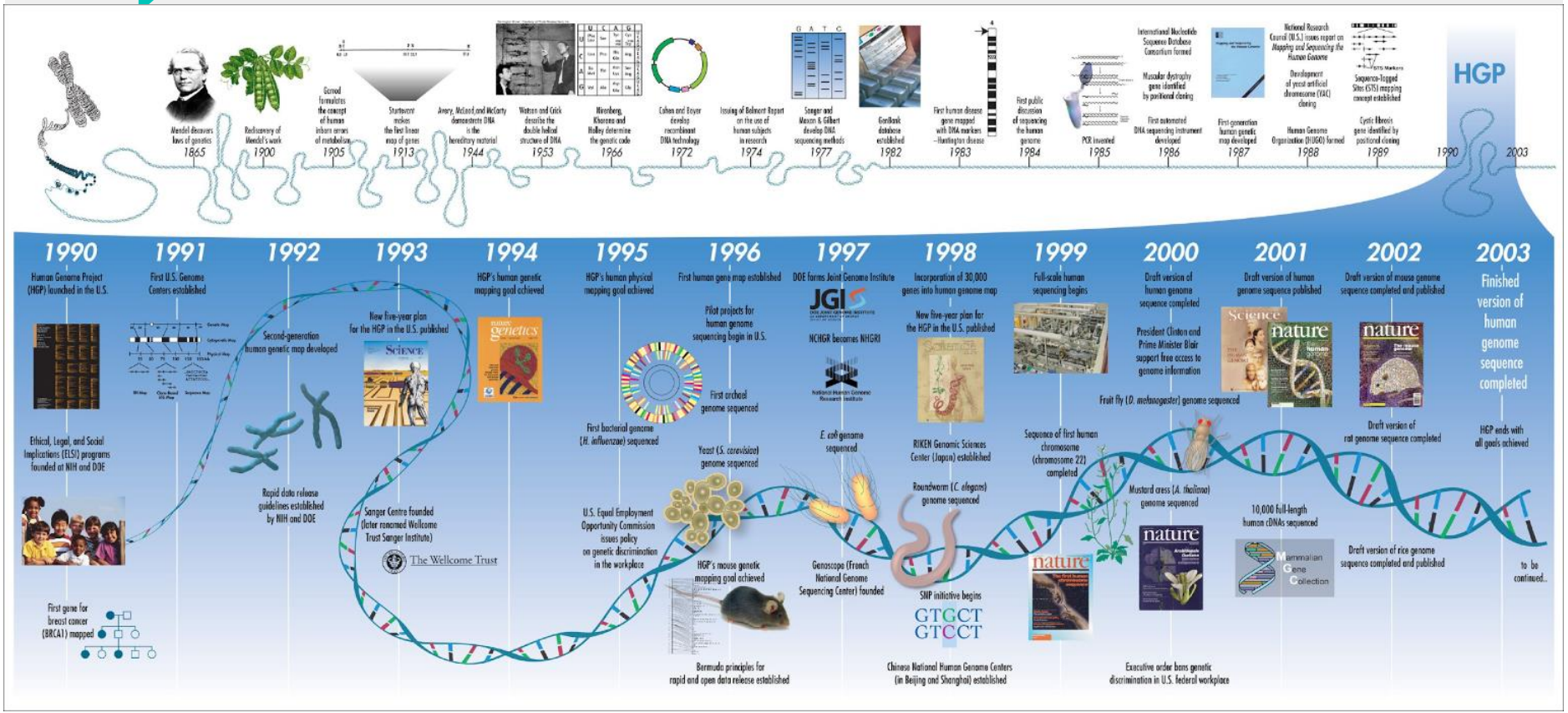
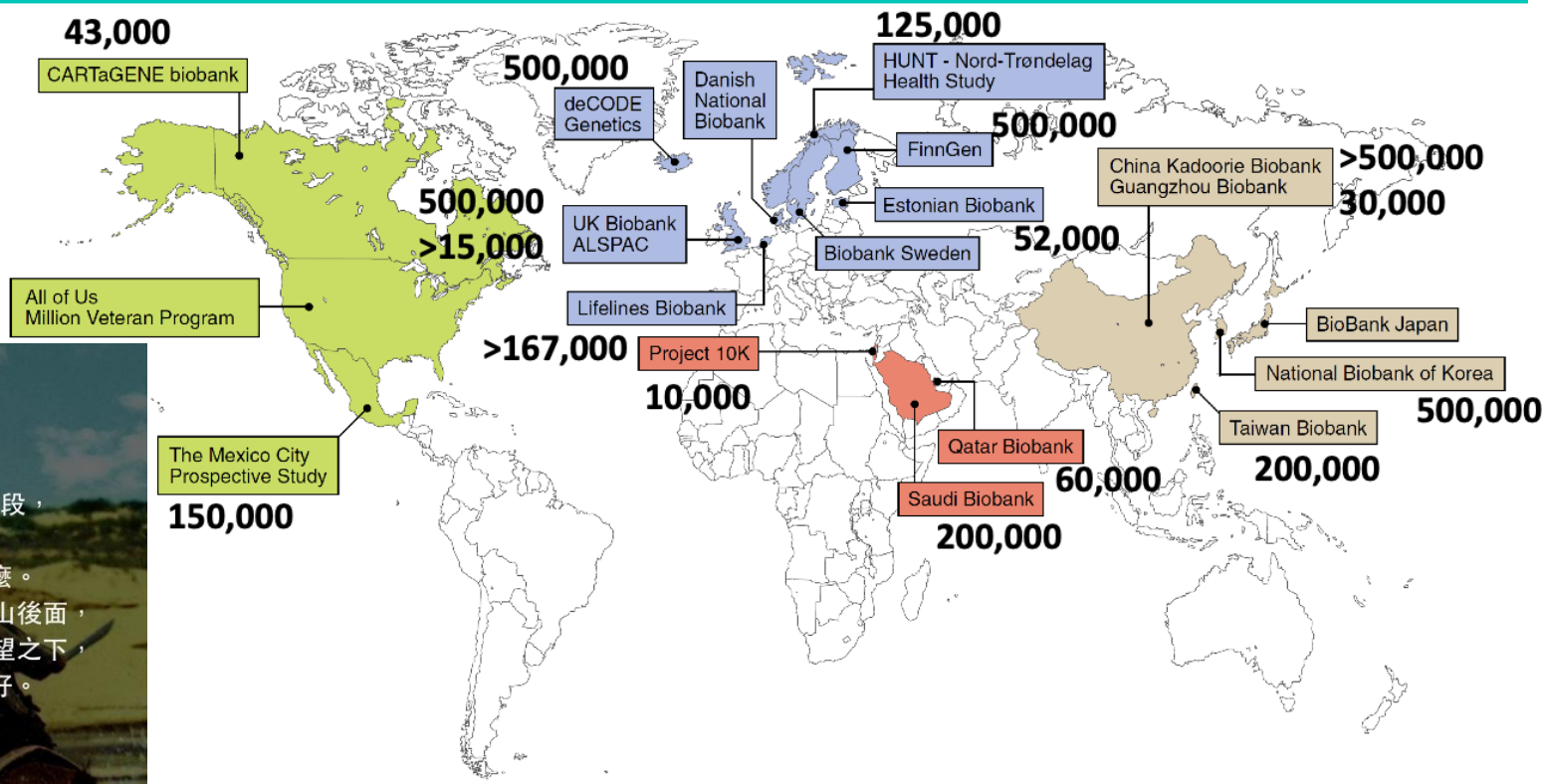
Taiwan Biobank

Su Ming-Wei wei@ibms.sinica.edu.tw

Human Genome Project



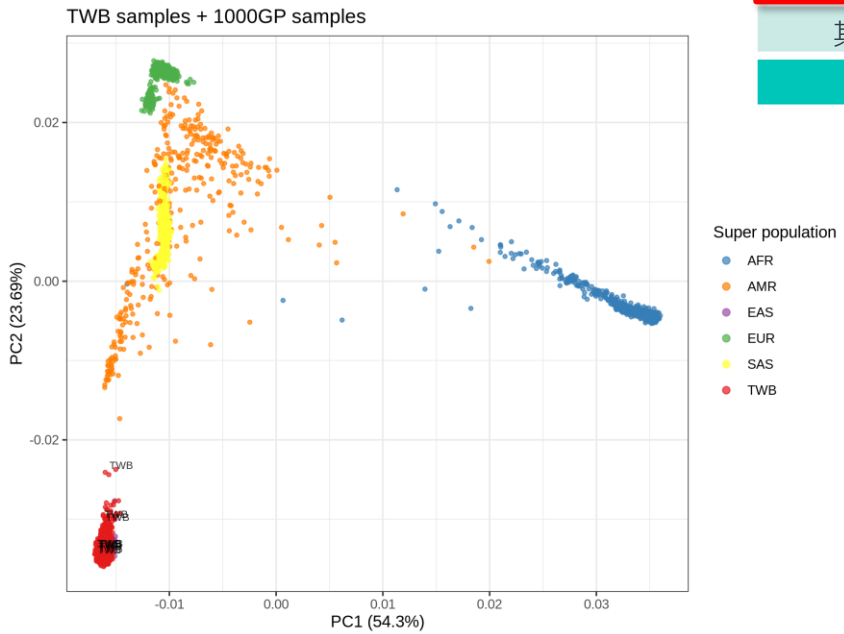
Global biobanks

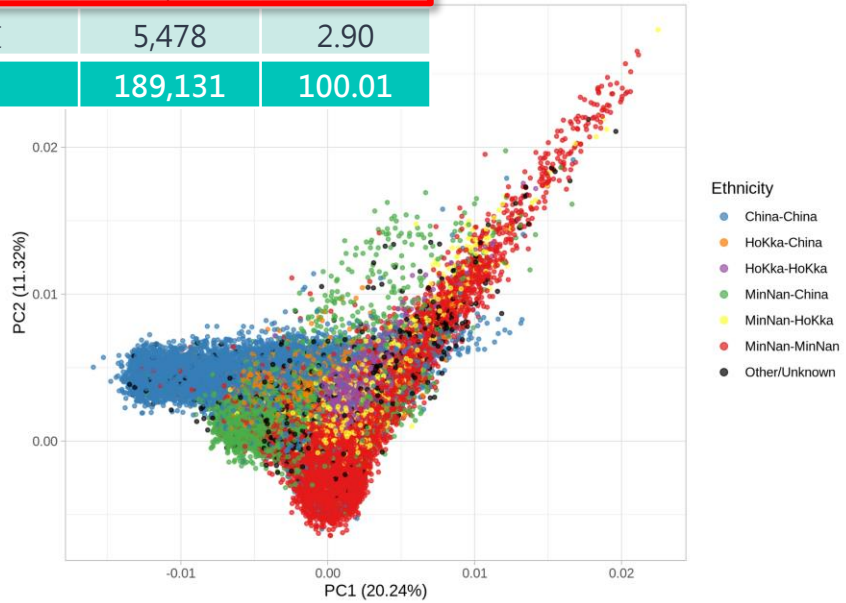


Shilo, et al. (2020) Nature Medicine, 26: 29-38

族群遺傳結構

自述父母親籍貫	人數	百分比
中國各省-中國各省	12,032	6.36
客家-中國各省	2,895	1.53
客家-客家	18,152	9.60
閩南-中國各省	12,795	6.77
閩南-客家	11,410	6.03
閩南-閩南	126,369	66.82
其他/不知道	5,478	2.90
加總	189,131	100.01





建構以大數據為基礎之國家精準健康架構

Biobanking 1.0
- 基礎設施建置
- 檢體收集及儲存

Biobanking 2.0
- 整合資源
- 標準化、一致化

Biobanking 3.0
- 數位化、資安
- 資料串檔、連結

Biobanking 4.0
- 創新模式
- 永續經營

有效運作 | 社會支持 | 國際連結 | 經營永續

檢體收集（實體庫） ▶ 檢體運用（整合平台） ▶ 檢體數位化（大數據）

Biobank 基盤建置 2012-2018
- 31家Biobank
- ~450萬檢體組數 (104-108年5月)

Biobank 整合平台 2019-
- 雲端整合平台
- 標準作業流程
- 聯合審查機制
- 加值服務

中央辦公室

跨庫申請案

引入國際大型產官學合作

沙盒示範（Sandbox）

建構精準健康大數據系統

★ 資料標準化模式

★ 健康大數據串聯

應用前瞻資訊技術開發工具，促進有效利用

水平整合

垂直整合

資料格式、關聯性資料庫

- 個人生理監測數據
- 環境資料
- 電子健康紀錄 (EHR)
- 臨床試驗資訊
- 長照資料
- 行為及調查資料
- 健保資料
- 健保雲端上傳資料(檢驗/影像/病摘)
- 政府登記資料(出生/死因/癌登)
- 警政社福內政資料
- 基因體資料
- Biobank資料

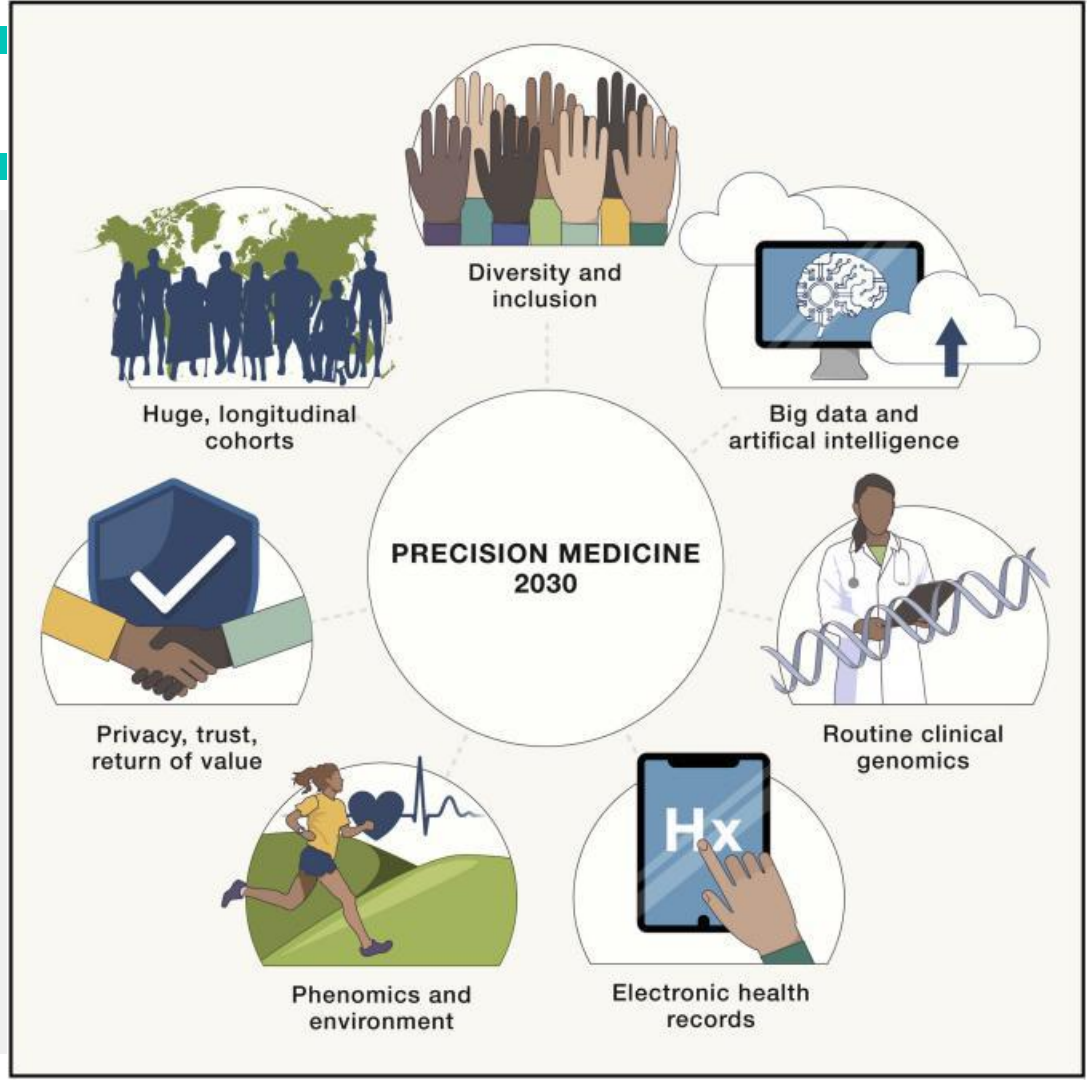
- 基因檢測及分析
- AI臨床應用
- 資安管理技術
- RWD / RWE
- 臨床決策系統
- 精準醫療健康
- 發展新興醫療科技
- 支援產業創新發展
- 增進全民健康福祉
- 接軌國際

Cell
Volume 184, Issue 6, 18 March 2021, Pages 1415-1419

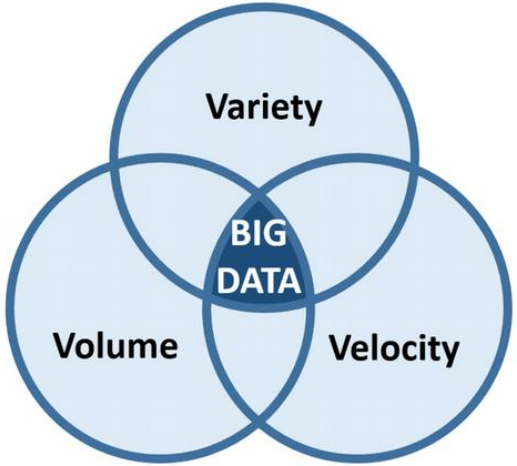
Commentary

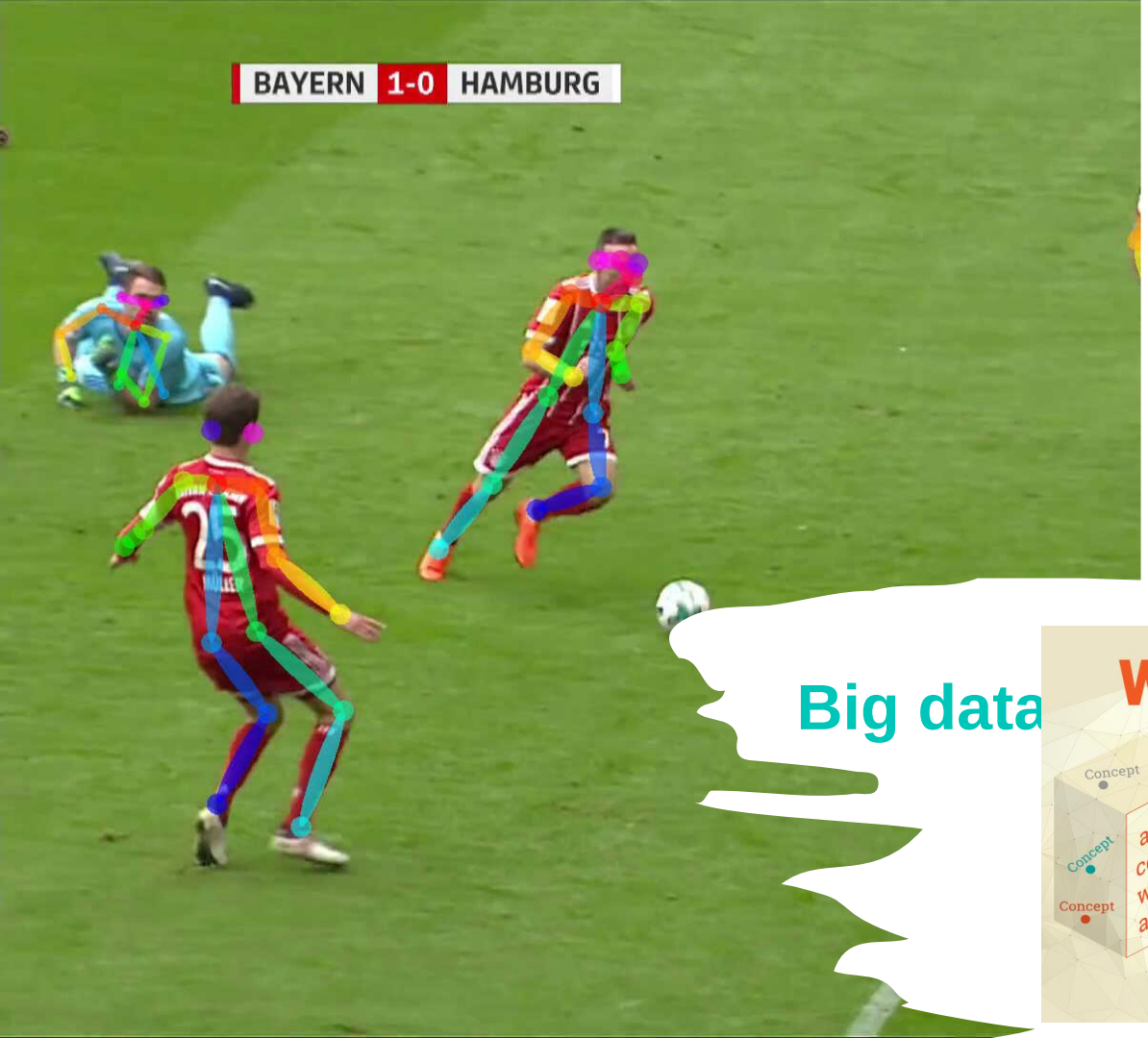
Precision medicine in 2030—seven ways to transform healthcare

Joshua C. Denny [1, 3], Francis S. Collins [2]



Data v.s. Big data in a geek way



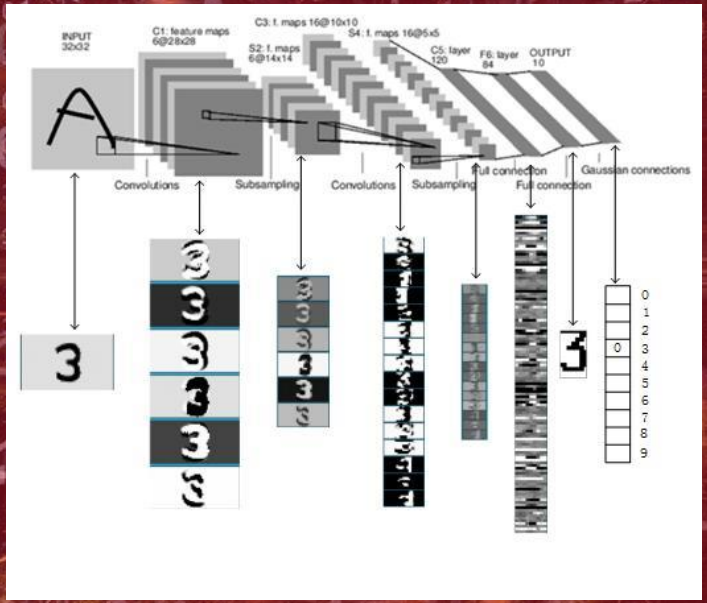

BAYERN 1-0 HAMBURG


PATIENT
CURARISE

Big data


What are Ontologies?
Concept
Concept
Concept
Concept
a set of concepts within a domain
Ontology
instances of objects
rules
attributes
axioms
classes

Deep Learning Meets Big data



Data = World

Why Taiwan ?

- Holo Taiwanese
- Hakka Taiwanese
- Mainianders
- Taiwanese indigenous proples

全民健康保險
NATIONAL HEALTH INSURANCE

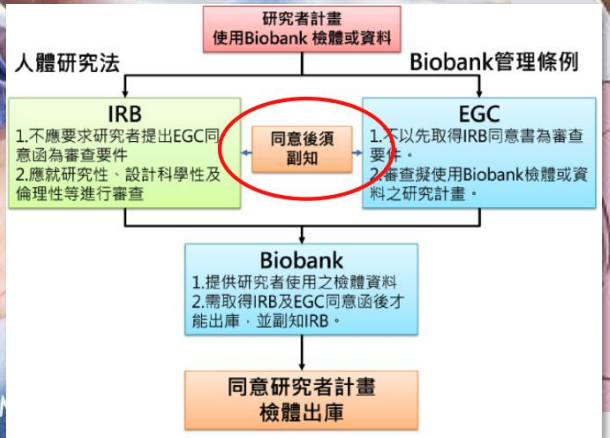
中華民國 總統府
Office of the President
Republic of China (Taiwan)
www.president.gov.tw

https://ogme.edu.tw/lc/culturalGroups
https://tendashsix.com/taiwan-medical-service-ranked-first/

TWB

- 主管機關：衛生福利部
- 設置者：中央研究院
- 依2010.2.3 總統府頒布執行之《人體生物資料庫管理條例》設置
- 設有倫理委員會，針對資料庫之管理等有關事項進行審查及監督
- 2012.10.24 取得設置許可
- 第一個主管機關核准設置之人體生物資料庫
- 2012.11.8 開始正式收案 預計目標20萬社區民眾

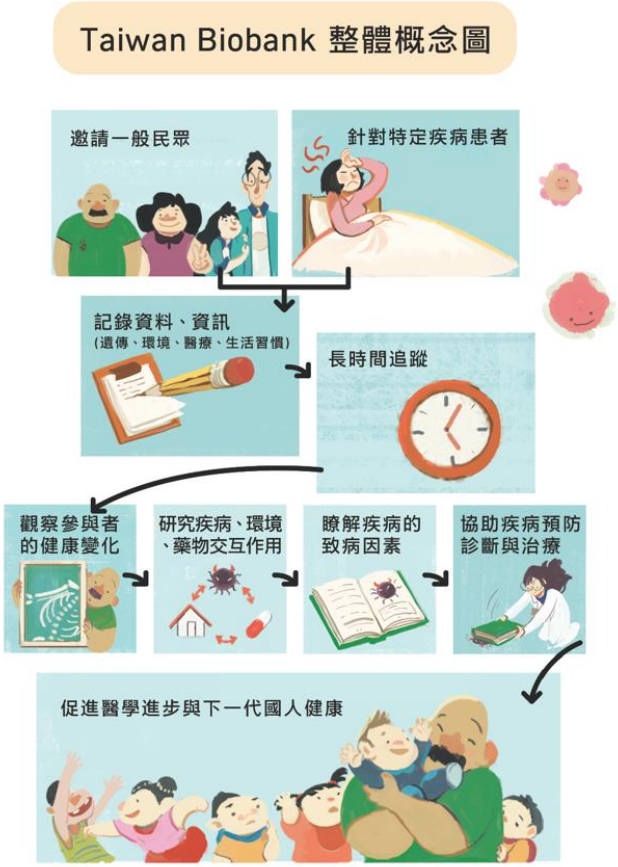
NOT ONLY YOU CARE!!



我們是誰？在做什

我們是中央研究院臺灣人體生物資料庫（Taiwan Biobank），
欲透過結合生活習慣、環境因子、臨床醫學與基因生物標幟等資訊，為生物醫學研究蒐集龐大的生物檢體與健康資訊，提供國內學者申請使用，目前已有諸多成果。

為籌建多元化的國家型健康大數據，因應多樣性研究，我們自2012年取得設立通過後即開始邀請20萬名一般民眾與部分疾病別之患者加入，收集其身體檢測、檢體、醫藥史、家族史、生活環境與習慣等健康相關資料、資訊，除了提供基礎研究資源外，更希望最終回饋於民，讓民眾共享健康碩果。



❶ 民眾自願參與

❷ 基礎資料變鑽石的過程

❸ 國際期刊發表

❹ 回饋於民、自主健康

檢體檢驗值

包含血液、尿液與糞便等之檢驗值。

生物檢體

包含DNA、血漿及尿液等。

身體檢測

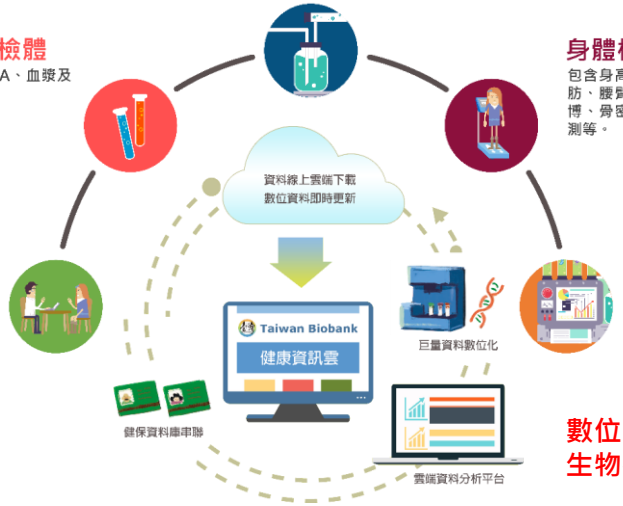
包含身高、體重、體脂肪、腰臀圍、血壓、脈博、骨密度與肺功能檢測等。

健康問卷

包含基本人口學、個人健康行為、生活環境、飲食狀況、家族疾病史、女性相關問題、中醫體質問項等。

檢體加值

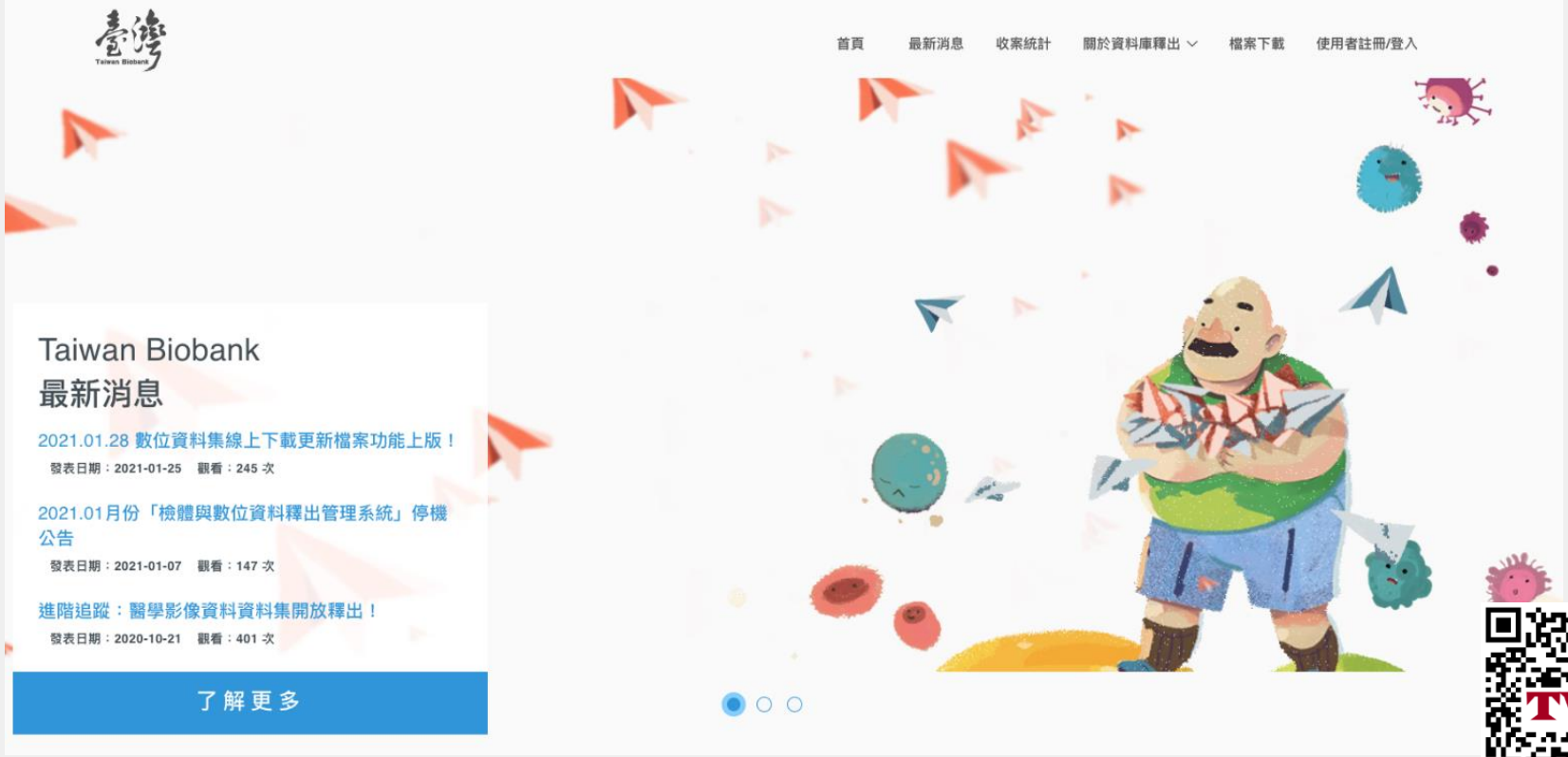
包含基因體學、表觀基因體學、代謝體學、環境毒物及影像資料等。



數位加值資訊
生物檢體釋出

基因體學 Genomics	表觀基因體學 Epigenomics	代謝體學 Metabolomics	環境毒物 Environmental Toxic	影像資料 Medical Image
✓ 全基因體定型資料 Whole-genome genotyping ✓ 全基因體定序資料 Whole-genome sequencing ✓ 人類白血球組織抗原分型資料HLA typing ✓ 腸道微生物菌相分析資料 ✓ Taiwan View	✓ 全基因體甲基化晶片資料 DNA methylation	✓ 血液代謝資料 Blood metabolome	✓ 尿液鄰苯二甲酸酯含量資料 Urinary Phthalate and Melamine Levels	✓ 維陪追蹤醫學影像資料 Medical Images in VIP Follow-up Stage

https://www.biobank.org.tw/index.php



https://reurl.cc/jqjqNZ

目前成果

✓ 已累計超過 25萬多人次參與

收案數

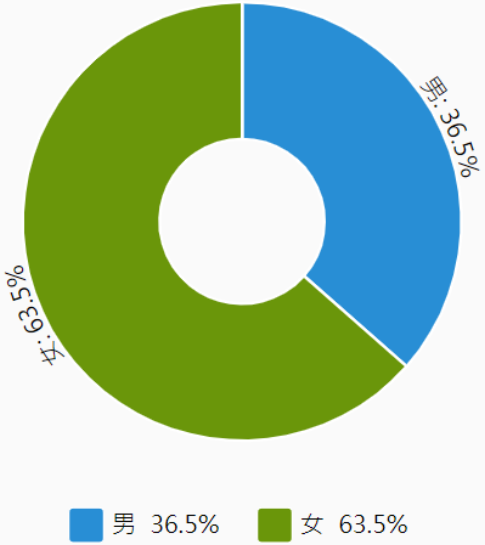
統計至2023年07月31日止

202,219 人參與

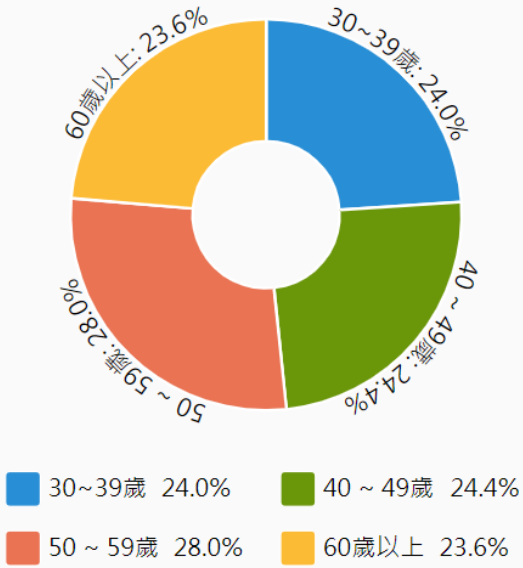
49,226 人完成第一輪追蹤

1,097 人完成第二輪追蹤

01. 性別分布

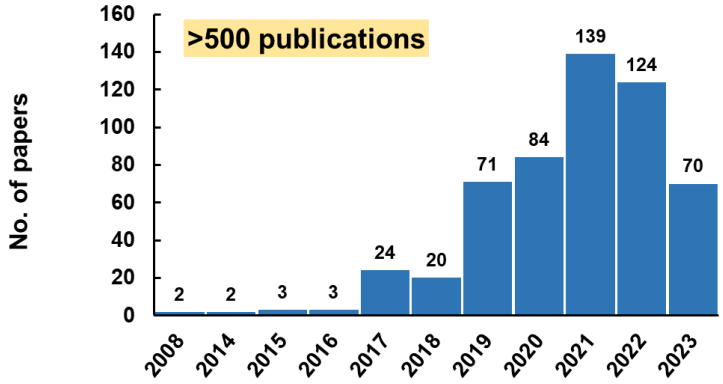


02. 年齡分布



✓ 透過Taiwan Biobank資料研究發表至國際期刊之文章已高達522篇以上

累計至 2023年7月

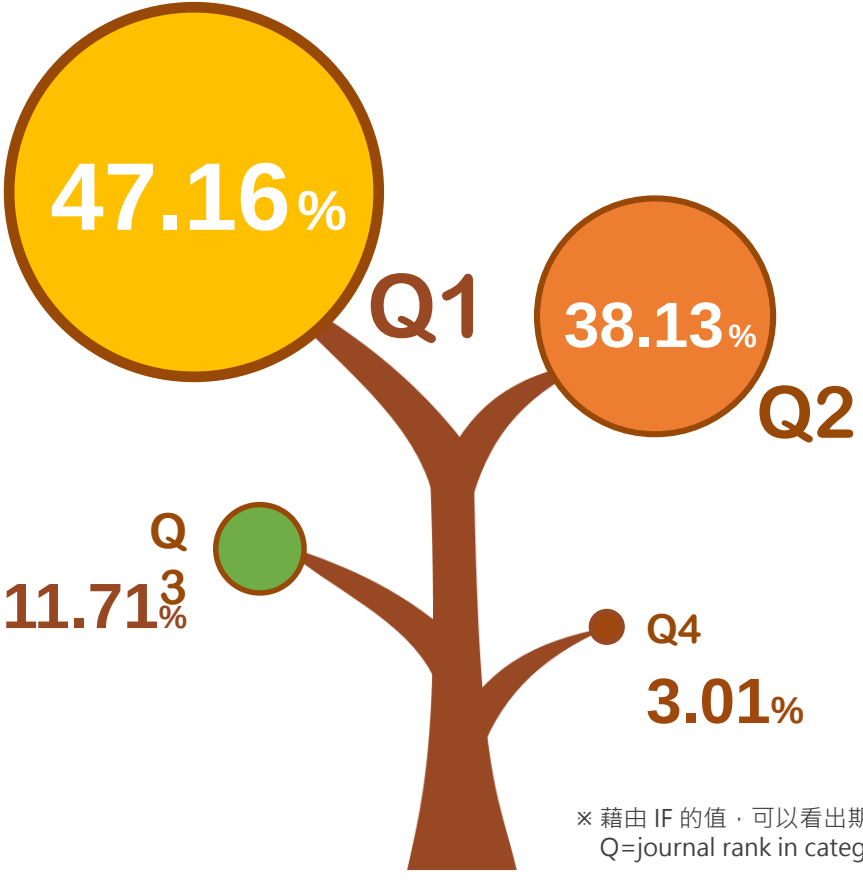


資料釋出成果 - 2023年

- Int. J. Mol. Sci. (2023) Splice-Site Variants in the Gene Encoding GABA-A Receptor Delta Subunit Are Associated with Amphetamine Use in Patients under Methadone Maintenance Treatment. **Lin, Y.-F.,et al.**
- J. Pers. Med. (2023) Whole-Exome Sequencing Identifies Genetic Variants for Severe Adolescent Idiopathic Scoliosis in a Taiwanese Population. **Lin, M.-R.,et al.**
- Mol Ther Oncolytics. (2023) Zika virus cleaves GSDMD to disseminate prognosticable and controllable oncolysis in a human glioblastoma cell model. **Kao, Y.-T.,et al.**
- J. Oncol. (2023) Clinicopathological Features and Oncological Outcomes of Early and Late Recurrence in Stage III Colorectal Cancer Patients after Adjuvant Oxaliplatin-Based Therapy. **Chang,Y.-T.,et al.**
- Am J Hum Genet. (2023) 15 years of GWAS discovery: Realizing the promise. **Abdellaoui A.,et al.**

https://www.biobank.org.tw/dt_release.php

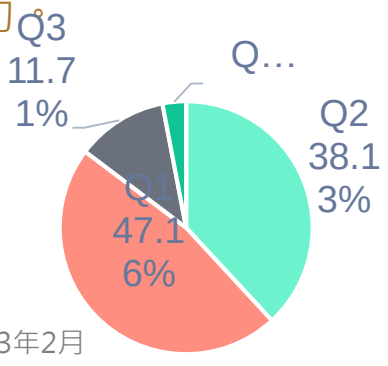
Taiwan Biobank 研究量能之基石



目前透過Taiwan Biobank資料研究發表至國際之期刊排名：

Q1高達47.16%、Q2達38.13%；
更有50.33%⬆篇數Impact Factor(IF)>5。

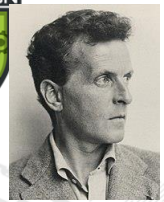
我們擁有完整時序性且精確可靠的健康大數據，可以增強學研能量之產值、減少研究者之人力時間成本，極具國際影響力。



資料統計截至2023年2月

※ 藉由 IF 的值，可以看出期刊在特定領域的排名、四分位區間、與百分位。
Q=journal rank in categories，Q1表示屬於排名前25%的期刊，以此類推。

Data Quality



世界是一切發生的事情
世界是事實的總和，而非事物的總和
發生的事情，即事實，就是諸事態的存在

Completeness

All data sets and data items are recorded?

缺失值

Uniqueness

Is there a single view of the data set?

某些資料須具獨特性，如身份證

Timeliness

Does data represent the correct time frame of events

Validity

Does data match th business rules?

Accuracy

Is data a proper reflection of the data set?

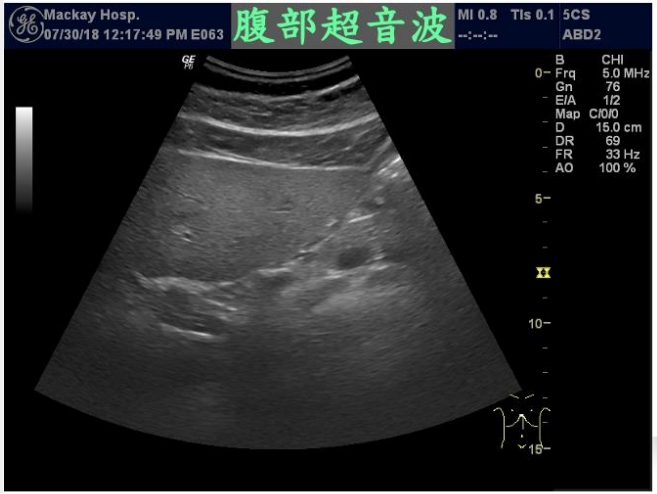
BMI > 100 ?

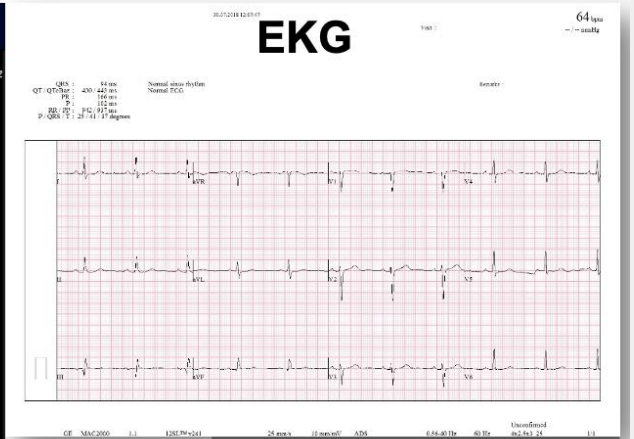
Consistency

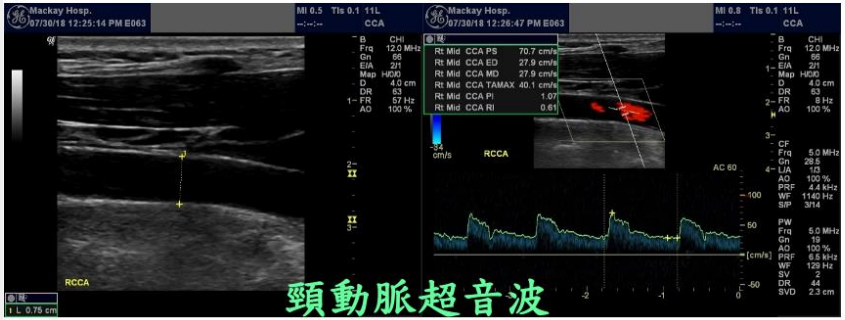
Is the data homogenous, can be compared across?

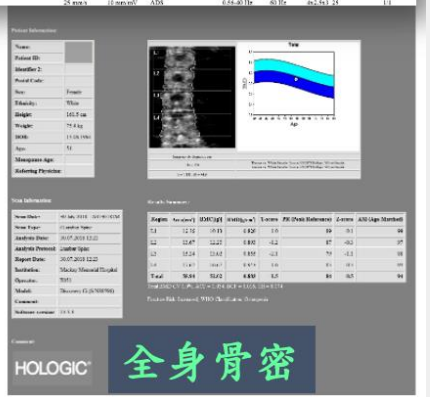
資料需有一致性描述，如胖 or 肥胖

advance follow-up program

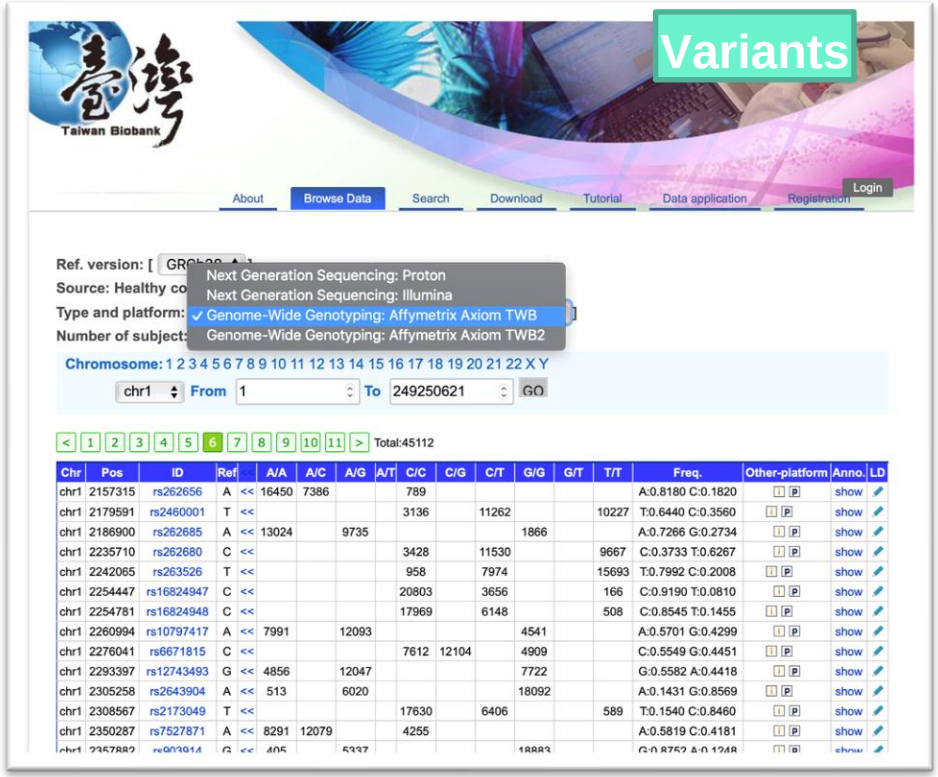






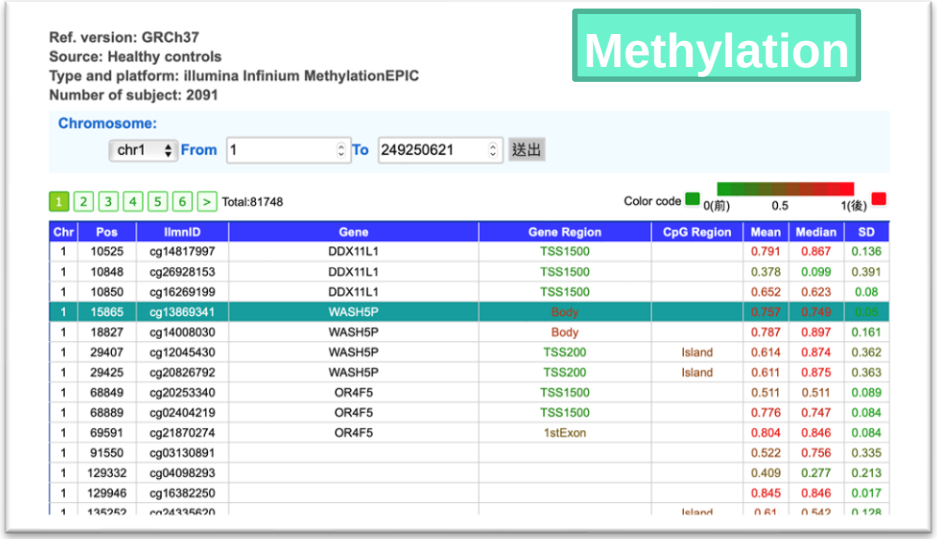


Taiwan View

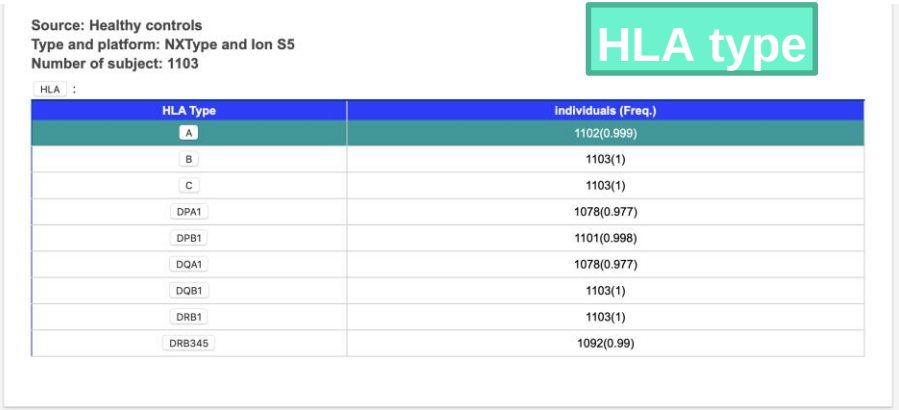


Chr	Pos	ID	Ref		A/A	A/C	A/G	A/T	C/C	C/G	C/T	G/G	G/T	T/T	Freq.	Other-platform	Anno.	LD
chr1	2157315	rs262656	A	<<	16450	7386			789						A:0.8180 C:0.1820	I P	show	
chr1	2179591	rs2460001	T	<<					3136		11262			10227	T:0.6440 C:0.3560	I P	show	
chr1	2186900	rs262685	A	<<	13024		9735					1866			A:0.7266 G:0.2734	I P	show	
chr1	2235710	rs262680	C	<<					3428		11530			9667	C:0.3733 T:0.6267	I P	show	
chr1	2242065	rs263526	T	<<					958		7974			15693	T:0.7992 C:0.2008	I P	show	
chr1	2254447	rs16824947	C	<<					20803		3656			166	C:0.9190 T:0.0810	I P	show	
chr1	2254781	rs16824948	C	<<					17969		6148			508	C:0.8545 T:0.1455	I P	show	
chr1	2260994	rs10797417	A	<<	7991		12093					4541			A:0.5701 G:0.4299	I P	show	
chr1	2276041	rs6671815	C	<<					7612	12104		4909			C:0.5549 G:0.4451	I P	show	
chr1	2293397	rs12743493	G	<<	4856		12047					7722			G:0.5582 A:0.4418	I P	show	
chr1	2305258	rs2643904	A	<<	513		6020					18092			A:0.1431 G:0.8569	I P	show	
chr1	2308567	rs2173049	T	<<					17630		6406			589	T:0.1540 C:0.8460	I P	show	
chr1	2350287	rs7527871	A	<<	8291	12079			4255						A:0.5819 C:0.4181	I P	show	
chr1	2357882	rs903914	G	<<	405		5337					18883			G:0.8752 A:0.1248	I P	show	

https://taiwanview.twbiobank.org.tw/index



Chr	Pos	IlmnID	Gene	Gene Region	CpG Region	Mean	Median	SD
1	10525	cg14817997	DDX11L1	TSS1500		0.791	0.867	0.136
1	10848	cg26928153	DDX11L1	TSS1500		0.378	0.099	0.391
1	10850	cg16269199	DDX11L1	TSS1500		0.652	0.623	0.08
1	15865	cg13869341	WASH5P	Body		0.757	0.749	0.05
1	18827	cg14008030	WASH5P	Body		0.787	0.897	0.161
1	29407	cg12045430	WASH5P	TSS200	Island	0.614	0.874	0.362
1	29425	cg20826792	WASH5P	TSS200	Island	0.611	0.875	0.363
1	68849	cg20253340	OR4F5	TSS1500		0.511	0.511	0.089
1	68889	cg02404219	OR4F5	TSS1500		0.776	0.747	0.084
1	69591	cg21870274	OR4F5	1stExon		0.804	0.846	0.084
1	91550	cg03130891				0.522	0.756	0.335
1	129332	cg04098293				0.409	0.277	0.213
1	129946	cg16382250				0.845	0.846	0.017
1	135252	cg24335620			Island	0.61	0.542	0.128



HLA Type	individuals (Freq.)
A	1102(0.999)
B	1103(1)
C	1103(1)
DPA1	1078(0.977)
DPB1	1101(0.998)
DQA1	1078(0.977)
DQB1	1103(1)
DRB1	1103(1)
DRB345	1092(0.99)

Taiwan View

PheWeb

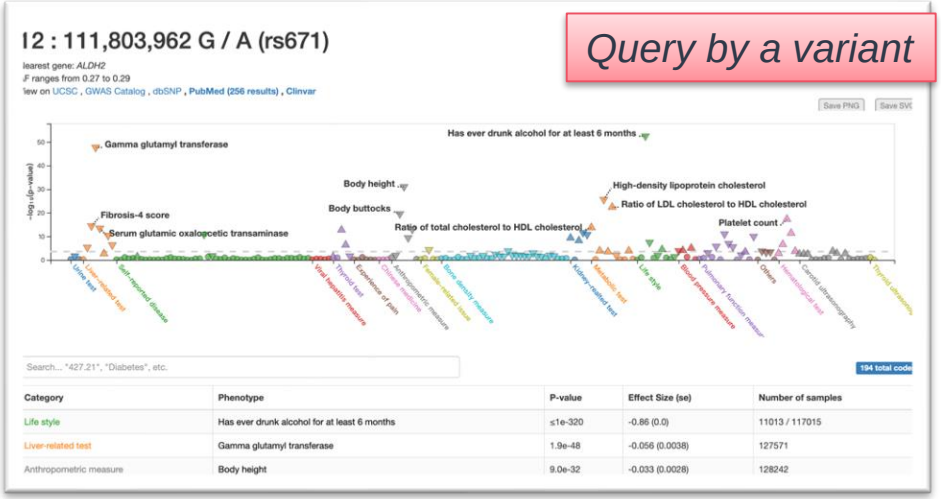
Source: Healthy controls
Type and platform: Genome-Wide Genotyping
Number of subject: 128775

Phenotypes Top Hits Random About

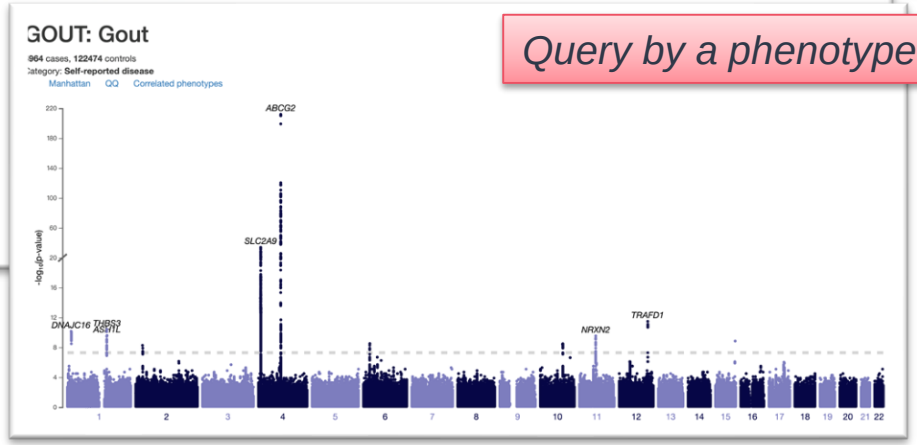
Browse GWAS results using SAIGE

Search for a variant, gene, or phenotype

Query by a variant



Query by a phenotype



Query by a gene

ALDH2

Phenotypes with the most-significant associations for this locus:

Top p-value in gene	Phenotype
≤1e-320	Has ever drunk alcohol for at least 6 months
8.4e-64	Platelet count
1.9e-48	Gamma glutamyl transferase
2.2e-45	Fibrosis-4 score
5.6e-32	Body height
6.5e-27	High-density lipoprotein cholesterol
1.3e-23	Ratio of LDL cholesterol to HDL cholesterol
1.7e-22	Red blood cell count
2.9e-20	Body buttocks
7.4e-15	Ratio of total cholesterol to HDL cholesterol

健康數據酷

https://healthy.twbiobank.org.tw/

data and tubes
https://www.youtube.com/watch?v=LmHJ1KFlW1s&t=2s

~1.5 ~3 million

ISO certificate

ISO/IEC 27001 for management system

TÜV NORD

CERTIFICATE

Management system as per
ISO/IEC 27001:2013

The Certification Body TÜV NORD CERT GmbH hereby confirms as a result of the audit, assessment and certification decision according to ISO/IEC 17021-1:2015, that the organization

TAIWAN BIOBANK
No. 99, Ln. 130, Sec. 1, Academia Rd., Nangang Dist., Taipei City 115202, Taiwan (R.O.C.)
2F,Building B, National Biotechnology Research Park

operates a management system in accordance with the requirements of ISO/IEC 27001 : 2013 and will be assessed for conformity within the 3 year term of validity of the certificate.

The information Security Management System of Information Management Department to provide the operation and maintenance, for Cohort and Dataland Management System Development and Maintenance, Intranet Network Support Service and Server Room

As per the Statement of Applicability (SOA), version 2.0, Effective date 2020-09-28

Certificate Registration No. 44 121 18820001
Audit Report No. IT241-R1

Valid from 2020-11-08
Valid until 2023-11-07
Initial Certification 2017-11-08

Certification Body
at TÜV NORD CERT GmbH

TUV ASIA PACIFIC LIMITED
Taiwan Branch
Room A1, 9F, No.333, Sec.2,
Tun Hua S. Rd.,
Taipei 10689 Taiwan, R.O.C.
2020-11-08

TÜV NORD CERT GmbH Langemarckstrasse 20 45141 Essen www.tuev-nord-cert.com

ISO/IEC 29100 for personal information protection

TÜV NORD

CERTIFICATE

Conformity of Personal Information Protection as Defined per
ISO 29100:2011

In accordance with TÜV NORD CERT procedures, it is hereby certified that

TAIWAN BIOBANK
No. 99, Ln. 130, Sec. 1, Academia Rd., Nangang Dist., Taipei City 115202, Taiwan (R.O.C.)
2F,Building B, National Biotechnology Research Park

with the locations according to the annex

applies a management system in line with the above standard for the following scope

Personal information protection of recruitment participant of Taiwan Biobank(working flow of personal information collection, processing and use) including:
(1) The Center of Taiwan Biobank
(2) Tri-Service General Hospital- Neihu Branch Station
(3) Tri-Service General Hospital- Tingjhou Branch Station
(4) Cheng Ching Hospital Chung Kang Branch Station
(5) Kaohsiung Municipal Ta-Tung Hospital Station
(6) Taichung Veterans General Hospital Station
(7) Kaohsiung Veterans General Hospital Station

Certificate Registration No. TW291800I
Audit Report No. IT-241

Valid from 2018-12-13
Valid until 2021-12-12
Issue Date 2020-11-18
Initial Certification 2018-12-13

Certification Body
at TÜV ASIA PACIFIC LIMITED Taiwan Branch

TUV ASIA PACIFIC LIMITED
Taiwan Branch
Room A1, 9F, No.333, Sec.2,
Tun Hua S. Rd.,
Taipei 10689 Taiwan, R.O.C.

This certification was conducted in accordance with the TUV ASIA PACIFIC LIMITED Taiwan Branch auditing and certification procedures based on TÜV NORD CERT auditing and certification procedures and is subject to regular surveillance audits.

www.tuvnord.com.tw

ROC TW/187_SA

TWB-NHIRD

1. 臺灣人體生物資料庫為生物醫學研究的目的而建立，且經參與者「事前同意」及完善的「事後退出」流程，保障參與者權利
2. IRB EGC 雙重保障

Diseases	Sex	TWB (%)	NHIRD (%)
Diabetes	Male	10.98	12.53
	Female	7.25	8.84
Hypertension	Male	29.12	26.7
	Female	16.86	17.08
Hyperlipidemia	Male	26.69	26.09
	Female	20.51	21.05

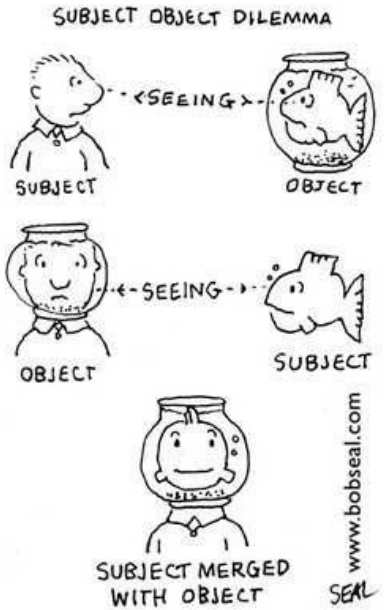
Genomics data = new experience = new world

健保資料庫是不同來源資料的合集
1. 非屬病歷、醫療及健康檢查的個人資料
2. 由醫院儀器所產生的健康檢查資料
3. 由醫師依其個人經驗判斷的病歷及醫療"資訊"

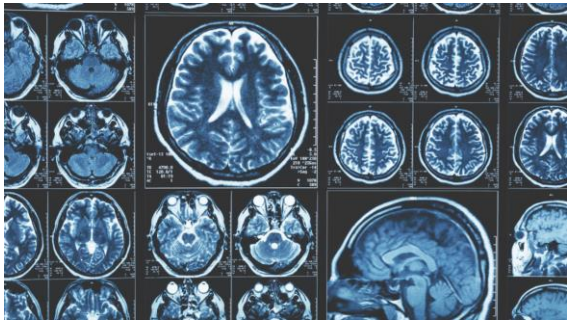
什麼是健保資料庫?

資料庫除國人的就醫、健保、出生、死亡紀錄外，加入更多社會福利相關資料，包括低收入戶家庭檔、原住民檔案等。

健保資料庫醫行嗎?
台灣人權促進會



Why medical image Ai growing so rapidly?? Tranditional experience



Evidence-based medicine

- 不同醫師依其不同的經驗對該個人的疾病狀態判斷可能不同
- 醫師的判斷又與該個人的自我認知不同
- 健保資料紀錄的是與醫療費用有關的資料

“心靈是塊白板”

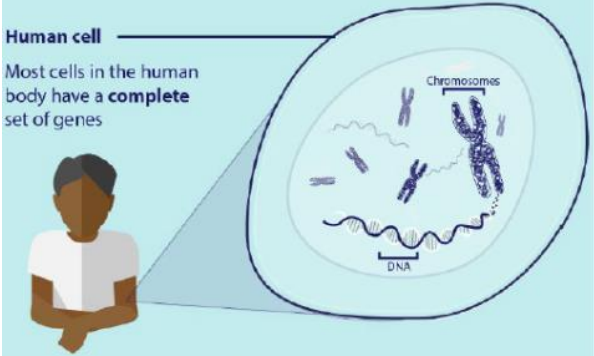
理性主義　經驗主義　德國觀念論（上）

一切知識和觀念都起源於經驗。人們能夠經驗外界事物，並對所觀察的事物加以反省，我們便得到知識。

Genetics 101

Human cell

Most cells in the human body have a **complete** set of genes



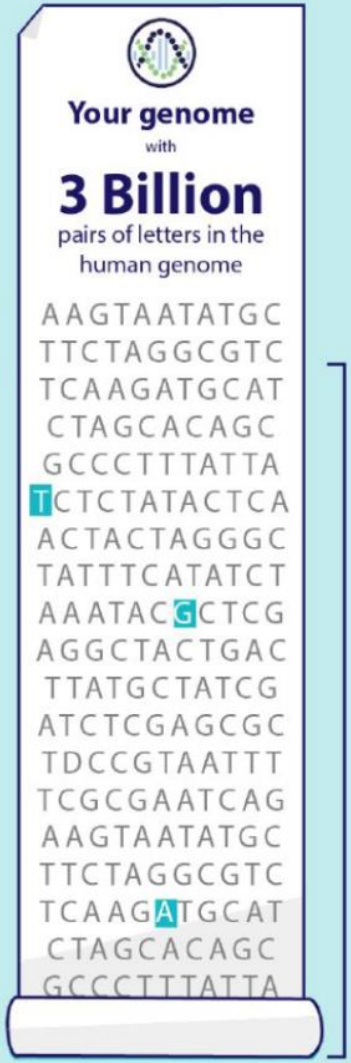
Your **genome** is one whole set of all your genes plus all the DNA between your genes.

There are around **20,000** genes in your **genome**

Now

We know that the non-gene (non-coding) parts of your genome may have a role to play so we look at the whole thing, every single letter, and how the different parts work together.

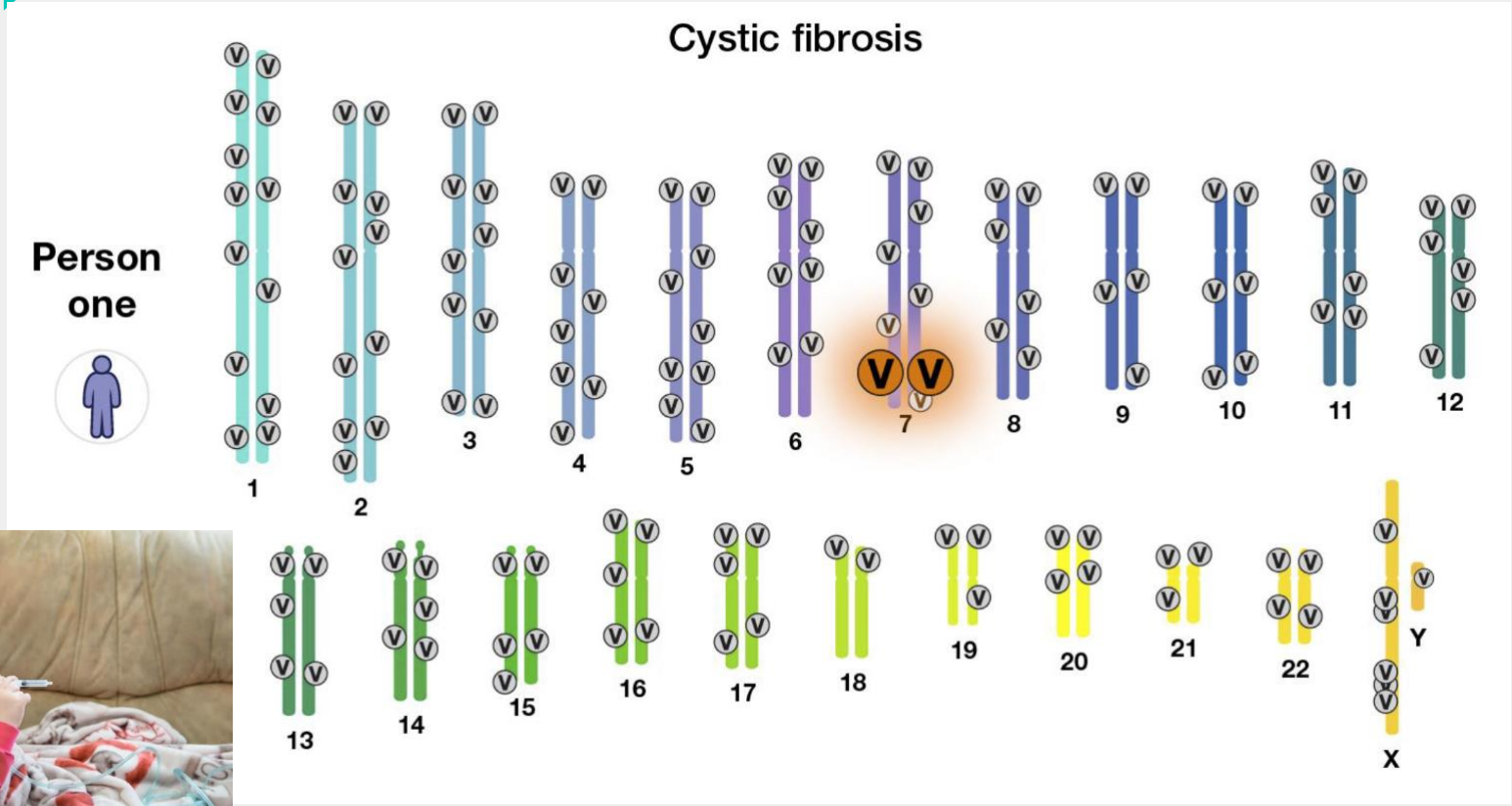
5%
were analysed
Only the active genes were looked at



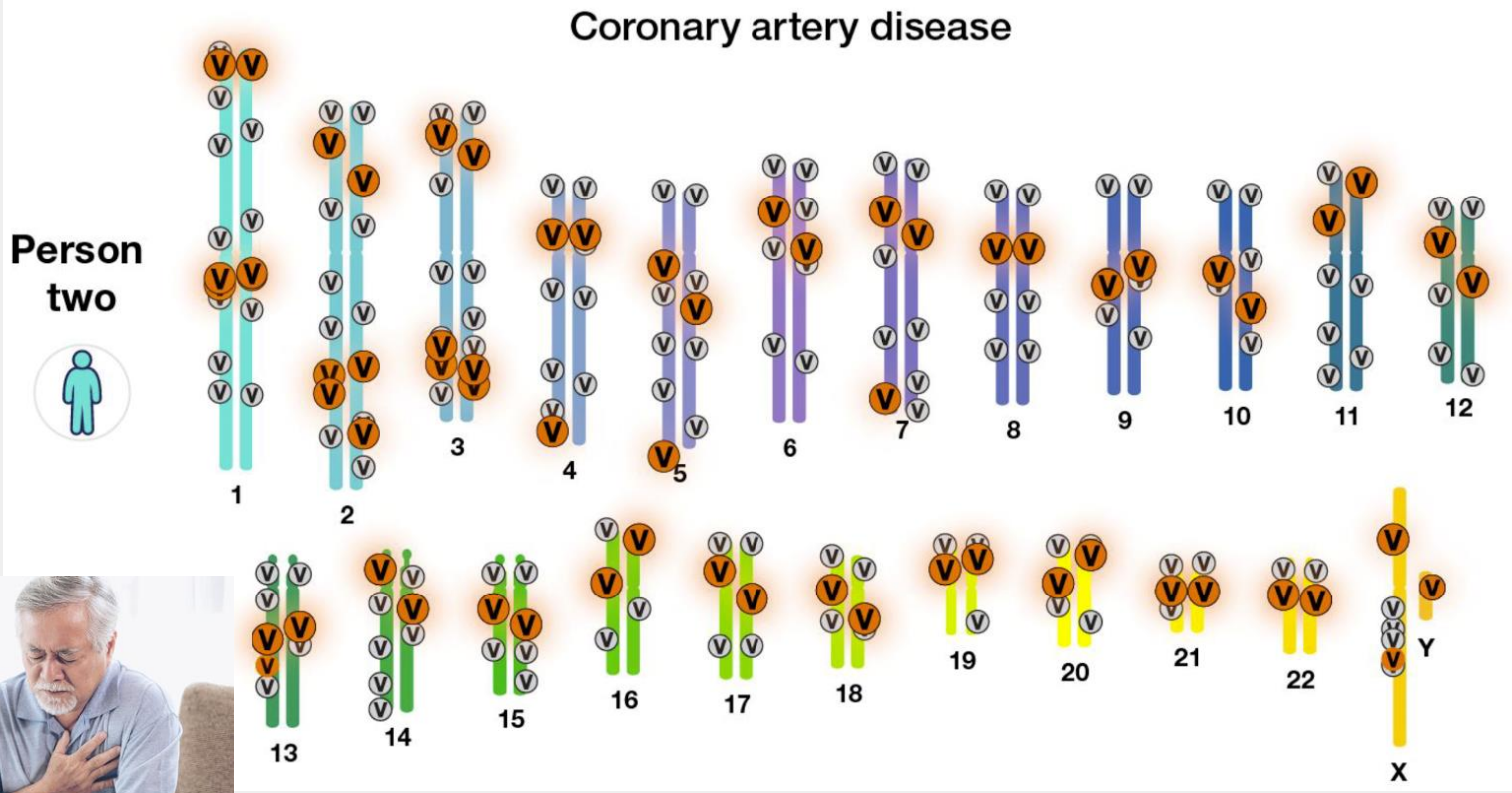
95%
were unused
All the non-gene sections that we didn't understand were disregarded as useless

BIOBANK • 健康世代

cystic fibrosis (囊狀纖維化) → transmembrane conductance regulator (CFTR) gene on chromosome 7.

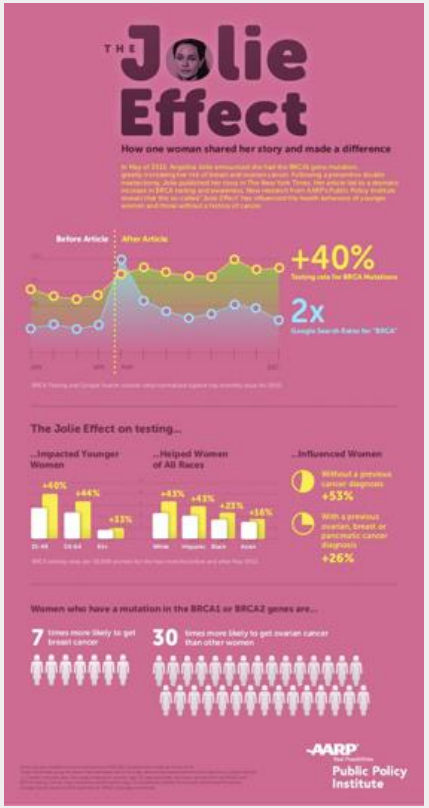


complex diseases (polygenic disease)

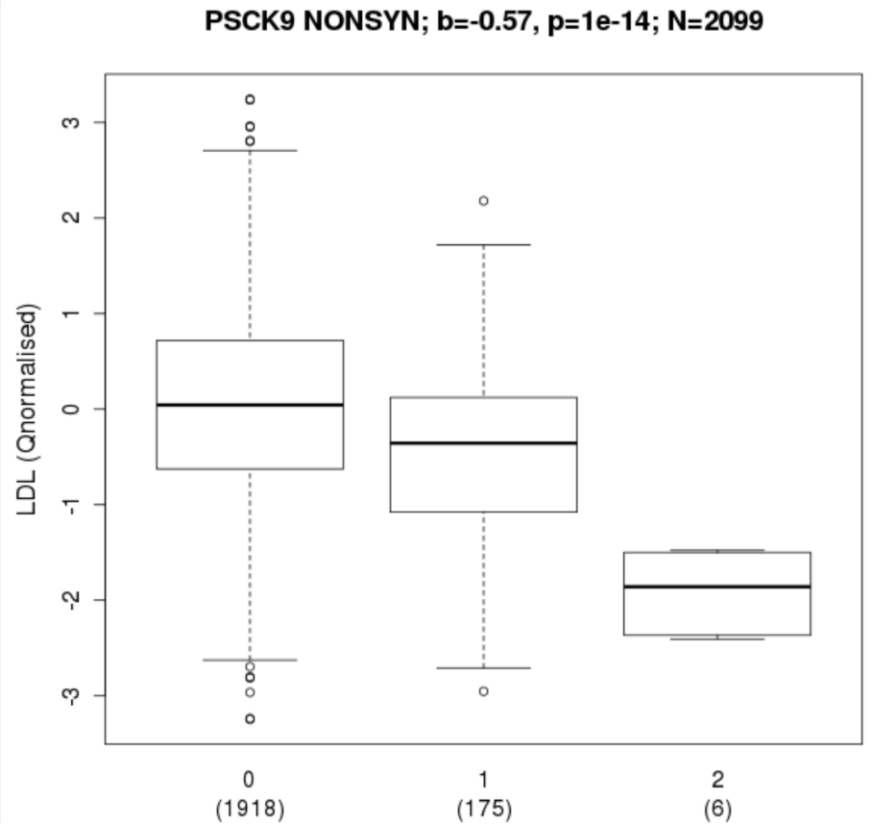


single gene

Jolie Effect

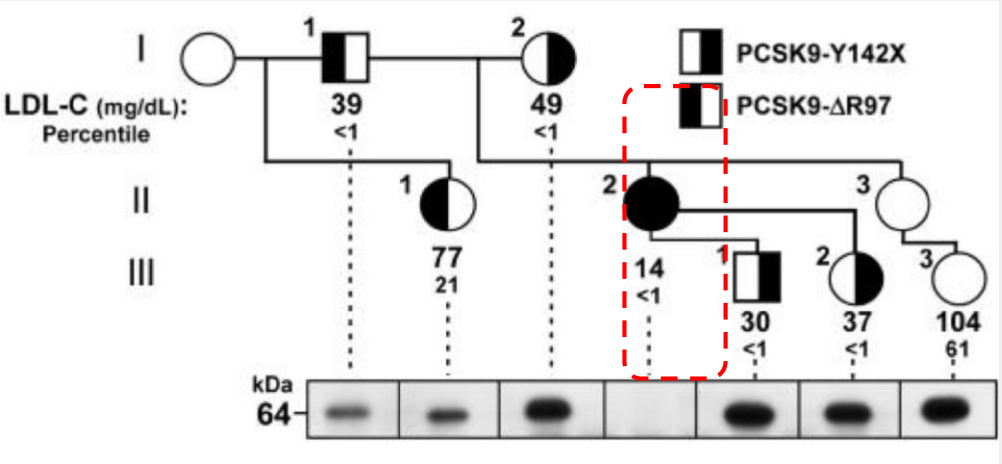


GENETIC VARIANT “RS11591147” IN PCSK9



- Carriers of **T** variant have **lower** levels of **LDL** cholesterol than carriers of G variant
- **LDL** is a strong risk factor for **heart disease**

A HUMAN KNOCK-OUT OF PCSK9 (2006)

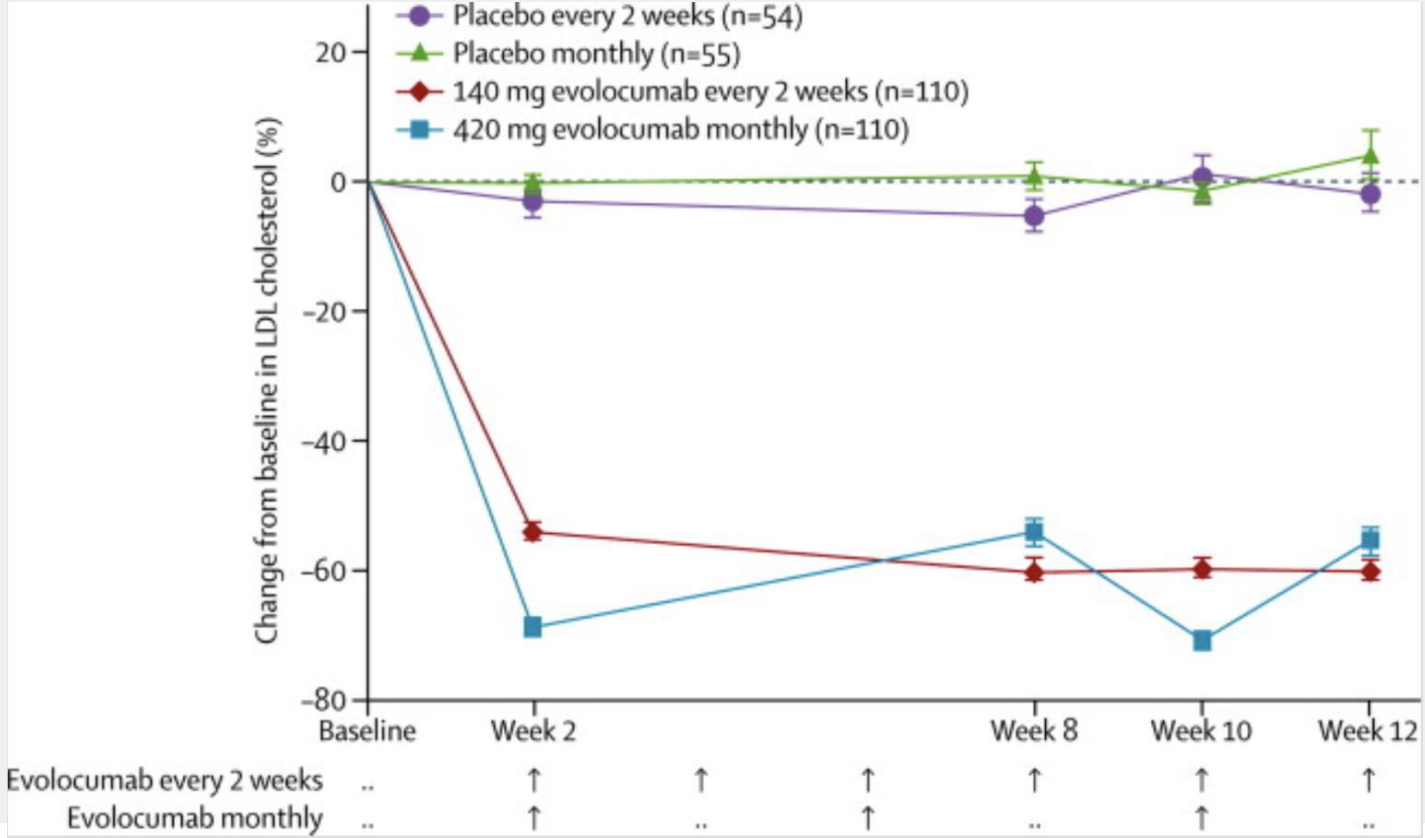


Individual **II.2 has zero working copies of PCSK9** gene

- no circulating PCSK9 and an LDL-C of only 14 mg/dL
- apparently healthy, fertile, normotensive, college-educated woman with normal liver and renal function tests who works as an aerobics instructor
- Why is this very interesting observation? **Inhibiting PCSK9 might be a safe way to reduce LDL**

Zhao et al. AJHG 2006

PCSK9 inhibition with evolocumab (AMG 145) in heterozygous familial hypercholesterolaemia (RUTHERFORD-2): a randomised, double-blind, placebo-controlled trial



Lancet Oct 2014

ORIGINAL ARTICLE

N Engl J Med 2017; 376:1713-1722

Evolocumab and Clinical Outcomes in Patients with Cardiovascular Disease

Marc S. Sabatine, M.D., M.P.H., Robert P. Giugliano, M.D., Anthony C. Keech, M.D., Narimon Honarpour, M.D., Ph.D., Stephen D. Wiviott, M.D., Sabina A. Murphy, M.P.H., Julia F. Kuder, M.A., Huei Wang, Ph.D., Thomas Liu, Ph.D., Scott M. Wasserman, M.D., Peter S. Sever, Ph.D., F.R.C.P., and Terje R. Pedersen, M.D. for the FOURIER Steering Committee and Investigators*

FDA Approves Amgen's Repatha (evolocumab) to Prevent Heart Attack and Stroke

Dec 1 2017

In the Repatha cardiovascular outcomes study (FOURIER), Repatha reduced the risk of heart attack by 27%, the risk of stroke by 21% and the risk of coronary revascularization by 22%..

precision medicine in Taiwan

Use of HLA-B*58:01 genotyping to prevent allopurinol-induced severe cutaneous adverse reactions in Taiwan: national prospective study

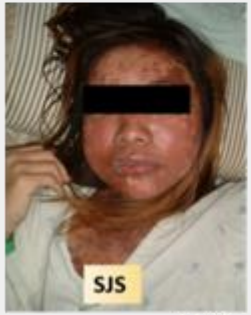
BMJ 2015;351:h4848 Sep 23, 2015

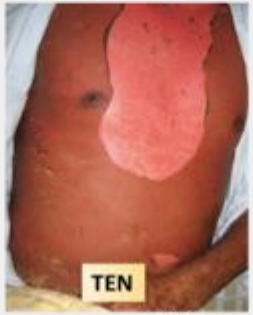
BMJ

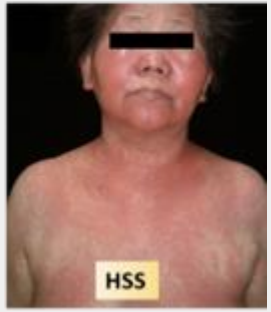
Gout

Allopurinol

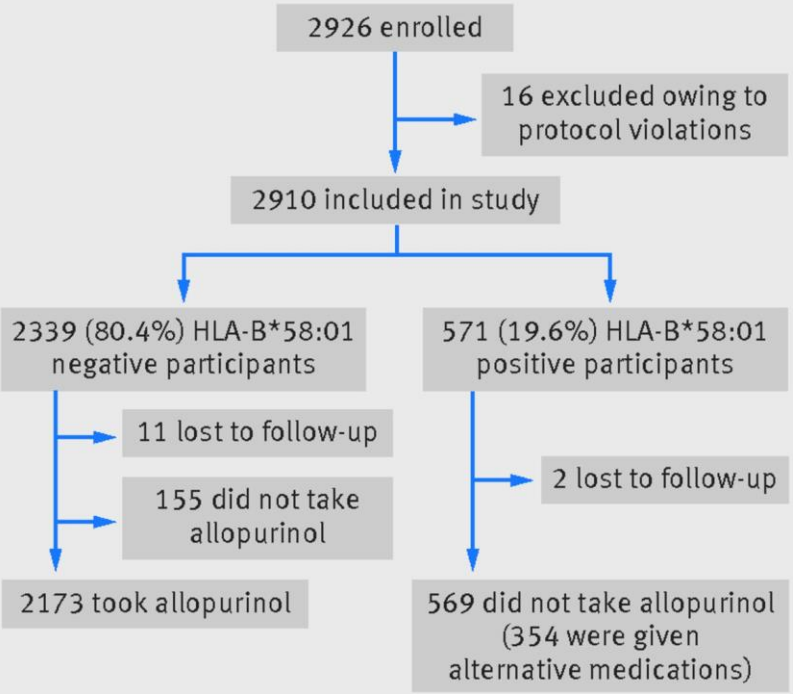
HLA-B*58:01







NHIRD-TWB → reduce health expenditures



Adverse event	HLA-B*58:01 positive participants receiving alternative drug treatment (n=354)	HLA-B*58:01 negative participants receiving allopurinol (n=2173)	Total (n=2910)
Mild cutaneous events			
Rash and itching	3*	94	97
Blisters	0	0	0
Oral ulcers	0	2	2
Rash, itching, oral ulcers, and fever	0	1	1
Rash, itching, and other adverse events	0	22	22

Records from the **National Health Insurance research database**

Incidence of SCARs ~0.3%

BMJ. 2015 Sep 23;351:h4848.

frequency distribution of pharmacogenetic phenotypes predicted by genotypes of TWB cohort

Gene	Drug	Rx[a]/year	EM	IM	PM	ADR[b] carrier rate
CYP2B6	Efavirenz	1,662,525	66.0%	30.5%	3.6%	
CYP2C19	Clopidogrel	63,664,076	39.8%	56.4%	3.8%	
CYP2C9	Celecoxib	65,058,810	93.6%	6.3%	0.1%	
CYP3A5	Tacrolimus	10,272,406	8.1%	40.6%	51.2%	
IL28	Peginterferon	40,941	88.6%	11.1%	0.3%	
NAT2	Isoniazid	7,885,251	28.8%	59.2%	12.0%	
SLCO1B1	Simvastatin	50,695,934	78.9%	19.9%	1.3%	
TPMT	Azathioprine	7,435,217	97.0%	2.9%	0.02%	
UGT1A1	Atazanavir	719,793	53.2%	39.8%	7.0%	
VKORC1	Warfarin	16,121,944	1.1%	19.2%	79.7%	
HLA-A*3101	Carbamazepine	17,078,849				2.0%
HLA-B*1502	Carbamazepine	17,078,849				4.1%
HLA-B*5701	Abacavir	3,049,217				0.2%
HLA-B*5801	Allopurinol	23,888,472				10.5%
MT-RNR1	Amikacin	321,561				4.7%

[a]Rx = prescriptions.

[b]*ADR* = adverse drug reactions.

NPJ Genom Med. 2021 Feb 11;6(1):10.

NOTCH3 cysteine-altering variant is an important risk factor for stroke in the Taiwanese population

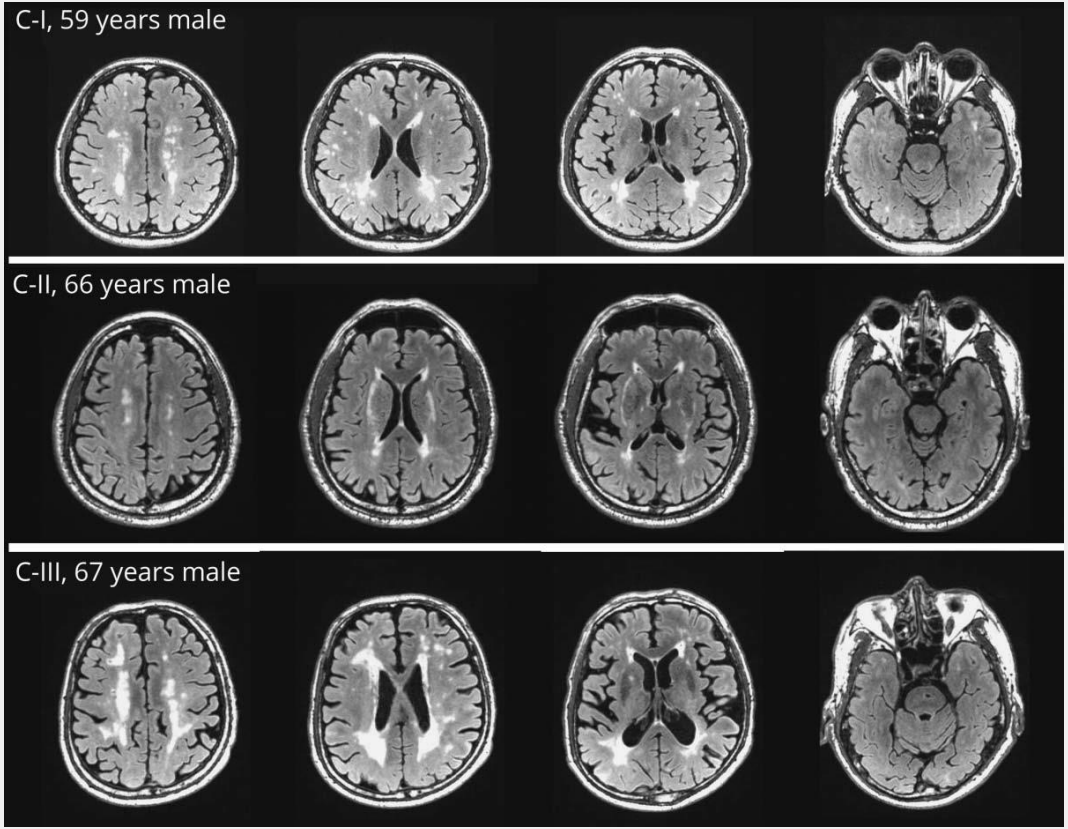
We **queried the Taiwan Biobank database** for cysteine-altering mutations in exons 2–24 of NOTCH3 within these genomes. The reference coding sequence of NOTCH3, NM_000435.3, was used for annotating the variants. (**p.R544C** (c.1630G>A), **p.C853Y** (c.2558G>A), and **p.C884Y** (c.2651G>A))

The cysteine-altering NOTCH3 variants identified from the Taiwan Biobank database were **genotyped in the control participants and patients with stroke** using the TaqMan genotyping assay

Only the NOTCH3 p.R544C variant was found in 4 individuals (TP-VGH (n =550))

	Controls			Patients with stroke		
	Taiwan biobank (n = 6,488)	TP-VGH (n = 550)	Biobank + TP-VGH (n = 7,038)	TP-VGH (n = 350)	TC-VGH (n = 450)	TP-VGH + TC-VGH (n = 800)
Male	2,293 (35.3)	230 (41.8)	2,523 (35.8)	248 (70.9)	327 (72.7)	575 (71.9)
Age, y	48.4 ± 10.9	56.7 ± 15.0	49.1 ± 11.5	64.4 ± 13.6	67.5 ± 13.2	66.2 ± 13.5
Hypertension	682 (10.6)	186 (33.8)	868 (12.4)	248 (70.9)	329 (73.1)	577 (72.1)
Diabetes	300 (4.6)	72 (13.1)	372 (5.3)	130 (37.1)	168 (37.3)	298 (37.3)
Hyperlipidemia	388 (6.0)	138 (25.1)	526 (7.5)	129 (36.9)	181 (40.2)	310 (38.8)
Smoking habit	1,228 (18.9)	98 (17.8)	1,326 (18.8)	122 (34.9)	136 (30.4)	258 (32.4)
Alcohol consumption	387 (6.0)	108 (19.6)	495 (7.0)	56 (16.0)	87 (23.2)	143 (19.7)
Family history of stroke	1,146 (17.8)	—	—	40 (14.1)	14 (4.0)	54 (8.5)
***NOTCH3* p.R544C mutation (+)**	56 (0.9)	4 (0.7)	60 (0.9)	8 (2.3)	9 (2.0)	17 (2.1)

Physical examination revealed that they were free of neurologic deficits. Three of them received brain MRI scans, and all had a variable degree of leukoencephalopathy



multi-gene

GWAS study

- 偏頭痛及阿茲海默氏症之全基因體關聯性研究-台北榮總
- 利用全基因關聯性研究於影響人類身高差異性之基因鑑定與功能性探討-中國醫藥大學
- 利用台灣生物資料庫探討成人肺功能變化之全基因體關聯研究-中央研究院

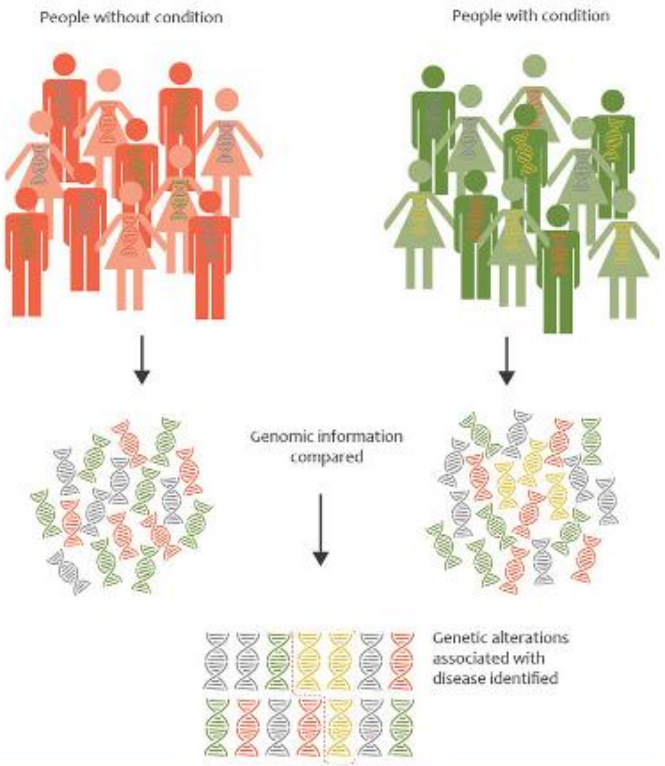
- Calculate polygenic risk score (PRS) for individual j using m SNPs

$$\mathrm{PRS}_j = \frac{\sum_{i=1}^{m} \ln(OR_i) \times SNP_{ij}}{m}$$

where

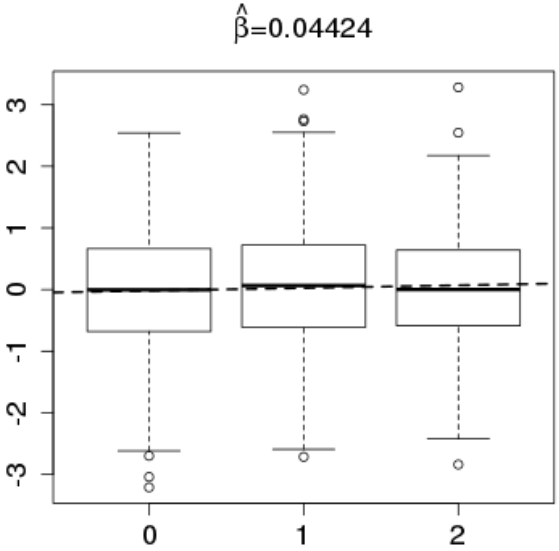
- $\ln(OR_i)$ = effect size for SNP i from discovery sample
- SNP_{ij} = number of risk alleles (0,1,2) for SNP i, individual j in target sample
- m = number of SNPs considered in test set

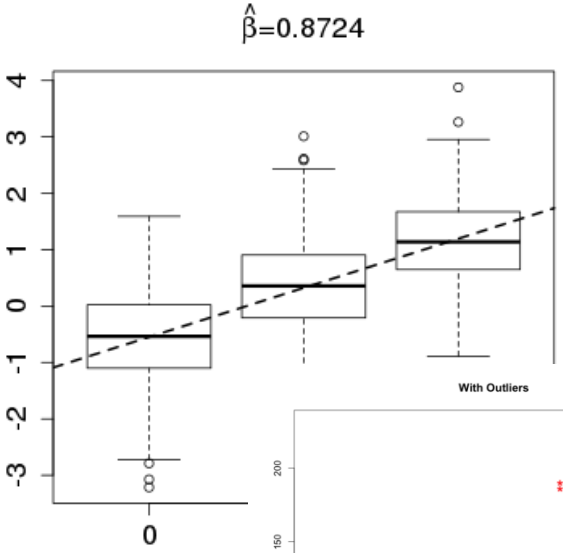
Purcell / ISC et al. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder Nature 2009

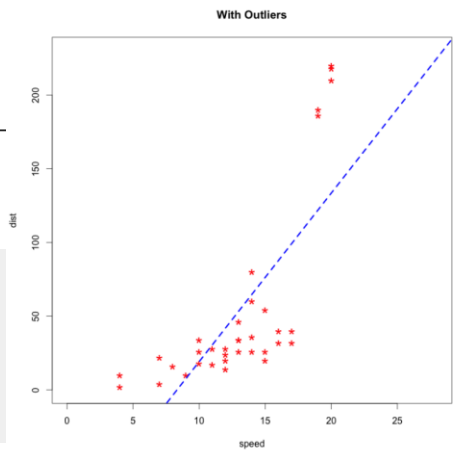


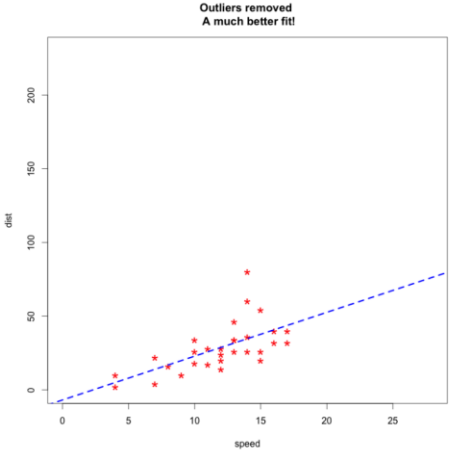
BIOBANK • 健康世代

association study linearMod <- lm(glucose ~ bmi, data=twb

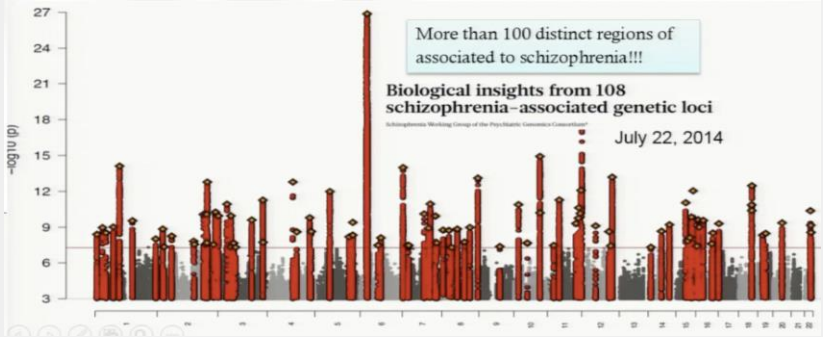








the power of sample size - schizophrenia | psychiatric genomics consortium



polygenic risk score

Discovery GWAS

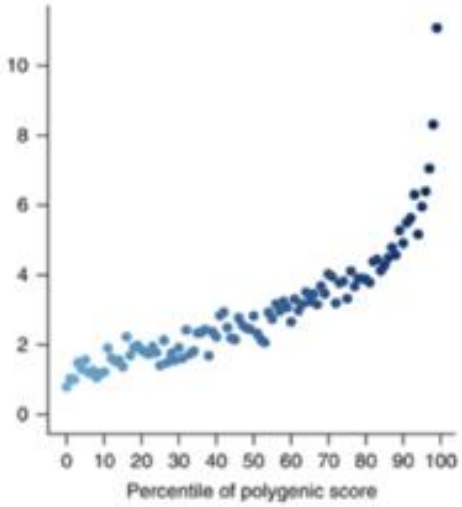
	Weight*	Risk Allele
SNP1	0.2	A
SNP2	-0.3	C
SNP3	0.1	G

Individual	Alleles SNP1	Alleles SNP2	Alleles SNP3
1	AT	AA	CG
2	AA	CA	GG
3	TT	AC	CG
4	TT	AA	GG
5	TA	CA	GC
6	AT	CA	CG
7	AA	AA	GG
8	AA	CC	CG
9	TA	CC	GC
10	AT	AA	CG

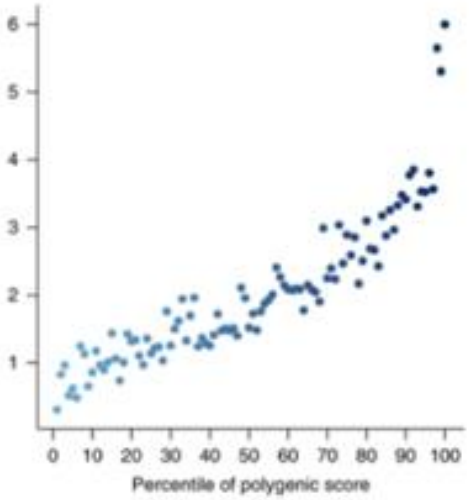
PRS:

Individual	SNP 1	SNP 2	SNP 3	**PRS**
1	0.2+0.0	0.0+0.0	0.0+0.1	**0.3**
2	0.2+0.2	-0.3+0.0	0.1+0.1	**0.3**
3	0.0+0.0	0.0-0.3	0.0+0.1	**-0.2**
4	0.0+0.0	0.0+0.0	0.1+0.1	**0.2**
5	0.0+0.2	-0.3+0.0	0.1+0.0	**0.0**
6	0.2+0.0	-0.3+0.0	0.0+0.1	**0.0**
7	0.2+0.2	0.0+0.0	+0.1+0.1	**0.6**
8	0.2+0.2	-0.3-0.3	0.0+0.1	**-0.1**
9	0.0+0.2	-0.3-0.3	0.1+0.0	**-0.3**
10	0.2+0.0	0.0+0.0	0.0+0.1	**0.3**

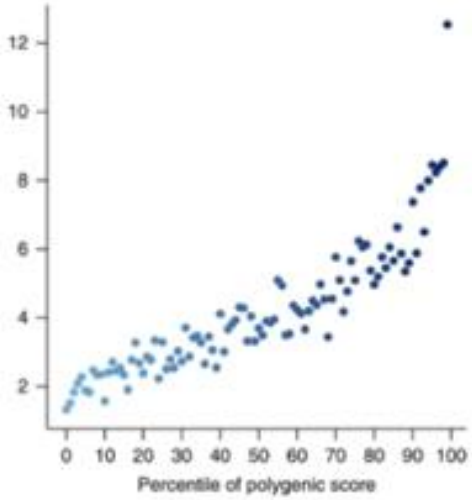
disease risk prediction



Prevalence of Coronary Artery Disease



Prevalence of Type 2 Diabetes

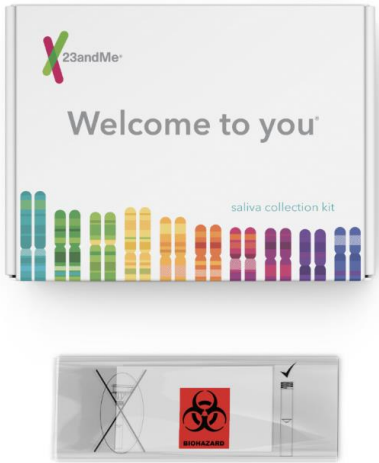


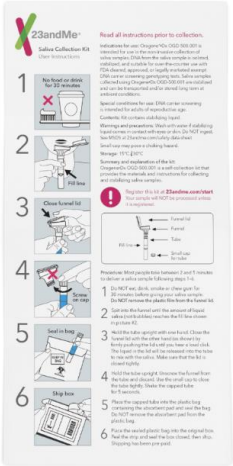
Prevalence of Breast Cancer

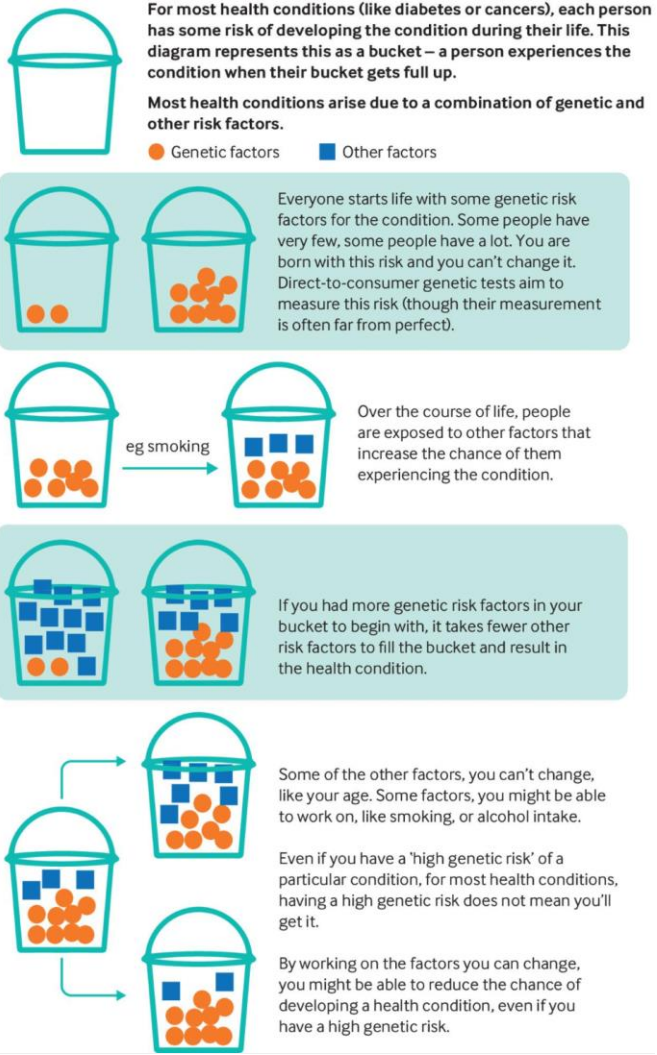
"…it is time to contemplate the inclusion of polygenic risk prediction in clinical care, and discuss relevant issues."

Khera et al. Nature Genetics 50, 1219–1224 (2018)

Direct to consumer genetic testing (DTG)

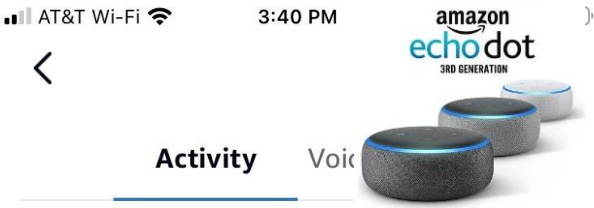






BMJ 2019;367:l5688

Rethinking data quality!! Big data create the world!!

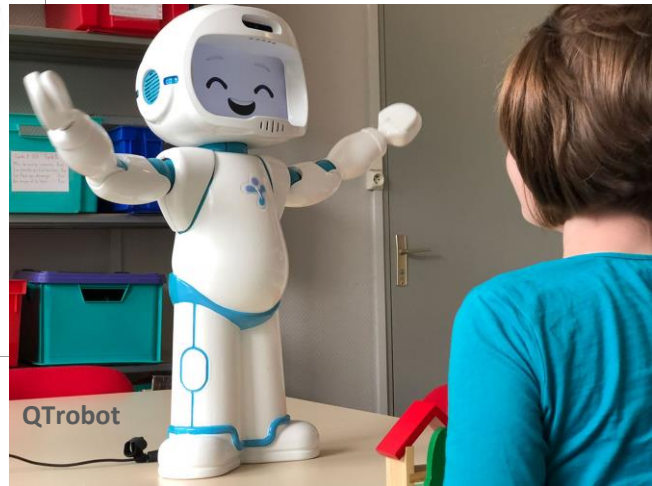


Tell me a challenge to do

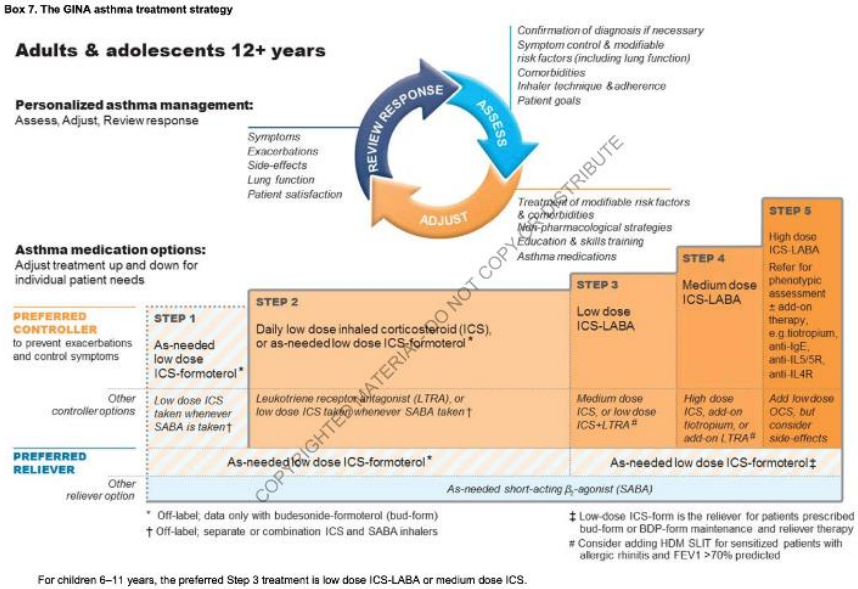
Here's something I found on the web. According to ourcommunitynow.com: The challenge is simple: plug in a phone charger about halfway into a wall outlet, then touch a penny to the exposed prongs.

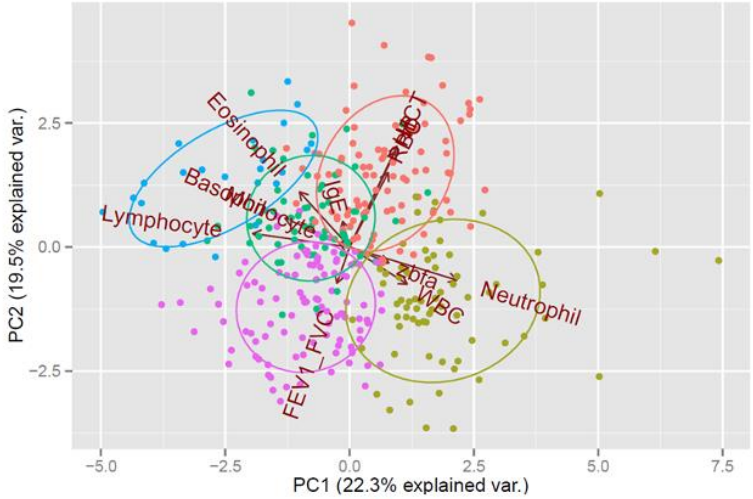
LEARN MORE ON OURCOMMUNITYNOW.COM

Therapy Robot Teaches **Social Skills** to Children With Autism.

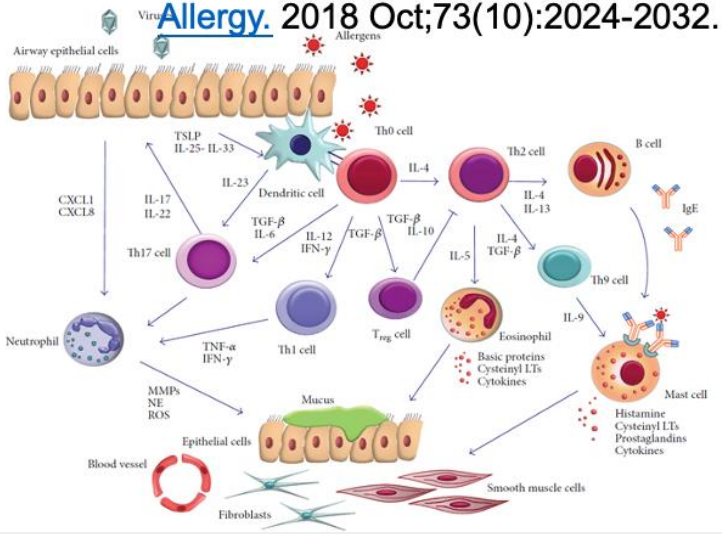


Precision medicine



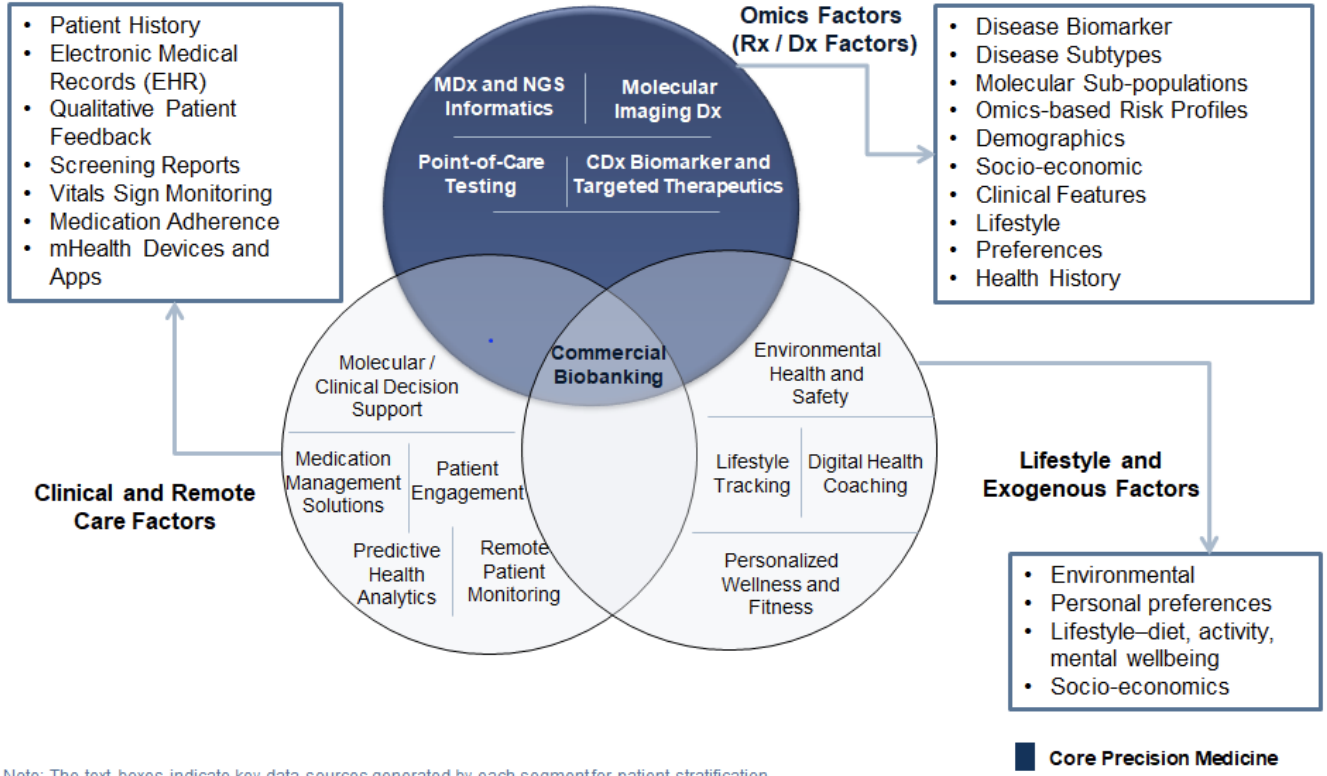


Allergy. 2018 Oct;73(10):2024-2032.



Nature 520, 609–611

Biobank - provide clinical research support that translates into bedside diagnostics and treatments, and advances research technologies into clinical applications



Lion man 35,000~40,000 years old

精準醫療並非只是醫學領域的轉變
而是人看待自己與形塑自己方式的轉變
是一個重新認識自己的過程