The American Journal of Human Genetics, Volume *108*

Supplemental information

Bayesian model comparison for rare-variant association studies

Guhan Ram Venkataraman, Christopher DeBoever, Yosuke Tanigawa, Matthew Aguirre, Alexander G. Ioannidis, Hakhamanesh Mostafavi, Chris C.A. Spencer, Timothy Poterba, Carlos D. Bustamante, Mark J. Daly, Matti Pirinen, and Manuel A. Rivas

Supplemental Figures

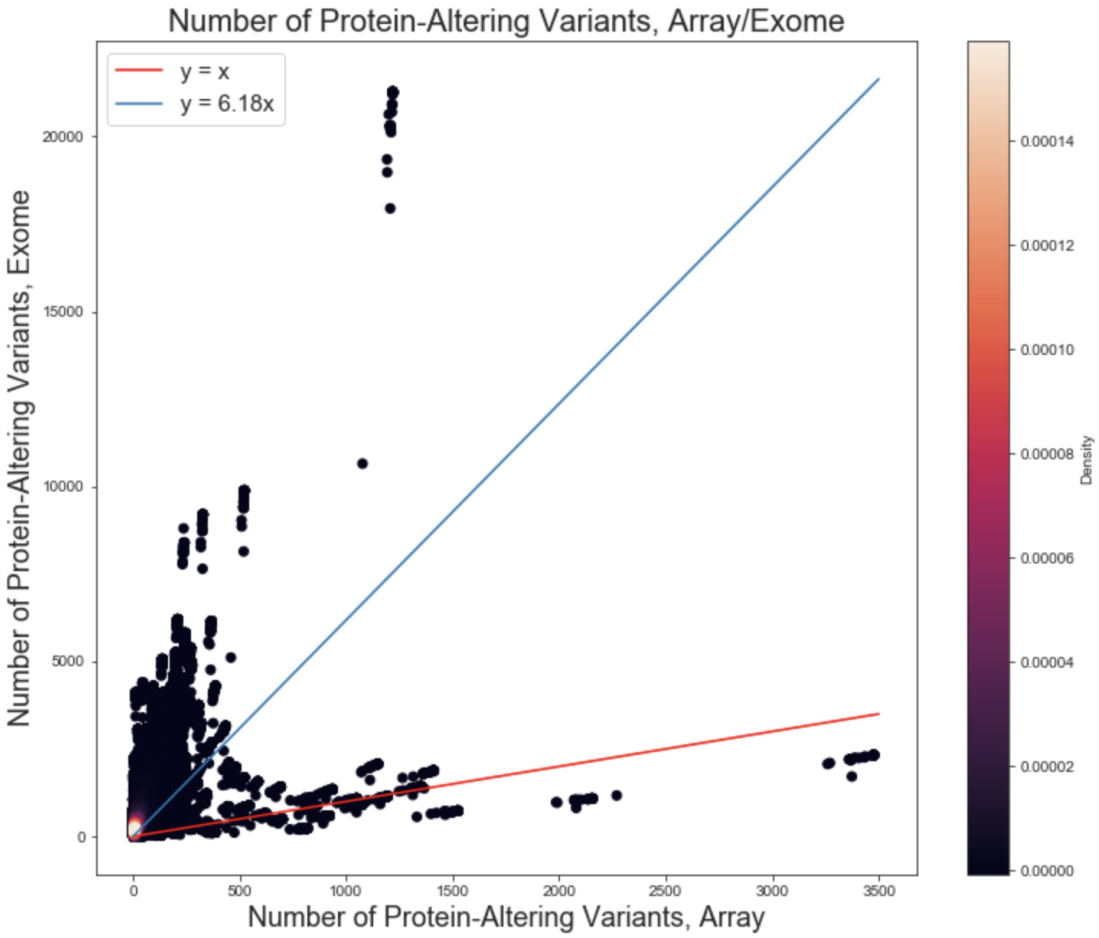


Figure S1. **From array to exome.** Scatterplot showing the increase in number of protein-altering variants in genes used in the analysis when comparing array (*x*-axis) to exome (*y*-axis) data. Data is taken from MRP calculations across 35 biomarker traits within the UK Biobank. Color shows density of points as according to colorbar (right).

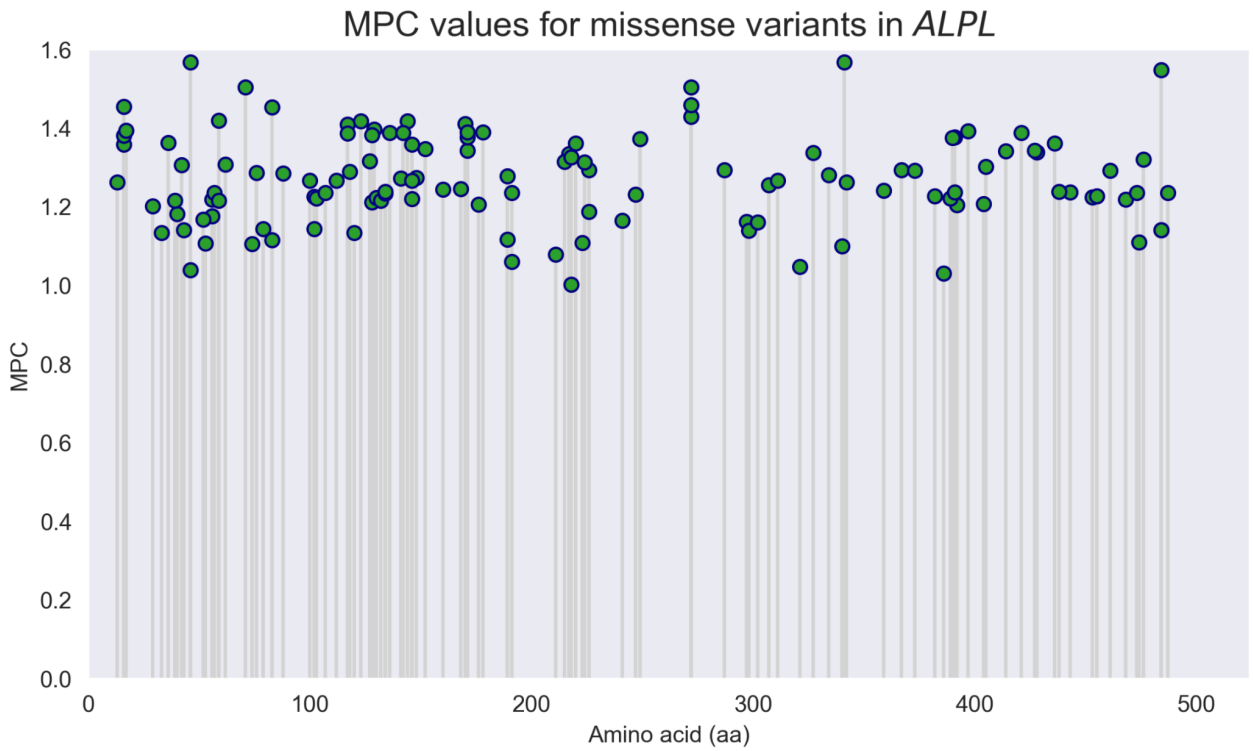


Figure S2. ***ALPL* gene plot.** Gene plot showing variants for which MPC pathogenicity information was incorporated, resulting in a power gain for *ALPL* gene that encodes alkaline phosphatase; for the Alkaline phosphatase phenotype, the incorporation of this information resulted in a $\log_{10}$BF gain of 34 (**Table S2**).

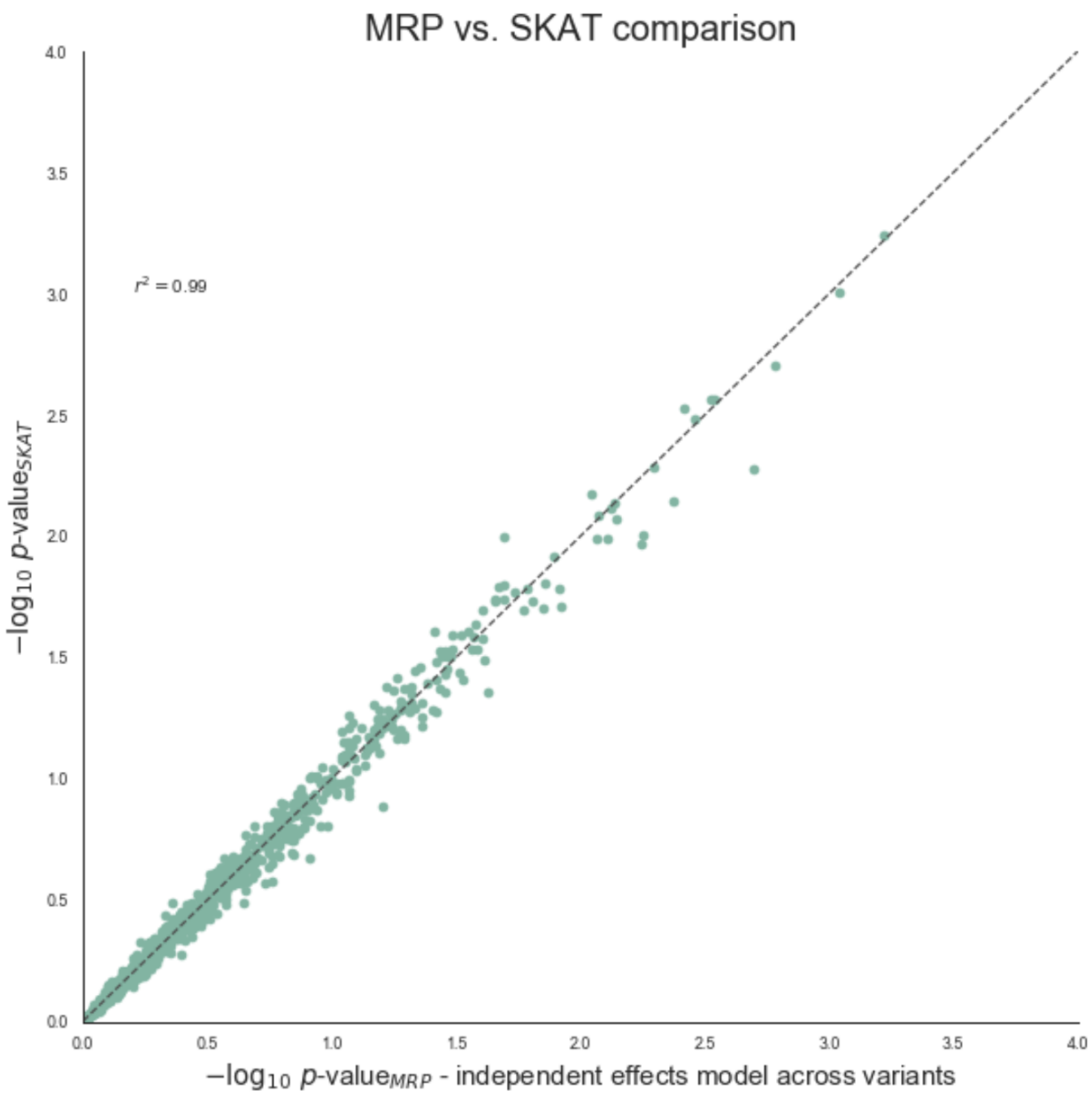


Figure S3. Comparisons of -$\log_{10}$ *p*-values: MRP with independent effects model across variants, SKAT. Correlation coefficient is 0.99. Line shown is $y = x$.

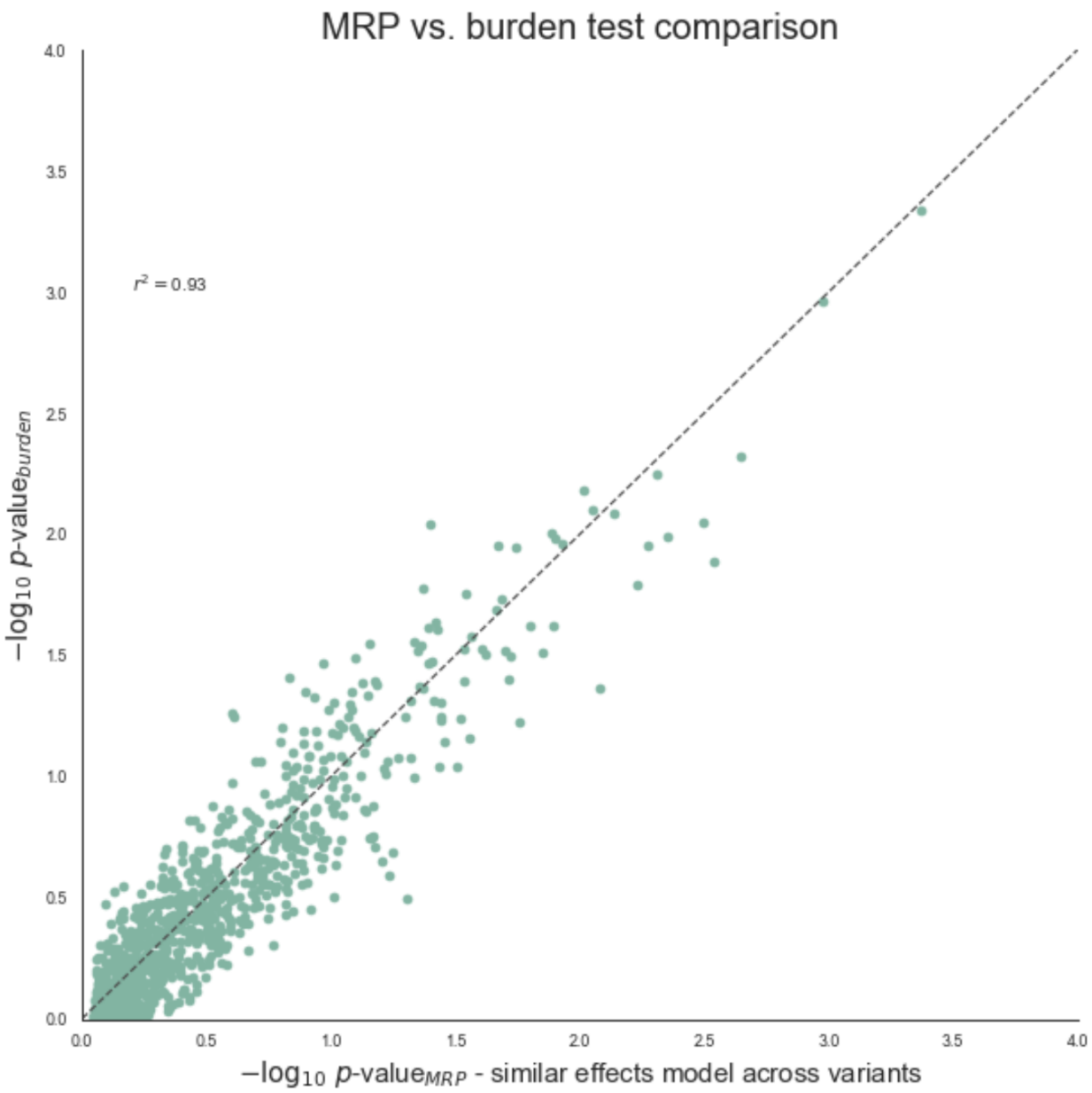


Figure S4. Comparisons of -$\log_{10}$ *p*-values: MRP with similar effects model across variants, burden test. Correlation coefficient is 0.93. Line shown is $y = x$.

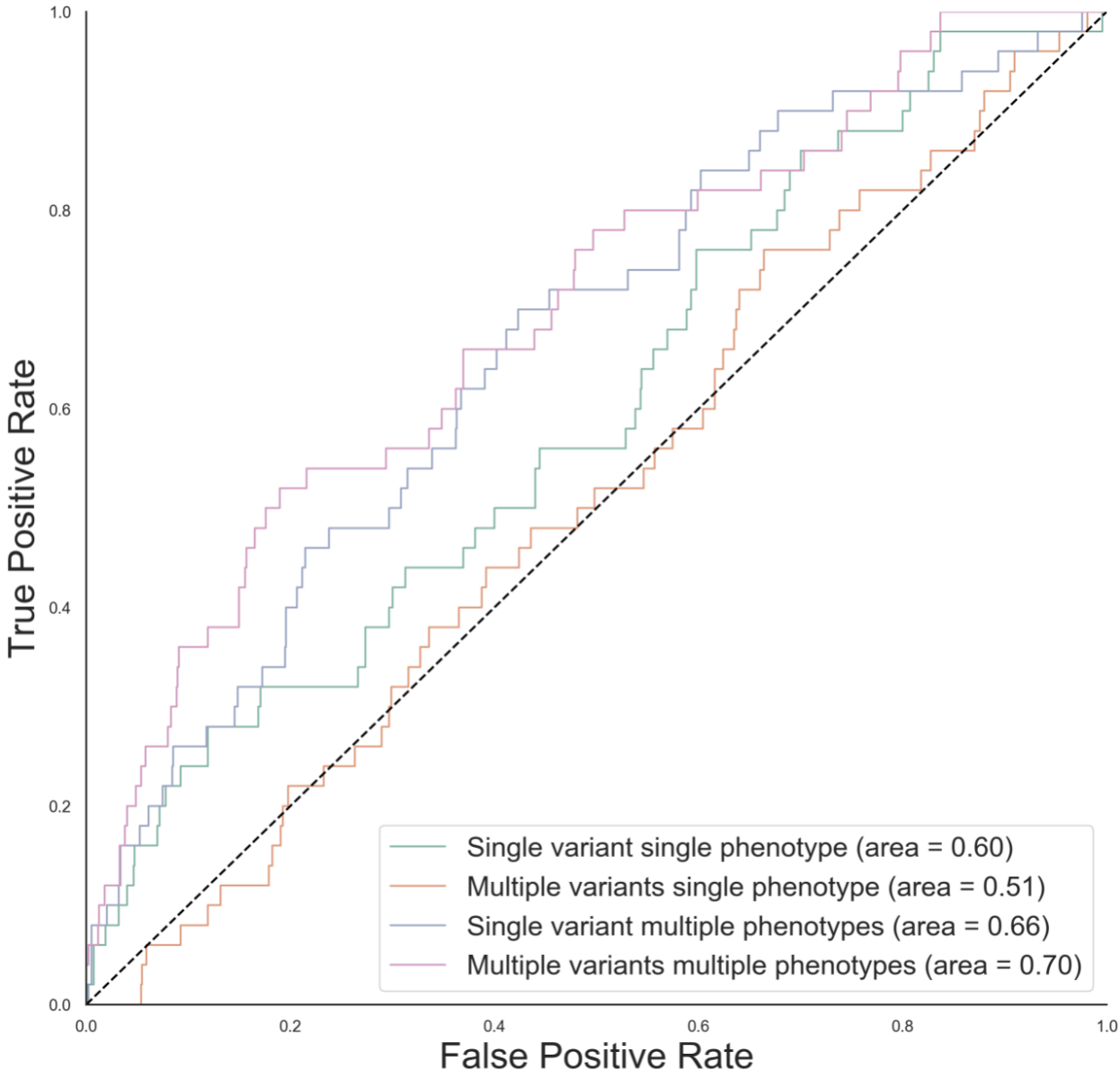


Figure S5. From single-variant and single-phenotype to multiple-variant and multiple-phenotype gene discovery: when only a single variant has true effect. ROC curves for detecting simulated gene association to any of the phenotypes using single variant/single phenotype association (turquoise) to multiple-variant and multiple-phenotype association (pink). If only a single variant has true effect, adding data from multiple variants demonstrates no improvement (and sometimes detriment).

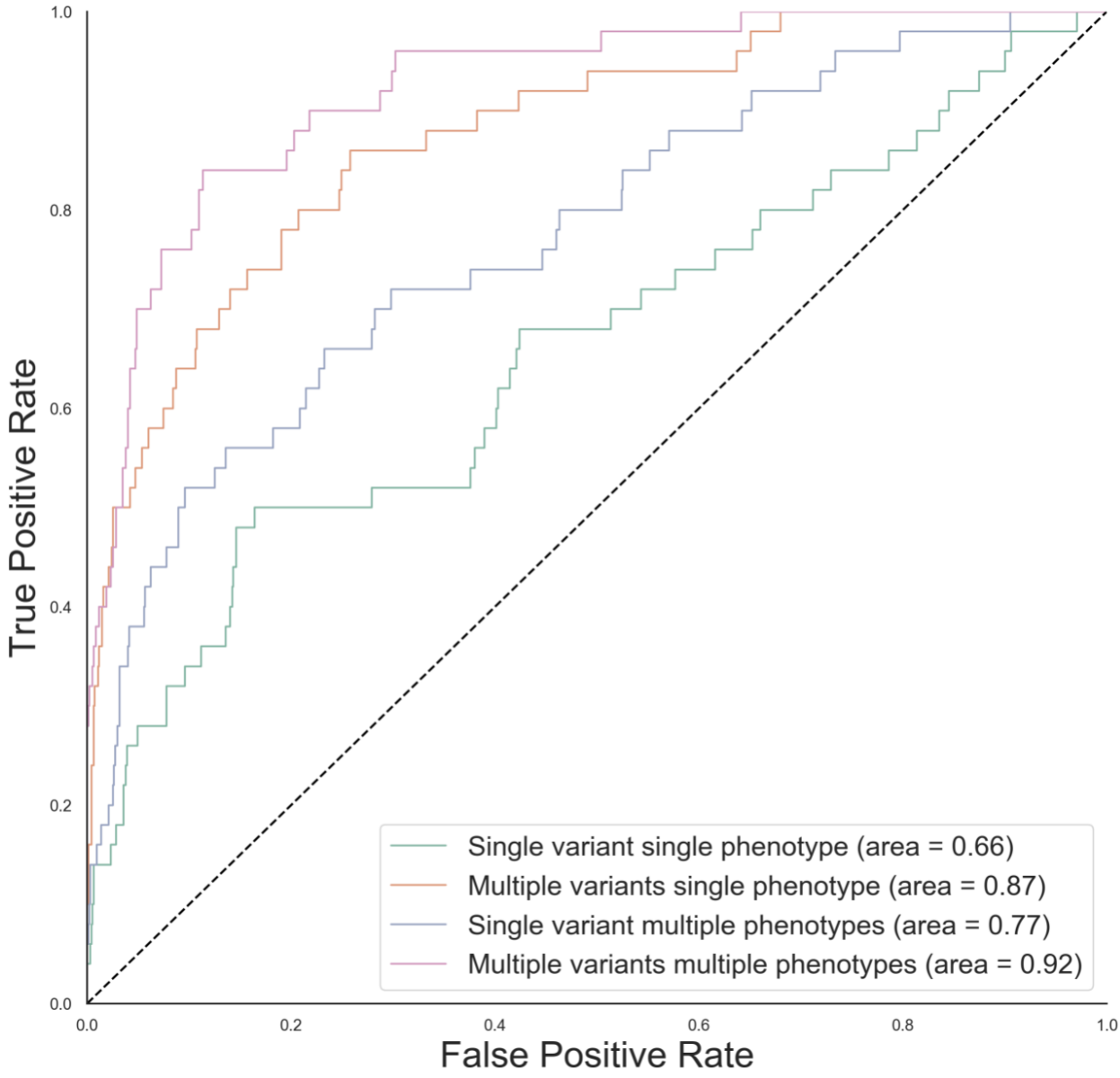
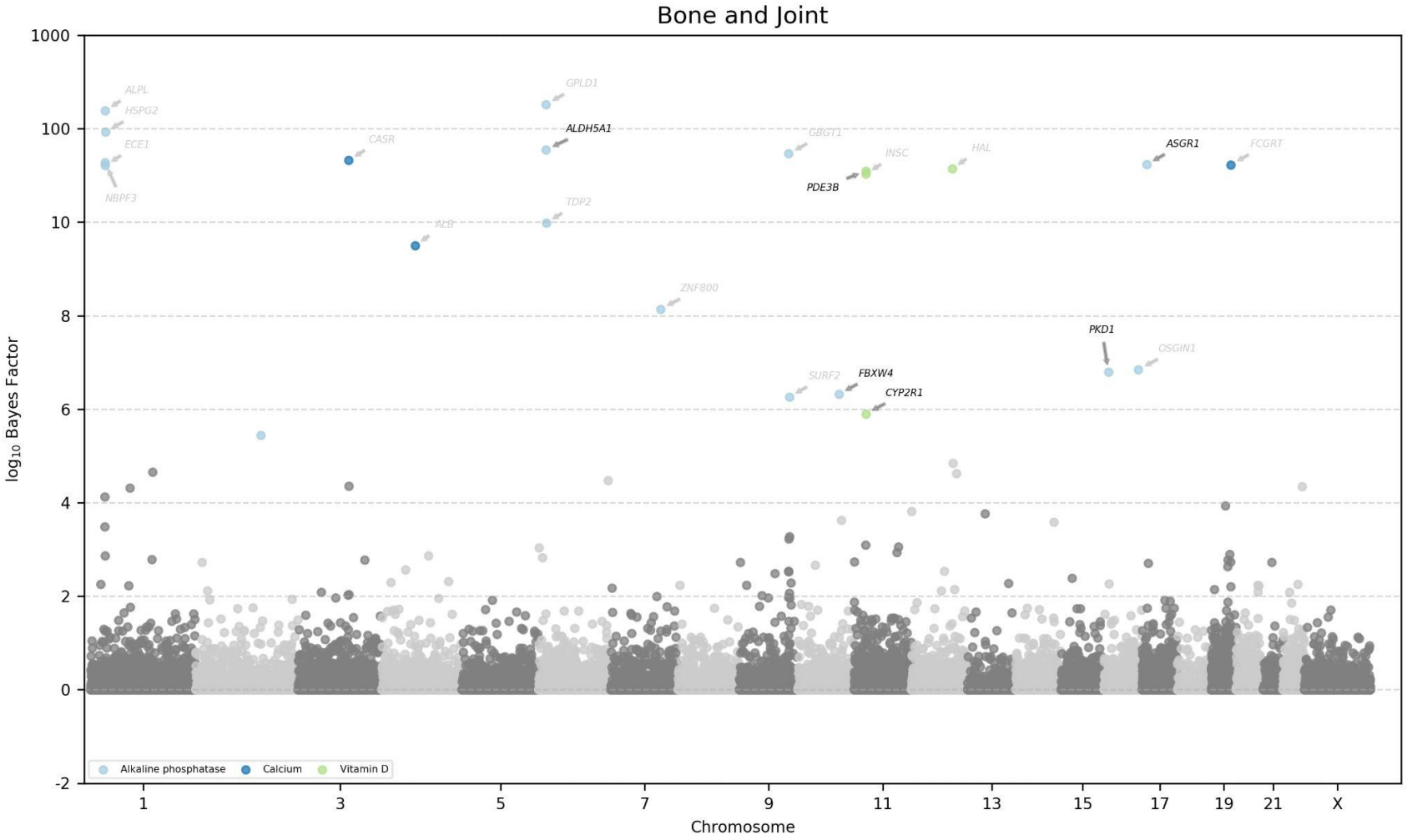
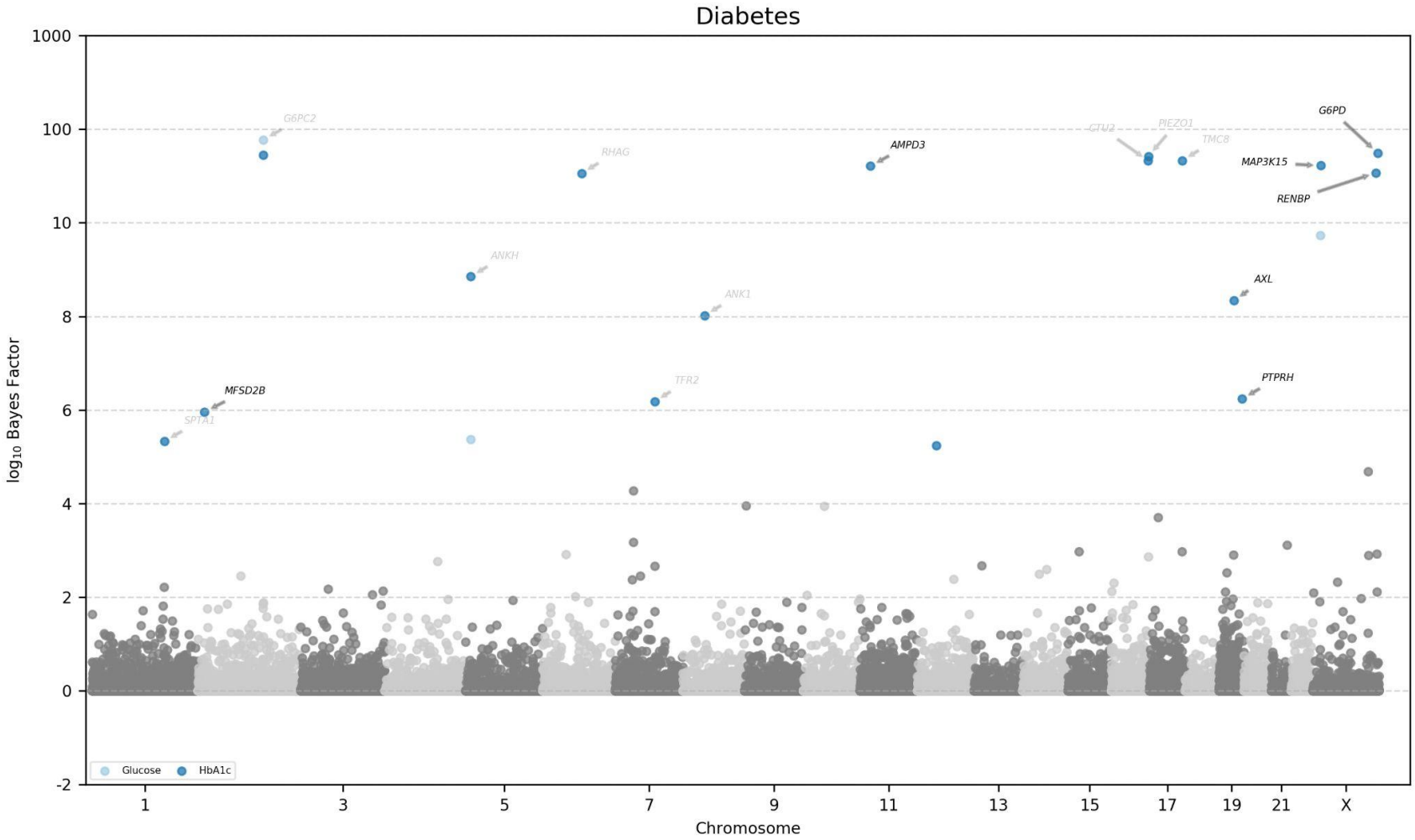
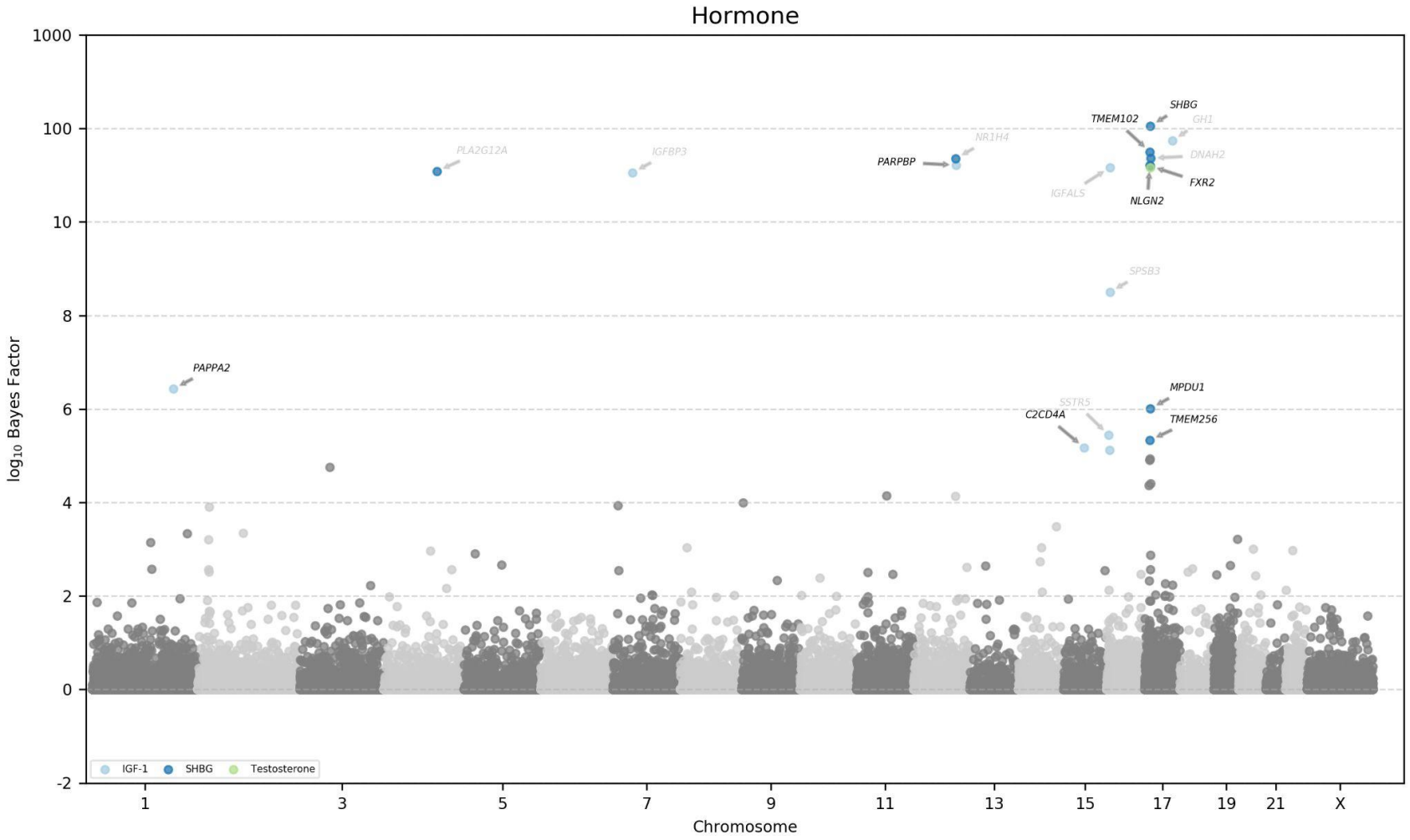
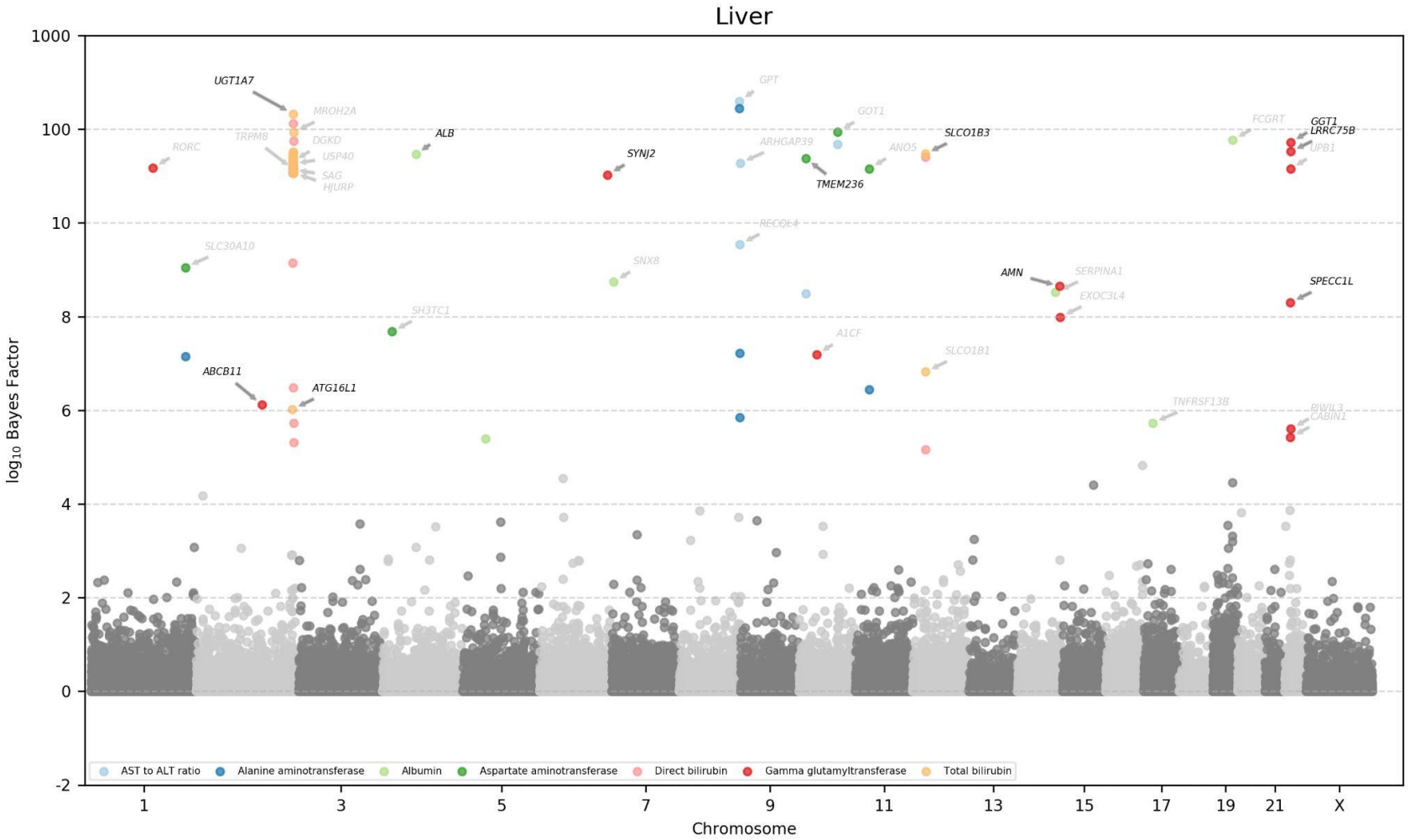


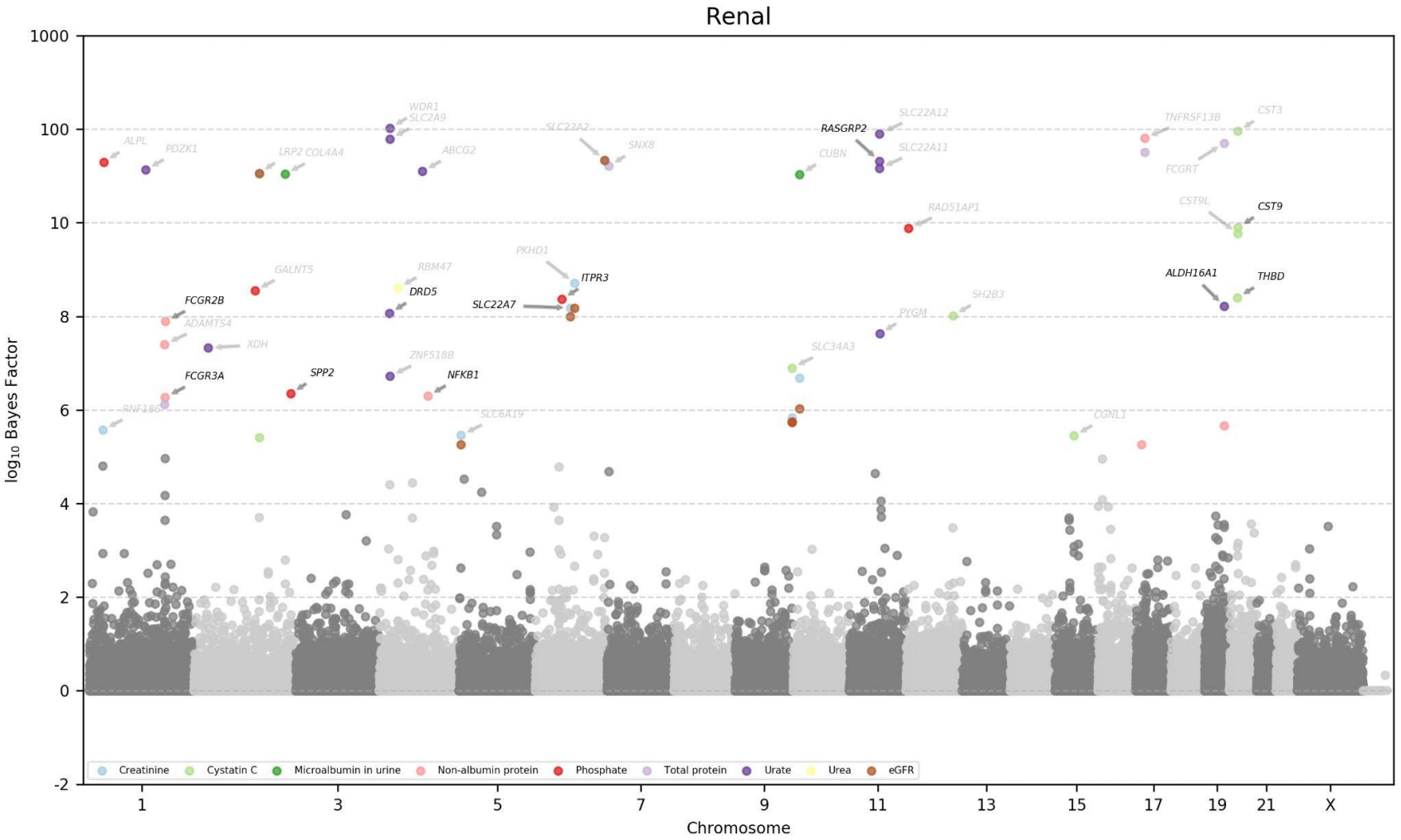
Figure S6. From single-variant and single-phenotype to multiple-variant and multiple-phenotype gene discovery: when phenotypes are independent. ROC curves for detecting simulated gene association to any of the phenotypes using single variant/single phenotype association (turquoise) to multiple-variant and multiple-phenotype association (pink). If phenotypes are truly independent, there may still be information across the genotypes that allows for improvement of performance, but not as drastic of an improvement as found when there is a clear correlation structure across phenotypes.


Bone and Joint
1000
100
10
8
6
4
2
0
-2
log10 Bayes Factor
ALPL
HSPG2
ECE1
NBPF3
CASR
GPLD1
ALDH5A1
TDP2
ALB
ZNF800
GBGT1
INSC
PDE3B
HAL
ASGR1
FCGRT
PKD1
OSGIN1
SURF2
FBXW4
CYP2R1
Alkaline phosphatase
Calcium
Vitamin D
1
3
5
7
9
11
13
15
17
19
21
X
Chromosome


Diabetes
1000
100
10
8
6
4
2
0
-2
log10 Bayes Factor
G6PC2
RHAG
AMPD3
GTU2
PIEZO1
TMC8
G6PD
MAP3K15
RENBP
ANKH
ANK1
AXL
MFSD2B
SPTA1
TFR2
PTPRH
Glucose
HbA1c
1
3
5
7
9
11
13
15
17
19
21
X
Chromosome


Hormone
1000
100
10
8
6
4
2
0
-2
log10 Bayes Factor
PAPPA2
PLA2G12A
IGFBP3
PARPBP
NR1H4
SHBG
TMEM102
GH1
DNAH2
FXR2
NLGN2
IGFALS
SPSB3
MPDU1
SSTR5
C2CD4A
TMEM256
IGF-1
SHBG
Testosterone
1
3
5
7
9
11
13
15
17
19
21
X
Chromosome


Liver
1000
100
10
8
6
4
2
0
-2
log10 Bayes Factor
UGT1A7
MROH2A
TRPM8
DGKD
RORC
USP40
SAG
HJURP
ALB
SYNJ2
GPT
GOT1
ARHGAP39
ANO5
SLCO1B3
TMEM236
FCGRT
GGT1
LRRC75B
UPB1
RECQL4
SLC30A10
SNX8
AMN
SERPINA1
EXOC3L4
SPECC1L
SH3TC1
A1CF
SLCO1B1
ABCB11
ATG16L1
TNFRSF13B
PIWIL3
CABIN1
AST to ALT ratio
Alanine aminotransferase
Albumin
Aspartate aminotransferase
Direct bilirubin
Gamma glutamyltransferase
Total bilirubin
1
3
5
7
9
11
13
15
17
19
21
X
Chromosome


Renal
1000
100
10
8
6
4
2
0
-2
log10 Bayes Factor
WDR1
SLC2A9
ALPL
PDZK1
LRP2
COL4A4
ABCG2
SLC22A2
SNX8
RASGRP2
SLC22A12
CUBN
SLC22A11
TNFRSF13B
CST3
FCGRT
RAD51AP1
CST9L
CST9
PKHD1
ITPR3
GALNT5
RBM47
DRD5
SLC22A7
FCGR2B
ADAMTS4
XDH
PYGM
SH2B3
ALDH16A1
THBD
SLC34A3
ZNF518B
SPP2
NFKB1
FCGR3A
RNF186
SLC6A19
CGNL1
Creatinine
Cystatin C
Microalbumin in urine
Non-albumin protein
Phosphate
Total protein
Urate
Urea
eGFR
1
3
5
7
9
11
13
15
17
19
21
X
Chromosome

Figures S7-11. Manhattan plots showing $\log_{10}$ BF under an independent effects variant model amongst protein-altering variants for 5 categories across 35 biomarkers. These include: Bone and Joint, Diabetes, Hormone, Liver, and Renal traits. Scale is logarithmic after $\log_{10}$ BF ≥ 10. Genes found in Sinnott-Armstrong, et.al.[1] are annotated in grey, whereas the other genes are annotated in black. Colors indicate different traits as indicated at the bottom of the plots.

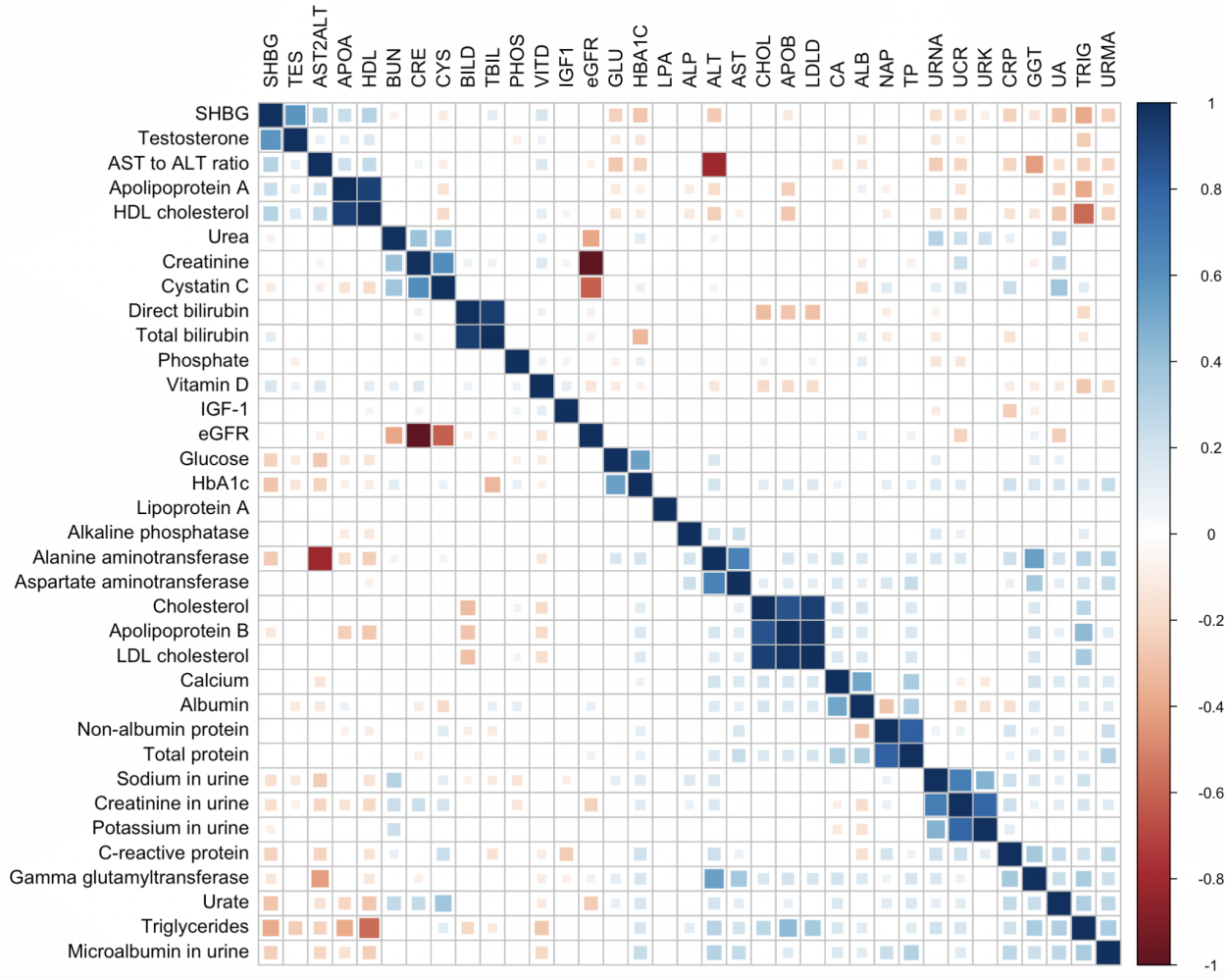


Figure S12. **LD-score regression-based genetic correlation plots of all 35 biomarkers included in the multi-trait analyses.** The traits are ordered by hierarchical clustering. Blue implies positive and red implies negative correlation coefficients as indicated by the colorbar (right).

Supplemental Tables

Trait	gene	Number of PAVs, array	log_{10}BF, array	Number of PAVs, exome	log_{10}BF, exome	log_{10}BF Difference
Total bilirubin	*UGT1A7*	5	1.2	247	213	211.8
Direct bilirubin	*UGT1A7*	5	0.6	228	133	132.4
Lipoprotein A	*PLG*	57	38.9	583	165	126.1
SHBG	*SHBG*	7	2.7	284	114	111.3
LDL cholesterol	*PCSK9*	94	4.0	759	99	95.0
Total bilirubin	*MROH2A*	33	4.6	1649	85.8	81.2
Apolipoprotein B	*PCSK9*	94	3.1	756	80.7	77.6
Cholesterol	*PCSK9*	94	4.0	760	80.9	76.9
IGF-1	*GH1*	5	2.1	301	55.1	53.0
Direct bilirubin	*MROH2A*	33	2.9	1497	55.7	52.8
Gamma glutamyltransferase	*GGT1*	5	0.008	545	52.1	52.1
Triglycerides	*ANGPTL3*	7	-0.02	337	39.9	39.9
Cholesterol	*ANGPTL3*	7	-0.6	337	34.3	34.9
Cholesterol	*APC*	1409	-34.7	1882	-0.5	34.2
LDL cholesterol	*APC*	1410	-33.7	1882	-0.5	33.2
Apolipoprotein B	*APC*	1405	-32.7	1876	-0.7	32.0
Total bilirubin	*UGT1A5*	12	2.0	225	33	31.0

Albumin	*APC*	1366	-31.6	1807	-1.2	30.4
Vitamin D	*APC*	1379	-29.8	1828	0.2	30.0
Creatinine	*APC*	1411	-31.4	1883	-1.9	29.5

Table S1. Genes with considerable power gain in exome data as compared to array data.

Trait	Gene	Number of PAVs	log_{10}BF without MPC	Number of MPC-augmented PAVs	Number of pLI-augmented PAVs	log_{10}BF with MPC	log_{10}BF Difference
Alkaline phosphatase	*ALPL*	198	126	93	0	160	34
Lipoprotein A	*LPA*	512	109	20	0	114	5
Apolipoprotein A	*APOA1*	102	11.7	30	0	15.7	4
HDL cholesterol	*APOA1*	103	9.36	30	0	13.2	3.84
Aspartate aminotransferase	*SLC30A10*	112	3.76	50	6	7.2	3.44
Phosphate	*ALPL*	192	10.9	91	0	14.3	3.4
Lipoprotein A	*IGF2R*	763	29.8	153	27	33.1	3.3
HDL cholesterol	*SCARB1*	220	5.45	66	0	8.29	2.84
Apolipoprotein B	*APOE*	142	5.48	60	0	8.27	2.79
Alanine aminotransferase	*SLC30A10*	112	2.94	50	6	5.56	2.62

Table S2. **Power comparison between variant annotation-based MRP and MPC/pLI-augmented MRP analyses across 35 biomarkers.** We see considerable gains in power in several gene/trait combinations.

# Studies	# Phenotypes	Phenotype Specification	Genetic Datatype	Average Runtime (DD:HH:MM)
Single	Single	Binary	Array	00:00:02
Single	Single	Quantitative	Array	00:00:08
Multiple	Single	Binary	Array	00:00:20
Multiple	Single	Quantitative	Array	00:00:40
Single	Single	Binary	Exome	00:02:00
Single	Single	Quantitative	Exome	00:08:00
Multiple	Single	Binary	Exome	00:05:00
Multiple	Single	Quantitative	Exome	01:00:00
Single	Multiple	Binary	Array	00:02:00
Single	Multiple	Quantitative	Array	00:05:00
Single	Multiple	Binary	Exome	02:00:00
Single	Multiple	Quantitative	Exome	05:00:00

Table S3. **Computation times for various MRP analyses.** One node with 16 cores and 200 GB RAM was used.

Supplemental References

1. Sinnott-Armstrong, N., Tanigawa, Y., Amar, D., Mars, N., Benner, C., Aguirre, M., Venkataraman, G.R., Wainberg, M., Ollila, H.M., Kiiskinen, T., et al. (2021). Genetics of 35 blood and urine biomarkers in the UK Biobank. Nat. Genet. *53*, 185–194.