SUPPLEMENTARY TABLES AND FIGURES

Exome array analysis identifies novel loci and low-frequency variants for insulin processing and secretion

Jeroen R Huyghe[1], Anne U Jackson[1], Marie P Fogarty[2], Martin L Buchkovich[2], Alena Stančáková[3], Heather M Stringham[1], Xueling Sim[1], Lingyao Yang[1], Christian Fuchsberger[1], Henna Cederberg[3], Peter S Chines[4], Tanya M Teslovich[1], Jane M Romm[5], Hua Ling[5], Ivy McMullen[5], Roxann Ingersoll[5], Elizabeth W Pugh[5], Kimberly F Doheny[5], Benjamin M Neale[6-8], Mark J Daly[9], Johanna Kuusisto[3], Laura J Scott[1], Hyun Min Kang[1], Francis S Collins[4], Gonçalo R Abecasis[1], Richard M Watanabe[10,11], Michael Boehnke[1,12], Markku Laakso[3,12], Karen L Mohlke[2,12]

1. Department of Biostatistics and Center for Statistical Genetics, University of Michigan, Ann Arbor, Michigan, USA.
2. Department of Genetics, University of North Carolina, Chapel Hill, North Carolina, USA.
3. Department of Medicine, University of Eastern Finland and Kuopio University Hospital, Kuopio, Finland.
4. Genome Technology Branch, National Human Genome Research Institute, Bethesda, Maryland, USA.
5. The Center for Inherited Disease Research, Johns Hopkins University, Baltimore, Maryland, USA.
6. The Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA.
7. The Center for Human Genetic Research, Massachusetts General Hospital, Boston, Massachusetts, USA.
8. Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, Massachusetts, USA.
9. Department of Genetics, Harvard University, Cambridge, Massachusetts, USA.
10. Department of Preventive Medicine, Keck School of Medicine of USC, Los Angeles, California, USA.
11. Department of Physiology and Biophysics, Keck School of Medicine of USC, Los Angeles, California, USA.

12. These authors jointly directed this work.

Contents

Supplementary Table 1. Characteristics of non-diabetic METSIM study participants analyzed

Variable	*n*	Mean	SD	Minimum	Maximum
Age (years)	8,229	57.1	7.1	45.0	74.0
BMI (kg/m^2)	8,229	26.8	3.8	16.2	51.6
Fasting glucose (mM)	8,229	5.71	0.48	3.50	6.90
Glucose at 120 min of an OGTT (mM)	8,219	6.06	1.69	1.40	11.00
Fasting proinsulin (pM)	8,227	13.8	6.7	2.3	90.5
Fasting insulin (pM)	8,225	49.7	35.5	6.0	611.4
Fasting proinsulin/insulin ratio	8,224	0.334	0.155	0.030	2.242
Early-phase[a] proinsulin AUC (pM × min)	8,191	669.0	313.8	135.0	3,730.5
Late-phase[b] proinsulin AUC (pM × min)	8,186	3,688.8	1,765.5	612.0	21,064.5
Early-phase[a] proinsulin AUC/insulin AUC ratio	8,175	0.118	0.053	0.015	0.686
Late-phase[b] proinsulin AUC/insulin AUC ratio	8,165	0.139	0.062	0.025	0.618
Insulinogenic index (pmol/mmol)[c]	8,161	131.7	218.3	-4,665.0	7,626.0
Disposition index[c]	8,170	163.4	71.9	34.0	1,304.6

mM, millimolar; OGTT, oral glucose tolerance test; pM, picomolar; AUC, area under the curve

[a] Early-phase: 0 to 30 minute of an OGTT

[b] Late-phase: 30 to 120 minute of an OGTT

[c] See Online Methods for definitions and references

Supplementary Table 2. Numbers of exome chip variants passing QC that are variable in 8,229 non-diabetic METSIM study participants

		MAF				Variants included in		
	Total	< 0.5%	0.5% to 5%	> 5%	Present in NHGRI GWAS catalog[a]	single-variant tests (MAF > 0.05%)	gene-based tests (MAF < 1%)	gene-based tests (MAF < 3%)
Nonsynonymous[b]	69,246	43,944 (B=20,521; P=7,110; D=14,706) (N=29,611; C=13,624)	13,311 (B=6,969; P=1,992; D=3,662) (N=9,726; C=3,305)	11,991 (B=8,526; P=1,246; D=1,591) (N=10,408; C=1,408)	187	40,801	48,295	54,764
Stop-introducing	1,152	926	151	75	3	511	989	1,054
Stop losses	67	43	7	17	0	43	46	49
Splice-site-disrupting	612	424	151	75	7	344	452	493
Synonymous	2,972	1,545	563	864	47	1,939	-	-
Other	15,815	840	1,182	13,793	4,037	15,391	-	-
Total	89,864	47,722	15,295	26,847	4,281	59,029	52,535	59,901

Number of monomorphic variants is 152,207 (62.9% of 242,071 variants passing QC and with Hardy-Weinberg equilibrium P value > 10^{-6}). Variants were annotated relative to GENCODE version 7 transcripts[1].

[a] http://www.genome.gov/gwastudies (accessed 17 June 2012)

[b] For nonsynonymous variants, for each frequency category, the number of variants predicted to be benign (B), possibly damaging (P) and probably damaging (D) based on PolyPhen-2 scores[2] (where available) are given. Also given are the number of variants predicted to be at non-conserved sites (N) and conserved sites (C) based on the PhyloP score[3] (where available). Variants with a PhyloP score > 3 were considered to be conserved.

Supplementary Table 3. Evidence for association with fasting proinsulin in 8,224 METSIM study participants at loci previously identified by genome-wide association studies

SNP	Nearby gene(s)	Chr	Position (bp)	Minor/ major allele	MAF	$\hat{\beta}$ ± SE	Effect size in SD units ± SE	Proportion of trait variance explained	*P* value
rs11603334	*ARAP1*	11	72,432,985	A/G	.253	.081 ± .006	.26 ± .02	.0259	6.1×10^{-45}
rs7944584[a]	*MADD*	11	47,336,320	T/A	.181	-.065 ± .006	-.21 ± .02	.0131	4.9×10^{-25}
rs7172432[b]	*VPS13C/C2CD4A/B*	15	62,396,389	A/G	.446	.042 ± .005	.14 ± .02	.0091	7.4×10^{-18}
rs11558471	*SLC30A8*	8	118,185,733	G/A	.398	-.036 ± .005	-.12 ± .02	.0064	1.3×10^{-12}
rs1051006[c]	*MADD*	11	47,306,585	A/G	.310	-.031 ± .006	-.10 ± .02	.0043	2.7×10^{-8}
rs6235	*PCSK1*	5	95,728,898	G/C	.282	.031 ± .005	.10 ± .02	.0040	2.3×10^{-8}
rs7903146	*TCF7L2*	10	11,4758,349	T/C	.177	.027 ± .006	.09 ± .02	.0022	3.0×10^{-5}
rs1549318	*LARP6*	15	71,109,147	T/C	.495	.015 ± .005	.05 ± .02	.0011	.0031
rs4790333	*SGSM2*	17	2,262,703	C/T	.472	-.013 ± .005	-.04 ± .02	.0009	.0076
rs9727115[d]	*SNX7*	1	99,177,253	A/G	.311	-.007 ± .005	-.02 ± .02	.0002	.18

Chr, chromosome; MAF, minor allele frequency. Positions are from NCBI Build 37 with allele labels from the forward strand. Fasting proinsulin levels were log-transformed and adjusted for fasting insulin, BMI, age, and age^2. Effects are reported for the minor allele. $\hat{\beta}$ coefficient units are ln(pmol/l). Results are given for the SNPs reported in Strawbridge et al.[4] or proxies thereof. For comparison, effect size in SD units (± SE) and proportion of trait variance explained for the low-frequency variants shown in Table 2 are as follows: rs150781447 (TBC1D30), 0.38 (± 0.06) and 0.0056, respectively; rs3824420 (KANK1), 0.26 (± 0.05) and 0.0040, respectively.

[a] Proxy of discovery SNP rs10501320 (D' = .996; r^2 = .959)

[b] Proxy of discovery SNP rs4502156 (D' = .986; r^2 = .926)

[c] Proxy of discovery SNP rs10838687 (D' = 1.000; r^2 = .798). In Strawbridge et al., this variant only reached significance after adjusting for the lead SNP rs7944584. Reported effect sizes and *P* value are for the adjusted analysis.

[d] In Strawbridge et al., rs9727115 only reached significance after adjusting for fasting glucose. The *P* value after adjusting for fasting glucose is .58.

Supplementary Table 4. Results of conditional analyses for fasting proinsulin for SNPs with association *P* value < 1×10^{-4} after conditioning on the lead SNP(s) from GWAS-identified signals

Locus	Chr	Position (bp)	SNP	SNP type	Minor/ major allele	MAF	Unadjusted *P* value	Conditional analysis			*D'*	r^2	Allele couple[a]	Association pattern[b]
								Conditioning SNP(s)	$\hat{\beta}$ (SE)	*P* value				
SGSM2	17	2,282,779	**rs61741902**	nonsyn	A/G	.014	8.9×10^{-10}	rs4790333	0.128 (0.021)	4.8×10^{-10}	.301	.001	G,T	
		2,262,703	rs4790333	intronic	C/T	.472	.0076	**rs61741902**	-0.014 (0.005)	.0039				
MADD	11	47,306,630	**rs35233100**	nonsyn	T/C	.037	7.6×10^{-15}	rs7944584	-0.054 (0.014)	.0001	1.000	.174	C,A	↓ pair
		47,336,320	rs7944584	intronic	T/A	.181	4.9×10^{-25}	**rs35233100**	-0.054 (0.007)	5.7×10^{-15}				
		47,306,630	**rs35233100**	nonsense	T/C	.037	7.6×10^{-15}	rs1051006	-0.105 (0.013)	5.0×10^{-16}	1.000	.017	T,G	↑ pair
		47,306,585	rs1051006	nonsyn	A/G	.310	.033	**rs35233100**	-0.017 (0.005)	.0016				
		47,306,585	rs1051006	nonsyn	A/G	.310	.033	rs7944584	-0.031 (0.006)	2.7×10^{-8}	1.000	.099	G,T	↑ pair
		47,336,320	rs7944584	intronic	T/A	.181	4.9×10^{-25}	rs1051006	-0.076 (0.007)	8.3×10^{-31}				
		47,306,630	**rs35233100**	nonsense	T/C	.037	7.6×10^{-15}	rs7944584 rs1051006	-0.054 (0.014)	.0001				
		47,306,585	rs1051006	nonsyn	A/G	.310	.033	rs7944584 **rs35233100**	-0.031 (0.006)	2.7×10^{-8}				
		47,336,320	rs7944584	intronic	T/A	.181	4.9×10^{-25}	rs1051006 **rs35233100**	-0.065 (0.007)	1.3×10^{-19}				

Chr, chromosome; MAF, minor allele frequency; SE, standard error; nonsyn, nonsynonymous.
Positions are from NCBI Build 37 with allele labels from the forward strand. Fasting proinsulin was log-transformed and adjusted for fasting insulin, BMI, age, and age^2. For conditional analysis, common SNP(s) representing GWAS signals are those reported by Strawbridge et al. (2010) and low-frequency nonsynonymous SNPs are those identified in this study. **Nonsynonymous SNPs with MAF < 5% are bolded.** Effects are for the conditional analysis and are reported for the minor allele. $\hat{\beta}$ coefficient units are ln(pmol/l). Sample sizes for these analyses ranged from 8,221 to 8,224. The LD metrics between SNP pairs *D'* and r^2 were estimated in 9,633 METSIM individuals passing genotype quality control.
[a] Allele couple: positively associated alleles.
[b] Up arrow indicates a SNP pair for which significance increases after conditioning on the other SNP, down arrow indicates the opposite.

Supplementary Table 5. Protein annotations of low frequency SNPs associated with insulin secretion traits

SNP (chr:pos)	Gene	Amino acid change	Location	Transcript	UniProt ID	PolyPhen-2 score[a]	SIFT score[b]	PhastCons score[c] (mammal/ vertebrate)	Additional annotations
rs61741902 (chr17:2282779)	*SGSM2*	V996I	Exon 23	ENST00000426855.1	**O43147-1**	.605 (P)	-	1/1	Located 30 amino acids from the Rab-GTPase Activating Protein TBC (Tre-2/Bub2/Cdc16) domain
		V1041I	Exon 24	ENST00000268989.3	**O43147-2**	.338 (B)	.06 (T)		
rs150781447 (chr12:65224220)	*TBC1D30*	R279C	Exon 5	ENST00000229088.6	**Q9Y2I9-1**	1.0(D)	-	.94/.94	In Rab-GTPase Activating Protein TBC (Tre-2/Bub2/Cdc16) domain
		R116C	Exon 4	ENST00000539867.1	**Q9Y2I9-2**	1.0(D)	-		
		R116C	Exon 4	ENST00000455166.1	**Q9Y2I9-3**	1.0(D)	-		
		R127C	Exon 5	ENST00000544190.1	F5GY74	-	-		
		R2C	Exon 5	ENST00000542120.1	F8VZ81	-	-		First coding exon; may interfere with Kozak sequence
		R2C	Exon 5	ENST00000539120.1	F5H7L7	-	-		
		-	Intron 3	ENST00000544457.1	F5H0E8	-	-		
		-	Intron 3	ENST00000411580.2	E7ES83	-	-		
rs3824420 (chr9:712766)	*KANK1*	R667H	Exon 3	ENST00000354485.5	Q5W0W1	0 (B)	.31(T)	.25/.04	
		R667H	Exon 7	ENST00000382303.1	**Q14678-1**	.001 (B)	-		
		R667H	Exon 3	ENST00000382297.2	**Q14678-1**	.001 (B)	-		
		R509H	Exon 2	ENST00000382293.3	**Q14678-2**	-	-		
		-	5'UTR	ENST00000397976.1	F5H7I5	-	-		
		-	5'UTR	ENST00000489369.1	-	-			
rs35658696 (chr5:102338811)	*PAM*	D563G	Exon 16	ENST00000438793.3	P19021-1	.986(D)	-	1/1	In peptidyl-alpha-hydroxyglycine alpha-amidating lyase domain
		D563G	Exon 16	ENST00000346918.2	P19021-4	.984 (D)	-		
		D563G	Exon 16	ENST00000304400.7	P19021-5	.992 (D)	-		
		D563G	Exon 16	ENST00000455264.2	P19021-3	.958 (D)	-		
		D466G	Exon 15	ENST00000274392.9	F8WE90	.984 (D)	.01(D)		
		D456G	Exon 15	ENST00000348126.2	P19021-2	.992 (D)	-		
		D335G	Exon 8	ENST00000379799.3	H7BYD9	.956 (D)	-		
		-	5'UTR	ENST00000511429.1	-	-	-		
		-	5'UTR	ENST00000379787.4	A6NMH0	-	-		
		-	3'UTR	ENST00000345721.2	F8W8D9	-	-		
rs36046591 (chr5:102537285)	*PPIP5K2*	S1207G	Exon 30	ENST00000321521.8	O43314-2	0 (B)	.43(T)	.9/.46	
		S1228G	Exon 31	ENST00000358359.3	O43314-1	0 (B)	-		
		S1263G	Exon 33	ENST00000451606.1	E9PGM8	0 (B)	-		
		S1207G	Exon 29	ENST00000414217.1	O43314-2	0 (B)	-		
		S383G	Exon 10	ENST00000509597.1	H0Y9S9	0 (B)	-		
		-	5'UTR	ENST00000504083.1	-	-	-		
		-	3'UTR	ENST00000513500.1	D6RFZ8	-	-		

Supplementary Table 5. (continued)

SNP	Gene	Amino acid change	Location	Transcript	UniProt ID	PolyPhen-2 score[a]	SIFT score[b]	PhastCons score[c] (mammal/ vertebrate)	Additional annotations
rs35233100 (chr11:47306630)	*MADD*	R766X	Exon 13	ENST00000342922.4	Q8WXG6-3	-	-	.92/.05	Internal exon; 117 bp upstream of exon-intron junction and may cause nonsense-mediated decay
		R766X	Exon 13	ENST00000395336.3	Q8WXG6-7	-	-		
		R766X	Exon 13	ENST00000311027.5	Q8WXG6-1	-	-		
		R766X	Exon 13	ENST00000349238.3	Q8WXG6-2	-	-		
		R766X	Exon 13	ENST00000402192.2	Q8WXG6-3	-	-		
		-	Intron 13	ENST00000407859.3	Q8WXG6-4	-	-		
		-	Intron 13	ENST00000395342.2	F8W9P9	-	-		
		-	Intron 13	ENST00000395344.3	B5MEE5	-	-		
		-	Intron 13	ENST00000406482.1	Q8WXG6-6	-	-		
		-	Intron 13	ENST00000402799.1	Q8WXG6-5	-	-		
		-	5'UTR	ENST00000524530.1	-	-	-		

Each SNP is annotated with all isoforms based on the comprehensive set of GENCODE version 7 gene transcripts[1].

[a] Polyphen-2 scores[2] (where available) range from 0 (B) to 1 (D). B=benign, P=possible damaging, D=probably damaging.

[b] SIFT scores[5] (where available) range from 0 (D) to 1 (T). All scores ≤ .05 are predicted to be damaging. T=tolerated, D=damaging.

[c] Average PhastCons score[6] for 3 codon nucleotides derived from alignments of 33 Mammals/46 Vertebrates. Scores range from 0 (not conserved) to 1 (very conserved). Data from UCSC genome browser.

Supplementary Table 6. Association results for SNPs in the *MADD* gene region showing suggestive or significant association for fasting proinsulin before and after conditioning on the three associated variants at *MADD* (rs7944584, rs1051006, and rs35233100)

								rs7944584[b]		rs1051006[c]		rs35233100[d]	
Locus	Chr	Position (bp)	SNP	Variant	MAF	*P* value	Conditional *P* value[a]	*D*'	r^2	*D*'	r^2	*D*'	r^2
OR5M11	11	56,310,222	rs628524	S171N	.258	3.7×10^{-6}	.11	.029	.001	.406	.128	.620	.042
ATG13	11	46,690,413	rs35619591	G434R	.030	6.1×10^{-10}	.0005	.970	.131	.970	.013	.599	.000
AGBL2	11	47,712,213	rs7941404	R349H	.118	4.7×10^{-21}	.0018	.694	.293	.348	.007	.813	.189
OR4S1	11	48,333,360	rs1483121	intergenic	.112	3.8×10^{-14}	.084	.619	.220	.402	.009	.798	.193

Chr, chromosome; MAF, minor allele frequency. Positions are from NCBI Build 37 with allele labels from the forward strand. Fasting proinsulin was log-transformed and adjusted for fasting insulin, BMI, age, and age^2.

[a] *P* values after conditioning on the three variants rs7944584, rs1051006, and rs35233100 at *MADD*; [b] Lead SNP at *MADD*; [c] Lead SNP of secondary GWAS signal at *MADD*; [d] Nonsense variant at *MADD*. LD between *MADD* variants: rs7944584 - rs1051006 (*D*' = 1.000; r^2 = .099), rs7944584 - rs35233100 (*D*' = 1.000; r^2 = .174), and rs1051006 - rs35233100 (*D*' = 1.000; r^2 = .017). The LD metrics between SNP pairs *D'* and r^2 were estimated in 9,633 METSIM individuals passing genotype quality control.

Supplementary Table 7. Minor allele frequencies of associated SNPs in other data sets

Gene	SNP	SNP type	METSIM[a]	1000G-EUR[b]	1000G-ASN[b]	1000G-AFR[b]	ESP-EA[c]	ESP-AA[d]
SGSM2	rs61741902	nonsynonymous	.014	.012	0	0	.014	.003
MADD	rs35233100	nonsense	.037	.046	0	.004	.063	.012
TBC1D30	rs150781447	nonsynonymous	.020	.015	0	0	0	0
KANK1	rs3824420	nonsynonymous	.029	.005	.163	.043	.001	.033
PAM	rs35658696	nonsynonymous	.053	.044	0	.002	.049	.007
PPIP5K2	rs36046591	nonsynonymous	.053	.041	0	.002	.048	.008

[a] Based on 9,633 METSIM individuals who passed genotype quality control; [b] Based on 1000 Genomes Project phase 1 genotype data from ~379 European ancestry individuals (EUR), ~286 Asian ancestry individuals (ASN), and ~246 African ancestry individuals (AFR); [c] Based on ~4300 unrelated European-American individuals from the NHLBI Exome Sequencing Project (ESP); [d] Based on ~2203 unrelated African-Americans individuals from the NHLBI Exome Sequencing Project (ESP). Frequencies for ESP obtained from the Exome Variant Server (http://evs.gs.washington.edu/EVS/).

SUPPLEMENTARY FIGURE LEGENDS

Supplementary Figure 1 Associations across the 19 insulin processing, insulin secretion, and glycemic traits for low-frequency variants in *SGSM2*, *MADD*, *TBC1D30*, *KANK1*, and *PAM*. Bars represent $-\log_{10} P$ value signed according to direction of effect. The effect allele is given and is the minor allele. For quantitative traits, results are based on 8,104 to 8,229 subjects and rows are ordered according to a complete linkage hierarchical clustering of traits based on the Pearson correlation matrix. Type 2 diabetes association results are based on 1,376 diabetic cases and 5,478 normal glycemic controls.

Supplementary Figure 2 Distribution of primary traits for each genotype for the low-frequency variants in *SGSM2*, *MADD*, *TBC1D30*, *KANK1*, and *PAM*. Traits were log-transformed and adjusted for BMI, age, and age^2. Fasting proinsulin was also adjusted for fasting insulin.

Supplementary Figure 3 Manhattan plots for the insulinogenic index (**a**) and disposition index (**b**) association analyses. Association results of the single-variant analysis ($-\log_{10} P$ values) are plotted against genomic position (NCBI Build 37). Loci previously associated with glycemic traits at genome-wide significance are denoted in blue and loci identified by the current study in red. Previously associated loci are based on Rung et al.[7], Saxena et al.[8], Dupuis et al.[9], and Scott et al.[10]. Insulinogenic index and disposition index were log-transformed and adjusted for fasting insulin, BMI, age, and age^2.

Supplementary Figure 4 Quantile-quantile plots (QQ) for gene-based association analysis results for fasting proinsulin. Tests were performed using SKAT-O[11] on trait residuals obtained by adjusting log-transformed fasting proinsulin for fasting insulin, age, age^2, and BMI, and for relatedness using GenABEL[12]. Results shown for variants with MAF < 3% (magenta) and MAF < 1% (gray).

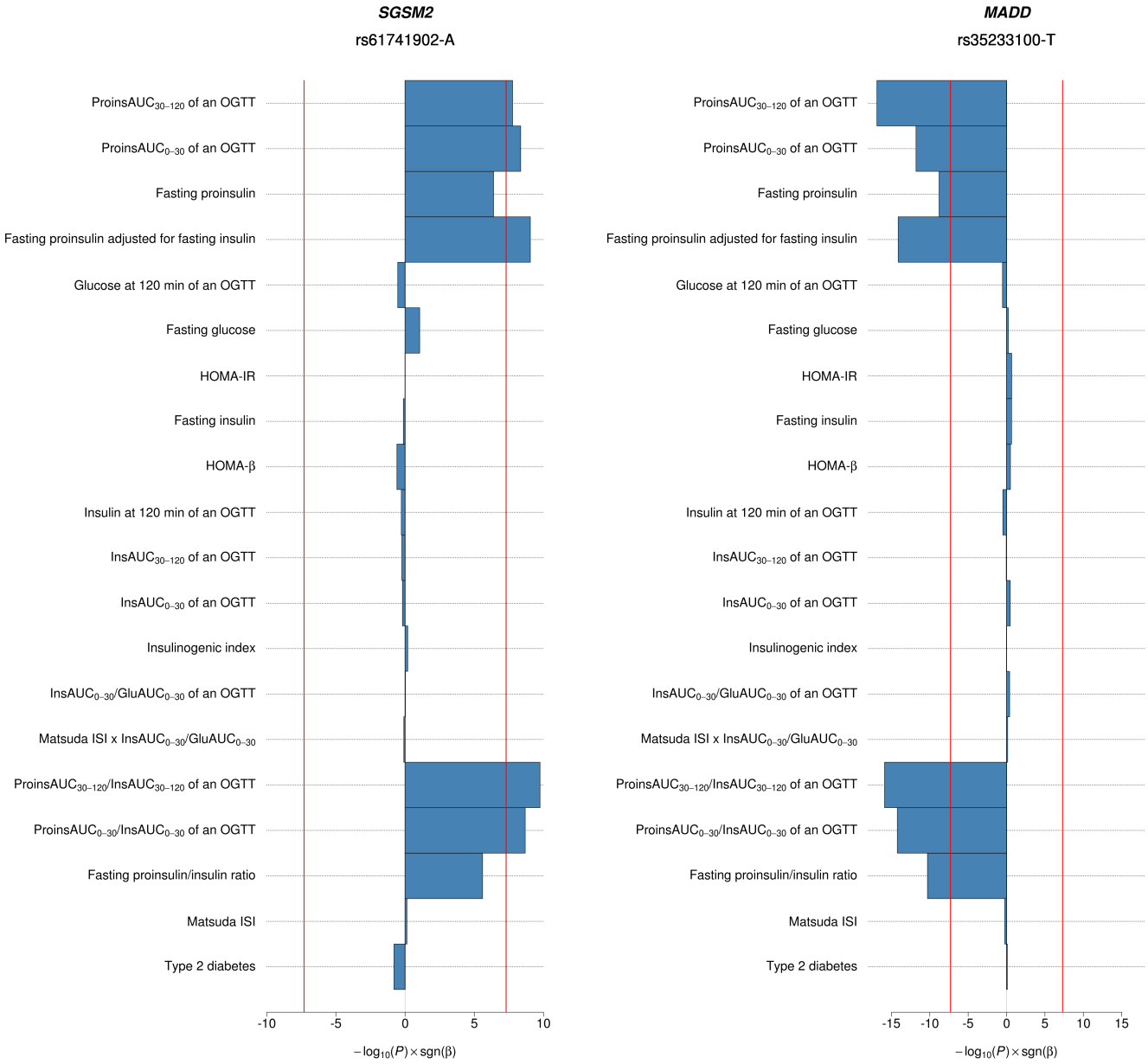
Supplementary Figure 5 Roles of described genes in G-protein signaling. SGSM2, MADD, TBC1D30 and KANK1 have been shown to regulate or function in G-protein signaling. GTP-binding proteins (G-proteins) are characterized by their ability to bind to and hydrolyze GTP and include members of the Rab, Rac, Rho, Rap, and other families. G-proteins are active when bound to GTP but inactive when bound to GDP. Guanine exchange factors (GEF's) catalyze the dissociation of GDP and the binding of GTP, thus promoting the active G-protein state. When bound to GTP, G-proteins remain active briefly and can activate or inactivate many other proteins. GTPase activator proteins (GAP's) turn off active G-proteins by promoting GTP hydrolysis and a return to the inactive state.

Supplementary Figure 6 Pearson correlations between quantitative traits. Rows and columns are ordered according to a complete linkage hierarchical clustering of traits based on the Pearson correlation matrix.

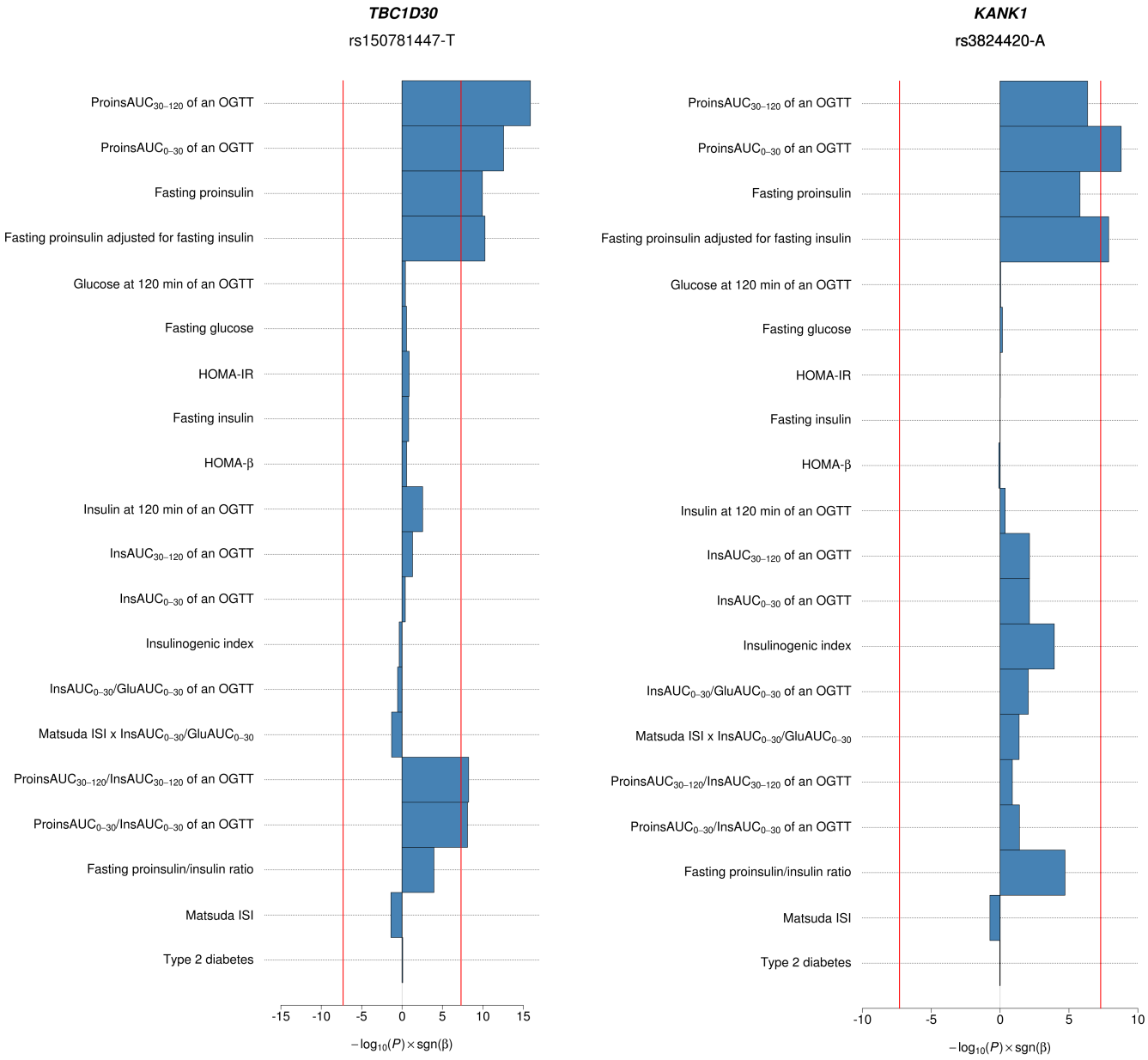
Supplementary Figure 7 Quantile-quantile (QQ) plots for single-variant analysis results of primary traits. Red dots correspond to all *P* values, blue dots to association results after removal of previously identified and novel genome-wide significant loci. Reported genomic control inflation factors were calculated after removal of previously identified and novel genome-wide significant loci.

Supplementary Figure 8 Comparison of gene-based association analysis results ($-\log_{10} P$ values) obtained by conducting tests using SKAT-O[11] on trait residuals unadjusted for relatedness versus trait residuals adjusted for relatedness using GenABEL[12].

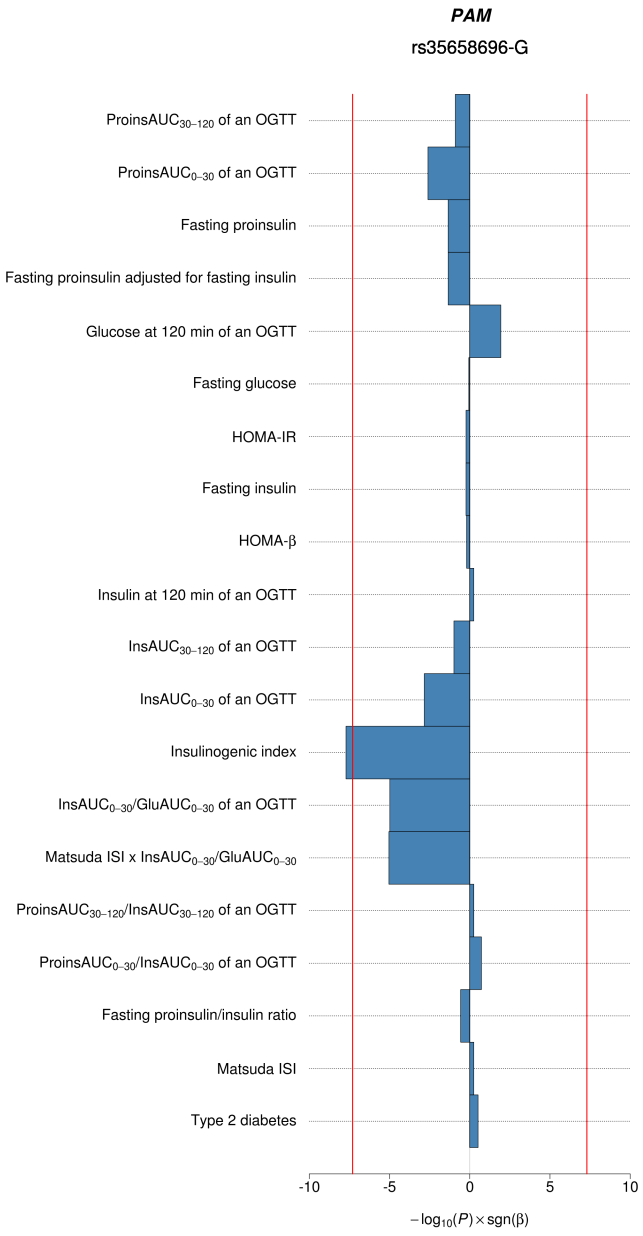
Supplementary Figure 1



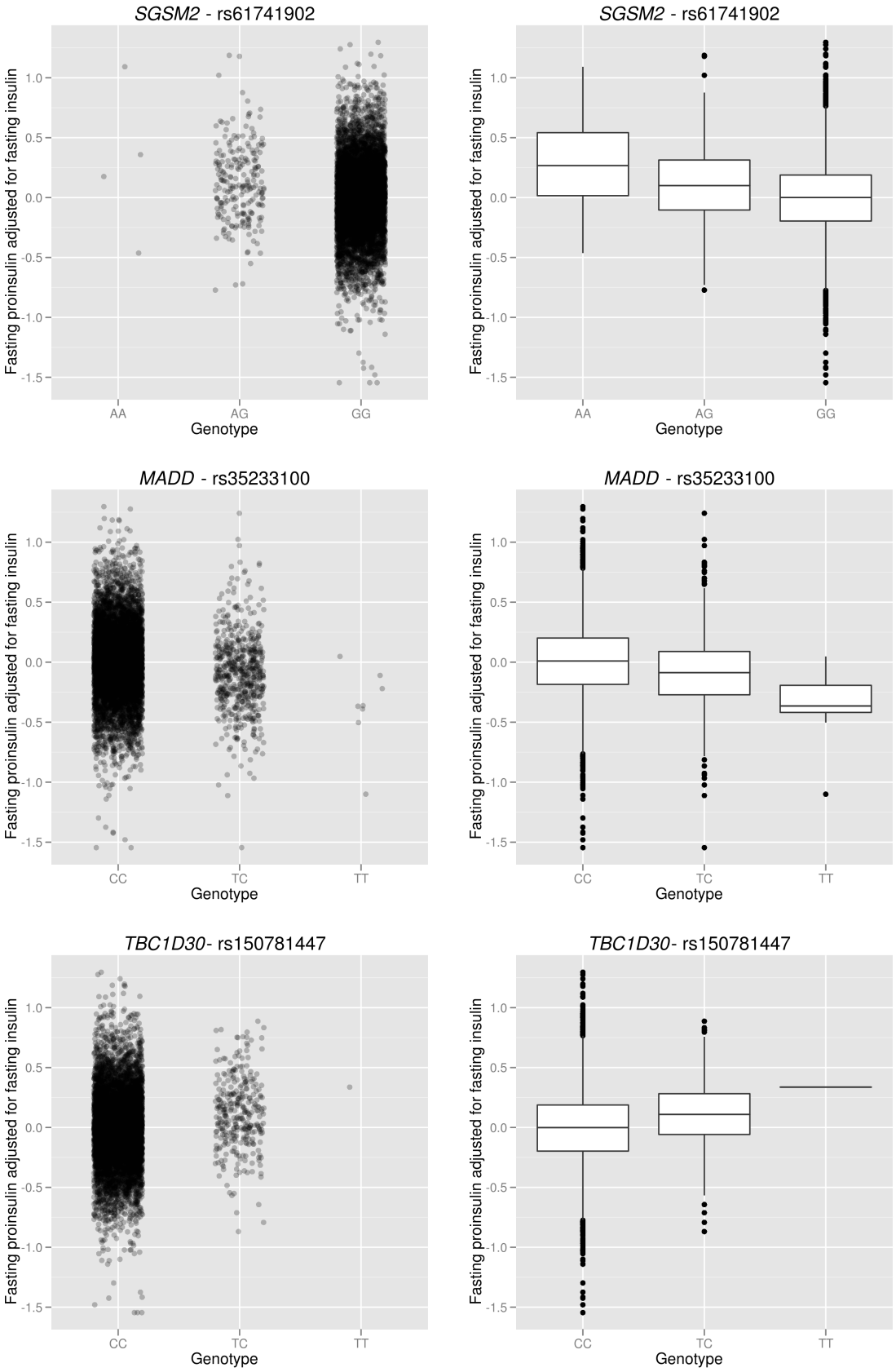
Supplementary Figure 1 cont'd



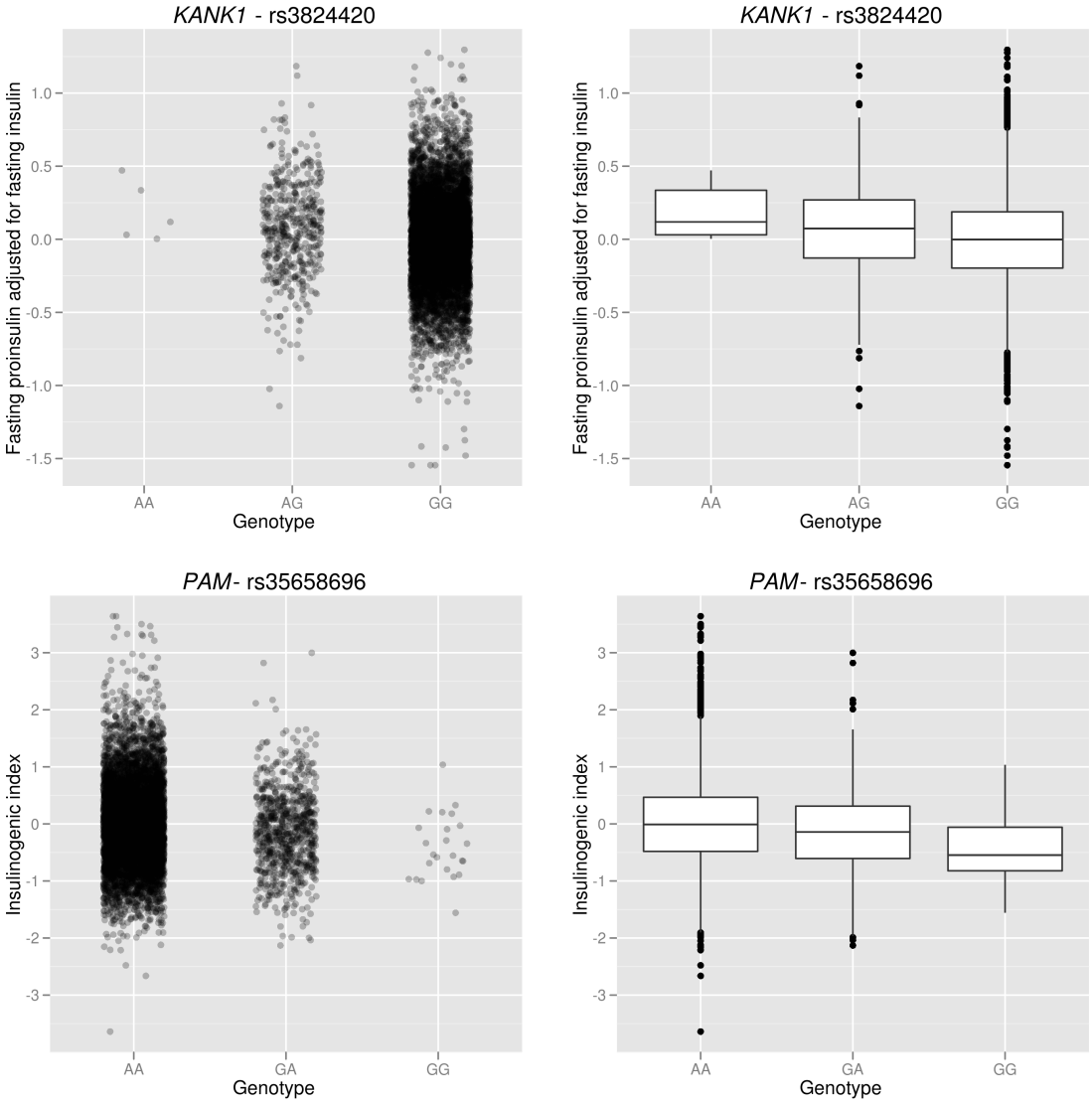
Supplementary Figure 1 cont'd



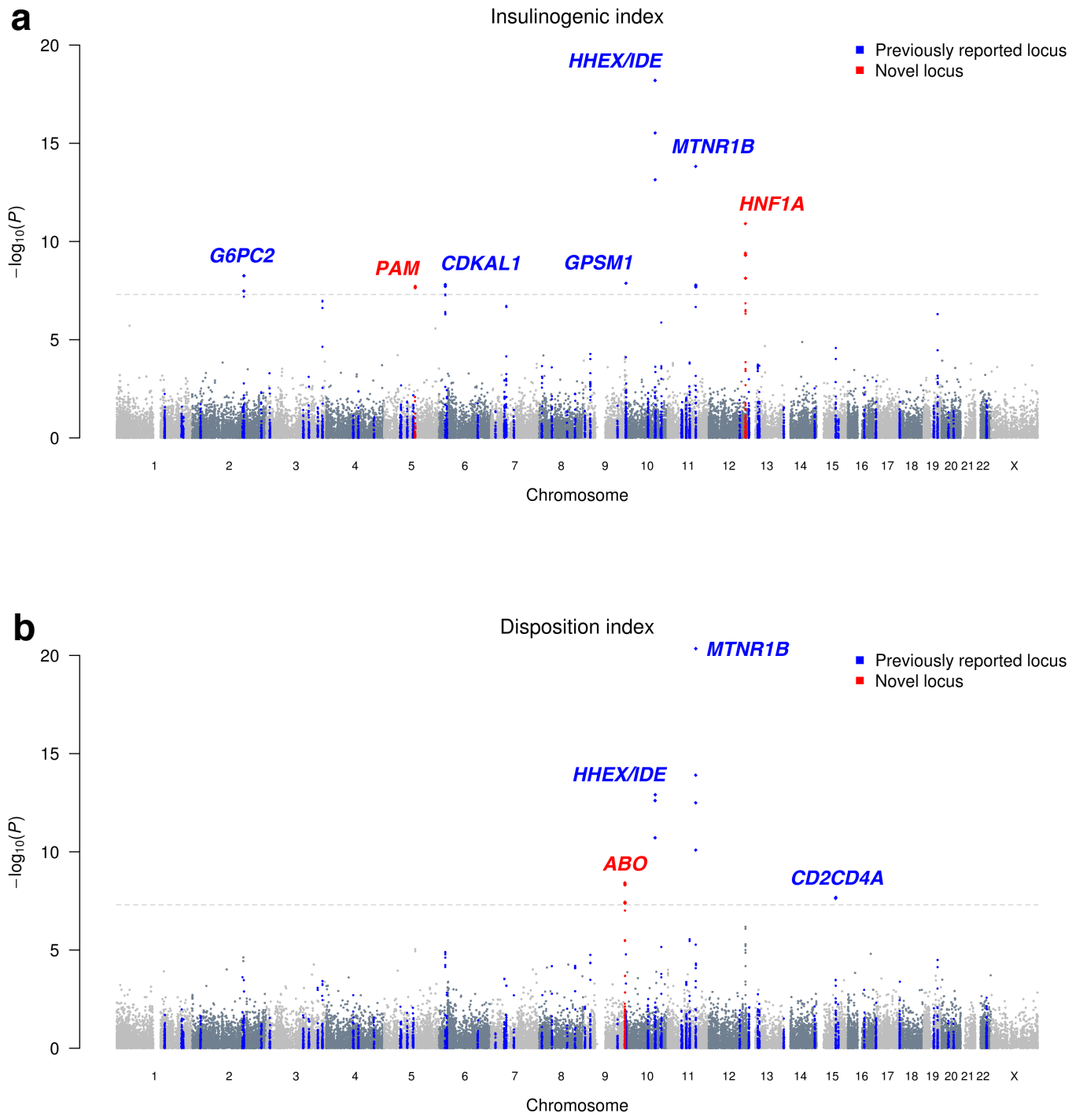
Supplementary Figure 2



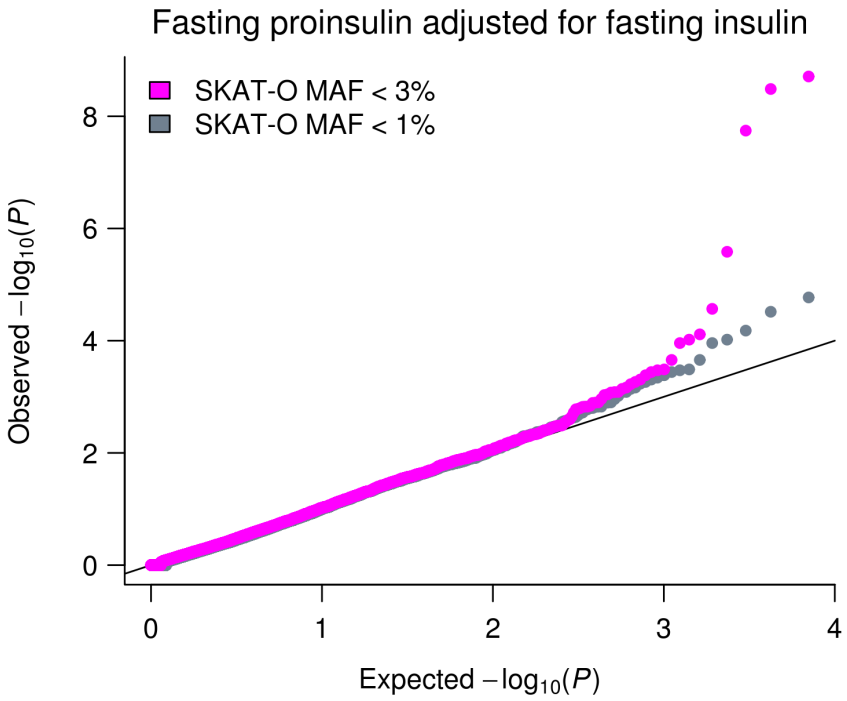
Supplementary Figure 2 cont'd



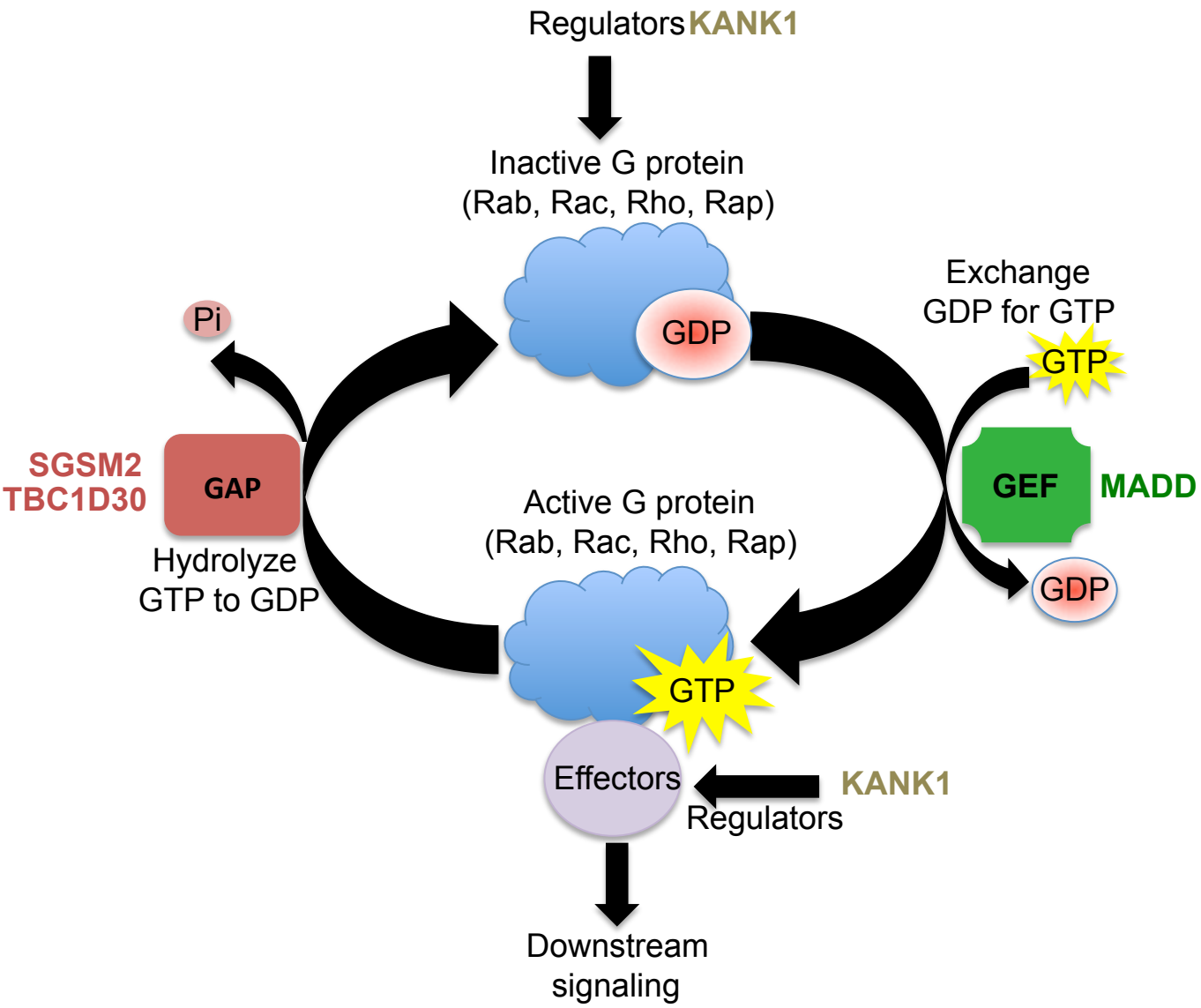
Supplementary Figure 3



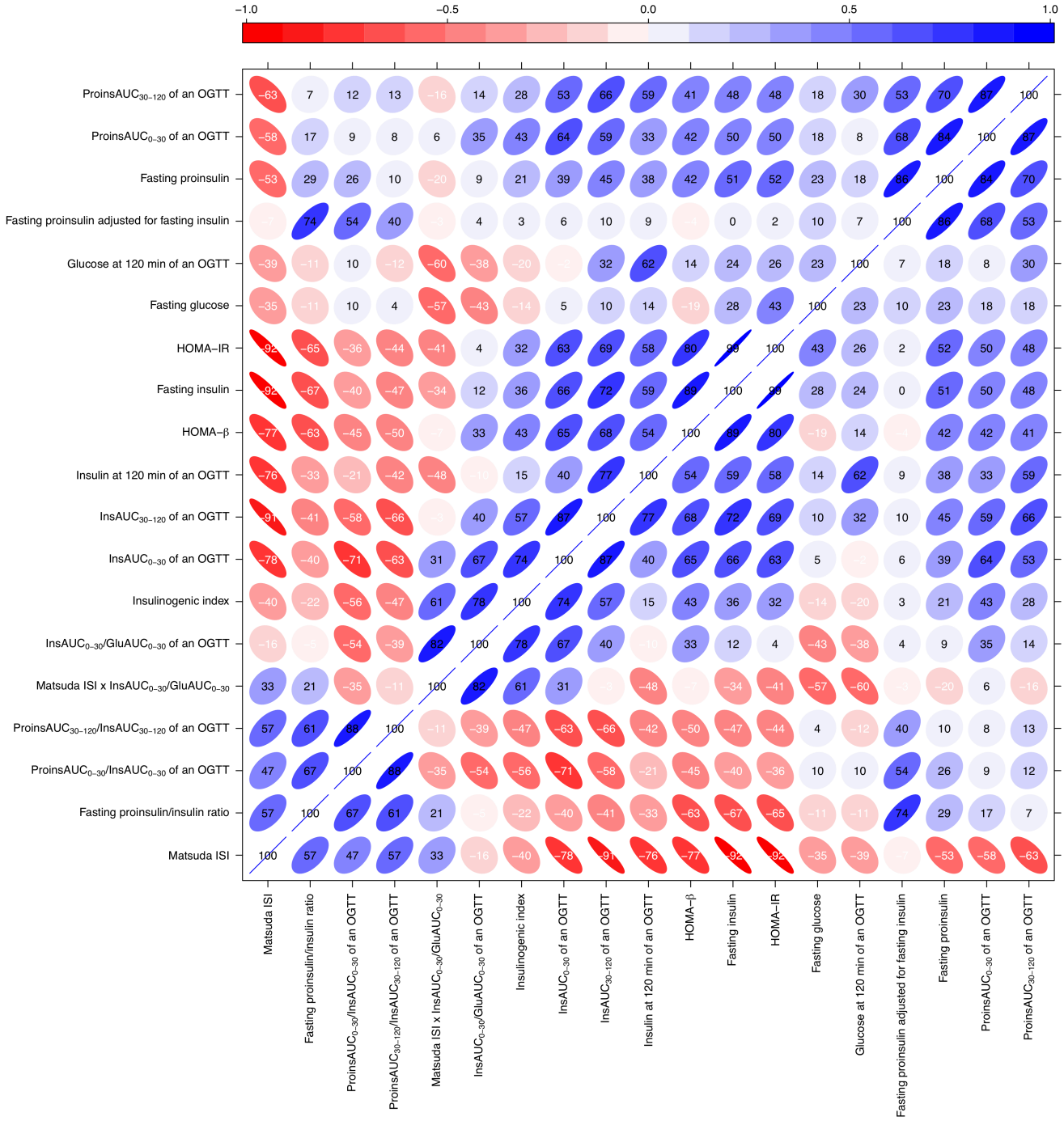
Supplementary Figure 4



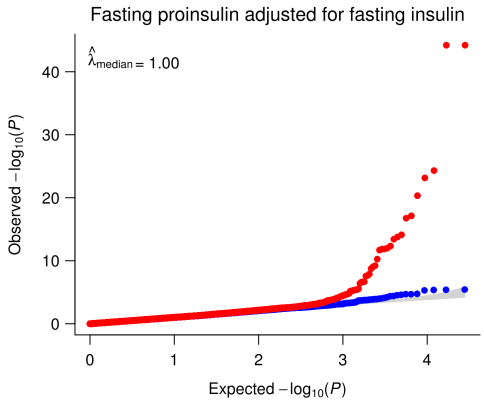
Supplementary Figure 5

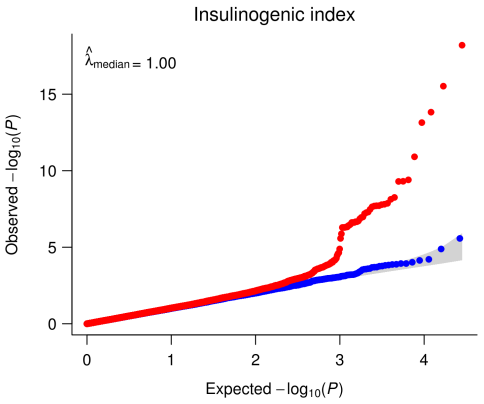


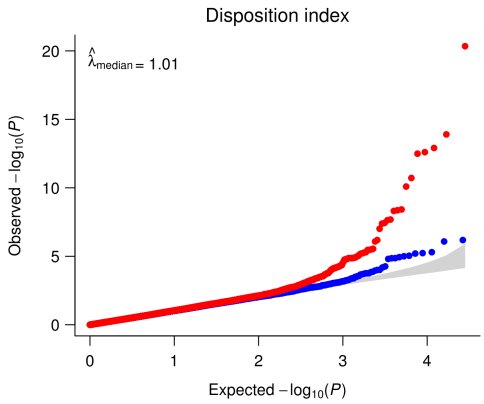
Supplementary Figure 6



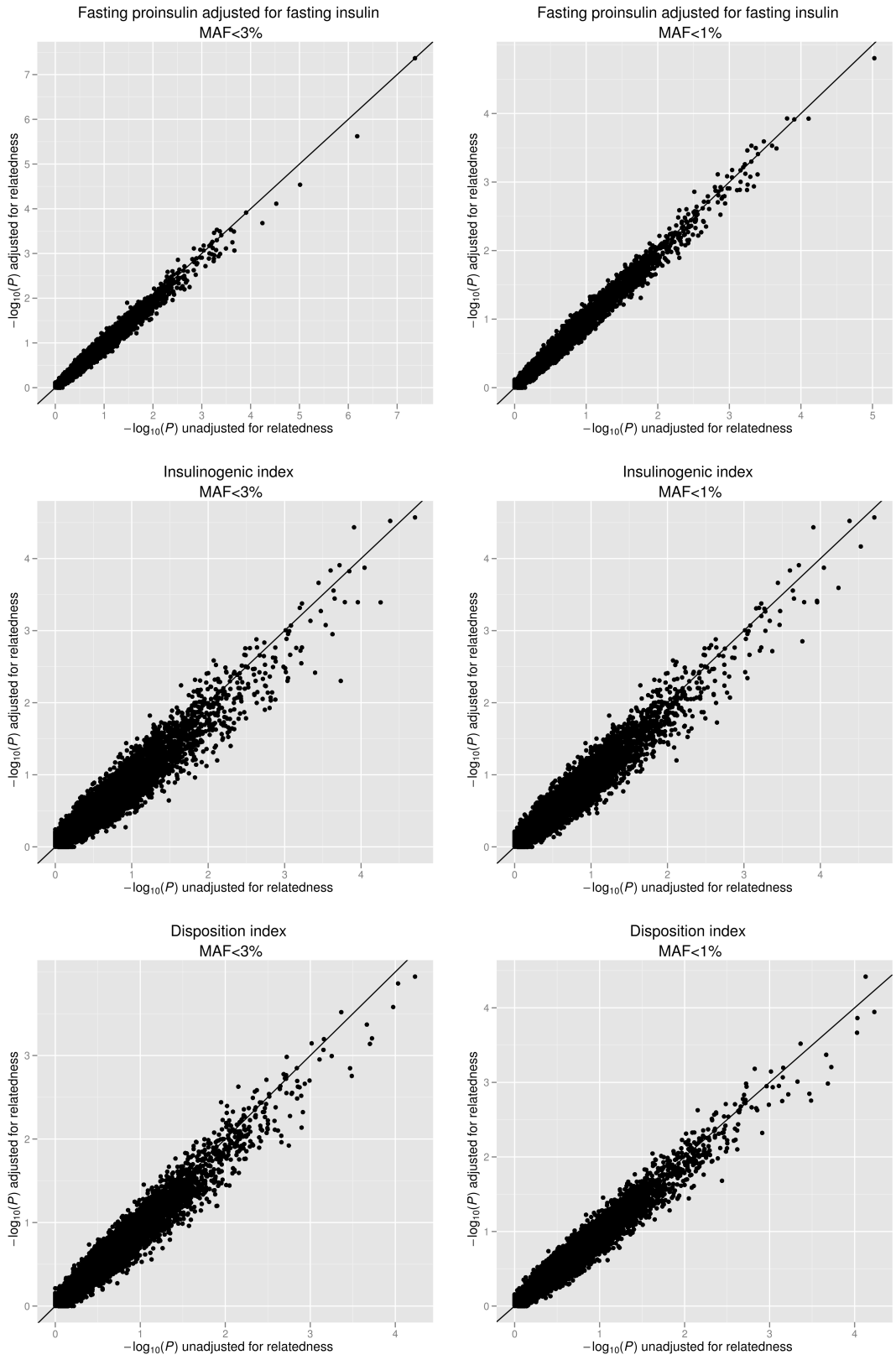
Supplementary Figure 7







Supplementary Figure 8



SUPPLEMENTARY REFERENCES

1. Harrow, J. *et al.* GENCODE: producing a reference annotation for ENCODE. *Genome Biol* **7 Suppl 1**, S4 1-9 (2006).
2. Adzhubei, I.A. *et al.* A method and server for predicting damaging missense mutations. *Nat Methods* **7**, 248-9 (2010).
3. Cooper, G.M. *et al.* Distribution and intensity of constraint in mammalian genomic sequence. *Genome Res* **15**, 901-13 (2005).
4. Strawbridge, R.J. *et al.* Genome-wide association identifies nine common variants associated with fasting proinsulin levels and provides new insights into the pathophysiology of type 2 diabetes. *Diabetes* **60**, 2624-34 (2011).
5. Kumar, P., Henikoff, S. & Ng, P.C. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc* **4**, 1073-81 (2009).
6. Siepel, A. & Haussler, D. Phylogenetic hidden Markov models. (2005). In R. Nielsen, ed., *Statistical Methods in Molecular Evolution*, pp. 325-351. Springer, New York.
7. Rung, J. *et al.* Genetic variant near IRS1 is associated with type 2 diabetes, insulin resistance and hyperinsulinemia. *Nat Genet* **41**, 1110-5 (2009).
8. Saxena, R. *et al.* Genetic variation in GIPR influences the glucose and insulin responses to an oral glucose challenge. *Nat Genet* **42**, 142-8 (2010).
9. Dupuis, J. *et al.* New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. *Nat Genet* **42**, 105-16 (2010).
10. Scott, R.A. *et al.* Large-scale association study using the Metabochip array reveals new loci influencing glycemic traits and provides insight into the underlying biological pathways. *Nat Genet* **44**, 991-1005 (2012).
11. Lee, S., Wu, M.C. & Lin, X. Optimal tests for rare variant effects in sequencing association studies. *Biostatistics* (2012).
12. Aulchenko, Y.S., Ripke, S., Isaacs, A. & van Duijn, C.M. GenABEL: an R library for genome-wide association analysis. *Bioinformatics* **23**, 1294-6 (2007).