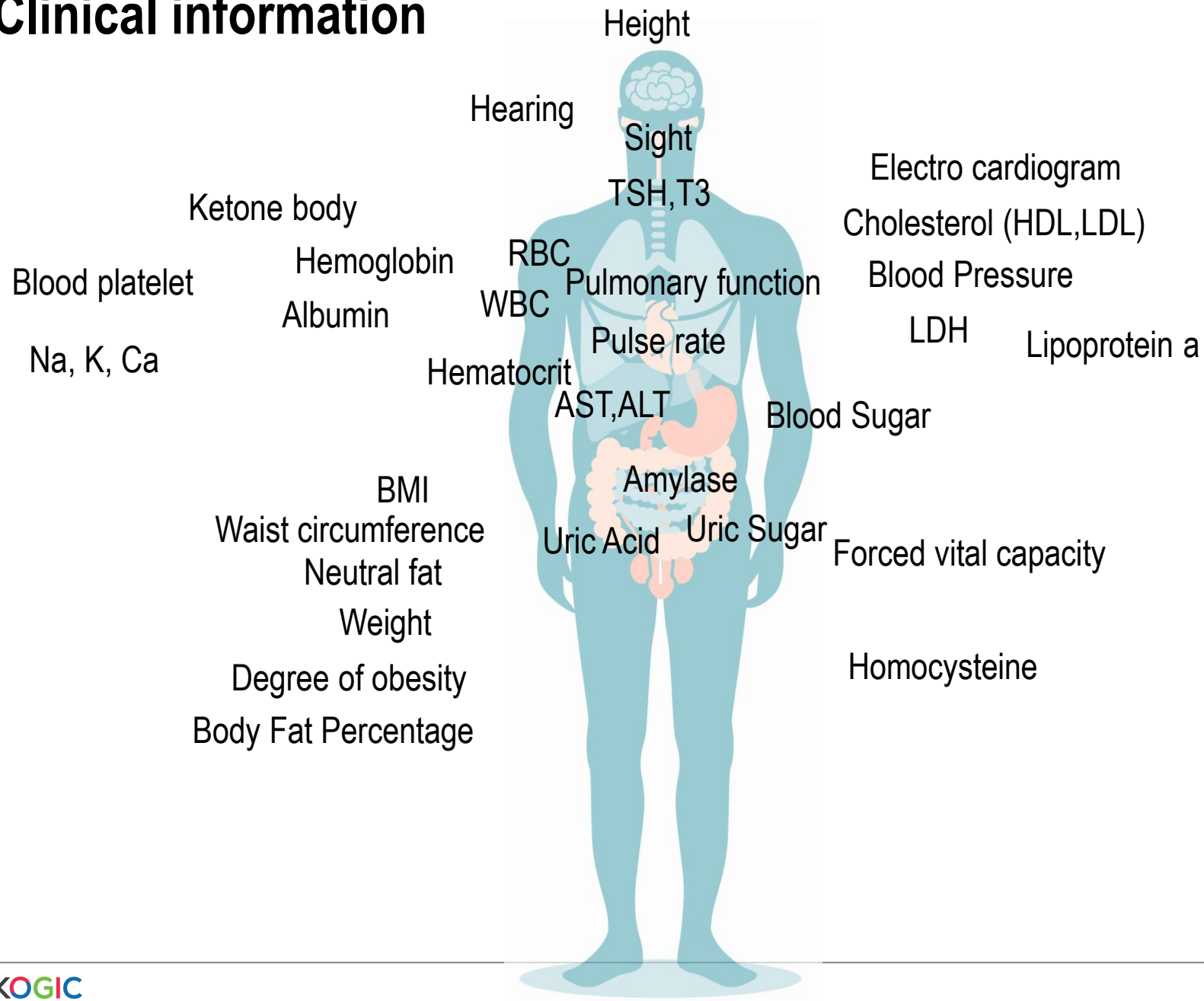


Whole Genome Processing Pipelines



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

Clinical information

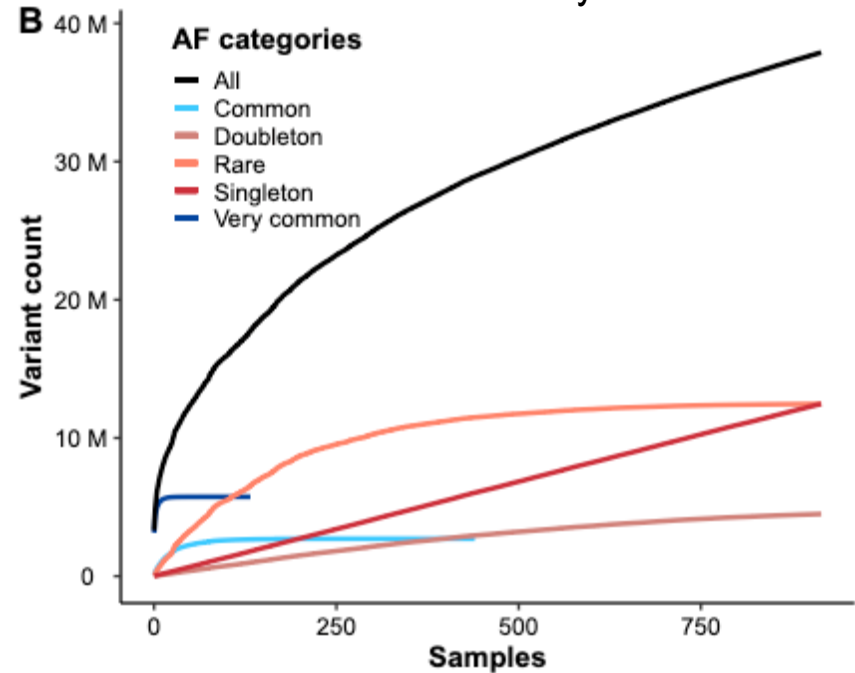
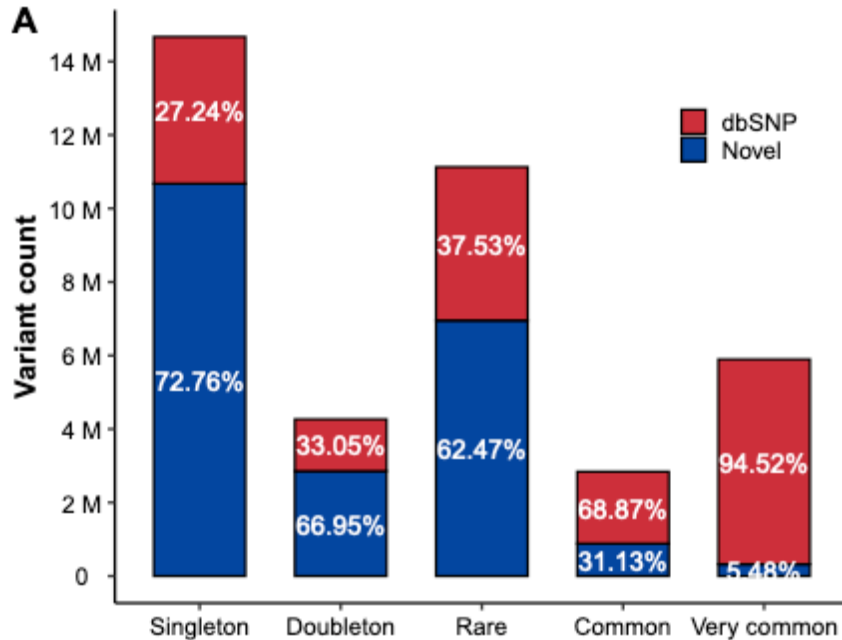


Variant Statistics

Rare: $\geq AC=3, <1\%$

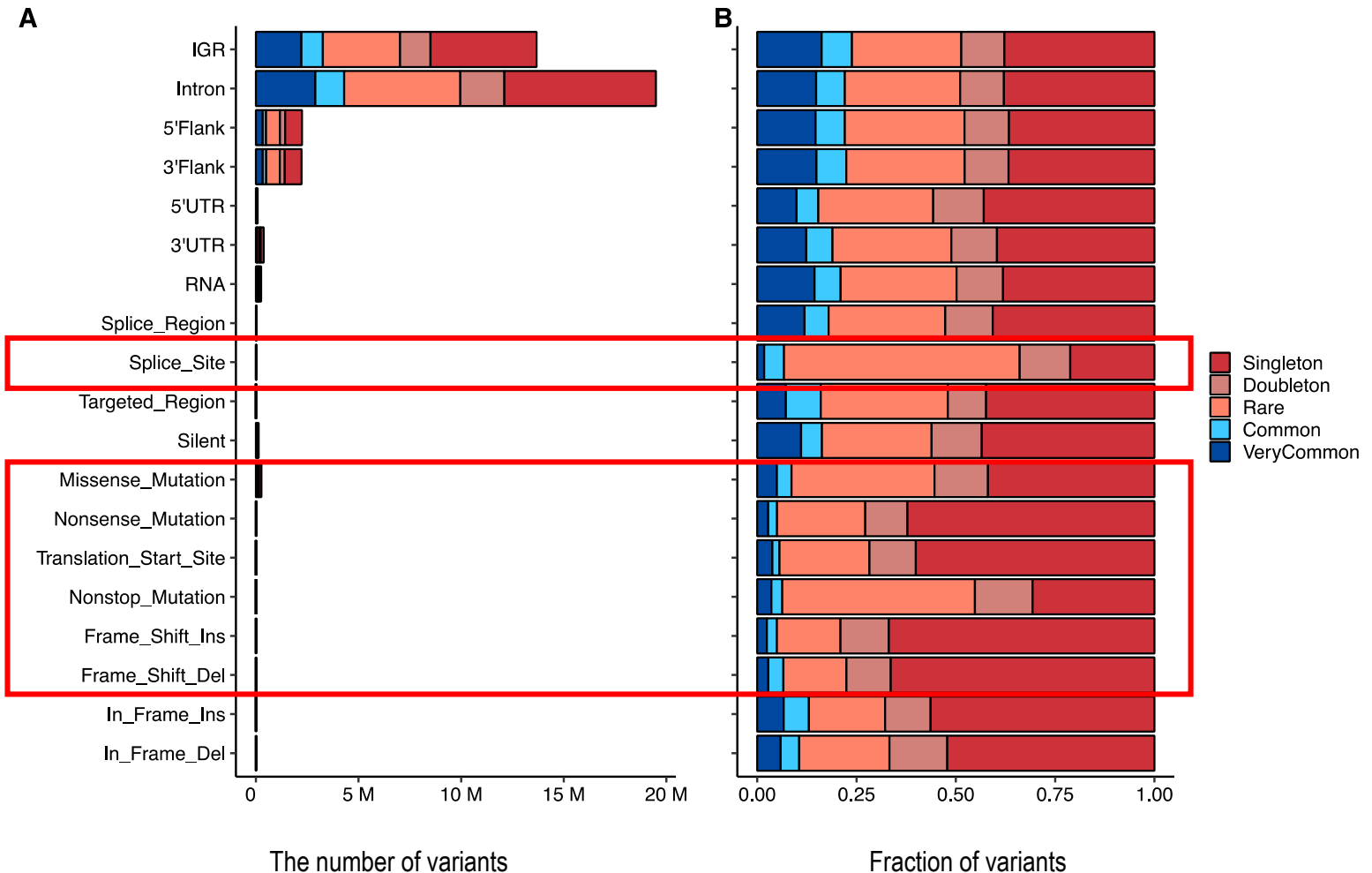
Common: $\geq 1\%, <5\%$

Very Common: $>5\%$



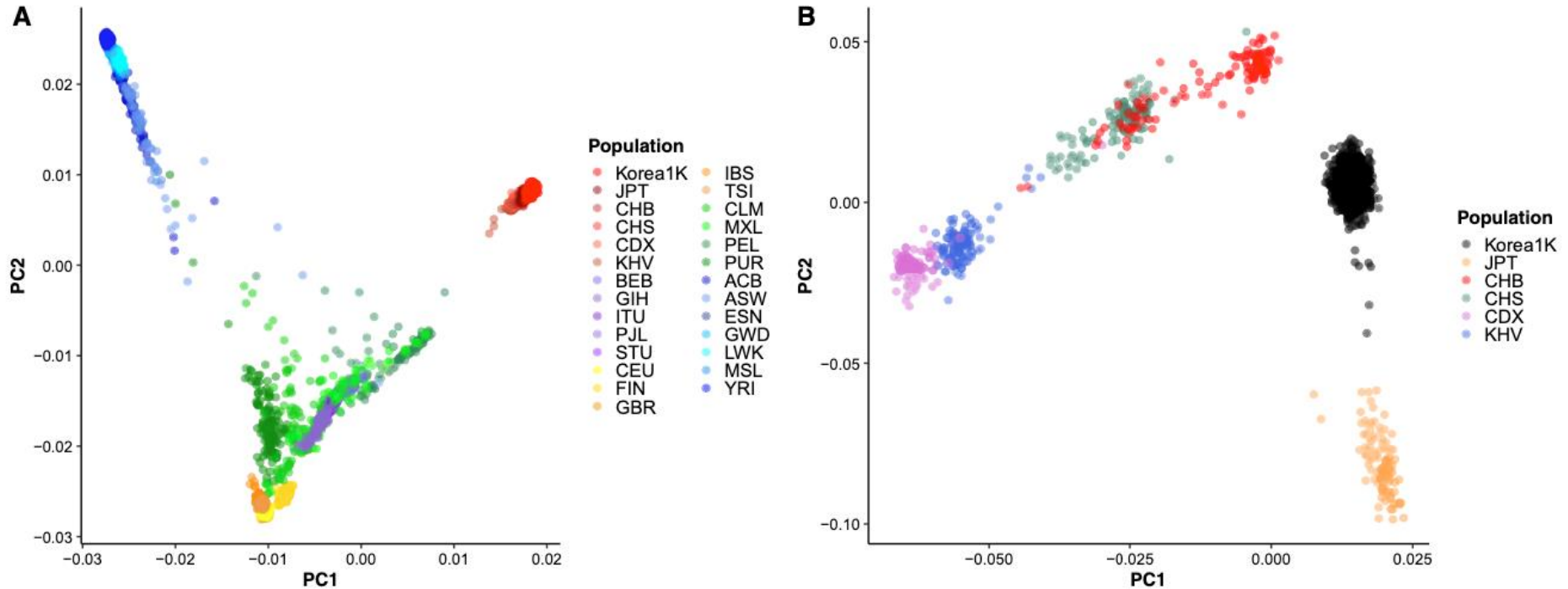
- 39,025,362 variants (13,462,495 singleton) from 1,094 Koreans and most of them are rare variants, although Korean population is known to be homogenous
- Korea1K covered most of very common variants, while more samples are needed to cover very rare variants.

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.

Drug response variants found in Korea1Ka

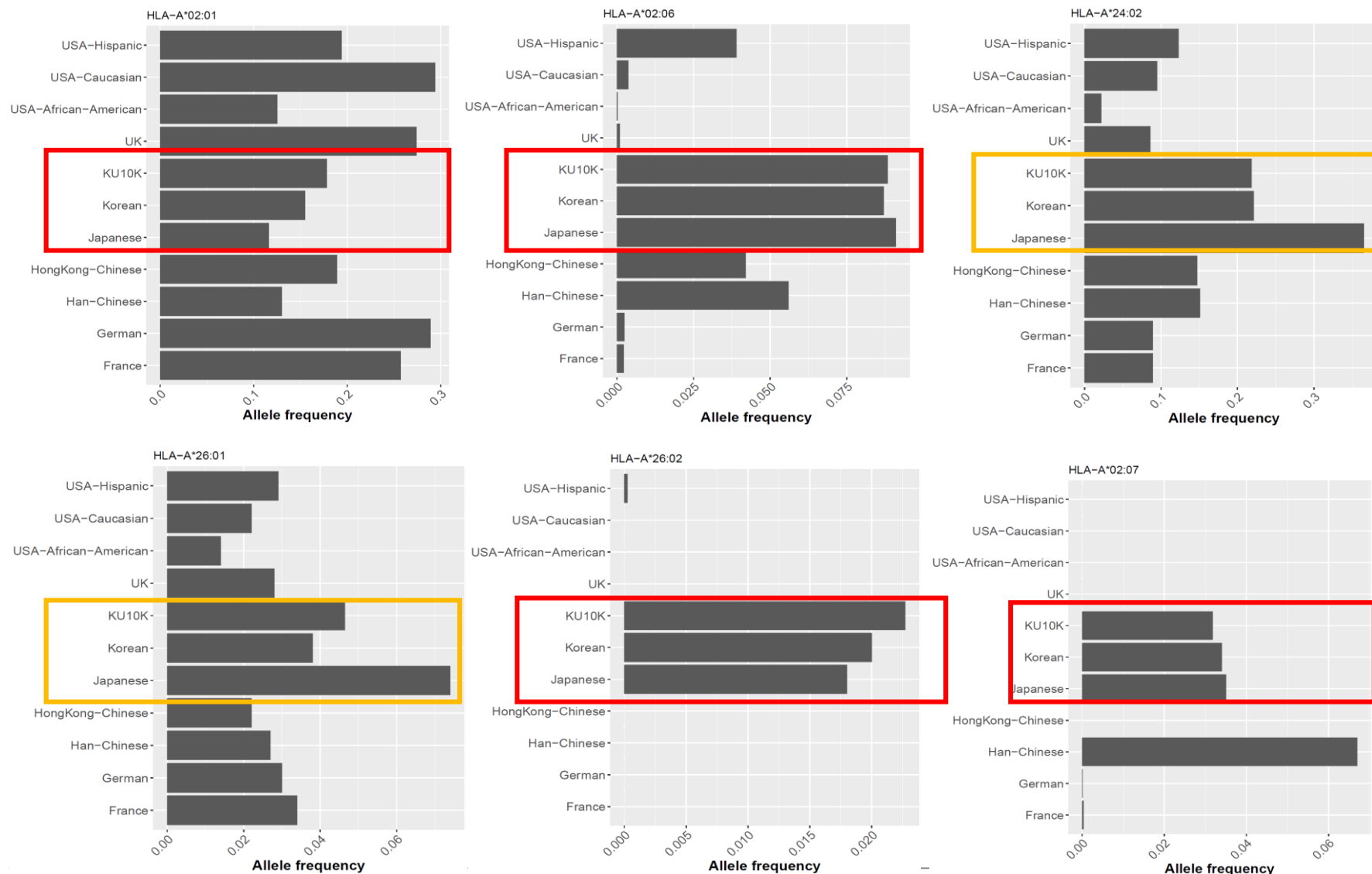
		Korea1K																																				
		0.98	0.76	0.66	0.64	0.64	0.63	0.6	0.6	0.59	0.59	0.54	0.43	0.42	0.41	0.4	0.4	0.27	0.27	0.26	0.26	0.25	0.18	0.16	0.16	0.14	0.14	0.13	0.12	0.12	0.05	0.02	0.02	0.01	0.01	0.01		
Population	CDX	0.95	0.8	0.66	0.67	0.53	0.7	0.37	0.77	0.72	0.72	0.58	0.52	0.38	0.28	0.45	0.36	0.22	0.27	0.25	0.31	0.28	0.22	0.33	0.27	0.1	0.27	0.07	0.15	0.05	0.03	0.05	0.03	0.04	0.01	0.01		
	CHS	0.96	0.79	0.69	0.67	0.6	0.68	0.46	0.6	0.68	0.68	0.61	0.52	0.45	0.32	0.3	0.42	0.26	0.28	0.25	0.31	0.3	0.19	0.16	0.24	0.12	0.19	0.08	0.11	0.16	0.05	0.04	0.03	0	0.03	0.01		
	CHB	0.97	0.75	0.7	0.67	0.55	0.62	0.54	0.62	0.68	0.67	0.55	0.48	0.39	0.33	0.35	0.5	0.31	0.32	0.25	0.31	0.33	0.18	0.16	0.16	0.1	0.17	0.09	0.12	0.13	0.04	0.03	0	0	0.01	0.01		
	KHV	0.94	0.77	0.68	0.65	0.57	0.68	0.31	0.65	0.75	0.75	0.57	0.61	0.46	0.25	0.39	0.46	0.34	0.25	0.25	0.35	0.28	0.2	0.22	0.25	0.15	0.19	0.11	0.15	0.06	0.04	0.06	0.03	0.04	0	0.02		
	JPT	0.98	0.72	0.66	0.64	0.68	0.65	0.56	0.67	0.57	0.57	0.44	0.49	0.34	0.43	0.49	0.35	0.32	0.28	0.3	0.22	0.19	0.1	0.22	0.13	0.12	0.1	0.11	0.1	0.07	0.02	0.02	0.02	0.01	0.01	0		
	EAS	0.96	0.76	0.68	0.66	0.59	0.67	0.45	0.66	0.68	0.68	0.55	0.52	0.4	0.32	0.39	0.42	0.29	0.28	0.26	0.3	0.28	0.18	0.22	0.21	0.12	0.18	0.09	0.12	0.1	0.03	0.04	0.02	0.02	0.01	0.01		
	SAS	0.64	0.66	0.56	0.54	0.51	0.68	0.2	0.6	0.36	0.34	0.45	0.46	0.53	0.64	0.42	0.36	0.1	0.44	0.52	0.3	0.42	0.29	0.38	0.22	0.23	0.07	0.17	0.51	0.07	0.11	0.35	0.02	0.06	0	0		
	EUR	0.56	0.63	0.55	0.55	0.71	0.6	0.2	0.65	0.46	0.42	0.39	0.36	0.35	0.54	0.16	0.21	0.09	0.5	0.52	0.25	0.51	0.33	0.24	0.09	0.16	0.02	0.4	0.47	0	0.07	0.45	0.03	0.12	0	0		
	AMR	0.63	0.69	0.58	0.58	0.68	0.72	0.17	0.71	0.69	0.66	0.46	0.33	0.26	0.29	0.2	0.33	0.14	0.38	0.28	0.35	0.45	0.48	0.37	0.1	0.14	0.11	0.28	0.58	0.04	0.04	0.36	0.06	0.06	0	0		
AFR	0.61	0.89	0.6	0.24	0.33	0.73	0.05	0.98	0.48	0.48	0.52	0.85	0.51	0.33	0.01	0.18	0.01	0.28	0.25	0.29	0.53	0.48	0.37	0.08	0.35	0.03	0.82	0.42	0	0	0.29	0.07	0.03	0	0			
		rs1208	rs25487	rs6166	rs6165	rs1042522	rs2228001	rs4961	rs5219	rs10246939	rs713598	rs1042713	rs20455	rs1056892	rs1726866	rs1799971	rs1800566	rs2231142	rs4680	rs1801394	rs3212986	rs1801274	rs1695	rs3745274	rs2011425	rs2234922	rs1799931	rs1056836	rs4880	rs116855232	rs1057910	rs1801280	rs1142345	rs2297595	rs186364861	rs111033610		

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.

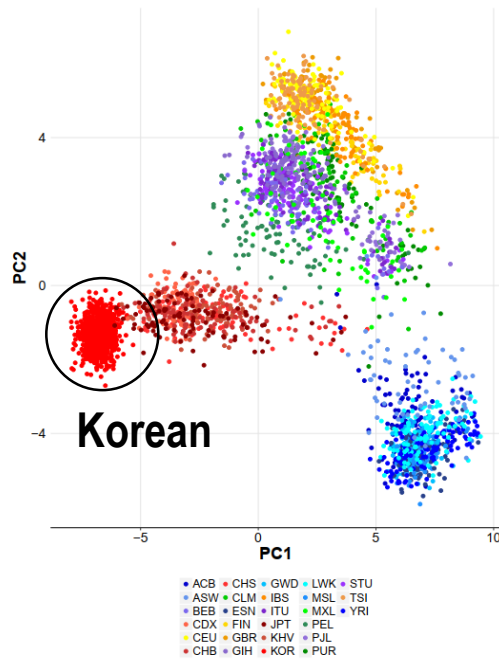
HLA typing by WGS dataa

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

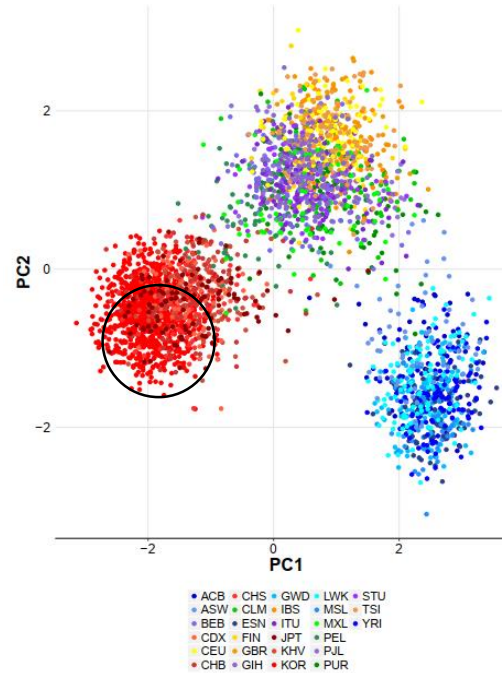


Population stratification confirmed by PCA with TEs in Korea1Ka

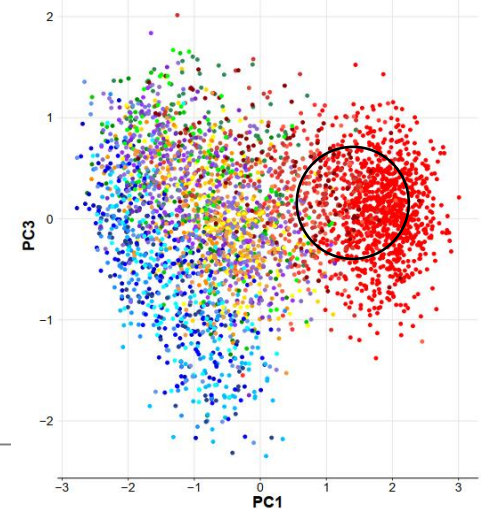
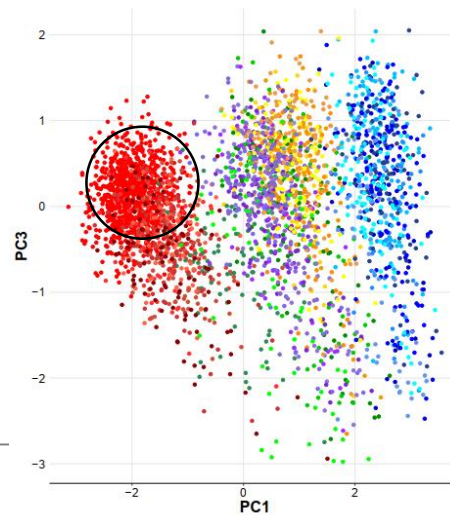
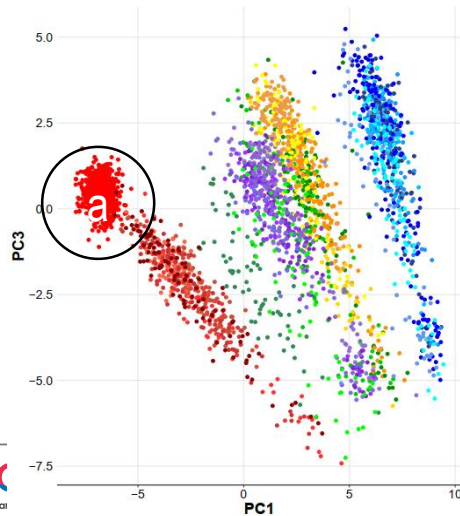
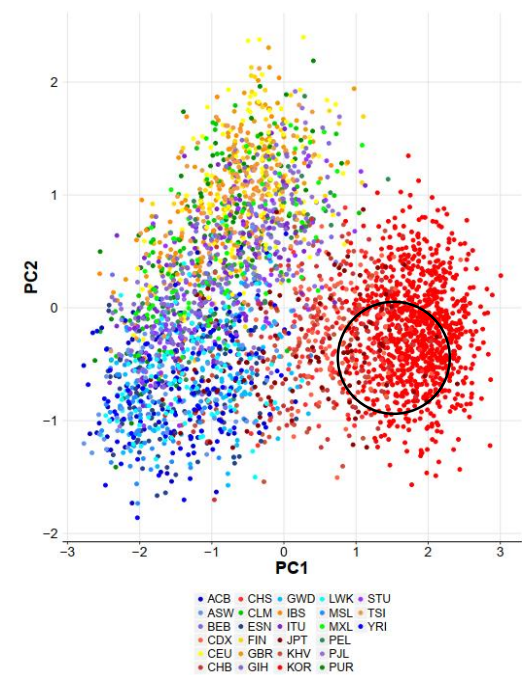
ALU



LINE1

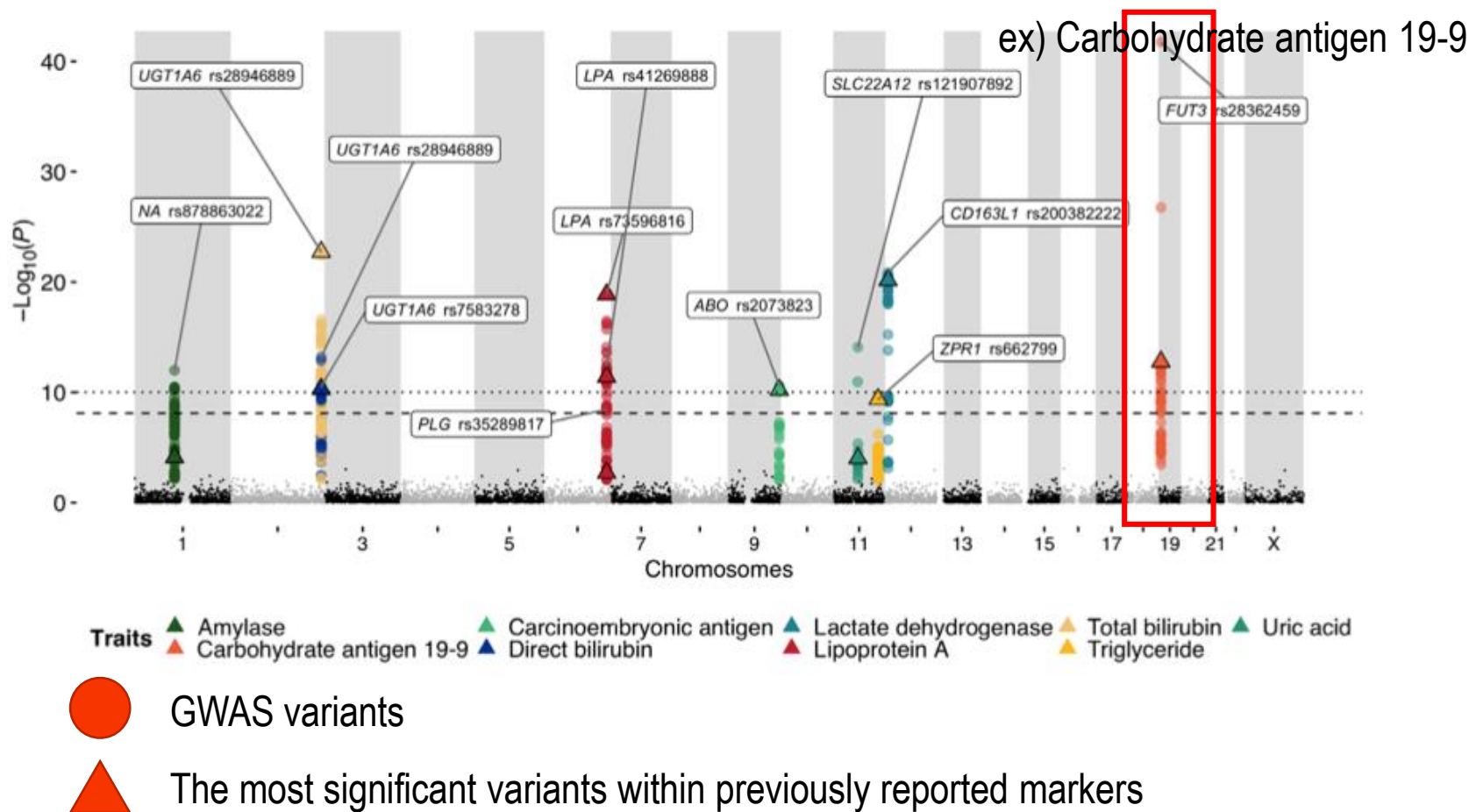


SVA

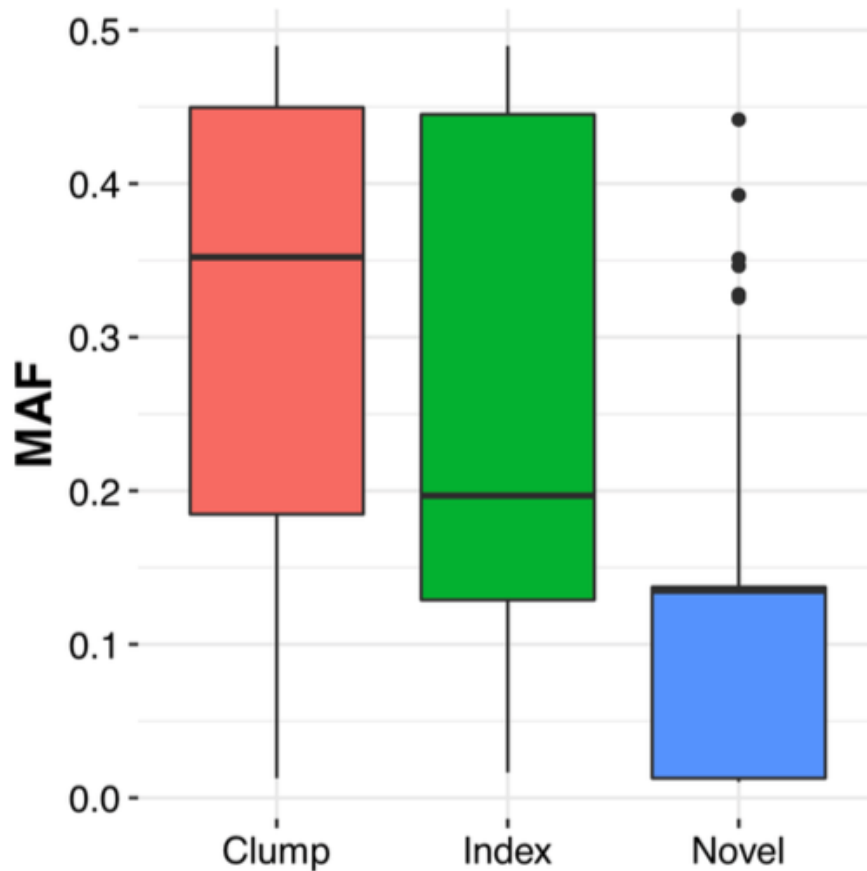


Whole-genome sequencing based GWAS identified more significant GWAS marker than chip-based

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.**

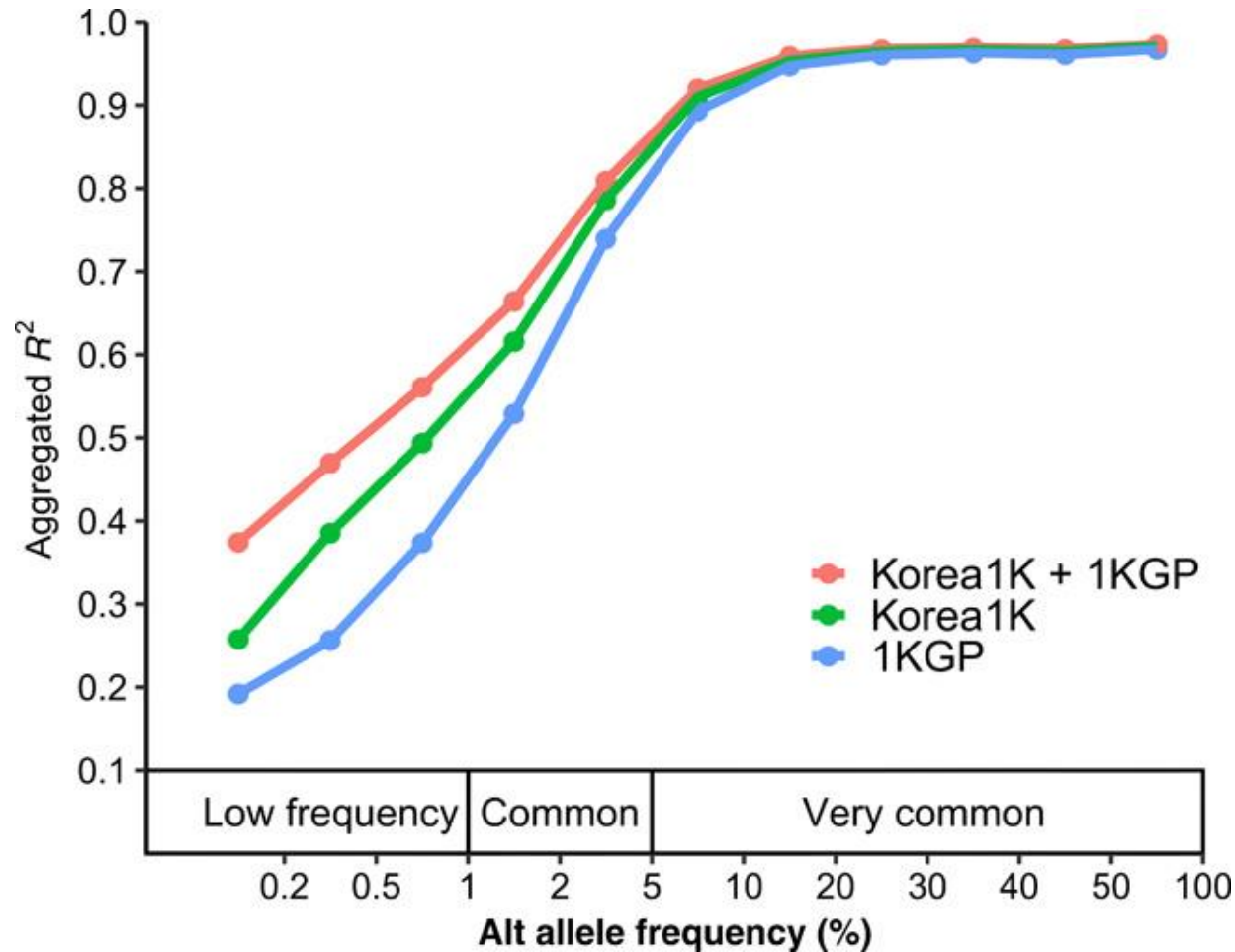


WGS data help identify low-frequency alleles and unreported GWAS loci



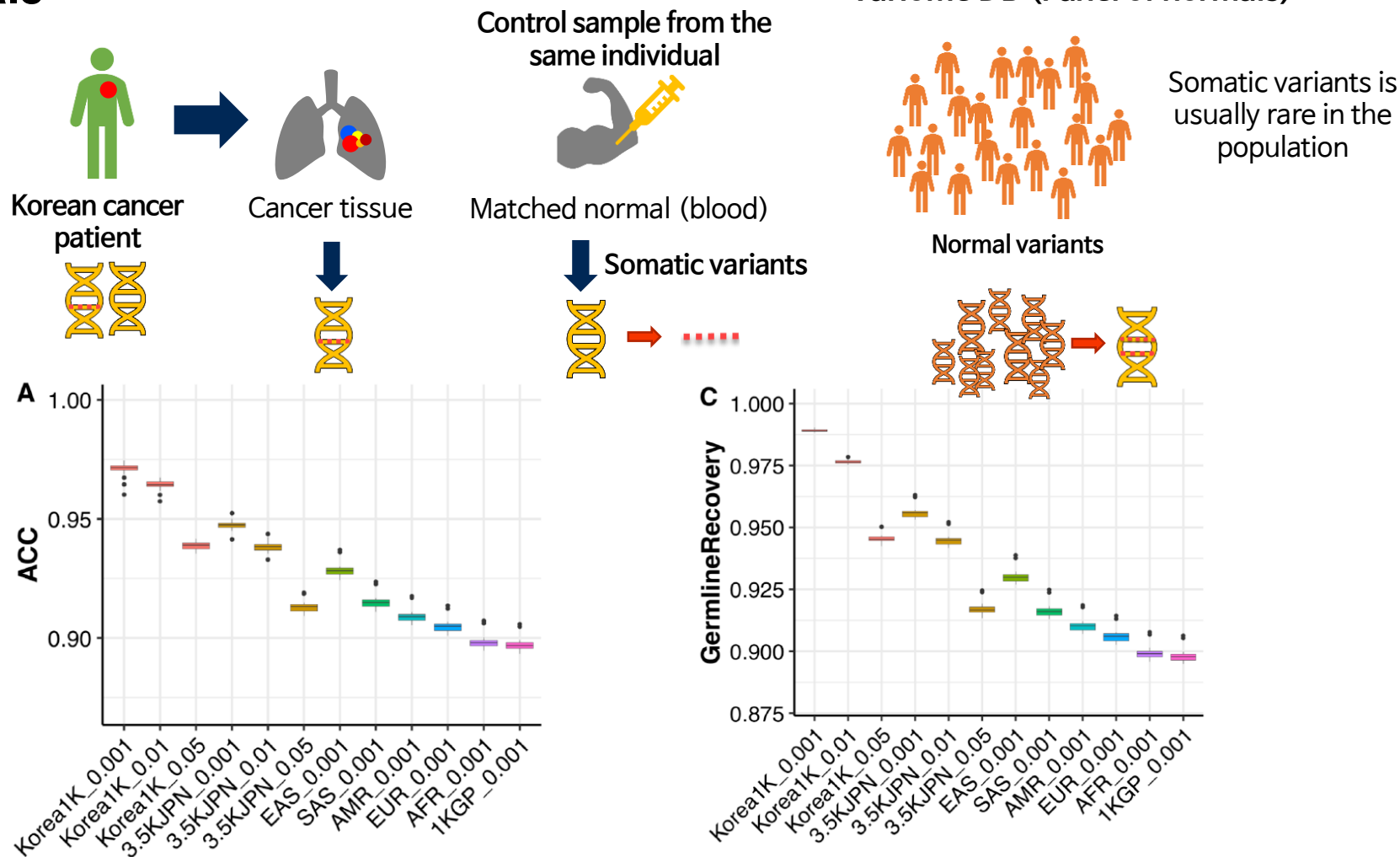
*The MAFs in the previously unidentified loci were markedly lower than those previously reported when we compared the MAFs of GWAS variants in these previously reported and the unreported loci (fig. S29). This means **that large-scale variomes from WGS data help identify low-frequency alleles and unreported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



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 Jang¹, Jinho Jang^{1,2}, Asta Blazyte¹, Changjae Kim^{1,3}, Yeonkyung Kim¹, Jungae Shim¹, Nayeong Kim¹, Yeo Jin Kim¹, Seung Gu Park¹, Jungeun Kim⁴, Yun Sung Cho³, Yeshin Park³, Hak-Min Kim^{1,2,3}, Byoung-Chul Kim³, Neung-Hwa Park^{5,6}, Eun-Seok Shin⁷, Byung Chul Kim³, Dan Bolser³, Andrea Manica⁸, Jeremy S. Edwards⁹, 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>

Copyright © 2020
The Authors, some
rights reserved;
exclusive licensee
American Association
for the Advancement
of Science. No claim to
original U.S. Government
Works. Distributed
under a Creative
Commons Attribution
NonCommercial
License 4.0 (CC BY-NC).

<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



ARTICLE

Received 24 Mar 2016 | Accepted 18 Oct 2016 | Published 24 Nov 2016

DOI: 10.1038/ncomms13637

OPEN

An ethnically relevant consensus Korean reference genome is a step towards personal reference genomes

Yun Sung Cho^{1,2,3,*}, Hyunho Kim^{4,*}, Hak-Min Kim^{1,2}, Sungwoong Jho³, JeHoon Jun^{3,4}, Yong Joo Lee⁴, Kyun Shik Chae⁵, Chang Geun Kim⁵, Sangsoo Kim⁶, Anders Eriksson⁷, Jeremy S. Edwards⁸, Semin Lee^{1,2}, Byung Chul Kim^{1,2}, Andrea Manica⁷, Tae-Kwang Oh⁹, George M. Church^{10,*} & Jong Bhak^{1,2,3,4,*}

개인 게놈시대를 향한 KOREF

한국인
표준게놈지도



www.koreanreference.org

❖ 한국인 참조표준 변이체(KoVariome) 고도화

www.nature.com/scientificreports

SCIENTIFIC REPORTS

OPEN

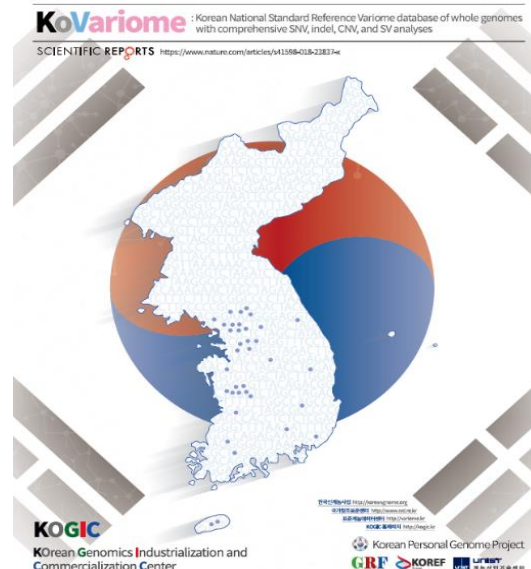
KoVariome: Korean National Standard Reference Variome database of whole genomes with comprehensive SNV, indel, CNV, and SV analyses

Received: 22 September 2017
Accepted: 16 March 2018
Published online: 04 April 2018

Jungeun Kim¹, Jessica A. Weber², Sungwoong Jho³, Jinho Jang^{4,5}, JeHoon Jun^{1,6}, Yun Sung Cho⁷, Hak-Min Kim^{1,6}, Hyunho Kim¹, Yumi Kim⁸, OkSung Chung^{1,6}, Chang Geun Kim⁵, Hyejin Lee¹, Byung Chul Kim¹, Kyudong Han⁹, InSung Koh¹, Kyun Shik Chae¹, Semin Lee^{1,6}, Jeremy S. Edwards¹⁰ & Jong Bhak^{1,6,*}

KoVariome : Korean National Standard Reference Variome database of whole genomes with comprehensive SNV, indel, CNV, and SV analyses

SCIENTIFIC REPORTS | https://www.nature.com/articles/s41598-018-22833-4



KOGIC
Korean Genomics Industrialization and Commercialization Center

Korean Personal Genome Project
KOREF
KOGIC

게놈분석 연구리포트 개발 및 리포팅 자동화 시스템 구축

수백명 단위의 대규모 연구 참여자들에게 게놈 연구 분석 결과 정보인 게놈 분석 리포트를
혜택으로 제공하기 위해 대량의 결과 데이터를 효율적으로 리포팅 할 수 있는 자동화 시
스템 구축

울산 만명 게놈 프로젝트 결과 리포트



- ◀ 유전자 나이 예측
염색체 길이로 세포노화 측정하여
생체나이 예측
- ◀ 나의 인종 분석
게놈 기반으로 인종/민족 구성요소
및 비율 분석
- ◀ 조상의 이동경로 분석
모계 및 부계의 인종/민족 분석
- ◀ 질병 및 신체적 특성 예측
질병 및 신체적 특성과 연관된
유전자 변이 상태 분석

Genome Report

관리자 메뉴

- 샘플
- 질병
- 비질병
- 프로젝트
- 보고서

샘플 목록

Show 10 entries

전체 참여자아이디 검색

프로젝트	참여자아이디	언어	성별	성인	리포트	PDF	관리
U10K	U10K-00921	국문	여성	N	국문 여성 아동 보고서		Make
U10K	U10K-00920	국문	여성	Y	국문 여성 성인 보고서		Make
U10K	U10K-00919	국문	남성	Y	국문 남성 성인 보고서		Make
U10K	U10K-00918	국문	여성	Y	국문 여성 성인 보고서		Make
U10K	U10K-00917	국문	여성	N	국문 여성 아동 보고서		Make
U10K	U10K-00916	국문	여성	N	국문 여성 아동 보고서		Make
U10K	U10K-00915	국문	남성	N	국문 남성 아동 보고서		Make
U10K	U10K-00914	국문	남성	N	국문 남성 아동 보고서		Make
U10K	U10K-00913	국문	여성	N	국문 여성 아동 보고서		Make
U10K	U10K-00912	국문	남성	N	국문 남성 아동 보고서		Make

Showing 1 to 10 of 316 entries

Previous 1 2 3 4 5 ... 32 Next

다운로드

게놈 데이터의 공유

❖ 20여건의 전장게놈 데이터 공유 진행(Korea1K)

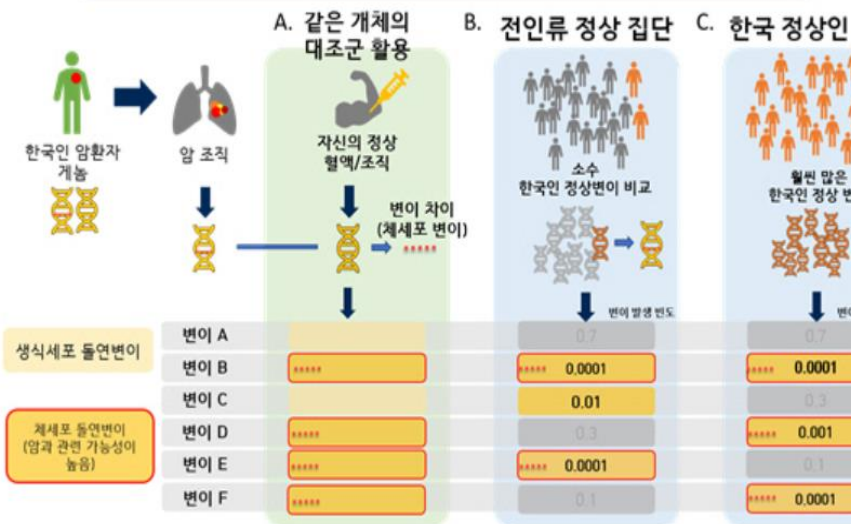
요청 기관	제공 날짜	요청 기관	제공 날짜
양산부산대병원	2020.05.13.	게놈연구재단	2020.09.07.
울산대 의대 약리학교실	2020.05.27.	성균관대 삼성융합의과학원/미 NIH	2020.10.04.
Fudan University	2020.06.09.	KAIST 의과학대학원	2020.11.25.
울산대병원	2020.06.11.	한양대학교	2020.12.03.
울산대 의대 생화학분자 생물학교실	2020.06.12.	UC Berkeley	2020.12.08.
연세대 의대 약리학교실	2020.06.12.	양산부산대병원	2020.12.08.
게놈연구재단	2020.06.15.	연세대 의대	분양절차 진행중
연세대 의대	2020.06.21.	UNST	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

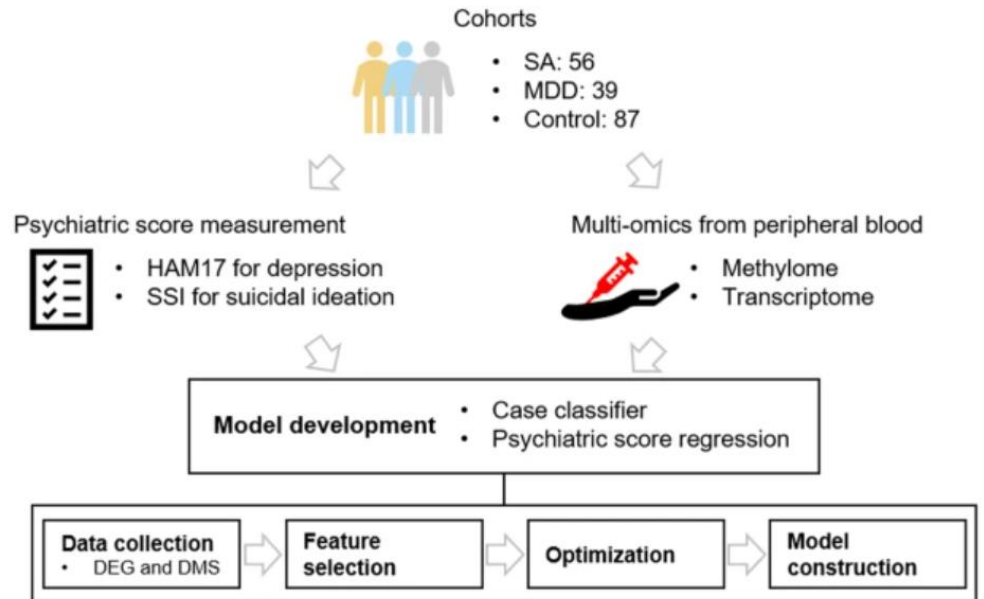
Korean Genome Project: 1,094 Korean personal genomes with clinical information

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

Korea1K 활용 한국인 암 분석 개선



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



다중오믹스 분석을 통해 우울증과 자살을 예측하는 기계학습 모델을 개발하여 치료에 활용할 수 있음

Korea10K Data → BioDataFarm

❖ Genome Based Biomedical

BBS 바이오 데이터팜

인체유래물은행
바이오데이터 수집, 분양, 관리

**Bio Engine
분석 플랫폼**

백신/치료제 후보물질

심혈관 진단마커

복합만성 진단마커

수집-검색 엔진

분석 엔진

시각화 엔진

국내 최고 수준 설비

11,868 core 312 TB 103 PB

데이터 제공 특례

데이터 수집 특례

21년 상반기 1만명 게놈 해독
22년 1만명 게놈분석 논문 공개
→ 한국인 표준게놈, 변이체 완성
특구사업 연계하여 특허출원 50건
기술이전 21건

공개 바이오 빅데이터 (20여종)

NCBI, GeneCards, Ensembl, Pfam, PDB, UniProt

감염병 완치자 바이오빅데이터 (300명)

Ulsan BioBank

급성심근경색

우울증

복합만성질환

위험 예측 진단 마커 발굴

기계학습 예측 모델 구축

질병예측

맞춤형 진단

맞춤형 치료

감염병 분석 플랫폼 구축

백신, 신약, 항체치료제 후보물질 발굴

감염병 진단기술 개발
(게놈기반 고감도 바이러스 진단)

팬데믹 대응 시스템 구축

독감 및 각종 감염 백신 펩타이드 후보물질 발굴

감염질환 치료 화학 신약 후보물질 발굴

바이러스 항체 치료제 후보물질 발굴

감염질환 신속/편리/정밀형 진단마커/시스템 제공

모든 新 감염병 즉각 연구분석 대응 시스템 제공

Korea10K Data → Personalized Medicine



FIRST IN CHANGE