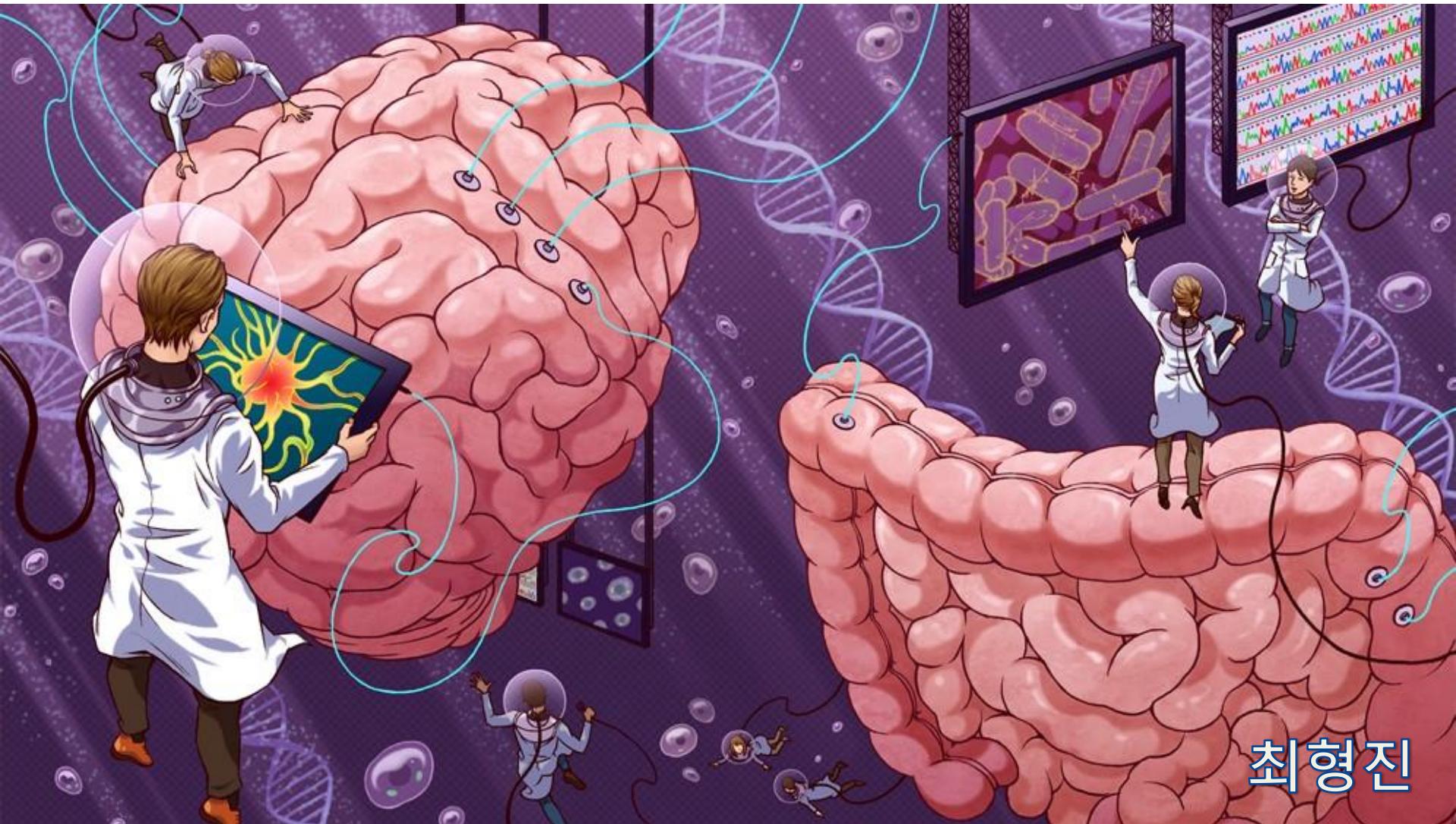


헬스케어 빅데이터로 무엇을 할 수 있는가?

연구와 임상 적용 경험을 중심으로

(유전체, 전자의무기록, 건강보험청구자료, 영상, 센서/모바일)



최형진

Contents

1. What is Healthcare Big Data?
2. Healthcare Big Data

- ① Genetic Data
- ② Electrical Health Records
- ③ National Healthcare Data
- ④ Medical Images
- ⑤ Sensor/Mobile Data

3. Clinical and Research Applications



ICE/ENDO 2014

JUNE 21-24



CHICAGO

Navigating Biomedical Big Data

미국 내분비 학회

Program: Symposia
Translational Session

Tuesday, June 24, 2014: 11:15 AM-12:45 PM

W178 (McCormick Place West Building)

Chair:

M Susan Smith, MS, PhD, OR National Primate Rsrch Ctr, Oregon Health & Science University, Beaverton, OR

11:15 AM S77-1

Big Data from Small Data: Linking & Accessing Biomedical Big Data
Maryann E Martone, PhD, UCSD

11:45 AM S77-2

Making Sense of Genomic Data to Advance Genomic Medicine
Alan R Shuldiner, MD, Department of Medicine, Univ of Maryland, Baltimore, MD

12:15 PM S77-3

Linking and Accessing Big Data for Translation to Clinical Medicine: The Cancer Genome Atlas
Thomas J Giordano, MD, PhD, Pathology, Internal Medicine, University of Michigan, Ann Arbor, MI

미국 골대사

ASBMR

Using
Osteo

Claims
EHR
Registry/Cohort
Trial/LTE
Sensor
SES/Zipcode/Geography
PROC(smartphones)
Genetic
Biospecimen



대한내분비학회



2014 대한내분비학회 학연산 및 추계심포지엄

- 일정 : 2014년 10월 30일(목) ~ 11월 1일(토)
- 장소 : 광주 김대중컨벤션센터

학연산심포지엄 III : Information Communication Technology for endocrinology clinic

10월 31일(금) 15:00-16:30



Big Data in Healthcare: 분산연구망을 이용한 국제공동연구

아주의대 의료정보학과 박래웅



인지 컴퓨팅을 활용한 의료혁신-IBM Watson for Healthcare

IBM Korea 배영우



Future of Medical Devices and Clinical Application for Diabetes and Endocrinology Diseases

Girish Srinivasan, Samsung Electronics, Korea

IBM

SAMSUNG

대한의료정보학회 추계학술대회 & 한중일 국제 의료정보학회

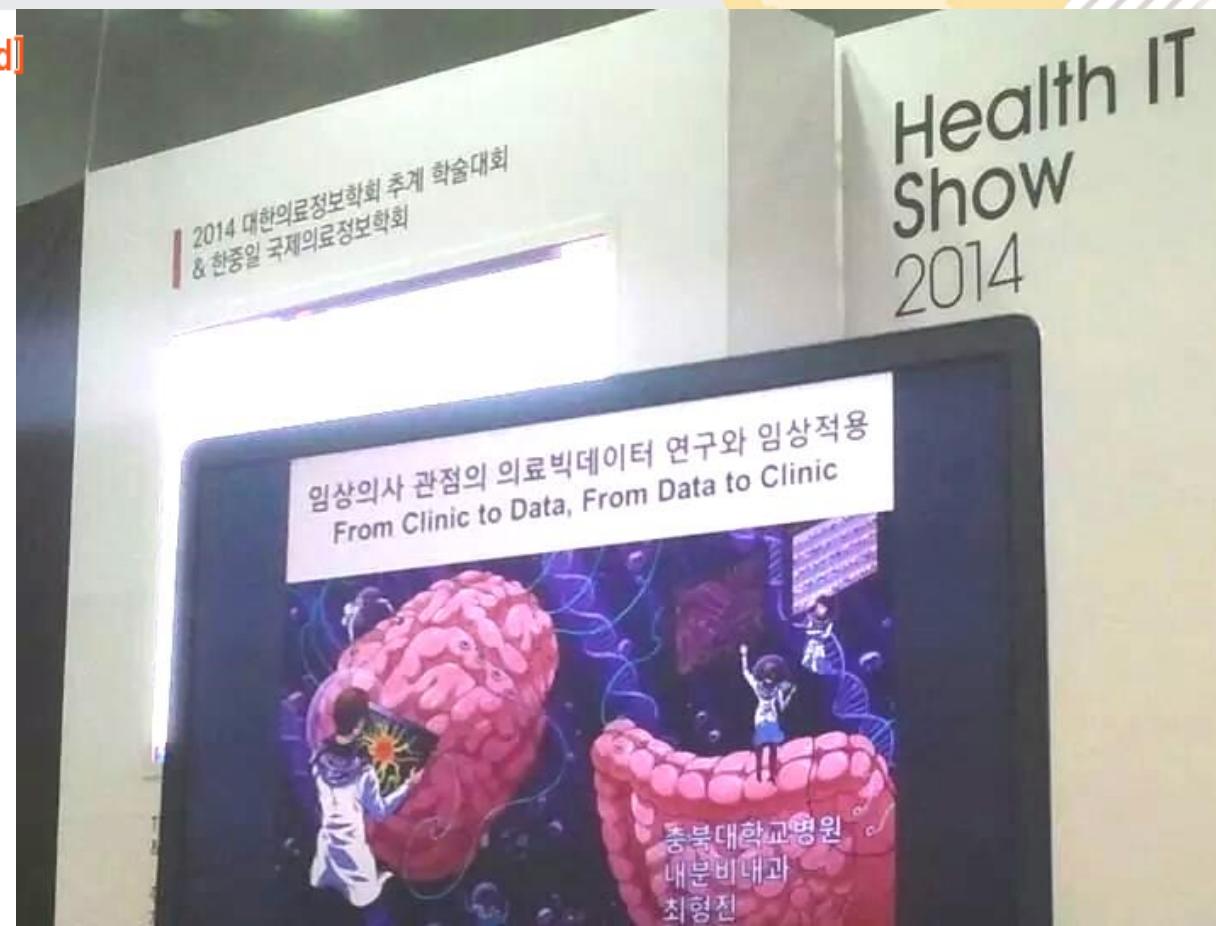
KOSMI 2014 & CJKMI 2014

HEALTHCARE IT PLATFORMS FOR ONE HEALTH



2014. 10. 27. [Mon] ~ 29. [Wed]

▶ 부산 벡스코 제2전시장 회의실



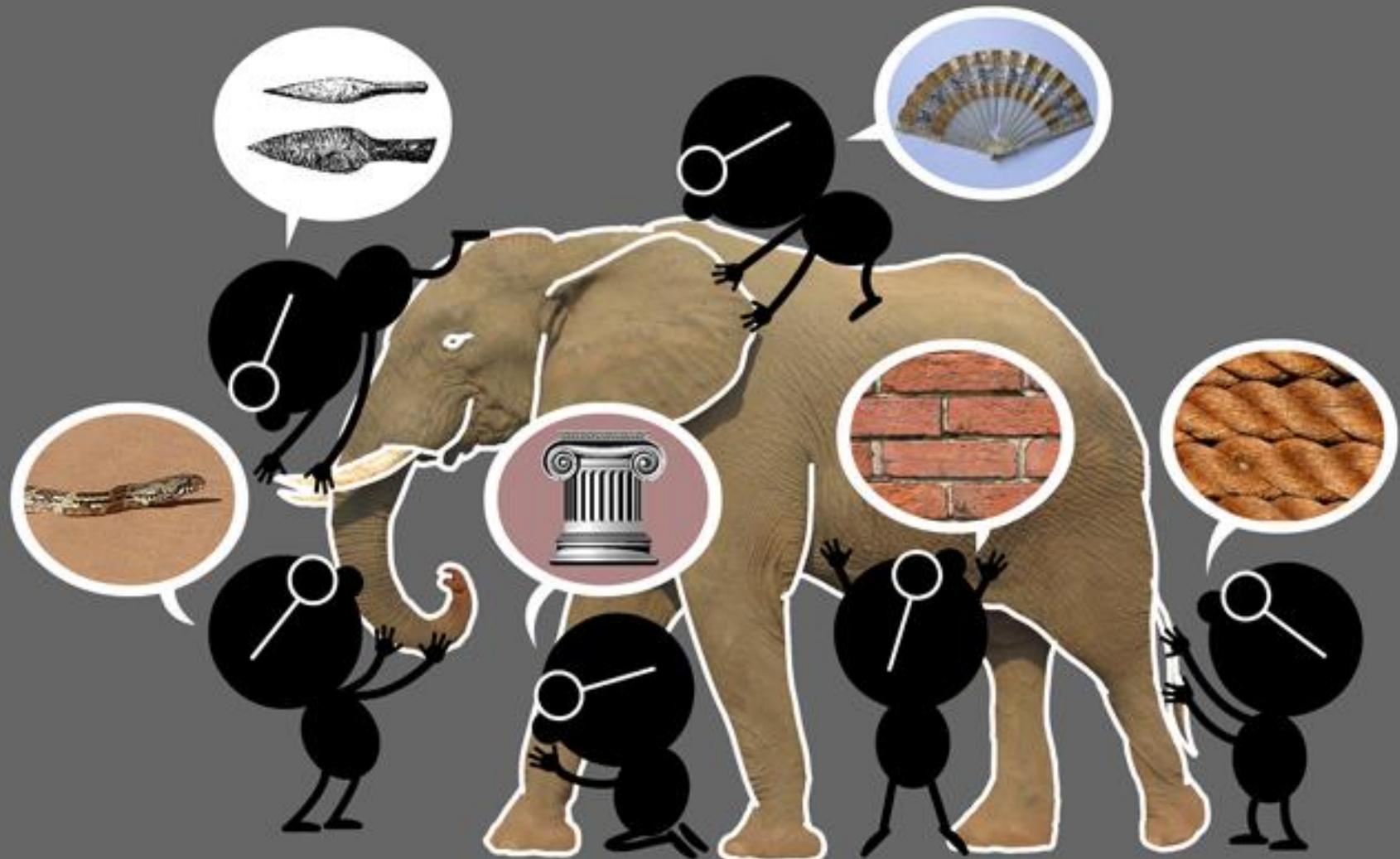
2014 대한골대사학회 제26차 추계학술대회

- 일시 : 2014년 11월 15일(토)
- 장소 : 가톨릭대학교 성의회관 마리아홀 1002호

15:40~16:05

Biomedical big data and osteoporosis research
최형진 (충북의대 내분비내과)

What is Big Data?



**Big data is like teenage sex:
everyone talks about it,
nobody really knows how to do it,
everyone thinks everyone else is
doing it, so everyone claims they
are doing it...**

(Dan Ariely)

Big Data Dimensions

Volume



Data at scale

Terabytes to petabytes of data

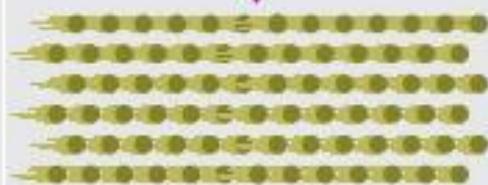
Variety



Data in many forms

Structured, unstructured, text, multimedia

Velocity



Data in motion

Analysis of streaming data to enable decisions within fractions of a second

Veracity



Data uncertainty

Managing the reliability and predictability of inherently imprecise data types

New Analysis Tool (R, Python, Matlab)
New Approach

Big Meal?



Volume



Variety



Velocity



Potential Use Cases for Big Data Analytics

Real time

Credit & Market Risk in Banks

(중환자실 심전도)

Fraud Detection (Credit Card) & Financial Crimes (AML) in Banks
(including Social Network Analysis)

Event-based Marketing in Financial Services and Telecoms

Markdown Optimization in Retail

Claims and Tax Fraud in Public Sector

Predictive Maintenance in Aerospace

Social Media Sentiment Analysis

Demand Forecasting in Manufacturing

Disease Analysis on Electronic Health Records

Traditional Data Warehousing

Text Mining

Video Surveillance/Analysis

Data Velocity

Batch

Structured

Semi-structured

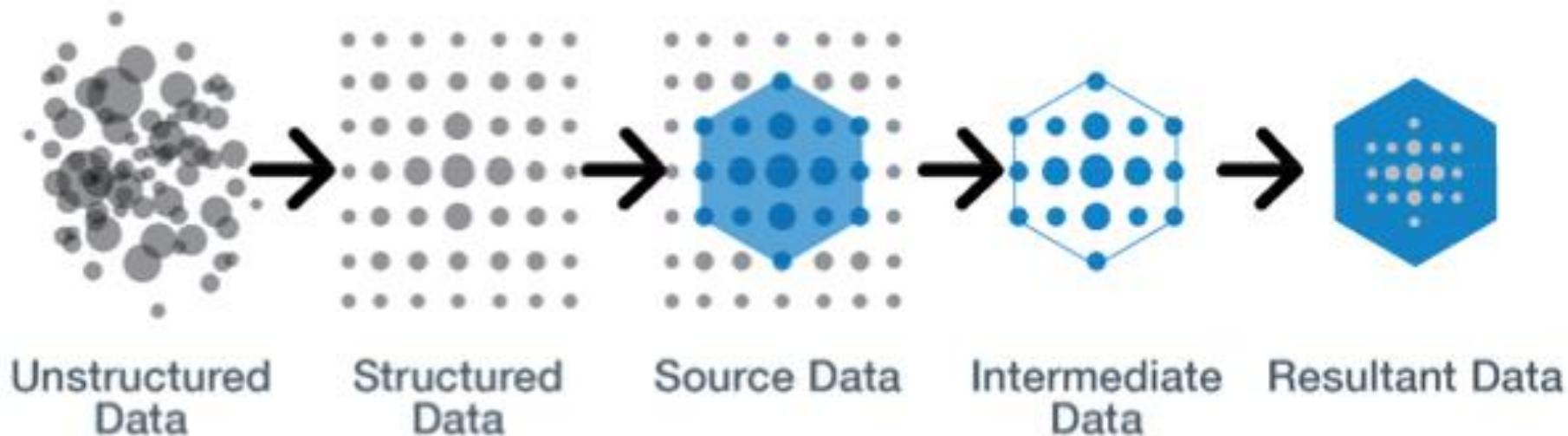
Unstructured

Data Variety

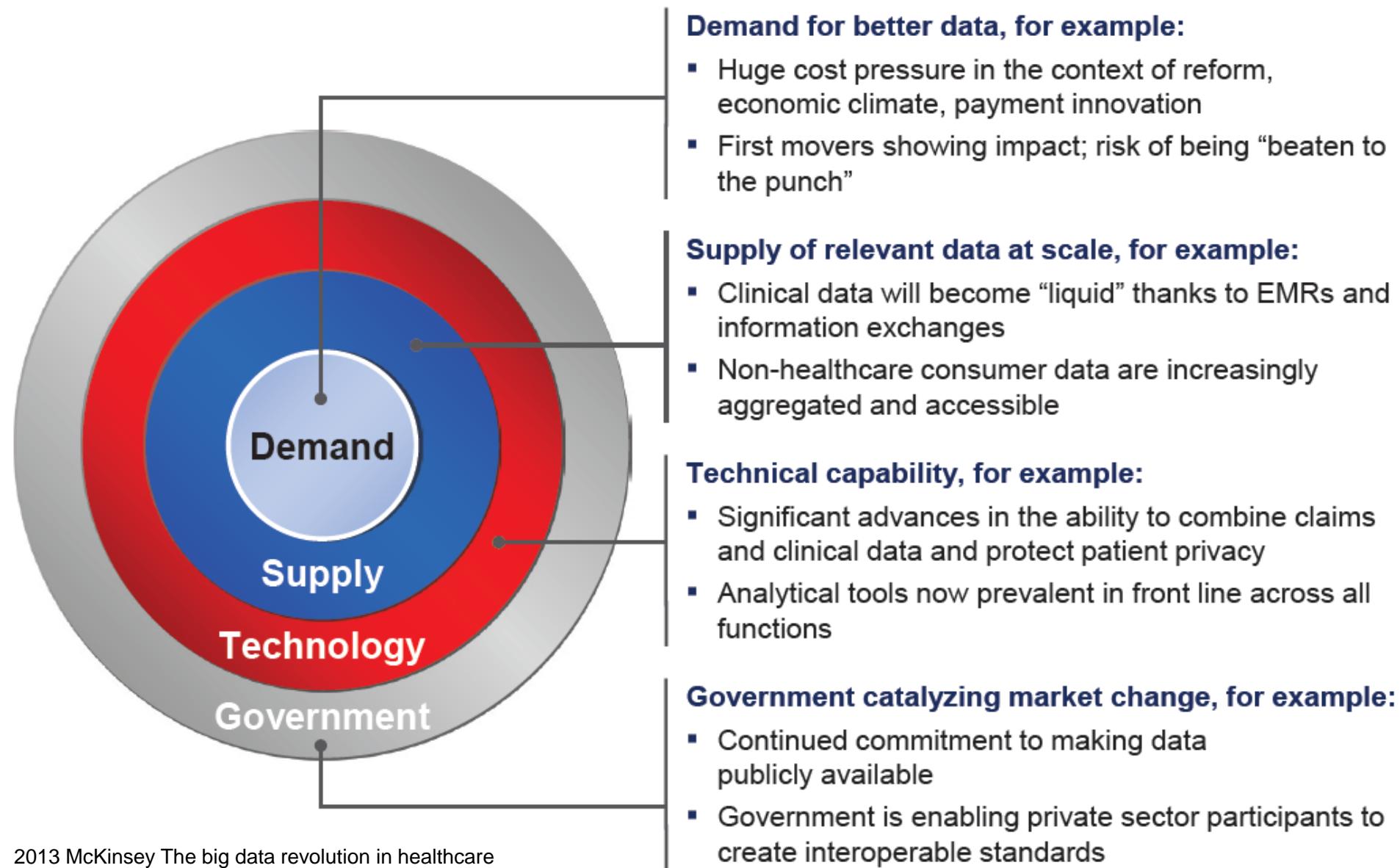
출처 : SAS and IDC

Data Variety

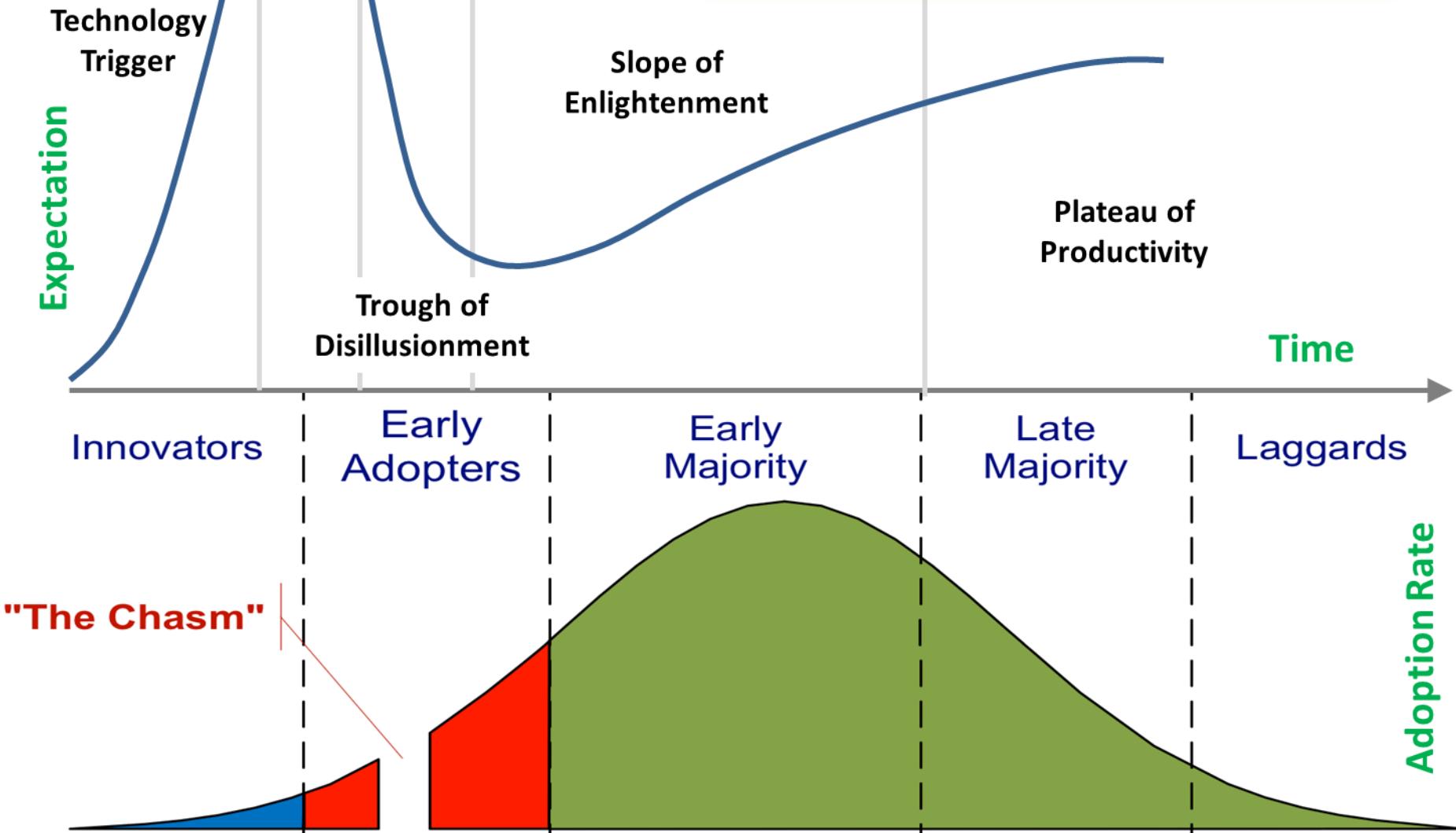
UNSTRUCTURED DATA TO RESULTS

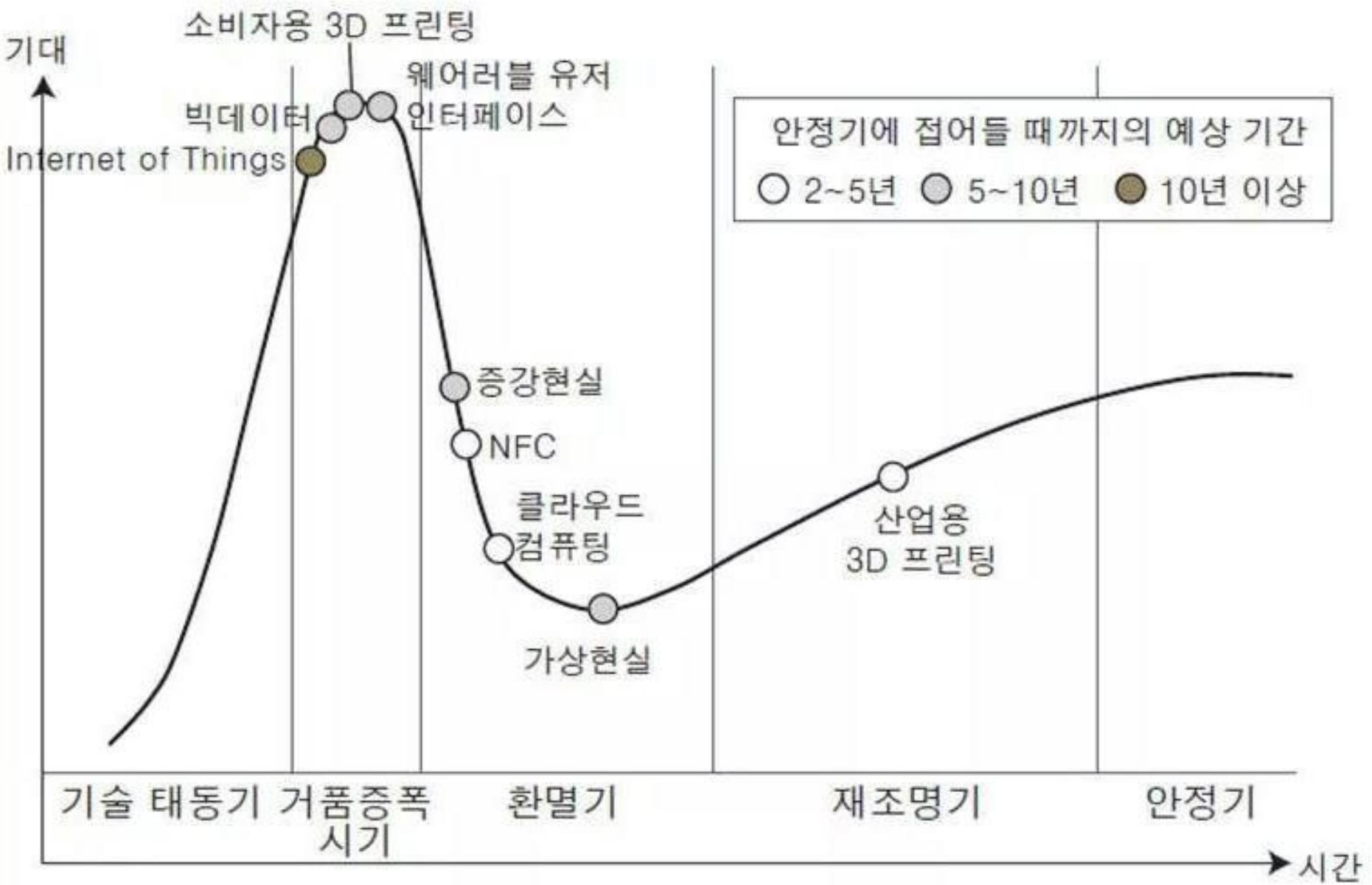


Tipping Point for Big Data Healthcare

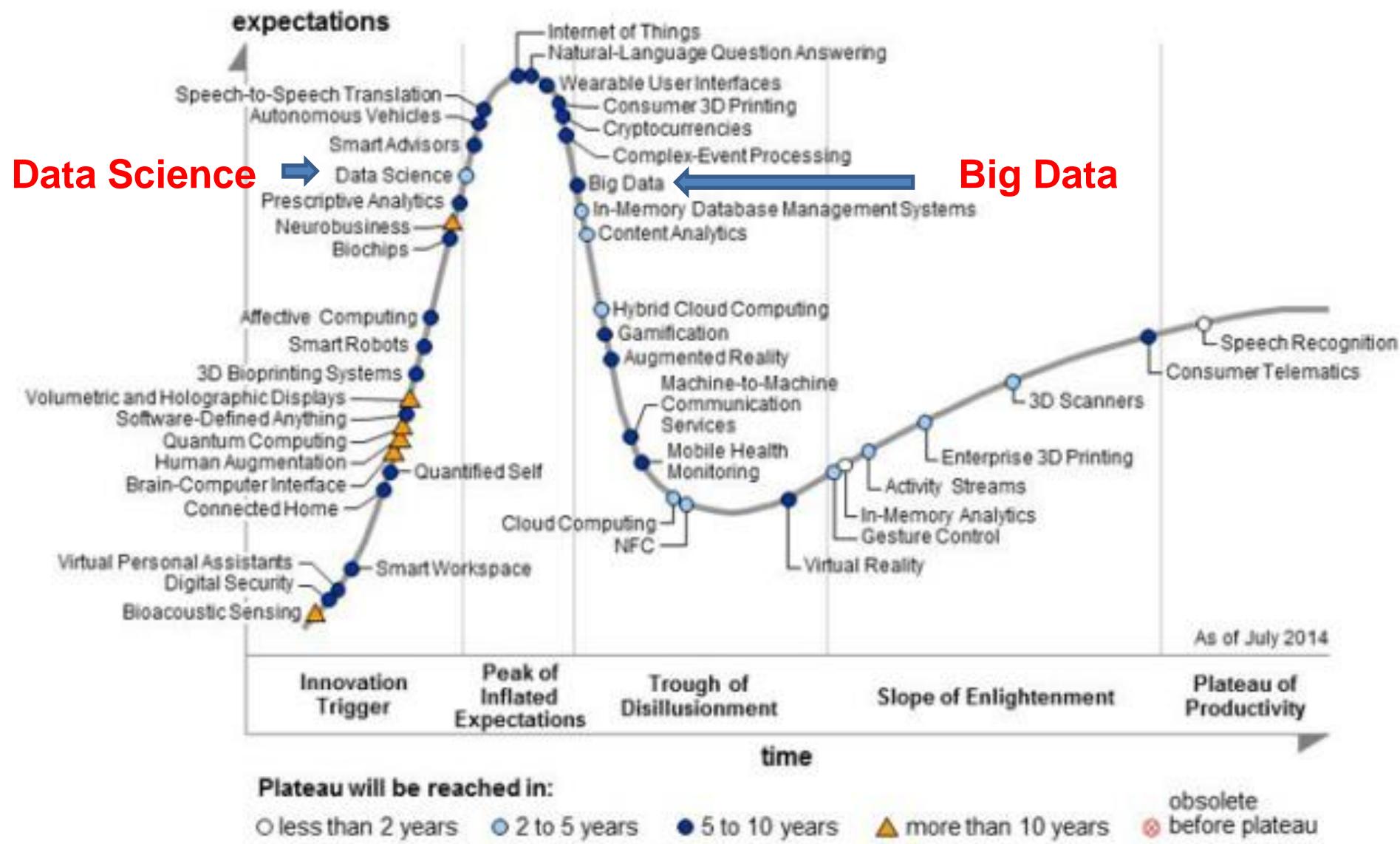


Hype Cycle and Technology Adoption Lifecycle Plotted together

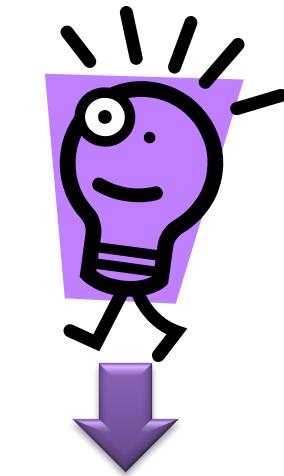




자료 : 가트너(2013년 7월)



Hypothesis Driven Science Data Driven Science



Hypothesis



Collect
Data



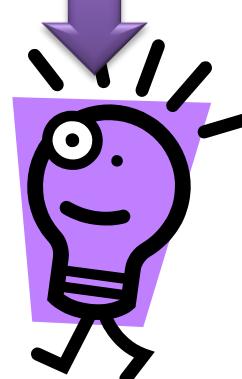
Analyze



Data

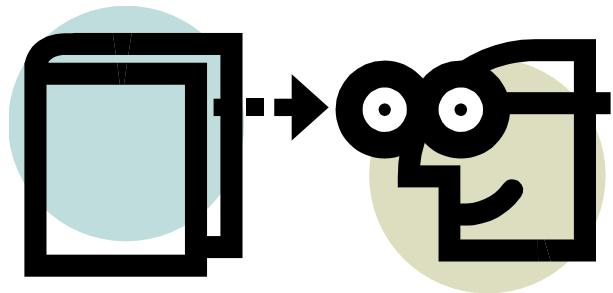


Analyze

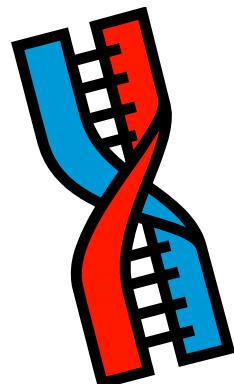


Generate
Hypothesis

Candidate Gene Approach

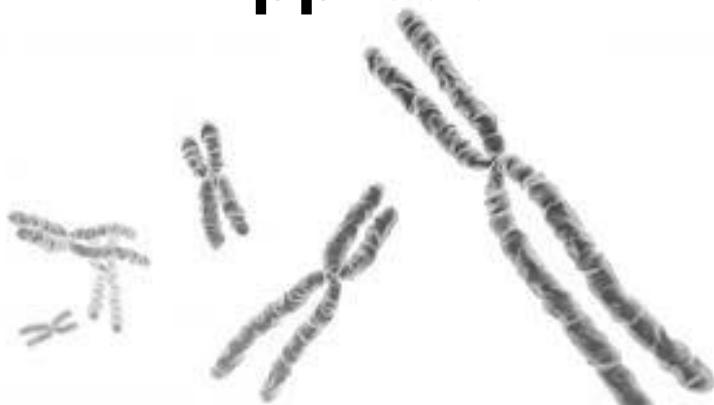


Choose a Gene
from Prior Knowledge

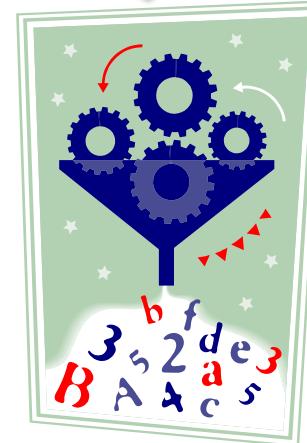


Analyze the Gene

Genome-wide Approach



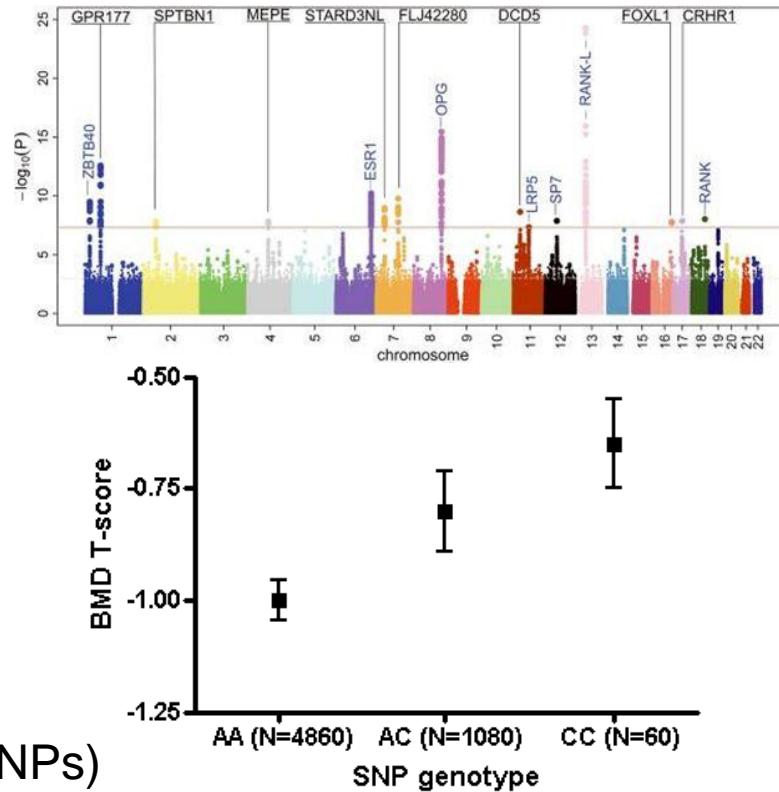
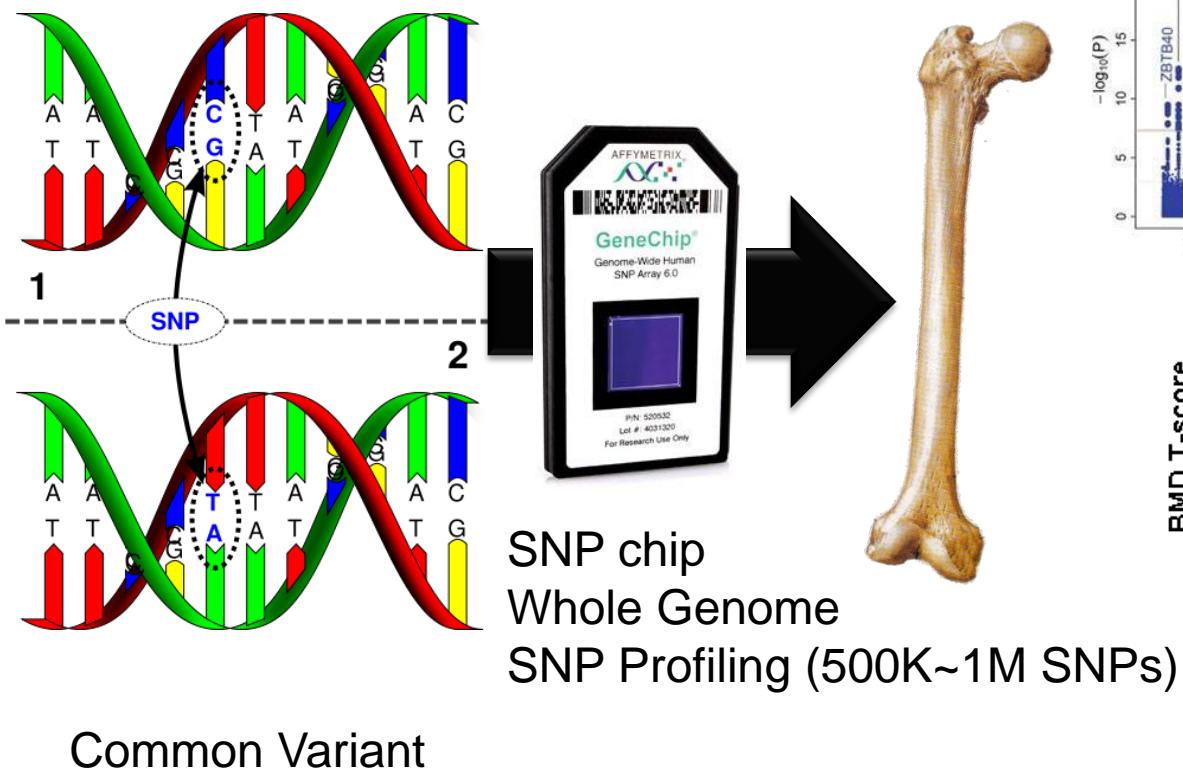
Analyze ALL Genes



Discover Novel Findings

GWAS

(Genome wide association study)



Recent largest GWAS GEFOS consortium

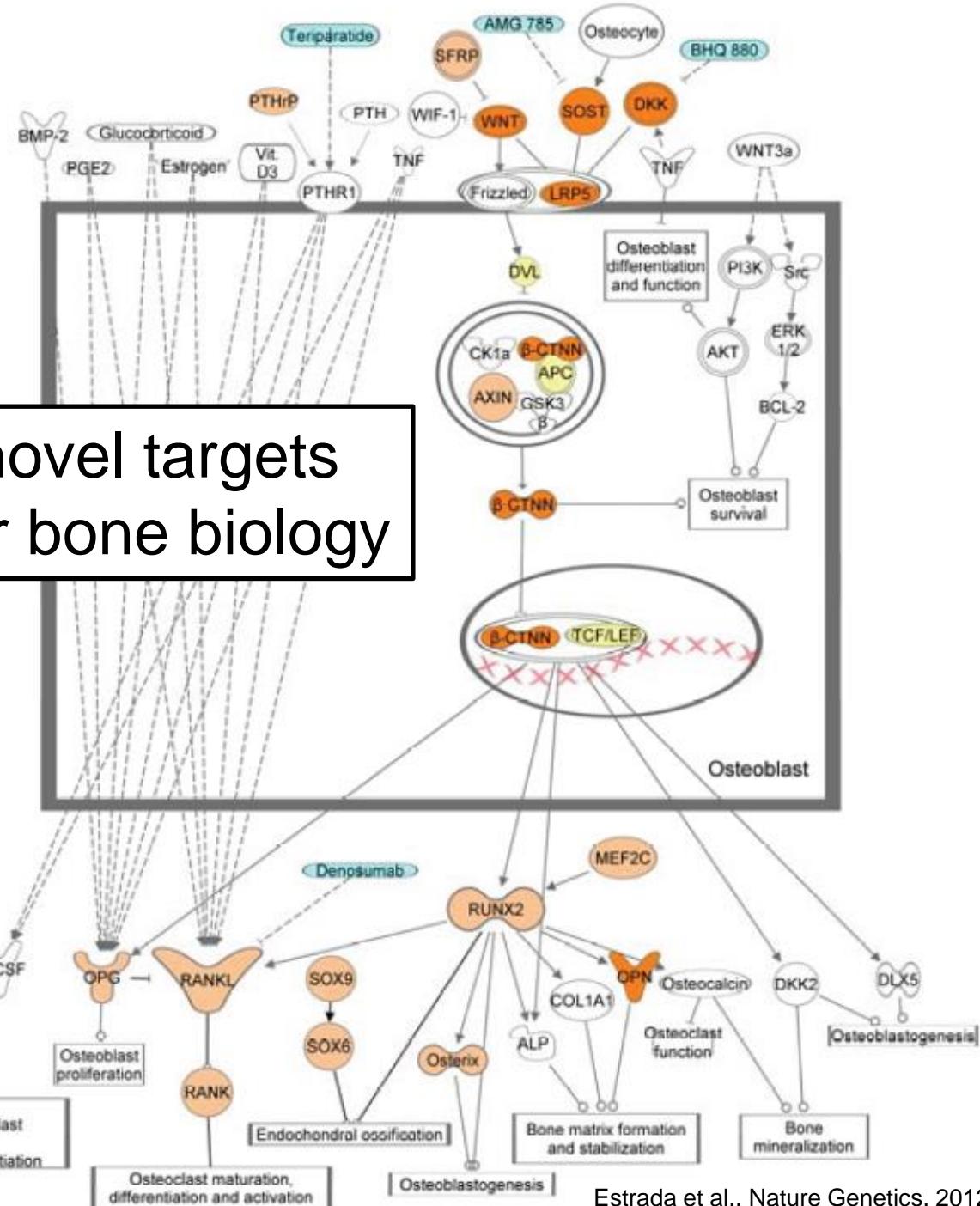


BMD discovery (stage 1)
meta-analysis of 17 BMD genome-wide association studies
($n = 32,961$)

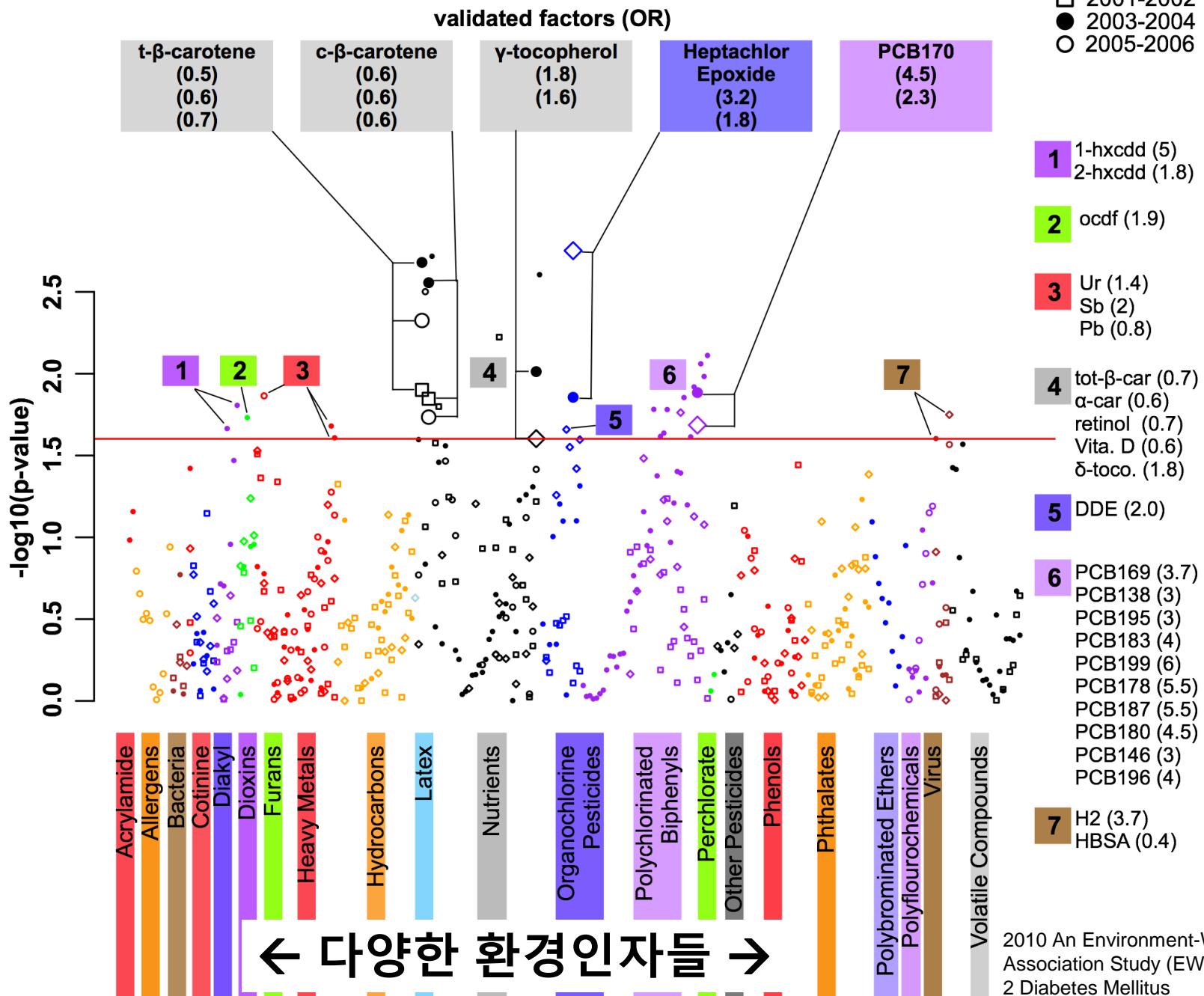
BMD replication (stage 2)
96 SNPs in 34 studies
(*de novo* and *in silico*)
($n = 50,933$)

Association of the BMD loci with fracture (stage 3)
96 SNPs in 50 studies
(*de novo* and *in silico*)
($n = 31,016$ cases and 102,444 controls)

+ novel targets
for bone biology

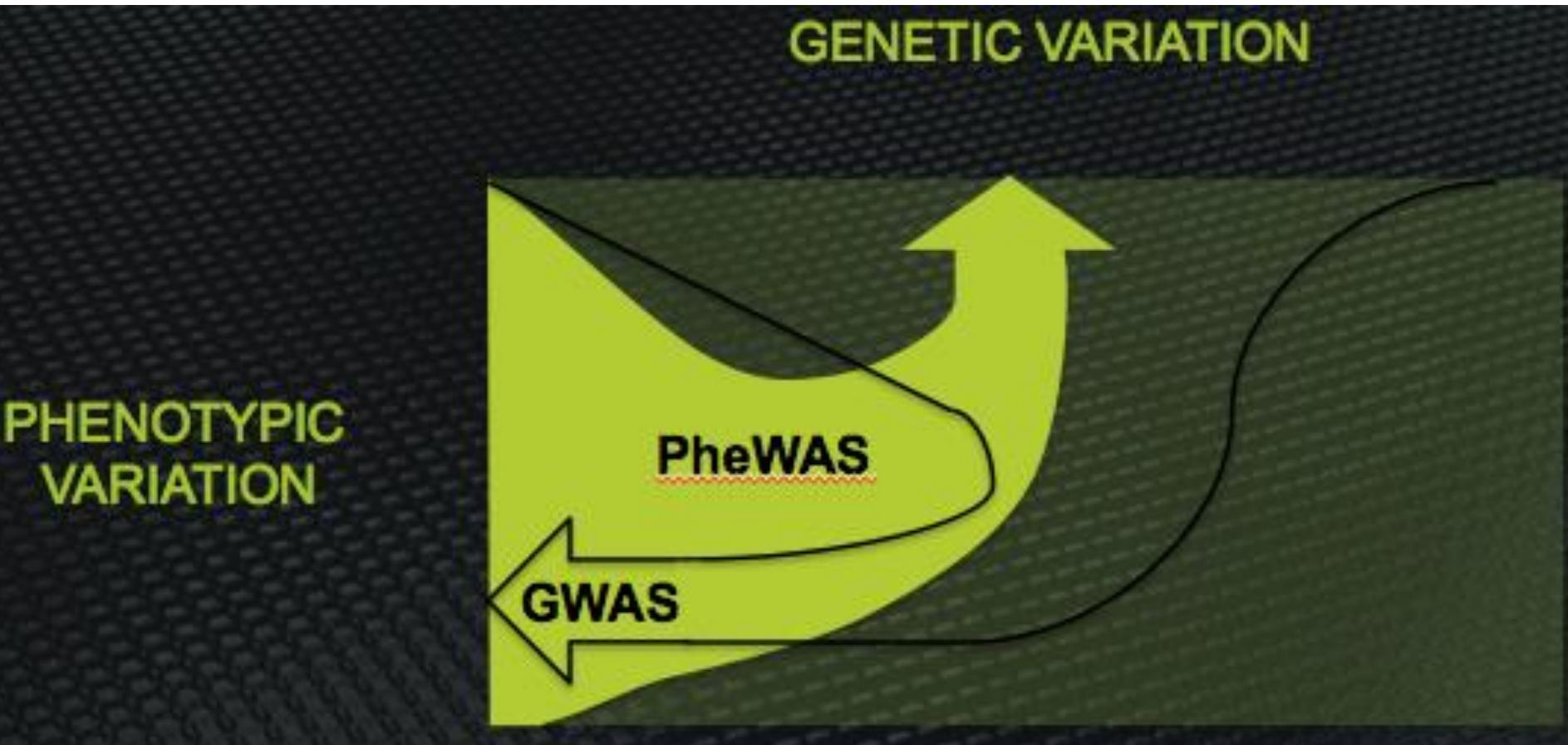


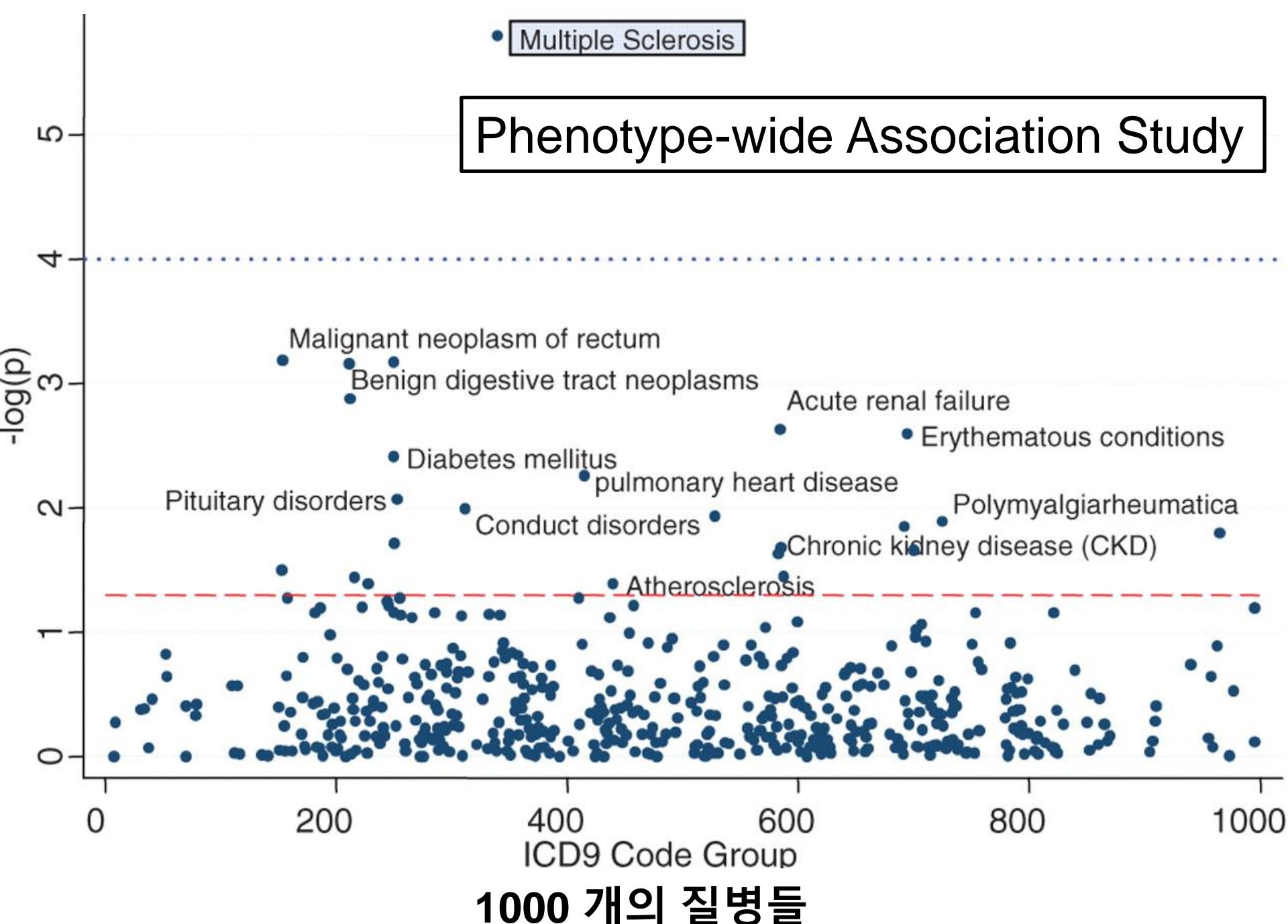
Environment-Wide Association Study (EWAS)



GWAS → PheWAS

Phenotype-wide Association Study

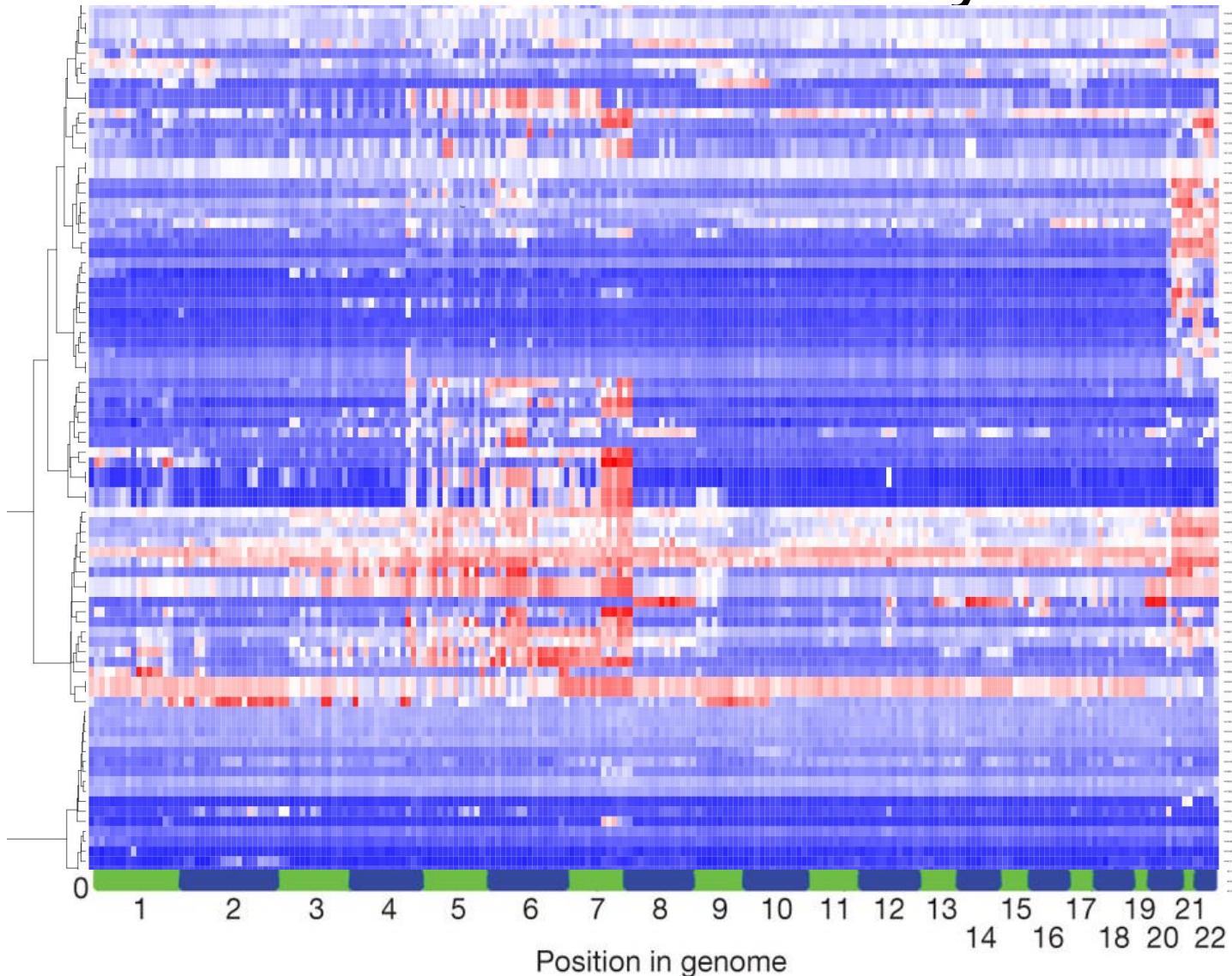




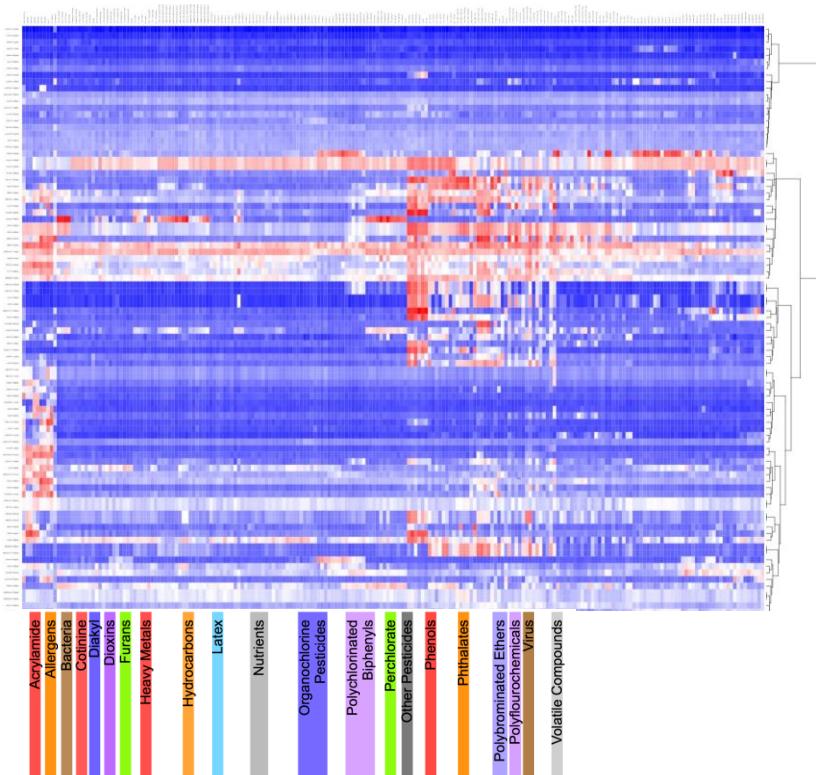
Bioinformatics. 2010 PheWAS: demonstrating the feasibility of a genome-wide scan to discover gene-disease associations.

Genome-wide Phenome-wide Association Study

Phenome-wide
(Lab, Diagnosis)

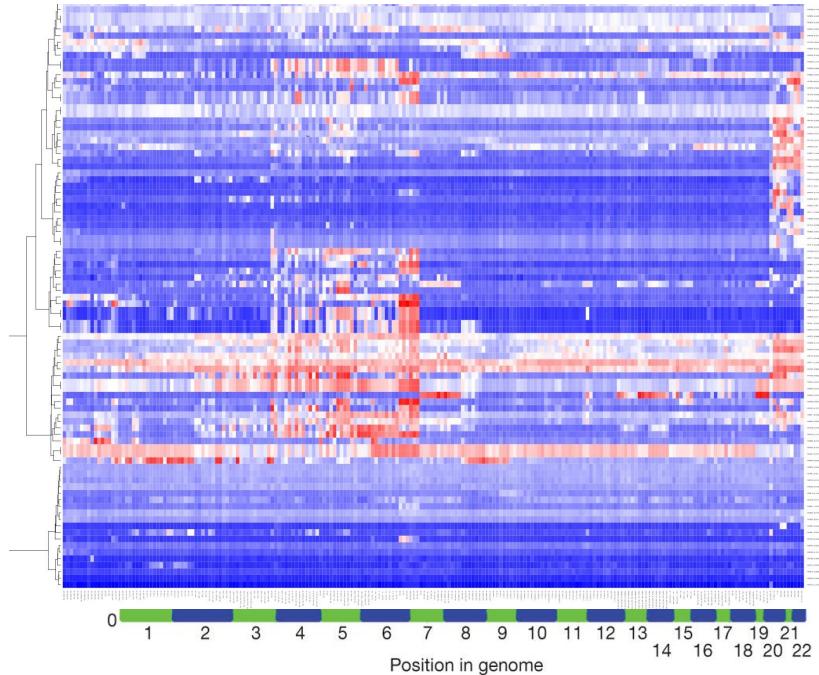


Genome-Envirome-Phenome-wide Association Study



Environment-wide
(Life style, diet, exercise, pollution)

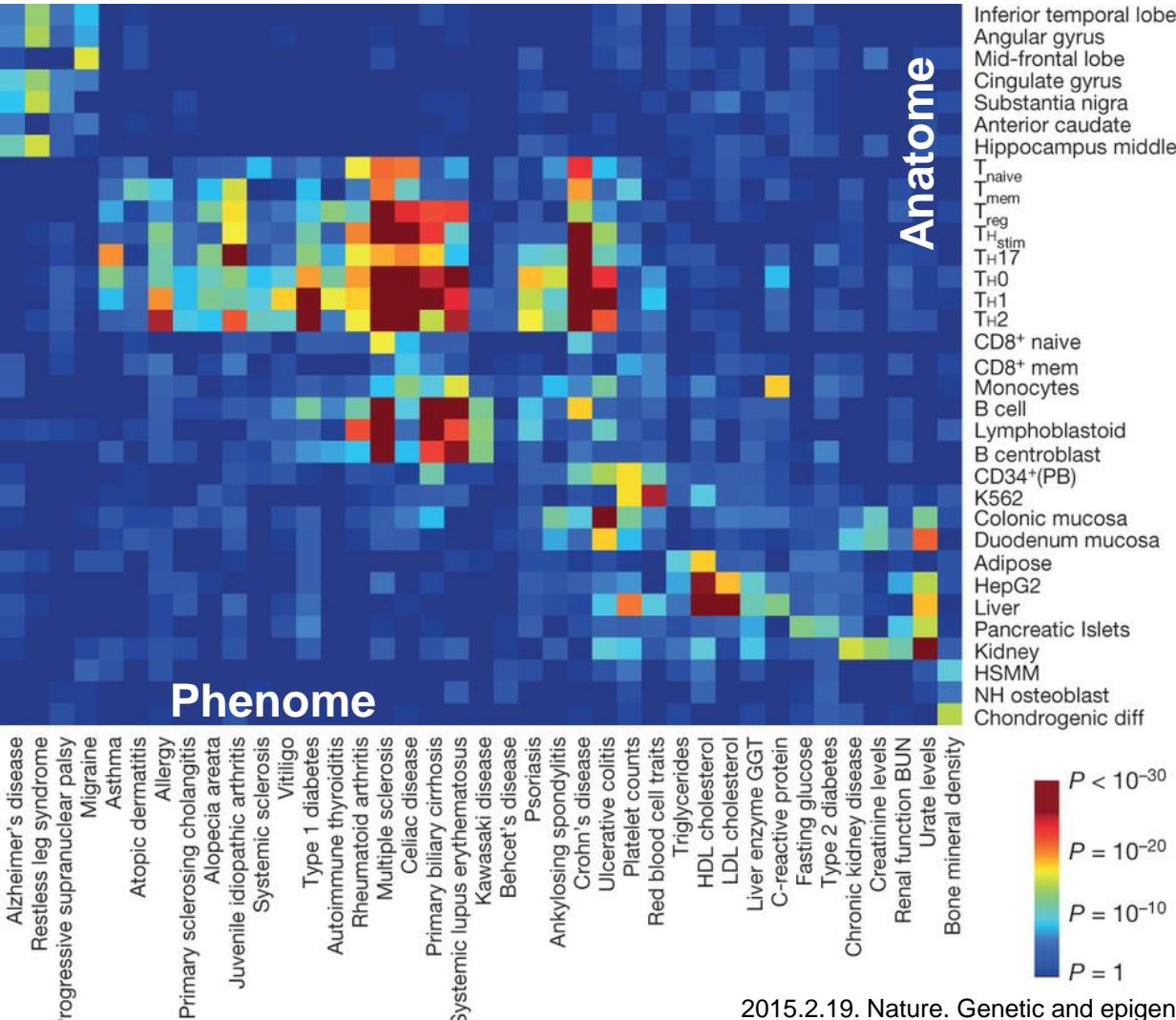
Phenome-wide
(Lab, Diagnosis)



Genome-wide

Proposal
(Choi)

Anatome-Phenome-wide Association Study



헬스케어 3.0 시대의 변화 키워드

기술혁신(Seeds)

IT 융복합 @+

- 디지털 이미징 처리
- 진단소프트웨어
- 의료정보 전자화

유전공학기술

- 유전자 해독
- 줄기세포 치료제
- 유전자 치료제

변화의 4대 키워드

일상
관리화

개인
맞춤화

진단 · 치료
미세화

환자
중심화

시장의 요구(Needs)

건강수명 연장

의료비 절감



주: 선의 굵기는 영향의 강도를 의미

Precision Medicine Initiative

\$215M

2015/1/30



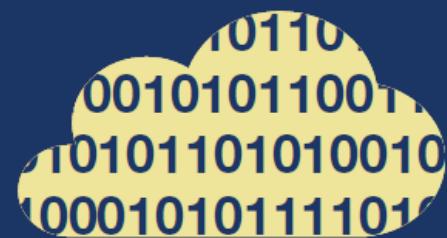
WHY NOW?

The **time is right** because of:

Sequencing
of the human
genome

Improved
technologies for
biomedical analysis

New tools
for using large
datasets



NEAR TERM GOALS

Intensify efforts to apply precision medicine to **cancer**.

Innovative **clinical trials**
of targeted drugs for
adult, pediatric cancers

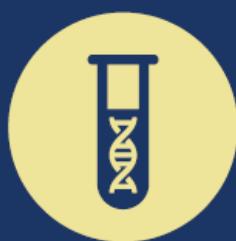
Use of
combination
therapies

Knowledge to
overcome **drug**
resistance



LONGER TERM GOALS

Create a research cohort of **> 1 million American volunteers** who will share genetic data, biological samples, and diet/lifestyle information, all linked to their electronic health records if they choose.



Research based upon the cohort data will:

- Advance **pharmacogenomics**, the right drug for the right patient at the right dose
- Identify new targets for **treatment and prevention**
- Test whether **mobile devices** can encourage healthy behaviors
- Lay **scientific foundation** for precision medicine for **many diseases**



Contents

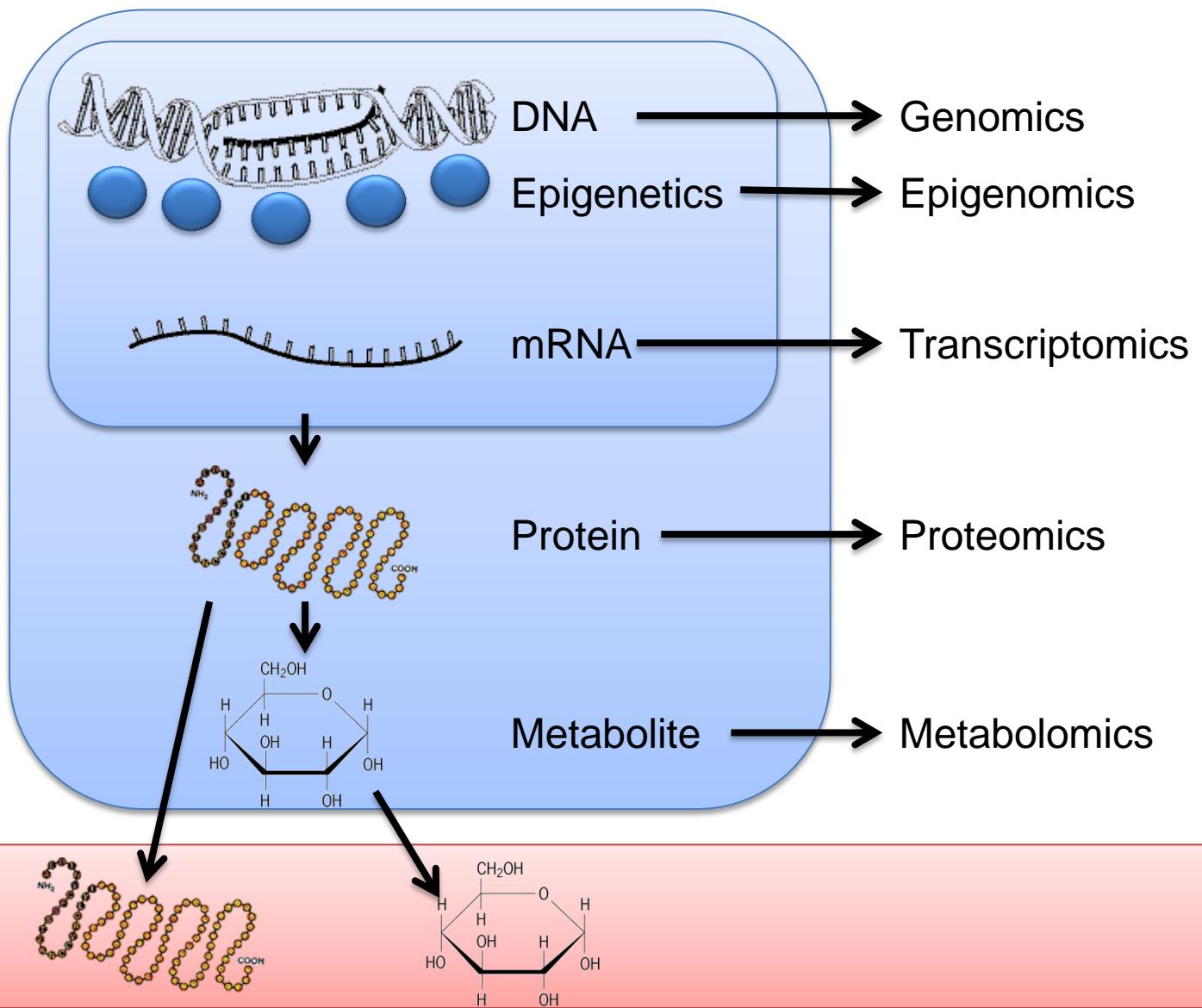
1. What is Healthcare Big Data?
2. Healthcare Big Data

- ① Genetic Data
- ② Electrical Health Records
- ③ National Healthcare Data
- ④ Medical Images
- ⑤ Sensor/Mobile Data

3. Clinical and Research Applications



Genetics Information and OMICs



Germline Genomics

Cancer Genomics

Germline mutation

WBC

DNA

Liver

Somatic mutation

Cancer

mRNA

Epigenomics

Transcriptomics

Proteomics

Epigenomics

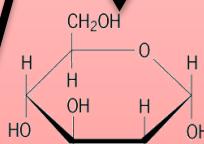
Transcriptomics

Proteomics

Epigenomics

Transcriptomics

Proteomics



Metabolomics



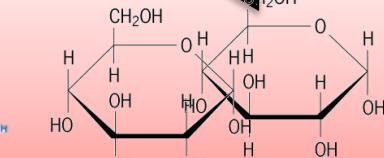
Proteomics



Metabolomics

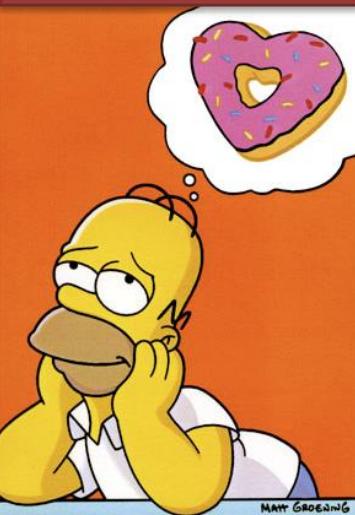


Metabolomics

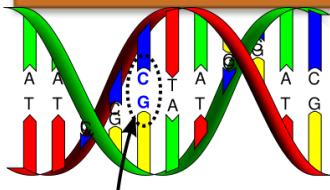


Diabetes and Epigenetics Study

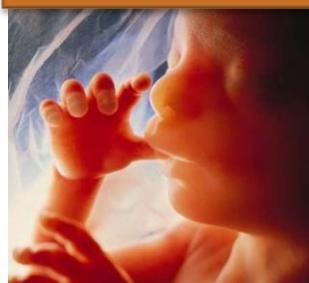
Food



Genetic Predisposition



Environmental Predisposition

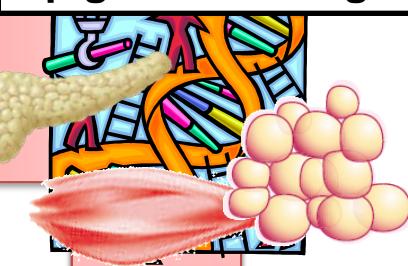


Exercise



Diabetes

1. Cause of Diabetes



2. Result of Diabetes

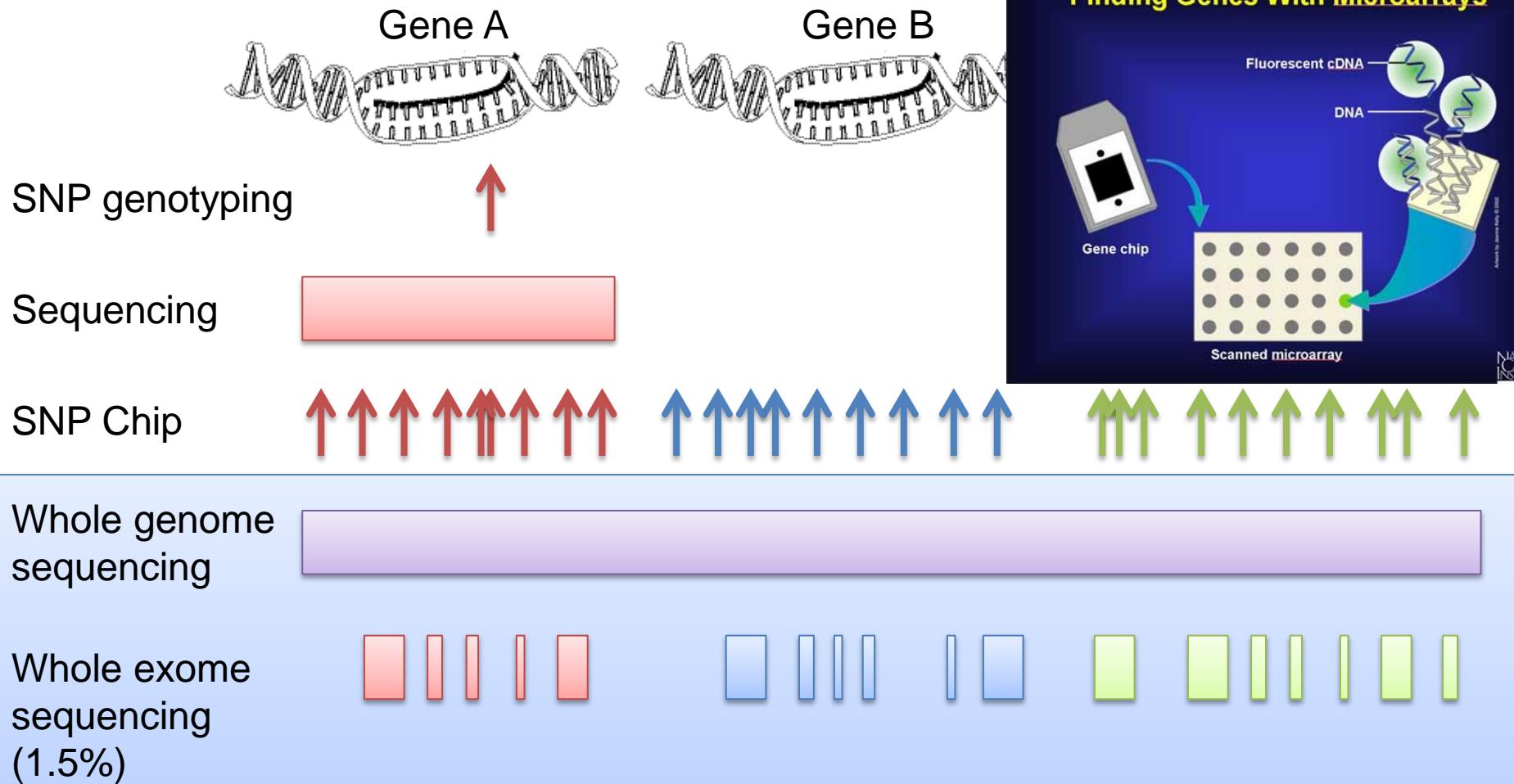
Epigenetic change



Diabetic Complications

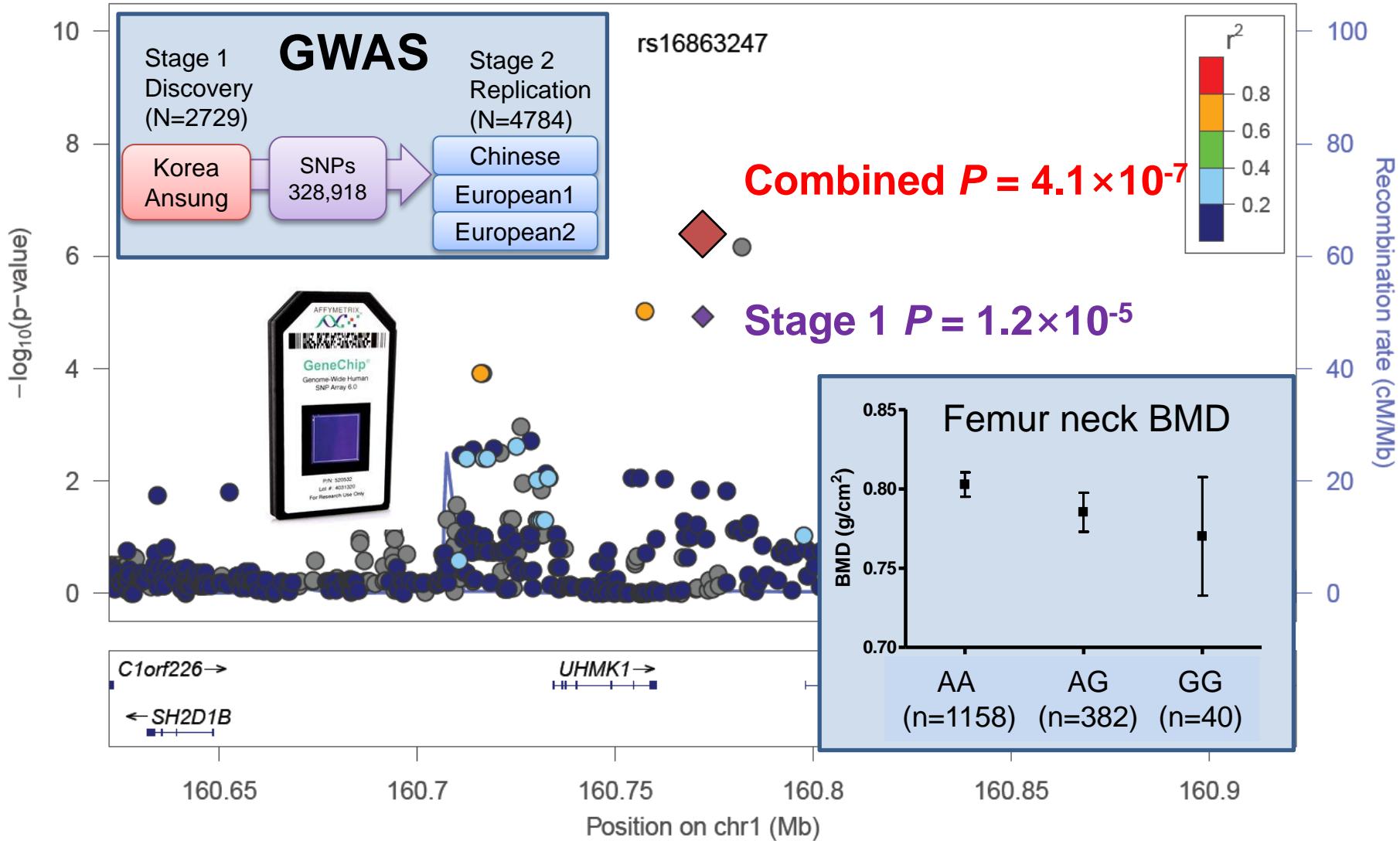


Genetic Variation Analyses

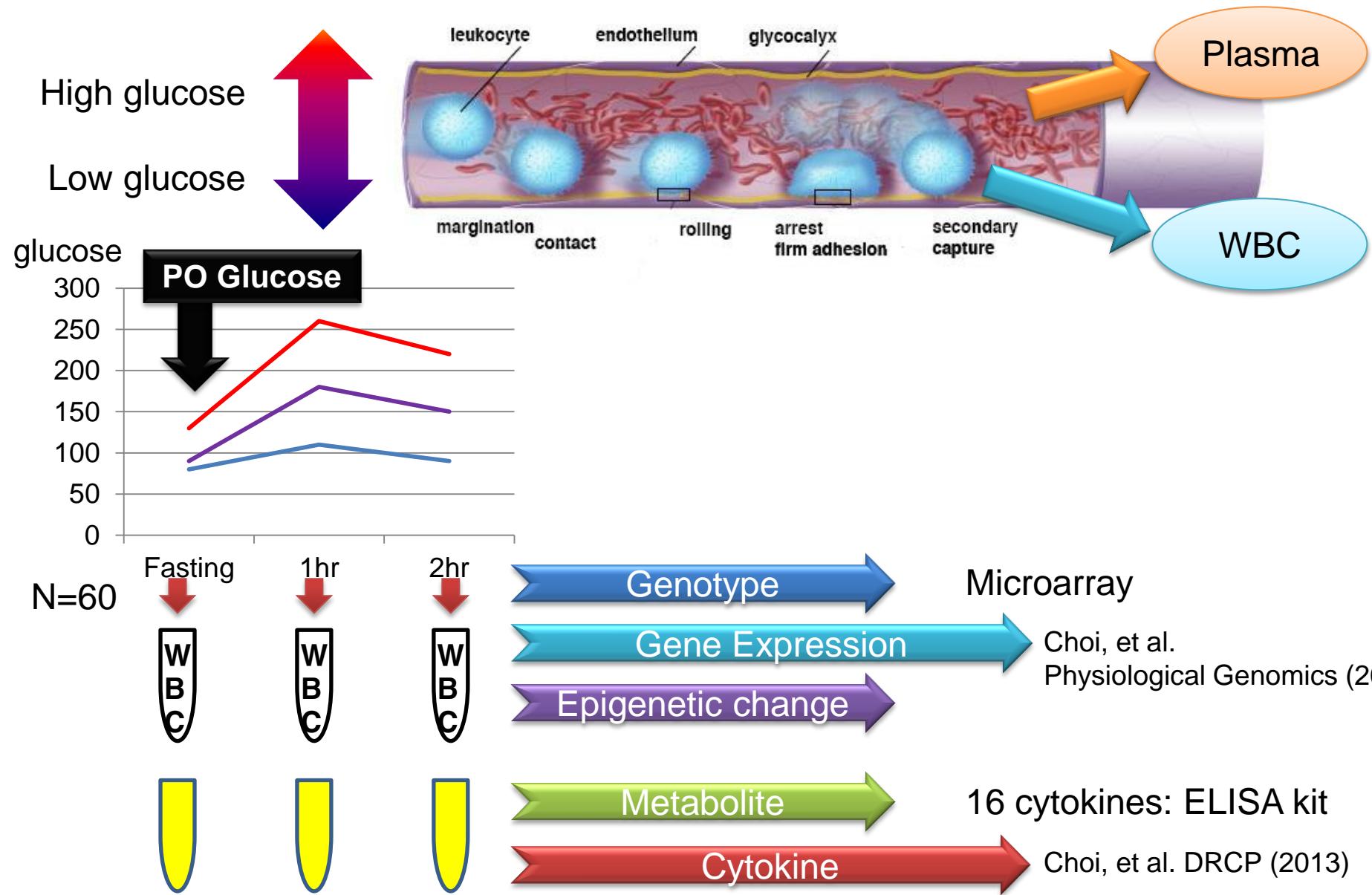


Next Generation Sequencing (NGS)

Bone Mineral Density GWAS



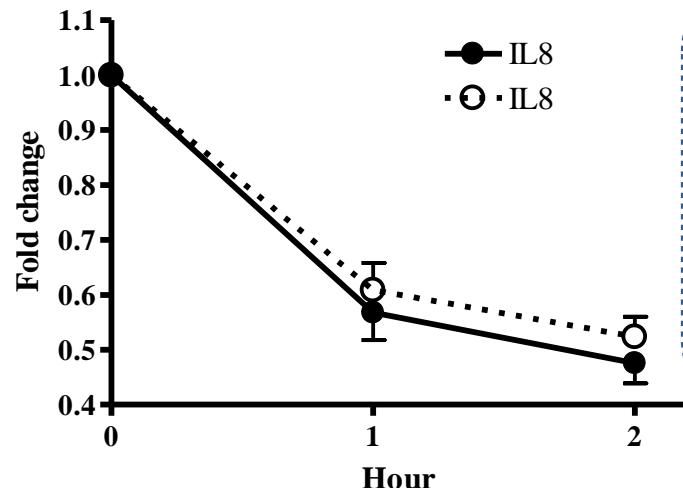
Human Blood, Cell Biospecimen Study



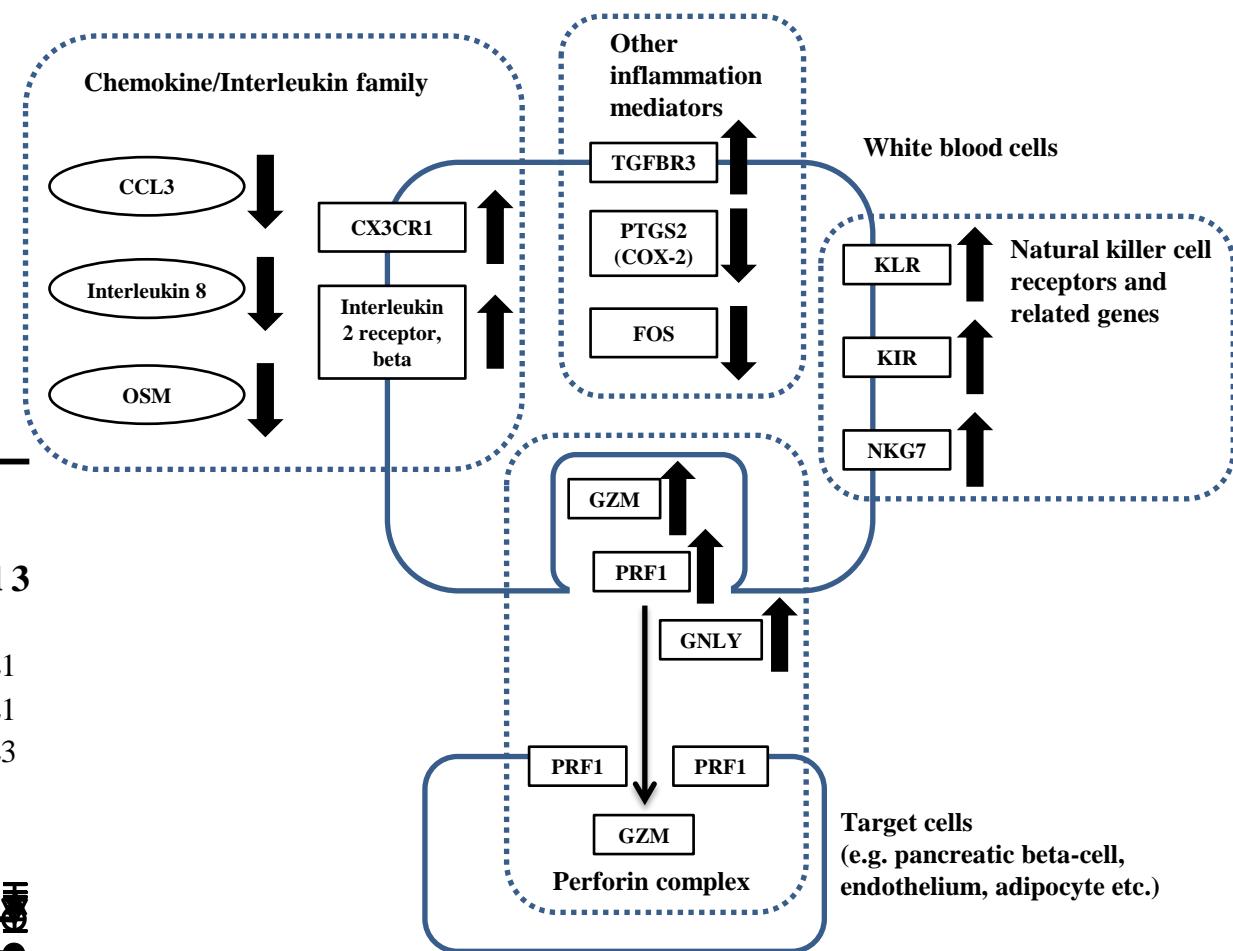
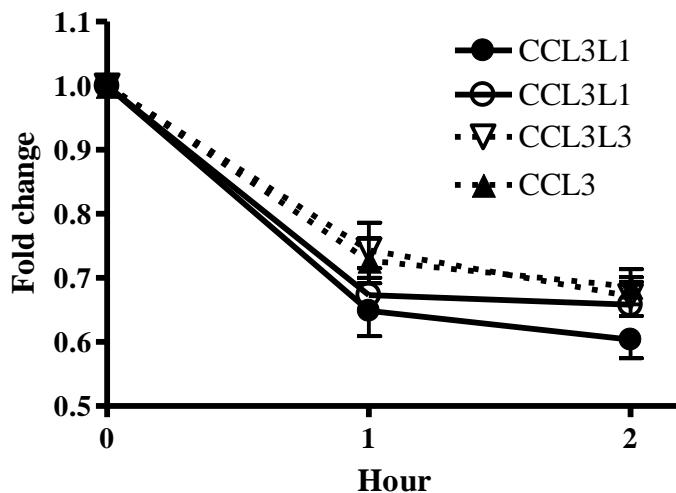
Transcriptome Response to Hyperglycemia

Microarray (N=60x3)

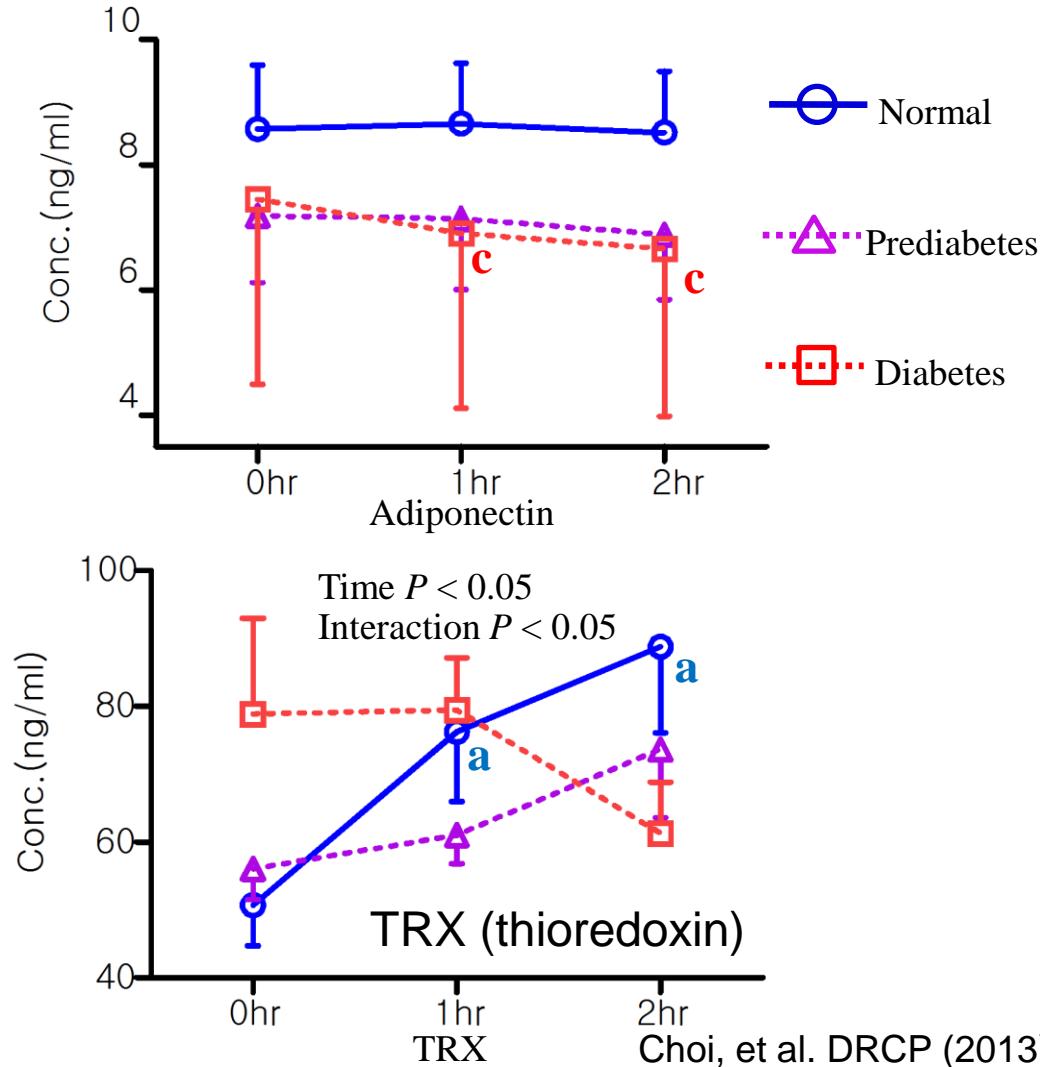
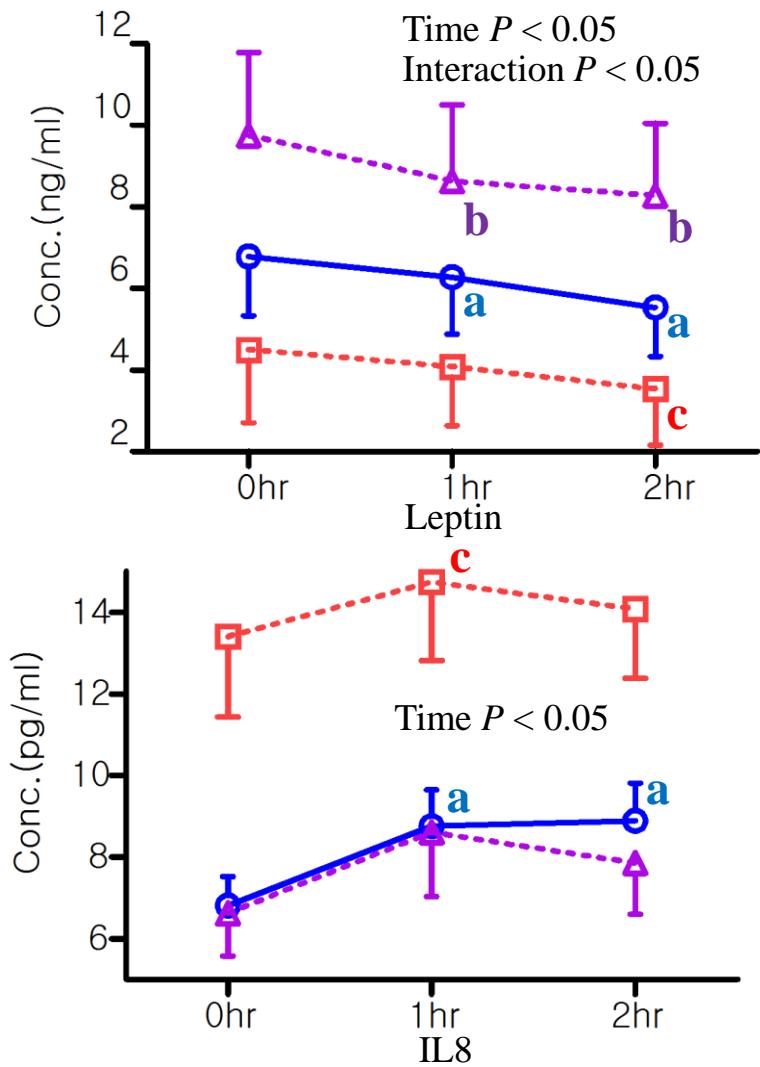
Interleukin 8



Chemokine (C-C motif) ligand 3



Postprandial Change of Cytokine Concentrations during OGTT



Metabolomics Profiling

- Quantification of 186 metabolites
 - Acylcarnitines
 - Amino Acids
 - Biogenic Amines
 - Hexoses (sum of Hexoses)
 - Phospho and Sphingolipids
 - Phosphatidylcholines
 - Lysophosphatidylcholines
 - Sphingomyelin



Absolute/DQ® p180 Kit
Targeted Metabolite
Identification and Quantification

Comprehensive Catalogues of Genomic Data

Variation in the human genome

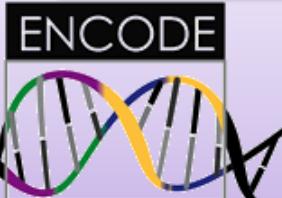


Four ethnic groups
(CEU, YRI, JPT, CHB, N=270)



Whole genome sequencing (N=1,000)

Functional elements



Disease-related variations



Mendelian (monogenic) diseases
(N=22,432)



GWAS catalog
Complex (multigenic) traits
(1926 publications and 13410 SNPs)

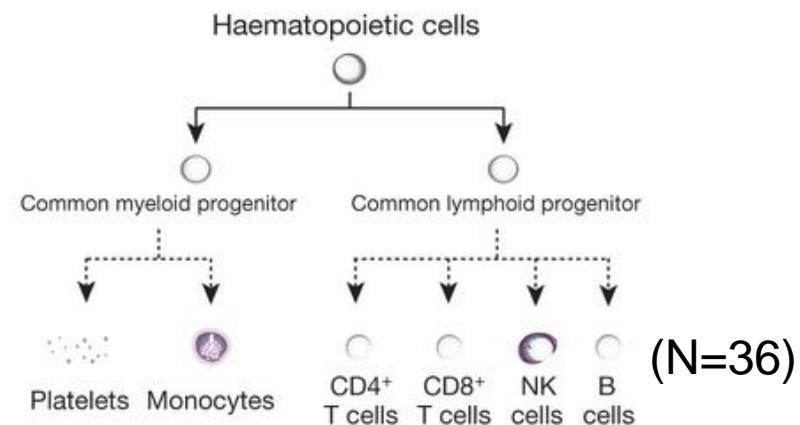
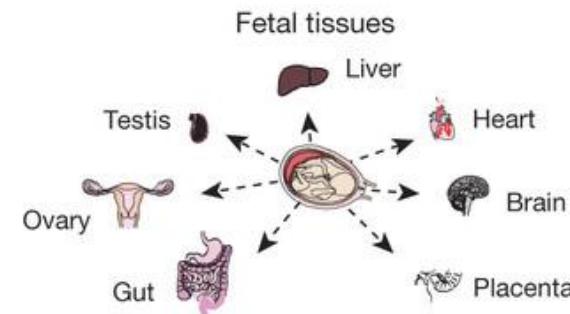
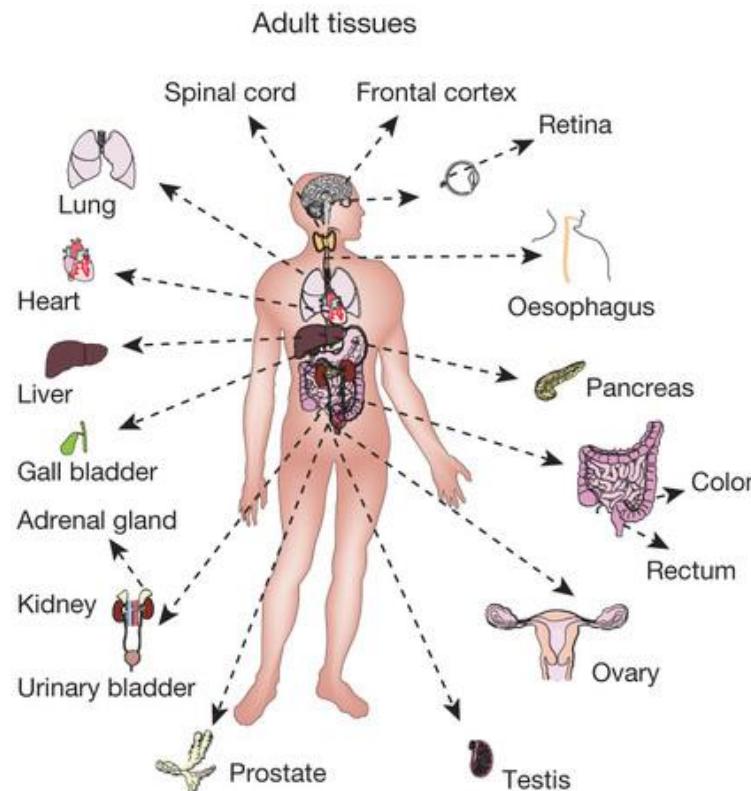
2014-06-29

Tissue Specific Expression

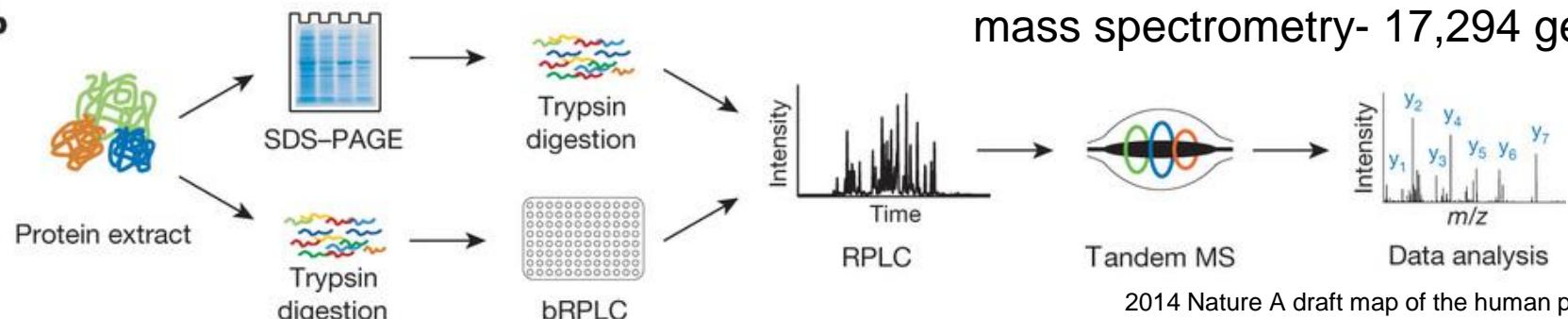


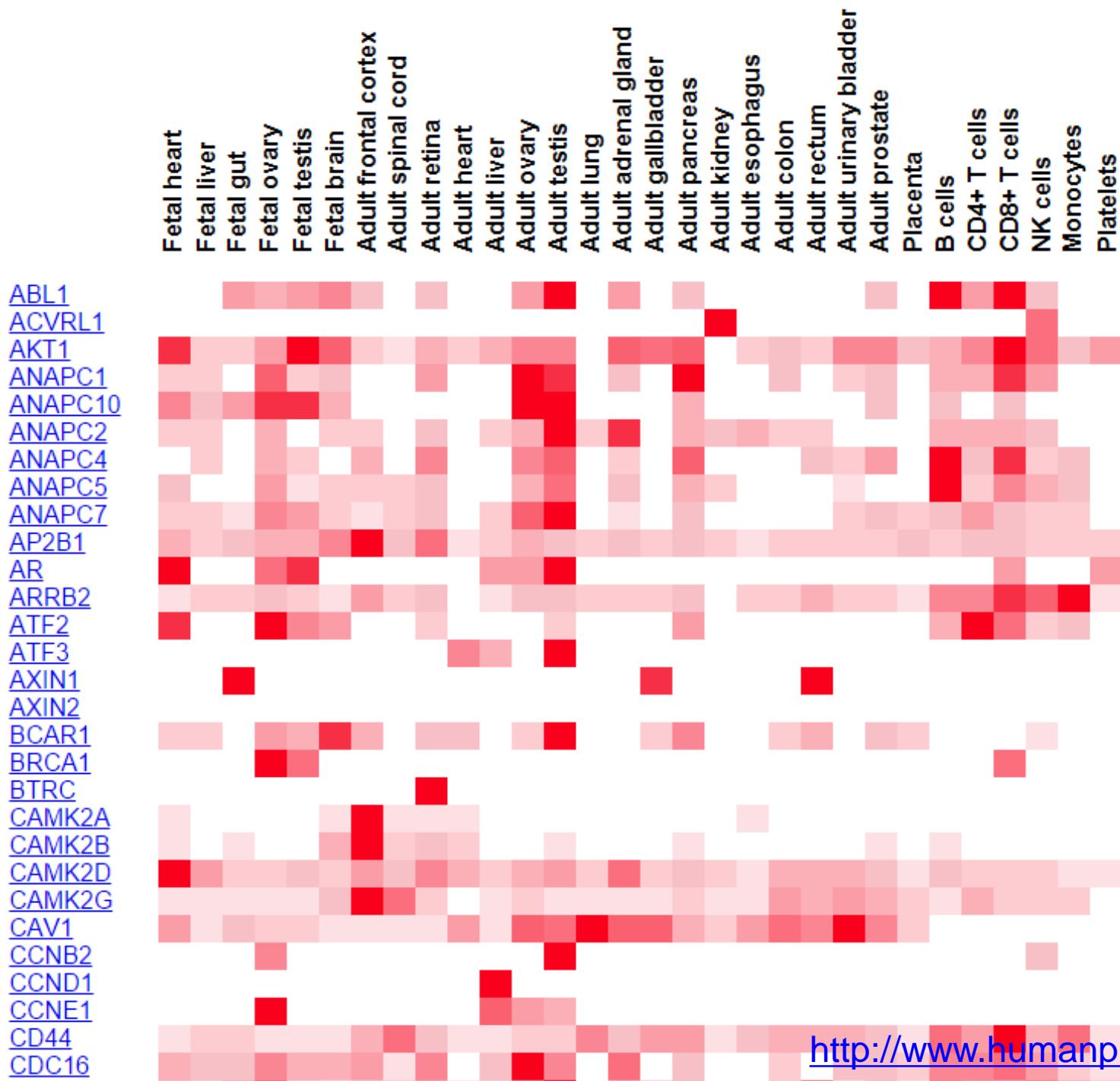
Adult/Fetal Tissues and Blood Cells

a

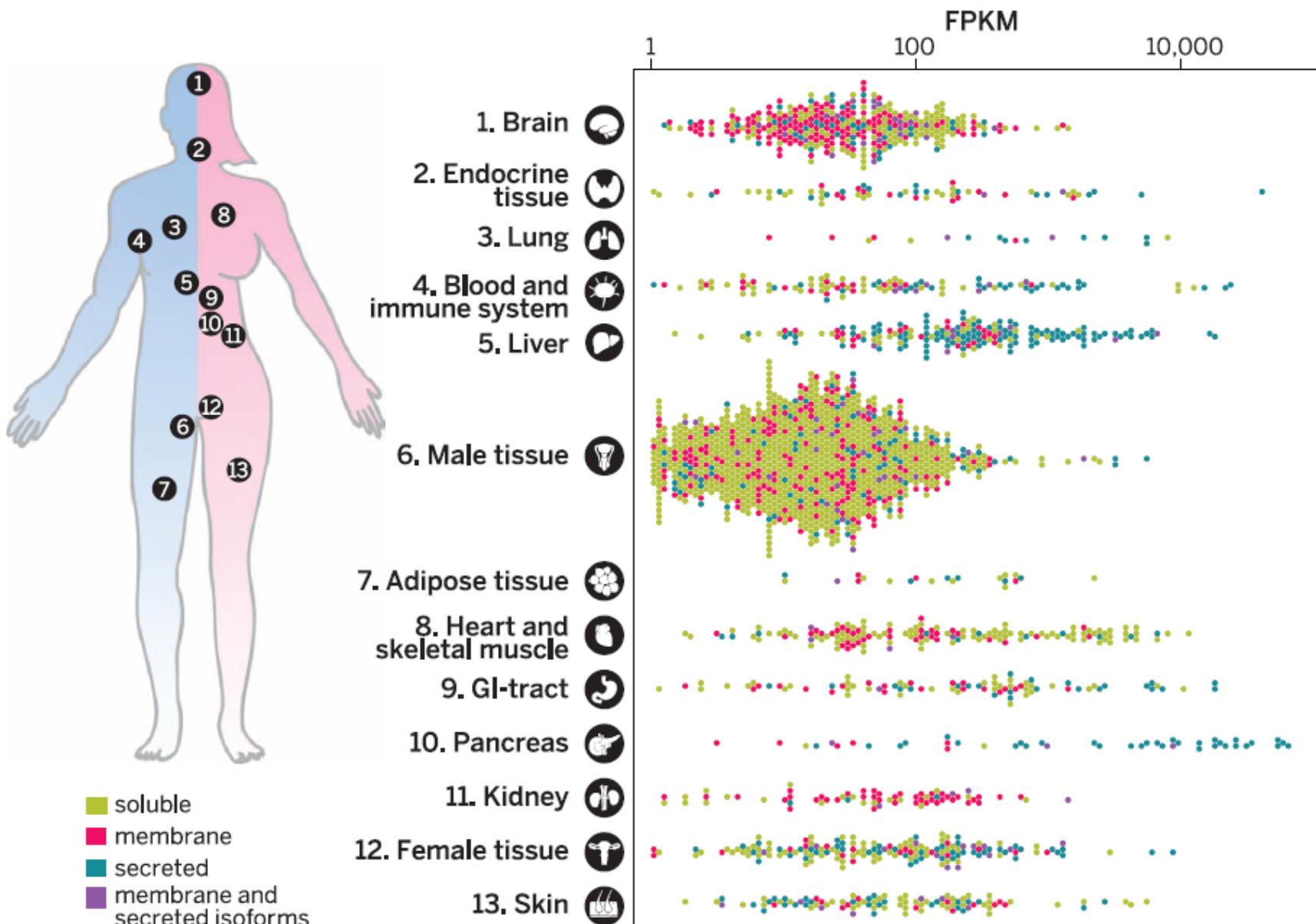


b

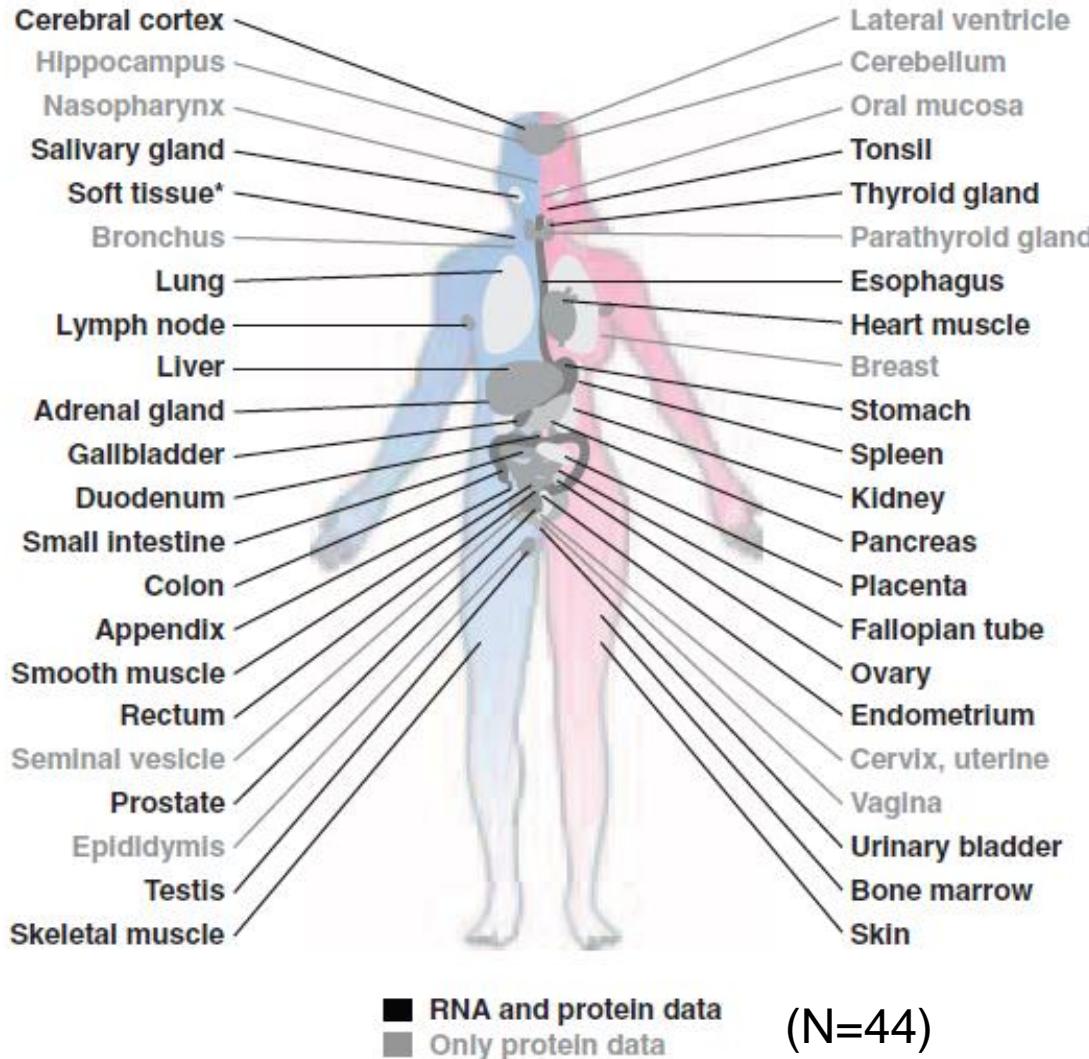




Tissue-based map of the human proteome



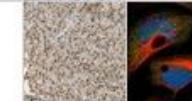
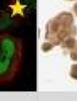
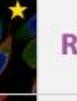
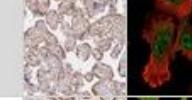
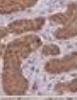
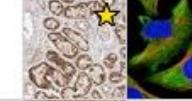
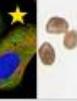
All Major Tissues/Organs All Proteins + All RNAs



1. Immunohistochemistry (IHC)
24,028 antibodies
(16,975 proteins)
→ >13 million IHC images

2. RNA-sequencing

391 GENES FOUND [? »](#)Limit search: [Genes with antibodies](#) | [Premium](#) | [Premium \(Tissue\)](#) | [Premium \(Subcell\)](#) | [Premium \(Cell line\)](#)Page **1** of 8 | [next »](#)[XML](#) | [RDF](#) | [TAB](#)

Gene	Gene description	x	Protein class	x	Tissue	Subcell	Cell line	Cancer	RNA tissue category	x
MAPK1	Mitogen-activated protein kinase 1		Cancer-related genes Cytoskeleton related proteins Enzymes FDA approved drug targets Plasma proteins RAS pathway related proteins							Expressed in all
MAP3K5	Mitogen-activated protein kinase kinase kinase 5		Enzymes Plasma proteins			N/A				Expressed in all
MAP3K4	Mitogen-activated protein kinase kinase kinase 4		Cancer-related genes Enzymes							Expressed in all
MAPK3	Mitogen-activated protein kinase 3		Cancer-related genes Cytoskeleton related proteins Enzymes FDA approved drug targets RAS pathway related proteins							Expressed in all
MAPKAPK3	Mitogen-activated protein kinase-activated protein kinase 3		Enzymes Predicted membrane proteins					N/A		Expressed in all
MAPK14	Mitogen-activated protein kinase 14		Cancer-related genes Enzymes							Expressed in all
MAPK6	Mitogen-activated protein kinase 6		Enzymes							Expressed in all
MAPK8IP2	Mitogen-activated protein kinase 8 interacting protein 2					N/A				Tissue enriched

MAPKAP1



GENE/PROTEIN

ANTIBODY/ANTIGEN

TISSUE ATLAS

STAINING OVERVIEW

Dictionary



Dictionary

TISSUE ATLAS ? »

Gene description Mitogen-activated protein kinase associated protein 1

RNA tissue category Expressed in all.

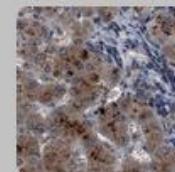
Protein summary Detected at High or Medium expression levels in 57 of 80 analyzed normal tissue cell types.

Protein expression General cytoplasmic expression at variable levels.

Predicted localization Intracellular

Protein evidence Evidence at protein level

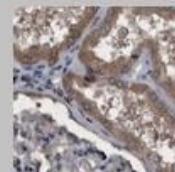
Protein reliability Supportive based on 2 antibodies.



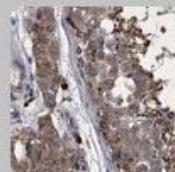
Liver



Colon



Kidney



Testis



Lymph node



Cerebral cortex

RNA

Expression (FPKM)

100 50 0

N/A

Organ system

Liver and pancreas

Liver

Gallbladder

Pancreas

Digestive tract (GI-tract)

Oral mucosa

Salivary gland

Esophagus

Stomach

Duodenum

Protein

Localization (score)

n l m h



MAPKAP1



GENE/PROTEIN

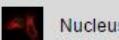
ANTIBODY/ANTIGEN

SUBCELL ATLAS

Dictionary

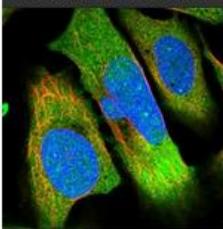


Cytoplasm



Nucleus

SUBCELL ATLAS - HUMAN ? »



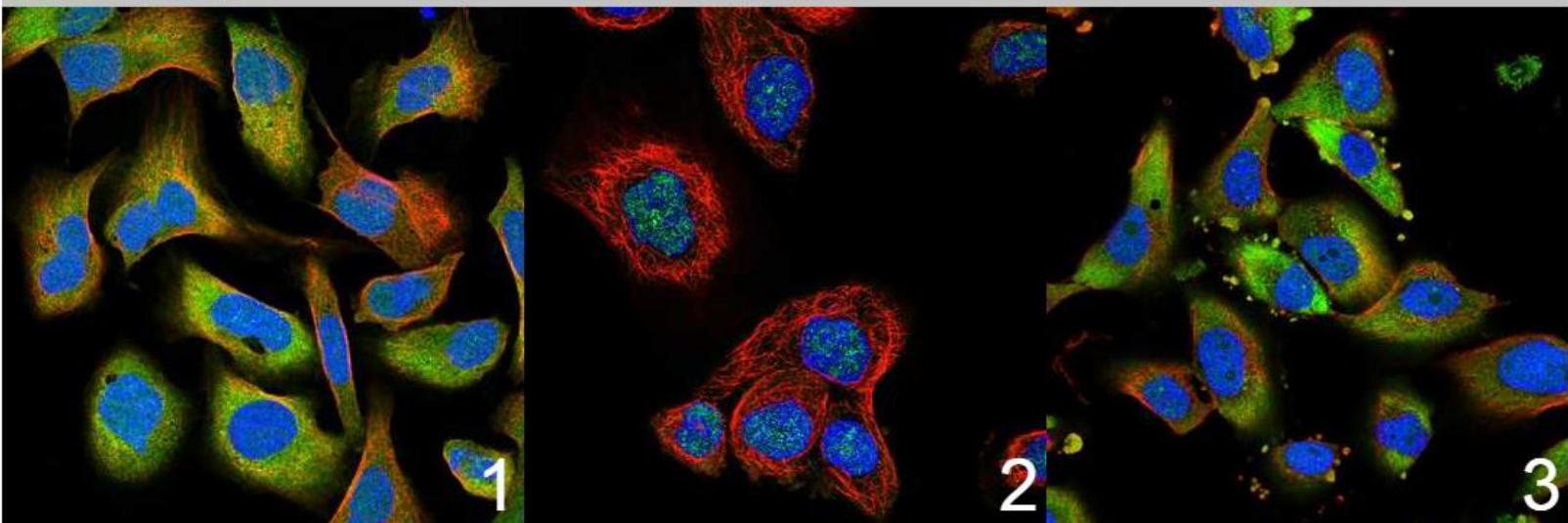
Summary
Localized to the nucleus & cytoplasm.

Main location
Nucleus, Cytoplasm

Reliability
Supportive

Protein evidence
Evidence at protein level

Assay summary
Analysis based on two antibodies, HPA029091 and HPA029092, using immunofluorescence in human cells



HPA029092: U-2 OS

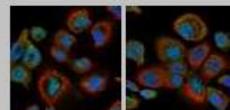
HPA029092: A-431

HPA029092: U-251 MG

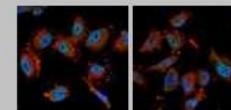
Toggle
channels
 Antibody Nucleus Intensity
 Microtubules ER

HPA029091

HPA029091 stained U-
2 OS
cells unspecifically



Cell line: A-431
Location: Nucleus but not
nucleoli, Cytoplasm
Literature validation: Supportive



Cell line: U-251 MG
Location: Nucleus but
not nucleoli
Literature validation:
Supportive

MAPKAP1



GENE/PROTEIN

ANTIBODY/ANTIGEN

SUBCELL ATLAS

Dictionary

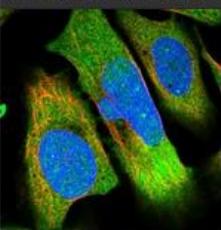
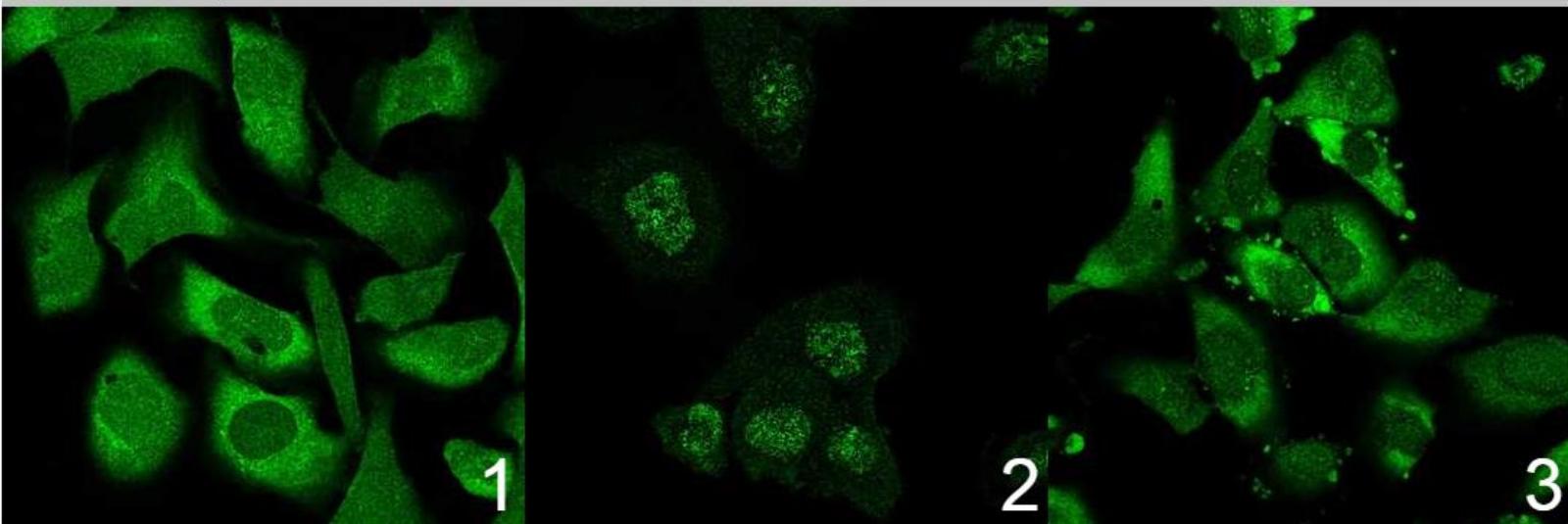


Cytoplasm



Nucleus

SUBCELL ATLAS - HUMAN ? »

Summary
Localized to the nucleus & cytoplasm.Main location
Nucleus, CytoplasmReliability
SupportiveProtein evidence
Evidence at protein levelAssay summary
Analysis based on two antibodies, HPA029091 and HPA029092, using immunofluorescence in human cells

HPA029092: U-2 OS

HPA029092: A-431

HPA029092: U-251 MG

Toggle
channels

Antibody

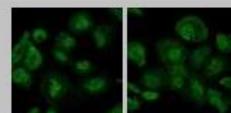
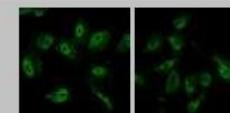
Microtubules

Nucleus

ER

Intensity

HPA029091

HPA029091 stained U-
2 OS
cells unspecificallyCell line: A-431
Location: Nucleus but not
nucleoli, Cytoplasm
Literature validation: SupportiveCell line: U-251 MG
Location: Nucleus but
not nucleoli
Literature validation:
Supportive

MAPKAP1



GENE/PROTEIN

ANTIBODY/ANTIGEN

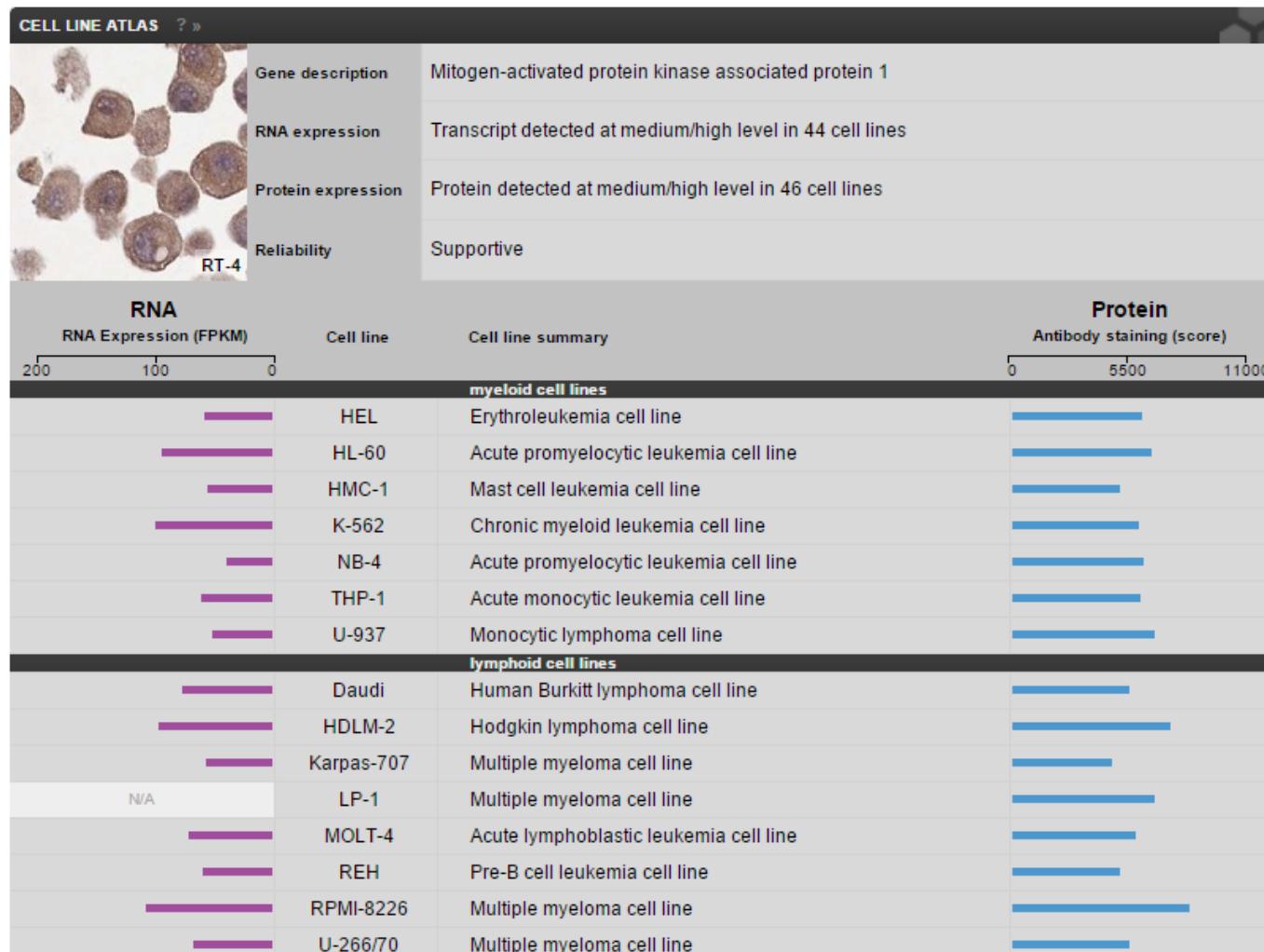
CELL ATLAS HPA029091

CELL ATLAS HPA029092

Dictionary



Dictionary



MAPKAP1



GENE/PROTEIN

ANTIBODY/ANTIGEN

CANCER ATLAS

STAINING OVERVIEW

Dictionary



Dictionary

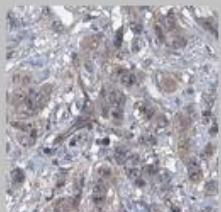
CANCER ATLAS ? »

Gene description Mitogen-activated protein kinase associated protein 1

Protein evidence Evidence at protein level



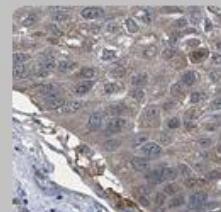
Colorectal cancer



Breast cancer



Prostate cancer



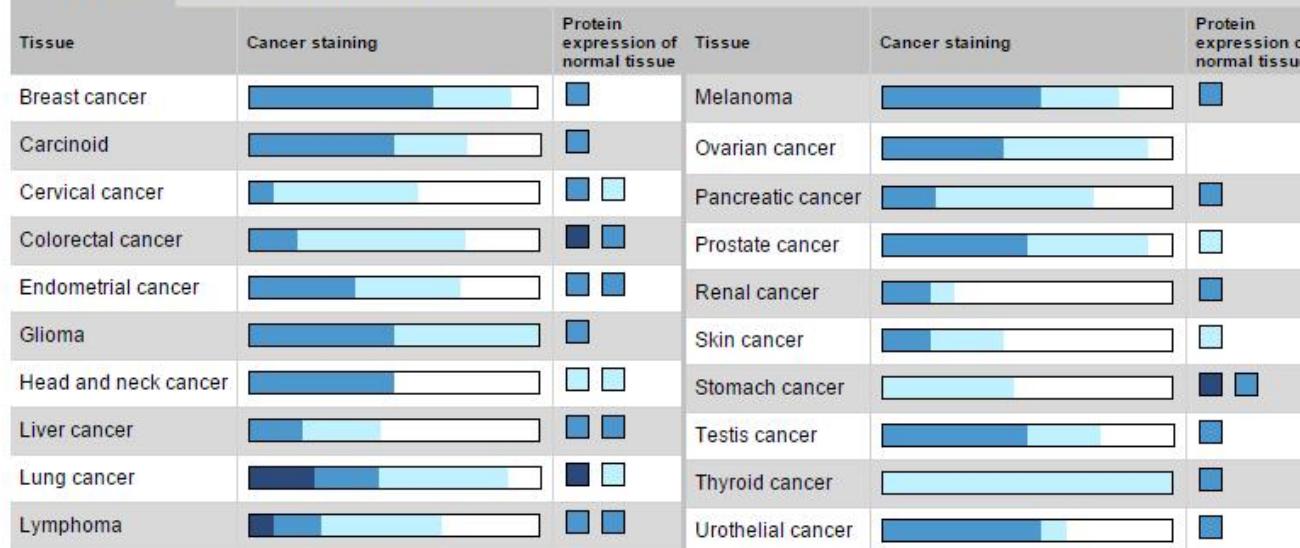
Lung cancer



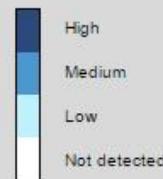
Liver cancer

STAINING SUMMARY - BASED ON HPA029092 ? »

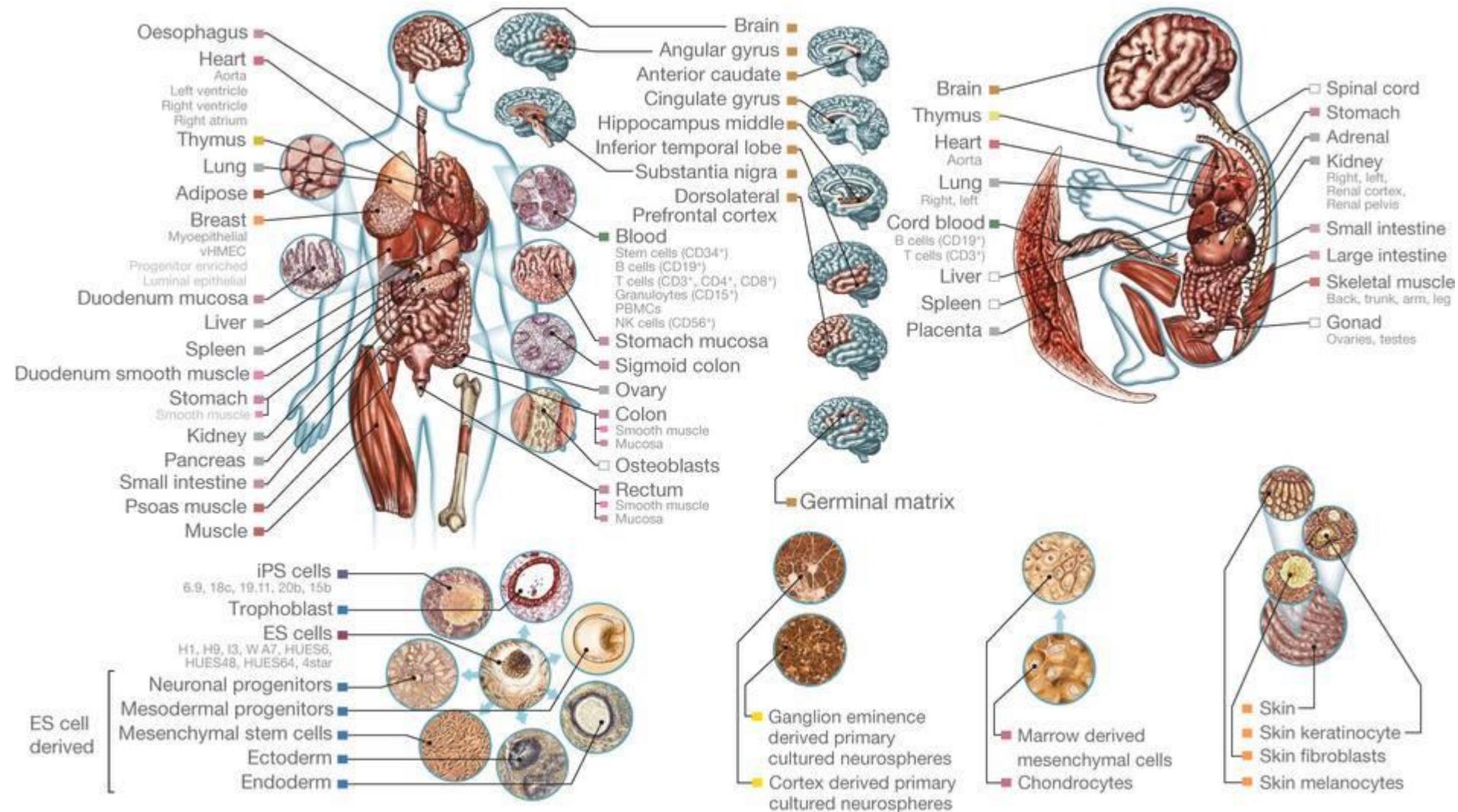
Staining summary Antibody staining in 70% of the cancers

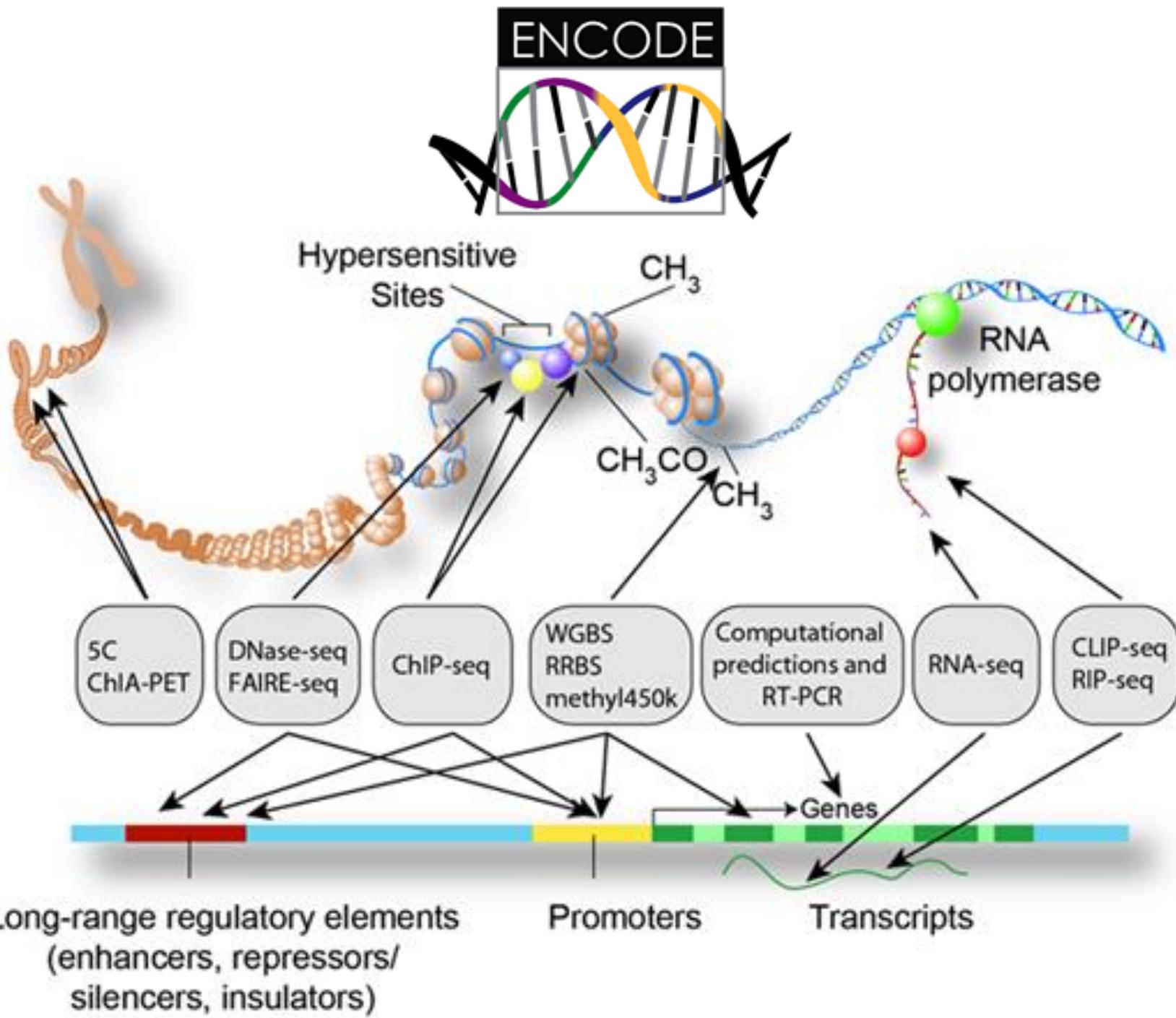


Level of antibody staining/expression

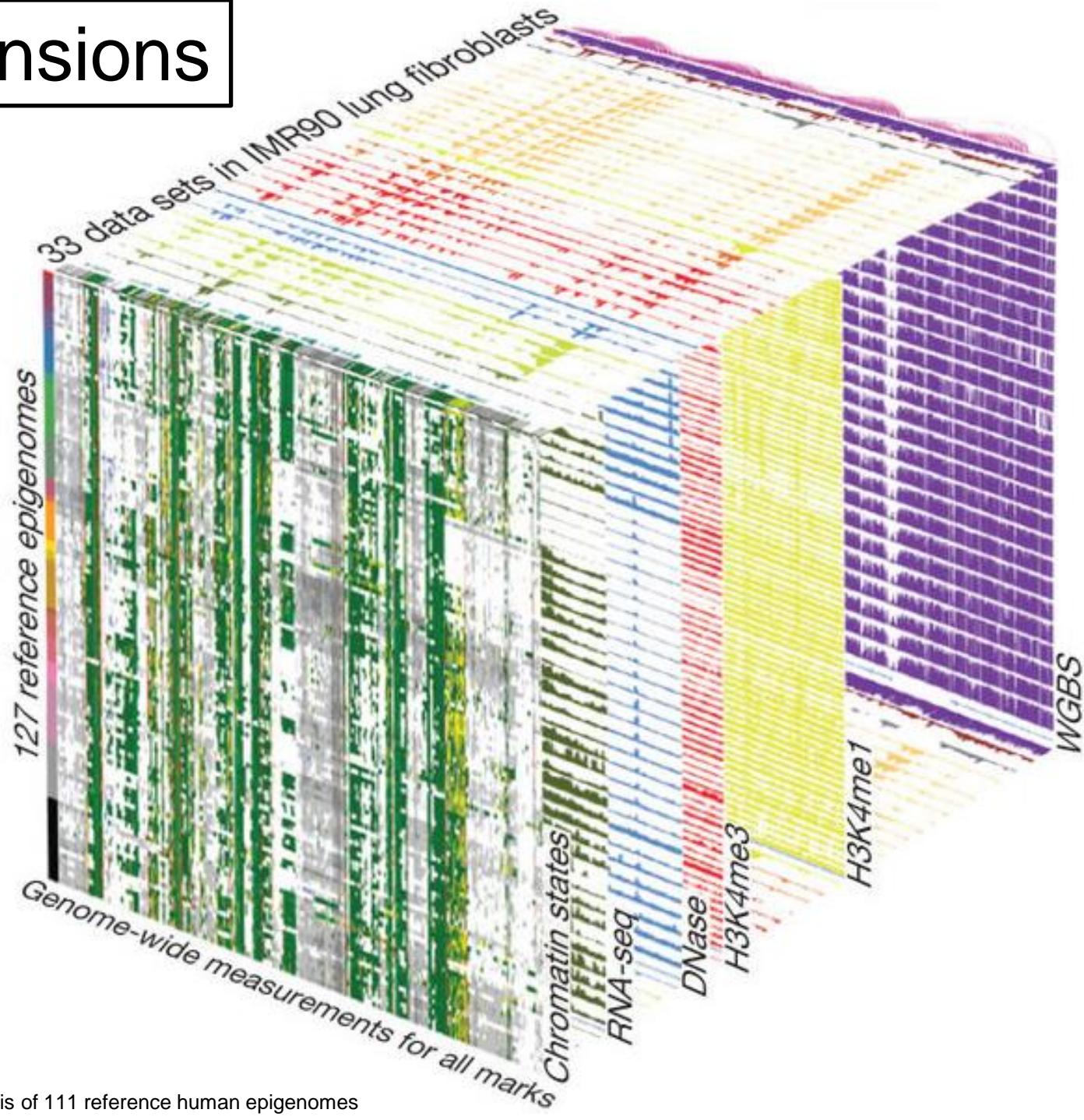


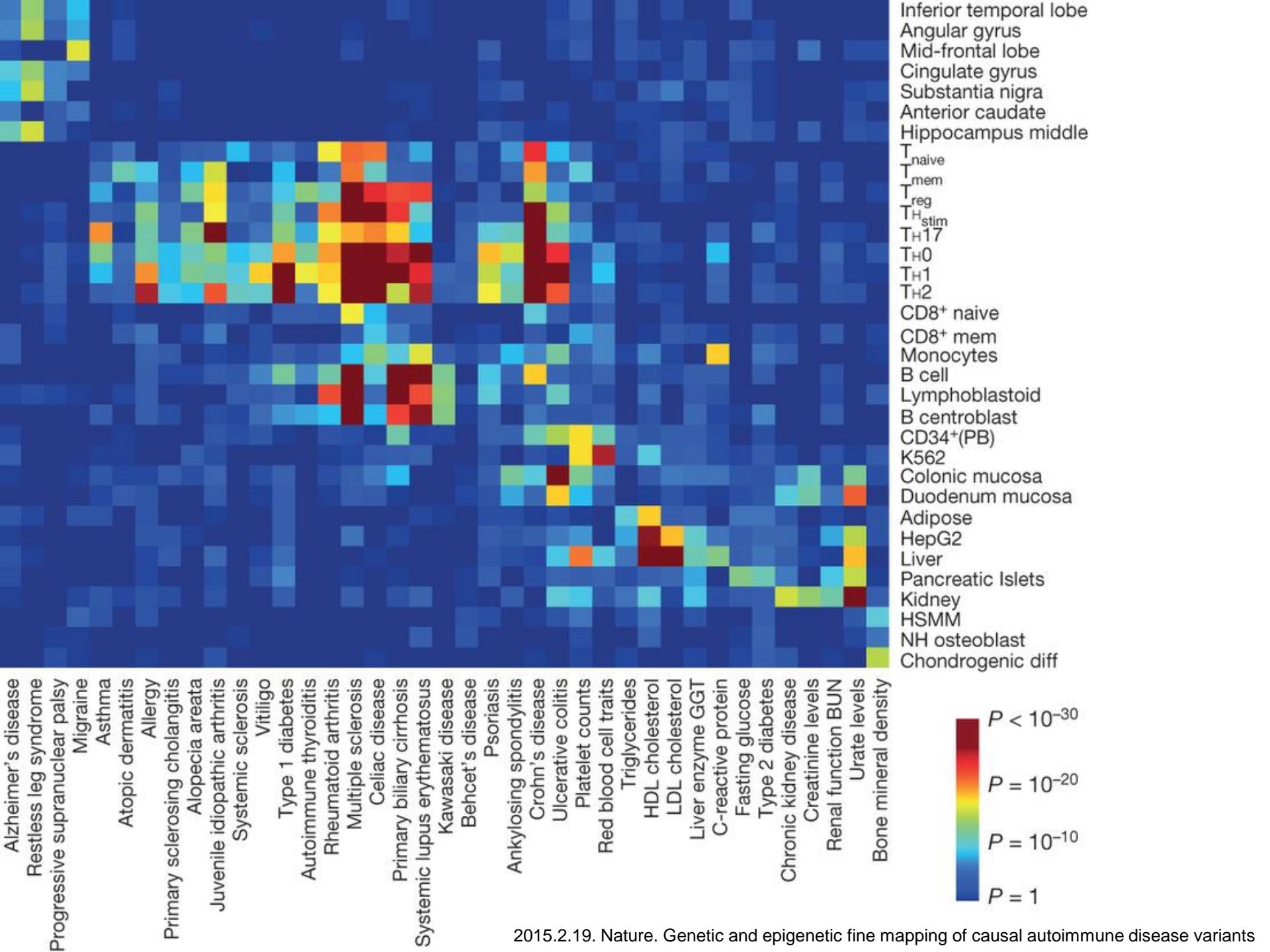
111 Reference Human Epigenomes



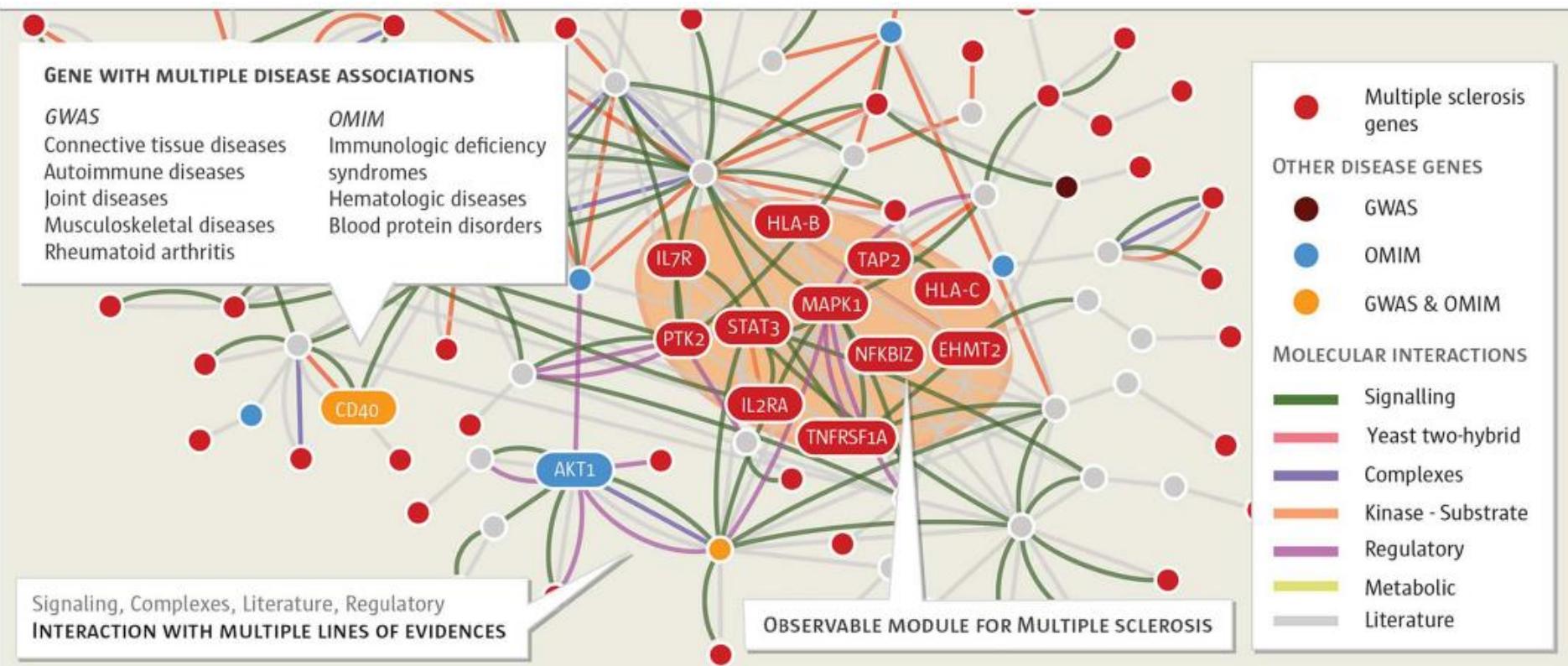


Data Dimensions



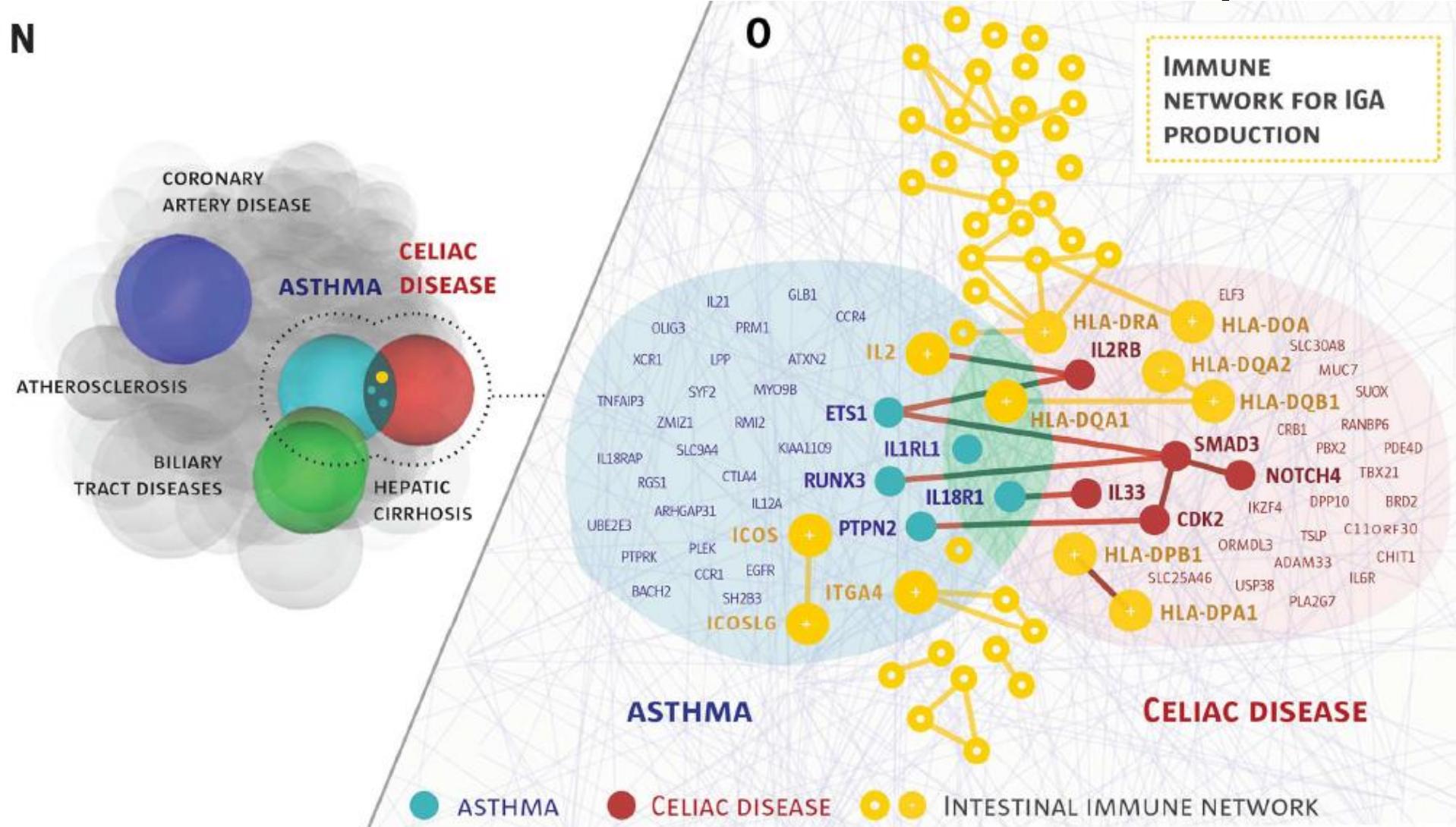


Building the Interactome



Network-based Model of Disease-disease Relationship

N



Genomic Study

Genome-wide Profiling

Transcriptomics
Proteomics
Metabolomics

Human Genome(DNA) Study

GWAS, Candidate gene study
Familial study
Linkage study

Novel Target Discovery

Novel Variant(SNP) Discovery

Genomic Medicine

Novel Disease Target

GENE for everyone

Diagnosis

Genetic counseling for rare diseases
Sensitive urine test, DM subtype

Treatment

Monoclonal antibody for osteoporosis

Prevention

Anti-oxidant

Personalized Medicine

VARIANT based individualization

Common Disease Risk

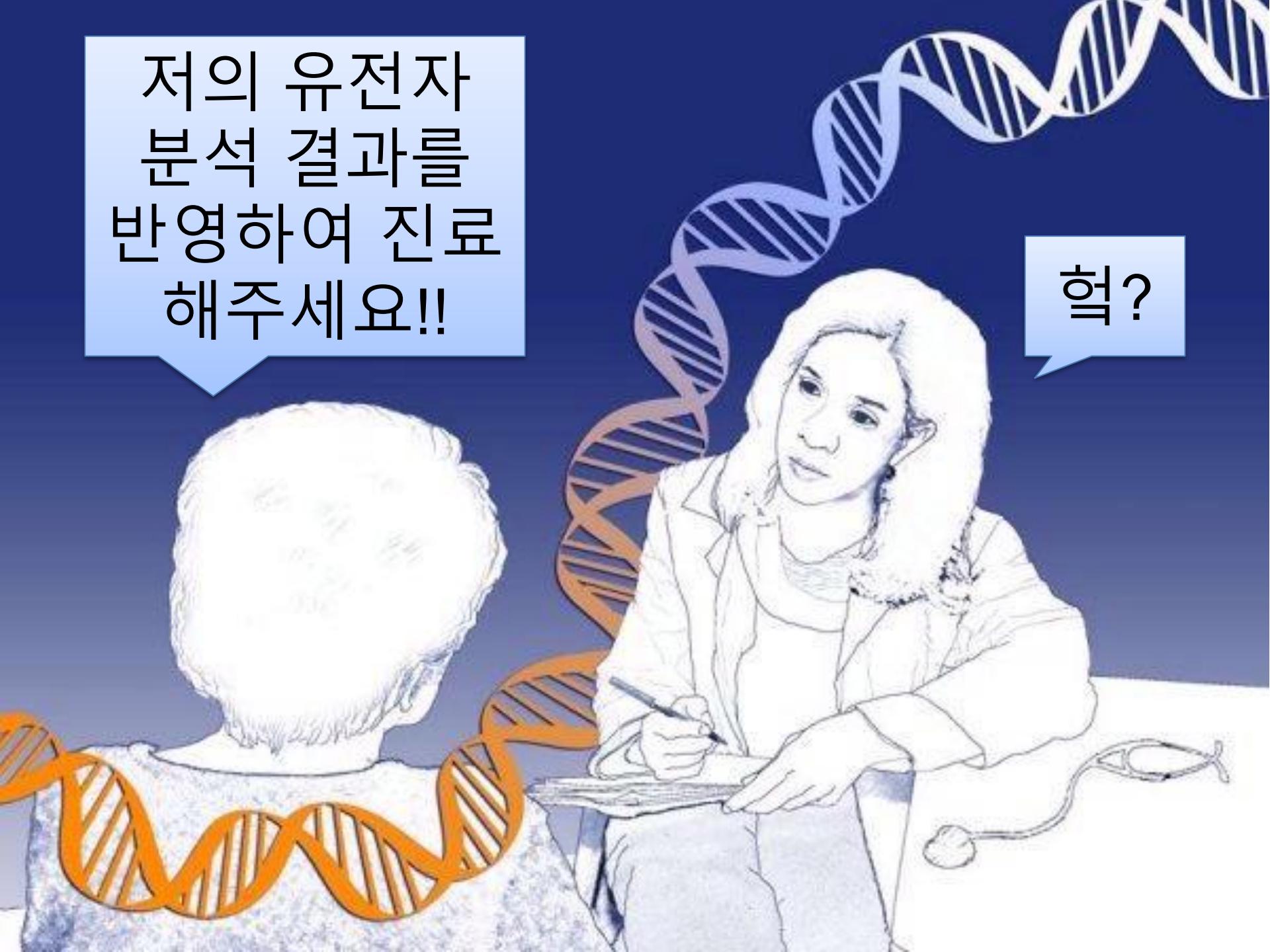
High risk of future osteoporosis
High risk of DM complications

Rare Disease Risk

Mendelian disease risk

Therapeutic Option

Non-responder of treatment
Severe side effect

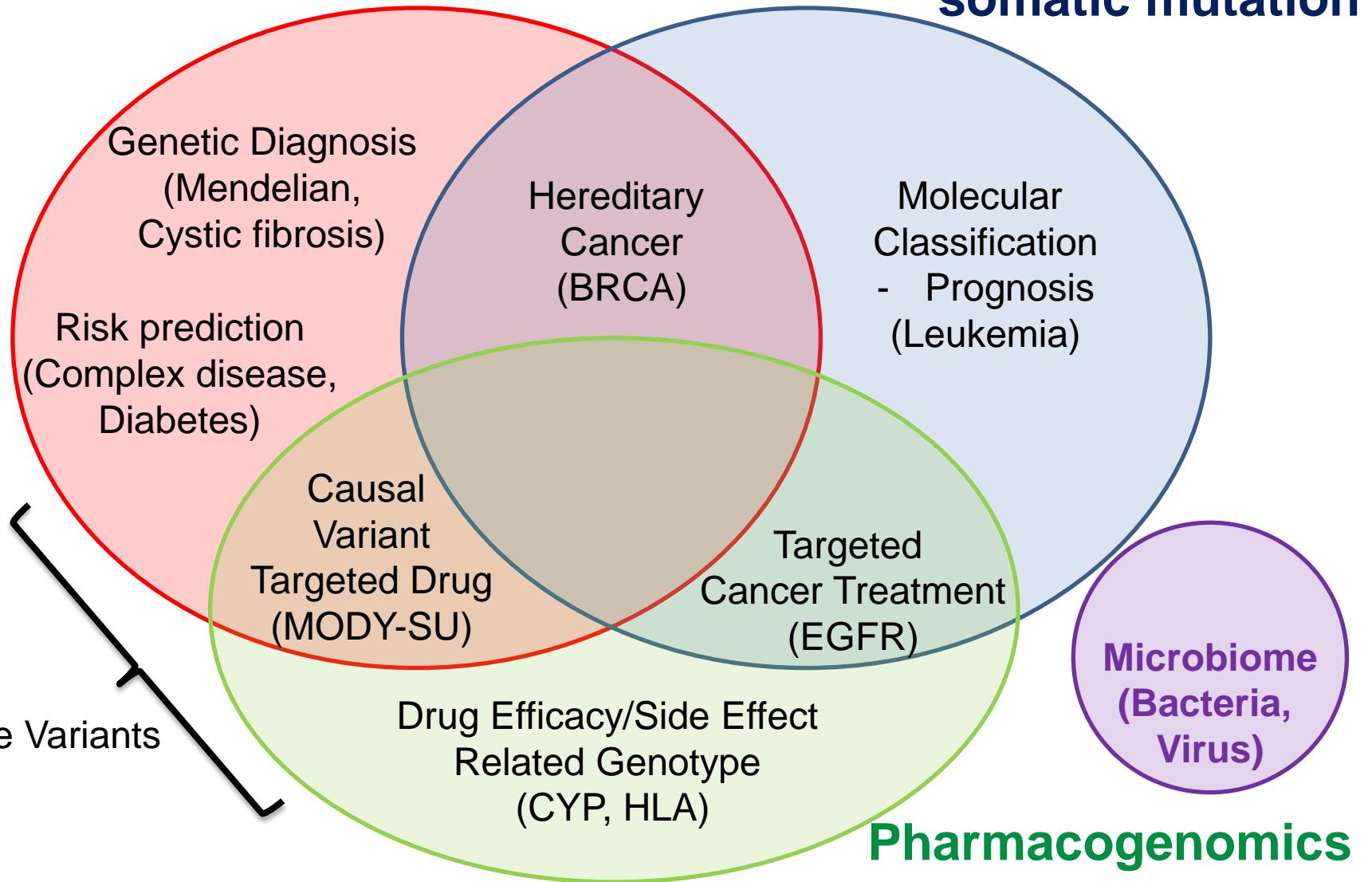


저의 유전자
분석 결과를
반영하여 진료
해주세요!!

헐?

Personalized Medicine

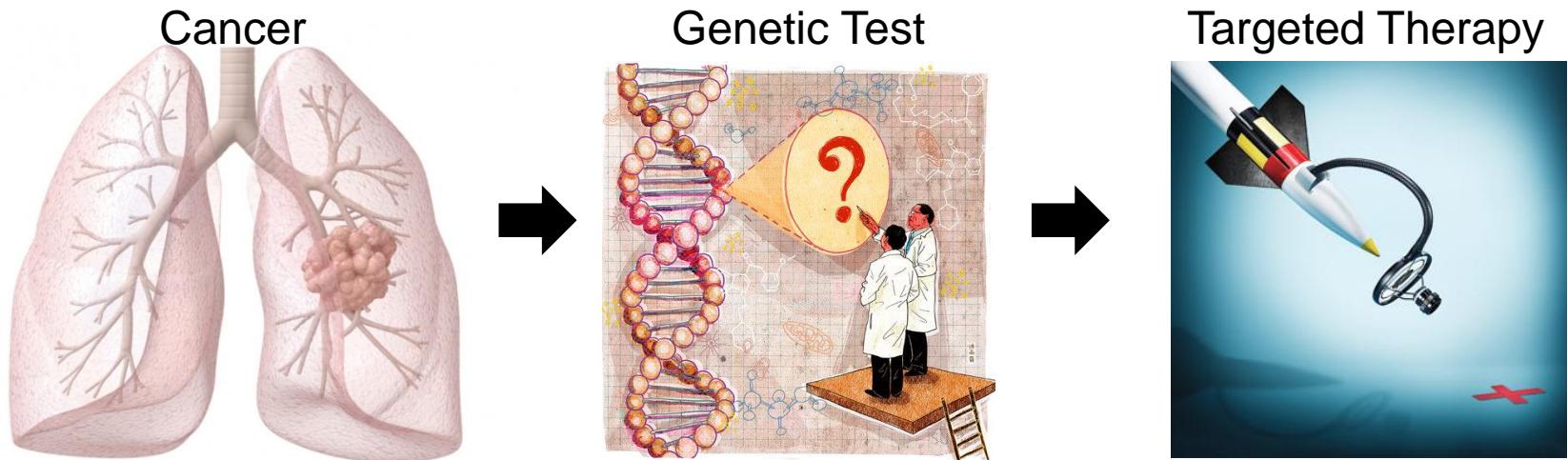
Disease genetic susceptibility



Cancer driver somatic mutation

Pharmacogenomics

Cancer Targeted Therapy



[뉴스](#) ▾[오피니언](#) ▾[경제](#) ▾[스포츠](#) ▾[연예](#) ▾[라이프](#) ▾[사회](#) ▾

의료 · 보건

[오늘의 세상] 삼성서울병원 "癌치료의 삼성전자 될 것"

김철중 의학전문기자 ▾

**MedicalTimes**

“유전체를 기반으로 한
맞춤형 항암치료를
5년 내 시작하겠다”

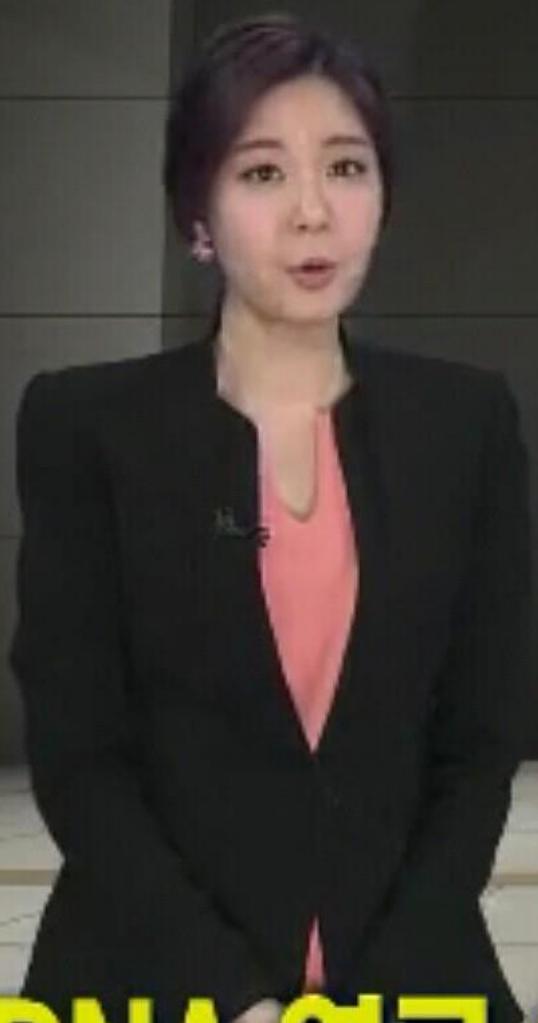
“5년 안에 모든 암환자
맞춤치료 실현하겠다”

2013.06.24



Germline Genomic Medicine

- Disease Susceptibility
 - Common Complex Disease
 - Rare Mendelian Disease
- Pharmacogenomics

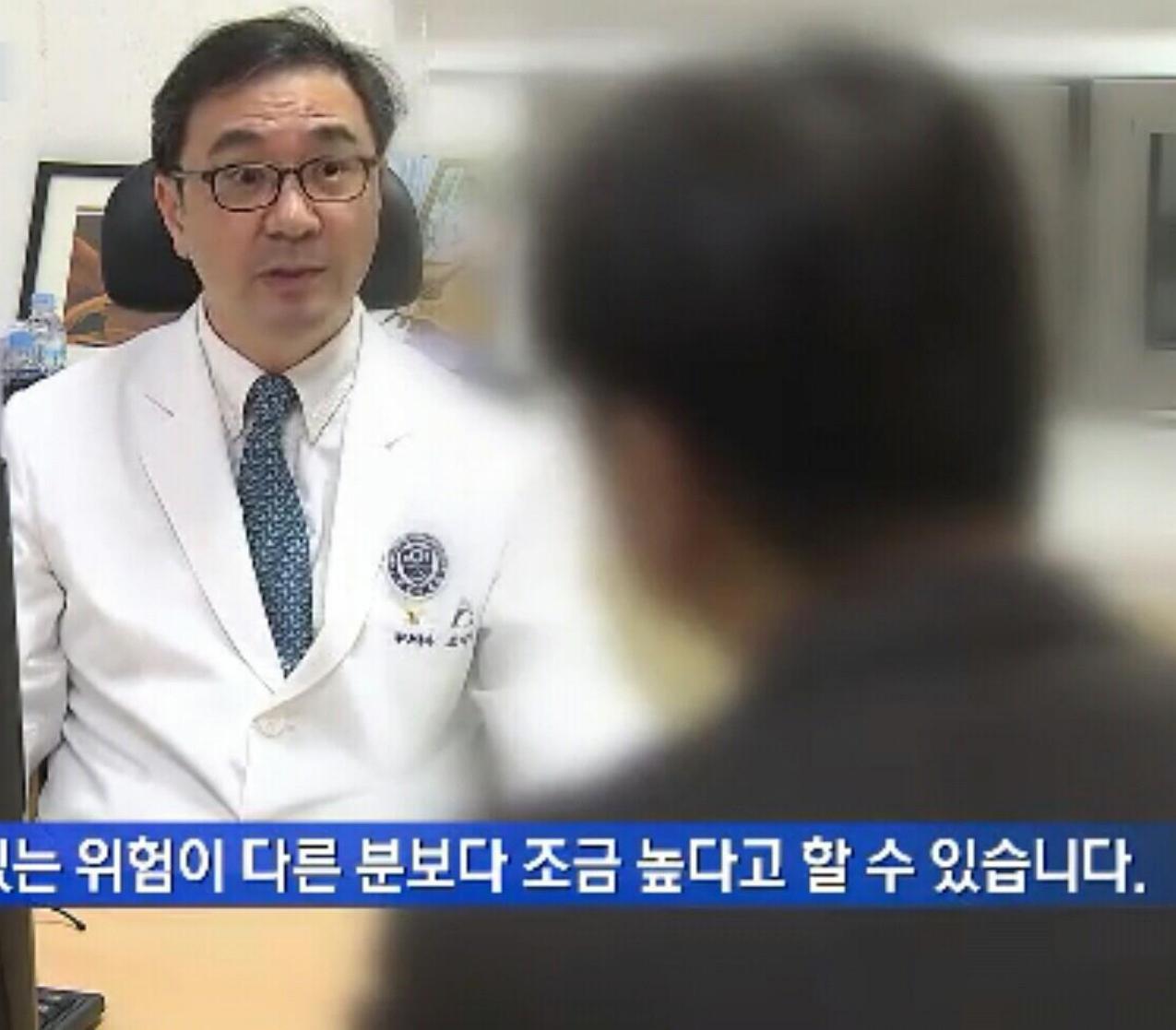


DNA 연구 60년…의술 혁명 예고



심층취재

DNA로 암 정복



암이 생길 수 있는 위험이 다른 분보다 조금 높다고 할 수 있습니다.

며칠 전 유전자 검사를 받은 40대 남성입니다.

혈액세포의 DNA 상태를 분석해 앞으로 암에 걸릴 위험이 있는지 여부를 판단할 수 있다고 합니다.



DNA로 암 정복

염색체 번호 유전자 이름

난소암

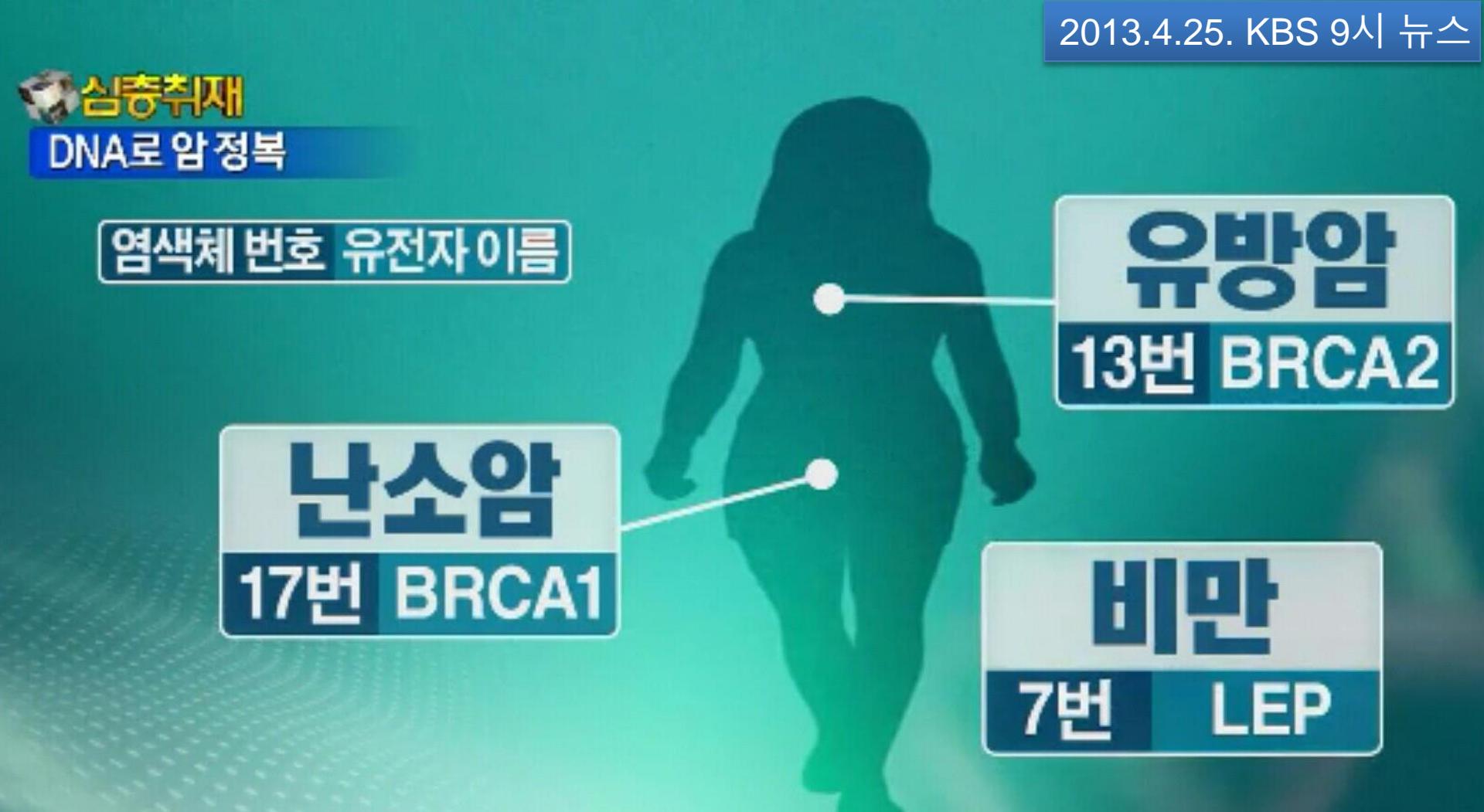
17번 BRCA1

유방암

13번 BRCA2

비만

7번 LEP



60년전 DNA의 구조가 밝혀진 이래 2003년 인간 유전자 지도가 완성됐고, 현재는 어떤 유전자가 어떤 질병을 일으키는지 분석도 80% 정도 끝난 상태입니다. 예를 들어 13번 염색체의 BRCA2 유전자에 이상이 생기면 유방암에 걸릴 확률이 높습니다. 또 17번 염색체 유전자는 난소암, 7번 염색체 유전자는 비만을 일으킵니다.



유한양행

헬로진 개인 유전체 분석 서비스 HELLO GENE



	암	일반 질환	신체적 특징	식품 및 약물 반응	희귀 질환 보인자 검사
헬로진 암	남성 12종				
	여성 14종				
헬로진 일반 질환		14종	2종		
헬로진 플래티넘 라이트	남성 12종	남성 12종			
	여성 13종	여성 11종			
헬로진 플래티넘	남성 13종	21종	6종	4종	
	여성 15종				
헬로진 노블	남성 13종	30종	10종	5종	40종
	여성 15종				

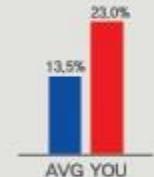
질환명	평균 유병률	유전적 위험도 (Odds Ratio)
1위 아토피	1,345 /10,000명	당신은 평균 인구보다 아토피에 대한 위험도가 1.7배 (2,300/10,000명) 높습니다.
2위 양극성 장애	129 /10,000명	당신은 평균 인구보다 양극성 장애에 대한 위험도가 1.4배 (182/10,000) 높습니다.
3위 우울증	129 /10,000명	당신은 평균 인구보다 우울증에 대한 위험도가 1.7배 (155/10,000명) 높습니다.

갑상선암 (Thyroid Cancer) 고위험군

당신의 질병 위험도

통합 위험도 **23%** (100명 당 23명)

일반인 평균 위험도 **13.5%** (100명 당 13.5명)



유전자형 분석 결과

유전자	유전 마커	질병 유전자	당신의 유전형	유전적 위험도
CH3L1	rs123	A	AG	1.71

30만원-200만원



23andMe

Variant Present

Disease Risks (100)

↑ Elevated Risks

Gallstones

Your Risk

11.1%

Average Risk

7.0%

Restless Legs Syndrome

2.5%

2.0%

[more »](#)

↓ Decreased Risks

Prostate Cancer

Your Risk

12.7%

Average Risk

17.8%

Alzheimer's Disease

4.9%

7.2%

Colorectal Cancer

4.2%

5.6%

[more »](#)[See all 100 risk reports...](#)

Carrier Status (24)

Hemochromatosis

Variant Present

Alpha-1 Antitrypsin Deficiency

Variant Absent

Bloom's Syndrome

Variant Absent

BRCA Cancer Mutations (Selected)

Variant Absent

Canavan Disease

Variant Absent

Cystic Fibrosis

Variant Absent

Familial Dysautonomia

Variant Absent

Factor XI Deficiency

Variant Absent

[See all 24 carrier status...](#)

Traits (50)

Alcohol Flush Reaction

Does Not Flush

Bitter Taste Perception

Can Taste

Earwax Type

Wet

Eye Color

Likely Brown

Hair Curl

Slightly Curlier Hair on Average

[See all 50 traits...](#)

Drug Response (19)

Warfarin (Coumadin®) Sensitivity

Increased

Abacavir Hypersensitivity

Typical

Alcohol Consumption, Smoking and Risk of Esophageal Cancer

Typical

Clopidogrel (Plavix®) Efficacy

Typical

Fluorouracil Toxicity

Typical

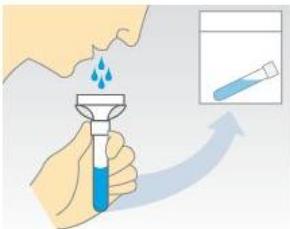
[See all 19 drug response...](#)



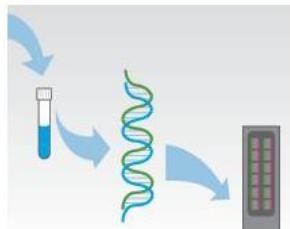
Here's what you do:



1. Order a kit from our [online store](#).



2. Register your kit, spit into the tube, and send it to the lab.



3. Our CLIA-certified lab analyzes your DNA in 6-8 weeks.



4. Log in and start exploring your genome.

**Living well starts with
knowing your DNA.**

[order now](#)

Our genes make us who we are, so naturally they impact our health. By knowing your DNA, you can take steps toward living a healthier life.



\$99



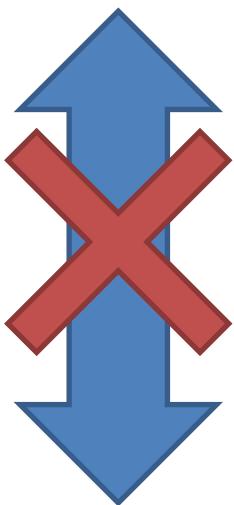
TV CF



(출처: [금창원 대표님 블로그](#))



23andMe



Navigenics

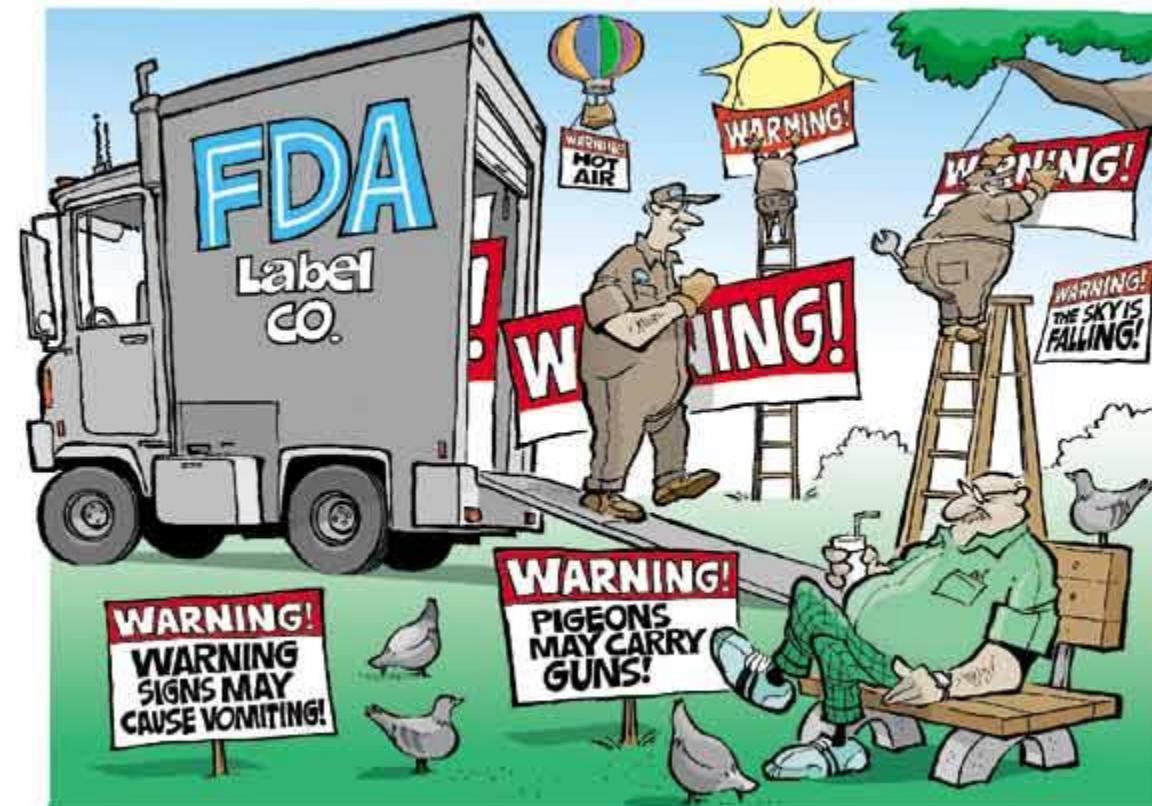
TABLE 1: PREDICTIONS FOR DISEASE RELATIVE RISKS FOR FIVE INDIVIDUALS

Disease	Female A	Female B	Female C	Male D	Male E
Breast cancer	↑↑	↑↑	↓↓		
Coeliac disease	↓↓	↓↓	↓↓	↓↓	↓↓
Colon cancer	==	==	=↓	↑↑	=↓
Crohn's disease	↓↑	↓↑	↓↓	↓↓	↓=
Heart attack	↓↓	=↓	=↓	=↓	↑↑
Lupus	↑↓	↓↓	↓↓	↑=	↑=
Macular degeneration	↓↓	↓↓	↑=	↓↓	↓↓
Multiple sclerosis	↑↑		↓↓	↓↓	↓↓
Prostate cancer				↑↑	↓↑
Psoriasis	↓↑		↑↓	↑↑	↓↓
Restless legs syndrome	=↓	↑↑	↓=	↓↑	↑↑
Rheumatoid arthritis	↑↑	↑↑	↓↓	↓↓	↑↑
Type 2 diabetes	↓↓	=↓	↓↓	↑↓	=↓

↑ increased risk ($RR > 1.05$), ↓ decreased risk (relative risk ($RR < 0.95$)), = average risk ($0.95 \leq RR \leq 1.05$). First prediction is from 23andMe; second prediction is from Navigenics. Different predictions are highlighted in beige.

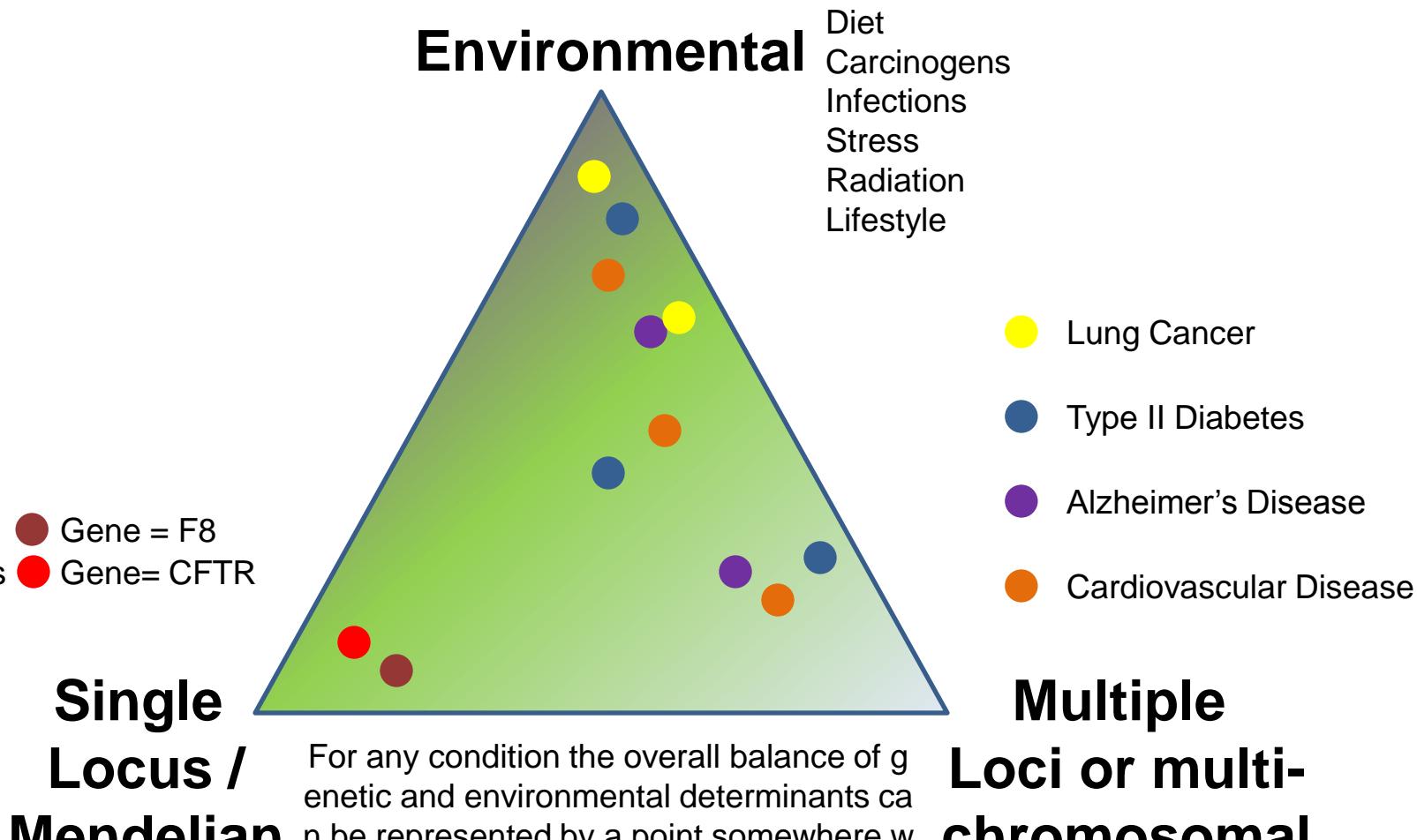
FDA Halt 23andMe

2013.11.22.



→ 2015.2.20. FDA Bloom Syndrome mutation 검사 23andMe 승인

Influence of Genetics on Human Disease



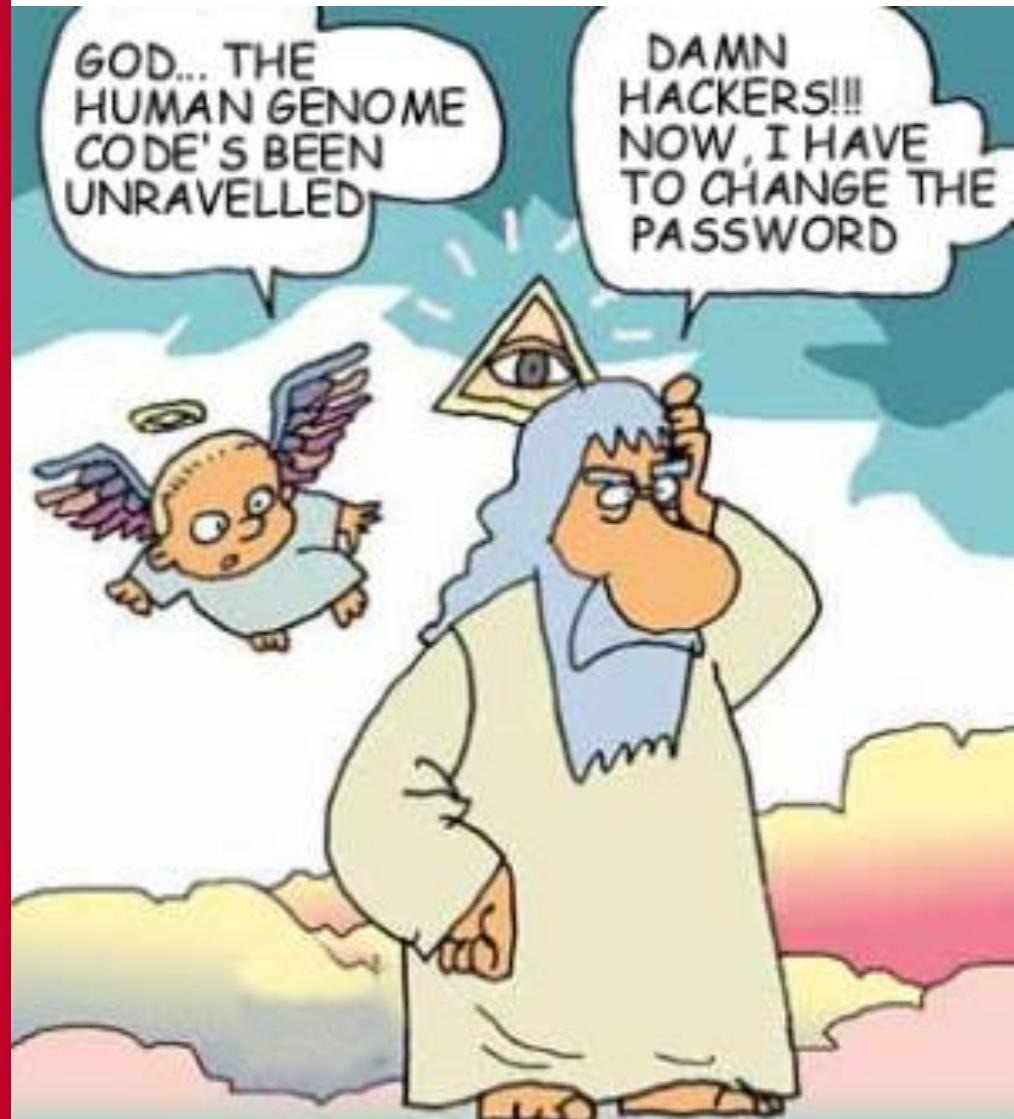
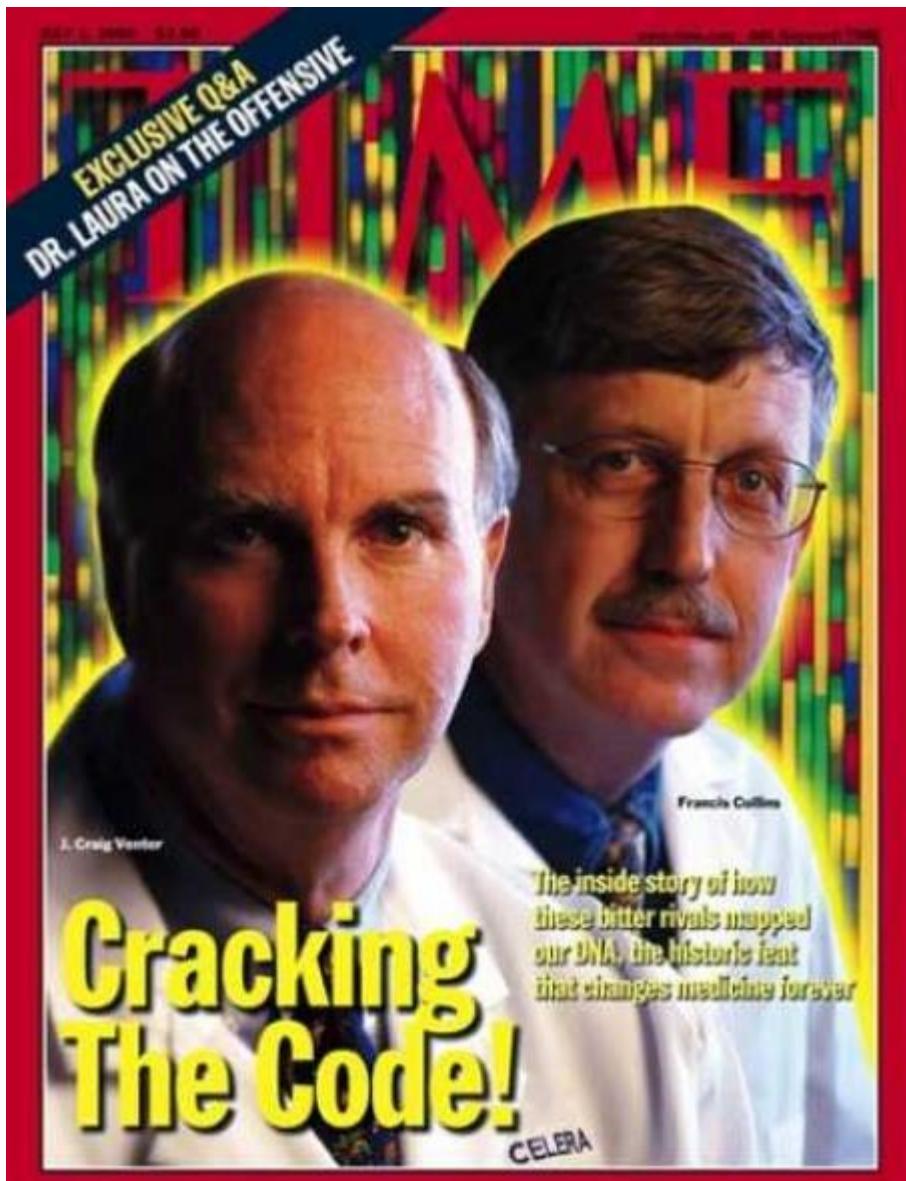
F8 = Coagulation Factor VIII

CFTR = Cystic Fibrosis Conductance Transmembrane Regulator

Germline Genomic Medicine

- Disease Susceptibility
 - Common Complex Disease
 - Rare Mendelian Disease
- Pharmacogenomics

Promise of Human Genome Project

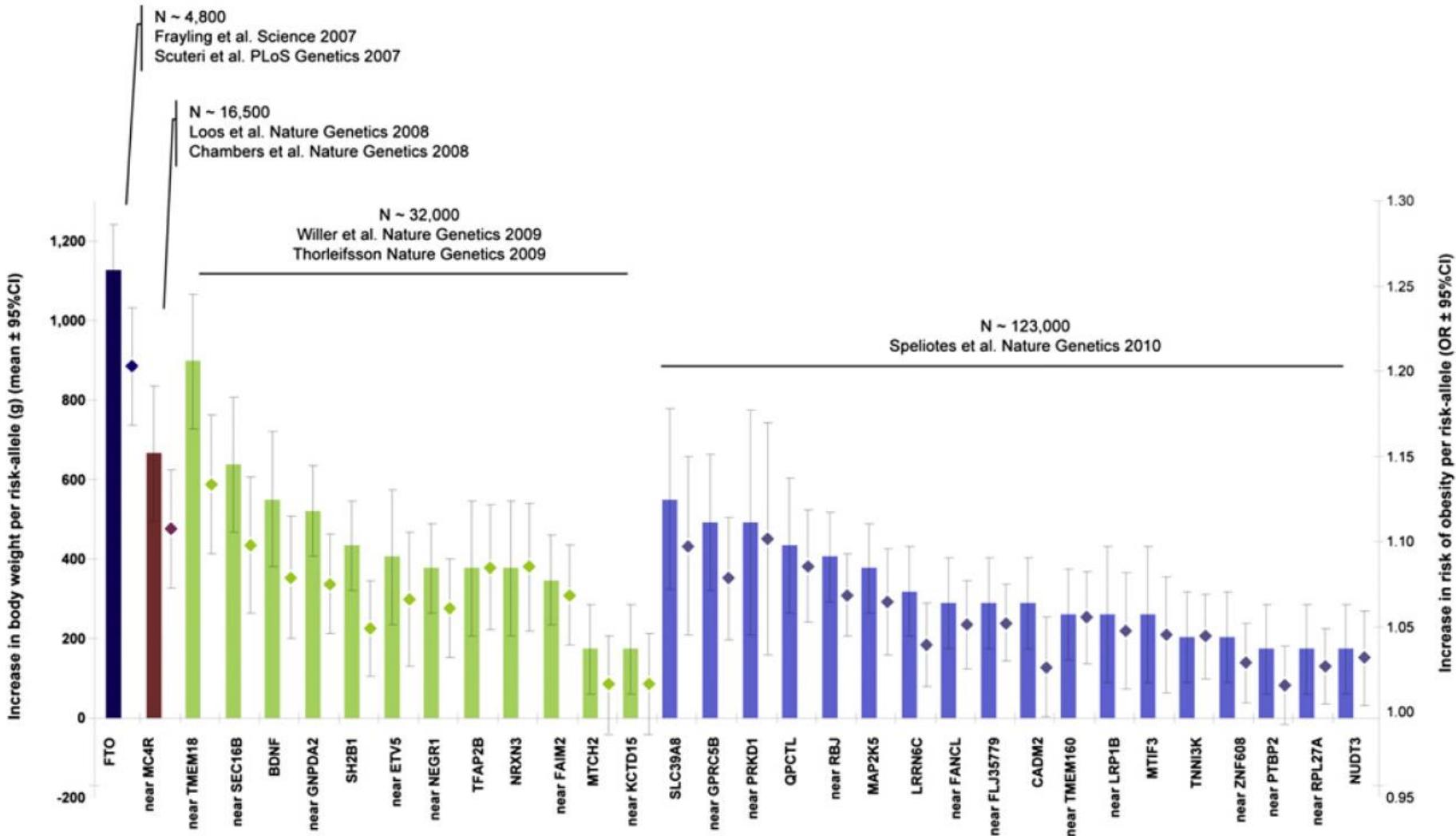


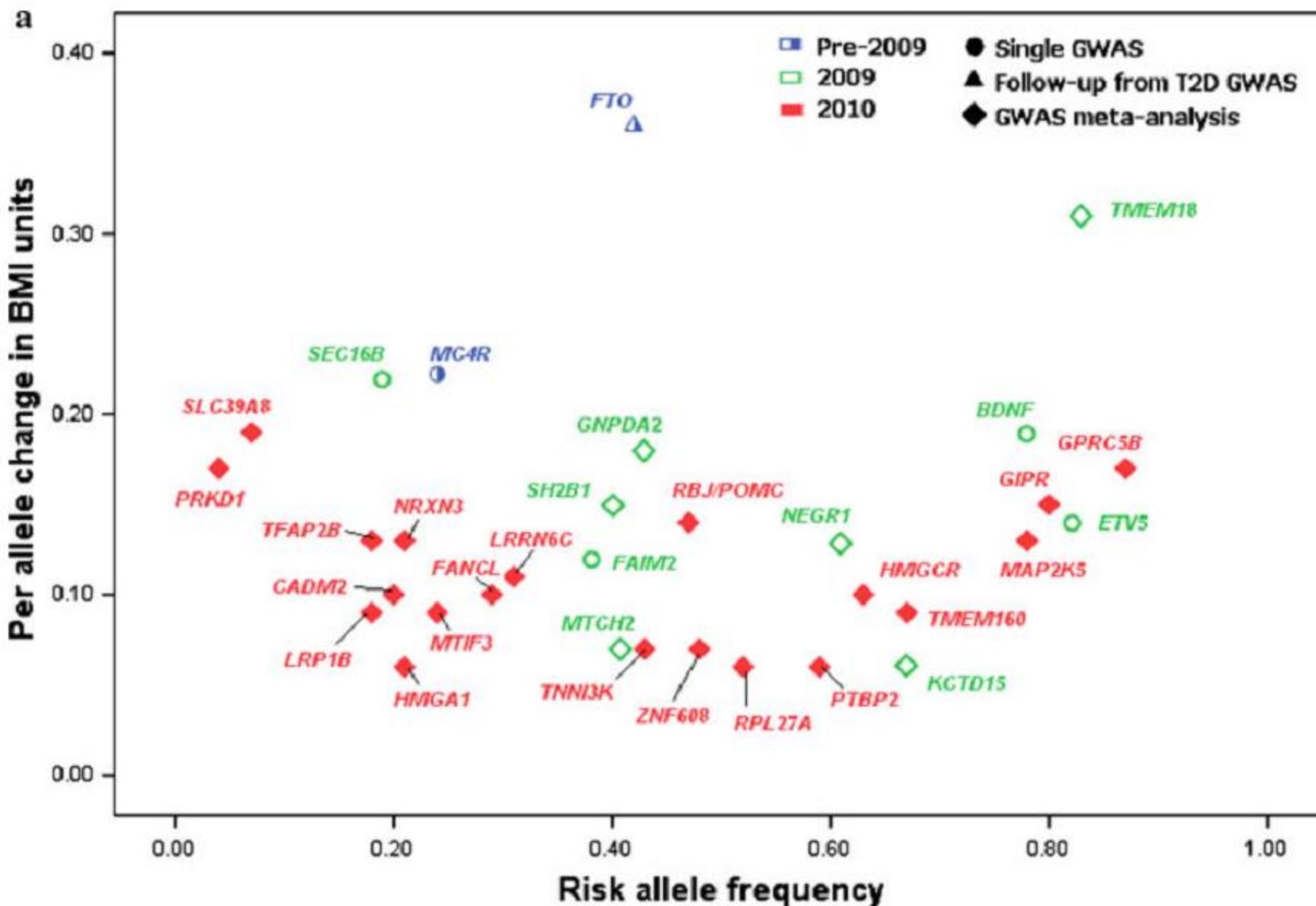


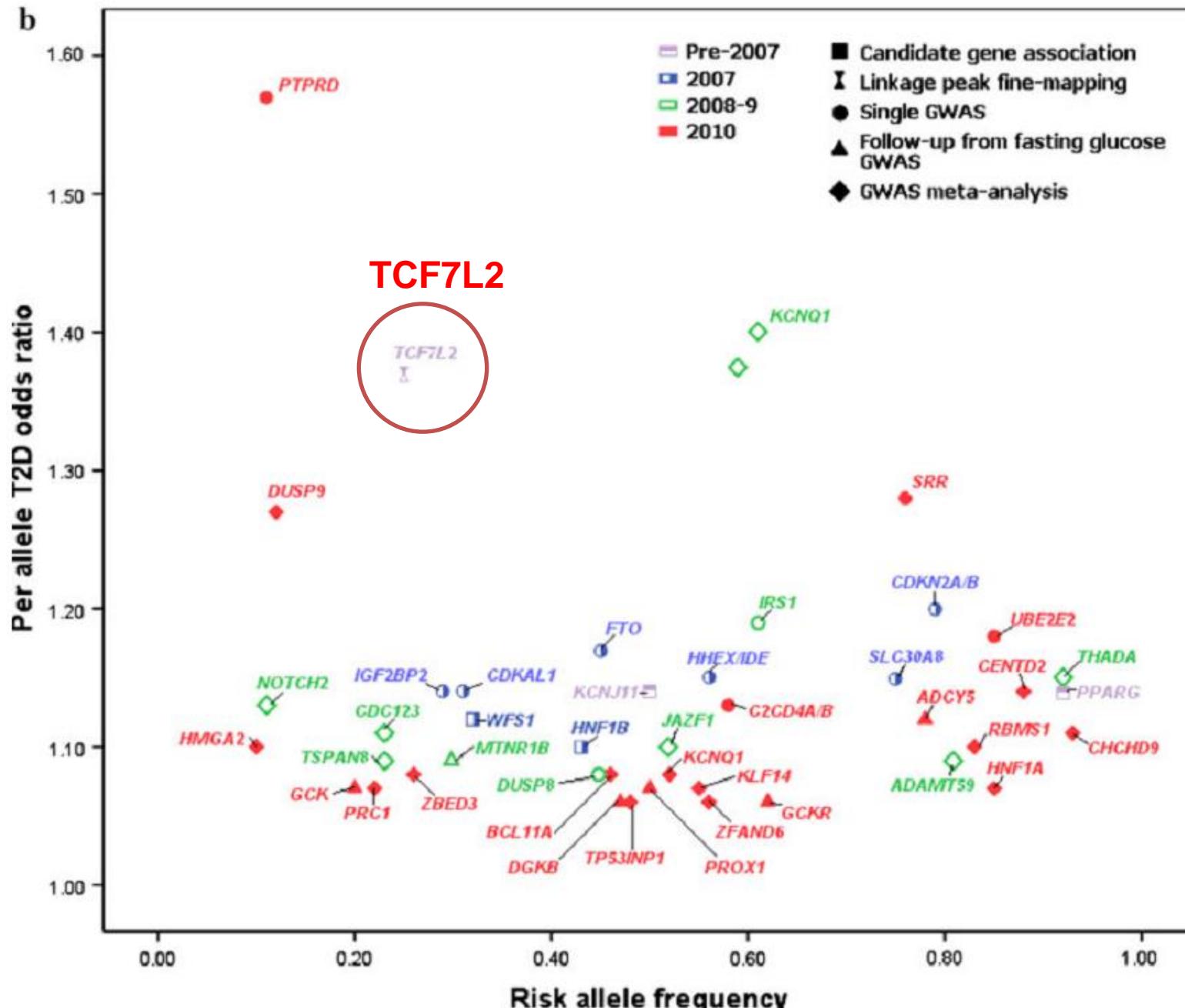
The case of the missing heritability

When scientists opened up the human genome, they expected to find the genetic components of common traits and diseases. But they were nowhere to be seen. **Brendan Maher** shines a light on six places where the missing loot could be stashed away.

Per-allele effect of BMI-associated loci on body weight



a

b

Jason L. Vassy,^{1,2,3} Marie-France Hivert,^{1,4,5} Bianca Porneala,⁶ Marco Dauriz,^{1,6,7} Jose C. Florez,^{1,8,9} Josée Dupuis,^{10,11} David S. Siscovick,¹² Myriam Fornage,¹³ Laura J. Rasmussen-Torvik,¹⁴ Claude Bouchard,¹⁵ and James B. Meigs^{1,6}



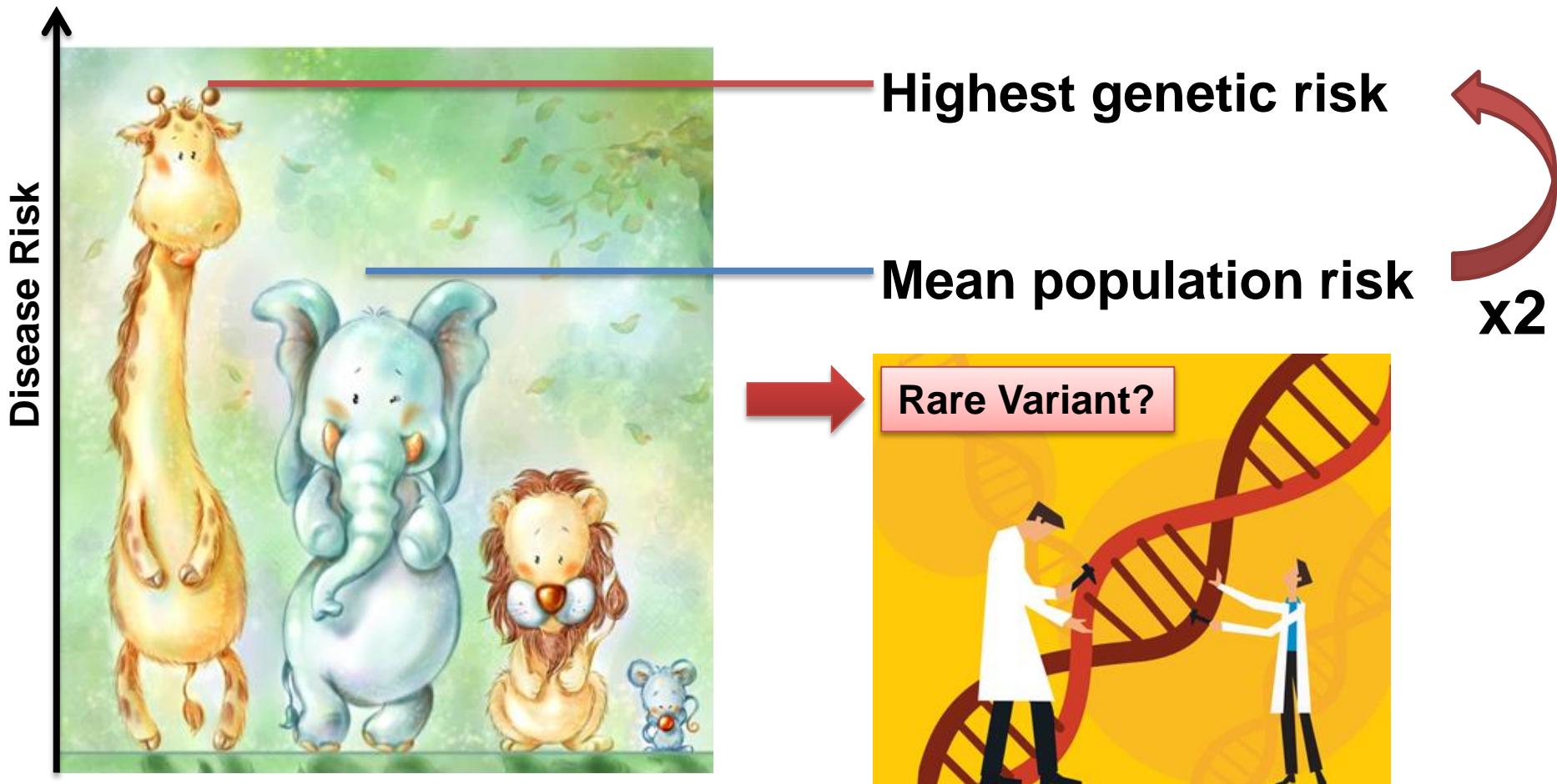
Polygenic Type 2 Diabetes Prediction at the Limit of Common Variant Detection

Diabetes 2014;63:2172–2182 | DOI: 10.2337/db13-1663

“At the end of the era of common variant discovery for T2D, polygenic scores can predict T2D in whites and blacks but do not outperform clinical models.

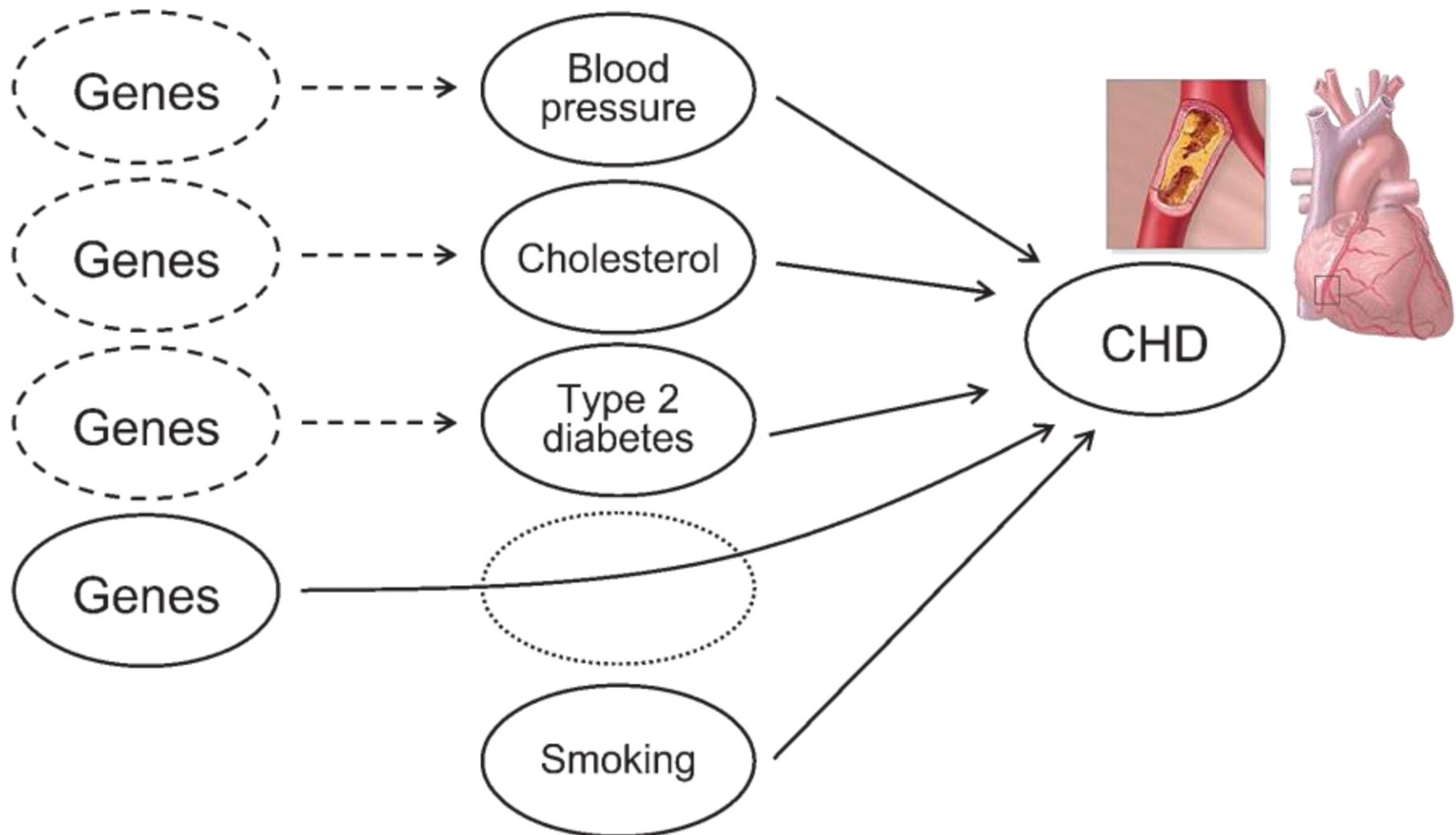
Further optimization of polygenic prediction may require novel analytic methods, including less common as well as functional variants.”

Predicting Complex Diseases

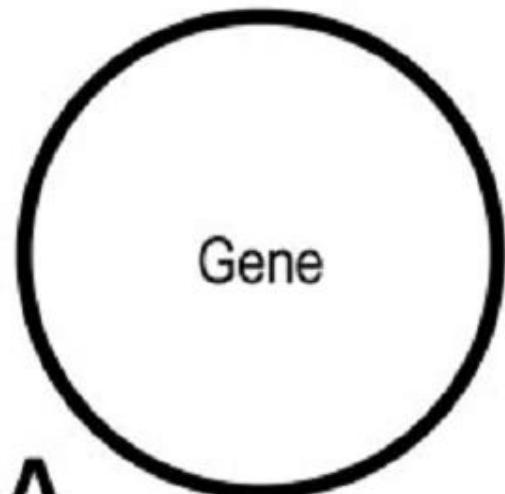


"For most diseases, it should be possible to identify the individuals with the highest genetic risk. However, if the aim is to identify individuals with just twice the mean population risk, we cannot currently do that with SNPs"

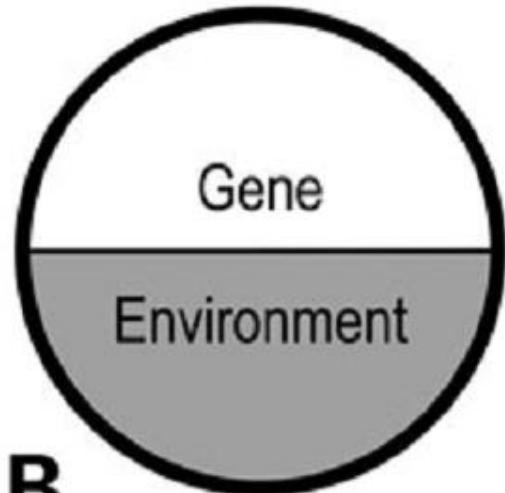
Gene-disease pathways



Monogenic disorders

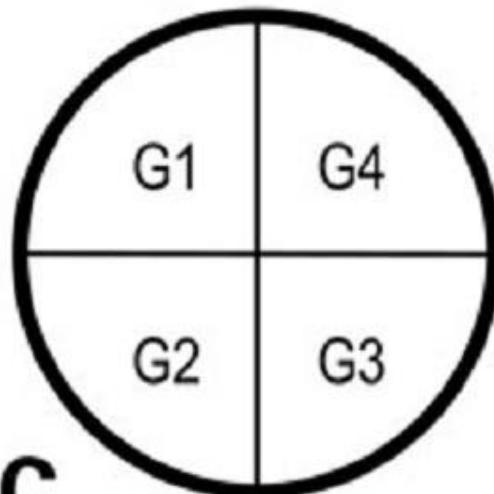


A

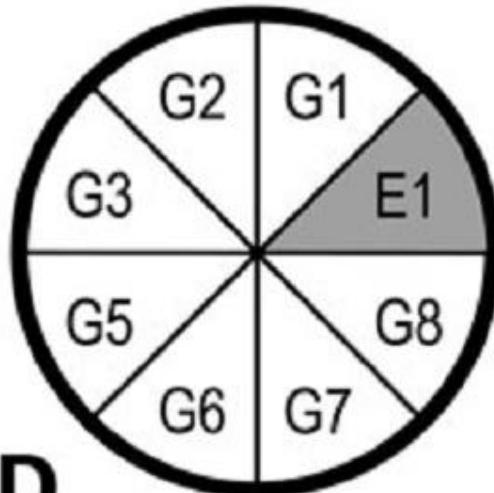


B

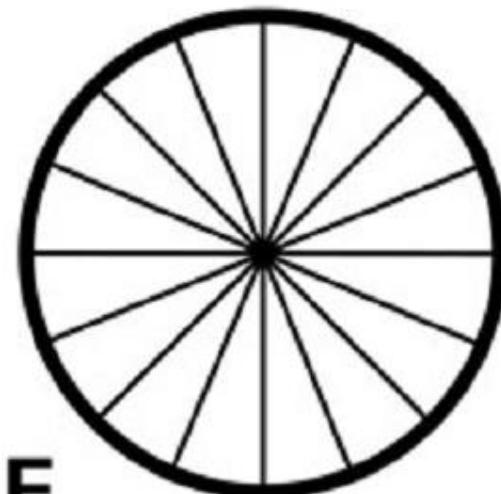
Complex diseases



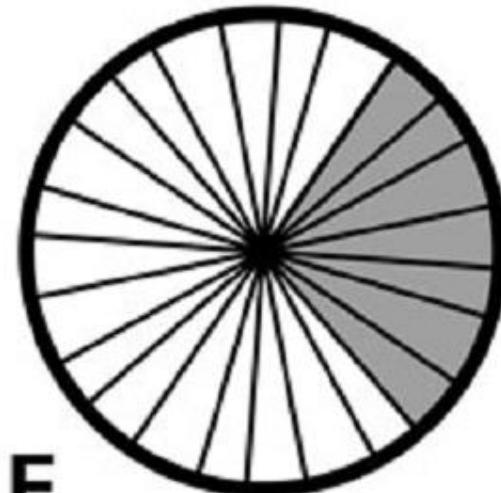
C



D



E



F

Diabetes ≠ Genetic Disease?

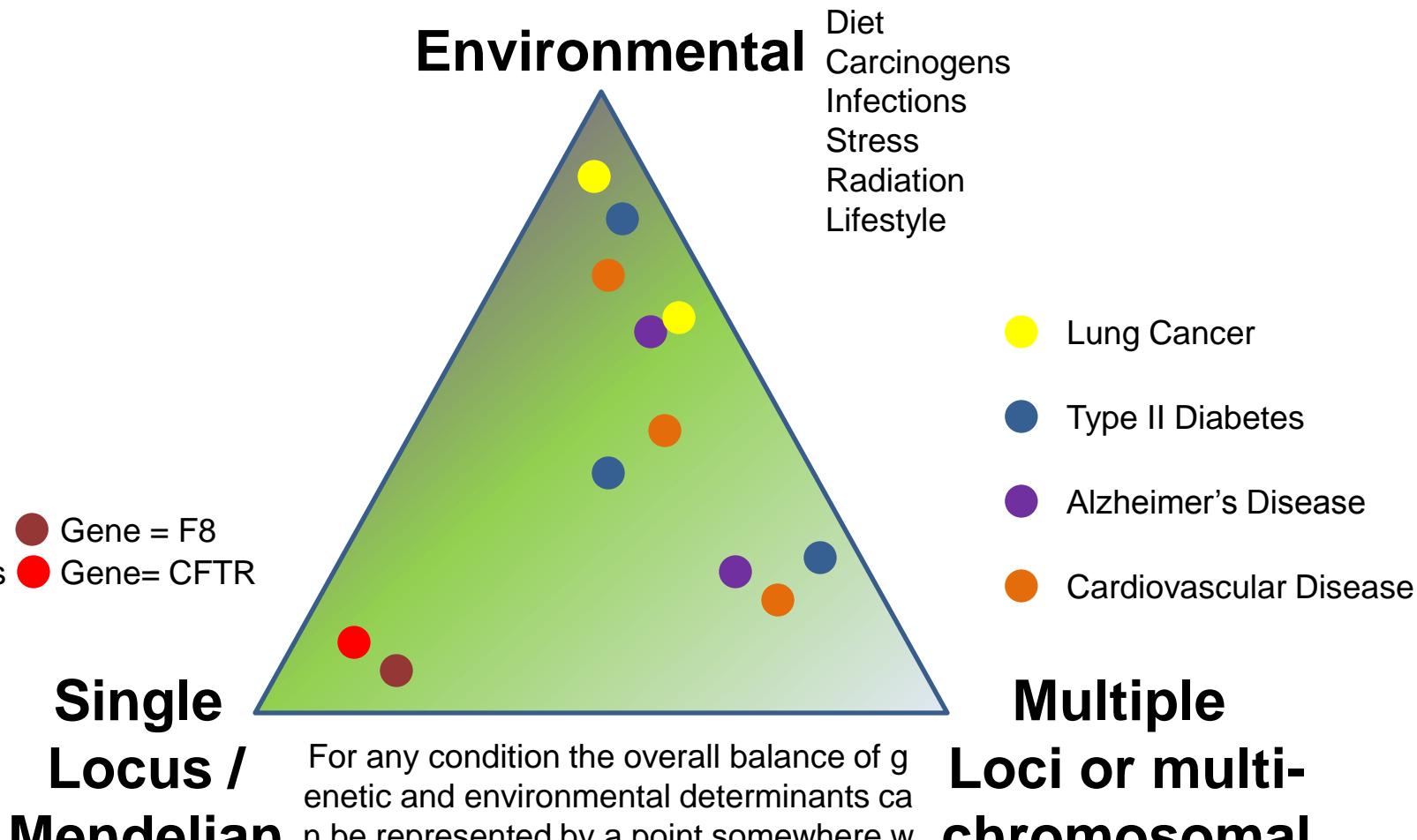
- Familial aggregation
 - Genetic influences?
 - Epigenetic influences
 - Intrauterine environment
 - Shared family environment?
 - Socioeconomic status
 - Dietary preferences
 - Food availability
 - Gut microbiome content
- Overestimated heritability
 - Phantom heritability



Germline Genomic Medicine

- Disease Susceptibility
 - Common Complex Disease
 - Rare Mendelian Disease
- Pharmacogenomics

Influence of Genetics on Human Disease

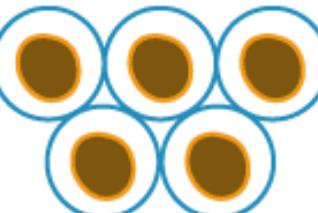


F8 = Coagulation Factor VIII

CFTR = Cystic Fibrosis Conductance Transmembrane Regulator



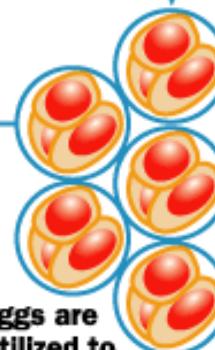
Egg donor is given fertility drugs



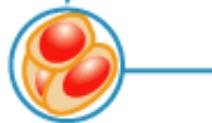
Multiple eggs are produced



Embryos analyzed for genetic defects



Eggs are fertilized to produce embryos



Only healthy embryos are injected into uterus

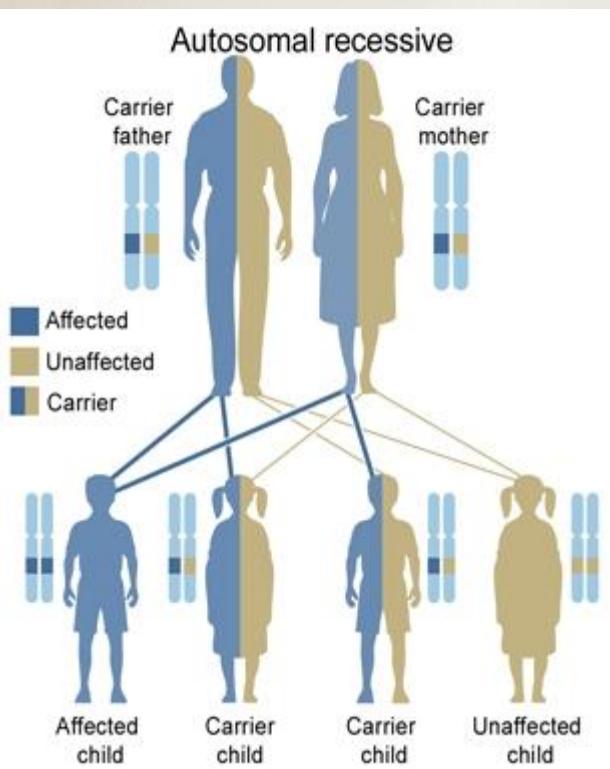


Mother gives birth to genetically healthy baby

The first step to a healthy pregnancy

Babies with genetic disease are often born to healthy parents.

The Counsyl Test gives you insight to a healthy future.



Laboratory Director
강현석



REVIEW ARTICLE

Elizabeth G. Phimister, Ph.D., *Editor*

Diagnostic Clinical Genome and Exome Sequencing

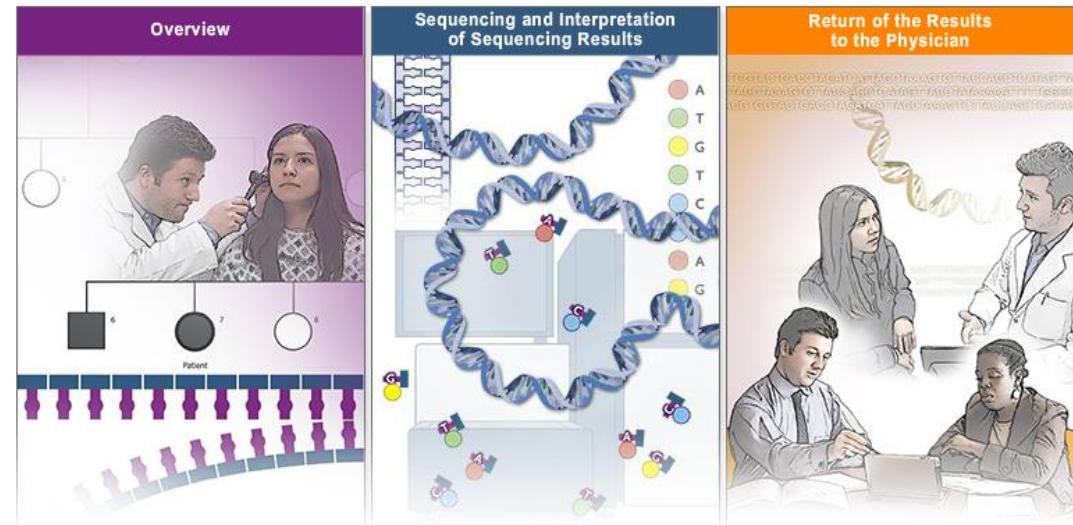
Leslie G. Biesecker, M.D., and Robert C. Green, M.D., M.P.H.

INTERACTIVE GRAPHIC | Genome and Exome Sequencing in Clinical Practice

Introduction

Mendelian (single-gene) genetic disorder

- Known single-gene candidates testing
- Whole Genome or Whole Exome Sequencing



CLINICAL DECISIONS
INTERACTIVE AT NEJM.ORG

Screening an Asymptomatic Person for Genetic Risk

OPTION 1

**Recommend Sequencing of
Cancer Genes Only, if Certain
Conditions Are Met**

Wylie Burke, M.D., Ph.D.

OPTION 2

**Recommend Whole-Genome
Sequencing**

David Dimmock, M.B., B.S.



VIEWPOINT

Returning Pleiotropic Results From Genetic Testing to Patients and Research Participants

JAMA February 26, 2014 Volume 311, Number 8



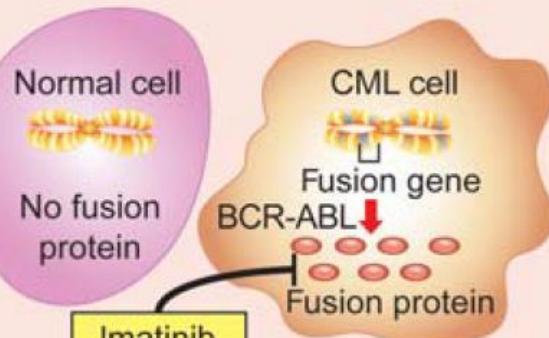
Germline Genomic Medicine

- Disease Susceptibility
 - Common Complex Disease
 - Rare Mendelian Disease
- Pharmacogenomics

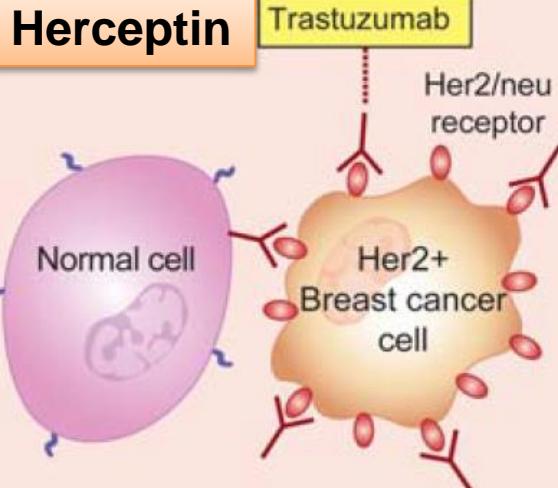
A

Testing for disease-specific or -enriched drug targets

Disease-specific drug target

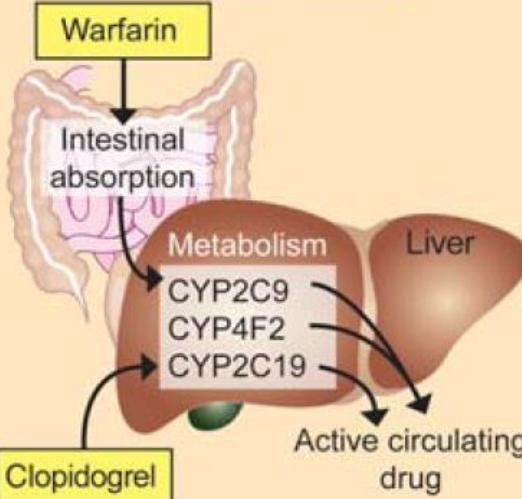


Disease-enriched drug target

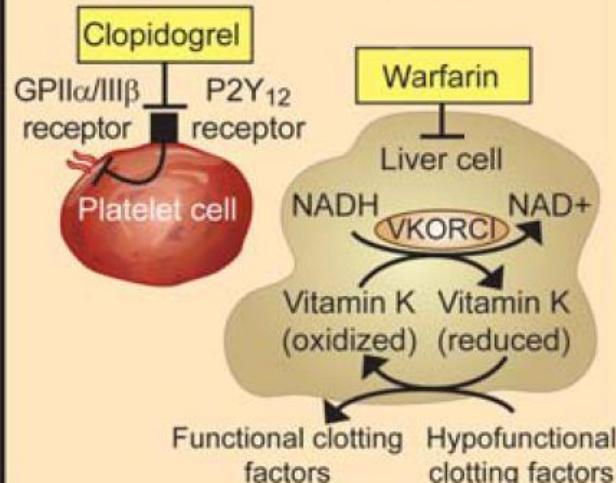
**B**

Estimating drug efficacy and dosing

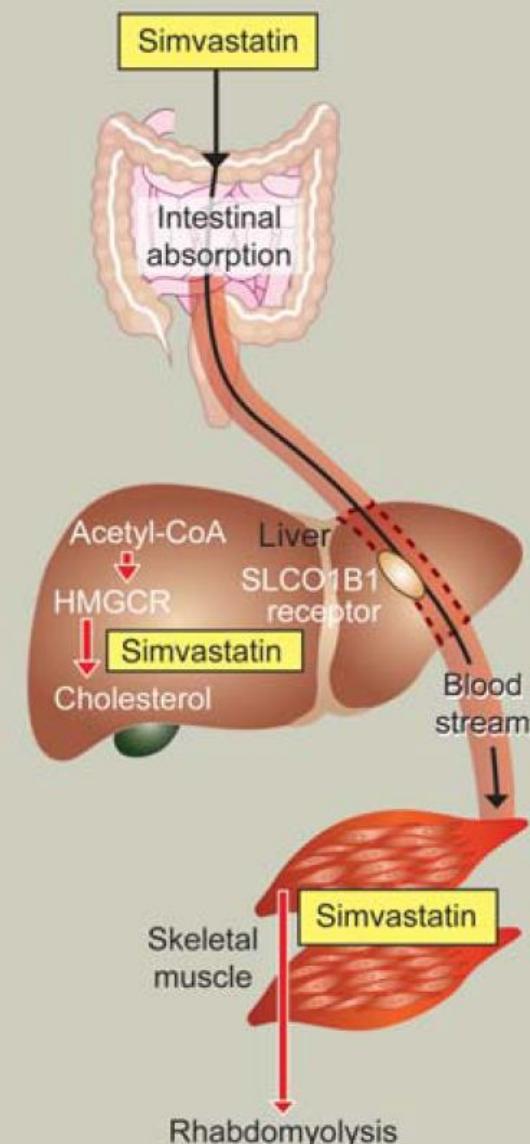
Pharmacokinetics



Pharmacodynamics

**C**

Predicting and avoiding side effects





THE ENDOCRINE SOCIETY'S 96th ANNUAL MEETING & EXPO

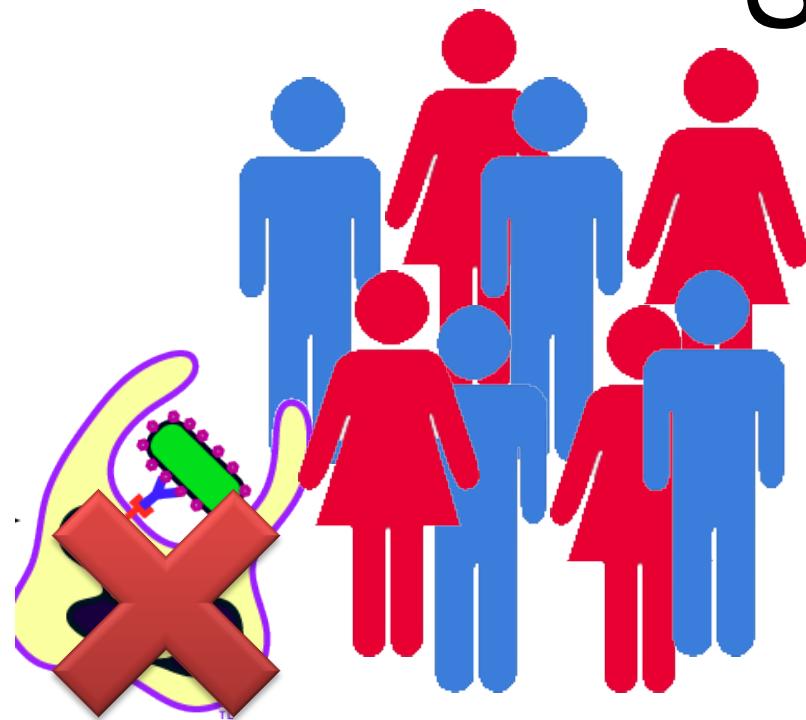
Whole-Exome Sequencing Reveals Strong Genetic Predictors of Antithyroid Drugs Induced Agranulocytosis

Sin-Gi Park¹, Jong Bhak¹, Jong-Soo Kim¹, Tae-Hyung Kim¹, Jee Hyun An², Sang Youl Rhee³,
Hee Kyung Kim⁴, Min Joo Kim⁵, Hyun Jeong Jeon⁶, Taekeun Oh⁶, Hyung Jin Choi⁶

(1)Theragen BiO Institute, TheragenEtex, Suwon, South Korea, (2)Division of Endocrinology and Metabolism, Department of Internal Medicine, Korea University Anam Hospital, Seoul, South Korea, (3)Department of Endocrinology and Metabolism, Kyung Hee University School of Medicine, Seoul, South Korea, (4)Internal Medicine, Chonnam National University Medical School, Gwangju, South Korea, (5)Center for Thyroid Cancer, Korea Cancer Center Hospital, Seoul, South Korea, (6)Department of Internal Medicine, Chungbuk National University College of Medicine, Cheongju, South Korea

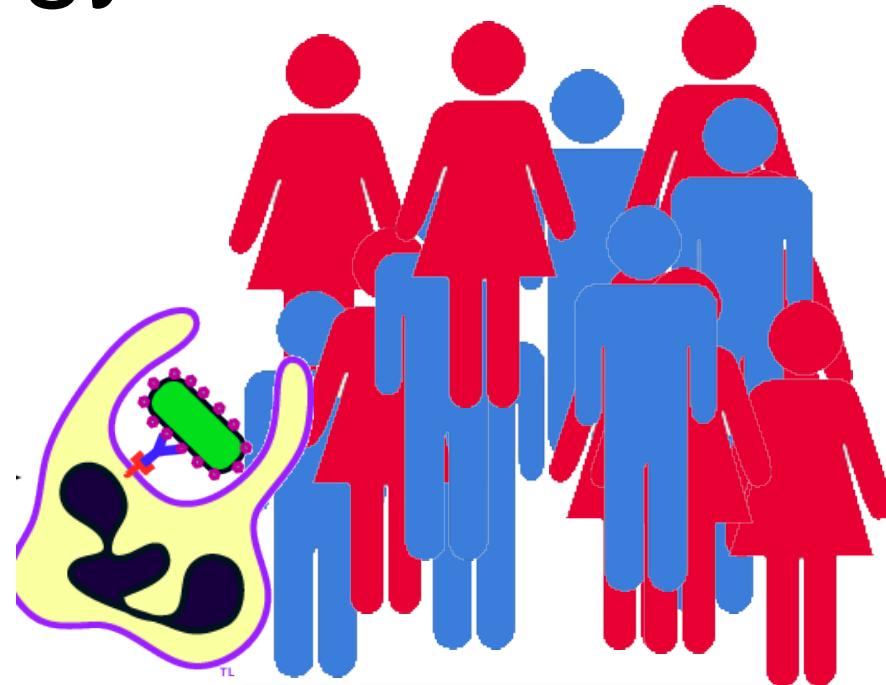


Strategy



Cases
Agranulocytosis

Stage 1
Discovery



Controls
Anti-thyroid drug
No side effect

Whole Exome Sequencing

Case (N=26) vs Control (N=392)

Stage 2
Validation

N=55

(Choi et. al. on going)

CPIC Pharmacogenetic Tests: 최형진

No	Drug (N= 10)	Gene (6 genes=8 biomarkers)	Target SNPs (N=12)	#5 (HJC)	Genotype Interpretation	Clinical Interpretation
1	Clopidogrel	CYP2C19	rs4244285 (G>A)	GG	*1/*1 (EM)	Use standard dose
			rs4986893 (G>A)	GG		
			rs12248560 (C>T)	CC		
2	Warfarin	VKORC1	rs9923231 (C>T)	TT	Low dose (higher risk of bleeding)	Warfarin dose=0.5~2 mg/day
		CYP2C9	rs1799853 (C>T)	CC		
			rs1057910 (A>C)	AC		
3	Simvastatin	SLCO1B1	rs4149056 (T>C)	TT	Normal	
4	Azathioprine (AP), MP, or TG	TPMT	rs1142345 (A>G)	AA	Normal	
5	Carbamazepine or Phenytoin	HLA-B*1502	rs2844682 (C>T)	CT	Normal	
			rs3909184 (C>G)	CC		
6	Abacavir	HLA-B*5701	rs2395029 (T>G)	TT	Normal	
7	Allopurinol	HLA-B*5801	rs9263726 (G>A)	GG	Normal	

Clopidogrel¹⁾: UM/EM=standard dose, IM/PM= consider alternative antiplatelet agent (eg. prasugrel/ticagrelor)

Warfarin²⁾: high dose=5~7 mg/day, medium dose=3~4 mg/day, low dose=0.5~2 mg/day



23andMe

\$99

N>800,000

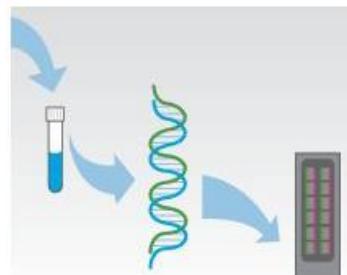
Here's what you do:



1. Order a kit from our [online store](#).



2. Register your [kit](#), spit into the tube, and send it to the lab.



3. Our CLIA-certified lab analyzes your DNA in 6-8 weeks.



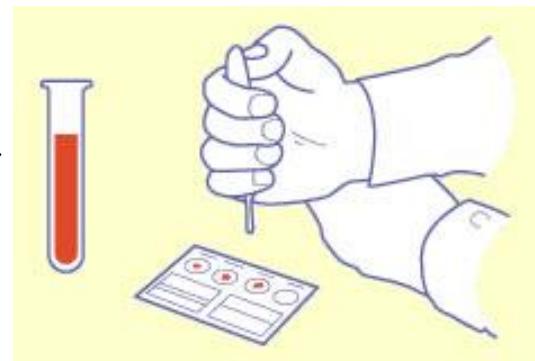
4. Log in and start exploring your genome.

Future of Genomic Medicine?

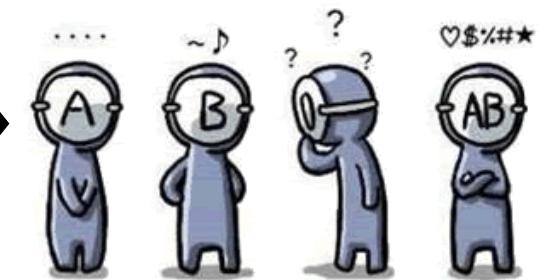
Without information



Test when needed



Know your type



Blood
type

Geno
type



Contents

1. What is Healthcare Big Data?

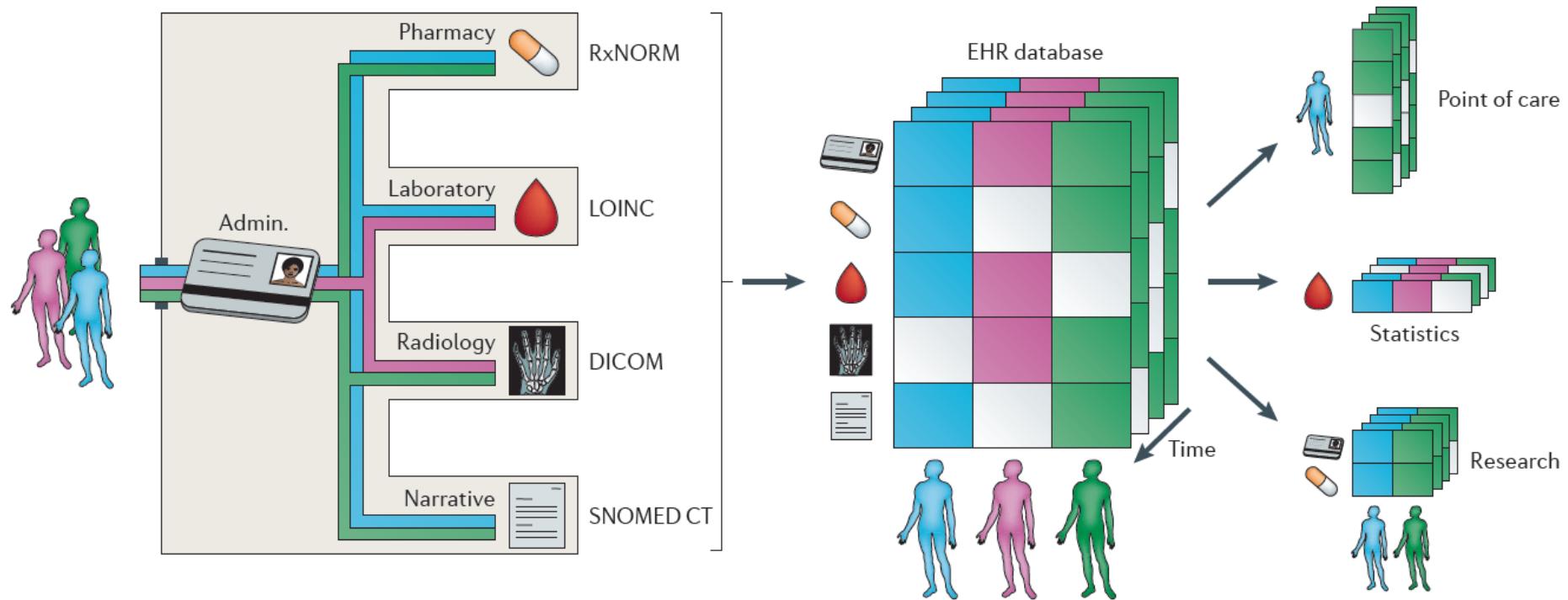
2. Healthcare Big Data

- ① Genetic Data
- ② Electrical Health Records
- ③ National Healthcare Data
- ④ Medical Images
- ⑤ Sensor/Mobile Data

3. Clinical and Research Applications



Electronic Health Records



Common EHR Data

	ICD	CPT	Lab	Medication	Clinical notes
Availability	High	High	High	Medium	Medium
Recall	Medium	Poor	Medium	Inpatient: High Outpatient: Variable	Medium
Precision	Medium	High	High	Inpatient: High Outpatient: Variable	Medium high
Format	Structured	Structured	Mostly structured	Structured and unstructured	Unstructured
Pros	Easy to work with, a good approximation of disease status	Easy to work with, high precision	High data validity	High data validity	More details about doctors' thoughts
Cons	Disease code often used for screening, therefore disease might not be there	Missing data	Data normalization and ranges	Prescribed not necessary taken	Difficult to process

International Classification of Diseases (ICD)
Current Procedural Terminology (CPT)

Medication Data

화일(E) 편집(E) 조회(Q) 경로(R) Frequency 용법(D) 정도(G) 기타 도구(I) CP/DRG 참(W)

Lab 타과 상병보기 상병 rem rem rem

환자처방요약 [REDACTED] M 1958-09-13 충청북도진천군광혜원면

환자처방요약

환자번호 [REDACTED] 전체(현재월~3개월)
입원(입원기간)
외래(외래진료월~3개월)

처방코드 분류선택 조회 종료

기간 2014-01-01 ~ 2014-03-31

I/O	처방구분	처방코드	처방명	년-월	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
0	5	7GLAPEN	Insulin Glargin Pen Type (La)	2014-02																															
				2014-03	-	-																													
0	5	7INGUPEN	Insulin Glulisine (Apidra Solost)	2014-01																															
				2014-02																															
				2014-03	-	-																													
0	6	ACYCL-0	Acyclovir (Vancarb Dint.5G/TU)	2014-01																															
0	6	BACT	Sulfamethoxazole/Trimethopri	2014-01																															

▼ 처방상세내역

외래처방 상세내역	값
처방코드	MYCONA36
처방명	Mycophenolate Na (Myfortic)
처방일자	2014-02-05
처방수량	1
처방횟수	2회
처방날수	21일
기본용량	360
처방투여량	360
수량*횟수*(투여량/기본용량)	2
처방실시의사	[REDACTED]

Lab Data

종료 진단검사의학 핵의학체외 핵의학체내 영상의학 병리검사 특수검사 영양검색 검사별처방목록 Lab출력

**처방일별
진료과별**

처방일자 **2014-02-26** 병록번호 [REDACTED]
 접수일자 성별/나이 M/55 접수일자 2014-03-24
 담당의사 접수시간 09:06:36 시계열 조회

총료 [X] **Refresh**

기간별 1년간 한달간 2013-03-26 ~ 2014-03-26 일반검사 접수 시간 **결과조회** **항목선택** **출력** **Exit**

검사명	참고치	14-03-24	14-02-24	14-02-03	14-01-21	14-01-13	14-01-06	13-12-30	13-12-30	13-12-24	13-12-20
22 CBC/WBC count	4.0~10.0	4.71	▼3.78	4.93	4.80	4.78	6.37	6.74		6.84	9.72
23 RBC count	M:4.2~6.3,F: 4.37		▼4.01	▼3.83	▼3.28	▼3.15	▼3.18	▼3.19		▼3.59	▼3.05
24 Hb	M:13~17,F:12	▼12.9	▼12.3	▼11.8	▼10.3	▼9.7	▼9.8	▼10.0		▼11.4	▼9.6
25 Hct	M:39~52,F:36~39.2		▼36.6	▼35.3	▼30.1	▼28.7	▼28.8	▼29.2		▼32.4	▼27.7
26 PL								172		196	179
27 ESR								▲13		▲25	
28 se								72.2		71.0	90.7
29 Ly								▼14.8		▼17.4	▼5.1
30 Mo								▲11.7		9.9	4.1
31 Eo								1.2		1.6	0.1
32 Ba								0.1		0.1	0.0
33 RD								▲15.9		▲15.9	▲16.3
34 PD								▼12.2		▼11.9	▼11.6
35 Chen											
36 Elect											
37 Ca											
38 Glu											
39 Cr											
40 eG											
41 Ur											
42 Ch											
43 Pr											
44 Alb											
45 AS											
46 AL											
47 Alk											
48 Bill											
49 Ph											
50 BUN	8.0~20.0	▲22.9	19.7					1.6		20.0	20.0
51 Sodium	135~145	141	145					40		127	142

HbA1c-NGSP

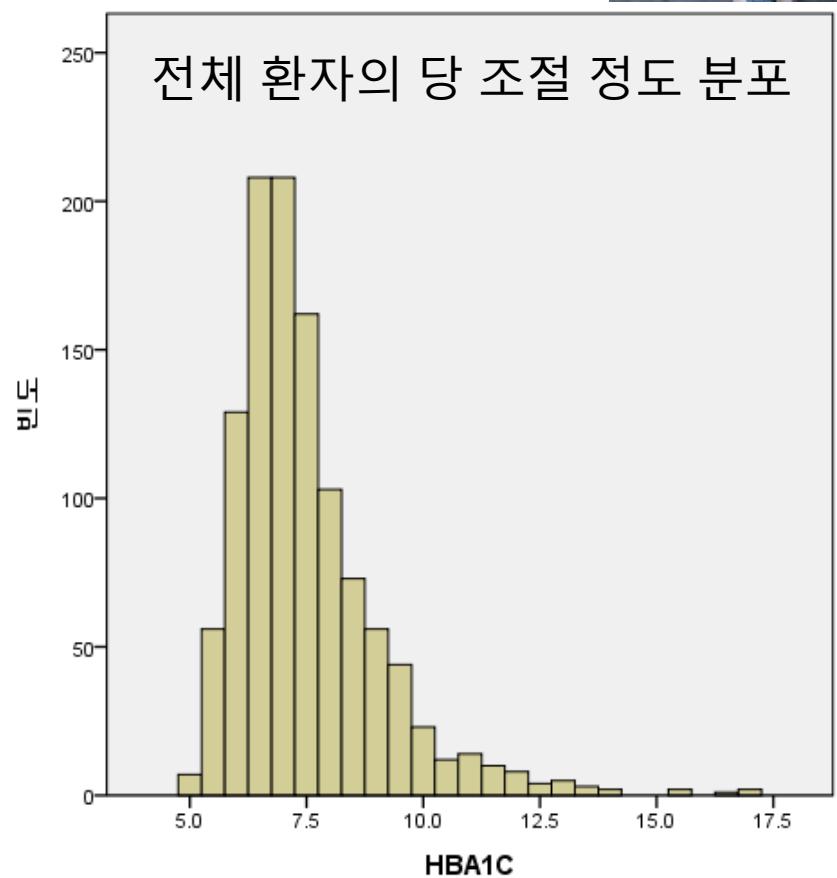
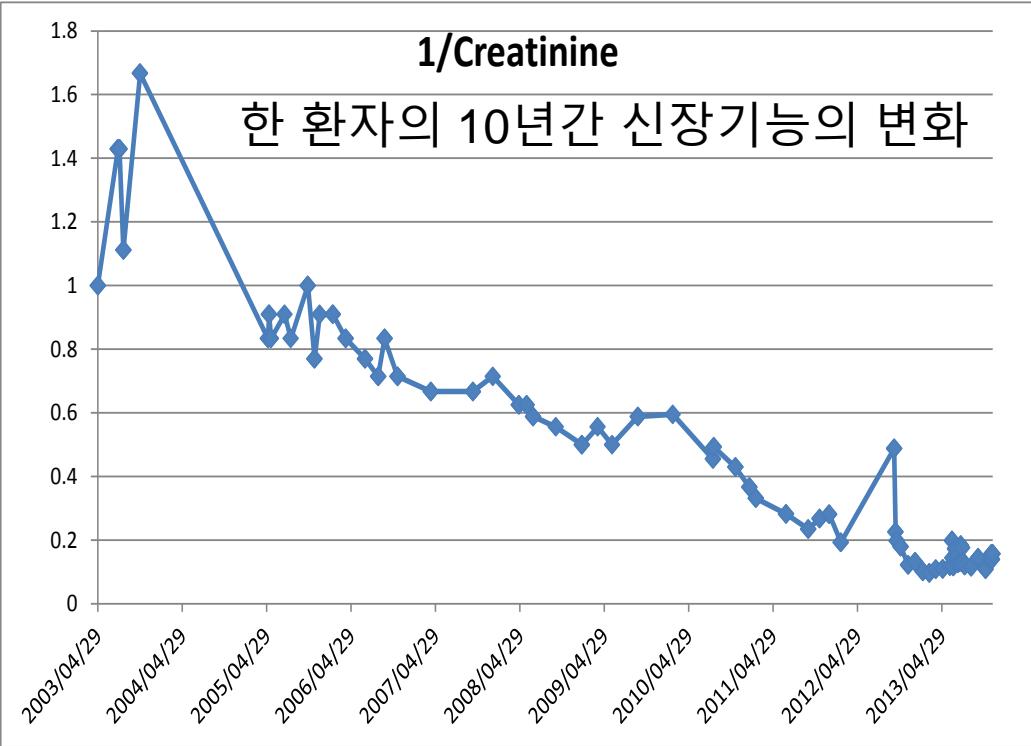
결과값

검사일

	A	B	C	D	E	F	G	H	I	J
1	PTNO	SNAME	SEX	AGE	JUPSUPDATE	ORDERCC	CODENAME	SUBCODE	SUBCODE	RDATA
2	012	윤	M	50	1998-11-06	NK8001	CBC(Em)	NK8011	WBC	6.4
3	012	윤	M	50	1998-11-06	NK8001	CBC(Em)	NK8013	Hb	13.8
4	012	윤	M	50	1998-11-06	NK8001	CBC(Em)	NK8015	Platelet	361
5	012	윤	M	50	1998-11-06	NK8101	Electrolyte	NK8130	Sodium	150
6	012	윤	M	50	1998-11-06	NK8101	Electrolyte	NK8131	Potassium	4.0
7	012	윤	M	50	1998-11-06	NK8101	Electrolyte	NK8132	Chloride	113
8	012	윤	M	50	1998-11-06	NK8101	Electrolyte	NK8133	Co2, Total	23
9	012	윤	M	50	1998-11-06	NK8142	Glucose	NK8142	Glucose	158
10	012	윤	M	50	1998-11-06	NK8143	Creatinine	NK8143	Creatinine	0.9
11	012	윤	M	50	1998-11-06	NK8144	BUN	NK8144	BUN	17
12	012	윤	M	50	1998-11-06	NK8146	AST	NK8146	AST	40
13	012	윤	M	50	1998-11-06	NK8147	ALT	NK8147	ALT	62
14	009	권	M	60	1999-03-26	NK3003	Electrolyte	NK3041	Sodium	143
15	009	권	M	60	1999-03-26	NK3003	Electrolyte	NK3041	Sodium	143
16	009	권	M	60	1999-03-26	NK3003	Electrolyte	NK3042	Potassium	5.3
17	009	권	M	60	1999-03-26	NK3003	Electrolyte	NK3042	Potassium	5.3
18	009	권	M	60	1999-03-26	NK3003	Electrolyte	NK3043	Chloride	108
19	009	권	M	60	1999-03-26	NK3003	Electrolyte	NK3043	Chloride	108
20	009	권	M	60	1999-03-26	NK3003	Electrolyte	NK3044	Total CO2	21
21	009	권	M	60	1999-03-26	NK3003	Electrolyte	NK3044	Total CO2	21
22	022	김	F	65	1999-06-01	NK3231	HbA1c	NK3231	HbA1c	6.14 %
23	022	김	F	65	1999-06-01	NK3231	HbA1c	NK3231	HbA1c	6.14 %
24	002	송	M	64	1999-06-02	NK3004	Lipid Batter	NK3015	Cholesterol	218
25	002	송	M	64	1999-06-02	NK3004	Lipid Batter	NK3081	Triglycerid	215

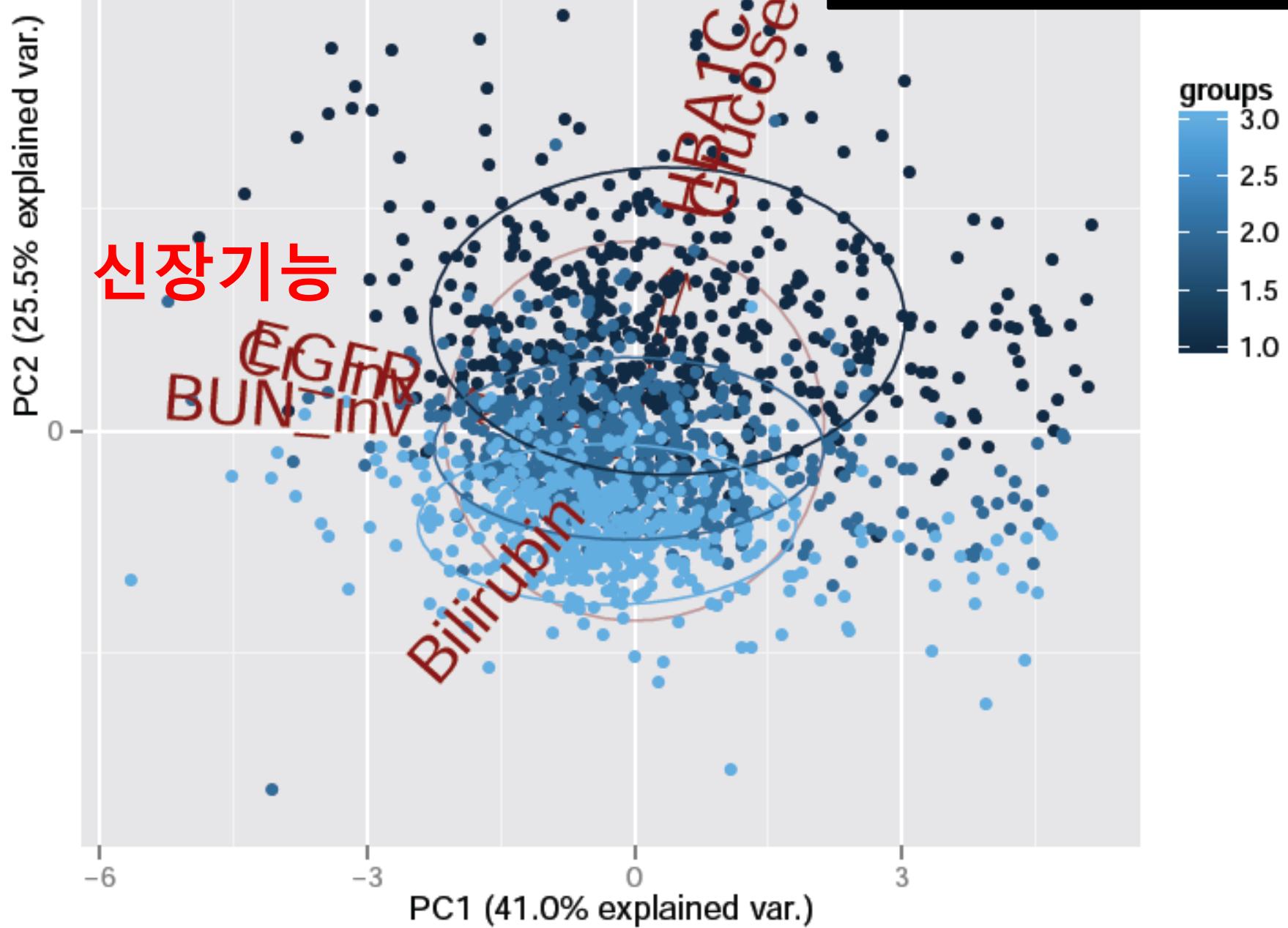
	A	B	C	D	E	F	G	H	I	J
1048549	061	최	M	55	2014-02-11	NK312060	Creatinine	NK3120	Creatinine	211.6
1048550	061	최	M	55	2014-02-11	NK312360	Micro Alb	NK312360	Micro Albumin,	
1048551	061	최	M	55	2014-02-11	NK312360	Micro Alb	NK312360	A/C ratio	17.3
1048552	061	최	M	55	2014-02-11	NK312360	Micro Alb	NK312360	Micro Alb	36.6
1048553	061	최	M	55	2014-02-11	NK312360	Micro Alb	NK312360	Creatinine	211.6
1048554	061	최	M	55	2014-02-11	NK3231	HbA1c	NK3231	HbA1c-NC	7.6
1048555	061	최	M	55	2014-02-11	NK3231	HbA1c	NK3231	HbA1c-NC	7.6
1048556	066	반	M	53	2014-02-11	NK3011	Calcium	NK3011	Calcium	9.2
1048557	066	반	M	53	2014-02-11	NK3014	Uric Acid	NK3014	Uric acid	7.2
1048558	066	반	M	53	2014-02-11	NK3031	Phosphorus	NK3031	Phosphorus	3.8
1048559	066	반	M	53	2014-02-11	NK8000	응급 검사:	NK8000 32	PLT. clumps	
1048560	066	반	M	53	2014-02-11	NK8000	응급 검사:	NK8001 19	NE%	61.0
1048561	066	반	M	53	2014-02-11	NK8000	응급 검사:	NK8011	WBC count	7.63
1048562	066	반	M	53	2014-02-11	NK8000	응급 검사:	NK8013	Hb	10.4
1048563	066	반	M	53	2014-02-11	NK8000	응급 검사:	NK8015	PLT count	152
1048564	066	반	M	53	2014-02-11	NK8101	응급 검사:	NK8130	Sodium,EM	134
1048565	066	반	M	53	2014-02-11	NK8101	응급 검사:	NK8131	Potassium	5.3
1048566	066	반	M	53	2014-02-11	NK8101	응급 검사:	NK8132	Chloride,EM	104
1048567	066	반	M	53	2014-02-11	NK8101	응급 검사:	NK8133	Total CO2	18
1048568	066	반	M	53	2014-02-11	NK8172	응급 검사:	NK8142	Glucose,EM	169
1048569	066	반	M	53	2014-02-11	NK8172	응급 검사:	NK8143	Creatinine	2.88
1048570	066	반	M	53	2014-02-11	NK8172	응급 검사:	NK81431	eGFR	24.5
1048571	066	반	M	53	2014-02-11	NK8172	응급 검사:	NK8144	BUN,EM	34.6
1048572	066	반	M	53	2014-02-11	NK8172	응급 검사:	NK8145	Protein,EM	7.0
1048573	066	반	M	53	2014-02-11	NK8172	응급 검사:	NK8146	AST,EM	13
1048574	066	반	M	53	2014-02-11	NK8172	응급 검사:	NK8147	ALT,EM	9
1048575	066	반	M	53	2014-02-11	NK8172	응급 검사:	NK8156	Albumin,EM	4.4
1048576	066	반	M	53	2014-02-11	NK8172	응급 검사:	NK8170	Bilirubin t	0.38

Big Data Analysis



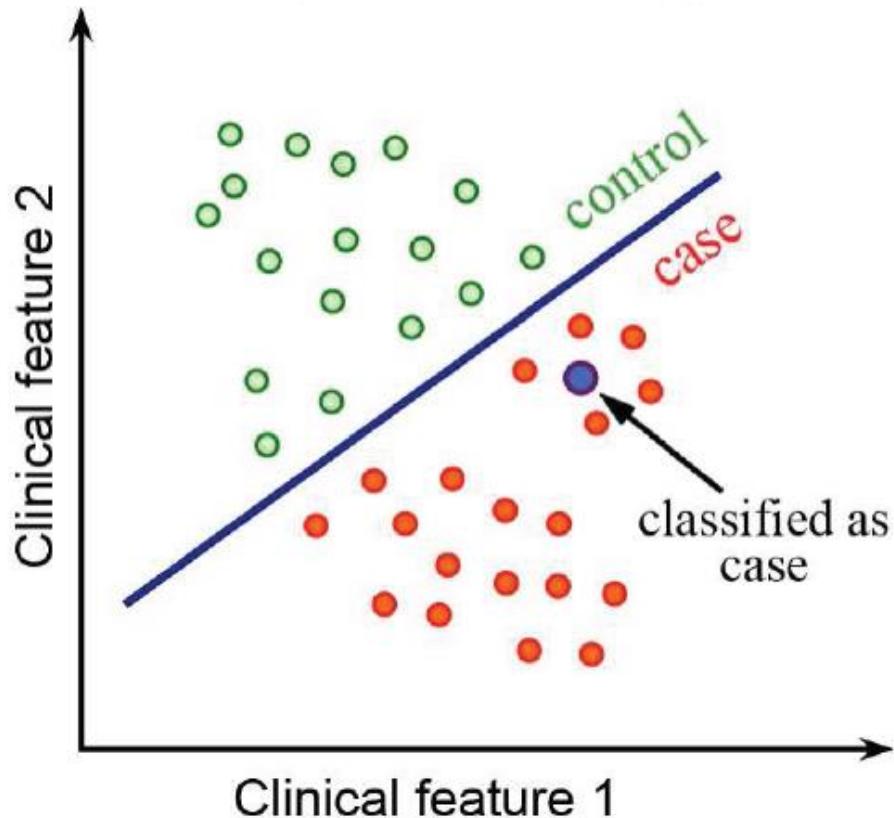
혈당

PCA Analysis

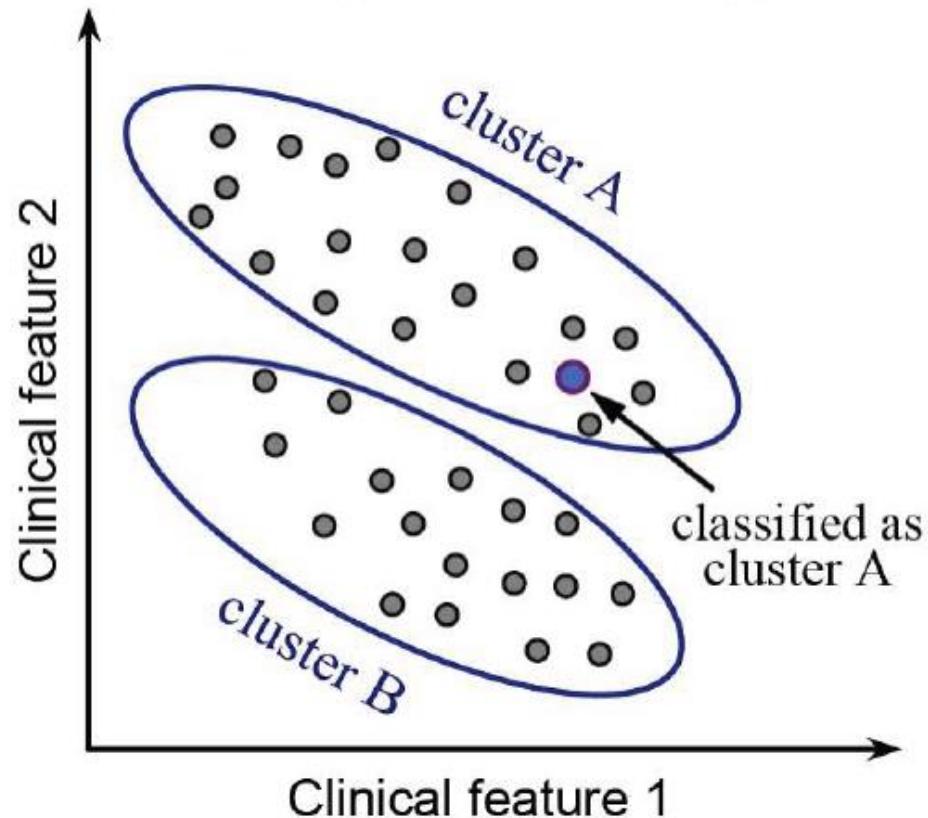


Machine Learning

Supervised learning



Unsupervised learning





환자메모 [재진][2014-06-11] HD: (-) POD:

F/47.6

환자메모

기록조회

기록작성

외래 재진기록지-IM-기본 [내과]

외래 재진기록지 [내과] [2014-06-11]

등록번호: [REDACTED] 성명: [REDACTED] 성별: 여 나이: 47
작성일자: 2014-06-11 09:19 진료과: 내분비대사(내과)

Subjective
약 잘 복용한다.
Graves

약 잊은 날이 하루 있다.

Objective
T3 S.B 111 78-182
Free T4 S.B 1.20 0.89-1.79
TSH S.B ▼0.09 0.17-4.05
TSH-R-Ab(TBII) S.B 4.2 1.00|하 :Negative

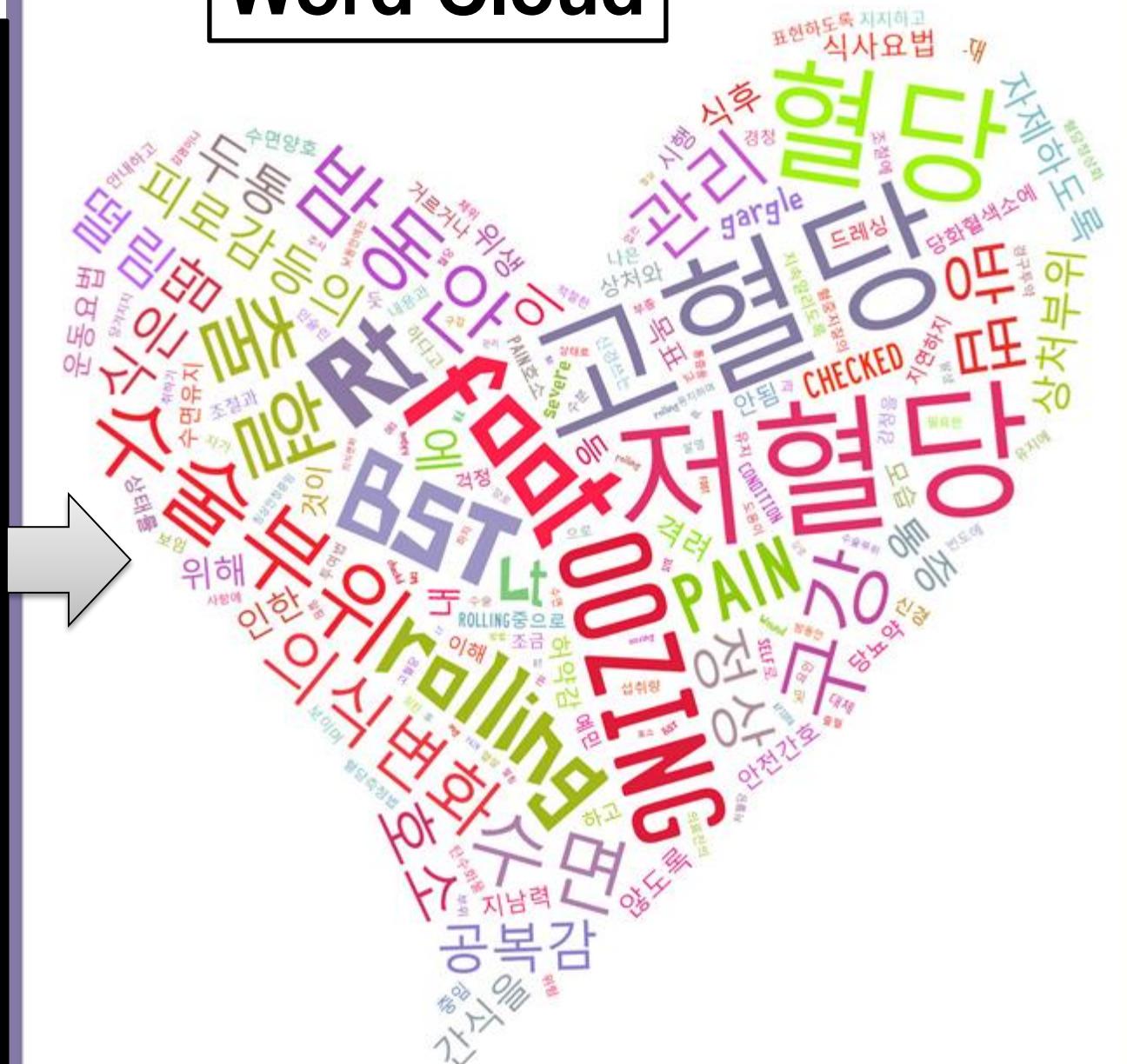
추정진단(Assessment)
갑상선 발증 또는 급성 발작을 동반하지 않은 미만성 고이터를 동반한 갑상선증증

계획(Plan)
유지 (2013.12.31. 시작)

간호기록지

Word Cloud

밤동안 저혈당수면 Lt.foot rolling Keep떨림, 식은땀, 현기증, 공복감, 두통, 피로감등의 저혈당에 저혈당이 있을 즉알려주도록 밤사이 특이호소 수면유지상처와 통증 상처부위 출혈 oozing, severe pain 알리도록 고혈당 처방된 당뇨식이의 중요성과 간식을 자제하도록 고혈당,,관리 방법 .당뇨약 이해 잘하고 수술부위 oozing Rt.foot rolling keep드레싱 상태를 고혈당 고혈당 의식변화 BST 387 checked.고혈당으로 인한 구강 내 감염 위해 식후 양치, gargle 등 구강 위생 격려.당뇨환자의 발관리 방법에 .목표 혈당, 목표 당화혈색소에 .식사를 거르거나 지연하지 않도록 .식사요법, 운동요법, 약물요법을 정확히 지키는 것이 중요을 .처방된 당뇨식이의 중요성과 간식을 자제하도록 고혈당,,관리 방법 .혈당 정상 범위임 rt foot rolling 중으로 pain호소 밤사이 수면양호걱정신경 예민감정변화 중임감정을 표현하도록 지지하고 경청기분상태 condition 조금 나은 듯 하다고 혈당 조절과 관련하여 신경쓰는 모습 보이며 혈당 self로 측정하는 모습 보임혈당 조절에 안내하고 불편감 지속알리도록고혈당 고혈당 의식변화 고혈당 허약감 지남력 혈당조절 안됨고혈당으로 인한 구강 내 감염 위해 식후 양치, gargle 등 구강 위생 격려.당뇨환자의 정기점검 내용과 빈도에 .BST 140 으로 저혈당 호소 밤동안 저혈당수면 Lt.foot rolling Keep떨림, 식은땀, 현기증, 공복감, 두통, 피로감등의 저혈당에 저혈당이 있을 즉알려주도록 pain 및 불편감 호소 WA 잘고혈당 고혈당 의식변화 고혈당 허약감 지남력 혈당조절 안됨식사요법, 운동요법, 약물요법을 정확히 지키는 것이 중요을 .저혈당/고혈당과 대처법에 .혈당정상화, 표준체중의 유지, 정상 혈중지질의 유지에 .고혈당,,관리 방법 .혈당측정법,인슐린 자가 투여법, 경구투약 .수분 섭취량,대체 탄수화물,의료진의 도움이 필요한 사항에 교혈당 정상 범위임수술부위 oozing Rt.foot rolling keep수술 부위 (출혈, 통증, 부종)수술부위 출혈 상처부위 oozing Wound 당거지지 않도록 적절한 체위 취하기 설명감염 발생 위험 요인 수술부위 출혈 밤동안 혼란다 호소 수면 약산 약호 rt foot rolling은지



Natural Language Processing (NLP)



GCS 양쪽
TENDENCY
DISINTEGRATION
PUPIL Pupil Rhythm
체위변경



건진시장 최강자 강북삼성 한국인 건강지도 그린다

100만 수진자 코호트 분석 완료…“NEJM도 노려볼만 하다”

이인복 기자 news@medicalestimes.com

■ 확대 ■ 축소 ■ 목록 ■ 메일 ■ 프린트 [t](#) [f](#)

기사입력 2015-02-24 05:50

|메디칼타임즈 이인복 기자| 1년에 17만명에 달하는 건강검진 수진자를 모으며 건진시장의 최강자 자리를 지키고 있는 강북삼성병원이 이를 기반으로 한국인 건강지도를 준비하고 있다.

총 100만명에 달하는 검진 수검자 코호트 분석을 통해 한국인의 건강 특성과 질병 정보는 물론 이에 대한 관리방안 이르는 방대한 빅데이터 분석을 내놓겠다는 의지다.



Contents

1. What is Healthcare Big Data?

2. Healthcare Big Data

- ① Genetic Data
- ② Electrical Health Records
- ③ National Healthcare Data
- ④ Medical Images
- ⑤ Sensor/Mobile Data

3. Clinical and Research Applications





골다공증질환의 의료이용 및 약제처방 양상에 관한 연구

연 구 진



건강보험심사평가원
Health Insurance Review & Assessment Service

연구보고서 2009-08

연 구 책 임 자 장 선 미 연구위원

공 동 연 구 자

박 찬 미 부연구위원
장 수 현 주임연구원
신 찬 수 대한골대사학회, 서울대학교병원 내분비내과 교수
하 용 찬 대한골대사학회, 중앙대학교병원 정형외과 교수

자 문 위 원

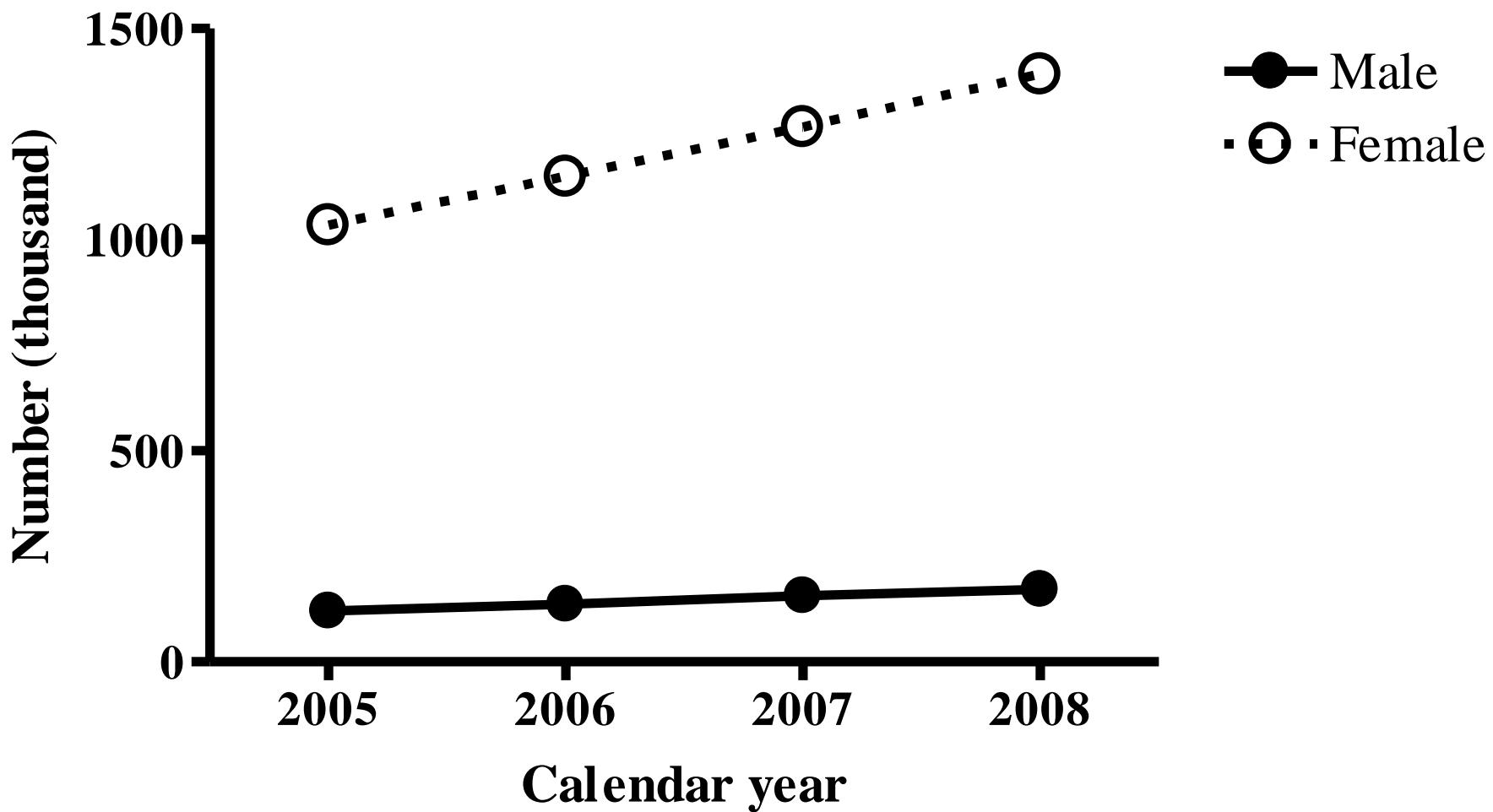
대한골대사학회 역학위원회
윤 현 구 관동대학교병원 내과 교수
김 덕 윤 경희대학교병원 핵의학과 교수
김 범 택 아주대학교병원 가정의학과 교수
이 영 균 분당서울대학교병원 정형외과 교수
이 성 수 가톨릭대학교병원 내과 교수
최 지 협 인제대학교 약물유전체연구센터 교수
최 형 진 질병관리본부 국립보건연구원

ORIGINAL ARTICLE

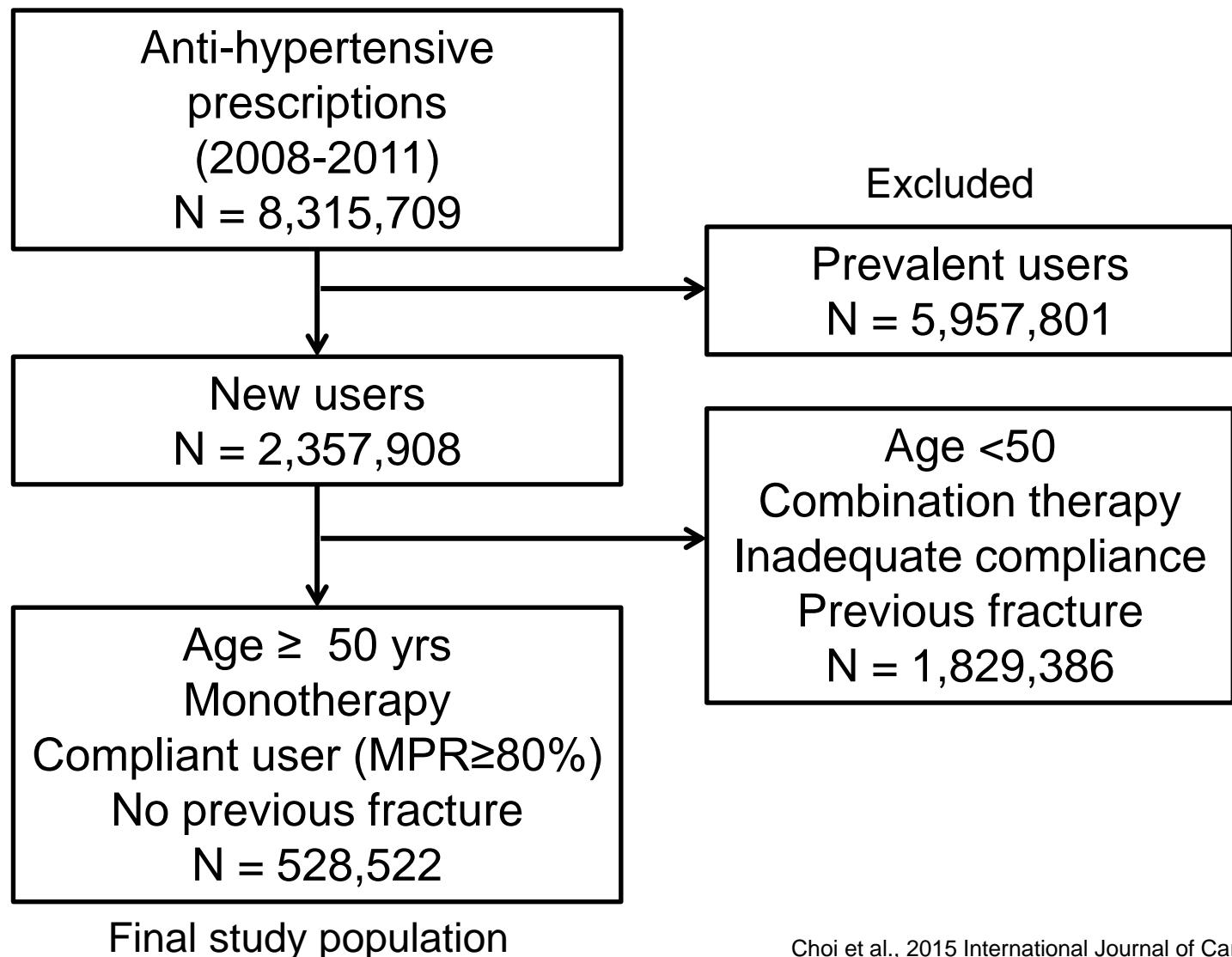
Burden of osteoporosis in adults in Korea: a national health insurance database study

Hyung Jin Choi · Chan Soo Shin · Yong-Chan Ha ·
Sun-Mee Jang · Su-Hyun Jang · Chan Mi Park ·
Hyun-Koo Yoon · Seong-Su Lee

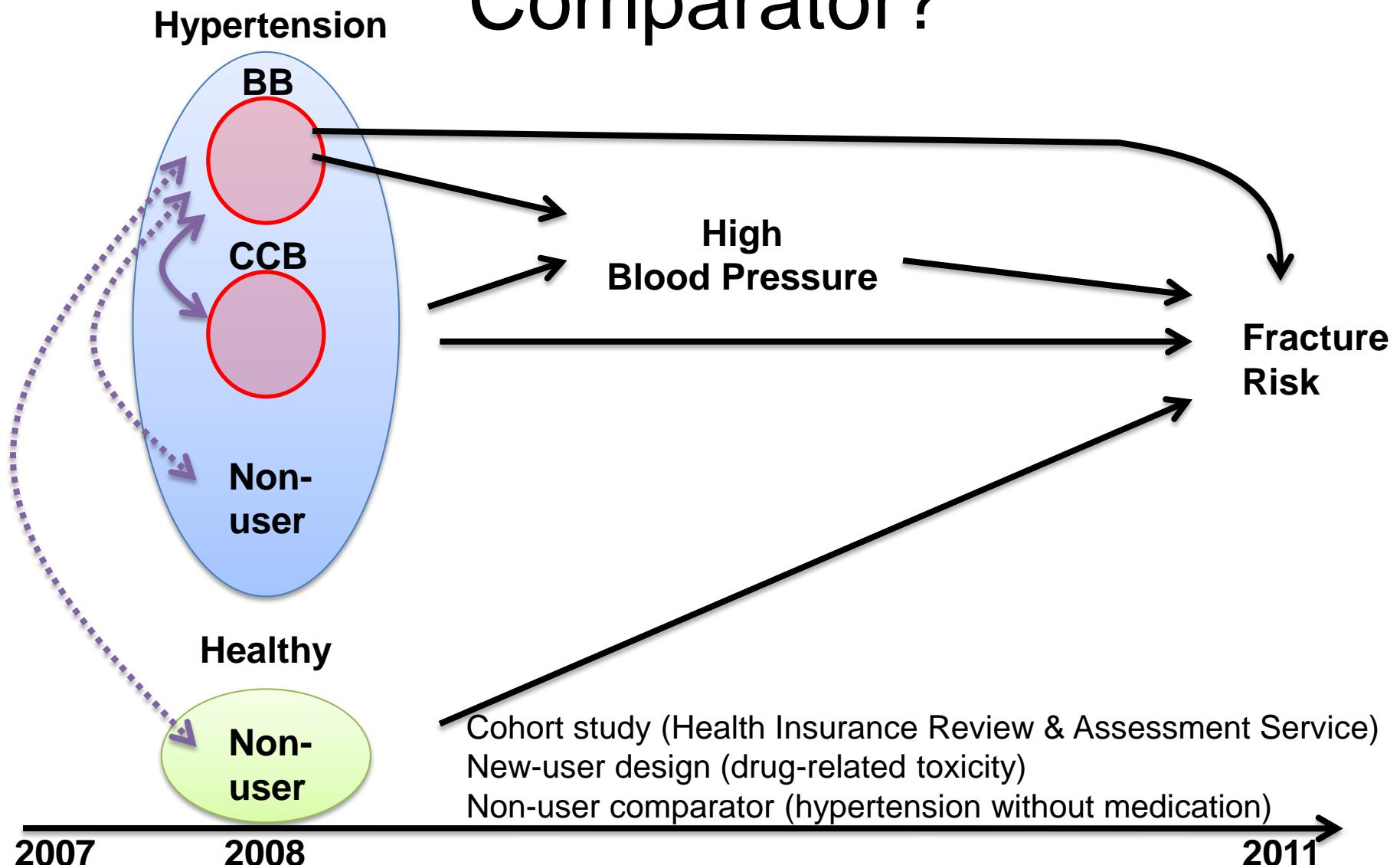
Number of patients with medical treatments related to osteoporosis in each calendar year

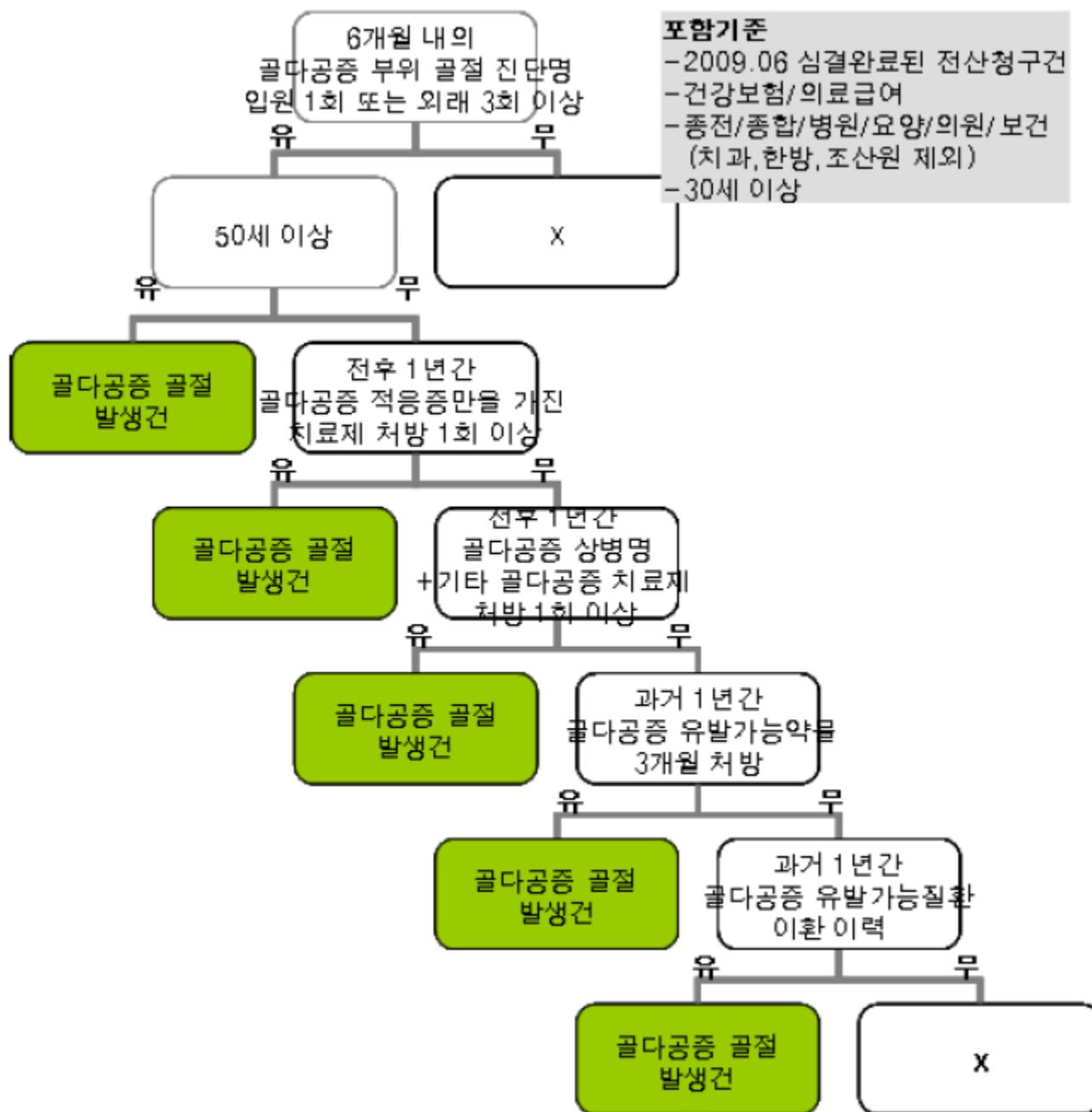


심평원 빅데이터 연구 고혈압약과 골절



Compare Fracture Risk Comparator?





〈그림 3-2〉 골다공증 골절의 조작적 정의

Distribution of ARB MPR (Histogram)

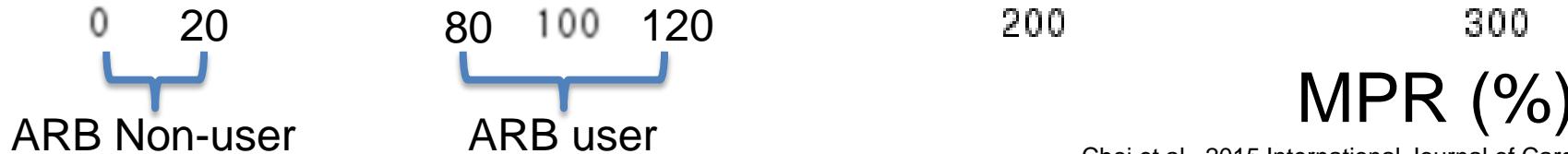
Medication Possession Ratio (MPR)

$$= \frac{\text{Total prescription days}}{\text{Observation days}}$$

$$\frac{350 \text{ days (Prescription)}}{365 \text{ days (Observation)}} = \text{MPR} \\ 96\%$$

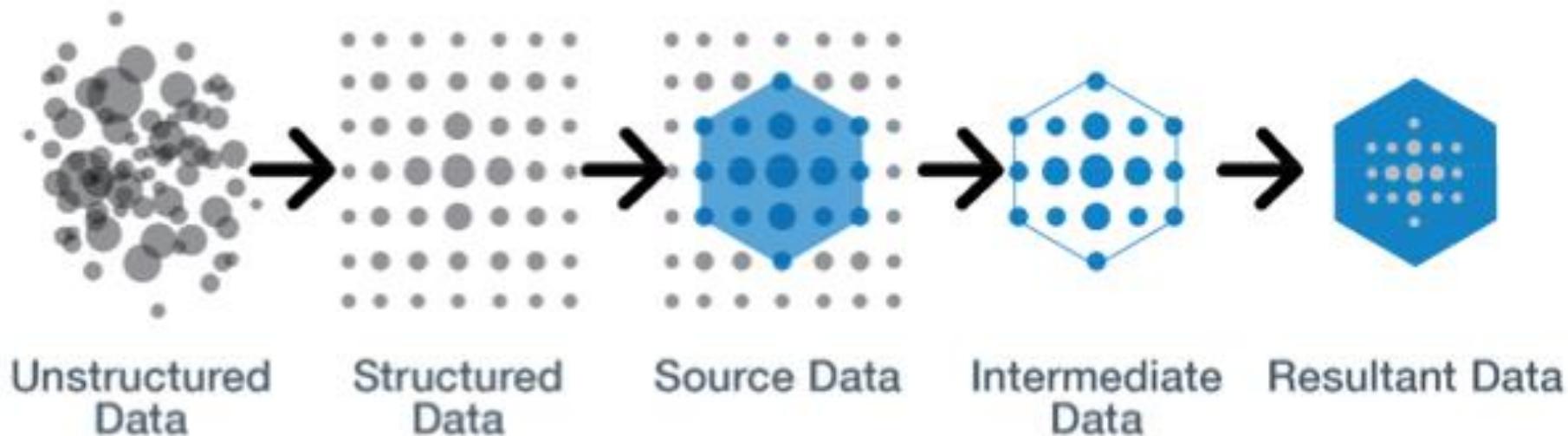
Frequency Density

3
2
1
0



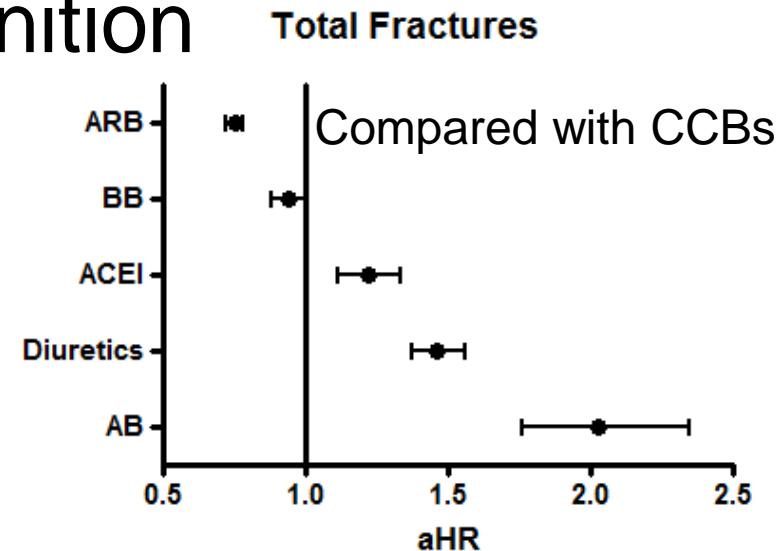
Data Variety

UNSTRUCTURED DATA TO RESULTS



Limitations

- Not a randomized control trial
- Confounding by indication
 - Treatment decisions
 - Age, gender, diabetes, co-morbidity
- Bias from non-users definition
 - HTN without medication
 - Poor compliance
- No BMD data



Overview of secondary data in public health by data source

	보험청구자료 National health claims data	건강검진 National health examination	통계청 Beneficiary	암센터 Cause of death	Cancer registry
Variable	NHIS/HIRA	NHIS		Statistics Korea	National Cancer Center
Socio-demographics	-	○	○	-	-
Health behavior	-	○	-	-	-
Disease	○	○	-	-	○
Pharmaceuticals	○	-	-	-	-
Lab data	-	○	-	-	-
Cause of death	-	-	Death only	○	-

NHIS (National Health Insurance Corporation): 국민건강보험공단 (보험공단)

HIRA (Health Insurance Review and Assessment Service) : 건강보험심사평가원 (심평원)

Contents

1. What is Healthcare Big Data?

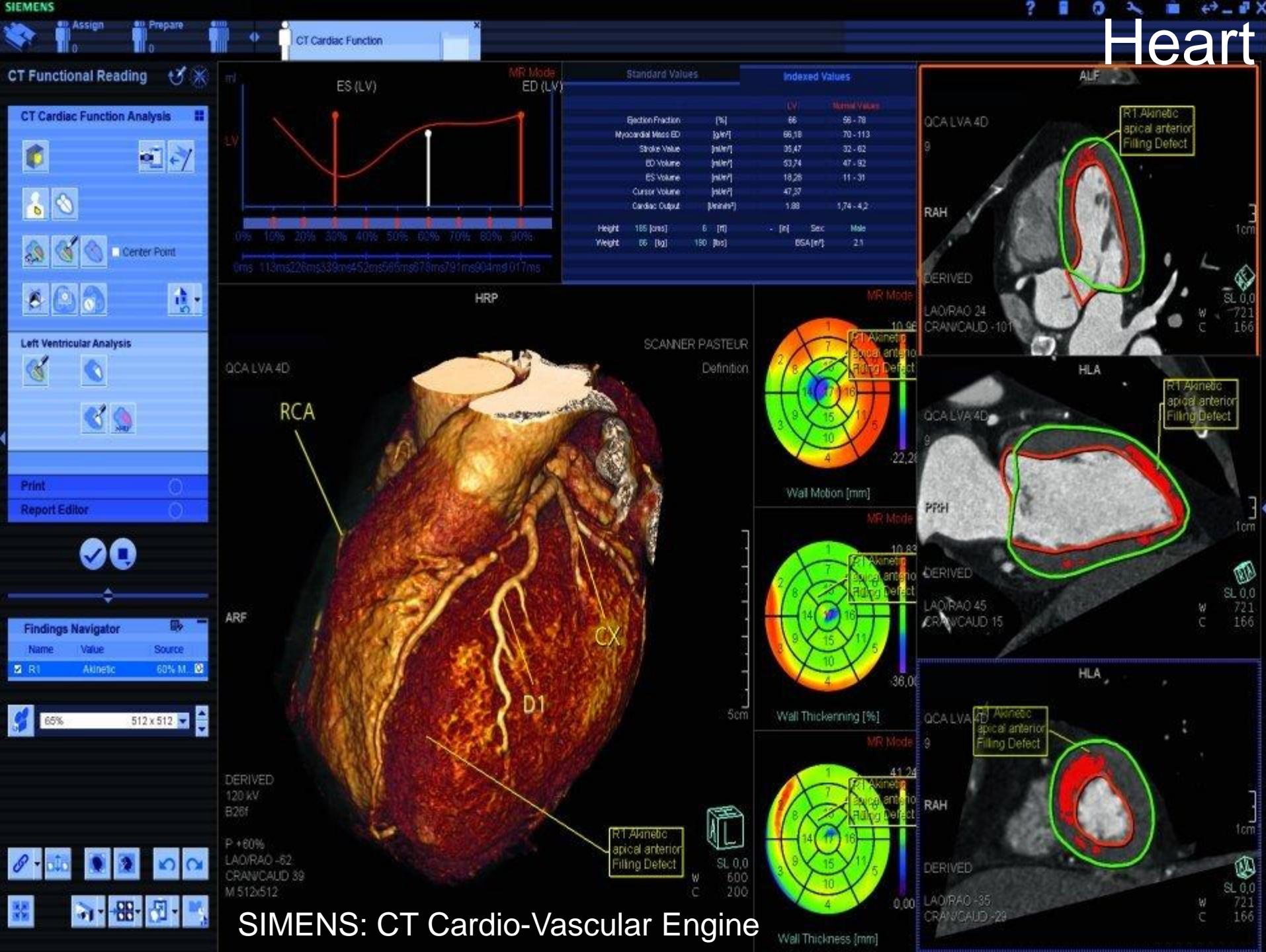
2. Healthcare Big Data

- ① Genetic Data
- ② Electrical Health Records
- ③ National Healthcare Data
- ④ Medical Images
- ⑤ Sensor/Mobile Data

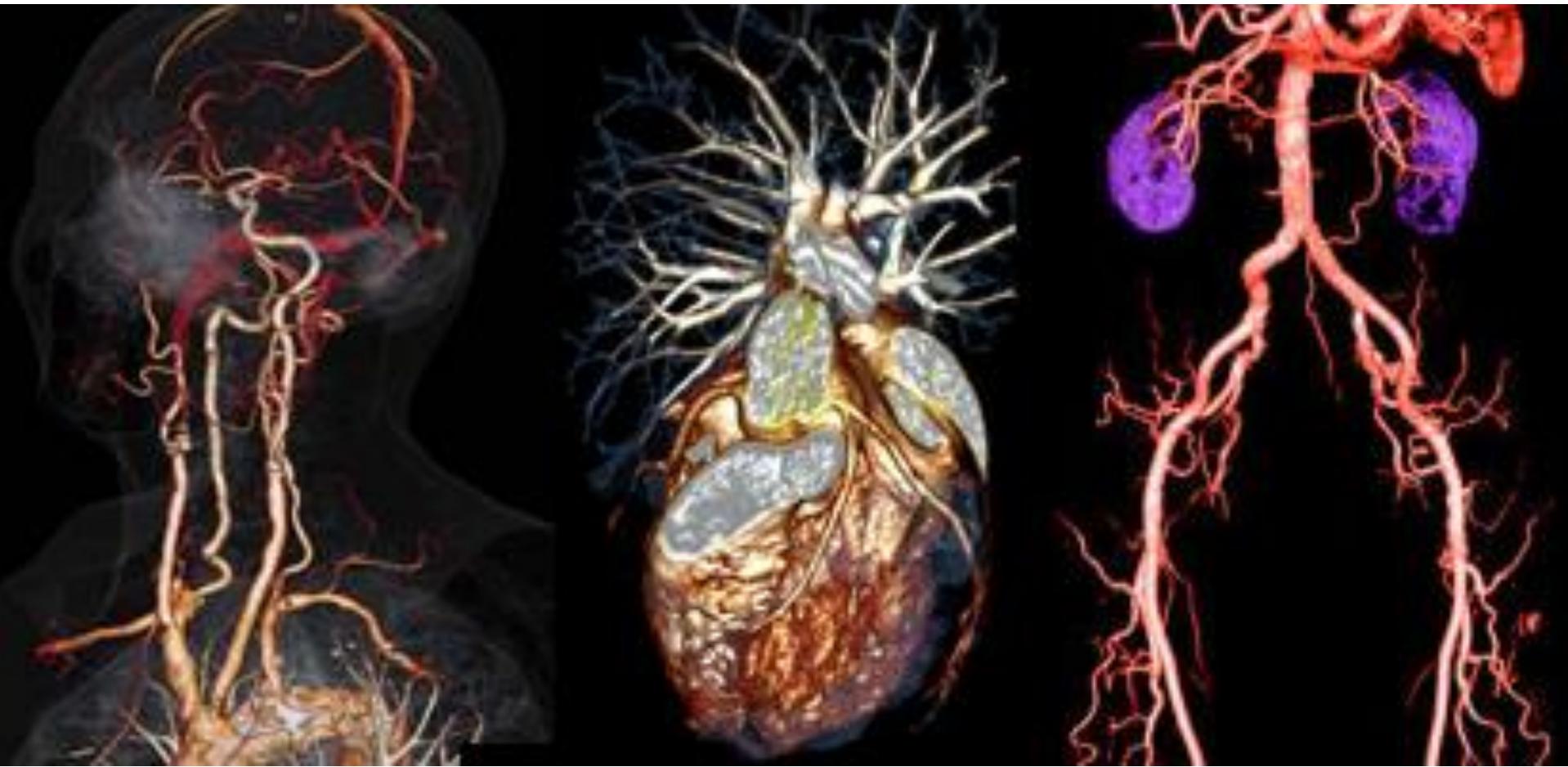
3. Clinical and Research Applications

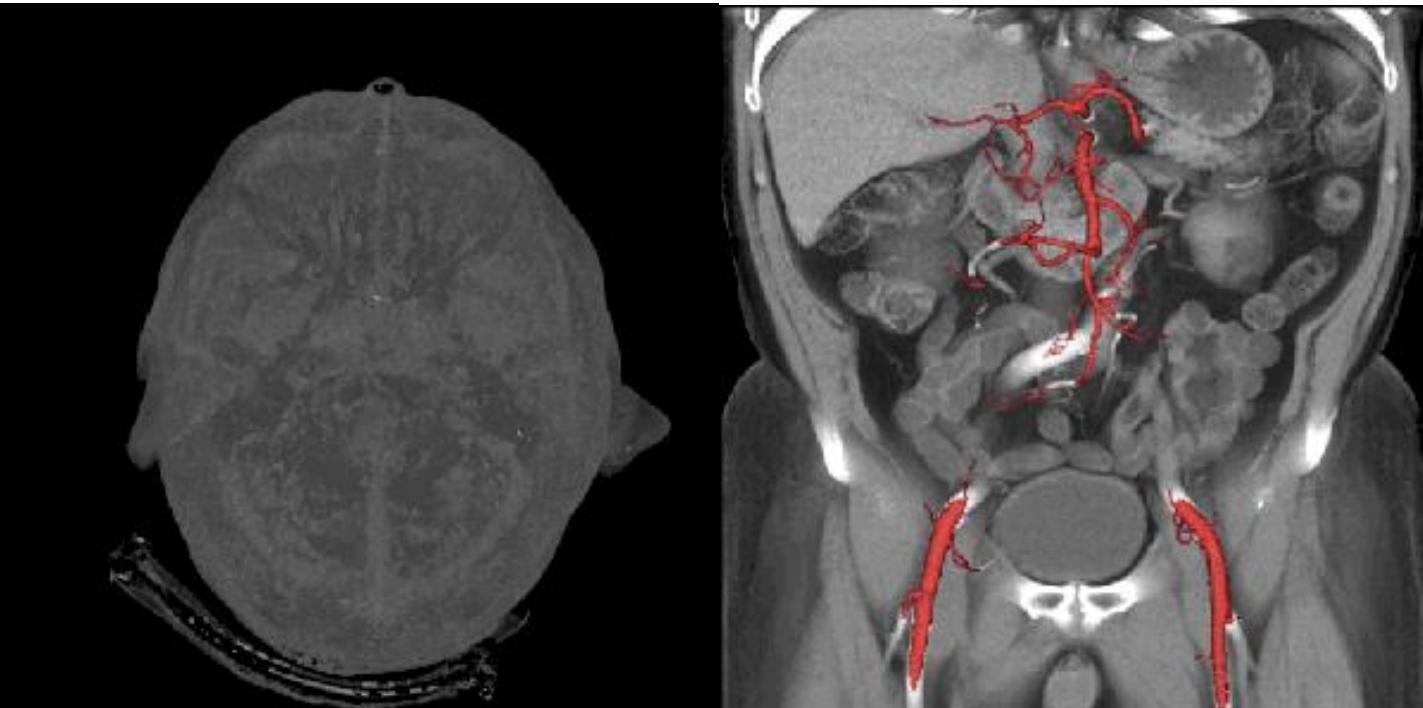


Heart



CT Angiography



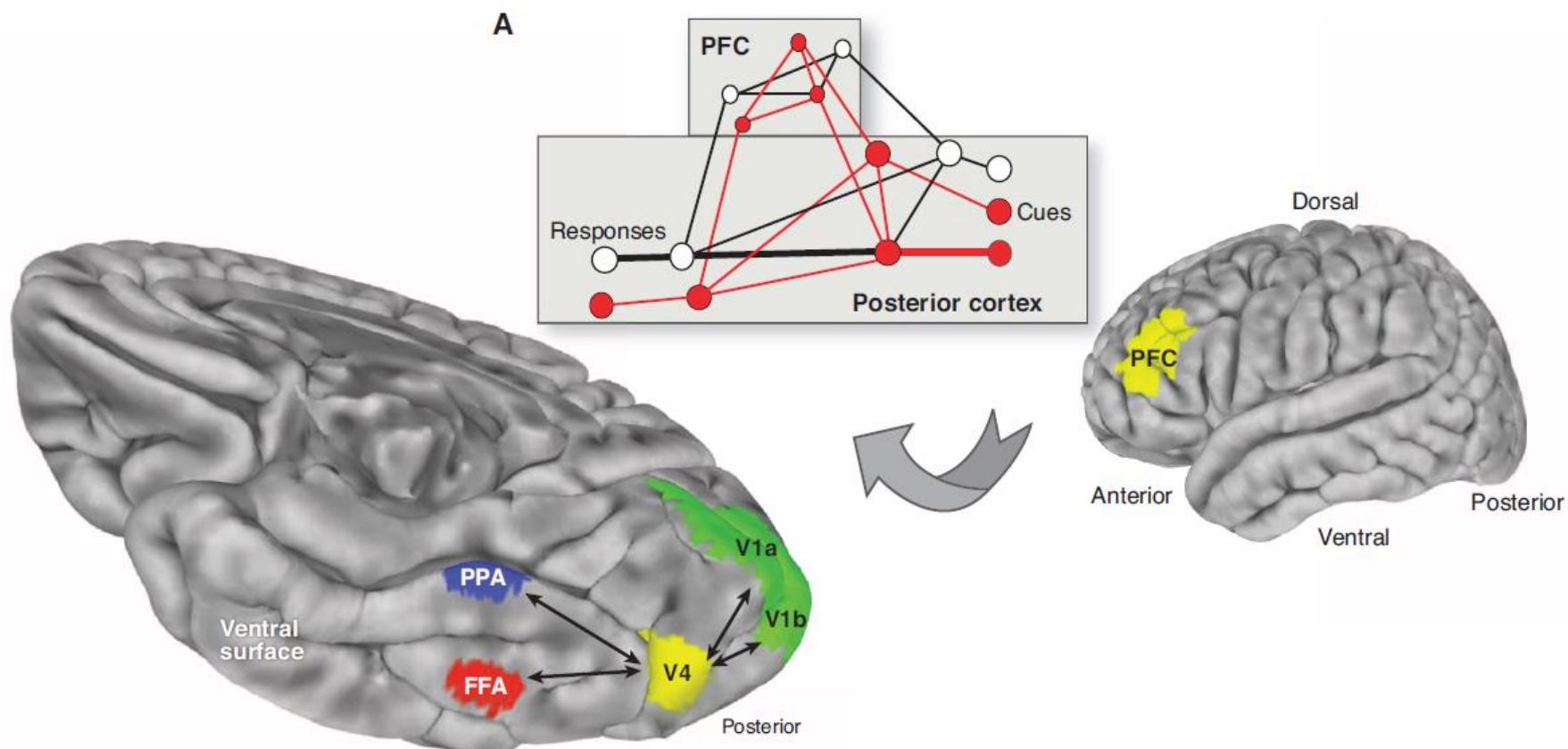


General
Electric
2015.1.8.

REVIEW

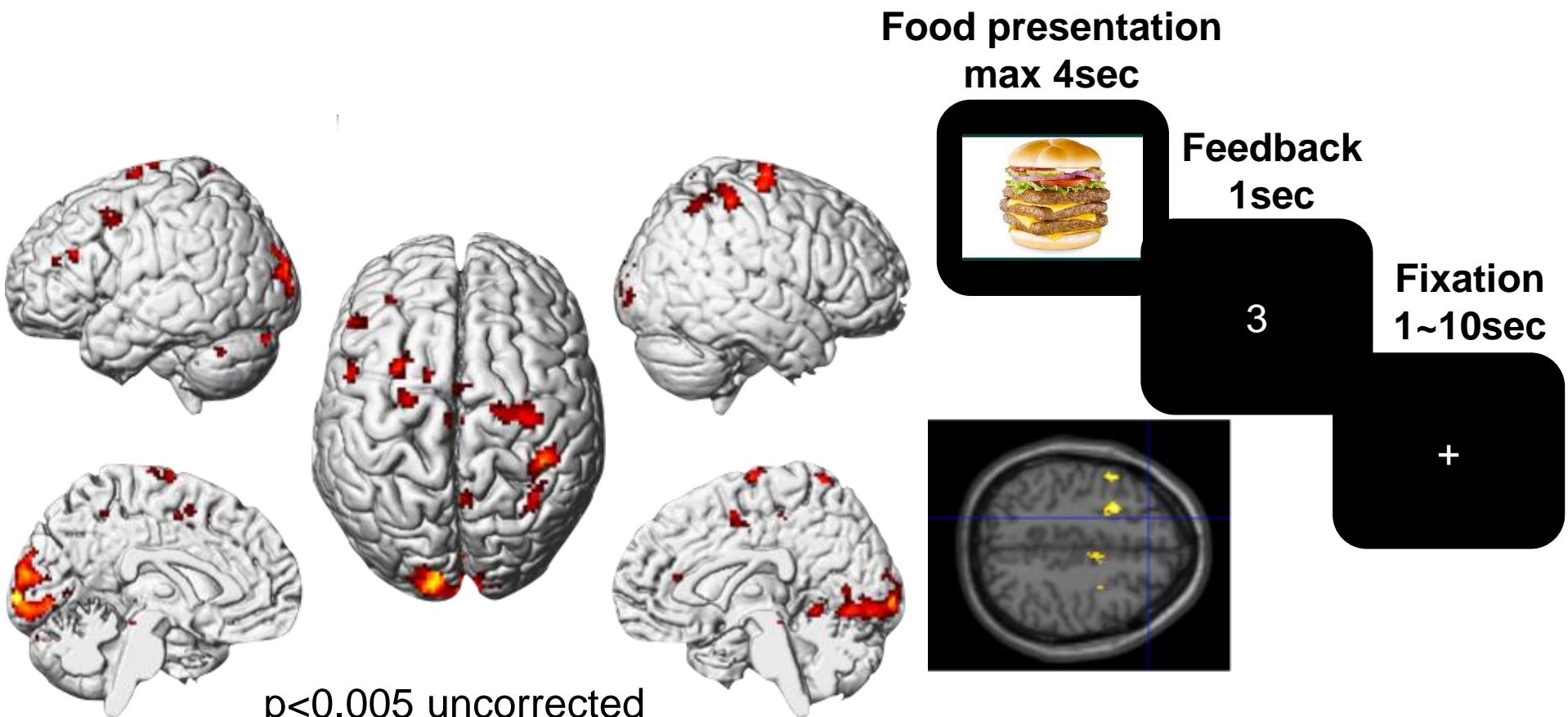
Science

Functional Interactions as Big Data in the Human Brain



Obesity fMRI Research Preliminary Pilot Results

Activation of visual, motor and prefrontal brain area in response to visual food cue



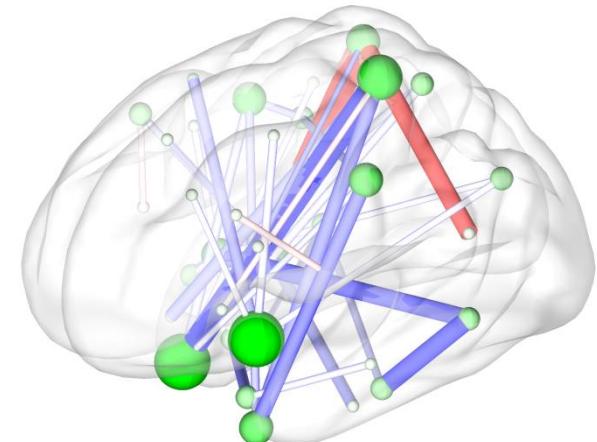
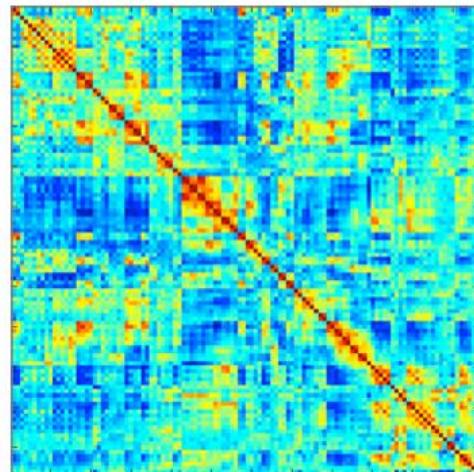
Resting state fMRI analysis

Complex network approach

Parcellation
116 brain regions
by AAL map

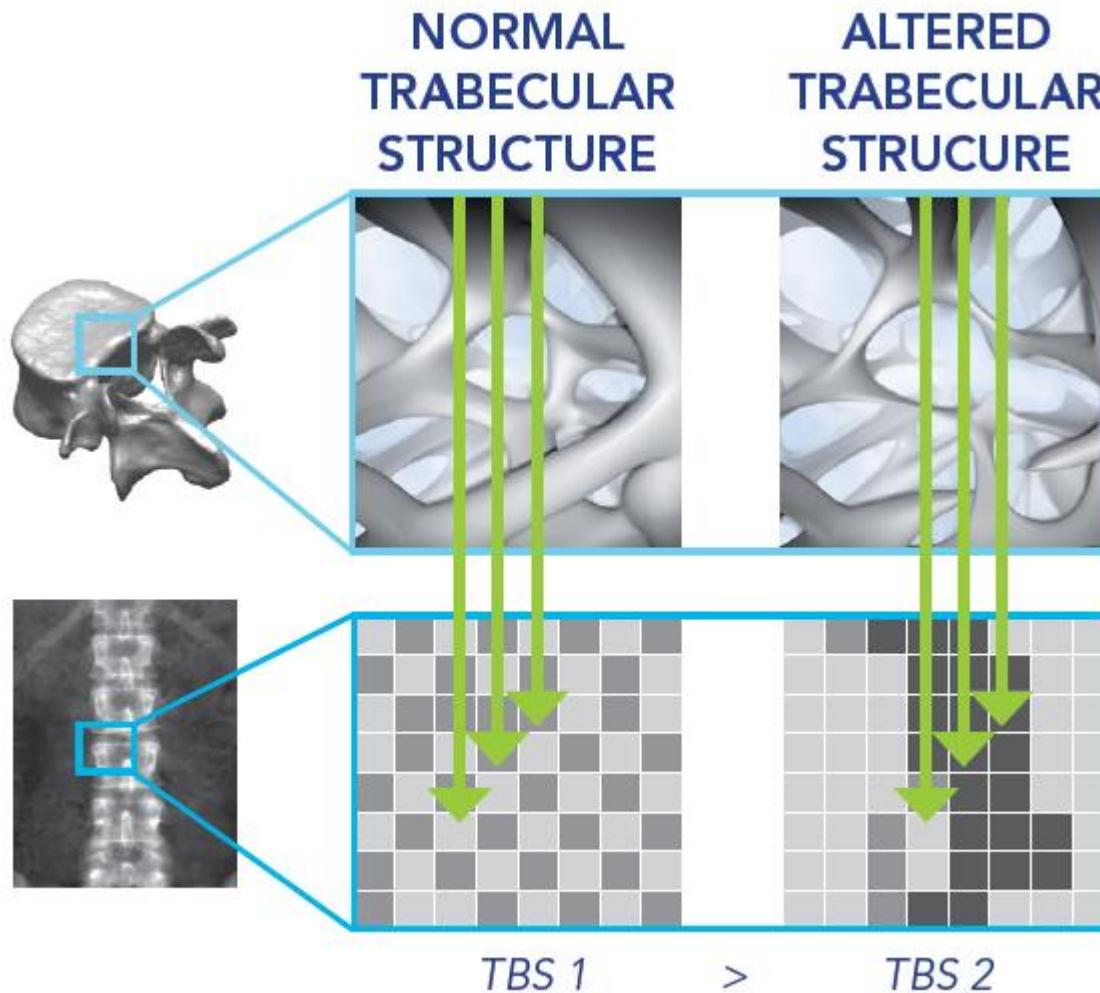
Adjacency matrix

Network properties



node degree, clustering coefficient

Bone Quality and TBS (Trabecular Bone Score)



Clinical Implication of TBS

Bone Mineral Density images

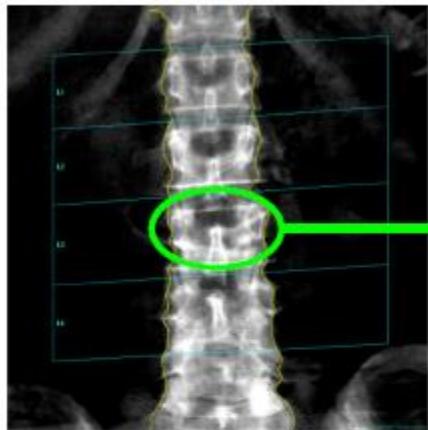
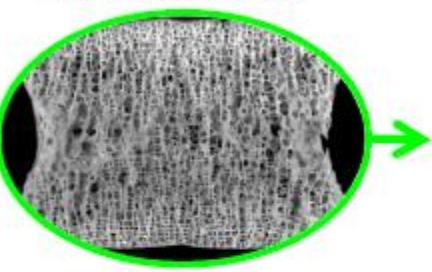


Illustration of a good microarchitecture



Identical BMD

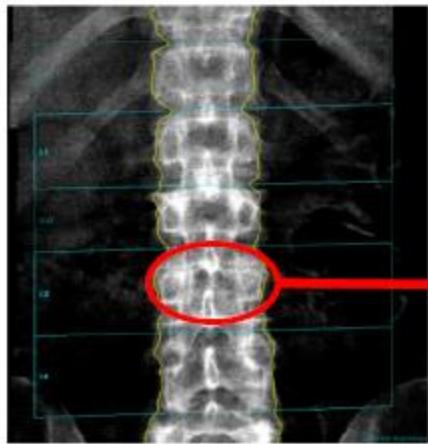
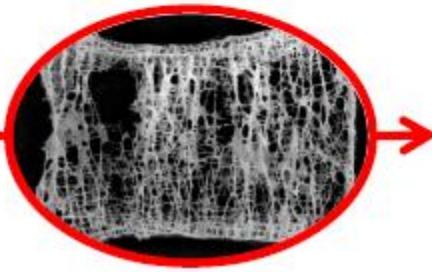
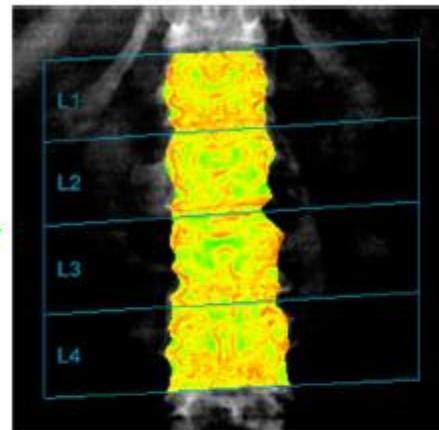


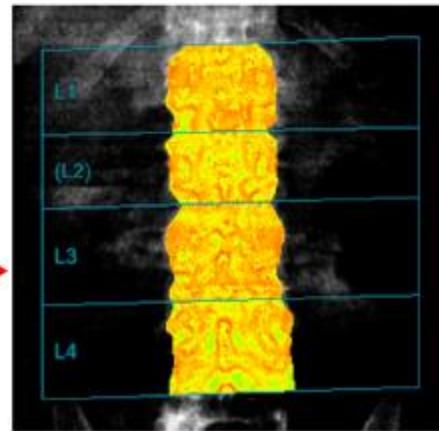
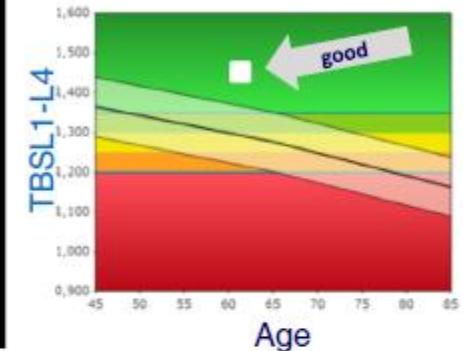
Illustration of a poor microarchitecture



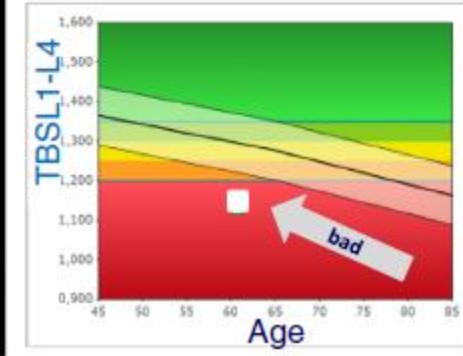
TBS Images and associated reference curves



TBS L1-L4: 1.457



TBS L1-L4: 1.132



Genome Big Data + Brain Big Data

NEUROGENOMICS

REVIEW

nature
neuroscience

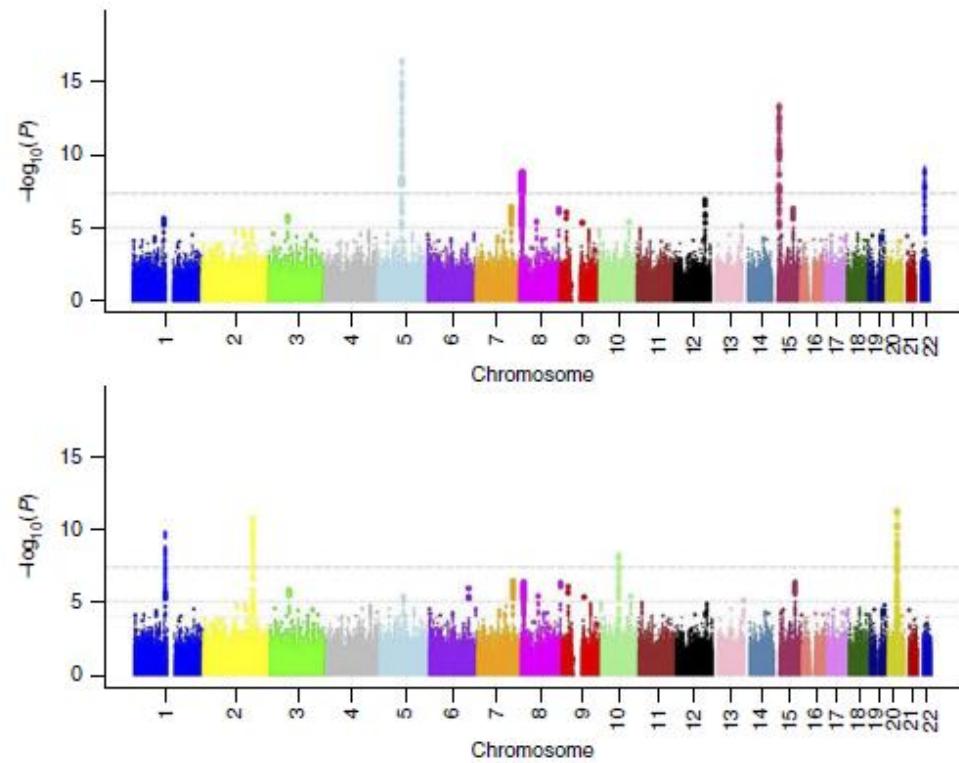
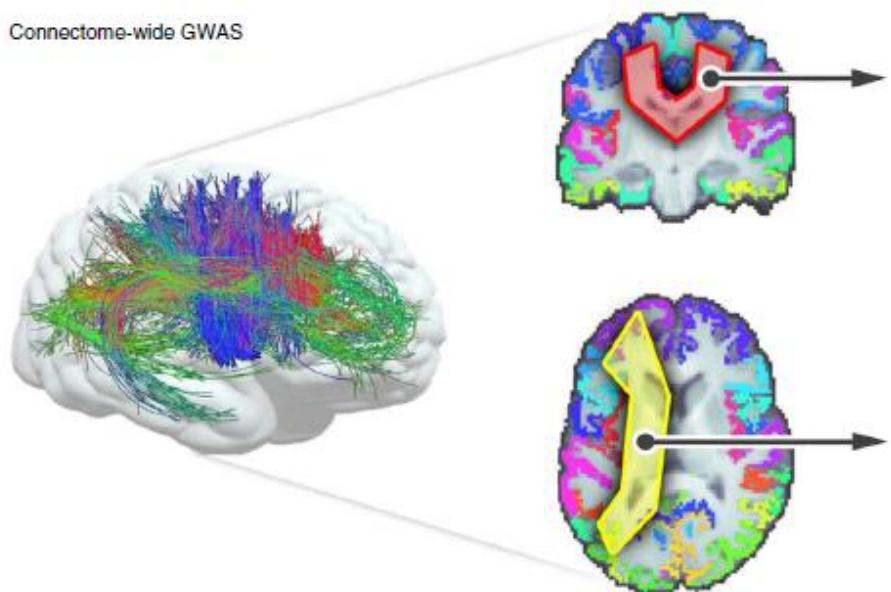
Whole-genome analyses of whole-brain data:
working within an expanded search space

Sarah E Medland^{1,13}, Neda Jahanshad^{2,3,13}, Benjamin M Neale^{4–6} & Paul M Thompson^{2,3,7–12}

Connectome-wide GWAS

b

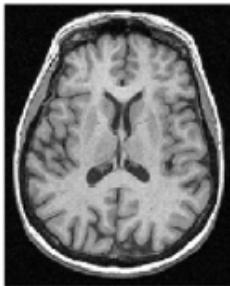
Connectome-wide GWAS



Imaging Protocols



Participant 1



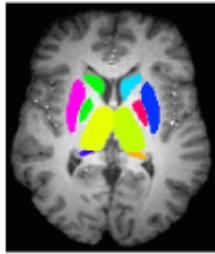
Participant N

T_1 -weighted structural images

MRI Image Processing



Freesurfer or FSL



Subcortical segmentation

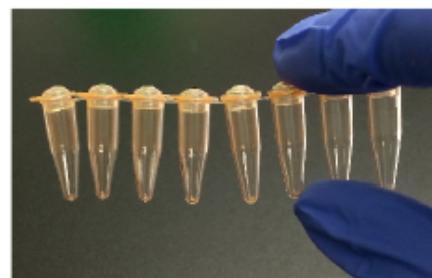


Registration to template to calculate ICV



Subcortical volumes calculated from segmentation

Genetics Protocols



Biological specimens collected for genotyping

Genotyping using commercially available platforms



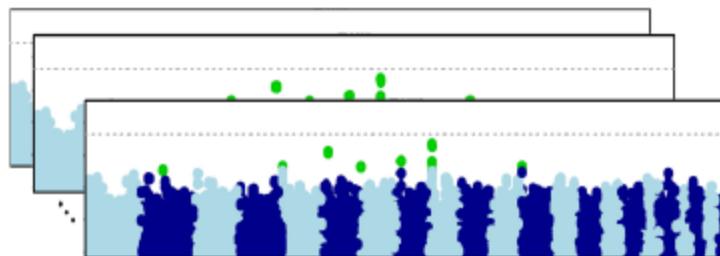
AGGAAGGGTTGGAACCCCGGGGCGGGGTTC
TTGGAAAAACCAAGAAATCAATCCTAGGGGTC
TGAGGGCTTTAGTTACCCCTCTGGGCCCGG
AATTTCGGGGCCAGAAATCGAAAATGGGAGT
TAATCTTACCTCCTCTGTAGCTTCTCCAA
CCGGCTTCCAAAGATCTCTTCCCCTGGGTGCG
GCCTTGGAGGACCTTAGAGGGGACTGGCCAA
AAAAAAATTCTGAATCCCTCTGGCTAGTTG
GAGAAGGCCAGTCGGTTGATC GGAGGGAAA
GAGGTTTTCCCCCTTTGGCCAAAAGGTTG
GTTTCTCCCCGGAGGTCTCCAAAACCAAGG
TTTTGAGATGGAAACTCTCTCGGCCGAAGT
CGAGGTCTGCCGGTCTTTTTCGGTTAG
TTTGAGAGAAGGGCTAGTGCTTCCAACCA
GGCCGGCCCTCTAGCCGAGGGGAGGAAAAG
GAGGTCTGGGTCTGGTCCCCAAGATCTTAGC
TACGACCAGGCCCAATCGGTCAACAACCAAG
AGTCGAGGCCAATCAGGGTTAACCTGGACAG
GTTCCCCCGGTTCCGGCCGAAGGGACCGGGGA
TTAAAGCCAAGAGGGGAATTTTAGCCCTTAAC
CCCTCGGAGTCCCCCAAGGTCAAAAGACTAA

Genome-wide genotypes were imputed to EUR reference panel from the 1000 Genomes Project (Phase 1 Version 3) following standardized protocols

nature

Published online
21 January
2015

Genome-wide association to imaging phenotypes (accounting for kinship in related samples)
(50 contributing sites with maximum N = 30,717)



Phenotypes (volume):

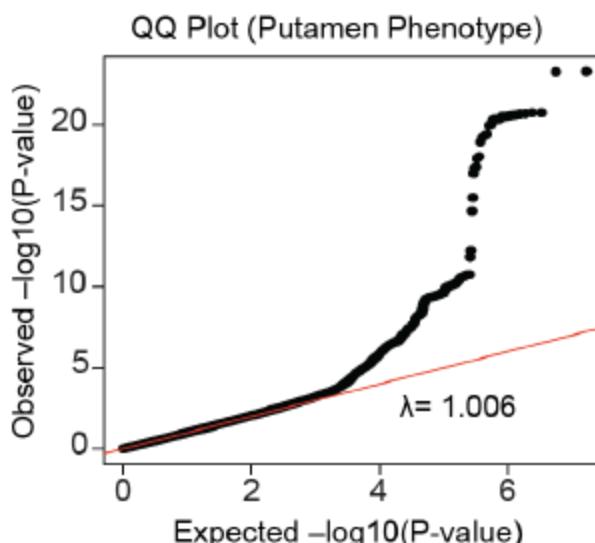
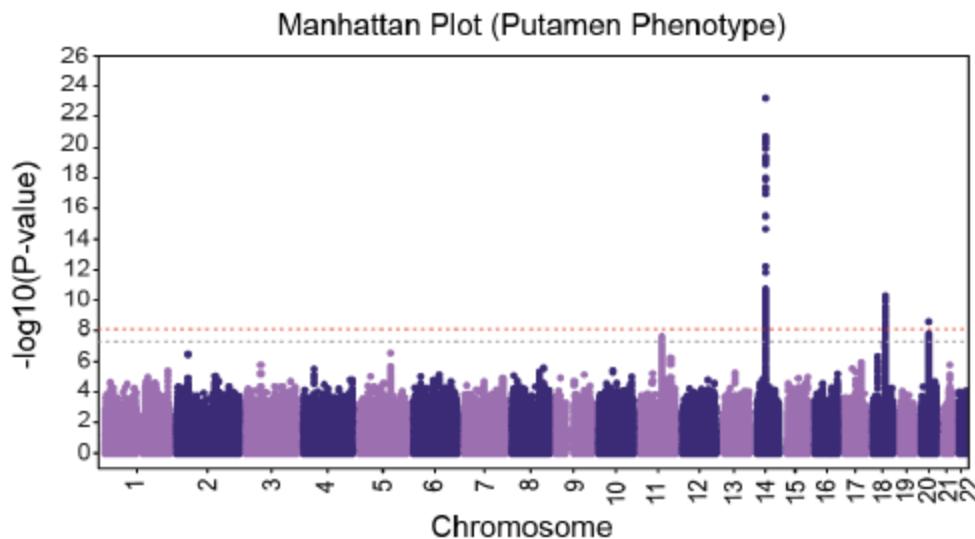
- ICV
- Nucleus Accumbens
- Amygdala
- Caudate
- Hippocampus
- Pallidum
- Putamen
- Thalamus

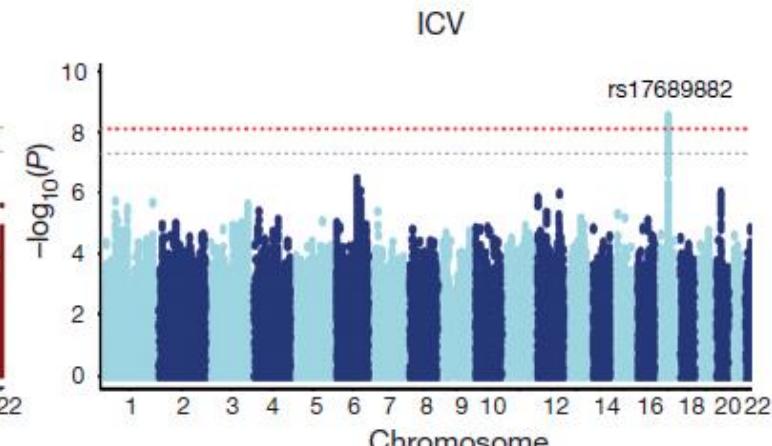
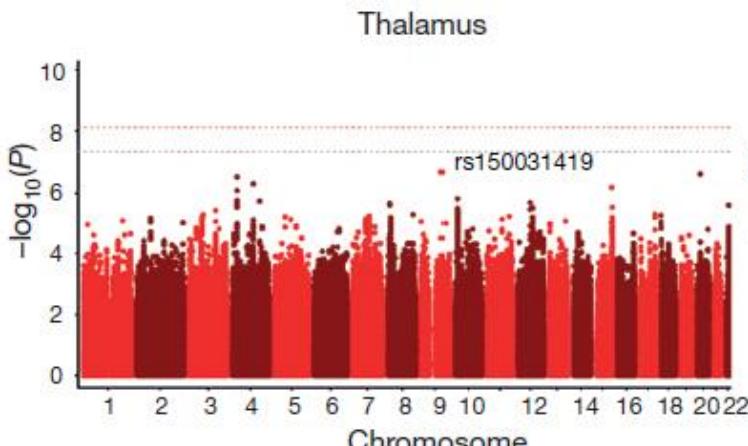
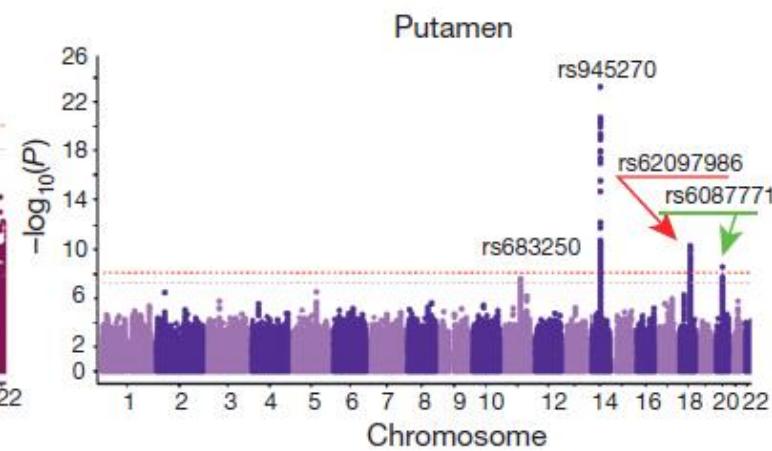
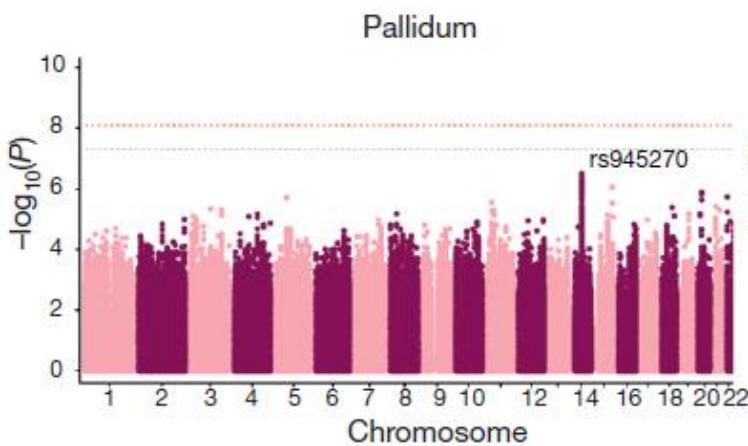
Covariates:

- ICV (for non-ICV phenotypes)
- Age
- Sex
- Age-squared
- 4 MDS
- Dummy covariate for scanner
- Patient status

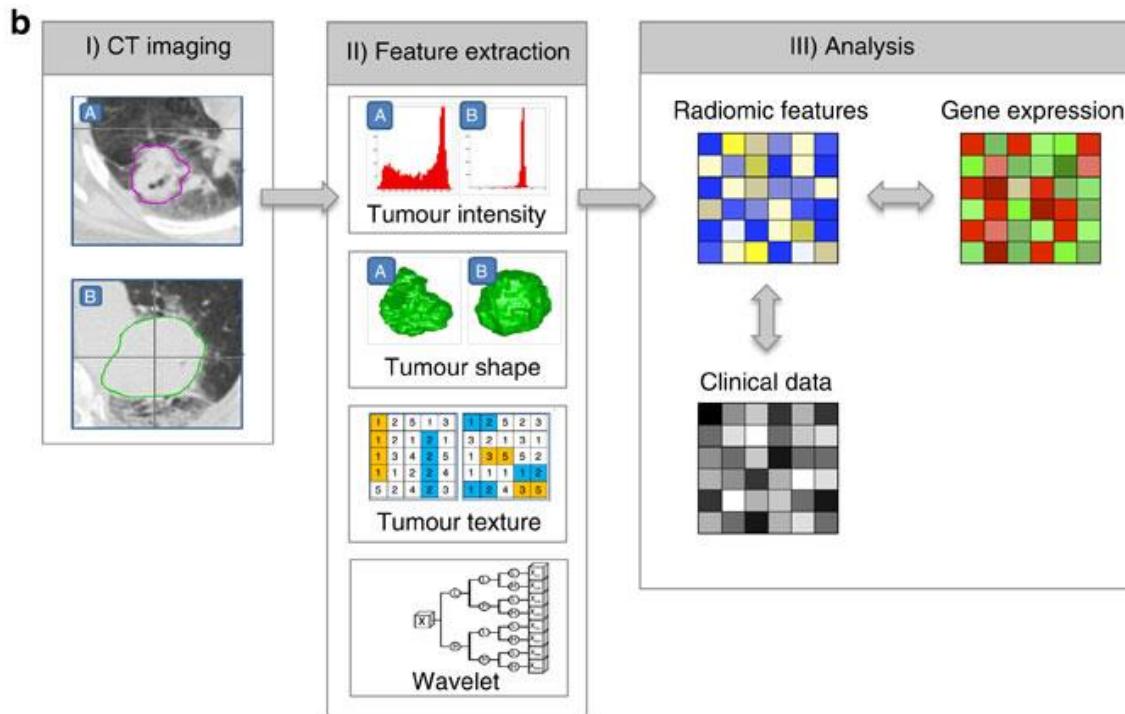
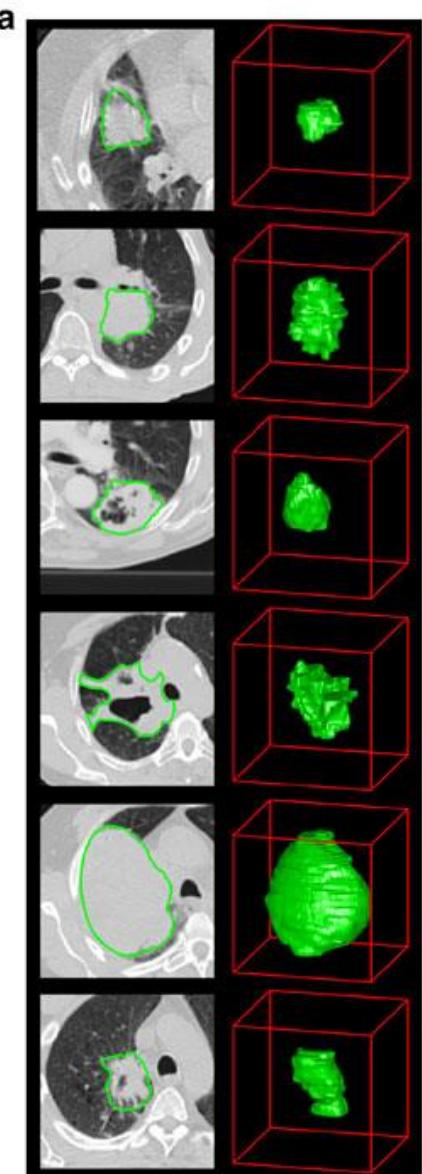
Quality Checking and Filtering (MAC < 10, R² < 0.5)

Meta-analyses
Fixed-effect, inverse variance-weighted model





Radiomics



Computer Whole Slide Image Analysis

a

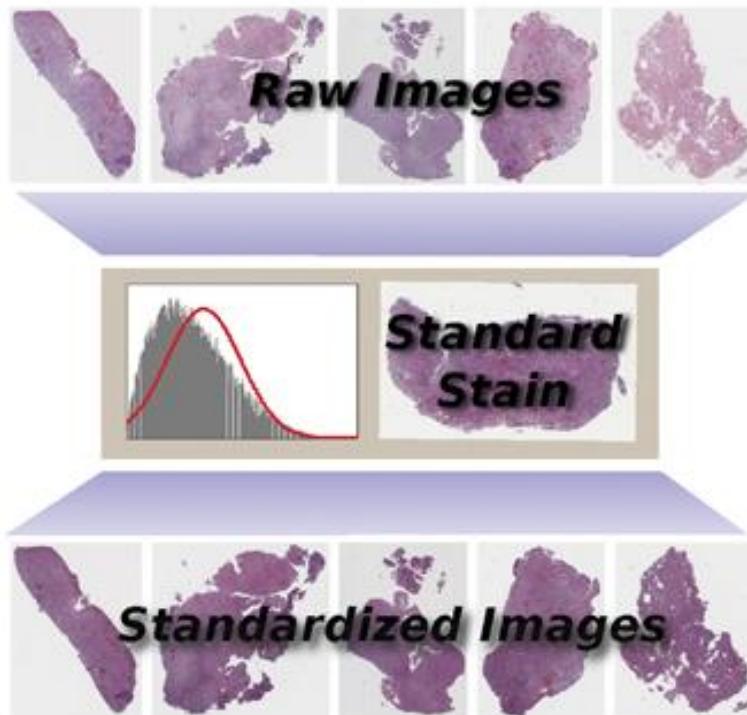


Image normalization

b

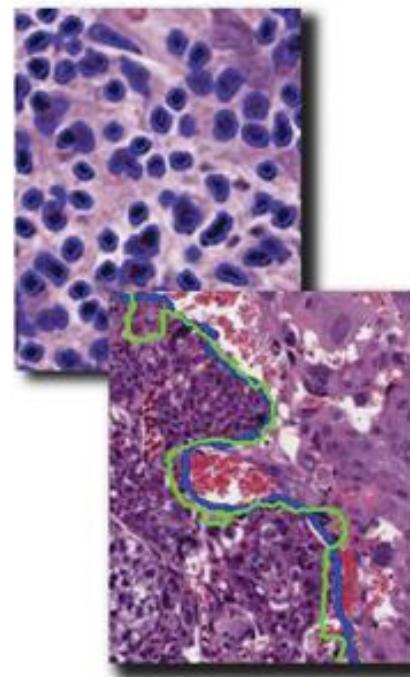
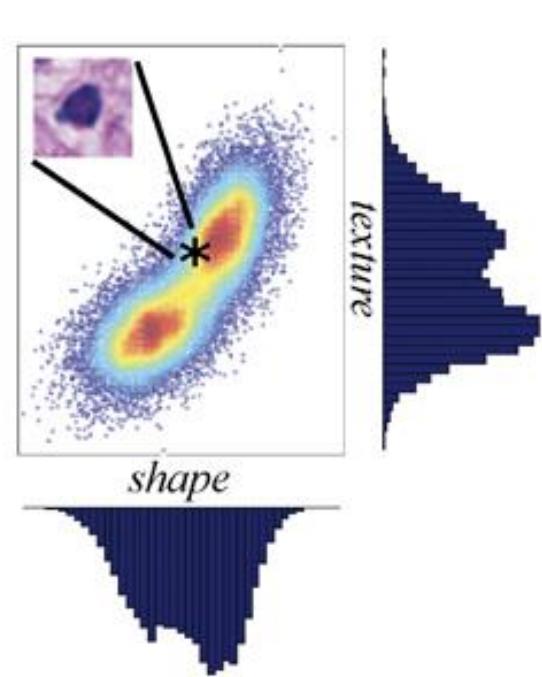


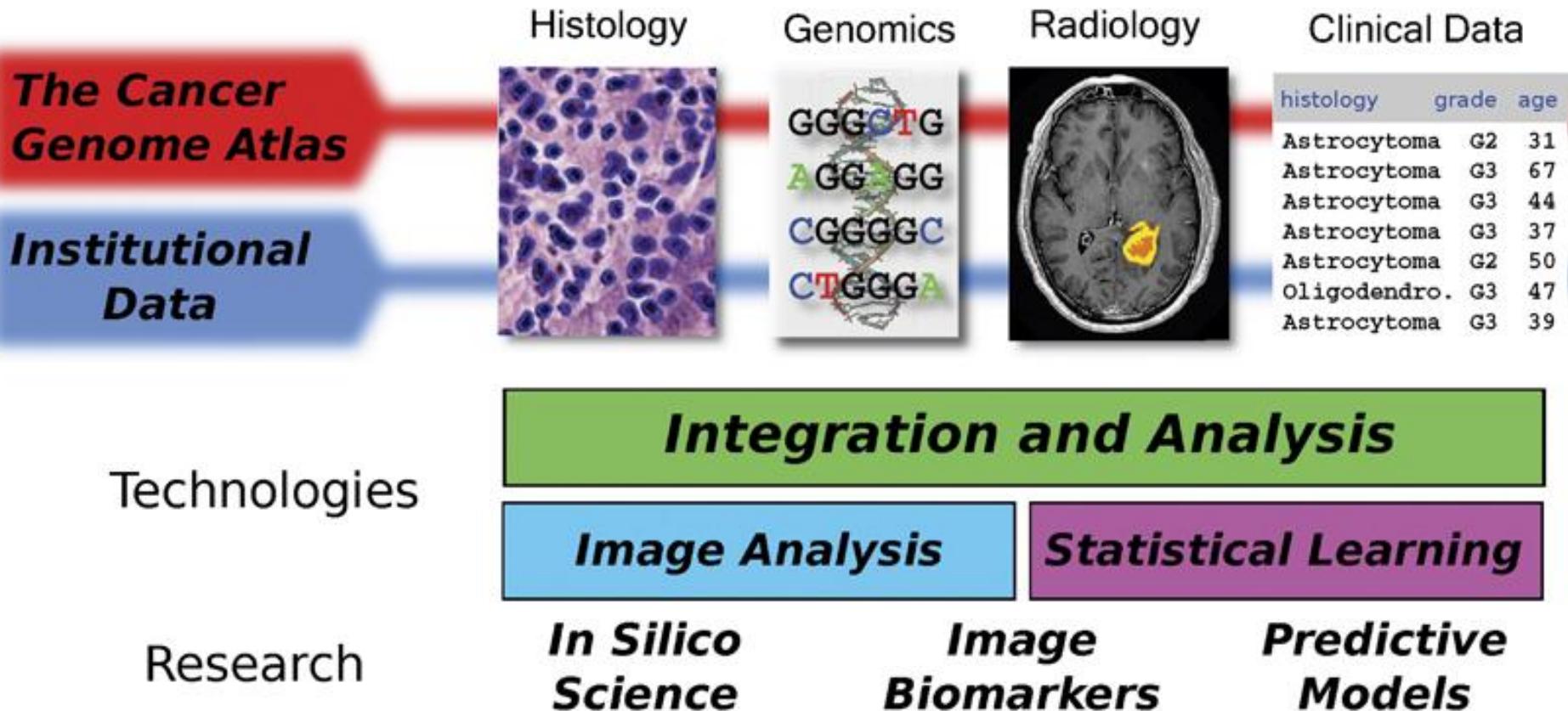
Image segmentation

c

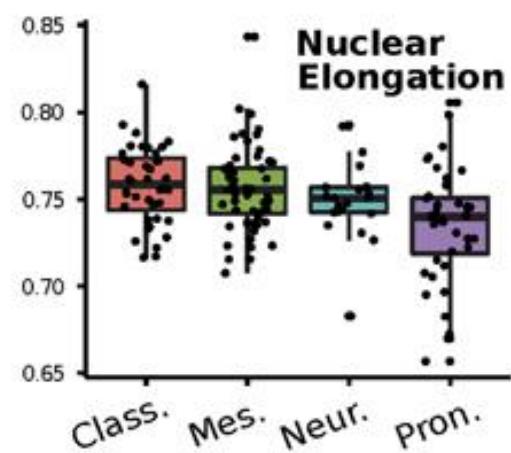
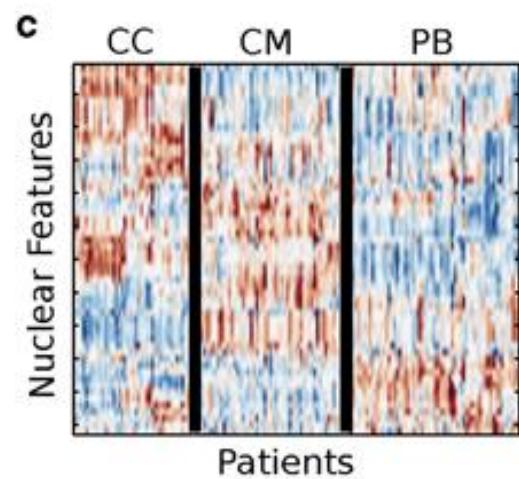
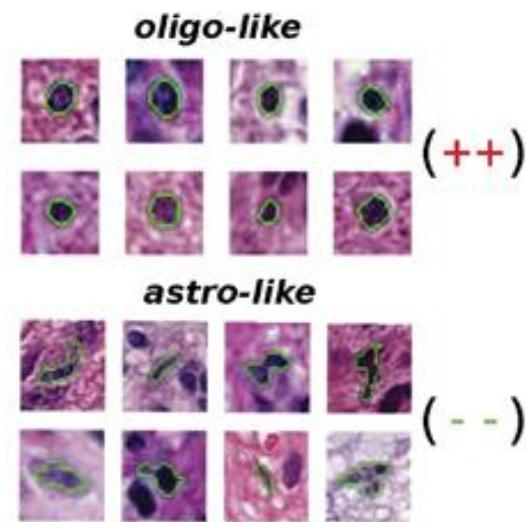
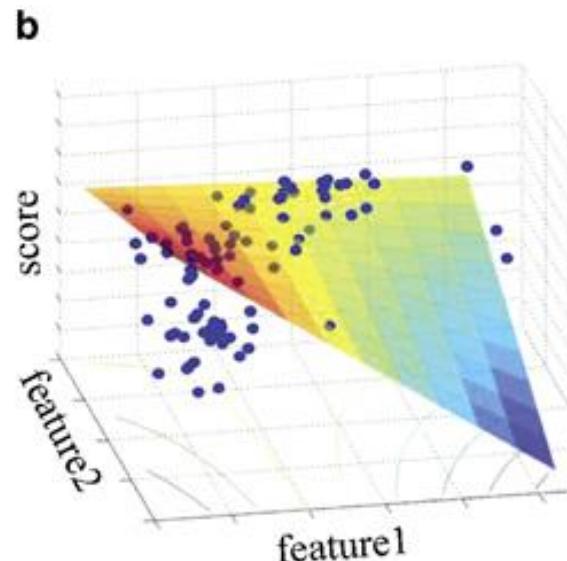
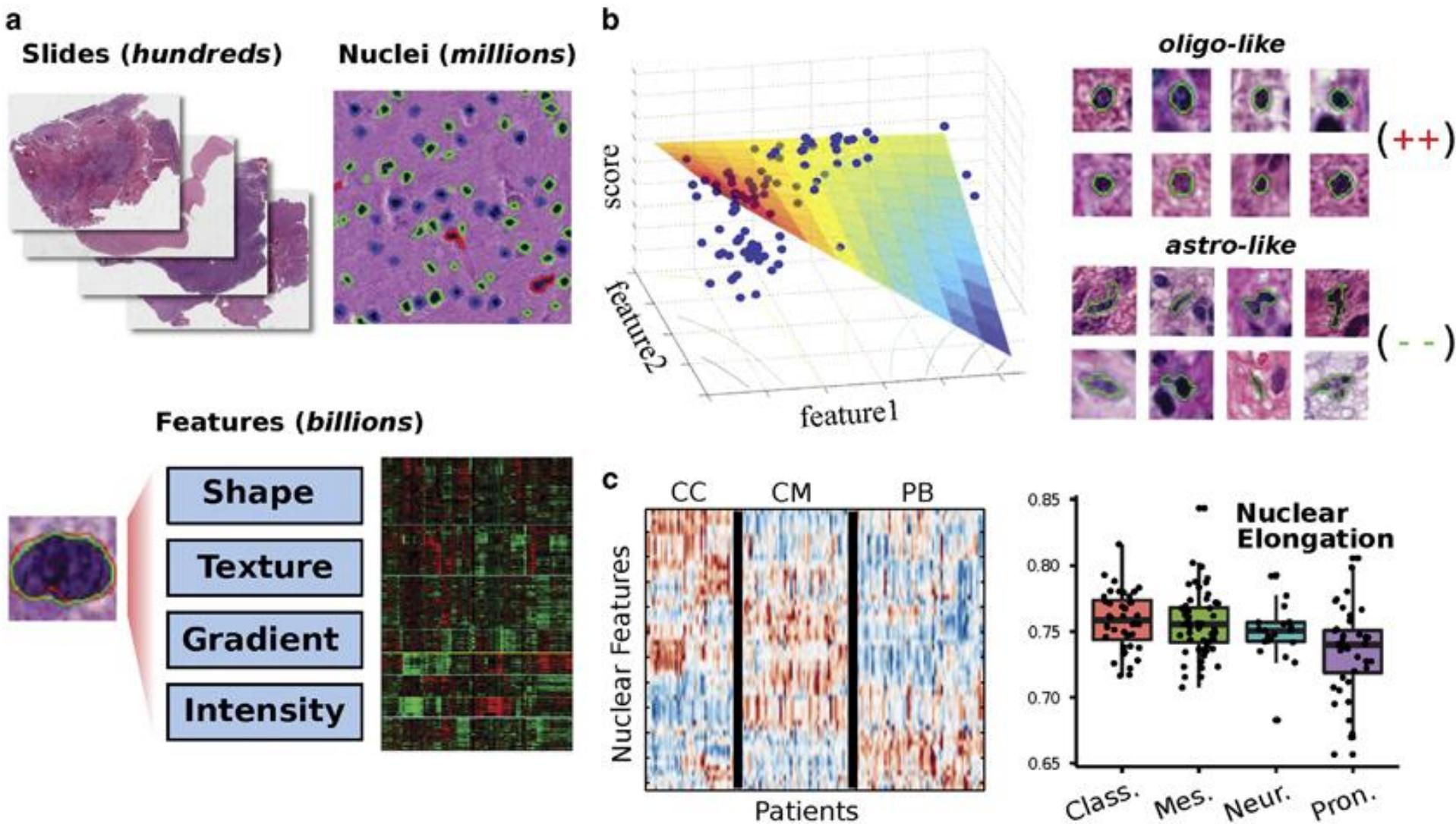


Feature extraction

Integration of quantitative histology with multifaceted clinical and genomic data



Quantitative nuclear morphometry



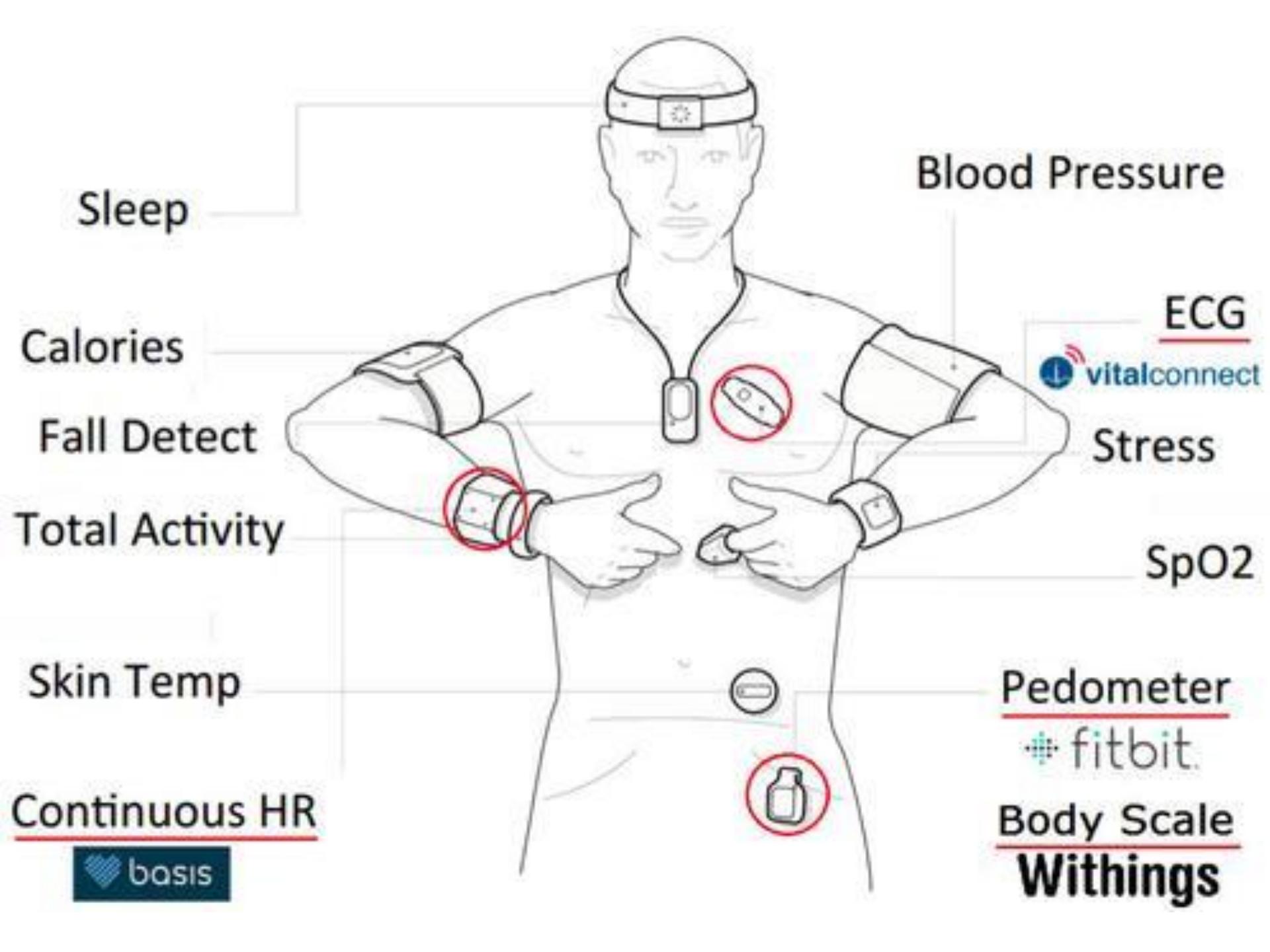
Contents

1. What is Healthcare Big Data?
2. Healthcare Big Data

- ① Genetic Data
- ② Electrical Health Records
- ③ National Healthcare Data
- ④ Medical Images
- ⑤ Sensor/Mobile Data

3. Clinical and Research Applications





Samsung Digital Health Platform



NFC 혈당측정기



)) NFC))



데이터베이스 기반 혈당관리

홈 | 내정보관리 | 사이트맵

혈당관리에 센스가 필요한 시간
센스다이어리

충북대병원 / 최형진님

당지킴이_예제보기 GO
당지킴이_예제보기 0개
[New] (멘토) 3개
0개
0개

고 김 김 남 박 전 최 최 최 최 활

27 28 29 30 31

혈당값다운로드

당뇨수첩입력

결과보기

혈당관리안내

커뮤니티

SENSDIARY 01 >
혈당값다운로드

혈당기 속 나의 혈당을 온라인
센스다이어리에서 관리하세요.



SENSDIARY 02 >
당뇨수첩입력

당뇨 관리, 이제는 쉽게하세요.
당뇨 치료, 관리가 최선입니다.



SENSDIARY 03 >
결과보기

내 혈당관리가 잘 되고 있는지
센스다이어리에서 확인하세요



1형당뇨혈당값

2형당뇨혈당값

그래프보기

혈당데이터 요약

혈당측정값전체보기

나의통계보기

메모보내기

▶ 내용보기

최근14일



2014-07-09



2014-07-23



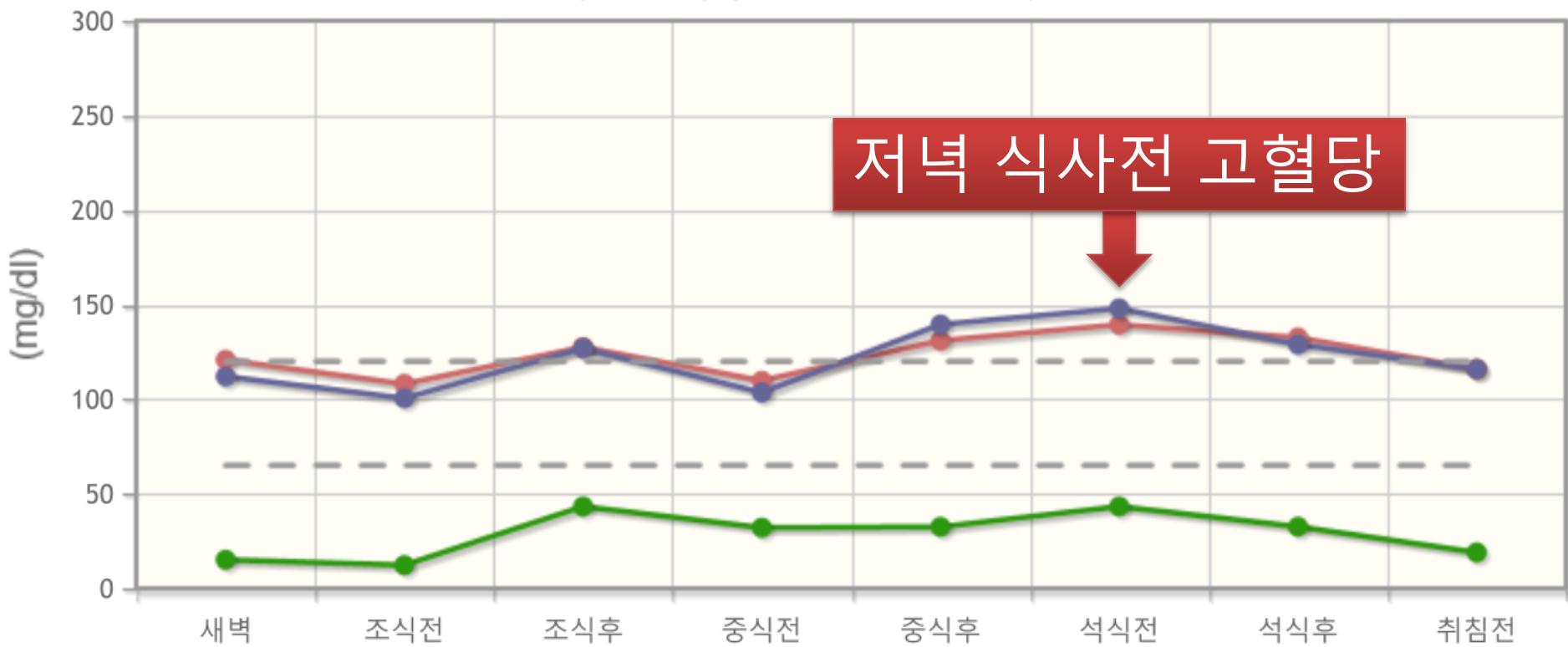
검색 >

날짜	새벽	조식전	조식후	중식전	중식후	석식전	석식후	취침전	메모
2014-07-23	0	93	0	0	0	0	0	0	
메모									
2014-07-22	112	108	151	111	142	0	129	0	
메모									
2014-07-21	0	105	127	119	108	0	138	0	
메모									
2014-07-20	109	115	169	88	130	109	113	0	
메모									
2014-07-19	125	134	117	101	83	0	119	0	
메모									
2014-07-18	136	0	155	116	191	199	191	0	
메모									
2014-07-17	0	0	119	116	0	190	99	0	
메모									
혈당값평균	121	111	140	109	131	166	132		고:주황색 저:청색 (단위 : mg/dL)

혈당 변화 실시간 모니터링

[2014-07-15 ~ 2014-07-29]

최근14일내 평균값 중앙값 표준편차



구글 헬스 앱 분야 매출 1위 '눔(noom)'



12:06 12:06 12:06

운동 음식 검색 운동

다이어트와 운동

김치

맑은 국, 탕류
김차국

기타 멀티성분 노란색 음식
김차진

찌개, 전골, 크림스프류
김치찌개

밥요리,면요리
김치김밥

저지방, 무지방 양념소스
김치국물

소세지,베이컨,가공육 또는 장기(간, 신장 등)
김치소세지볶음

운동정보

그래프

걸음 2569

운동거리 1.46 마일

칼로리 138

속도 3.1

전원 버튼을 눌러 화면을 꽉십시오.

업종 0:13:58

섭취하신 칼로리량

100 kcal

식단 분포도 ②

사용자의 아침



100% 0% 0%

이상적 비율



50% 35% 15%

색상 분포도 이해하기**핸드폰을 꼭 챙겨주세요**

Noom 코치는 회원님의 걷는 횟수를 매일 측정합니다. 밧데리 소모량도 크지 않으니 핸드폰을 바지 주머니 또는 가방에 쑥 넣고 다니세요.


 목표 걸음수에 도달하면 알림받기

Date	Time	name	foodType	calories	unit	amount
2014-08-09	0	미역국	0	23	1국그릇 (300ml)	105 g
2014-08-09	0	잡곡밥	0	80	1/4공기 (52.5g)	52 g
2014-08-09	0	열무김치	0	3	1/4소접시 (8.75g)	9 g
2014-08-09	0	파프리카	0	6	1/2개 (33.25g)	35 g
2014-08-09	0	토란대무침	0	28	1/2소접시(46.5g)	46 g
2014-08-09	1	복숭아	0	91	1개 (269g)	268 g
2014-08-09	2	마른오징어	2	88	1/4마리 (25g)	25 g
2014-08-09	2	파프리카	0	6	1/2개 (33.25g)	35 g
2014-08-09	2	저지방우유	1	72	1컵 (200ml)	180 g
2014-08-09	2	복숭아	0	183	2개 (538g)	538 g
2014-08-09	3	복숭아	0	91	1개 (269g)	268 g
2014-08-09	3	파프리카	0	6	1/2개 (33.25g)	35 g
2014-08-09	4	파프리카	0	6	1/2개 (33.25g)	35 g
2014-08-09	4	식빵	1	92	1장 (33g)	33 g
2014-08-09	4	삶은옥수수	1	197	1개 반 (150g)	150 g
2014-08-09	4	복숭아	0	91	1개 (269g)	268 g
2014-08-09	4	저지방우유	1	72	1컵 (200ml)	180 g
2014-08-10	0	복숭아	0	91	1개 (269g)	268 g
2014-08-10	0	저지방우유	1	36	1/2컵 (100ml)	90 g
2014-08-10	0	두부	0	20	1/4인분 (25g)	25 g
2014-08-10	0	견과류	2	190	1/4 컵 (50g)	31 g
2014-08-10	0	파프리카	0	11	1개 (66.5g)	65 g
2014-08-10	2	방울토마토	0	8	4개 (60g)	60 g

식사량 실시간 모니터링

점심 과식

저녁 금식



운동량 실시간 모니터링

걸음 수

25000

20000

15000

10000

5000

0

2014-08-09

2014-08-10

2014-08-11

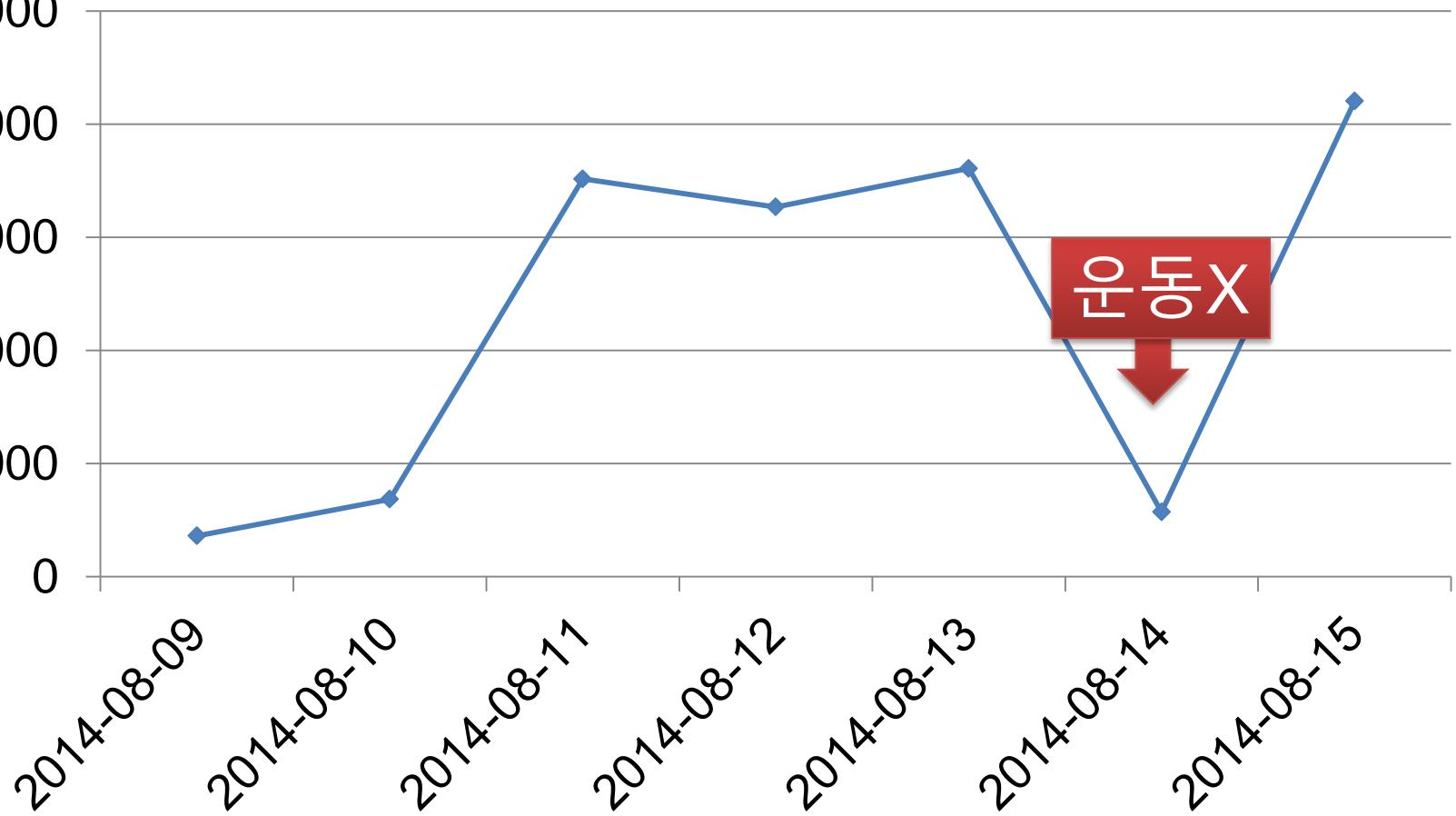
2014-08-12

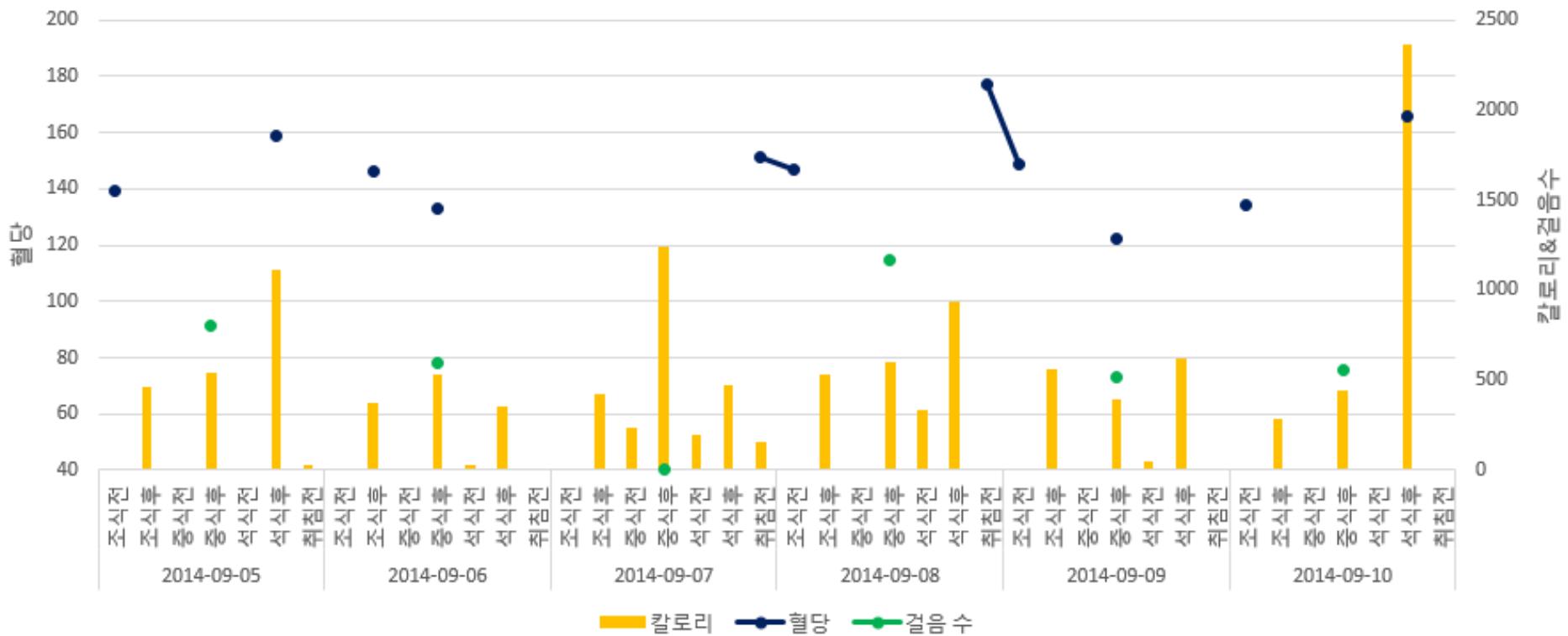
2014-08-13

2014-08-14

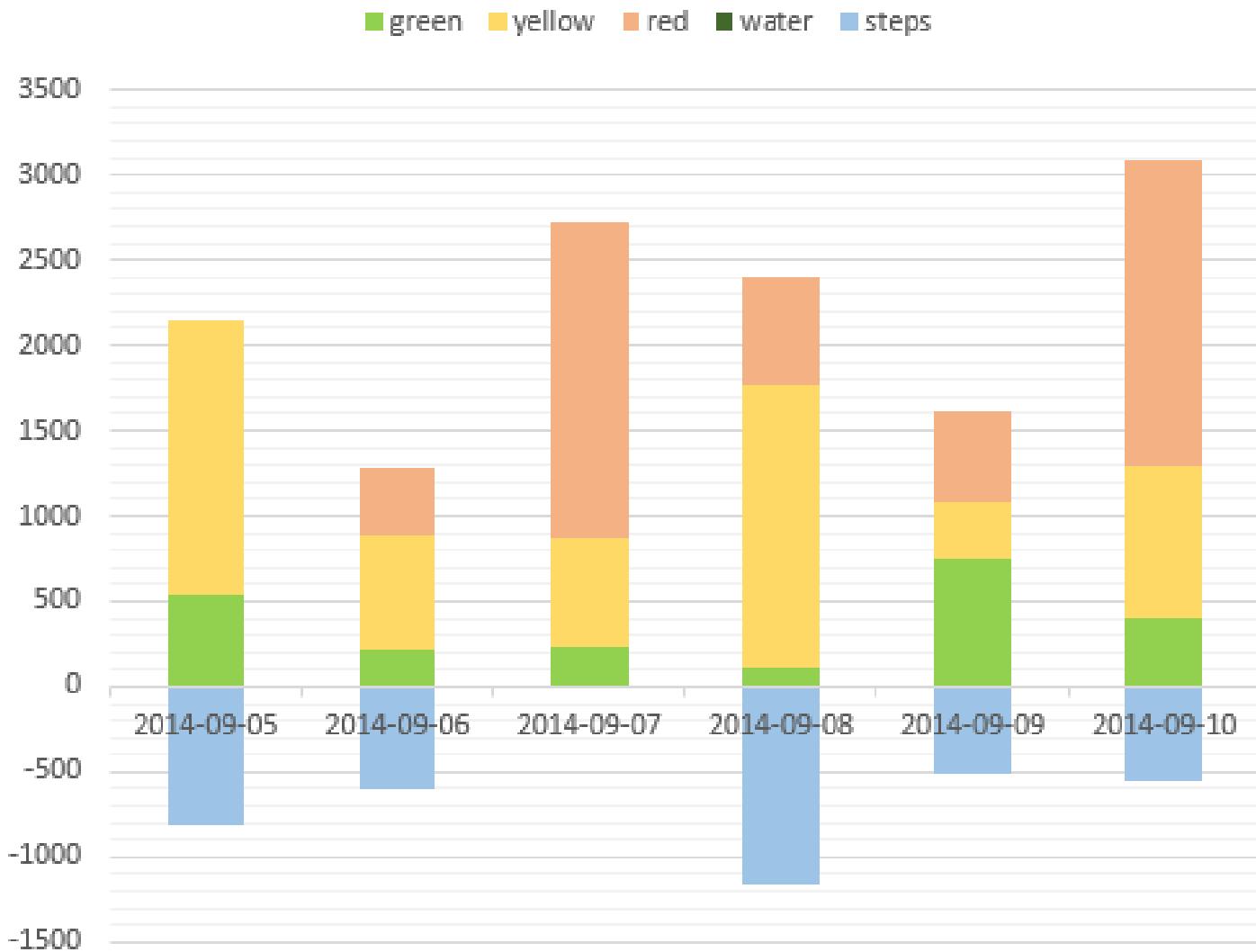
2014-08-15

운동X





식사
운동



스마트폰 활용 당뇨병 통합관리



진료 ↓ ↑ 처방
2-3달 간격

상담 ↓ ↑ 교육
매주/필요시

자가관리 ↗ 전송 ↘
조회/분석



의사

평가 회의
매주/필요시

교육간호사



최윤섭의 Healthcare Innovation

헬스케어, 바이오테크놀러지, 신약개발, IT & 헬스케어 컨버전스, 기업가정신

HOME DIGITAL HEALTHCARE PERSONALIZED I

피트니스 트레커를 이용한 환자 모니터링: Fitbit의 다 발성 경화증 치료 활용

Yoon Sup Choi

January 21, 2015

Big Data, Digital Healthcare

0 Comments



Like

Share

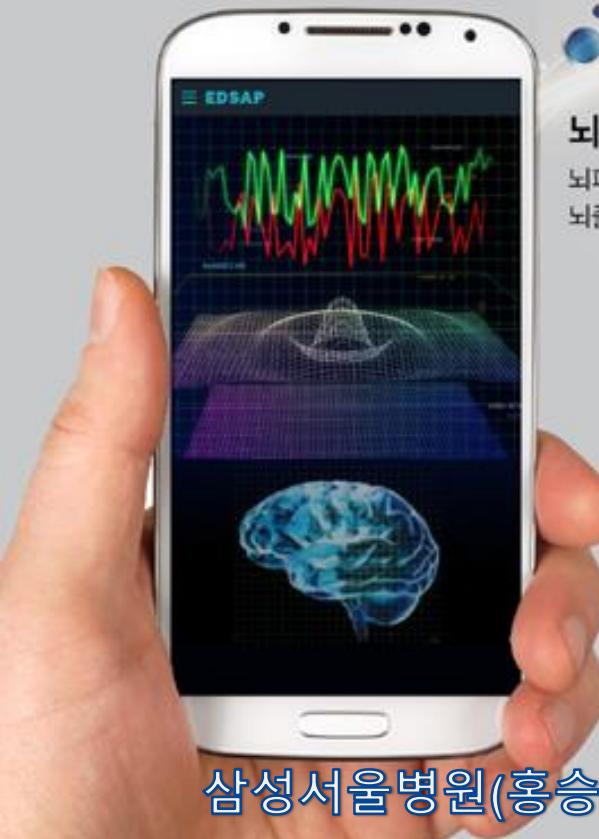
39 people like this. Be the first of your friends.

최근 **다국적 제약사 Biogen Idec은 피트니스 트레커 Fitbit** 를 자가면역질환의 일종인 **다발성 경화증 (multiple sclerosis, MS)** 환자들의 모니터링과 신약개발에 활용하겠다고 발표하였습니다.

뇌졸중 발생 위험을 감지하는 모바일 뇌파 센싱 시스템

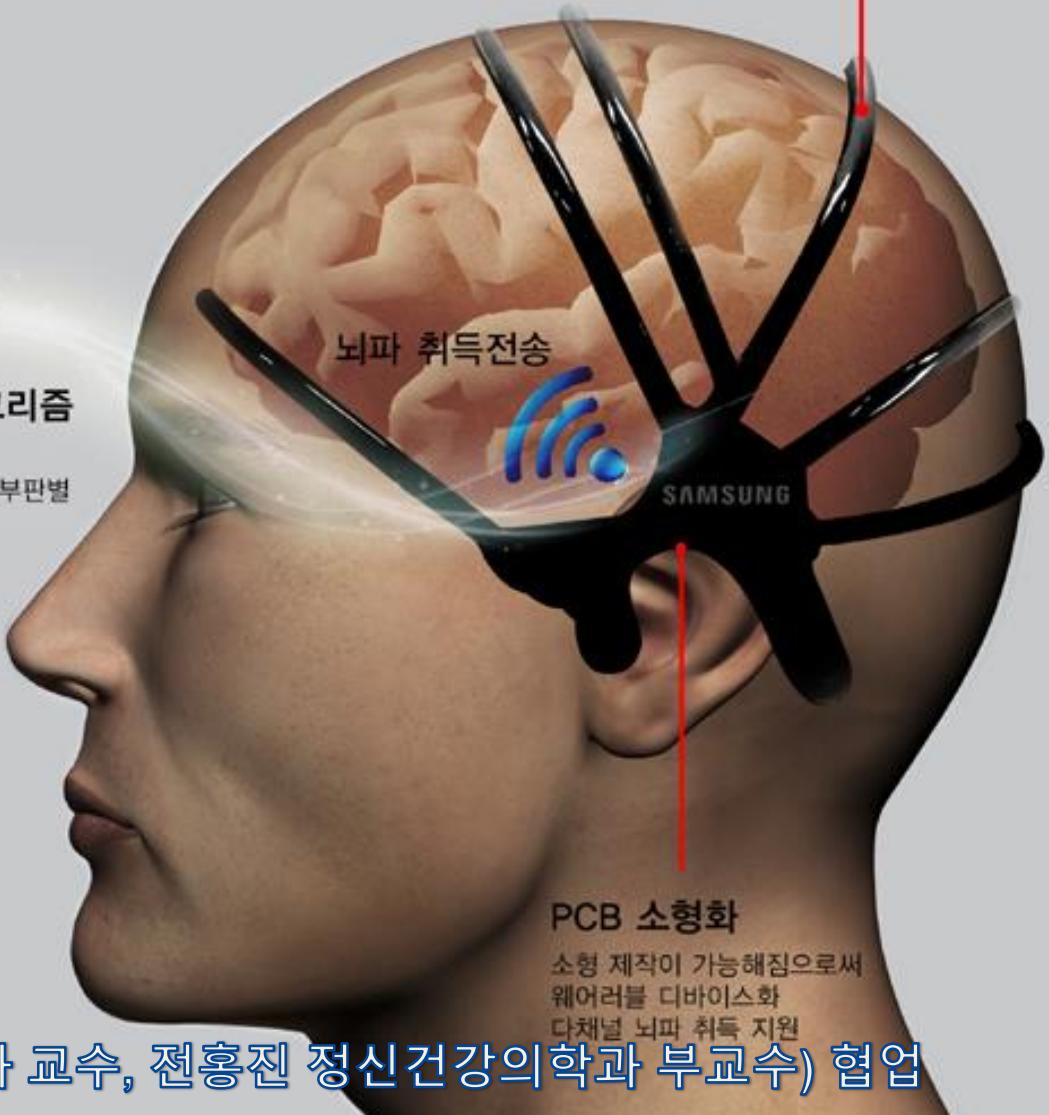
모바일 알람 솔루션

헤드셋에서 측정한 뇌파를 모바일에서
분석하여 위험여부 전달



뇌파 분석 알고리즘

뇌파를 분석하여
뇌졸중 발생 위험 여부판별



센서 개선

소프트 드라이 센서 적용으로
금속 센서 대비 착용감 대폭 향상

PCB 소형화

소형 제작이 가능해짐으로써
웨어러블 디바이스화
다채널 뇌파 취득 지원

삼성서울병원(홍승봉 신경과 교수, 전홍진 정신건강의학과 부교수) 협업

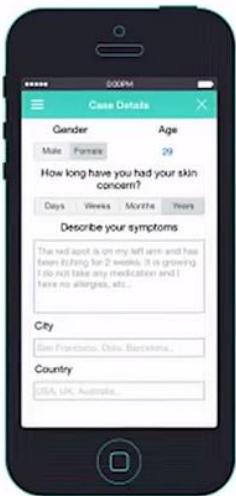
피부 사진 원격 진단/처방

How it Works



SNAP

Using our app our web form, take 2 pictures of the effected area- an overview and close-up photo.



SEND

Once your information has been filled out, simply pay the \$40 case fee, and your case will be sent to a First Derm Dermatologist.



REVIEW

Our dermatologists will carefully review your case and send you an initial triaging opinion. 70% of cases submitted to First Derm have been recommended to treat at home or with over the counter medications.

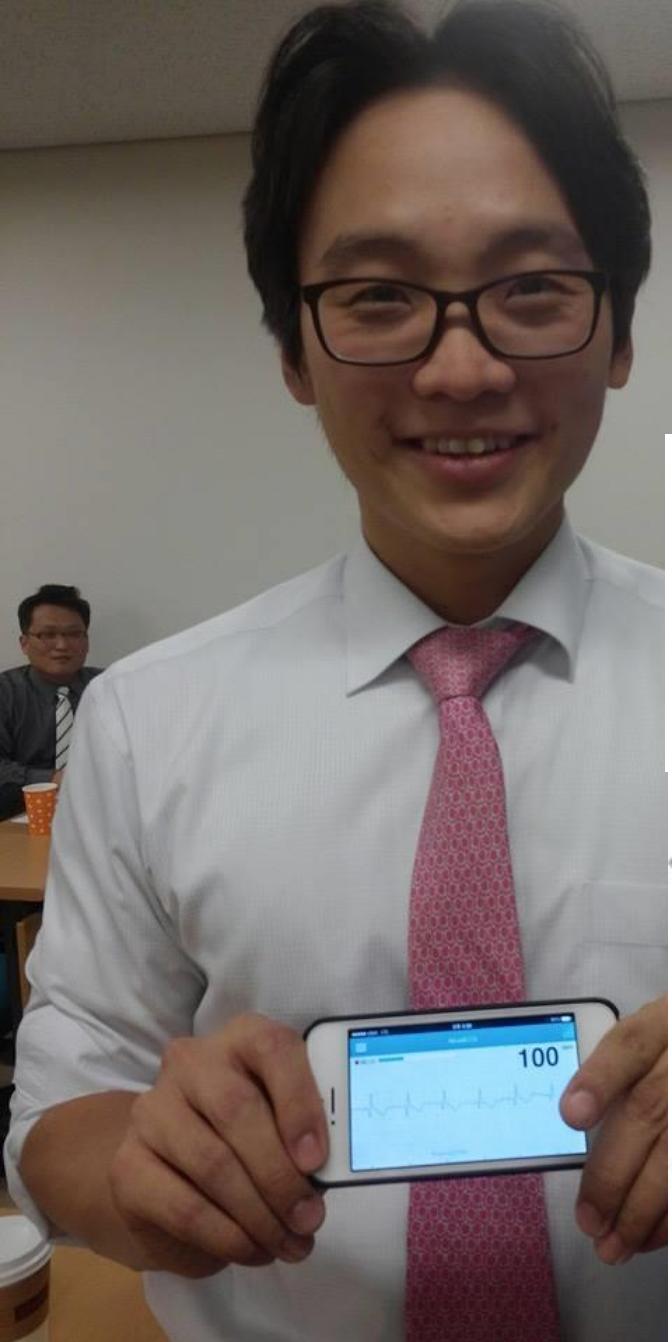


FIRST DERM

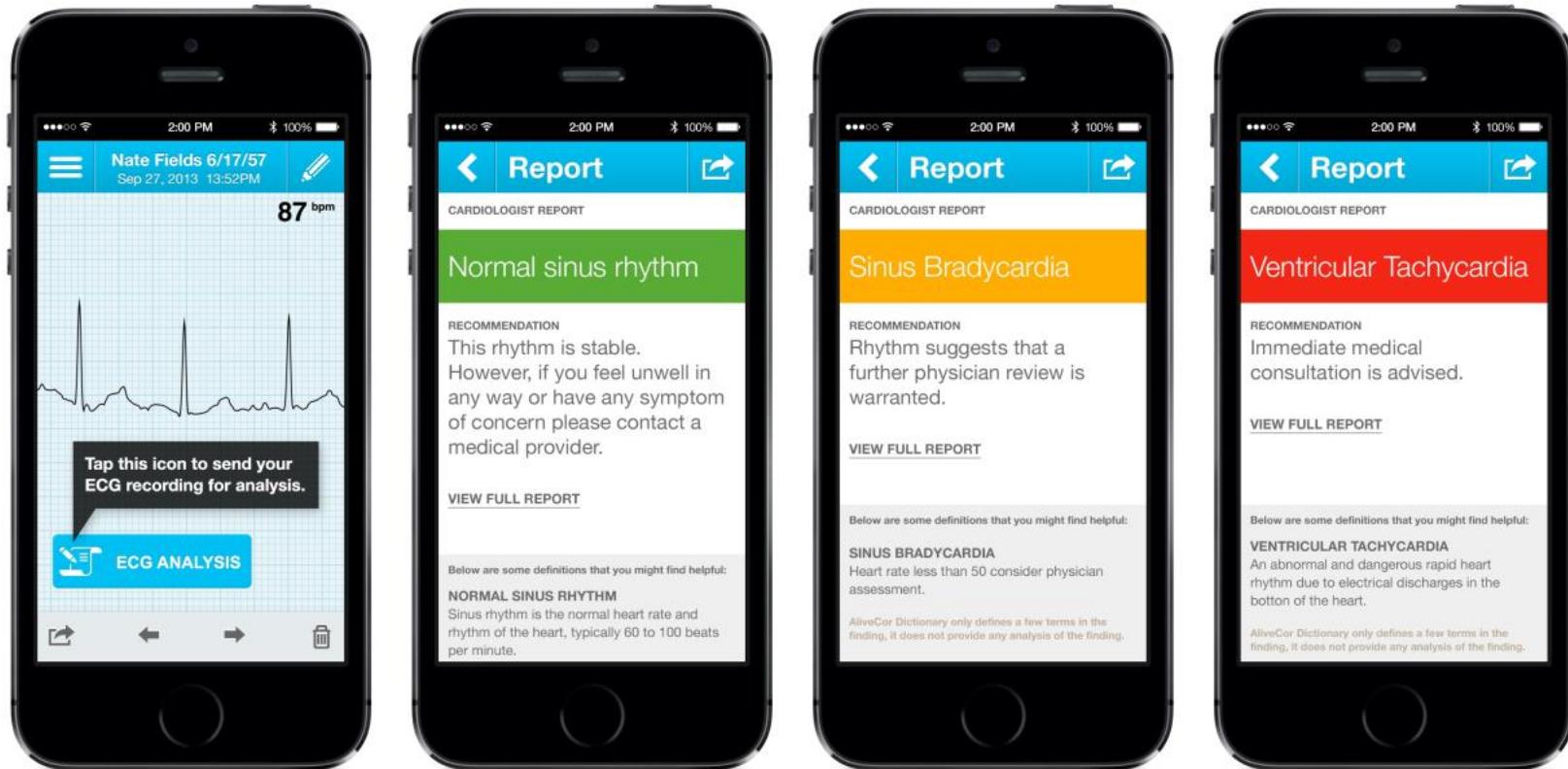


VISIT

In 30% of our cases, we recommend you see a doctor. Our app allows you to find your closest dermatologist or pediatrician.



심전도 원격 진단/처방



AliveInsights™ ECG Analysis Service

Contents

1. What is Healthcare Big Data?

2. Healthcare Big Data

- ① Genetic Data
- ② Electrical Health Records
- ③ National Healthcare Data
- ④ Medical Images
- ⑤ Sensor/Mobile Data

3. Clinical and Research Applications



Clinical and Research Applications

1. Clinical Decision Support
2. Public Health Surveillance
3. Personal Health Record
4. Healthcare Data Integration

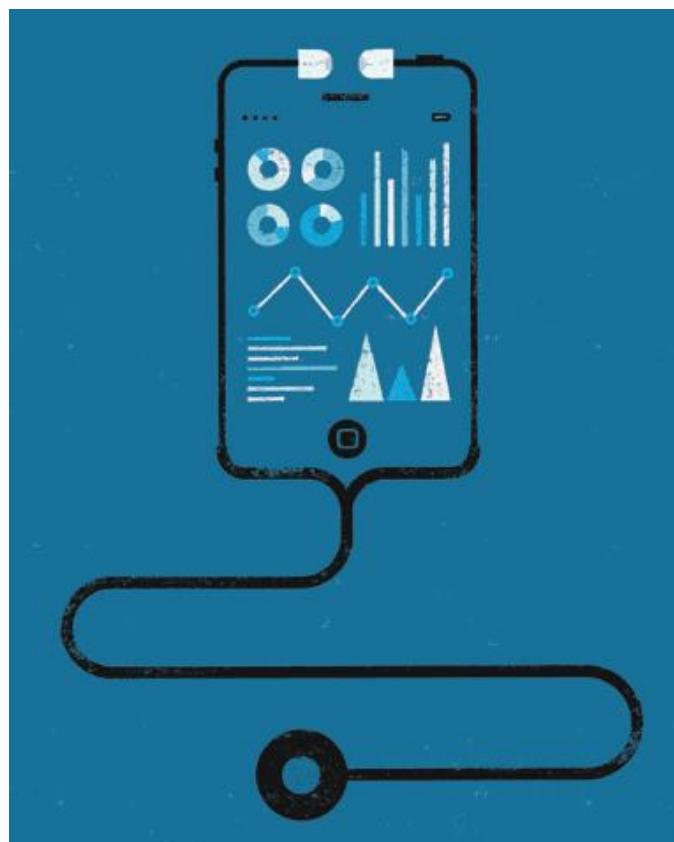
Clinical and Research Applications

1. Clinical Decision Support
2. Public Health Surveillance
3. Personal Health Record
4. Healthcare Data Integration



Data-Driven Health Care

New technologies promise a flood of molecular, environmental, and behavioral information about patients. Will all that data make medicine better?



뉴스

- ▶ 행정/법률
- ▶ 의원/병원
- ▶ 학술/학회
- ▶ 의대/전공의
- ▶ 제약
- ▶ **의료기기/IT**
- ▶ 약국/유통
- ▶ 월드뉴스
- ▶ 한방
- ▶ 간호
- ▶ 기획/특집
- ▶ e-談

홈 > 뉴스 > 의료기기/IT



처방 내리는 컴퓨터 '왓슨', 한국 상륙 임박

서울시, 왓슨 연구소와 업무협약 예정…국내 의료계 영향력 예고

2015.01.10 06:08 입력

미국 MD앤더슨 암 센터(MD Anderson Cancer Center) 의사들에게 최적의 치료법을 권고해 주는 IBM의 인지컴퓨팅 시스템 '왓슨'(Watson)이 한국에 들어올 것으로 보인다.

서울시의 '미 동부권 R&D 연구소 유치 계획(안)'에 따른 것인데, 최근 IBM 왓슨연구소와 드렉셀대학이 '스마트에이징'(Smart Aging) 분야 공동연구를 진행한다는 투자 의향을 전달했다.

시는 이르면 이달 중 업무협약이 체결될 예정이라고 밝혔다. 왓슨연구소가 서울에 들어설 경우 상암 DMC단지가 유력할 것이라는 전망이다. 향후 투자 규모 및 임대 조건 등은 MOU를 맺은 뒤 구체화 될 것으로 전해졌다.

Risk Prediction

As-Is



현재 환자모니터에서 기록되는
로그성 데이터는
단순히 수집만 되고 있으며,
활용이 되지 않음

To-Be



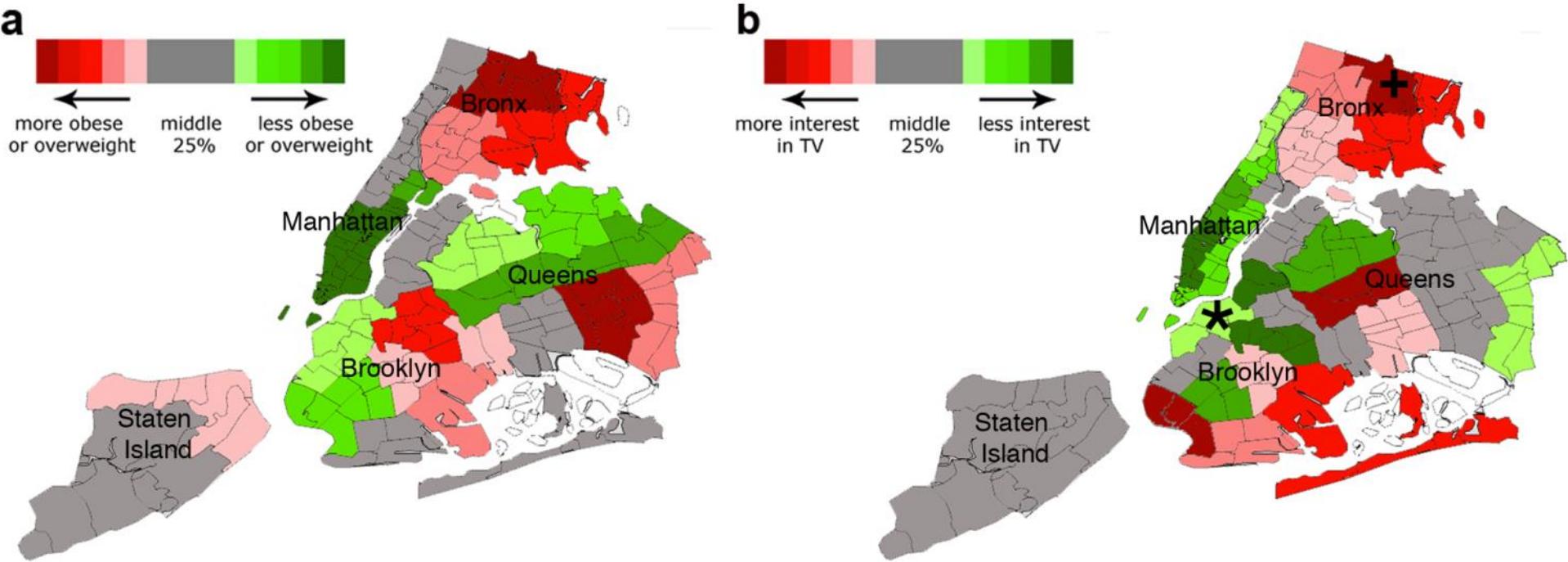
저장된 데이터를 분석,
심실부정맥 발병을 예측하여
중환자의 생존율 증가

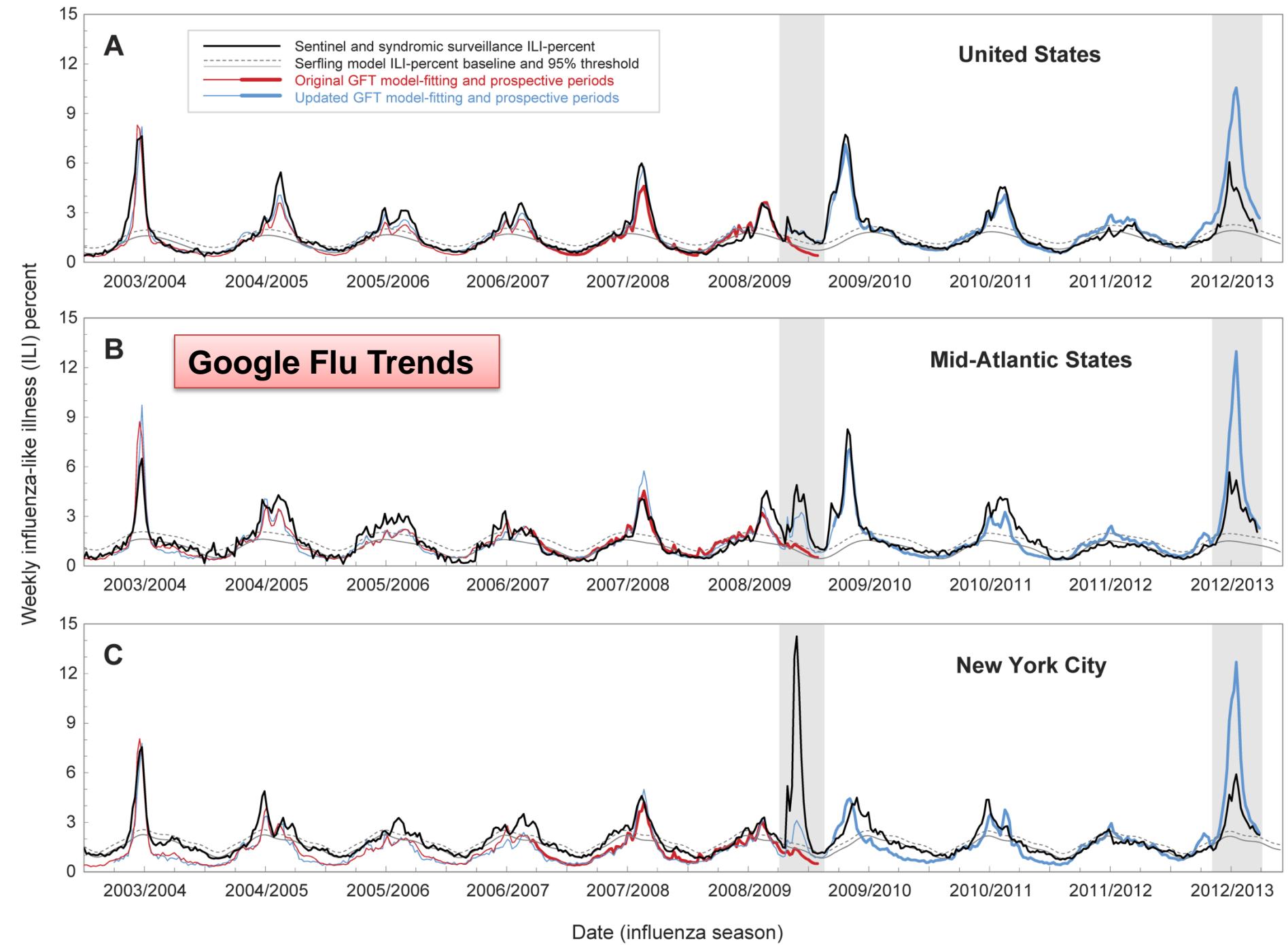
Clinical and Research Applications

1. Clinical Decision Support
2. Public Health Surveillance
3. Personal Health Record
4. Healthcare Data Integration



Social Network and Obesity Prevalence





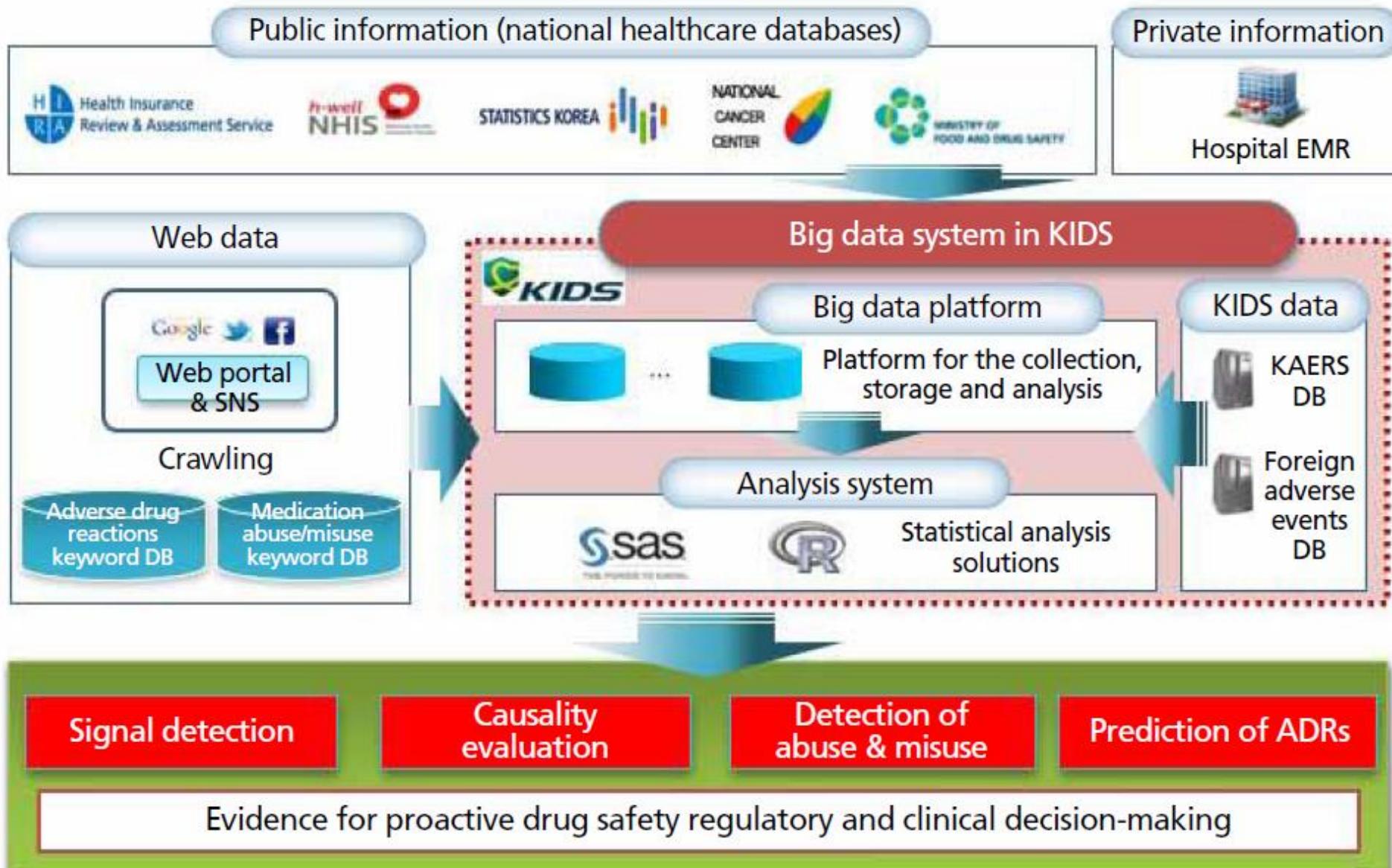
환경부, 건강보험 빅데이터 시스템과 연계

아토피·비염 등 환경질환 상관관계 찾는다

미현근 기자 bass007@dt.co.kr | 입력: 2014-12-02 15:13 | 수정: 2014-12-02 18:16

- 환경부 빅데이터 시스템 도입
- 비염, 천식, 아토피 등 환경질환 개선
- 국민건강보험공단의 진료 기록
- 기후 자료
- 국민건강보험공단 자료(국민건강정보 DB, 코호트 DB) 연계
- 이상기후·대기오염에 따른 건강영향 분석
- 지역별 환경 위험요인 자료 DB구축
- 위치정보를 이용한 환경성·심혈관계 질환 위해지도(risk map), 질병지도(disease map) 작성
- 국민건강보험공단 2002~2013 건강보험 가입자 전체 DB
- 100만명 장기추적자료
- 개인별 질환 악화·변화- 폭염/미세먼지 등 환경유해인자와 상관성 검증

Big data platform model by Korea Institute of Drug Safety and Risk Management



Clinical and Research Applications

1. Clinical Decision Support
2. Public Health Surveillance
3. Personal Health Record
4. Healthcare Data Integration



사이버상의 헬스 아바타



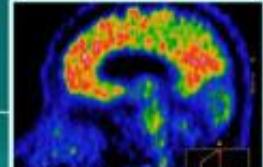
유전자 정보

특정 암 유발 또는 가족이
잘 걸리는 질병 유전자 자료 입력



심전도

U헬스 의료기기를 통해 집에서
체크한 검사 정보 자동 전송



의료영상

병원에서 질병 발견과 치료
목적으로 활용한 영상 전송



일상활동 정보

스마트폰 등으로 운동량,
설취 칼로리 양 자동 전송



진료 기록

병원에서 환자의 병세를 쉽게
알 수 있는 핵심 자료 입력



투약 기록

약국에서 환자가 구매한 약물
기록 입력

현실세계에서 평가

그래픽=양민성 기자 in7@chosun.com



의사

여러 진료과의 의사가
집속해서 환자가 어떤
질병에 걸릴 위험이 있는지
분석하고 치료 방향 제시.



간호사, 영양사,
운동처방사 등
환자의 질병 관리
문제점 분석, 생활
습관 개선 방향 제시.



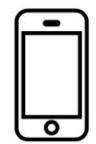
생명공학자
유전자 정보 분석하여
어떤 질병이 위험하고
무슨 암을 조심해야
하는지 예측.



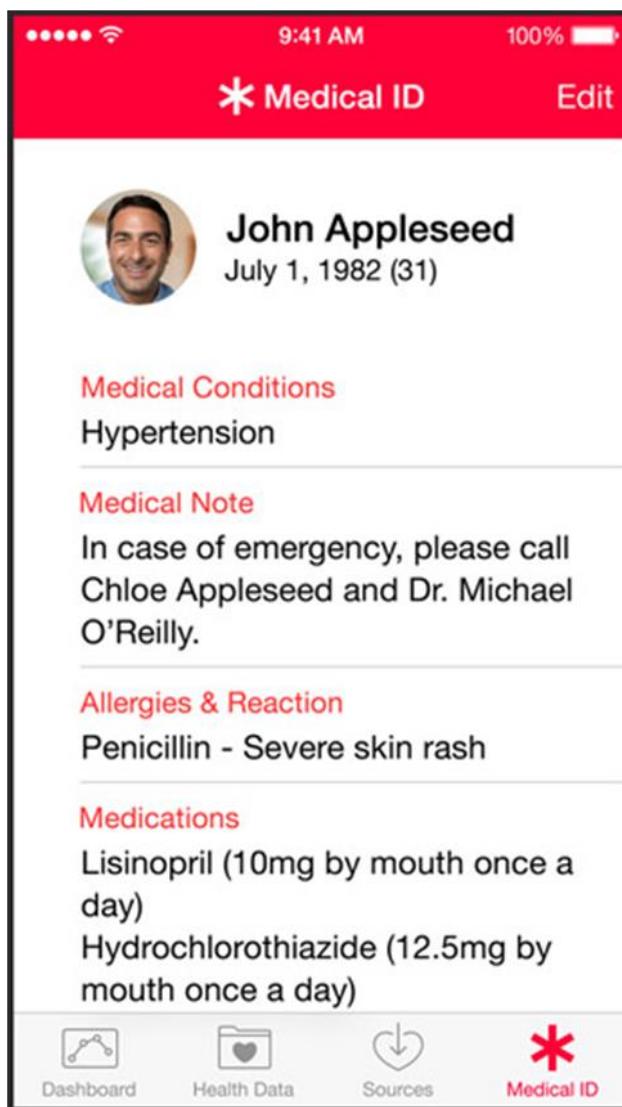
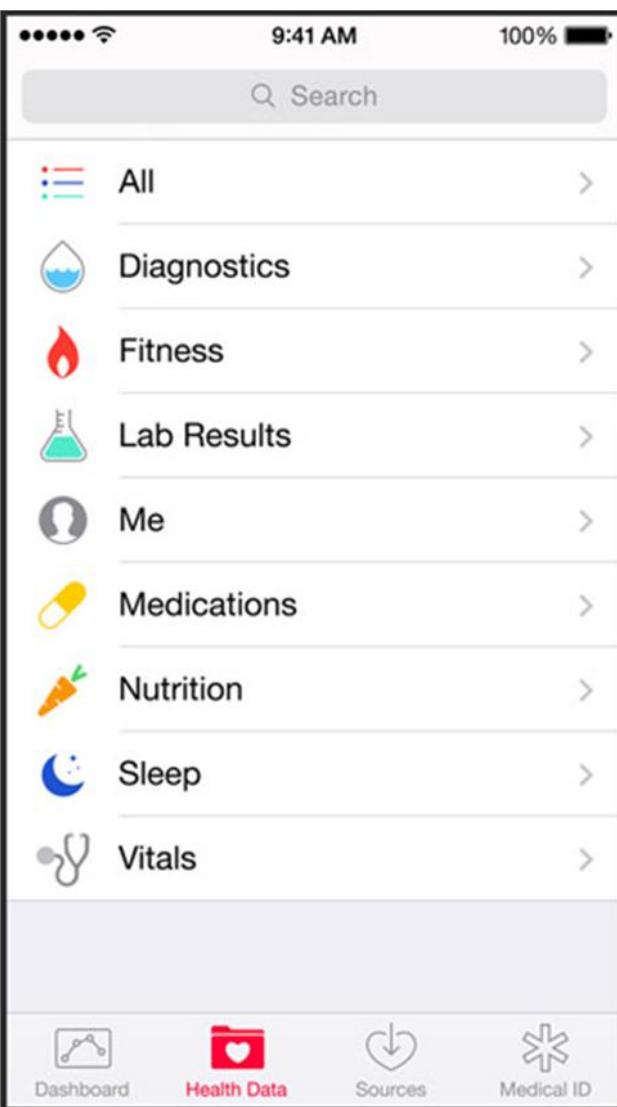
의료정보학자
환자의 의료
기록을 분석
하고 통합 관리,



Apple Health App



Made for
iPhone



HealthKit 관련 에코시스템 확대

• 기타 업체

• Allscripts

• Mount Sinai
• Cleveland Clinic
• Johns Hopkins
• Kaiser Permanente

써드 파티 웨어러블
단말 및 의료 기기업체

EMR 업체
(Epic Systems)

전문 의료 기관
(Mayo Clinic 외 다수)

헬스 및 피트니스
활동 정보

헬스 및 피트니스
활동 관리

전자의무기록

통합 데이터

전문 의료 서비스



헬스 및 피트니스
활동 정보

iPhone

iWatch

소비자

→ 개인 정보
→ 서비스

August 01, 2014

Notice

바이오 전문가의 꿈 키워요

분야별뉴스 | 과학기술 | 보건 · 의학

병원간 의료정보 교류 시대 연다

K-헬스케어 3.0 프로젝트 시범사업 추진

Digital Hospital

- 전자 의료 기록 시스템
- 처방 전달 시스템
- 의료영상 저장정보 시스템
- 통합 환자 중심 환경



u-Health

- 실시간 건강 관리 서비스
- 스마트 케어 처방/관리
- 모바일 의료 솔루션

Health 2.0

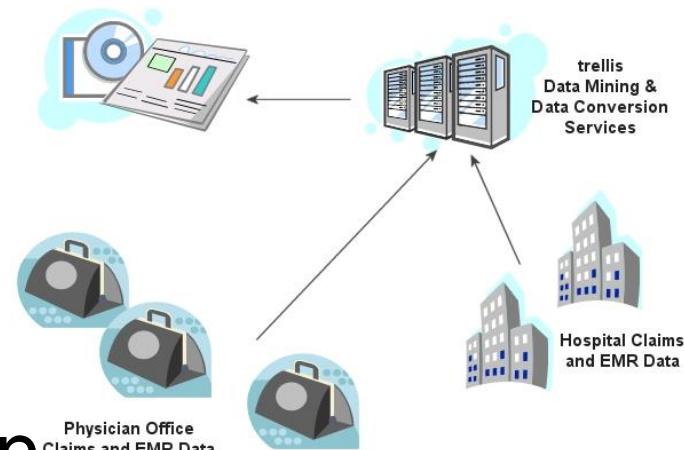
- 개인 맞춤형 의료서비스
- 웹기반 진료 정보 공유

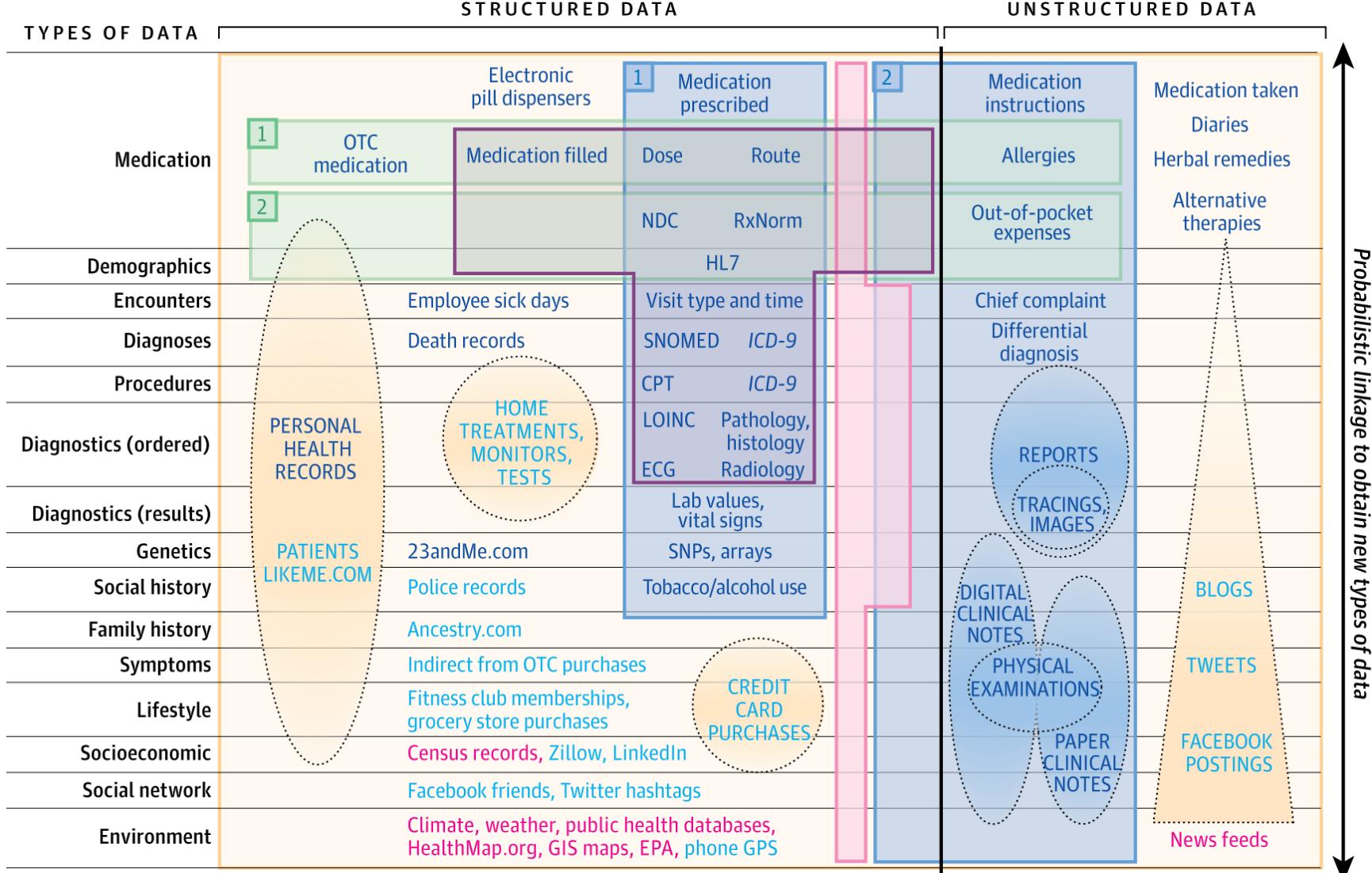
여름휴가를 맞아 가족들과 여행을 떠난 40대 후반의 김 모 씨는 중증의 천식 환자다. 그래서 기침이 나지 않도록 늘 조심을 한다. 그런데 이번 여행길에서는 챙겨간 천식약이 모자라 복용량이 충분치 못했던 까닭에 호흡곤란이 시작되면서 고초를 겪게 됐다.

가족들은 호흡곤란으로 고로워하는 김 씨를 근처의 종합병원으로 긴급히 후송하였다. 불행 중 다행으로 김 씨는 지난달 '의료정보교류 서비스'에 가입해 두었던 터라, 이내 의료진은 의료정보교류 시스템을 통해 김 씨의 과거 진료기록을 신속하게 확인할 수 있었다. 그 결과 혈액검사 등 간단한 검사만으로 빠르고 정확한 진단과 처방을 내릴 수 있었다.

Clinical and Research Applications

1. Clinical Decision Support
2. Public Health Surveillance
3. Personal Health Record
4. Healthcare Data Integration





Examples of biomedical data

1 Pharmacy data
2 Claims data

1 Health care center (electronic health record) data
2 Registry or clinical trial data

Data outside of health care system

Ability to link data to an individual

- Easier to link to individuals
- Harder to link to individuals
- Only aggregate data exists

Data quantity



More Less

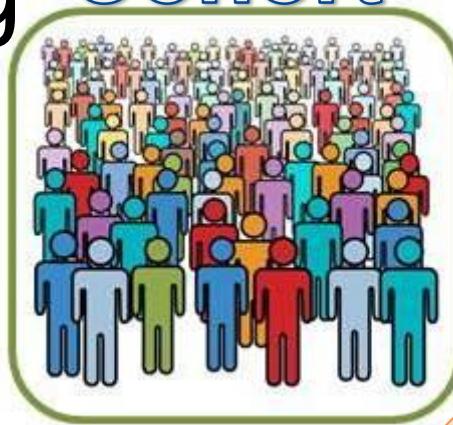
Pros and Cons of Different Data Sources

Pros and Cons of Different Data Sources

	RCT	<u>Observational Cohort/Registry</u>	<u>EHR Data</u>	<u>Claims Data</u>
Typical Questions the Data Is Optimized to Address	Efficacy	Variable	± Effectiveness Practice Patterns	± Effectiveness Safety Practice Patterns
Data Capture	+++ phenotype ++ completeness	+++ phenotype + completeness	++ phenotype ++ completeness	+ phenotype +++ completeness
Internal Validity	+++	++	++	±±
External Validity	-	±±	±±	+++
Data Acquisition Costs	+	++	++	+++

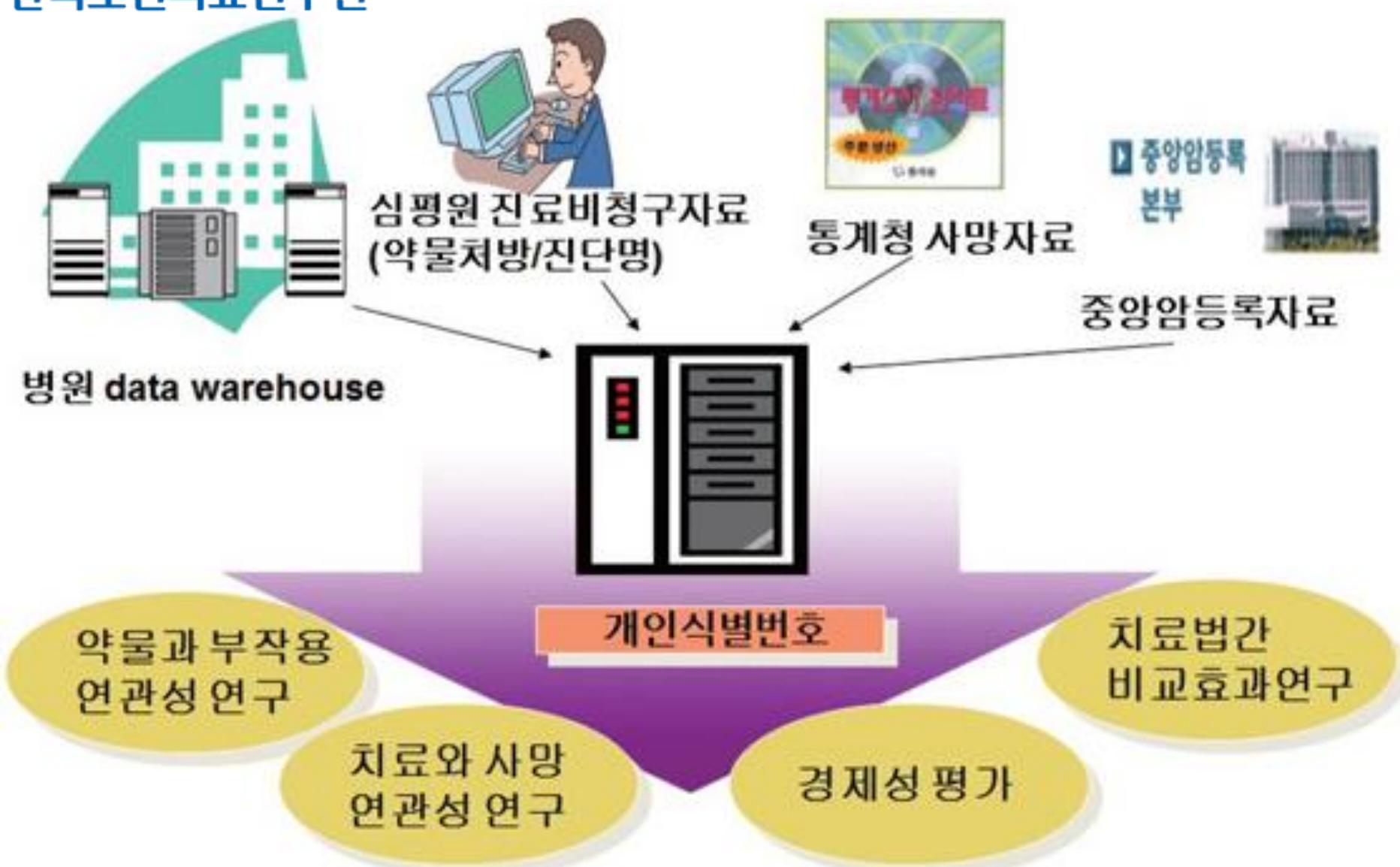
Key: + = fair; ++ = good; +++ = very good; ±± = variable

Data Sharing Cohort



The diagram illustrates the data sharing process. A central orange double-headed arrow connects the 'Cohort' (represented by the crowd icon) and the 'HOSPITAL' (represented by the building icon). From each end of this central arrow, a single-headed orange arrow points to specific data sources: 'National health claims data' (NHIS/HIRA), 'National health examination' (NHIS), 'Beneficiary' (Statistics Korea), 'Cause of death' (Statistics Korea), and 'Cancer registry' (National Cancer Center).

Variable	National health claims data	National health examination	Beneficiary	Cause of death	Cancer registry
	NHIS/HIRA	NHIS	Statistics Korea	Statistics Korea	National Cancer Center
Socio-demographics	-	○	○	-	-
Health behavior	-	○	-	-	-
Disease	○	○	-	-	○
Pharmaceuticals	○	-	-	-	-
Lab data	-	○	-	-	-
Cause of death	-	-	Death only	○	-



NATURE | NEWS



US big-data health network launches aspirin study

PCORI clinical-research initiative will collect information on some 30 million people.

Sara Reardon

06 August 2014



US\$10-million pilot study N=30M
Patient-Centered Outcomes Research Institute

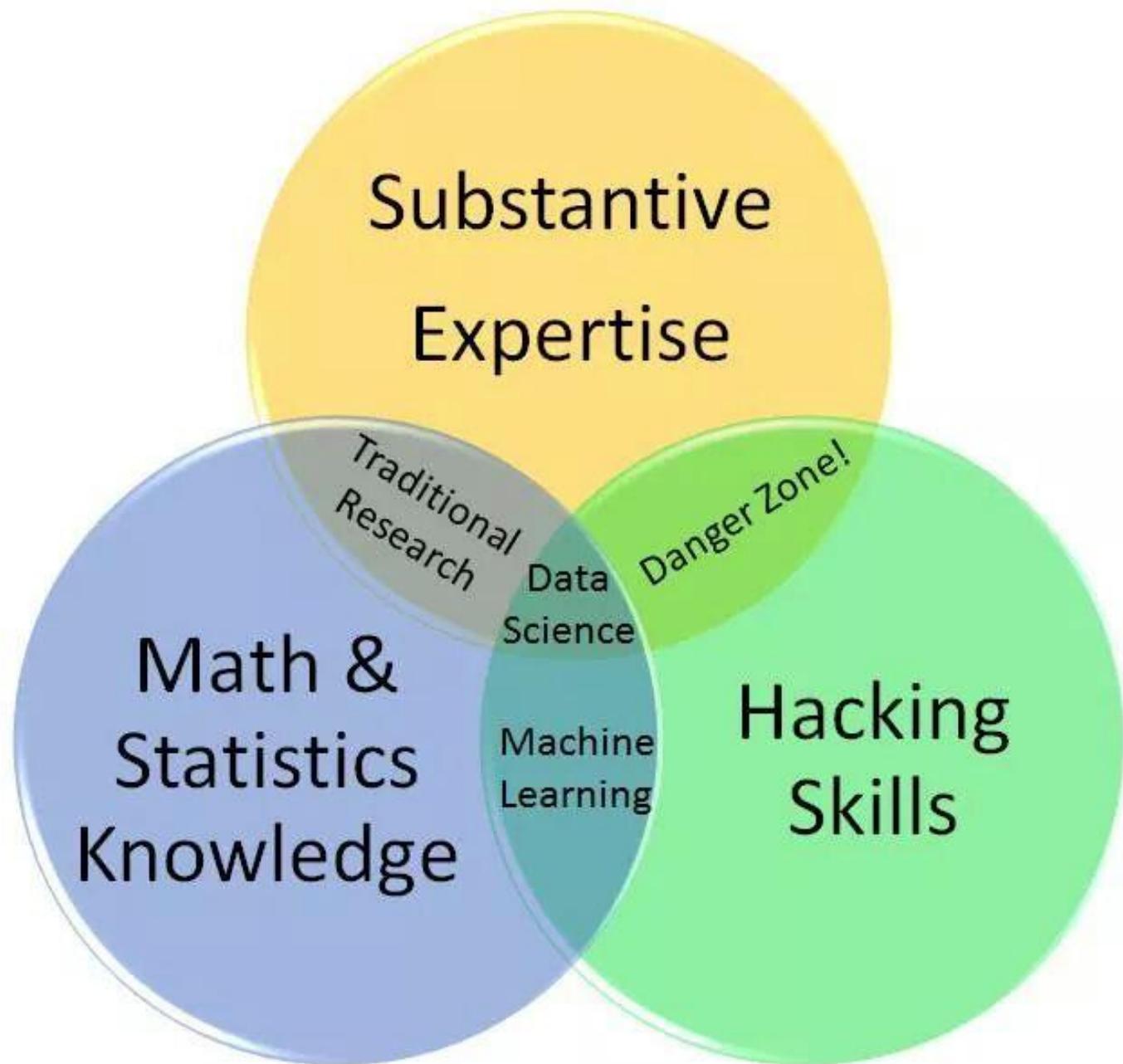
insurance claims, blood tests, medical histories



Clinical and Research Applications

1. Clinical Decision Support
2. Public Health Surveillance
3. Personal Health Record
4. Healthcare Data Integration
5. How to Prepare for Big Data Driven Healthcare

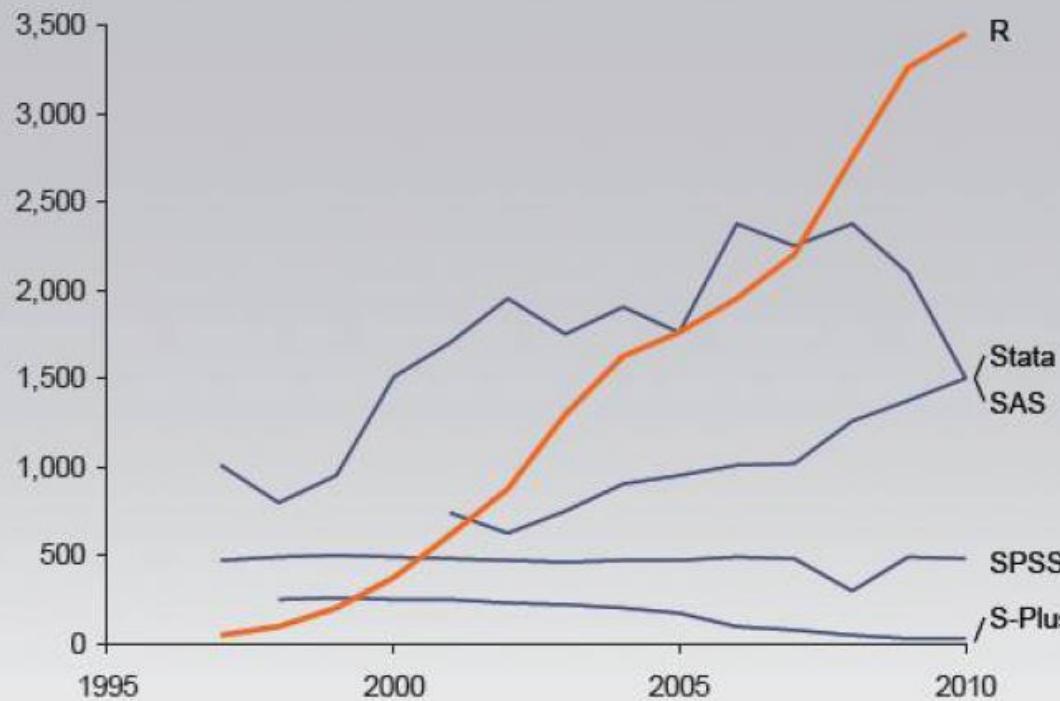




R vs. Stata vs. SAS

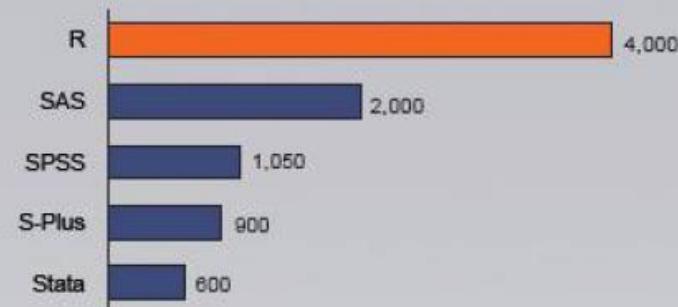
Internet Discussion

Mean monthly traffic on email discussion list



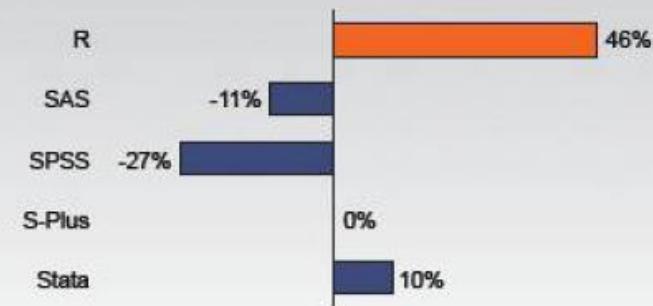
Web Site Popularity

Number of links to main web site



Scholarly Activity

Google Scholar hits ('05-'09 CAGR)



주요 통계 소프트웨어 특징 비교

특징	Stata	SPSS	SAS	R
학습 곡선	가파름/완만함	완만함/평坦함	매우 가파름	매우 가파름
사용자 환경	프로그래밍/마우스 클릭	대부분 마우스 클릭	프로그래밍	프로그래밍
데이터 관리	매우 강력	보통	매우 강력	매우 강력
데이터 분석	강력	강력	강력/만능	강력/만능
그래픽	매우 좋음	매우 좋음	좋음	탁월
비용	저렴(평생 라이선스, 업그레이드만 갱신)	고가(업그레이드까지 갱신 필요 없음, 장기간 라이센스)	고가(매년 갱신)	오픈 소스

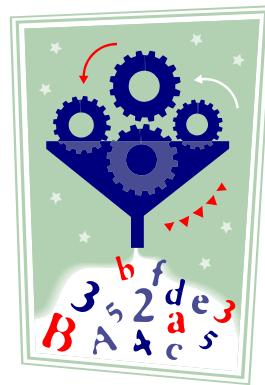
출처: Oscar Torres-Reyna, Getting Started in R~Stata Notes on Exploring Data
<http://dss.princeton.edu/training/RStata.pdf>

빅데이터 연구 적용

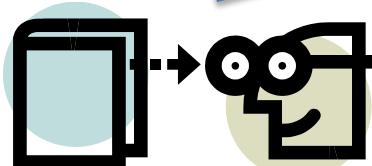
전통적인 관점 연구



Large scale
(unstructured)
data

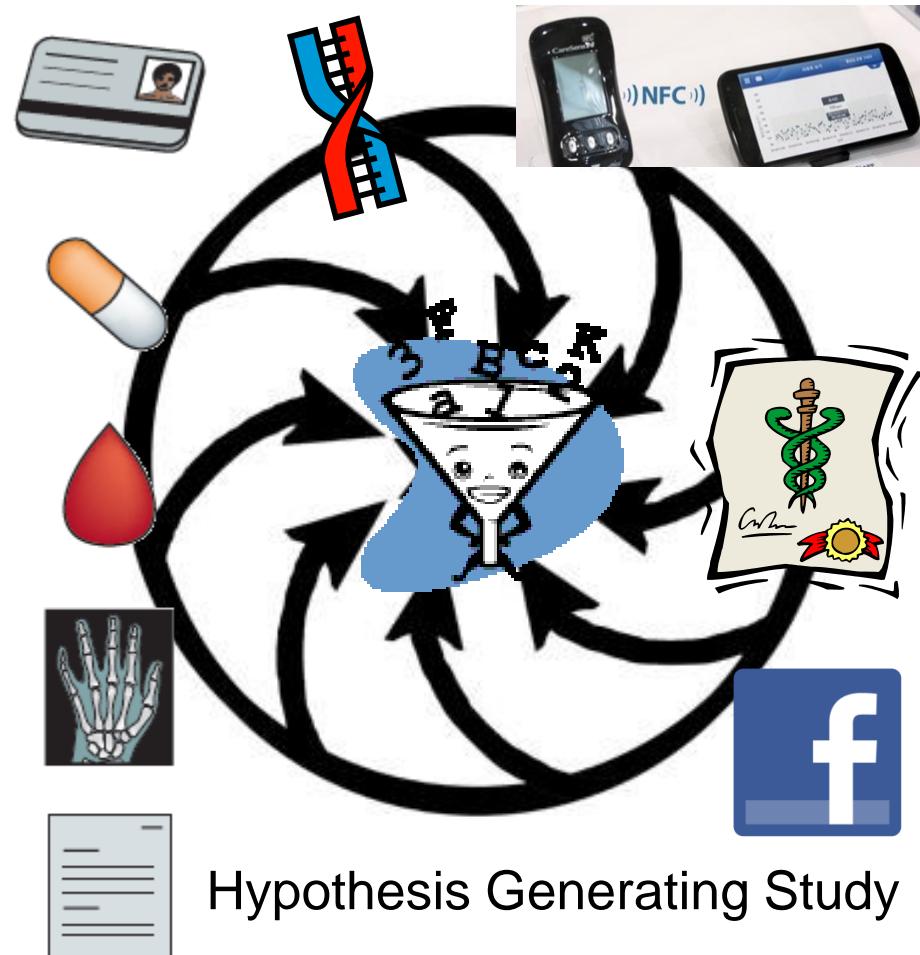


Summary
(Modify)



Classical hypothesis driven study

새로운 관점 연구



Hypothesis Generating Study