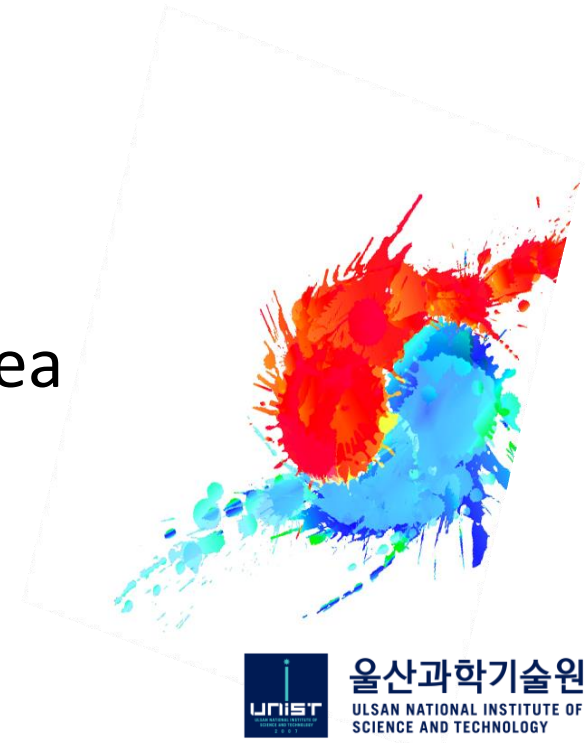


유전체 기술을 이용한 심혈관 질환 진단방법

2019. 12. 14.

전성원 (Sungwon Jeon)

KOGIC, UNIST, South Korea

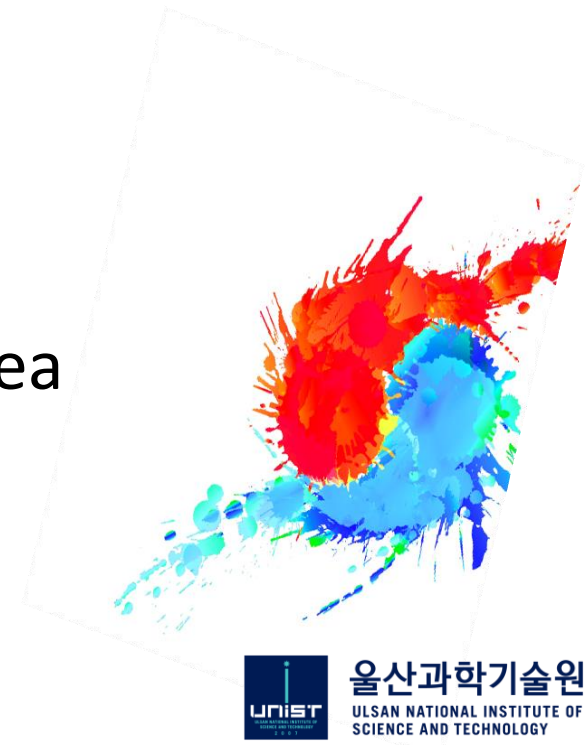


유전체 기술을 이용한 심혈관 질환 진단방법

2019. 12. 14.

전성원 (Sungwon Jeon)

KOGIC, UNIST, South Korea



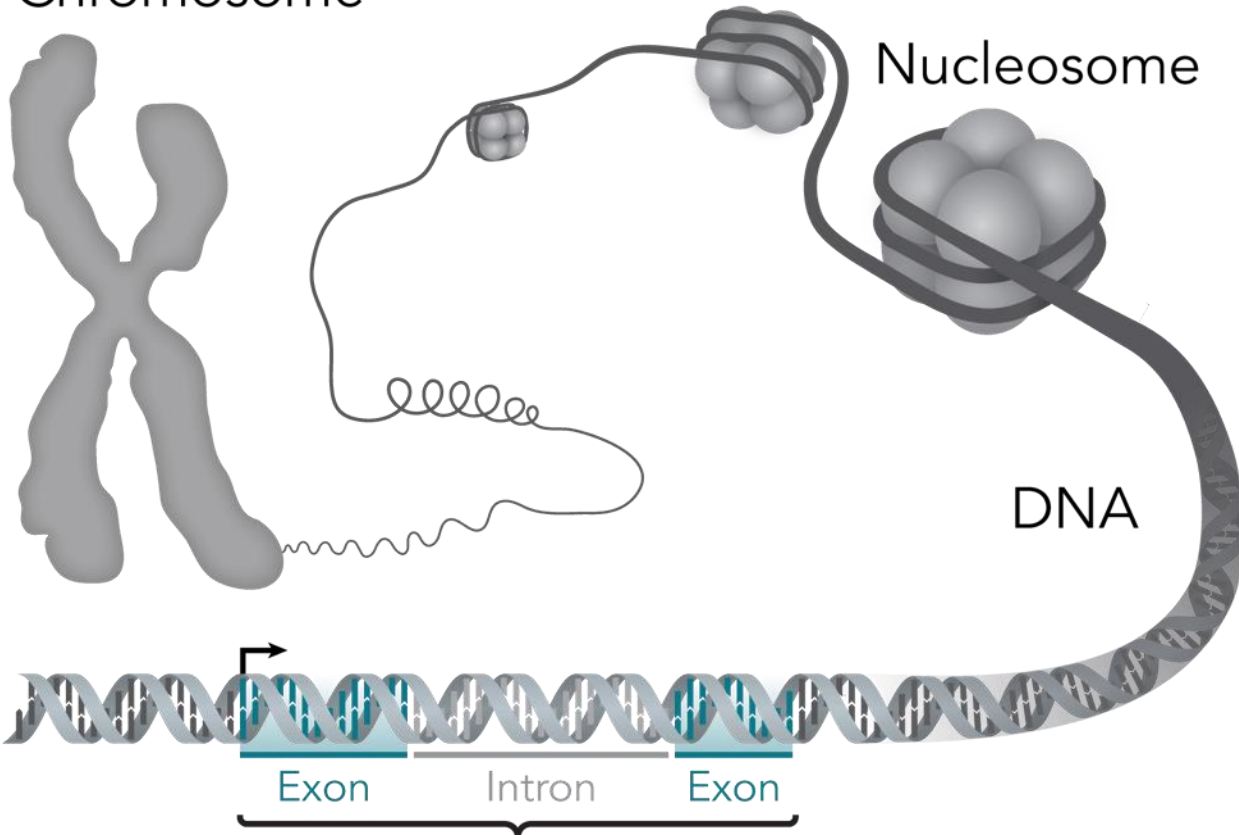
인간 게놈

유전체: 30억 염기 서열 (사람경우)

Chromosome

Nucleosome

DNA



Exon

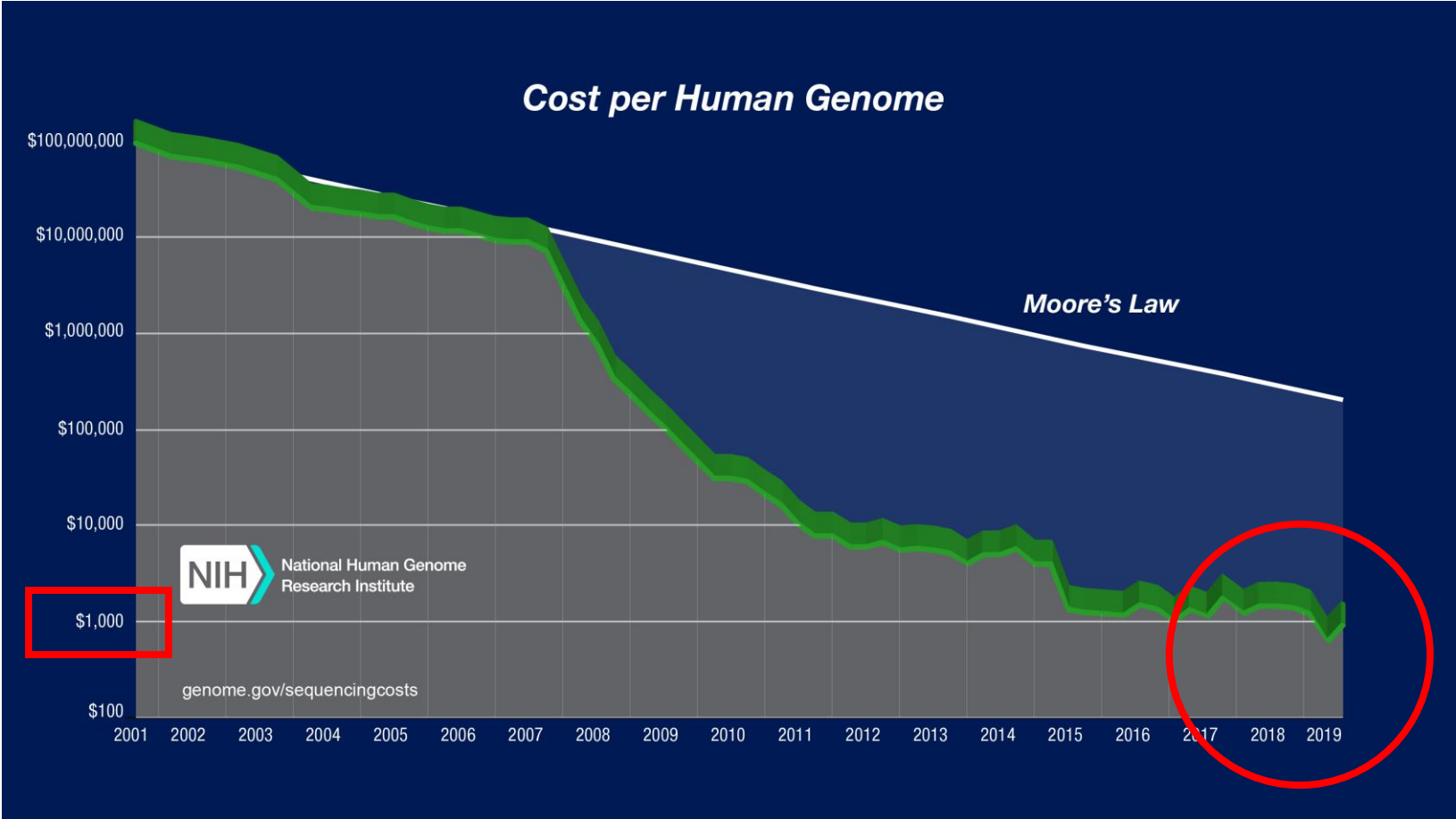
Intron

Exon

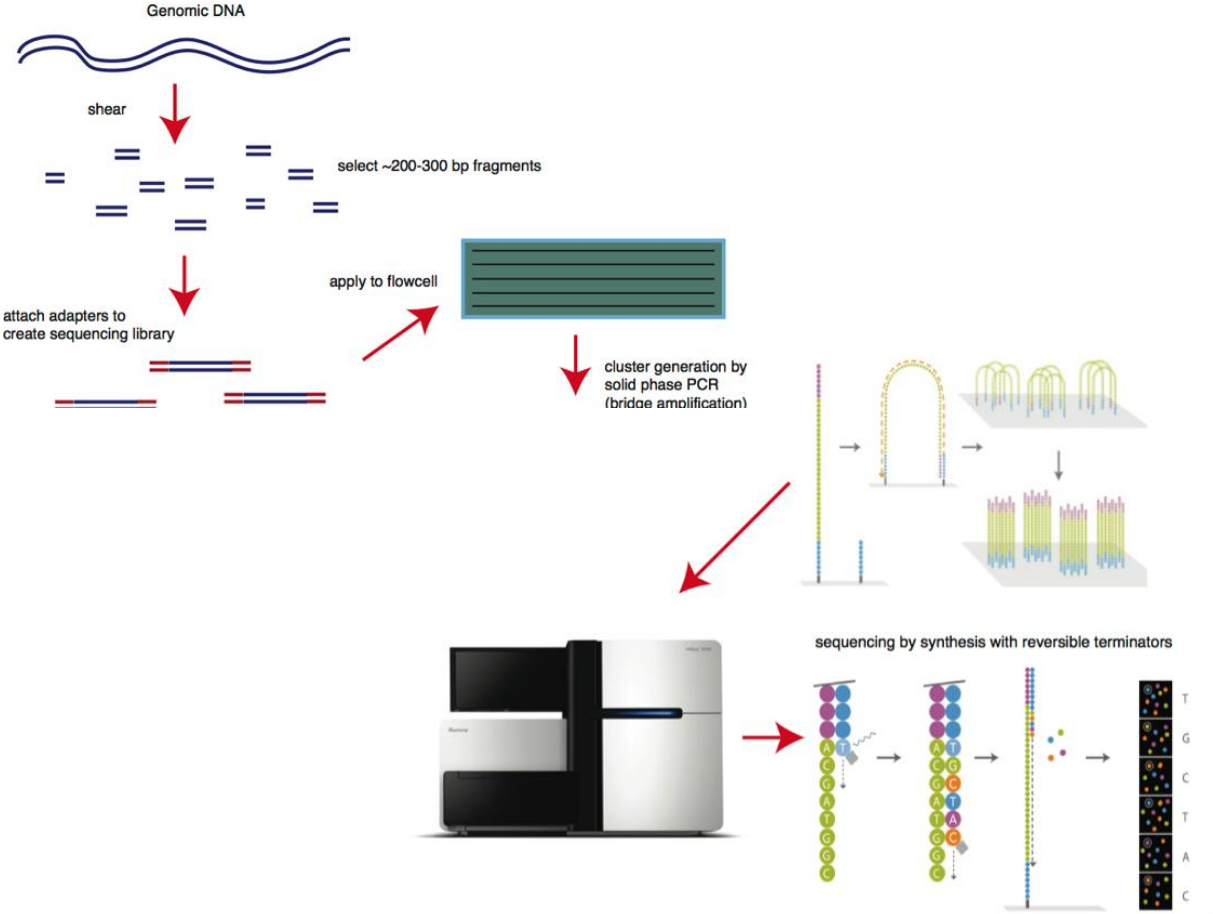
Gene

유전자: 기능성 단백질을 암호화 하는 영역 (2만개)

Human whole genome sequencing cost

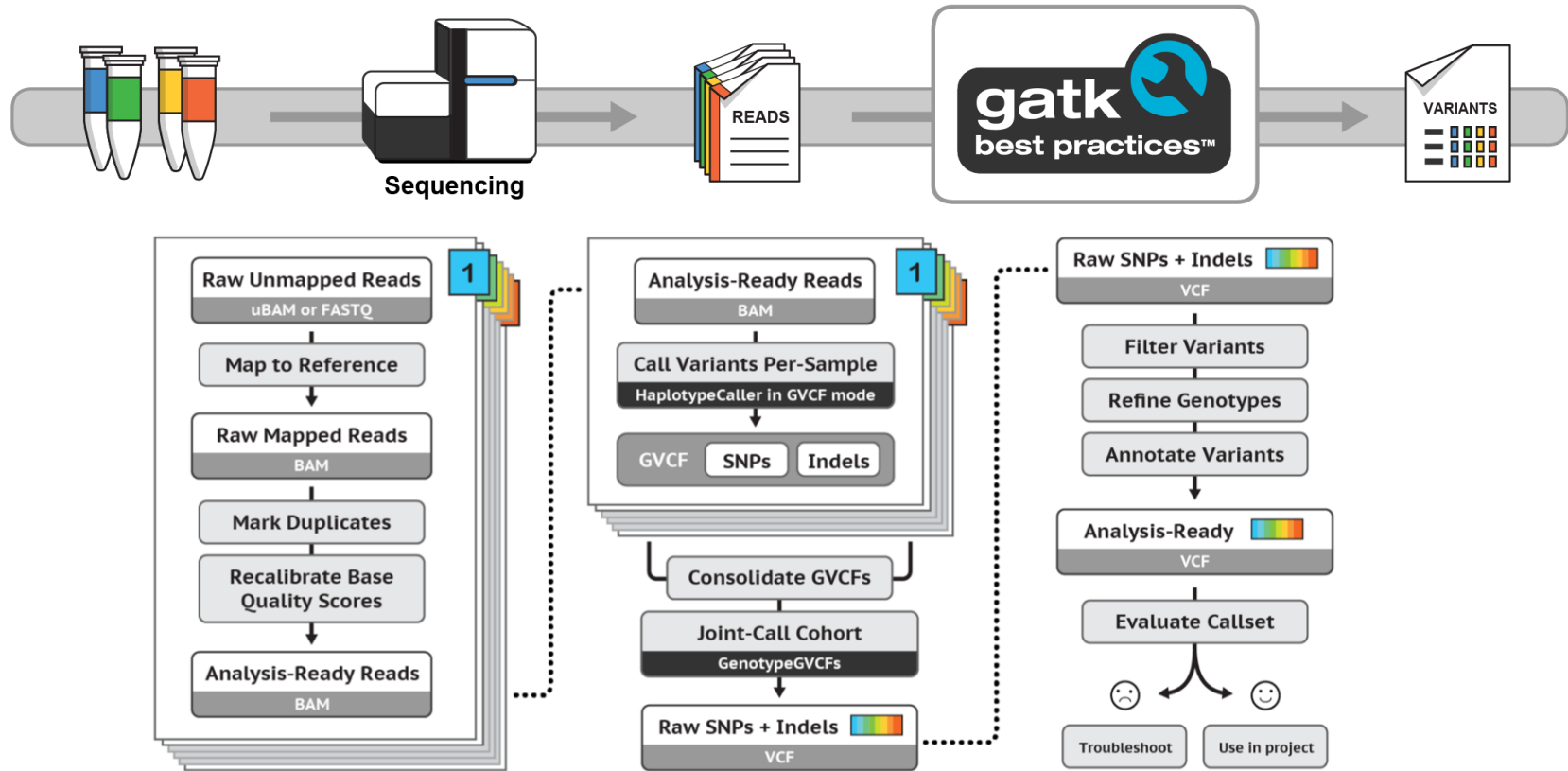


Next generation sequencing (NGS)



Bioinformatics (생정보학)

인간 게놈 분석 Pipeline



36 threads => 36 hours

<https://software.broadinstitute.org/gatk/>

Human whole genome sequencing

Human genome
sequencing data

$$3\text{Gb} \times 30\text{배수} = 90\text{Gb}$$



생정보 기초분석 이후

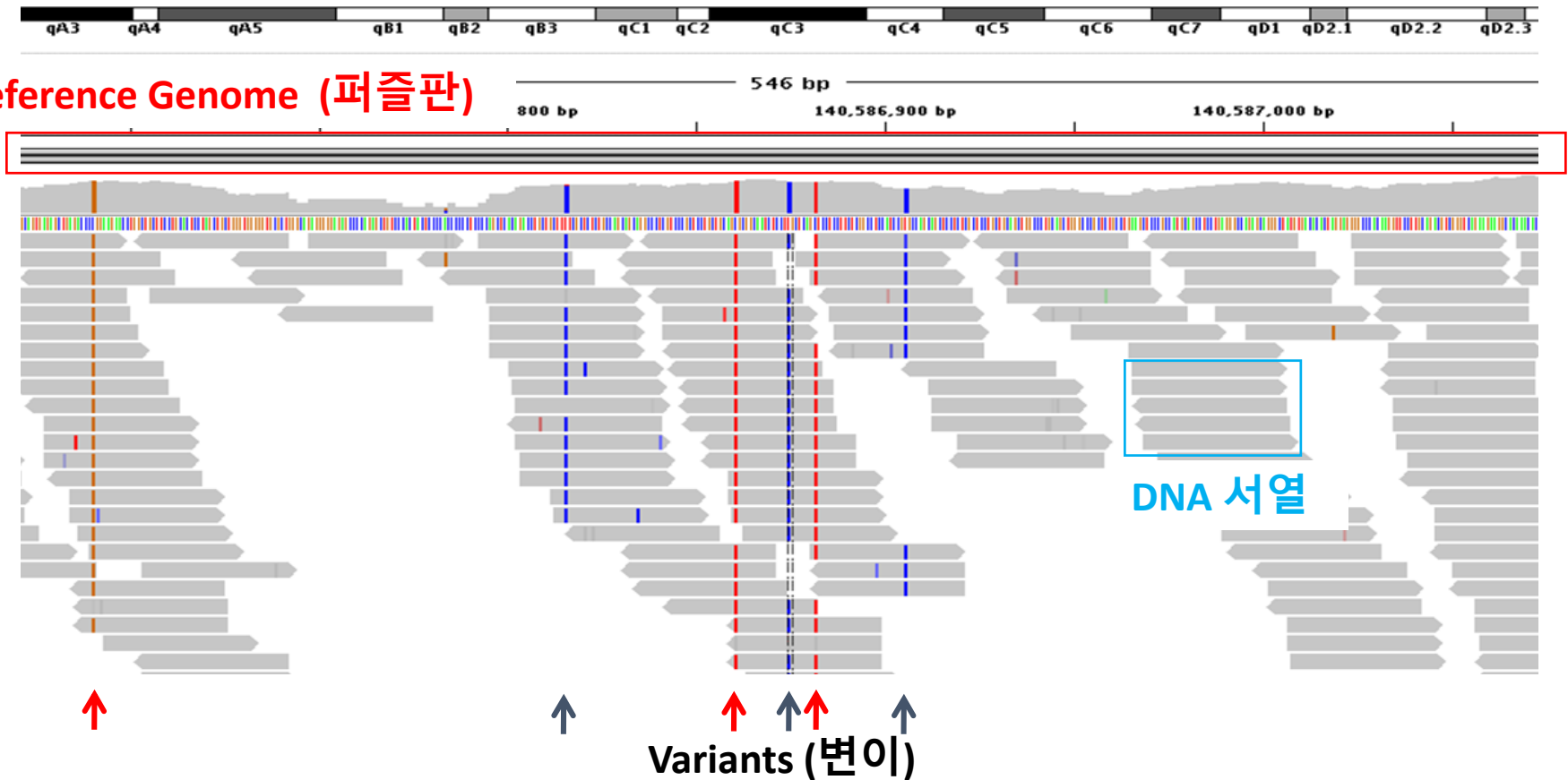
0.5~1TB/sample



인간 게놈 분석

유전체: 30억 염기 서열 (사람경우)

Reference Genome (퍼즐판)

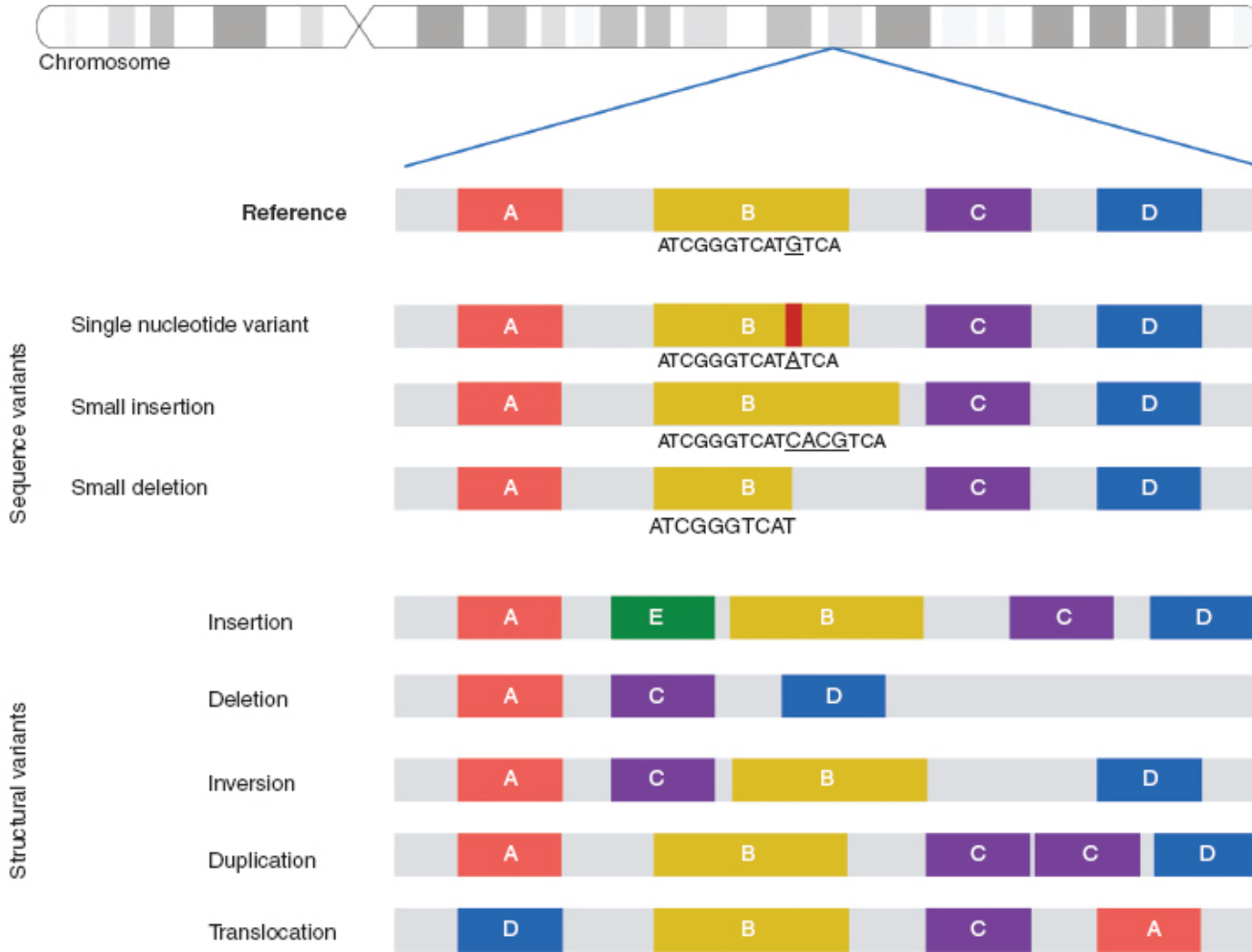


변이: 표준 서열과 다른 염기 서열
변이체: 총 변이의 합

보통 30X 정도는 해야한다.

게놈 변이

<https://neupsykey.com/genetics-3/>



Variome (변이체)

프로그램

```
1. jsungwon@lion: /BIO/Research/Project2/TGI-PAPGI-Genome-2016-05/Store/Panel_From_Veronika (ssh)
AA_AM AD_006 0 0 0 -9 G G G T G A T C A A A G G G A A T T G G G G C T C T A G T C T C C A T C T C
AA_AM AD_015 0 0 0 -9 A A T T G A T C A G G G G G A A C T A G G G T T C C A A T T T T C C T T T C
AA_AM AD_061 0 0 0 -9 G A T T A A T C G G A A G G A A T T A G A G C T C T A G T C T C C C T T C C
AA_AM AD_064 0 0 0 -9 G A T T G A T C A A G G G G A A C T G G G G C T T T G G C C C C A A T C C C
AA_AM AD_066 0 0 0 -9 G A T T A A T T A G A G G G A A T T A G G G C T T T G G C C C C A A T C C C
AA_AM AD_076 0 0 0 -9 G A T T A A T C A G A G G G A A C T A G G G C T T T G G C C T C C A T T C C
AA_AM AD_500 0 0 0 -9 G A G T A A T C G G A G G G A A T T A G G G C T T T G G C C C C A A C C C C
AA_AM AD_505 0 0 0 -9 G A G T A A T C A G A A G G A A T T A G A G C T T T G G C C C C C A T T C C
AA_AM AD_510 0 0 0 -9 G A T T A A C C G G A G G G A A T T A G A G C T T T G G C C C C A A T C C C
AA_AM AD_511 0 0 0 -9 G A T T A A C C G G A G G G A A T T A G A G C T T T G G C C C C A A T C C C
AA_AM AD_512 0 0 0 -9 A A T T G A C C G G A G G G A A C T G G G G T T T T G G C C C C A A C C C C
AA_AM AD_523 0 0 0 -9 G A T T G A T C A G A G G G A A T T G G G G C C T T G G C C C C A A T T C C
Abkhasian_SC abh24 0 0 0 -9 G A T T A A C C G G G G G G A A C T G G G G T T T T G G C C C C A A C C
Abkhasian_SC abh27 0 0 0 -9 G A G T A A T C G G A G G G A A T T G G G G C T T T G G C C C C A A T C
Abkhasian_SC abh41 0 0 0 -9 A A T T G A C C G G A A G G A A C T G G A G C T T T G G C C C C A A C C
Abkhasian_SC abh107 0 0 0 -9 G A T T A A C C G G G G G G A A T T A G A G C C T T G G C C C C A A T C
Abkhasian_SC abh119 0 0 0 -9 A A T T A A T C A G A A G G A A C T A G A G T T T T G G C C C C A A C C
Abkhasian_SC abh122 0 0 0 -9 A A G G A A T C A G A A G G G A C C G G G G C T T T G G C C C C A A C C
Abkhasian_SC abh133 0 0 0 -9 A A T T A A C C G G A G G G G A C T G G G G T T T T G G C C C C A A C C
Abkhasian_SC abh147 0 0 0 -9 A A T T A A C C G G G G G G A A T T G G G G T T T T G G C C C C A A C C
Abkhasian_SC abh154 0 0 0 -9 G A G G A A T T G G A G G G A A T T A G G G C T T T G G C C C C A A C C
Adygei_NC HGDP01381 0 0 0 -9 A A T T G A C C A G A G G G A A C T A G A G T T T T G G C C C C A A
Adygei_NC HGDP01382 0 0 0 -9 A A G T A A C C G G G G G G A A T T G G G G T T C C A A T T C C A A
Adygei_NC HGDP01383 0 0 0 -9 A A G T A A T C G G G G G G A A T T G G G G T T T T G G C C C C A A
Adygei_NC HGDP01385 0 0 0 -9 A A T T A A C C G G A G G G A A C T G G G G C T T T G G C C C C A A
Adygei_NC HGDP01386 0 0 0 -9 G A G G A A T T A G G G G G A A T T A G G G C T T T G G C C C C A A
Adygei_NC HGDP01387 0 0 0 -9 G A G T G A T C G G A G G G A A T T G G G G C T T T G G C C C C A A
Adygei_NC HGDP01396 0 0 0 -9 A A T T G A T C G G A G G G A A T T A G G G C T T T G G C C C C A A
Adygei_NC HGDP01397 0 0 0 -9 A A T T A A C C G G A A G G G A C C A G G G C T C T A G T C C C A A
Adygei_NC HGDP01398 0 0 0 -9 A A G T G A T C G G A G G G A A C C G G G G T T T T G G C C C C A A
Adygei_NC HGDP01399 0 0 0 -9 A A G T A A C C G G G G G G G A C T G G G G T T T T G G C C C C A A
Adygei_NC HGDP01400 0 0 0 -9 A A T T A A C C G G A G G G A A C T G G G G C T C T A G T C C C A A
Adygei_NC HGDP01401 0 0 0 -9 A A T T G A C C G G A A G G A A T T A G G G C C T T G G C C C C A A
Adygei_NC HGDP01402 0 0 0 -9 A A G G G A T T G G A G G G A A C T G G G G C T T T G G C C C C A A
Adygei_NC HGDP01403 0 0 0 -9 G A G T A A T C G G A G G G A A C T G G G G T T T T G G 0 0 C C A A
Adygei_NC HGDP01404 0 0 0 -9 A A T T G A T C G G A G G G G A C T G G G G C T T T G G C C C C A A
Adygei_NC NA13619 0 0 0 -9 A A T T A A C C A G A G G G A A T T G G G G T T C T A G T C C C A A C C
Adygei_NC NA13626 0 0 0 -9 A A T T G A T C A G A A G G A A C T A G A G T T T T G G C C C C A A C C
Ajv58_ancient Ajv58_md 0 0 0 -9 A A T T 0 0 0 0 G G A A G G A A T T G G 0 0 T T T T G G C C C C A A
```

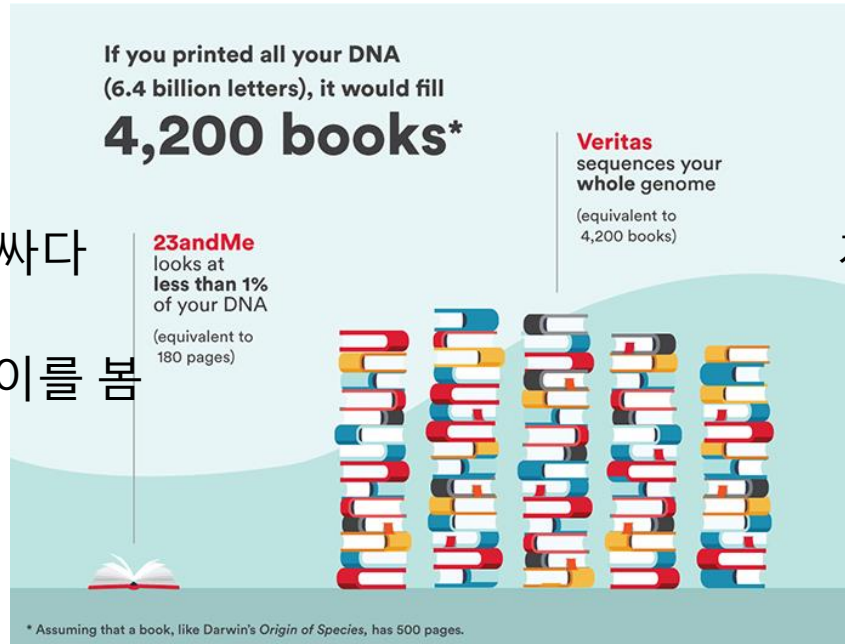
변이 정보

DNA Chip vs Whole genome sequencing (WGS)

DNA chip

가격이 상대적으로 싸다

대체로 Common 변이를 봄



WGS

게놈전체에 대한 변이 파악가능

개인 특이적인 변이도 발견가능

And this matters because having your **whole genome** sequenced means:

1. More Useful Info

90%+ of relevant DNA is distributed across your genome

2. More Actionable Insights

Make better health & lifestyle decisions with clinical-grade results

3. A Resource for Life

Sequence your genome once and learn more and more as science progresses

유전체기술로 심혈관질환 연구

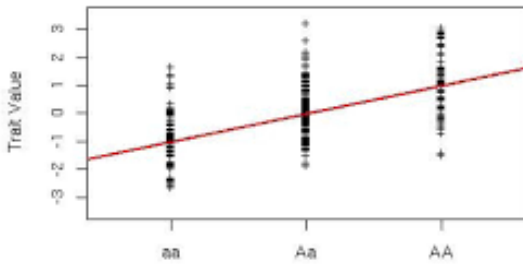
예측?

진단?

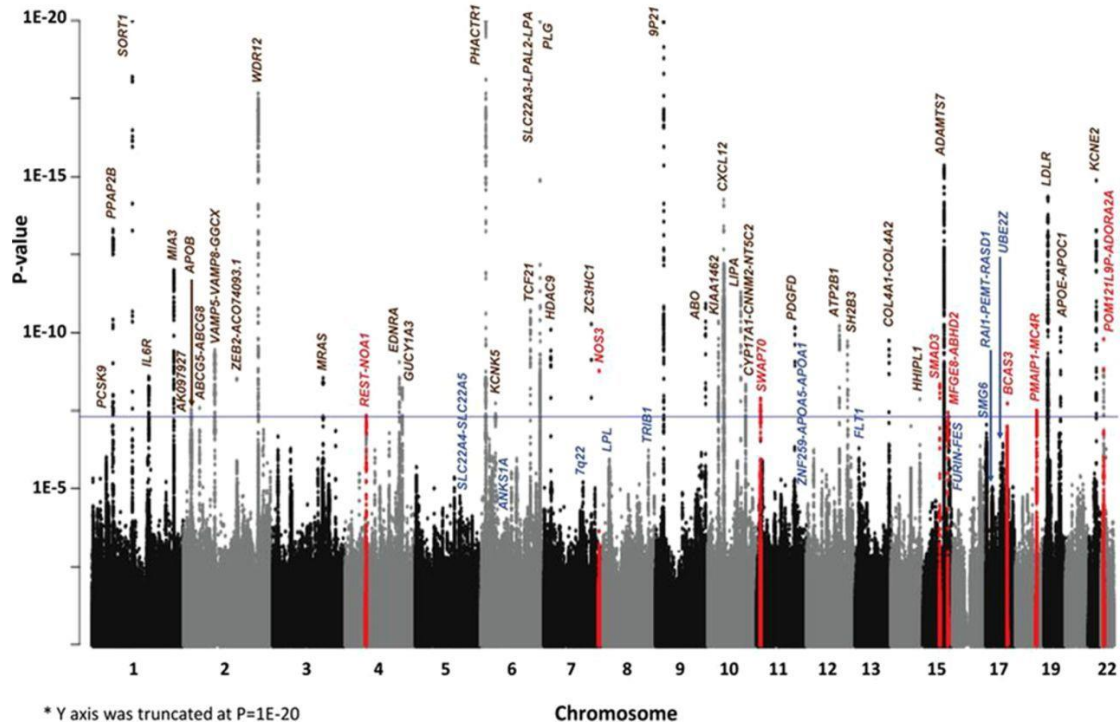
Genome wide association study (GWAS)

Unrelated Samples

$$\hat{y}_i = \mu + \hat{\beta} x_i$$



Manhattan plot of CAD additive meta-analysis results



관련있는 변이 발견

CVD GWAS catalog variants

Refine search results ^

P Publications

18

T Traits

199

Catalog stats

- Last data release on 2019-11-21
- 4298 publications
- 118283 SNPs
- 161525 associations
- Genome assembly GRCh38.p13
- dbSNP Build 152
- Ensembl Build 96

Search results for *cardiovascular disease*



GWAS Catalog

The NHGRI-EBI Catalog of published genome-wide association studies

cardiovascular disease

Examples: breast carcinoma, rs7329174, Yao, 2q37.1, HBS1L, 6:16000000-25000000

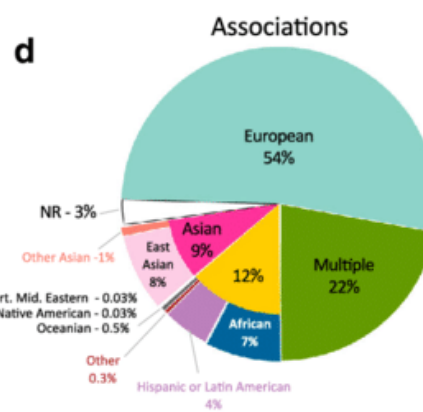
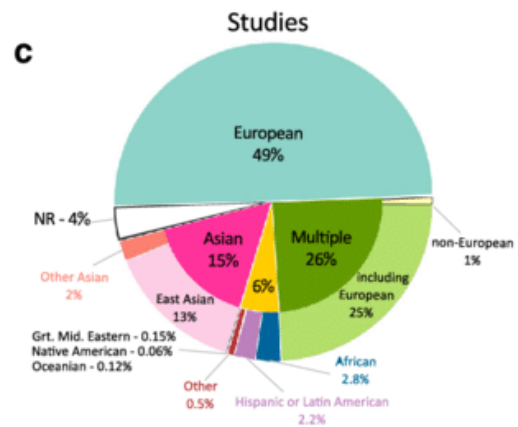
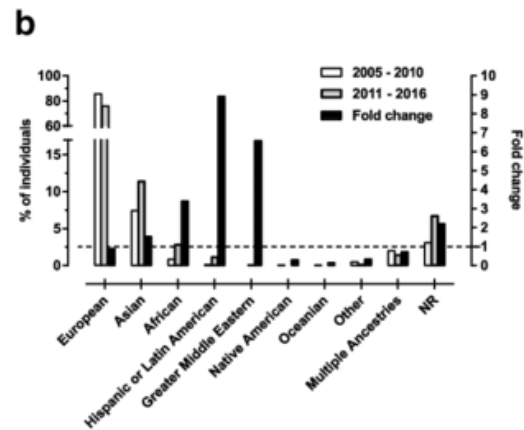
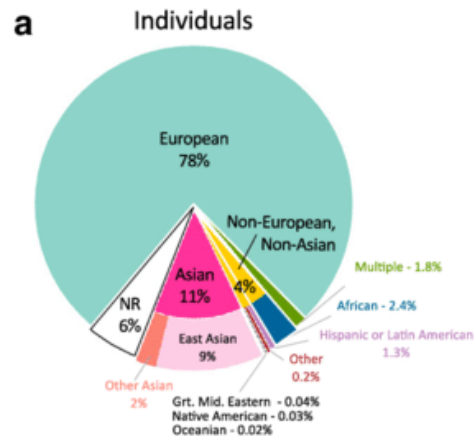
Cardiovascular disease biomarker measureme

cardiovascular disease biomarkers, such as ST2 cardiac biomarker and C-reactive protein, are used for cardiovascular disease and as predictors for therapeutic responses

Associations **6384** Studies **338**

4,298 publications, 118,282 SNPs from GWAS catalog

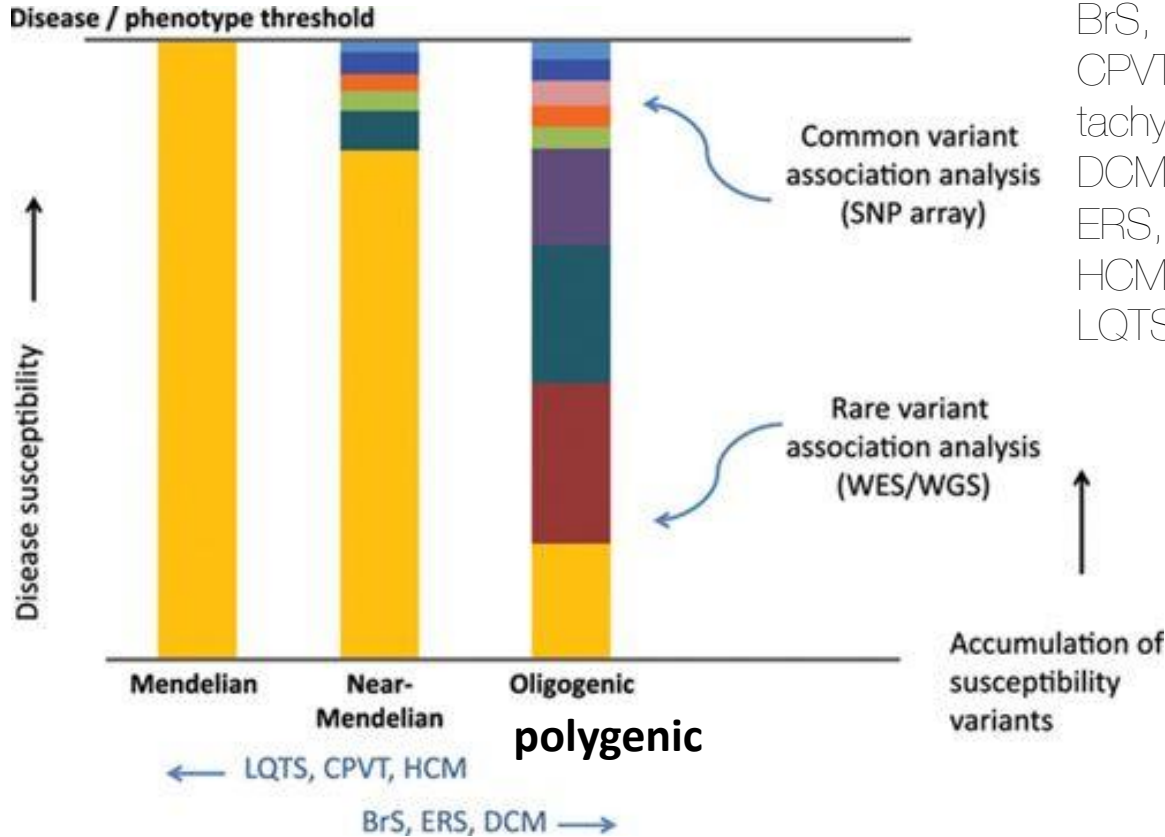
Skewed population distribution



한국인, 아시아인 연구의 부족

Monogenic vs polygenic

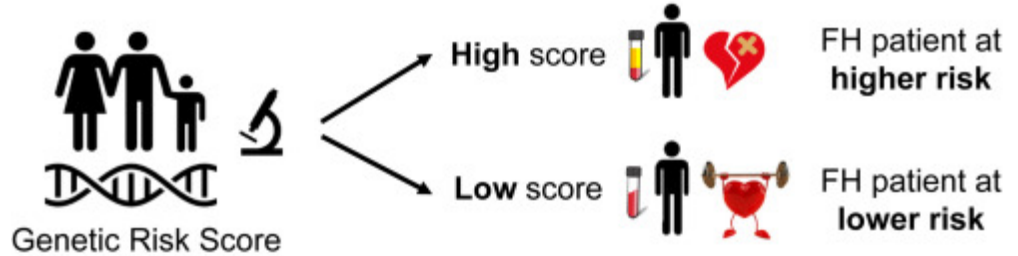
Bezzina, Connie R., Najim Lahrouchi, and Silvia G. Priori. "Genetics of sudden cardiac death." *Circulation research* 116.12 (2015): 1919-1936.



BrS, Brugada syndrome
 CPVT, catecholaminergic polymorphic ventricular tachycardia
 DCM, dilated cardiomyopathy
 ERS, early repolarization syndrome
 HCM, hypertrophic cardiomyopathy
 LQTS, long-QT syndrome

Polygenic (genetic) risk score

$$\hat{S} = \sum_{j=1}^m X_j \hat{\beta}_j$$



Polygenic factors can influence the FH phenotype

CLINICAL APPLICATION OF PRS

A polygenic risk score (PRS) is calculated from many small genetic variants, and can often be modified by lifestyle factors.

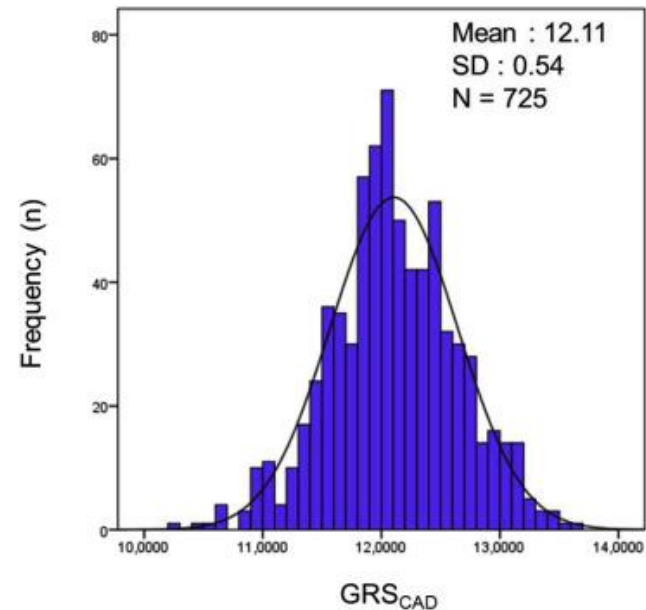
<https://www.nature.com/articles/d42473-019-00270-w>

Some gene variants confer very high risk of getting a particular disease - but these are rare.

Some gene variants confer very small risk of getting a particular disease - these are common.

PRS uses the sum of all known common variants to calculate an overall risk of getting a particular disease.

PRS combined with lifestyle and clinical factors which modify this risk, can inform treatment decisions and the need to intervene.



AI for CVD Risk prediction

RESEARCH ARTICLE

Cardiovascular disease risk prediction using automated machine learning: A prospective study of 423,604 UK Biobank participants

Ahmed M. Alaa^{1*}, Thomas Bolton^{2,3}, Emanuele Di Angelantonio^{2,3}, James H. F. Rudd⁴, Mihaela van der Schaar^{1,5,6}

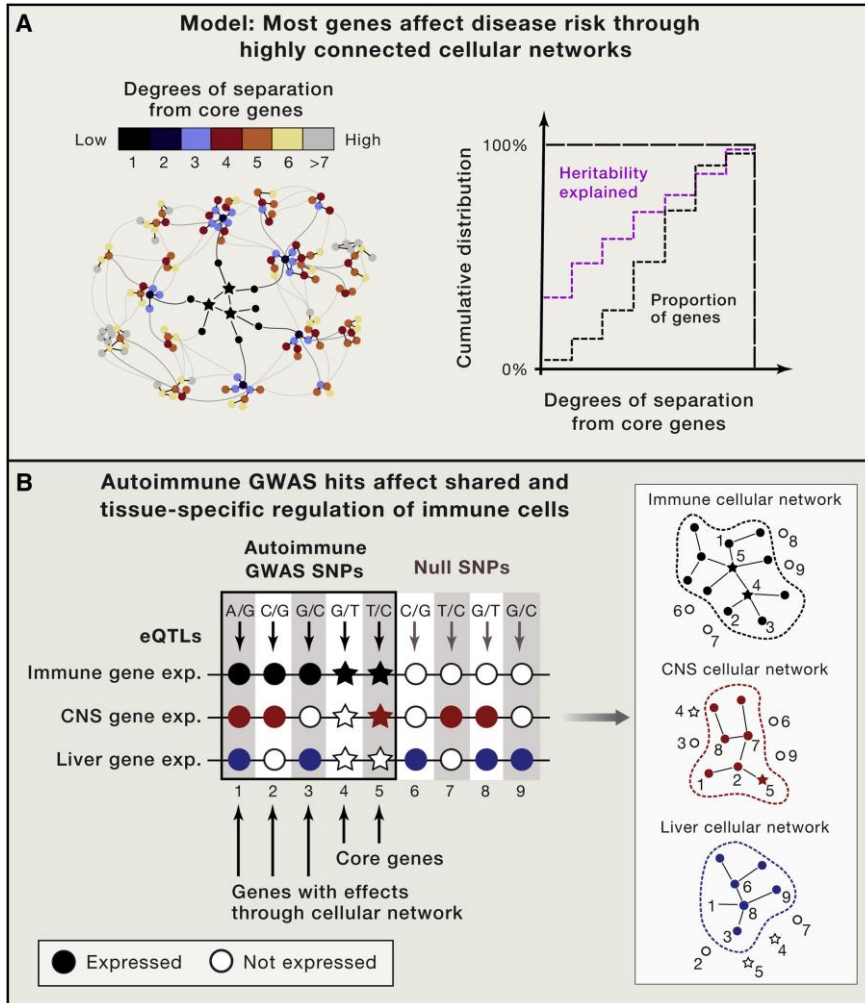
1 University of California Los Angeles, Los Angeles, California, United States of America, 2 Department of Public Health and Primary Care, University of Cambridge, Cambridge, United Kingdom, 3 National Institute for Health Research (NIHR) Blood and Transplant Research Unit (BTRU) in Donor Health and Genomics, University of Cambridge, Cambridge, United Kingdom, 4 Department of Cardiovascular Medicine, University of Cambridge and Cambridge University Hospitals NHS Foundation Trust, Cambridge, United Kingdom, 5 University of Oxford, Oxford, United Kingdom, 6 Alan Turing Institute, London, United Kingdom

* ahmedmalaa@ucla.edu

Model	AUC-ROC	Absolute AUC-ROC Change
Framingham Score	0.724 ± 0.004	Baseline model
Cox PH Model (7 core variables)	0.734 ± 0.005	+ 1.0%
Cox PH Model (all variables)	0.758 ± 0.005	+ 3.4%
Support Vector Machines	0.709 ± 0.061	- 1.5%
Random Forest	0.730 ± 0.004	+ 0.6%
Neural Networks	0.755 ± 0.005	+ 3.1%
AdaBoost	0.759 ± 0.004	+ 3.5%
Gradient Boosting	0.769 ± 0.005	+ 4.5%
AutoPrognosis (7 core variables)	0.744 ± 0.005	+ 2.0%
AutoPrognosis (369 non-lab. variables)	0.761 ± 0.005	+ 3.7%
AutoPrognosis (104 lab. variables)	0.735 ± 0.008	+ 1.1%
AutoPrognosis (all variables)	0.774 ± 0.005	+ 5.0%

Omnigenic model

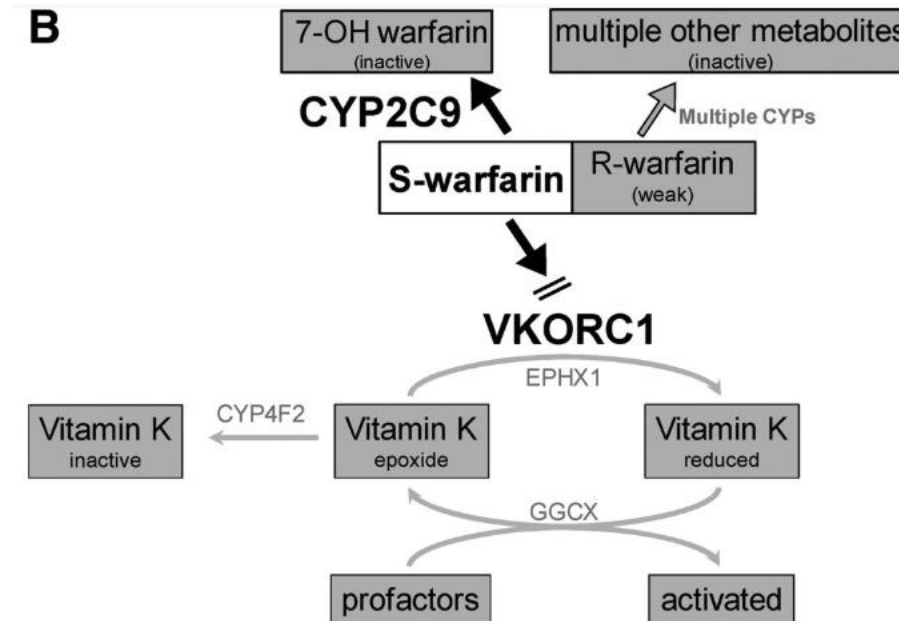
Boyle, Evan A., Yang I. Li, and Jonathan K. Pritchard. "An expanded view of complex traits: from polygenic to omnigenic." *Cell* 169.7 (2017): 1177-1186.



We propose that [gene regulatory networks](#) are sufficiently interconnected such that all genes expressed in disease-relevant cells are liable to affect the functions of core disease-related genes and that most [heritability](#) can be explained by effects on genes outside core pathways.

무한히 많은 작은 효과들

Pharmacogenomics : Warfarin

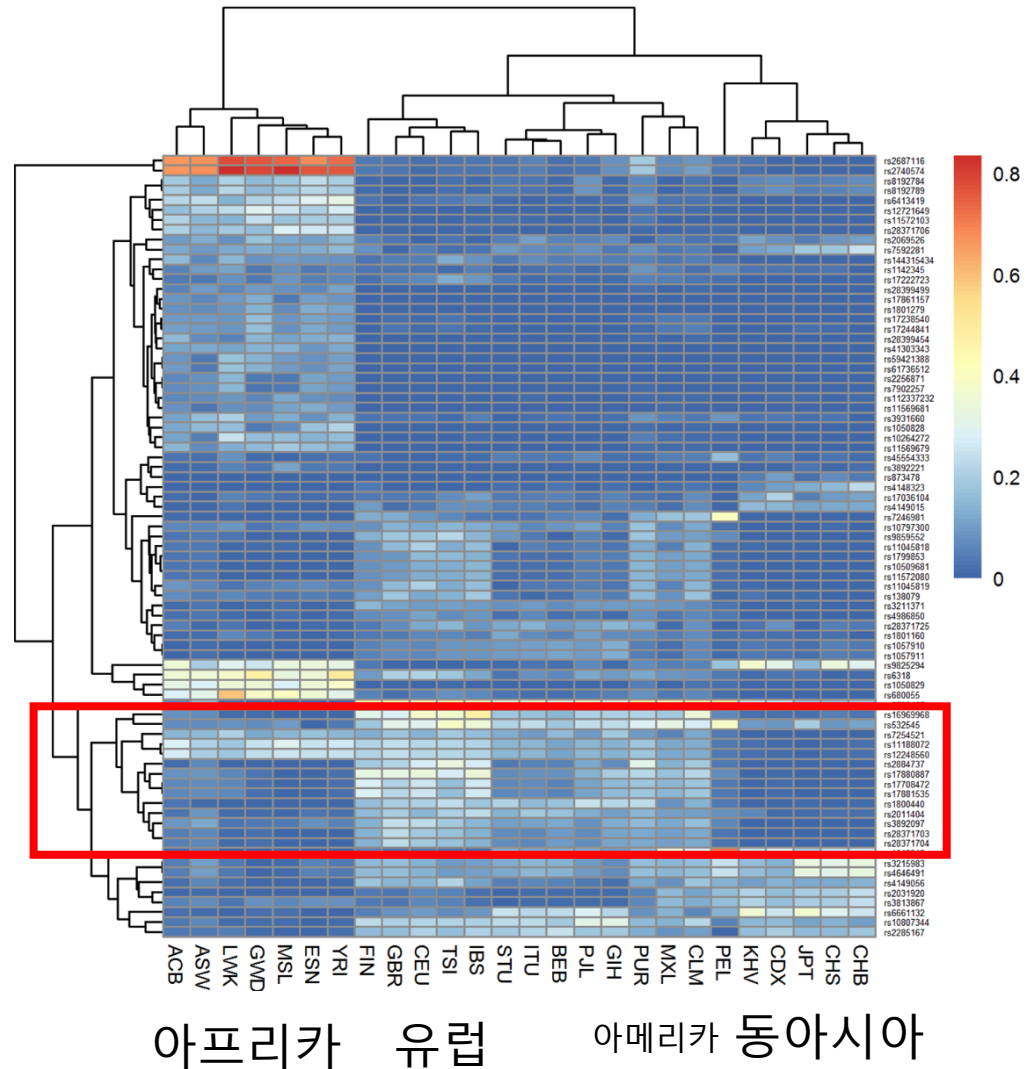


variants in CYP2C9 and VKORC1 contribute to variable warfarin dosage.

Roden, Dan M., et al. "Cardiovascular pharmacogenomics." *Circulation research* 109.7 (2011): 807-820.

Pharmacogenomics

인종마다 변이의 빈도가 차이남



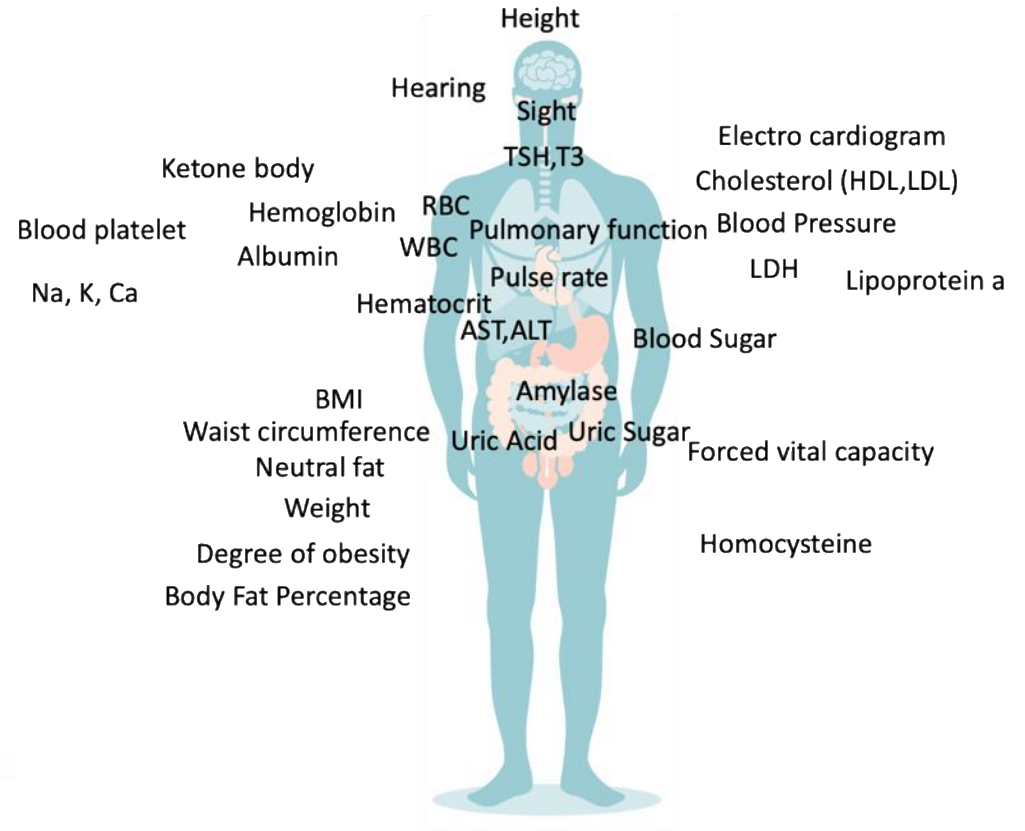
Minor Allele Frequency heatmap for 79 Pharmacogenomics variants that are common ($MAF > 0.1$) and rare ($MAF < 0.005$) in at least one 1KG population.

Korean genome database

울산 만명게놈 프로젝트



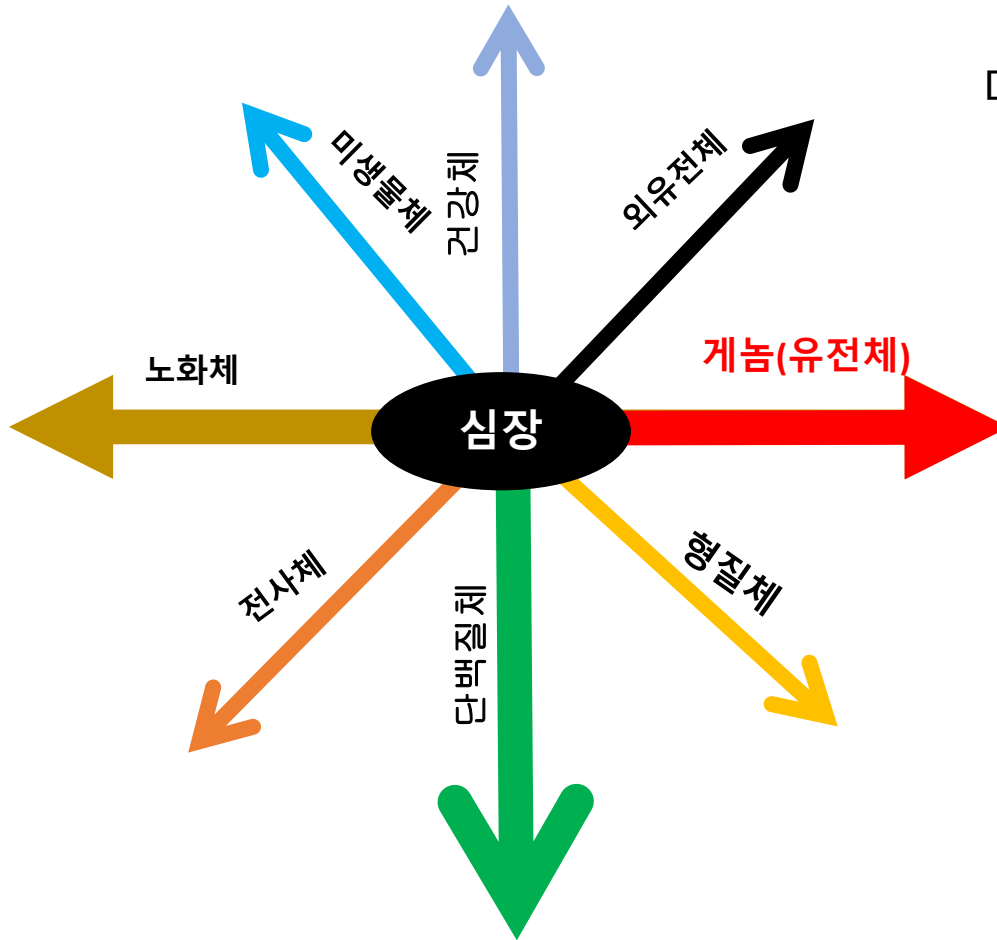
국민 게놈



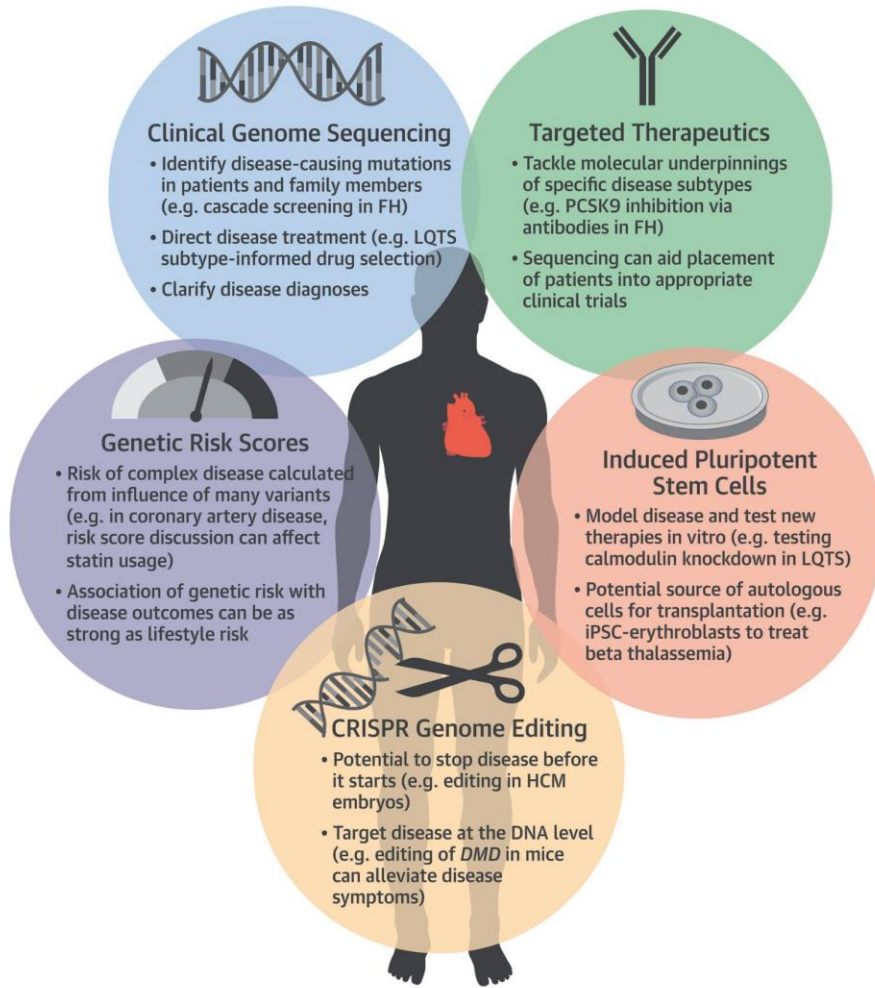
건강 검진 정보

Multi-omics

다중 오믹스를 통한 심장게놈 연구



Precision/Personalized medicine



개인 게놈 정보를
활용한
정밀/맞춤의학

State-of-the-art genetic technologies are revolutionizing cardiovascular precision medicine.

Dainis, Alexandra M., and Euan A. Ashley. "Cardiovascular precision medicine in the genomics era." *JACC: Basic to Translational Science* 3.2 (2018): 313-326.