SUPPLEMENTARY INFORMATION

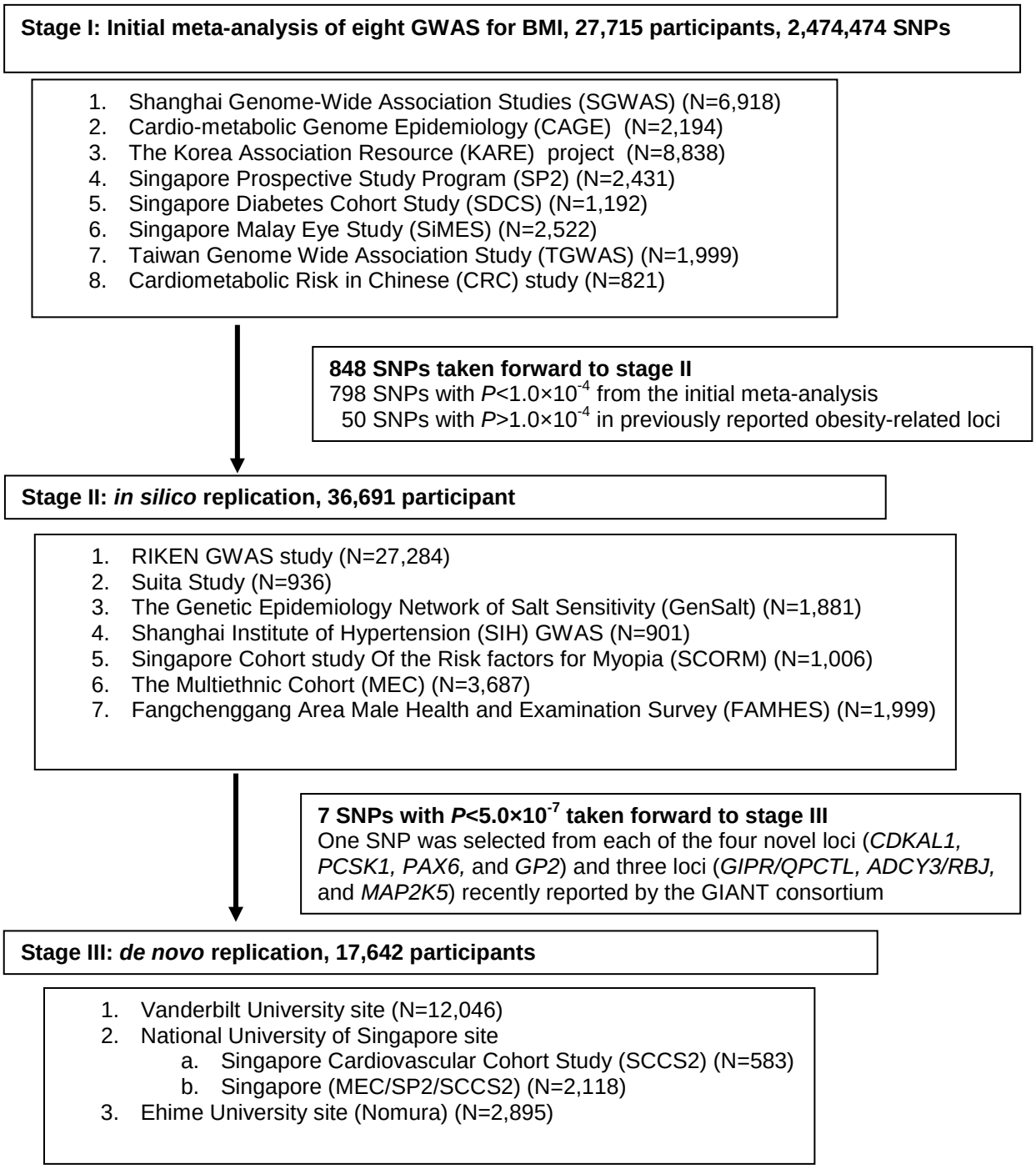
Meta-analysis of genome-wide association studies in East Asians identifies novel genetic variants associated with body mass index

Wanqing Wen, Yoon Shin Cho, Wei Zheng, Rajkumar Dorajoo, Norihiro Kato, Lu Qi, Chien-Hsiun Chen, Ryan J. Delahanty, Yukinori Okada, Yasuharu Tabara, Dongfeng Gu, Dingliang Zhu, Christopher A. Haiman, Zengnan Mo, Yu-Tang Gao, Seang Mei Saw, Min Jin Go, Fumihiko Takeuchi, Li-Ching Chang, Yoshihiro Kokubo, Jun Liang, Mei Hao, Loic Le Marchand, Yi Zhang, Yanling Hu, Tien Yin Wong, Jirong Long, Bok-Ghee Han, Michiaki Kubo, Ken Yamamoto, Mei-Hsin Su, Tetsuro Miki, Brian E. Henderson, Huaidong Song, Aihua Tan, Jiang He, Daniel P.K. (Peng-Keat) Ng, Qiuyin Cai, Tatsuhiko Tsunoda, Fuu-Jen Tsai, Naoharu Iwai, Gary K. Chen, Jiajun Shi, Jianfeng Xu, Xue Ling Sim, Yong-Bing Xiang, Shiro Maeda, Rick T.H. Ong, Chun Li, Yusuke Nakamura, Tin Aung, Naoyuki Kamatani, Jian Jun Liu, Wei Lu, Mitsuhiro Yokota, Mark Seielstad, Cathy S.J. Fann, The GIANT Consortium, Jer-Yuarn Wu, Jong-Young Lee, Frank Hu, Toshihiro Tanaka, E. Shyong Tai, Xiao Ou Shu

1. SUPPLEMENTARY FIGURE AND TABLES

1.1. Supplementary Figure

Supplementary Figure 1. Study design.



1.2. Supplementary Tables

Supplementary Table 1. Description of participating studies in the initial meta-analysis of GWAS (stage I) and in the replication studies (stage II, III).

Supplementary Table 2. Descriptive characteristics of participating studies in the initial meta-analysis of GWAS (stage I) and in the replication studies (stage II, III).

Supplementary Table 3. Information on genotyping methods, quality control of SNPs, imputation, and statistical analysis in stage I studies.

Supplementary Table 4. Observed associations of BMI with SNPs in all previously and newly identified loci in East Asian populations.

Supplementary Table 5. Comparison and conditional analysis of the SNPs identified by this study and by the GIANT consortium.

Supplementary Table 6. Detailed results by study for the three newly identified loci (*CDKAL1*, *PSCK1*, and *GP2*) and one suggested locus (*PAX6*) associated with BMI variation in East Asian populations.

Supplementary Table 7. Observed associations between BMI and seven SNPs in data obtained from the GIANT consortium.

Supplementary Table 8. Associations between the identified SNPs and obesity (BMI>=27.5) in East Asian populations.

Supplementary Table 9. Observed associations between BMI and SNPs in the four newly identified loci in stage I and II data.

Supplementary Table 10. Observed associations in stage I data for BMI and selected relevant SNPs.

Supplementary Table 1. Description of participating studies in the initial meta-analysis of GWAS (stage I) and in the replication studies (stage II, III)

Study Name	Study design	Study participants
Stage I. Initial meta-analysis of GWAS		
Shanghai genome-wide association studies (SGWAS)	population-based	1) 2,895 breast cancer cases, 2) 828 endometrial cancer cases, 3) 886 type 2 diabetes cases, 4) 2,309 controls, all unrelated
Cardio-metabolic Genome Epidemiology (CAGE)	Case-control	1) 790 type 2 diabetes cases, 2) 1,404 controls, all unrelated
The Korea Association Resource (KARE) project	population-based	8,838 unrelated participants with BMI and genotyping data from a total 10,038 participant of the KARE project
Singapore Prospective Study Program (SP2)	population-based	2,431 random unrelated individuals
Singapore Diabetes Cohort Study (SDCS)	type 2 diabetes cases	1,992 type 2 diabetes cases, all unrelated
Singapore Malay Eye Study (SiMES)	population-based	2,522 random unrelated individuals
Taiwan Genome Wide Association Study (TGWAS)	Case-control	1) 999 type 2 diabetes cases, 2) 1,000 controls, all unrelated
Cardiometabolic Risk in Chinese (CRC) study	population-based	821 unrelated healthy participants
Stage II: *in silico* replication studies		
REKIN GWAS study	hospital-based	27,284 unrelated Japanese subjects enrolled in the BioBank Japan Project
Suita Study	population-based	933 unrelated healthy adults
The Genetic Epidemiology Network of Salt Sensitivity (GenSalt)*	population-based	1,881 individuals, some family members included
Shanghai Institute of Hypertension (SIH) GWAS	hospital-based	446 hypertension cases and 455 controls, all unrelated
Singapore Cohort study Of the Risk factors for Myopia (SCORM)	population-based	1,006 unrelated children with age 9 years
The Multiethnic Cohort (MEC)	population-based	1) 876 breast cancer cases, 2) 980 prostate cancer cases, 3) 1,831 controls, all unrelated
Fangchenggang Area Male Health and Examination Survey (FAMHES)	population-based	1,999 unrelated healthy men
Stage III: *de novo* replication studies		
Vanderbilt University site	population-based	1) 3,975 breaset cancer cases, 2) 1,586 type 2 diabetes cases, 3) 6,485 controls, all unrelated
Singapore Cardiovascular Cohort Study (SCCS2)	population-based	583 unrelated Chinese selected from SCCS2
Singapore (MEC/SP2/SCCS2)	population-based	2,118 unrelated Malays selected from MEC, SP2, SCCS2
Ehime site (Nomura)	population-based	2,895 random unrelated individuals

* A mixed model was used to take account for the family structure.

Supplementary Table 2. Descriptive characteristics of participating studies in the initial meta-analysis of GWAS (stage I) and in the replication studies (stage II, III)

Study name	Population	Sample size (women %)	Means/Standard deviations		Obesity (%)
			Age	BMI*	
Stage I. Initial meta-analysis of GWAS					
SGWAS	Chinese	6918(100.0)	50.9/8.7	24.3/3.7	18.0%
CAGE	Japanese	2194(36.5)	66.0/8.2	23.6/3.2	11.0%
KARE	Korean	8838(52.7)	52.2/8.9	24.6/3.1	16.6%
SP2	Chinese	2431(53.3)	48.0/11.3	22.9/3.7	10.2%
SDCS	Chinese	1992(49.8)	64.3/10.2	25.3/3.9	25.2%
SiMES	Malay	2522(50.5)	59.0/11.0	26.4/5.1	36.8%
TGWAS	Chinese	1999(49.7)	55.2/15.1	24.7/3.8	19.4%
CRC	Chinese	821(76.9)	45.0/9.0	22.2/2.2	0.9%
Stage II: *in silico* replication studies					
RIKEN	Japanese	27284(46.6)	62.9/12.0	22.8/3.6	9.3%
Suita	Japanese	933(58.4)	59.0/7.0	22.7/2.9	6.1%
GenSalt	Chinese	1881(47.2)	38.7/9.5	23.4/3.2	10.6%
SIH	Chinese	901(49.6)	54.6/6.4	22.4/1.8	0.0%
SCORM	Chinese	1006(49.0)	9.0/0	17.1/2.9	NA
MEC	Japanese	3687(46.2)	61.8/8.1	24.6/3.5	16.8%
FAMHES	Chinese	1999(0)	36.5/7.6	23.3/3.4	10.8%
Stage III: *de novo* replication studies					
Vanderbilt	Chinese	12046(88.0)	55.3/9.7	24.4/3.5	17.0%
SCCS2_Chinese	Chinese	583(49.6)	58.0/10.2	23.5/3.6	13.9%
MEC/SP2/SCCS2	Malay	2118(59.6)	49.7/12.1	26.4/5.2	37.2%
Nomura	Japanese	2895(56.6)	61.1/14.0	23.4/3.2	10.3%

* Height and weight measured in all studies, except the RIKEN study, in which height and weight were self reported.

Supplementary Table 3. Information on genotyping methods, quality control of SNPs, imputation, and statistical analysis in stage I studies

	Genotyping				Imputation				Statistical analysis		
		Inclusion criteria				Inclusion criteria				Inflation	SNPs in meta-
Study name	Platform	MAF	Call rate	p for HWE	Software	MAF	Quality	Reference	Software	factor (λ)	analysis
Stage I. Initial meta-analysis of GWAS											
SGWAS	Affy 6.0	>1%	>95%	>1e-6	MACH	>1%	r2-hat>0.3	HapMap rel. 22, build 36, CHB+JPT	PLINK	1.075	2,318,347
CAGE	Illumina610/ 550	>1%	>95%	>1e-6	BEAGLE	>1%	r2-hat>0.3	HapMap rel. 24, build 36, CHB+JPT	PLINK	1.009	1,997,238
KARE	Affy 5.0	>1%	>90%	>1e-7	IMPUTE	>1%	proper-info>0.5	HapMap rel. 22, build 36, CHB+JPT	PLINK	1.050	1,850,000
SP2	Illumina1M/ 610/ 550	>1%	>95%	>1e-4	IMPUTE	>1%	proper-info>0.5	HapMap rel. 22, build 36, CHB+JPT	SNPTEST	1.018	2,415,597
SDCS	Illumina1M/ 610	>1%	>95%	>1e-4	IMPUTE	>1%	proper-info>0.5	HapMap rel. 22, build 36, CHB+JPT	SNPTEST	1.007	2,415,597
SiMES	Illumina610	>1%	>95%	>1e-4	IMPUTE	>1%	proper-info>0.5	HapMap rel. 22, build 36, CHB+JPT+CEU+YRI	SNPTEST	1.046	1,934,324
TGWAS	Illumina 550	>1%	>95%	>1e-6	MACH	>5%	r2-hat>0.3	HapMap rel. 22, build 36, CHB+JPT	PLINK	1.007	2,217,616
CRC	Illumina610	>1%	>95%	>1e-5	MACH	>1%	r2-hat>0.3	HapMap rel. 22, build 36, CHB+JPT	ProbABEL	1.004	2,249,369
Stage II: *in silico* replication studies											
RIKEN	Illumina610	>1%	>99%	>1e-6	MACH	>1%	r2-hat>0.7	HapMap rel. 22, build 36, CHB+JPT	R	1.123	1,533
Suita	Illumina550	>1%	>95%	>1e-6	MACH	>1%	r2-hat>0.3	HapMap rel. 22, build 36, CHB+JPT	PLINK	1.000	1,338
GenSalt	Affy 6.0	>1%	>90%	>1e-6	MACH	>1%	r2-hat>0.3	HapMap rel. 22, build 36, CHB+JPT	R	1.050	1,409
SIH	Illumina610	>1%	>97%	>1e-5	MACH	>1%	r2-hat>0.3	HapMap rel. 22, build 36, CHB+JPT	PLINK	1.000	1,516
SCORM	Illumina550	>1%	>95%	>1e-4	IMPUTE	>1%	proper-info>0.5	HapMap rel. 22, build 36, CHB+JPT	SNPTEST	1.022	1,373
MEC	Illumina Infinium 660W	>1%	>95%	>1e-6	MACH	>1%	r2-hat>0.3	HapMap rel. 21, build 36, CHB+JPT	C++	1.050	1,344
FAMHES	Illumina HumanOmni1	>1%	>95%	>1e-4	IMPUTE	>1%	proper-info>0.5	HapMap rel. 24, build 36, CHB+JPT	SAS	1.036	1,354

Supplementary Table 4. Observed associations of BMI with SNPs in all previously identified loci in East Asian populations

Gene	Chr	SNP	Allele Effect/Other	EAF Asian	EAF European	Stage I β(SE)	Stage I p	Stage II β(SE)	Stage II p	Stage I & II β(SE)	Stage I & II p	Stage III β(SE)	Stage III p	Final P	Explained variance	Reference
Previously identified loci for BMI from studies of European ancestry populations																
FTO	16	rs17817449	G/T	0.17	0.46	9.02(1.31)	6.13E-12	7.92(1.06)	8.18E-14	8.46(0.79)	4.60E-27			4.60E-27	0.18%	1-5,7,8
SEC16B	1	rs574367	T/G	0.20	0.27	7.58(1.13)	2.38E-11	5.93(0.92)	1.28E-10	6.57(0.72)	9.47E-20			9.47E-20	0.11%	4,8
MC4R	18	rs6567160	C/T	0.21	0.28	5.82(1.08)	6.92E-08	5.51(0.93)	3.35E-09	5.64(0.71)	2.76E-15			2.76E-15	0.10%	3-5,7,8
GIPR/QPCTL	19	rs11671664	G/A	0.50	0.92	4.89(1.12)	1.29E-05	4.22(0.76)	2.57E-08	4.43(0.63)	2.80E-12	3.30(1.13)	3.57E-03	5.94E-14	0.09%	New
		rs2287019	C/T	0.79	0.83	4.77(1.52)	1.74E-03	0.38(0.97)	6.96E-01	1.59(0.82)	5.26E-02			5.26E-02	0.00%	8
ADCY3/RBJ	2	rs6545814	G/A	0.45	0.42	3.84(0.88)	1.20E-05	3.26(0.76)	1.62E-05	3.50(0.58)	1.62E-09	4.59(1.04)	1.05E-05	1.35E-13	0.05%	New
		rs11676272	G/A	0.45	0.46	3.74(0.88)	2.26E-05	3.70(0.80)	4.02E-06	3.72(0.60)	5.88E-10			5.88E-10	0.07%	New
		rs713586	C/T	0.46	0.47	2.59(1.07)	1.56E-02	3.14(0.84)	1.75E-04	2.94(0.67)	1.02E-05			1.02E-05	0.05%	8
BDNF	11	rs6265	C/T	0.44	0.81	3.98(0.91)	1.18E-05	4.97(0.83)	2.72E-09	4.53(0.62)	3.56E-13			3.56E-13	0.12%	4,8
MAP2K5	15	rs4776970	A/T	0.22	0.62	5.31(1.09)	1.10E-06	2.55(0.90)	4.63E-03	3.64(0.70)	2.24E-07	3.76(1.26)	2.90E-03	2.33E-09	0.02%	New
		rs2241423	G/A	0.38	0.20	4.21(0.90)	2.78E-06	2.19(0.84)	9.06E-03	3.10(0.62)	6.02E-07			6.02E-07	0.02%	8
GNPDA2	4	rs10938397	G/A	0.29	0.45	3.69(1.17)	1.60E-03	3.72(0.85)	1.30E-05	3.71(0.70)	9.69E-08			9.69E-08	0.06%	5,8
TFAP2B	6	rs4715210	T/C	0.21	0.16	5.32(1.21)	1.12E-05	3.05(0.91)	7.64E-04	3.84(0.73)	1.61E-07			1.61E-07	0.03%	8
TMEM18	2	rs11127485	T/C	0.91	0.85	6.78(1.61)	2.63E-05	3.92(1.33)	3.26E-03	5.05(1.04)	1.16E-06			1.16E-06	0.03%	4,5,7,8
TMEM160	19	rs3810291	A/G	0.24	0.67	3.67(1.24)	3.05E-03	3.34(1.09)	2.18E-03	3.48(0.83)	2.66E-05			2.66E-05	0.04%	8
TNNI3K	1	rs1514175	A/G	0.79	0.46	2.79(1.03)	4.92E-03	2.98(1.02)	3.69E-03	2.94(0.74)	6.38E-05			6.38E-05	0.03%	8
MTCH2	11	rs3817334	T/C	0.31	0.41	2.45(0.94)	9.50E-03	2.34(0.85)	5.90E-03	2.39(0.64)	1.82E-04			1.82E-04	0.02%	5,8
FAIM2	12	rs7138803	A/G	0.31	0.35	2.55(0.99)	9.78E-03	1.63(0.81)	4.49E-02	1.99(0.63)	1.74E-03			1.74E-03	0.01%	8
FLJ35779	5	rs2112347	T/G	0.44	0.62	1.50(0.88)	8.71E-02	2.01(0.84)	1.66E-02	1.77(0.61)	3.92E-03			3.92E-03	0.02%	8
SH2B1	16	rs4788102	A/G	0.13	0.38	2.04(1.33)	1.24E-01	2.65(1.15)	2.14E-02	2.40(0.88)	6.51E-03			6.51E-03	0.02%	4,5,8
RPL27A	11	rs4929949	C/T	0.43	0.60	0.87(1.07)	4.13E-01	2.34(0.90)	9.66E-03	1.75(0.70)	1.23E-02			1.23E-02	0.03%	8
NEGR1	1	rs2568958	A/G	0.92	0.64	2.78(1.61)	8.54E-02	1.87(1.51)	2.16E-01	2.28(1.12)	4.06E-02			4.06E-02	0.01%	4,5,8
SFRS10/ETV5	3	rs7647305	C/T	0.95	0.80	2.45(2.03)	2.27E-01	2.68(2.13)	2.07E-01	2.56(1.49)	8.50E-02			8.50E-02		4,8
NUDT3	6	rs206936	G/A	0.56	0.18	2.93(0.88)	8.84E-04	-1.12(0.84)	1.80E-01	0.75(0.61)	2.23E-01			2.23E-01		8
LRRN6C	9	rs10968576	G/A	0.19	0.33	1.48(1.40)	2.91E-01	0.60(1.03)	5.59E-01	0.90(0.84)	2.83E-01			2.83E-01		8
KCTD15	19	rs29941	G/A	0.23	0.68	1.65(1.04)	1.11E-01	0.03(0.94)	9.78E-01	0.74(0.71)	2.98E-01			2.98E-01		4,5,8
NPC1	18	rs1805081	C/T	0.27	0.47	-2.07(1.27)	1.01E-01	-0.03(0.86)	9.75E-01	-0.65(0.72)	3.68E-01			3.68E-01		6
MAF	16	rs1424233	T/C	0.71	0.56	1.00(0.94)	2.89E-01	0.19(0.92)	8.34E-01	0.58(0.67)	3.87E-01			3.87E-01		6
PTBP2	1	rs1555543	C/A	0.87	0.59	1.13(1.35)	4.01E-01	0.06(1.15)	9.55E-01	0.50(0.88)	5.74E-01			5.74E-01		8
MTIF3	13	rs4771122	G/A	0.22	0.26	-0.14(2.22)	9.51E-01	0.57(1.01)	5.69E-01	0.46(0.92)	6.20E-01			6.20E-01		8
FANCL	2	rs887912	T/C	0.00	0.32											8
LRP1B	2	rs2890652	C/T	NA	0.13											8
CADM2	3	rs13078807	G/A	0.00	0.23											8
SLC39A8	4	rs13107325	T/C	0.00	0.10											8
ZNF608	5	rs4836133	A/C	NA	NA											8
PTER	10	rs10508503	C/T	1.00	0.91											6
PRKD1	14	rs11847697	T/C	0.00	0.04											8
NRXN3	14	rs10150332	C/T	0.00	0.26											8
GPRC5B	16	rs12444979	C/T	1.00	0.94											8
Previously identified loci for early-onset extreme obesity from studies of European ancestry populations																
TNKS-MSRA	8	rs17150703	G/A	0.83	0.89	1.48(1.24)	2.30E-01	1.32(1.04)	2.07E-01	1.38(0.81)	8.68E-02			8.68E-02		7
SDCCAG8	1	rs2783963	G/A	0.94	0.88	3.73(2.08)	7.31E-02	0.42(1.86)	8.22E-01	1.85(1.40)	1.88E-01			1.88E-01		7

β: Effect of SNPs per allele (in percentage) on BMI, derived from meta-analysis.

p: Derived from meta-analysis adjusted for for both study-specific inflation factors (for stage I, II) and also the estimated inflation factor for the stage I meta-analysis statistic (for the combined data).

NA: Not available in the data.

EAF: Effect allele frequency, estimated from stage I and II studies for Asian and obtained from the HapMap data for Europeans.

MAF: Minor allele frequency, estimated from stage I and II studies for Asian and obtained from the HapMap data for Europeans.

The effect sizes obtained from stage II data were used to estimate the explained variance.

Final P: Combined all available data from three stages.

Explained variance: Only calculated forthose SNPs with the final p<0.05.

Supplementary Table 5. Comparison and conditional analysis of the SNPs identified by this study and by the GIANT consortium

Locus	SNP	Number	Allele	Model with single SNP		Model with both SNPs		EAF		RSQ		D'	
			Effect/Other	β(SE)	p	β(SE)	p	Asian	European	Asian	European	Asian	European
ADCY3	rs6545814	57817	G/A	3.66(0.59)	4.45E-10	3.81(1.26)	2.46E-03	0.45	0.42				
	rs713586	57817	C/T	2.72(0.63)	1.82E-05	0.96(1.26)	4.44E-01	0.50	0.47	0.742	0.683	0.975	0.961
MAP2K5	rs4776970	55574	A/T	3.91(0.71)	3.69E-08	3.64(1.03)	3.81E-04	0.22	0.62				
	rs2241423	55574	G/A	3.06(0.60)	3.28E-07	1.41(0.88)	1.10E-01	0.38	0.80	0.536	0.533	1.000	1.000
GIPR	rs11671664	48985	G/A	4.78(0.65)	1.45E-13	4.38(0.68)	1.09E-10	0.50	0.92				
	rs2287019	48985	C/T	1.86(0.81)	2.18E-02	0.53(0.85)	5.37E-01	0.79	0.83	0.026	0.264	0.331	0.679

SNPs rs6545814, rs4776970, rs11671664 were identified by this study, and rs713586, rs2241423, rs2287019 were identified by the GIANT consortium.

β: Effect of SNPs per allele (in percentage) on BMI, derived from meta-analysis.

p: Derived from meta-analysis without adjustment for study-specific inflation factors.

EAF: Effect allele frequency, estimated from stage I and II studies for Asian and obtained from the HapMap data for Europeans.

RSQ: r-squared for LD measurement according to HapMap data.

Supplementary Table 6. Detailed results by study for the three newly identified loci (*CDKAL1*, *PSCK1*, *and GP2*) and one suggested locus (*PAX6*) associated with BMI variation in East Asian populations

		in *CDKAL1* (6p22.3) rs9356744 (T/C)			near *PCSK1* (5q15) rs261967 (C/A)			near *GP2* (16p12.3) rs12597579 (C/T)			near *PAX6* (11p13) rs652722 (C/T)		
Study	**Number**	**EAF**	**β(SE)**	**p**	**EAF**	**β(SE)**	**p**	**EAF**	**β(SE)**	**p**	**EAF**	**β(SE)**	**p**
SGWAS	6918	0.59	3.50(1.60)	2.71E-02	0.42	6.50(1.60)	3.69E-05	0.73	4.66(1.77)	8.57E-03	0.63	2.66(1.63)	1.02E-01
CAGE	2194	0.58	4.78(2.99)	1.09E-01	0.40	5.49(3.04)	7.07E-02	0.83	-1.69(3.90)	6.65E-01	0.61	4.62(3.03)	1.28E-01
KARE	8838	0.52	3.14(1.55)	4.28E-02	0.43	0.62(1.53)	6.87E-01	0.79	4.05(1.83)	2.71E-02	0.61	5.04(1.52)	9.21E-04
SP2	2431	0.66	1.05(2.66)	6.93E-01	0.40	3.77(2.63)	1.52E-01	0.75	5.59(2.89)	5.30E-02	0.60	4.20(2.55)	9.97E-02
SDCS	1992	0.62	10.85(3.14)	5.56E-04	0.42	3.91(3.17)	2.17E-01	0.76	1.50(3.62)	6.79E-01	0.60	1.52(3.18)	6.34E-01
SiMES	2522	0.63	1.05(2.85)	7.13E-01	0.48	3.71(2.79)	1.84E-01	0.89	5.54(4.27)	1.94E-01	0.61	2.75(2.84)	3.34E-01
TGWAS	1999	0.62	19.38(11.71)	9.79E-02	0.41	34.23(11.48)	2.87E-03	0.73	22.62(12.94)	8.06E-02	0.62	27.75(11.76)	1.83E-02
CRC	821	0.57	9.73(10.48)	3.52E-01	0.41	4.85(10.40)	6.40E-01	0.71	19.17(11.46)	9.40E-02	0.65	-18.10(11.24)	1.07E-01
Stage I	***27715***		***3.71(0.89)***	***3.21E-05***		***3.89(0.89)***	***1.22E-05***		***4.16(1.05)***	***7.13E-05***		***3.73(0.89)***	***2.84E-05***
RIKEN	27284	0.58	3.21(0.85)	1.74E-04	0.39	3.87(0.87)	8.78E-06	0.82	4.22(1.11)	1.51E-04	0.61	2.57(0.87)	3.08E-03
Suita	933	0.59	1.93(4.72)	6.82E-01	0.39	6.65(4.73)	1.60E-01	0.80	5.76(5.72)	3.14E-01	0.62	10.47(4.80)	2.94E-02
GenSalt	1881	0.58	-0.28(3.27)	9.31E-01	0.44	2.98(3.27)	3.62E-01	0.76	4.42(3.66)	2.29E-01	0.67	1.61(3.48)	6.43E-01
SIH	901	0.59	5.98(4.78)	2.11E-01	0.40	-2.33(4.90)	6.35E-01	0.71	1.17(5.15)	8.20E-01	0.62	1.14(4.72)	8.09E-01
SCORM	1006	0.64	2.52(4.56)	5.81E-01	0.41	6.47(4.40)	1.42E-01	0.73	9.37(4.85)	5.31E-02	0.62	9.53(4.48)	3.35E-02
MEC	3687	0.55	5.20(2.30)	2.39E-02	0.40	3.08(2.35)	1.90E-01	0.82	0.40(3.03)	8.95E-01	0.60	1.62(2.32)	4.84E-01
FAMHES	1999	0.65	5.81(3.26)	7.53E-02	0.45	4.13(3.07)	1.79E-01	0.77	5.56(3.71)	1.34E-01	0.59	2.19(3.12)	4.83E-01
Stage II	***37691***		***3.40(0.76)***	***7.67E-06***		***3.77(0.77)***	***9.36E-07***		***4.09(0.96)***	***2.07E-05***		***2.75(0.77)***	***3.70E-04***
Shanghai	12046	0.60	3.07(1.26)	1.53E-02	0.42	1.05(1.25)	4.00E-01	0.72	0.89(1.39)	5.22E-01	0.63	0.63(1.28)	6.24E-01
SCCS2-Chinese	583	0.66	13.00(6.02)	3.13E-02	0.40	-9.14(5.88)	1.20E-01	\	\	\	\	\	\
MEC/SP2/SCCS2	2118	0.65	1.58(3.11)	6.11E-01	0.48	-2.13(3.00)	4.77E-01	\	\	\	\	\	\
Nomura	2895	0.59	2.62(2.65)	3.23E-01	\	\	\	0.81	7.63(3.36)	2.35E-02	0.64	5.69(2.76)	3.93E-02
Stage III	***17642***		***3.13(1.06)***	***3.02E-03***		***0.22(1.13)***	***8.46E-01***		***1.87(1.29)***	***1.45E-01***		***1.53(1.16)***	***1.89E-01***
Combined	**83048**		**3.43(0.51)**	**2.00E-11**		**3.05(0.52)**	**5.13E-09**		**3.59(0.63)**	**1.02E-08**		**2.83(0.53)**	**7.65E-08**
Test for homogeneity across studies				**6.46E-01**			**5.71E-02**			**6.41E-01**			**2.98E-01**
Exclude SCROM			3.45(0.51)	1.36E-11		3.01(0.52)	7.41E-09		3.50(0.63)	2.26E-08		2.76(0.53)	1.49E-07
Female	48123		3.07(0.65)	2.18E-06		2.53(0.66)	1.20E-04		3.63(0.78)	3.04E-06		2.65(0.67)	7.05E-05
Male	34925		4.80(0.76)	2.54E-10		3.75(0.78)	1.57E-06		3.46(0.96)	3.14E-04		2.66(0.78)	6.58E-04
Test for homogeneity across genders				8.37E-02			2.33E-01			8.91E-01			9.92E-01
Chinese	34112		3.82(0.80)	2.05E-06		3.12(0.80)	8.78E-05		3.25(0.90)	2.88E-04		2.02(0.82)	1.34E-02
Japanese	36993		3.45(0.76)	6.08E-06		3.97(0.81)	1.04E-06		3.77(0.99)	1.44E-04		3.06(0.78)	8.85E-05
Karea	8838		3.14(1.55)	4.28E-02		0.62(1.53)	6.87E-01		4.05(1.83)	2.71E-02		5.04(1.52)	9.21E-04
Malay	3105		1.29(2.10)	5.39E-01		1.00(2.04)	6.25E-01		5.54(4.27)	1.94E-01		2.75(2.84)	3.36E-01
Test for homogeneity across populations				7.26E-01			1.85E-01			9.31E-01			3.65E-01
After exclusion of subjects with type 2 diabetes													
	69512		2.95(0.54)	4.01E-08		2.87(0.55)	1.53E-07		3.36(0.66)	3.06E-07		2.60(0.55)	2.29E-06
After exclusion of subjects with type 2 diabetes and cancers													
	59958		2.95(0.59)	5.00E-07		2.67(0.60)	8.65E-06		3.10(0.73)	2.10E-05		2.95(0.60)	9.37E-07

Genotype shown in the order of effect allele/other allele

EAF: Effect allele frequency.

β: Effect of SNPs per allele (in percentage) on BMI, derived from meta-analysis.

p: Derived from meta-analysis adjusted for for both study-specific inflation factors (for stage I, II) and also the estimated inflation factor for the stage I meta-analysis statistic (for the combined data).

Supplementary Table 7. Observed associations between BMI and seven SNPs in data obtained from the GIANT consortium

Nearest gene (distance from gene)	SNP	Allele Effect/Other	EAF in European	Number	β(SE)	p	Explained variance
ADCY3	rs6545814	G/A	0.43	123,822	2.28(0.47)	1.67E-06	0.03%
PCSK1 (81.3kb)	rs261967	C/A	0.43	123,864	1.50(0.47)	1.58E-03	0.01%
CDKAL1	rs9356744	T/C	0.68	123,862	0.67(0.51)	1.86E-01	0.00%
PAX6 (78.3kb)	rs652722	C/T	0.74	123,820	-0.22(0.54)	6.76E-01	0.00%
MAP2K5	rs4776970	A/T	0.64	123,864	2.74(0.49)	2.17E-08	0.03%
GP2 (63.9kb)	rs12597579	C/T	0.95	123,662	2.86(1.10)	9.00E-03	0.01%
GIPR	rs11671664	G/A	0.88	93,233	2.85(0.88)	1.24E-03	0.02%

Supplementary Table 8. Associations between the newly identified SNPs and obesity (BMI>=27.5) in East Asian populations

Nearest gene (distance from gene)	Top SNP	Allele	Stage I		Stage II		de novo		All		
		Effect/Other	OR (95% CI)	p	OR (95% CI)	p	OR (95% CI)	p	OR (95% CI)	p	p_H
PCSK1 (81.3kb)	rs261967	C/A	1.06(1.01-1.12)	1.56E-02	1.07(1.01-1.12)	1.38E-02	1.00(0.94-1.06)	9.59E-01	1.05(1.02-1.08)	3.11E-03	6.90E-01
CDKAL1	rs9356744	T/C	1.07(1.02-1.13)	6.45E-03	1.11(1.06-1.17)	4.61E-05	1.11(1.04-1.17)	5.96E-04	1.10(1.06-1.13)	4.21E-09	1.23E-01
PAX6 (78.3kb)	rs652722	C/T	1.07(1.02-1.13)	2.90E-03	1.07(1.02-1.13)	1.14E-02	0.98(0.91-1.04)	4.76E-01	1.05(1.02-1.08)	1.76E-03	3.07E-01
GP2 (63.9kb)	rs12597579	C/T	1.08(1.02-1.14)	9.12E-03	1.06(0.99-1.14)	7.33E-02	1.03(0.96-1.11)	4.07E-01	1.06(1.02-1.10)	1.71E-03	8.21E-01

OR (95% CI): Odds ratios and 95% confidence intervals for per allele effect of SNPs on obesity, derived from meta-analysis.

p: Derived from meta-analysis adjusted for study-specific inflation factors.

p_H: p for homogeneity test across studies in both discovery and replication stages.

Supplementary Table 9. Observed associations between BMI and SNPs in the four newly identified loci in stage I and II data

Gene (distance from gene)	SNP	Chr	Position	Allele	Stage I		Stage II		Stage I & II combined		EAF	LD	
				Effect/Other	β(SE)	p	β(SE)	p	β(SE)	p		RSQ	D'
5q15	rs261967	5	95876006	C/A	3.89(0.89)	1.22E-05	3.77(0.77)	9.36E-07	3.82(0.59)	8.50E-11	0.41	1.00	1.00
PCSK1(81.3kb)	rs263349	5	95877380	G/T	3.87(0.89)	1.48E-05	3.77(0.77)	9.57E-07	3.81(0.59)	9.54E-11	0.41	1.00	1.00
	rs3853212	5	95875104	C/T	3.81(0.88)	1.66E-05	3.78(0.77)	9.08E-07	3.79(0.59)	9.77E-11	0.41	1.00	1.00
	rs1837269	5	95884900	C/T	3.93(0.89)	9.31E-06	3.71(0.77)	1.38E-06	3.80(0.59)	1.04E-10	0.41	1.00	1.00
	rs7737742	5	95871003	C/T	3.83(0.89)	1.68E-05	3.78(0.77)	8.95E-07	3.80(0.59)	1.08E-10	0.41	1.00	1.00
	rs10077823	5	95872086	A/G	3.81(0.89)	1.79E-05	3.78(0.77)	8.79E-07	3.79(0.59)	1.16E-10	0.41	1.00	1.00
	rs261964	5	95873832	C/T	3.82(0.89)	1.73E-05	3.78(0.77)	9.25E-07	3.79(0.59)	1.16E-10	0.41	1.00	1.00
	rs6556926	5	95874369	T/A	3.81(0.89)	1.89E-05	3.78(0.77)	8.98E-07	3.79(0.59)	1.18E-10	0.41	1.00	1.00
	rs2611742	5	95882257	C/T	3.89(0.89)	1.15E-05	3.72(0.77)	1.34E-06	3.79(0.59)	1.22E-10	0.41	1.00	1.00
	rs6882366	5	95890449	C/T	3.84(0.89)	1.48E-05	3.75(0.77)	1.07E-06	3.79(0.59)	1.22E-10	0.40	1.00	1.00
	rs11951673	5	95886768	C/T	3.83(0.89)	1.56E-05	3.72(0.77)	1.29E-06	3.77(0.59)	1.54E-10	0.40	1.00	1.00
	rs1026534	5	95880847	C/G	3.82(0.89)	1.68E-05	3.72(0.77)	1.32E-06	3.76(0.59)	1.65E-10	0.41	1.00	1.00
	rs10476682	5	95884014	A/G	3.82(0.89)	1.71E-05	3.71(0.77)	1.39E-06	3.76(0.59)	1.74E-10	0.41	1.00	1.00
	rs261966	5	95875343	C/T	3.85(0.89)	1.44E-05	3.67(0.77)	1.75E-06	3.75(0.59)	1.91E-10	0.41	1.00	1.00
6p22.3	rs9356744	6	20793465	T/C	3.71(0.89)	3.21E-05	3.39(0.76)	7.67E-06	3.52(0.58)	1.60E-09	0.58		
CDKAL1	rs7766070	6	20794552	C/A	3.73(0.89)	2.89E-05	3.40(0.76)	7.42E-06	3.53(0.58)	1.42E-09	0.58	1.00	1.00
	rs9368219	6	20782670	C/T	3.62(0.89)	5.24E-05	3.56(0.78)	4.66E-06	3.58(0.59)	1.42E-09	0.58	0.96	1.00
	rs9368222	6	20794975	C/A	3.74(0.90)	3.25E-05	3.39(0.76)	7.68E-06	3.53(0.59)	1.70E-09	0.58	1.00	1.00
11p13	rs652722	11	31862110	C/T	3.73(0.89)	2.84E-05	2.75(0.77)	3.70E-04	3.15(0.59)	8.73E-08	0.61		
PAX6(78.3kb)	rs667156	11	31873795	T/C	3.86(0.89)	1.30E-05	2.82(0.82)	5.80E-04	3.28(0.61)	7.49E-08	0.61	1.00	1.00
	rs623102	11	31883012	C/A	3.79(0.89)	1.98E-05	2.71(0.77)	4.47E-04	3.16(0.59)	8.53E-08	0.61	0.95	0.98
	rs602907	11	31878590	T/C	3.87(0.88)	1.21E-05	2.61(0.77)	7.15E-04	3.14(0.59)	8.72E-08	0.61	1.00	1.00
	rs621420	11	31875774	A/G	3.89(0.89)	1.10E-05	2.62(0.77)	6.87E-04	3.15(0.59)	9.42E-08	0.61	1.00	1.00
	rs602074	11	31878738	C/T	3.91(0.89)	1.14E-05	2.58(0.77)	8.24E-04	3.13(0.59)	1.08E-07	0.60	0.98	1.00
	rs623312	11	31864040	G/A	3.64(0.89)	4.04E-05	2.76(0.77)	3.42E-04	3.13(0.59)	1.13E-07	0.61	1.00	1.00
	rs673946	11	31874175	G/A	3.85(0.89)	1.35E-05	2.59(0.77)	7.67E-04	3.11(0.59)	1.25E-07	0.61	1.00	1.00
	rs597660	11	31871790	G/T	3.78(0.89)	2.01E-05	2.63(0.77)	6.48E-04	3.11(0.59)	1.35E-07	0.61	1.00	1.00
	rs683028	11	31879151	A/G	3.73(0.89)	2.86E-05	2.66(0.77)	5.61E-04	3.10(0.59)	1.39E-07	0.61	0.98	1.00
	rs2440238	11	31864562	T/G	3.67(0.89)	3.57E-05	2.69(0.77)	4.77E-04	3.10(0.59)	1.46E-07	0.61	1.00	1.00
	rs666822	11	31873845	A/G	3.69(0.93)	6.77E-05	2.63(0.77)	6.50E-04	3.05(0.60)	3.80E-07	0.61	1.00	1.00
16p12.3	rs12597579	16	20165368	C/T	4.16(1.05)	7.13E-05	4.09(0.96)	2.07E-05	4.12(0.72)	9.27E-09	0.80		
GP2(63.9kb)													

β: Effect of SNPs per allele (in percentage) on BMI, derived from meta-analysis.

p: Derived from meta-analysis adjusted for for both study-specific inflation factors (for stage I, II) and also the estimated inflation factor for the stage I meta-analysis statistic (for stage I and II combined data).

p_H: p for homogeneity test across studies in stage I and II.

EAF: Effect allele frequency, estimated from stage I and II studies.

RSQ and D': r-squared and D' for the measurement of LD with the top SNP in Asians according to HapMap data.

Supplementary Table 10. Observed associations in stage I data for BMI and selected relevant SNPs

Loci	SNP	Chr	Position	Allele	β(SE)	p	EAF	
				Effect/Other			Asian	European
PCSK1	rs6235	5	95754654	G/C	1.45(1.01)	1.51E-01	0.31	0.27
PCSK1	rs6234	5	95754730	C/G	1.32(1.01)	1.89E-01	0.31	0.27
CDKAL1	rs4712526	6	20771014	T/A	3.32(0.88)	1.75E-04	0.57	0.69
GPR139/GP2	rs12598578	16	20161977	G/C	3.95(1.05)	1.63E-04	0.77	0.83

β: Effect of SNPs per allele (in percentage) on BMI, derived from meta-analysis.

p: Derived from meta-analysis adjusted for study-specific inflation factors.

EAF: Effect allele frequency, estimated from stage I studies for Asian and obtained from the HapMap data for Europeans.

2. SUPPLEMENTARY NOTE

2.1.Description of Participating Studies

2.1.1. Stage I – genome-wide association meta-analysis

Shanghai genome-wide association studies (SGWAS)

The Shanghai genome-wide association studies include participants of the Shanghai Breast Cancer Study (SBCS), Shanghai Endometrial Cancer Study (SECS), Shanghai Breast Cancer Survival Study (SBCSS), and Shanghai Women's Health Study (SWHS). The SBCS and SECS are population-based case-control studies, and the SBCSS and SWHS are ongoing population-based prospective cohort studies. All participants of these studies were recruited in Shanghai during the same time period using similar study protocols. Structured questionnaires including the same core questions were used to collected information on sociodemographic factors, reproductive history, lifestyle factors, and dietary habits. Anthropometrics, including weight, height, and waist and hip circumferences were taken by trained interviewers.

(1) The SBCS, described in detail elsewhere[1, 2], included two recruitment phases. During the initial phase (SBCS-I), 1,459 breast cancer patients and 1,556 controls were recruited between 1996 and 1998 through a rapid case-ascertainment system and the population-based Shanghai Cancer Registry. Blood samples were obtained from 1,193 (82%) cases and 1,310 (84%) controls. The second phase of subject recruitment (SBCS-II) was conducted between 2002 and 2005 using a protocol similar to the one used during the initial phase. A total of 1,989 incident cases and 1,918 community controls were recruited. The majority of cases (n = 1,932, 97.1%) and controls (n = 1,857, 96.8%) provided a blood sample or an exfoliated buccal cell sample. The age range of participants was 20-70 years with an average age of 50 years.

(2) The SECS, described in detail elsewhere[3], included 1,204 endometrial cancer cases and 1,212 controls aged 30-69 years. The cases were newly diagnosed and identified through the population-based Shanghai Cancer Registry between 1997 and 2003. The controls were randomly selected from the female residents of urban Shanghai through the Shanghai Resident Registry. Of the 1,204 cases and 1,212 controls, 857 cases and 837 controls donated a blood sample.

(3) The SBCSS, described in detail elsewhere[4], is a population-based cohort study that recruited a total of 5,042 breast cancer cases diagnosed between April 1, 2002 and December 31, 2006 approximately six months after cancer diagnosis. In-person interviews were conducted to collect information on known breast cancer risk factors and anthropometrics by using a protocol and questionnaire similar to that used in the SBCS. Buccal cell samples were collected from 96% of study participants using a modified mouthwash method. Because of a time overlap in subject recruitment for the SBCS-II and the SBCSS, 1,469 breast cancer patients participated in both studies. DNA samples from SBCSS cases were scanned and contributed to the current study.

(4) The SWHS, described in detail elsewhere[5], is a population-based cohort study of approximately 75,000 women who were aged 40 to 70 years at study enrollment and resided in the seven geographically defined communities; 56,832 (75.8%) provided a blood sample. After approximately 10 years of follow-up to date, multiple disease outcomes, including breast cancer and diabetes, have been observed.

Participants included in the Shanghai genome-wide association studies were selected from the four studies described above. They constitute three case groups and a control group. The three case groups were: (i) the breast cancer group included 2,909 cases (2,065 from the SBCS, 351 from the SBCSS, and 493 from the SWHS); (ii) the endometrial cancer group

included 830 cases from the SECS, and (iii) the diabetes case group included 886 newly diagnosed patients from the SWHS. The control group included 2,311 women (2,081 controls from the SBCS and 230 women from the SWHS) and served as the common control for three disease groups. Thus, a total of 6,936 women were included in the Shanghai genome-wide association studies, of whom 6,918 women with BMI measurement are included in the current analysis.

Anthropometric measurements taken at study recruitment for controls and for SWHS participants were used in the analysis. For cancer patients, self-reported weight at one year prior to diagnosis was used in the current analysis. Genomic DNA was isolated from peripheral blood for all participants except for 351 women from the SBCSS whose genomic DNA was extracted from buccal cells.

Genotyping methods and quality control (QC): Genomic DNA was extracted from buffy coats by using a Qiagen DNA purification kit (Valencia, CA) or Puregene DNA purification kit (Minneapolis, MN) according to the manufacturers' instructions and then used for genotyping assays. The GWAS genotyping was performed at the Vanderbilt Microarray Shared Resource (VMSR) and Affymetrix Research Services Laboratory (ARSL) using the Affymetrix Genome-Wide Human SNP Array 6.0 (Affy6.0) platform, following Affymetrix's protocols. In each of the 96-well plates for Affymetrix SNP 6.0 genotyping, three positive QC samples purchased from Coriell Cell Repositories (http://ccr.coreill.org/) were included. SNP data obtained from 227 positive quality control samples showed a very high concordance rate of called genotypes (mean, 99.85%; median, 100%). SNPs that showed genotyping call rates or concordance rate of less than 95% among the QC samples were excluded and the remaining samples were recalled by using Birdseed calling algorithm v2. In addition, a series of datasets were used to assess cross-

genotyping platform validation. These included the following sets of SNPs that are on the Affymetrix SNP Array 6.0 and had been genotyped previously using various platforms for a subset of subjects included in the GWAS scan: 1) 669 SNPs genotyped for 1,035 subjects by using Affymetrix Target Genotyping System; 2) 17 SNPs genotyped for 1,091 subjects by Taqman; and 3) 251 SNPs genotyped for 108 subjects by using Sequenom. These three sets of SNPs served as cross-platform sample verification during the laboratory process. The mean concordance rates were 99.5%, 98.5%, and 98.9% for Affymetrix Targeted Genotyping, Taqman, and Sequenom, when compared with the Affymetrix SNP Array 6.0.

Marker exclusion criteria: The following quality control criteria were applied to assure the data quality of each SNP: 1) MAF <0.01; 2) call rate <95% in both VMAR and ARSL data; 2) bad genotyping cluster; 3) concordance rate <95% among duplicated QC samples in both VMAR and ARSL data; 4) concordance rate ≥95% among duplicated QC samples within VMAR and within ARSL data, but <95% in combined VMSR and ARSL data; 5) significant difference in allele frequency (P<0.01) between the breast cancer cases genotyped in the VMSR and ARSL data and the breast cancer controls genotyped in VMSR and ARSL, and the differences were in the same direction in breast cancer cases and controls; 6) SNPs with a call rate <95% or QC concordance rate <95% or bad genotyping cluster within VMSR data were set to missing in the VMSR dataset. These criteria were also used for the ARSL dataset. After applying the QC filter, 690,947 SNPs remained for the analyses and were used for imputation.

Individual exclusion criteria: The gender of all study subjects was confirmed to be female according to the X chromosome genotyping data. Multidimensional scaling (MDS) analyses based on pairwise IBS showed that all subjects in the present study were clustered closely with HapMap Asians. We also excluded samples that had: 1) call rate <95%; 2) contaminated samples

(based on inbreeding coefficients), samples with mixed-up labels, or duplicated samples based on IBD estimate; 3) first-degree relatives, such as parent-offspring and full siblings based on IBD estimate. A total of 21 subjects were excluded. Other relationships (21 half-sibling/aunt-niece pairs and 26 first-cousins/grandchild-grandparent pairs) remained in the dataset.

Imputation methods. Genotypes were imputed using the program MACH (http://www.sph.umich.edu/csg/abecasis/MACH/download/), which determines the probability distribution of missing genotypes conditional on a set of known haplotypes, while simultaneously estimating the fine-scale recombination map. Phased autosome SNPs data from the HapMap Phase II Asian (release 22) were used as reference. To test for associations between the imputed SNP data with BMI, linear regression (additive model) was used, in which SNPs were represented by the expected allele count, an approach that takes into account the degree of uncertainty of genotype imputation (http://www.sph.umich.edu/csg/abecasis/MACH/download/).

The Korea Association Resource (KARE) project

The KARE project has been described previously[6]. This project was initiated in 2007 to undertake a large-scale GWAS. A total of 10,038 participants aged between 40 and 69 years were recruited through two population-based prospective cohort studies conducted in the Ansung (n=5,018) and Ansan (n=5,020) areas of South Korea. Both cohorts were designed to allow longitudinal prospective studies and adopted the same investigational strategy. More than 260 traits have been extensively examined through epidemiological surveys, physical examinations, and laboratory tests. Included in the current analysis were 8,838 participants with anthropometric measurement and genomic genotyping data.

Genotyping methods and quality control: The majority of genomic DNA genotyped on the Affymetrix Genome-Wide Human SNP array 5.0 were isolated from peripheral blood drawn from the Ansung and Ansan cohort participants. Where DNA samples for genotyping were inadequate, DNA extracted from Epstein-Barr virus-immortalized lymphoblastoid cell lines (LCL) was substituted. DNA samples with low concentration were amplified before genotyping according to the manufacturer's protocol (Qiagen). Genotyping performance was identical for these three DNA sources. Bayesian Robust Linear Modeling using the Mahalanobis Distance (BRLMM) Genotyping Algorithm was used for genotype calling of 500,568 SNPs. Samples with high missing genotype call rates (>4%), high heterozygosity (>30%), or gender inconsistencies, and those obtained from individuals who had developed any kind of cancer were excluded from subsequent analyses along with related or identical individuals whose computed average pairwise identity-by-state value was higher than that estimated from first-degree relatives of Korean sib-pair samples. Samples whose genotype-inferred sex disagreed with clinical records were re-tested for sex confirmation using the SNaPshot Multiplex System (Applied Biosystems). Markers with high missing gene call rates (>5%), low MAF (<0.01), or significant deviation from Hardy-Weinberg equilibrium ($P < 1 \times 10^{-6}$) were excluded.

Singapore Prospective Study Program (SP2)

Population: The Singapore Prospective Study Program (SP2) is a cross-sectional study of 6,968 adult Singaporean Chinese, Malay and Asian-Indian men and women, aged 24-95 years. Individuals who participated in previous cross sectional studies, the Thyroid and Heart Study 1982–1984 [7], National Health Survey 1992 [8], National University of Singapore Heart Study 1993–1995 [9], or National Health Survey 1998 [10], were invited to participate. All studies

involved a random sample of individuals from the Singapore population, aged 24 to 95 years, with disproportionate sampling stratified by ethnicity to increase the number of minority ethnic groups (Malays and Asian Indians). Subjects who were successfully re-contacted and gave informed consent answered a questionnaire and attended a clinic examination. Height (m) and weight (kg) were measured similarly in all datasets using standard protocols and were used to derive BMI as weight over height-squared (kg/m^2). Institutional review board approval was provided by the National Healthcare Group domain specific review board.

Singapore Diabetes Cohort Study (SDCS)

Population: The Singapore Diabetes Cohort Study is a research initiative led by the National University of Singapore together with the National Healthcare Group Polyclinics, National University Hospital Singapore, and Tan Tock Seng Hospital. Its primary aim is to identify genetic and environmental risk factors for diabetic complications, especially diabetic nephropathy, and to develop novel biomarkers for tracking disease progression. Since 2004, all adult (aged 18 years or more) type 2 diabetes patients (men and women) seen at the polyclinics and hospitals have been invited to participate in the cohort. Questionnaire and clinical data from consented patients were obtained together with bio-specimens (blood and urine archived at -80°C). The participation rate is more than 90% and to date, there are more than 5,000 patients in the SDCS [11]. Height (m) and weight (kg) were measured as described for the SP2 and were used to derive BMI (kg/m^2). This study was approved by the National University of Singapore Institutional Review Board.

Singapore Malay Eye Study (SiMES)

Population: SiMES is a population-based, cross-sectional study of Malay men and

women (N = 3,280), aged 40-80 years living in Singapore. Of the 4,168 eligible participants invited, 3,280 participated in the study (participation rate: 78.7%). Age-stratified random sampling of all Malay adults aged 40–80 years residing in 15 residential districts in the southwestern part of Singapore was performed. Height (m) and weight (kg) were measured as described for the SP2 and were used to derive BMI (kg/m^2). Details on the study participants and methods have been published previously[12]. This study was approved by the Singapore Eye Research Institute Institutional Review Board.

Genotyping methods and quality control of Singapore datasets: Genotyping assays for the SDCS and SP2 were conducted together as these studies were originally part of a Singaporean Chinese case-control study of type 2 diabetes. A total of 3,066 Chinese adults from the SP2 were genotyped using 1Mduov3 (N=1,016), HumanHap 610Quad (N=1,467), and Hap550 (N=583) and 2,210 Chinese adult diabetes cases from the SDCS were genotyped using 1Mduov3 (n=1,015) and HumanHap 610Quad (n=1,195) Beadchips® (http://www.illumina.com/). A total of 1,155 Chinese children from SCORM were genotyped using Illumina Hap550 (n=418) and Hap550Duo arrays (n=737), while all SiMES Malay adults (n=3,072) were genotyped using HumanHap 610Quad arrays.

Duplicate samples were plated for the SDCS, SP2 and SCORM studies. A total of 8 duplicate samples were available for the SDCS study, while a total of 198 duplicates were available for the SP2 study. The average SNP concordance rate between chips for the post-quality control (QC) duplicated samples was computed based on 531,805 post-QC common SNPs between 1Mduov3 and 610 Quad chips and 496,653 post-QC common SNPs between the 1Mduov3 and 550 chips. The mean concordance was >95%, and 5 discrepant SNPs were removed (rs10953303, rs11075260, rs1447826, rs274646, and rs430794). For the SCORM study,

19 duplicates on the Hap550 and Hap550Duo arrays had >98% concordance. For each array in each cohort, a first round of clustering was performed with the proprietary clustering files from Illumina (GenCall). Samples achieving a 99% call rate were subsequently used to generate local clusterfiles (GenTrain), which were used for a final round of genotype calling. A threshold of 0.15 was implemented on the GenCall score to decide on the confidence of the assigned genotypes.

Samples were removed based on the following conditions: sample call rates of less than 95%, excessive heterozygosity, cryptic relatedness, discordant ethnic membership, or gender discrepancy. Bivariate plots of sample call rates and heterozygosity, defined as the proportion of heterozygous calls of all valid autosomal genotypes in an individual, were used to assess the overall distribution of missingness and heterozygosity across all the samples. Identity-by-state measures were performed by pair-wise comparison of samples to detect cryptic relatedness such as monozygotic twins, full-sibling pairs, and parent-offspring pairs. One sample from each relationship was excluded from further analysis and where duplicate samples had been genotyped in different SNP-arrays, samples from the denser array was retained. Population structure ascertainment to prevent confounding of study results was performed by using principal component analysis (PCA) with 4 panels from the International HapMap Project (http://hapmap.ncbi.nlm.nih.gov) and the Singapore Genome Variation Project (http://www.nus-cme.org.sg/SGVP) with a thinned set of SNPs to reduce linkage disequilibrium (LD). Individuals who showed ethnic membership discordant from their self-reported ethnicity were excluded from the analysis. For the SiMES Malays that showed a continuous cloud suggesting some degree of admixture, the first two principal components were useful for correction of population structure in association testing in SiMES. The Eigenstrat PCA program [13] with a threshold set at eight

standard deviations was used to remove samples that were possible outliers in the SCORM dataset. Subsequently, samples with gender discrepancies between the genetically inferred gender from Beadstudio and clinic-reported gender were identified and removed. A total of 1,006 SCORM samples, 2,431 SP2 samples, 1,992 SDCS samples, and 2,522 SiMES samples with BMI data were available after sample QC procedures.

We excluded sex and mitochondrial SNPs, together with gross Hardy-Weinberg equilibrium (HWE) outliers (p-value $<1x10^{-4}$). SNPs that were monomorphic or rare (minor allele frequency (MAF) <1%) and SNPs with low call-rates (<95%) were also excluded. Where more than one chip was used for genotyping, Mantel-extension tests were carried out to detect differences in allele frequencies of SNPs between the chips; 62 such SNPs were detected in the SP2 and 69 such SNPs were detected in the SDCS and removed from the analyses.

Imputation methods: Imputation procedures were performed using IMPUTE v0.5.0. [14] and genotype calls were based on HapMap Phase 1 and 2 East-Asian samples (CHB and JPT) of NCBI build 36 for all Chinese samples (SCORM, SP2, SDCS) [15]. For the Malay and Asian-Indian samples, all HapMap reference panels (CEU, YRI and JPT+CHB) on NCBI build 36 were used for imputation to better capture local patterns of haplotype variations [16]. Actual genotyped calls were placed back into the files and only imputed SNPs, with a posterior probability ≥0.90 and call-rate ≥95% were used. A total of 1,816,934 SNPs from SCORM, 1,555,870 SNPs from SiMES, 1,745,788 SNPs from SP2, and 1,791,569 SNPs from SDCS were available for subsequent analyses after imputation and QC procedures.

Cardio-metabolic Genome Epidemiology (CAGE)

The Cardio-metabolic Genome Epidemiology (CAGE) Network is an ongoing collaborative

effort to investigate genetic and environmental factors and their interactions affecting cardiometabolic traits/disorders among Asians, including the Japanese[17, 18]. CAGE participants were recruited in a population-based or hospital-based setting, depending on the design of the member studies. Participation rates varied among the member studies (approximately from 25% in the community-based survey to 80% in the work place-based survey). From this network sample, a total of 2,194 Japanese samples (1,394 men and 800 women; age range 32 –95, median 66 years) were used for a genome-wide association study (GWAS) of measures of obesity, such as body mass index. These participants were enrolled at four separate sites in Japan, the Tokyo, Nagoya, Osaka, and Shimane districts. Participants' height and body weight were measured by trained personnel using standard anthropometric techniques. In the CAGE Network, all participants provided written informed consent and studies were approved by local Research Ethics Committees and/or Institutional Review boards.

Genotyping methods and quality control: Genotyping was performed with Infinium HumanHap550/Human610-Quad BeadArray (Illumina, San Diego, CA, USA), which interrogated 550K/610K SNPs, according to the manufacturer's protocol. This set of SNP markers reportedly captures 87% of common SNPs with an LD coefficient of $r^2 > 0.8$ in the HapMap JPT and CHB populations (according to the manufacturer's brochure). Assay accuracy and reproducibility were measured by using DNA from CEU samples genotyped as part of the HapMap project [http://www.hapmap.org]. Genotype calling was performed using BeadStudio software (Illumina) and genotype calls with a 'GenCall' Score <0.53 were dropped from the analysis. The GenCall Score measures the reliability of genotype calls based on the clustering of dye intensities (www.illumina.com/downloads/GenCallTechSpotlight.pdf).

QC of SNPs and samples was performed as previously described[18]. Briefly, data cleaning and analysis were performed using PLINK software (version 1.06) [19]. Among the assayed SNPs, we excluded SNPs for the following: 1) genotype call rate <0.95; 2) significant ($P<10^{-6}$) deviation from HWE; or 3) MAF, <0.01. The remaining 451,382 SNPs were analyzed in the genome scan. The average call rate for the 451,382 QC'd SNPs was 99.7% in 2,194 samples tested for an association with obesity.

Taiwan Genome Wide Association Study (TGWAS):

This study included 1,000 random samples from the Han-Chinese Cell and Genome Bank in Taiwan and 999 participants from the Taiwan type 2 diabetes study [20]. The Han-Chinese Cell and Genome Bank in Taiwan[21] includes more than 3,300 healthy controls who were recruited via a stratified, 3-staged probability clustering sampling scheme through a registry of all the 329 non-aboriginal townships or city districts in Taiwan. The study was approved by the institutional review board of Academia Sinica, Taiwan and written, informed consent was obtained from all participants. The Taiwan type 2 diabetes study included a total of 2,798 unrelated individuals (1,440 males and 1,358 females, participation rate ~60%) with type 2 diabetes, aged >20 years[20]. All of the type 2 diabetes cases were diagnosed according to medical records and fasting plasma glucose levels by using American Diabetic Association criteria. The study was approved by the institutional review board and the ethics committee of each institution. Written, informed consent was obtained from each participant in accordance with institutional requirements and Declaration of Helsinki Principles. For this meta-analysis, 999 subjects recruited by the China Medical University Hospital, Taichung, Taiwan, were analyzed.

Genotyping and QC: Genomic DNA was extracted from peripheral blood using the Puregene DNA isolation kit (Gentra Systems, Minneapolis, MN, USA). Whole genome genotyping using the Illumina HumanHap550-Duo BeadChip was performed by deCODE Genetics (Reykjavík, Iceland). Genotype calling was performed using the standard procedure implemented in BeadStudio (Illumina, Inc., San Diego, CA, USA), with the default parameters suggested by the platform manufacturer. QC of genotype data was performed as described in detail elsewhere[20, 22]. In brief, for each sample genotyped in this study, the average call rate was 99.92±0.12%. After applying stringent QC criteria, high-quality genotypes for 516,737 SNPs (92.24%) were obtained, with an average call rate of 99.92±0.24%. SNPs were excluded if they: 1) were nonpolymorphic among both cases and controls, 2) had a total call rate <95% among cases and controls combined, 3) had a MAF <5% and a total call rate <99% among cases and controls combined, or 4) had significant deviation from HWE among the controls ($P<10^{-7}$).

Cardiometabolic Risk in Chinese (CRC) study:

The Cardiometabolic Risk in Chinese (CRC) Study is a community-based health examination survey of 6,431 individuals (aged 18-93 years; 53.7% men) who were randomly selected from residents living in the urban area of Xuzhou, China in 2009. Written consents were obtained from all the participants. The study was reviewed and approved by the ethics committee of the Central Hospital of Xuzhou, Affiliated Hospital of Medical School of Southeast University, Nanjing, China. A total of 811 study samples were included in a GWAS that was carried out on Illumina Human660-Quad BeadChips at the Chinese National Human Genome Center in Shanghai, China. Genotype clustering was conducted with Illumina BeadStudio 3.3 software. Height was measured to the nearest 0.5 cm without shoes and body weight was

measured to the nearest 100 grams without shoes. BMI was calculated as weight (in kilograms) divided by height (in meters) squared.

2.1.2. Stage II – *in silico* Replication Studies

BioBank Japan (RIKEN)

The replication study by BioBank Japan ncludes 27,284 subjects enrolled in the BioBank Japan Project[23] at the Institute of Medical Science, the University of Tokyo. These subjects were recruited from 12 medical institutes in Japan including Osaka Medical Center for Cancer and Cardiovascular Diseases, The Cancer Institute Hospital of Japanese Foundation for Cancer Research, Juntendo University, Tokyo Metropolitan Geriatric Hospital, Nippon Medical School, Nihon University School of Medicine, Iwate Medical University, Tokushukai Hospitals, Shiga University of Medical Science, Fukujuji Hospital, National Hospital Organization Osaka National Hospital, and Iizuka Hospital. Subjects with dialysis treatment, or who were determined to be of non-Japanese origin by self-report or by PCA or our previous studies[24, 25] were not excluded. Clinical information of the subjects including age (mean ± SD; 62.9 ± 12.0), gender (46.6 % for female), and BMI (mean ± SD; 22.8 ± 3.59) were collected by a standard questionnaire. All participants provided written informed consent as approved by the ethical committees of the RIKEN Yokohama Institute and the Institute of Medical Science, University of Tokyo.

Genotyping and quality control: We used the data of GWAS for 29 diseases which were performed for the BioBank Japan Project[24] In the GWAS, genotyping was performed using the Illumina HumanHap610-Quad Genotyping BeadChip (Illumina, CA, USA). After excluding the subjects with call rates lower than 0.98, we excluded SNPs with call rates lower than 0.99, SNPs

with ambiguous call, or non-autosomal SNPs. We excluded closely related subjects by using identity-by-state (IBS). For each pair with 1st or 2nd degree of kinship, we excluded the member of the pair with lower call rates. We also excluded subjects whose ancestries were estimated to be distinct from the other subjects by using PCA performed by EIGENSTRAT version 2.0[13]. We performed PCA for the genotype data of our study along with the genotype data of unrelated European (CEU), African (YRU), and East-Asian (Japanese and Han Chinese; JPT + CHB) individuals obtained from the Phase II HapMap database (release 24)[26]. Based on the PCA plot, we excluded the outliers in terms of ancestry from JPT + CHB clusters. We then excluded the SNPs with MAF < 0.01 or the SNPs with exact P-value of the Hardy-Weinberg equilibrium test < 1.0×10^{-7}.

Genotype imputation was performed using MACH 1.0[27] in two step procedure. The JPT and CHB individuals obtained from Phase II HapMap database (release 24)[26] were used as references. In the first step, recombination and error rate maps were estimated using 500 subjects randomly selected from the GWAS data. In the second step, genotype imputation of all subjects was conducted using the rate maps estimated in the first step. We excluded the imputed SNPs with MAF < 0.01 or Rsq values < 0.7.

The Genetic Epidemiology Network of Salt Sensitivity (GenSalt)

The GenSalt study participants were recruited from six sites in rural areas of northern China from October 2003 to July 2005[28]. The selection of these study sites was based on the homogeneity of the study population with regard to ethnicity and environmental exposures, including lifestyle, nutritional factors, and habitual dietary intake. The residents in these regions are of the Han ethnicity, the ethnic majority in China. A community-based blood pressure

screening was conducted among persons aged 18-60 years in the study villages to identify potential probands and their families for the study. Those with a mean systolic blood pressure between 130-160 mmHg and/or diastolic blood pressure between 85-100 mmHg and no use of antihypertensive medications and their spouses, siblings, and offspring were recruited as volunteers for a dietary intervention study. In general, individuals who had stage-2 hypertension, secondary hypertension, use of antihypertensive medications, history of clinical cardiovascular disease, diabetes, chronic kidney disease, pregnancy, or heavy alcohol use were excluded from the study. A total of 1,906 individuals (1,010 men and 896 women) met the eligibility criteria for the dietary intervention study[28]. Of these individuals, 1,843 (96.7%) completed the entire 21-day dietary intervention and included in the GWAS. The completeness of the study questionnaire data, blood pressure and anthropometric data, blood and urine sample collection is near 100%[28]. The institutional review boards at all participating institutes approved the study, and written, informed consent was obtained from each participant.

A standard questionnaire was administered by trained staff at the baseline examination to collect information on demographic characteristics, personal and family medical history, and lifestyle risk factors (including cigarette smoking, alcohol consumption, and physical activity). Three blood pressure measurements were obtained each morning during the 3-day baseline examination by trained and certified observers using a random–zero sphygmomanometer according to a standard protocol [28]. Blood pressure was measured with the participant in a seated position after 5 minutes of rest. In addition, participants were advised to avoid consumption of alcohol, coffee or tea, or cigarettes and exercise for at least 30 minutes prior to their blood pressure measurements. Body weight, height, and waist circumference were measured twice with the participant in light indoor clothing without shoes. Waist circumference was measured one cm

above the participant's navel during minimal respiration. Overnight (≥8 hours) fasting blood specimens were obtained for measurement of glucose and lipids. Plasma glucose was measured using a modified hexokinase enzymatic method (Hitachi automatic clinical analyser, model 7060, Japan). Concentrations of total cholesterol, HDL-cholesterol, and triglycerides were assessed enzymatically using commercially available reagents. Concentration of LDL-cholesterol was calculated by means of the Friedewald equation for participants who had less than 400 mg/dL triglycerides: LDL cholesterol=total cholesterol–HDL cholesterol–triglycerides/5.

Lymphocytic DNA samples were obtained from GenSalt family members (probands, parents, spouses, siblings, and offspring). Genome-wide SNPs were genotyped using Affymetrix® Genome-Wide Human Array 6.0 at the Affymetrix genotyping facility. After removing sex-linked SNPs, mitochondrial SNPs, and 'unassigned' SNPs that had no annotated chromosomal location, 871,166 SNPs were chosen for examination. Strict procedures for extensive QC were used to check the data for any obvious errors, remove uninformative data, and find and remove all Mendelian errors in three stages. In stage 1, we removed subjects with gender discrepancies between reported sex and that estimated by PLINK[19] and those who had potential pedigree errors, as determined by GRR [29]. In stage 2, we removed monomorphic SNPs, Affymetrix 'housekeeping' SNPs, SNPs with missing rates of >25% or MAF of <1%. In the final stage, Mendelian errors were found and removed using PLINK[19] and PedCheck [30]. After the QC process, 820,015 autosomal SNPs from 1,881 subjects remained. An additional 1,792,556 SNPs were imputed from a HapMap reference panel using 90 subjects from the JPT and CHB populations. The QC processes removed imputed SNPs with R^2<0.3, MAF<1%, Hardy-

Weinberg p-value $<10^{-6}$, or Mendelian errors. Finally, 2,216,774 autosomal SNPs were used for GWAS analyses.

To estimate the association between BMI and SNPs, a mixed linear model was used for association test and the family structure was taken into account by treating it as a random effect.

Singapore Cohort study Of the Risk factors for Myopia (SCORM)

SCORM [15] is a prospective study consisting of 1,979 boys and girls in grades 1-3 (aged 7-9 years) who were recruited from three schools in Singapore (located in the eastern, western and northern regions). Children with serious medical conditions such as heart disorders and leukemia were excluded from the study (N=94). Participants were followed up yearly and height (m) and weight (kg) were measured as described for the SP2 and were used to derive BMI (kg/m^2). For the purpose of this study, BMI as calculated from measurements taken at the visits when the children were 9 years of age were used to minimize pubertal effects of later pediatric age. Ethics approval was granted by the institutional review boards of the Singapore Eye Research Institute and the National University of Singapore. Data and samples were collected only after informed consent was given by the participants' parents/guardians. The genotyping and quality control protocols are same as described for the SP2.

Suita Study

The Suita study was based on the residents of Suita City, an urban city located in Osaka, the second largest area of Japan[31]. The sample consisted of 14,200 men and women (aged 30–79 years at enrollment), stratified by gender and 10-year age groups (10 groups with 1,420 participants in each group) who had been randomly selected from the municipal population

registry. All participants were invited by letter to attend regular follow-up examinations (every 2 years). Among initially included 5,098 participants, DNA samples were available for 3,310 participants 3,228 participants. Subjects (n=933, age range: 42–73 years) who were not receiving medication for hypertension were selected for the GWAS. (Illumina Sentrix Human Hap550 BeadChipAll). After exclusions for deviations from Hardy–Weinberg equilibrium ($P < 10^{-6}$), minor allele frequency < 0.01%, and SNP call rate <0.95, 482,162 markers were used for imputation. Imputation was performed with MACH (v. 1.0.16) using phased CHB+JPT haplotypes from HapMap Phase 2 (release 22). All study procedures were approved by the ethics committee of National Cerebral and Cardiovascular Center. Informed consent was obtained from all participating subjects.

Shanghai Institute of Hypertension (SIH) GWAS

This is a case-control GWAS investigating the genetic variants associated with hypertension. The study included a total of 446 hypertension cases (224 men/222 women) and 455 controls (230 men, 225 women) aged 37-84 years, which were Chinese Han origin residing in the Shanghai metropolitan area and gave written informed consent to donating blood samples for genetic analysis and related assays. This study was approved by the Ethics Committee of Ruijin Hospital. To be eligible, cases had to meet the following criteria: 1) have at least two family members with hypertension; 2) with a systolic blood pressure measurement of >140 mmHg and/or a diastolic blood pressure measurement of >90 mmHg (an average of three blood pressure measurements was used) or using anti-hypertension medication for long term; 3) diagnosed with hypertension at age >30 years; 4) have a fasting blood glucose level of <7 mmol/L and no family members with a diagnosis of diabetes; and 5) have a BMI <27 kg/m^2.

Controls had to meet the following eligibility criteria: 1) normal blood pressure measurement (systolic blood pressure <=120 mmHg and diastolic blood pressure <=80 mmHg); 2) be aged >40 years; and 3) have a BMI <27 kg/m^2.

Genotyping and QC: Participants were genotyped using the Illumina Human-Hap 610K BeadChip platform. We removed individuals with genotyping call rates of less than 98% and kept the higher call rate for duplicate samples (IBS>99.99%) or first-degree relatives (4 pairs identified as IBD>0.45). Population structure was evaluated using principal component analysis and the study sample conformed with HapMap3 CHB+CHD populations. After QC procedures and using the more stringent case-control definition used by the AGEN consortium, 384 cases (196 men/188 women) and 337 controls (203 men/134 women) were included. Markers from chrY and chrM were excluded. We required markers on autosomes and chrX to pass the following criteria: 1) have a MAF >1% ; 2) have missing genotypes <3% per individual (and <1% for SNPs with MAF <5%); 3) HWE p >1e−5 in controls. After applying these QC filters, 502,847 high-quality SNPs remained. To facilitate imputation, we aligned markers to the HapMap CHB+JPT panel (Phased, Release 22) to make them appear on the same strand. Markers were excluded for the following reasons: 1) alleles inconsistent with HapMap or 2) allele frequency differences with the HapMap JPT+CHB panel >20%, leaving 502,700 markers for imputation. MACH 1.0.16 was used to impute genotypes using HapMap CHB+JPT Phased genotypes (Release 22) as the reference. We only kept SNPs with a MAF >1% and with MACH RSQR >0.3 for association analyses. A total of 2,245,153 genotyped and imputed SNPs were available for the final data set. The genomic inflation factors were 1.01 for genotyped markers of systolic blood pressure, diastolic blood pressure, and hypertension and imputed markers of

systotlic blood pressure and diastolic blood pressure; 1.02 for imputed markers of hypertension. All results are reported without genome control correction.

The Multiethnic Cohort (MEC)

Population: The MEC includes 215,251 men and women aged 45-75 years at recruitment from Hawaii and California [32]. The cohort was assembled in 1993-1996 by mailing a self-administered, 26-page questionnaire to persons identified primarily through driver's license files. Identification of incident cancer cases occurred by regular linkage with the Hawaii Tumor Registry and the Los Angeles County Cancer Surveillance Program; both NCI-funded Surveillance, Epidemiology, and End Results registries. Collection of biospecimens from incident breast and prostate cases began in California in 1995 and in Hawaii in 1997 and a biorepository was established between 2001 and 2006, which includes 67,000 MEC participants. The participation rates for providing a blood sample are >60%. Linkage with the tumor registries is complete through December 31, 2008 in Hawaii and December 31, 2007 in California. The case-control study of breast cancer among Japanese-ancestry participants of the MEC has been conducted and includes 889 cases and 830 control. In addition, a case-control study of prostate cancer among Japanese-ancestry participants of the MEC has been conducted and includes 1,033 cases and 1,042 controls. Genotyping data from these participants of these two studies are included in the current study.

Genotyping methods and quality control: Genotyping of breast cancer samples was conducted using the Illumina Infinium 660W bead array at the University of Southern California. Genotyping of the prostate cancer samples was conducted using the Illumina Infinium 660W bead array at the Broad Institute. Following genotyping, samples within each study were

removed based on the following exclusion criteria: 1) unknown replicates across studies; 2) call rates <95%; 3) samples with >10% mean heterozygosity on the X chromosome and/or <10% mean intensity on the Y chromosome; 4) ancestry outliers (<5% African ancestry as determined by STRUCTURE), and; 5) samples that were related (>50% correlation as determined by an empirical kinship matrix). To assess genotyping reproducibility, we included 138 replicate samples; the average concordance rate was 99.95% (≥99.93% for all pairs). Of the 7 SNPs examined, 2 (rs652722 and rs11671664) were directly genotyped. The call rates for these two SNPs were >99.9% in both studies and HWE p-values were ≥ 0.25 for these SNPs.

The analysis of these 7 SNPs in stage 3 was conducted among 876 Japanese breast cancer cases and 826 controls and among 980 Japanese prostate cancer cases and 1,005 controls.

Imputation: We carried out genome-wide imputation using the software MACH [27]. Phased haplotype data from the founders of the Japanese (JPN) HapMap Phase 2 samples were used to infer LD patterns in order to impute ungenotyped markers. The Rsq metric, defined as the observed variance divided by the expected variance, provides a measure of the quality of the imputation at any SNP. Of the 7 SNPs examined in the replication phase, 5 were imputed (rs6545814, rs261967, rs9356744, rs4776970 and rs12597579). The Rsq metric for these SNPs was ≥0.992.

Fangchenggang Area Male Health and Examination Survey (FAMHES) study

Design, Population Recruitment, and Sample Size: The FAMHES is a population-based study conducted among non-institutionalized Chinese men aged 17 to 88 years in Guangxi and was designed to investigate the effects of environmental and genetic factors and their interaction with the development of age-related chronic disease. A comprehensive demographic and health

survey was conducted among 4,303 continuous men who participated in a large-scale physical examination in Fangchenggang First People's Hospital Medical Centre from September 2009 to December 2009. All participants provided written informed consents and the study received local ethics committee approval. The current study was confined to men aged 20 to 69 years. All participants reported themselves to be of Chinese southern Han origin and to be free of diabetes mellitus, coronary heart disease, stroke, hyperthyroidism, rheumatoid arthritis, cancer, and impaired hepatic or renal function. For the association study, 2,018 men with anthropometric measurements were included.

Source of anthropometric information: Anthropometric measurements were performed by trained personnel using a standardized protocol. Body weight and height were measured without shoes to the nearest 0.1 kg and 0.1 cm, respectively. Body mass index (BMI) was then calculated as weight (kg)/height (m^2). Waist circumference was measured midway between the lowest rib and the iliac crest to the nearest 0.1 cm, and hip circumference was taken over the widest part of the gluteal region. The waist to hip ratio (WHR) was then calculated as waist circumference (cm)/hip circumference (cm).

Genotyping platform: The study populations were genotyped using the Illumina HumanOmni1-Quad BeadChip (Illumina, CA, USA). All of these experiments were performed at the same institute by the same technical staff. 19 individuals were excluded for suspected first-degree relative of an included individual based on genome-wide genotype data. Individuals with call rate <95% were excluded from analysis. SNPs with HWE P< 10-4 or call rate <95% or MAF <1% were excluded from analysis.

2.1.3. Stage III – *de novo* Replication Study

Vanderbilt University site

The *de novo* genotyping conducted in Vanderbilt University included 10,602 Chinese women and 1,444 Chinese men, a total of 12,046 individuals. The women included 3,975 breast cancer cases from SBCS, SBCSS, 864 type 2 diabetes cases from SWHS, and 5,763 controls from SBCS, SECS, and SWHS. These studies were described in the above description of the SGWAS. The men included 722 type 2 diabetes cases and 722 age-matched controls selected from participants of Shanghai Men's Health Study (SMHS)[33], a population-based cohort study of approximately 61,000 men who were aged 40 to 74 years at study enrollment and resided in the seven geographically defined communities. Anthropometrics, including weight, height, and waist and hip circumferences were taken by trained interviewers in all those parent studies. The individuals selected for *de novo* genotyping were independent of the subjects in the SGWAS. The genotyping was performed using the Sequenom iPLEX MassARRAY® .

National University of Singapore site

A total of 2,701 individuals recruited in three cohorts were included in the *de novo* replication study. Design of these cohort studies are described below.

i. Multi-ethnic cohort (MEC)

The Multi-ethnic cohort (MEC) is a population-based prospective cohort that is part of the Singapore Consortium of Cohort Studies. Participants in the cohort completed a questionnaire, providing information on demographics, socio-economic status and medical history. They were subsequently invited to attend a clinical examination where anthropometric measures, blood pressure, fasting blood glucose and lipids measurements are collected. Height (m) and weight (kg) was measured similarly in all datasets using standard protocols and used to derive BMI as

weight over height-squared (kg/m^2). Participants gave broad consent for i) future biomedical research, ii) access to their medical records and iii) linkages to various registries. 1,200 ethnic Malay samples from this cohort were used for *de novo* genotyping using the Sequenom iPLEX MassARRAY® platform.

ii. Singapore Cardiovascular Cohort Study (SCCS2)

The Singapore Cardiovascular Cohort Study is a population-based prospective study combining two cross-sectional surveys – the 1992 National Health Survey[8] and the National University of Singapore Heart Study (1993–1995) [9] and has been previously described [34]. Both these surveys were of a random sample of all Singapore residents, with disproportionate sampling by ethnic group to increase the number of Malays and Asian Indians relative to Chinese. Height (m) and weight (kg) measures and derivation of BMI (kg/m^2) was performed as indicated above. For this study, genotyping using the Sequenom iPLEX MassARRAY® platform was performed on Chinese (N=583) and Malay (N=197) samples only.

iii. Singapore Prospective Study Program (SP2):

The Singapore Prospective Study Program (SP2), described above, contributed 721 Malay samples for *de novo* genotyping of SNPs using the Sequenom iPLEX MassARRAY® platform.

Ehime University site

The Nomura study of Ehime University is a longitudinal epidemiological study based on the Nomura Town residents of a community of 11,000 inhabitants, a largely rural community located in Ehime Prefecture, Japan[35]. A total of 2,895 participants (1,256 men and 1,639 women)

were recruited through a community-based annual medical check-up process for self-employees, including farmers and foresters, employees of small companies, and elderly without fixed employment. Body height (cm) and weight (kg) were measured at the medical check-up. Genotyping was performed with the TaqMan probe assay (Applied Biosystems Co., Ltd., Foster City, CA) using commercially available primers and probes purchased from the Assay-on-Demand system.

2.2. Author List from the GIANT Consortium

Association analyses of 249,796 individuals reveal eighteen new loci associated with body mass index (Speliotes et al Nature Genetics 2010)

Elizabeth K. Speliotes**,** Cristen J. Willer, Sonja I. Berndt, Keri L. Monda, Gudmar Thorleifsson, Anne U. Jackson, Hana Lango Allen, Cecilia M. Lindgren**,** Jian'an Luan, Reedik Mägi, Joshua C. Randall, Sailaja Vedantam**,** Thomas W. Winkler, Lu Qi**,** Tsegaselassie Workalemahu, Iris M. Heid**,** Valgerdur Steinthorsdottir, Heather M. Stringham, Michael N. Weedon, Eleanor Wheeler, Andrew R. Wood, Teresa Ferreira, Robert J. Weyant, Ayellet V. Segrè**,** Karol Estrada**,** Liming Liang**,** James Nemesh, Ju-Hyun Park, Stefan Gustafsson, Tuomas O. Kilpeläinen, Jian Yang, Nabila Bouatia-Naji**,** Tõnu Esko**,** Mary F. Feitosa, Zoltán Kutalik**,** Massimo Mangino, Soumya Raychaudhuri**,** Andre Scherag, Albert Vernon Smith**,** Ryan Welch, Jing Hua Zhao, Katja K. Aben, Devin M. Absher, Najaf Amin, Anna L. Dixon, Eva Fisher, Nicole L. Glazer, Michael E. Goddard**,** Nancy L. Heard-Costa, Volker Hoesel, Jouke-Jan Hottenga, Åsa Johansson**,** Toby Johnson**,** Shamika Ketkar, Claudia Lamina**,** Shengxu Li, Miriam F. Moffatt, Richard H. Myers, Narisu Narisu, John R.B. Perry, Marjolein J. Peters**,** Michael Preuss, Samuli Ripatti**,** Fernando Rivadeneira**,** Camilla Sandholt, Laura J. Scott, Nicholas J. Timpson, Jonathan P. Tyrer, Sophie van Wingerden, Richard M. Watanabe**,** Charles C. White, Fredrik Wiklund, Christina Barlassina, Daniel I. Chasman**,** Matthew N. Cooper, John-Olov Jansson, Robert W. Lawrence, Niina Pellikka**,** Inga Prokopenko**,** Jianxin Shi, Elisabeth Thiering, Helene Alavere, Maria T. S. Alibrandi, Peter Almgren, Alice M. Arnold**,** Thor Aspelund**,** Larry D. Atwood, Beverley Balkau**,** Anthony J. Balmforth, Amanda J. Bennett, Yoav Ben-Shlomo, Richard N. Bergman, Sven Bergmann**,** Heike Biebermann, Alexandra I.F. Blakemore, Tanja Boes, Lori L. Bonnycastle, Stefan R. Bornstein, Morris J. Brown, Thomas A. Buchanan**,** Fabio Busonero, Harry Campbell, Francesco P. Cappuccio, Christine Cavalcanti-Proença**,** Yii-Der Ida Chen, Chih-Mei Chen, Peter S. Chines, Robert Clarke, Lachlan Coin, John Connell, Ian N.M. Day, Martin den Heijer**,** Jubao Duan, Shah Ebrahim**,** Paul Elliott**,** Roberto Elosua, Gudny Eiriksdottir, Michael R. Erdos, Johan G. Eriksson**,** Maurizio F. Facheris**,** Stephan B. Felix, Pamela Fischer-Posovszky, Aaron R. Folsom, Nele Friedrich, Nelson B. Freimer, Mao Fu, Stefan Gaget**,** Pablo V. Gejman, Eco J.C. Geus, Christian Gieger, Anette P. Gjesing, Anuj Goel**,** Philippe Goyette, Harald Grallert, Jürgen Gräßler, Danielle M. Greenawalt, Christopher J. Groves, Vilmundur Gudnason**,** Candace

Guiducci, Anna-Liisa Hartikainen, Neelam Hassanali, Alistair S. Hall, Aki S. Havulinna, Caroline Hayward, Andrew C. Heath, Christian Hengstenberg, Andrew A. Hicks, Anke Hinney, Albert Hofman, Georg Homuth, Jennie Hui, Wilmar Igl, Carlos Iribarren, Bo Isomaa, Kevin B. Jacobs, Ivonne Jarick, Elizabeth Jewell, Ulrich John, Torben Jørgensen, Pekka Jousilahti, Antti Jula, Marika Kaakinen, Eero Kajantie, Lee M. Kaplan, Sekar Kathiresan, Johannes Kettunen, Leena Kinnunen, Joshua W. Knowles, Ivana Kolcic, Inke R. König, Seppo Koskinen, Peter Kovacs, Johanna Kuusisto, Peter Kraft, Kirsti Kvaløy, Jaana Laitinen, Olivier Lantieri, Chiara Lanzani, Lenore J. Launer, Cecile Lecoeur, Terho Lehtimäki, Guillaume Lettre, Jianjun Liu, Marja-Liisa Lokki, Mattias Lorentzon, Robert N. Luben, Barbara Ludwig, MAGIC, Paolo Manunta, Diana Marek, Michel Marre, Nicholas G. Martin, Wendy L. McArdle, Anne McCarthy, Barbara McKnight, Thomas Meitinger, Olle Melander, David Meyre, Kristian Midthjell, Grant W. Montgomery, Mario A. Morken, Andrew P. Morris, Rosanda Mulic, Julius S. Ngwa, Mari Nelis, Matt J. Neville, Dale R. Nyholt, Christopher J. O'Donnell, Stephen O'Rahilly, Ken K. Ong, Ben Oostra, Guillaume Paré, Alex N. Parker, Markus Perola, Irene Pichler, Kirsi H. Pietiläinen, Carl G.P. Platou, Ozren Polasek, Anneli Pouta, Suzanne Rafelt, Olli Raitakari, Nigel W. Rayner, Martin Ridderstråle, Winfried Rief, Aimo Ruokonen, Neil R. Robertson, Peter Rzehak, Veikko Salomaa, Alan R. Sanders, Manjinder S. Sandhu, Serena Sanna, Jouko Saramies, Markku J. Savolainen, Susann Scherag, Sabine Schipf, Stefan Schreiber, Heribert Schunkert, Kaisa Silander, Juha Sinisalo, David S. Siscovick, Jan H. Smit, Nicole Soranzo, Ulla Sovio, Jonathan Stephens, Ida Surakka, Amy J. Swift, Mari-Liis Tammesoo, Jean-Claude Tardif, Maris Teder-Laving, Tanya M. Teslovich, John R. Thompson, Brian Thomson, Anke Tönjes, Tiinamaija Tuomi, Joyce B.J. van Meurs, Gert-Jan van Ommen, Vincent Vatin, Jorma Viikari, Sophie Visvikis-Siest, Veronique Vitart, Carla I. G. Vogel, Benjamin F. Voight, Lindsay L. Waite, Henri Wallaschofski, G.Bragi Walters, Elisabeth Widen, Susanna Wiegand, Sarah H. Wild, Gonneke Willemsen, Daniel R. Witte, Jacqueline C. Witteman, Jianfeng Xu, Qunyuan Zhang, Lina Zgaga, Andreas Ziegler, Paavo Zitting, John P. Beilby, I. Sadaf Farooqi, Johannes Hebebrand, Heikki V. Huikuri, Alan L. James, Mika Kähönen, Douglas F. Levinson, Fabio Macciardi, Markku S. Nieminen, Claes Ohlsson, Lyle J. Palmer, Paul M. Ridker, Michael Stumvoll, Jacques S. Beckmann, Heiner Boeing, Eric Boerwinkle, Dorret I. Boomsma, Mark J. Caulfield, Stephen J. Chanock, Francis S. Collins, L. Adrienne Cupples, George Davey Smith, Jeanette Erdmann, Philippe Froguel, Henrik Grönberg, Ulf Gyllensten, Per Hall, Torben Hansen, Tamara B. Harris, Andrew T. Hattersley, Richard B. Hayes, Joachim Heinrich, Frank B. Hu, Kristian Hveem, Thomas Illig, Marjo-Riitta Jarvelin, Jaakko Kaprio, Fredrik Karpe, Kay-Tee Khaw, Lambertus A. Kiemeney, Heiko Krude, Markku Laakso, Debbie A. Lawlor, Andres Metspalu, Patricia B. Munroe, Willem H. Ouwehand, Oluf Pedersen, Brenda W. Penninx, Annette Peters, Peter P. Pramstaller, Thomas Quertermous, Thomas Reinehr, Aila Rissanen, Igor Rudan , Nilesh J. Samani, Peter E.H. Schwarz, Alan R. Shuldiner, Timothy D. Spector, Jaakko Tuomilehto, Manuela Uda, André Uitterlinden, Timo T. Valle, Martin Wabitsch, Gérard Waeber, Nicholas J. Wareham, Hugh Watkins, on behalf of Procardis Consortium, James F. Wilson, Alan F. Wright, M.Carola Zillikens, Nilanjan Chatterjee, Steven A. McCarroll, Shaun Purcell, Eric E. Schadt, Peter M. Visscher, Themistocles L. Assimes, Ingrid B. Borecki, Panos Deloukas, Caroline S. Fox, Leif C. Groop, Talin Haritunians, David J. Hunter, Robert C. Kaplan, Karen L. Mohlke, Jeffrey R. O'Connell, Leena Peltonen, David Schlessinger, David P. Strachan, Cornelia M. van Duijn, H.-Erich Wichmann, Timothy M. Frayling, Unnur Thorsteinsdottir, Gonçalo R. Abecasis, Inês Barroso, Michael Boehnke, Kari Stefansson, Kari E. North, Mark I. McCarthy, Joel N. Hirschhorn, Erik Ingelsson, Ruth J.F. Loos

2.3. Supplementary References

1. Gao,Y.T. *et al.* Association of menstrual and reproductive factors with breast cancer risk: results from the Shanghai Breast Cancer Study. *Int. J. Cancer* **87**, 295-300 (2000).

2. Zheng,W. *et al.* Genome-wide association study identifies a new breast cancer susceptibility locus at 6q25.1. *Nat. Genet.* **41**, 324-328 (2009).

3. Wen,W. *et al.* The modifying effect of C-reactive protein gene polymorphisms on the association between central obesity and endometrial cancer risk. *Cancer* **112**, 2409-2416 (2008).

4. Shu,X.O. *et al.* Soy food intake and breast cancer survival. *JAMA* **302**, 2437-2443 (2009).

5. Zheng,W. *et al.* The Shanghai Women's Health Study: rationale, study design, and baseline characteristics. *Am. J. Epidemiol.* **162**, 1123-1131 (2005).

6. Cho,Y.S. *et al.* A large-scale genome-wide association study of Asian populations uncovers genetic factors influencing eight quantitative traits. *Nat. Genet.* **41**, 527-534 (2009).

7. Hughes,K. *et al.* Cardiovascular diseases in Chinese, Malays, and Indians in Singapore. II. Differences in risk factor levels. *J. Epidemiol. Community Health* **44**, 29-35 (1990).

8. Tan,C.E., Emmanuel,S.C., Tan,B.Y., & Jacob,E. Prevalence of diabetes and ethnic differences in cardiovascular risk factors. The 1992 Singapore National Health Survey. *Diabetes Care* **22**, 241-247 (1999).

9. Hughes,K., Aw,T.C., Kuperan,P., & Choo,M. Central obesity, insulin resistance, syndrome X, lipoprotein(a), and cardiovascular risk in Indians, Malays, and Chinese in Singapore. *J. Epidemiol. Community Health* **51**, 394-399 (1997).

10. Cutter,J., Tan,B.Y., & Chew,S.K. Levels of cardiovascular disease risk factors in Singapore following a national intervention programme. *Bull. World Health Organ* **79**, 908-915 (2001).

11. Ng,D.P. *et al.* Reduced GFR and albuminuria in Chinese type 2 diabetes mellitus patients are both independently associated with activation of the TNF-alpha system. *Diabetologia* **51**, 2318-2324 (2008).

12. Foong,A.W. *et al.* Rationale and methodology for a population-based study of eye diseases in Malay people: The Singapore Malay eye study (SiMES). *Ophthalmic Epidemiol.* **14**, 25-35 (2007).

13. Price,A.L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* **38**, 904-909 (2006).

14. Marchini,J., Howie,B., Myers,S., McVean,G., & Donnelly,P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat. Genet.* **39**, 906-913 (2007).

15. Saw,S.M. *et al.* A cohort study of incident myopia in Singaporean children. *Invest Ophthalmol. Vis. Sci.* **47**, 1839-1844 (2006).

16. Chambers,J.C. *et al.* Common genetic variation near MC4R is associated with waist circumference and insulin resistance. *Nat. Genet.* **40**, 716-718 (2008).

17. Takeuchi,F. *et al.* Blood pressure and hypertension are associated with 7 loci in the Japanese population. *Circulation* **121**, 2302-2309 (2010).

18. Takeuchi,F. *et al.* Confirmation of multiple risk Loci and genetic impacts by a genome-wide association study of type 2 diabetes in the Japanese population. *Diabetes* **58**, 1690-1699 (2009).

19. Purcell,S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559-575 (2007).

20. Tsai,F.J. *et al.* A genome-wide association study identifies susceptibility variants for type 2 diabetes in Han Chinese. *PLoS. Genet.* **6**, e1000847 (2010).

21. Pan,W.H. *et al.* Han Chinese cell and genome bank in Taiwan: purpose, design and ethical considerations. *Hum. Hered.* **61**, 27-30 (2006).

22. Lee,M.T. *et al.* Genome-wide association study of bipolar I disorder in the Han Chinese population. *Mol. Psychiatry*(2010).

23. Nakamura,Y. The BioBank Japan Project. *Clin. Adv. Hematol. Oncol.* **5**, 696-697 (2007).

24. Okada,Y. *et al.* Common variations in PSMD3-CSF3 and PLCB4 are associated with neutrophil count. *Hum. Mol. Genet.* **19**, 2079-2085 (2010).

25. Kamatani,Y. *et al.* Genome-wide association study of hematological and biochemical traits in a Japanese population. *Nat. Genet.* **42**, 210-215 (2010).

26. The International HapMap Project. *Nature* **426**, 789-796 (2003).

27. Li,Y., Willer,C., Sanna,S., & Abecasis,G. Genotype imputation. *Annu. Rev. Genomics Hum. Genet.* **10**, 387-406 (2009).

28. GenSalt: rationale, design, methods and baseline characteristics of study participants. *J. Hum. Hypertens.* **21**, 639-646 (2007).

29. Abecasis,G.R., Cherny,S.S., Cookson,W.O., & Cardon,L.R. GRR: graphical representation of relationship errors. *Bioinformatics.* **17**, 742-743 (2001).

30. O'Connell,J.R. & Weeks,D.E. PedCheck: a program for identification of genotype incompatibilities in linkage analysis. *Am. J. Hum. Genet.* **63**, 259-266 (1998).

31. Mannami,T., Konishi,M., Baba,S., Nishi,N., & Terao,A. Prevalence of asymptomatic carotid atherosclerotic lesions detected by high-resolution ultrasonography and its relation to cardiovascular risk factors in the general population of a Japanese city: the Suita study. *Stroke* **28**, 518-525 (1997).

32. Kolonel,L.N. *et al.* A multiethnic cohort in Hawaii and Los Angeles: baseline characteristics. *Am. J. Epidemiol.* **151**, 346-357 (2000).

33. Jurj,A.L. *et al.* Reproducibility and validity of the Shanghai Men's Health Study physical activity questionnaire. *Am. J. Epidemiol.* **165**, 1124-1133 (2007).

34. Lee,J. *et al.* Should central obesity be an optional or essential component of the metabolic syndrome? Ischemic heart disease risk in the Singapore Cardiovascular Cohort Study. *Diabetes Care* **30**, 343-347 (2007).

35. Tabara,Y. *et al.* Common variants in the ATP2B1 gene are associated with susceptibility to hypertension: the Japanese Millennium Genome Project. *Hypertension* **56**, 973-980 (2010).