

Genome-wide association studies: implication for public health

Surakameth Mahasirimongkol, MD MSc

Biobank Japan Summary at 2006 November 30

Hyperlipidemia	33,312	Myoma uteri	4,678	Leukemia	1,173
Diabetes	32,506	Pollinosis	4,656	Cervical cancer	1,105
Cataract	13,541	Prostate cancer	4,406	Hepatitis B	1,073
Cerebral infarction	13,280	Glaucoma	4,233	Esophageal cancer	1,057
Stable angina pectoris	12,152	Unstable angina pectoris	3,521	Uterine corpus cancer	901
Arrhythmia	12,135	Lung cancer	3,519	Nephrotic syndrome	885
Myocardial Infarction	11,020	Rheumatoid arthritis	3,477	Ovarian cancer	805
Bronchial asthma	7,084	Atopic dermatitis	2,403	ALS	746
Cardiac failure	5,640	COPD	2,311	Tuberculosis	664
Colorectal cancer	5,520	Cerebral aneurysm	2,291	Keloid	652
Breast cancer	5,433	Liver cirrhosis	1,978	ILD	571
Gastric cancer	5,386	Hyperthyroidism	1,947	Drug-induced hypersensitivity	486
Urinary stone	5,344	Arteriosclerotic obliterans	1,847	Febrile seizures	407
Osteoporosis	5,286	Epilepsy	1,831	CCC	406
Periodontitis	5,019	Liver cancer	1,826	Pancreatic cancer	404
Hepatitis C	4,780	Endometriosis	1,450	Total	235,147

IC asked 192,332 patients
 IC obtained 165,221 patients
 (85.9%)

Total Cases 235,147 cases
 Withdrawn 150 individuals

UK-biobank

- **UK Biobank**

[Accessibility](#) | [Home](#) | [About Us](#) | [Site map](#) | [Health Professionals](#)



Call 0800 0 276 276
Confirm Appointment

Improving the health of future generations

Your feedback is important to us, tell us what you think

Welcome to UK Biobank

Have you received your letter inviting you to participate in UK Biobank? Do you want to do something good today? Not just for yourself, but for our children and our children's children? Signing up to take part in UK Biobank's unique research project is easy to do. If you are happy with your appointment time you can confirm it online now. Just click the button below. A map and more information about your visit to an assessment centre are also a click away. If you want more information or to change your appointment time, that's simple too - just call our Participant Resource Centre on the number above (Monday to Saturday 8am-7pm, calls are recorded). UK Biobank has Quality Management System Standard [ISO 9001:2000 accreditation](#).

If you are looking for more information about the project, just follow the links on the left.

Recruitment Update: 4pm Thursday 6 March 2008
79,083

Confirm your appointment


Assessment centre location maps

What happens at an assessment centre



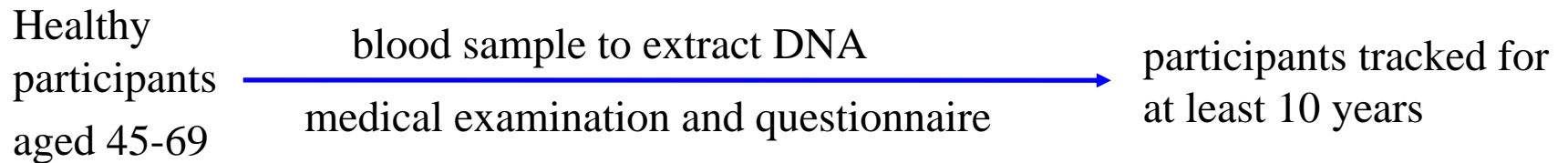
TV presenter Nick Ross supports UK Biobank

welcometrust

- 
- [Confirm appointment](#)
 - [Assessment centre locations](#)
 - [UK Biobank: what is it?](#)
 - [About assessments](#)
 - [Why is it important](#)
 - [Your questions answered](#)
 - [Our supporters](#)
 - [Meet the team](#)
 - [Ethical standards](#)
 - [Latest Developments](#)
 - [Research round up](#)
 - [GPs](#)

Sources of Data : Large-scale studies

- UK Biobank
- A study of genes, environment and health
- www.ukbiobank.ac.uk

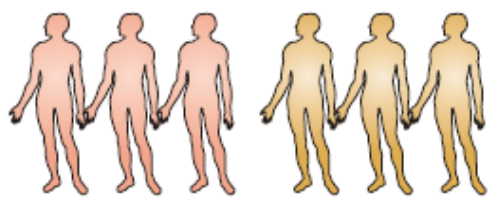


- The Wellcome Trust: £28 million
- The Medical Research Council: £28 million
- UK Department of Health: £5 million
- Information to be made **freely available** to researchers!

What is GWAS?

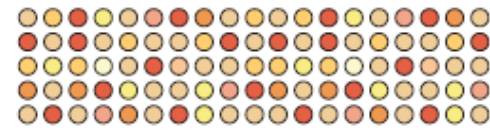
Large cohort of cases and controls ($n > 1,000$)

- Matched for confounding variables, such as race, ethnicity and sex
- Stratified in order to maximize signals



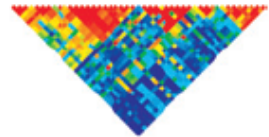
Microarray-based SNP genotyping

- ~1 million random marker SNPs or
- ~25,000 risk-enhancing SNPs (for example, nsSNPs)



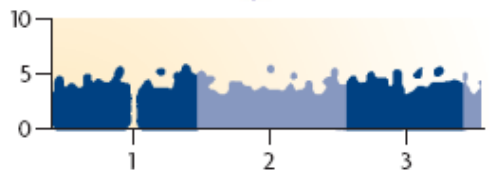
Derivation of haplotypes

- Predicated on International HapMap



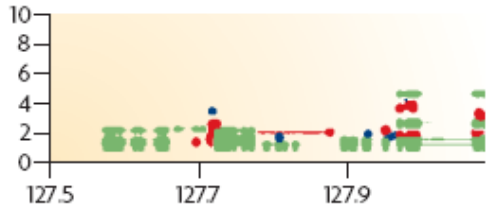
Detection of association signals

- χ^2 or similar test
- Uncorrected $P < 10^{-7}$ or false discovery rate-like correction



Fine mapping of association signal (see FIG. 2)

- Directed genotyping of additional SNPs in region
- Fine mapping of LD in region of association
- Empirical derivation of haplotypes
- Examination of effect of stratification, if available



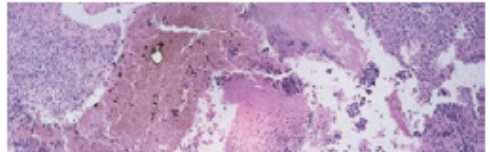
Replication of association

- Large independent cohort of cases and controls ($n > 1,000$)
- Genotyping of nominated candidate SNPs (< 20)
- χ^2 or similar test; replication of initial signal

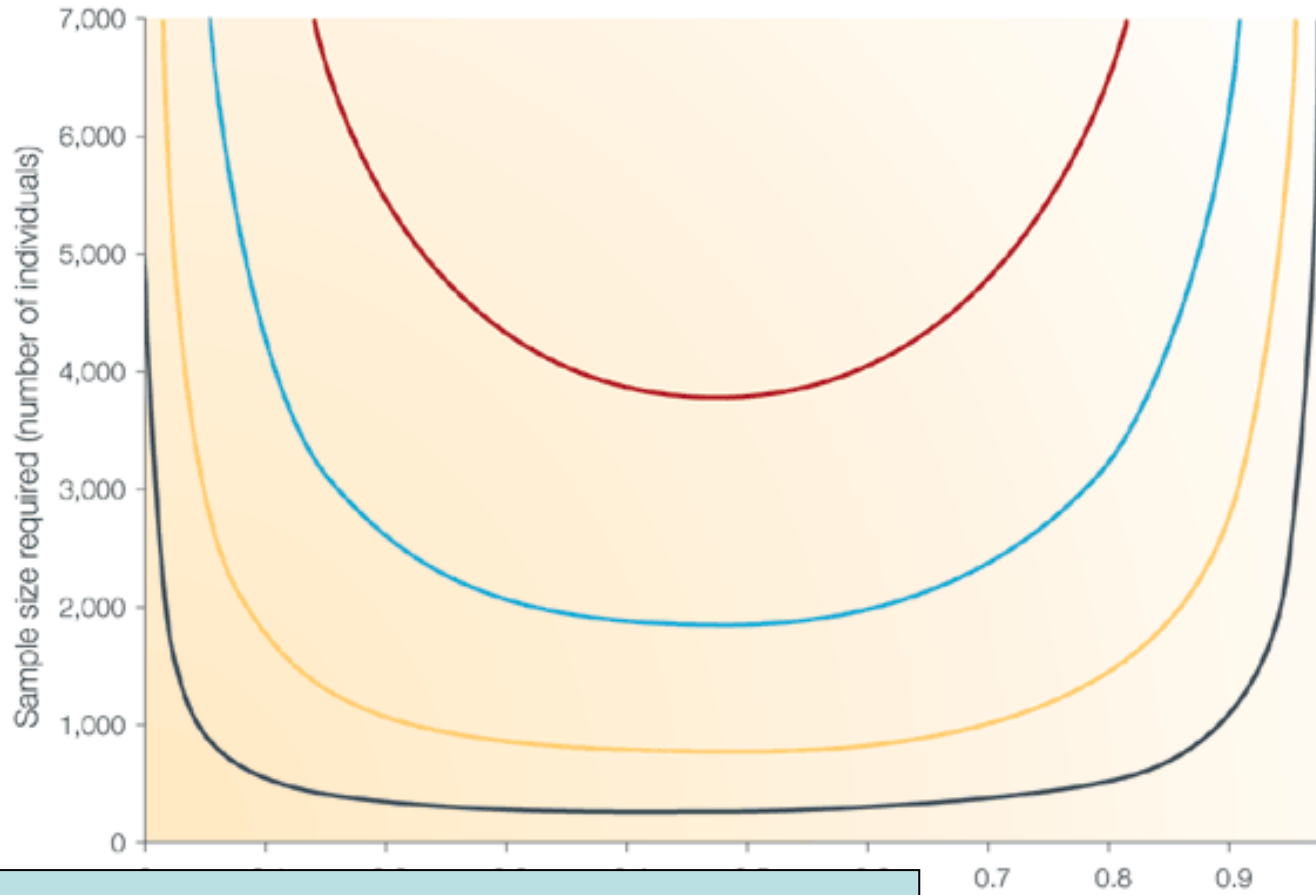
Genotypes	CC	AA	CA	Total
Cases observed	59	27	98	184
Controls observed	60	89	36	185
Total	119	116	134	369

Biological validation of association

- Identification of risk-enhancing variant
- Examination of functional consequence of variant
- Determination of mechanism of risk-enhancement



Samples require vs AF at difference odds ratios



power of 80%, $P < 10^{-6}$

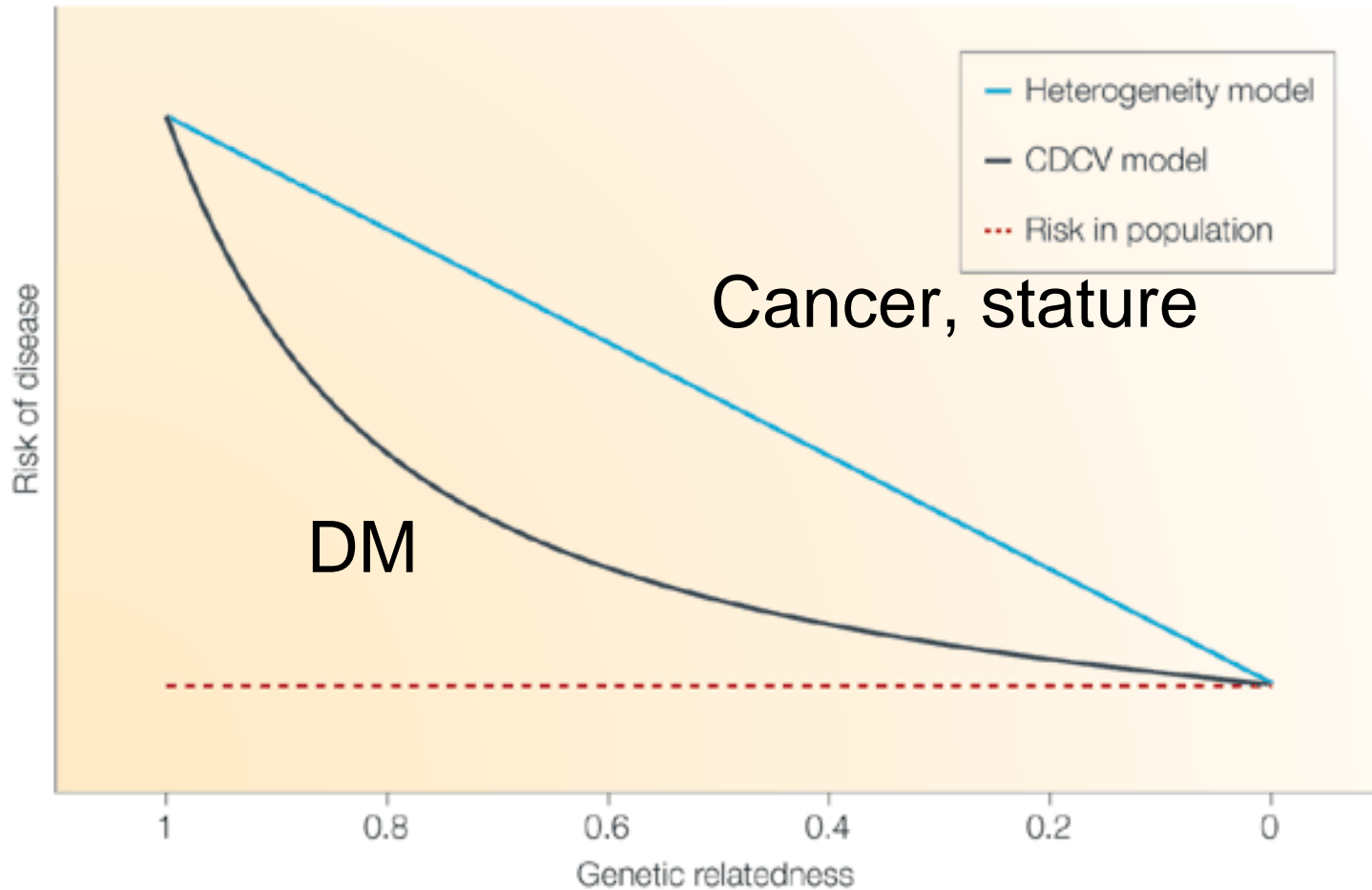
allele

1.2 (red), 1.3 (blue), 1.5 (yellow) and 2 (black)

CDCV vs heterogeneity model


- Common disease common variants
 - Interaction of common alleles cause diseases
- Heterogeneity model
 - Multiple rare alleles cause diseases

Genetic relatedness vs Risk of disease



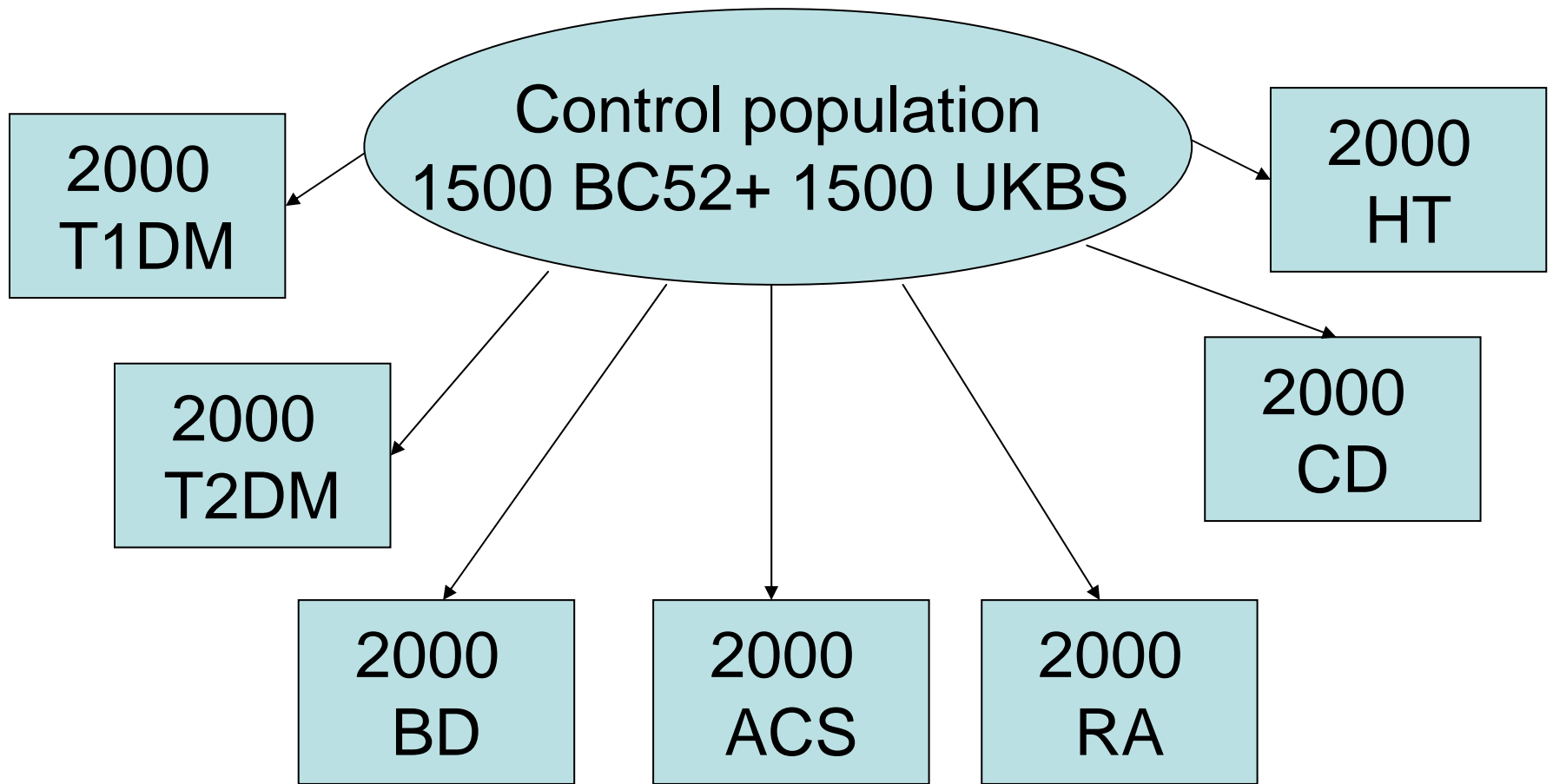
Example from WTCCC study

- กลุ่มโรค bipolar
- โรคเบาหวานชนิดที่ 1
- โรคเบาหวานชนิดที่ 2
- โรคข้ออักเสบรูมาตอยด์
- โรคความดันโลหิตสูง
- โรคหลอดเลือดหัวใจอุดตัน และ
- โรคลำไส้อักเสบ(Chron's disease)ในกลุ่มประชากรอังกฤษ
- โดยทำการศึกษาเปรียบเทียบระหว่างกับ
- กลุ่มประชากรควบคุม 3000 คน
 - กลุ่มประชากรควบคุมจาก กลุ่มผู้บริจาคเลือด 1,500 คน(UKBS) และ
 - กลุ่มประชากรชาวอังกฤษที่เกิดเมื่อปี ค.ศ. 1952 จำนวน 1,500 คน(BC52)



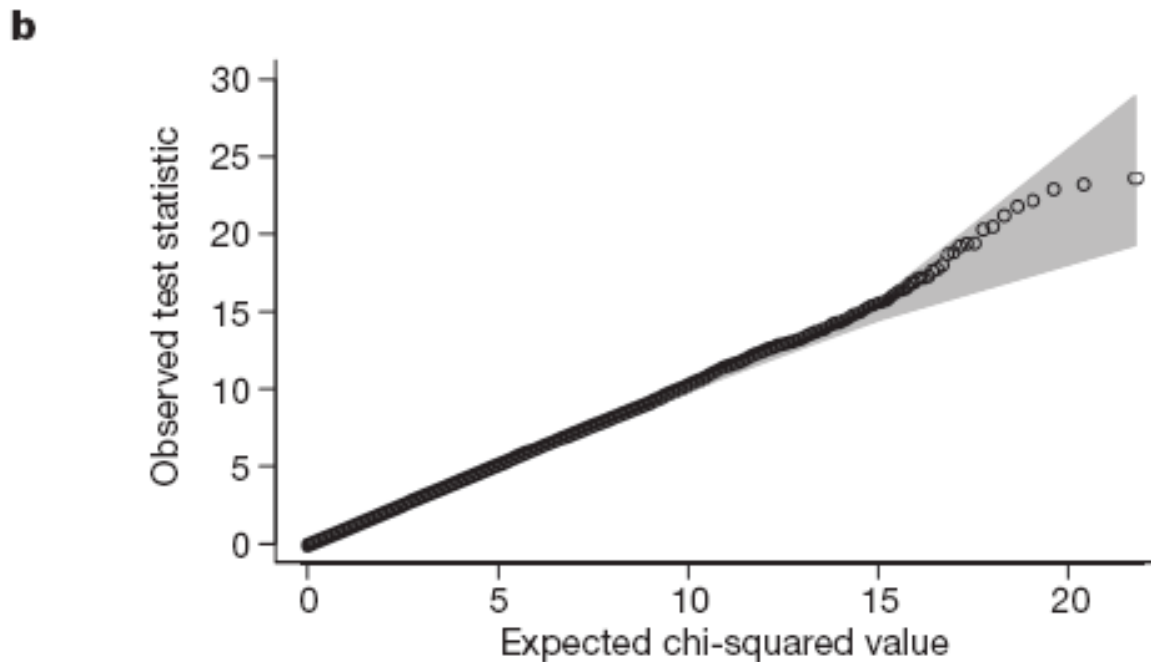
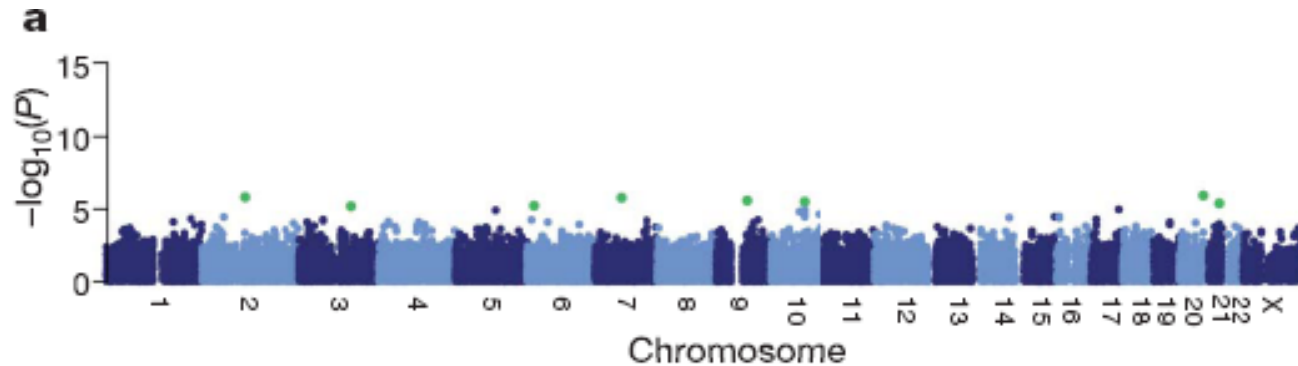
กลุ่มประชากร
ที่เป็นโรคโรค
ละ 2,000 คน

Schematic diagram of WTCCC

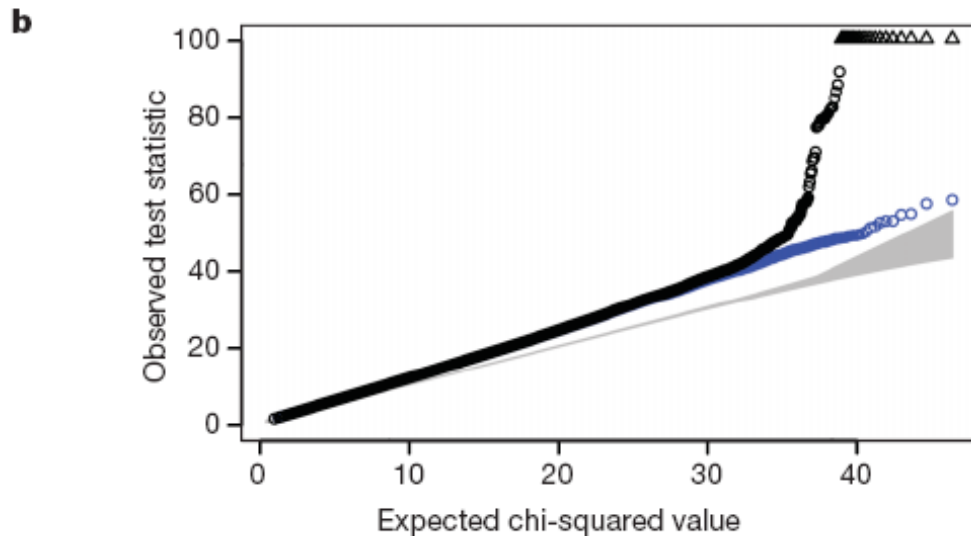
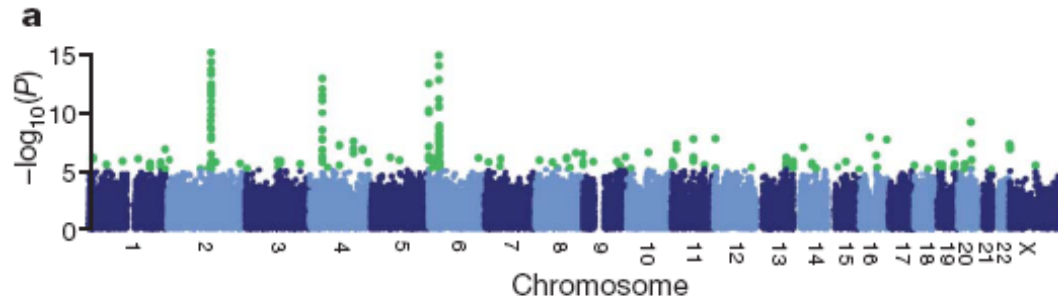


$$17,000 * 20,000 = 340,000,000$$

QQ plot of χ^2 statistics comparing two control populations



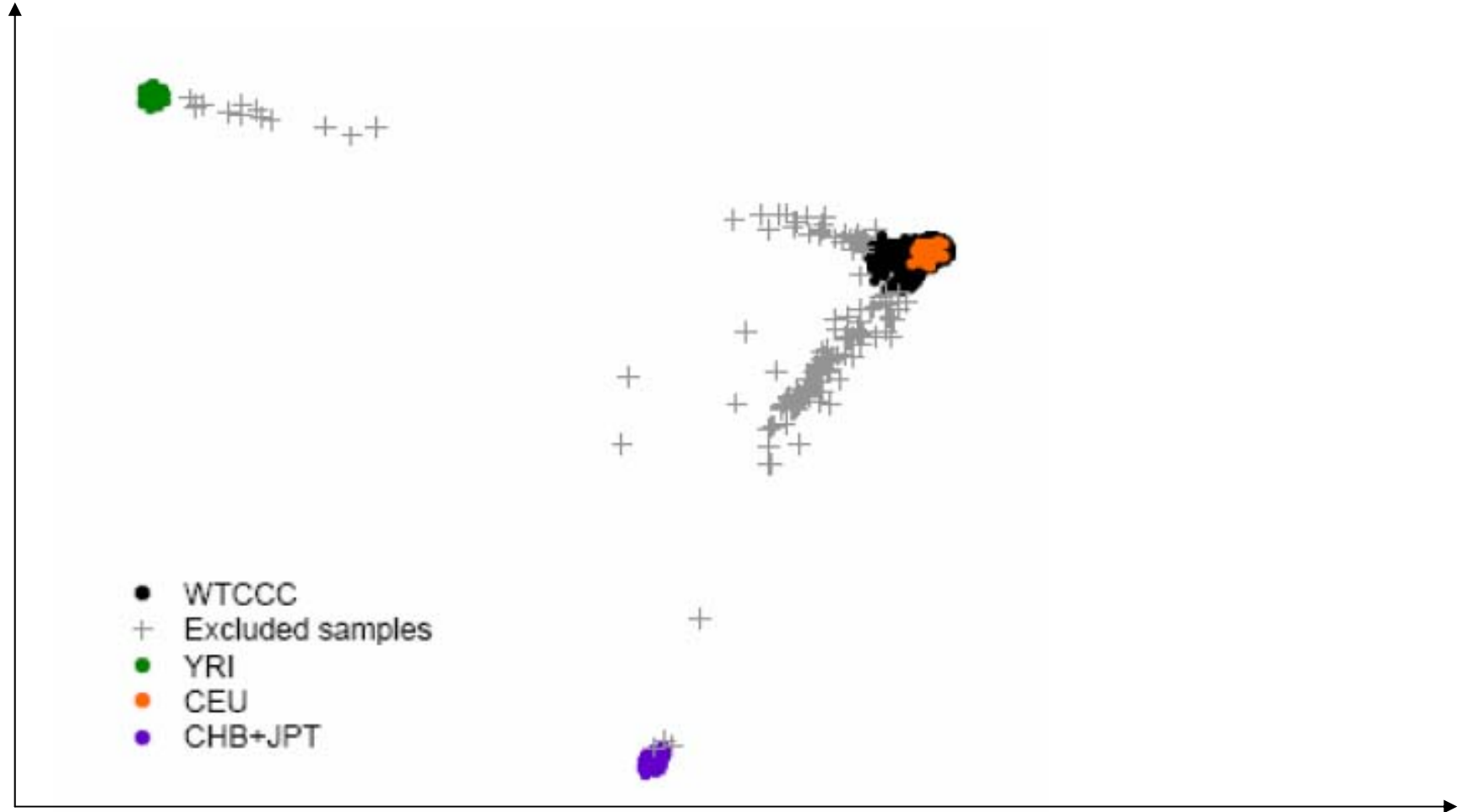
Comparison of populations from 12 different geographic areas



- Northern
- East & West Ridings
- North Midland
- Eastern
- Southeastern
- Southern
- Southwestern
- Wales
- Midlands
- Scotland
- London
- Northwestern



Correction of population stratification



Results of association studies

Table 3 | Regions of the genome showing the strongest association signals

Collection	Chromosome	Region (Mb)	SNP	Trend P value	Genotypic P value	$\log_{10}(\text{BF})$, additive	$\log_{10}(\text{BF})$, general	Risk allele	Minor allele	Heterozygote odds ratio	Homozygote odds ratio	Control MAF	Case MAF
Standard analysis													
BD	16p12	23.3–23.62	rs420259	2.19×10^{-04}	6.29×10^{-08}	1.96	4.79	A	G	2.08 (1.60–2.71)	2.07 (1.6–2.69)	0.282	0.248
CAD	9p21	21.93–22.12	rs1333049	1.79×10^{-14}	1.16×10^{-13}	11.66	11.19	C	C	1.47 (1.27–1.70)	1.9 (1.61–2.24)	0.474	0.554
CD	1p31	67.3–67.48	rs11805303	6.45×10^{-13}	5.85×10^{-12}	10.07	9.41	T	T	1.39 (1.22–1.58)	1.86 (1.54–2.24)	0.317	0.391
CD	2q37	233.92–234	rs10210302	7.10×10^{-14}	5.26×10^{-14}	11.11	11.28	T	C	1.19 (1.01–1.41)	1.85 (1.56–2.21)	0.481	0.402
CD	3p21	49.3–49.87	rs9858542	7.71×10^{-07}	3.58×10^{-08}	4.24	5.22	A	A	1.09 (0.96–1.24)	1.84 (1.49–2.26)	0.282	0.331
CD	5p13	40.32–40.66	rs17234657	2.13×10^{-13}	1.99×10^{-12}	10.41	9.89	G	G	1.54 (1.34–1.76)	2.32 (1.59–3.39)	0.125	0.181
CD	5q33	150.15–150.31	rs1000113	5.10×10^{-08}	3.15×10^{-07}	5.36	5.01	T	T	1.54 (1.31–1.82)	1.92 (0.92–4.00)	0.067	0.098
CD	10q21	64.06–64.31	rs10761659	2.68×10^{-07}	1.75×10^{-06}	4.69	4.13	G	A	1.23 (1.05–1.45)	1.55 (1.3–1.84)	0.461	0.406
CD	10q24	101.26–101.32	rs10883365	1.41×10^{-08}	5.82×10^{-08}	5.91	5.48	G	G	1.2 (1.03–1.39)	1.62 (1.37–1.92)	0.477	0.537
CD	16q12	49.02–49.4	rs17221417	9.36×10^{-12}	3.98×10^{-11}	8.93	8.47	G	G	1.29 (1.13–1.46)	1.92 (1.58–2.34)	0.287	0.356
CD	18p11	12.76–12.91	rs2542151	4.56×10^{-08}	2.03×10^{-07}	5.42	5.00	G	G	1.3 (1.14–1.48)	2.01 (1.46–2.76)	0.163	0.208
RA	1p13	113.54–114.16	rs6679677	4.90×10^{-26}	5.55×10^{-25}	22.36	21.99	A	A	1.98 (1.72–2.27)	3.32 (1.93–5.69)	0.096	0.168
RA	6	MHC	rs6457617*	3.44×10^{-76}	5.18×10^{-75}	74.84	73.18	T	T	2.36 (1.97–2.84)	5.21 (4.31–6.30)	0.489	0.685
T1D	1p13	113.54–114.16	rs6679677	1.17×10^{-26}	5.43×10^{-26}	23.07	22.83	A	A	1.82 (1.59–2.09)	5.19 (3.15–8.55)	0.096	0.169
T1D	6	MHC	rs9272346*	2.42×10^{-134}	5.47×10^{-134}	141.9	142.2	A	G	5.49 (4.83–6.24)	18.52 (27.03–12.69)	0.387	0.150
T1D	12q13	54.64–55.09	rs11171739	1.14×10^{-11}	9.71×10^{-11}	8.89	8.24	C	C	1.34 (1.17–1.54)	1.75 (1.48–2.06)	0.423	0.493
T1D	12q24	109.82–111.49	rs17696736	2.17×10^{-15}	1.51×10^{-14}	12.53	11.88	G	G	1.34 (1.16–1.53)	1.94 (1.65–2.29)	0.424	0.506
T1D	16p13	10.93–11.37	rs12708716	9.24×10^{-08}	4.92×10^{-07}	5.15	4.70	A	G	1.19 (0.97–1.45)	1.55 (1.27–1.89)	0.350	0.297
T2D	6p22	20.63–20.84	rs9465871	1.02×10^{-06}	3.34×10^{-07}	4.15	3.98	C	C	1.18 (1.04–1.34)	2.17 (1.6–2.95)	0.178	0.218
T2D	10q25	114.71–114.81	rs4506565	5.68×10^{-13}	5.05×10^{-12}	10.14	9.43	T	T	1.36 (1.2–1.54)	1.88 (1.56–2.27)	0.324	0.395
T2D	16q12	52.36–52.41	rs9939609	5.24×10^{-08}	1.91×10^{-07}	5.35	5.05	A	A	1.34 (1.17–1.52)	1.55 (1.3–1.84)	0.398	0.453
Multi-locus analysis													
T1D	4q27	123.26–123.92	rs6534347	4.48×10^{-07}	1.83×10^{-06}	5.15	4.69	A	A	1.30 (1.10–1.55)	1.49 (1.25–1.78)	0.351	0.402
T1D	12p13	9.71–9.86	rs3764021	7.19×10^{-05}	5.08×10^{-08}	2.12	4.55	C	T	1.57 (1.38–1.79)	1.48 (1.25–1.75)	0.467	0.426
Sex differentiated analysis													
RA	7q32	130.80–130.84	rs11761231	3.91×10^{-07}	1.37×10^{-06}	-	-	G	A	1.44 (1.19–1.75)	1.64 (1.35–1.99)	0.375	0.327
Combined cases													
RA+T1D	10p15	6.07–6.17	rs2104286	5.92×10^{-08}	2.52×10^{-07}	5.26	4.45	T	C	1.35 (1.11–1.65)	1.62 (1.34–1.97)	0.286	0.245

Implication: Small molecule drug targets

Table 2 | **Loci and variants associated with multiple diseases in GWA studies**

Locus	Variant (rs)	Disease	Refs
PTPN22	6679677	RA	18
		T1DM	18
IL2RA	2104286	RA	18
		T1DM	18
PTPN2	2542151	T1DM	51
		CD	18,46
TCF2	4430796	T2DM	53
		PC	53
FTO	9939609	T2DM	18
		Obesity	71
APOE	4420638	Triglyceride level	54
		Alzheimer's disease	90
8q24	6983267	PC	79, 81, 113
		CC	111–113
IL23R	11209026	CD	43
		Psoriasis	21

Above variants are associated $P < 5 \times 10^{-7}$. CC, colorectal cancer; CD, Crohn's disease; PC, prostate cancer; RA, rheumatoid arthritis; T1DM, type 1 diabetes mellitus; T2DM, type 2 diabetes mellitus.

Some GWAS examples

Genome-wide association study shows *BCL11A* associated with persistent fetal hemoglobin and amelioration of the phenotype of β -thalassemia

Manuela Uda^a, Renzo Galanelli^b, Serena Sanna^a, Guillaume Lettre^{2b}, Vijay G. Sankaran^{2f}, Weimin Chen¹, Gianluca Usala^a, Fabio Busonero^a, Andrea Maschio^a, Giuseppe Albal^a, Maria Grazia Piras^a, Nataschia Sestu^a, Sandra Lai^a, Mariano Del^a, Antonella Mulas^a, Laura Crisponi^a, Silvia Naitza^a, Isadora Asunis^a, Manila Delana^a, Ramalah Nagaraja^{2g}, Lucia Perseu^a, Stefania Satta¹, Maria Dolores Cipollina¹, Carla Sollaino¹, Paolo Mol¹, Joel N. Hirschhorn^{2b}, Stuart H. Orkin^{2f, h}, Gonçalo R. Abecasis¹, David Schlessinger^{2a}, and Antonio Cao^{2f, h}

^aIstituto di Neurogenetica e Neurofarmacologia, Consiglio Nazionale delle Ricerche, Cittadella Universitaria di Monserrato, Monserrato, Cagliari, Italy 09042; ^bClinica Pediatrica, Ospedale Microcitemico, Via Jenner s/n 09121 Cagliari, Dipartimento di Scienze Biomediche e Biotecnologie, Università degli Studi di Cagliari, 09121 Cagliari, Italy; ^cChildren's Hospital of Boston, 300 Longwood Avenue, Boston, MA 02115; ^dBroad Institute of Harvard and Massachusetts Institute of Technology, 7 Cambridge Center, Cambridge, MA 02142; ^eDepartment of Pediatric Oncology, Dana-Farber Cancer Institute, Boston, MA 02115; ^fCenter for Statistical Genetics, Department of Biostatistics, University of Michigan, 1420 Washington Heights, Ann Arbor, MI 48109; and ^gGerontology Research Center, National Institute on Aging, 5600 Nathan Shock Drive, Baltimore, MD 21224

Contributed by Stuart H. Orkin, December 12, 2007 (sent for review November 14, 2007)

Research

Open Access

Genome-wide association with select biomarker traits in the Framingham Heart Study

Emelia J Benjamin^{*1,2,3,4}, Josée Dupuis^{1,4}, Martin G Larson^{1,5}, Kathryn L Lunetta^{1,4}, Sarah L Booth⁶, Diddahally R Govindaraju^{1,2}, Sekar Kathiresan^{7,8}, John F Keaney Jr^{2,3}, Michelle J Keyes^{1,5}, Jing-Ping Lin¹⁰, James B Meigs⁹, Sander J Robins^{1,2}, Jian Rong^{1,4}, Renate Schnabel^{1,2}, Joseph A Vita^{2,3}, Thomas J Wang⁸, Peter WF Wilson¹¹, Philip A Wolf^{1,2} and Ramachandran S Vasan^{1,2,3}

Genome-Wide Association for Methamphetamine Dependence

Convergent Results From 2 Samples

George R. Uhl, MD, PhD; Tomas Drgon, PhD; Qing-Rong Liu, PhD; Catherine Johnson, MSc; Donna Walther, MSc; Tokutaro Komiyama, MD; Mutsuo Harano, MD; Yoshimoto Sekine, MD, PhD; Toshiya Inada, MD; Norio Ozaki, MD, PhD; Masaomi Iyo, MD, PhD; Nakao Iwata, MD, PhD; Mitsuhiro Yamada, MD; Ichiro Sora, MD, PhD; Chih-Ken Chen, MD, PhD; Hsing-Cheng Liu, MD, PhD; Hiroshi Ujike; Shih-Ku Lin, MD

Private services



23andMe genetics just got personal.

[sign in](#) | [claim codes](#) | [help](#)
[Visit our blog!](#)

[order now](#)

[try a demo](#)

[replay animation]

1866: Gregor Mendel discovers the laws of inheritance.

200,000 years ago: *Homo sapiens* walks the Earth.

2003: The Human Genome Project maps a single person's genome.

2007: 23andMe introduces the first Personal Genome Service.

Unlock the secrets of your own DNA. Today.

175,000 years ago: The mother of all present-day humans is born in Africa.

1953: Watson and Crick uncover the double-helix structure of DNA.

Welcome to 23andMe, a web-based service that helps you read and understand your DNA. After providing a saliva sample using an at-home kit, you can use our interactive tools to shed new light on your distant ancestors, your close family and most of all, yourself.

[✉ Sign up for our Email Newsletter](#)

news

What's new at 23andMe

Feb 21, 2008: Personal Genome Service enhanced with more [Gene Journal](#) content, a new [Paternal Ancestry](#) feature, and free [demo accounts](#).

ETHAN
HAWKE

UMA
THURMAN

JUDE
LAW

G A T T A C A

There is no gene for the human spirit



JenzScans

Implication: Disease prevention

- Increased cost-effectiveness of prevention strategy by gene-based risks stratifying
 - *Streptococcus pneumoniae* vaccination
 - Latent TB treatment
 - HPV vaccination
- Increased surveillance in high genetic risks

Implication: Personalized medicine

- US FDA recommended to test HLA-B for
 - Carbamazepine in Asian (HLA-B*1502)
 - Abacavir in Caucasian(HLA-B*5707)
 - TPMT genetic test for 5-FU
 - CYP2*9 and VKORC for Warfarin dosing
- Selection of effective drug
- Avoidance of side effect

Pharmacogenomics case study

HLA-B*5701 Screening for Hypersensitivity to Abacavir

The NEW ENGLAND JOURNAL of MEDICINE

ORIGINAL ARTICLE

HLA-B*5701 Screening for Hypersensitivity to Abacavir

Simon Mallal, M.B., B.S., Elizabeth Phillips, M.D., Giampiero Carosi, M.D., Jean-Michel Molina, M.D., Cassy Workman, M.B., B.S., Janez Tomažič, M.D., Eva Jägel-Guedes, M.D., Sorin Rugina, M.D., Oleg Kozyrev, M.D., Juan Flores Cid, M.D., Phillip Hay, M.B., B.S., David Nolan, M.B., B.S., Sara Hughes, M.Sc., Arlene Hughes, Ph.D., Susanna Ryan, Ph.D., Nicholas Fitch, Ph.D., Daren Thorborn, Ph.D., and Alastair Benbow, M.B., B.S., for the PREDICT-1 Study Team*

NEJM; Feb 2008

HLA-B*5701 Screening for Hypersensitivity to Abacavir

- Double-blind, prospective, randomized
- 1956 patients from 19 countries
- Naïve to abacavir
- Randomly assigned to
 - Prospective screening of HLA-B*5701 or
 - Standard-of-care approach without HLA-B testing (control group)
- Observe for 6 weeks
- Epicutaneous patch testing with abacavir in HSR

Statistical analysis

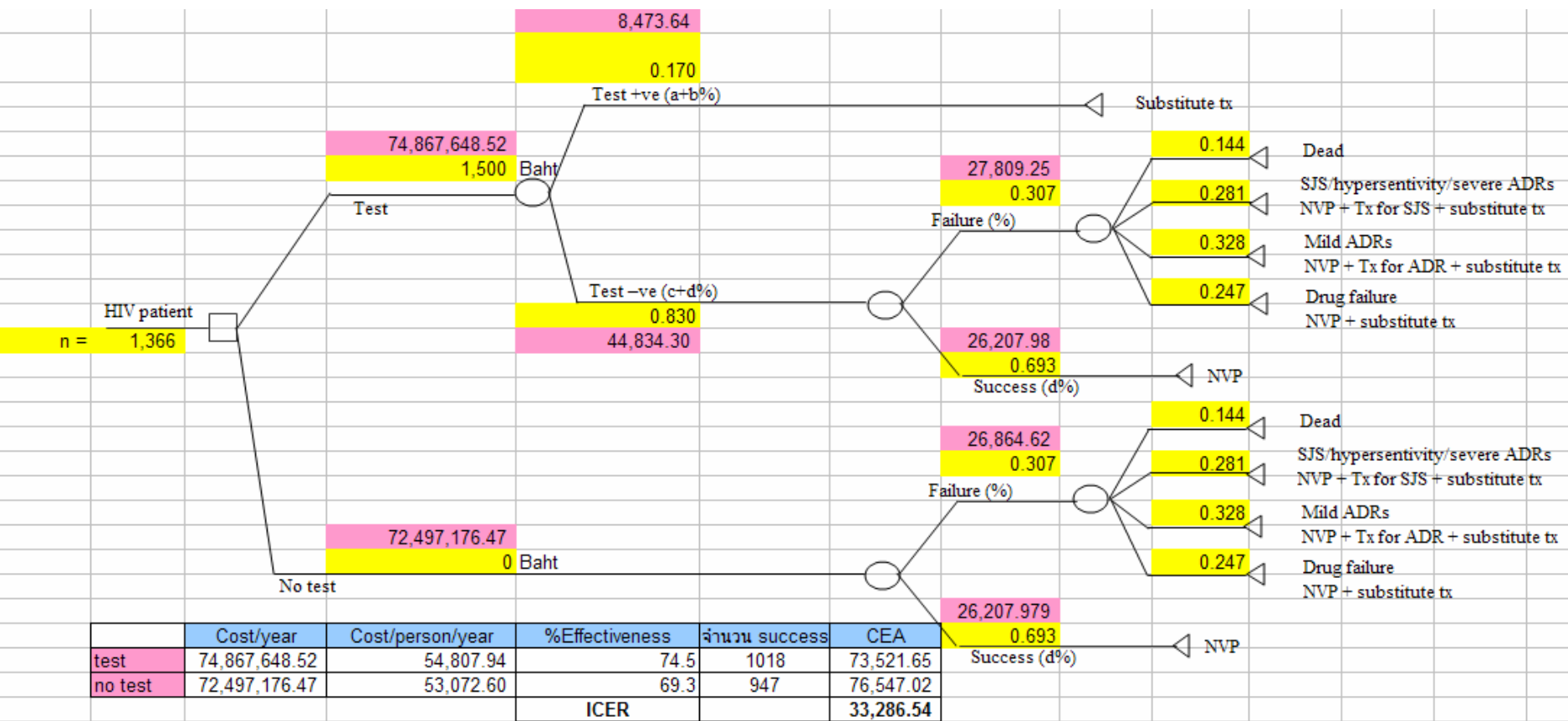
- Two primary endpoint
 - Clinical hypersensitivity
 - Lab-confirmed hypersensitivity
- Logistic regression (exact)
- Sensitivity, specificity of HLA-B calculated in control group

Group	Positive for HLA-B*5701	Negative for HLA-B*5701	Total	Performance Characteristics for Hypersensitivity Reaction
	<i>number of patients</i>			<i>percent (95% CI)</i>
Clinically diagnosed hypersensitivity reaction				
population that could be evaluated				
with hypersensitivity reaction	30	36	66	Sensitivity: 45.5 (33.1–57.9)
without hypersensitivity reaction	19	762	781	Specificity: 97.6 (96.2–98.9) PPV: 61.2 (46.2–74.8) NPV: 95.5 (93.8–96.8)
Subgroup				
with hypersensitivity reaction	29	32	61	Sensitivity: 47.5 (34.6–60.4)
without hypersensitivity reaction	19	638	657	Specificity: 97.1 (95.5–98.6) PPV: 60.4 (45.3–74.2) NPV: 95.2 (93.3–96.7)
Immunologically confirmed hypersensitivity reaction				
population that could be evaluated				
with hypersensitivity reaction	23	0	23	Sensitivity: 100 (85.2–100)
without hypersensitivity reaction	25	794	819	Specificity: 96.9 (95.5–98.2) PPV: 47.9 (33.3–62.8) NPV: 100 (99.5–100)
Subgroup				
with hypersensitivity reaction	22	0	22	Sensitivity: 100 (84.6–100)
without hypersensitivity reaction	25	666	691	Specificity: 96.4 (94.7–98.0) PPV: 46.8 (32.1–61.9) NPV: 100 (99.4–100)

Identification of genetic risks of NVP-induced rash

No.	Marker or Combination			case (n=131)		control (n=177)		p-value	OR	sensitivity %	specificity %	PPV %	NPV %
				risk	non-risk	risk	non-risk						
1	<i>B*3505</i>			22	109	2	175	2.51E-07	17.66	17	99	79	83
2	<i>B*3505</i>	<i>SNP7</i>		33	98	6	171	1.21E-08	9.60	25	97	65	84
3	<i>B*3505</i>	<i>A*2407</i>		28	103	9	168	2.51E-05	5.07	21	95	51	83
4	<i>SNP7</i>			15	116	5	172	3.91E-03	4.45	11	97	50	81
5	<i>B*3505</i>	<i>A*2407</i>	<i>SNP7</i>	38	93	13	164	7.29E-07	5.15	29	93	50	84
6	<i>B*3505</i>	<i>SNP3</i>		32	99	11	166	8.44E-06	4.88	24	94	50	83
7	<i>B*3505</i>	<i>SNP3</i>	<i>SNP7</i>	40	91	14	163	3.08E-07	5.12	31	92	49	84
8	<i>SNP3</i>			31	100	11	166	1.59E-05	4.68	24	94	49	83
9	<i>SNP3</i>	<i>SNP7</i>		39	92	14	163	6.19E-07	4.94	30	92	48	84
10	<i>B*3505</i>	<i>SNP4/5</i>	<i>SNP7</i>	38	93	15	162	3.08E-06	4.41	29	92	46	84

Cost-effectiveness analysis

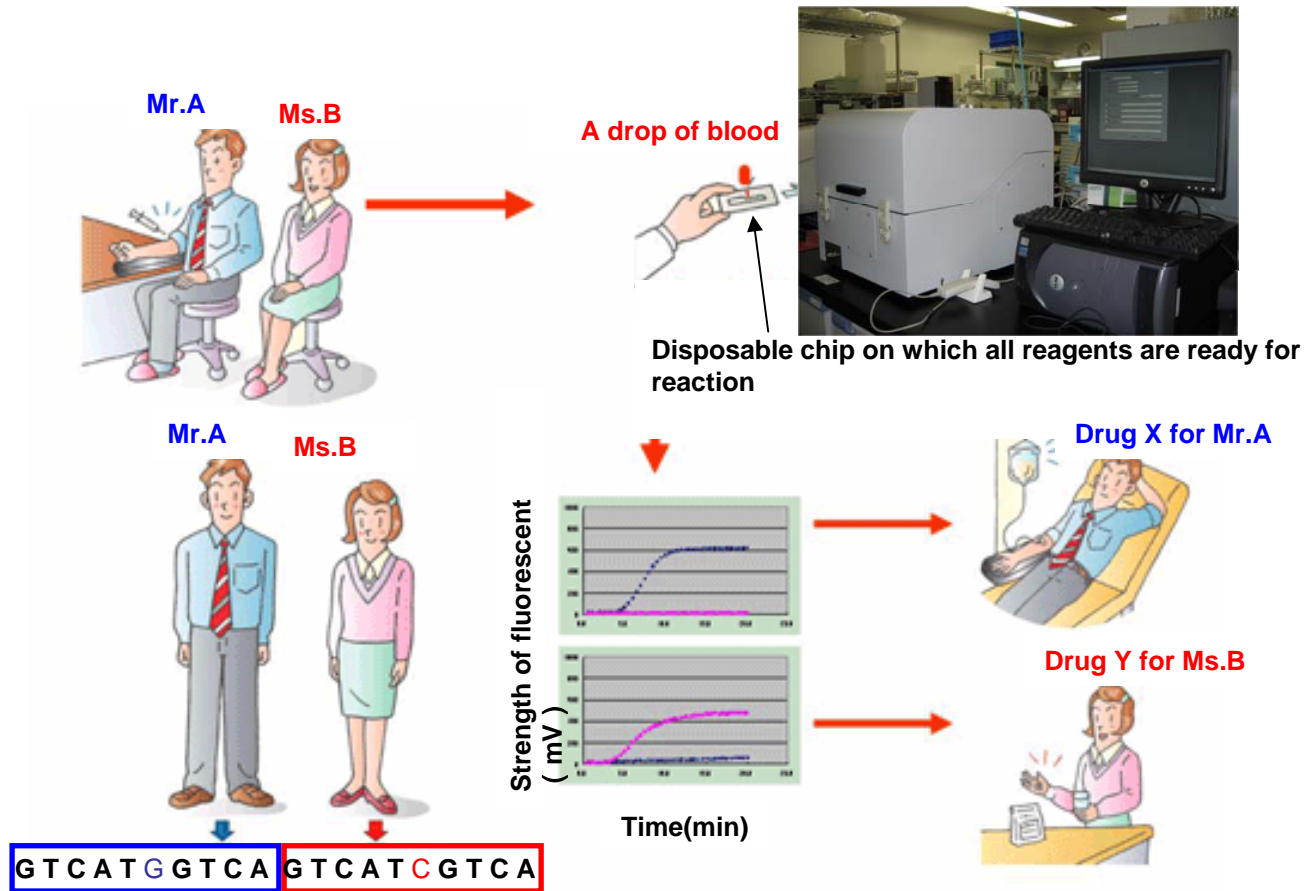


Controlled trials for Screening for NVP-induced rash

- Double-blind controlled trial
- Screening genetic tests(HLA-B+additional markers)
- Primary endpoints
 - Expected 25% to 50% rashes reduction
 - Epicutaneous test
- Control groups confirmed sensitivity and specificity

Development of a fully-automated SNP genotyping System

(Collaboration with TOPPAN and SHIMADZU)



This system requires a drop of blood for genotyping and provides genotype information within 80 minutes.