Identification of genetic basis for Type 2 Diabetes in East Asian populations

Yoon Shin Cho

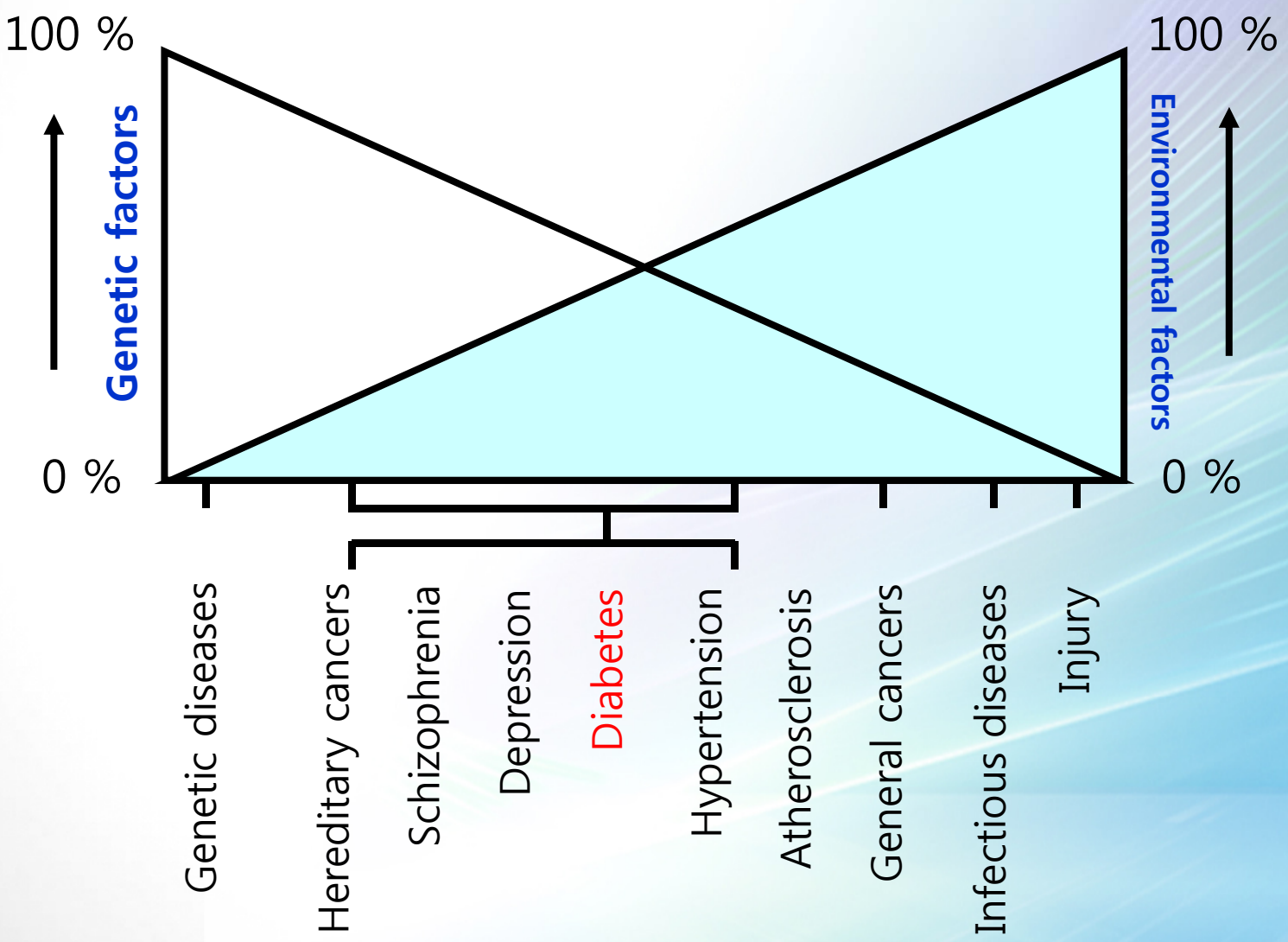
Hallym University

Department of Biomedical Science

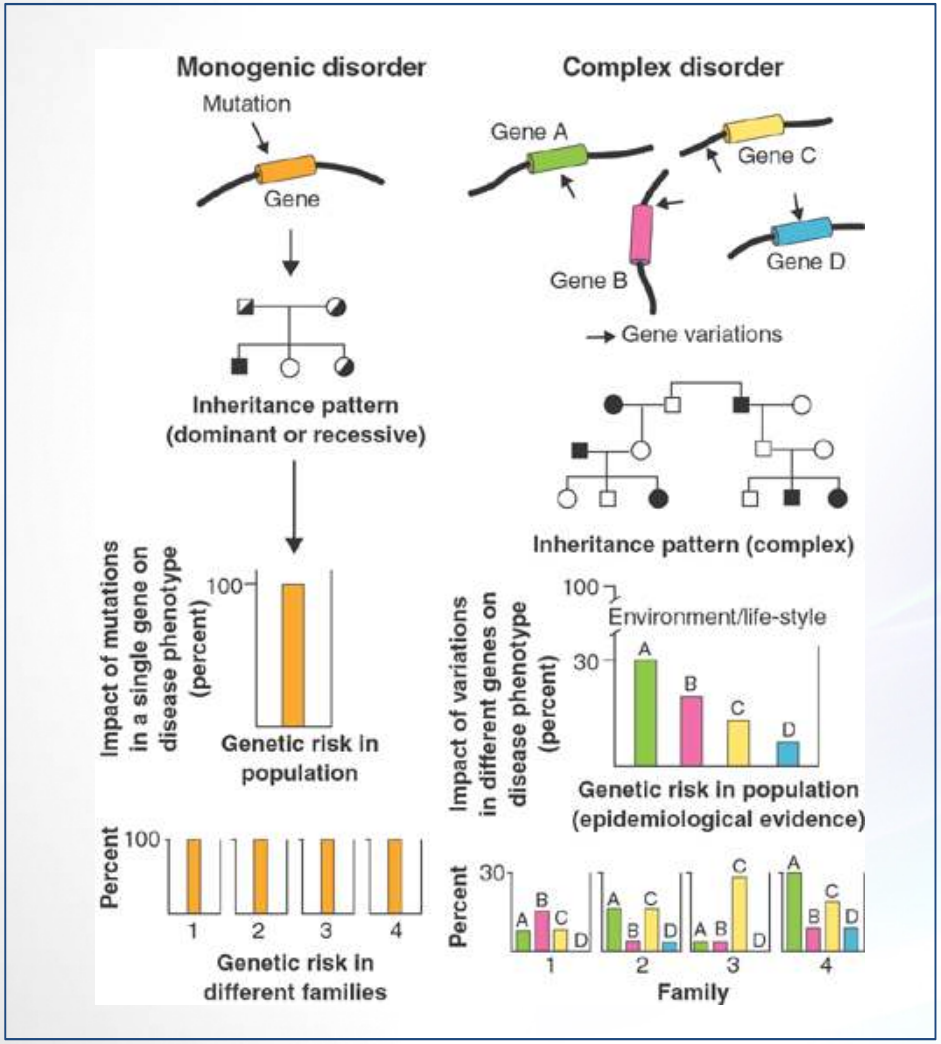
HALLYM UNIVERSITY

Risk factors for complex diseases and traits

(correlation between genetic and environmental factors)

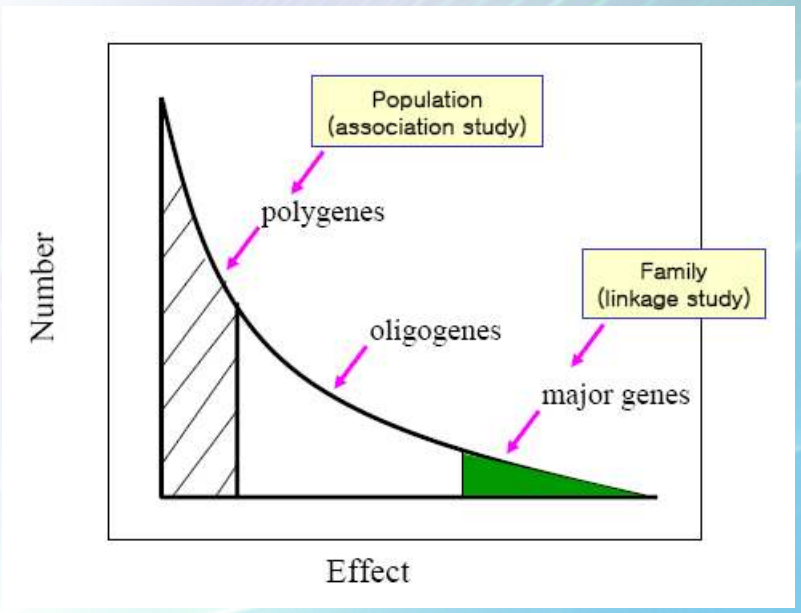


Ways to identify genetic factors for diseases



(From Peltonen and McKusick 2001)

Disease Genomics :
Study of human diseases based on genetic information in the human genome



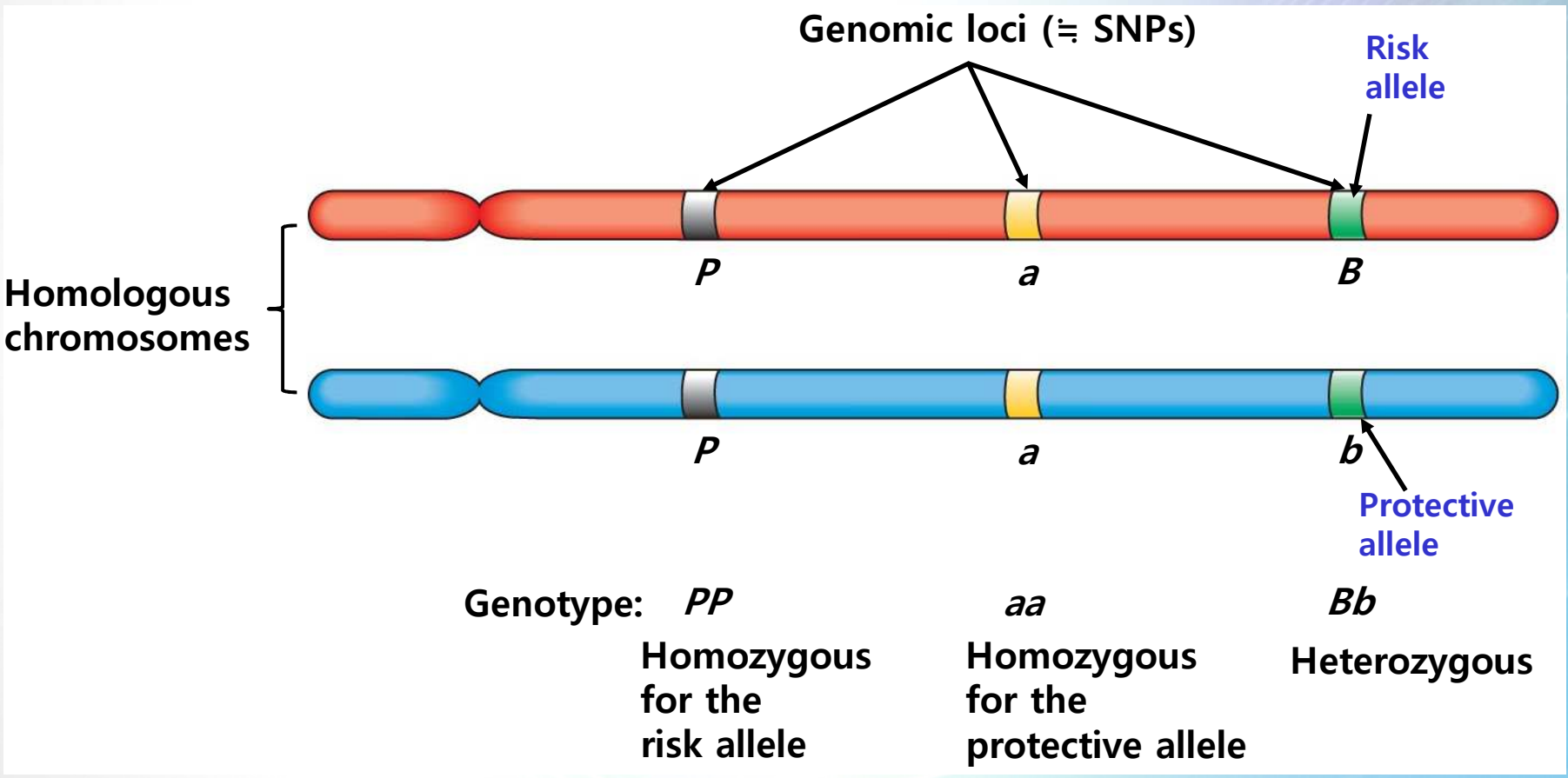
Number of disease determinants by effect

DNA genetic variation

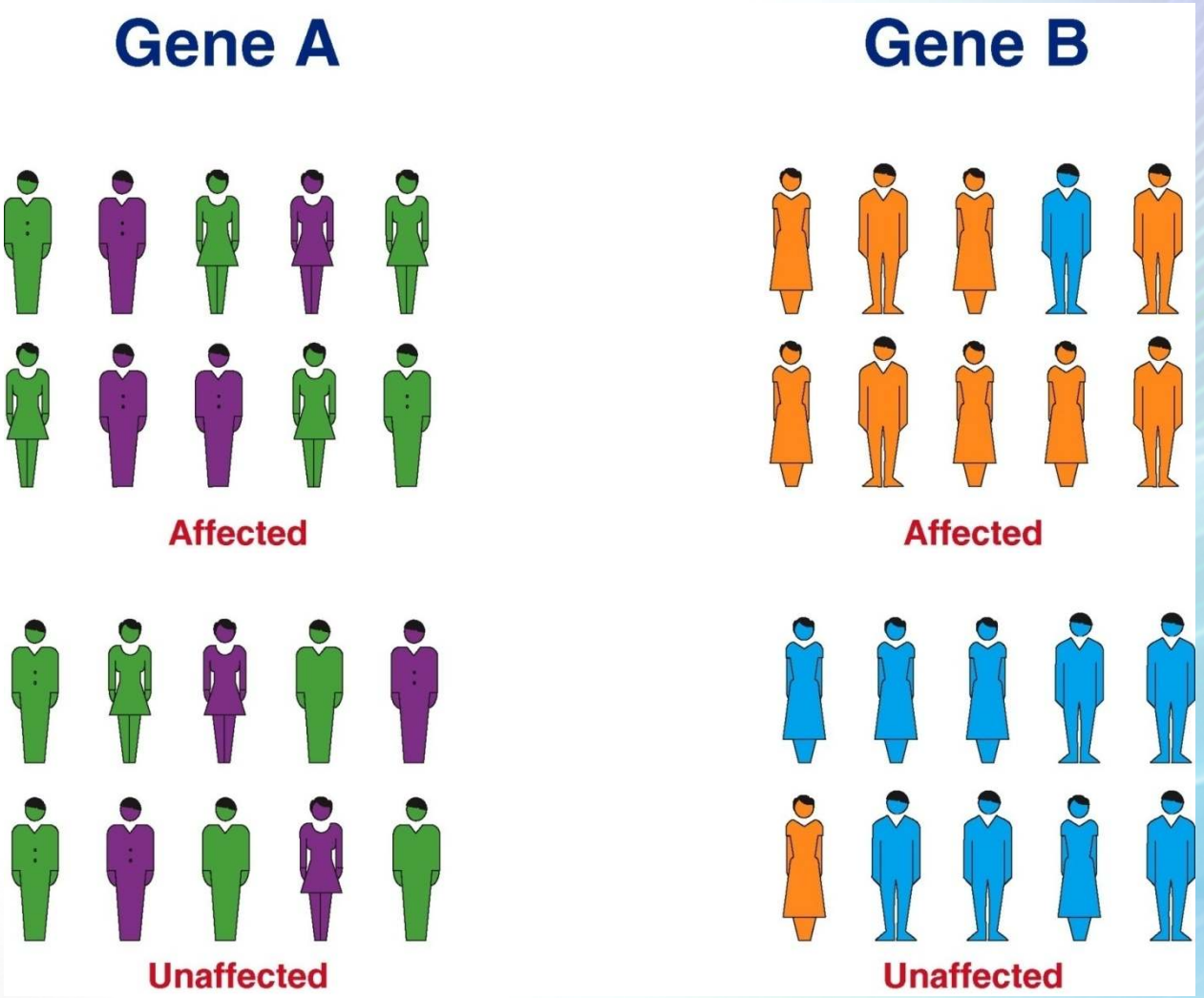
- Insertions/Deletions, Translocation
- Copy number variation (CNV)
- **Single Nucleotide Polymorphism (SNP)**
  - 0.1% difference btw individuals
  - No. of SNP = ~23,650,000

    (based on dbSNP build 131, 2011.5)

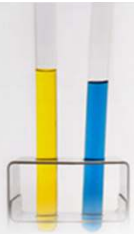
Genetic variations in homologous chromosomes

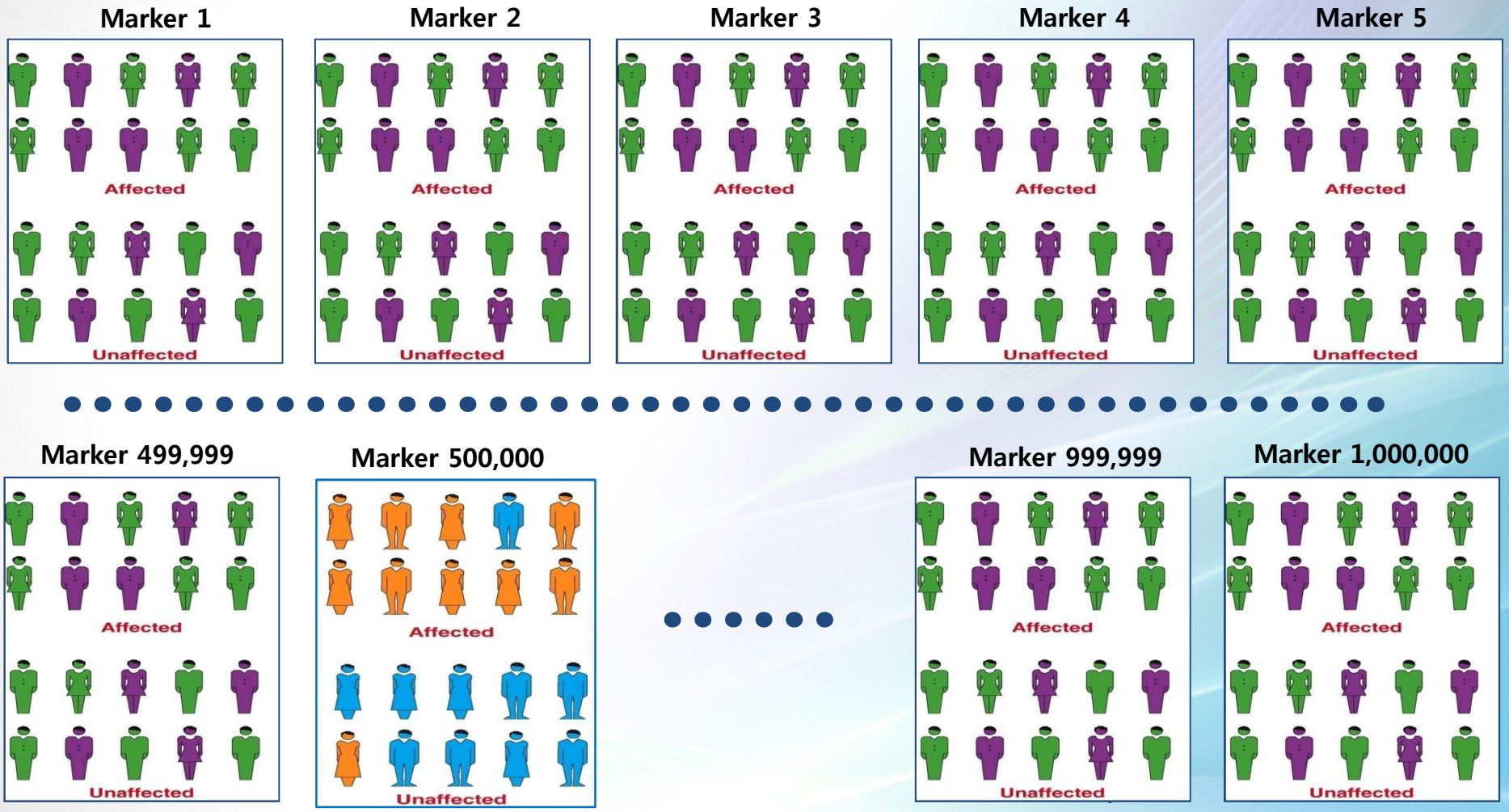


Testing for disease-marker (SNP) association



Genome Wide Association Study (GWAS)

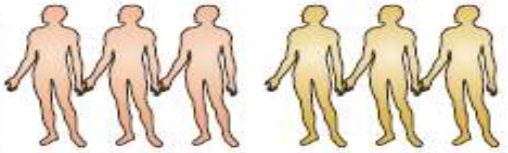




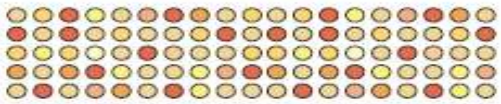
Association analysis between genetic markers (SNPs) across the entire genome and phenotypes in a large number of samples

Overview of the general design and workflow of GWAS

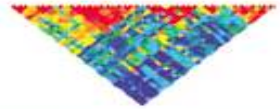
Large cohort of cases and controls (*n* >1,000)
- Matched for confounding variables, such as race, ethnicity and sex
- Stratified in order to maximize signals



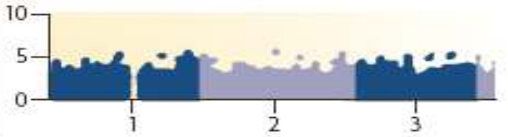
Microarray-based SNP genotyping
- ~1 million random marker SNPs or ~25,000 risk-enhancing SNPs (for example, nsSNPs)



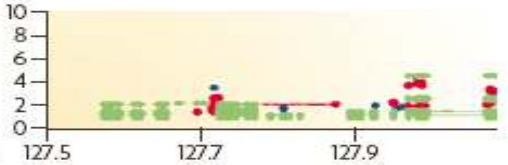
Derivation of haplotypes
- Predicated on International HapMap



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

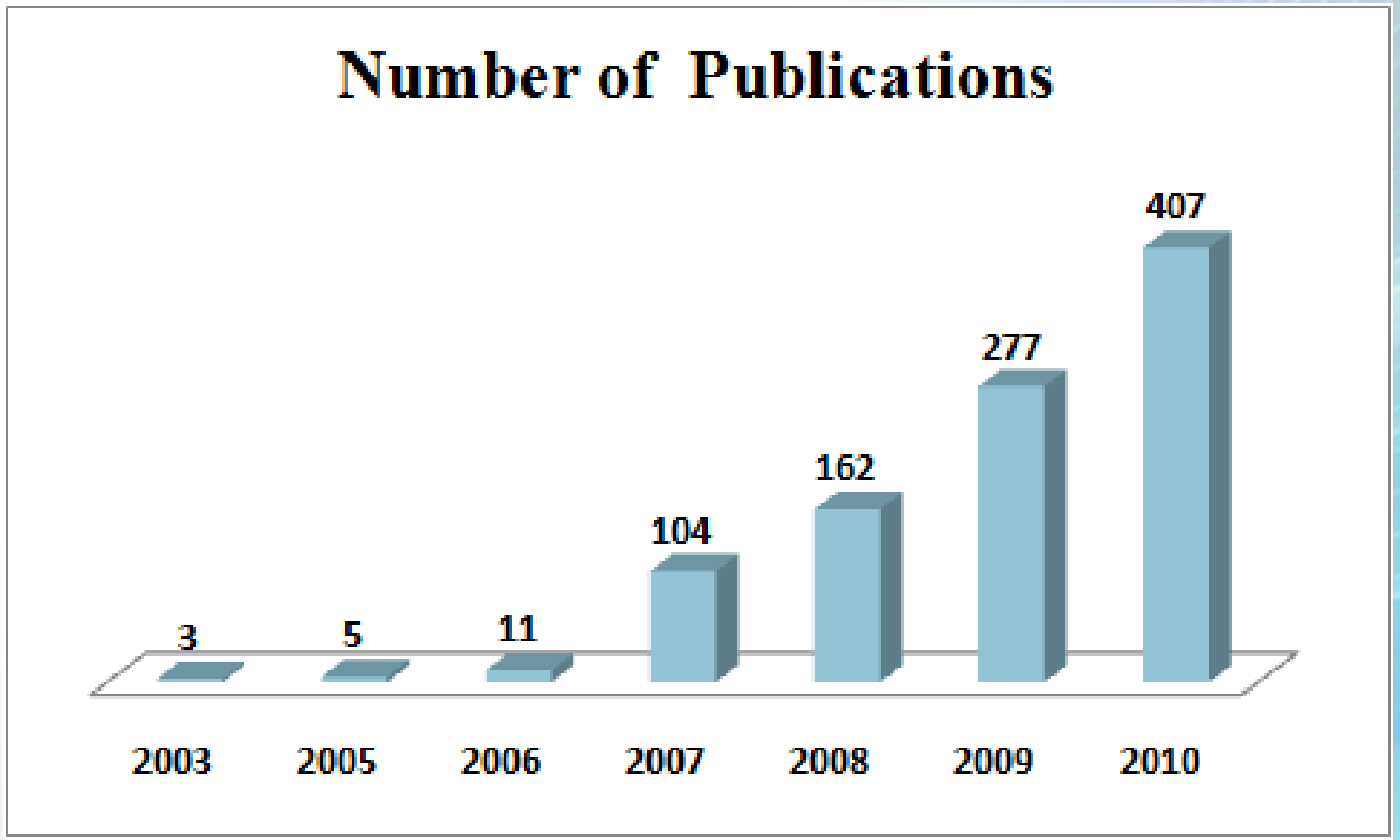
Genotypes	CC	AA	CA	Total
Cases observed	59	27	98	184
Controls observed	60	89	36	185
Total	119	116	134	369

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

(Kingsmore et al., Nature Reviews Drug Discovery. 2008)

GWAS publications (by Dec 30, 2010)

Genome Wide Association Study (GWAS)



Modified data from HuGE Navigator

Advances for GWA analysis

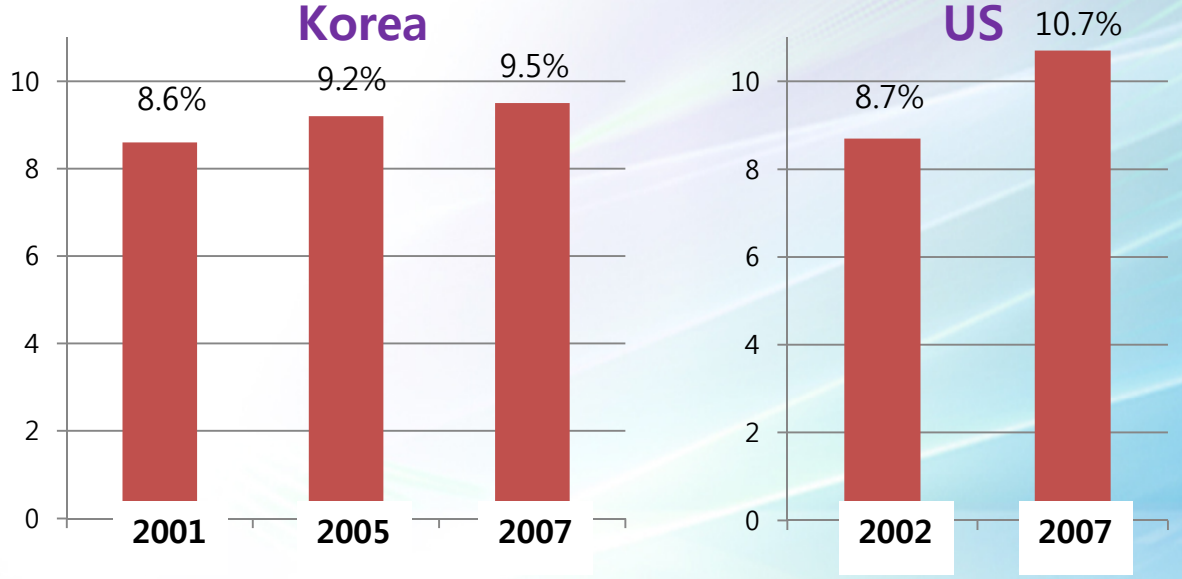
Advances in genotyping technology

Better understanding of patterns of human sequence variation

Sample collections of adequate size

Why a Genetic Study of T2D?

- **T2D strongly familial**
- **T2D huge, growing public health problem worldwide**
  - Risk factors : renal failure, retinopathy, peripheral neuropathy, cardiovascular diseases
  - Death by diabetes : 22.9/100,000 in Korea in 2007 (5th ranked)
  - Prevalence rate growing rapidly

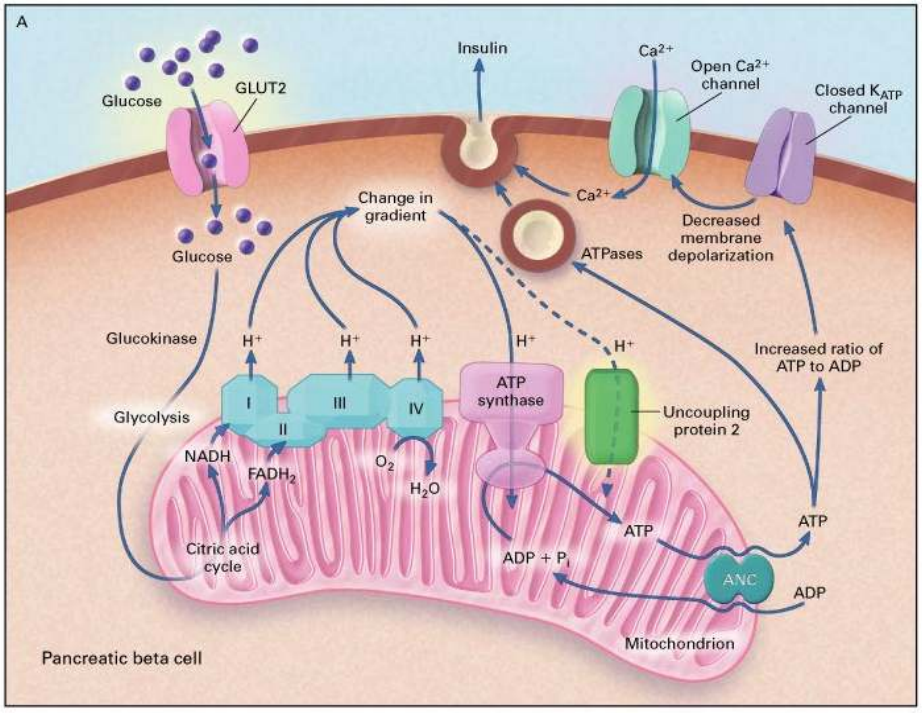


Age above 30 years old & FPG>126mg/dL or diabetes medication treatment

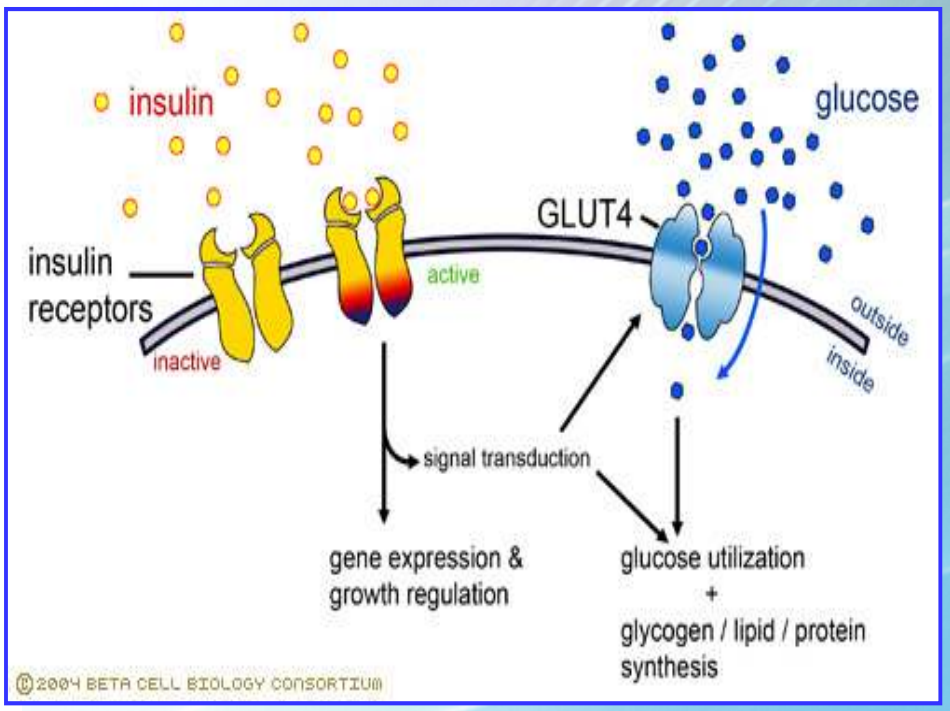
(2008 Korea National Health and Nutrition Examination Survey)

Pathobiology of T2D

Glucose-stimulated secretion of insulin



Insulin-mediated glucose uptake (adipocyte/skeletal muscles)



Defect in Insulin secretion in pancreatic beta cells

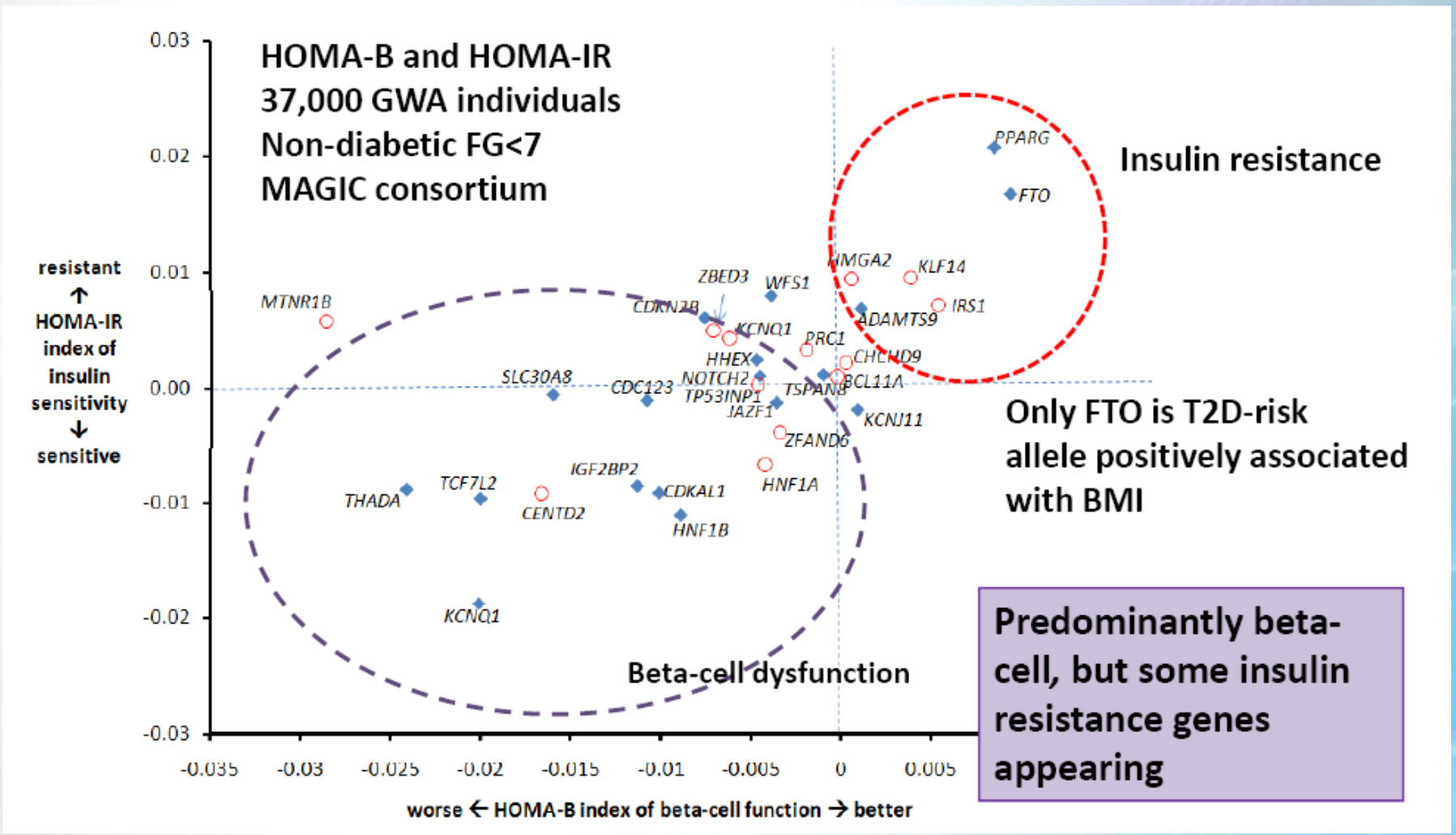
Insulin resistance in adipocytes & skeletal muscle cells

Efforts to identify susceptibility loci for T2D

2000	*PPARG*			
2003	*KCNJ11*			
2006	*TCF7L2*			
2007	*FTO*	*SLC30A8*	*HHEX/IDE*	*WFS1*
	CDKAL1	*IGF2BP2*	9p21	*HNF1B*
2008	*MTNR1B*	*KCNQ1*	*TSPAN8*	*THADA*
	ADAMTS9	*NOTCH2*	*CAMK1D*	*JAZF1*
2009	*TP53INP1*	*KLF14*	*ZBED3*	*HMGA2*
	BCL11A	*CHCHD9*	*HNF1A*	*IRS1*
2010	*DGKB*	*GCKR*	*GCK*	*CENTD2*
	ADCY5	*PROX1*	*DUSP8*	*SRR*
	DUSP9	*ZFAND6*	*PRC1*	*PTPRD*
	UBE2E2	*C2CD4A-C2CD4B*	*SPRY2*	*CDC123/CAMK1D*
2011	*GRB14*	*ST6GAL1*	*VPS26A*	*HMG20A*
	AP3S2	*HNF4A*		

So far, 49 T2D loci have been identified by candidate gene studies, GWAS for T2D, GWAS of related traits, GWA meta-analysis for T2D

Physiological mechanisms implied from European T2D loci



(Adapted from McCarthy)

Why GWAS in Korean (or Asian) population ?

- 대부분의 전장유전체연관분석연구는 유럽인을 대상으로 이루어져 왔음

- 특정 유전적 변이는 다른 인구집단에 따라 특정 질환 혹은 형질에 상이한 영향을 미칠 수 있음
  1. Allele 빈도의 차이
  2. LD 구조의 차이
  3. 환경적 차이 (예: GGI, GEI, selection, drift)
  4. 각 집단에서 새롭게 생겨난 돌연변이

KARE Project

The 1st Korea GWAS of population-based cohorts

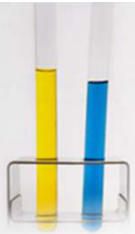
Objectives of KARE (Korea Association REsoure) project

1. Genome-wide association study (GWAS) 수행을 통한
   - Quantitative trait(QT)들에 영향을 미치는 유전요인 발굴
   - 생활습관질환(예, 제2형 당뇨)에 영향을 미치는 유전요인 발굴
2. 국내 유전체 연구의 기반확립
   - 한국인 10,000 여명에 대한 500,000 SNP 정보 확보
   - 유전체 정보의 DB화 및 국내연구자들에게 유전체 정보제공

KARE phenotypes

Phenotypes for KARE study		
Diseases	- T2D	
	- Hypertension	
	- Osteoporosis (bone mineral density)	
	- Obesity	
	- Metabolic syndrome	
	- Dyslipidemia	
Health related quantitative traits	- T2D related traits	Fasting plasma glucose/insulin, Glucose/insulin OGTT 60, Glucose/insulin OGTT 120, HBA1C, HOMAB, HOMA_IR
	- Plasma lipids	LDLC, HDLC, TG, TCHL
	- Blood pressure	SBP, DBP
	- Liver enzymes	γ–GTP, AST, ALT
	- Kidney function related traits	Creatinine, eGFR, Plasma_albumin, BUN
	- Obesity related traits	Weight, Waist, BMI, WHR, Height
	- Pulse rate	Pulse rate
	- Hematological traits	Hemoglobin, Hematocrit, Red blood cell count, White blood cell count

Korea GWAS for Type 2 Diabetes (KARE study)



Korean Genome Epidemiology Study (KoGES) (10,038 subjects)

↓

Genotyping w/ Affymetrix 5.0 (500,568 SNPs)

↓

Data Filtering (8,842 subjects/352,228 SNPs)

↓

Association Analyses (T2D case-control) **(1,042/2,943)**

↓

Identification of significant SNPs

↓

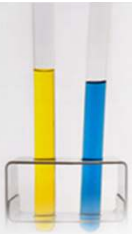
Replication study from independent population (Health2) **(1,216/1,352)**

Results of GWAS for T2D in KARE study

SNP ID	CHR	Gene	Stage 1 (KARE-GWAS) (1042/2943)			Stage 2 (replication) (1216/1352)			All Korean (stage 1 + stage 2) (2258/4295)	
			MAF	OR (95% CI)	*P*-value	MAF	OR (95% CI)	*P*-value	OR (95% CI)	*P*-value
SNPs showing the strong evidence of association										
rs7754840	6	*CDKAL1*	0.53/0.46	1.30 (1.17-1.45)	1.06E-06	0.51/0.46	1.22 (1.08-1.38)	1.68E-03	1.26 (1.17-1.37)	5.04E-09
rs10811661	9	*CDKN2A/B*	0.40/0.45	0.79 (0.71-0.88)	1.41E-05	0.39/0.45	0.78 (0.68-0.89)	2.24E-04	0.79 (0.72-0.86)	2.07E-08
rs1106XXXX	12	*Gene 1*	0.15/0.19	0.70 (0.61-0.81)	1.96E-06	0.16/0.19	0.74 (0.62-0.88)	4.73E-04	0.72 (0.64-0.80)	6.68E-09
rs207XXXX	12	*Gene 2*	0.12/0.16	0.66 (0.57-0.78)	3.81E-07	0.12/0.16	0.73 (0.61-0.88)	1.08E-03	0.69 (0.61-0.78)	3.04E-09
SNPs showing the moderate evidence of association										
rs4376068	3	*IGF2BP2*	0.32/0.28	1.26 (1.12-1.41)	7.42E-05	0.30/0.26	1.20 (1.04-1.38)	1.36E-02	1.23 (1.13-1.35)	2.47E-06
rs3821964	4	*BMPR1B*	0.45/0.50	0.80 (0.72-0.89)	4.19E-05	0.47/0.49	0.87 (0.76-0.99)	2.96E-02	0.83 (0.76-0.90)	8.41E-06
rs6882351	5	5p13.1b	0.33/0.39	0.78 (0.69-0.87)	1.31E-05	0.36/0.39	0.85 (0.75-0.97)	1.63E-02	0.81 (0.74-0.88)	1.30E-06
rs10258075	7	*INSIG1*	0.16/0.12	1.39 (1.19-1.61)	2.30E-05	0.13/0.11	1.22 (1.00-1.49)	4.72E-02	1.32 (1.17-1.49)	4.92E-06
rs2868088	20	*HNF4A*	0.41/0.46	0.80 (0.72-0.89)	3.55E-05	0.42/0.44	0.87 (0.76-0.99)	4.14E-02	0.82 (0.76-0.90)	7.39E-06

E Asian T2D GWA meta-analysis to identify more T2D loci

Stage	Representative	Study	Ethnic group	Sample size		
				case	control	total
Stage 1	KNIH	KARE	Korean	1042	2943	3985
	NUS	SP1	Chinese	1082	1006	2088
		SP2	Chinese	928	939	1867
		SiMES	Malay	794	1240	2034
	IMCJ	IMCJ	Japanese	931	1404	2335
	Vanderbilt U	Shanghai	Chinese	1019	1710	2729
	Taiwan	Taiwan	Chinese	997	999	1996
	UNC	CLHNS	Philipino	159	1624	1783
	total			**6952**	**11865**	**18817**
Stage 2	RIKEN/Tokyo U	BBJ	Japanese	4885	3779	8664
	KNIH	Health2	Korean	1183	1305	2488
	SJTU	Shanghai	Chinese	190	198	388
	total			**6258**	**5282**	**11540**
Stage 3	IMCJ	IMCJ	Japanese	5253	6160	11413
	SJTU	Shanghai	Chinese	3410	3412	6822
	CUHK	CUHK	Chinese	1500	1500	3000
	NTUH	NTUH	Chinese	1500	1500	3000
	SNU	SNU	Korean	600	700	1300
	total			**12263**	**13272**	**25535**
Overall	**AGEN**	**AGEN-T2D**	**East Asian**	**25473**	**30419**	**55892**



Stage1 : Discovery

✓**GWA meta-analysis combining 8 T2D GWA studies**
(6,952 cases vs. 11,865 controls)

$P < 5 \times 10^{-4}$

Stage2 : *in silico* replication

✓**Validation of 4,014 SNPs selected from Stage1**
(303 lead SNPs + their proxy SNPs)
in 3 T2D GWA studies (6,258 cases vs. 5,282 controls)
✓**Combined meta-analysis (Stages 1+2)**

$P < 10^{-5}$

Stage3 : *de novo* replication

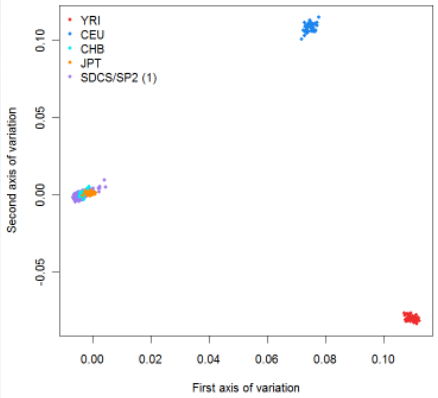
✓**Validation of 19 SNPs selected from Stage2**
in 5 T2D studies (12,263 cases vs. 13,272 controls)
✓**Combined meta-analysis (Stages 1+2+3)**

$P < 5 \times 10^{-8}$

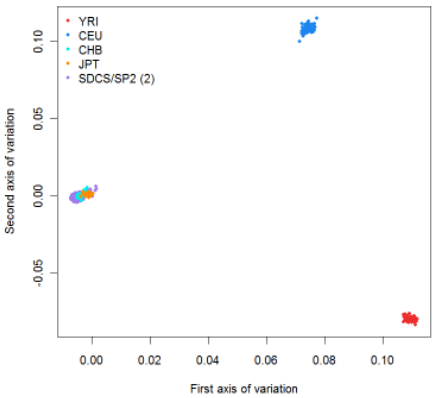
8 novel T2D SNPs

Principal Component Analysis (PCA) in all individuals from each stage 1 component study and 270 individuals from HapMap DATA

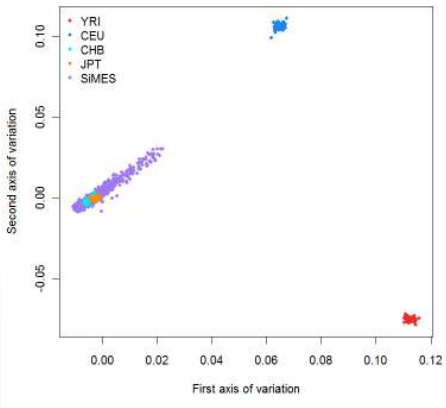
(A) SDCS/SP2 (1), Chinese



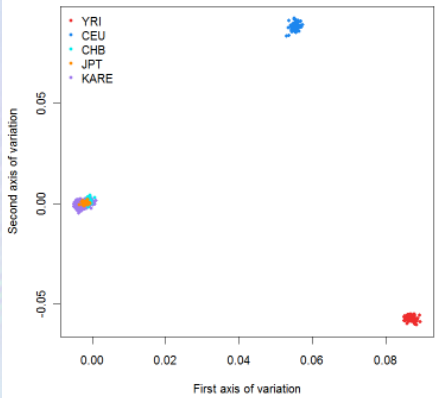
(B) SDCS/SP2 (2), Chinese



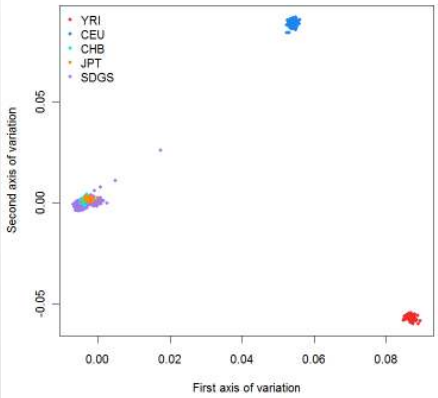
(C) SiMES, Malays



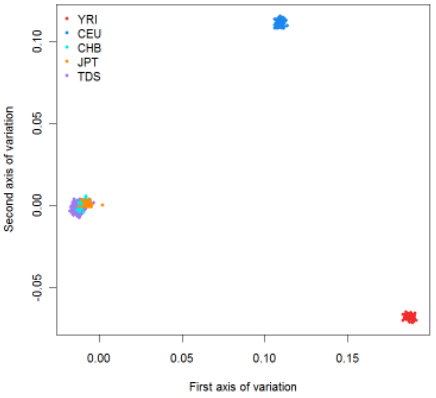
(D) KARE, Korean



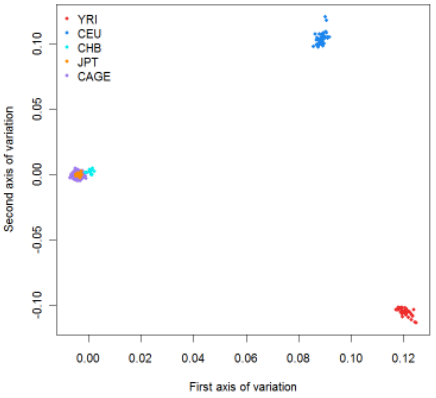
(E) SDGS, Chinese



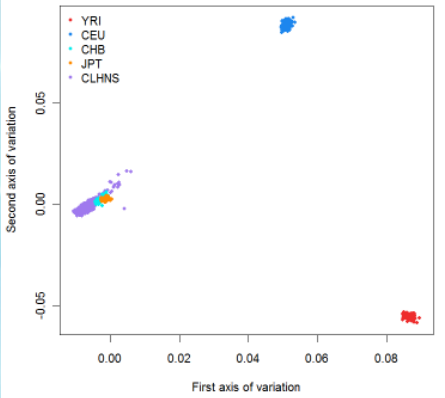
(F) TDS, Taiwanese



(G) CAGE, Japanese



(H) CLHNS, Filipino



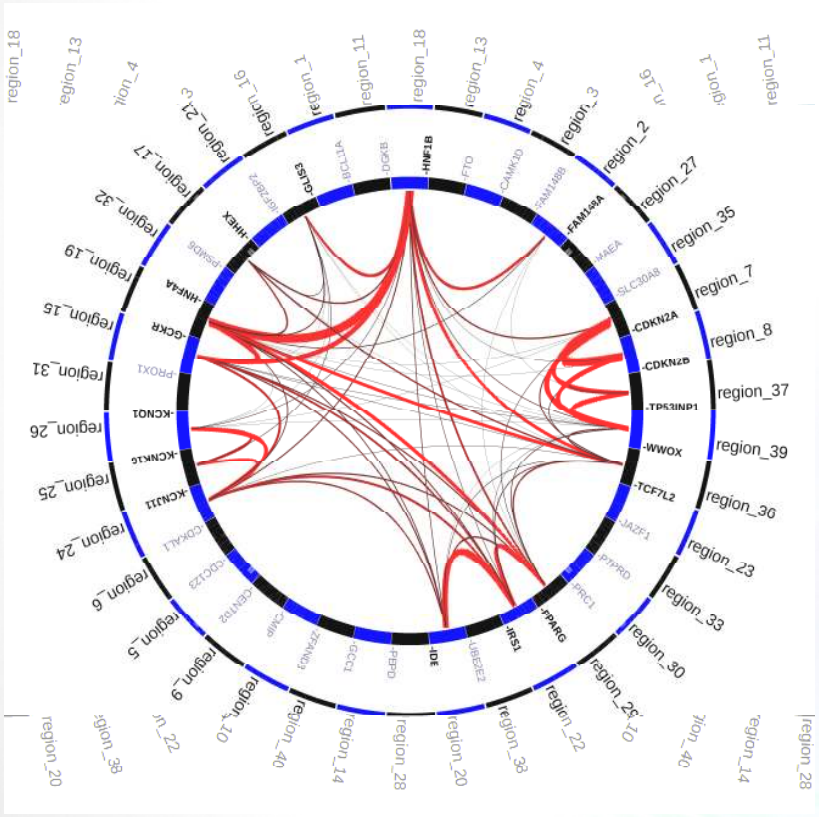
Summary of GWA meta-analysis for T2D

Candiate gene	Risk allele	Other allele	Combined (stage 1+2+3) RAF (HapMap JPT/CHB)	OR (CI)	*P*-value	DIAGRAM+ RAF (HapMap CEU)	OR (CI)	*P*-value	power
			up to 25,079 cases and 29,611 controls			up to 8,130 cases and 38,987 controls			
Loci showing strong evidence of association with T2D									
MAEA	c	g	0.58	1.13 (1.10-1.16)	1.57E-20	NA	NA	NA	NA
GLIS3	a	g	0.41	1.10 (1.07-1.13)	1.99E-14	0.54	1.04 (1.00-1.08)	6.43E-02	0.62
HNF4A	g	t	0.48	1.09 (1.07-1.12)	1.12E-11	0.18	1.07 (1.01-1.13)	1.47E-02	0.86
GCC1	g	a	0.79	1.11 (1.07-1.14)	4.96E-11	0.56	0.99 (0.95-1.03)	4.89E-01	0.09
PSMD6	c	t	0.61	1.09 (1.06-1.12)	8.41E-11	0.76	1.02 (0.97-1.07)	4.45E-01	0.16
ZFAND3	c	t	0.27	1.12 (1.08-1.16)	2.06E-10	0.14	0.97 (0.90-1.04)	4.00E-01	0.23
PEPD	a	g	0.56	1.10 (1.07-1.14)	1.30E-08	0.6	1.02 (0.98-1.06)	3.61E-01	0.2
KCNK16	t	g	0.42	1.08 (1.05-1.11)	2.30E-08	0.47	NA	NA	NA
Loci showing moderate evidence of association with T2D									
CMIP	c	t	0.8	1.08 (1.05-1.12)	2.84E-07	0.99	1.20 (1.01-1.42)	3.33E-02	0.52
WWOX	t	c	0.32	1.08 (1.05-1.12)	9.49E-07	0.02	1.20 (0.95-1.52)	1.22E-01	0.87

The power was estimated given the 8,130 cases/38,987 controls, DIAGRAM+ ORs, T2D prevalency of 10% and RAF in HapMap (CEU) for α=0.05.

Pathway enrichments

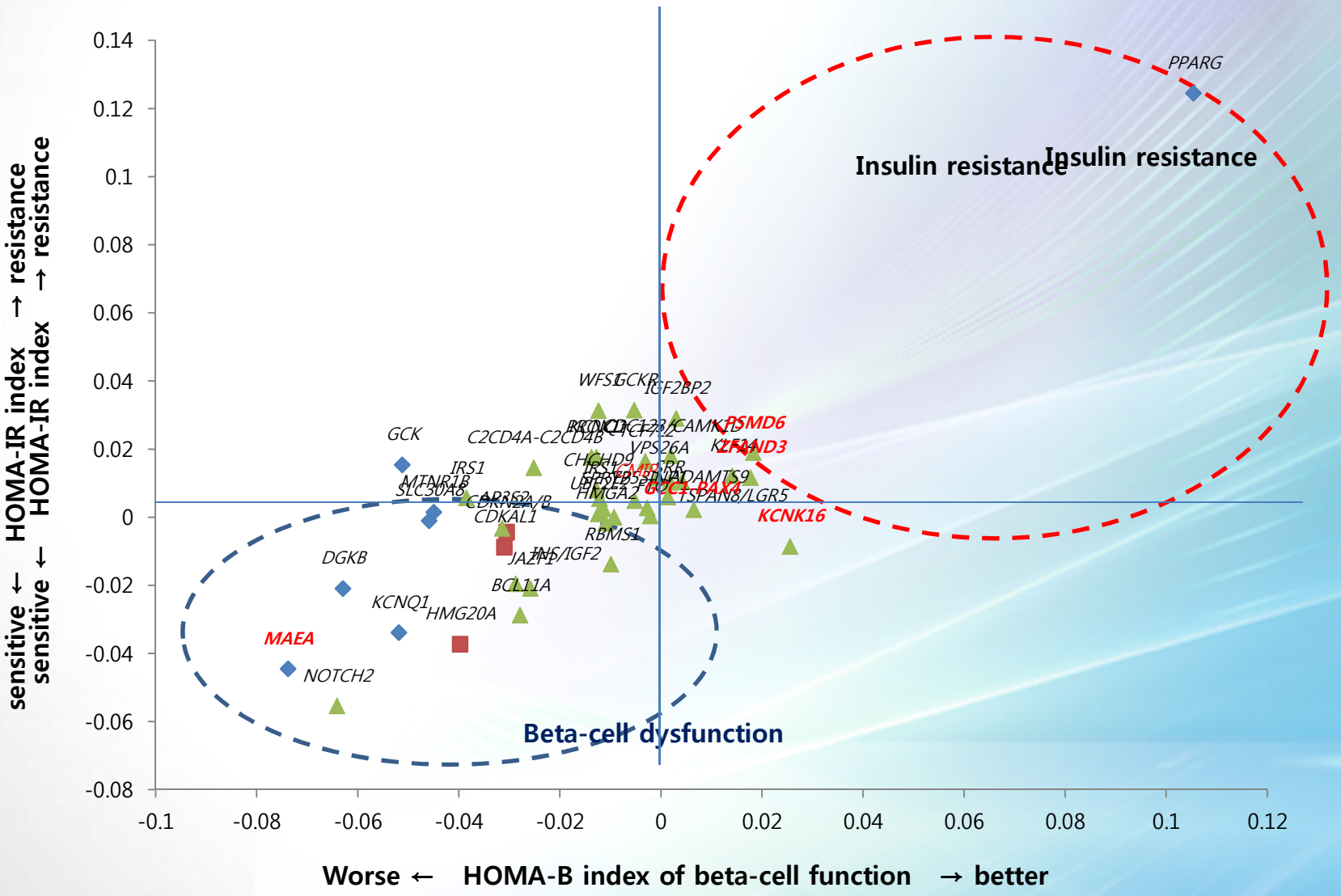
GRAIL
Pubmed abstract mining



Some connectivity but no clear hits to latent mechanisms

GENE	GRAIL P_{Gene} value	SELECTED SIMILAR GENES (Rank in parantheses)
HNF1B	1.52E-06	HNF4A(3), GCKR(98), TCF7L2(111), GLIS3(224), PPARG(340), PROX1(355), PEPD(389), FAM148B(412), WWOX(433), HHEX(474), CDKN2A(608), FAM148A(652), IRS1(1463), KCNJ11(1465), CDKN2B(1732), CAMK1D(1993)
CDKN2A	8.54E-06	CDKN2B(1), WWOX(18), TP53INP1(64), CENTD2(320), PEPD(349), HNF1B(460), TCF7L2(606), PPARG(1059), PSMD6(1245), PTPRD(1299), BCL11A(1414), PROX1(1479), CAMK1D(1482), HNF4A(1663)
IDE	2.05E-05	IRS1(5), PEPD(50), PPARG(160), SLC30A8(278), KCNJ11(367), HNF4A(378), HNF1B(407), CAMK1D(670), GCKR(681), TCF7L2(702), WWOX(1202), PSMD6(1437)
GCKR	2.07E-05	HNF4A(18), HNF1B(24), PEPD(138), IRS1(147), KCNJ11(177), PPARG(196), IDE(581), WWOX(834), TCF7L2(896), CAMK1D(1021), GLIS3(1512), PROX1(1733)
HNF4A	2.17E-05	HNF1B(3), GCKR(85), TCF7L2(121), PPARG(127), PROX1(348), HHEX(394), PEPD(575), WWOX(822), IRS1(833), CAMK1D(1138), GLIS3(1161), BCL11A(1497), KCNJ11(1717)
CDKN2B	2.96E-05	CDKN2A(1), WWOX(29), TP53INP1(273), BCL11A(401), PEPD(407), JAZF1(462), CAMK1D(652), TCF7L2(751), HNF1B(863), PROX1(1039), HNF4A(1521), PTPRD(1623), PSMD6(1807), PPARG(1984)
IRS1	0.001055	IDE(90), PPARG(102), KCNJ11(354), HNF4A(472), SLC30A8(609), TCF7L2(623), HNF1B(748), CAMK1D(807), WWOX(830), PEPD(888), IGF2BP2(902), GCKR(922)
KCNJ11	0.0011244	KCNQ1(103), IRS1(229), KCNK16(267), PPARG(365), HNF1B(376), HNF4A(395), GCKR(491), IDE(655), PEPD(782), TCF7L2(828), SLC30A8(1790), WWOX(1994)
FAM148A	0.0031708	FAM148B(1), HNF1B(64), HNF4A(492), CAMK1D(747), TP53INP1(1040), PPARG(1327)
GLIS3	0.003893	HNF1B(174), ZFAND3(196), JAZF1(383), SLC30A8(403), HNF4A(500), BCL11A(611), PROX1(694), CAMK1D(802), HHEX(952), WWOX(1174), PEPD(1286), TCF7L2(1646)
TCF7L2	0.0051165	HNF4A(185), HNF1B(192), WWOX(344), PPARG(505), HHEX(524), PROX1(634), CAMK1D(756), CDKN2A(810), PEPD(1009), IRS1(1058), CDKN2B(1378), IGF2BP2(1877), BCL11A(1921)
PPARG	0.0068465	IRS1(71), HNF4A(142), HNF1B(371), TCF7L2(451), WWOX(531), PEPD(683), KCNJ11(1125), CDKN2A(1187), CAMK1D(1326), IDE(1536), GCKR(1878)
WWOX	0.0094217	CDKN2A(50), TP53INP1(112), CDKN2B(167), PTPRD(514), PEPD(650), PROX1(1311), CAMK1D(1440), TCF7L2(1454), BCL11A(1847)
TP53INP1	0.0214116	WWOX(39), CDKN2A(82), CDKN2B(425), CAMK1D(662), PEPD(1009), PROX1(1243), CDC123(1946)
HHEX	0.0318255	PROX1(97), HNF4A(420), HNF1B(492), TCF7L2(496), BCL11A(614), CAMK1D(1140), WWOX(1253), PEPD(1295), GLIS3(1474)
KCNK16	0.0330698	KCNJ11(74), KCNQ1(130), CENTD2(514), GLIS3(686), CAMK1D(889), PEPD(1010), SLC30A8(1657)
KCNQ1	0.0332554	KCNJ11(78), PEPD(350), KCNK16(496), HNF1B(597), CAMK1D(757), CDKN2A(1101), WWOX(1109), GLIS3(1652), PROX1(1662), HNF4A(1673)

Physiological mechanisms of E Asian T2D loci



GWA meta-analysis for T2D in E Asian populations

LETTERS

nature genetics

Meta-analysis of genome-wide association studies identifies eight new loci for type 2 diabetes in east Asians

Yoon Shin Cho[1,46], Chien-Hsiun Chen[2,3,46], Cheng Hu[4,46], Jirong Long[5,46], Rick Twee Hee Ong[6,46], Xueling Sim[7,46], Fumihiko Takeuchi[8,46], Ying Wu[9,46], Min Jin Go[1,46], Toshimasa Yamauchi[10,46], Yi-Cheng Chang[11,46], Soo Heon Kwak[12,46], Ronald C W Ma[13,46], Ken Yamamoto[14,46], Linda S Adair[15], Tin Aung[16,17], Qiuyin Cai[5], Li-Ching Chang[2], Yuan-Tsong Chen[2], Yutang Gao[18], Frank B Hu[19], Hyung-Lae Kim[1,20], Sangsoo Kim[21], Young Jin Kim[1], Jeannette Jen-Mai Lee[22], Nanette R Lee[23], Yun Li[9,24], Jian Jun Liu[25], Wei Lu[26], Jiro Nakamura[27], Eitaro Nakashima[27,28], Daniel Peng-Keat Ng[22], Wan Ting Tay[16], Fuu-Jen Tsai[3], Tien Yin Wong[16,17,29], Mitsuhiro Yokota[30], Wei Zheng[5], Rong Zhang[4], Congrong Wang[4], Wing Yee So[13], Keizo Ohnaka[31], Hiroshi Ikegami[32], Kazuo Hara[10], Young Min Cho[12], Nam H Cho[33], Tien-Jyun Chang[11], Yuqian Bao[4], Åsa K Hedman[34], Andrew P Morris[34], Mark I McCarthy[34,35], DIAGRAM Consortium[36], MuTHER Consortium[36], Ryoichi Takayanagi[37,47], Kyong Soo Park[12,38,47], Weiping Jia[4,47], Lee-Ming Chuang[11,39,47], Juliana C N Chan[13,47], Shiro Maeda[39,47], Takashi Kadowaki[10,47], Jong-Young Lee[1,47], Jer-Yuarn Wu[2,3,47], Yik Ying Teo[6,7,22,25,41,47], E Shyong Tai[22,42,43,47], Xiao Ou Shu[5,47], Karen L Mohlke[9,47], Norihiro Kato[8,47], Bok-Ghee Han[1,47] & Mark Seielstad[25,44,45,47]

Cho et al. *Nature Genetics, 2012*

Limitation in GWAS

Estimation of heritability and number of loci for several complex traits

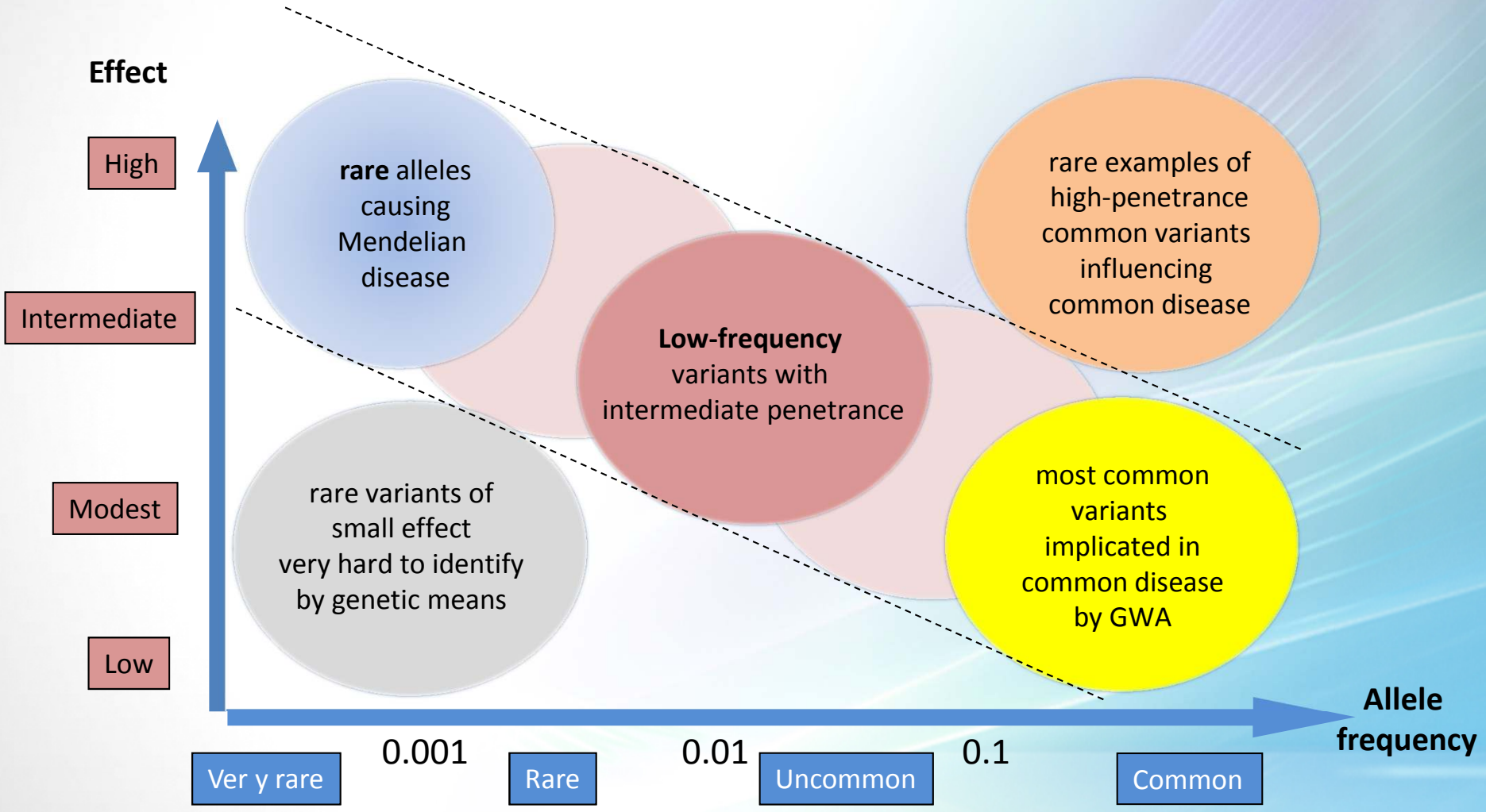
Table 1 | GWAS for common diseases and traits

Phenotype	Number of GWAS loci	Proportion of heritability explained (%)*
Type 1 diabetes	41	~60
Fetal haemoglobin levels	3	~50
Macular degeneration	3	~50
Type 2 diabetes	39	20–25
Crohn's disease	71	20–25
LDL and HDL levels	95	20–25
Height	180	~12

HDL, high-density lipoprotein; LDL, low-density lipoprotein.
* Fraction of heritability explained is calculated by dividing the phenotypic variance explained by variants at loci identified by GWAS by the total heritability as inferred from epidemiological parameters.

(Lander, Nature 2011)

Atlas of susceptibility



(Adapted from McCarthy)

Thus, future studies to explain missing heritability

1. GWA meta-analysis & ethnic specific GWAS

- More common variants
- ethnic specific variants

2. Fine mapping of candidate T2D loci or Exome sequencing

- rare variants
- causal variants

3. Structural variants

- CNVs
- indels

4. Others

- GXG interaction, GXE interaction
- epigenetic modifications

The path for disease genomics

	Method	Technology	Main Purpose
2005	Candidate Gene Approach for Association Analysis	Genotyping (1~1000 SNPs)	- identification of disease associated loci - identification of causal variation for disease
2007	Genome Wide Association Study (GWAS)	High throughput genotyping (> 500K SNPs)	- identification of disease associated loci - identification of causal variation for disease (rarely)
2011	Genome-Wide Association Meta-Analysis (GWA MA)	Imputation/meta-analysis (> 1.5~3 M imputed/ genotyped SNPs)	- identification of disease associated loci - identification of causal variation for disease (rarely)
	Targeted Resequencing for Disease Loci	NGS (Next Generation Sequencing)	- identification of causal variation for disease
2012	Exome Sequencing	NGS	- identification of causal variation for disease
future	Whole Genome Sequencing	NGS	- identification of causal variation for disease

Acknowledgements

KNIH

Young Jin Kim

Min Jin Go

Ji Hee Oh

Jong-Young Lee

Bok-Ghee Han

Ajou Univ

Nam H. Cho

Korea Univ

Chol Shin

Shanghai Jiao Tong U

Cheng Hu

Weiping Jia

WTCCC

Mark McCarthy

Andrew Morris

IMCJ

Norihiro Kato

RIKEN

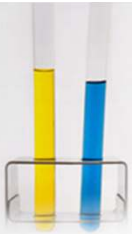
Shiro Maeda

Naoyuki Kamatani

Univ of Tokyo

Takashi Kadowaki

All AGEN members



Thank you!