당뇨병에서의 유전자연구

SNP 연구입문

2012년 7월 14일

연세대학교 의과대학 내과학교실

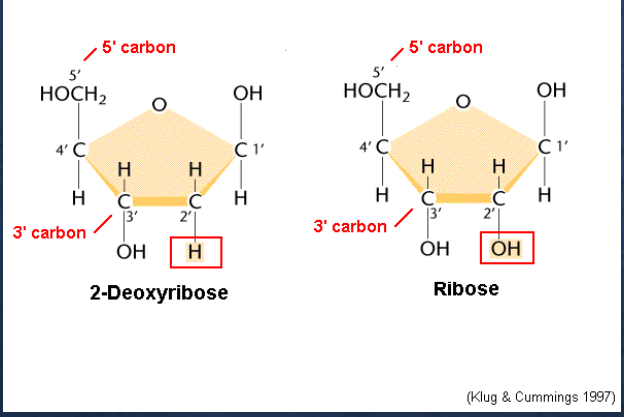
강 은 석

GENE (유전자)

생명체에서 유전의 분자단위로, 생명체 안에서 기능을 갖는 polypeptide나 RNA chain을 Coding 하는 DNA나 RNA

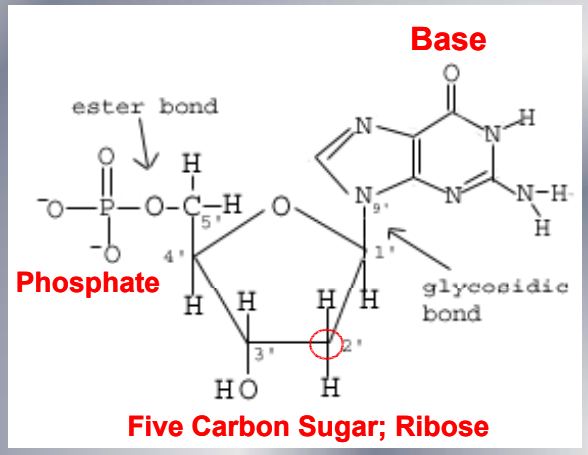
GENE (유전자)

DNA vs. RNA



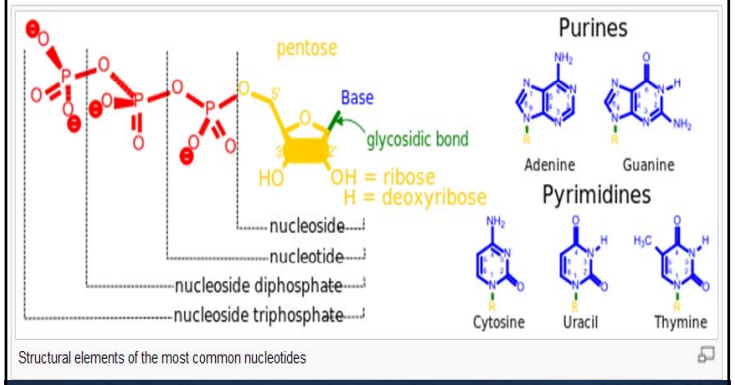
Five Carbon sugar; Ribose

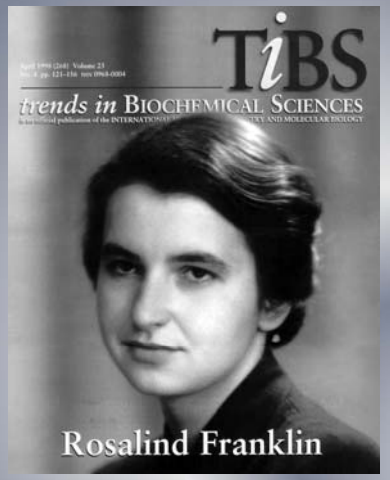
Nucleotide



DNA Double Strands

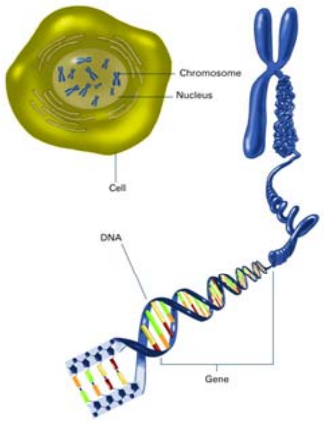
Nucleotides & Base





1920-1958

Chromosome (염색체)and DNA



Chromosome (염색체)

Unsorted Human Chromosomes

Male Karyotype

Female Karyotype

44+XY OR 44+XX

GENOME (유전체)

= GENe + chromosOME

유전체 = 유전자 + 염색체

Chromosome(염색체) 부위 명명법

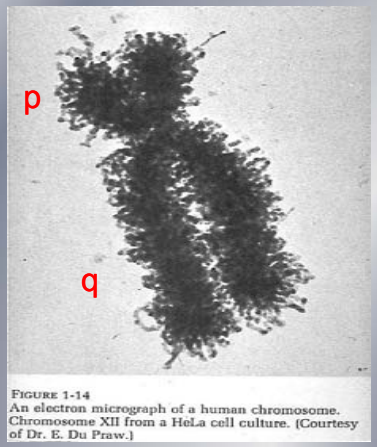


FIGURE 1-14
An electron micrograph of a human chromosome. Chromosome XII from a HeLa cell culture. (Courtesy of Dr. E. Du Praw.)

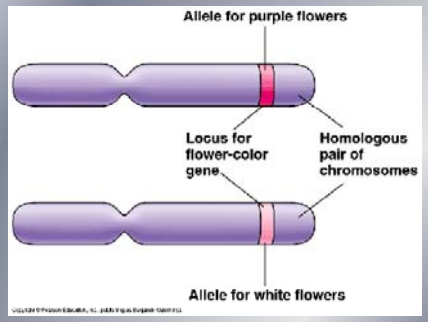
예)
12p23.1

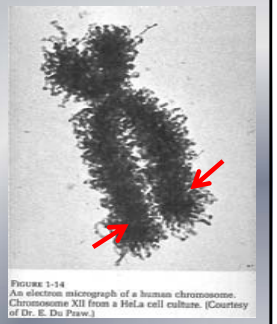
12 = chromosome 12
p = short arm

2 = region
3 = band

.1 = sub-band

Allele (대립유전자)



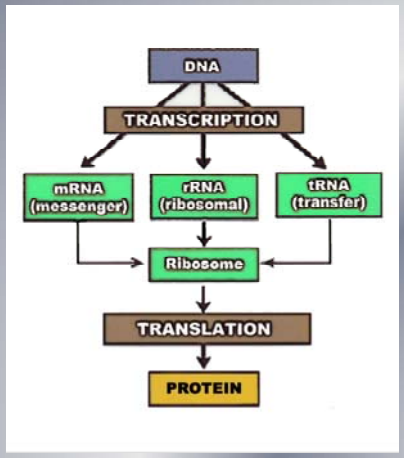


같은 염색체내의 동일 유전자의 다른 부위

Each *locus* on a chromosome has alternative versions of a *gene* called *alleles.*

You inherit one allele from each parent.

Central Dogma (센트럴 도그마)

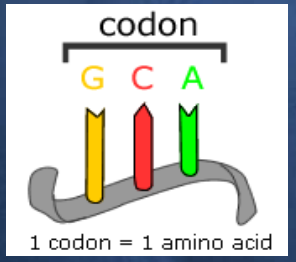


Human Genome (인간 유전체)

Chromosome	Genes	Total base pairs	Sequenced base pairs[1]
1	4,220	247,199,719	224,999,719
2	1,491	242,751,149	237,712,649
3	1,550	199,446,827	194,704,827
4	446	191,263,063	187,297,063
5	609	180,837,866	177,702,766
6	2,281	170,896,993	167,273,993
7	2,135	158,821,424	154,952,424
8	1,106	146,274,826	142,612,826
9	1,920	140,442,298	120,312,298
10	1,793	135,374,737	131,624,737
11	379	134,452,384	131,130,853
12	1,430	132,289,534	130,303,534
13	924	114,127,980	95,559,980
14	1,347	106,360,585	88,290,585
15	921	100,338,915	81,341,915
16	909	88,822,254	78,884,754
17	1,672	78,654,742	77,800,220
18	519	76,117,153	74,656,155
19	1,555	63,806,651	55,785,651
20	1,008	62,435,965	59,505,254
21	578	46,944,323	34,171,998
22	1,092	49,528,953	34,893,953
X (sex chromosome)	1,846	154,913,754	151,058,754
Y (sex chromosome)	454	57,741,652	25,121,652
Total	32,185	3,079,843,747	2,857,698,560

- 염색체 수 22x2+2=46개
- 유전자 수: 32,000개
- 염기서열 수:
  30억 Nucleotides
- 유전자의 <5%만이
  Active

CODON (코돈)



CODON (코돈): 유전정보의 최소 단위

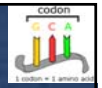
1 Codon = 3 Nucleotide

1 Codon = 1 amino acid

Number of Possible Codons:

4 (A,T,C,G) x 4 (A,T,C,G) x 4 (A,T,C,G) = 64개

Genetic Code (유전암호)



아미노산의 수 = 20

코돈의 수 = 64

3개 Codon (TAA, TGA, TAG): Stop Codon
61개 Codon: 아미노산을 Coding
ATG Codon: Methionine을 Coding하는 Start Codon

Genetic Code

Universal Genetic Code Chart
Messenger RNA Codons and Amino Acids for Which They Code

First base	Second base U	Second base C	Second base A	Second base G	Third base
U	UUU PHE	UCU SER	UAU TYR	UGU CYS	U
	UUC PHE	UCC SER	UAC TYR	UGC CYS	C
	UUA LEU	UCA SER	UAA STOP	UGA STOP	A
	UUG LEU	UCG SER	UAG STOP	UGG TRP	G
C	CUU LEU	CCU PRO	CAU HIS	CGU ARG	U
	CUC LEU	CCC PRO	CAC HIS	CGC ARG	C
	CUA LEU	CCA PRO	CAA GLN	CGA ARG	A
	CUG LEU	CCG PRO	CAG GLN	CGG ARG	G
A	AUU ILE	ACU THR	AAU ASN	AGU SER	U
	AUC ILE	ACC THR	AAC ASN	AGC SER	C
	AUA ILE	ACA THR	AAA LYS	AGA ARG	A
	AUG MET or START	ACG THR	AAG LYS	AGG ARG	G
G	GUU VAL	GCU ALA	GAU ASP	GGU GLY	U
	GUC VAL	GCC ALA	GAC ASP	GGC GLY	C
	GUA VAL	GCA ALA	GAA GLU	GGA GLY	A
	GUG VAL	GCG ALA	GAG GLU	GGG GLY	G

A mapping of amino acids and stop signals to DNA codons

Amino Acid/Signal	Codons	Amino Acid/Signal	Codons
A	GCT, GCC, GCA, GCG	C	TGT, TGC
D	GAT, GAC	E	GAA, GAG
F	TTT, TTC	G	GGT, GGC, GGA, GGG
H	CAT, CAC	I	ATT, ATC, ATA
K	AAA, AAG	L	TTA, TTG, CTT, CTC, CTA, CTG
M	ATG	N	AAT, AAC
P	CCT, CCC, CCA, CCG	Q	CAA, CAG
R	CGT, CGC, CGA, CGG, AGA, AGG	S	TCT, TCC, TCA, TCG, AGT, AGC
T	ACT, ACC, ACA, ACG	V	GTT, GTC, GTA, GTG
W	TGG	Y	TAT, TAC
START	ATG	STOP	TAA, TGA, TAG

Genetic Variations
(유전자 변이)

Germline mutation, not somatic mutation

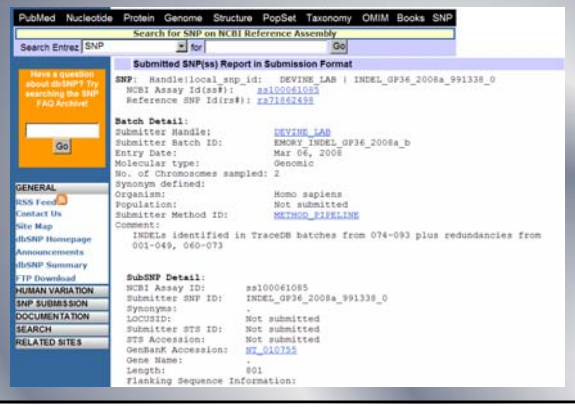
DNA Genetic Variations

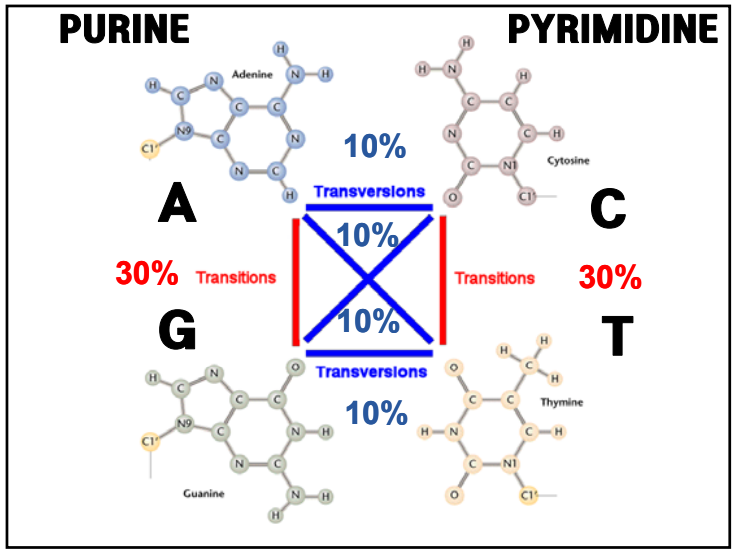
1. **Copy Number Variation**
   1) STR (Short Tandem Repeat)
      = microsatellite 반복단위 2-7 base pair
   2) VNTR (Variable Number of Tandem Repeat)
      = 반복단위 14-70 base pair
2. **Insertion/Deletion, Translocation**
   예) Frameshift
3. **SNP (Point mutation)**
   Silent (Synonymous) mutation: No AA Change
   Missense (Non-Synonymous) mutation: AA Change
   Nonsense mutation = Stop Codon

SNP
(Single Nucleotide Polymorphism)

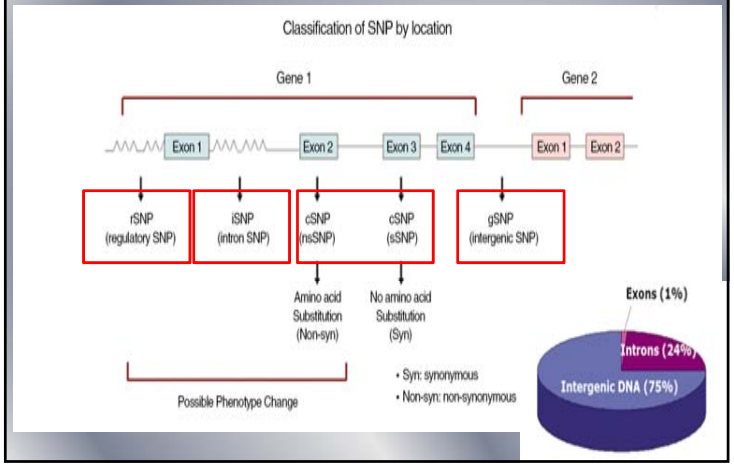
- SNP: DNA 염기서열에서 한 개의 염기서열의 차이를 보이는 유전적 변화
- Polymorphism: 1% 이상의 빈도로 존재하는 2개의 대립 염기서열(Bi-allelic)변이
  - Common polymorphism (MAF > 5%)
  - Rare polymorphism (MAF = 1-5%)
- Mutation (MAF < 1%)
- Useful genetic marker
  No. of SNP: 2,365만개
  dbSNP build 131 (2011.5)

SNP ID
Submitted SNP (SS) & Reference SNP (RS)





Classification of SNP by Location



MAF (Minor Allele Frequency)

대립 유전자형의 빈도가 낮은 것의 비율

SNP마커의 유전적 다양성을 표현하는 표준지표

MAF > 5% : Common polymorphism

MAF 1-5% : Rare polymorphism

MAF < 1% : Mutation

SNP 명명법

Nucleotide Numbering
ATG Translation Start Site기준: ATG의 A=+1
ATG codon 5`앞 염기서열 -1
No base 0

SNP 명명법

g for genomic sequence; g85T>G

c for cDNA sequence; c85T>G

m for mitochondrial sequence; m55T>G

r for RNA sequence; r63u>a

p for protein sequence; pR325W

SNP 명명법

Intron; IVS (InterVening Sequence)
앞쪽 인트론:
앞의 exon끝 번호 + exon 끝부터의 염기 수
77+1G
(exon번호 알 때 IVS1+1G)

뒤쪽 인트론:
뒤의 exon시작 번호 + exon 앞부터 염기 수
78-2A
(exon번호 알 때 IVS1-2A)

SNP 명명법

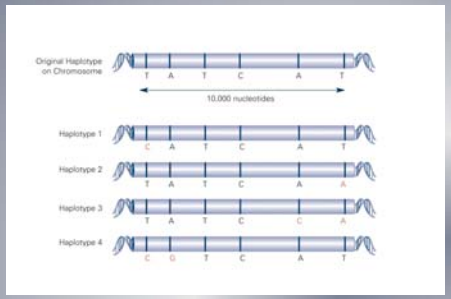
치환 Substitution
86A>G; 86번째 염기가 A에서 G로
IVS2-2A>C; intron 2번의 -2위치에서 A가 C로

결실 Deletion
76_78delACT; 76번과 78번사이에서 ACT결손

삽입 Insertion
76_77insT; 76번과 77번사이에 T 삽입

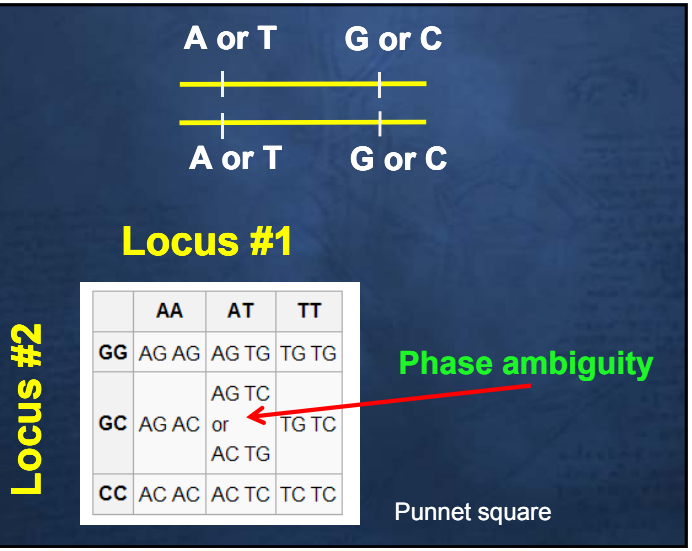
중복 Duplication
76_77dupCT; 76번과 77번사이에 CT중복

Haplotype



같은 chromosome내에서 함께 유전되는 경향이 있는 인접한 SNP들의 집합

Haplotype



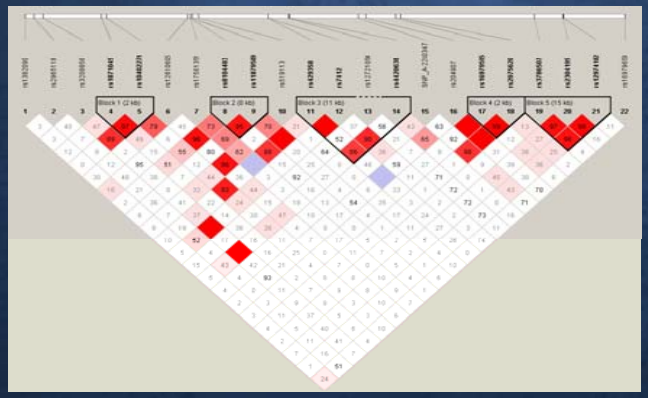
Haplotyping

Haplotyping 방법

Molecular Haplotyping
(분자생물학적 방법)
가계도분석, 정확, 비용, 많은 시간소요

In Silico Haplotyping (통계적 방법)

Haplotype-Tagging SNP (htSNP) And HapMap (Haplotype Map)



Haplotype Program

Haploview (Broad Institute; Harvard and MIT)
http://www.broadinstitute.org/scientific-community/science/programs/medical-and-population-genetics/haploview/haploview

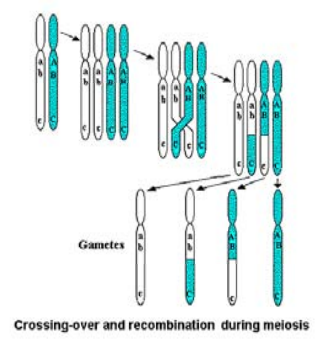
Haplotyper (Harvard University)
http://www.people.fas.harvard.edu/~junliu/Haplo/click.html

PHASE2 (Univ. of Chicago, Metthew Stephens)
http://www.stat.washington.edu/stephens/phasefaq.html

SAS Genetics
http://sas.com

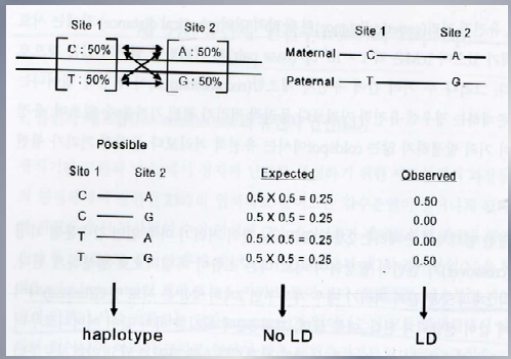
Recombination & Linkage

Recombination (재조합)



감수분열 과정에서 염색체의 일부가 서로 교환하는 교차 (crossover)가 일어나 많은 대립유전자들이 새로운 조합을 가지게 되는 것

Linkage



LD란 두 loci에서 일어나는 대립유전자들 쌍의 이론적인 예측치와 실제측정간의 차이 (deviation)를 나타냄.

Linkage (연관)

두 개의 SNP간의 거리가 매우 가까우면 2개의 SNP는 서로 연관되어 있어서 다음 세대에 같이 전달

Linkage Equilibrium (연쇄평형): 유전자들이 독립적으로 배합되어 있는 상태, 서로 다른 locus에 있는 유전자의 대립형질은 서로 독립적으로 나타남. 따라서, haplotype 빈도는 각 대립형질 빈도의 곱.

Linkage Disequilibrium (연쇄비평형): 유전자들이 의존적으로 배합되어 있는 상태

LD를 측정하는 방법

D : Linkage disequilbrium coefficient

$D = P(AB) - P(A)P(B)$

D=0 if LE. -0.25<D<+0.25

D' : D/D_{max}

D'=0 if LE

D'=1 if complete LD (no recombination)

$0<D'<1$ if variable LD with recombination

r^2 : LD correlation coefficient

$r^2 = D^2/p_1p_2q_1q_2$

$0<r^2<1$

r^2=1 if perfect LD

if r^2>0.33, strong LD

기초 통계유전학

Genotype / Allele Frequency

Genotype	N
AA	300
AB	500
BB	200
Total	1,000

Allele		N
A	300*2 + 500	1,100
B	200*2 + 500	900
Total		2,000

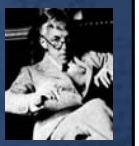
Allele, Genotype and Phenotype

Phenotype → Genotype → Gene/Allele

Phenotype	Genotype	Gene/Allele
A형	AA, AO, OA	A
B형	BB, BO, OB	B
AB형	AB	O
O형	OO	
M형	MM	M
N형	NN	N
MN형	MN	

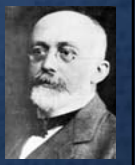
Hardy-Weinburg Equilbrium (Genotype 검증)

정의
무작위교배를 하는 큰 집단에서 유전자와 유전자형의 빈도는 세대를 거듭하여도 변하지 않고 평형을 이루게 된다.



Godfrey Hardy
1877~1947

조건
① 교배는 무작위적으로 이루어져야 한다 (Random mating).
② 돌연변이는 생기지 않는다(no mutation).
③ 이입과 이출이 없다(no migration).
④ 개체군에는 선택이 작용하지 않는다(no natural selection).
⑤ 표본집단의 크기가 크다(population size is infinite).

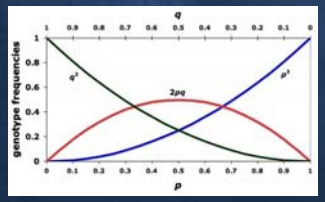


Wilhelm Weinberg
1862-1937

Hardy-Weinburg Equilibrium 검증 방법

기대 값과 실제로 관측한 값의 차이를 X2 (chi-square test로 검증)

P-value <0.05 이면 deviated from HWE



질병유전자 발굴 방법

Hardy-Weinburg Equilbrium 에서 deviation된 경우는?

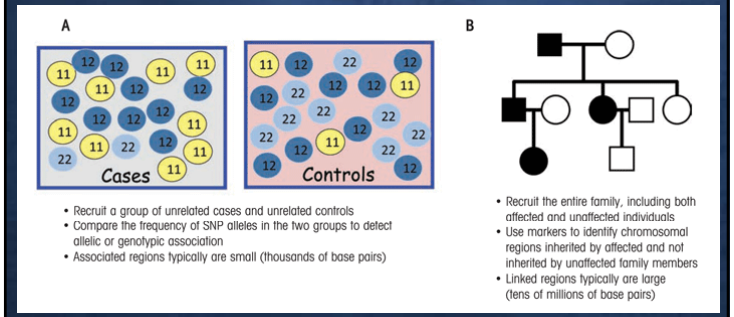
Heterozygote Excess

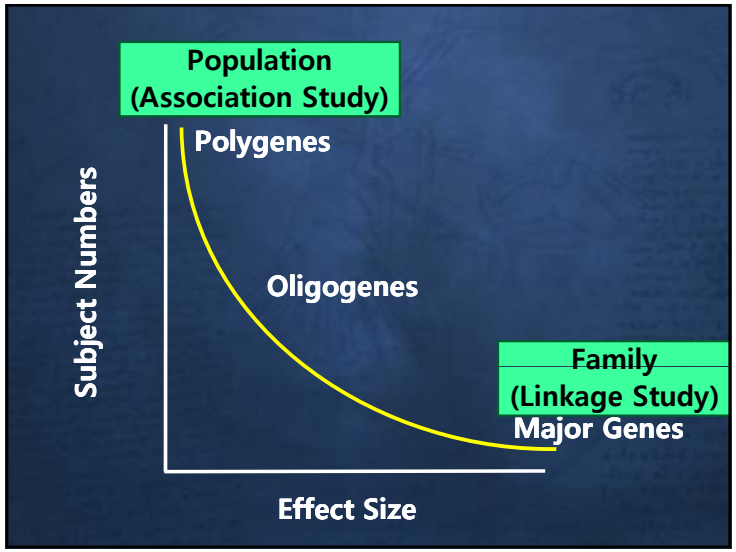
다른 생존율의 차이에서 기인 (differential survival)
Genotyping error (nonspecific assay)

Homozygote Excess

Population 문제일 가능성
(이질적인 집단의 시료; population stratification)
치사 유전자 (Null Allele)
Genotyping error (nonspecific assay)

Association vs. Linkage Study





Association Study 단계

1. 실험설계
   - 연구대상 질환선정
   - 연구대상 시료 수 결정
   - 정상과 질병 구분 기준 결정
2. 시료 및 데이터 수집
   - 유전체 시료수집
   - 임상정보 수집
3. 유전자형 조사 (Genotyping)
   - 후보유전자 및 후보 SNP선정 및 유전형조사
4. 통계분석
   - HWE 검증, 단일 SNP연관성분석, LD구조분석, Haplotype연관성분석
   - Multiple Comparison 보정

Association Study의 의미

1. 질병의 직접적 원인이 되는 경우
2. 질병 원인유전자와 Linkage 된 경우 – Marker
3. 특정유전자형과 질병이 자연적으로 선택된 경우
4. 집단간의 유전적 조성차이에 의한 연관성 (population stratification)
5. False Positive, Type 1 Error

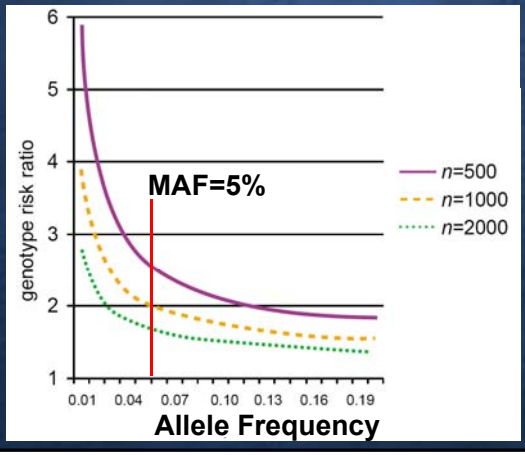
Association Study 방법

통계의 유의성 (POWER) = 1- beta (type 2 error)
Type 2 Error = False negative

통계적 유의성 (POWER)에 영향을 주는 요소
1. 시료의 수 – 절대적임.
2. 유전모델 – dominant or recessive
3. Allele frequency
4. Relative Risk

→ 임의로 조절 가능한 것은 시료의 수(Sample Size) 뿐임

Association Study 방법
Sample Size



Association Study
재현성 (Replication)

통계적 반복실험 (Statistical Replication)
사용된 시료를 나누어 재분석
다른 시료에서 동일한 분석

기능적 반복실험 (Functional Replication)

기존 연구결과가 재현되지 않을 때의 원인
기존 연구가 False Positive
재현 연구의 False Negative
두 연구에 사용된 시료에 대한 집단의 차이

Association Study – Phenotype
정상과 질병 구분 기준 결정

대부분의 연관성연구에서는 특정기준을 중심으로 질환자(Case) 와 정상인(Control)을 구분함.

정상인 중에는 아직 질병이 발병하지 않았지만 이후에 질병이 발생할 수 있음.

예) 당뇨병의 진단기준을 공복혈당 126 mg/dl로 했을 때, 혈당 125와 127의 차이는?

예) 정상인으로 분류된 40세 남자, 10년 뒤에도 정상일까?

Association Study
개체 선별법

집단을 기초로 한 전향적 조사, Cohort 연구
→ 상대적 위험도 Relative Risk, RR

병원을 기초로 한 후향적 조사
→ 오즈비 Odds Ratio, OR

Association Study

데이터의 종류	그룹의 특성	통계분석 방법
범주형 Categorical	2 Groups	X^2 test
	2 Groups (small group)	Fisher's exact test
	2 Groups (adjustment)	Logistic regression
연속형 Continuous	2 Groups	t-Test
	3 or more Groups	ANOVA
	3 or more Groups (adjusment)	Multiple regression

Genetic Background

- Familial Aggregation of Diabetes (both parents-offspring 40%)
- Twin Studies (70-90%)
- Genetic Syndromes Associated with Diabetes

It is clear that T2D has a strong genetic component.

당뇨병에서의 유전자연구

Story of Pima Indians

( ) prevalence of diabetes

(54%)

Arizona Pima

(6.3%)

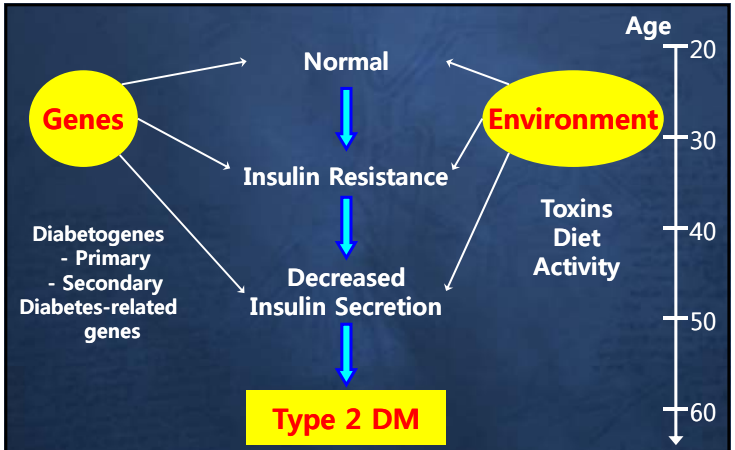
Mexican Pima

Are genes responsible for Type 2 Diabetes and Obesity?

Genetics of T2DM

Geneticist's Nightmare

Neel J. Diabetes mellitus: a geneticist's nightmare. In: Creutzfeldt W, Kobberling J, Neel JV, eds. The genetics of diabetes mellitus. Springer-Verlag, **1976**:1-11.



Natural History of Pre-type 2 Diabetes and the Impact of genes and the Environment in this Process.

Genetics of T2DM

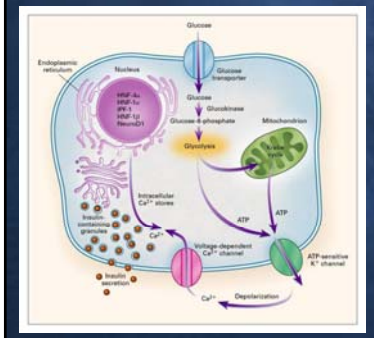
- Is genetically heterogeneous
- Is almost certainly polygenic
- Strong gene/gene and gene/environmental interactions play important roles in development of T2D

Common Gene, Common Disease and Complex Disease Phenotyping

당뇨병 발병에 관여하는 유전적인 원인은
질병발생 위험도가 그리 크지 않으면서
비교적 흔한 유전자들의 변이일 가능성이 크다.

Monogenic Causes

MODY
(Maturity Onset Diabetes of the Young)



Uncommon form of T2D
(<5% of all T2D)

Autosomal dominant inheritance

Early onset of hyperglycemia; < 25 years
Impairment in insulin secretion

caused by a mutation in a different gene that is directly involved with beta cell function

Slow onset of symptoms, Absence of obesity, No ketosis, No beta cell autoimmunity.

Family-Based Linkage Analysis

장점: **Linkage analysis exhibits its maximal power in identifying loci implicated in rare (MODY, Neonatal Diabetes, ...)**

단점: **Loses precision for common forms of the diseases (T2DM, ...)**

Polygenic Causes

연구방법

1. Candidate Gene Association Study
   1) Functional Candidate Gene
      eg. PPARG, KCNJ1
   2) Positional Candidate Gene
      eg. Calpain 10, *TCF7L2*
2. Genome Wide Association Study

Candidate-Gene Association Study

장점: Known Target Gene (may play a role in T2DM Pathogenesis)

단점: Lack of Consistency – replication issue
Lack of Power
Acceptance of low p-value threshold (0.05)
False Positive

Contributed to Pathogenesis of Diabetes
Successful target for anti diabetes medication
- PPARG – P12A, KCNJ11 – E23K

Candidate Gene Association Study

Positional Cloning

Calpain 10 Gene

Reported in a Mexican-American population
Not robustly replicated in other ethnic groups

Fine Mapping Linkage Analysis

TCF7L2

Chromosome 10 region showed linkage to T2D
Fine mapping – TCF7L2
5 SNPS and 1 tetranucleotide repeat polymorphism (DG10S478)
Replicated many ethnic groups ethnic groups

GWAS is facilitated by

Completion of Human Genome Project

Completion of International HapMap Project

Advance in Genotyping Technology

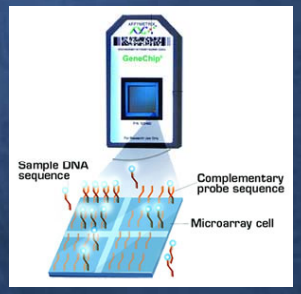
Advance in Computer Technology

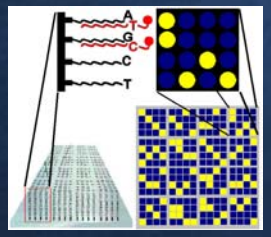
Genome Wide Association Studies

3.9 million SNPs in 270 DNA samples from 4 different ethnic groups

90 Yoruba individuals (30 parent-parent-offspring trios) from Ibadan, Nigeria (YRI)
90 individuals (30 trios) of European descent from Utah (CEU)
45 Han Chinese individuals from Beijing (CHB)
45 Japanese individuals from Tokyo (JPT)

Quantified by LD → Tagging SNP

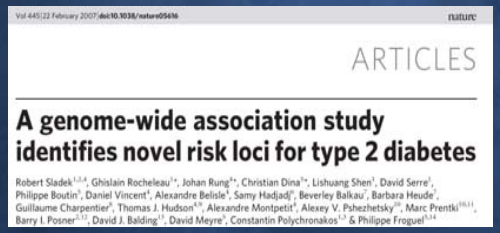




Illumina
HumanOmni5-Quad BeadChip Kit
4.3 million SNP loci

Affymetrix
Genome-Wide Human SNP Array 6.0
More than 906,600 SNPs

The First GWAS for T2D in 2007



nature

ARTICLES

A genome-wide association study identifies novel risk loci for type 2 diabetes

Robert Sladek, Ghislain Rocheleau, Johan Rung, Christian Dina, Lishuang Shen, David Serre, Philippe Boutin, Daniel Vincent, Alexandre Belisle, Samy Hadjadj, Beverley Balkau, Barbara Heude, Guillaume Charpentier, Thomas J. Hudson, Alexandre Montpetit, Alexey V. Pshezhetsky, Marc Prentki, Barry I. Posner, David J. Balding, David Meyre, Constantin Polychronakos & Philippe Froguel

French population
661 T2DM and 614 Control
392,935 SNPs
4 new loci and confirm *TCF7L2*

Sladek R, et al. *Nature 445: 881-885, 2007*

Successful GWAS Conditions

1. **Sufficient sample size (at least 1,000 each of cases and controls)**

2. **P-value < 5 x 10^{-8} (Genome wide significance)**

3. **Confirmation of association by independent replication studies**

McCarthy MI, et al. *Nature Reviews Genetics 9*:356-369, 2008

The First GWAS for T2D

SNP	Chromosome	Position (nucleotides)	Risk allele	Major allele	MAF (case)	MAF (ctrl)	Odds ratio (het)	Odds ratio (hom)	PAR	I_A	Stage 2 pMAX	Stage 2 pMAX (perm)	Stage 1 pMAX	Stage 1 pMAX (perm)	Nearest gene
rs7903146	10	114,748,339	T	C	0.406	0.293	1.65 ± 0.19	2.77 ± 0.50	0.28	1.0546	1.5×10^{-34}	$<1.0 \times 10^{-7}$	3.2×10^{-17}	$<3.3 \times 10^{-10}$	*TCF7L2*
rs13266634	8	118,253,964	C	C	0.254	0.301	1.18 ± 0.25	1.53 ± 0.31	0.24	1.0089	6.1×10^{-8}	5.0×10^{-7}	2.1×10^{-5}	1.8×10^{-5}	*SLC30A8*
rs1111875	10	94,452,862	G	G	0.358	0.402	1.19 ± 0.19	1.44 ± 0.24	0.19	1.0069	3.0×10^{-6}	7.4×10^{-5}	9.1×10^{-6}	7.3×10^{-5}	*HHEX*
rs7923837	10	94,471,897	G	G	0.335	0.377	1.22 ± 0.21	1.45 ± 0.25	0.20	1.0065	7.5×10^{-6}	2.2×10^{-5}	3.4×10^{-6}	2.5×10^{-6}	*HHEX*
rs7480010	11	42,203,294	G	A	0.336	0.301	1.14 ± 0.13	1.40 ± 0.25	0.08	1.0041	1.1×10^{-4}	2.9×10^{-4}	1.5×10^{-5}	1.2×10^{-5}	LOC387761
rs3740878	11	44,214,378	A	A	0.240	0.277	1.26 ± 0.29	1.46 ± 0.33	0.24	1.0046	1.2×10^{-4}	2.8×10^{-4}	1.8×10^{-4}	1.3×10^{-5}	*EXT2*
rs11037909	11	44,212,190	T	T	0.240	0.271	1.27 ± 0.30	1.47 ± 0.33	0.25	1.0045	1.8×10^{-4}	4.5×10^{-4}	1.8×10^{-5}	1.3×10^{-5}	*EXT2*
rs1113132	11	44,209,979	C	C	0.237	0.267	1.15 ± 0.27	1.36 ± 0.31	0.19	1.0044	3.3×10^{-4}	8.1×10^{-4}	3.7×10^{-6}	2.9×10^{-5}	*EXT2*

SLC30A8, HHEX, LOC387761, EXT2, TCF7L2

Sladek R, et al. *Nature 445: 881-885, 2007*

The Second GWAS for T2D

deCODE genetics, Iceland

nature genetics

A variant in *CDKAL1* influences insulin response and risk of type 2 diabetes

SLC30A8, HHEX, CDKAL1

Steinthorsdottir V, et al. *Nat Genet 39: 770-775, 2007*

FUSION

Finland-United States Investigation of NIDDM

A Genome-Wide Association Study of Type 2 Diabetes in Finns Detects Multiple Susceptibility Variants

Laura J. Scott,[1] Karen L. Mohlke,[2] Lori L. Bonnycastle,[3] Cristen J. Willer,[1] Yun Li,[1] William L. Duren,[1] Michael R. Erdos,[2] Heather M. Stringham,[1] Peter S. Chines,[3] Anne U. Jackson,[1] Ludmila Prokunina-Olsson,[2] Chia-Jen Ding,[1] Amy J. Swift,[3] Narisu Narisu,[3] Tianle Hu,[1] Randall Pruim,[4] Rui Xiao,[1] Xiao-Yi Li,[1] Karen N. Conneely,[1] Nancy L. Riebow,[1] Andrew G. Sprau,[1] Maurine Tong,[1] Peggy P. White,[1] Kurt N. Hetrick,[1] Michael W. Barnhart,[1] Craig W. Bark,[1] Janet L. Goldstein,[1] Lee Watkins,[1] Fang Xiang,[1] Jouko Saramies,[4] Thomas A. Buchanan,[7] Richard M. Watanabe,[8,9] Timo T. Valle,[10] Leena Kinnunen,[10,11] Gonçalo R. Abecasis,[2] Elizabeth W. Pugh,[1] Kimberly F. Doheny,[1] Richard N. Bergman,[9] Jaakko Tuomilehto,[10,11,12] Francis S. Collins,[2*] Michael Boehnke[2*]

SLC30A8, HHEX, CDKAL1, IGFBP2, CDKN2A/B, PPARG P12A, KCNJ11 E23K

Science 316: 1341-1345. 2007

WTTCCC/UKT2D

Wellcome Trust Case Control Consortium/ United Kingdom Type 2 Diabetes Genetics consortium

Replication of Genome-Wide Association Signals in UK Samples Reveals Risk Loci for Type 2 Diabetes

Eleftheria Zeggini,[1,2*] Michael N. Weedon,[3,4*] Cecilia M. Lindgren,[1,2*] Timothy M. Frayling,[3,4*] Katherine S. Elliott,[2] Hana Lango,[3,4] Nicholas J. Timpson,[2,5] John R. B. Perry,[3,4] Nigel W. Rayner,[1,2] Rachel M. Freathy,[3,4] Jeffrey C. Barrett,[2] Beverley Shields,[4] Andrew P. Morris,[2] Sian Ellard,[4,6] Christopher J. Groves,[1] Lorna W. Harries,[4] Jonathan L. Marchini,[7] Katharine R. Owen,[1] Beatrice Knight,[4] Lon R. Cardon,[2] Mark Walker,[8] Graham A. Hitman,[9] Andrew D. Morris,[10] Alex S. F. Doney,[10] The Wellcome Trust Case Control Consortium (WTCCC),† Mark I. McCarthy,[1,2‡§] Andrew T. Hattersley[3,4‡]

SLC30A8, HHEX, CDKAL1, IGFBP2, CDKN2A/B, PPARG P12A, KCNJ11 E23K

Science 316: 1336-1341. 2007

DGI

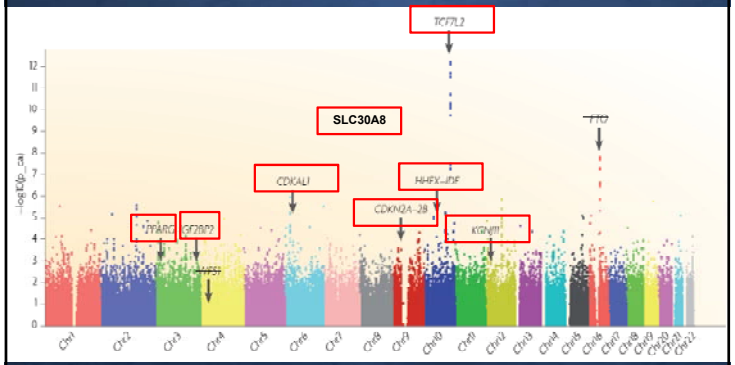
Diabetes Genetics Initiative of Broad Institute of Harvard and MIT, Lund University, and Novartis Institutes of BioMedical Research

Genome-Wide Association Analysis Identifies Loci for Type 2 Diabetes and Triglyceride Levels

Diabetes Genetics Initiative of Broad Institute of Harvard and MIT, Lund University, and Novartis Institutes for BioMedical Research*†

SLC30A8, HHEX, CDKAL1, IGFBP2, CDKN2A/B, PPARG P12A, KCNJ11 E23K, FTO

Science 316: 1331-1336. 2007



Association statistics from one of the 8 T2DM GWAS

Meta-analysis – DIAGRAM Plus

DIAGRAM + (combined with diagram cohort)
2,426,886 SNPs
2,000 subjects of European origin

Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis

34,412 cases and 59,925 controls
12 new T2D association signals

Nat Genet 42(7):579-589, 2010

Meta-analysis of Initial GWASs

Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes

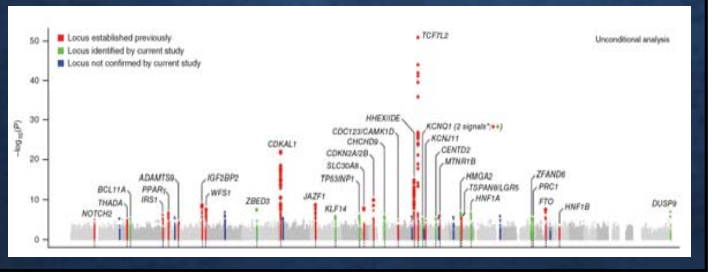
the Diabetes Genetics Replication And Meta-analysis (DIAGRAM) Consortium

5 additional loci:
JAZF1, CDC123-CAMK1D, TSPAN8-LGR5, THADA, ADAMTS9, NOTCH2

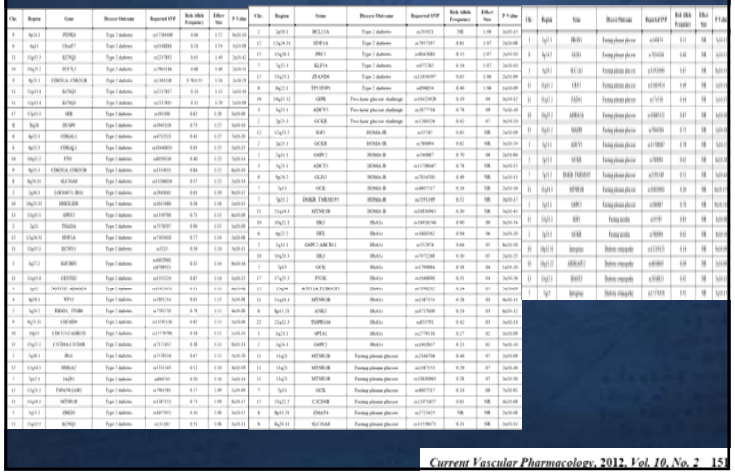
4,549 cases and 5,579 controls
2.2 million SNPs

Nat Genet 40: 638-645, 2008

BCL11A, ZBED3, KLF14, TP53INP1, CHCHD9, KCNQ1, CENTD2, HMGA2, HNF1A, ZFAND6, PRC1, DUSP9 (X-chromosomal association)



Identified Genome-Wide Associations



Current Vascular Pharmacology, 2012, Vol. 10, No. 2 151

GWAS for Glycemic Traits

A polymorphism within the *G6PC2 gene is associated with fasting plasma glucose* levels. ***Science 320: 1085-1088, 2008***

Variants in *MTNR1B influence fasting glucose* levels. ***Nat Genet 41: 77-81.*** **2009**

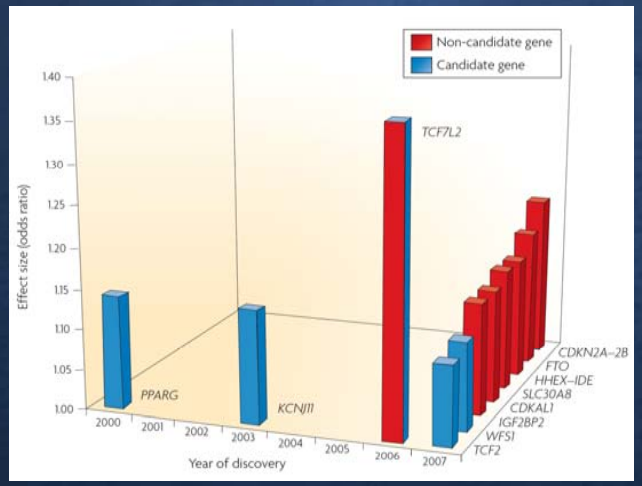
A variant near *MTNR1B is associated with increased fasting* plasma glucose levels and type 2 diabetes risk. ***Nat Genet 41: 89-94. 2009***

Variations in the *G6PC2/ABCB11 genomic region are associated with* fasting glucose levels. ***J Clin Invest 118: 2620-2628.*** **2008**

A common haplotype of the glucokinase gene alters fasting glucose and birth weight: association in six studies and population-genetics analyses. ***Am J Hum Genet 79:*** **991-1001. 2006**

New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. ***Nat Genet 42: 105-116.*** **2010**

Genetic variation in *GIPR influences* the glucose and insulin responses to an oral glucose challenge. ***Nat Genet 42: 142-148. 2010***



Meta-Analyses of Glucose and Insulin-related traits Consortium (MAGIC)

Examined 21 GWAS to identify loci associating with fasting glucose, fasting insulin, HOMA-β, and HOMA-IR

76,558 individuals from 34 additional cohorts,

9 new loci in or near *ADCY5, MADD, CRY2, ADRA2A, FADS1, PROX1, SLC2A2, GLIS3, and C2CD4B were found to* be associated with fasting glucose.

Nat Genet 42: 105-116. **2010**

GWAS for Glycemic Traits

Fasting glucose
G6PC2, MTNR1B, GCK, *ADCY5, MADD, CRY2, ADRA2A, FADS1, PROX1, SLC2A2, GLIS3, and C2CD4B*

Fasting insulin and HOMA-IR.
GCKR, IGF1

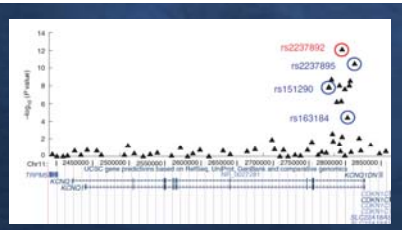
2-hour postprandial glucose
GIPR, ADCY5, GCKR, VPS13C, TCF7L2

Common GWAS genes of T2D and Glycemic traits
MTNR1B, GCK, ADCY5, PROX1, DGKB-TMEM195, GCKR

Nat Genet 42: 105-116. 2010

GWAS for T2D in Japanese #2

Variants in *KCNQ1* are associated with susceptibility to type 2 diabetes mellitus



187 T2D and 752 controls

Nat Genet 40: 1092-1097, 2008

GWAS for T2D in Japanese #1

SNPs in *KCNQ1* are associated with susceptibility to type 2 diabetes in East Asian and European populations

KCNQ1 (potassium voltage-gated channel, KQT-like subfamily, member 1) to be a strong candidate for conferring susceptibility to type 2 diabetes

194 T2D and 1,558 controls

Nat Genet 40: 1098-1102, 2008

GWAS for T2D in Japanese #3

A genome-wide association study in the Japanese population identifies susceptibility loci for type 2 diabetes at *UBE2E2* and *C2CD4A-C2CD4B*

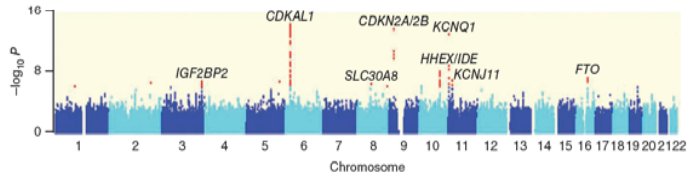
4,470 T2D and 3,071controls
459,359 SNPs
stage 1, 4,470 cases and 3,071 controls
stage 2, 2,886 cases and 3,087 controls
stage 3, 3,622 cases and 2,356 controls

UBE2E2 is not associated with T2D in Europeans

Nat Genet 40: 864-869, 2010

GWAS for T2D in KOREA ?

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



Meta-analysis of 8 T2D GWAS (6,952 T2D, 11,865 controls) with a stage 2 *in silico replication* analysis (5,843 T2D, 4,574 controls) and a stage 3 *de novo* replication analysis (12,284 T2D, 13,172 controls).

8 new T2D loci reaching GW significance *GLIS3, PEPD, FITM2-R3HDML-HNF4A, KCNK16, MAEA, GCC1-PAX4, PSMD6 and ZFAND3*.

Nat Genet 40: 864-869, 2011

Current Limitations for GWAS

1. Considerable number of uncaptured SNPs
   No. of SNP: 2,365만개
   SNP Chip 4백만

2. GWAS p-value may produce type 2 errors
   (false negative results)

3. Low frequency (MAF<1%) risk variants with large effects could be missed

What have GWAS brought about so far?

1. Identified T2D loci are associated more frequently with beta cell function rather than insulin resistance
   (Only GCKR, PPARG, FTO, KLF14 – associated with HOMA-IR)

2. Missing Heritability
   GWAS explain only 10%(-20%) of the known heritability in twin study.

3. Translation of T2D genetics into clinical Practice
   3-1. Disease Prediction and Prevention
   3-2. identifying novel therapeutic targets

Endo J 58: 723-739, 2011