Nonadditive Effects of Genes in Human Metabolomics

Yakov A. Tsepilov,*,†,‡,§, So-Youn Shin,††,‡‡ Nicole Soranzo,††,§§ Tim D. Spector,*** Cornelia Prehn,†††
Jerzy Adamski,†††,‡‡‡,§§§ Gabi Kastenmüller,***,**** Rui Wang-Sattler,§,** Konstantin Strauch,‡,††††
Christian Gieger,‡,**,§ Yurii S. Aulchenko,*,† and Janina S. Ried‡**

*Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences, 630090 Novosibirsk, Russia, †Novosibirsk State University, 630090 Novosibirsk, Russia, ‡Institute of Genetic Epidemiology, §Research Unit of Molecular Epidemiology, **Institute of Epidemiology II, †††Institute of Experimental Genetics, Genome Analysis Center, and ****Institute of Bioinformatics and Systems Biology, Helmholtz Zentrum München–German Research Center for Environmental Health, 85764 Neuherberg, Germany, ††Human Genetics, Wellcome Trust Sanger Institute, Genome Campus, Hinxton, CB10 1HH, United Kingdom, ‡‡MRC Integrative Epidemiology Unit, School of Social and Community Medicine, University of Bristol, Bristol, BS8 1TH, United Kingdom, §§Department of Haematology, University of Cambridge, Cambridge, CB2 0AH, United Kingdom, ***Department of Twin Research and Genetic Epidemiology, King's College London, London, WC2R 2LS, United Kingdom, ‡‡‡Institute of Experimental Genetics, Life and Food Science Center Weihenstephan, Technische Universität München, 85354 Freising-Weihenstephan, Germany, §§§German Center for Diabetes Research, 85764 Neuherberg, Germany, ††††Institute of Medical Informatics, Biometry and Epidemiology, Chair of Genetic Epidemiology, Ludwig-Maximilians-Universität, 85764 Neuherberg, Germany

ABSTRACT Genome-wide association studies (GWAS) are widely applied to analyze the genetic effects on phenotypes. With the availability of high-throughput technologies for metabolite measurements, GWAS successfully identified loci that affect metabolite concentrations and underlying pathways. In most GWAS, the effect of each SNP on the phenotype is assumed to be additive. Other genetic models such as recessive, dominant, or overdominant were considered only by very few studies. In contrast to this, there are theories that emphasize the relevance of nonadditive effects as a consequence of physiologic mechanisms. This might be especially important for metabolites because these intermediate phenotypes are closer to the underlying pathways than other traits or diseases. In this study we analyzed systematically nonadditive effects on a large panel of serum metabolites and all possible ratios (22,801 total) in a population-based study [Cooperative Health Research in the Region of Augsburg (KORA) F4, $N = 1{,}785$]. We applied four different 1-degree-of-freedom (1-df) tests corresponding to an additive, dominant, recessive, and overdominant trait model as well as a genotypic model with two degree-of-freedom (2-df) that allows a more general consideration of genetic effects. Twenty-three loci were found to be genome-wide significantly associated (Bonferroni corrected $P \leq 2.19 \times 10^{-12}$) with at least one metabolite or ratio. For five of them, we show the evidence of nonadditive effects. We replicated 17 loci, including 3 loci with nonadditive effects, in an independent study (TwinsUK, $N = 846$). In conclusion, we found that most genetic effects on metabolite concentrations and ratios were indeed additive, which verifies the practice of using the additive model for analyzing SNP effects on metabolites.

KEYWORDS genome-wide association studies; nonadditive models; KORA; metabolomics; genotypic model

IN RECENT years, genome-wide association studies (GWAS) have identified many loci that affect quantitative traits or diseases. The development of high-throughput molecular biology technologies enabled the measurement of hundreds of metabolites that are small molecules and intermediate products of metabolism. Gieger *et al.* (2008) published the first GWAS using metabolite concentrations as outcome. Other studies followed with great success (Hicks *et al.* 2009; Kolz *et al.* 2009; Tanaka *et al.* 2009; Illig *et al.* 2010; Nicholson *et al.* 2011; Suhre *et al.* 2011; Demirkan *et al.* 2012; Kettunen *et al.* 2012; Suhre and Gieger 2012; Shin *et al.* 2014). Many loci identified by these studies provide a deeper insight into underlying pathways.

doi: 10.1534/genetics.115.175760
Manuscript received February 22, 2015; accepted for publication May 4, 2015; published Early Online May 14, 2015.
Supporting information is available online at www.genetics.org/lookup/suppl/doi:10.1534/genetics.115.175760/-/DC1
[1]Corresponding author: Institute of Genetic Epidemiology, Helmholtz Zentrum München, 85764, Neuherberg, Germany. E-mail: janina.s.ried@gmx.de

Like GWAS in general, most GWAS on metabolites analyzed the additive effects of SNPs on metabolite levels and neglected other possible genetic effects. One reason for this, besides computational simplification, is the question of power; if different genetic models were applied in parallel, it would be necessary to account for multiple testing, which can reduce the power. Several studies that analyzed additive *vs.* nonadditive SNP effects support the widely used assumption of additive SNP effects. Hill *et al.* (2008) found in different empirical studies that there is mainly additive variance for complex traits. They did not disclaim the existence of nonadditive effects on the level of genes but concluded that those in most cases do not considerably affect the variance at the phenotype level. Other studies show that gene expression is controlled by predominantly additive genes (Powell *et al.* 2013).

In contrast, there is a long-running discussion about the incidence and contribution of nonadditive effects (Fisher 1928a, b; Wright 1929; Haldane 1930; Kacser and Burns 1981; Orr 1991; Porteous 1996). Fisher's theory of dominance assumed that dominance is a result of selection that aligns the effect of the heterozygote genotype with the "normal" homozygote genotype (Fisher 1928a, b). Wright (1929) and Haldane (1930) discussed the idea that the dominance is based on physiologic factors and that selection for genotype modifiers is not the primary origin. Subsequent theories tended to favor Wright's view without completely excluding the idea that Fisher's argument may hold in some circumstances (Kacser and Burns 1981; Orr 1991). Kacser and Burns (1981) proposed that dominance is a "consequence of the kinetic properties of enzyme-catalyzed pathways." Their modern biochemical view on dominant/recessive effects was based on the observation that the increase in reaction rate does not depend linearly on the enzyme activity or concentration but is slowed down toward the upper end. The interpretation of these observations is that a single enzyme cannot be thought to be independent but is embedded in a whole system of "fluxes" that mutually lessen responses to increases in enzyme activity (Kacser and Burns 1981). This theory of dominance may have especially relevance for metabolites because metabolites are sometimes direct products of enzymatic reactions and hence closer to gene action than clinical phenotypes. If a SNP has a nonadditive effect on a metabolite level, the assumption of an additive effect will reduce the power to detect the association. Therefore, previous GWAS may have overseen SNP associations with nonadditive effects on the metabolite profile.

The objective of this study was to investigate nonadditive effects on serum metabolite levels on a genome-wide scale to answer the question of whether genetic analysis of metabolites benefits from the analysis of nonadditive effects. To our knowledge, this is the first study that systematically investigates nonadditive SNP effects on a large panel of serum metabolites in a population-based study.

Materials and Methods

KORA study

The Cooperative Health Research in the Region of Augsburg (KORA) studies are based on several cross-sectional population-based cohorts from the region of Augsburg in southern Germany (Wichmann *et al.* 2005). The cohort KORA F4 is the follow-up survey (2006–2008) of the baseline survey KORA S4 that was conducted from 1999 to 2001. All study protocols were approved by the ethics committee of the Bavarian Medical Chamber (Bayerische Landesärztekammer), and all participants gave written informed consent. Concentrations of 163 metabolites were measured in 3061 participants in KORA F4 using flow injection electrospray ionization tandem mass spectrometry and the Absolute*IDQ* p150 Kit (Biocrates Life Sciences AG, Innsbruck, Austria) in the fasting serum of individuals in KORA F4. Details of the measurement methods and quality control and a list of the metabolites were given previously (Illig *et al.* 2010; Ried *et al.* 2012; Römisch-Margl *et al.* 2012). Genotyping was performed with the Affymetrix 6.0 SNP Array (534,174 SNP markers after quality control), with further imputation using HapMap2 (Release 22) as a reference panel, resulting in a total of 1,717,498 SNPs (details given in Kolz *et al.* 2009). For 1785 samples, both metabolite concentrations and genotypes were available.

TwinsUK study

The TwinsUK study is a British adult twin registry (Moayyeri *et al.* 2012). The participants were recruited from the general population through national media campaigns in the United Kingdom. It was shown that participants are comparable to age-matched population singletons in terms of disease-related and lifestyle characteristics (Andrew *et al.* 2001). Ethics approval was obtained from the Guy's and St. Thomas' Hospital Ethics Committee, and written informed consent was obtained from every participant in the study. A total of 2277 individuals of European ancestry (1073 singletons and 602 dizygotic twins) was genotyped using the Illumina Hap317K Chip. The metabolite measurements were performed with the same targeted metabolomics platform and the identical protocol as for the KORA study was applied at the Genome Analysis Center of the Helmholtz Zentrum München. For more detailed description, see Römisch-Margl *et al.* (2012) and Menni *et al.* (2013).

Methods of analysis, transformation, and study-specific quality control

For the GWAS analysis, we used standard tests implemented in the genome-wide feasible generalized least squares (GWFGLS) function with Wald approximation of the MixABEL package, which is a part of the GenABEL suite of programs (Aulchenko *et al.* 2007) for statistical genomics. For the analysis of imputed data, the regression was performed on genotype probabilities. The genomic control method was applied to correct for a possible inflation of

the test statistics. We used functions implemented in the GenABEL package (PGC and VIFGC) (Tsepilov *et al.* 2013). The lambda for all traits was between 1.00 and 1.03.

We analyzed concentrations of all metabolites (151) and all possible ratios among pairs (22,650) and applied inverse-normal transformation (Beasley *et al.* 2009). The analysis was adjusted for sex, age, and batch.

Only SNPs that had a call rate ≥ 0.95, $R^2 \geq 0.3$, Hardy-Weinberg equilibrium (HWE) $P \geq 10^{-6}$, and minor allele frequency (MAF) ≥ 0.1 were regarded. Moreover, we excluded SNPs that showed one of the three possible genotypes for fewer than 30 individuals.

To reduce the computational effort, we decided to calculate GWAS using directly genotyped SNPs only (482,616 SNPs), selected SNPs with a liberal *P*-value threshold $\leq 5 \times 10^{-7}$, analyzed for these hits all imputed SNPs in a 500-kb region ($\pm$250 kb), and chose the most strongly associated SNP in this region. After that we compared all identified SNPs and checked whether they were located on the same chromosome and less than 250 kb apart, we considered these SNPs as one locus and chose the most associated SNP (lowest *P*-value) and trait as a top SNP-trait combination. Then regional association plots were made to check for other hits within a 1000-kb region.

Models for nonadditive effects

The widely used model of *additive* SNP effects assumes that the effect of the genotype on the (quantitative) trait changes linearly with the number of copies of the effect allele (Figure 1). This change can be either increasing or decreasing depending on the allele that was chosen to be the effect allele. We did not choose the effect and other allele independent of the allele frequency. Therefore, the effect allele can be the minor or major allele. All alleles refer to the plus strand. If additive SNP effects are assumed, the three possible genotypes of a biallelic marker with the alleles a_1 and a_2, where a_2 denotes the effect allele, can be written in numeric representation as 0 (a_1a_1), 0.5 (a_1a_2), and 1 (a_2a_2). These numbers correspond to the relative genotype effects. If a single copy of the effect allele has the same effect on a trait as two copies, the model is called *dominant*. Correspondingly, the numeric representation of the genotypes are 0 (a_1a_1), 1 (a_1a_2), and 1 (a_2a_2). The important point is that the effect is the same for the heterozygous and effect allele homozygous genotype, and again, this effect can be either increasing or decreasing. This model of genotype effects is sometimes named *complete dominance* to distinguish it from situations in which the absolute relative effect of the heterozygous genotype is higher than 0.5 but lower than 1. We focused on complete dominance and use the term *dominance* in the sense of *complete dominance* in the following. The complementary model, that zero copies of the effect allele have the same effect as one copy, is called *recessive* [0 (a_1a_1), 0 (a_1a_2), and 1 (a_2a_2)]. It is clear that if the effect allele has a dominant effect on the trait, the other allele has a recessive effect (Figure 1) and vice versa. Another model is that both homozygous genotypes have the same effect on a trait, and the heterozygous genotype either increases or decreases the trait. This model is called *overdominant* (Figure 1), with the numerical representation of the genotypes being 0 (a_1a_1), 1 (a_1a_2), and 0 (a_2a_2). Other forms of overdominance are possible; therefore, this special case is sometimes called *pure overdominance*. We use in the following the term *overdominance* interchangeably with *pure overdominance*. These four trait models are implemented as tests with one degree-of-freedom (df) in a regression. It is also possible to regard different effects for all three genotypes to allow for more complex effects. This more general model has two df and is called *genotypic* in the following. It enables us to analyze additive and the described nonadditive effect models because those are special cases of the genotypic model (Figure 1) (Zheng *et al.* 2006).

Two-step approach for identification of nonadditive effects

First, we conducted a genome-wide screening for effects on metabolite concentrations and ratios with a 2-df test for the genotypic model. For each locus identified in this step, we calculated statistics for the four 1-df models mentioned earlier and used two tests for comparison: the likelihood ratio test (LRT) (Huelsenbeck and Crandall 1997) and the Akaike information criterion (AIC) (Akaike 1974). The null hypothesis (H_0) for the LRT is that the 2-df test is not significantly better than the 1-df tests; the alternative (H_1) is that the 2-df test is significantly better. This means that if the *P*-value was below the threshold (for this study, $0.05/20 = 0.0025$ after Bonferroni correction), H_0 could be rejected. If the LRT rejects H_0 for all 1-df models, in other words, the genotypic model is found to have a better fit than all 1-df models, it will be interpreted as evidence for a genetic effect that is neither additive nor can be described by dominant, recessive, or overdominance effects. Moreover, if the LRT rejects H_0 for the additive 1-df test but neither for the dominant, recessive, nor overdominant 1-df tests, it also will be interpreted as evidence of nonadditivity.

One-step approach to identify nonadditive effects with 1-df models

We performed a genome-wide screening for the same data as described earlier but applied the additive, recessive, dominant, and overdominant models directly for the screen.

Replication

For replication, we used the TwinsUK study. The same transformation was applied to the phenotypes as for the KORA study. The replication threshold was chosen as $P = 0.05/20 = 0.0025$ with Bonferroni correction for the genotypic model and $P = 0.05/22 \approx 0.0023$ for the additive model. We used the same SNP and the same metabolite level or ratio for the replication test that we identified in the discovery step.

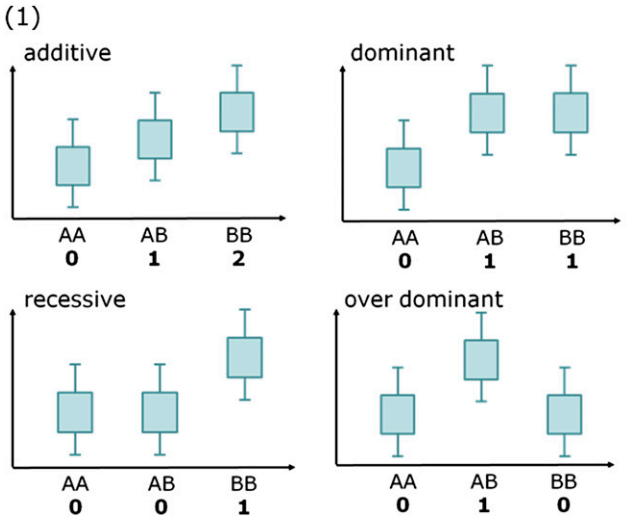


Figure 1 Schematic overview of SNP effects on quantitative phenotypes. (1) Additive, dominant, recessive, and overdominant models. The genotype effect assumed in these models can be coded by a single variable. Values of this variable are given in bold below the corresponding genotypes. (2) Genotypic model. The genotype effect assumed in this model is coded by two variables (Var1 and Var2).

Results

Two-step approach for identification of nonadditive effects

By screening with the genotypic model, we found 20 loci that were significantly associated with at least one metabolite or metabolite ratio with a Bonferroni-corrected significance level of $5 \times 10^{-8}/22{,}801 \approx 2.19 \times 10^{-12}$. These loci are presented in Table 1 (all regional association plots for these loci are given in Supporting Information, File S1). Sixteen loci could be replicated in the TwinsUK study (Bonferroni-corrected $P = 0.05/20 = 0.0025$). It should be noted that SNP rs715 was in high linkage disequilibrium (LD; $R^2 = 0.912$) with SNP rs7422339 that had a stronger association in KORA ($P = 5.19 \times 10^{-74}$), but it was neither genotyped nor imputed in the TwinsUK study. Therefore, SNP rs715 was selected as proxy for this locus and could be successfully replicated. SNP rs6970485 was neither genotyped nor imputed in the TwinsUK study. No proxy SNP with good quality could be identified, so we could not analyze this SNP in the replication step.

For each locus we compared the 1-df models to the genotypic model with the LRT and AIC among each other. Results of these analyses are given in Table 2. For 14 of the identified loci, the genotypic model did not fit significantly better than the additive model but better than all other 1-df models (indicated in Table 2 with a single "a" in column LRT for KORA), which can be interpreted as evidence of additivity. Regarding the AIC, the best model for those loci was additive (10 loci) or genotypic (4 loci represented by rs273913, rs174547, rs1077989, and rs603424). For two loci (rs4902242 and rs7200543), the genotypic model did not show a significantly better fit (LRT) than two 1-df models (additive/recessive and additive/dominant, respectively); the AIC identified the best model as genotypic for rs4902242 and additive for rs7200543. The remaining four loci showed clear evidence of nonadditive effects. For the rs715 locus, the genotypic model did not fit significantly better than the recessive model (with the AIC indicating a best-fitting genotypic model). This finding was replicated in the TwinsUK study. For two loci (represented by rs2066938 and rs7601356), the genotypic model yielded a significantly better fit than all 1-df models, which also could be observed in the TwinsUK study. For SNP rs6970485, the genotypic model did not fit significantly better than the dominant model (LRT), and the AIC indicated a best-fitting dominant model as well. However, this finding could not be replicated in the TwinsUK study because there the SNP was neither genotyped nor imputed.

One- step approach to identify nonadditive effects with 1-df model

By screening with the additive model, we identified the 20 loci that were found with the genotypic model (partly represented by other SNPs and metabolite ratios) and 2 additional loci (represented by rs477992 and rs1374804) (Table 4). Regional association plots for these loci can be found in File S1. For both SNPs, the *P*-value for the genotypic model was slightly above the threshold (rs477992: $P = 6.43 \times 10^{-12}$; rs1374804: $P = 1.43 \times 10^{-11}$). The new locus rs1374804 could not be replicated in the TwinsUK study.

The results of the additive model on the identical data were part of a previous publication (Illig *et al.* 2010). Only the locus represented by rs477992 (our study)/rs541503 (previous publication) was not found in the screening with the genotypic model. The *P*-value for rs541503 (genotypic model) was slightly above the threshold ($P = 3.88 \times 10^{-11}$). Some loci identified in this study are represented by other SNPs than those reported in the previously published GWAS. Therefore, we summarized the results of the genotypic model for exactly the same SNPs reported in the published additive GWAS (Illig *et al.* 2010) (Table 3, Table 4). Apart from SNP rs541503, only one additional SNP, rs11158519, was not significantly associated with the previously reported metabolite or metabolite ratio. This locus

Table 1 Results for GWAS with genotypic model

					KORA sample		TwinsUK sample		
SNP	Metabolite (ratio)	Chr	Position	Gene	AF	*P*-value genotypic	AF	*P*-value genotypic	Replicated
rs7552404	C12/C10	1	75,908,534	*ACADM*	0.300	1.69×10^{-72}	0.314	1.89×10^{-29}	*
rs7601356	C9/PC.ae.C30.0	2	210,764,902	*ACADL*	0.632	1.24×10^{-70}	0.649	6.86×10^{-28}	*
rs715	Gly/Gln	2	211,251,300	*CPS1*	0.687	4.28×10^{-69}	0.703	1.12×10^{-48}	*
rs8396	C7.DC/C10	4	159,850,267	*ETFDH*	0.707	5.98×10^{-26}	0.678	3.14×10^{-17}	*
rs2046813	PC.ae.C42.5/PC.ae.C44.5	4	186,006,153	*ACSL1*	0.688	6.29×10^{-17}	0.687	1.18×10^{-3}	*
rs273913	C5/PC.ae.C34.1	5	131,689,055	*SLC22A4*	0.405	1.60×10^{-16}	0.351	4.19×10^{-2}	
rs3798719	PC.aa.C42.5/PC.aa.C40.3	6	11,144,811	*ELOVL2*	0.248	5.01×10^{-32}	0.234	4.01×10^{-4}	*
rs12356193	C0	10	61,083,359	*SLC16A9*	0.166	2.18×10^{-27}	0.161	1.20×10^{-7}	*
rs603424	C16.1/C14	10	102,065,469	*PKD2L1*	0.801	3.70×10^{-18}	0.818	1.99×10^{-2}	
rs174547	PC.aa.C36.3/PC.aa.C36.4	11	61,327,359	*FADS1*	0.701	2.29×10^{-208}	0.649	2.09×10^{-76}	*
rs2066938	C3/C4	12	119,644,998	*ACADS*	0.270	1.73×10^{-159}	0.257	2.17×10^{-67}	*
rs4902242	PC.aa.C28.1/PC.ae.C40.2	14	63,299,842	*SYNE2*	0.849	2.00×10^{-35}	0.872	4.78×10^{-15}	*
rs1077989	PC.ae.C32.1/PC.ae.C34.1	14	67,045,575	*PLEKHH1*	0.463	6.80×10^{-42}	0.472	4.05×10^{-18}	*
rs4814176	SM..OH..C24.1/SM..OH..C22.1	20	12,907,398	*SPTLC3*	0.364	2.69×10^{-31}	0.416	9.69×10^{-9}	*
rs6970485	lysoPC.a.C28.0/PC.aa.C26.0	7	11,752,704	*THSD7A*	0.354	1.21×10^{-47}	—	—	
rs1894832	Ser/Trp	7	56,144,740	*LOC389493*	0.508	1.98×10^{-12}	0.511	4.02×10^{-3}	
rs2657879	His/Gln	12	55,151,605	*GLS2*	0.207	2.89×10^{-14}	0.186	1.90×10^{-6}	*
rs7200543	PC.aa.C36.2/PC.aa.C38.3	16	15,037,471	*NTAN1*	0.312	7.45×10^{-16}	0.277	1.66×10^{-6}	*
rs1466448	SM.C18.1/SM.C16.1	19	8,195,519	*CERS4*	0.222	7.01×10^{-16}	0.194	3.90×10^{-10}	*
rs5746636	xLeu/Pro	22	17,276,301	*DGCR6*	0.236	2.98×10^{-20}	0.273	2.40×10^{-3}	*

This table reports the results for 20 loci that had a significant *P*-value for the genotypic model in KORA F4 ($P < 2.19 \times 10^{-12}$). Sixteen loci were replicated in the TwinsUK data ($P < 0.0025$) and are marked with an asterisk in the last column. Per locus, the SNP and metabolite or metabolite ratio with the lowest *P*-value in KORA is given. The table is divided into two parts: the upper part reports loci that were already found in the previously published GWAS on the same data (Illig *et al.* 2010); the lower part shows novel loci. AF, allele frequency of the effect allele; Chr, chromosome.

was identified in the genotypic screening with an association of an SNP 134 kb away from the GWAS SNP and another ratio of phosphatidylcholines. The remaining 13 SNPs showed significant *P*-values for both the genotypic model and the additive model (significance level: 2.19×10^{-12}). A possible explanation for the observed differences, despite data of the same individuals being used in both analyses, is that in the published GWAS a slightly different quality control of genotypes and metabolites was applied compared to the present analysis. For 18 of the 20 loci that were identified with both screening for additive and genotypic effects, the same SNP and metabolite (ratio) were found with the strongest association at this locus. Only for two loci was another SNP-metabolite combination observed for the additive screening. It should be noted that using an additive model, we were able to replicate the rs1894832 locus, which could not be replicated using the genotypic model.

The screening also was performed with the recessive and dominant models (File S3). Even with the use of a liberal threshold ($5 \times 10^{-8}/22{,}801$) instead of a strict one [$5 \times 10^{-8}/(22{,}801 \times 4)$], we identified no additional loci. From the 20 loci identified by the genotypic model, 14 were found with the recessive model and 18 with the dominant model. Use of the overdominant model identified 10 of the 20 described loci and one additional association between an SNP (rs219040) on the seventh chromosome ($P < 3.94 \times 10^{-13}$) and the ratio C5.1/C6.1. This SNP was located close to the gene *STEAP2-AS1* (non-protein-coding antisense RNA1 gene) that hardly relates to metabolism control. Its *P*-value of HWE was close to the threshold of quality exclusion ($P < 1.03 \times 10^{-5}$), and it could not be replicated ($P < 8.37 \times 10^{-1}$) in the TwinsUK data.

Novel loci with additive effects

We identified and replicated five novel loci (represented by SNPs rs1466448, rs7200543, rs2657879, rs5746636, and rs1894832) that were not reported in the published additive-effects GWAS (Illig *et al.* 2010). According to the LRT and AIC, the best model for these SNPs was additive in our analyses.

The locus including SNP rs1466448 is located in the region of the gene *CERS4*, which encodes the enzyme ceramide synthase, involved in the biosynthesis of ceramides [a simple form of sphingolipids consisting of sphingosine (or some of its derivatives) and fatty acid]. We observed an association with the ratio of sphingomyelins SM.C18.1 and SM.C16.1; this seems to be consistent with the gene function of *CERS4*. The locus including rs7200543 is located in the region of the gene *NTAN1*, encoding the N-terminal asparagine amidase protein. It is associated with the ratio of phosphatidylcholines PC.aa.C36.2 and PC.aa.C38.3. There is no direct connection between gene function and the associated trait. Two loci (rs2657879 and rs5746636) are settled in regions near genes involved in amino acid metabolism—*GLS2* (encoding glutaminase enzyme) and *PRODH* (encoding proline dehydrogenase), respectively. The functions of the identified genes were consistent with the associated traits: rs2657879 is associated with the ratio of histidine and glutamine, and rs5746636 is associated with the ratio of leucine/isoleucine to proline. The locus

Table 2 Results of analysis of nonadditive effects

				Genotypic model			Additive model		KORA study		TwinsUK study		
SNP	**Metabolite/ratio**	**A/B**	**AF**	**Beta(BB) (SE)**	**Beta(AB) (SE)**	***P*-value**	**Beta (SE)**	***P*-value**	**LRT**	**AIC**	**LRT**	**AIC**	**Replicated**
rs7552404	C12/C10	G/A	0.30/0.31	1.150(0.076)	0.642(0.046)	1.69×10^{-72}	0.600(0.033)	1.84×10^{-73}	a	a	a	a	*
rs7601356	C9/PC.ae.C30.0	C/T	0.63/0.65	−1.225(0.068)	−0.961(0.068)	1.24×10^{-70}	−0.523(0.032)	1.57×10^{-58}	g	g/d	g	g/a	*
rs715	Gly/Gln	C/T	0.69/0.70	−0.937(0.082)	−0.139(0.084)	4.28×10^{-69}	−0.590(0.036)	1.00×10^{-61}	r	g/r	r,a	r	*
rs8396	C7.DC/C10	C/T	0.71/0.68	−0.760(0.086)	−0.360(0.087)	5.98×10^{-26}	−0.388(0.036)	4.54×10^{-27}	a	a	r,a	g/r	*
rs2046813	PC.ae.C42.5/PC.ae.C44.5	C/T	0.69/0.69	0.647(0.084)	0.360(0.084)	6.29×10^{-17}	0.309(0.036)	7.11×10^{-18}	a	a	a,r,d	a	*
rs273913	C5/PC.ae.C34.1	T/C	0.41/0.35	0.606(0.071)	0.203(0.052)	1.60×10^{-16}	0.283(0.034)	1.28×10^{-16}	a	g/a	a,d,r,o	a	
rs3798719	PC.aa.C42.5/PC.aa.C40.3	T/C	0.25/0.23	−0.983(0.100)	−0.418(0.048)	5.01×10^{-32}	−0.453(0.038)	6.78×10^{-33}	a	a	a,d,r	a	*
rs12356193	C0	G/A	0.17/0.16	−1.145(0.167)	−0.483(0.053)	2.18×10^{-27}	−0.507(0.046)	2.42×10^{-28}	a	a	a,d	a	*
rs603424	C16.1/C14	A/G	0.80/0.82	0.872(0.121)	0.558(0.124)	3.70×10^{-18}	0.362(0.041)	1.42×10^{-18}	a	g/a	a,r,o,d	a	
rs174547	PC.aa.C36.3/PC.aa.C36.4	C/T	0.70/0.65	−1.872(0.068)	−1.019(0.069)	2.29×10^{-208}	−0.904(0.029)	5.98×10^{-209}	a	g/a	g	g/a	*
rs2066938	C3/C4	G/A	0.27/0.26	−1.942(0.077)	−0.649(0.043)	1.73×10^{-159}	−0.832(0.032)	7.49×10^{-149}	g	g/a	g	g/a	*
rs4902242	PC.aa.C28.1/PC.ae.C40.2	C/T	0.85/0.87	−1.019(0.181)	−0.356(0.189)	2.00×10^{-35}	−0.619(0.049)	3.77×10^{-36}	a,r	g/a	r,a	r	*
rs1077989	PC.ae.C32.1/PC.ae.C34.1	C/A	0.46/0.47	−0.888(0.066)	−0.529(0.056)	6.80×10^{-42}	−0.450(0.033)	1.99×10^{-42}	a	g/a	a,d	g/a	*
rs4814176	SM..OH..C24.1/SM..OH..C22.1	T/C	0.36/0.42	0.792(0.074)	0.422(0.050)	2.69×10^{-31}	0.403(0.034)	2.00×10^{-32}	a	a	d,a	d	*
rs6970485	lysoPC.a.C28.0/PC.aa.C26.0	C/T	0.35/-	0.980(0.107)	0.877(0.063)	1.21×10^{-47}	0.635(0.047)	4.66×10^{-41}	d	d	—	—	
rs1894832	Ser/Trp	C/T	0.51/0.51	0.491(0.067)	0.302(0.061)	1.98×10^{-12}	0.245(0.034)	4.02×10^{-13}	a	a	a,r,d	a	
rs2657879	His/Gln	G/A	0.21/0.19	0.747(0.128)	0.303(0.050)	2.89×10^{-14}	0.328(0.042)	4.44×10^{-15}	a	a	d,o,a	d	*
rs7200543	PC.aa.C36.2/PC.aa.C38.3	G/A	0.31/0.28	0.531(0.078)	0.328(0.049)	7.45×10^{-16}	0.287(0.035)	1.40×10^{-16}	a,d	a	a,d	a	*
rs1466448	SM.C18.1/SM.C16.1	C/A	0.22/0.19	−0.775(0.120)	−0.305(0.049)	7.01×10^{-16}	−0.337(0.041)	1.31×10^{-16}	a	a	a,d	a	*
rs5746636	xLeu/Pro	T/G	0.24/0.27	0.705(0.103)	0.379(0.049)	2.98×10^{-20}	0.366(0.039)	2.64×10^{-21}	a	a	a,r,d	a	*

The GWAS results for the genotypic and additive model for 20 identified loci in KORA are reported. Loci that could be replicated in the TwinsUK data are marked with an asterisk in the last column. In the LRT column, all 1-df models are given for which the genotypic model was not significantly better (in descending order of *P*-value). The genotypic model is indicated in this column if it led to a significantly better fit than all 1-df models. In the AIC column, the best model regarding AIC is named and separated by a slash from the best 1-df model if the best model is genotypic (r, recessive; a, additive; d, dominant; o, overdominant; g, genotypic). The table is divided into two parts: the upper part reports loci that were already found in the previously published GWAS on the same data (Illig *et al.* 2010); the lower part reports new loci. A, other allele; AF, allele frequency of the effect allele for KORA/for TwinsUK; B, effect allele; Beta(BB) (SE) and Beta(AB) (SE), estimated beta for BB and AB genotypes for genotypic model with standard error; Beta (SE), estimated beta for SNP with standard error.

Table 3 Results for SNPs reported in previous GWAS on the same data (Illig *et al.* 2010)

Best SNP reported in Illig *et al.* (2010)	Best metabolite/ratio reported in Illig *et al.* (2010)	Chr	Position	A/B	*P*-value genotypic	*P*-value additive	AF (B)	Gene	Best SNP of genotypic screening	Best metabolite/ratio of genotypic screening	Dist. (in bp)
rs211718	C12/C10	1	75,879,263	T/C	1.9×10^{-72}	1.6×10^{-73}	0.3	*ACADM*	rs7552404	C12/C10	29,271
rs541503	Orn/Ser	1	120,009,820	C/T	3.9×10^{-11}	5.0×10^{-11}	0.63	*PHGDH*	—	—	—
rs2286963	C9/C10.2	2	210,768,295	G/T	1.3×10^{-65}	5.8×10^{-60}	0.63	*ACADL*	rs7601356	C9/PC.ae.C30.0	3,393
rs2216405	Gly/PC.ae.C38.2	2	211,325,139	G/A	5.6×10^{-31}	3.3×10^{-31}	0.19	*CPS1*	rs715	Gly/Gln	73,839
rs8396	C14.1.OH/C10	4	159,850,267	C/T	3.4×10^{-23}	3.9×10^{-24}	0.71	*ETFDH*	rs8396	C7.DC/C10	0
rs2046813	PC.ae.C44.5/PC.ae.C42.5	4	186,006,153	C/T	8.3×10^{-17}	9.8×10^{-18}	0.69	*ACSL1*	rs2046813	PC.ae.C42.5/PC.ae.C44.5	0
rs272889	Val/C5	5	131,693,277	A/G	4.4×10^{-15}	5.1×10^{-16}	0.62	*SLC22A4*	rs273913	C5/PC.ae.C34.1	4,222
rs9393903	PC.aa.C40.3/PC.aa.C42.5	6	11,150,895	A/G	6.9×10^{-32}	1.0×10^{-32}	0.75	*ELOVL2*	rs3798719	PC.aa.C42.5/PC.aa.C40.3	6,084
rs7094971	C0	10	61,119,570	G/A	3.7×10^{-21}	3.2×10^{-22}	0.15	*SLC16A9*	rs12356193	C0	36,211
rs603424	C14/C16.1	10	102,065,469	A/G	5.4×10^{-18}	1.6×10^{-18}	0.8	*SCD*	rs603424	C16.1/C14	0
rs174547	PC.aa.C36.3/PC.aa.C36.4	11	61,327,359	C/T	6.3×10^{-209}	1.5×10^{-209}	0.7	*FADS1*	rs174547	PC.aa.C36.3/PC.aa.C36.4	0
rs2014355	C3/C4	12	119,659,907	C/T	8.9×10^{-121}	1.7×10^{-110}	0.73	*ACADS*	rs2066938	C3/C4	14,909
rs11158519	PC.ae.C38.1/PC.aa.C28.1	14	63,434,338	A/G	4.4×10^{-6}	$8.0 \times 10-7$	0.86	*SYNE2*	rs4902242	PC.aa.C28.1/PC.ae.C40.2	134,496
rs7156144	PC.ae.C32.1/PC.ae.C34.1	14	67,049,466	A/G	2.1×10^{-31}	4.8×10^{-32}	0.59	*PLEKHH1*	rs1077989	PC.ae.C32.1/PC.ae.C34.1	3,891
rs168622	SM..OH..C24.1/SM.C16.0	20	12,914,089	T/G	3.5×10^{-21}	3.1×10^{-22}	0.36	*SPTLC3*	rs4814176	SM..OH..C24.1/SM..OH..C22.1	6,691

This table summarizes the results for the 15 previously reported SNPs and corresponding traits (Illig *et al.* 2010). Results are given for the additive and genotypic models at these SNPs in our analyses. Moreover, per locus, the best SNP and metabolite or metabolite ratio identified with the genotypic screening is given along with the distance to the SNP reported in Illig *et al.* (2010). One locus was not found with the genotypic screening. A, other allele; AF, allele frequency of the effect allele; B, effect allele; Chr, chromosome.

including rs1894832, which was replicated only for the additive model, is located near the genes *PSPH* and *PHKG1* (which code for phosphoserine phosphatase and phosphorylase kinase, respectively). We cannot say which gene is causal for this association without additional analysis, but both genes are likely to be connected with the associated trait (serine and tryptophan ratio).

Loci with nonadditive effects

We identified two loci (rs2066938 and rs7601356) for which the genotypic model fit significantly better than all 1-df tests. The box plots for these two loci (Figure 2A) show that the underlying genetic models seem to be between the additive and dominant/recessive models. Assuming the mean metabolite concentration of all samples with the increasing homozygous genotype as 100% and the mean of all samples with decreasing homozygous genotype as 0%, the mean of all samples with the heterozygous genotype is expected to be approximately 50% for the additive model and 100 or 0% for the recessive/dominant model. For the two mentioned SNPs, rs2066938 and rs7601356, we observed 67 and 22% for the heterozygous genotype, respectively. We made a similar observation at loci for which the genotypic model fit significantly better not for one but for two 1-df models regarding LRT (rs4902242 and rs7200543). The percentages of the heterozygous genotype were 61 and 62%, respectively (Figure 2B). We also found one locus for which the genotypic model had the best AIC followed by the AIC for the recessive model (rs715), but the LRT showed that the genotypic model was not significantly better than the recessive model. The percentage of heterozygotes was 82% (Figure 2C)—less than 100%. For SNP rs6970485 with the dominant model fitting best (according to AIC and LRT), the box plot is shown in Figure 2D.

Discussion

In this study we systematically analyzed nonadditive genetic effects on human serum metabolites using data from two independent studies—KORA F4 and TwinsUK. We observed that most of the identified loci had additive effects on the associated metabolites or metabolite ratios. This validates the standard approach of many GWAS assuming additive effects. However, we found evidence of a nonadditive effect for four loci: either the genotypic model was better than all 1-df models or the genotypic model was not better than another 1-df model but was better than the additive model. This provided refined knowledge about the associated loci and additional insight into the genetic determination of metabolites.

Additive genetic effects on metabolites

Most of the loci identified in our analysis showed additive effects on the associated metabolites. Only four nonadditive genetic effects on metabolites were identified. Our findings are in line with theories about the prevalence of additivity

Table 4 Results for GWAS with additive model

				KORA sample			TwinsUK sample		gene info	
SNP	Metabolite (ratio)	Chr	Position	A/B	AF	*P*-value additive	AF	*P*-value additive	Gene	Gene pos.
rs7552404	C12/C10	1	75,908,534	G/A	0.300	1.84×10^{-73}	0.314	1.25×10^{-30}	ACADM	Downstream*
3223rs477992	Ser/Orn	1	120,059,099	A/G	0.702	5.75×10^{-13}	0.658	8.11×10^{-9}	*PHGDH*	inGene*
rs2286963	C9/PC.ae.C34.0	2	210,768,295	G/T	0.629	1.68×10^{-60}	0.651	2.14×10^{-31}	*ACADL*	inGene*
rs715	Gly/Gln	2	211,251,300	C/T	0.687	1.00×10^{-61}	0.703	4.37×10^{-49}	*CPS1*	Upstream*
rs8396	C7.DC/C10	4	159,850,267	C/T	0.707	4.54×10^{-27}	0.678	2.72×10^{-17}	*ETFDH*	inGene*
rs2046813	PC.ae.C42.5/PC.ae.C44.5	4	186,006,153	C/T	0.688	7.11×10^{-18}	0.687	2.71×10^{-4}	*ACSL1*	Upstream*
rs270613	C5/Val	5	131,668,482	A/G	0.606	8.03×10^{-17}	0.647	2.57×10^{-3}	*SLC22A4*	inGene
rs3798719	PC.aa.C42.5/PC.aa.C40.3	6	11,144,811	T/C	0.248	6.78×10^{-33}	0.234	1.32×10^{-4}	*ELOVL2*	inGene*
rs12356193	C0	10	61,083,359	G/A	0.166	2.42×10^{-28}	0.161	1.70×10^{-8}	*SLC16A9*	inGene*
rs603424	C16.1/C14	10	102,065,469	A/G	0.801	1.42×10^{-18}	0.818	5.36×10^{-3}	*PKD2L1*	inGene
rs174547	PC.aa.C36.3/PC.aa.C36.4	11	61,327,359	C/T	0.701	5.98×10^{-209}	0.649	1.00×10^{-74}	*FADS1*	inGene*
rs2066938	C3/C4	12	119,644,998	G/A	0.270	7.49×10^{-149}	0.257	1.59×10^{-63}	*ACADS*	Downstream*
rs4902242	PC.aa.C28.1/PC.ae.C40.2	14	63,299,842	C/T	0.849	3.77×10^{-36}	0.872	1.30×10^{-15}	*SYNE2*	Upstream*
rs1077989	PC.ae.C32.1/PC.ae.C34.1	14	67,045,575	C/A	0.463	1.99×10^{-42}	0.472	5.18×10^{-18}	*PLEKHH1*	Downstream*
rs4814176	SM..OH..C24.1/SM..OH..C22.1	20	12,907,398	T/C	0.364	2.00×10^{-32}	0.416	7.67×10^{-8}	*SPTLC3*	Downstream*
rs1374804	Ser/Gly	3	127,391,188	A/G	0.638	1.70×10^{-12}	0.619	1.03×10^{-1}	*ALDH1L1*	Upstream
rs6970485	lysoPC.a.C28.0/PC.aa.C26.0	7	11,752,704	C/T	0.354	4.66×10^{-41}	—	—	*THSD7A*	inGene
rs1894832	Ser/Trp	7	56,144,740	C/T	0.508	4.02×10^{-13}	0.511	8.92×10^{-4}	*LOC389493*	Downstream*
rs2657879	His/Gln	12	55,151,605	G/A	0.207	4.44×10^{-15}	0.186	1.57×10^{-6}	*GLS2*	inGene*
rs7200543	PC.aa.C36.2/PC.aa.C38.3	16	15,037,471	G/A	0.312	1.40×10^{-16}	0.277	2.55×10^{-7}	*NTAN1*	Downstream*
rs1466448	SM.C18.1/SM.C16.1	19	8,195,519	C/A	0.222	1.31×10^{-16}	0.194	5.12×10^{-11}	*CERS4*	inGene*
rs5746636	xLeu/Pro	22	17,276,301	T/G	0.236	2.64×10^{-21}	0.273	8.94×10^{-4}	*DGCR6*	inGene*

This table reports the results for 22 loci that had a significant *P*-value in KORA ($p < 2.19 \times 10^{-12}$) for the additive model. Eighteen loci were replicated in the TwinsUK data ($P < 0.0023$). Per locus, the SNP and metabolite or metabolite ratio with the lowest *P*-value in KORA is given. The table is divided into two parts: the upper part consists of loci that were already found in Illig *et al.* (2010); the lower part shows new loci. SNPs that were replicated are marked by an asterisk. AF, allele frequency of the effect allele; B, effect allele; Chr, chromosome.

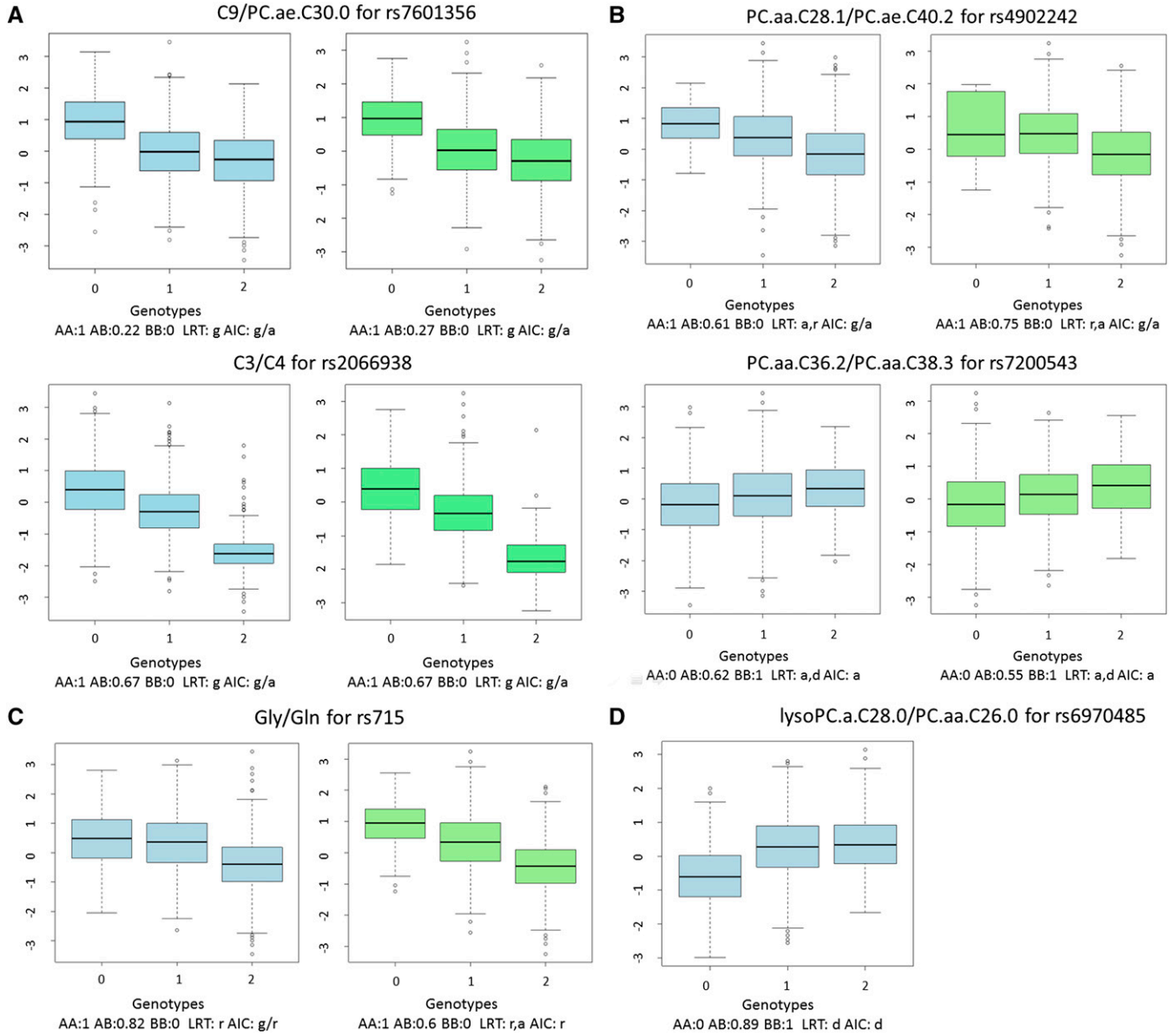


Figure 2 Box plots for loci with nonadditive effects. These box plots show the distribution of phenotypes among the different genotypes. The bold line is the median of the trait's per-genotype group. Box plots for KORA data are printed in blue, for TwinsUK data in green. Additional information is provided below each plot. First, the mean of the metabolite (ratio) at the heterozygous genotype is expressed as a percentage if the mean at the homozygous increasing allele genotype is taken as 100% and the mean at the homozygous decreasing allele genotype is taken as 0%. Second, information about the best model concerning the LRT and AIC is given (r, recessive; a, additive; d, dominant; o, overdominant; g, genotypic). For the LRT, all 1-df models are given for which the genotypic model did not fit significantly better (in descending order of *P*-value). If the genotypic model led to a significantly better fit than all 1-df models according to the LRT, this is indicated by "g." Regarding the AIC, the best model is named and separated by a slash from the best 1-df model if the best model is genotypic. (A) Box plots for rs7601356 and rs2066938. For these loci, the best model was genotypic. (B) Box plots for rs4902242 and rs7200543. For these loci, the genotypic model did not fit significantly better than two 1-df models regarding the LRT. (C) Box plot for rs715. For this locus, the best model was recessive. (D) Box plot for rs6970485. For this locus, the best model was dominant. This SNP was not available in the TwinsUK data.

(most of the trait variance is indeed additive), as described in the introduction (Hill *et al.* 2008; Powell *et al.* 2013). Based on previous publications, the detection of mainly additive genetic effects could be the result of the strong disease model distortion to additive effects resulting from LD between the causal variant and the tagged marker (Zondervan and Cardon 2004; Vukcevic *et al.* 2011). It was shown that LD between causal and tagged marker loci also distorts the observed genetic model because the power to detect such departures drops off very quickly with decreasing LD. We conducted simulations to elucidate the influence of such genotypic noise (LD with the casual variant) and phenotypic noise (the bias of measurements etc.) and obtained similar results (File S2). Therefore, we can expect larger

nonadditive effects (even if most of the trait variance is additive) in studies using genetic data with higher density, *e.g.*, sequencing data. It should be noted that all inferences were made under the assumption that only one causal gene is in LD with the studied marker allele. Situations in which there are several nonadditive QTLs at one locus are not considered here. Another point that we would like to mention is that the number of samples that carry each of the three possible genotypes of an SNP determines the potential to detect nonadditivity. A low number of samples with one particular genotype would lower the power to detect a potential nonadditive effect. To account for the extreme cases, we have excluded SNPs that showed one of the three possible genotypes in only a few samples (<30), but SNPs above this threshold might have low power as well.

Furthermore, we observed that even SNPs with nonadditive effects could be detected with the use of an additive model. The additive model can be seen as an approximation to the recessive and dominant model, which works well when the recessive allele is prevalent. In contrast, not all loci that were detected with the additive model could be identified with the genotypic model. This emphasizes even more that the general assumption of additive effects in the analysis of metabolite concentrations and ratios is acceptable and leads to good results.

Additionally, there are alternative methods for the analysis of genetic effects if the mode of inheritance is unknown, for example, the maximin efficiency robust test (MERT) (Gastwirth 1985) and the maximum (MAX) test (Davies 1987). They were previously compared with each other as well as with the genotypic test (Zheng *et al.* 2006; Loley *et al.* 2013). All three strategies gain in power under different conditions. We decided to use the genotypic test for several reasons—availability of fast software, easy validation, minimal time of calculation among the three strategies and no restriction regarding allele frequency (Loley *et al.* 2013), and applicability for quantitative traits and imputed data. The genotypic test is also robust if the underlying model is overdominant (Loley *et al.* 2013; Tsepilov *et al.* 2013).

Nonadditive effects

If not only the association per se but also the underlying genetic model is of interest, studying nonadditive effects can provide deeper insight. In our analysis, we found two loci with evidence for the genetic model being dominant (rs6970485) or recessive (rs715). The knowledge of these canonical biologic models helps us to understand the genetic control of these loci more deeply. The fact that the genotypic model showed the best result for two other loci (rs2066938 and rs7601356) could be caused by different reasons. There could be some technical impacts introduced by genotype or phenotype measurements or even the result of model distortion by LD mentioned earlier. We have demonstrated the robustness of nonadditive models for several types of phenotypic transformations—different combinations of covariate correction, log-transformation, and Gaussianization (File S1).

The genotypic model also could represent the effect of hidden nonlinear interactions or—following the Kacser and Burns theory (Kacser and Burns 1981; Phadnis and Fry 2005)—could be the consequence of a reciprocal instead of linear response due to mutual influence in embedded systems.

Genotypic model as screening step: advantages and disadvantages

Our proposed strategy to identify nonadditive effects using a genotypic model for screening could be transferred to other analyses. It is especially promising to use the genotypic model instead of an additive model when an essential part of gene effects is known to be nonadditive (Zheng *et al.* 2006; Loley *et al.* 2013). In comparison with the best 1-df model, the genotypic model may have a reduced power to detect associated loci because the association test involves 2-df. Still, it can be used as a screening tool that reduces the amount of multiple testing and is more efficient in terms of computing time than calculating each 1-df model separately on a genome-wide scale (File S3). For high-dimensional data such as metabolomics, the latter is an important consideration.

Conclusion

In conclusion, we confirmed the practice of using the additive model for analyzing SNP effects on metabolites. In our study, most of the associations were indeed additive, and even if another genetic model approximated the effect better, the association could be detected by an additive model. Our study brings some light into the discussion of prevalence of nonadditive effects in the genetic control of metabolites.

Acknowledgments

We thank Werner Römisch-Margl, Julia Scarpa, and Katharina Sckell for metabolomics measurements performed at the Helmholtz Centrum München, Genome Analysis Center, Metabolomics Core Facility. The KORA study was initiated and financed by the Helmholtz Zentrum München–German Research Center for Environmental Health, which is funded by the German Federal Ministry of Education and Research and by the State of Bavaria. Furthermore, KORA research was supported within the Munich Center of Health Sciences, Ludwig-Maximilians-Universität, as part of LMUinnovativ. The TwinsUK study was funded by the Wellcome Trust and the European Community's Seventh Framework Programme (FP7/2007–2013). The study also receives support from the National Institute for Health Research BioResource Clinical Research Facility and Biomedical Research Centre based at Guy's and St. Thomas' NHS Foundation Trust and King's College London. T.D.S. is holder of an ERC Advanced Principal Investigator award. SNP genotyping was performed by the Wellcome Trust Sanger Institute and National Eye Insti-

tute via the National Institutes for Health/Center for Inherited Disease Research. Moreover, the research leading to these results received funding by a grant from the Helmholtz Association (Helmholtz–Russia Joint Research Group 310) and from the European Union's Seventh Framework Programme (FP7-Health-F5-2012) under grant agreement 305280 (MIMOmics). The work of S.-Y.S. was supported by a Post-Doctoral Research Fellowship from the Oak Foundation. The work of Y.S.A. was supported by a grant from the Russian Foundation for Science (RSCF 14-14-00313). The work of Y.A.T. was supported by a grant from the Russian Foundation for Basic Research–Helmholtz Joint Research Group (12-04-91322). The work of J.A. was supported in part by a grant from the German Federal Ministry of Education and Research to the German Center Diabetes Research. The research of N.S. was supported by the Wellcome Trust (grants WT098051 and WT091310), the EU FP7 (EPIGENESYS grant 257082 and BLUEPRINT grant HEALTH-F5-2011-282510).

Author contributions: C.G., J.A., N.S., T.D.S., K.S., G.K., R.W.-S., and Y.S.A conceived and designed the experiments. R.W.-S., T.D.S., C.P., J.A., and G.K. performed the experiments. Y.A.T., S.-Y.S., and J.S.R. analyzed the data. C.G., C.P., J.A., N.S., T.D.S., K.S., G.K., R.W.-S., and Y.S.A. contributed reagents, materials, and analysis tools. Y.A.T., J.S.R., K.S., Y.S.A., and C.G. wrote the manuscript.

Literature Cited

Akaike, H., 1974 A new look at the statistical model identification. IEEE Trans. Automat. Contr. 19: 716–723.

Andrew, T., D. J. Hart, H. Snieder, M. de Lange, T. D. Spector *et al.*, 2001 Are twins and singletons comparable? A study of disease-related and lifestyle characteristics in adult women. Twin Res. 4: 464–477.

Aulchenko, Y. S., S. Ripke, A. Isaacs, and C. M. van Duijn, 2007 GenABEL: an R library for genome-wide association analysis. Bioinformatics 23: 1294–1296.

Beasley, T. M., S. Erickson, and D. B. Allison, 2009 Rank-based inverse normal transformations are increasingly used, but are they merited? Behav. Genet. 39: 580–595.

Davies, R. B., 1987 Hypothesis testing when a nuisance parameter is present only under the alternative. Biometrika 74: 33–43.

Demirkan, A., C. M. van, Duijn, P. Ugocsai, A. Isaacs, P. P. Pramstaller *et al.*, 2012 Genome-wide association study identifies novel loci associated with circulating phospho- and sphingolipid concentrations. PLoS Genet. 8: e1002490.

Fisher, R., 1928a The possible modification of the response of the wild type to recurrent mutations. Am. Nat. 62: 115–126.

Fisher, R., 1928b Two further notes on the origin of dominance. Am. Nat. 62: 571–574.

Gastwirth, J. L., 1985 The use of maximin efficiency robust tests in combining contingency tables and survival analysis. J. Am. Stat. Assoc. 80: 380–384.

Gieger, C., L. Geistlinger, E. Altmaier, M. Hrabé de Angelis, F. Kronenberg *et al.*, 2008 Genetics meets metabolomics: a genome-wide association study of metabolite profiles in human serum. PLoS Genet. 4: e1000282.

Haldane, J. B. S., 1930 A note on Fisher's theory of the origin of dominance and a correlation between dominance and linkage. Am. Nat. 64: 87–90.

Hicks, A. A., P. P. Pramstaller, A. Johansson, V. Vitart, I. Rudan *et al.*, 2009 Genetic determinants of circulating sphingolipid concentrations in European populations. PLoS Genet. 5: e1000672.

Hill, W. G., M. E. Goddard, and P. M. Visscher, 2008 Data and theory point to mainly additive genetic variance for complex traits. PLoS Genet. 4: e1000008.

Huelsenbeck, J. P., and K. A. Crandall, 1997 Phylogeny estimation and hypothesis testing using maximum likelihood. Annu. Rev. Ecol. Syst. 28: 437–466.

Illig, T., C. Gieger, G. Zhai, W. Römisch-Margl, R. Wang-Sattler *et al.*, 2010 A genome-wide perspective of genetic variation in human metabolism. Nat. Genet. 42: 137–141.

Kacser, H., and J. A. Burns, 1981 The molecular basis of dominance. Genetics 97: 639–666.

Kettunen, J., T. Tukiainen, A.-P. Sarin, A. Ortega-Alonso, E. Tikkanen *et al.*, 2012 Genome-wide association study identifies multiple loci influencing human serum metabolite levels. Nat. Genet. 44: 269–276.

Kolz, M., T. Johnson, S. Sanna, A. Teumer, V. Vitart *et al.*, 2009 Meta-analysis of 28,141 individuals identifies common variants within five new loci that influence uric acid concentrations. PLoS Genet. 5: e1000504.

Loley, C., I. R. König, L. A. Hothorn, and A. Ziegler, 2013 A unifying framework for robust association testing, estimation, and genetic model selection using the generalized linear model. Eur. J. Hum. Genet. 21: 1442–1448.

Menni, C., G. Zhai, A. MacGregor, C. Prehn, W. Römisch-Margl *et al.*, 2013 Targeted metabolomics profiles are strongly correlated with nutritional patterns in women. Metabolomics 9: 506–514.

Moayyeri, A., C. J. Hammond, D. J. Hart, and T. D. Spector, 2012 The UK Adult Twin Registry (TwinsUK Resource). Twin Res. Hum. Genet. 16: 144–149.

Nicholson, G., M. Rantalainen, J. V. Li, A. D. Maher, D. Malmodin *et al.*, 2011 A genome-wide metabolic QTL analysis in Europeans implicates two loci shaped by recent positive selection. PLoS Genet. 7: e1002270.

Orr, H. A., 1991 A test of Fisher's theory of dominance. Proc. Natl. Acad. Sci. USA 88: 11413–11415.

Phadnis, N., and J. D. Fry, 2005 Widespread correlations between dominance and homozygous effects of mutations: implications for theories of dominance. Genetics 171: 385–392.

Porteous, J. W., 1996 Dominance—one hundred and fifteen years after Mendel's paper. J. Theor. Biol. 182: 223–232.

Powell, J. E., A. K. Henders, A. F. McRae, J. Kim, G. Hemani *et al.*, 2013 Congruence of additive and non-additive effects on gene expression estimated from pedigree and SNP data. PLoS Genet. 9: e1003502.

Ried, J. S., A. Döring, K. Oexle, C. Meisinger, J. Winkelmann *et al.*, 2012 PSEA: Phenotype Set Enrichment Analysis—a new method for analysis of multiple phenotypes. Genet. Epidemiol. 36: 244–252.

Römisch-Margl, W., C. Prehn, R. Bogumil, C. Röhring, K. Suhre *et al.*, 2012 Procedure for tissue sample preparation and metabolite extraction for high-throughput targeted metabolomics. Metabolomics 8: 133–142.

Shin, S.-Y., E. B. Fauman, A.-K. Petersen, J. Krumsiek, R. Santos *et al.*, 2014 An atlas of genetic influences on human blood metabolites. Nat. Genet. 46: 543–550

Suhre, K., and C. Gieger, 2012 Genetic variation in metabolic phenotypes: study designs and applications. Nat. Rev. Genet. 13: 759–769.

Suhre, K., H. Wallaschofski, J. Raffler, N. Friedrich, R. Haring *et al.*, 2011 A genome-wide association study of metabolic traits in human urine. Nat. Genet. 43: 565–569.

Tanaka, T., J. Shen, G. R. Abecasis, A. Kisialiou, J. M. Ordovas *et al.*, 2009 Genome-wide association study of plasma polyunsaturated fatty acids in the InCHIANTI Study. PLoS Genet. 5: e1000338.

Tsepilov, Y. A., J. S. Ried, K. Strauch, H. Grallert, C. M. van Duijn *et al.*, 2013 Development and application of genomic control methods for genome-wide association studies using non-additive models. PLoS ONE 8: e81431.

Vukcevic, D., E. Hechter, C. Spencer, and P. Donnelly, 2011 Disease model distortion in association studies. Genet. Epidemiol. 290: 278–290.

Wichmann, H.-E., C. Gieger, and T. Illig, 2005 KORA-gen—resource for population genetics, controls and a broad spectrum of disease phenotypes. Gesundheitswesen 67(Suppl. 1): 26–30.

Wright, S., 1929 Fisher's theory of dominance. Am. Nat. 63: 274– 279.

Zheng G., B. Freidlin, and J. L. Gastwirth, 2007 Comparison of robust tests for genetic association using case-control studies, pp. 253–265 in *Optimality: The Second Erich L. Lehmann Symposium*, edited by Javier Rojo (IMS Lecture Notes–Monograph Series Vol. 49). Institute of Mathematical Statistics, Beachwood, OH.

Zondervan, K. T., and L. R. Cardon, 2004 The complex interplay among factors that influence allelic association. Nat. Rev. Genet. 5: 89–100.

Communicating editor: N. R. Wray

GENETICS

Supporting Information
www.genetics.org/lookup/suppl/doi:10.1534/genetics.115.175760/-/DC1

Nonadditive Effects of Genes in Human Metabolomics

Yakov A. Tsepilov, So-Youn Shin, Nicole Soranzo, Tim D. Spector, Cornelia Prehn, Jerzy Adamski, Gabi Kastenmüller, Rui Wang-Sattler, Konstantin Strauch, Christian Gieger, Yurii S. Aulchenko, and Janina S. Ried

DOI: 10.1534/genetics.115.175760

File S1

Table S1 - Analysis results of effects for significant loci obtained for different phenotype transformations.

Supporting Figures S1– Regional association plots for genotypic model for KORA data.

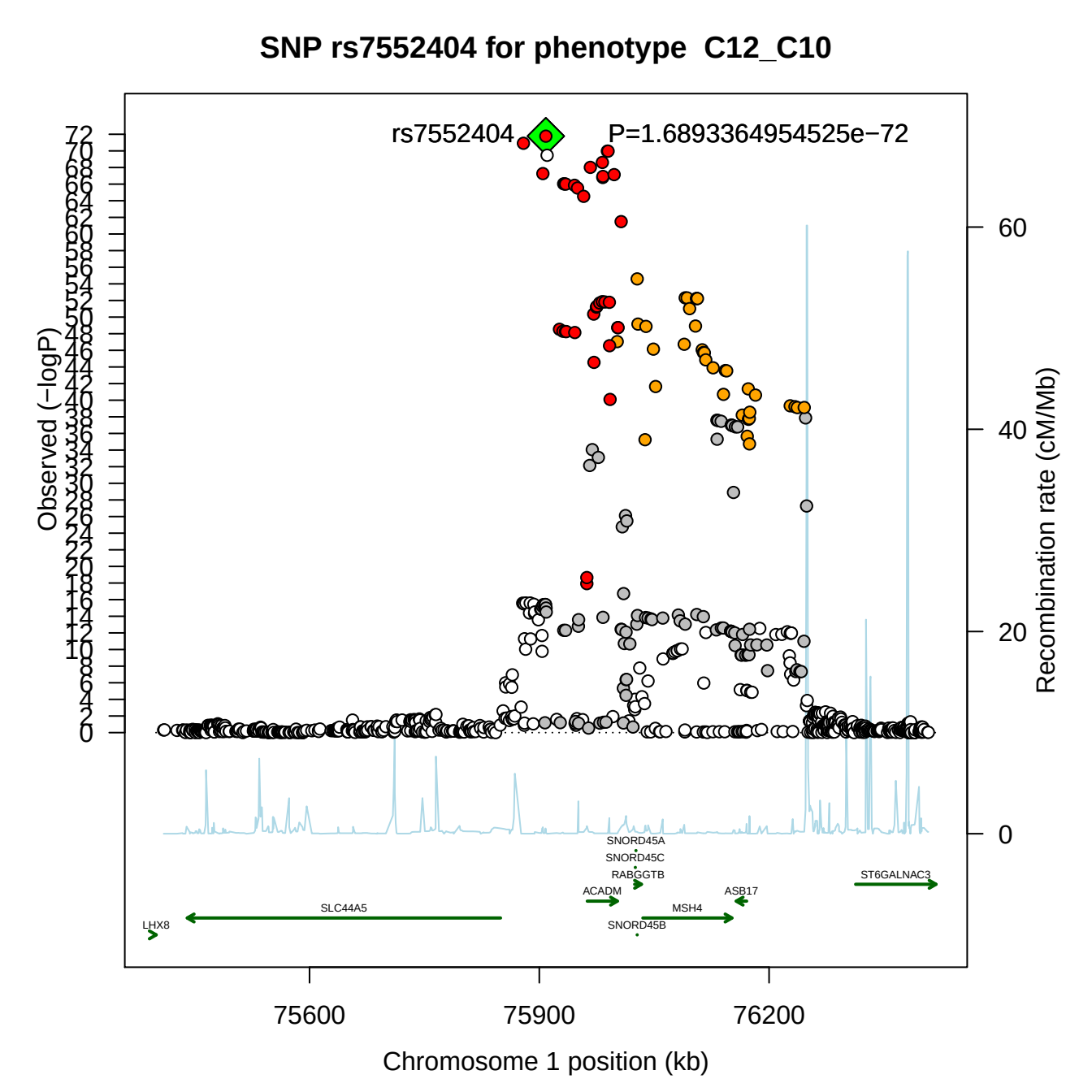
Supporting Figures S2 – Regional association plots for additive model for KORA data.

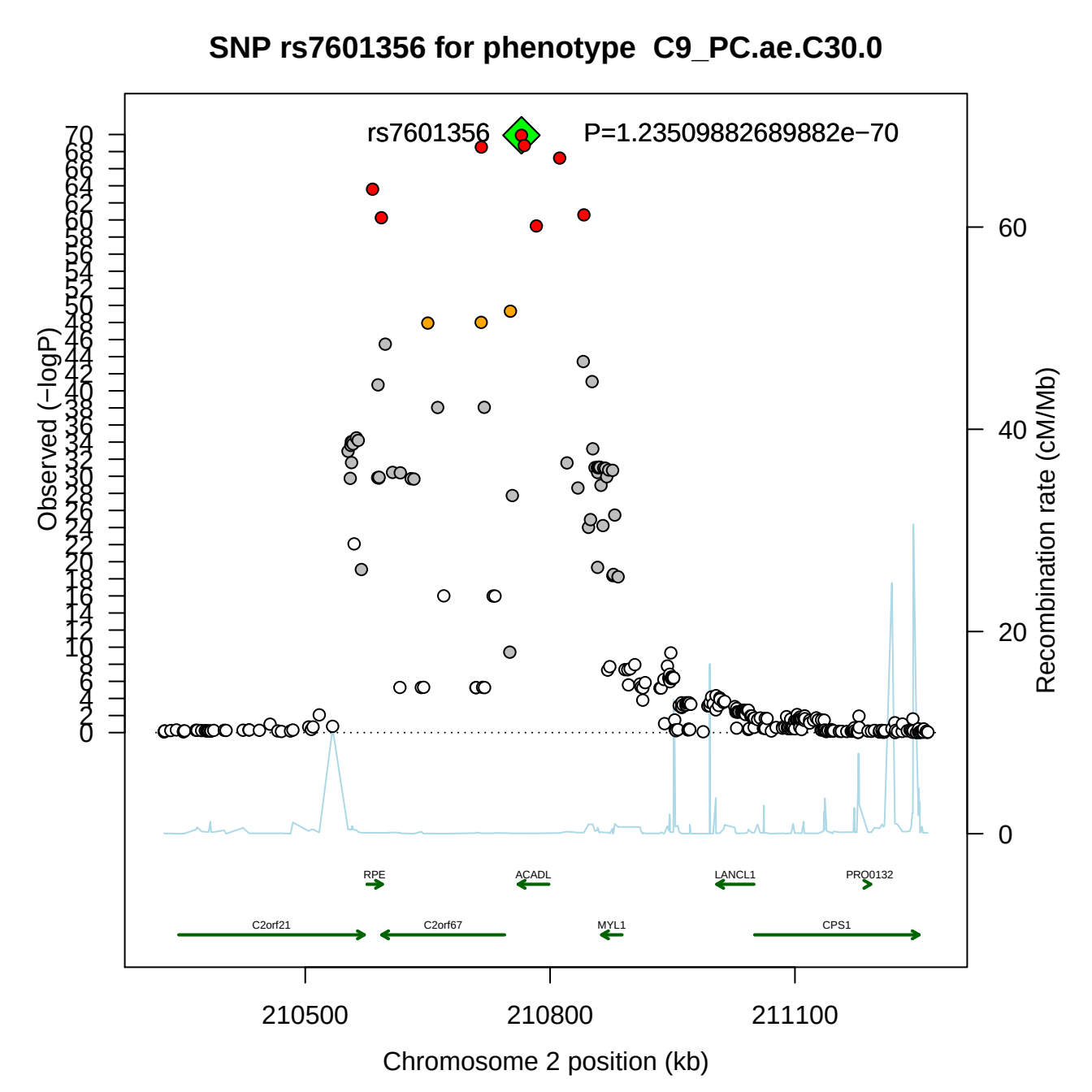
Table S1. Results of analysis of non-additive effects for diffrent phenotype transformations. The GWAS results for the genotypic and additive model for twenty identified loci in KORA are reported. Four transformations were used. Each letter represents the consecutive order of transformation: **G** – gaussinisation, **R** – correction for age, batch and sex (residuals), **L** – logariphmization (thus, for example, **LR** – consecutive transformation using logariphmization and then correction for age, batch and sex; in our study **RG** transformation was used). For each transformation the results for LRT and AIC tests are reported. In the LRT column all 1 df models are given for which the genotypic model was not significantly better (in descending order of p-value). In the AIC column the best model regarding AIC is named and separated by a slash from the best 1 df model, if the best model is genotypic. r, a, d, o, g – recessive, additive, dominant, over-dominant, genotypic, respectively. The table is divided into the two parts: the upper part reports loci that were already found in [3] the previously published GWAS on the same data, the lower part reports new loci.

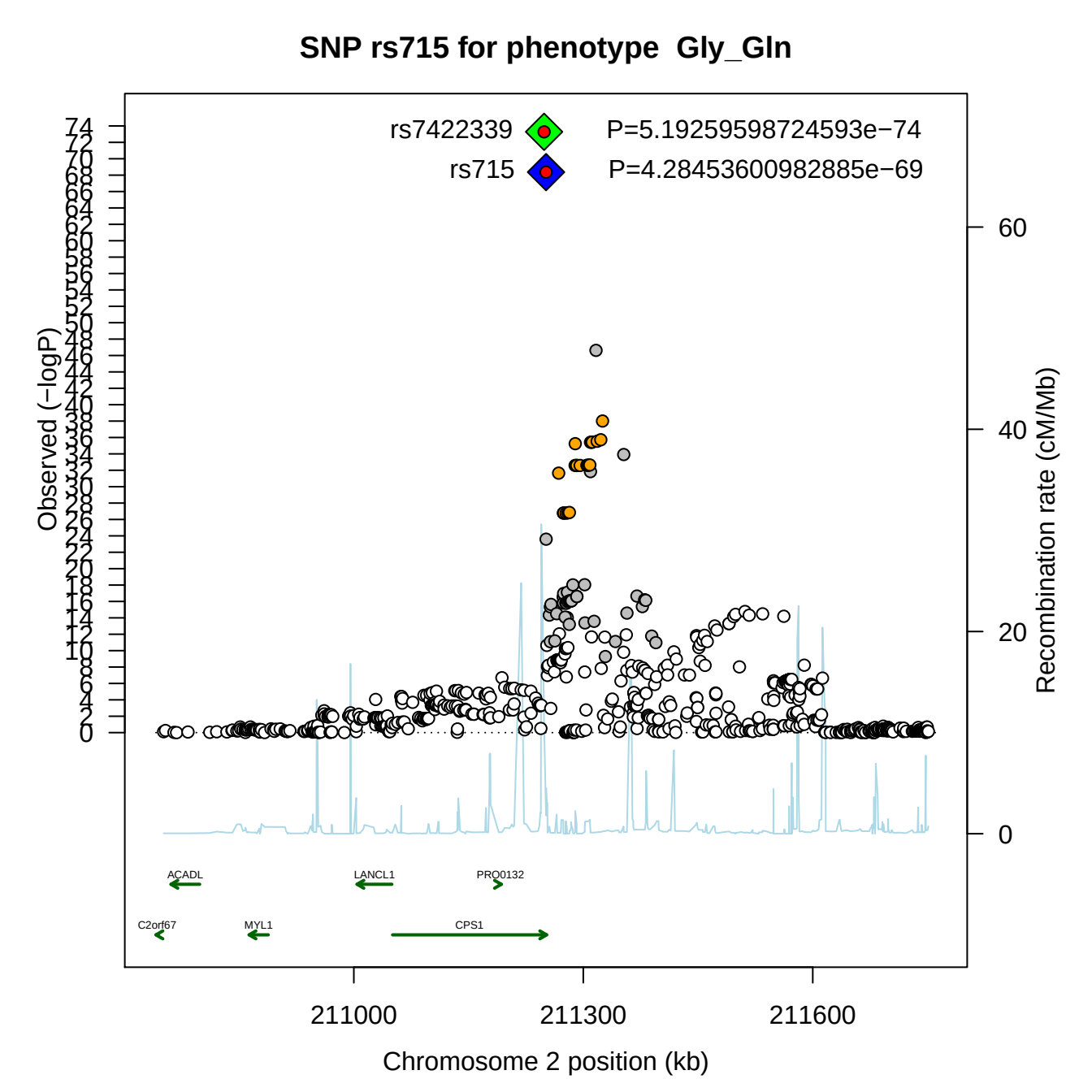
				GR		**LR**		**LRG**		**R**		**RG**	
SNP	**metabolite (ratio)**	**chr**	**position**	**LRT**	**AIC**	**LRT**	**AIC**	**LRT**	**AIC**	**LRT**	**AIC**	**LRT**	**AIC**
rs7552404	C12/C10	1	75,908,534	a	a	a	g/a	a	a	a	a	a	a
rs7601356	C9/PC.ae.C30.0	2	210,764,902	g	g/d	g	g/d	g	g/d	g	g/d	g	g/d
rs715	Gly/Gln	2	211,251,300	r	g/r	r	g/r	r	g/r	r	g/r	r	g/r
rs8396	C7.DC/C10	4	159,850,267	a	a	a	a	a	a	a	a	a	a
rs2046813	PC.ae.C42.5/PC.ae.C44.5	4	186,006,153	a	a	a	a	a	a	a	a	a	a
rs273913	C5/PC.ae.C34.1	5	131,689,055	a	g/a	a	g/a	a	g/a	a	g/a	a	g/a
rs3798719	PC.aa.C42.5/PC.aa.C40.3	6	11,144,811	a	a	a	a	a	a	a	a	a	a
rs12356193	C0	10	61,083,359	a	a	a	a	a	a	a	a	a	a
rs603424	C16.1/C14	10	102,065,469	a	g/a	a	g/a	a	g/a	a	g/a	a	g/a
rs174547	PC.aa.C36.3/PC.aa.C36.4	11	61,327,359	a	g/a	a	g/a	a	g/a	g	g/a	a	g/a
rs2066938	C3/C4	12	119,644,998	g	g/a	g	g/a	g	g/a	g	g/a	g	g/a

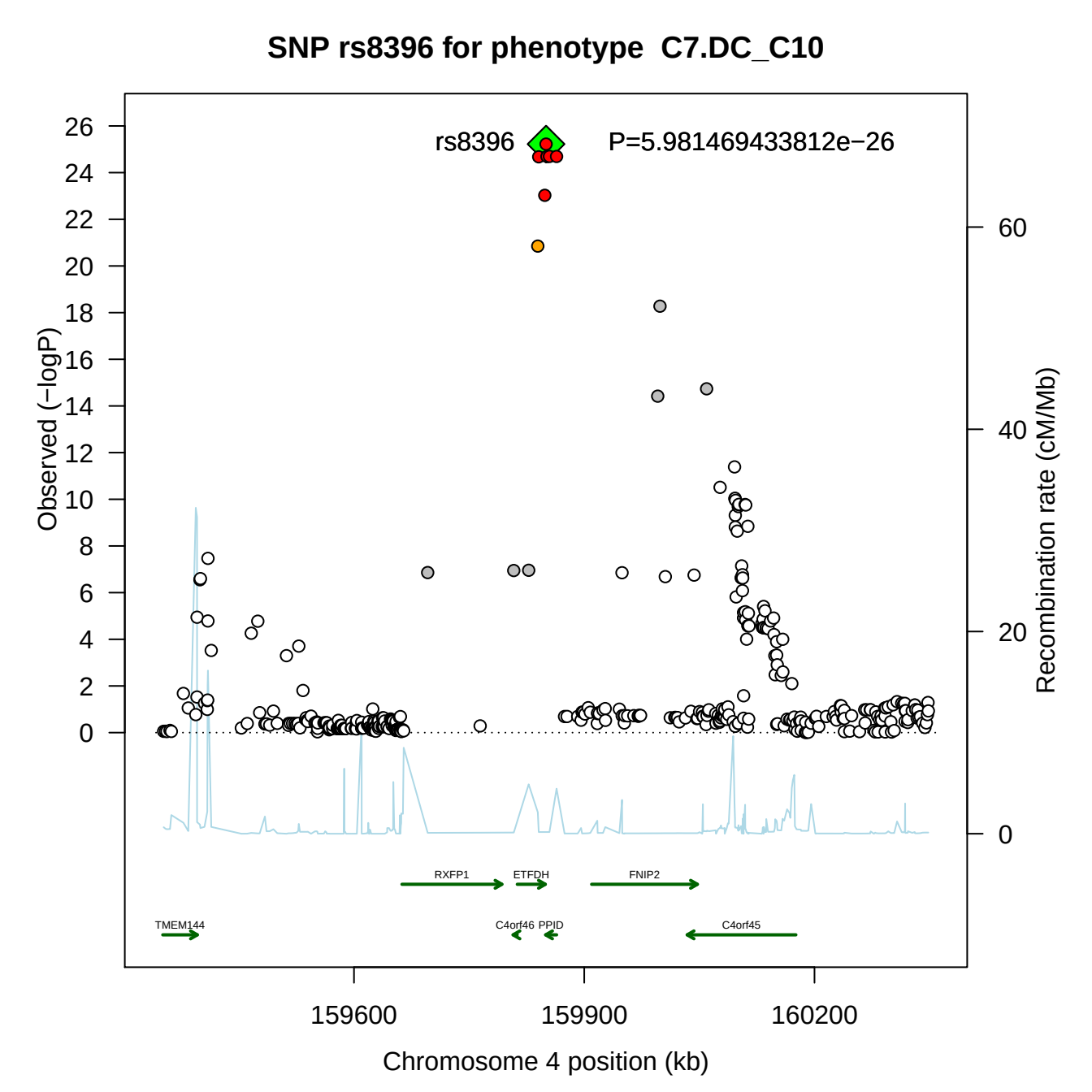
rs4902242	PC.aa.C28.1/PC.ae.C40.2	14	63,299,842	a,r	g/a	a,r	g/a	a,r	g/a	a,r	a	a,r	g/a
rs1077989	PC.ae.C32.1/PC.ae.C34.1	14	67,045,575	a	g/a	a	g/a	a	g/a	a	g/a	a	g/a
rs4814176	SM..OH..C24.1/SM..OH..C22.1	20	12,907,398	a	a	a	a	a	a	a	a	a	a
rs6970485	lysoPC.a.C28.0/PC.aa.C26.0	7	11,752,704	d	d	d	d	d	d	d	g/d	d	d
rs1894832	Ser/Trp	7	56,144,740	a	a	a	a	a	a	a	a	a	a
rs2657879	His/Gln	12	55,151,605	a	a	a	a	a	a	a	a	a	a
rs7200543	PC.aa.C36.2/PC.aa.C38.3	16	15,037,471	a,d	a	a,d	a	a,d	a	a,d	g/a	a,d	a
rs1466448	SM.C18.1/SM.C16.1	19	8,195,519	a	a	a	a	a	a	a	a	a	a
rs5746636	xLeu/Pro	22	17,276,301	a,d	a	a,d	a	a,d	a	a	a	a	a

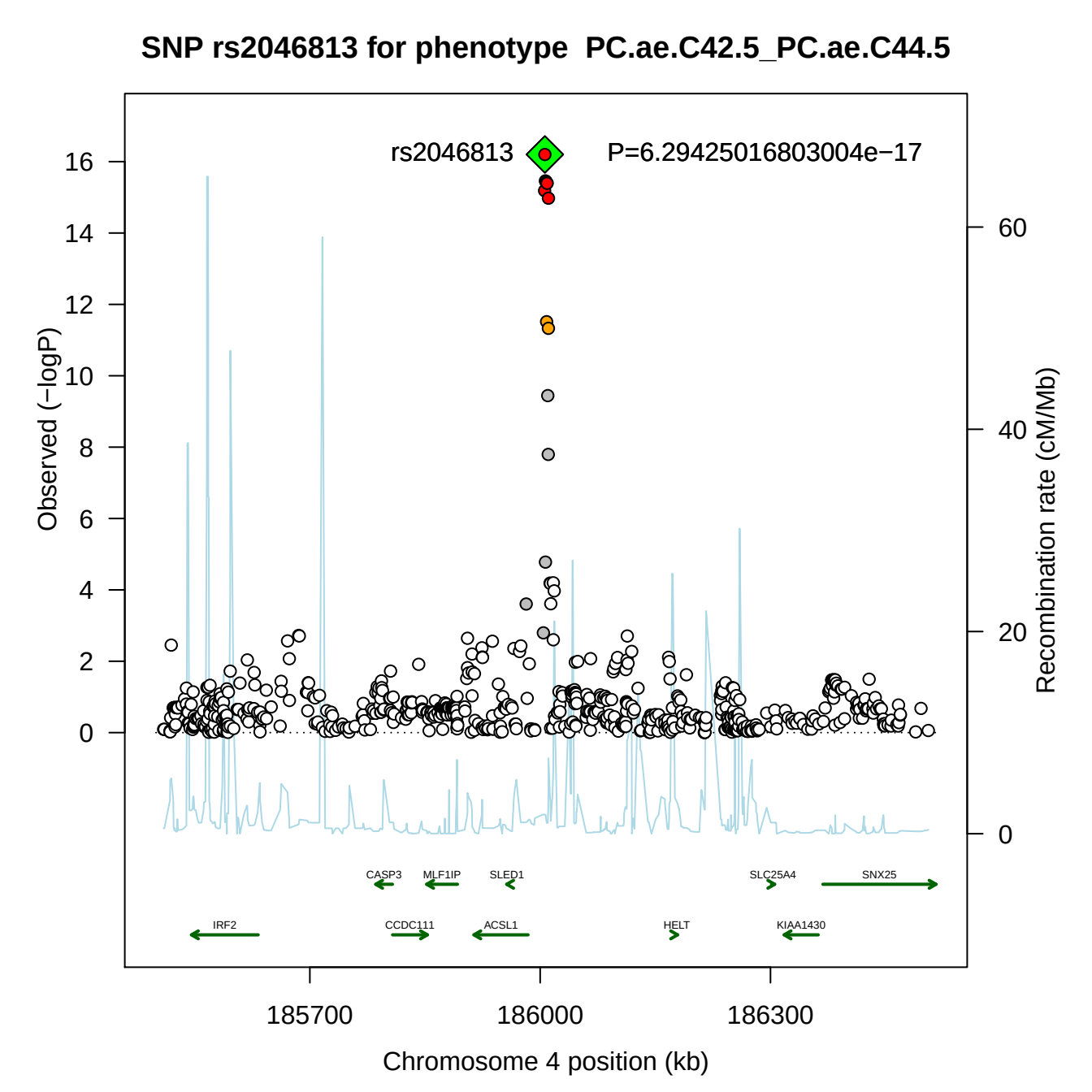
Supporting Figures S1. Regional association plots for genotypic model for KORA data.

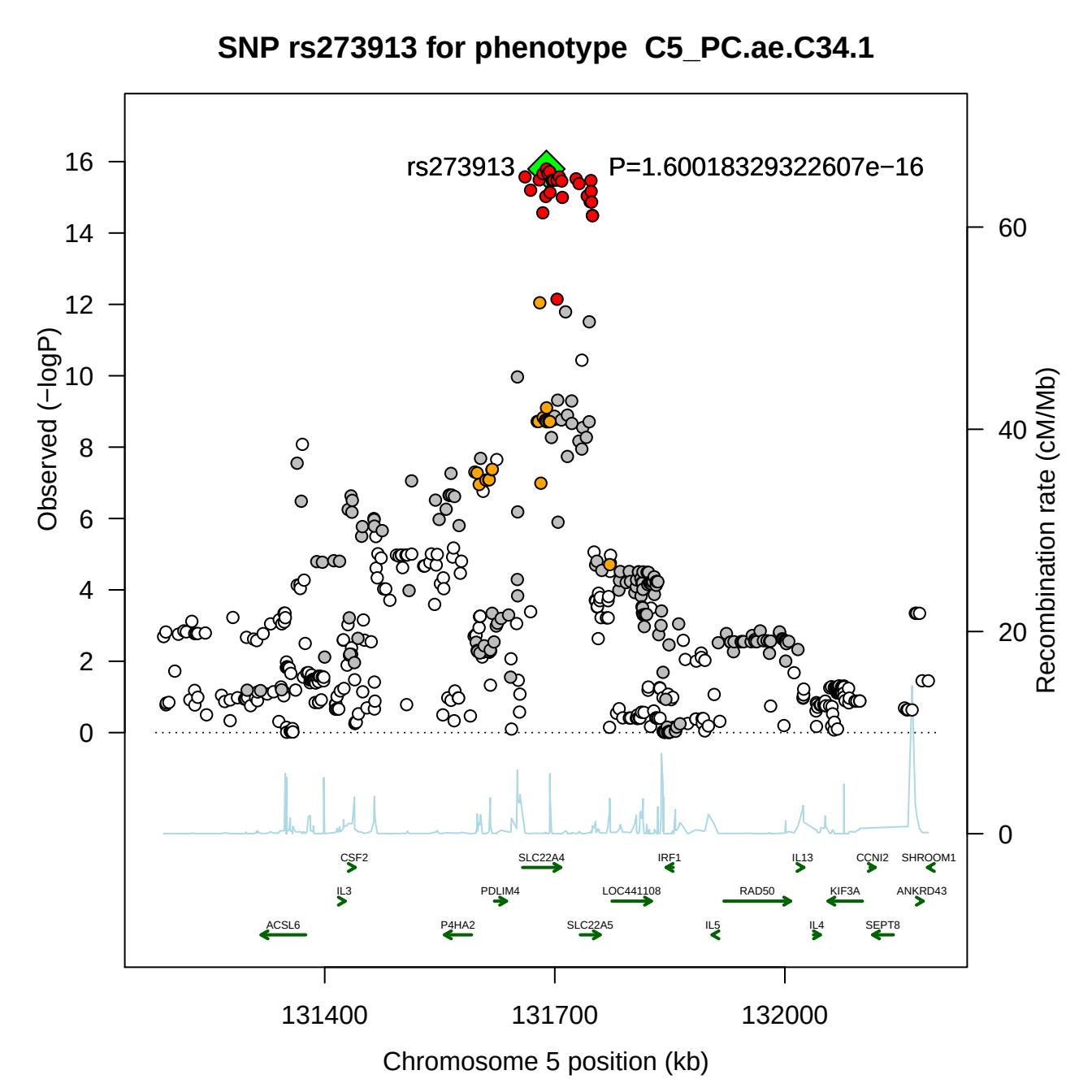

SNP rs7552404 for phenotype C12_C10
rs7552404
P=1.6893364954525e−72
Observed (−logP)
Recombination rate (cM/Mb)
Chromosome 1 position (kb)
75600
75900
76200
SNORD45A
SNORD45C
RABGGTB
ACADM
ASB17
ST6GALNAC3
SLC44A5
MSH4
LHX8
SNORD45B

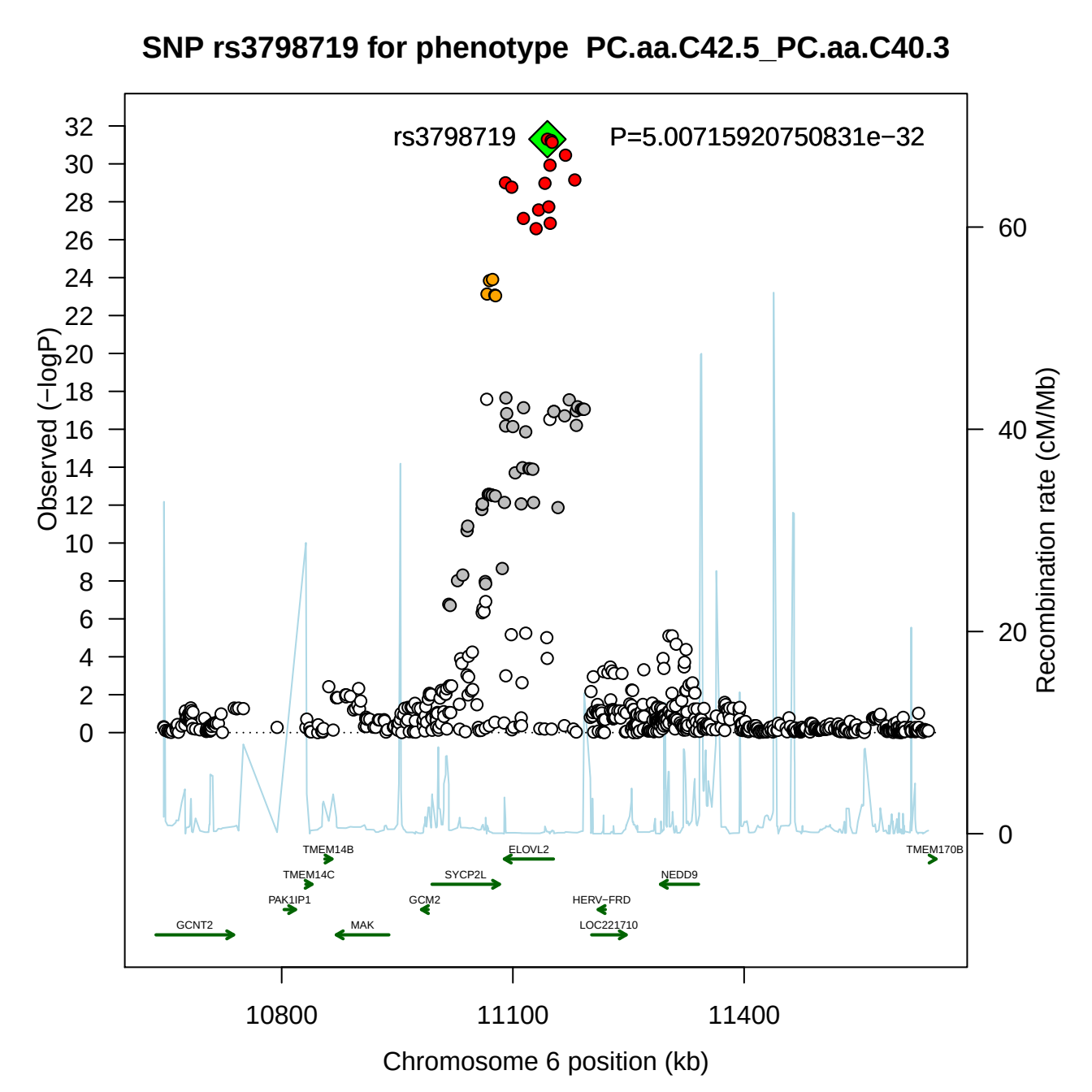

SNP rs7601356 for phenotype C9_PC.ae.C30.0
rs7601356
P=1.23509882689882e−70
Observed (−logP)
Recombination rate (cM/Mb)
RPE
ACADL
LANCL1
PRO0132
C2orf21
C2orf67
MYL1
CPS1
210500
210800
211100
Chromosome 2 position (kb)

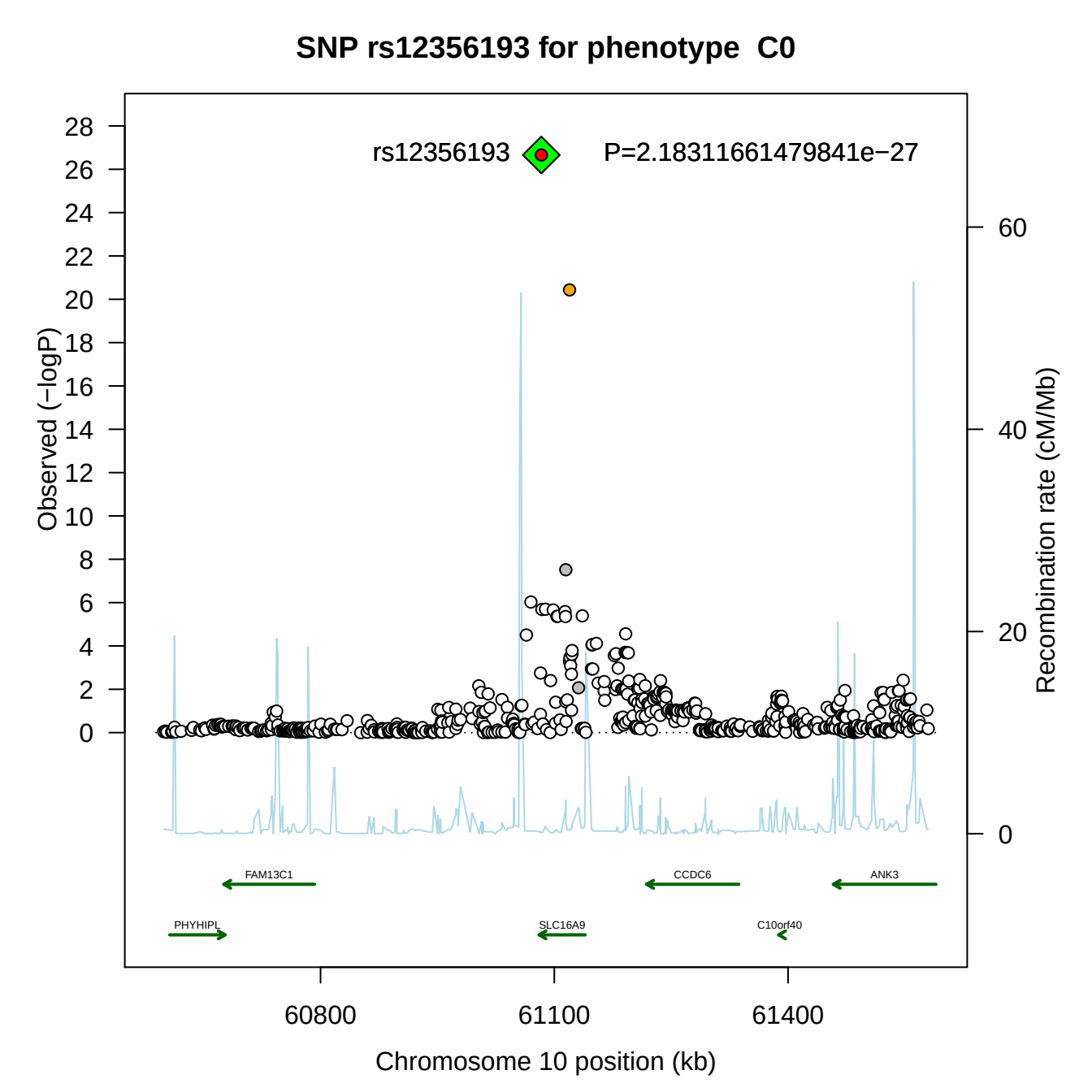

SNP rs715 for phenotype Gly_Gln
rs7422339
P=5.19259598724593e−74
rs715
P=4.28453600982885e−69
Observed (−logP)
Recombination rate (cM/Mb)
ACADL
LANCL1
PRO0132
C2orf67
MYL1
CPS1
211000
211300
211600
Chromosome 2 position (kb)

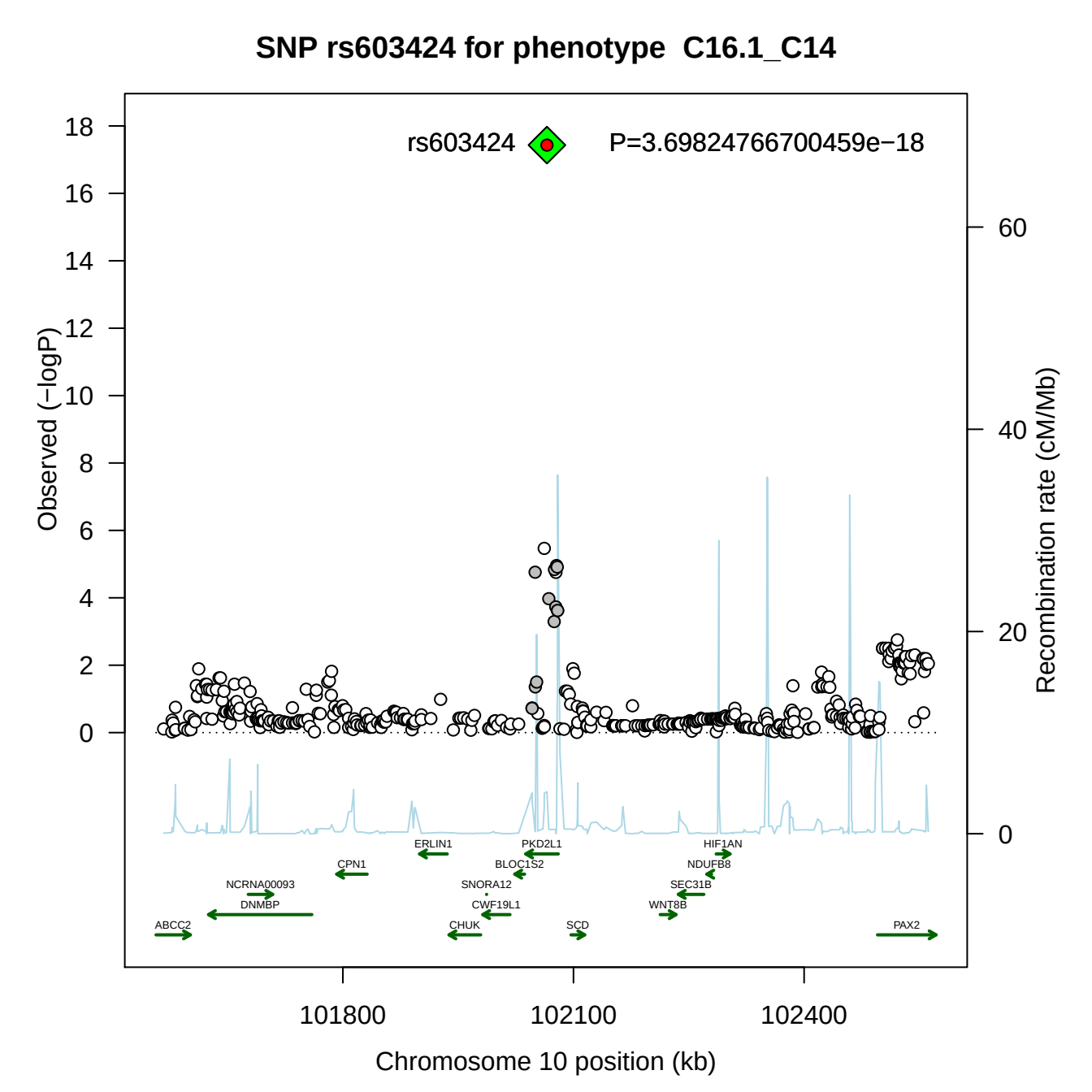

SNP rs8396 for phenotype C7.DC_C10
rs8396
P=5.981469433812e−26
Observed (−logP)
Recombination rate (cM/Mb)
Chromosome 4 position (kb)
159600
159900
160200
RXFP1
ETFDH
FNIP2
TMEM144
C4orf46
PPID
C4orf45

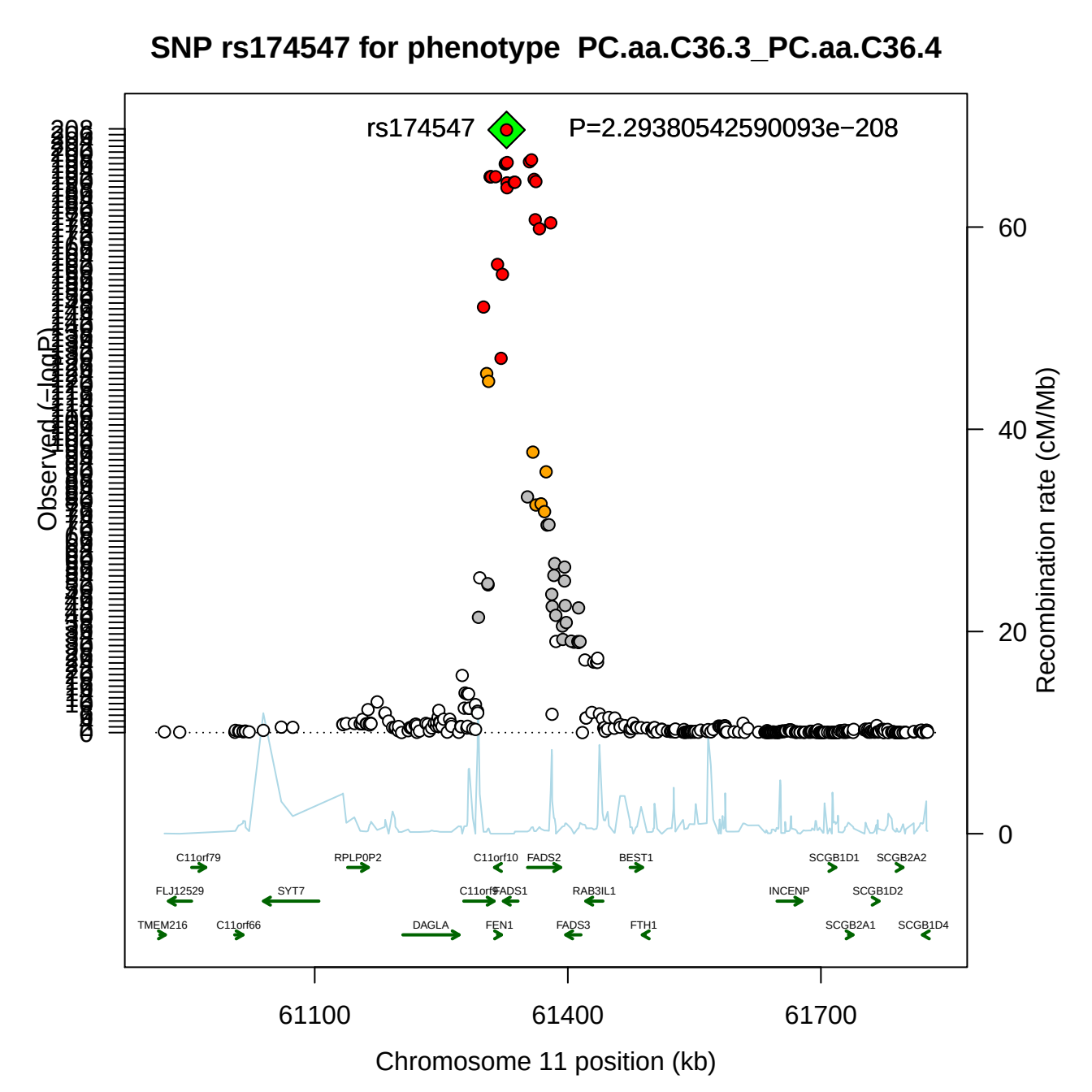

SNP rs2046813 for phenotype PC.ae.C42.5_PC.ae.C44.5
rs2046813
P=6.29425016803004e−17
Observed (−logP)
Recombination rate (cM/Mb)
0
2
4
6
8
10
12
14
16
0
20
40
60
CASP3
MLF1IP
SLED1
SLC25A4
SNX25
IRF2
CCDC111
ACSL1
HELT
KIAA1430
185700
186000
186300
Chromosome 4 position (kb)

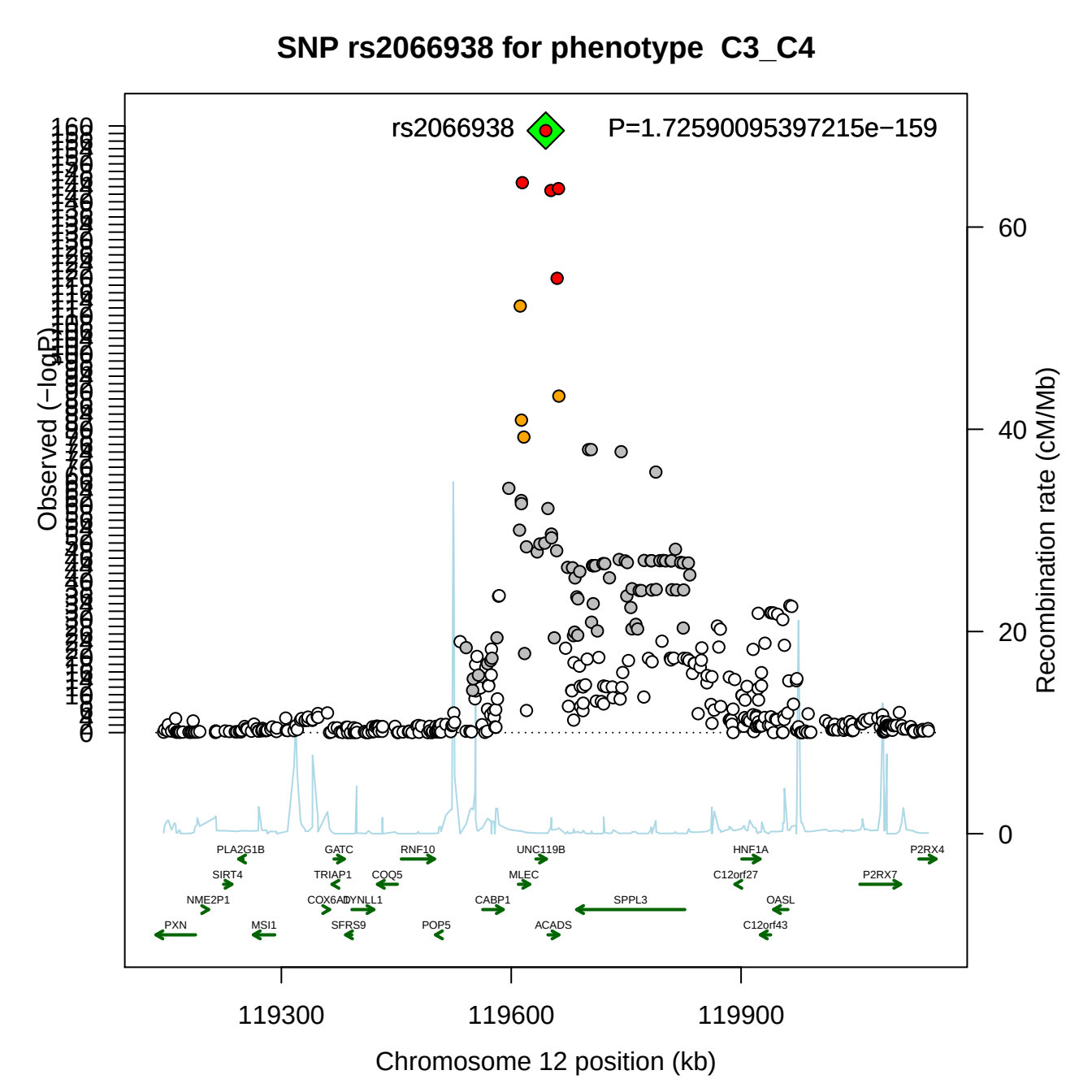

SNP rs273913 for phenotype C5_PC.ae.C34.1
rs273913
P=1.60018329322607e−16
Observed (−logP)
Recombination rate (cM/Mb)
Chromosome 5 position (kb)
131400
131700
132000
CSF2
SLC22A4
IRF1
IL13
CCNI2
SHROOM1
IL3
PDLIM4
LOC441108
RAD50
KIF3A
ANKRD43
ACSL6
P4HA2
SLC22A5
IL5
IL4
SEPT8

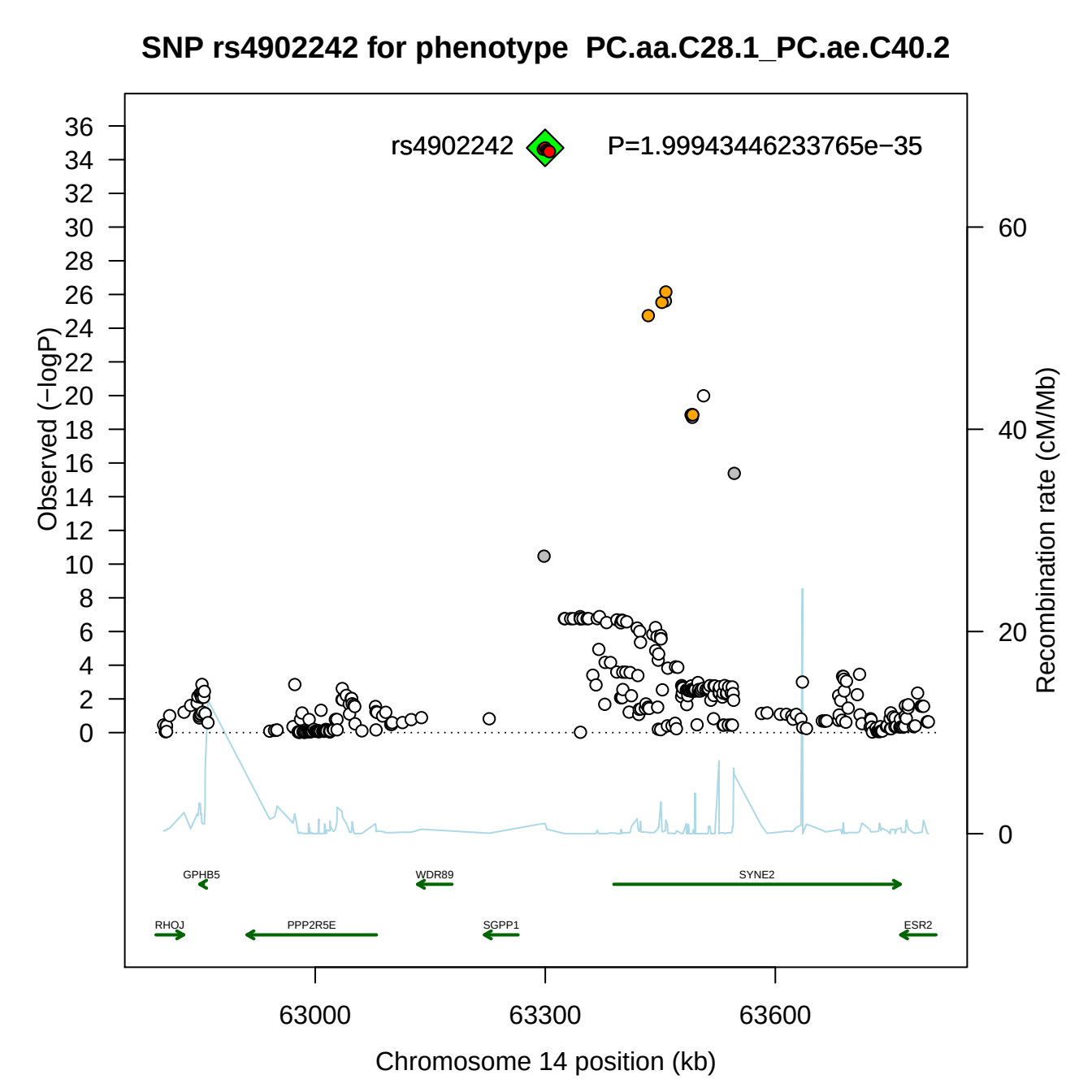

SNP rs3798719 for phenotype PC.aa.C42.5_PC.aa.C40.3
rs3798719
P=5.00715920750831e−32
Observed (−logP)
Recombination rate (cM/Mb)
Chromosome 6 position (kb)
10800
11100
11400
TMEM14B
TMEM14C
PAK1IP1
GCNT2
MAK
GCM2
SYCP2L
ELOVL2
HERV−FRD
LOC221710
NEDD9
TMEM170B

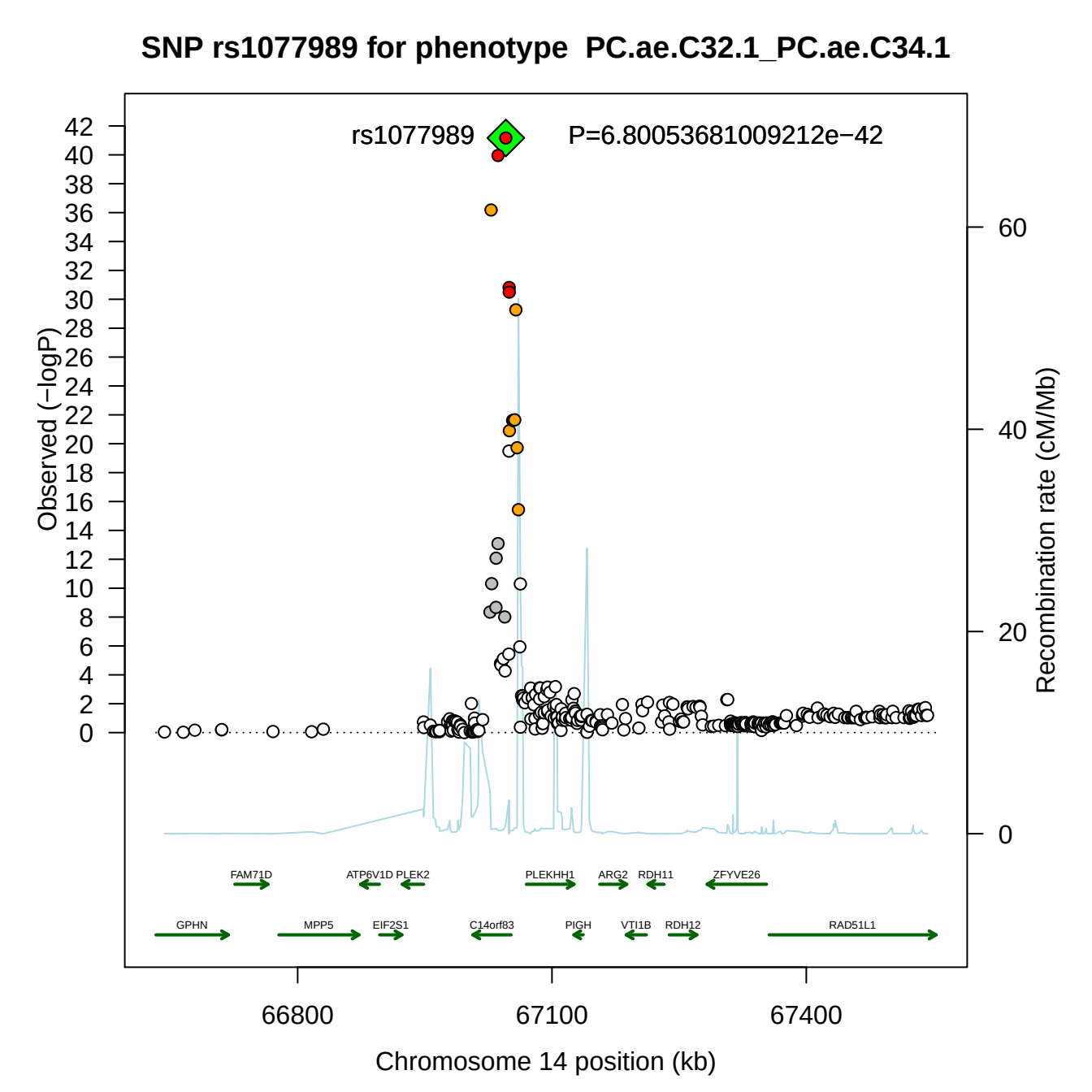

SNP rs12356193 for phenotype C0
rs12356193
P=2.18311661479841e−27
Observed (−logP)
Recombination rate (cM/Mb)
0
2
4
6
8
10
12
14
16
18
20
22
24
26
28
0
20
40
60
FAM13C1
CCDC6
ANK3
PHYHIPL
SLC16A9
C10orf40
60800
61100
61400
Chromosome 10 position (kb)

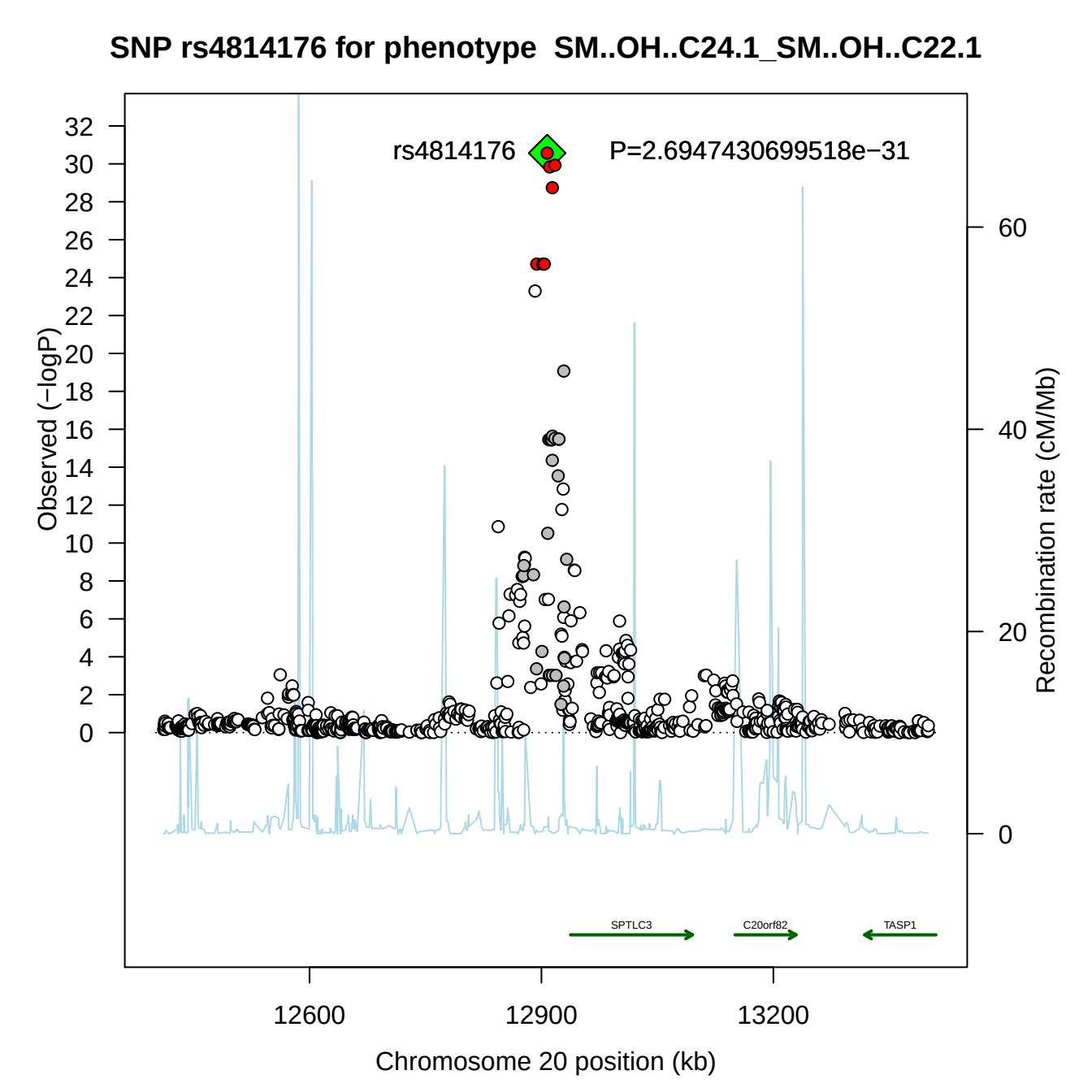

SNP rs603424 for phenotype C16.1_C14
rs603424
P=3.69824766700459e−18
Observed (−logP)
Recombination rate (cM/Mb)
Chromosome 10 position (kb)
101800
102100
102400
ERLIN1
PKD2L1
HIF1AN
CPN1
BLOC1S2
NDUFB8
NCRNA00093
SNORA12
SEC31B
DNMBP
CWF19L1
WNT8B
ABCC2
CHUK
SCD
PAX2

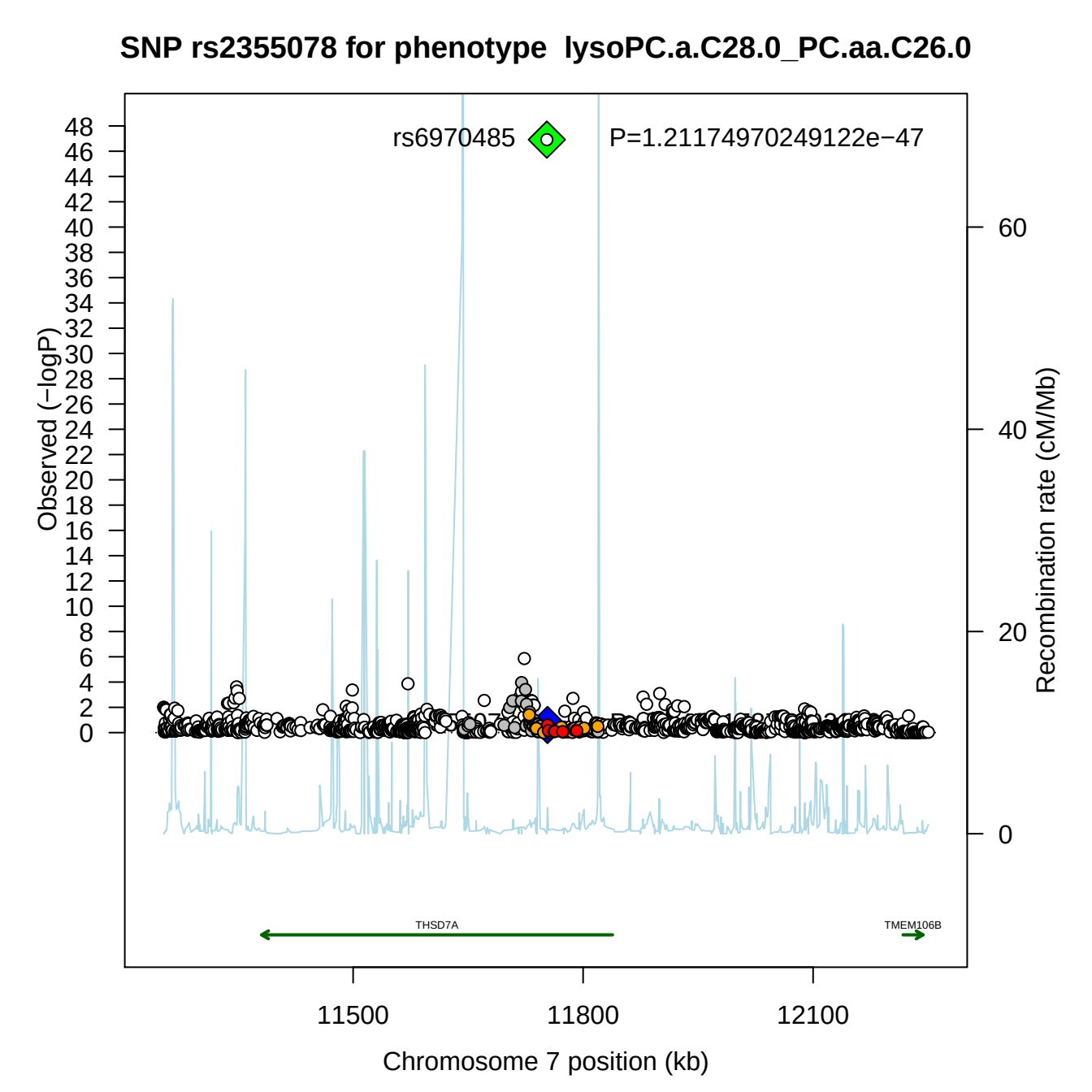

SNP rs174547 for phenotype PC.aa.C36.3_PC.aa.C36.4
rs174547
P=2.29380542590093e−208
Observed (−logP)
Recombination rate (cM/Mb)
60
40
20
0
C11orf79
RPLP0P2
C11orf10
FADS2
BEST1
SCGB1D1
SCGB2A2
FLJ12529
SYT7
C11orf9
FADS1
RAB3IL1
INCENP
SCGB1D2
TMEM216
C11orf66
DAGLA
FEN1
FADS3
FTH1
SCGB2A1
SCGB1D4
61100
61400
61700
Chromosome 11 position (kb)

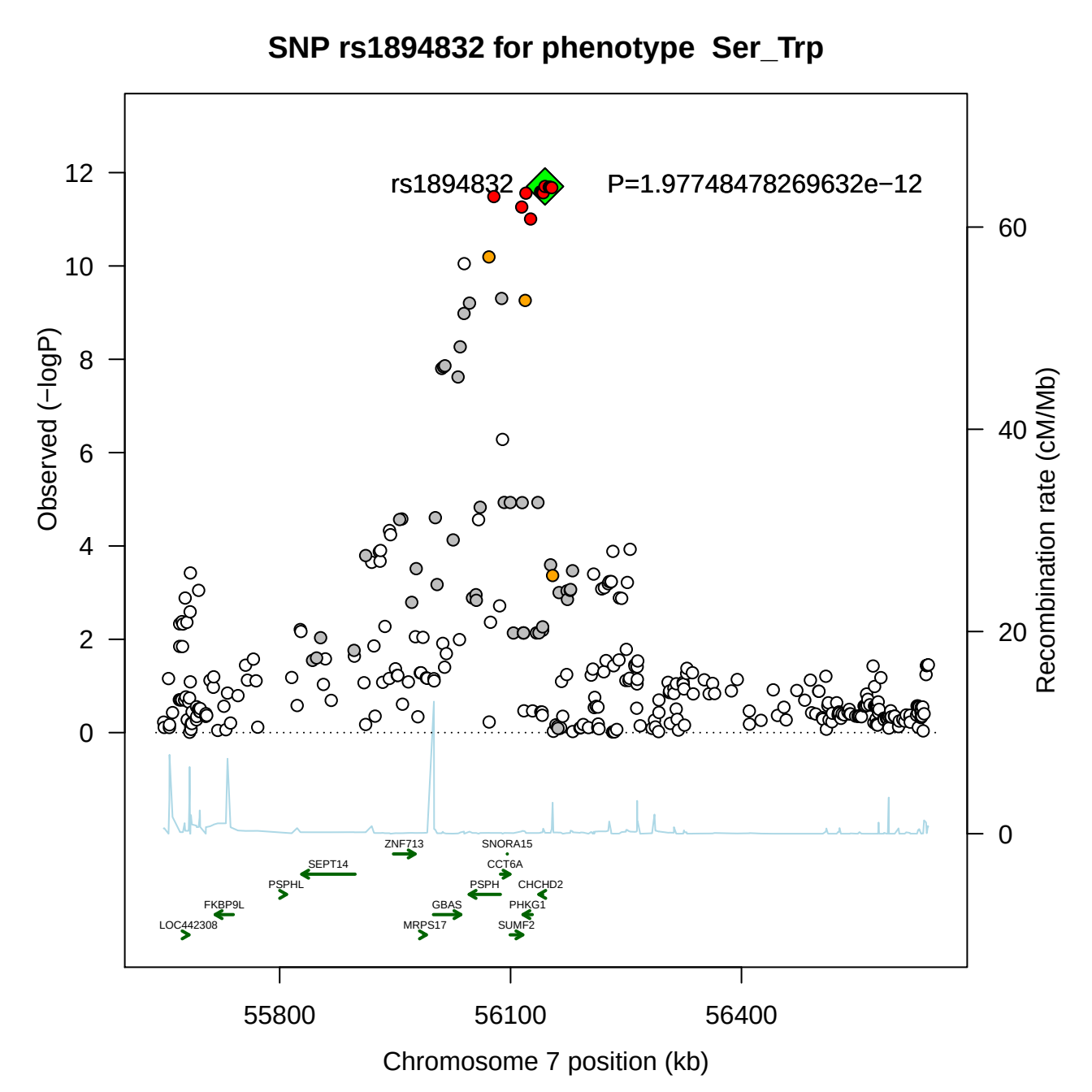

SNP rs2066938 for phenotype C3_C4
rs2066938
P=1.72590095397215e−159
Observed (−logP)
Recombination rate (cM/Mb)
0
20
40
60
PLA2G1B
GATC
RNF10
UNC119B
HNF1A
P2RX4
SIRT4
TRIAP1
COQ5
MLEC
C12orf27
P2RX7
NME2P1
COX6AP1
DYNLL1
CABP1
SPPL3
OASL
PXN
MSI1
SFRS9
POP5
ACADS
C12orf43
119300
119600
119900
Chromosome 12 position (kb)

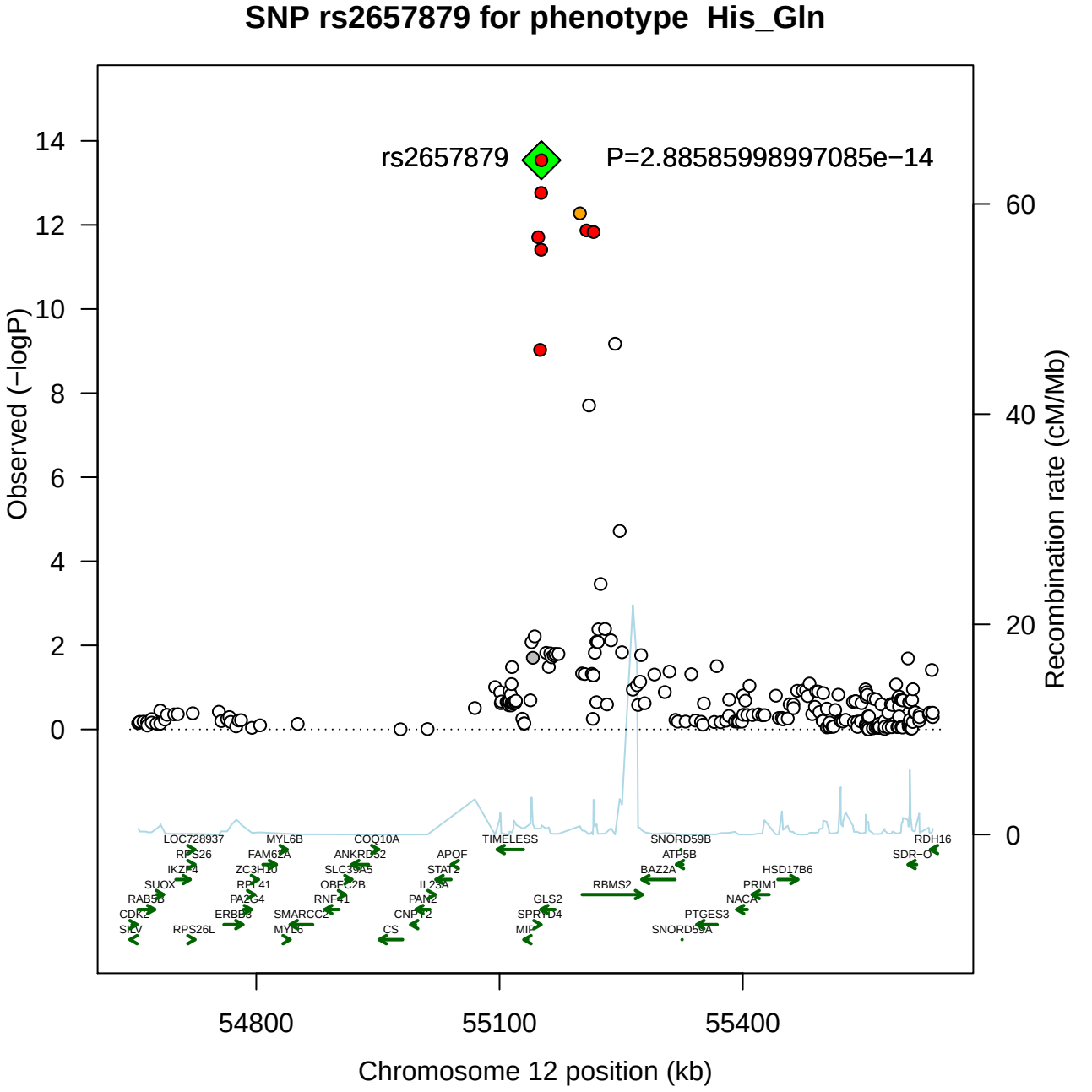

SNP rs4902242 for phenotype PC.aa.C28.1_PC.ae.C40.2
rs4902242
P=1.99943446233765e−35
Observed (−logP)
Recombination rate (cM/Mb)
0
2
4
6
8
10
12
14
16
18
20
22
24
26
28
30
32
34
36
0
20
40
60
GPHB5
WDR89
SYNE2
RHOJ
PPP2R5E
SGPP1
ESR2
63000
63300
63600
Chromosome 14 position (kb)

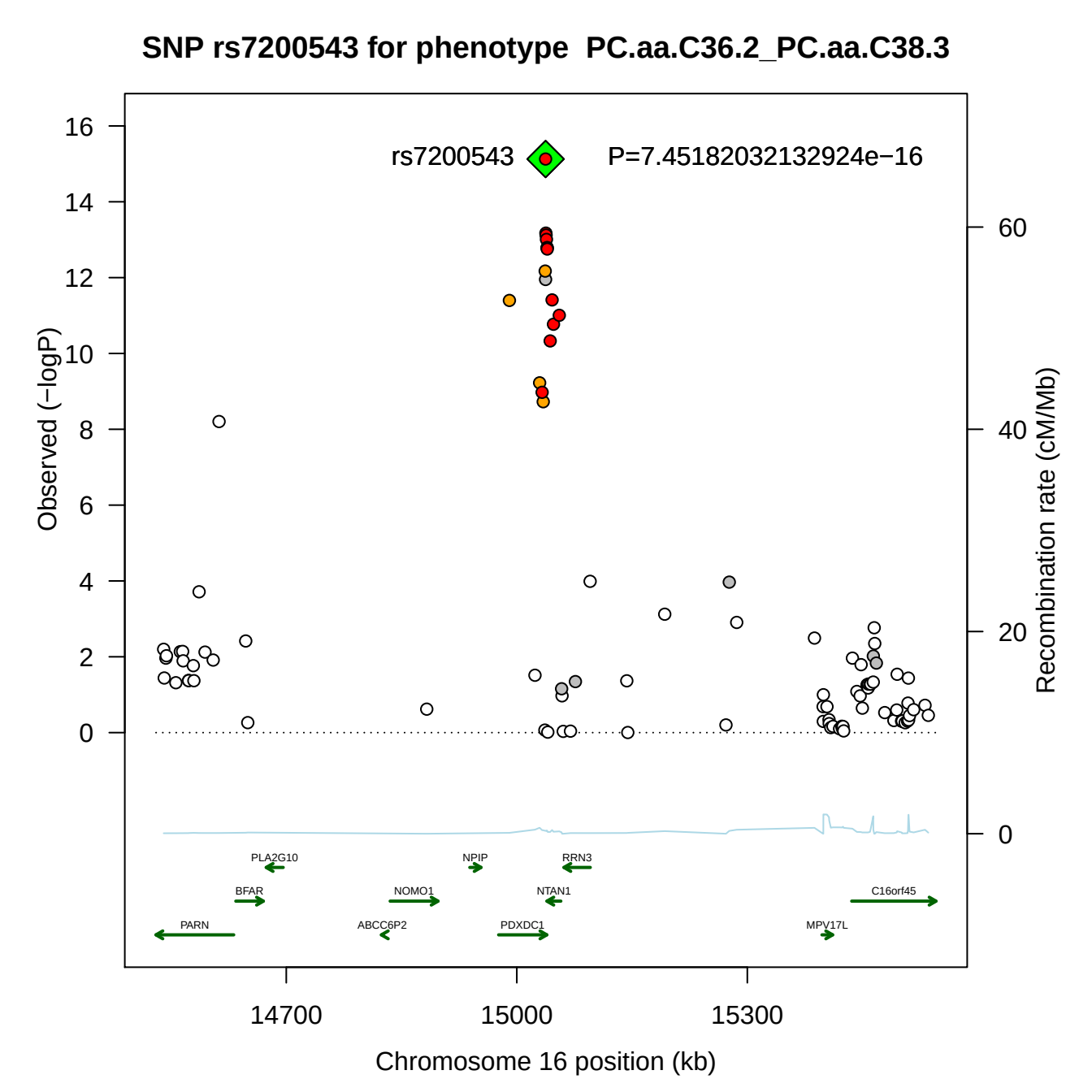

SNP rs1077989 for phenotype PC.ae.C32.1_PC.ae.C34.1
rs1077989
P=6.80053681009212e−42
Observed (−logP)
Recombination rate (cM/Mb)
FAM71D
ATP6V1D
PLEK2
PLEKHH1
ARG2
RDH11
ZFYVE26
GPHN
MPP5
EIF2S1
C14orf83
PIGH
VTI1B
RDH12
RAD51L1
66800
67100
67400
Chromosome 14 position (kb)

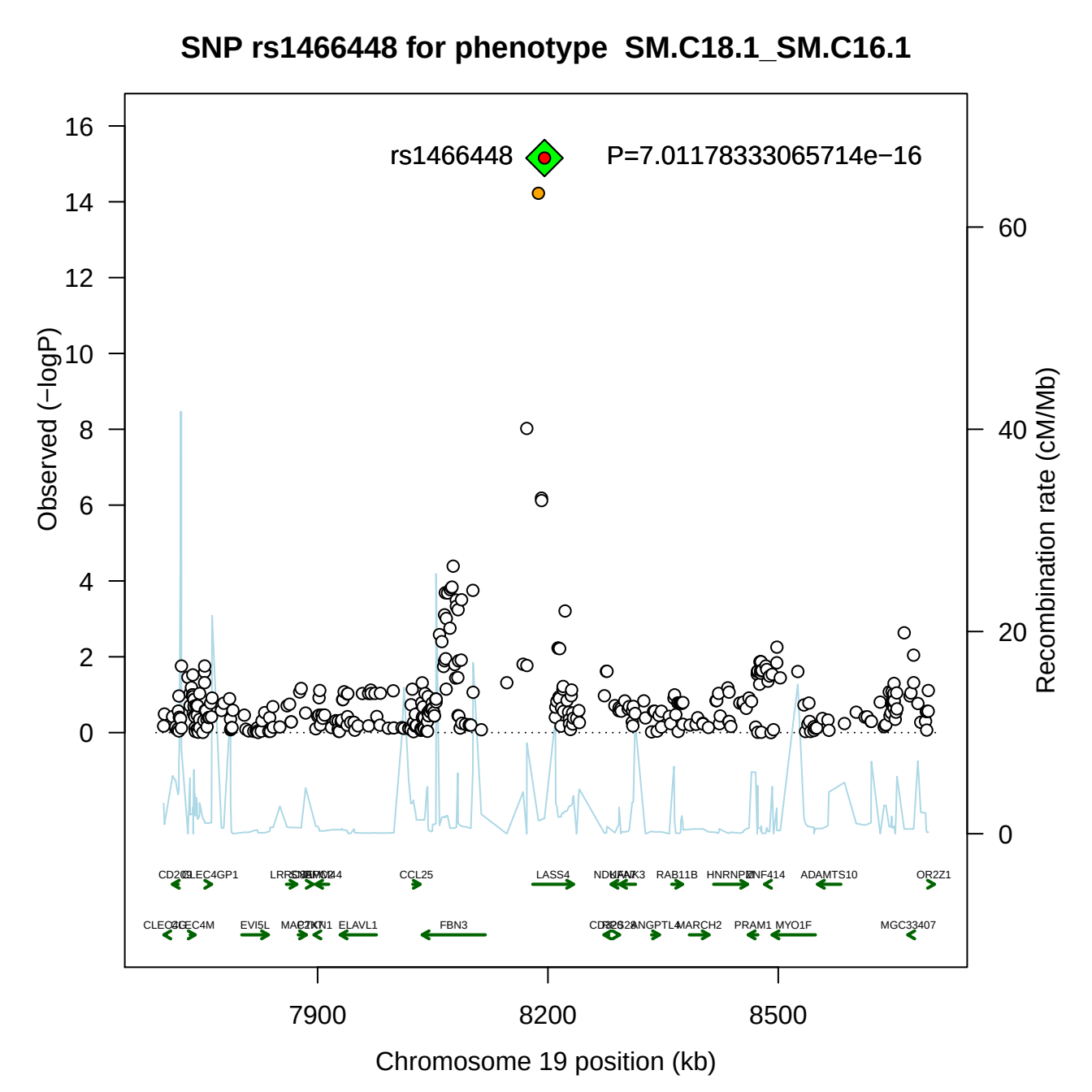

SNP rs4814176 for phenotype SM..OH..C24.1_SM..OH..C22.1
rs4814176
P=2.6947430699518e−31
Observed (−logP)
Recombination rate (cM/Mb)
SPTLC3
C20orf82
TASP1
12600
12900
13200
Chromosome 20 position (kb)

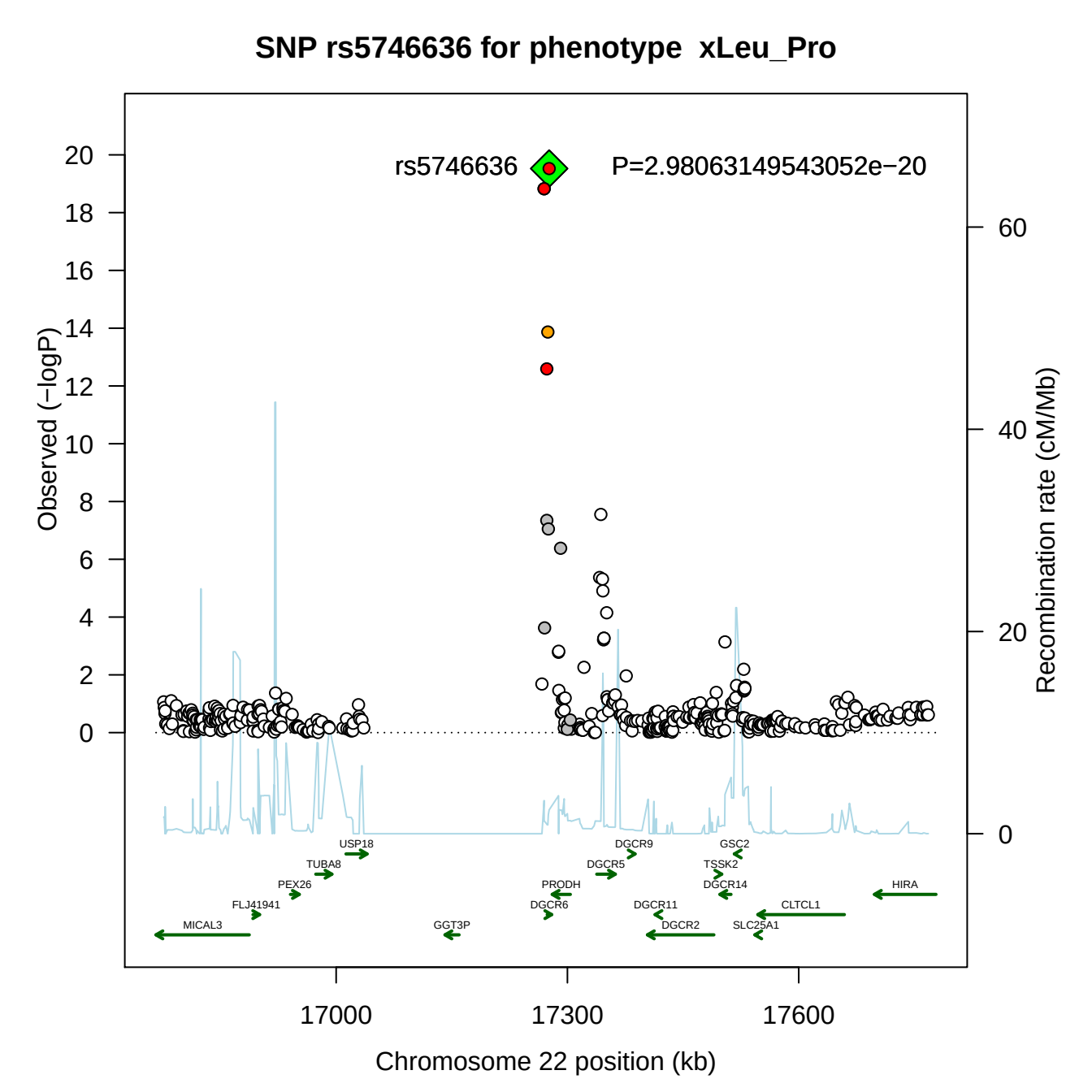

SNP rs2355078 for phenotype lysoPC.a.C28.0_PC.aa.C26.0
rs6970485
P=1.21174970249122e−47
Observed (−logP)
Recombination rate (cM/Mb)
THSD7A
TMEM106B
11500
11800
12100
Chromosome 7 position (kb)


SNP rs1894832 for phenotype Ser_Trp
rs1894832
P=1.97748478269632e−12
Observed (−logP)
Recombination rate (cM/Mb)
0
2
4
6
8
10
12
0
20
40
60
55800
56100
56400
Chromosome 7 position (kb)
ZNF713
SNORA15
SEPT14
CCT6A
PSPHL
PSPH
CHCHD2
FKBP9L
GBAS
PHKG1
LOC442308
MRPS17
SUMF2

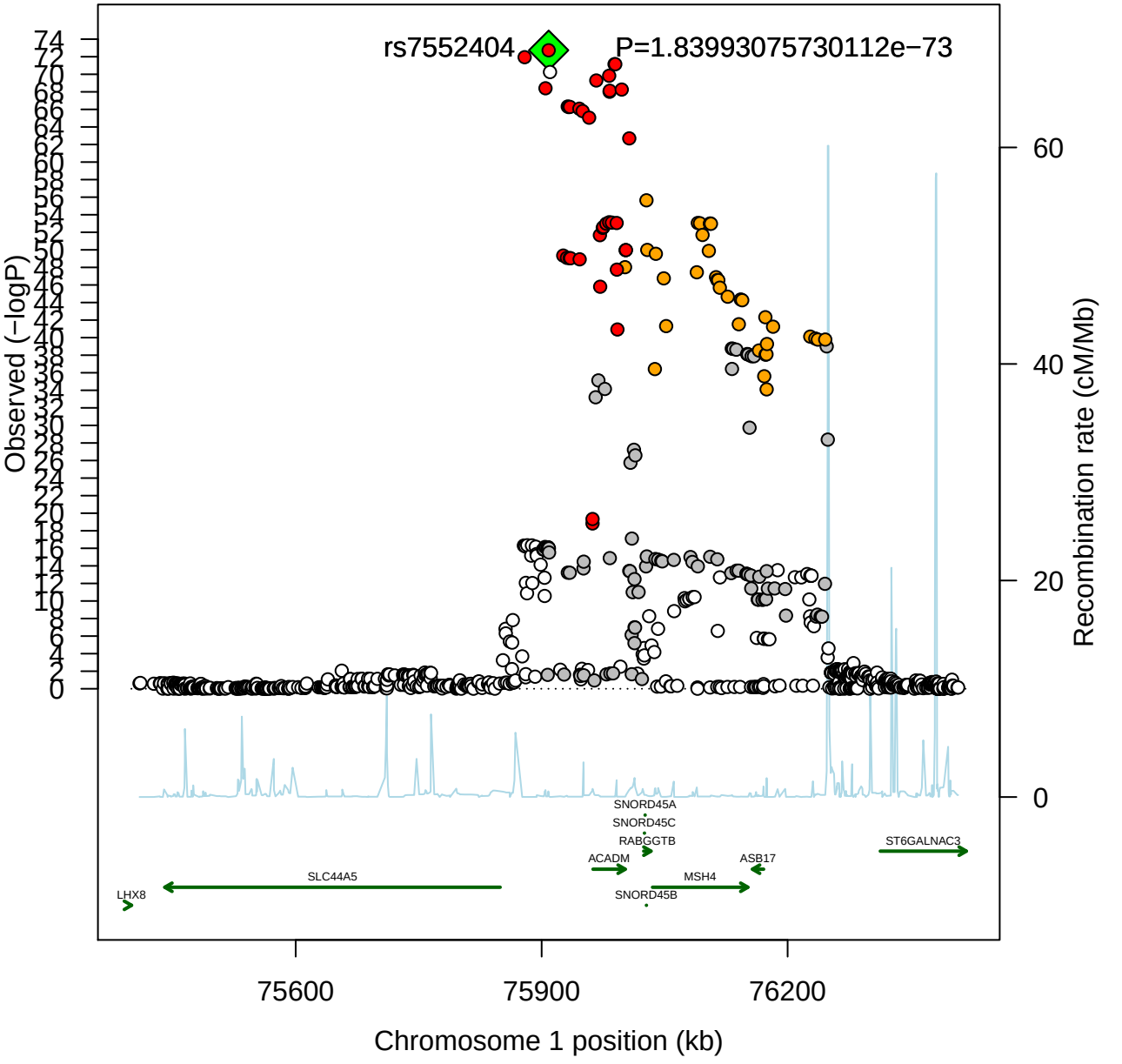

SNP rs2657879 for phenotype His_Gln
rs2657879
P=2.88585998997085e−14
Observed (−logP)
Recombination rate (cM/Mb)
Chromosome 12 position (kb)
54800
55100
55400
LOC728937
MYL6B
COQ10A
TIMELESS
SNORD59B
RDH16
RPS26
FAM62A
ANKRD52
APOF
ATP5B
SDR−O
IKZF4
ZC3H10
SLC39A5
STAT2
BAZ2A
HSD17B6
SUOX
RPL41
OBFC2B
IL23A
RBMS2
PRIM1
RAB5B
PA2G4
RNF41
PAN2
GLS2
NACA
CDK2
ERBB3
SMARCC2
CNP72
SPRYD4
PTGES3
SIV
RPS26L
MYL6
CS
MIP
SNORD59A

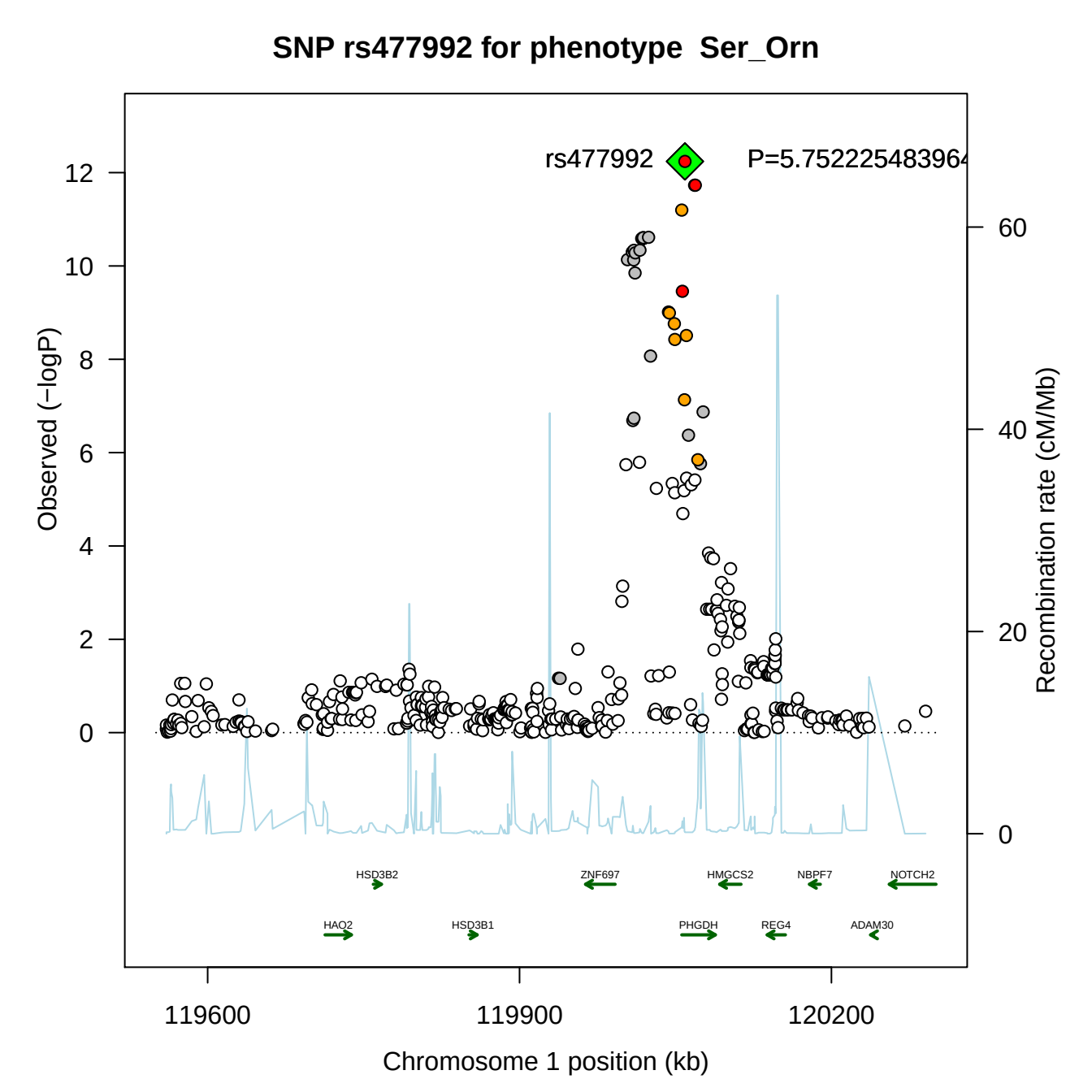

SNP rs7200543 for phenotype PC.aa.C36.2_PC.aa.C38.3
rs7200543
P=7.45182032132924e−16
Observed (−logP)
Recombination rate (cM/Mb)
0
2
4
6
8
10
12
14
16
0
20
40
60
PLA2G10
NPIP
RRN3
BFAR
NOMO1
NTAN1
C16orf45
PARN
ABCC6P2
PDXDC1
MPV17L
14700
15000
15300
Chromosome 16 position (kb)

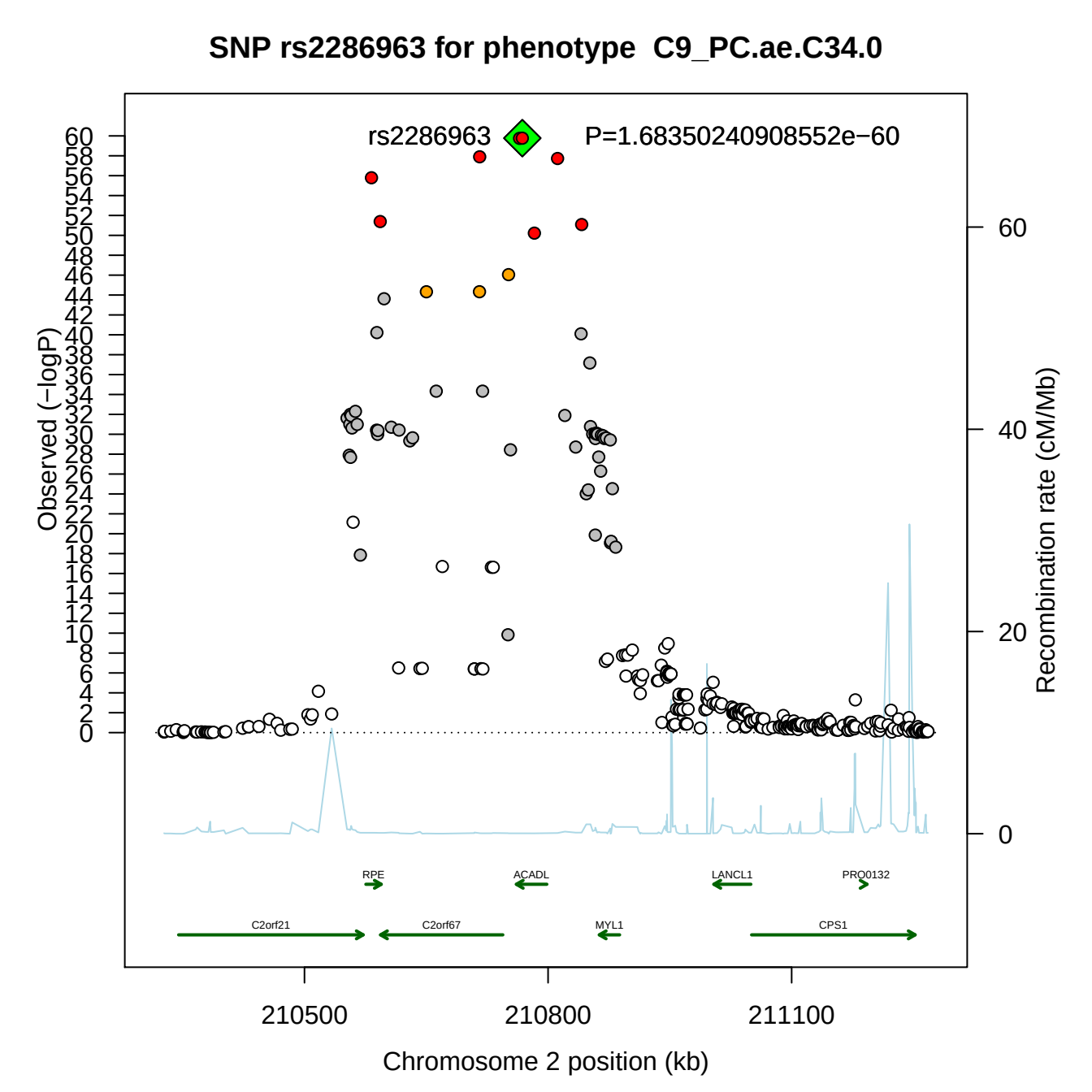

SNP rs1466448 for phenotype SM.C18.1_SM.C16.1
rs1466448
P=7.01178333065714e−16
Observed (−logP)
Recombination rate (cM/Mb)
Chromosome 19 position (kb)
7900
8200
8500
CCL25
LASS4
RAB11B
ADAMTS10
OR2Z1
EVI5L
FBN3
MARCH2
PRAM1
MYO1F
MGC33407

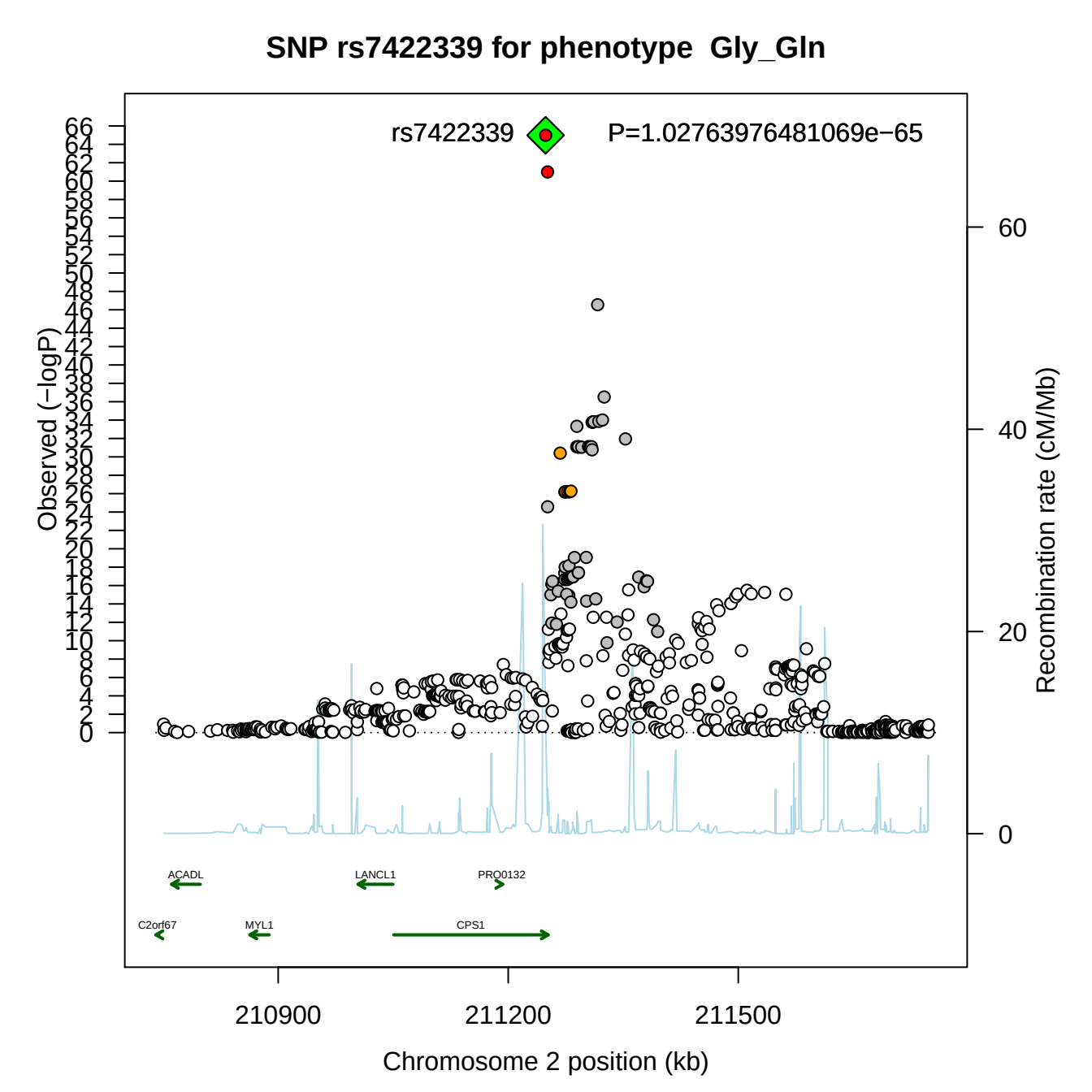

SNP rs5746636 for phenotype xLeu_Pro
rs5746636
P=2.98063149543052e−20
Observed (−logP)
Recombination rate (cM/Mb)
0
2
4
6
8
10
12
14
16
18
20
0
20
40
60
USP18
TUBA8
PEX26
FLJ41941
MICAL3
GGT3P
DGCR9
DGCR5
PRODH
DGCR6
DGCR11
DGCR2
GSC2
TSSK2
DGCR14
SLC25A1
CLTCL1
HIRA
17000
17300
17600
Chromosome 22 position (kb)

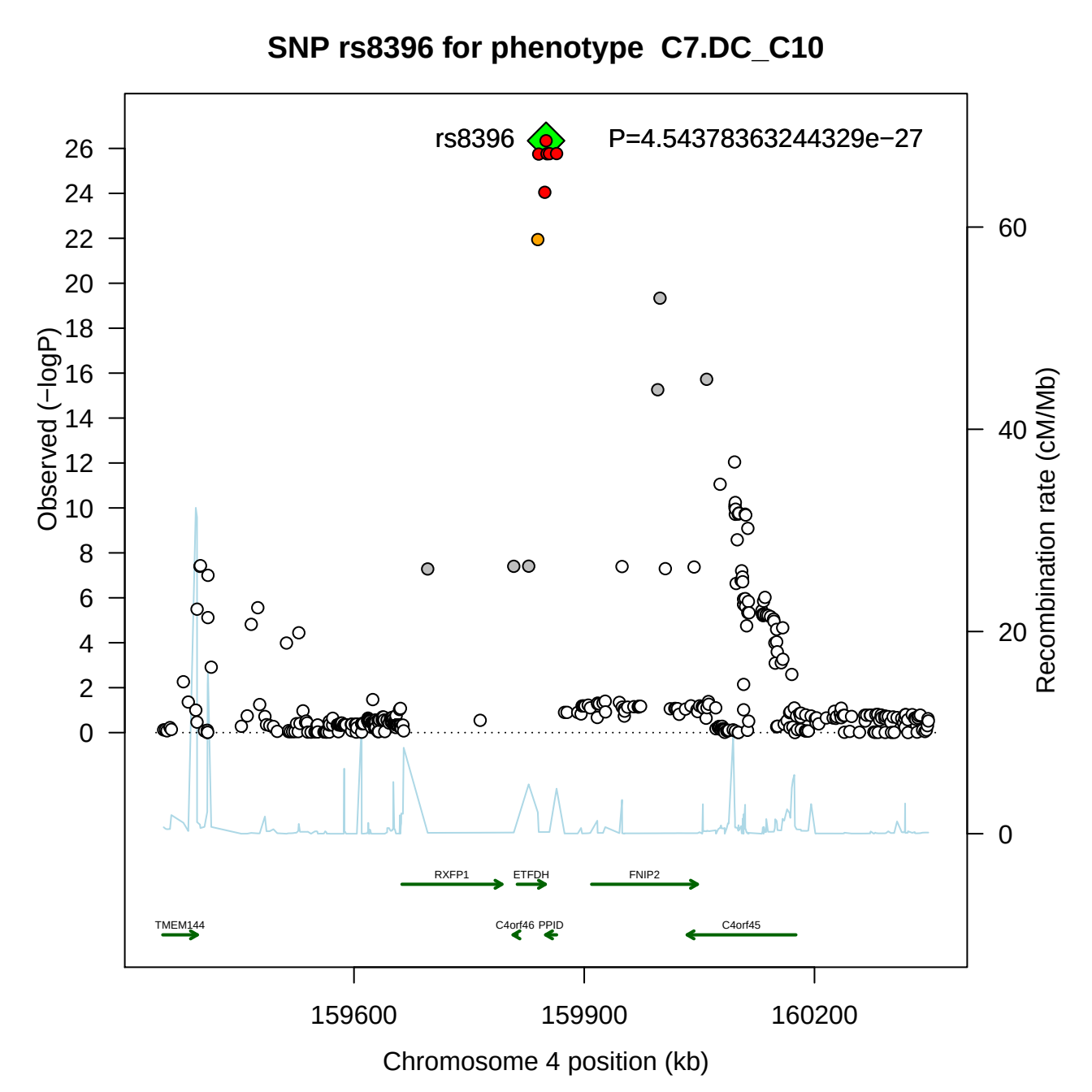
Supporting Figures S2. Regional association plots for additive model for KORA data.

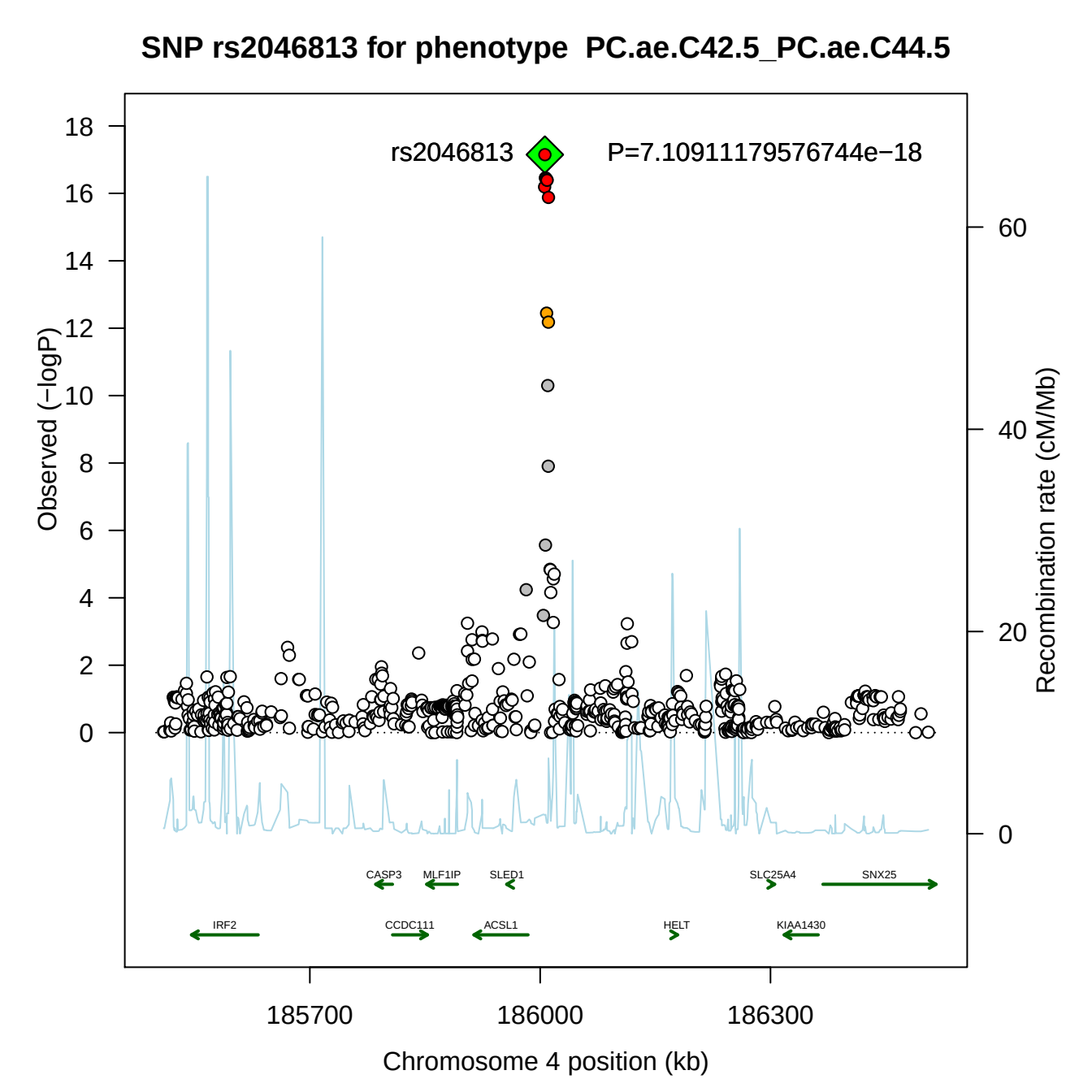

SNP rs7552404 for phenotype C12_C10
rs7552404
P=1.83993075730112e−73
Observed (−logP)
Recombination rate (cM/Mb)
Chromosome 1 position (kb)
75600
75900
76200
SNORD45A
SNORD45C
RABGGTB
ACADM
ASB17
ST6GALNAC3
SLC44A5
MSH4
LHX8
SNORD45B

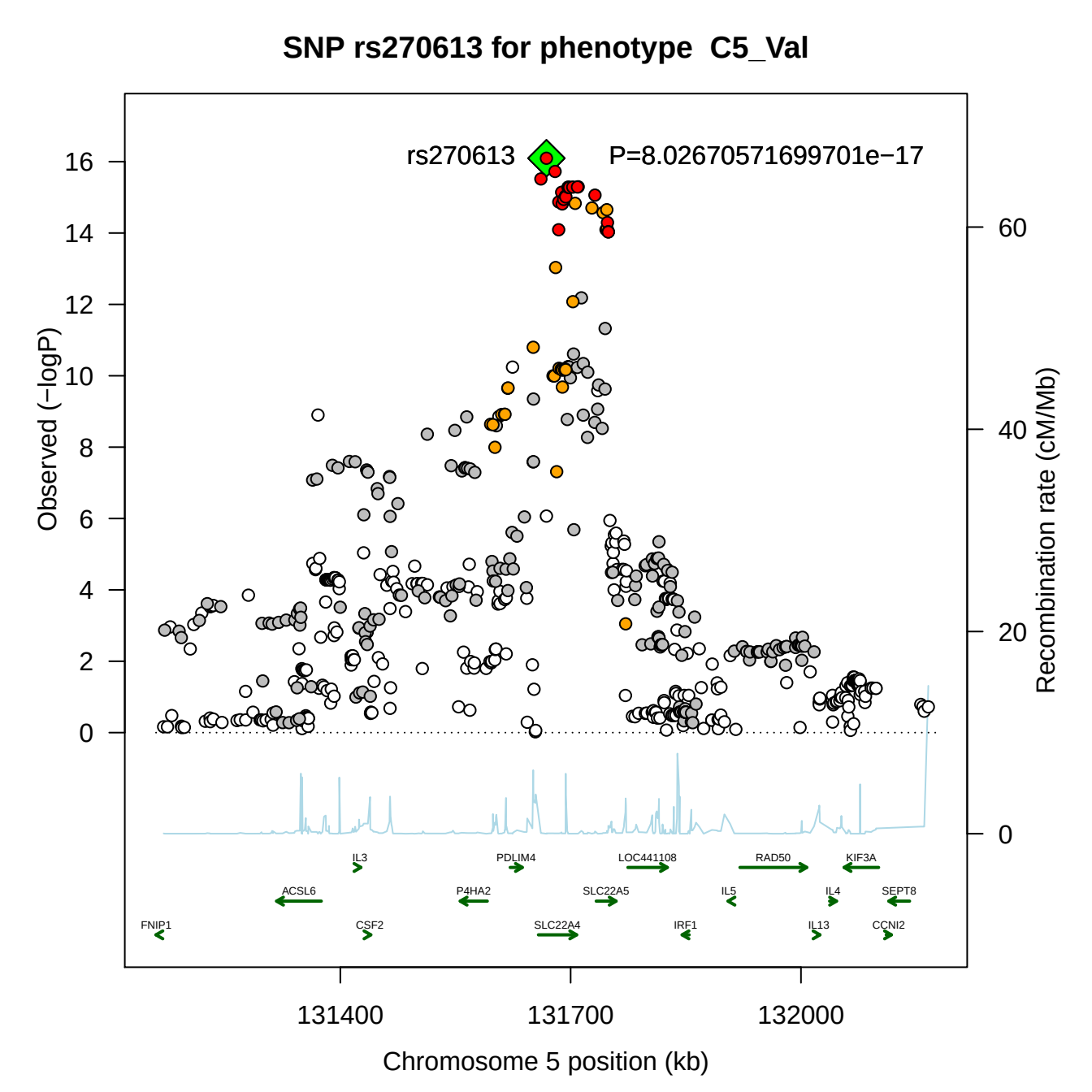

SNP rs477992 for phenotype Ser_Orn
rs477992
P=5.752225483964
Observed (−logP)
Recombination rate (cM/Mb)
0
2
4
6
8
10
12
0
20
40
60
HSD3B2
ZNF697
HMGCS2
NBPF7
NOTCH2
HAO2
HSD3B1
PHGDH
REG4
ADAM30
119600
119900
120200
Chromosome 1 position (kb)

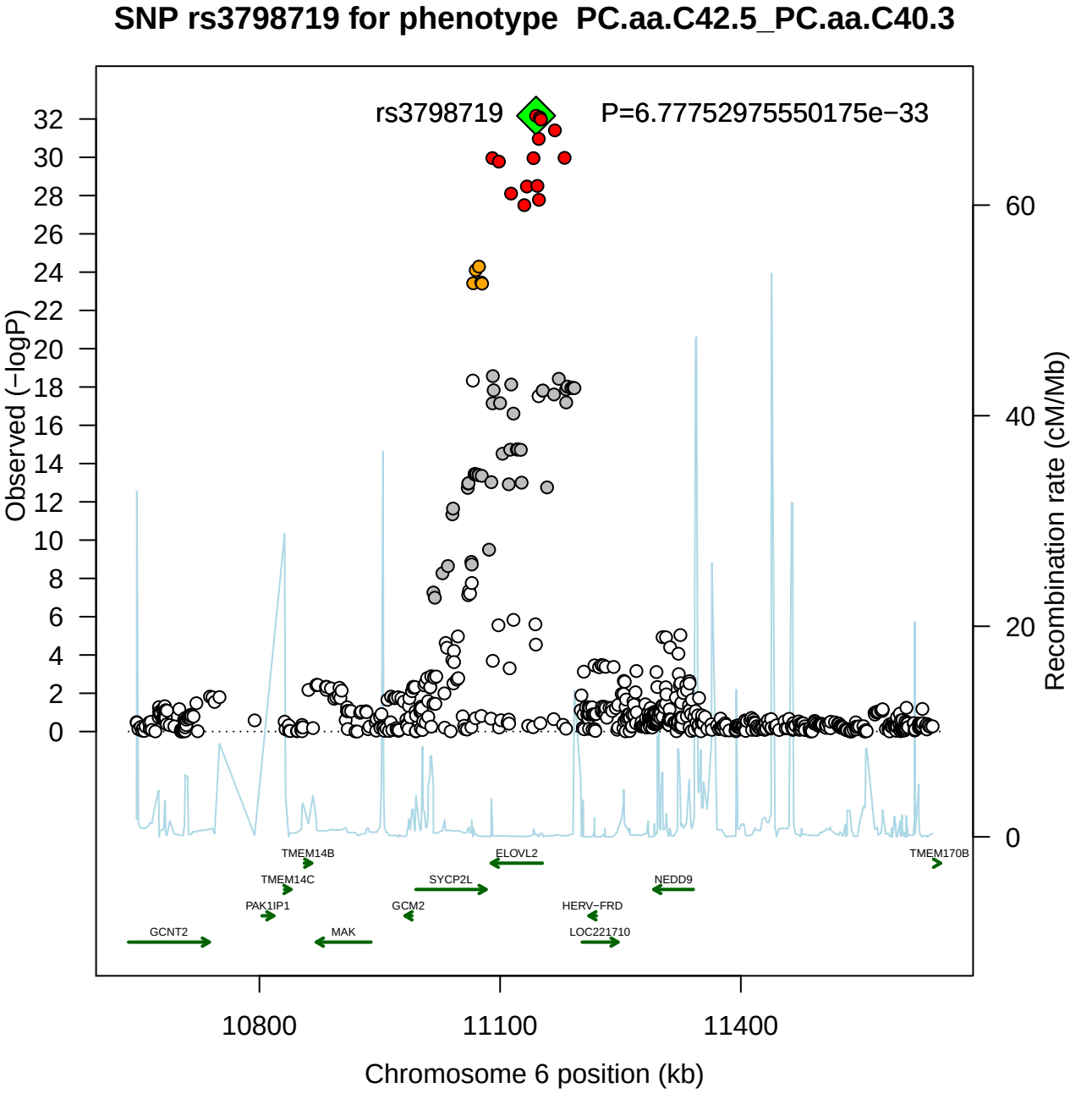

SNP rs2286963 for phenotype C9_PC.ae.C34.0
rs2286963
P=1.68350240908552e−60
Observed (−logP)
Recombination rate (cM/Mb)
RPE
ACADL
LANCL1
PRO0132
C2orf21
C2orf67
MYL1
CPS1
210500
210800
211100
Chromosome 2 position (kb)

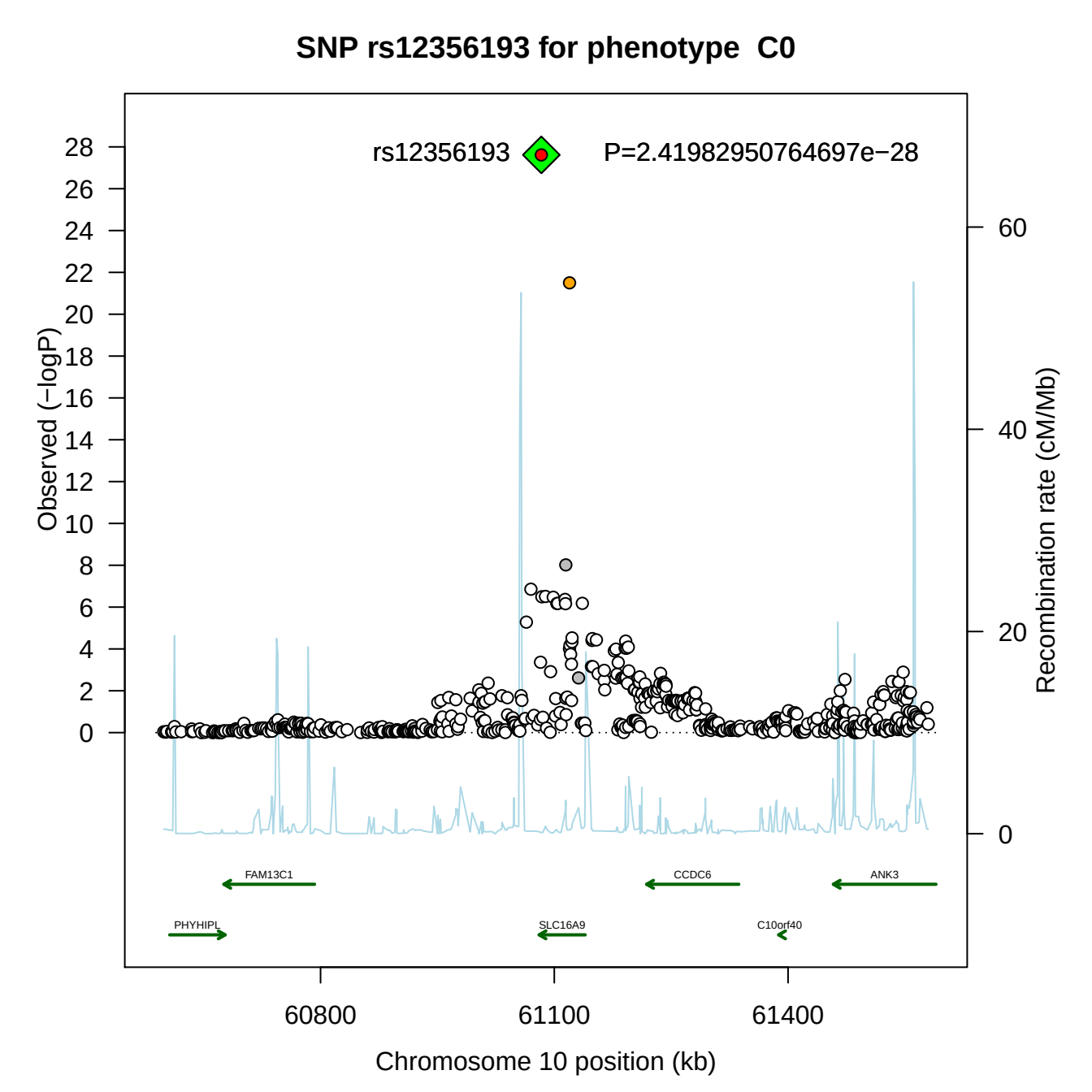

SNP rs7422339 for phenotype Gly_Gln
rs7422339
P=1.02763976481069e−65
Observed (−logP)
Recombination rate (cM/Mb)
ACADL
LANCL1
PRO0132
C2orf67
MYL1
CPS1
210900
211200
211500
Chromosome 2 position (kb)

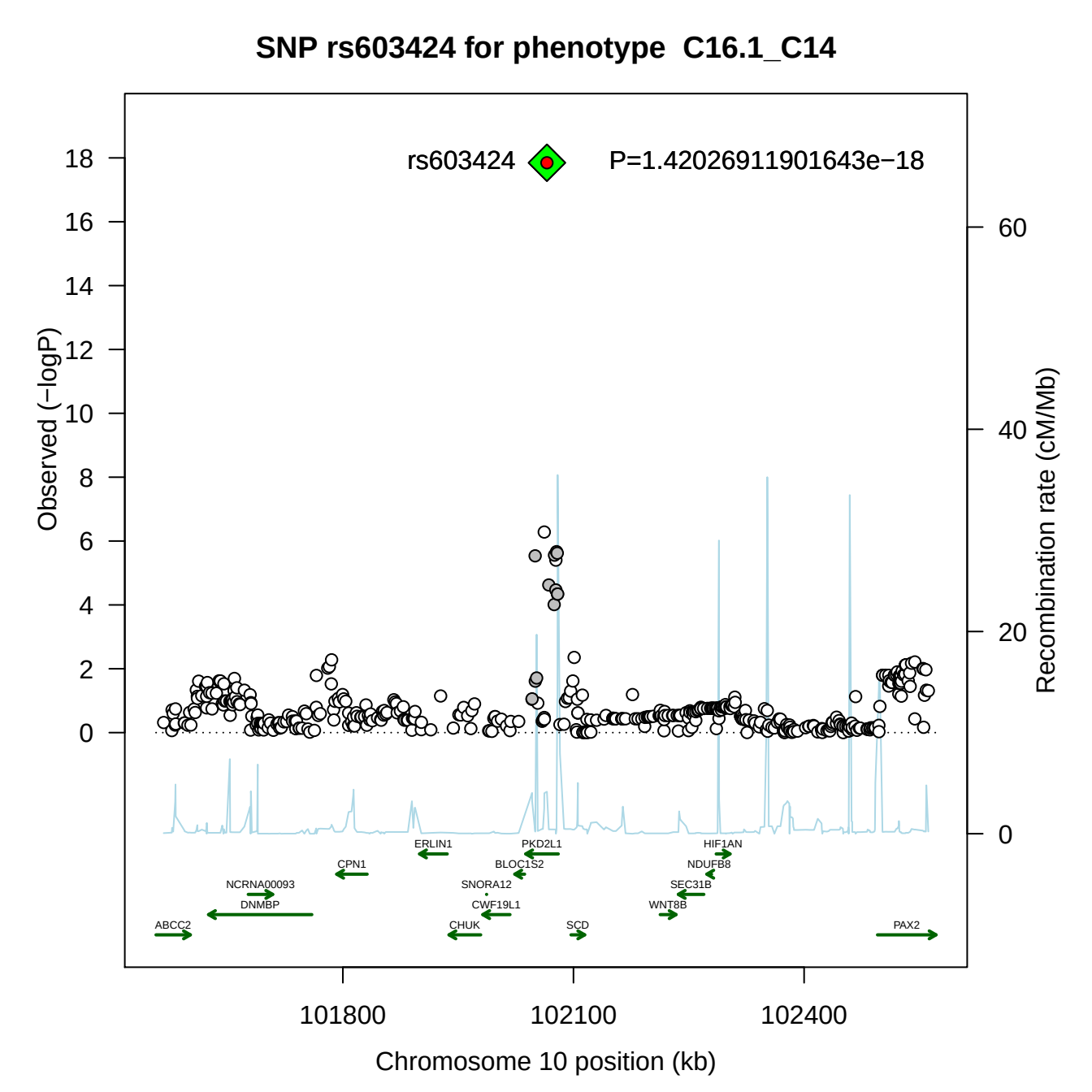

SNP rs8396 for phenotype C7.DC_C10
rs8396
P=4.54378363244329e−27
Observed (−logP)
Recombination rate (cM/Mb)
0
2
4
6
8
10
12
14
16
18
20
22
24
26
0
20
40
60
RXFP1
ETFDH
FNIP2
TMEM144
C4orf46
PPID
C4orf45
159600
159900
160200
Chromosome 4 position (kb)

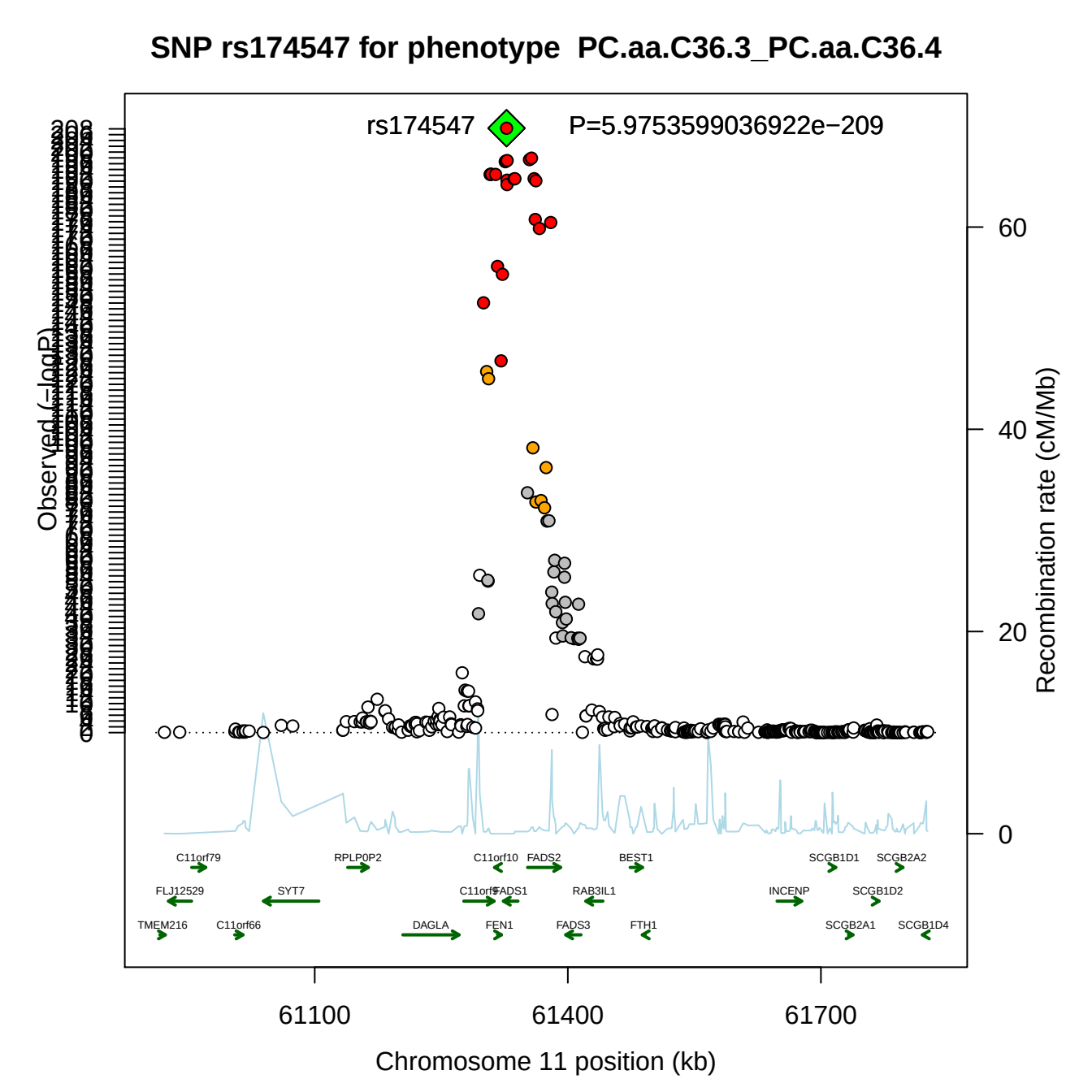

SNP rs2046813 for phenotype PC.ae.C42.5_PC.ae.C44.5
rs2046813
P=7.10911179576744e−18
Observed (−logP)
Recombination rate (cM/Mb)
0
2
4
6
8
10
12
14
16
18
0
20
40
60
CASP3
MLF1IP
SLED1
SLC25A4
SNX25
IRF2
CCDC111
ACSL1
HELT
KIAA1430
185700
186000
186300
Chromosome 4 position (kb)

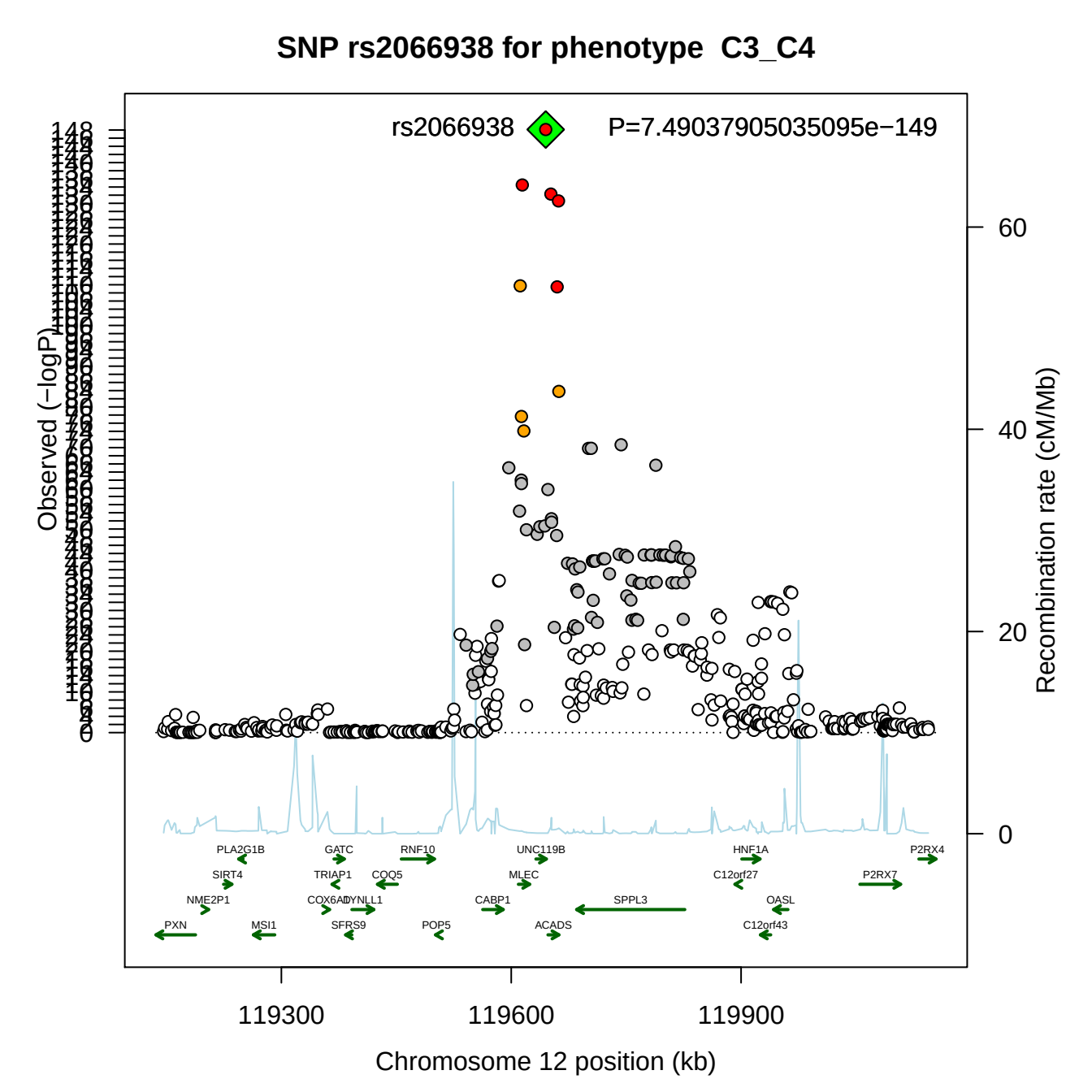

SNP rs270613 for phenotype C5_Val
rs270613
P=8.02670571699701e−17
Observed (−logP)
Recombination rate (cM/Mb)
0
2
4
6
8
10
12
14
16
0
20
40
60
IL3
PDLIM4
LOC441108
RAD50
KIF3A
ACSL6
P4HA2
SLC22A5
IL5
IL4
SEPT8
FNIP1
CSF2
SLC22A4
IRF1
IL13
CCNI2
131400
131700
132000
Chromosome 5 position (kb)

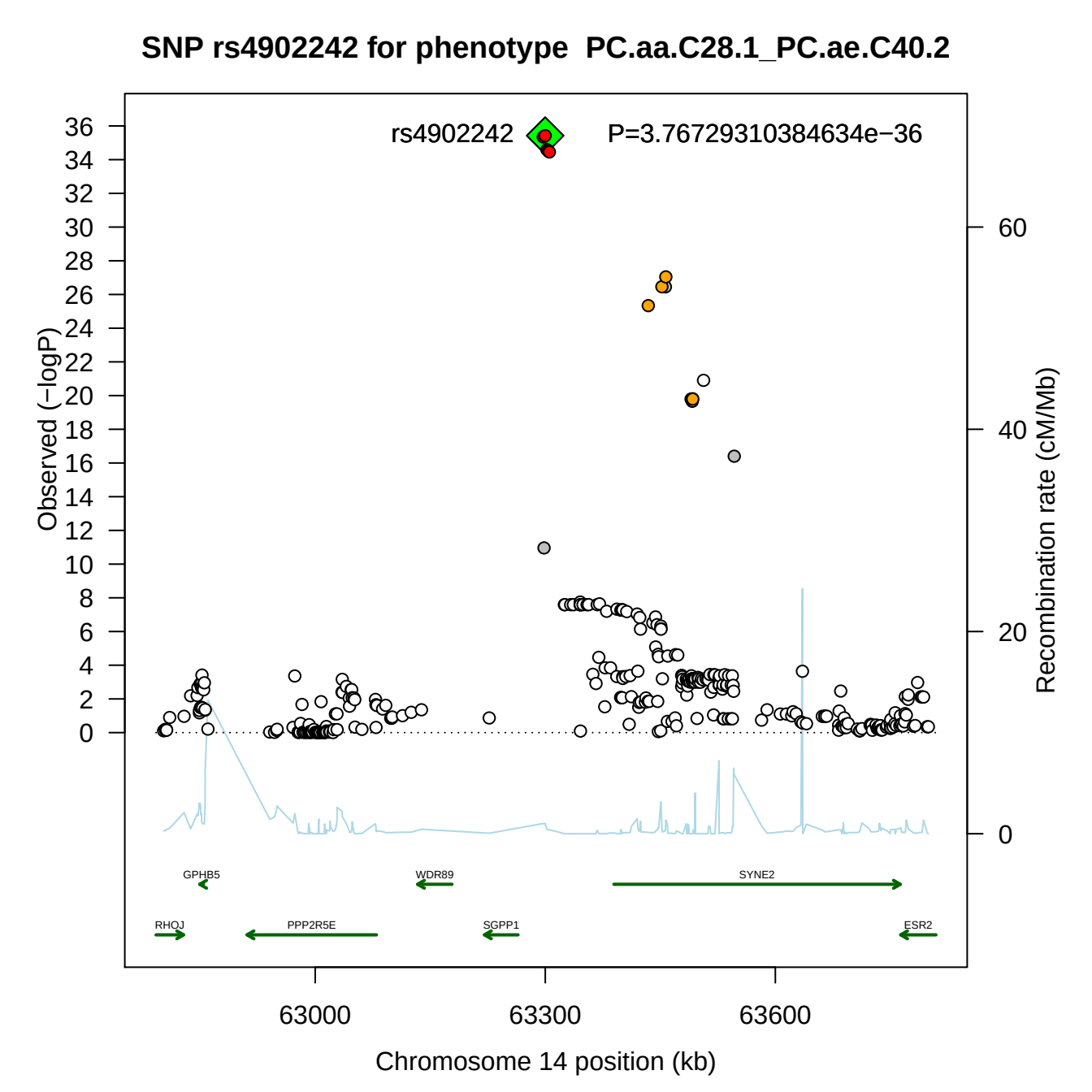

SNP rs3798719 for phenotype PC.aa.C42.5_PC.aa.C40.3
rs3798719
P=6.77752975550175e−33
Observed (−logP)
Recombination rate (cM/Mb)
Chromosome 6 position (kb)
10800
11100
11400
TMEM14B
TMEM14C
PAK1IP1
GCNT2
MAK
GCM2
SYCP2L
ELOVL2
HERV−FRD
LOC221710
NEDD9
TMEM170B

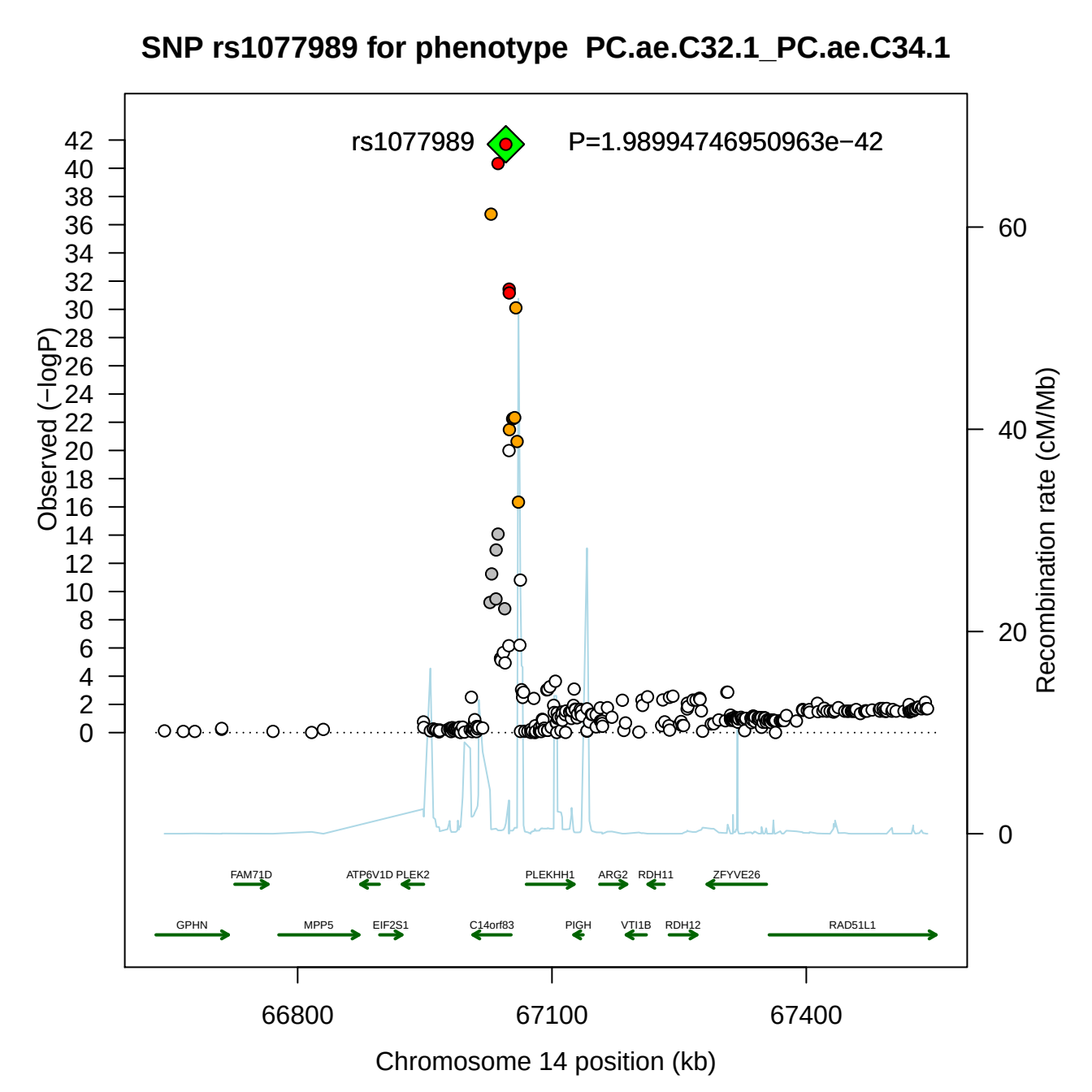

SNP rs12356193 for phenotype C0
rs12356193
P=2.41982950764697e−28
Observed (−logP)
Recombination rate (cM/Mb)
28
26
24
22
20
18
16
14
12
10
8
6
4
2
0
60
40
20
0
FAM13C1
CCDC6
ANK3
PHYHIPL
SLC16A9
C10orf40
60800
61100
61400
Chromosome 10 position (kb)

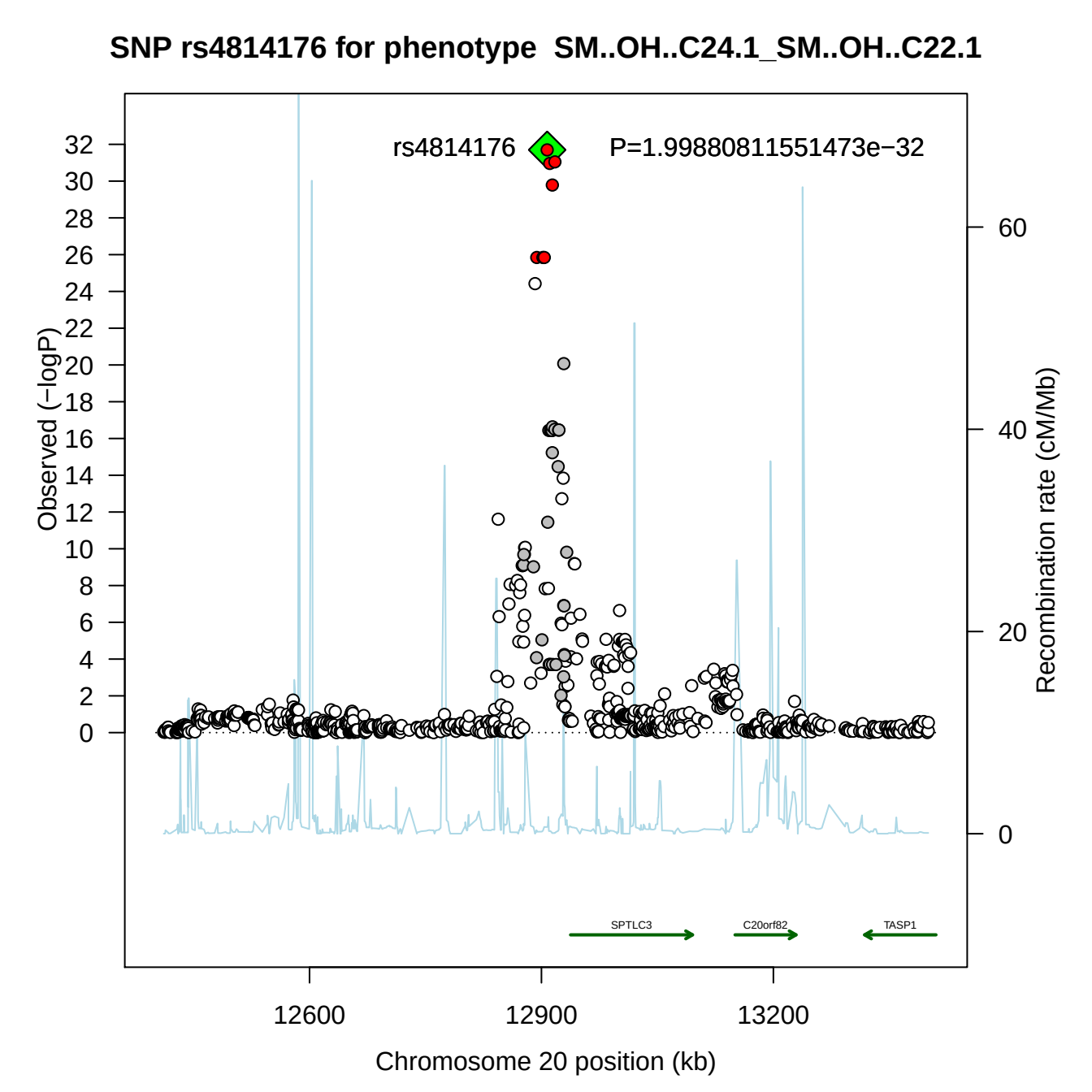

SNP rs603424 for phenotype C16.1_C14
rs603424
P=1.42026911901643e−18
Observed (−logP)
Recombination rate (cM/Mb)
Chromosome 10 position (kb)
101800
102100
102400
ERLIN1
PKD2L1
HIF1AN
CPN1
BLOC1S2
NDUFB8
NCRNA00093
SNORA12
SEC31B
DNMBP
CWF19L1
WNT8B
ABCC2
CHUK
SCD
PAX2

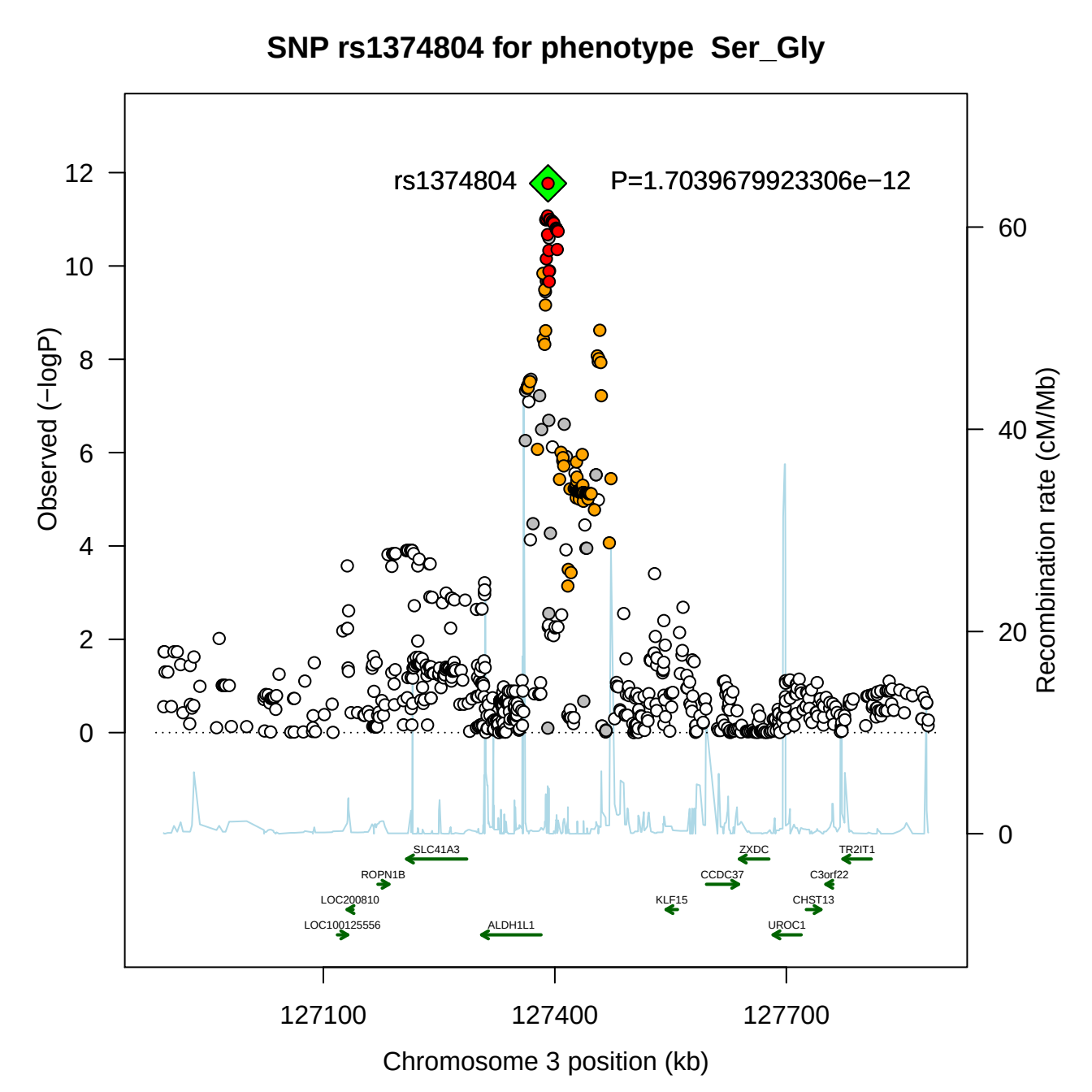

SNP rs174547 for phenotype PC.aa.C36.3_PC.aa.C36.4
rs174547
P=5.9753599036922e−209
Observed (−logP)
Recombination rate (cM/Mb)
0
20
40
60
C11orf79
RPLP0P2
C11orf10
FADS2
BEST1
SCGB1D1
SCGB2A2
FLJ12529
SYT7
C11orf9
FADS1
RAB3IL1
INCENP
SCGB1D2
TMEM216
C11orf66
DAGLA
FEN1
FADS3
FTH1
SCGB2A1
SCGB1D4
61100
61400
61700
Chromosome 11 position (kb)

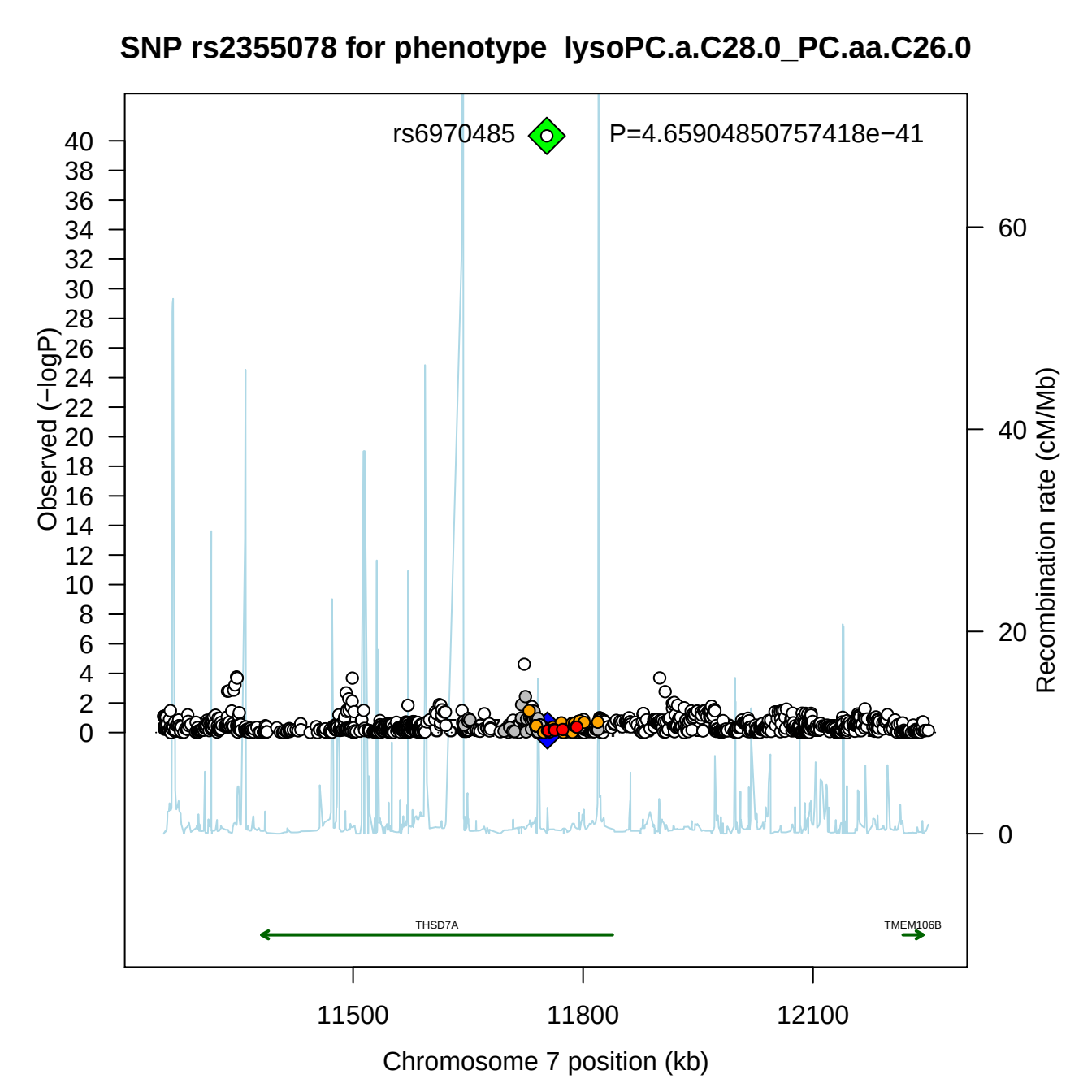

SNP rs2066938 for phenotype C3_C4
rs2066938
P=7.49037905035095e−149
Observed (−logP)
Recombination rate (cM/Mb)
60
40
20
0
PLA2G1B
GATC
RNF10
UNC119B
HNF1A
P2RX4
SIRT4
TRIAP1
COQ5
MLEC
C12orf27
P2RX7
NME2P1
COX6A1
DYNLL1
CABP1
SPPL3
OASL
PXN
MSI1
SFRS9
POP5
ACADS
C12orf43
119300
119600
119900
Chromosome 12 position (kb)

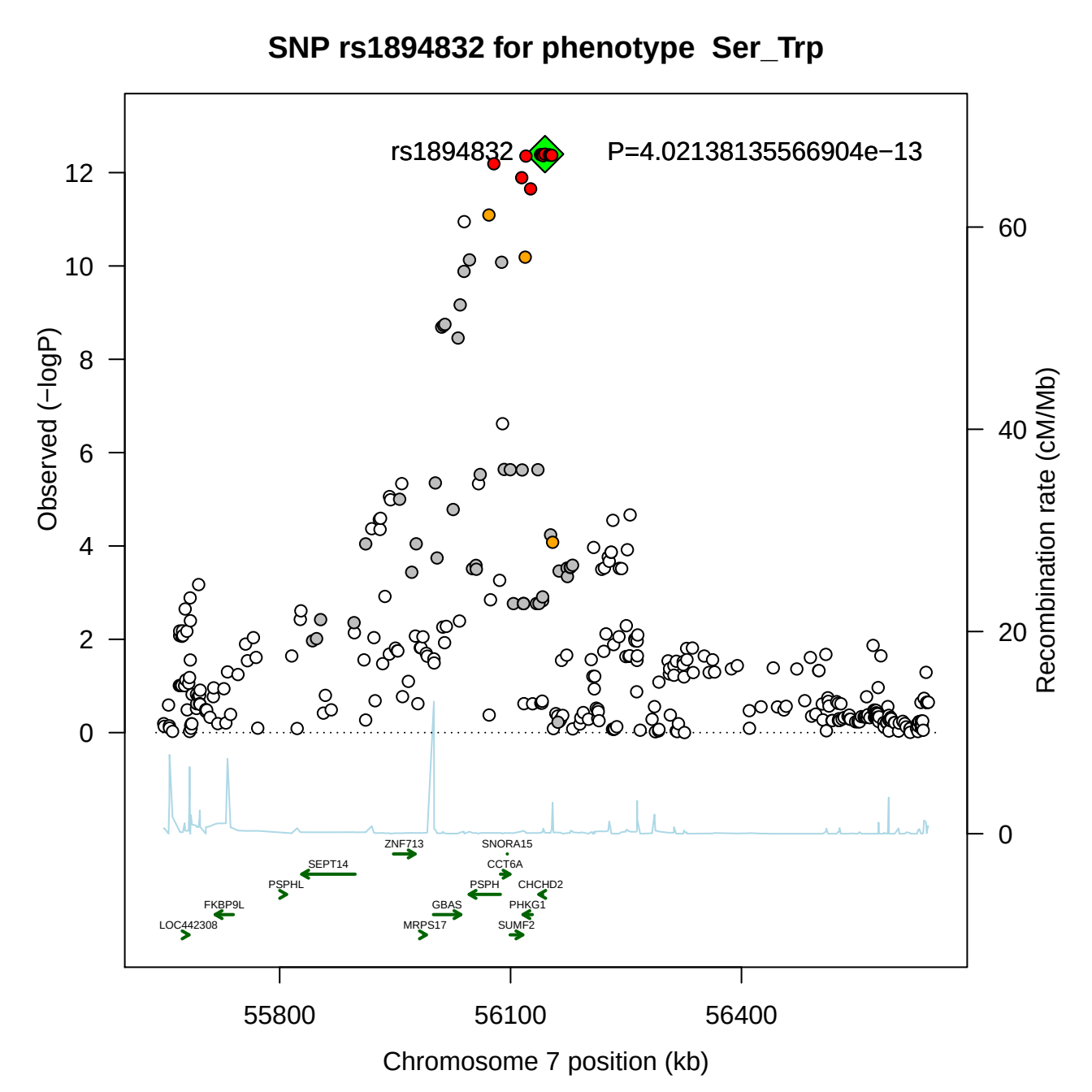

SNP rs4902242 for phenotype PC.aa.C28.1_PC.ae.C40.2
rs4902242
P=3.76729310384634e−36
Observed (−logP)
Recombination rate (cM/Mb)
Chromosome 14 position (kb)
63000
63300
63600
GPHB5
WDR89
SYNE2
RHOJ
PPP2R5E
SGPP1
ESR2

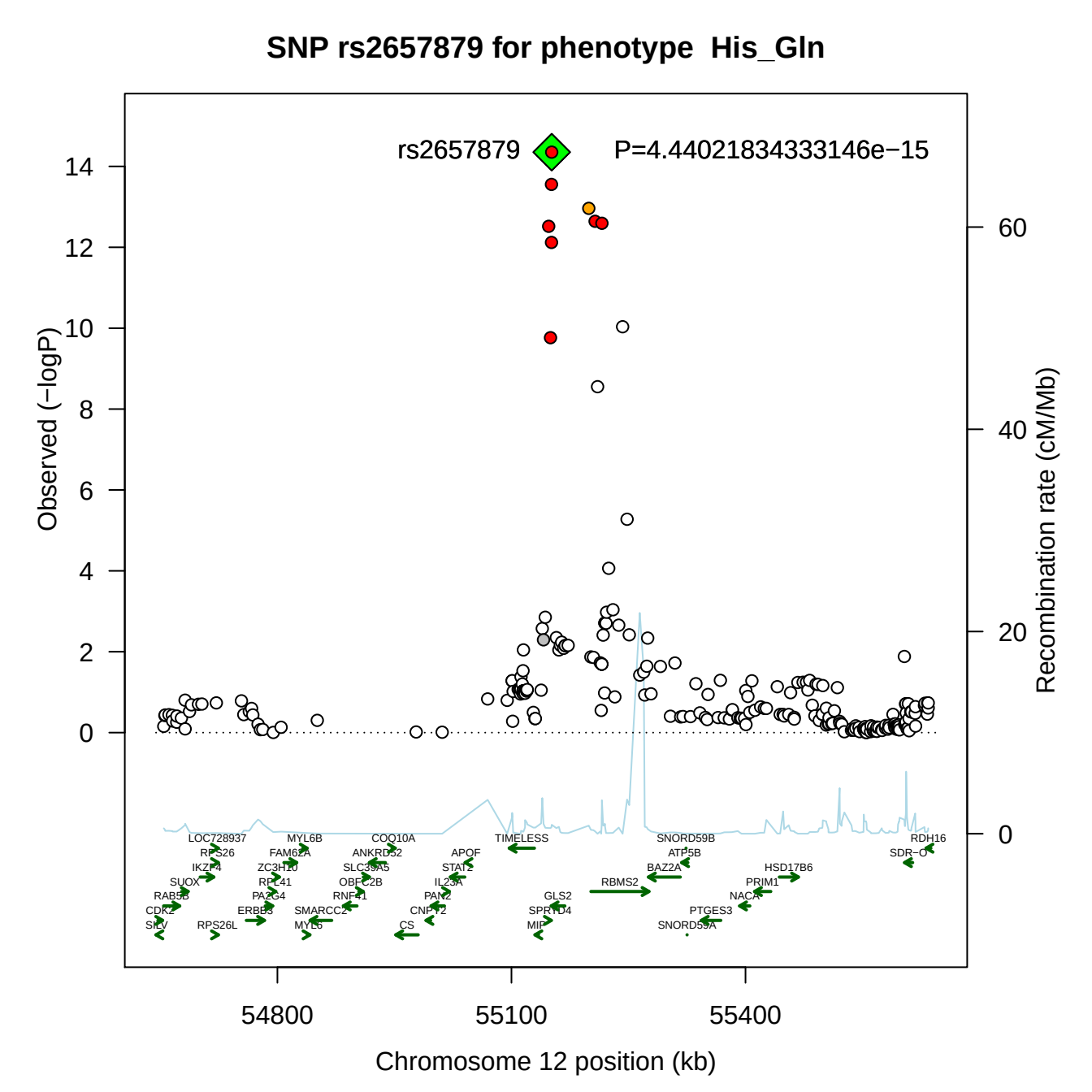

SNP rs1077989 for phenotype PC.ae.C32.1_PC.ae.C34.1
rs1077989
P=1.98994746950963e−42
Observed (−logP)
Recombination rate (cM/Mb)
FAM71D
ATP6V1D
PLEK2
PLEKHH1
ARG2
RDH11
ZFYVE26
GPHN
MPP5
EIF2S1
C14orf83
PIGH
VTI1B
RDH12
RAD51L1
66800
67100
67400
Chromosome 14 position (kb)

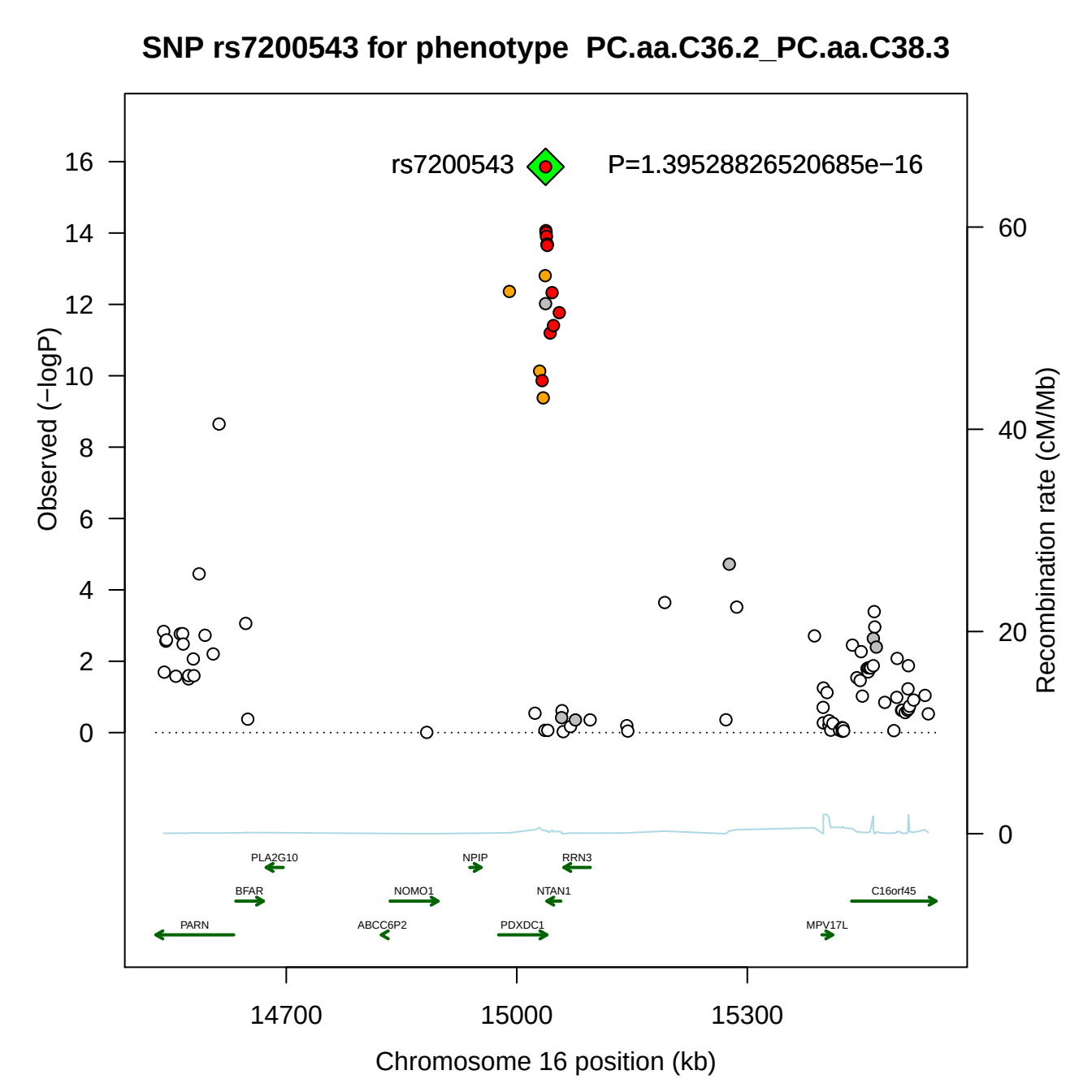

SNP rs4814176 for phenotype SM..OH..C24.1_SM..OH..C22.1
rs4814176
P=1.99880811551473e−32
Observed (−logP)
Recombination rate (cM/Mb)
SPTLC3
C20orf82
TASP1
12600
12900
13200
Chromosome 20 position (kb)

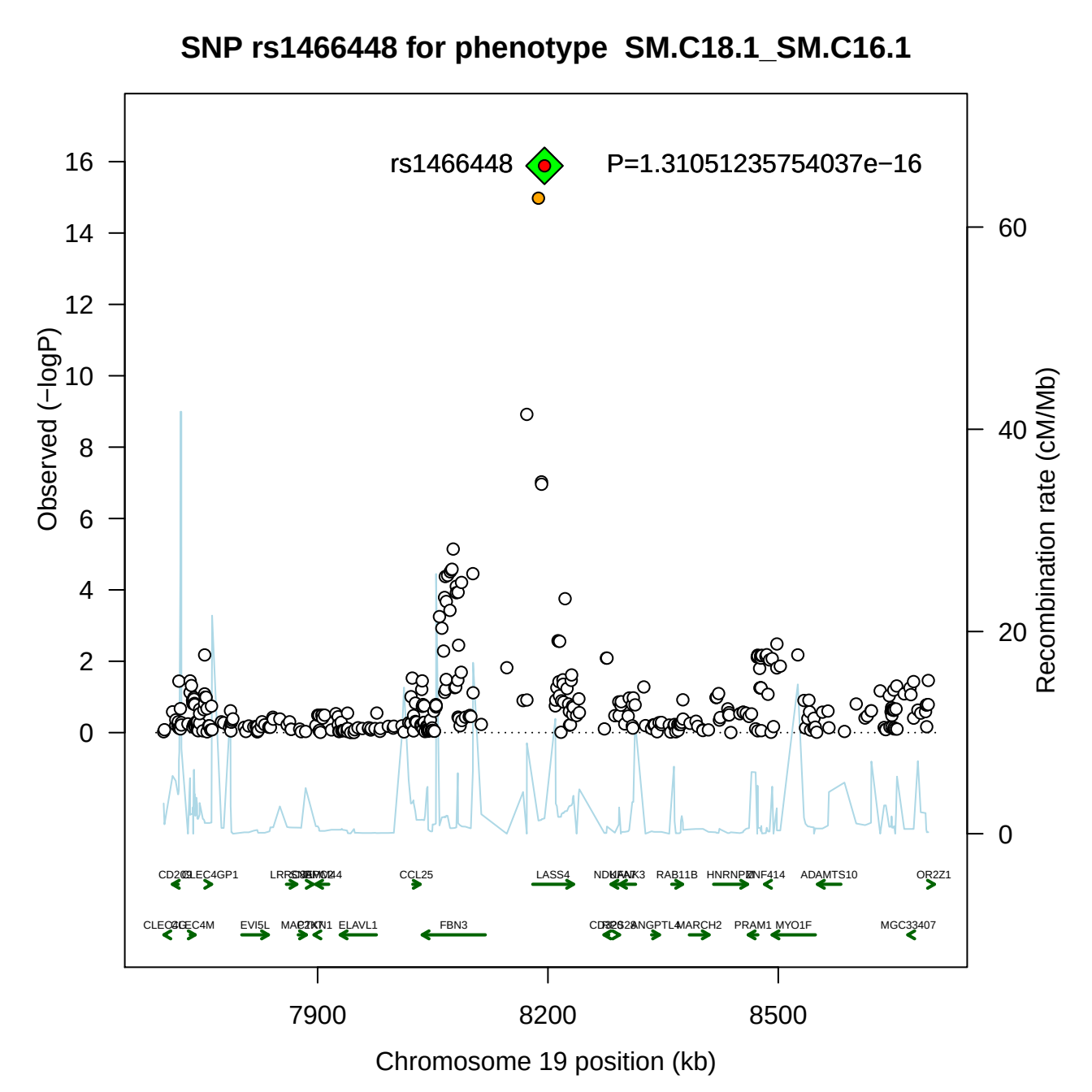

SNP rs1374804 for phenotype Ser_Gly
rs1374804
P=1.7039679923306e−12
Observed (−logP)
Recombination rate (cM/Mb)
0
2
4
6
8
10
12
0
20
40
60
SLC41A3
ROPN1B
LOC200810
LOC100125556
ALDH1L1
ZXDC
CCDC37
KLF15
C3orf22
CHST13
UROC1
TR2IT1
127100
127400
127700
Chromosome 3 position (kb)

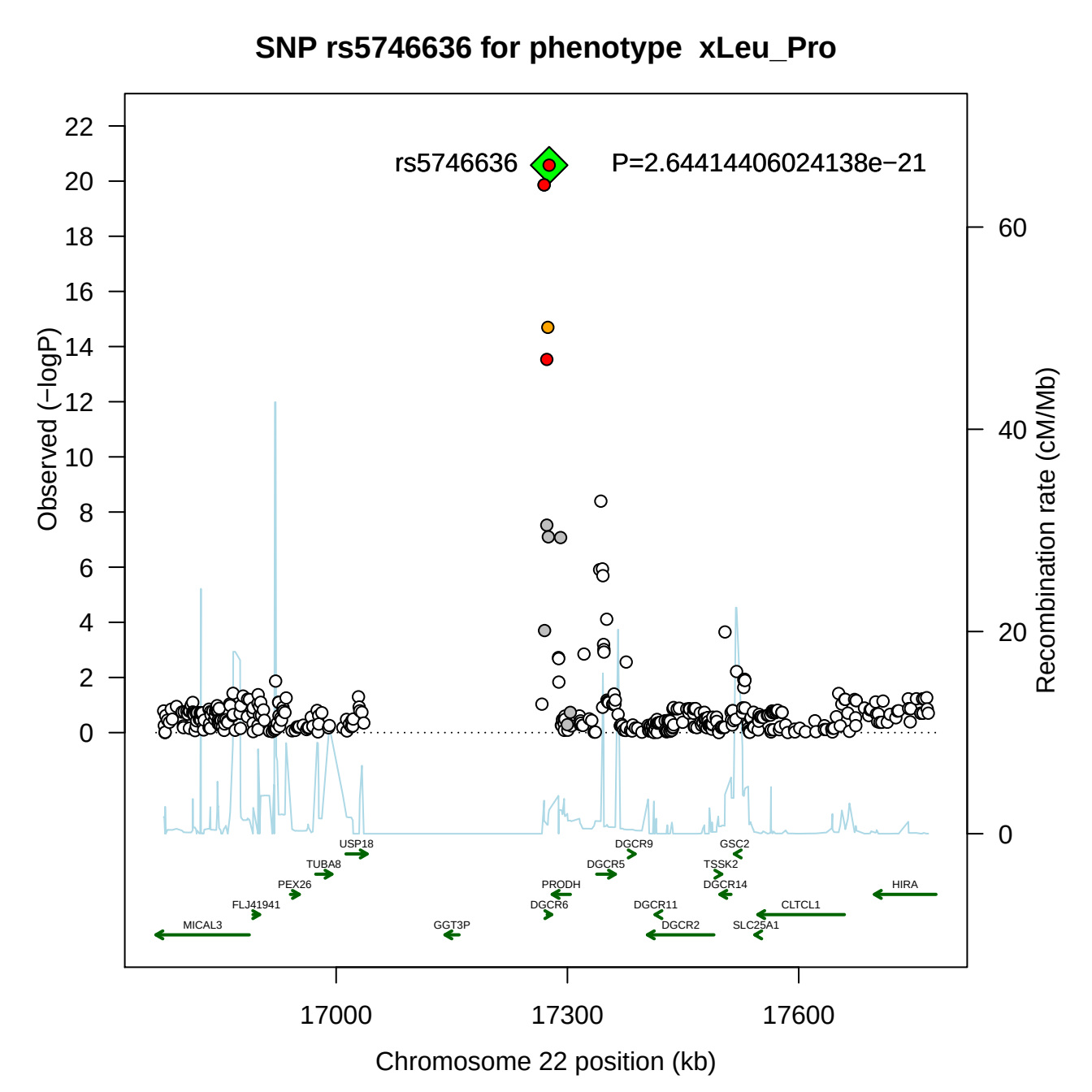

SNP rs2355078 for phenotype lysoPC.a.C28.0_PC.aa.C26.0
rs6970485
P=4.65904850757418e−41
Observed (−logP)
Recombination rate (cM/Mb)
THSD7A
TMEM106B
11500
11800
12100
Chromosome 7 position (kb)


SNP rs1894832 for phenotype Ser_Trp
rs1894832
P=4.02138135566904e−13
Observed (−logP)
Recombination rate (cM/Mb)
0
2
4
6
8
10
12
0
20
40
60
ZNF713
SNORA15
SEPT14
CCT6A
PSPHL
PSPH
CHCHD2
FKBP9L
GBAS
PHKG1
LOC442308
MRPS17
SUMF2
55800
56100
56400
Chromosome 7 position (kb)


SNP rs2657879 for phenotype His_Gln
rs2657879
P=4.44021834333146e−15
Observed (−logP)
Recombination rate (cM/Mb)
Chromosome 12 position (kb)
54800
55100
55400
LOC728937
MYL6B
COQ10A
TIMELESS
SNORD59B
RDH16
RPS26
FAM62A
ANKRD52
APOF
ATP5B
SDR-O
IKZF4
ZC3H10
SLC39A5
STAT2
BAZ2A
HSD17B6
SUOX
RPL41
OBFC2B
IL23A
RBMS2
PRIM1
RAB5B
PAZ2G4
RNF41
PAN2
GLS2
NACA
CDK2
ERBB3
SMARCC2
CNP12
SPRYD4
PTGES3
SILV
RPS26L
MYL6
CS
MIP
SNORD59A


SNP rs7200543 for phenotype PC.aa.C36.2_PC.aa.C38.3
rs7200543
P=1.39528826520685e−16
Observed (−logP)
Recombination rate (cM/Mb)
Chromosome 16 position (kb)
14700
15000
15300
PLA2G10
NPIP
RRN3
BFAR
NOMO1
NTAN1
C16orf45
PARN
ABCC6P2
PDXDC1
MPV17L


SNP rs1466448 for phenotype SM.C18.1_SM.C16.1
rs1466448
P=1.31051235754037e−16
Observed (−logP)
Recombination rate (cM/Mb)
Chromosome 19 position (kb)
7900
8200
8500
CCL25
LASS4
RAB11B
ADAMTS10
OR2Z1
EVI5L
ELAVL1
FBN3
ANGPTL4
MARCH2
PRAM1
MYO1F
MGC33407


SNP rs5746636 for phenotype xLeu_Pro
rs5746636
P=2.64414406024138e−21
Observed (−logP)
Recombination rate (cM/Mb)
Chromosome 22 position (kb)
17000
17300
17600
USP18
TUBA8
PEX26
FLJ41941
MICAL3
GGT3P
DGCR9
DGCR5
PRODH
DGCR6
DGCR11
DGCR2
GSC2
TSSK2
DGCR14
SLC25A1
CLTCL1
HIRA

File S2

Simulations

We conducted simulations for elucidate how phenotypic and genotypic noise can influence on the model of inheritance of the marker locus in case of non-additive model of casual variant. This section is divided into the two parts: phenotypic noise and genotypic correlations.

Phenotypic noise

We assumed that measured traits (in this study metabolites are measured by mass-spectrometry technology) are highly correlated with real biological phenotypes that possibly could be controlled by non-additive genes. In other words we always induce some noise to our measurements, in this case studied phenotype could be highly correlated but not the same with the phenotype that is controlled by non-additve genes. So we checked how this noise can influence onto the model of inheritance of associated variant.

For each round we simulated genotype as binomial distribution with fixed chosen frequency (we used three frequencies: 0.25, 0.5 and 0.75), than we simulated the associated phenotype that controlled by this gene. The number of individuals was set to 2000.

Liability values for original phenotype were simulated as a sum of independent quantitative trait locus (QTL), polygenic effects and environmental component. The heritability coefficient was set to 0.7. Based on its allele frequency (AF), the QTL effect was assigned in a way that the SNP was accounted for 5% of total liability variance, than we recoded genotypes regarding to the simulated model and multiply it by estimated beta. To model the polygenic effect, 50 markers were randomly simulated,

and based on their allele frequencies; effects were assigned in such way that each of the SNPs explained the same fraction of non-QTL heritability. Environmental component was estimated as norm distribution with zero mean and standard deviation (sd) equal to 0.3.

We simulated phenotypes for recessive, dominant and over-dominant models. Correlated traits were calculated as sum of original phenotype and vector of norm distribution with chosen sd equal to the value from range of fixed sd (0, 0.5, 1, 1.5, 2). Thus we gained four correlated traits and one original. 1000 simulation cycles were performed.

Results are shown in Table S1 and presented in Figure S1. In table results are shown only for phenotypes with correlation not less than 0.5.

Genotypic correlations

If the chip technology is used, SNPs with high LD to the casual variant are detected in most cases. So we checked how does a non-additive model change if we will analyse not-targeted SNP instead of the targeted. The scheme of simulation was the same as for phenotypic correlation simulations described above. Instead of simulation of correlated phenotypes we used simulation of correlated genotypes with the same fixed frequency. We first simulated the casual variant as binomial distribution with fixed frequency. Than we randomly changed genotypes to get correlated vector of genotypes, controlling the frequency and HW equilibrium, until desired value of correlation is obtained. We had three correlated genotypes (r^2 was equal to 0.7, 0.8, 0.9) and one original. Results are shown in Table S2 and presented in Figure S2.

Also we used real genotypes and simulated phenotypes. We chose locus represented by SNP rs419291 on chromosome 5 according to the regional association plot – this locus was mostly abounded by SNPs that were highly correlated with the most

associated SNP (rs419291) in our original study. In 1 Mb locus around casual variant (rs419291) SNPs with QC properties from original study were chosen within r square range from 0.39 to 1 (8 SNPs). We simulated phenotype with casual variant for recessive, additive, dominant and over-dominant model as was described before and then calculated test-statistic for each chosen SNP. Model selection was performed using minimal p-value within tested models. For each simulated model we checked the difference between negative log p-values of simulated model and other models, including the genotypic one. For each SNP and simulated model the mean and standard deviation was estimated. Results are shown in Table S3. Number of simulations cycles was equal to 100.

Results

From all scenarios of phenotypic and genotypic simulations we can judge that the difference between simulated non-additive model and additive model becomes smaller with bigger noise in case of phenotypic simulations or less LD in case of the genotypic. For our chosen simulation parameters according to LRT genotypic model becomes not significantly better than additive when the association becomes not significant for all models. For all cases the most associated model was the simulated one. As expected, the 2df genotypic test has less power in all cases than the test for simulated model but is robust compared with other models, especially for over-dominant model, which is hardly detectable by additive model.

Table S1. Results of simulations of non-additive effects for correlated with original phenotype traits. In all cells means with standard deviation are shown. Rows "–log(BM)", "–log(A)", "–log(G)" show the logarithm of p-value for simulated, additive, genotypic models, respectively. Row "LRT(A,G)" shows the p-value of LRT between additive and genotypic models.

	AF		Original	0.89+-0	0.71+-0.01	0.55+-0.02
Recessive	0.25	-log(BM)	23.4+-4.61	18.78+-4.08	12.15+-3.23	7.72+-2.49
		-log(A)	9.9+-3.09	8.05+-2.7	5.39+-2.17	3.49+-1.65
		-log(G)	22.52+-4.58	17.94+-4.05	11.41+-3.19	7.09+-2.43
		LRT(A,G)	14.53+-3.39	11.69+-3.07	7.63+-2.43	5+-1.96
	0.5	-log(BM)	23.34+-4.49	18.82+-4.02	11.97+-3.17	7.8+-2.59
		-log(A)	15.93+-3.69	12.91+-3.29	8.33+-2.57	5.47+-2.11
		-log(G)	22.46+-4.47	18+-3.99	11.24+-3.13	7.18+-2.54
		LRT(A,G)	8.41+-2.65	6.87+-2.43	4.49+-1.9	3.1+-1.62
	0.75	-log(BM)	23.48+-4.32	18.93+-3.86	12.13+-3.19	7.8+-2.48
		-log(A)	20.15+-4	16.28+-3.59	10.49+-2.94	6.76+-2.31
		-log(G)	22.59+-4.27	18.09+-3.83	11.38+-3.15	7.15+-2.42
		LRT(A,G)	4.21+-1.79	3.49+-1.64	2.38+-1.33	1.68+-1.08
Dominant	0.25	-log(BM)	23.71+-4.41	19.08+-3.98	12.24+-3.17	7.8+-2.46
		-log(A)	20.31+-4.06	16.41+-3.63	10.56+-2.9	6.8+-2.26
		-log(G)	22.82+-4.37	18.23+-3.95	11.51+-3.12	7.16+-2.41
		LRT(A,G)	4.28+-1.86	3.5+-1.68	2.43+-1.43	1.65+-1.15
	0.5	-log(BM)	23.69+-4.28	19.06+-3.95	12.13+-3.11	7.9+-2.46
		-log(A)	15.91+-3.49	12.86+-3.2	8.38+-2.57	5.56+-2.06
		-log(G)	22.81+-4.24	18.23+-3.9	11.41+-3.06	7.27+-2.4
		LRT(A,G)	8.8+-2.72	7.17+-2.44	4.62+-1.96	3.1+-1.58
	0.75	-log(BM)	23.63+-4.85	18.96+-4.23	12.12+-3.4	7.78+-2.63
		-log(A)	9.85+-3.13	8.04+-2.76	5.27+-2.21	3.51+-1.69
		-log(G)	22.75+-4.82	18.12+-4.2	11.4+-3.36	7.13+-2.58
		LRT(A,G)	14.8+-3.56	11.89+-3.18	7.73+-2.62	5.02+-2.03
Over-dominant	0.25	-log(BM)	23.66+-4.34	19.04+-3.92	12.14+-3.22	7.92+-2.58
		-log(A)	10+-2.99	8.12+-2.69	5.36+-2.15	3.65+-1.74
		-log(G)	22.78+-4.31	18.2+-3.89	11.42+-3.16	7.29+-2.51
		LRT(A,G)	14.69+-3.36	11.89+-3.06	7.66+-2.49	5.06+-2
	0.5	-log(BM)	23.58+-4.42	19.03+-3.98	11.98+-3.1	7.74+-2.45
		-log(A)	0.43+-0.43	0.45+-0.43	0.46+-0.47	0.43+-0.42
		-log(G)	22.69+-4.39	18.2+-3.94	11.26+-3.06	7.1+-2.39
		LRT(A,G)	23.58+-4.42	19.03+-3.98	11.98+-3.1	7.74+-2.45
	0.75	-log(BM)	23.5+-4.35	18.92+-3.94	12.13+-3.15	7.84+-2.54
		-log(A)	9.82+-2.89	8.02+-2.6	5.28+-2.03	3.55+-1.69
		-log(G)	22.62+-4.32	18.09+-3.91	11.4+-3.1	7.2+-2.48
		LRT(A,G)	14.7+-3.51	11.88+-3.16	7.73+-2.56	5.05+-1.98

Figure S1. The graphs of p-values for different models against the trait's correlation. For each graph three lines are presented: the red one represents the simulated model, the blue one – genotypic model, the green – additive. Horizontal line represents the threshold $5x10^{-8}$. Abbreviations r, a, d, o, g are recessive, additive, dominant and over-dominant model, respectively.

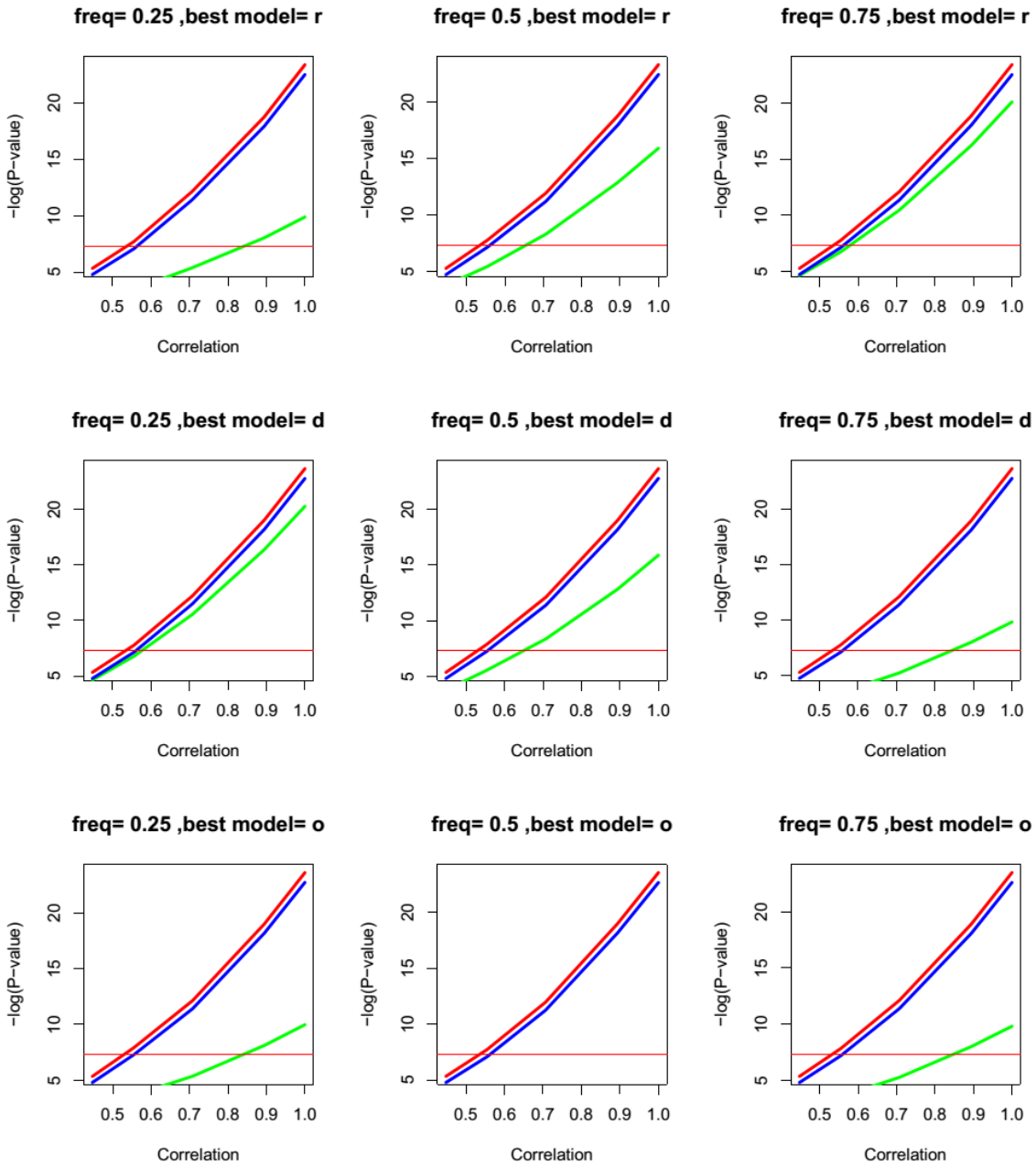


Table S2. Results of simulations of non-additive effects of SNPs, correlated with casual variant. In all cells means with standard deviation are shown. Rows "–log(BM)", "–log(A)", "–log(G)" show the logarithm of p-value for simulated, additive, genotypic models, respectively. Row "LRT(A,G)" shows the p-value of LRT between additive and genotypic models.

	AF		Target SNP	0.9	0.8	0.7
Recessive	0.25	-log(BM)	23.9+-4.63	19.78+-4.22	15.82+-3.73	12.36+-3.33
		-log(A)	9.94+-2.95	8.3+-2.69	6.77+-2.38	5.39+-2.14
		-log(G)	23+-4.59	18.92+-4.2	15.01+-3.68	11.62+-3.27
		LRT(A,G)	14.97+-3.44	12.44+-3.19	9.97+-2.77	7.84+-2.5
	0.5	-log(BM)	23.57+-4.34	19.53+-3.86	15.64+-3.49	12.18+-3.14
		-log(A)	15.95+-3.51	13.28+-3.13	10.73+-2.85	8.4+-2.49
		-log(G)	22.67+-4.3	18.67+-3.82	14.83+-3.46	11.45+-3.1
		LRT(A,G)	8.61+-2.59	7.2+-2.34	5.81+-2.16	4.64+-1.97
	0.75	-log(BM)	23.44+-4.27	19.36+-3.9	15.64+-3.68	12.1+-3.23
		-log(A)	20.16+-3.93	16.65+-3.55	13.53+-3.4	10.5+-2.97
		-log(G)	22.57+-4.22	18.53+-3.85	14.85+-3.64	11.36+-3.17
		LRT(A,G)	4.18+-1.83	3.56+-1.72	2.91+-1.49	2.34+-1.3
Dominant	0.25	-log(BM)	23.6+-4.35	19.56+-3.88	15.66+-3.62	12.31+-3.11
		-log(A)	20.37+-4.04	16.92+-3.59	13.57+-3.4	10.72+-2.91
		-log(G)	22.7+-4.31	18.7+-3.85	14.87+-3.59	11.58+-3.07
		LRT(A,G)	4.1+-1.76	3.47+-1.62	2.89+-1.49	2.34+-1.33
	0.5	-log(BM)	23.57+-4.33	19.51+-3.89	15.63+-3.7	12.18+-3.08
		-log(A)	15.98+-3.6	13.32+-3.27	10.68+-3.05	8.38+-2.52
		-log(G)	22.69+-4.3	18.68+-3.86	14.84+-3.66	11.45+-3.04
		LRT(A,G)	8.6+-2.69	7.16+-2.42	5.87+-2.18	4.66+-1.98
	0.75	-log(BM)	23.51+-4.56	19.41+-4.16	15.68+-3.8	12.26+-3.34
		-log(A)	9.92+-3.05	8.34+-2.81	6.77+-2.45	5.43+-2.13
		-log(G)	22.64+-4.54	18.58+-4.14	14.89+-3.77	11.5+-3.3
		LRT(A,G)	14.62+-3.44	12.06+-3.17	9.84+-2.91	7.68+-2.52
Over-dominant	0.25	-log(BM)	23.61+-4.07	19.61+-3.88	15.66+-3.53	12.22+-3.11
		-log(A)	9.84+-2.78	8.31+-2.63	6.75+-2.29	5.35+-2.04
		-log(G)	22.72+-4.05	18.76+-3.86	14.86+-3.49	11.48+-3.06
		LRT(A,G)	14.78+-3.35	12.28+-3.13	9.82+-2.84	7.74+-2.52
	0.5	-log(BM)	23.47+-4.35	19.46+-4.06	15.62+-3.57	12.18+-3.15
		-log(A)	0.43+-0.44	0.43+-0.44	0.44+-0.45	0.43+-0.43
		-log(G)	22.58+-4.33	18.62+-4.03	14.83+-3.53	11.44+-3.11
		LRT(A,G)	23.47+-4.35	19.47+-4.06	15.62+-3.56	12.18+-3.15
	0.75	-log(BM)	23.47+-4.23	19.48+-3.83	15.68+-3.63	12.14+-3.17
		-log(A)	9.86+-2.76	8.28+-2.53	6.72+-2.27	5.31+-2.09
		-log(G)	22.58+-4.19	18.63+-3.79	14.88+-3.58	11.4+-3.12
		LRT(A,G)	14.62+-3.53	12.16+-3.16	9.88+-2.94	7.7+-2.48

Figure S2. The graphs of p-values for different models against the genotype's correlation. For each graph three lines are presented: the red one represents the simulated model, the blue one – genotypic model, the green – additive. Horizontal line represents the threshold $5x10^{-8}$. Abbreviations r, a, d, o, g are recessive, additive, dominant and over-dominant model, respectively.

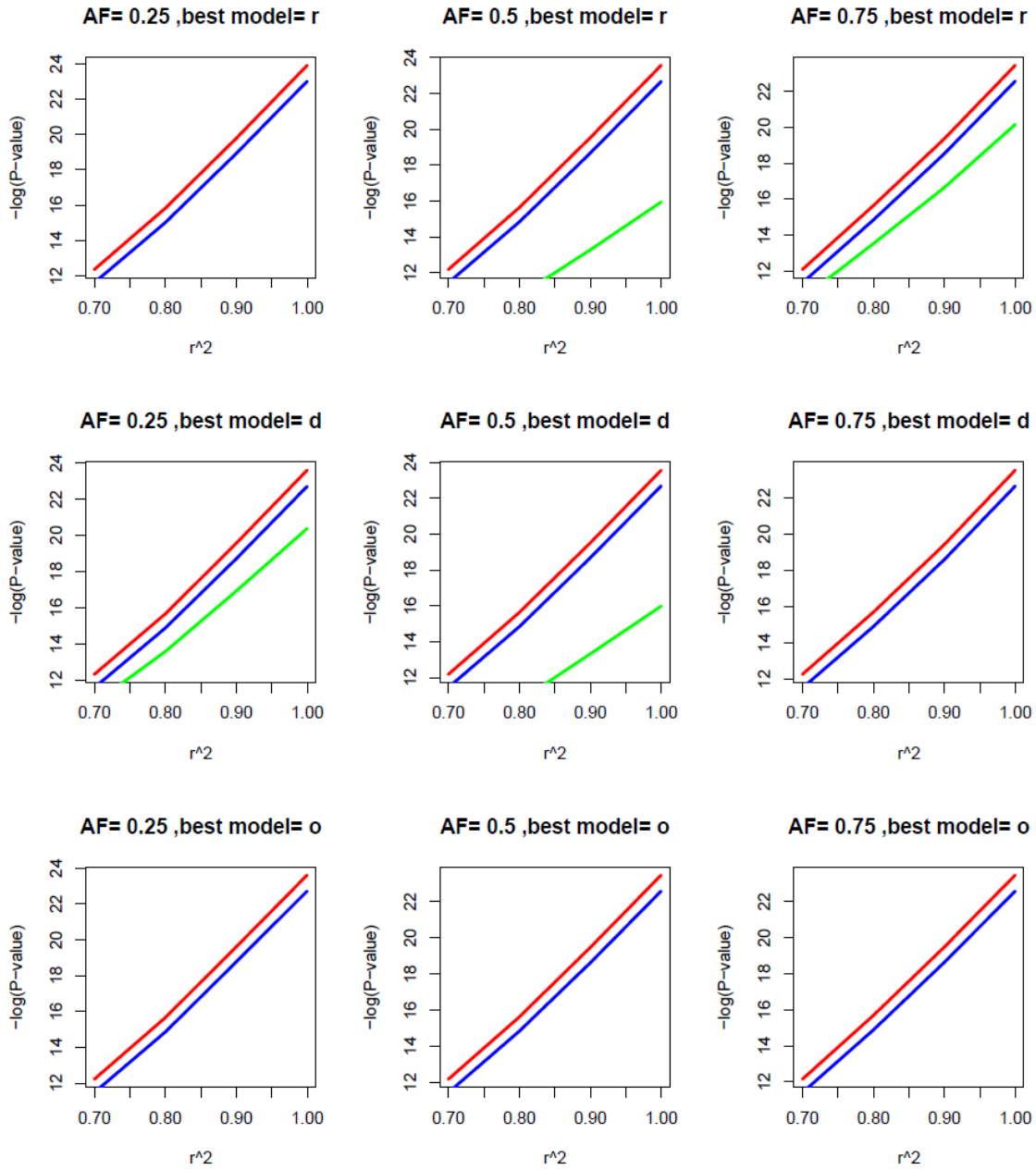


Table S3. Results of simulations of non-additive effects of SNPs, correlated with casual variant, using real genotype data. Table presents the difference in negative log p-values of simulated model and other 1 df and 2 df models. Eight SNPs were chosen according to their r^2 with casual variant (r^2=1). For each simulated model (denoted as R, A, D, O for recessive, additive, dominant and over-dominant) negatve log p-values of association test under the simulated model (denoted as r, a, d, o), and differences with other models (denoted as da, dr, dd, do, dg for difference with additive model, recessive, dominant, over-dominant and genotypic respectively).

		R^2							
Simulated model		1	0.98	0.92	0.82	0.76	0.6	0.51	0.39
R	r	20.8+-4.16	20.23+-4.09	11.76+-3.13	19.04+-4.22	17.23+-3.89	6.71+-2.17	6.02+-1.98	5.82+-2.03
	da	8.25+-2.58	8.02+-2.54	3.12+-2.18	7.49+-2.71	6.55+-2.73	0.83+-1.59	0.59+-1.45	0.57+-1.42
	dd	17.37+-3.85	17.01+-3.76	8.05+-2.92	16.21+-4.02	14.46+-3.9	3.81+-2.12	3.29+-1.89	3.29+-1.85
	do	17.28+-3.94	16.81+-3.9	11.28+-3.22	15.22+-3.81	14.1+-3.46	6.3+-2.23	5.64+-2.04	5.43+-2.04
	dg	0.88+-0.28	0.86+-0.28	0.21+-0.69	0.83+-0.29	0.76+-0.36	0.06+-0.71	0.03+-0.71	0.05+-0.64
A	a	6.03+-2.27	5.84+-2.19	4.4+-1.85	5.36+-2.07	5.25+-2.08	3.09+-1.58	2.85+-1.45	2.8+-1.39
	dr	2.17+-1.4	2.14+-1.37	1.98+-1.37	1.89+-1.31	1.97+-1.3	1.32+-1.21	1.23+-1.04	1.22+-1.03
	dd	1.25+-0.96	1.3+-0.94	0.69+-0.8	1.3+-0.89	1.22+-0.93	0.45+-0.73	0.43+-0.64	0.43+-0.67
	do	5.03+-2.14	4.91+-2.07	2.85+-1.61	4.64+-1.97	4.43+-2.02	2.01+-1.4	1.88+-1.3	1.91+-1.27
	dg	0.6+-0.27	0.6+-0.27	0.5+-0.25	0.59+-0.25	0.58+-0.27	0.38+-0.3	0.41+-0.25	0.41+-0.26
D	d	19.47+-3.61	18.87+-3.62	12.7+-2.89	16.69+-3.41	16.37+-3.31	8.48+-2.57	7.6+-2.33	7.15+-2.23
	da	4.21+-1.87	3.96+-1.85	1.86+-1.31	3.36+-1.82	3.07+-1.78	0.88+-1.24	0.67+-1.12	0.48+-1.18
	dr	16.19+-3.33	15.83+-3.37	10.65+-2.72	13.82+-3.25	13.55+-3.13	6.49+-2.56	5.72+-2.19	5.21+-2.18
	do	8.96+-2.74	8.79+-2.65	4.13+-1.77	8.5+-2.45	8.23+-2.45	3.6+-1.54	3.28+-1.53	3.45+-1.52
	dg	0.85+-0.24	0.84+-0.25	0.76+-0.28	0.79+-0.3	0.76+-0.34	0.57+-0.4	0.53+-0.41	0.44+-0.49

O	o	18.01+-3.74	17.19+-3.65	8.7+-2.57	15.25+-3.36	14.29+-3.07	4.46+-1.58	3.78+-1.37	3.51+-1.35
	dr	14.84+-3.69	14.01+-3.62	6.69+-2.58	12.2+-3.19	11.6+-2.9	3.61+-1.6	3.04+-1.38	2.82+-1.29
	dd	8.13+-2.55	7.44+-2.41	3.46+-1.53	6.54+-2.42	5.63+-2.36	1.25+-1.15	0.91+-1.06	0.76+-1.07
	da	15.66+-3.44	14.83+-3.29	6.81+-2.24	13.29+-3.2	12.09+-3.05	3.02+-1.51	2.4+-1.41	2.2+-1.44
	dg	0.81+-0.33	0.8+-0.34	0.5+-0.48	0.76+-0.39	0.72+-0.42	0.47+-0.37	0.43+-0.35	0.4+-0.36

File S3

Results of testing for recessive, dominant, over-dominant models and MAX-test

Additionally to overall calculations we calculated statistic for recessive, dominant and over-dominant models separately. But even if we use liberal threshold (5e-8/22801) instead of strict (5e-8/(22801*4)), nothing additionally could be found, except one SNP on seventh chromosome rs219040 using over-dominant model (p-value<3.94E-13) that was associated with C5.1/C6.1 ratio. This SNP located close to the gene STEAP2-AS1 (non-protein coding antisense RNA1 gene) that hardly relates to the metabolism control. Its p-value of HW was close to the threshold (p-value< 1.03E-05) and it was not replicated (p-value<8.37E-01), so we further not considered this SNP into overall score of non-additive effects.

Table S1. Results for GWAS with recessive model. This table reports the results for fourteen loci that had a significant p-value for the recessive model in KORA F4 (P-value <2.19x10^{-12}).

chr: chromosome; AF – allele frequency of the effect allele.

				KORA sample		TwinsUK sample		
SNP	**metabolite (ratio)**	**chr**	**position**	**AF**	**p-value recessive**	**AF**	**p-value recessive**	**gene**
rs11161521	C8/C12	1	75,988,918	0.70	2.74E-62	0.69	3.98E-28	*ACADM*
rs7558218	C9/PC.ae.C30.0	2	210,811,690	0.36	3.25E-61	0.35	2.94E-23	*ACADL*
rs7422339	Gly/Gln	2	211,248,752	0.69	7.37E-75	-	-	*CPS1*
rs8396	C7.DC/C10	4	159,850,267	0.71	3.75E-23	0.68	1.92E-17	*PPID*
rs2046813	PC.ae.C42.5/PC.ae.C44.5	4	186,006,153	0.69	8.14E-14	0.69	1.88E-03	*SLED1*
rs273913	C5/PC.ae.C34.1	5	131,689,055	0.41	4.00E-14	0.35	8.12E-02	*SLC22A4*
rs3798723	PC.aa.C42.5/PC.aa.C40.3	6	11,149,706	0.75	4.76E-26	-	-	*ELOVL2*
rs603424	C14/C16.1	10	102,065,469	0.80	9.86E-15	0.82	1.53E-02	*PKD2L1*
rs174547	PC.aa.C36.3/PC.aa.C36.4	11	61,327,359	0.70	1.02E-145	0.65	8.09E-44	*FADS1*
rs2066938	C4/C3	12	119,644,998	0.27	5.22E-99	0.26	7.29E-39	*ACADS*
rs7156144	PC.ae.C32.1/PC.ae.C34.1	14	67,049,466	0.59	1.14E-27	0.57	6.80E-14	*SGPP1*
rs1741	PC.aa.C38.3/PC.aa.C36.2	16	15,037,852	0.69	1.98E-13	0.72	2.07E-06	*NTAN1*
rs364585	SM..OH..C24.1/SM.C24.0	20	12,910,718	0.64	9.12E-28	0.59	1.02E-12	*SPTLC3*
rs5747922	xLeu/Pro	22	17,269,755	0.77	6.63E-19	0.73	6.18E-03	*DGCR6*

Table S2. Results for GWAS with dominant model. This table reports the results for eighteen loci that had a significant p-value for the dominant model in KORA F4 (P-value <2.19x10^{-12}).

chr: chromosome; AF – allele frequency of the effect allele.

				KORA sample		TwinsUK sample		
SNP	**metabolite (ratio)**	**chr**	**position**	**AF**	**p-value dominant**	**AF**	**p-value dominant**	**gene**
rs7552404	C8/C12	1	75,908,534	0.30	5.48E-64	0.31	4.02E-28	*ACADM*
rs7601356	C9/PC.ae.C30.0	2	210,764,902	0.63	2.78E-64	0.65	1.43E-23	*ACADL*
rs2216405	Gly/Gln	2	211,325,139	0.19	8.84E-40	0.16	1.15E-19	*CPS1*
rs12505475	C7.DC/C10	4	159,854,694	0.29	4.37E-23	0.33	1.51E-17	*PPID*
rs4862429	PC.ae.C42.5/PC.ae.C44.5	4	186,006,834	0.31	1.25E-13	0.31	1.65E-03	*SLED1*
rs270605	C5/PC.ae.C34.1	5	131,679,710	0.60	4.71E-14	0.65	8.17E-02	*SLC22A4*
rs3798719	PC.aa.C42.5/PC.aa.C40.3	6	11,144,811	0.25	3.73E-26	0.23	1.24E-03	*ELOVL2*
rs6970485	PC.aa.C26.0/PC.ae.C38.1	7	11,752,704	0.35	2.33E-17	-	-	*THSD7A*
rs12356193	C0	10	61,083,359	0.17	4.51E-25	0.16	4.25E-08	*SLC16A9*
rs174556	PC.aa.C36.3/PC.aa.C36.4	11	61,337,211	0.27	2.78E-144	0.32	4.65E-46	*FADS1*
rs1043011	Gln/Met	12	55,151,307	0.21	4.03E-13	0.19	4.67E-04	*GLS2*
rs3916	C3/C4	12	119,661,655	0.73	2.46E-97	0.75	4.07E-36	*ACADS*
rs4902243	PC.aa.C28.1/PC.ae.C40.2	14	63,303,996	0.17	3.66E-36	0.14	4.24E-17	*SGPP1*
rs1077989	PC.ae.C32.1/PC.ae.C34.1	14	67,045,575	0.46	3.60E-35	0.47	3.99E-17	*PLEKHH1*
rs7200543	PC.aa.C36.2/PC.aa.C38.3	16	15,037,471	0.31	2.14E-15	0.28	1.47E-06	*NTAN1*
rs1466448	SM.C16.1/SM.C18.1	19	8,195,519	0.22	1.45E-13	0.19	1.75E-10	*CERS4*
rs4814176	SM..OH..C24.1/SM.C24.0	20	12,907,398	0.36	6.70E-28	0.42	3.69E-13	*SPTLC3*
rs5746636	xLeu/Pro	22	17,276,301	0.24	3.80E-19	0.27	4.62E-03	*DGCR6*

Table S3. Results for GWAS with over-dominant model. This table reports the results for eleven loci that had a significant p-value for the over-dominant model in KORA F4 (P-value <2.19×10^{-12}).

chr: chromosome; AF – allele frequency of the effect allele.

				KORA sample		TwinsUK sample		
SNP	**metabolite (ratio)**	**chr**	**position**	**AF**	**p-value**	**AF**	**p-value**	**gene**
rs7365179	C10/C12	1	76,096,212	0.22	1.74E-30	0.24	5.01E-15	*ACADM*
rs12468576	C5.M.DC/C9	2	210,662,236	0.20	2.07E-14	0.22	1.65E-04	*ACADL*
rs7422339	Gly/Gln	2	211,248,752	0.69	3.26E-42	-	-	*CPS1*
rs3756963	PC.aa.C42.6/PC.aa.C38.5	6	11,130,140	0.76	1.13E-13	-	-	*ELOVL2*
rs6970485	lysoPC.a.C28.0/PC.aa.C26.0	7	11,752,704	0.35	3.47E-29	-	-	*THSD7A*
rs2190401	C5.1/C6.1	7	89,504,946	0.76	3.94E-13	0.78	8.37E-01	*STEAP2-AS1*
rs12356193	C0	10	61,083,359	0.17	9.27E-18	0.16	3.15E-06	*SLC16A9*
rs968567	PC.aa.C36.3/PC.aa.C36.4	11	61,352,140	0.15	6.05E-58	0.19	9.20E-13	*FADS1*
rs12310160	C3/C4	12	119,584,265	0.86	2.02E-26	0.85	2.76E-10	*ACADS*
rs7157785	PC.ae.C40.2/PC.aa.C28.1	14	63,305,309	0.17	2.08E-30	0.17	3.59E-12	*SGPP1*
rs4508668	SM.C24.0/SM..OH..C24.1	20	12,903,601	0.32	2.64E-13	0.37	2.06E-04	*SPTLC3*

Table S4. Results of MAX test for loci found by genotypic model.

chr: chromosome; AF – allele frequency of the effect allele. MAX_KORA and MAX_TUK – p-values for MAX test for KORA and TwinsUK respectively.

SNP	Trait	chr	Pos	Freq_KORA	g_pval_KORA	Freq_TUK	g_pval_TUK	MAX_KORA*	MAX_TUK*
rs7552404	C12/C10	1	75,908,534	0.300	1.69E-72	0.314	1.89E-29	0.00E+00	0.00E+00
rs7601356	C9/PC.ae.C30.0	2	210,764,902	0.632	1.24E-70	0.649	6.86E-28	0.00E+00	0.00E+00
rs715	Gly/Gln	2	211,251,300	0.687	4.28E-69	0.703	1.12E-48	0.00E+00	0.00E+00
rs8396	C7.DC/C10	4	159,850,267	0.707	5.98E-26	0.678	3.14E-17	0.00E+00	0.00E+00
rs2046813	PC.ae.C42.5/PC.ae.C44.5	4	186,006,153	0.688	6.29E-17	0.687	1.18E-03	0.00E+00	4.46E-04
rs273913	C5/PC.ae.C34.1	5	131,689,055	0.405	1.60E-16	0.351	4.19E-02	0.00E+00	2.21E-02
rs3798719	PC.aa.C42.5/PC.aa.C40.3	6	11,144,811	0.248	5.01E-32	0.234	4.01E-04	0.00E+00	1.11E-04
rs12356193	C0	10	61,083,359	0.166	2.18E-27	0.161	1.20E-07	0.00E+00	7.82E-06
rs603424	C16.1/C14	10	102,065,469	0.801	3.70E-18	0.818	1.99E-02	0.00E+00	1.47E-02
rs174547	PC.aa.C36.3/PC.aa.C36.4	11	61,327,359	0.701	2.29E-208	0.649	2.09E-76	0.00E+00	0.00E+00
rs2066938	C3/C4	12	119,644,998	0.270	1.73E-159	0.257	2.17E-67	0.00E+00	0.00E+00
rs4902242	PC.aa.C28.1/PC.ae.C40.2	14	63,299,842	0.849	2.00E-35	0.872	4.78E-15	0.00E+00	2.46E-07
rs1077989	PC.ae.C32.1/PC.ae.C34.1	14	67,045,575	0.463	6.80E-42	0.472	4.05E-18	0.00E+00	0.00E+00
rs4814176	SM..OH..C24.1/SM..OH..C22.1	20	12,907,398	0.364	2.69E-31	0.416	9.69E-09	0.00E+00	5.96E-09
rs6970485	lysoPC.a.C28.0/PC.aa.C26.0	7	11,752,704	0.354	1.21E-47	-	-	0.00E+00	9.82E-01
rs1894832	Ser/Trp	7	56,144,740	0.508	1.98E-12	0.511	4.02E-03	3.32E-13	1.97E-03
rs2657879	His/Gln	12	55,151,605	0.207	2.89E-14	0.186	1.90E-06	2.58E-13	7.33E-03
rs7200543	PC.aa.C36.2/PC.aa.C38.3	16	15,037,471	0.312	7.45E-16	0.277	1.66E-06	0.00E+00	5.65E-07
rs1466448	SM.C18.1/SM.C16.1	19	8,195,519	0.222	7.01E-16	0.194	3.90E-10	1.78E-15	1.88E-08
rs5746636	xLeu/Pro	22	17,276,301	0.236	2.98E-20	0.273	2.40E-03	0.00E+00	9.81E-04

*Value 0.00E+00 means that p-value<1e-15