2022년도 제16회 통계유전학워크숍

국립보건연구원 National Institute of Health

- 강좌명: 한국인칩을 이용한 대규모 유전체 정보 분석
- 일시: 07.20.(수) ~ 07.21. (목)

일자	시간	강의 내용	담당 강사명	구분
Day 1 07.20(수)	Session 1 (90분)	한국인칩사업 소개	김봉조	강의
	Session 2 (80분)	한국인칩 컨텐츠 특징 및 성능 소개	김영진	강의
	Session 3 (60분)	한국인칩 유전체정보 정도관리 (1)	황미영	강의
	Session 4 (90분)	한국인칩 유전체정보 분석 (1)	김영진	강의
	Session 5 (70분)	한국인칩 유전체정보 분석결과 해석 (1)	김영진	강의
Day 2 07.21(목)	Session 6 (09:20-10:20)	유전체정보 분석 기초 (실습 환경 준비, 리눅스 기초 등)	황미영	실습
	Session 7 (10:30-12:30)	한국인칩 유전체정보 정도관리 (2)	황미영	실습
	Session 8 (14:00-15:30)	한국인칩 유전체정보 분석 (2)	장혜미	실습
	Session 9 (15:40-17:10)	한국인칩 유전체정보 분석결과 해석 (2)	정인혜	실습

국립보건연구원
National Institute of Health

한국인칩사업 소개
(The Korea Biobank Array project)

김 봉 조 (국립보건연구원 미래의료연구부 유전체연구기술개발과)

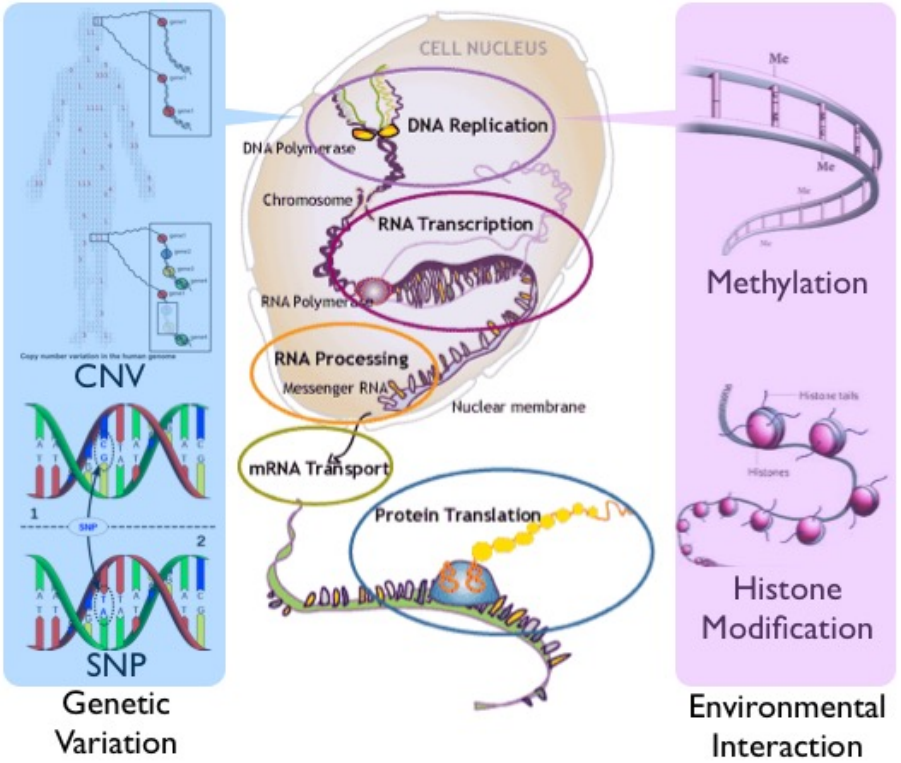
목차

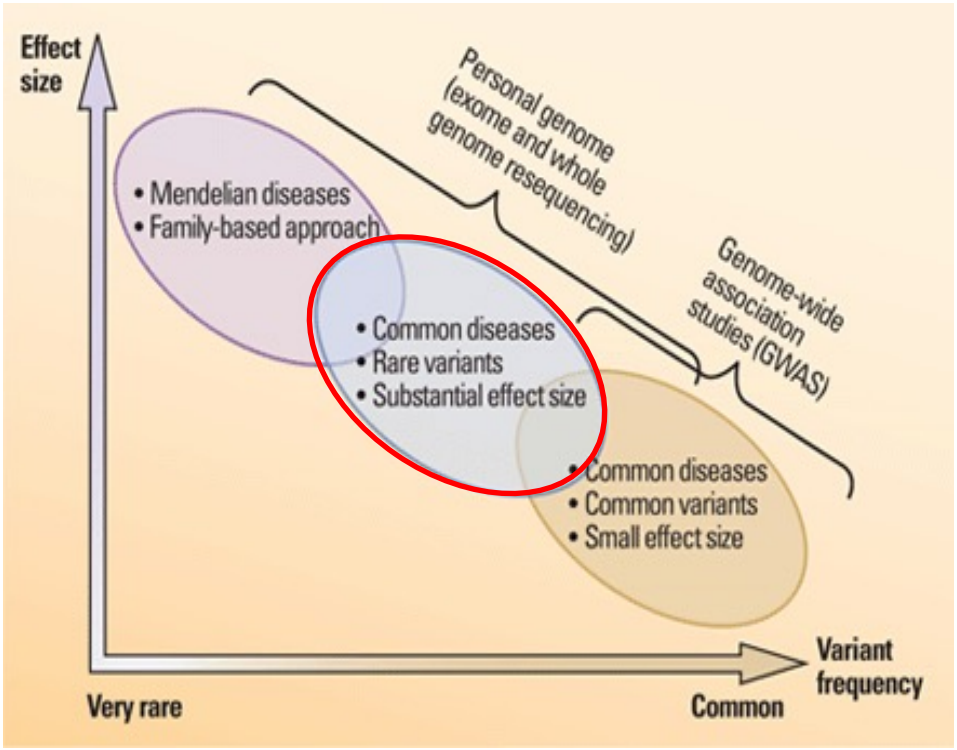
- 배경
  - 질병의 원인 : 유전 및 환경요인
  - 대규모 유전체 연구 동향 및 필요성
- 한국인칩사업 소개
  - 한국인칩 제작과정
  - 추진경과
  - 한국인칩 컨소시엄 현황 및 진행사항
- 분양된 유전체정보 현황 및 신청방법

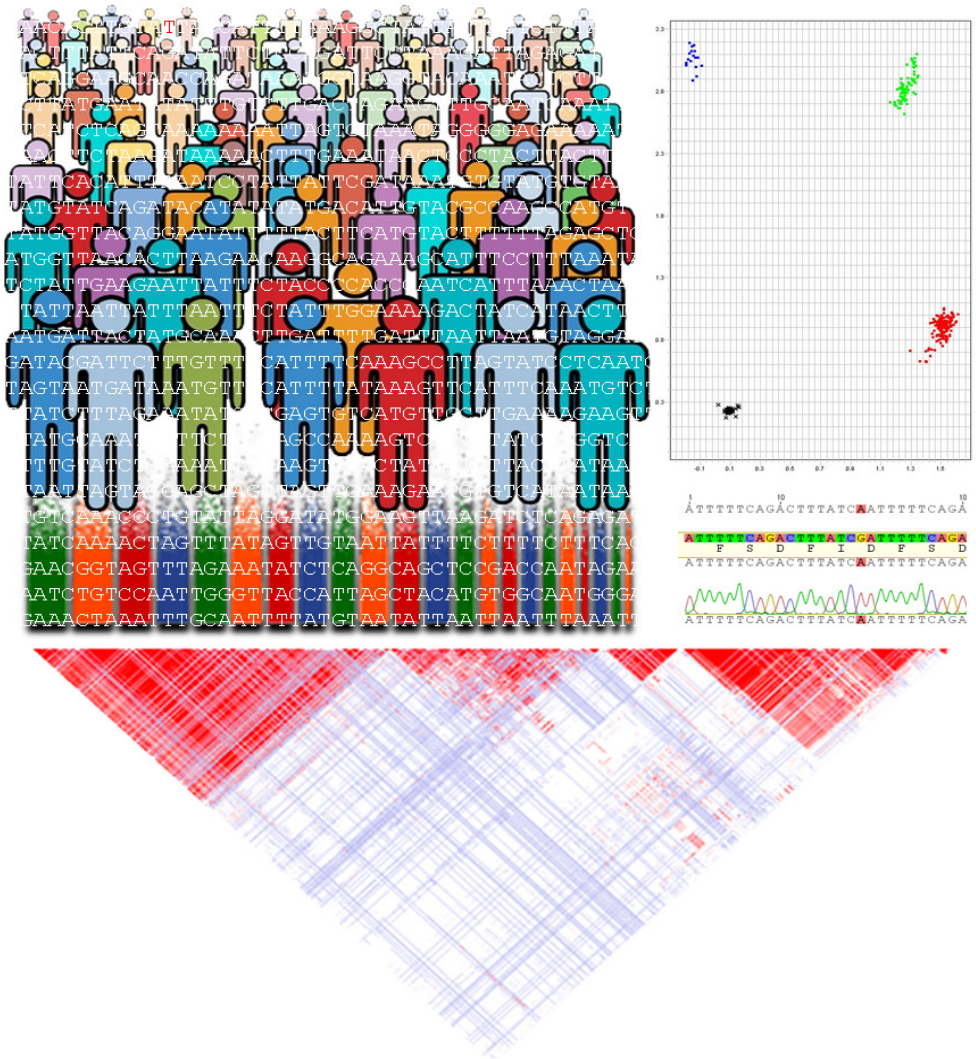
한국인칩 사업 배경

질병의 원인: 유전 및 환경 요인

- 질병은 유전 및 환경 요인의 복합적인 상호 작용에 의해서 발생함
- 이러한 요인 발굴을 통해 개인에 최적화된 치료를 할 수 있음(정밀의료)



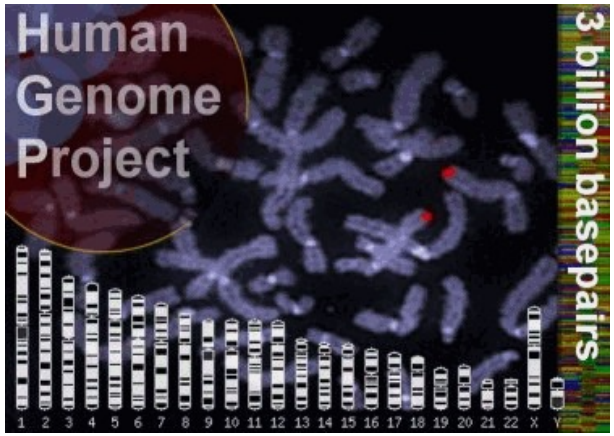


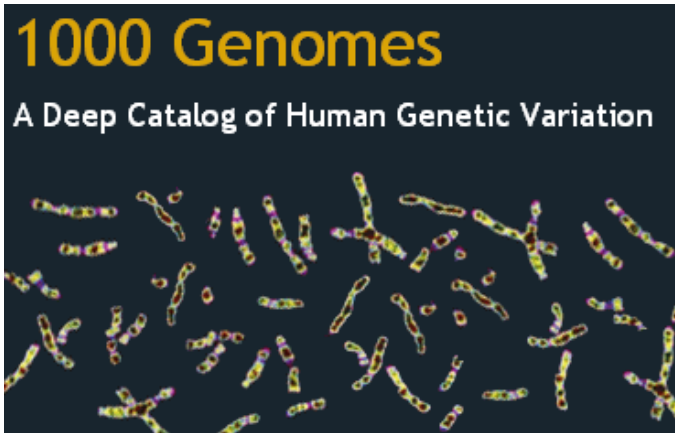


- 유전체정보 생산 및 분석
- 질병 연관 유전요인 발굴
- 유전정보 기반 질병 예측 및 예방

그림. 유전체분석연구

Exploring the human genome



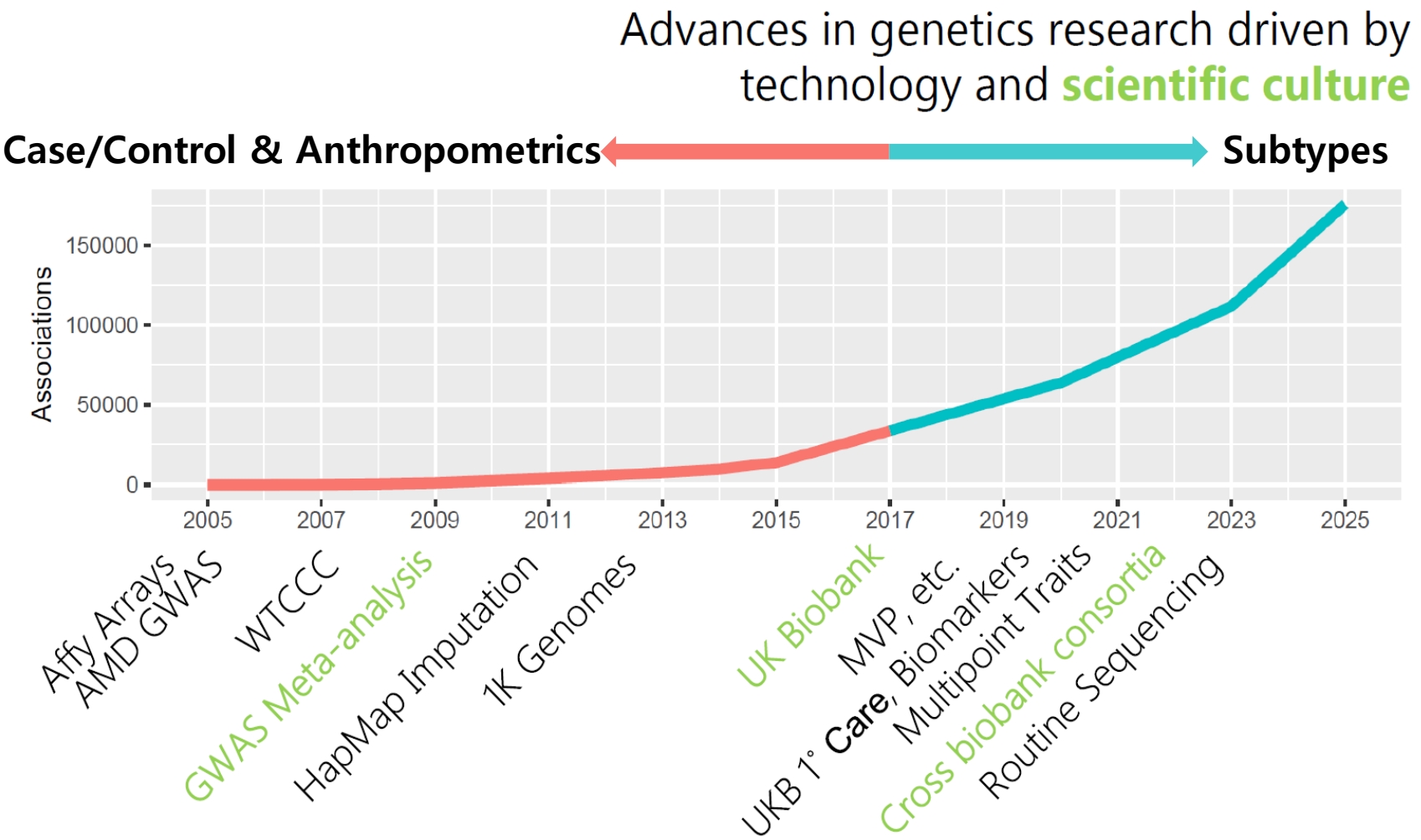


2002 2008

Sanger sequencing, targeted genotyping → Genome-wide genotyping (GWAS) → Exome sequencing → Genome sequencing

Jacques Fellay

Advances in genetics research



Nelson, ASHG 2017

Catalog of Genome-Wide Association Studies

국립보건연구원 National Institute of Health

16 years after 1st GWAS on AMD in 2005

2021. 05,
약 **5,037**편 논문
약 **257,351** SNPs 보고
(약48,000 SNPs, 2017.08)

GWAS catalog: www.ebi.ac.uk/gwas

Mega-cohorts: Current status

국립보건연구원 National Institute of Health

- Alternative genomic analysis tools are warranted to bridge the gap between SNP arrays and NGS in large-scale genomic study of complex traits.
  - Custom whole-genome genotyping array + Imputation (Sequencing-based panel)

UK Biobank Axiom Array (500K)

Million Veteran Project (>1M)

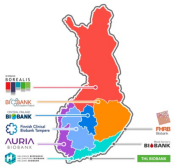
All of Us project (Precision Medicine) (~300K)

Kaiser Permanente (100K)

Taiwan Biobank Array (100K)

Japonica Array

Kadoorie Biobank Array (up to 500K)



FinnGen1 Axiom array Release#3 N=135,638

한국인유전체역학조사(KoGES) 코호트

국립보건연구원 National Institute of Health

코호트	구분	조사 시작 연도	참여자 수 (명)	세부내용
일반인 기반 코호트	지역사회기반	2001	10,030	• 안산, 안성에 거주하는 40~60세 남녀 대상 모집 • 2018년 8차 추적조사 완료
	도시기반	2004	173,209	• 도시 지역 의료기관 검진 센터 내원한 40세 이상 남녀 대상 모집 • 2007 ~ 2016년 추적조사 수행
	농촌기반	2005	28,337	• 농촌 지역에 거주하는 40세 이상 남녀 대상 모집 • 2007 ~ 2016년 6개 지역 중심 추적조사 수행
유전-환경 모델 코호트	쌍둥이 및 가족	2005	3,202	• 쌍둥이와 그 가족을 대상으로 모집 • 2008 ~ 2014년 추적조사 수행
	국내이주자	2006	7,191	• 한국인과의 결혼으로 국내에 이주한 아시아 국가 출신의 여성 및 배우자, 자녀를 대상으로 모집 • 2012 ~ 2014년 추적조사 수행
	국제협력	2005	3,556	• 일본(고베, 오사카) 및 중국(장춘, 연변)에 거주하는 한국인으로 이주 후 15년 이상인 사람과 자손 및 현지인을 대상으로 모집

국립중앙인체자원은행 저장시설

국립보건연구원
National Institute of Health

100만명분 이상 인체자원 수용가능, 현재 약 70% 운용 중

초저온저장고(-80℃)

최대 186대 설치

DNA 300만명 보관가능

* 3 vial/명 보관시

극저온저장고(-150℃)

최대 474대 설치

혈액자원 79만명 보관가능

* 20vial/명 보관시

The power is all of numbers

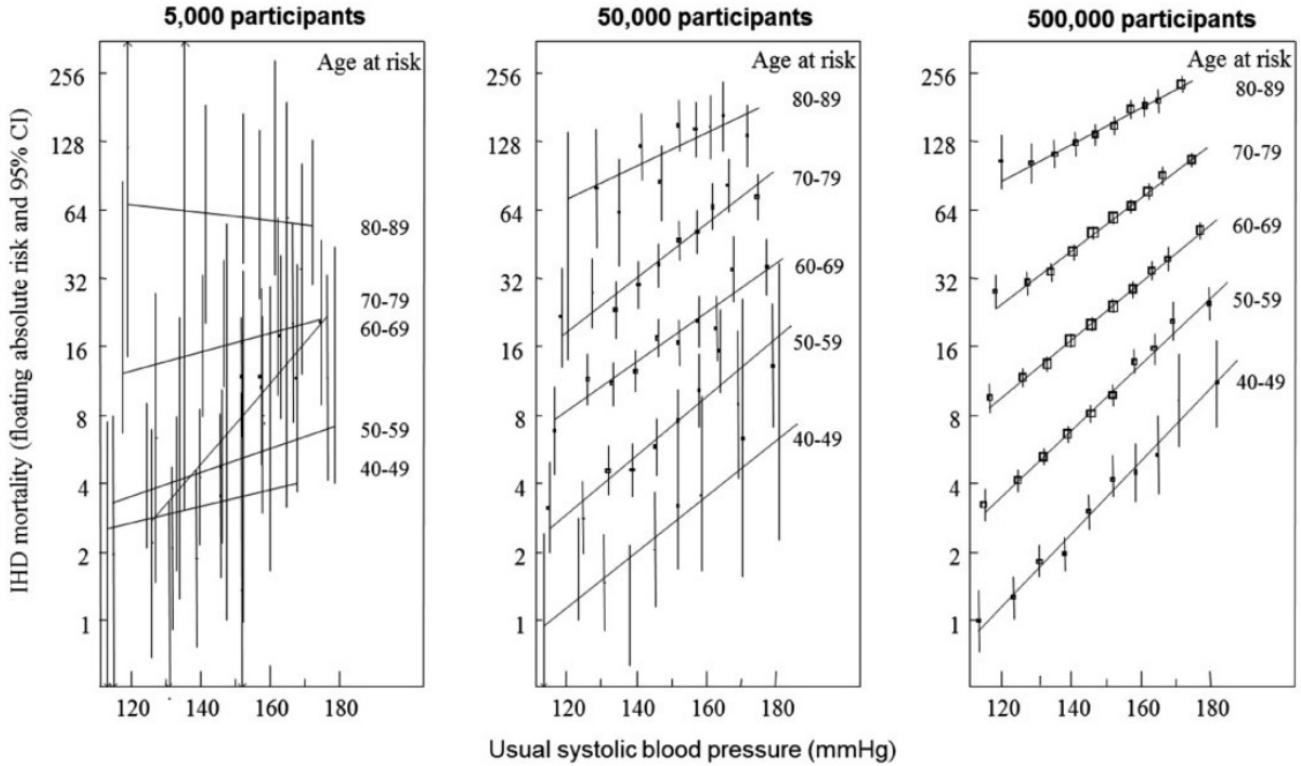
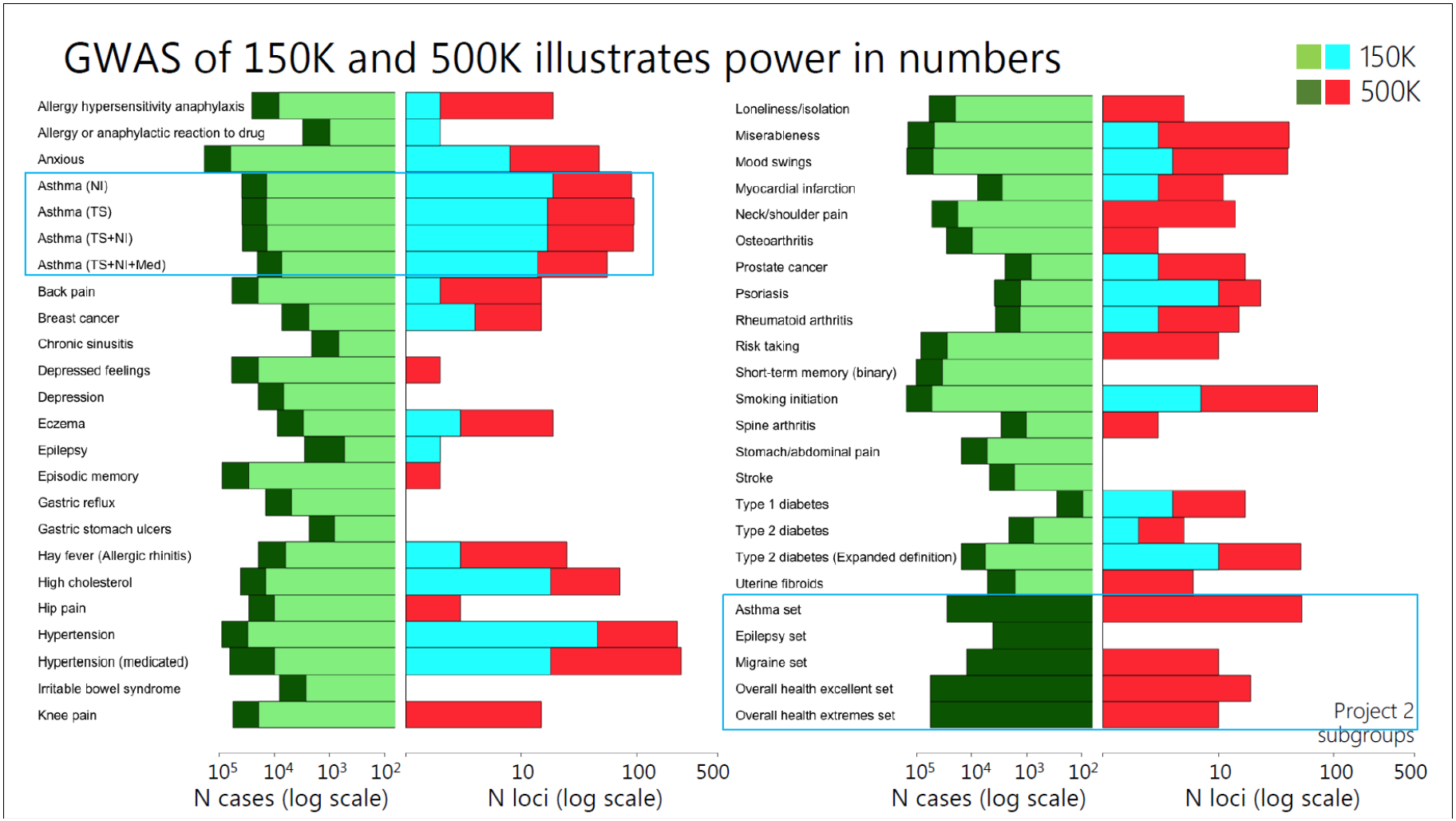


Figure 1 Absolute risk of ischaemic heart disease mortality by usual systolic blood pressure and age at risk in 5000, 50 000, and 500 000 participants. Unpublished figure containing data from the Prospective Studies Collaboration, obtained through personal communication. CI, confidence interval; IHD, ischaemic heart disease.

Eur Heart J. 2017. doi: 10.1093/eurheartj/ehx254

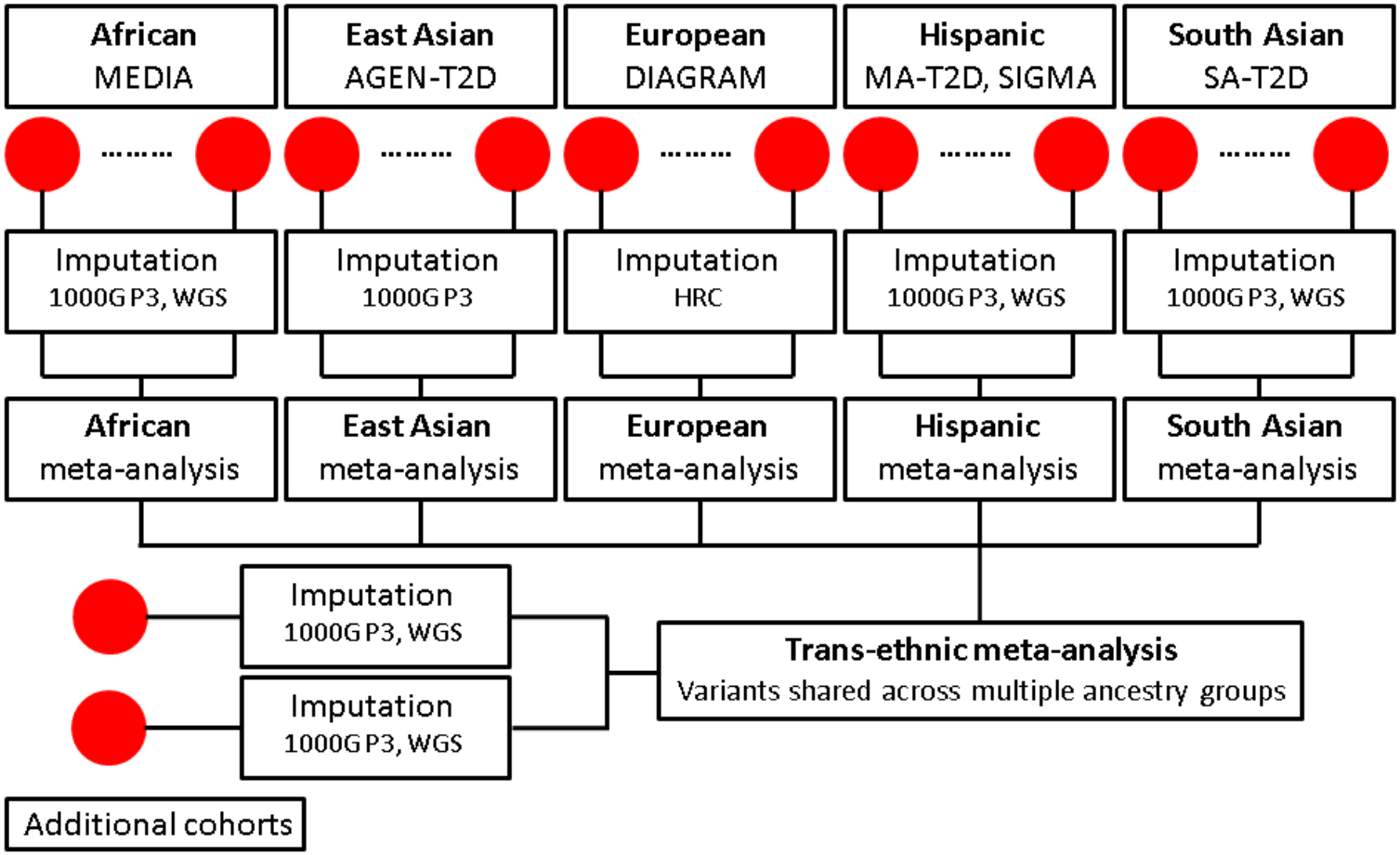
Mega-cohorts: Power in numbers



Nelson, ASHG 2017

DIAMANTE

(DIAbetes Meta-Analysis of Trans-Ethnic association studies)



DIAMANTE: sample sizes

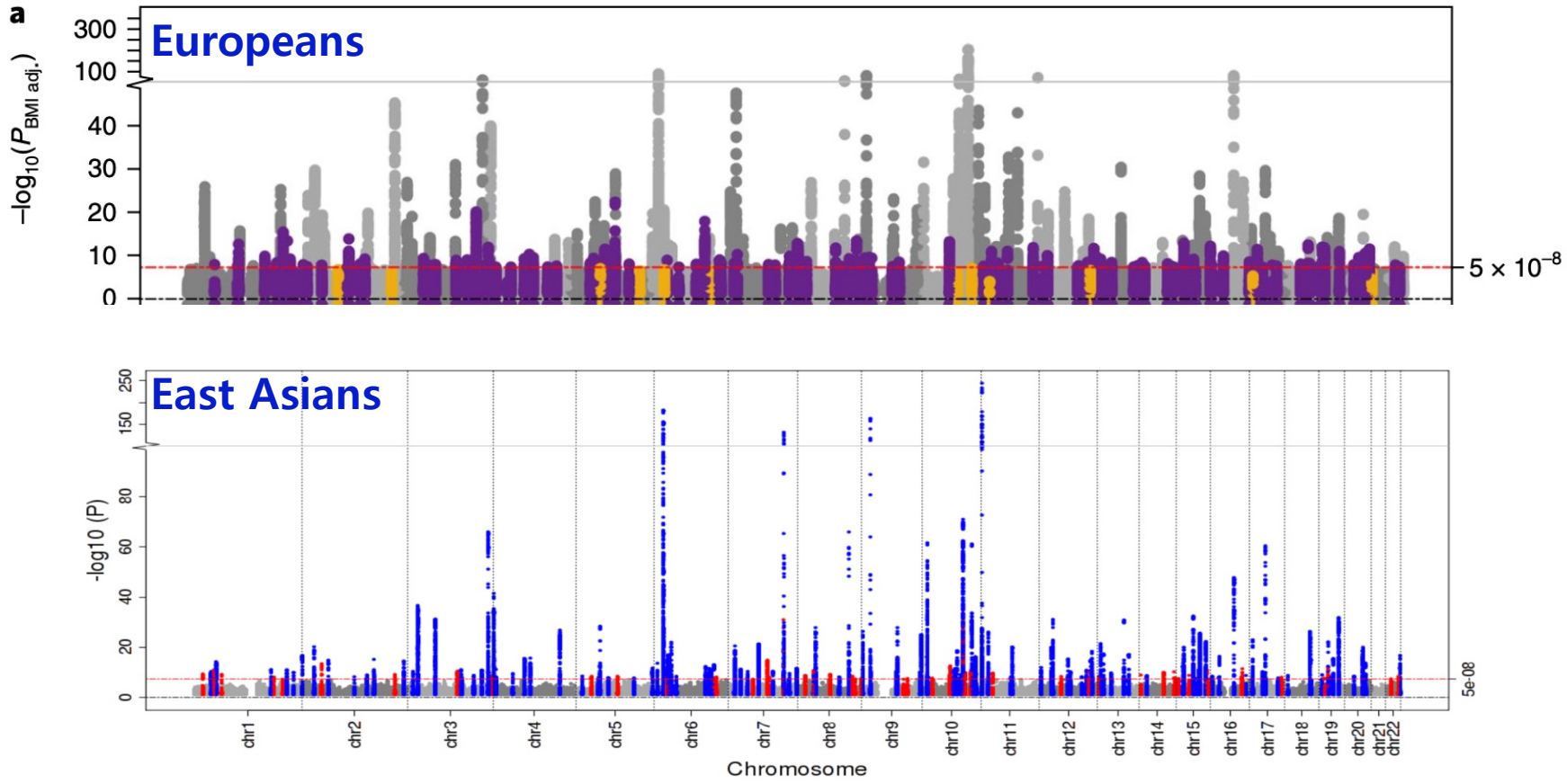
국립보건연구원
National Institute of Health

- A largest diabetes genetics study 1.49 million samples

Ancestry	# of cohorts	# of Cases	# of Controls
African	22	15,487	23,709
East Asian	**23**	**77,418**	**356,122**
European	32	80,154	853,816
Hispanic/Latino	14	12,385	21,423
South Asian	15	16,540	32,952
Total	**106**	**201,984**	**1,288,022**

DIAMANTE: Europeans & East-Asian, T2D GWAS

- EUR: 243 T2D loci, 135 new loci
- EAS: 183 T2D loci, 61 new loci
- In total, 332 lead variants in both datasets can be compared



DIAMANTE: European & East-Asian papers

국립보건연구원
National Institute of Health

European, N ~ 900K

nature genetics
ARTICLES
https://doi.org/10.1038/s41588-018-0241-6

Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps

We expanded GWAS discovery for type 2 diabetes (T2D) by combining data from 898,130 European-descent individuals (9% cases), after imputation to high-density reference panels. With these data, we (i) extend the inventory of T2D-risk variants (243 loci, 135 newly implicated in T2D predisposition, comprising 403 distinct association signals); (ii) enrich discovery of lower-frequency risk alleles (80 index variants with minor allele frequency <5%, 14 with estimated allelic odds ratio >2); (iii) substantially improve fine-mapping of causal variants (at 51 signals, one variant accounted for >80% posterior probability of association (PPA)); (iv) extend fine-mapping through integration of tissue-specific epigenomic information (islet regulatory annotations extend the number of variants with PPA >80% to 73); (v) highlight validated therapeutic targets (18 genes with associations attributable to coding variants); and (vi) demonstrate enhanced potential for clinical translation (genome-wide chip heritability explains 18% of T2D risk; individuals in the extremes of a T2D polygenic risk score differ more than ninefold in prevalence).

Array-based genome-wide association studies (GWAS) have identified ~140 loci influencing the risk of T2D[1–3]. Follow-up of these genetic discoveries has been compromised by the incomplete coverage of the most frequently used genotyping arrays, the imperfect performance of the reference panels available for imputation, extensive local linkage disequilibrium (LD), and inadequate sample sizes. These factors together have limited the power to detect low-frequency alleles with population-scale effects, to deliver clinically relevant risk prediction, and to define molecular mechanisms involved in disease predisposition. Here, we address the limitations of previous studies by combining GWAS from ~900,000 Europeans with dense, high-quality imputation, producing the most comprehensive view to date of the genetic contribution to T2D with respect to locus discovery, causal-variant resolution, and mechanistic insight.

Results

Study overview. We combined data from 32 GWAS, including 74,124 T2D cases and 824,006 controls of European ancestry. The effective sample size (N_{eff}) of 231,436 represents a 3.2-fold increase in N_{eff} relative to the largest previous genome-wide study of T2D risk in Europeans[3]. After harmonized quality control, 31 of the 32 GWAS were imputed using 64,976 whole-genome-sequenced haplotypes from the Haplotype Reference Consortium (HRC)[4]; the exception was the deCODE GWAS, which was imputed with a population-specific reference panel of 30,440 Icelandic haplotypes[5] (Methods and Supplementary Table 1). We conducted T2D-association analyses with and without adjustment for body-mass index (BMI).

Discovery of novel loci for T2D susceptibility. We tested for T2D association with ~27 million variants passing quality-control filters, ~21 million of which had a minor allele frequency (MAF) <5%. Our meta-analysis identified variants at 231 loci reaching genome-wide significance ($P < 5 \times 10^{-8}$) in the BMI-unadjusted analysis (N_{eff} 231,436) and 152 in the smaller (N_{eff} 157,401) BMI-adjusted analysis. Of the 243 loci identified across these two analyses, 135 mapped outside regions previously implicated in T2D risk (Methods, Fig. 1 and Supplementary Table 2).

Among samples not included in previous discovery efforts (42,734 cases and 497,261 controls), we replicated associations (directionally consistent, $P < 0.05$) at 126 of 140 previously reported T2D loci, including all 106 regions first discovered in European-only or transancestry efforts[3,6–8] and 20 initially reported in studies of non-European individuals[9,10]. The 14 loci not replicated were all first identified in non-European-ancestry samples: at five, the reported lead variant had MAF <1% in Europeans.

Multiple association signals at T2D-susceptibility loci. Across the 243 associated loci, we identified 160 additional signals at 'locus-wide' significance ($P < 10^{-5}$; Methods), 110 of which were within previously reported T2D loci. Overall, we observed one signal at 151 loci, and two to ten signals at the remaining 92 loci (Supplementary Table 2), for a total of 403 distinct T2D-association signals.

We observed the first evidence for multiple signals at the *TCF7L2* locus. In addition to rs7903146, the largest-effect common-variant signal for T2D in Europeans, we detected seven secondary signals, each represented by noncoding index variants (0.5% < MAF < 47.6%, 1.05 < odds ratio (OR) < 1.36).

In the ~1-Mb telomeric region of chromosome 11 that encompasses the (previously annotated) *INS–IGF2* and *KCNQ1* loci, we detected 15 distinct signals (0.15% < MAF < 42.8%, 1.03 < OR < 1.68). This multiplicity of signals in a region notable for complex imprinting effects, and several strong biological candidates (*INS*, *IGF2*, *KCNQ1*, and *CDKN1C*), illustrates a previously unrecognized degree of complexity in the risk-variant architecture at this locus.

The effects of BMI and sex. At most T2D loci, there were only minimal differences in the estimated T2D effect sizes between BMI-adjusted and unadjusted models (Methods and Fig. 2). However, at index SNPs for 41 signals (mapping to 21 known and 16 novel loci), we observed significant differences in effect sizes between BMI-adjusted and unadjusted analyses ($P_{adj} < 0.00012$, corrected for 403

Mahajan *et al.* Nature Genetics 2018

East Asian, N ~ 434K

Article

Identification of type 2 diabetes loci in 433,540 East Asian individuals

https://doi.org/10.1038/s41586-020-2263-3
Received: 24 June 2019
Accepted: 2 March 2020
Check for updates

Cassandra N. Spracklen[1,2,92], Momoko Horikoshi[3,92], Young Jin Kim[4,92], Kuang Lin[5,92], Fiona Bragg[6], Sanghoon Moon[4], Ken Suzuki[2,6,7,8], Claudia H. T. Tam[9,10], Yasuharu Tabara[11], Soo-Heon Kwak[12], Fumihiko Takeuchi[13], Jirong Long[14], Victor J. Y. Lim[15], Jin-Fang Chai[15], Chien-Hsiun Chen[16], Masahiro Nakatochi[17], Jie Yao[18], Hyeok Sun Choi[19], Apoorva K. Iyengar[1], Hannah J. Perrin[1], Sarah M. Brotman[1], Martijn van de Bunt[20,21], Anna L. Gloyn[20,21,22,90], Jennifer E. Below[23,24], Michael Boehnke[25], Donald W. Bowden[26,27], John C. Chambers[28,29,30,31,32], Anubha Mahajan[20,21,91], Mark I. McCarthy[20,21,22,91], Maggie C. Y. Ng[23,26], Lauren E. Petty[23,24], Weihua Zhang[29,30], Andrew P. Morris[31,33,34], Linda S. Adair[35], Masato Akiyama[6,36,37], Zheng Bian[38], Juliana C. N. Chan[9,10,39,40], Li-Ching Chang[16], Miao-Li Chee[41], Yii-Der Ida Chen[38], Yuan-Tsong Chen[16], Zhengming Chen[5], Lee-Ming Chuang[42,43], Shufa Du[35], Penny Gordon-Larsen[35], Myron Gross[44], Xiuqing Guo[18], Yu Guo[38], Sohee Han[4], Annie-Green Howard[45], Wei Huang[46], Yi-Jen Hung[47,48], Mi Yeong Hwang[4], Chii-Min Hwu[49,50], Sahoko Ichihara[51], Masato Isono[13], Hye-Mi Jang[4], Guozhi Jiang[9,10], Jost B. Jonas[52], Yoichiro Kamatani[6,53], Tomohiro Katsuya[54,55], Takahisa Kawaguchi[11], Chiea-Chuen Khor[56,57], Katsuhiko Kohara[58], Myung-Shik Lee[59,60], Nanette R. Lee[61], Liming Li[62,63], Jianjun Liu[56,64], Andrea O. Luk[9,10], Jun Lv[62,63], Yukinori Okada[6,65], Mark A. Pereira[66], Charumathi Sabanayagam[41,67,68], Jinxiu Shi[48], Dong Mun Shin[4], Wing Yee So[9,39], Atsushi Takahashi[6,69], Brian Tomlinson[9,70], Fuu-Jen Tsai[71], Rob M. van Dam[15,64], Yong-Bing Xiang[72,73], Ken Yamamoto[74], Toshimasa Yamauchi[3], Kyungheon Yoon[4], Canqing Yu[62,63], Jian-Min Yuan[75,76], Liang Zhang[41], Wei Zheng[14], Michiya Igase[77], Yoon Shin Cho[19], Jerome I. Rotter[78,79], Ya-Xing Wang[80], Wayne H. H. Sheu[48,50,81], Mitsuhiro Yokota[82], Jer-Yuarn Wu[16], Ching-Yu Cheng[41,67,68], Tien-Yin Wong[41,67,68], Xiao-Ou Shu[14], Norihiro Kato[13], Kyong-Soo Park[12,83,84], E-Shyong Tai[15,64,85], Fumihiko Matsuda[11], Woon-Puay Koh[15,86], Ronald C. W. Ma[9,10,39,40], Shiro Maeda[3,87,88], Iona Y. Millwood[6,89], Juyoung Lee[4], Takashi Kadowaki[3] ✉, Robin G. Walters[5,89] ✉, Bong-Jo Kim[4] ✉, Karen L. Mohlke[1] ✉ & Xueling Sim[38] ✉

Meta-analyses of genome-wide association studies (GWAS) have identified more than 240 loci that are associated with type 2 diabetes (T2D)[1,2]; however, most of these loci have been identified in analyses of individuals with European ancestry. Here, to examine T2D risk in East Asian individuals, we carried out a meta-analysis of GWAS data from 77,418 individuals with T2D and 356,122 healthy control individuals. In the main analysis, we identified 301 distinct association signals at 183 loci, and across T2D association models with and without consideration of body mass index and sex, we identified 61 loci that are newly implicated in predisposition to T2D. Common variants associated with T2D in both East Asian and European populations exhibited strongly correlated effect sizes. Previously undescribed associations include signals in or near *GDAP1*, *PTF1A*, *SIX3*, *ALDH2*, a microRNA cluster, and genes that affect the differentiation of muscle and adipose cells[3]. At another locus, expression quantitative trait loci at two overlapping T2D signals affect two genes—*NKX6-3* and *ANK1*—in different tissues[4–6]. Association studies in diverse populations identify additional loci and elucidate disease-associated genes, biology, and pathways.

Type 2 diabetes is a common metabolic disease that is primarily caused by insufficient insulin production and/or secretion by the pancreatic β cells, and insulin resistance in peripheral tissues[7]. Most genetic loci associated with T2D have been identified in populations of European (EUR) ancestry, including a recent meta-analysis of GWAS of nearly 900,000 individuals of European ancestry that identified more than 240 loci that influence risk of T2D[1]. Differences in allele frequency between ancestries affect the power to detect associations within a population, particularly among variants that are rare or monomorphic in one population but more frequent in another[2,3]. Although smaller than studies of European populations, a recent T2D meta-analysis of data from almost 200,000 Japanese individuals identified 28 additional loci[2]. The relative

A list of affiliations appears at the end of the paper.

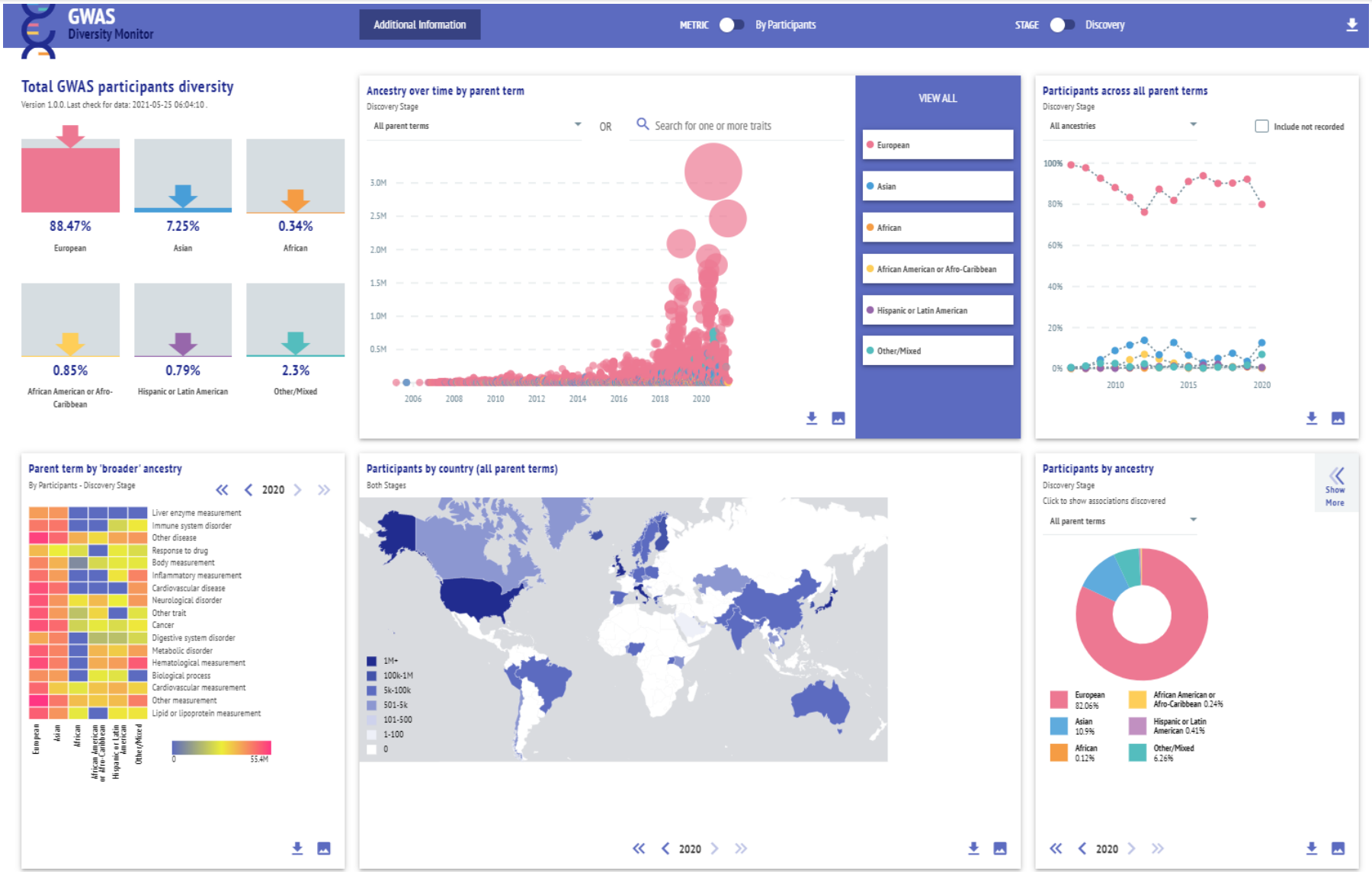
Spracklen, Horikoshi, Kim, Lin *et al.* Nature 2020

한국인칩사업 소개

(The Korea Biobank Array optimized for Korean Genome)

인종별 GWAS 연구 동향

국립보건연구원
National Institute of Health

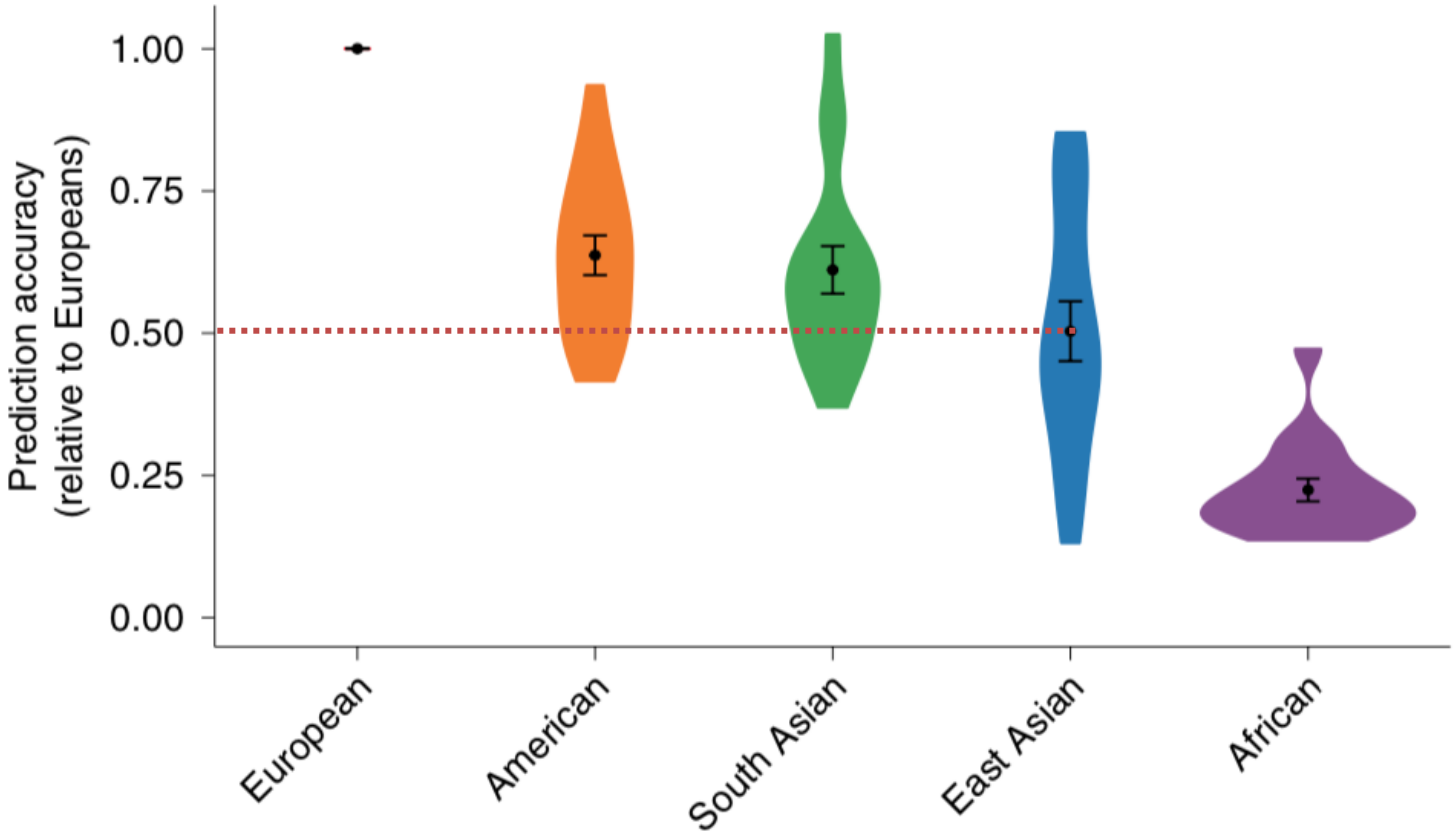


Nature Genetics **52**, 242–243 (2020)

Challenges in Clinical practice

국립보건연구원
National Institute of Health

- About 88% of GWAS were conducted in Europeans
  - Europeans are 16% of worldwide populations
- European based GRS showed poor predictability in non-Europeans



Martin et al. Nature Genetics 2019

한국인칩 필요성 : 기존 연구기법의 한계

국립보건연구원 National Institute of Health

- 한국인 만성질환 유전체 연구를 위한 대규모 인구집단 유전체 연구의 기존 연구기법의 문제점 대두
  - 유전변이 칩 : 서양인 중심 설계, 한국인 염기서열정보 미반영
    * 낮은 genomic coverage(~75%, 1KG ASN, MAF 5% 기준)
  - 차세대염기서열분석 기법
    * 높은 계산력과 유전변이 칩 대비 수십 배의 분석 시간 요구

- 이러한 한계 극복을 위한 인종 특이칩 제작
  - 인종 별 염기서열 정보 기반, 각 인종의 질환 유전체 연구에 최적화
    * 인종별 1000게놈 프로젝트 phase 3 서양인(503명), 동아시아인(504명)
  - 낮은 비용 (기존칩 대비 약 3-5배, NGS 대비 약 10배 절감)

한국인(유전체)칩 제작 과정 : 마커선별 기준 및 방법

국립보건연구원 National Institute of Health

KORV1.0

About 2,500 Sequencing data, 60,000 chip data Korean samples
Public data (1,000 genomes, ESP data etc.)

Genome-Wide Tagging

Functional Variants

GWAS markers

Pharmacogenomic markers

Marker Screening

1. Validate markers (AxiomGD 26M)
2. Design unvalidated markers
3. Check for suspected or non-unique SNPs

Select the Best Performing markers
(2.6M Axiom custom Chip)

Variants selection

833K Korean chip(KOV1.0)제작, 7,000명 유전체데이터 생산 베타테스트(검증)

한국인칩 제작 과정

국립보건연구원
National Institute of Health

칩제작과정

유전변이 칩 시범 제작

Re-design

정확도 99.5%
재현성 99.5%

- Imputation Backbone
  - MAF > 5%
  - genome coverage > 85% (r^2>0.8)
- Exonic variants, eQTL, ADME markers
- GWAS hits, forensic markers, etc.
- 유전변이 수는 80만개 이상

한국인칩 제작 (Korea Biobank Array)

- Call rate > 99%
- 원시데이터 가공, 정도관리 및 기초분석

한국인칩 특징(장점) 및 활용방안

국립보건연구원 National Institute of Health

- 장점
  - 높은 genomic coverage (95% for MAF ≥ 5%)
  - 기능변이 (약 2천 5백명 시퀀싱 기반)
  - 기 보고된 GWAS loci (GWAS cotalog)
  - 낮은 가격
  - 데이터 분석을 위한 정도관리(QC) 및 분석 파이프라인 제공
  - 한국인칩을 이용한 유전체데이터 표준화

- 활용방안
  - 다양한 질환 연관 신규 유전변이 발굴
  - 기 보고된 변이의 Replication study
  - 약물유전체연구 등 임상분야 활용

한국인칩사업 - 추진경과

국립보건연구원
National Institute of Health

- 한국인칩사업 계획안 마련 ('13)
- 국내 유전체 관련 학회 임원 모임을 통해 사업 홍보 및 의견 수렴
  - 한국유전체학회 운영위원회 ('14.4.21, 30여명 참석)
  - 생물정보학학회 확대이사회 ('14.4.18, 40여명 참석)
- 한국인칩사업 설명회 개최
  - 일정 : '14.4.24, 15:00-18:00, 유전체센터 세미나실 (510호)
  - 참석자 : 60여명 (패널위원 5명)
  - 패널회의 안건 : 칩 디지인, 공개 및 활용 등

그림. 한국인칩사업 설명회 개최 관련 그림

한국인칩사업 - 추진경과

- 자문회의 구성 ('14.5)
  - 목적 : 한국인칩 제작을 위한 컨텐츠 구성
  - 자문단 구성(20명) : 유전체학, 생물정보학, 의학 등 관련 분야 전문가로 구성
  - 논의사항 : 컨텐츠 구성 및 확보 방법, 컨소시엄 구성 및 역할

- 한국인칩 성과(제작) 발표회

그림. 한국인칩 성과(제작) 발표회 관련 그림

한국인칩사업 - 한국인칩 제작 발표

보건복지부

보 도 참 고 자 료

배 포 일	2015.5.13 / (총 5 매)	담당부서	보건의료기술개발과 질병관리본부 형질연구과
보건의료기술개발과	정 영 기 과장	전 화	044-202-2920
	정 지 원 연구관		044-202-2923
형질연구과	김 봉 조 과장	전 화	043-719-8870
	김 영 진 연구원		043-719-8894

만성질환 규명의 열쇠, 한국인 맞춤형 칩 제작

– 한국인칩 제작 발표 –

The Science Times
June 19,2015

Notice 대중소 기업간 IoT·제조 협업 생태계 구축 합의

분야별뉴스 | 과학기술 | 신소재 · 신기술

유전정보 담은 '한국인칩' 첫 선

한국인 만성질환의 치료 길 열려

한국인에게 흔히 발생하는 암이나 당뇨, 그리고 고혈압과 같은 만성질환의 유전적 요인을 근본적으로 규명할 수 있는 길이 열렸다.

최근 발표된 한국인칩 시제품 (c) 국립보건연구원

보건복지부는 최근 한국인의 질병예방 예측 및 미래 유전체의학의 조기실현을 위해, 맞춤형 유전체칩인 '한국인칩'을 제작하여 생산한다고 밝혔다.한국인칩이란 한국인만의 특이적인 유전체 정보가 담겨있는 일종의 반도체칩을 말한다.

한국인이 갖고 있는 유전변이 중 단백질 기능에 영향을 주는 20만개의 유전변이와 한국인의 유전체를 대표하는 60만개의 유전변이가 담겨 있어 한국인 질병 유전체 연구에 최적화된 칩이라 할 수 있다.

한국인칩 컨소시엄 운영 (17년 ~)

국립보건연구원 National Institute of Health

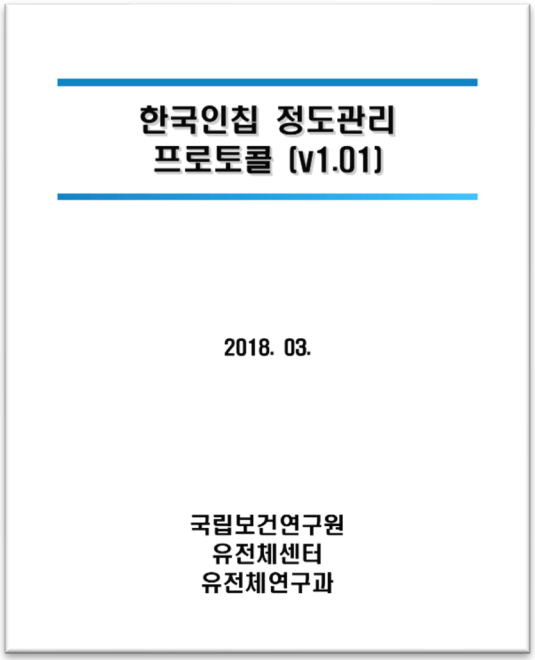
- 한국인칩 유전체정보 활용을 위한 민간 전문가 주도의 한국인칩 컨소시엄 운영 (한국유전체학회 주관)
  - 참여 연구자 간 공유 및 협력의 장으로 활용
  - 한국인칩 데이터 정도관리 및 분석 방법 교육 (워크숍 및 교육 11회 개최)
  - 45개 기관 145명 참여(23개 대학교, 9개 병원, 5개 연구소, 8개 기업)
- 한국인칩 유전체정보 생산 규모: 약 30만명 이상
  - (국립보건연구원) 코호트 기반 약 20만명 생산·정도관리 및 공개 중
  - (국내 연구자) 수요조사 약 10만명 생산하여 다양한 질병관련 유전체정보 생산 및 분석 중

한국인칩 지원 : 홈페이지, 분석프로토콜, 한국인칩백서

국립보건연구원
National Institute of Health

- 홈페이지 운영: www.koreanchip.org
- 한국인칩 정도관리·분석 프로토콜 개발·공개 및 데이터분석 교육 및 워크숍
- 한국인칩사업 백서 발간 및 배포

<한국인칩사업 홈페이지>



<프로토콜 개발·교육>

<한국인칩 백서 발간>

SCIENTIFIC REPORTS

OPEN

The Korea Biobank Array: Design and Identification of Coding Variants Associated with Blood Biochemical Traits

Received: 26 July 2018
Accepted: 11 December 2018
Published online: 04 February 2019

Sanghoon Moon[1], Young Jin Kim[1], Sohee Han[1], MiYeong Hwang[1], Dong Mun Shin[1], Min Young Park[3], Yontao Lu[4], Kyungheon Yoon[1], Hye-Mi Jang[1], Yun Kyoung Kim[1], Tae-Joon Park[1], Dae Sub Song[2], Jae Kyung Park[2], Jong-Eun Lee[3] & Bong-Jo Kim[1]

Moon et al. Scientific Reports 2019

한국인칩 활용 성과

국립보건연구원 National Institute of Health

- ‘22.03. 기준 71편 논문 발표

연번	논문 제목	저널명	연도
1	Identification of type 2 diabetes loci in 433,540 East Asian individuals	Nature	2020
2	Genome-wide association study in a Korean population identifies six novel susceptibility loci for rheumatoid arthritis	Ann Rheum Dis	2020
3	Dietary Protein and Fat intake affects diabetes risk with CDKAL1 genetic variants in Korean Adults	Int J Mol Sci	2020
4	Causal Associations Between Serum Bilirubin levels and decreased stroke risk: A two-sample mendelian randomization study	Arterioscler Thromb Vasc Biol	2020
5	A novel susceptiblity locus near GRIK2 associated ith erosive esophagitis in a Korean cohort	Clin Transl Gastroenterol	2020
6	Role of aldehyde dehydrogenases, alcohol dehydrogenase 1B genotype, alcohol consumption, and their combination in breast cancer in East Asian women	Sci Rep	2020
7	Combination of Genome-Wide Polymorphisms and Copy Number Variations of Pharmacogenes in Korean	J Personalized Medicine	2021
8	A Genome-Wide Association Study of Novel Genetic Variants Associated With Anthropometric Traits in Korean	Frontiers in Neurology	2021
9	A Genome-Wide Association Study Hypertensive Kidney Disease on Korean Man	Genes	2021

한국인(유전체)칩 상용화

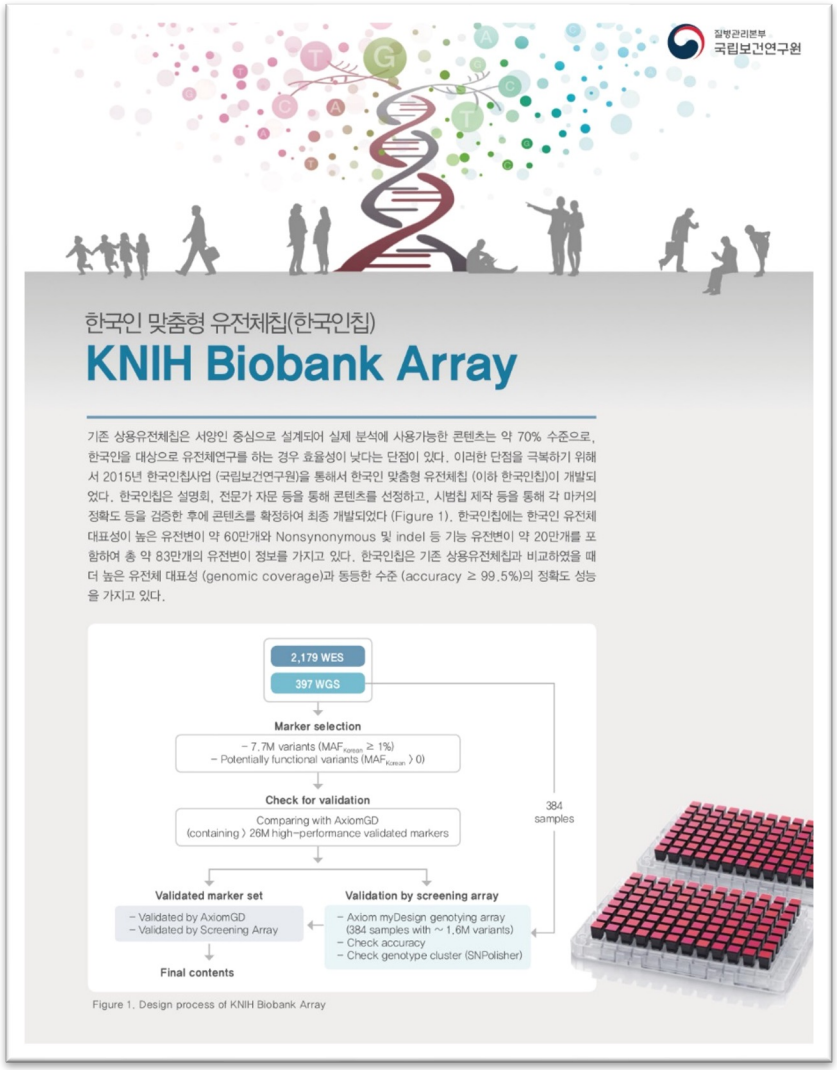
- 한국인칩 상표권(KNIH Korea Biobank Array) 출원
- ‘19년 국내 6개 사업체/기관 기술이전
  (디엔에이링크, 테라젠이텍스, 마크로젠, 보훈병원, 클리노믹스, SCL헬스케어)

Axiom KORV1.0 96 – Array Plate
550604425442907241665 6

한국인(유전체)칩 상용화

국립보건연구원
National Institute of Health

- 한국인칩 소개 브로슈어 (http://koreanchip.org)



질병관리본부 국립보건연구원

한국인 맞춤형 유전체칩(한국인칩)

KNIH Biobank Array

기존 상용유전체칩은 서양인 중심으로 설계되어 실제 분석에 사용가능한 콘텐츠는 약 70% 수준으로, 한국인을 대상으로 유전체연구를 하는 경우 효율성이 낮다는 단점이 있다. 이러한 단점을 극복하기 위해서 2015년 한국인칩사업 (국립보건연구원)을 통해서 한국인 맞춤형 유전체칩 (이하 한국인칩)이 개발되었다. 한국인칩은 설명회, 전문가 자문 등을 통해 콘텐츠를 선정하고, 시범칩 제작 등을 통해 각 마커의 정확도 등을 검증한 후에 콘텐츠를 확정하여 최종 개발되었다 (Figure 1). 한국인칩에는 한국인 유전체 대표성이 높은 유전변이 약 60만개와 Nonsynonymous 및 indel 등 기능 유전변이 약 20만개를 포함하여 총 약 83만개의 유전변이 정보를 가지고 있다. 한국인칩은 기존 상용유전체칩과 비교하였을 때 더 높은 유전체 대표성 (genomic coverage)과 동등한 수준 (accuracy ≥ 99.5%)의 정확도 성능을 가지고 있다.

Figure 1. Design process of KNIH Biobank Array

한국인 맞춤형 유전체칩(한국인칩)
KNIH Biobank Array

Table 1 Contents of KNIH Biobank Arrary

Category	Description	# of variants
GWAS markers	Tag SNPs for genome-wide coverage (Autosomal and X chromosome)	595,957
	NHGRI GWAS catalog	7,635
Function variants	nonsynonymous SNPs, Indels, etc.	196,944
	eQTL	16,097
Pharmacogenomic	Absorption, distribution, metabolism, and excretion (ADME) markers	1,818
Immune related	Human Leukocyte Antigen (HLA) region	6,605
	Killer immunoglobulin-like receptor (KIR) region	1,495
Fingerprint	Fingerprint/sample tracking	248
MT/Y Chromosome	Common mitochondrial DNA variants	178
	Y chromosome markers	806
Total # of variants		827,783

Table 2 Genomic coverage of KNIH Biobank Array

Category	Total (MAF≥0.01)	Common (MAF≥0.05)	Less common (0.01≤MAF<0.05)
KNIH Biobank Array	89.86	95.38	73.65
A Chip	87.09	94.48	65.42
B Chip	84.38	92.27	61.24

*Genomic coverage was calculated based on imputation data and 8.7M variants (MAF≥0.01, 1KG P3 East Asians)

Table 3 Accuracy of KNIH Biobank Array

Compared platforms	Overlapped with KNIH Biobank Array		Accuracy, %	
	Subject	Marker	Overall	Hetero
Affymetrix Genome-wide human SNP array 5.0	6,949	41,246	99.8	99.5
Illumina HumanExome BeadChip v1.1	5,793	34,683	99.9	99.7
Exome sequencing (Illumina Hiseq 2000)	155	90,020	99.8	99.7

Accuracy: # of True genotypes / # of Total genotypes
Overall: Overall accuracy, Hetero: Accuracy of heterozygotes

한국인칩은 국립보건연구원 유전체연구과에서 디자인하였으며, 문의사항은 아래로 연락주세요.
e-mail koreanchip@gmail.com | homepage www.koreanchip.org

질병관리본부 국립보건연구원

최신 유전체칩 개발 동향

국립보건연구원 National Institute of Health

- Customized SNP chip 기반 최신 염기서열정보 반영

Arrays	Target Diseases	Main purpose	# of Contents	GW Tagging	Description
UK BioBank	Complex Diseases	Discovery	820K	Yes	UK BioBank 500,000 samples
Kaiser BioBank	Complex Diseases	Discovery	650K	Yes	Kaiser BioBank 100,000 samples
Taiwan Biobank (v1, v2)	Complex Diseases	Discovery	650~680K	Yes	Taiwan BioBank 100,000 samples
Million Veteran Program (MVP)	Complex Diseases	Discovery	700K	Yes	MVP ~1 million samples
Global Screening Array	Complex Diseases	Discovery	700K	Yes	Various consortia (designed by illumina)
Global Diversity Array	**Complex Diseases**	**Discovery**	**1.8M**	**Yes**	**All of Us project 300K samples**

한국인(유전체)칩 v2.0 개발 계획

국립보건연구원 National Institute of Health

- 한국인 정밀의료 연구를 위한 임상 유전변이 확대 및 최신 염기서열 정보 반영

	한국인칩 v1.1	한국인칩 v2.0
마커 수	약 83만	약 180만 (↑2.2배)
임상 연구 마커	• 약물반응 • 단백질 변형 • 만성질환 연관 변이 (~2013년까지) • 약 20만개	• 약물반응 변이 확대 • 단백질 변형 변이 확대 • 만성·희귀질환 연관 변이 (~2021년까지 최신 정보) • 임상 유전변이 • 약 100만개 (↑5배)
한국인 유전체 대표성	• 빈도 5% 이상: 95% • 빈도 1-5%: 73% • 500명 WGS 기반	• 빈도 5% 이상: 97% • 빈도 1-5%: 85% (↑12%) • 5,000명 WGS 기반
사용된 염기서열 정보 수	약 2,500 (동아시아인 및 한국인)	약 28,000 (동아시아인 및 한국인)

분양 유전체정보 현황 및 신청방법

분양 유전체 정보 현황 - 한국인유전체역학조사사업(KoGES)

국립보건연구원 National Institute of Health

※ 각 코호트별 조사단계를 클릭하시면 해당 분양코드북으로 연결됩니다. (통계기준: '21.3.)

대상자	조사단계/조사연도	역학정보		유전정보		
		참가자 수	변수 수	참가자 수	Platform	종류
지역사회 기반 코호트 (안성/안산) 내용 확인하기	1기 (기반)('01-'02년)	10,030	1,851	8,840	Affy. 5.0	· SNP · Imputation - HapMap - 1,000 genome
				1,000	CGH array chip	· CNV
				100	Illumina Hiseq	· Exome Seq
				50	Infinium HumanMethy450	· Methylation
				2,426	Exome chip	· SNP
				5,493	Korea Biobank Array**	· SNP(VCF, Genotype)
	2기 (1차 추적)('03-'04년)	8,603	1,571	-	-	-
	3기 (2차 추적)('05-'06년)	7,515	2,347	2,580	MetIDQ p180	· 대사체
	4기 (3차 추적)('07-'08년)	6,688	1,788	-	-	-
	5기 (4차 추적)('09-'10년)	6,665	2,046	450*	Infinium HumanMethy450	· Methylation
				5,098	Exome chip	· SNP (Genotype)
				1,528	Infinium Methylation EPIC array 850K	· Methylation
	6기 (5차 추적)('11-'12년)	6,238	2,122	-	-	-
	7기 (6차 추적)('13-'14년)	5,906	1,942	-	-	-
	8기 (7차 추적)('15-'16년)	6,318	1,634	-	-	-
	9기 (8차 추적)('17-'18년)	6,157	1,143	-	-	-
	반복 추적조사 통합자료 1기(기반)-9기(8차추적) ('01-'17년)	10,030	4,119	-	-	-

출처.질병관리청 국립보건연구원 국립중앙인체자원은행

분양 유전체정보 현황 - 한국인유전체역학조사사업(KoGES)

국립보건연구원 National Institute of Health

도시 기반 코호트 내용 확인하기	기반(04-13년)	173,205	1,232	3,693	Affy. 6.0	· SNP (Genotype)
				3,432	Exome chip	· SNP (Genotype)
				58,698	Korea Biobank Array**	· SNP(VCF, Genotype)
				822	Infinium Methylation EPIC array 850K	· Methylation
	1차 추적(12-16년)	65,614	1,047	-	-	-
농촌 기반 코호트 내용 확인하기	기반(05-11년)	28,337	1,030	3,666	Illumina Omni1	· SNP (Genotype)
				1,816	Affy. 6.0	· SNP (Genotype)
				3,068	Exome chip	· SNP (Genotype)
				8,105	Korea Biobank Array*	· SNP (VCF, Genotype)
	1차 추적(07-14년)	12,463	857	-	-	-
	2차 추적(08-16년)	11,399	857	-	-	-
	3차 추적(11-16년)	-	-	-	-	-
	4차 추적(14-16년)	-	-	-	-	-

지역사회, 도시, 농촌 기반 코호트 대상자 **72,299명**에 대한 한국인칩 유전체정보 분양 중

출처.질병관리청 국립보건연구원 국립중앙인체자원은행

분양 유전체정보 현황 – 기타 사업

국립보건연구원 National Institute of Health

대상자	조사연도	역학정보		유전정보		
		참가자 수	변수 수	참가자 수	Platform	종류
한국인 참조 유전체 정보	'12-'13년	-	-	397†	Illumina Hiseq	· whole genome seq (BAM, VCF)
유방암 환자군	'08년	-	-	2,165	Affy. 6.0	· SNP(Genotype)
위암 환자군	'08년	-	-	803	Affy. 6.0	· SNP(Genotype)
다낭난소증후군(PCOS)	'10년	-	-	422	Illumina Omni1	· SNP (Genotype)
알츠하이머병/파킨슨병 치매환자	'13-'14년	526	56	526‡	genotyping (taqman, sanger sequencing)	· SNP (LRRK2, APOE)

출처.질병관리청 국립보건연구원 국립중앙인체자원은행

유전체정보 분양 서식 및 신청방법

국립보건연구원
National Institute of Health

국립중앙인체자원은행은 한국인유전체역학조사사업(KoGES), 국민건강영양조사사업 등 질병관리청 국립보건연구원에서 수행하고 있는 사업을 통해 인구집단 기반 인체자원을 수집하고 있으며, 수집된 인체자원을 국내 보건의료 R&D에 활용될 수 있도록 연구자들에게 분양하고 있습니다.

- 인체자원 분양 이용은 국내 거주하는 한국인 연구자로서 IRB 승인(또는 면제) 받은 연구과제를 수행하는 연구책임자라면 누구나 분양 신청할 수 있으며, 국립중앙인체자원은행 분양위원회의 심의 이후 분양 승인이 결정되면 인체자원을 분양 받을 수 있습니다.

국립중앙인체자원은행 분양위원회

신청대상

국가연구개발사업 또는 국가기관, 정부출연기관, 민간 연구기관의 지원을 받아 연구과제를 수행하는 연구자
※ 인체유래물의 경우 연구비를 지원받는 과제에 한하여 분양

필요서류

인체자원이용계획서 ↓ 서약서 ↓ 개인정보수집·이용동의서 ↓
IRB 심의결과서 또는 면제확인서(자율서식) 연구계획서(자율서식)

신청방법

온라인 분양신청
가입방법 보기 ↓
인체자원 분양데스크(HuBIS_Desk) 바로가기

문의)
인체자원연구지원센터
ARS 1661-9070

출처.질병관리청 국립보건연구원 국립중앙인체자원은행

유전체정보 분양 서식 및 신청방법

국립보건연구원
National Institute of Health

분양신청 방법 및 절차

인체자원분양데스크 회원가입

- 질병보건통합관리시스템 사용자 가입 및 '인체자원 분양데스크 연구자 User' 권한 신청 *

* 1일 2회(11시, 17시) 관리자가 권한 승인 처리 진행

가입방법 보기

출처.질병관리청 국립보건연구원 국립중앙인체자원은행

“한국인칩을 이용한 대규모 유전체정보 분석” 강좌가 유전체연구에 도움이 되기를 바랍니다.

감사합니다