Supplement to:

Genome-wide association study of serum metabolites in the African American Study of Kidney Disease and Hypertension

Shengyuan Luo, MBBS, MHS (1,2)*, Elena V. Feofanova, MS (3)*, Adrienne Tin, MS, PhD (1,2), Sarah Tung (4), Eugene P. Rhee, MD (5), Josef Coresh, MD, PhD (1), Dan E. Arking, PhD (6), Aditya Surapaneni, PhD (1,2), Pascal Schlosser, PhD (7), Yong Li, MD (7), Anna Köttgen, MD, MPH (1, 7), Bing Yu, PhD (3), Morgan E. Grams, MD, PhD (1,2,8)

1) Department of Epidemiology, Johns Hopkins University Bloomberg School of Public Health, Baltimore MD, USA
2) Welch Center for Prevention, Epidemiology, and Clinical Research, Johns Hopkins University, Baltimore MD, USA
3) Human Genetics Center, The University of Texas Health Science Center at Houston, Houston, TX, USA
4) Johns Hopkins Whiting School of Engineering, Baltimore MD, USA
5) Division of Nephrology, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA
6) McKusick-Nathans Institute, Department of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, USA
7) Institute of Genetic Epidemiology, Faculty of Medicine and Medical Center – University of Freiburg, Freiburg, Germany
8) Division of Nephrology, Department of Medicine, Johns Hopkins University, Baltimore MD, USA

Running head: GWAS of serum metabolites in AASK

* These authors are co-first authors.

Corresponding author:

Shengyuan Luo, MBBS, MHS
shengyuan.luo@jhu.edu
2024 E. Monument, Suite 2-500
Baltimore, MD 21205

Table S1. Statistically significant associations (N=42) between index SNPs representing 23 genomic loci and 38 serum metabolites in AASK.

Distinct index SNP-metabolite association	**Distinct Loci**	**Chromosomal location**	**Mapped genes**	**Index SNP**	**Coded/ Reference Allele**	**Coded Allele Frequency**	**Associated metabolite**	**Beta coefficient (p value)**	**Beta coefficient (p value) in the ARIC Study***
1	1	1:45907026-47417369	*CYP4A11, FAAH*	rs4507958	T/A	0.373	10-undecenoate (11:1n1)	0.42 (3.78E-14)	NA
2				rs324418	G/A	0.361	N-palmitoylglycine	0.42 (6.26E-12)	NA
3	2	1:53444415-54444415	*HSPB11, LRRC42, DIO1, TMEM59*	rs79319225	A/G	0.029	3-methyladipate	0.75 (1.29E-10)	-0.12 (0.43)
4	3	1:118887090-119887090	*HAO2*	rs7528838	T/A	0.252	alpha-hydroxyisovalerate	-0.57 (4.18E-19)	NA
5				rs7528838	T/A	0.252	2-hydroxy-3-methylvalerate	-0.47 (1.03E-14)	-0.27 (5.97E-07)
6	4	2:73105659-74173773	*ALMS1, ALMS1P, NAT8, TPRKB*	rs7587577	T/C	0.547	N-acetylasparagine	0.72 (1.47E-41)	NA
7				rs7587577	T/C	0.547	N-acetyl-3-methylhistidine	0.39 (6.11E-13)	NA
8				rs11126412	T/A	0.392	N-acetyl-1-methylhistidine	0.52 (5.40E-23)	NA
9				rs11126412	T/A	0.392	N-acetylphenylalanine	0.36 (2.42E-11)	NA
10				rs13431529	C/G	0.535	N2-acetyllysine	0.61 (3.26E-31)	0.62 (1.18E-45)
11				rs13409366	G/A	0.547	N-acetylcitrulline	0.68 (7.26E-37)	NA
12				rs13409366	G/A	0.547	N-acetylglutamine	0.62 (6.84E-29)	NA
13				rs13409366	G/A	0.547	N-acetylkynurenine	0.33 (7.26E-37)	NA
14				rs6546854	G/A	0.547	N-acetylleucine	0.53 (1.44E-21)	0.57 (1.39E-34)

15				rs10168931	A/G	0.54	N-acetylarginine	0.66 (1.13E-32)	0.68 (2.85E-49)
16				rs10206899	C/T	0.545	N-delta-acetylornithine	-0.69 (7.24E-39)	NA
17	5	2:129396140-130396140	*UGT1A3*, *UGT1A4*, *UGT1A10*	rs951389793	T/TA	0.014	cysteinylglycine	1.12 (1.42E-10)	0.01 (0.97)
18	6	2: 233257337-234264663	AC114812.9, RPL17P11, *UGT1A3*, *UGT1A4*, *UGT1A5*, *UGT1A6*, *UGT1A7*, *UGT1A8*, *UGT1A9*, *UGT1A10*	rs1976391	G/A	0.453	biliverdin	0.42 (3.69E-17)	NA
19				rs4148325	T/C	0.454	bilirubin (E,E)	0.36 (3.82E-12)	NA
20	7	2:236306573-237306573	AC011286.1, AC112715.2	rs62184450	C/A	0.029	cysteinylglycine	0.77 (1.80E-10)	-0.20 (0.18)
21	8	3:51489004-52620872	*POC1A*, *ACY1*, *ABHD14A-ACY1*	rs73835707	C/T	0.021	N-acetylmethionine	1.38 (5.33E-13)	1.31 (7.12E-15)
22				rs73835707	C/T	0.021	N-acetylmethionine sulfoxide	1.24 (6.06E-13)	NA
23				rs2229152	T/C	0.022	N-acetylalanine	1.26 (2.21E-12)	NA
24	9	4:185748013-186748013	*KLKB1*, *CYP4V2*, *F11*, *CSMD1*, *PTHLH*	rs1973612	T/C	0.587	histidylalanine	0.36 (5.08E-11)	0.04 (0.16)
25	10	5:34536259-35536259	*AGXT2*	rs344514	C/T	0.822	3-aminoisobutyrate	0.96 (1.33E-46)	NA
26	11	6:125264108-126264108	*HEY2*, *NCOA7*	rs60132035	A/G	0.016	sphingomyelin (d18:1/20:1, d18:2/20:0)	1.35 (1.14E-10)	-0.23 (0.27)

27	12	7:100192604-101692604	*MOGAT3, AP1S1*	rs7780766	T/C	0.297	N-acetylhistidine	-0.50 (6.32E-15)	NA
28	13	7:141183499-142183499	*KIAA1147*	rs139981949	T/C	0.018	glycohyocholate	1.21 (9.15E-12)	-0.17 (0.38)
29	14	8:90241098-91241098	*TMEM64*	rs76541360	T/A	0.072	1-arachidonoyl-GPI	0.73 (2.93E-12)	0.16 (0.10)
30	15	8:143564265-144564265	*OPLAH, CTD-3065J16.9, EXOSC4, GPAA1, KIAA1875, MAF1*	rs59095288	C/T	0.309	6-oxopiperidine-2-carboxylate	0.62 (1.92E-22)	0.50 (3.37E-24)
31	16	9:128422557-129422736	*LRRC8A, PHYHD1, SH3GLB2*	rs55758160	G/A	0.548	2'-O-methyluridine	0.72 (1.97E-42)	NA
32				rs57294583	G/A	0.547	2'-O-methylcytidine	0.39 (1.97E-12)	0.80 (6.88E-69)
33	17	11:48528971-49528971	*PTPRJ, FOLH1*	rs142514677	C/CAG	0.059	N-acetyl-aspartyl-glutamate	-0.79 (5.35E-14)	-0.67 (1.57E-09)
34	18	11:54393863-55393863	*OR4C1*	rs150556827	T/C	0.056	N-acetyl-aspartyl-glutamate	-0.86 (1.11E-12)	-0.48 (0.04)
35	19	11:61320833-62320833	*FADS1, FADS2, TMEM258*	rs174564	G/A	0.119	1-arachidonoyl-GPC (20:4n6)	-0.54 (3.31E-11)	NA
36	20	11:67144985-68144985	*ACY3*, AP003385.2	rs115780269	T/C	0.025	N-acetyltryptophan	1.64 (1.11E-22)	1.23 (1.72E-10)
37				rs115780269	T/C	0.025	N-acetylkynurenine	1.07 (2.18E-12)	NA
38				rs115780269	T/C	0.025	N-acetyltyrosine	1.42 (4.15E-17)	1.24 (2.79E-10)
39				rs115780269	T/C	0.025	N-acetylphenylalanine	1.39 (8.18E-16)	NA
40	21	12:20701434-21701434	*SLCO1B1*	rs114419265	T/C	0.048	glycochenodeoxycholate glucuronide	0.93 (1.44E-14)	NA
41	22	12:120242216-121242216	*ACADS, SPPL3,*	rs34708625	C/T	0.061	ethylmalonate	0.90 (5.02E-14)	NA

			UNC119B, RP11-173P15.7						
42	23	17:77706736-78706736	*AFMID*	rs114080902	T/C	0.043	N-formylanthranilic acid	1.00 (1.80E-13)	0.89 (9.13E-15)

* Replication analysis was performed in the ARIC Study for associations that were determined to be novel.

Table S2. Correlations between index SNPs and metabolites within distinct genomic loci.

<table>
<tr><th>Distinct Loci</th><th>Chromosomal location</th><th>Index SNP</th><th>Associated metabolite</th><th colspan="3">Correlation between SNPs</th><th colspan="3">Correlations between metabolites</th></tr>
<tr><td rowspan="2">1</td><td rowspan="2">1:45907026-47417369</td><td>rs4507958</td><td>10-undecenoate (11:1n1)</td><td colspan="3" rowspan="2">-0.03</td><td colspan="3" rowspan="2">0.15</td></tr>
<tr><td>rs324418</td><td>N-palmitoylglycine</td></tr>
<tr><td>2</td><td>1:53444415-54444415</td><td>rs79319225</td><td>3-methyladipate</td><td colspan="3">NA</td><td colspan="3">NA</td></tr>
<tr><td rowspan="2">3</td><td rowspan="2">1:118887090-119887090</td><td>rs7528838</td><td>alpha-hydroxyisovalerate</td><td colspan="3" rowspan="2">(same SNP)</td><td colspan="3" rowspan="2">0.89</td></tr>
<tr><td>rs7528838</td><td>2-hydroxy-3-methylvalerate</td></tr>
<tr><td rowspan="11">4</td><td rowspan="11">2:73105659-74173773</td><td>rs7587577</td><td>N-acetylasparagine</td><td colspan="3" rowspan="11">Range: 0.73 to 1.0
Median, 0.98
25th-75th (0.73-1.0)</td><td colspan="3" rowspan="11">Range: -0.08 to 0.78;
Median, 0.54
25th-75th (0.40-0.61)</td></tr>
<tr><td>rs7587577</td><td>N-acetyl-3-methylhistidine</td></tr>
<tr><td>rs11126412</td><td>N-acetyl-1-methylhistidine</td></tr>
<tr><td>rs11126412</td><td>N-acetylphenylalanine</td></tr>
<tr><td>rs13431529</td><td>N2-acetyllysine</td></tr>
<tr><td>rs13409366</td><td>N-acetylcitrulline</td></tr>
<tr><td>rs13409366</td><td>N-acetylglutamine</td></tr>
<tr><td>rs13409366</td><td>N-acetylkynurenine</td></tr>
<tr><td>rs6546854</td><td>N-acetylleucine</td></tr>
<tr><td>rs10168931</td><td>N-acetylarginine</td></tr>
<tr><td>rs10206899</td><td>N-delta-acetylornithine</td></tr>
<tr><td>5</td><td>2:129396140-130396140</td><td>rs951389793</td><td>cysteinylglycine</td><td colspan="3">NA</td><td colspan="3">NA</td></tr>
<tr><td rowspan="2">6</td><td rowspan="2">2: 233257337-234264663</td><td>rs1976391</td><td>biliverdin</td><td colspan="3" rowspan="2">1.0</td><td colspan="3" rowspan="2">0.75</td></tr>
<tr><td>rs4148325</td><td>bilirubin (E,E)</td></tr>
<tr><td>7</td><td>2:236306573-237306573</td><td>rs62184450</td><td>cysteinylglycine</td><td colspan="3">NA</td><td colspan="3">NA</td></tr>
<tr><td rowspan="3">8</td><td rowspan="3">3:51489004-52620872</td><td>rs73835707</td><td>N-acetylmethionine</td><td>-</td><td></td><td></td><td>-</td><td></td><td></td></tr>
<tr><td>rs73835707</td><td>N-acetylmethionine sulfoxide</td><td>same SNP</td><td>-</td><td></td><td>0.59</td><td>-</td><td></td></tr>
<tr><td>rs2229152</td><td>N-acetylalanine</td><td>0.96</td><td>0.96</td><td>-</td><td>0.41</td><td>0.45</td><td>-</td></tr>
<tr><td>9</td><td>4:185748013-186748013</td><td>rs1973612</td><td>histidylalanine</td><td colspan="3">NA</td><td colspan="3">NA</td></tr>
<tr><td>10</td><td>5:34536259-35536259</td><td>rs344514</td><td>3-aminoisobutyrate</td><td colspan="3">NA</td><td colspan="3">NA</td></tr>
<tr><td>11</td><td>6:125264108-126264108</td><td>rs60132035</td><td>sphingomyelin (d18:1/20:1, d18:2/20:0)</td><td colspan="3">NA</td><td colspan="3">NA</td></tr>
<tr><td>12</td><td>7:100192604-101692604</td><td>rs7780766</td><td>N-acetylhistidine</td><td colspan="3">NA</td><td colspan="3">NA</td></tr>
<tr><td>13</td><td>7:141183499-142183499</td><td>rs139981949</td><td>glycohyocholate</td><td colspan="3">NA</td><td colspan="3">NA</td></tr>
<tr><td>14</td><td>8:90241098-91241098</td><td>rs76541360</td><td>1-arachidonoyl-GPI</td><td colspan="3">NA</td><td colspan="3">NA</td></tr>
</table>

<table>
<tr><td>15</td><td>8:143564265-144564265</td><td>rs59095288</td><td>6-oxopiperidine-2-carboxylate</td><td>NA</td><td colspan="4">NA</td></tr>
<tr><td rowspan="2">16</td><td rowspan="2">9:128422557-129422736</td><td>rs55758160</td><td>2'-O-methyluridine</td><td rowspan="2">1.00</td><td colspan="4" rowspan="2">0.07</td></tr>
<tr><td>rs57294583</td><td>2'-O-methylcytidine</td></tr>
<tr><td>17</td><td>11:48528971-49528971</td><td>rs142514677</td><td>N-acetyl-aspartyl-glutamate</td><td>NA</td><td colspan="4">NA</td></tr>
<tr><td>18</td><td>11:54393863-55393863</td><td>rs150556827</td><td>N-acetyl-aspartyl-glutamate</td><td>NA</td><td colspan="4">NA</td></tr>
<tr><td>19</td><td>11:61320833-62320833</td><td>rs174564</td><td>1-arachidonoyl-GPC (20:4n6)</td><td>NA</td><td colspan="4">NA</td></tr>
<tr><td rowspan="4">20</td><td rowspan="4">11:67144985-68144985</td><td>rs115780269</td><td>N-acetyltryptophan</td><td rowspan="4">(same SNP)</td><td>-</td><td></td><td></td><td></td></tr>
<tr><td>rs115780269</td><td>N-acetylkynurenine</td><td>0.63</td><td>-</td><td></td><td></td></tr>
<tr><td>rs115780269</td><td>N-acetyltyrosine</td><td>0.77</td><td>0.59</td><td>-</td><td></td></tr>
<tr><td>rs115780269</td><td>N-acetylphenylalanine</td><td>0.73</td><td>0.69</td><td>0.76</td><td>-</td></tr>
<tr><td>21</td><td>12:20701434-21701434</td><td>rs114419265</td><td>glycochenodeoxycholate glucuronide</td><td>NA</td><td colspan="4">NA</td></tr>
<tr><td>22</td><td>12:120242216-121242216</td><td>rs34708625</td><td>ethylmalonate</td><td>NA</td><td colspan="4">NA</td></tr>
<tr><td>23</td><td>17:77706736-78706736</td><td>rs114080902</td><td>N-formylanthranilic acid</td><td>NA</td><td colspan="4">NA</td></tr>
</table>

Table S3. Allele frequencies in general populations.

Distinct locus number	Index SNP	Coded allele in AASK	Coded allele frequency in AASK	Coded allele frequency in general populations
1	rs4507958	T/A	0.373	African = 0.368 European = 0.129
1	rs324418	G/A	0.361	African = 0.381 European = 0.256
2	rs79319225	A/G	0.029	African = 0.032 European = 0.0001
3	rs7528838	T/A	0.252	African = 0.273 European = 0.001
4	rs7587577	T/C	0.547	African = 0.548 European = 0.228
4	rs11126412	T/A	0.392	African = 0.372 European = 0.217
4	rs13431529	C/G	0.535	African = 0.531 European = 0.217
4	rs13409366	G/A	0.547	African = 0.548 European = 0.231
4	rs6546854	G/A	0.547	African = 0.548 European = 0.230
4	rs10168931	A/G	0.540	African = 0.529 European = 0.215
4	rs10206899	C/T	0.545	African = 0.556 European = 0.223
5	rs951389793	T/TA	0.014	NA
6	rs1976391	G/A	0.453	African = 0.449 European = 0.318
6	rs4148325	T/C	0.454	African = 0.458 European = 0.325
7	rs62184450	C/A	0.029	African = 0.030 European = 0.081
8	rs73835707	C/T	0.021	African = 0.016 European = 0.000
8	rs2229152	T/C	0.022	African = 0.018 European = 0.000
9	rs1973612	T/C	0.587	African = 0.601 European=0.519
10	rs344514	C/T	0.822	African = 0.801 European=0.998
11	rs60132035	A/G	0.016	African = 0.015 European = 0.003
12	rs7780766	T/C	0.297	African = 0.292 European = 0.082
13	rs139981949	T/C	0.018	African = 0.023 European = 0.000
14	rs76541360	T/A	0.072	African = 0.071 European = 0.000
15	rs59095288	C/T	0.309	African = 0.340

				European = 0.069
16	rs55758160	G/A	0.548	African = 0.515 European = 0.393
17	rs142514677	C/CAG	0.059	African = 0.056 European = 0.000
18	rs150556827	T/C	0.056	African = 0.057 European = 0.000
19	rs174564	G/A	0.119	African = 0.118 European = 0.346
20	rs115780269	T/C	0.025	African = 0.017 European = 0.000
21	rs114419265	T/C	0.048	African = 0.035 European = 0.000
22	rs34708625	C/T	0.061	African = 0.064 European = 0.264
23	rs114080902	T/C	0.043	African = 0.044 European = 0.000

Table S4. Pathway enrichment analysis.

Super-pathway	Total N of metabolites evaluated	N of metabolites associated with >=1 SNP	Fisher Exact p-value	Permutation p-value*
Amino acid	180 (28%)	24 (63%)	7E-06	0.005
Nucleotide	34 (5%)	3 (8%)	0.448	0.345
Cofactors and vitamins	23 (4%)	2 (5%)	0.624	0.335
Peptide	35 (5%)	1 (3%)	0.714	0.765
Lipid	349 (54%)	8 (21%)	1E-04	0.995
Other	31 (4%)	0 (0%)	-	-
Total	652	38	-	-

Table S5. Summary of published metabolite GWAS literature.

Study	Metabolites	Metabolite platform	Study population, N	African Americans in study population	CKD in study population and/or (e)GFR	Main findings	Overlap in findings with present study
Yousri 2018[S1]	1303 serum metabolites	Metabolon	Consanguineous Middle Eastern population (N=996)	0%	General population	21 common and 12 rare variant metabolic quantitative trait loci found and replicated, 7 of the common mQTLs were novel, all 12 of the rare mQTLs were also novel	*ACADS* and ethylmalonate, *PHYHD1* and 2'-O-methyluridine
Li 2018[S2]	139 serum metabolites, 41 urine metabolites, pairwise ratios, and fractional excretions	Biocrates Life Sciences AG, Austria	GCKD (N=1143)	0%	All participants had CKD, mean eGFR = 51 mL/min/1.73 m^2	Genome-wide significance for 25 serum metabolites, 2 urine metabolites, 259 serum and 14 urine metabolite ratios	*NAT8* and N-acetylornithine
Long 2017[S3]	644 serum metabolites including 156 unknown metabolites	Metabolon	TwinsUK registry (N=1960)	0%	General population	Genetic sequence variations at 101 loci associated with the levels of 246 (38%) metabolites (P $\leq 1.9\times10^{-11}$)	*ACADS* and ethylmalonate, *NAT8* and N-acetylasparagine, *A* , *NAT8* and N-acetyl-1-methylhistidine, *NAT8* and N-acetylphenylalanine, *NAT8* and N-acetylcitrulline, *NAT8* and N-acetylglutamine, *NAT8* and N-delta-acetylornithine, *FADS2* and 1-arachidonoyl-GPC

							(20:4n6), *CYP4A11* and N-palmitoylglycine, *FAAH* and N-palmitoylglycine, *CYP4A11* and 10-undecenoate (11:1n1), *AP1S1* and N-acetylhistidine, *AGXT2* and 3-aminoisobutyrate, *UGT1A* and biliverdin, *UGT1A* and bilirubin (E,E), *FADS2* and 1-arachidonoyl-GPC (20:4n6)
Davis 2017[S4]	72 blood lipid and lipoprotein traits	In-house platform	The METSIM study (N=8372)	0%	General population	Five novel genome-wide lipid and lipoprotein subclass association signals near established loci at *HIF3A*, *ADAMTS3*, *PLTP*, *LCAT*, and *LIPG*	None
Rhee 2016[S5]	217 plasma metabolites	Broad Institute (FHS) and Metabolon (ARIC)	FHS Offspring cohort (N=2076) and ARIC European Americans (N=1528)	0%	General population	Associations between *GMPS* and xanthosine, *HAL* and histidine, *PAH* and phenylalanine, and *UPB1* and ureidopropionate	None
Yu 2016[S6]	70 serum amino acids	Metabolon	ARIC African Americans (N discovery=1872) and Europeans (N replication=1552)	100% discovery, 0% replication	General population, mean eGFR = 104.33 mL/min/1.73 m^2 for the discovery	14 genetic loci-amino acid significant associations found and replicated, 4 of which were novel (DDC and 3-methoxytyrosine, VNN1 and acisoga, ACY1	*ALMS1P* and N-acetylphenylalanine, *ACY1* and N-acetylalanine

					cohort and 91.12 mL/min/1.73 m^2 for the replication cohort	and N-acetylalanine, ACY1 and N-acetylthreonine)	
Yu 2016[S7]	308 serum metabolites	Metabolon	ARIC African Americans (N=1361)	100%	General population	Loss-of-function variants influence metabolites, some of which are risk predictors or diagnostic biomarkers of disease. For example, loss of function mutations in *SLCO1B1* elevate the levels of hexadecanedioate, a fatty acid significantly associated with increased blood pressure levels and risk of incident heart failure	None
Kettunen 2016[S8]	123 blood lipids and metabolites	In-house platform	14 European cohorts (N=24925)	0%	General population	Eight novel loci for amino acids, pyruvate and fatty acids. *LPA* locus linked with cardiovascular risk	None
Yet 2016[S9]	43 serum metabolites named for the same compounds across Biocrates and Metabolon platforms	Biocrates and Metabolon	TwinsUK registry (N=1001)	0%	General Population	16 metabolites showed consistent genetic associations and appear to be robustly measured across metabolomic platforms	*FADS1* and 1-arachidonoyl-GPC (20:4n6)

Burkhardt 2015[S10]	26 amino acids, 36 acylcarnitines and 34 metabolite ratios in blood	In-house platform	LIFE Leipzig Heart Study (N=2107)	0%	General population / CAD patients	Six novel loci associated with blood levels of total acylcarnitine, arginine (both on chromosome 6; rs12210538, rs17657775), propionylcarnitine (chromosome 10; rs12779637), 2-hydroxyisovalerylcarnitine (chromosome 21; rs1571700), stearoylcarnitine (chromosome 1; rs3811444), and aspartic acid traits (chromosome 8; rs750472)	None
Kraus 2015[S11]	63 blood metabolites	In-house platform	CATHGEN study (N=3512)	19.1%	Cardiac catherization patients	The UPS arm of the ER stress pathway was implicated in CVD pathogenesis	None
Draisma 2015[S12]	129 serum metabolites	Biocrates Life Sciences AG, Austria	7 European cohorts (N discovery=7478, N replication=1182)	0%	General population	5 of 59 identified SNP-metabolite associations were novel, which were located in or near genes encoding metabolite transporter proteins or enzymes (*SLC22A16*, *ARG1*, *AGPS* and *ACSL1*), and demonstrated biomedical or pharmaceutical importance	None

Demirkan 2015[S13]	42 serum metabolites	In-house platform	ERF (N=2482)	0%	General population	Replicated three known loci in the metabolome wide significance: *CPS1* with glycine, *PRODH* with proline, *SLC16A9* with carnitine and a novel association between *DMGDH* and dimethyl-glycine. Three novel, suggestively significant loci: *TNP1* with pyruvate, *KCNJ16* with 3-hydroxybutyrate and 2p12 locus with valine	None
Ried 2014[S14]	151 serum metabolites using BIOCRATES AbsoluteIDQTM p150 kit and 193 serum metabolites measured with technique by Metabolon	Biocrates and Metabolon	KORA F4 (N=1814) and TwinsUK (N not described)	0%	General population	12 new genes were identified and validated using phenotype set enrichment analysis	None
Yu 2014[S15]	308 serum metabolites	Metabolon	ARIC African Americans (N=1260)	100%	Mean eGFR = 105 mL/min/1.73 m^2	14 genes mapped within 19 loci, including 12 enzyme-encoding genes (*KLKB1*, *SIAE*, *CPS1*, *NAT8*, *ACE*, *GATM*, *ACY3*, *ACSM2B*, *THEM4*, *ADH4*, *UGT1A*, *TREH*) and a polycystin protein gene (*PKD2L1*); 4 potential disease-associated paths including *NAT8*->N-acetylornithine->CKD, *TREH*->trehalose->diabetes	*UGT1A* and biliverdin, *UGT1A* and bilirubin (E,E), *ACY3* and N-acetylphenylalanine, *NAT8* and N-acetylornithine

Shin 2014[S16]	486 blood metabolites including 309 known and 177 unknown metabolites	Metabolon	KORA (N=1768) and TwinsUK (N=6054)	0%	General population	Genome-wide significant associations at 145 metabolic loci and their biochemical connectivity with more than 400 metabolites in human blood	*CYP4A11* and 10-undecenoate, *HAO2* and alpha-hydroxyisovalerate, *NAT8* and N-acetylornithine, *UGT1A1* and biliverdin
Rhee 2013[S17]	217 plasma metabolites	In-house platform	FHS (N=2076)	0%	General Population	31 genetic loci associated with plasma metabolites, including 23 that have not previously been reported. Results highlight the role of AGXT2 in cholesterol ester and triacylglycerol metabolism	*AGXT2* and β-aminoisobutyric acid
Hong 2013[S18]	6138 unique molecular features from UPLC–MS were retained for genetic association testing. The molecular identities of the most strongly replicating features for each locus were determined according to a defined workflow	In-house platform	CAPS study (N discovery=492, N replication=489)	0%	Prostate cancer patients	Seven loci (*PYROXD2*, *FADS1*, *PON1*, *CYP4F2*, *UGT1A8*, *ACADL*, and *LIPC*) with associated sequence variants contributing significantly to trait variance for one or more metabolites. Regional mQTL enrichment analyses implicated two loci that included *FADS1* and a novel locus near *PDGFC*. Biological pathway analysis implicated *ACADM*, *ACADS*, *ACAD8*, *ACAD10*, *ACAD11*, and *ACOXL*, reflecting significant enrichment of genes with	*UGT1A8* and bilirubin; *SLCO1B1* and glycochenodeeoxycholic acid 3-glucuronide

						acyl-CoA dehydrogenase activity	
Kettunen 2012[S19]	216 serum metabolic phenotypes (including metabolite levels and ratios)	In-house platform	5 Finnish population-based cohorts (N=8330)	0%	General population	31 loci, including 11 for which there have not been previous reports of associations to a metabolic trait or disorder. Analysis of twin pairs showed higher heritability of metabolic measures reported than conventional phenotypes	None
Tukiainen 2012[S20]	117 serum metabolites, 99 derived measures and 4 enzymatic lipid measures	In-house platform	5 Finnish population-based cohorts (N=8330)	0%	General population	New metabolic or genetic associations were identified for 30 of the 95 known lipid loci	None
Krumsiek 2012[S21]	225 unknown metabolites	Metabolon	German KORA F4 study (N=1768)	0%	General population	Overlaying inferred genetic associations, metabolic networks, and knowledge-based pathway information, testable hypotheses were formed for the biochemical identities of 106 unknown metabolites. Nine concrete predictions experimentally confirmed as proof of principle	*UGT1A* and bilirubin
Inouye 2012[S22]	130 metabolites and 11 metabolic networks derived using "agglomerative hierarchical clustering"	In-house platform	Young Finns Study (N=1905) and the NFBC66 (N=4703)	0%	General population	34 genomic loci at genome-wide significance, of which 7 are novel. In comparison to univariate tests, multivariate	None

						association analysis identified nearly twice as many significant associations in total	
Demirkan 2012[S23]	Plasma levels of 24 sphingomyelins, 9 ceramides, 57 phosphatidylcholines, 20 lysophosphatidylcholines, 27 phosphatidylethanolamines (PE), and 16 PE-based plasmalogens, as well as their proportions in each major class	In-house platform	5 European family-based studies (N=4034)	0%	General population	Four novel loci significantly associated with phospholipids (*PAQR9*, *AGPAT1*, *PKD2L1*, *PDXDC1*) and two with sphingolipids (*PLD2* and *APOE*). Associations observed between *FADS1-2-3* and carotid intima media thickness, *AGPAT1* and type 2 diabetes, and APOA1 and coronary artery disease	None
Suhre 2011[S24]	250 serum metabolites, >30,000 ratios	Metabolon	KORA F4 (N=1768) and TwinsUK (N=1052)	0%	General population	37 independent loci reached genome-wide significance in meta-analysis, 23 of which described new genetic associations with metabolic traits. 14 of the new associations were replicated and extended knowledge of known genetically determined metabotypes	*NAT8* and N-acetylornithine, *UGT1A* and bilirubin (E,E), *CYP4A* and 10-undecenoate (11:1n1)
Mittelstrass 2011[S25]	131 serum metabolites	Biocrates Life Sciences AG, Austria	KORA F3 and F4 (N=3300)	0%	General population	Genome-wide significant sex-specific differences SNPs in the *CPS1* locus for glycine	None

Nicholson 2011[S26]	526 plasma or urine metabolite peaks, 4 were driven largely by SNP variation, of which 3 were unambiguously identified and one was partially identified	In-house technique and Biocrates	MolTWIN (N=142 female twins), the MolOBB cohort, Oxford Biobank (N=69)	0%	General population	Replicated mQTL findings of previous study (Illig 2010) and discovered a previously uncharacterized yet substantial familial component of variation in metabolite levels in addition to the heritability contribution from the corresponding mQTL effects	None (*NAT8* was associated with urine N-acetylornithine, *AGXT2* with urine 3-amino-isobutyrate)
Illig 2010[S27]	163 serum metabolites and all 26,406 possible metabolite ratios	Biocrates Life Sciences AG, Austria	KORA (N discovery=1809), TwinsUK (N replication=422)	0%	General population	Eight out of nine replicated loci (*FADS1*, *ELOVL2*, *ACADS*, *ACADM*, *ACADL*, *SPTLC3*, *ETFDH* and *SLC16A9*), the genetic variant is located in or near genes encoding enzymes or solute carriers whose functions match the associating metabolic traits. The use of metabolite concentration ratios as proxies for enzymatic reaction rates reduced the variance and yielded robust statistical associations	None
Gieger 2008[S28]	363 serum metabolites	Biocrates Life Sciences AG, Austria	KORA S3 (N=284)	0%	General population	4 genetic variants in genes coding for enzymes (*FADS1*, *LIPC*, *SCAD*, *MCAD*) where the	None

						corresponding metabolic phenotype clearly matches the biochemical pathways in which these enzymes are active	

ARIC, Atherosclerosis Risk in Communities; CAPS, Cancer of the Prostate in Sweden; CATHGEN, Catherization Genetics; ERF, Erasmus Rucphen Family; FHS, Framingham Heart Study; KORA, Cooperative Health Research in the Augsburg Region; METSIM, Metabolic Syndrome in Men.

Table S6. Novel Index SNP-metabolite associations in AASK and replication in African Americans in the ARIC Study (N=818).

index SNP	Associated metabolites	Metabolite measured in the ARIC Study	Coded allele/reference allele in the ARIC Study	Coded allele frequency in the ARIC Study	Metabolite missingness in the ARIC Study	Beta coefficient in the ARIC Study (P-value)	Replicated?
rs79319225	3-methyladipate	Yes	A/G	0.03	145/818	-0.12 (0.43)	No
rs7528838	2-hydroxy-3-methylvalerate	Yes	T/A	0.27	0/818	-0.27 (5.97E-07)	Yes
rs13431529	N2-acetyllysine	Yes	C/G	0.52	232/818	0.62 (1.18E-45)	Yes
rs13409366	N-acetylkynurenine	No	NA	NA	NA	NA	NA
rs6546854	N-acetylleucine	Yes	G/A	0.54	8/818	0.57 (1.39E-34)	Yes
rs10168931	N-acetylarginine	Yes	A/G	0.54	0/818	0.68 (2.85E-49)	Yes
rs951389793	cysteinylglycine	Yes	T/TA	0.01	0/818	0.01 (0.97)	No
rs62184450	cysteinylglycine	Yes	C/A	0.03	0/818	-0.20 (0.18)	No
rs73835707	N-acetylmethionine	Yes	C/T	0.02	0/818	1.31 (7.12E-15)	Yes
rs73835707	N-acetylmethionine sulfoxide	No	NA	NA	NA	NA	NA
rs1973612	histidylalanine	Yes	T/C	0.59	738/818	0.04 (0.16)	No
rs60132035	sphingomyelin (d18:1/20:1, d18:2/20:0)	Yes	A/G	0.01	0/818	-0.23 (0.27)	No
rs139981949	glycohyocholate	Yes	T/C	0.01	261/818	-0.17 (0.38)	No
rs76541360	1-arachidonoyl-GPI	Yes	T/A	0.07	0/818	0.16 (0.10)	No
rs59095288	6-oxopiperidine-2-carboxylate	Yes	C/T	0.34	7/818	0.50 (3.37E-24)	Yes
rs57294583	2'-O-methylcytidine	Yes	G/A	0.53	9/818	0.80 (6.88E-69)	Yes
rs142514677	N-acetyl-aspartyl-glutamate	Yes	C/CAG	0.04	124/818	-0.67 (1.57E-09)	Yes
rs150556827	N-acetyl-aspartyl-glutamate	Yes	T/C	0.01	124/818	-0.48 (0.04)	No
rs115780269	N-acetyltryptophan	Yes	C/T	0.02	40/818	1.23 (1.72E-10)	Yes
rs115780269	N-acetylkynurenine	No	NA	NA	NA	NA	NA
rs115780269	N-acetyltyrosine	Yes	T/C	0.02	33/818	1.24 (2.79E-10)	Yes
rs114080902	N-formylanthranilic acid	Yes	T/C	0.04	0/818	0.89 (9.13E-15)	Yes

Table S7. Most likely causal genes for novel variant-metabolite associations identified in AASK and replicated in the ARIC Study and whether genes are enzyme encoding.

Gene	Enzyme encoding?
HAO2	Yes
NAT8	Yes
ACY1	Yes
EXOSC4	Yes
GPAA1	No
PHYHD1	Yes
ACY3	Yes
AFMID	Yes

Table S8. Interactions between index single nucleotide polymorphisms and glomerular filtration rate.

Distinct locus number	Genomic locus number; chromosomal location; mapped genes	Index SNP	Index SNP coded allele/reference allele	Index SNP coded allele frequency	Associated metabolites	Beta coefficient for index SNP (P-value)*	Beta coefficient for measured GFR (P-value)	Beta coefficien for interaction term between index SNP and GFR (P-value)*
3	**1:118887090-119887090; *HAO2***	**rs7528838**	**T/A**	**0.252**	**2-hydroxy-3-methylvalerate**	**-0.238 (6.81E-13)**	**0.001 (0.515)**	**0.007 (0.006)**
4	2:73105659-74173773; ALMS1, ALMS1P, NAT8, TPRKB	rs13431529	C/G	0.535	N2-acetyllysine	-0.198 (4.94E-08)	-0.018 (1.33E-15)	-0.004 (0.165)
		rs6546854	G/A	0.547	N-acetylleucine	-0.058 (0.001)	-0.008 (1.52E-15)	0.001 (0.553)
		rs10168931	A/G	0.540	N-acetylarginine	-0.125 (9.06E-08)	-0.013 (2.17E-18)	-0.001 (0.583)
8	3:51489004-52620872; *POC1A, ACY1, ABHD14A-ACY1*	rs73835707	C/T	0.021	N-acetylmethionine	1.096 (8.20E-12)	-0.019 (9.06E-13)	0.011 (0.495)
15	8:143564265-144564265; *OPLAH, CTD-3065J16.9, EXOSC4, GPAA1, KIAA1875, MAF1*	rs59095288	C/T	0.309	6-oxopiperidine-2-carboxylate	0.311 (5.73E-23)	-0.006 (2.73E-04)	-0.004 (0.090)
16	9:128422557-129422736; *LRRC8A, PHYHD1, SH3GLB2*	rs57294583	G/A	0.547	2'-O-methylcytidine	0.007 (0.809)	0.005 (0.003)	-0.003 (0.162)
17	11:48528971-49528971; *PTPRJ, FOLH1*	rs142514677	C/CAG	0.059	N-acetyl-aspartyl-glutamate (NAAG)	-0.391 (6.03E-12)	-0.005 (1.90E-04)	0.006 (0.172)
20	11:67144985-68144985;	rs115780269	T/C	0.025	N-acetyltryptophan	1.56 (2.09E-34)	-0.011 (3.61E-07)	0 (0.963)
					N-acetyltyrosine	1.165 (1.21E-23)	-0.014 (7.27E-12)	0.016 (0.089)

	ACY3, AP003385.2							
23	17:77706736-78706736; *AFMID*	rs114080902	T/C	0.043	N-formylanthranilic acid	0.671 (8.18E-22)	-0.01 (1.50E-10)	0 (0.969)

* **Bold font** indicates nominally significantly stronger ($P<0.05$) index-SNP association with lower measured GFR.

Beta coefficient for index SNP.

Table S9. Metabolites with replicated novel associations and correlation coefficients with GFR in AASK.

Associated metabolites	Spearman correlation coefficient
2-hydroxy-3-methylvalerate	0.03
N2-acetyllysine	-0.28
N-acetylleucine	-0.35
N-acetylarginine	-0.34
N-acetylmethionine	-0.21
6-oxopiperidine-2-carboxylate	-0.12
2'-O-methylcytidine	0.08
N-acetyl-aspartyl-glutamate (NAAG)	-0.08
N-acetyltryptophan	-0.16
N-acetyltyrosine	-0.18
N-formylanthranilic acid	-0.23

Table S10. Metabolite missingness and index SNP-metabolite associations after excluding missing values, re-quantifying metabolite level as an ordinary variable (1=missing, 2=below median, 3=equal to or above median), or re-imputing missing values using missing data principal components and a k-nearest neighbor method for metabolites detected in <90% of samples in AASK.

Distinct variant-metabolite association	index SNP	Associated metabolites	Metabolite missingness in AASK	Beta coefficient with missing values imputed with lowest detectable value (P-value)	Beta coefficient after excluding missing values (P-value)	Beta coefficient after re-quantifying metabolite level as ordinal variable (P-value)	Beta coefficient after imputing missing values using missing data principal components and a 25-nearest neighbor method (P-value)
1	rs4507958	10-undecenoate (11:1n1)	0/619	NA*	NA	NA	NA
2	rs324418	N-palmitoylglycine	26/619	NA	NA	NA	NA
3	rs79319225	3-methyladipate	438/619	**0.75 (1.29E-10)**	**0.76 (8.10E-05)**	**0.76 (8.10E-05)**	**0.54 (5.19E-04)**
4	rs7528838	alpha-hydroxyisovalerate	0/619	NA	NA	NA	NA
5	rs7528838	2-hydroxy-3-methylvalerate	1/619	NA	NA	NA	NA
6	rs7587577	N-acetylasparagine	0/619	NA	NA	NA	NA
7	rs7587577	N-acetyl-3-methylhistidine	101/619	**0.39 (6.11E-13)**	**0.35 (1.09E-07)**	**0.16 (6.59E-07)**	**0.54 (1.92E-22)**
8	rs11126412	N-acetyl-1-methylhistidine	10/619	NA	NA	NA	NA
9	rs11126412	N-acetylphenylalanine	0/619	NA	NA	NA	NA
10	rs13431529	N2-acetyllysine	56/619	NA	NA	NA	NA
11	rs13409366	N-acetylcitrulline	1/619	NA	NA	NA	NA
12	rs13409366	N-acetylglutamine	0/619	NA	NA	NA	NA
13	rs13409366	N-acetylkynurenine	177/619	**0.33 (7.26E-37)**	**0.39 (5.29E-09)**	**0.15 (1.03E-05)**	**0.32 (9.27E-09)**
14	rs6546854	N-acetylleucine	1/619	NA	NA	NA	NA
15	rs10168931	N-acetylarginine	1/619	NA	NA	NA	NA
16	rs10206899	N-delta-acetylornithine	0/619	NA	NA	NA	NA
17	rs951389793	cysteinylglycine	468/619	**1.12 (1.42E-10)**	**0.68 (0.02)**	**0.11 (0.23)**	**0.90 (2.2E-04)**
18	rs1976391	biliverdin	2/619	NA	NA	NA	NA

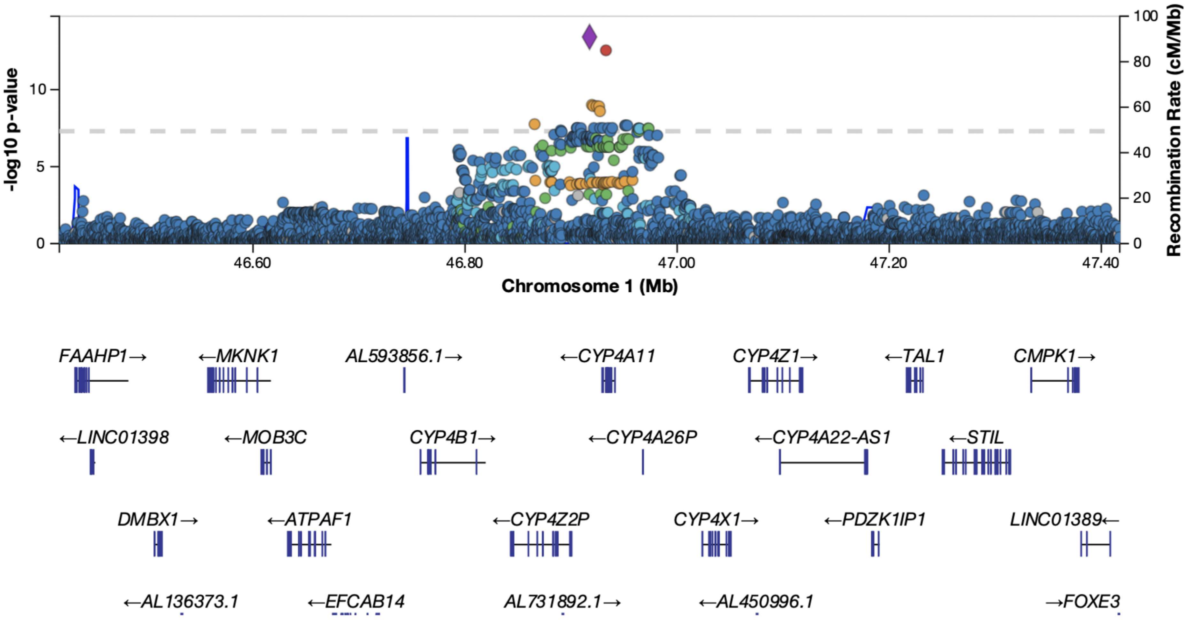
19	rs4148325	bilirubin (E,E)	4/619	NA	NA	NA	NA
20	rs62184450	cysteinylglycine	468/619	**0.77 (1.80E-10)**	0.44 (0.03)	0.10 (0.36)	**0.44 (0.009)**
21	rs73835707	N-acetylmethionine	18/619	NA	NA	NA	NA
22	rs73835707	N-acetylmethionine sulfoxide	252/619	**1.24 (6.06E-13)**	**0.99 (7.38E-07)**	**0.36 (3.89E-04)**	**1.08 (2.53E-08)**
23	rs2229152	N-acetylalanine	0/619	NA	NA	NA	NA
24	rs1973612	histidylalanine	3/619	NA	NA	NA	NA
25	rs344514	3-aminoisobutyrate	0/619	NA	NA	NA	NA
26	rs60132035	sphingomyelin (d18:1/20:1, d18:2/20:0)	0/619	NA	NA	NA	NA
27	rs7780766	N-acetylhistidine	3/619	NA	NA	NA	NA
28	rs139981949	glycohyocholate	112/619	**1.21 (9.15E-12)**	**1.09 (2.52E-08)**	**0.41 (5.31E-05)**	**1.19 (1.38E-09)**
29	rs76541360	1-arachidonoyl-GPI	0/619	NA	NA	NA	NA
30	rs59095288	6-oxopiperidine-2-carboxylate	18/619	NA	NA	NA	NA
31	rs55758160	2'-O-methyluridine	12/619	NA	NA	NA	NA
32	rs57294583	2'-O-methylcytidine	22/619	NA	NA	NA	NA
33	rs142514677	N-acetyl-aspartyl-glutamate	49/619	NA	NA	NA	NA
34	rs150556827	N-acetyl-aspartyl-glutamate	49/619	NA	NA	NA	NA
35	rs174564	1-arachidonoyl-GPC (20:4n6)	0/619	NA	NA	NA	NA
36	rs115780269	N-acetyltryptophan	45/619	NA	NA	NA	NA
37	rs115780269	N-acetylkynurenine	177/619	**1.07 (2.18E-12)**	**1.08 (2.13E-09)**	**0.37 (9.45E-05)**	**1.16 (1.60E-11)**
38	rs115780269	N-acetyltyrosine	76/619	**1.42 (4.15E-17)**	**1.48 (2.40E-16)**	**0.38 (4.74E-05)**	**1.48 (5.43E-17)**
39	rs115780269	N-acetylphenylalanine	0/619	NA	NA	NA	NA
40	rs114419265	glycochenodeoxycholate glucuronide	54/619	NA	NA	NA	NA
41	rs34708625	ethylmalonate	0/619	NA	NA	NA	NA
42	rs114080902	N-formylanthranilic acid	5/619	NA	NA	NA	NA

* NA, sensitivity analysis not conducted for the index SNP-metabolite association because of low level of missingness (<10%) of the metabolite.

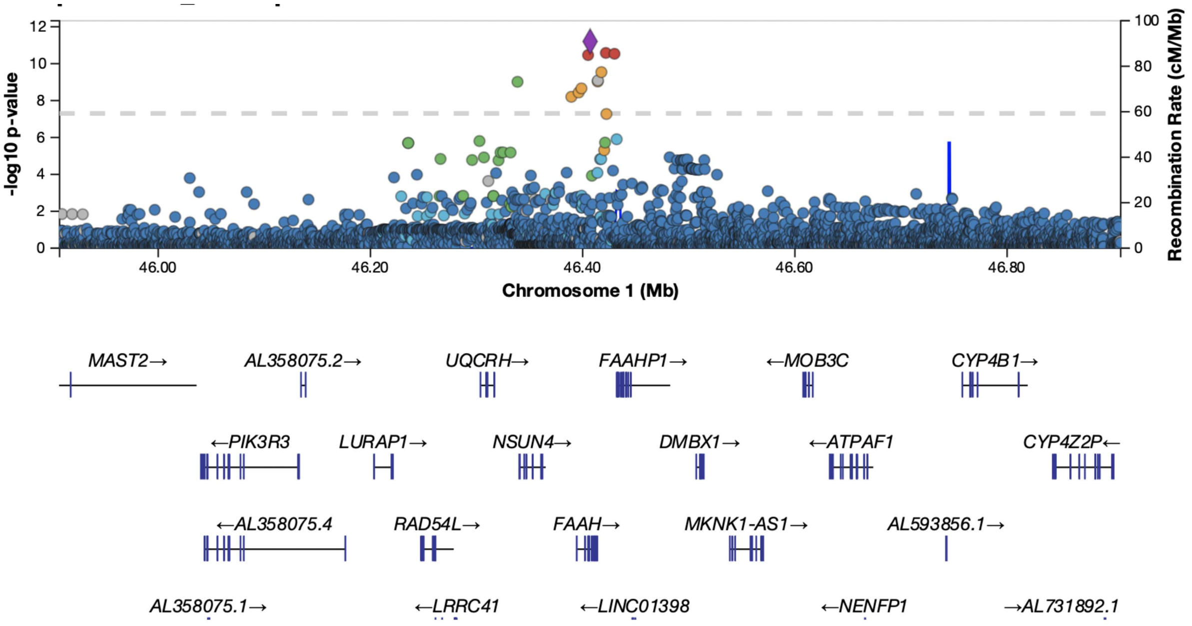
Bold font indicates statistical significance ($P<0.05/9$).

Figure S1. Regional association plots for locus-metabolite associations (N=42).

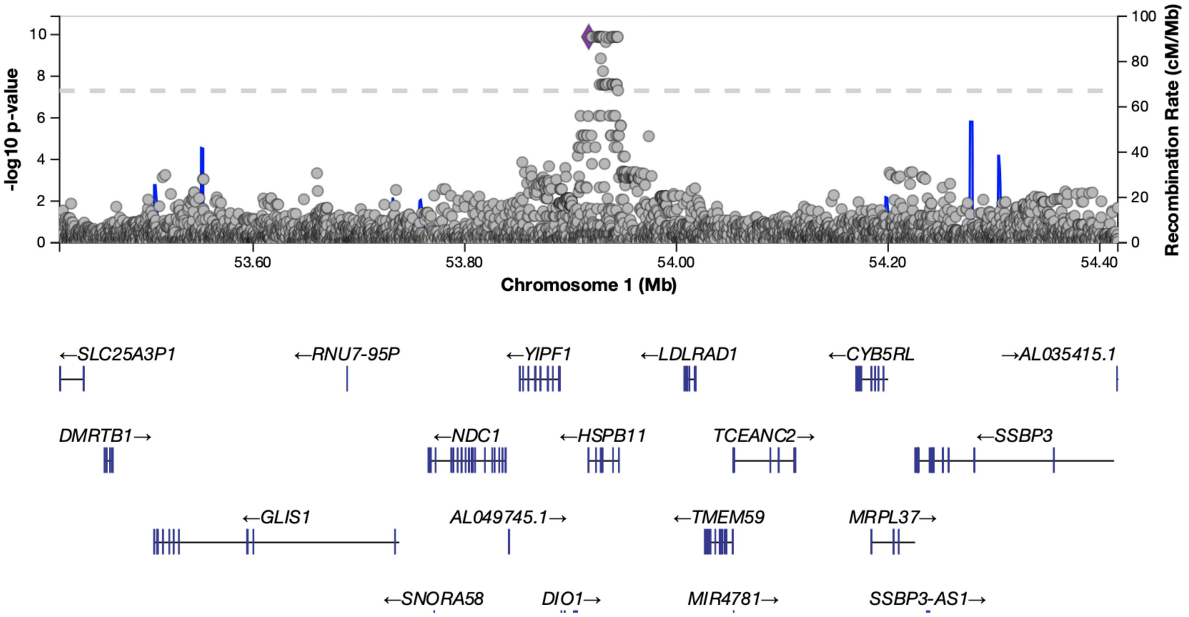
a. Locus 1, rs4507958 and 10-undecenoate (11:1n1).



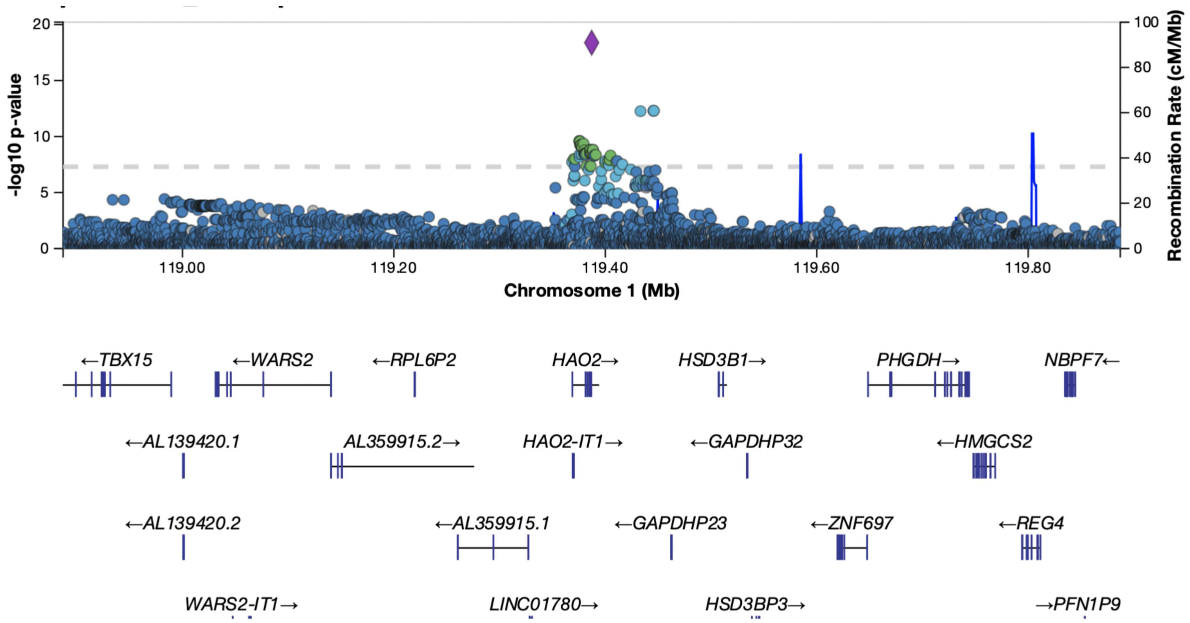
b. Locus 1, rs324418 and N-palmitoylglycine.



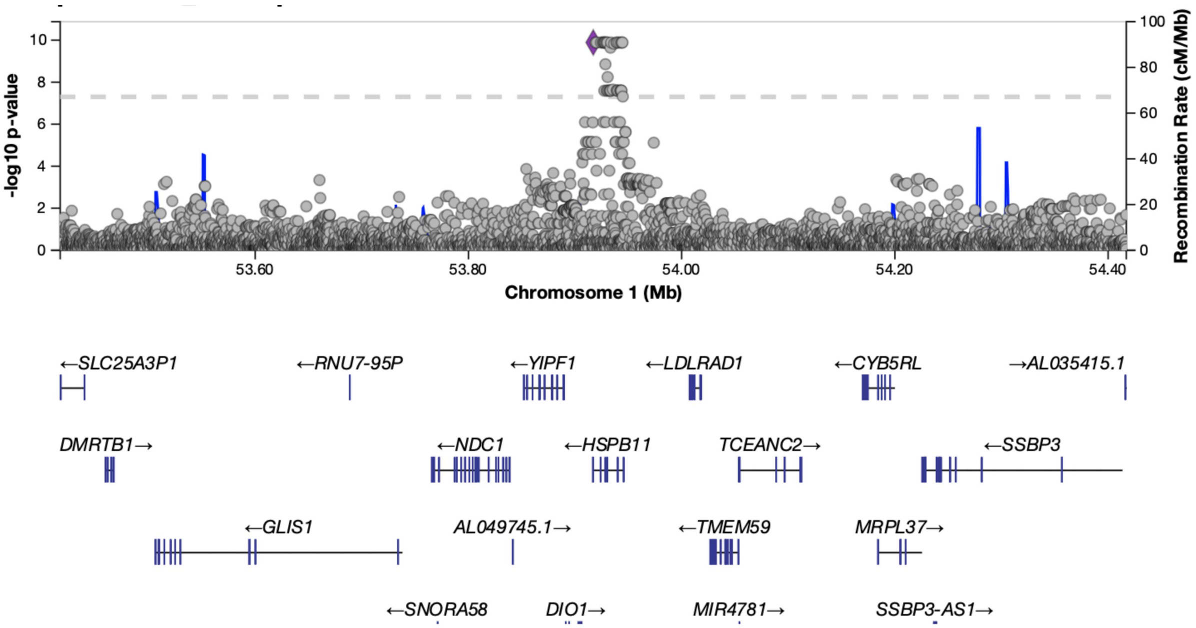
c. Locus 2, rs79319225 and 3-methyladipate.



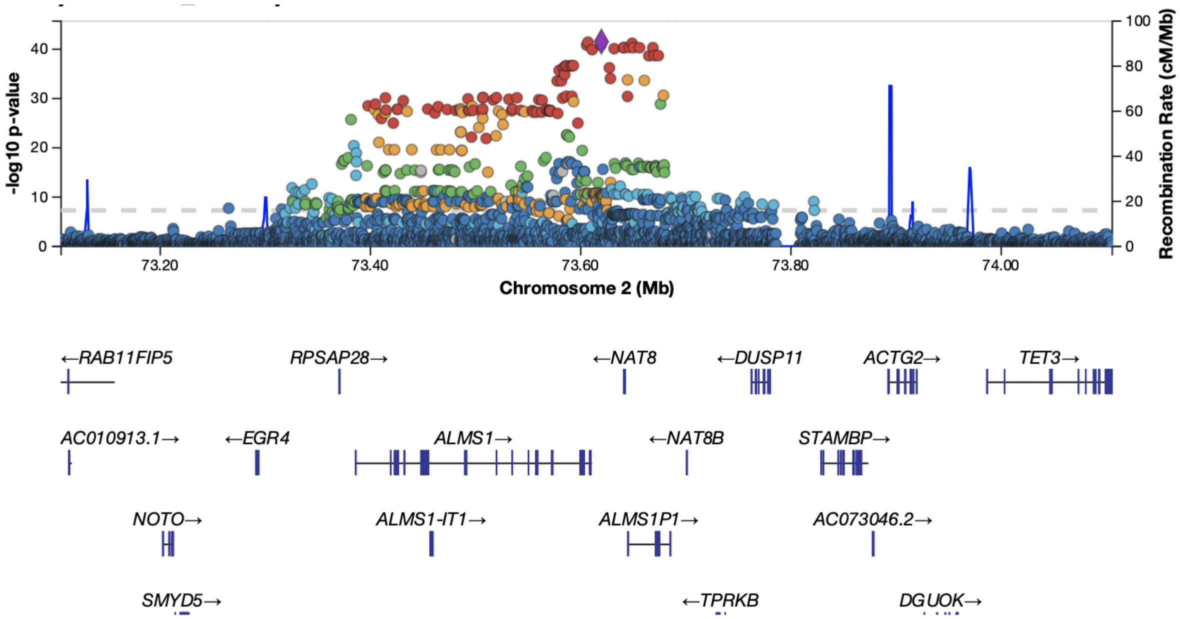
d. Locus 3, rs7528838 and alpha-hydroxyisovalerate.



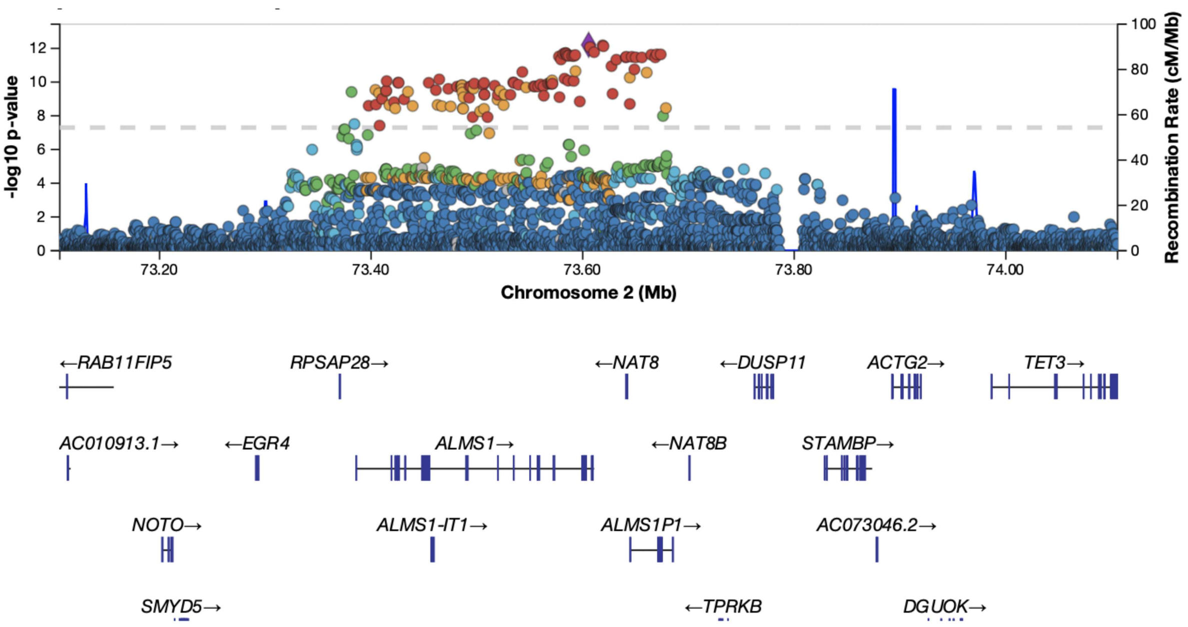
e. Locus 3, rs7528838 and 2-hydroxy-3-methylvalerate.



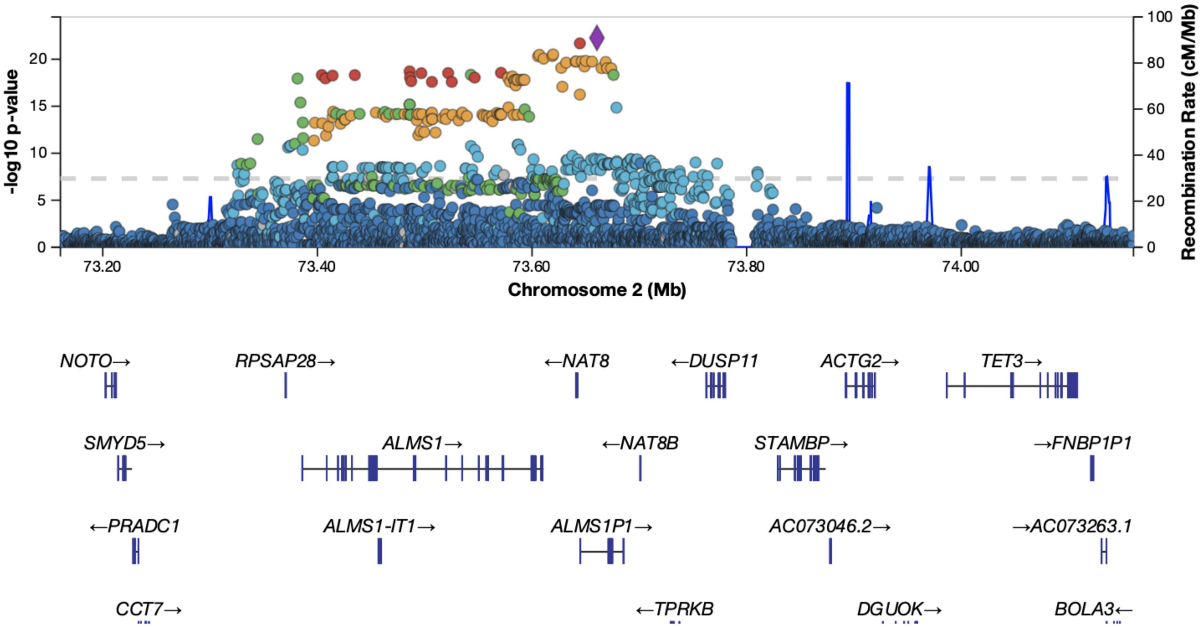
f. Locus 4, rs7587577 and N-acetylasparagine.



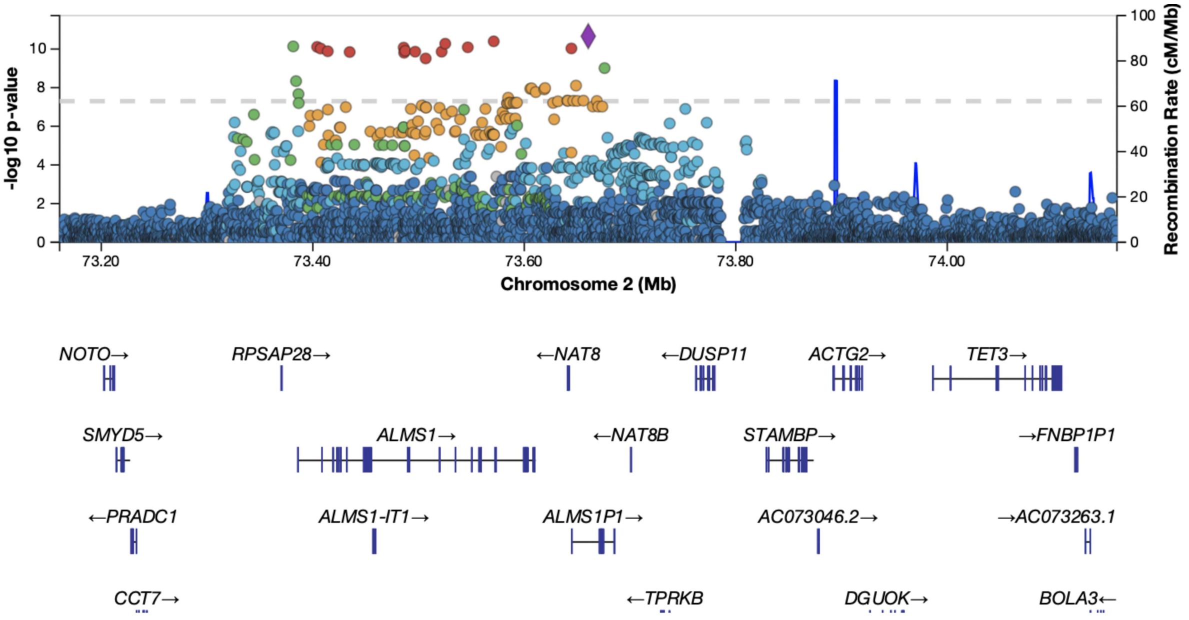
g. Locus 4, rs7587577 and N-acetyl-3-methylhistidine.



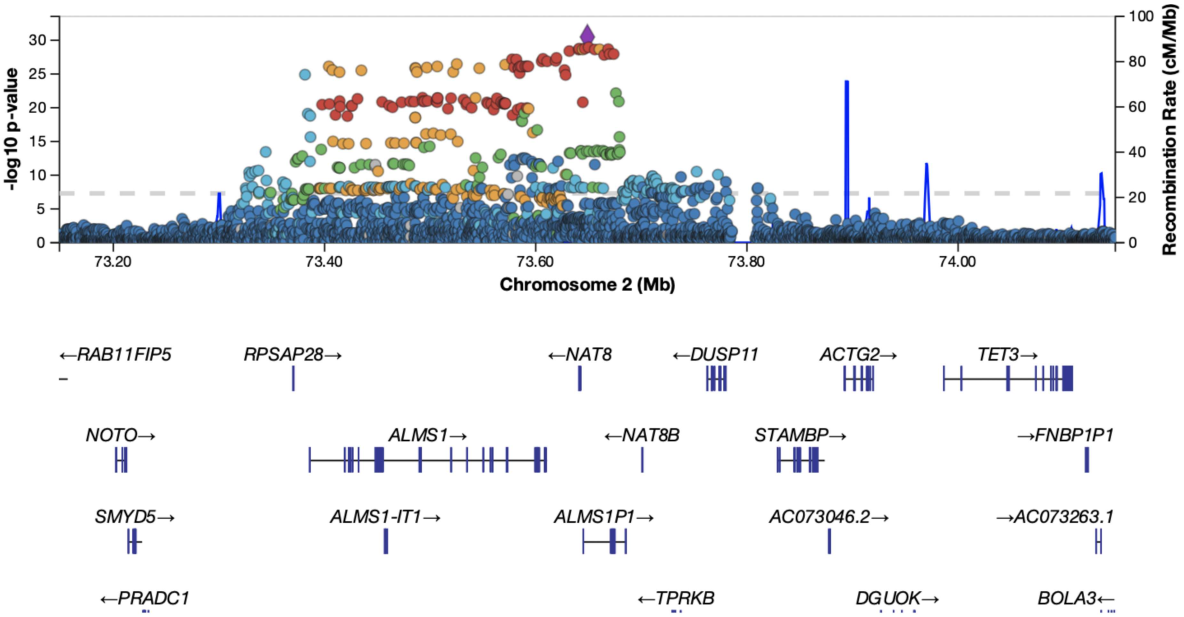
h. Locus 4, rs11126412 and N-acetyl-1-methylhistidine.



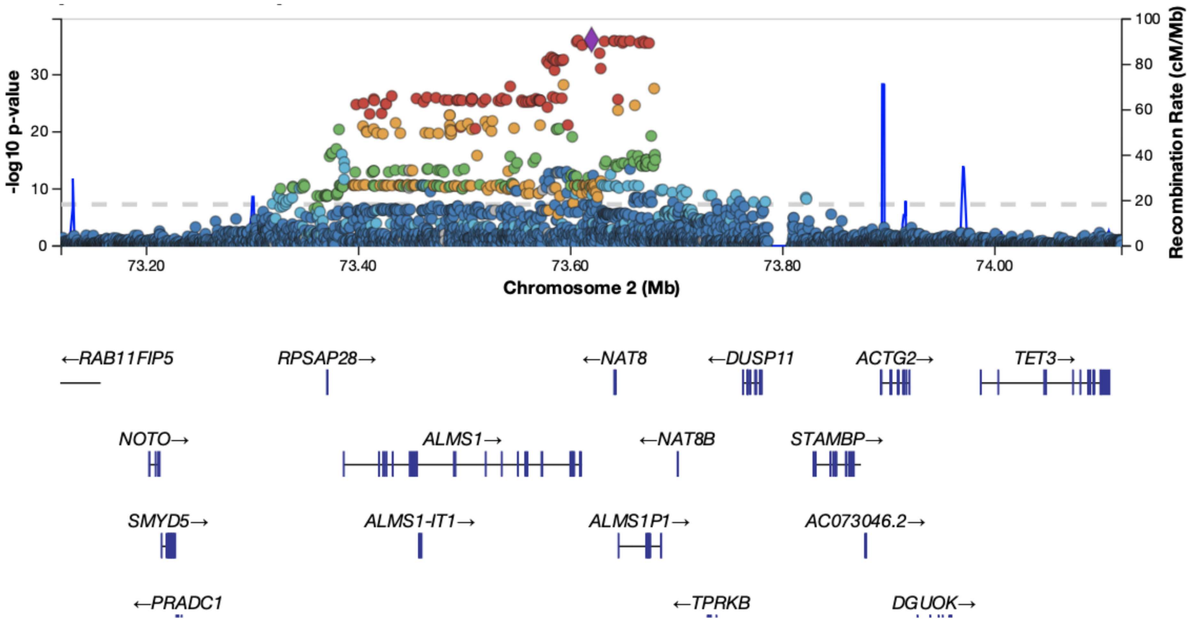
i. Locus 4, rs11126412 and N-acetylphenylalanine.



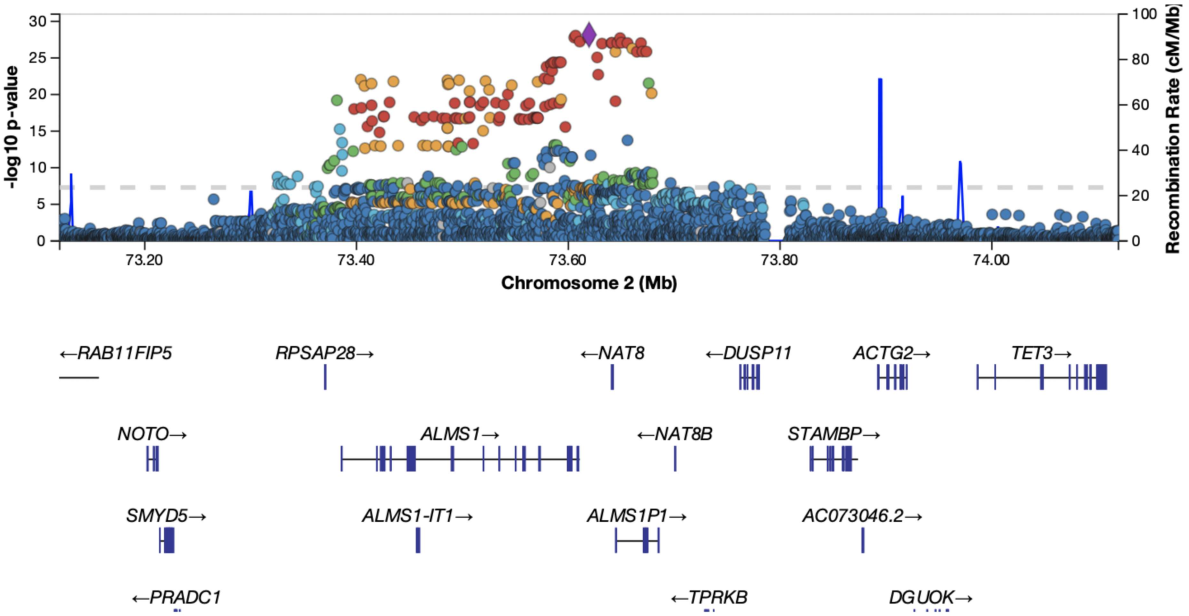
j. Locus 4, rs13431529 and N2-acetyllysine.



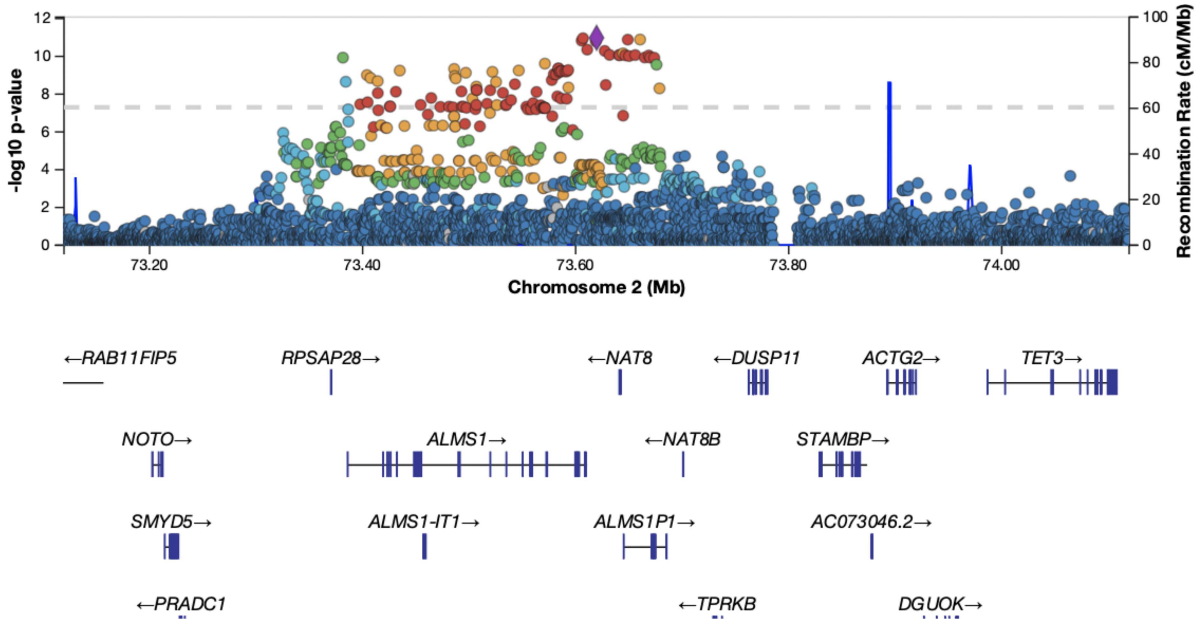
k. Locus 4, rs13409366 and N-acetylcitrulline.



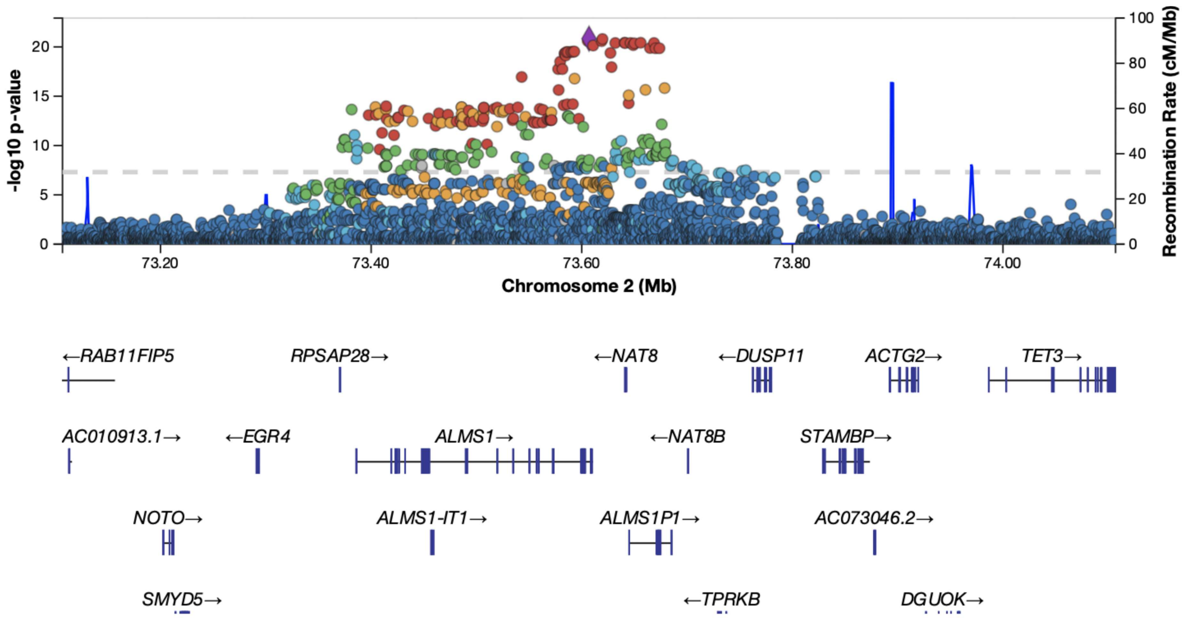
l. Locus 4, rs13409366 and N-acetylglutamine.



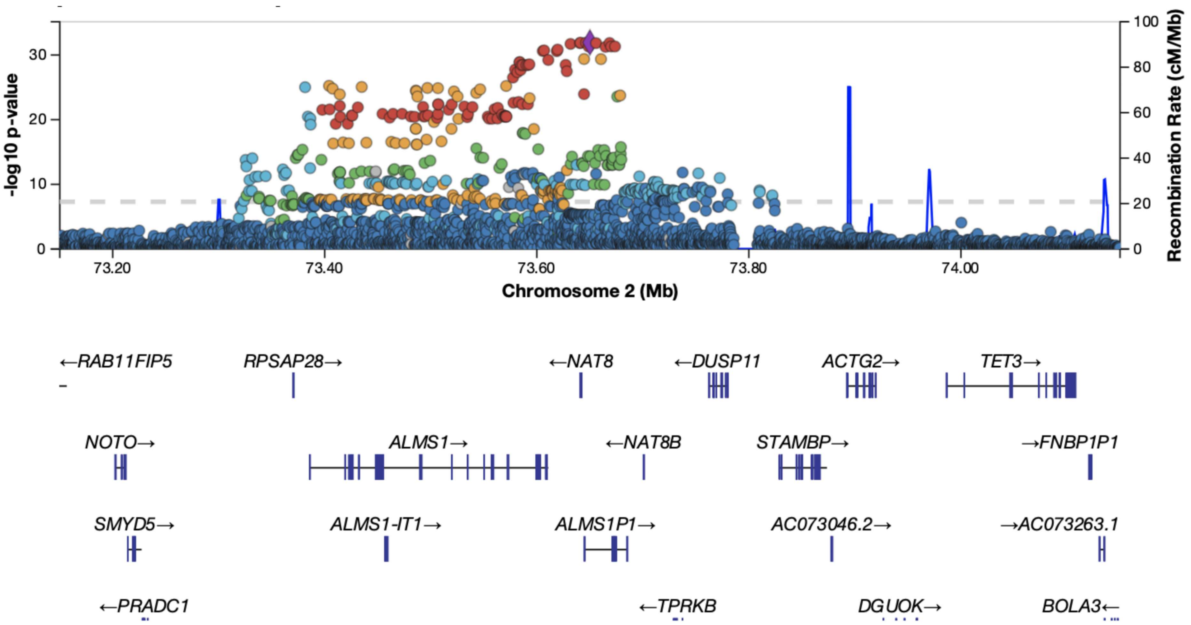
m. Locus 4, rs13409366 and N-acetylkynurenine.



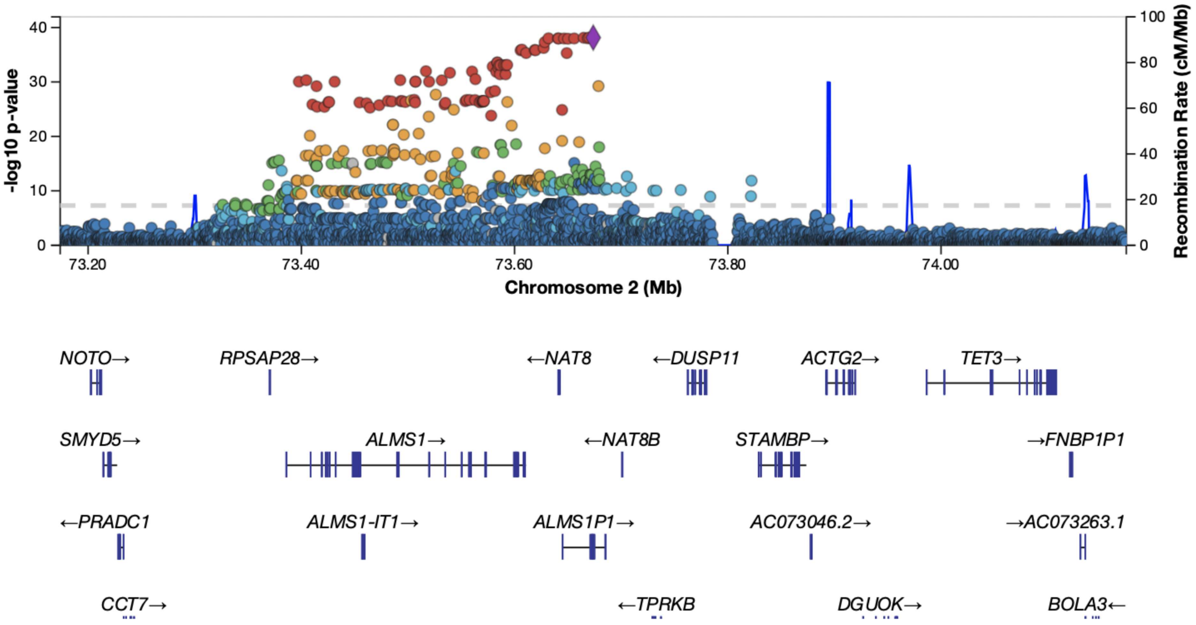
n. Locus 4, rs6546854 and N-acetylleucine.



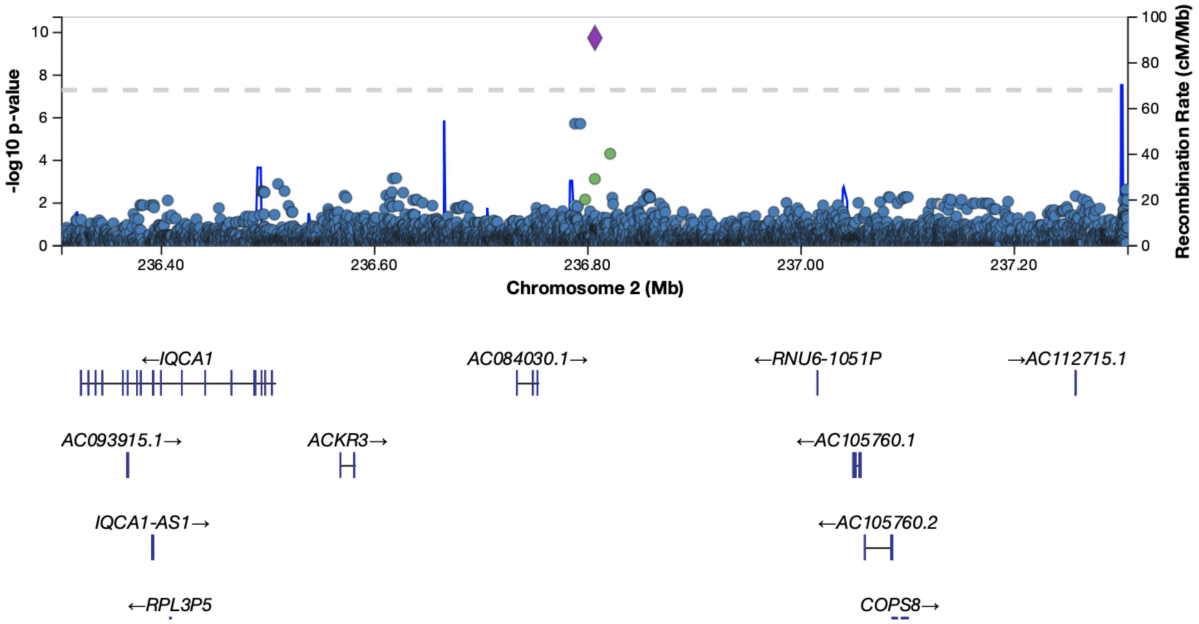
o. Locus 4, rs10168931 and N-acetylarginine.



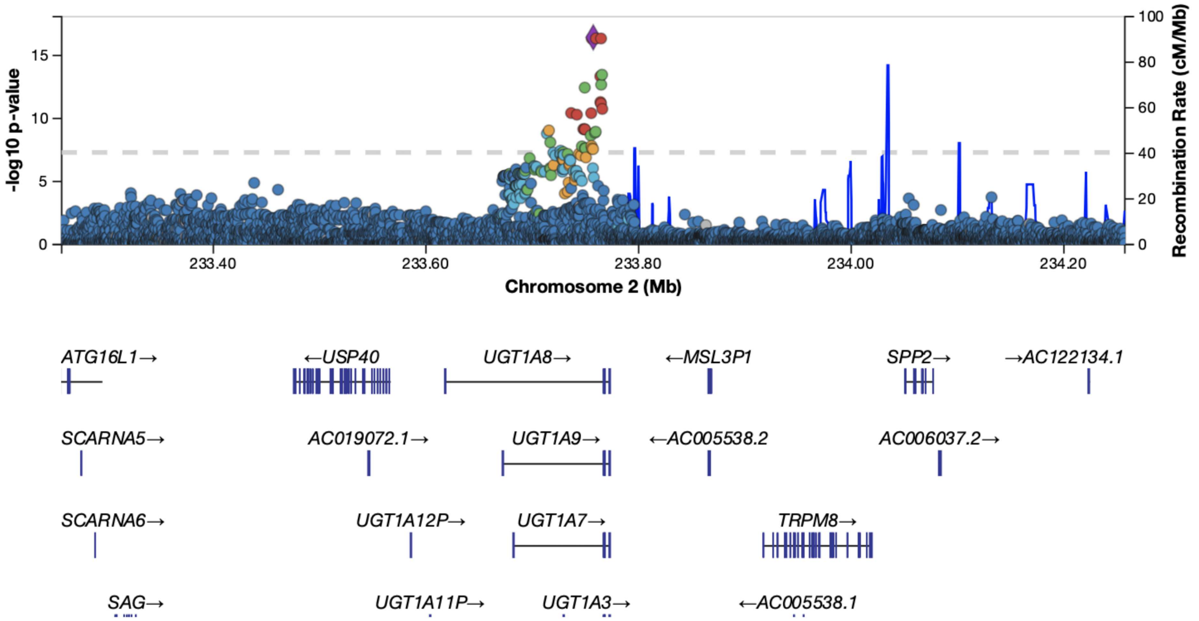
p. Locus 4, rs10206899 and N-delta-acetylornithine.



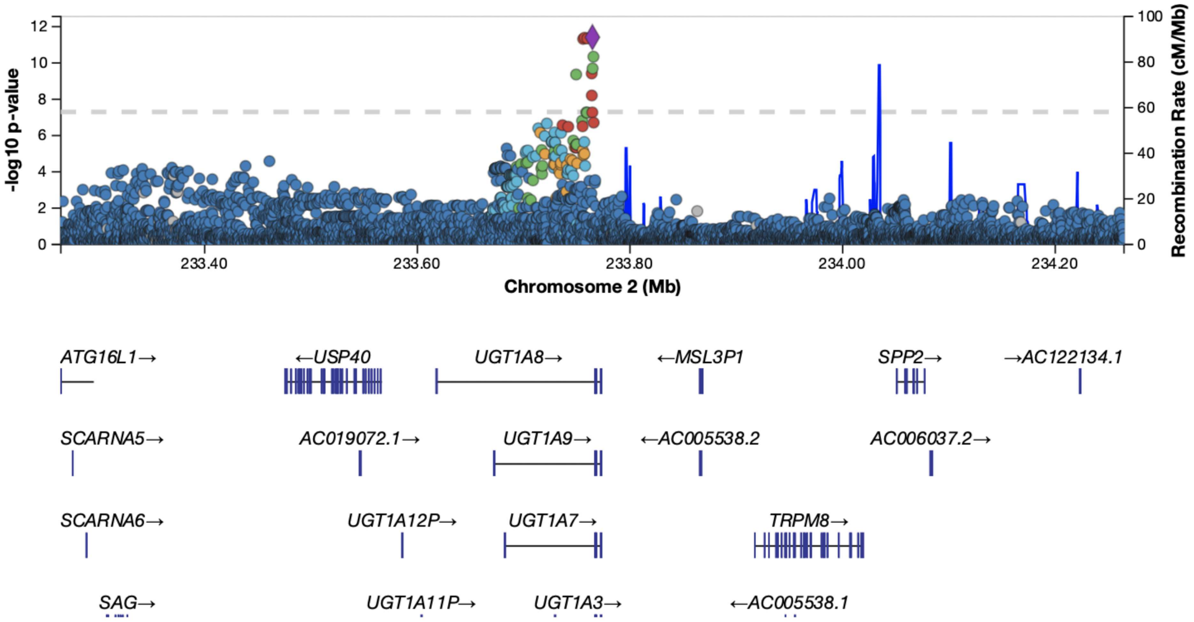
q. Locus 5, rs951389793 and cysteinylglycine.



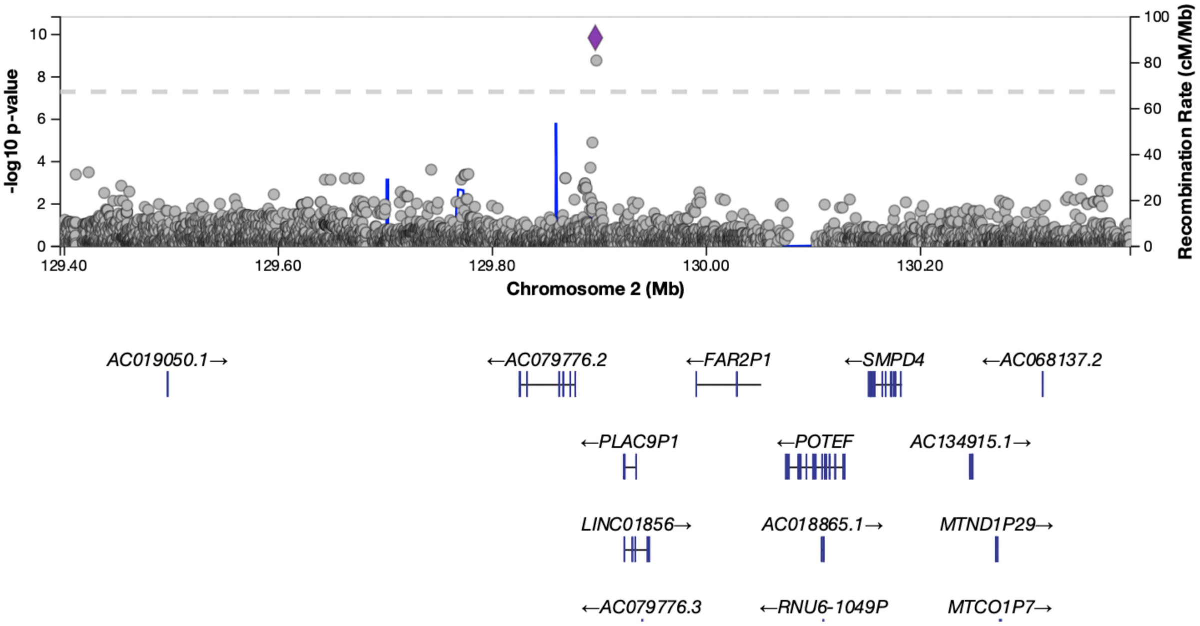
r. Locus 6, rs1976391 and biliverdin.



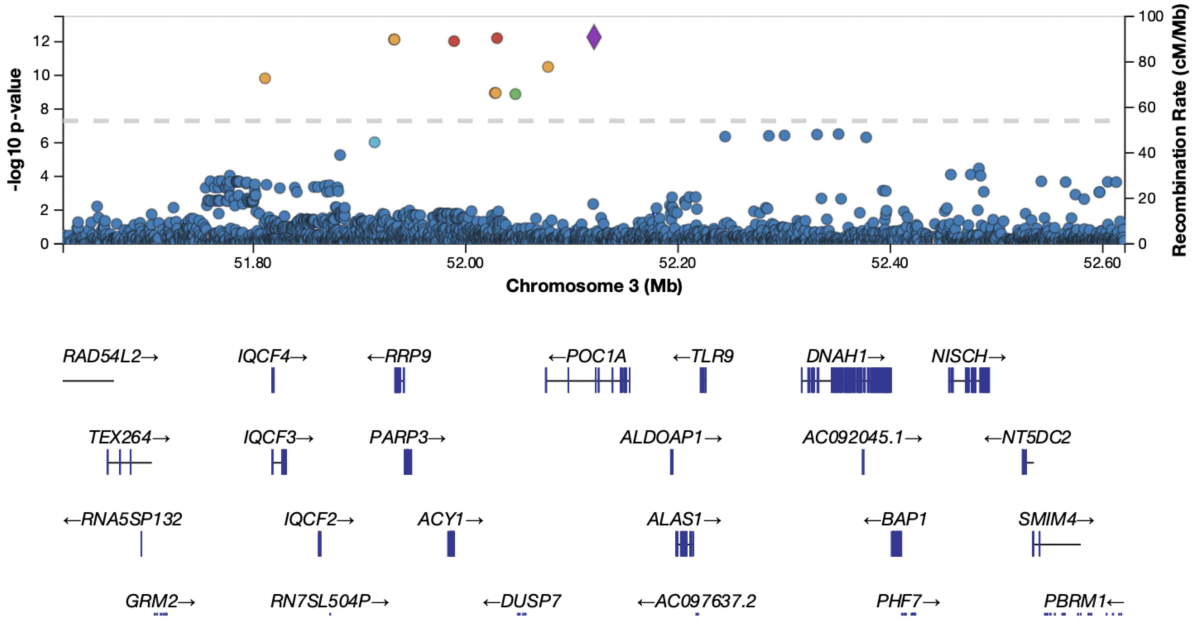
s. Locus 6, rs4148325 and bilirubin (E,E).



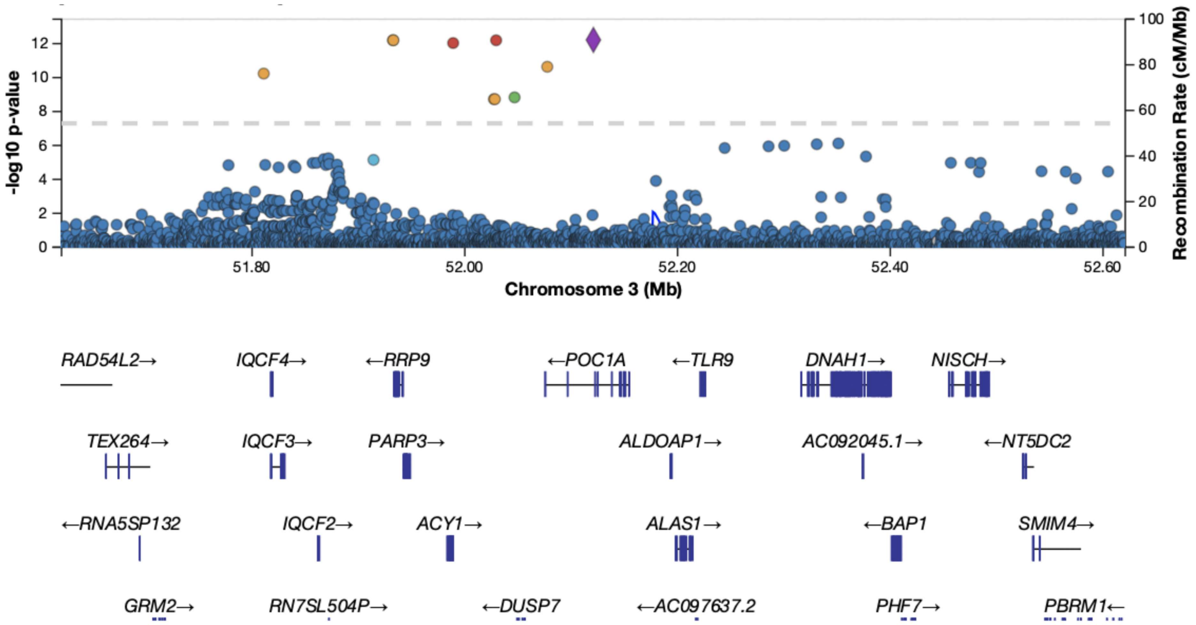
t. Locus 7, rs62184450 and cysteinylglycine.



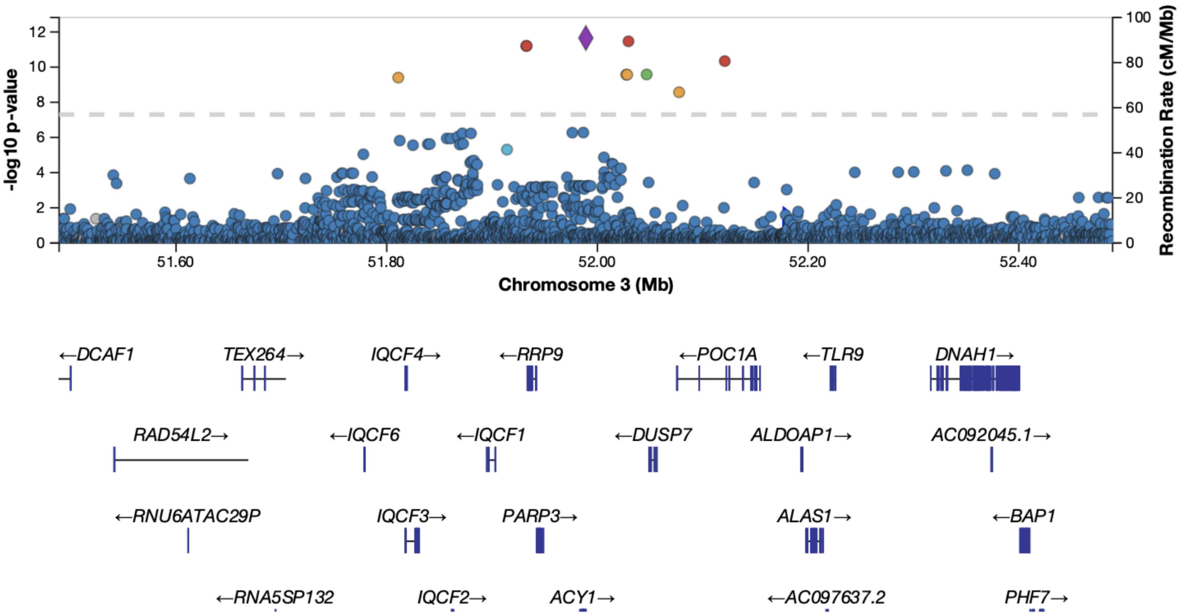
u. Locus 8, rs73835707 and N-acetylmethionine.



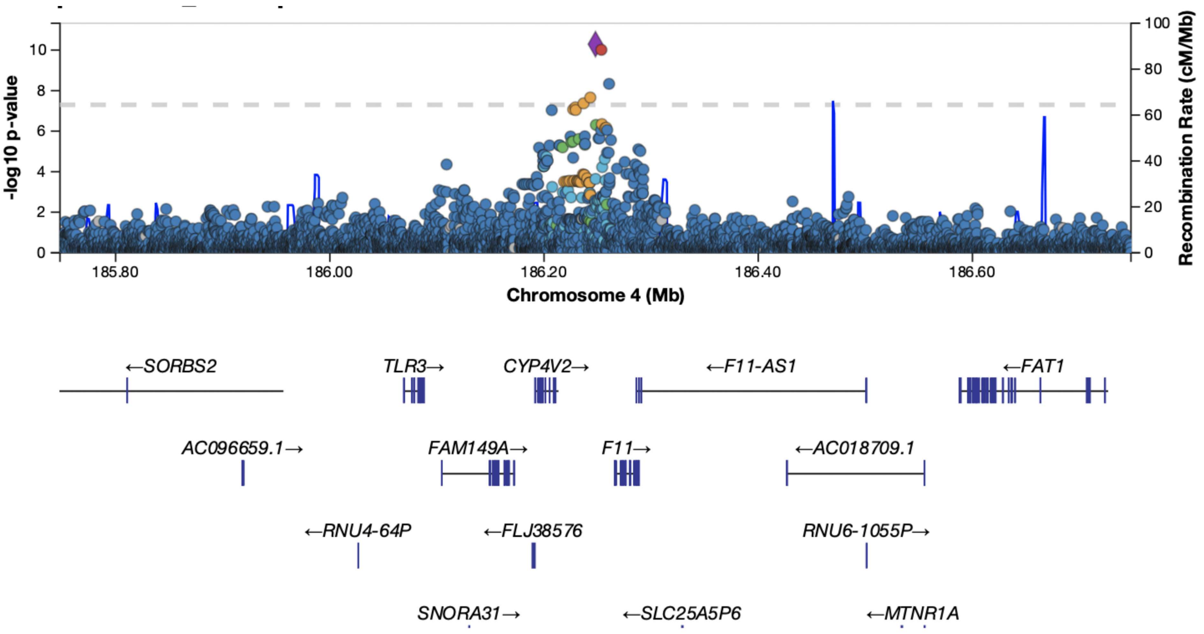
v. Locus 8, rs73835707 and N-acetylmethionine sulfoxide.



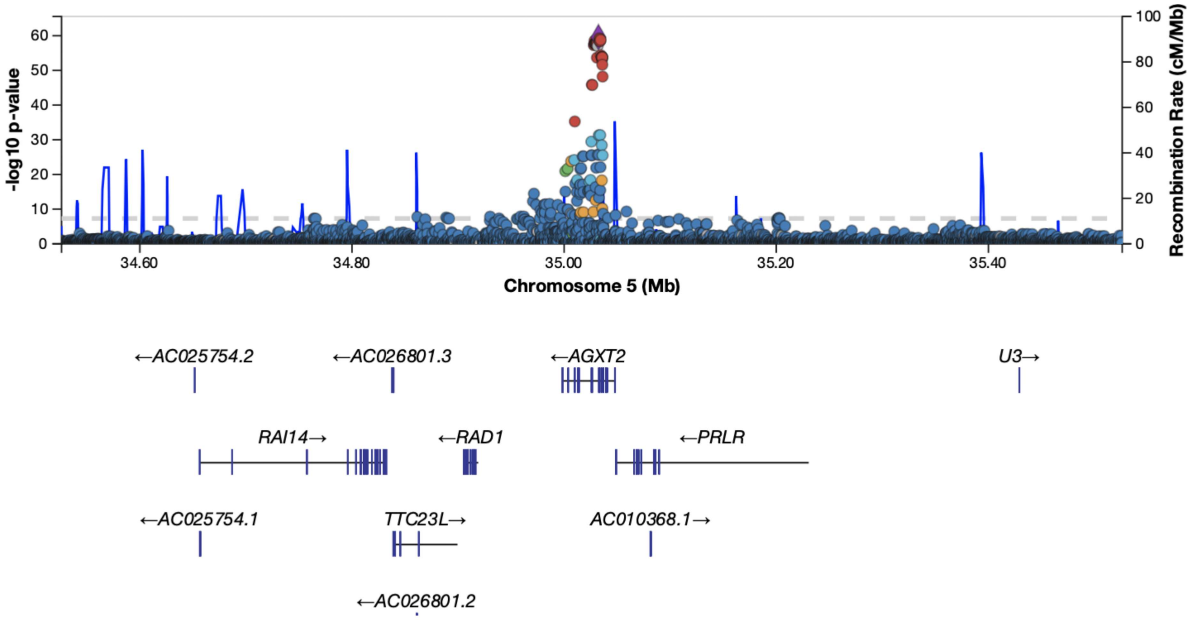
w. Locus 8, rs2229152 and N-acetylalanine.



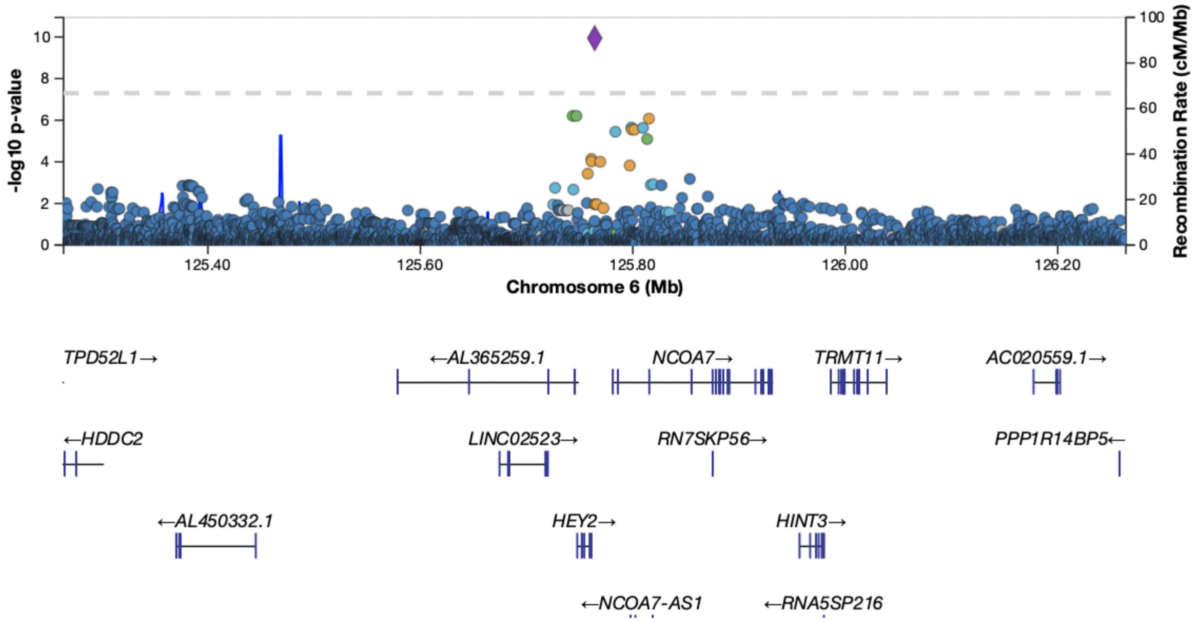
x. Locus 9, rs1973612 and histidylalanine.



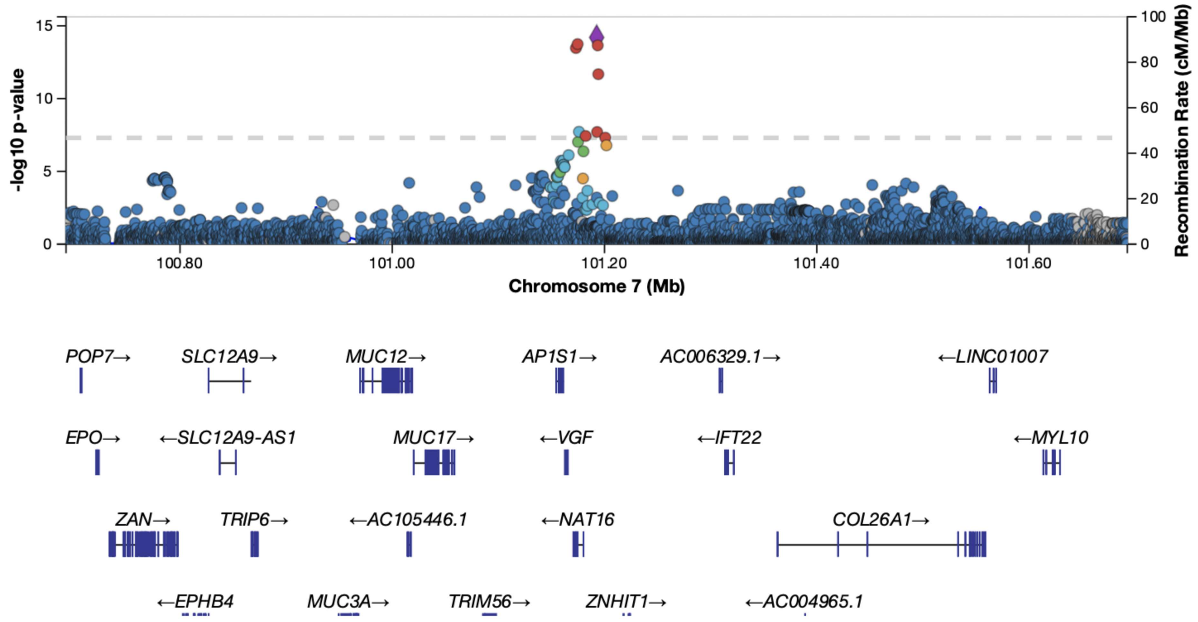
y. Locus 10, rs344514 and 3-aminoisobutyrate.



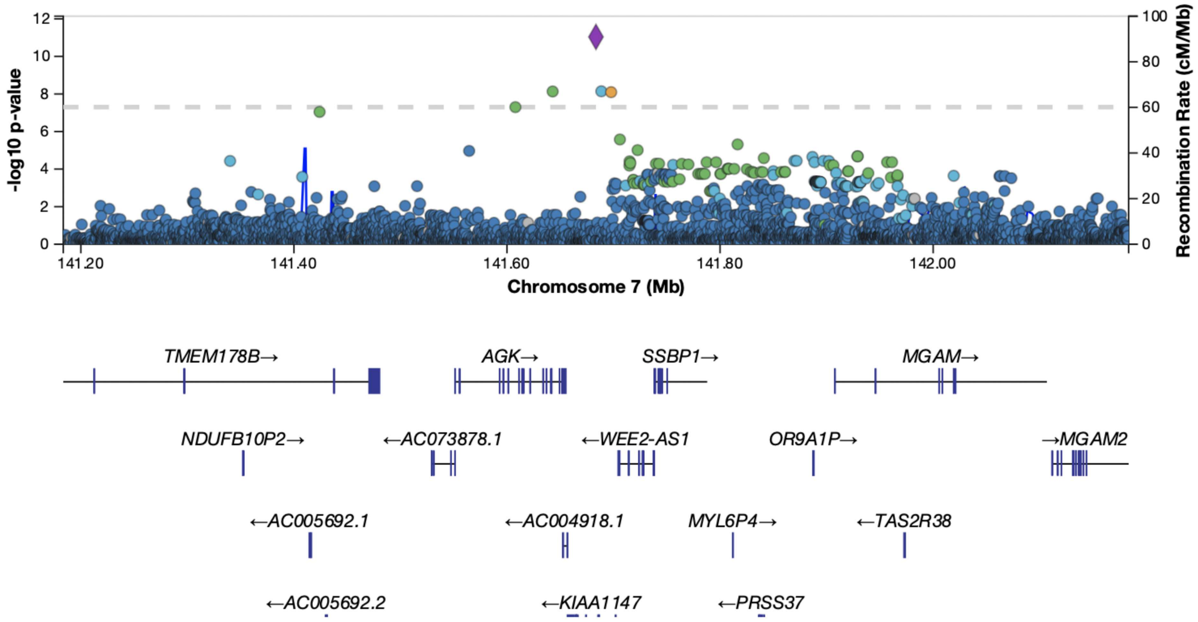
z. Locus 11, rs60132035 and sphingomyelin (d18:1/20:1, d18:2/20:0).



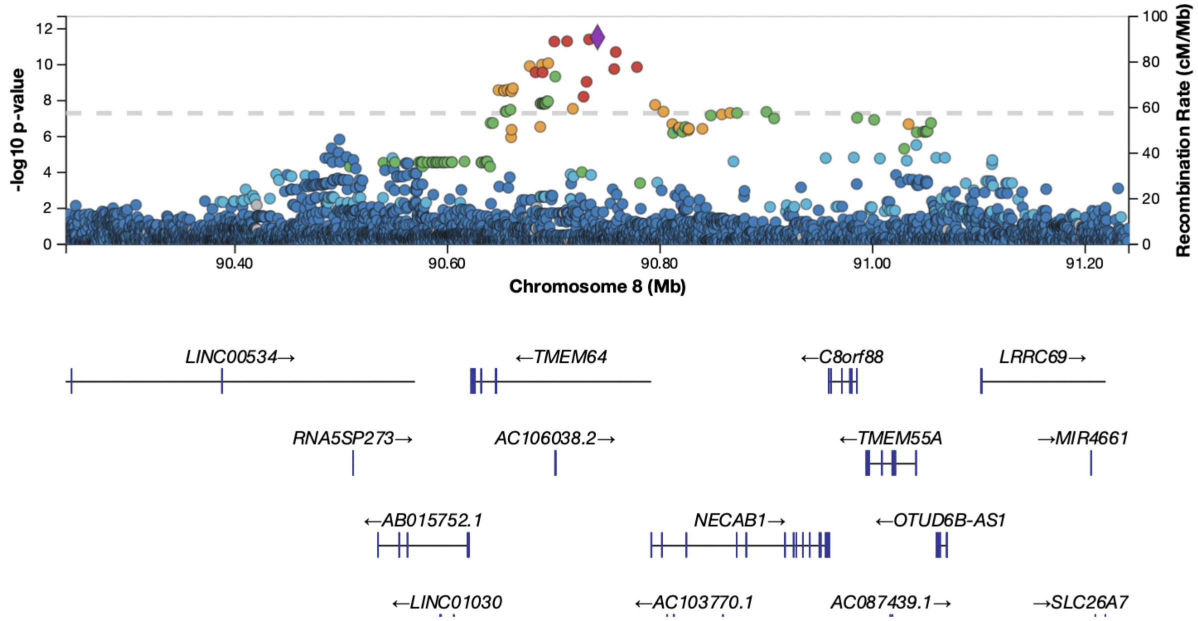
aa. Locus 12, rs7780766 and N-acetylhistidine.



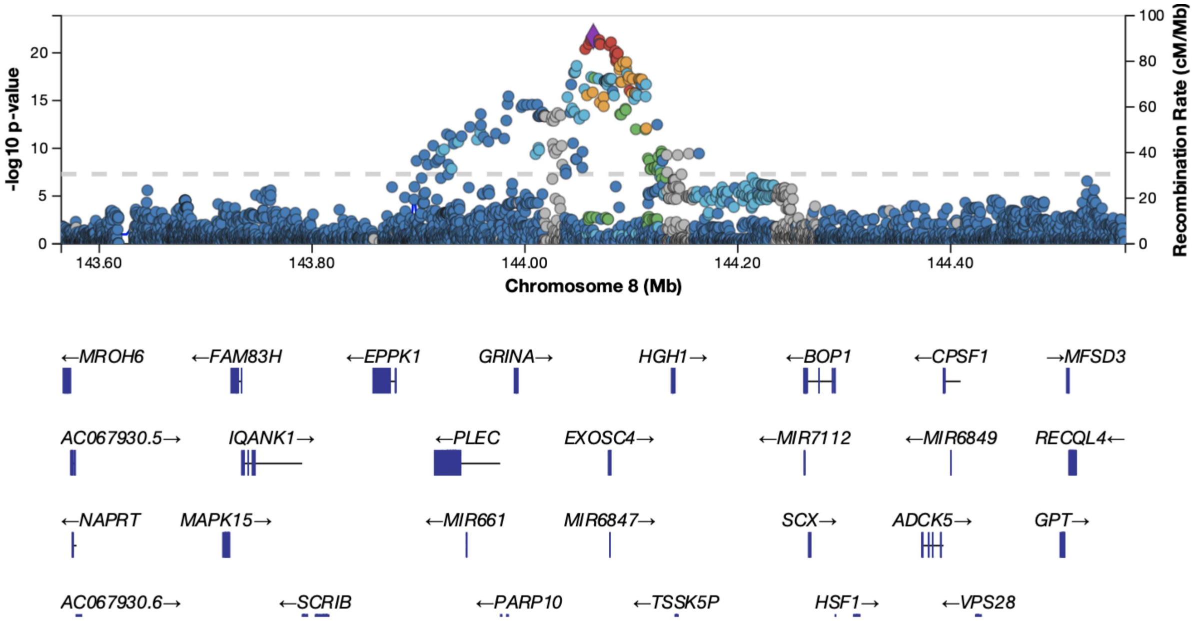
bb. Locus 13, rs139981949 and glycohyocholate.



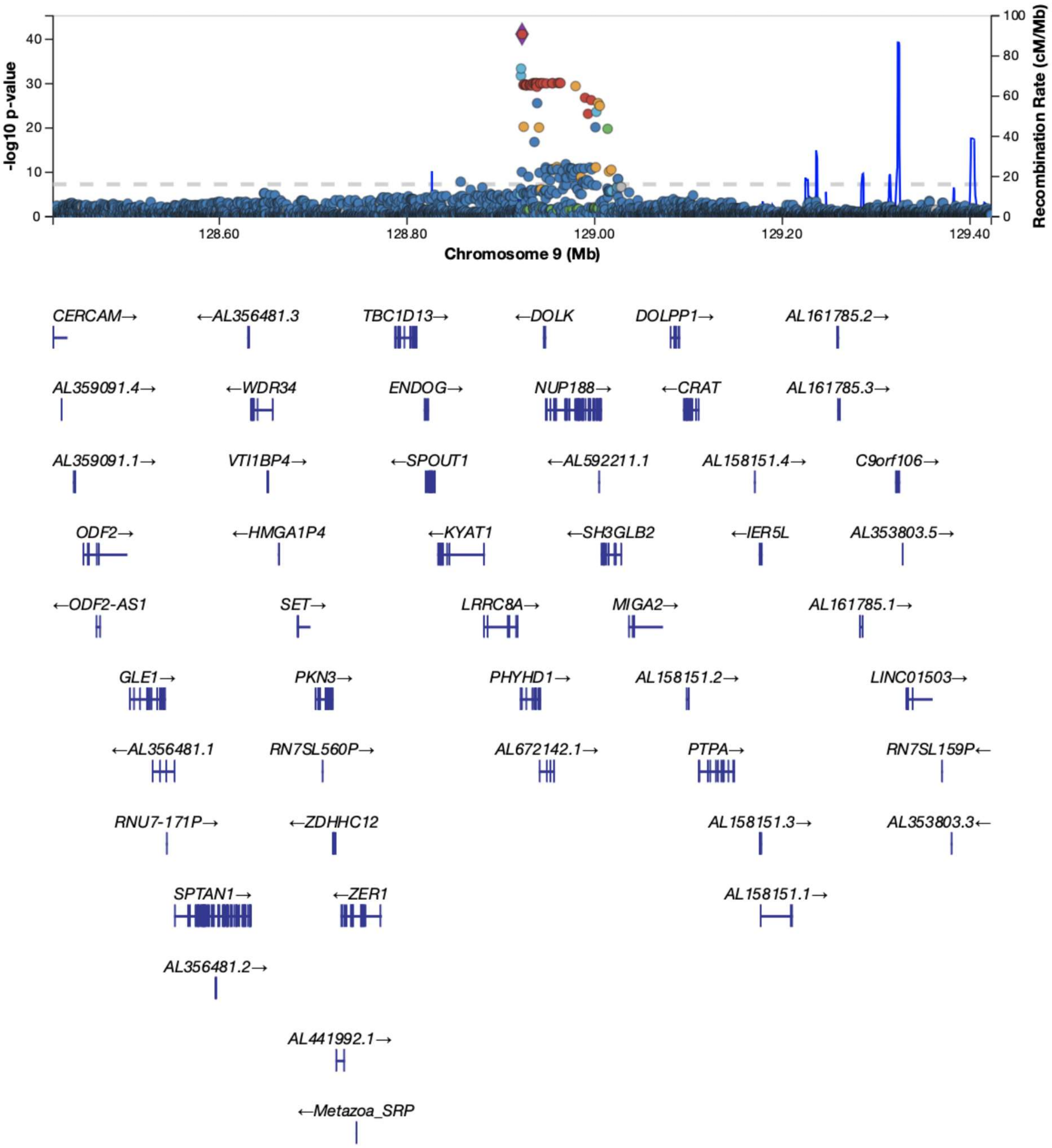
cc. Locus 14, rs76541360 and 1-arachidonoyl-GPI (20:4).



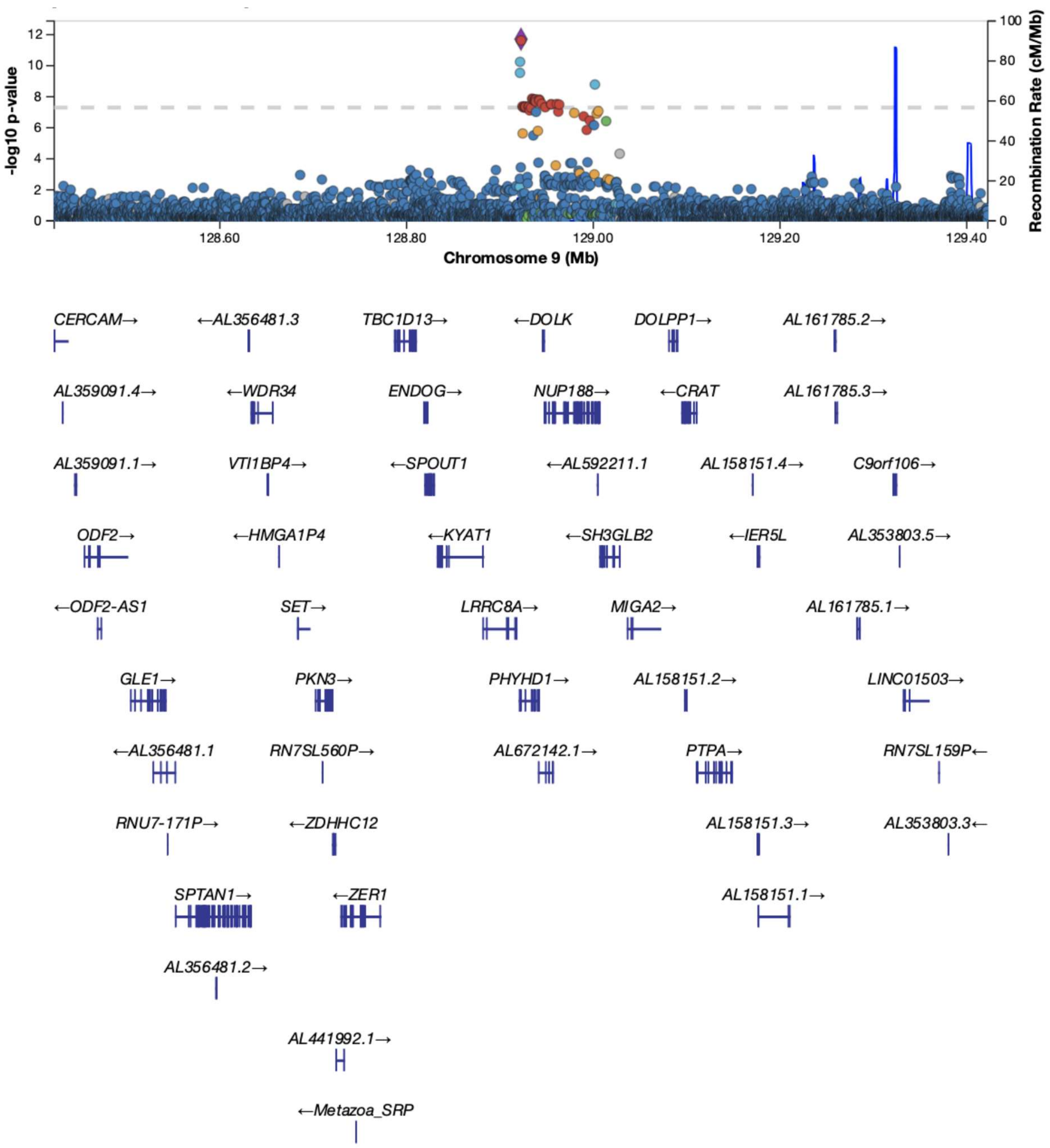
dd. Locus 15, rs59095288 and 6-oxopiperidine-2-carboxylate.



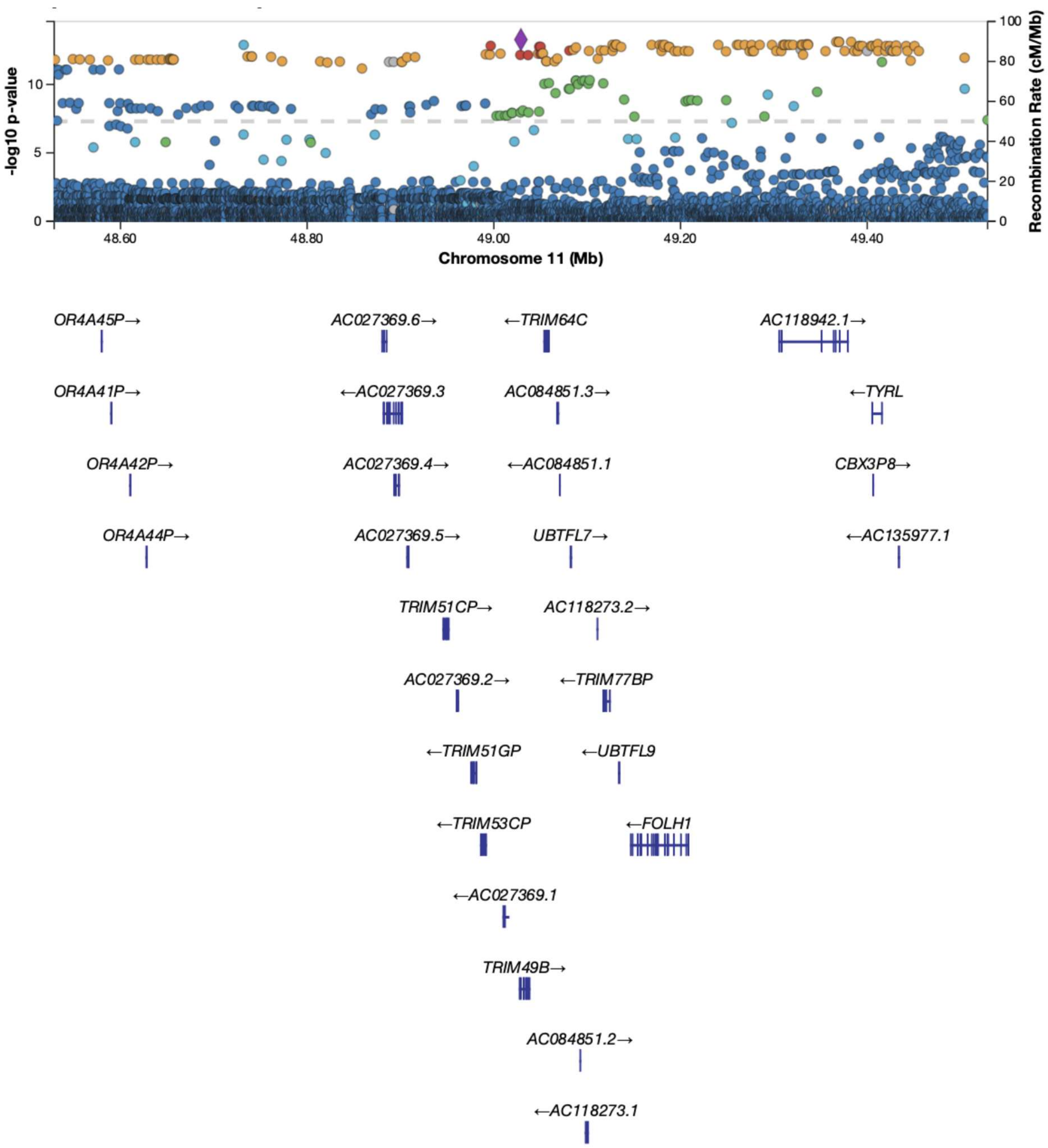
ee. Locus 16, rs55758160 and 2'-O-methyluridine.



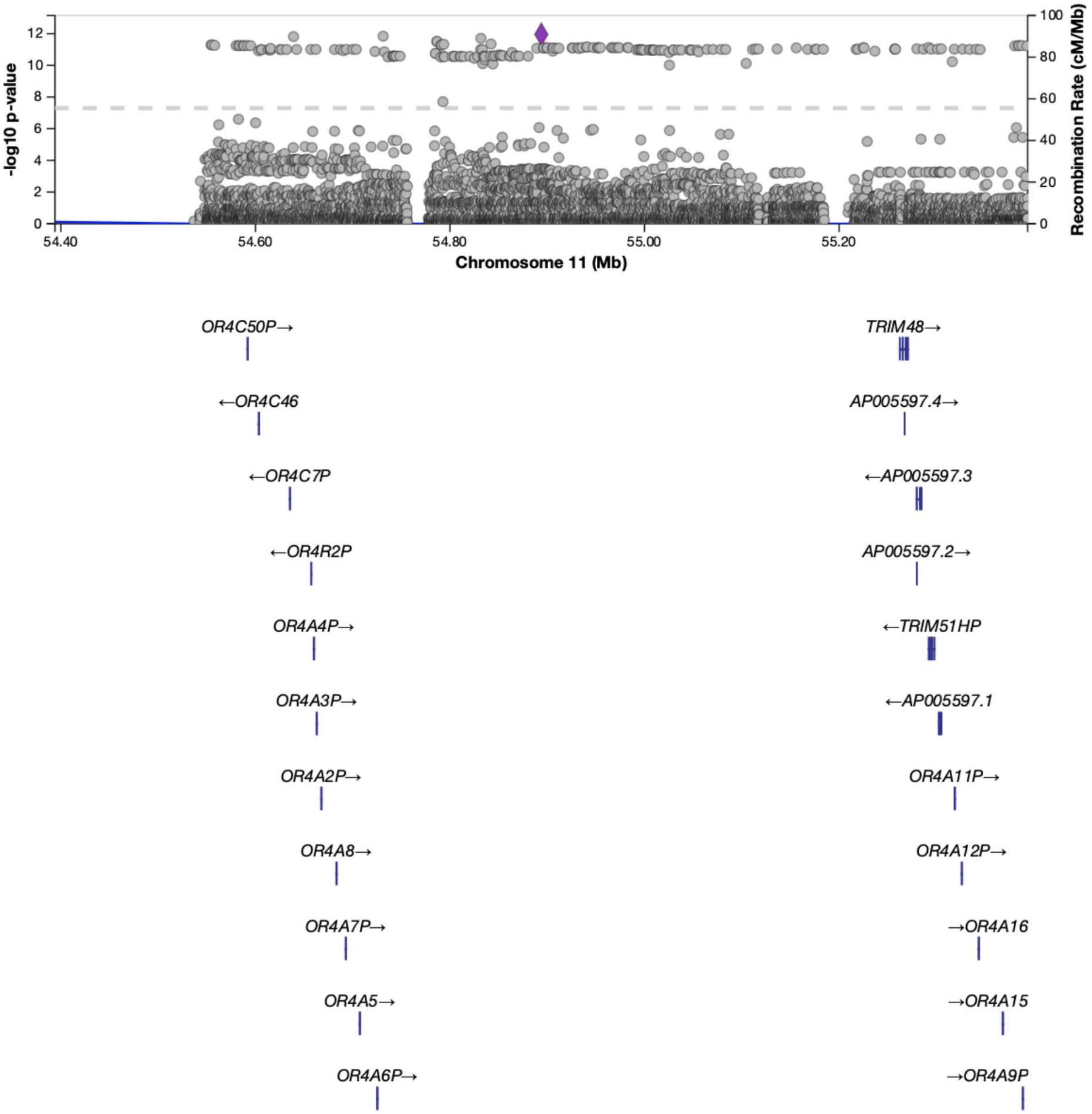
ff. Locus 16, rs57294583 and 2'-O-methylcytidine.



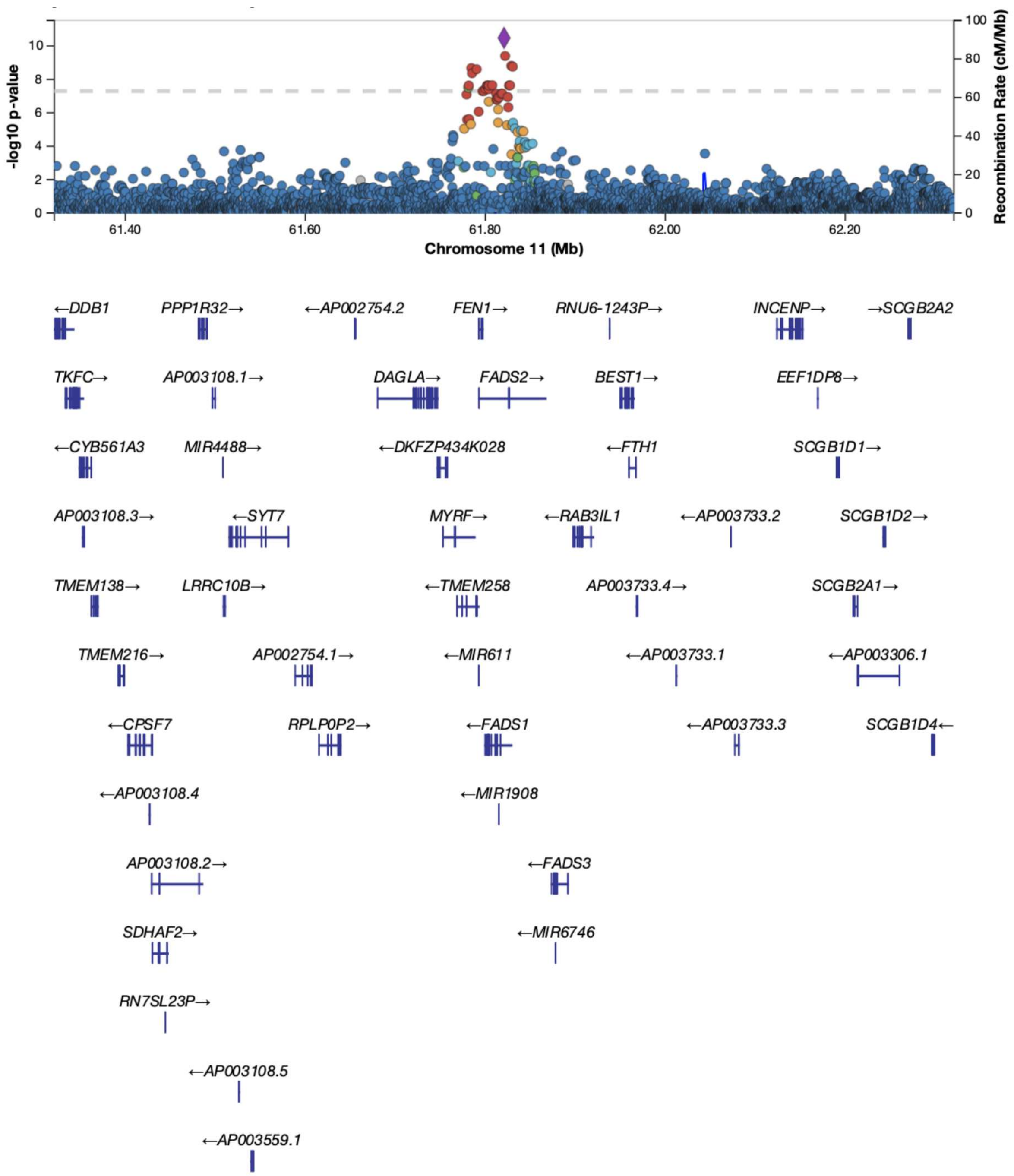
gg. Locus 17, rs142514677 and N-acetyl-aspartyl-glutamate (NAAG).



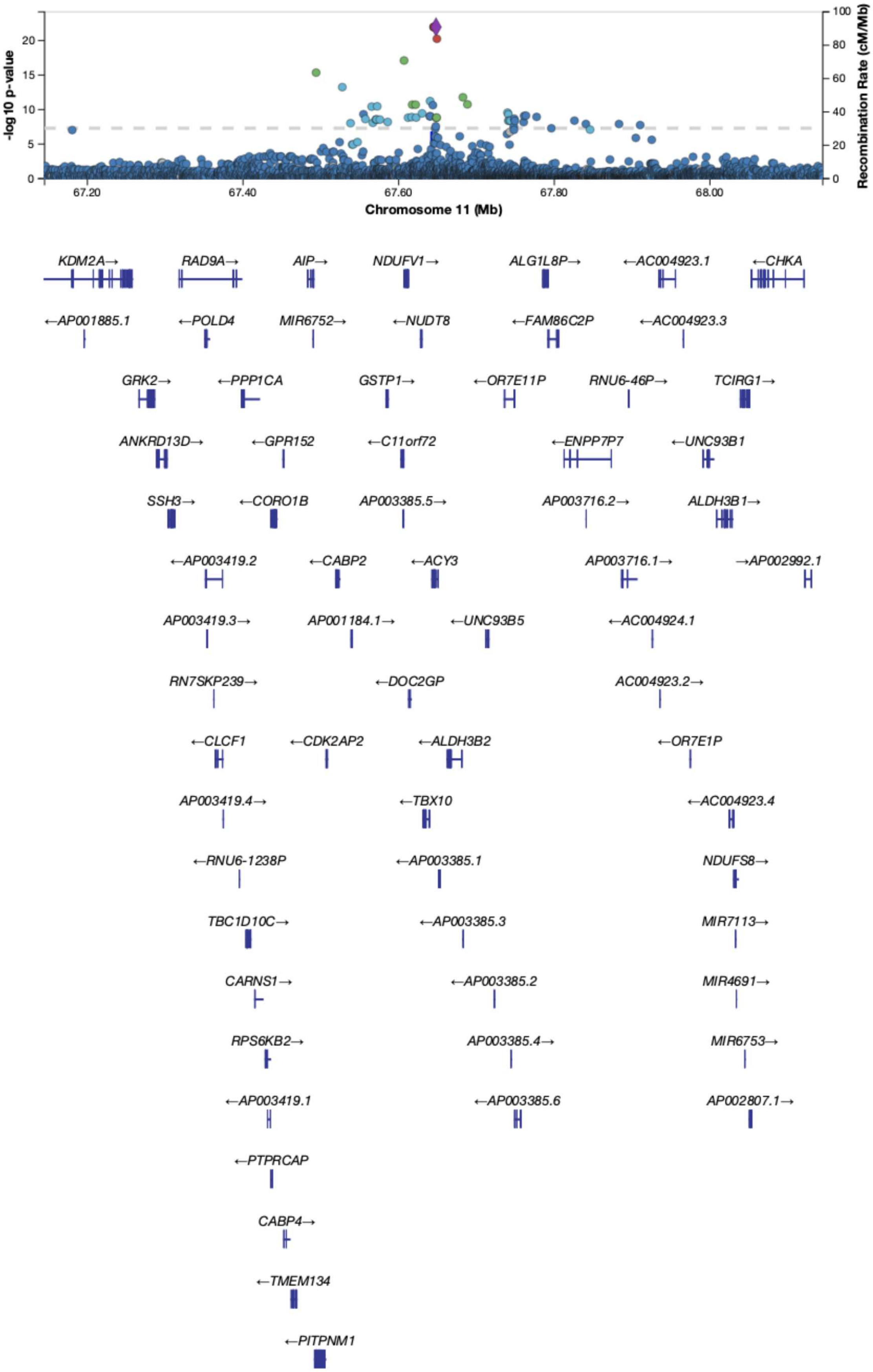
hh. Locus 18, rs150556827 and N-acetyl-aspartyl-glutamate (NAAG).



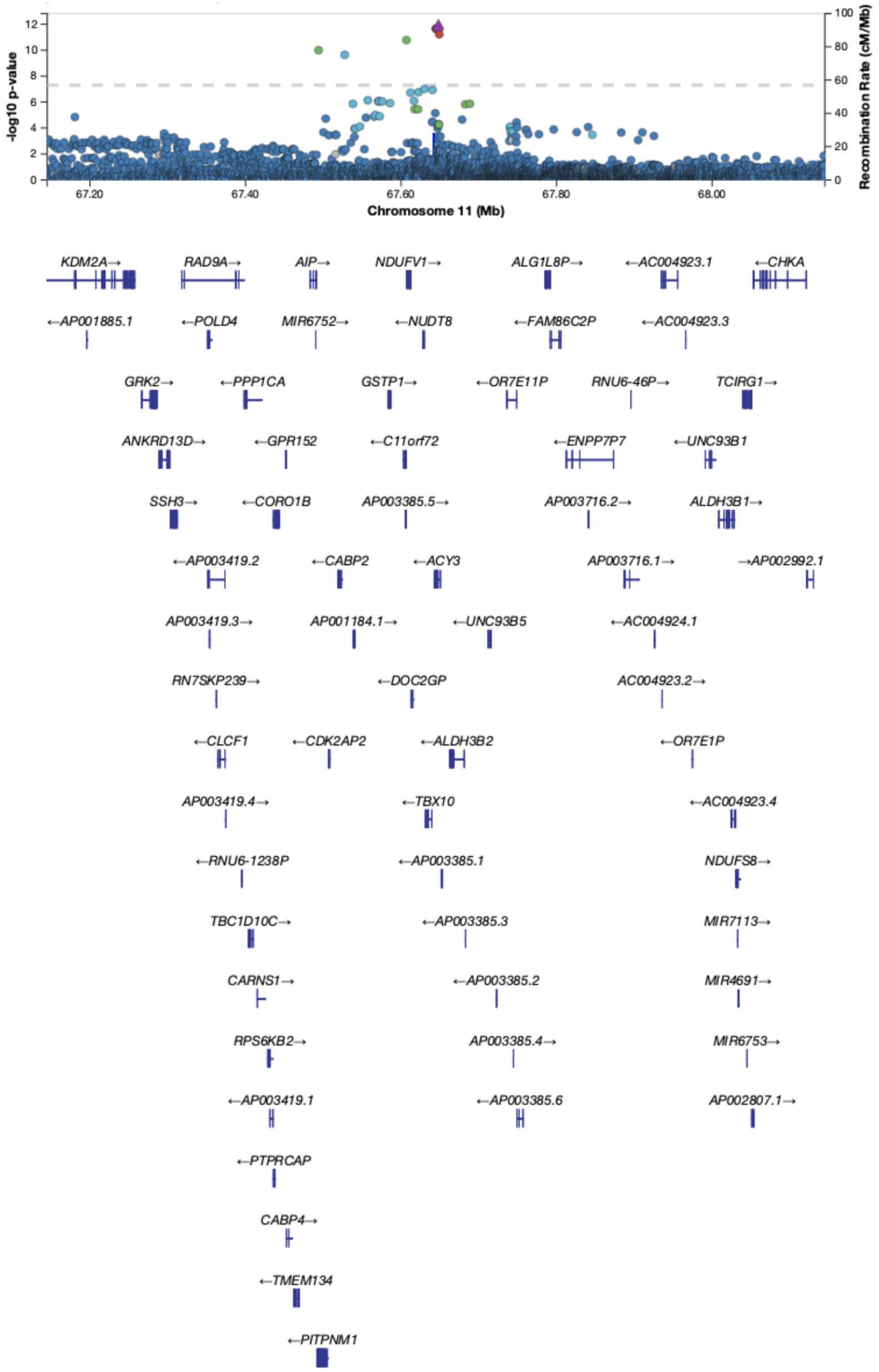
ii. Locus 19, rs174564 and 1-arachidonoyl-GPC (20:4n6).



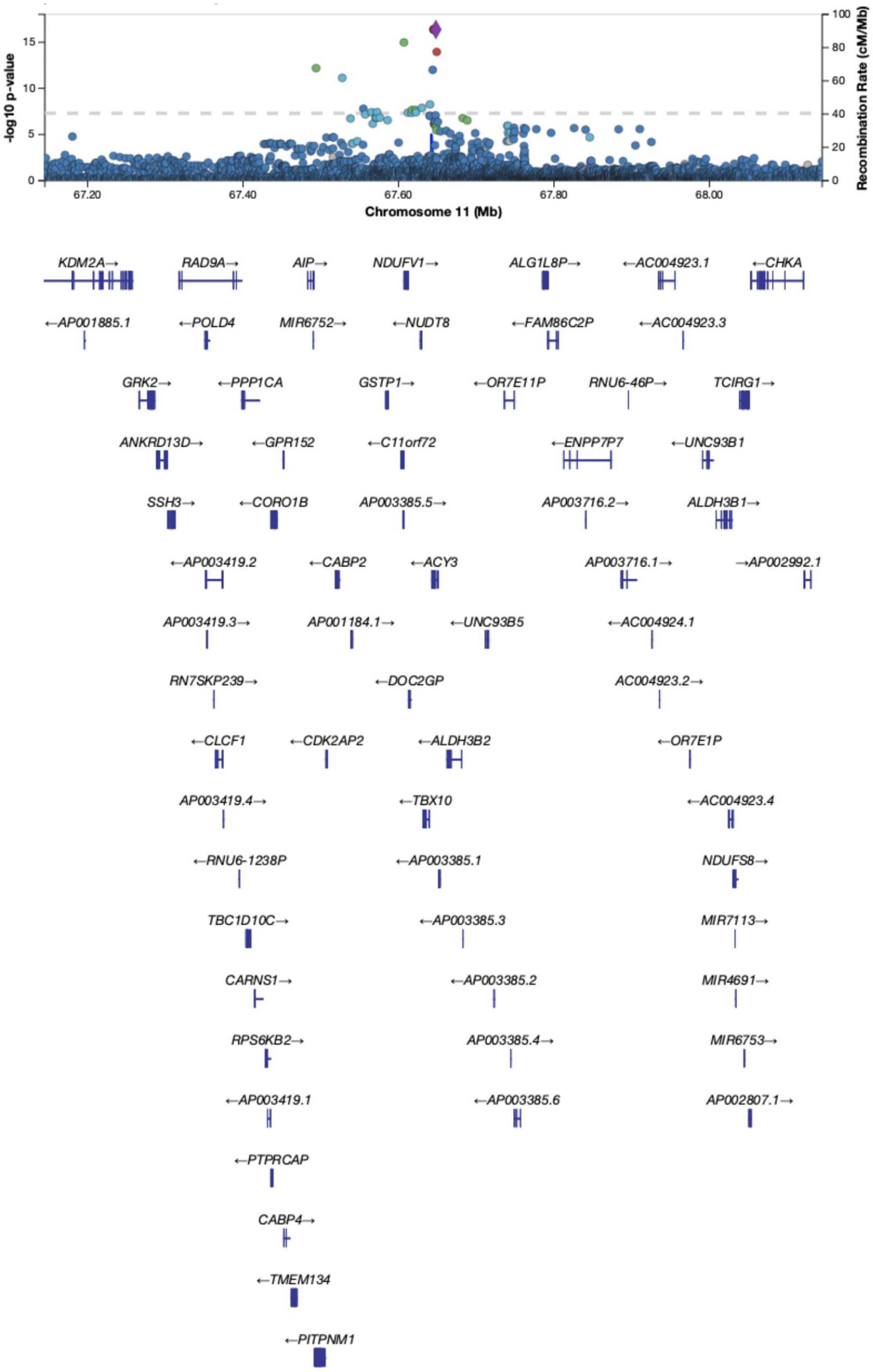
jj. Locus 20, rs115780269 and N-acetyltryptophan.



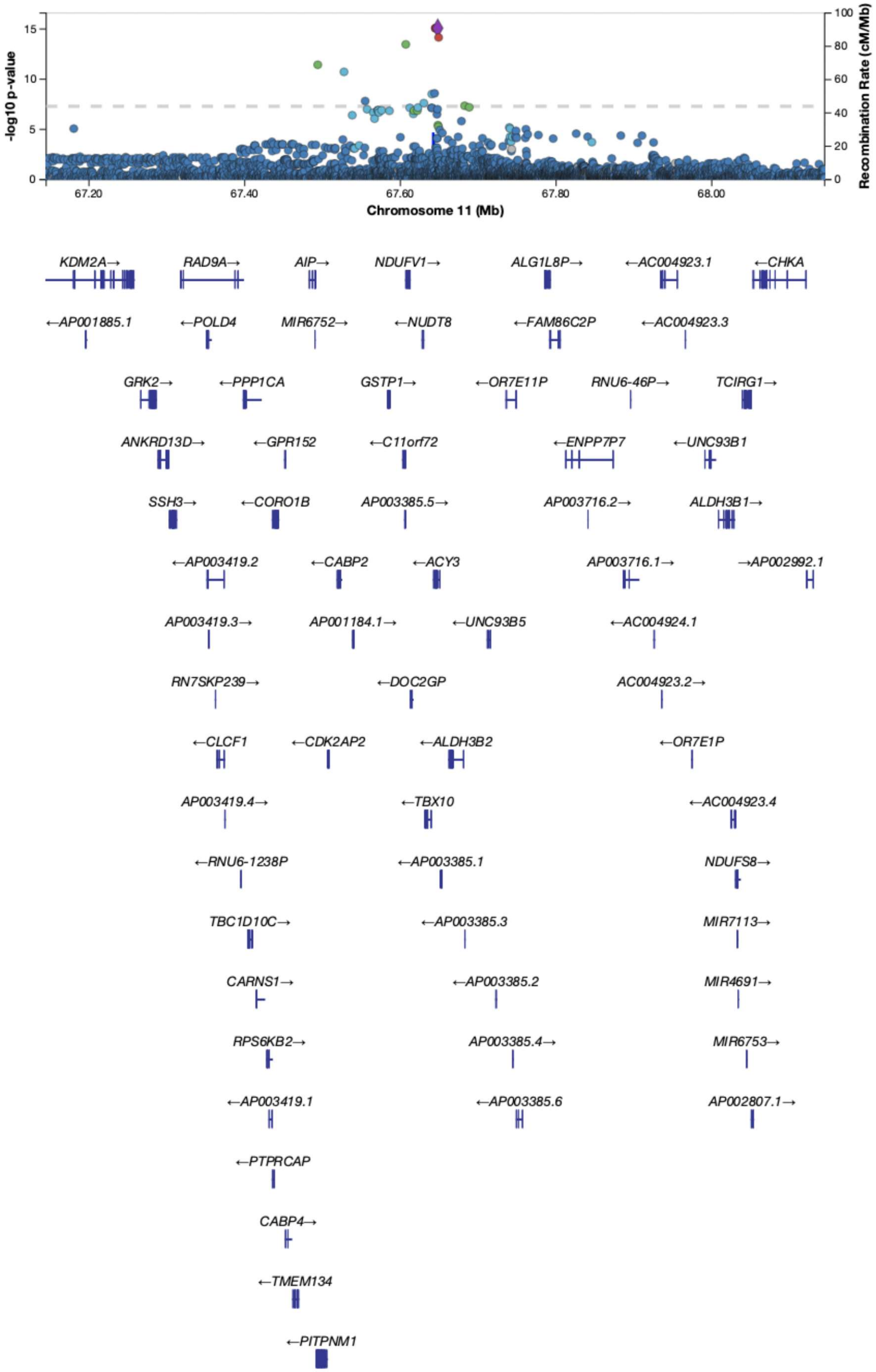
kk. Locus 20, rs115780269 and N-acetylkynurenine.



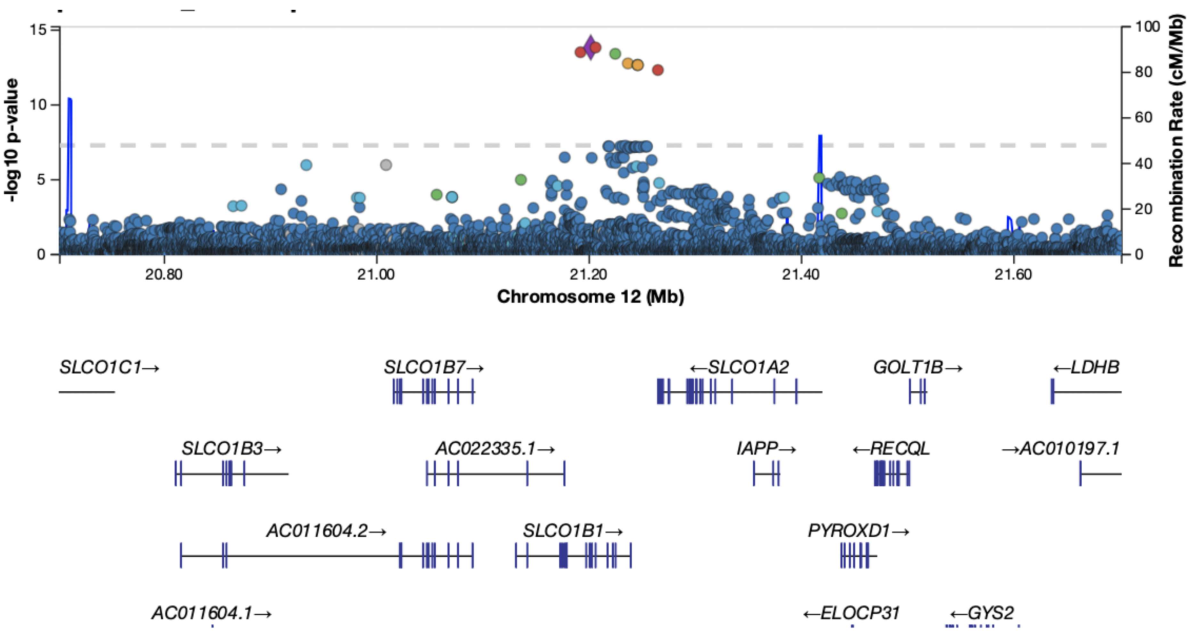
ll. Locus 20, rs115780269 and N-acetyltyrosine.



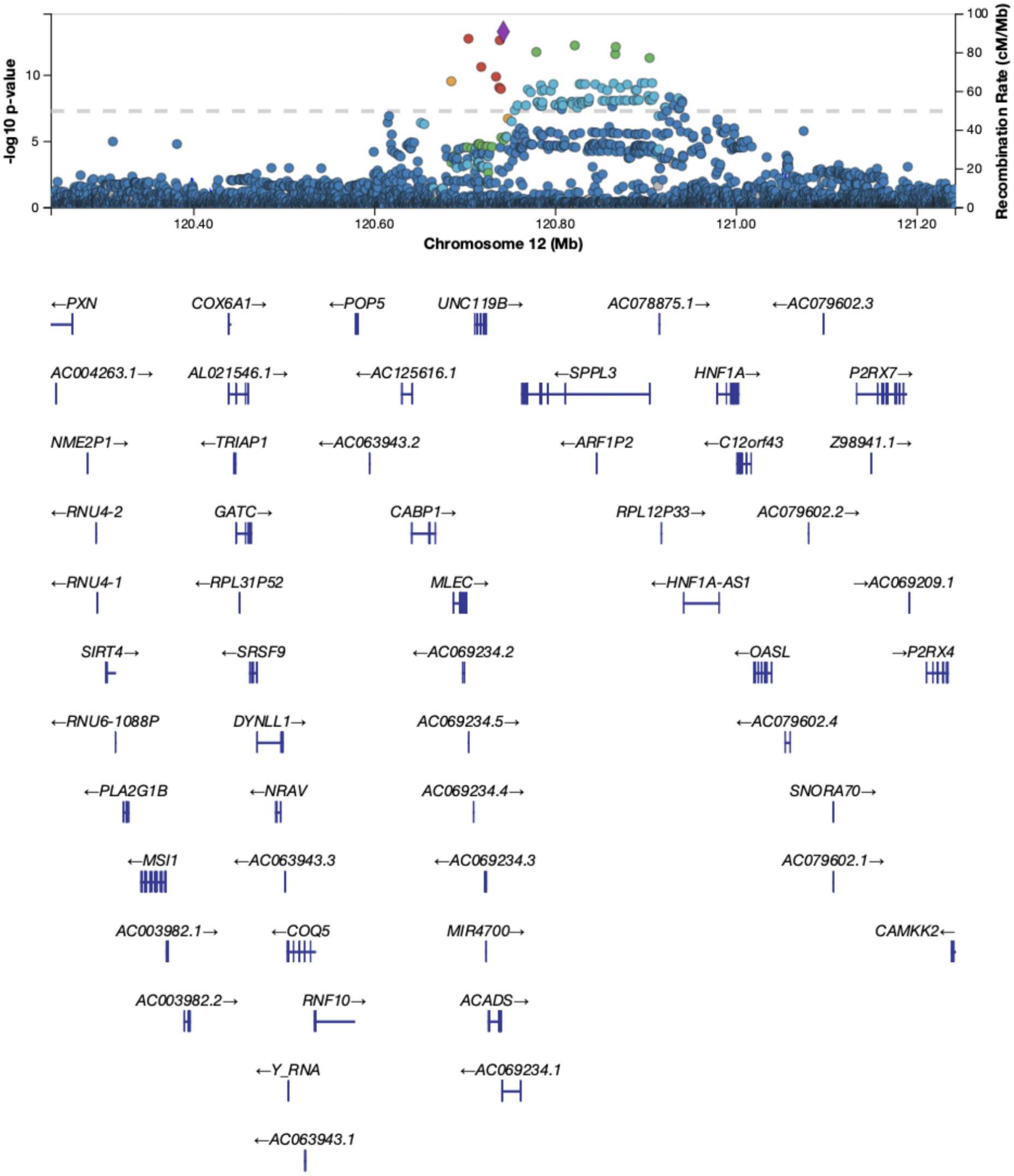
mm. Locus 20, rs115780269 and N-acetylphenylalanine.


-log10 p-value
15
10
5
0
100
80
60
40
20
0
Recombination Rate (cM/Mb)
67.20
67.40
67.60
67.80
68.00
Chromosome 11 (Mb)
KDM2A→
RAD9A→
AIP→
NDUFV1→
ALG1L8P→
←AC004923.1
←CHKA
←AP001885.1
←POLD4
MIR6752→
←NUDT8
←FAM86C2P
←AC004923.3
GRK2→
←PPP1CA
GSTP1→
←OR7E11P
RNU6-46P→
TCIRG1→
ANKRD13D→
←GPR152
←C11orf72
←ENPP7P7
←UNC93B1
SSH3→
←CORO1B
AP003385.5→
AP003716.2→
ALDH3B1→
←AP003419.2
←CABP2
←ACY3
AP003716.1→
→AP002992.1
AP003419.3→
AP001184.1→
←UNC93B5
←AC004924.1
RN7SKP239→
←DOC2GP
AC004923.2→
←CLCF1
←CDK2AP2
←ALDH3B2
←OR7E1P
AP003419.4→
←TBX10
←AC004923.4
←RNU6-1238P
←AP003385.1
NDUFS8→
TBC1D10C→
←AP003385.3
MIR7113→
CARNS1→
←AP003385.2
MIR4691→
RPS6KB2→
AP003385.4→
MIR6753→
←AP003419.1
←AP003385.6
AP002807.1→
←PTPRCAP
CABP4→
←TMEM134
←PITPNM1

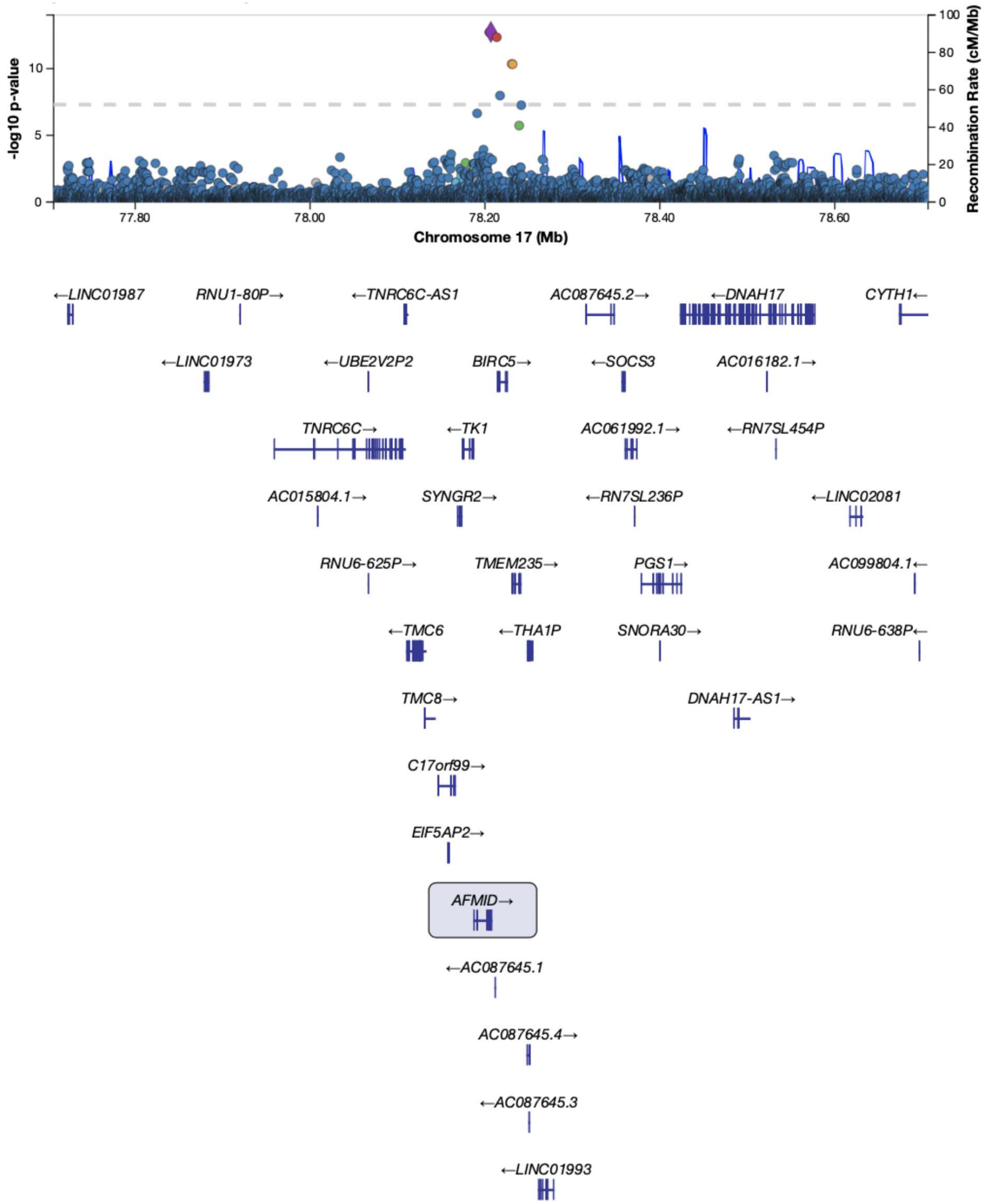
nn. Locus 21, rs114419265 and glycochenodeoxycholate glucuronide.



oo. Locus 22, rs34708625 and ethylmalonate.

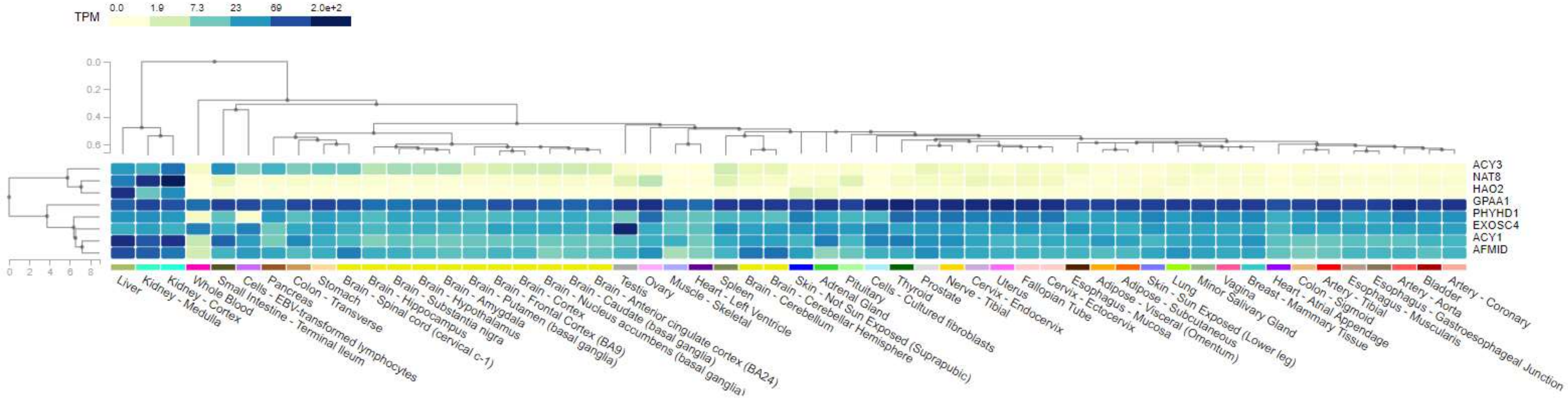


pp. Locus 23, rs114080902 and N-formylanthranilic acid.



Color-coding indicates degree of linkage disequilibrium with index SNP: Purple indicates index SNP, red indicates SNPs with $r^2 \geq 0.8$, orange indicates SNPs with $r^2 \geq 0.6$, green indicates SNPs with $r^2 \geq 0.4$, light blue indicates SNPs with $r^2 \geq 0.2$, dark blue indicates SNPs with $r^2 \geq 0$, and grey indicates SNPs with unknown r^2.

Figure S2. Gene expression in different tissue types for the eight likely causal genes for the novel genetic loci-variant associations. (Source: GTEX version 8).



Reference:

S1. Yousri NA, Fakhro KA, Robay A*, et al.* Whole-exome sequencing identifies common and rare variant metabolic QTLs in a Middle Eastern population. *Nat Commun* 2018; **9:** 333.

S2. Li Y, Sekula P, Wuttke M*, et al.* Genome-Wide Association Studies of Metabolites in Patients with CKD Identify Multiple Loci and Illuminate Tubular Transport Mechanisms. *Journal of the American Society of Nephrology : JASN* 2018; **29:** 1513-1524.

S3. Long T, Hicks M, Yu HC*, et al.* Whole-genome sequencing identifies common-to-rare variants associated with human blood metabolites. *Nature genetics* 2017; **49:** 568-578.

S4. Davis JP, Huyghe JR, Locke AE*, et al.* Common, low-frequency, and rare genetic variants associated with lipoprotein subclasses and triglyceride measures in Finnish men from the METSIM study. *PLoS genetics* 2017; **13:** e1007079-e1007079.

S5. Rhee EP, Yang Q, Yu B*, et al.* An exome array study of the plasma metabolome. *Nature communications* 2016; **7:** 12360-12360.

S6. Yu B, de Vries PS, Metcalf GA*, et al.* Whole genome sequence analysis of serum amino acid levels. *Genome Biol* 2016; **17:** 237.

S7. Yu B, Li AH, Metcalf GA*, et al.* Loss-of-function variants influence the human serum metabolome. *Sci Adv* 2016; **2:** e1600800-e1600800.

S8. Kettunen J, Demirkan A, Wurtz P*, et al.* Genome-wide study for circulating metabolites identifies 62 loci and reveals novel systemic effects of LPA. *Nat Commun* 2016; **7:** 11122.

S9. Yet I, Menni C, Shin S-Y*, et al.* Genetic Influences on Metabolite Levels: A Comparison across Metabolomic Platforms. *PLoS One* 2016; **11:** e0153672-e0153672.

S10. Burkhardt R, Kirsten H, Beutner F*, et al.* Integration of Genome-Wide SNP Data and Gene-Expression Profiles Reveals Six Novel Loci and Regulatory Mechanisms for Amino Acids and Acylcarnitines in Whole Blood. *PLoS genetics* 2015; **11:** e1005510-e1005510.

S11. Kraus WE, Muoio DM, Stevens R*, et al.* Metabolomic Quantitative Trait Loci (mQTL) Mapping Implicates the Ubiquitin Proteasome System in Cardiovascular Disease Pathogenesis. *PLoS genetics* 2015; **11:** e1005553-e1005553.

S12. Draisma HHM, Pool R, Kobl M*, et al.* Genome-wide association study identifies novel genetic variants contributing to variation in blood metabolite levels. *Nature communications* 2015; **6:** 7208-7208.

S13. Demirkan A, Henneman P, Verhoeven A*, et al.* Insight in genome-wide association of metabolite quantitative traits by exome sequence analyses. *PLoS genetics* 2015; **11:** e1004835-e1004835.

S14. Ried JS, Shin SY, Krumsiek J*, et al.* Novel genetic associations with serum level metabolites identified by phenotype set enrichment analyses. *Human molecular genetics* 2014; **23:** 5847-5857.

S15. Yu B, Zheng Y, Alexander D*, et al.* Genetic determinants influencing human serum metabolome among African Americans. *PLoS genetics* 2014; **10:** e1004212.

S16. Shin SY, Fauman EB, Petersen AK*, et al.* An atlas of genetic influences on human blood metabolites. *Nature genetics* 2014; **46:** 543-550.

S17. Rhee EP, Ho JE, Chen MH*, et al.* A genome-wide association study of the human metabolome in a community-based cohort. *Cell metabolism* 2013; **18:** 130-143.

S18. Hong M-G, Karlsson R, Magnusson PKE*, et al.* A genome-wide assessment of variability in human serum metabolism. *Hum Mutat* 2013; **34:** 515-524.

S19. Kettunen J, Tukiainen T, Sarin AP*, et al.* Genome-wide association study identifies multiple loci influencing human serum metabolite levels. *Nature genetics* 2012; **44:** 269-276.

S20. Tukiainen T, Kettunen J, Kangas AJ*, et al.* Detailed metabolic and genetic characterization reveals new associations for 30 known lipid loci. *Human molecular genetics* 2012; **21:** 1444-1455.

S21. Krumsiek J, Suhre K, Evans AM*, et al.* Mining the unknown: a systems approach to metabolite identification combining genetic and metabolic information. *PLoS genetics* 2012; **8:** e1003005-e1003005.

S22. Inouye M, Ripatti S, Kettunen J*, et al.* Novel Loci for metabolic networks and multi-tissue expression studies reveal genes for atherosclerosis. *PLoS genetics* 2012; **8:** e1002907-e1002907.

S23. Demirkan A, van Duijn CM, Ugocsai P*, et al.* Genome-wide association study identifies novel loci associated with circulating phospho- and sphingolipid concentrations. *PLoS genetics* 2012; **8:** e1002490-e1002490.

S24. Suhre K, Shin SY, Petersen AK*, et al.* Human metabolic individuality in biomedical and pharmaceutical research. *Nature* 2011; **477:** 54-60.

S25. Mittelstrass K, Ried JS, Yu Z*, et al.* Discovery of sexual dimorphisms in metabolic and genetic biomarkers. *PLoS genetics* 2011; **7:** e1002215-e1002215.

S26. Nicholson G, Rantalainen M, Li JV*, et al.* A genome-wide metabolic QTL analysis in Europeans implicates two loci shaped by recent positive selection. *PLoS genetics* 2011; **7:** e1002270-e1002270.

S27. Illig T, Gieger C, Zhai G*, et al.* A genome-wide perspective of genetic variation in human metabolism. *Nature genetics* 2010; **42:** 137-141.

S28. Gieger C, Geistlinger L, Altmaier E*, et al.* Genetics meets metabolomics: a genome-wide association study of metabolite profiles in human serum. *PLoS genetics* 2008; **4:** e1000282-e1000282.