基因型鑑定與精準醫療

李玲慧
中央研究院生物醫學研究所
國家基因體醫學研究中心

精準醫療的目標



量身訂製的治療與
健康照護計畫

個人化的疾病治療

↓

疾病風險管控

↓

健康促進

精準醫療面面觀



Cancer:
Sequence analysis
(Somatic mutation)
- Tailored treatment
- Outcome prediction



Rare genetic diseases:
Sequence (Genotype)
analysis (Germline
mutation)
- Diagnosis
- Tailored management
- Family planning



Common diseases:
Genotype analysis
(Germline variants)
- Disease risk prediction
- Tailored management
- Drug response

演講大綱

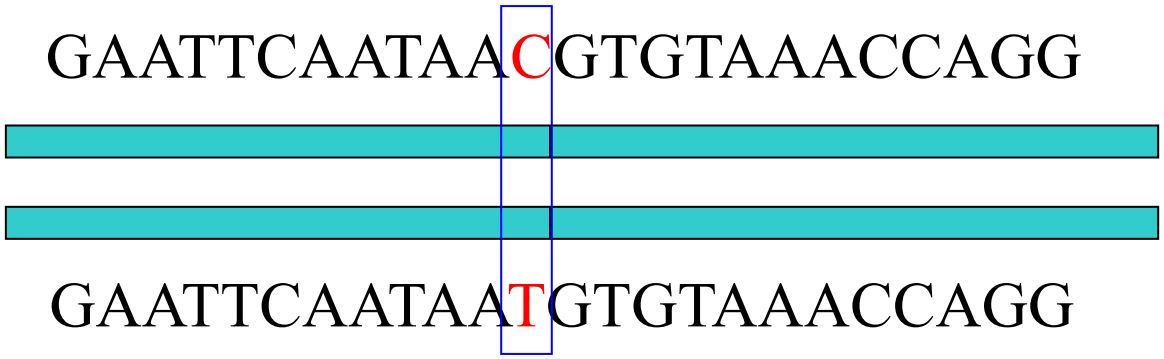
- 基因型鑑定平台



- 全基因體關聯性分析(GWAS)與多基因風險分數(PRS)

Single nucleotide polymorphisms (SNP)
(單一核苷酸多形性)

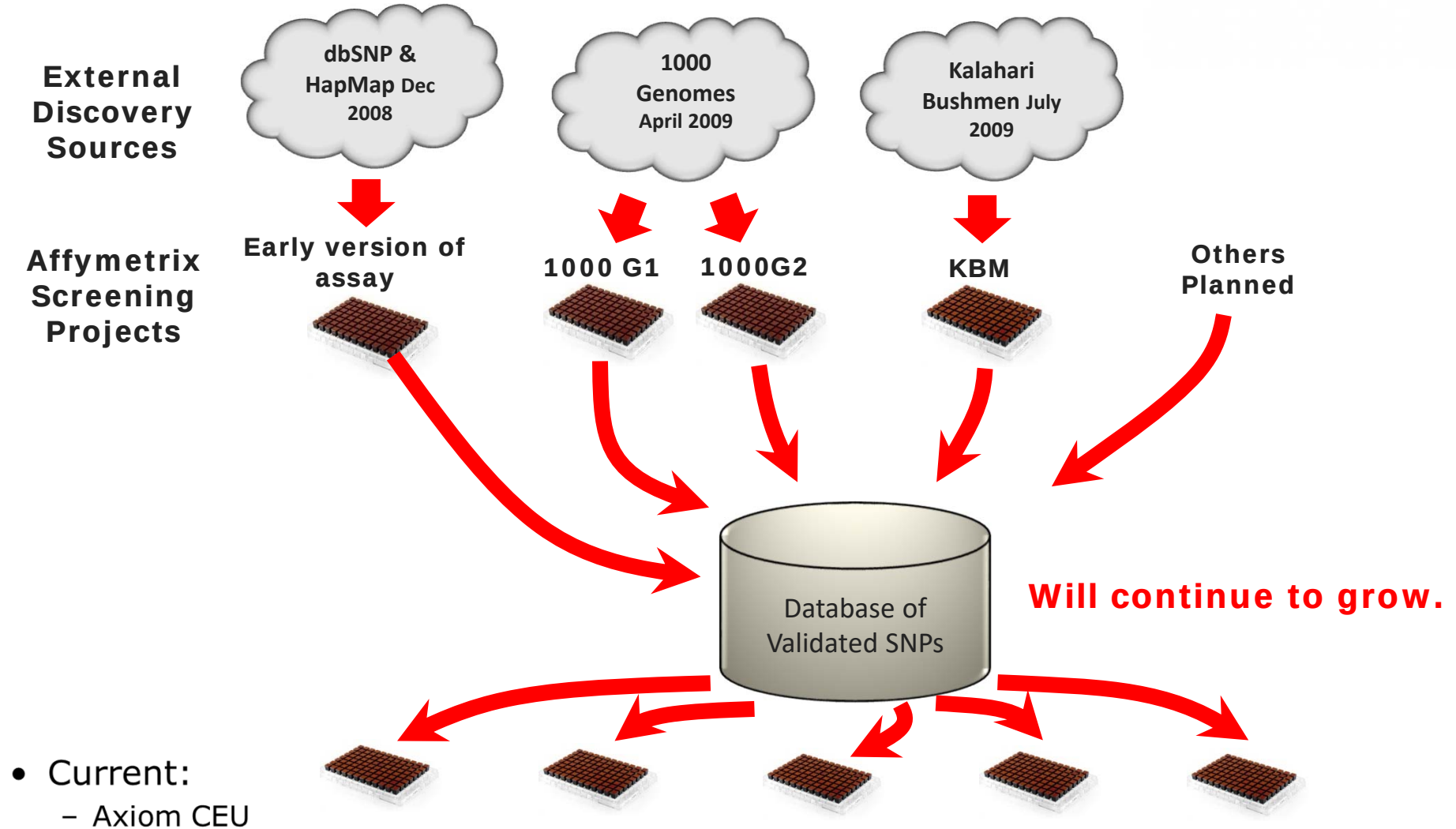
Single nucleotide polymorphisms (SNPs) are DNA sequence variations that occur when a single nucleotide (A,T,C,or G) in the genome sequence is changed. This is the most frequent form (1/500-1000 bp) of sequence variation in human genome.



A genotype (基因型) is the combination of alleles (對偶基因) for a particular gene or locus.
The process of determining a genotype is called genotyping (基因型鑑定).

Axiom SNP Database

ThermoFisher SCIENTIFIC



- Current:
  - Axiom CEU
  - Axiom ASI
  - Axiom CHB

TWB2 SNP Array

GWAS imputation grid
- East and South Asia population and Taiwan biobank
- Common & low frequency SNPs (down to 1%)

Disease related variants
- Susceptibility genes for complex diseases
- Variants for neurological disorders including Alzheimer's and Parkinson's diseases
- Variants associated with cardiometabolic traits
- Variants associated with diabetes

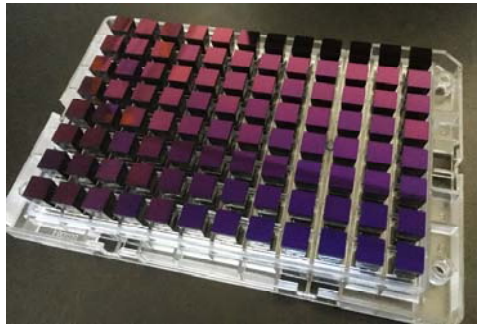
Common and rare cancer variants
- From NHGRI-EBI GWAS catalog and publications
- Markers for rare missense variants

Functional variants
- Loss of function (LoF) variants
- eQTLs
- Rare and nonsynonymous coding variants

~686K markers

"Mapping" variants

"Actionable" variants



Variants with known associations to human health and traits
- Pathogenic per ClinVar
- ACMG published list of genes
- BRCA1/2, CFTR, DMD, APOE

NHGRI-GWAS Catalog
- Covering the complete NHGRI catalog of published GWAS

Immune and transplant related variants
- Human leukocyte antigen (HLA)
- Killer-cell immunoglobin-like receptors (KIR)
- Autoimmune and inflammatory variants

Pharmacogenomic variants
- ADME markers from the CPIC
- Genetic variants associated with drug efficacy and adverse events

TWB2晶片的準確性

Concordance of genotypes by TWB2 Array and Sequencing

Minor Allele Frequency (%)	Number of Markers	Concordance Rate (S.D.)
0 - 0.1	110,497	99.6% (5.7)
0.1 - 1	47,631	99.8% (3.8)
1 - 5	204,981	99.9% (0.9)
5 - 10	73,263	99.8% (1.9)
10 - 20	77,125	99.8% (2.5)
29 - 30	54,953	99.7% (3.1)
30 - 40	47,449	99.6% (4.0)
40 - 50	44,707	99.4% (6.3)
Total	660,606	99.8% (3.5)

Consistency of imputed ABO and HLA types in parent-child pairs

	ABO	HLA-A	HLA-B	HLA-C	HLA-DPB1	HLA-DQB1	HLA-DRB1
Consistency	99.90%	97.80%	97.40%	98.90%	99.20%	99.70%	97.50%

基因型鑑定資料來源：台灣人體生物資料庫

台灣漢人疾病基因變異點頻率分析

Table 2. Recessive genetic disorders with the highest carrier rates in the TWB cohort.

Recessive genetic disorder	Gene	Carrier rate
G6PD deficiency	G6PD	2.49%
Citrullinemia type II	SLC25A13	1.94%
Wilson disease	ATP7B	1.77%
Pendred syndrome	SLC26A4	1.70%
Krabbe Disease	GALC	1.67%
Nonsyndromic hearing loss and deafness	GJB2	1.59%
Nagashima-type palmoplantar keratosis	SERPINB7	1.43%
Primary carnitine deficiency	SLC22A5	0.90%
Hereditary spastic paraplegia 5	CYP7B1	0.83%
Congenital hypothyroidism	TSHR	0.71%
Sitosterolemia	ABCG5	0.66%
Beta Thalassemia	HBB	0.59%
Total iodide organification defect	TPO	0.56%
Joubert Syndrome	CEP290	0.51%
Usher syndrome	USH2A	0.50%
Phenylketonuria	PAH	0.48%
Mucolipidosis type III	GNPTAB	0.44%
Waardenburg syndrome	EDNRB	0.40%
Congenital Disorder of Glycosylation 1a	PMM2	0.40%
Glutaric aciduria 1	GCDH	0.38%

Table 3. Autosomal dominant disorders with the highest allele frequencies in the cohort.

Dominant genetic disorder	Gene	Allele freq.
DFNA2 nonsyndromic hearing loss	KCNQ4	1.24%
Hereditary pancreatitis	SPINK1, PRSS1	1.05%
Familial hypercholesterolemia	APOB, LDLR	0.89%
Cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy type 1 (CADASIL)	NOTCH3	0.88%
Familial hypertrophic cardiomyopathy	TCAP	0.45%
Spastic paraplegia 4	SPAST	0.29%
Maturity-onset diabetes of the young type 3	HNF1A	0.17%

Table 4. Cancer susceptibility conditions with the highest fraction of affected individuals in the cohort.

Cancer susceptibility syndromes	Gene	Allele freq.
Hereditary cancer-predisposing syndrome	CHEK2 + RAD51	0.58%
Juvenile polyposis syndrome	SMAD4	0.44%
Adenomatous polyposis coli	APC	0.42%
Breast-ovarian cancer, familial 1	BRCA1	0.41%
Li-Fraumeni syndrome 1	TP53	0.40%
Lynch syndrome	MSH6	0.40%
MYH-associated polyposis	MUTYH	0.40%
Neuroblastoma	KIF1B	0.40%
PTEN hamartoma tumor syndrome	PTEN	0.40%
Prostate cancer	EHBP1	0.40%
Breast-ovarian cancer, familial 2	BRCA2	0.29%

npj Genomic Medicine (2021) 10

TWB2晶片的改版

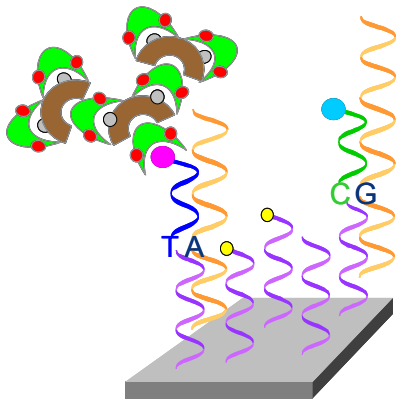
TWB2晶片在經過十萬個樣本的實測後

1. 刪除表現不佳的SNP位點
2. 刪除大部份allele frequency ≤ 0.1%的SNP位點
3. 增加臨床醫師提供的疾病相關SNP位點
4. 改善CNV probes的分佈，增強涵蓋基因的DNA片段之探針數目
5. 增加藉由Taiwan Reference Genome Project得到專屬台灣漢人的新增DNA序列片段中的SNP位點

TWB2.0 and TPM 2.0 Array
(For Research Use Only)

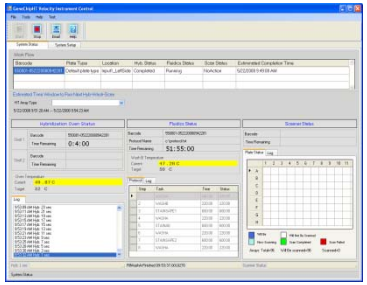
	TWB2.0 (TPM)	TPM 2.0
SNP	686,463 SNPs	743,227 SNPs
SNPs 挑選之來源	■ TWB Array，共104,909個SNP點。 ■ 根據TWB Array的使用經驗及Taiwan Biobank 1000名參與者次世代定序的結果，設計446K個SNP點。 ■ 與臨床顯著相關的SNP點約105K個。 ■ Thermo Fisher Scientific歷年替各國Biobank設計的疾病相關標記SNP。 ■ 其他與藥物反應、藥物代謝相關，如MHC，PGX等基因上的SNP點。 ■ 偵測拷貝數變異（CNV）的SNP點。	■ 根據Taiwan Biobank次世代定序及Han Chinese次世代定序的結果，設計596K個SNP點。 ■ 與臨床顯著相關的SNP點約124K個。 ■ Thermo Fisher Scientific歷年替各國Biobank設計的疾病相關標記SNP。 ■ 其他與藥物反應、藥物代謝相關，如MHC，PGX等基因上的SNP點。 ■ 偵測拷貝數變異（CNV）的SNP點。
【註】兩款 array plate相同的SNP點約496K個。		

Thermo Fisher Axiom Genotyping Solution



NEW ASSAY





COMPREHENSIVE USER-FRIENDLY SOFTWARE SUITE

96 ARRAY PLATE



600K SNPs PER “PEG”



Axiom assay process

Produce an array with probes specific to a genomic position of interest

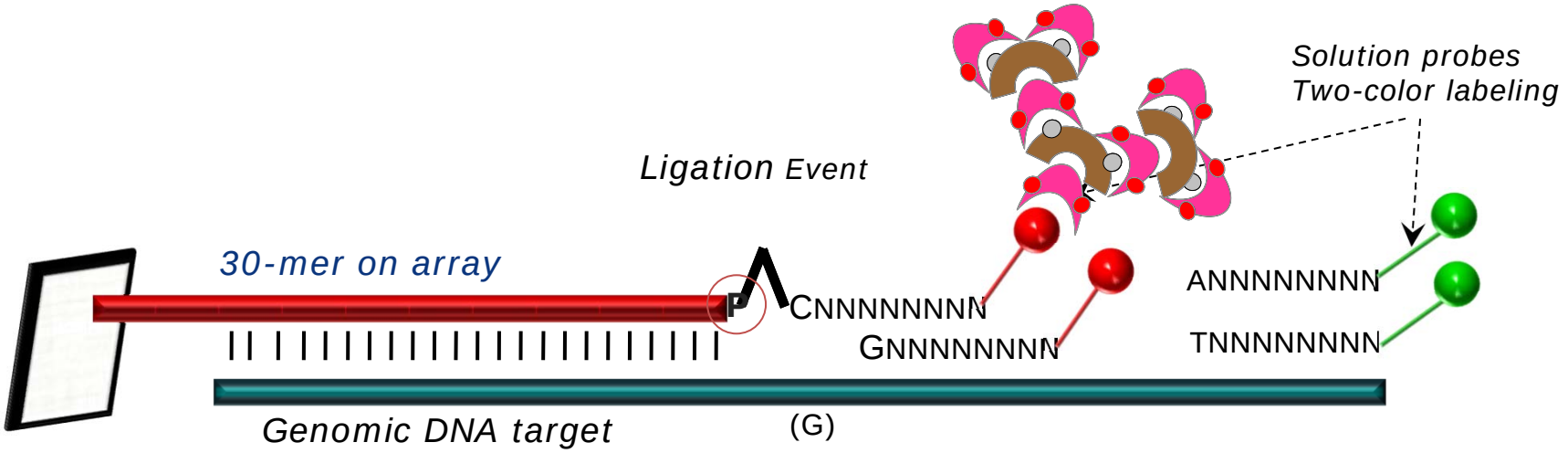
Amplify total genomic DNA, fragment it and hybridize to the array

Hybridize solution probes (9 mer) that pair a specific “hapten” to a specific base

Use DNA ligation to covalently bind only the correct base

Wash stringently to remove solutions probes that are not ligated, and target DNA

Stain, fix and scan



Axiom genotyping workflow

Monday



•Start DNA amplification on 3 plates of genomic DNA (plates A, B, and C)

•Hands-on time:

Tuesday



▪ Fragment and precipitate 2 plates (A & B) of amplified DNA

▪ Freeze plate C overnight

▪Hands-on time:

Wednesday



▪ Re-suspend 2 plates (A & B) of precipitated DNA

▪ Fragment and precipitate plate C

▪Begin hybridization on plate A

▪Hands-on time

Thursday



▪ Begin hybridization on Plate B
▪Re-suspend plate C
▪Load GT reagents for Plate A
▪Begin hybridization on plate C
▪Begin scanning Plate A
▪ Hands-on time:

Friday

•Load GT reagents for Plate B
•Load GT Reagents for Plate C
•Begin scanning plates B & C

• Hands-on time



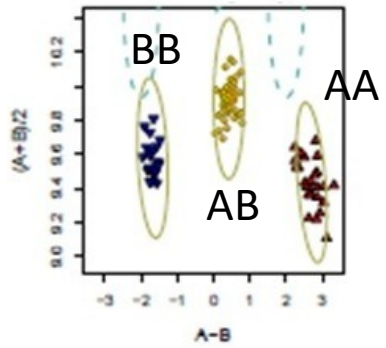






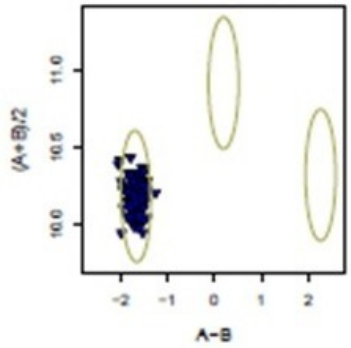
Genotype calling

Poly High Resolution



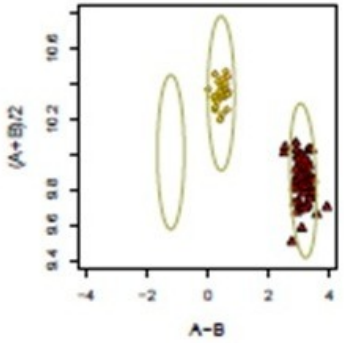
- Good cluster Resolution
- At least 2 examples of minor allele

Mono High Resolution



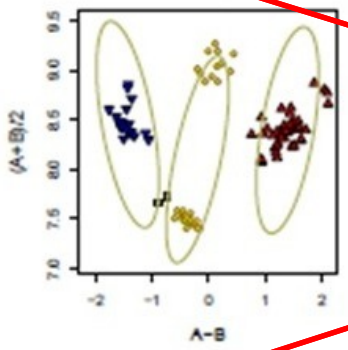
- All genotyped samples are monomorphic

No Minor Homozygote



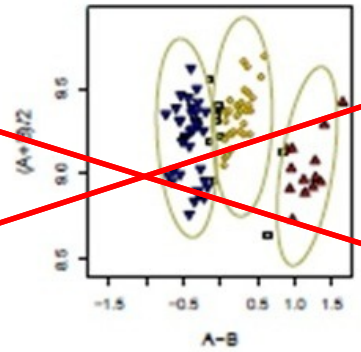
- Less than 2 examples of the minor allele

Off Target Variant



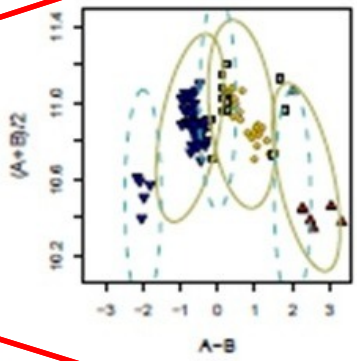
- Off Target Variants can be subjected to the off target variant calling algorithm

Call Rate Below Threshold



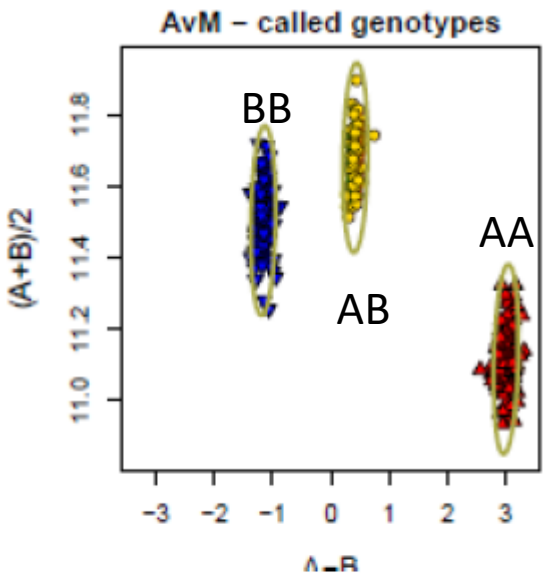
- Call Rate is below threshold, but all other cluster properties are normal

Other

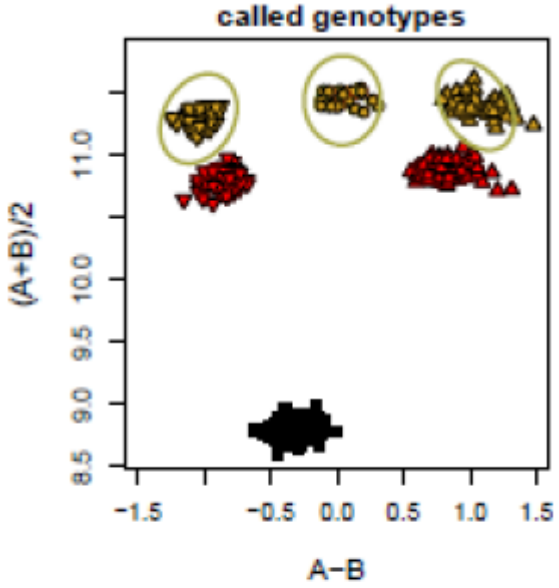


- One or more cluster properties are below threshold values

Genotype calling and CNV detection

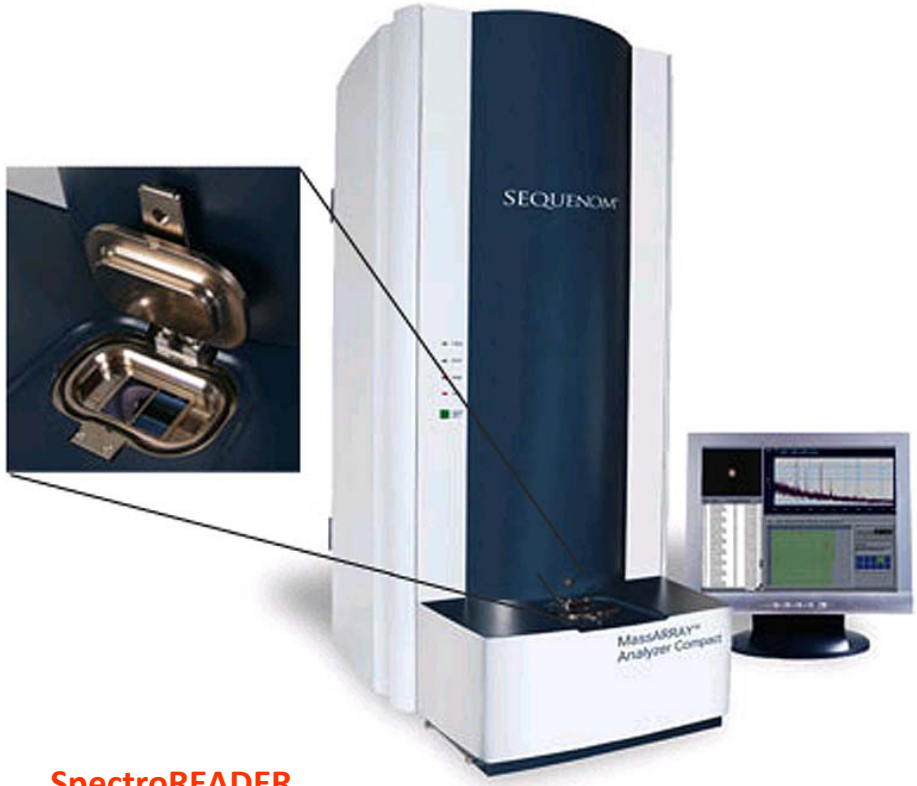


Bi-allelic variation

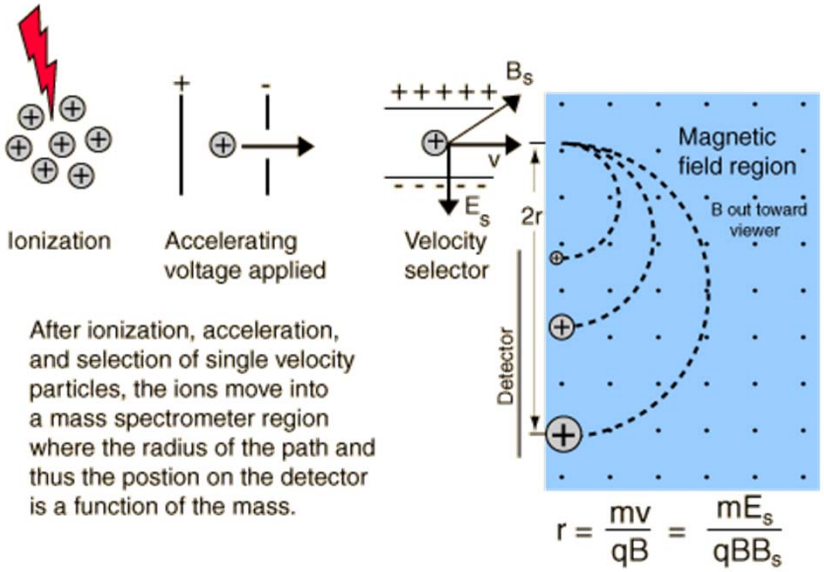


Copy number aware GT

Candidate gene/loci SNP Genotyping using **Sequenom MassARRAY System**



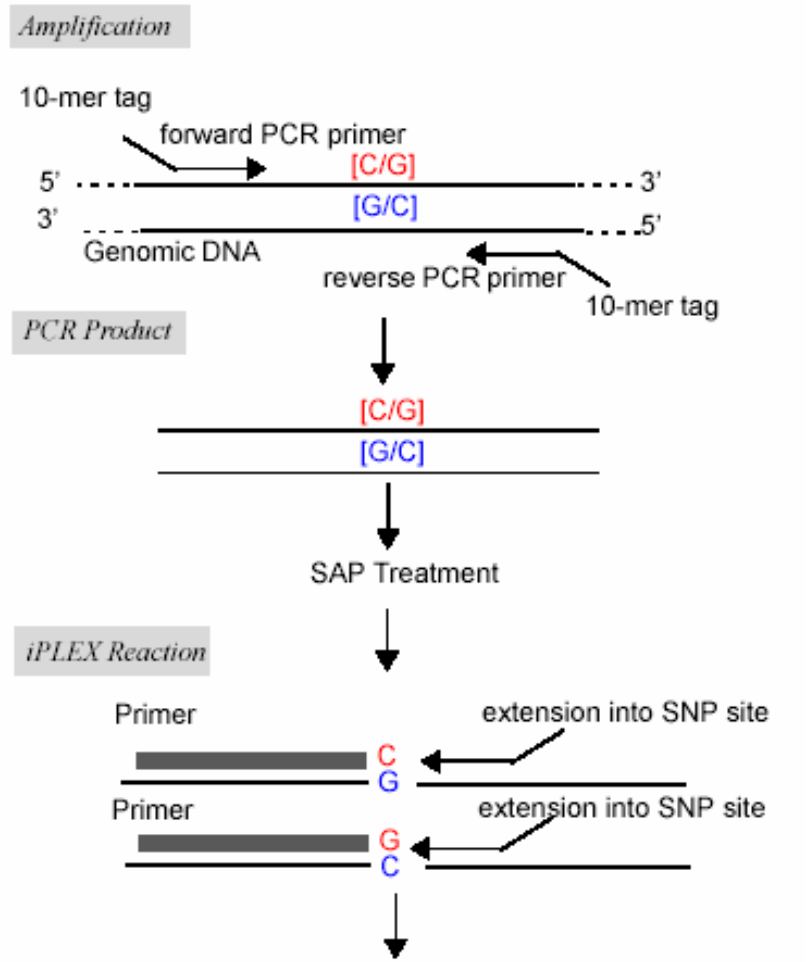
SpectroREADER

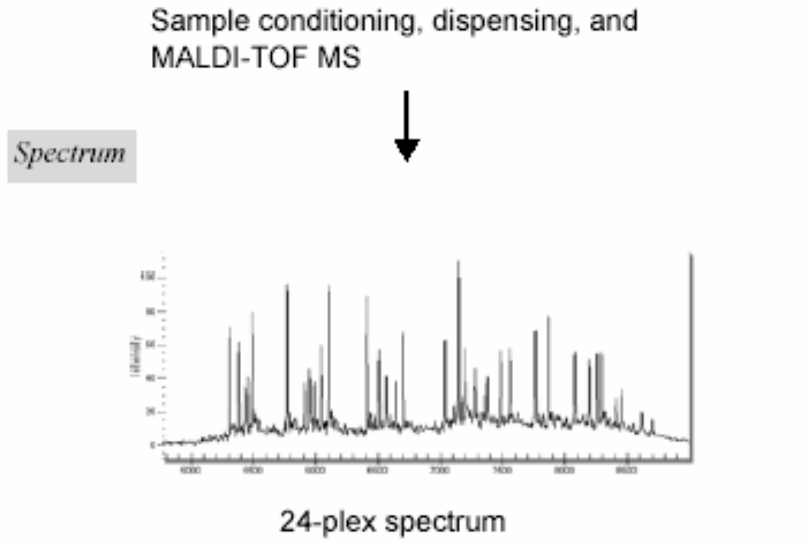


MALDI-TOF MS

(Matrix-Assisted Laser Desorption/ Ionization Time-Of-Flight) Mass Spectrometry

iPLEX SNP Genotyping





iPLEX ™

Novel Mass-Modified terminators

Terminator	A	C	G	T
A	0	-24	16	55.9
C	24	0	40	79.9
G	-16	-40	0	39.9
T	-55.9	-79.9	-39.9	0

演講大綱

- 基因型鑑定平台



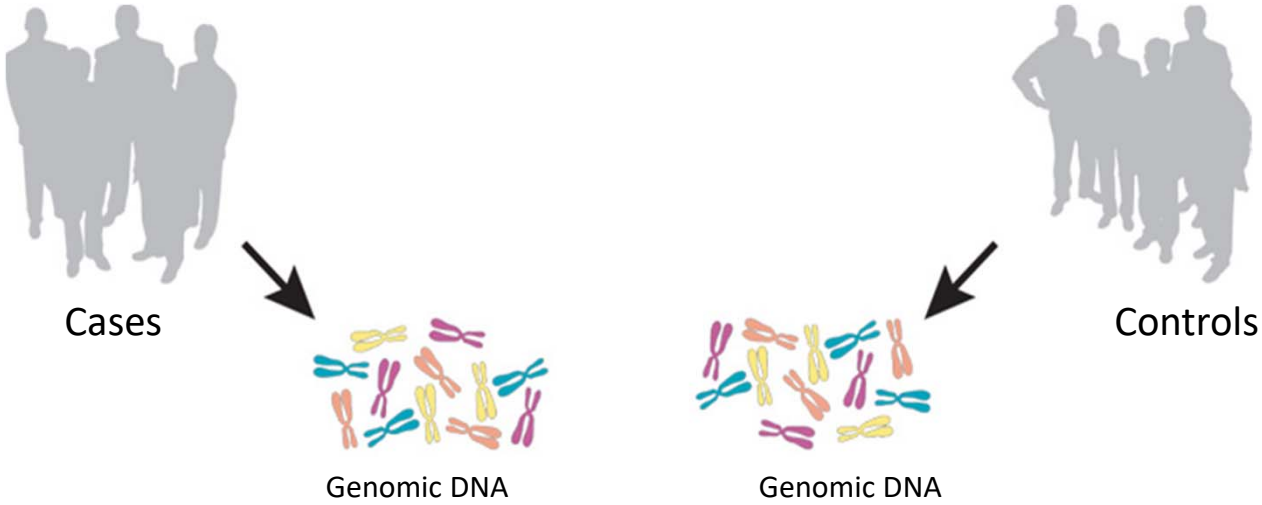
- 全基因體關聯性分析(GWAS)與多基因風險分數(PRS)

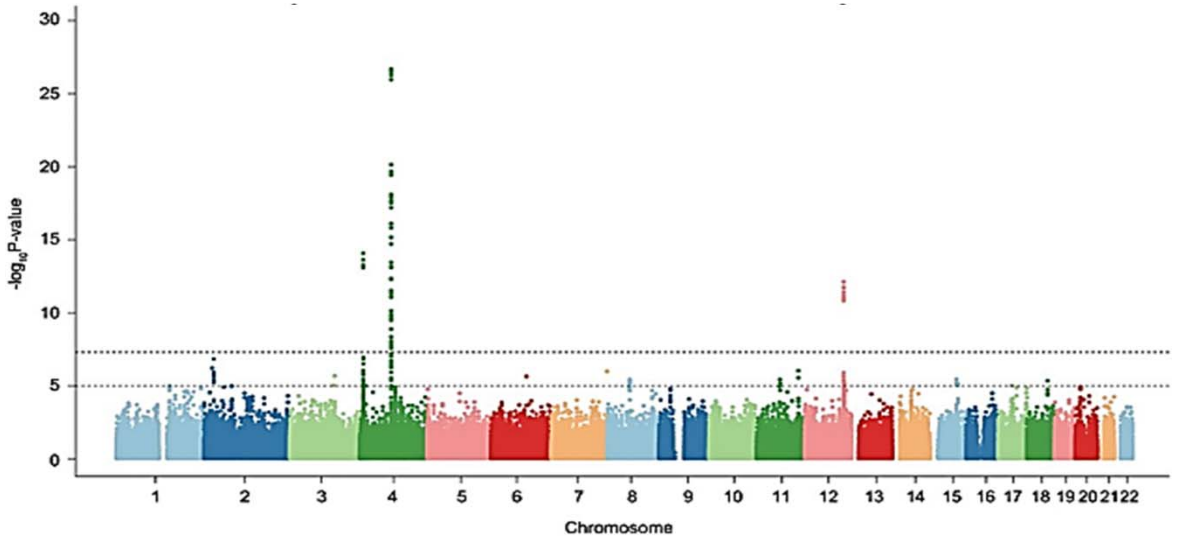
利用基因型鑑定進行基因體醫學研究的類型

- 尋找疾病易感性基因 (disease susceptibility genes)
- 尋找藥物不良反應的遺傳標記(genetic markers for adverse drug reaction)
- 尋找藥物反應有效性的遺傳標記(genetic variants responsible for drug efficacy)
- 尋找與癌症發展相關的染色體異常(chromosomal alteration and genetic mutation of cancer genes)
- 偵測血型、HLA分型、藥物代謝基因基因型、罕見疾病基因突變、遺傳性癌症基因突變

全基因體關聯性分析(GWAS)

Do SNPs anywhere on the genome associate with trait, disease susceptibility, or drug response?





多基因風險分數（polygenic risk score）

- 從GWAS得到與複雜特徵(complex traits)或疾病(disease)相關聯的SNPs數目通常都很多
- 這些SNPs對每個特徵或疾病的貢獻度通常都很小
- 在這些SNPs的變異中只有少部分是影響特徵或造成疾病的原因

→每一個SNP對於特徵或疾病的預測能力很有限

Table 1 Estimation of phenotypic variance explained from genetic relationships among unrelated individuals by restricted maximum likelihood

		No. SNPs	$L(H_0)^a$	$L(H_1)^b$	LRT[c]	σ_g^2 (s.e.)	σ_e^2 (s.e.)	σ_P^2 (s.e.)	$h^{2\,d}$ (s.e.)
295K SNPs	Raw	294,831	−1950.89	−1936.12	29.53	0.445 (0.084)	0.546 (0.082)	0.991 (0.023)	0.449 (0.083)
	Adj.[e]	294,831	−1950.89	−1936.12	29.53	0.532 (0.101)	0.458 (0.098)	0.991 (0.023)	0.537 (0.100)
295K/516K SNPs[f]	Raw	294,831/516,345	−1950.89	−1935.94	29.89	0.449 (0.085)	0.536 (0.083)	0.986 (0.022)	0.456 (0.085)
	Adj.	294,831/516,345	−1950.89	−1935.87	30.04	0.536 (0.101)	0.449 (0.099)	0.985 (0.022)	0.544 (0.101)

The proportion of phenotypic variance explained by the SNPs as 0.45 (s.e. = 0.08), a nearly tenfold increase relative to the 5% explained by published and validated individual SNPs.

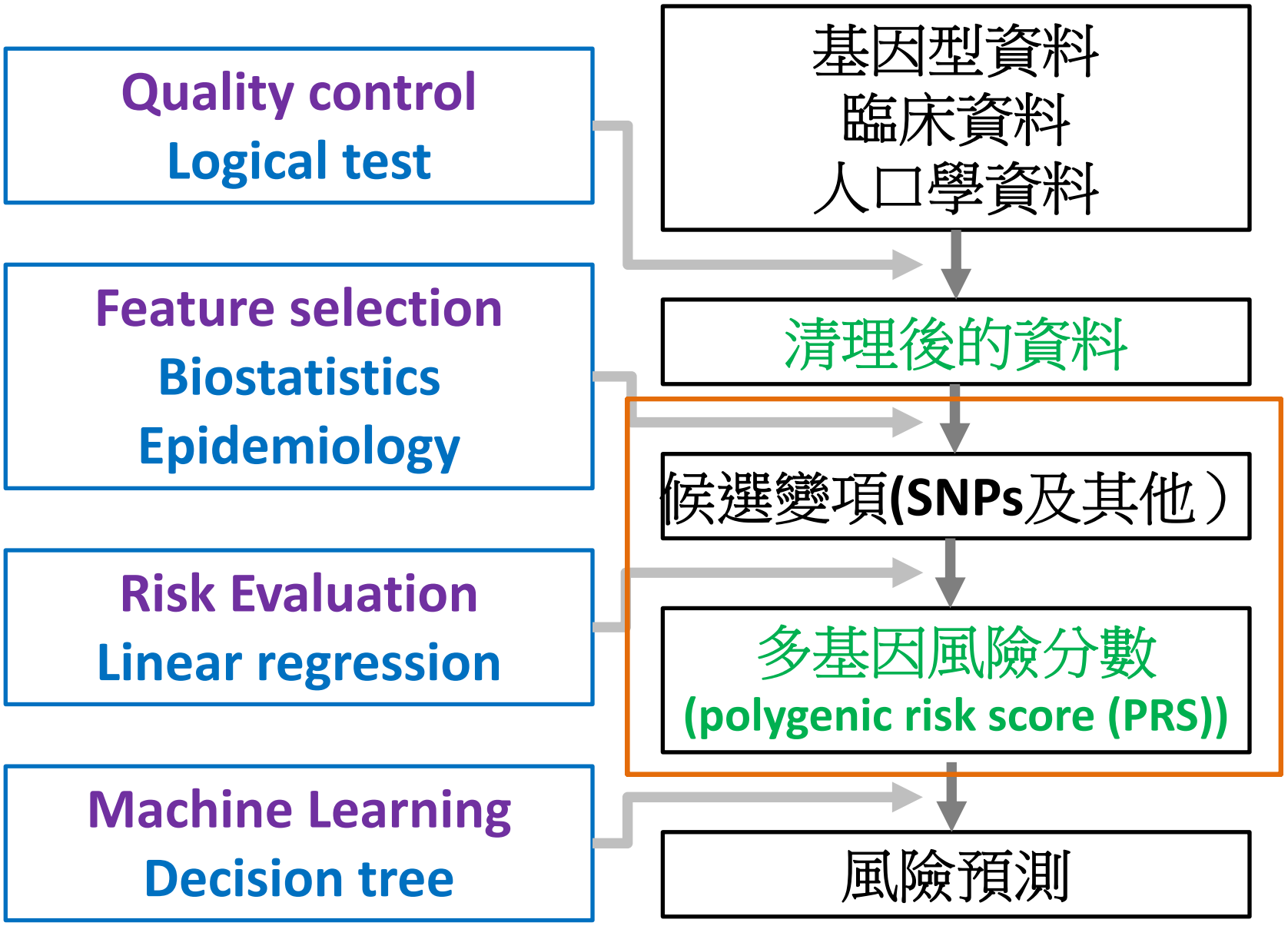
Nat Genet 2010, 42:565-569

Genetic Risk for Disease

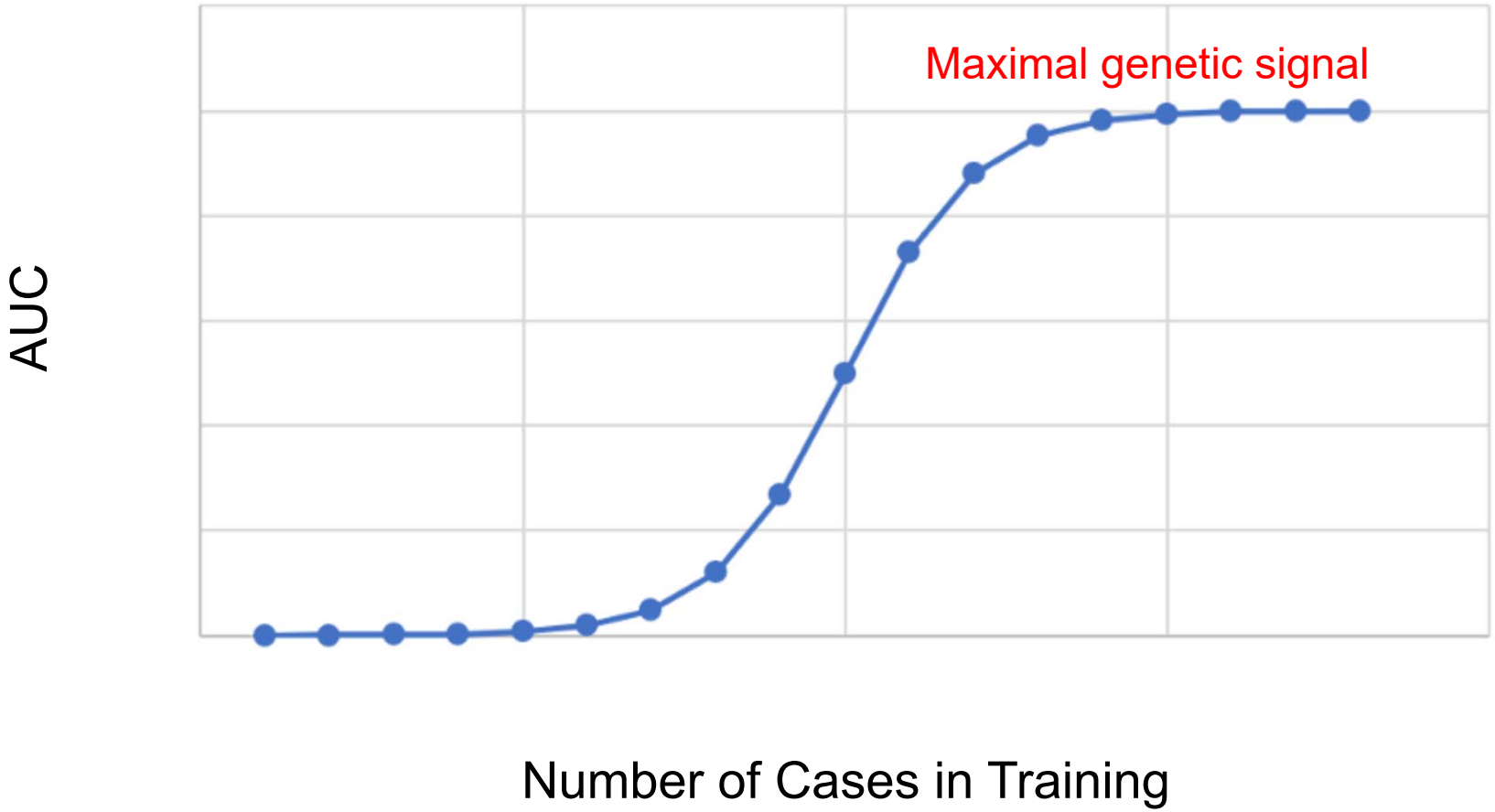
	Rare Mutations	Polygenic Risk Scores (409K Eur)	Action
Coronary Artery Disease	0.4%	8%	Early testing, Cholesterol control
Type 2 Diabetes	0.1%	3.5%	Early screening Diet & weight control
Breast Cancer	0.25%	1.5%	Early screening

Khera et al., Nat Genet. 2018; 50:1412-1425

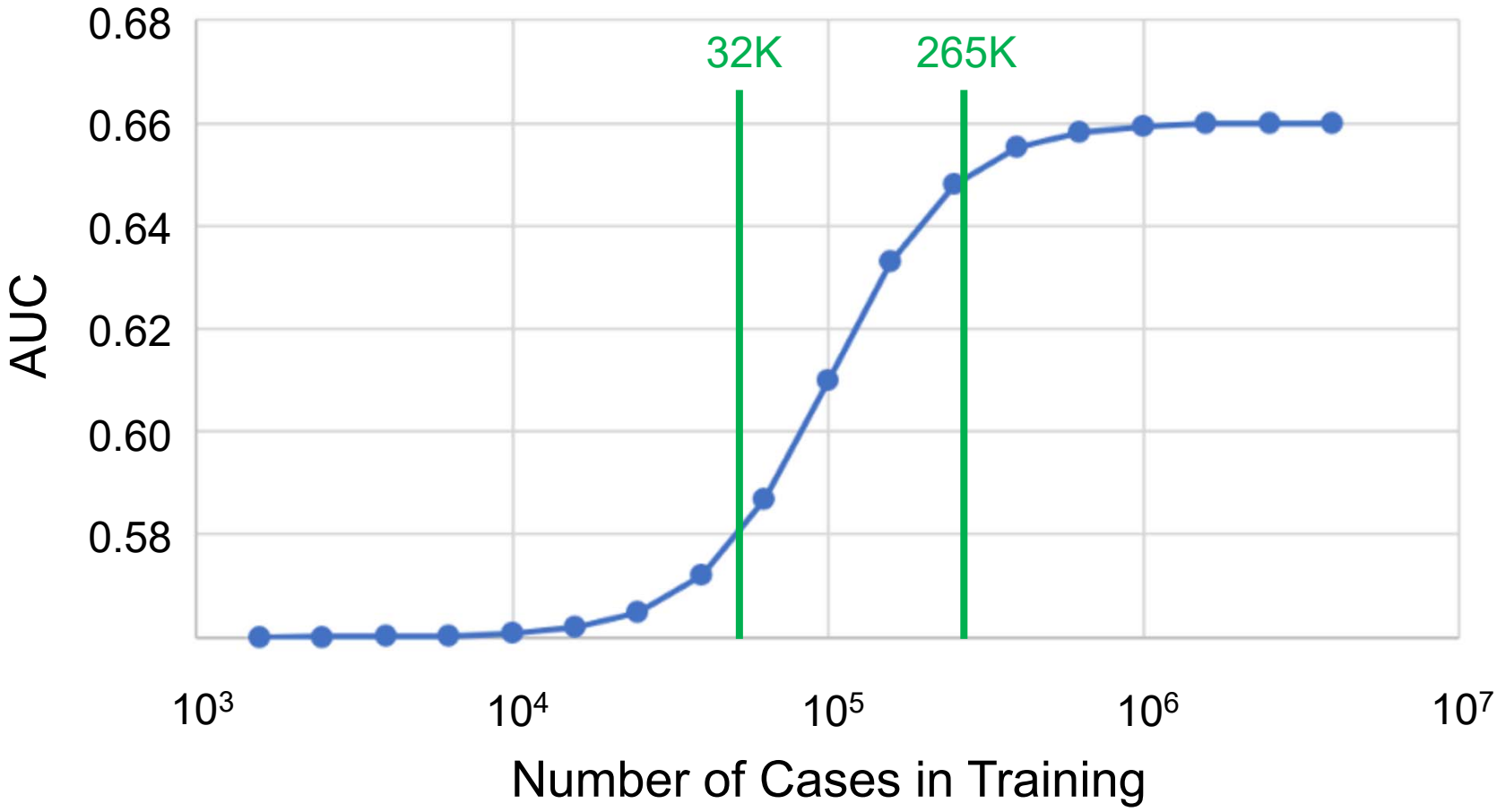
疾病風險預測模式的建立



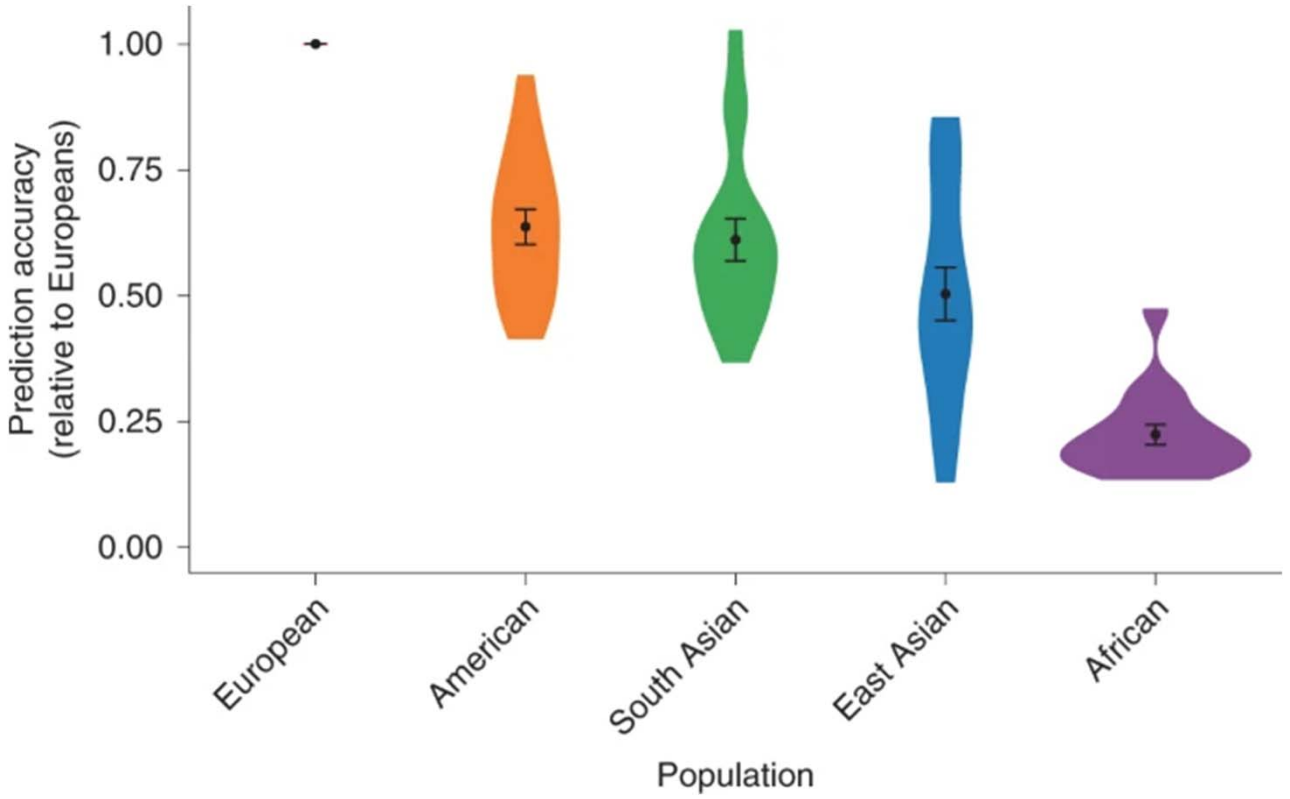
Polygenic Risk Score Accuracy



PRS for Hypertension



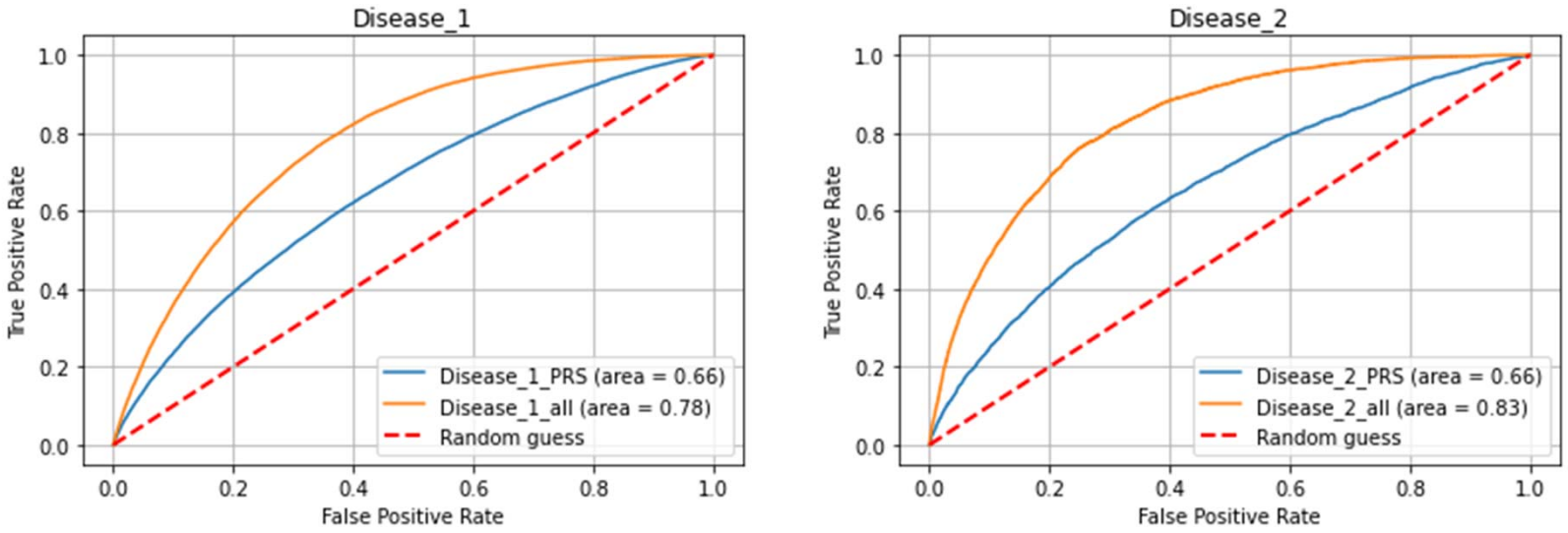
Limited generalizability of PRS across populations



Base data and Target data: Europeans in UK Biobank

Nat Genet 51, 584–591 (2019).

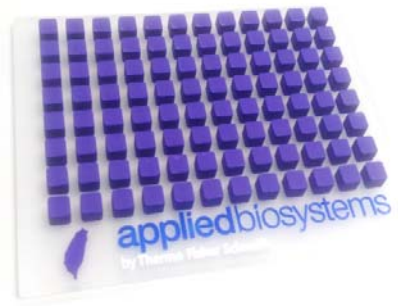
多因子風險預測模式



ROC analysis of prediction accuracy by PRS or
PRS plus Principal Components, Age, Sex

疾病風險預測:
整合風險因子計算器

Assess



Genetic data



Personal monitor data



Environmental data

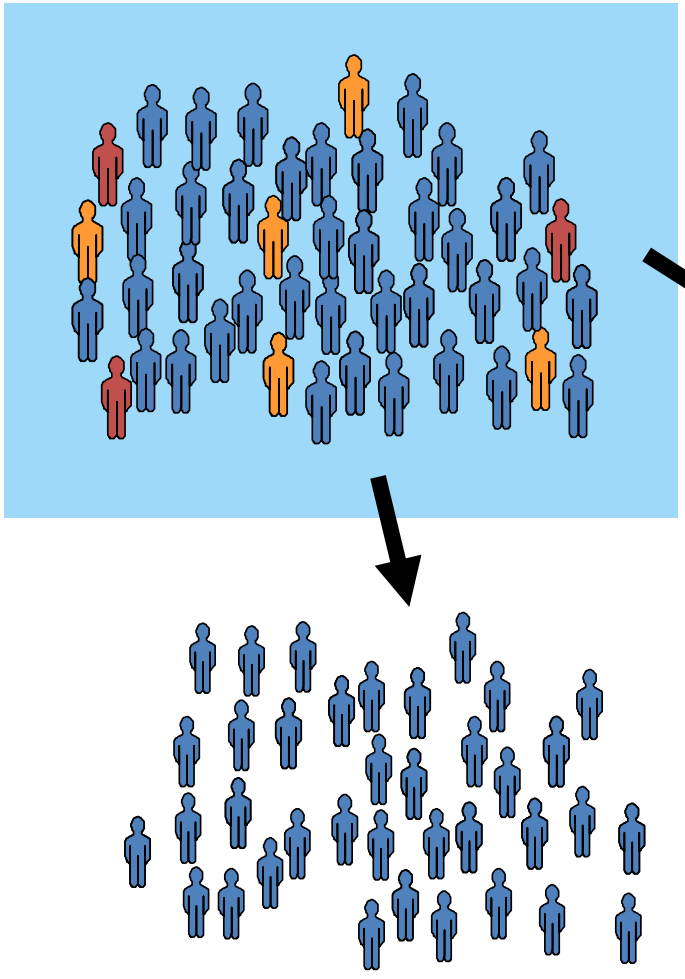


Lifestyle data

Risk Score	Risk Level
0 -5	Low
6 - 10	Moderate
11 -15	High
16 -20	Very High
≥ 21	Extreme

Integrated risk

量身訂製的疾病風險管控



High Risk Patients

Change life style for prevention

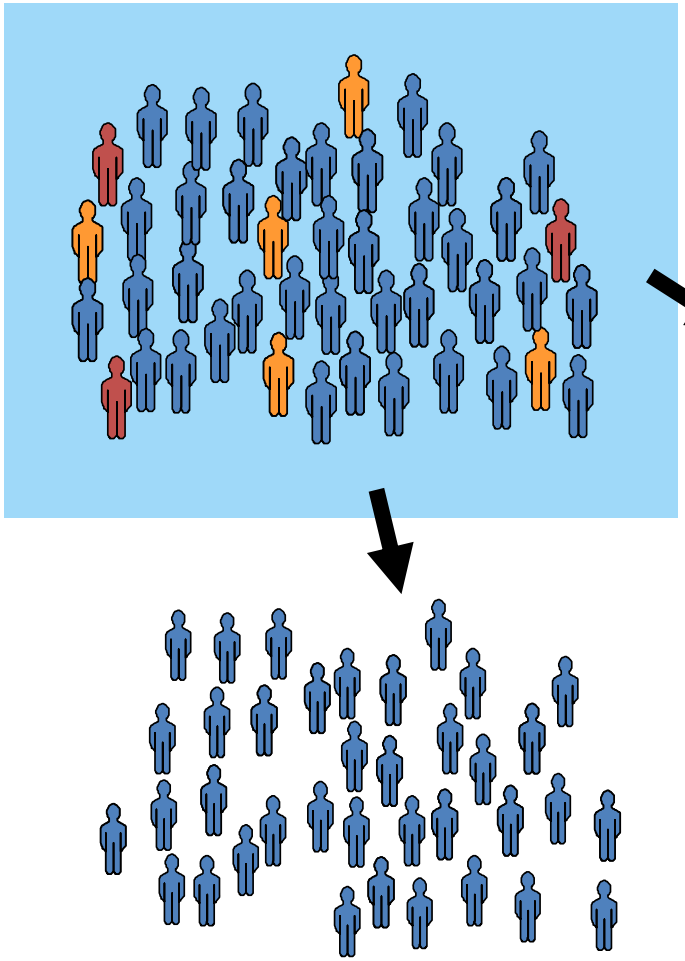
Screening tests early

Treat before symptomatic

Low Risk Patients

Screening tests late

量身訂製的用藥指引



Remove
Non-Responders
and Toxic Responders

Treat
Responders Not
Predisposed to Toxicity