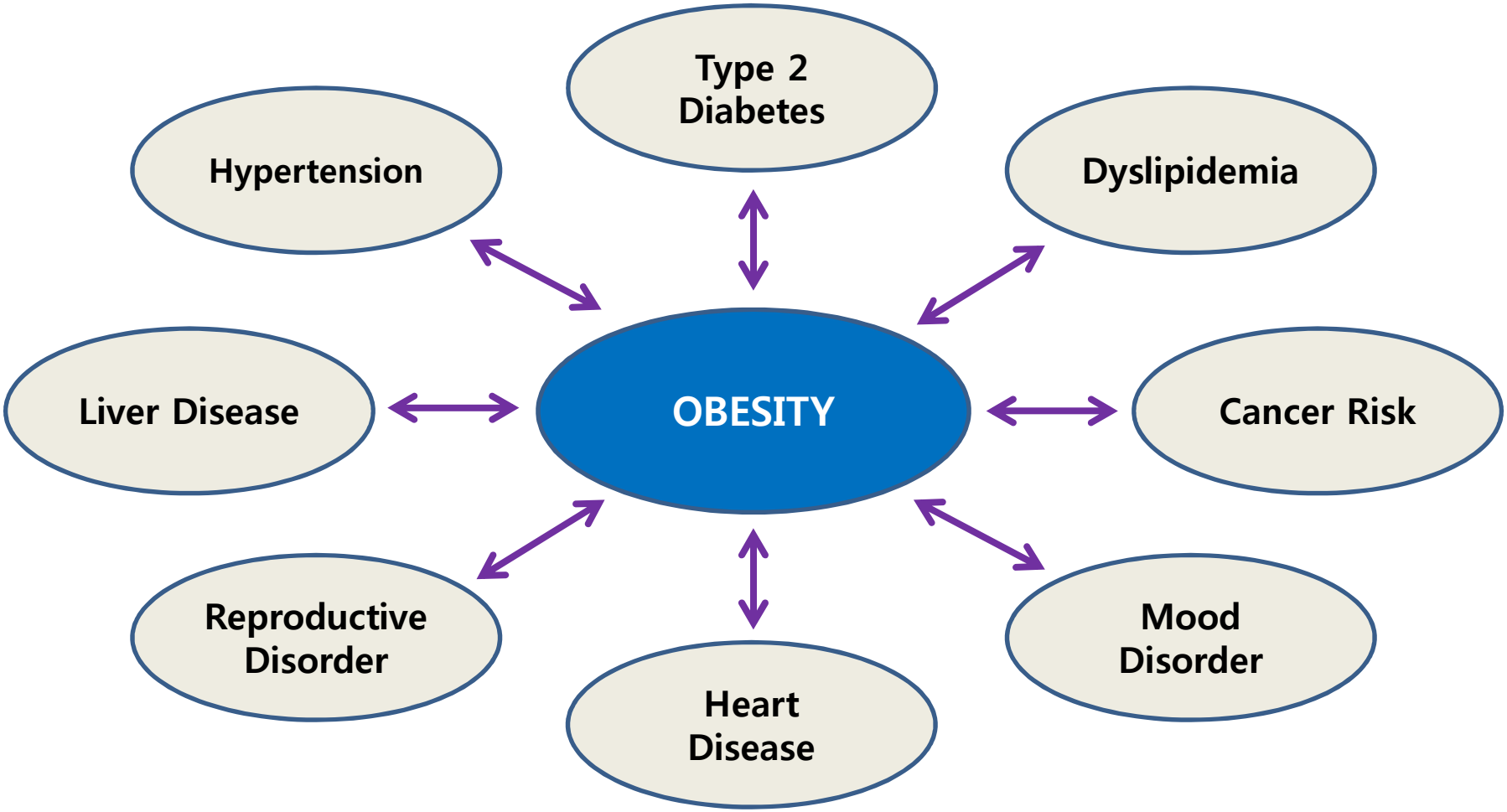
Identification of low-frequency variants influencing body mass index

Yoon Shin Cho

Hallym University

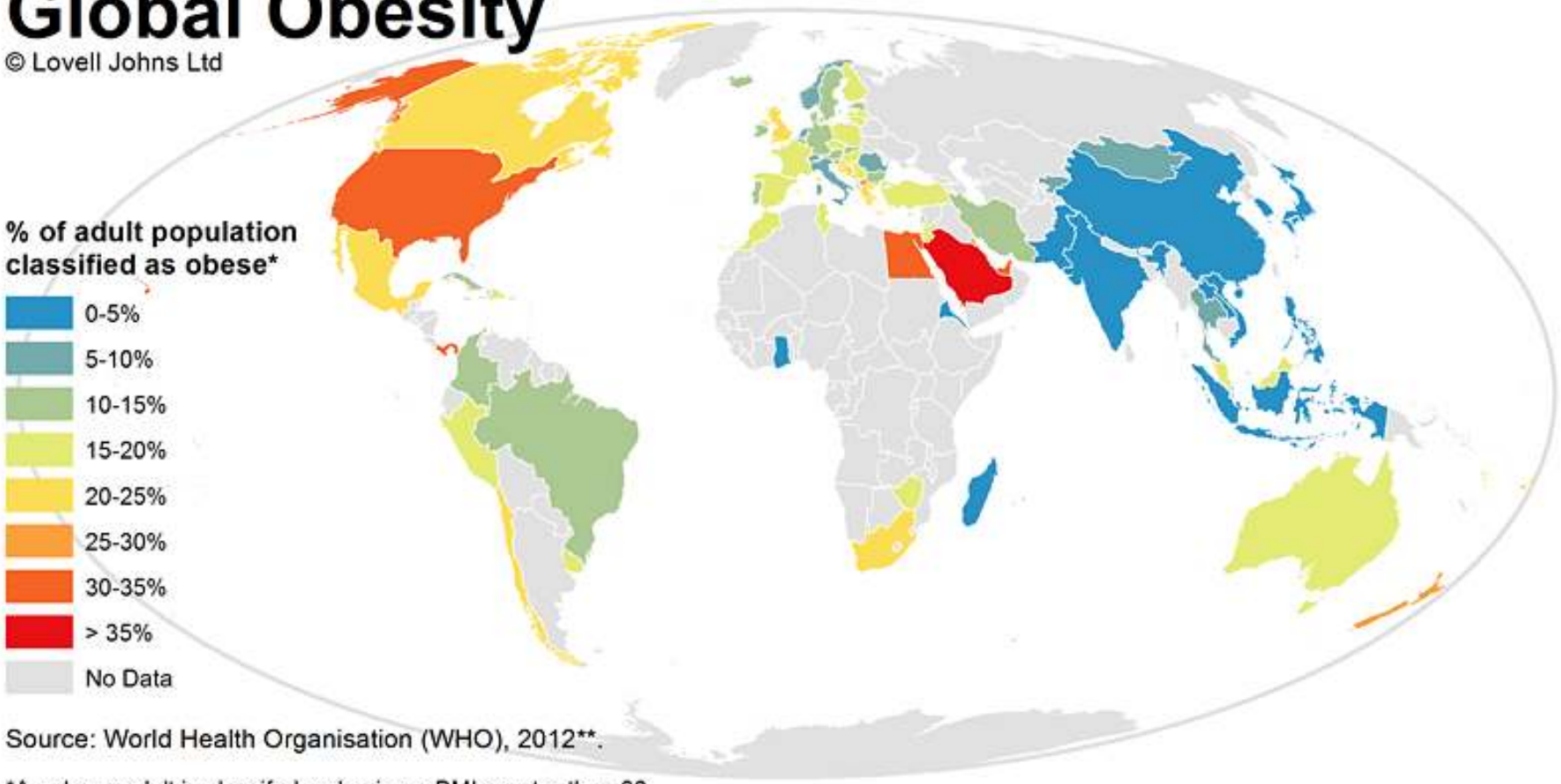
Obesity is a medical condition in which excess body fat has accumulated to the extent that it may have a negative effect on health, leading to reduced life expectancy and/or increased health problems.



✓ **BMI (Body Mass Index)** = mass (kg) / (height (m))^2

Category	BMI range – kg/m^2	
	WHO	Korea
Very severely underweight	< 15	
Severely underweight	15.0 ~ 16.0	
Underweight	16.0 ~ 18.5	< 18.5
Normal (healthy weight)	18.5 ~ 24.9	18.5 ~ 22.9
Overweight	25 ~ 29.9	≥ 23
Obese Class I (Moderately obese)	30 ~ 34.9	25 ~ 29.9
Obese Class II (Severely obese)	35 ~ 39.9	30 ~ 34.9
Obese Class III (Very severely obese)	> 40	> 35

Global Obesity

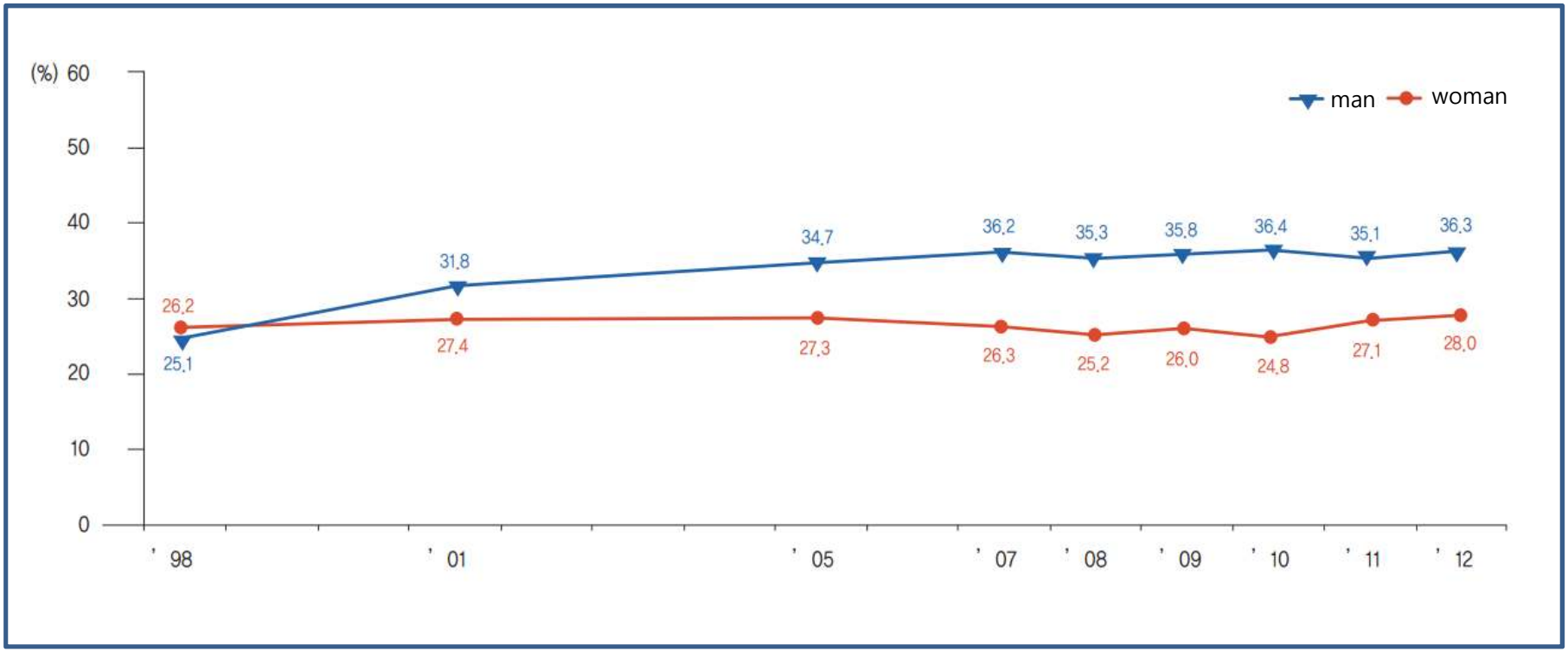


Source: World Health Organisation (WHO), 2012**.

*An obese adult is classifed as having a BMI greater than 30.
**The map uses the latest available data which varies in year of data collection.

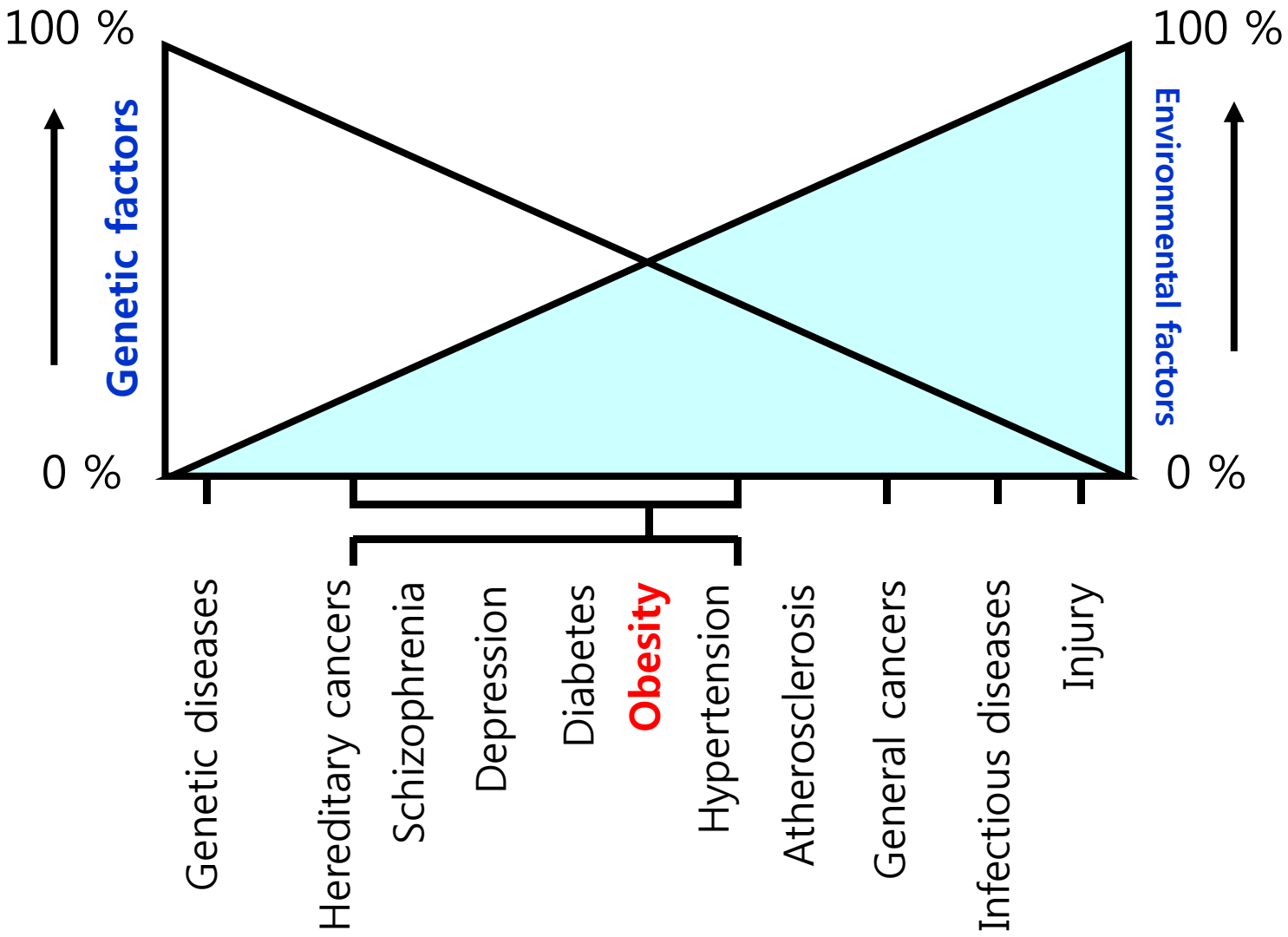
Adult obesity prevalence in Korea

- BMI ≥ 25 kg/m^2
- Age ≥ 19 years

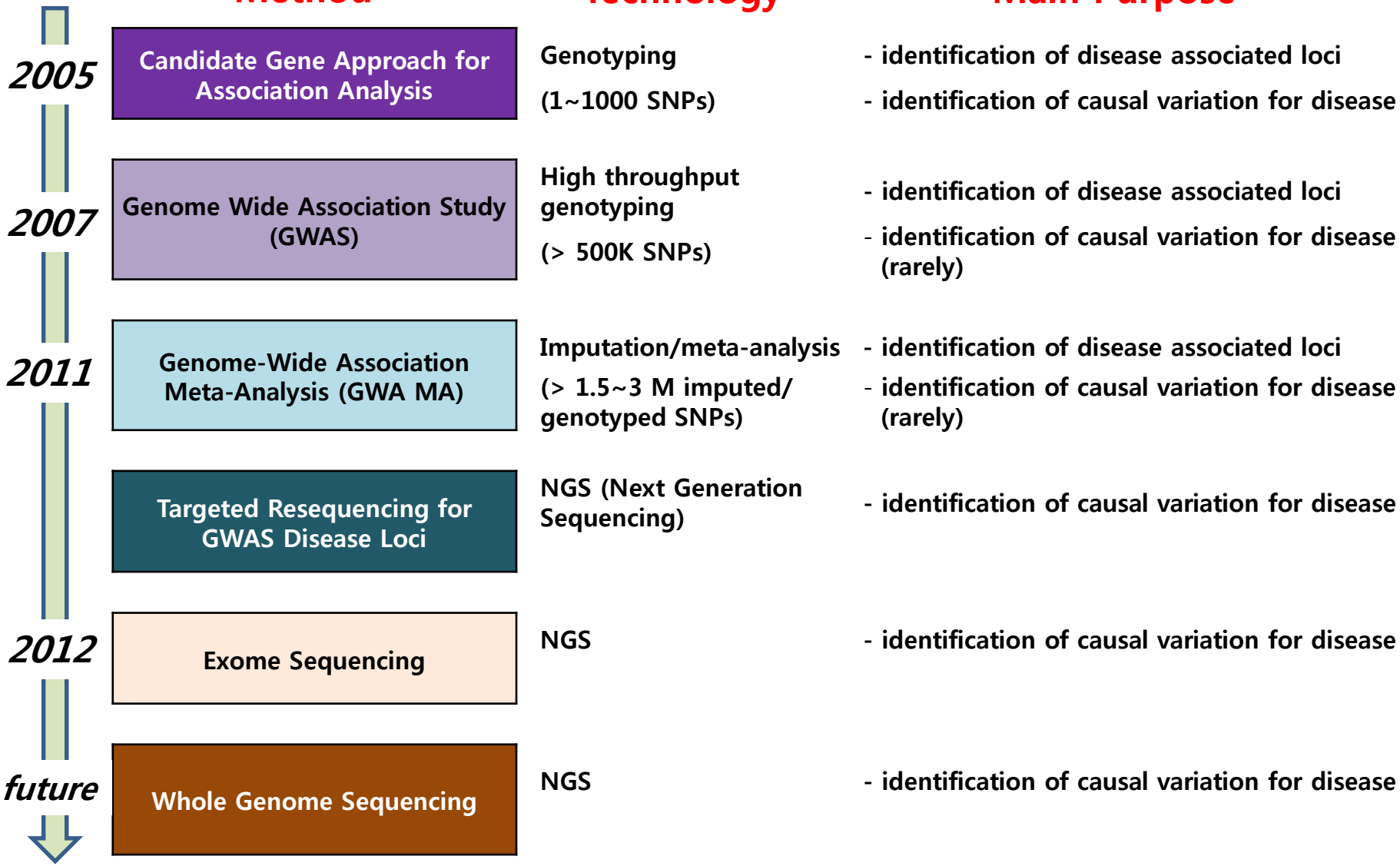


Source : Korean national health and nutrition examination survey 2012

Risk factors for obesity

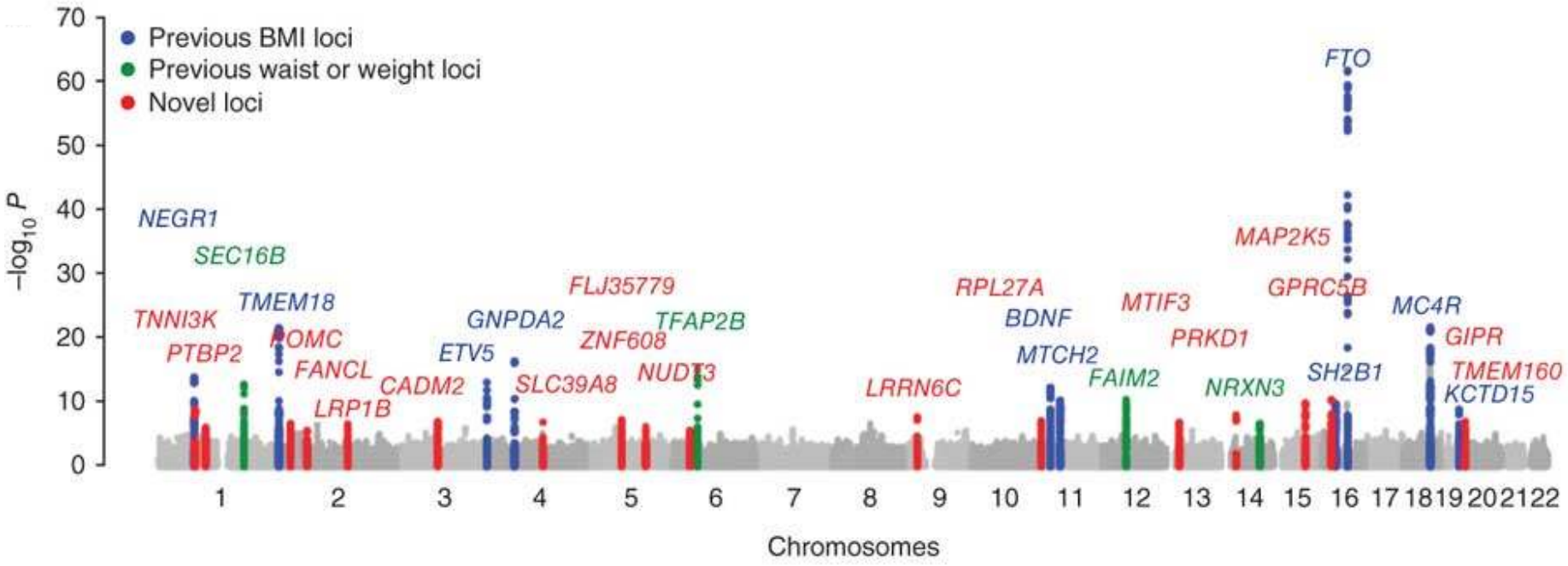


Methodological transition of genomics for complex traits



GWAS for BMI

- GIANT study : *Nat Genet* (2010)
  → 14 known loci + 18 new loci
  → Explain < 2% of the variation in BMI



- AGEN study : *Nat Genet* (2012)
  → 7 known loci + 3 new loci
  → Explain ~0.87% of the variation in BMI

LETTERS

nature genetics

Meta-analysis identifies common variants associated with body mass index in east Asians

Wanqing Wen[1–3,53], Yoon-Shin Cho[4,5,53], Wei Zheng[1–3,53], Rajkumar Dorajoo[6,7,53], Norihiro Kato[8,53], Lu Qi[9,53], Chien-Hsiun Chen[10,11,53], Ryan J Delahanty[1–3], Yukinori Okada[12,13], Yasuharu Tabara[14], Dongfeng Gu[15], Dingliang Zhu[16–19], Christopher A Haiman[20], Zengnan Mo[21], Yu-Tang Gao[22], Seang-Mei Saw[23], Min-Jin Go[4], Fumihiko Takeuchi[8], Li-Ching Chang[10], Yoshihiro Kokubo[24], Jun Liang[25], Mei Hao[26], Loïc Le Marchand[27], Yi Zhang[16–18], Yanling Hu[28], Tien-Yin Wong[29–31], Jirong Long[1–3], Bok-Ghee Han[4], Michiaki Kubo[32], Ken Yamamoto[33], Mei-Hsin Su[10], Tetsuro Miki[34], Brian E Henderson[20], Huaidong Song[35], Aihua Tan[36], Jiang He[26], Daniel P-K Ng[23], Qiuyin Cai[1–3], Tatsuhiko Tsunoda[37], Fuu-Jen Tsai[11], Naoharu Iwai[38], Gary K Chen[20], Jiajun Shi[1–3], Jianfeng Xu[39], Xueling Sim[40], Yong-Bing Xiang[22], Shiro Maeda[41], Rick T H Ong[6,42], Chun Li[43], Yusuke Nakamura[44], Tin Aung[29,30], Naoyuki Kamatani[12], Jian-Jun Liu[6], Wei Lu[45], Mitsuhiro Yokota[46], Mark Seielstad[6,47], Cathy S J Fann[10], The Genetic Investigation of ANthropometric Traits (GIANT) Consortium[48], Jer-Yuarn Wu[10,11,54], Jong-Young Lee[4,54], Frank B Hu[9,49,54], Toshihiro Tanaka[50,54], E Shyong Tai[23,51,52,54] & Xiao-Ou Shu[1–3,54]

- 45 GWAS loci for BMI have been identified (by 2012)

Limitations in GWAS

2) Unexplained heritability

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

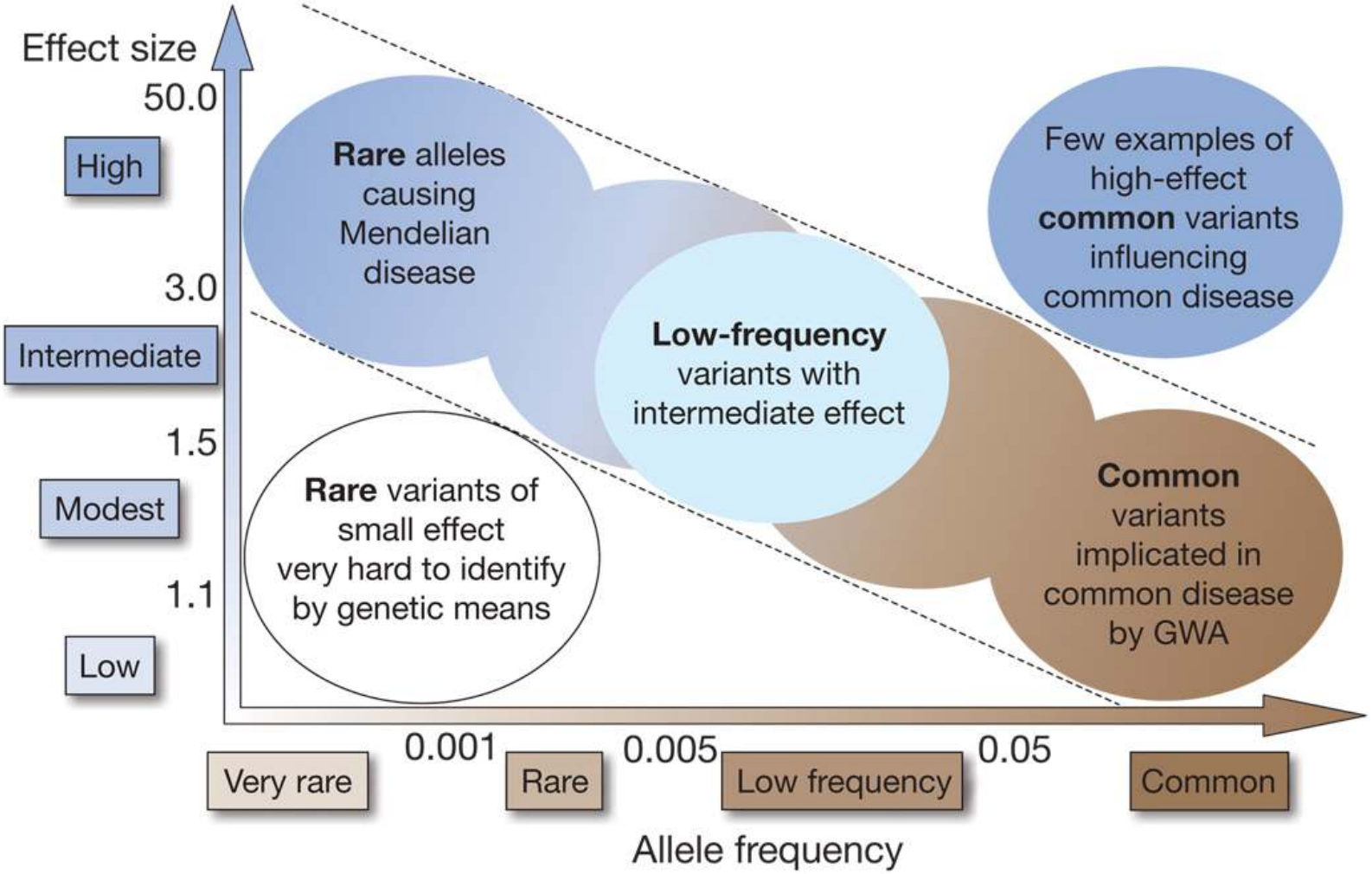
HDL, high-density lipoprotein; LDL, low-density lipoprotein.
* Fraction of heritability explained is calculated by dividing the phenotypic variance explained by variants at loci identified by GWAS by the total heritability as inferred from epidemiological parameters.

(Lander *Nature,* 2011)

※ BMI : The 32 established SNPs explain < 2% of heritability

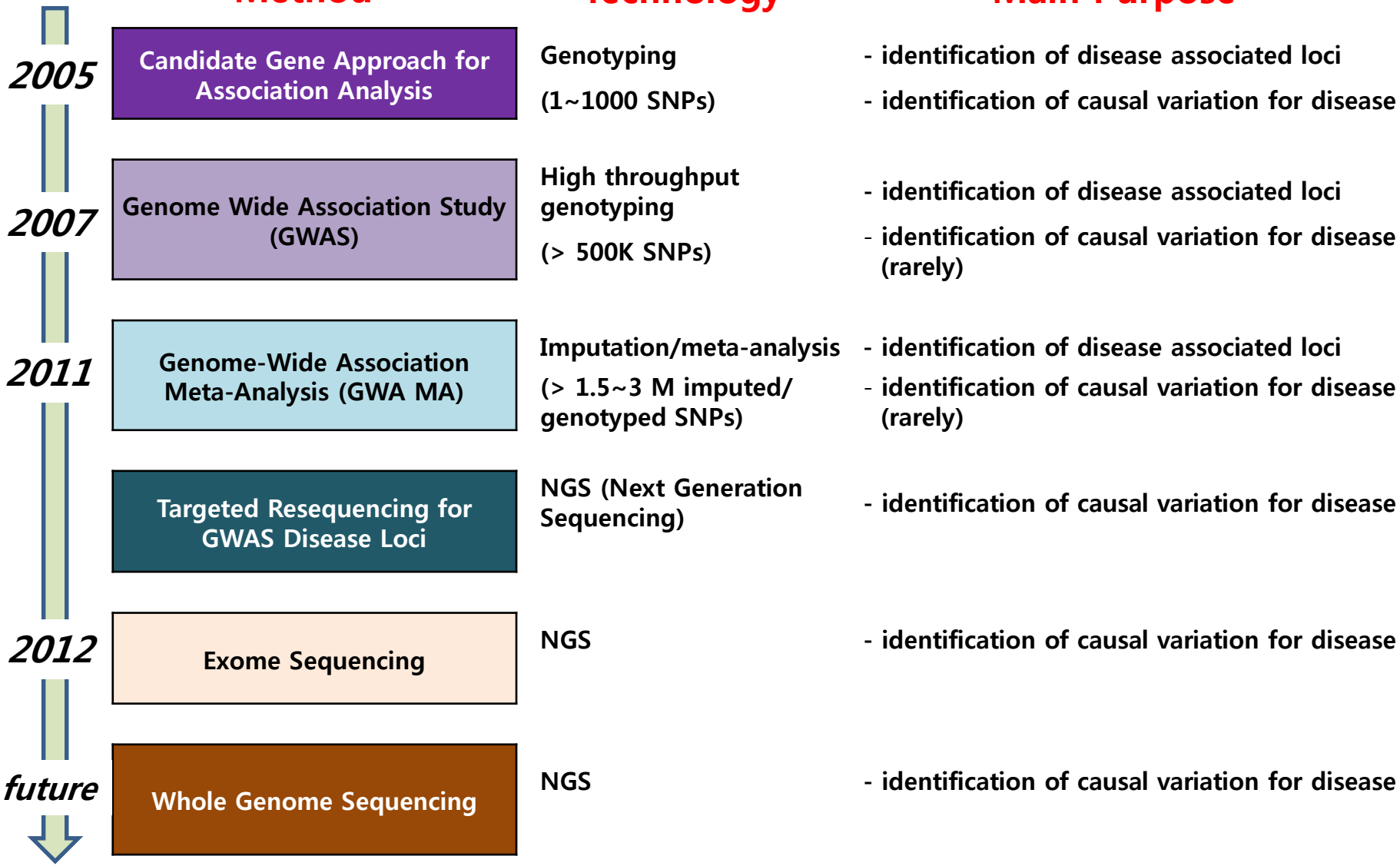
2) Most GWA genes reported are not necessarily causal

Allele frequency and effect sizes for genetic factors



(Manolio et al. *Nature,* 2009)

Methodological transition of genomics for complex traits



Low frequency and rare variants

- ✓ **Low allele frequency** : low frequency variant (1% ~ 5%)

  rare variant (usually less than 1%)

- ✓ **Low power** : for most analyses, due to less variation of observations
- ✓ **High false positive rate** : for some model-based analyses, due to sparse distribution of data, unstable/biased parameter estimation and inflated p-value.
- ✓ **How do we analyze them?**

  → *Still an open area of research*

Identification of BMI loci from exome sequencing data in Korea

✓ General specification of exome sequencing experiment

- 918 samples (620 DM & 298 non-DM samples) from SNUH
  (directed by Dr. Kyong Soo Park)
- Exon capture kit : Agilent SureSelect Human All Exon V4+UTR (70 Mb)
- Sequencing platform : Illumina Hiseq 2000
- Variant calling method : GATK UnifiedGenotyper
- Average read depth : ~ 200X (read length : 101 bp)

Clinical characteristics of study samples

Mean (±SD)

Variables	DM samples (N = 620)	Non-DM sample (N = 298)	Total samples (N = 918)
Age	56.5* (±9.1)	67.0* (±7.4)	59.9 (±9.9)
Sex (M/F)	286/334	134/164	420/498
BMI	24.4* (±2.8)	23.6* (±3.1)	24.2 (±2.9)
Height	161* (±8.6)	158.1* (±8.5)	160.1 (±8.7)
Weight	63.5* (±9.5)	59.2* (±9.5)	62.1 (±9.7)
Waist	86.0* (±7.7)	83.2* (±8.3)	85.3 (±8.0)
Hip	95.4* (±5.6)	93.3* (±6.6)	94.9 (±5.9)
WHR	0.91 (±0.07)	0.89 (±0.08)	0.90 (±0.06)

* $P < 0.01$

Analysis methods for low frequency (or rare) variants

✓ **QT analyses for BMI**

(1) Single variant test by EMMAX test

(Efficient Mixed-Model Association eXpedited)

- based on linear mixed model to correct for relatedness
- takes advantage of the fact that each locus explains only a small fraction of complex traits

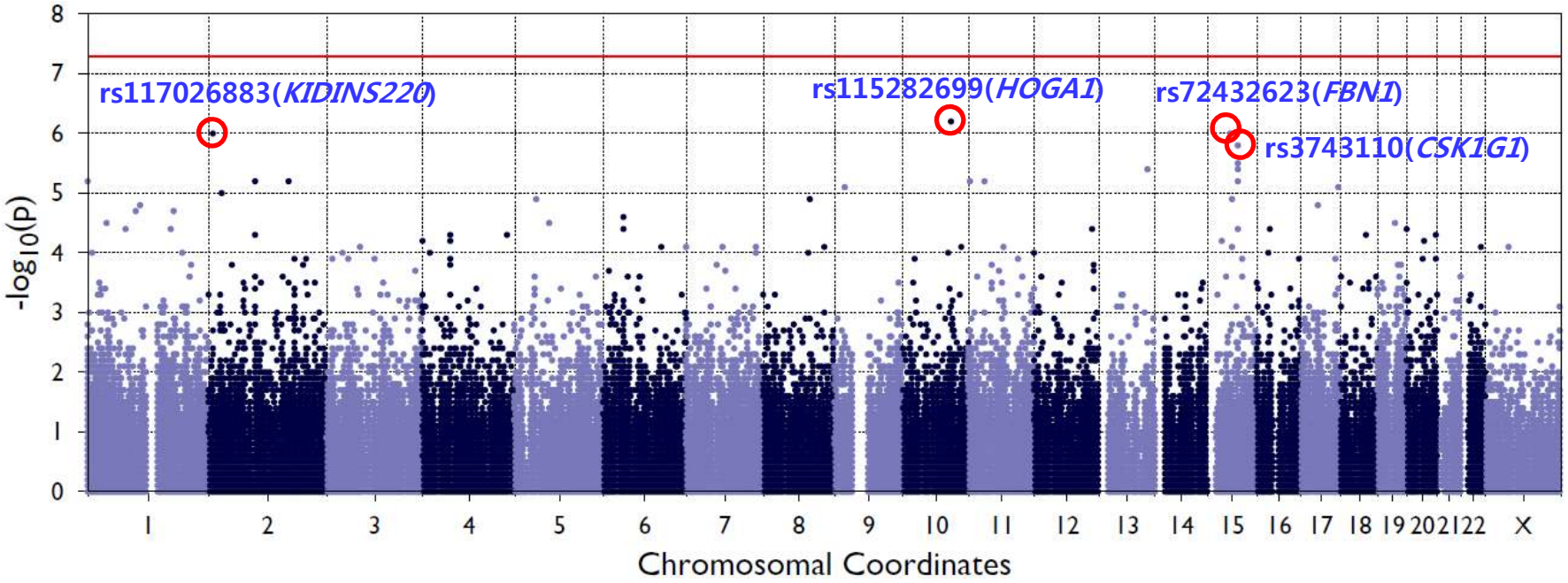
(2) Gene-based test by SKAT-O test (Sequence Kernel Association Test)

- SNP-set (e.g., a gene or a region) level test for association between a set of rare (or common) variants and dichotomous or quantitative phenotypes
- has been shown to perform well for rare causal variants

1. Analysis in all 918 samples

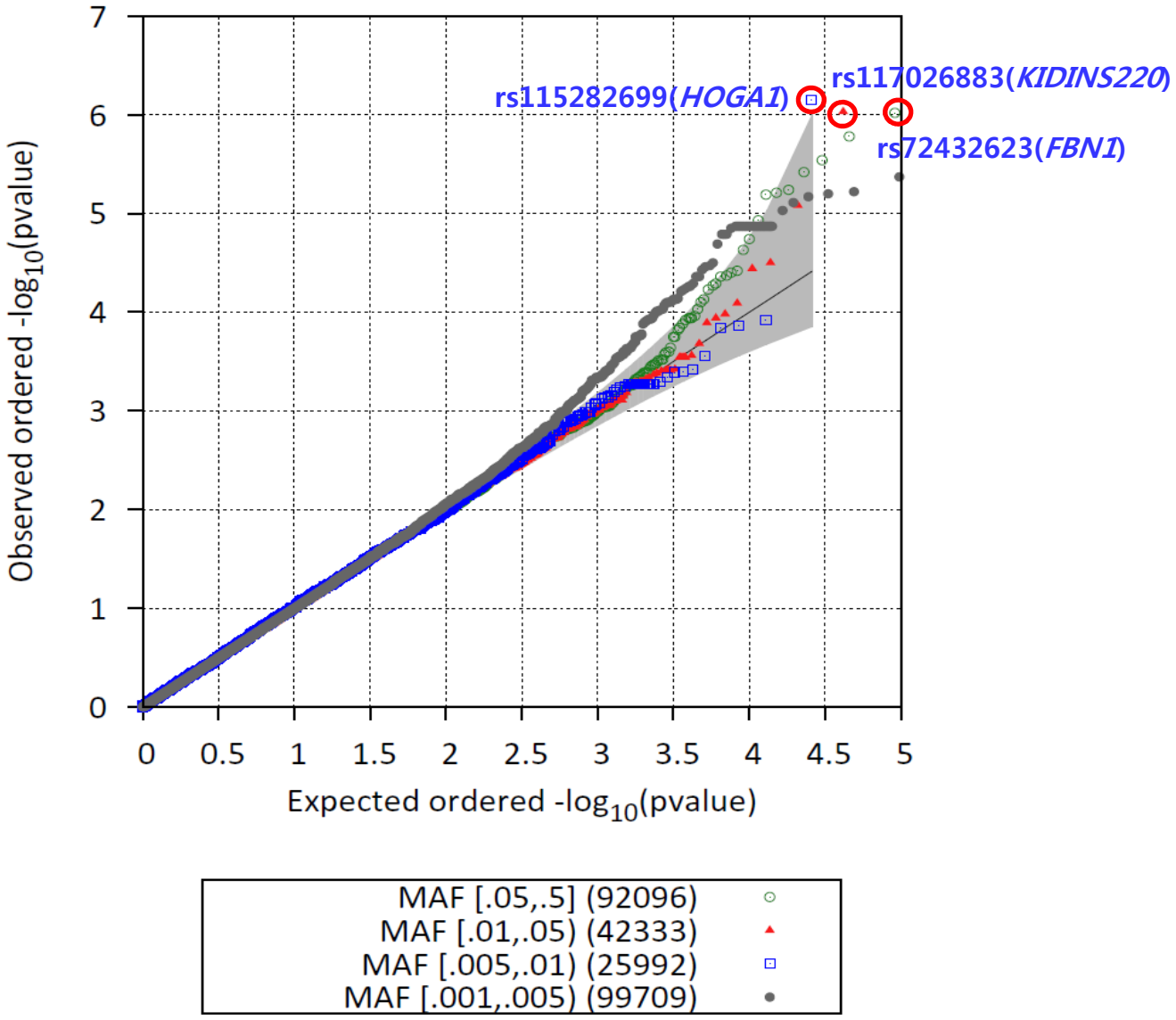
1-A. Single variant test

Manhattan plot of single variant test results for BMI association



※ Analysis was adjusted for DM, AGE & SEX.
※ Analysis reported total 641,489 variants (genome-wide significant *P*-value<~7.79E-08)
(number of MAF > 0.05 : 92,101 variants)

QQ plot of single variant test results for BMI association

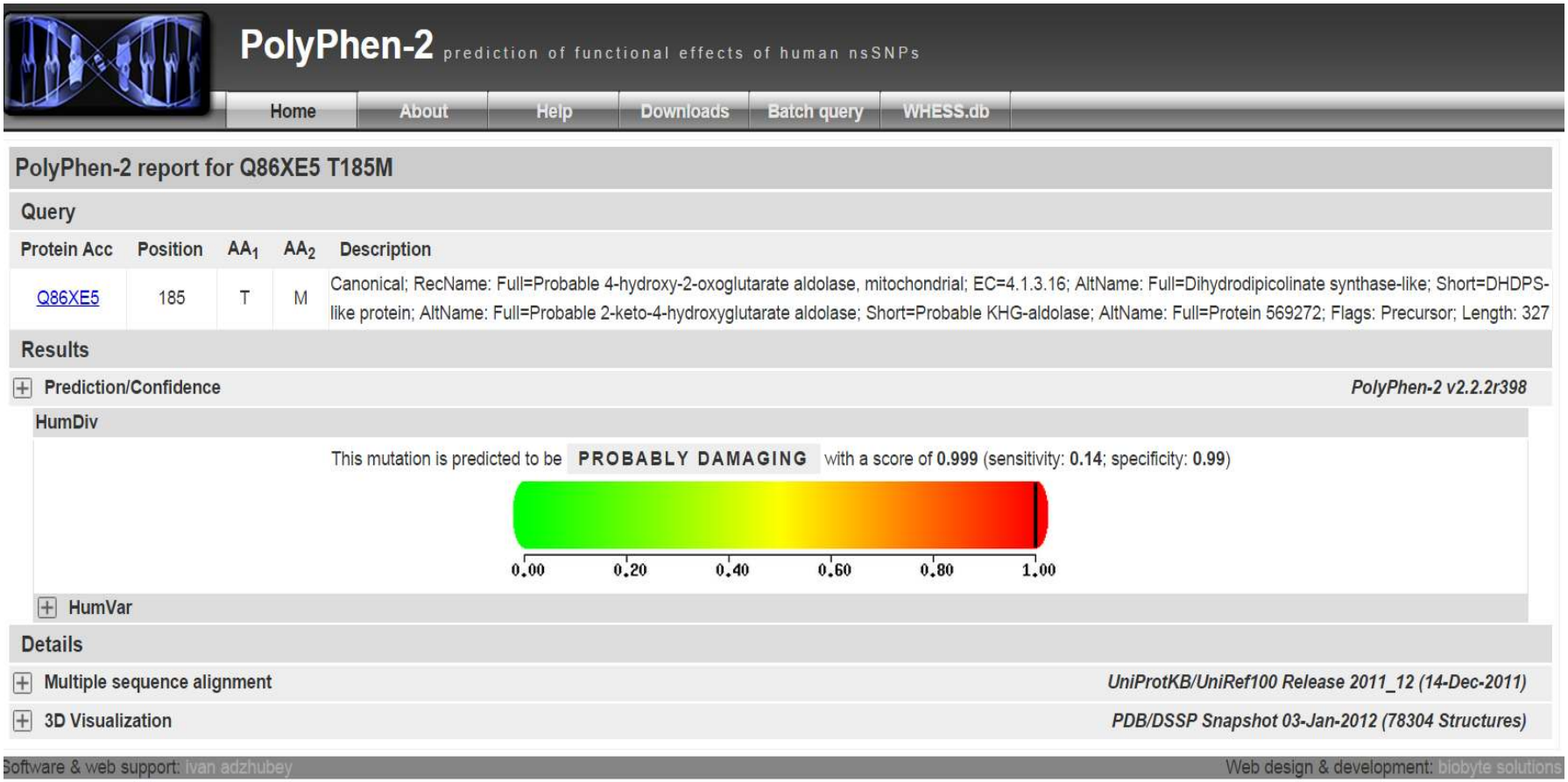


Top 10 loci associated with BMI identified by single variant test

SNP	Gene	Function	Chr	Position	Minor/ Major	MAF	Beta±SE	P-value
rs115282699	***HOGA1***	T185M	10	99,359,522	T/C	0.006	4.353±0.872	7.08E-07
rs117026883	***KIDINS220***	3′-UTR	2	8,870,019	G/T	0.025	2.170±0.439	9.29E-07
rs72432623	*FBN1*	intron	15	48,755,450	-/A	0.418	0.524±0.106	9.50E-07
rs3743110	*CSNK1G1*	3′-UTR	15	64,462,463	G/A	0.053	1.444±0.299	1.64E-06
-	***ZIC5***		13	100,623,754	A/G	0.002	7.699±1.663	4.21E-06
-	*COBLL1*		2	165,694,012	-/A	0.448	0.452±0.099	5.69E-06
rs72555790	***RRM1***	intron	11	4,144,647	T/G	0.004	4.670±1.025	5.92E-06
rs61742095	*FER1L5*	V1662V	2	97,366,088	C/T	0.157	0.837±0.184	6.07E-06
rs200016628	***KLHL17***	G419G	1	898,717	T/C	0.004	4.646±1.023	6.30E-06
rs7167612	*TRIP4*	E302E	15	64,737,225	G/A	0.050	-1.406±0.310	6.42E-06

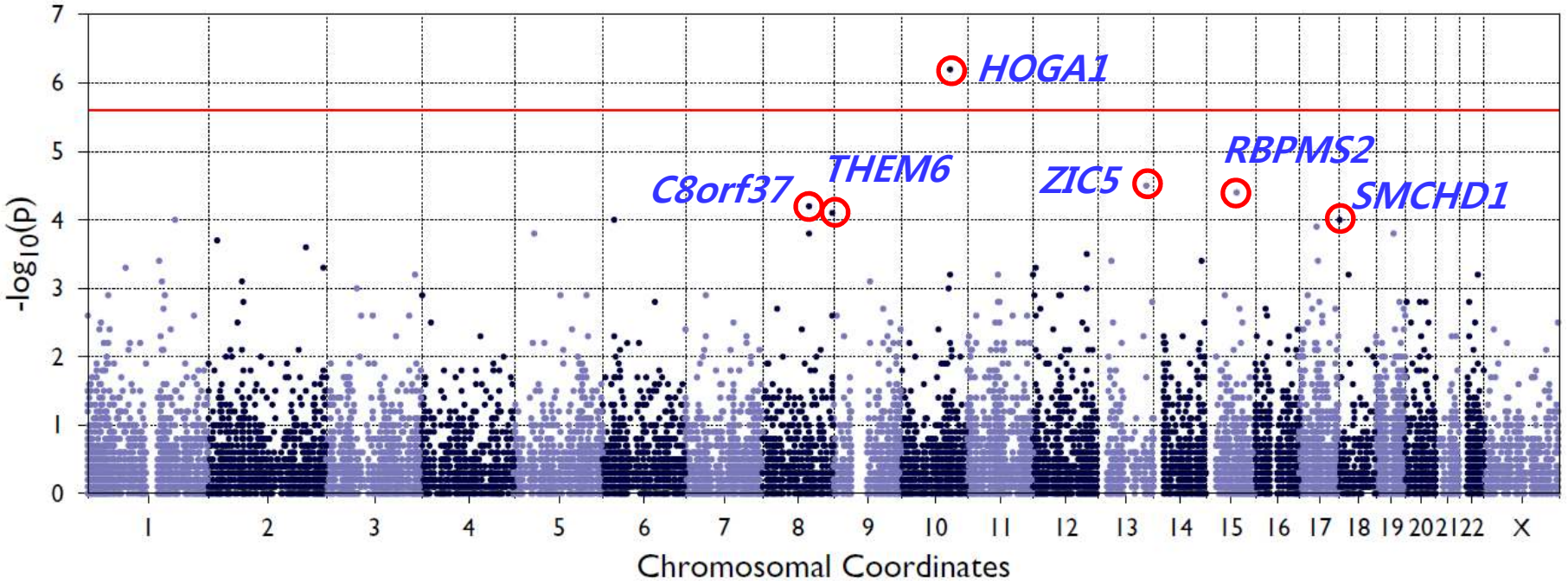
In silico functional test of HOGA1 (T185M) by PolyPhen → rs115282699

(http://genetics.bwh.harvard.edu/pph2/index.shtml)



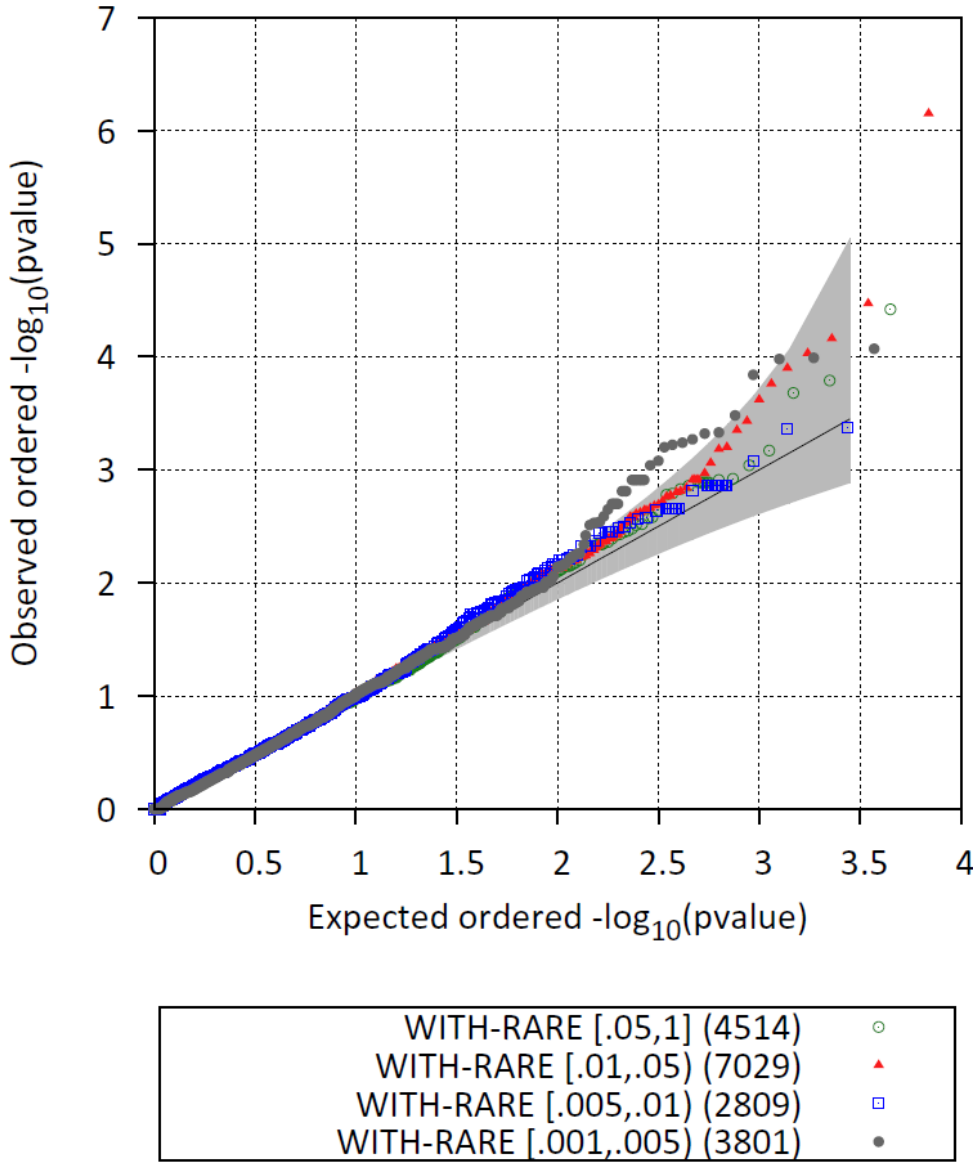
1-B. Gene-based test

Manhattan plot of gene-based test results for BMI association



※ Analysis was adjusted for DM, AGE & SEX.
※ Analysis reported total 18,379 loci (genome-wide significant *P*-value<~2.72E-06)

QQ plot of gene-based test results for BMI association



Top 10 genes associated with BMI identified by gene-based test

Gene	Position	# of variants/ gene set	P-value
HOGA1	chr10:99344508-99371274	6	7.07E-07
ZIC5	chr13:100617663-100623754	5	3.31E-05
RBPMS2	chr15:65041622-65041622	1	3.73E-05
C8orf37	chr8:96264434-96281381	5	6.80E-05
THEM6	chr8:143808820-143809059	3	8.39E-05
SMCHD1	chr18:2666885-2772321	15	9.32E-05
RGS16	chr1:182569431-182571141	3	1.02E-04
HIST1H2AD	chr6:26199306-26199396	3	1.04E-04
MLLT6	chr17:36864132-36878964	8	1.26E-04
NDUFAF6	chr8:96037261-96047772	4	1.42E-04

The functional relevance of identified BMI genes → not clear yet

2. Analysis in 298 non-DM samples

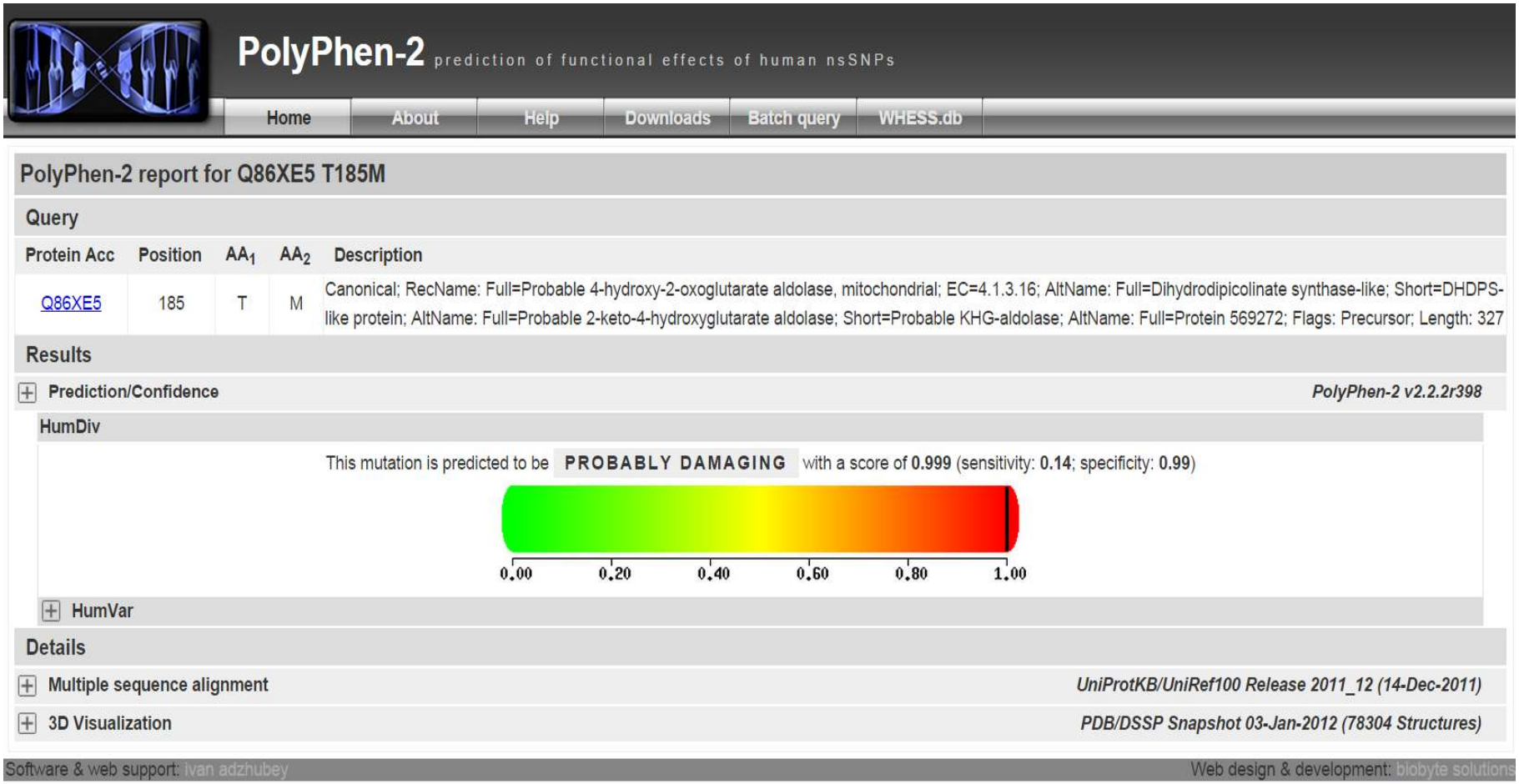
2-A. Single variant test (Top 10 loci associated with BMI)

SNP	Gene	Function	Chr	Position	Minor/ Major	MAF	Beta±SE	P-value
rs143364138	*HIST1H3A*	P67L	6	26,020,917	T/C	0.00503	7.999±1.690	3.43E-06
rs111991092	*PPP1R16B*	D422D	20	37,546,997	T/C	0.00503	7.940±1.690	4.06E-06
rs72555790	*RRM1*	intron	11	4,144,647	T/G	0.00839	5.981±1.318	8.32E-06
rs373470890	*IL17C*	I135I	16	88,706,291	T/C	0.01858	3.933±0.905	1.91E-05
rs146005389	*ATP5O*	ds	21	35,267,969	A/G	0.00839	5.650±1.319	2.51E-05
rs149501146	*GBA2*	P465P	9	35,740,009	T/C	0.00503	7.289±1.702	2.51E-05
-	*TRAPPC2L*	Exon	16	88,923,565	A/G	0.00671	-6.292±1.477	2.74E-05
-	*TRIO*	intron	5	14,488,594	C/CCTT	0.00503	7.176±1.707	3.47E-05
rs4648095	*NFKB1*	intron	4	103,527,876	C/T	0.05872	2.227±0.533	3.83E-05
-	*CYP46A1*	Exon	14	100,193,034	T/G	0.00507	7.144±1.711	3.93E-05

※ Analysis was adjusted for AGE & SEX.

In silico functional test of HIST1H3A (P67L) by PolyPhen → rs143364138

(http://genetics.bwh.harvard.edu/pph2/index.shtml)



2-B. Gene-based test (Top 10 genes associated with BMI)

Gene	Position	# of variants/ gene set	P-value
CCIN	chr9:36169539-36170940	8	2.94E-06
HIST1H3A	chr6:26020883-26021090	4	3.28E-05
ZDHHC16	chr10:99211893-99215808	2	3.29E-05
TRAPPC2L	chr16:88923565-88926368	1	4.47E-05
CYP46A1	chr14:100193034-100193041	1	4.74E-05
TMC1	chr9:75315438-75445568	3	4.78E-05
PTPRT	chr20:40709557-41419897	9	1.26E-04
PRRT4	chr7:127991133-127999697	7	1.32E-04
CCDC77	chr12:518533-550107	1	1.47E-04
TTC4	Chr1:1:55183290-55207181	1	1.47E-04

※ Analysis was adjusted for AGE & SEX.

The functional relevance of identified BMI genes → not clear yet

3. Analysis in 620 DM samples

3-A. Single variant test (Top 10 loci associated with BMI)

SNP	Gene	Function	Chr	Position	Minor/ Major	MAF	Beta±SE	P-value
-	*DMXL2*	-	15	51,806,766	-/A	0.46527	0.565±0.105	9.95E-08
rs202022056	*SCUBE1*	V521V	22	43,616,580	T/C	0.00242	7.863±1.615	1.44E-06
rs72432623	*FBN1*	intron	15	48,755,450	-/A	0.38493	0.594±0.126	2.87E-06
rs145006588	*ZNF680*	3'-UTR	7	63,981,237	G/A	0.00404	5.830±1.258	4.42E-06
-	*ZNF107*	ds	7	64,170,968	T/-	0.00405	5.826±1.259	4.48E-06
-	*COBLL1*	intron	2	165,694,012	-/A	0.40953	0.548±0.119	4.54E-06
rs3743110	*CSNK1G1*	3'-UTR	15	64,462,463	G/A	0.04685	1.746±0.380	5.11E-06
rs60566469	*BCKDHB*	3'-UTR	6	81,054,353	T/C	0.16236	0.987±0.218	7.02E-06
rs7167612	*TRIP4*	E532E	15	64,737,225	G/A	0.04523	-1.779±0.393	7.16E-06
rs79326109	*HEATR5B*	3'-UTR	2	37,208,293	A/T	0.00323	6.312±1.407	8.69E-06

※ Analysis was adjusted for AGE & SEX.

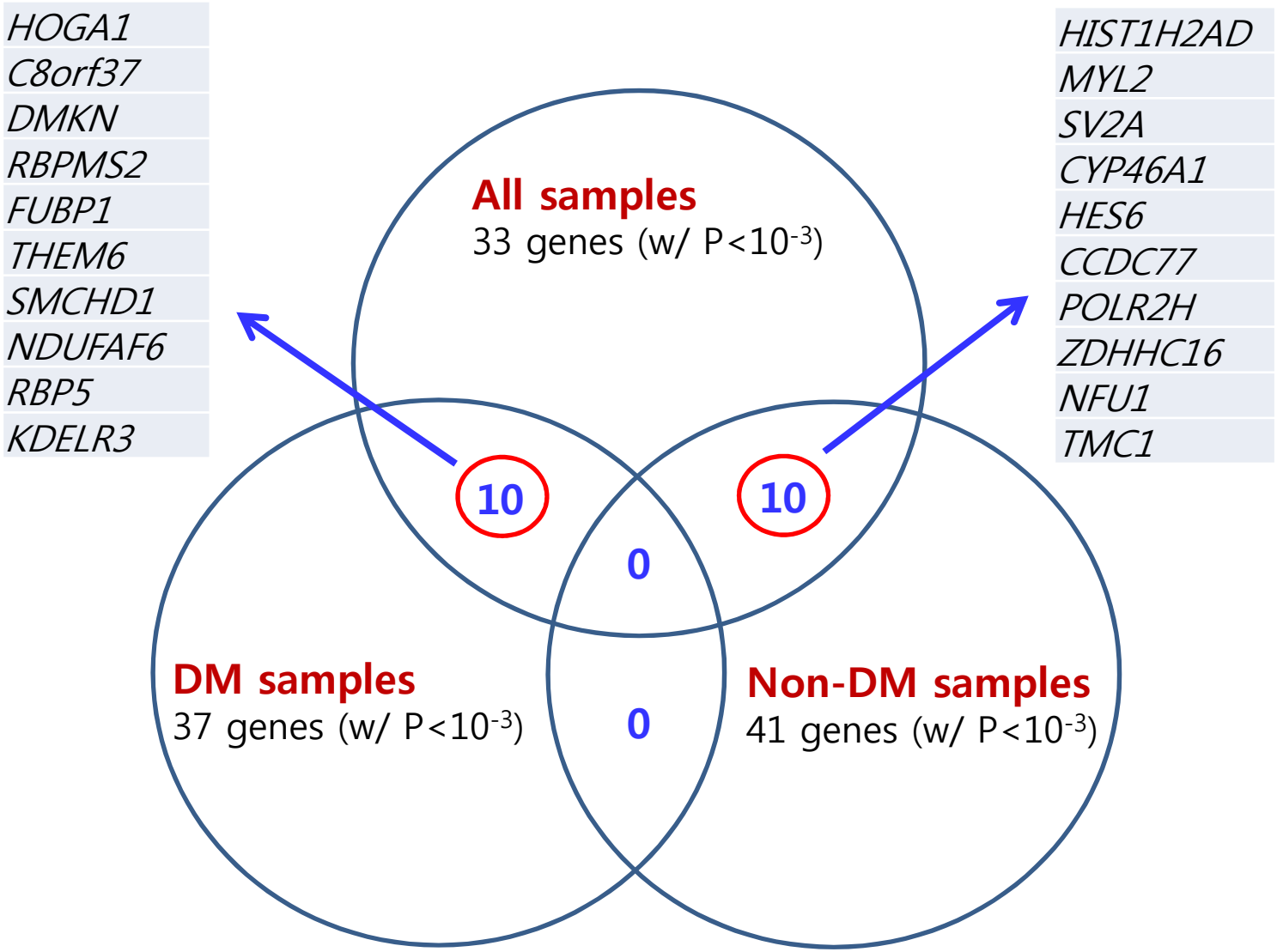
3-B. Gene-based test (Top 10 genes associated with BMI)

Gene	Position	# of variants/ gene set	P-value
RBPMS2	chr15:65041622-65041622	1	1.15E-05
HOGA1	chr10:99344508-99371274	6	1.54E-05
C8orf37	chr8:96264434-96281381	3	3.06E-05
PSMB6	chr17:4700993-4701334	1	5.10E-05
FUBP1	chr1:78426038-78430402	1	5.10E-05
MCM3	chr6:52129468-52148194	8	7.82E-05
AC093726.4	chr7:154862150-154862748	2	8.41E-05
NDUFAF6	chr8:96037261-96047772	4	8.78E-05
RBP5	chr12:7280874-7281359	2	2.27E-04
SMCHD1	chr18:2666885-2772321	12	3.41E-04

※ Analysis was adjusted for AGE & SEX.

The functional relevance of identified BMI genes → not clear yet

Overlappings of BMI genes among stratified sample groups



BMI genes in non-DM condition

- ✓ *HIST1H2AD* (*histone cluster 1, H2ad*) : core component of nucleosome
- ✓ *MYL2* (*myosin, light chain 2, regulatory, cardiac, slow*)
- ✓ *SV2A* (*synaptic vesicle glycoprotein 2A*) : control of regulated secretion in neural and endocrine cells
- ✓ *CYP46A1* (*cytochrome P450, family 46, subfamily A, polypeptide 1*) : monooxygenases involved in drug metabolism and synthesis of cholesterol, steroids and other lipids.
- ✓ *HES6* (hes family bHLH transcription factor 6) : a subfamily of basic helix-loop-helix transcription repressors
- ✓ *CCDC77, POLR2H, ZDHHC16, NFU1, TMC1, etc.*

BMI genes in DM condition

- ✓ *HOGA1* (*4-hydroxy-2-oxoglutarate aldolase 1*) : involved in kidney function
- ✓ *C8orf37* (*Chromosome 8 Open Reading Frame 37*) : unknown function
- ✓ *DMKN* (*dermokine*) : soluble regulator of keratinocyte differentiation
- ✓ *RBPMS2* (*RNA binding protein with multiple splicing 2*)
- ✓ *FUBP1* (*far upstream element (FUSE) binding protein 1*) : act both as activator and repressor of transcription
- ✓ *THEM6* (*thioesterase superfamily member 6*)
- ✓ *SMCHD1, NDUFAF6, RBP5, KDELR3, etc.*

Summary and future plans

1. Several low frequency BMI variants were identified in both DM and non-DM conditions.

2. One locus (*HOGA1*) showed the strong evidence of association with BMI (reaching genome-wide significance).

3. Identified BMI genes are lack of functional relatedness with obesity.

4. Validation of findings is required.

✓ by *de novo* genotyping in independent population samples

✓ by using ~1000 exome seq data from Child Obesity Cohort (in KNIH)

✓ by biological validation : gene knock-out/down, overexpression

Acknowledgements

SNUH

Prof. Kyong Soo Park

Prof. Soo-Heon Kwak

Hallym U

Dr. Jisun Lim

Kwang Sub Kim