Multiomic integration analysis identifies atherogenic metabolites mediating between novel immune genes and cardiovascular risk: Additional file 1

Authors

Robert Carreras-Torres[1,2†], Iván Galván-Femenía[3,4†], Xavier Farré[4,5], Beatriz Cortés[4,5], Virginia Díez-Obrero[2,6], Anna Carreras[4], Ferran Moratalla-Navarro[2,6,7,8], Susana Iraola-Guzmán[4,5], Natalia Blay[4,5], Mireia Obón-Santacana[2,6,7], Víctor Moreno[2,6,7,8,#], Rafael de Cid[4,5,#]

[1]Digestive Diseases and Microbiota Group, Girona Biomedical Research Institute (IDIBGI), Salt, 17190, Girona, Spain.

[2]ONCOBELL Program, Bellvitge Biomedical Research Institute (IDIBELL), L'Hospitalet de Llobregat, 08908, Barcelona, Spain.

[3]Institute for Research in Biomedicine (IRB Barcelona), The Barcelona Institute for Science and Technology, Barcelona, Spain.

[4]Genomes For Life - GCAT lab, CORE Program, Germans Trias i Pujol Research Institute (IGTP), Can Ruti Campus, Badalona, Barcelona, Spain.

[5]Grup de Recerca en Impacte de les Malalties Cròniques i les seves Trajectòries (GRIMTra) (IGTP), Badalona, Spain.

[6]Unit of Biomarkers and Susceptibility (UBS), Oncology Data Analytics Program (ODAP), Catalan Institute of Oncology (ICO), L'Hospitalet del Llobregat, 08908, Barcelona, Spain.

[7]Consortium for Biomedical Research in Epidemiology and Public Health (CIBERESP), 28029, Madrid, Spain.

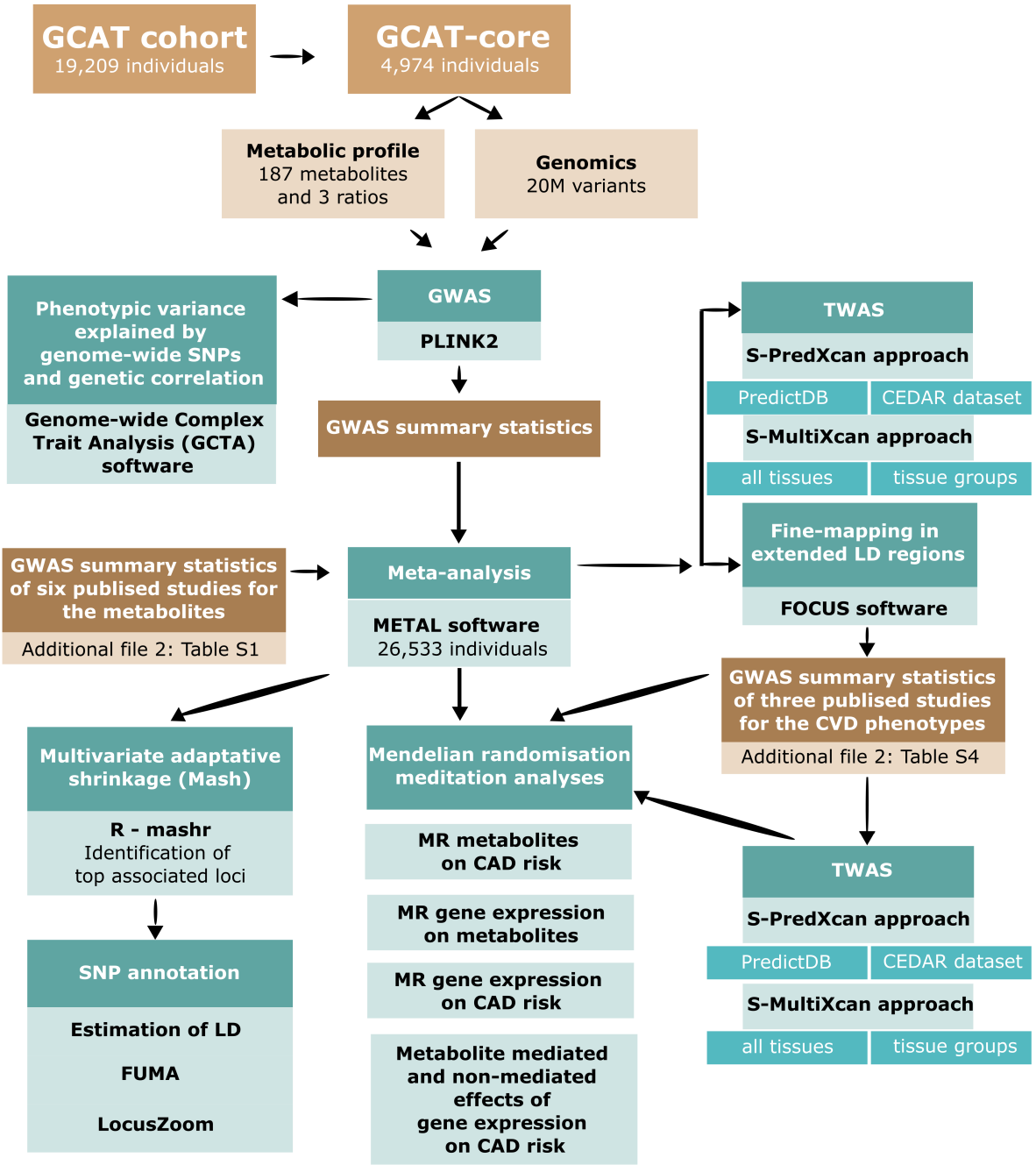
[8]Department of Clinical Sciences, University of Barcelona, Barcelona, Spain.

[†]These authors contributed equally to this work: Robert Carreras-Torres, Iván Galván-Femenía

[#]Corresponding authors: rdecid@igtp.cat; v.moreno@iconcologia.net

Workflow diagram of the performed analyses

GWAS: Genome-wide association study. LD: Linkage disequilibrium. FUMA: Functional mapping and association. TWAS: Transcriptome-wide association study. MR: Mendelian randomisation.



GCAT Metabolome characterization - Analytical Methods

Selected plasma samples were stored at -80ºC until they were analyzed. Absolute quantification for 192 metabolites, lipids and lipoproteins were quantified by GC-MS (30), LC-MS (121), NMR (14, lipidomics) and NMR (27, lipoproteins). Sample extraction and profiling was performed by the Centre for Omic Sciences (COS) Joint Unit of the Universitat Rovira i Virgili-Eurecat, as follows.

Sample extraction

For **lipoprotein and chylomicron characterization** by NMR, 200 μl of plasma samples were mixed with 300 μL of phosphate buffer (50 mM, pH 7.4) and 50 μl of D_2O in 96-well format plates without any further extraction process (final concentration around 10% deuterium oxide (D_2O)). This mixture was transferred to a 5 mm NMR tube and kept at 4ºC in the NMR sample changer (Sample Jet®, Bruker®) until their analysis.

For **metabolite extraction**, the Bravo Automated Liquid Handling Platform from Agilent Technologies was used to extract plasma samples in 96-well format plates. Aqueous extractions of 250 μl of plasma were performed with a methanol/water solution (8:1, v,v) for GC-MS analysis and lipidic extractions of 100 μl of plasma were performed by a biphasic extraction with methanol/methyl-*tert*-butyl ether (MTBE) for NMR and LC-MS analysis. In both cases, solvents were added automatically to the samples and after the appropriate shaking and centrifugation steps; the supernatants were dispensed in 96-well plates and stored until analyzed by GC-MS, LC-MS and NMR. Internal standards for GC-MS and LC-MS were previously dispensed by the robot to the same plates where supernatants were collected.

Sample analysis

For **GC-MS analysis**, were more polar metabolites were determined, a protein precipitation extraction will be made by adding 5.6 volumes of MeOH:H_2O (8:1) to a volume of 100 μl of plasma. This mixture is stirred and centrifuged. Then, the supernatants were dispensed in new 96 well plates that contain internal standard mixture. This plate that contains the extracts and internal standards is evaporated to dryness with a vacuum centrifugation system (Sped Vac), then a double derivatization of the metabolites with methoxiamine and trimethylsilil was performed to increase the volatilization of the compounds and to protect the alcohol, amine, carbonyl and carboxylic acid groups. Finally, the sample was analyzed by gas chromatography coupled to a time of flight high resolution mass spectrometry (GC-MS)

For **LC-MS analysis and NMR analyses**, where lipid metabolites were determined, a liquid-liquid extraction was made by adding a ternary solvent system (methanol/methyl-*tert*-butyl ether/water). These solvents are automatically and sequentially added to a volume of 100 μl of plasma with agitation stages between them and final centrifugation to promote phase separation. Then, for (1) lipidomic analysis by ultra-high resolution liquid chromatography (UHPLC) coupled to a time of flight high resolution mass spectrometry (LC-MS), a small aliquot of the supernatant (organic phase) was dispensed and diluted 1:10 with methanol in a new 96-well plates containing deuterated internal standards for each family of lipids (Lipidomix SPLASH from Avanti Polar Lipid), and for (2) lipidomic analysis by Nuclear Magnetic Resonance (NMR), a second aliquot of the supernatant (organic phase) was dispensed in new 96 well plates that was evaporated to dryness with a vacuum centrifugation system. Afterwards, they were reconstituted with a solution of CD_3Cl: CD_3OD with 4% D_2O and 0.01% TMS (0.067 mM, Eretic Signal 6,166 mM) and analyzed by proton NMR (1H-NMR).

Metabolome profiling

Lipoprotein NMR. 1H NMR spectra were recorded at 310 K on a Bruker Avance III 600 spectrometer operating at a proton frequency of 600.20 MHz using a 5 mm PBBO broadband gradient probe. A set of 1H NMR spectra were acquired on each sample consisting on: A standard one-dimensional (1D) spectrum for all samples (n=5200), a two-dimensional (2D) diffusion experiment for the initial subset of 200 samples that was used to build a model for analysis of chylomicrons (n=200) and 1D diffusion-edited spectrum for all samples (n=5200).

For 1D Standard Spectra, One-dimensional 1H pulse experiments were carried out using a nuclear Overhauser effect spectroscopy (NOESY)-presaturation sequence (RD-90º– τ 1–90º– τ m –90º ACQ) to suppress

the residual water peak. τ 1 (recycle delay time) was set to 4μs and τm (mixing time) was 100 ms. The 90° pulse length was calibrated for each sample. The result of the 90° pulse calibration for each sample was transferred to the other NMR diffusion experiments. The spectral width was 30 ppm, and a total of 64 transients were collected into 64K data points for each 1H spectrum.

For 2D NMR diffusion experiment (diffusion measurement), the relaxation delay was set to 2s, the FIDs were collected into 64K complex data points, and 16 scans were acquired on each transient. The gradient pulse strength was increased from 5 to 95% of the maximum strength of the probe (53.5 G cm-1 -0.535 T m-1) in 16 steps, where the linear gradient pulse strength was linearly distributed. A diffusion time (Δ) of 150 ms and bipolar sine shaped gradient pulses of length 12 ms (δ) were applied to obtain a reasonable amount of diffusion of the lipoproteins and chylomicrons signals in the raw serum. The total experiment time was 17 min per sample.

For 1D NMR diffusion experiment (diffusion calculation), 1D diffusion spectra were measured using a diffusion-edited pulse sequence with bipolar gradients and the longitudinal eddy-current delay (LED) scheme with two spoil gradients (ledbpgp2s1d: RD–90º–G1–180º–G1–90º–G2– Δ –90º–G1–180º–G1–90º–G2– τ –90º–acquire FID) and water suppression achieved through irradiation of the water signal during the recycle delay (RD). The parameters used for each experiments were the following: 16 scans, 4 dummy scans, a spectrum width of 20.0 ppm, 32K time and frequency domain data points, a line broadening of 1 Hz, a RD of 2 s, a diffusion time (Δ) of 150 ms, a gradient strength (G1) of 34.5 Gauss per cm, and an eddy current delay (τ) of 50 ms. All NMR spectra were phased, baseline-corrected and referenced using TopSpin software (version 3.2, Bruker®).

Characterization of the lipoprotein profile

The analysis of the lipoprotein profile was based on the measurement of diffusion by NMR as described by Liu et al. Briefly, the method measured the different methyl lipoprotein peaks overlapped in the NMR spectrum, which were afterwards deconvoluted mathematically to obtain the area of each lipoprotein subtraction. The deconvolution was performed using the calculated NMR-measured diffusion coefficients corresponding to each NMR sub-peak and, given the relationship between diffusion coefficient and particle hydrodynamic radius, it was possible to assign each peak to a lipoprotein fraction. The size of lipoprotein subtractions was calculated based on the distribution of diffusion coefficient.

Chylomicron analysis. The measurement of NMR diffusion allowed distinguishing and measuring the fraction of chylomicrons due to their different size from the rest of lipoproteins. Chylomicron larger size resulted in a signal at a higher chemical shift (0.853 ppm) and a very small diffusion value (< 1x10-7 cm2 s-1). Samples with very high TG values were deconvoluted and the amount of TG present in the chylomicron fraction was calculated.

Lipid 1H-NMR profiling. 1H NMR spectra were recorded at 300K on an Avance III 600 spectrometer (Bruker®, Germany) operating at a proton frequency of 600.20 MHz using a 5 mm PBBO gradient probe. Lipidic samples were measured and recorded in procno 11 using a simple presaturation sequence (RD–90°–ACQ zgpr pulse program in Bruker®) to eliminate the residual water moisture of deuterated methanol. Solvent presaturation with irradiation power of 50 Hz was applied during recycling delay (RD = 5s) and mixing time. The 90° pulse length was calibrated for each sample and varied from 10.42 to 11.49 ms. The spectral width was 12kHz (20ppm), and a total of 64 transients were collected into 64k data points for each 1H spectrum. The exponential line broadening applied before Fourier transformation was of 0.3Hz. The frequency domain spectra were manually phased and baseline-corrected using TopSpin software (version 2.1, Bruker). After pre-processing and visual checking of NMR dataset, specific 1H regions of diacylglycerols, triglycerides and total lipids based on terminal methyl and methylene signals were identified by in the spectra using a comparison into AMIX 3.9 software. Curated identified regions across the spectra were integrated using the same AMIX 3.9 software package and exported to Excel spreadsheet in order to give absolute concentrations.

Lipid LC-MS profiling and quantification. The lipid species present in plasma samples were determined by ultra-high performance chromatography coupled to quadrupole-time of flight high resolution mass spectrometry UHPLC-qTOF from Agilent Technologies (6550). The ionization was performed in positive electrospray and mass calibration reference was used along all the analyses to maintain the mass accuracy

below 5 ppm. Lipids were separated on C18 reversed phase column (Kinetex C18-EVO from Phenomenex) and a ternary mobile phase (water/methanol/2-propanol) was used. The quantification of each lipid was made by an internal standard calibration method using one analytical standard and one deuterated internal standard for each lipid family (lysophosphatidylcholines, phosphatidylcholines, phosphatidylethanolamines and triglycerides).

Polar GC-MS profiling and quantification. Sample was analyzed in a 7890A Series gas chromatograph coupled to a 7000 GC-qTOF from Agilent Technologies. Chromatographic column was a J&W Scientific HP5-MS (30 m x 0.25 mm i.d., 0.25 μm film) (Agilent Technologies), and helium 99.999% purity was used as a carrier gas. Ionization was done by electronic impact (EI), recording data in Full Scan mode. Quantification was performed by internal standard calibration, using the corresponding analytical standard for each determined metabolite and a deuterated internal standard depending on the family of metabolite. Internal standards used were succinic acid-d4, glycerol-13C3, norvaline, L-methionine-(carboxy-13C, methyl-d3), D-glucose-13C6, myristic-d27 acid and alpha-tocopherol-d6.

Publications used to generate the additional GWAS summary statistics for metabolites

Rhee EP, Ho JE, Chen M, Shen D, Larson MG, Ghorbani A, et al. A Genome-Wide Association Study of the Human Metabolome in a Community-Based Cohort. Cell Metab. 2014;18:130–43.

This study generated genome-wide summary statistics for metabolite levels from 2,076 participants of the FHS Offspring cohort. Sampling and genotyping details of this cohort were described in detail in the study and are briefly described in the cohort details section.

As described in this study, metabolomic profiling of these samples were done as follows. Briefly, amino acids, amino acid derivatives, urea cycle intermediates, nucleotides, and other positively charged polar metabolites were profiled using 10 ml of plasma. Lipids were profiled using 10 ml of plasma. For organic acids, sugars, bile acids, and other negatively charged polar metabolites, 30 ml of plasma were used and MS data were acquired using ESI and MRM in the negative ion mode.

Shin S, Fauman EB, Petersen A, Krumsiek J, Santos R, Huang J, et al. An atlas of genetic influences on human blood metabolites. Nat Genet. 2014;46.

This study generated genome-wide summary statistics for metabolite levels from 6,056 participants of the TwinsUK, and from 1,768 KORA F4 participants, of which 486 overlap between two cohorts. Sampling and genotyping details of these cohorts were described in detail in the study and are briefly described in the cohort details section.

For this study, non-targeted MS analysis was performed at Metabolon, Inc. Extracts were subjected to either GC-MS or UPLC-MS/MS. The UPLC-MS/MS platform utilized a Waters Acquity UPLC and a ThermoFisher LTQ mass spectrometer, which included an electrospray ionization source and a linear ion-trap mass analyzer operated at nominal mass resolution. The samples destined for analysis by GC-MS were analyzed on a Thermo-Finnigan Trace DSQ MS operated at unit mass resolving power with electron impact ionization and a 50–750 atomic mass unit scan range.

Draisma HHM, Pool R, Kobl M, Jansen R, Petersen A, Vaarhorst AAM, et al. Genome-wide association study identifies novel genetic variants contributing to variation in blood metabolite levels. Nat Commun. 2015;6:7208.

This study generated genome-wide summary statistics for metabolite levels from 8,660 participants of different cohort studies; NTR (1,372), QIMR (858), TwinsUK (1,235), LLS (657), EGCUT (1,124), KORA F4 (1,814), ERF (992) & KORA S4 (1,182). Sampling and genotyping details of these cohorts were described in detail in the study and are briefly are briefly described in the cohort details section.

Metabolomic profiling for this study was centralized. Targeted metabolomics measurements were performed using electrospray–flow injection analysis–tandem mass spectrometry methods and the Biocrates AbsoluteIDQ p150 kit (BIOCRATES Life Sciences AG). For all cohorts, metabolite measurements were carried out at the Metabolomics Platform of the Genome Analysis Center at the Helmholtz Zentrum München, Germany. In brief, the used metabolomic measurement technique is based on a targeted profiling scheme that is used to quantitatively screen for known small-molecule metabolites by multiple reaction monitoring, neutral loss and precursor-ion scans. Internal standards served as reference for the calculation of all metabolite concentrations, which are reported as micromolar.

Kettunen J, Demirkan A, Würtz P, Draisma HHM, Haller T, Rawal R, et al. Genome-wide study for circulating metabolites identifies 62 loci and reveals novel systemic effects of LPA. Nat Commun. 2016;7:11122.

This study generated genome-wide summary statistics for metabolite levels from 24,925 participants of different cohort studies; EGCUT (3,287), ERF (2,118), FTC (664), FR97 (3,661), COROGENE (828), GenMets (572), HBCS (708), KORA (1,745), LLS (2,227), NTR (1,192), NFBC 1966 (4,709), PredictCVD (374), PROTE (597), YFS (2,390). Sampling and genotyping details of these cohorts were described in detail in the study and are briefly described in the cohort details section.

In this study, the quantitative high-throughput NMR metabolomics platform was used to quantify 123 human blood metabolite measures that represent a broad molecular signature of systemic metabolism.

Long T, Hicks M, Yu H, Biggs WH, Kirkness EF, Menni C, et al. Whole-genome sequencing identifies common-to-rare variants associated with human blood metabolites. Nat Genet. 2017;49:568–78.

This study generated genome-wide summary statistics for metabolite levels from 1,960 participants of the TwinsUK. Sampling and genotyping details of this cohort were described in detail in the study and are briefly described in the cohort details section.

For this study, NMR spectroscopy was used for metabolomic profiling. This technique was employed to measure a wide range of metabolites in blood samples, including lipids, lipoproteins, and small molecules. The study linked the NMR-based metabolomic profiles with whole-genome sequencing (WGS) data.

Gallois A, Mefford J, Ko A, Vaysse A, Julienne H, Ala-korpela M, et al. A comprehensive study of metabolite genetics reveals strong pleiotropy and heterogeneity across time and context. Nat Commun. 2019;10:4788.

This study generated genome-wide summary statistics for metabolite levels from 6,253 participants of the METSIM. Sampling and genotyping details of this cohort were described in detail in the study and are briefly described in the cohort details section.

For this study, 228 serum metabolites (lipids, lipoproteins, amino acids, fatty acids, and other low molecular weight metabolites) measurements were made with NMR at baseline.

Cohorts used to generate the additional GWAS summary statistics for metabolites

The FHS Offspring cohort (https://www.framinghamheartstudy.org/) is a prospective, observational, community-based cohort. A total of 2,076 participants of European descent who attended the fifth examination (1991–1995) and underwent metabolic profiling and genome-wide genotyping were included in this analysis. All participants provided informed consent and the study protocol was approved by the Boston University Medical Center IRB. Clinical Assessment Participants underwent a comprehensive medical history, physical examination, and anthropometry at the fifth examination. Blood samples were collected after an overnight fast, immediately centrifuged, and stored at -80^{o}C until assayed.

Genome-wide genotyping was conducted using the Affymetrix 500K mapping array and the Affymetrix 50K gene-focused MIP array. Imputation of 2.5 million SNPs was then performed (HapMap CEU population, release 22, build 36; http://hapmap.org) using a hidden Markov model that was implemented in MACH (version 1.0.15).

The TwinsUK cohort (https://twinsuk.ac.uk/) is the largest adult twin registry in the UK, established in 1992 to support research on the genetic and environmental factors that influence common age-related diseases. Based at King's College London, the registry includes over 15,000 twins aged 16 to 98, predominantly female. Blood samples were collected from the TwinsUK cohort after at least 6 hours of fasting predominantly overnight.

Genotyping of the TwinsUK dataset was done with a combination of Illumina arrays (HumanHap300, HumanHap610Q, 1M-Duo and 1.2MDuo). Imputation was performed using the IMPUTE software package (v2) using a stepwise procedure. First, the sparser HumanHap300 dataset was imputed to the HumanHap610Q content using phased TwinsUK HumanHap610Q haplotypes as reference. Subsequently, the combined panel was imputed using reference haplotypes from the HapMap2 project (rel 22, combined CEU+YRI+ASN panels).

The KORA (Cooperative Health Research in the Region of Augsburg) cohort (https://www.helmholtz-munich.de/index.php) is a large-scale, long-term research project conducted in southern Germany. It

focuses on understanding the epidemiology and risk factors for chronic diseases, including cardiovascular diseases, diabetes, and other metabolic and respiratory conditions. The KORA S4 is one of the main baseline surveys in the KORA study, conducted between 1999 and 2001. It involved over 4,000 participants from the Augsburg region, aged 25 to 74 years, and gathered extensive data on cardiovascular and metabolic health. Participants underwent clinical assessments and provided biological samples for future analysis. The KORA F4 cohort specifically is a follow-up study of the KORA S4 baseline survey to study the progression of diseases over time. Blood samples in KORA F4 were collected between 2006 and 2008, as part of the follow-up examination, when participants aged 32 to 81 years.

Blood was drawn in the morning after a period of overnight fasting. Genotyping of the KORA population was carried out using the Affymetrix GeneChip array 6.0. Imputation was done using IMPUTE v0.4.2 12 based on HapMap2.

The NTR (Netherlands Twin Register) cohort (http://www.tweelingenregister.org/) is a large, longitudinal research project that gathers genetic and environmental data from twins and their families in the Netherlands. Established in 1987 by the Vrije Universiteit (VU) Amsterdam, its main purpose is to study how genetic and environmental factors influence traits and diseases such as mental health, behavior, personality, and physical health.Blood samples were collected in a standardized manner after overnight fasting.

Genotyping was performed on the Affymetrix 6.0, Affymetrix Perlegen 5.0, Illumina 370, Illumina 660 and Illumina Omni Express 1M platforms. Genotypes were imputed against the reference set using IMPUTE v2.

The EGCUT (Estonian Genome Center, University of Tartu) cohort is a large biobank and longitudinal population study focused on the health and genetics of the Estonian population. It was established to investigate how genetic and environmental factors contribute to common diseases, complex traits, and overall human health. This cohort consists of over 200,000 gene donors, representing approximately 20% of Estonia's adult population. The PROTE sub-cohort specifically focuses on genomic data related to metabolic traits and diseases. Blood samples were collected as follows: venous blood was drawn, DNA was isolated by precipitation, and samples were stored in liquid nitrogen.

The ERF (Erasmus Rucphen Family) cohort (http://www.gefos.org/?q=content/erasmus-rucphen-family-study-2) is a family-based cohort study that is embedded in the Genetic Research in Isolated Populations (GRIP) program in the South West of the Netherlands. The aim of this program was to identify genetic risk factors in the development of complex disorders. For the ERF study, 22 families that had at least five children baptized in the community church between 1850-1900 were identified with the help of genealogical records. All living descendants of these couples and their spouses were invited to take part in the study. Data collection started in June 2002 and was finished in February 2005 (n=2065).

The LLS (Leiden Longevity Study) (https://www.lumc.nl/en/afdelingen/biomedical-data-sciences/molecular-epidemiology/ageing/), is a study in which 421 long-lived families participate. The participants in this study are 944 90-year-old brother-sister pairs, 1,671 of their offspring and 744 offspring partners. The partners form the control group and represent the general population. Blood samples were taken at baseline for extraction of DNA and the determination of non-fasted serum parameters.

The QIMR (Queensland Institute of Medical Research) adolescent study comprised twins and their non-twin siblings living in south-east Queensland, Australia. Most (98% by self-report) are of mixed European ancestry, mainly from the British Isles. The participants were not selected on the basis of any disease or other outcome. Blood samples were collected at the end of testing sessions. The time since the last meal was recorded. DNA samples from each individual were genotyped on the Illumina 610-Quad Beadchip by the Scientific Services Division at deCODE Genetics Iceland.

The FCT (Finnish twin cohort) is a population-based cohort longitudinal study of five consecutive birth cohorts of Finnish twins born between 1983 and 1987 and five consecutive birth cohorts of Finnish twins born between 1975 and 1979. The baseline data collection started in 1991 and was completed in 1996, with a response rate ~ 88%. Subsequent follow-up assessments were made following a similar approach. For both the baseline and follow-up assessments included surveys of health habits and attitudes, symptom checklists, personality scales, and social relationships. In addition, blood samples were taken from all twins during a visit to the twin research clinic in Helsinki (Finland) at the last follow-up (young adulthood) for DNA and biochemistry analyses.

The FINRISK is a long-term, population-based health survey conducted in Finland since 1972. Its primary goal is to study risk factors for chronic diseases, particularly cardiovascular diseases. The survey collects data on lifestyle, diet, blood pressure, cholesterol levels, and other health-related variables from random samples of Finnish adults every five years. For each cross-sectional survey, a representative random sample was selected from 25- to 74-year-old inhabitants of different regions in Finland. Along the different above studies, there were included included eligible individuals from FINRISK surveys conducted in 1992, 1997 (FR97), 2002, and 2007. Biological samples are also collected for further analysis. FINRISK samples were genotyped in three separate batches: DILGOM, PredictCVD and FINRISK.

The HBCS (Helsinki Birth Cohort Study) is a longitudinal study that investigates how early-life factors, including prenatal and early postnatal conditions, influence health outcomes later in life. Established in 2001, the study includes over 13,000 individuals born in Helsinki between 1934 and 1944. Researchers use medical records, questionnaires, and physical assessments to examine links between early development and risks for chronic diseases such as cardiovascular disease, diabetes, and mental health conditions. The HBCS aims to understand the long-term effects of early-life factors on aging and disease.

The GenMets (Genetics of Metabolic Syndrome) study is a population-based cohort primarily focused on understanding the genetic basis of metabolic disorders. This cohort includes individuals who have been extensively phenotyped for various metabolic traits, such as insulin resistance, obesity, lipid profiles, and blood pressure. Persons with known diabetes were excluded. Participants provided blood samples in fasting conditions.

The NFBC1966 (Northern Finland Birth Cohort 1966) is a large, longitudinal birth cohort study that tracks individuals born in the provinces of Oulu and Lapland in northern Finland. The cohort was established to investigate how early-life factors influence long-term health outcomes, with a particular focus on genetic, environmental, and social influences on health throughout life. Mothers living in the two northern-most provinces of Finland were invited to participate if they had expected delivery dates during 1966. Individuals still living in the Helsinki area or Northern Finland (N = 5923) were asked at age 31 to participate in a detailed biological and medical examination as well as a questionnaire.

The Cardiovascular Risk in YFS (Young Finns Study) is a population based prospective cohort study. It was conducted at five medical schools in Finland (Turku, Helsinki, Kuopio, Tampere and Oulu), with the aim of studying the levels of cardiovascular risk factors in children and adolescents in different parts of the country. The latest follow-up was conducted in 2007. The serum samples for this metabolomics study were collected at this latest follow up.

The COROGENE study (Genetic Predisposition of Coronary Heart Disease in Patients Verified with Coronary Angiogram) focuses on understanding the genetic underpinnings of coronary heart disease. Recruitment for the study occurred between 2006 and 2008 at several hospitals in Finland, primarily including Helsinki University Central Hospital. It specifically involves patients who have undergone coronary angiograms, a diagnostic procedure used to visualize the heart's blood vessels and assess for blockages.

The METSIM cohort (Metabolic Syndrome in Men) is a large, population-based study conducted in Finland. The cohort consists of over 10,000 Finnish men, aged 45–73 years, who were recruited from the Kuopio region in Eastern Finland between 2005 and 2010. The study aimed at investigating the genetic and environmental factors contributing to metabolic syndrome, type 2 diabetes, cardiovascular disease, and related metabolic traits. The study focuses on men and provides insights into the genetic determinants of various metabolic disorders. All samples were genotyped for 665,478 SNPs using the Illumina OmniExpress chip.

Cohorts from GWAS summary statistics for cardiovascular risk

The CARDIoGRAMplusC4D 1000 Genomes-based GWAS is a meta-analysis of GWAS studies of mainly European, South Asian, and East Asian, descent imputed using the 1000 Genomes phase 1 v3 training set with 38 million variants. The study interrogated 9.4 million variants and involved 60,801 CAD cases and 123,504 controls. Data as published in: CARDIoGRAMplusC4D Consortium, M Nikpey, A Goel, H Won,

LM Hall C. Willenborg, S Kanoni, D Saleheen et al. A comprehensive 1000 Genomes–based genome-wide association meta-analysis of coronary artery disease. Nat Genet 2015 47:1121-1130.

The UK Biobank is a large-scale biomedical database and research resource in the United Kingdom, established to improve the prevention, diagnosis, and treatment of a wide range of serious and life-threatening diseases. It contains extensive genetic, lifestyle, and health information from approximately 500,000 volunteer participants, aged 40-69, who were recruited between 2006 and 2010. The data uses was published under the following reports: P van der Harst, N Verweij. Identification of 64 Novel Genetic Loci Provides an Expanded View on the Genetic Architecture of Coronary Artery Disease. Circ Res 2018 122(3):433-443; L Jiang, Z Zheng, H Fang, J Yang. A generalized linear mixed model association tool for biobank-scale data. Nat Genet 2021 53(11):1616-1621.

The FinnGen is a large-scale public-private research project in Finland aimed at understanding the genetic basis of diseases by combining genetic data with extensive health records from Finnish biobanks and national health registries. The study is one of the largest genetic research projects in the world, and it seeks to generate new insights into disease mechanisms, advance personalized medicine, and improve public health. The project aims to analyze genetic data from 500,000 Finnish individuals, representing approximately 10% of Finland's population. The cohort is recruited from Finnish biobanks and linked to comprehensive health information. Data as published in: MI Kurki, J Karjalainen, P Palta, et al. FinnGen provides genetic insights from a well-phenotyped isolated population. Nature 613, 508–518 (2023).

List of plots for each metabolite

Distribution plot of metabolite levels among the 4,974 GCAT individuals

QQ plot of GWAS meta-analyses

QQ plot of TWAS overall multi-tissue results

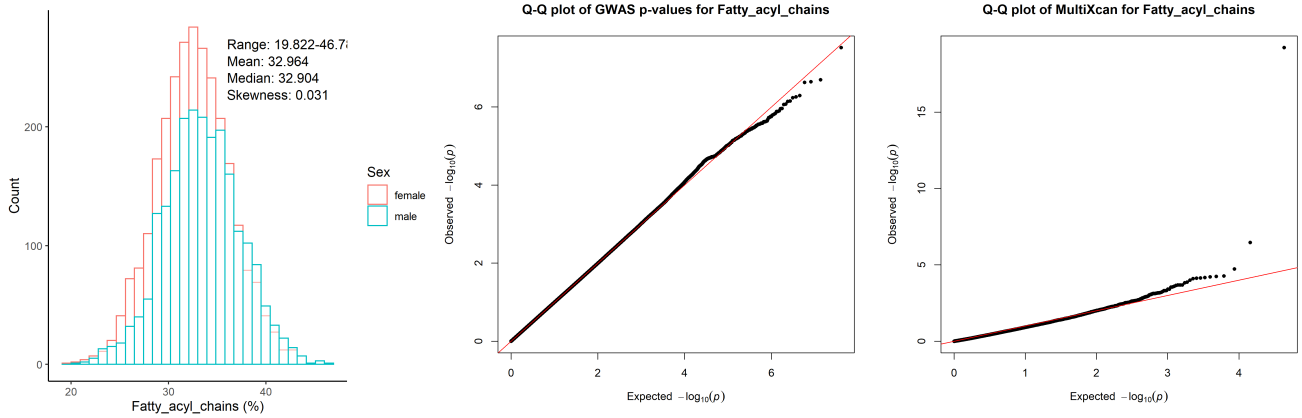
Manhattan plot of GWAS meta-analyses

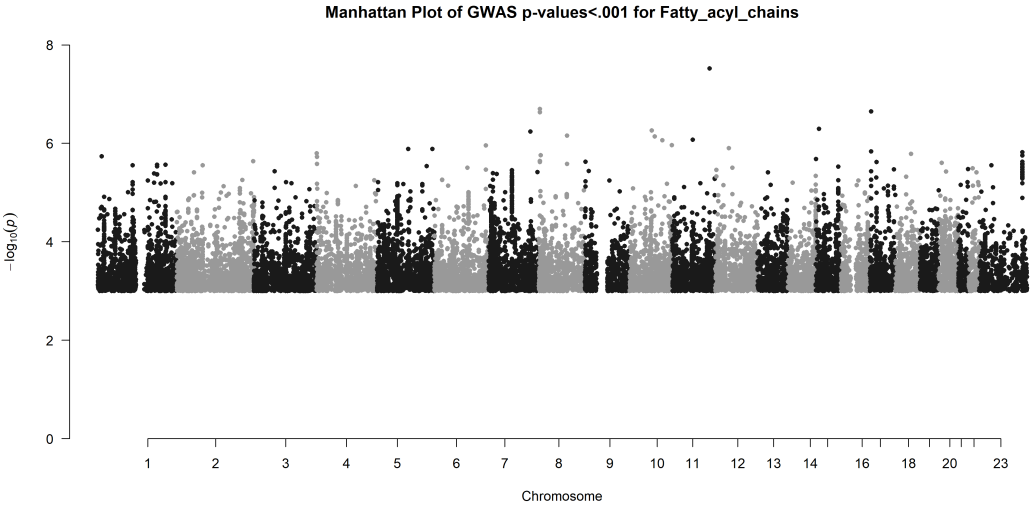
Manhattan plot of TWAS overall multi-tissue results with significant fine-mapped genes (in green)

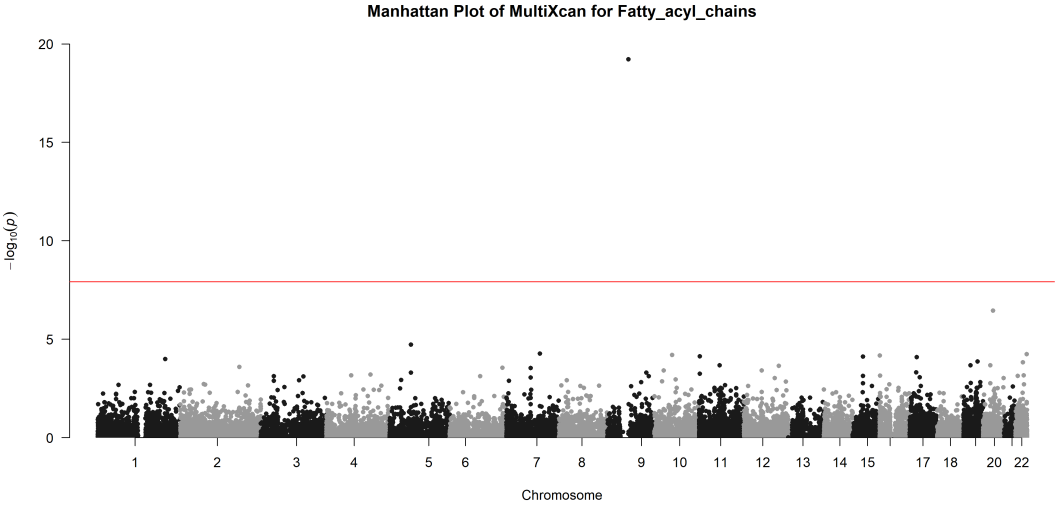
LocusZoom plots for the top locus-metabolite associations (350 in total)

Fatty acyls

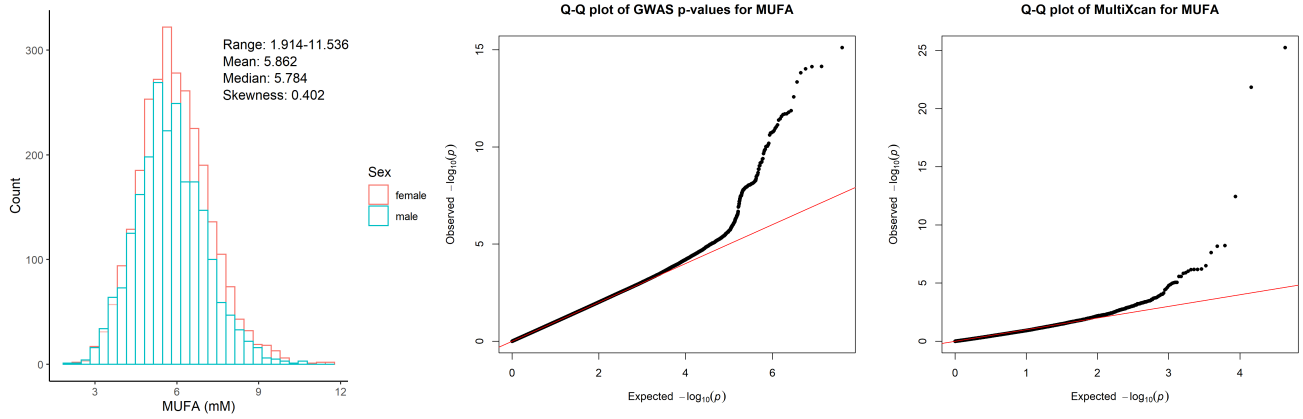
Fatty acyl chains (%)

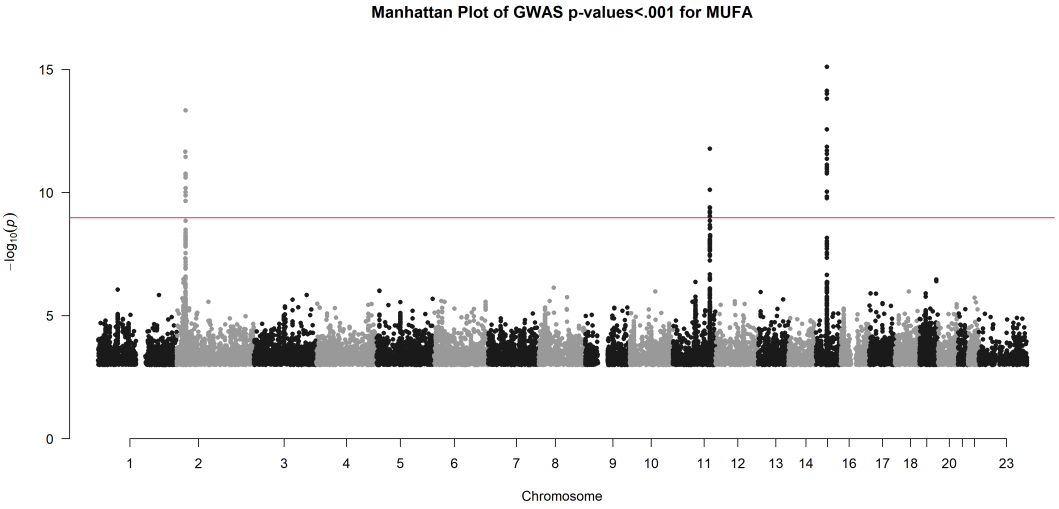


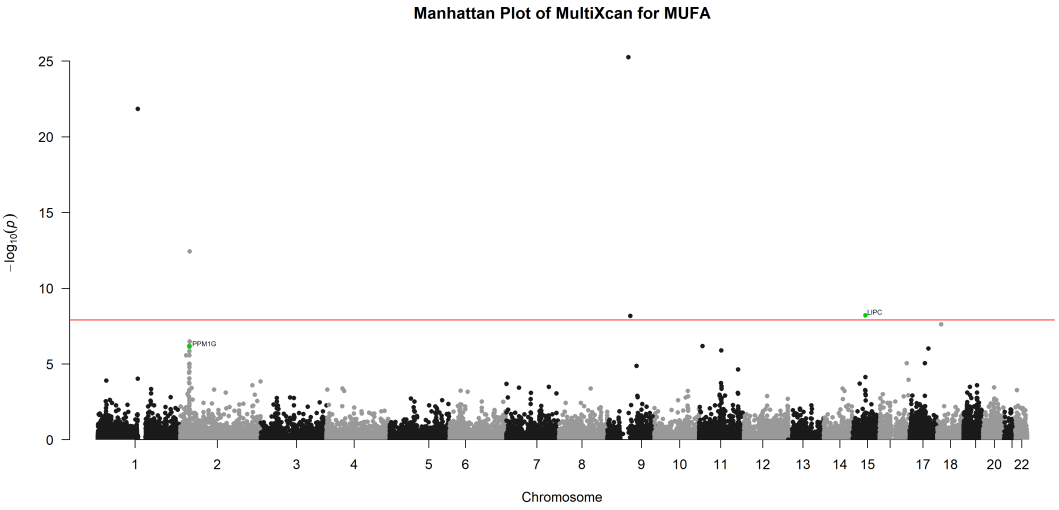




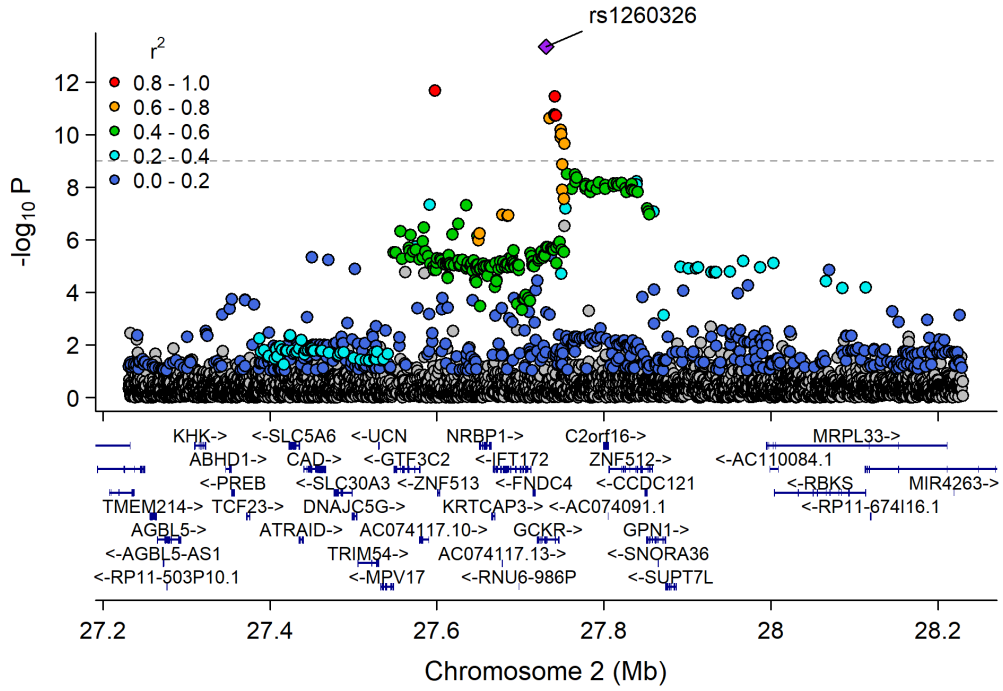
Monounsaturated fatty acids (mM)



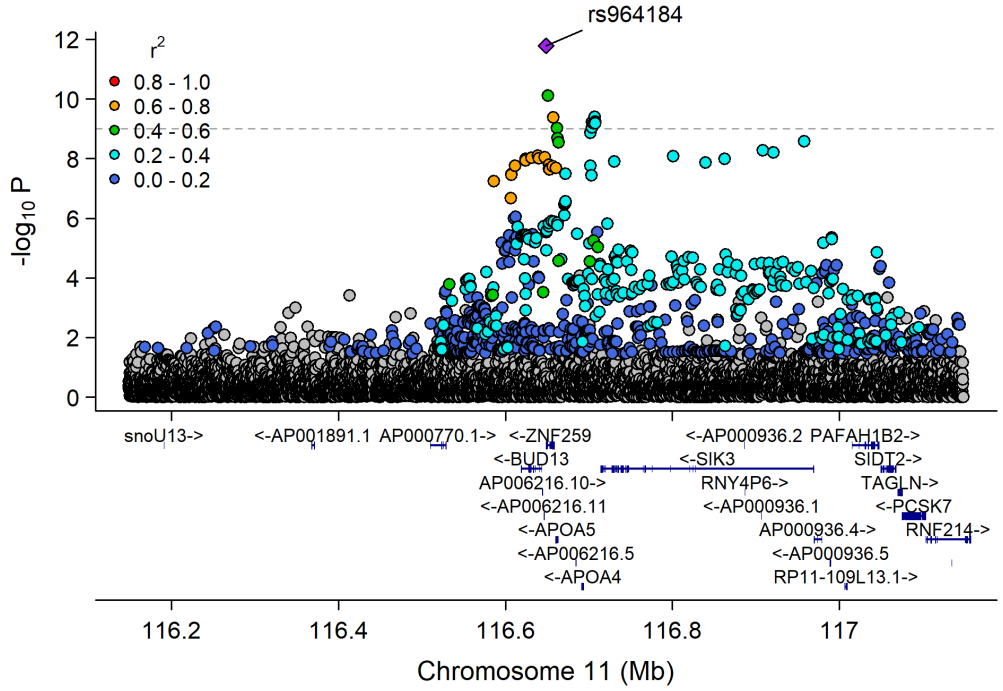




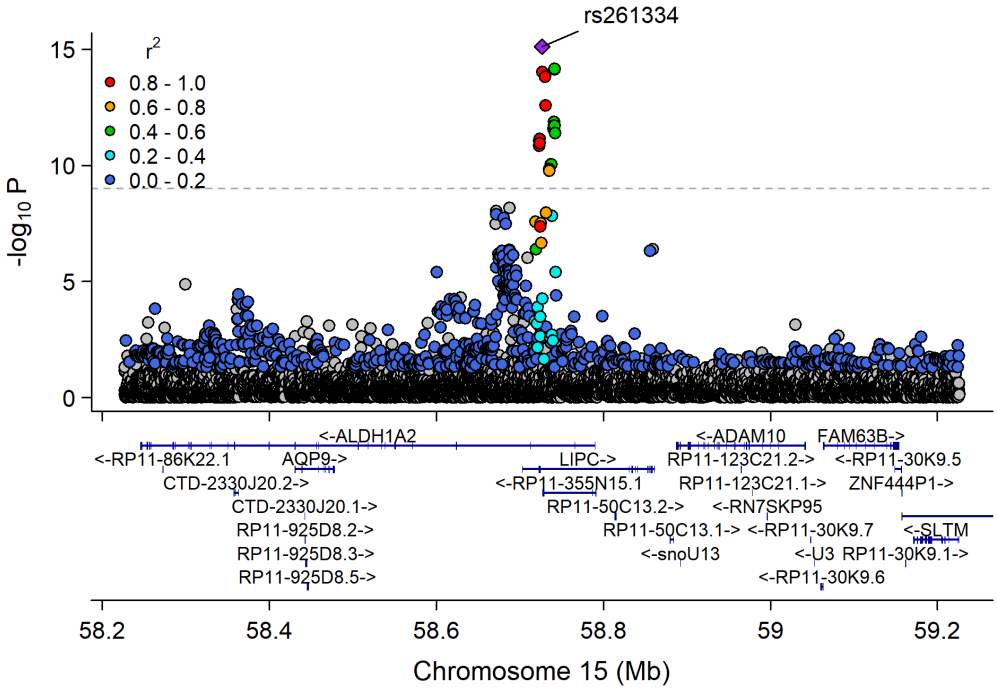
Chr2_26753815_28597624



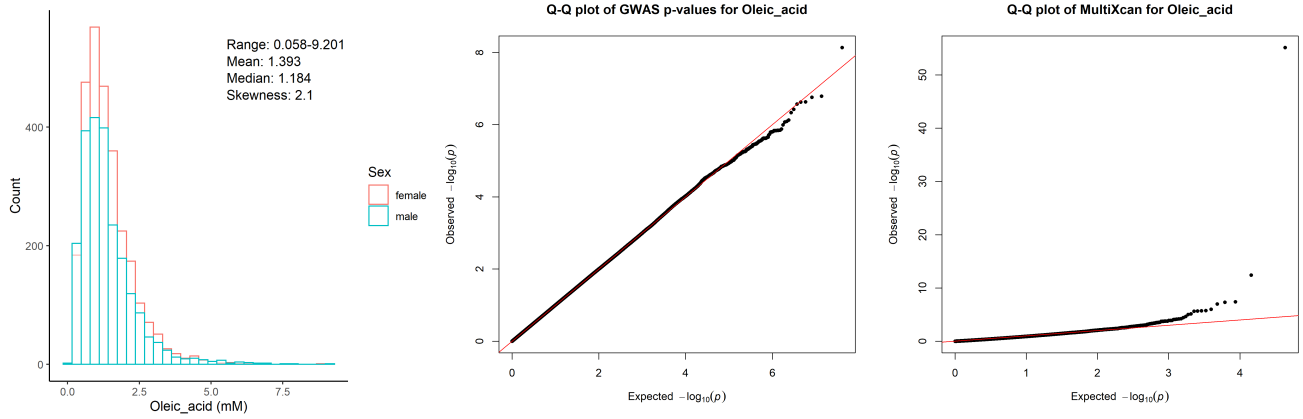
Chr11_115541901_117633315

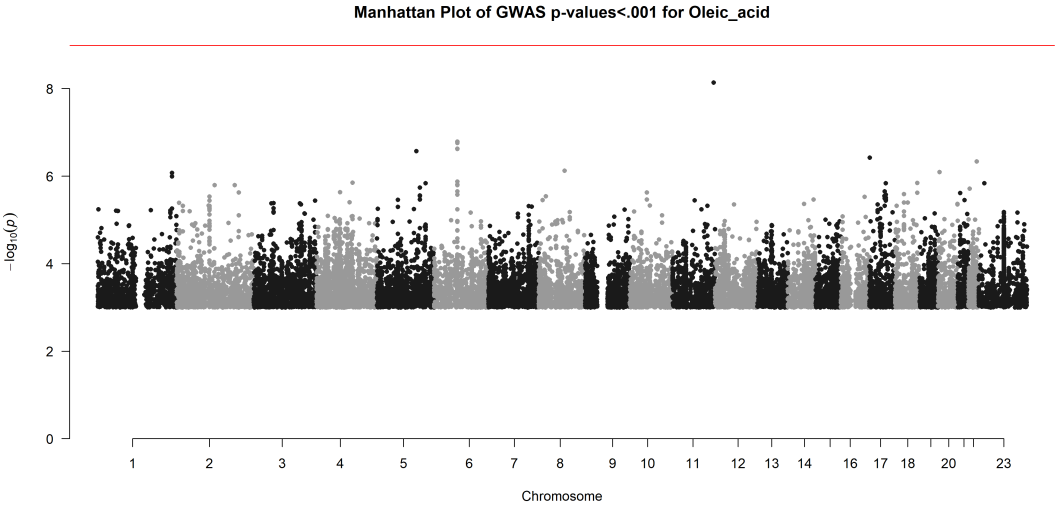


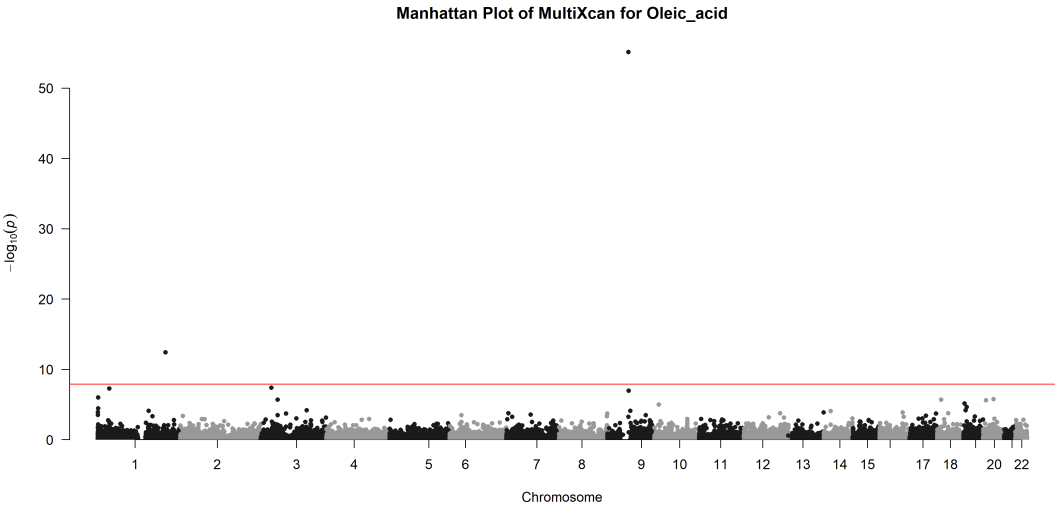
Chr15_57658798_60490883



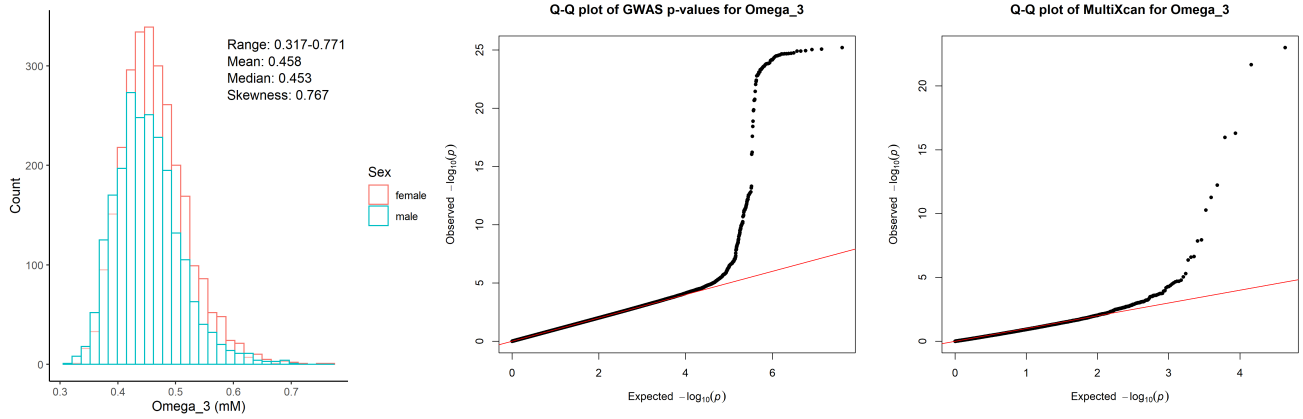
Oleic acid (mM)

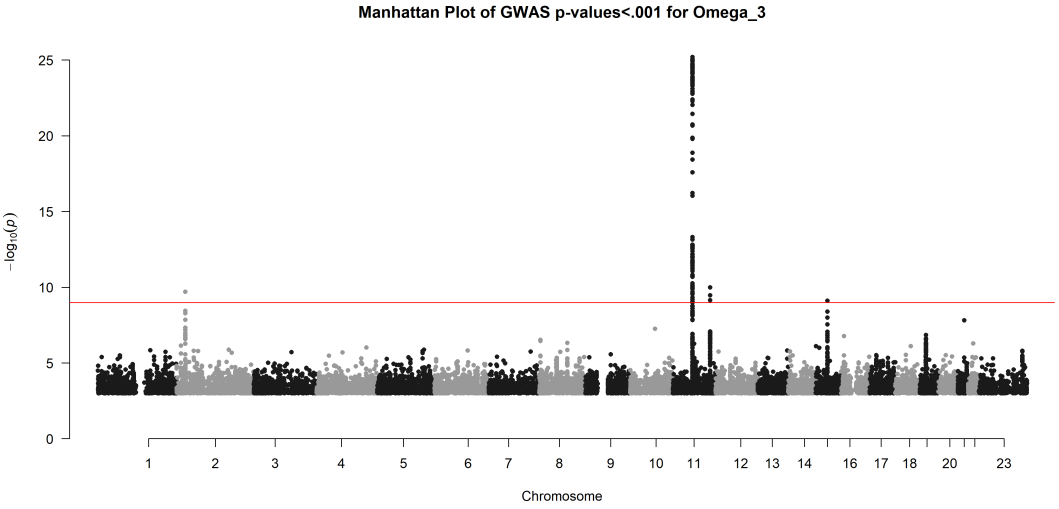


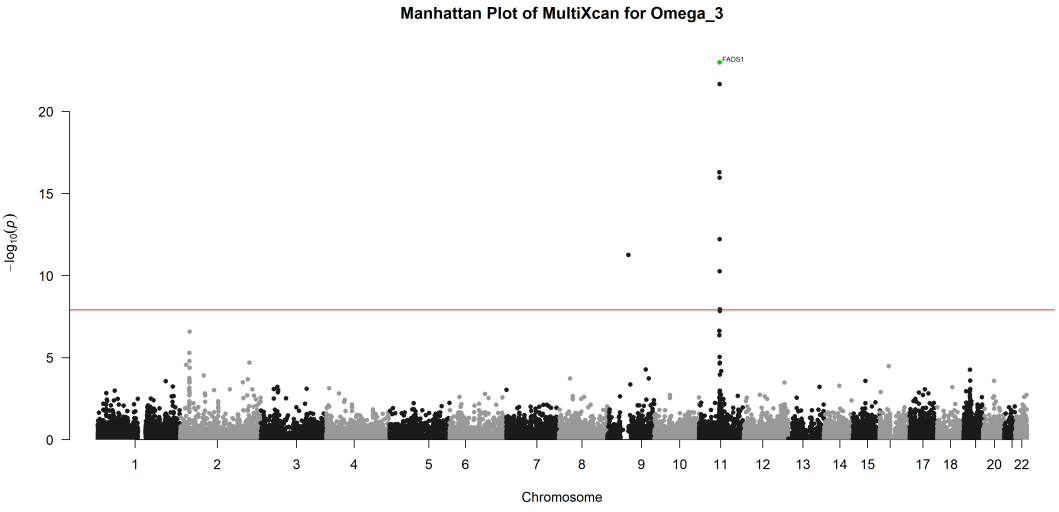




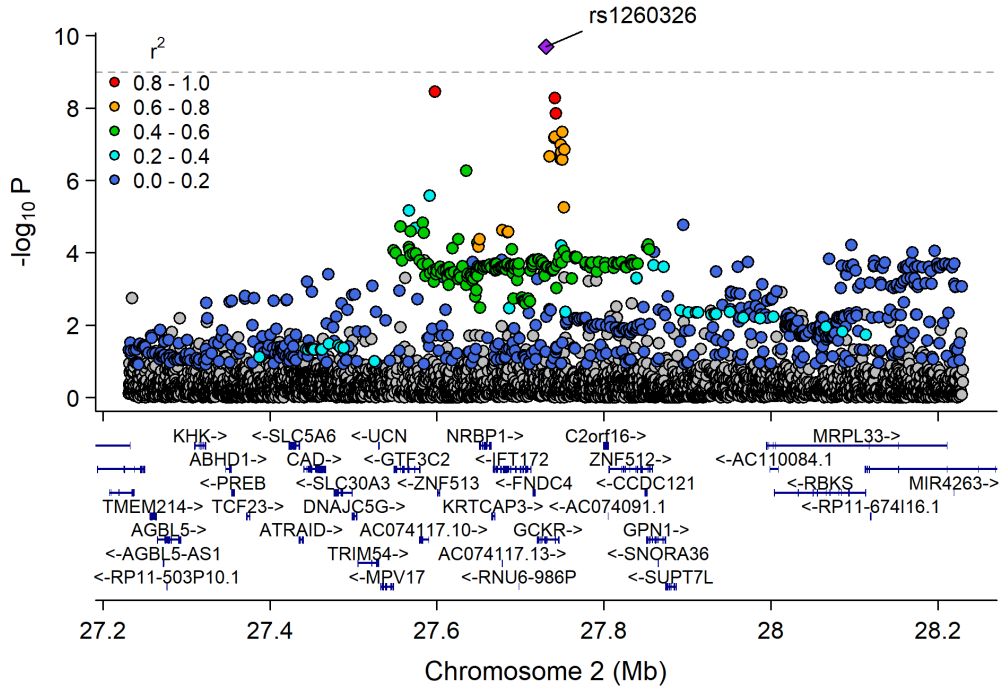
Omega-3 (mM)


Range: 0.317-0.771
Mean: 0.458
Median: 0.453
Skewness: 0.767
Count
Omega_3 (mM)
Sex
female
male
Q-Q plot of GWAS p-values for Omega_3
Observed $-\log_{10}(p)$
Expected $-\log_{10}(p)$
Q-Q plot of MultiXcan for Omega_3
Observed $-\log_{10}(p)$
Expected $-\log_{10}(p)$

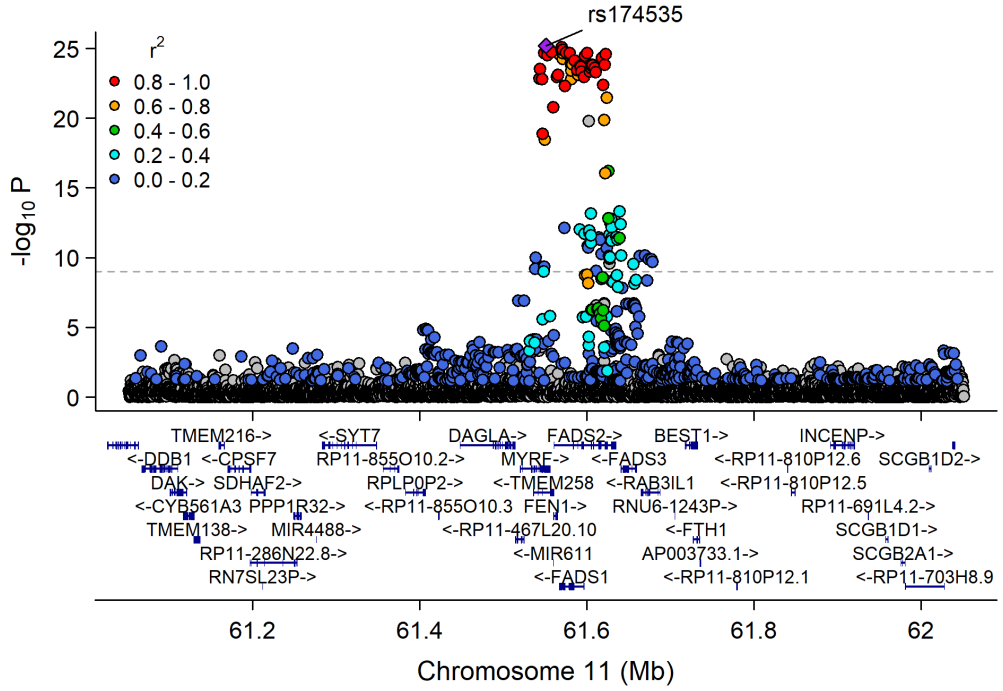

Manhattan Plot of GWAS p-values<.001 for Omega_3
$-\log_{10}(p)$
Chromosome


Manhattan Plot of MultiXcan for Omega_3
FADS1
$-\log_{10}(p)$
Chromosome

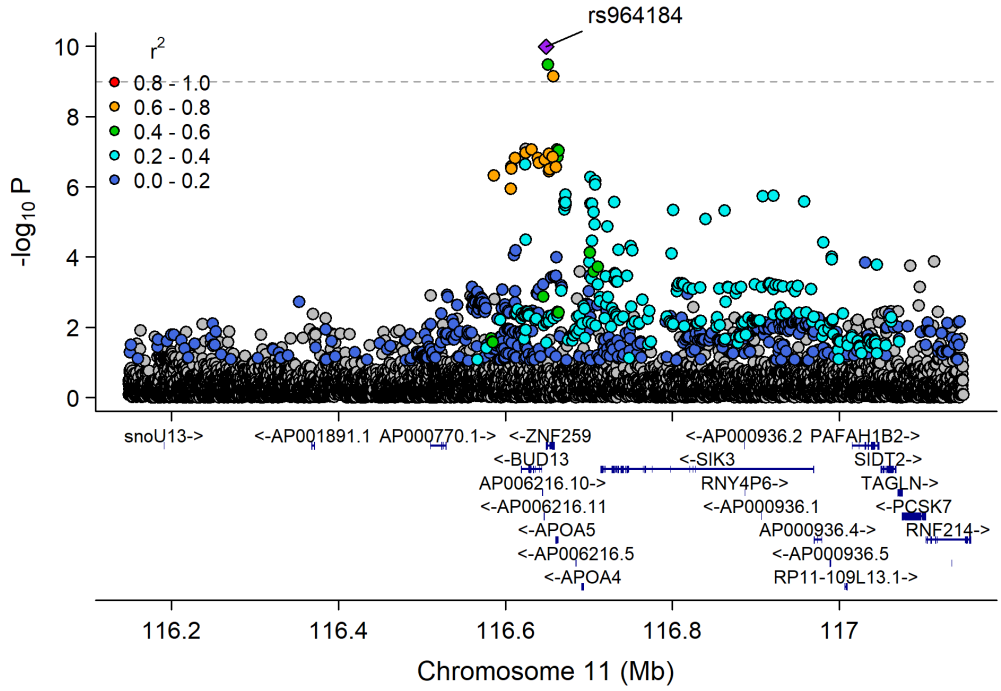
Chr2_26753815_28597624



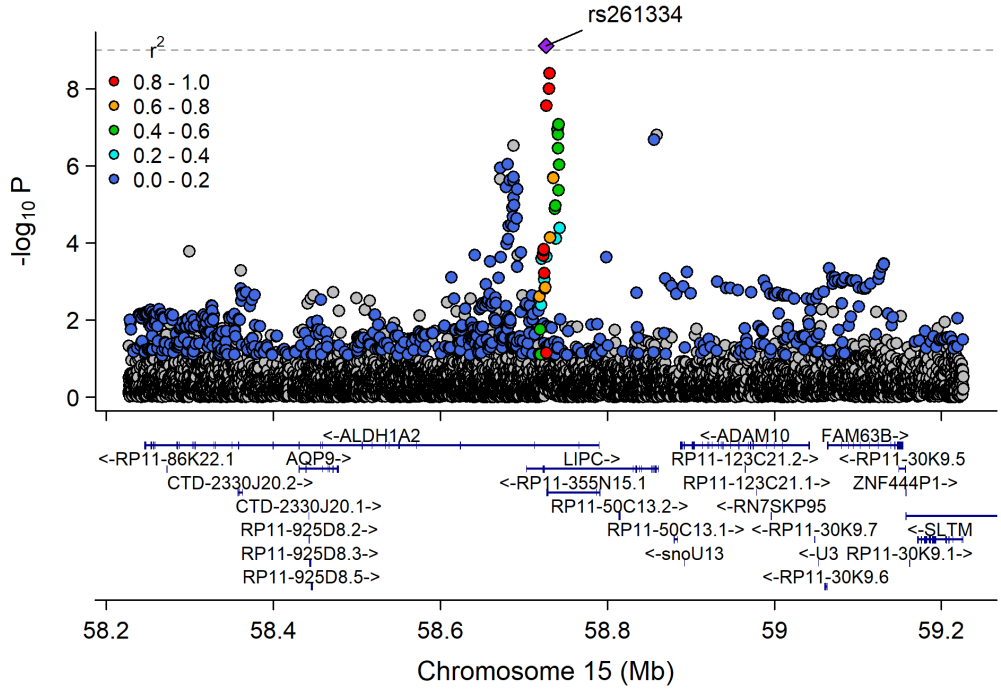
Chr11_59978355_62914375



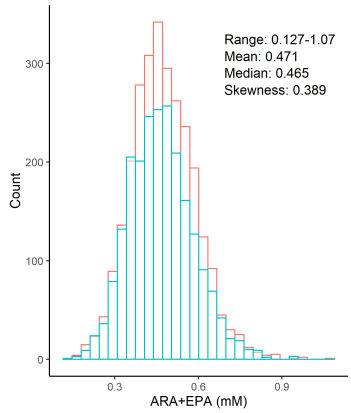
Chr11_115541901_117633315

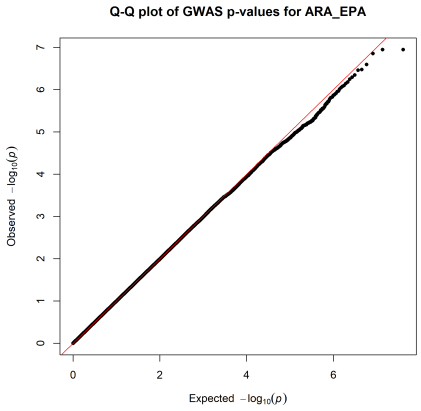


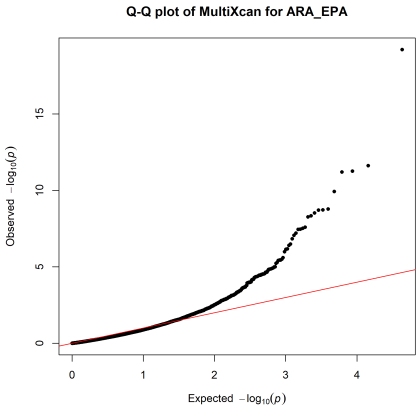
Chr15_57658798_60490883

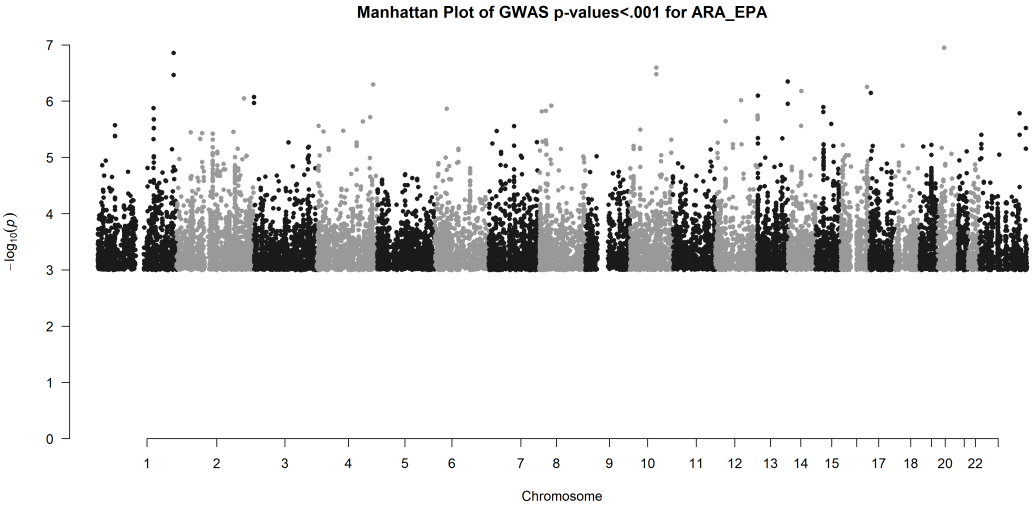


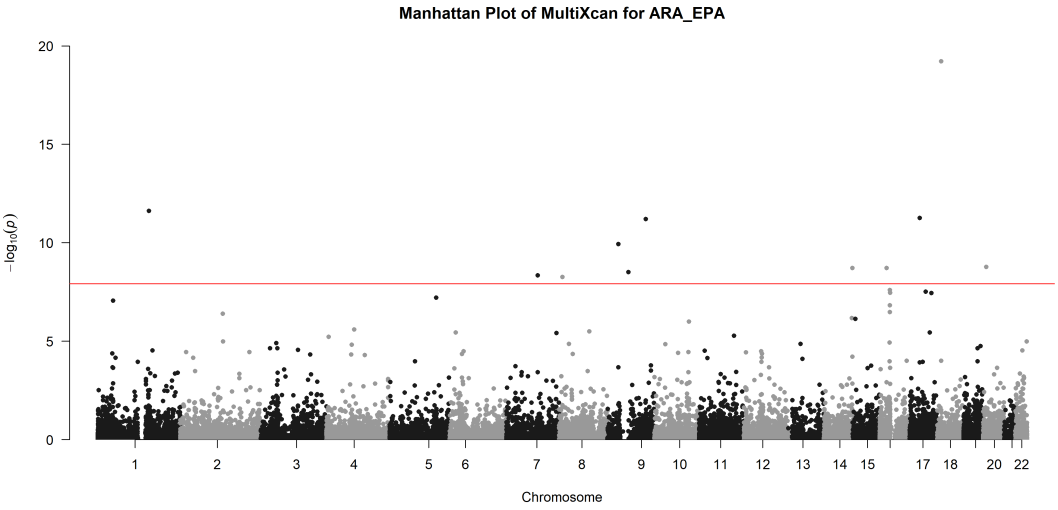
Arachidonic acid + Eicosapentaenoic acid (mM)



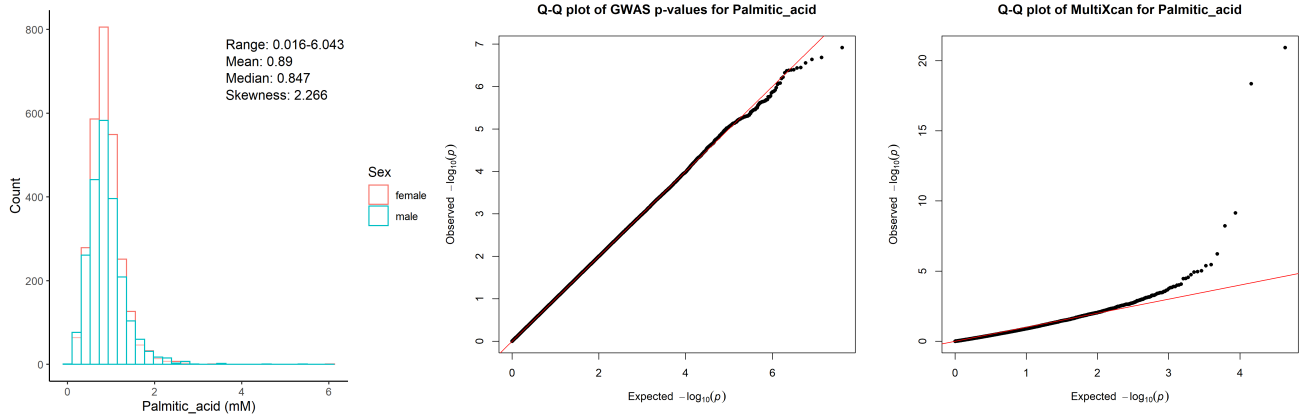


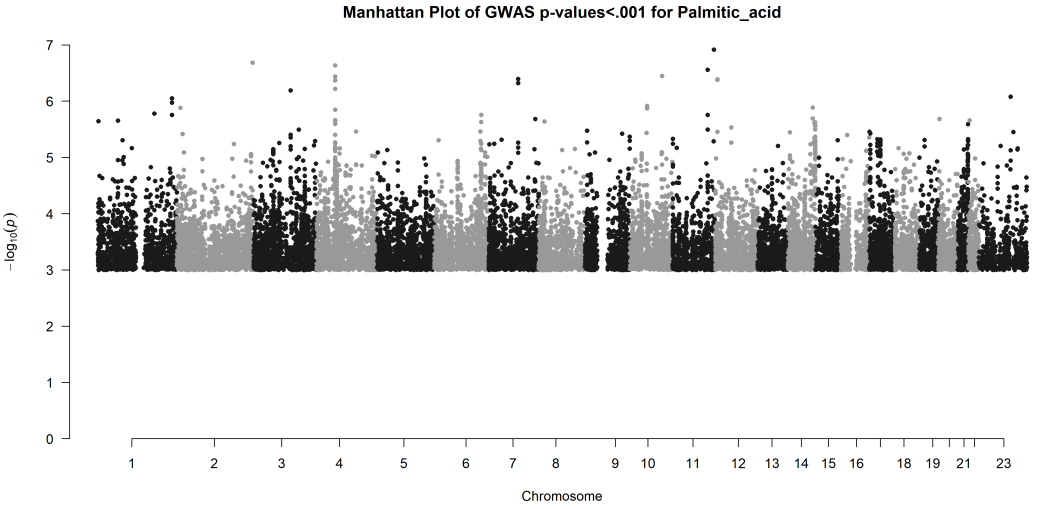


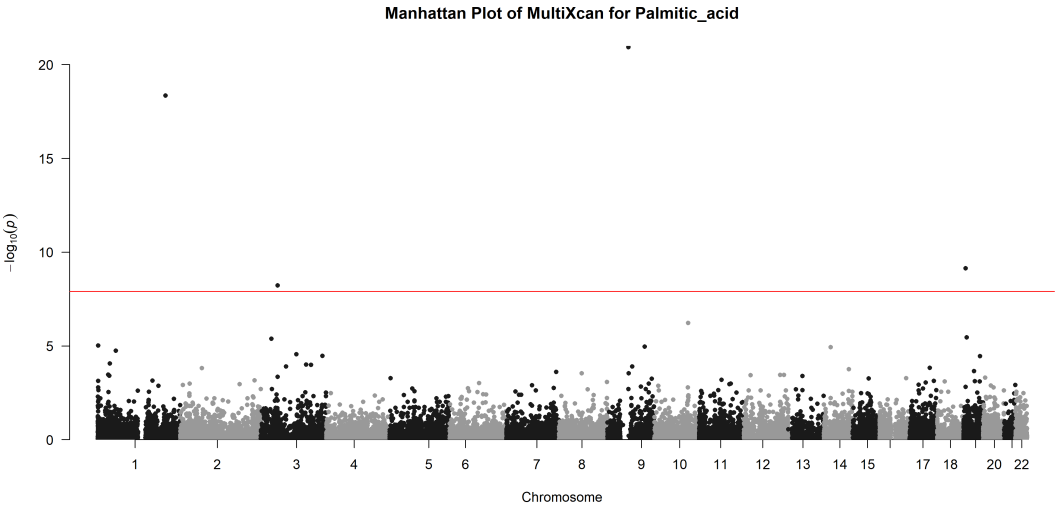




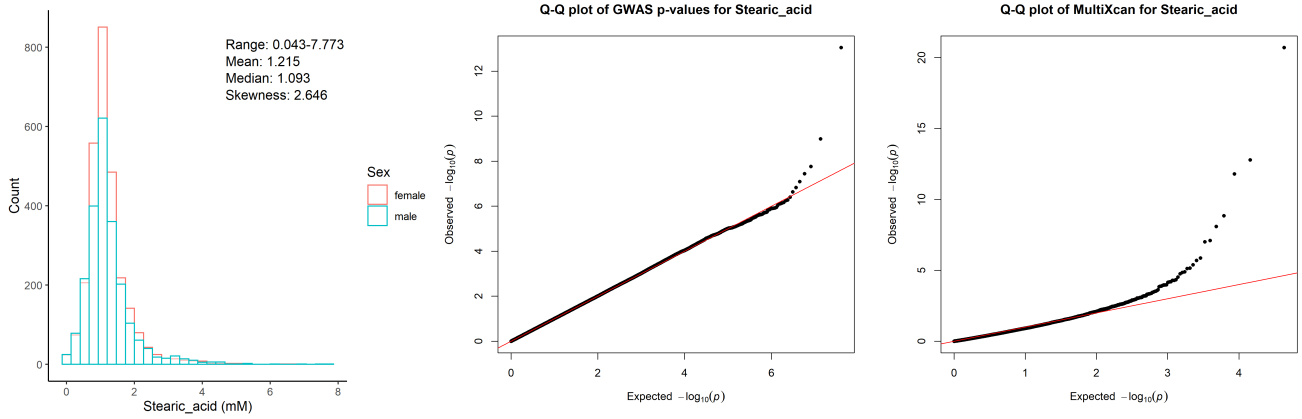
Palmitic acid (mM)

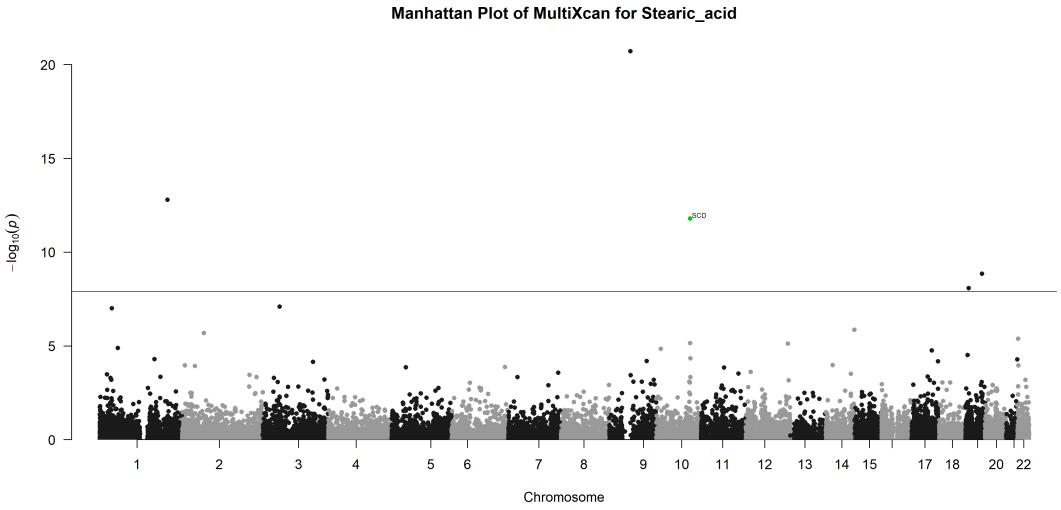




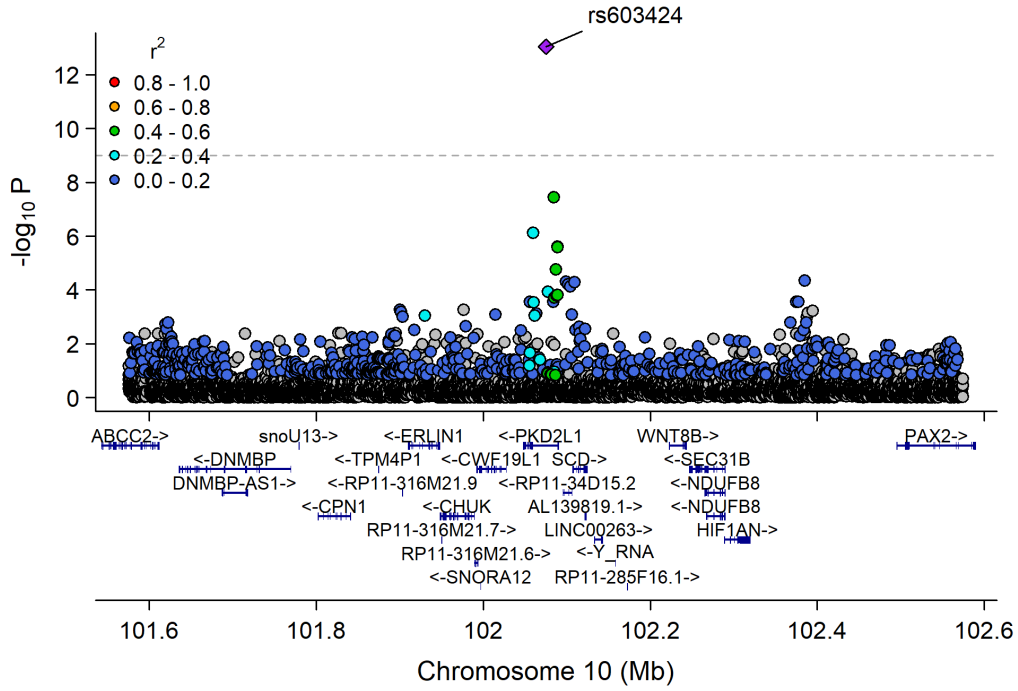


Stearic acid (mM)

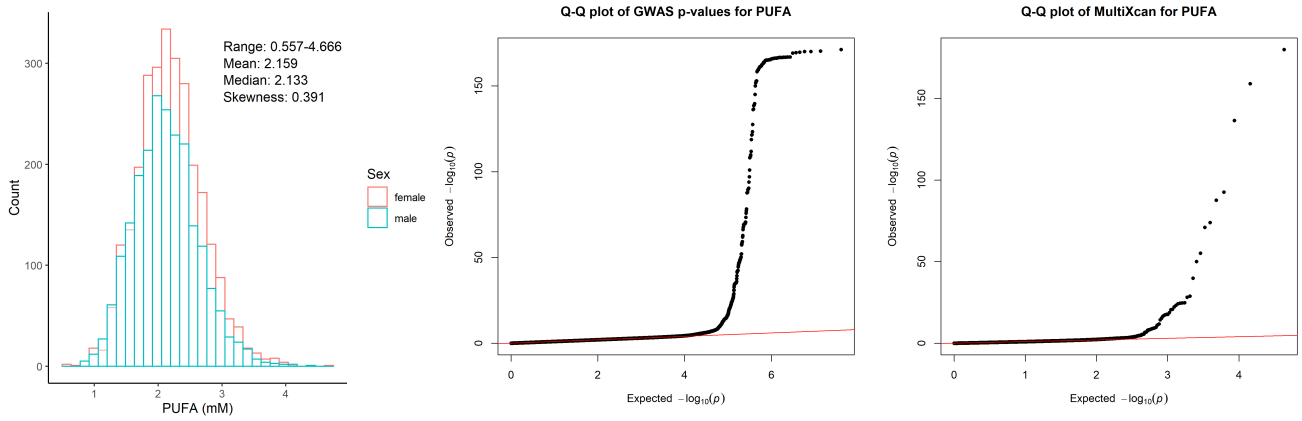


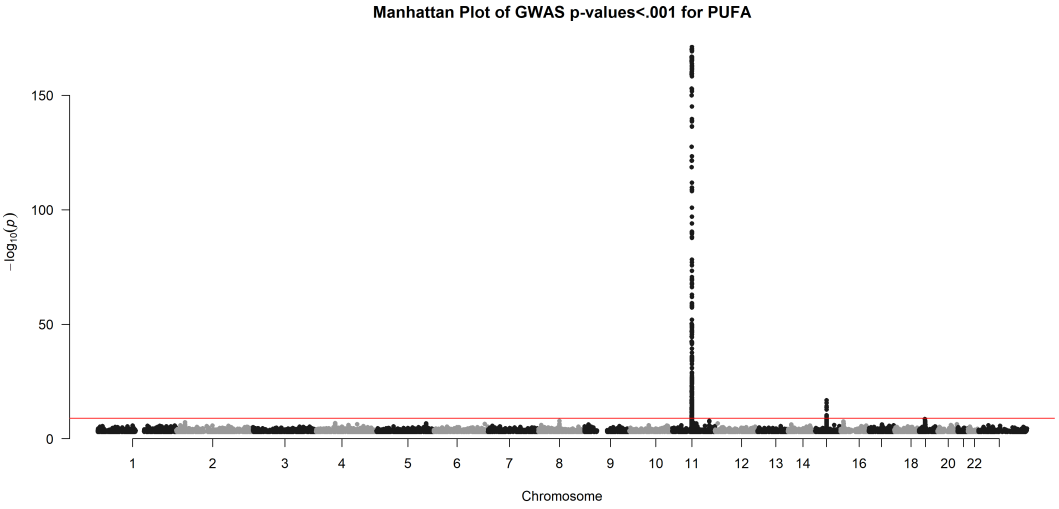


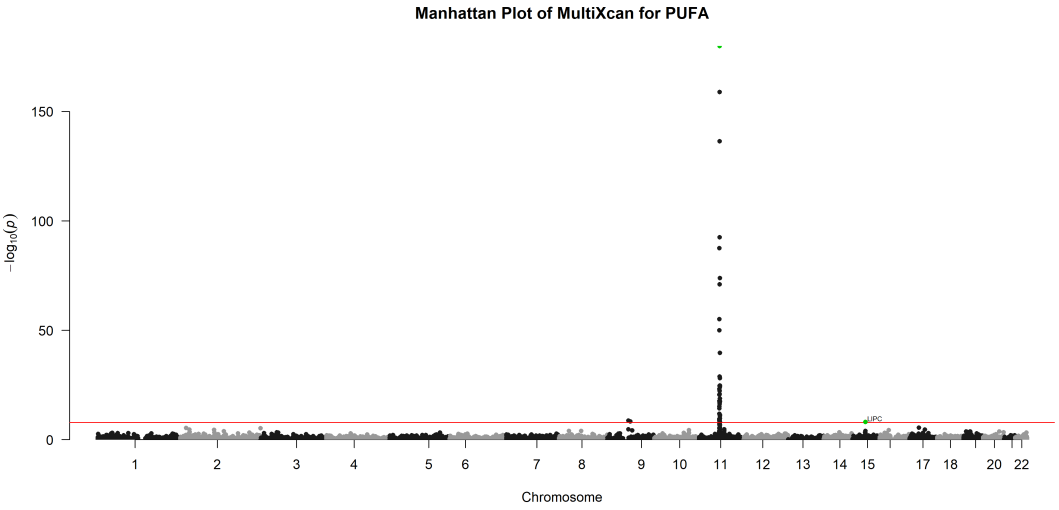
Chr10_101403323_102214579



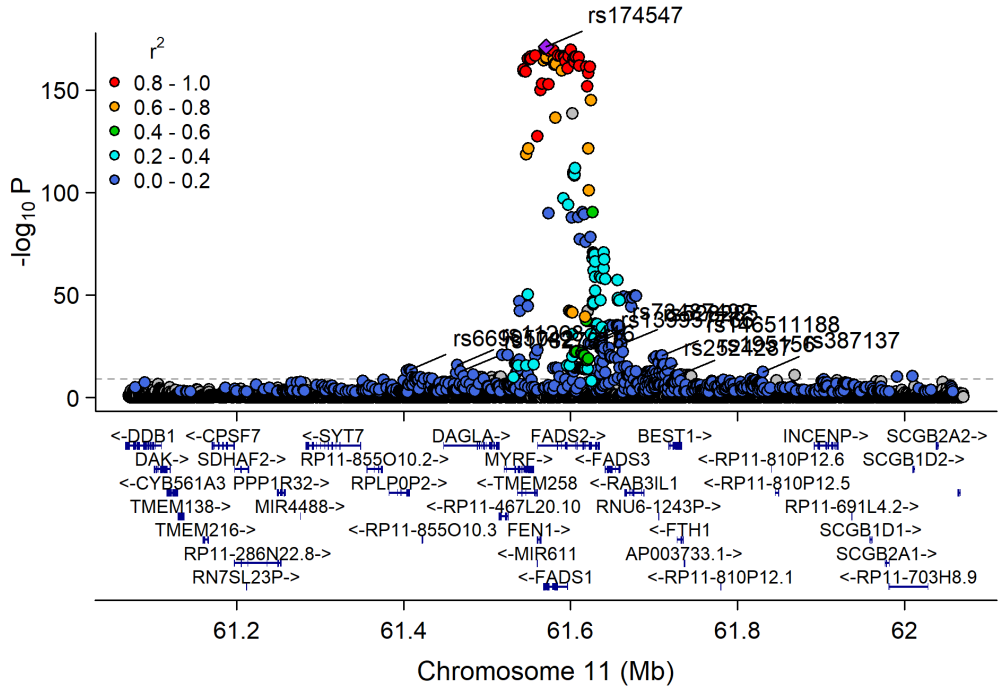
Polyunsaturated fatty acids (mM)



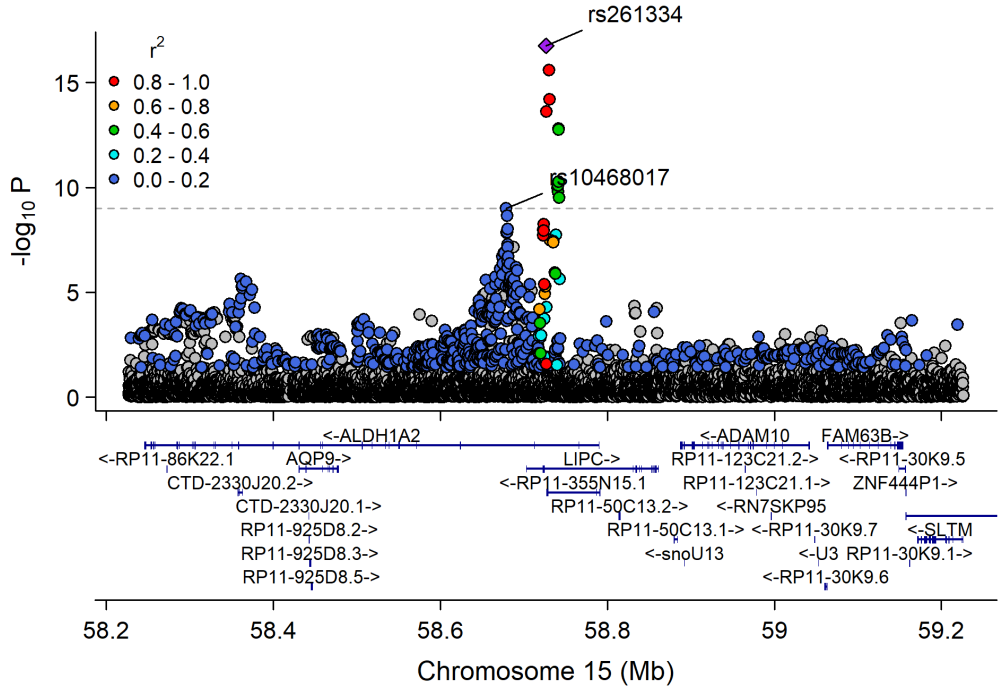




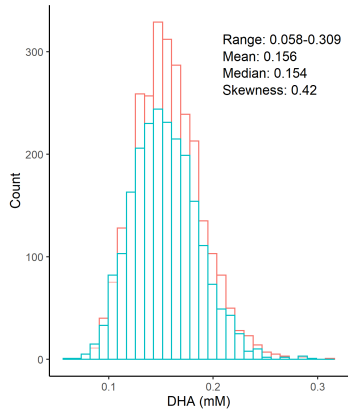
Chr11_59978355_62914375

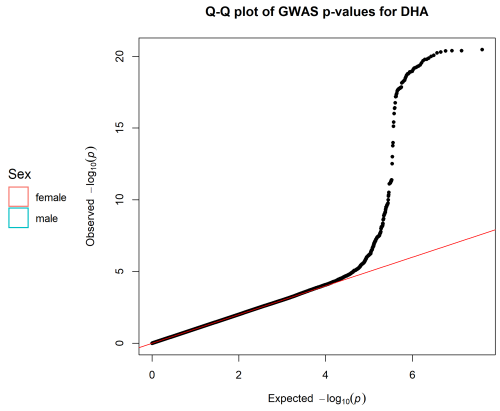


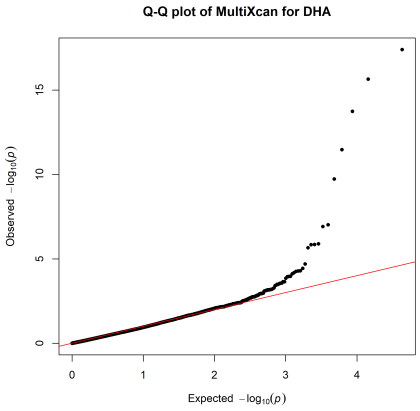
Chr15_57658798_60490883

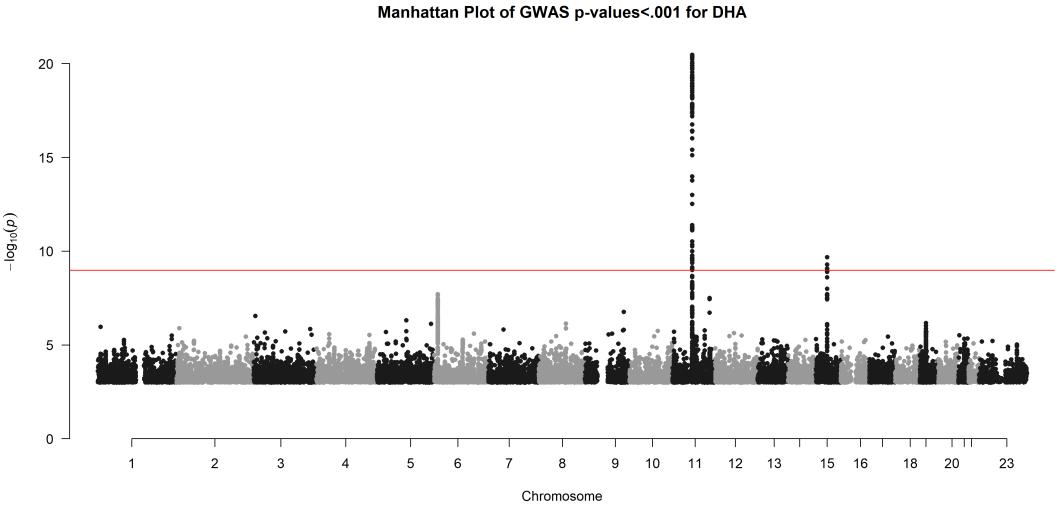


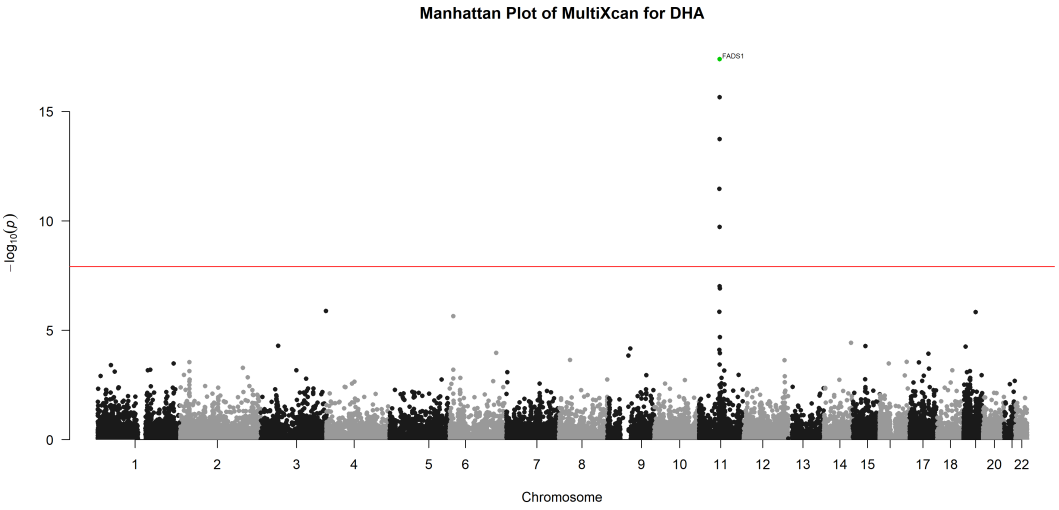
Docosahexaenoic acid (mM)



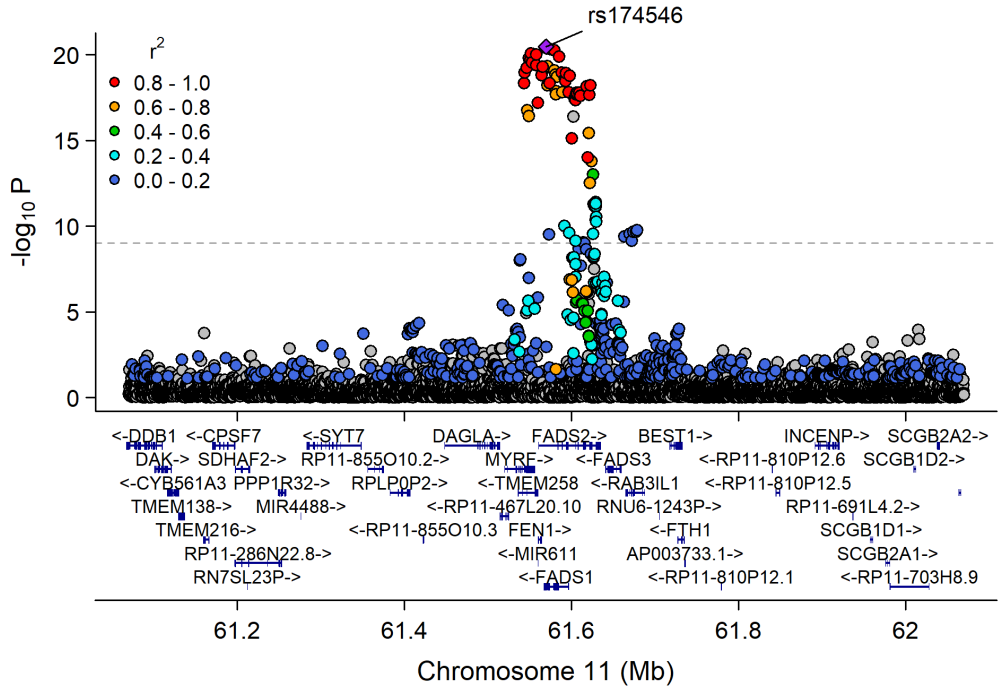




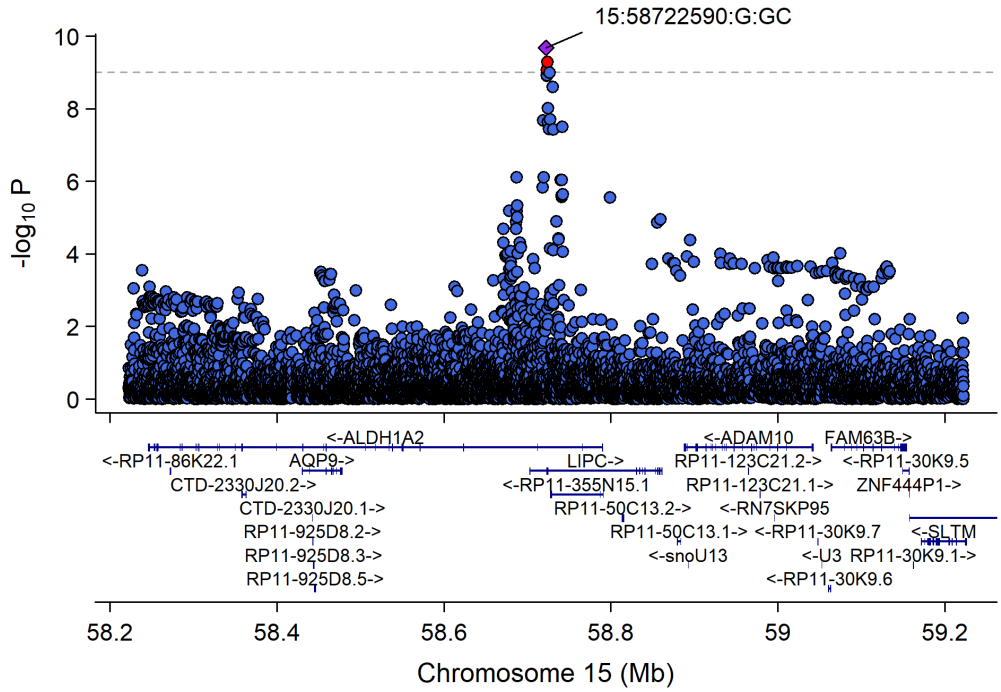




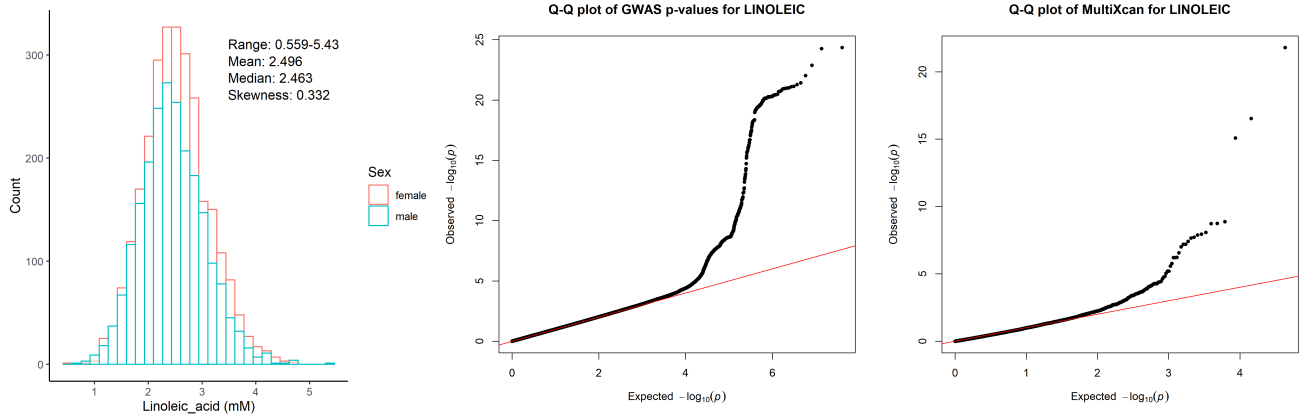
Chr11_59978355_62914375

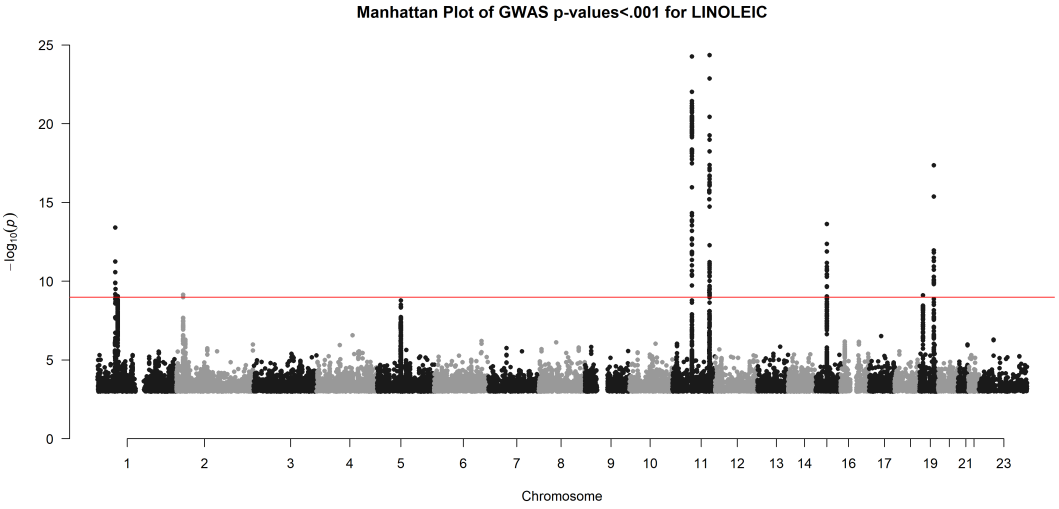


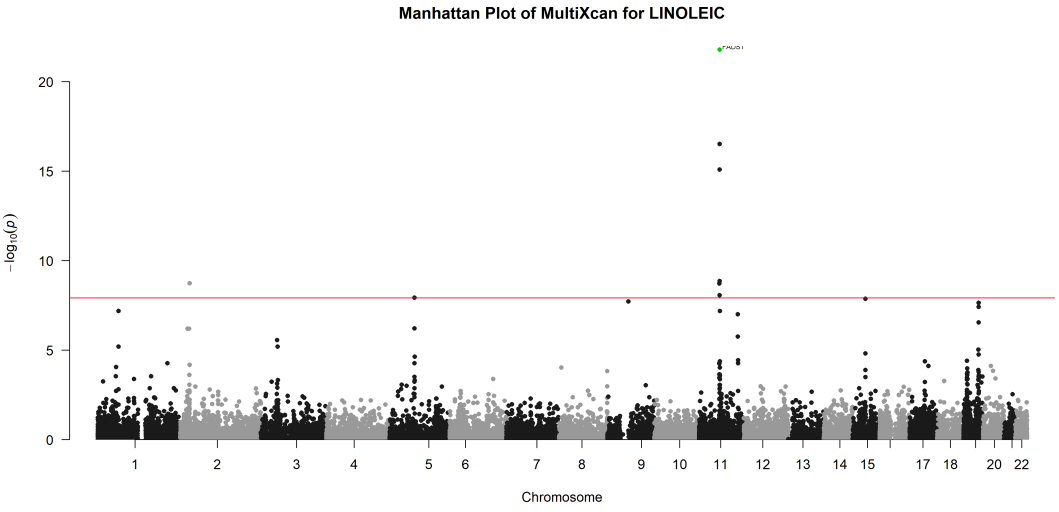
Chr15_57658798_60490883



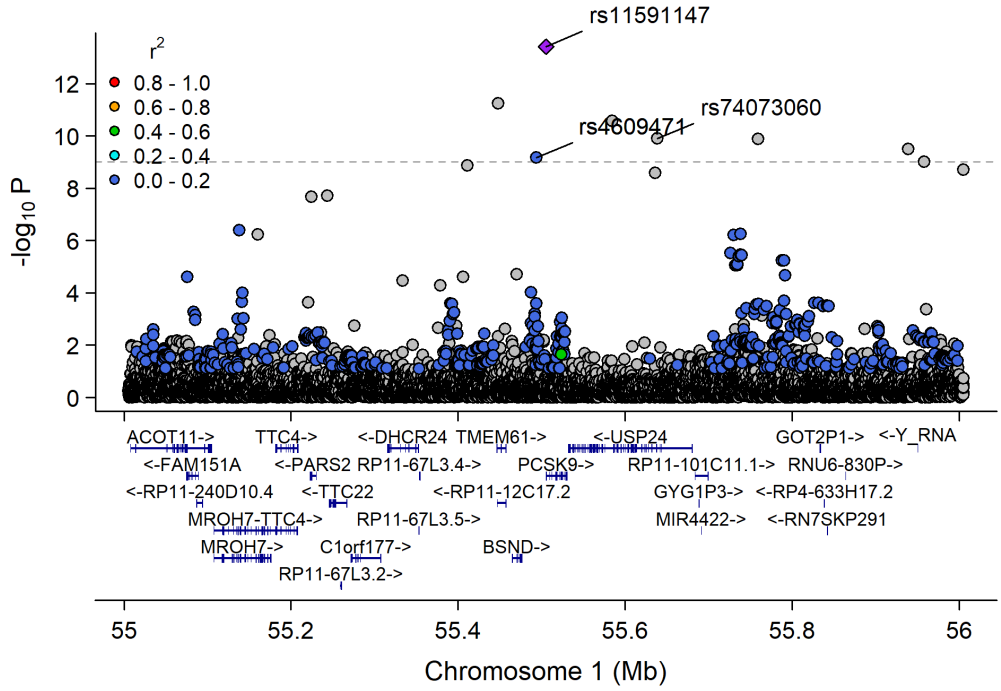
Linoleic acid (mM)



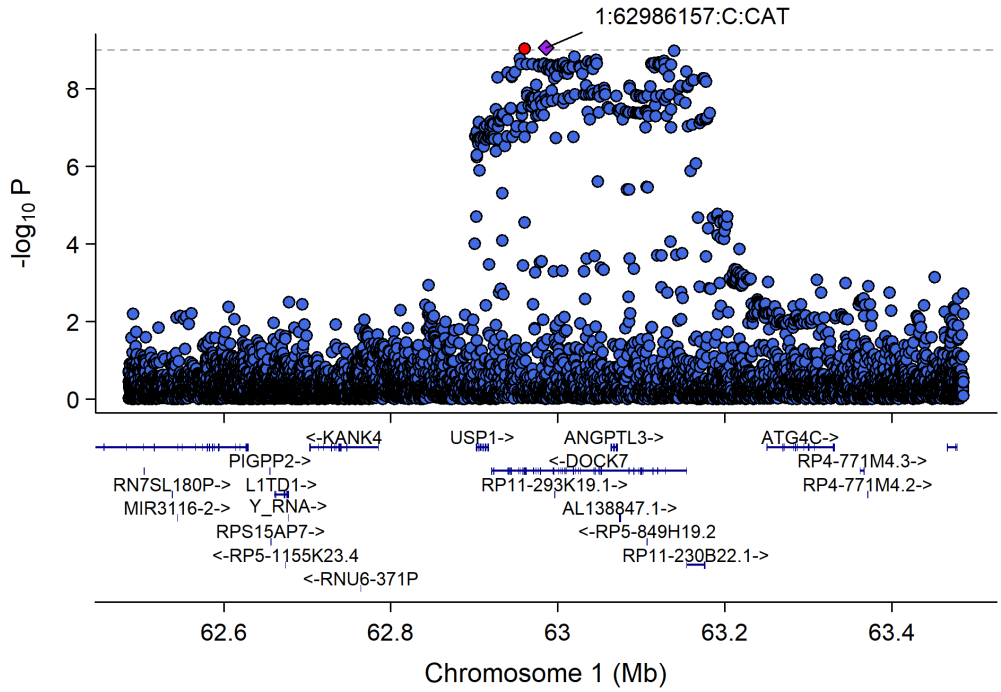




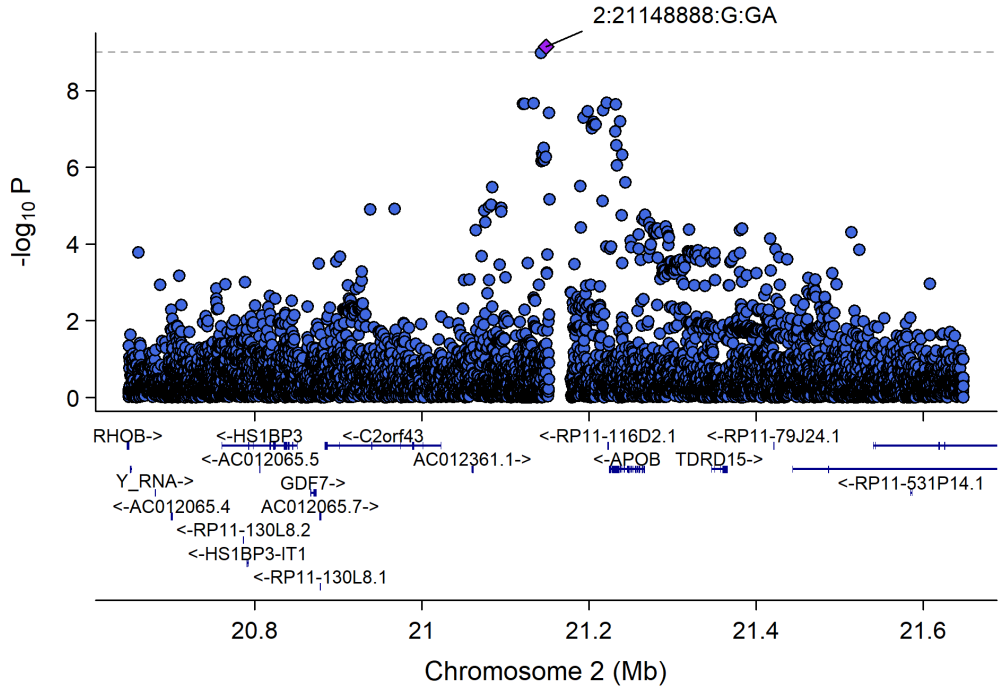
Chr1__53272879__57300396



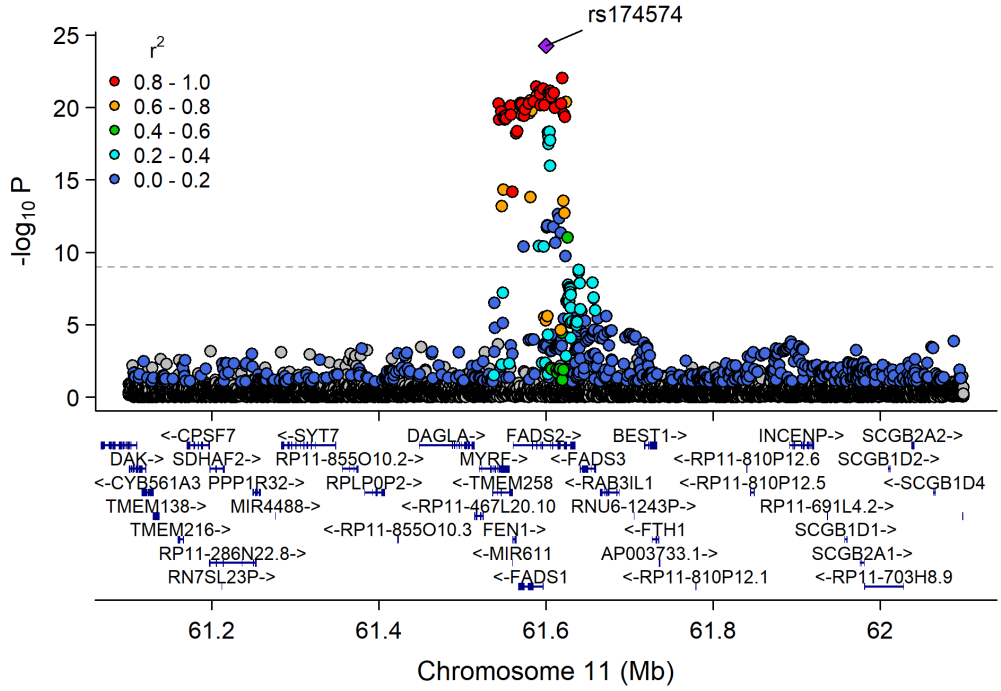
Chr1__62833402__63375116



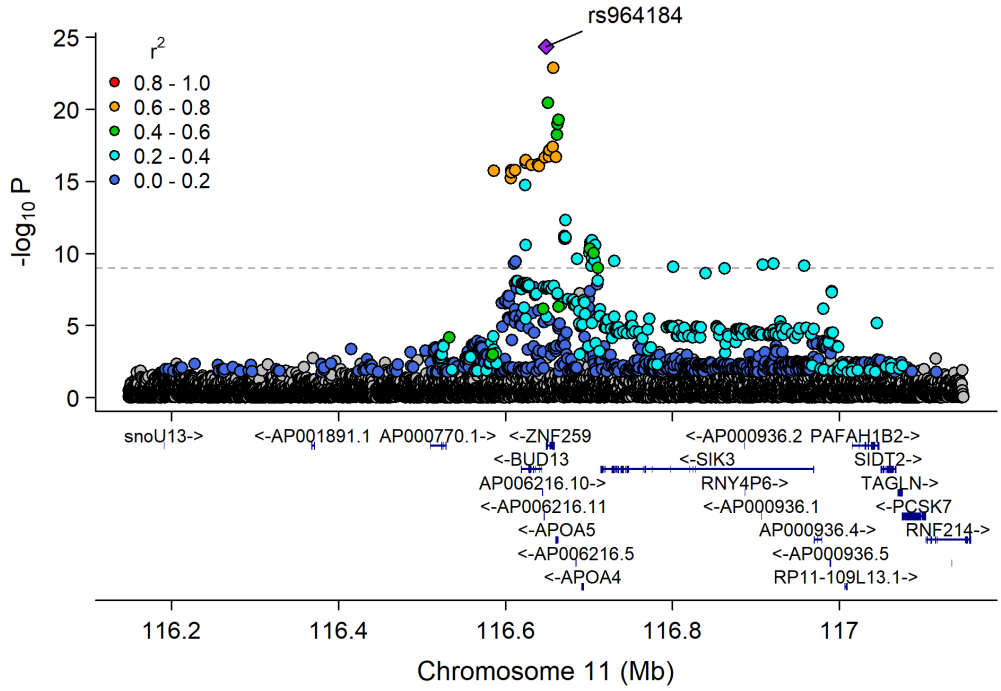
Chr2_19947287_22427492



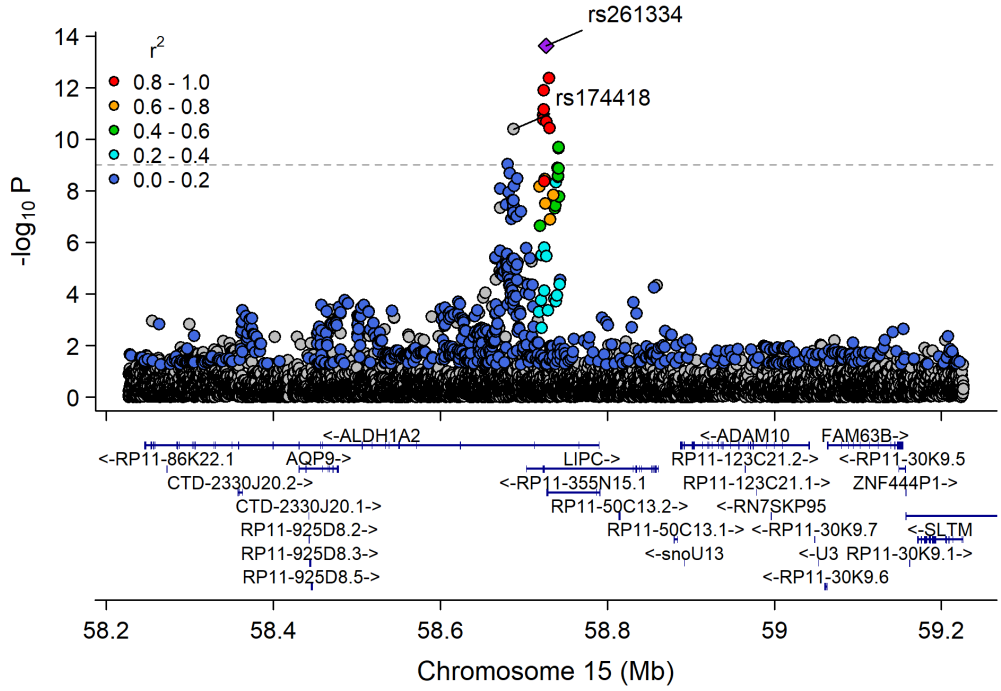
Chr11_59978355_62914375



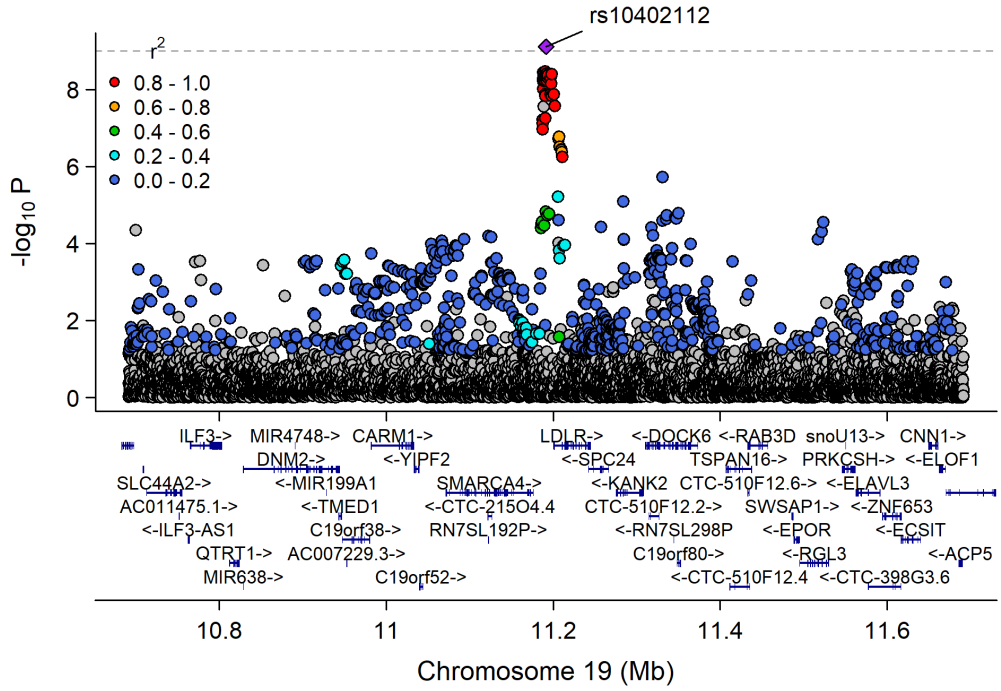
Chr11_115541901_117633315



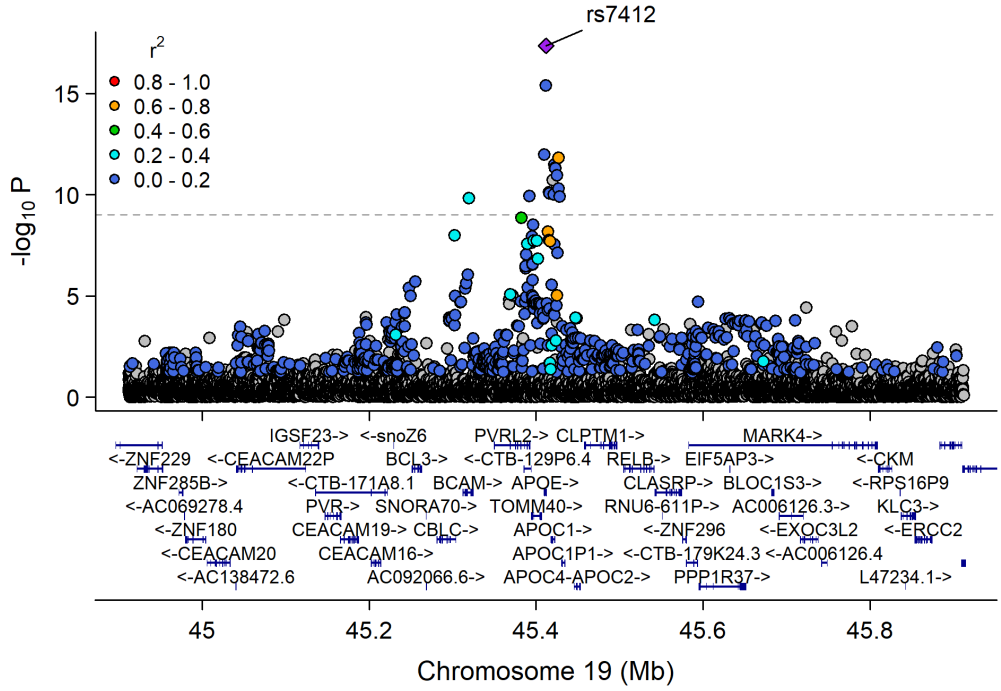
Chr15_57658798_60490883



Chr19_9807999_12255594

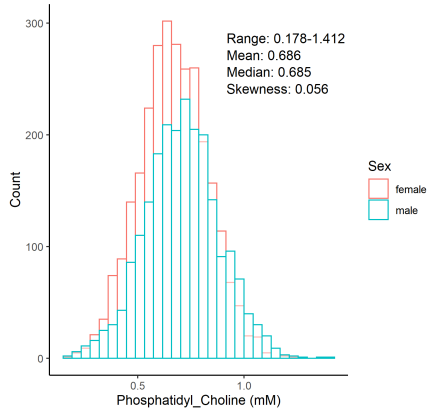


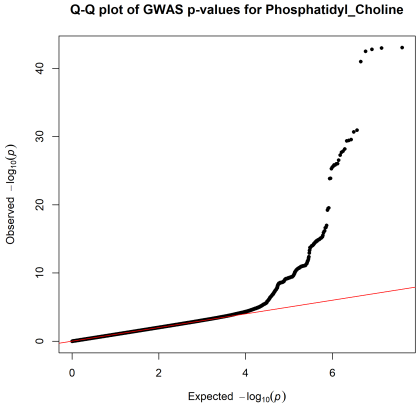
Chr19_44063363_46637375

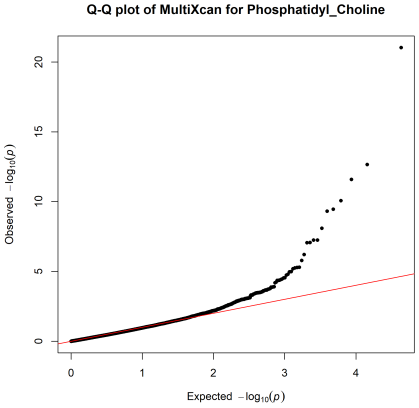


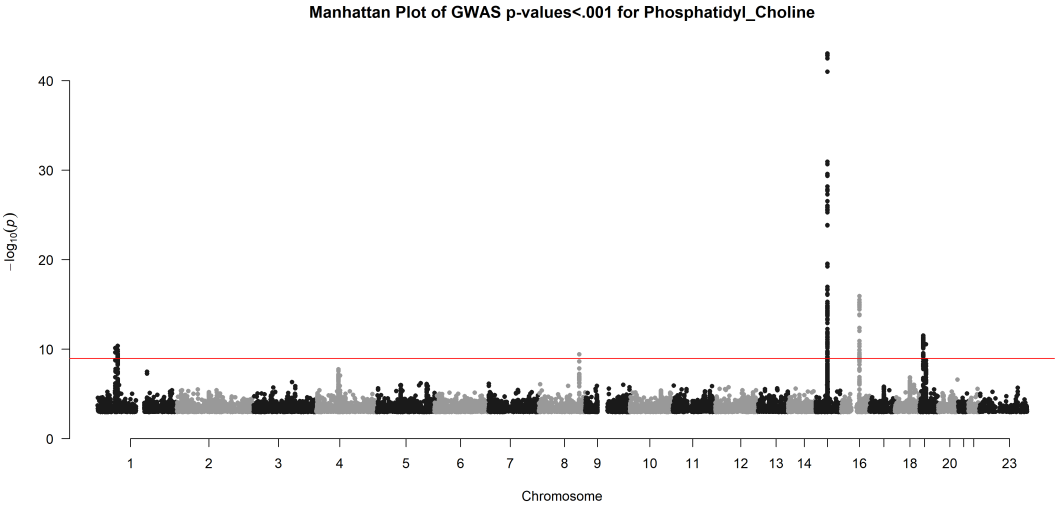
Glycerophospholipids

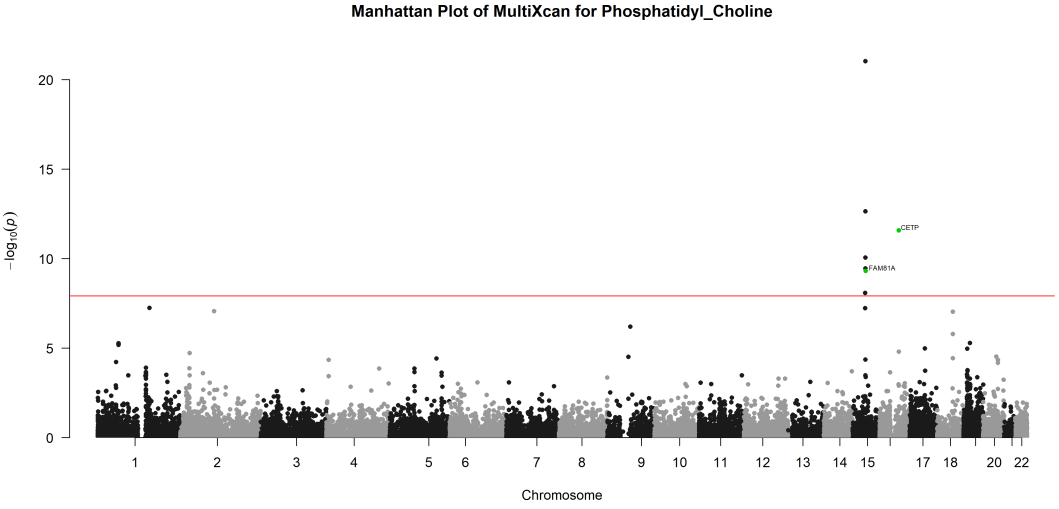
Phosphatidylcholine (mM)



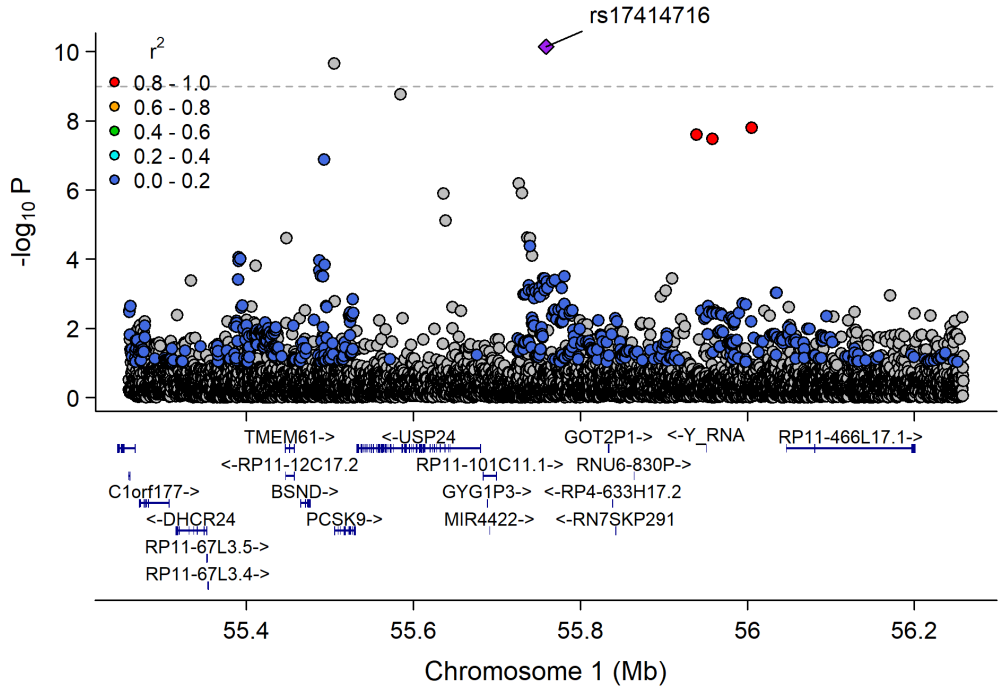




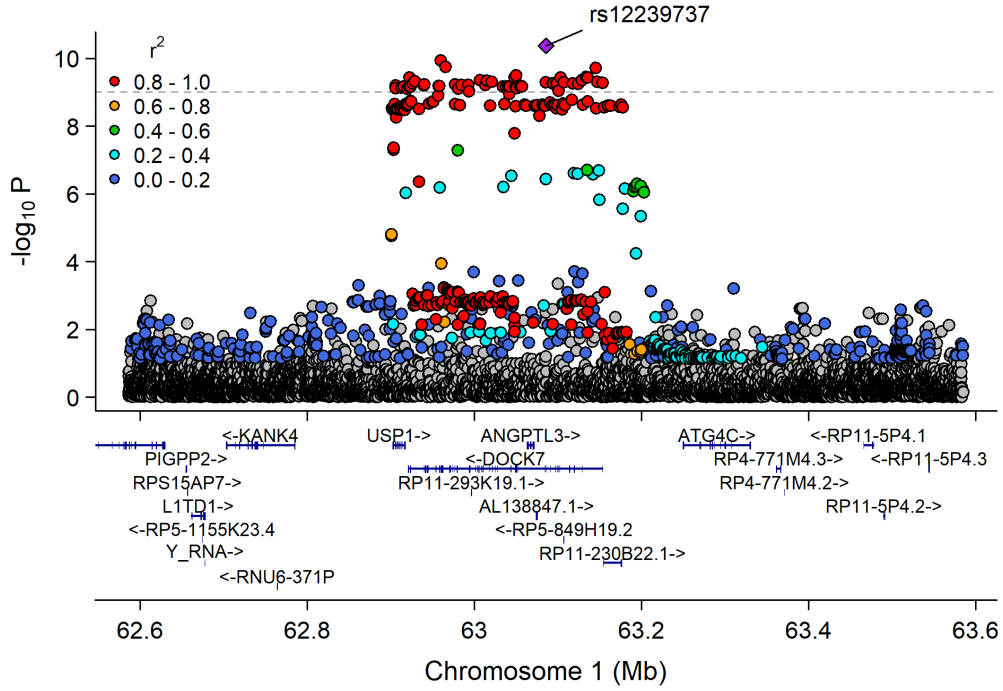




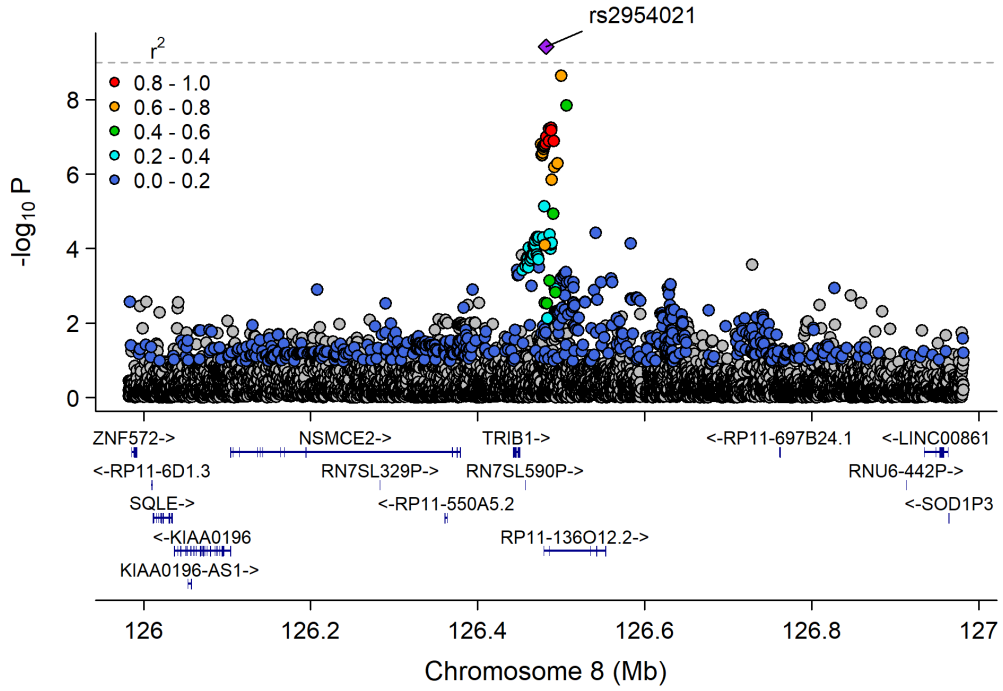
Chr1__53272879__57300396



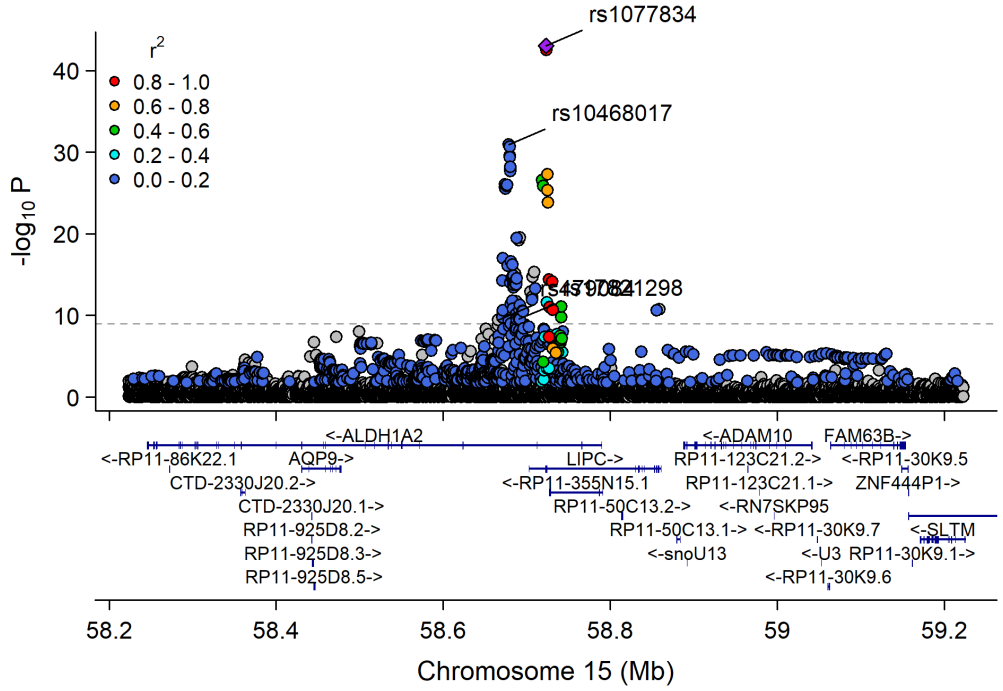
Chr1__62833402__63375116



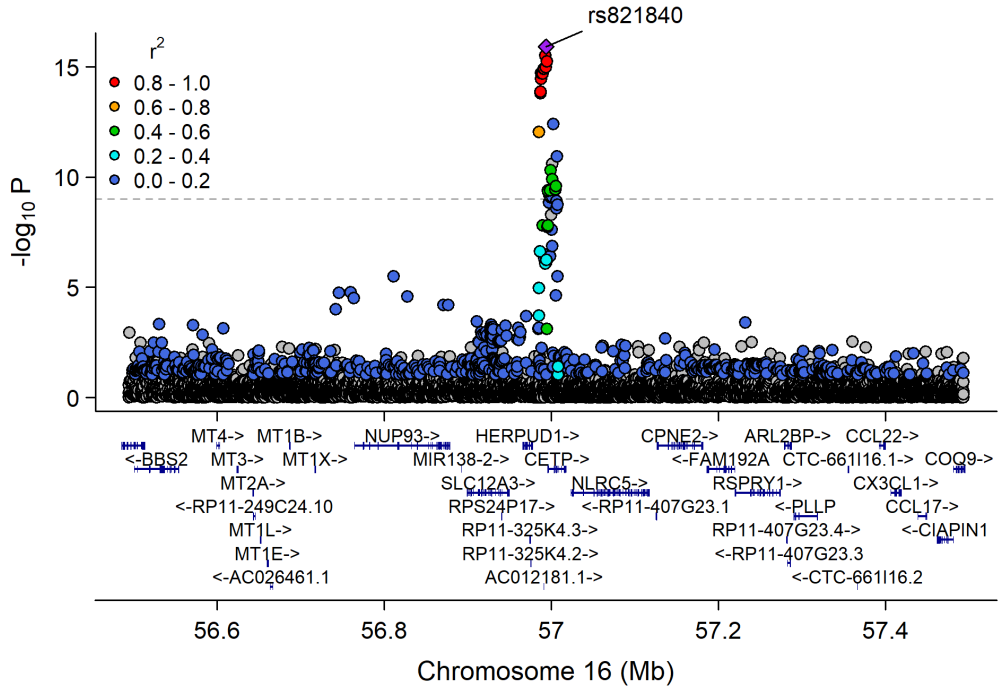
Chr8_126435663_126578061



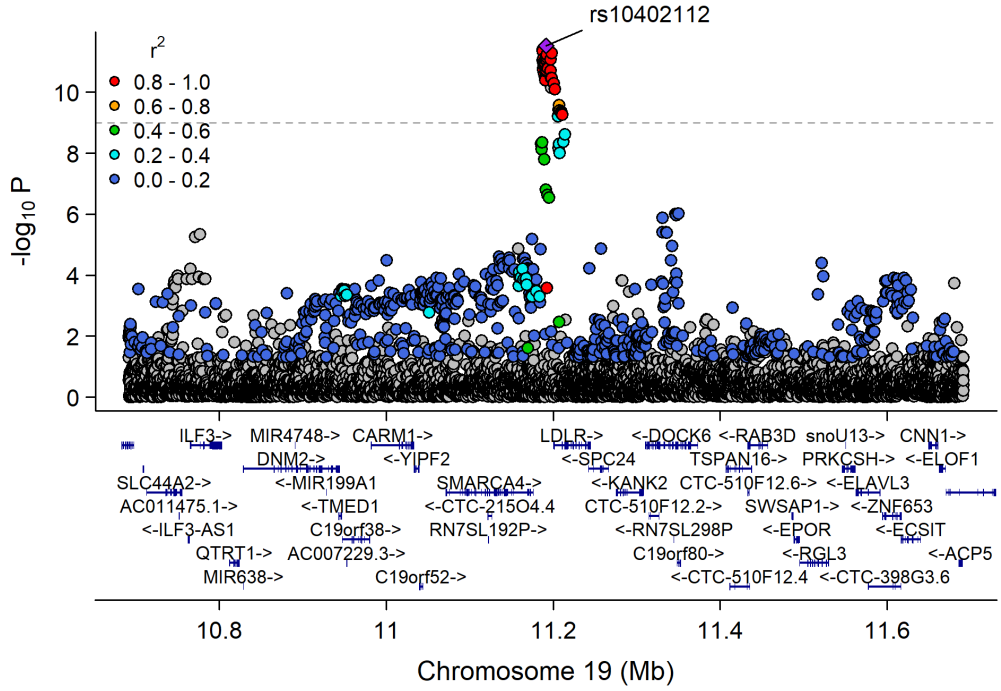
Chr15_57658798_60490883



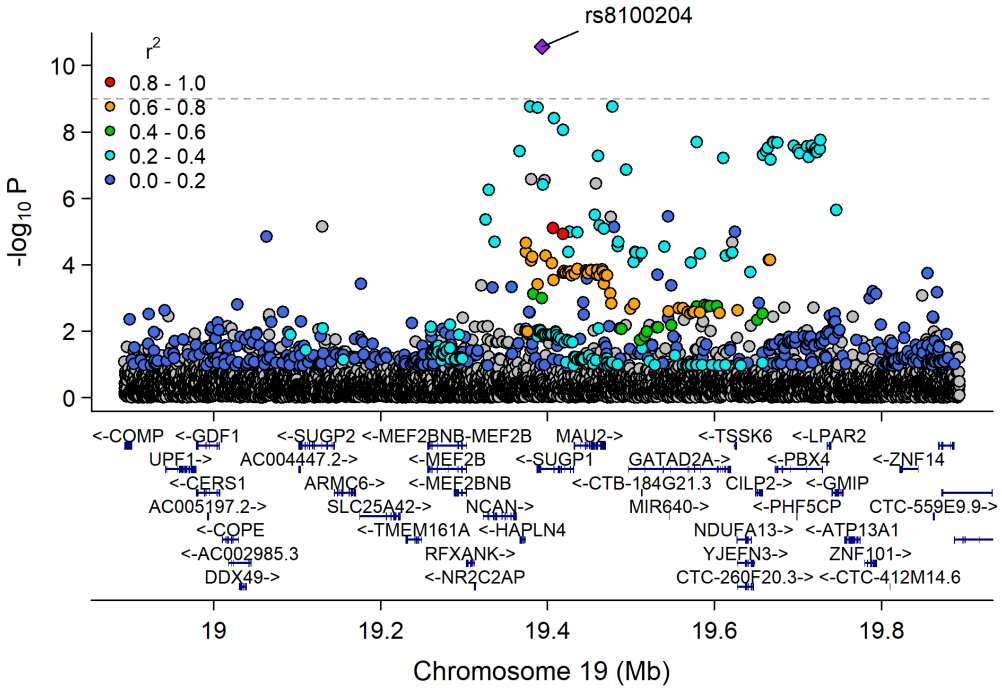
Chr16_55870822_57992421



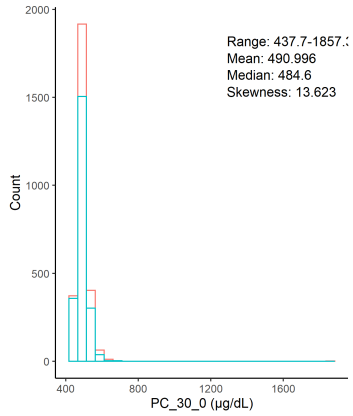
Chr19_9807999_12255594

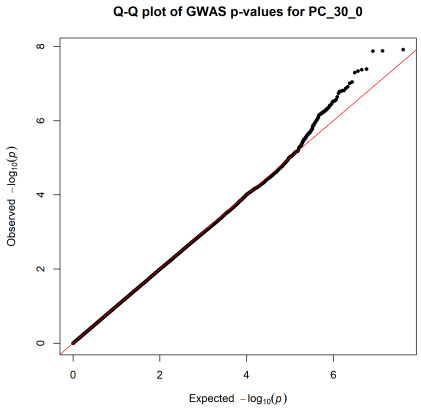


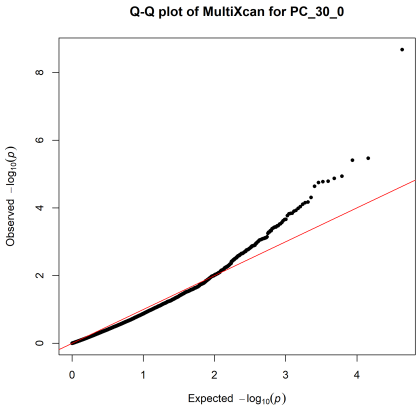
Chr19_18370495_20841464

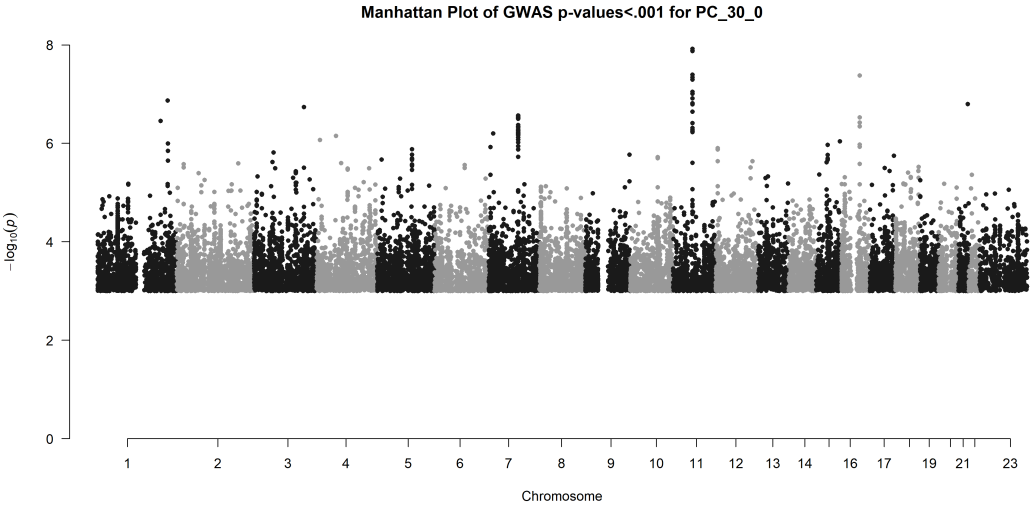


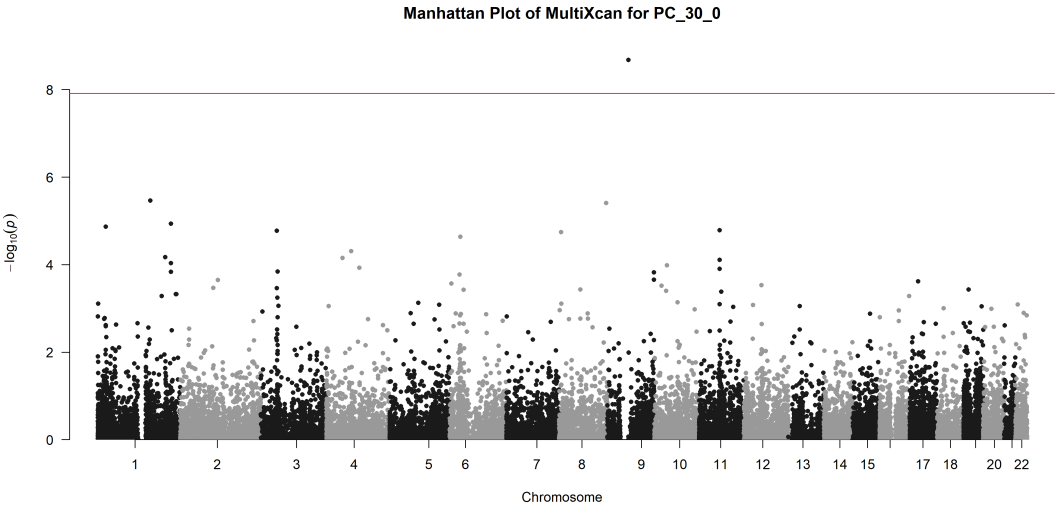
Phosphatidylcholine diacyl C30:0 (µg/dL)



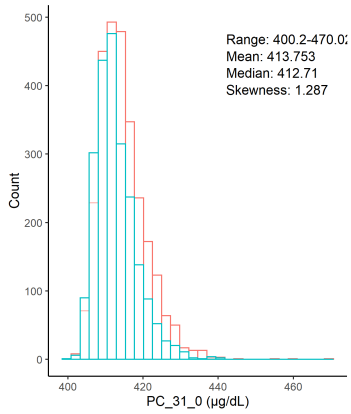


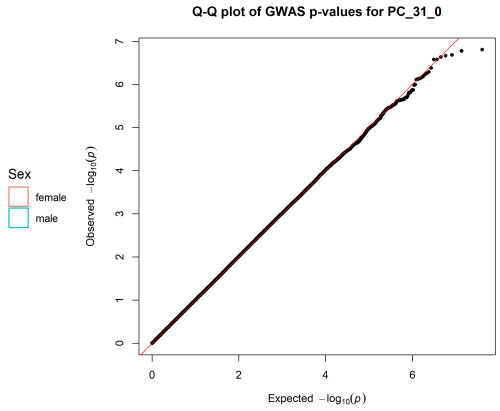


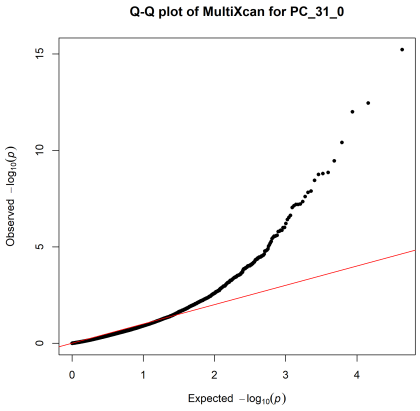


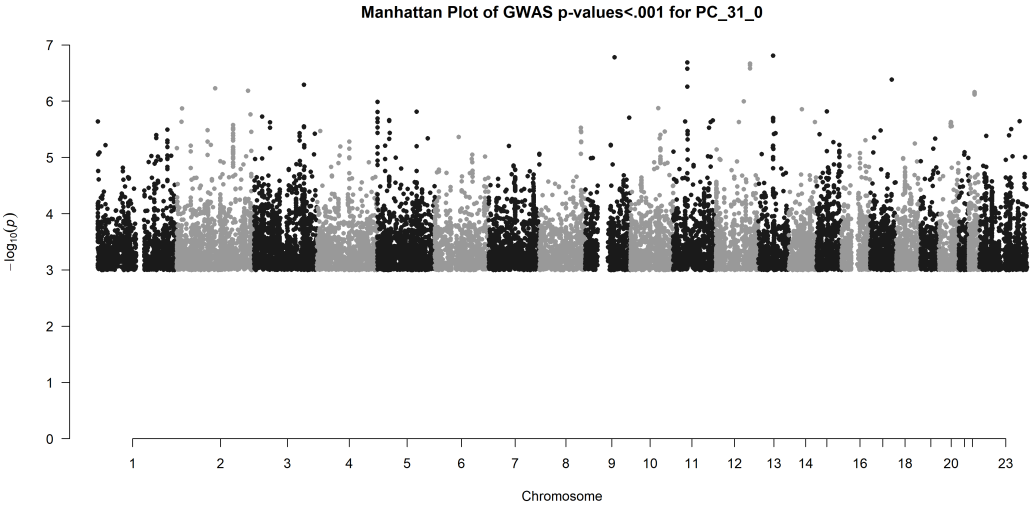


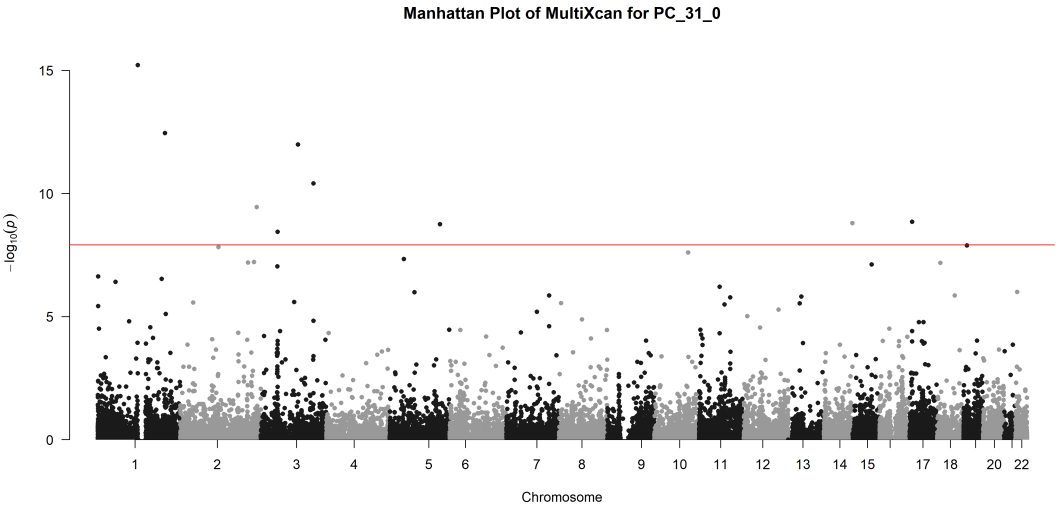
Phosphatidylcholine diacyl C31:0 (µg/dL)



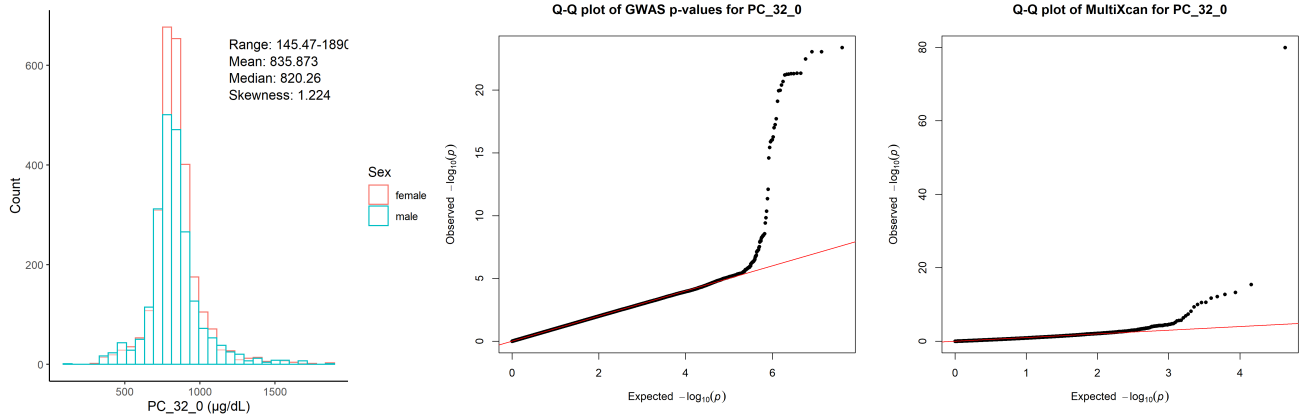


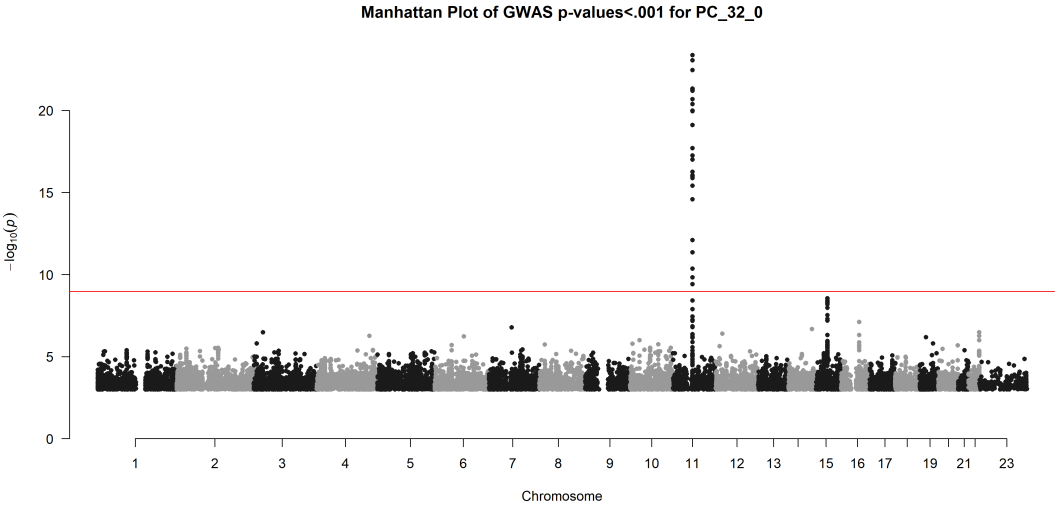


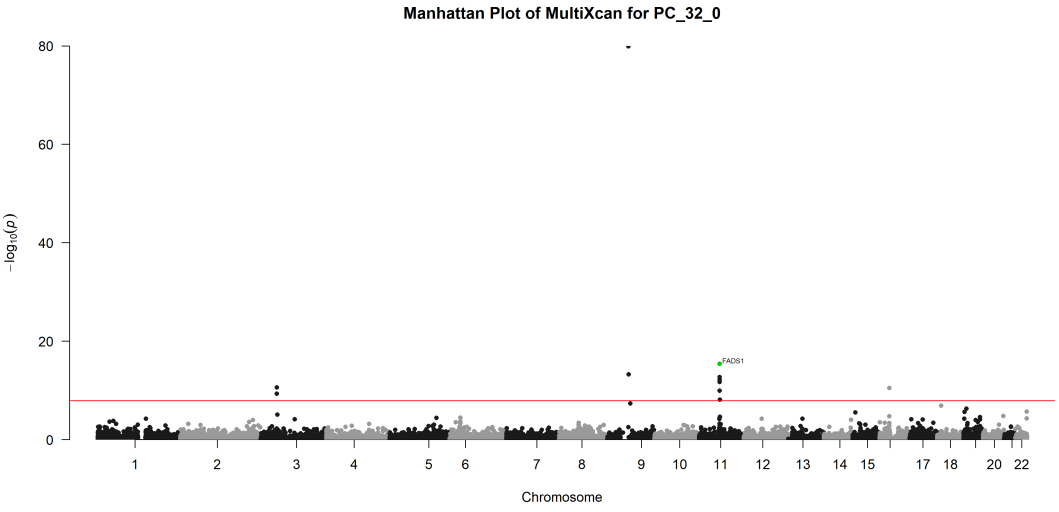




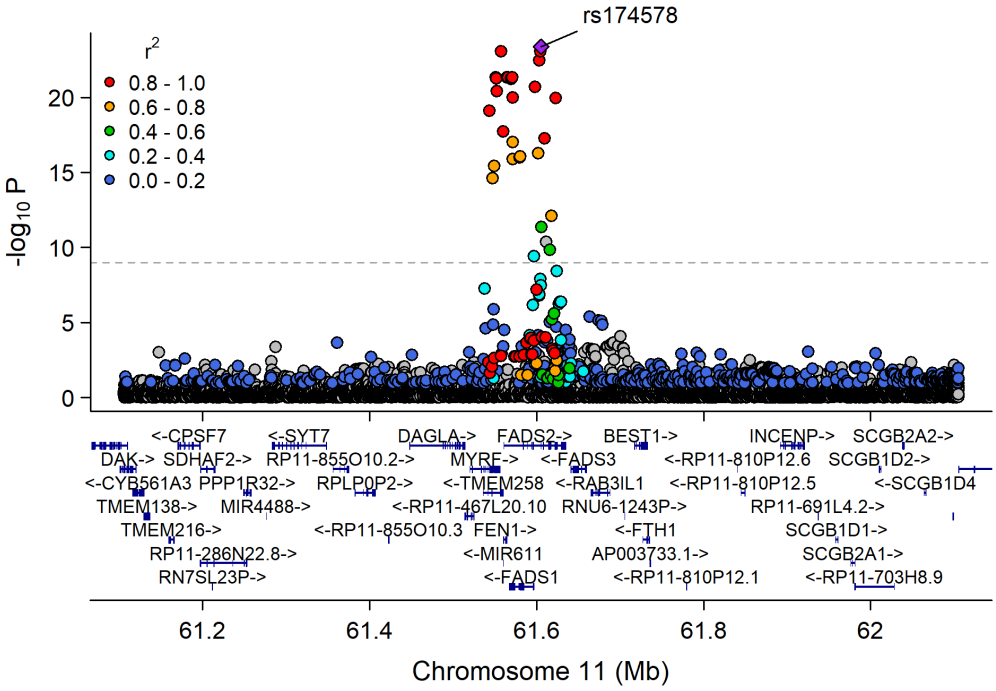
Phosphatidylcholine diacyl C32:0 (µg/dL)



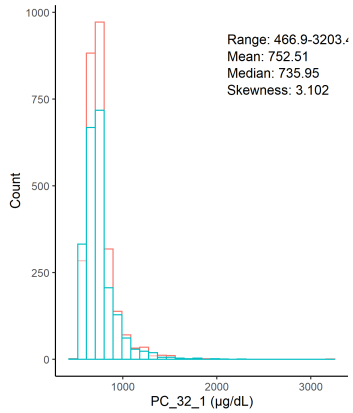


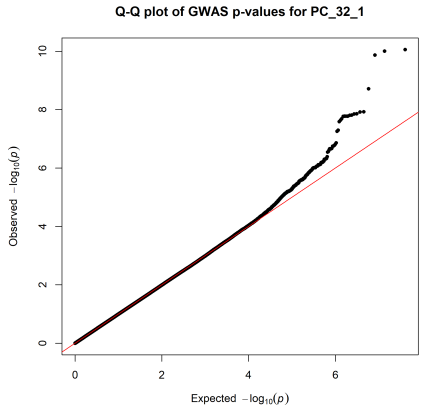


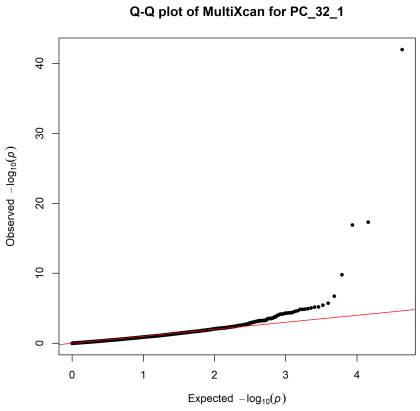
Chr11_59978355_62914375

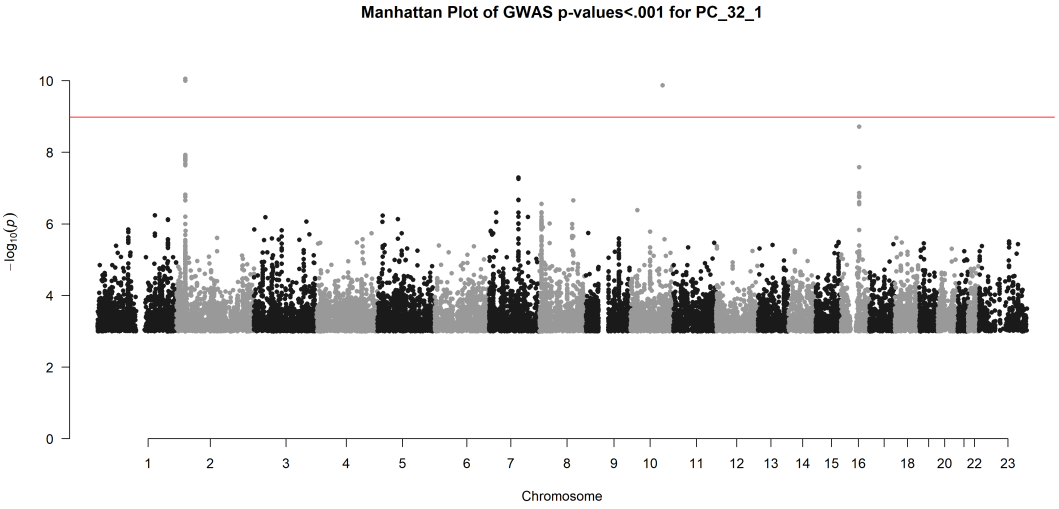


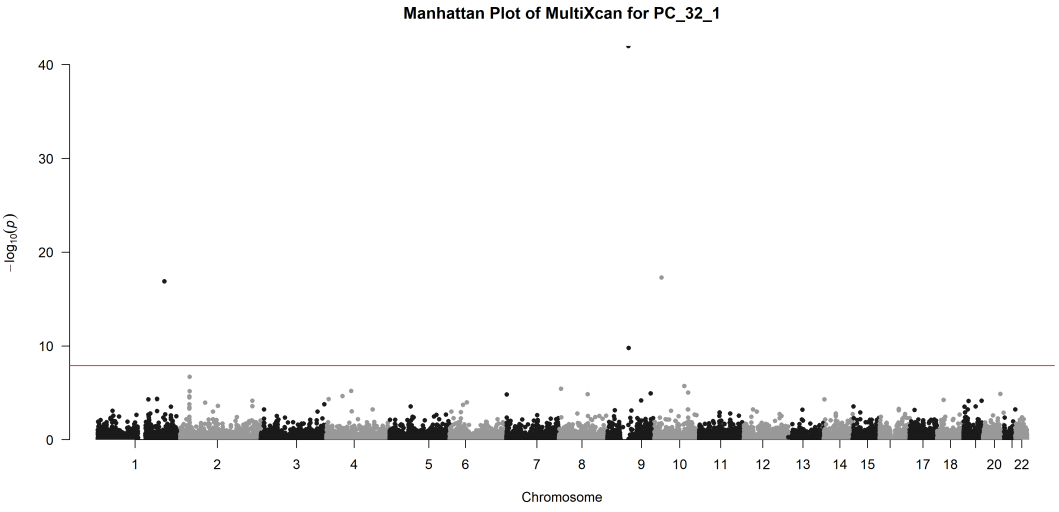
Phosphatidylcholine diacyl C32:1 (µg/dL)



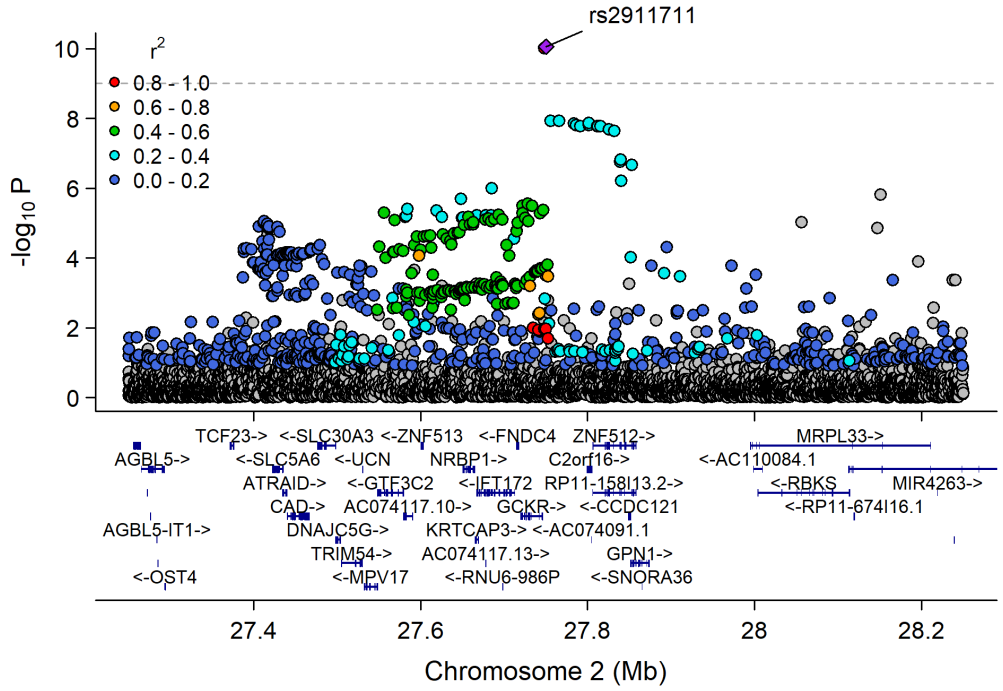




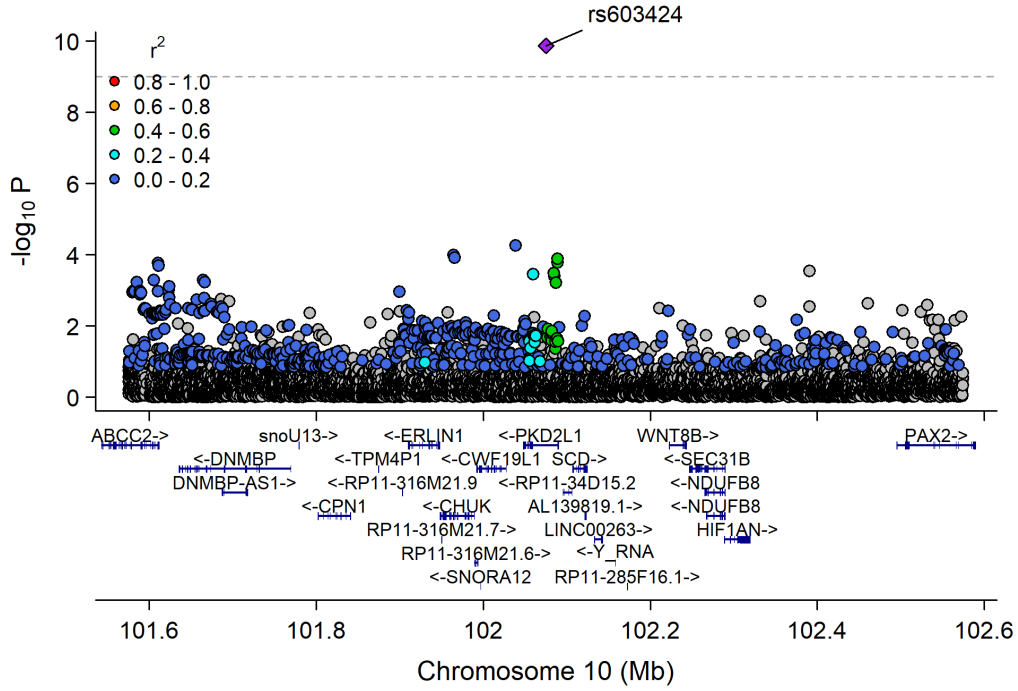




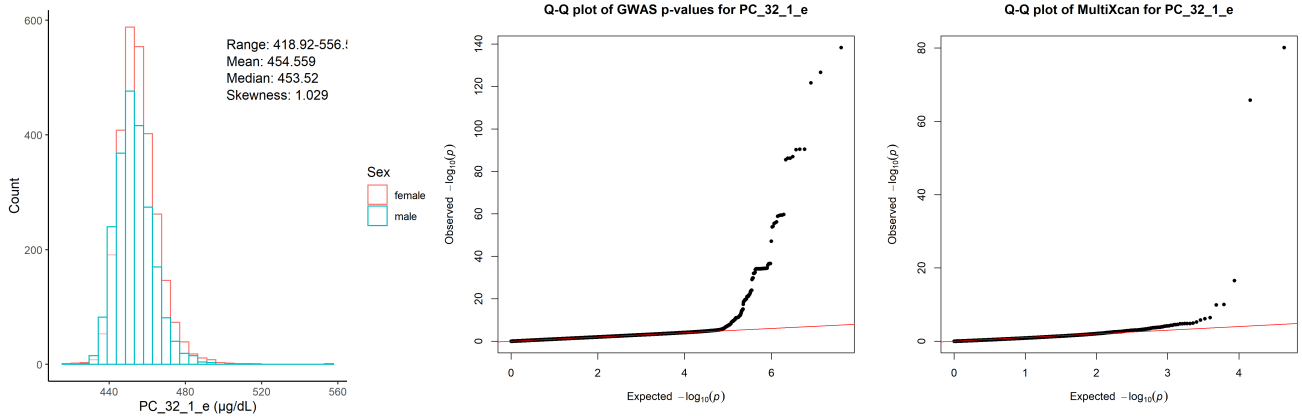
Chr2_26753815_28597624

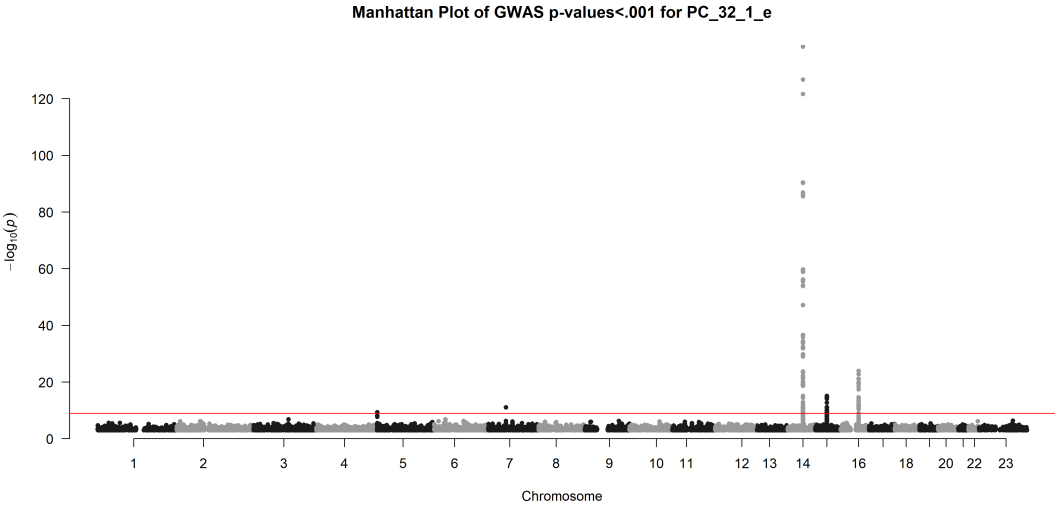


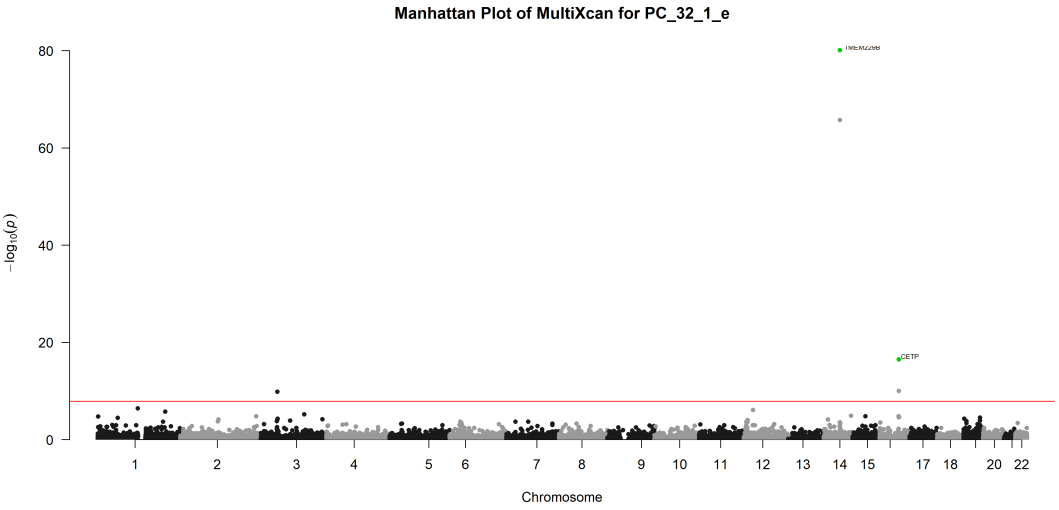
Chr10_101403323_102214579



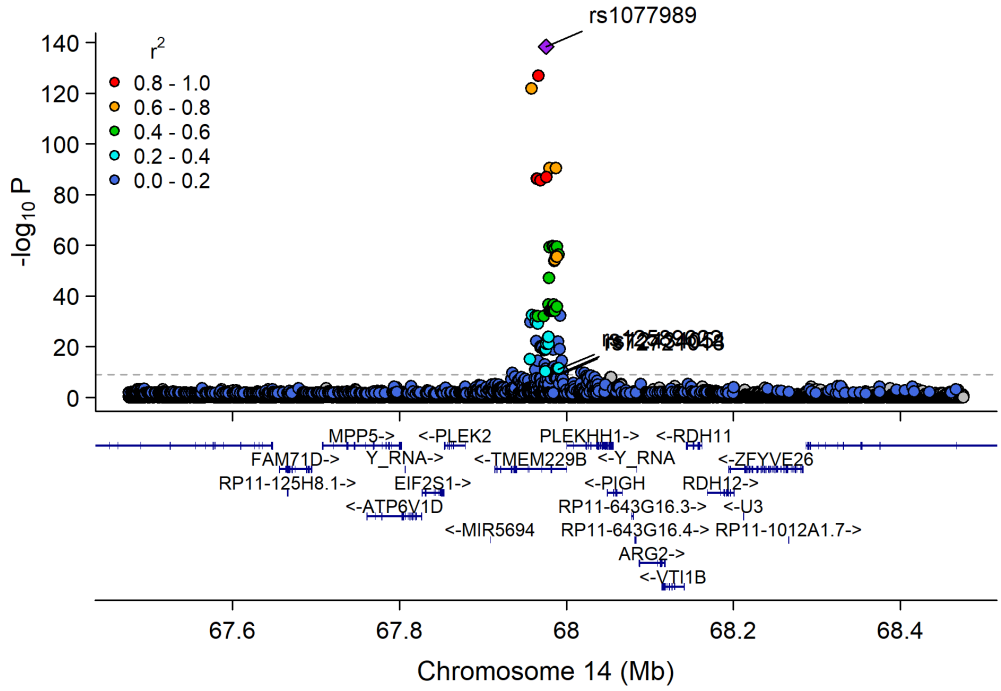
Phosphatidylcholine acyl-alkyl C32:1 (µg/dL)



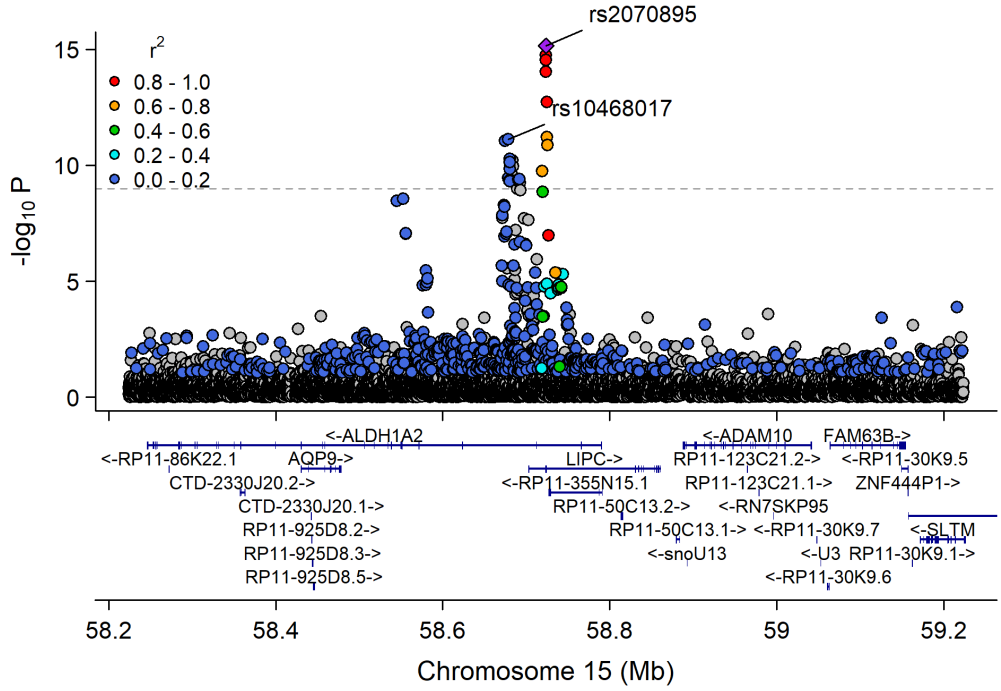




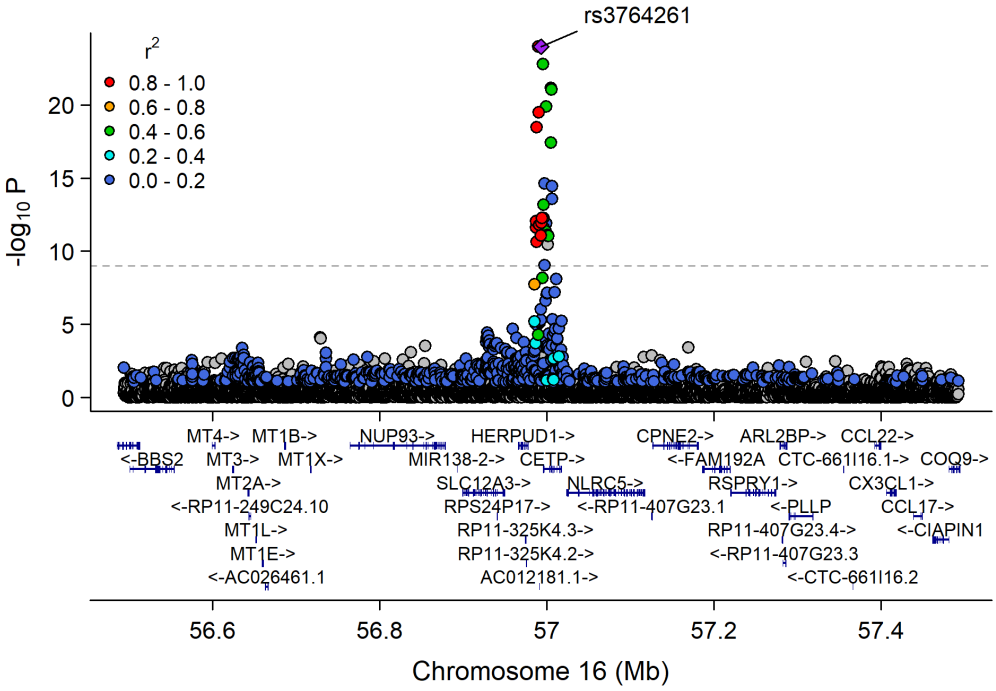
Chr14_66997265_68972354



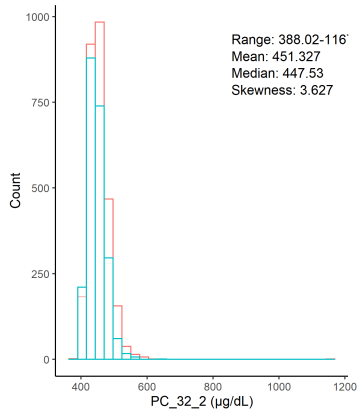
Chr15_57658798_60490883

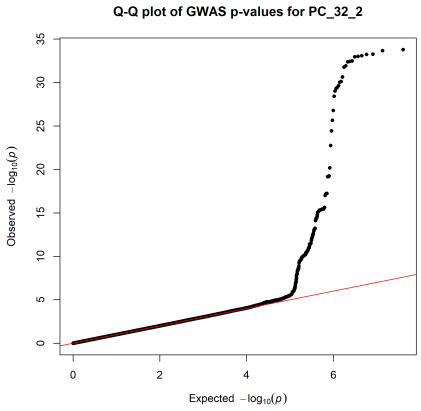


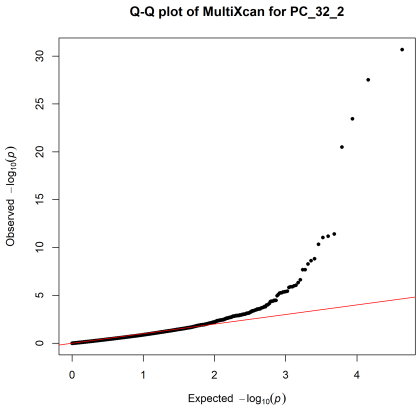
Chr16_55870822_57992421

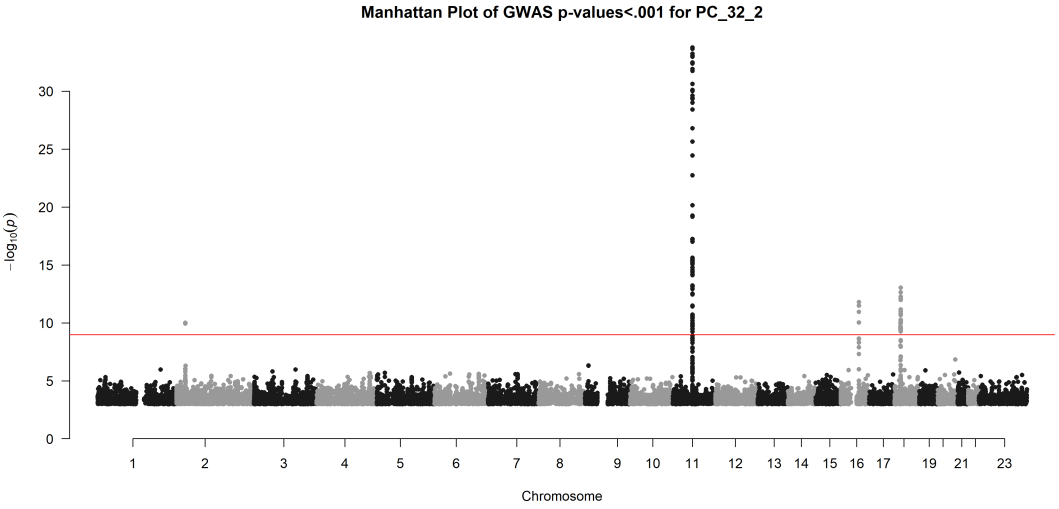


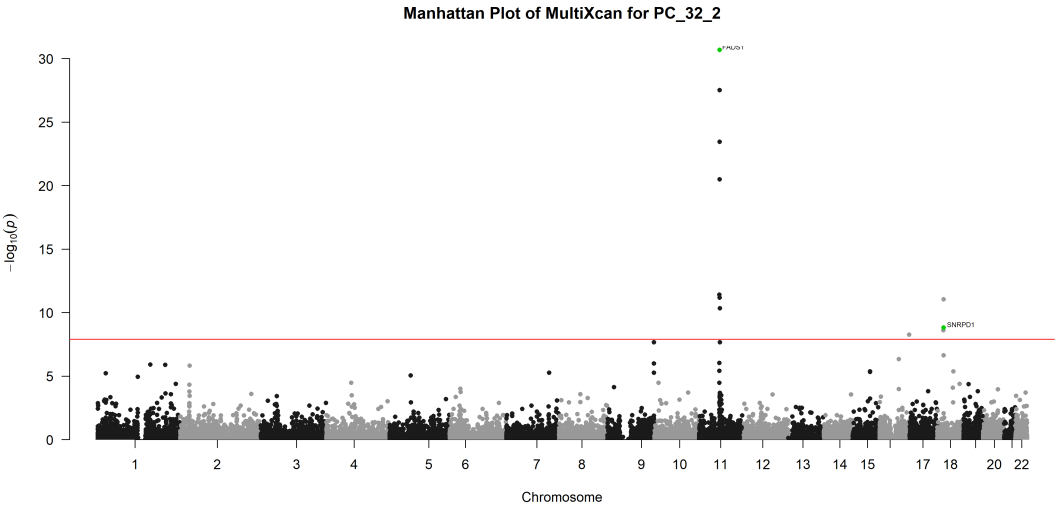
Phosphatidylcholine diacyl C32:2 (µg/dL)



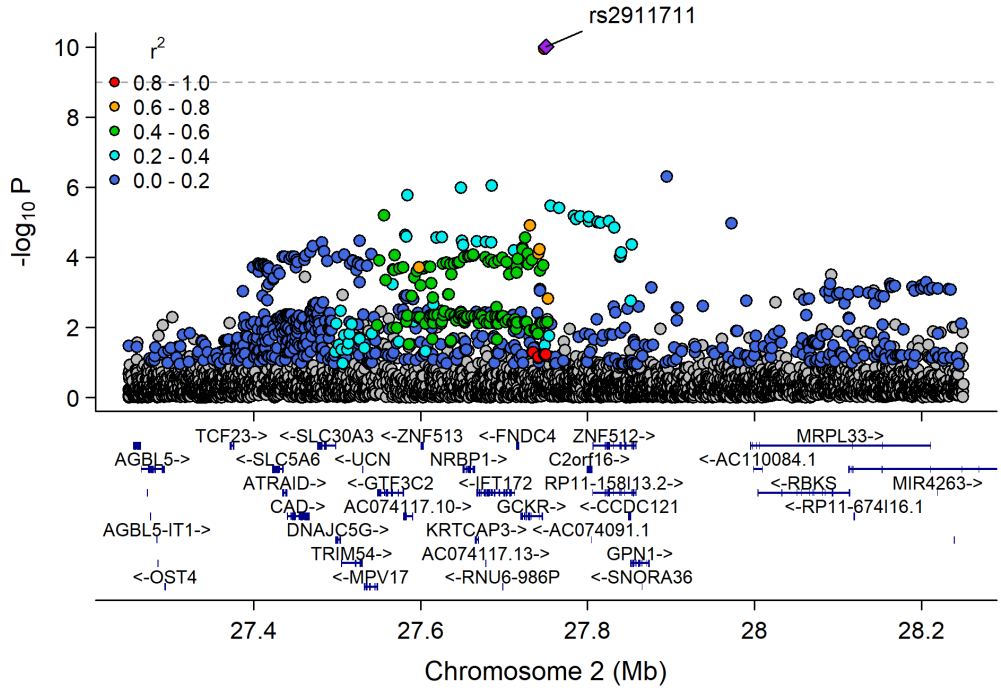




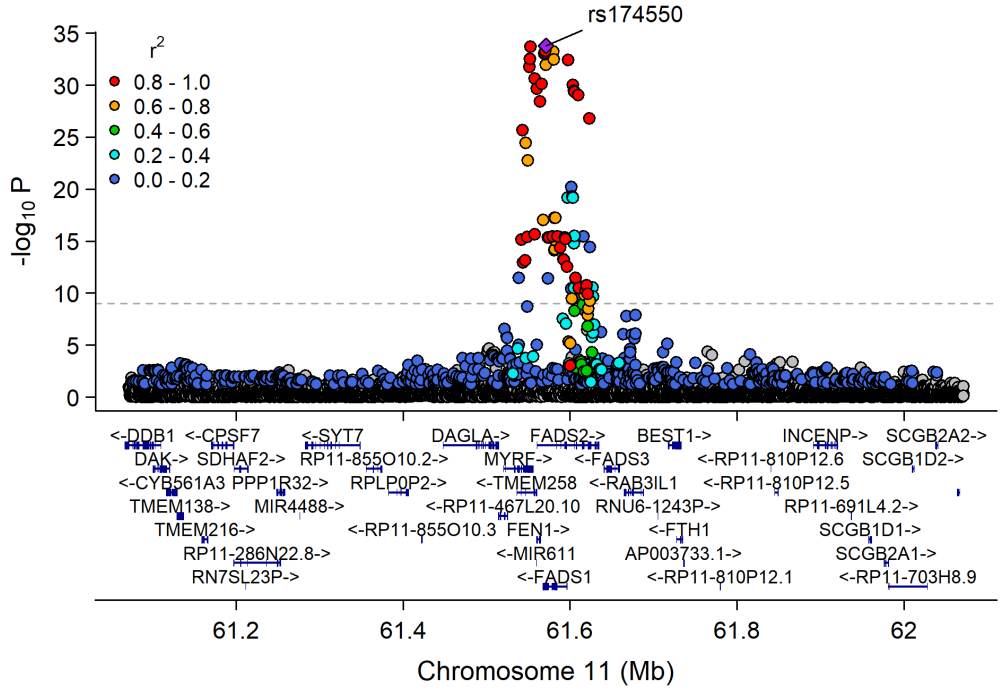




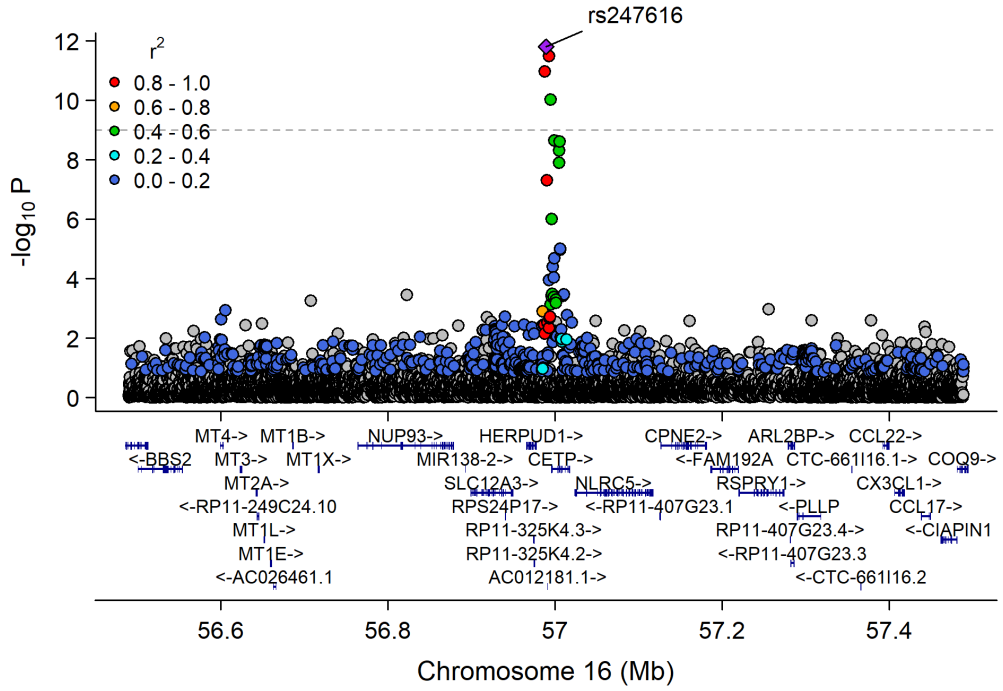
Chr2_26753815_28597624



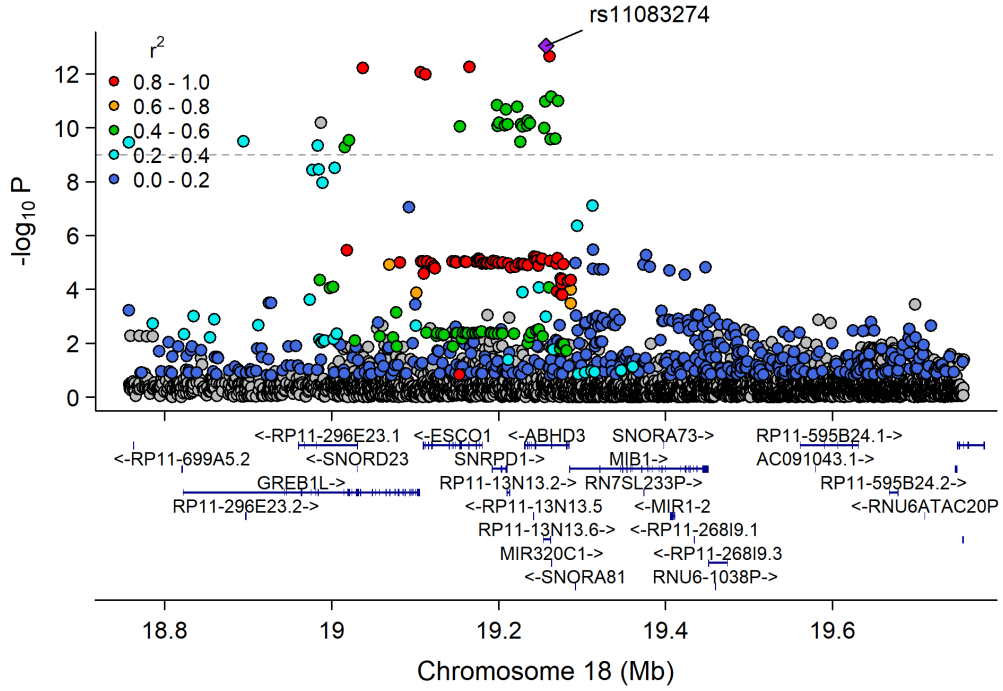
Chr11_59978355_62914375



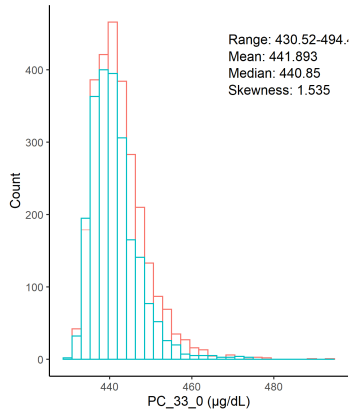
Chr16_55870822_57992421

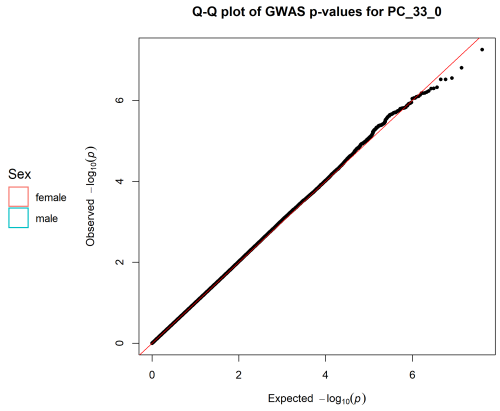


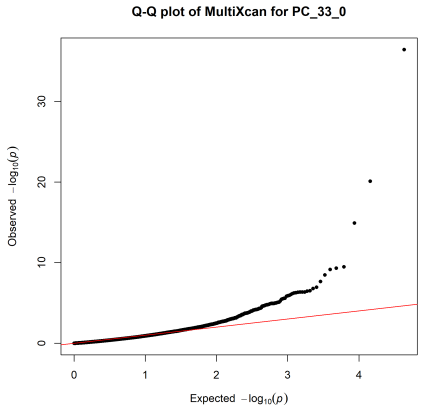
Chr18_18514118_19486319

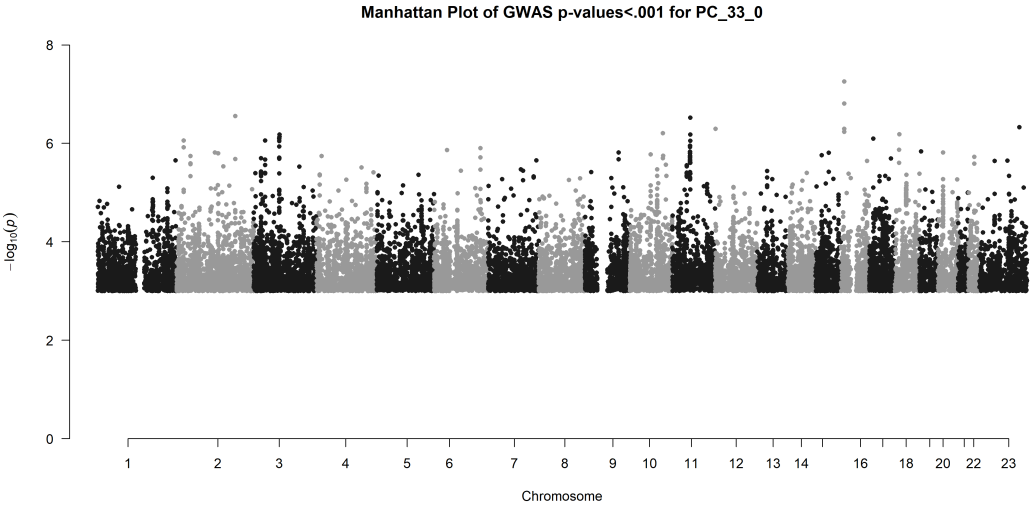


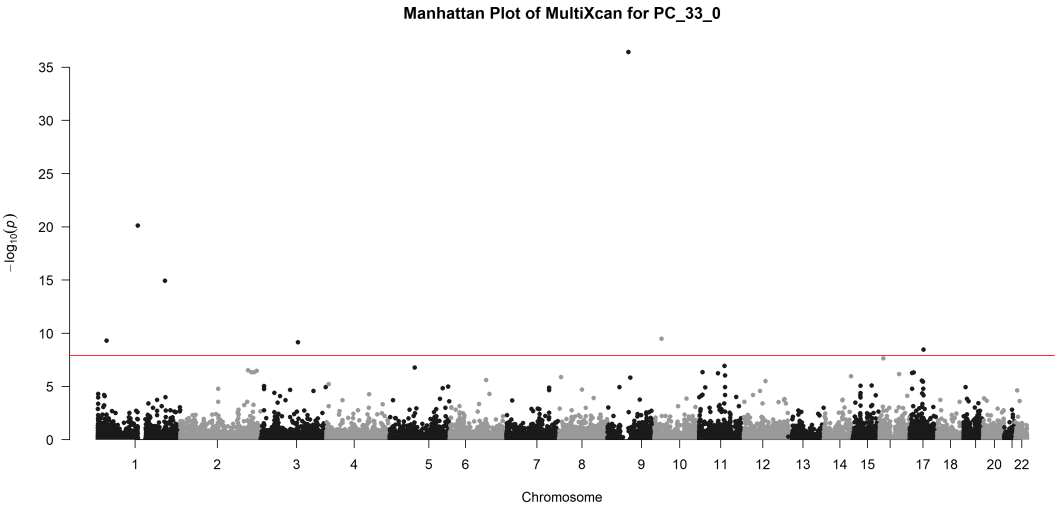
Phosphatidylcholine diacyl C33:0 (µg/dL)



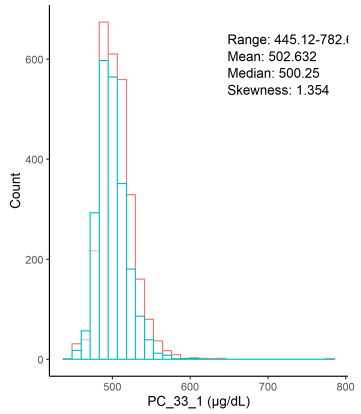


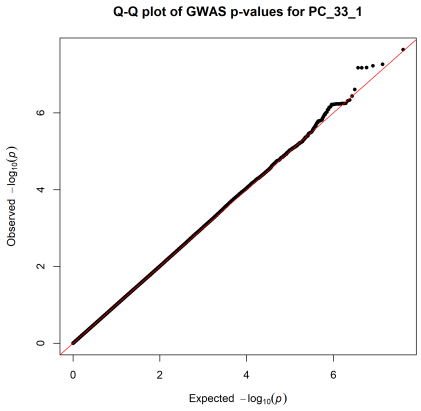


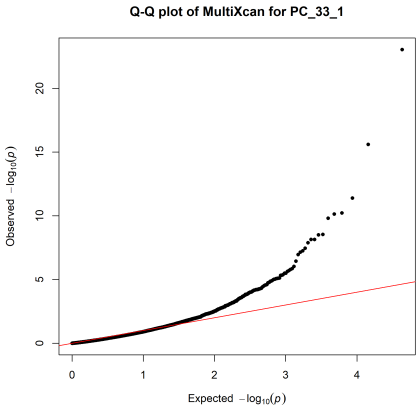


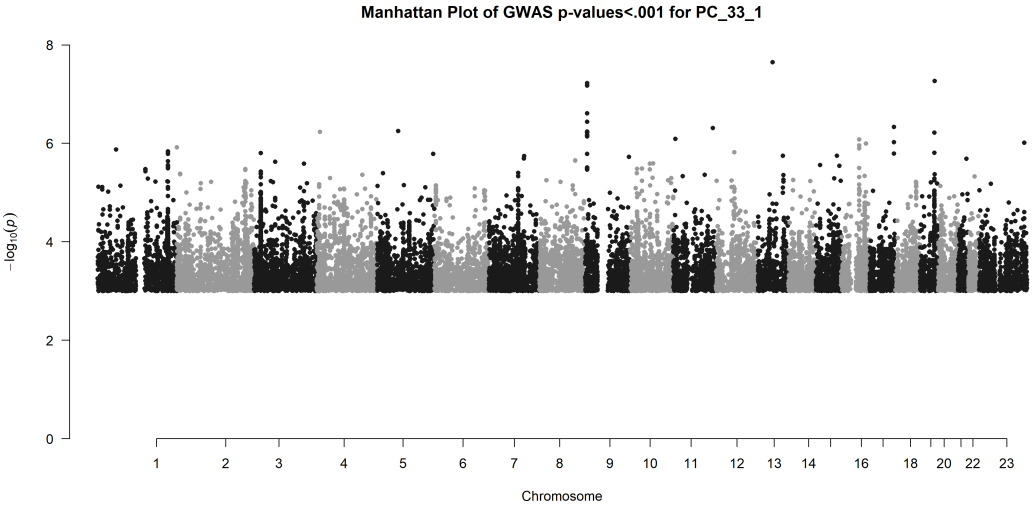


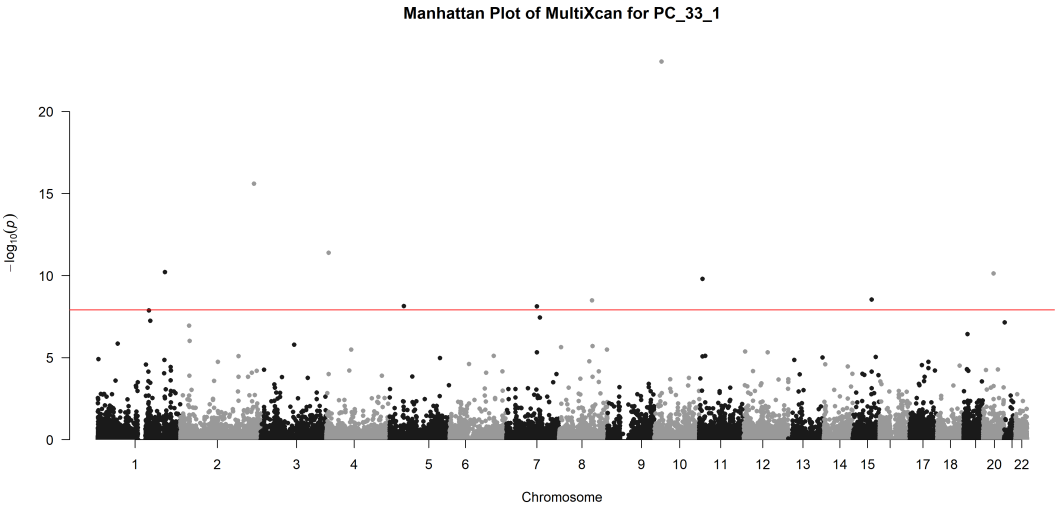
Phosphatidylcholine diacyl C33:1 (µg/dL)



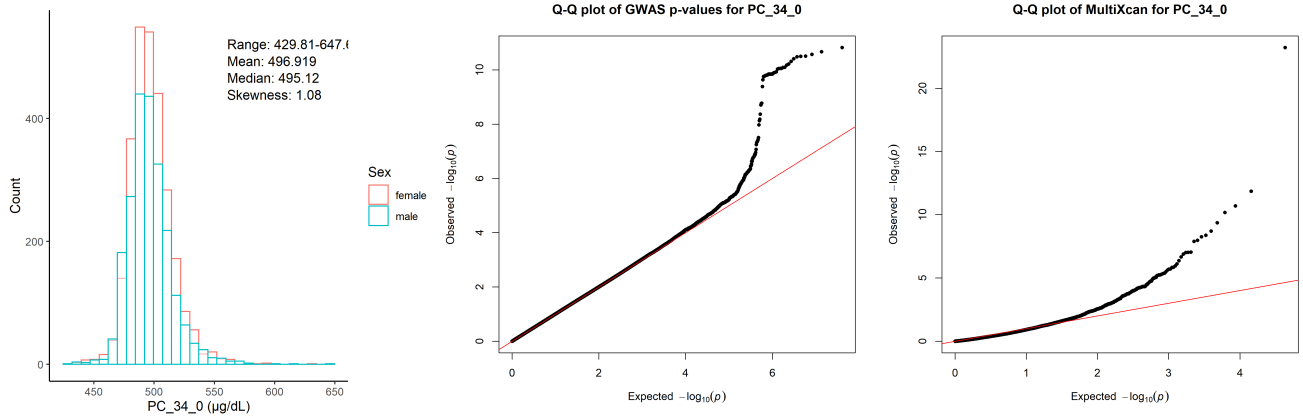


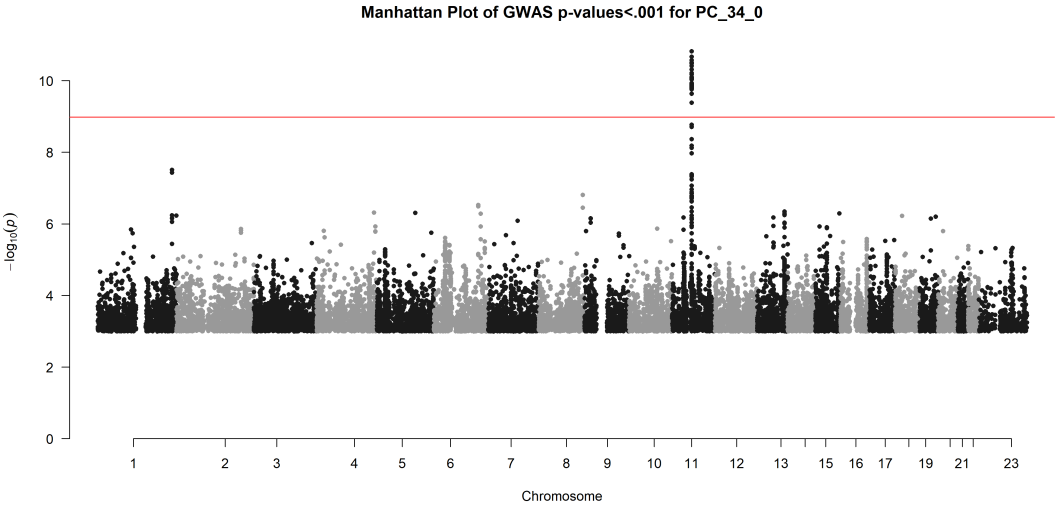


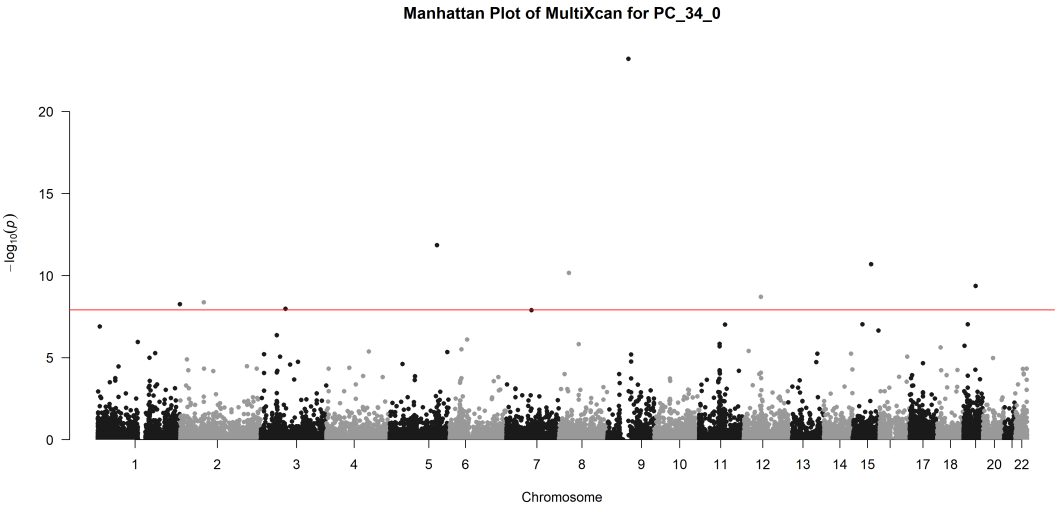




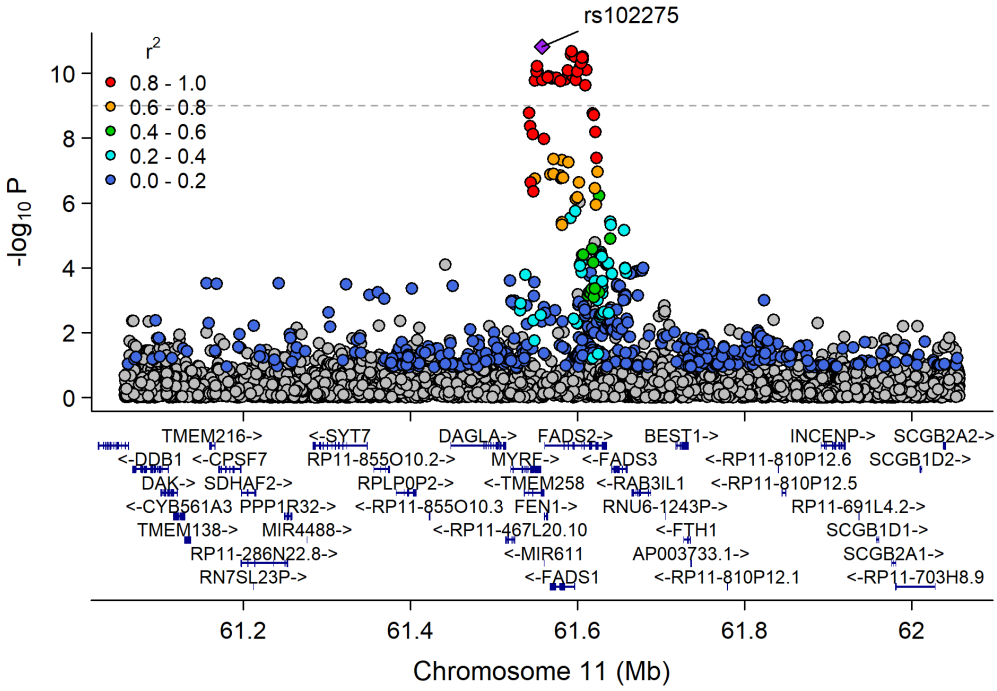
Phosphatidylcholine diacyl C34:0 (µg/dL)



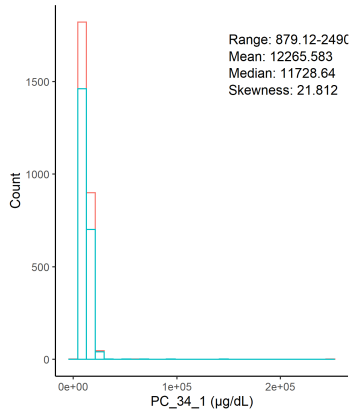


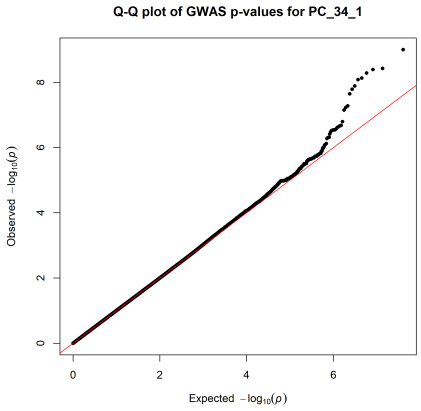


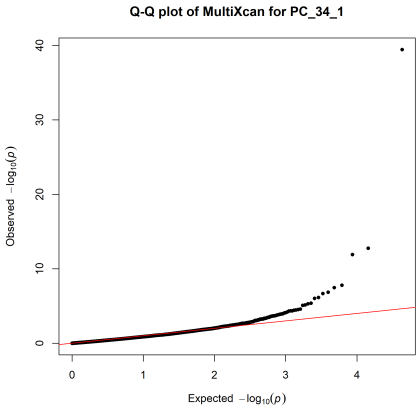
Chr11_59978355_62914375

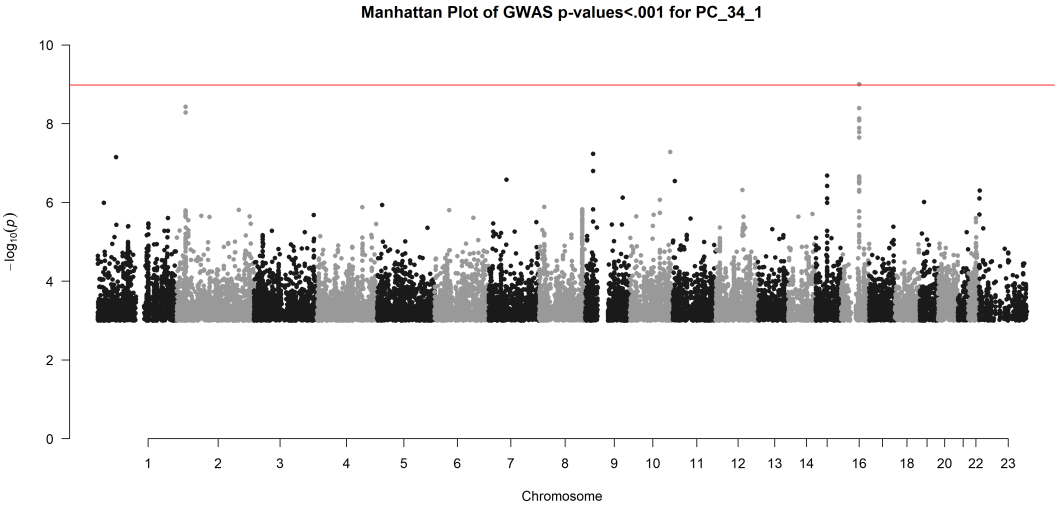


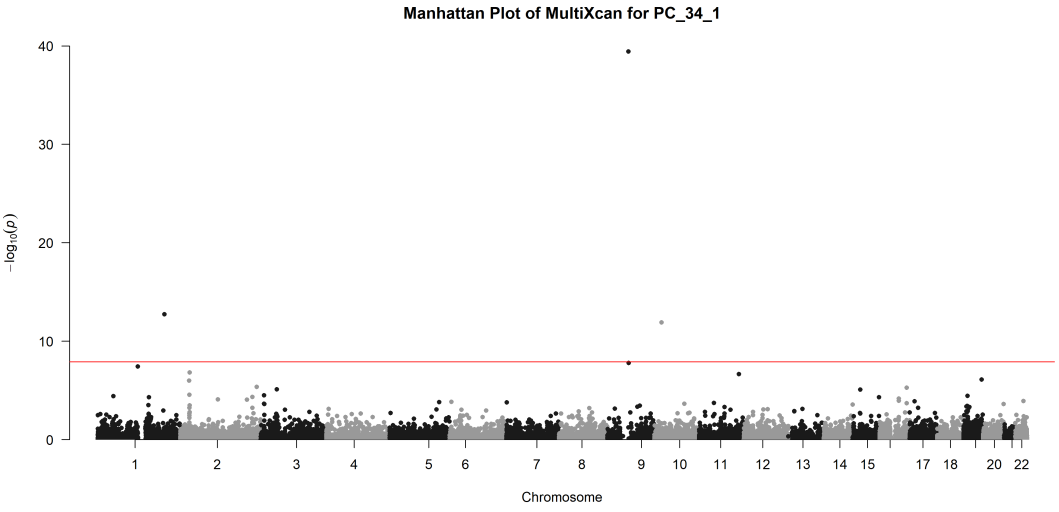
Phosphatidylcholine diacyl C34:1 (µg/dL)



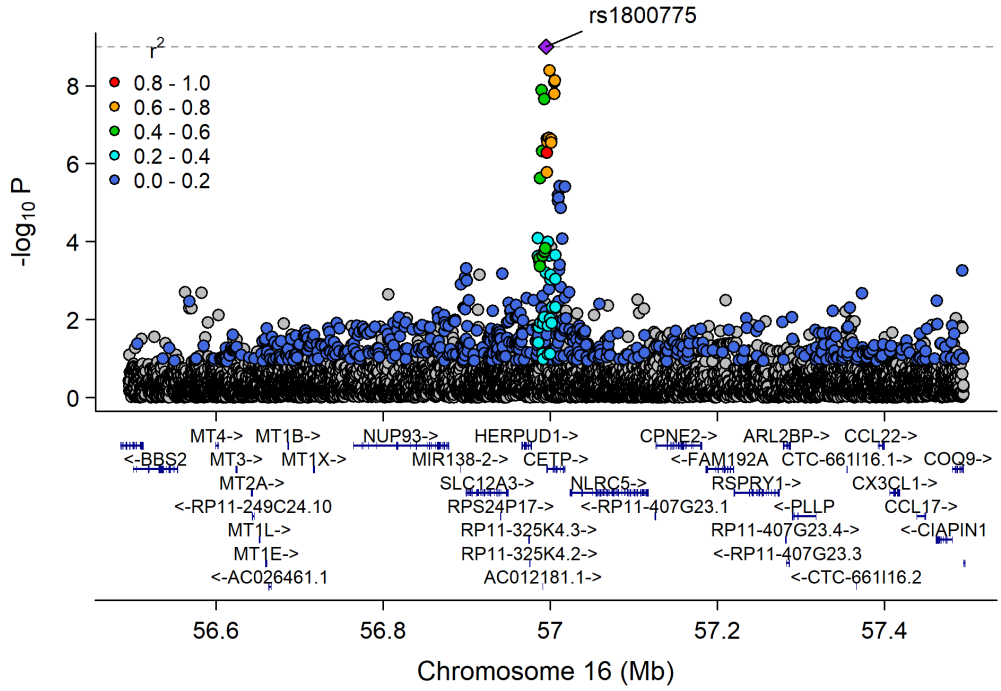




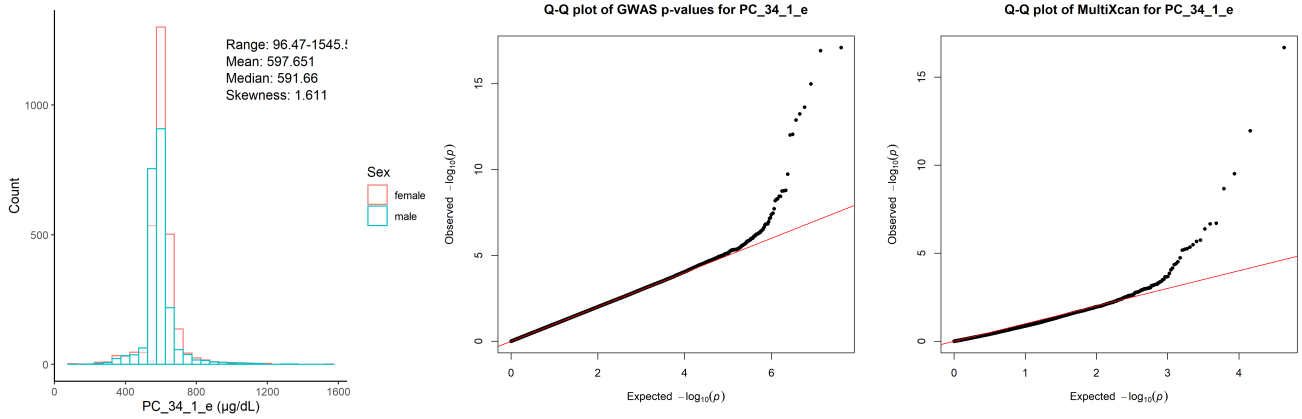


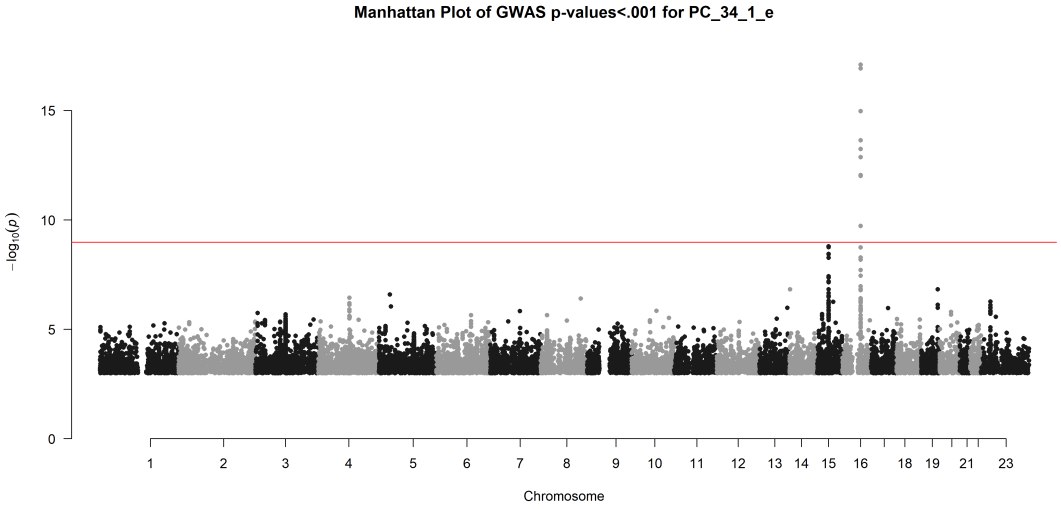


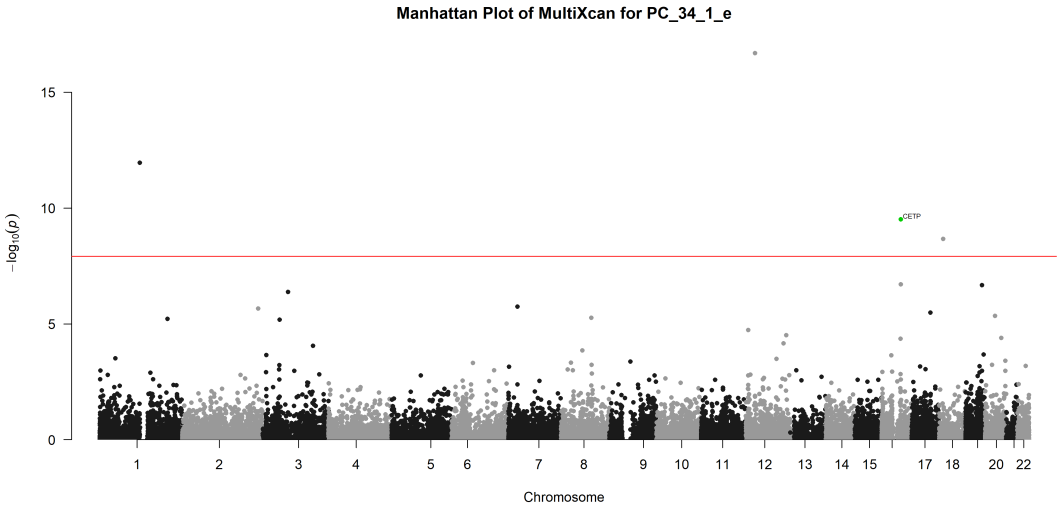
Chr16_55870822_57992421



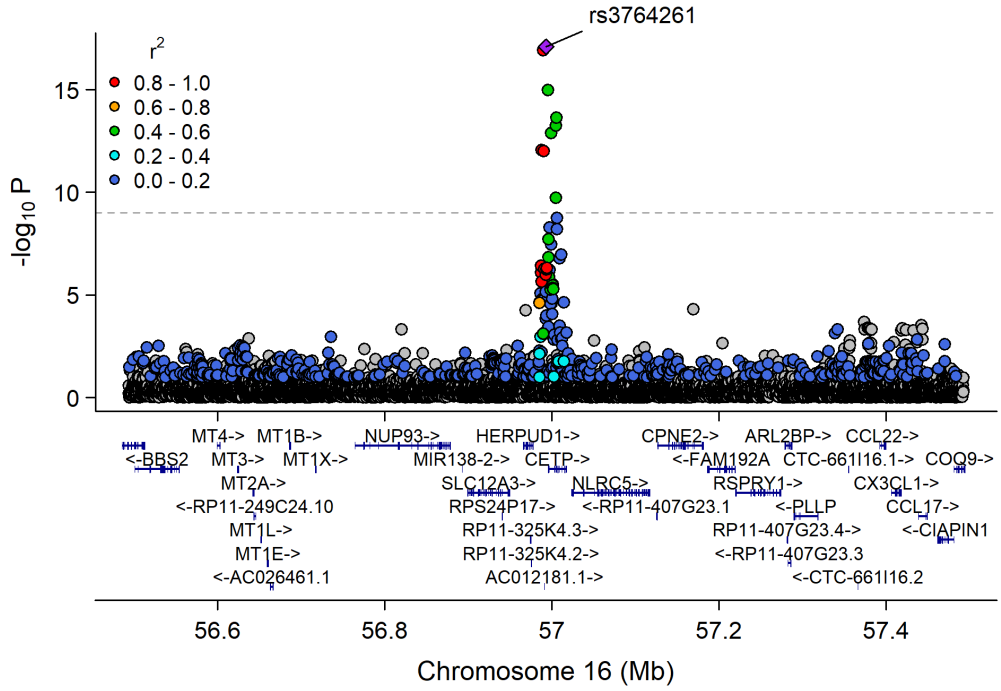
Phosphatidylcholine acyl-alkyl C34:1 (µg/dL)



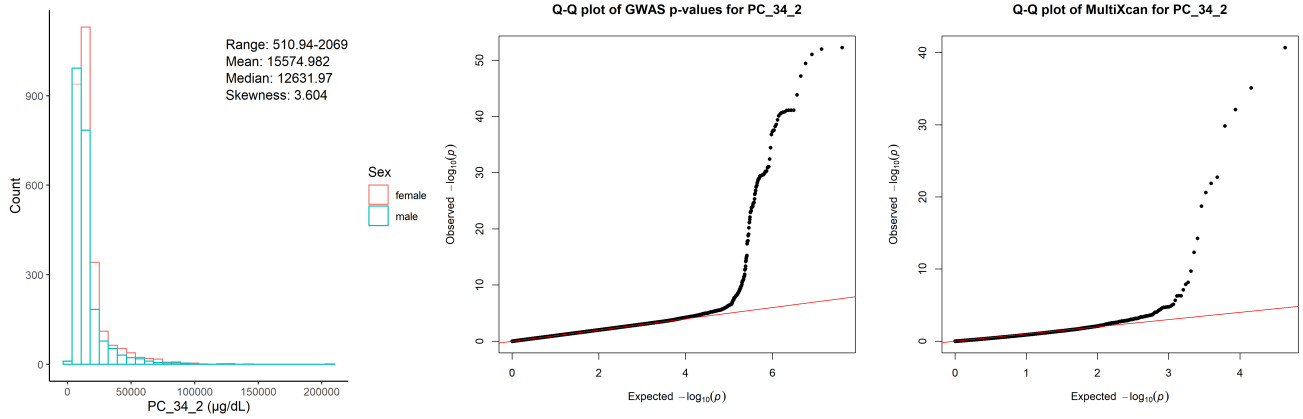


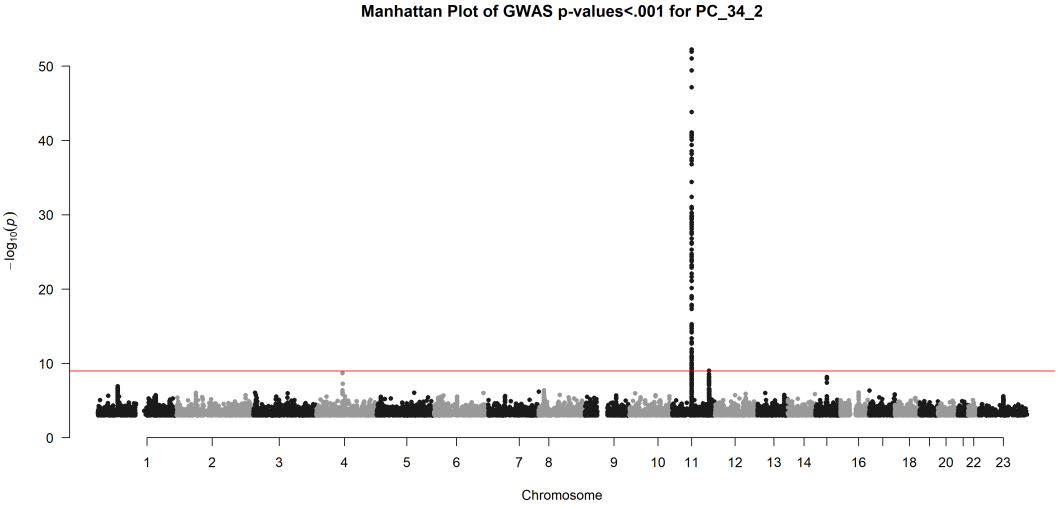


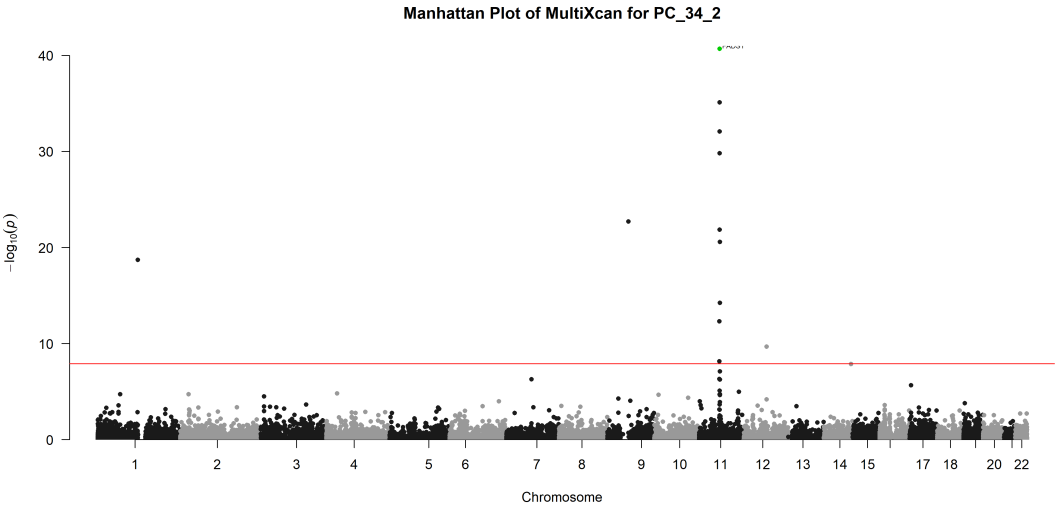
Chr16_55870822_57992421



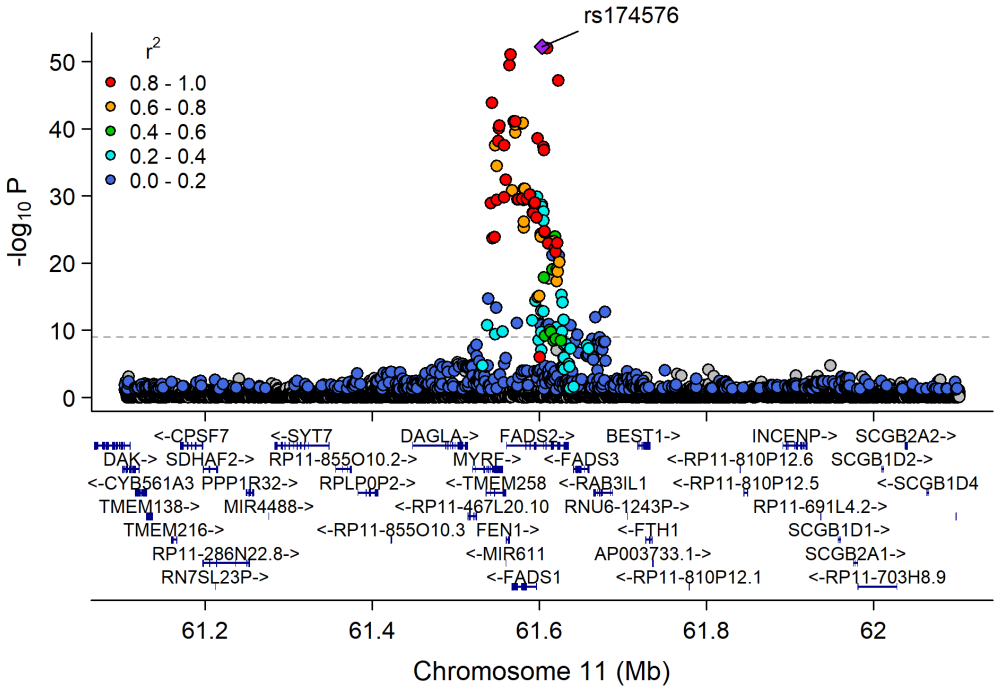
Phosphatidylcholine diacyl C34:2 (µg/dL)



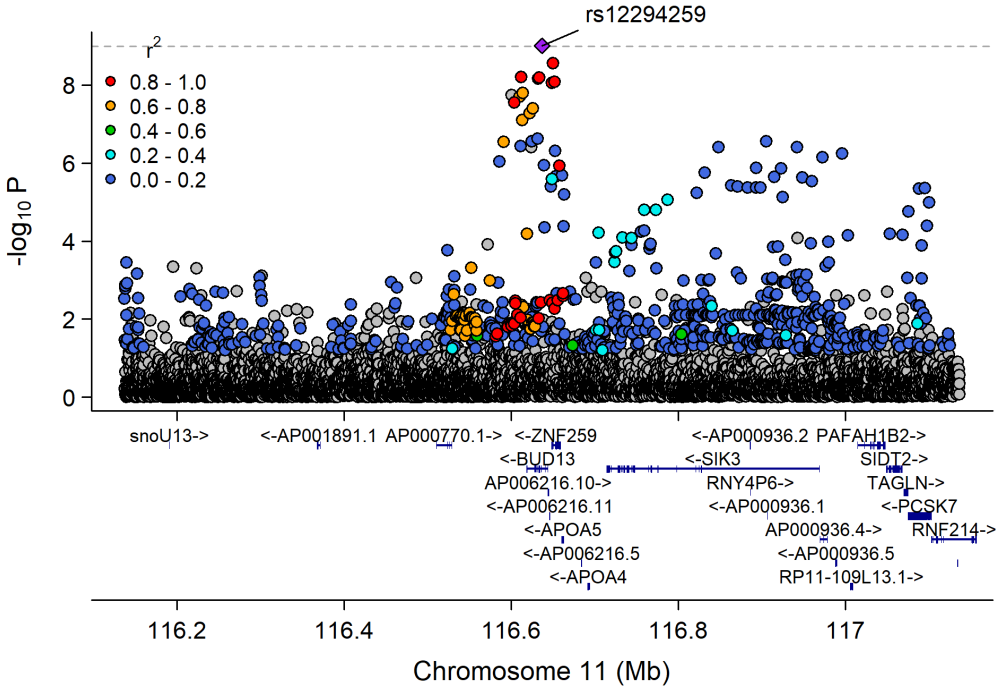




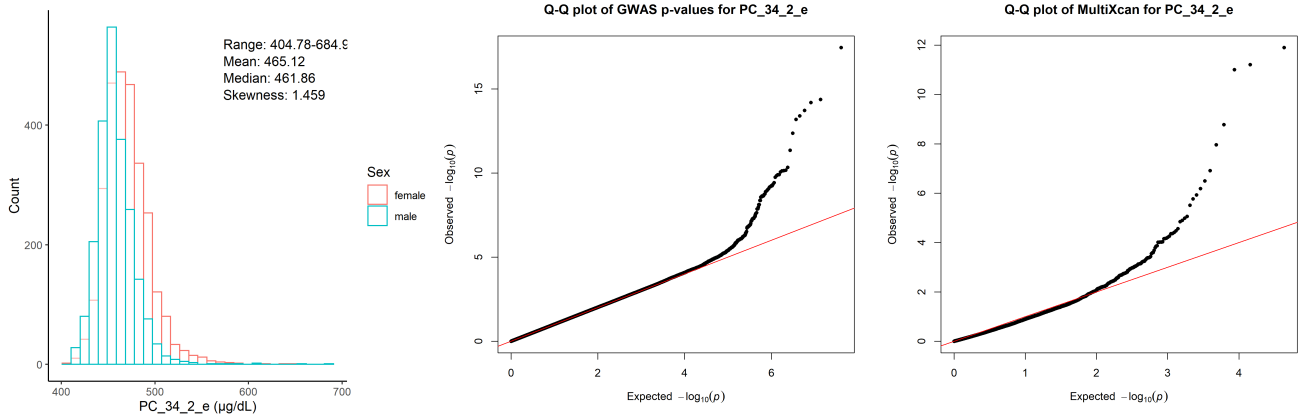
Chr11_59978355_62914375

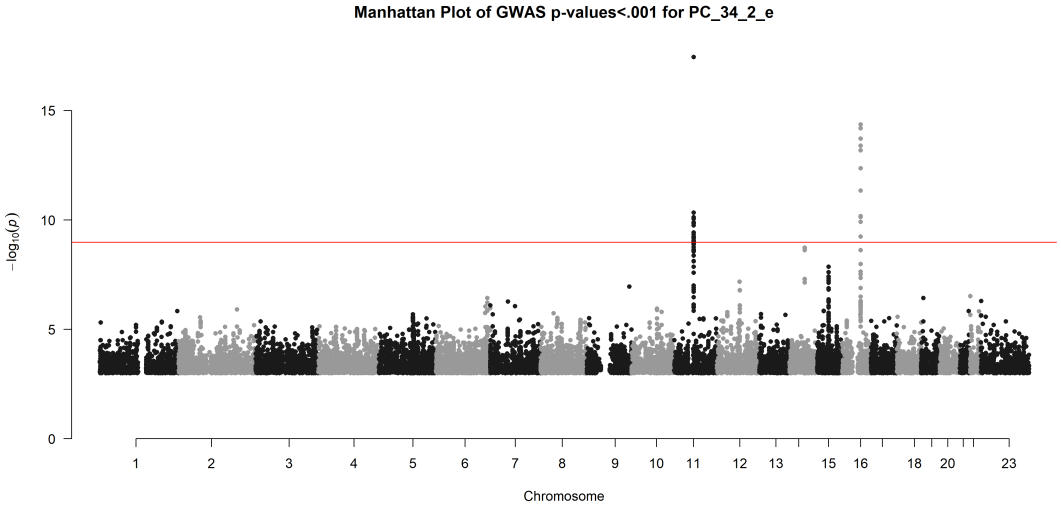


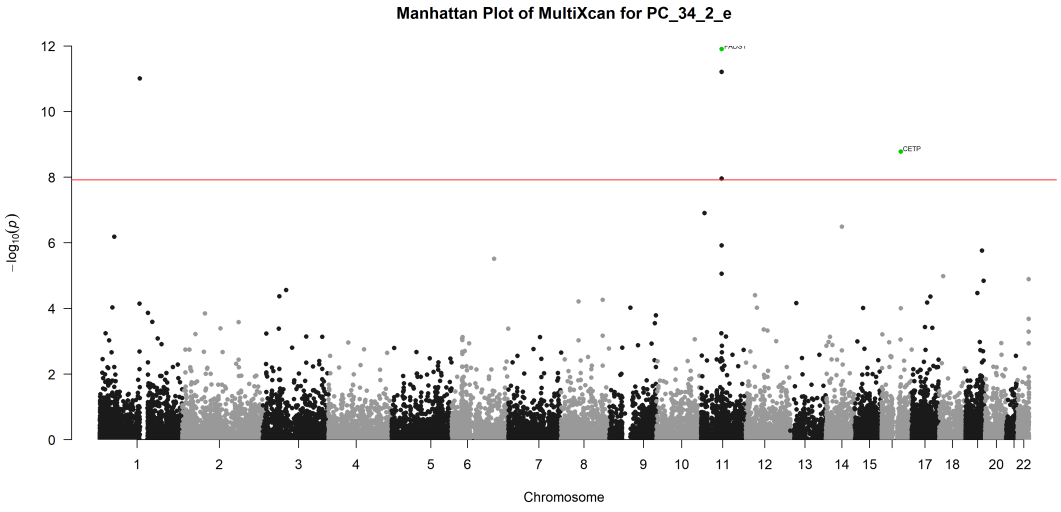
Chr11_115541901_117633315



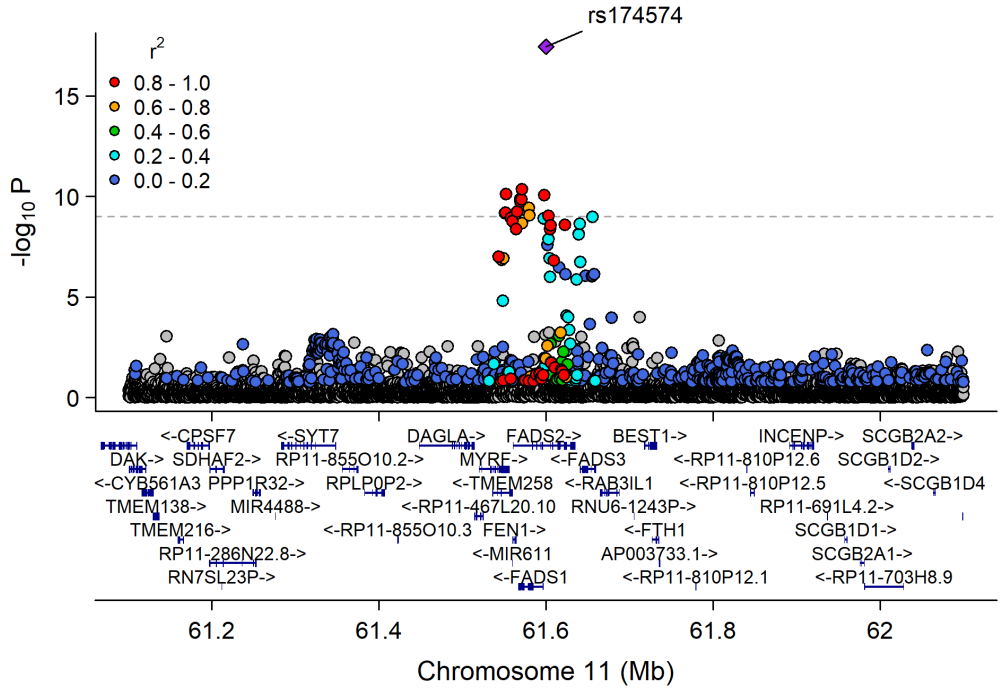
Phosphatidylcholine acyl-alkyl C34:2 (µg/dL)



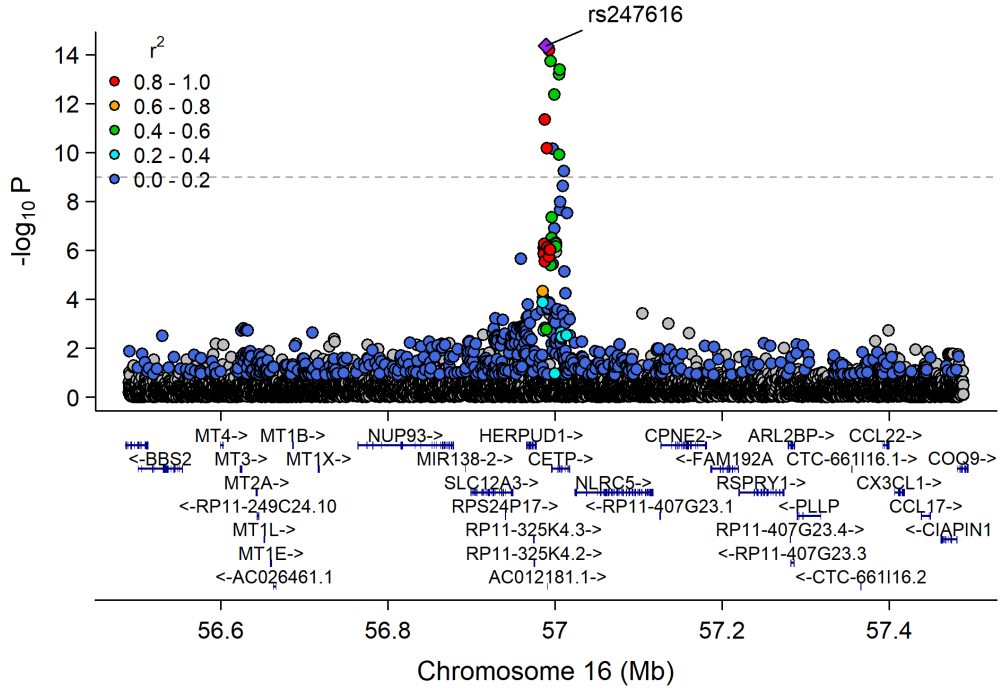




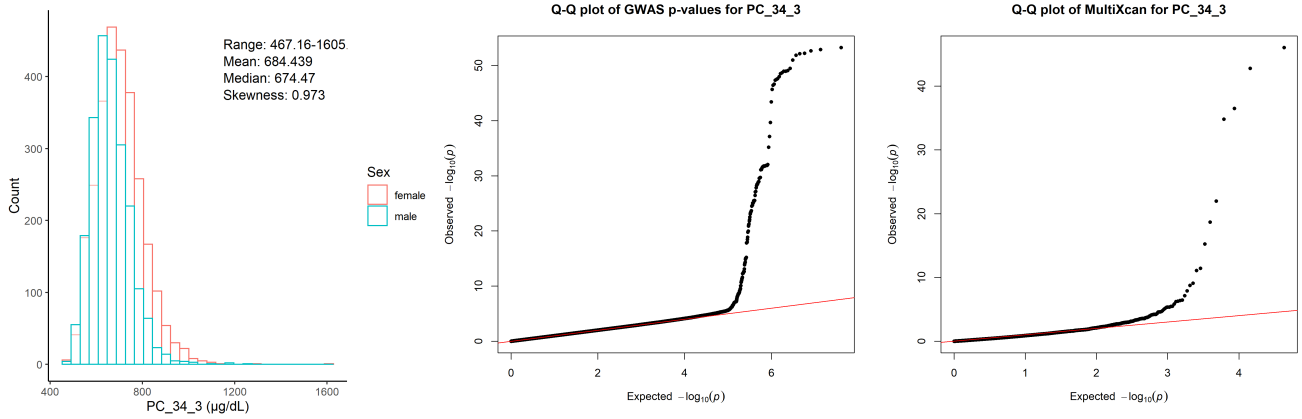
Chr11_59978355_62914375

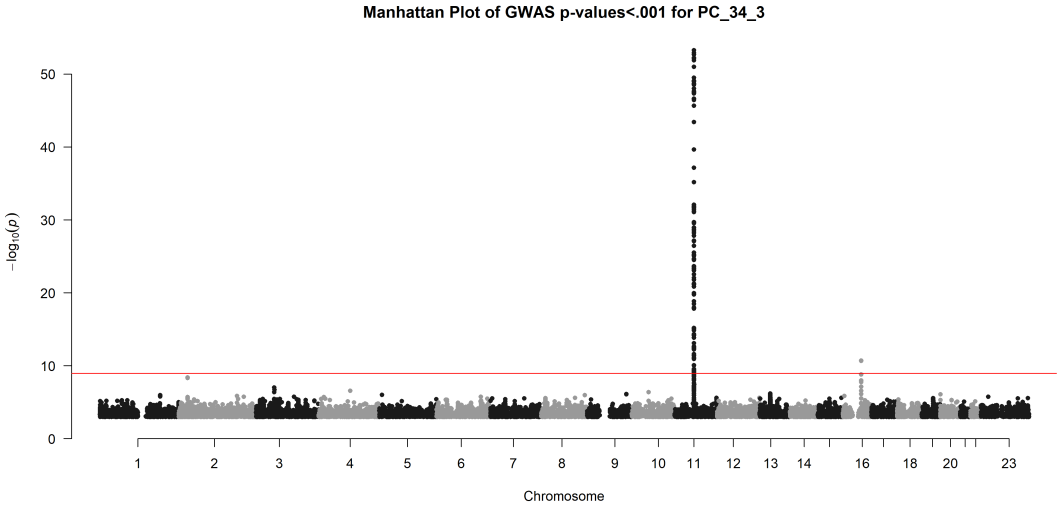


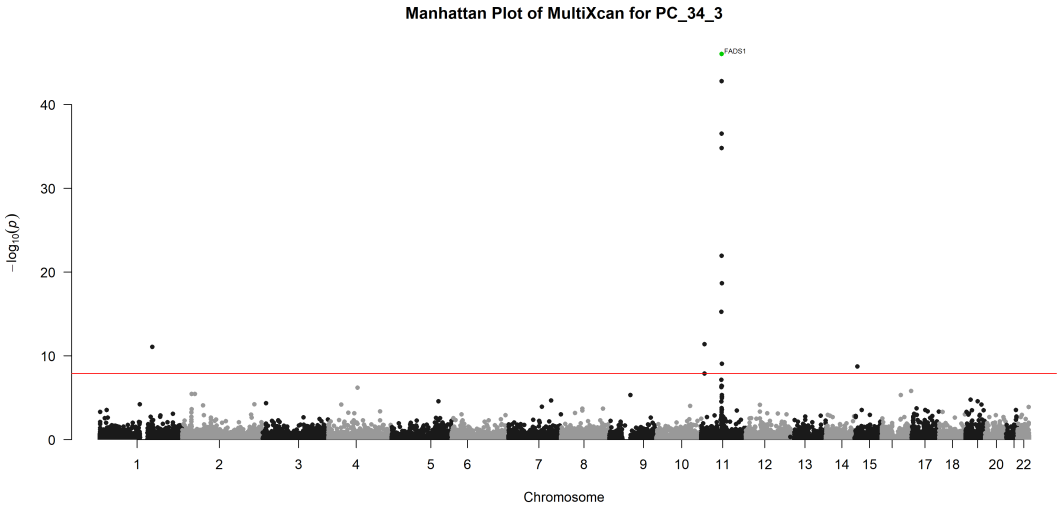
Chr16_55870822_57992421



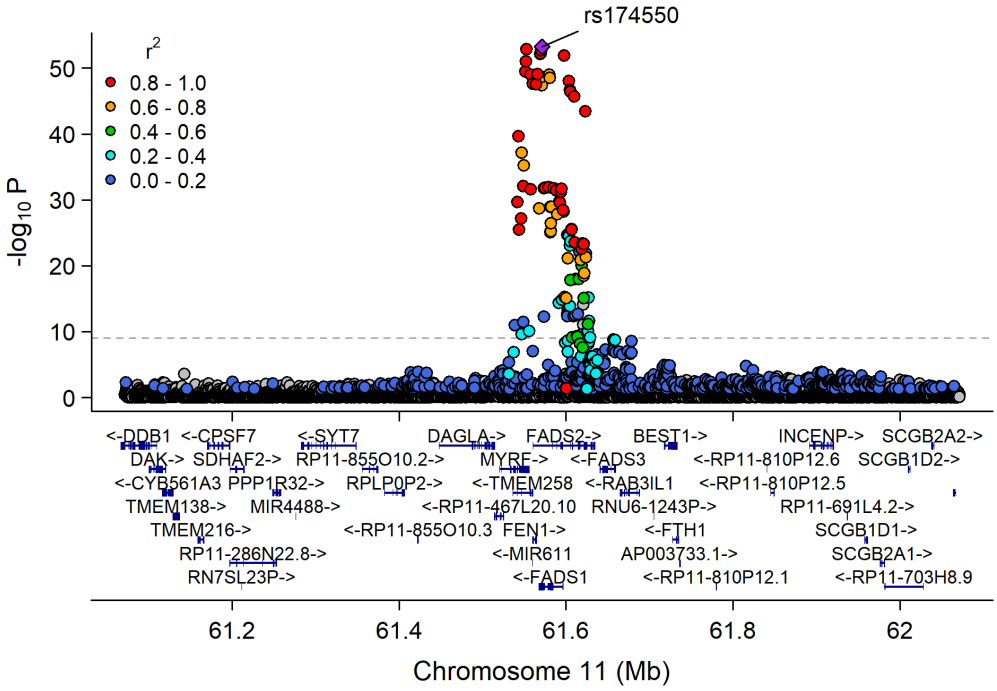
Phosphatidylcholine diacyl C34:3 (µg/dL)



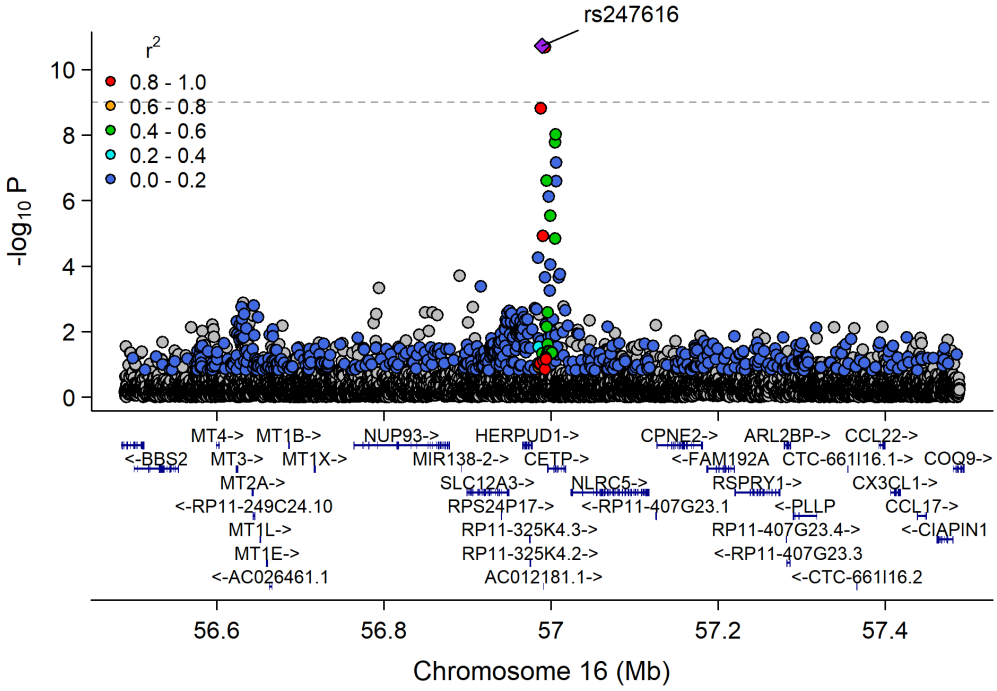




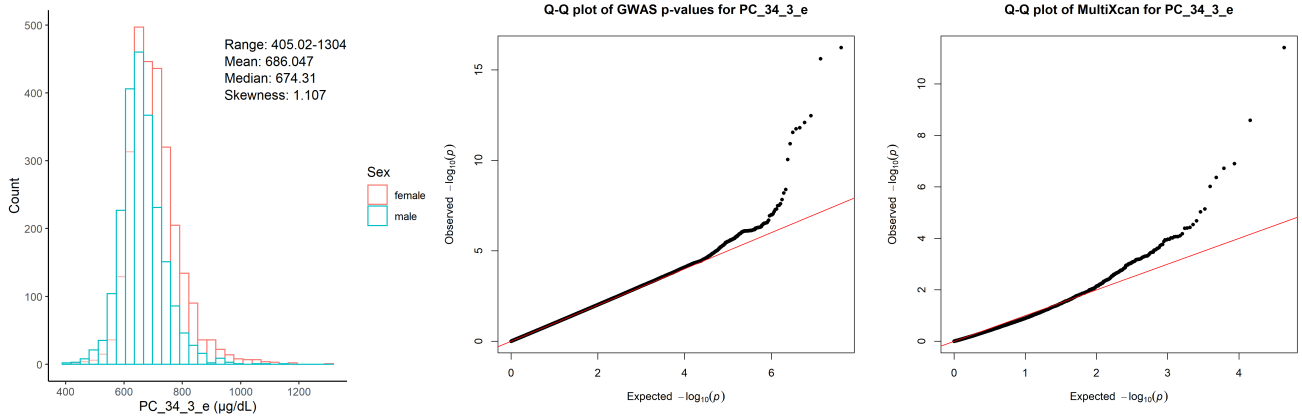
Chr11_59978355_62914375

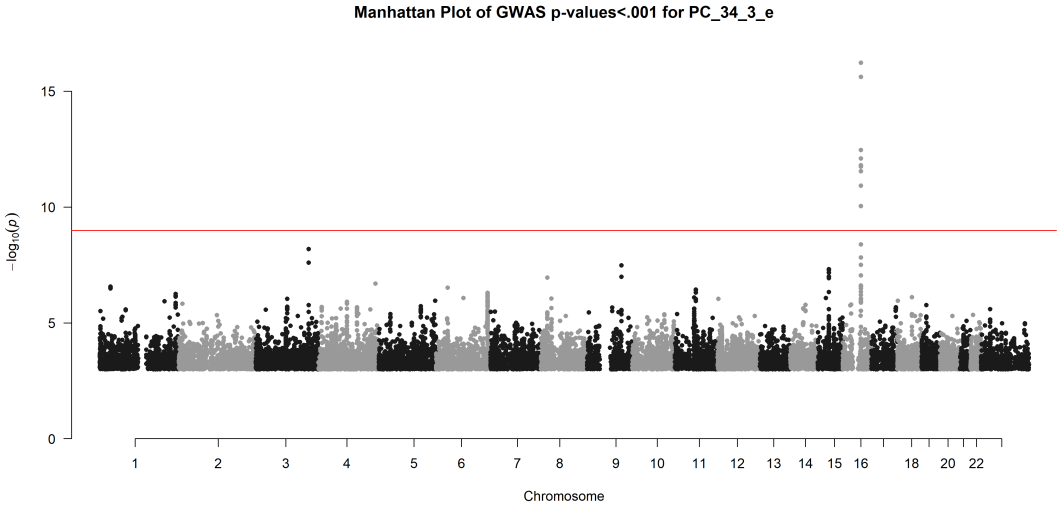


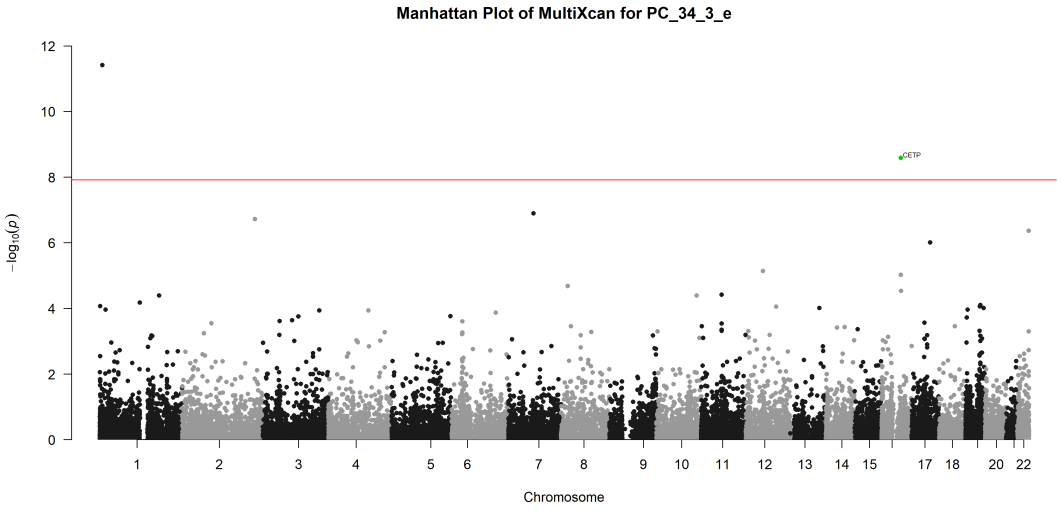
Chr16_55870822_57992421



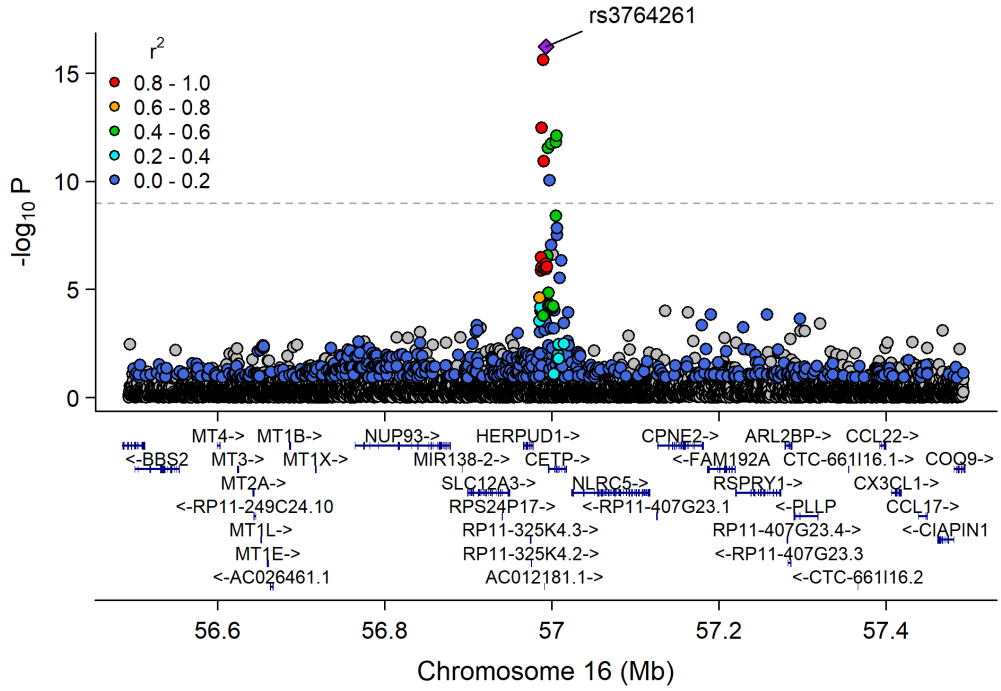
Phosphatidylcholine acyl-alkyl C34:3 (µg/dL)



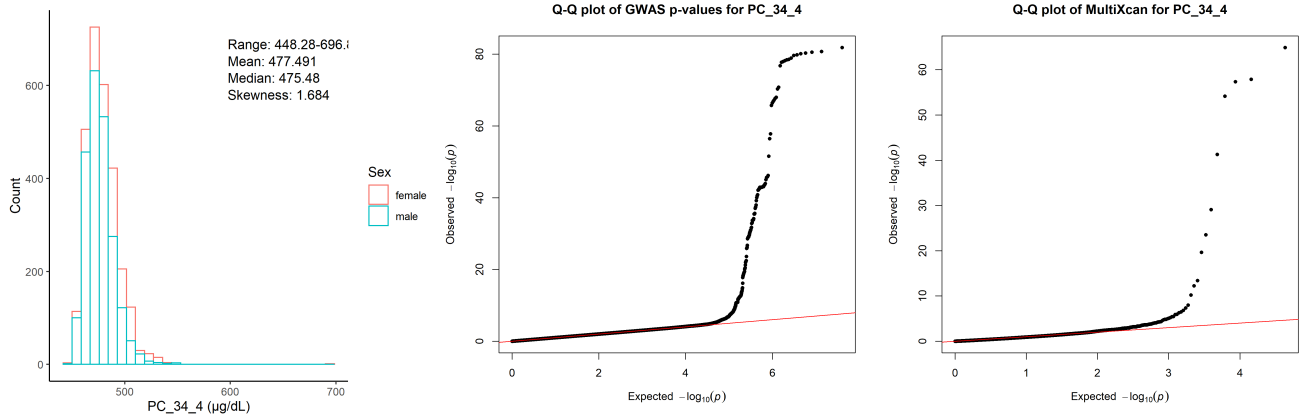


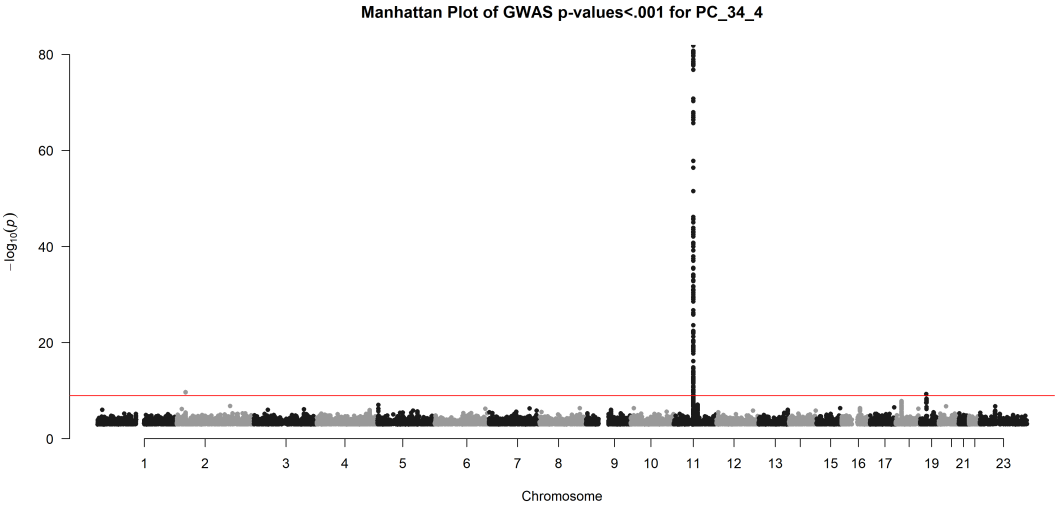


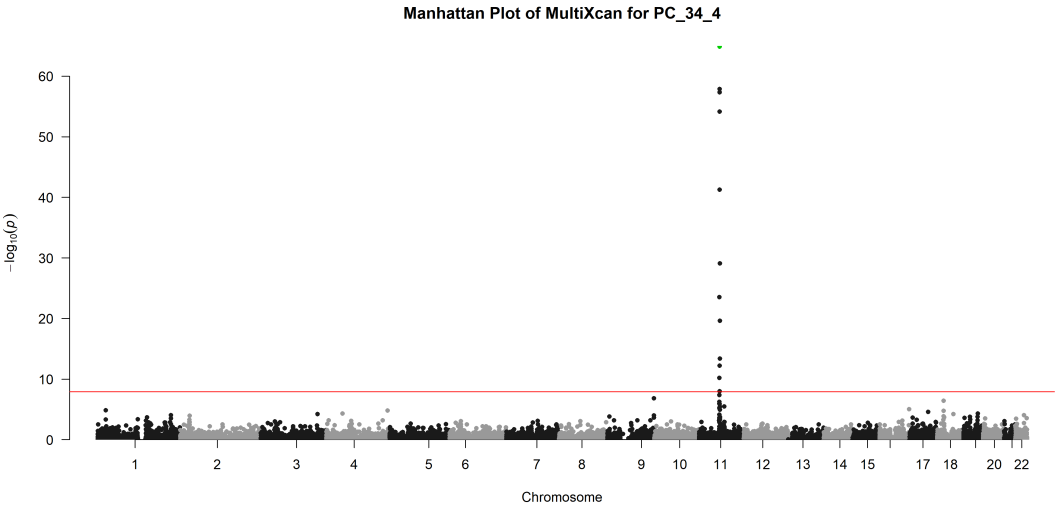
Chr16_55870822_57992421



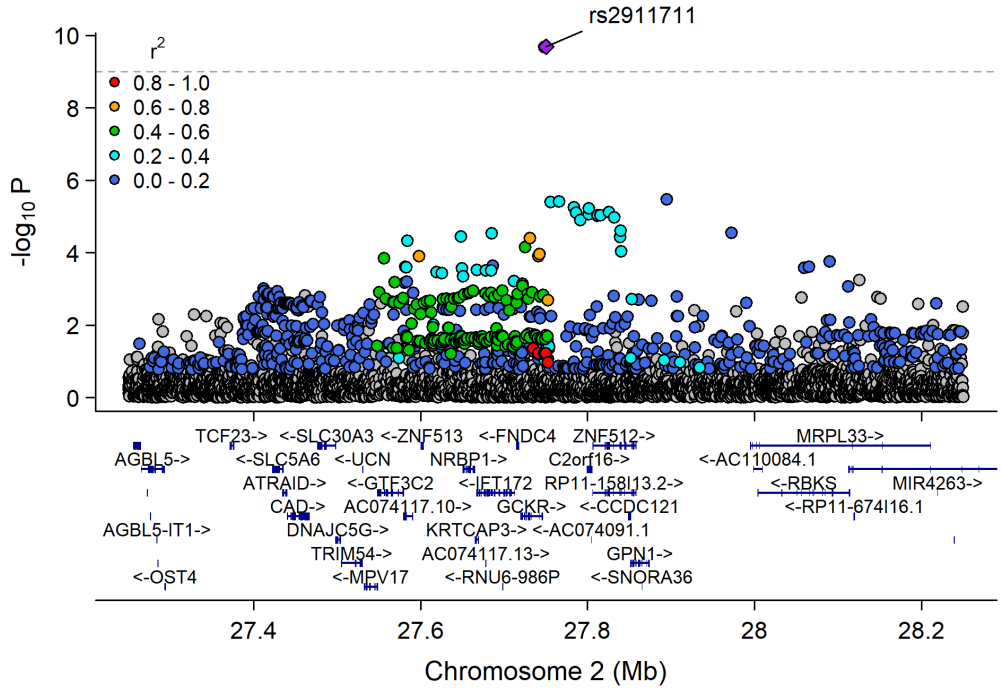
Phosphatidylcholine diacyl C34:4 (µg/dL)



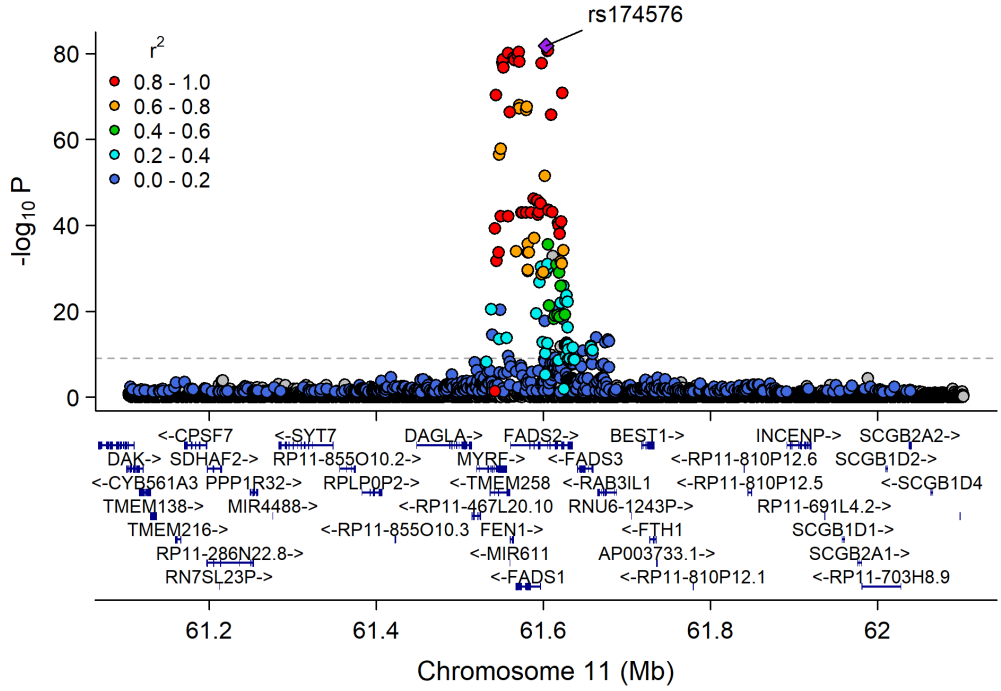




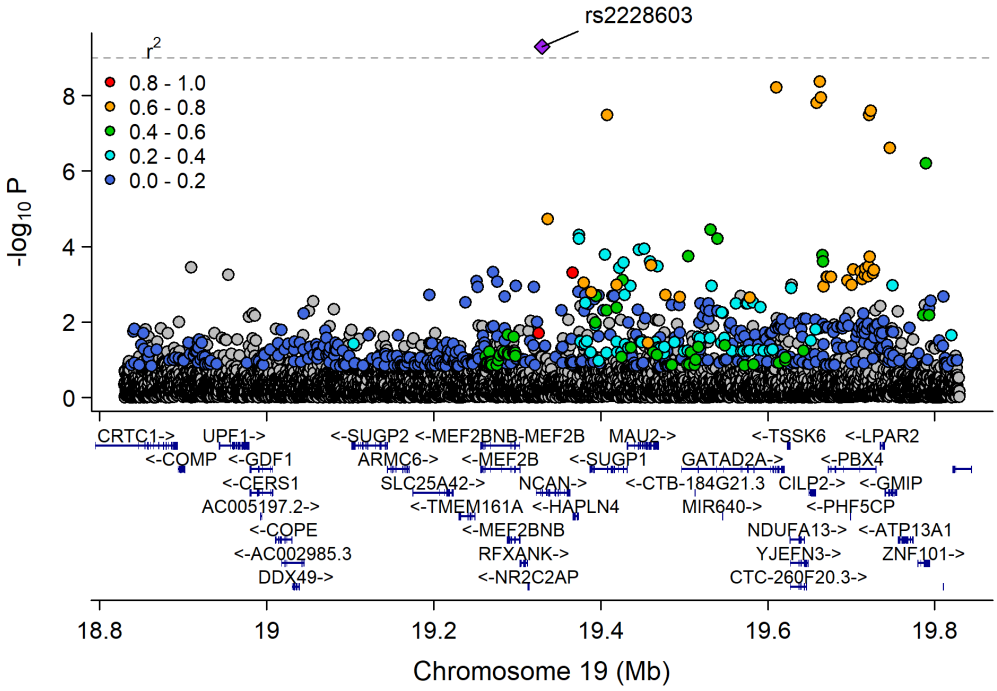
Chr2_26753815_28597624



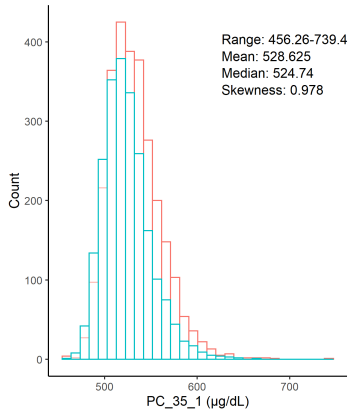
Chr11_59978355_62914375

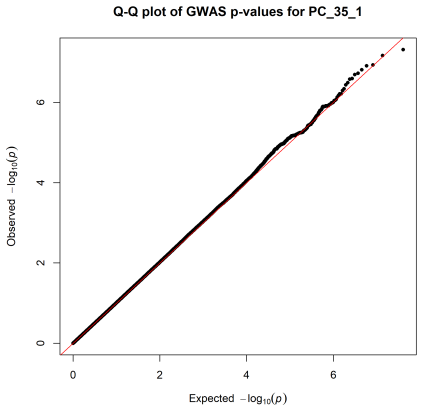


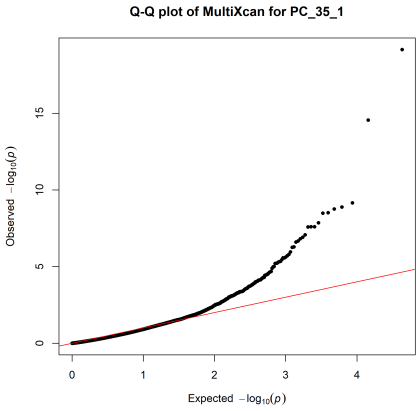
Chr19_18370495_20841464

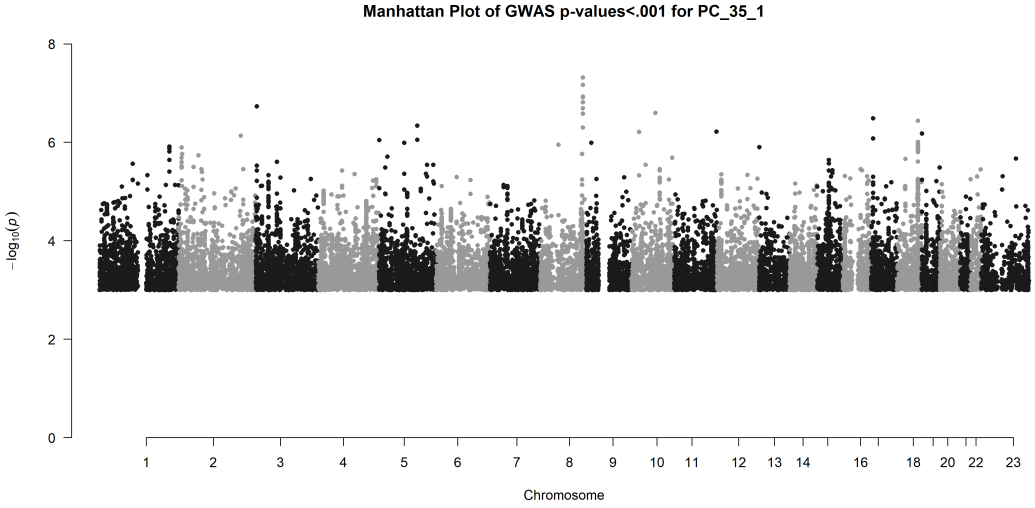


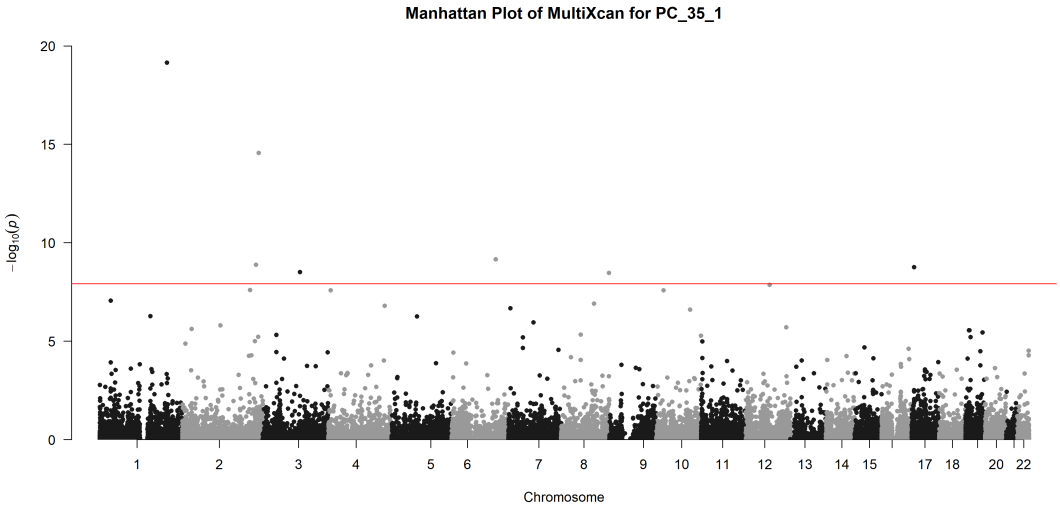
Phosphatidylcholine diacyl C35:1 (µg/dL)



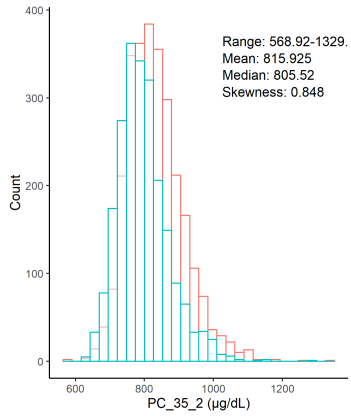


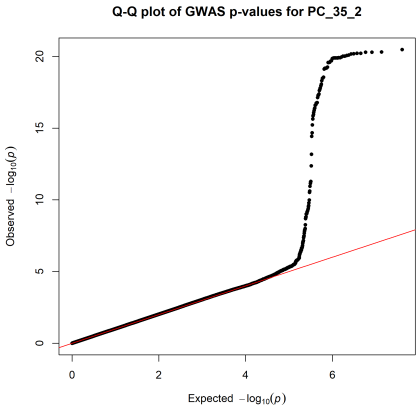


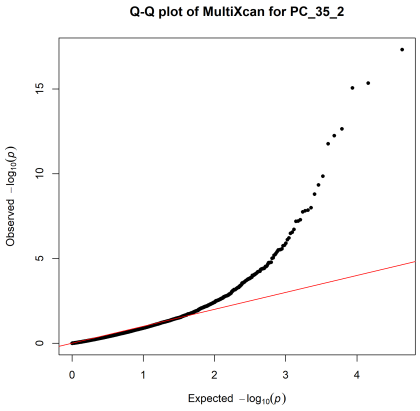


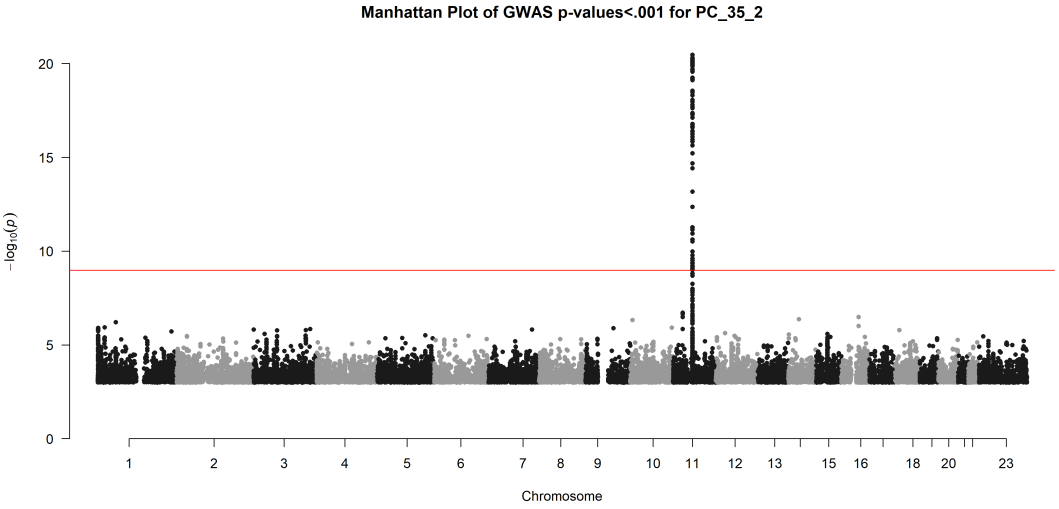


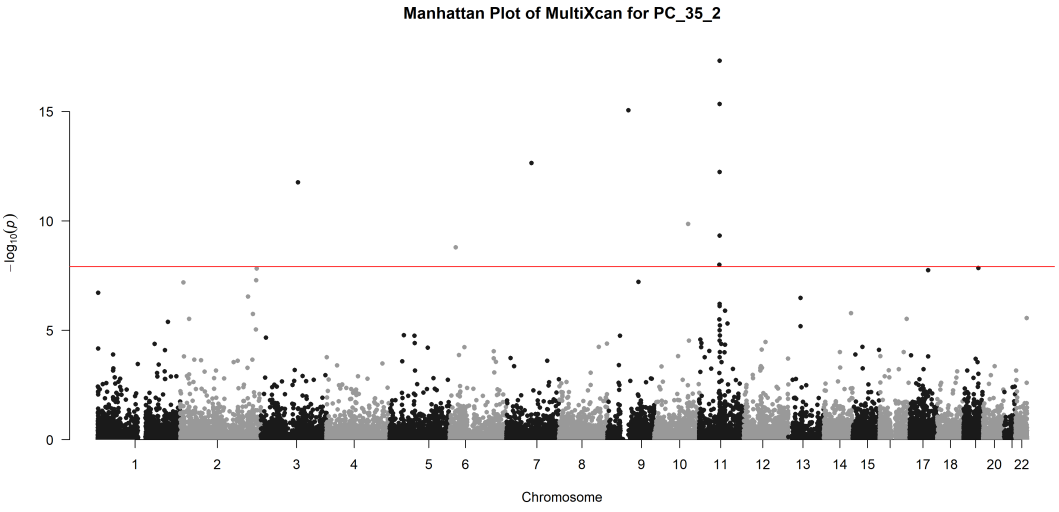
Phosphatidylcholine diacyl C35:2 (µg/dL)



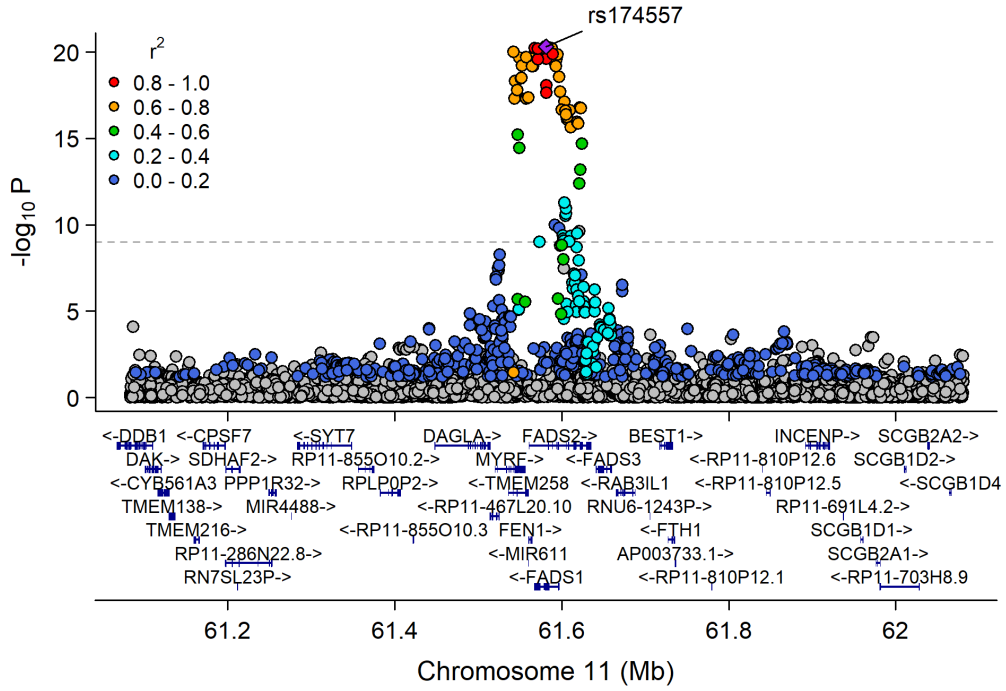




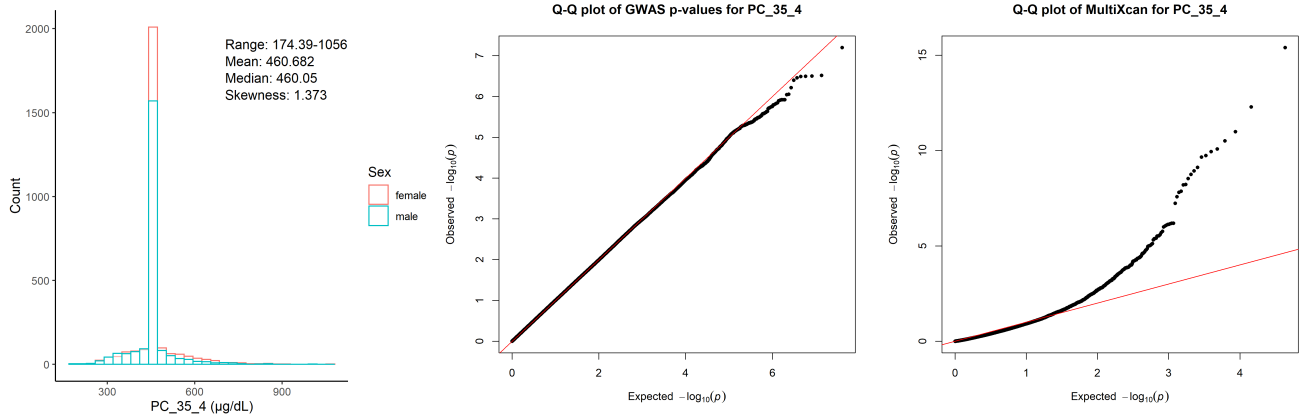


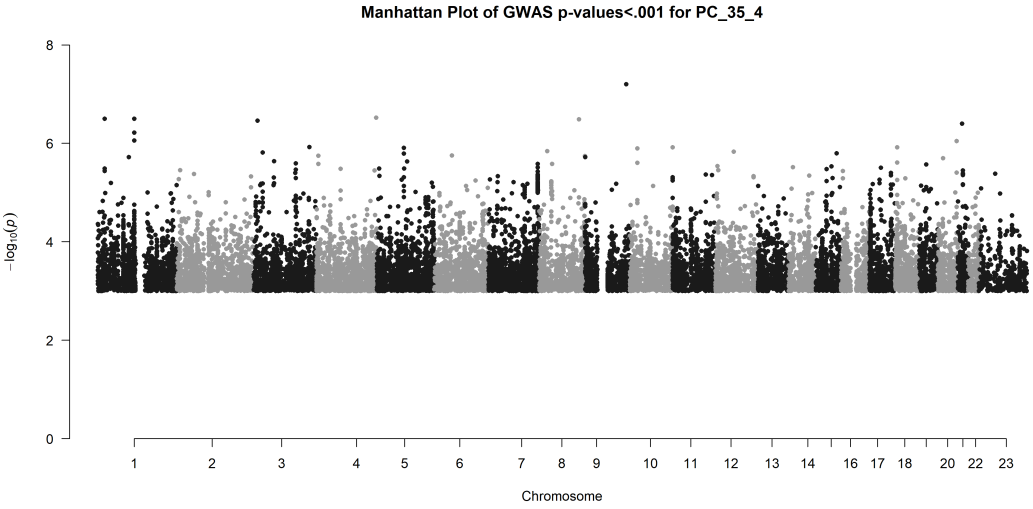


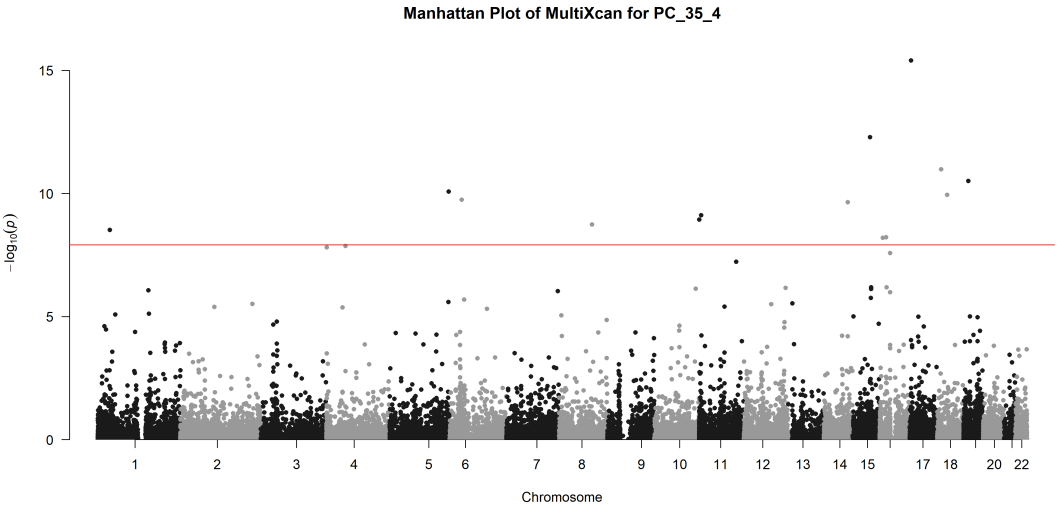
Chr11_59978355_62914375



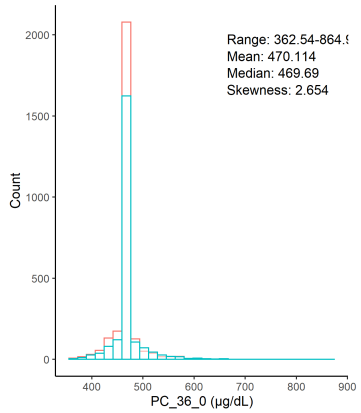
Phosphatidylcholine diacyl C35:4 (µg/dL)

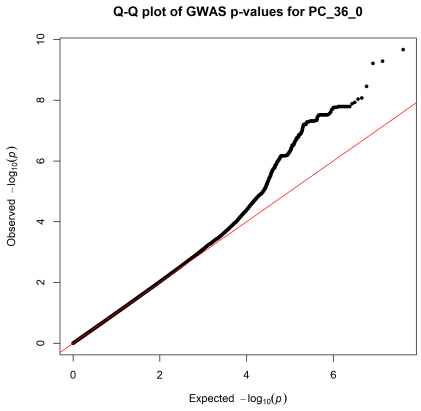


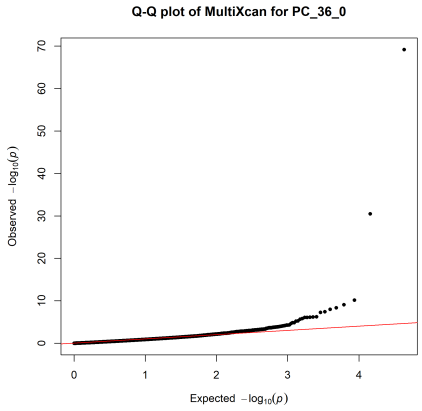


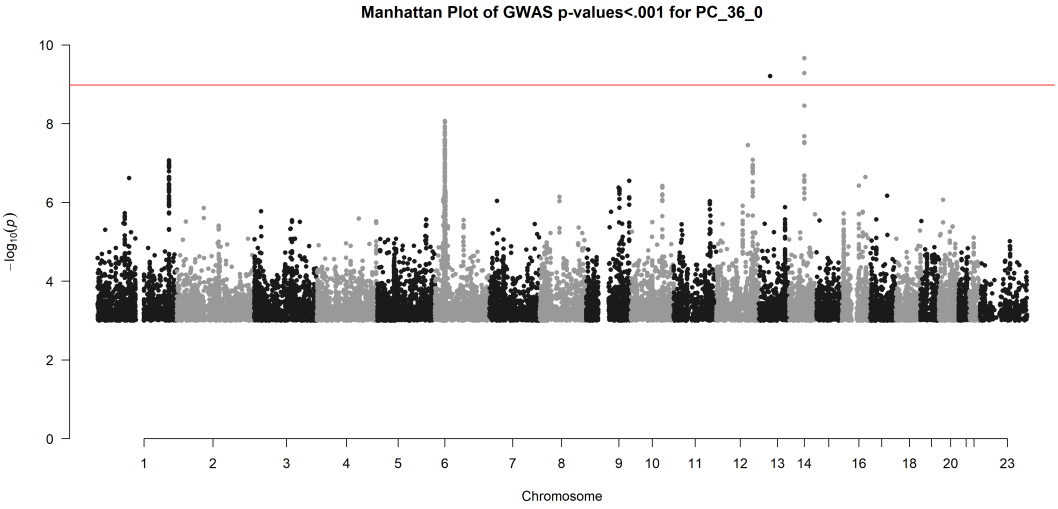


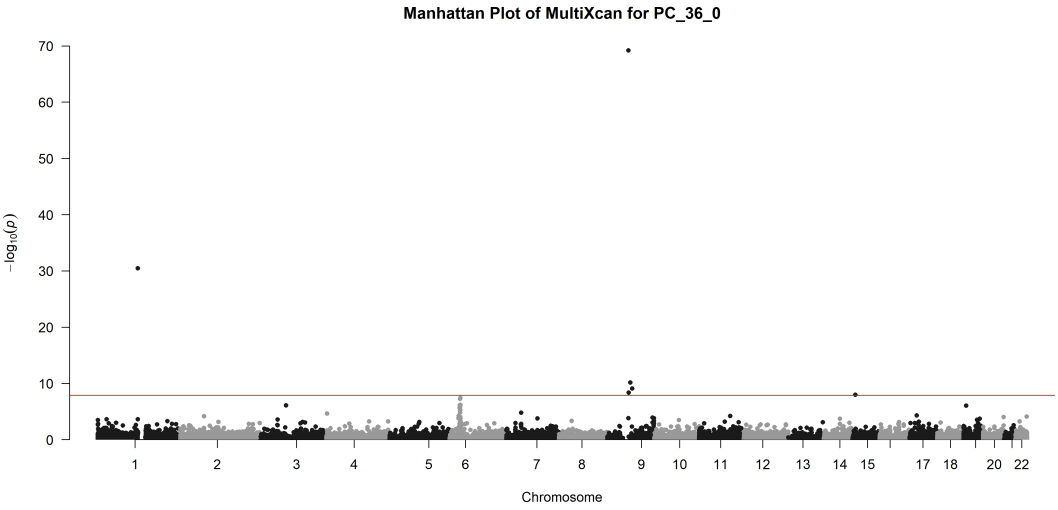
Phosphatidylcholine diacyl C36:0 (µg/dL)



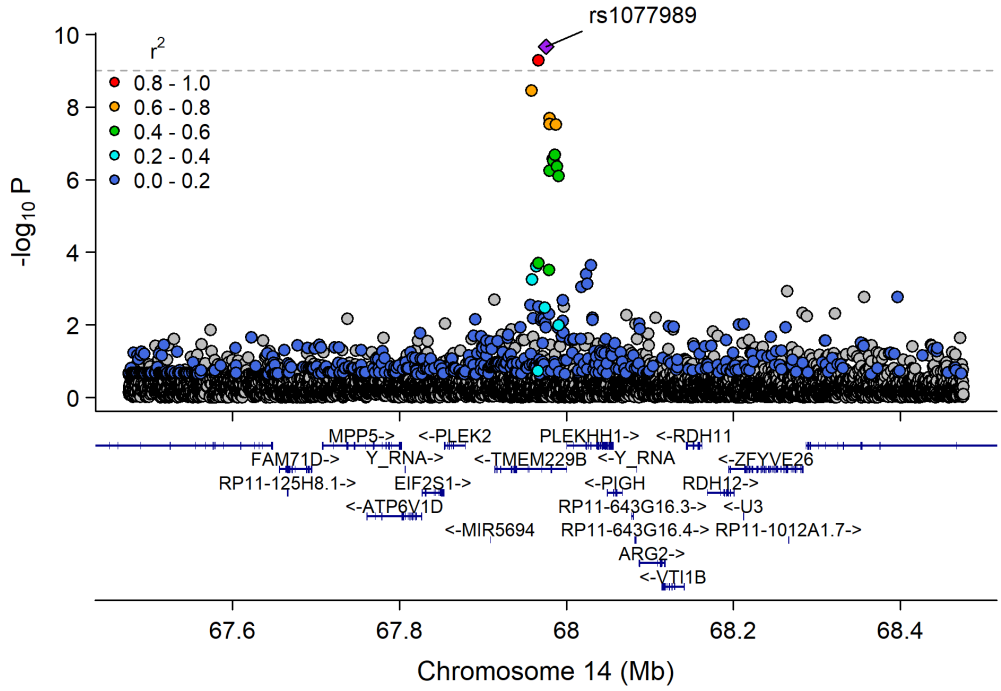




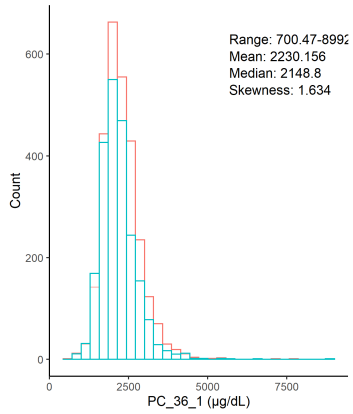


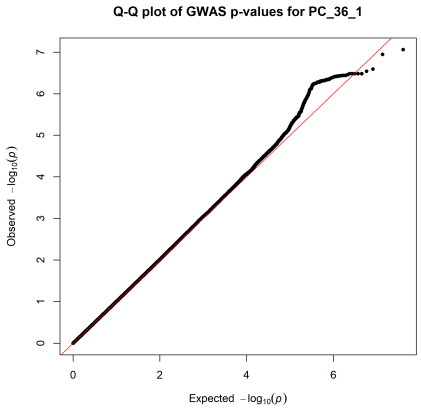


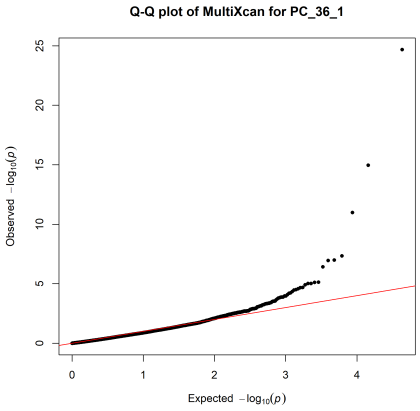
Chr14__66997265__68972354

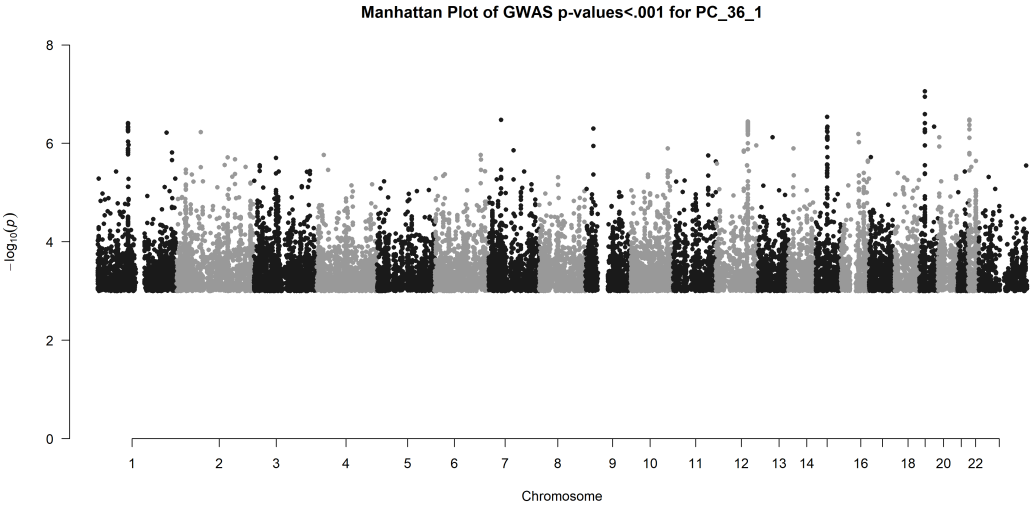


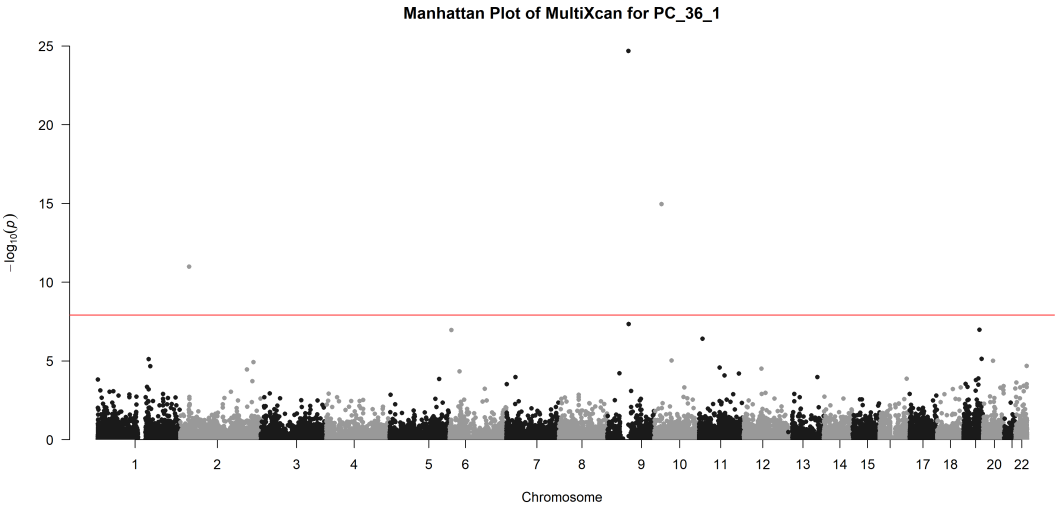
Phosphatidylcholine diacyl C36:1 (µg/dL)



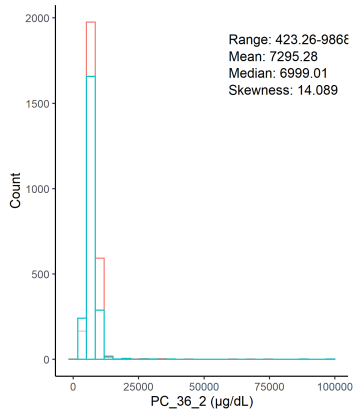


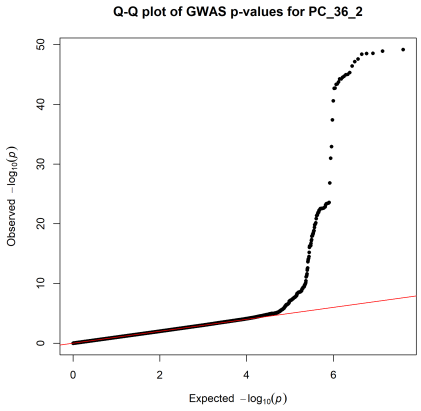


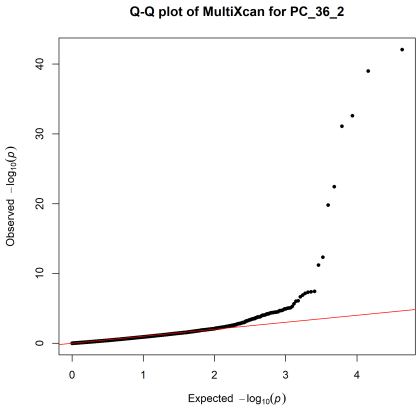


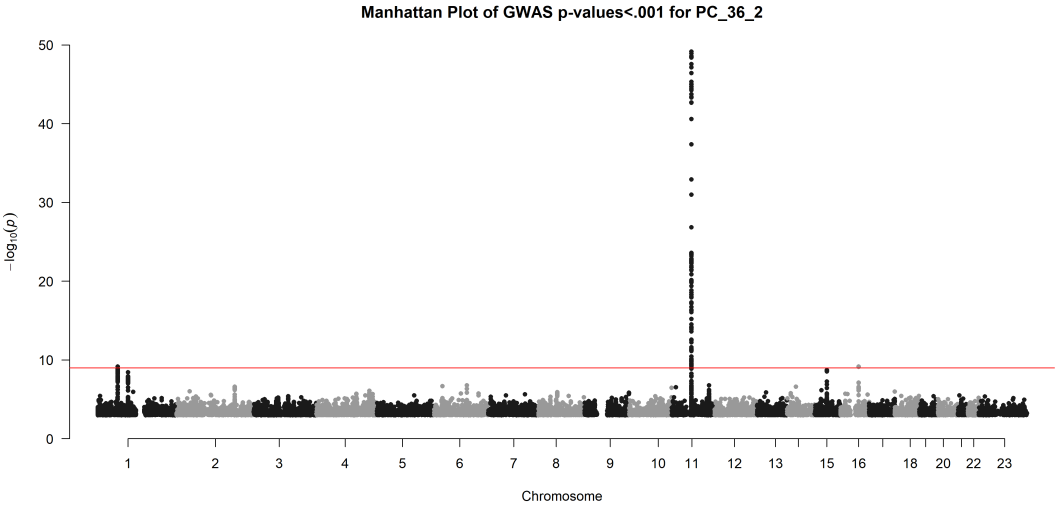


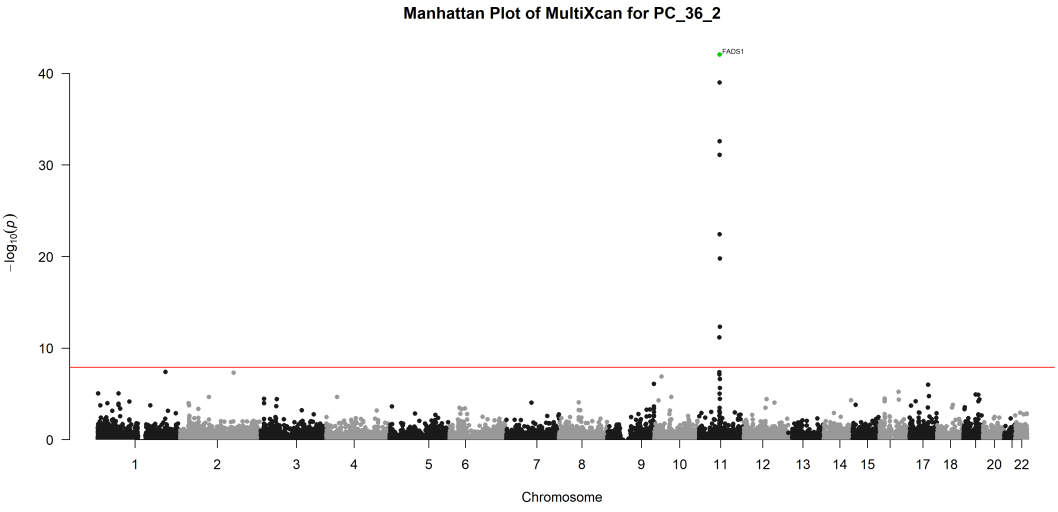
Phosphatidylcholine diacyl C36:2 (µg/dL)



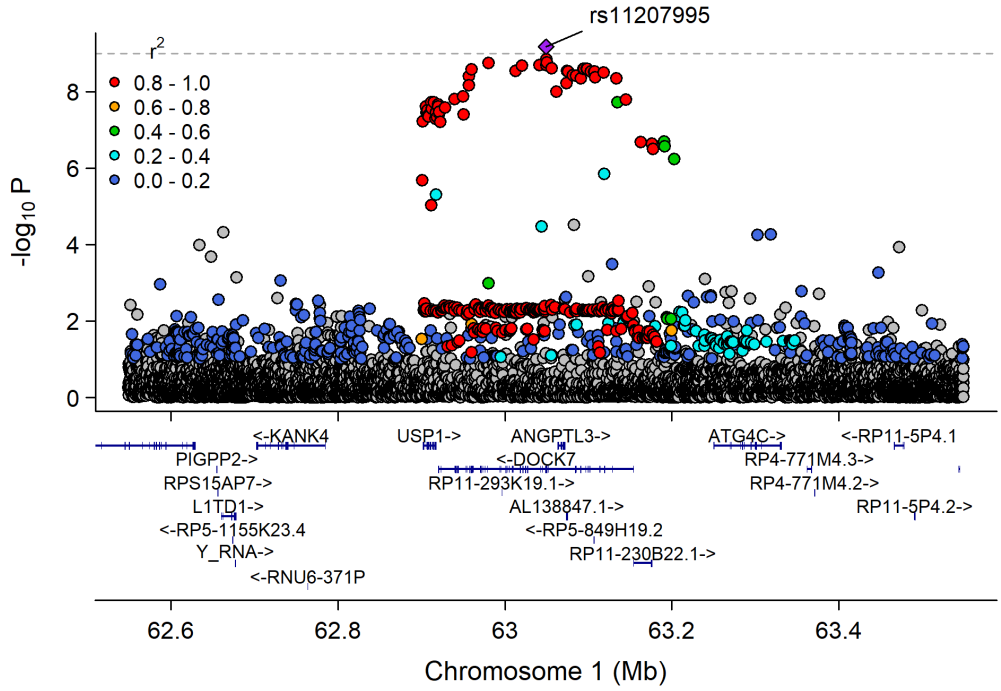




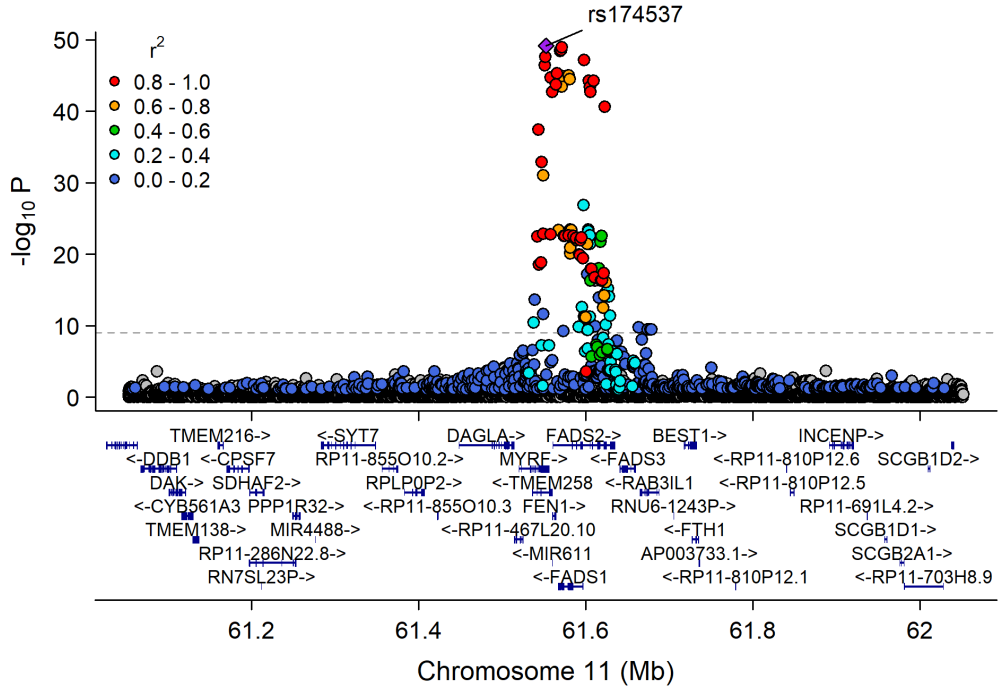




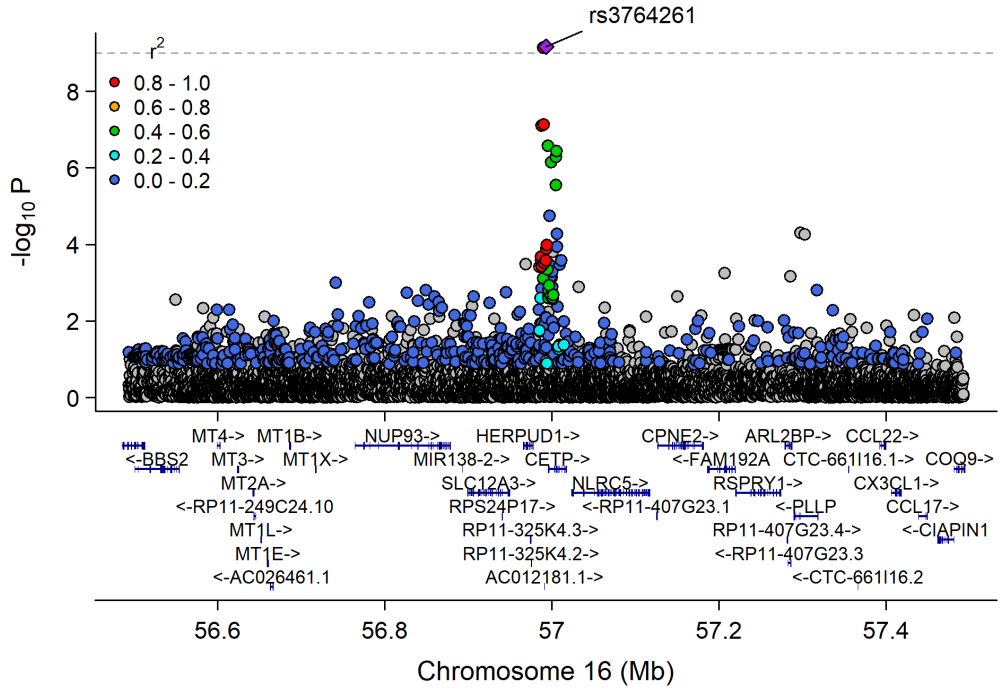
Chr1_62833402_63375116



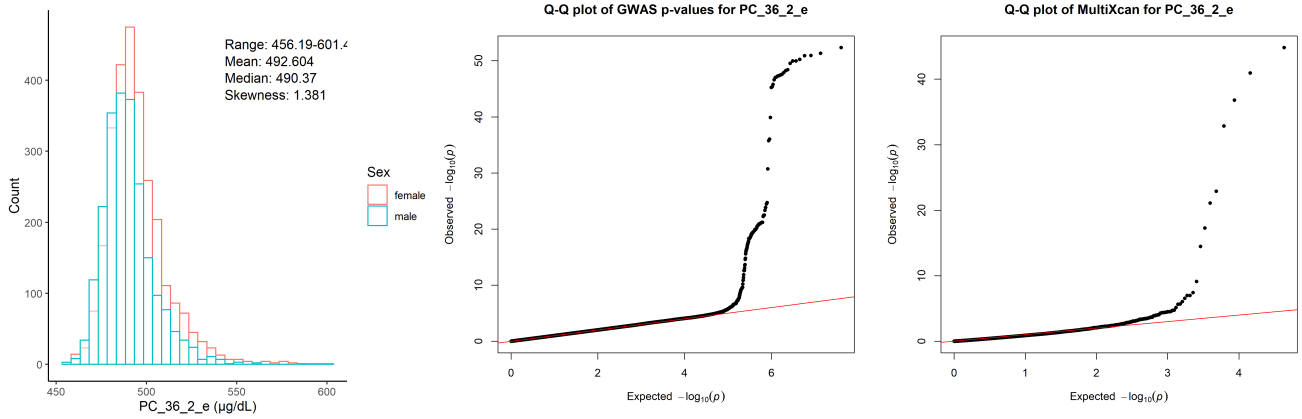
Chr11_59978355_62914375

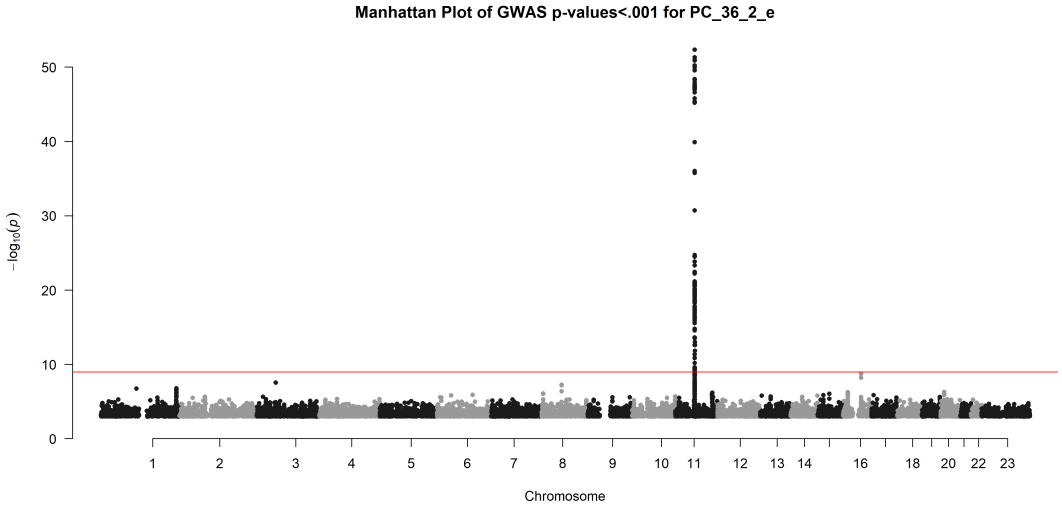


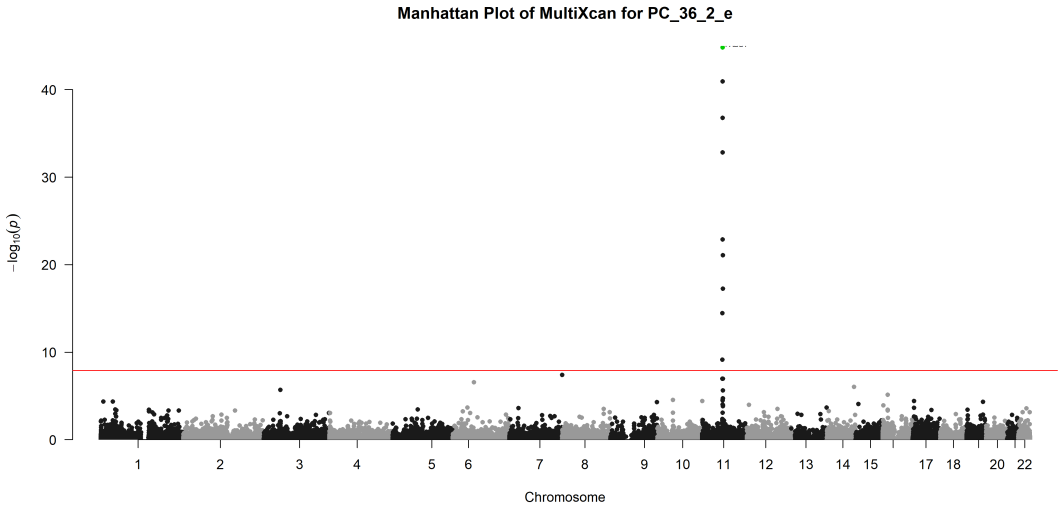
Chr16__55870822__57992421



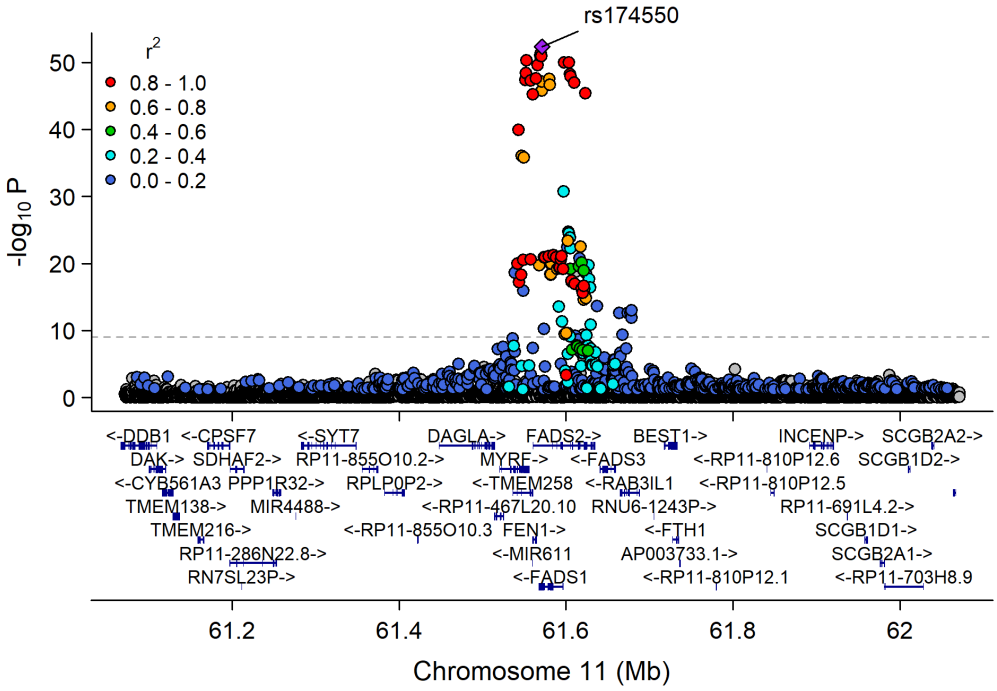
Phosphatidylcholine acyl-alkyl C36:2 (µg/dL)



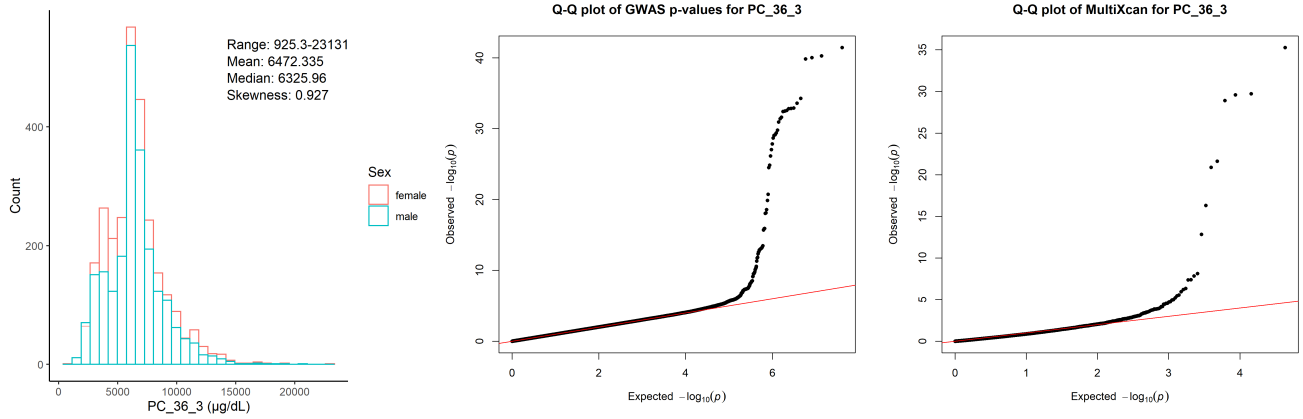


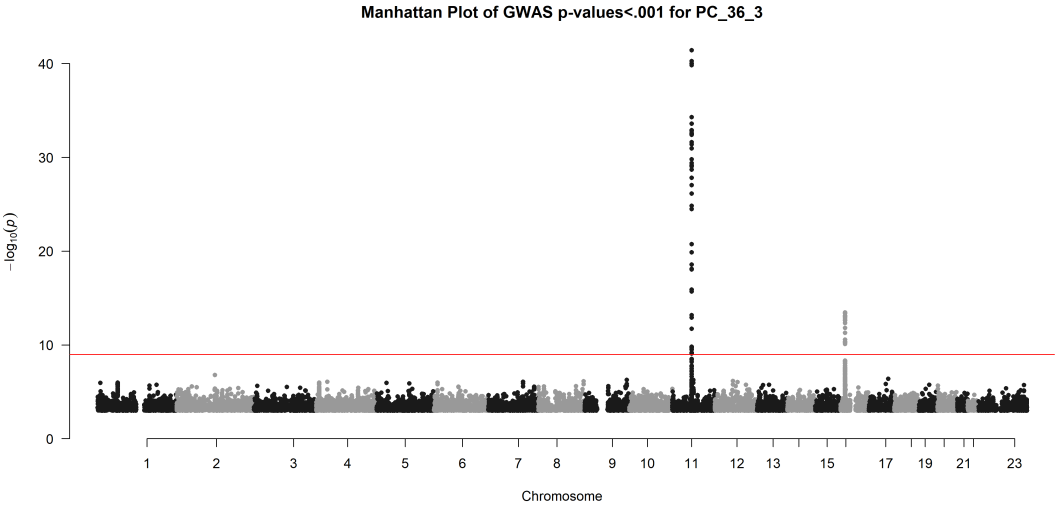


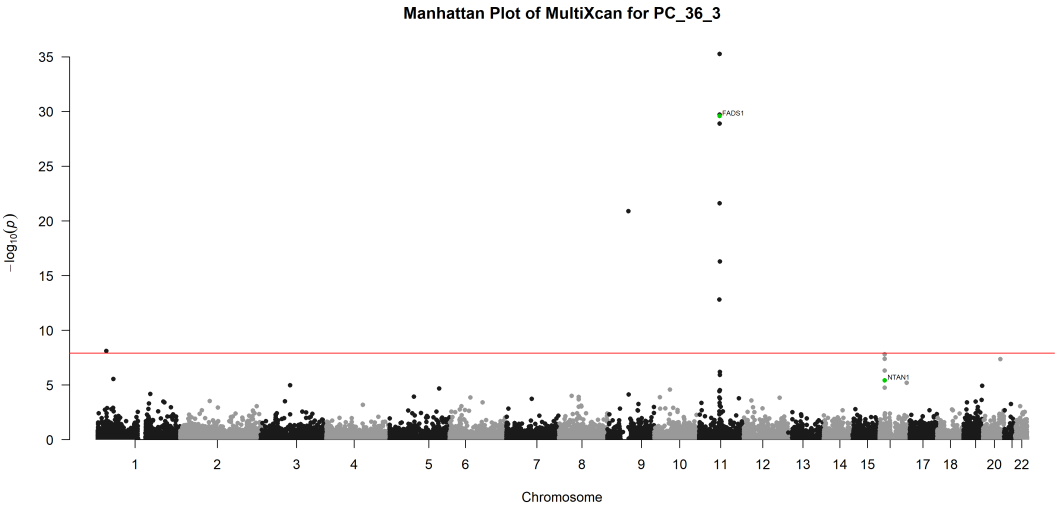
Chr11_59978355_62914375



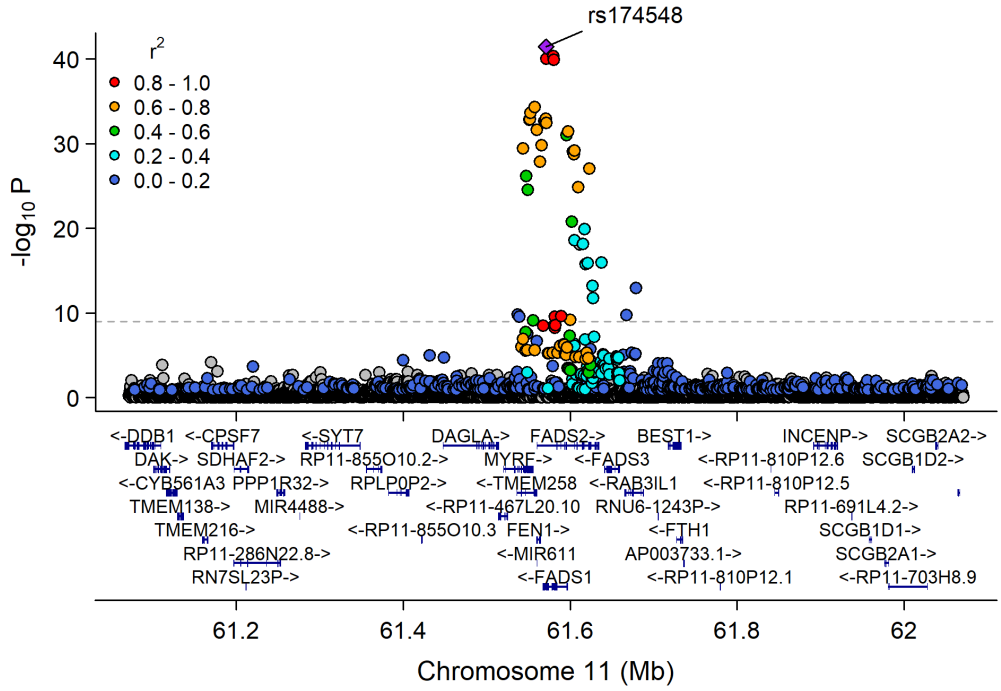
Phosphatidylcholine diacyl C36:3 (µg/dL)



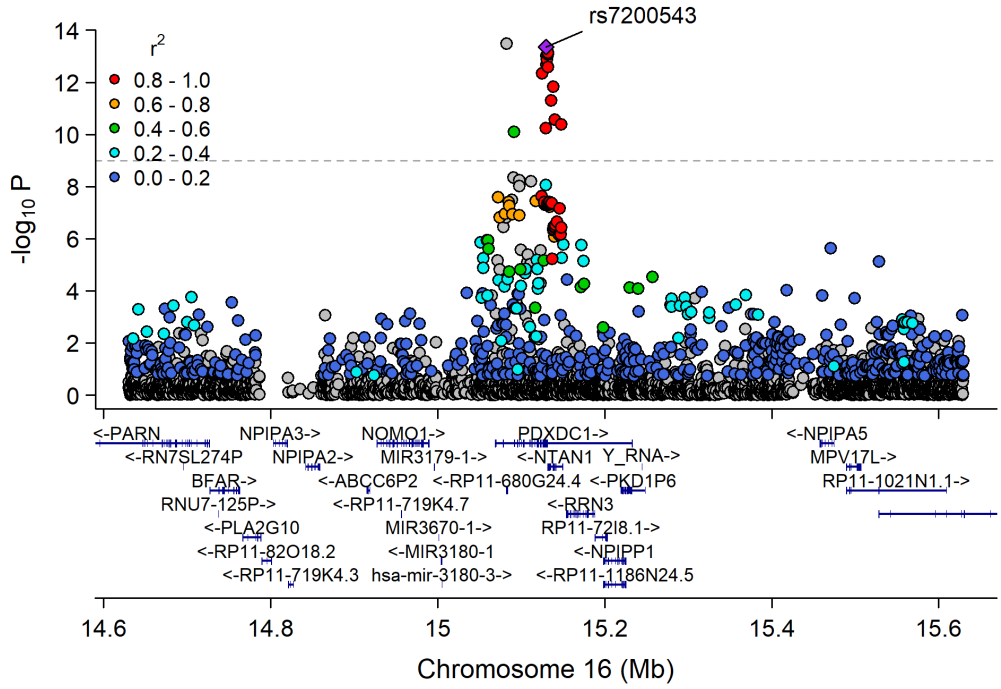




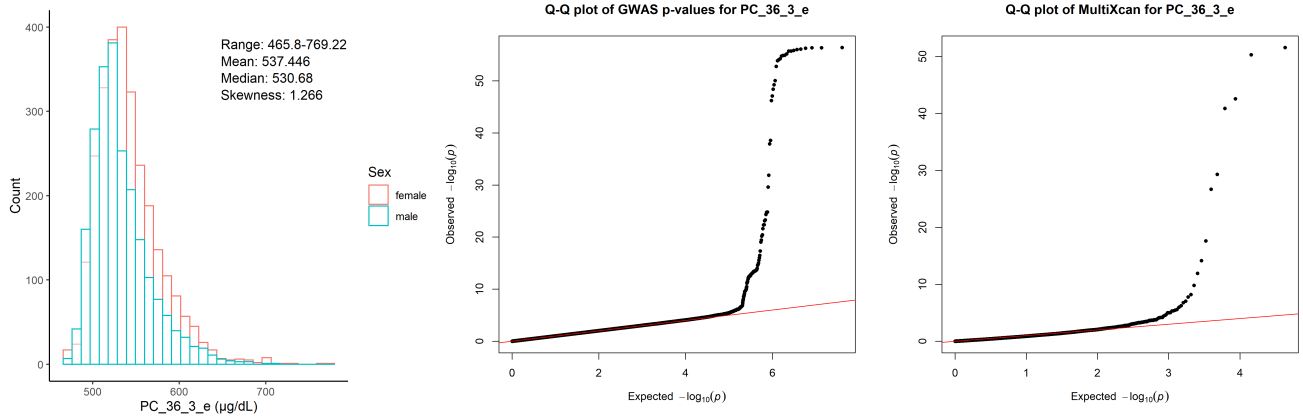
Chr11_59978355_62914375

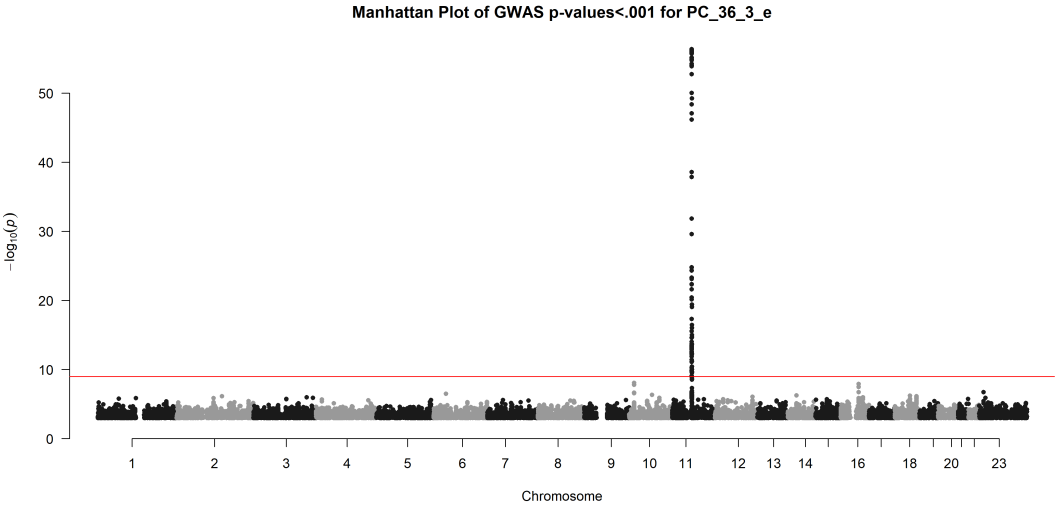


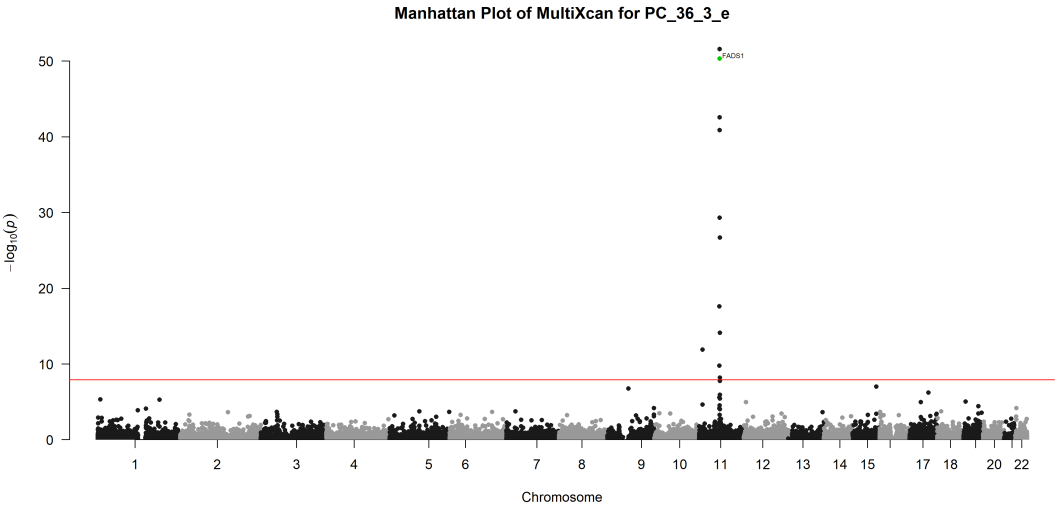
Chr16_14393707_15885866



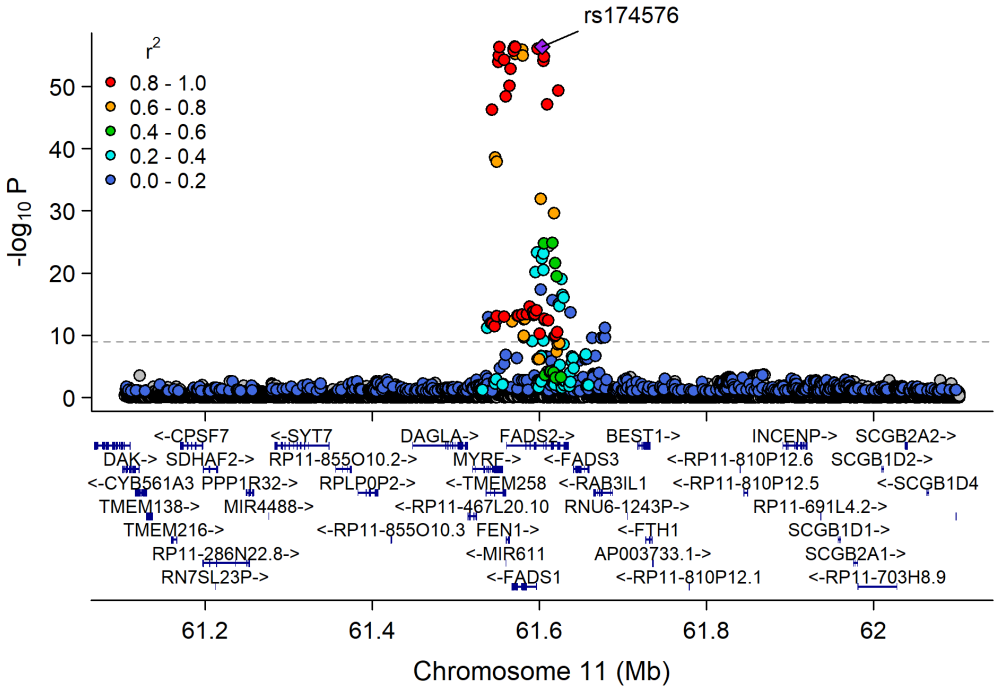
Phosphatidylcholine acyl-alkyl C36:3 (µg/dL)



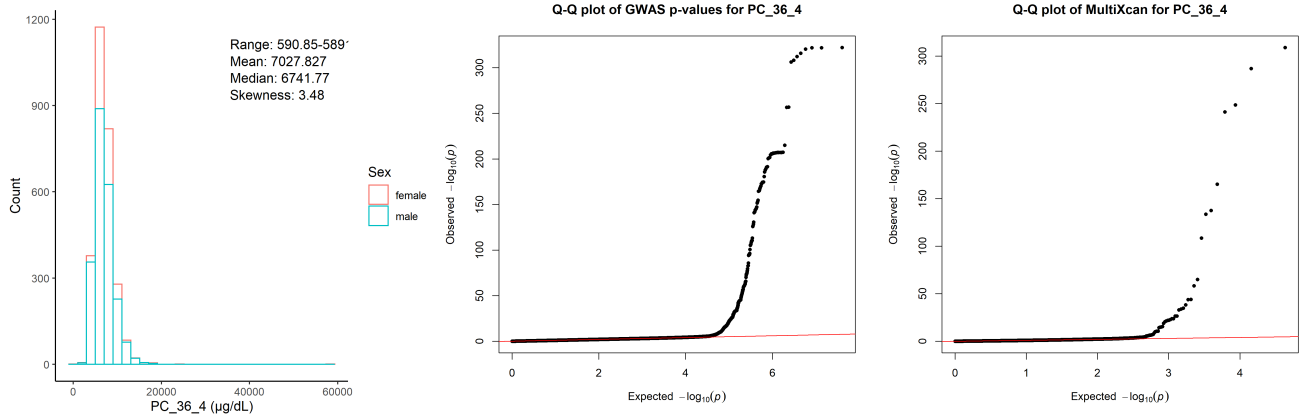


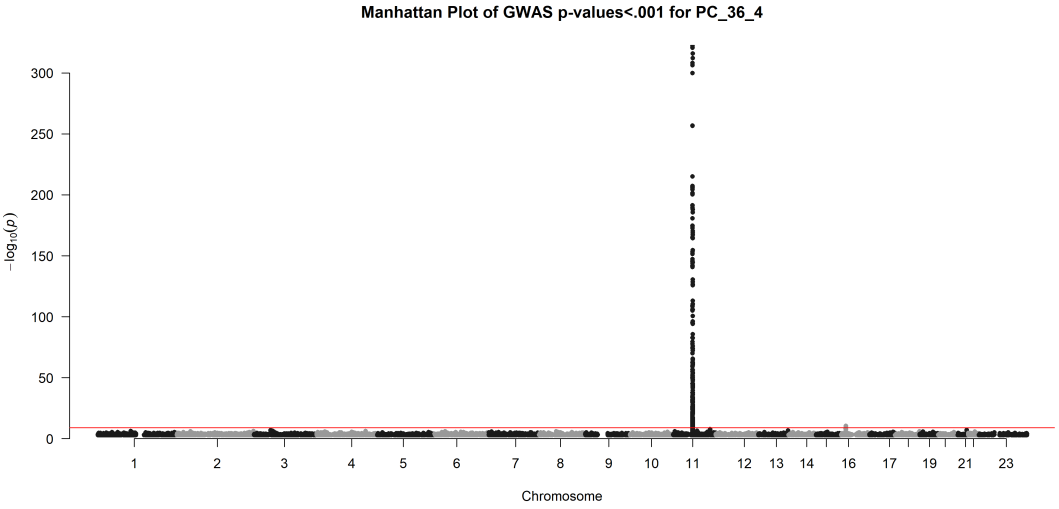


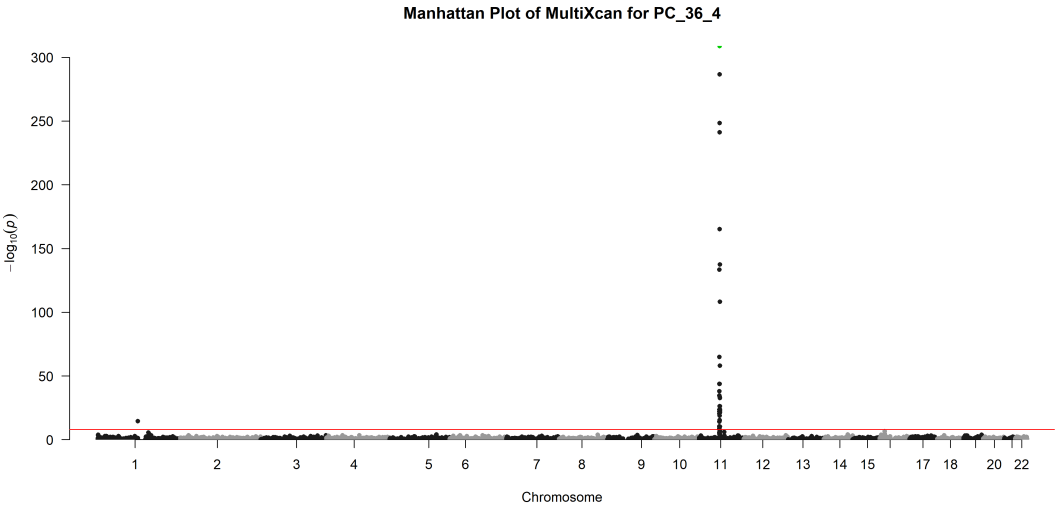
Chr11_59978355_62914375



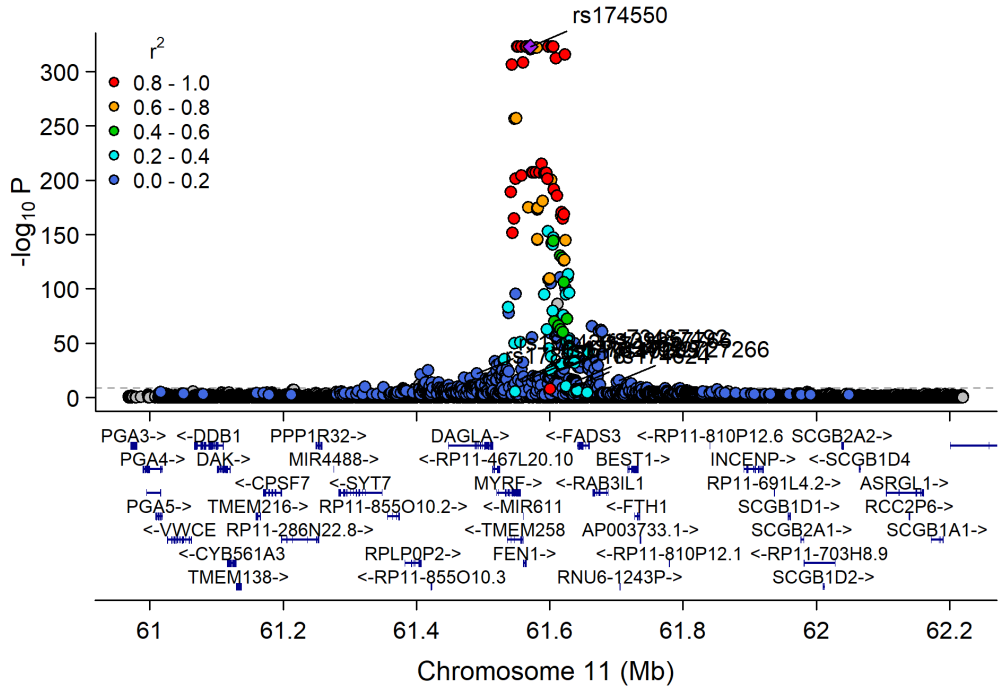
Phosphatidylcholine diacyl C36:4 (µg/dL)



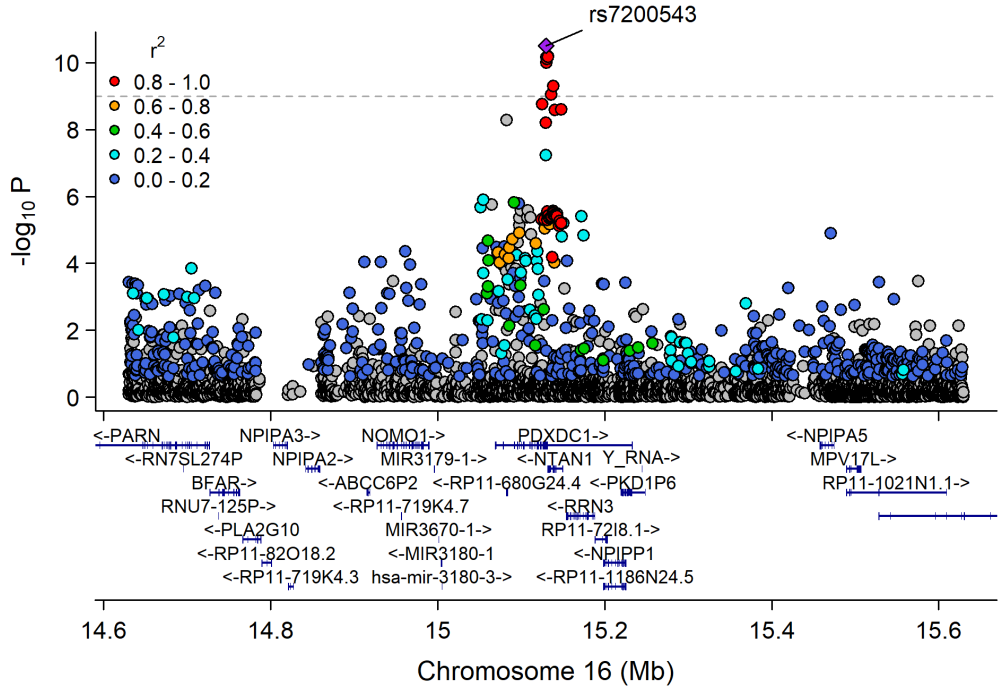




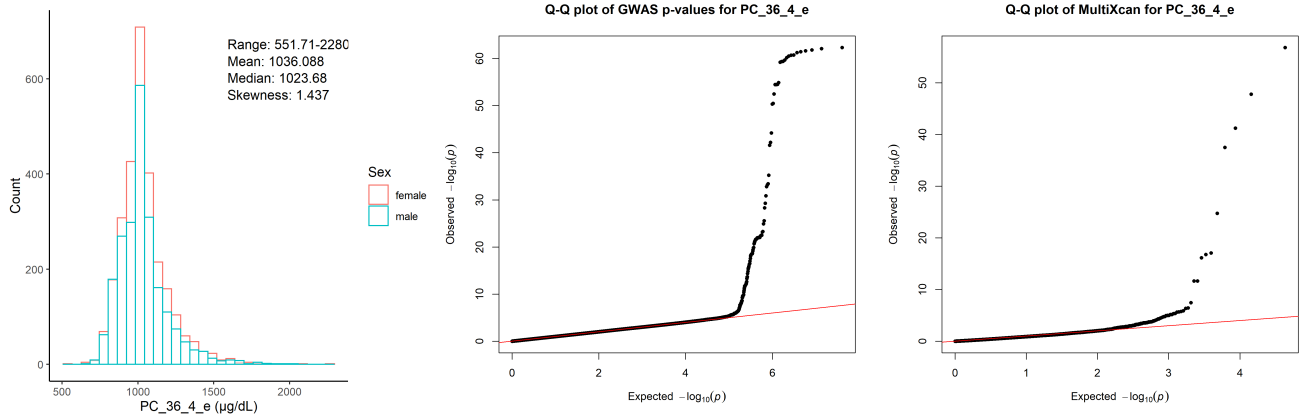
Chr11_59978355_62914375

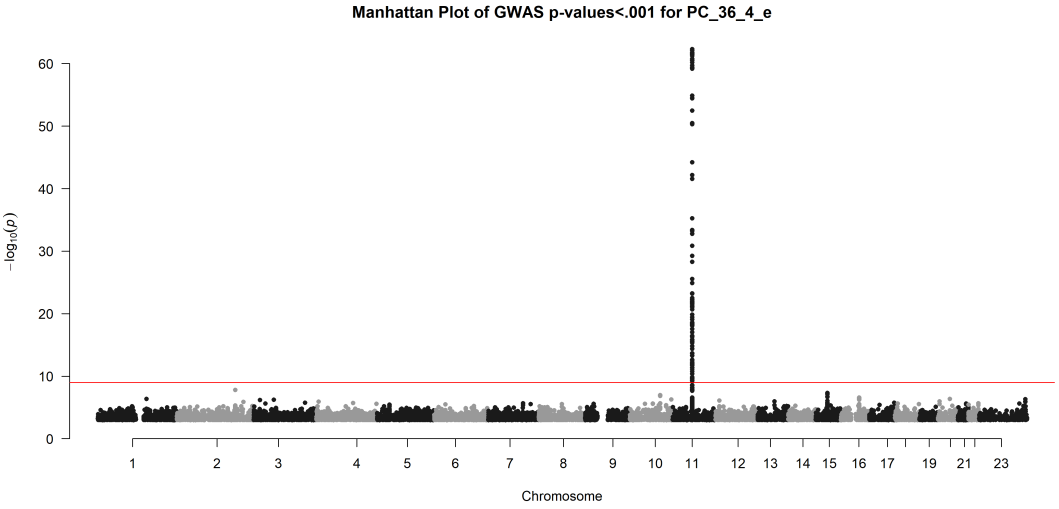


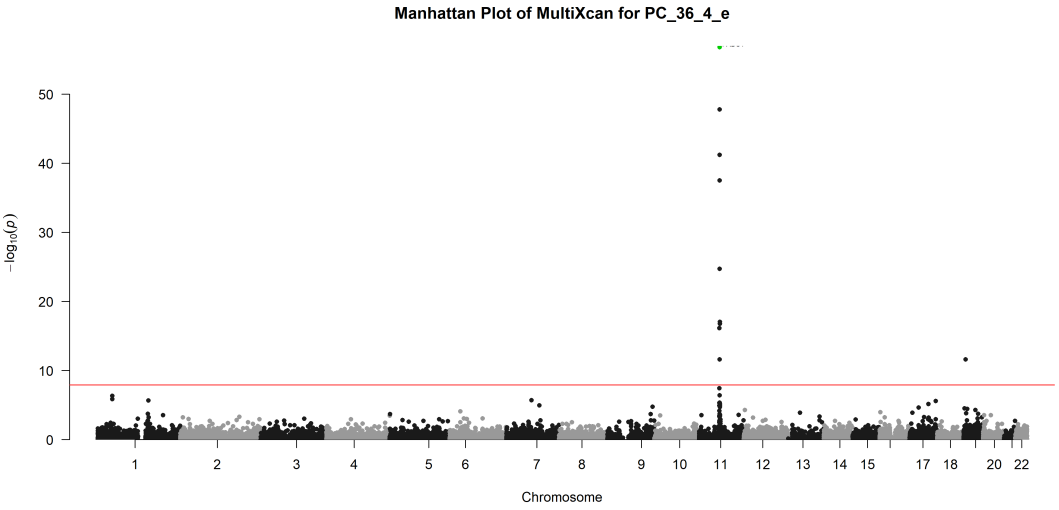
Chr16_14393707_15885866



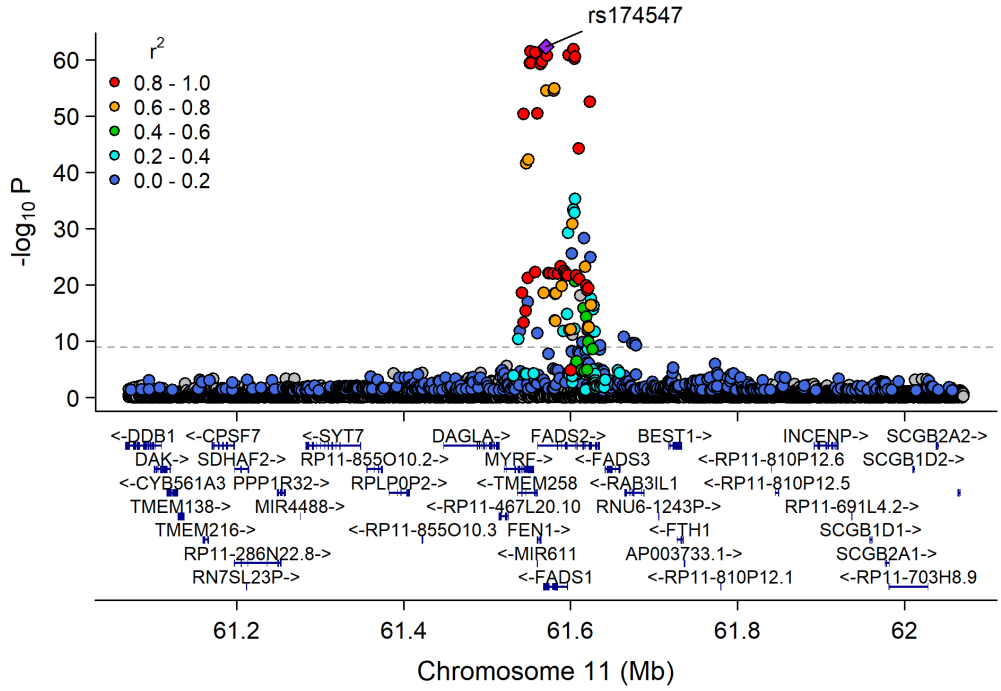
Phosphatidylcholine acyl-alkyl C36:4 (µg/dL)



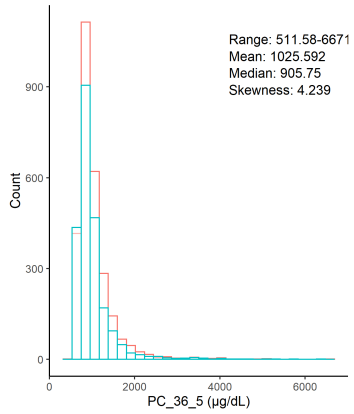


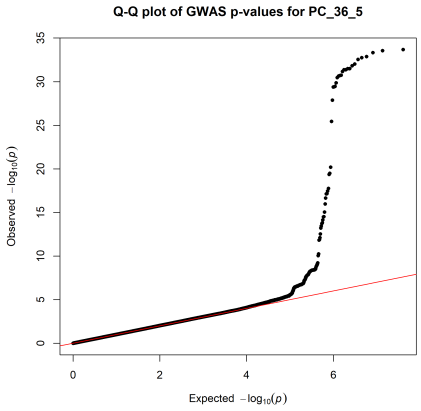


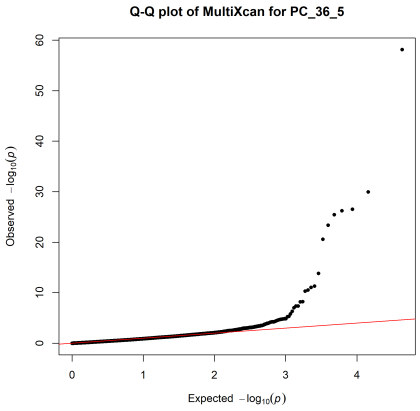
Chr11__59978355__62914375

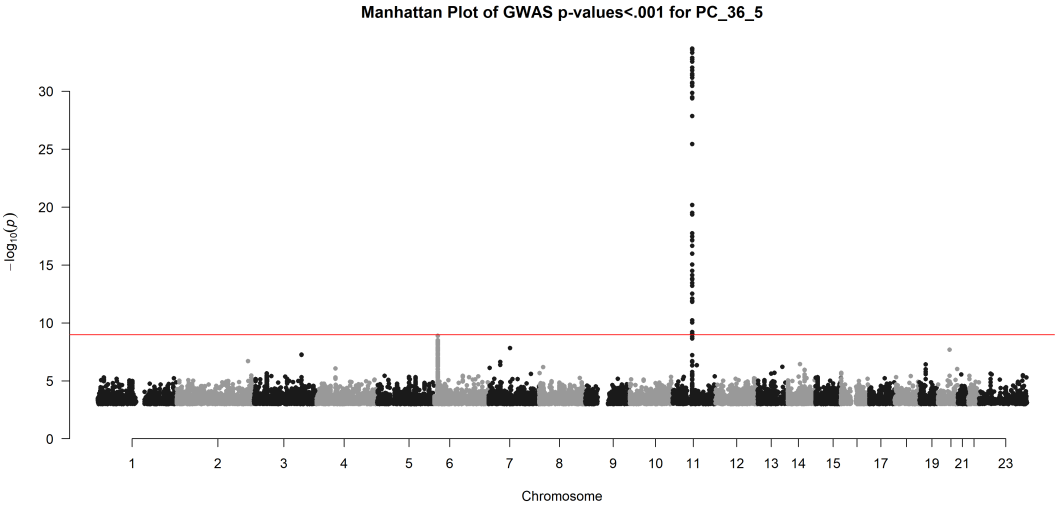


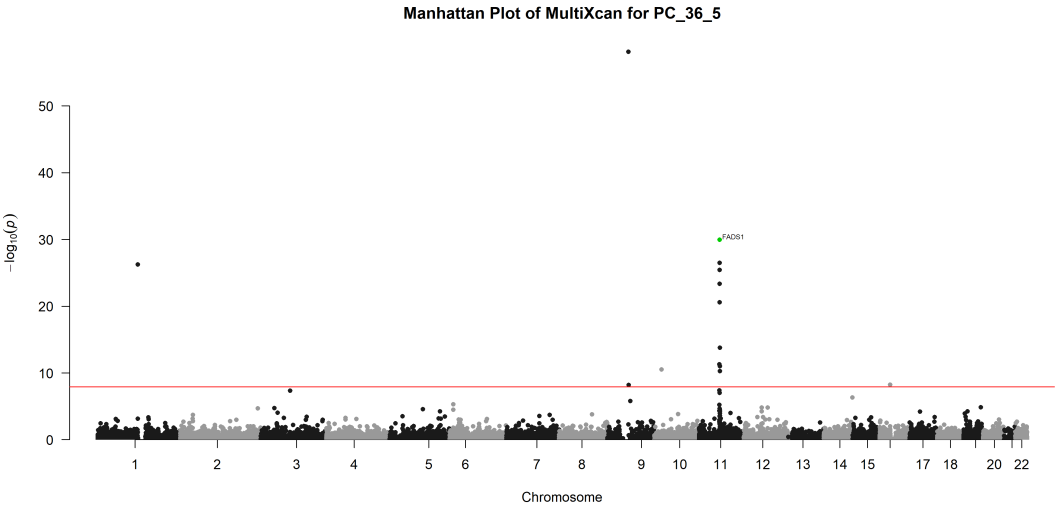
Phosphatidylcholine diacyl C36:5 (µg/dL)



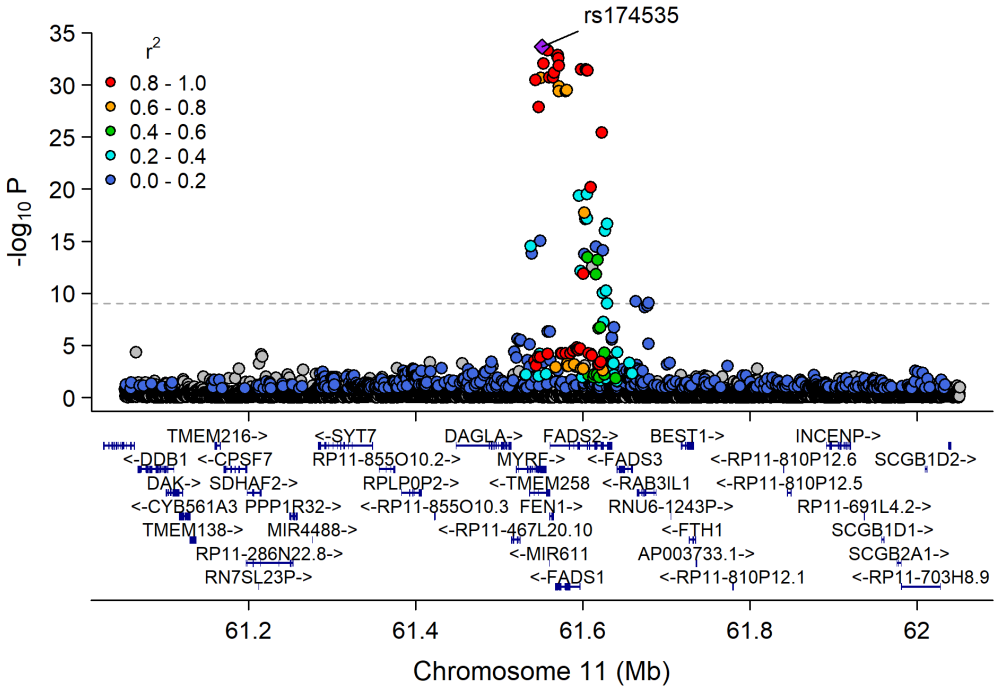




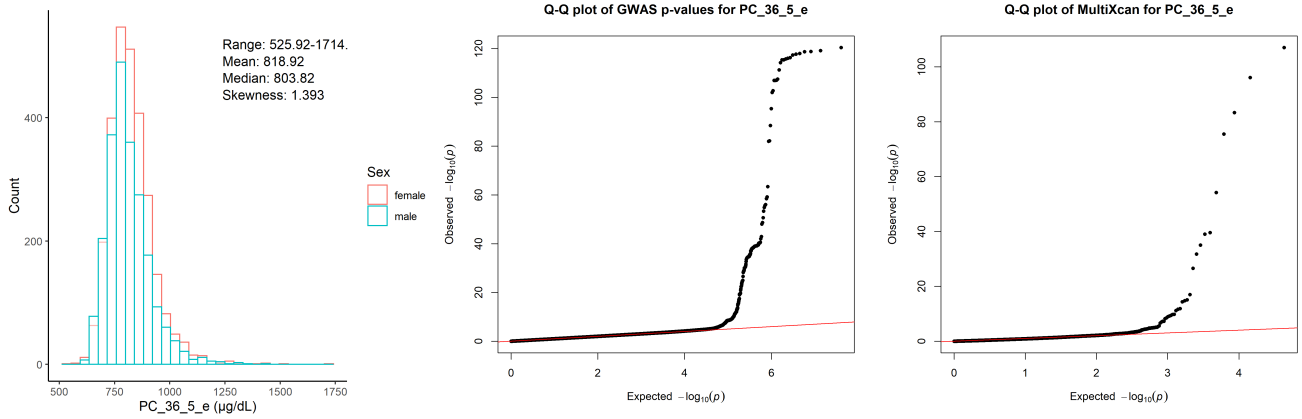


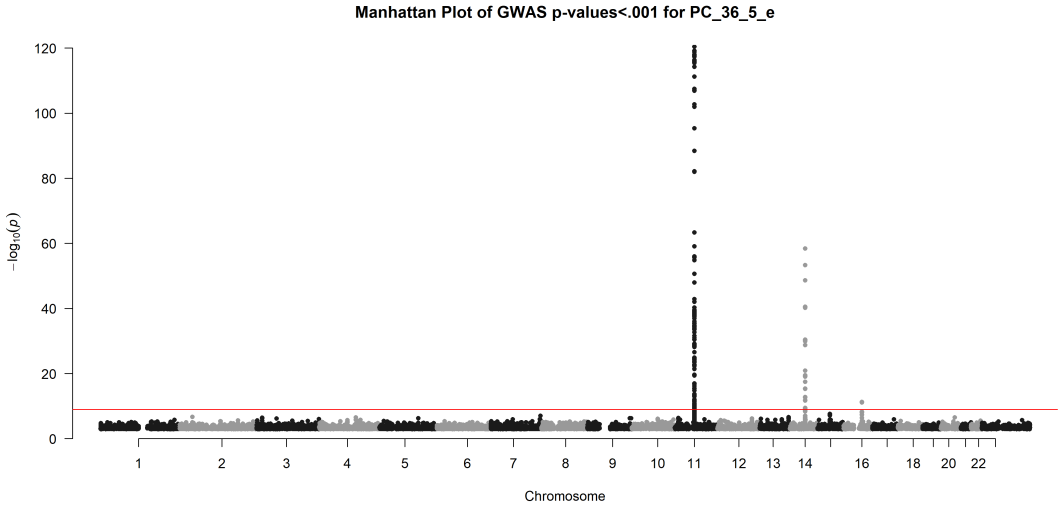


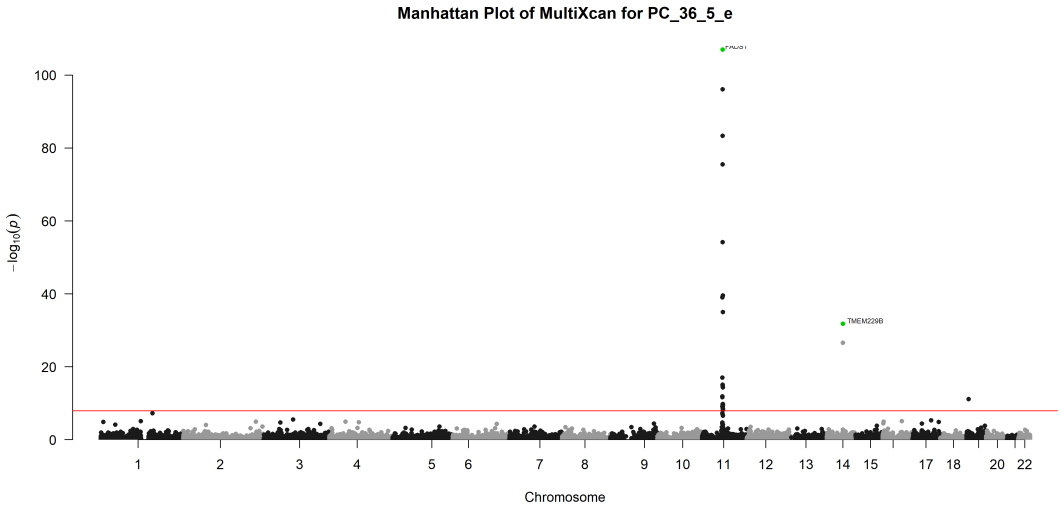
Chr11__59978355__62914375



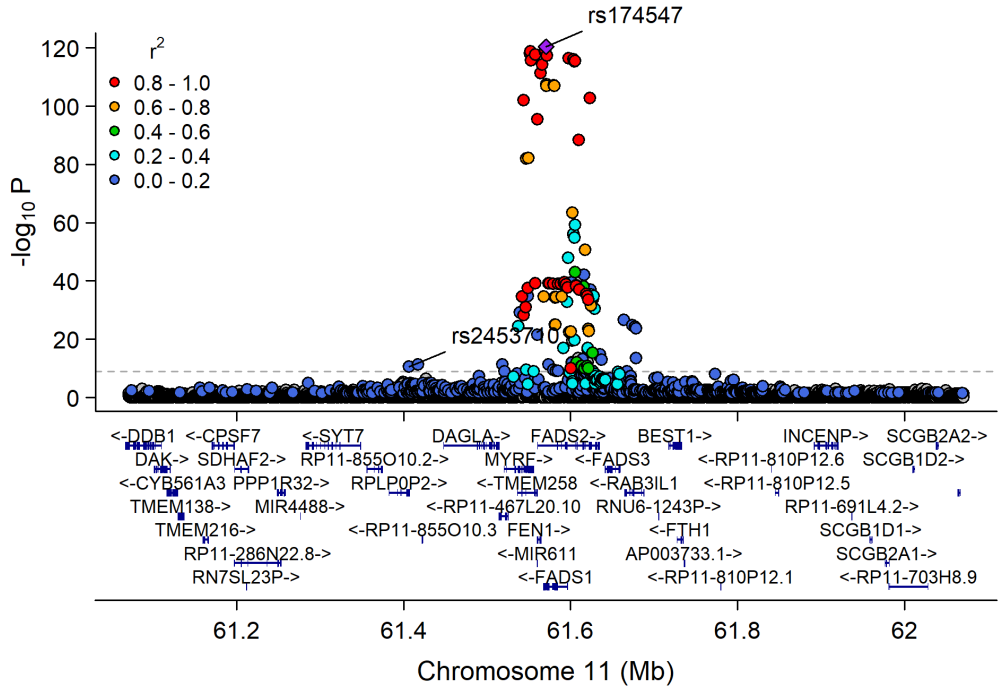
Phosphatidylcholine acyl-alkyl C36:5 (µg/dL)



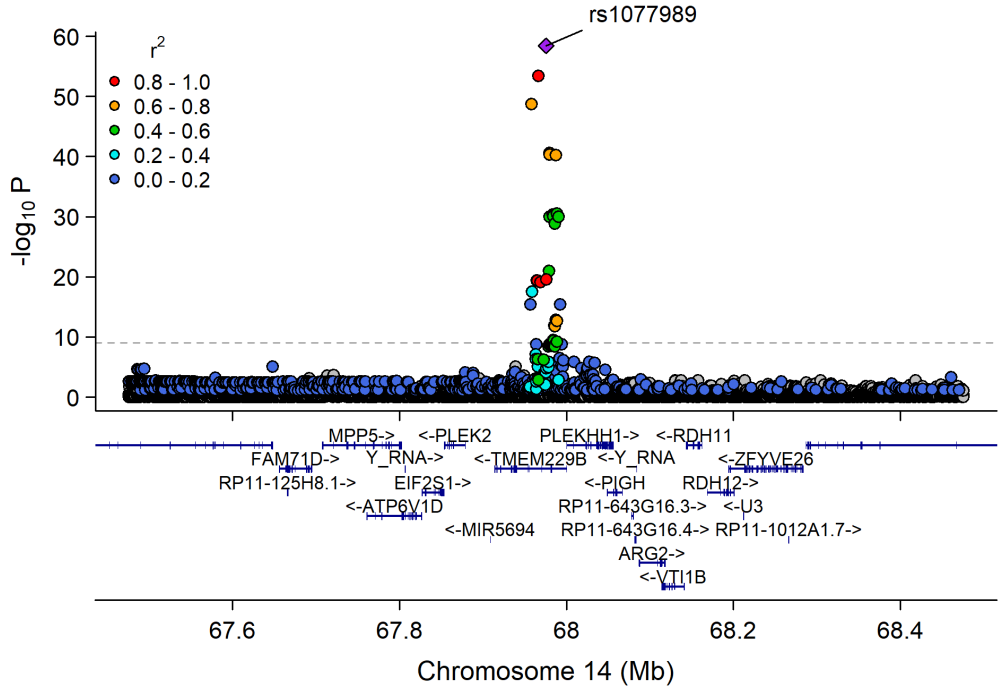




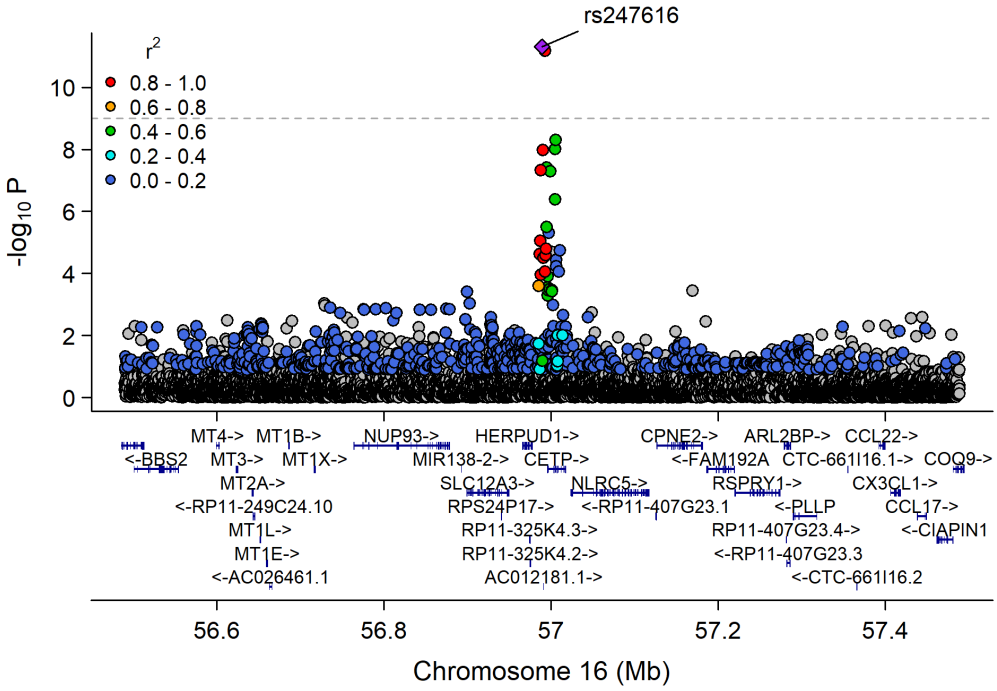
Chr11_59978355_62914375



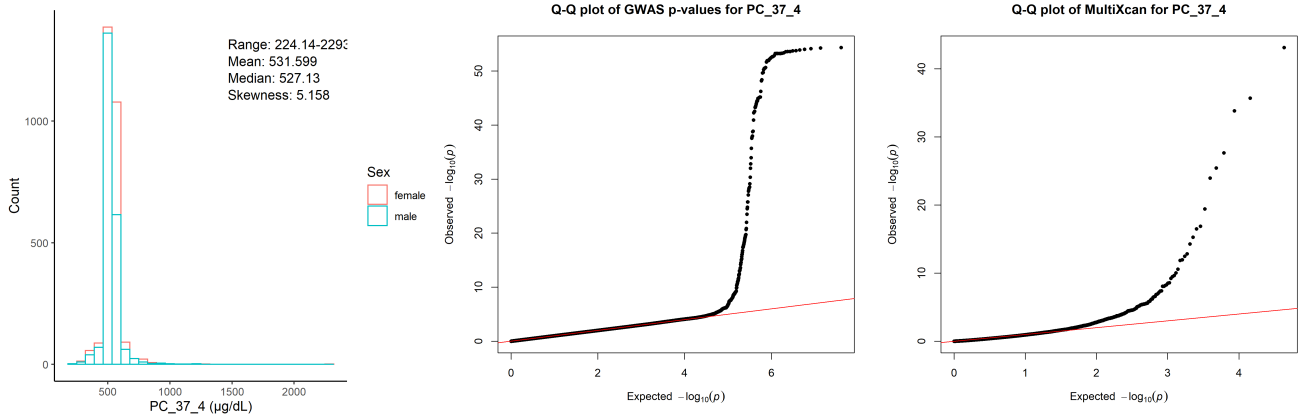
Chr14_66997265_68972354

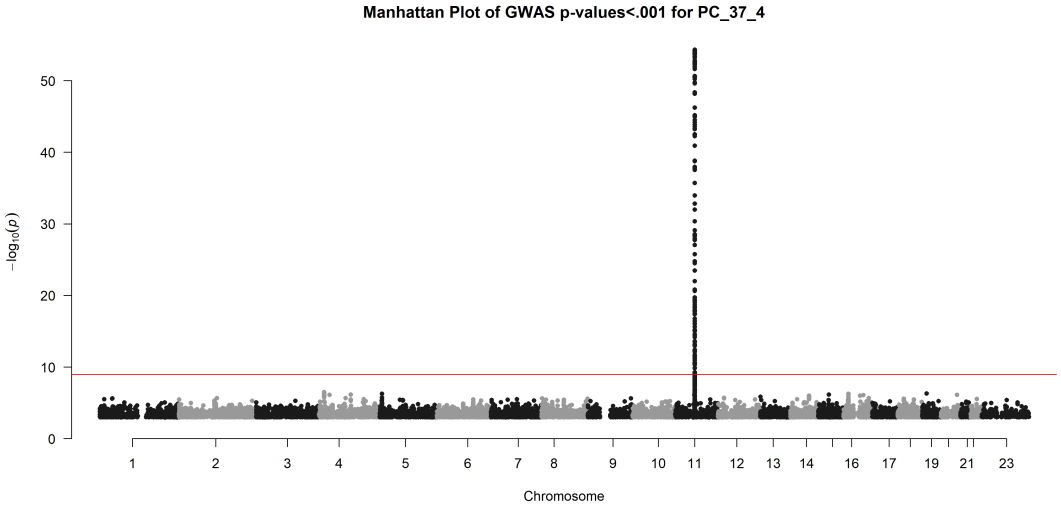


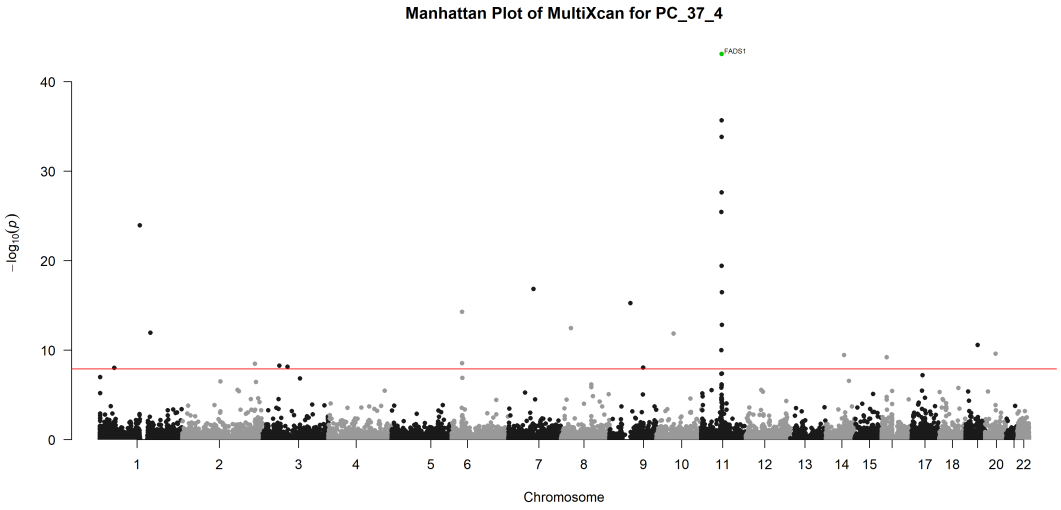
Chr16_55870822_57992421



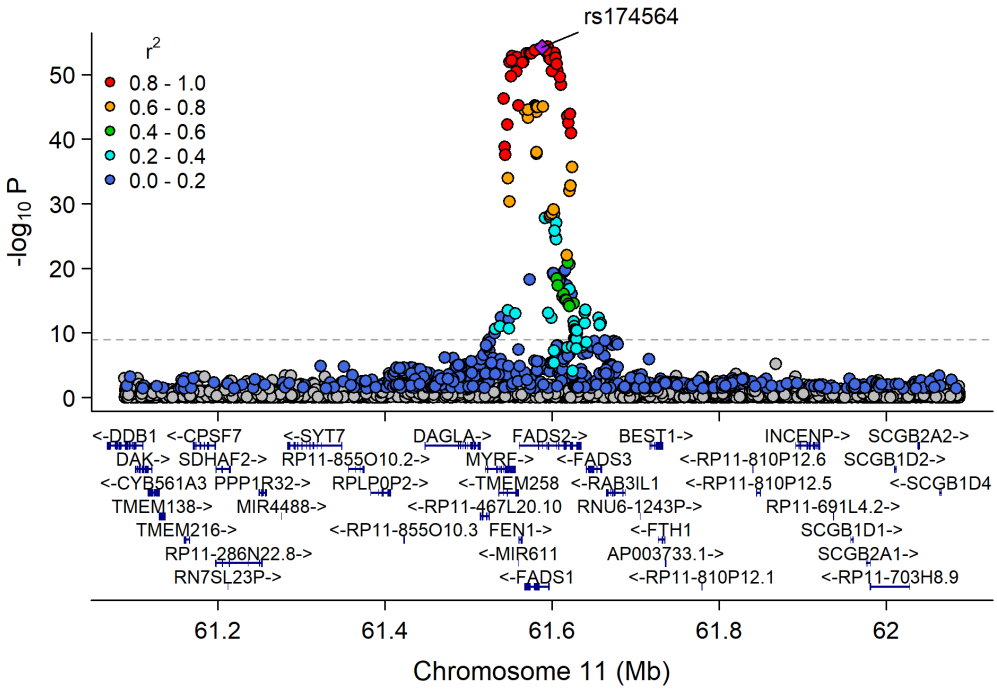
Phosphatidylcholine diacyl C37:4 (µg/dL)



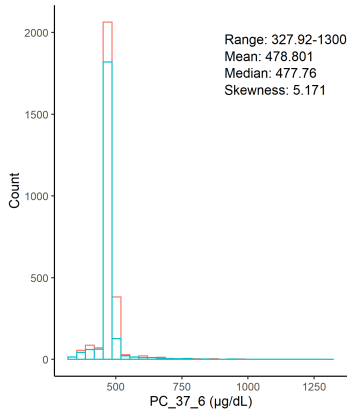


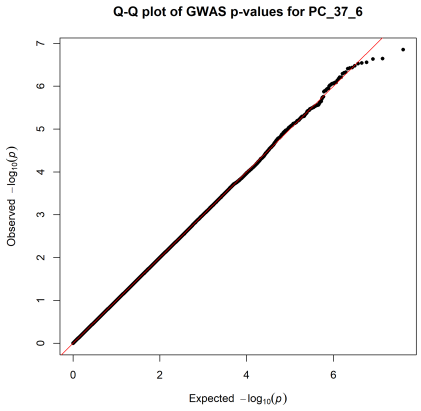


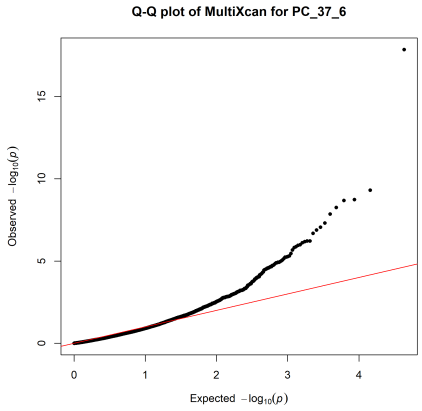
Chr11_59978355_62914375

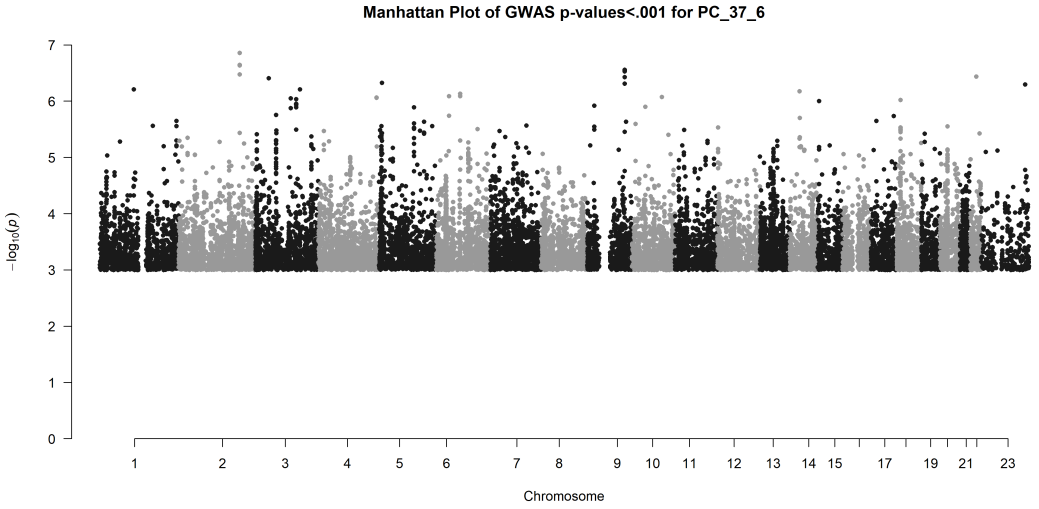


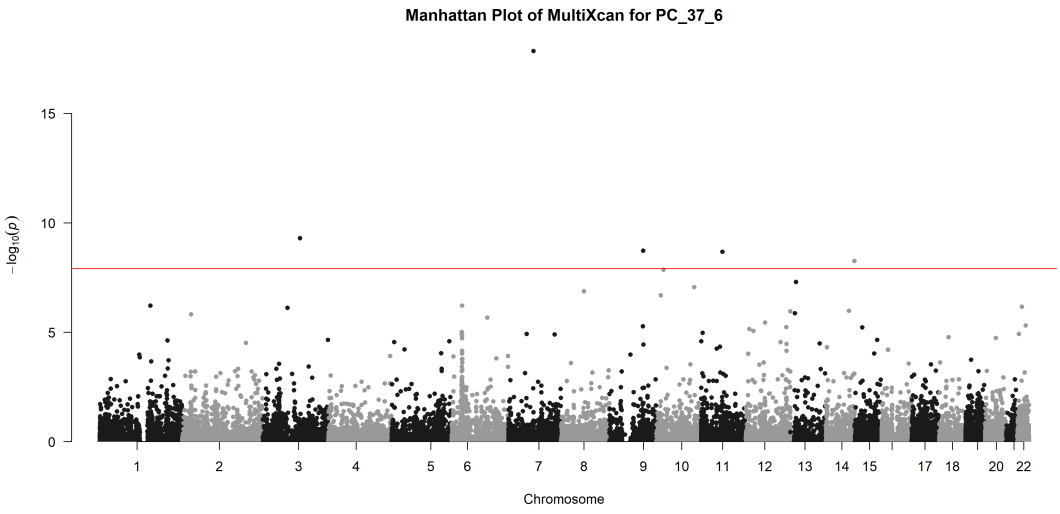
Phosphatidylcholine diacyl C37:6 (µg/dL)



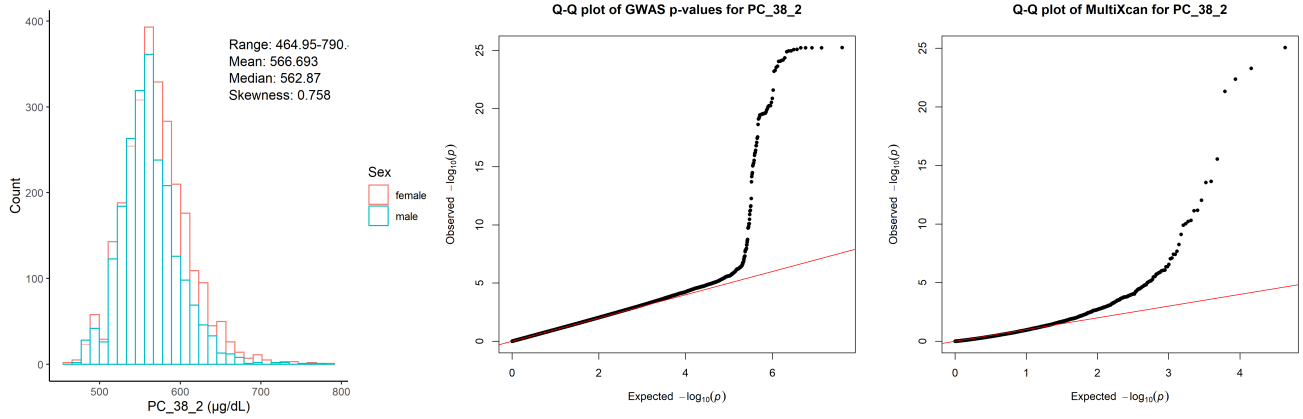


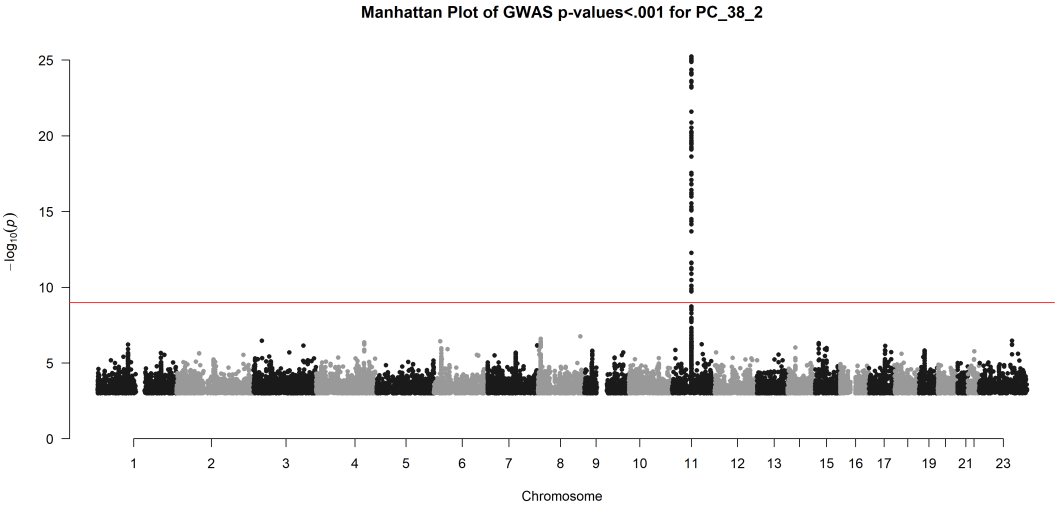


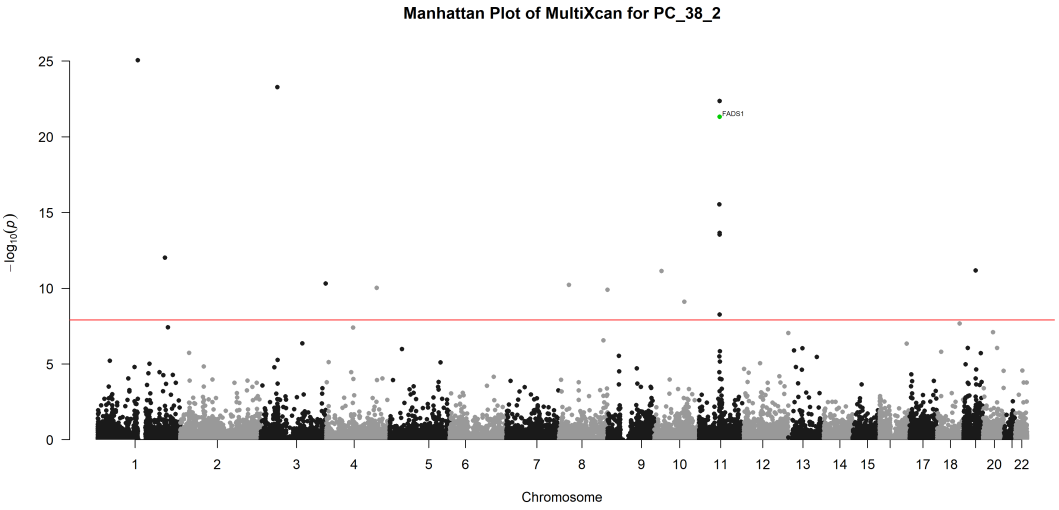




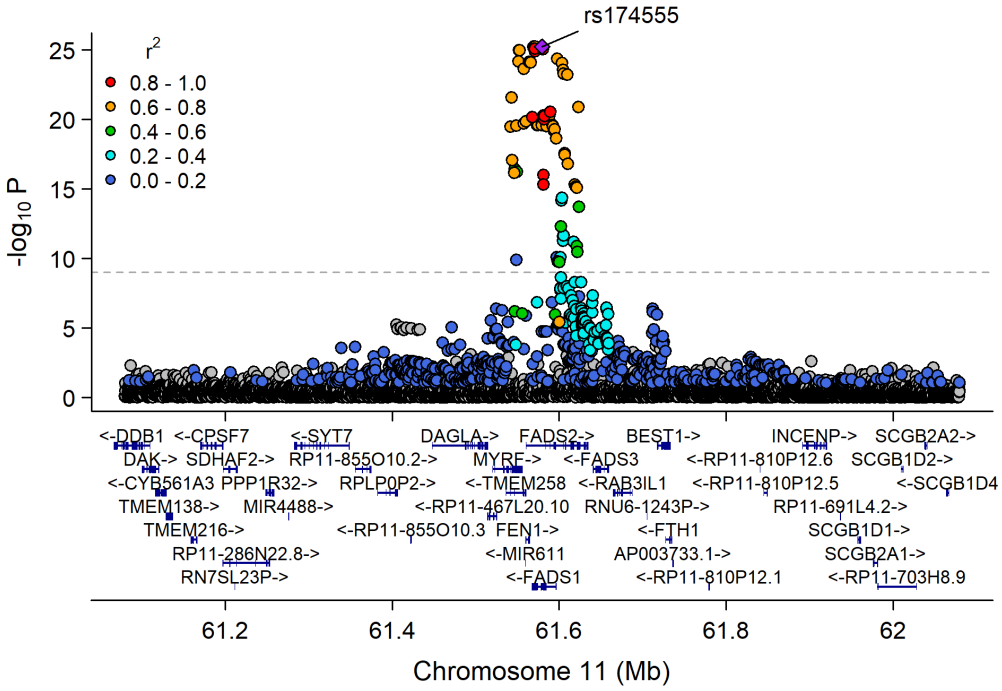
Phosphatidylcholine diacyl C38:2 (µg/dL)



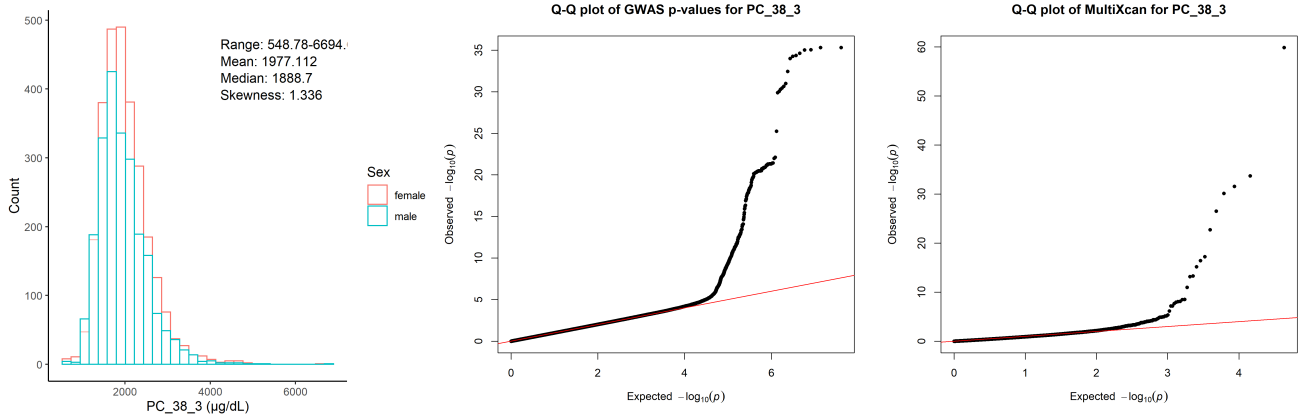


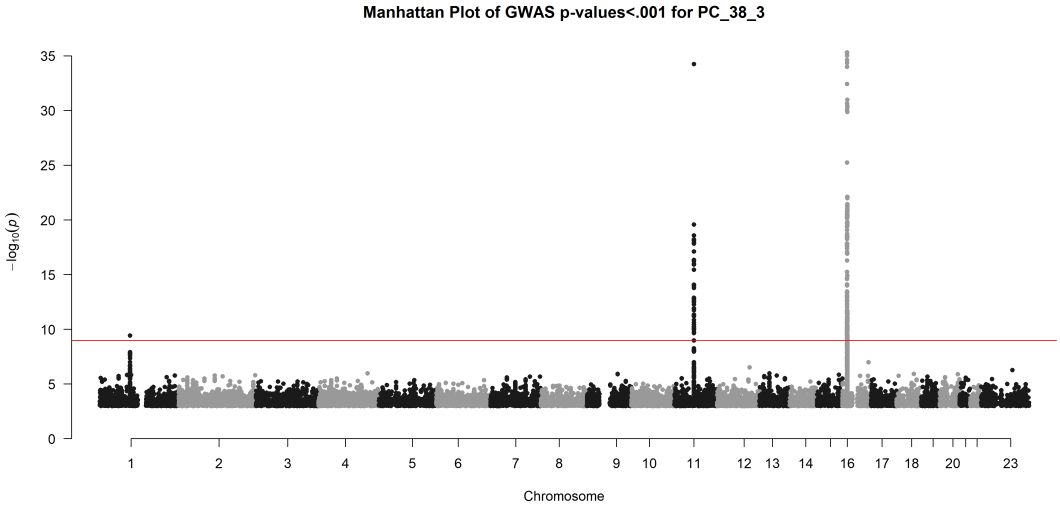


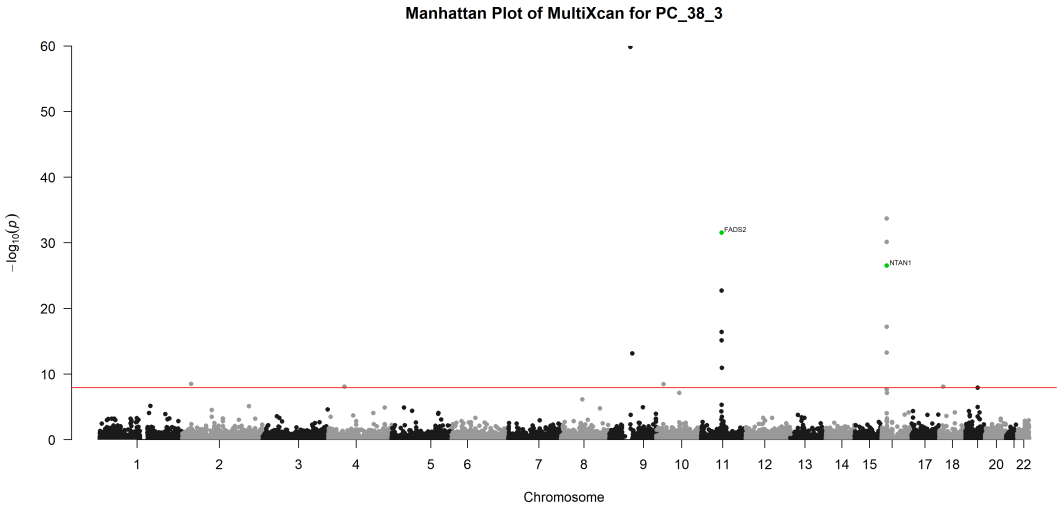
Chr11_59978355_62914375



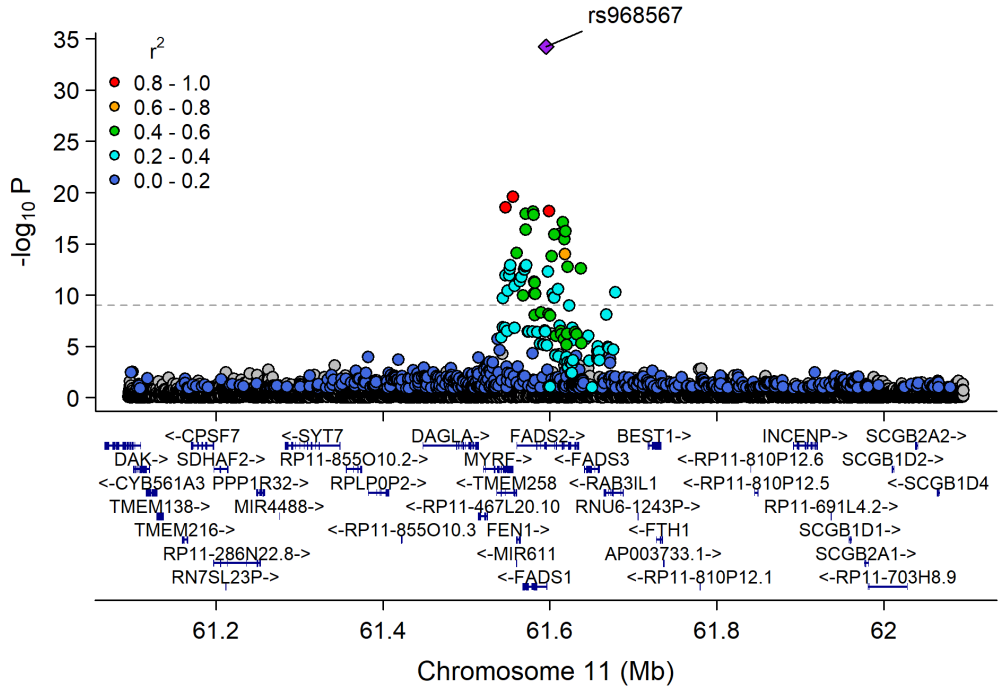
Phosphatidylcholine diacyl C38:3 (µg/dL)



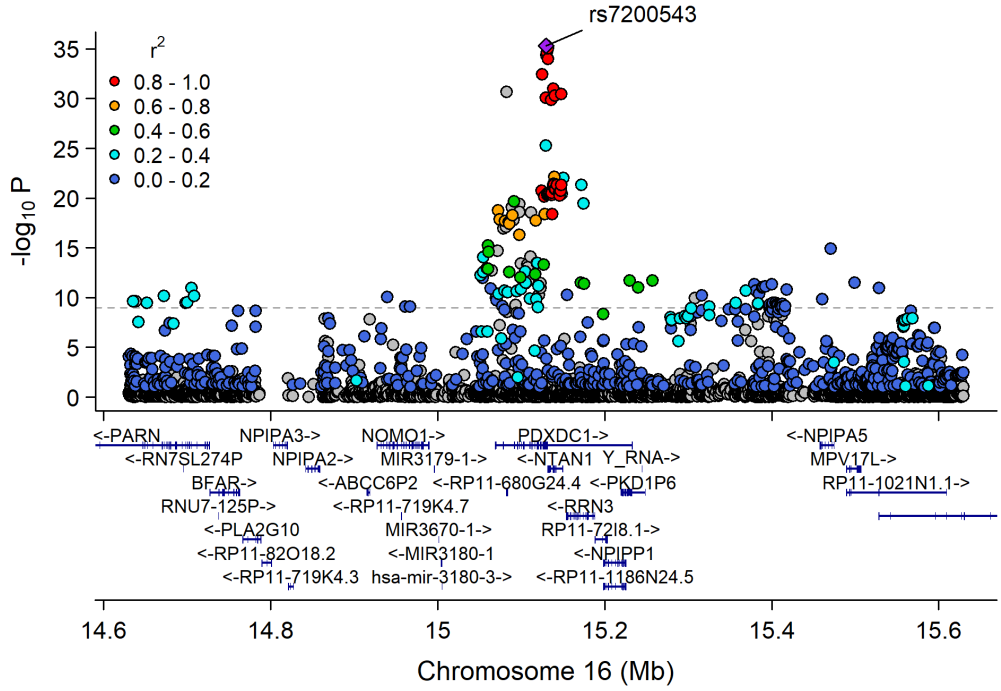




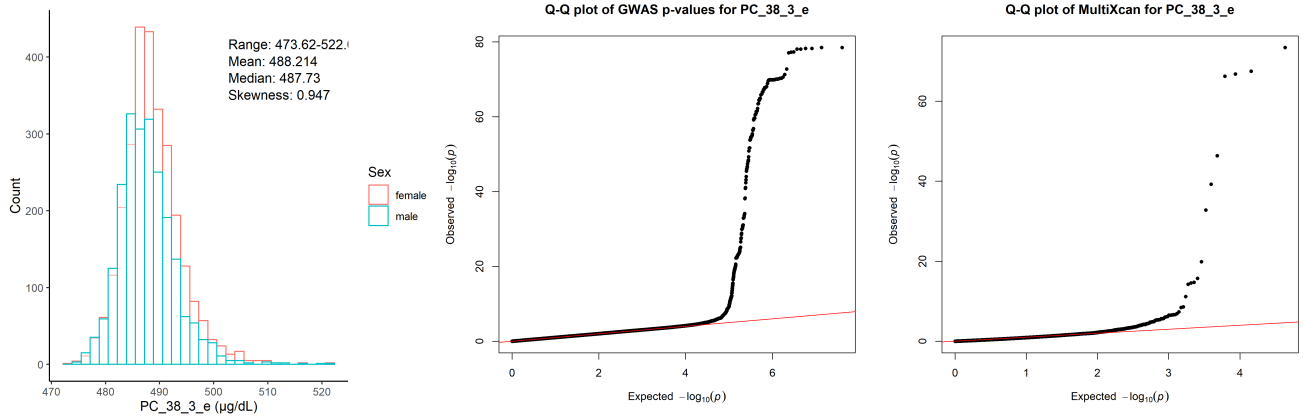
Chr11_59978355_62914375

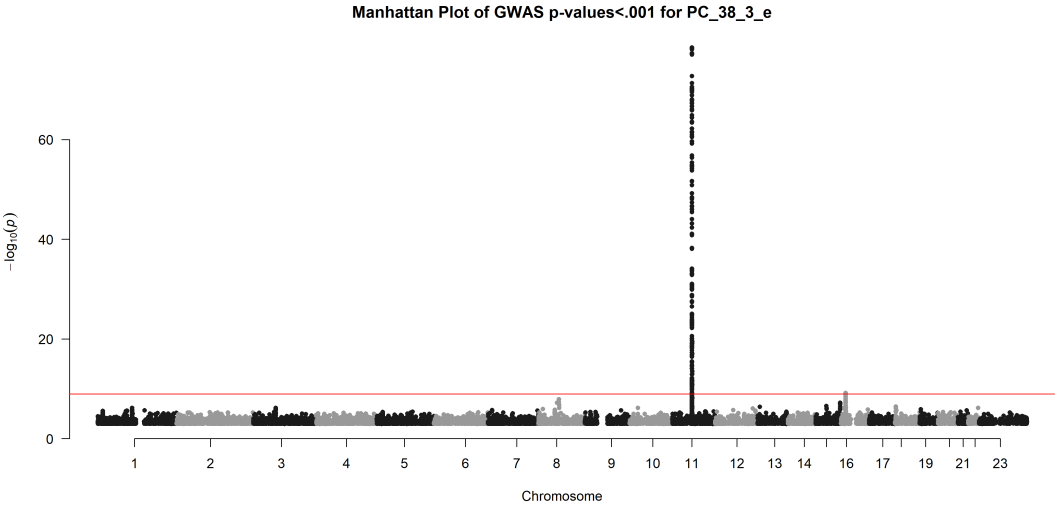


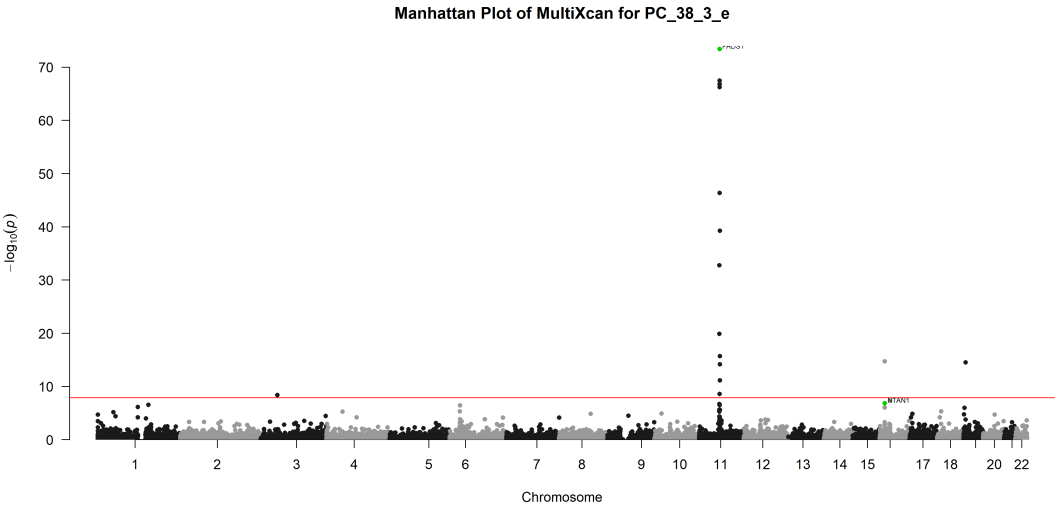
Chr16_14393707_15885866



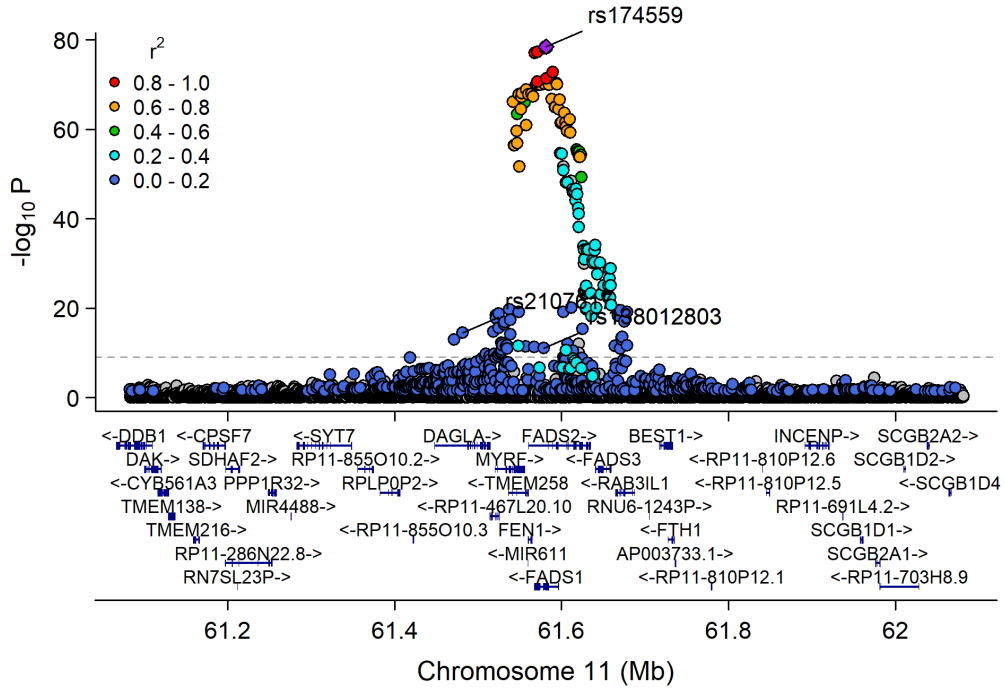
Phosphatidylcholine acyl-alkyl C38:3 (µg/dL)



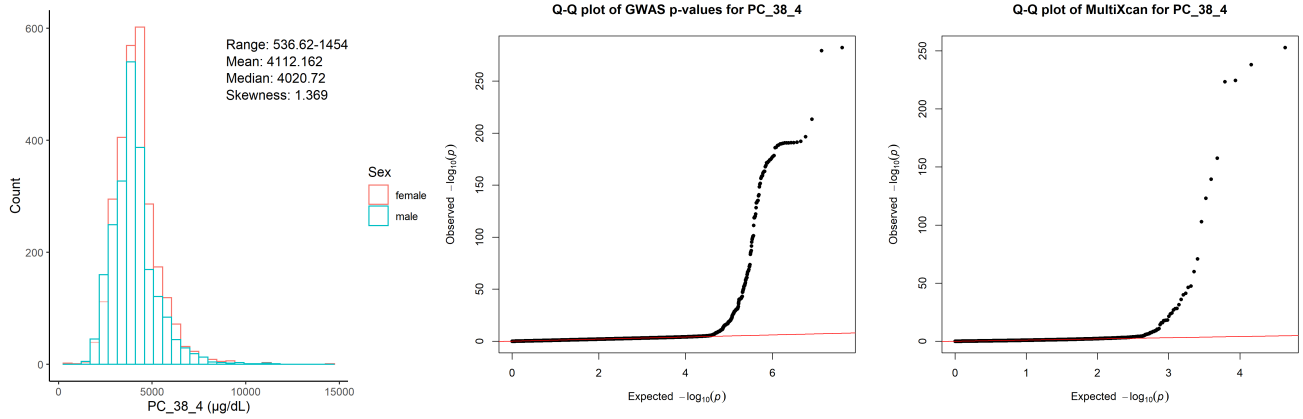


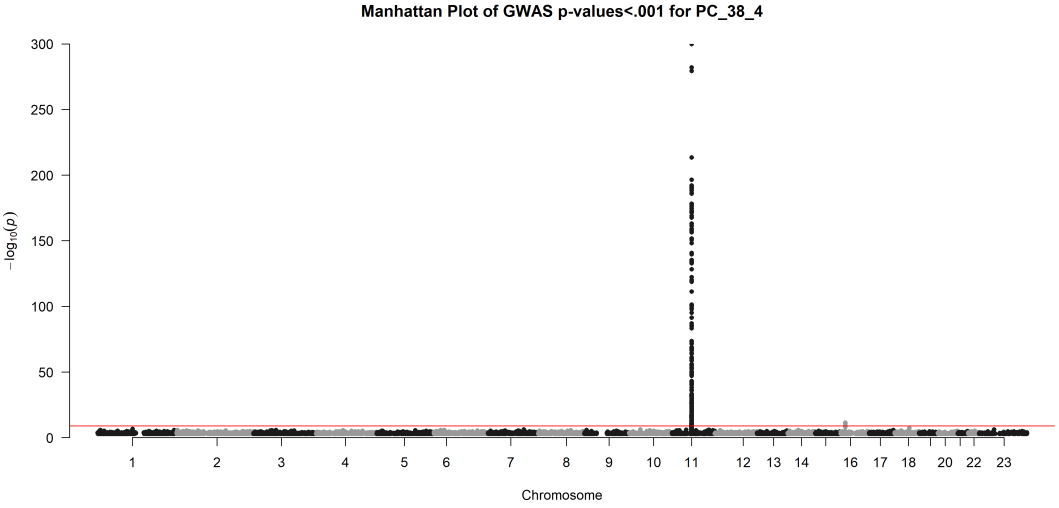


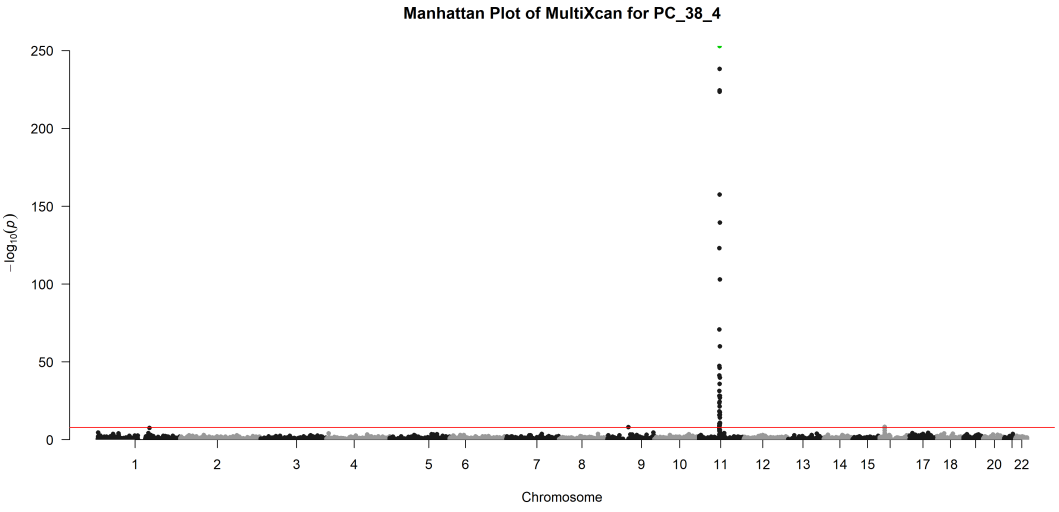
Chr11_59978355_62914375



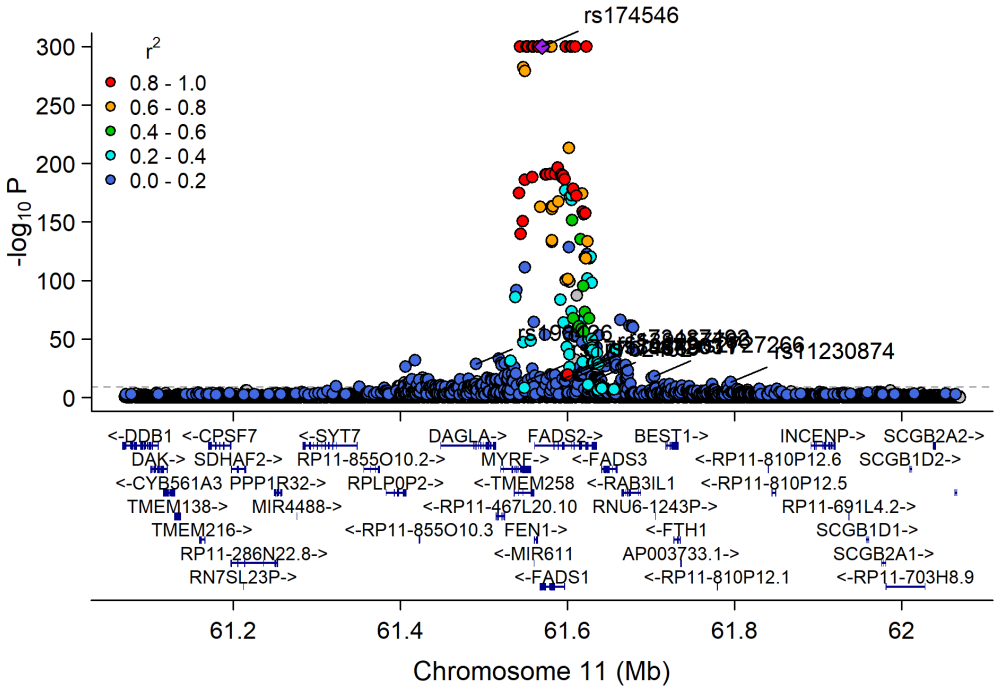
Phosphatidylcholine diacyl C38:4 (µg/dL)



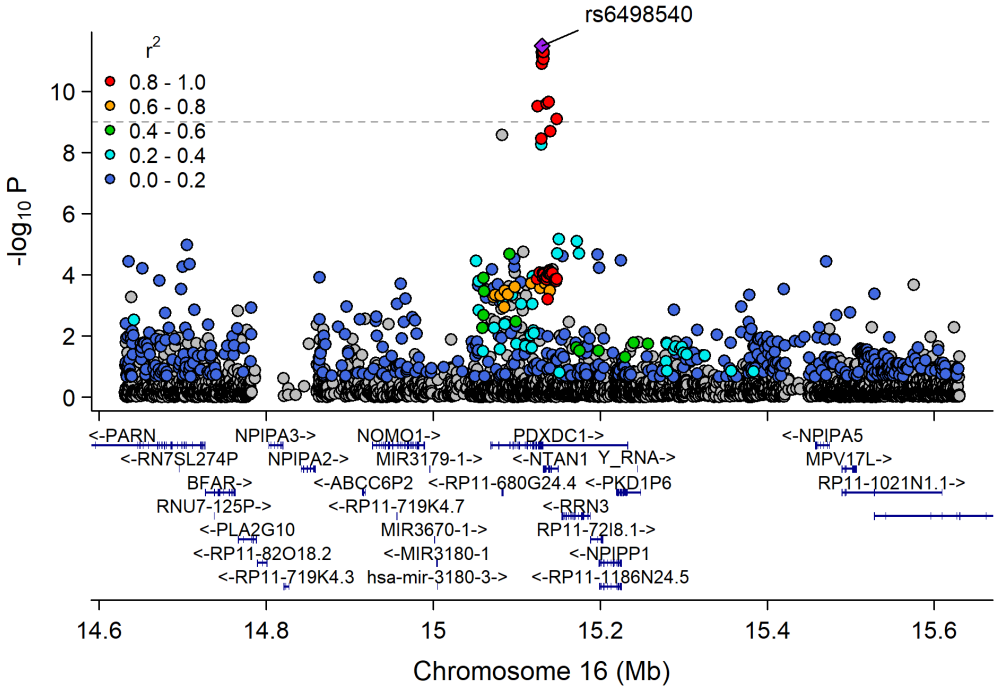




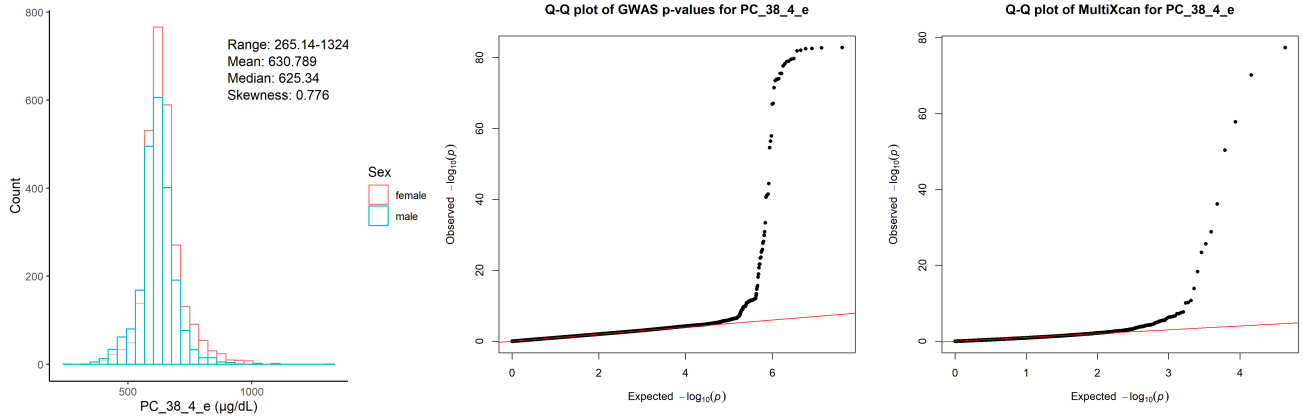
Chr11_59978355_62914375

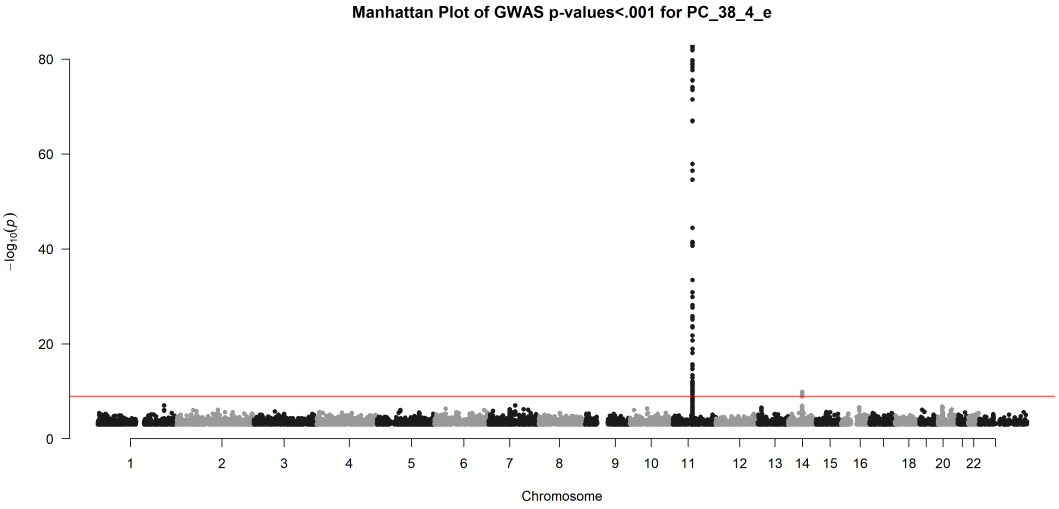


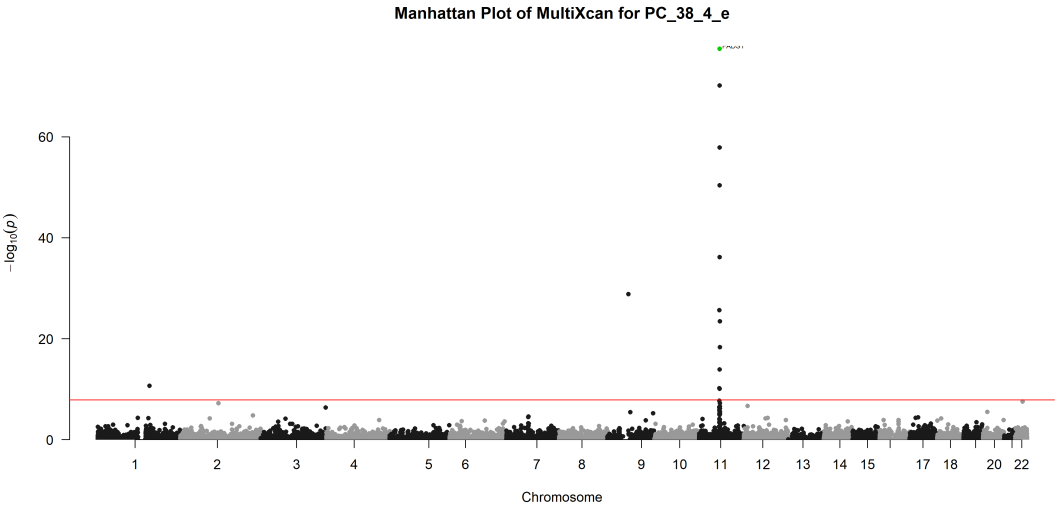
Chr16_14393707_15885866



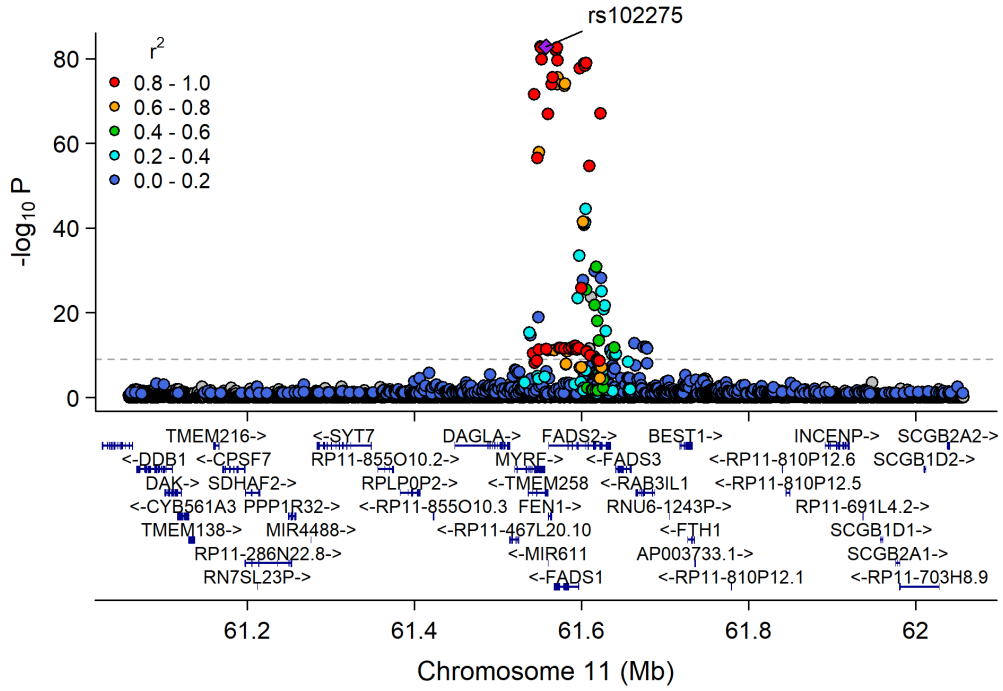
Phosphatidylcholine acyl-alkyl C38:4 (µg/dL)



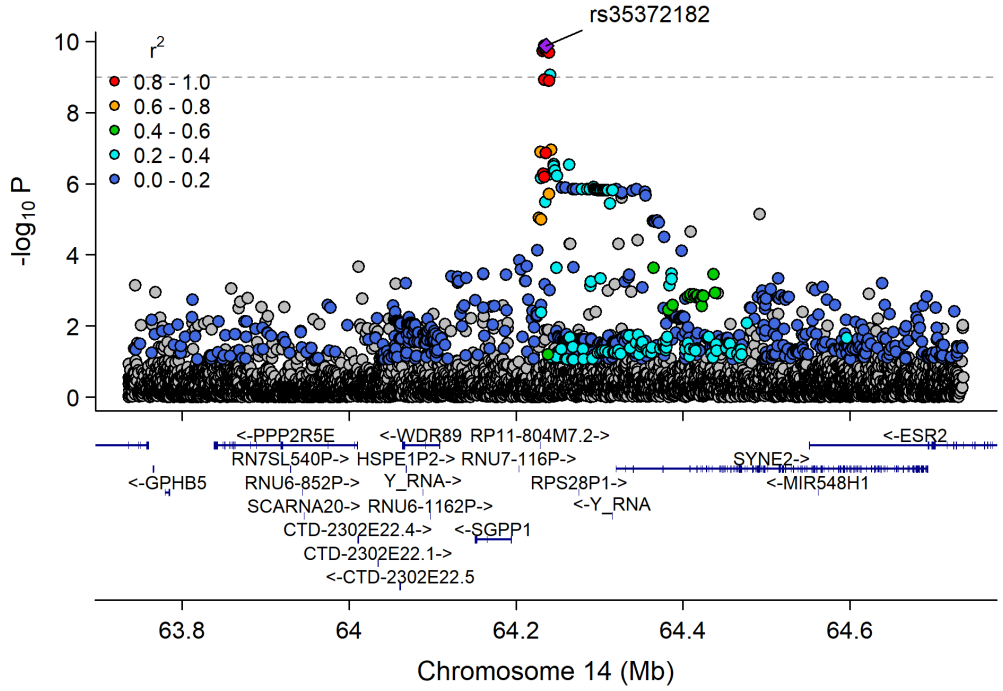




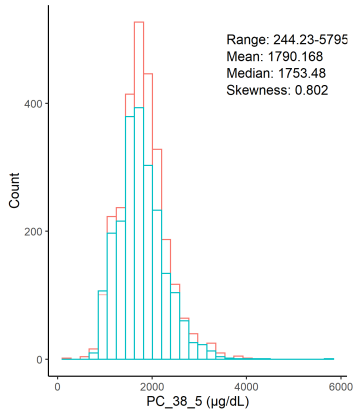
Chr11_59978355_62914375

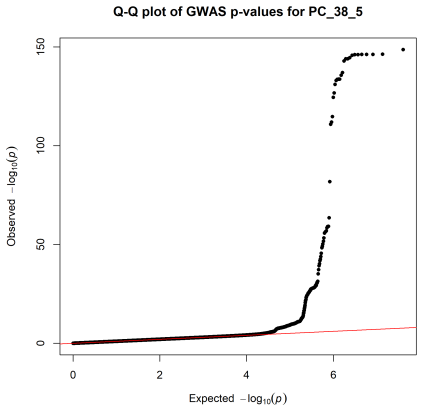


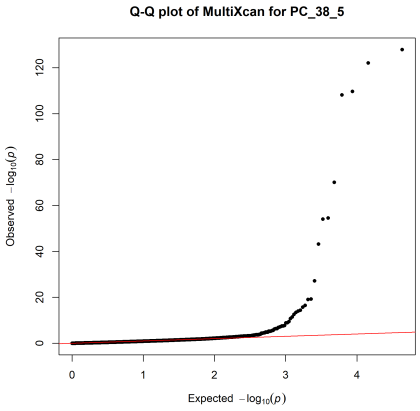
Chr14_63168899_65811394

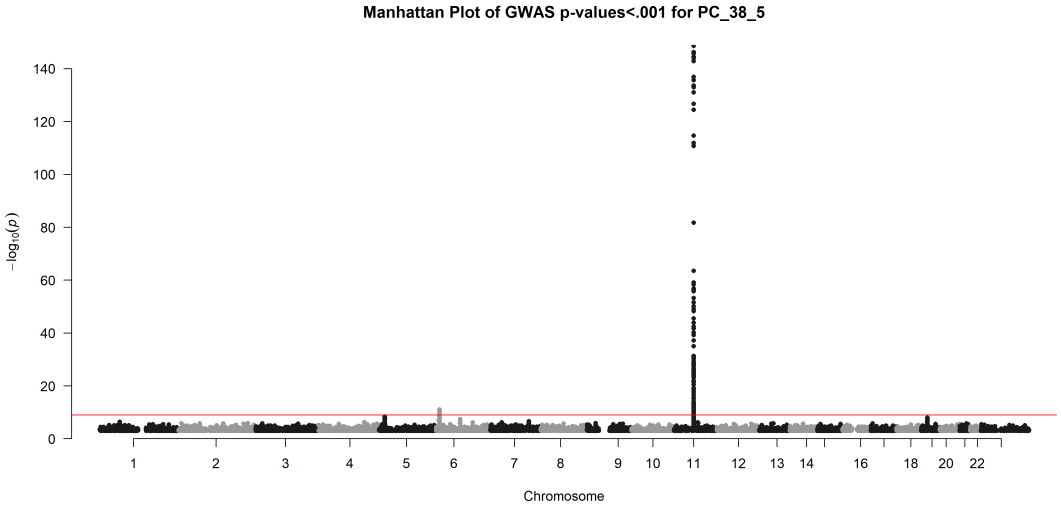


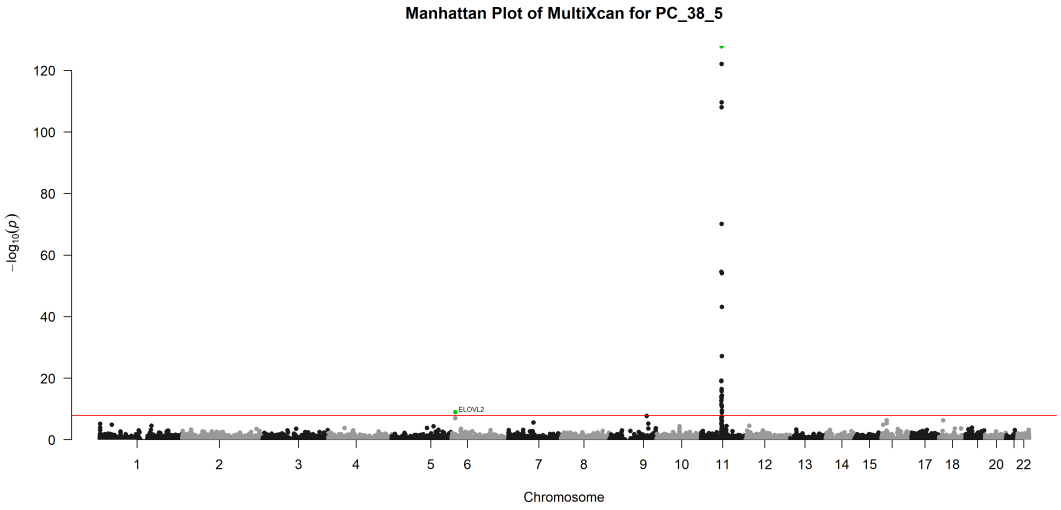
Phosphatidylcholine diacyl C38:5 (µg/dL)



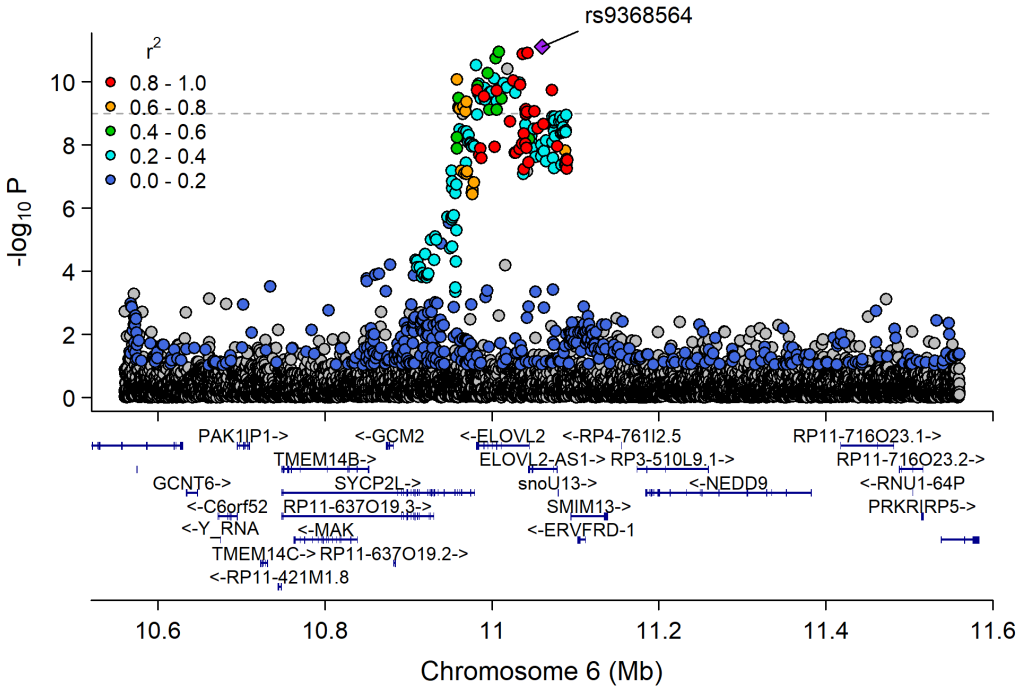




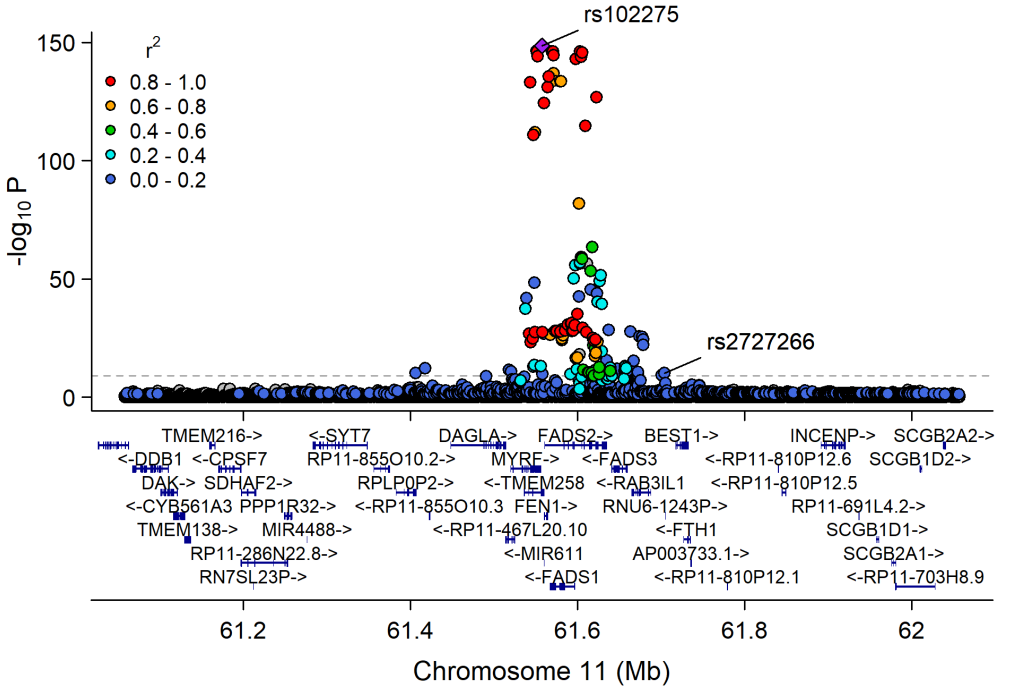




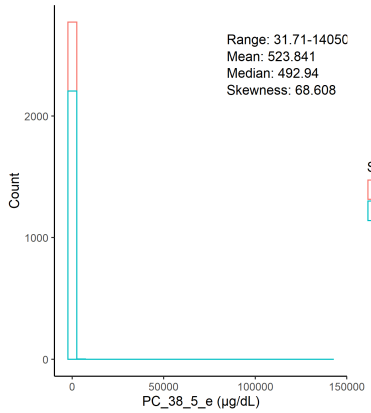
Chr6_10903850_12000017

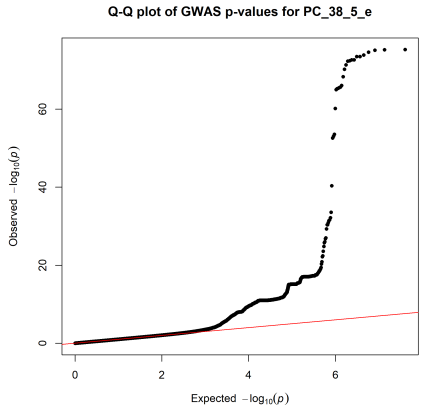


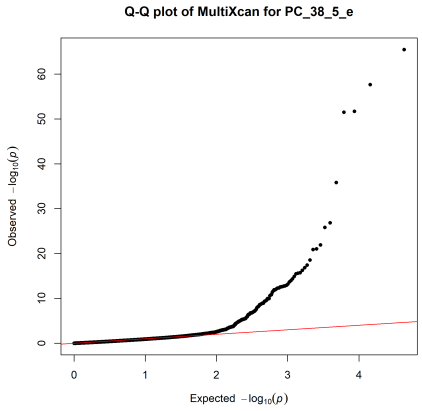
Chr11_59978355_62914375

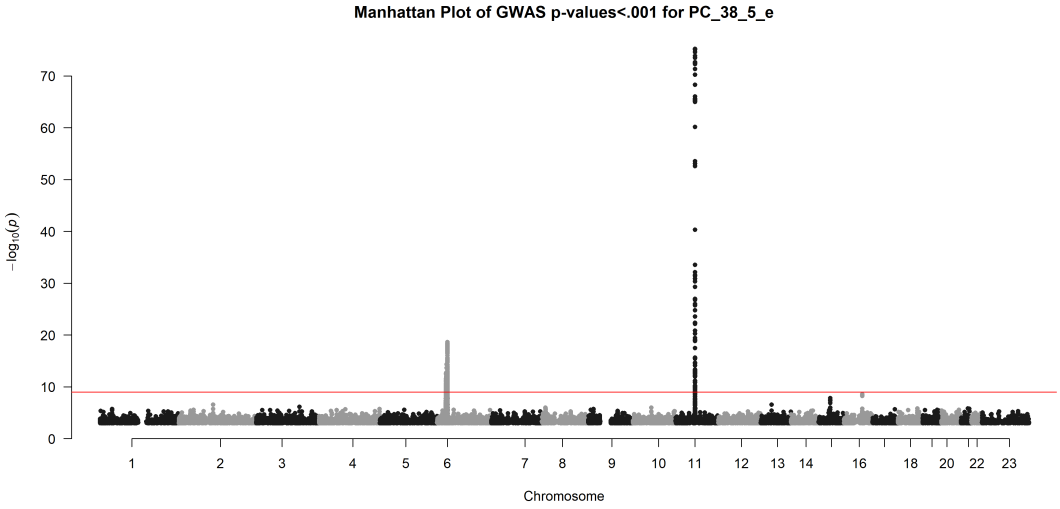


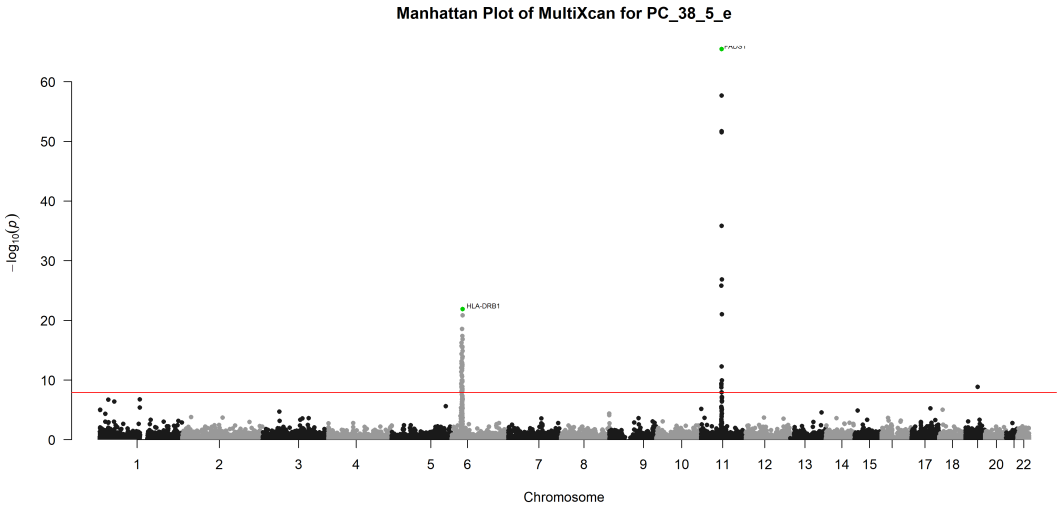
Phosphatidylcholine acyl-alkyl C38:5 (µg/dL)



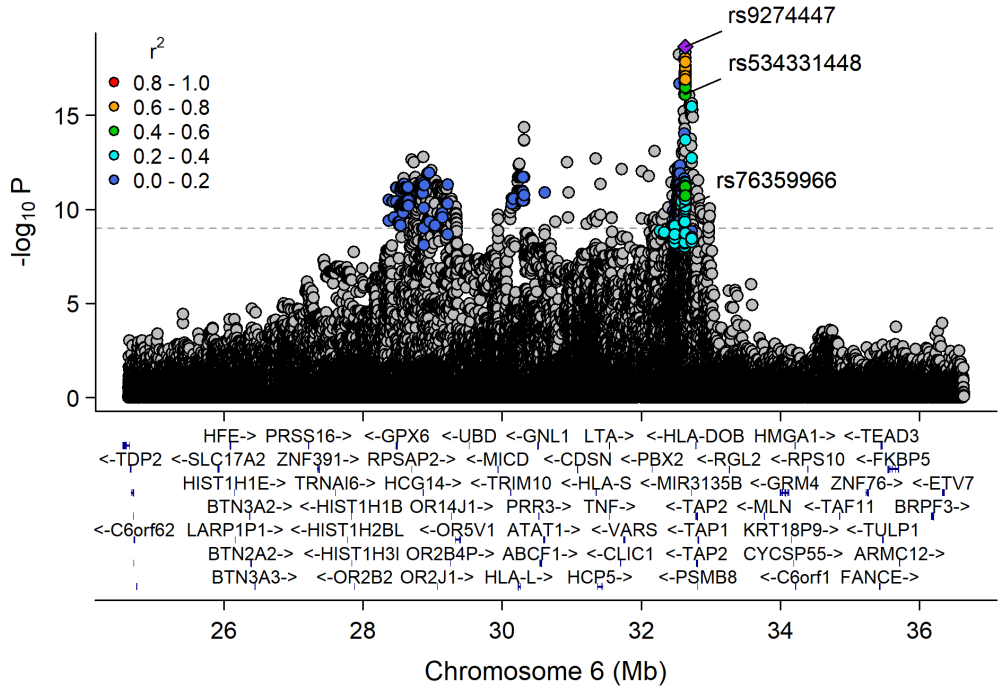




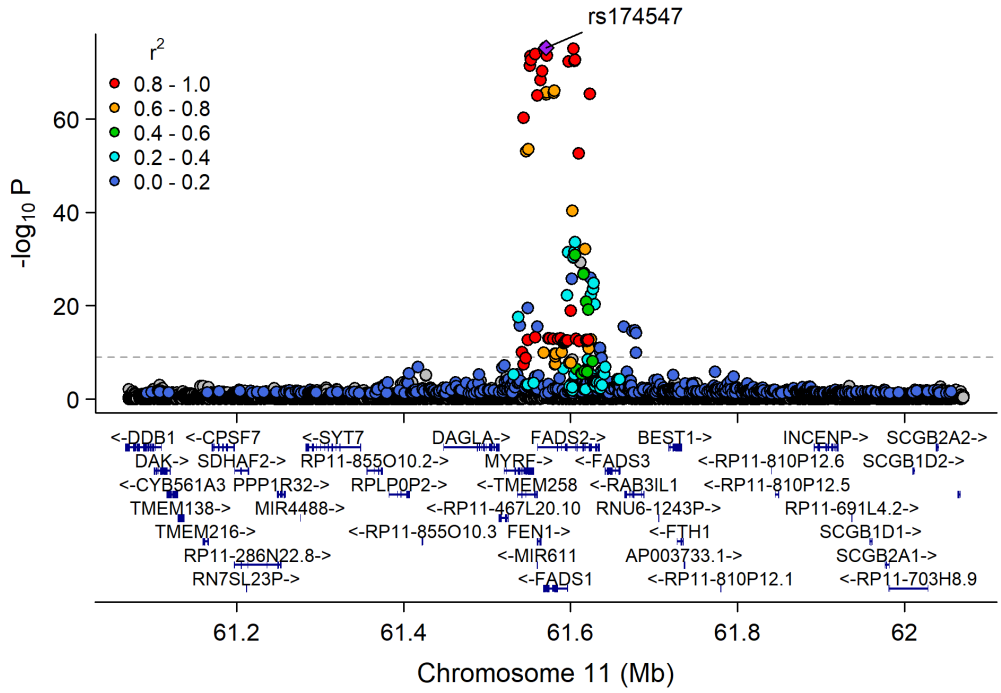




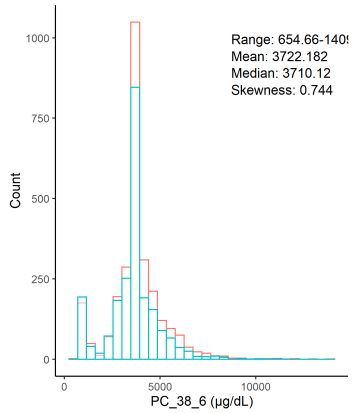
Chr6_24042708_34003583

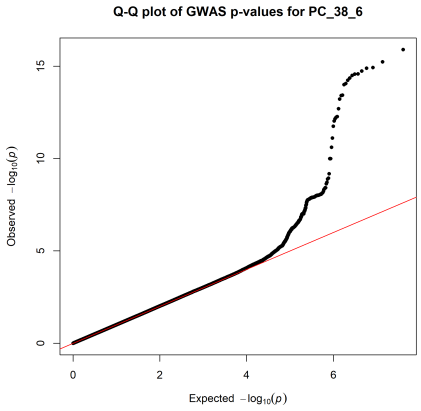


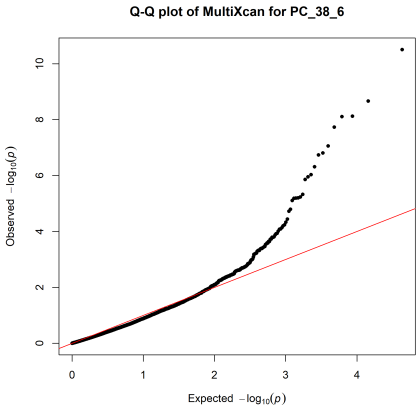
Chr11_59978355_62914375

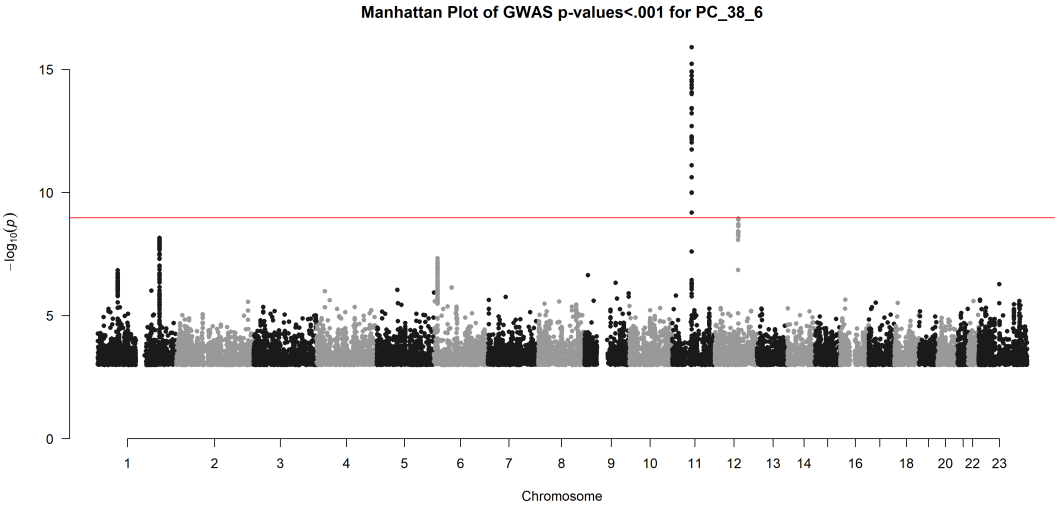


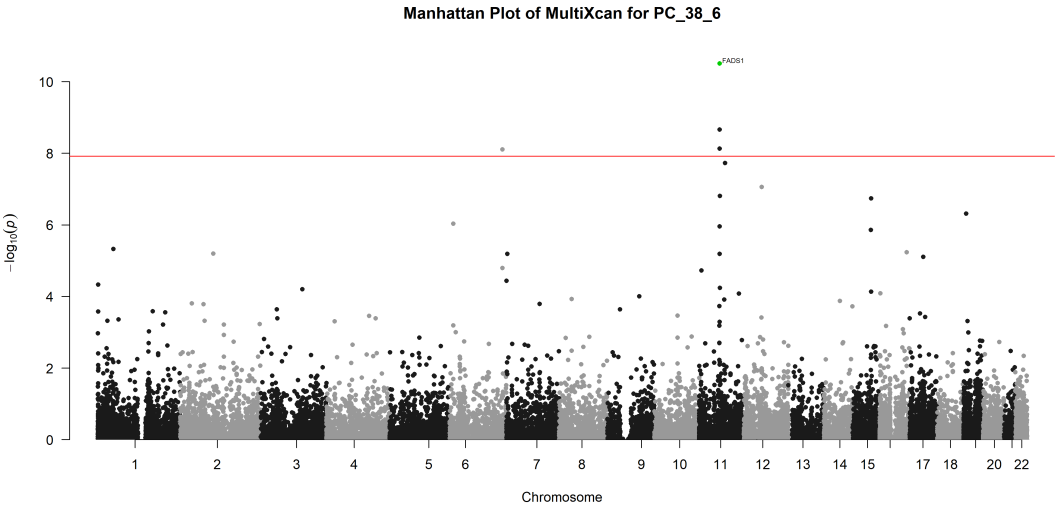
Phosphatidylcholine diacyl C38:6 (µg/dL)



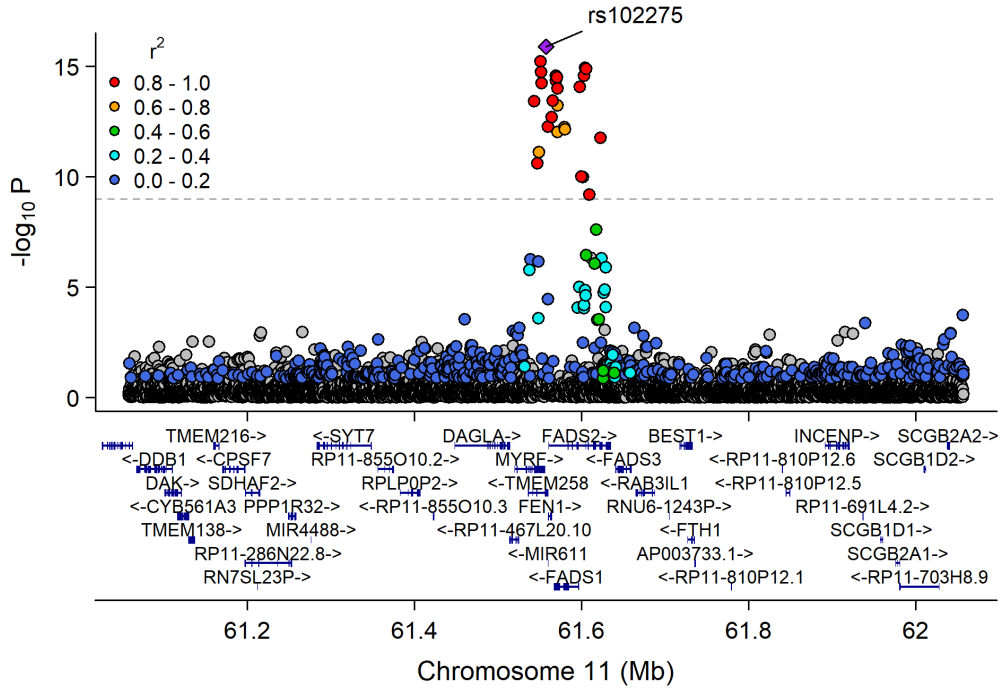




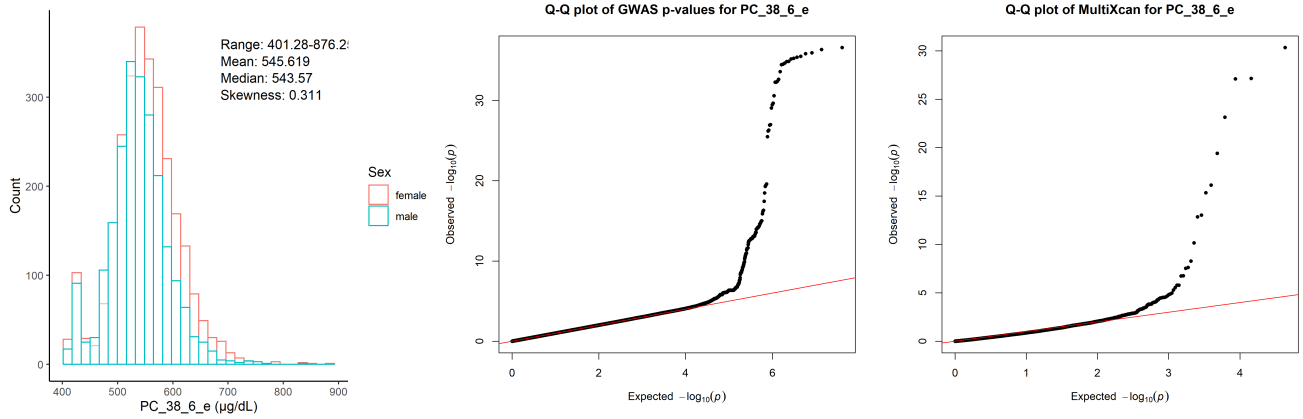


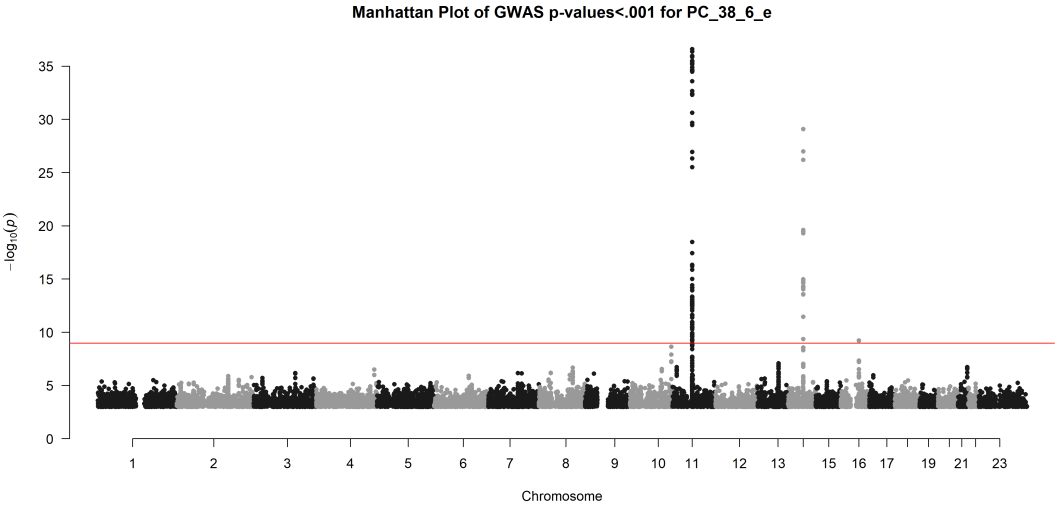


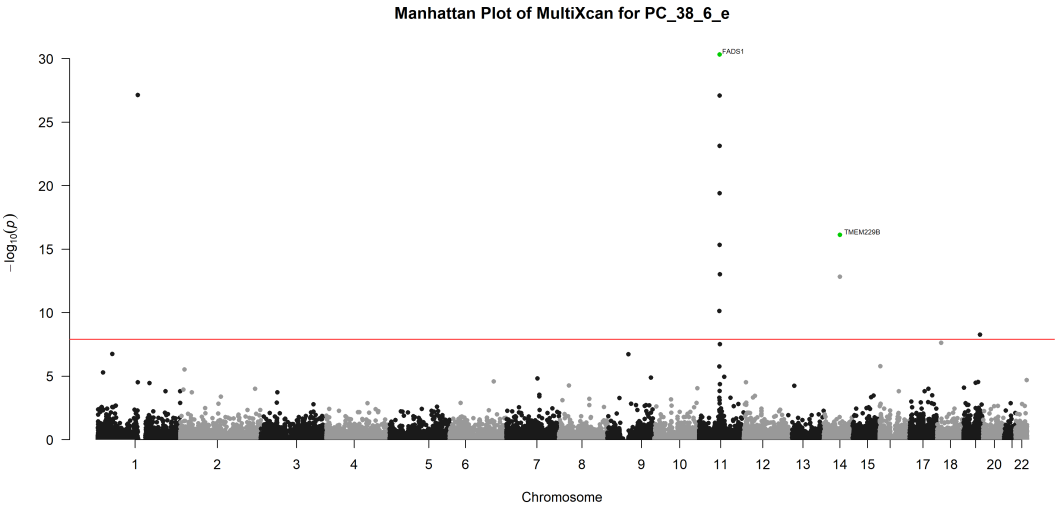
Chr11_59978355_62914375



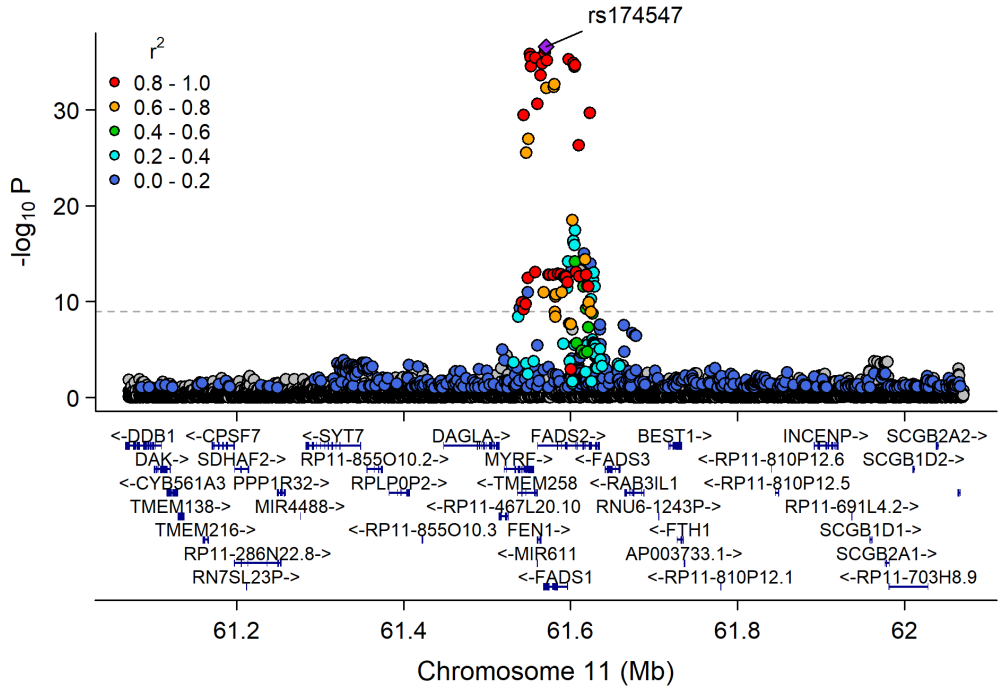
Phosphatidylcholine acyl-alkyl C38:6 (µg/dL)



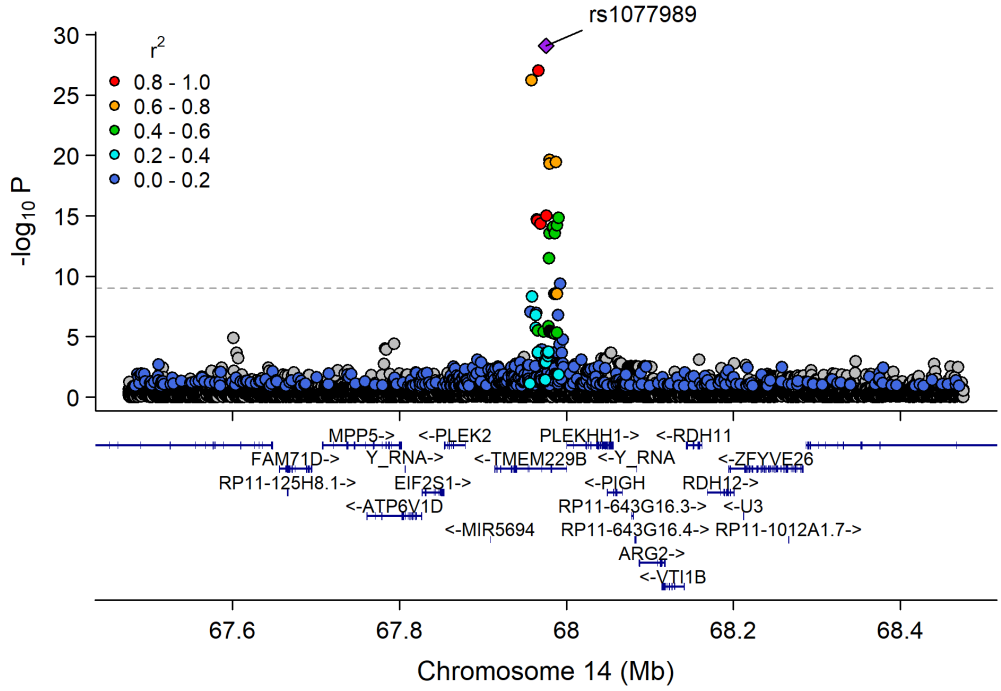




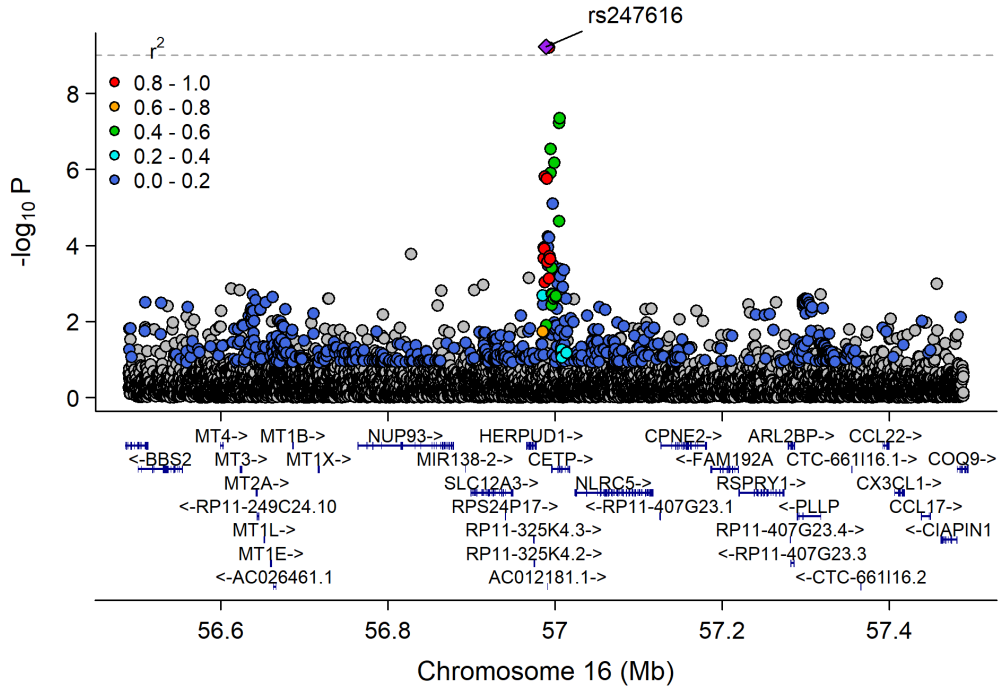
Chr11_59978355_62914375



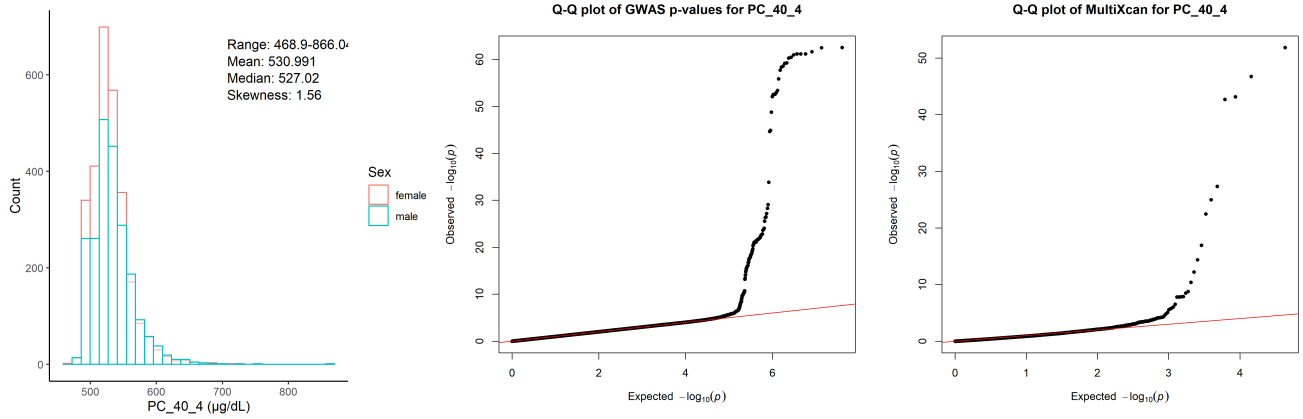
Chr14_66997265_68972354

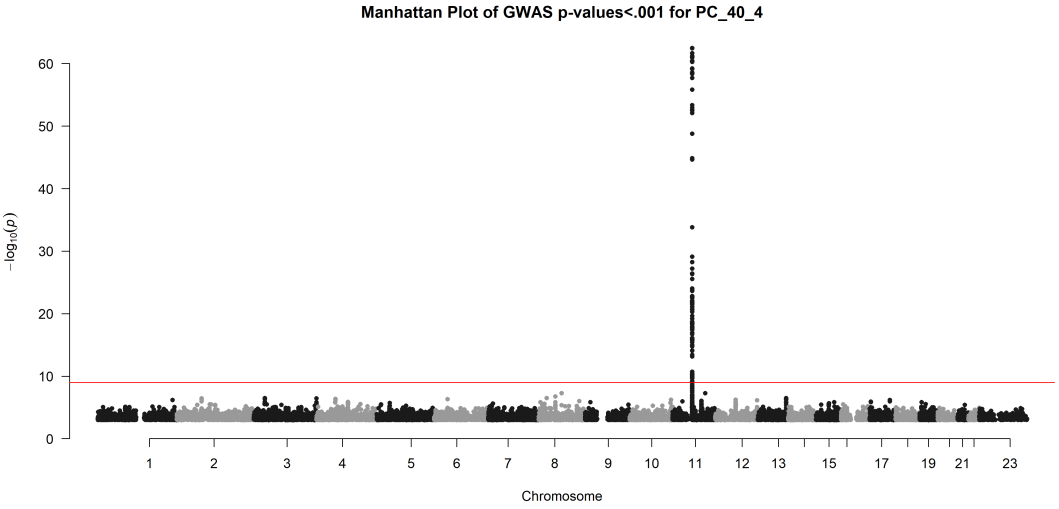


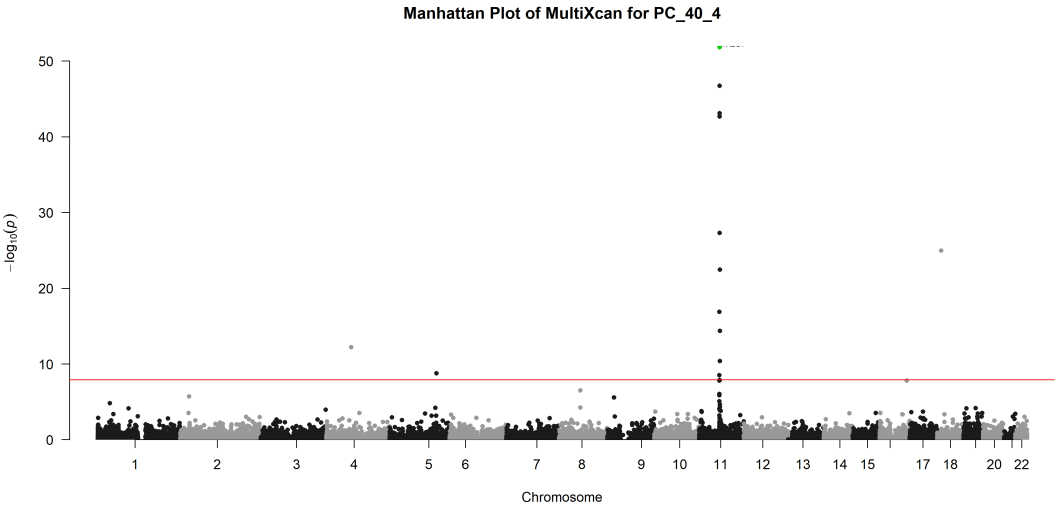
Chr16_55870822_57992421



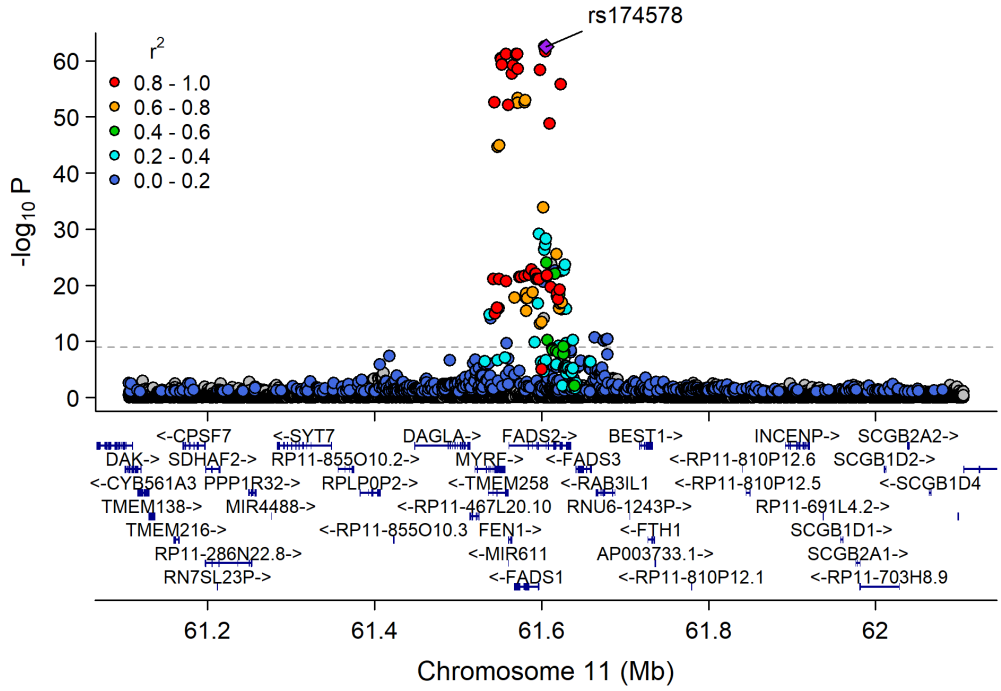
Phosphatidylcholine diacyl C40:4 (µg/dL)



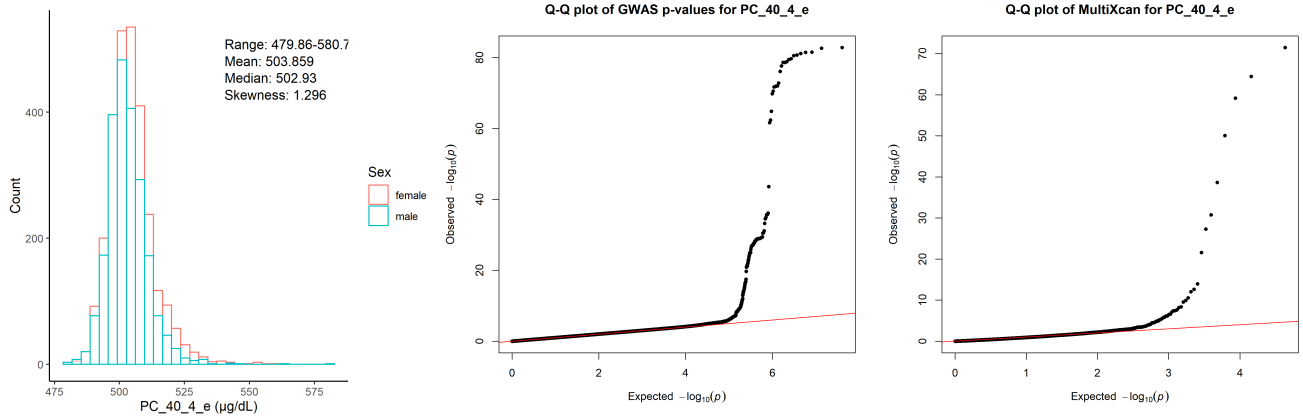


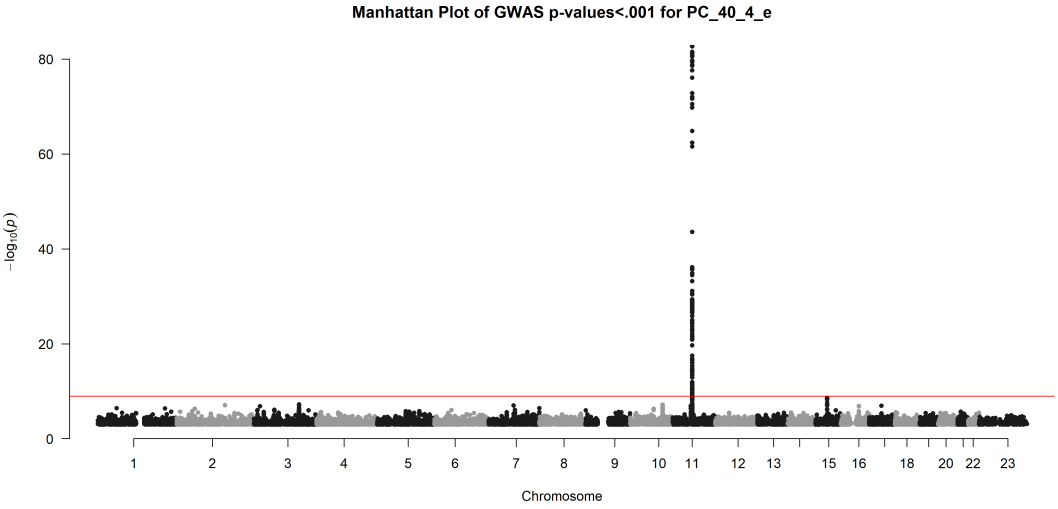


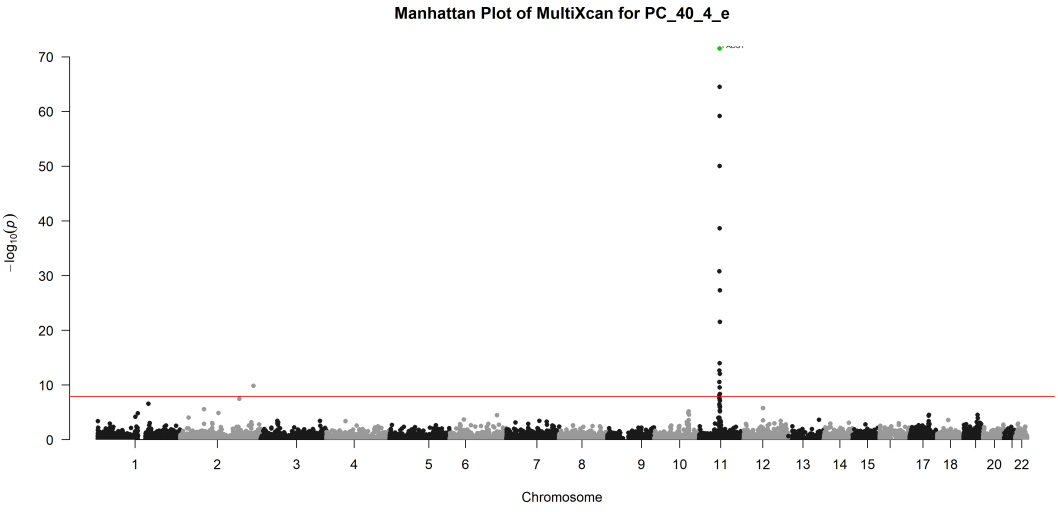
Chr11_59978355_62914375



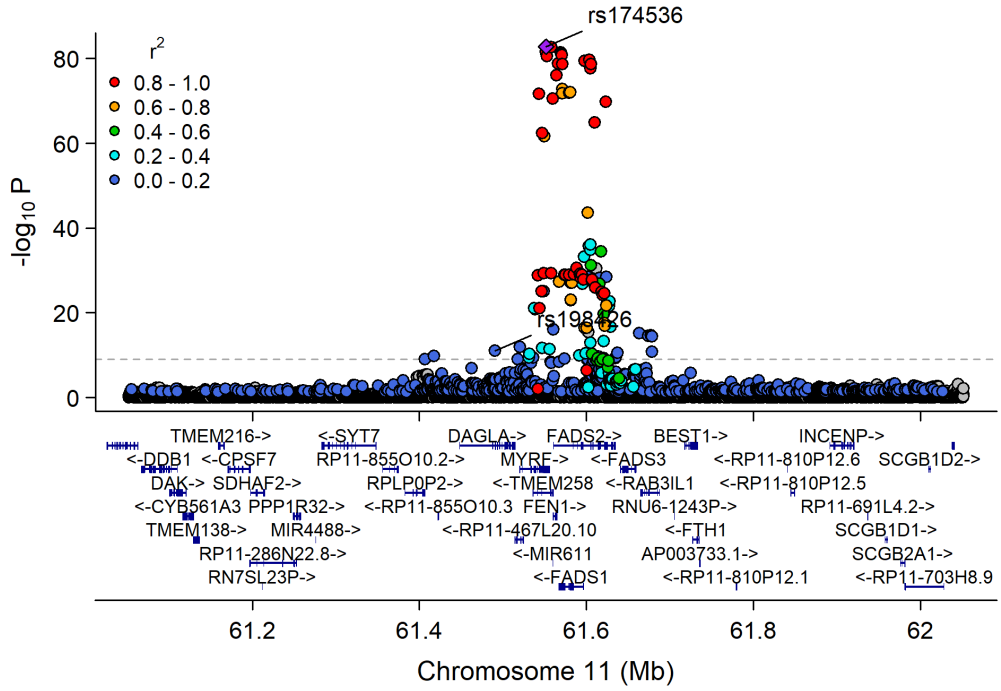
Phosphatidylcholine acyl-alkyl C40:4 (µg/dL)



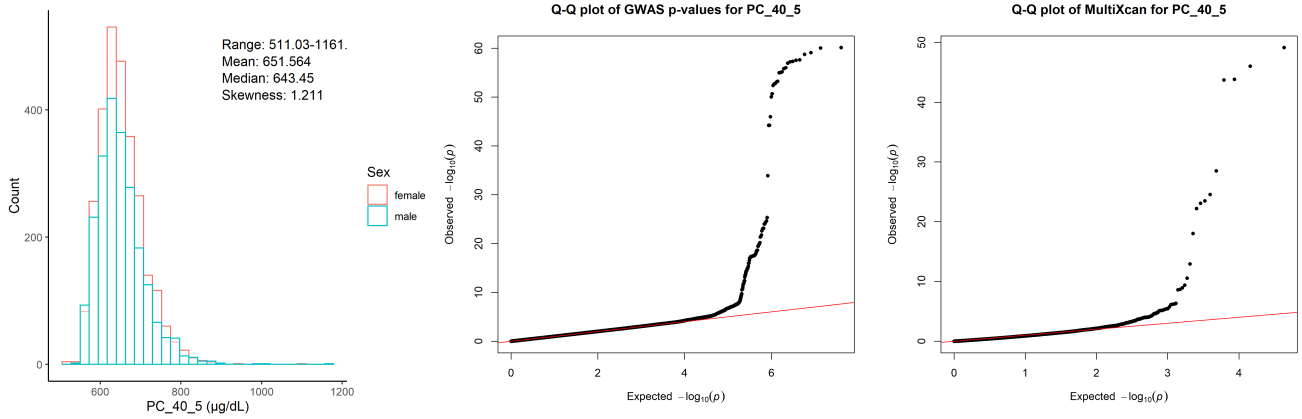


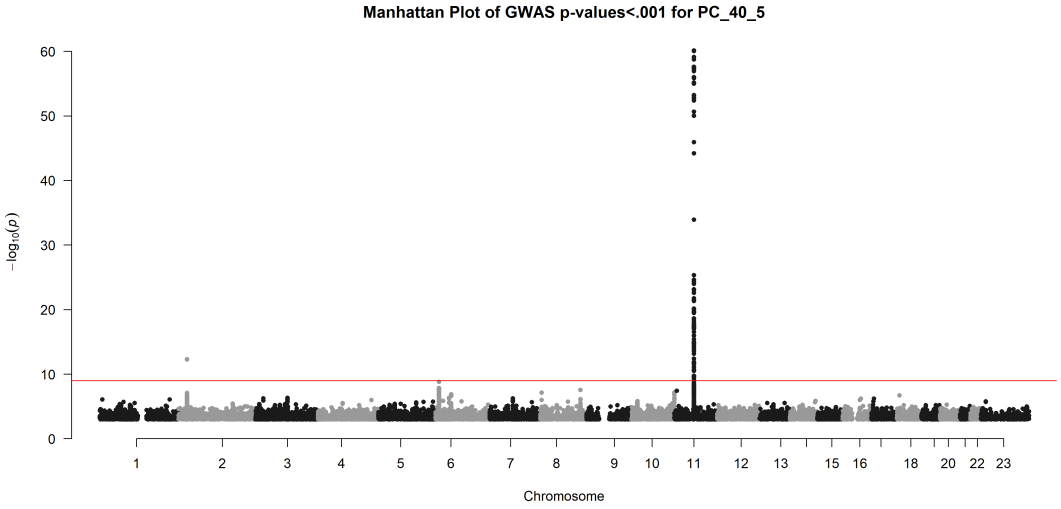


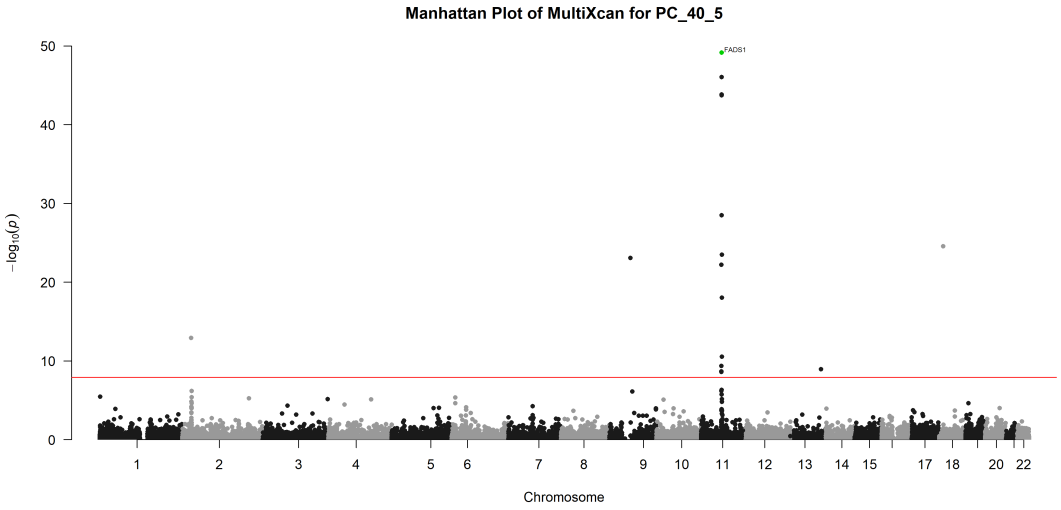
Chr11_59978355_62914375



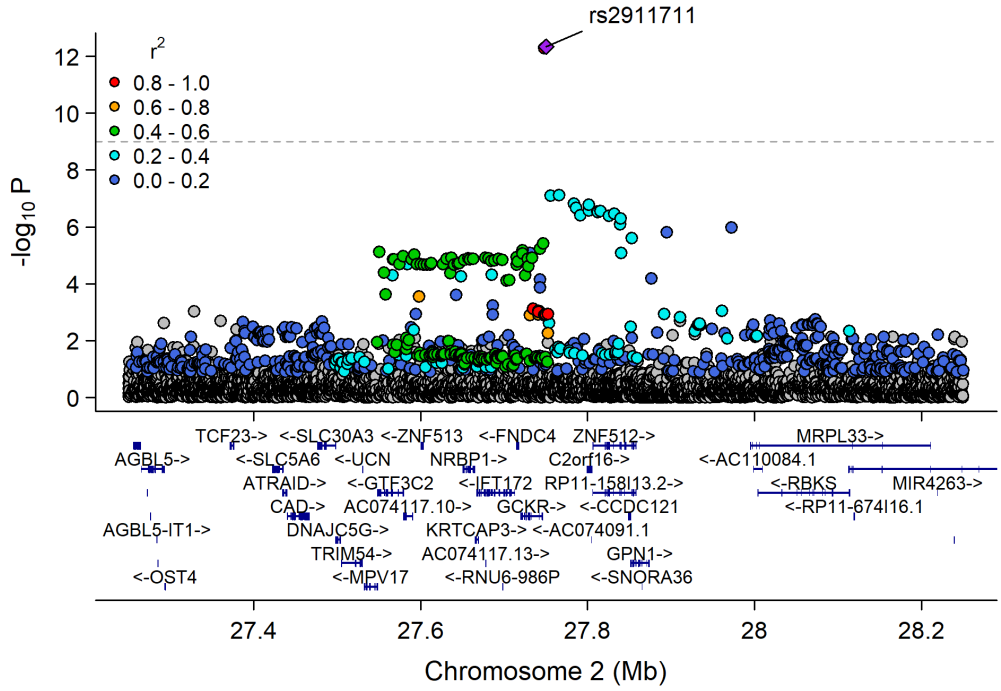
Phosphatidylcholine diacyl C40:5 (µg/dL)



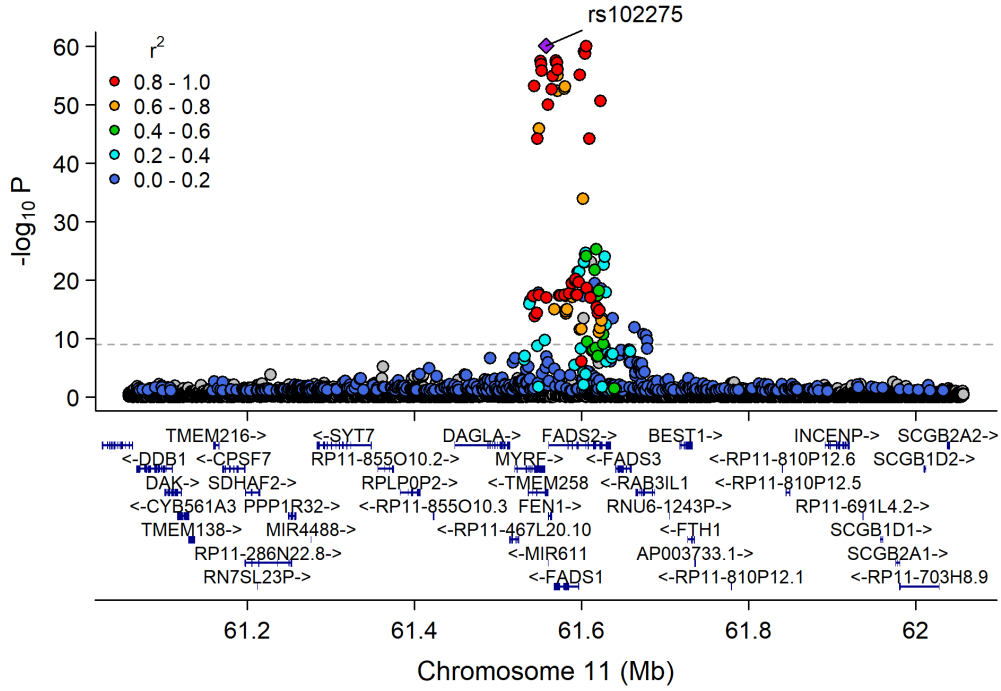




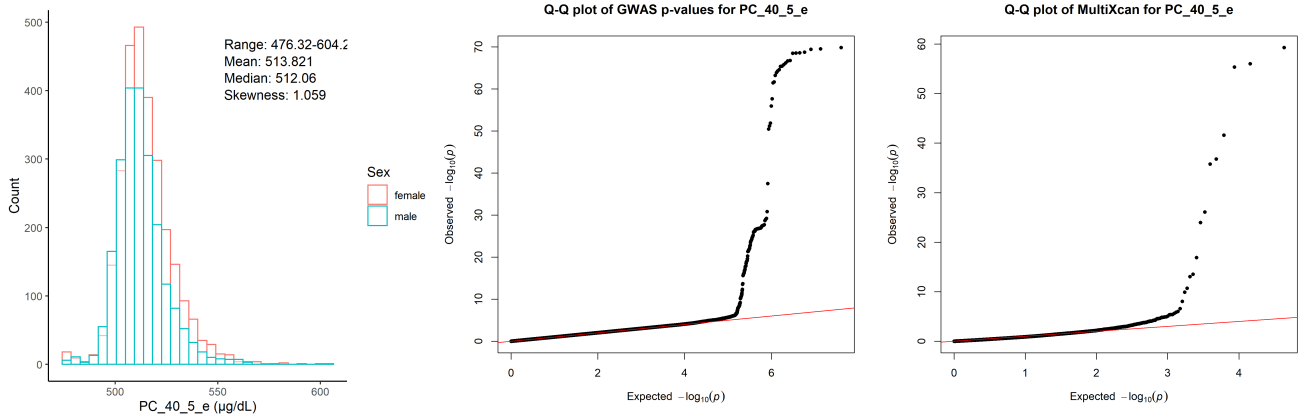
Chr2_26753815_28597624

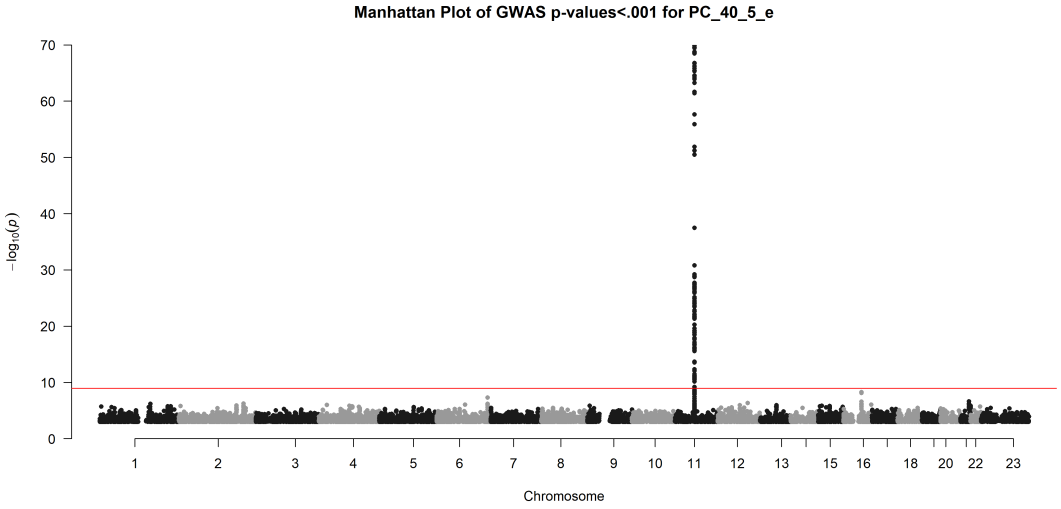


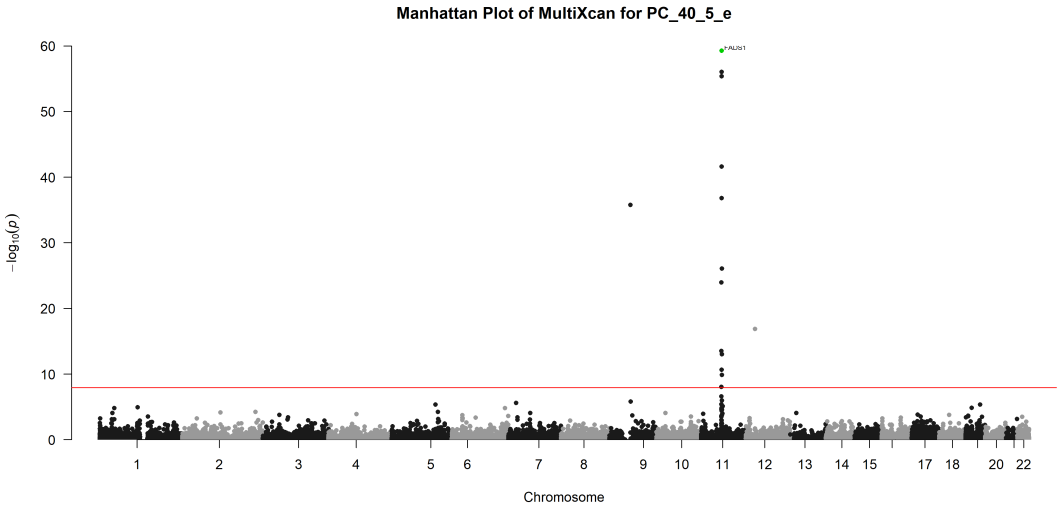
Chr11_59978355_62914375



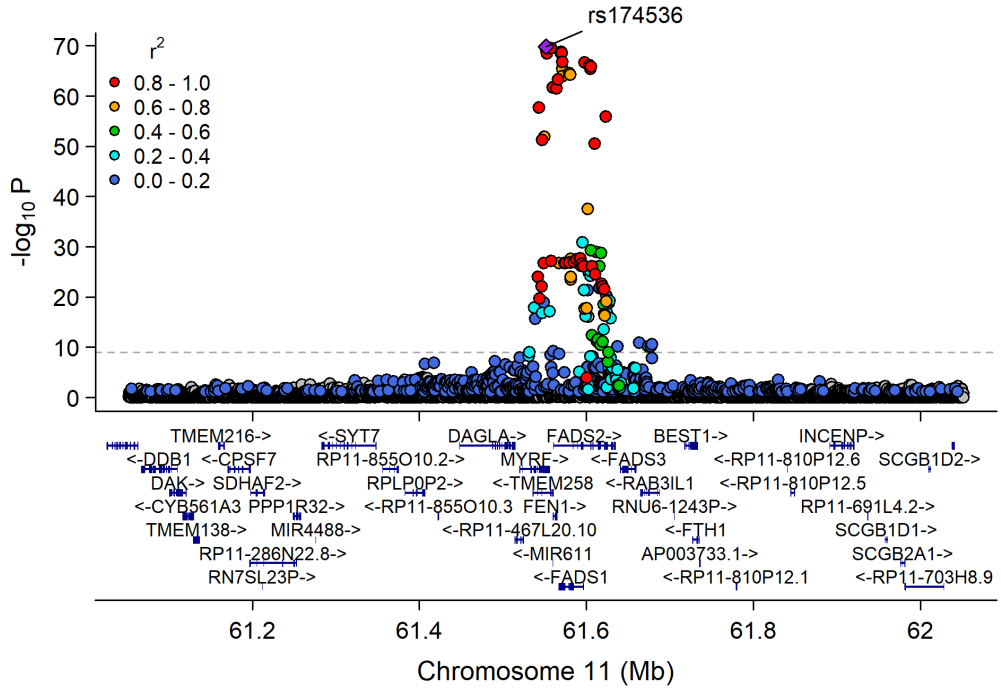
Phosphatidylcholine acyl-alkyl C40:5 (µg/dL)



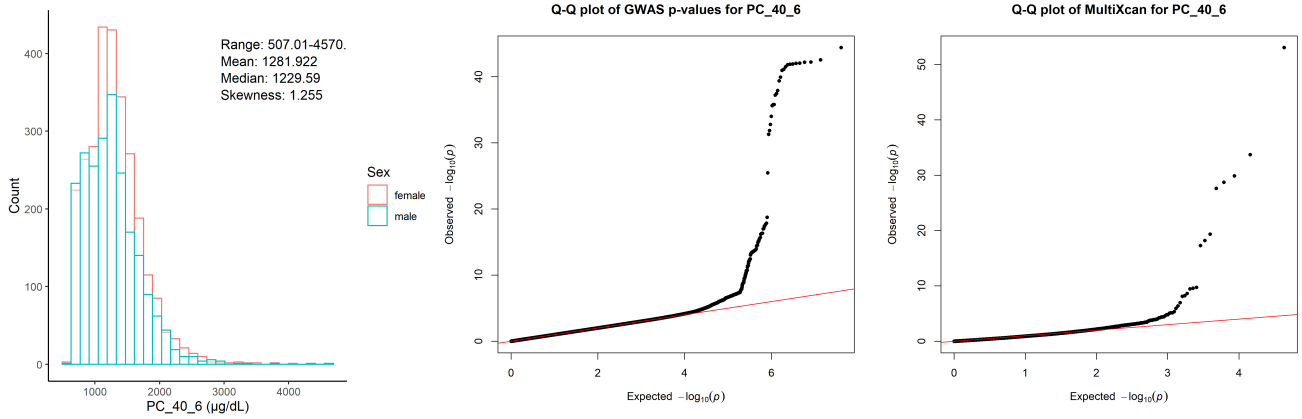


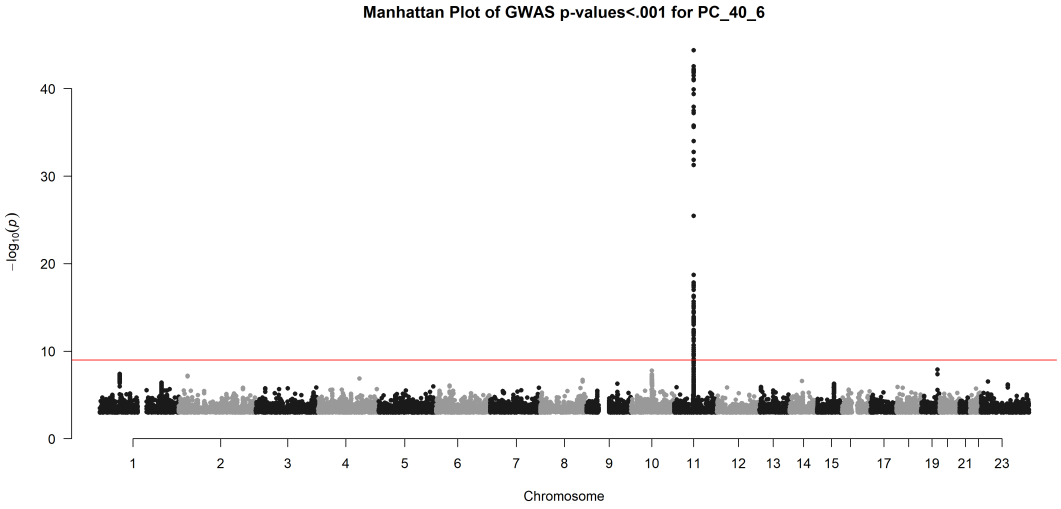


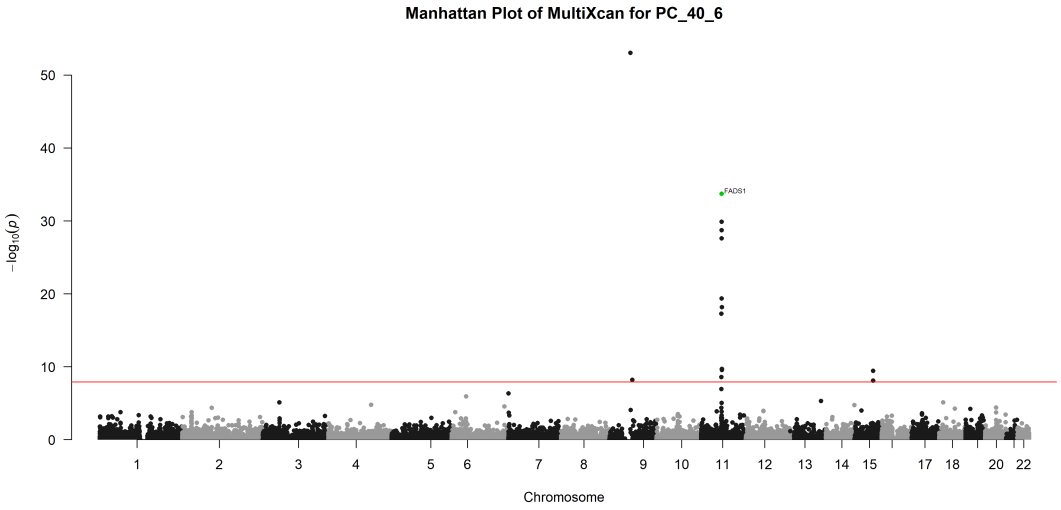
Chr11__59978355__62914375



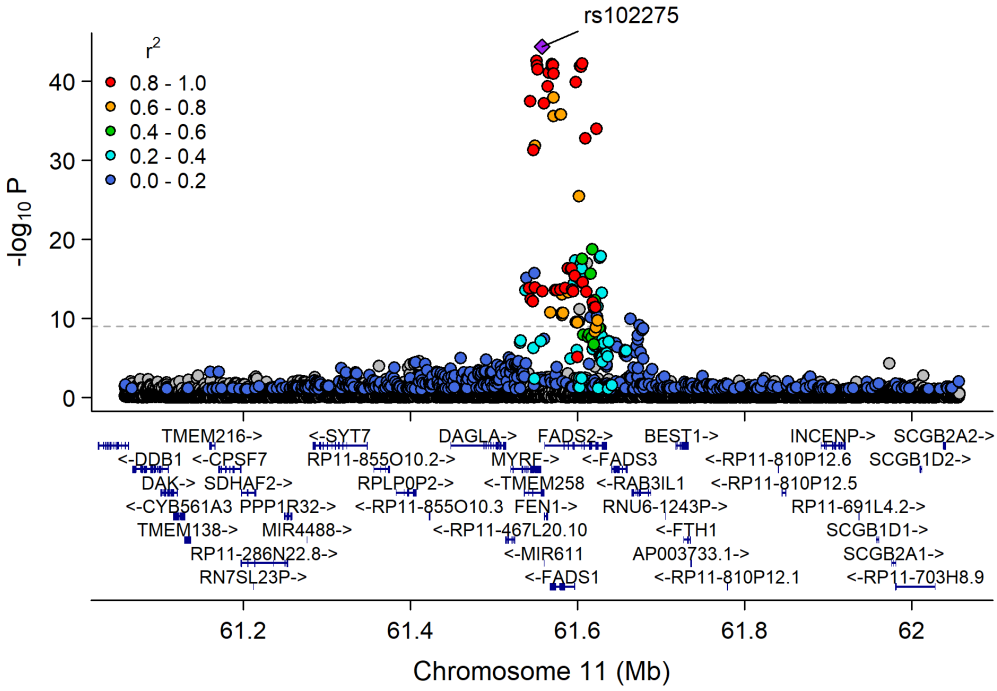
Phosphatidylcholine diacyl C40:6 (µg/dL)



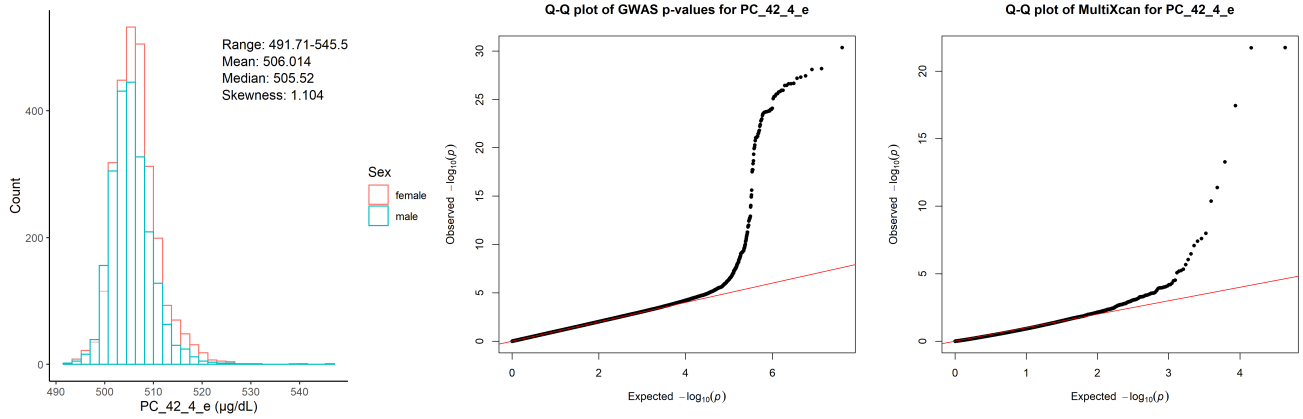


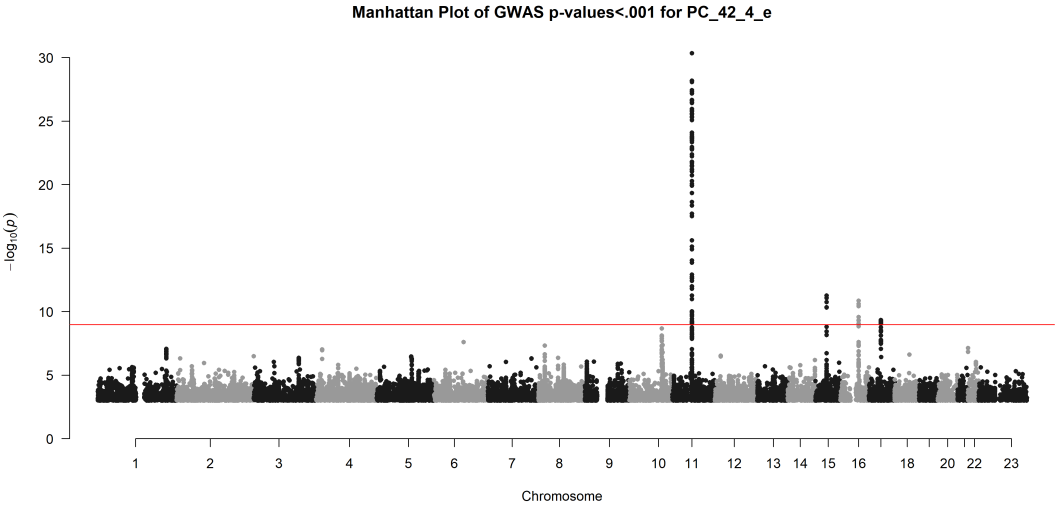


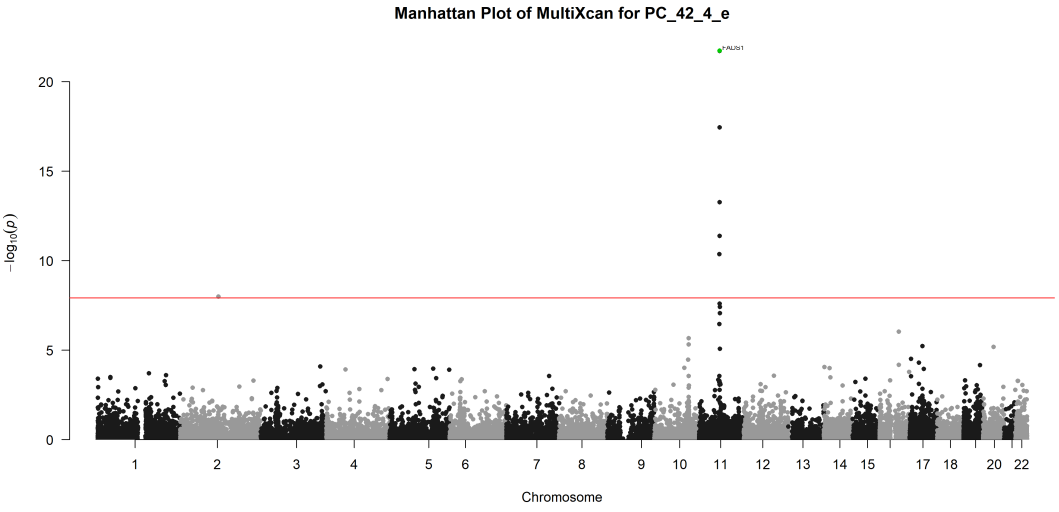
Chr11__59978355__62914375



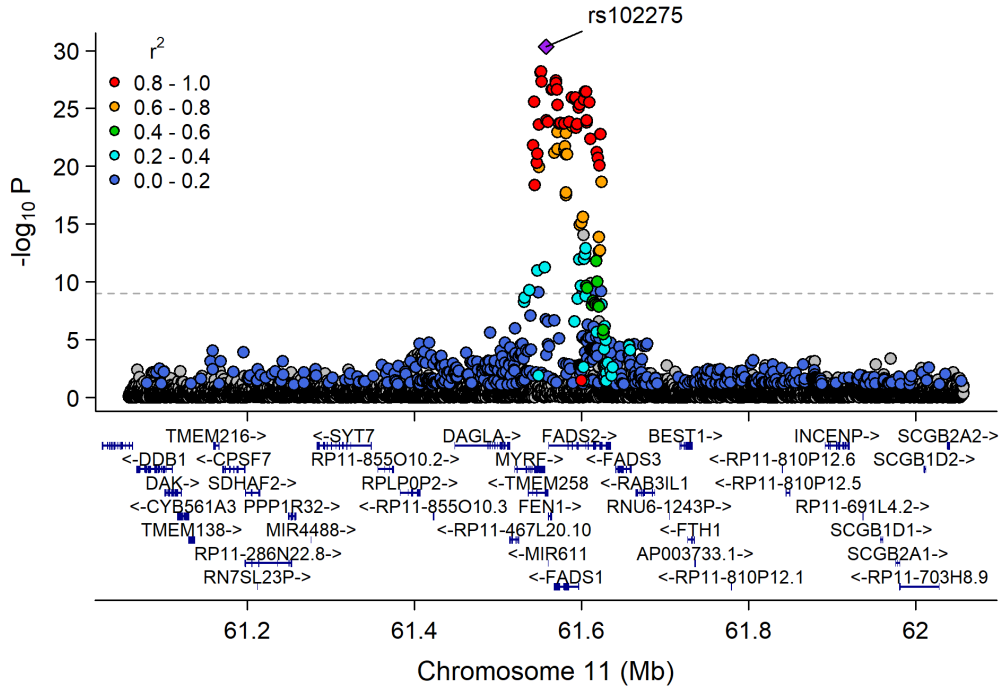
Phosphatidylcholine acyl-alkyl C42:4 (µg/dL)



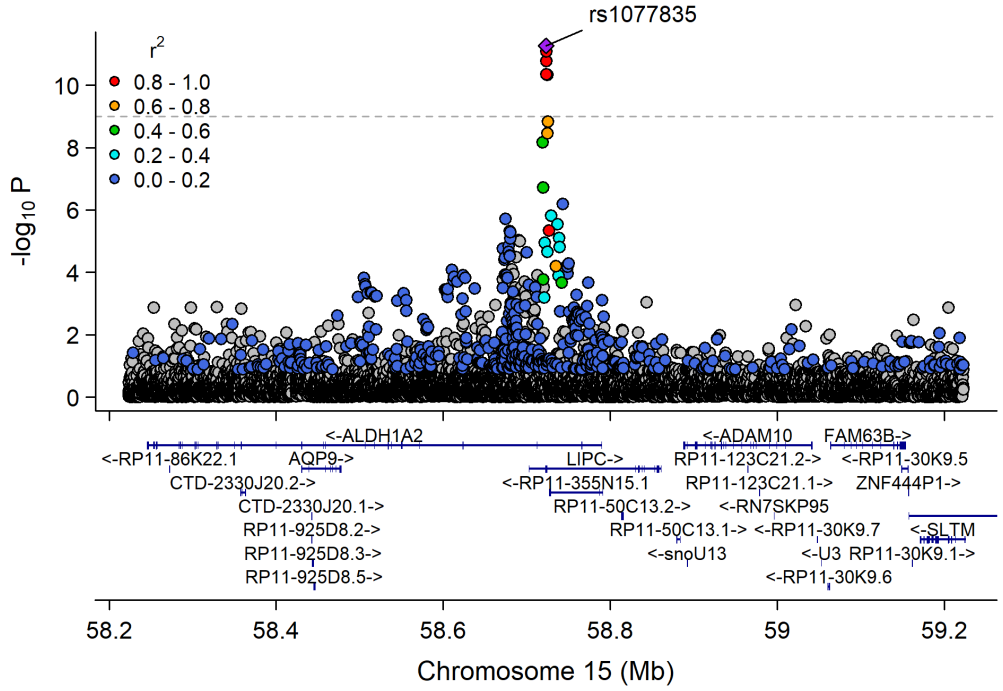




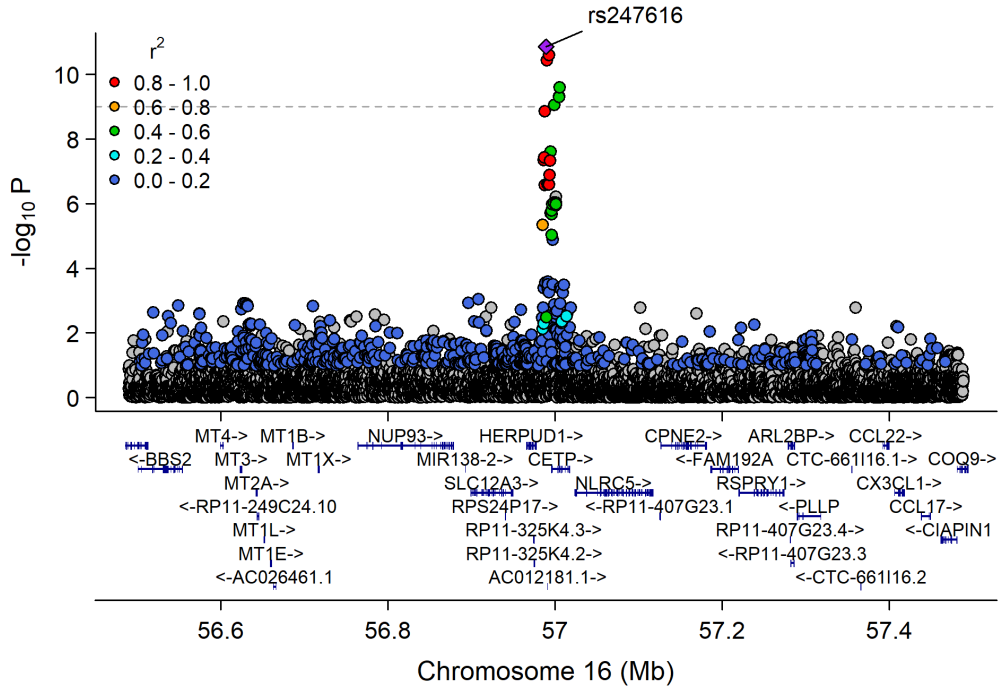
Chr11_59978355_62914375



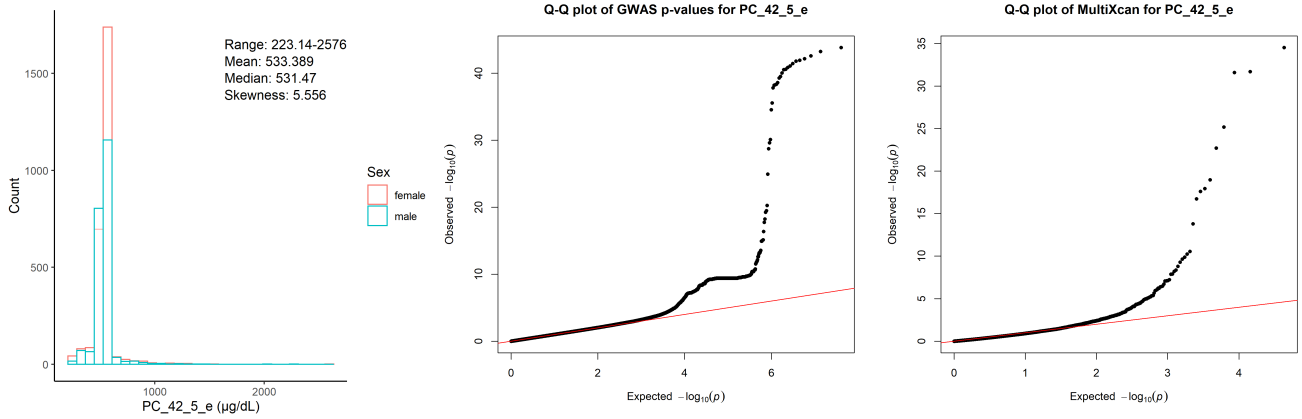
Chr15_57658798_60490883

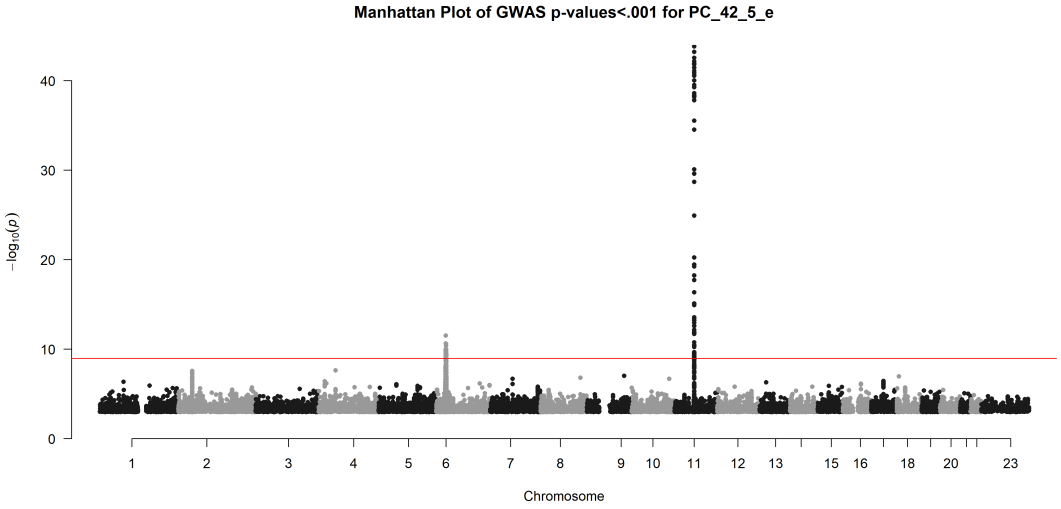


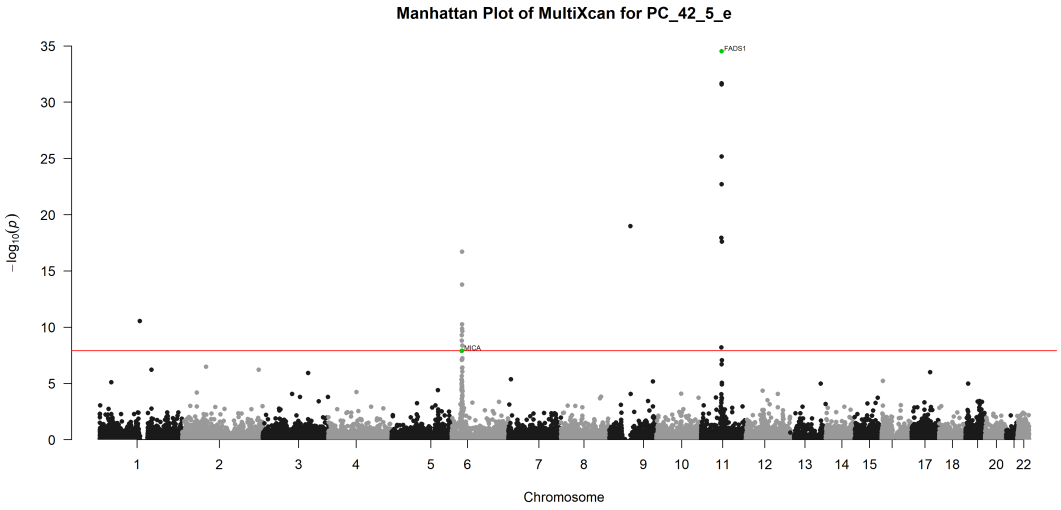
Chr16_55870822_57992421



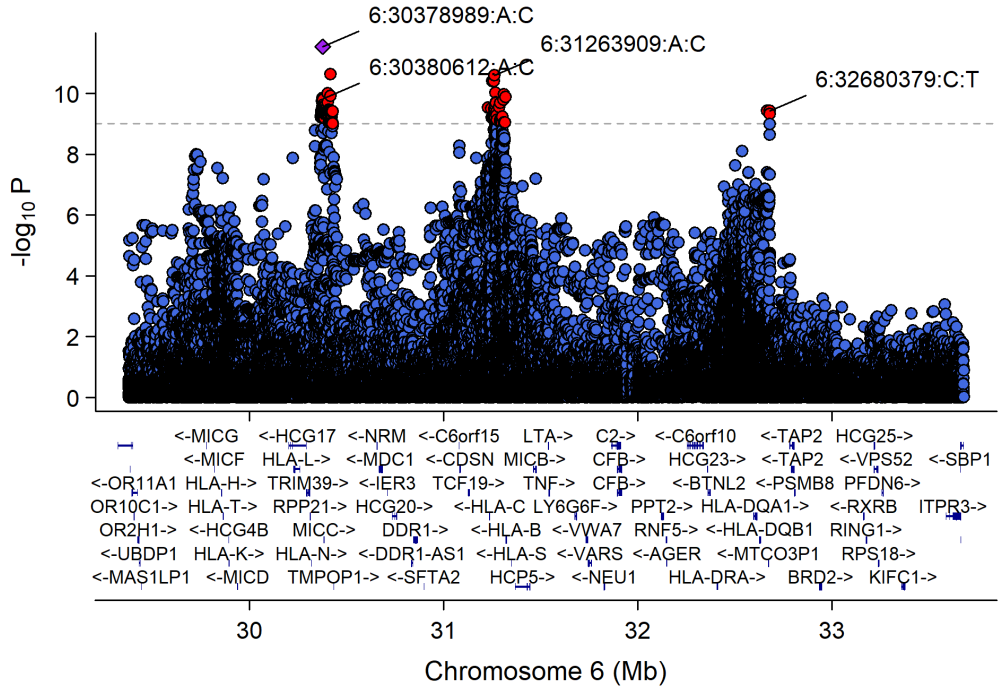
Phosphatidylcholine acyl-alkyl C42:5 (µg/dL)



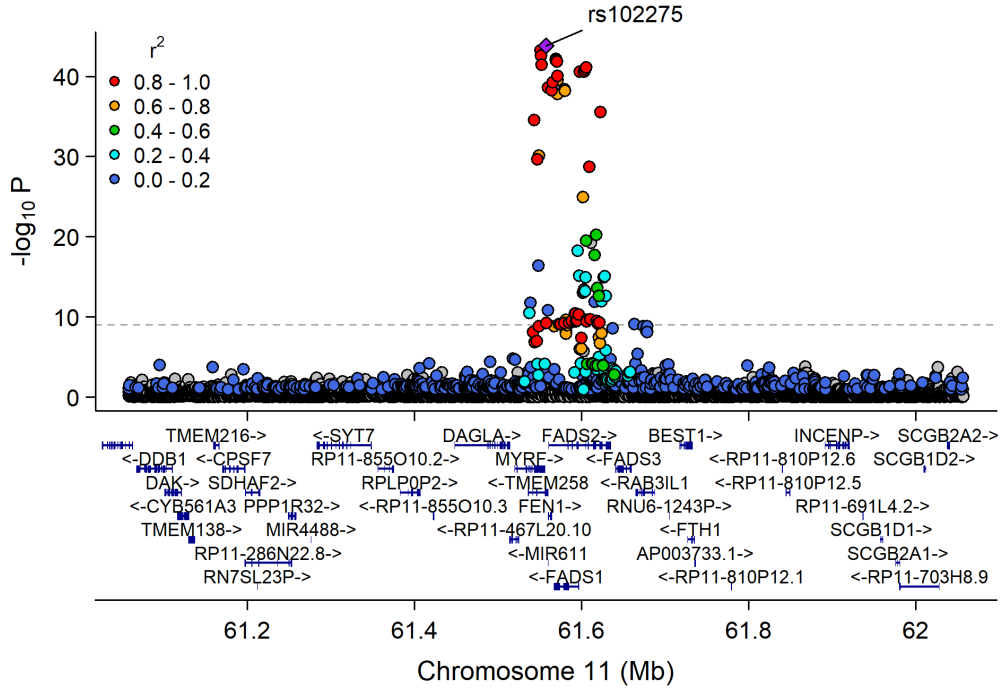




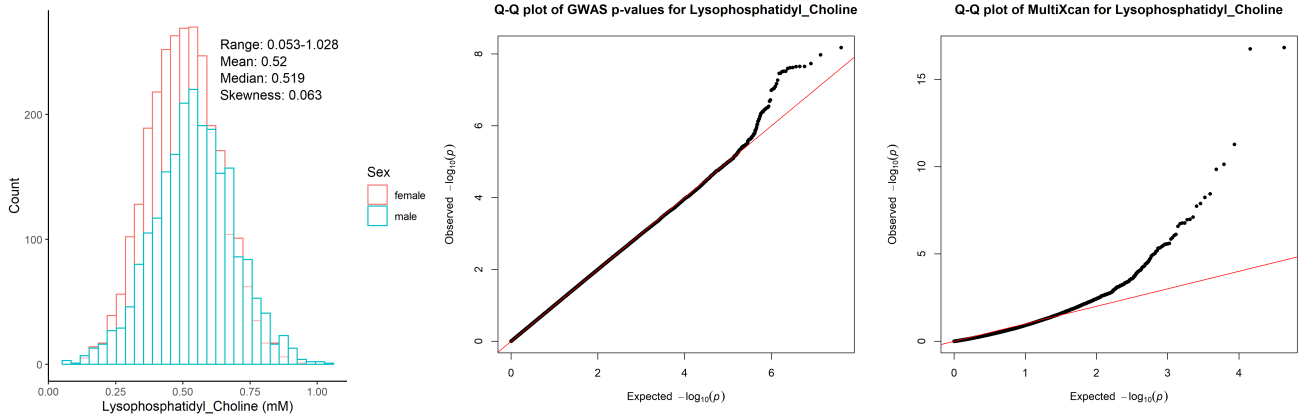
Chr6_24042708_34003583

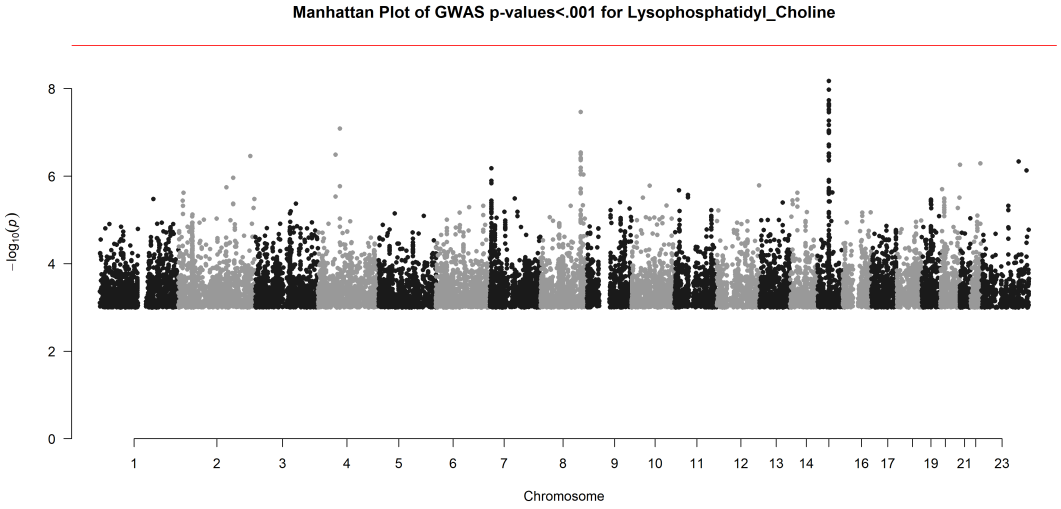


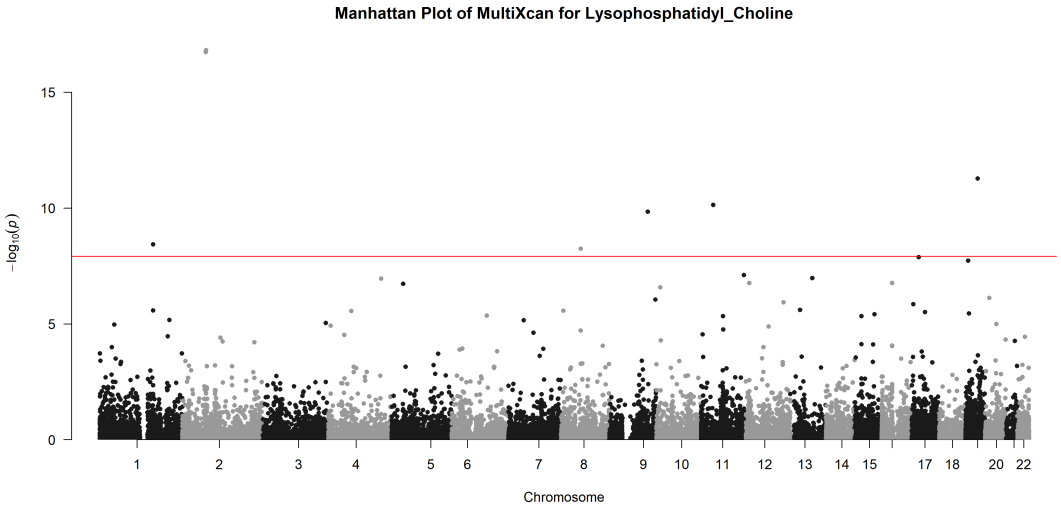
Chr11_59978355_62914375



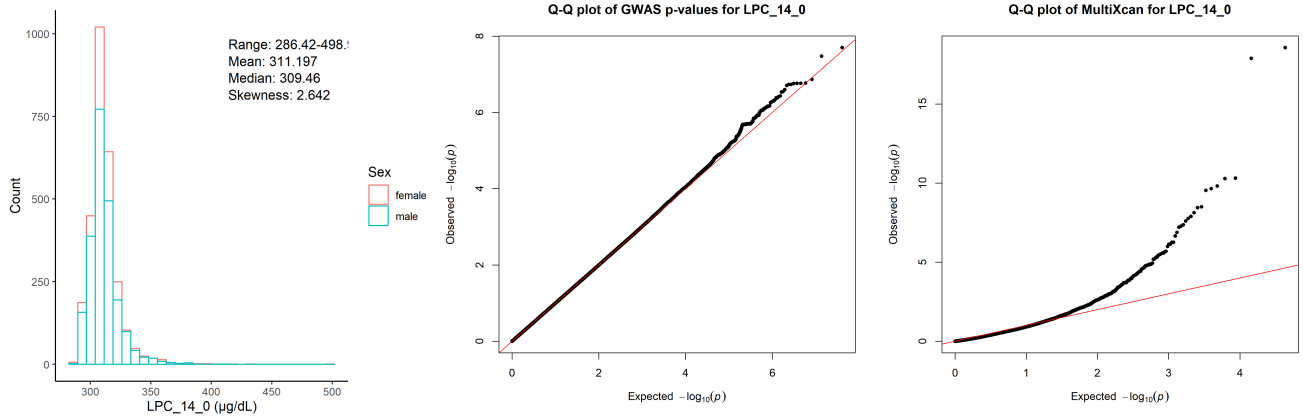
Lysophosphatidylcholine (mM)

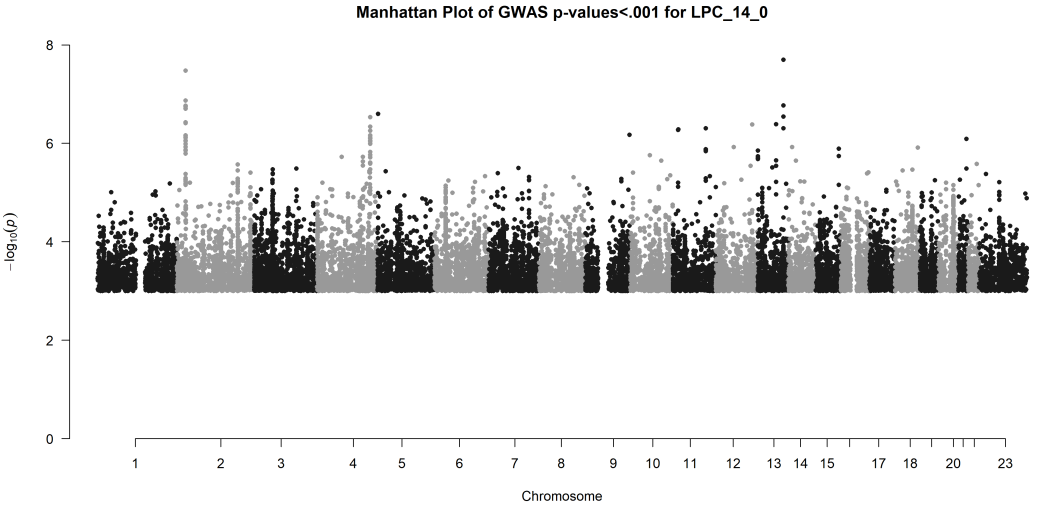


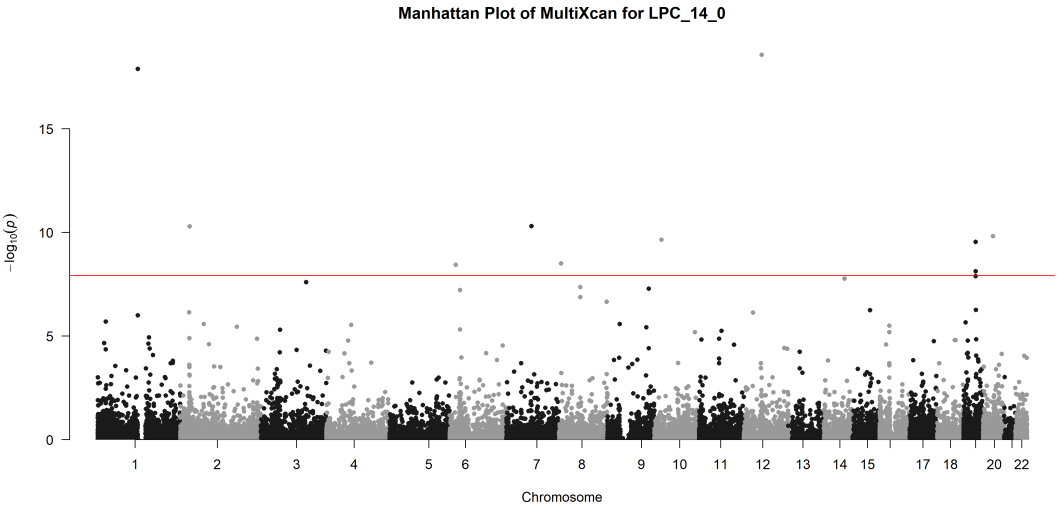




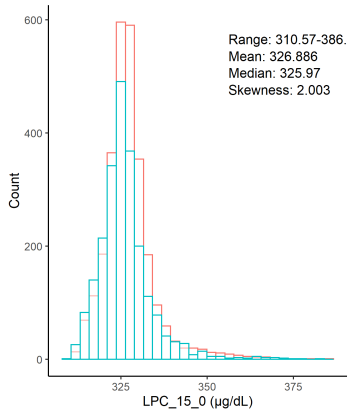
Lysophosphatidylcholine acyl C14:0 (µg/dL)

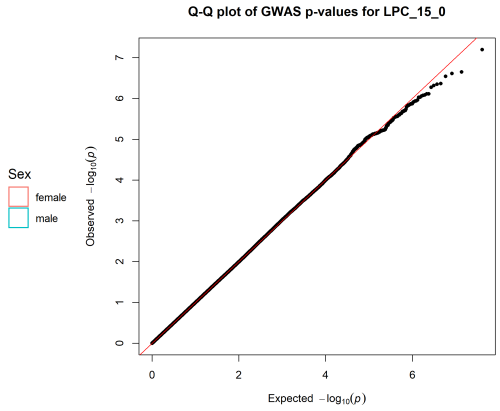


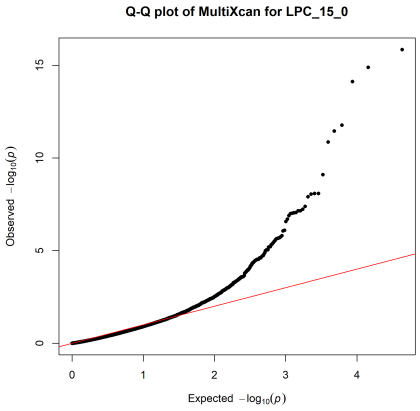


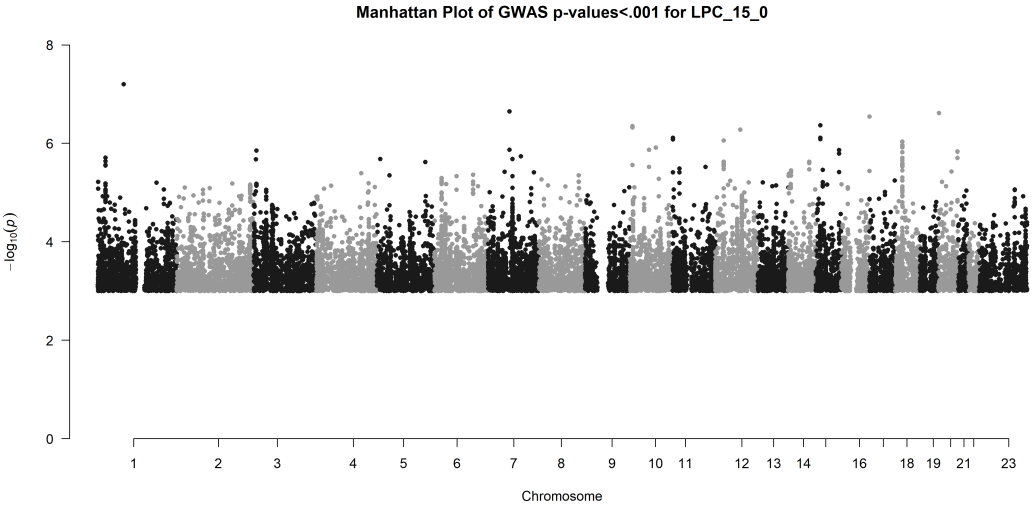


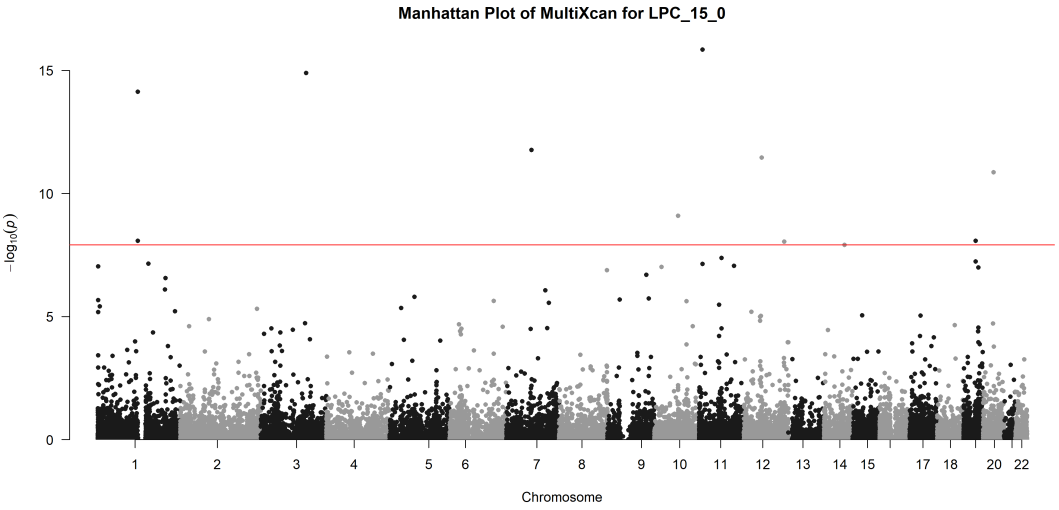
Lysophosphatidylcholine acyl C15:0 (µg/dL)



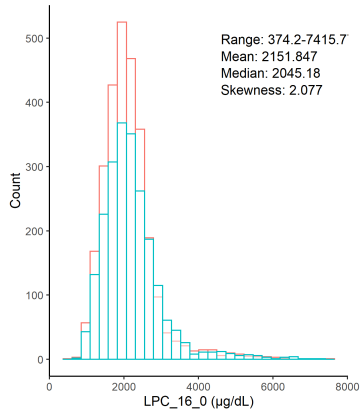


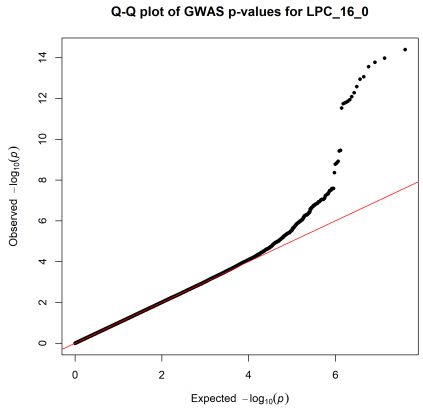


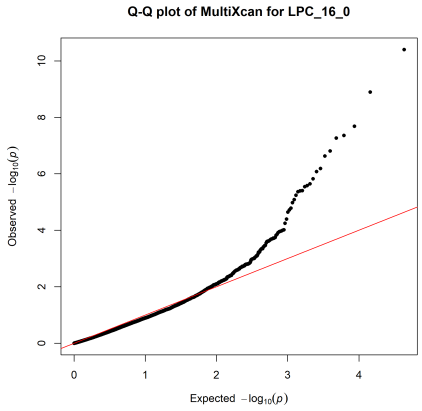


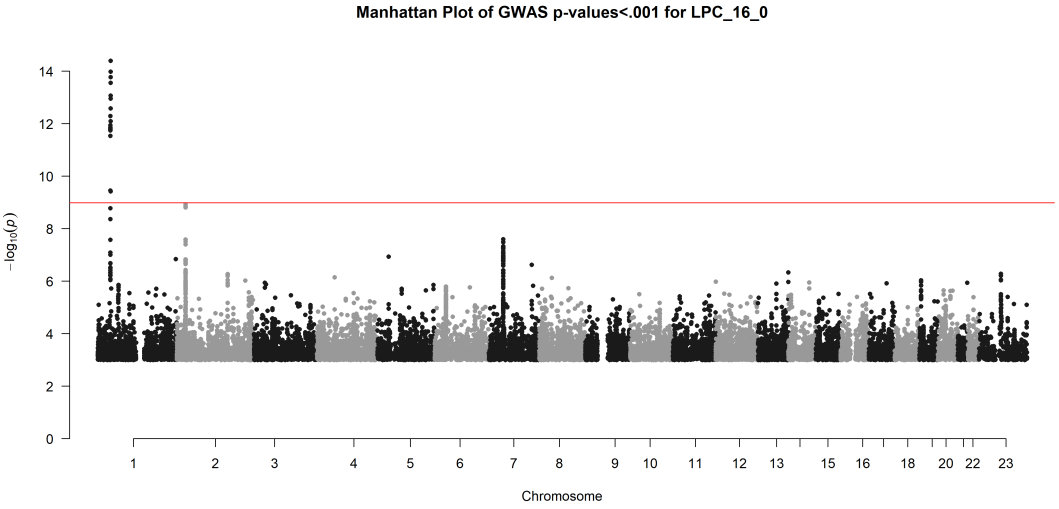


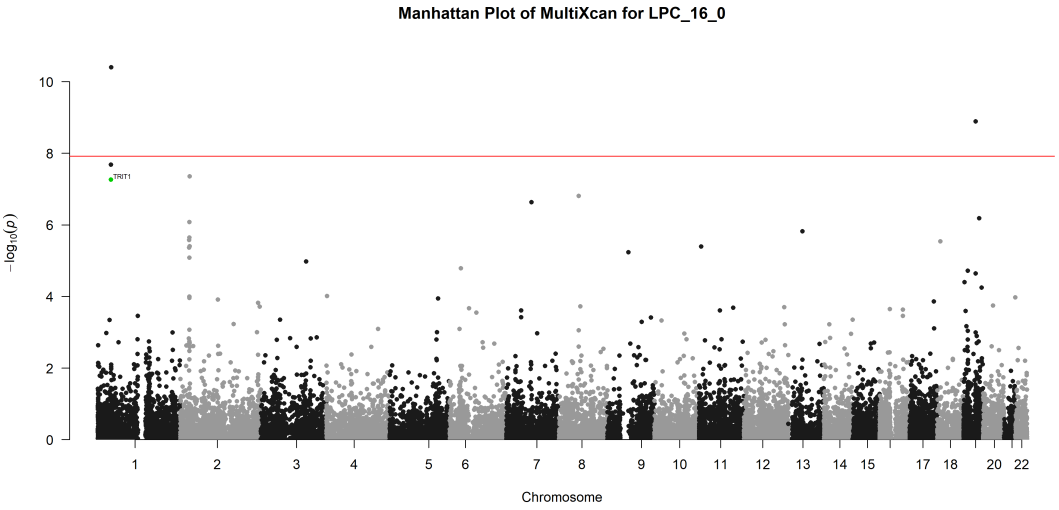
Lysophosphatidylcholine acyl C16:0 (µg/dL)



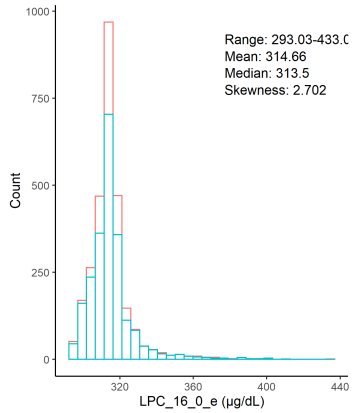


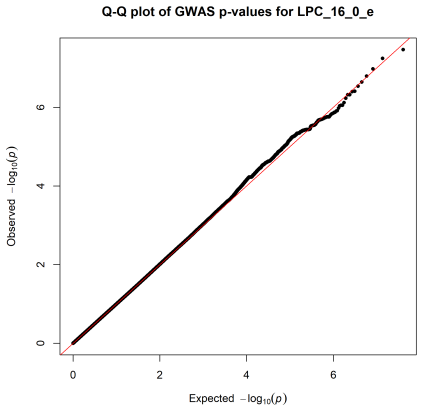


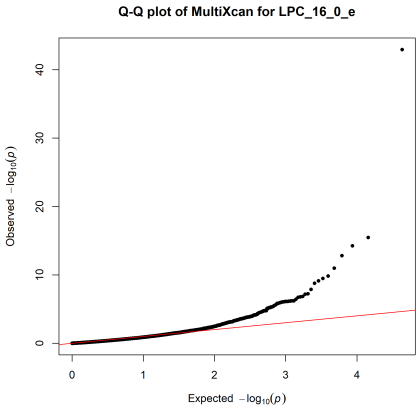


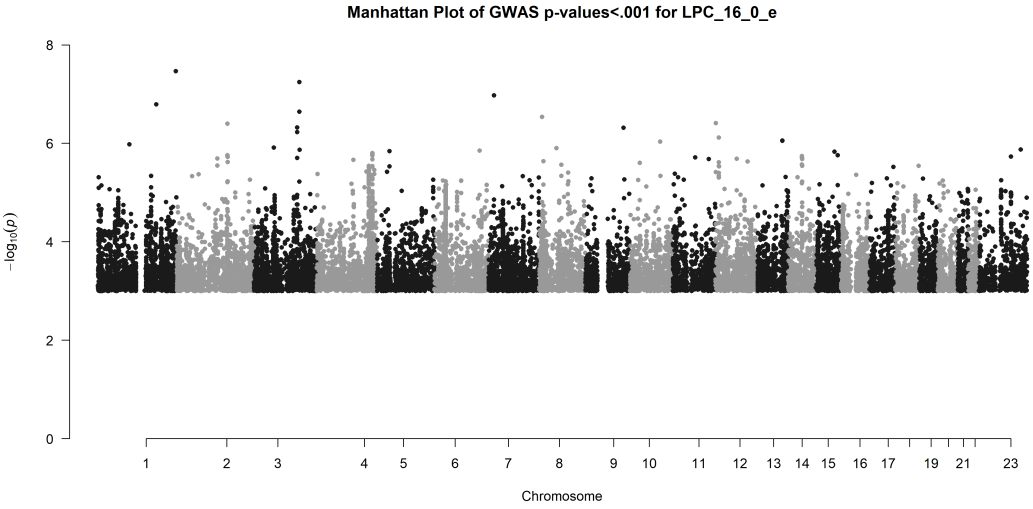


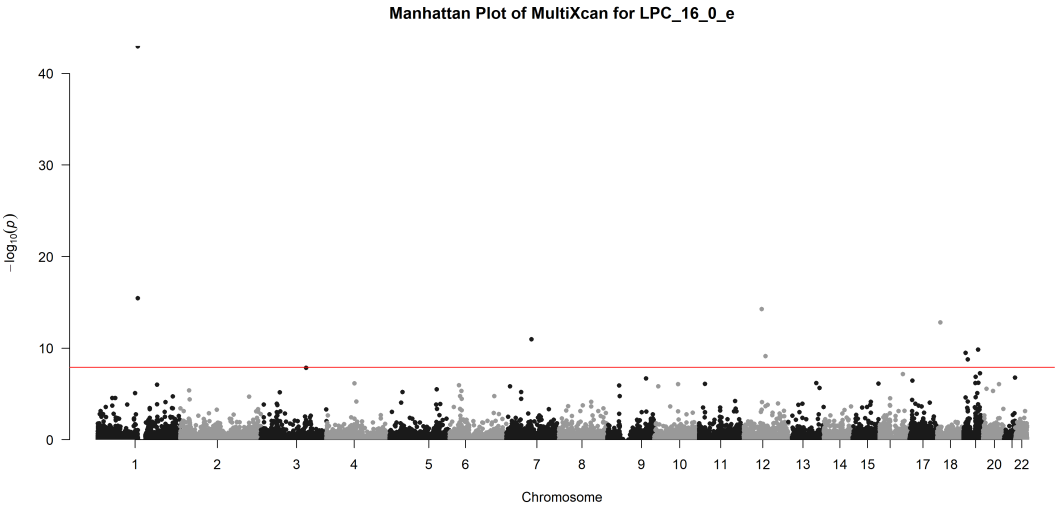
Lysophosphatidylcholine alkyl C16:0 (µg/dL)



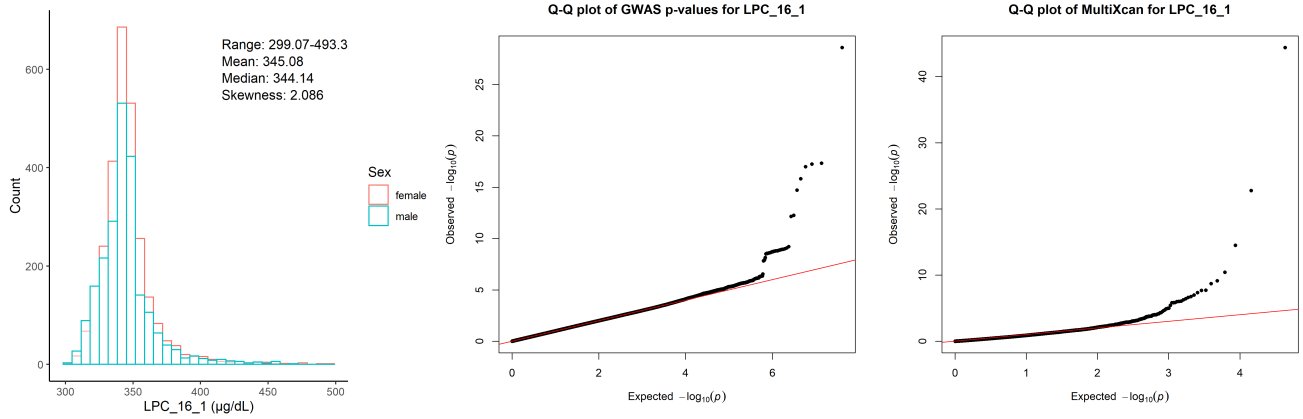


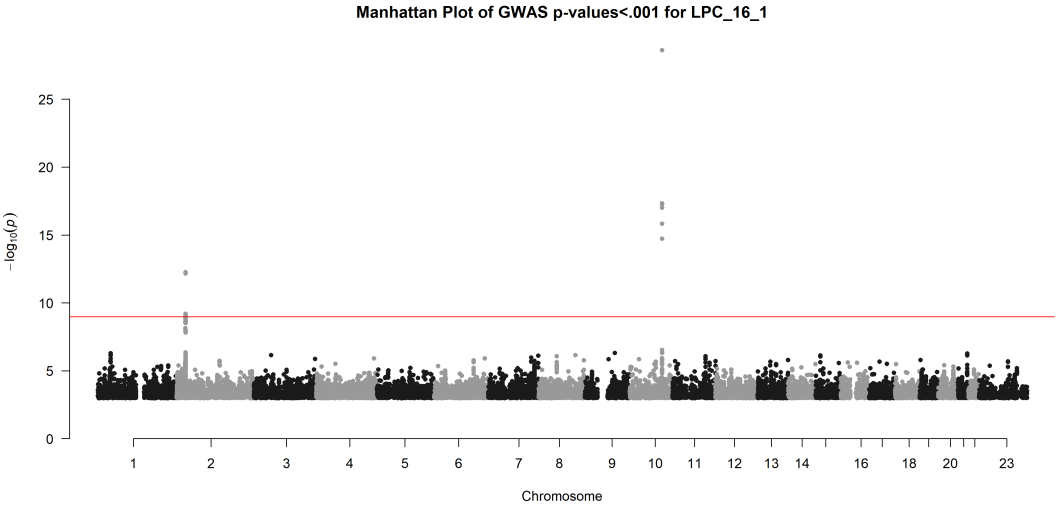


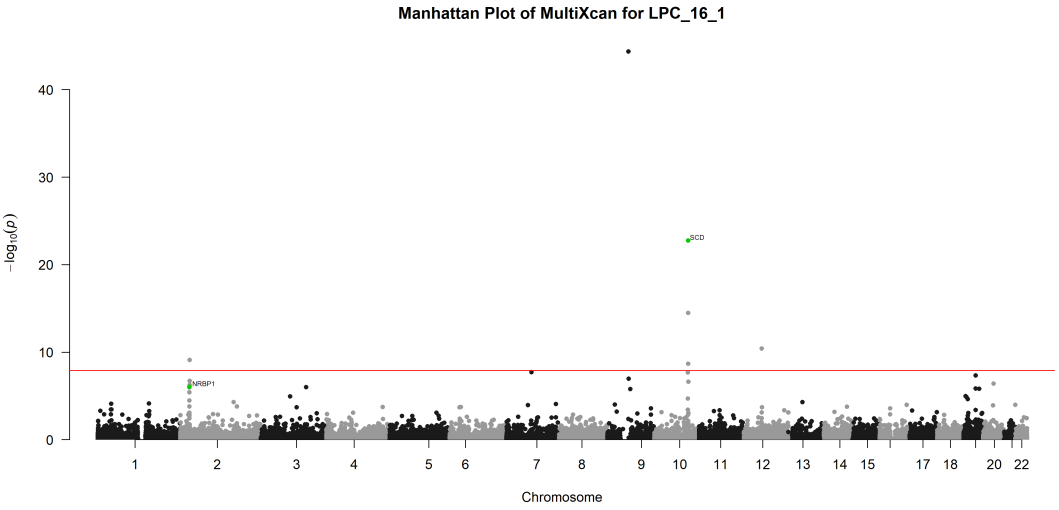




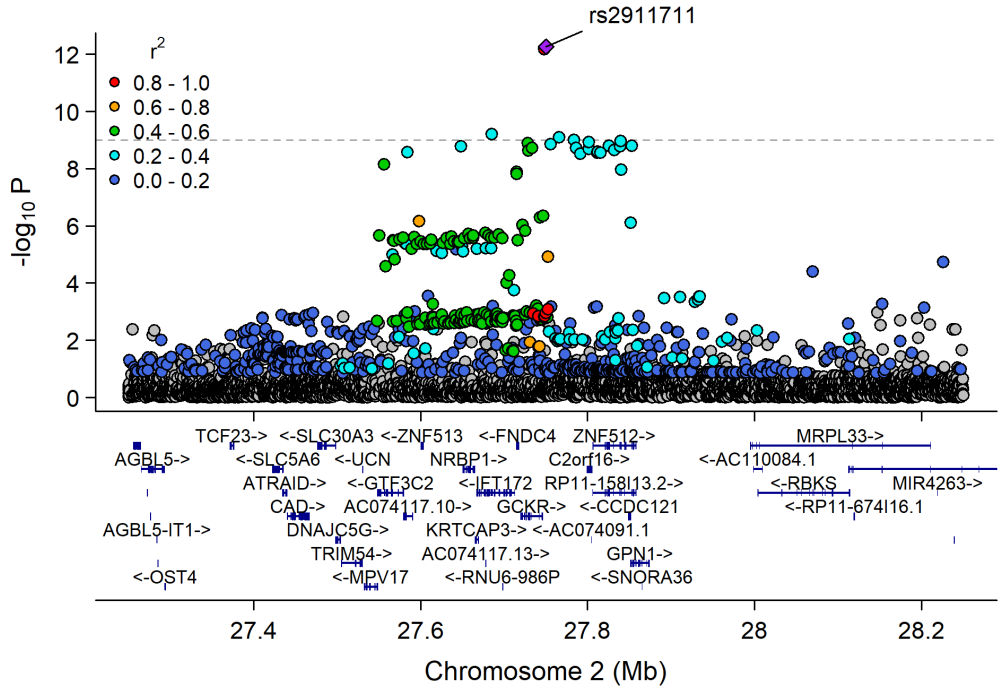
Lysophosphatidylcholine acyl C16:1 (µg/dL)



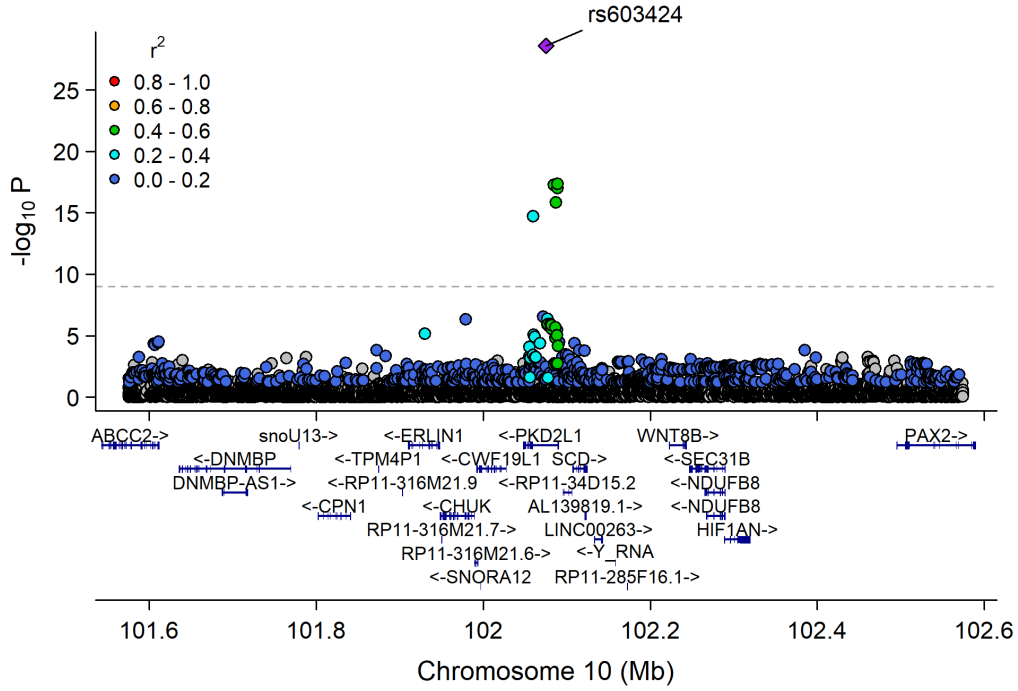




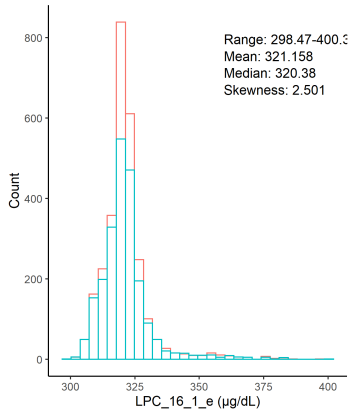
Chr2_26753815_28597624

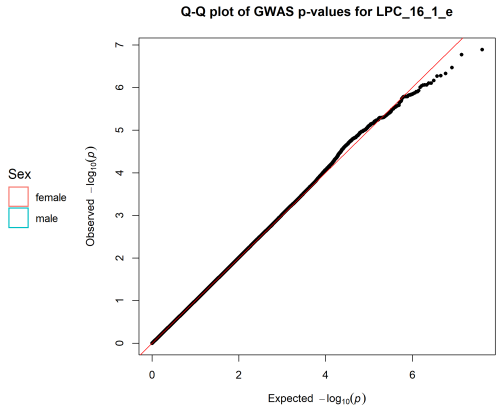


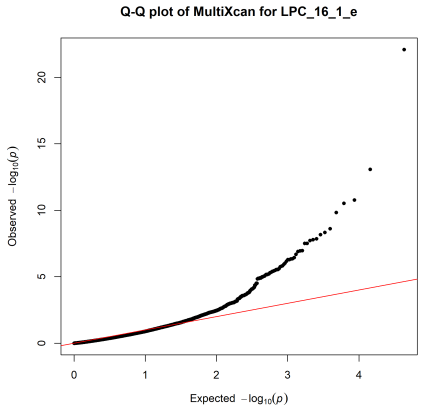
Chr10_101403323_102214579

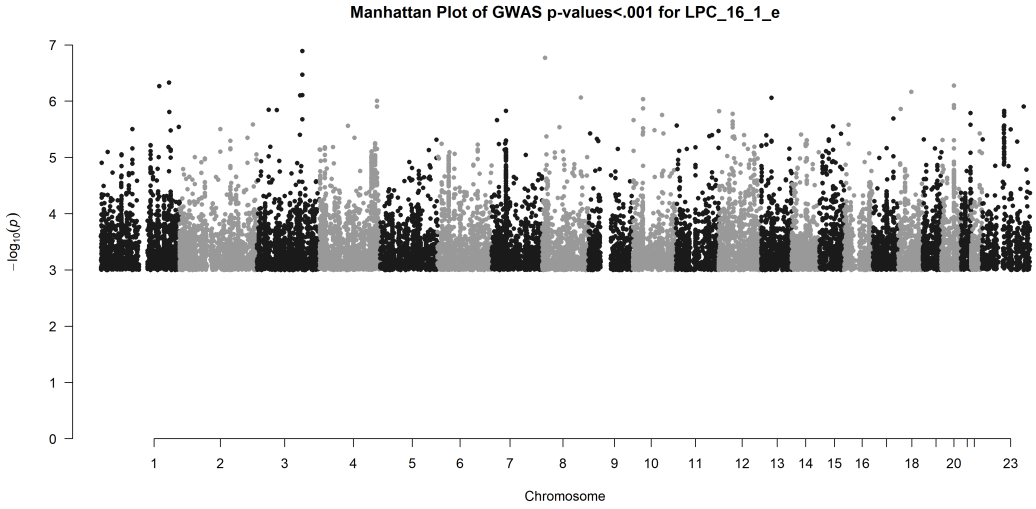


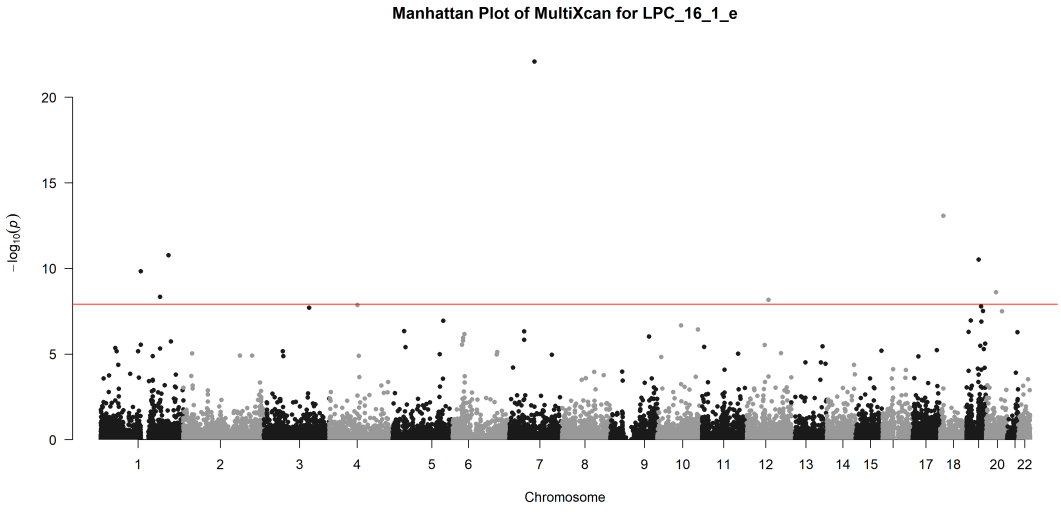
Lysophosphatidylcholine alkyl C16:1 (µg/dL)



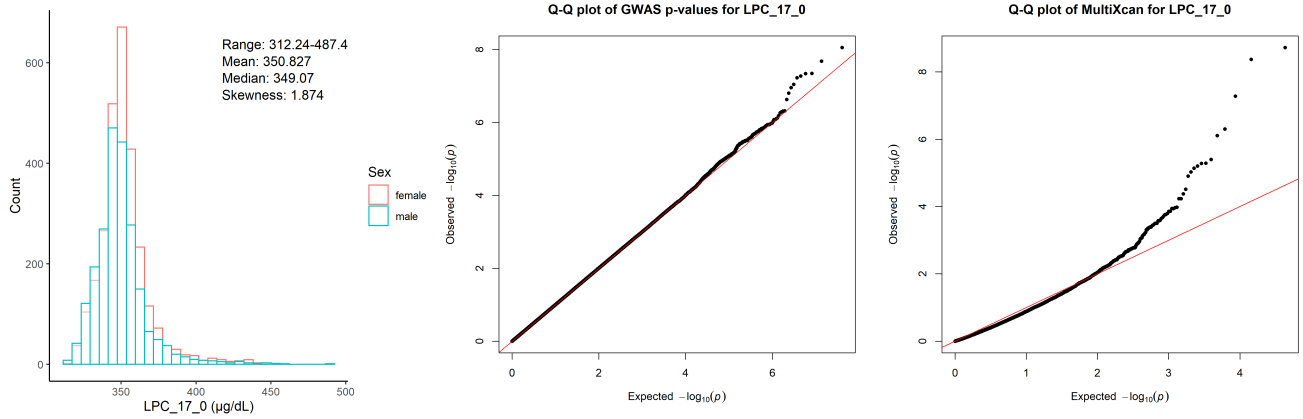


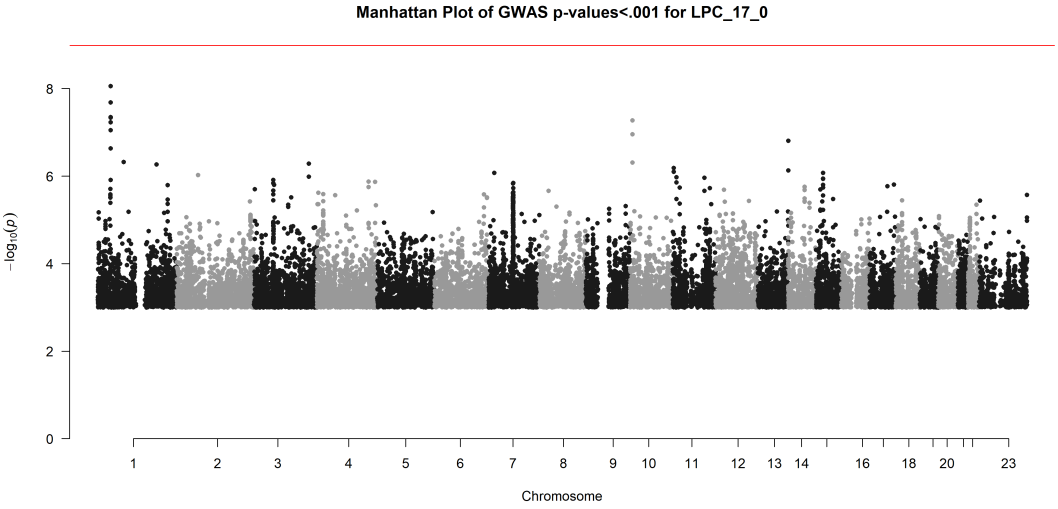


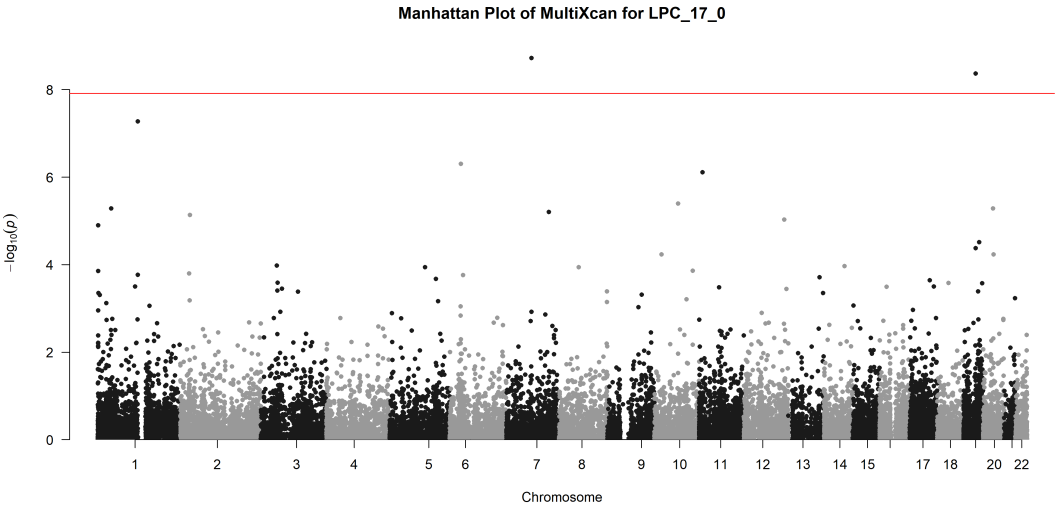




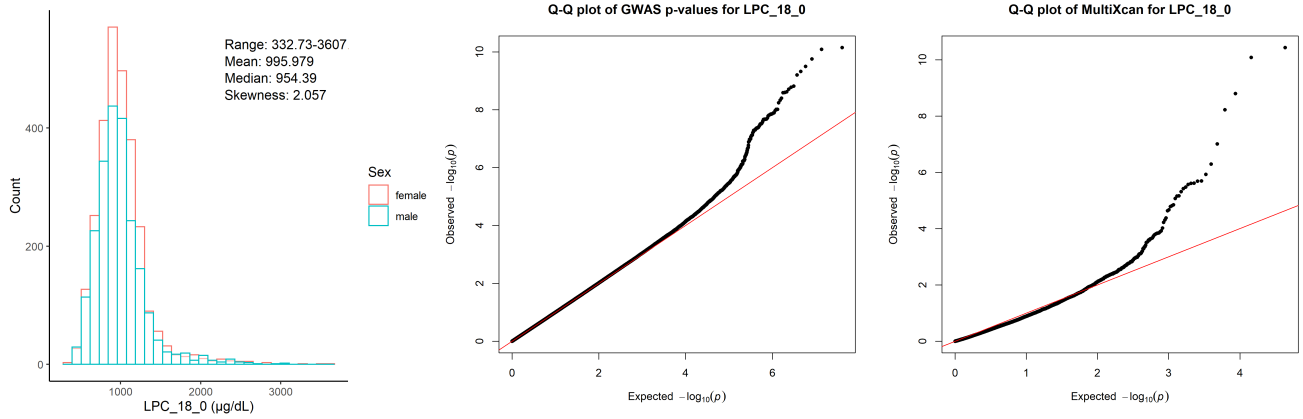
Lysophosphatidylcholine acyl C17:0 (µg/dL)

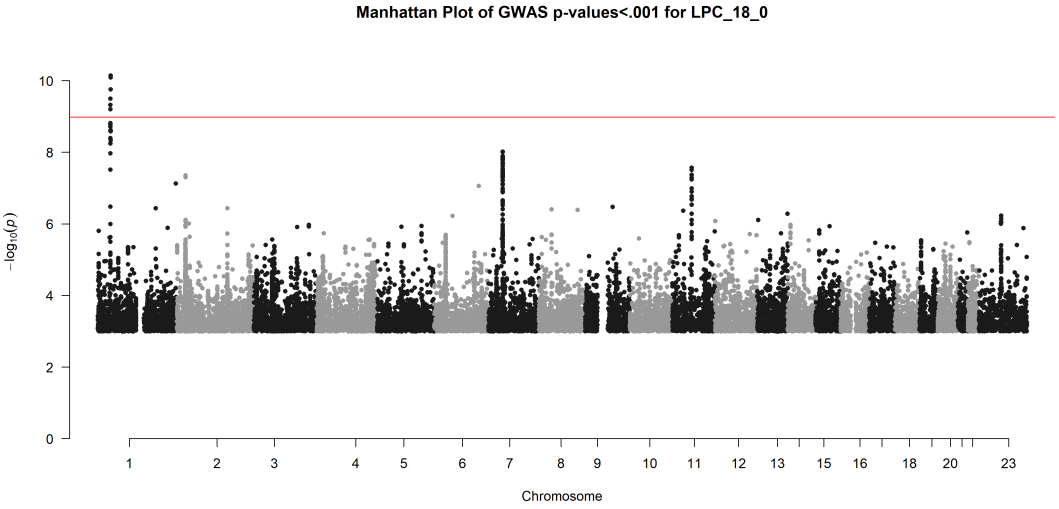


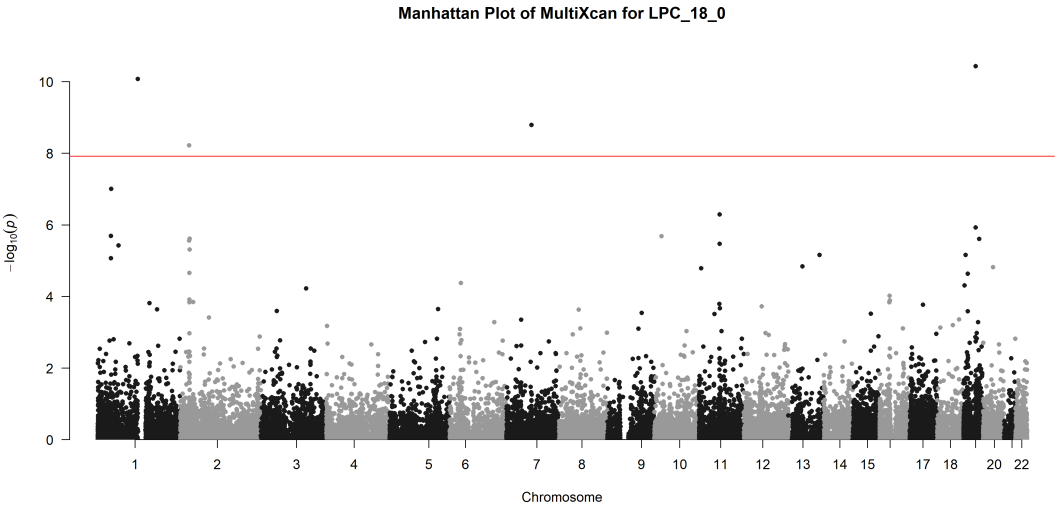




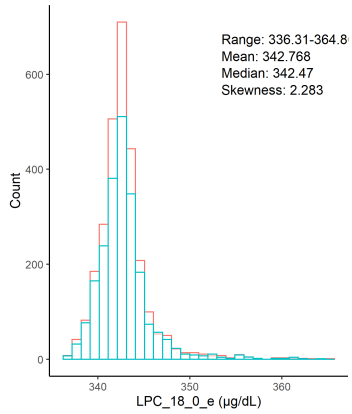
Lysophosphatidylcholine acyl C18:0 (µg/dL)

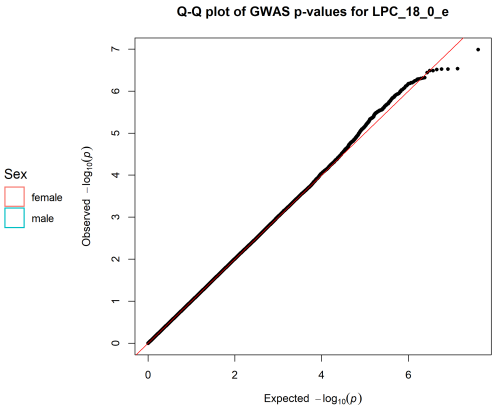


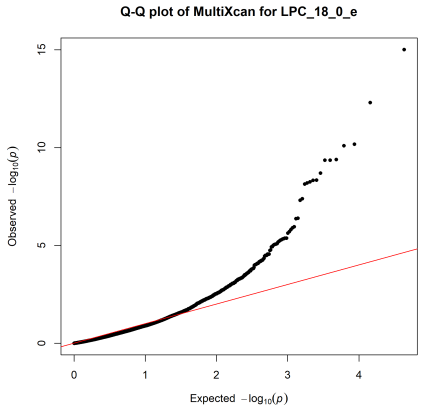


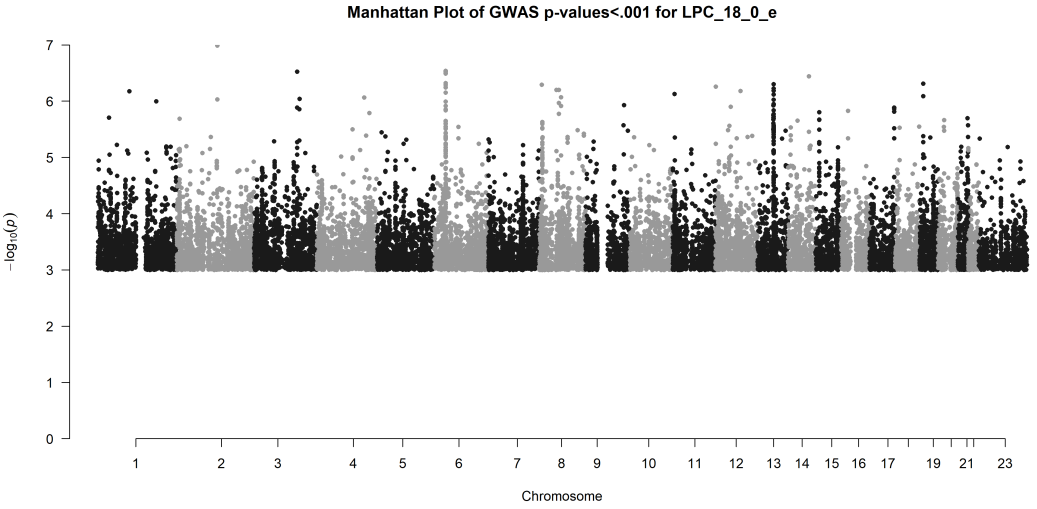


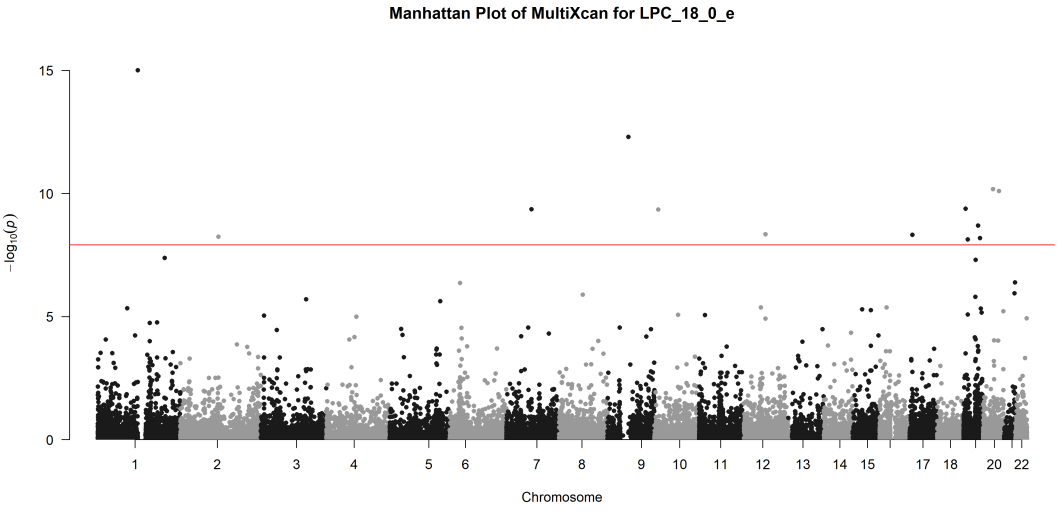
Lysophosphatidylcholine alkyl C18:0 (µg/dL)



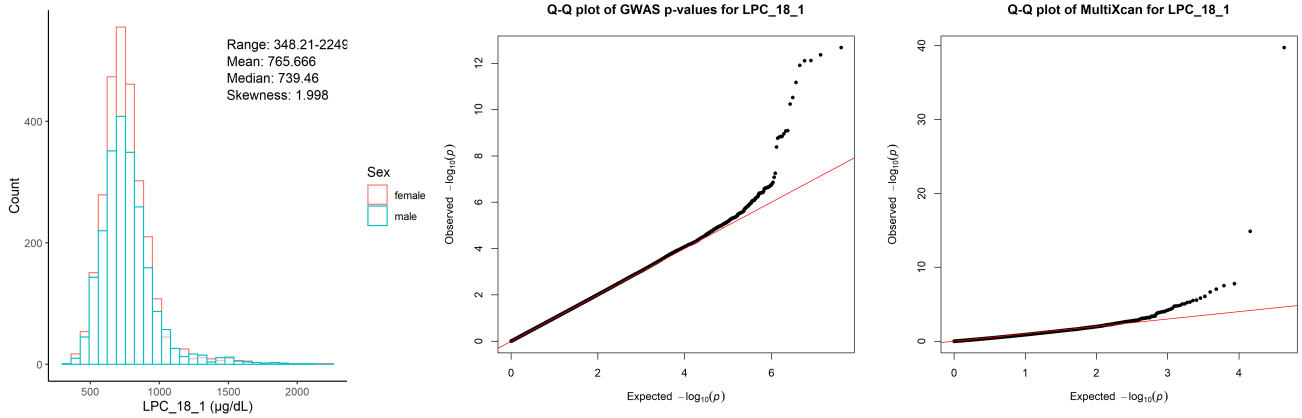


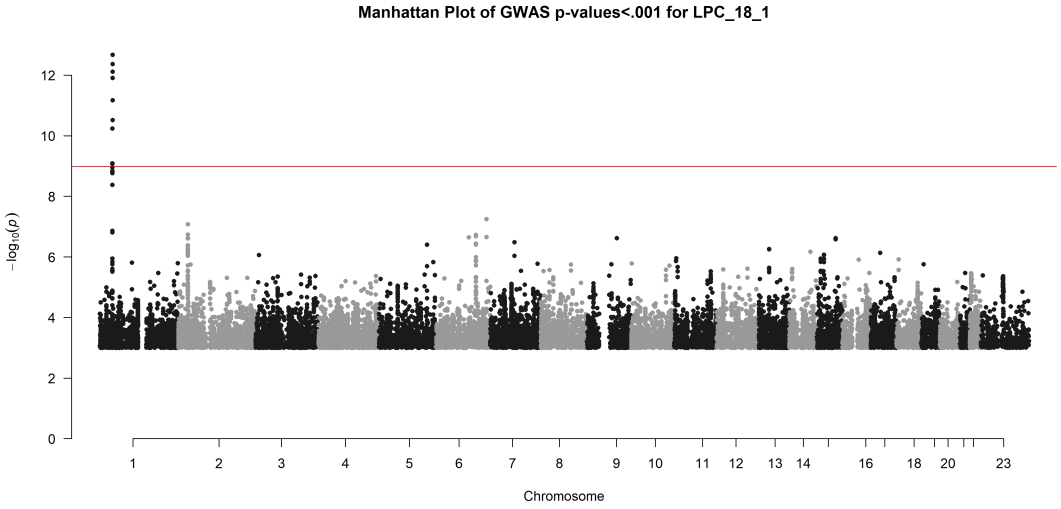


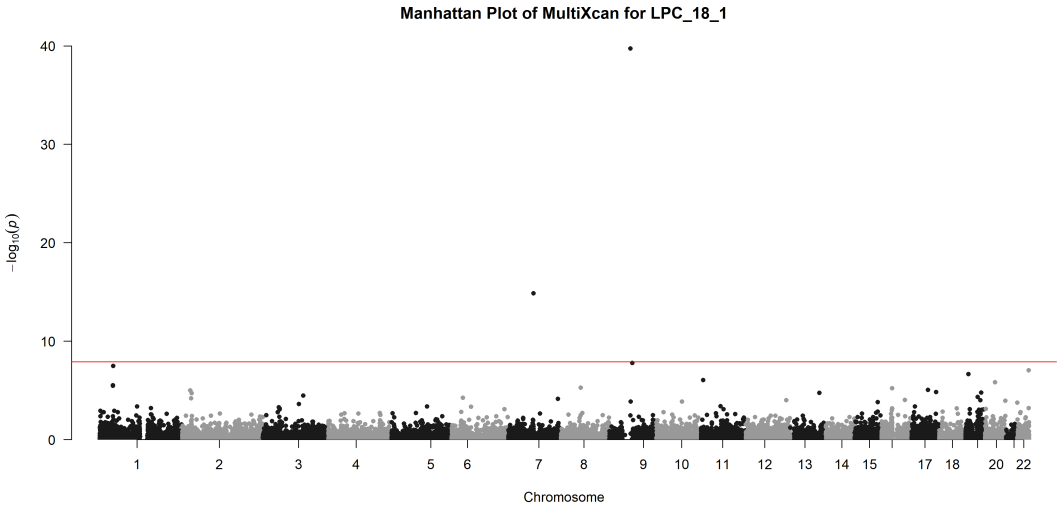




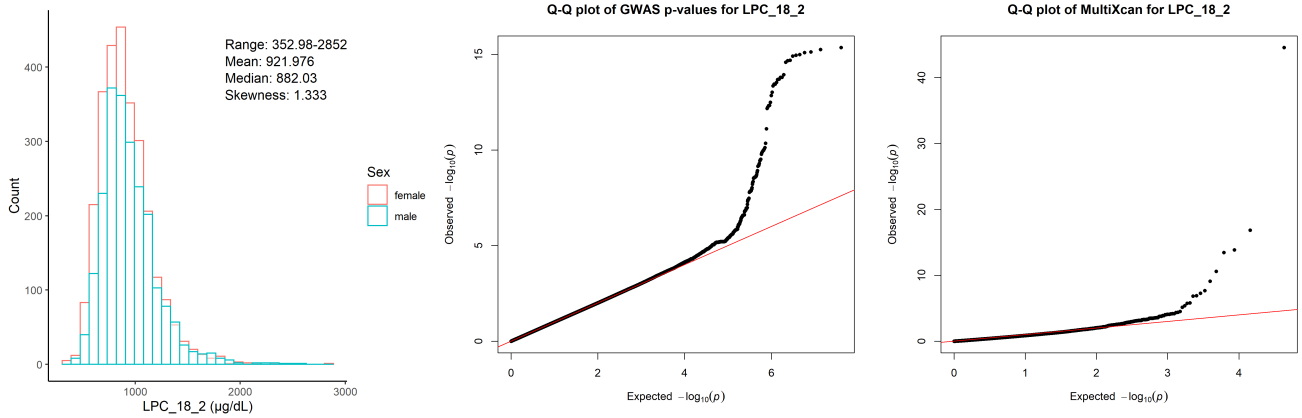
Lysophosphatidylcholine acyl C18:1 (µg/dL)

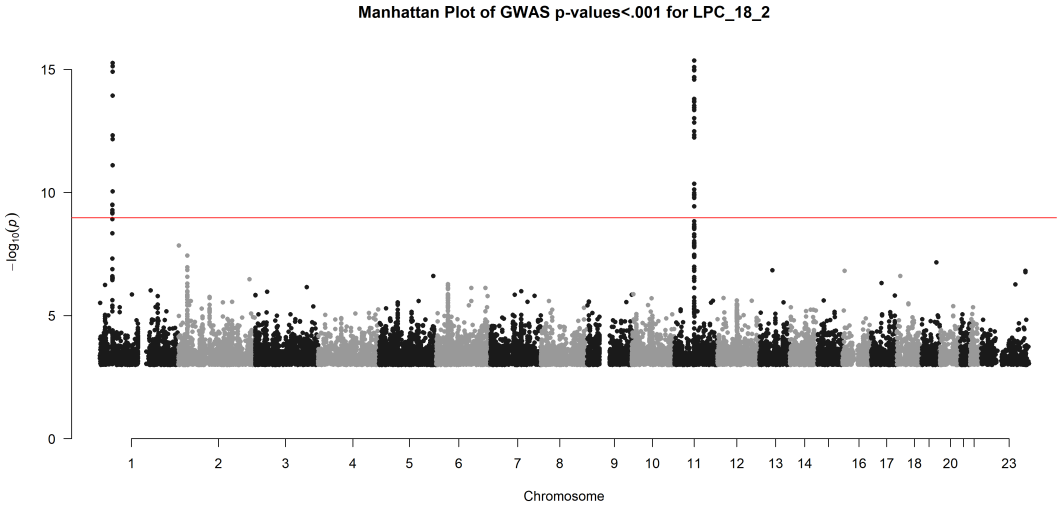


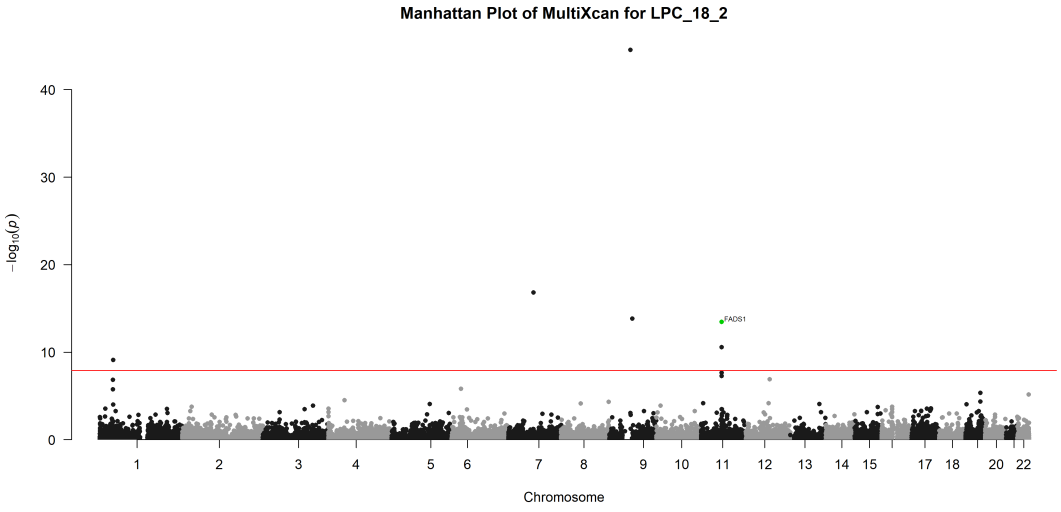




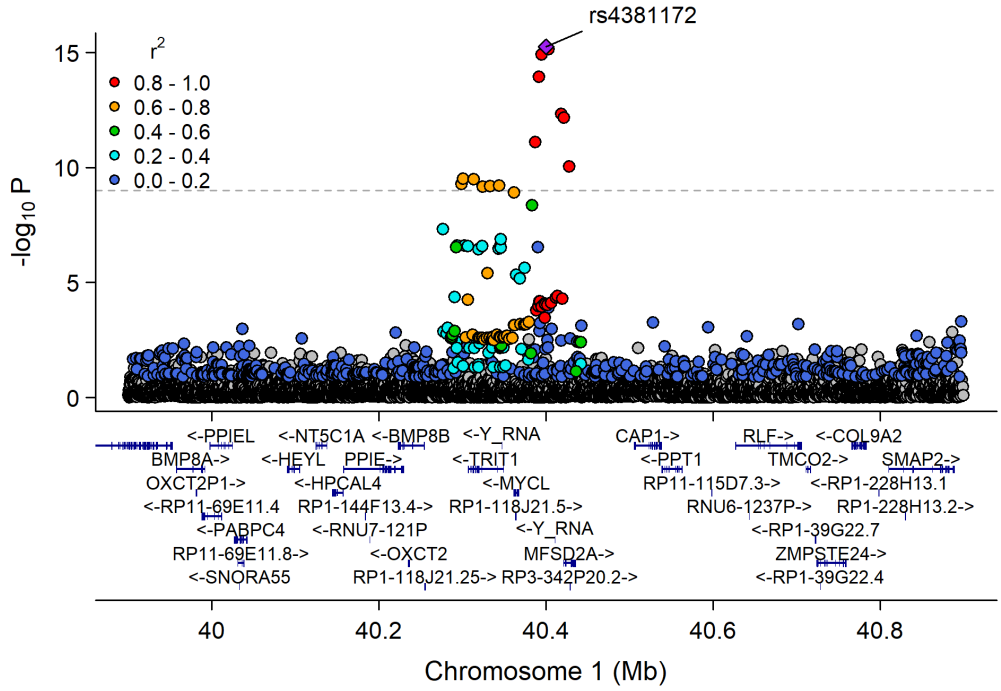
Lysophosphatidylcholine acyl C18:2 (µg/dL)



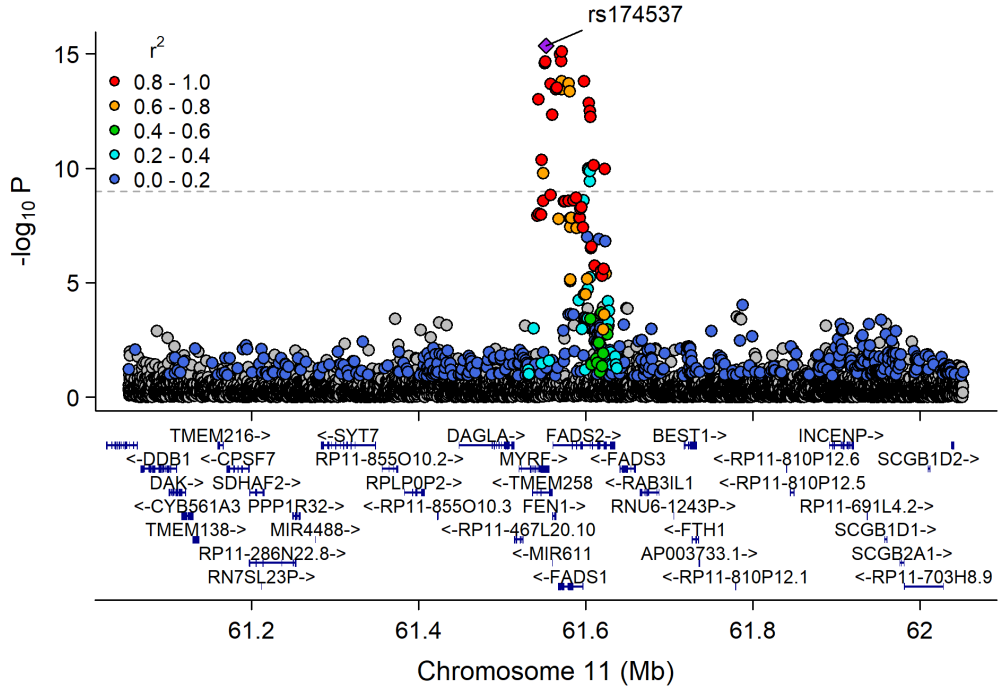




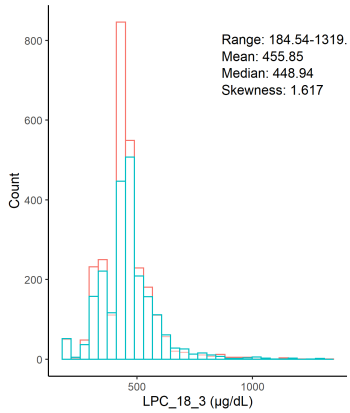
Chr1__40208386__40441964

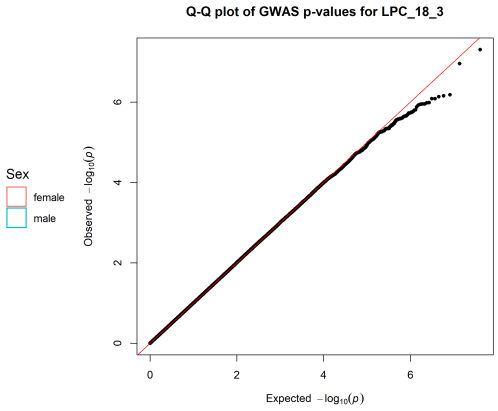


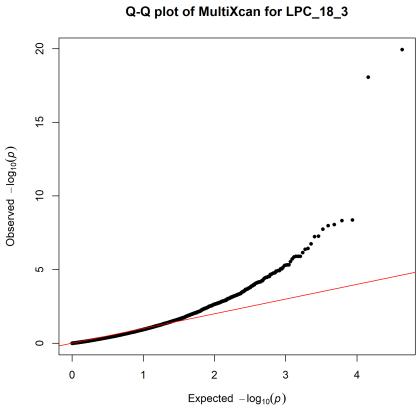
Chr11__59978355__62914375

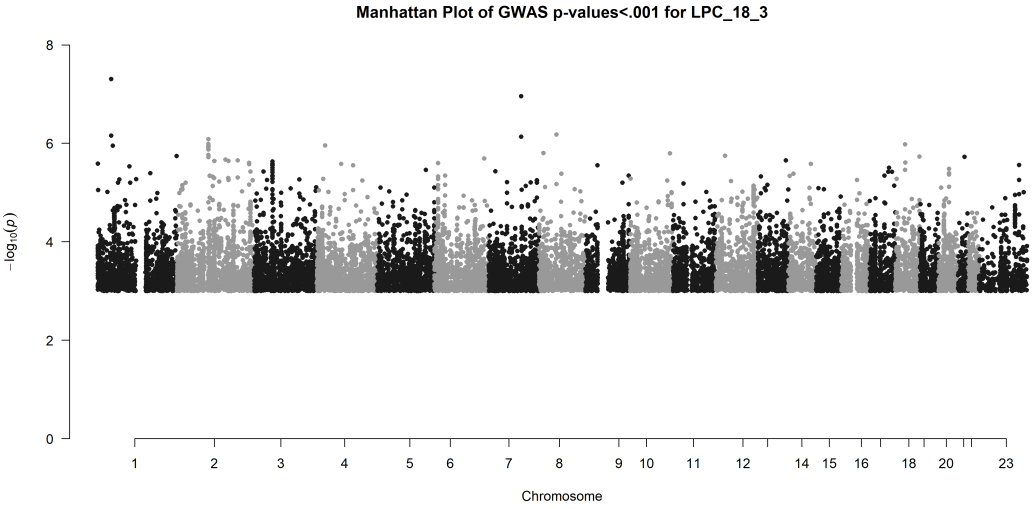


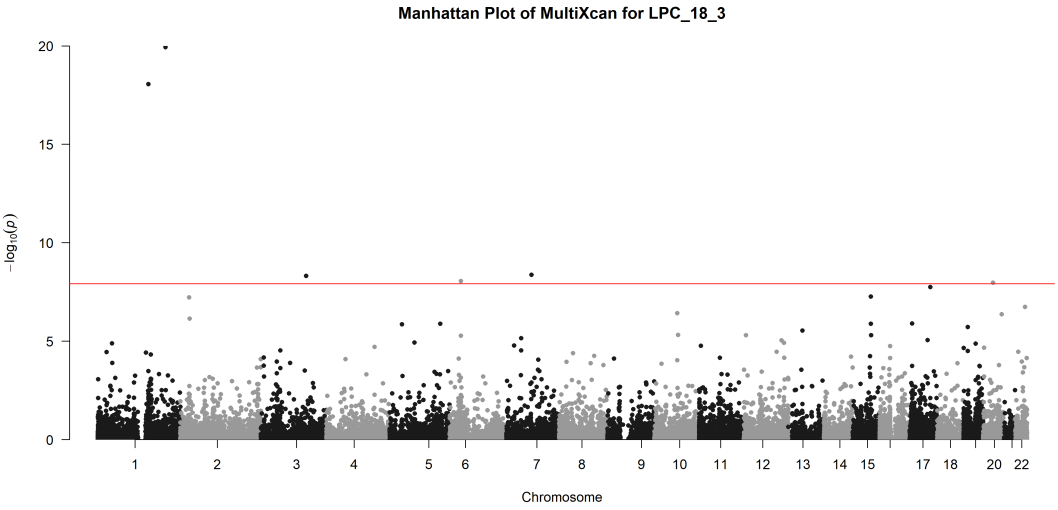
Lysophosphatidylcholine acyl C18:3 (µg/dL)



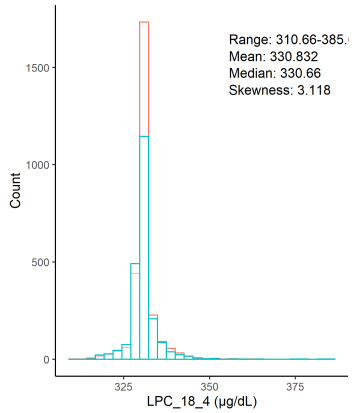


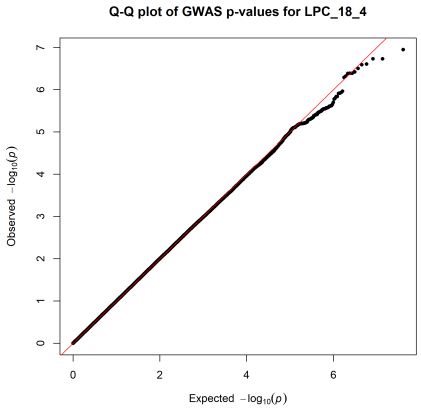


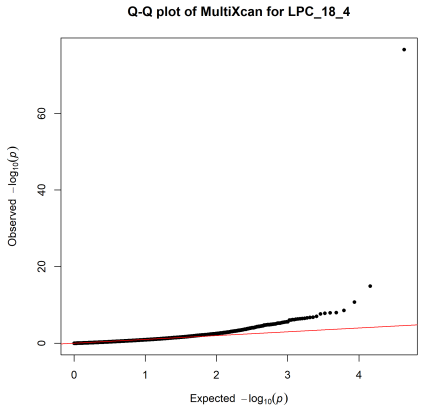


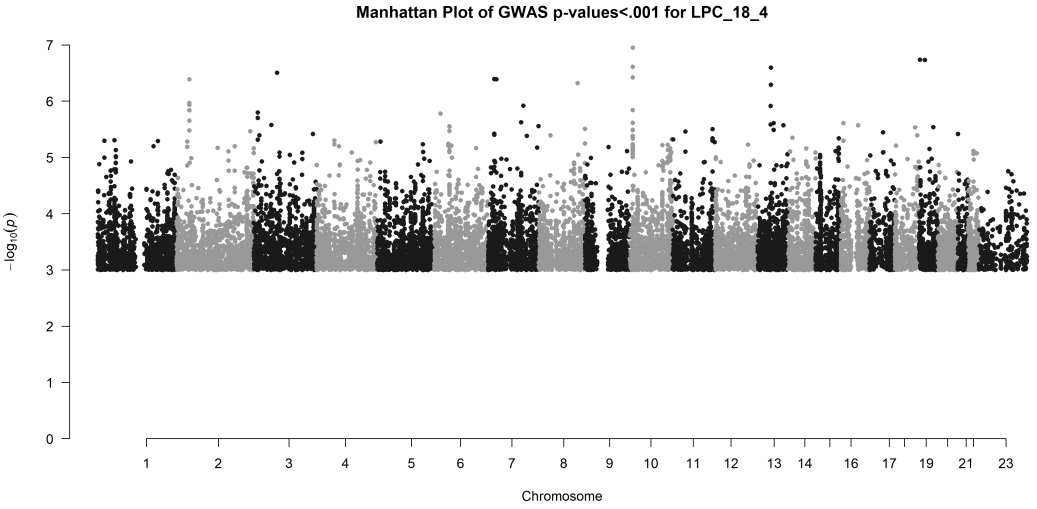


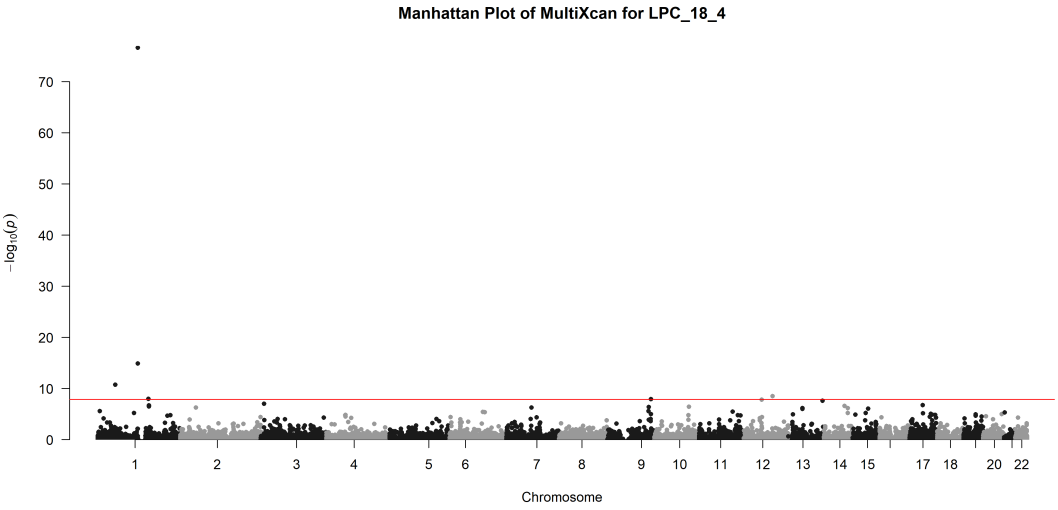
Lysophosphatidylcholine acyl C18:4 (µg/dL)



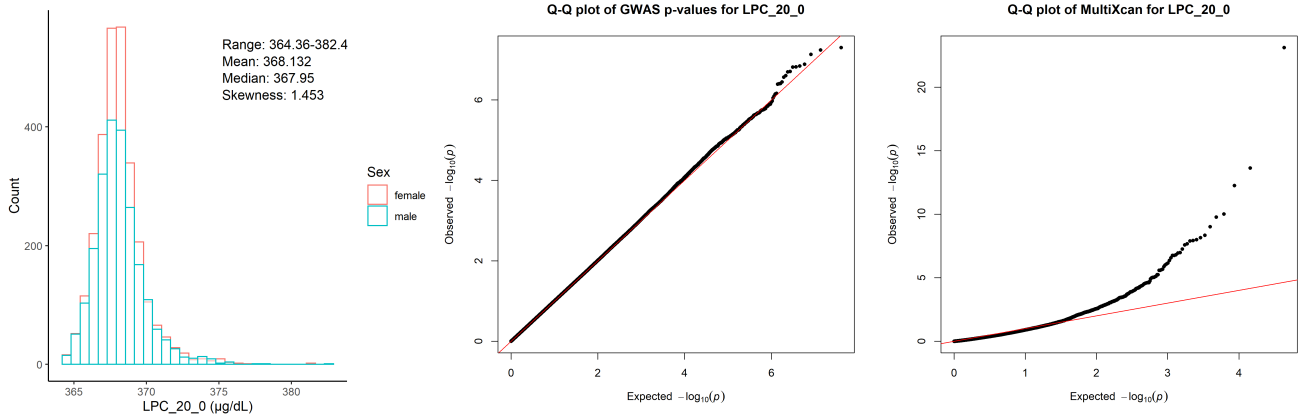


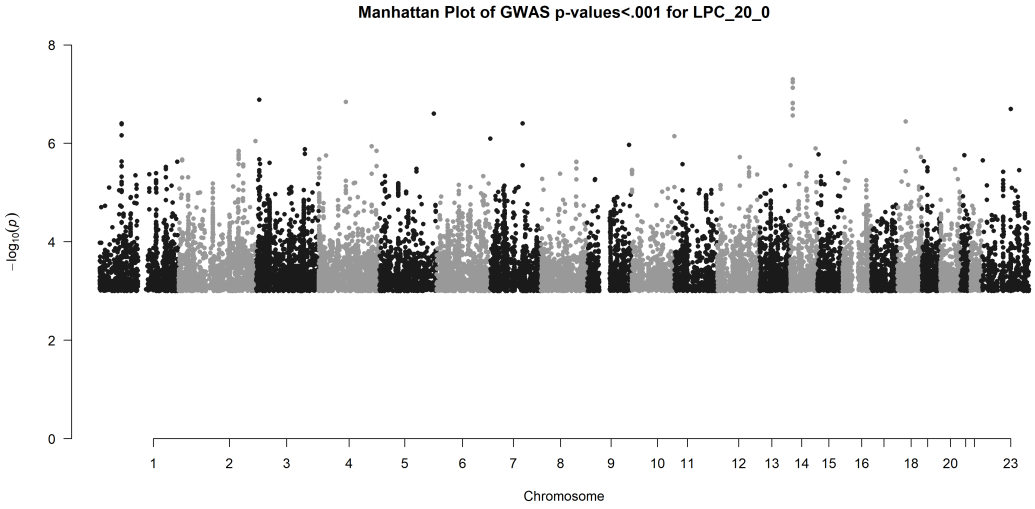


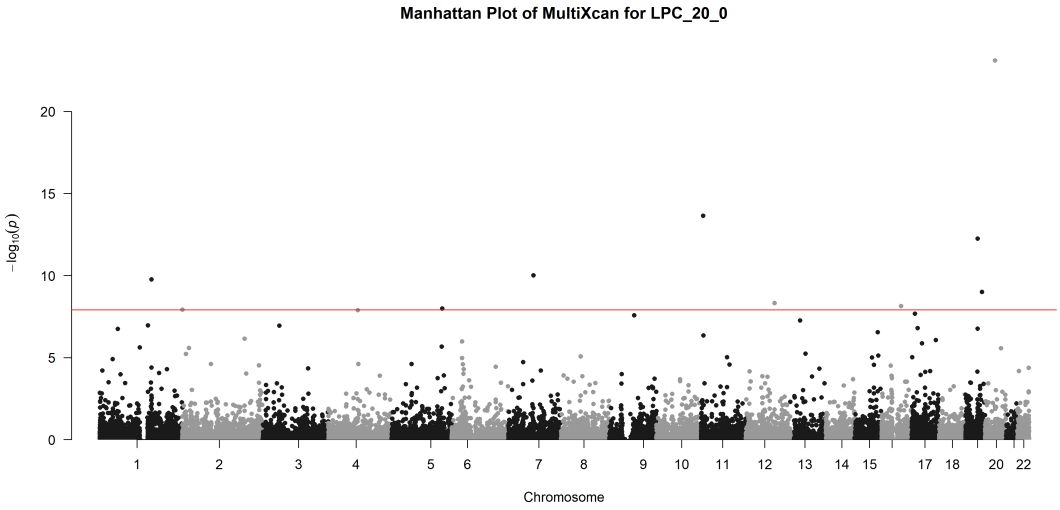




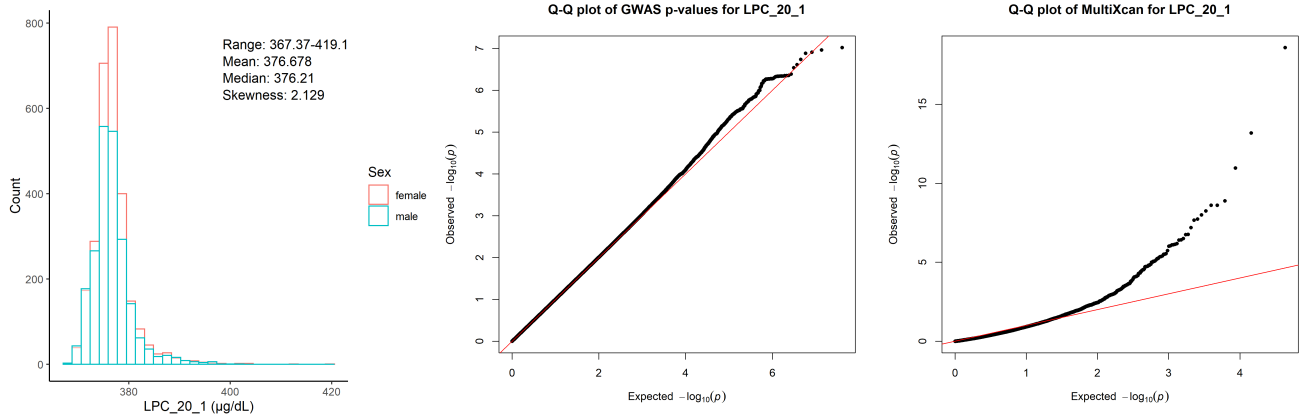
Lysophosphatidylcholine acyl C20:0 (µg/dL)

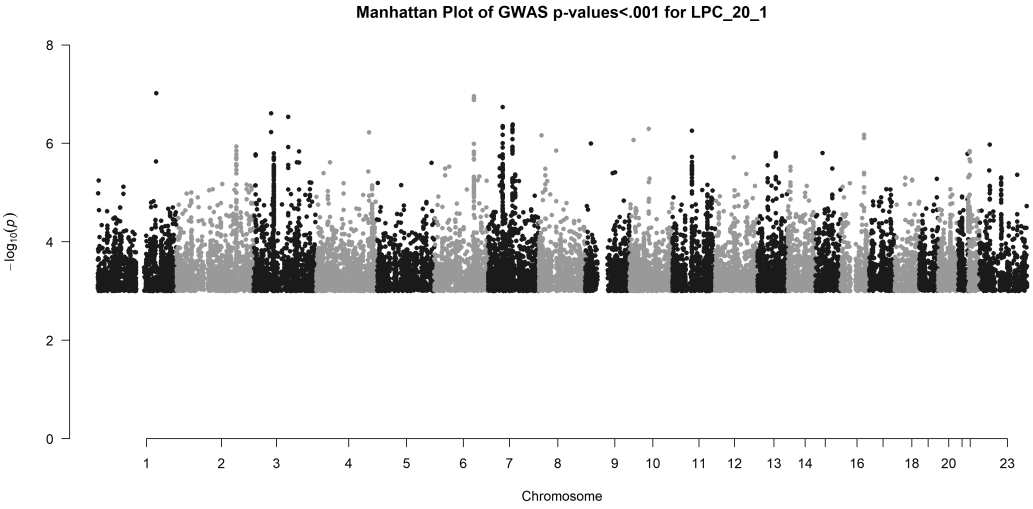


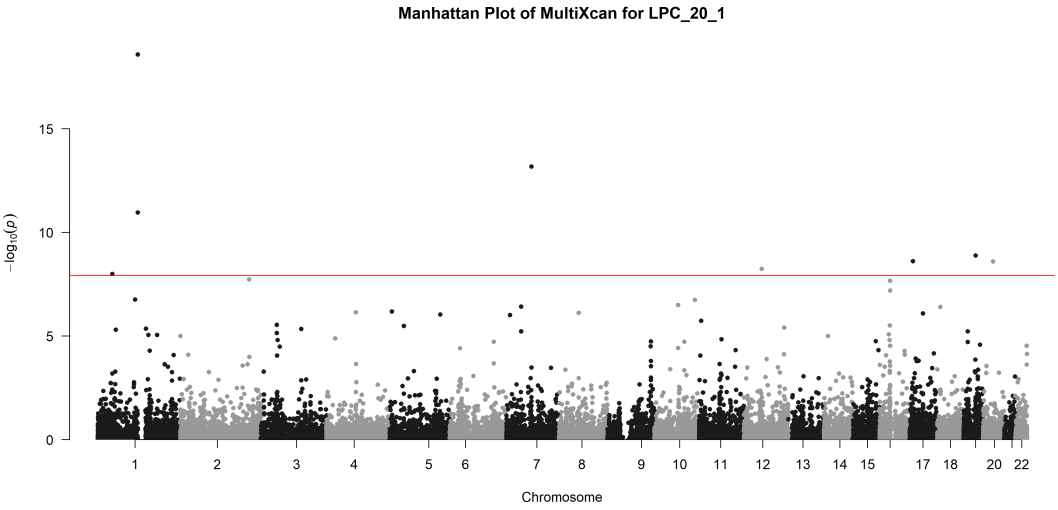




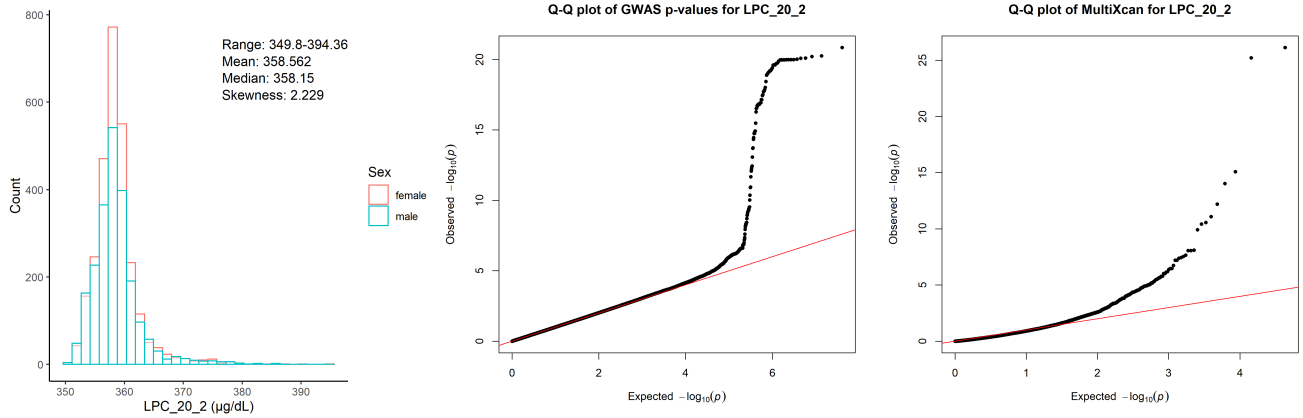
Lysophosphatidylcholine acyl C20:1 (µg/dL)

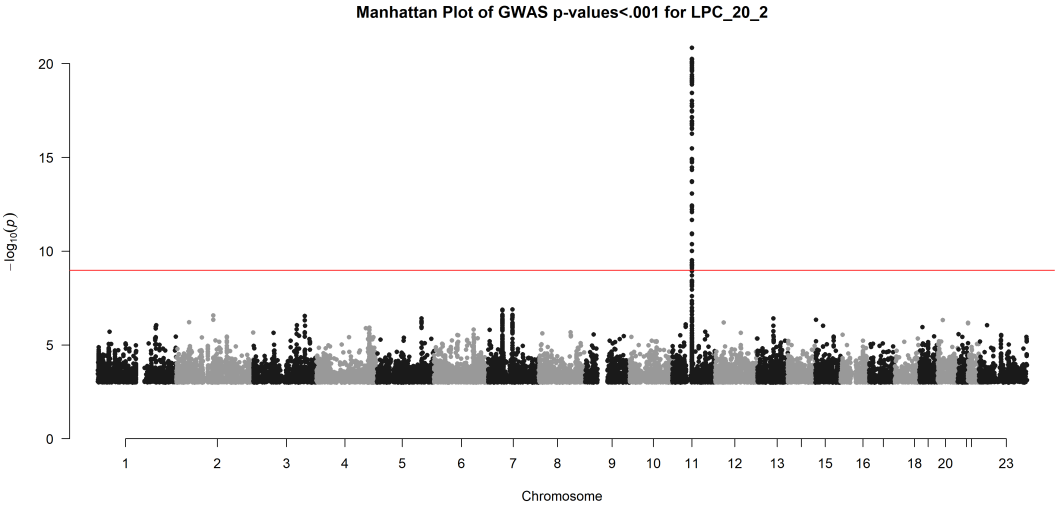


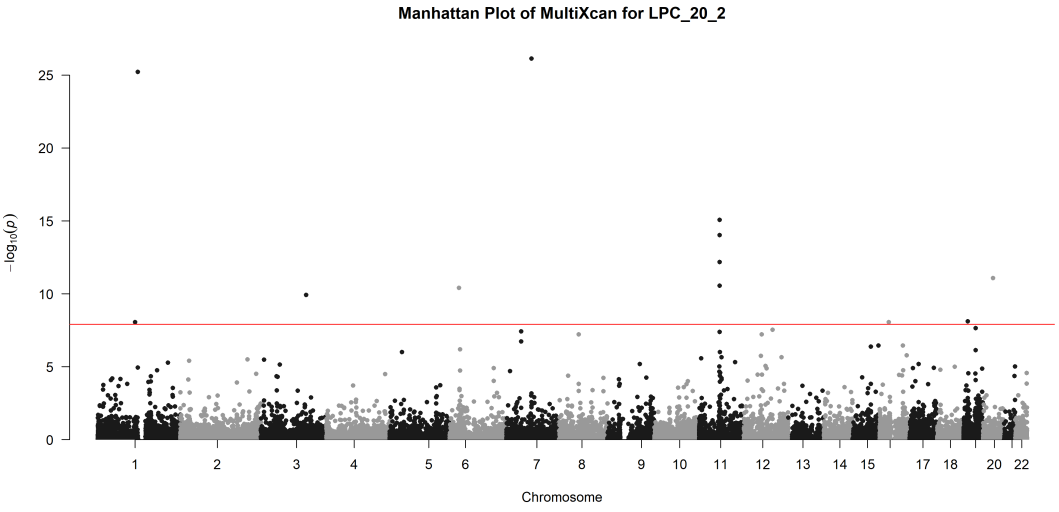




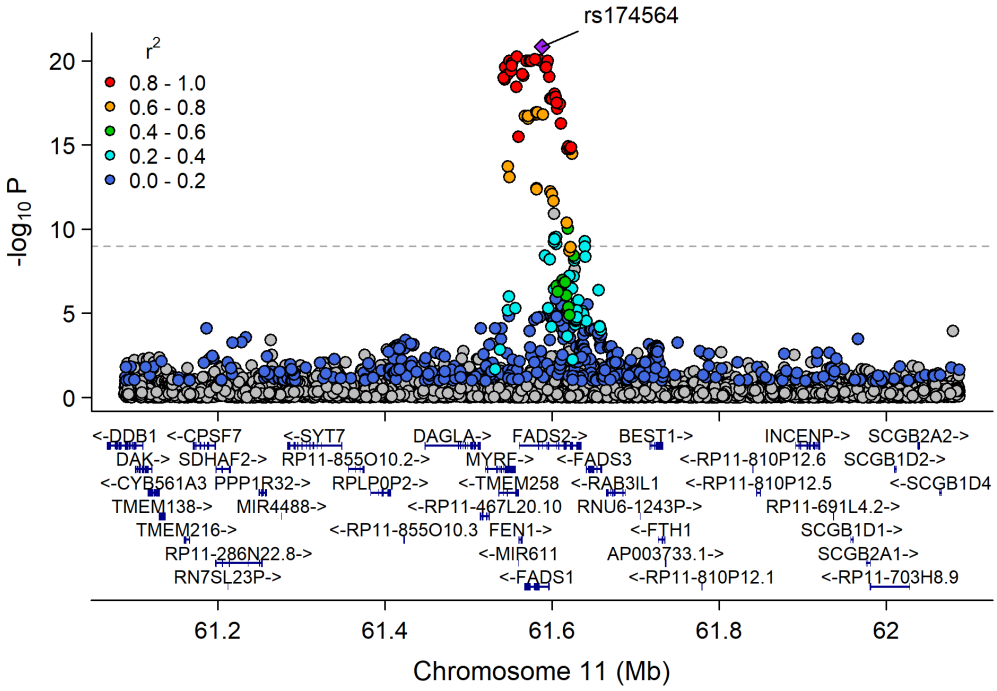
Lysophosphatidylcholine acyl C20:2 (µg/dL)



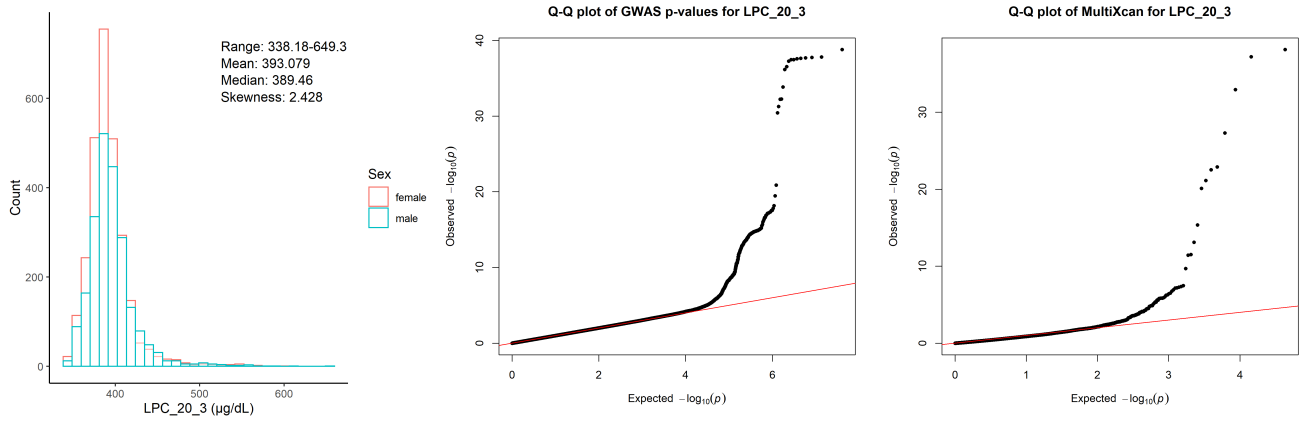


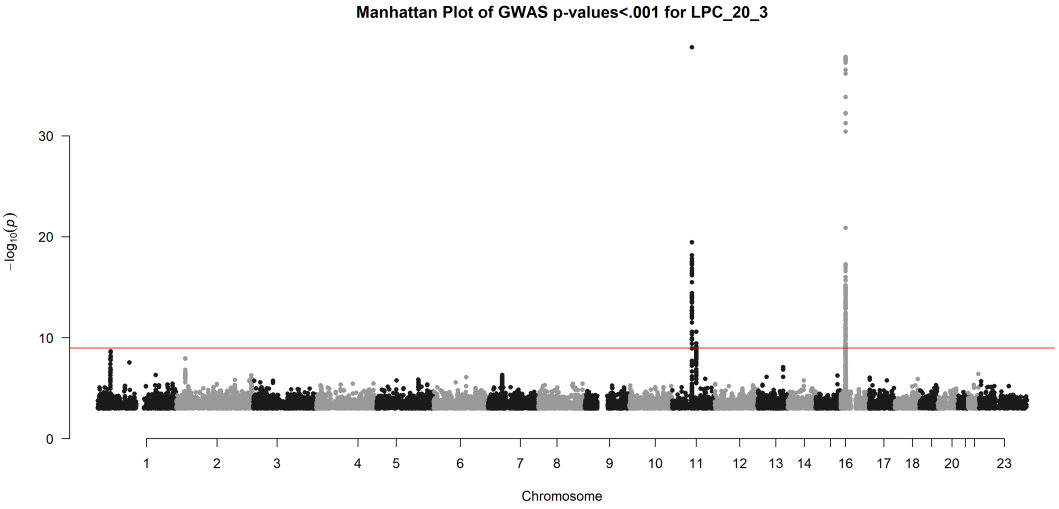


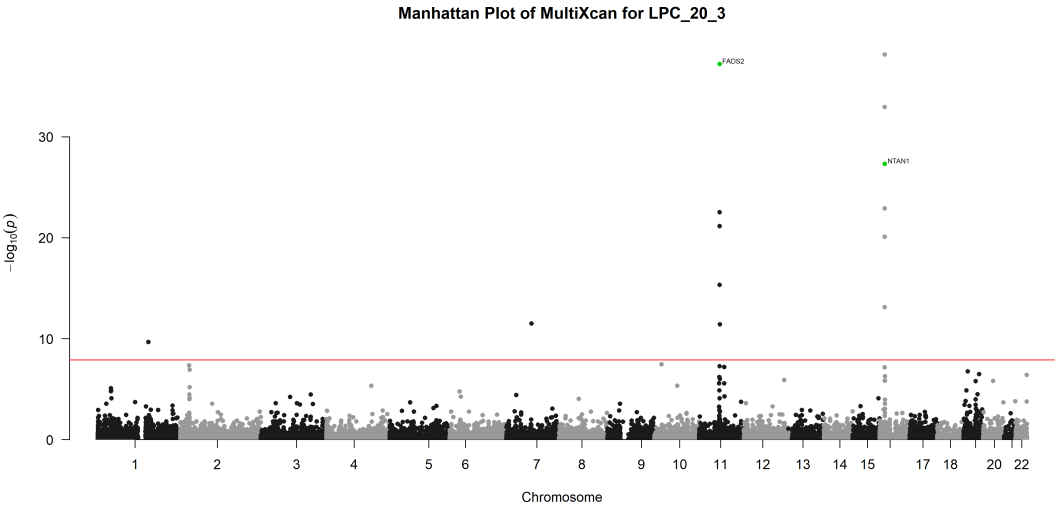
Chr11__59978355__62914375



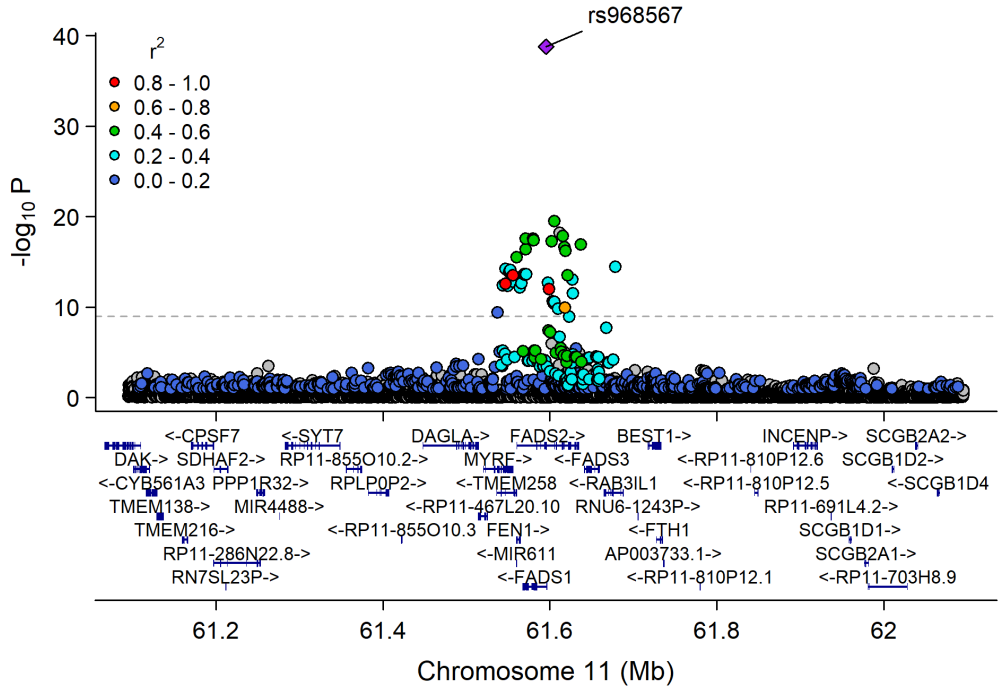
Lysophosphatidylcholine acyl C20:3 (µg/dL)



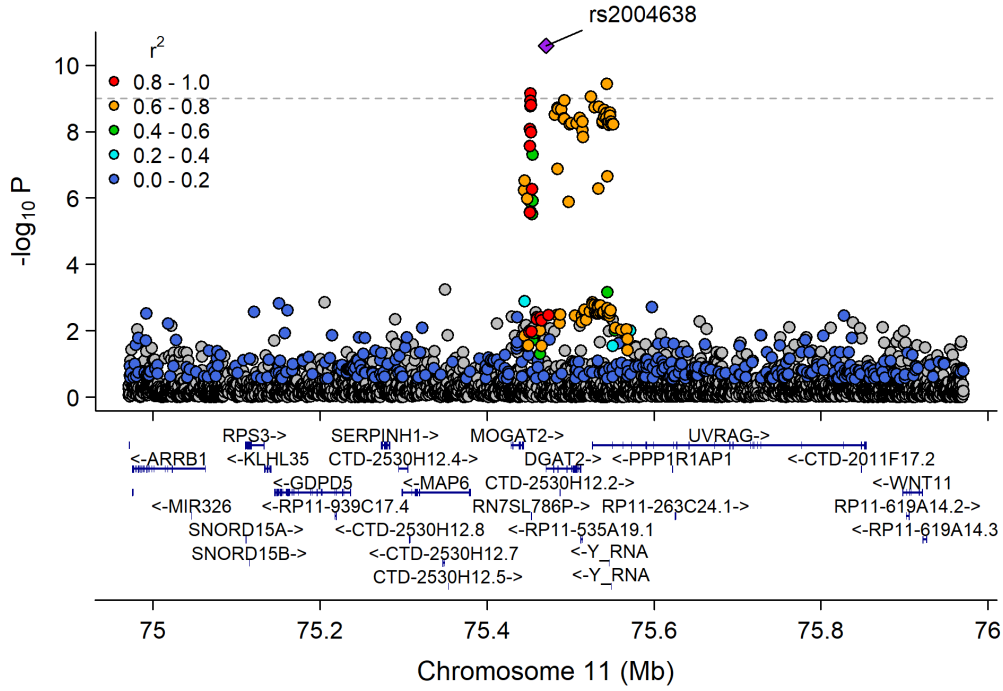




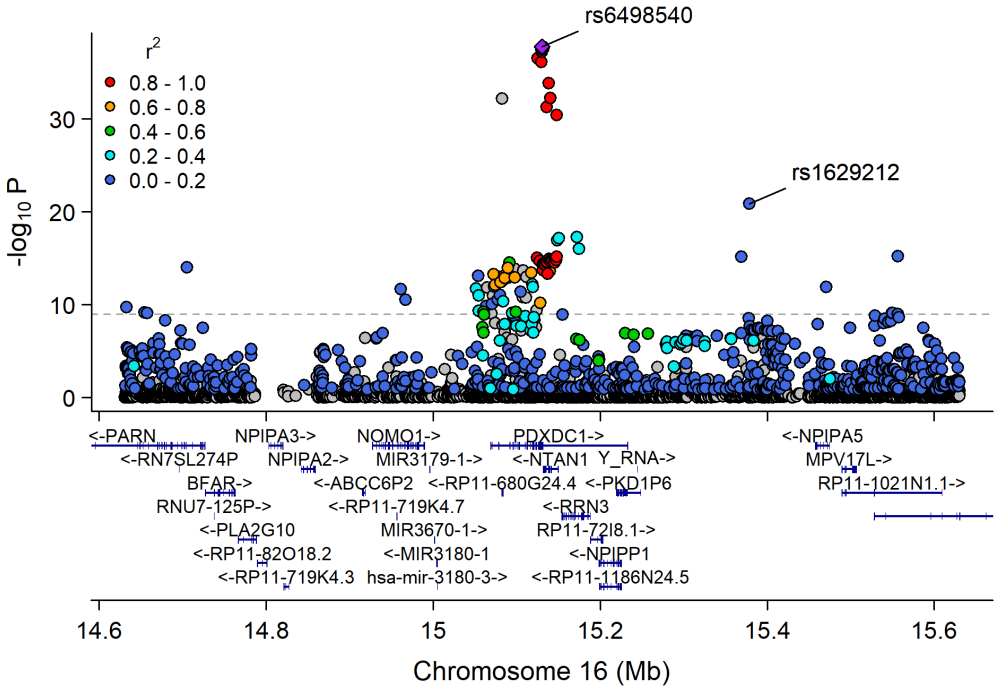
Chr11_59978355_62914375



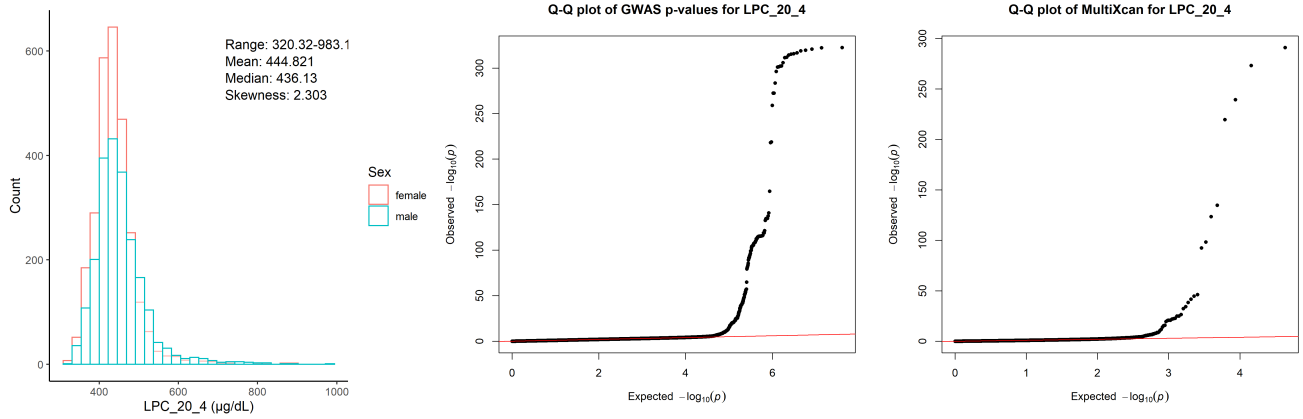
Chr11_75180957_75845682

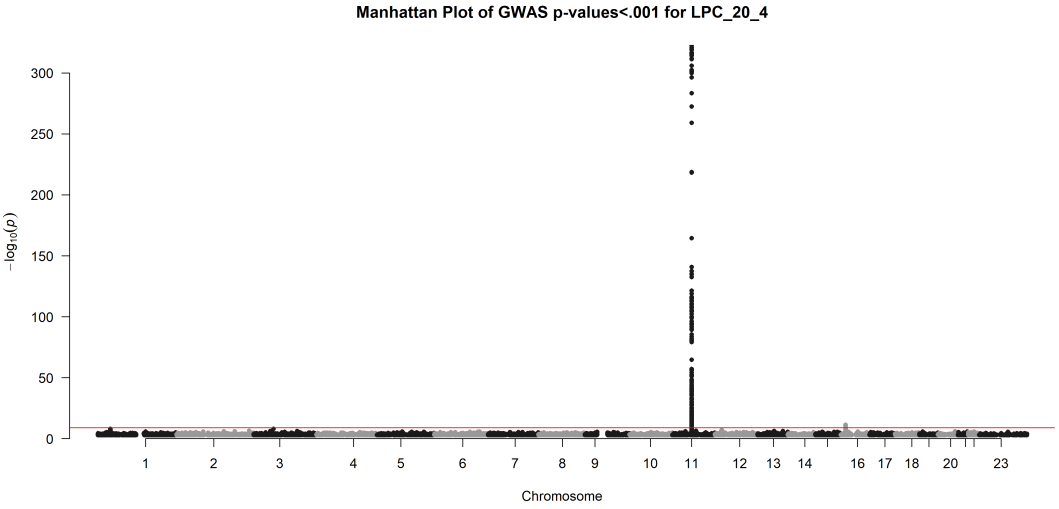


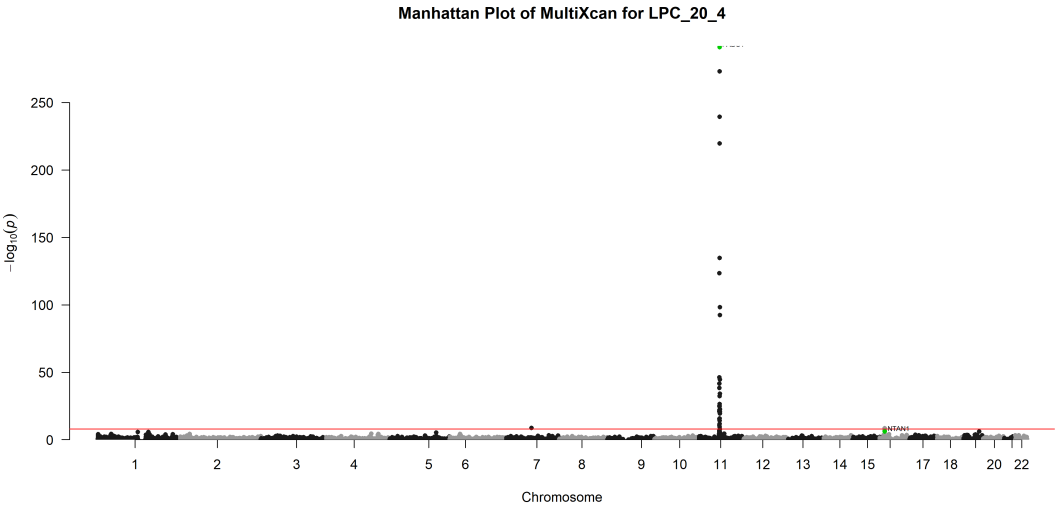
Chr16_14393707_15885866



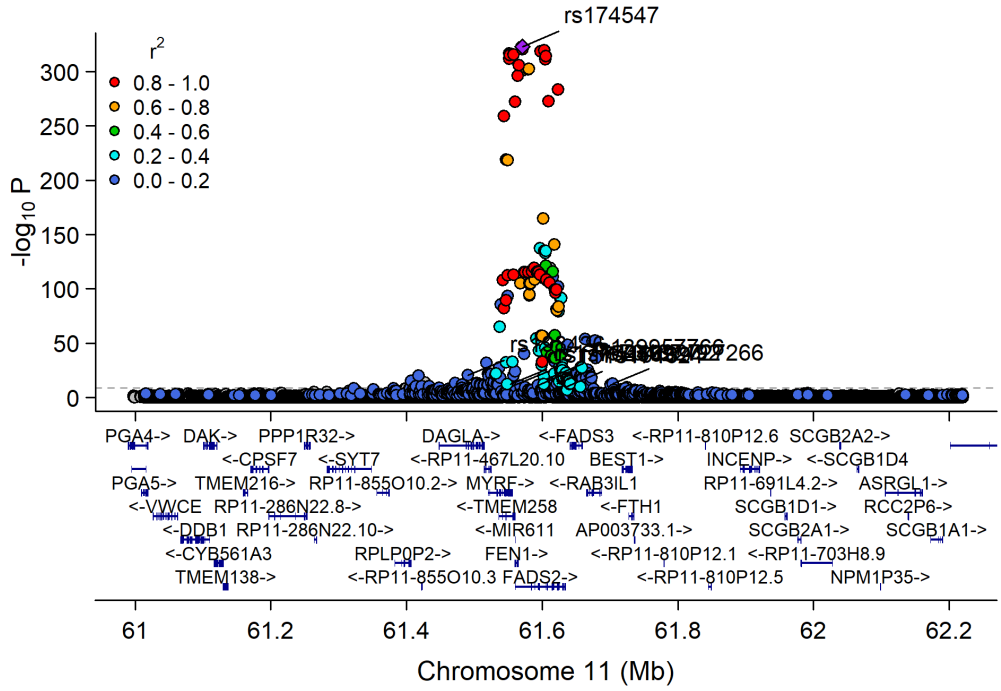
Lysophosphatidylcholine acyl C20:4 (µg/dL)



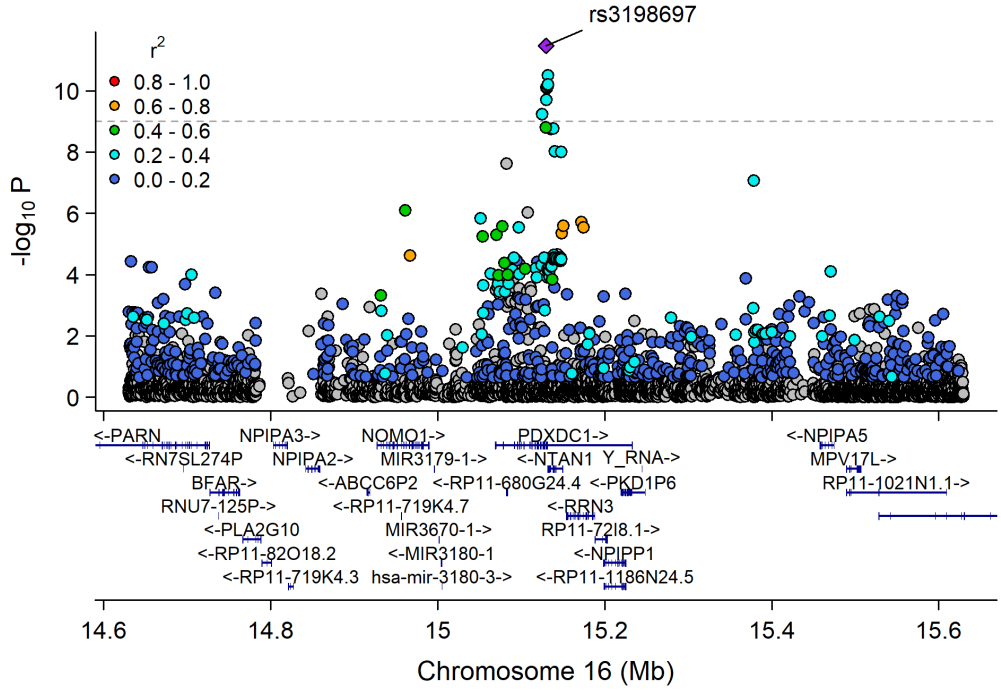




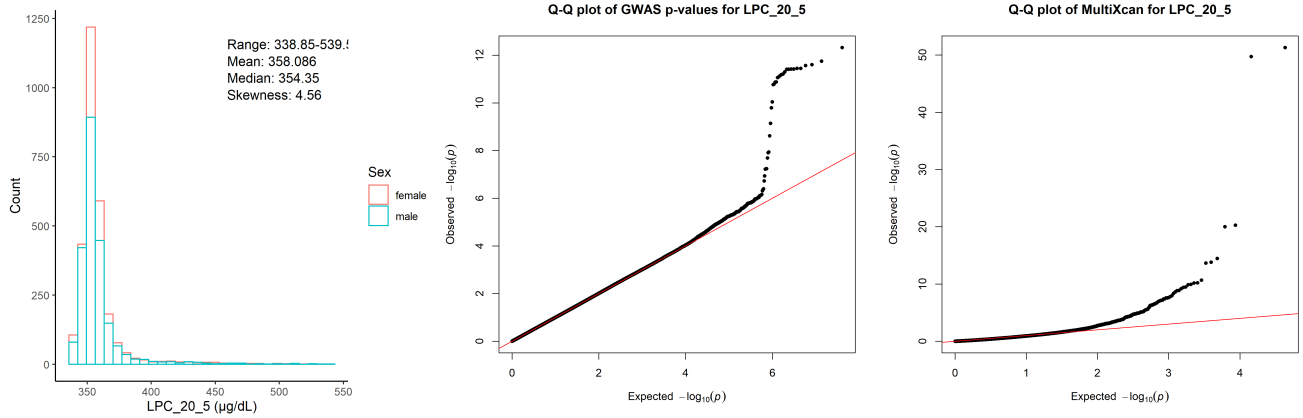
Chr11_59978355_62914375

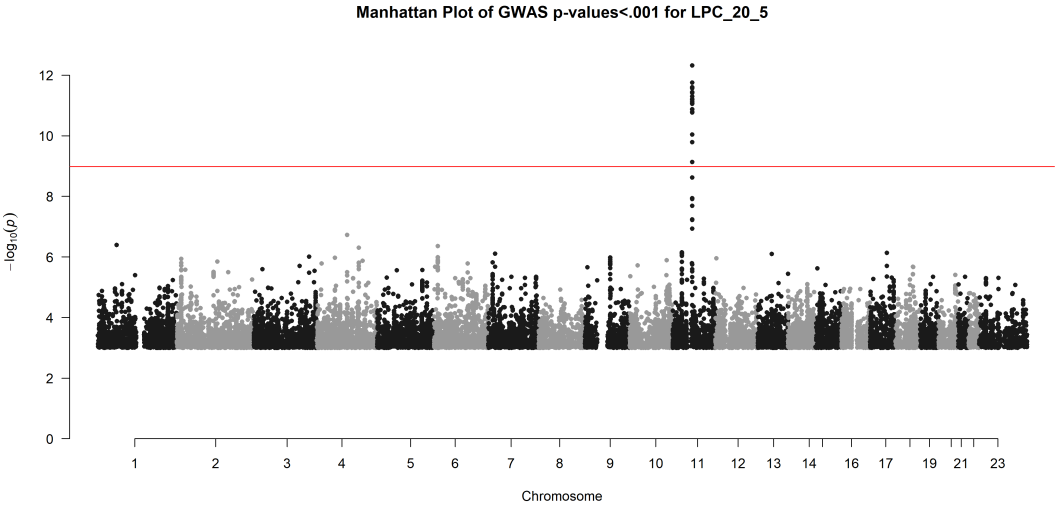


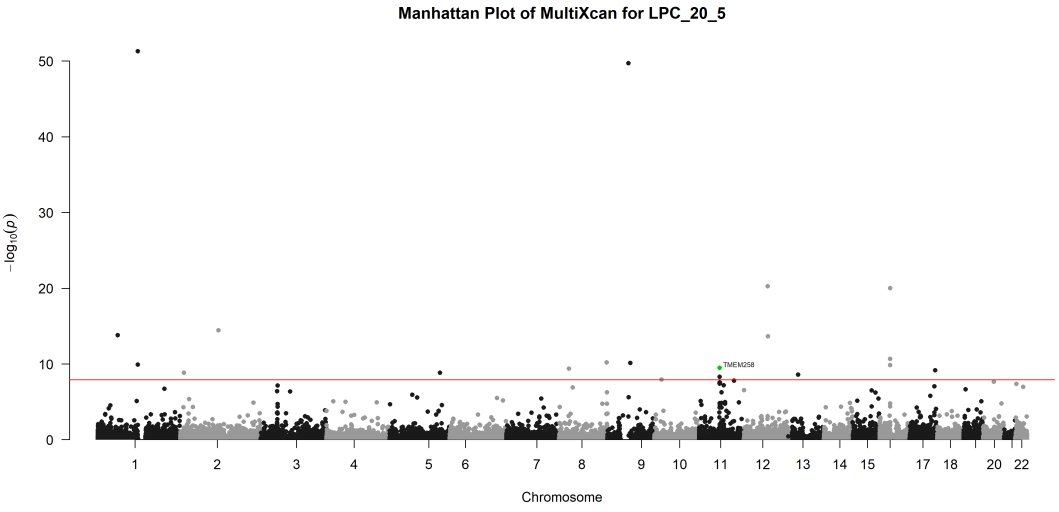
Chr16_14393707_15885866



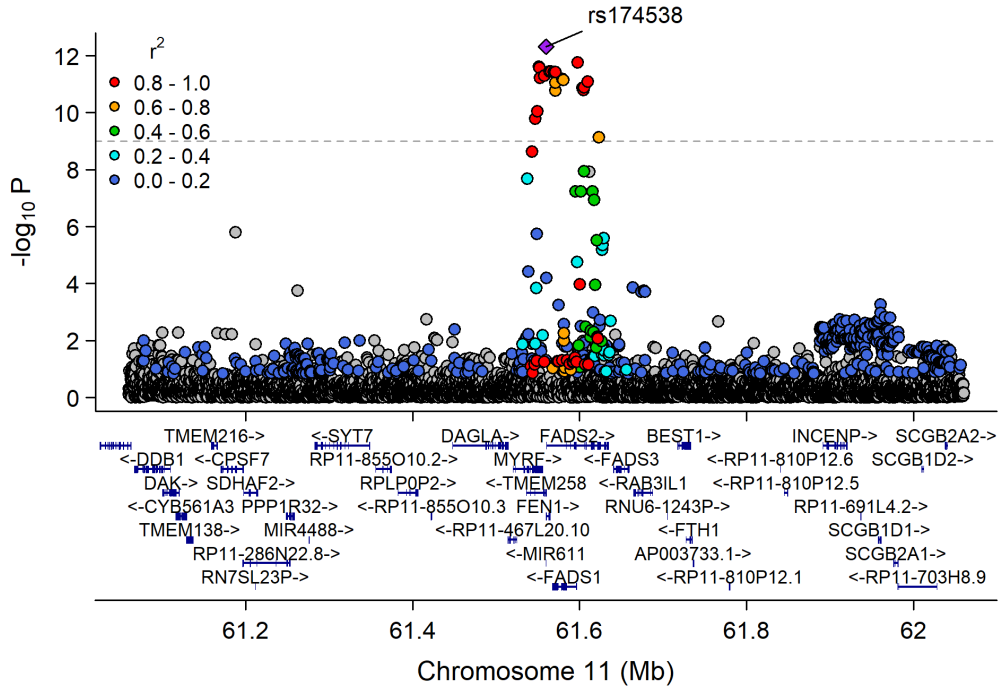
Lysophosphatidylcholine acyl C20:5 (µg/dL)



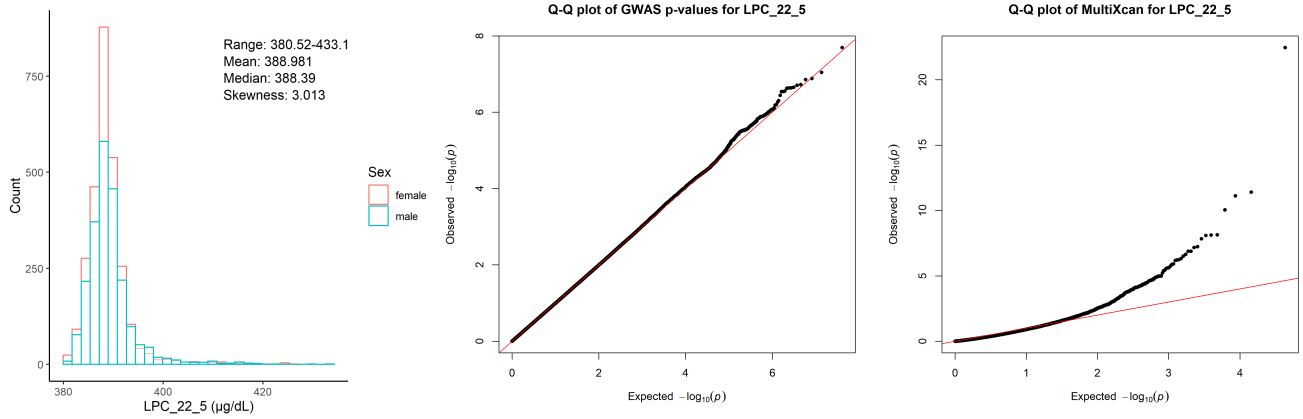


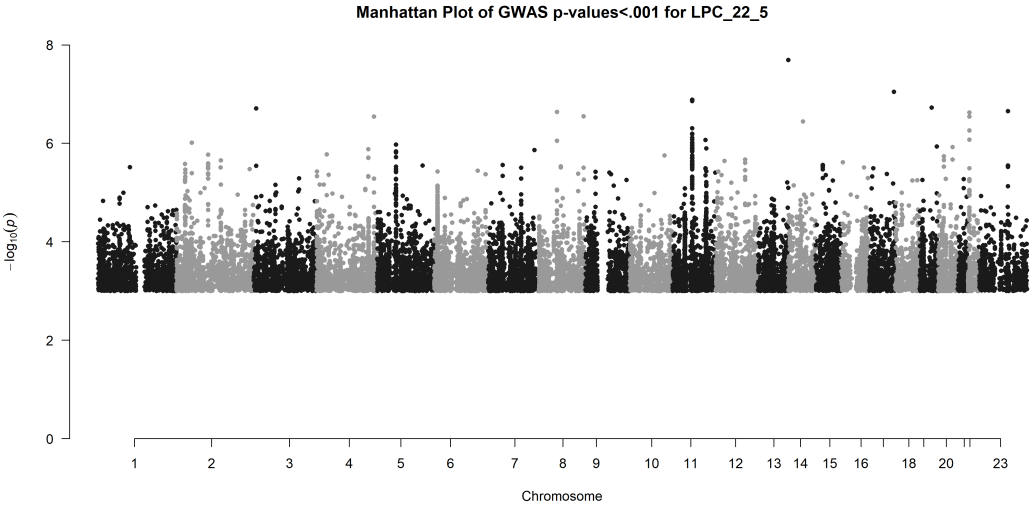


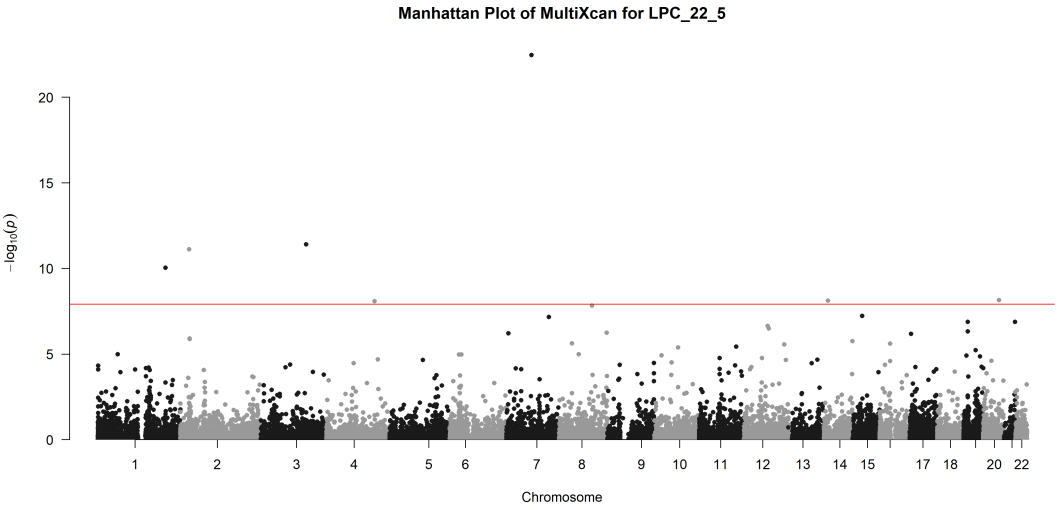
Chr11__59978355__62914375



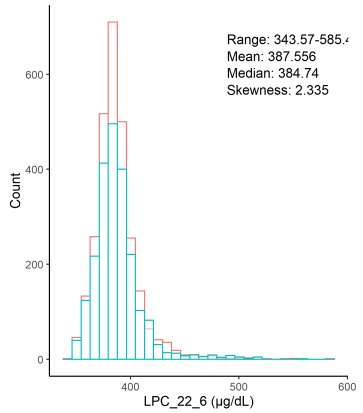
Lysophosphatidylcholine acyl C22:5 (µg/dL)

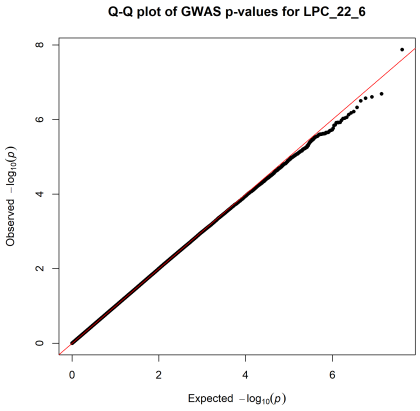


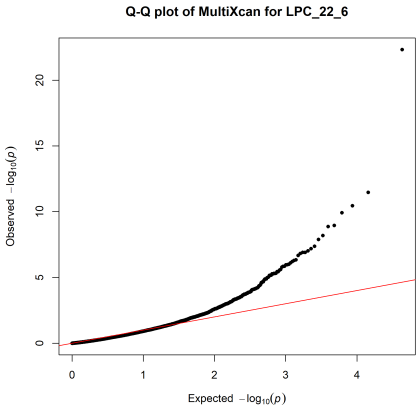


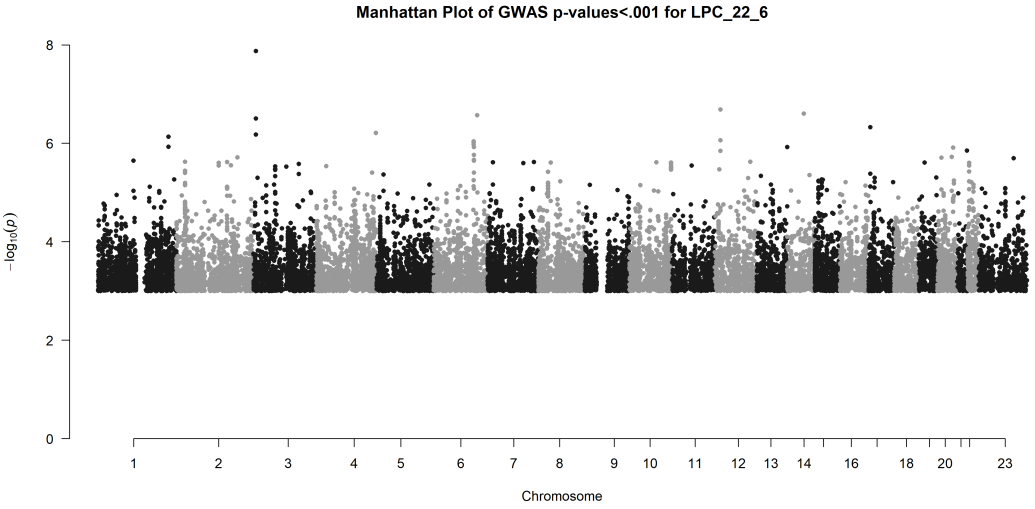


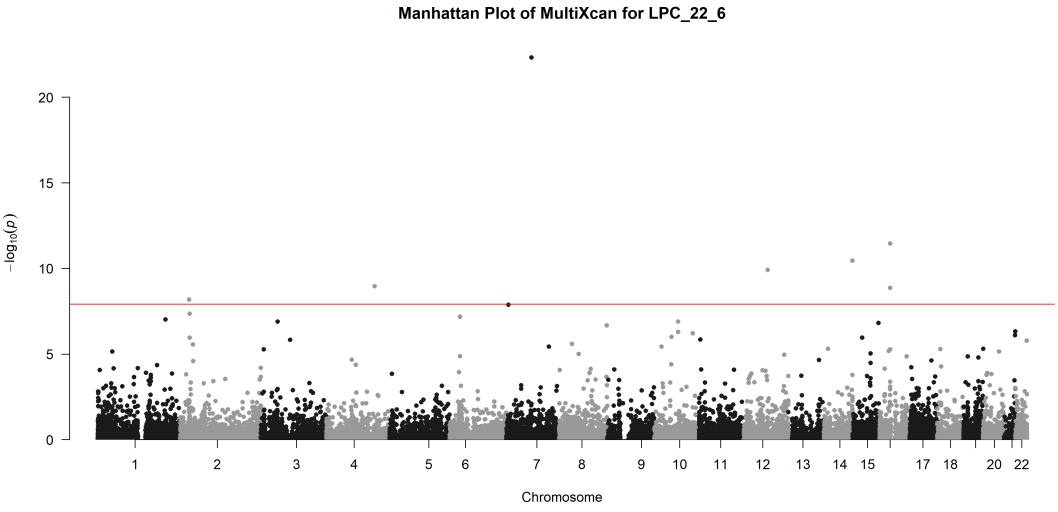
Lysophosphatidylcholine acyl C22:6 (µg/dL)



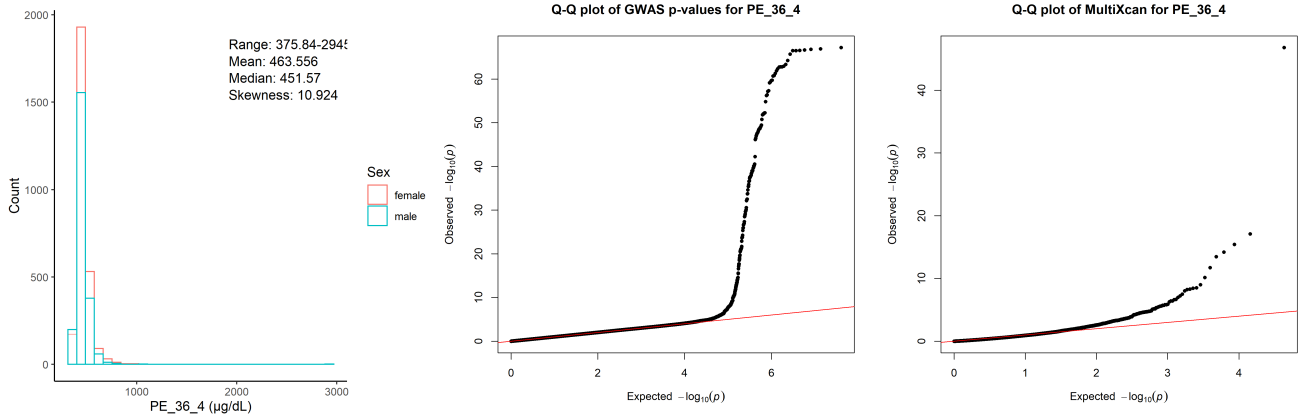


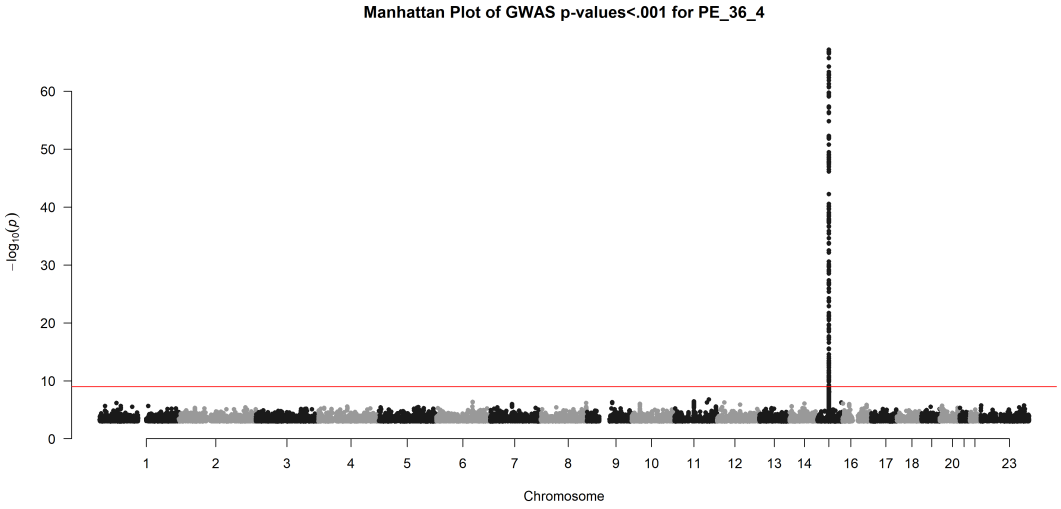


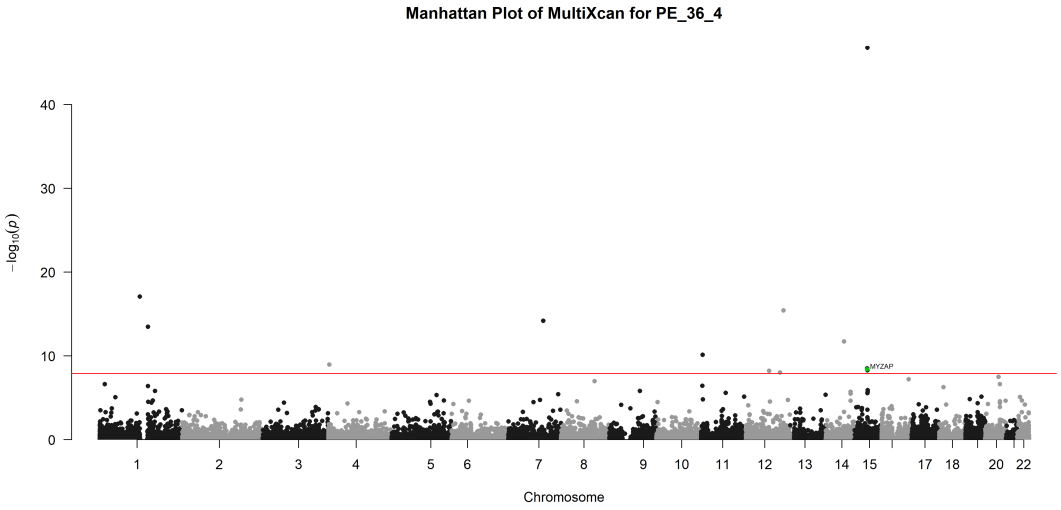




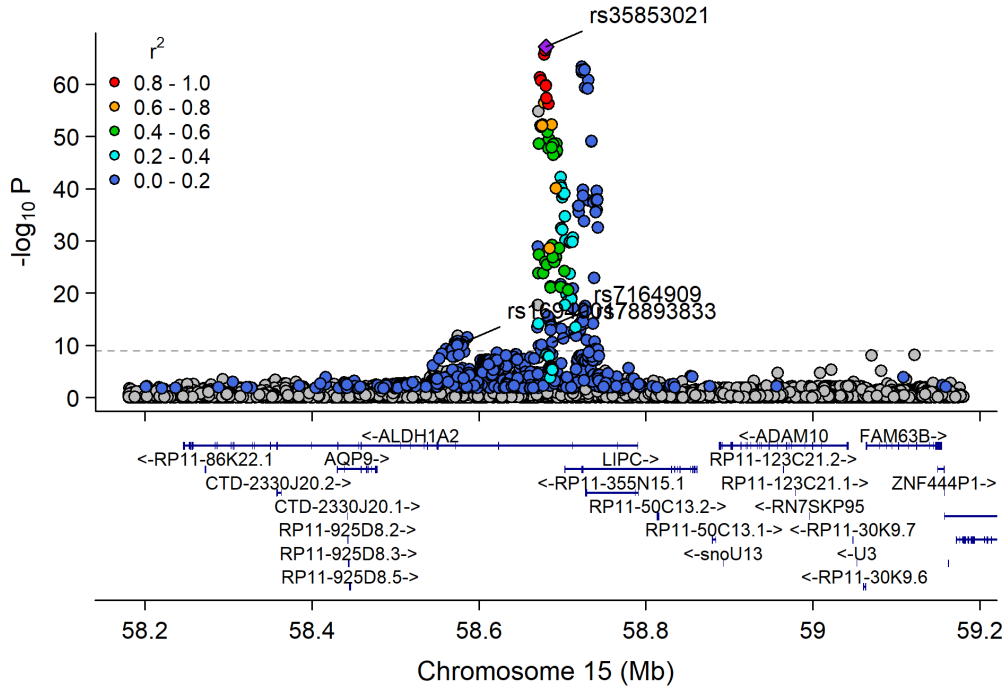
Phosphatidylethanolamine acyl C36:4 (µg/dL)



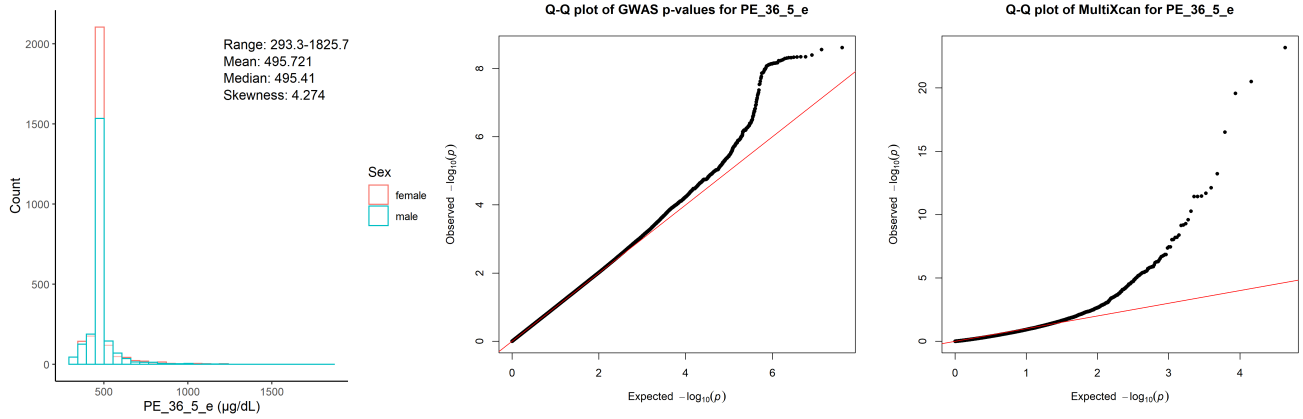


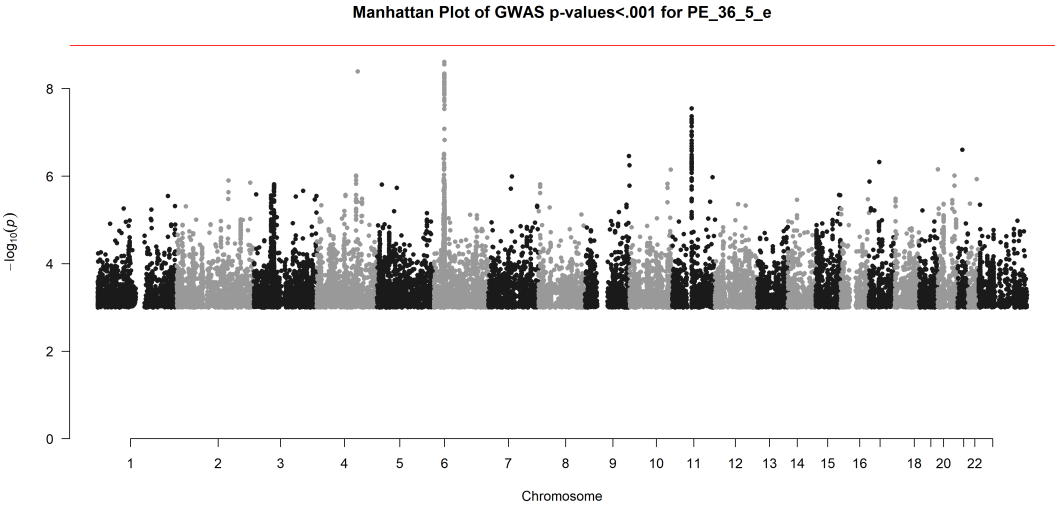


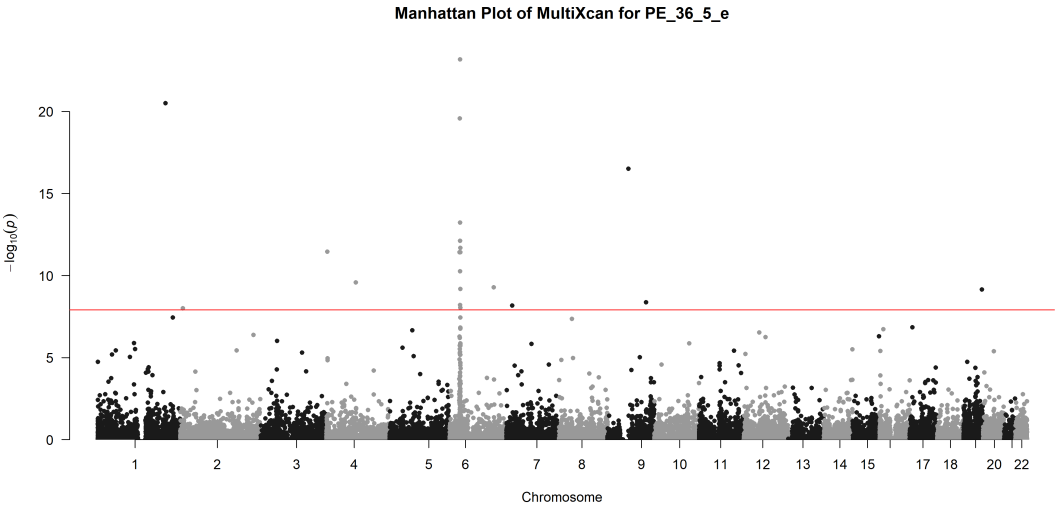
Chr15_57658798_60490883



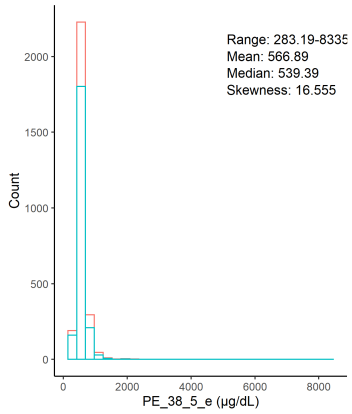
Phosphatidylethanolamine alkyl C36:5 (µg/dL)

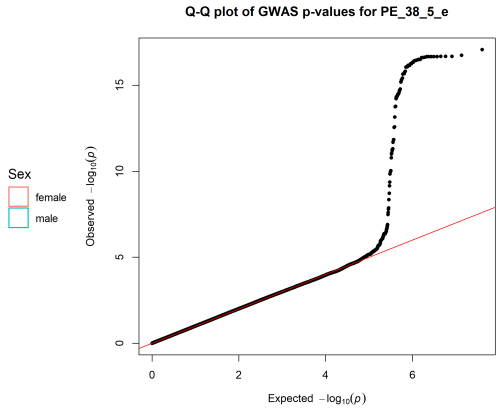


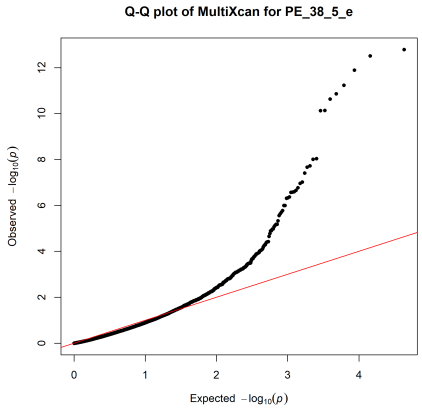


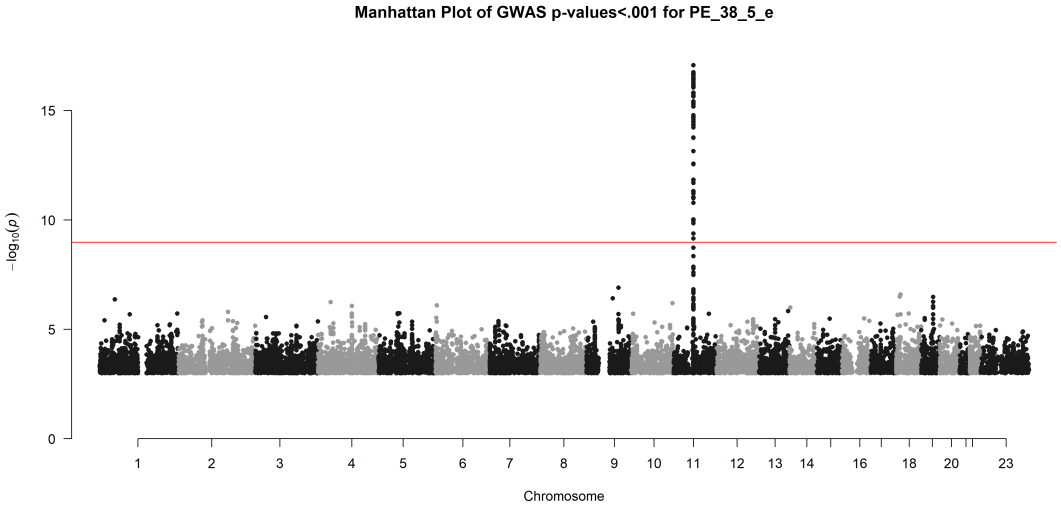


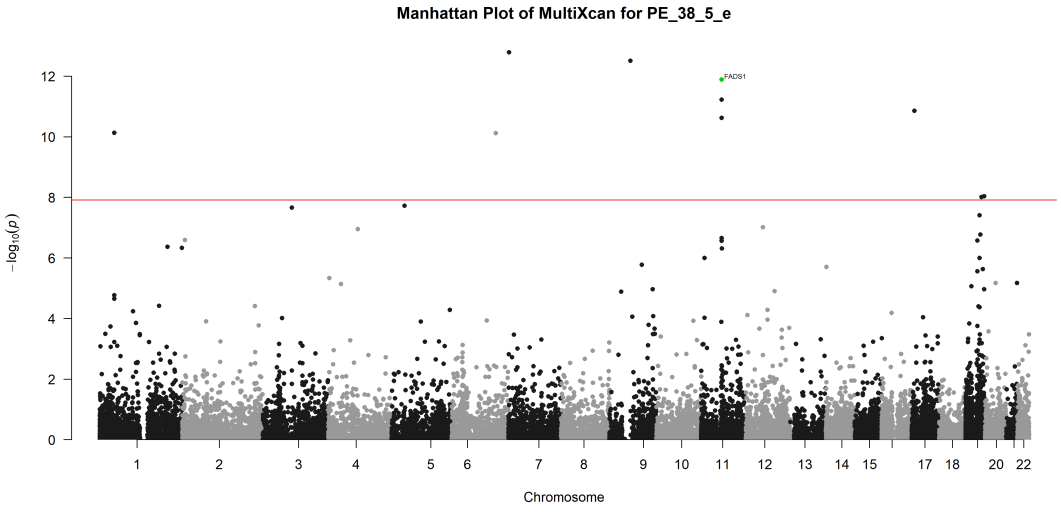
Phosphatidylethanolamine alkyl C38:5 (µg/dL)



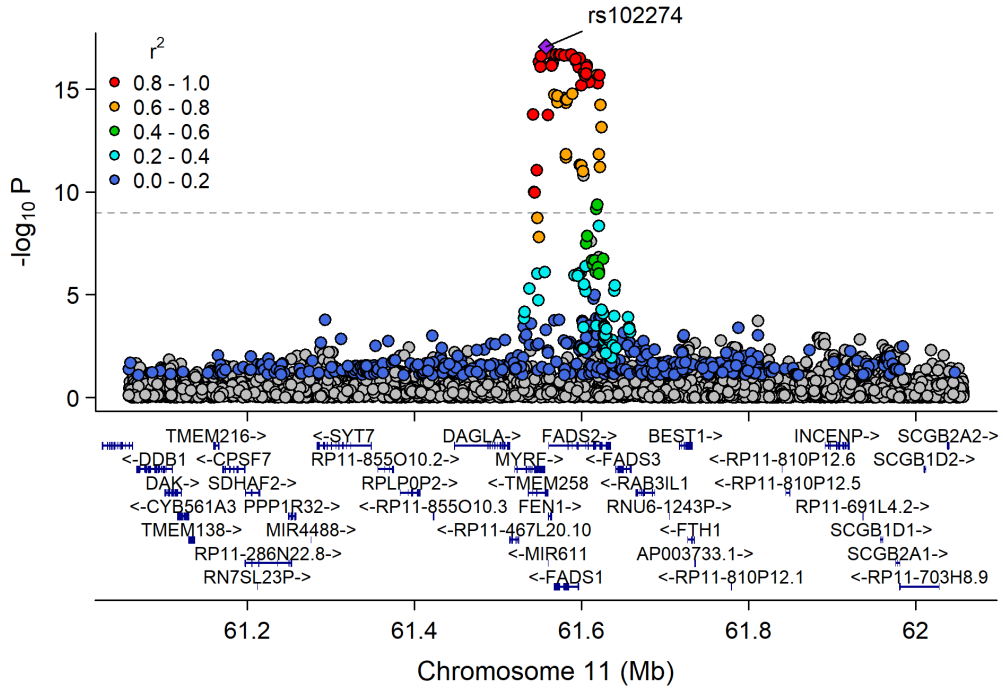




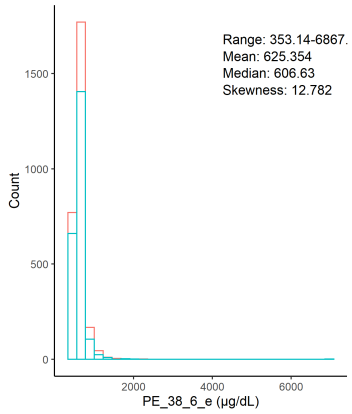


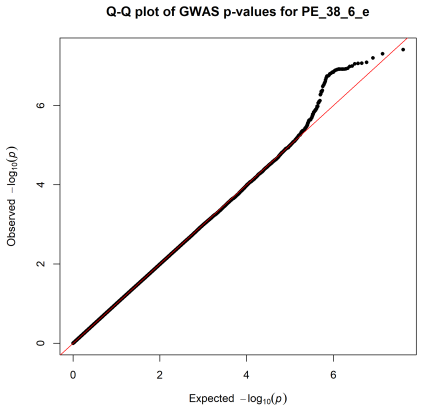


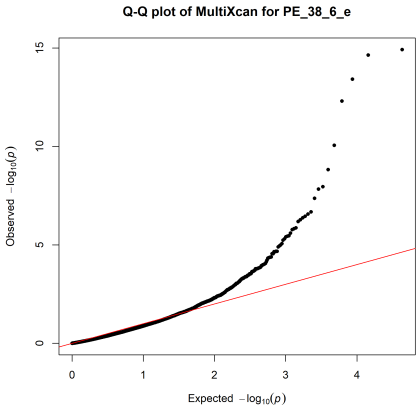
Chr11_59978355_62914375

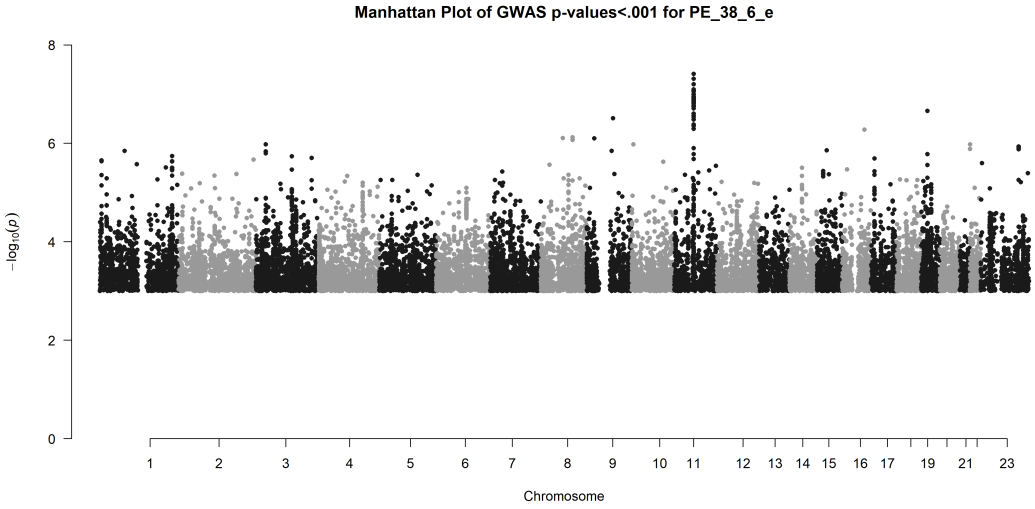


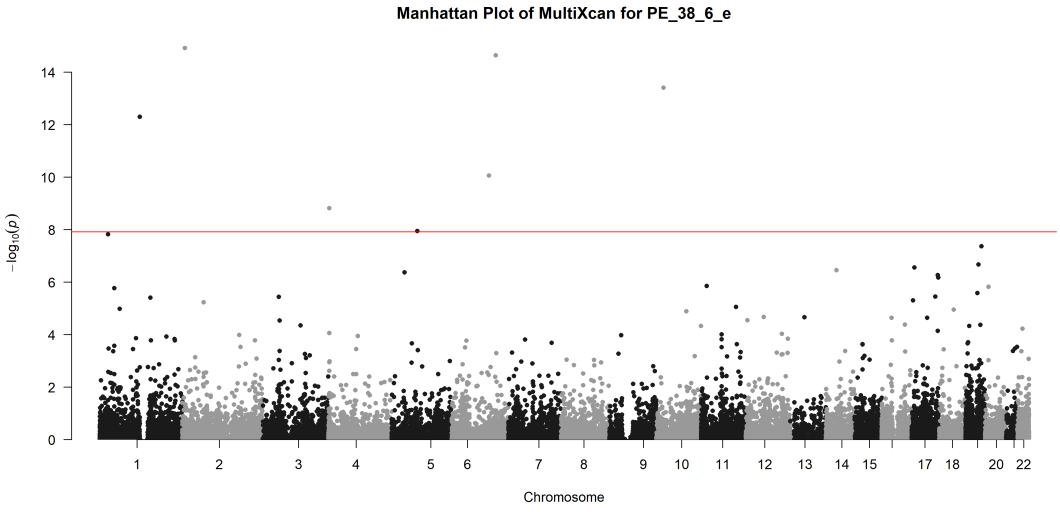
Phosphatidylethanolamine alkyl C38:6 (µg/dL)





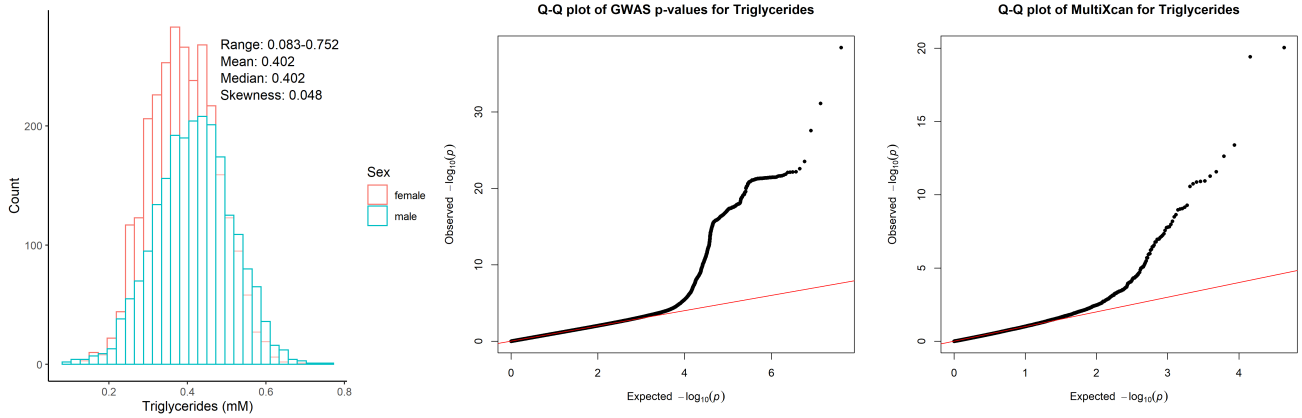


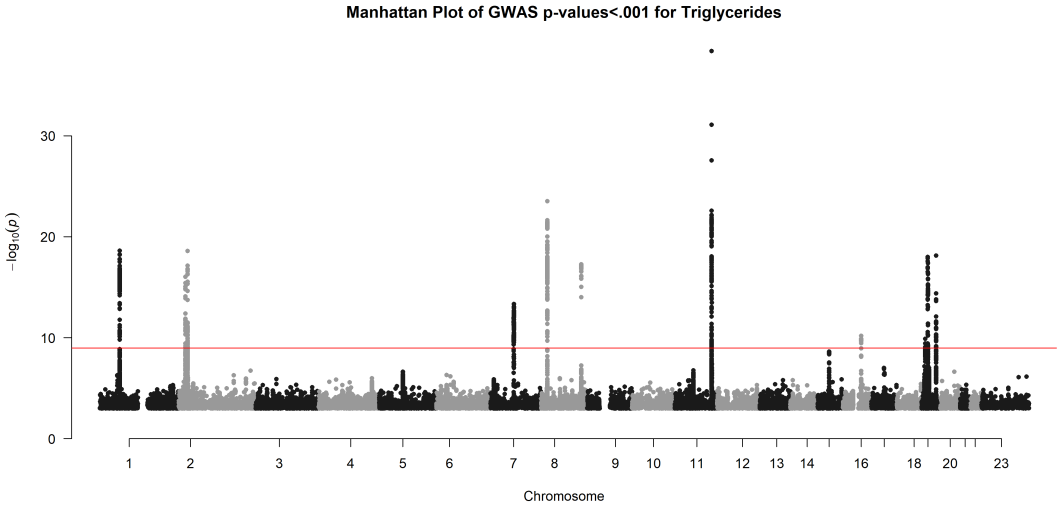


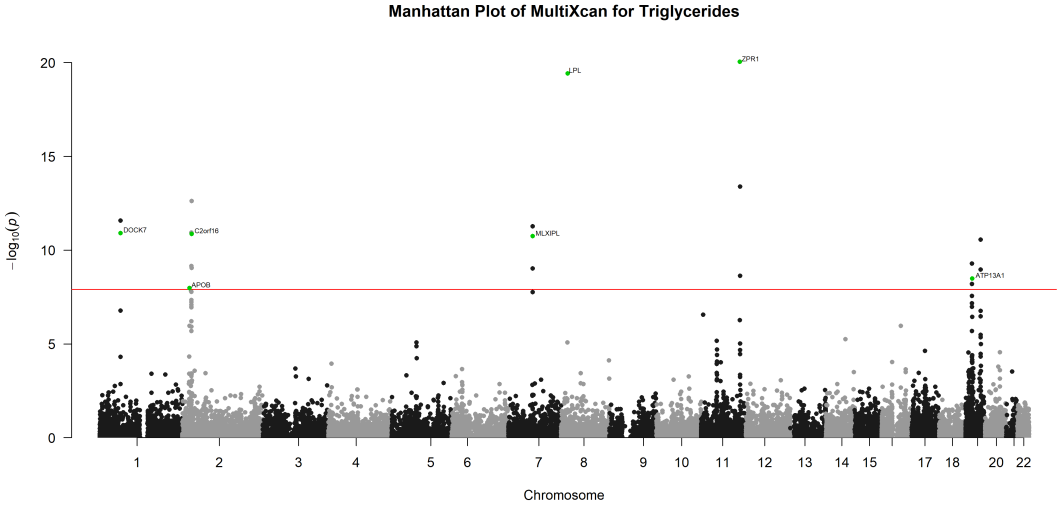


Glycerolipids

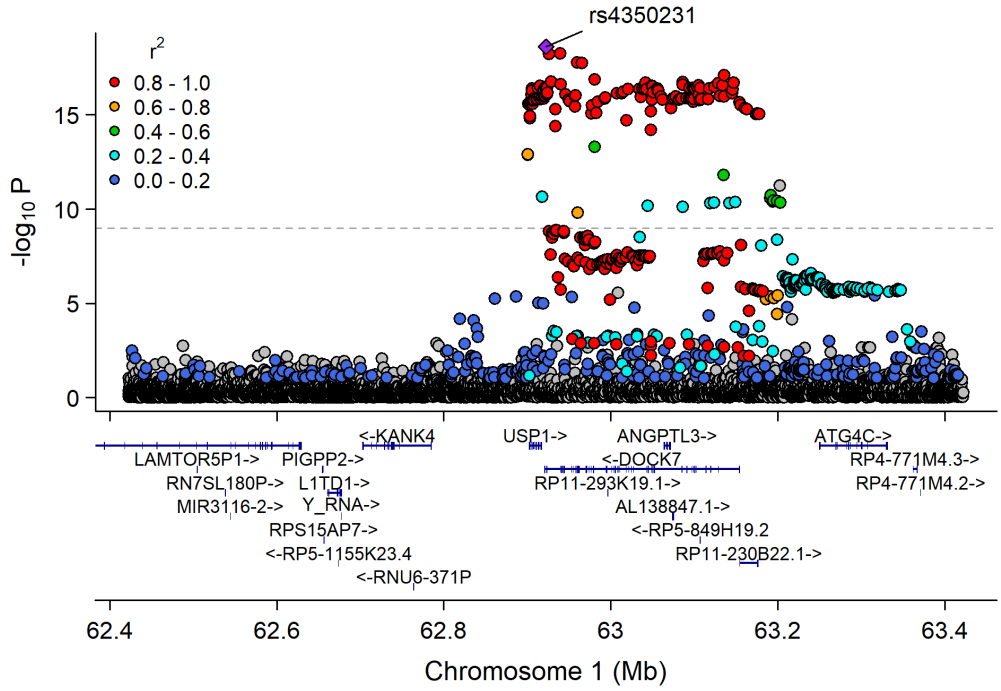
Triglycerides (mM)



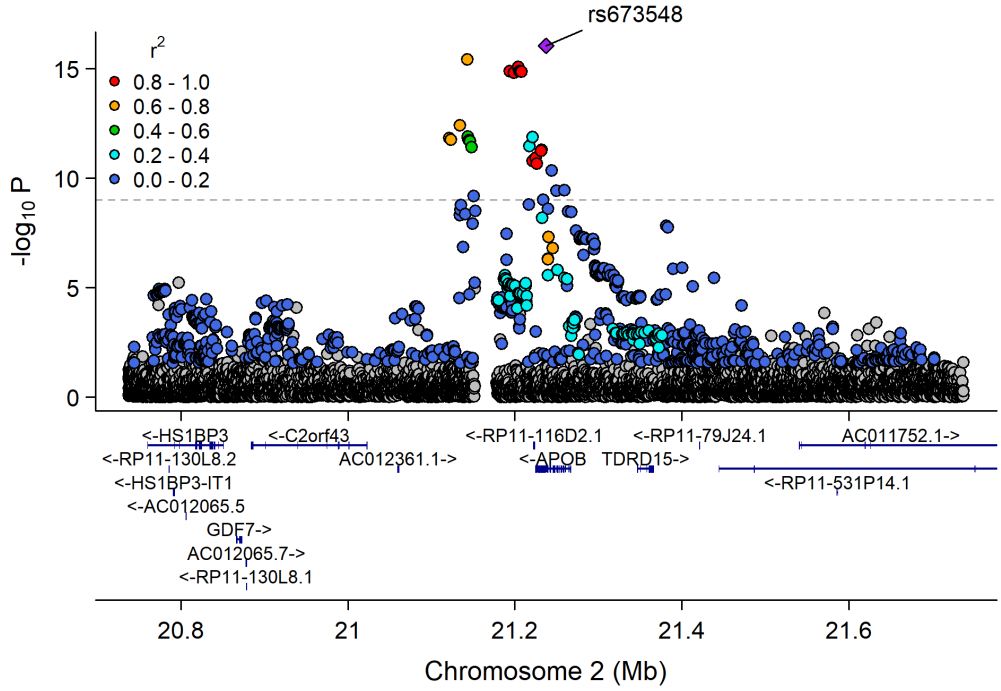




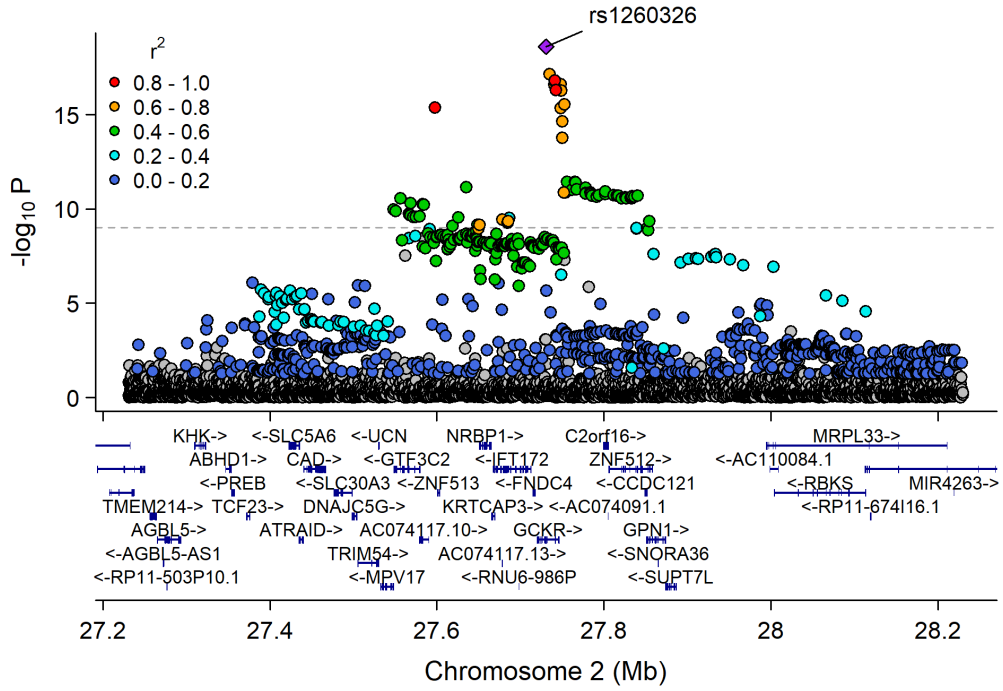
Chr1_62833402_63375116



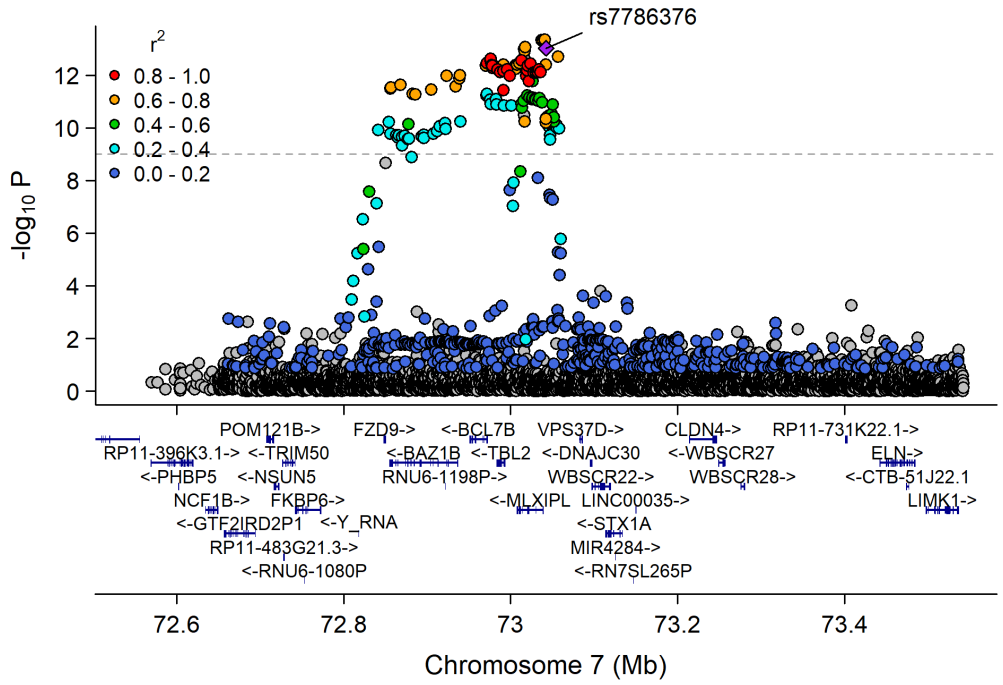
Chr2_19947287_22427492



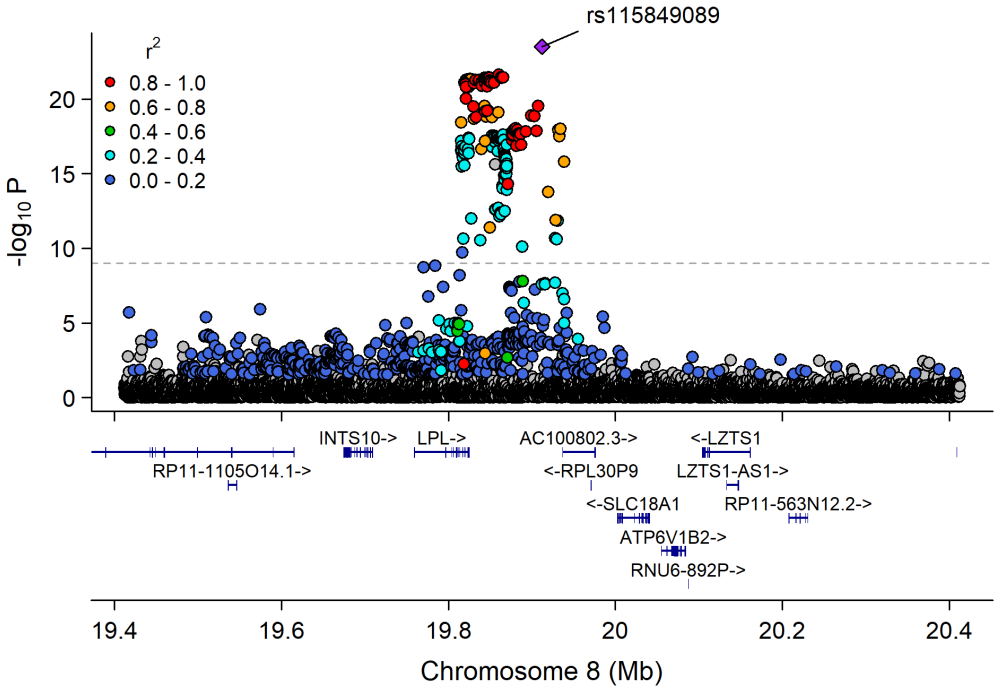
Chr2_26753815_28597624



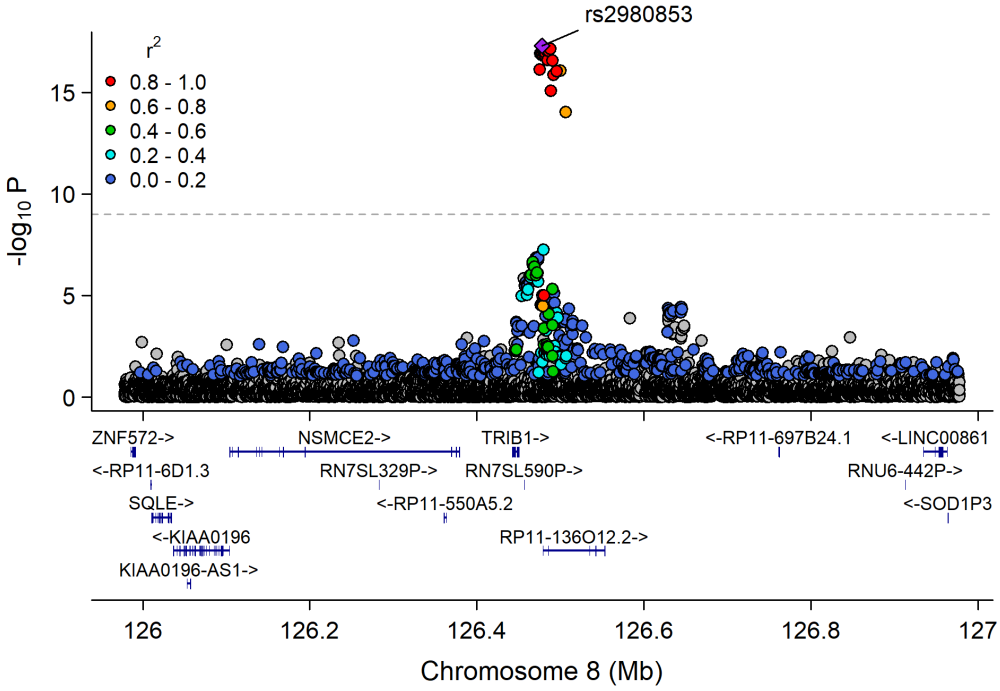
Chr7_72070325_73133300



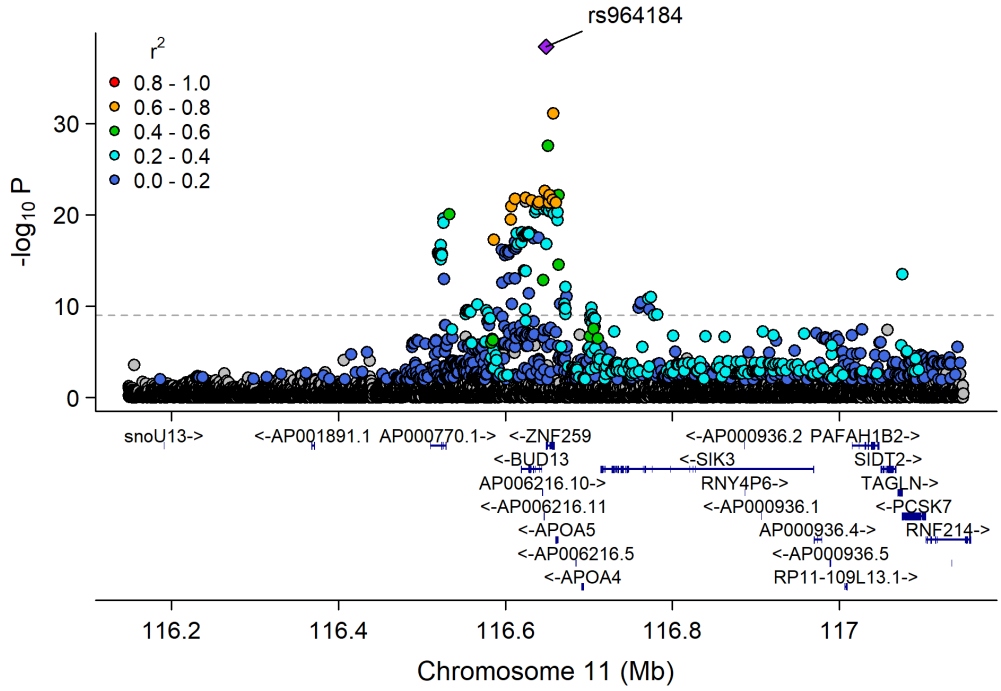
Chr8__18582620__20792744



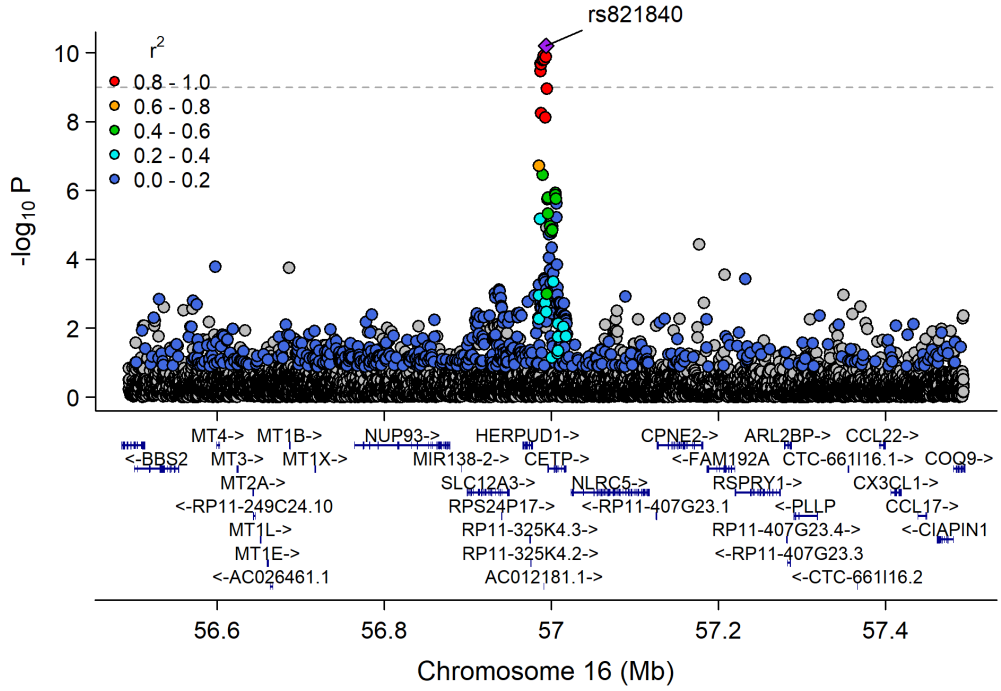
Chr8__126435663__126578061



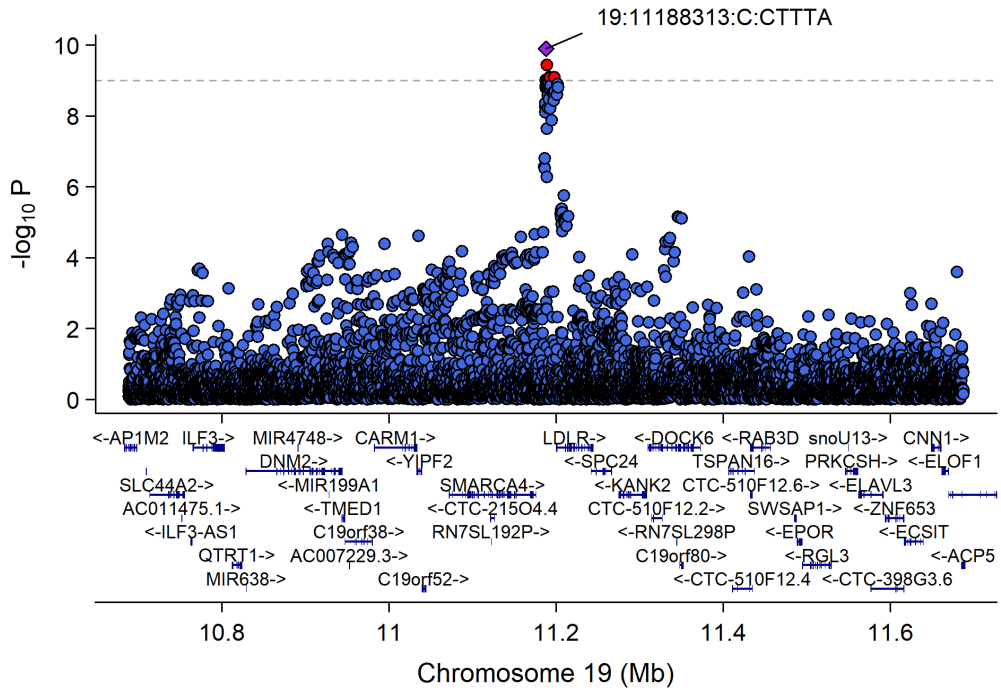
Chr11_115541901_117633315



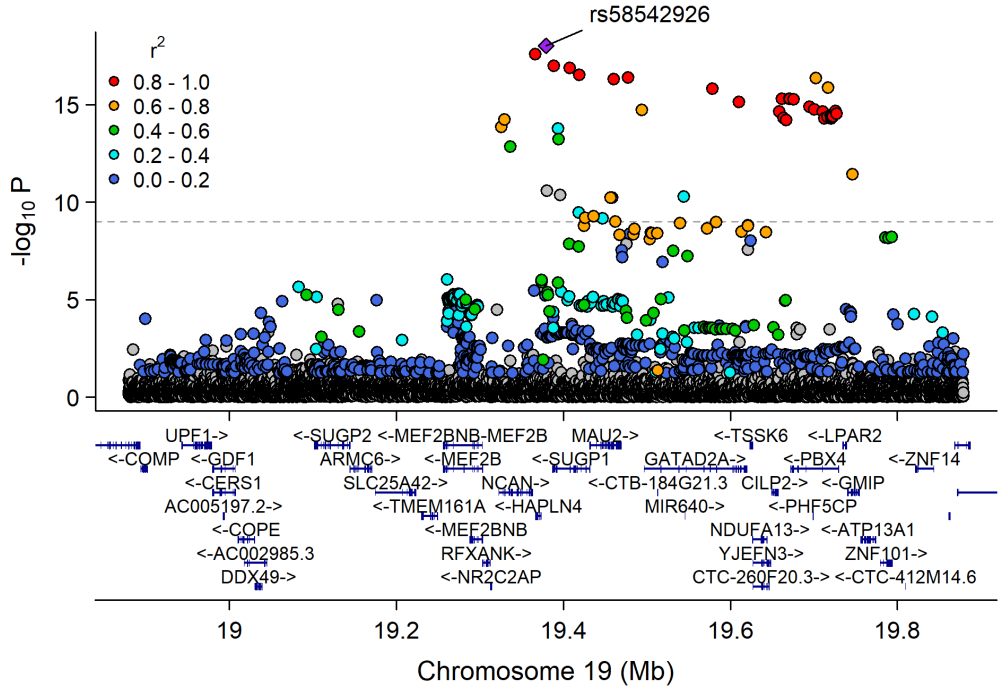
Chr16_55870822_57992421



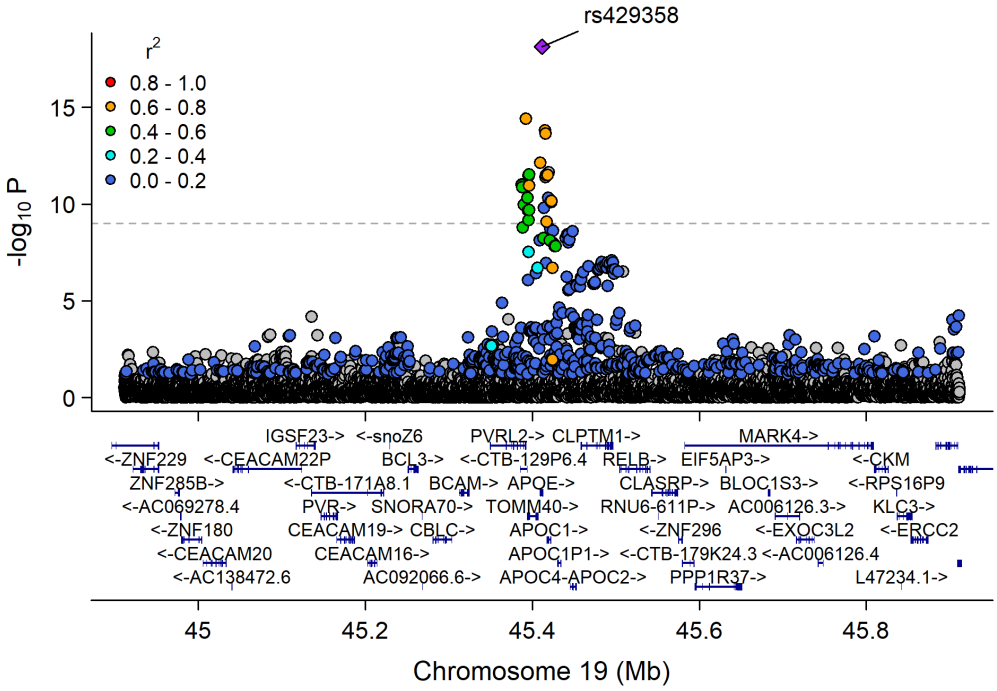
Chr19_9807999_12255594



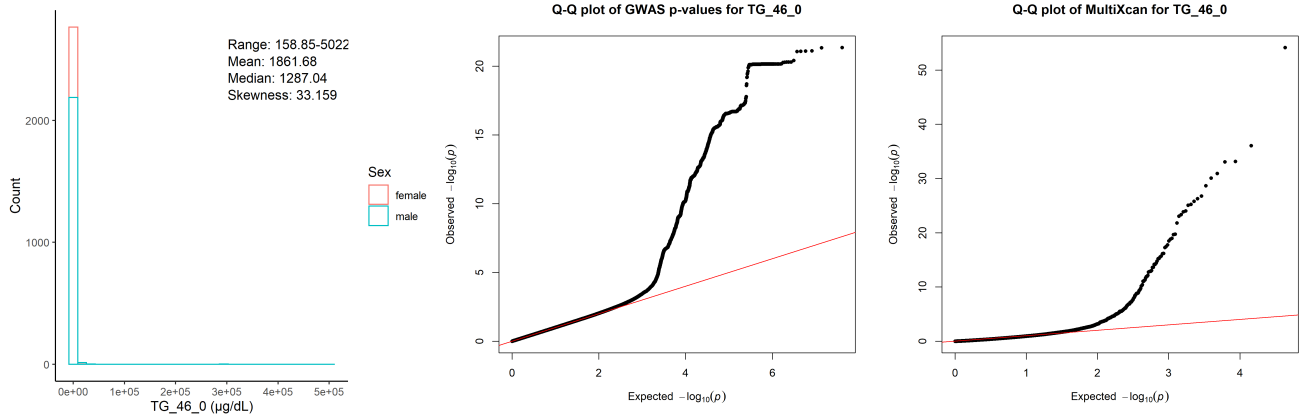
Chr19_18370495_20841464

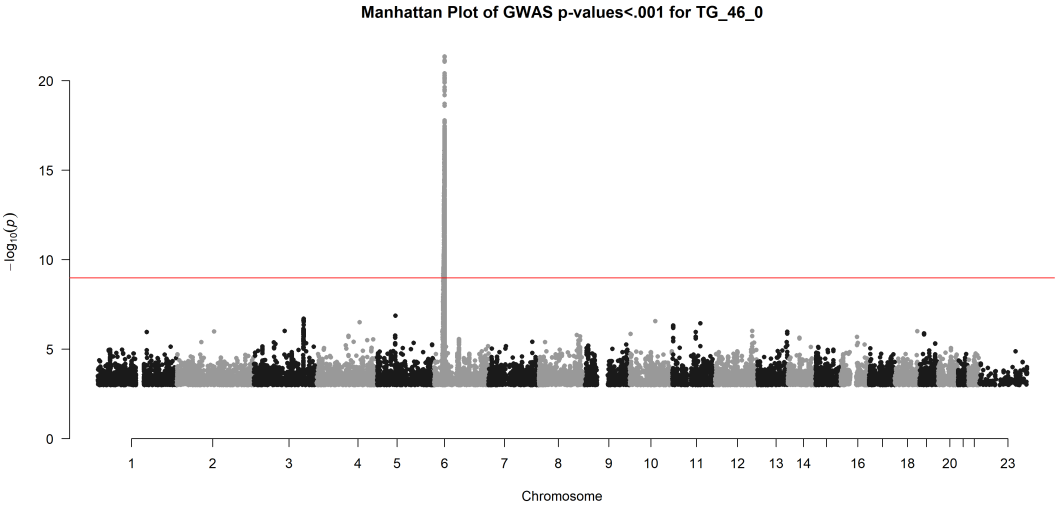


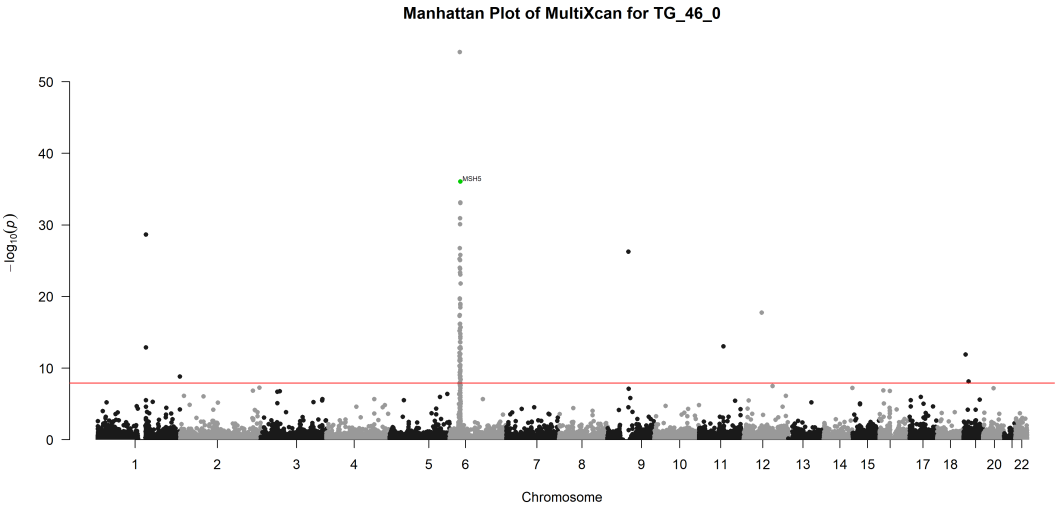
Chr19__44063363__46637375



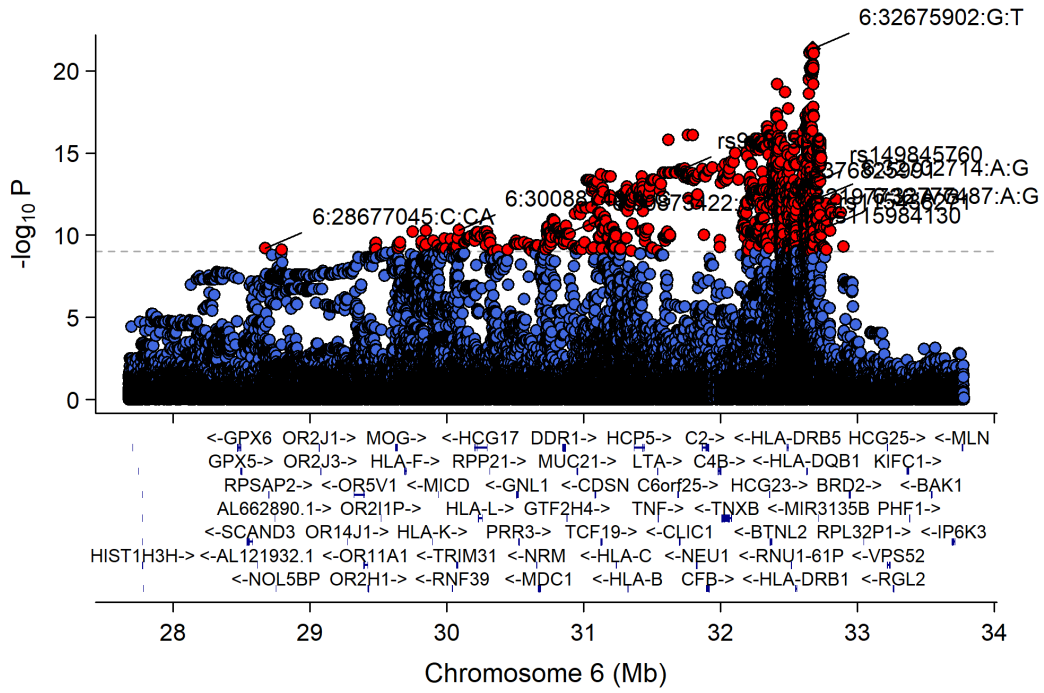
Triglycerol C46:0 (µg/dL)



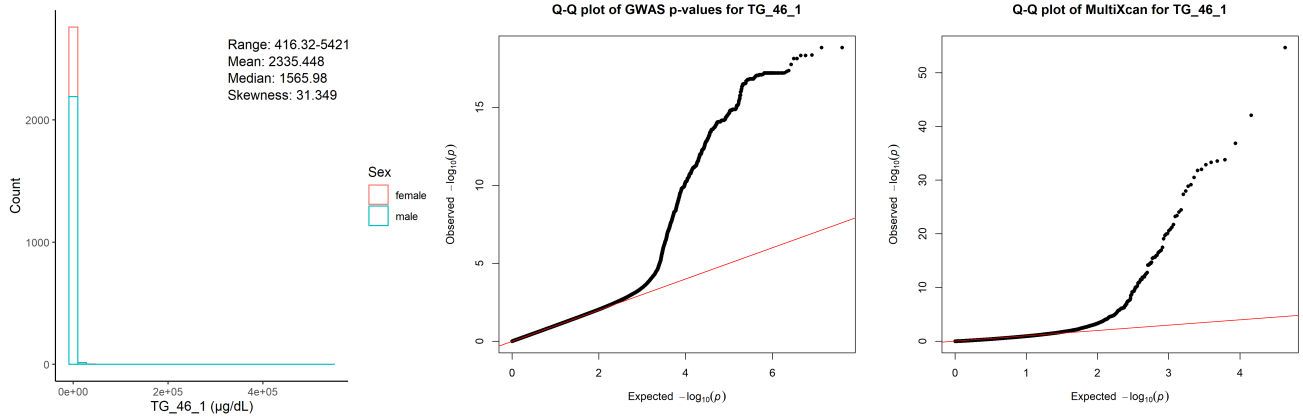


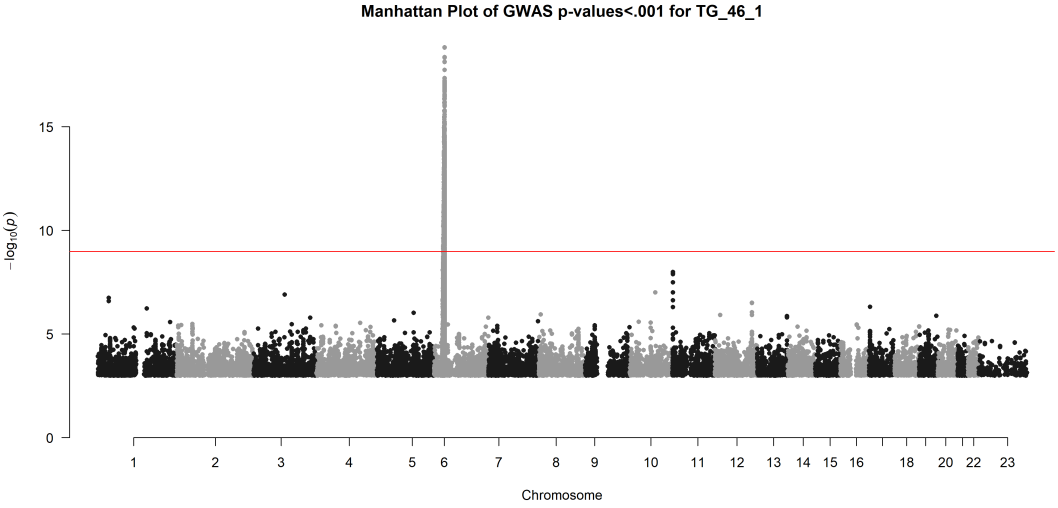


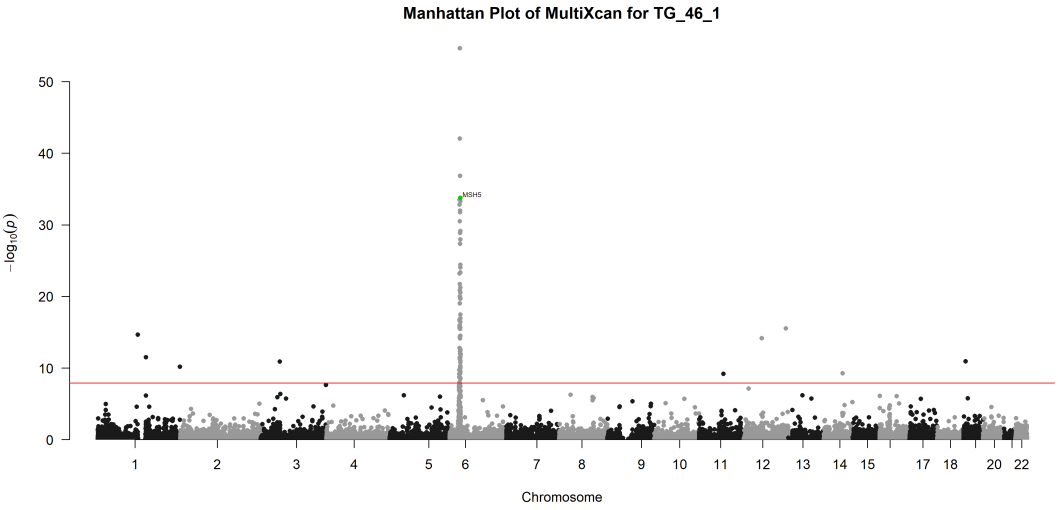
Chr6__24042708__34003583



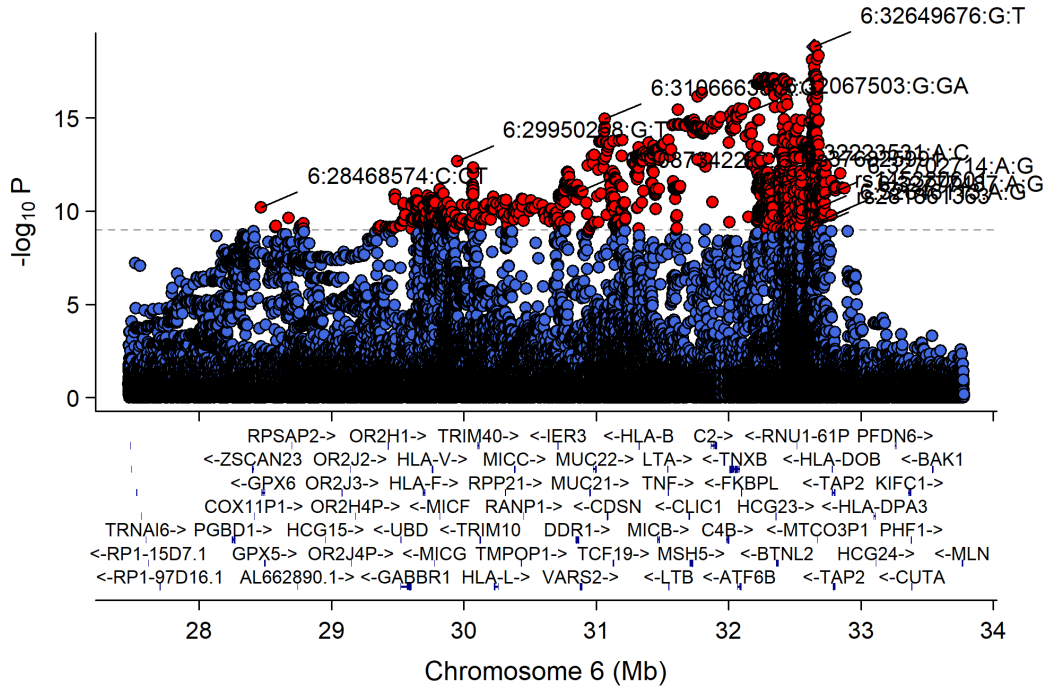
Triglycerol C46:1 (µg/dL)



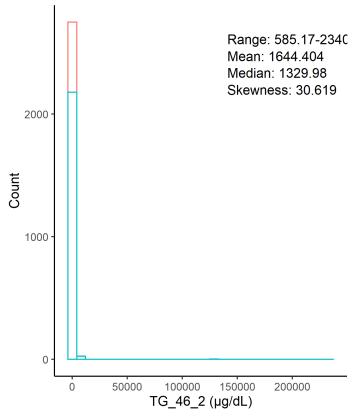


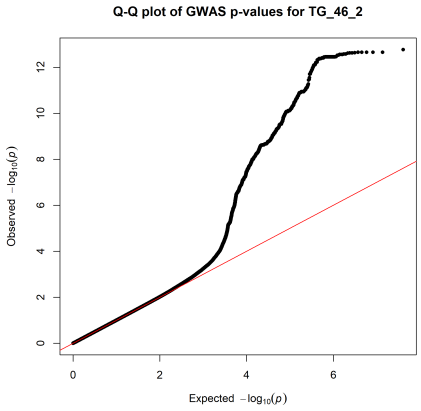


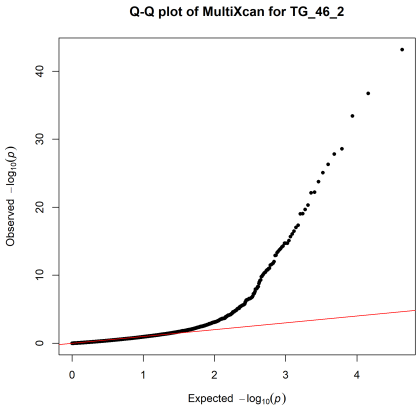
Chr6__24042708__34003583

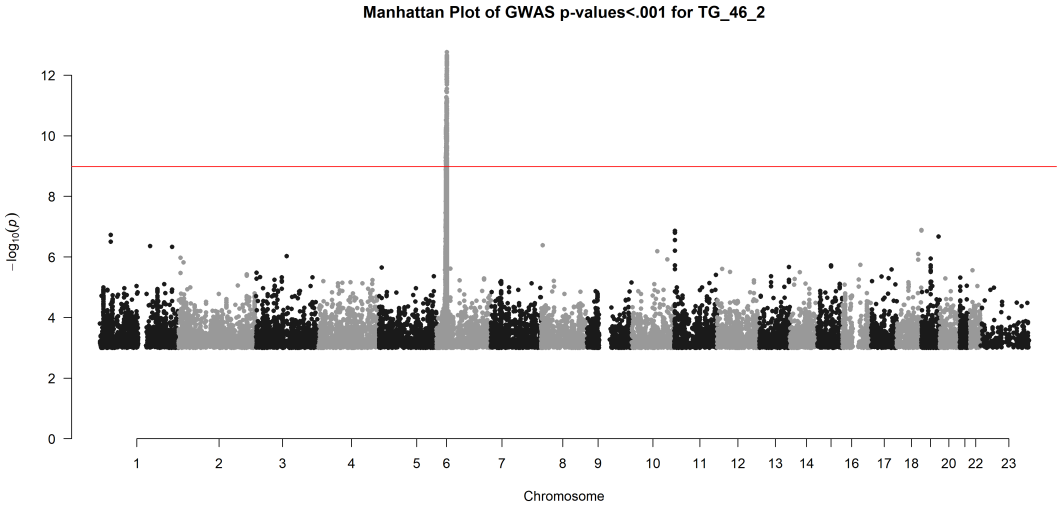


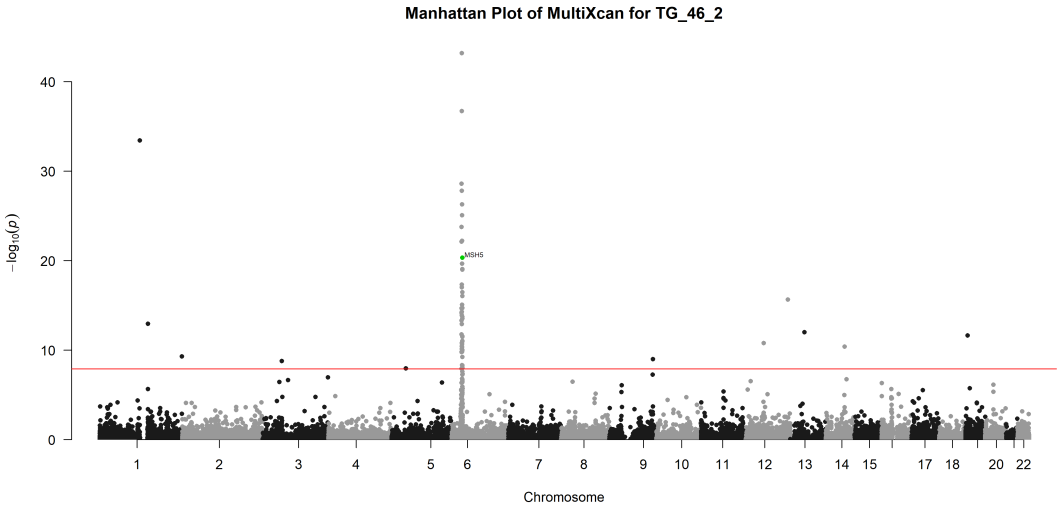
Triglycerol C46:2 (µg/dL)



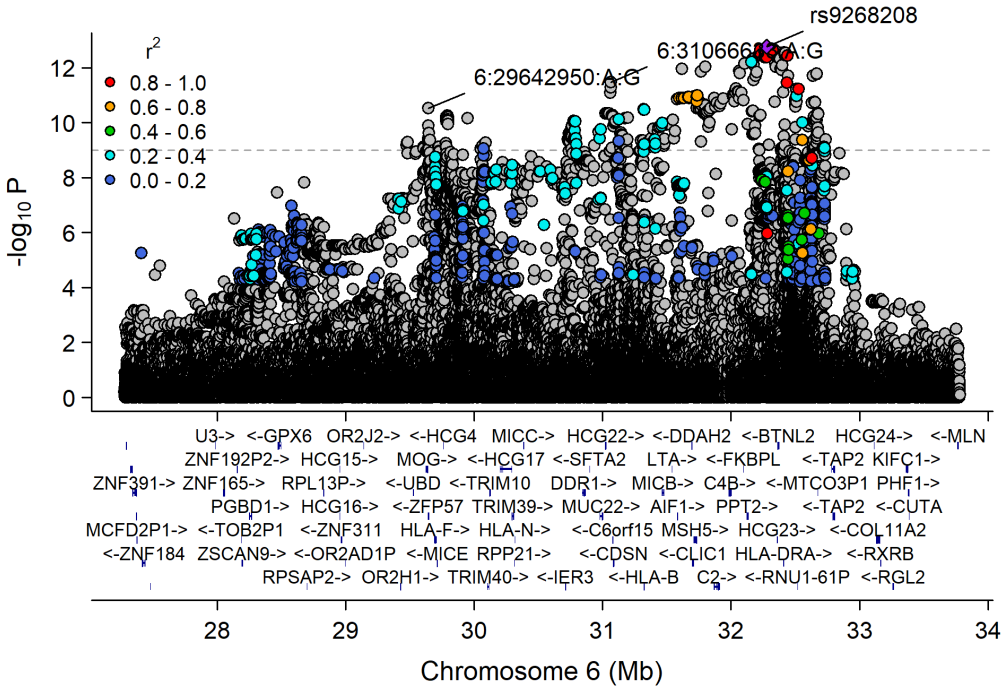




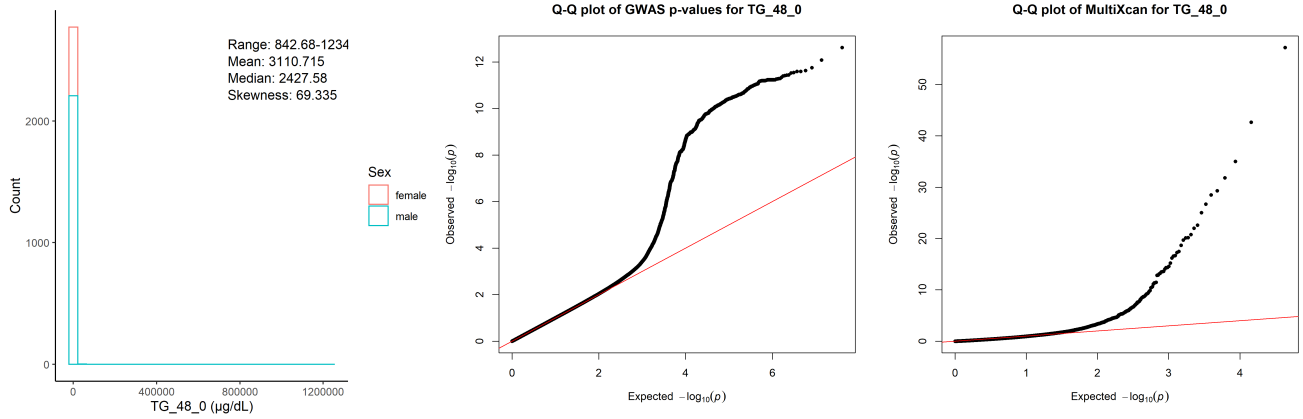


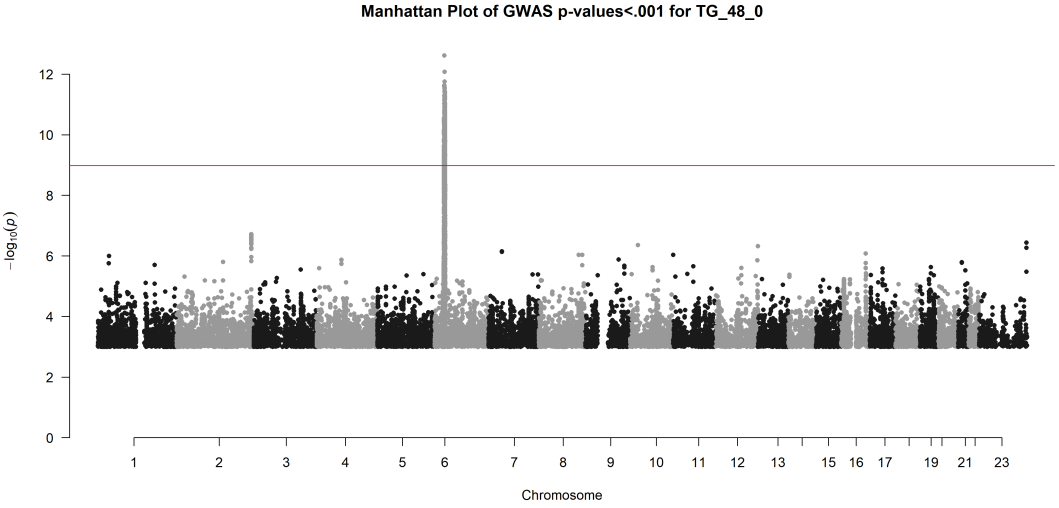


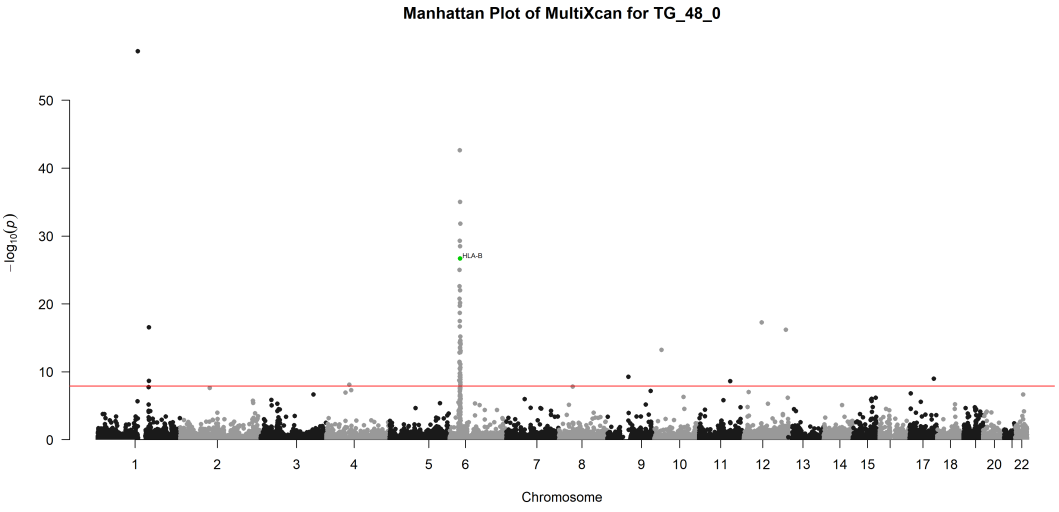
Chr6__24042708__34003583



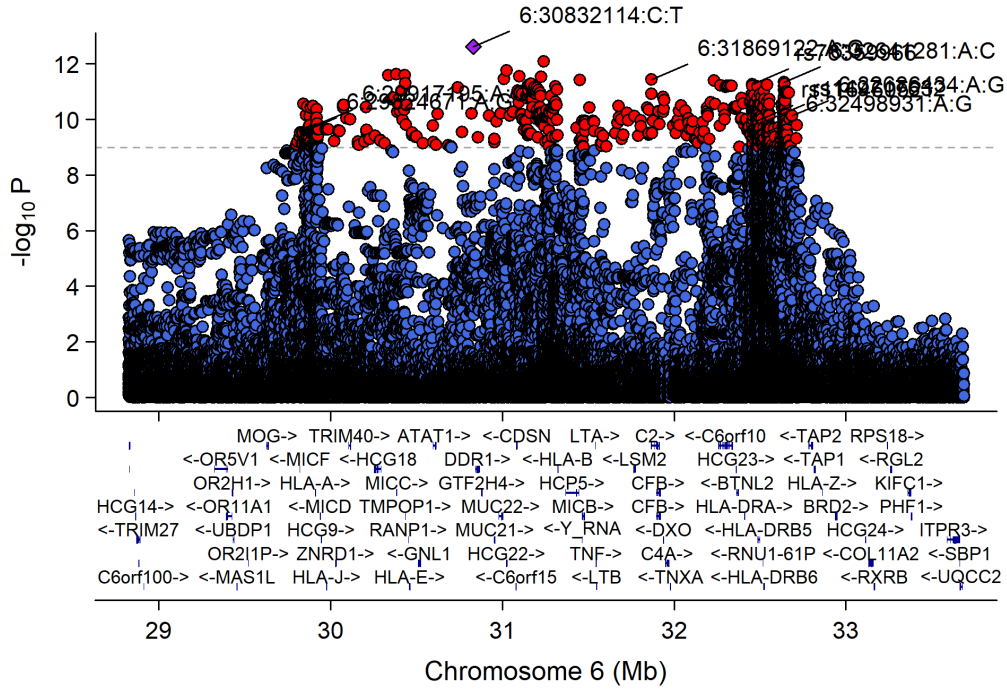
Triglycerol C48:0 (µg/dL)



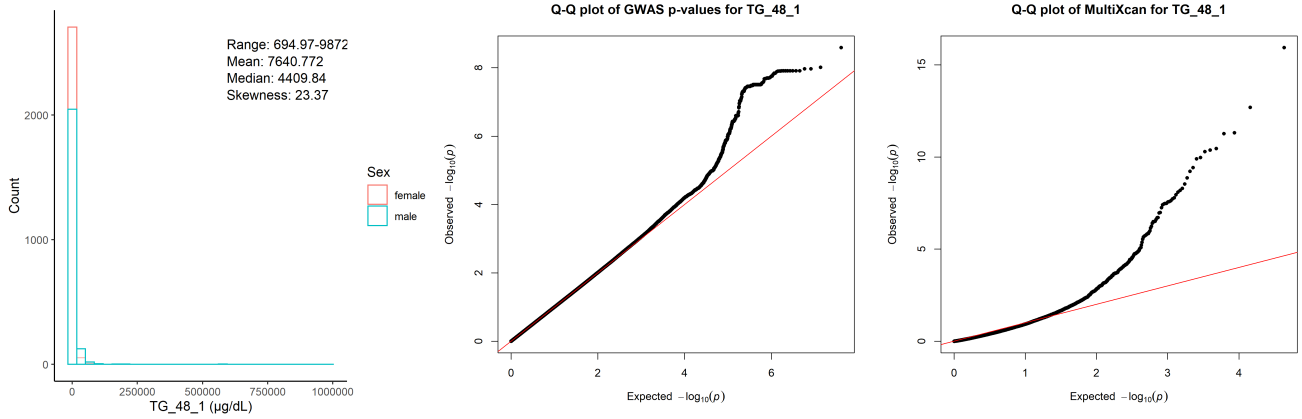


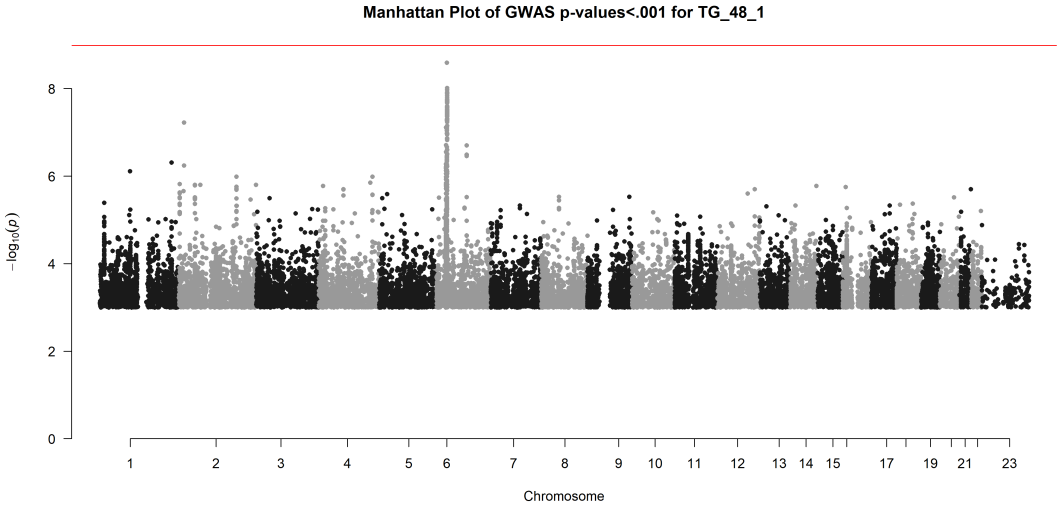


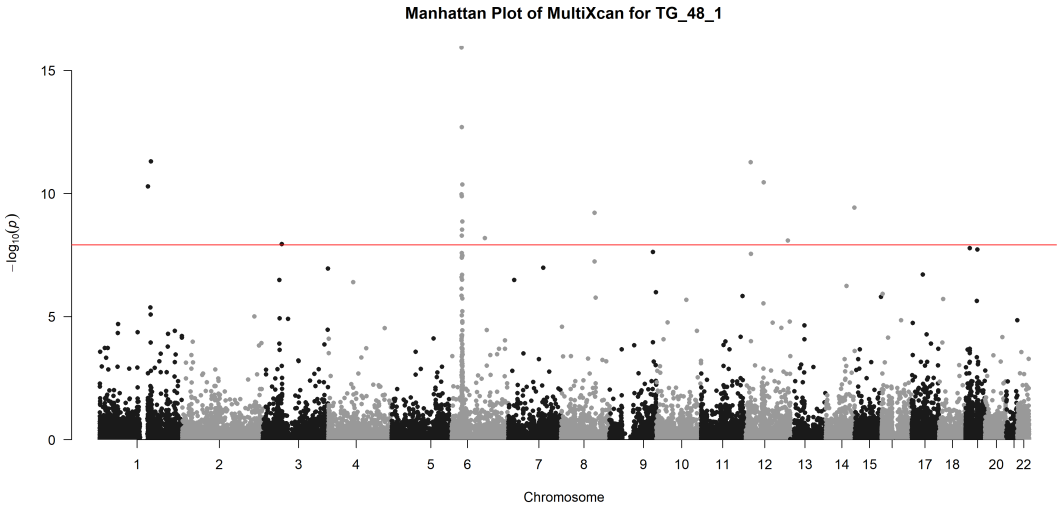
Chr6__24042708__34003583



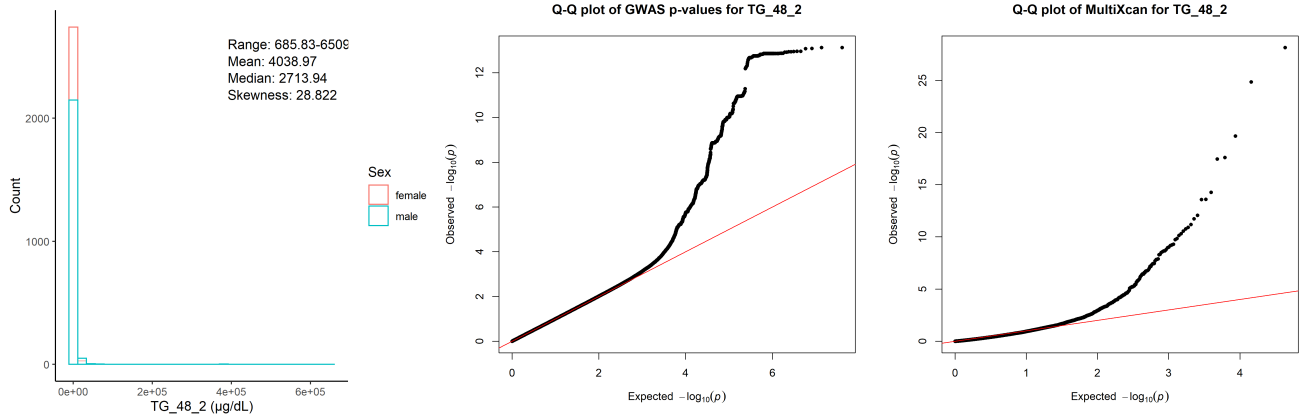
Triglycerol C48:1 (µg/dL)

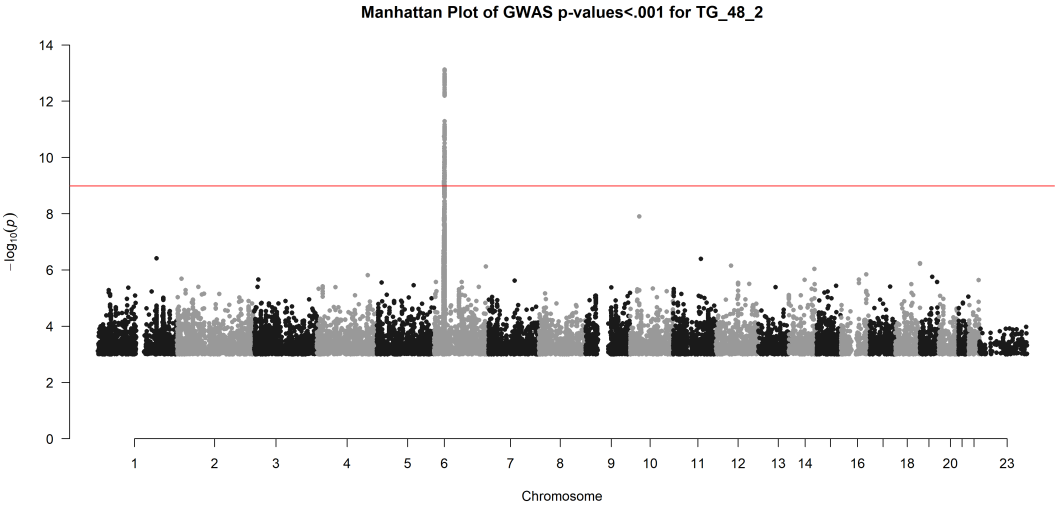


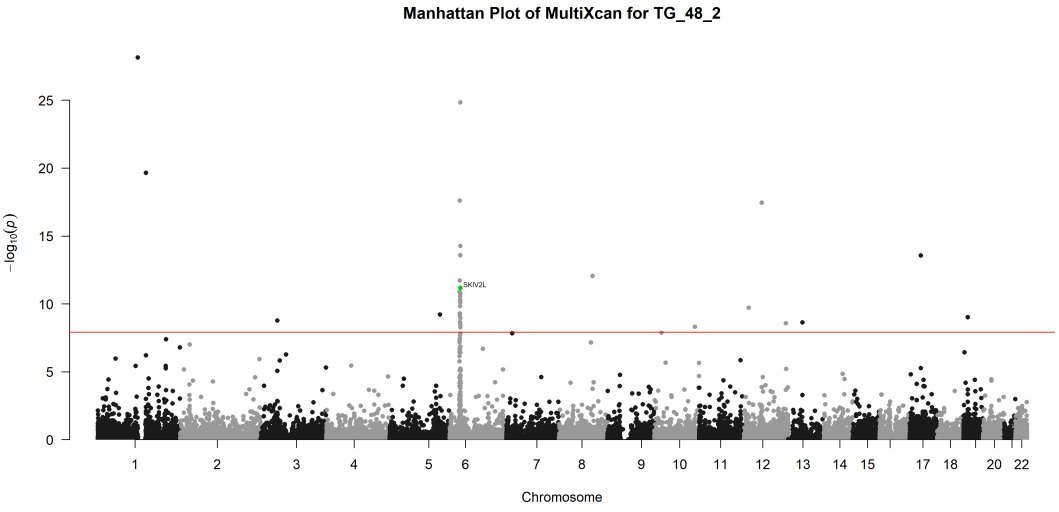




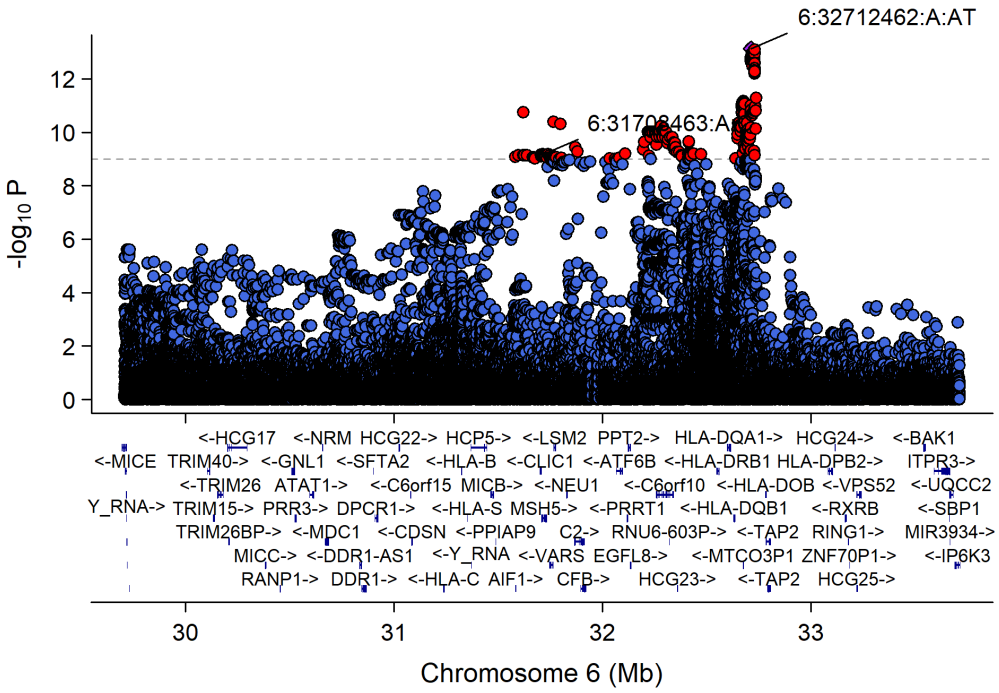
Triglycerol C48:2 (µg/dL)



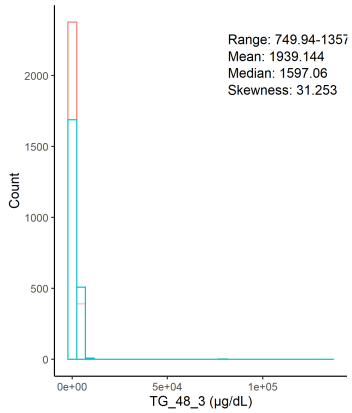


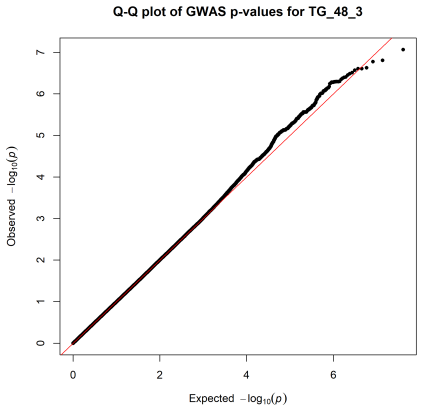


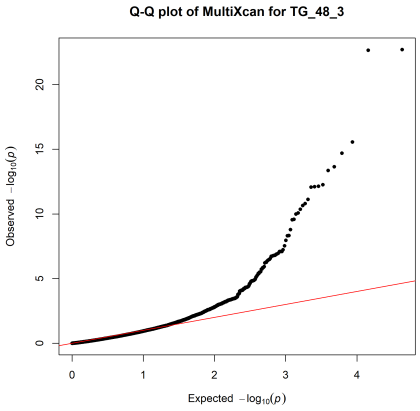
Chr6__24042708__34003583

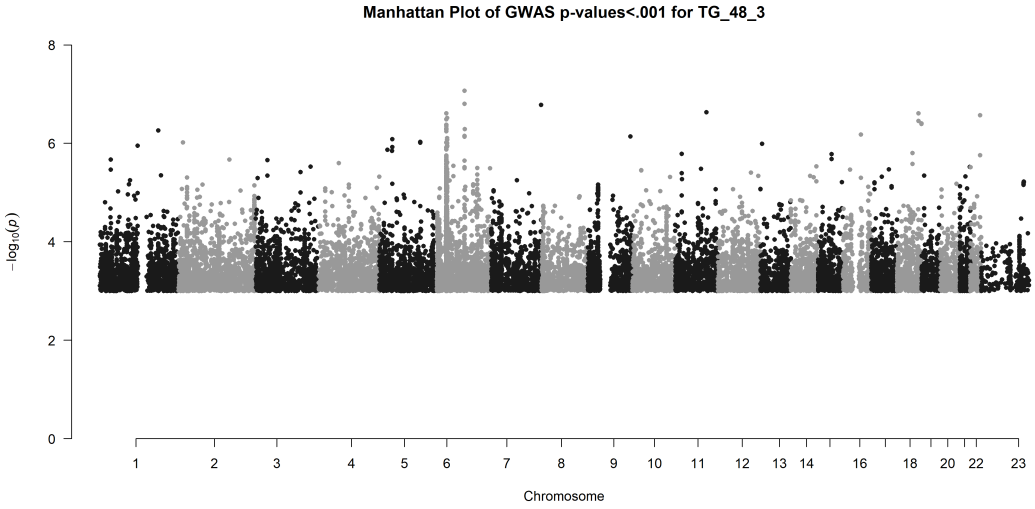


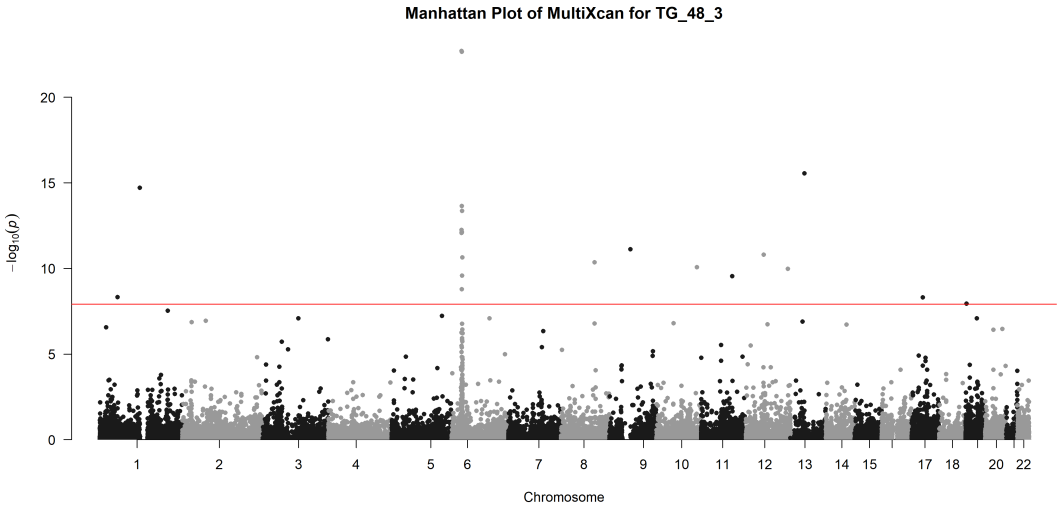
Triglycerol C48:3 (µg/dL)



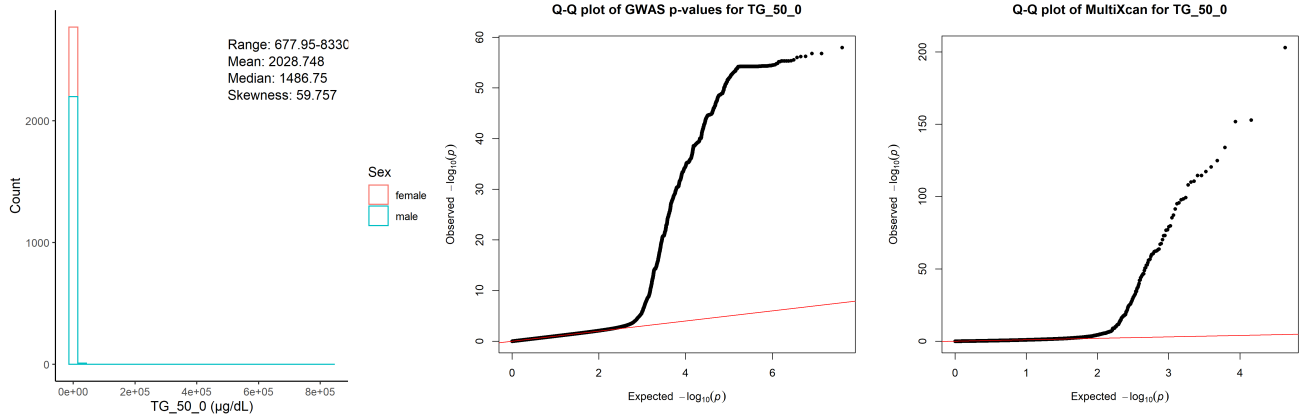


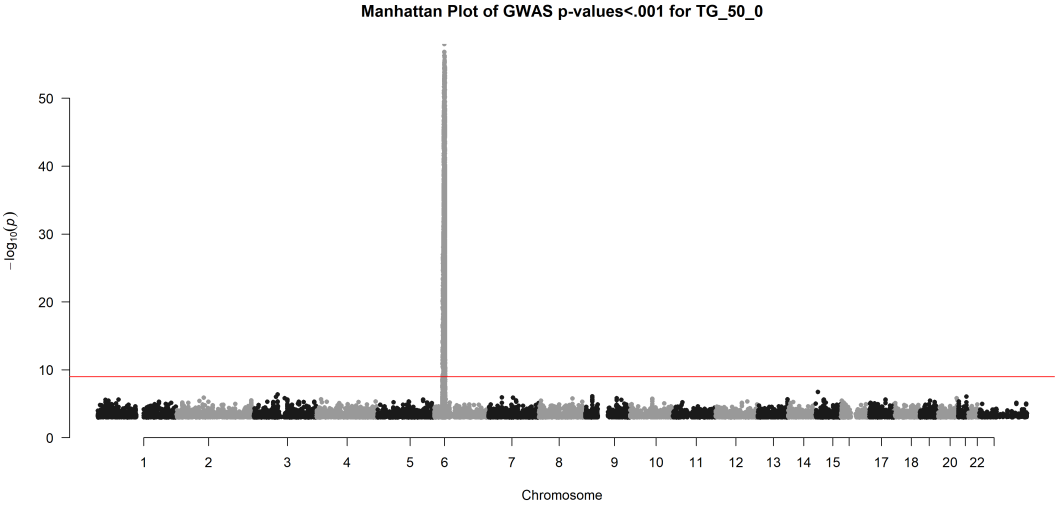


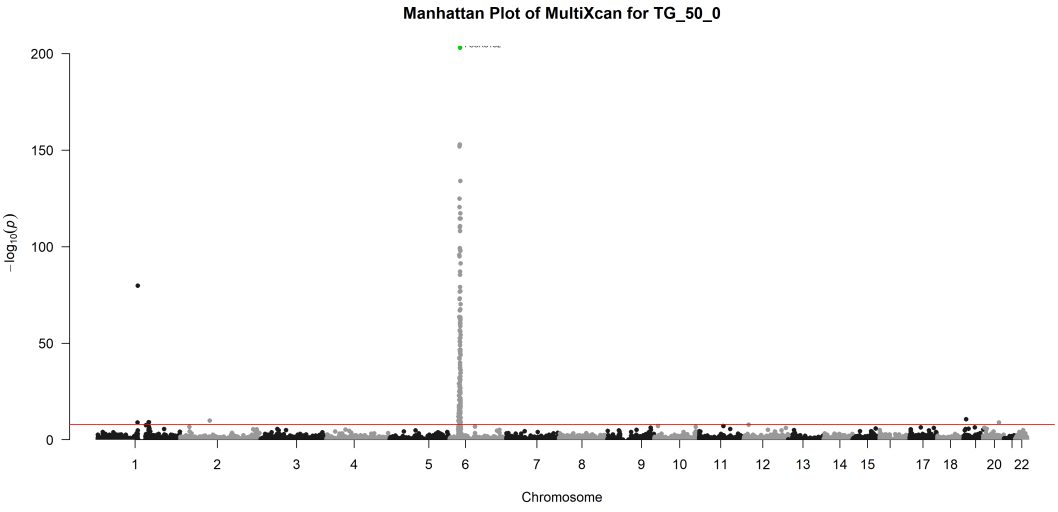




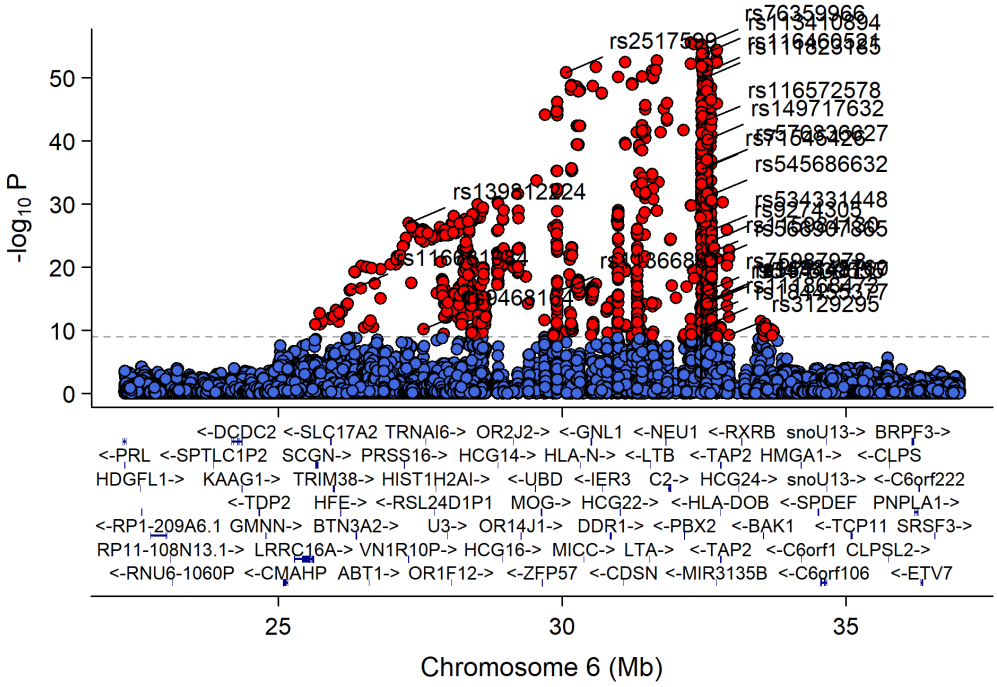
Triglycerol C50:0 (µg/dL)



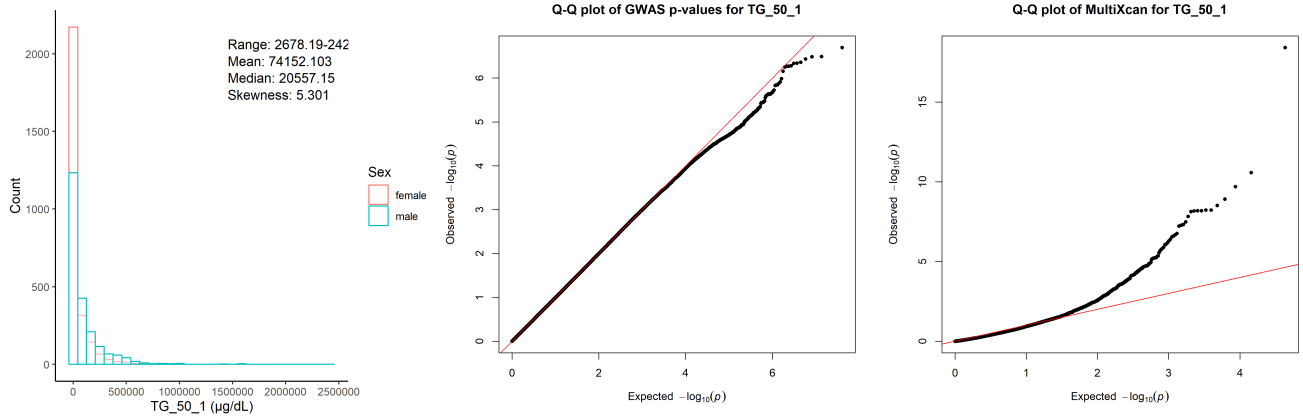


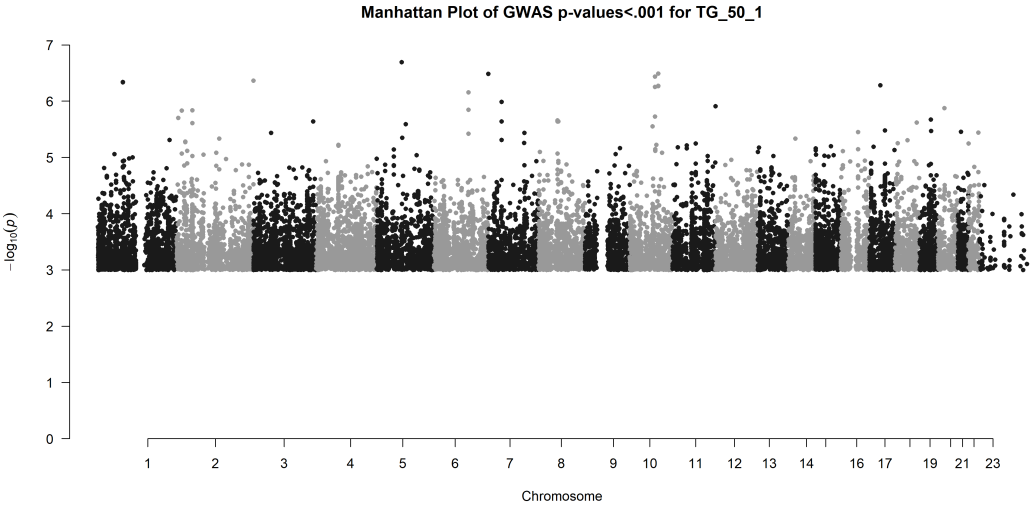


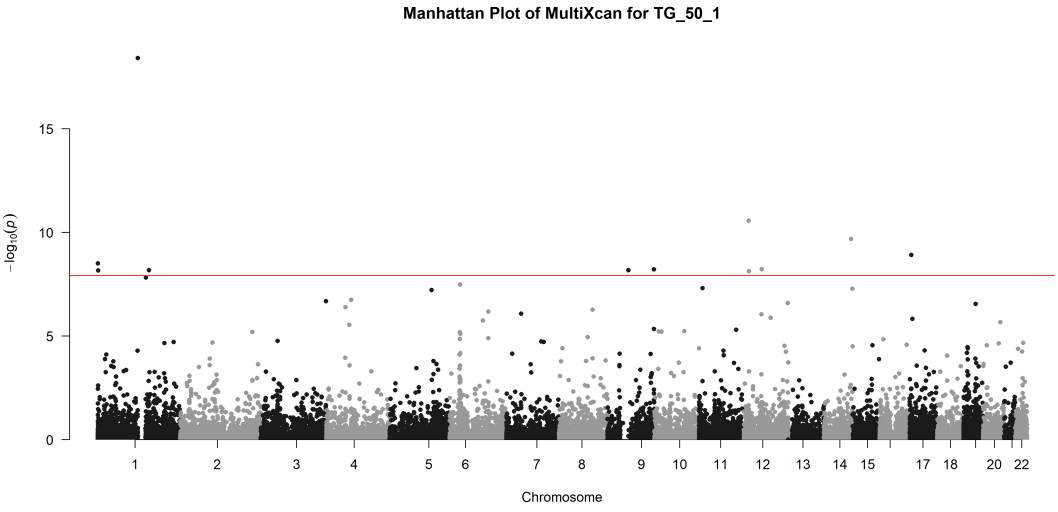
Chr6__24042708__34003583



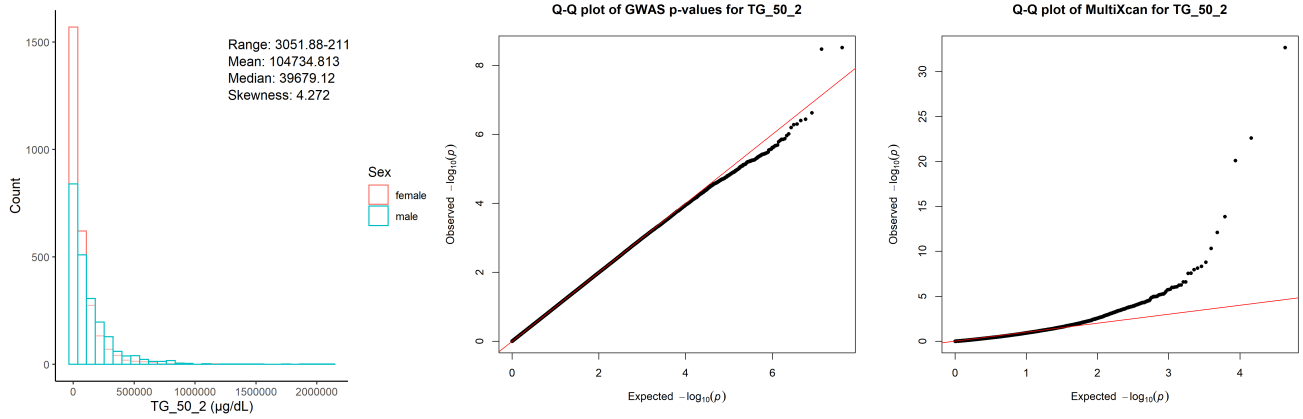
Triglycerol C50:1 (µg/dL)

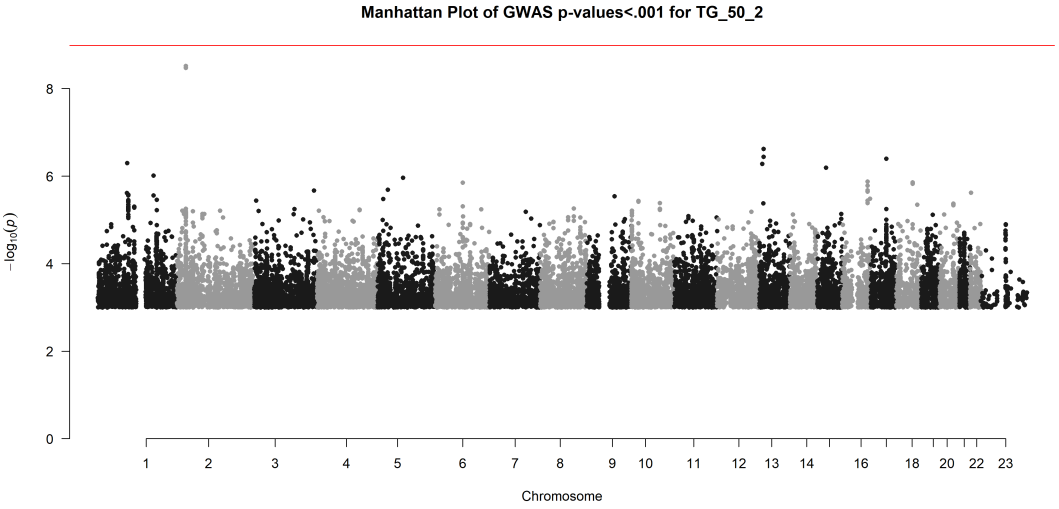


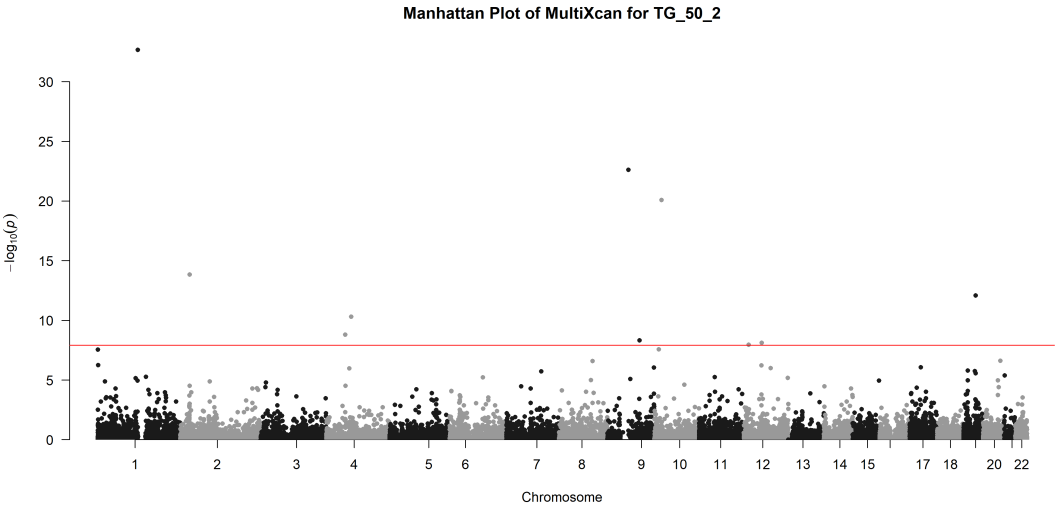




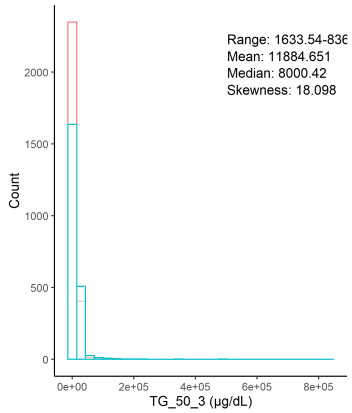
Triglycerol C50:2 (µg/dL)

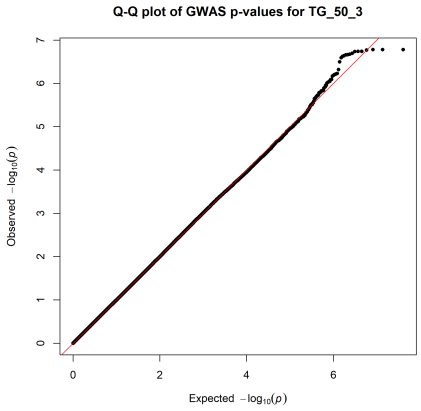


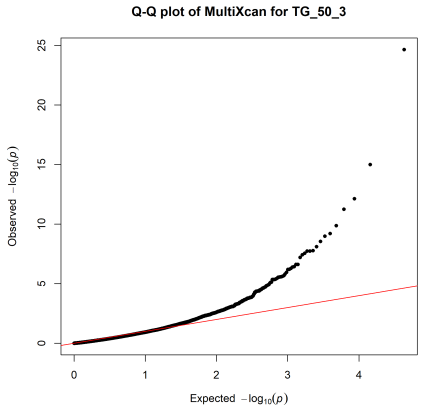


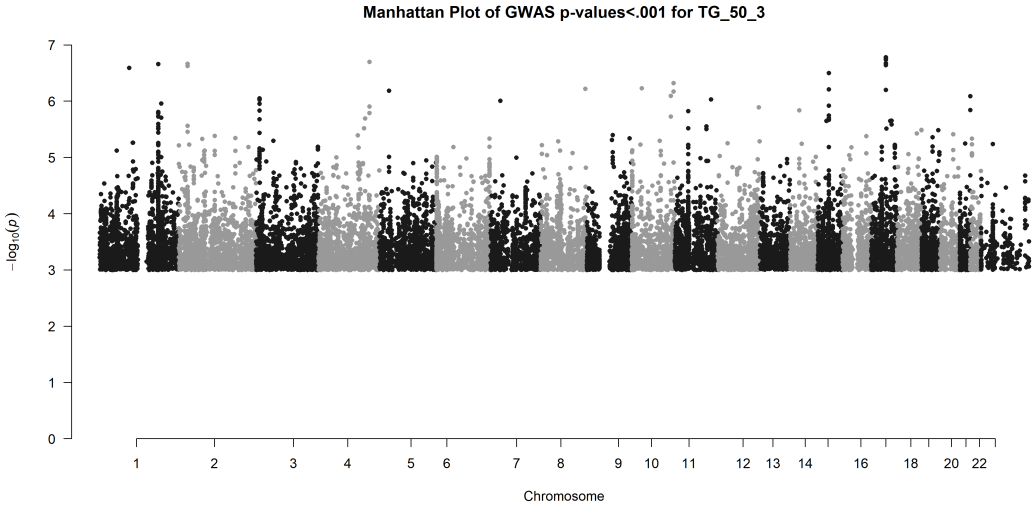


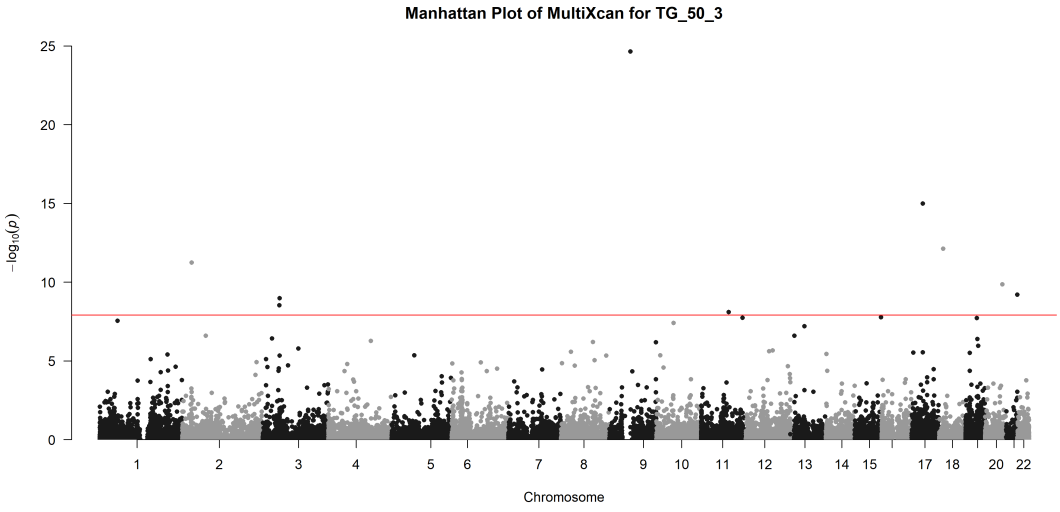
Triglycerol C50:3 (µg/dL)



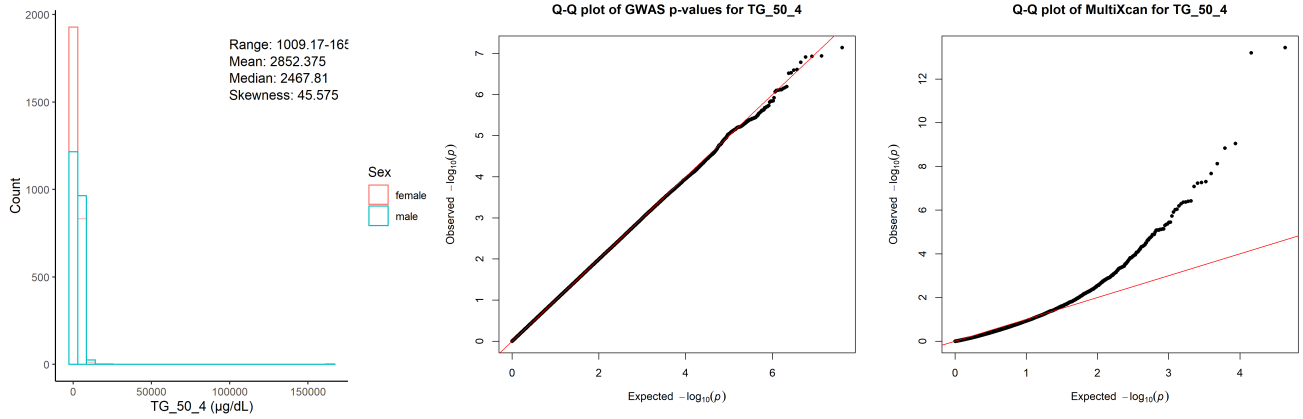


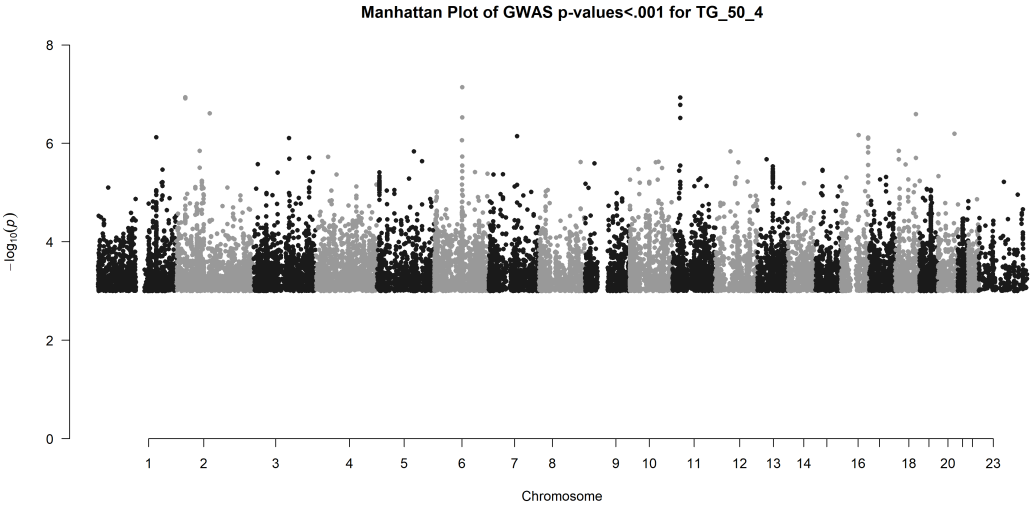


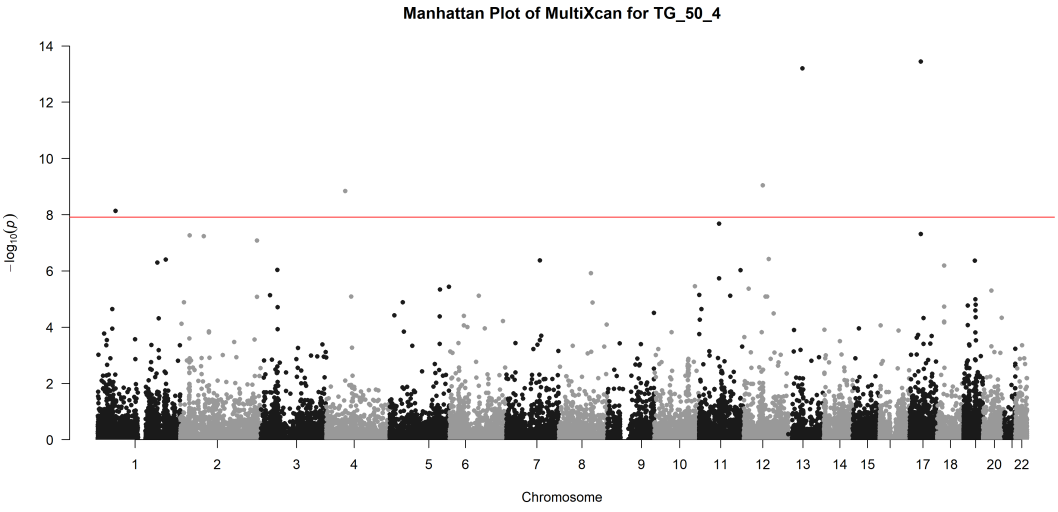




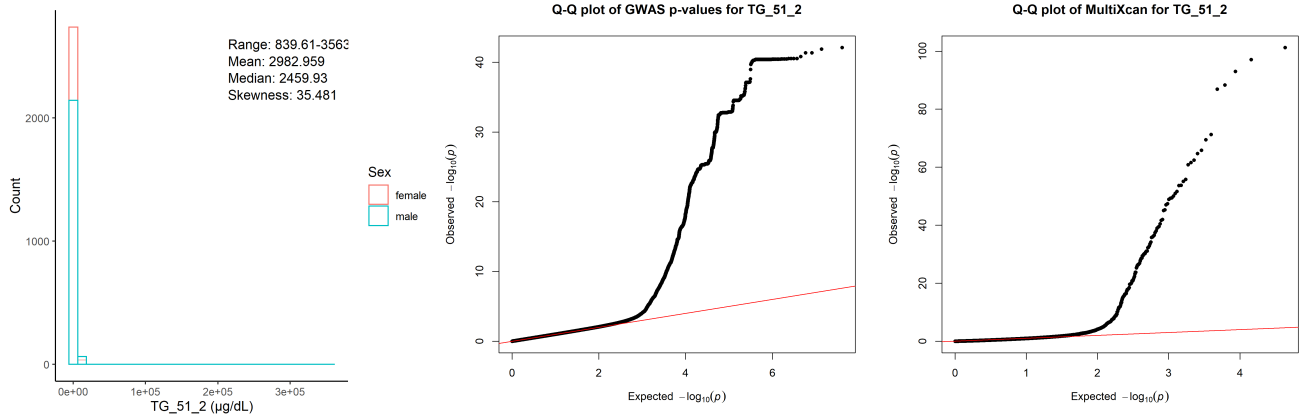
Triglycerol C50:4 (µg/dL)

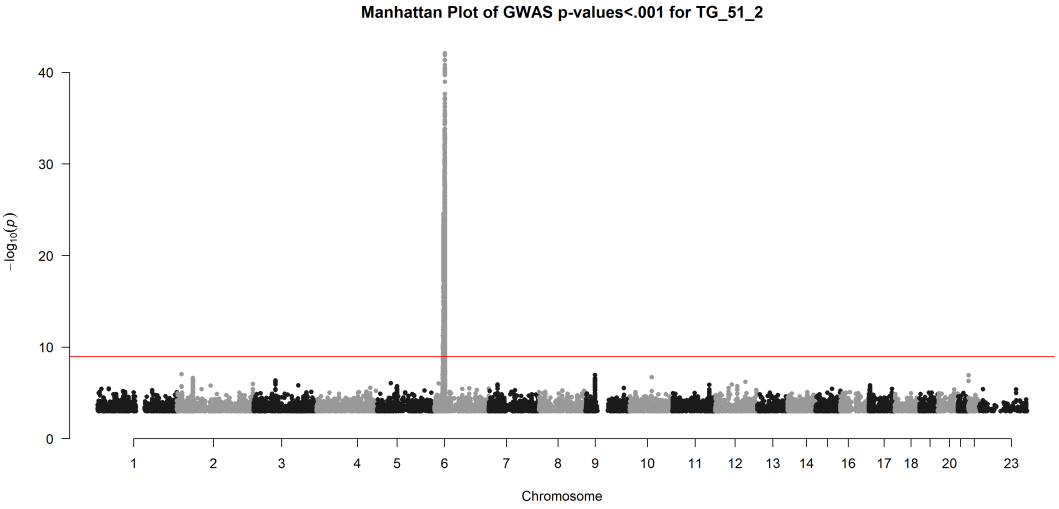


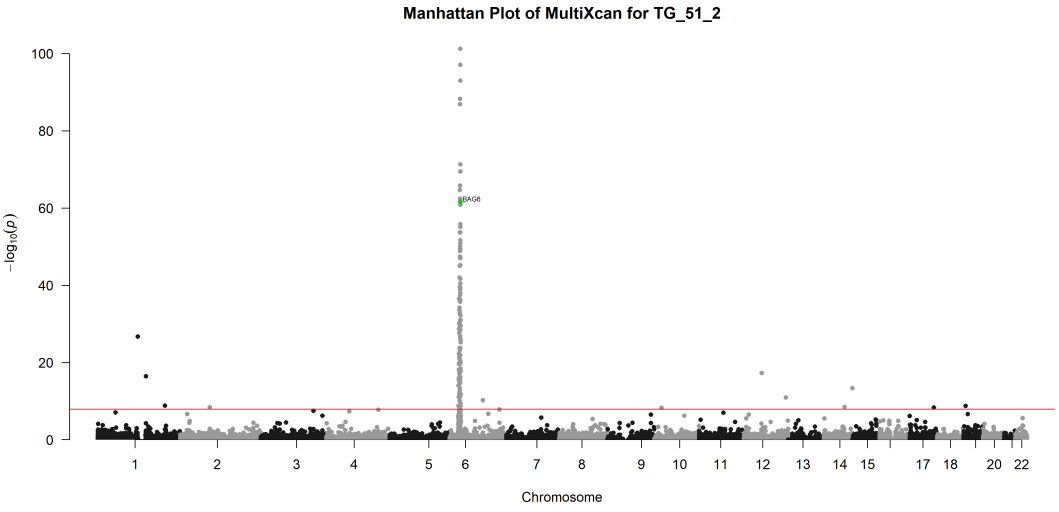




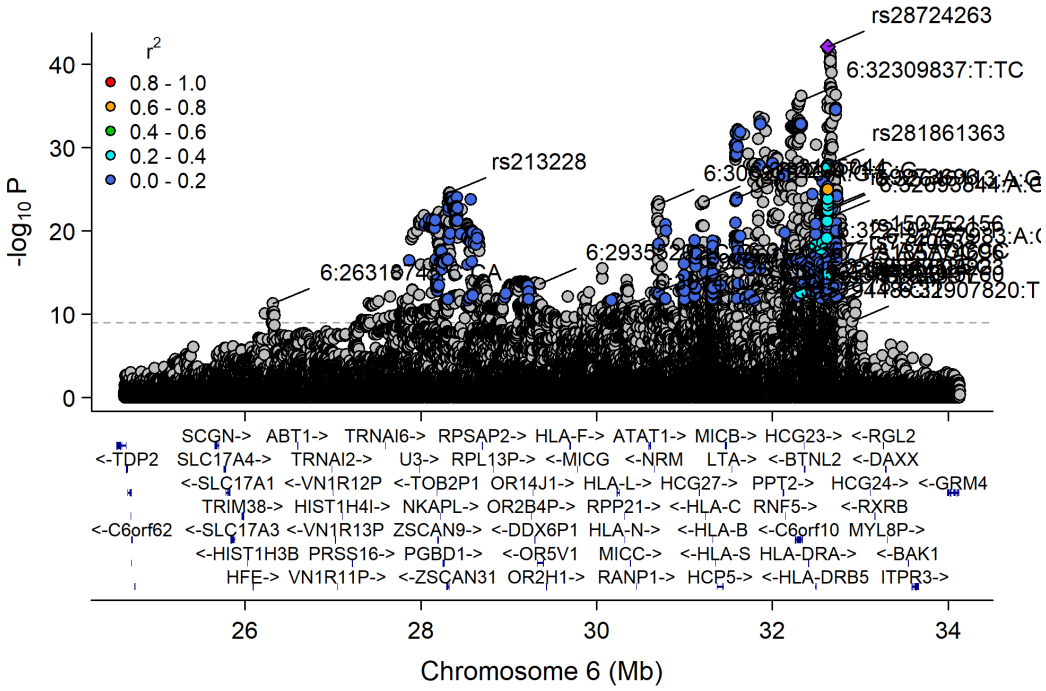
Triglycerol C51:2 (µg/dL)



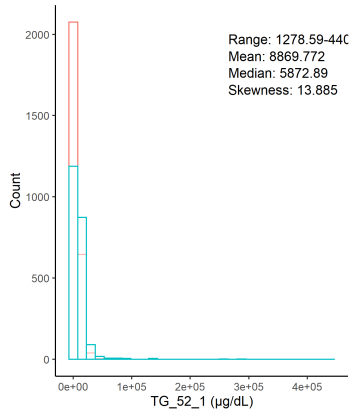


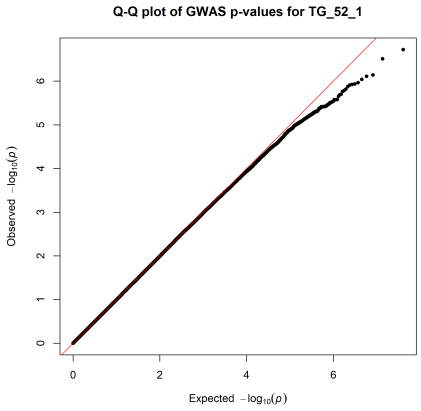


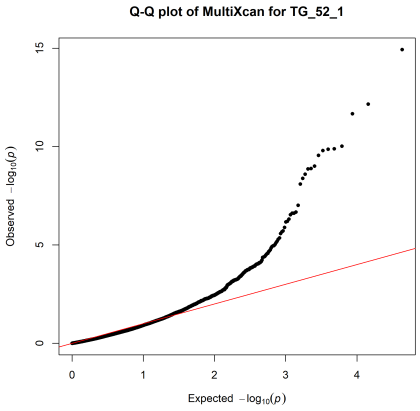
Chr6__24042708__34003583

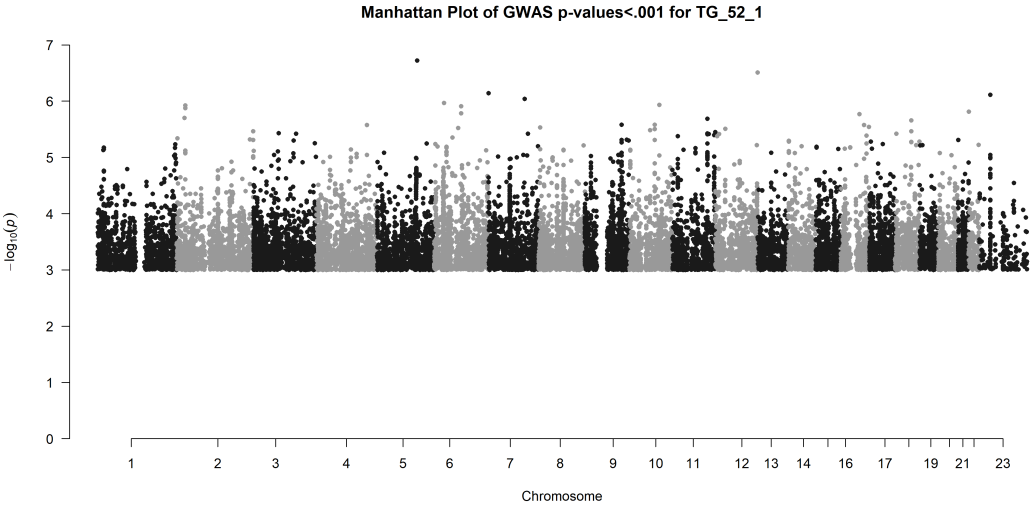


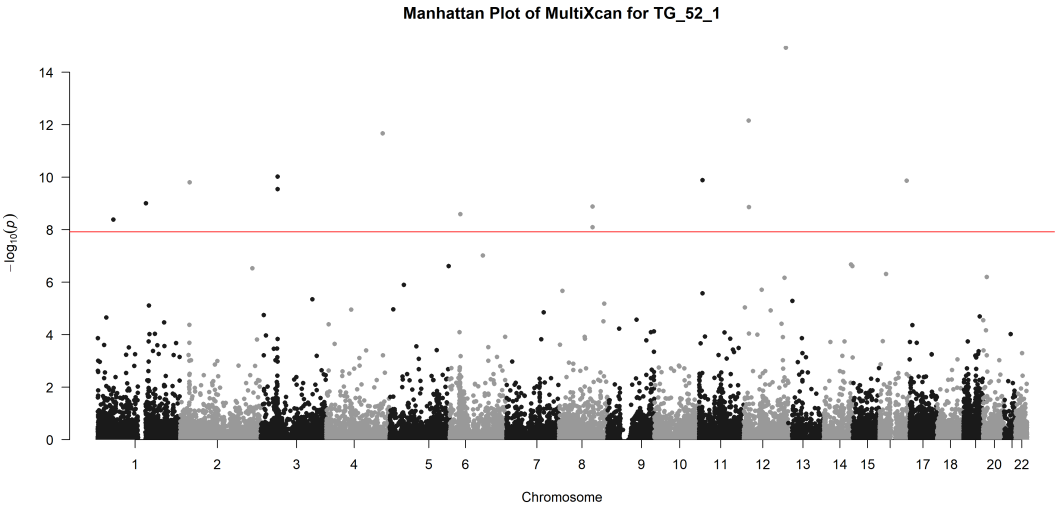
Triglycerol C52:1 (µg/dL)



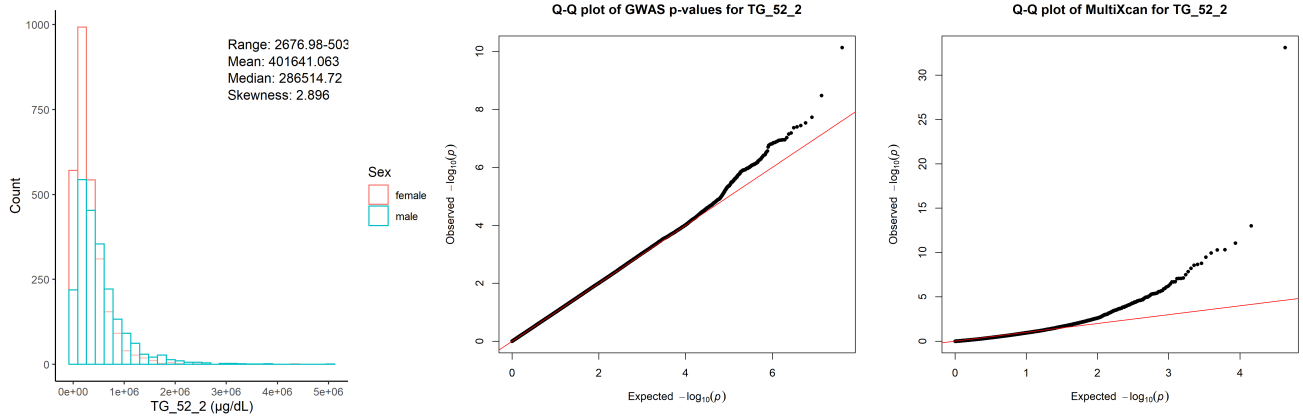


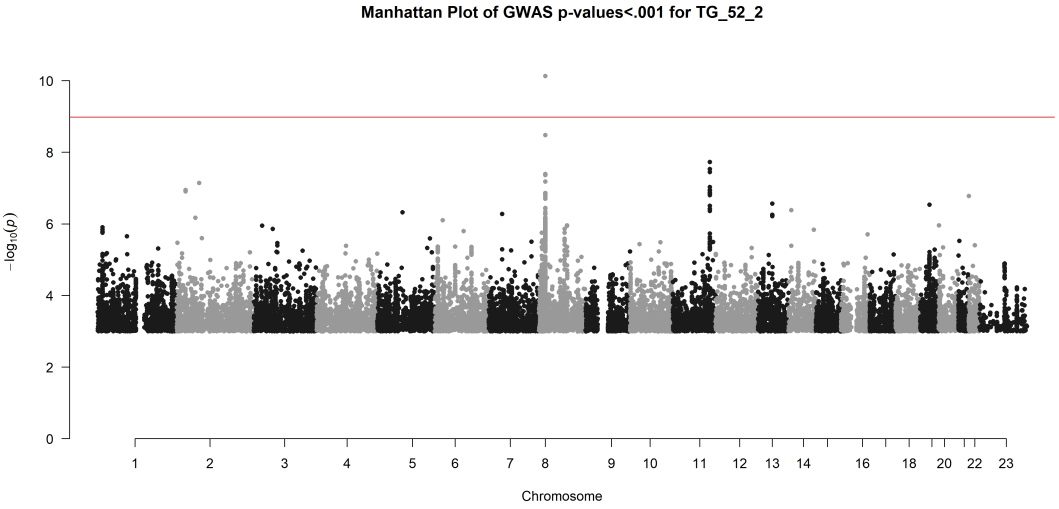


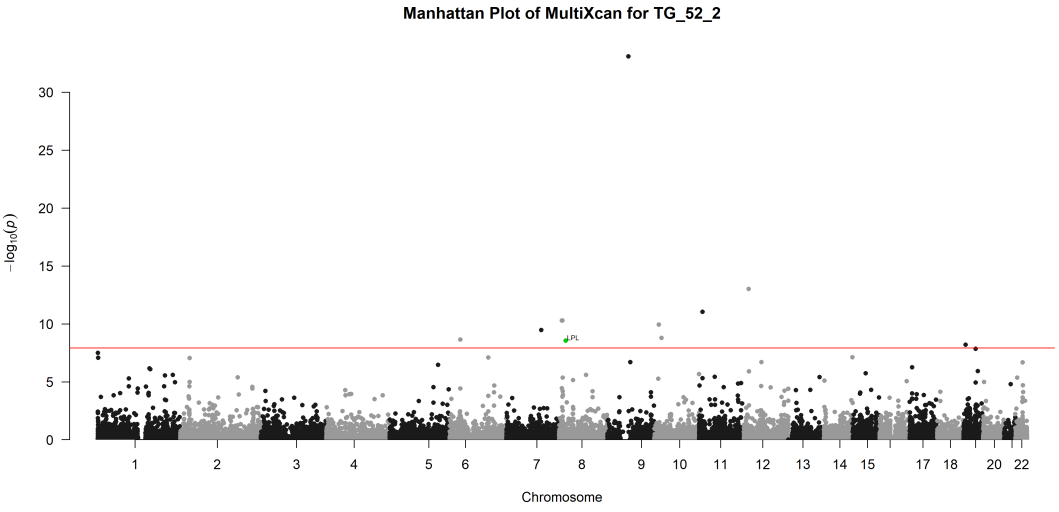




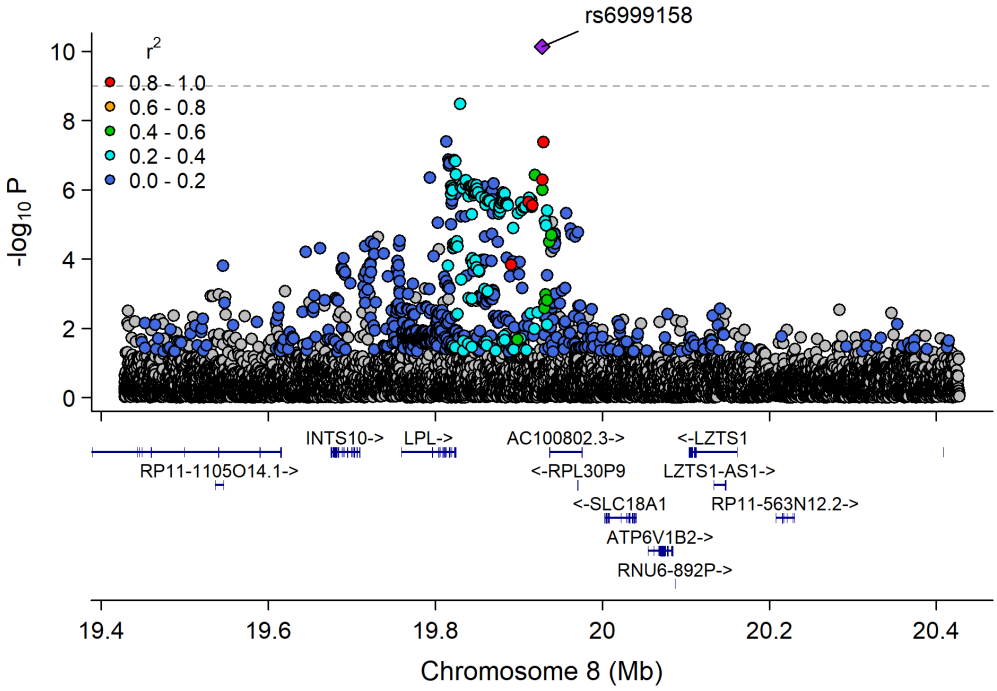
Triglycerol C52:2 (µg/dL)



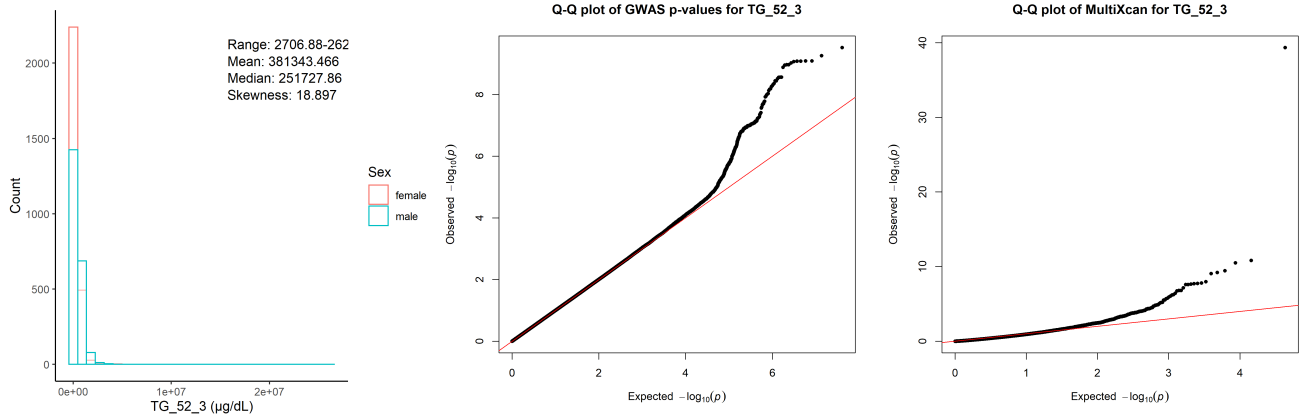


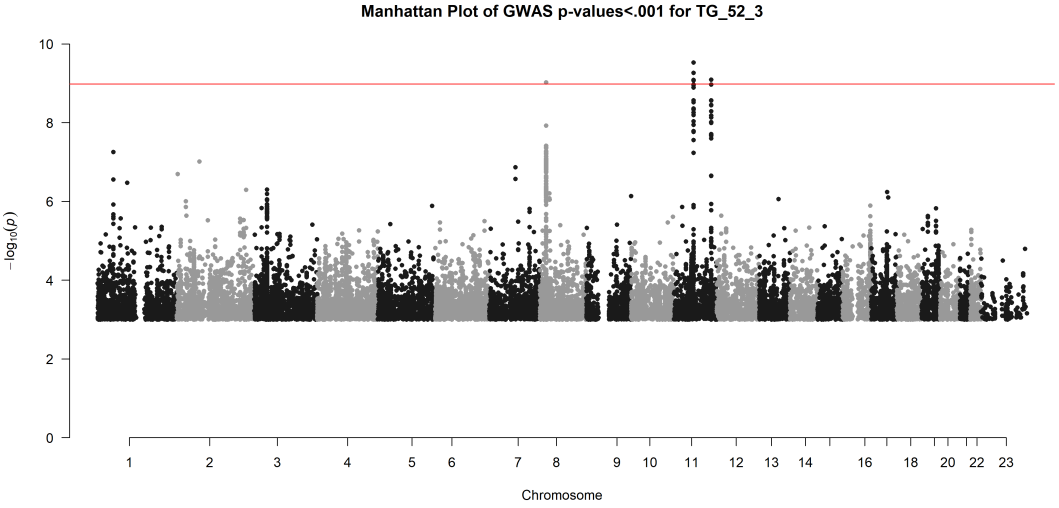


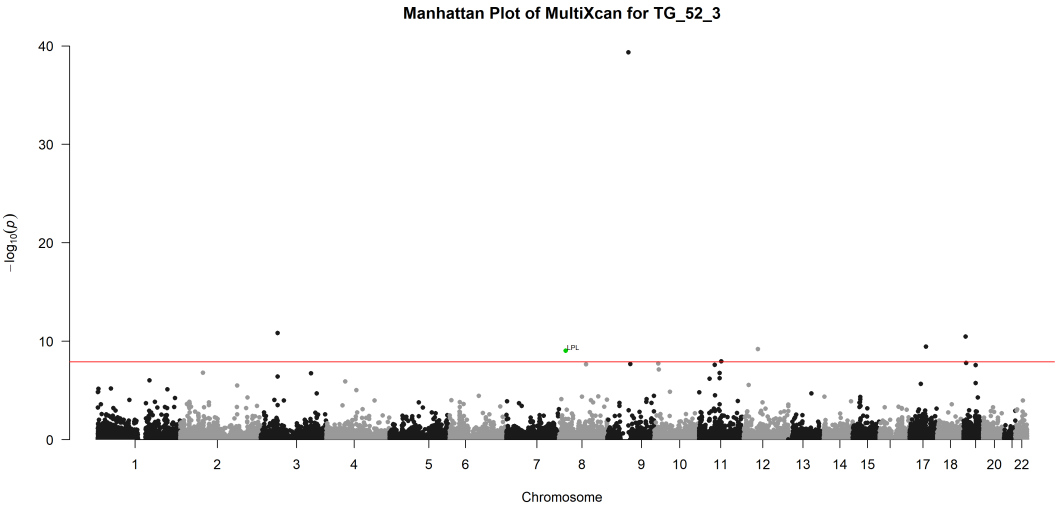
Chr8_18582620_20792744



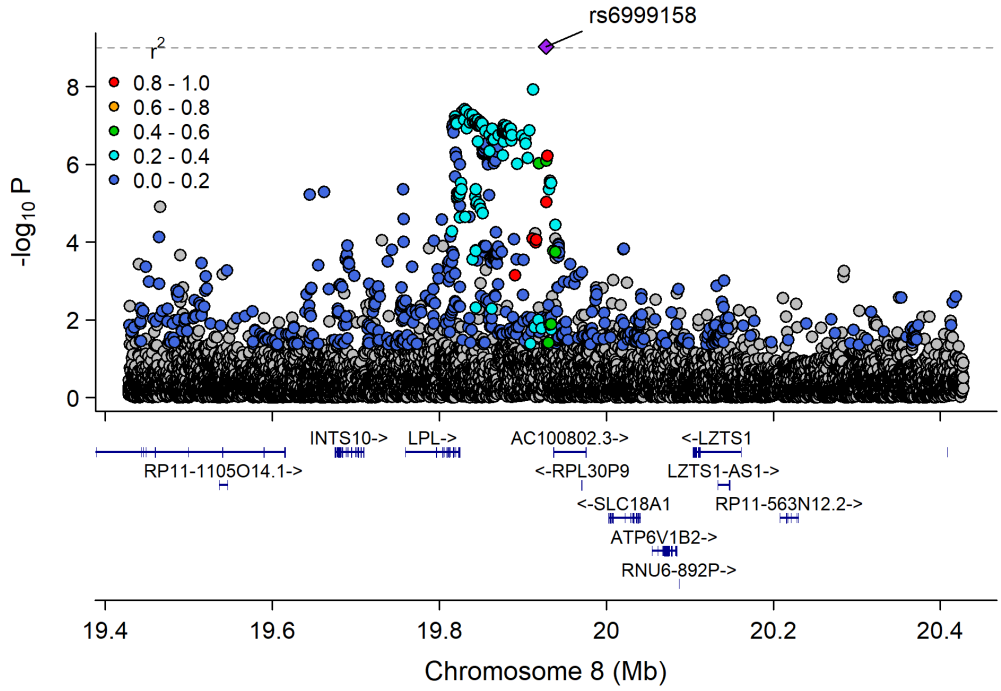
Triglycerol C52:3 (µg/dL)



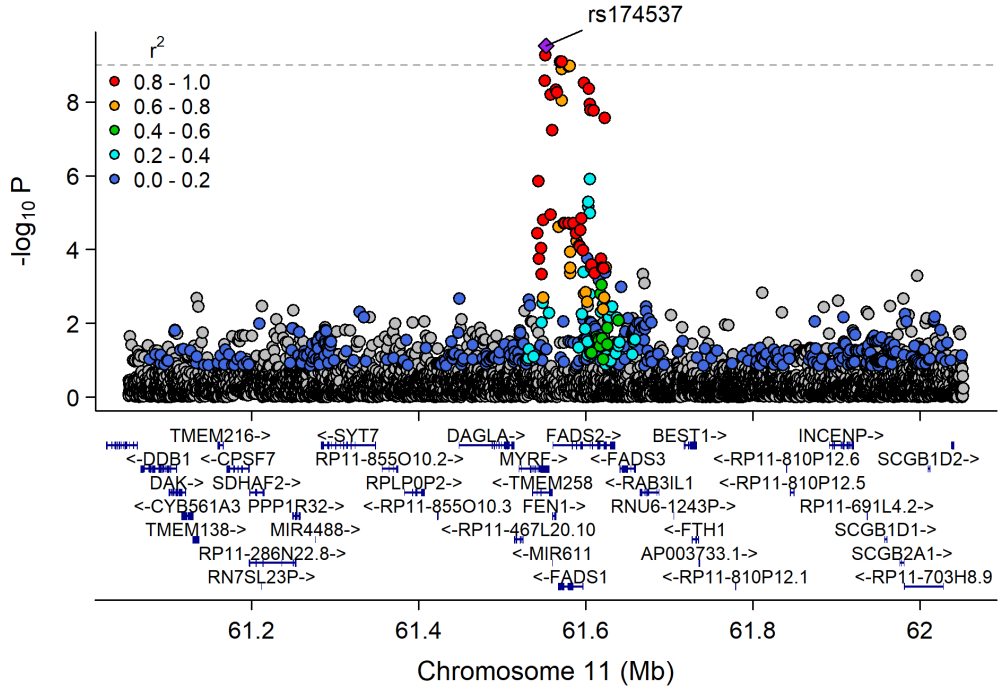




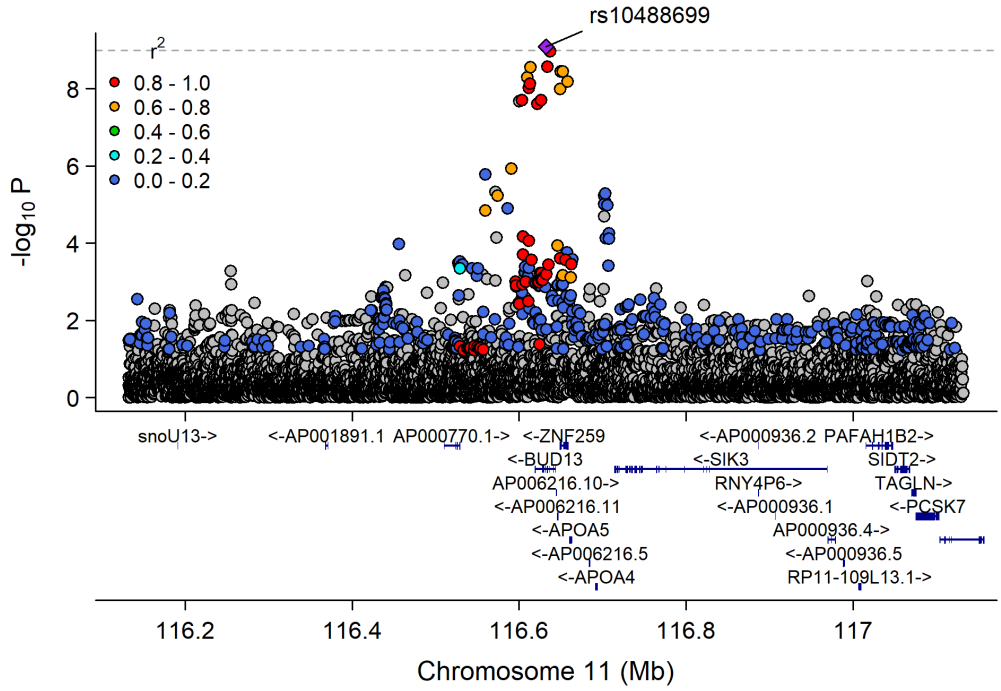
Chr8_18582620_20792744



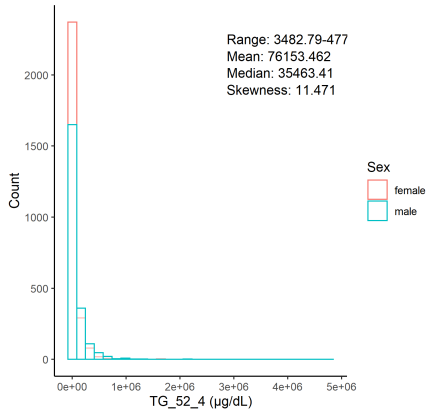
Chr11_59978355_62914375

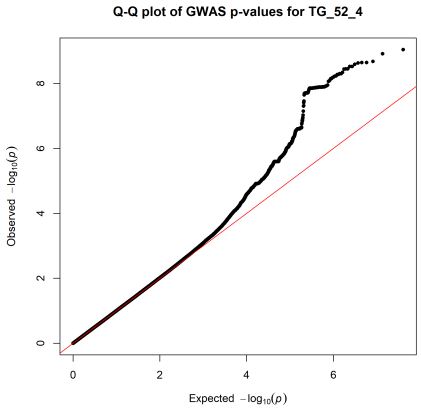


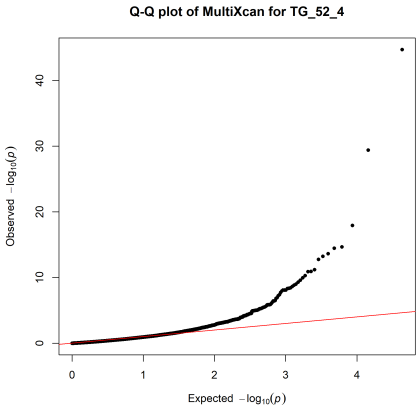
Chr11_115541901_117633315

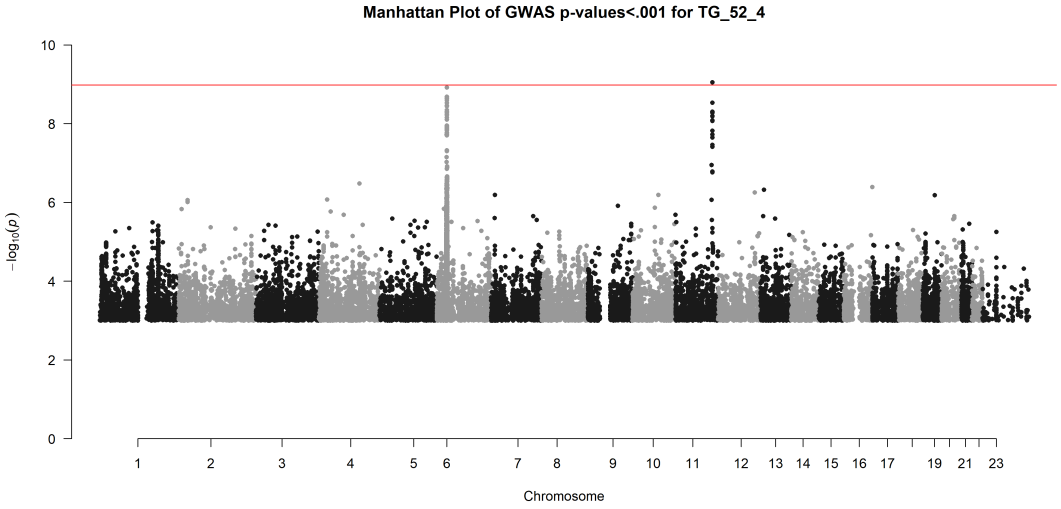


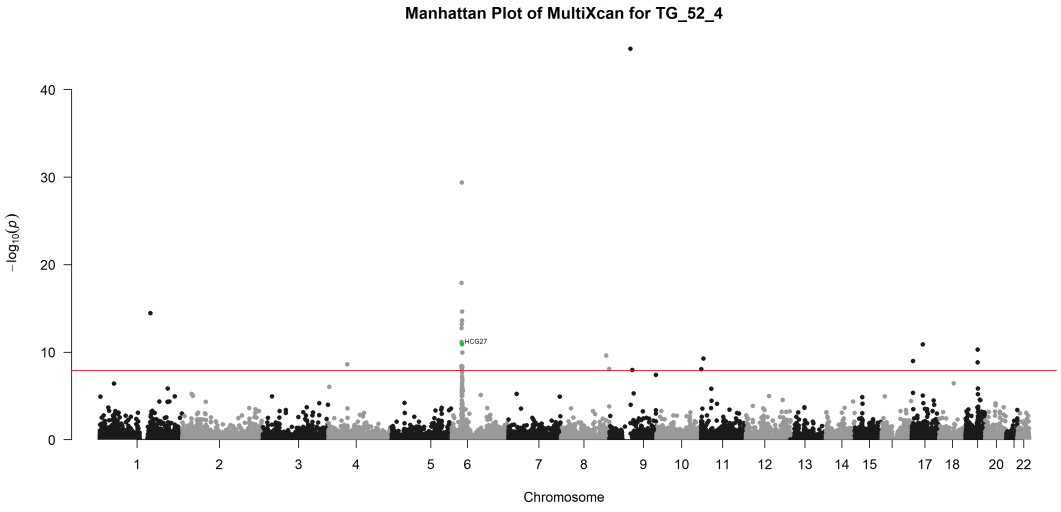
Triglycerol C52:4 (µg/dL)



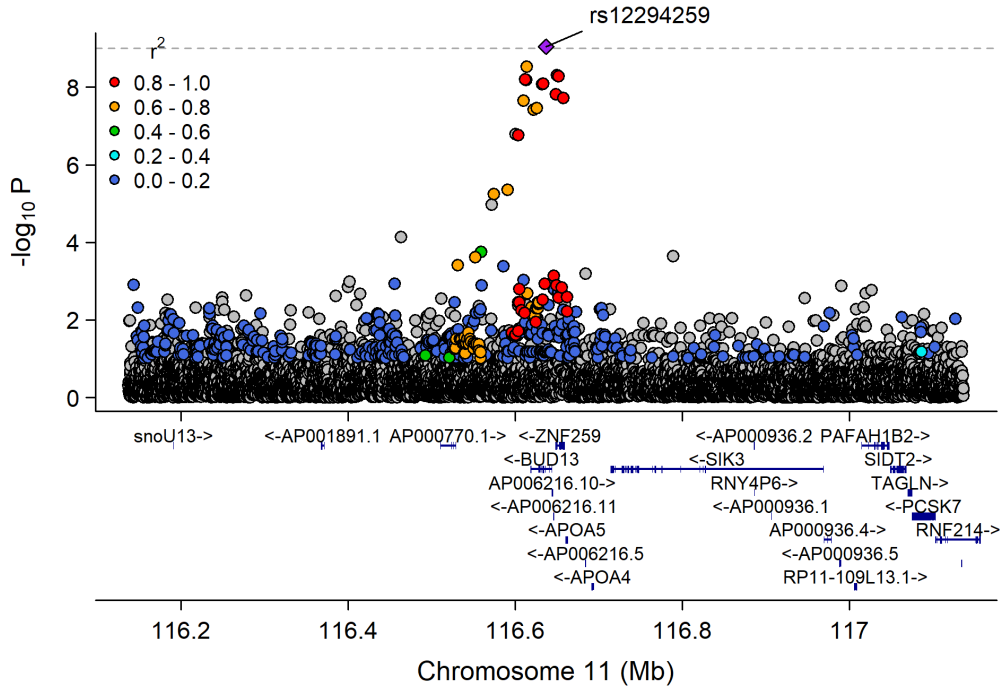




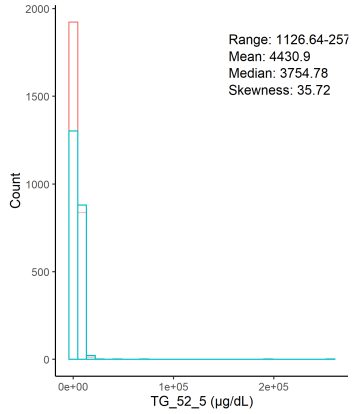


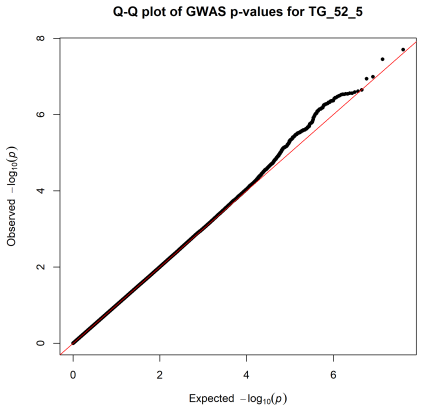


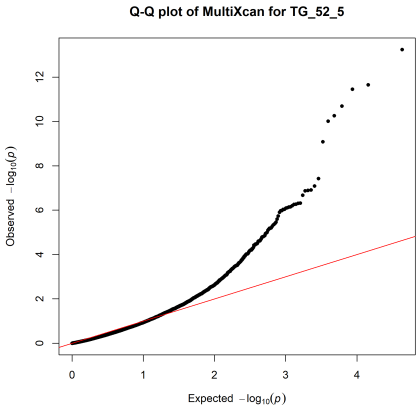
Chr11_115541901_117633315

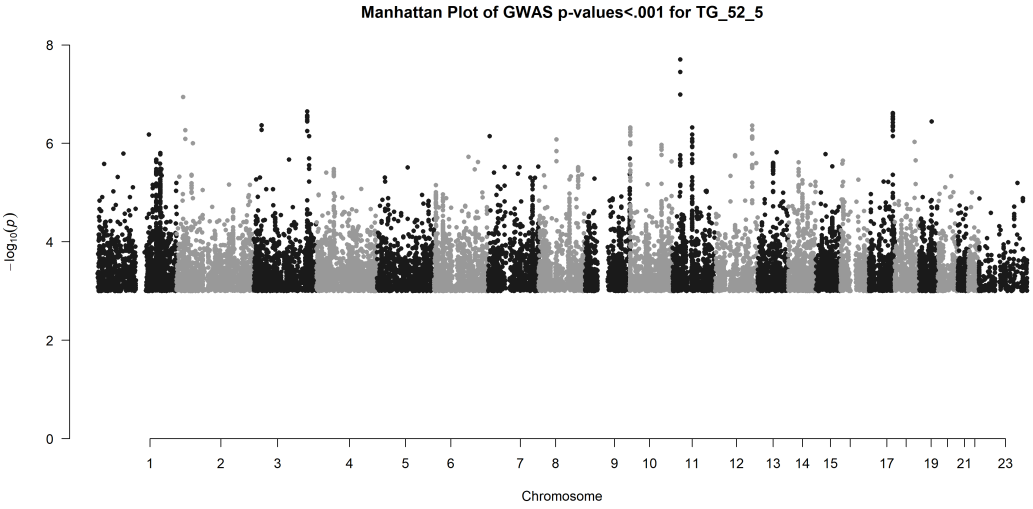


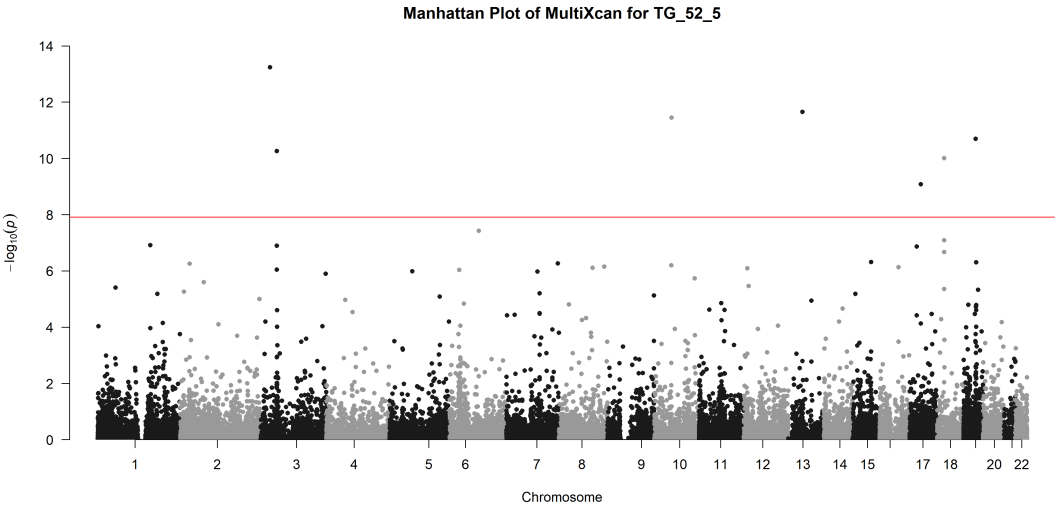
Triglycerol C52:5 (µg/dL)



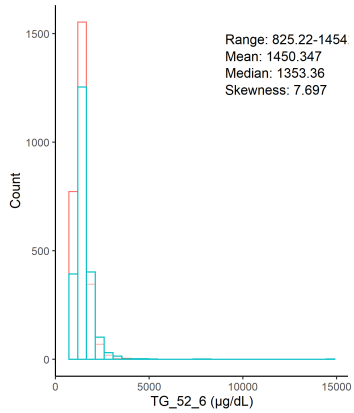


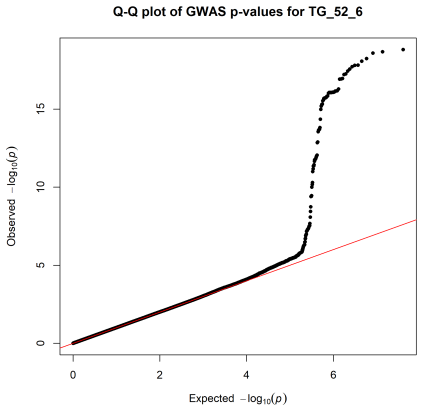


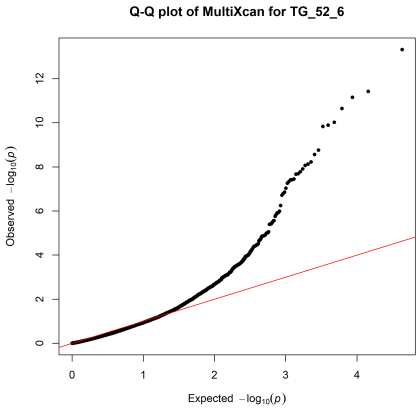


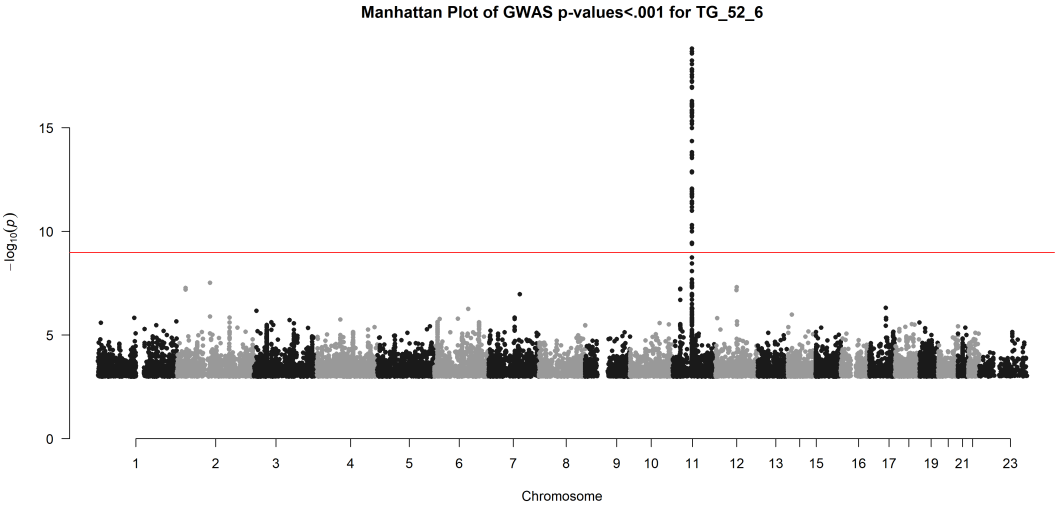


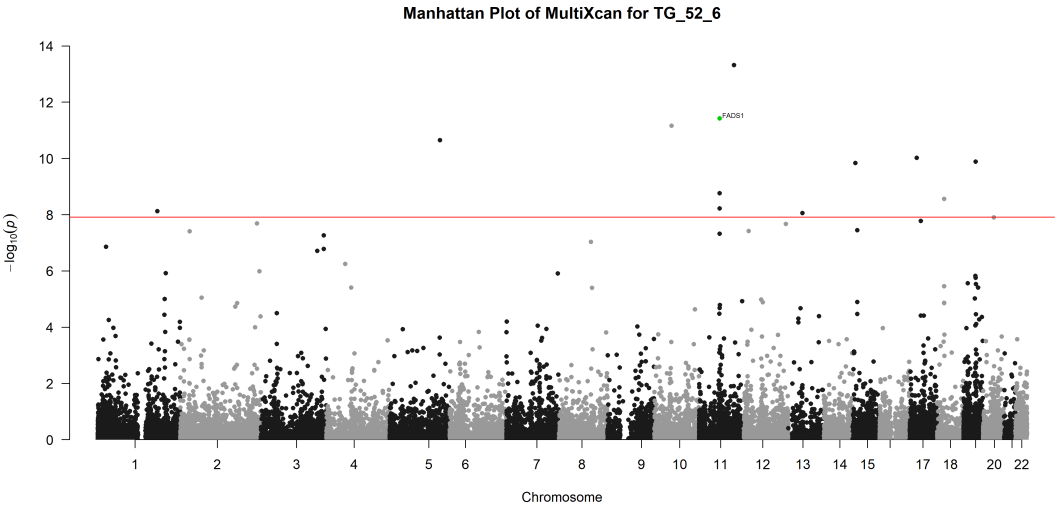
Triglycerol C52:6 (µg/dL)



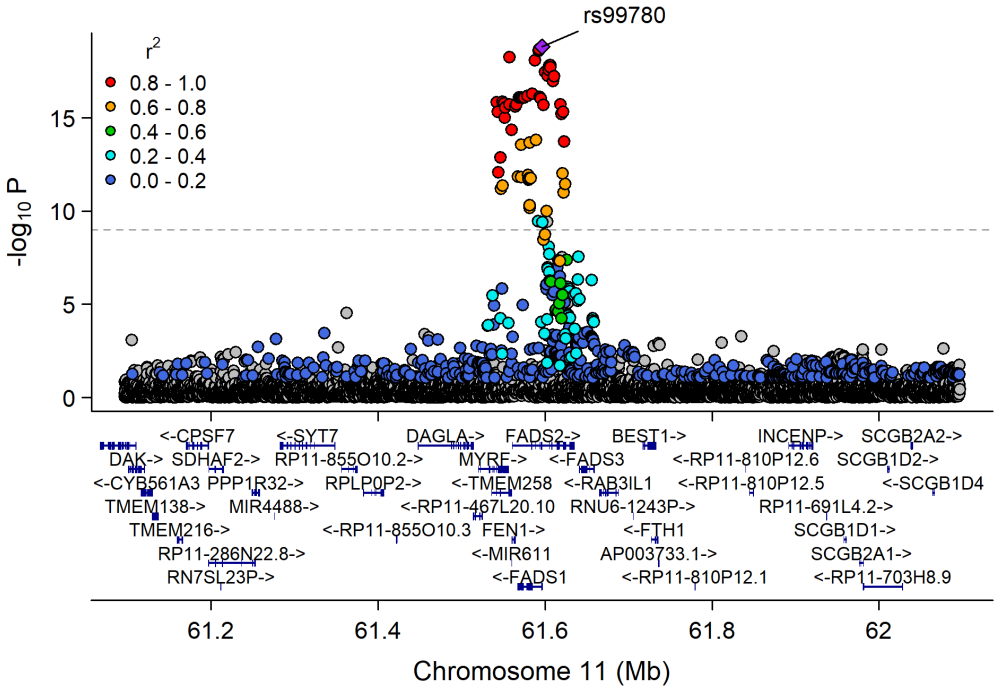




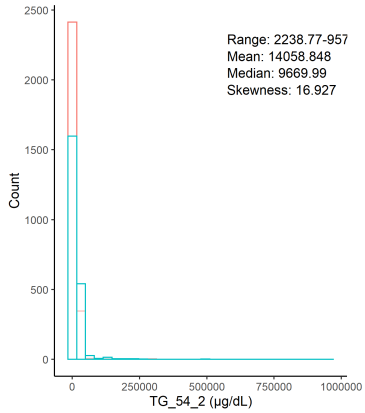


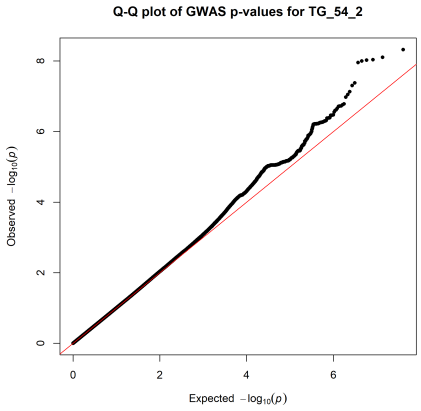


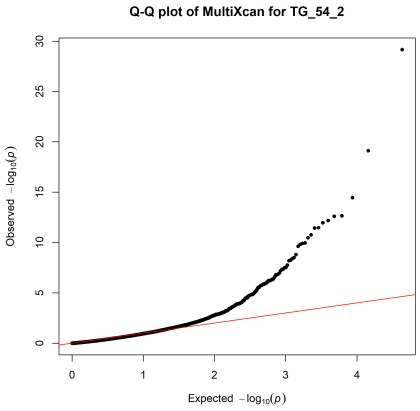
Chr11_59978355_62914375

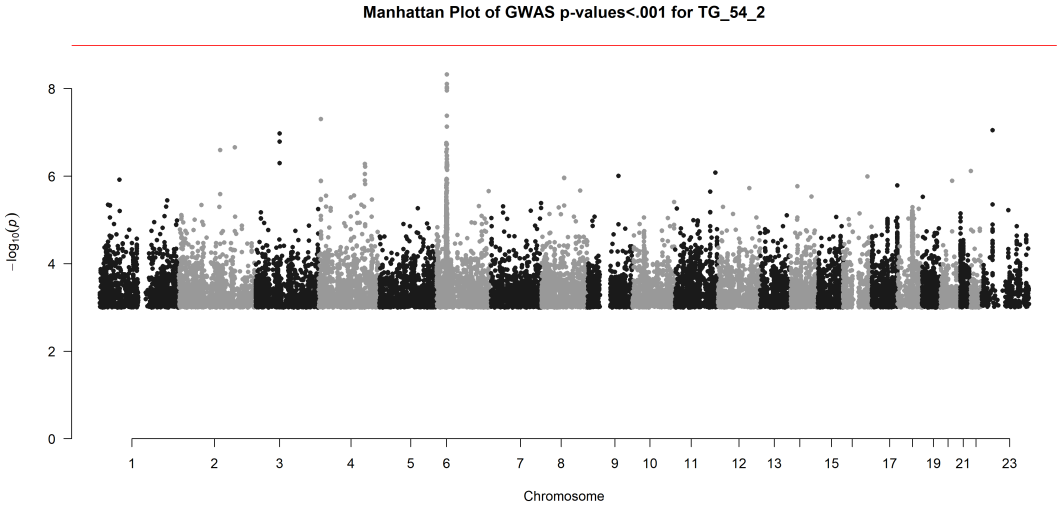


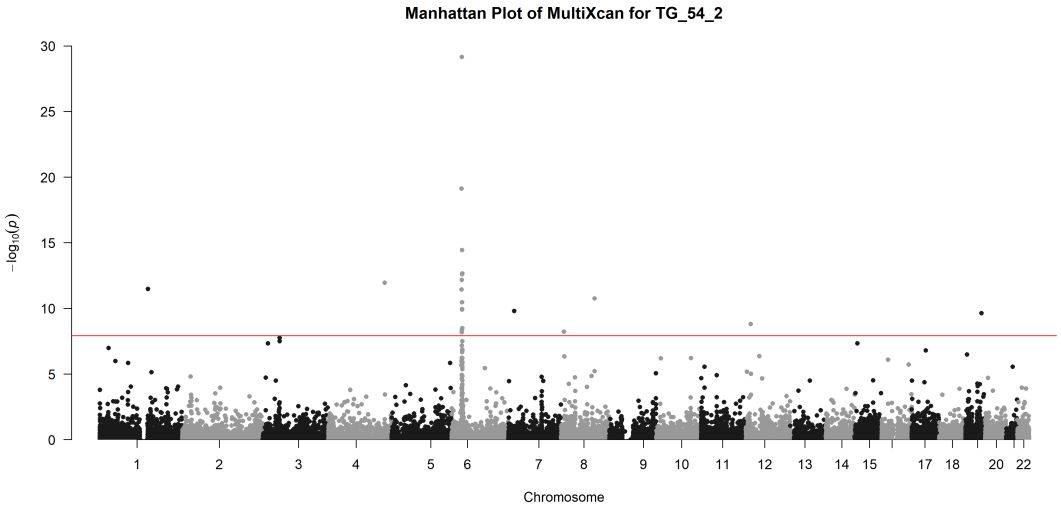
Triglycerol C54:2 (µg/dL)



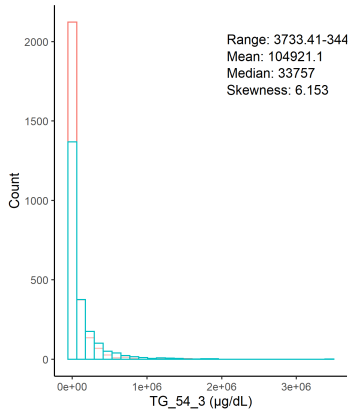


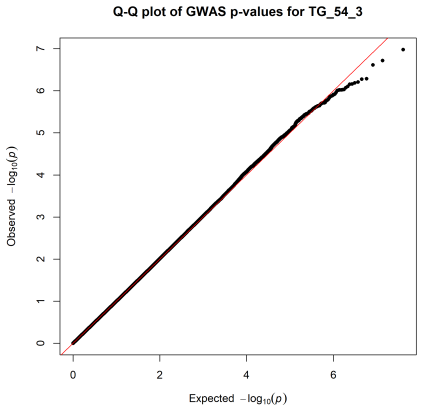


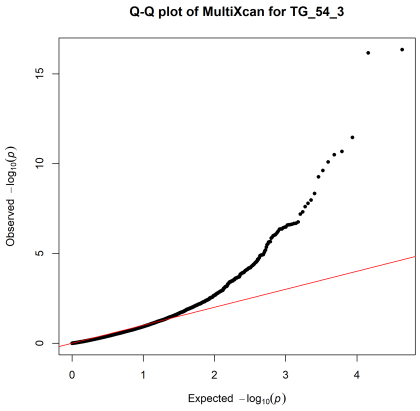


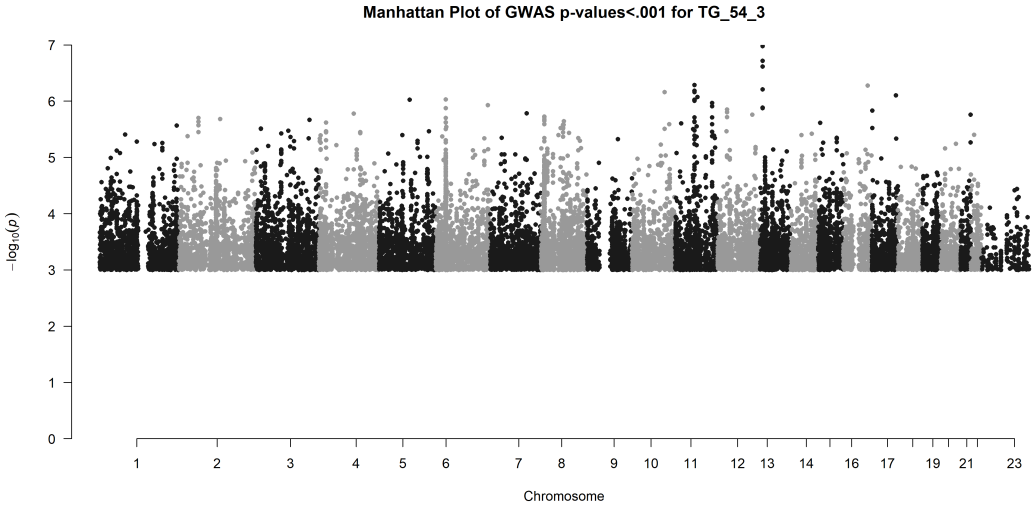


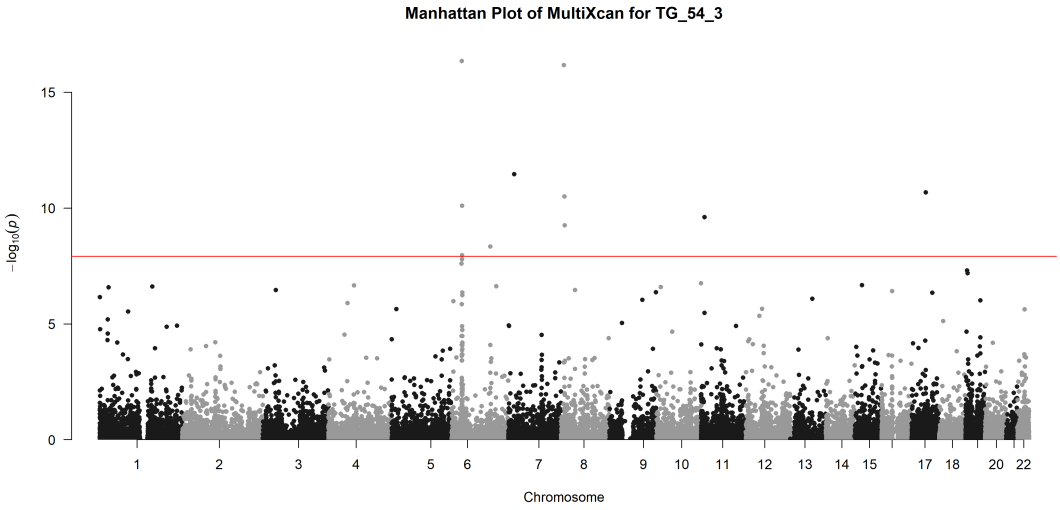
Triglycerol C54:3 (µg/dL)



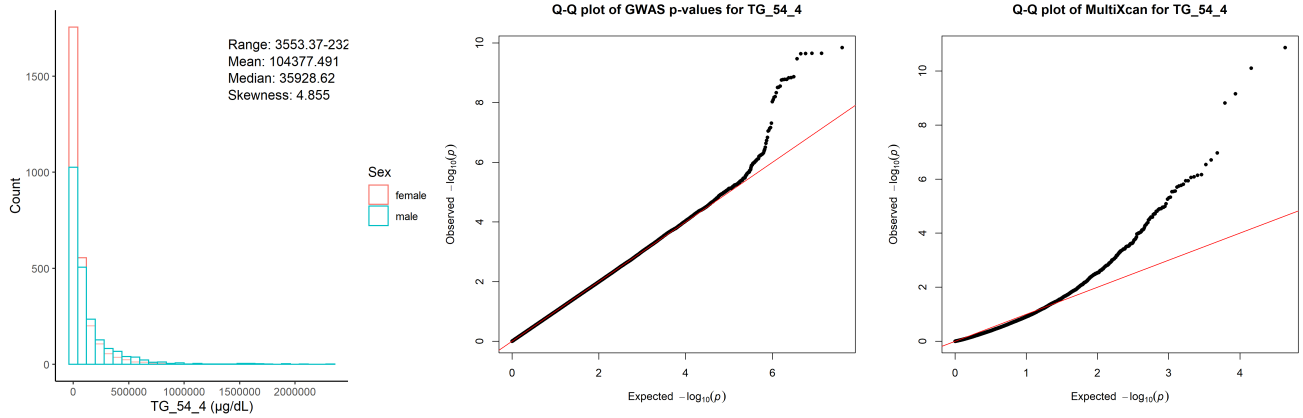


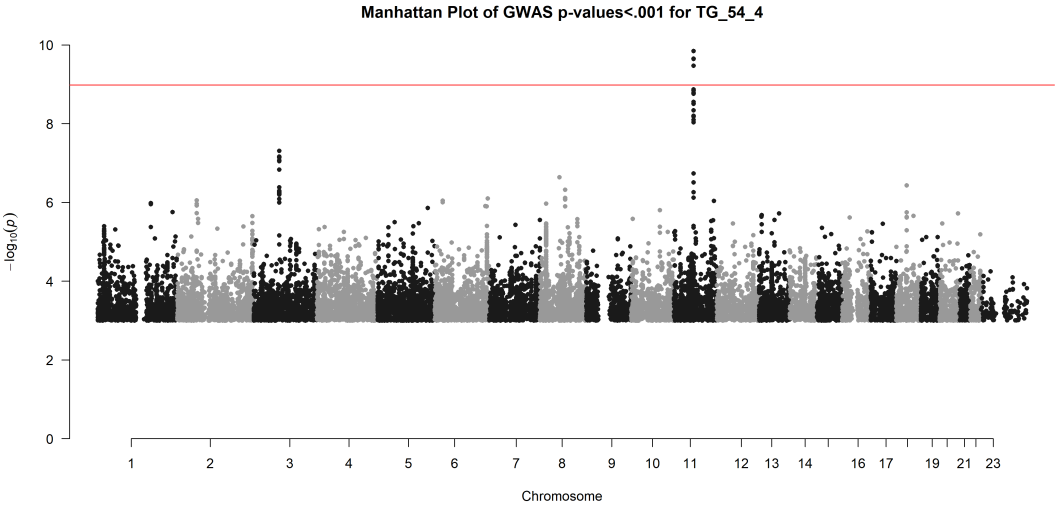


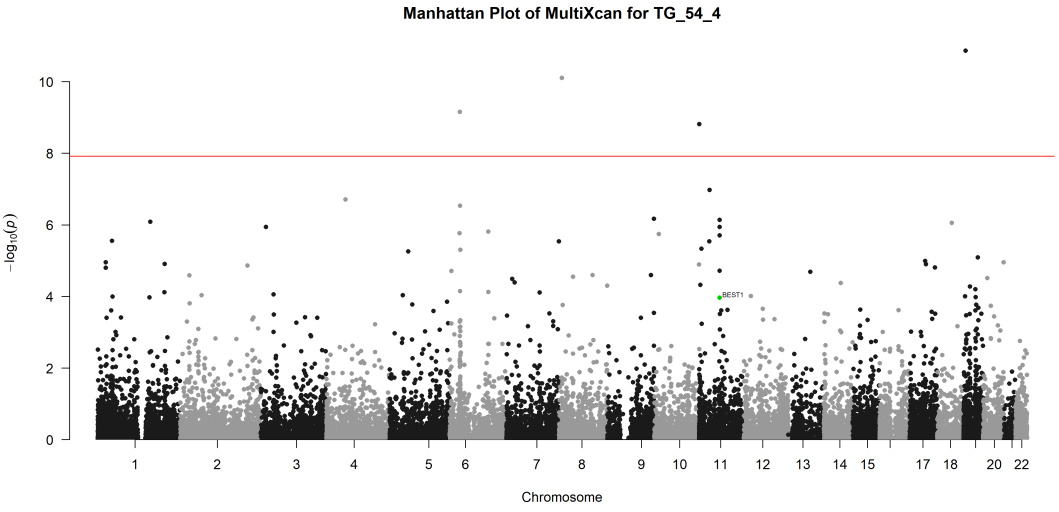




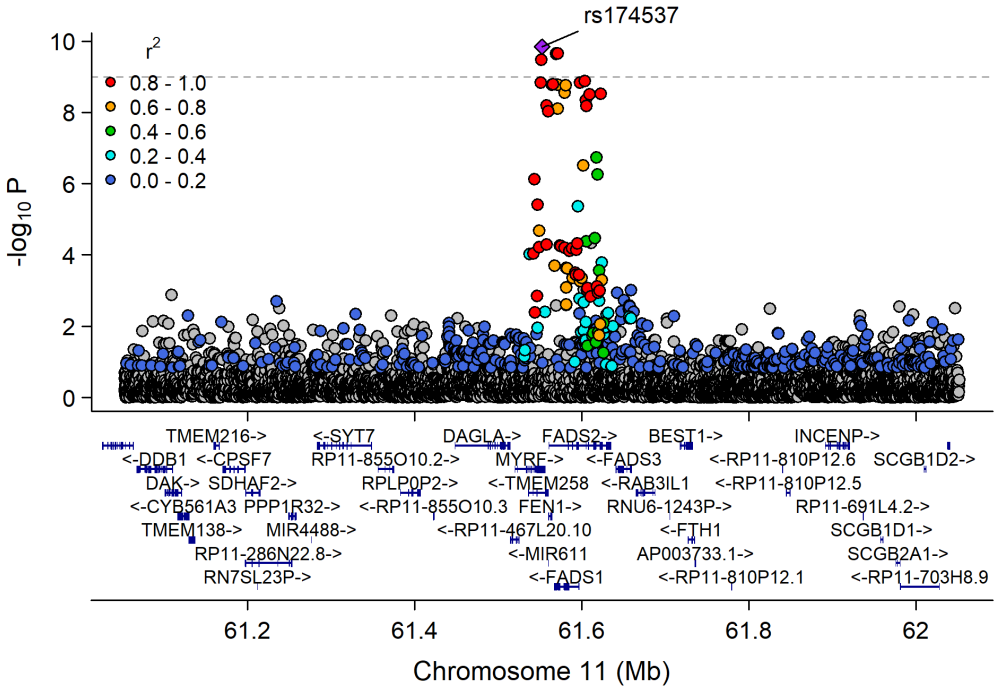
Triglycerol C54:4 (µg/dL)



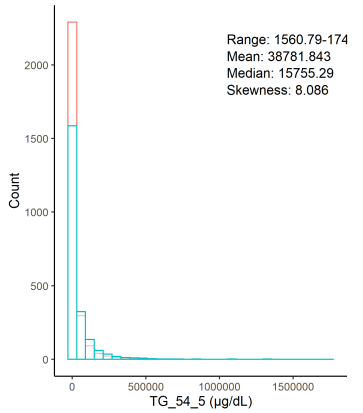


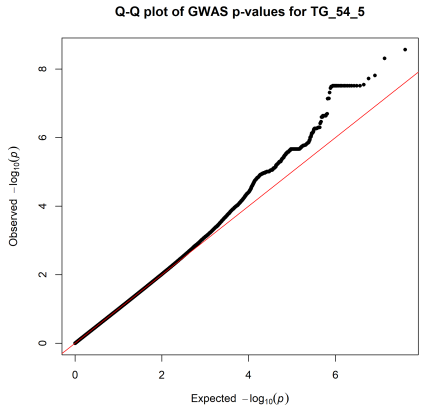


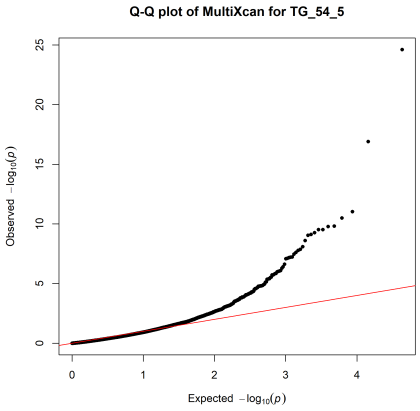
Chr11_59978355_62914375

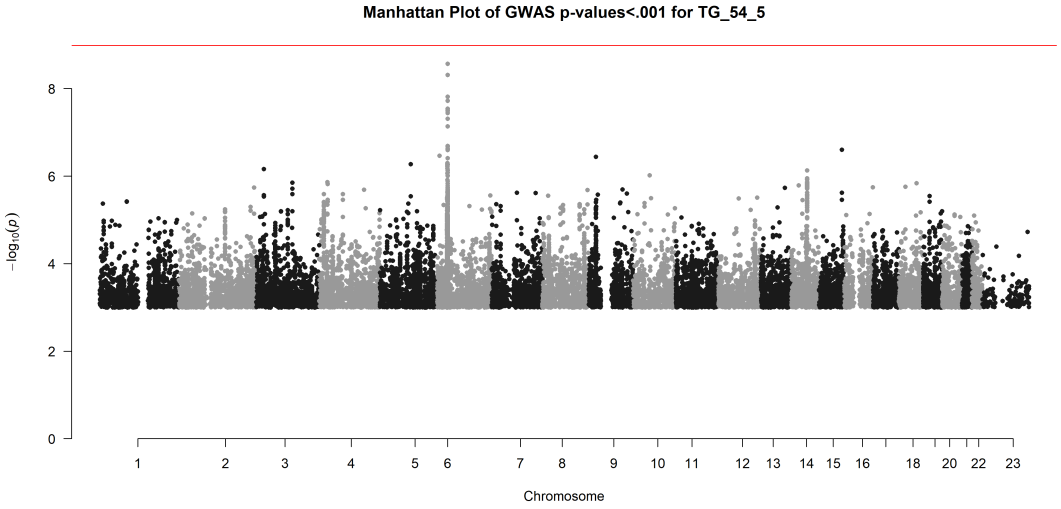


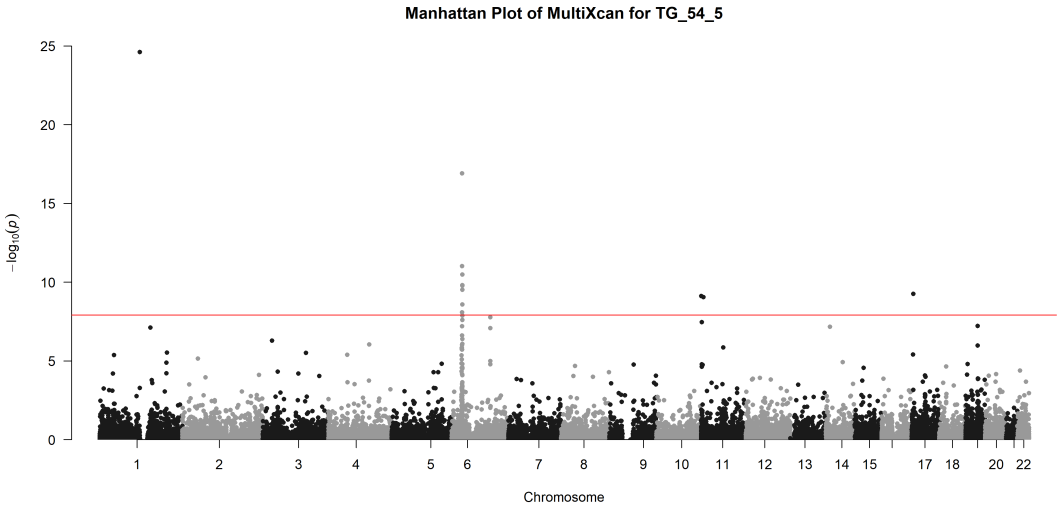
Triglycerol C54:5 (µg/dL)



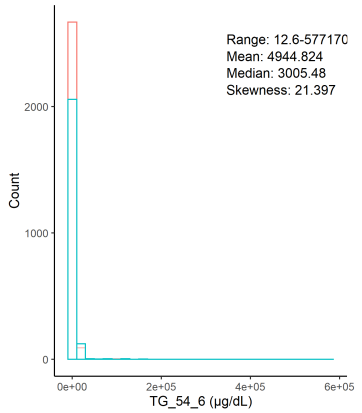


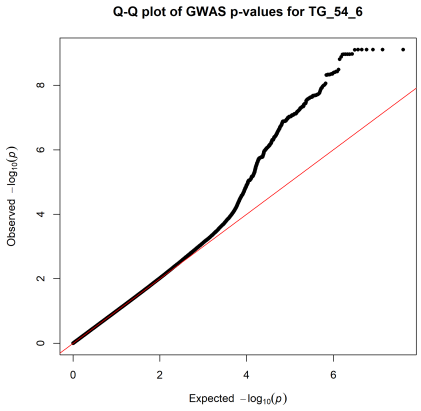


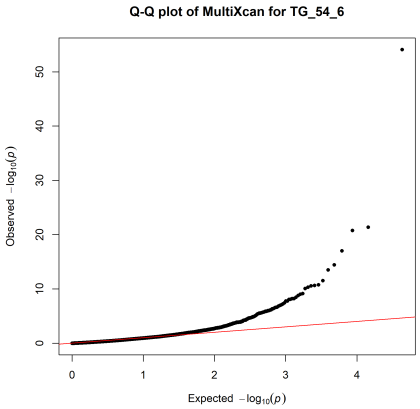


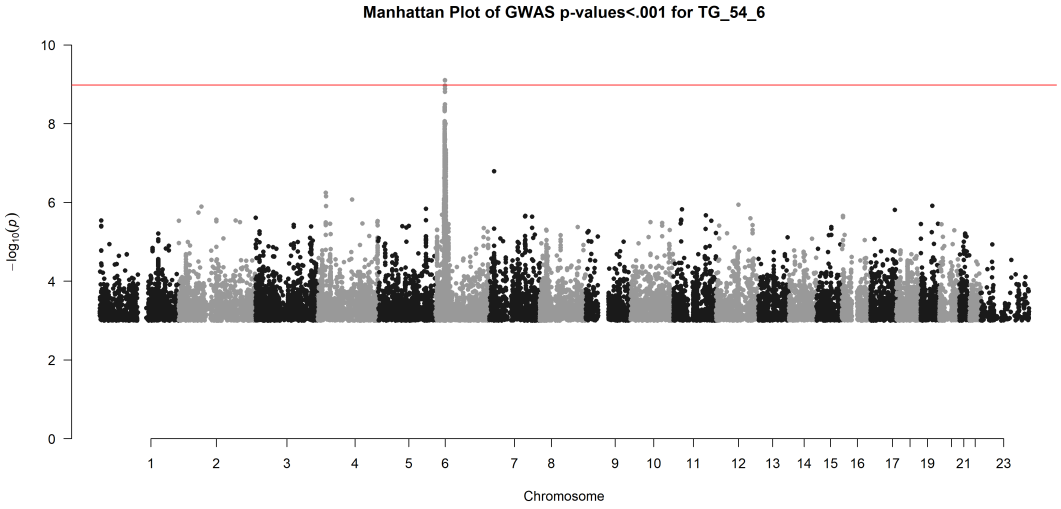


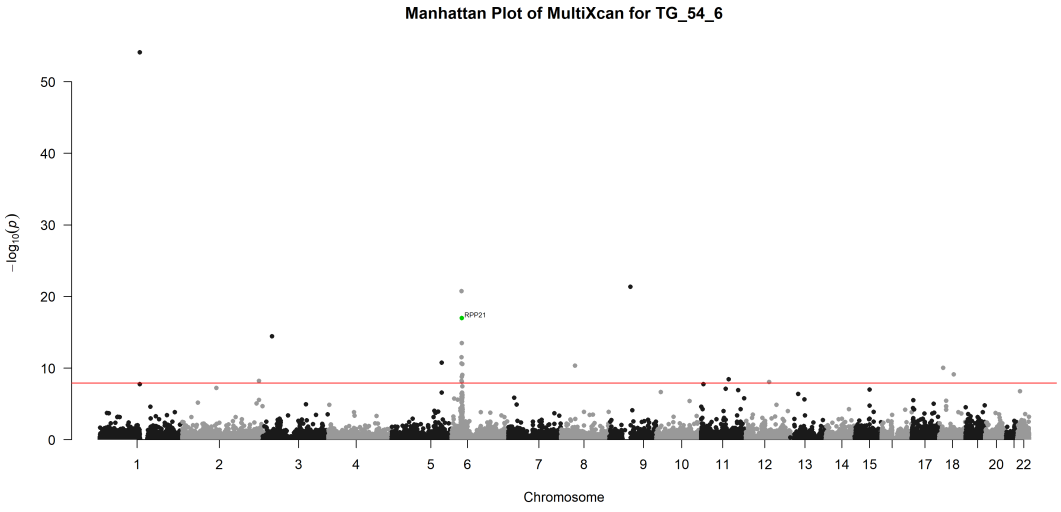
Triglycerol C54:6 (µg/dL)



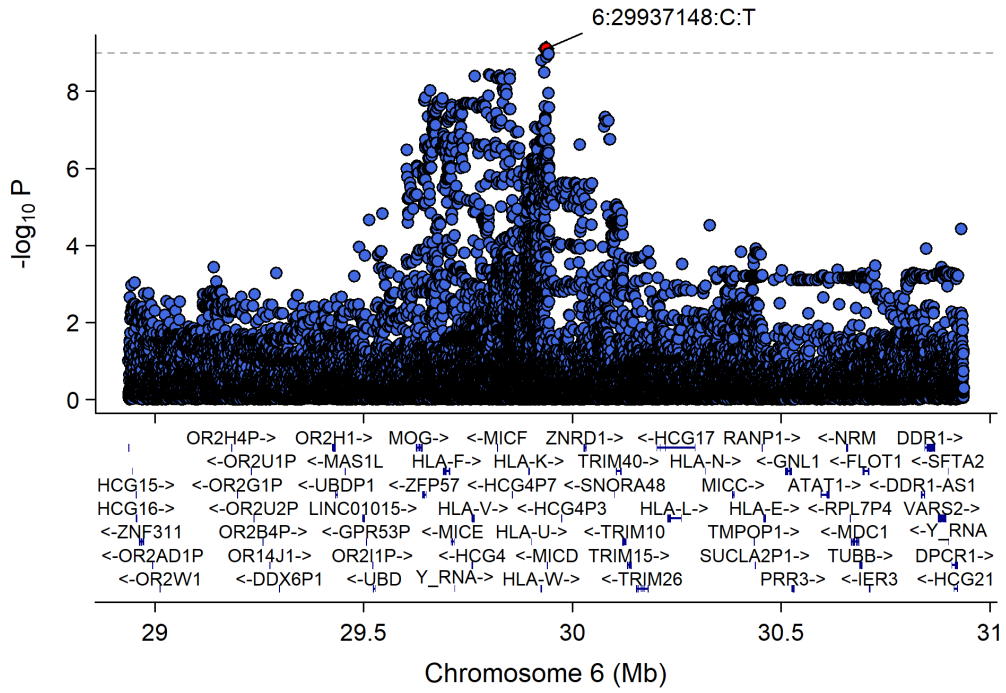




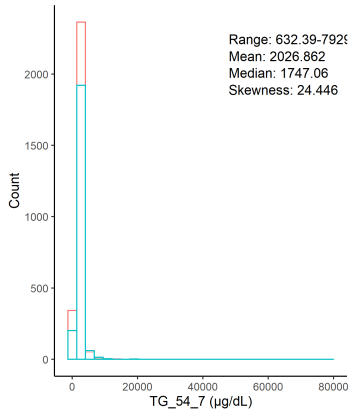


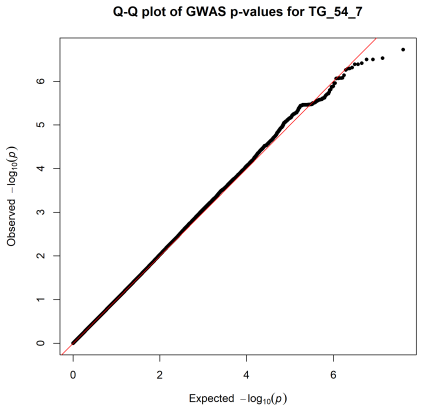


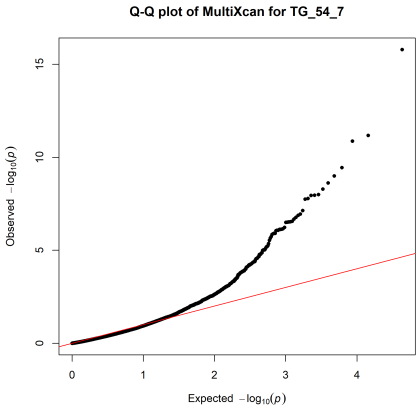
Chr6_24042708_34003583

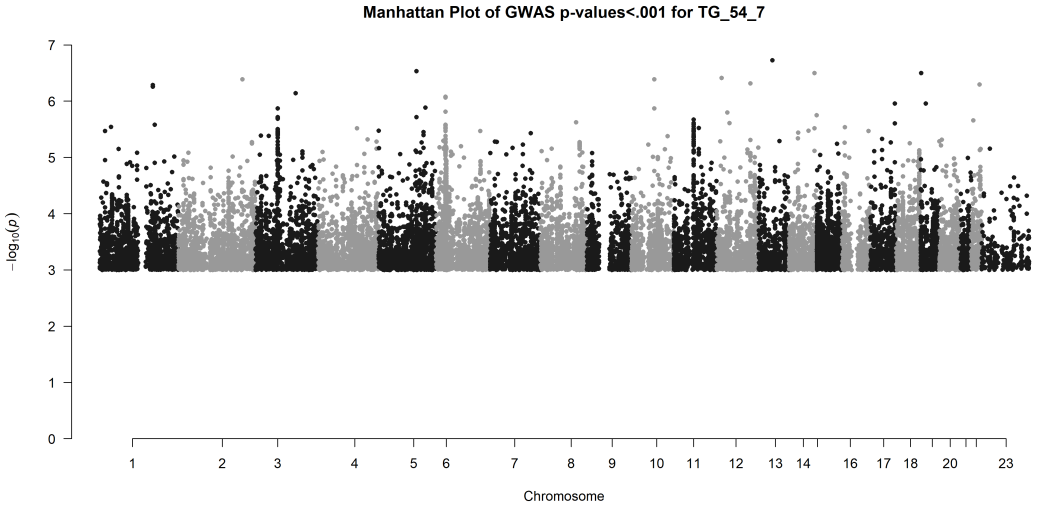


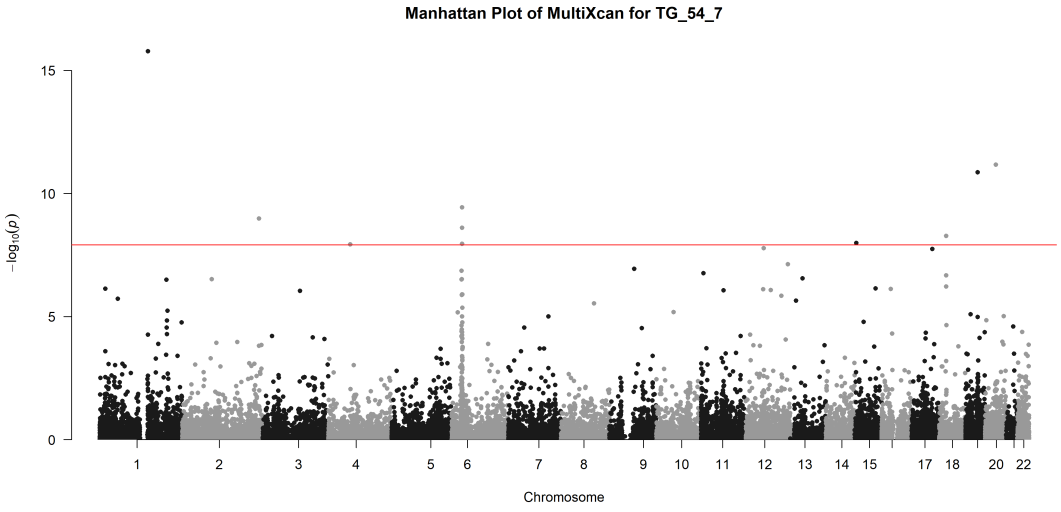
Triglycerol C54:7 (µg/dL)





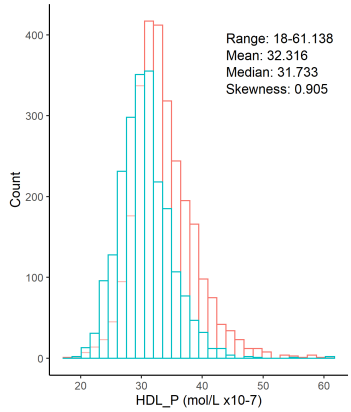


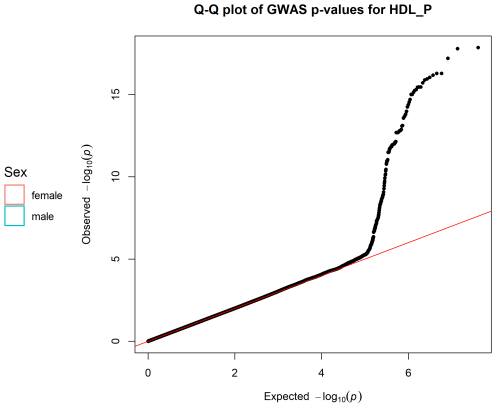


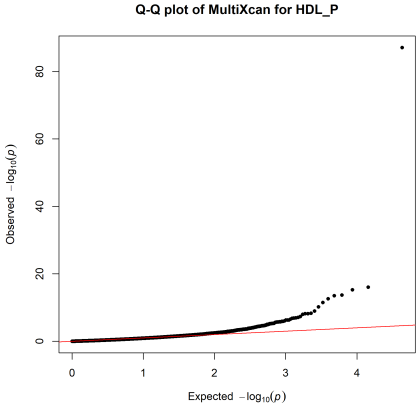


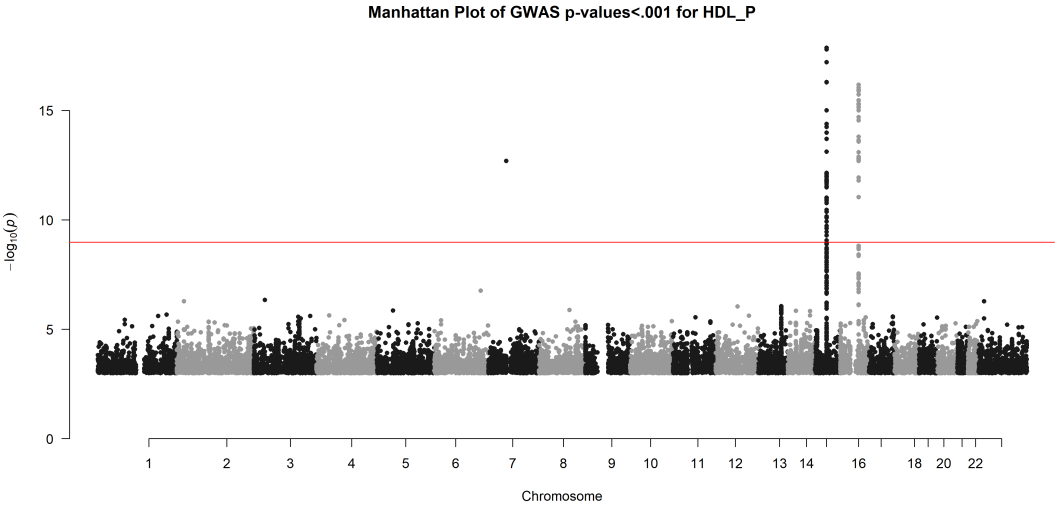
Lipoprotein lipids

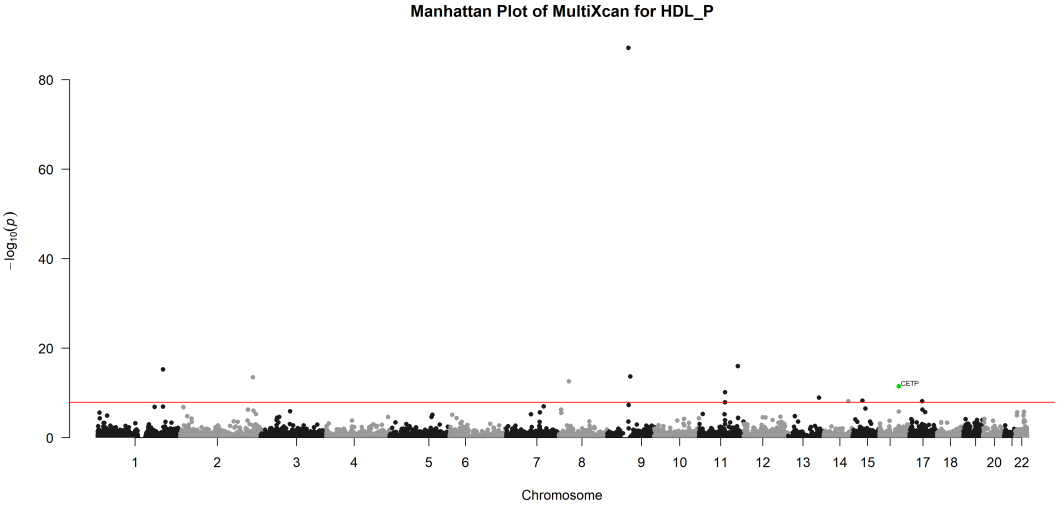
Concentration of HDL particles (mol/l)



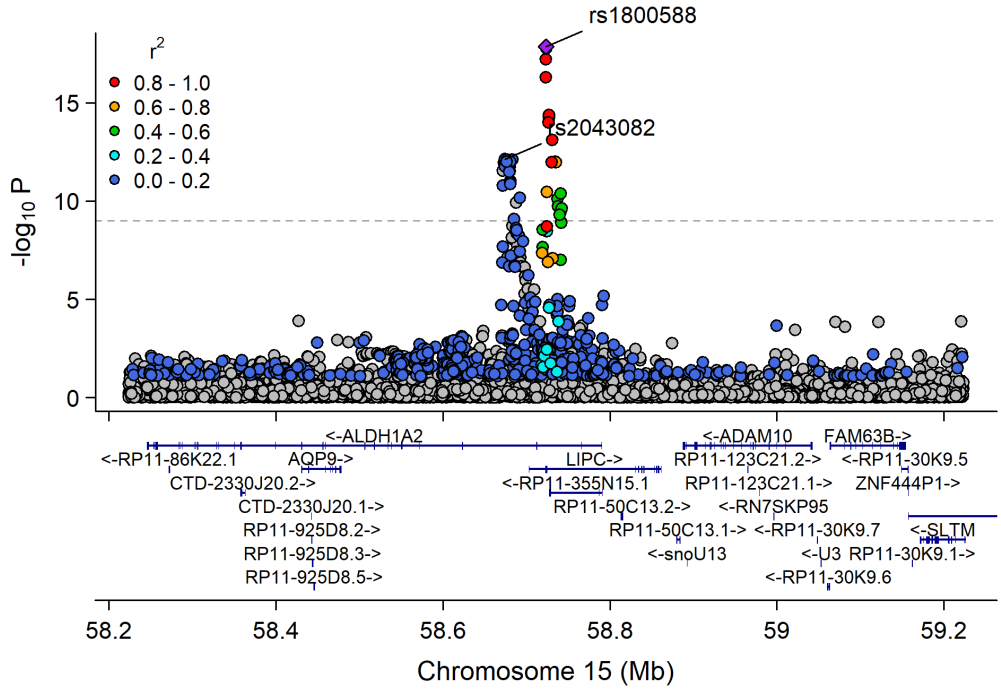




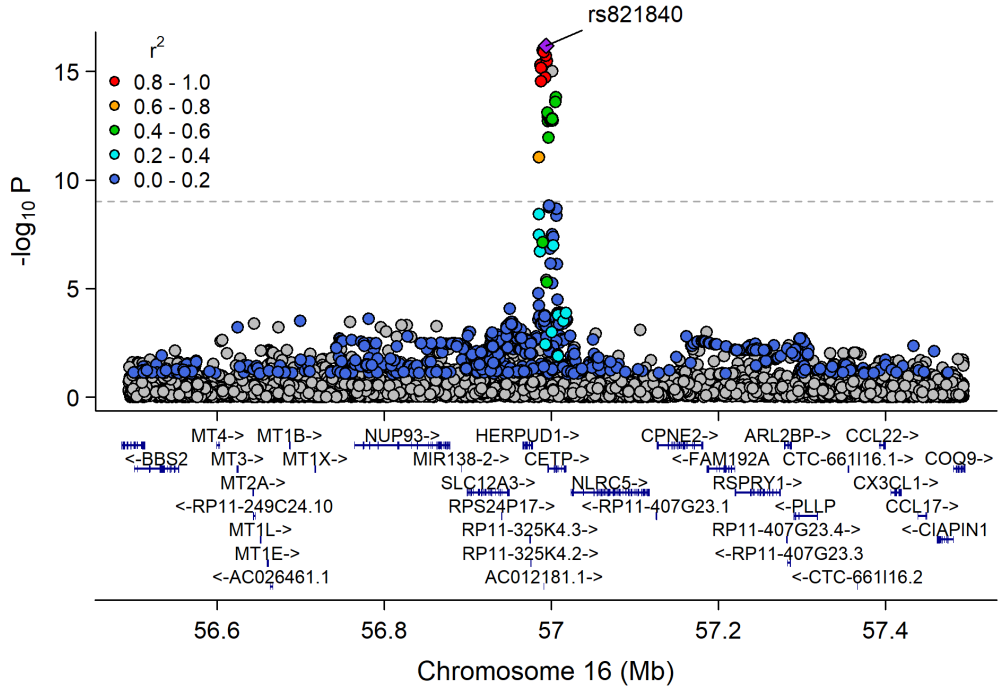




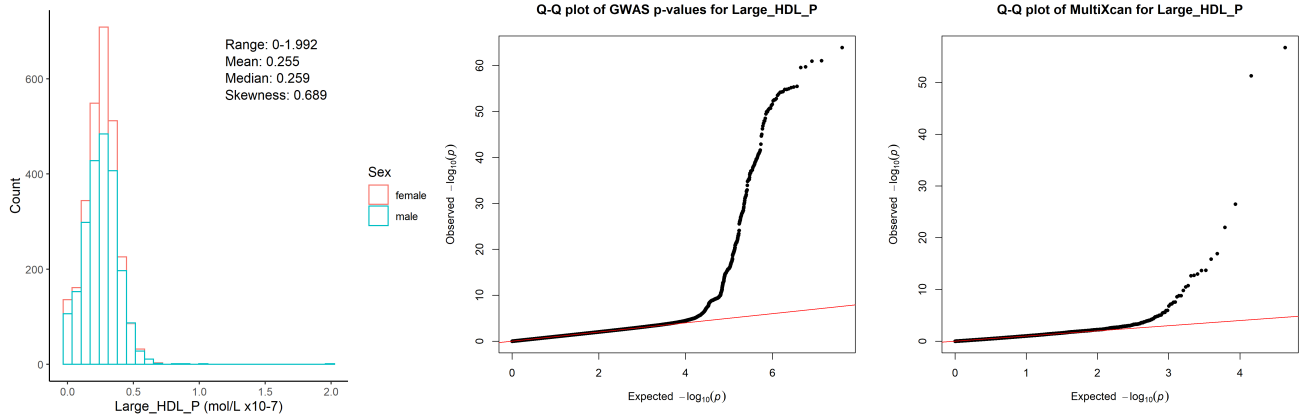
Chr15_57658798_60490883

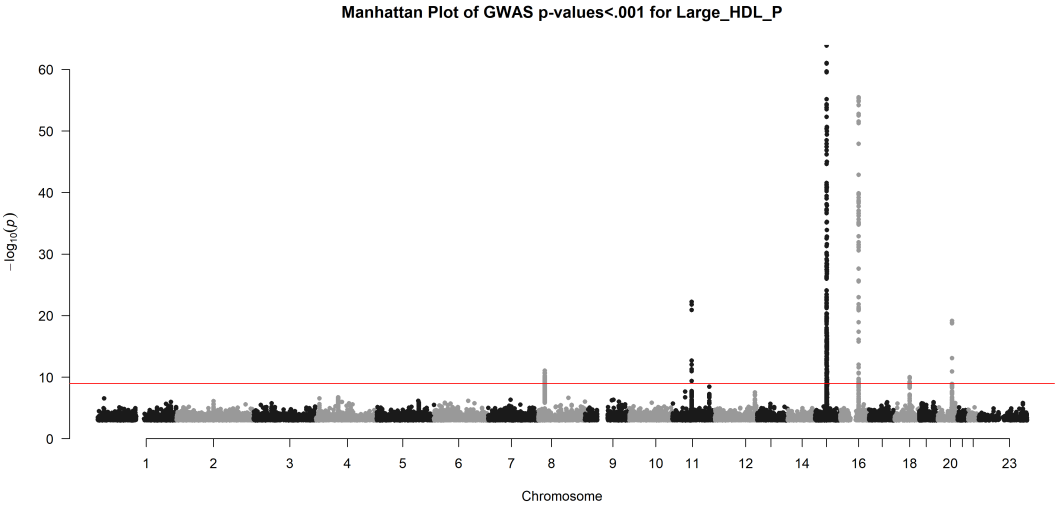


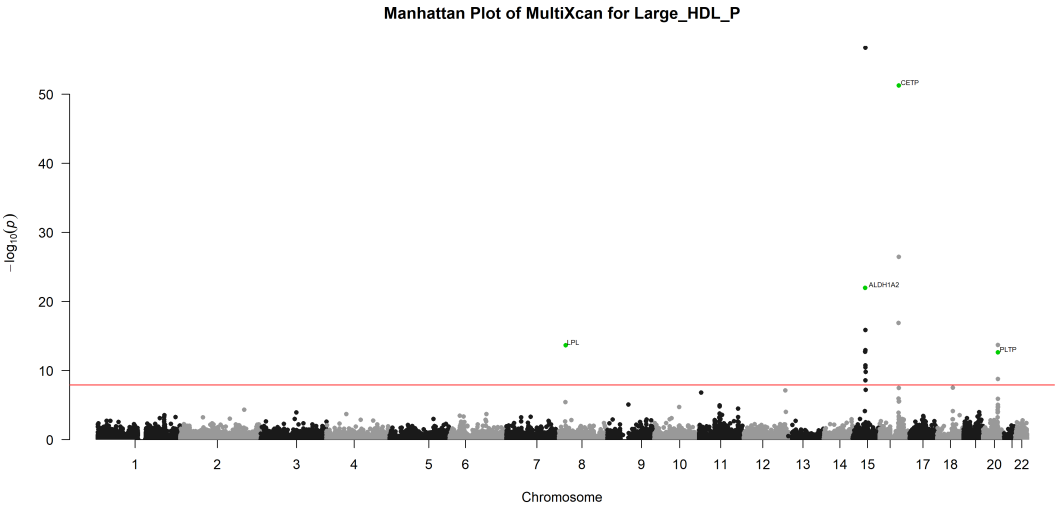
Chr16_55870822_57992421



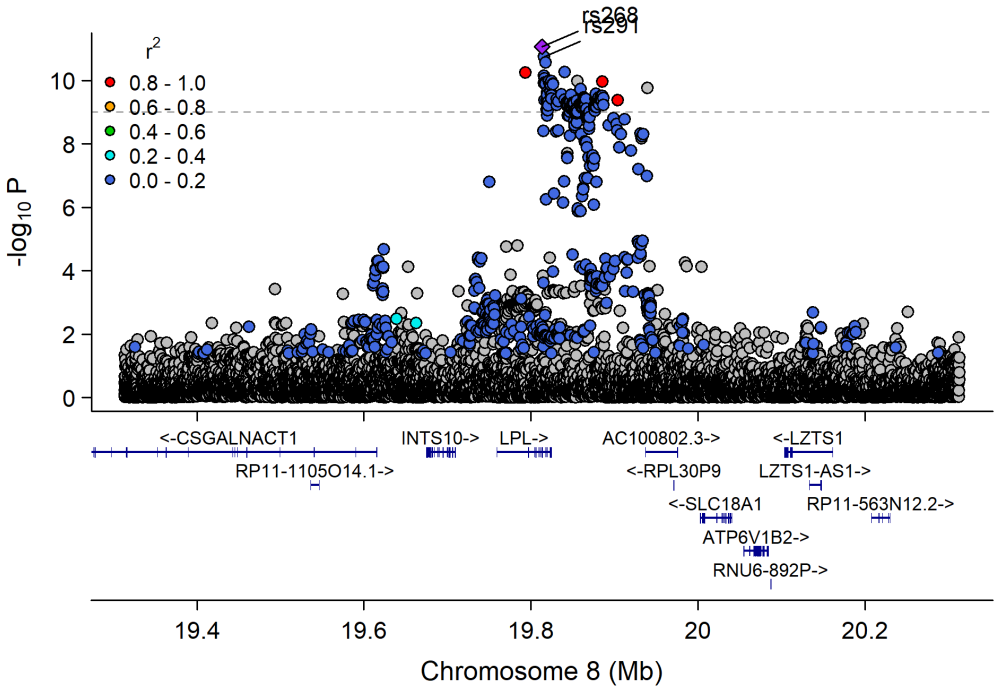
Concentration of large HDL particles (mol/l)



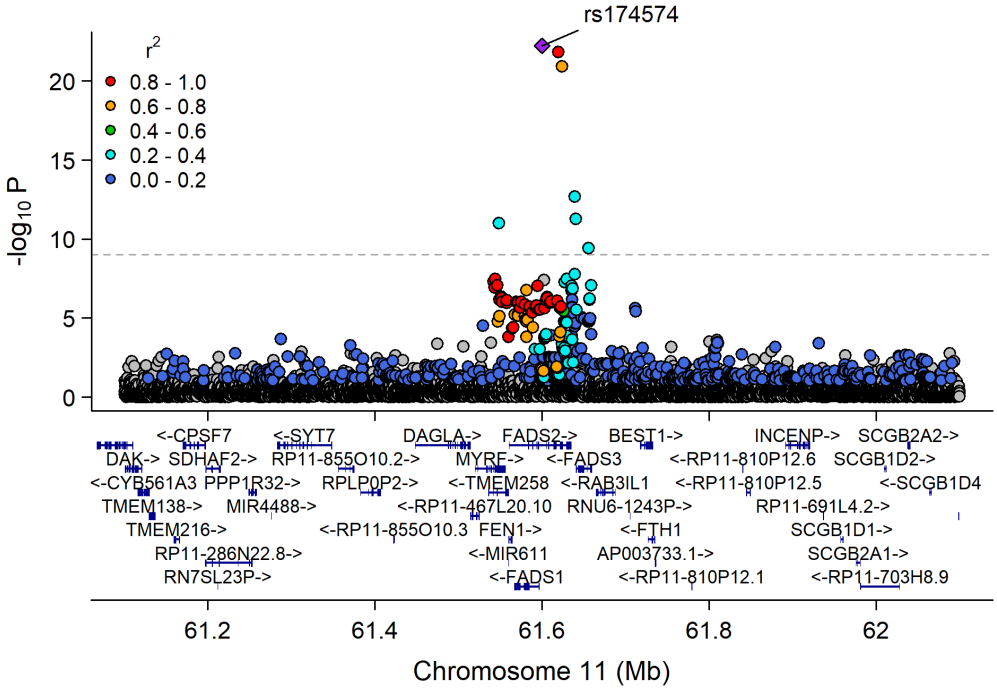




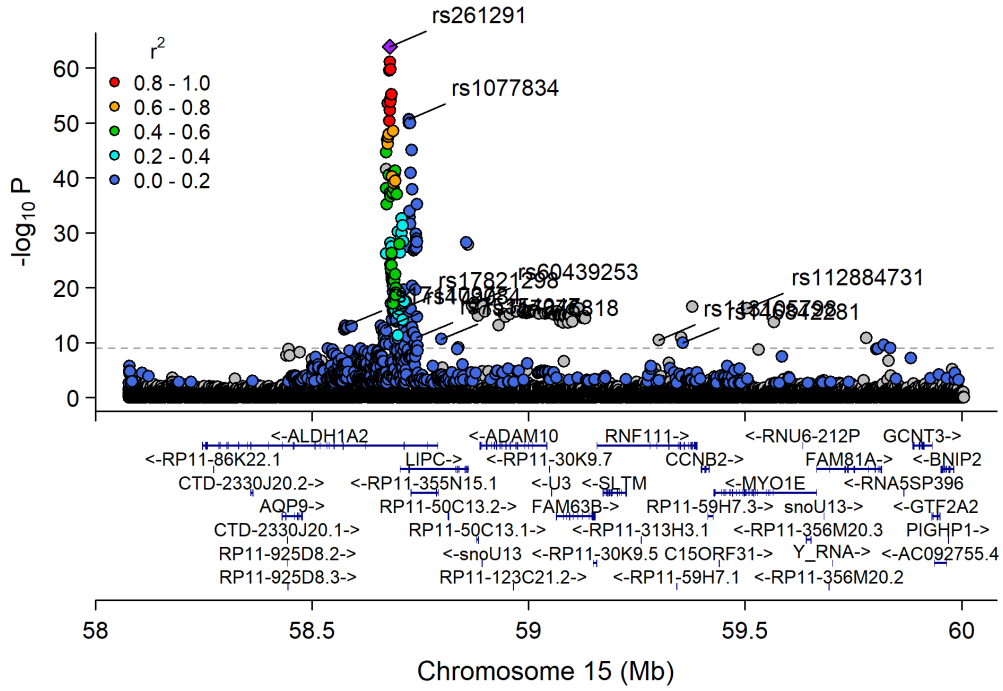
Chr8_18582620_20792744



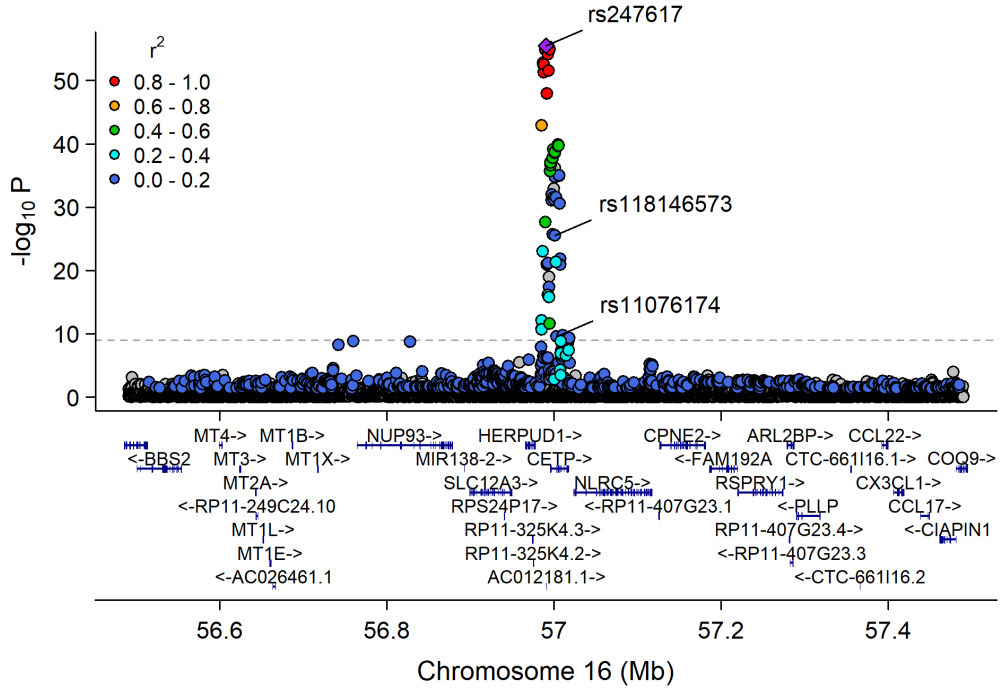
Chr11_59978355_62914375



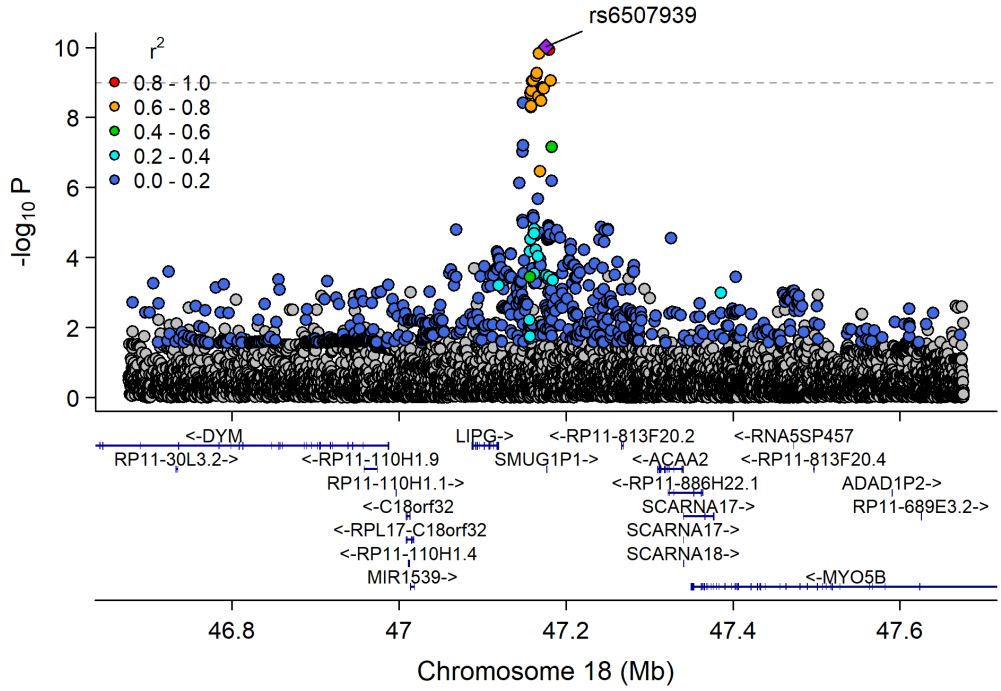
Chr15_57658798_60490883



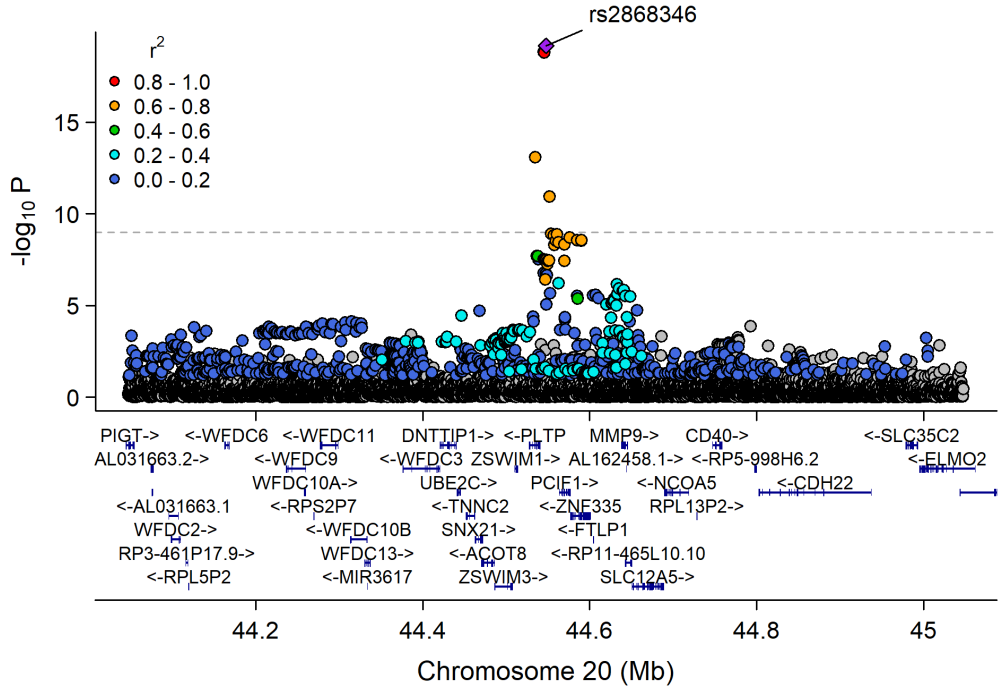
Chr16_55870822_57992421



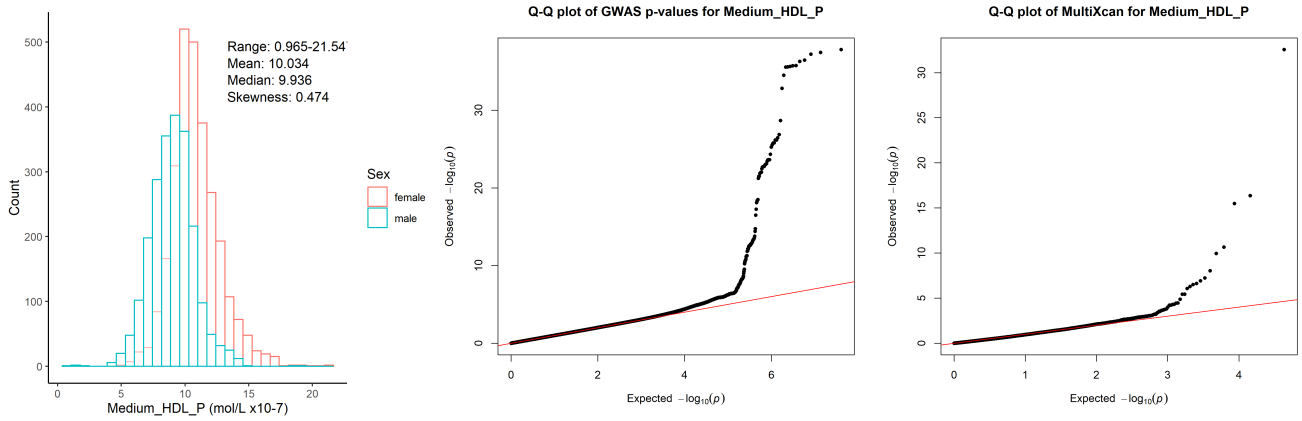
Chr18_46534938_47960800

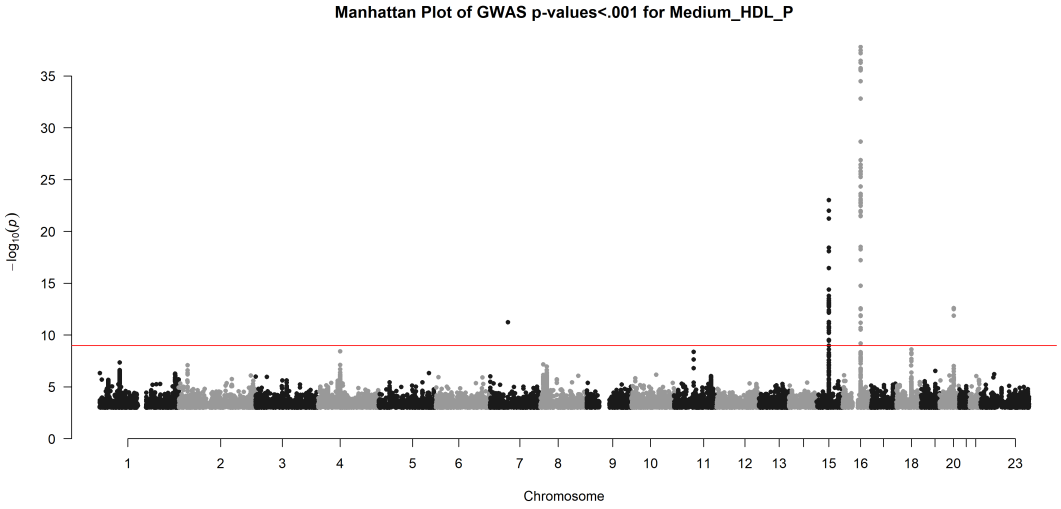


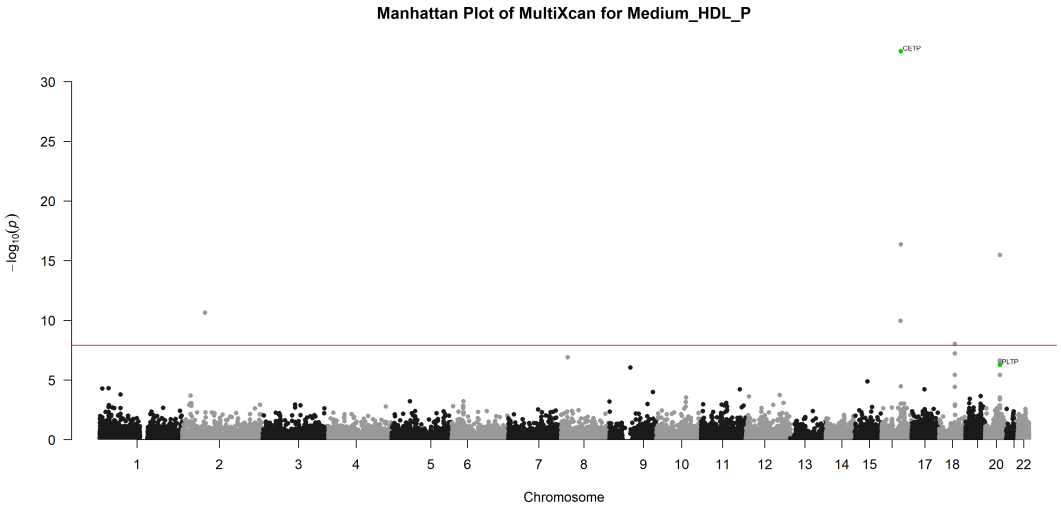
Chr20_42232254_45611445



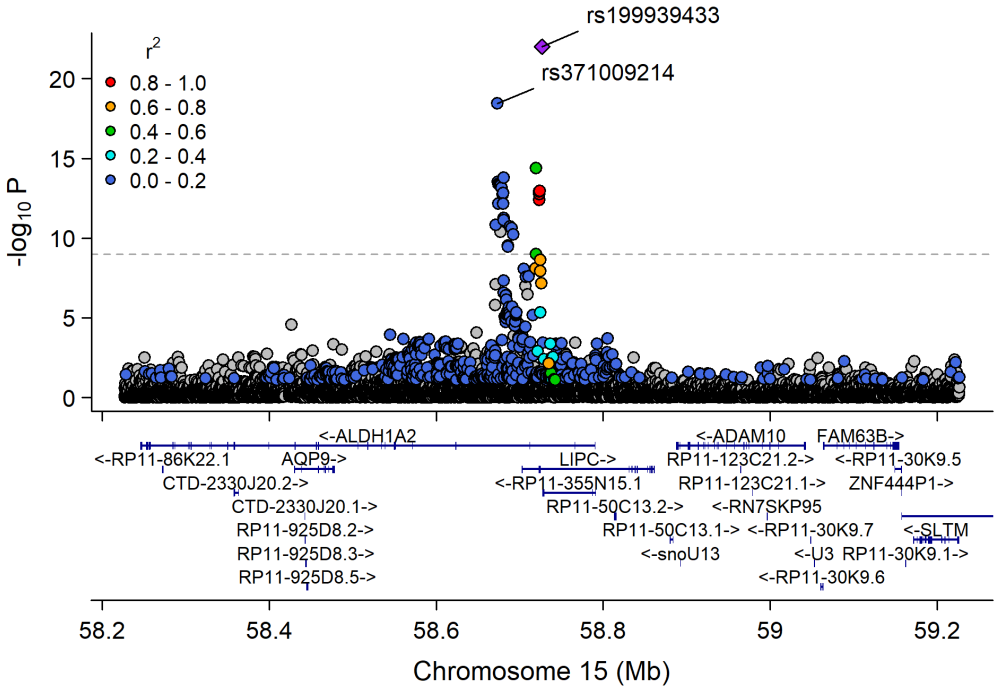
Concentration of medium HDL particles (mol/l)



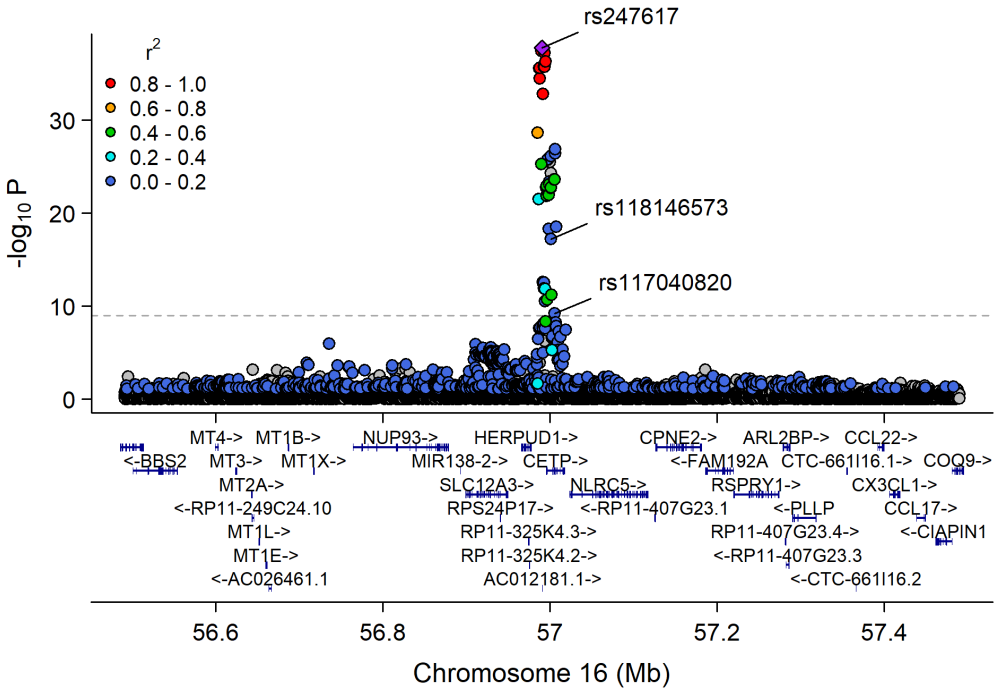




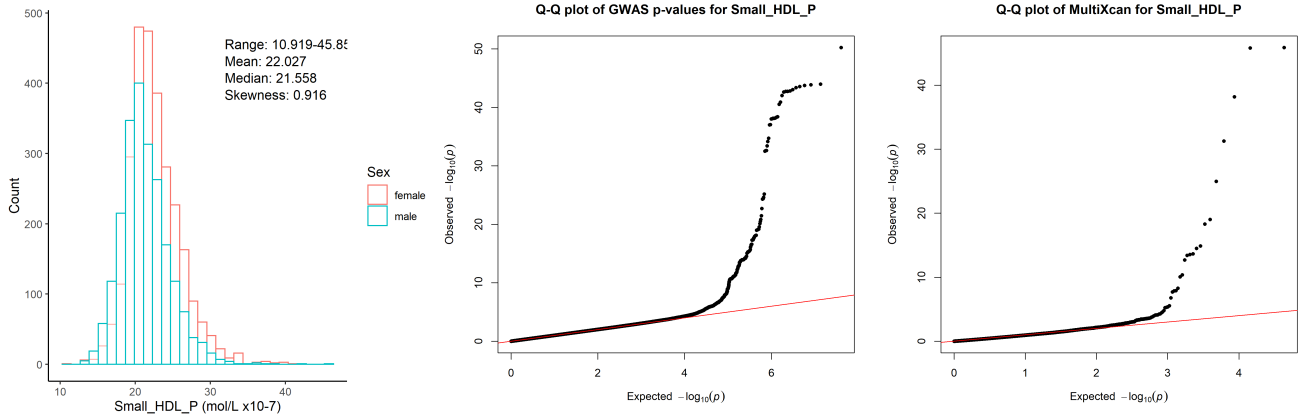
Chr15_57658798_60490883

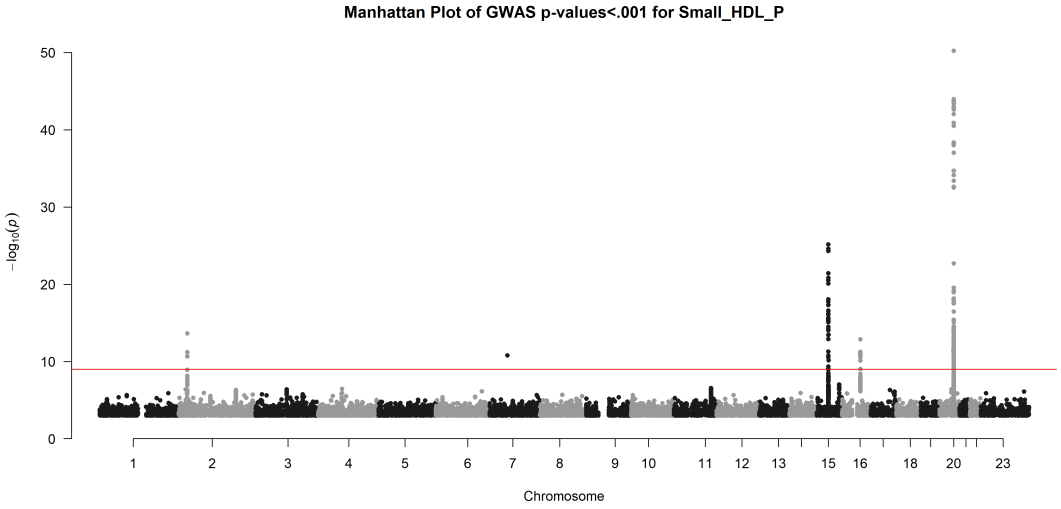


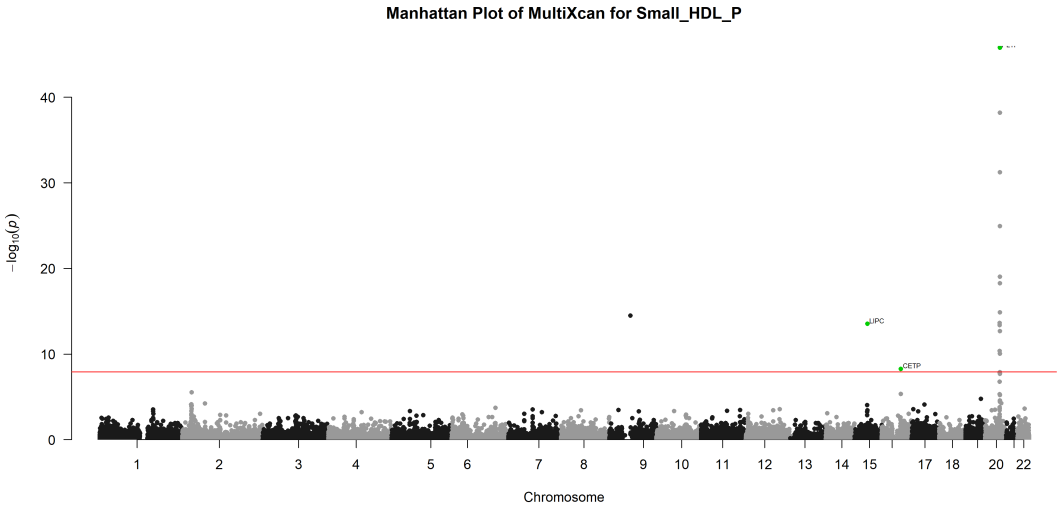
Chr16_55870822_57992421



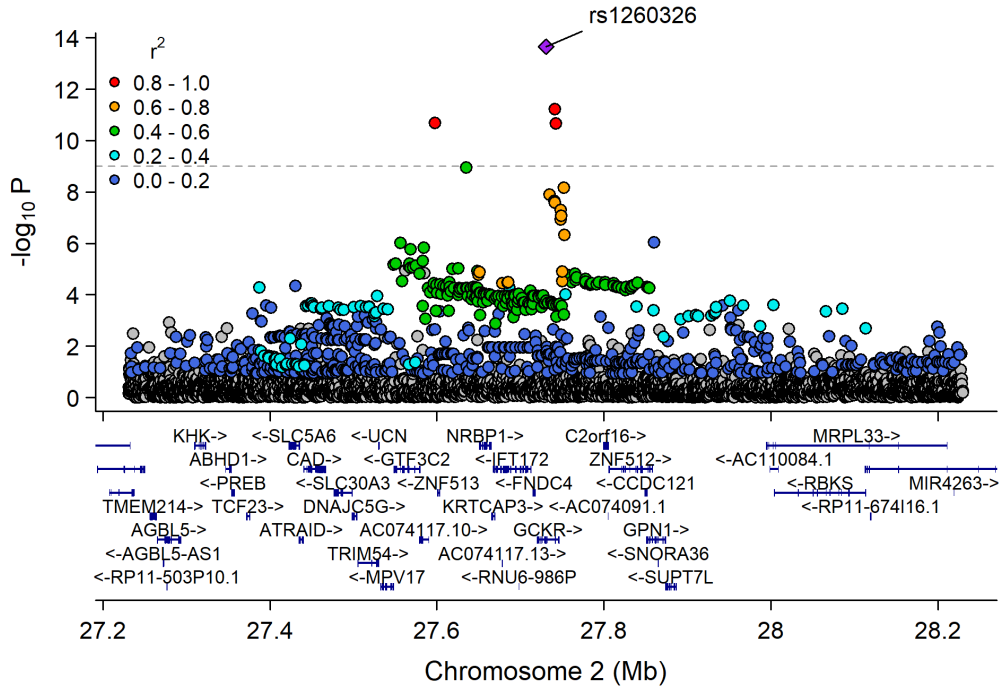
Concentration of small HDL particles (mol/l)



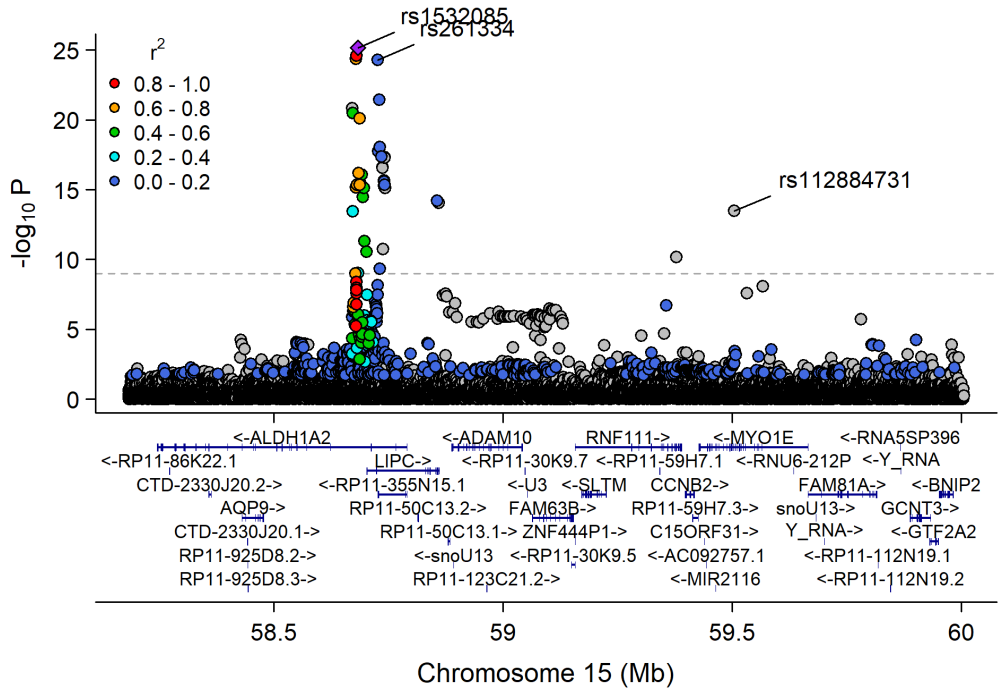




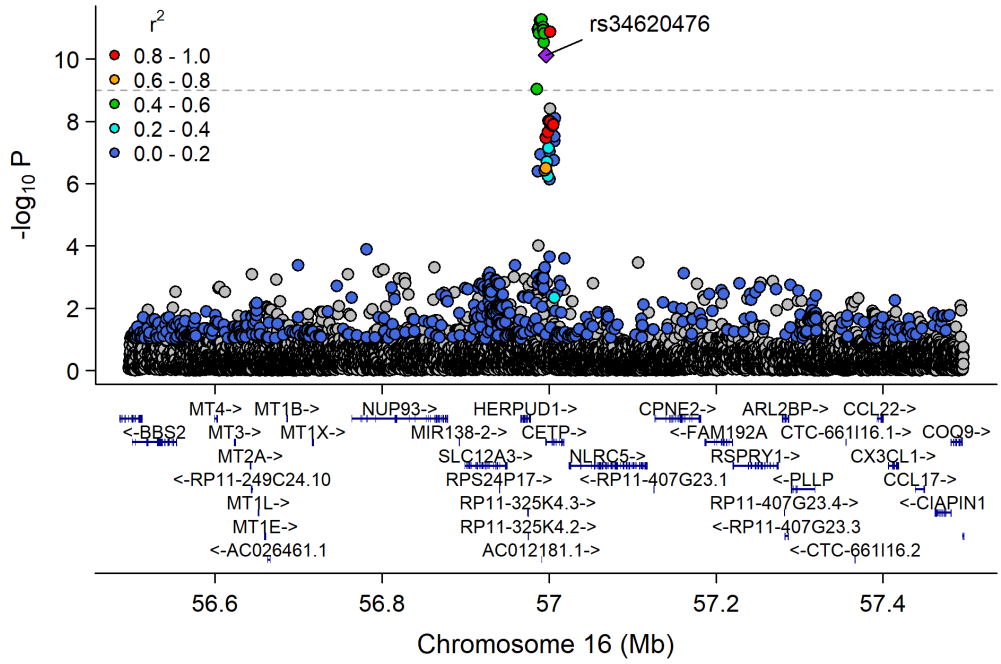
Chr2_26753815_28597624



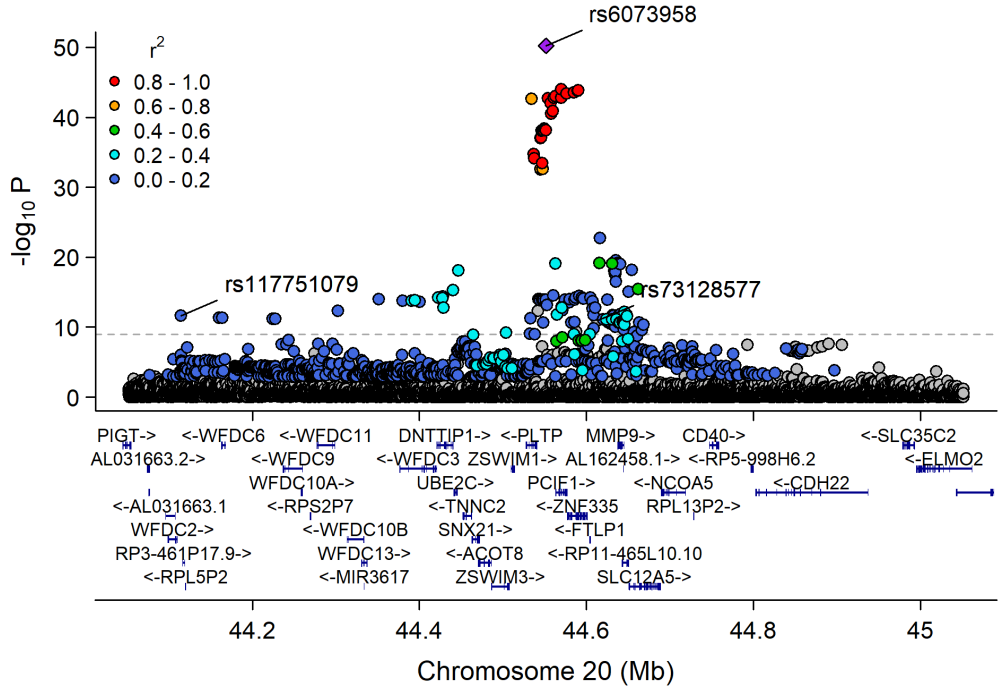
Chr15_57658798_60490883



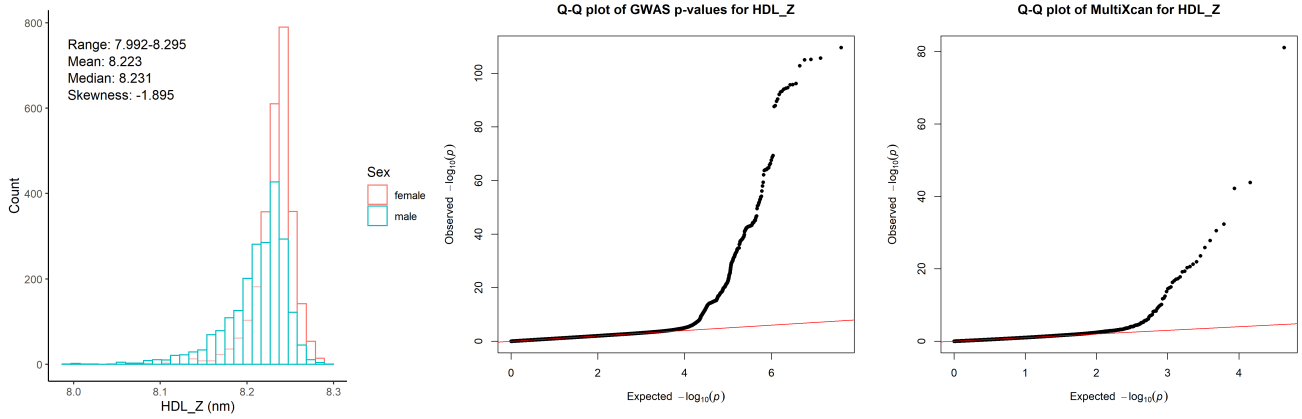
Chr16_55870822_57992421

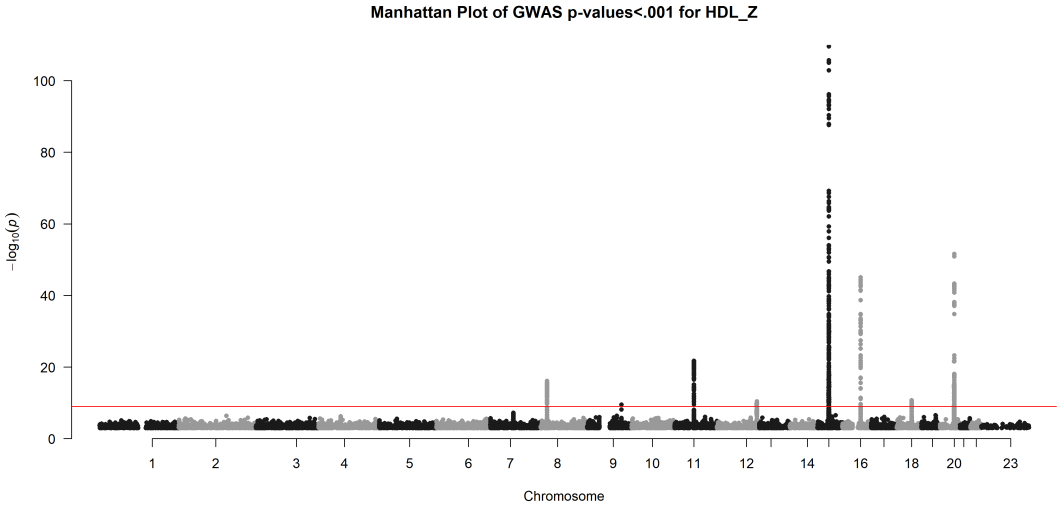


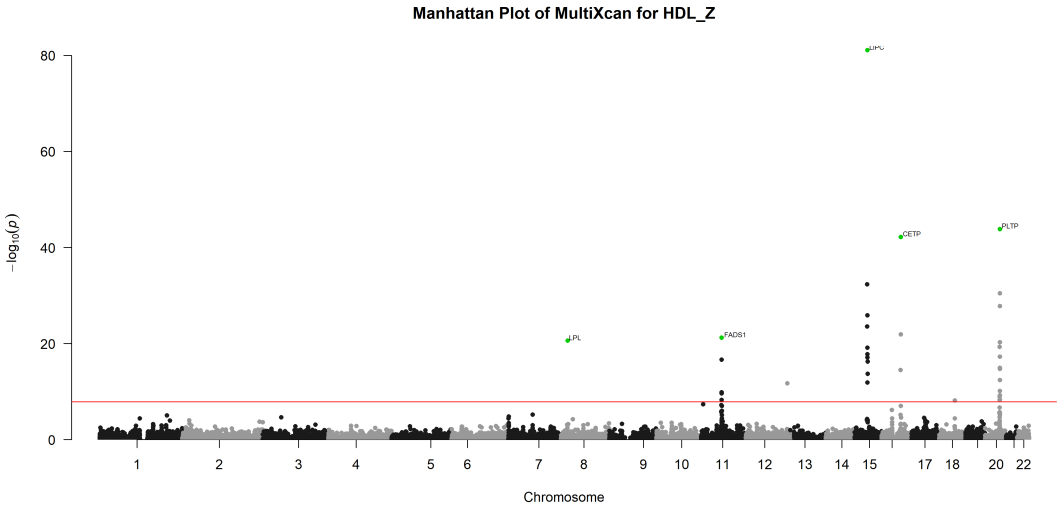
Chr20_42232254_45611445



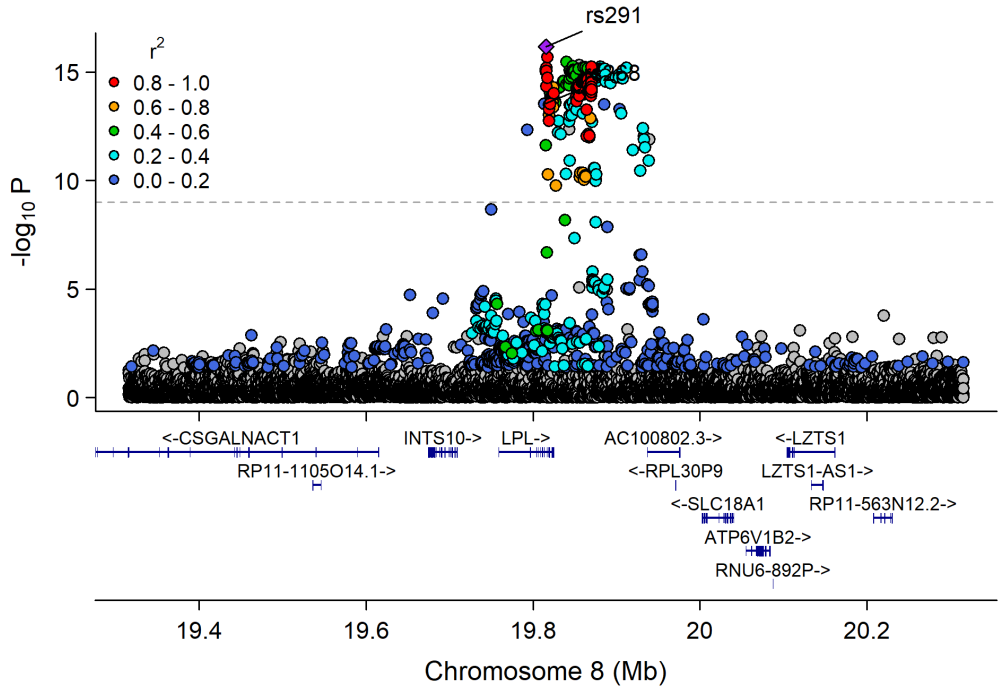
Mean diameter for HDL particles (nm)



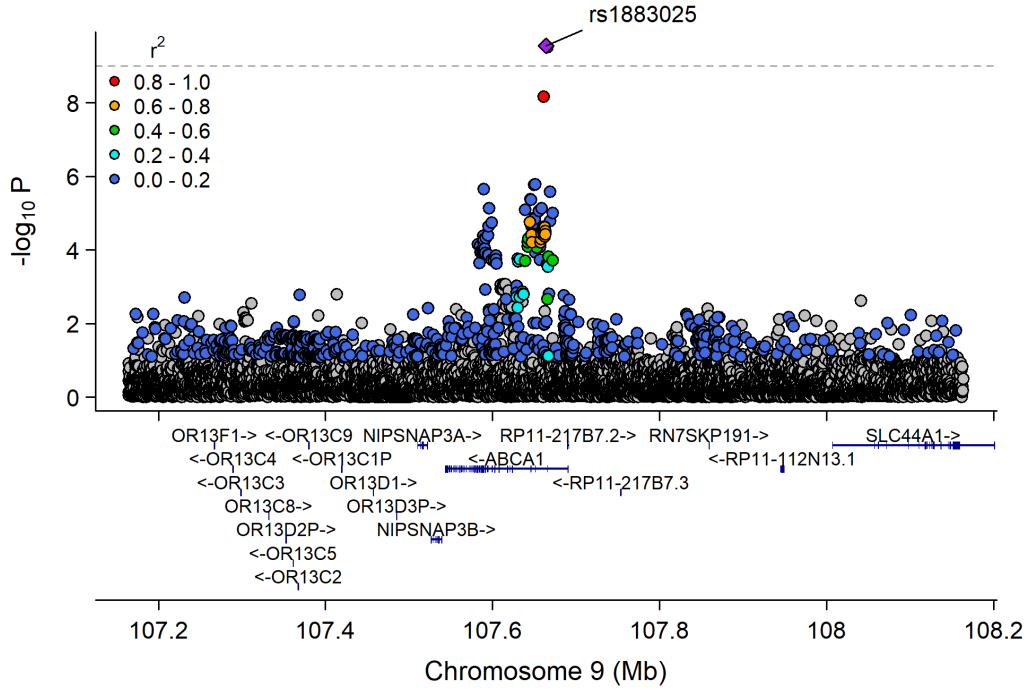




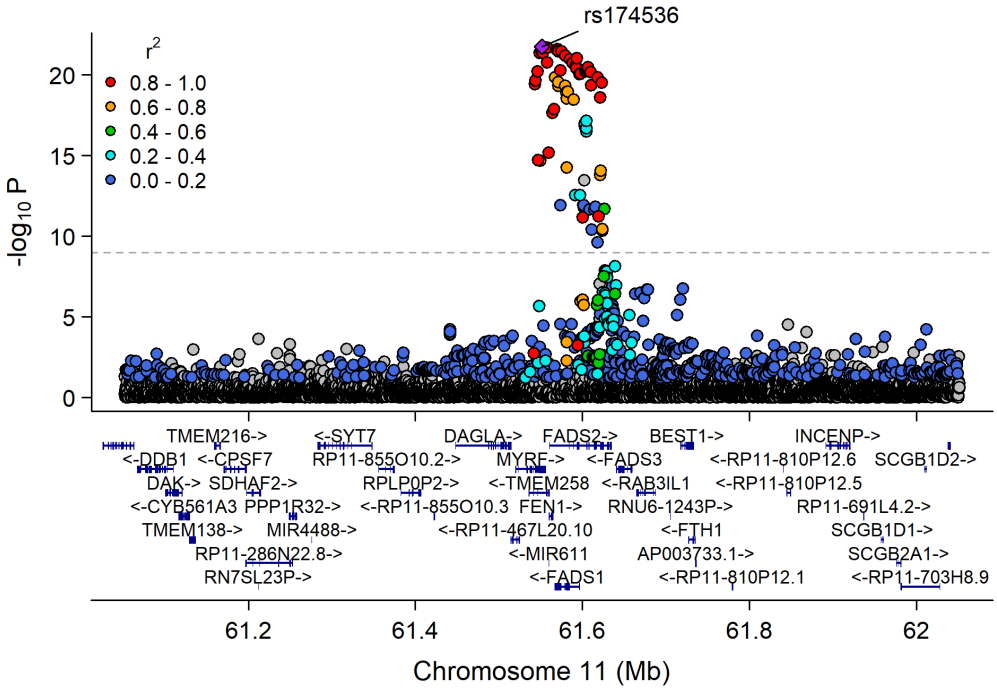
Chr8_18582620_20792744



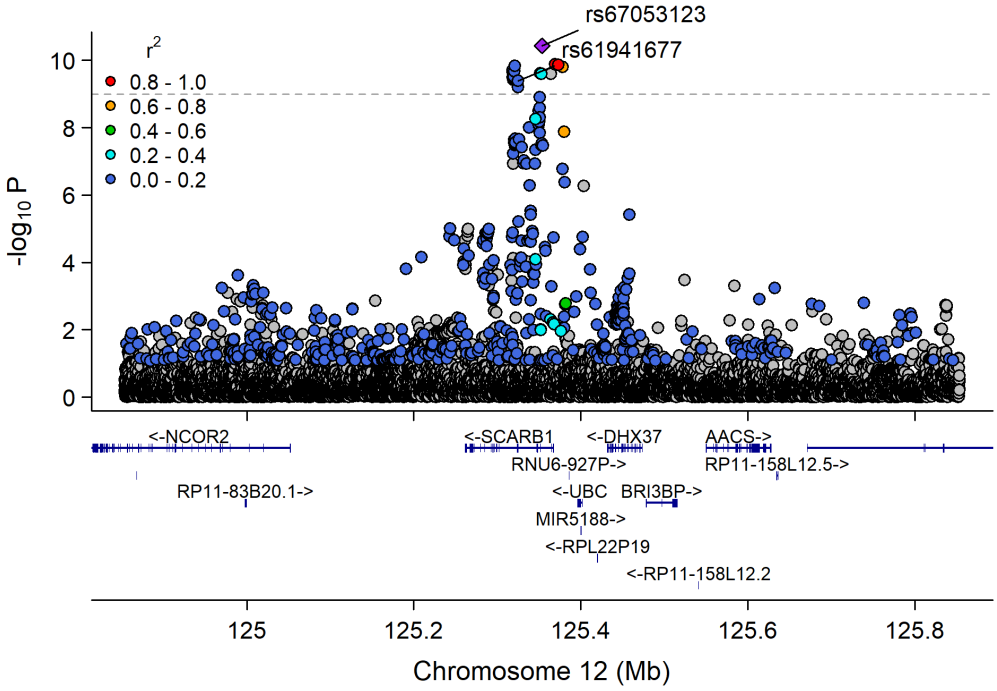
Chr9_107523863_107672526



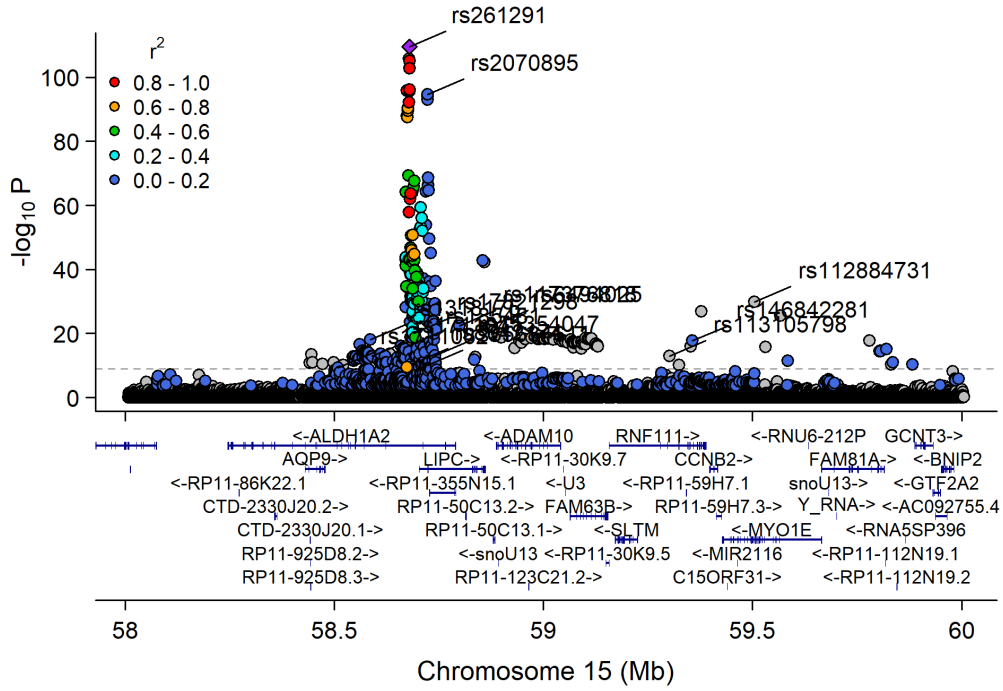
Chr11_59978355_62914375



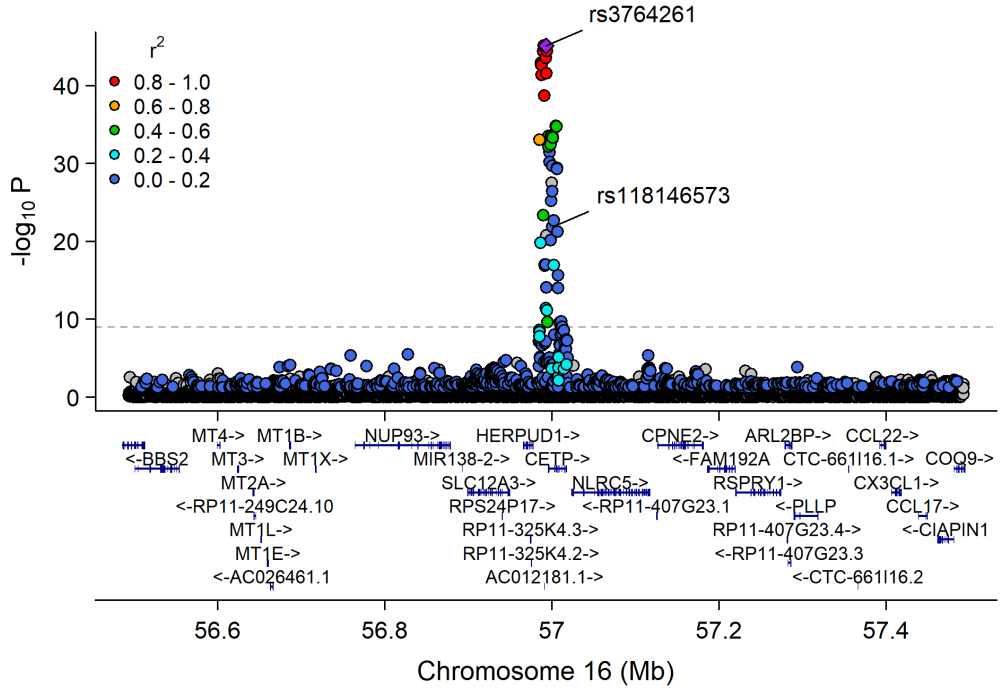
Chr12_125138706_125512423



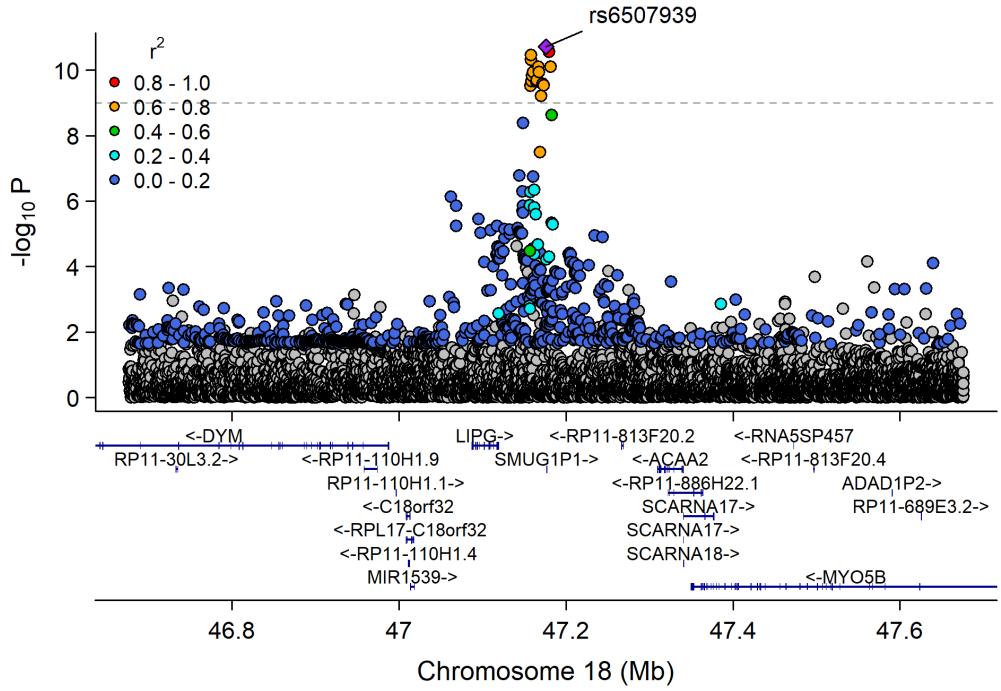
Chr15_57658798_60490883



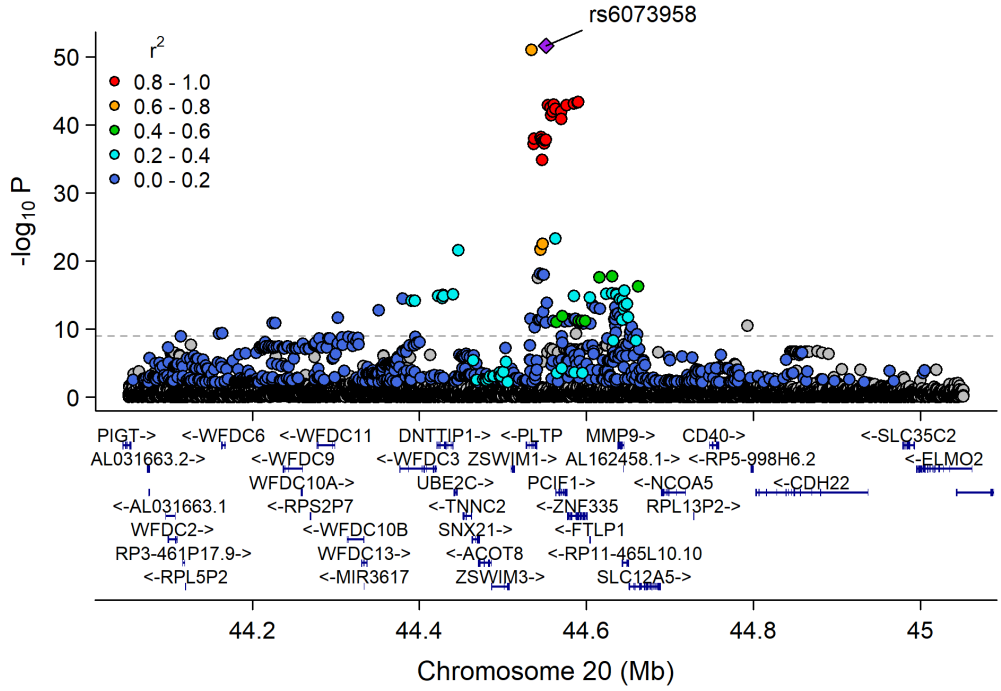
Chr16_55870822_57992421



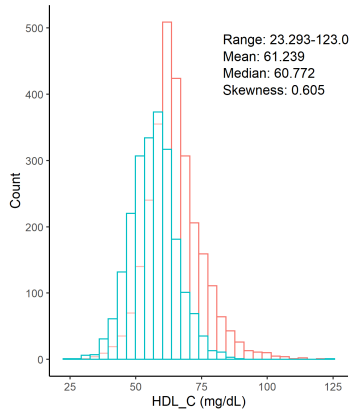
Chr18_46534938_47960800

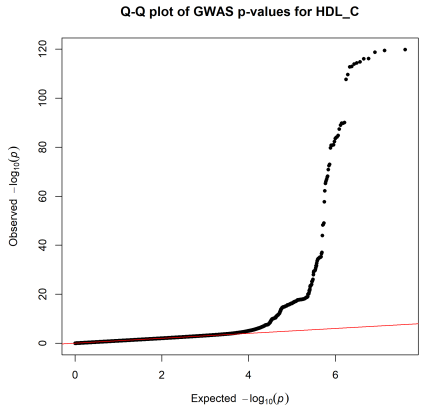


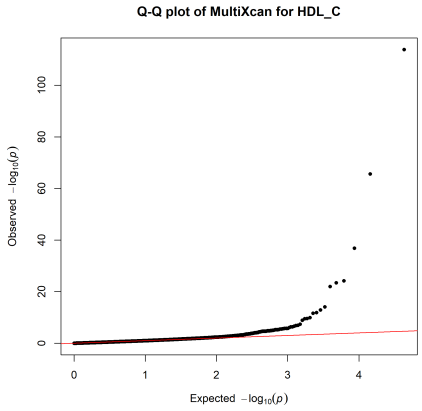
Chr20_42232254_45611445

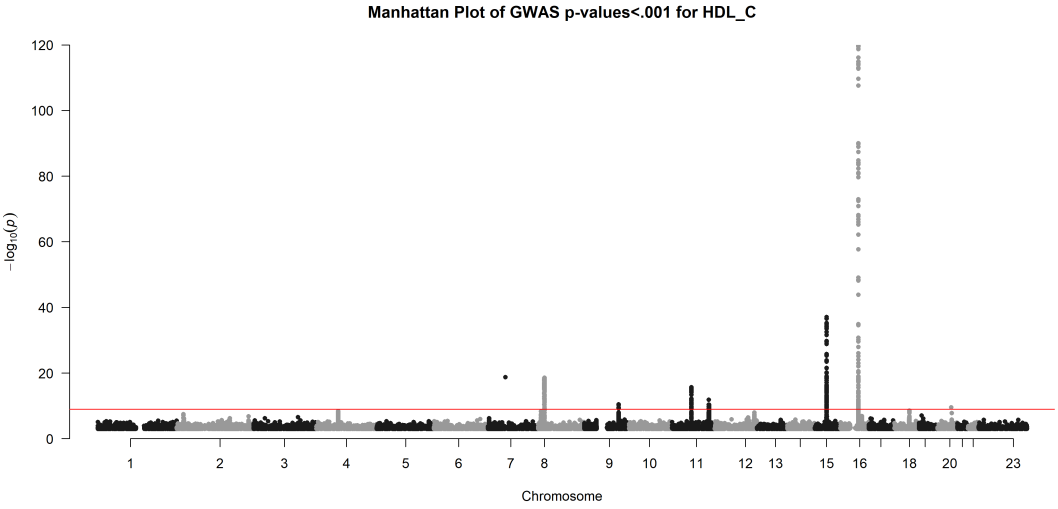


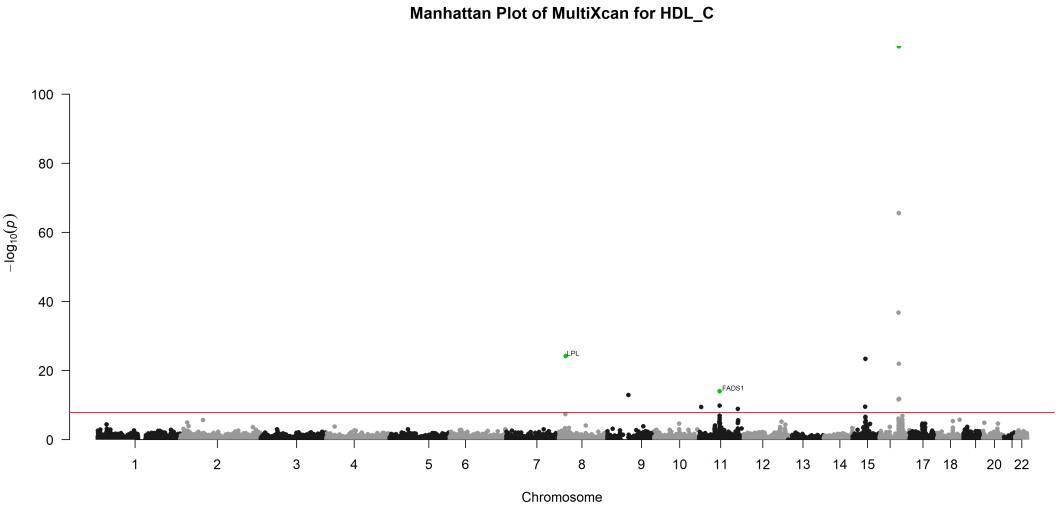
Total cholesterol in HDL (mg/dL)



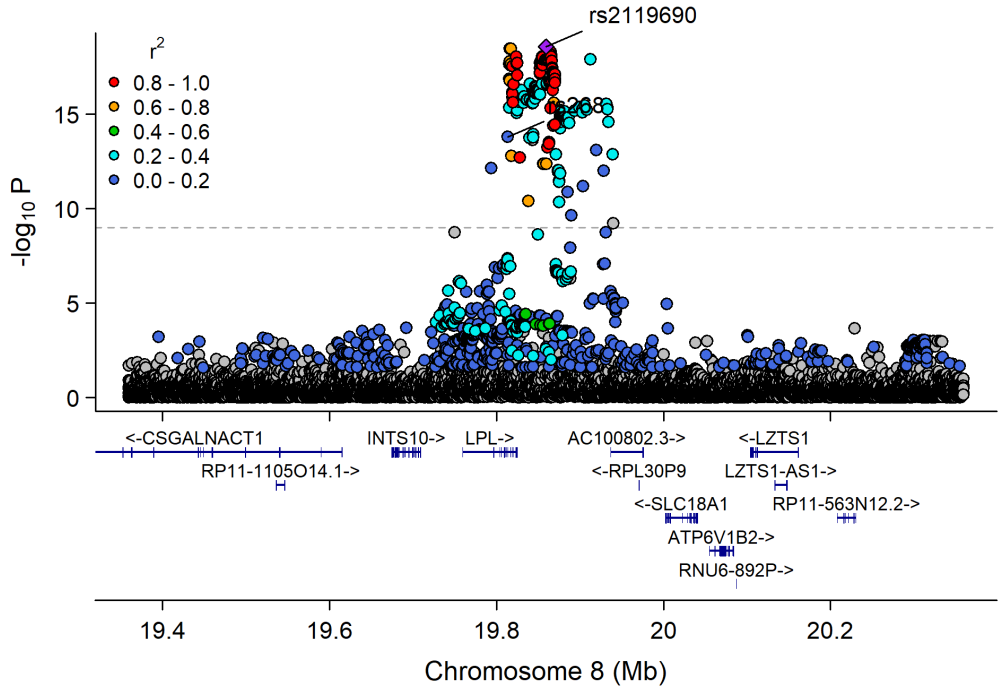




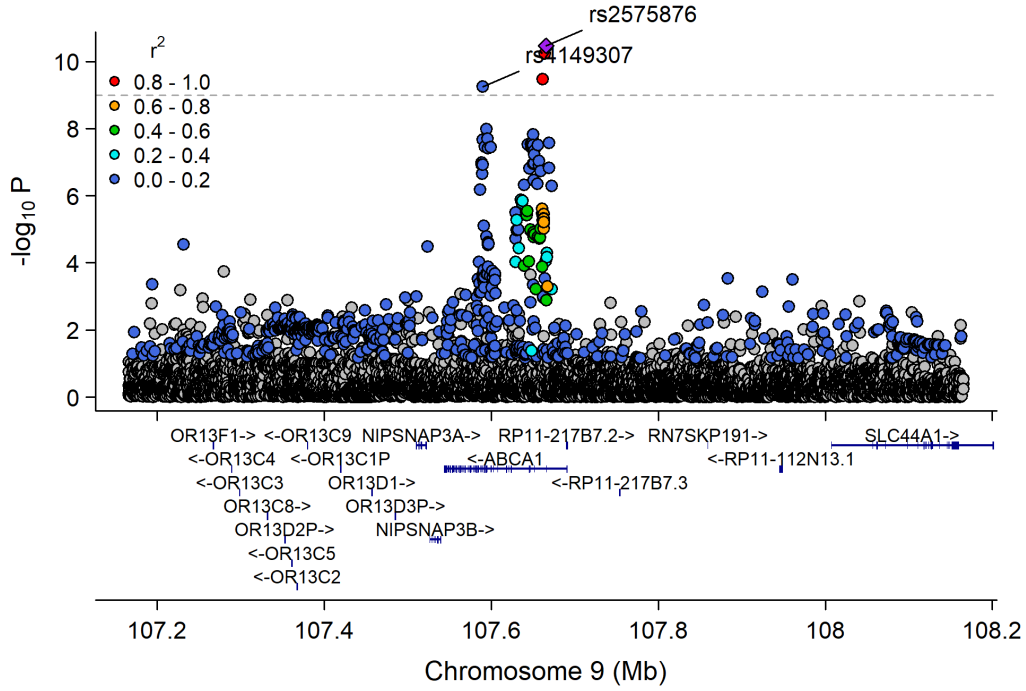




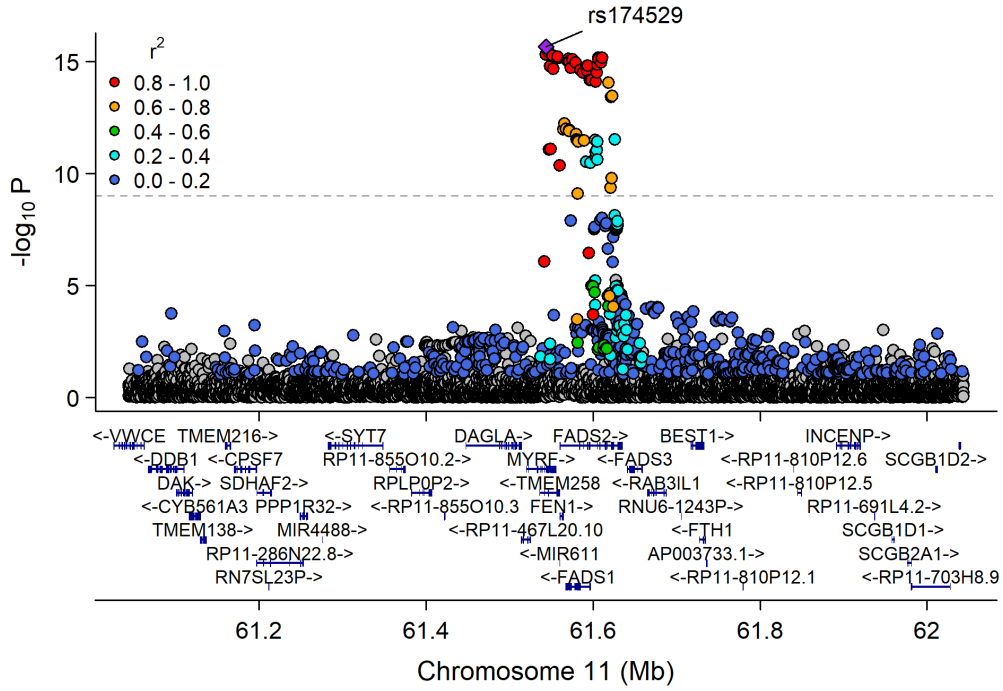
Chr8_18582620_20792744



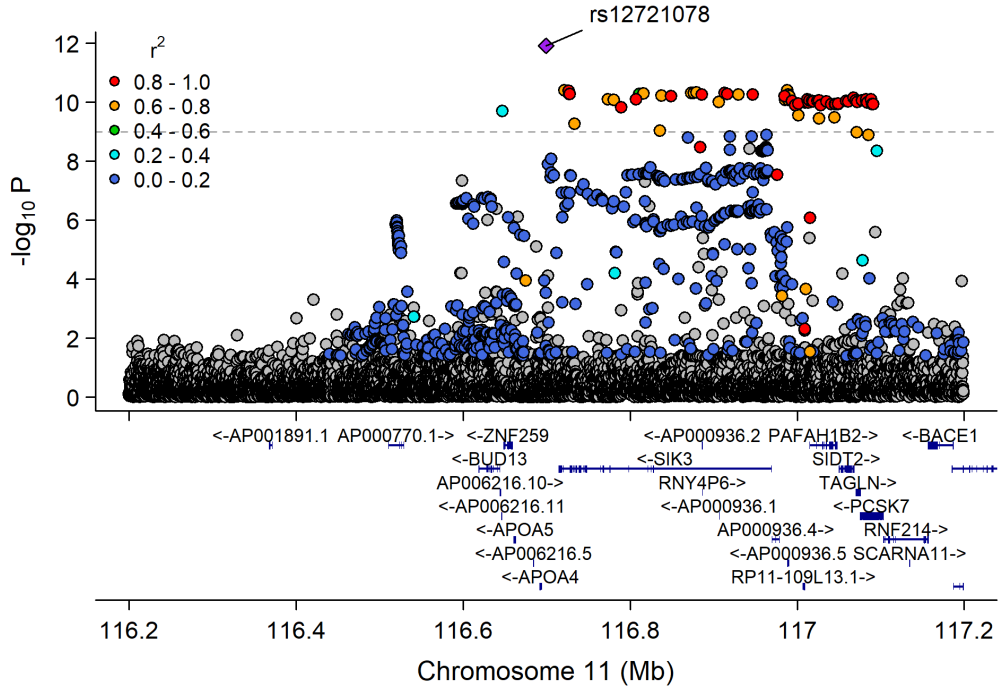
Chr9_107523863_107672526



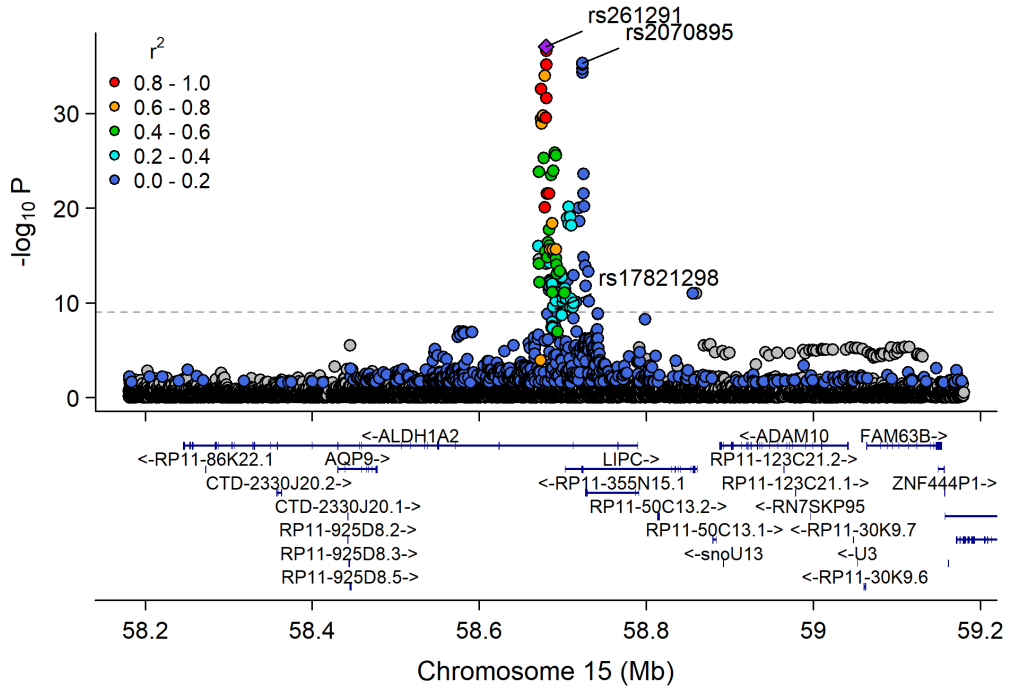
Chr11_59978355_62914375



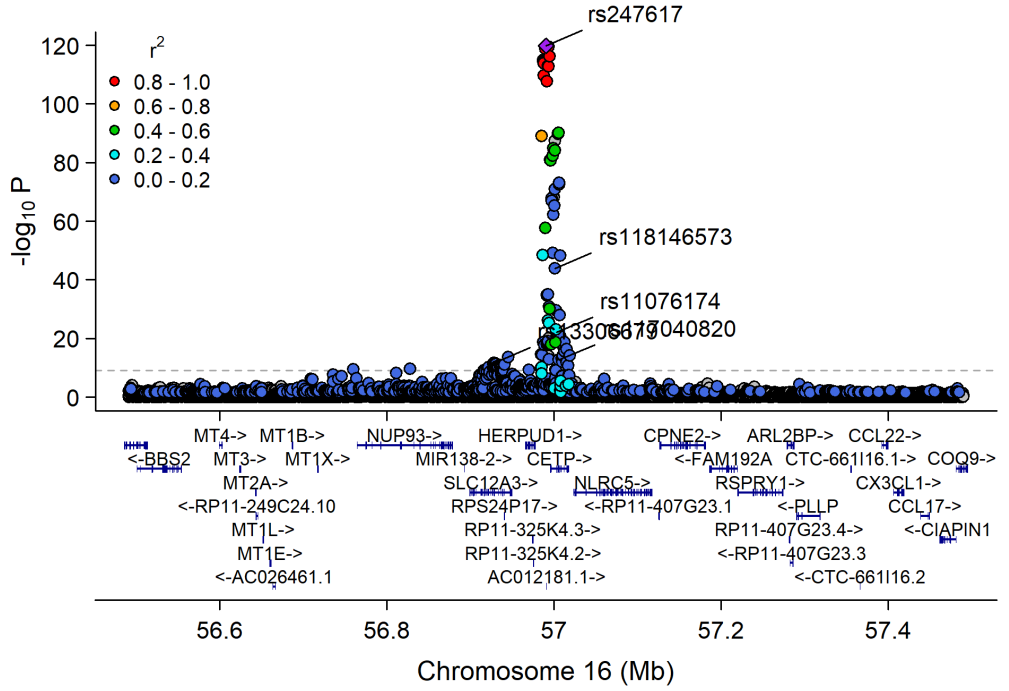
Chr11_115541901_117633315



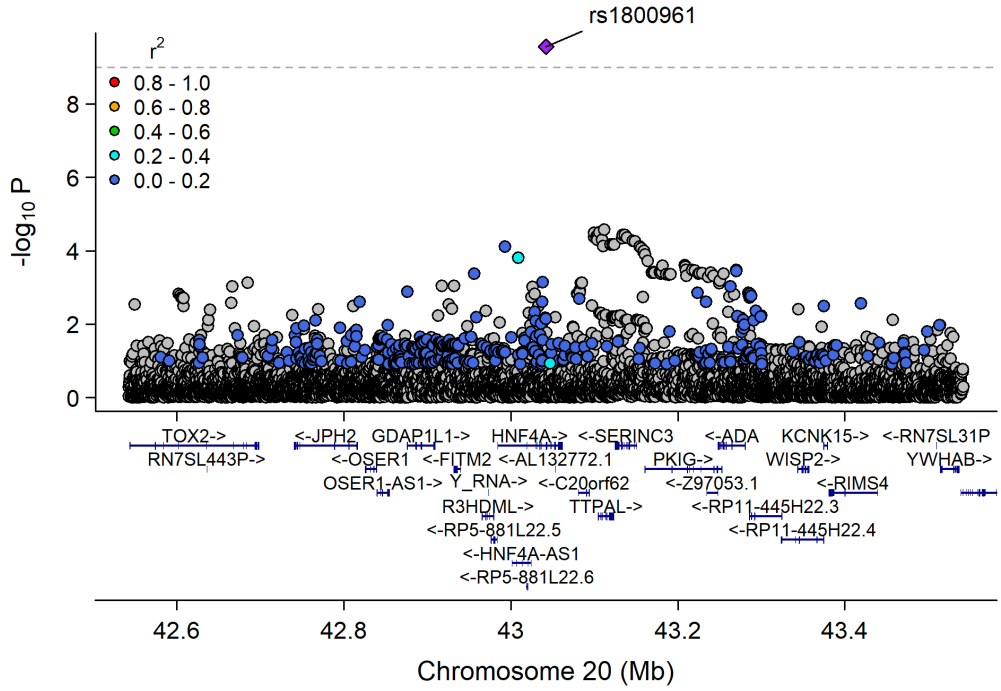
Chr15_57658798_60490883



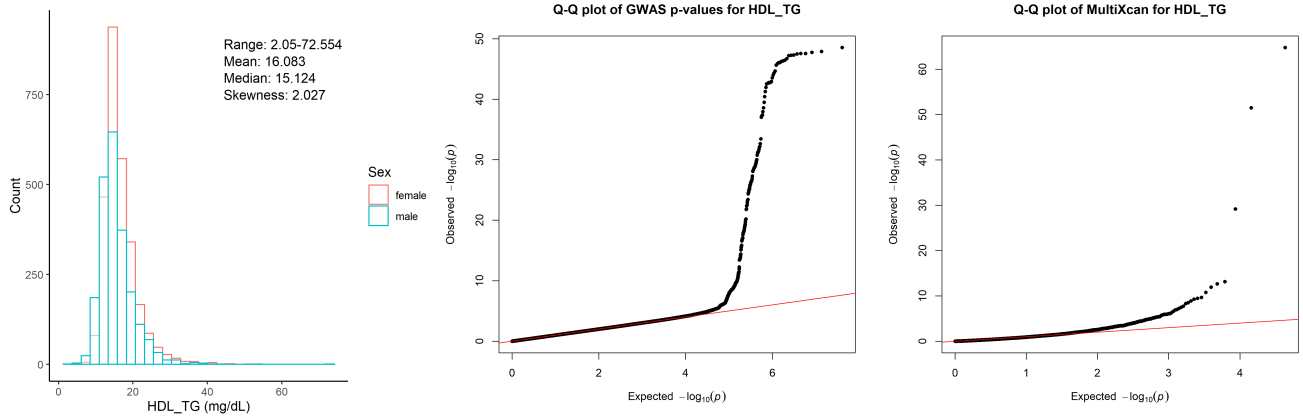
Chr16_55870822_57992421

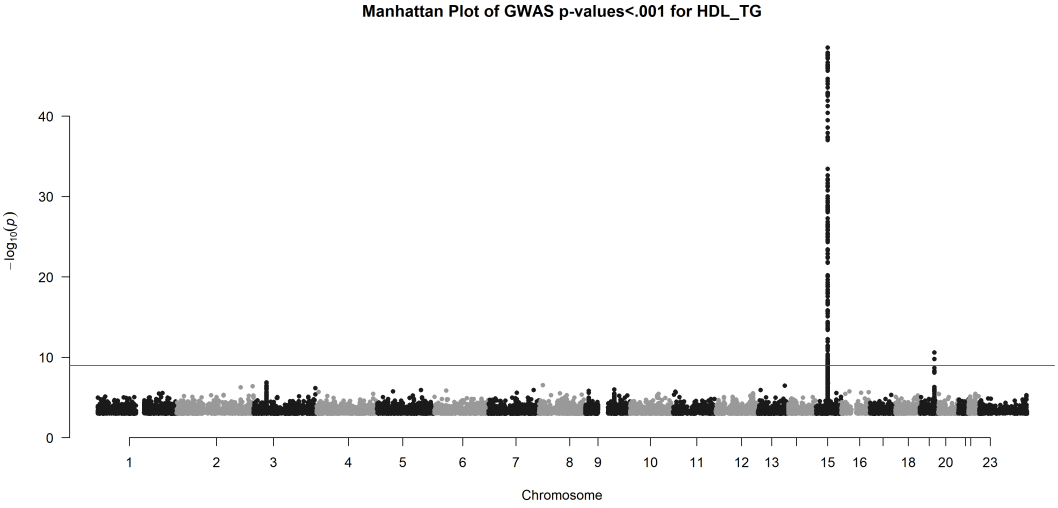


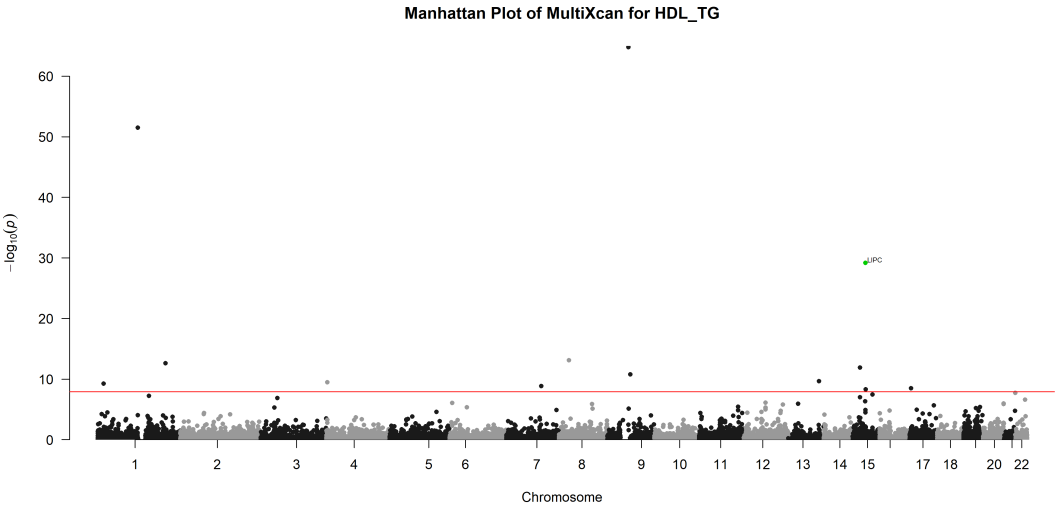
Chr20_42232254_45611445



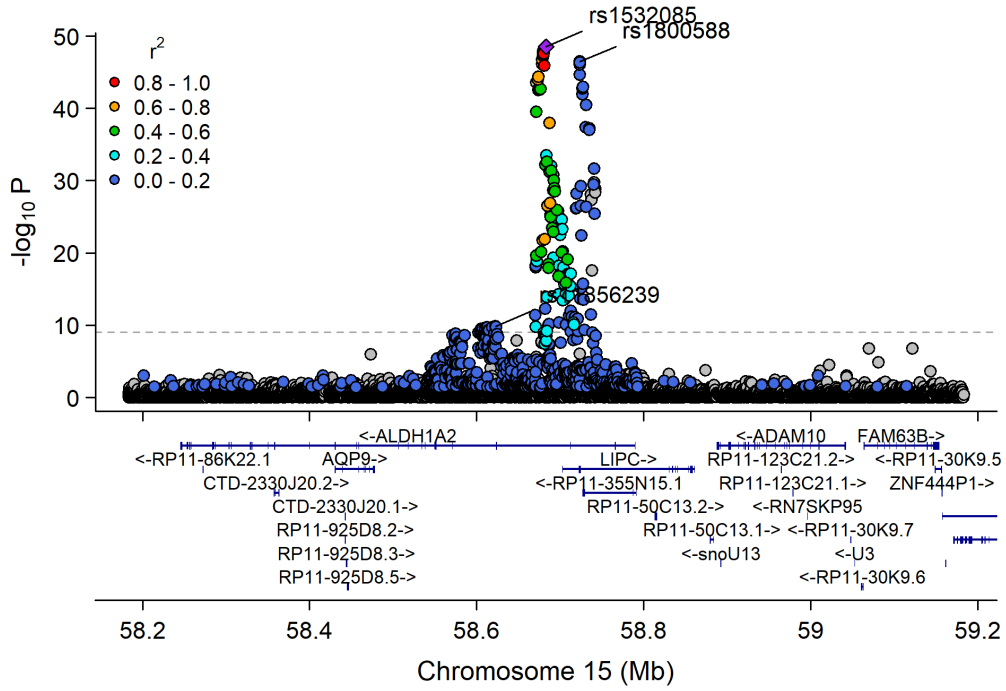
Triglycerides in HDL (mg/dL)



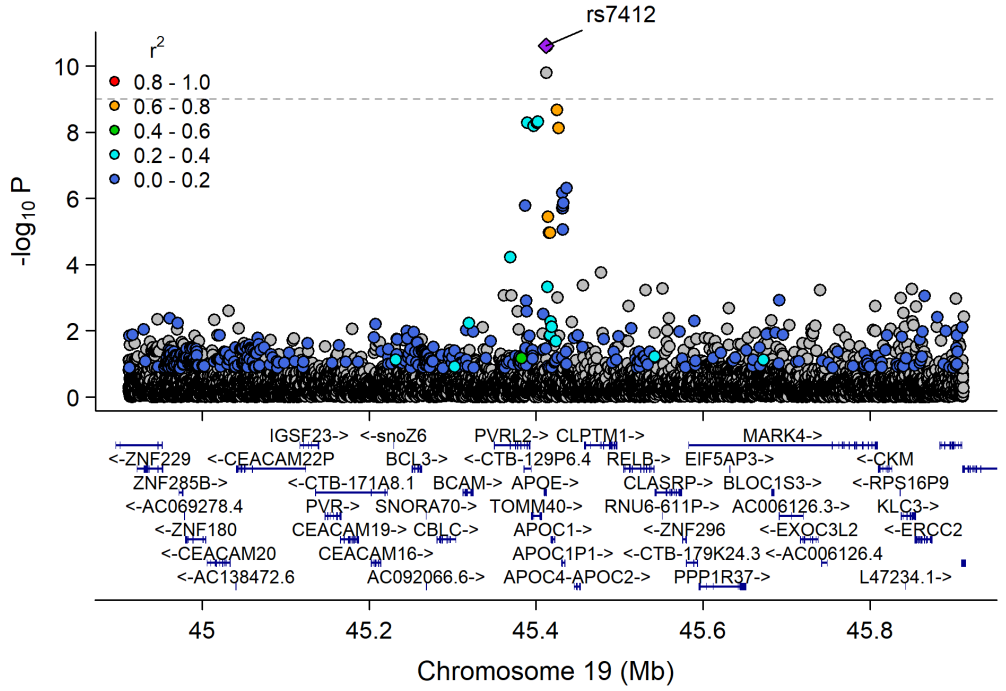




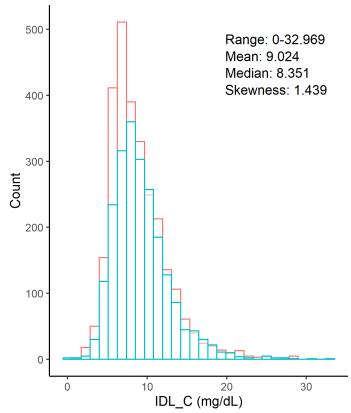
Chr15_57658798_60490883

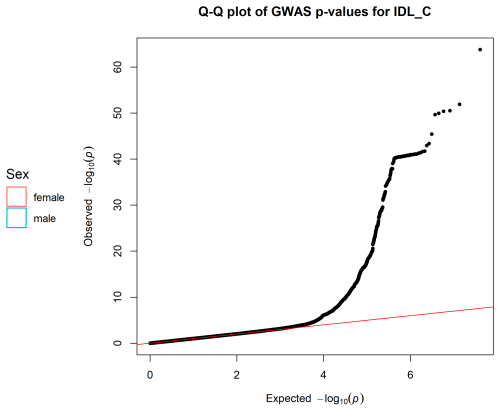


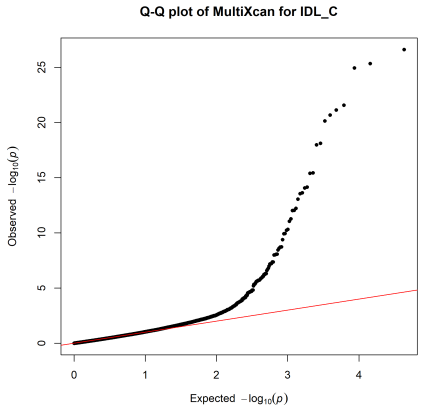
Chr19_44063363_46637375

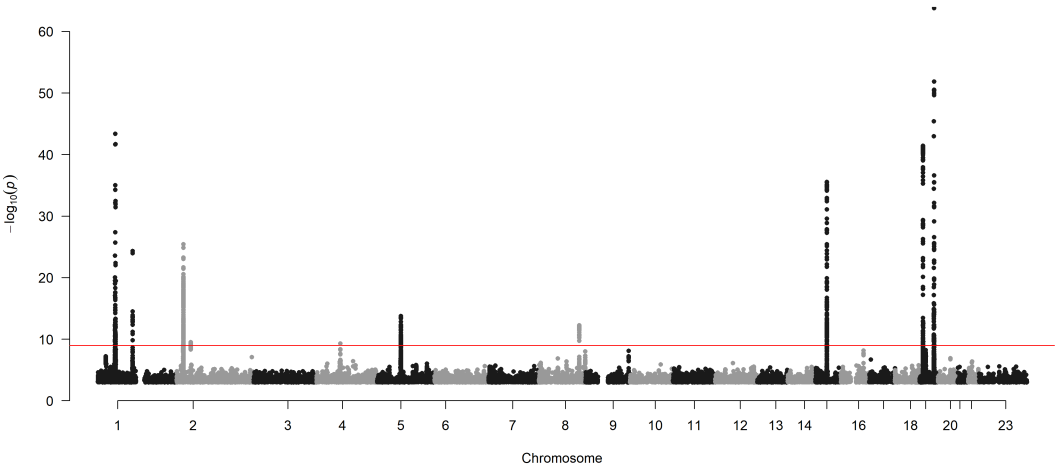


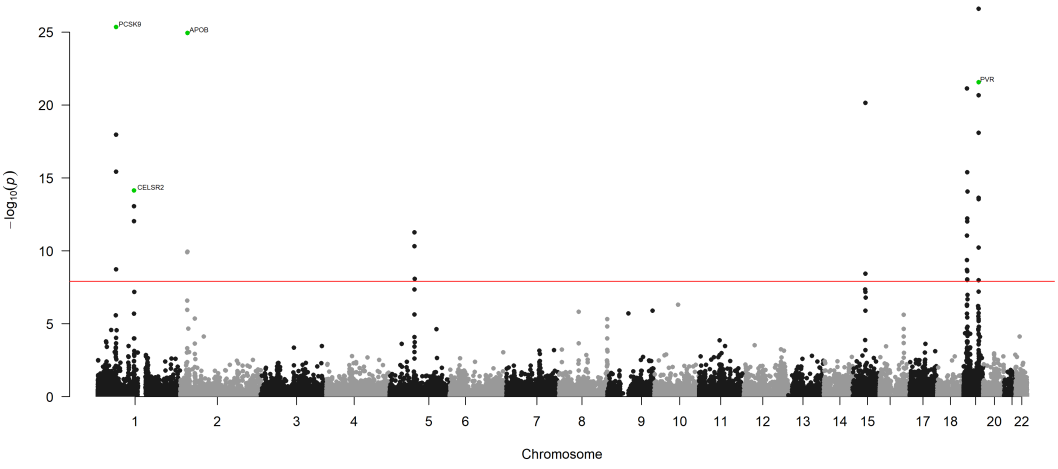
Total cholesterol in IDL (mg/dL)



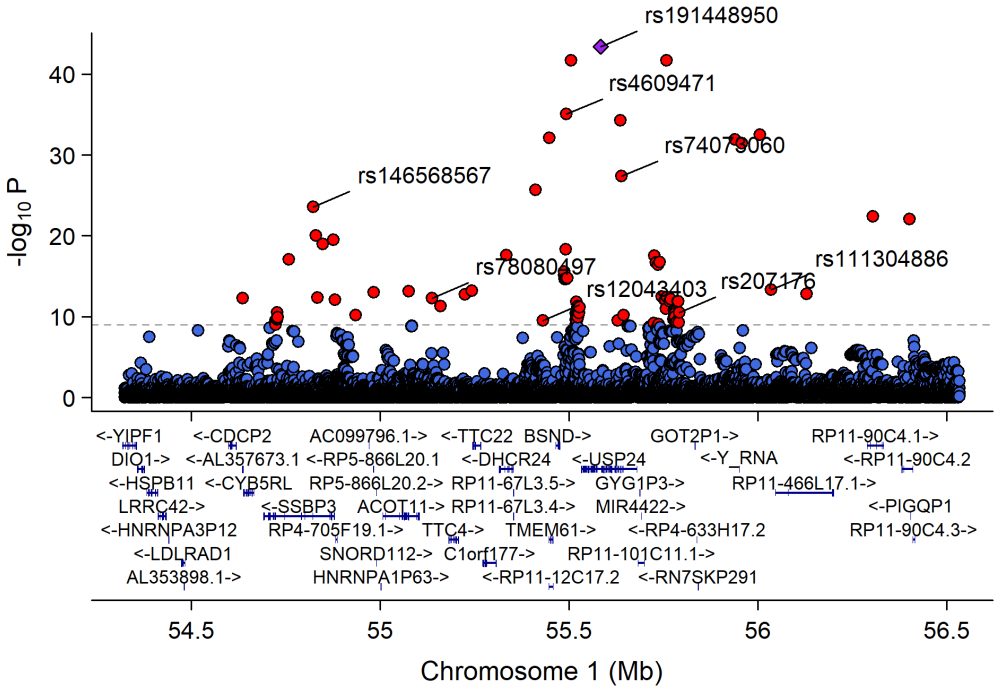




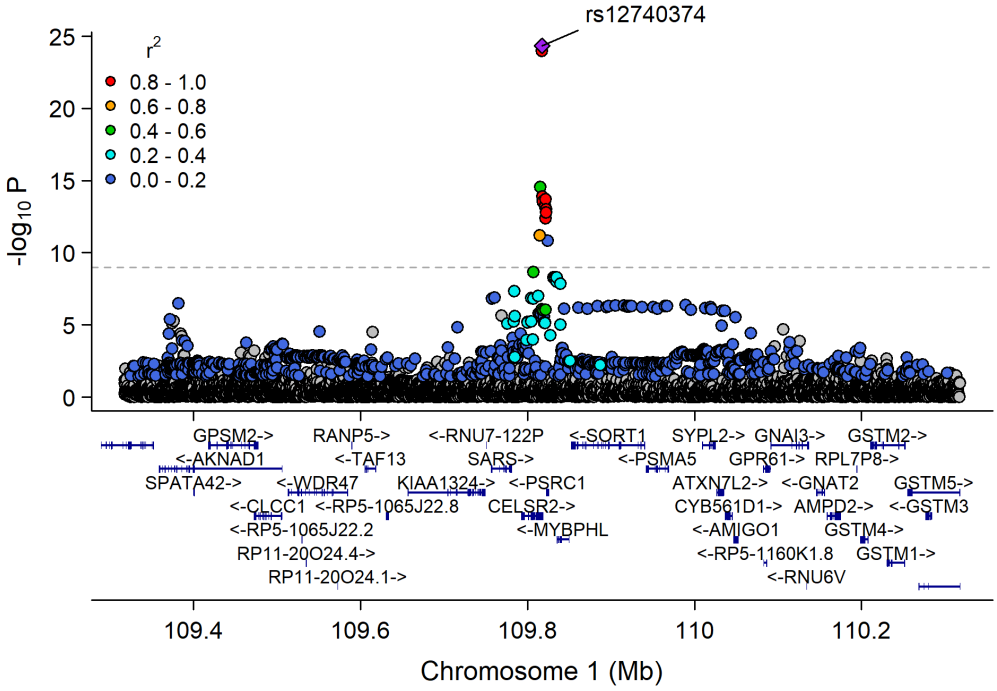




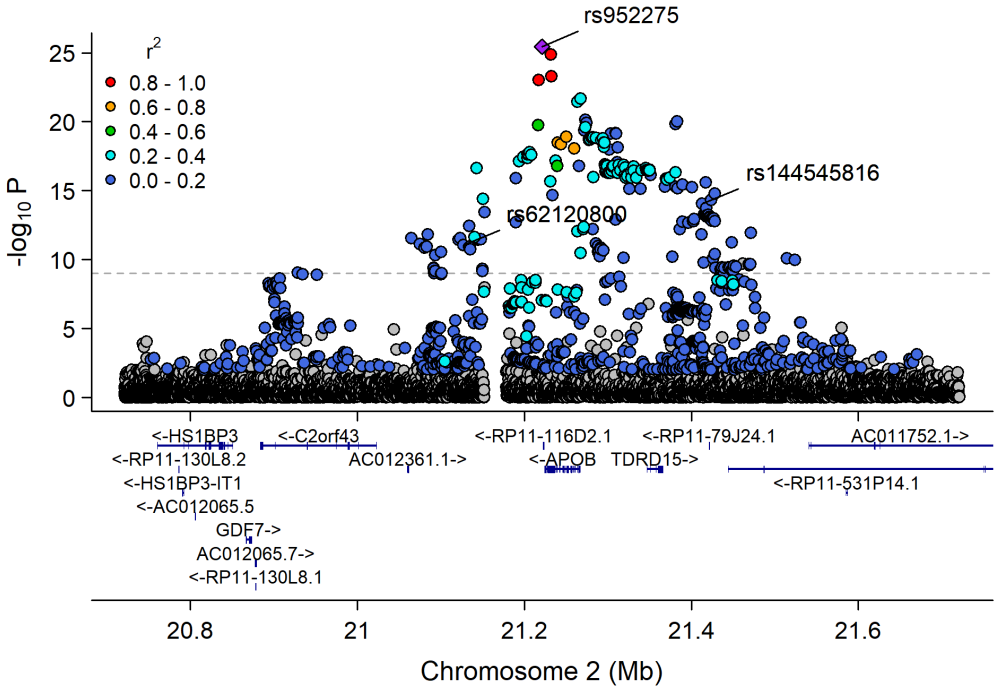
Chr1__53272879__57300396



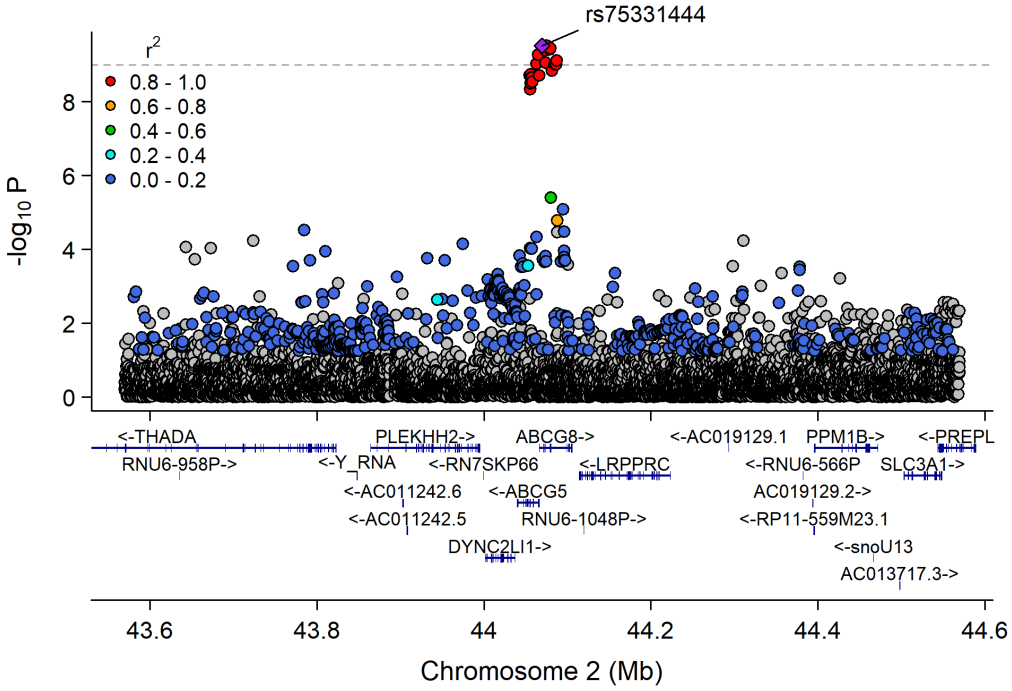
Chr1__109696238__110162190



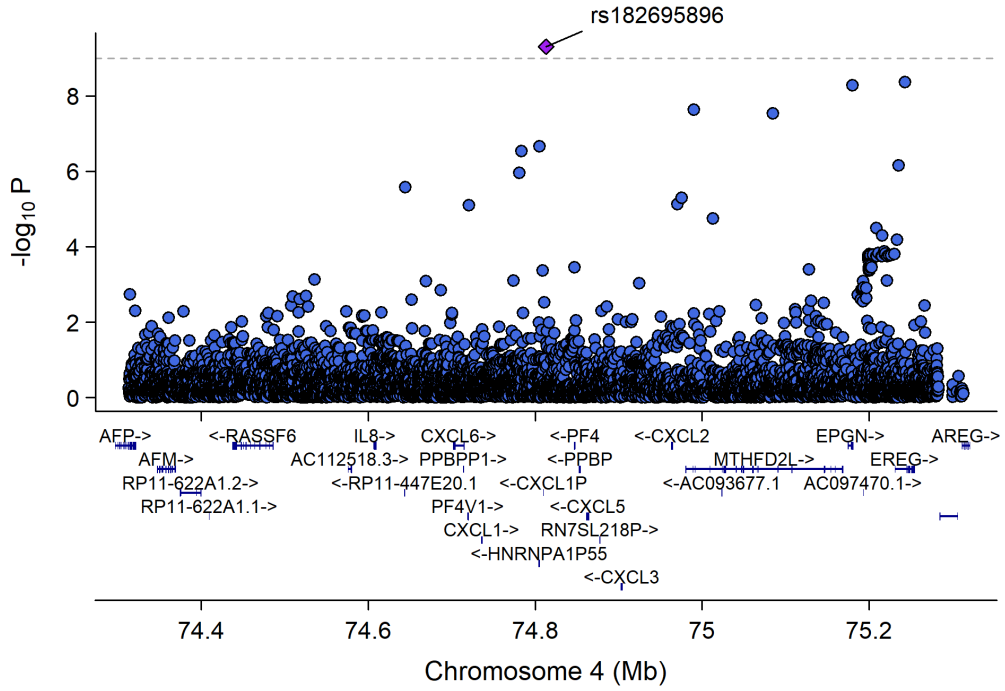
Chr2_19947287_22427492



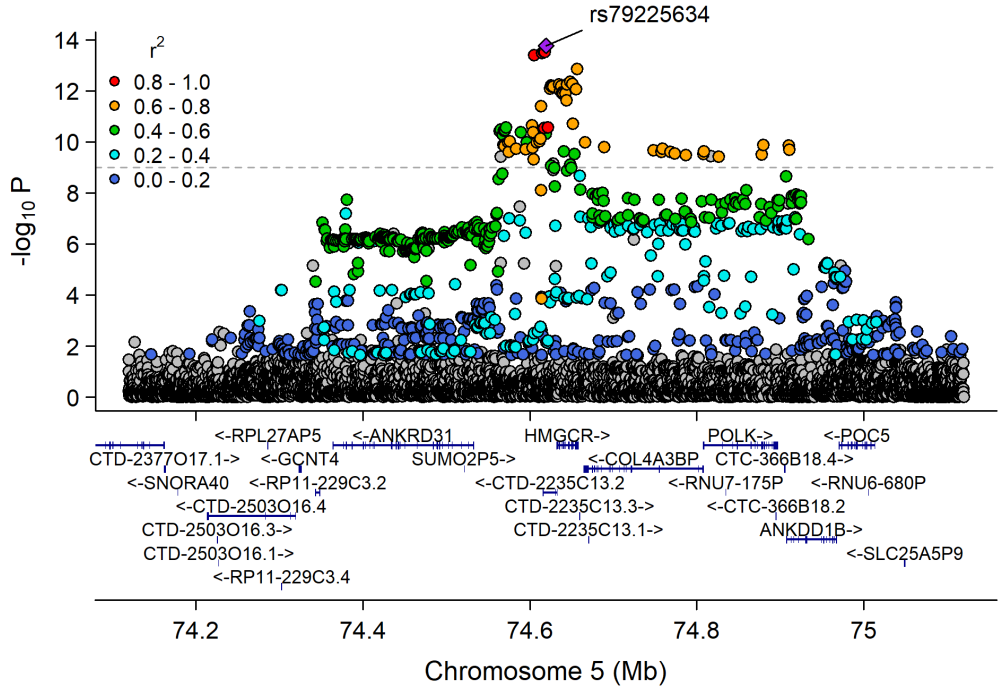
Chr2_43065153_45056024



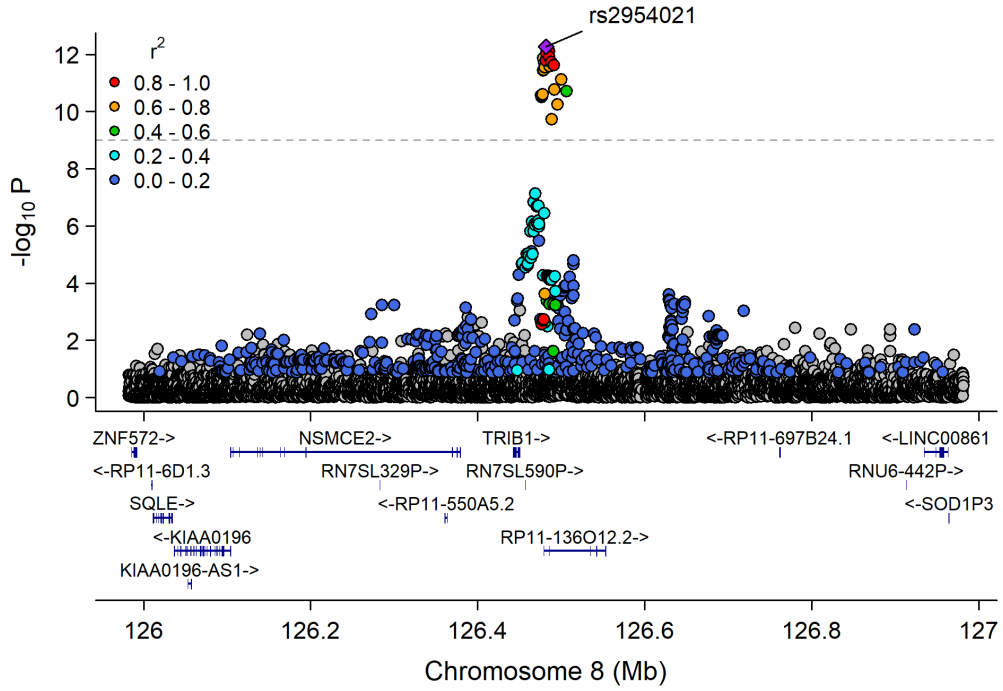
Chr4_71881824_76681229



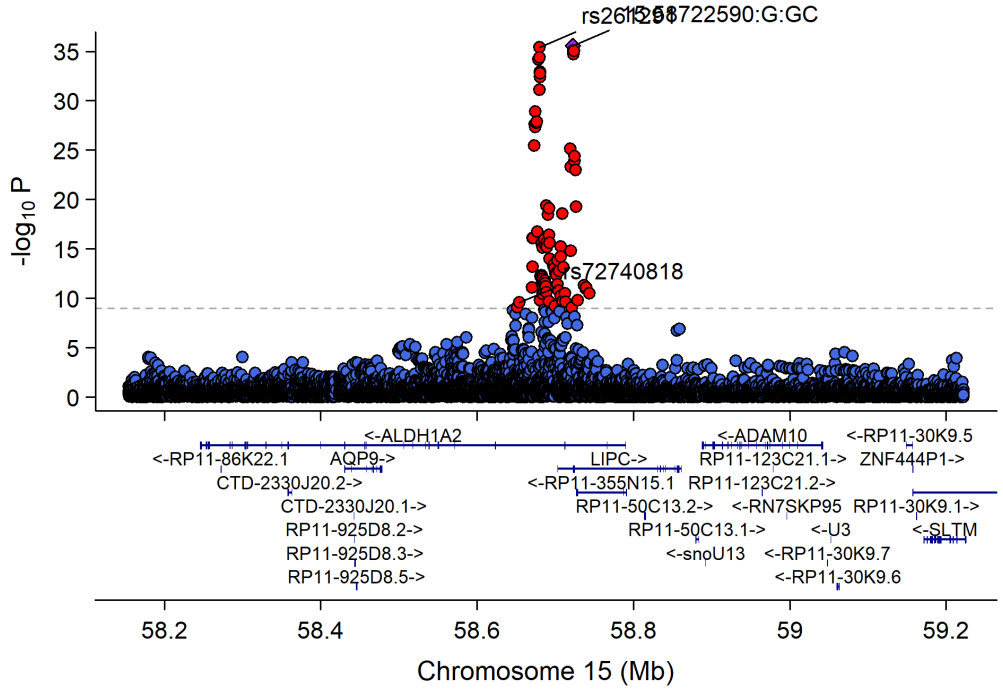
Chr5_74242002_75221582



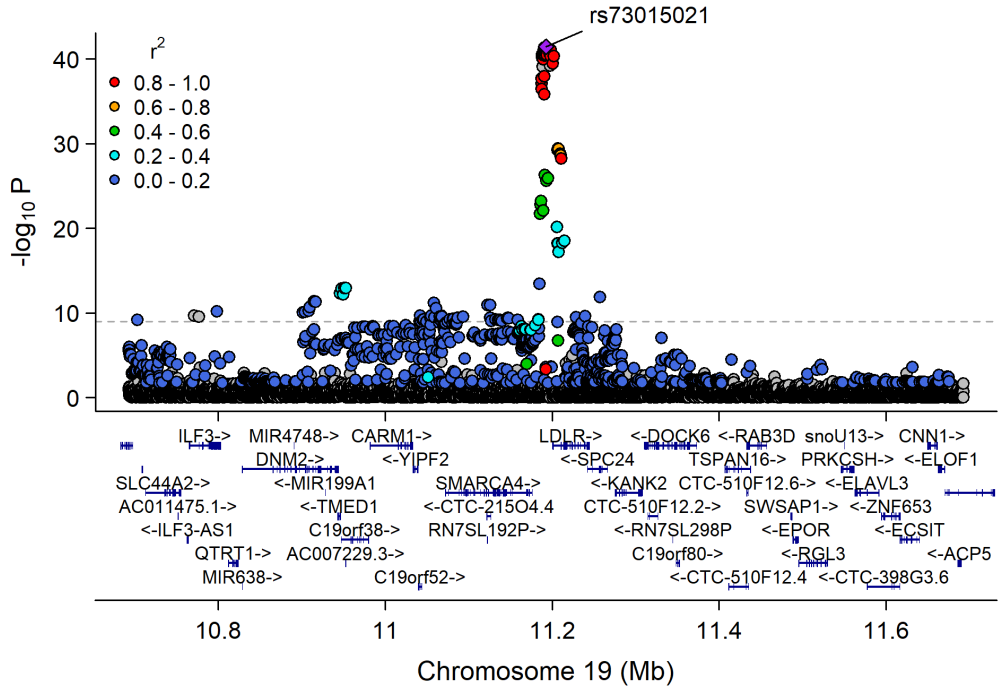
Chr8__126435663__126578061



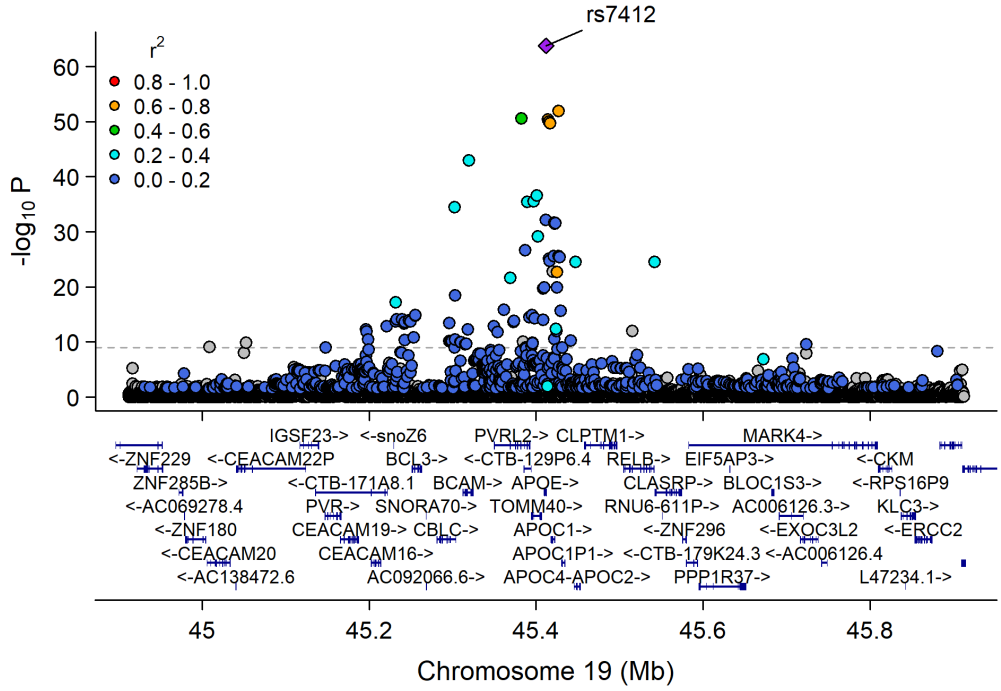
Chr15__57658798__60490883



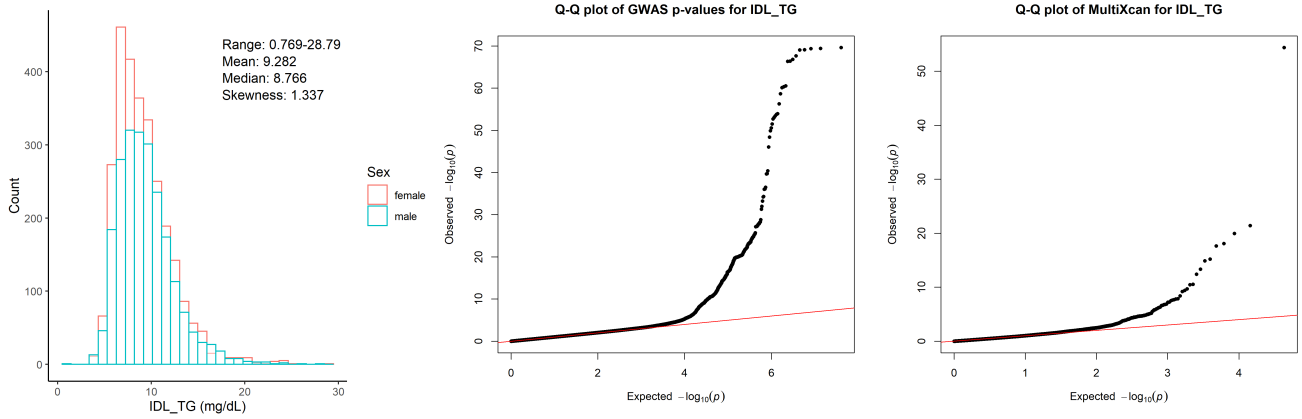
Chr19_9807999_12255594

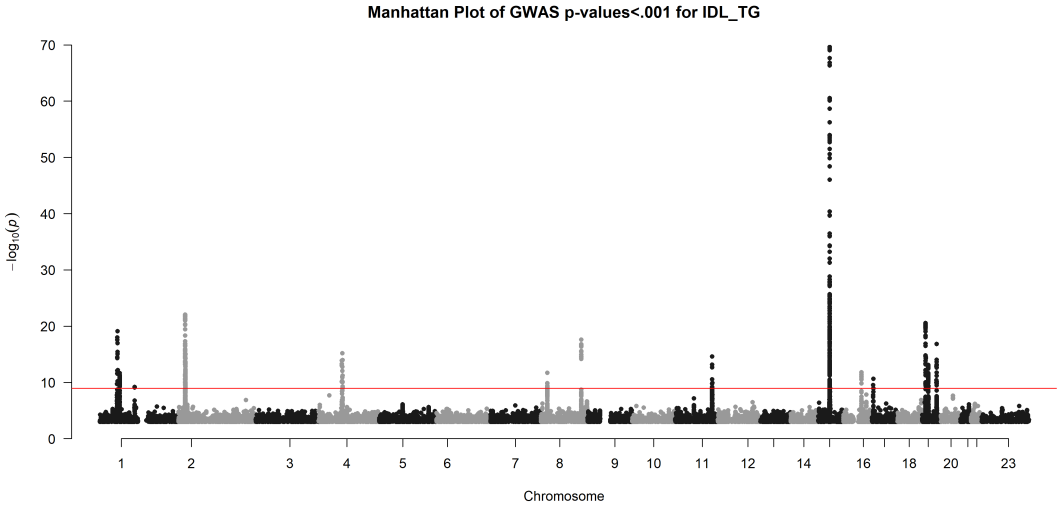


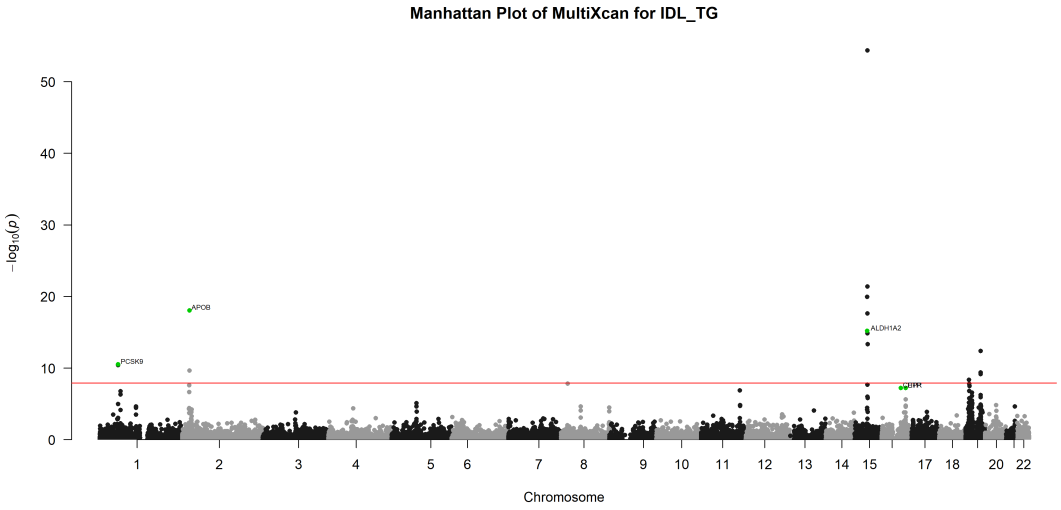
Chr19_44063363_46637375



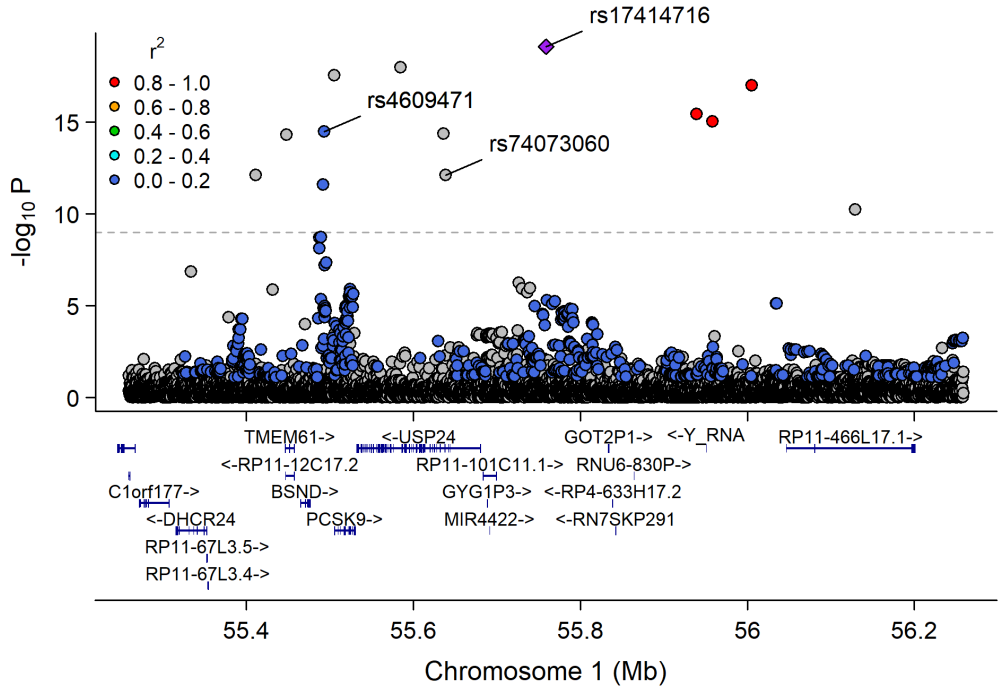
Triglycerides in IDL (mg/dL)



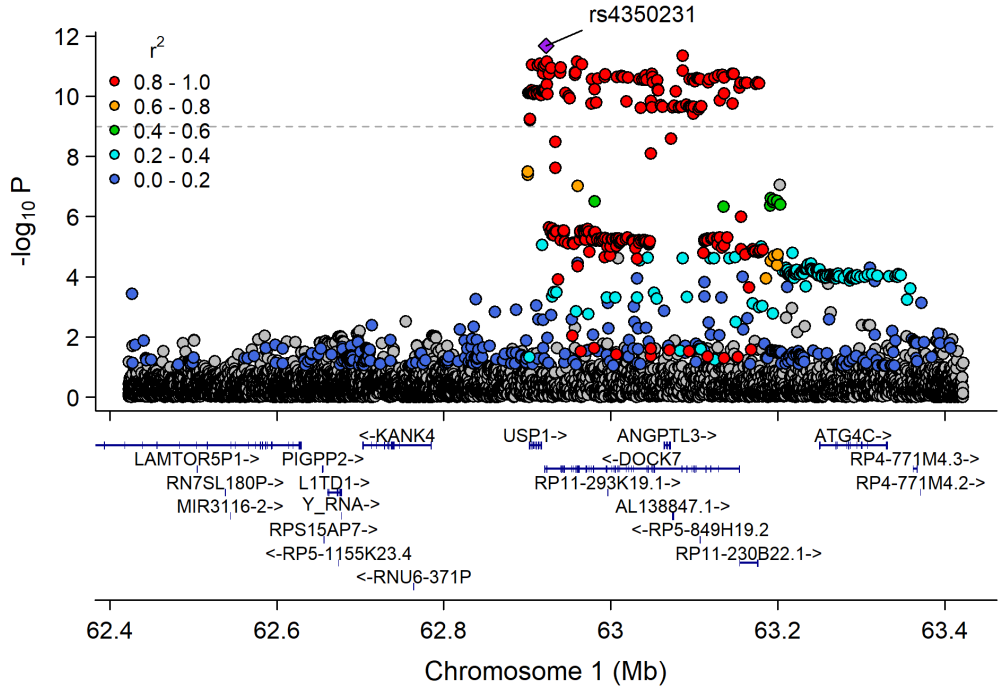




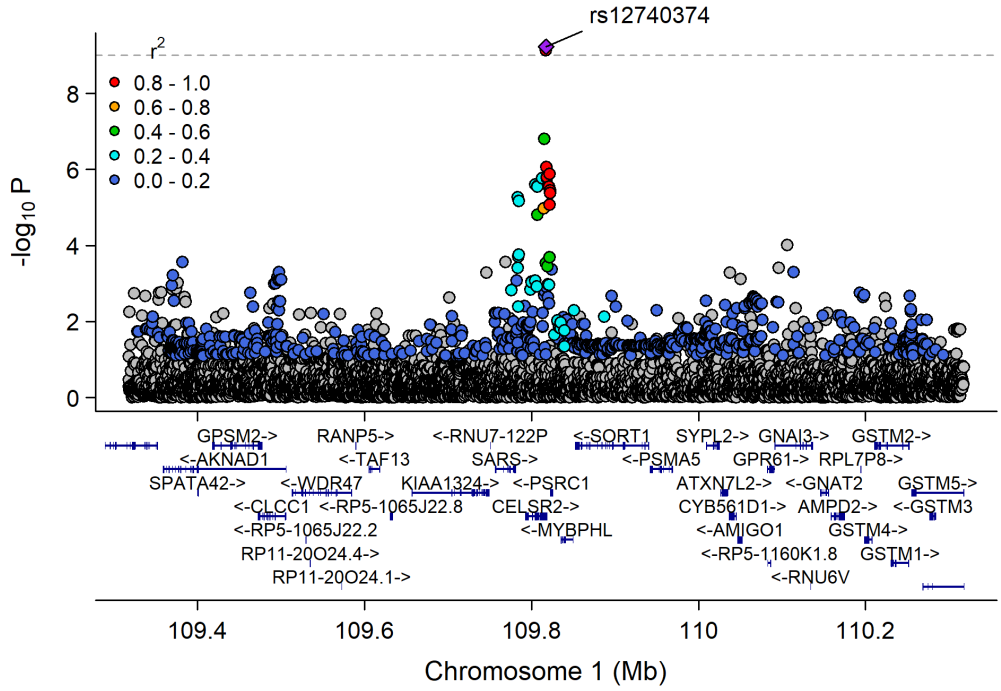
Chr1_53272879_57300396



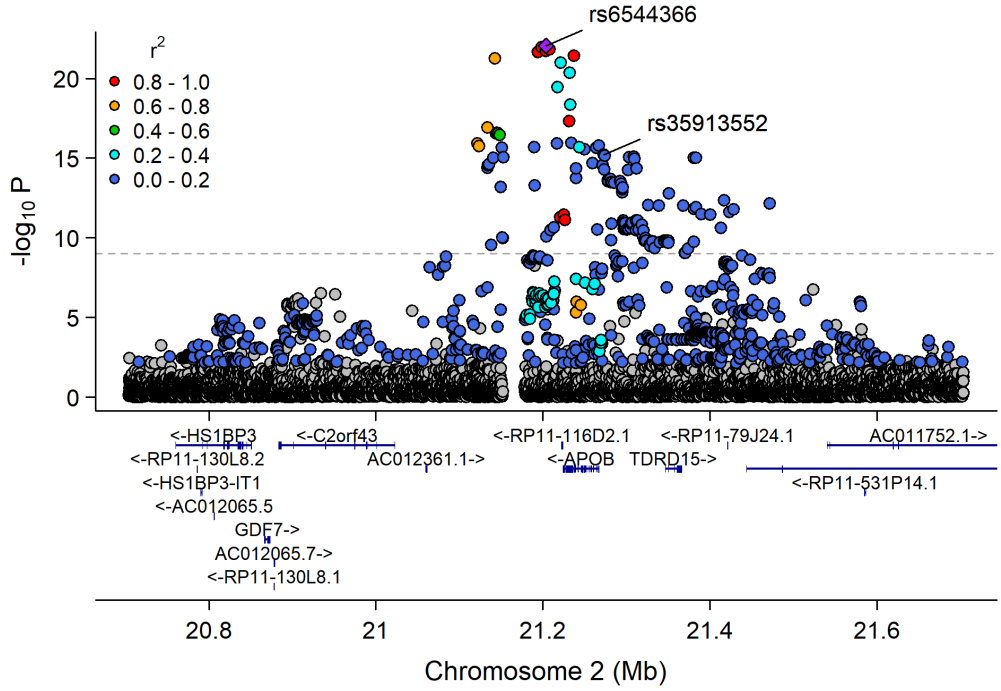
Chr1_62833402_63375116



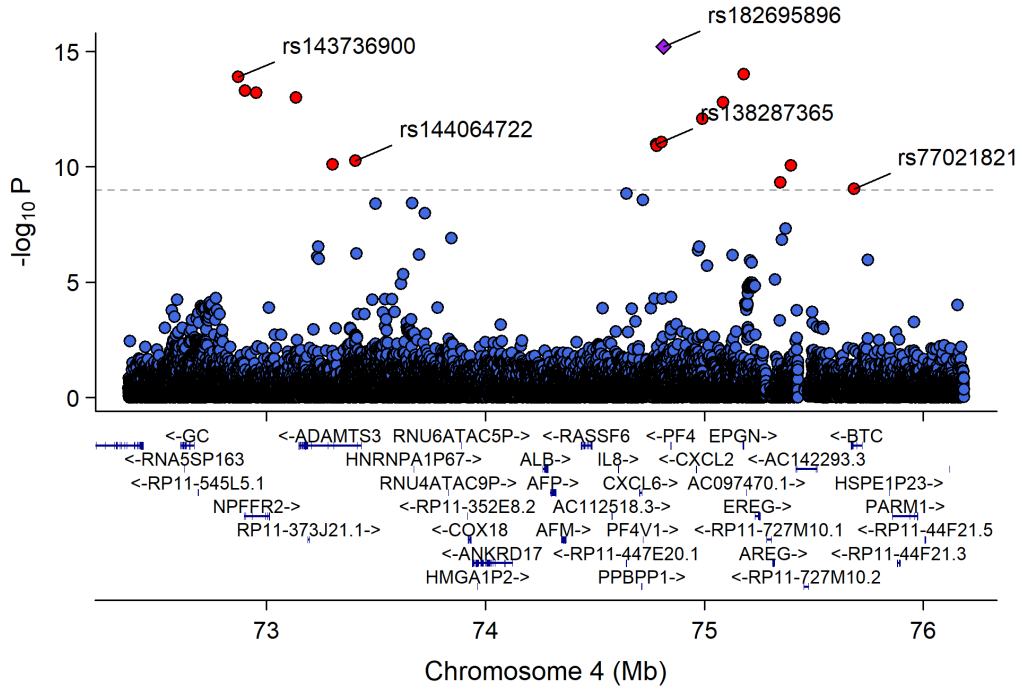
Chr1_109696238_110162190



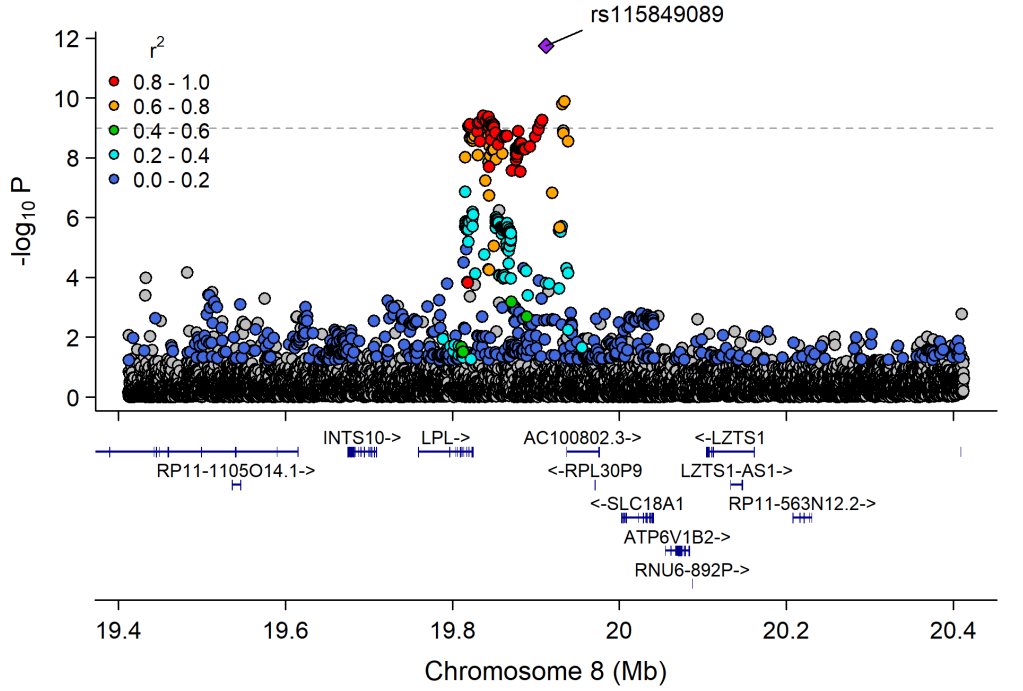
Chr2_19947287_22427492



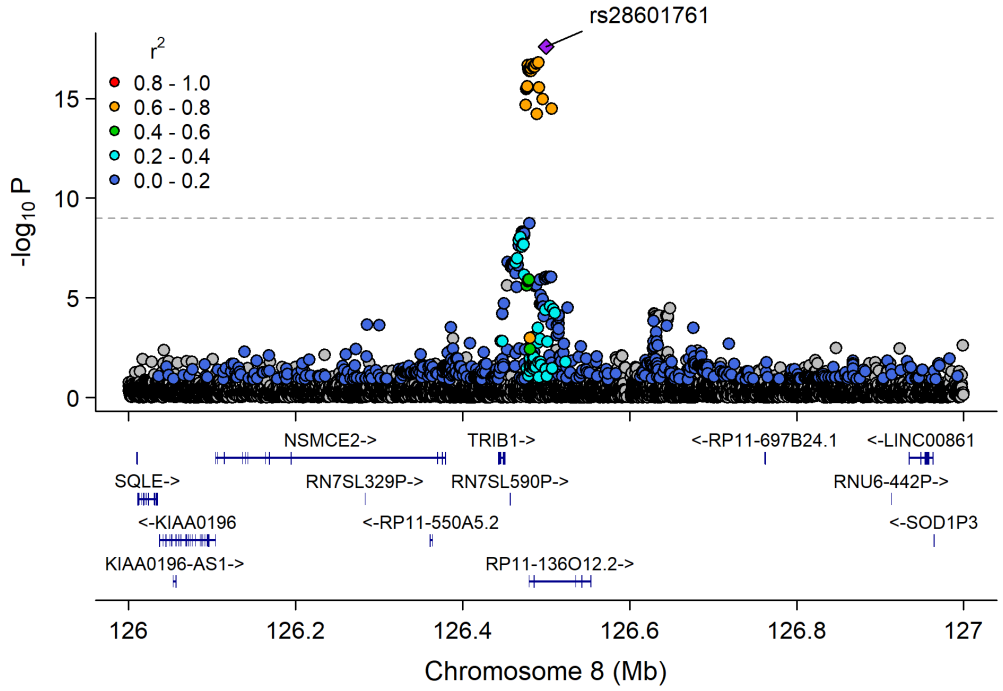
Chr4_71881824_76681229



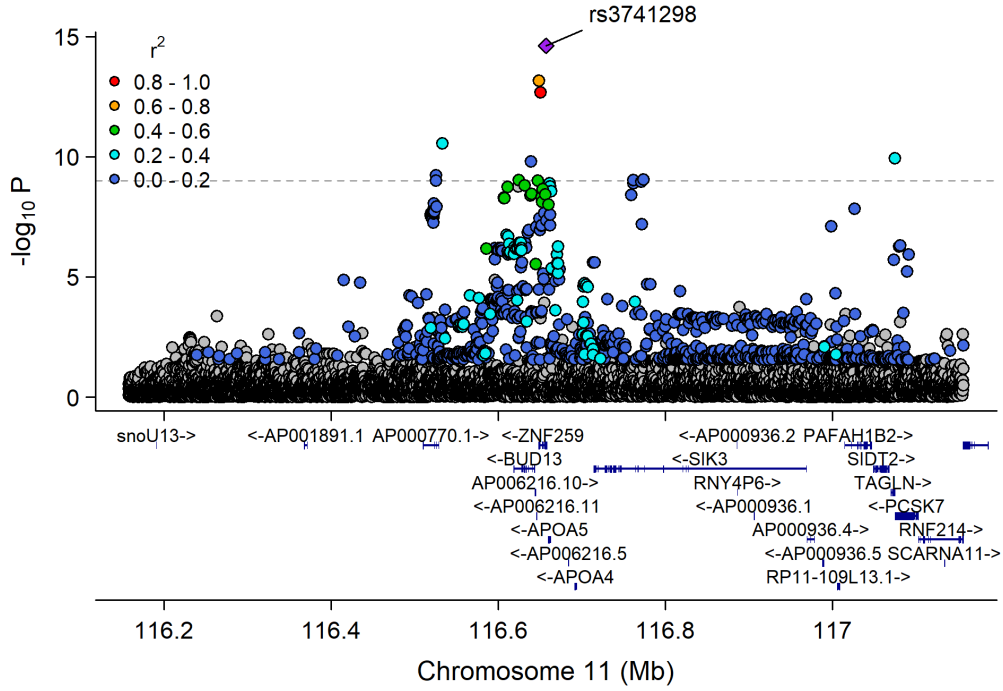
Chr8_18582620_20792744



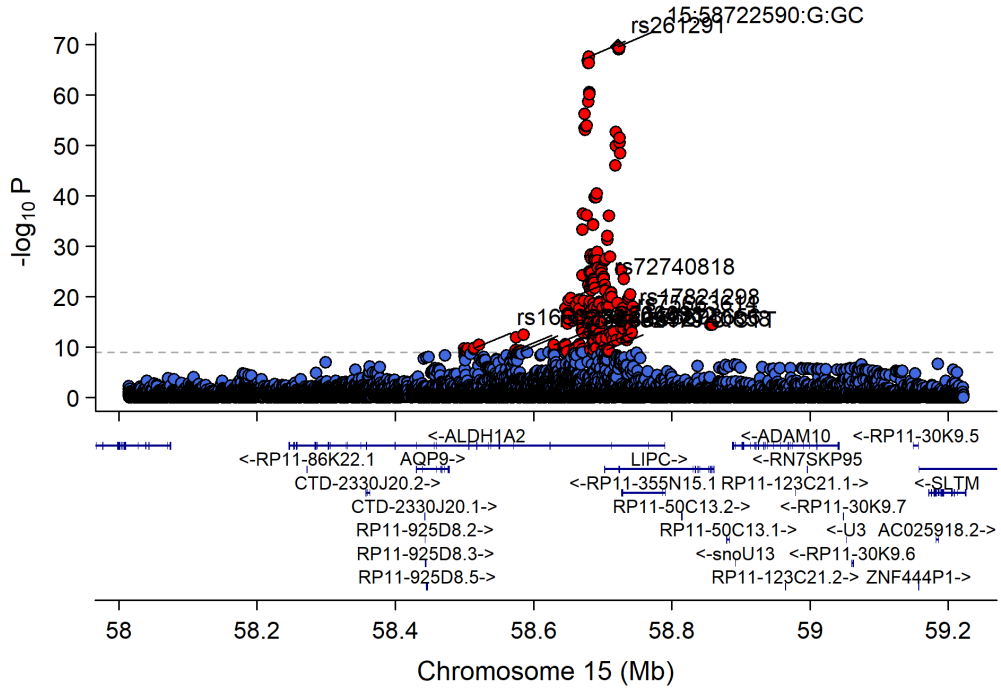
Chr8_126435663_126578061



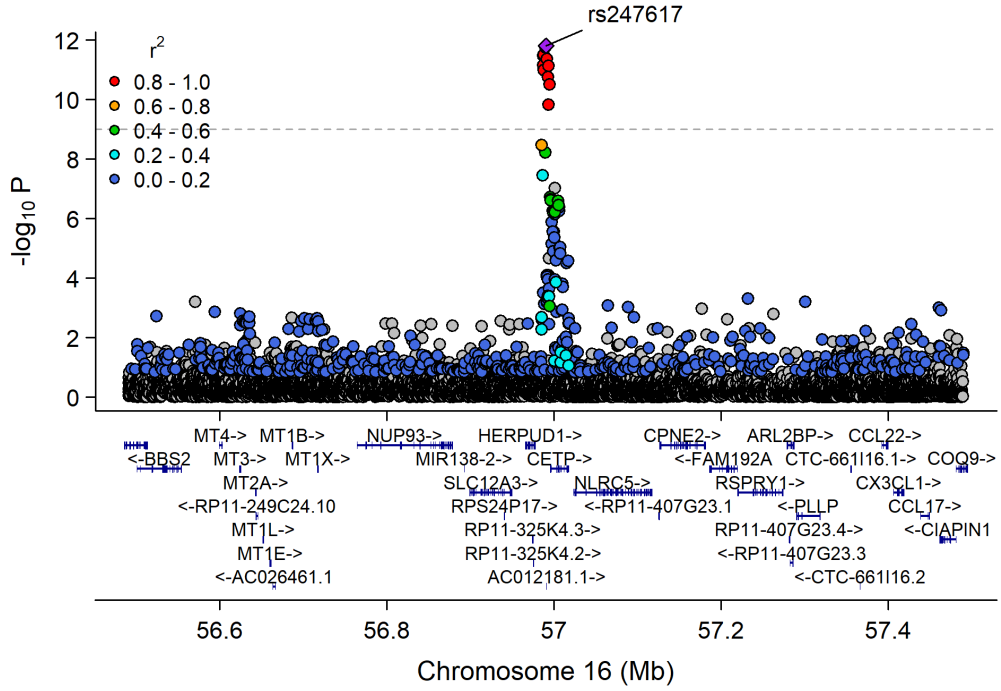
Chr11_115541901_117633315



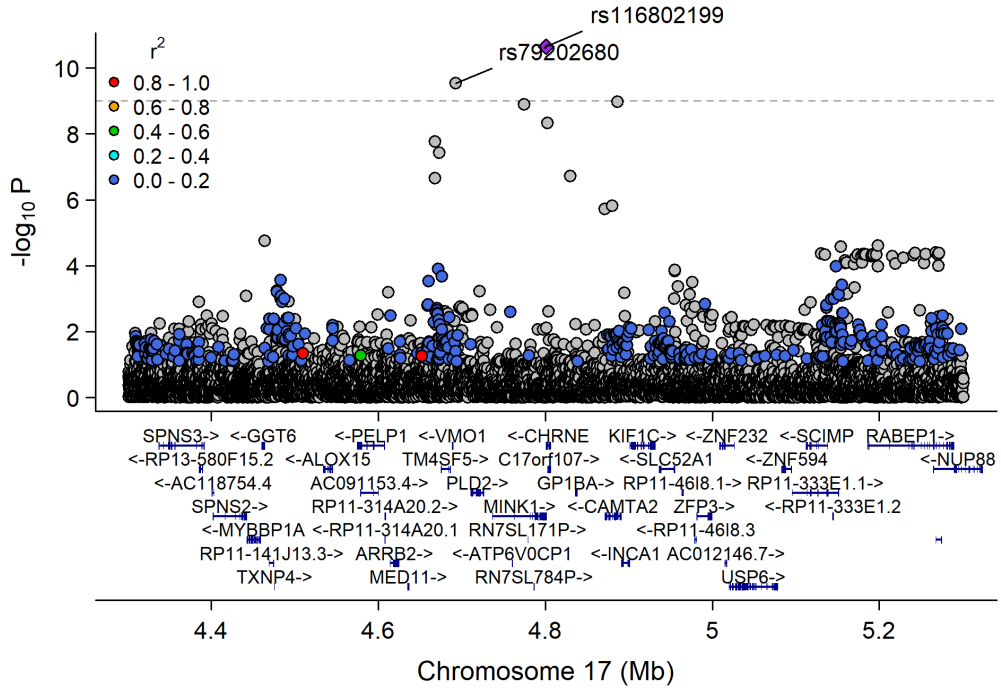
Chr15_57658798_60490883



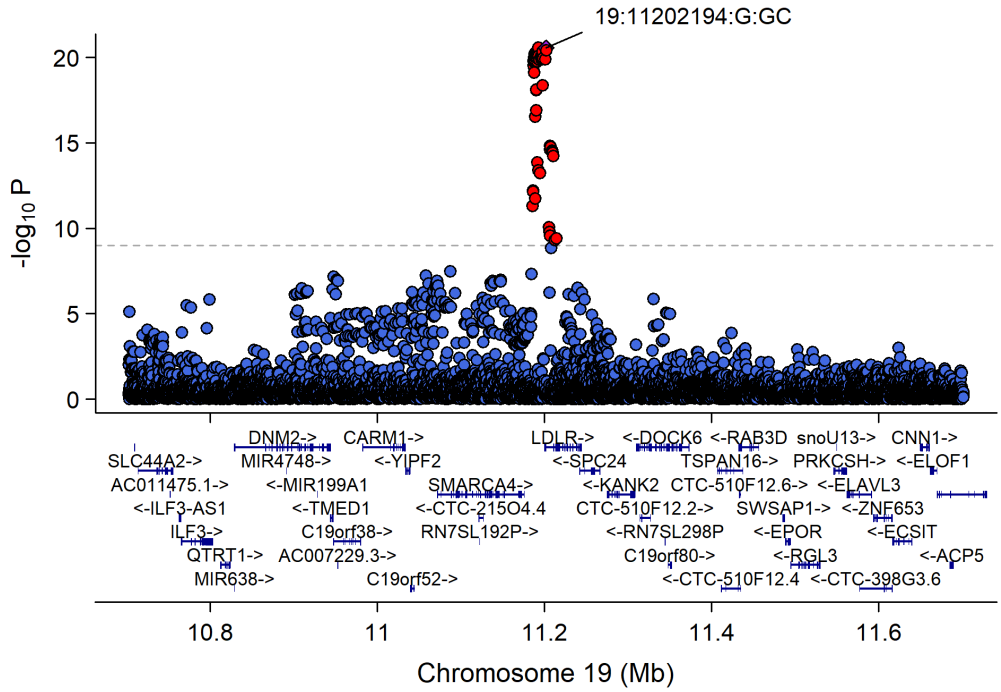
Chr16_55870822_57992421



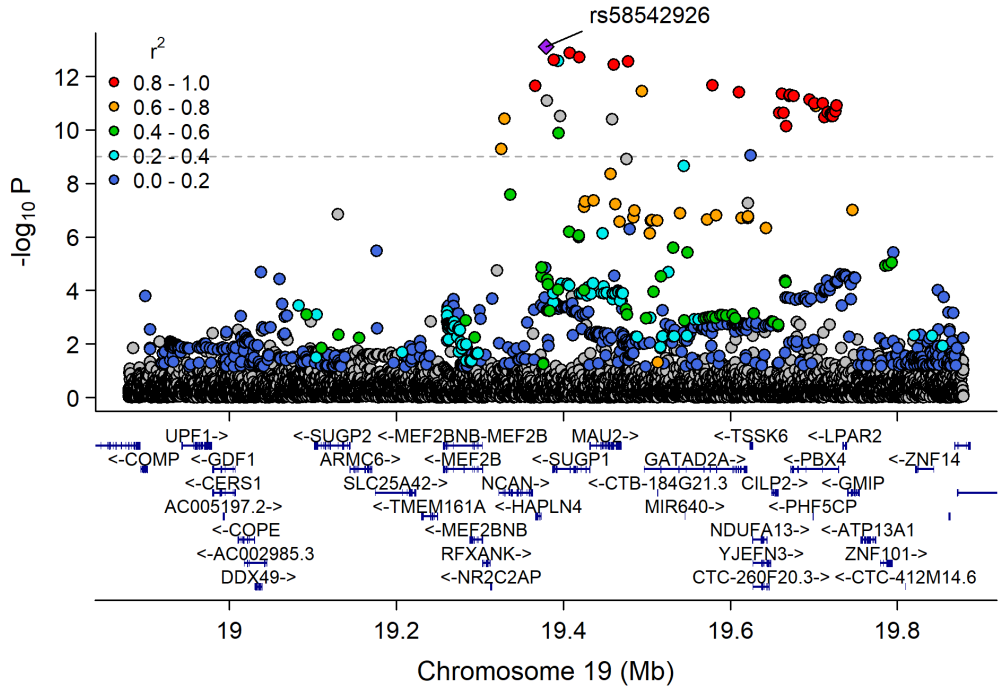
Chr17_3704190_5793348



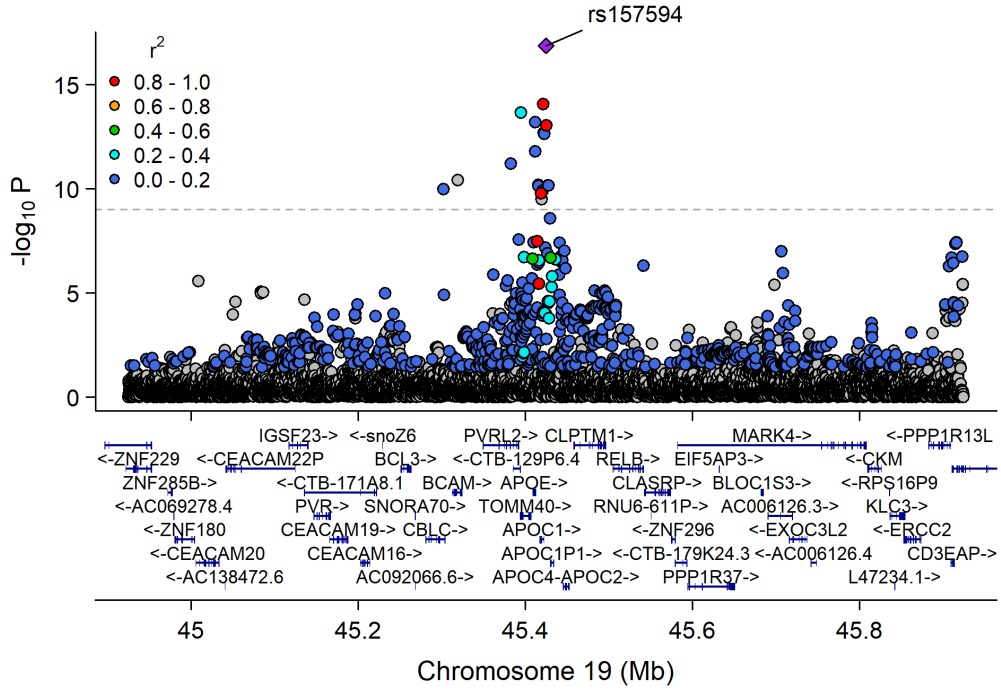
Chr19_9807999_12255594



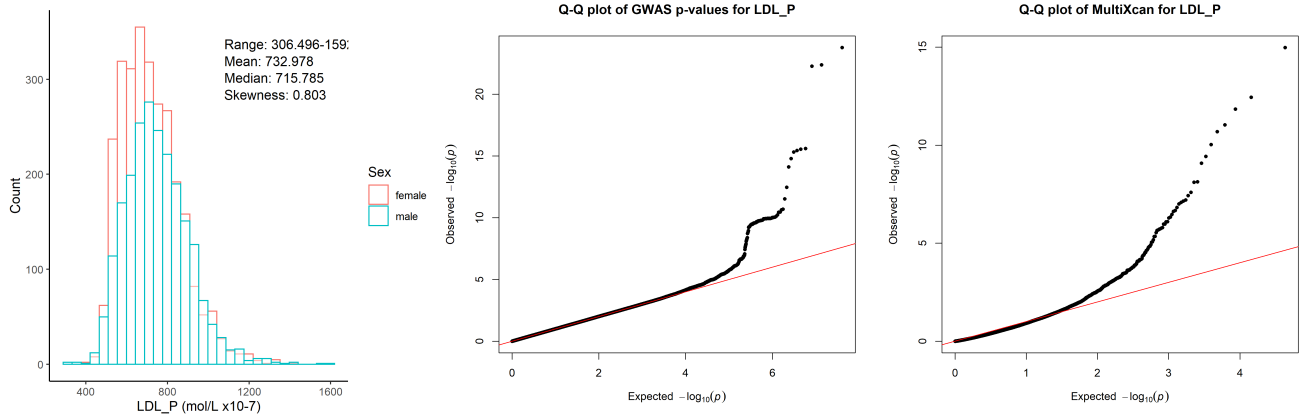
Chr19_18370495_20841464

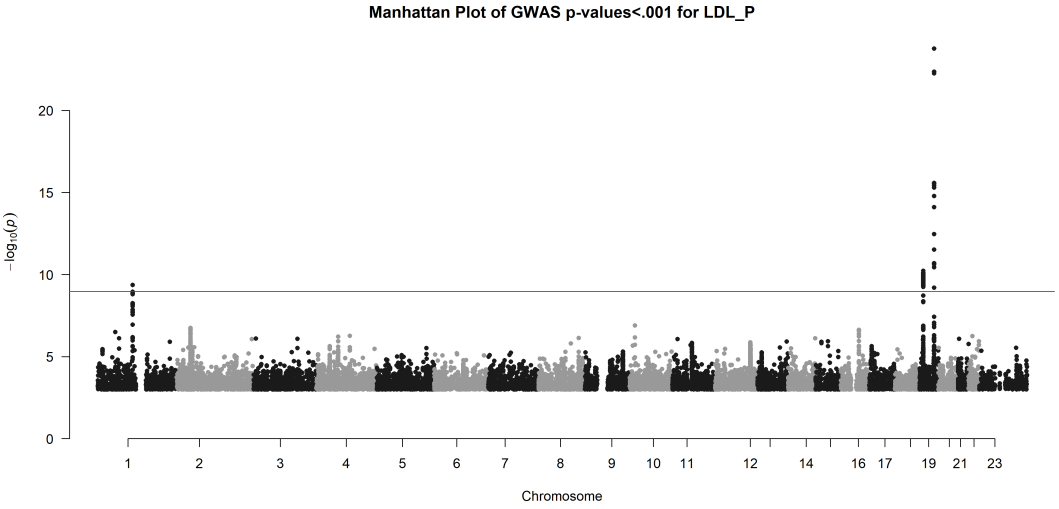


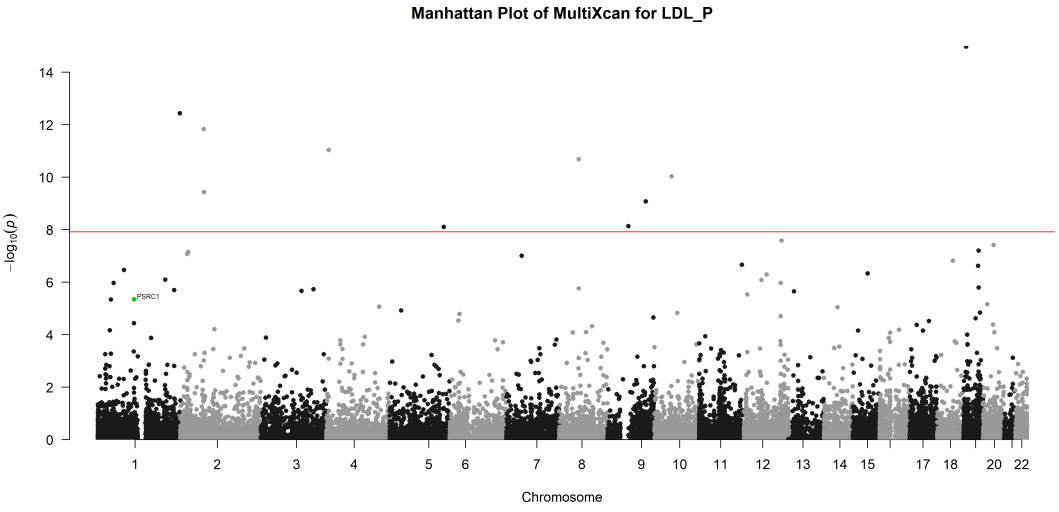
Chr19_44063363_46637375



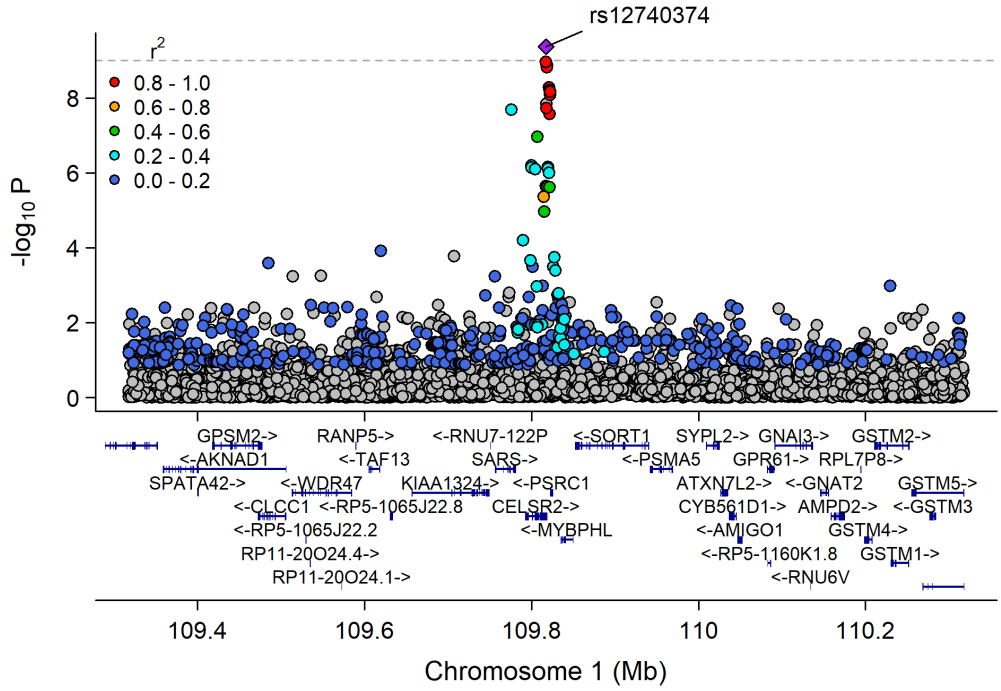
Concentration of LDL particles (mol/l)



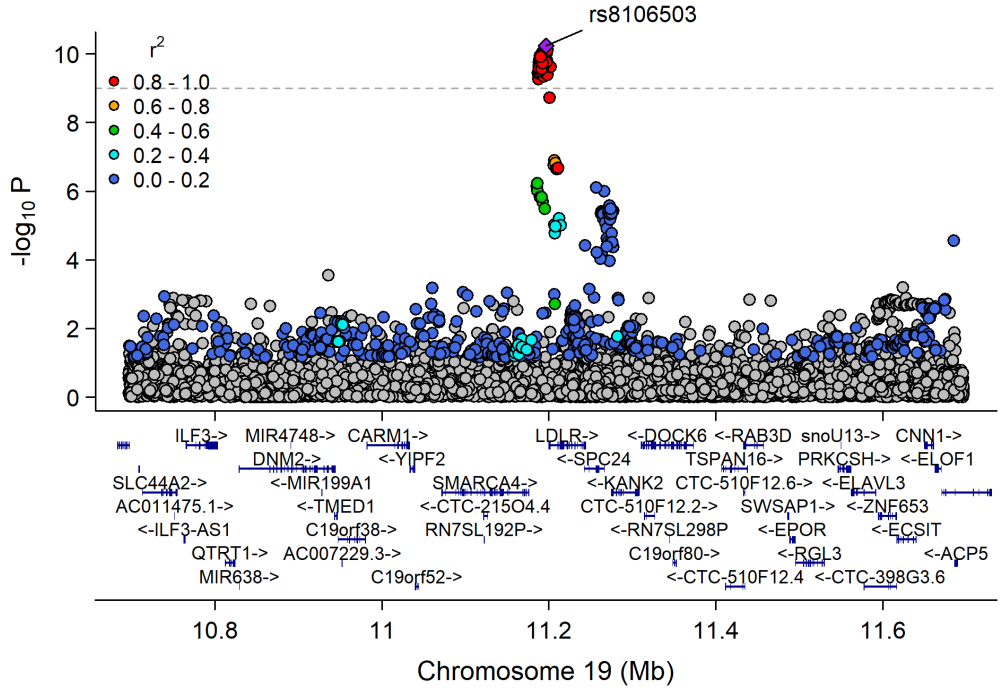




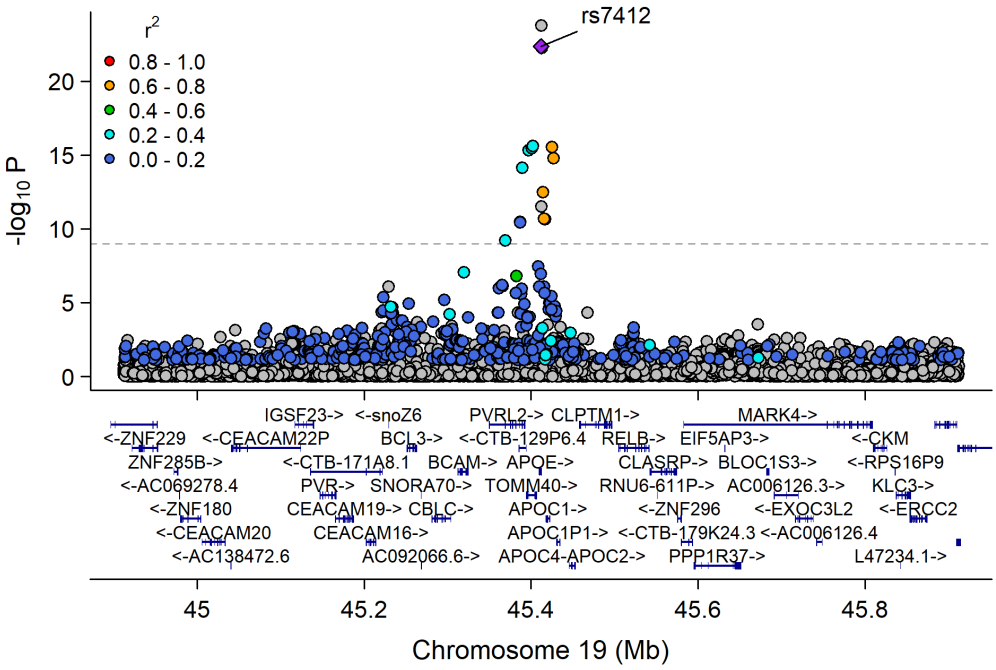
Chr1__109696238__110162190



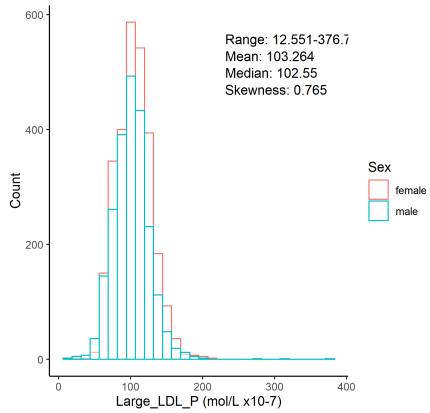
Chr19__9807999__12255594

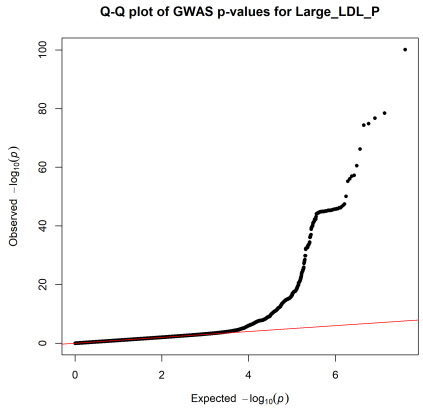


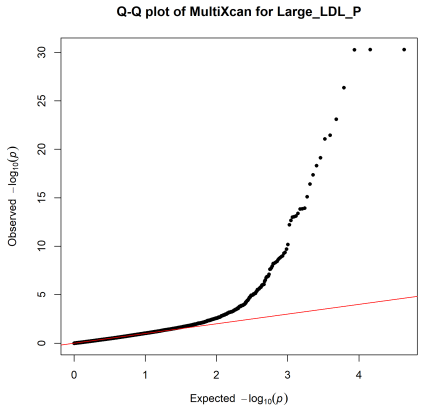
Chr19__44063363__46637375

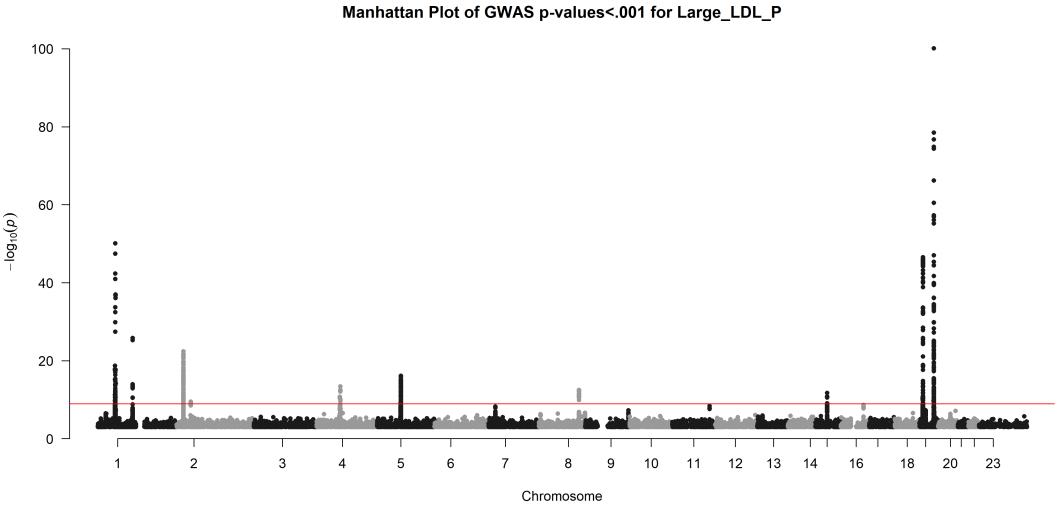


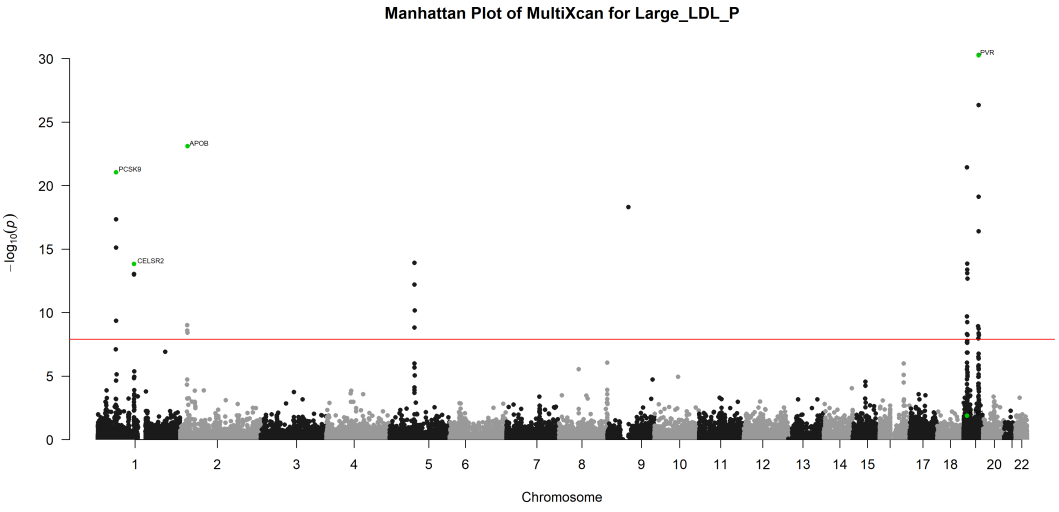
Concentration of large LDL particles (mol/l)



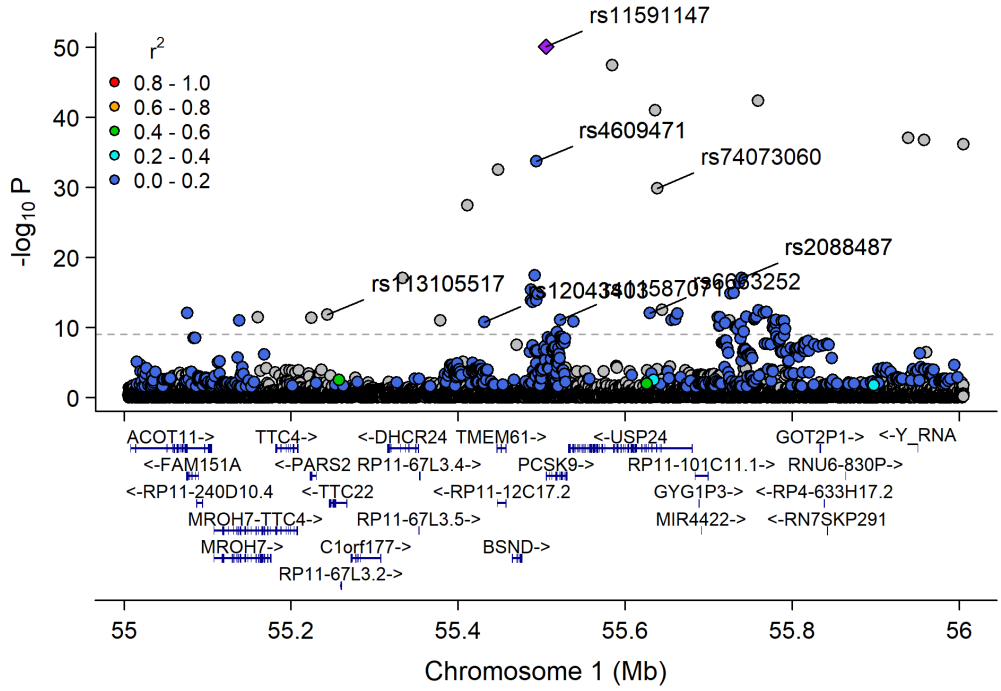




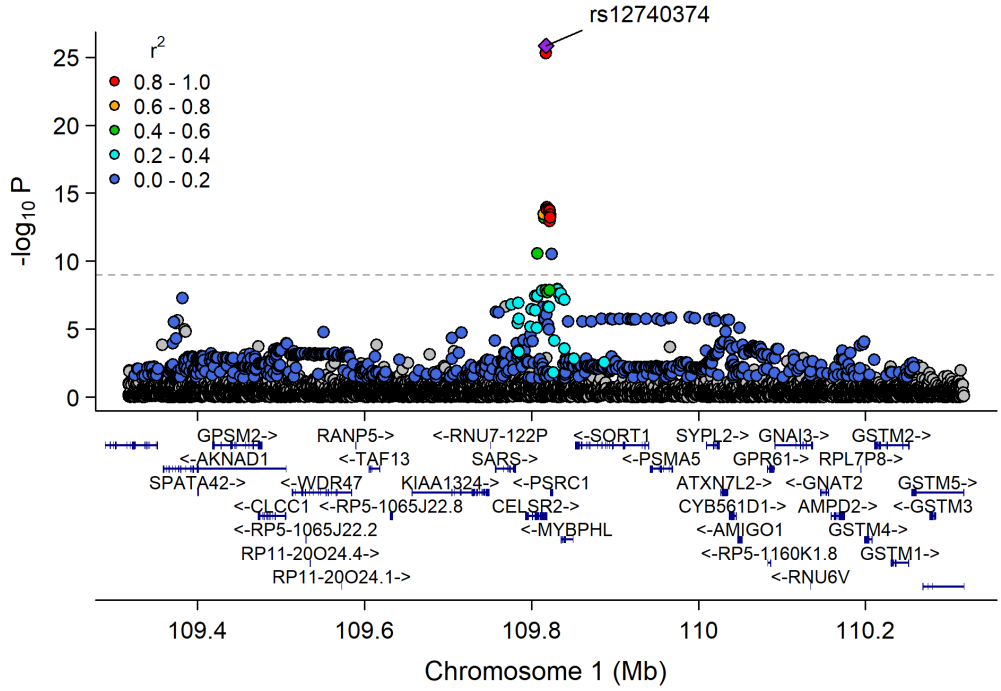




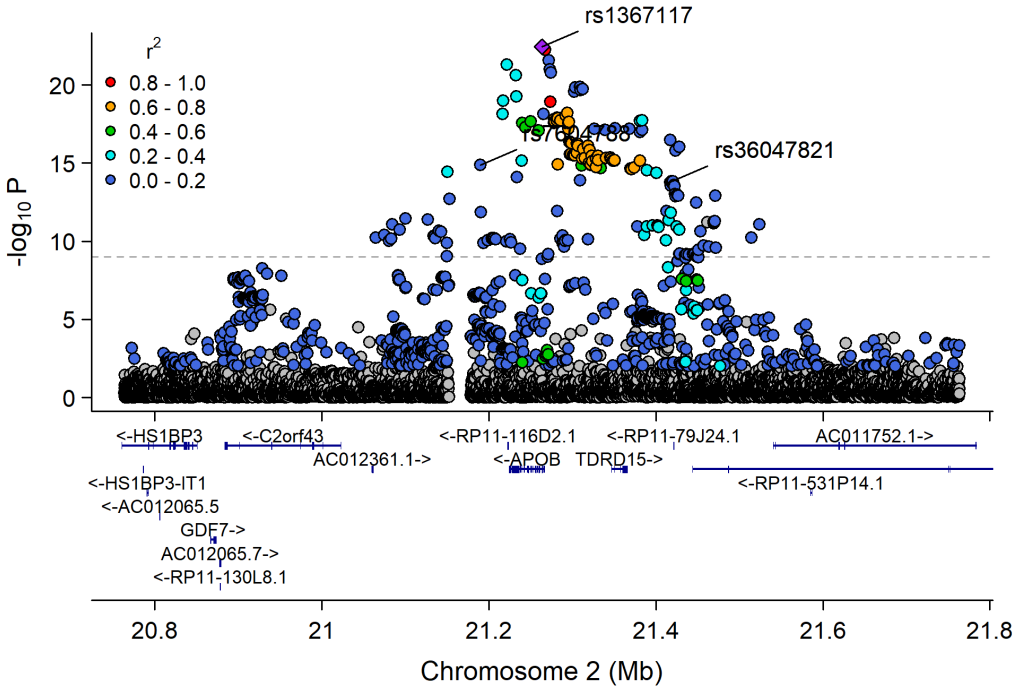
Chr1__53272879__57300396



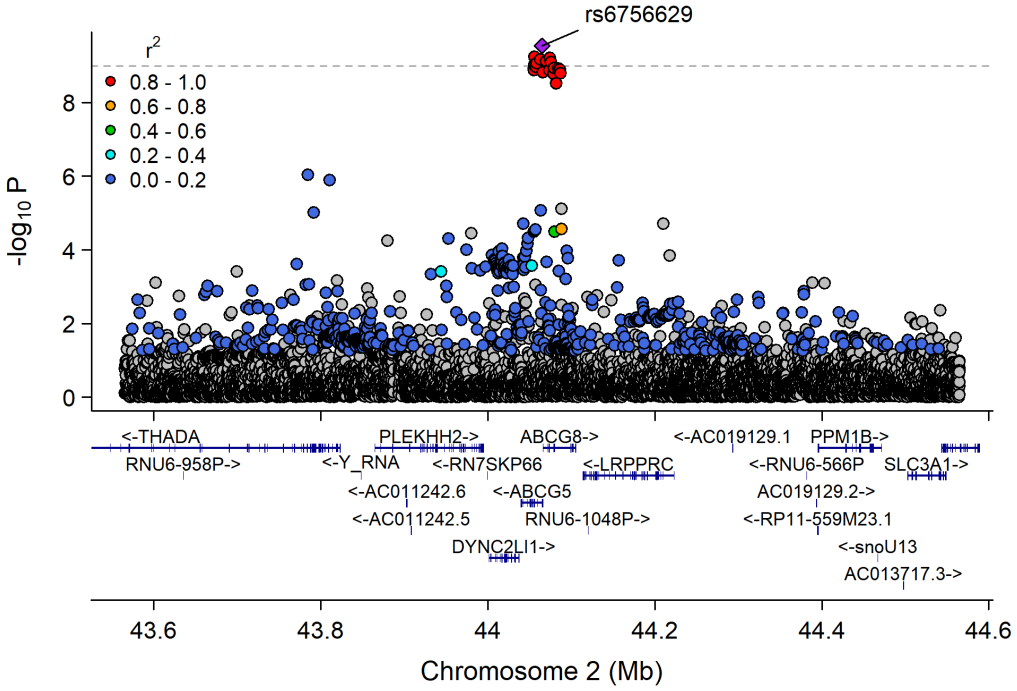
Chr1__109696238__110162190



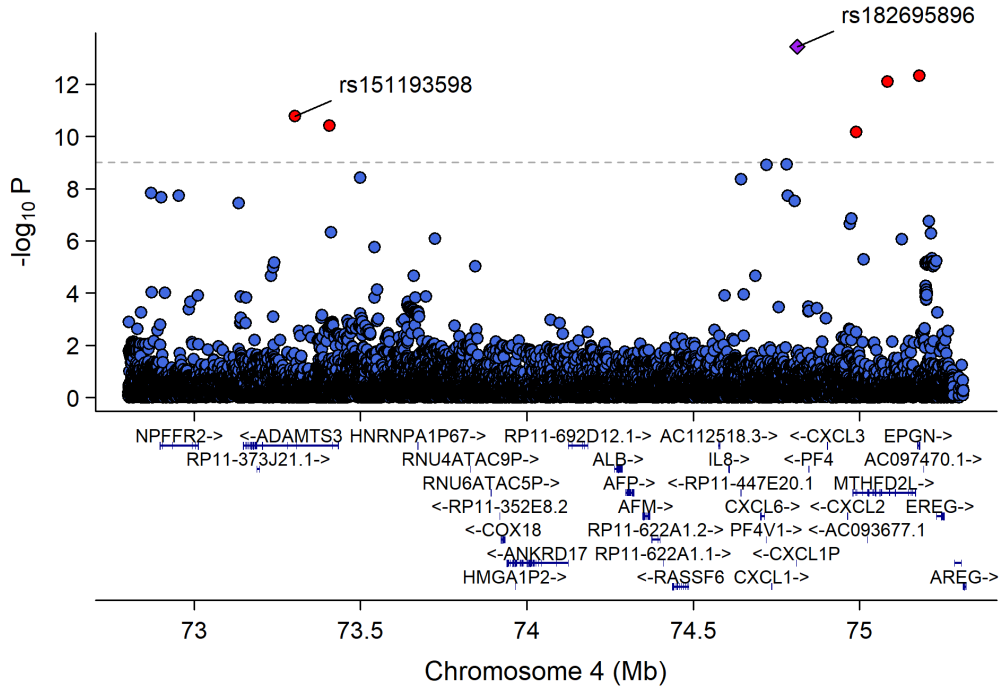
Chr2_19947287_22427492



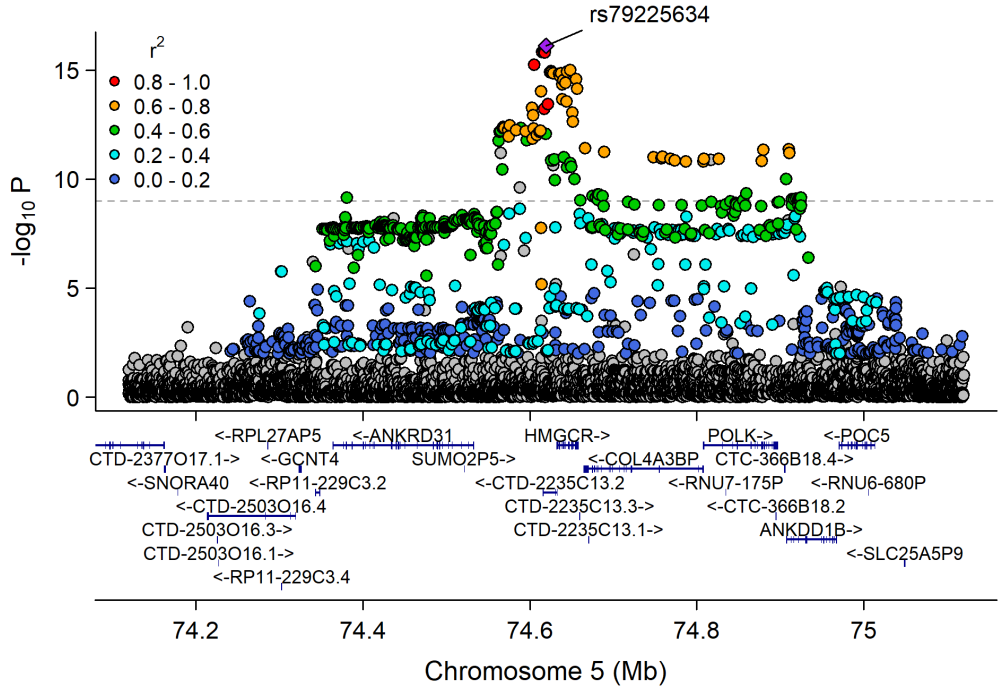
Chr2_43065153_45056024



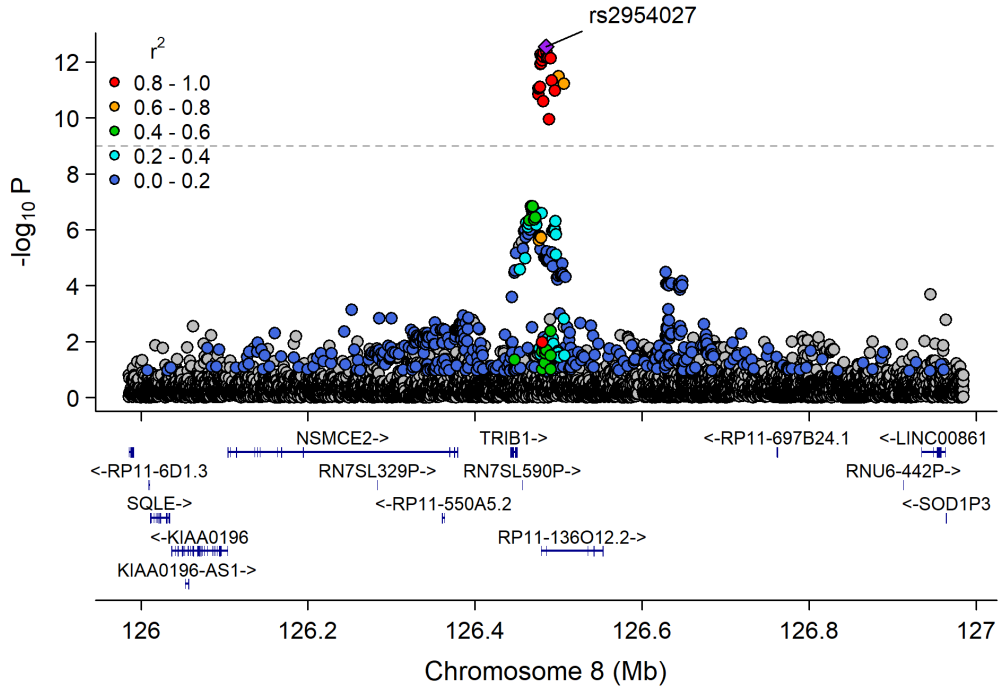
Chr4_71881824_76681229



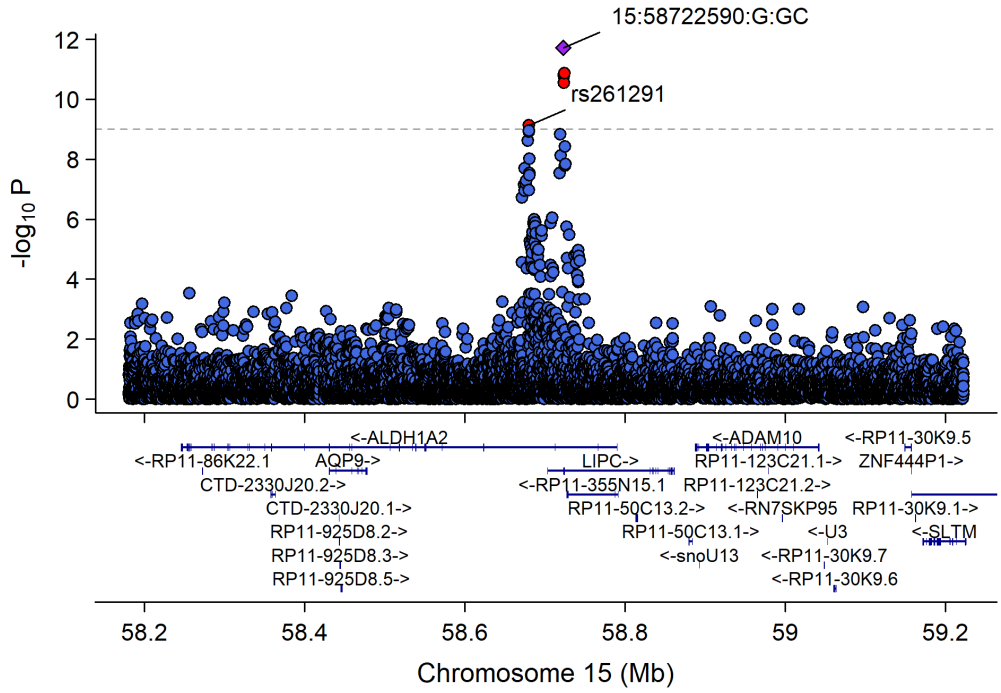
Chr5_74242002_75221582



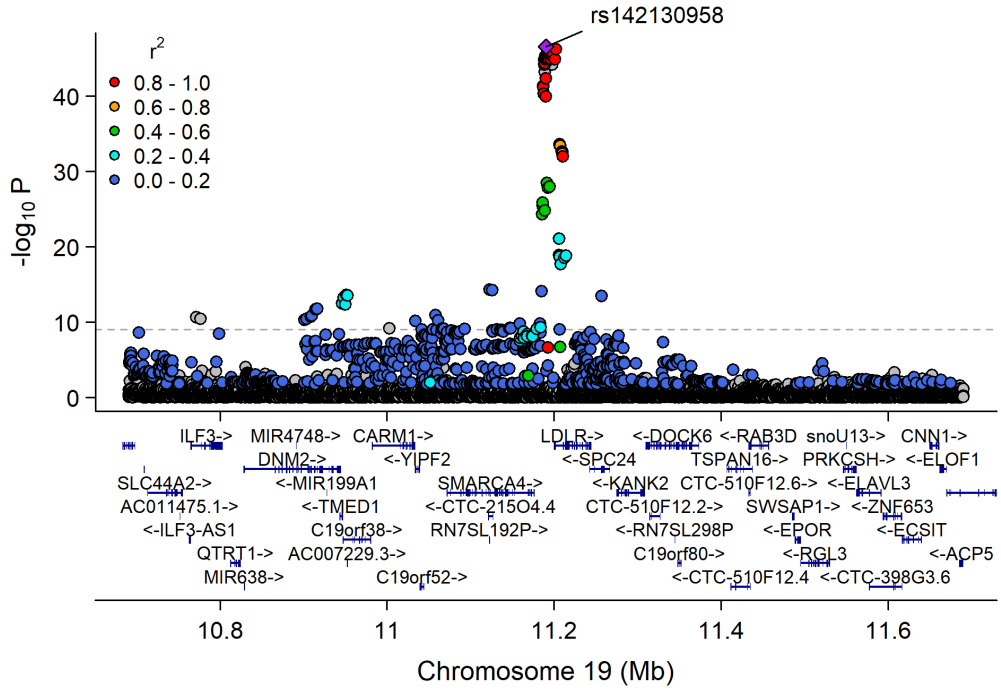
Chr8_126435663_126578061



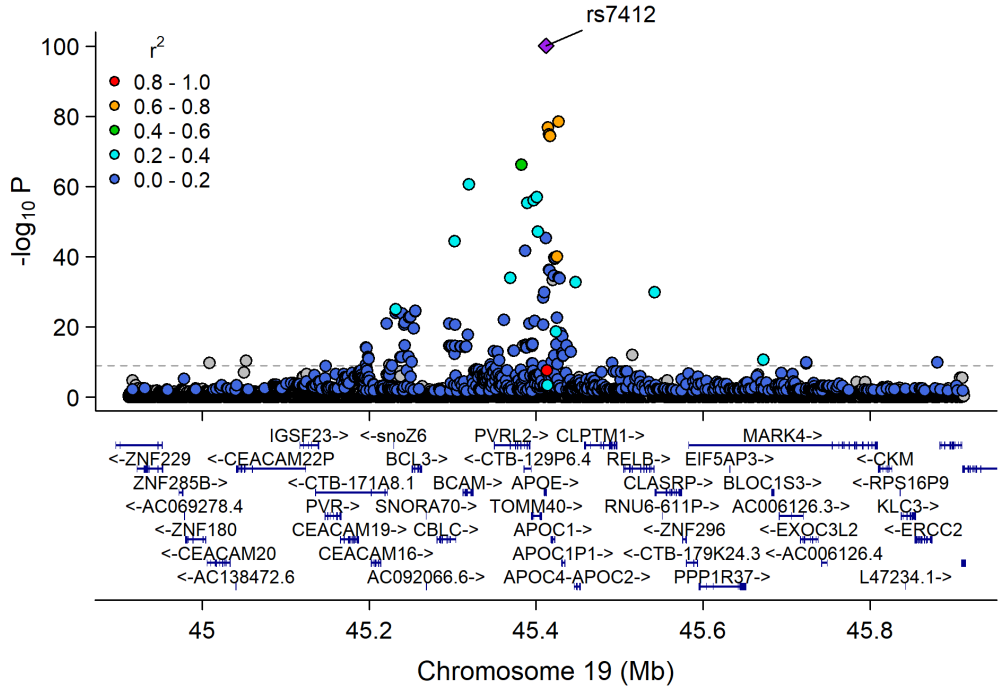
Chr15_57658798_60490883



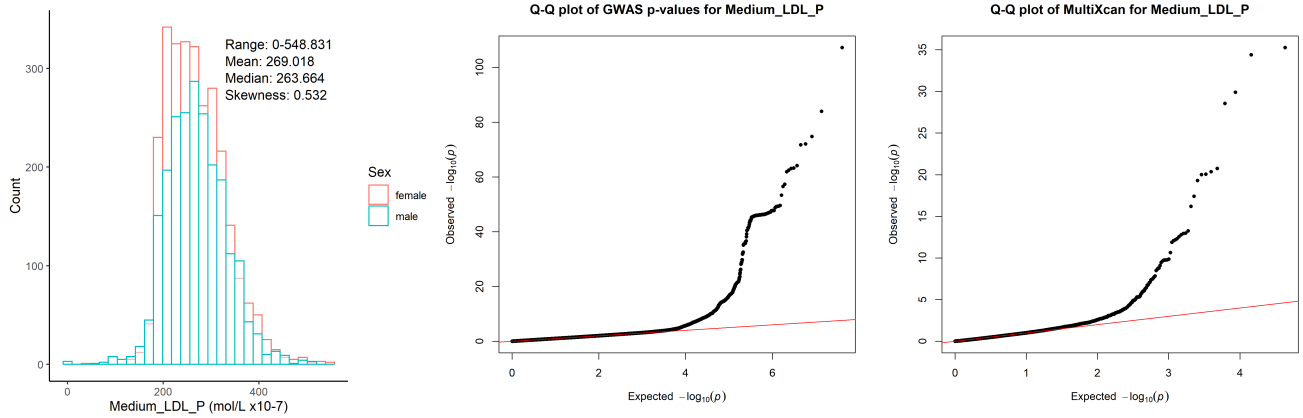
Chr19_9807999_12255594

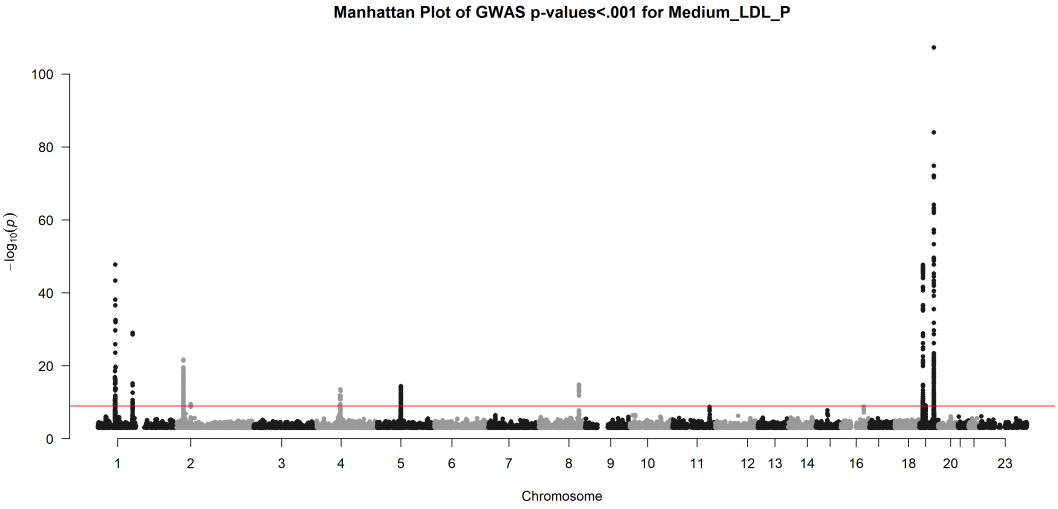


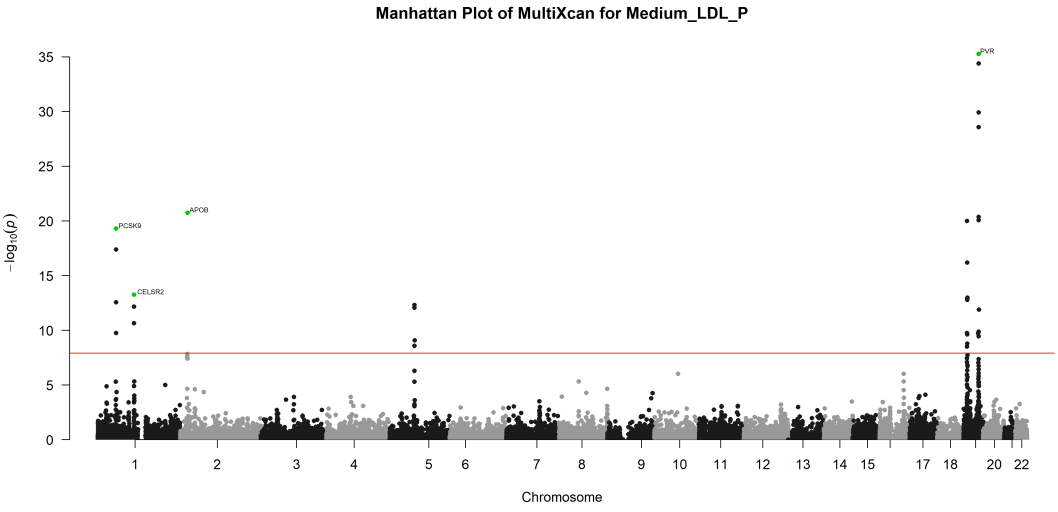
Chr19_44063363_46637375



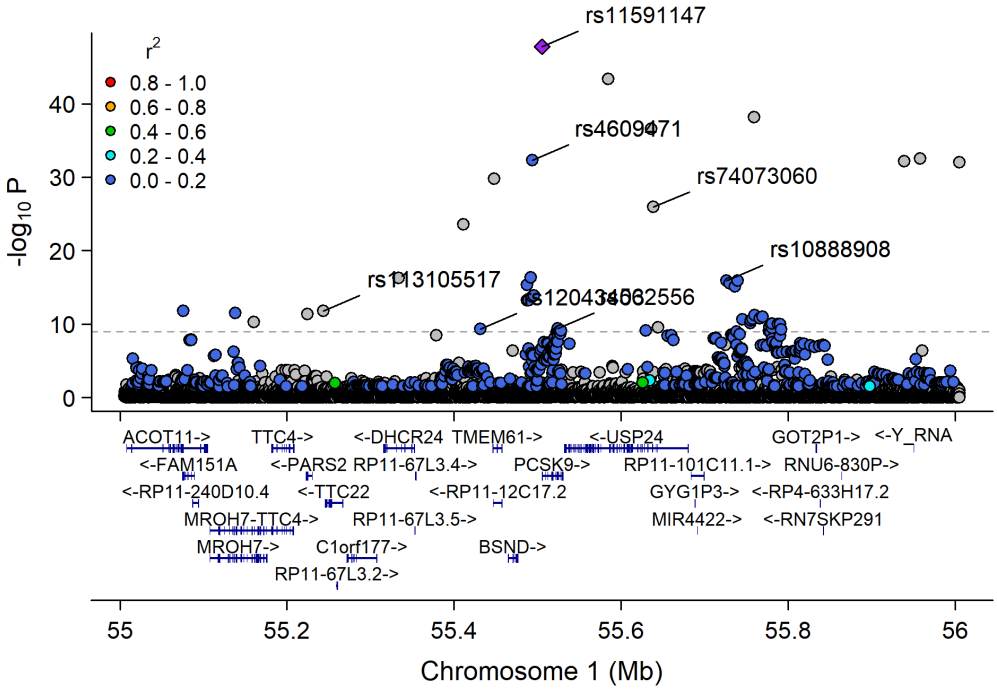
Concentration of medium LDL particles (mol/l)



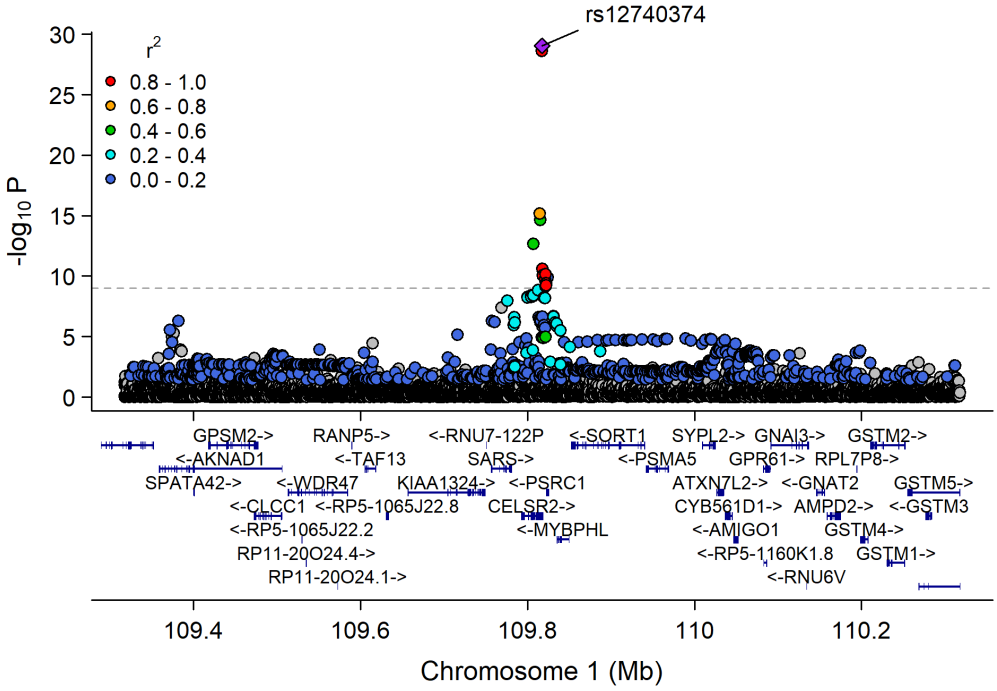




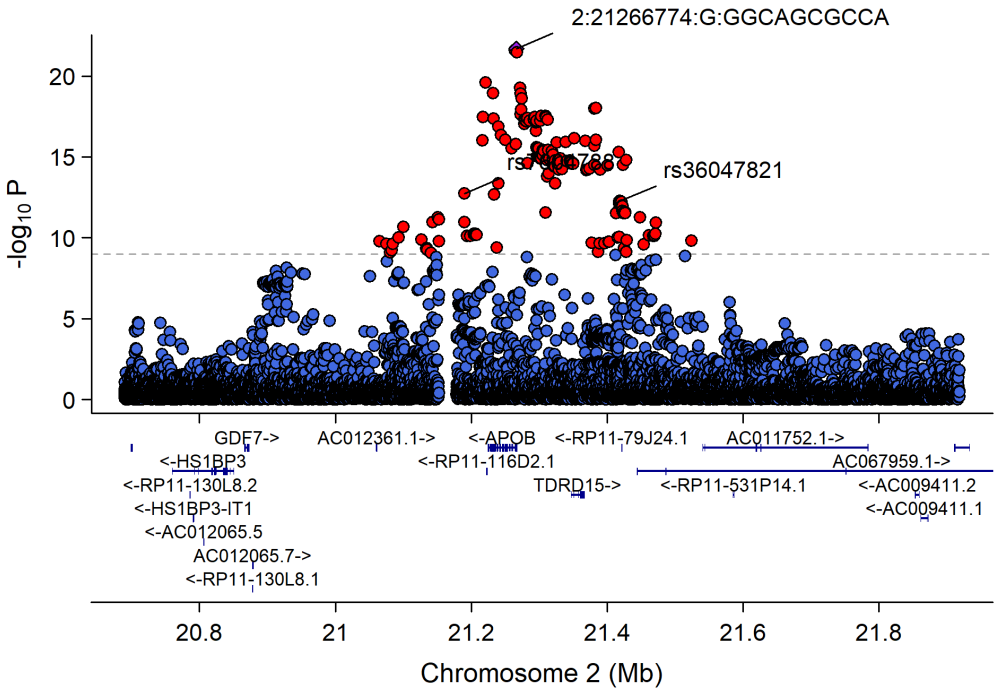
Chr1_53272879_57300396



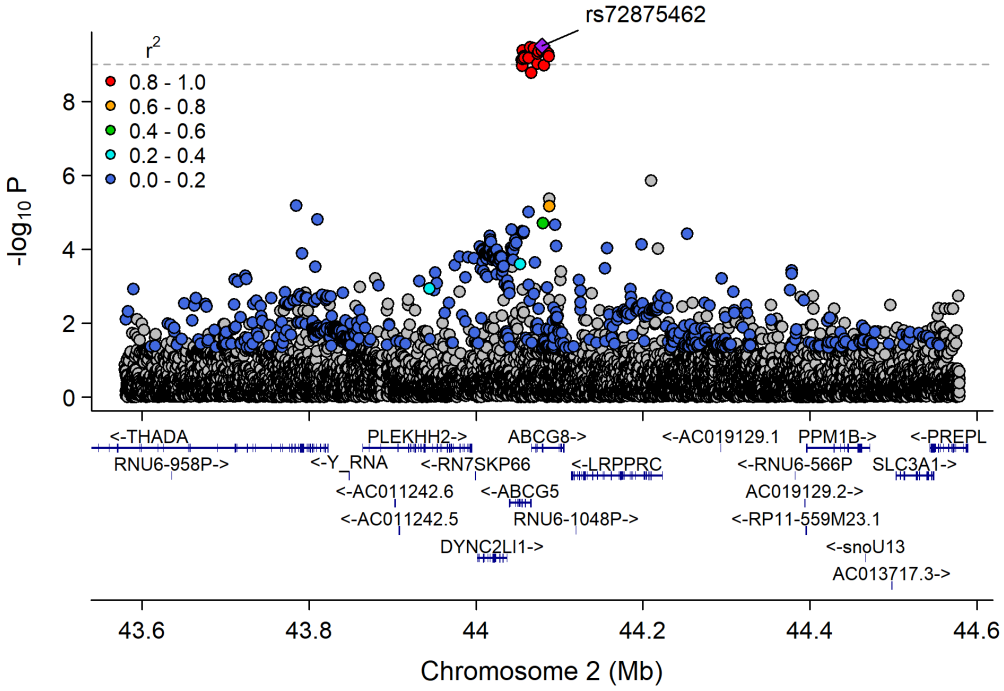
Chr1_109696238_110162190



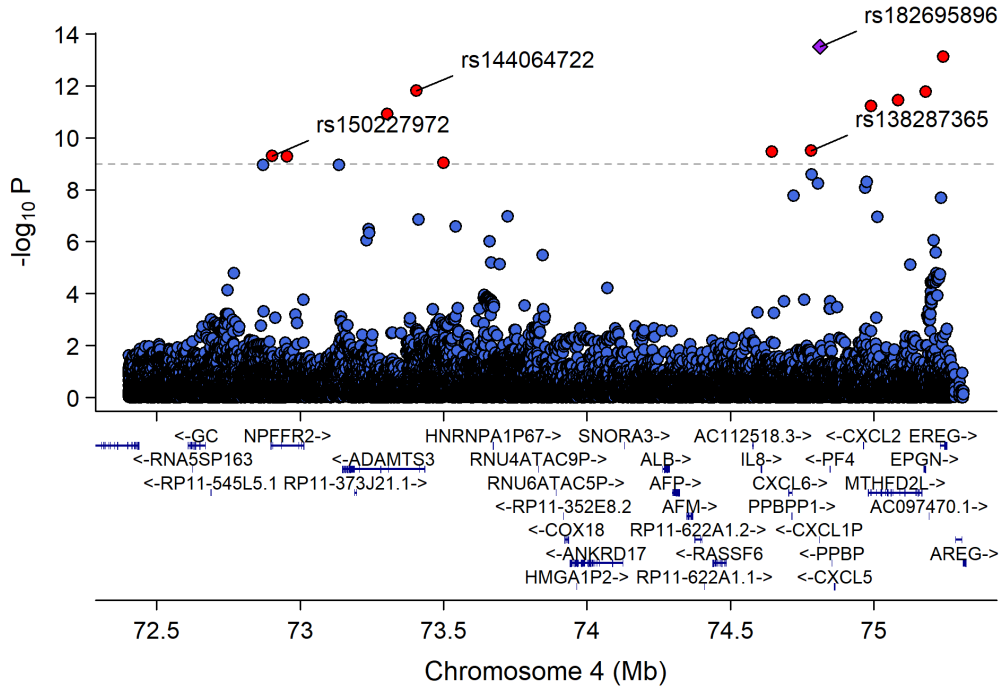
Chr2_19947287_22427492



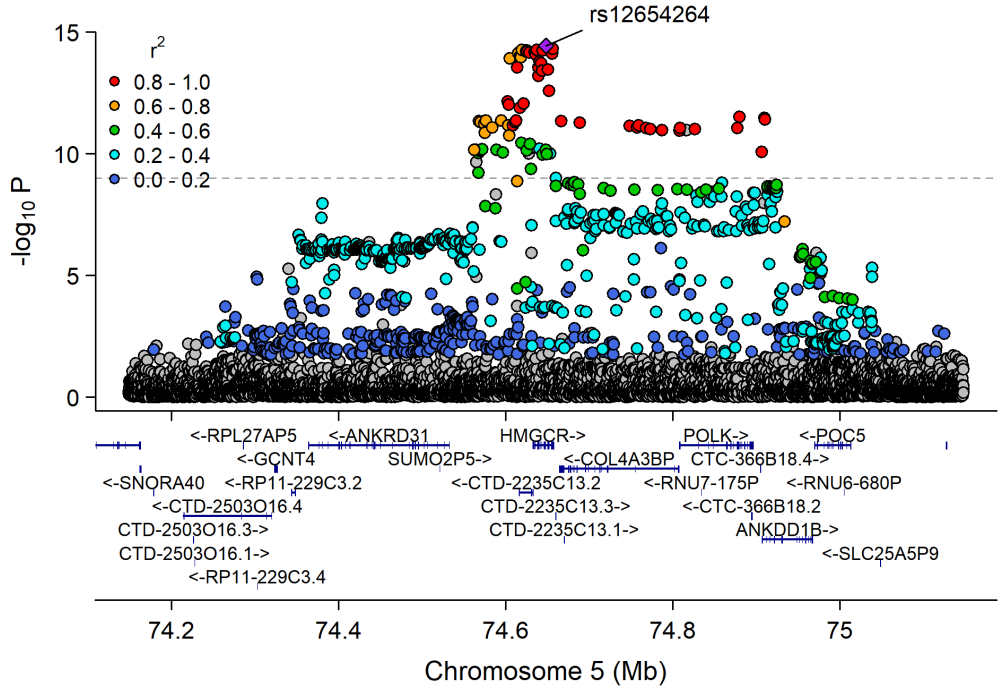
Chr2_43065153_45056024



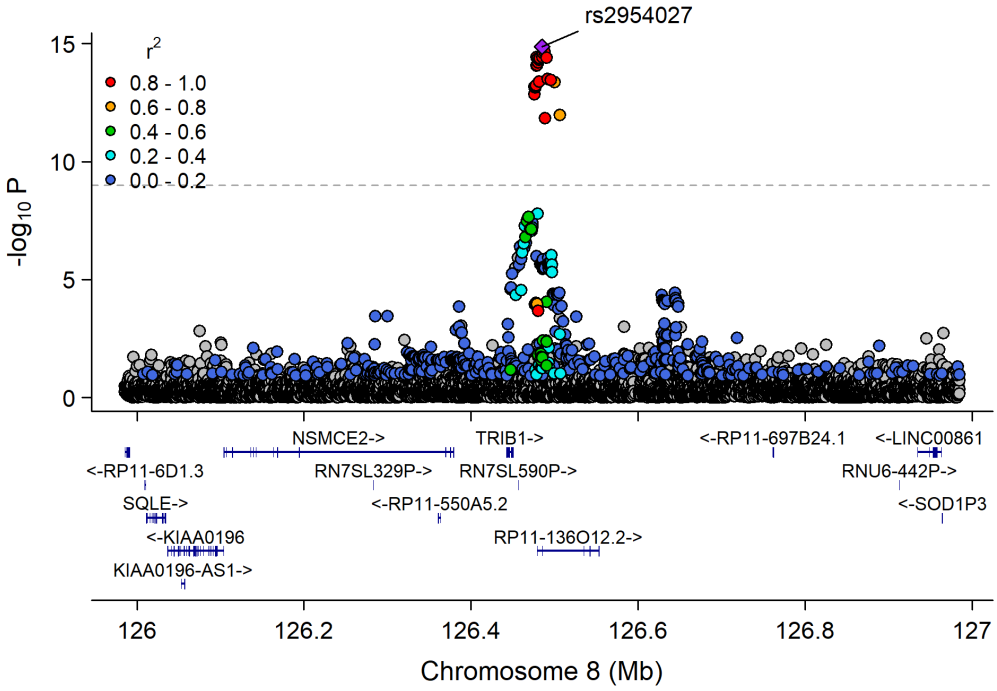
Chr4_71881824_76681229



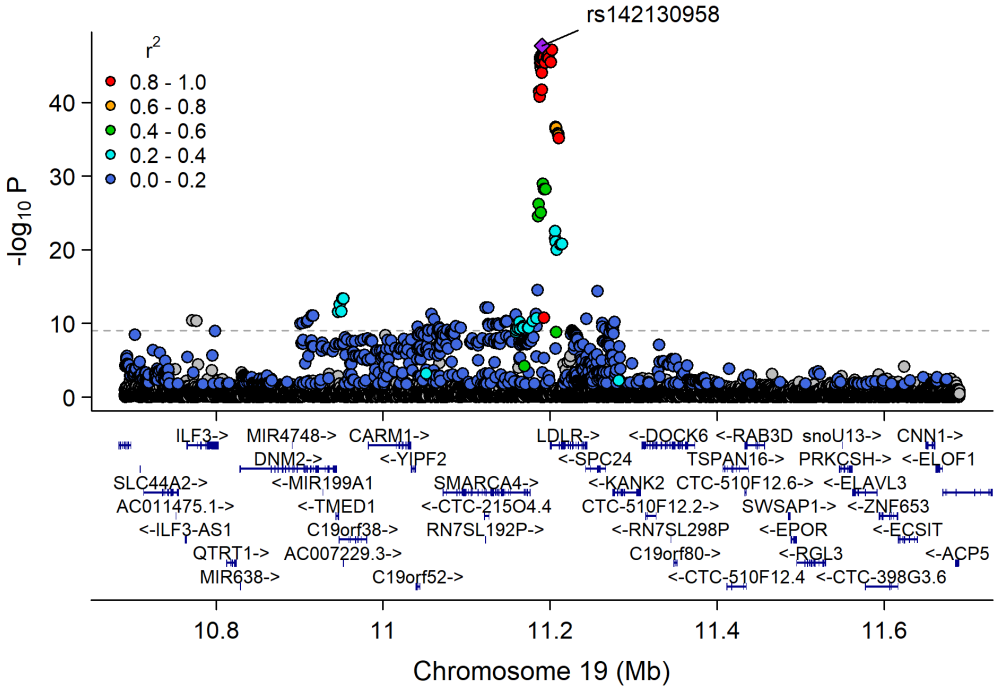
Chr5_74242002_75221582



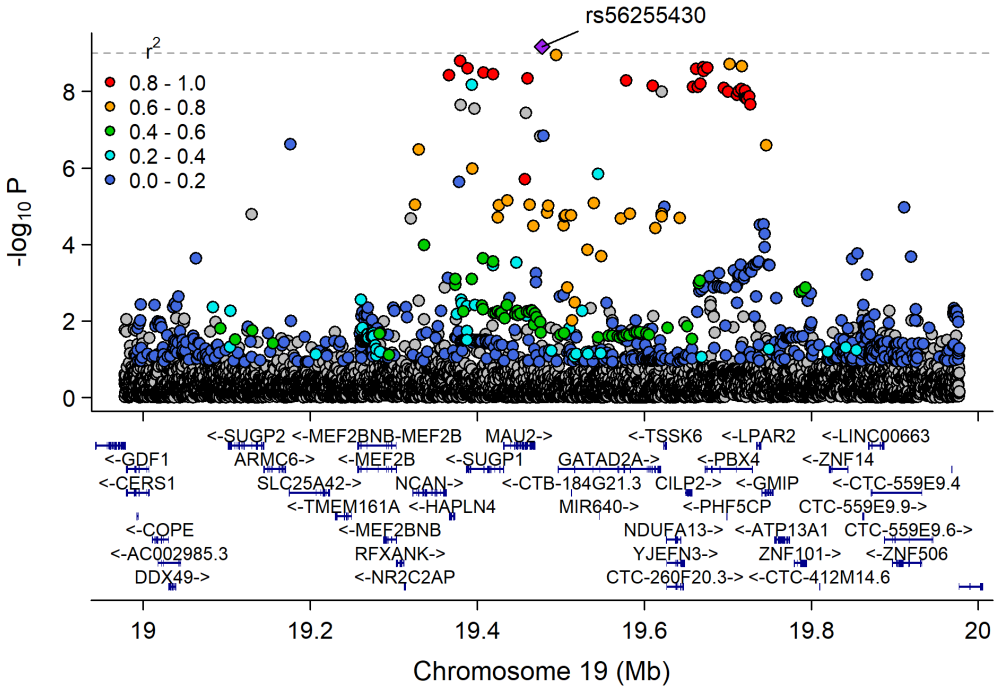
Chr8_126435663_126578061



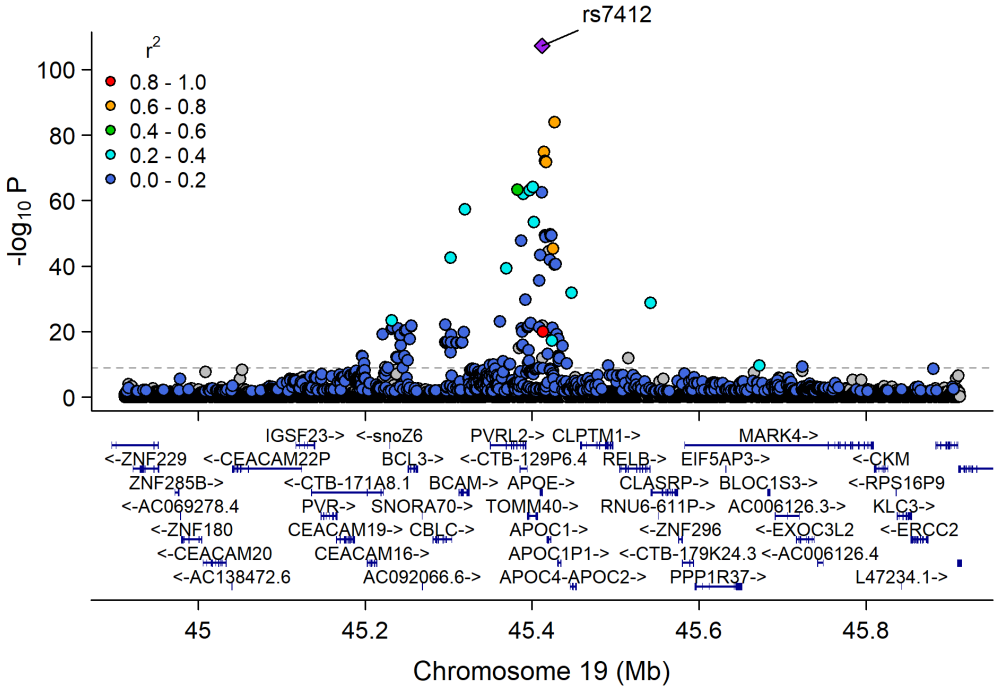
Chr19_9807999_12255594



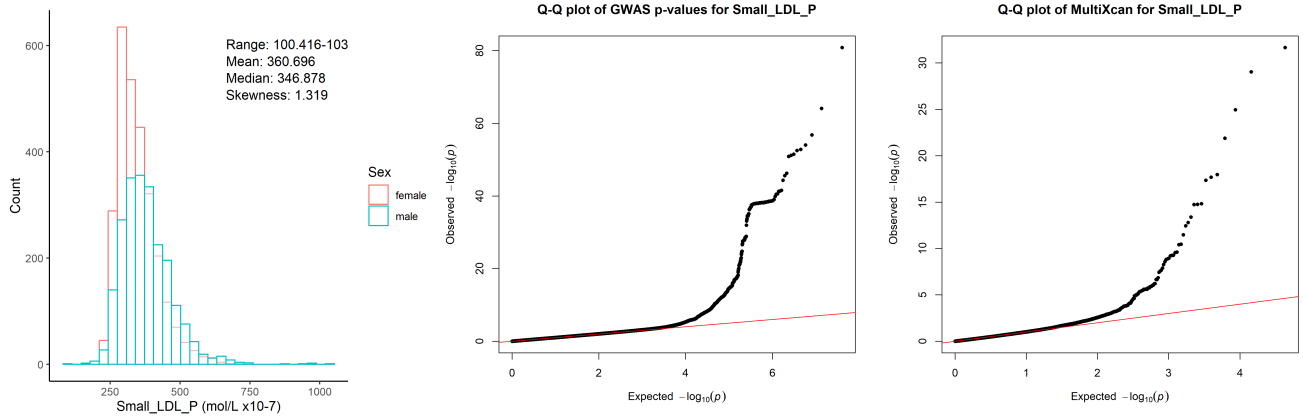
Chr19_18370495_20841464

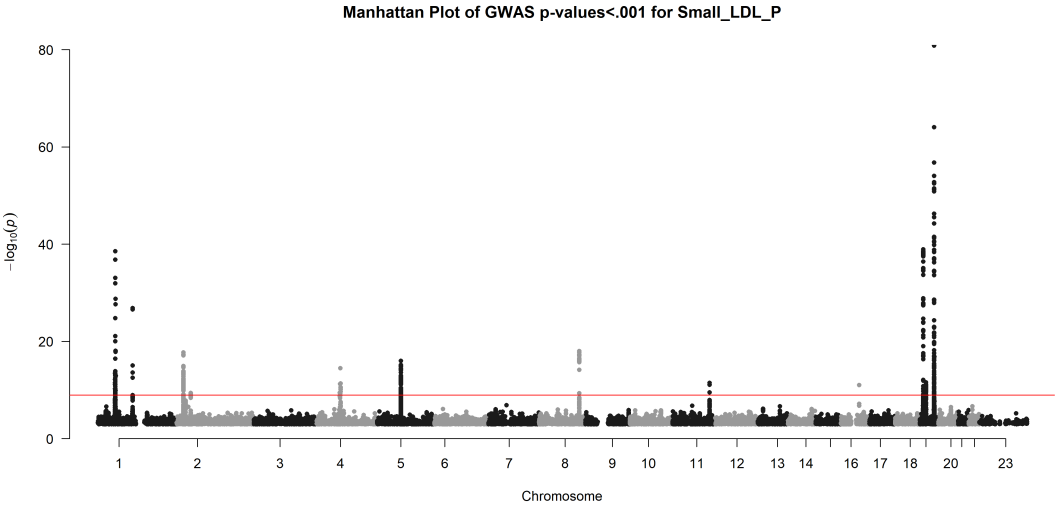


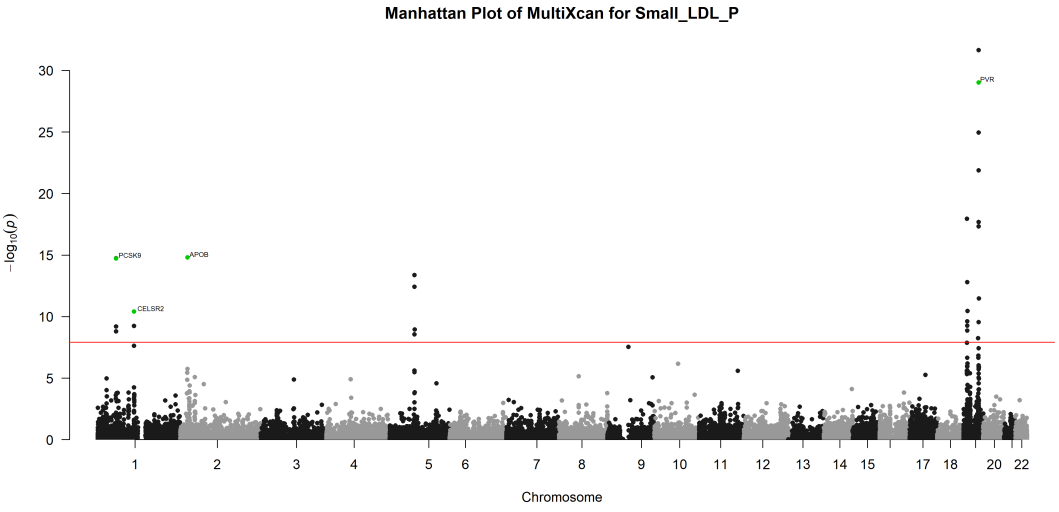
Chr19_44063363_46637375



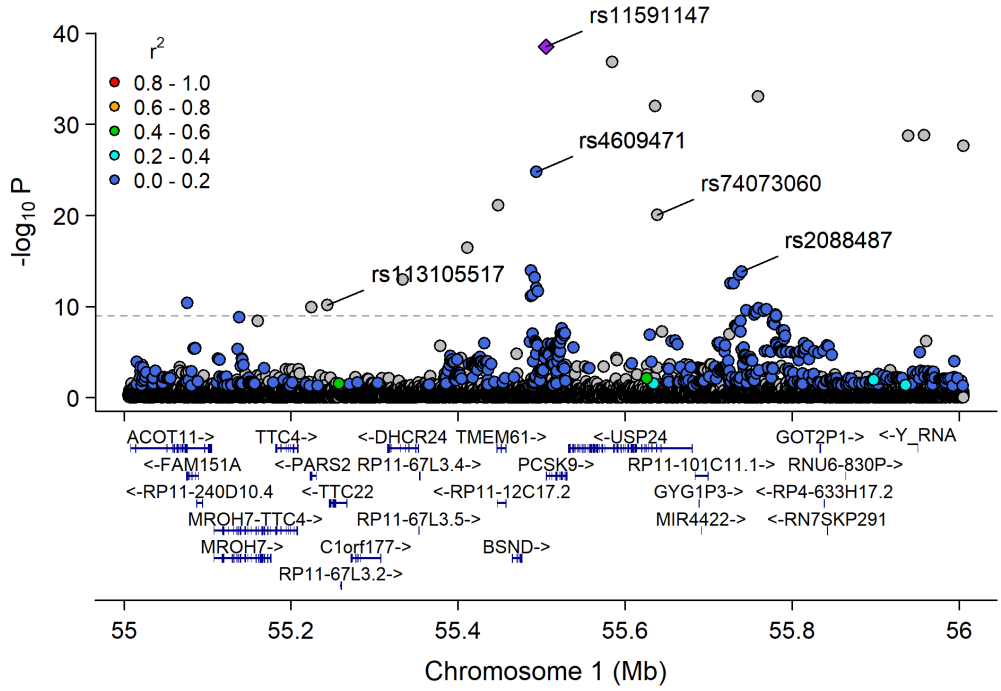
Concentration of small LDL particles (mol/l)



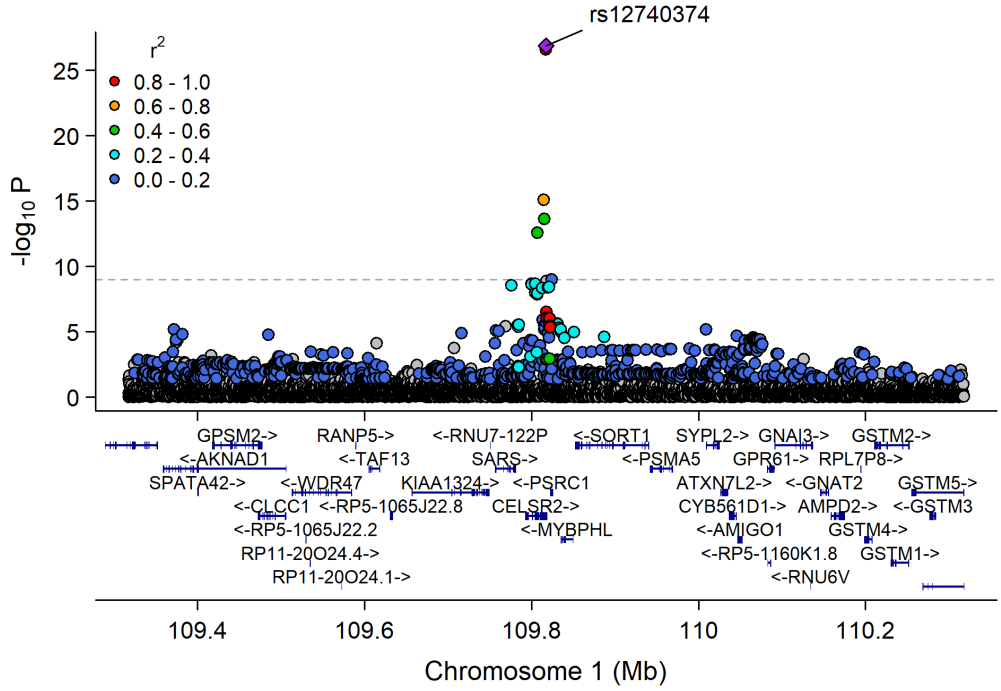




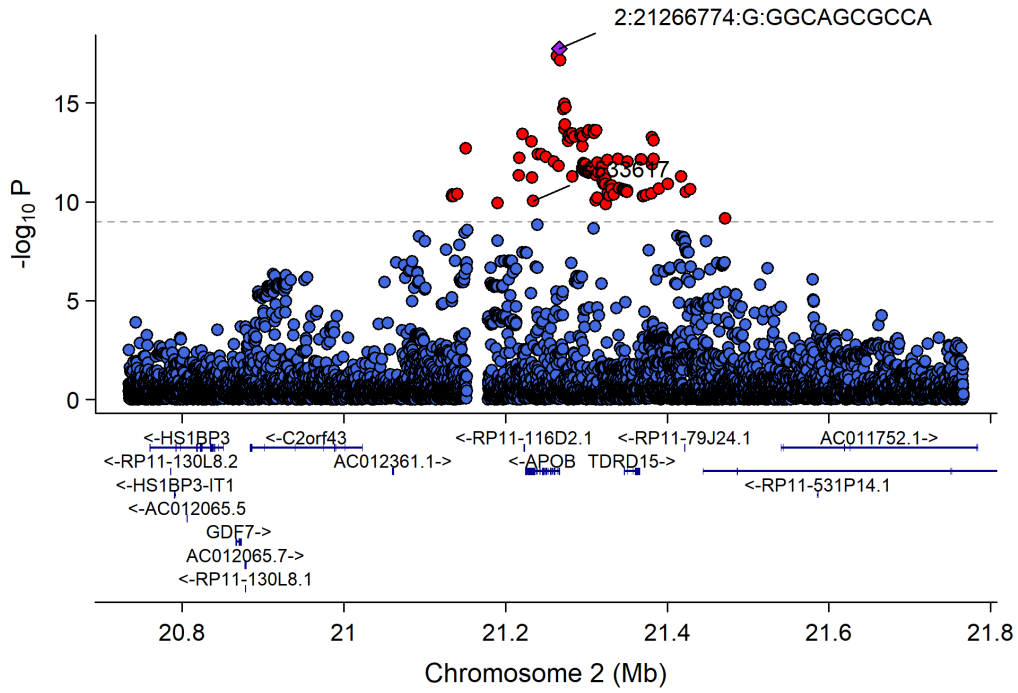
Chr1__53272879__57300396



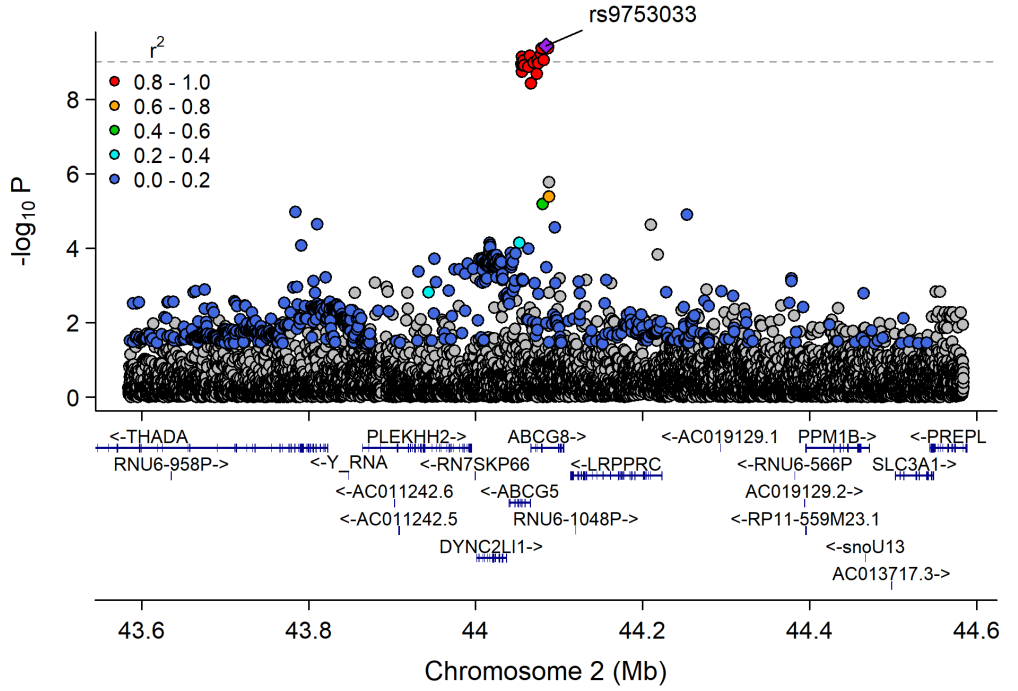
Chr1__109696238__110162190



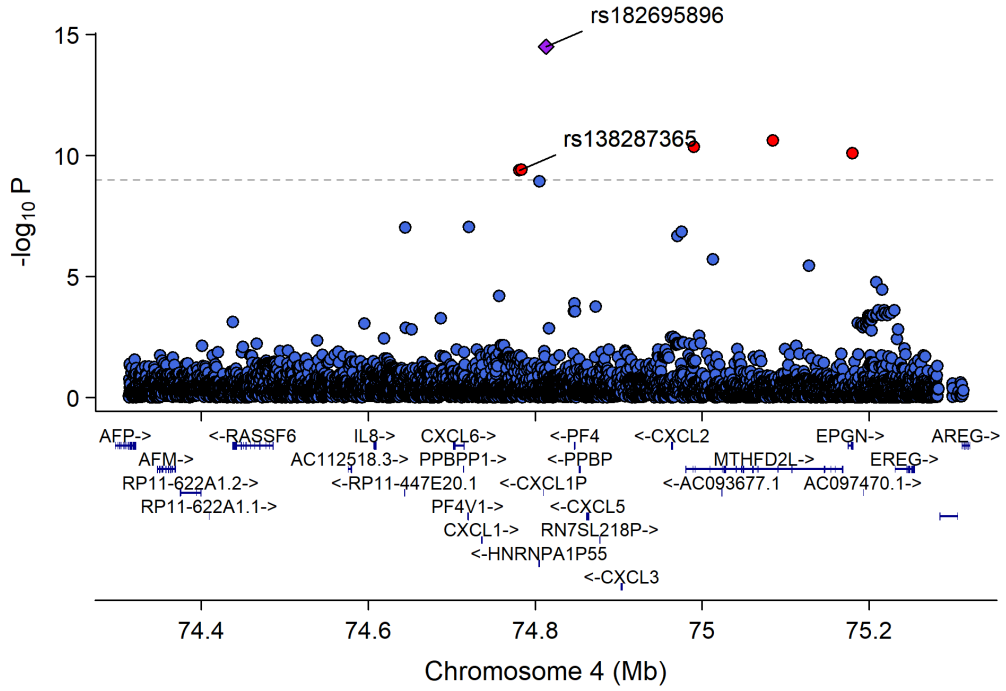
Chr2_19947287_22427492



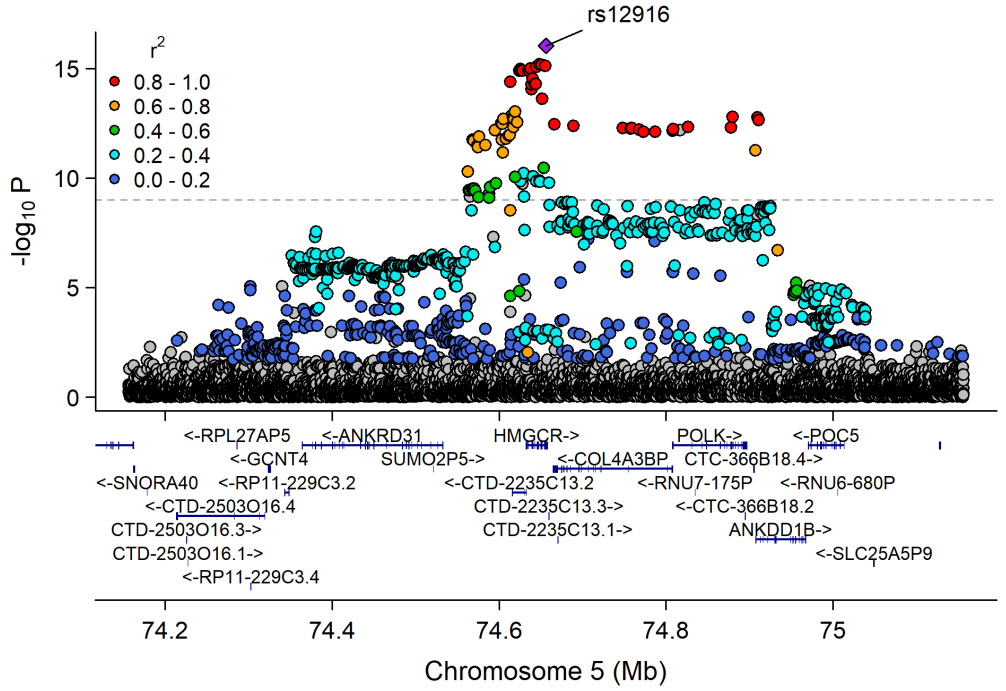
Chr2_43065153_45056024



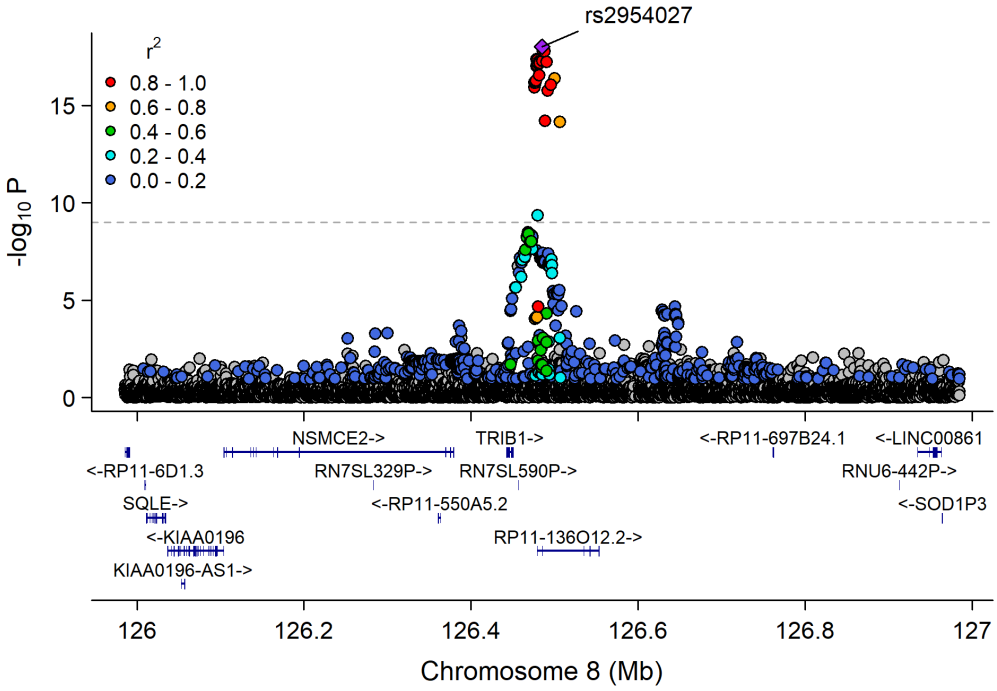
Chr4_71881824_76681229



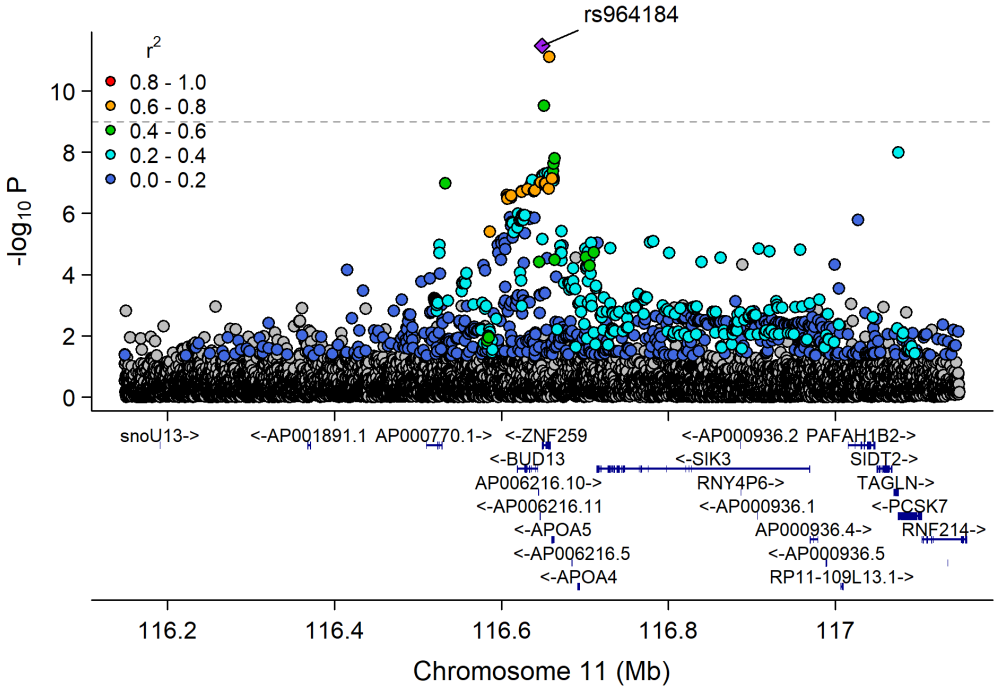
Chr5_74242002_75221582



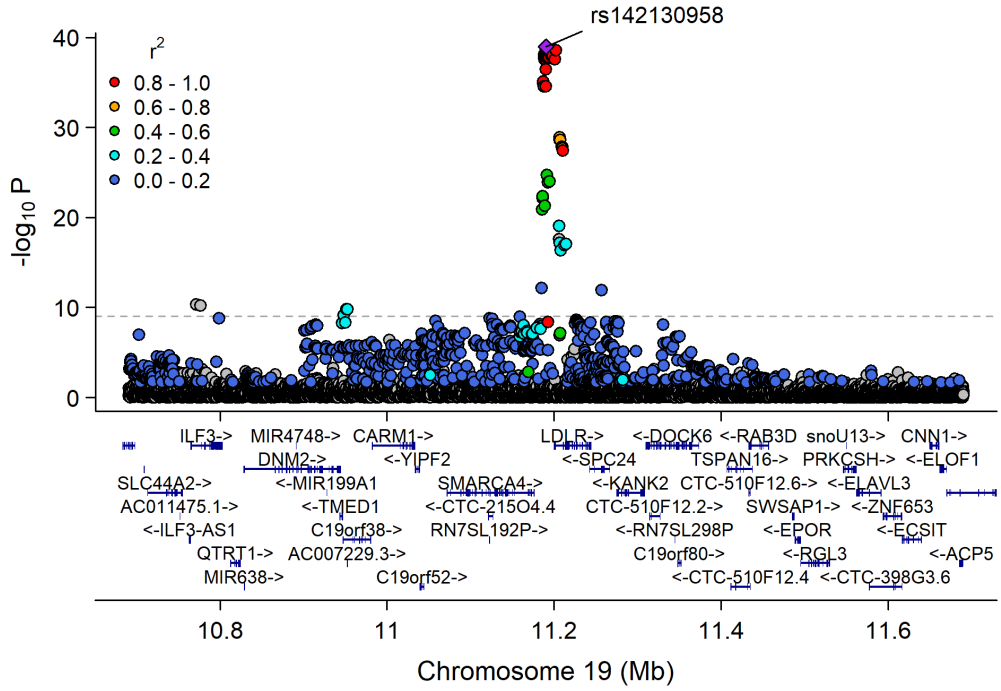
Chr8_126435663_126578061



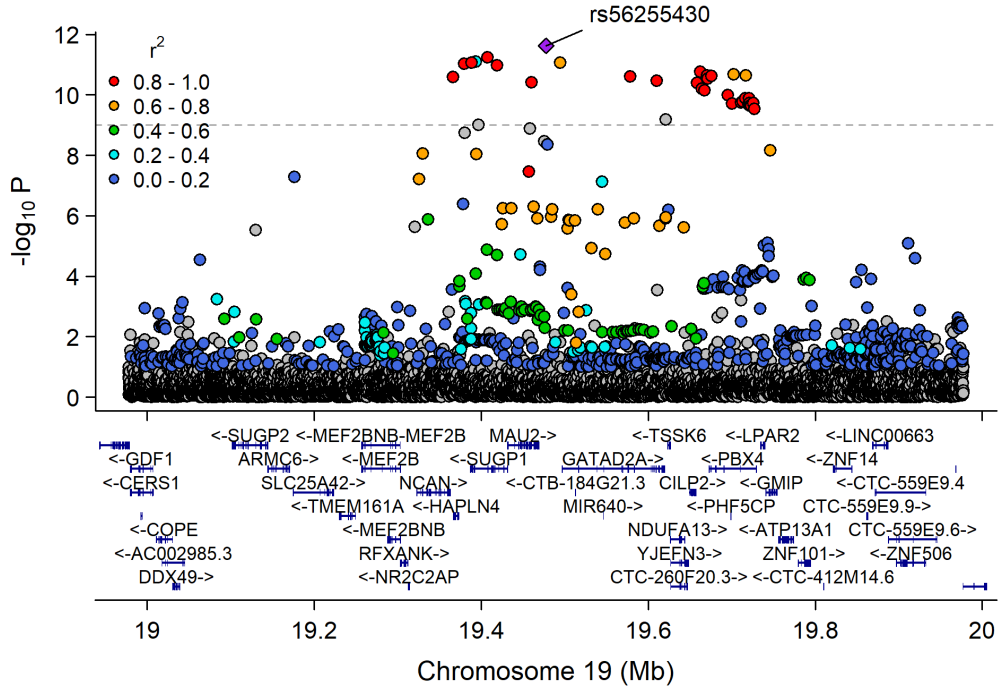
Chr11_115541901_117633315



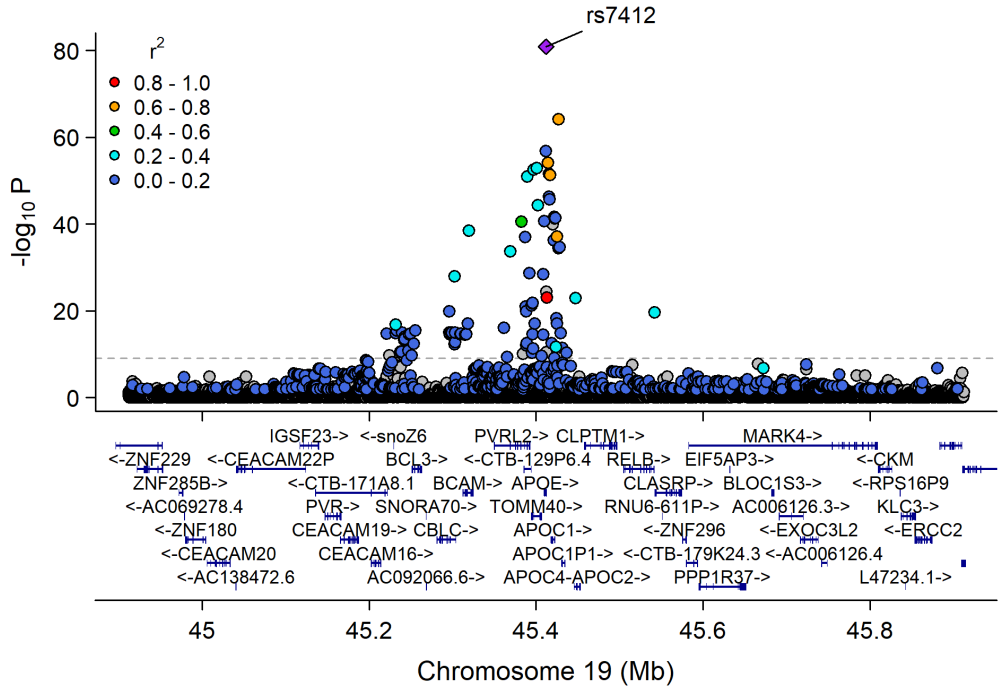
Chr19_9807999_12255594



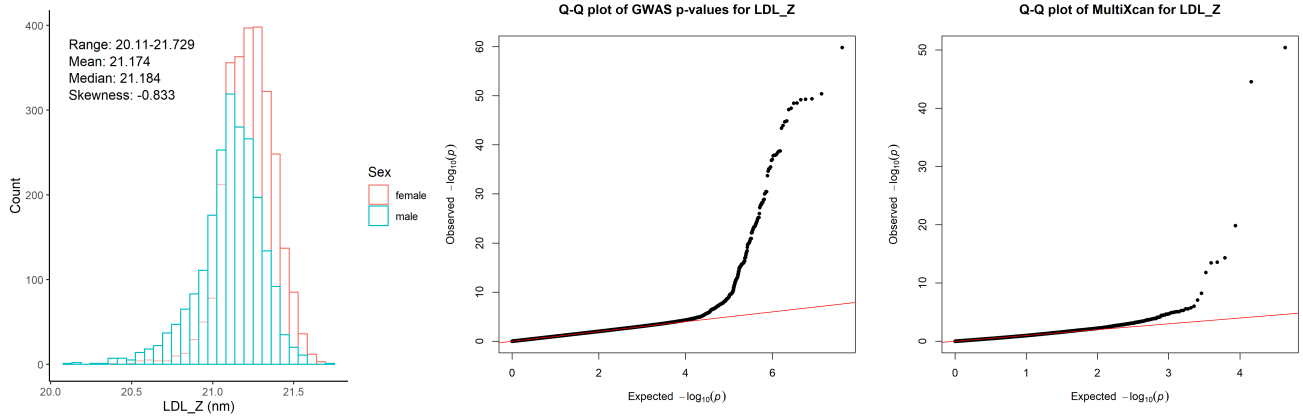
Chr19_18370495_20841464

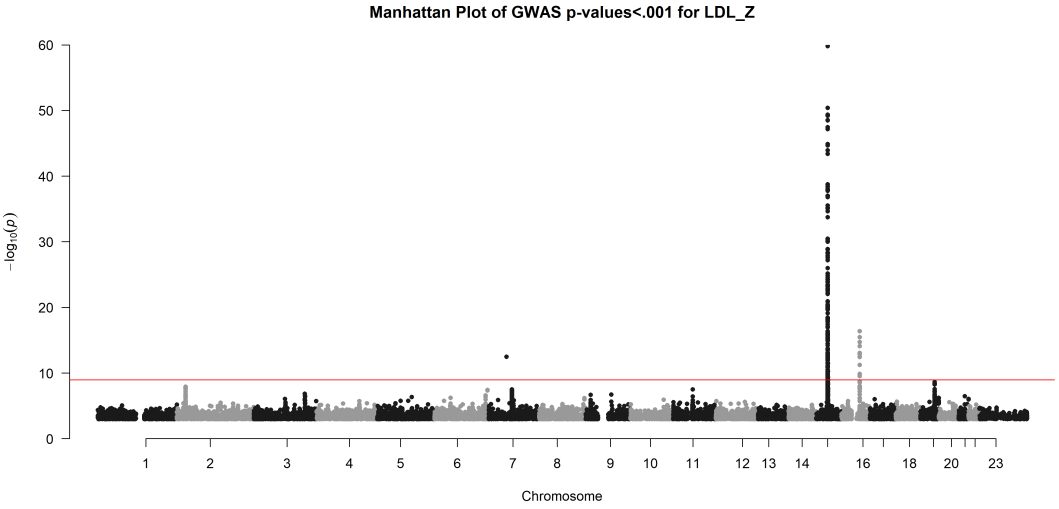


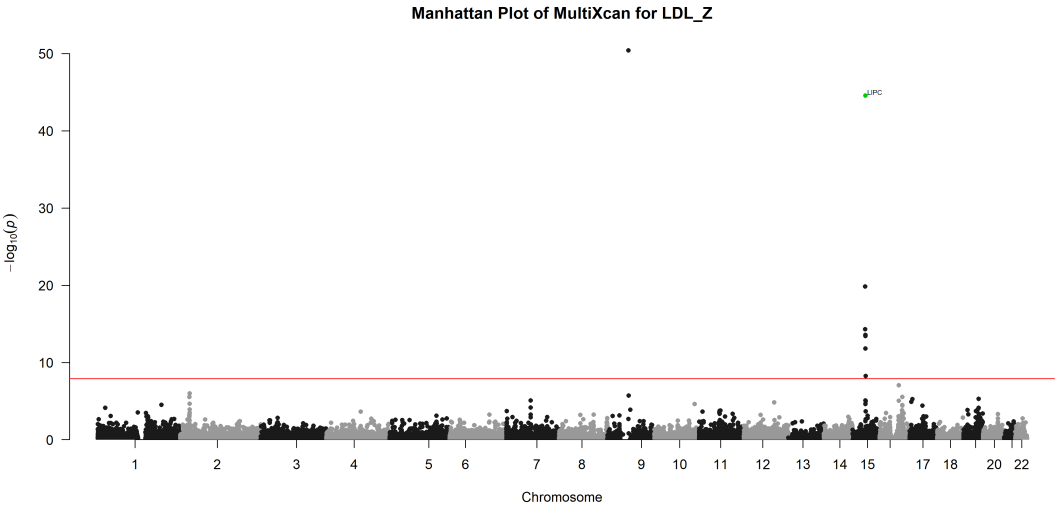
Chr19__44063363__46637375



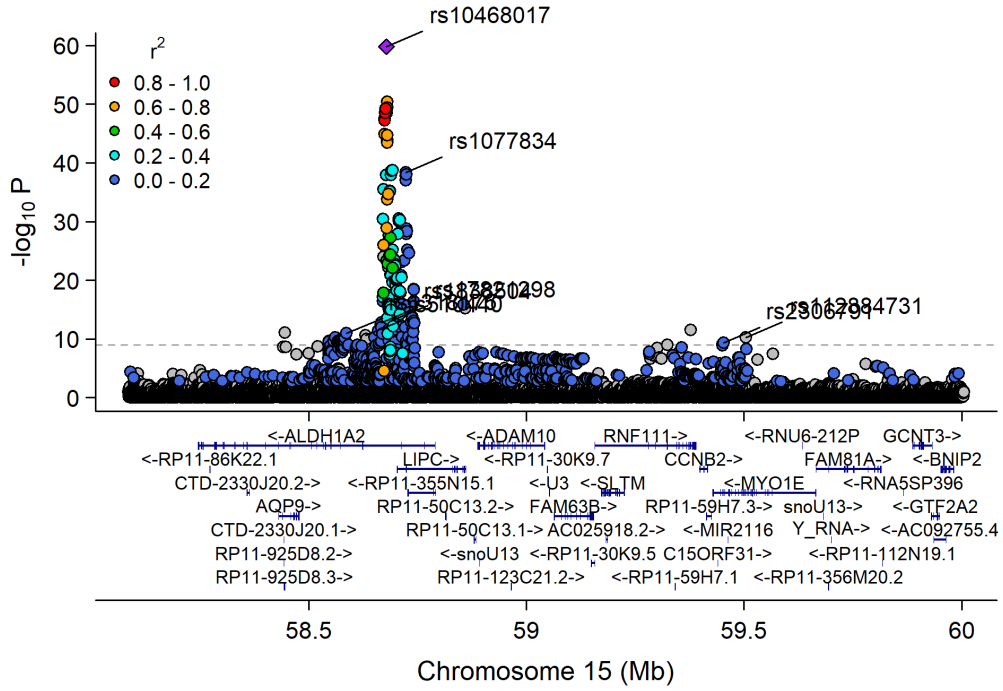
Mean diameter for LDL particles (nm)



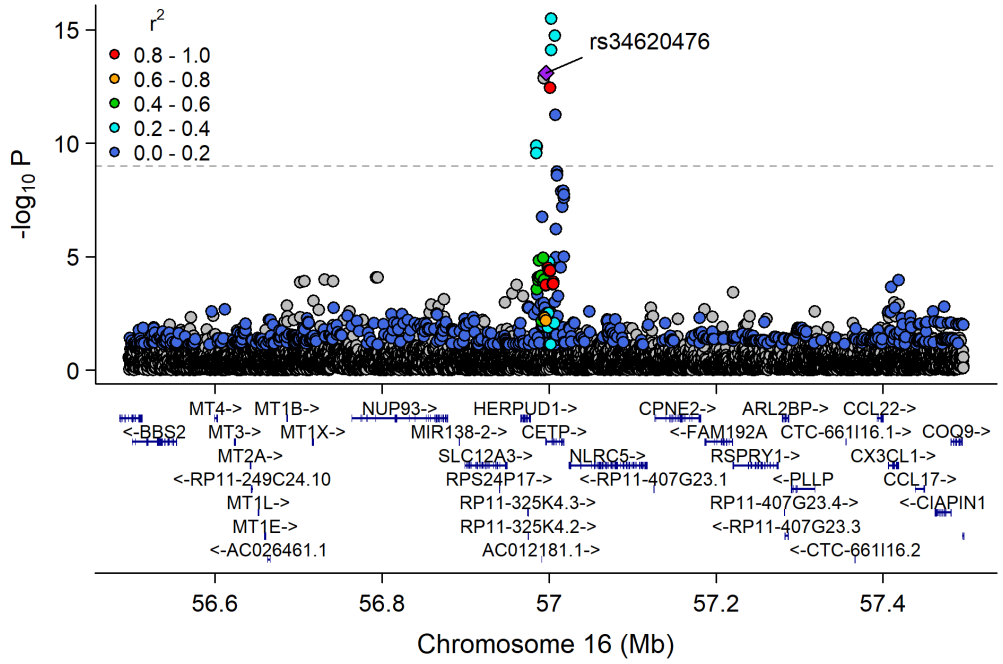




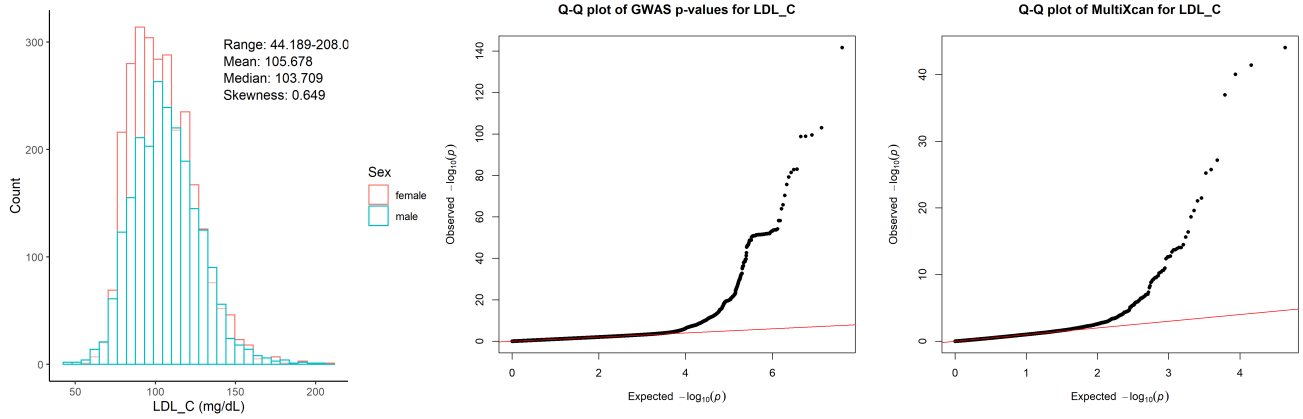
Chr15_57658798_60490883

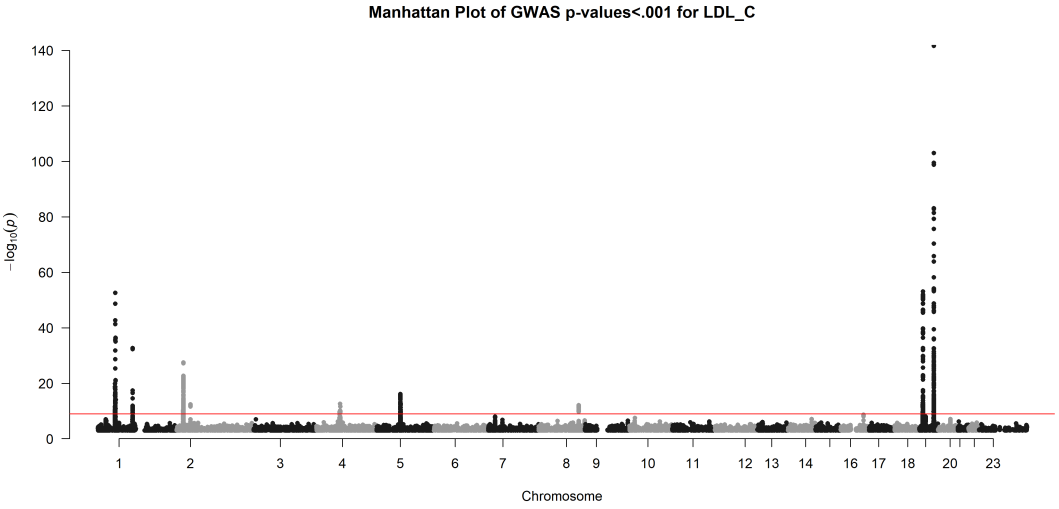


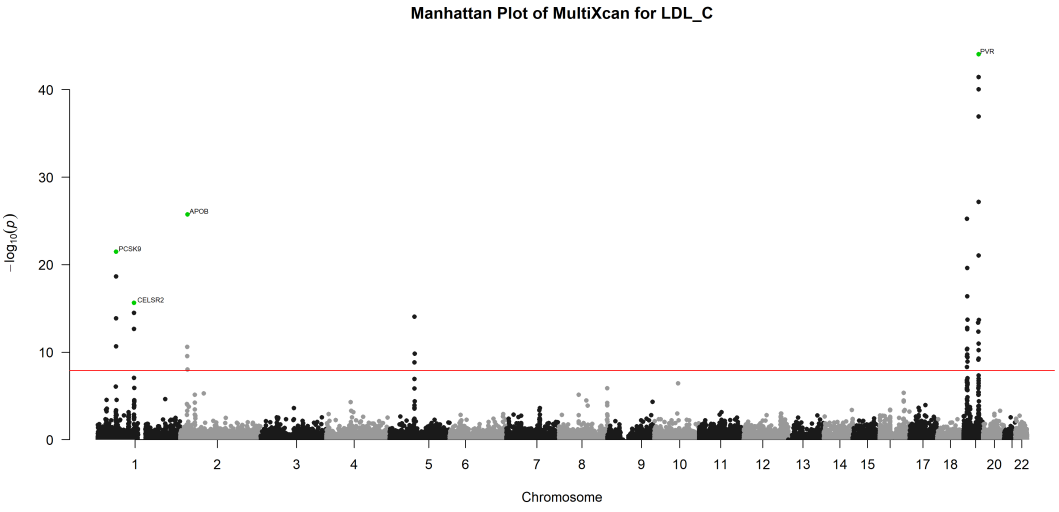
Chr16_55870822_57992421



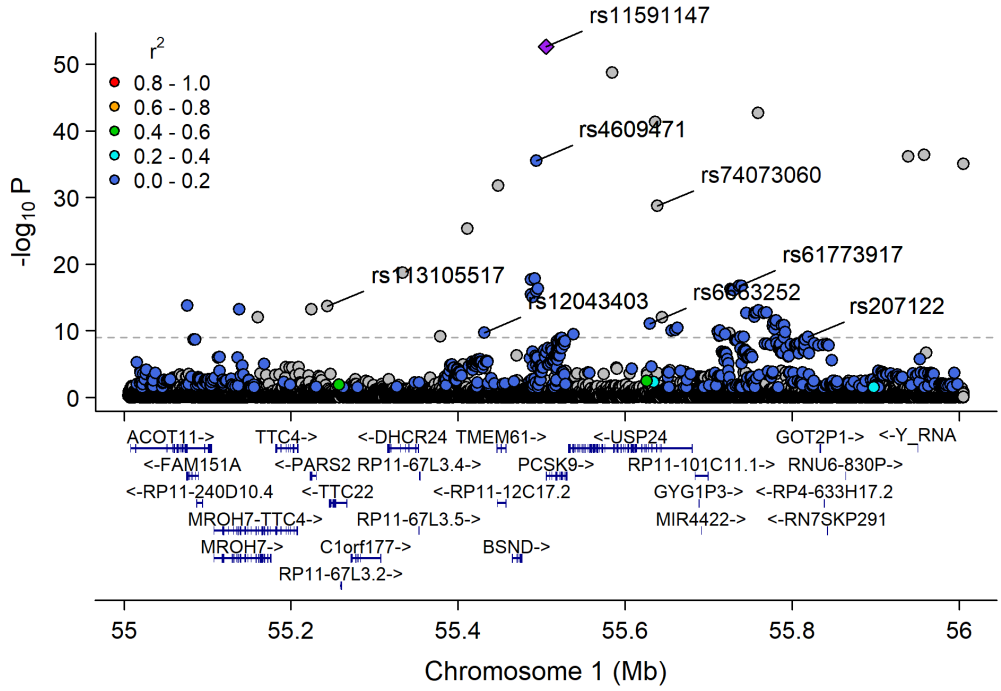
Total cholesterol in LDL (mg/dL)



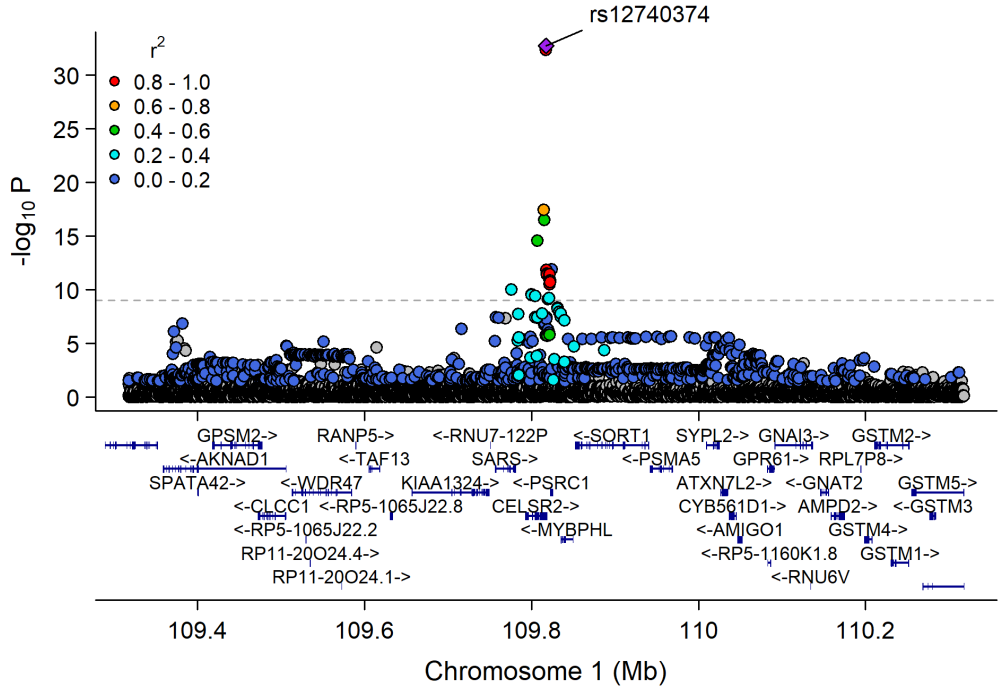




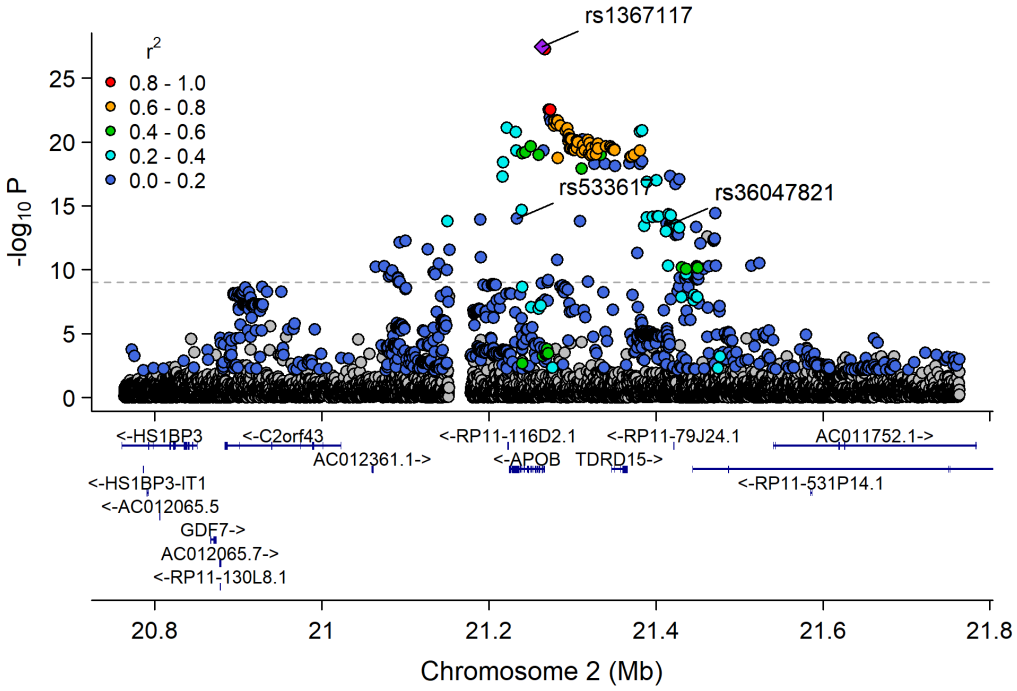
Chr1__53272879__57300396



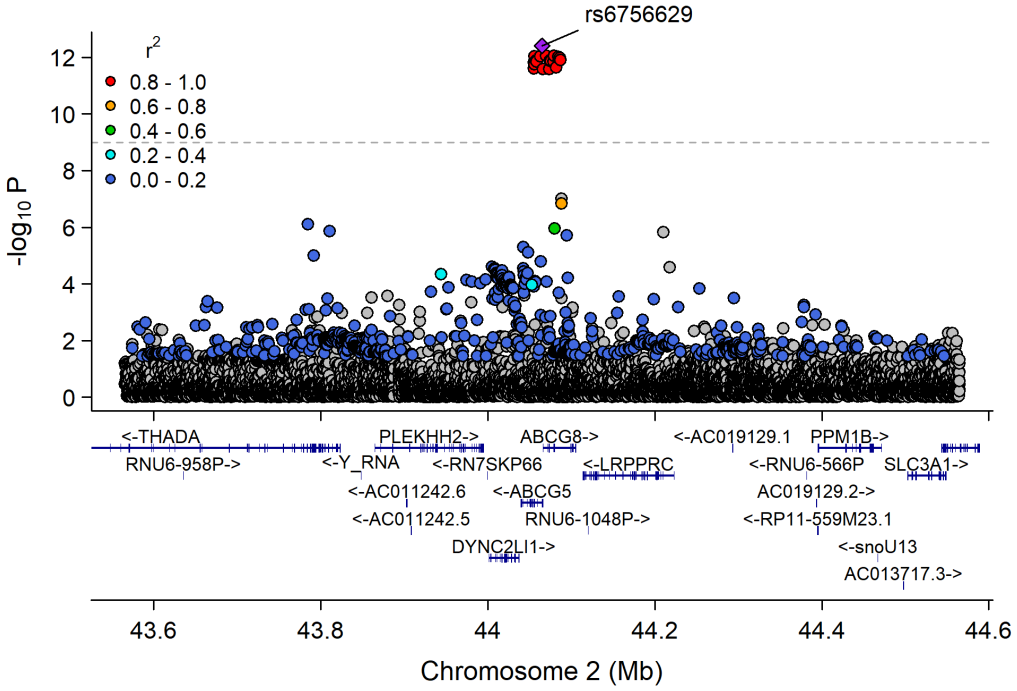
Chr1__109696238__110162190



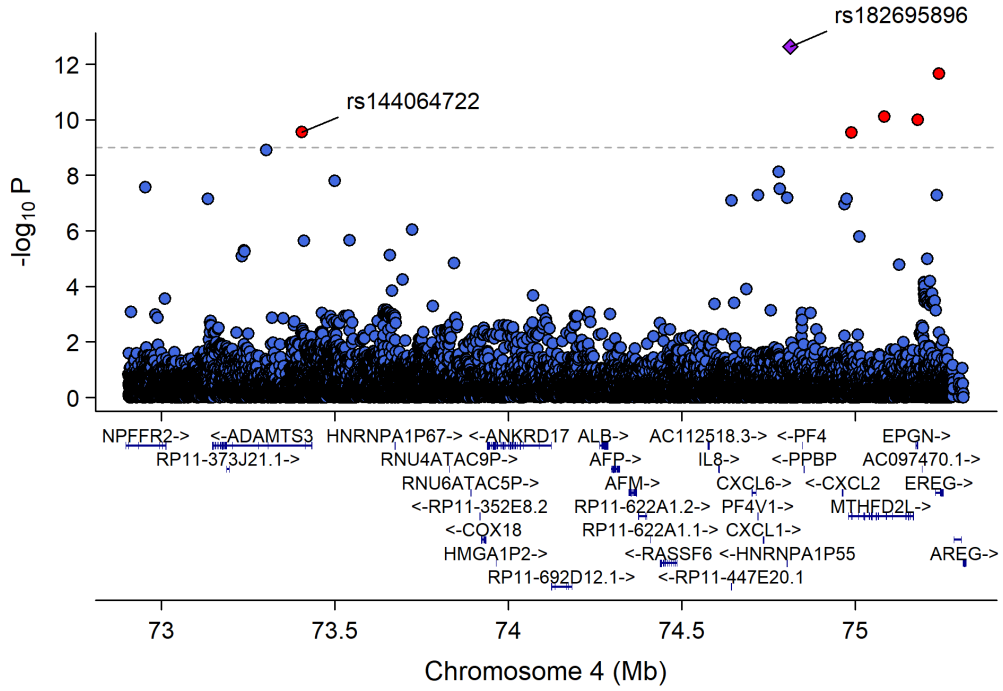
Chr2_19947287_22427492



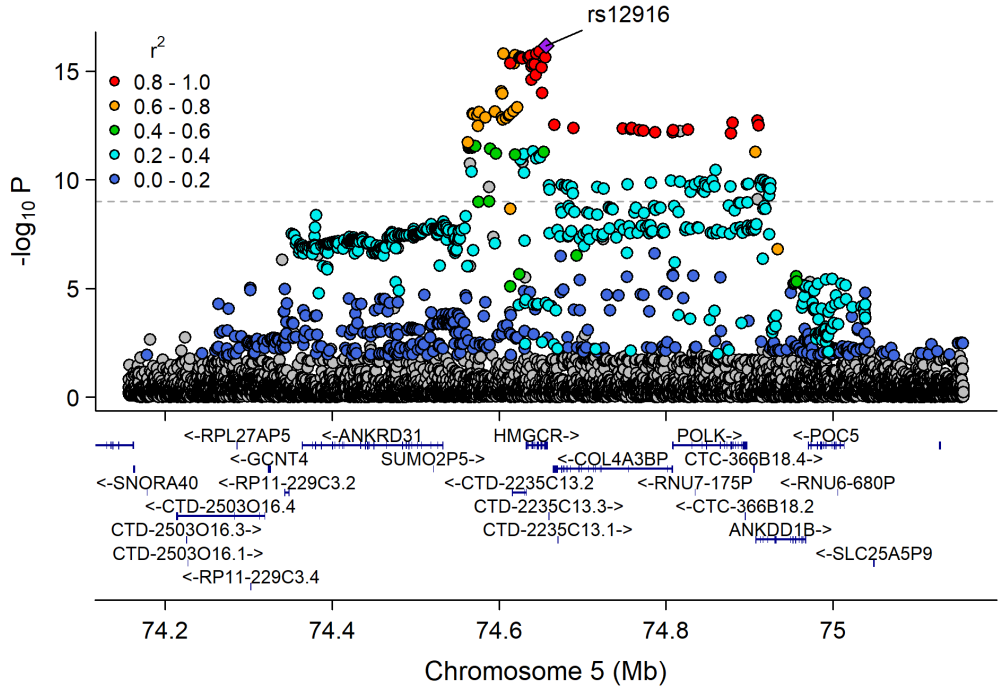
Chr2_43065153_45056024



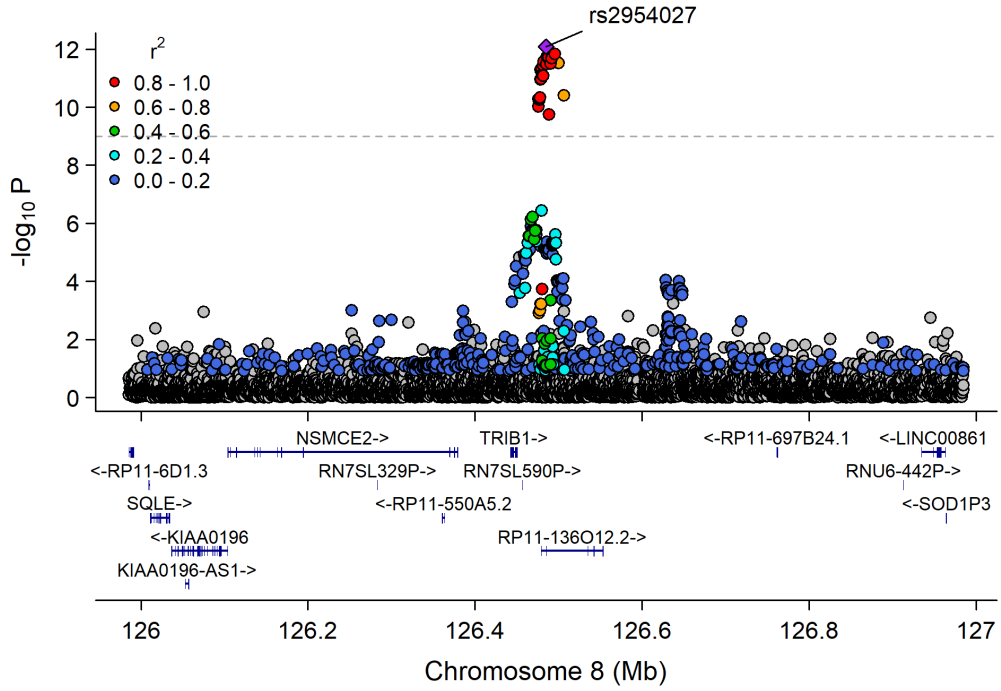
Chr4__71881824__76681229



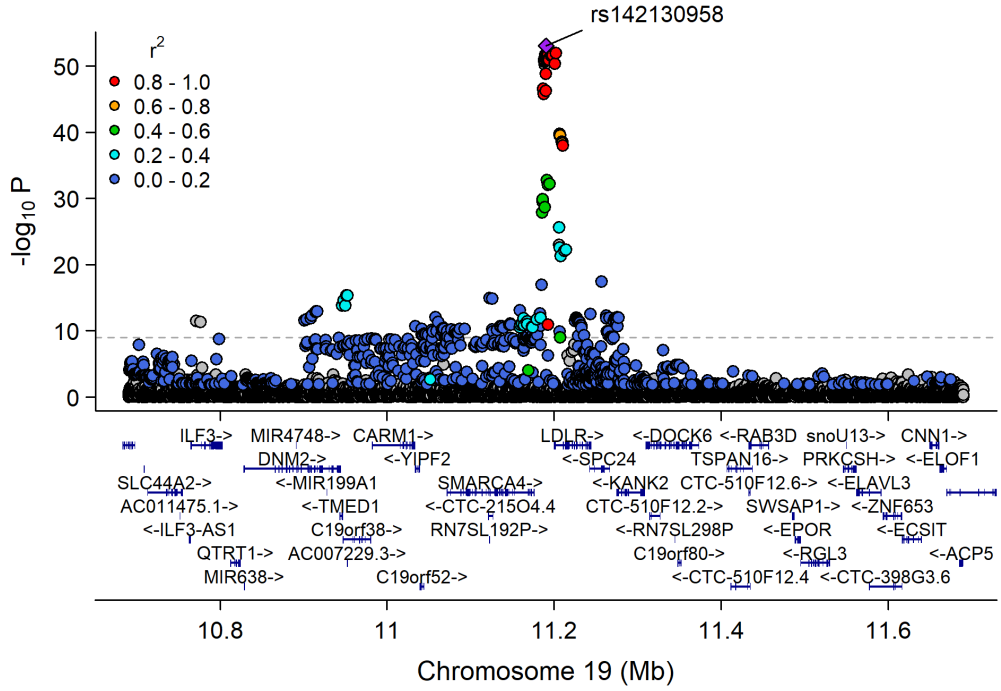
Chr5__74242002__75221582



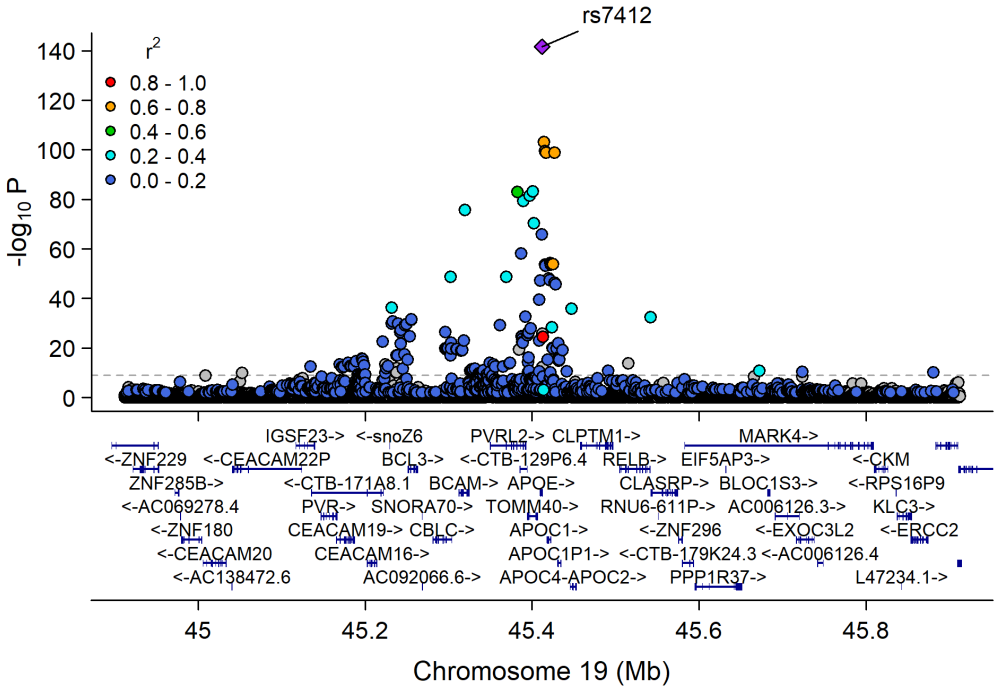
Chr8_126435663_126578061



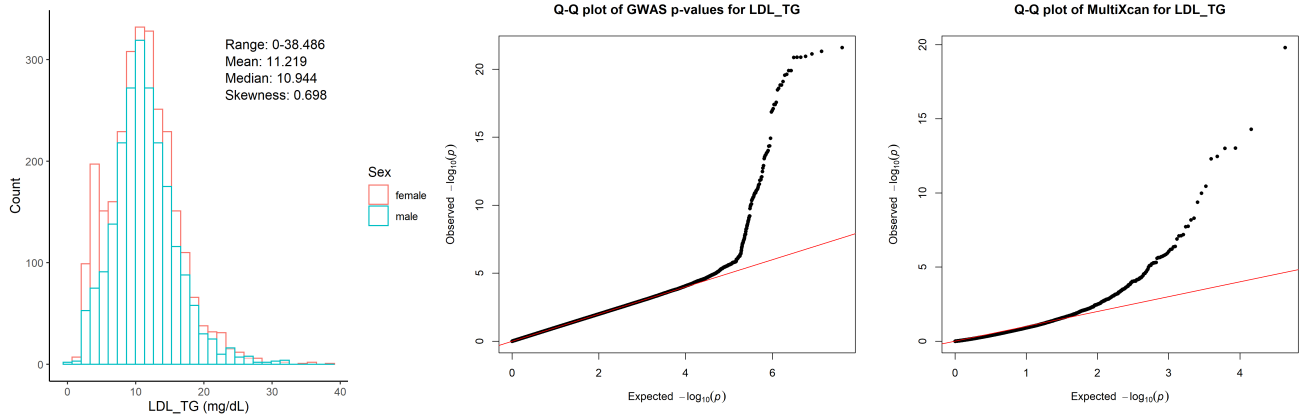
Chr19_9807999_12255594

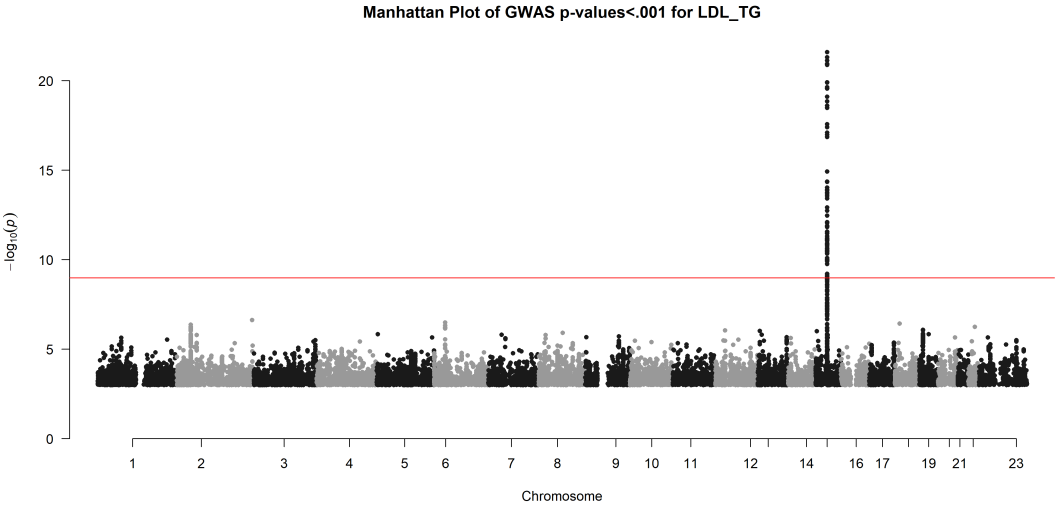


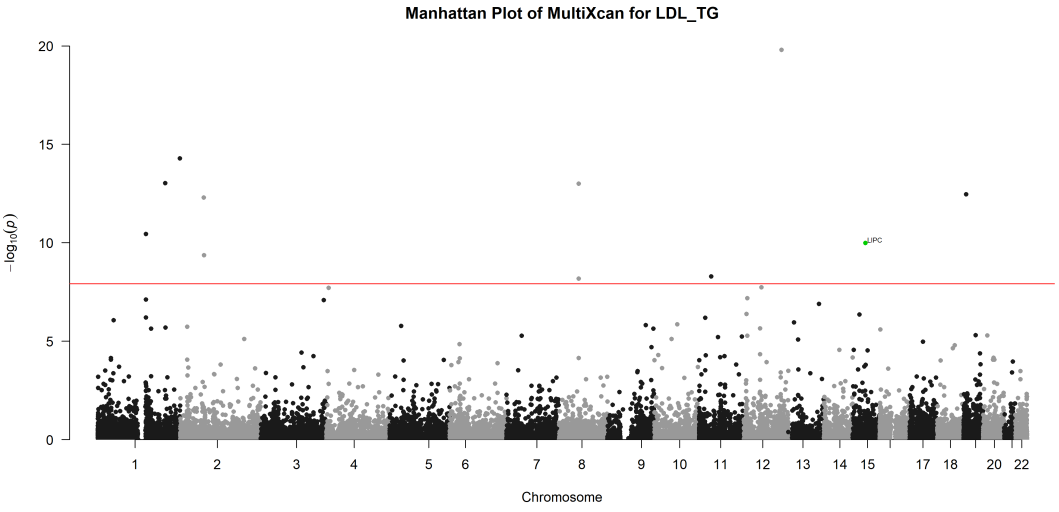
Chr19__44063363__46637375



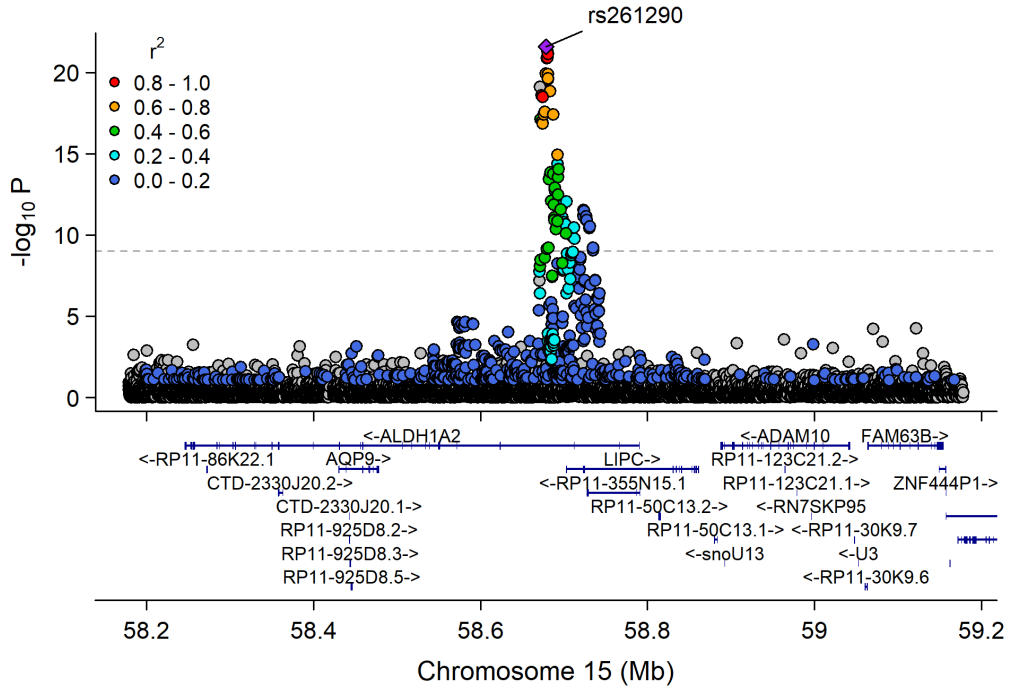
Triglycerides in LDL (mg/dL)



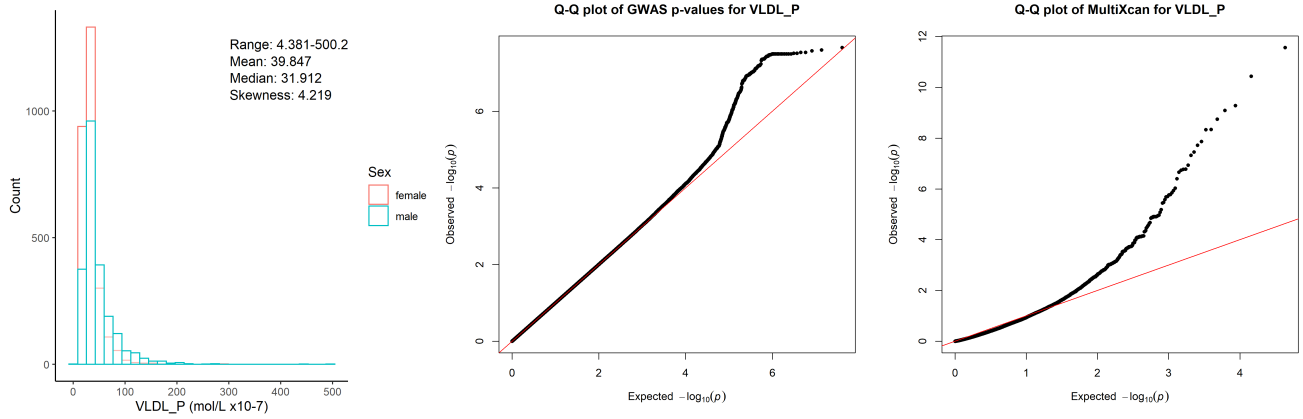


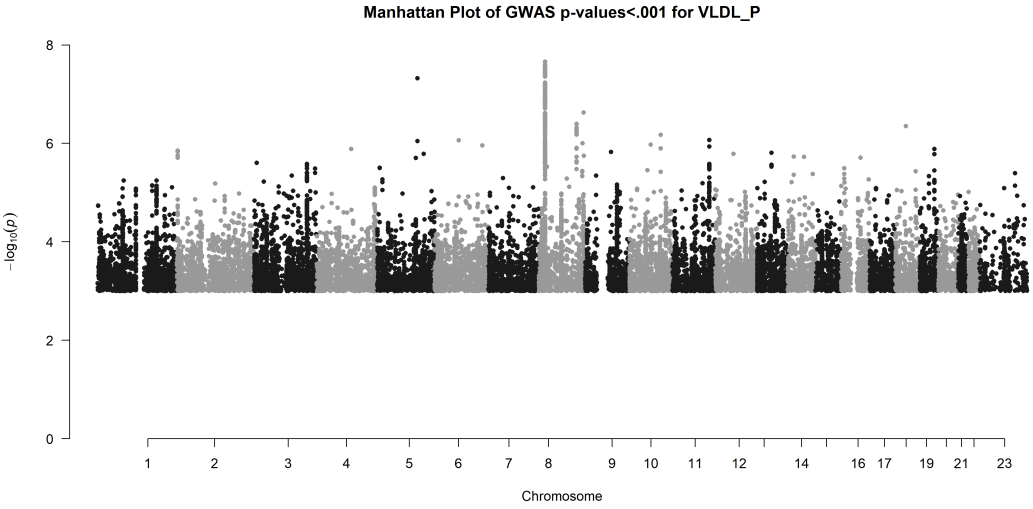


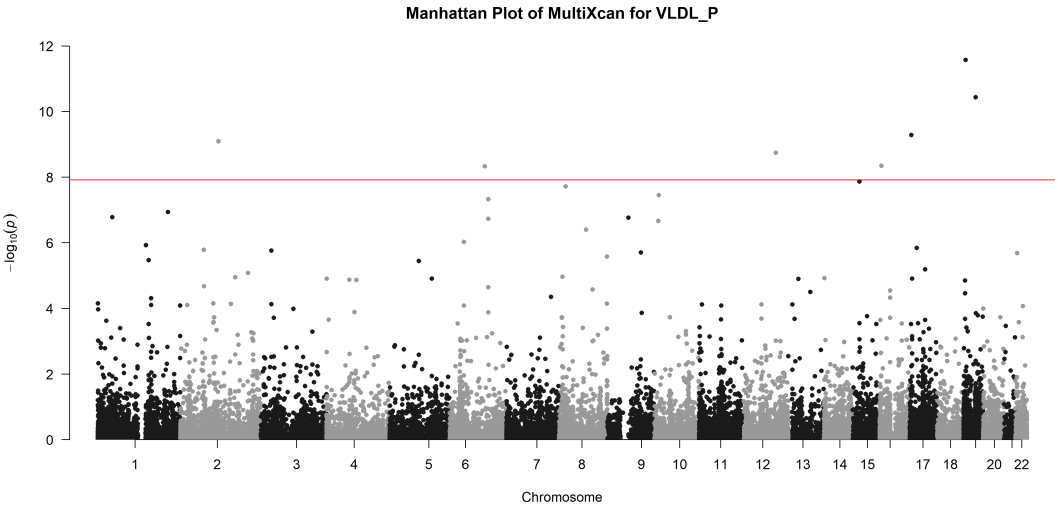
Chr15_57658798_60490883



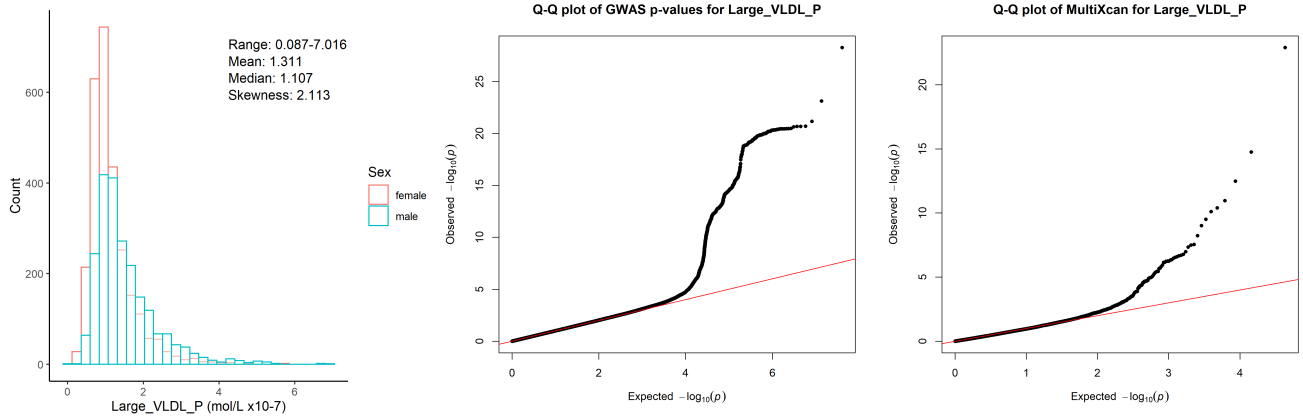
Concentration of VLDL particles (mol/l)

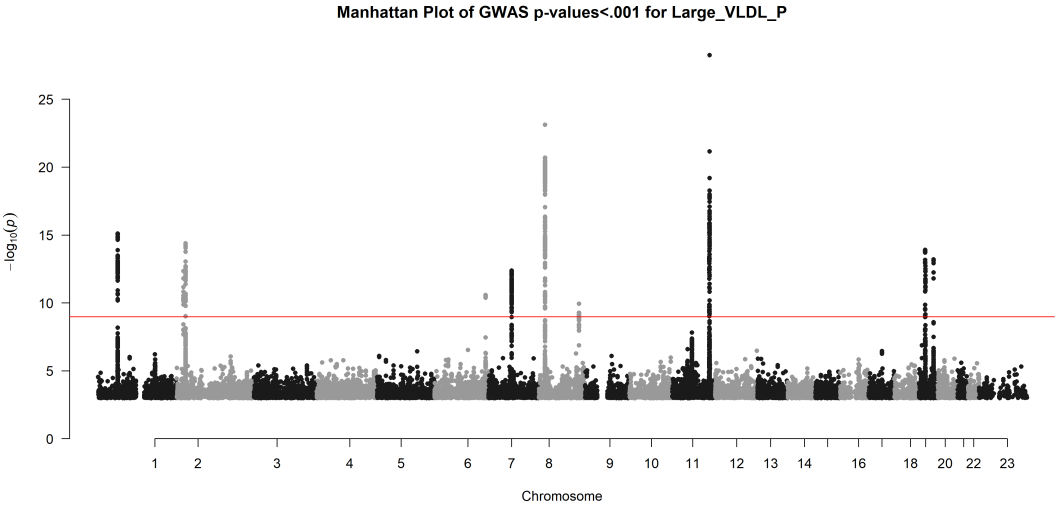


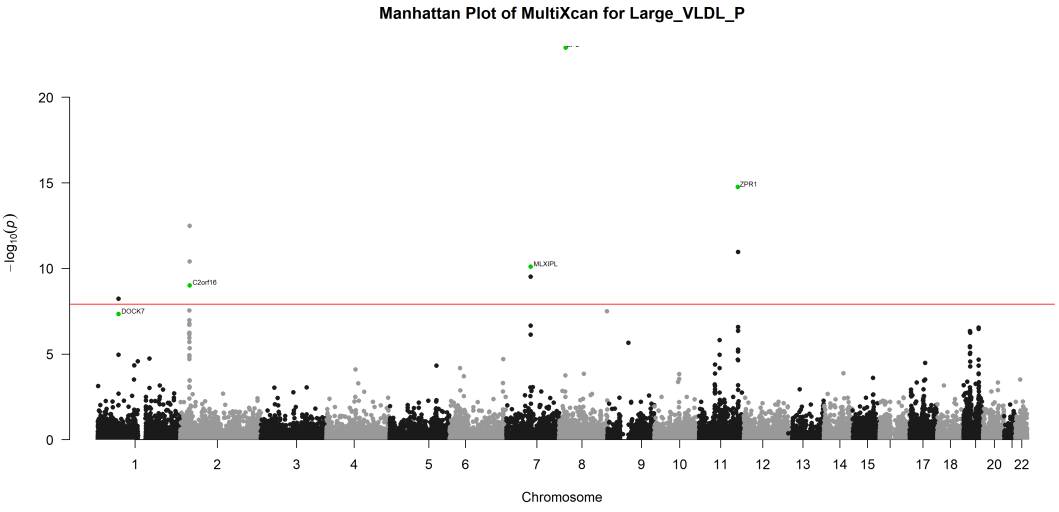




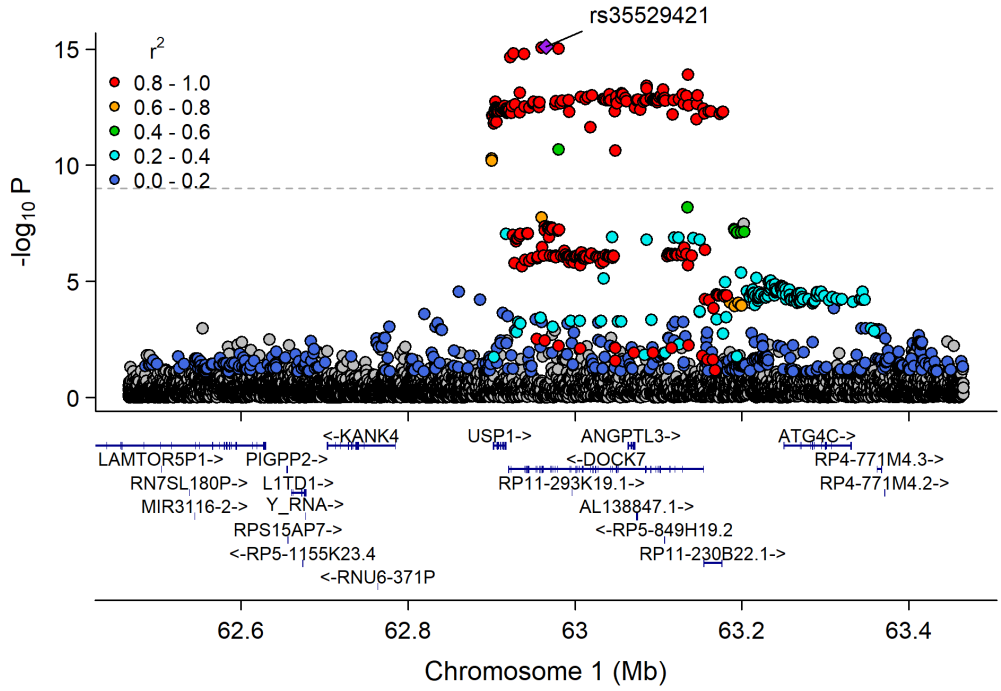
Concentration of large VLDL particles (mol/l)



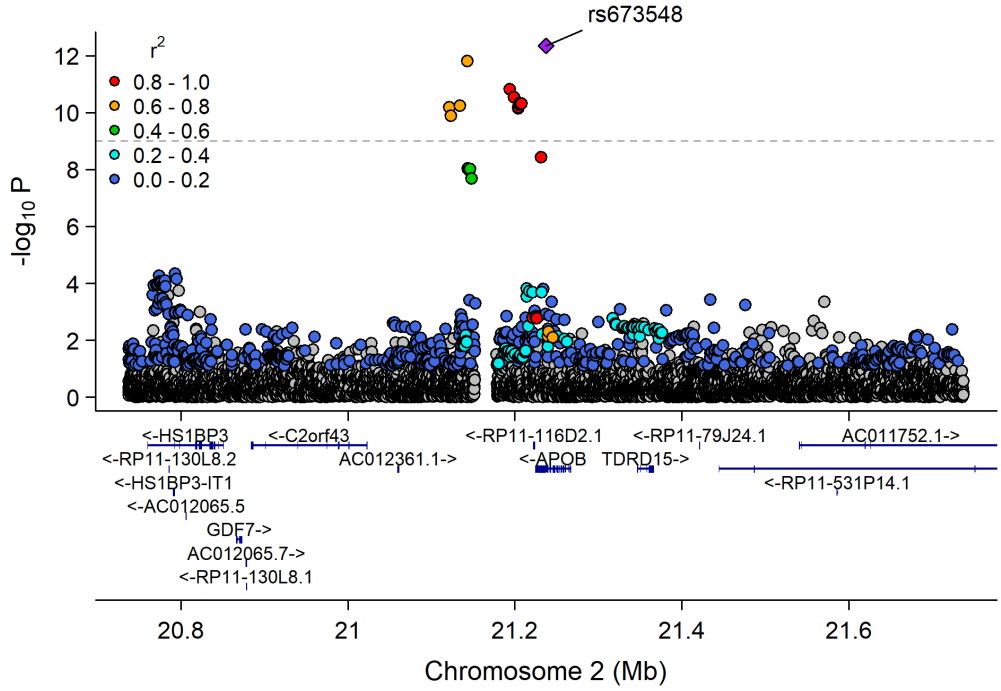




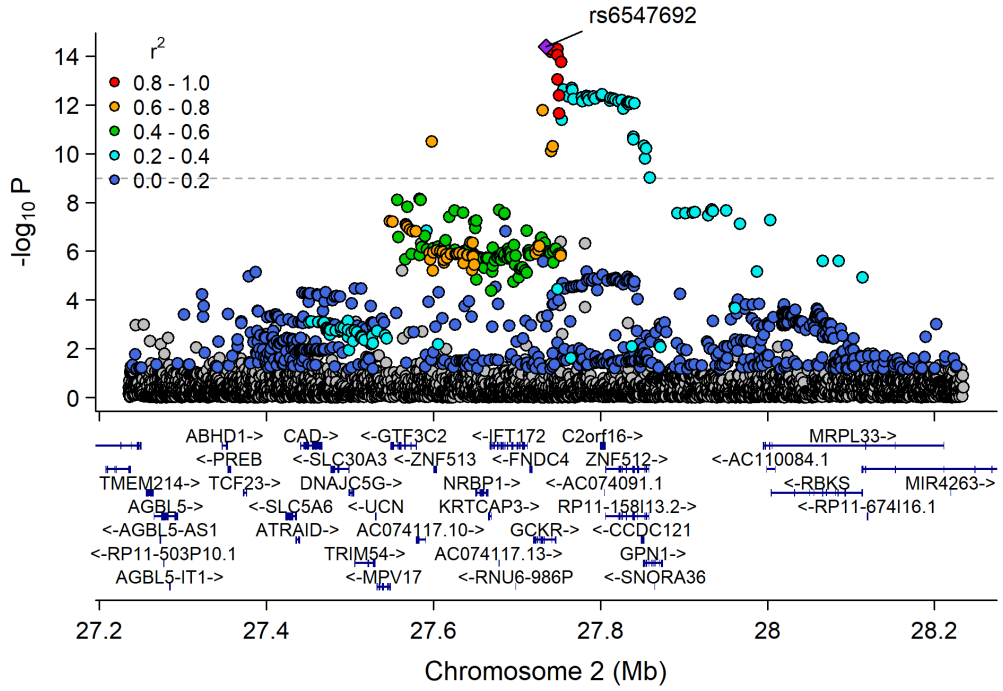
Chr1_62833402_63375116



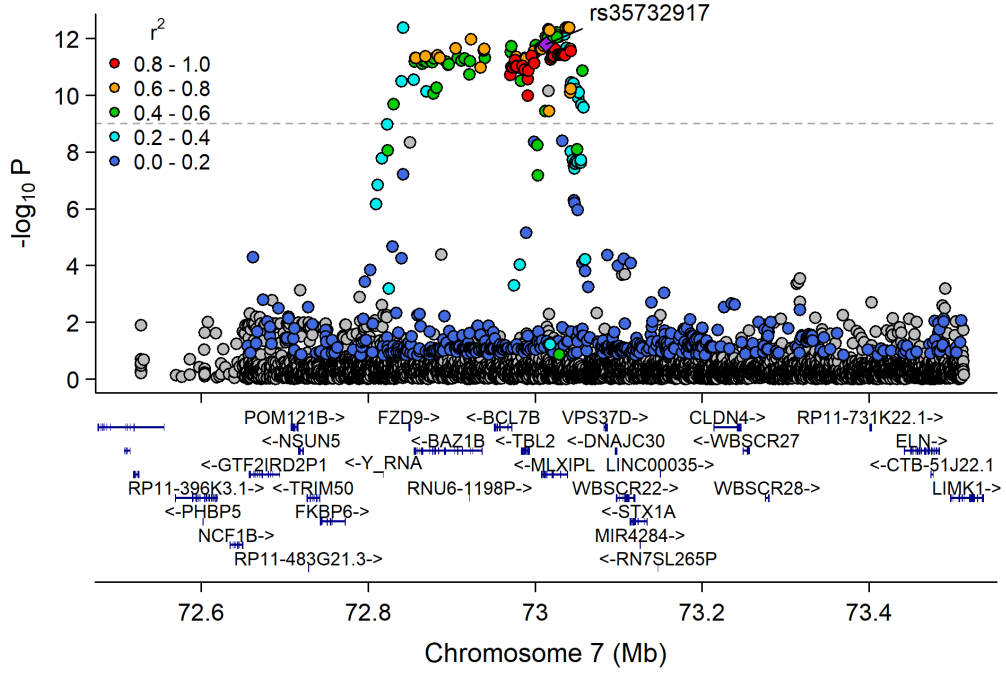
Chr2_19947287_22427492



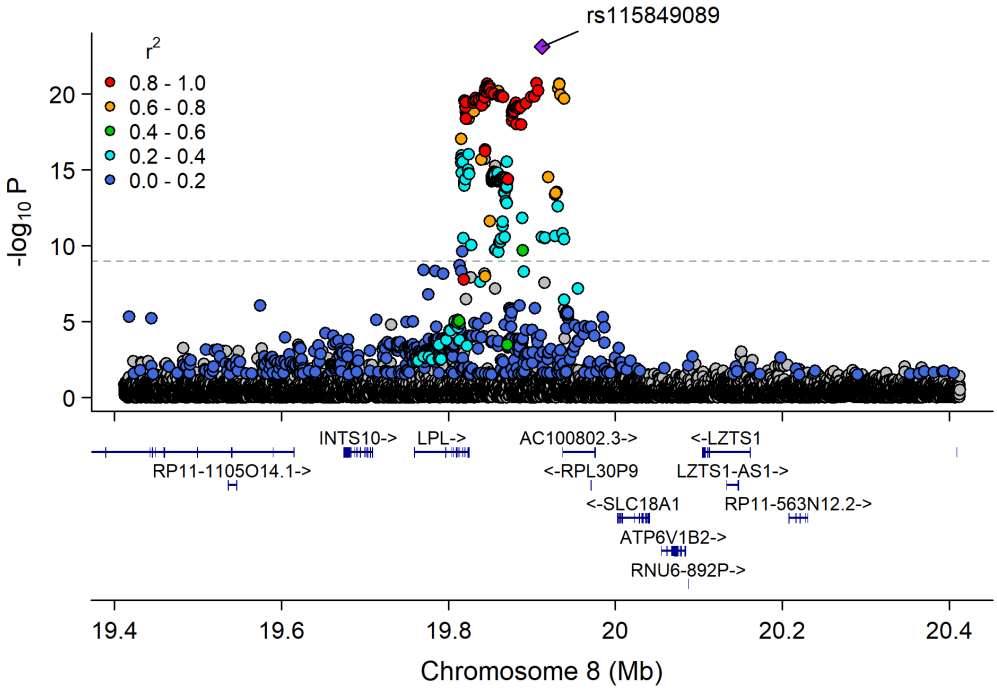
Chr2_26753815_28597624



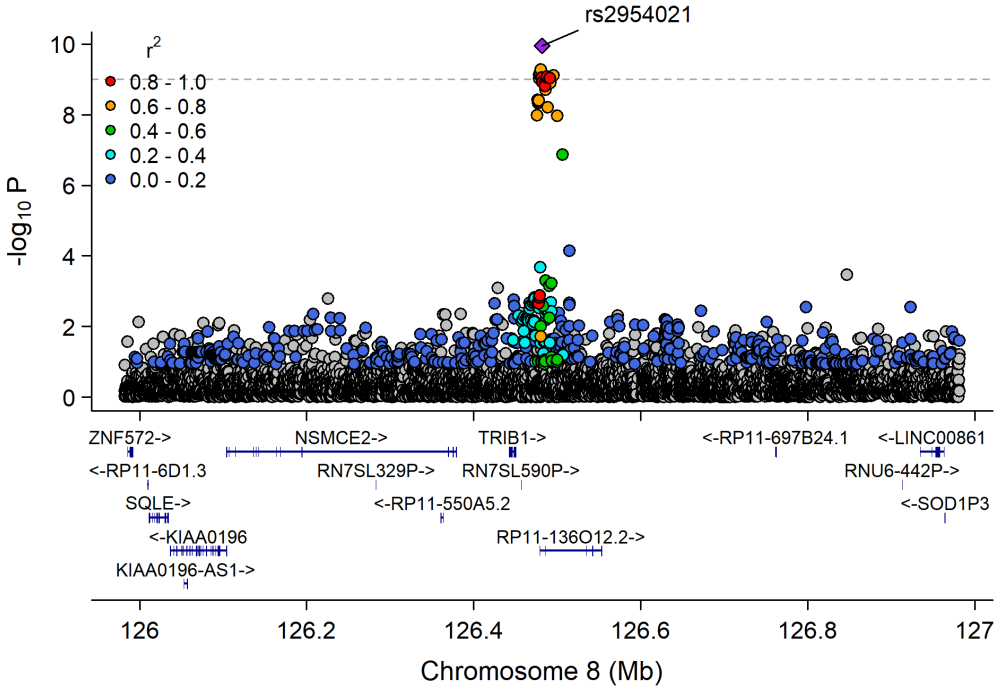
Chr7_72070325_73133300



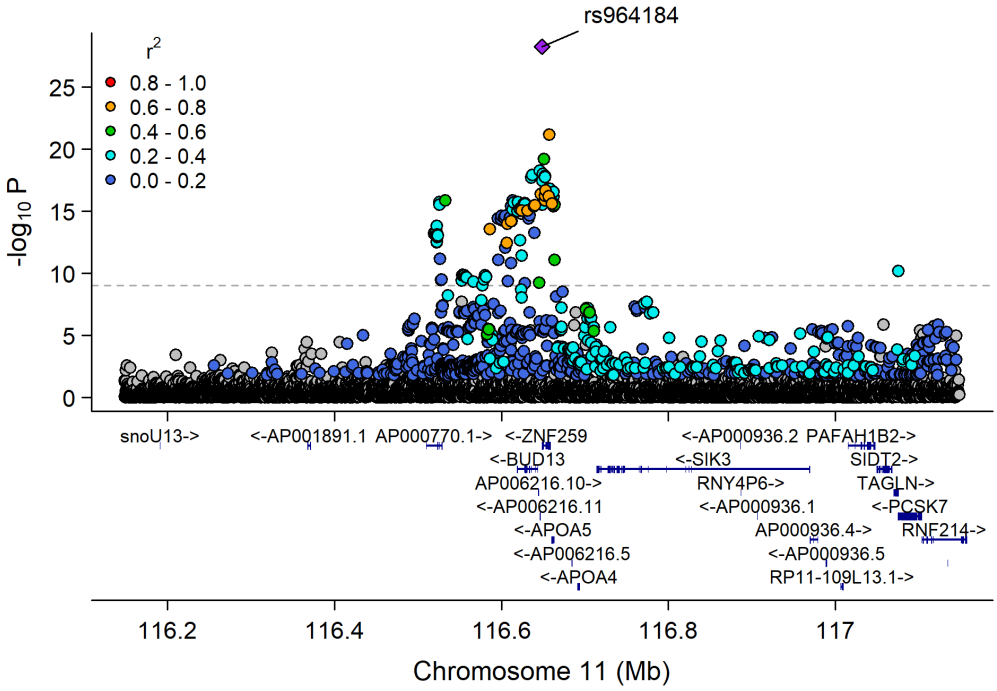
Chr8__18582620__20792744



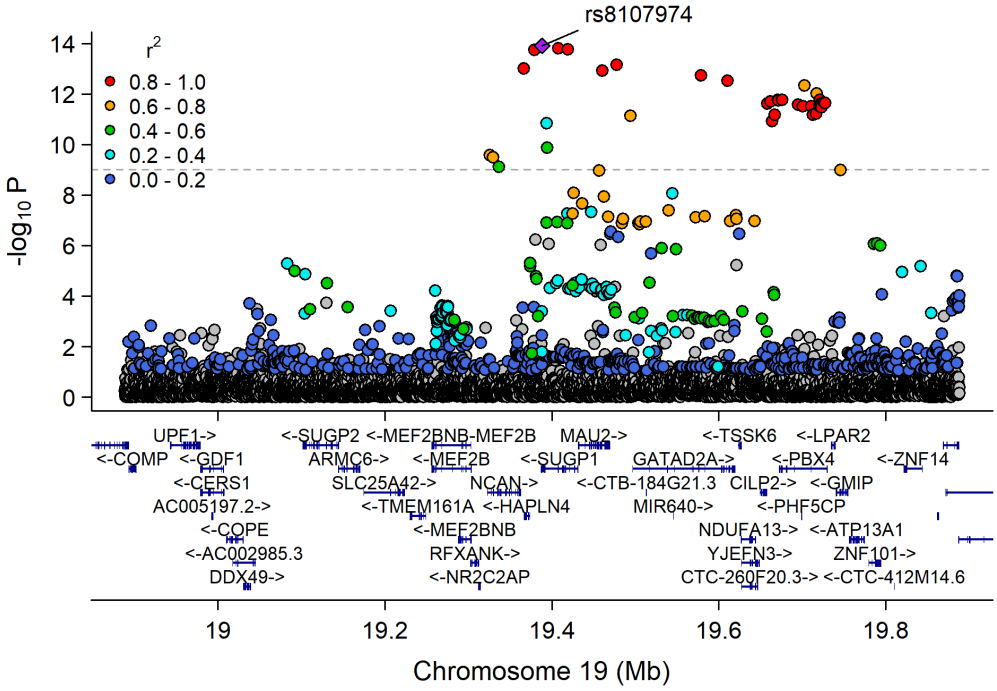
Chr8__126435663__126578061



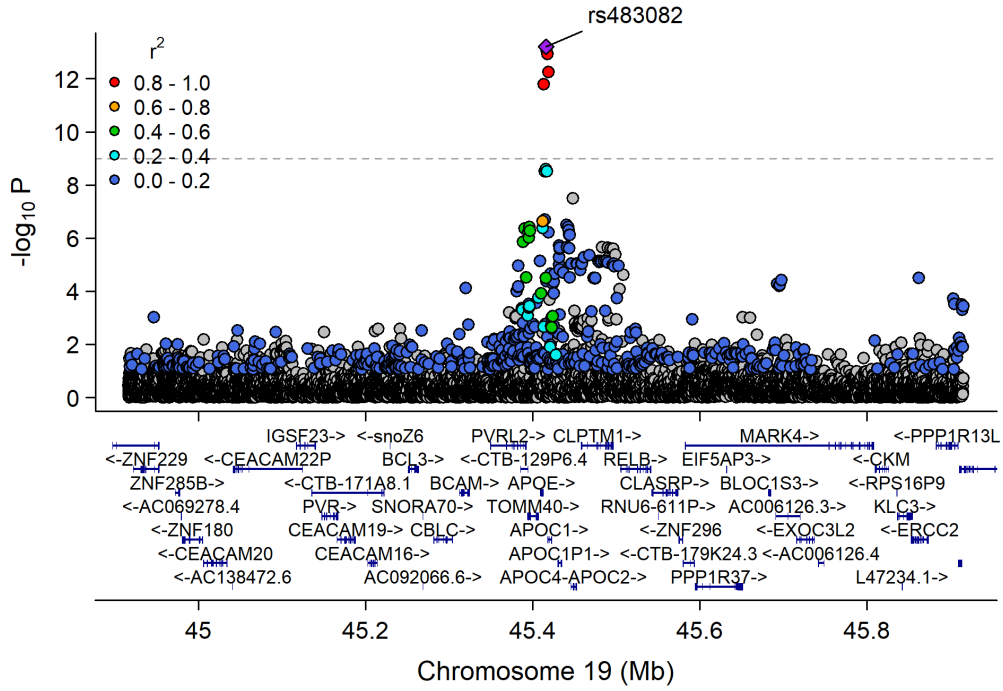
Chr11_115541901_117633315



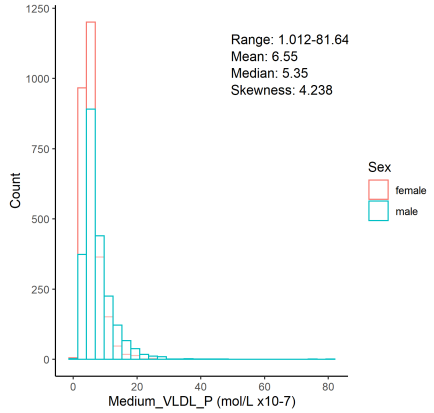
Chr19_18370495_20841464

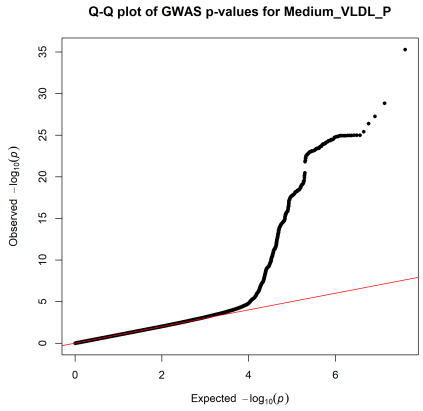


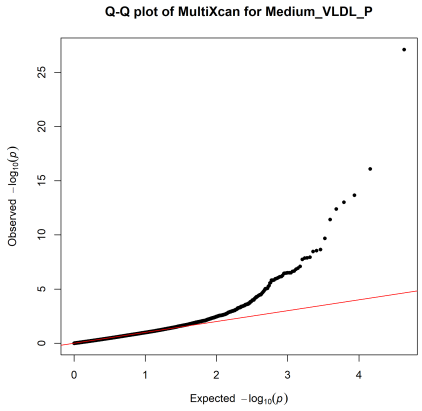
Chr19__44063363__46637375

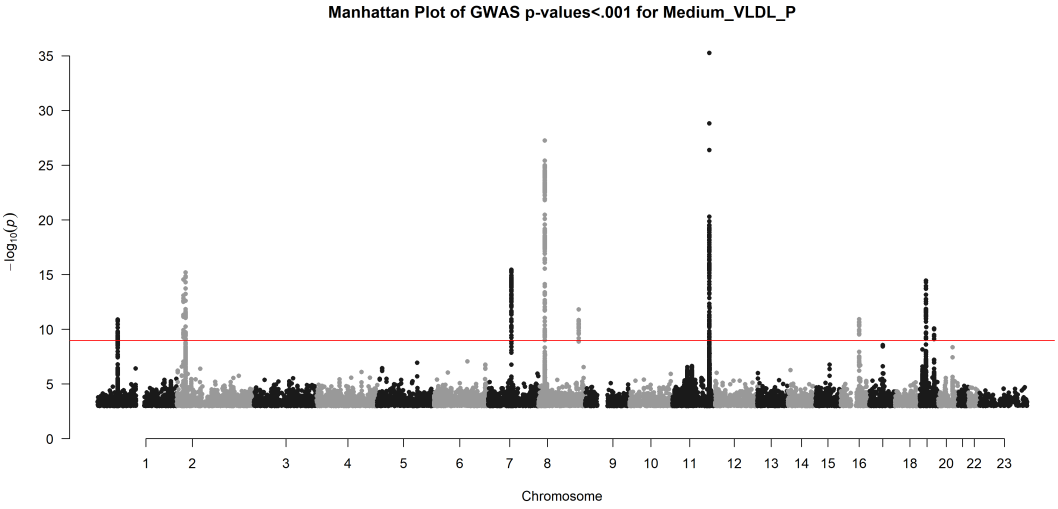


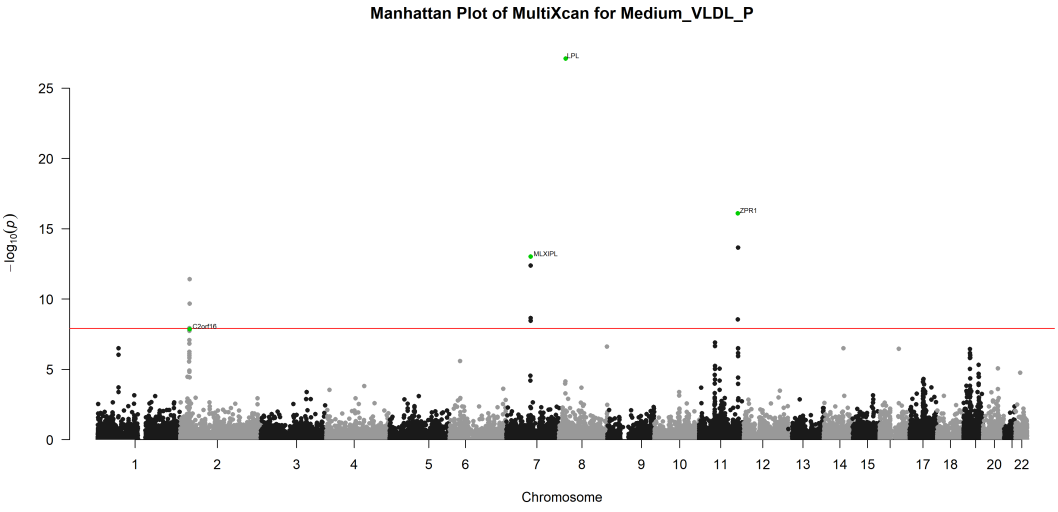
Concentration of medium VLDL particles (mol/l)



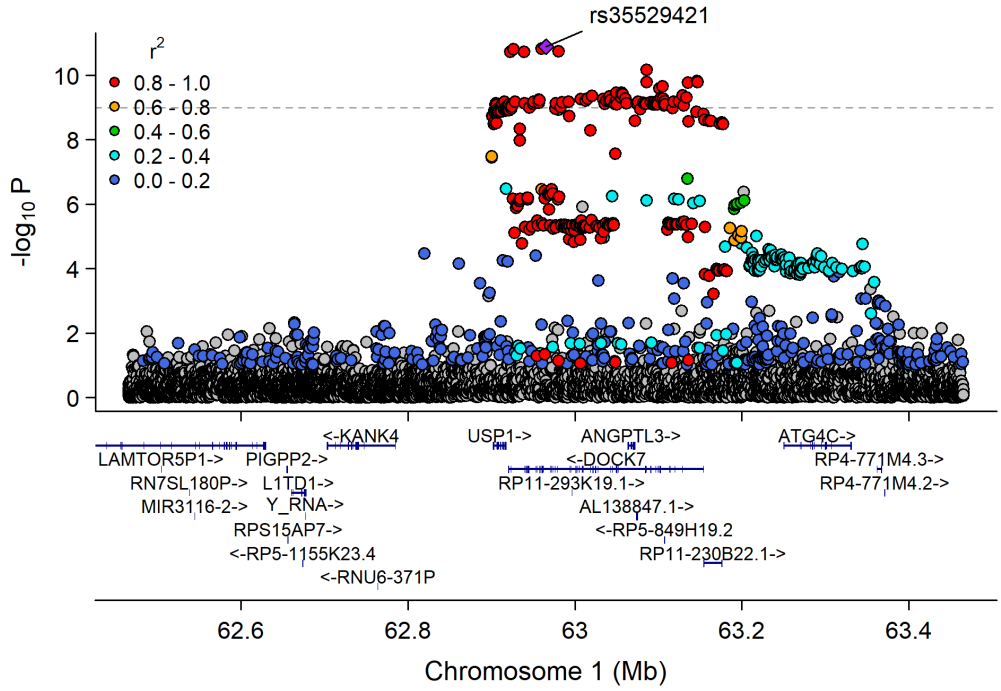




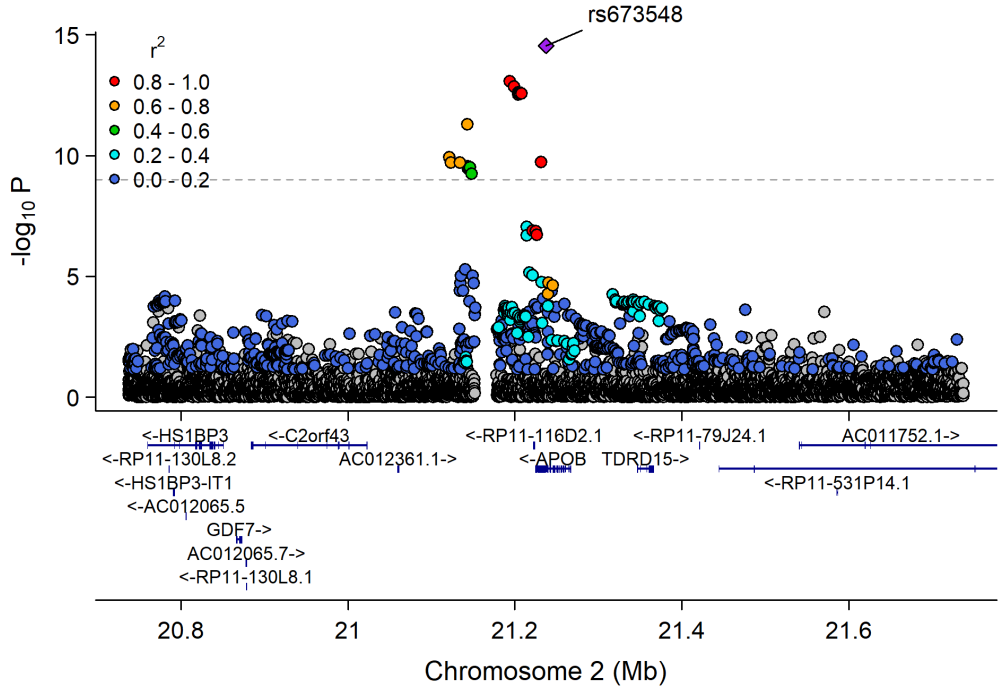




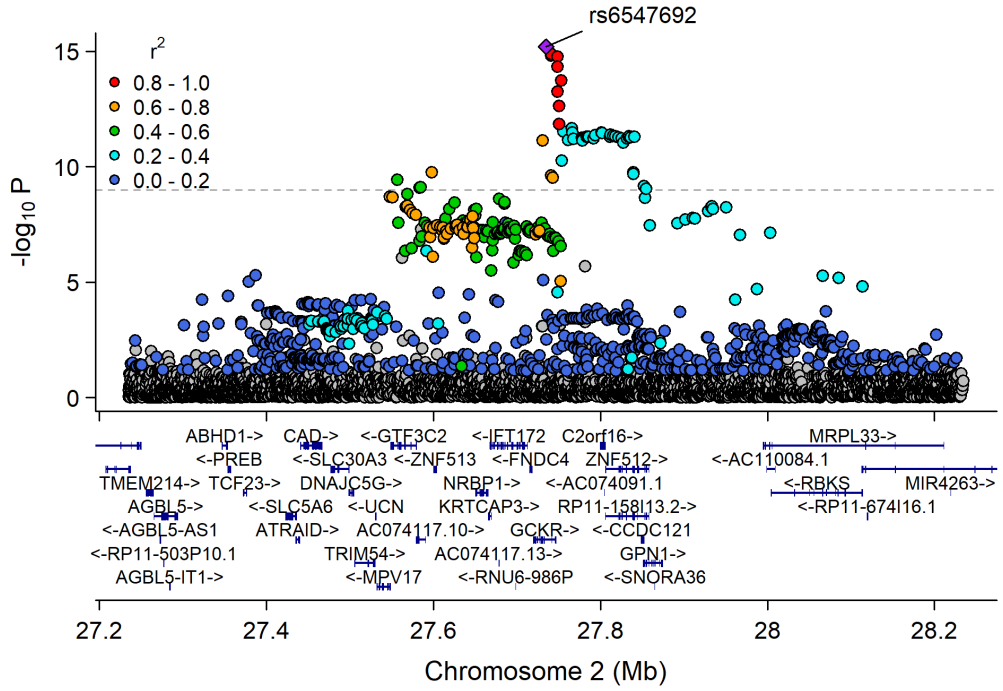
Chr1_62833402_63375116



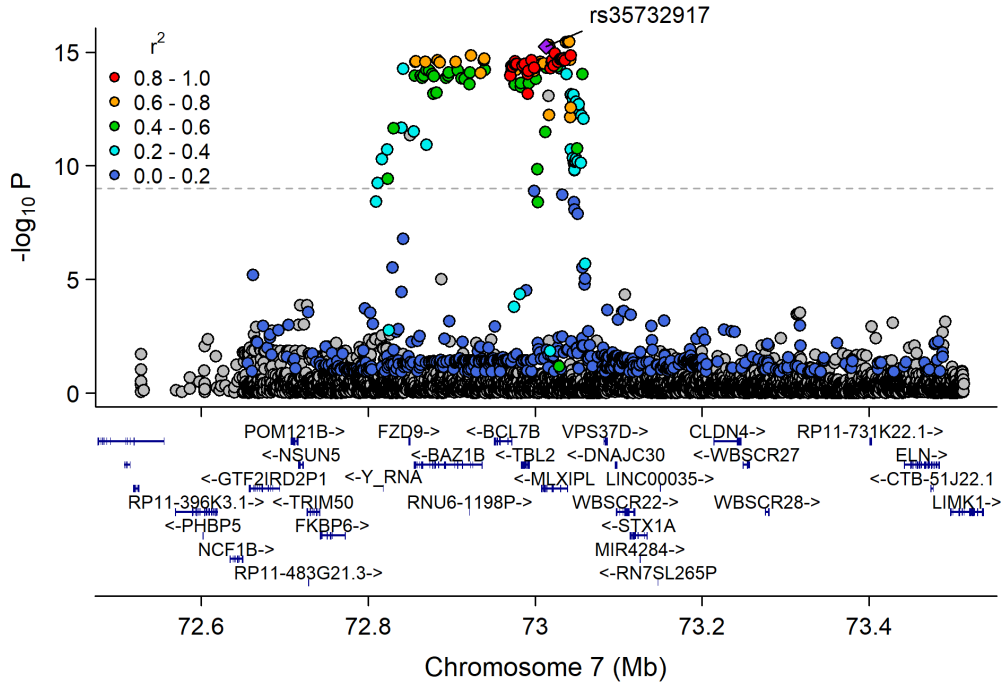
Chr2_19947287_22427492



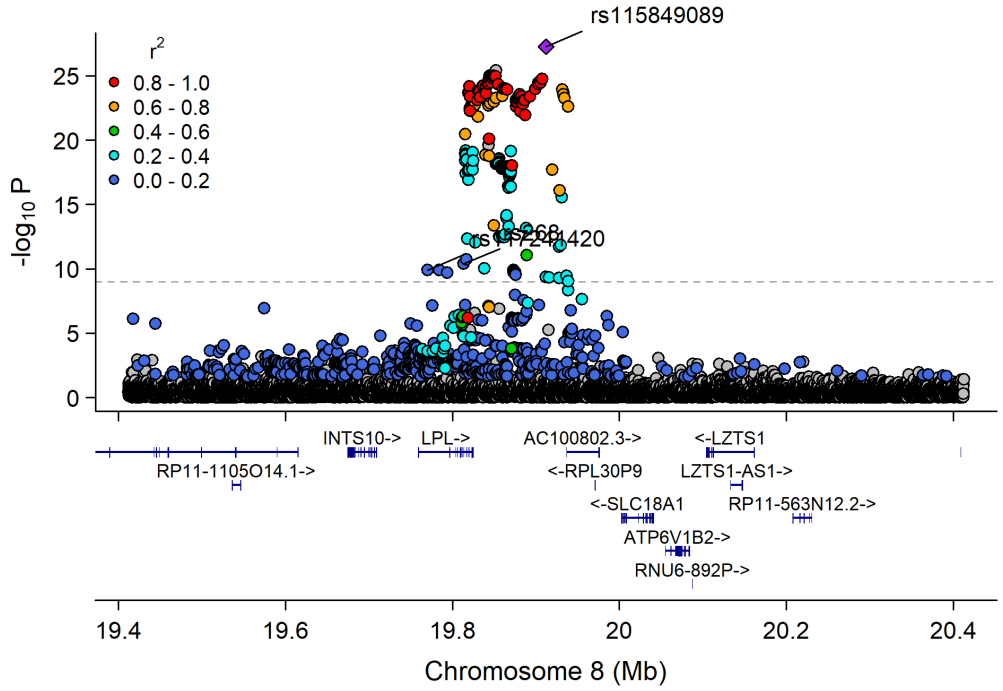
Chr2_26753815_28597624



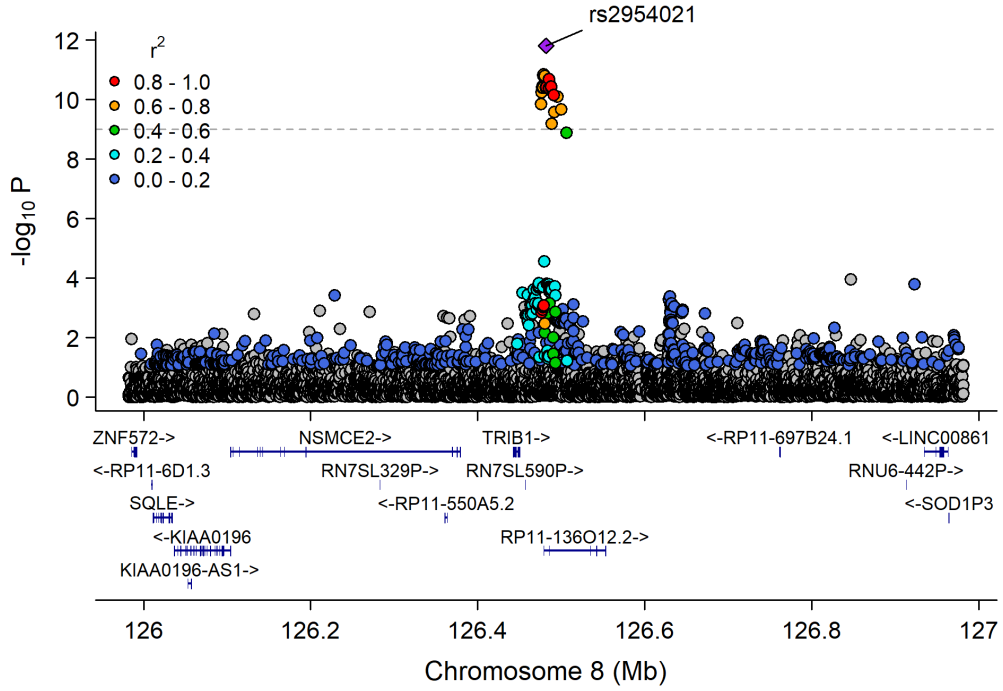
Chr7_72070325_73133300



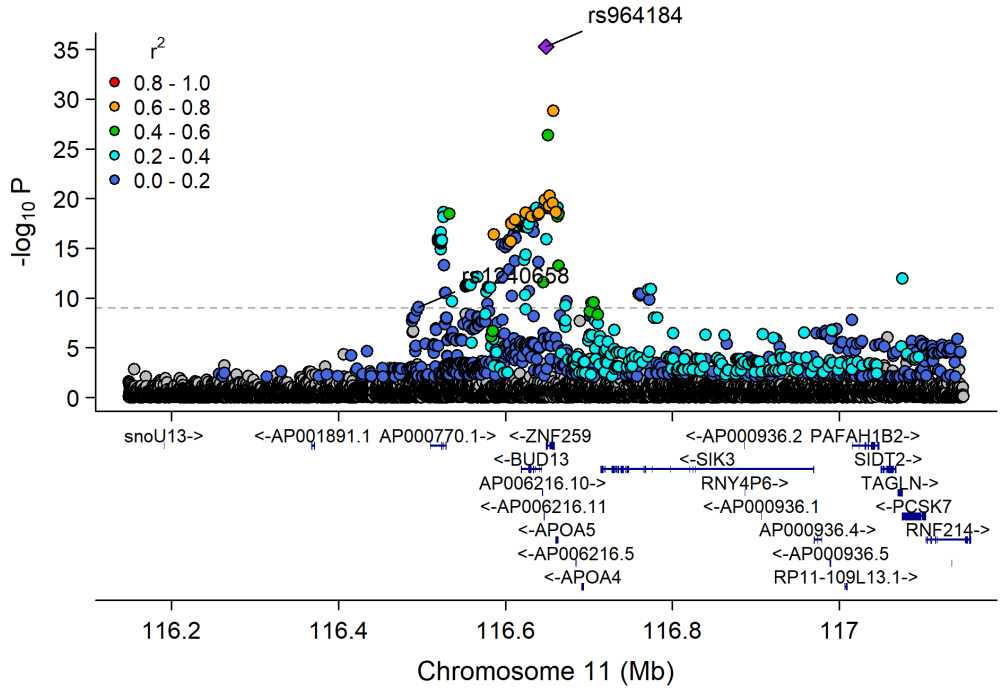
Chr8__18582620__20792744



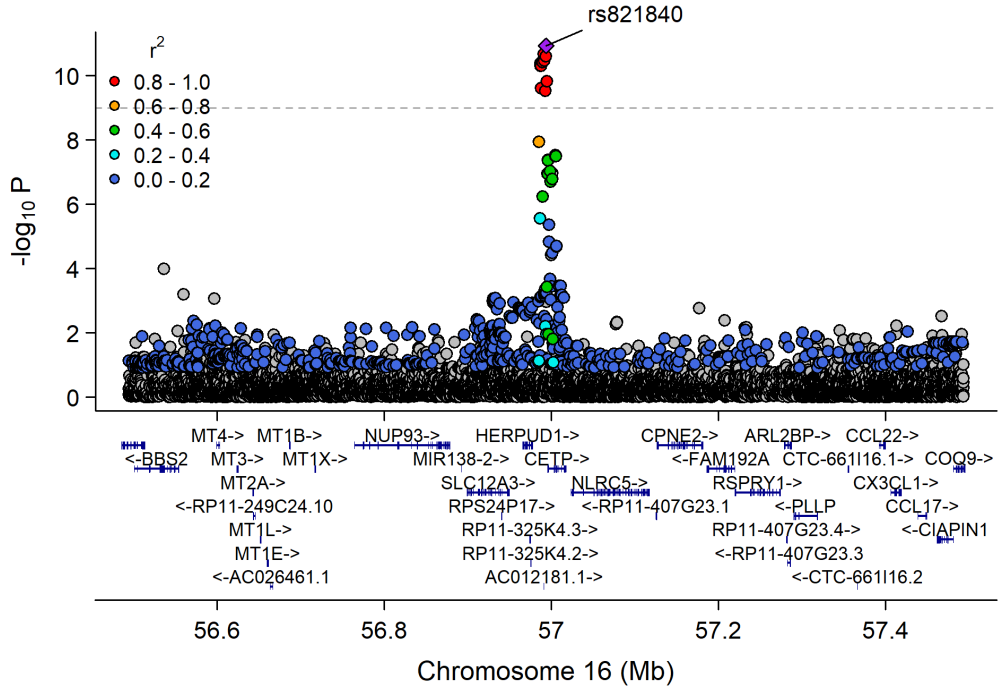
Chr8__126435663__126578061



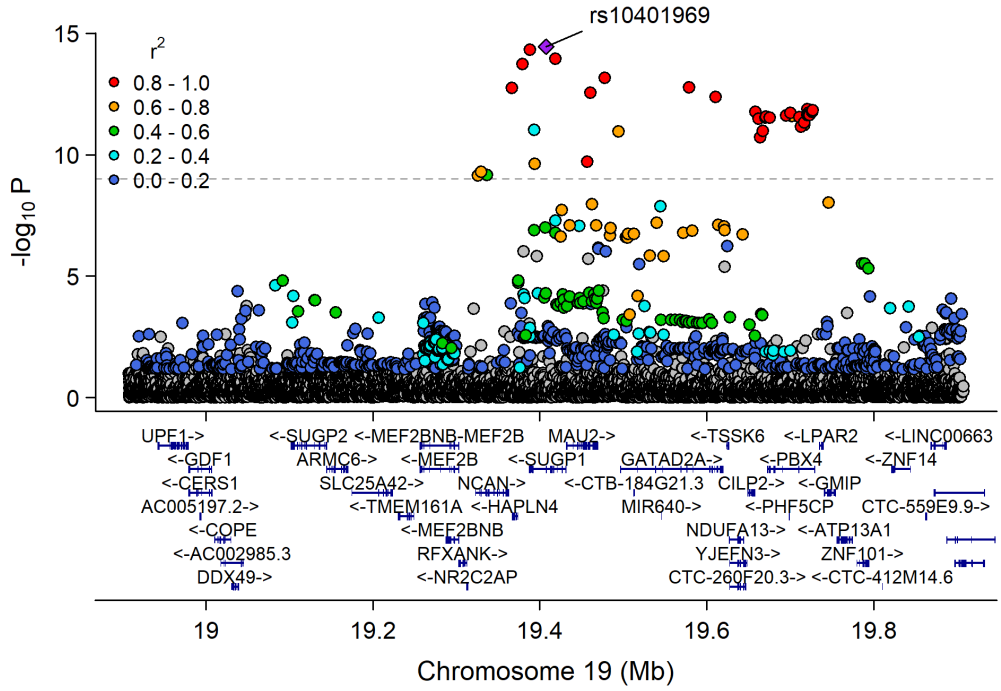
Chr11_115541901_117633315



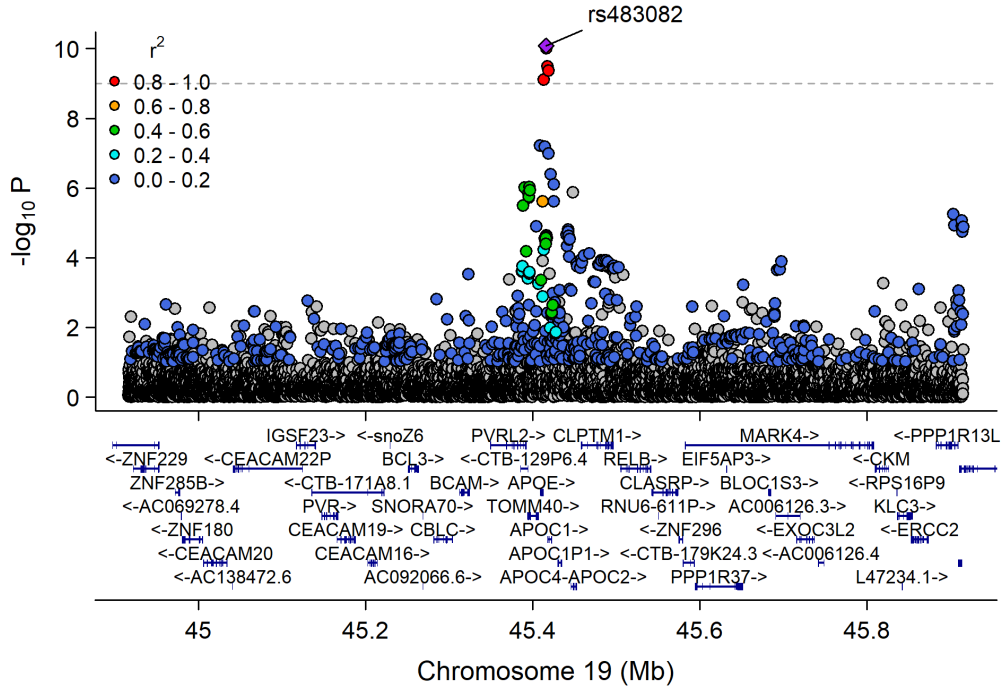
Chr16_55870822_57992421



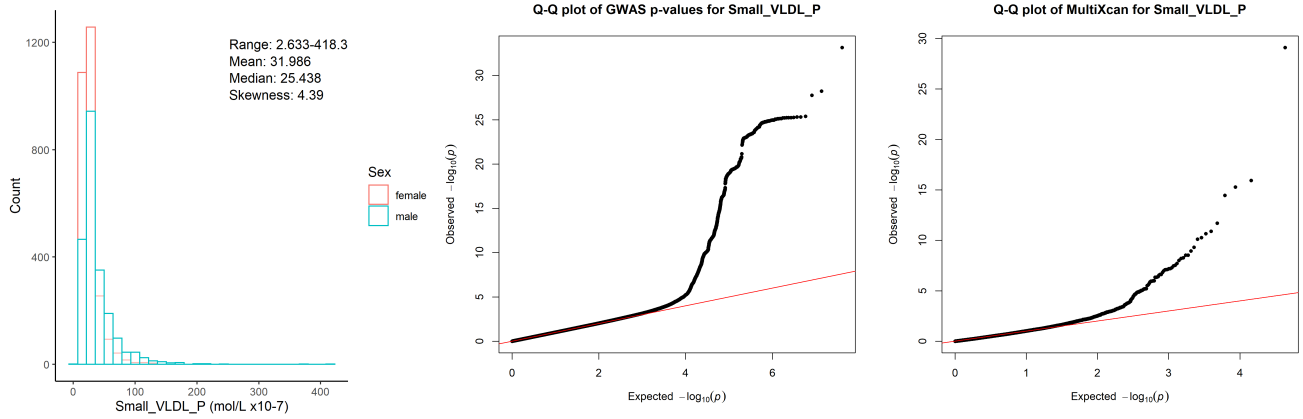
Chr19_18370495_20841464

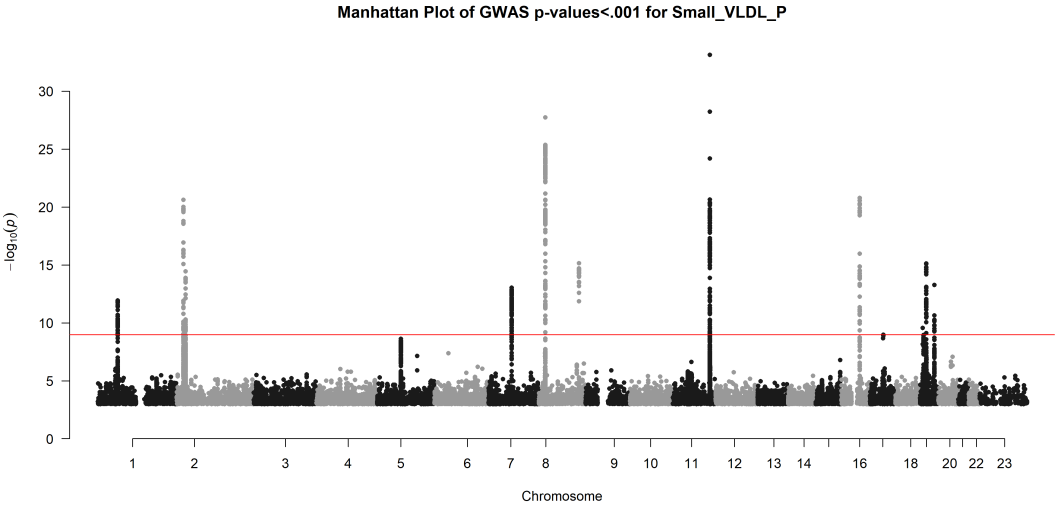


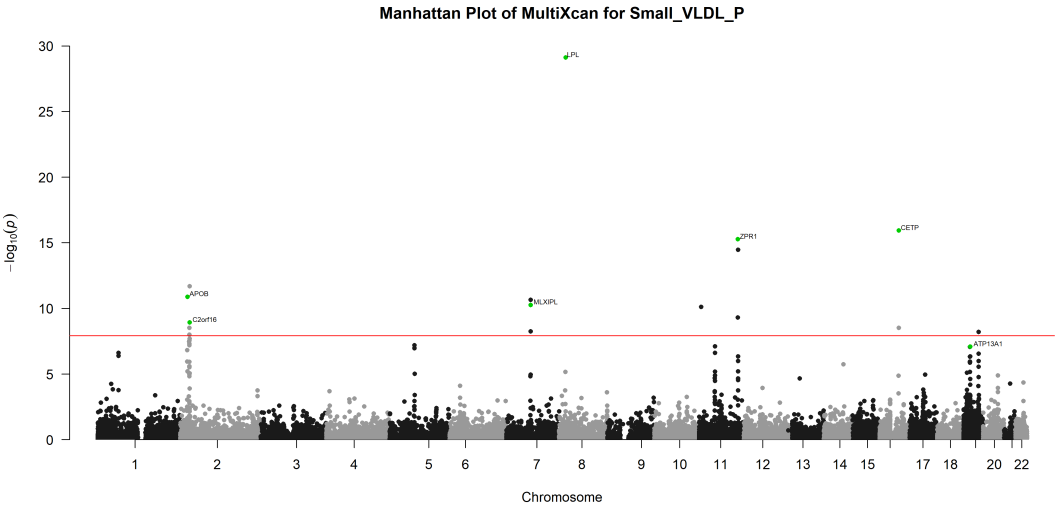
Chr19_44063363_46637375



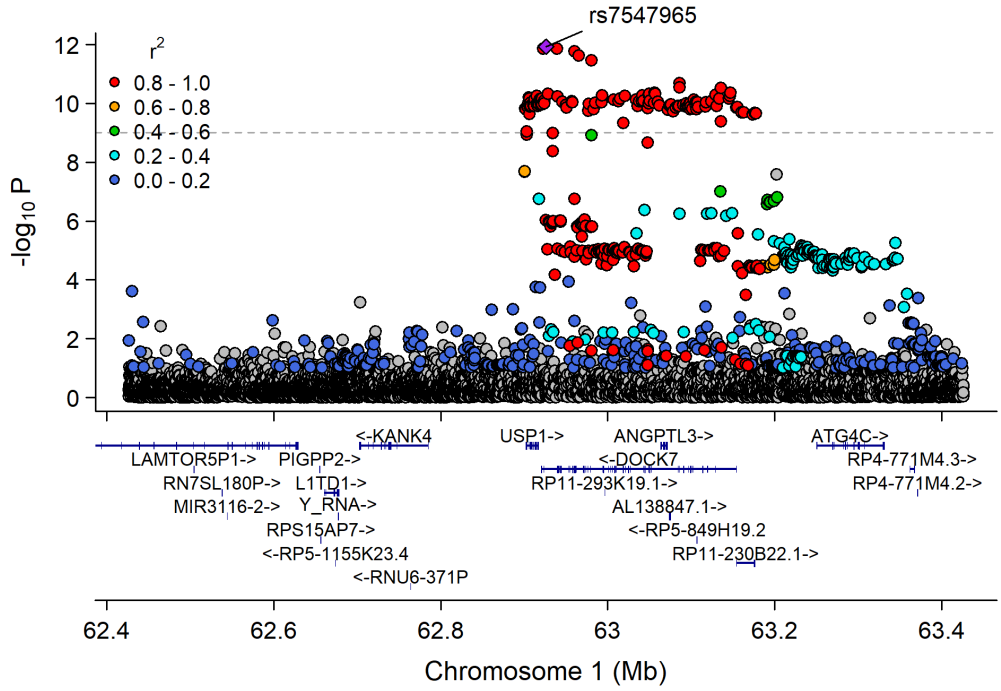
Concentration of small VLDL particles (mol/l)



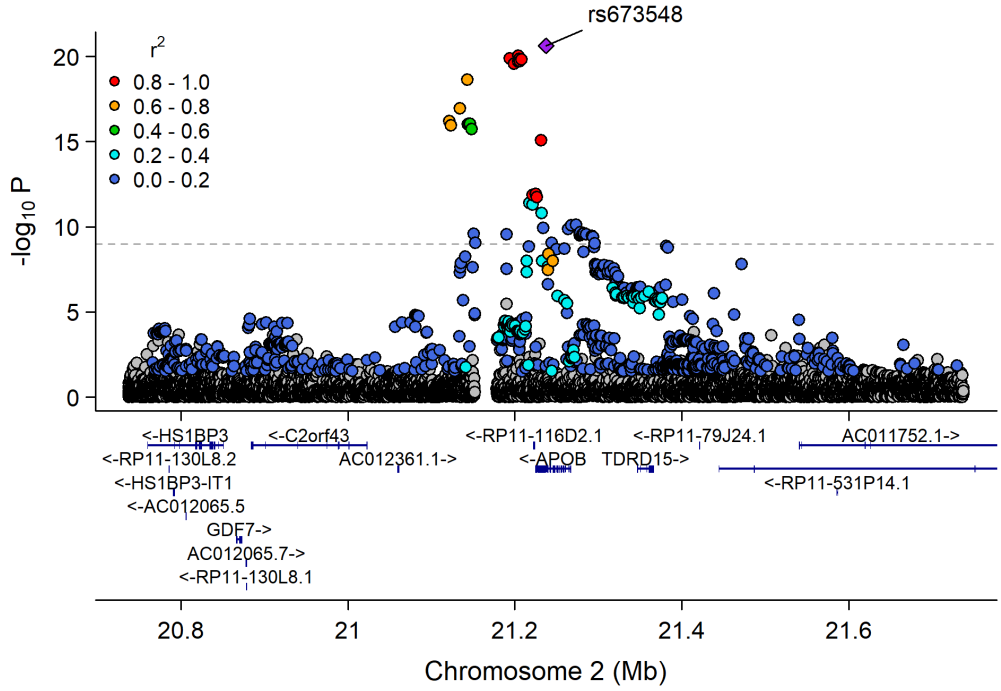




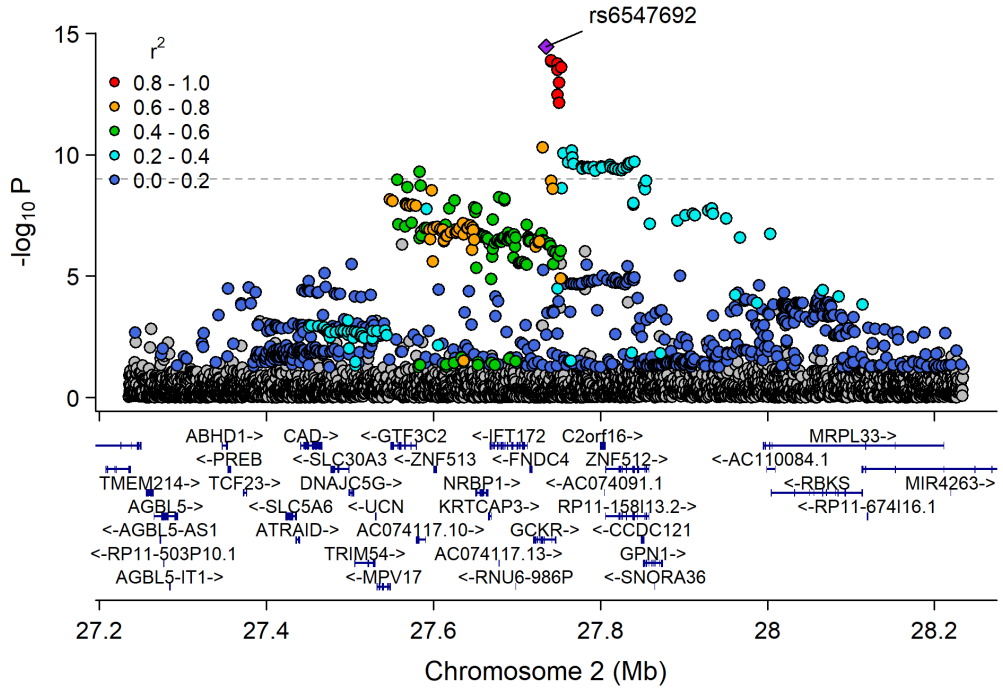
Chr1_62833402_63375116



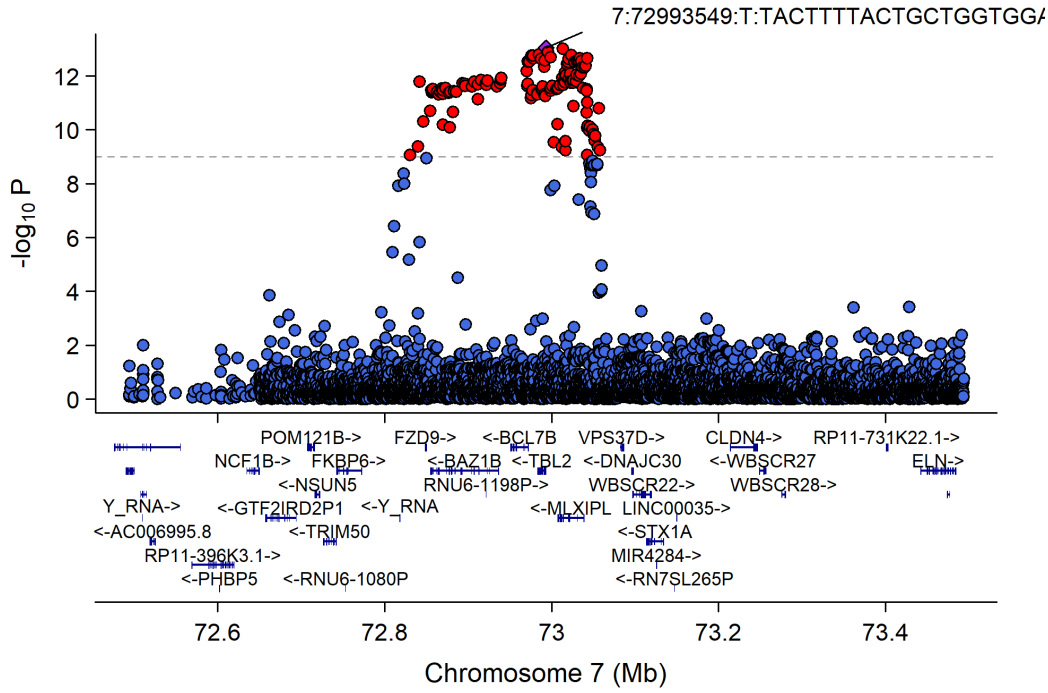
Chr2_19947287_22427492



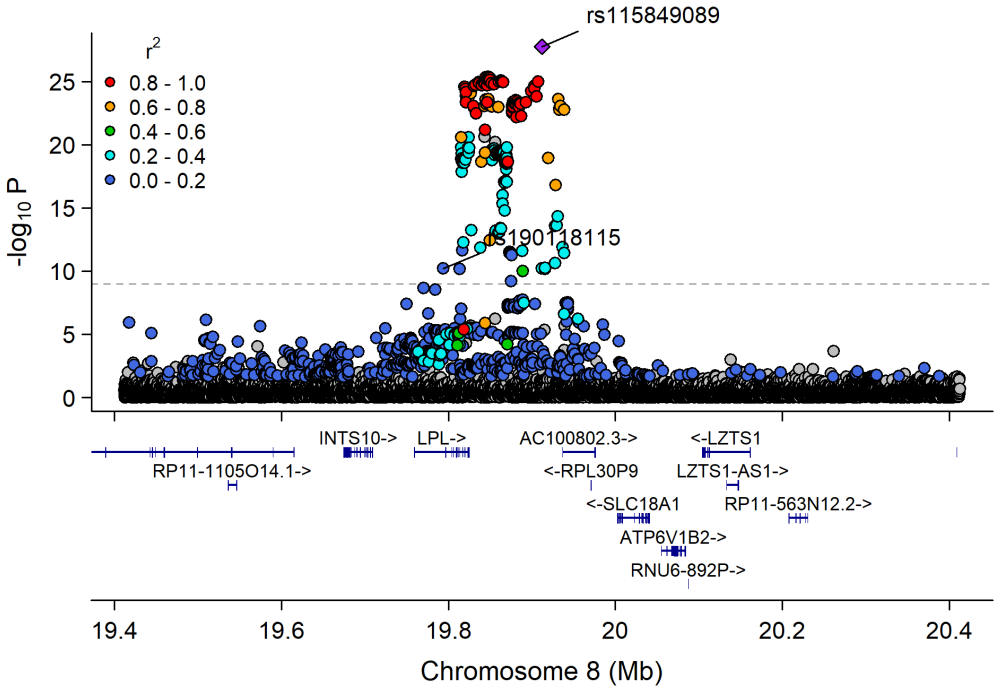
Chr2_26753815_28597624



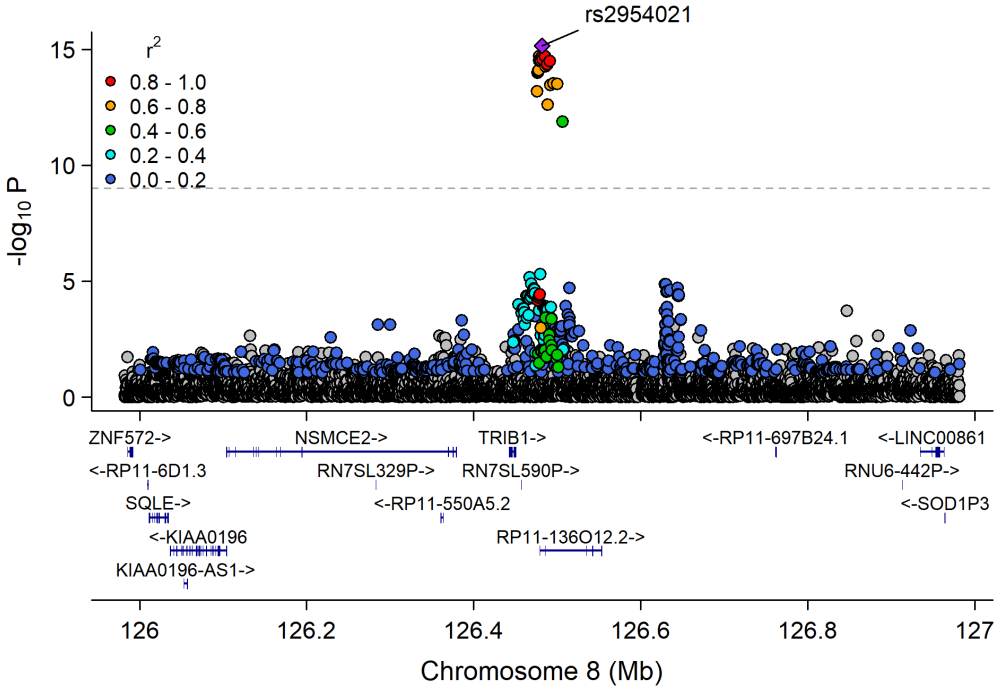
Chr7_72070325_73133300



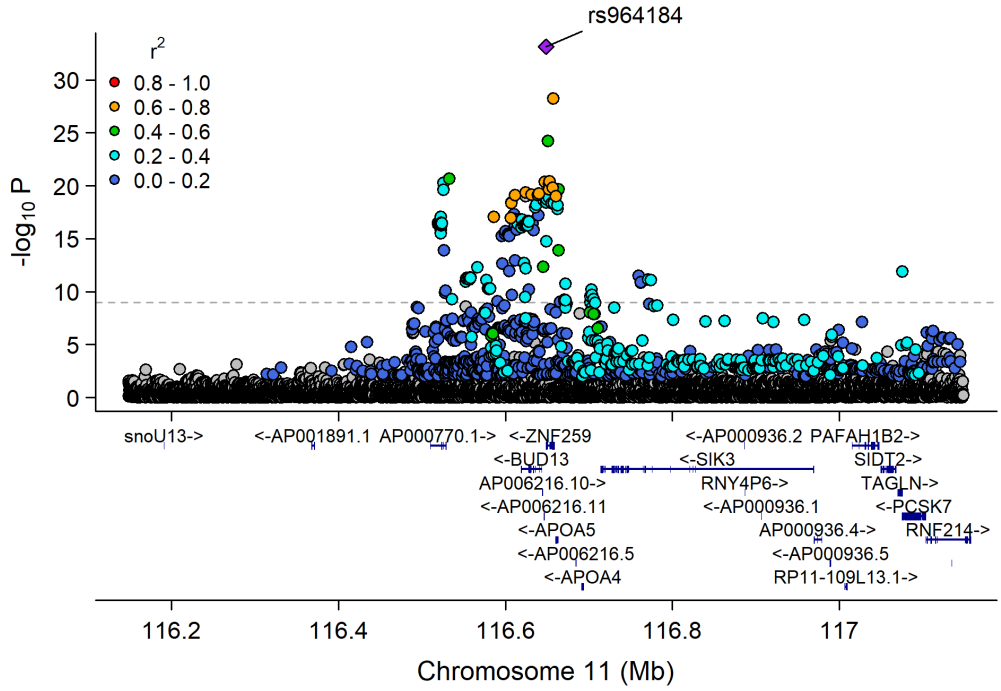
Chr8__18582620__20792744



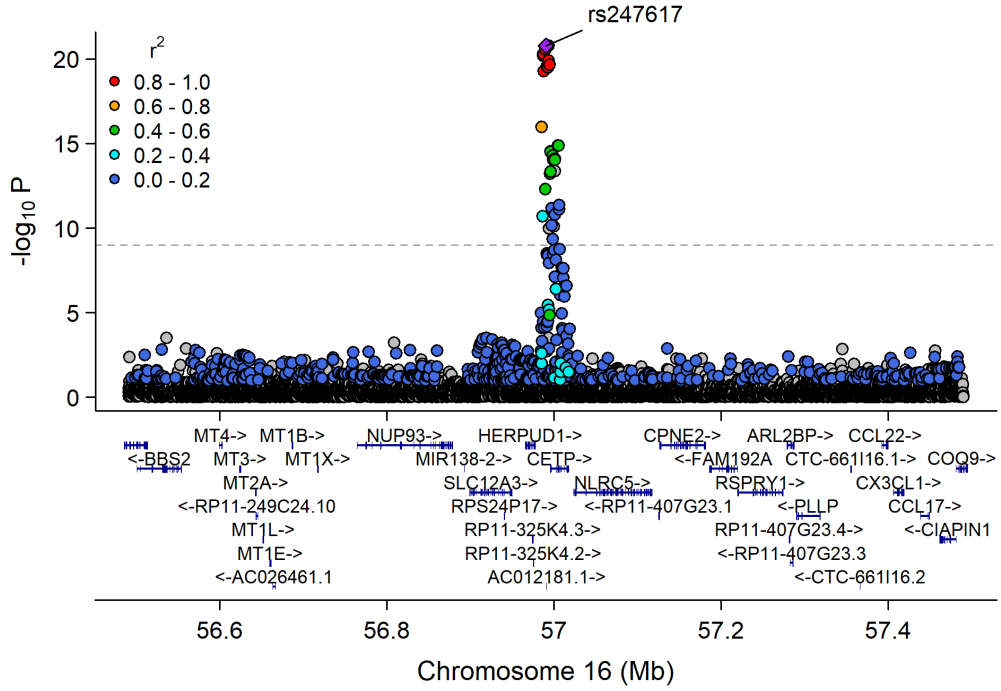
Chr8__126435663__126578061



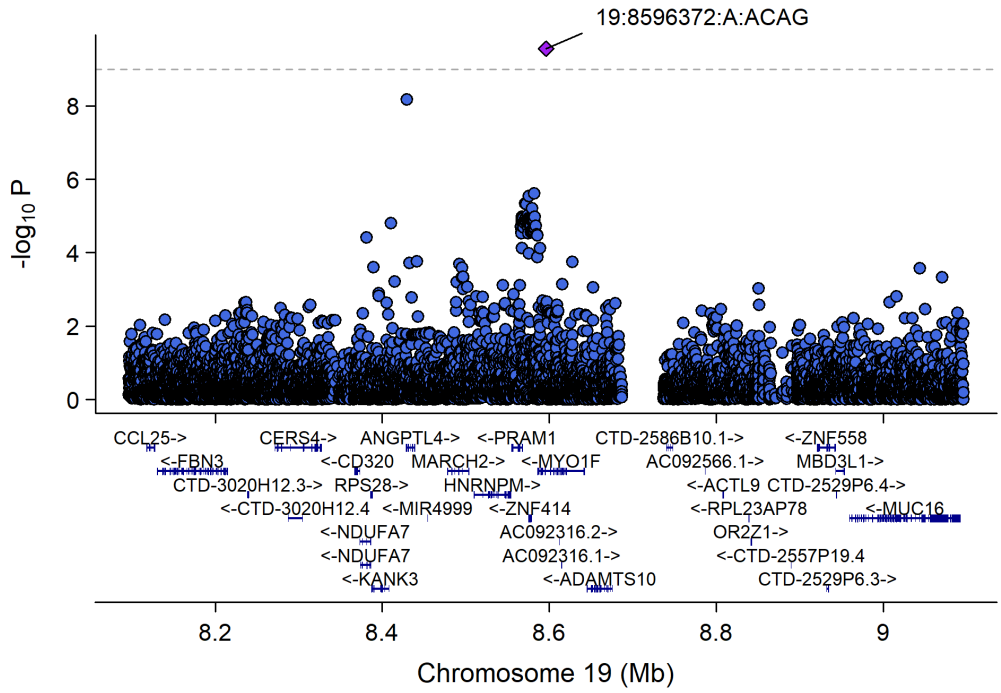
Chr11_115541901_117633315



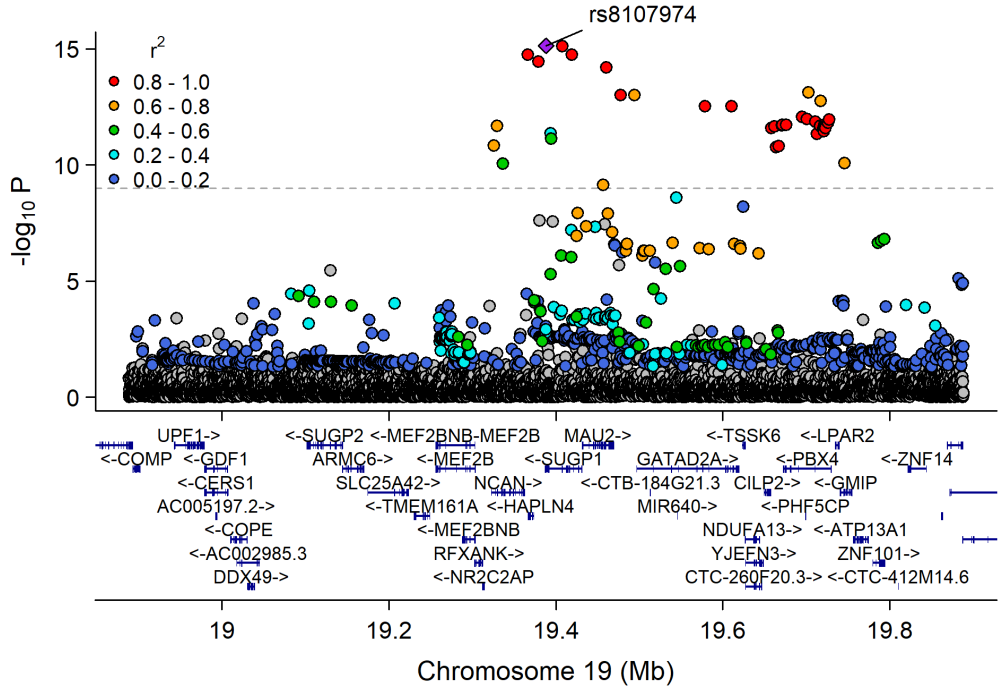
Chr16_55870822_57992421



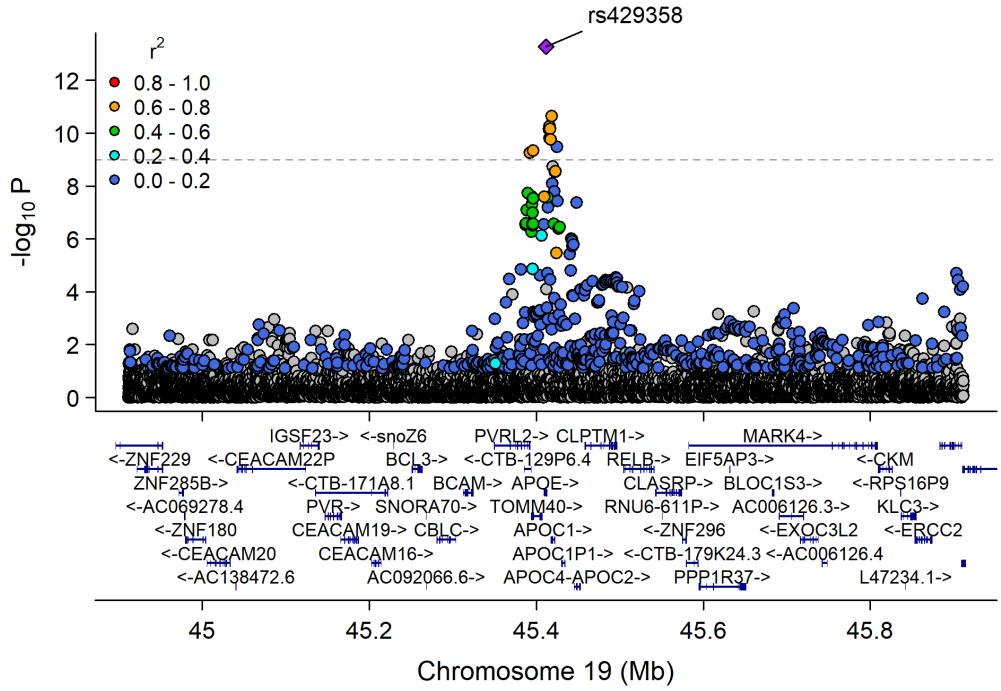
Chr19__7600546__9567827



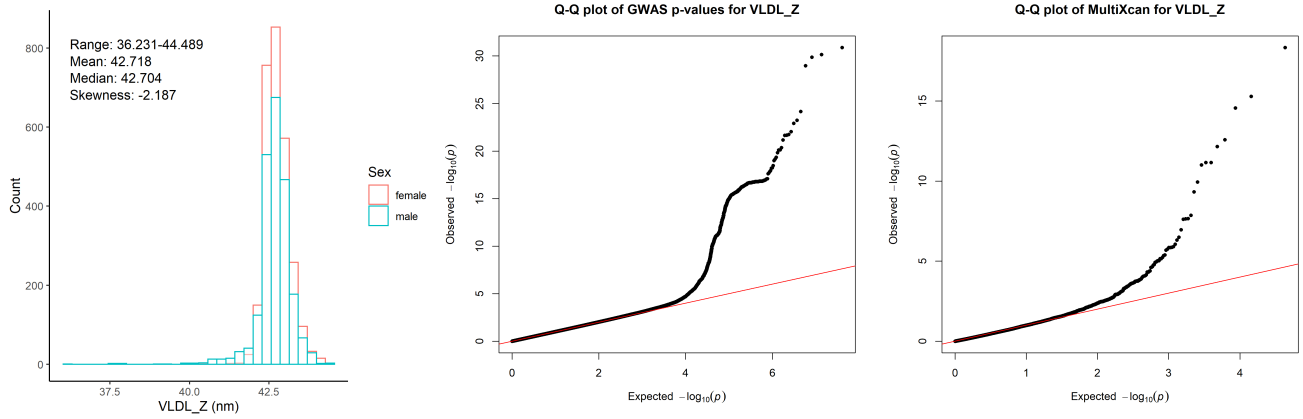
Chr19__18370495__20841464

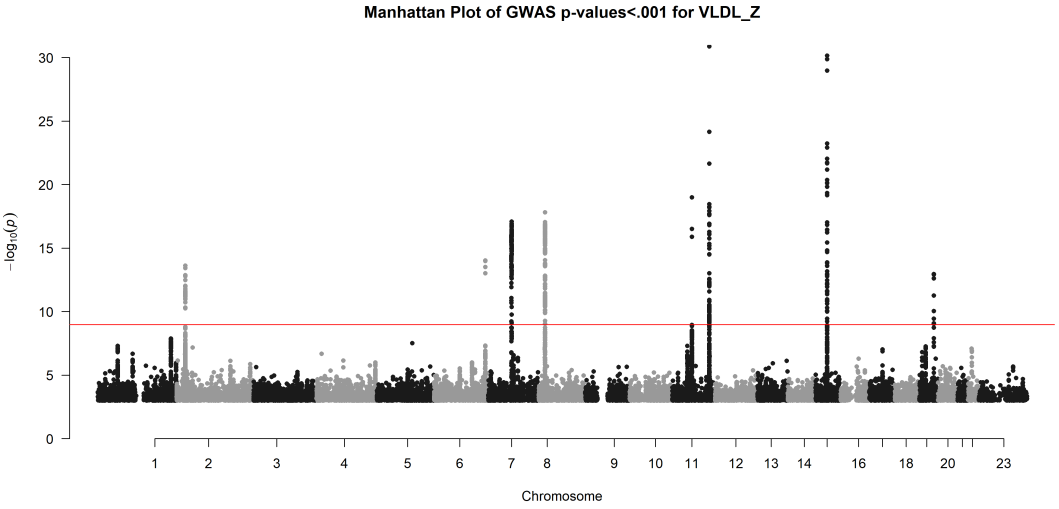


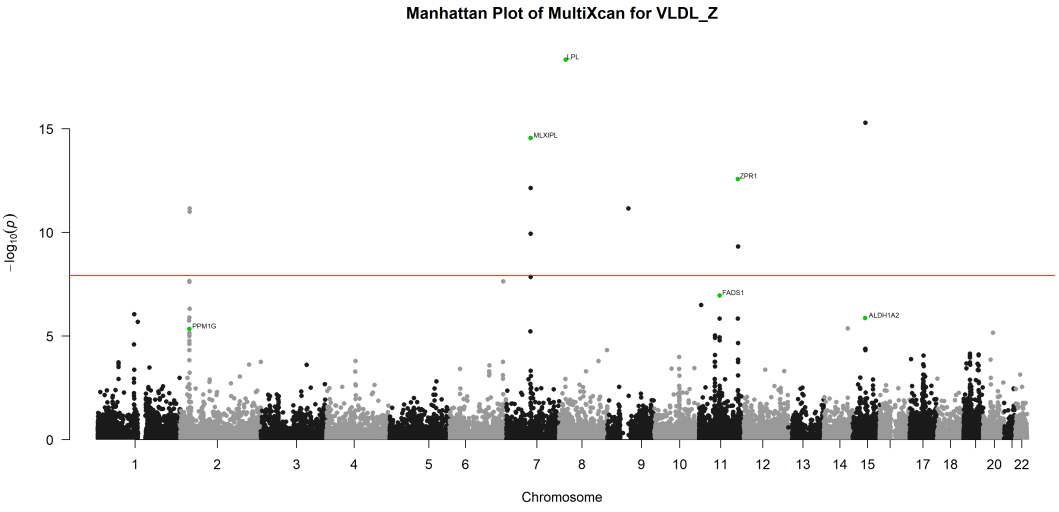
Chr19__44063363__46637375



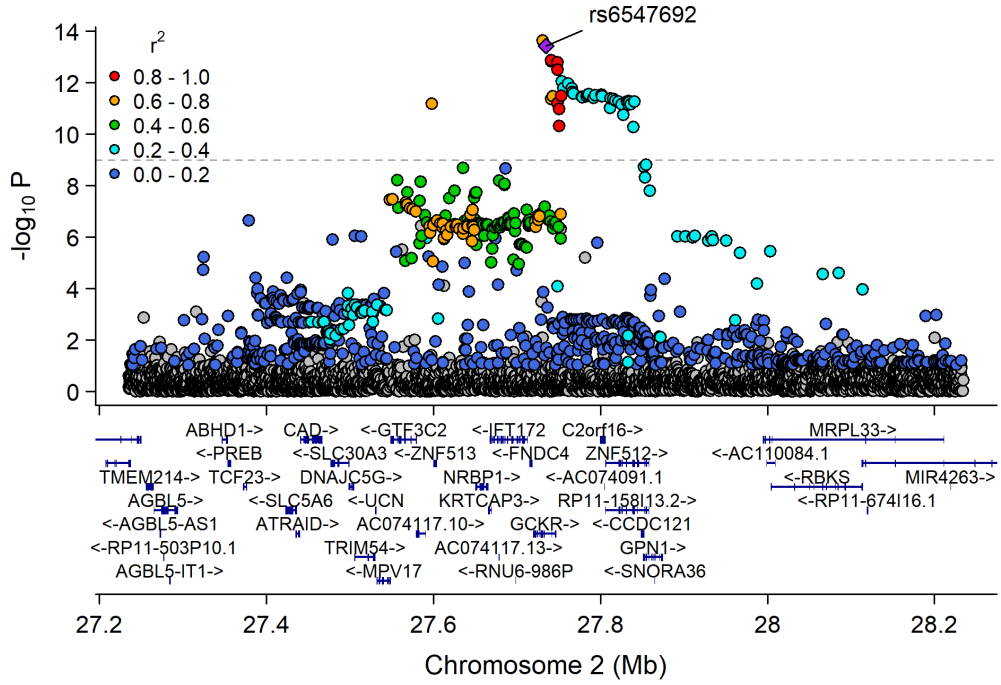
Mean diameter for VLDL particles (nm)



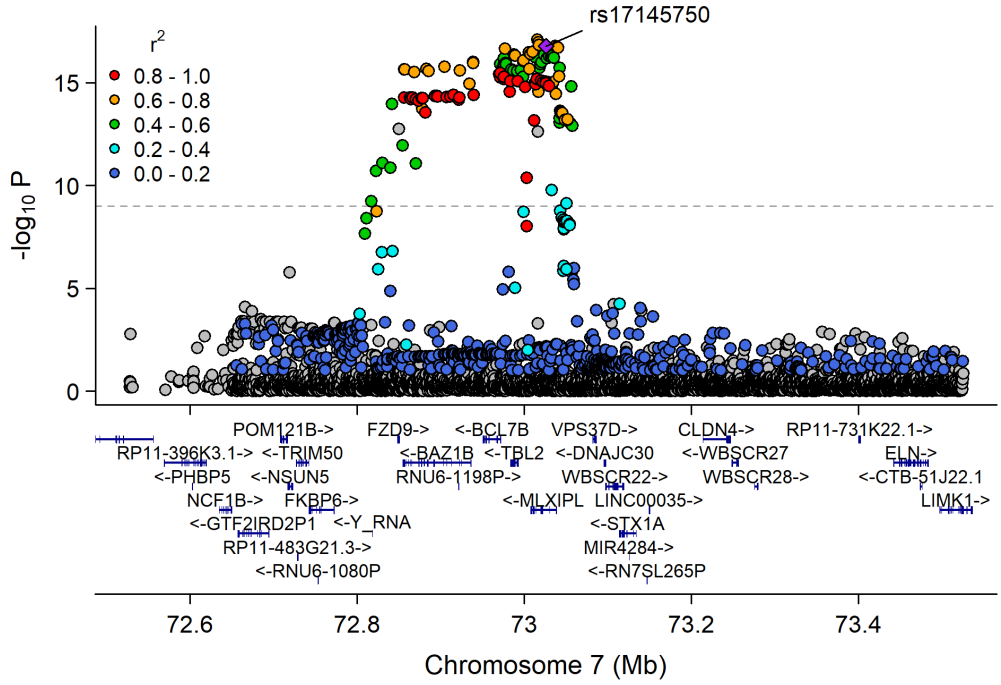




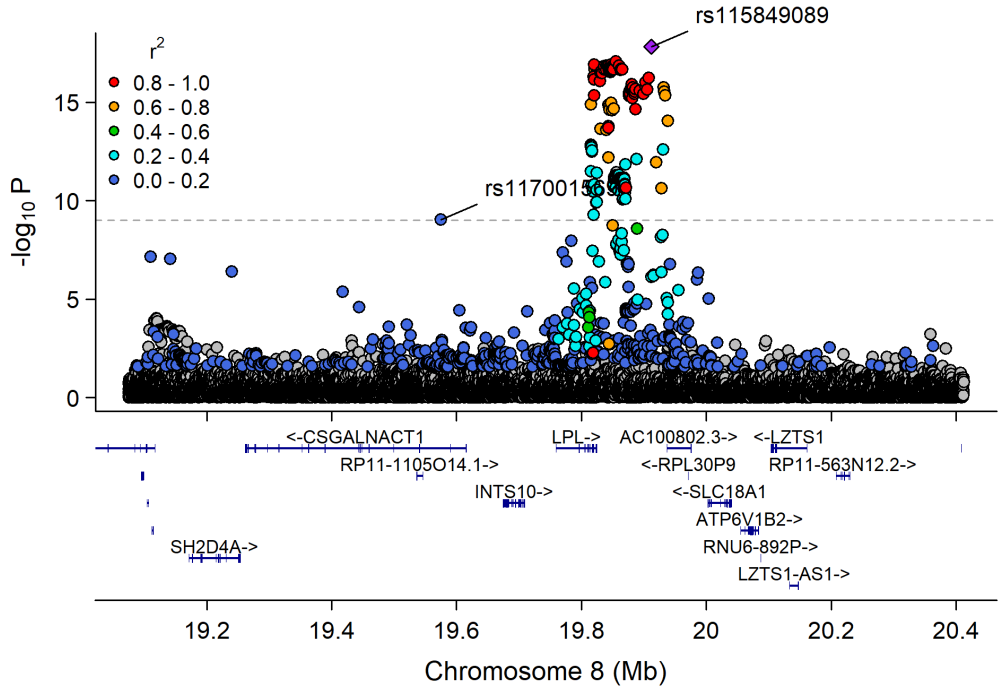
Chr2_26753815_28597624



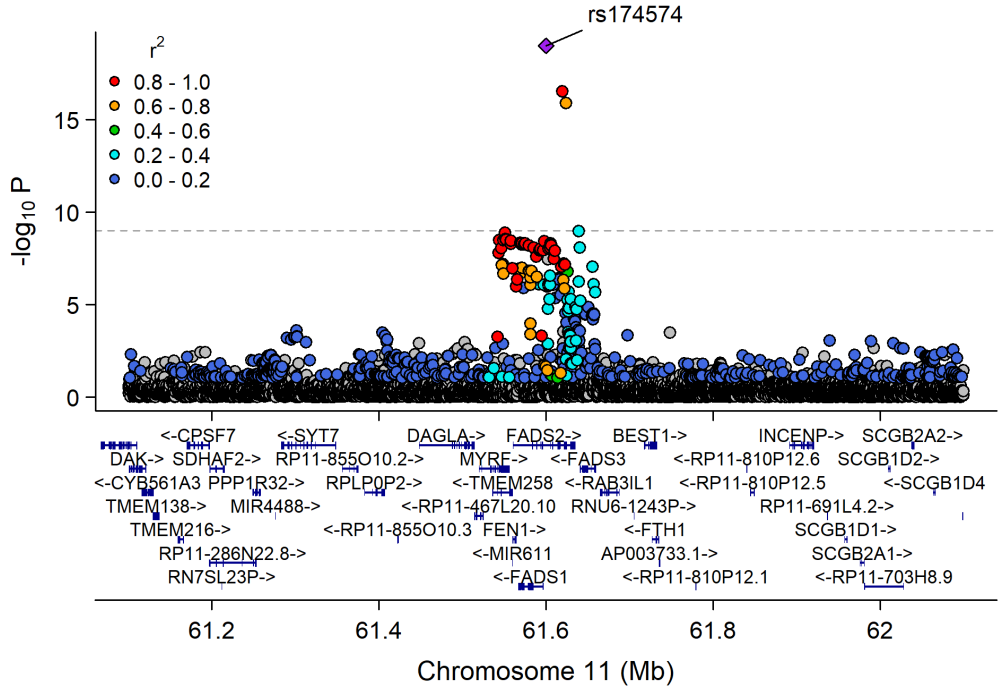
Chr7_72070325_73133300



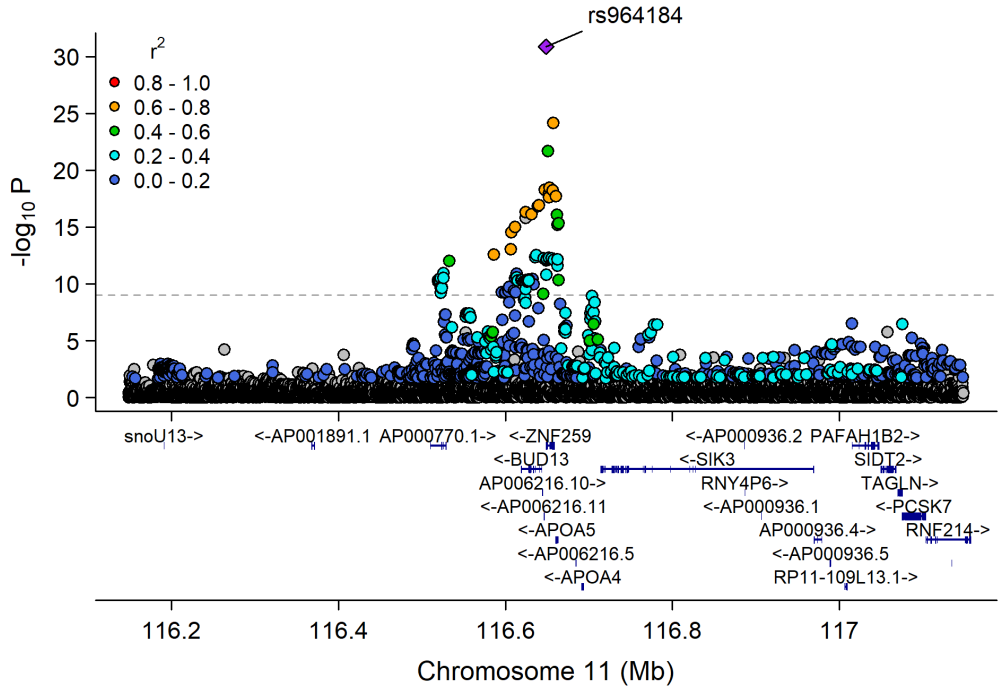
Chr8_18582620_20792744



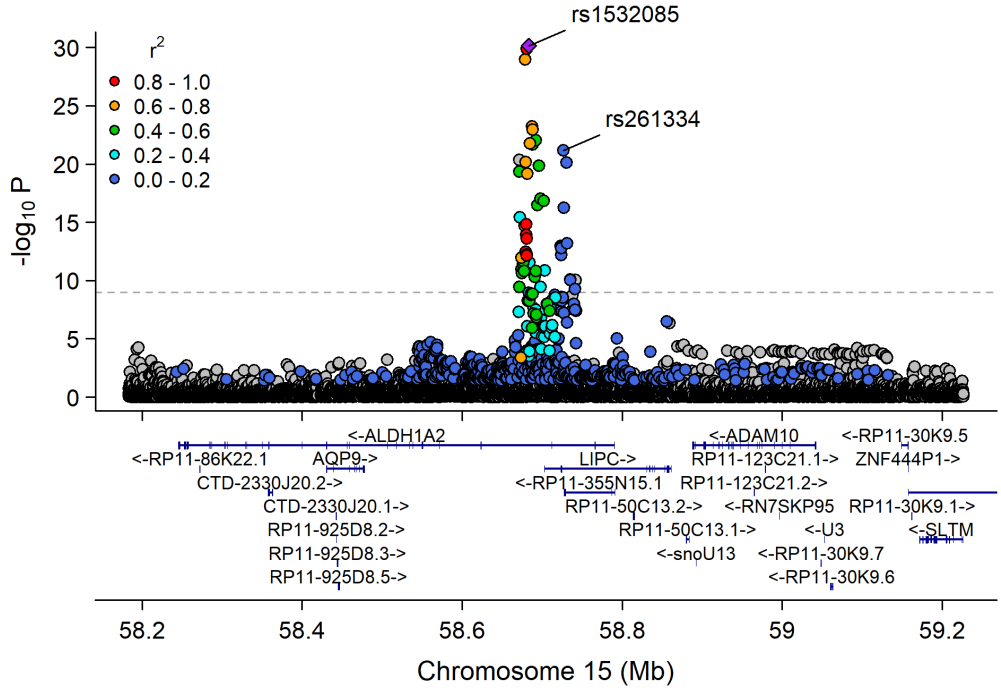
Chr11_59978355_62914375



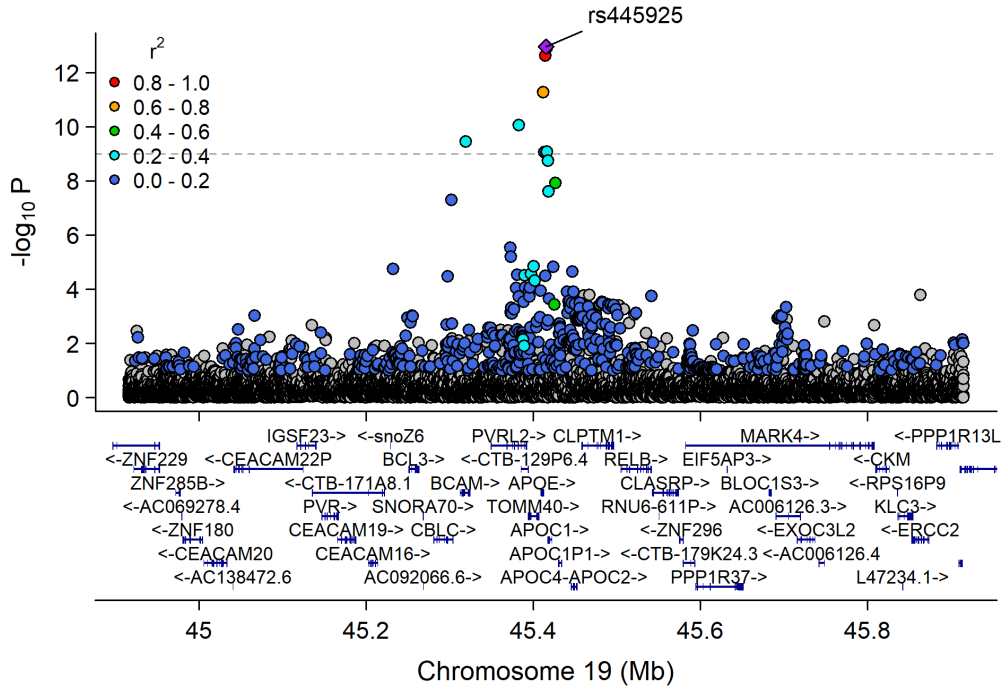
Chr11_115541901_117633315



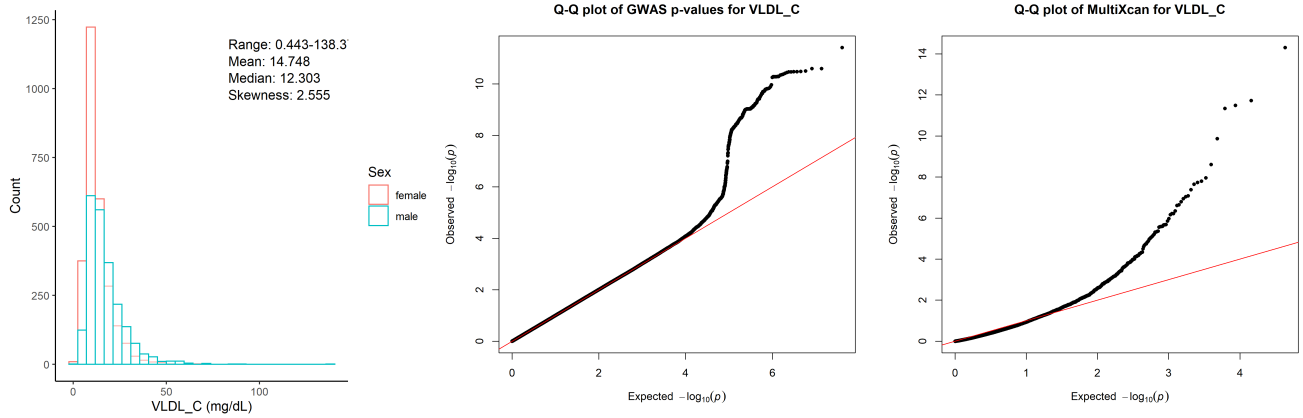
Chr15_57658798_60490883

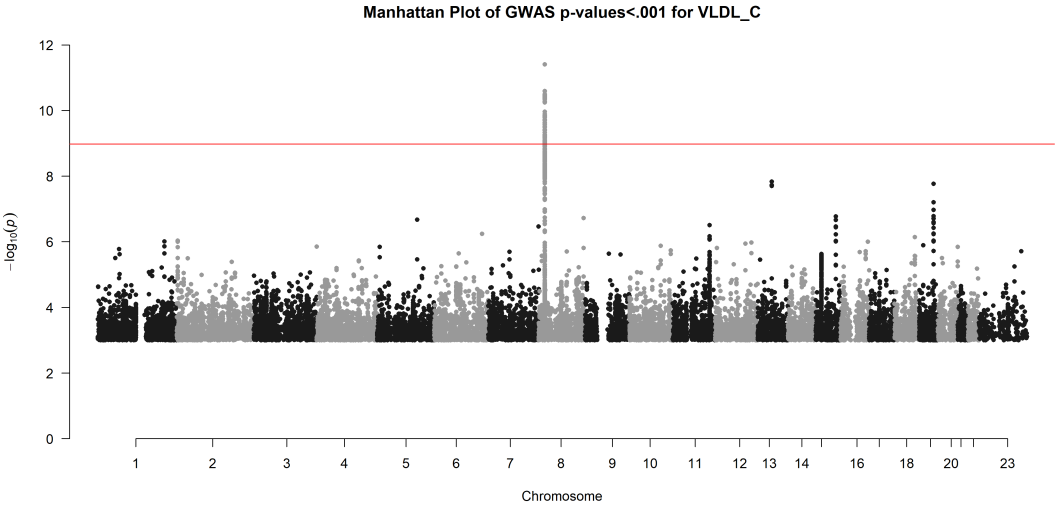


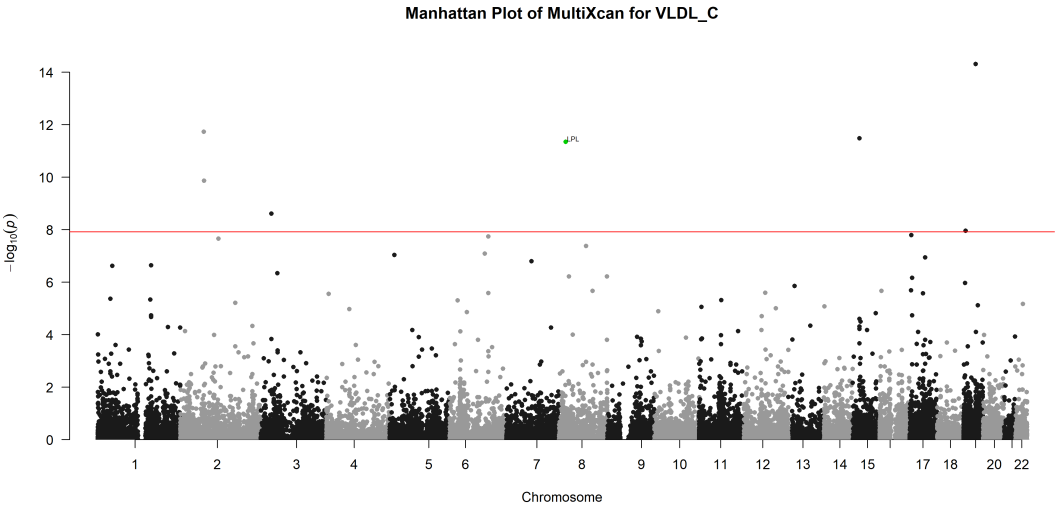
Chr19__44063363__46637375



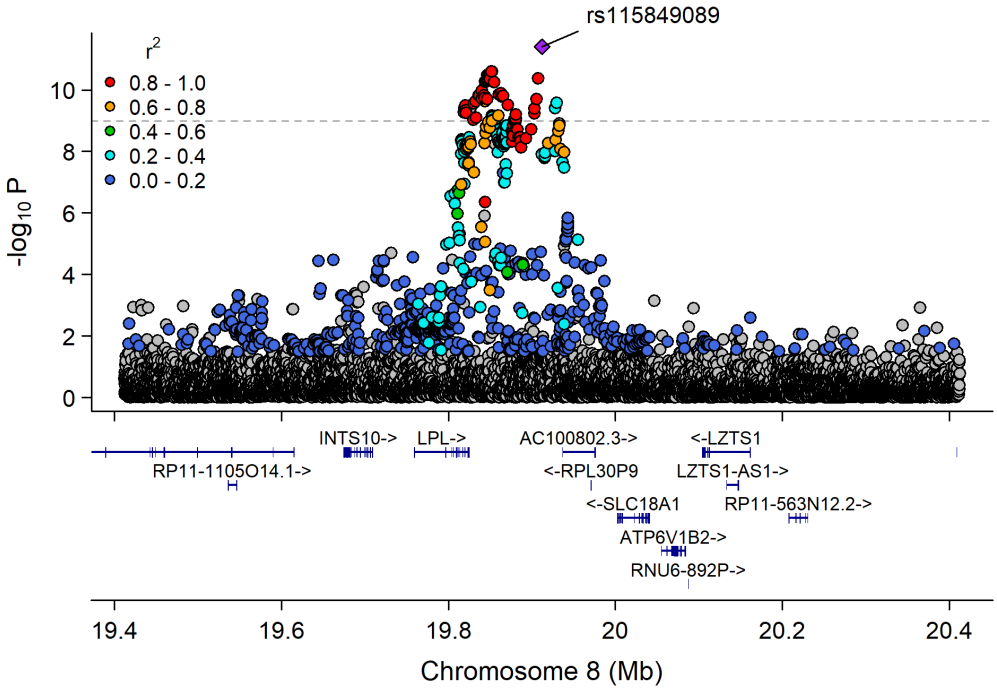
Total cholesterol in VLDL (mg/dL)



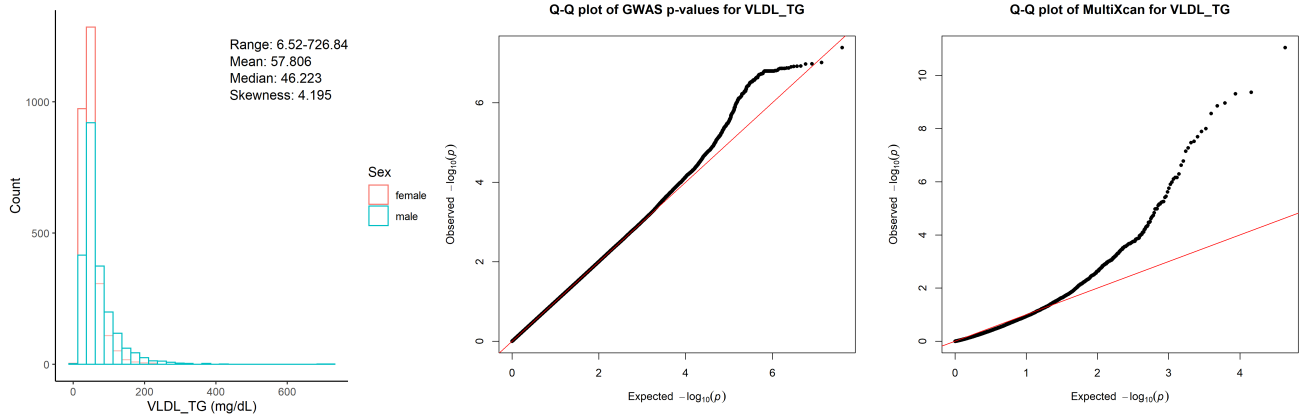


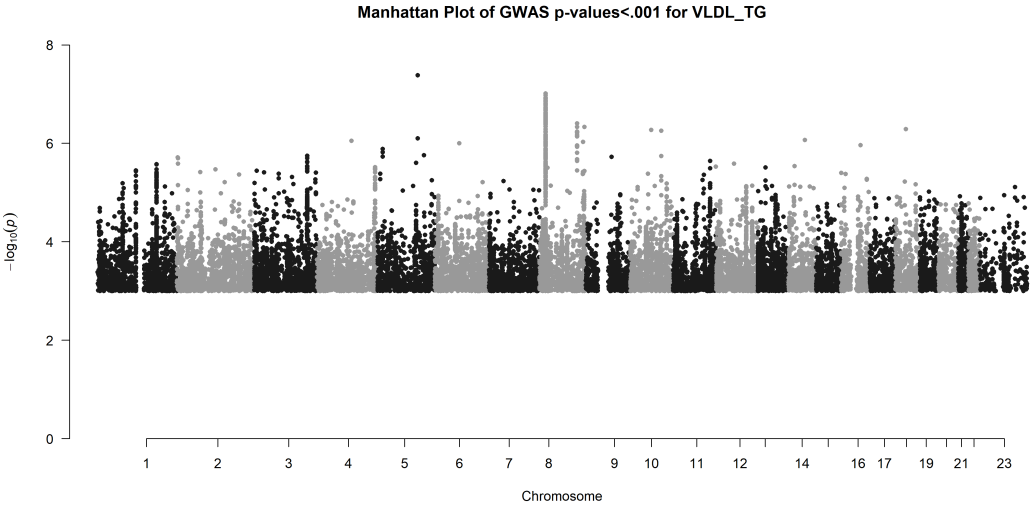


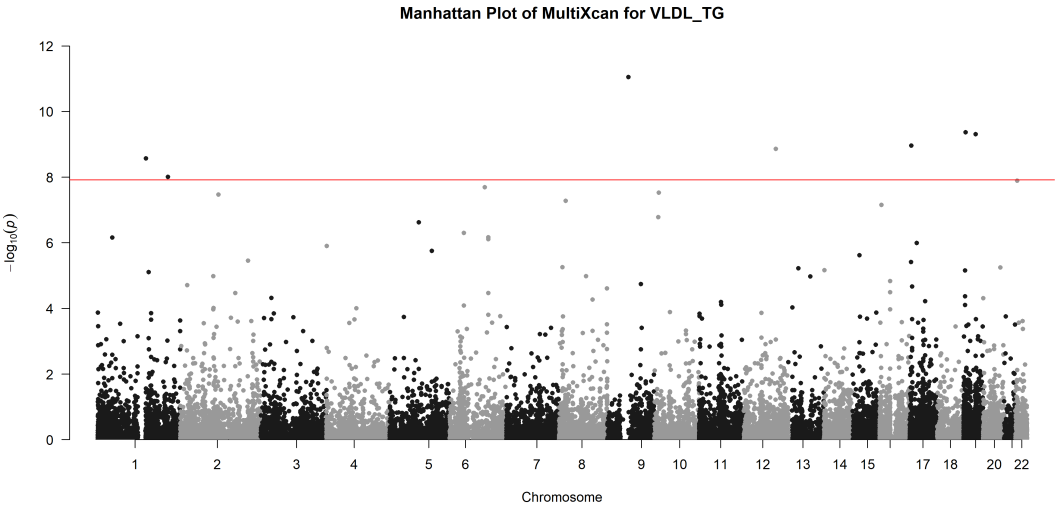
Chr8__18582620__20792744



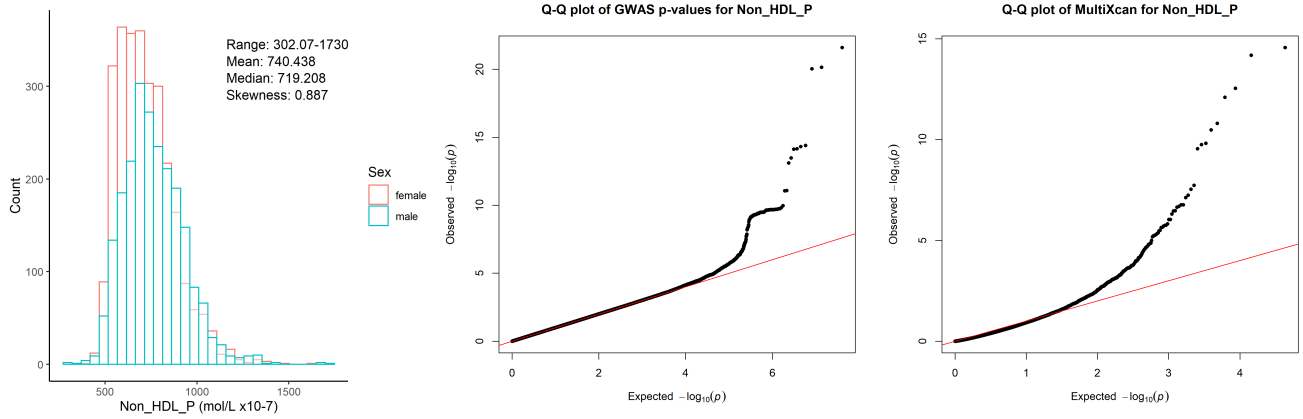
Triglycerides in VLDL (mg/dL)

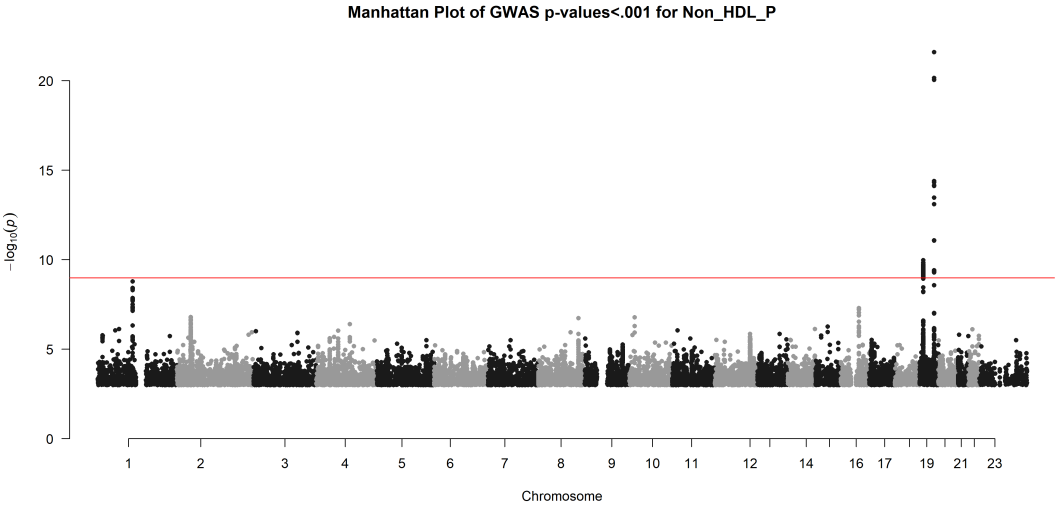


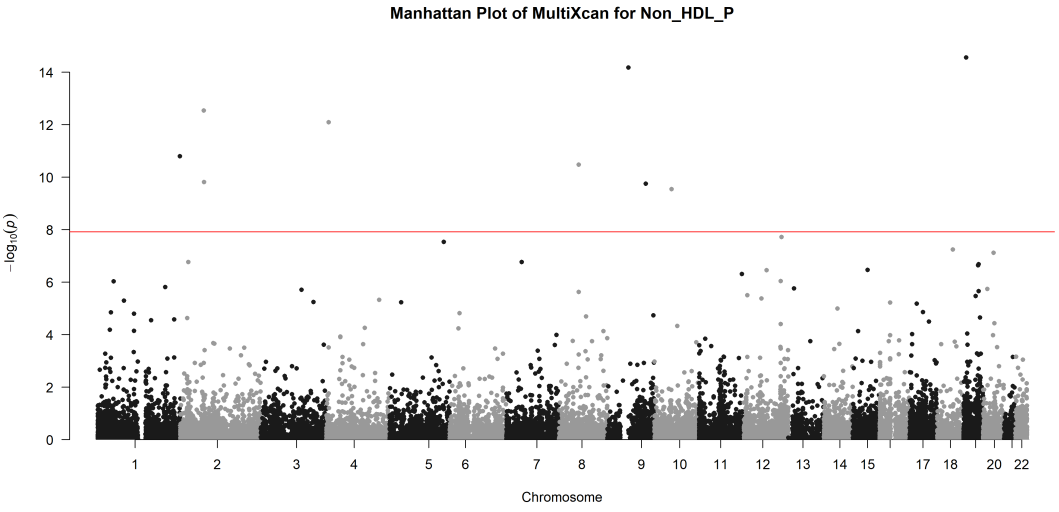




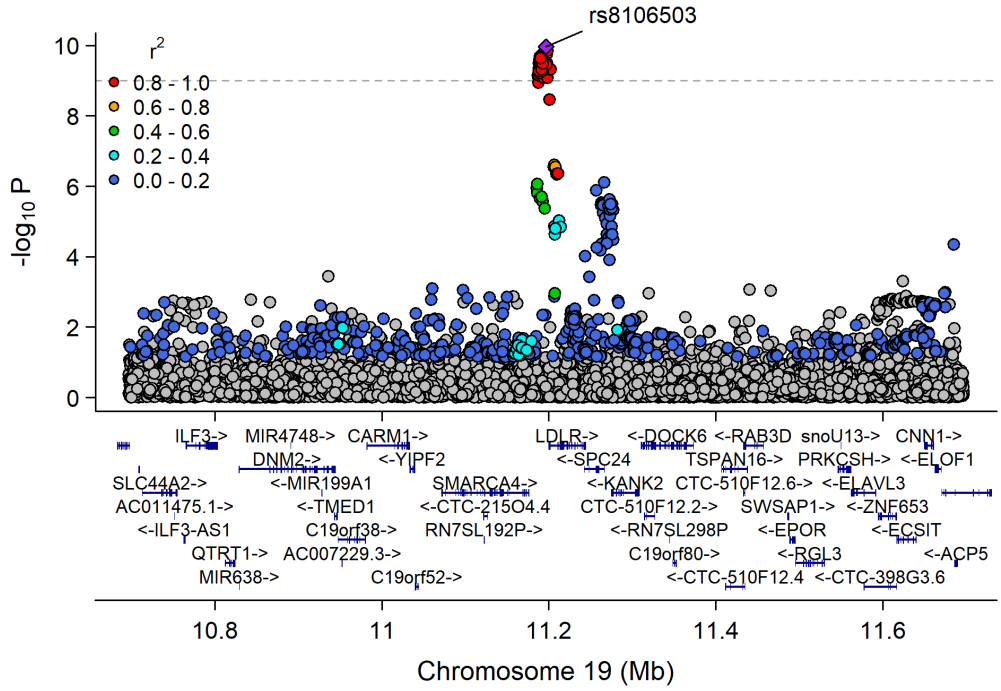
Concentration of non-HDL particles (mol/l)



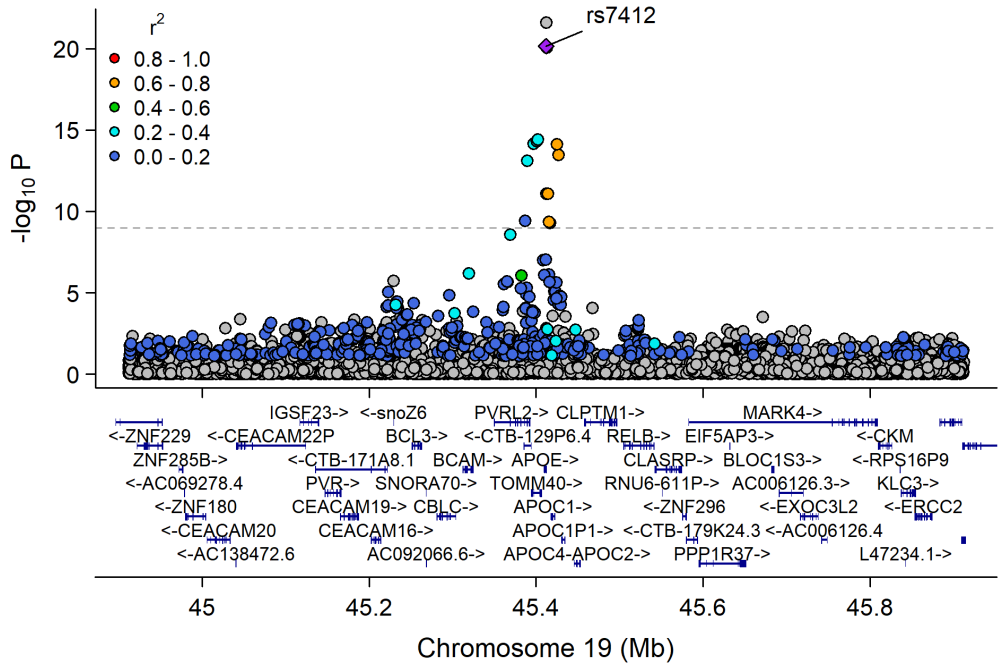




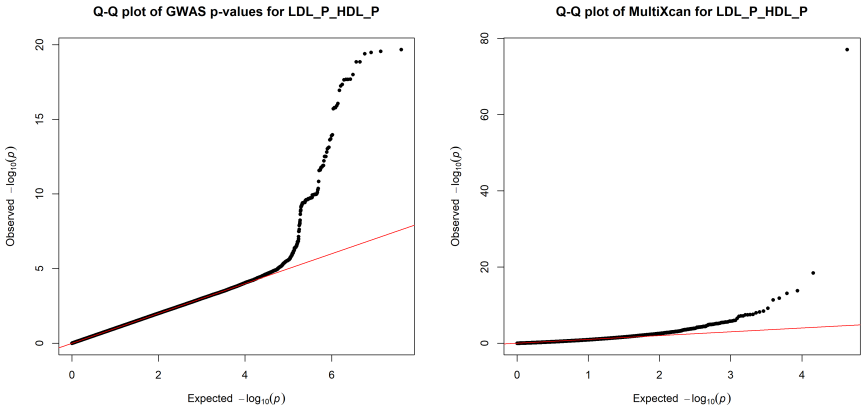
Chr19__9807999__12255594

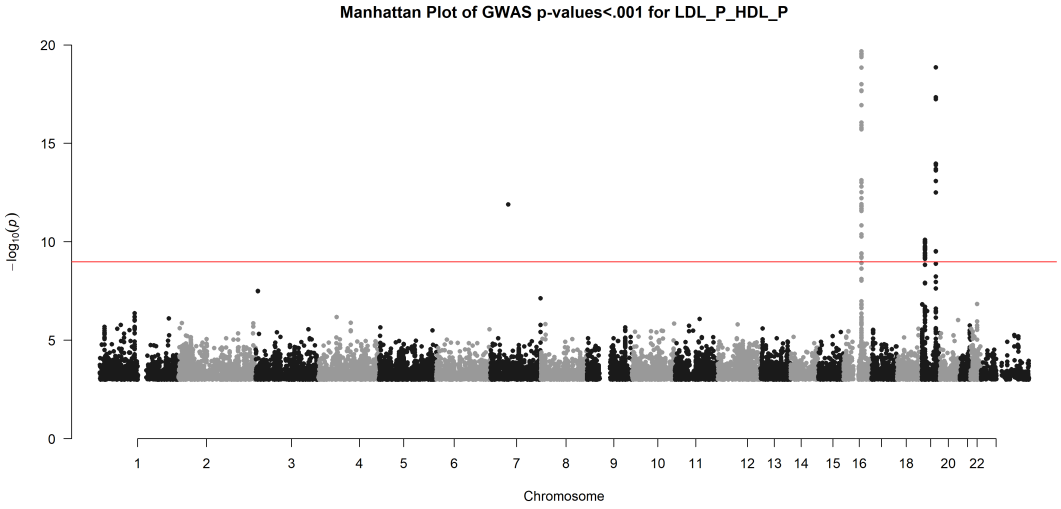


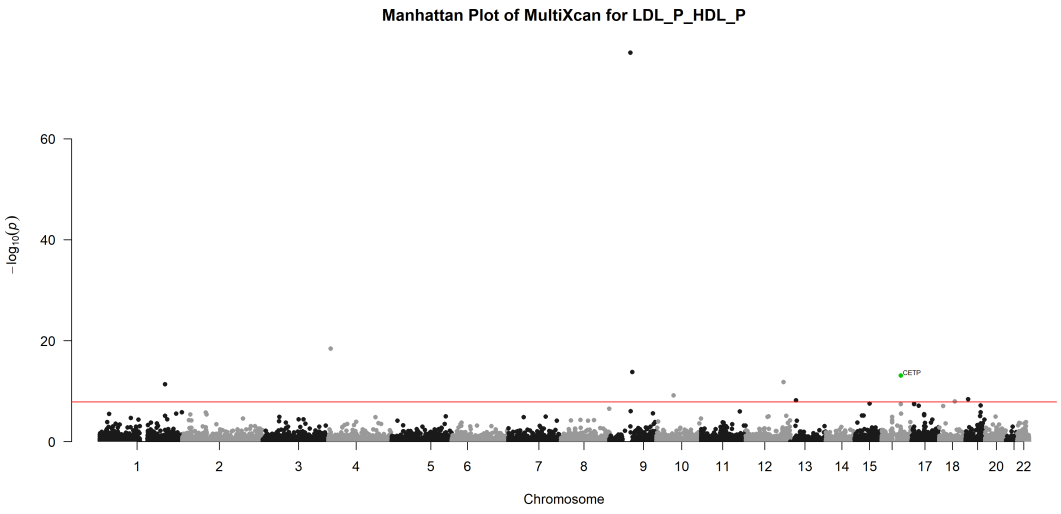
Chr19__44063363__46637375



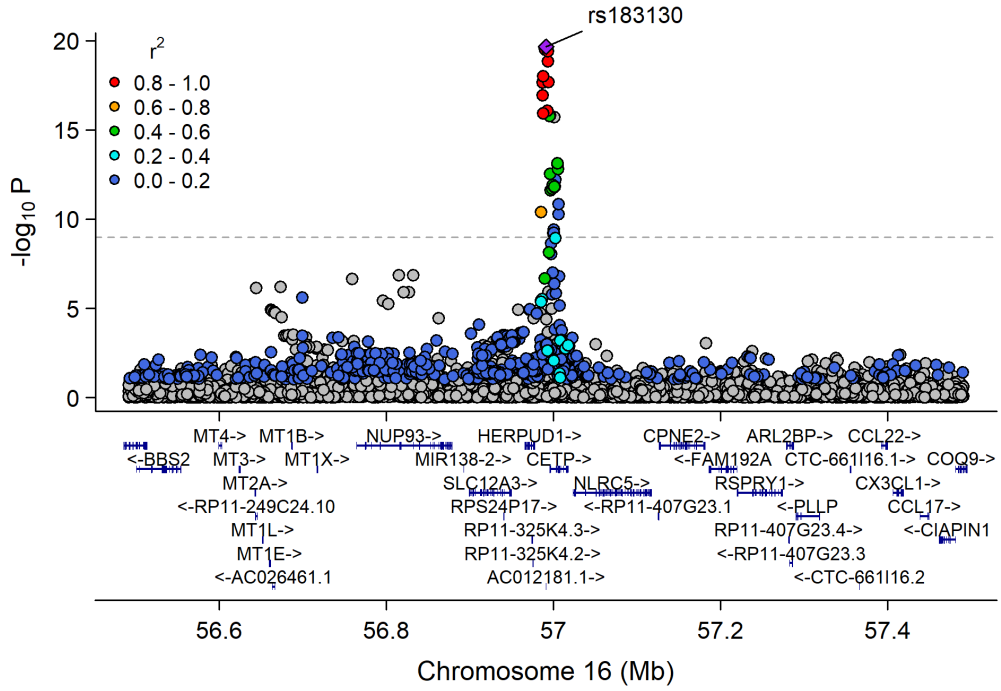
Ratio of concentration of LDL particles to HDL particles



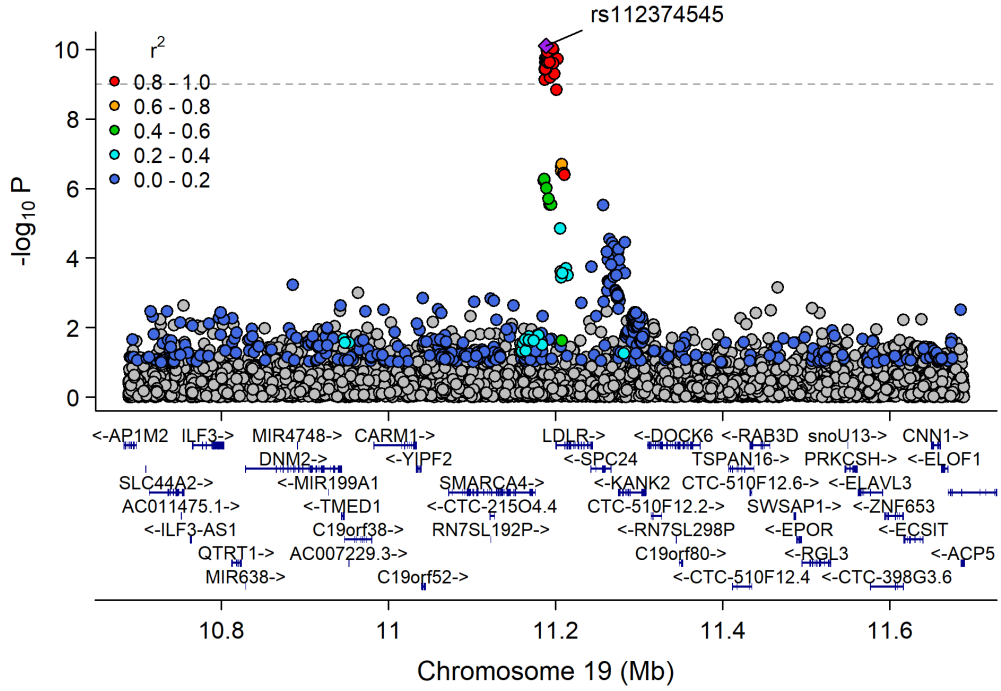




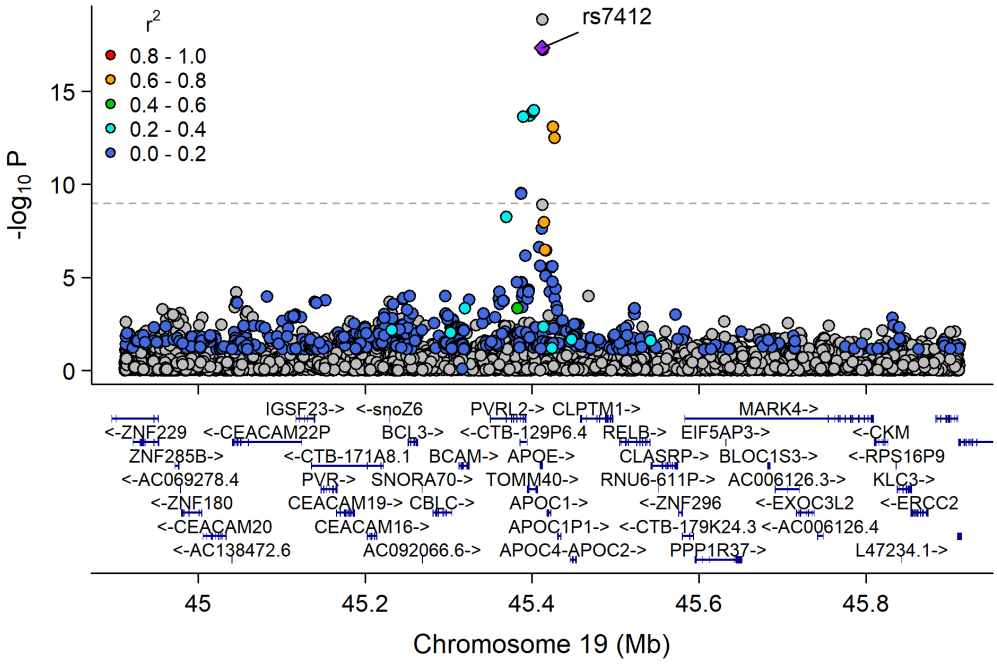
Chr16_55870822_57992421



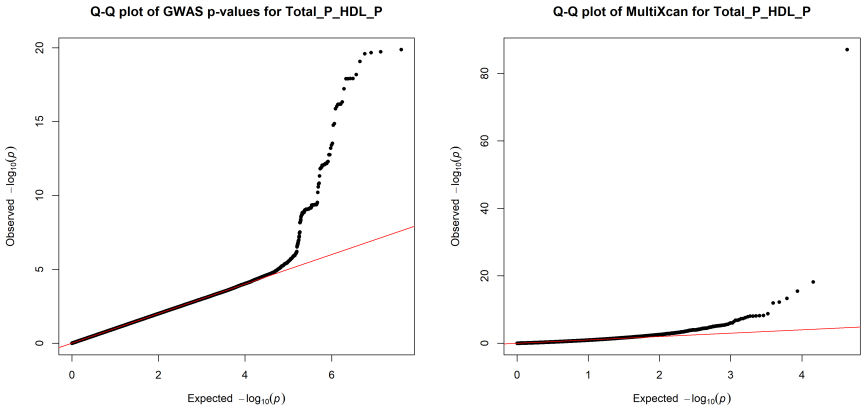
Chr19_9807999_12255594

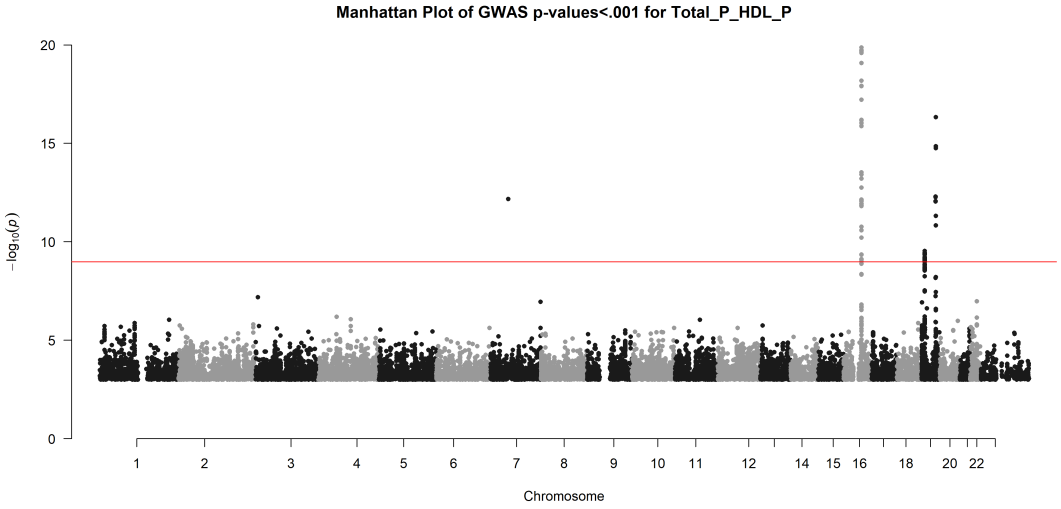


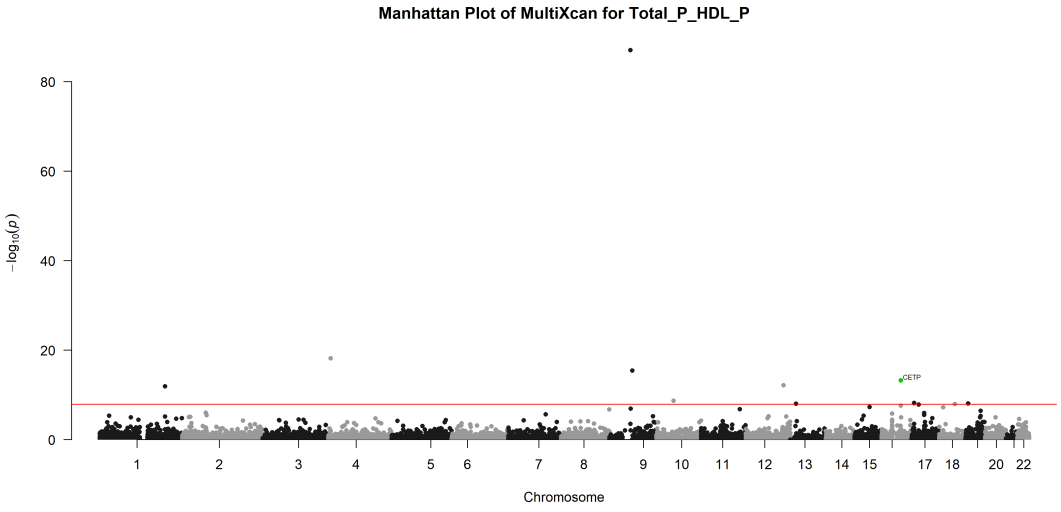
Chr19__44063363__46637375



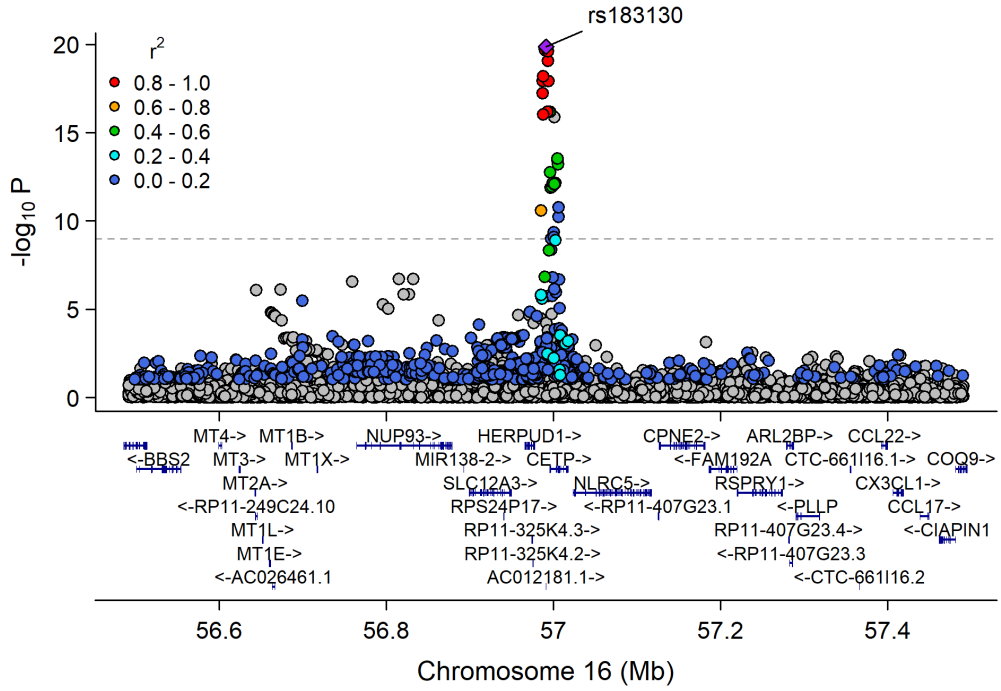
Ratio of concentration of total particles to HDL particles



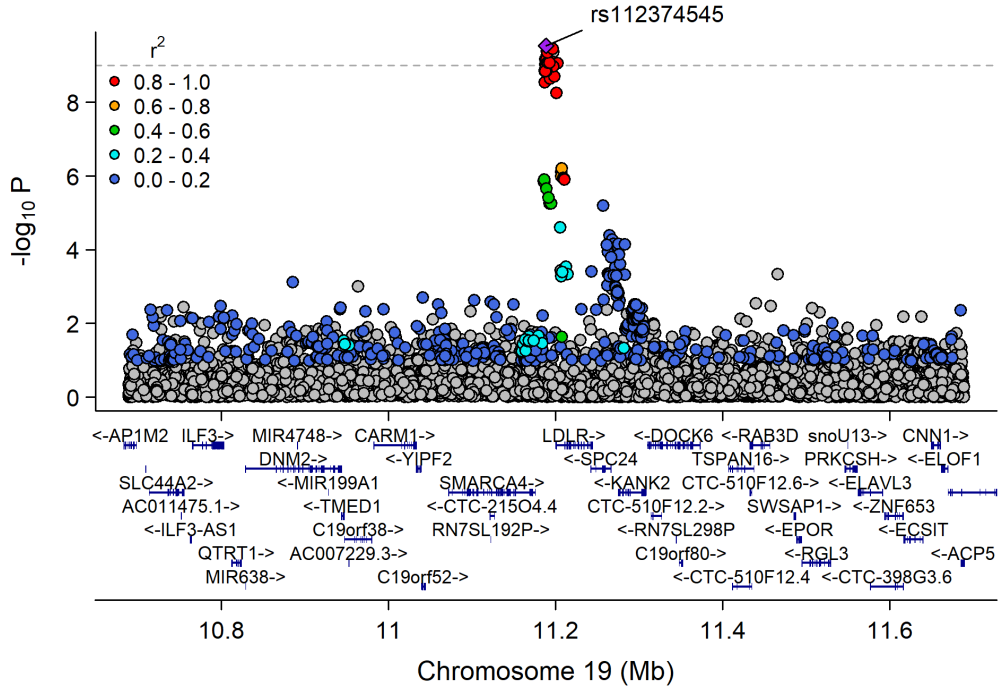




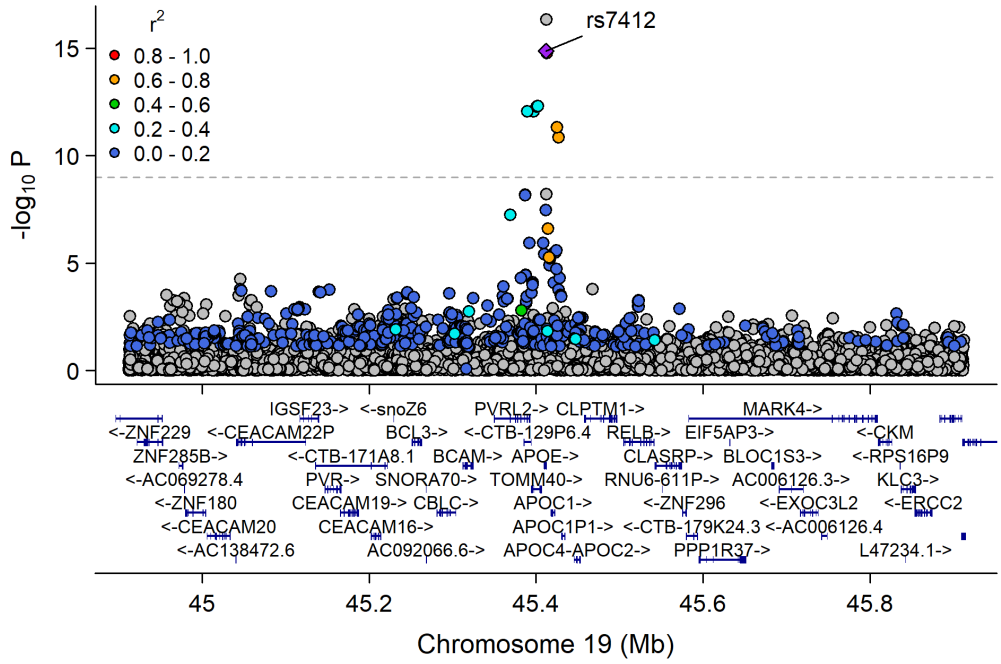
Chr16_55870822_57992421



Chr19_9807999_12255594

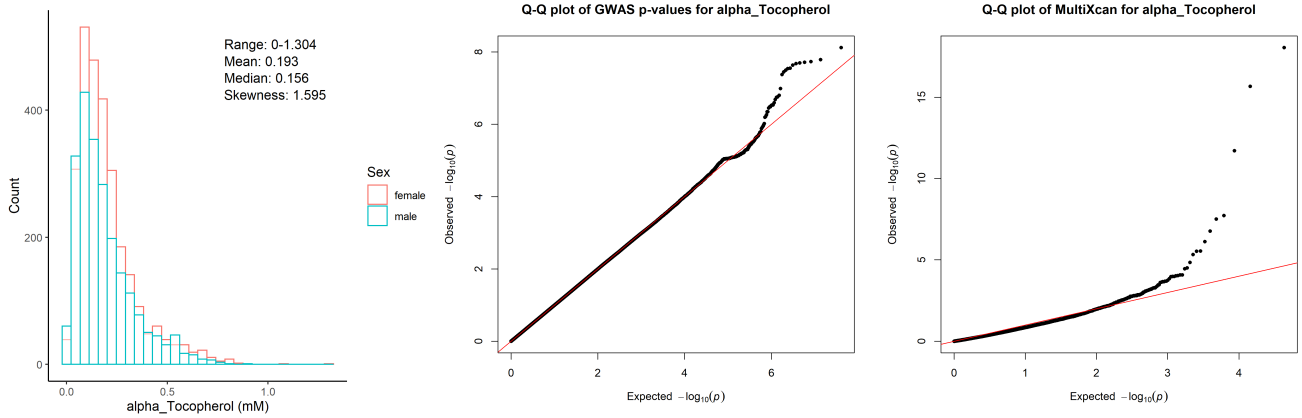


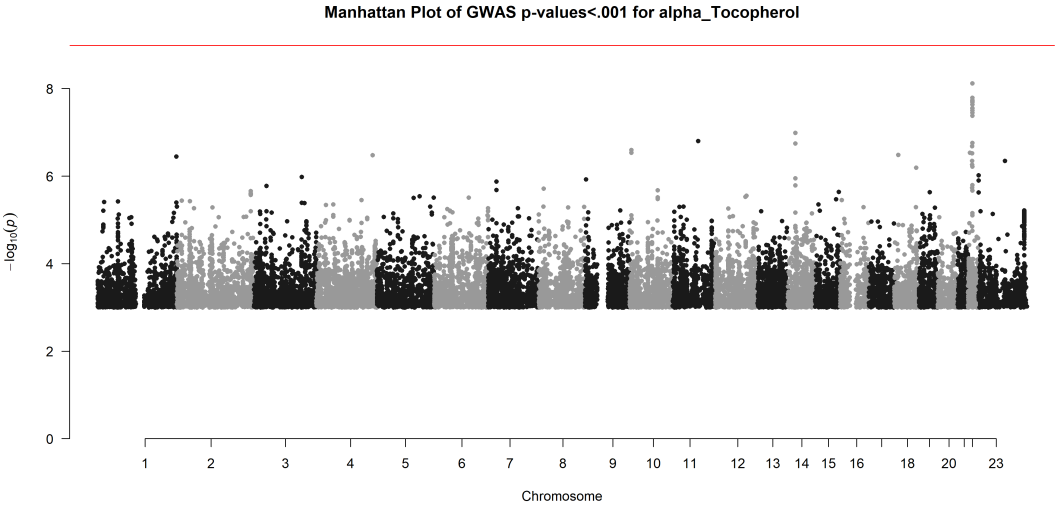
Chr19_44063363_46637375



Prenols

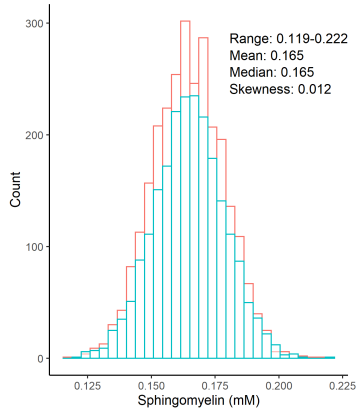
Alpha-Tocopherol (mM)

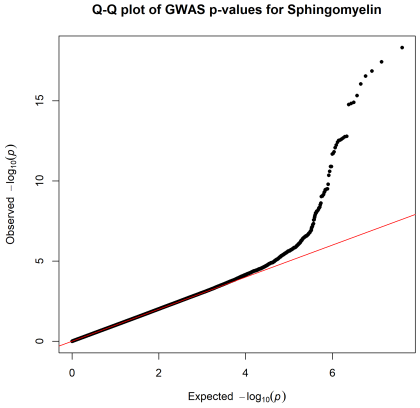


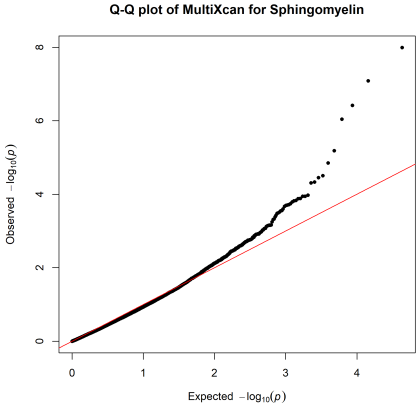


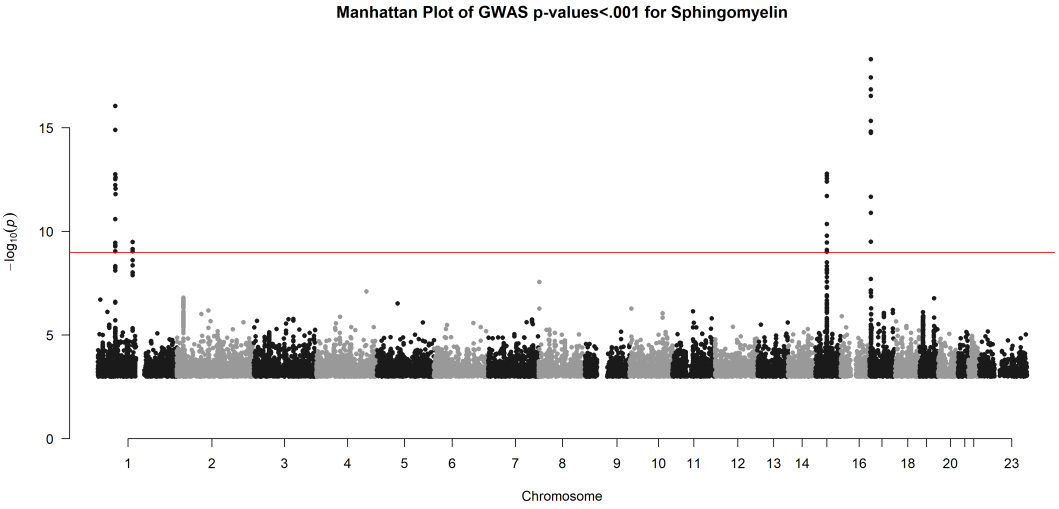
Sphingolipids

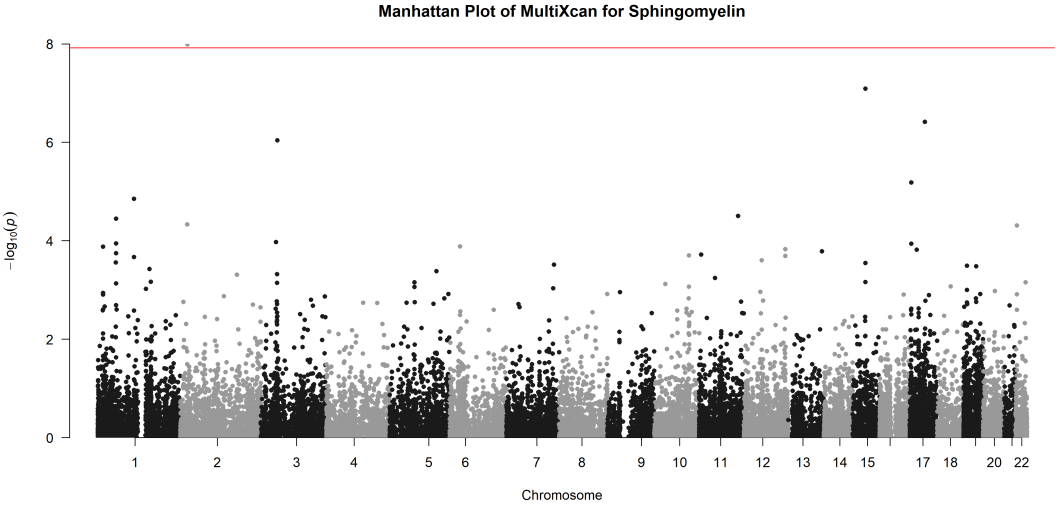
Sphingomyeline (mM)



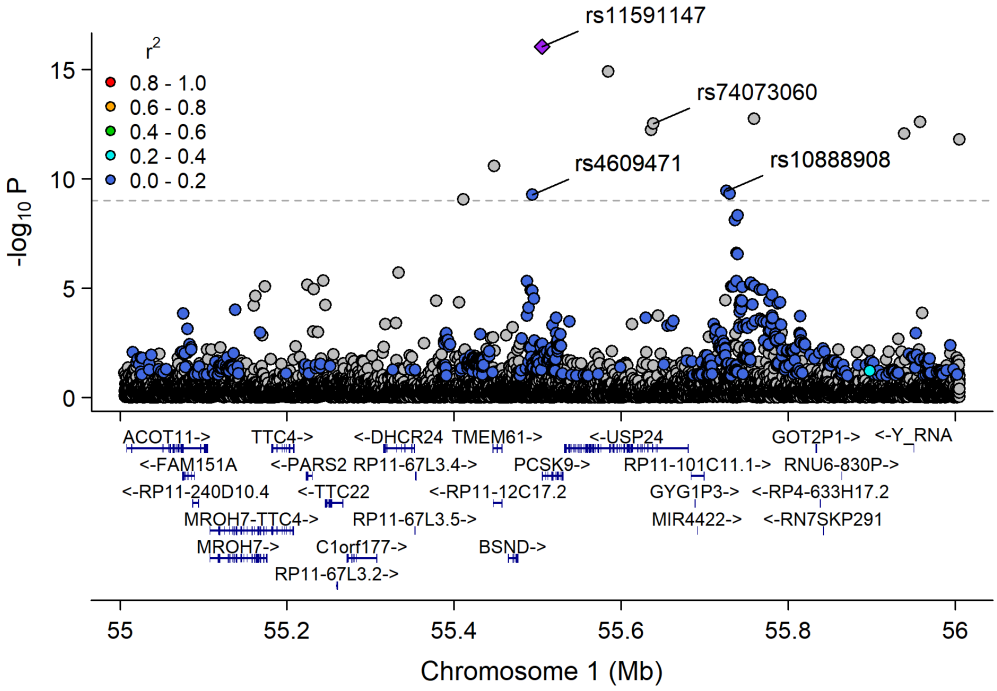




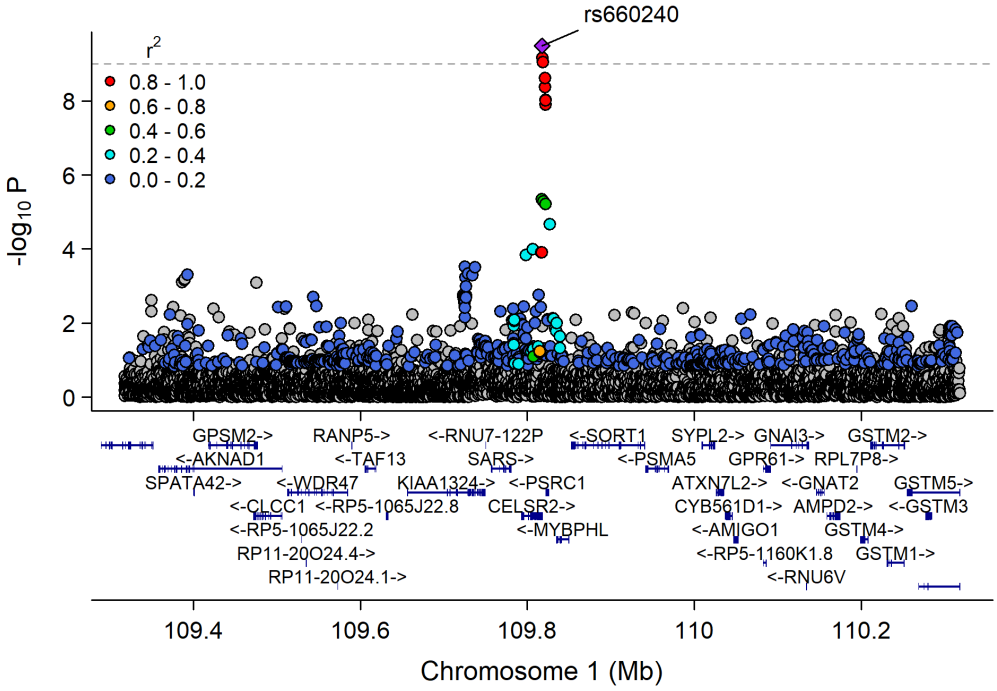




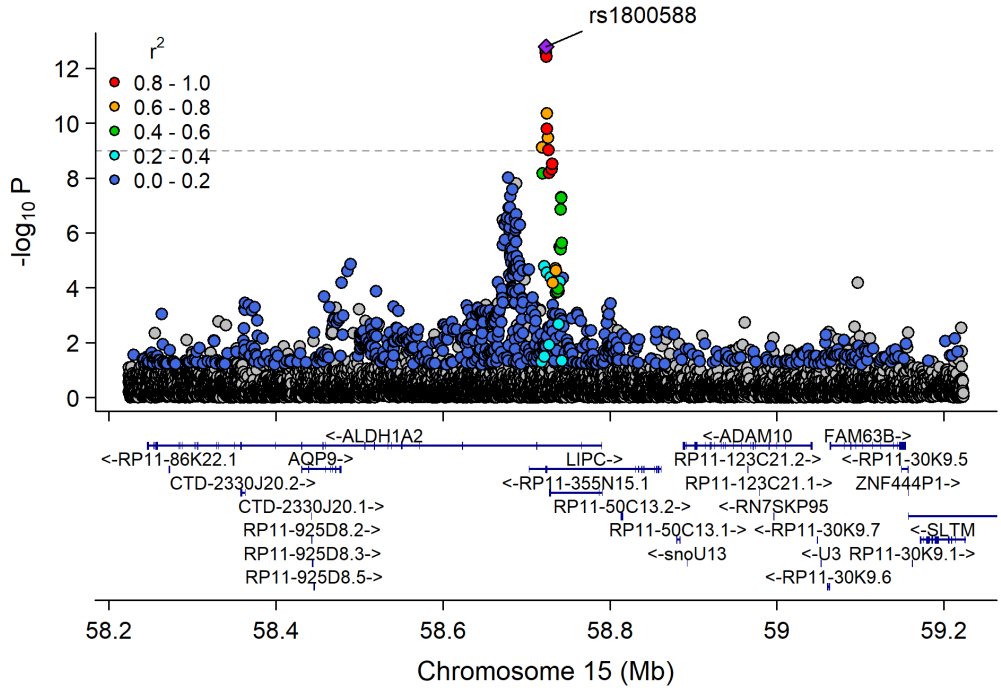
Chr1__53272879__57300396



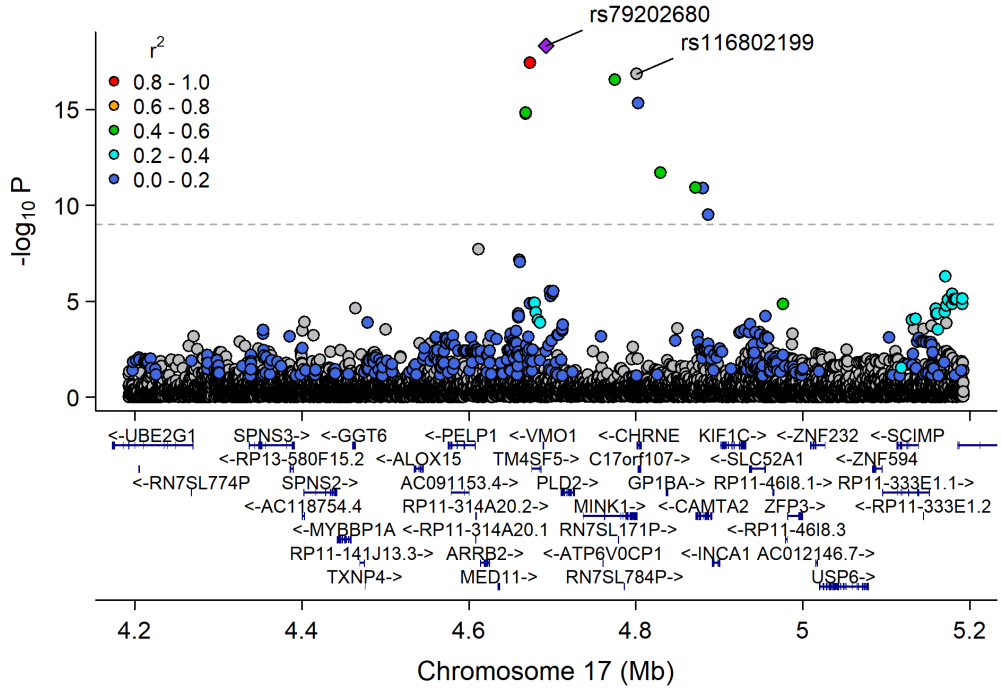
Chr1__109696238__110162190



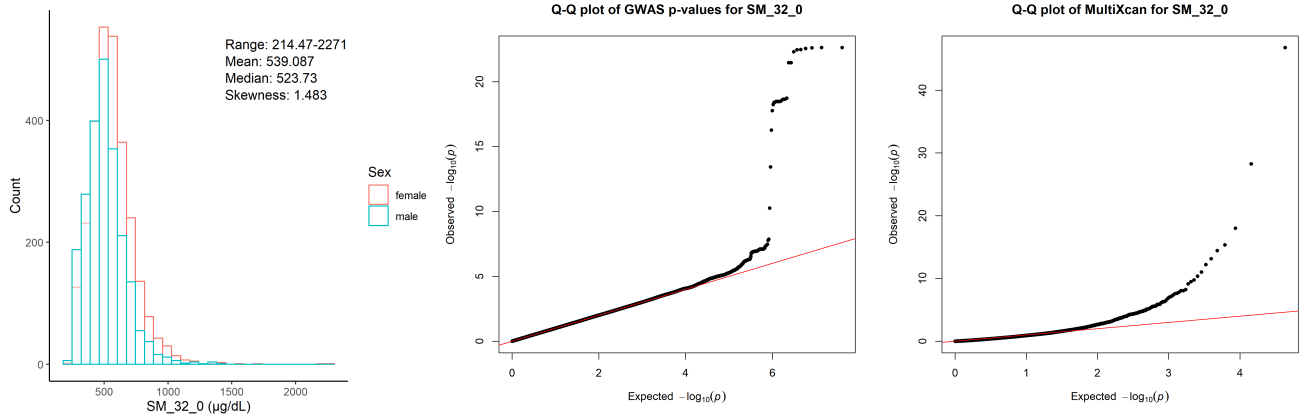
Chr15_57658798_60490883

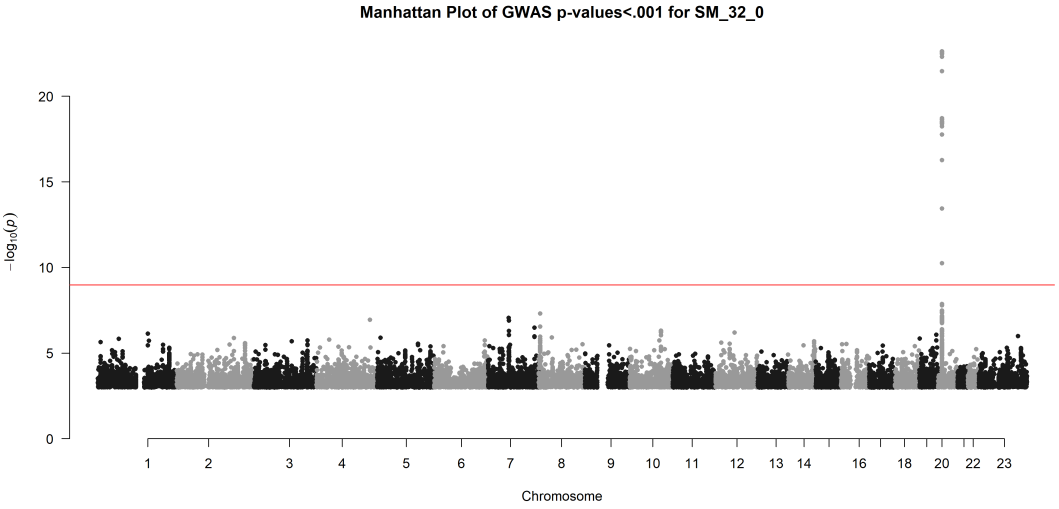


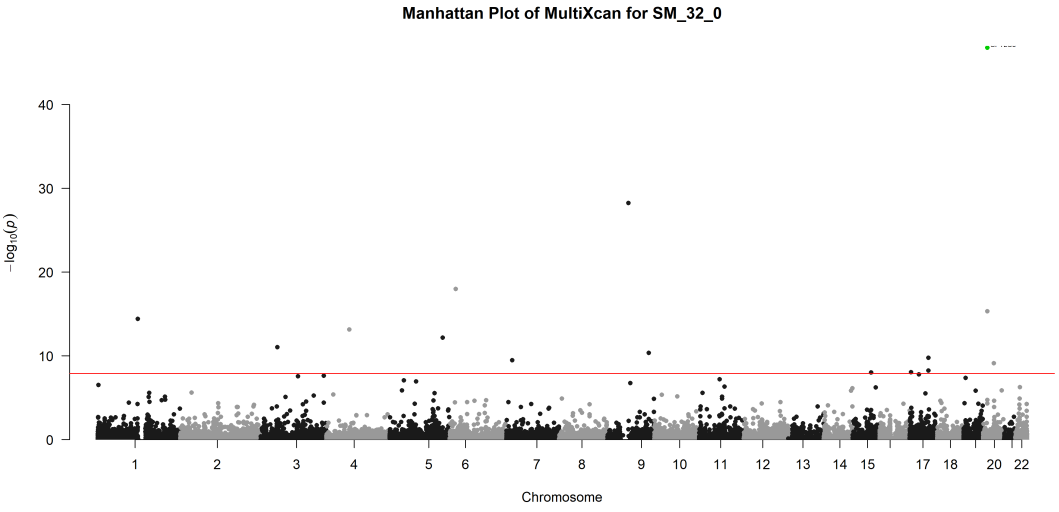
Chr17_3704190_5793348



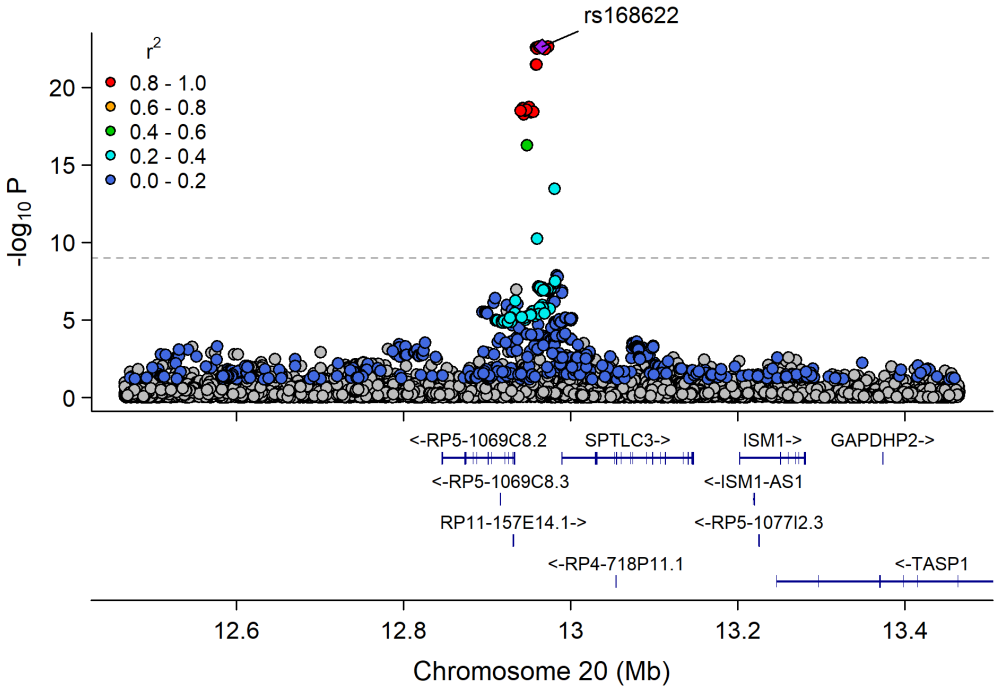
Sphingomyeline C32:0 (µg/dL)



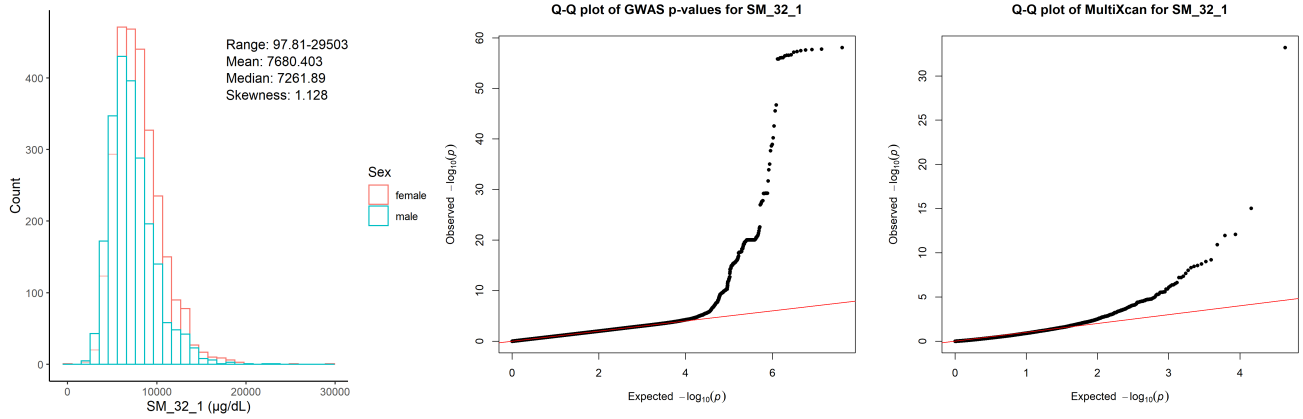


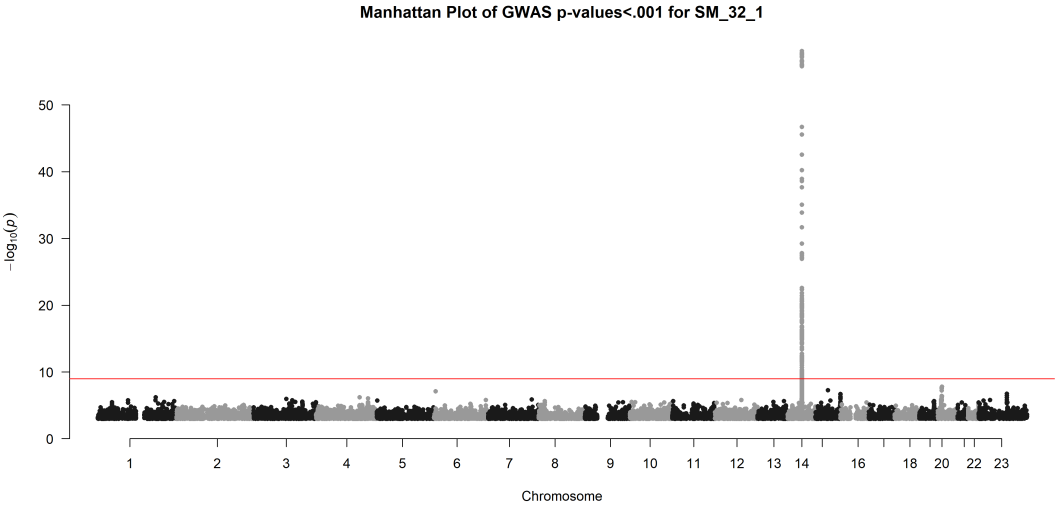


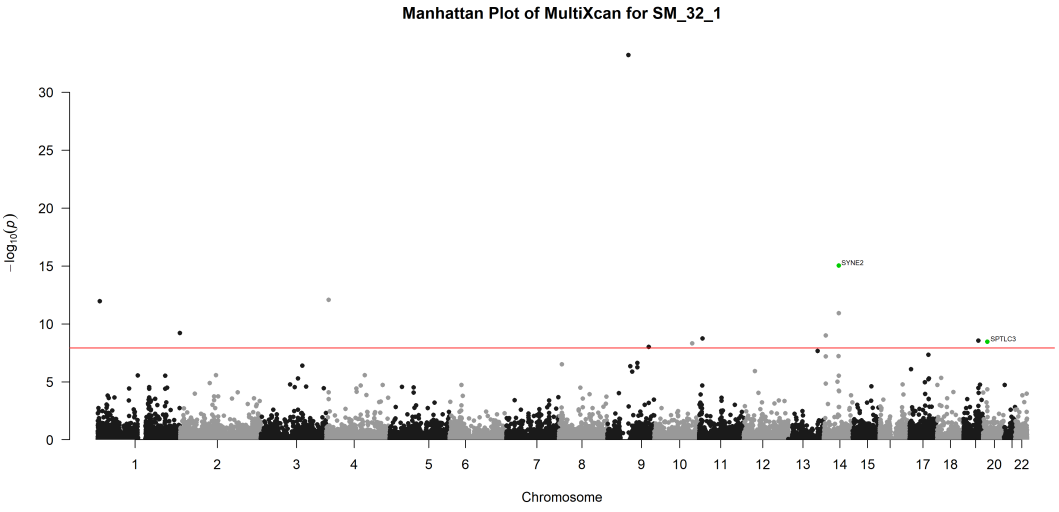
Chr20_12814090_13047914



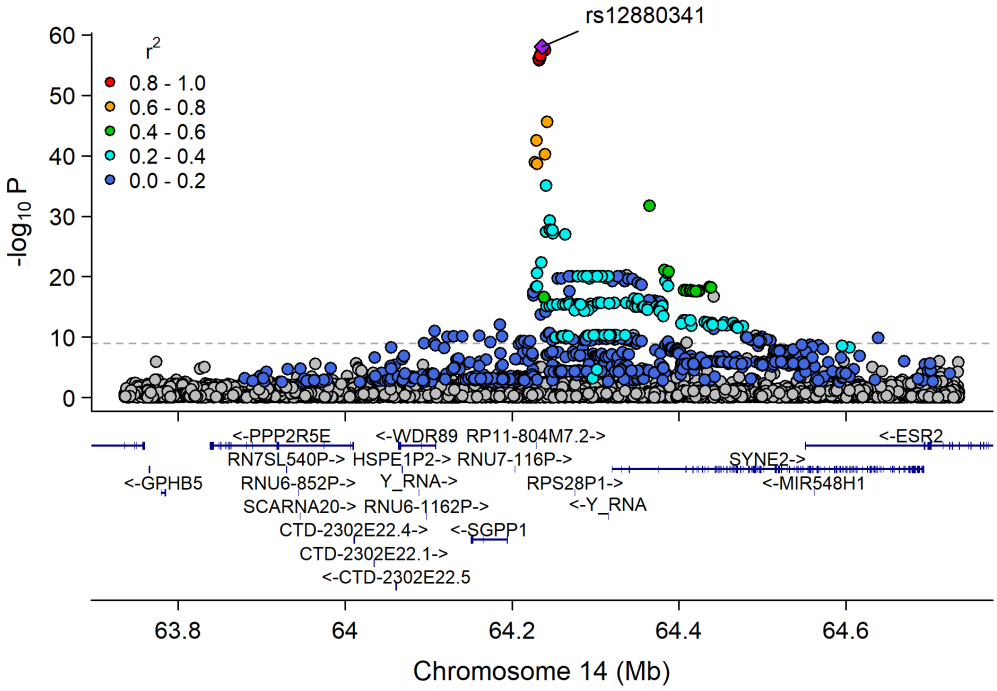
Sphingomyeline C32:1 (µg/dL)



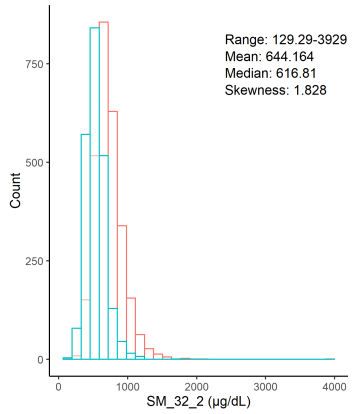


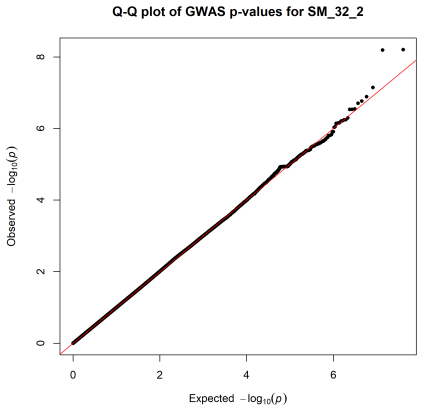


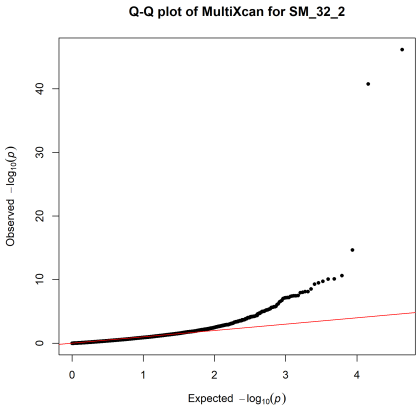
Chr14__63168899__65811394

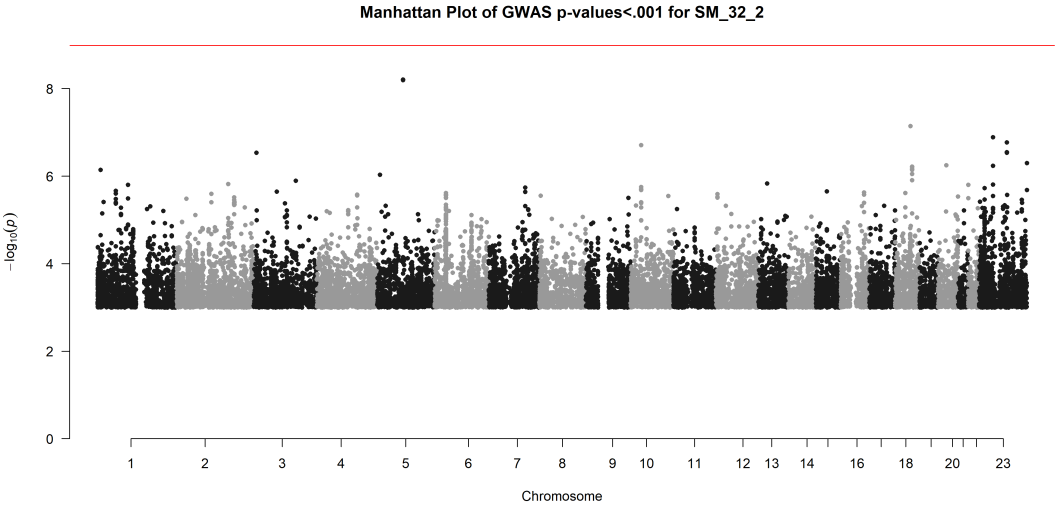


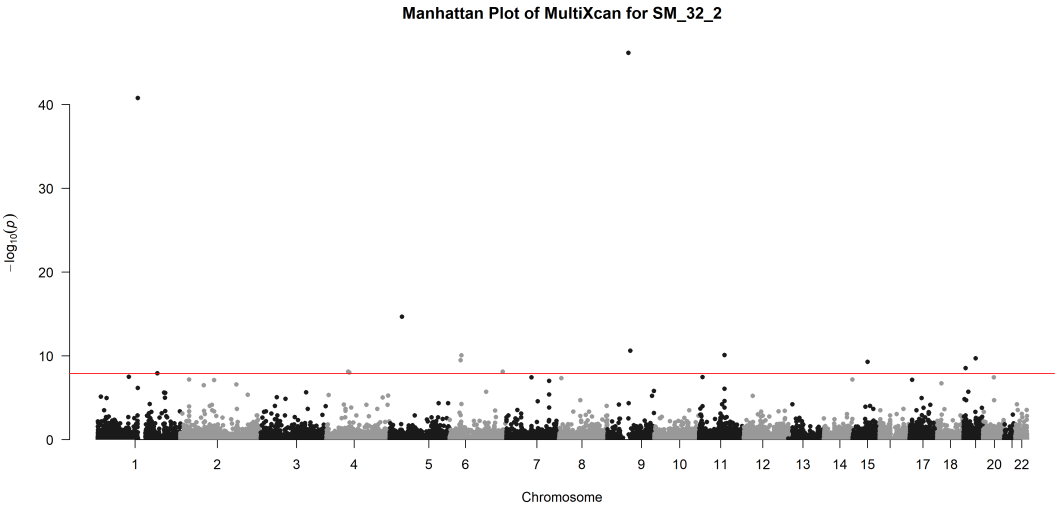
Sphingomyeline C32:2 (µg/dL)



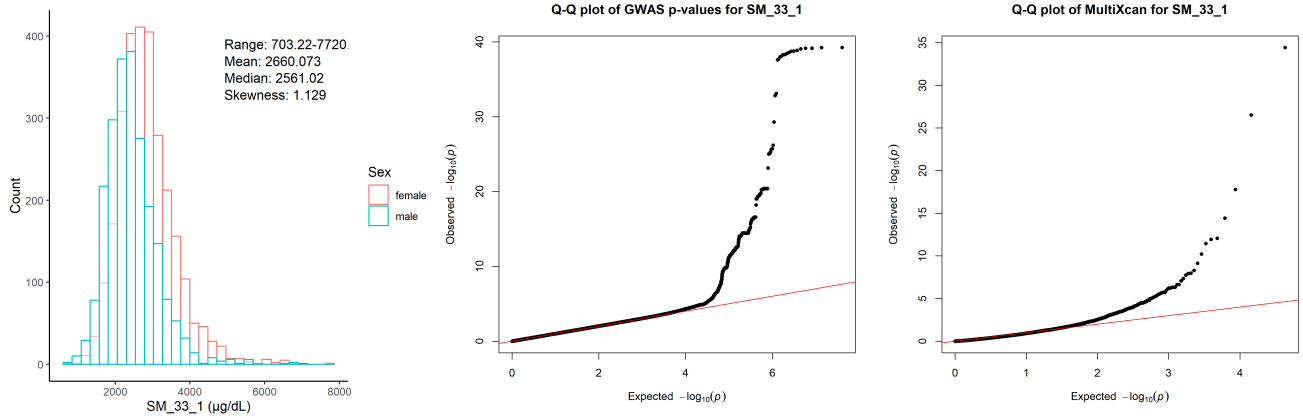


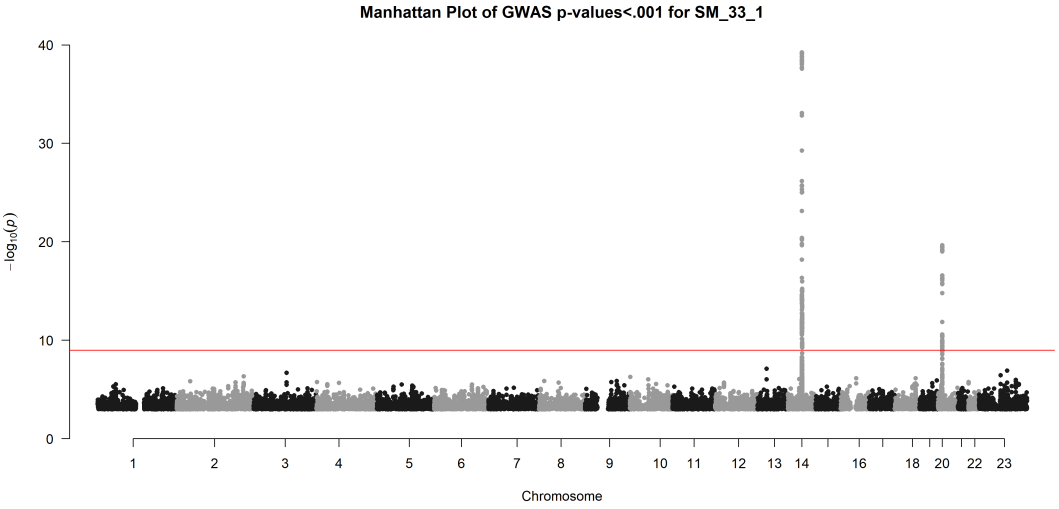


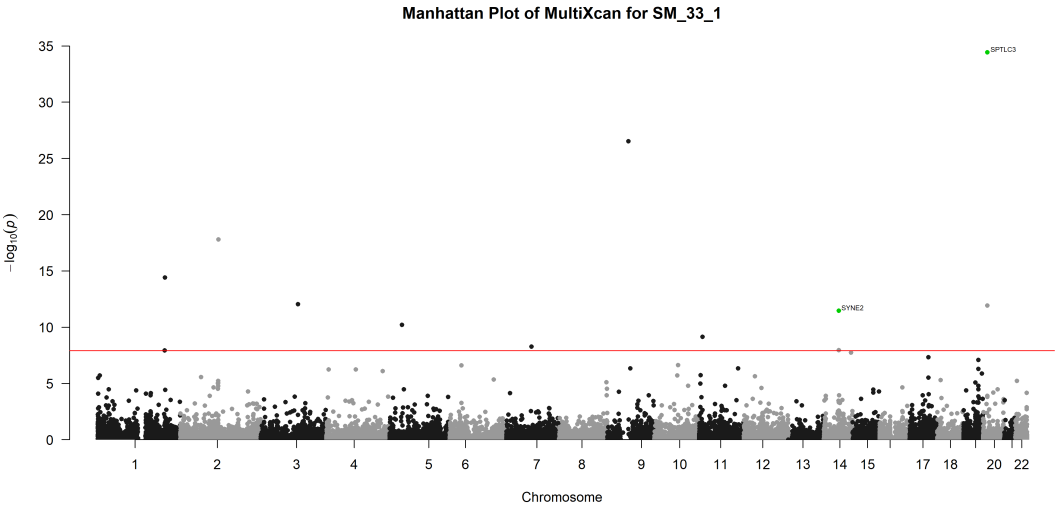




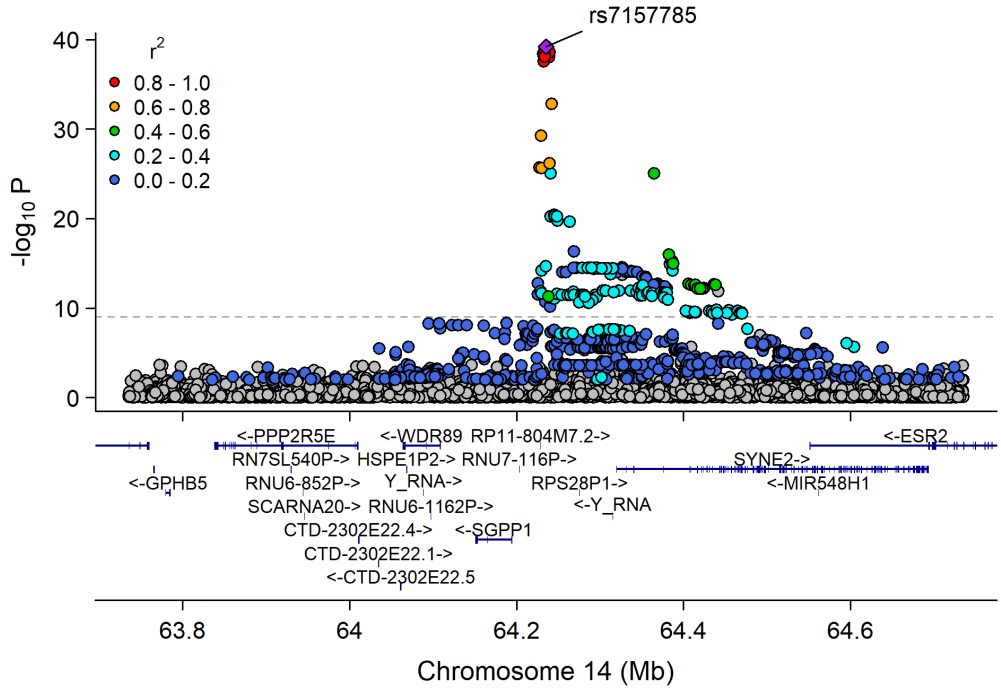
Sphingomyeline C33:1 (µg/dL)



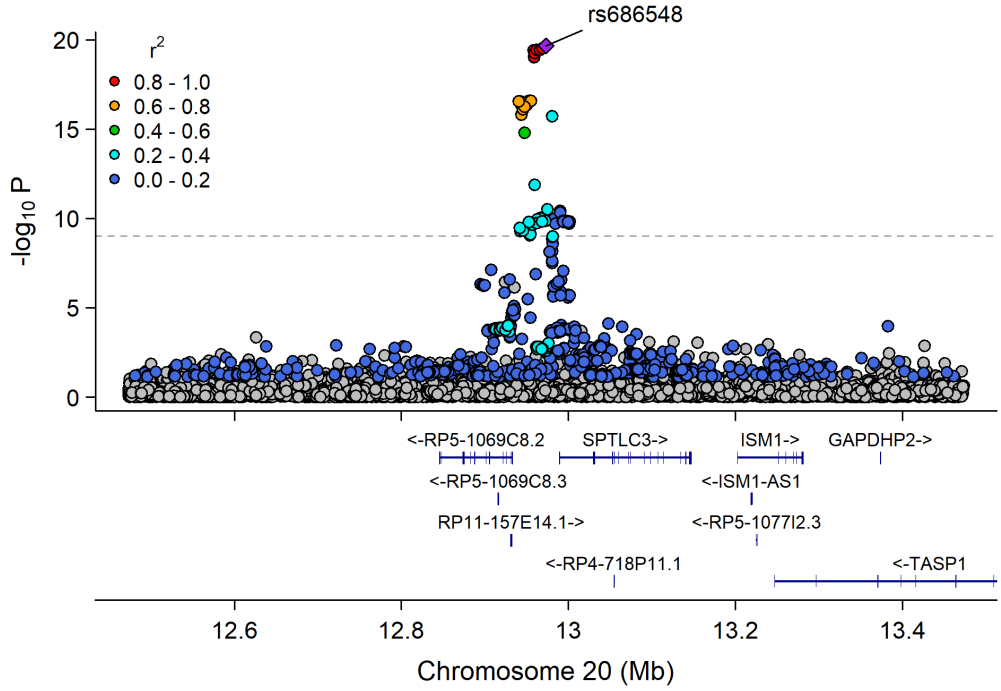




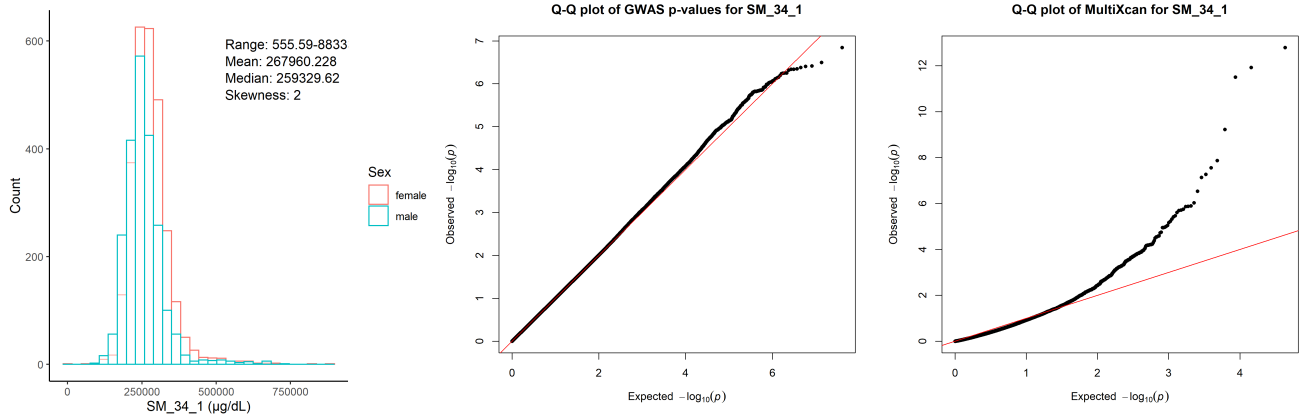
Chr14__63168899__65811394

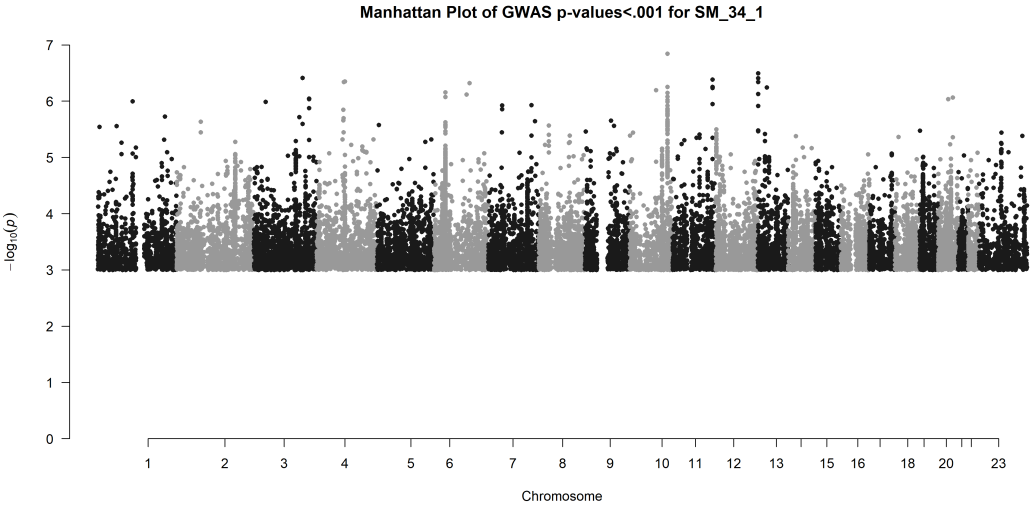


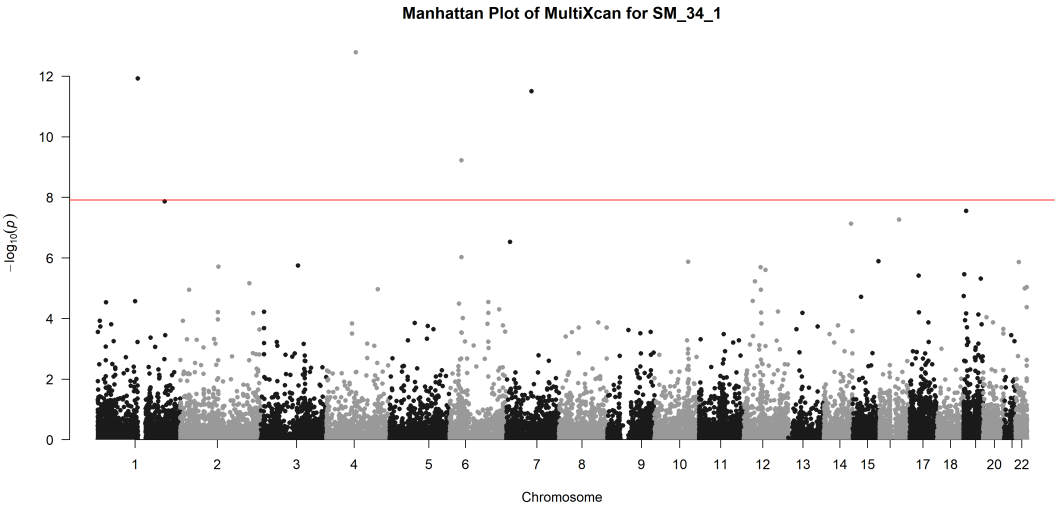
Chr20__12814090__13047914



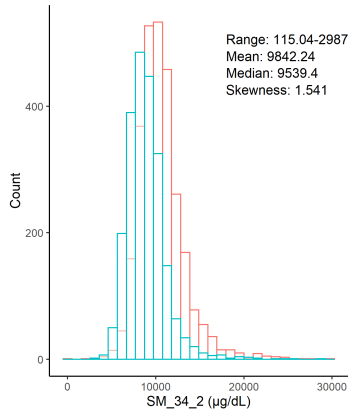
Sphingomyeline C34:1 (µg/dL)

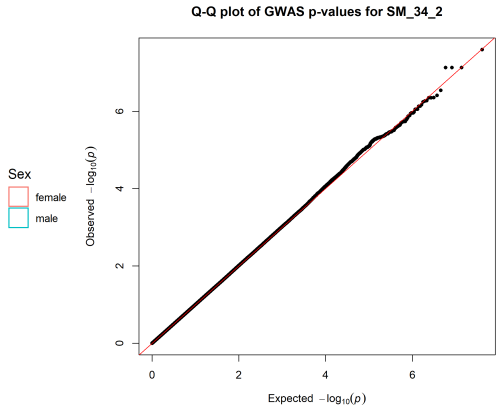


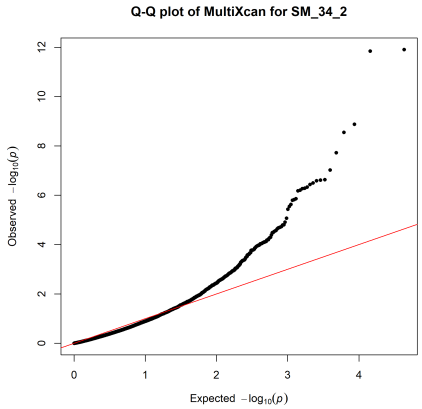


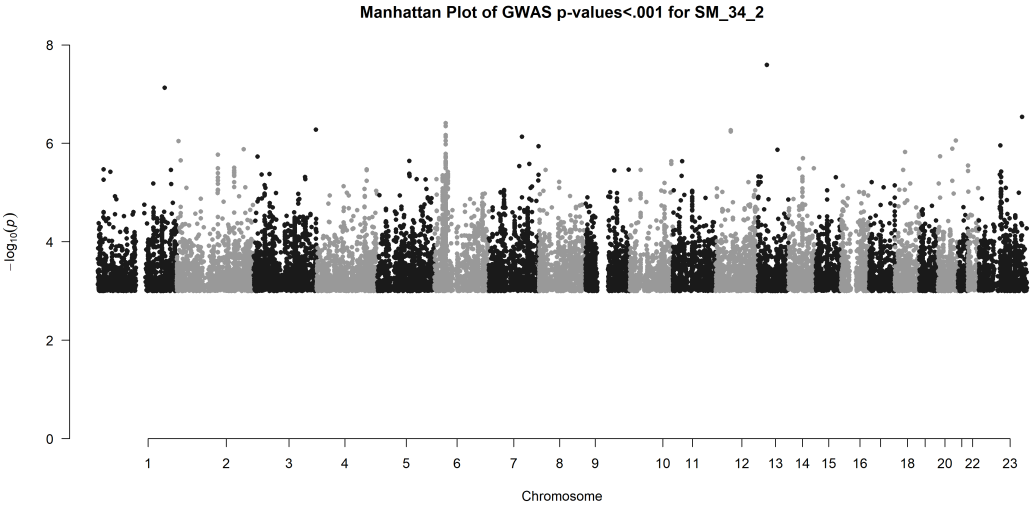


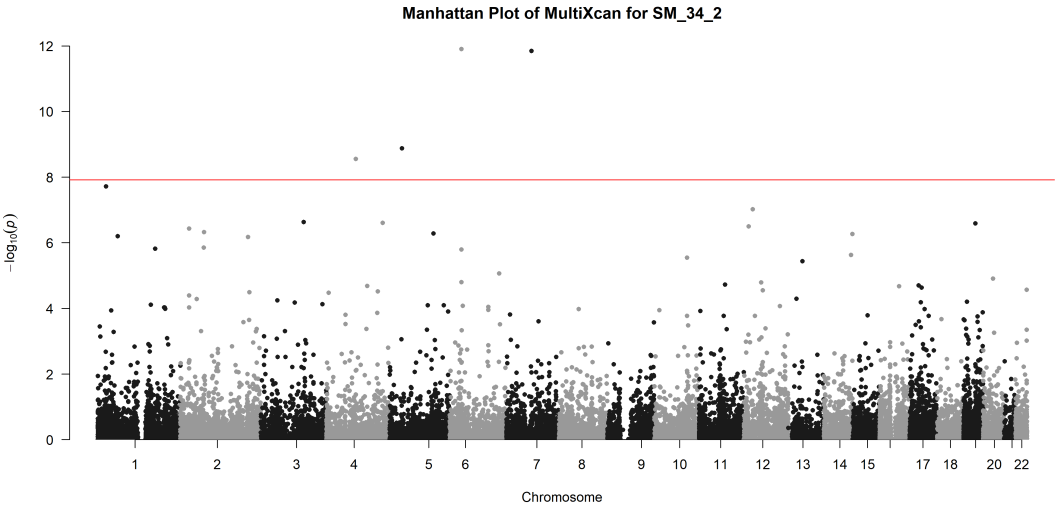
Sphingomyeline C34:2 (µg/dL)



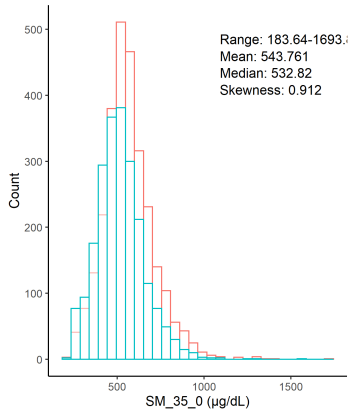


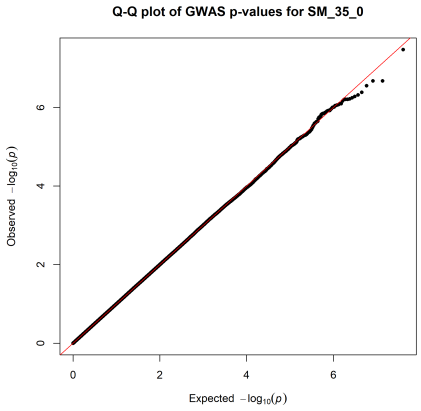


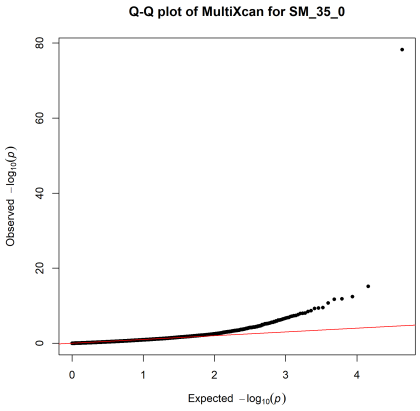


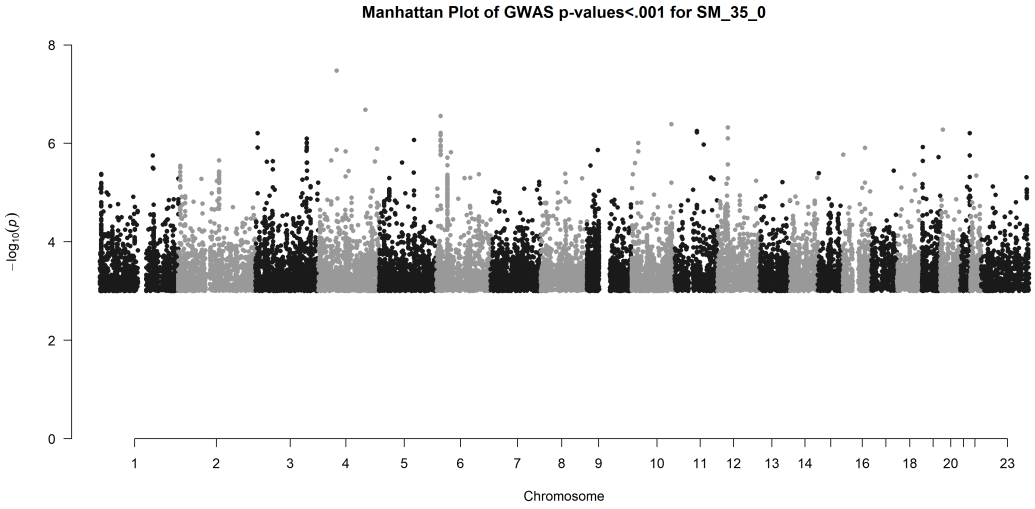


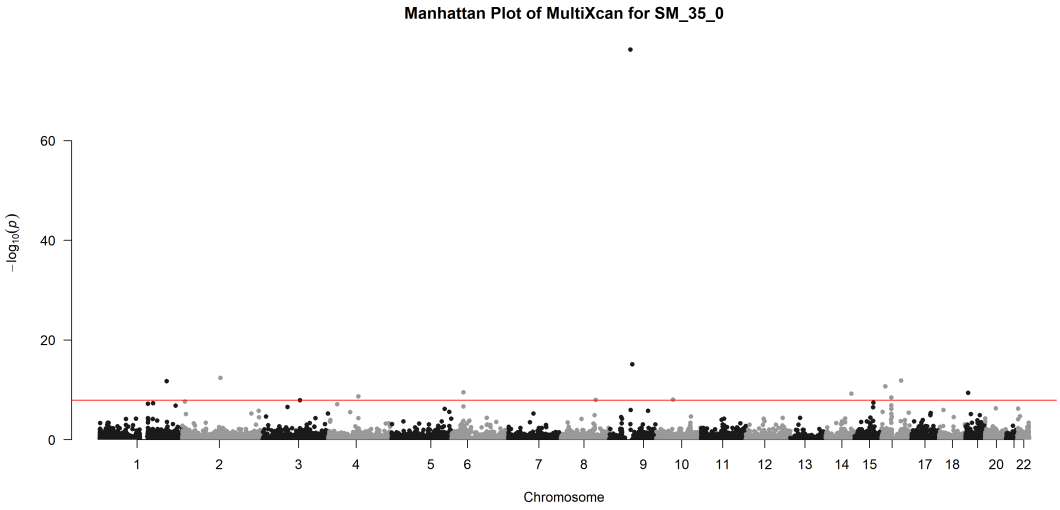
Sphingomyeline C35:0 (µg/dL)



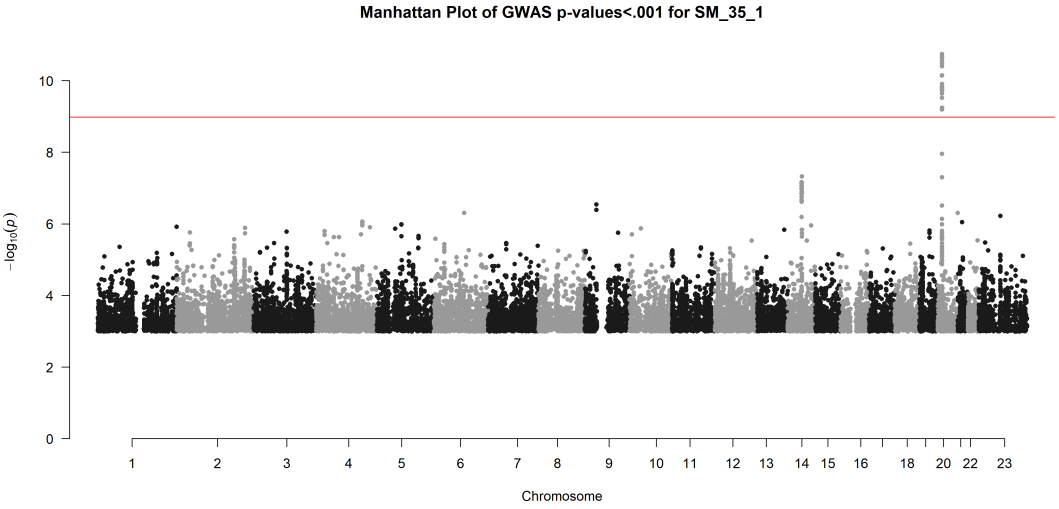


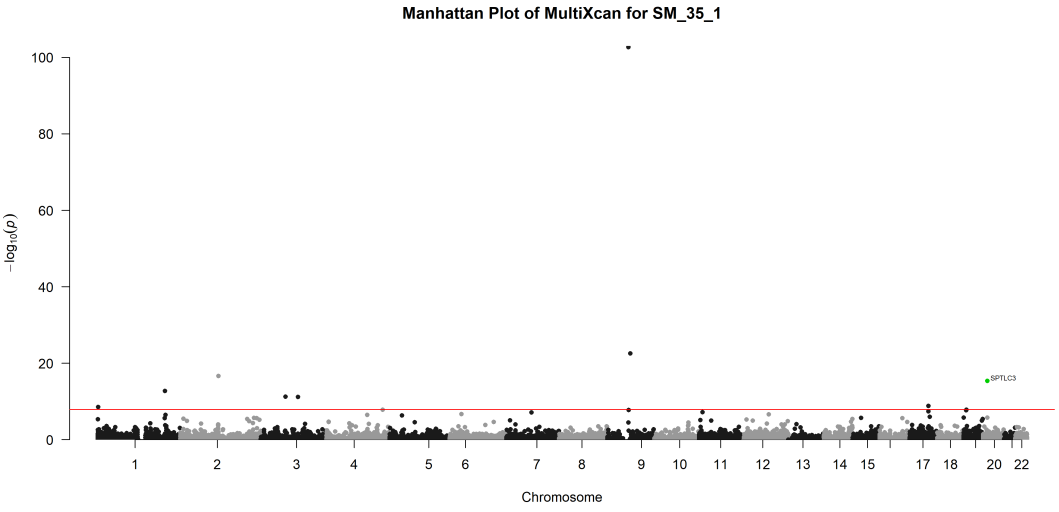




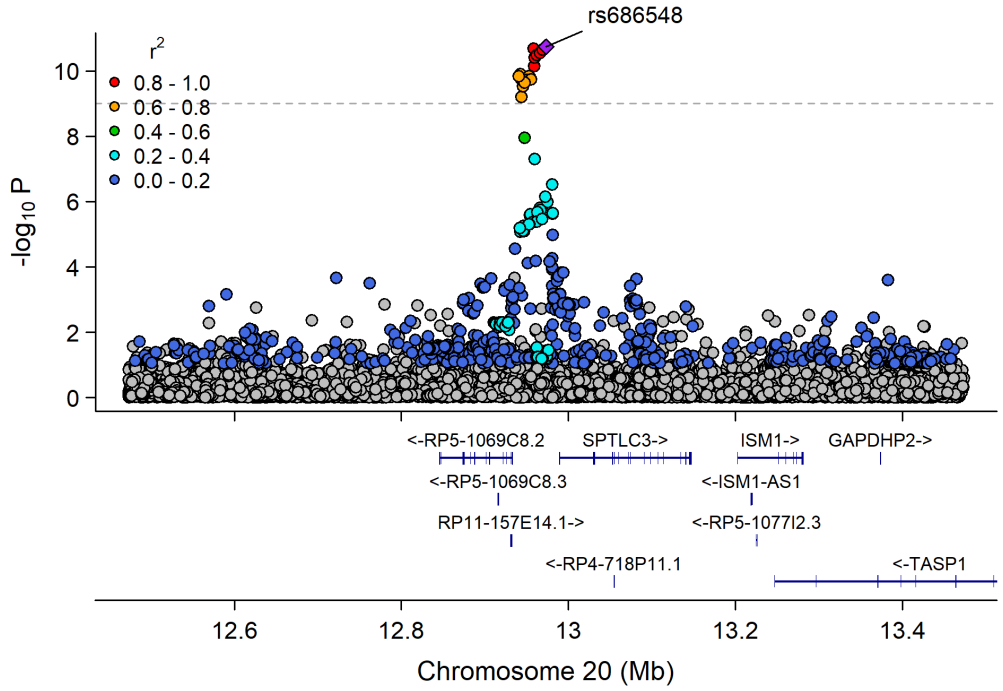


Sphingomyeline C35:1 (µg/dL)

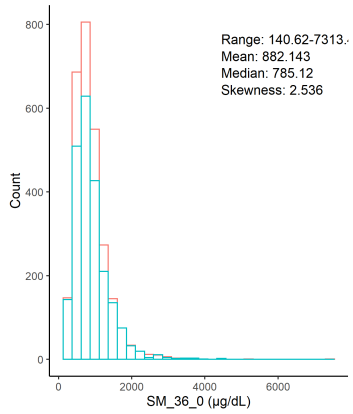


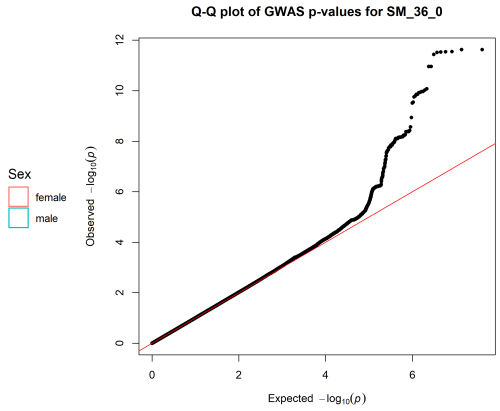


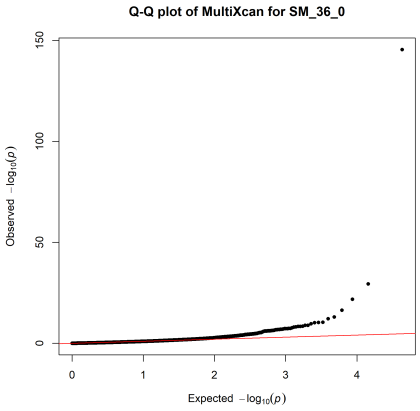
Chr20_12814090_13047914

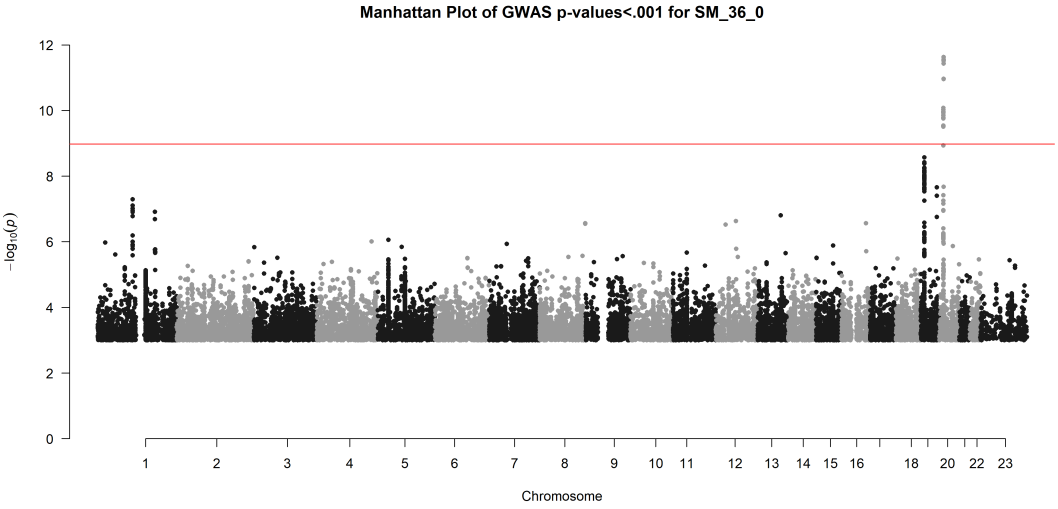


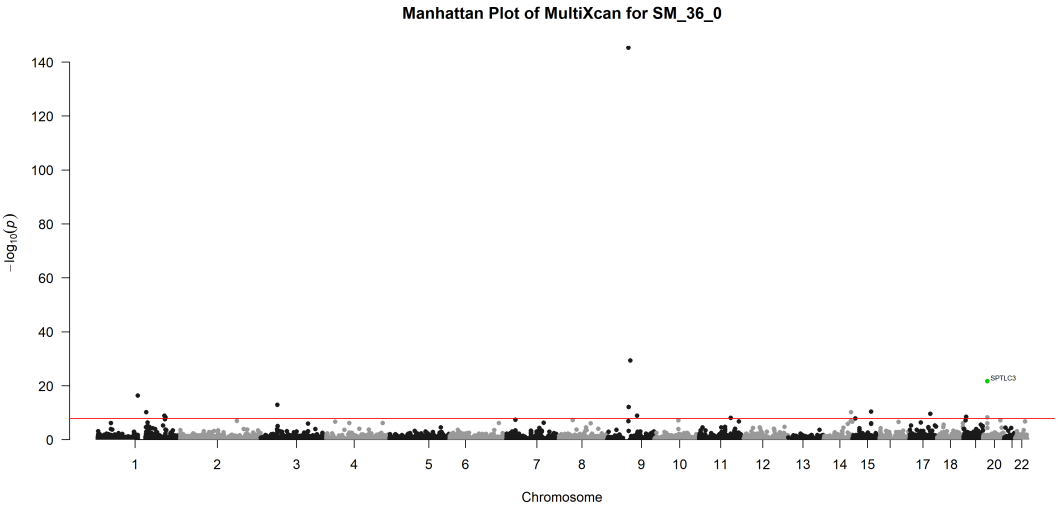
Sphingomyeline C36:0 (µg/dL)



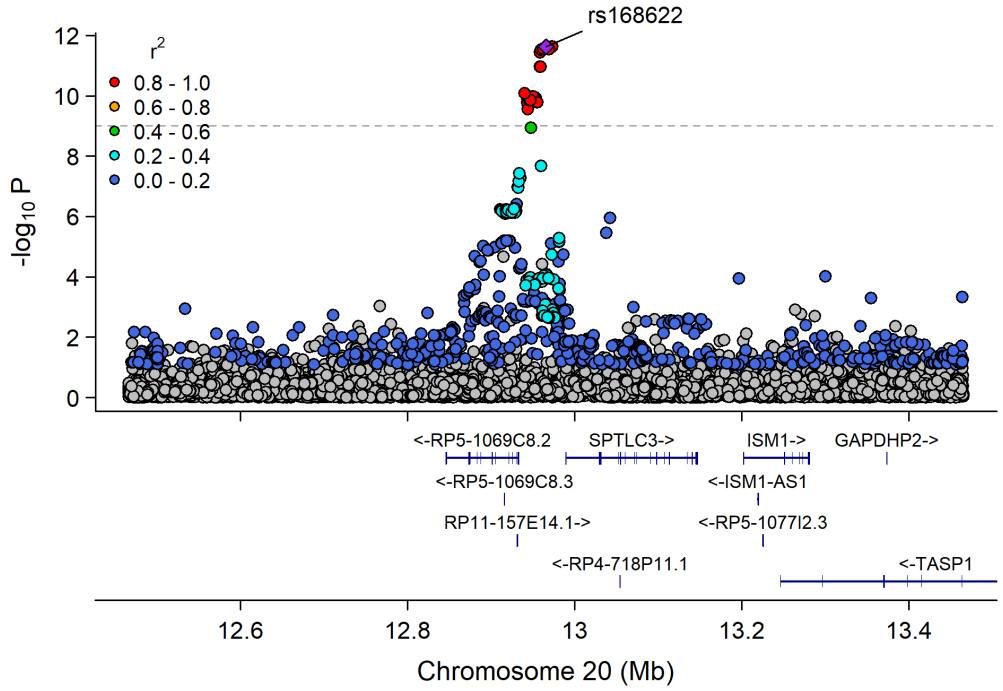




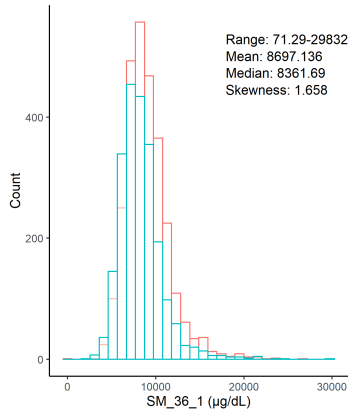


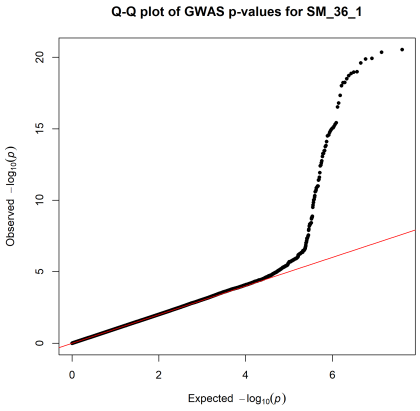


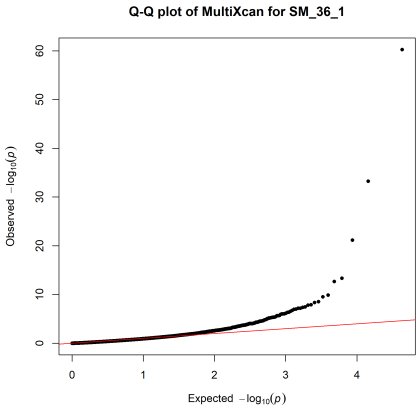
Chr20_12814090_13047914

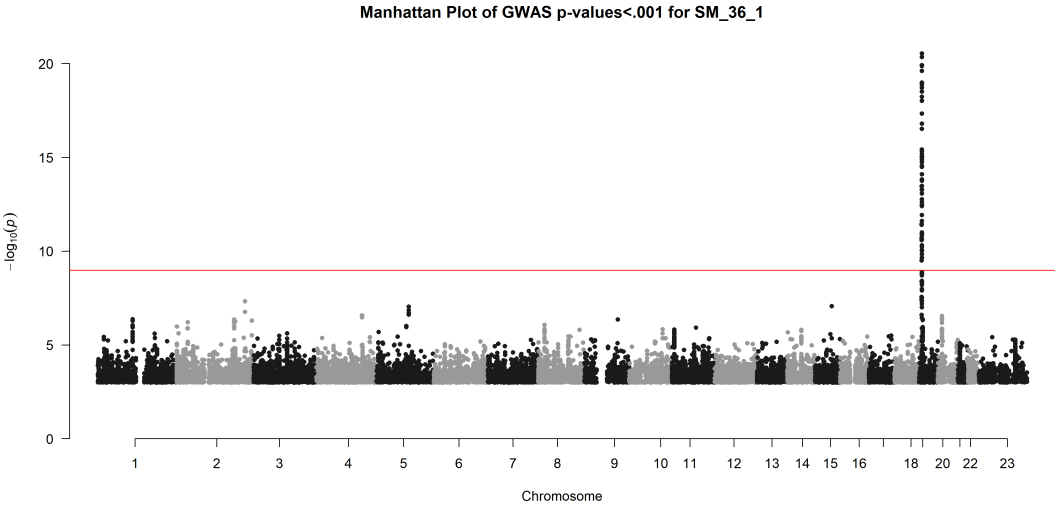


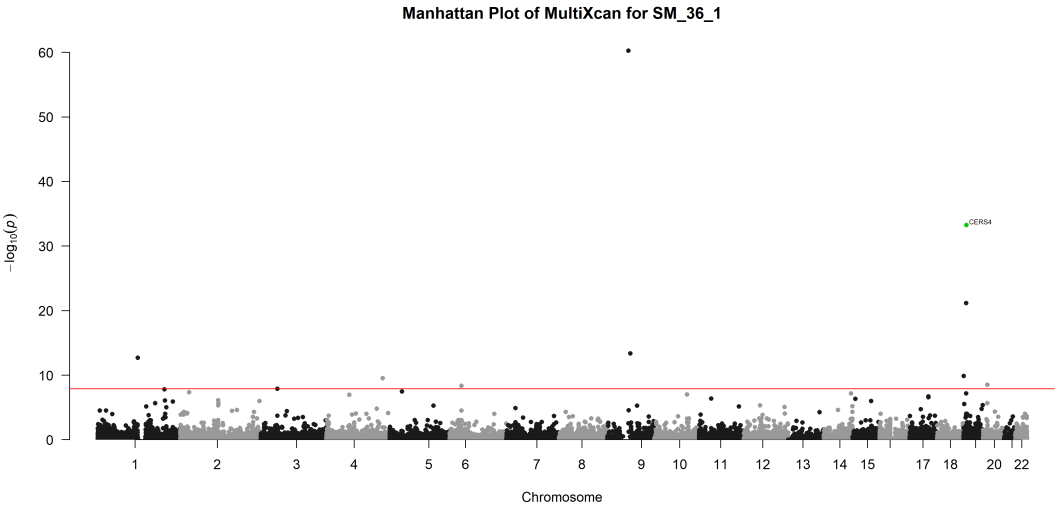
Sphingomyeline C36:1 (µg/dL)



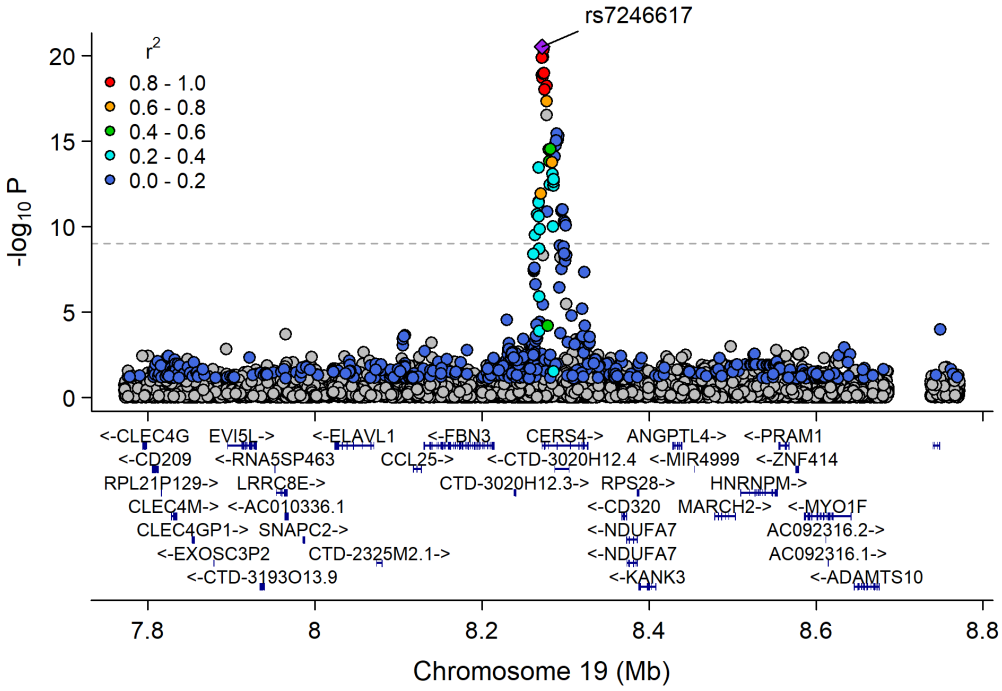




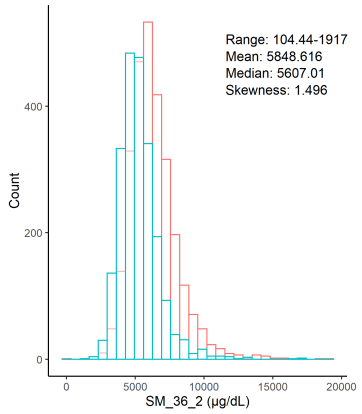


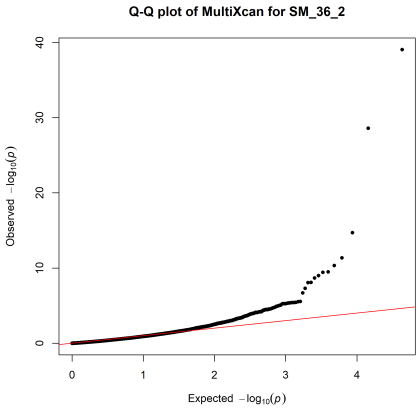


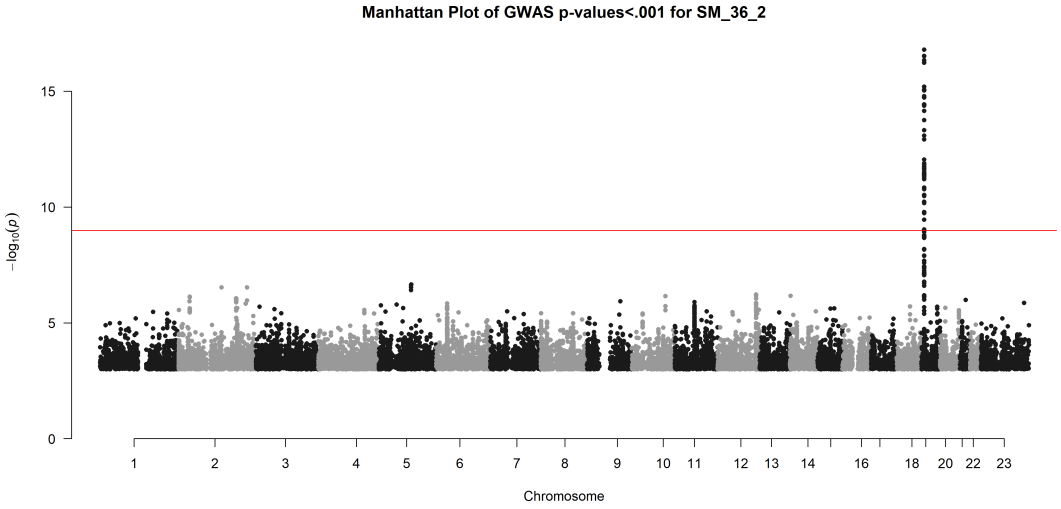
Chr19__7600546__9567827

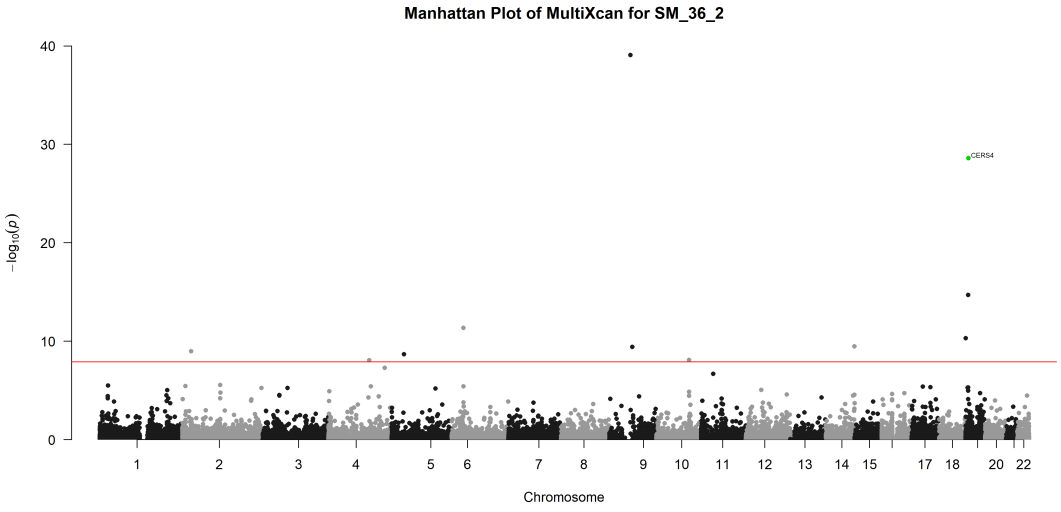


Sphingomyeline C36:2 (µg/dL)

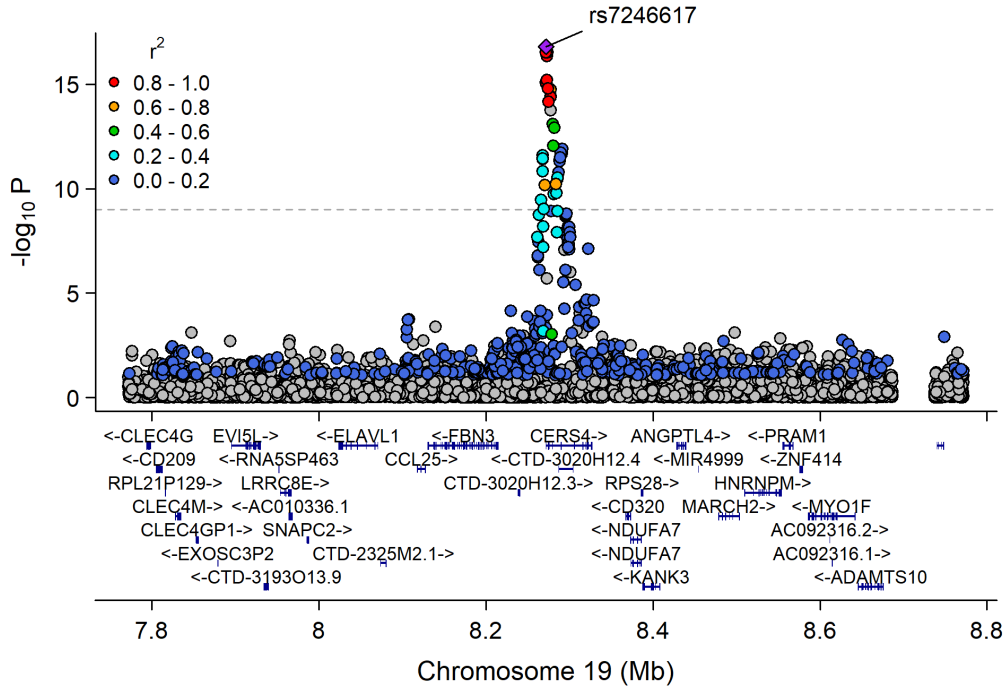




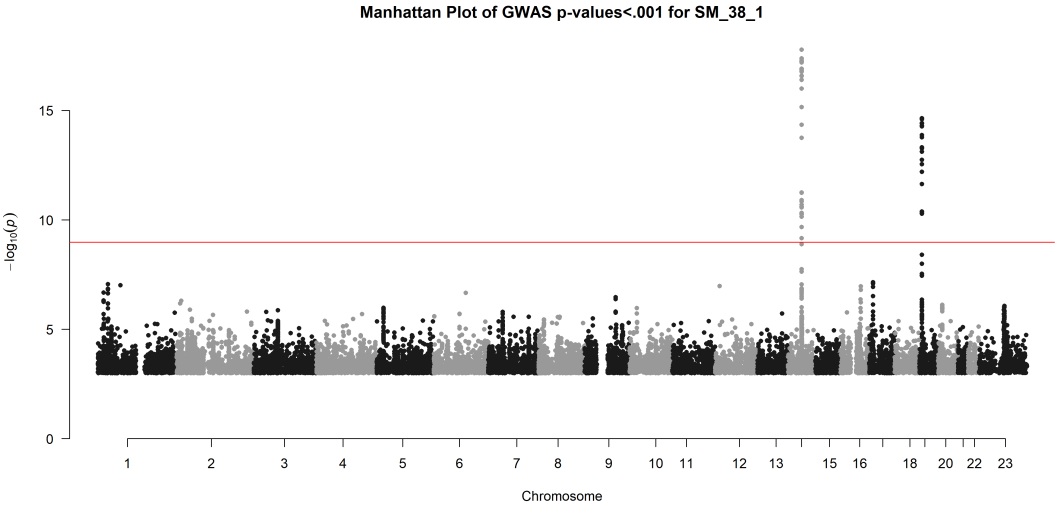


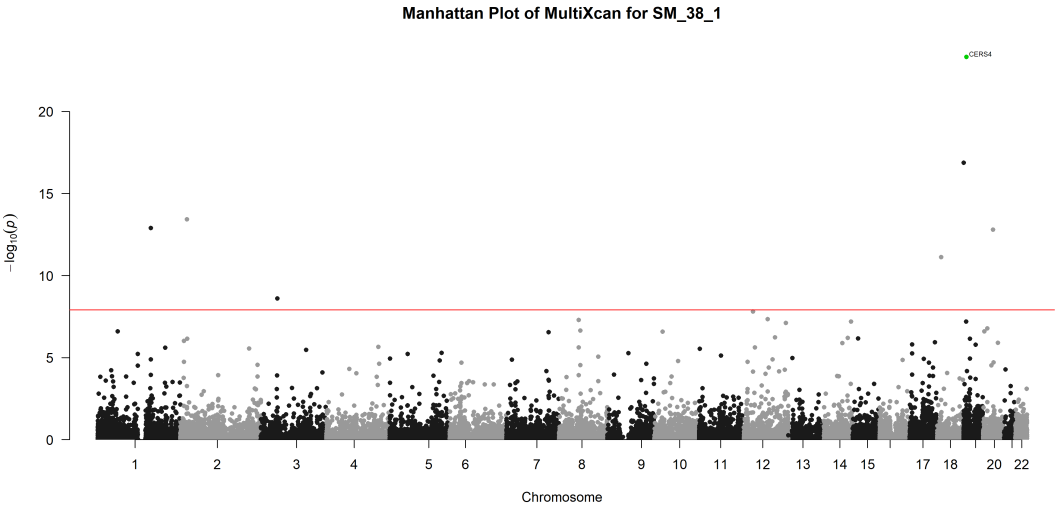


Chr19_7600546_9567827

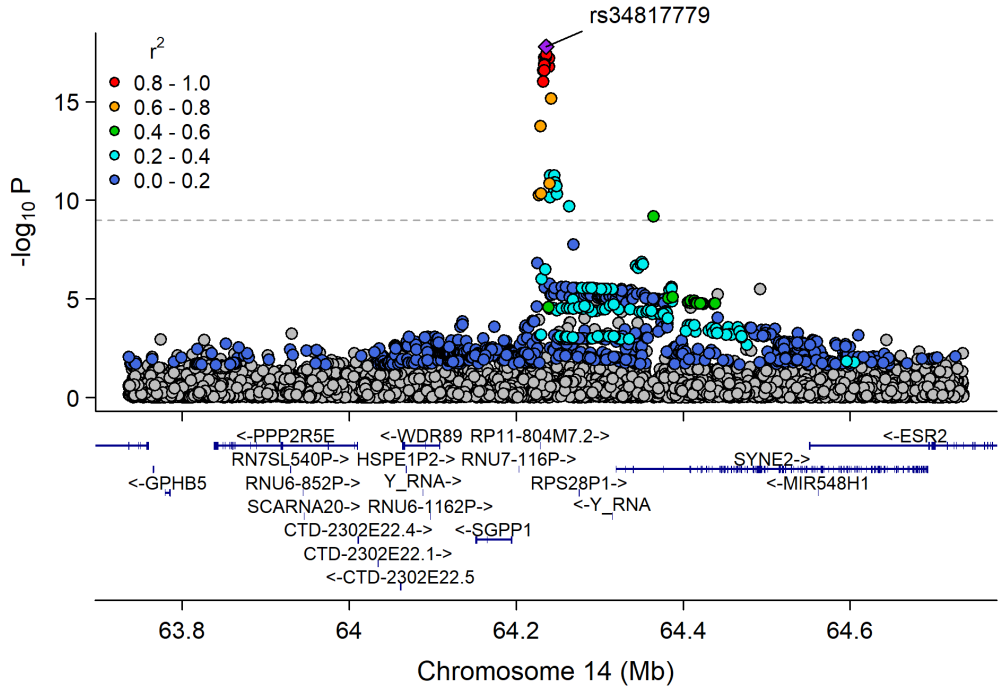


Sphingomyeline C38:1 (µg/dL)

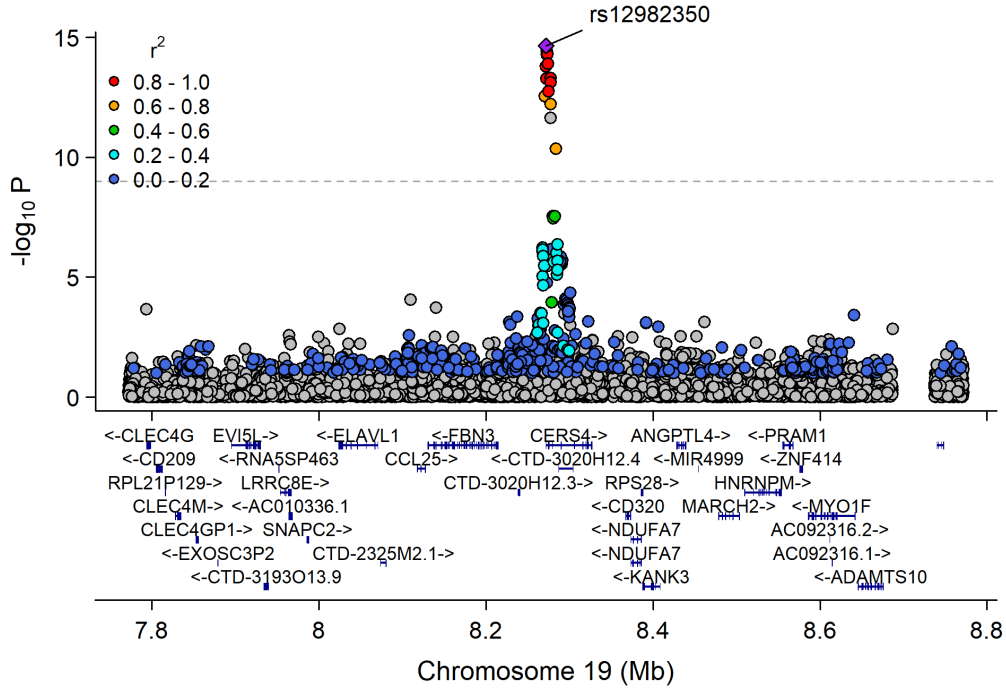




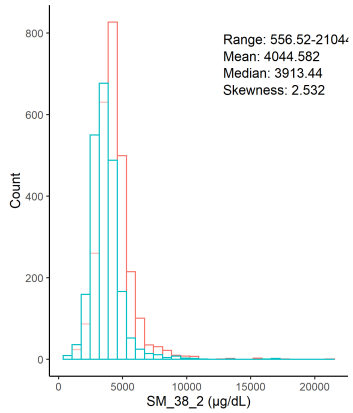
Chr14_63168899_65811394

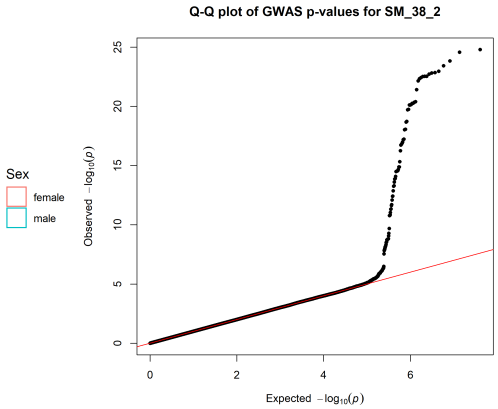


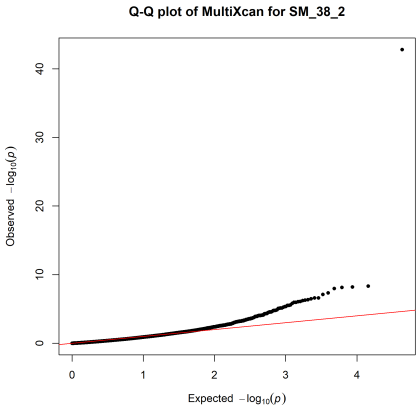
Chr19_7600546_9567827

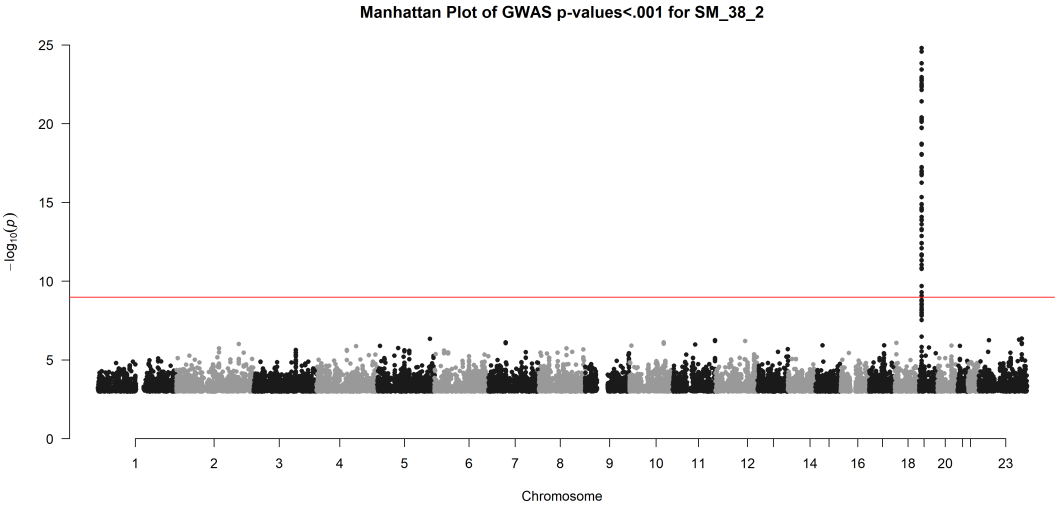


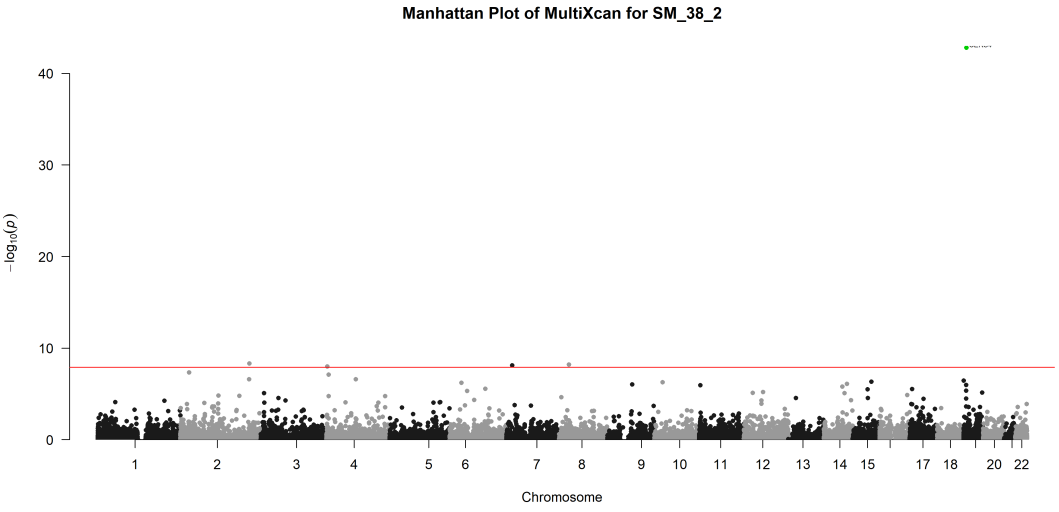
Sphingomyeline C38:2 (µg/dL)



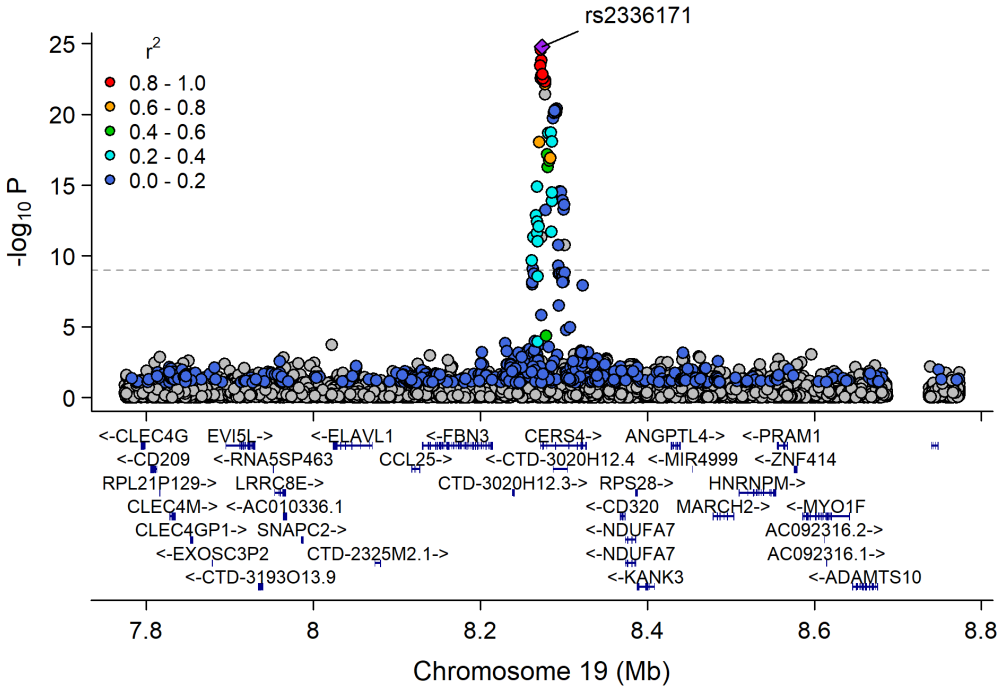




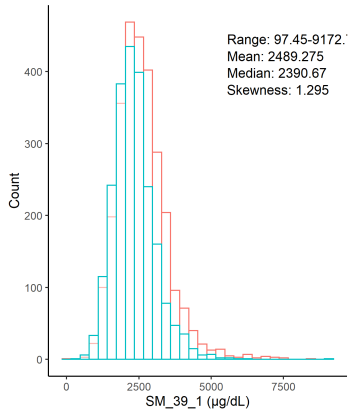


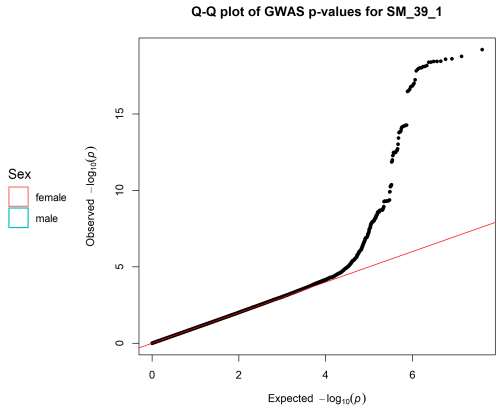


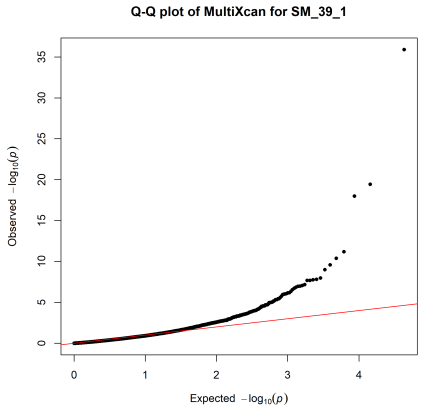
Chr19_7600546_9567827

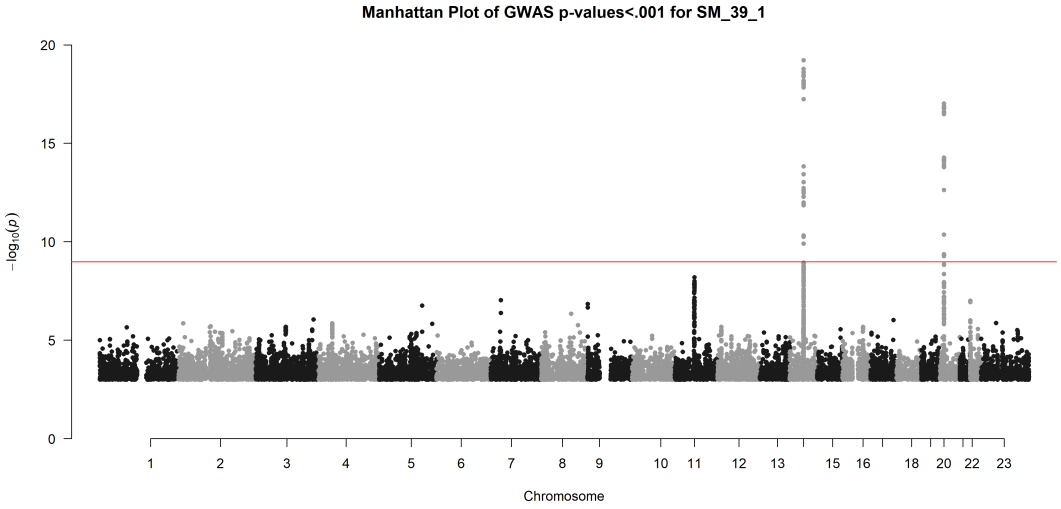


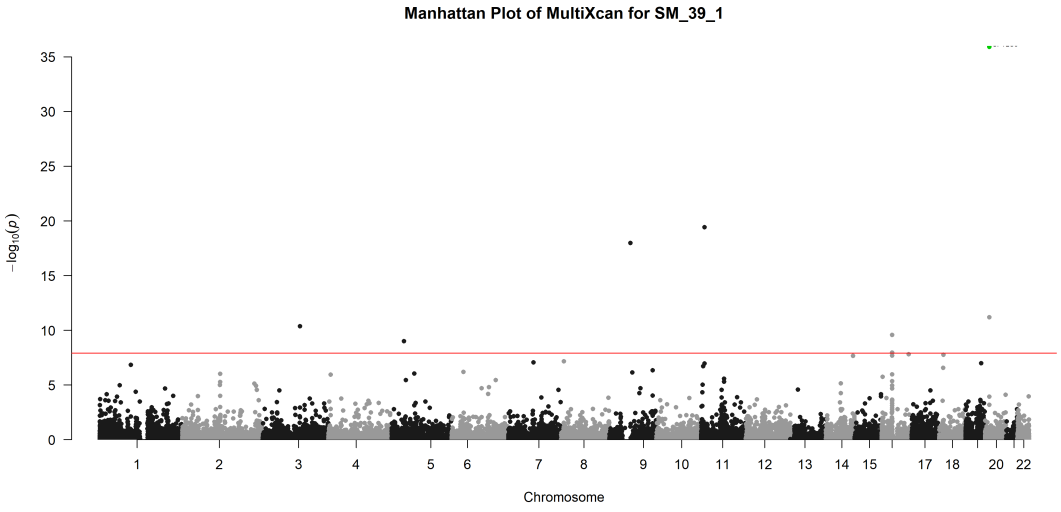
Sphingomyeline C39:1 (µg/dL)



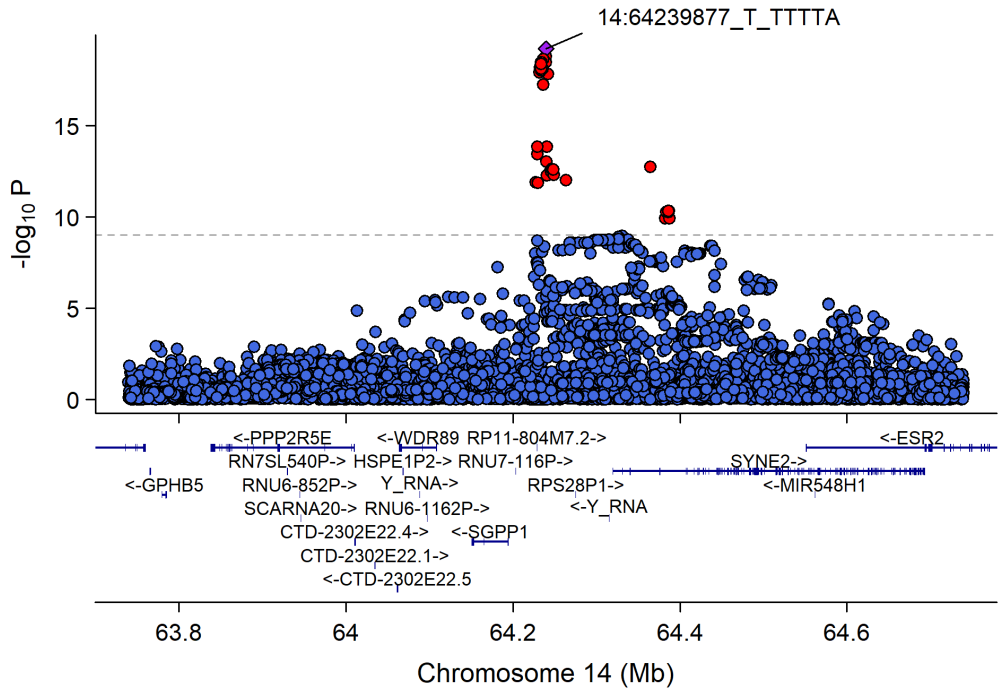




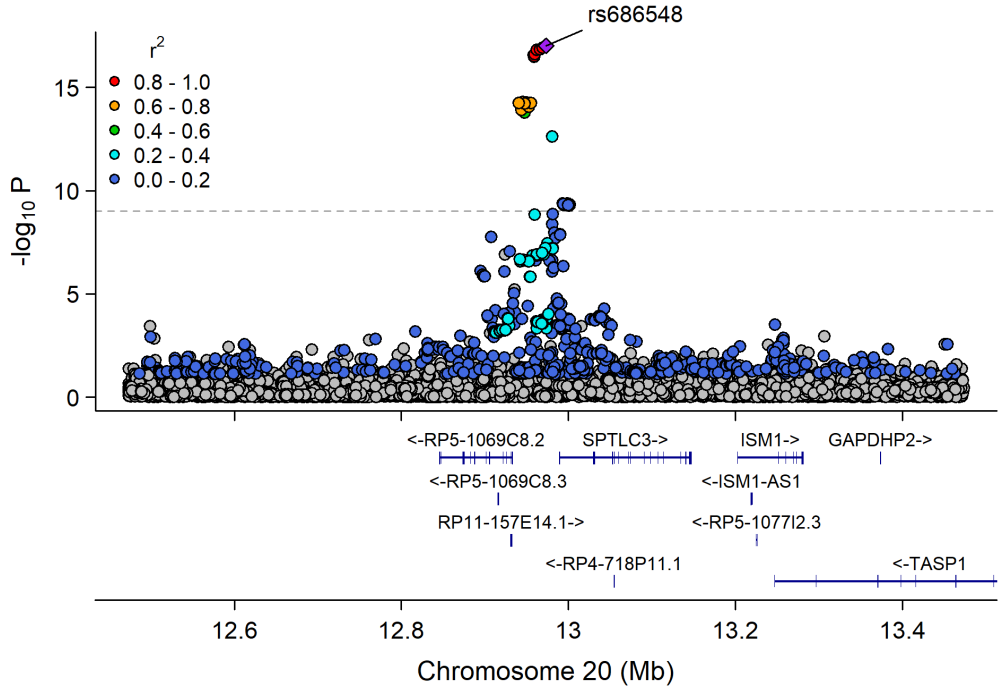




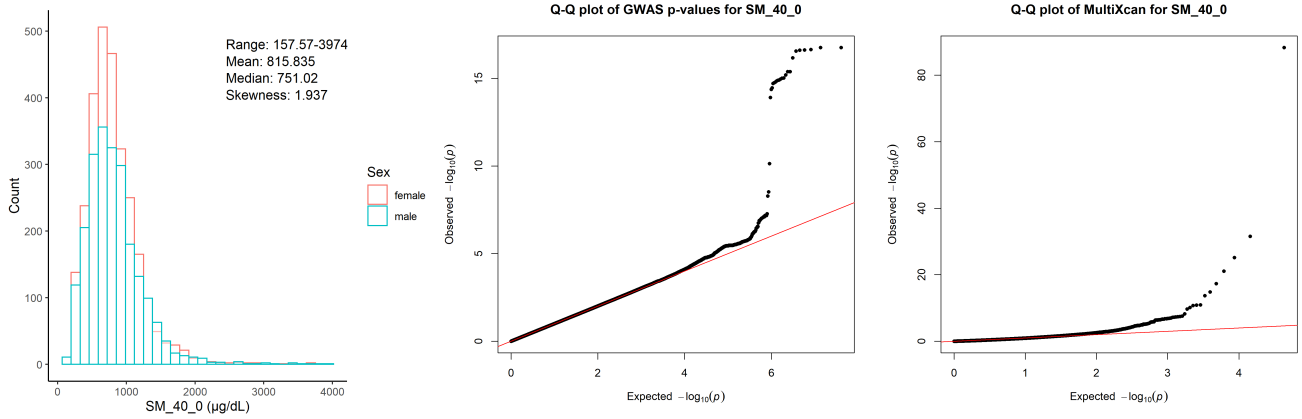
Chr14__63168899__65811394

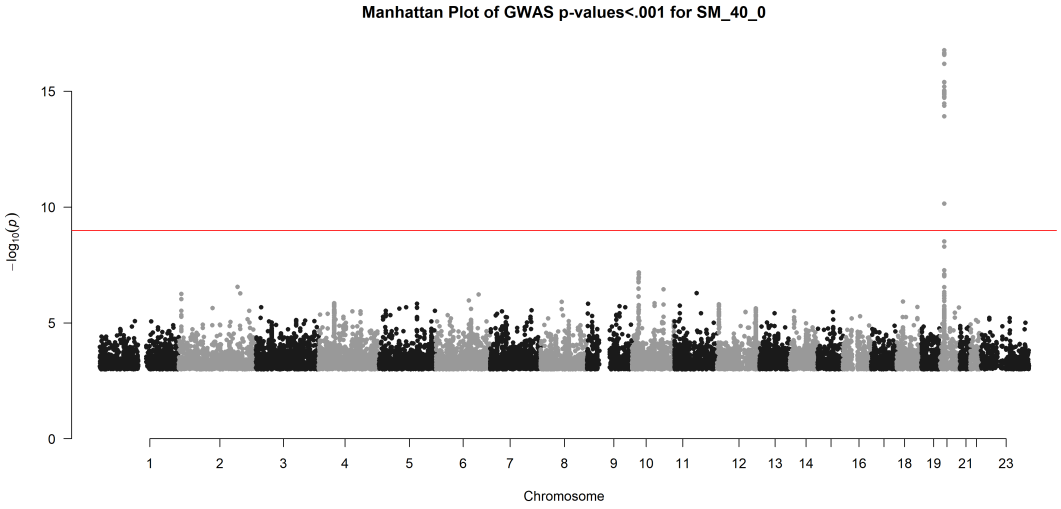


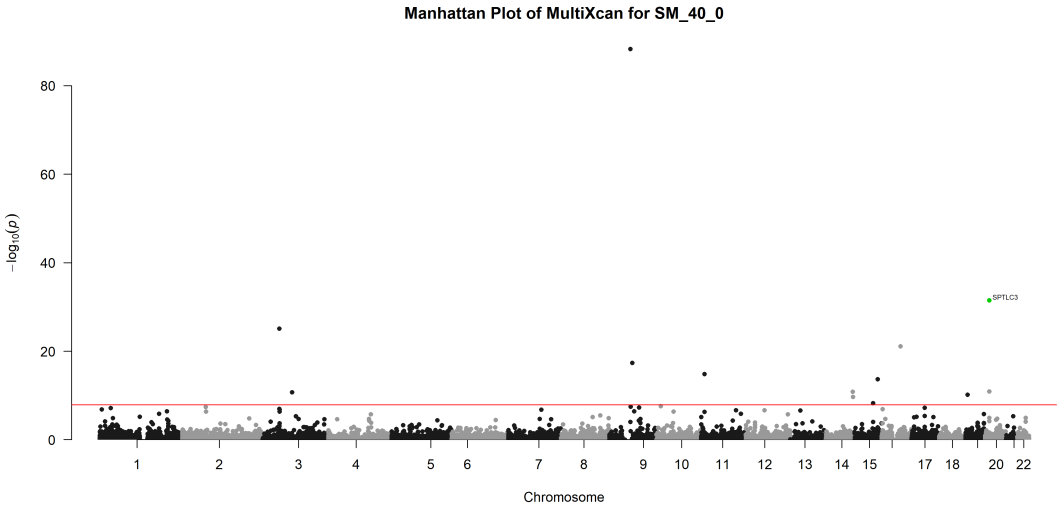
Chr20__12814090__13047914



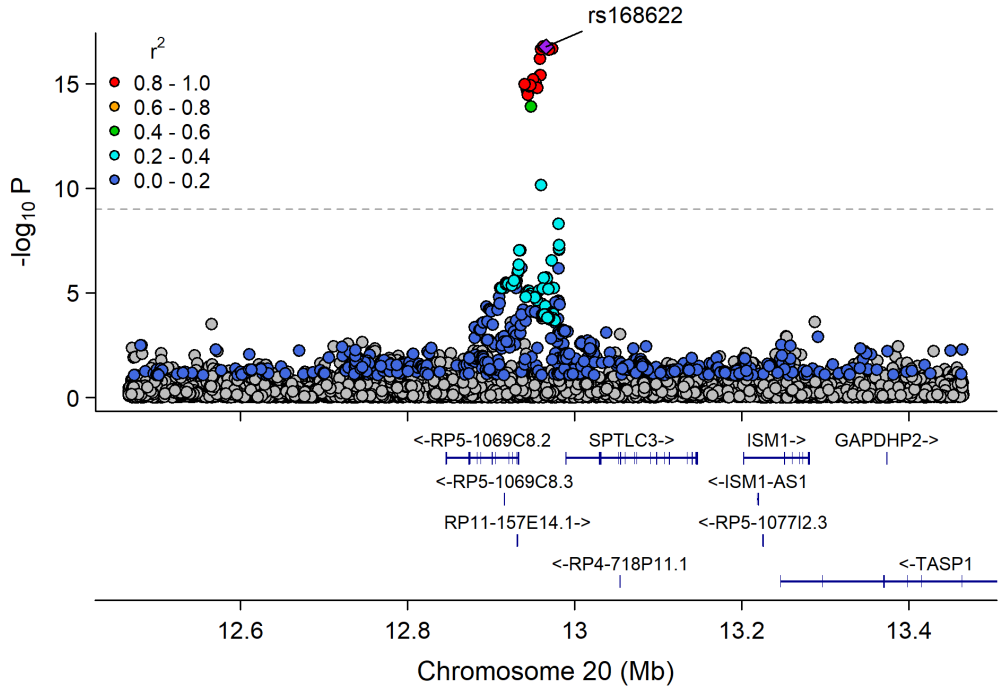
Sphingomyeline C40:0 (µg/dL)



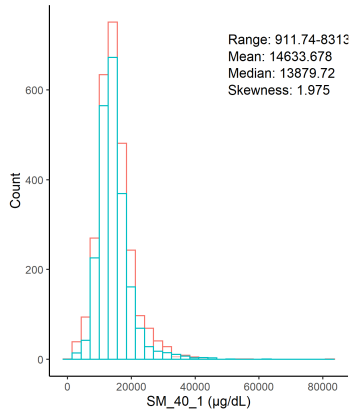


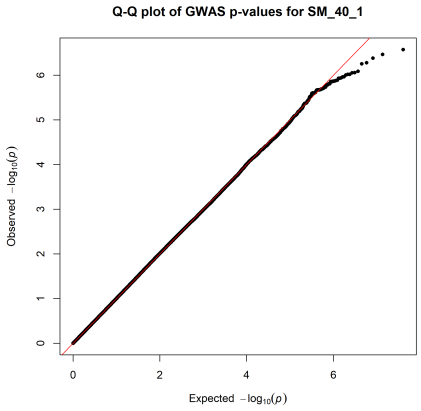


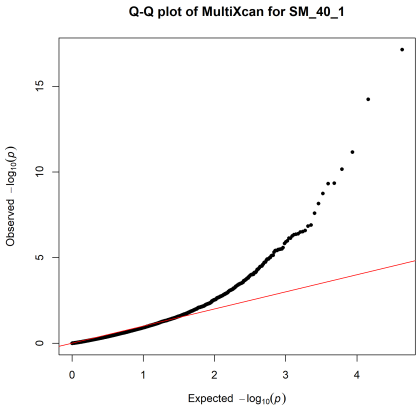
Chr20_12814090_13047914

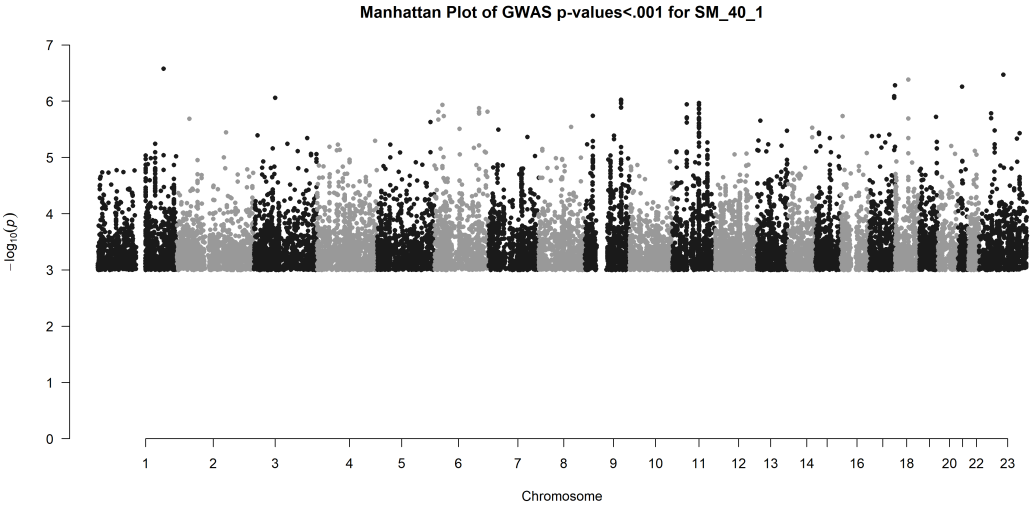


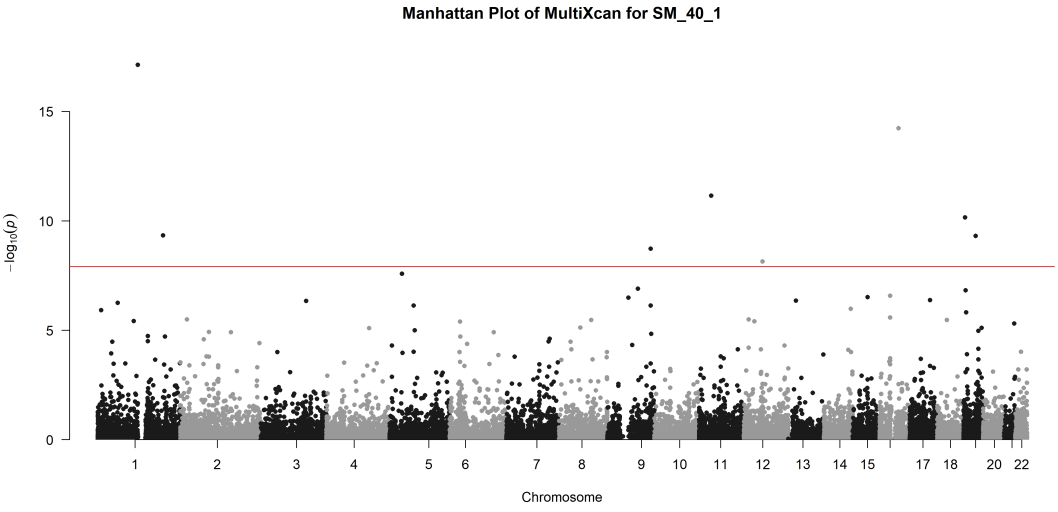
Sphingomyeline C40:1 (µg/dL)



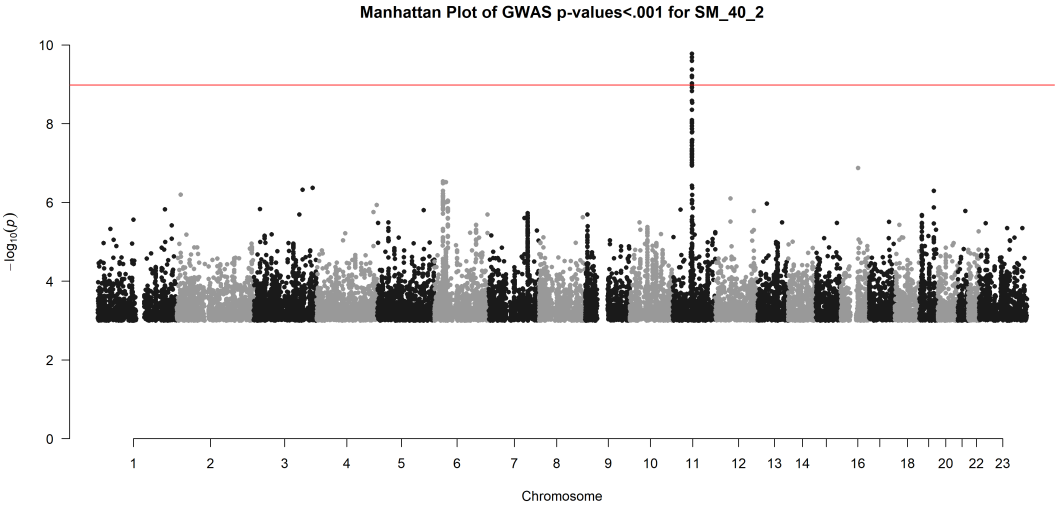


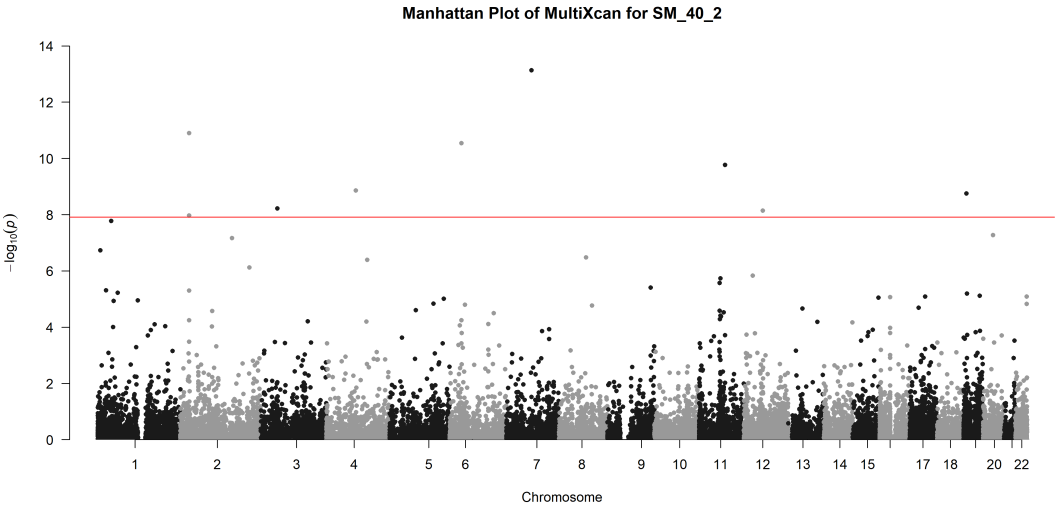




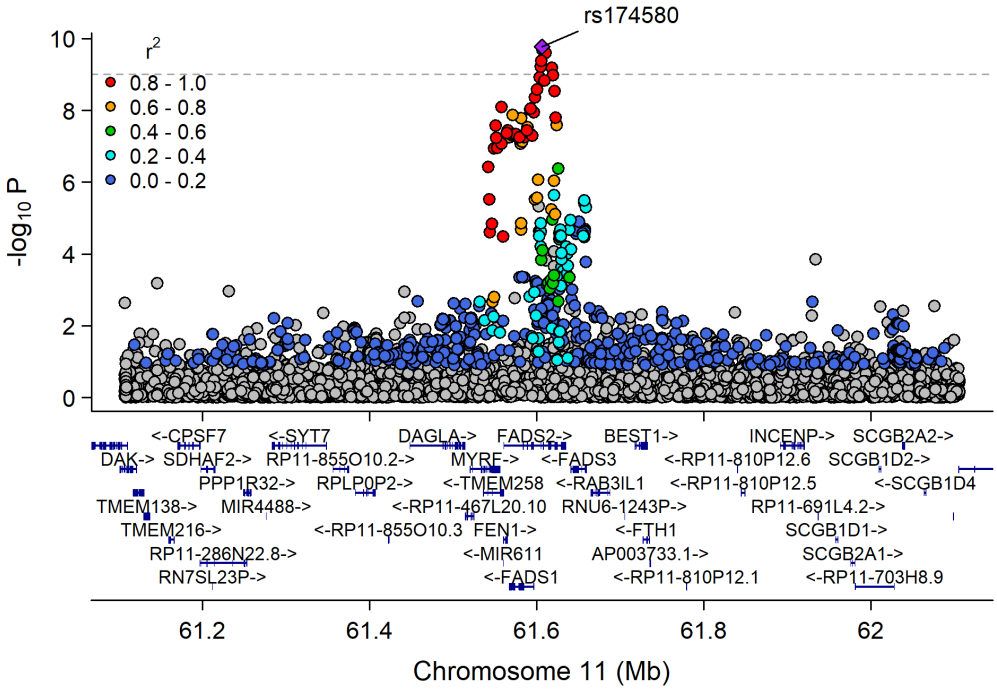


Sphingomyeline C40:2 (µg/dL)

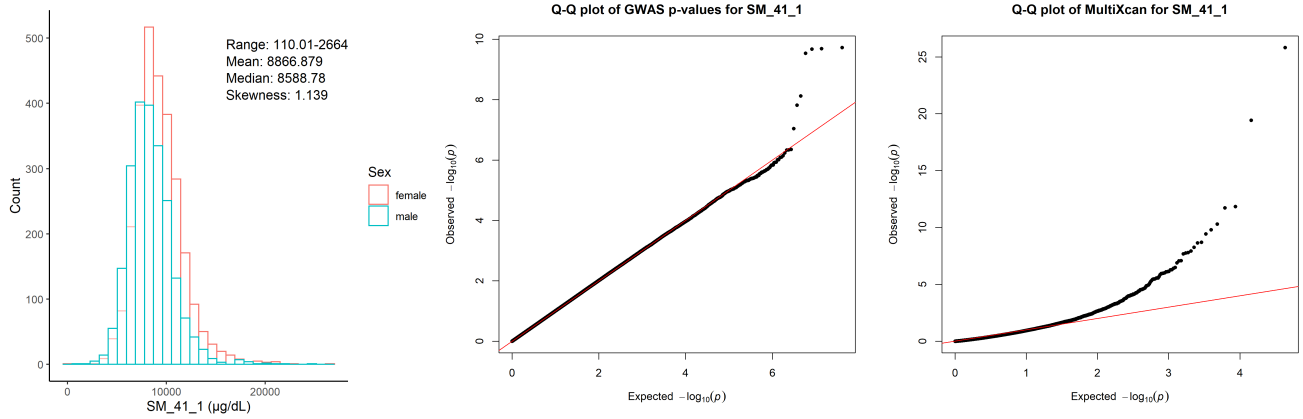


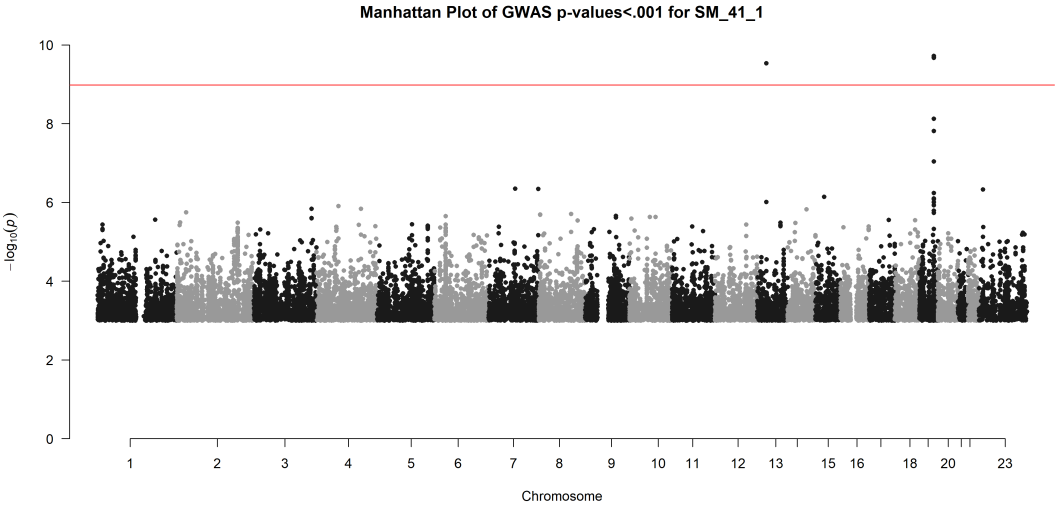


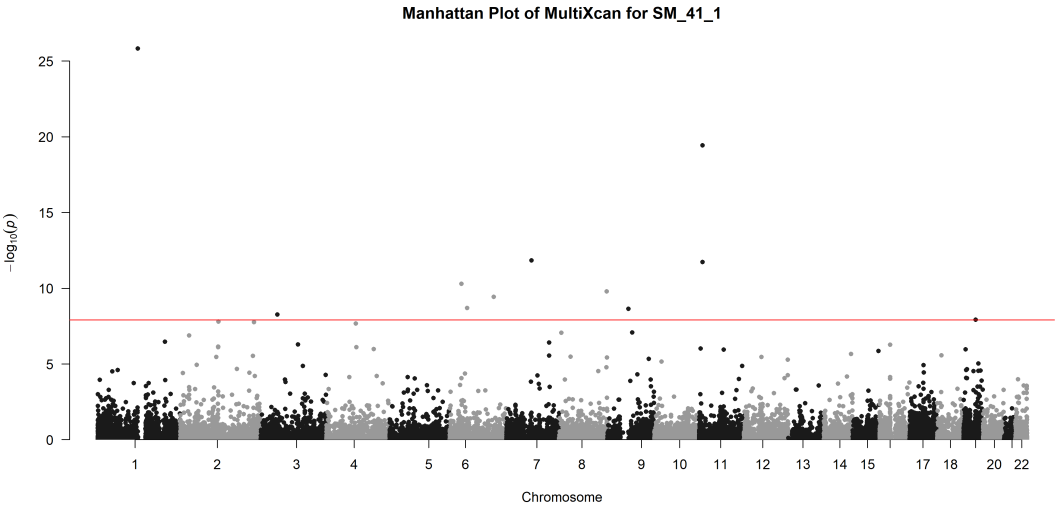
Chr11__59978355__62914375



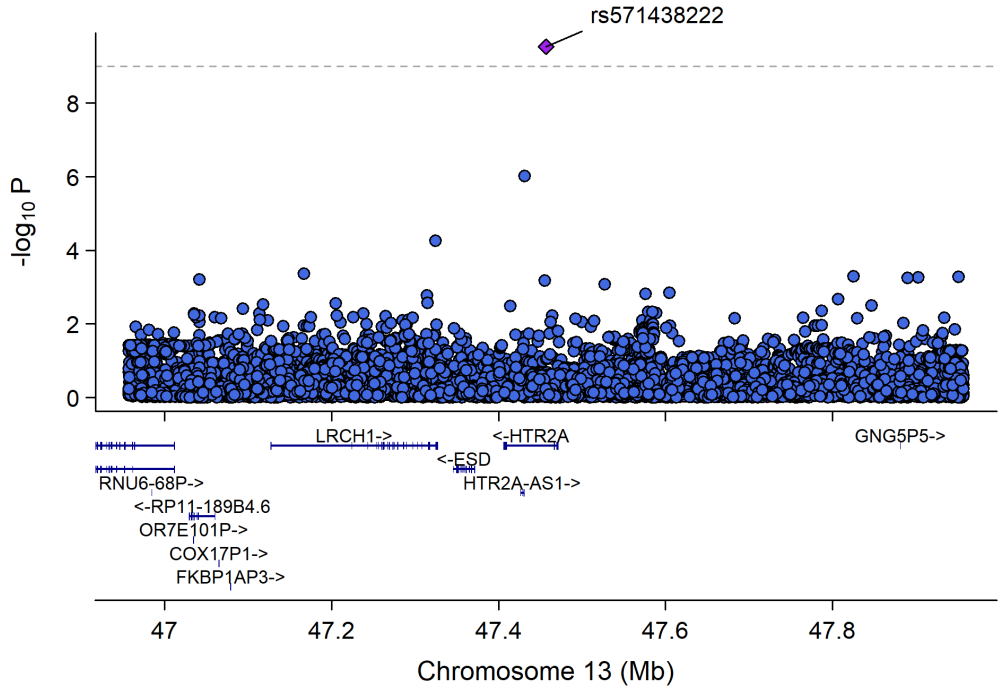
Sphingomyeline C41:1 (µg/dL)



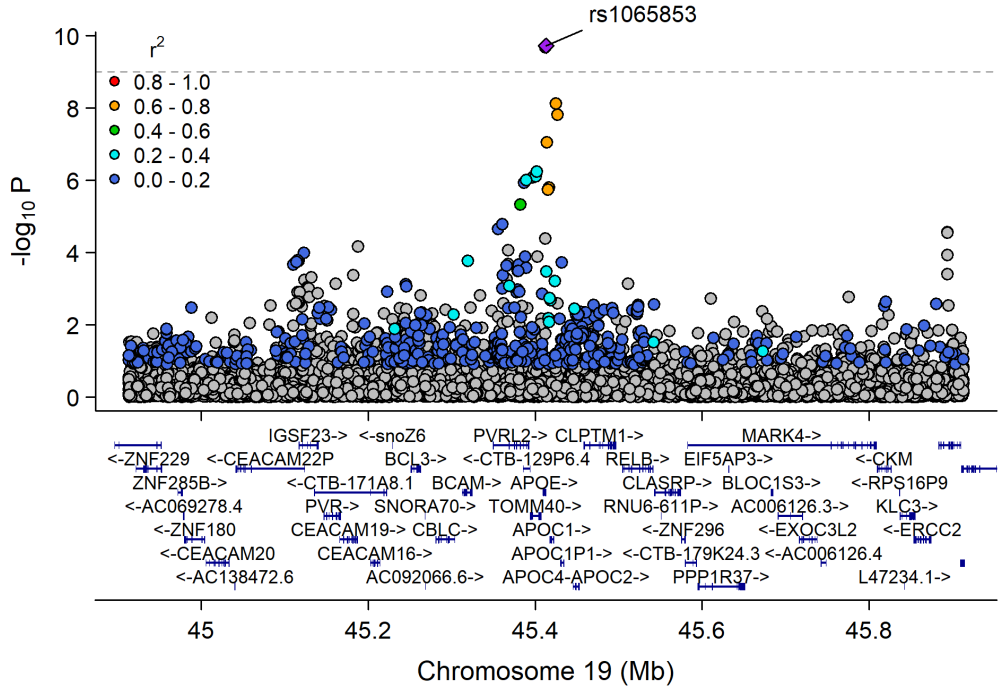




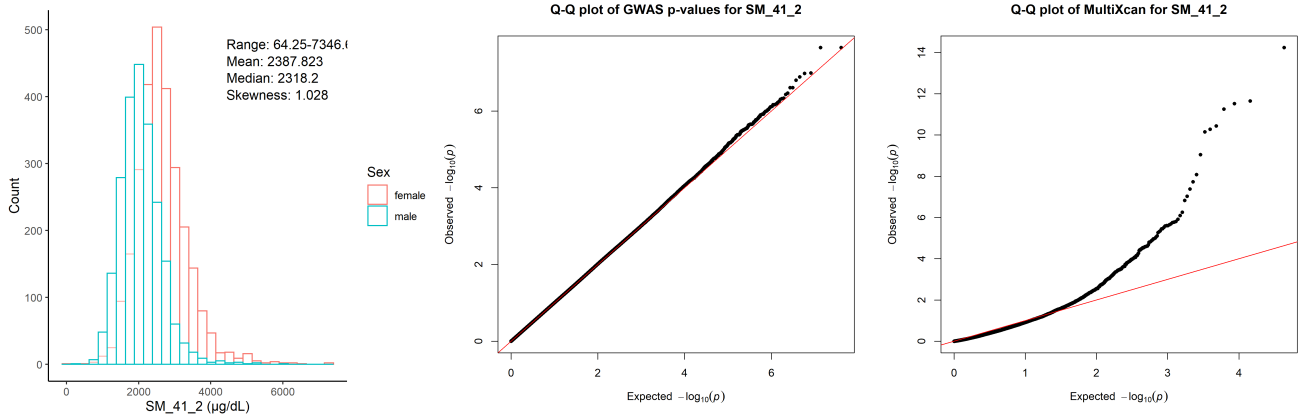
Chr13_46593728_48441802

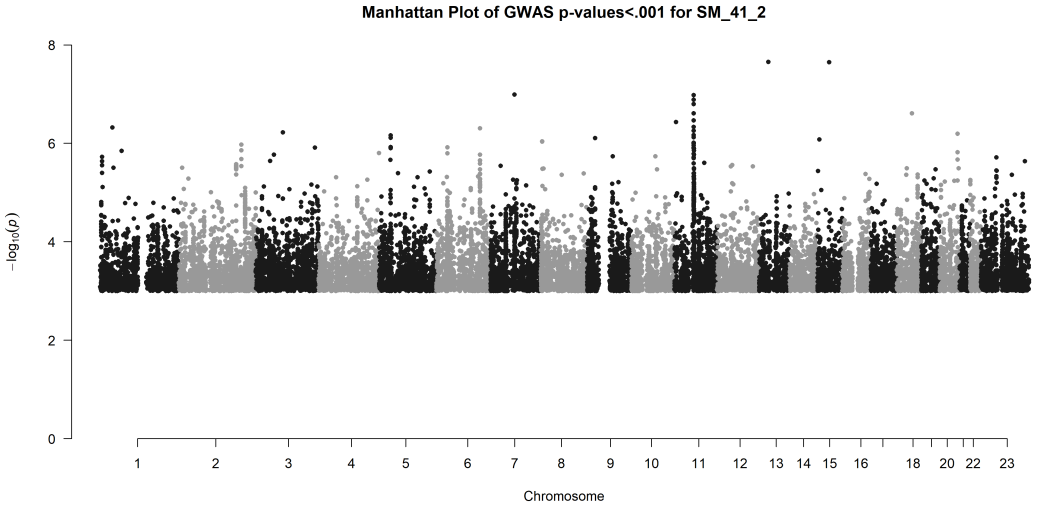


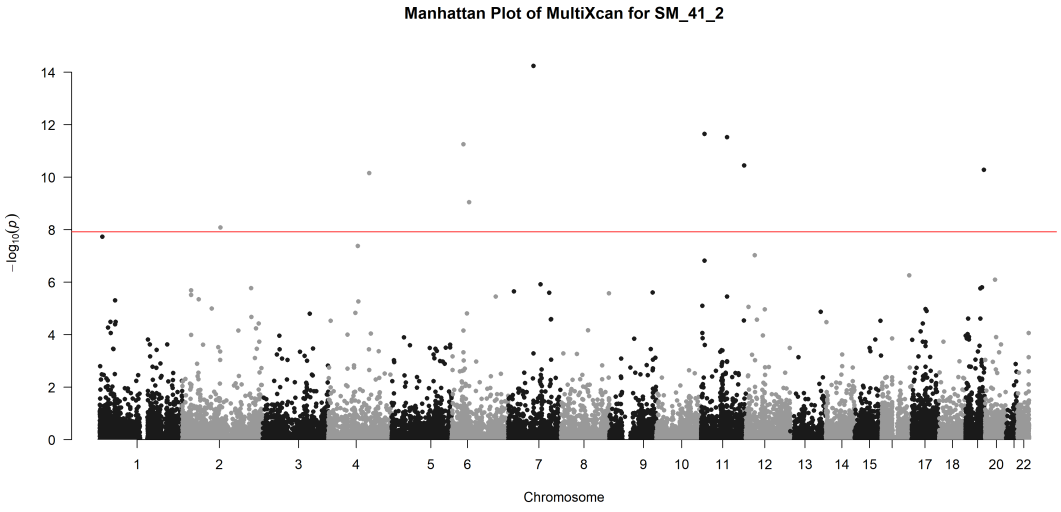
Chr19_44063363_46637375



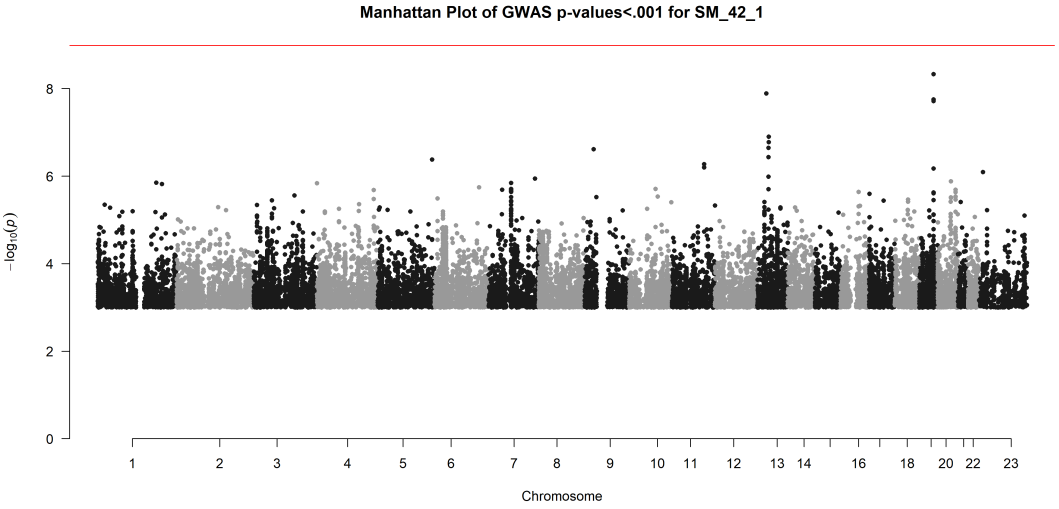
Sphingomyeline C41:2 (µg/dL)

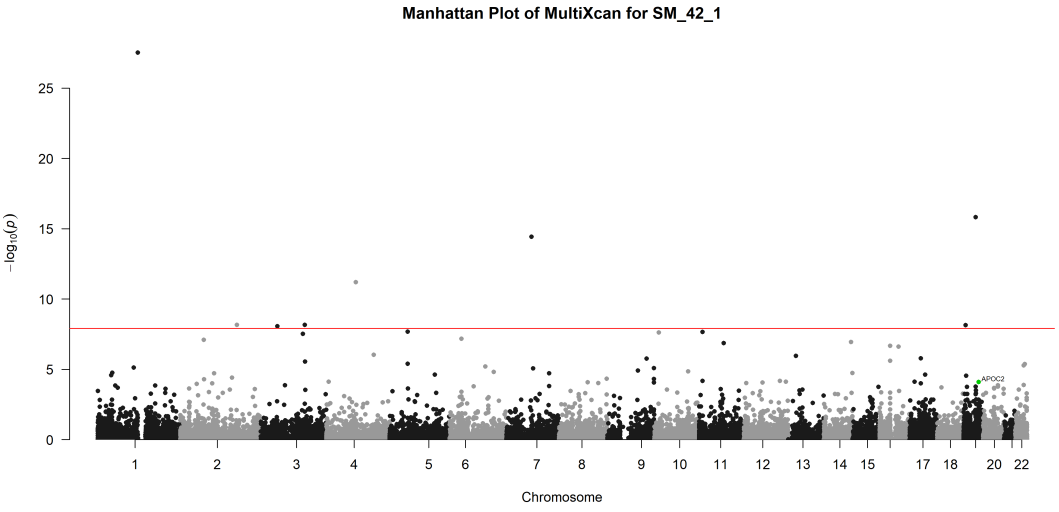




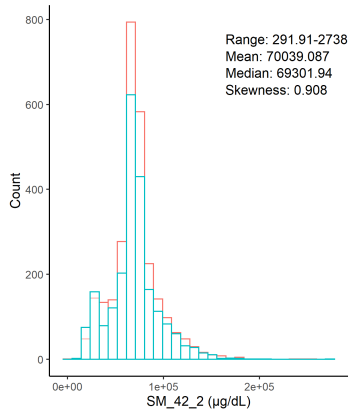


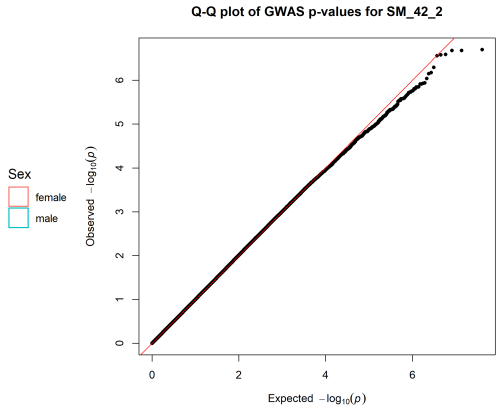
Sphingomyeline C42:1 (µg/dL)

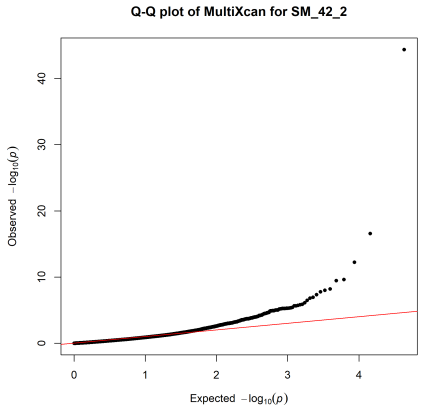


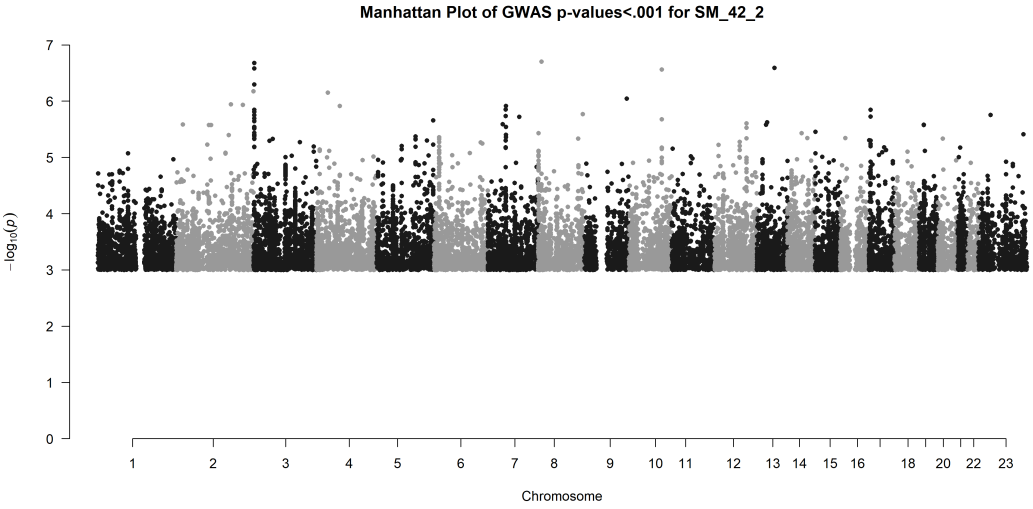


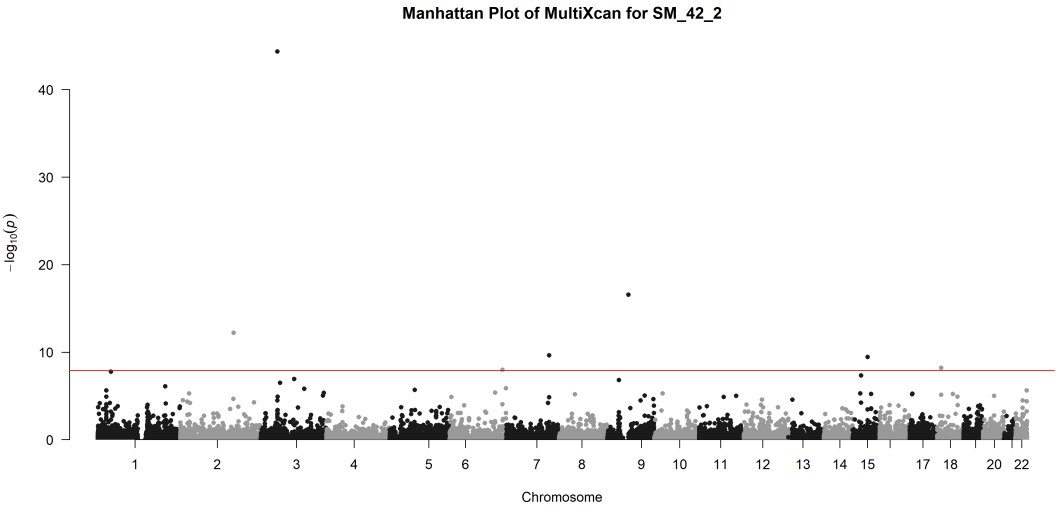
Sphingomyeline C42:2 (µg/dL)



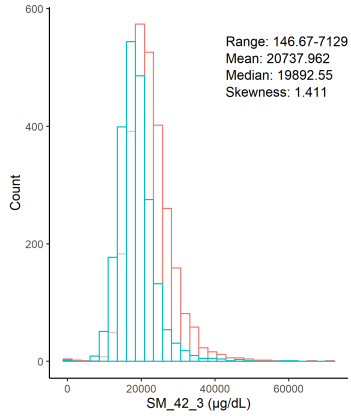


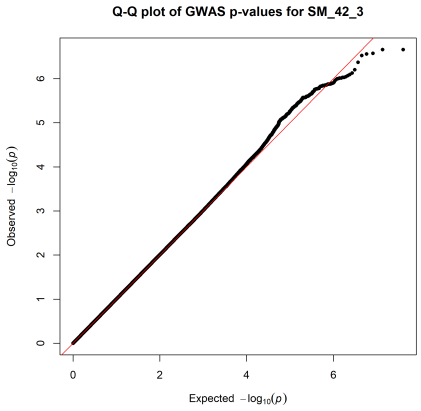


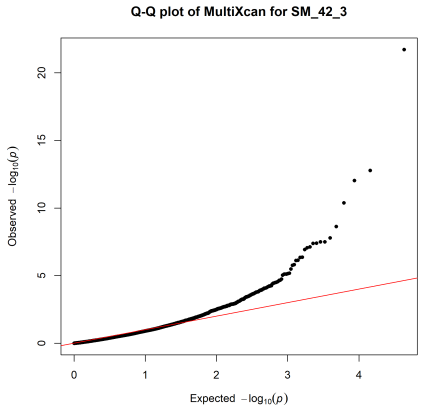


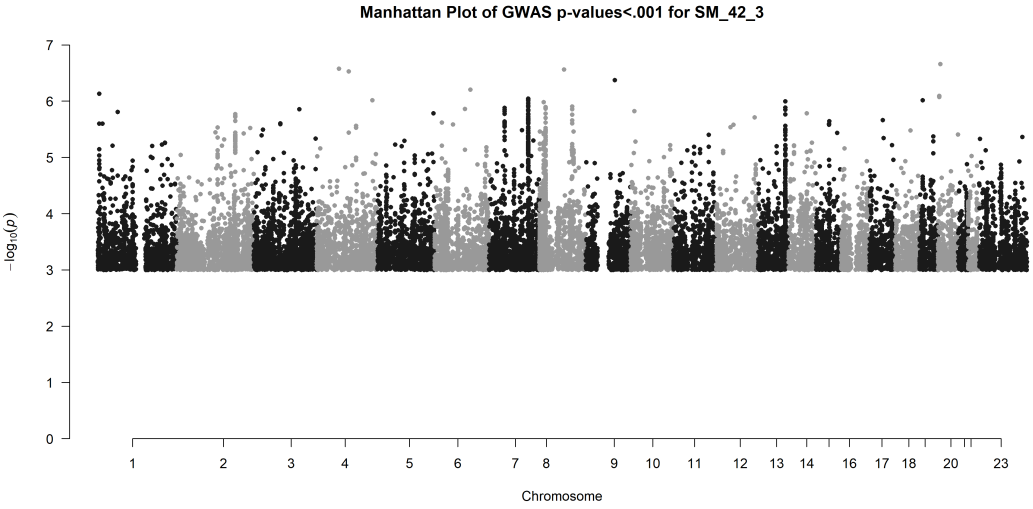


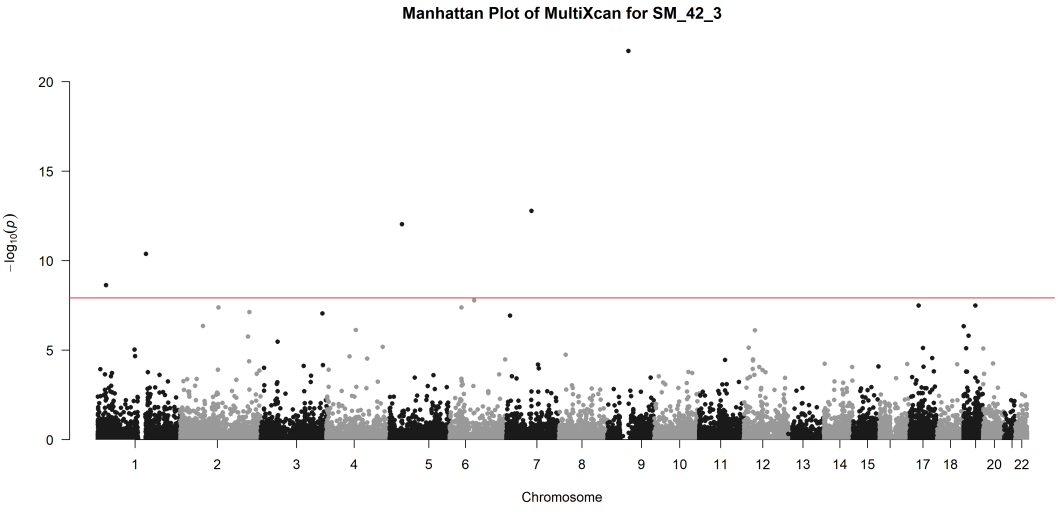
Sphingomyeline C42:3 (µg/dL)



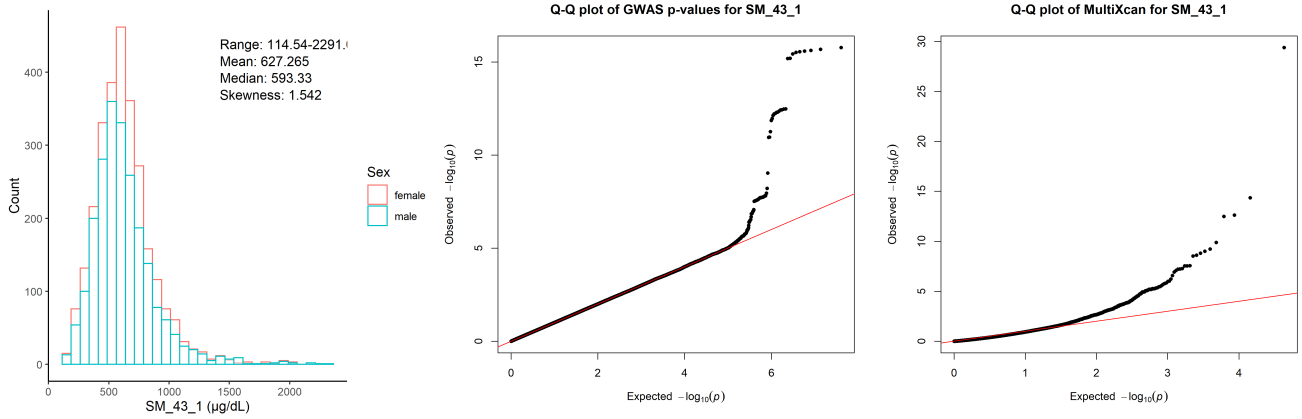


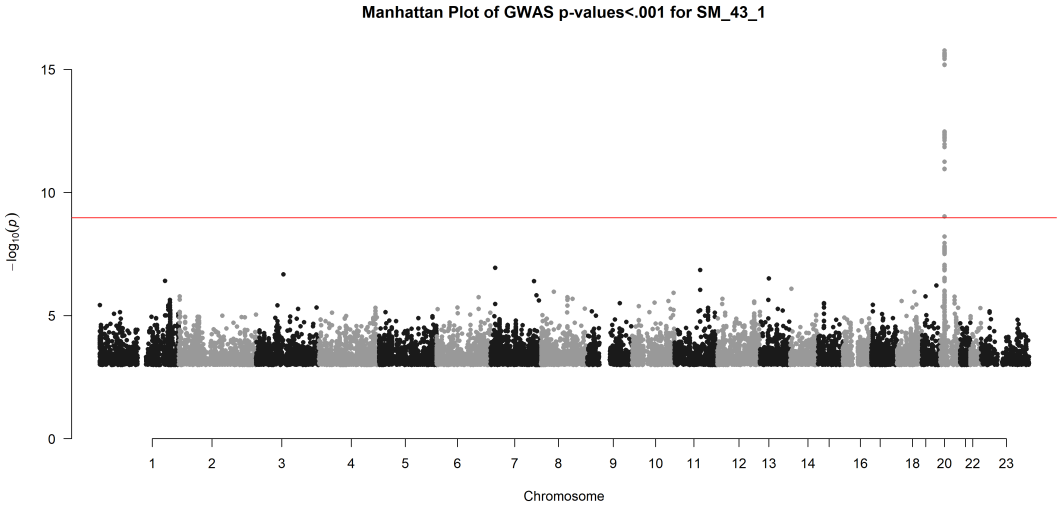


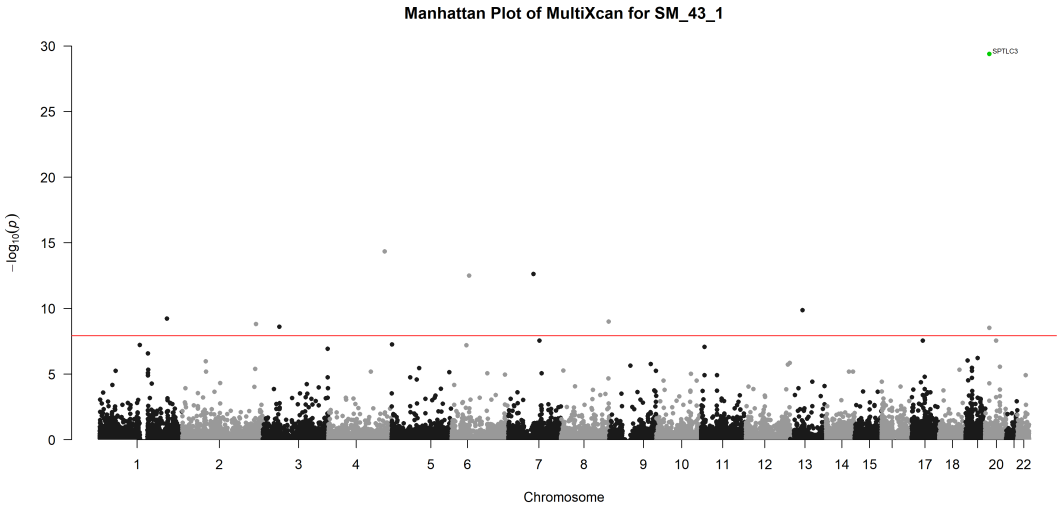




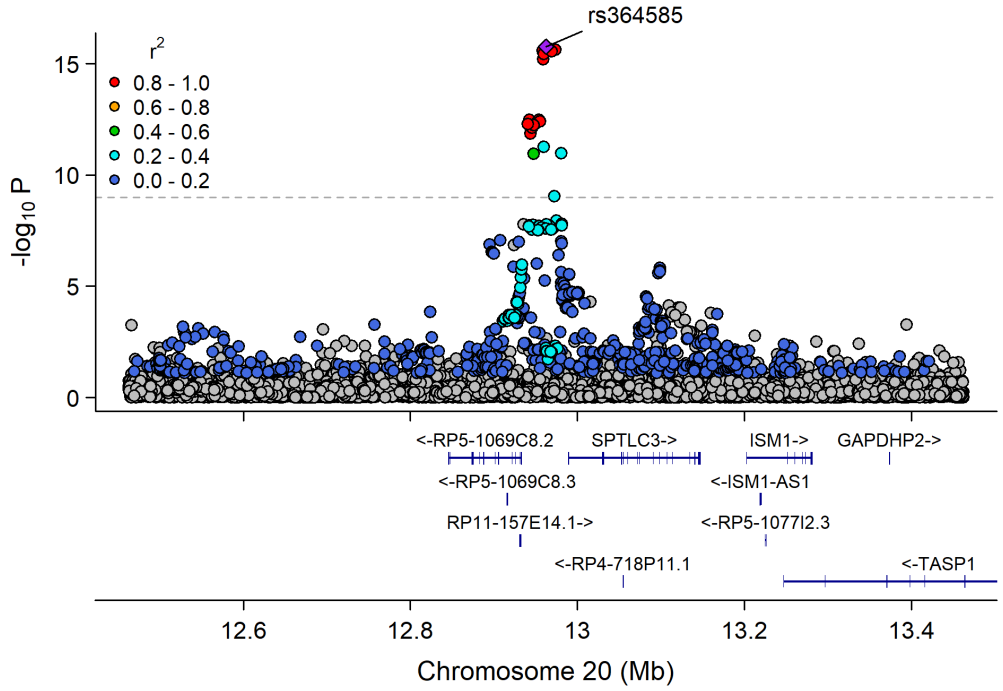
Sphingomyeline C43:1 (µg/dL)



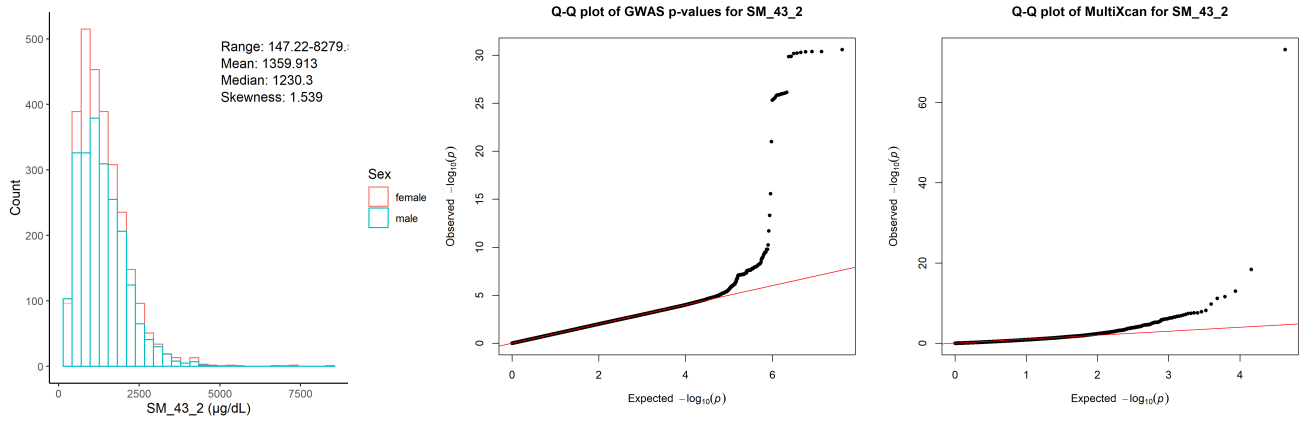


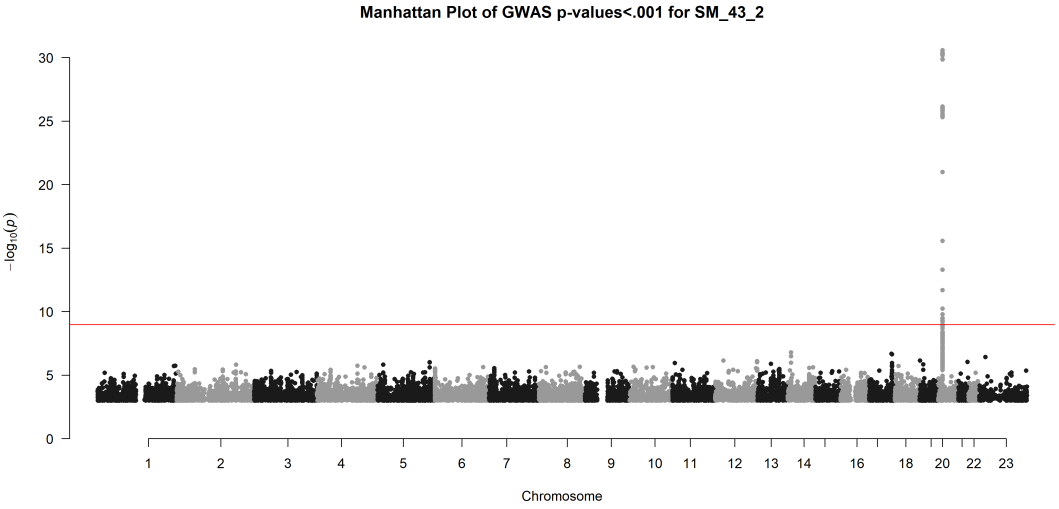


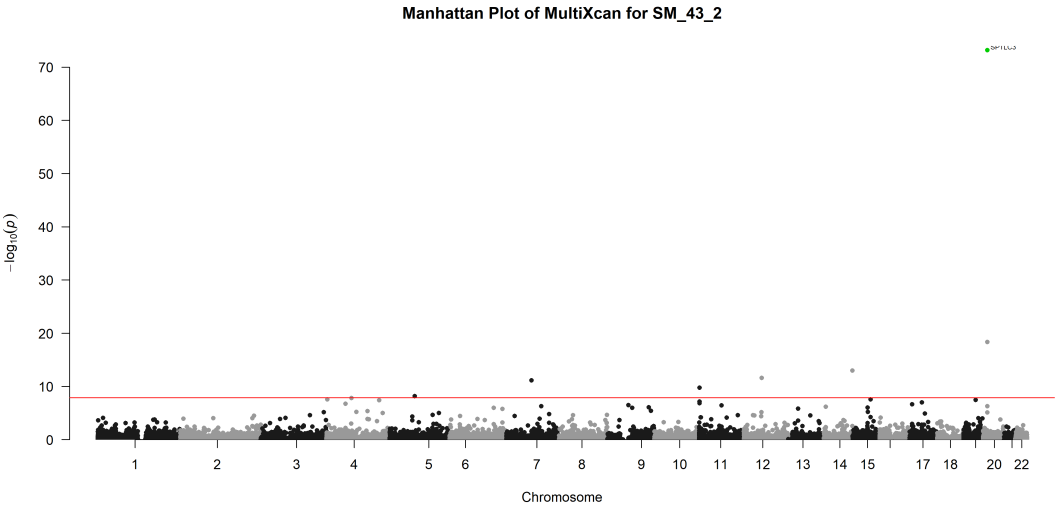
Chr20_12814090_13047914



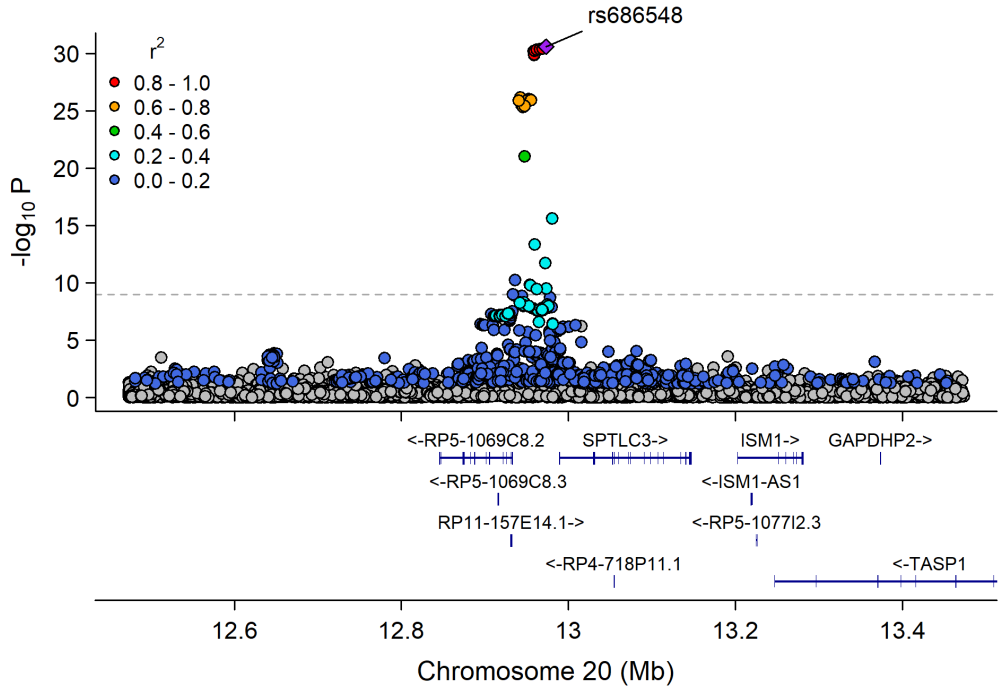
Sphingomyeline C43:2 (µg/dL)





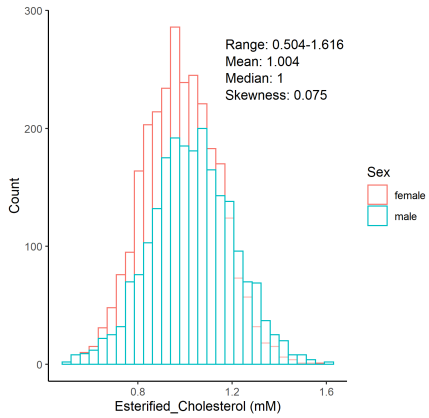


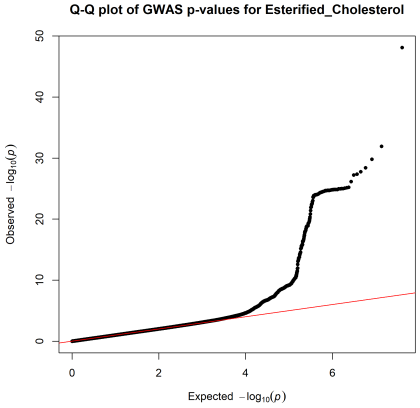
Chr20_12814090_13047914

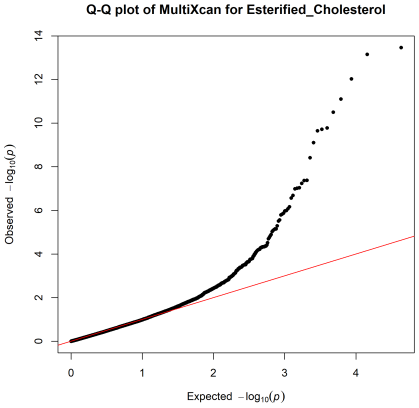


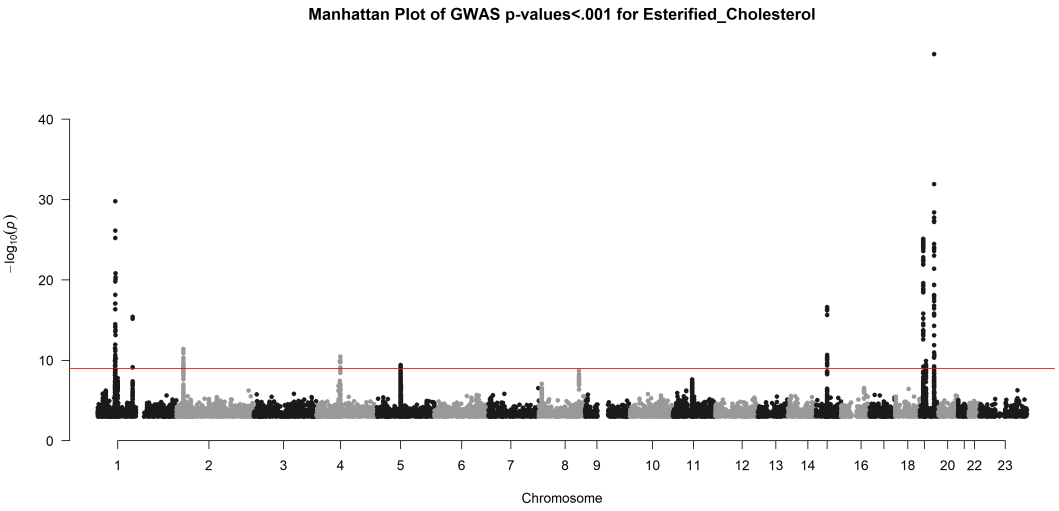
Steroids

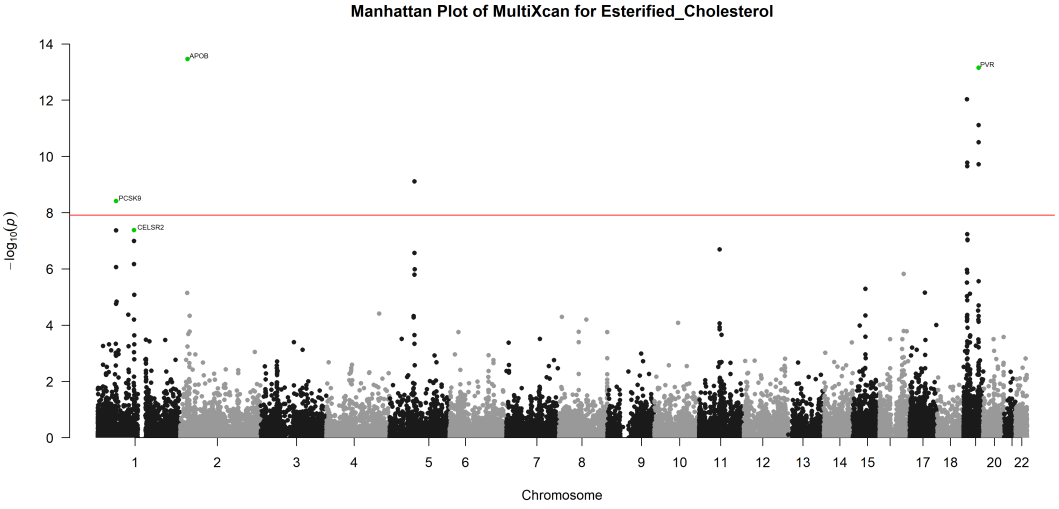
Esterified cholesterol (mM)



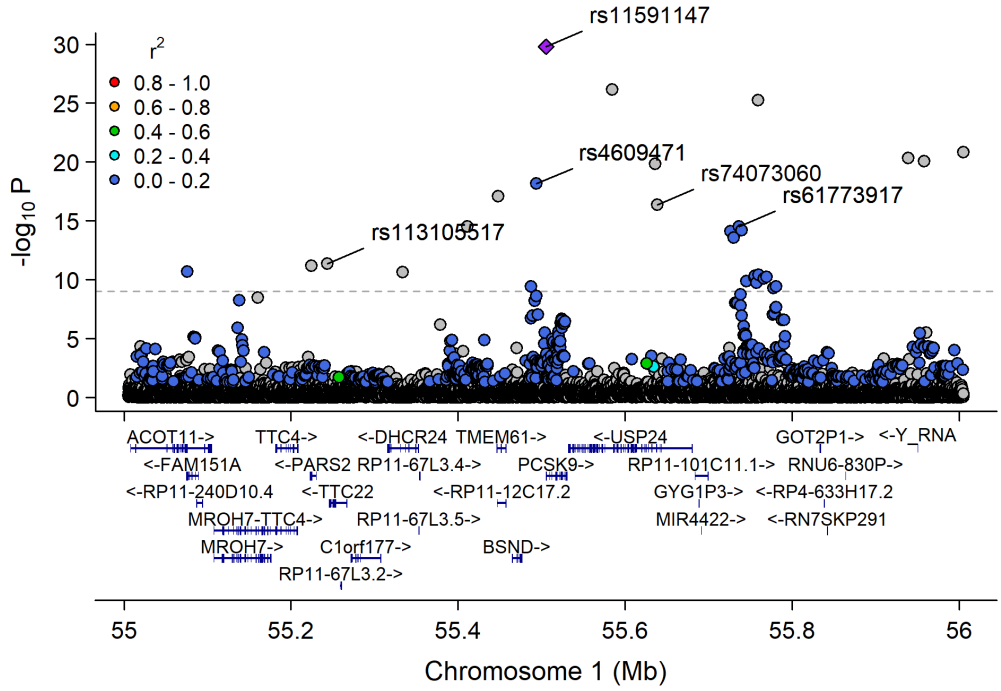




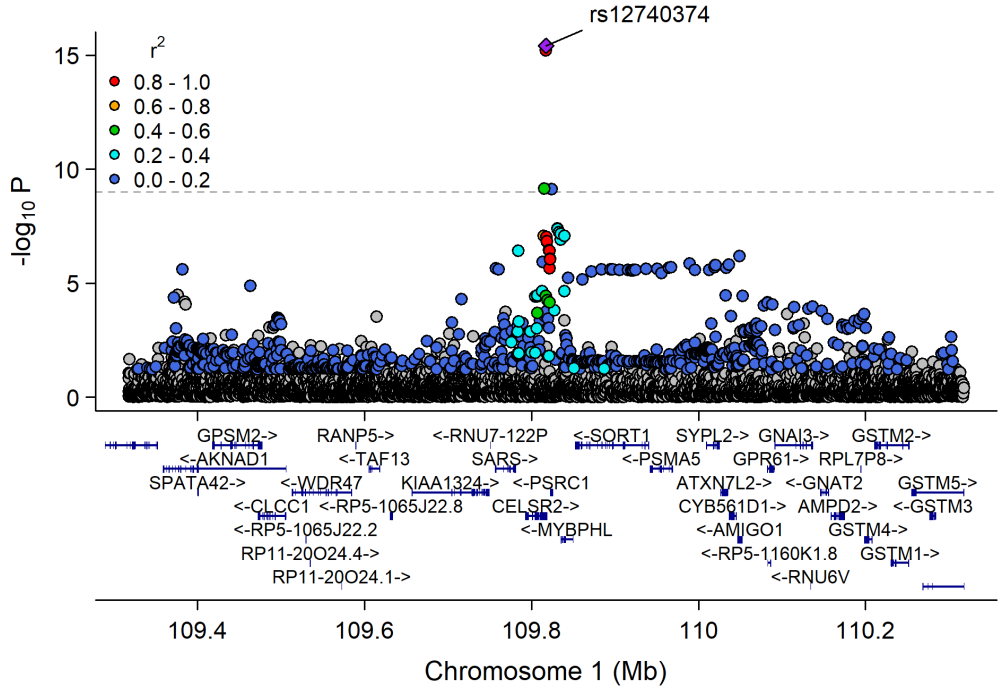




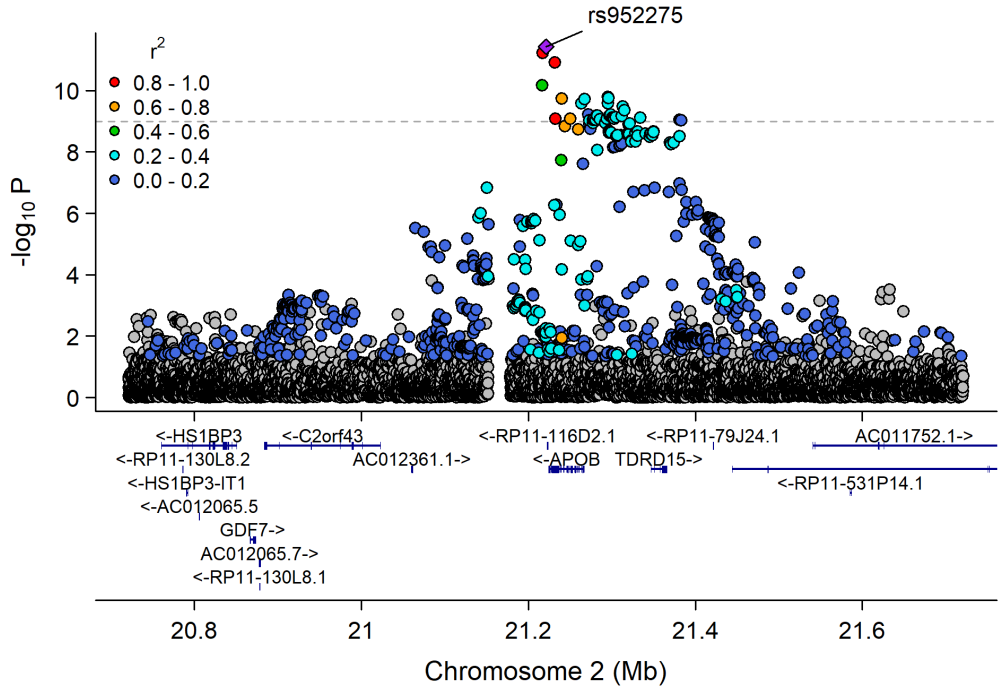
Chr1__53272879__57300396



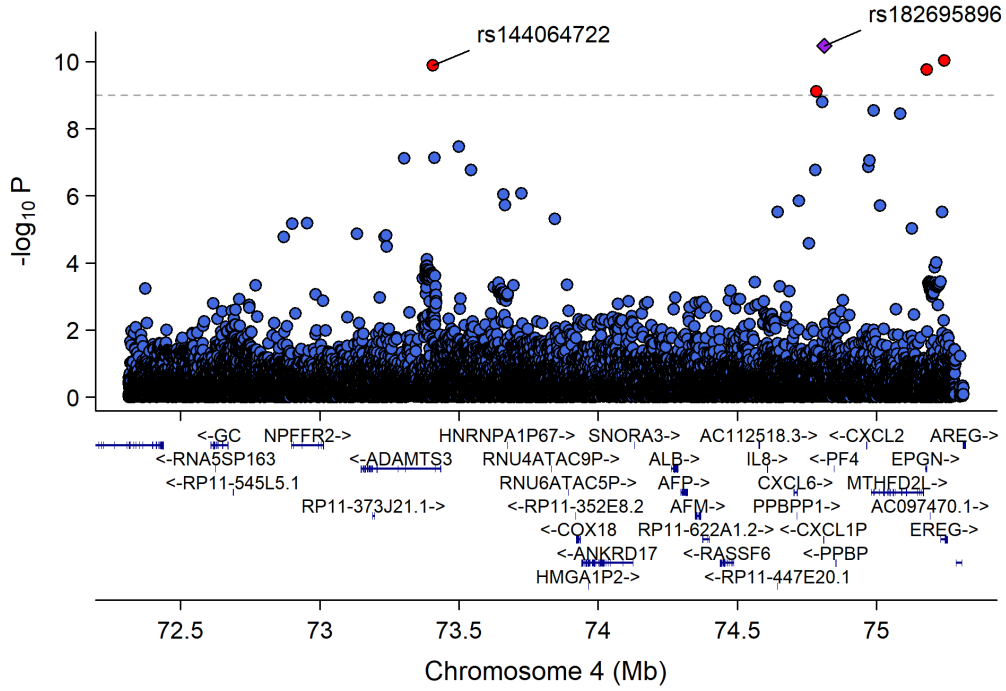
Chr1__109696238__110162190



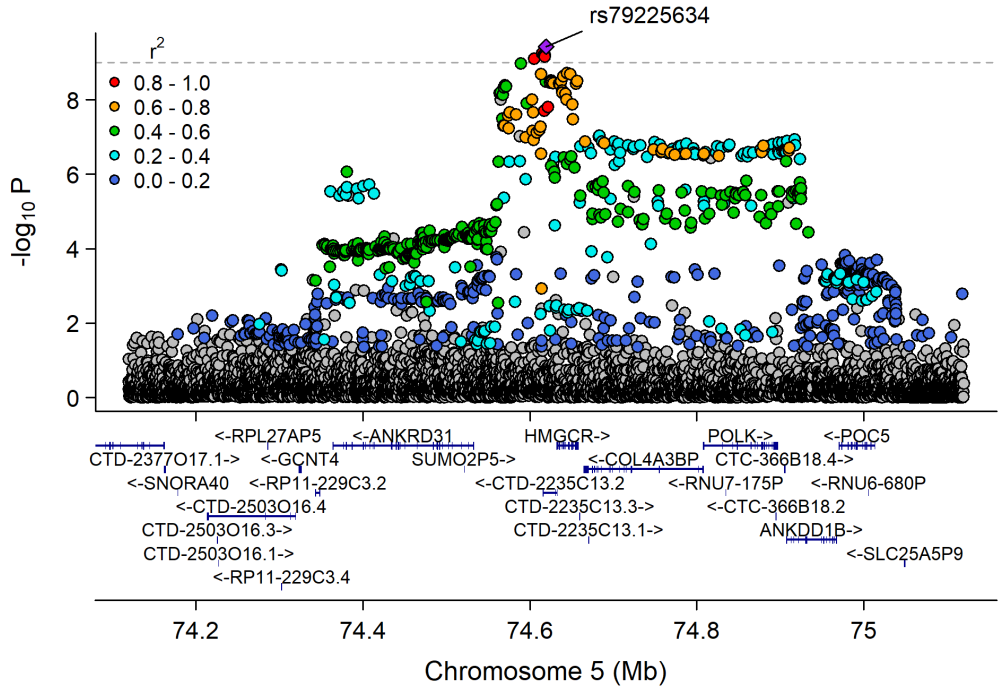
Chr2_19947287_22427492



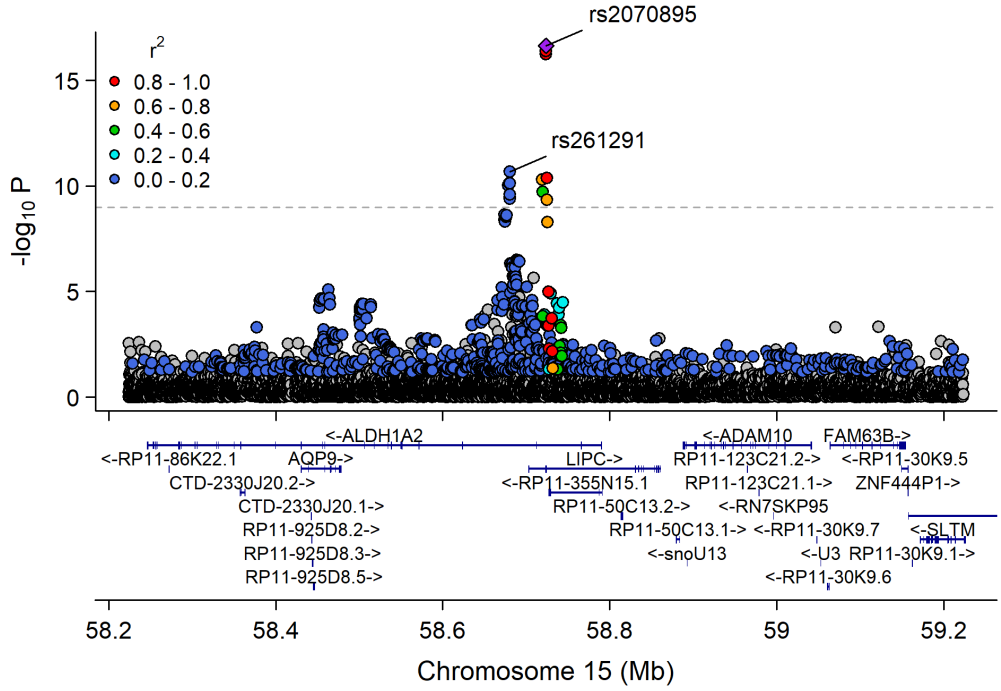
Chr4_71881824_76681229



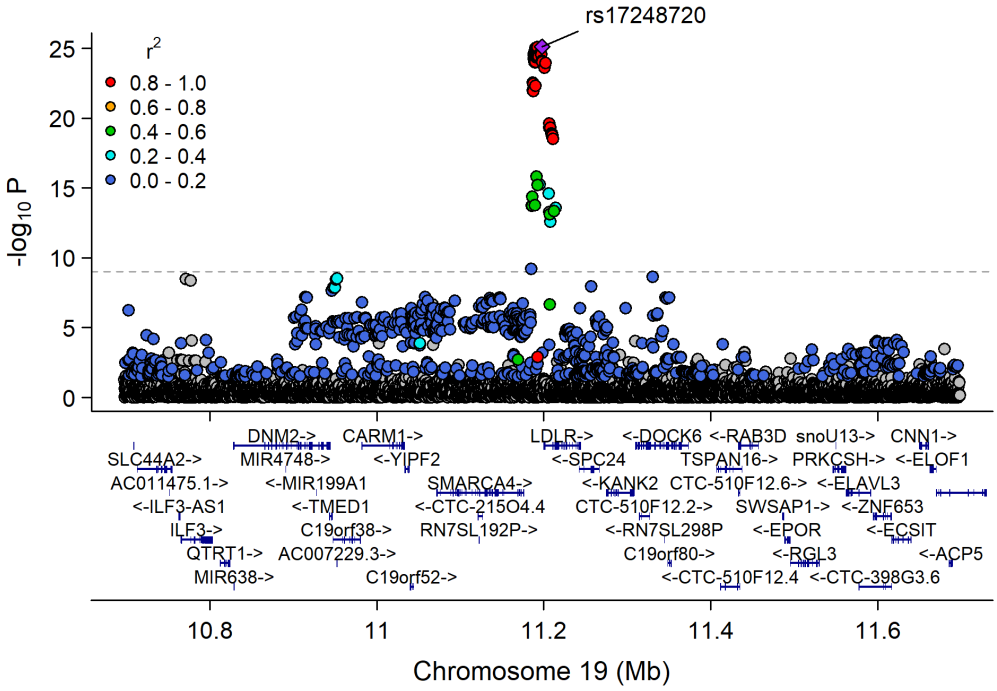
Chr5_74242002_75221582



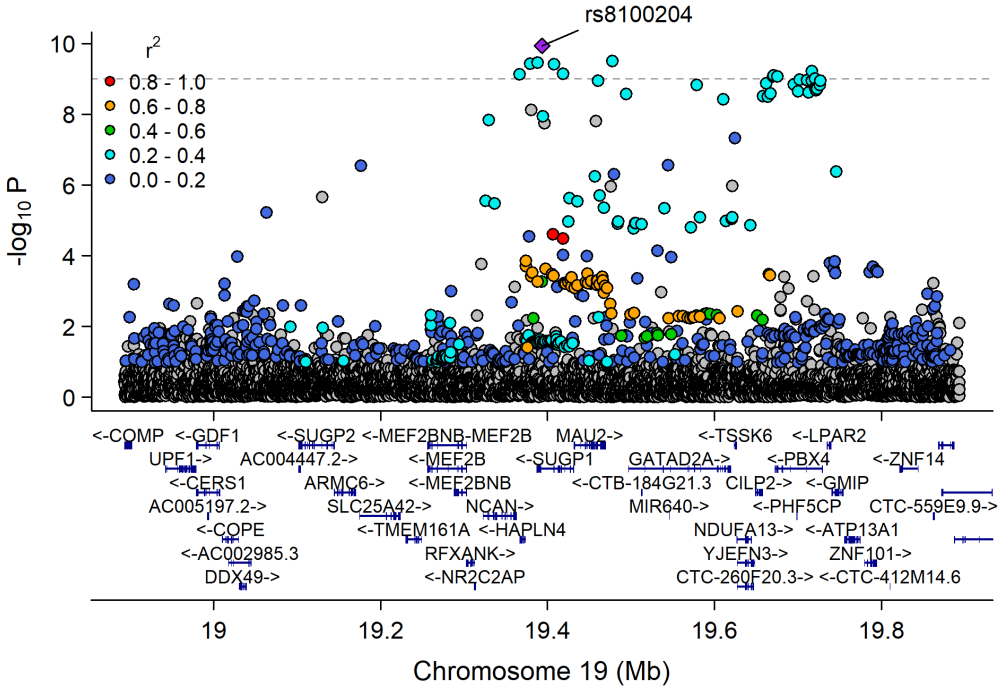
Chr15_57658798_60490883



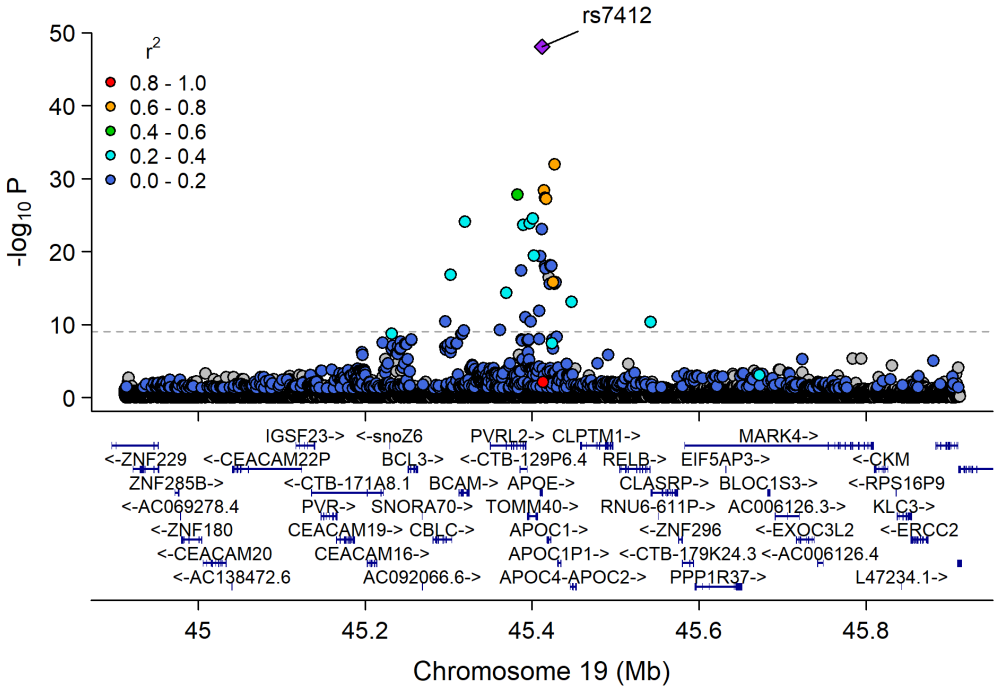
Chr19_9807999_12255594



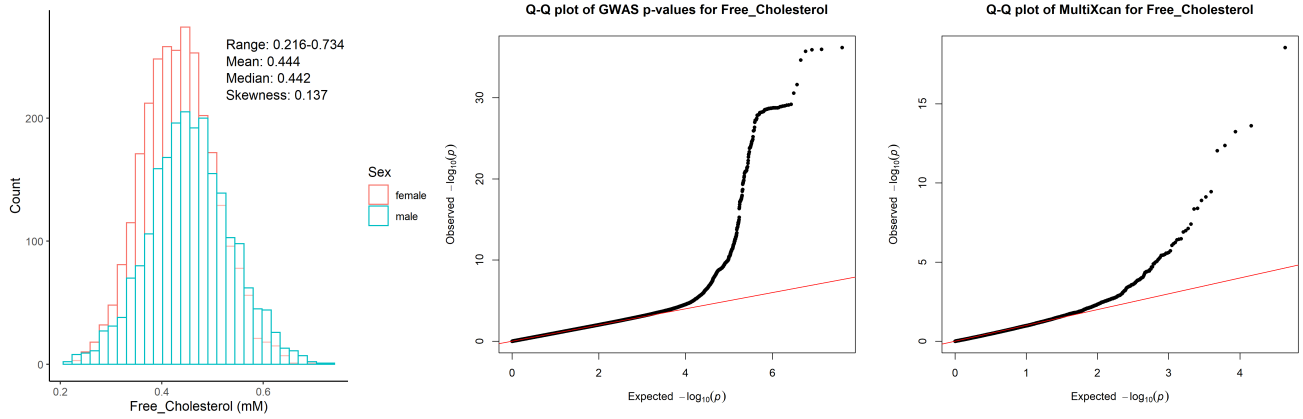
Chr19_18370495_20841464

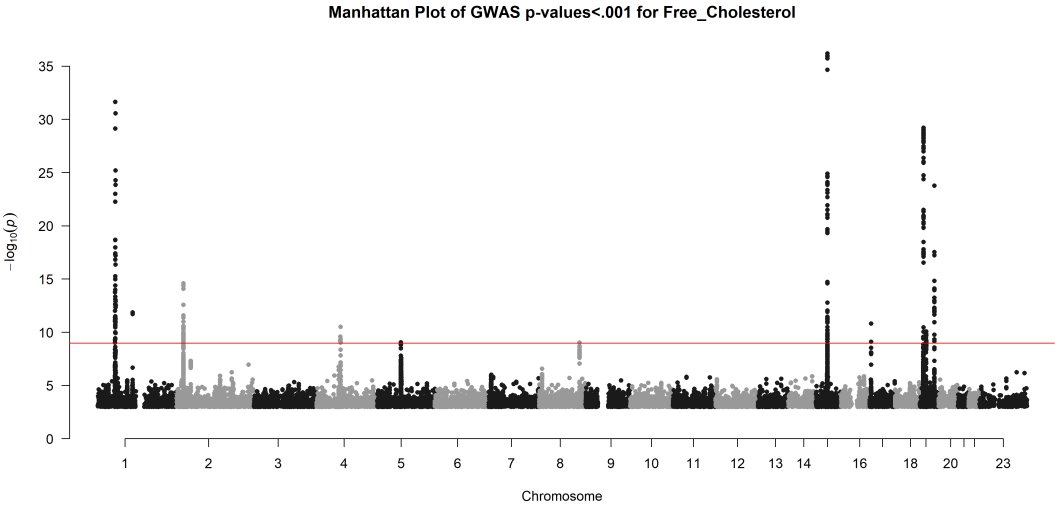


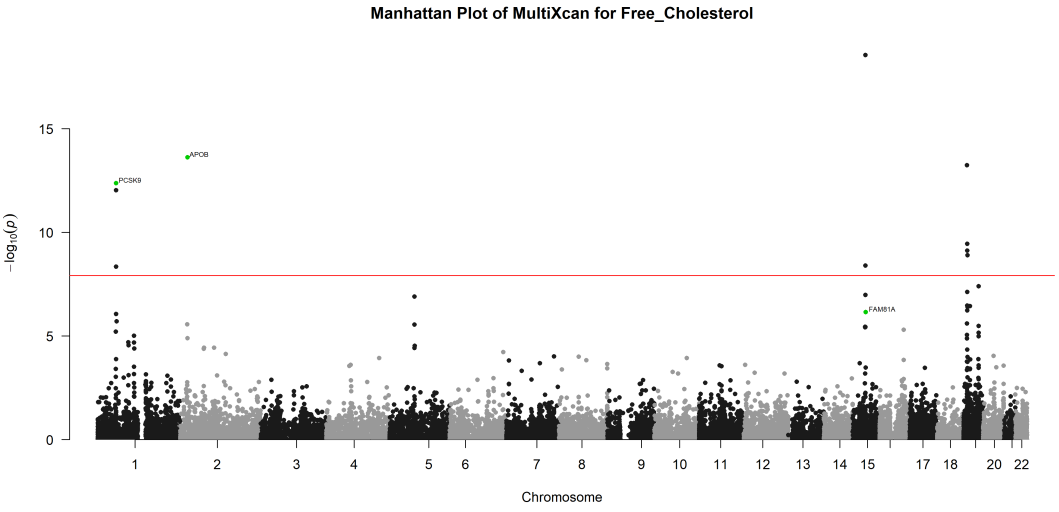
Chr19__44063363__46637375



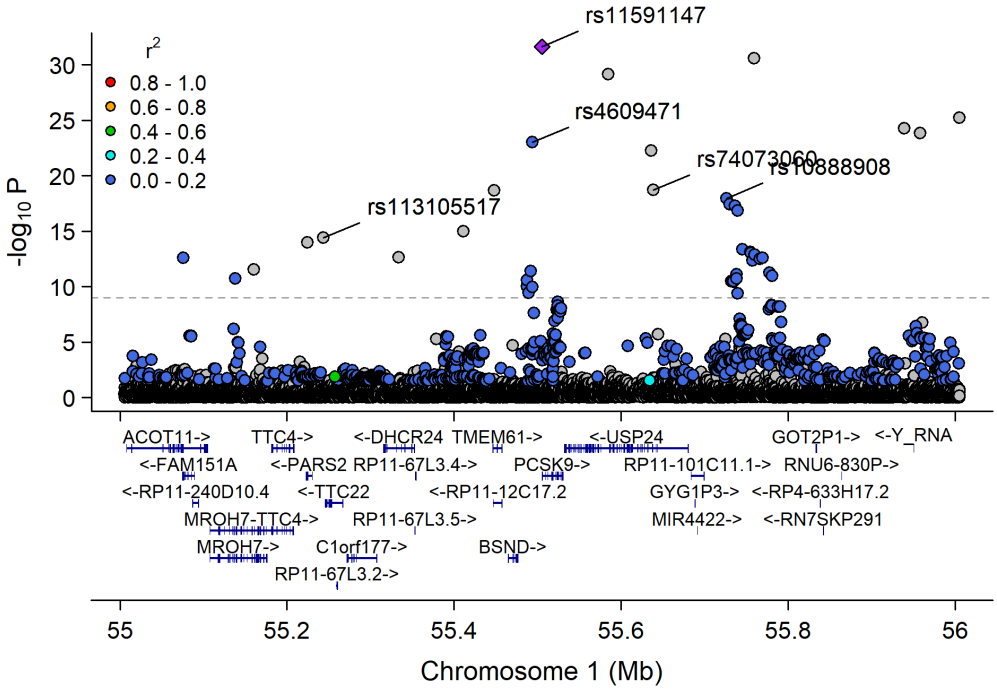
Free cholesterol (mM)



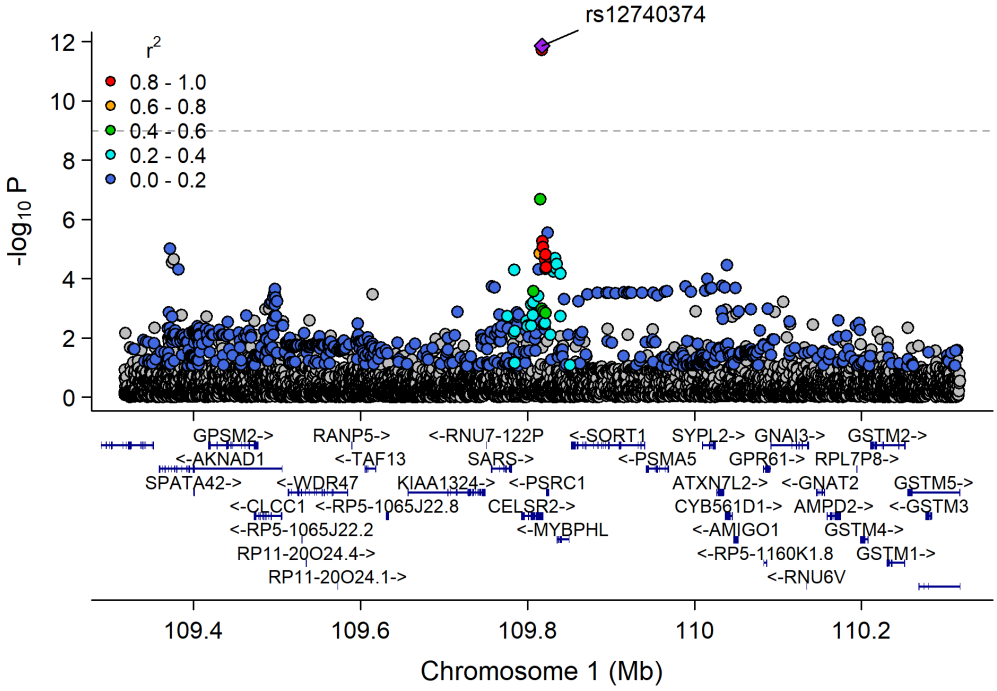




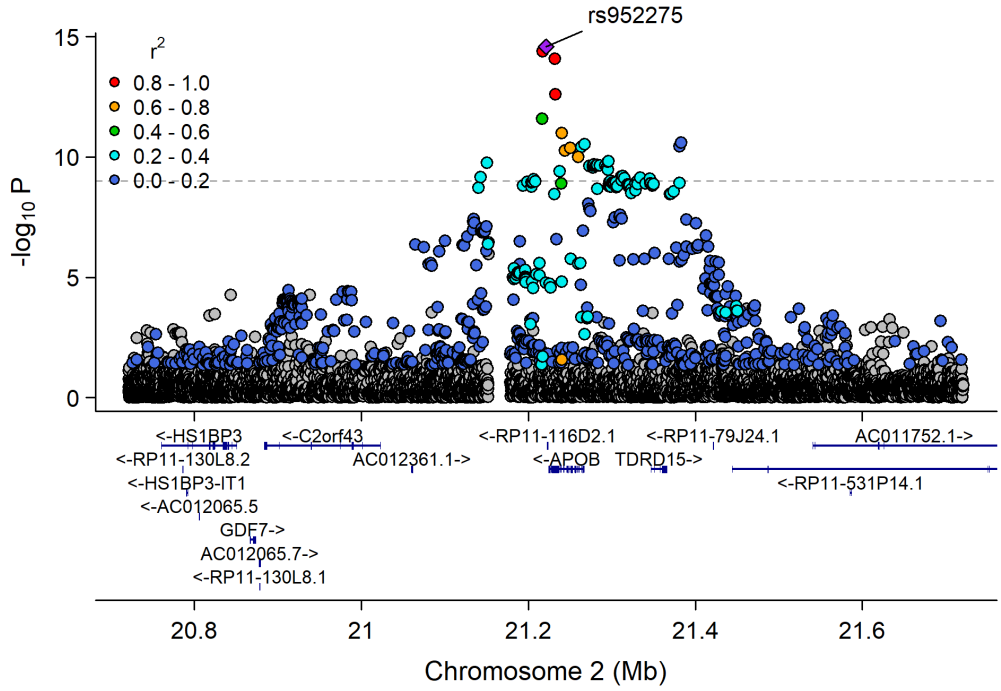
Chr1__53272879__57300396



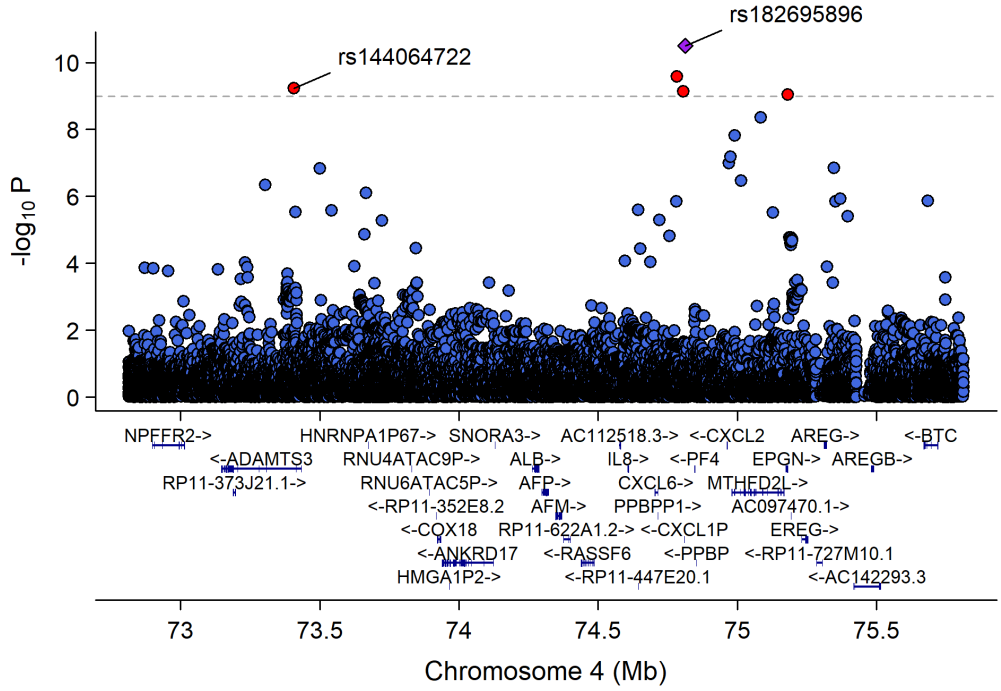
Chr1__109696238__110162190



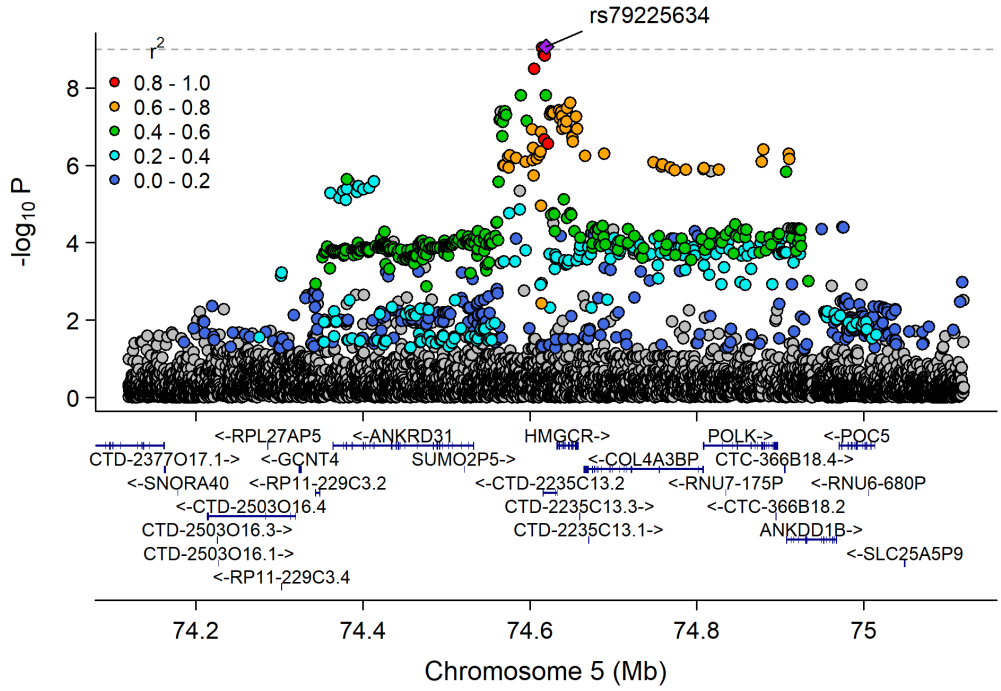
Chr2_19947287_22427492



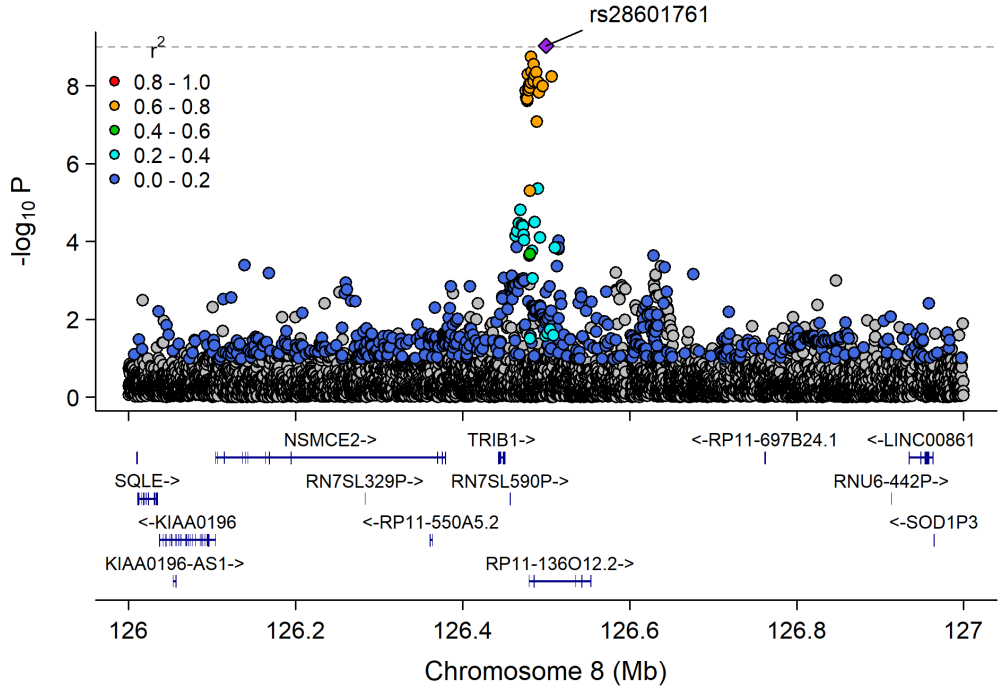
Chr4_71881824_76681229



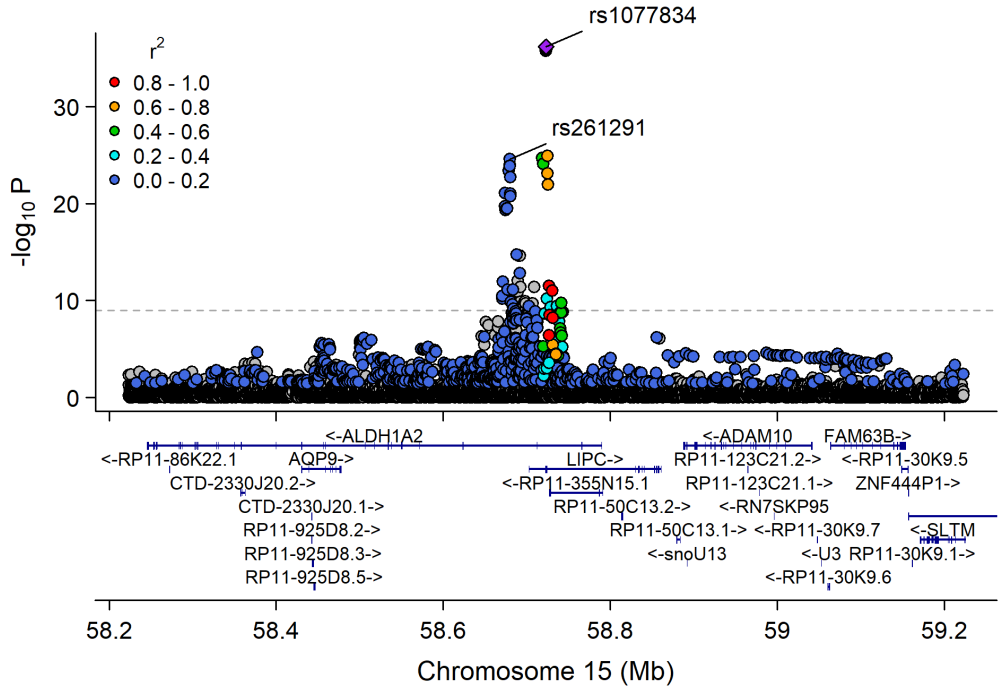
Chr5_74242002_75221582



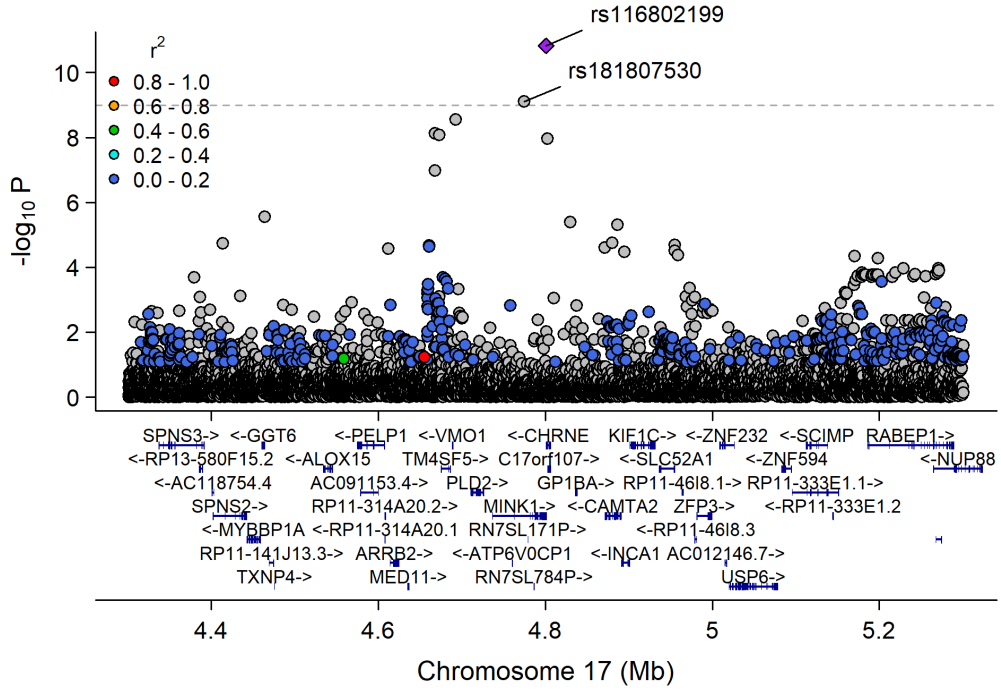
Chr8_126435663_126578061



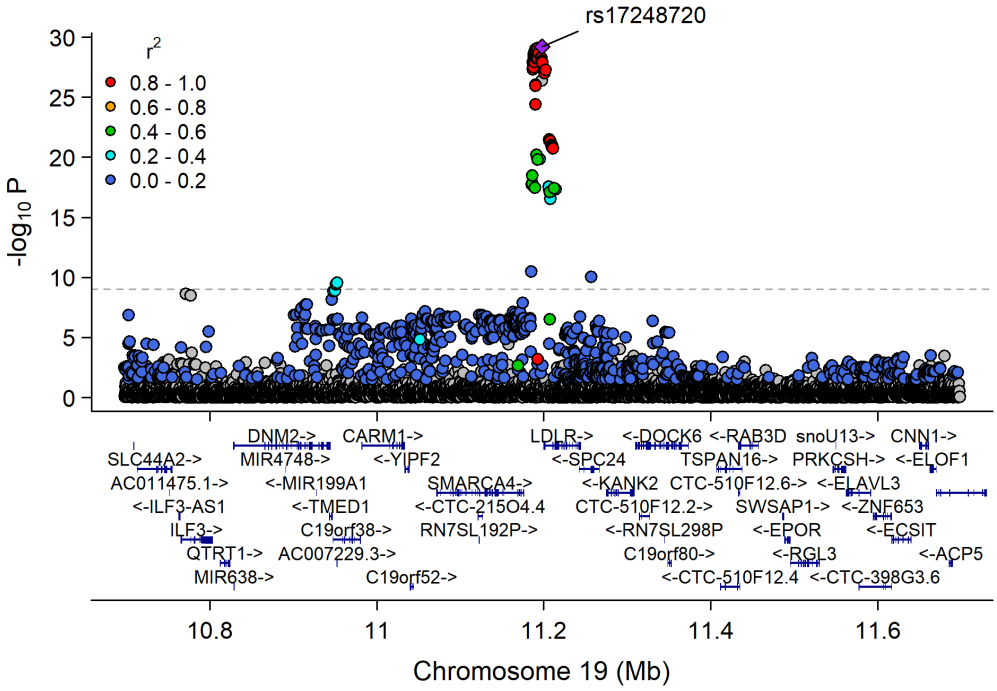
Chr15_57658798_60490883



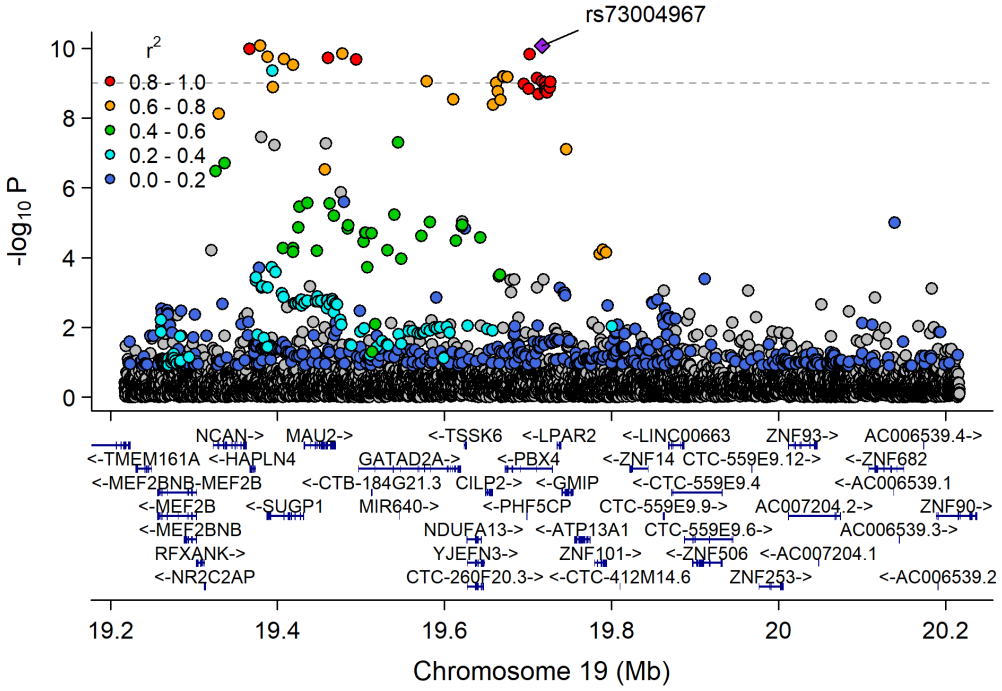
Chr17_3704190_5793348



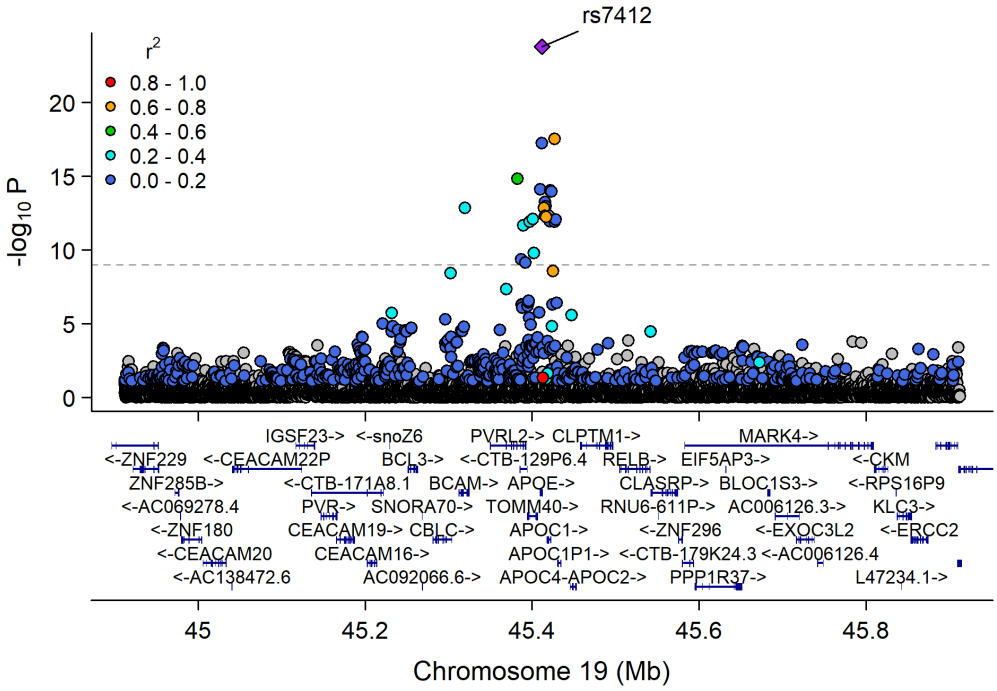
Chr19_9807999_12255594



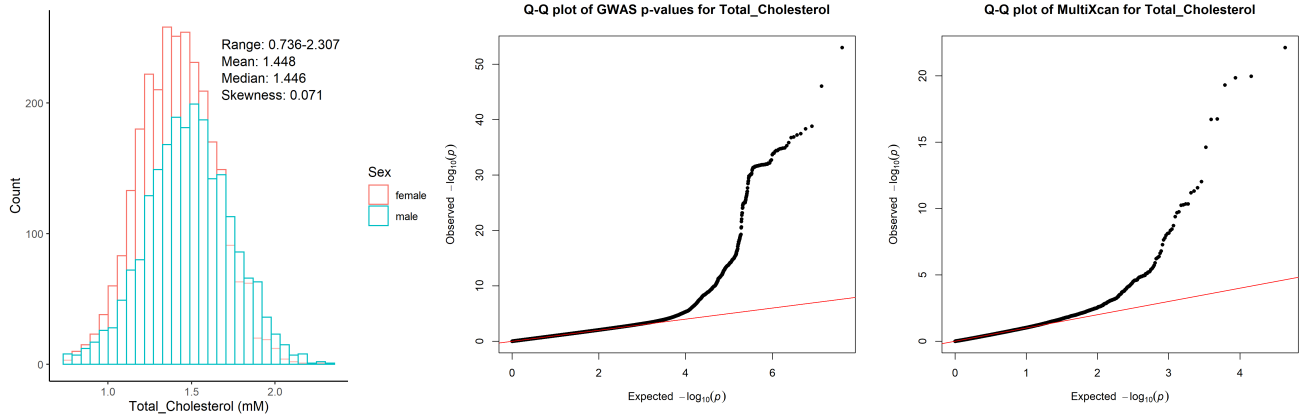
Chr19_18370495_20841464

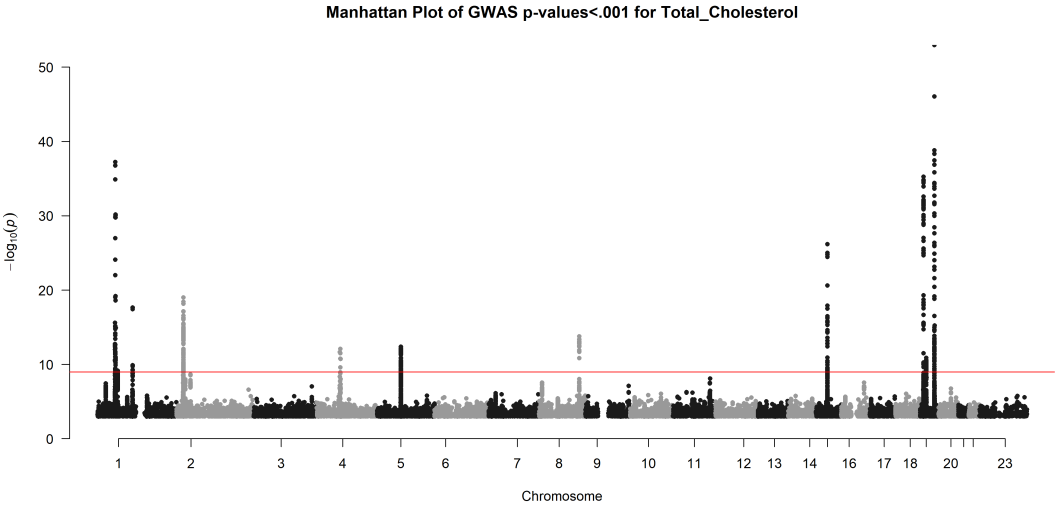


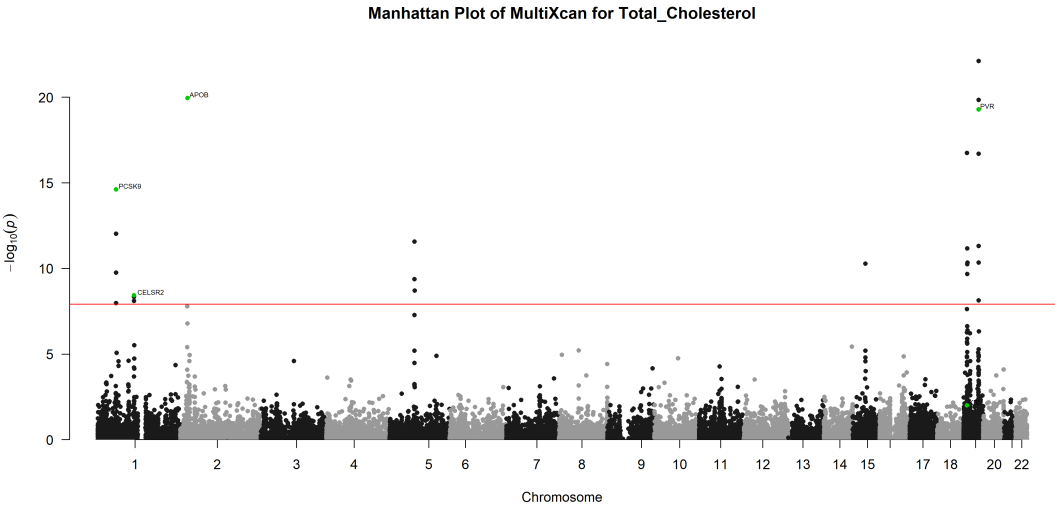
Chr19__44063363__46637375



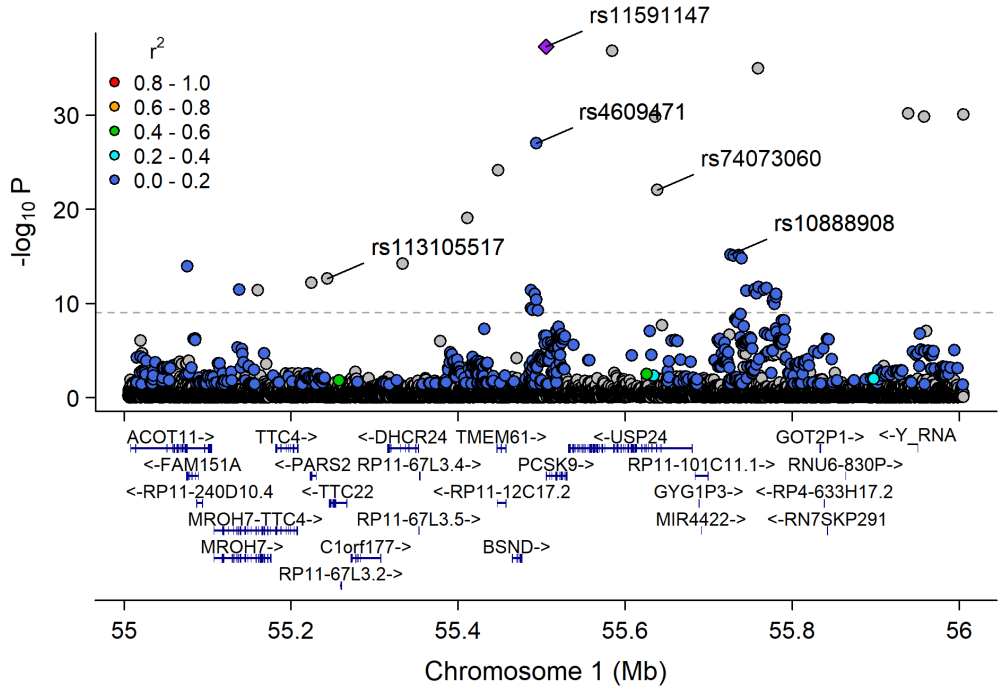
Total cholesterol (mM)



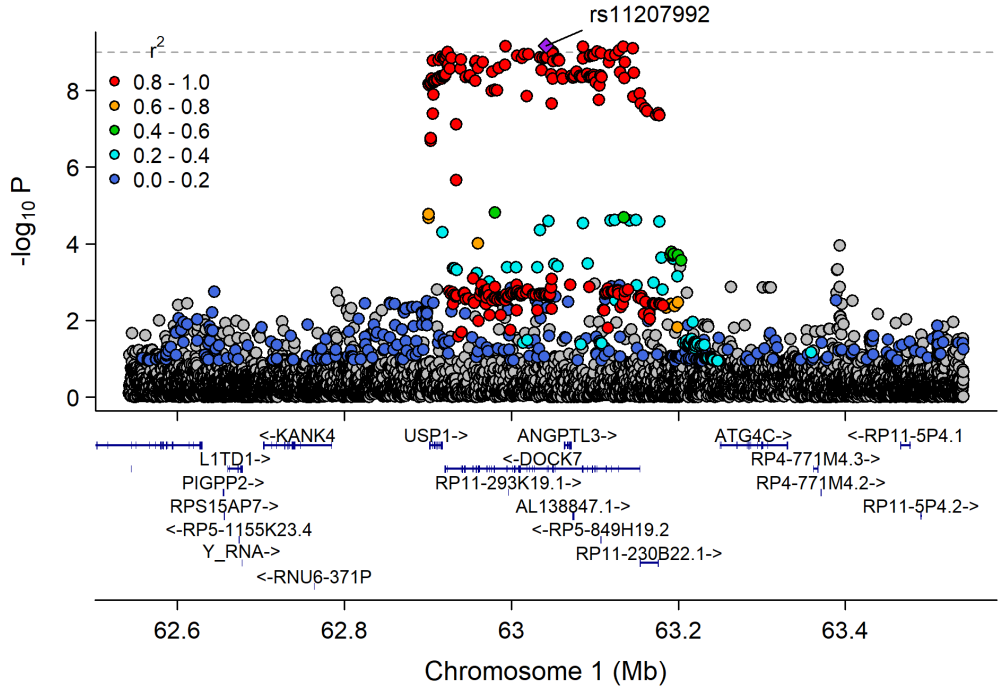




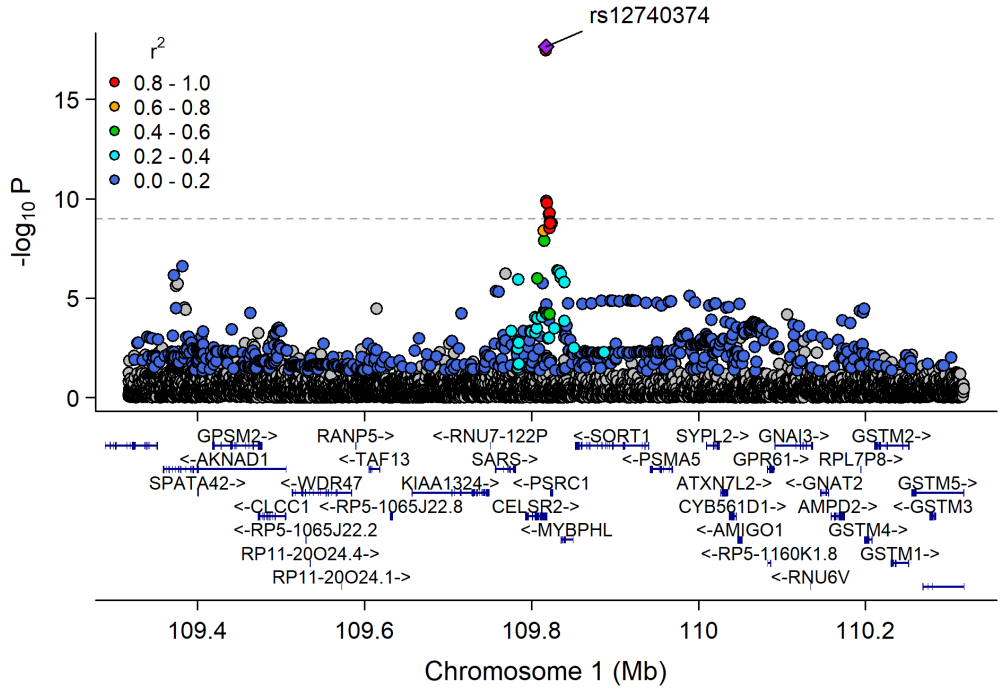
Chr1__53272879__57300396



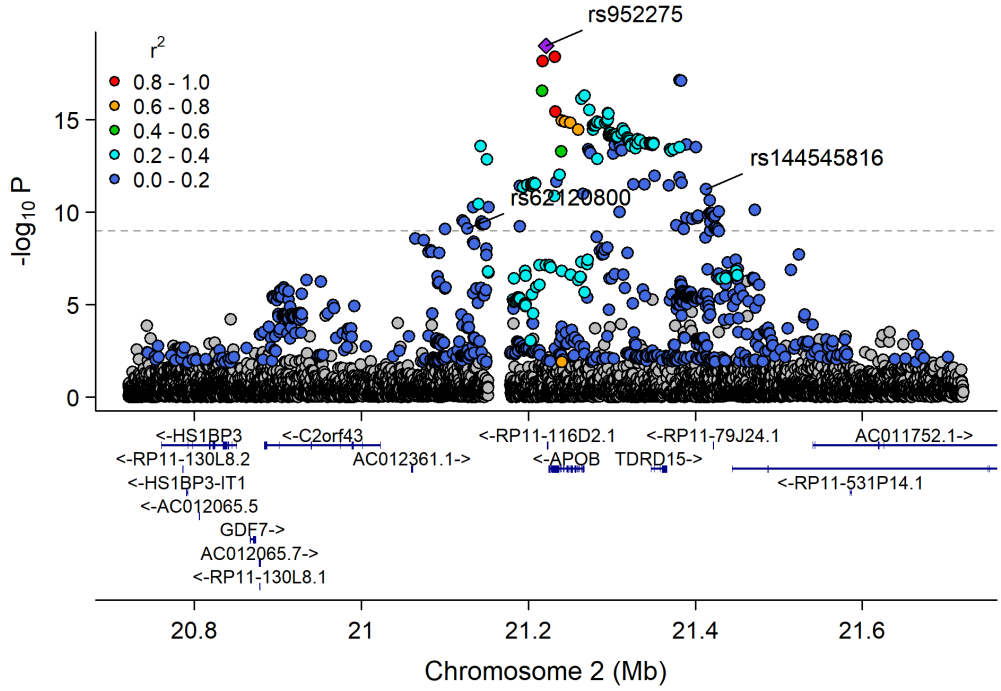
Chr1__62833402__63375116



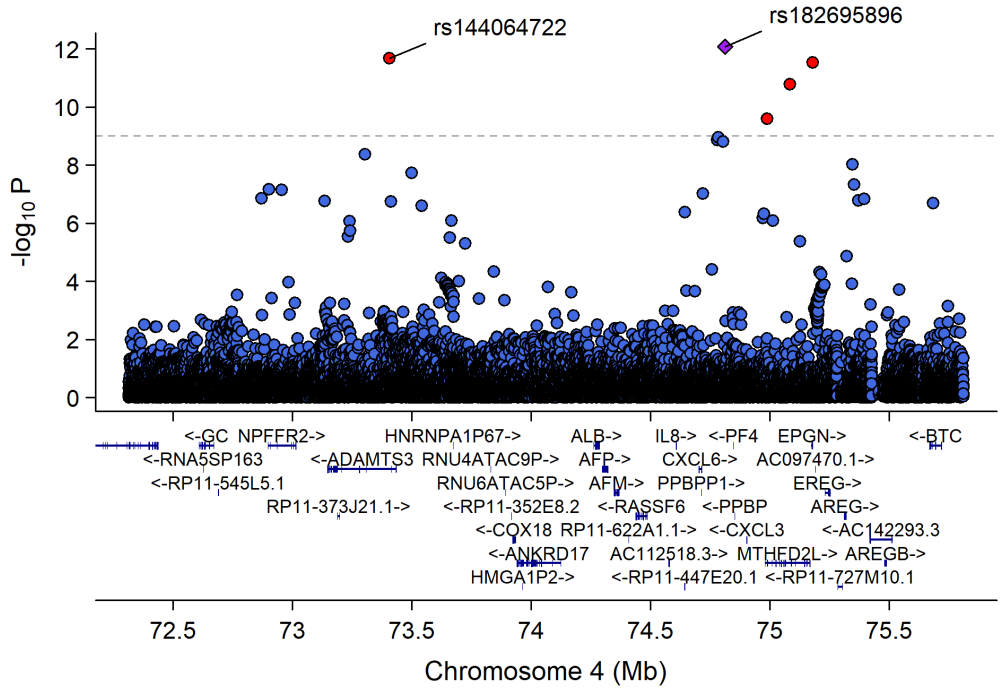
Chr1__109696238__110162190



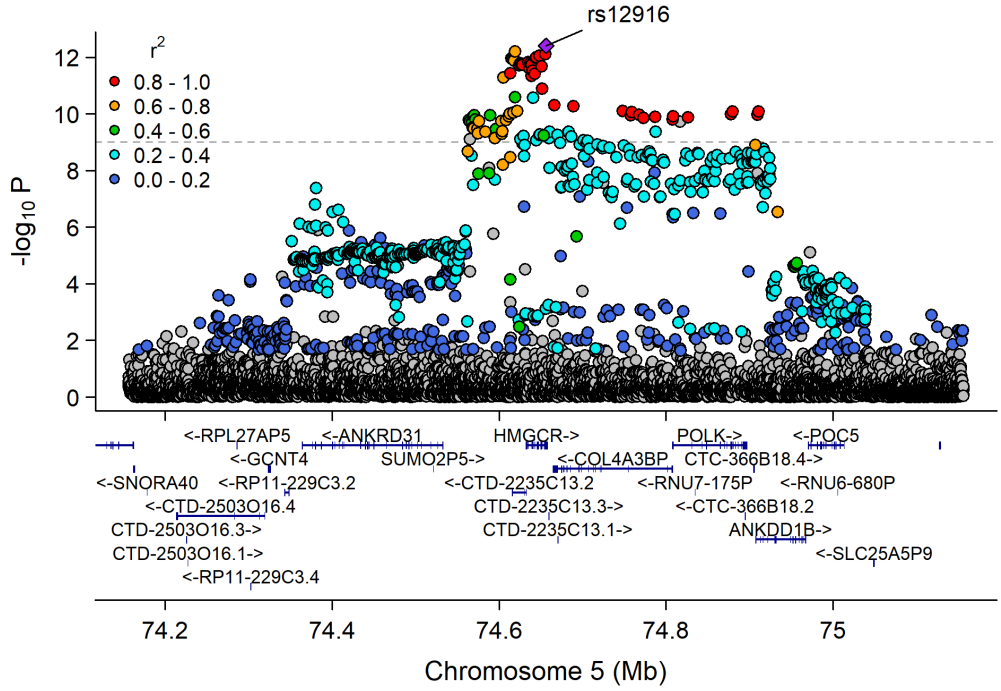
Chr2__19947287__22427492



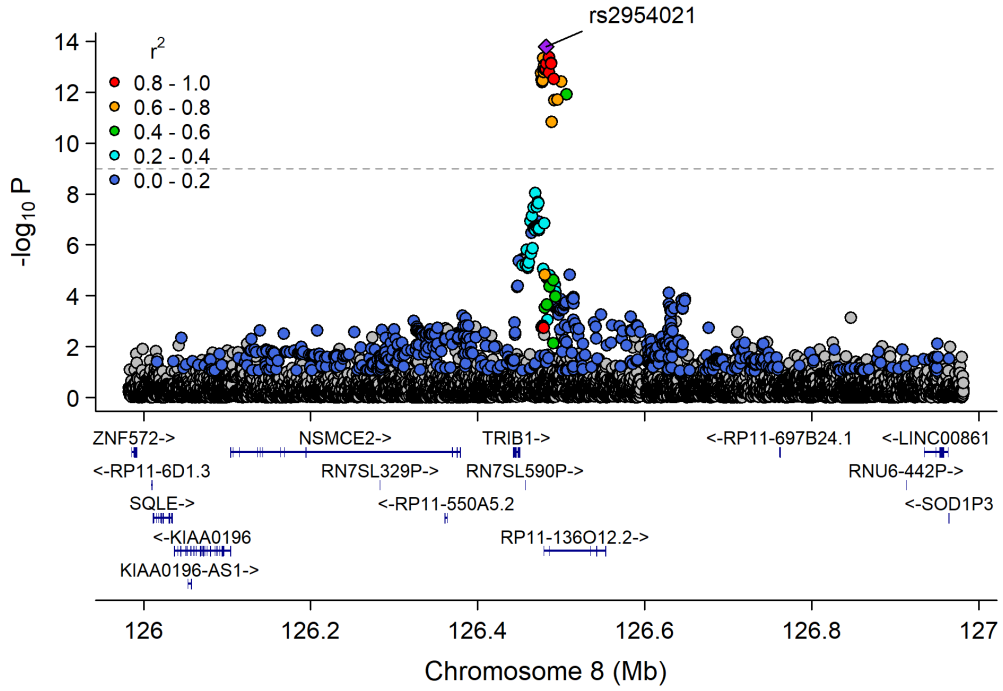
Chr4_71881824_76681229



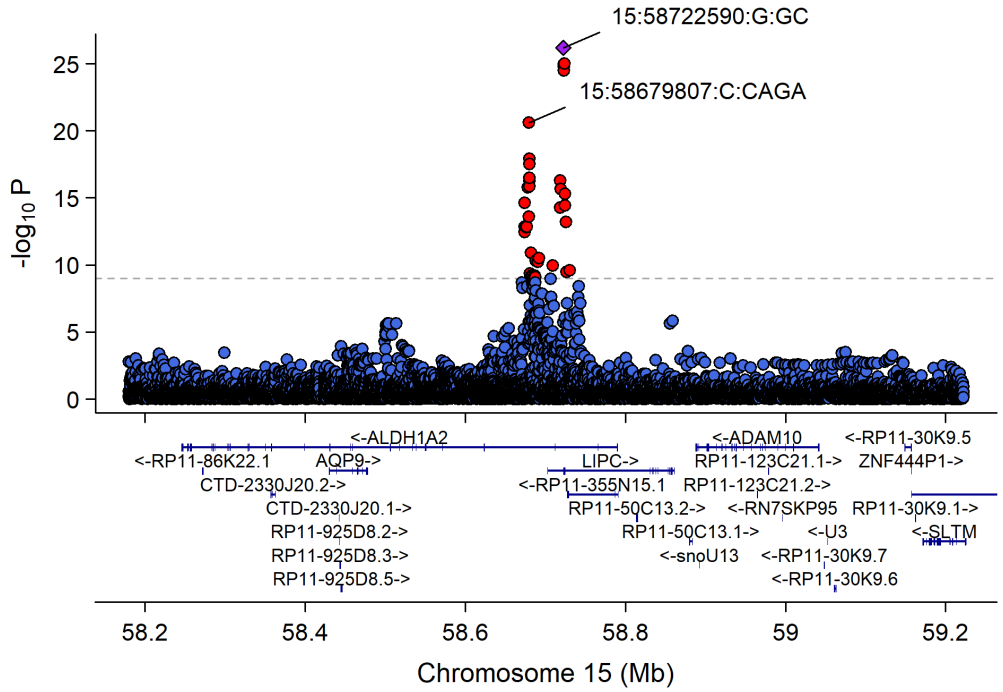
Chr5_74242002_75221582



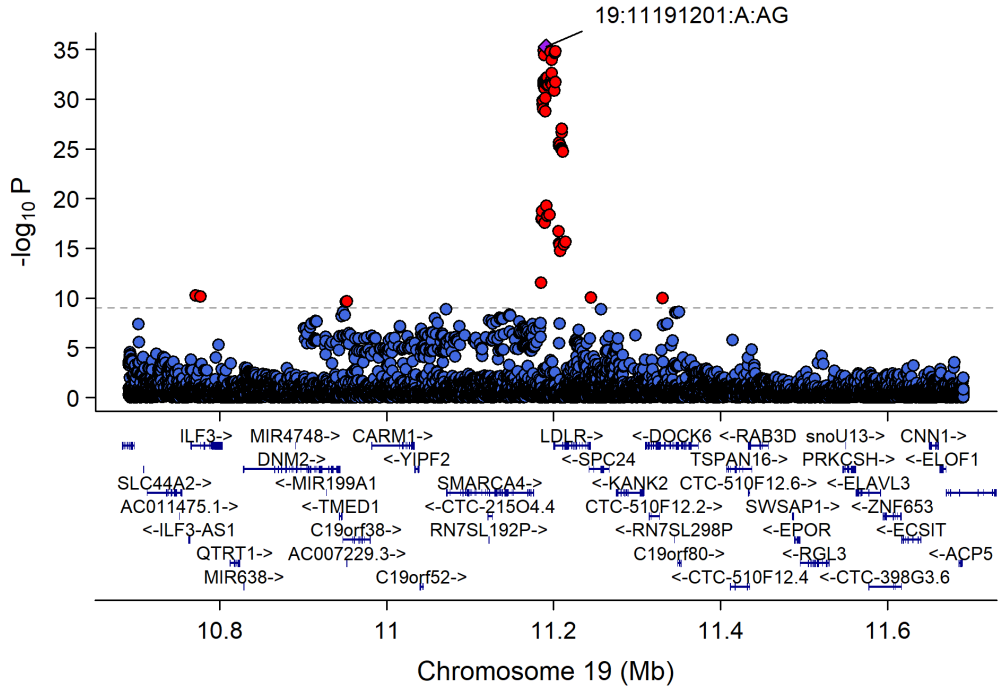
Chr8_126435663_126578061



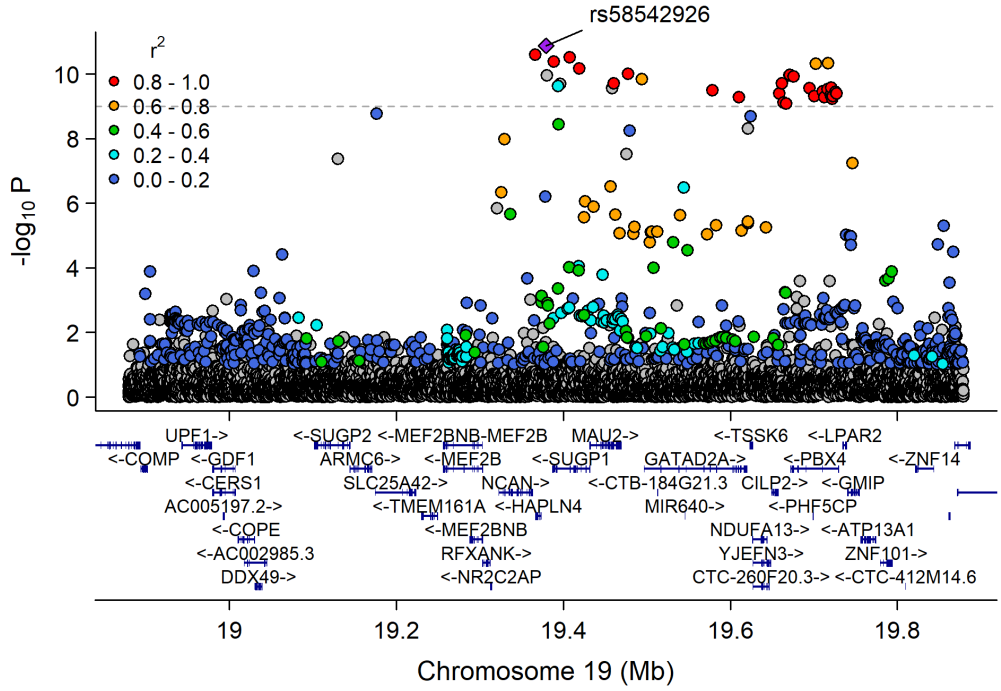
Chr15_57658798_60490883



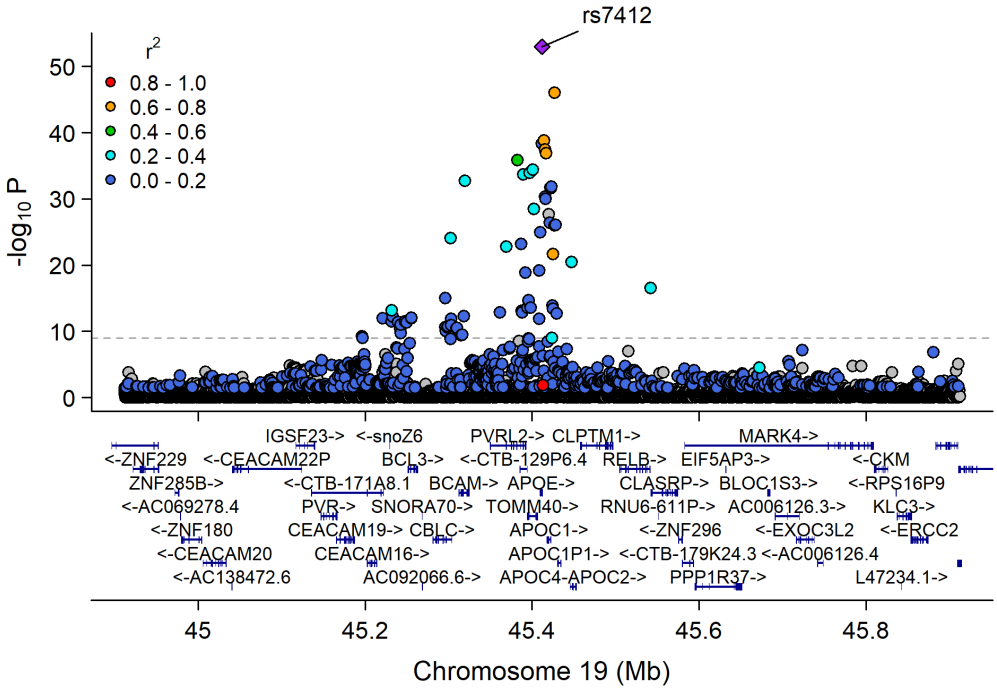
Chr19__9807999__12255594



Chr19__18370495__20841464

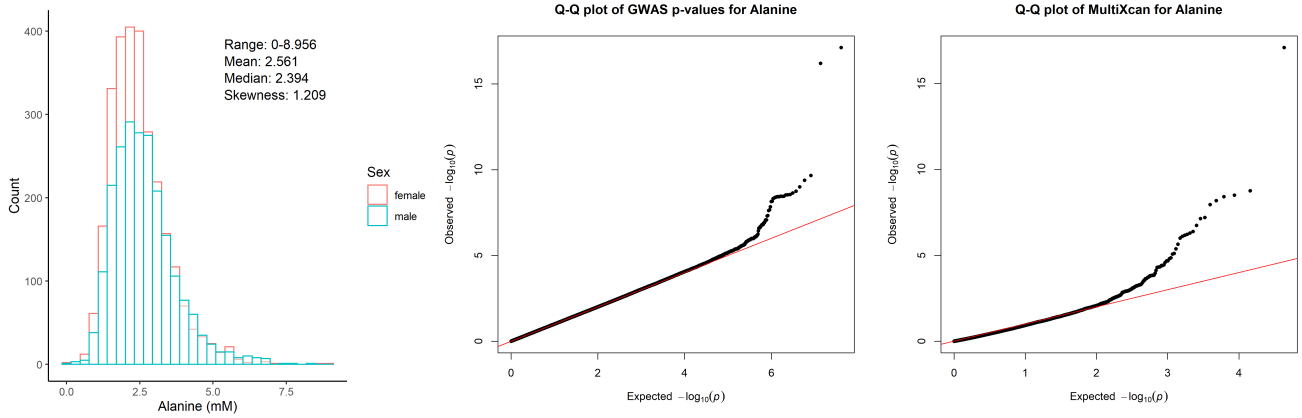


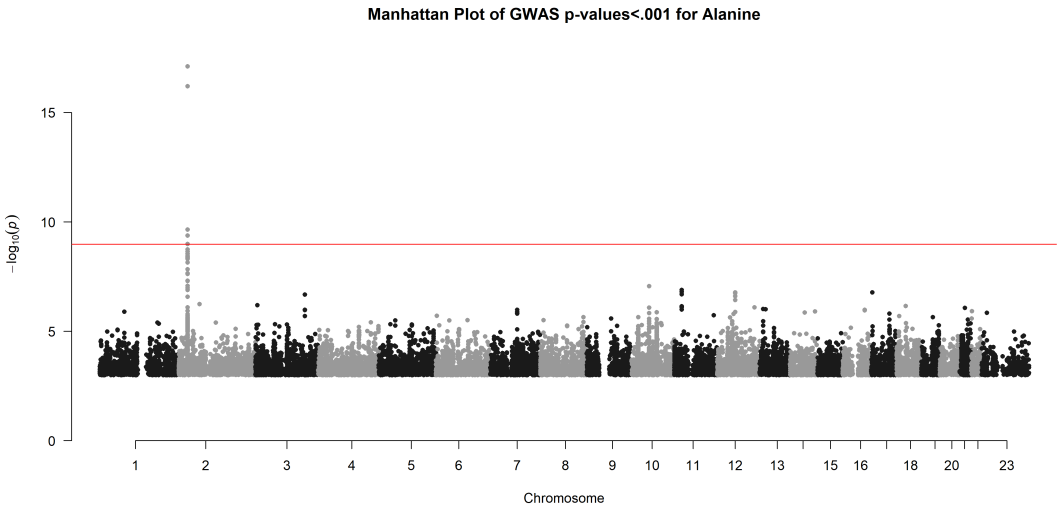
Chr19__44063363__46637375

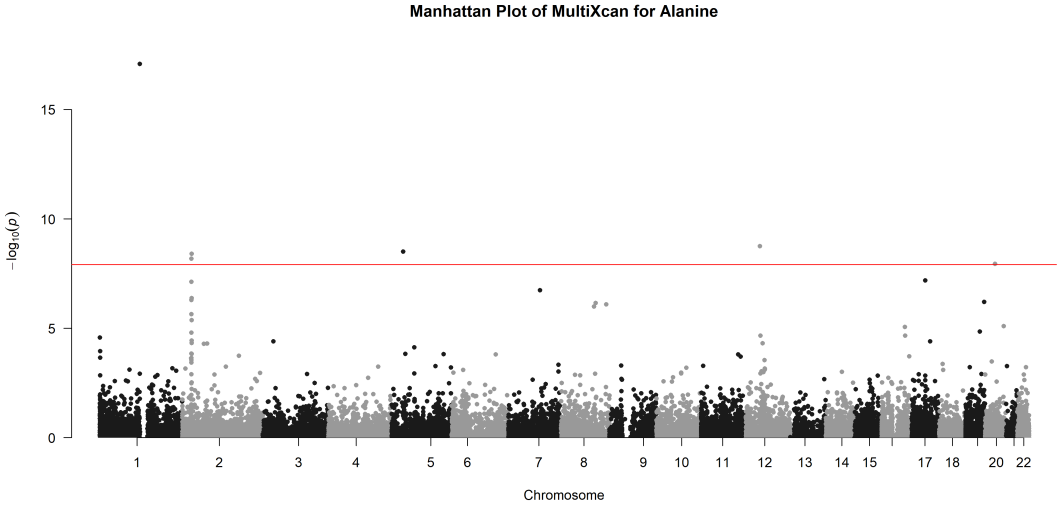


Carboxylic acids

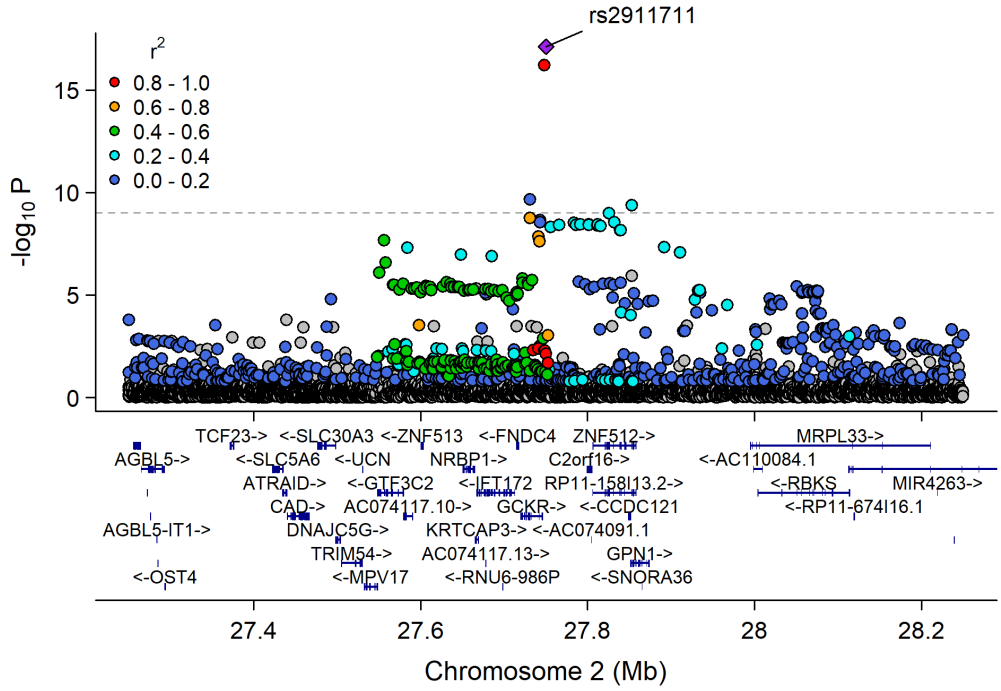
Alanine (mM)



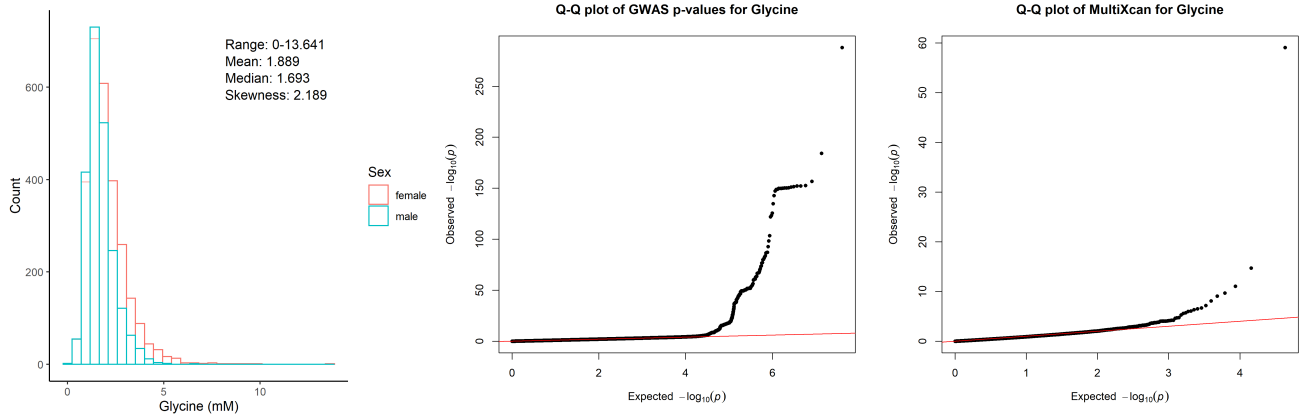


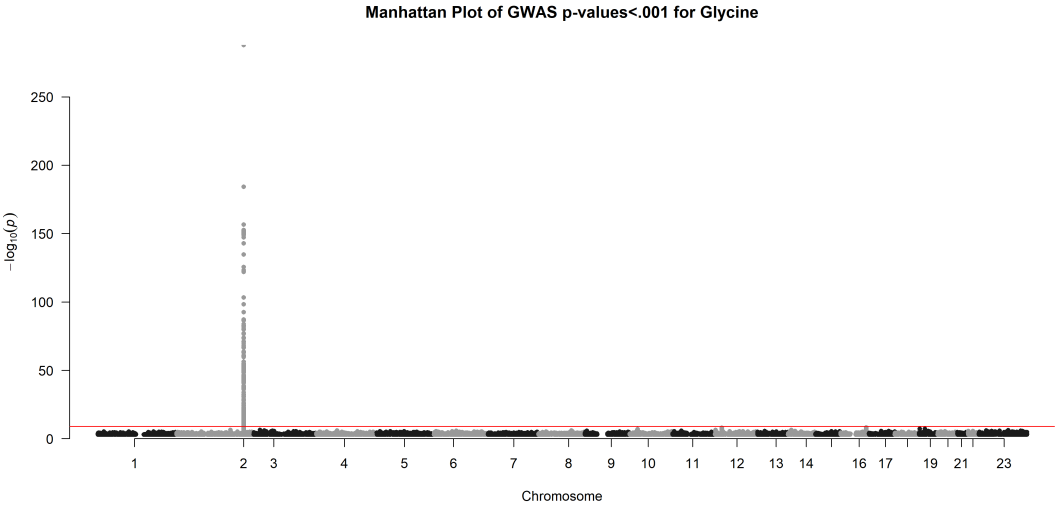


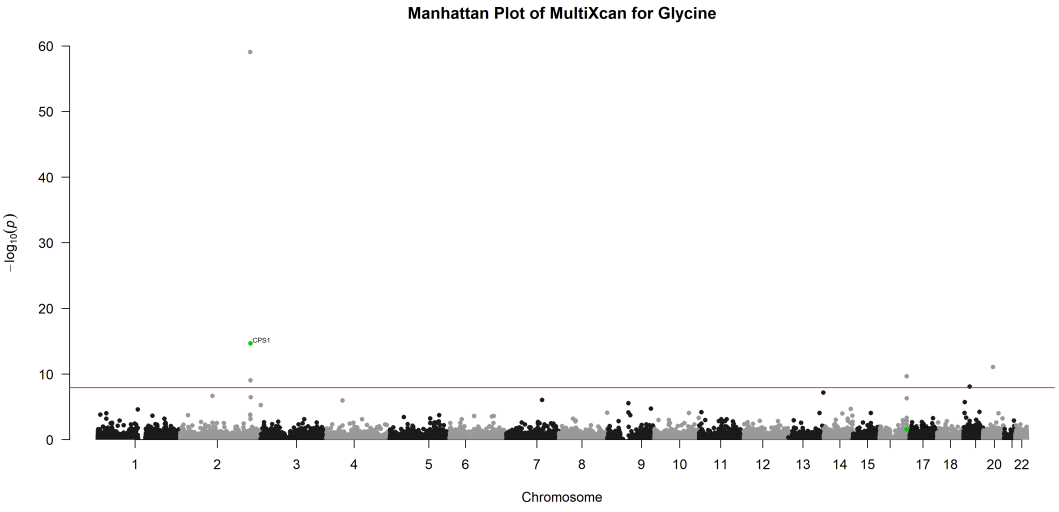
Chr2__26753815__28597624



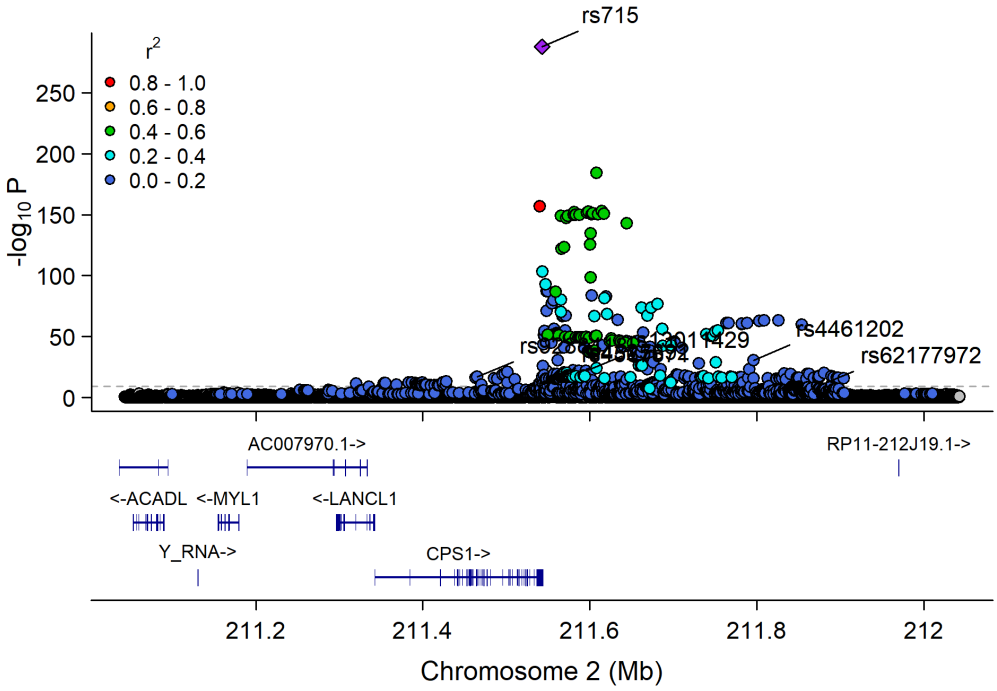
Glycine (mM)


Range: 0-13.641
Mean: 1.889
Median: 1.693
Skewness: 2.189
Count
Glycine (mM)
Sex
female
male
Q-Q plot of GWAS p-values for Glycine
Observed $-\log_{10}(p)$
Expected $-\log_{10}(p)$
Q-Q plot of MultiXcan for Glycine
Observed $-\log_{10}(p)$
Expected $-\log_{10}(p)$

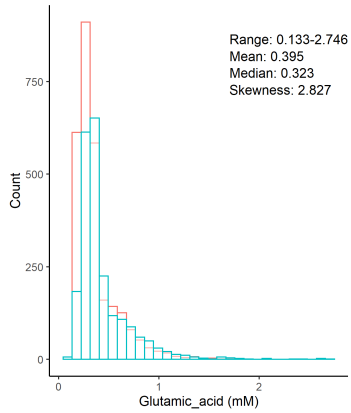

Manhattan Plot of GWAS p-values<.001 for Glycine
$-\log_{10}(p)$
Chromosome

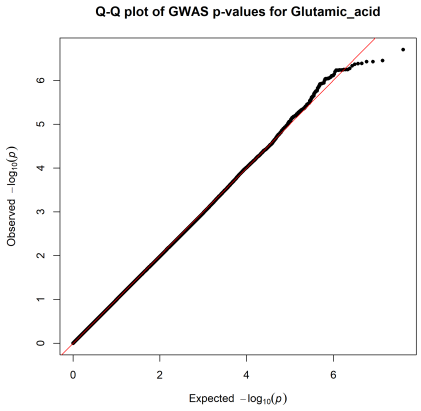

Manhattan Plot of MultiXcan for Glycine
$-\log_{10}(p)$
CPS1
Chromosome

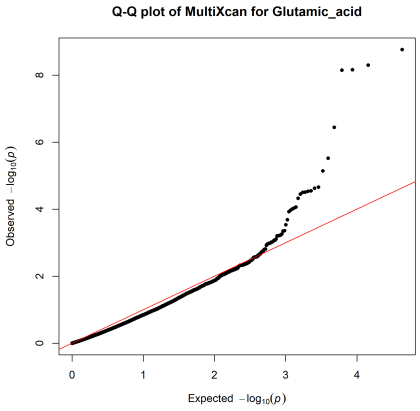
Chr2__210617214__212618476

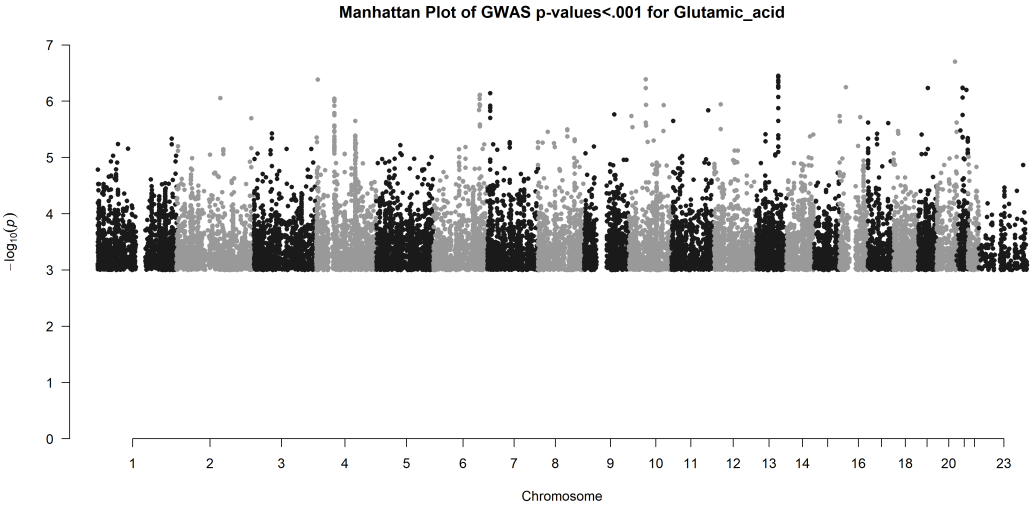


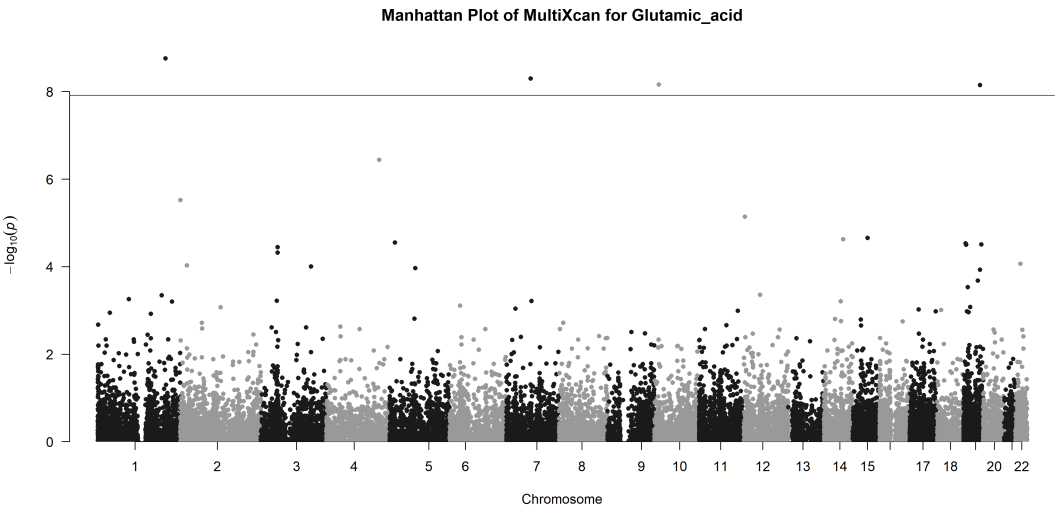
Glutamic acid (mM)



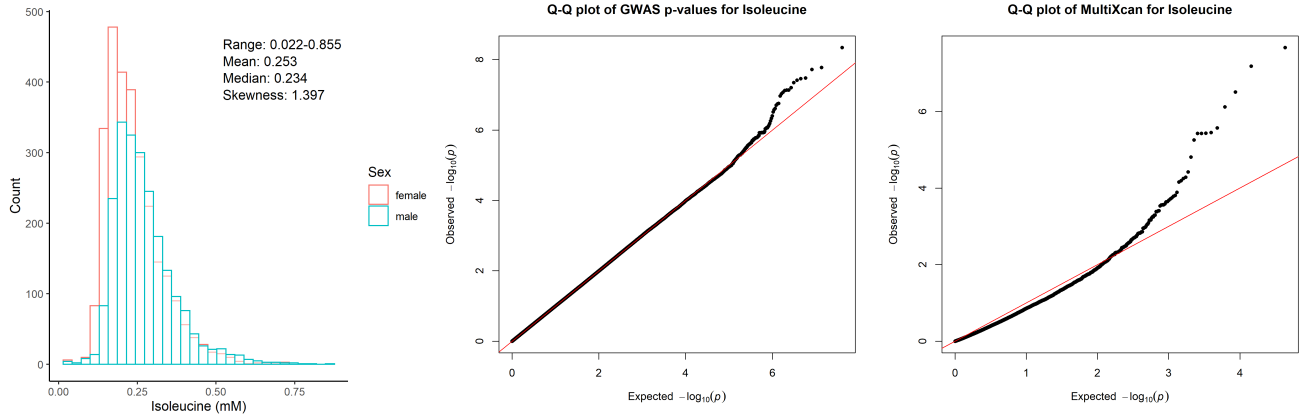


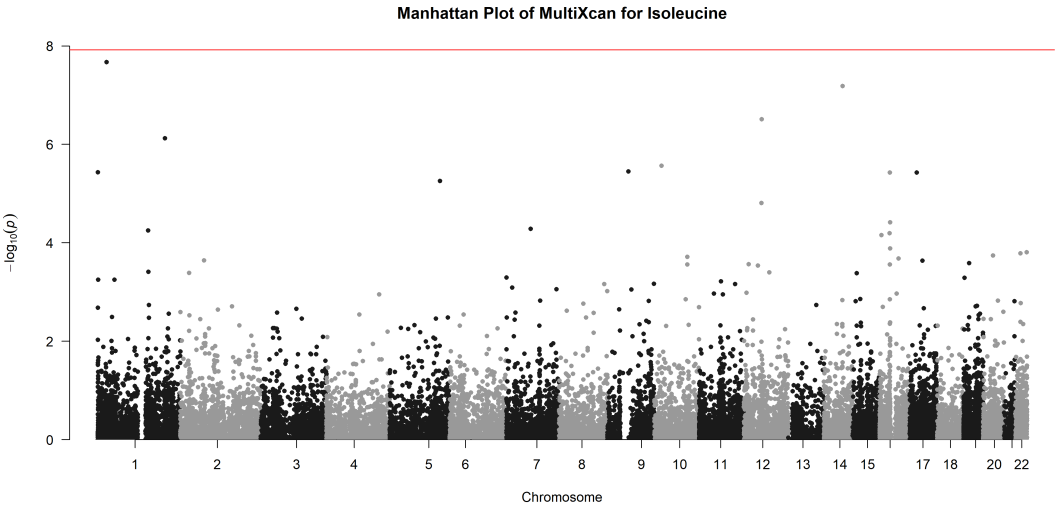




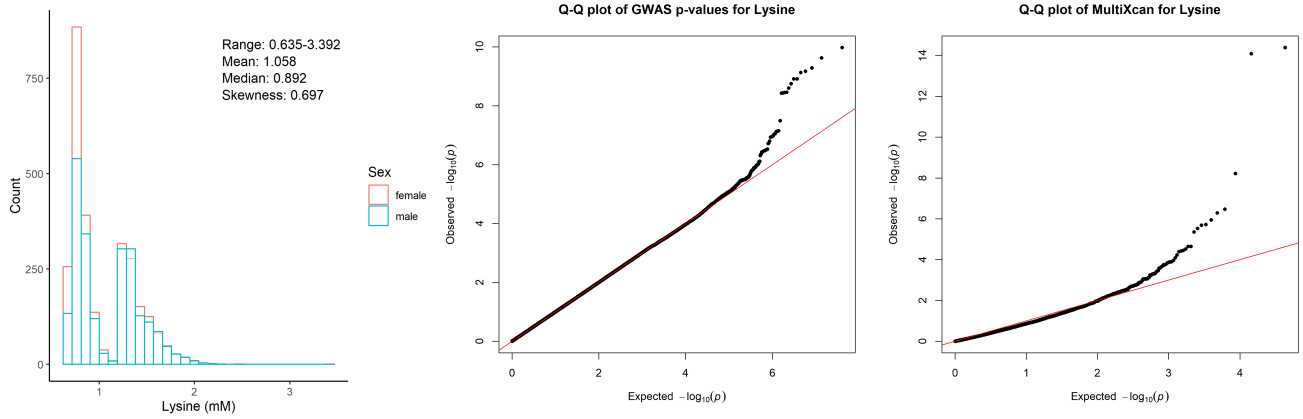


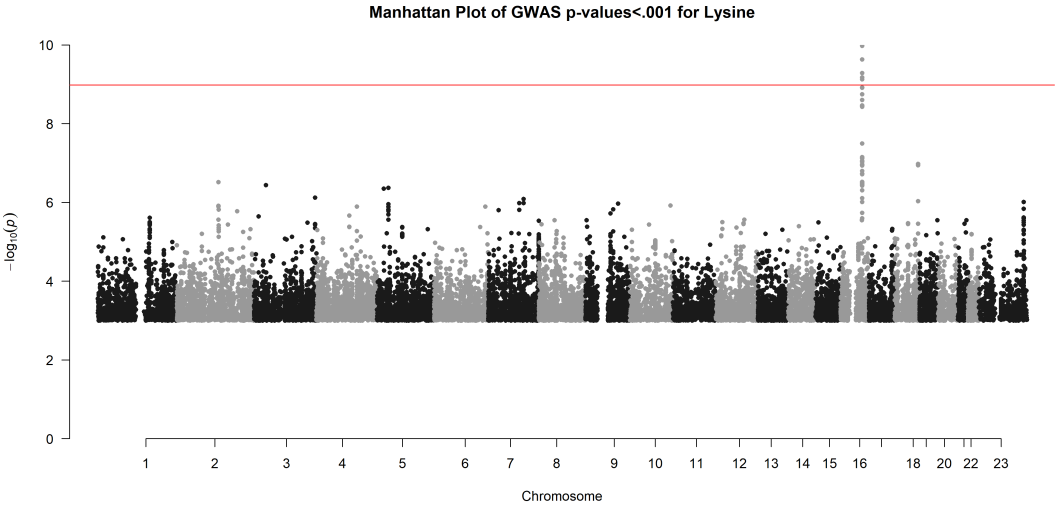
Isoleucine (mM)

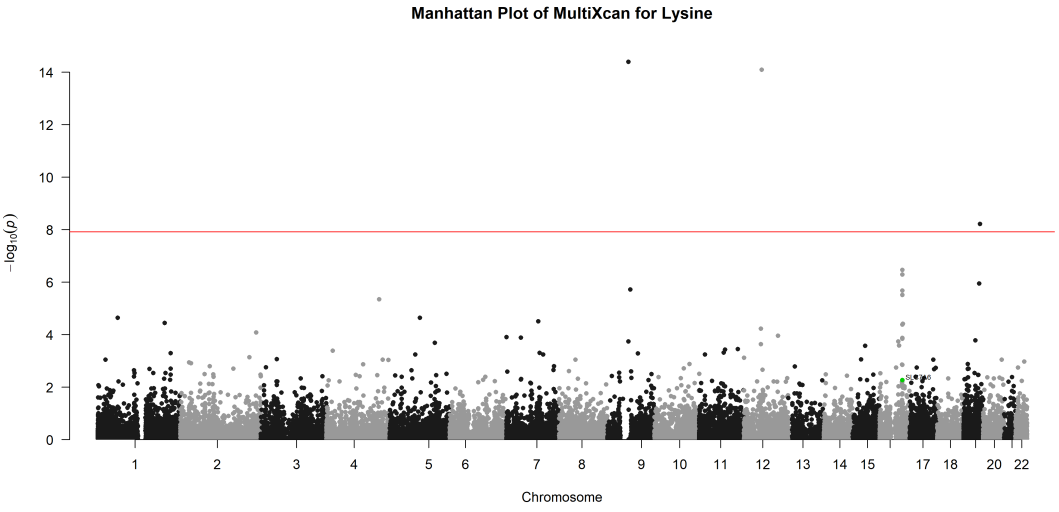




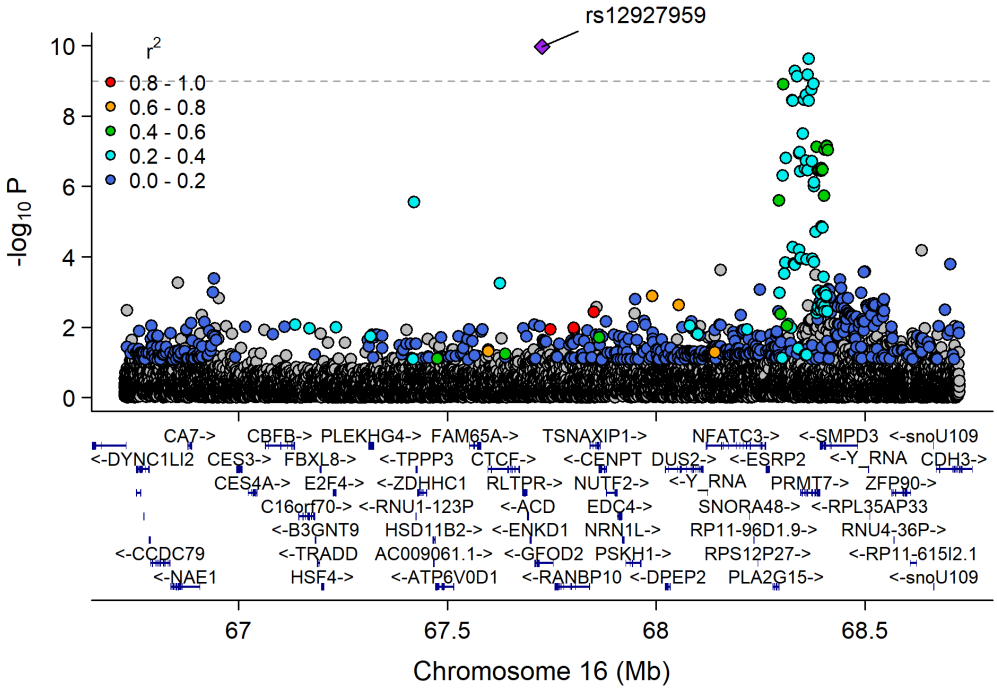
Lysine (mM)


Range: 0.635-3.392
Mean: 1.058
Median: 0.892
Skewness: 0.697
Count
Lysine (mM)
Sex
female
male
Q-Q plot of GWAS p-values for Lysine
Observed $-\log_{10}(p)$
Expected $-\log_{10}(p)$
Q-Q plot of MultiXcan for Lysine
Observed $-\log_{10}(p)$
Expected $-\log_{10}(p)$

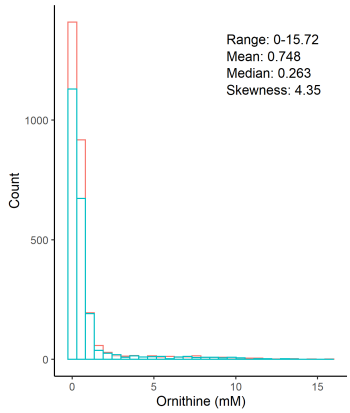

Manhattan Plot of GWAS p-values<.001 for Lysine
$-\log_{10}(p)$
Chromosome

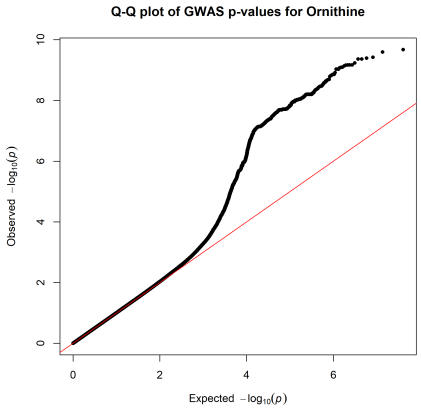

Manhattan Plot of MultiXcan for Lysine
$-\log_{10}(p)$
Chromosome

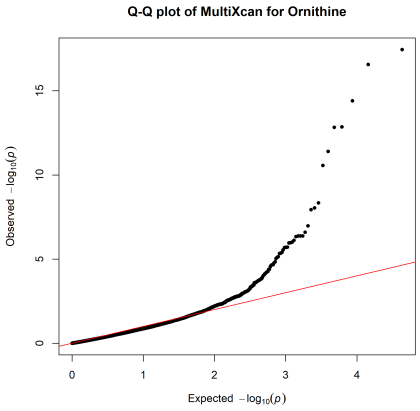
Chr16__66748612__68530942

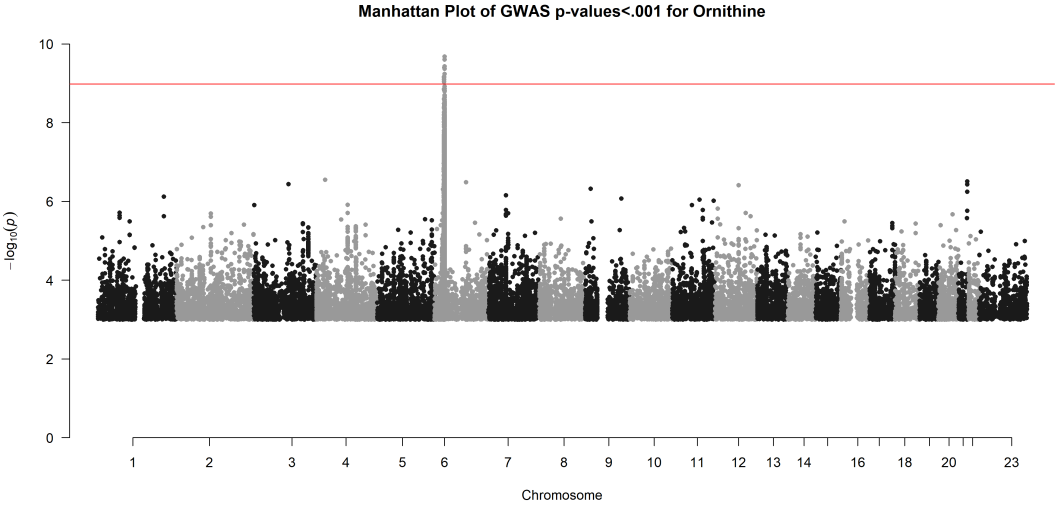


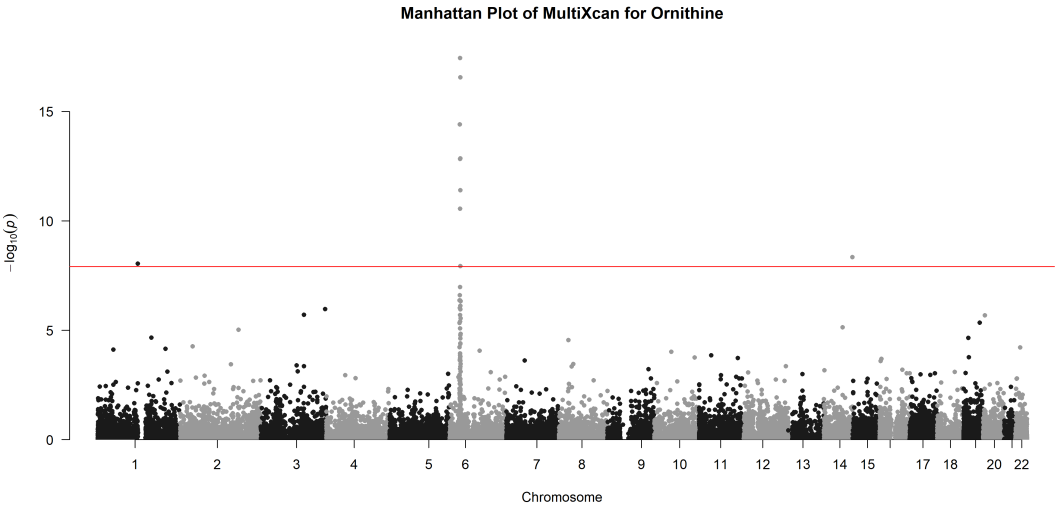
Ornithine (mM)



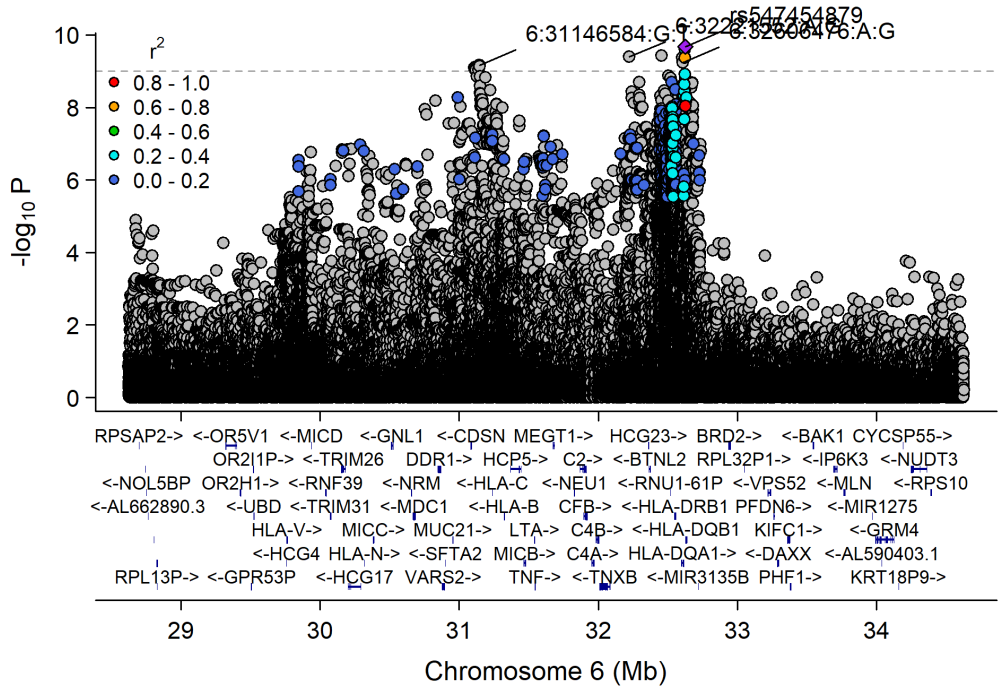




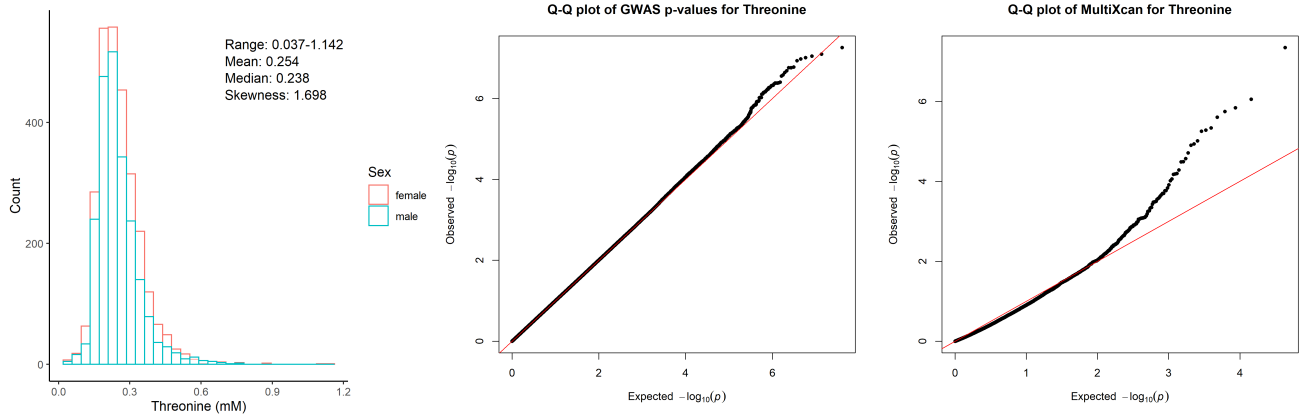


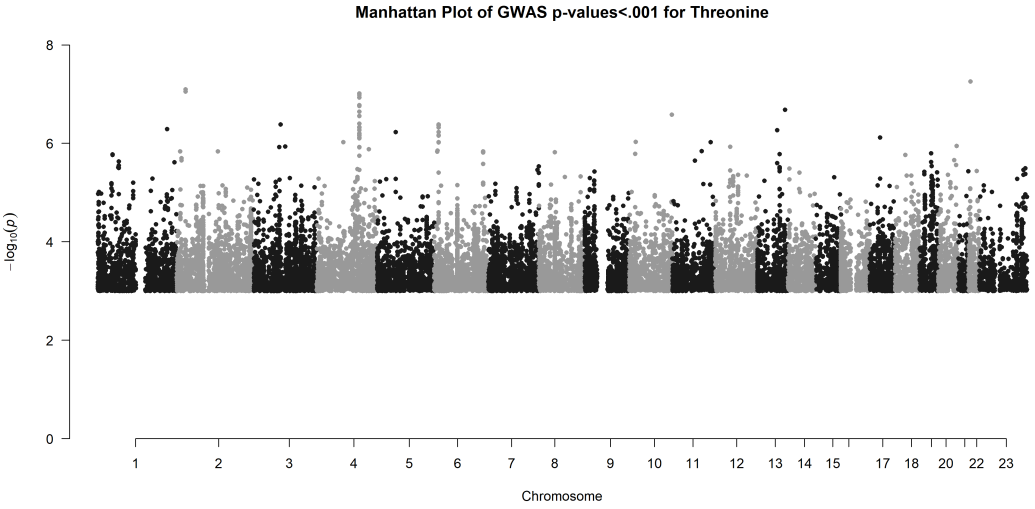


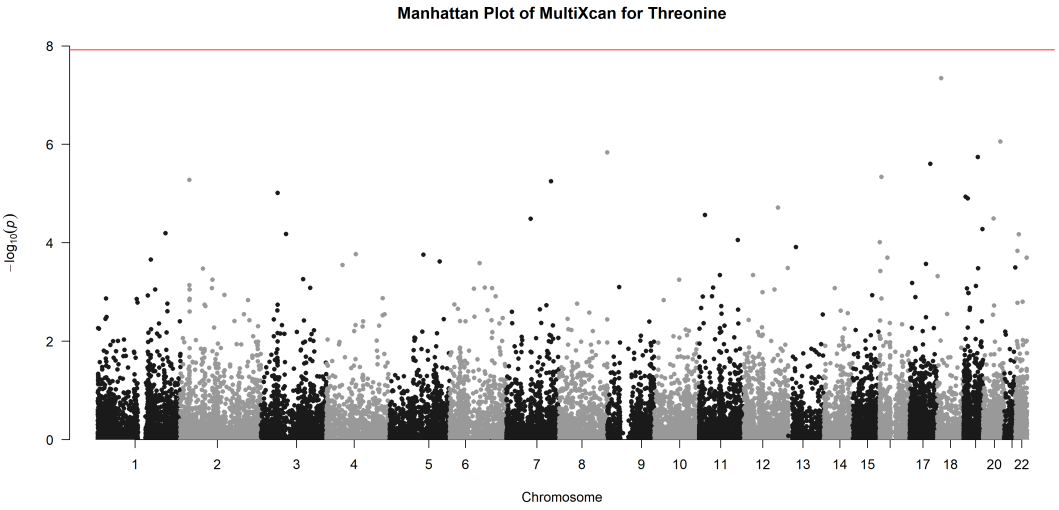
Chr6__24042708__34003583



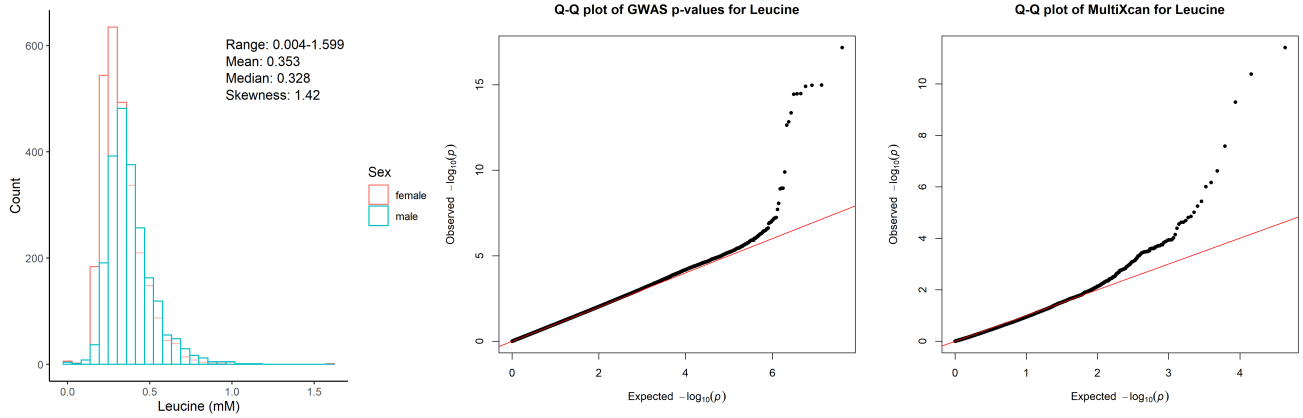
Threonine (mM)

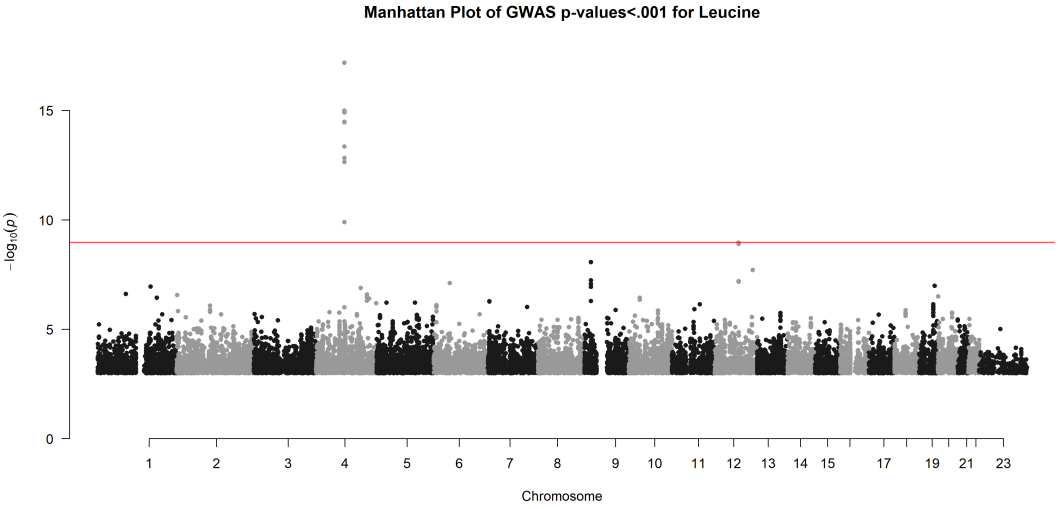


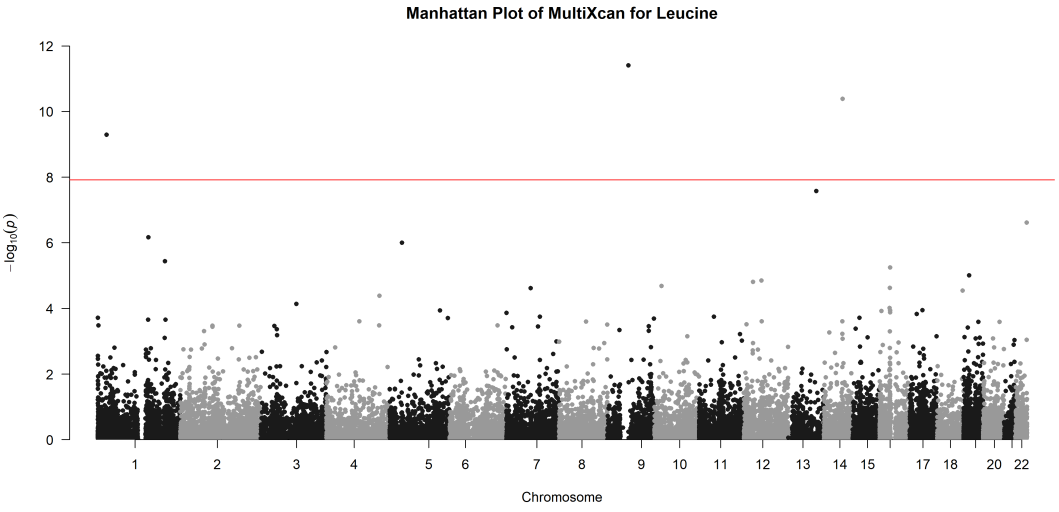




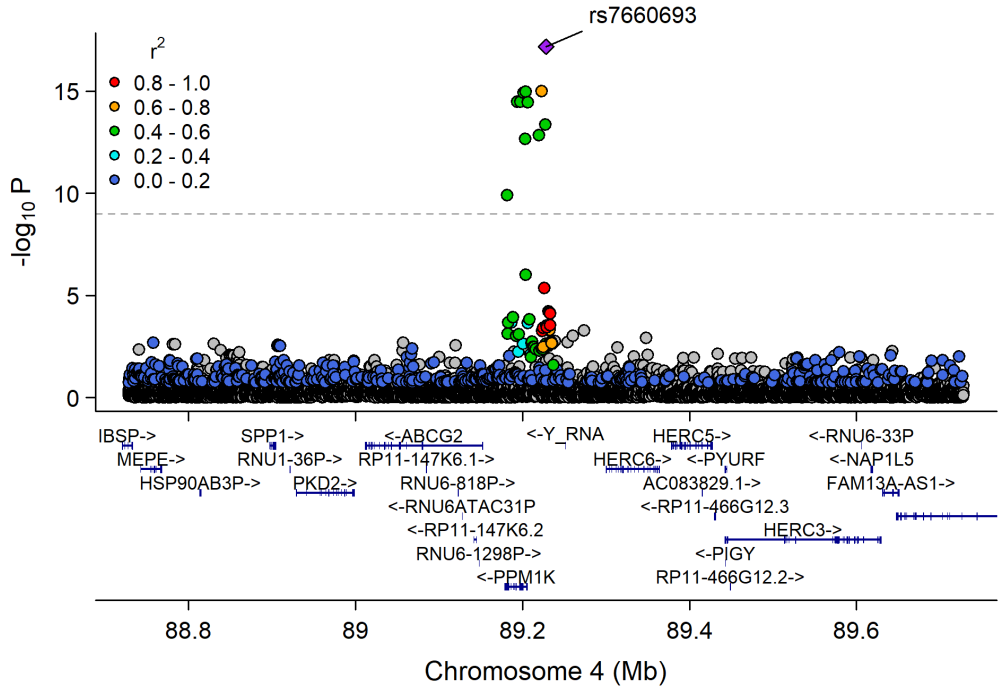
Leucine (mM)


Range: 0.004-1.599
Mean: 0.353
Median: 0.328
Skewness: 1.42
Count
Leucine (mM)
Sex
female
male
Q-Q plot of GWAS p-values for Leucine
Observed $-\log_{10}(p)$
Expected $-\log_{10}(p)$
Q-Q plot of MultiXcan for Leucine
Observed $-\log_{10}(p)$
Expected $-\log_{10}(p)$

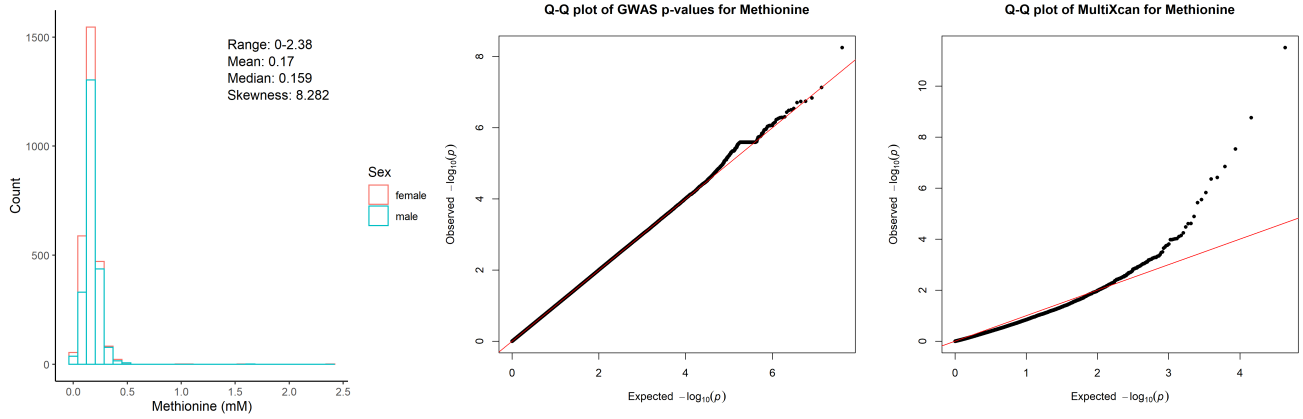

Manhattan Plot of GWAS p-values<.001 for Leucine
$-\log_{10}(p)$
Chromosome

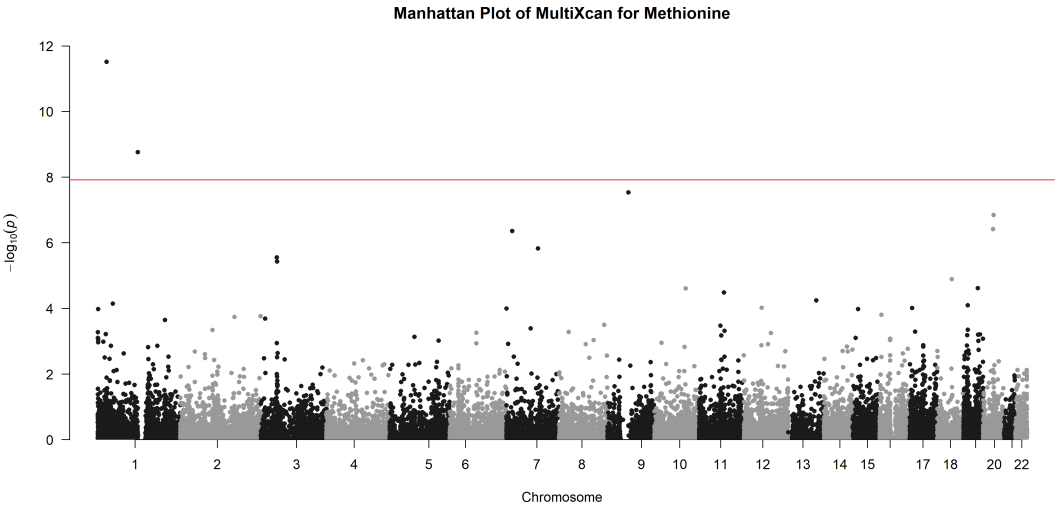

Manhattan Plot of MultiXcan for Leucine
$-\log_{10}(p)$
Chromosome

Chr4__89101913__89295163

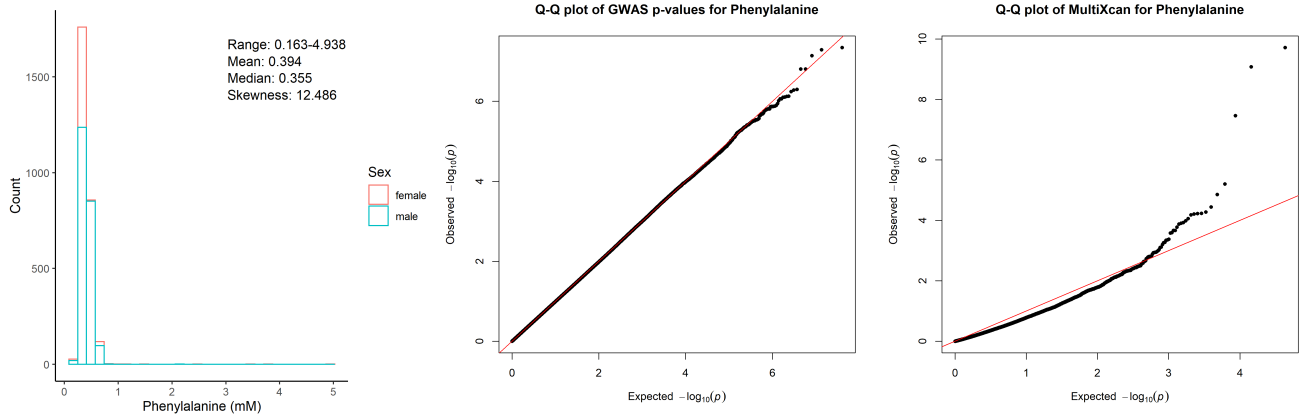


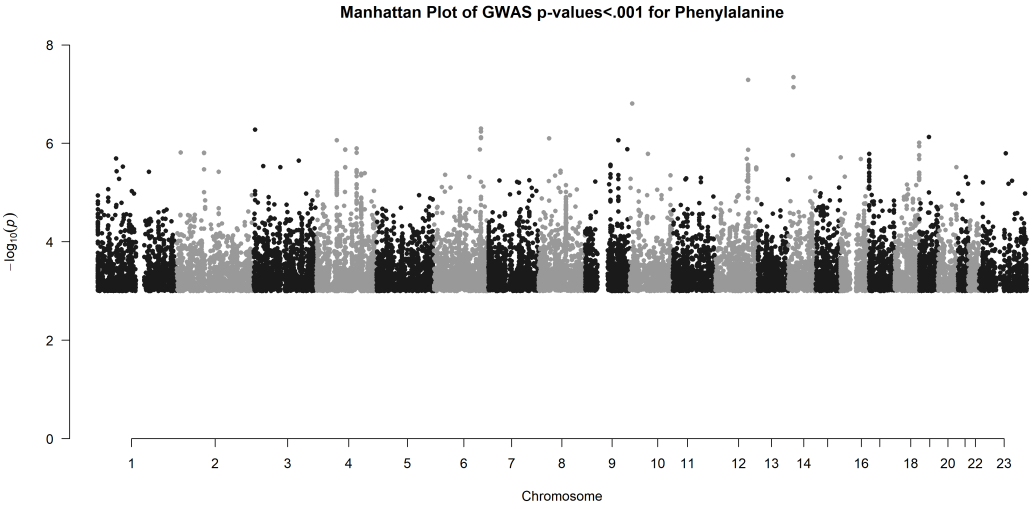
Methionine (mM)

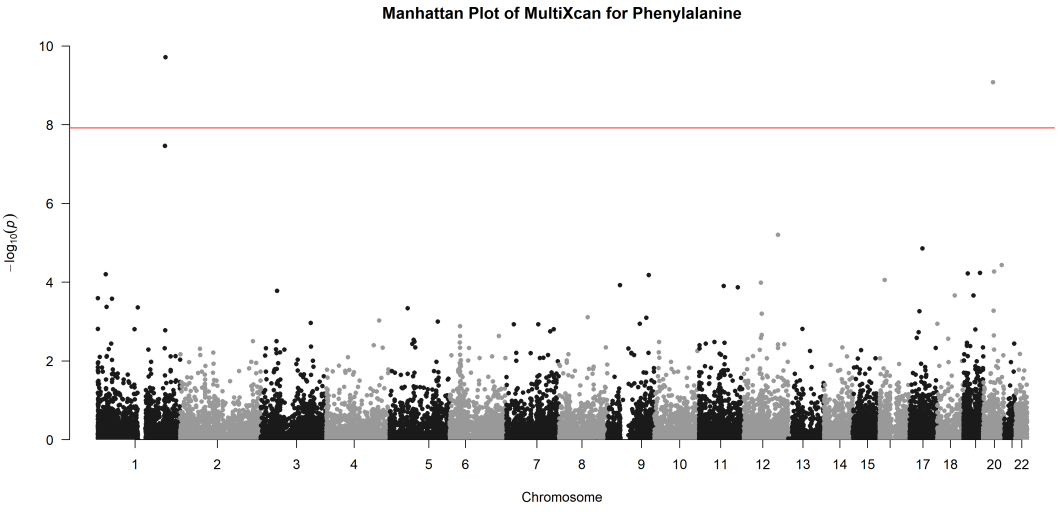




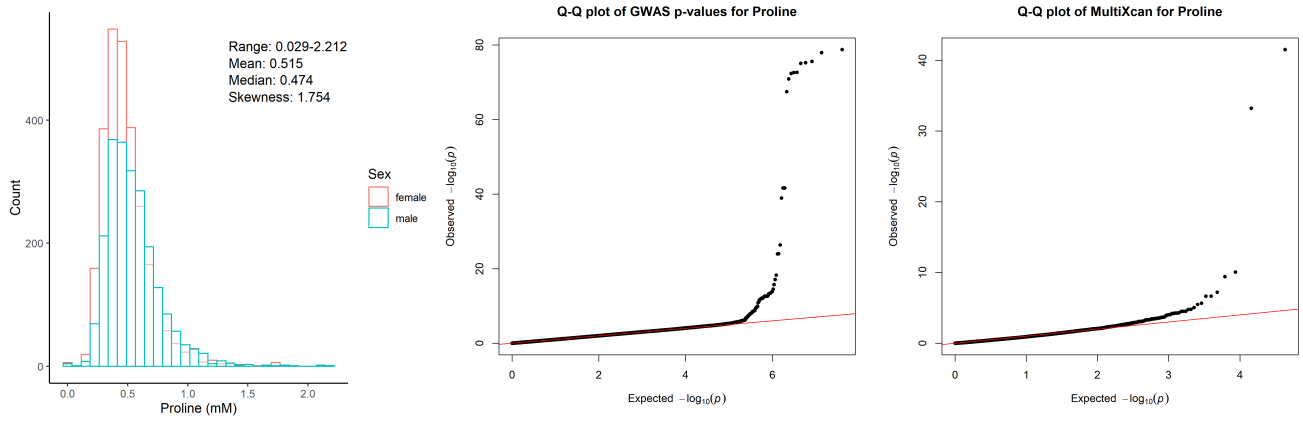
Phenylalanine (mM)

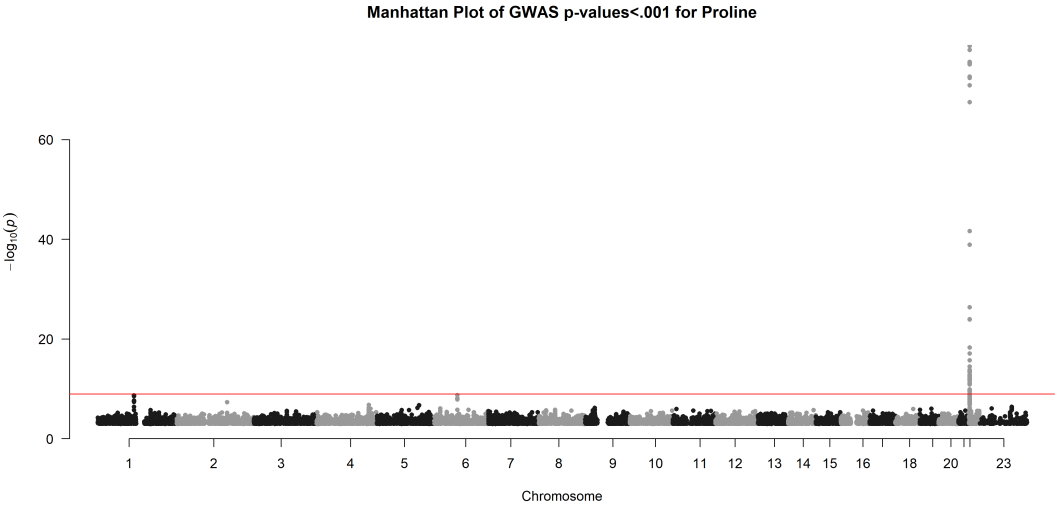


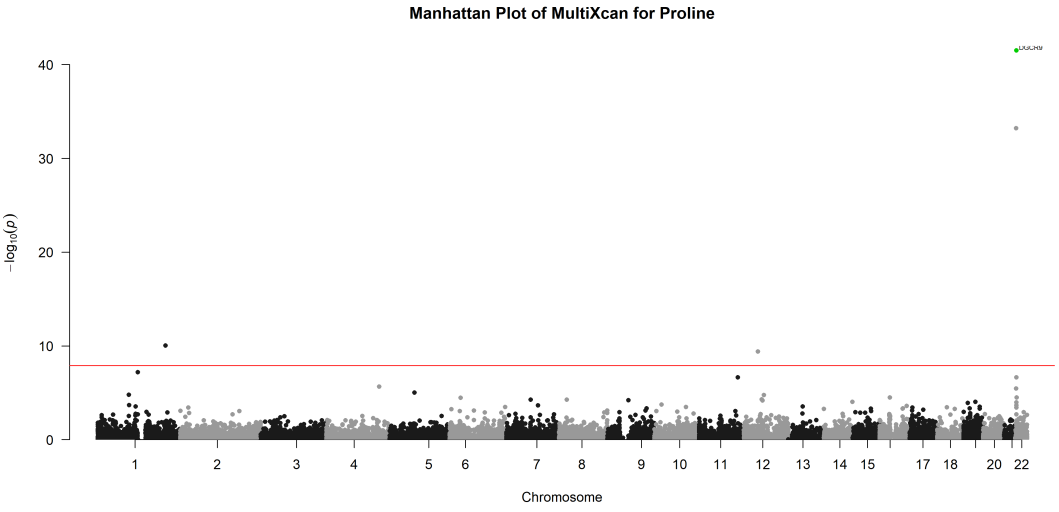




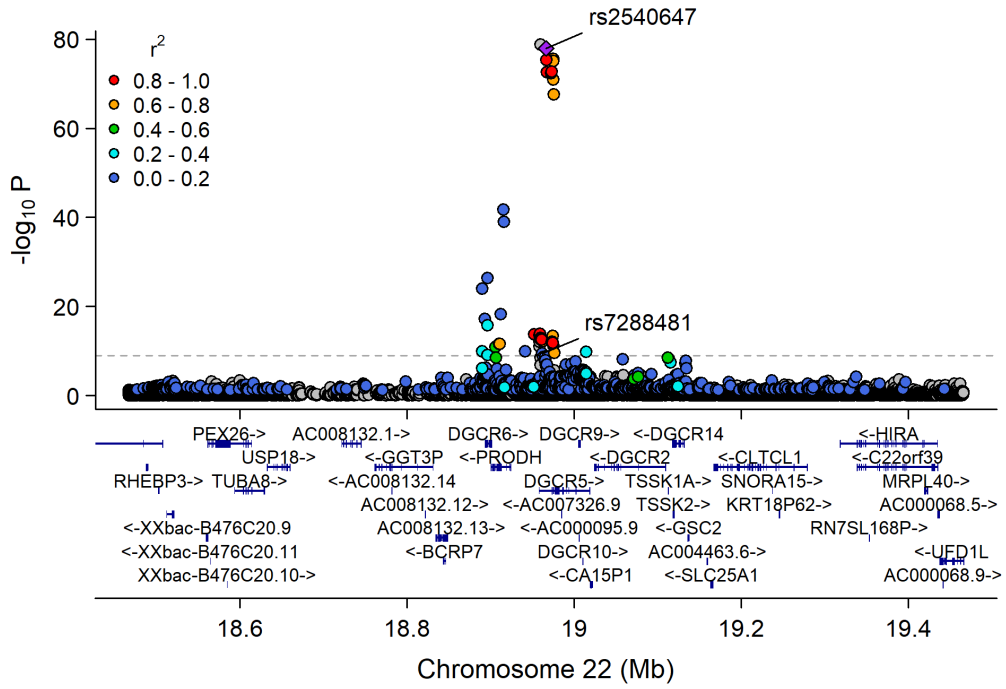
Proline (mM)


Range: 0.029-2.212
Mean: 0.515
Median: 0.474
Skewness: 1.754
Count
Proline (mM)
Sex
female
male
Q-Q plot of GWAS p-values for Proline
Observed $-\log_{10}(p)$
Expected $-\log_{10}(p)$
Q-Q plot of MultiXcan for Proline
Observed $-\log_{10}(p)$
Expected $-\log_{10}(p)$

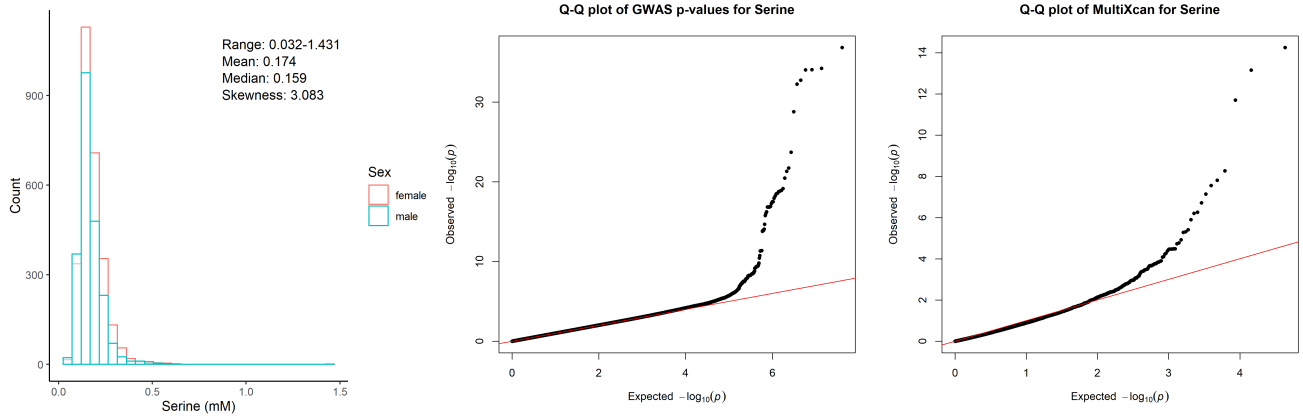

Manhattan Plot of GWAS p-values<.001 for Proline
$-\log_{10}(p)$
Chromosome

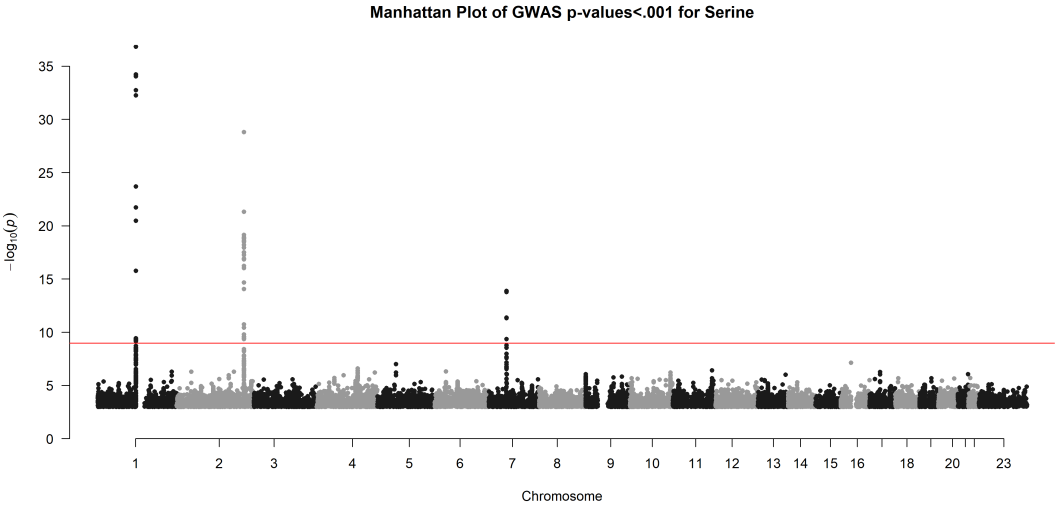

Manhattan Plot of MultiXcan for Proline
$-\log_{10}(p)$
Chromosome

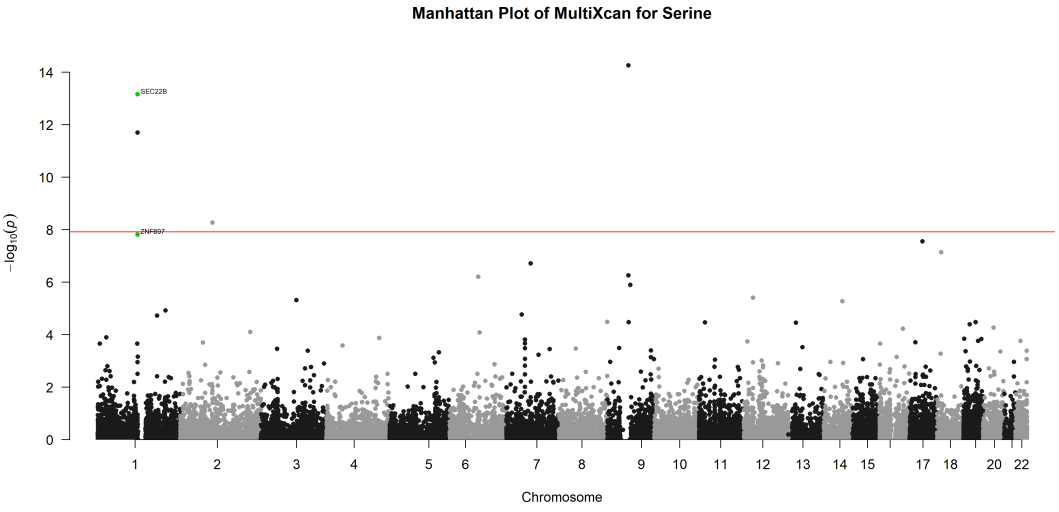
Chr22_18721052_19750317



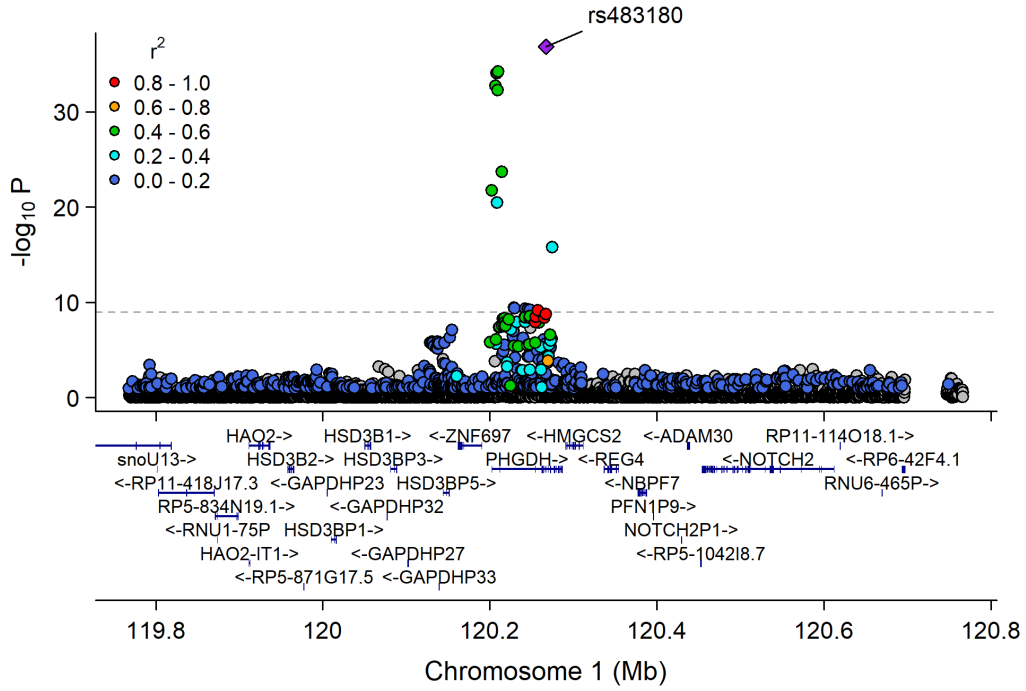
Serine (mM)


Range: 0.032-1.431
Mean: 0.174
Median: 0.159
Skewness: 3.083
Count
Serine (mM)
Sex
female
male
Q-Q plot of GWAS p-values for Serine
Observed $-\log_{10}(p)$
Expected $-\log_{10}(p)$
Q-Q plot of MultiXcan for Serine
Observed $-\log_{10}(p)$
Expected $-\log_{10}(p)$

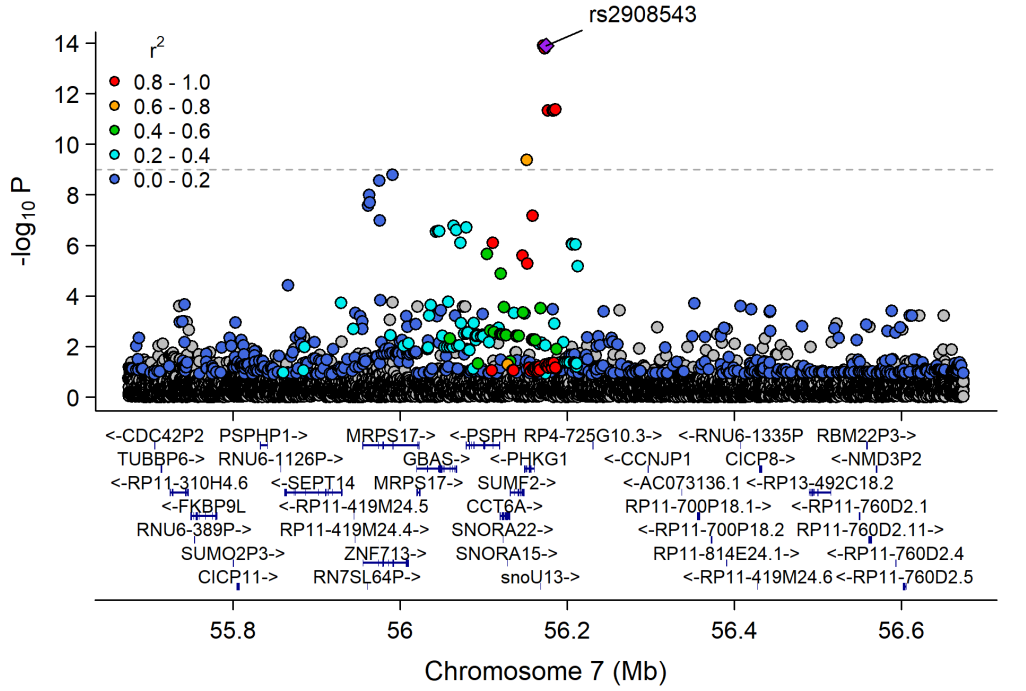

Manhattan Plot of GWAS p-values<.001 for Serine
$-\log_{10}(p)$
Chromosome


Manhattan Plot of MultiXcan for Serine
SEC22B
ZNF697
$-\log_{10}(p)$
Chromosome

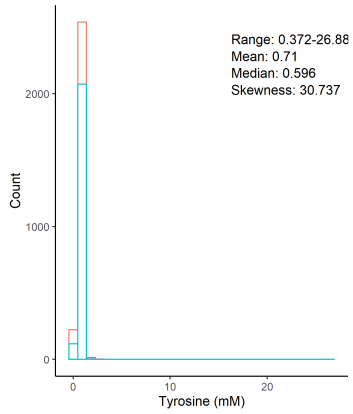
Chr1_119992162_120392714

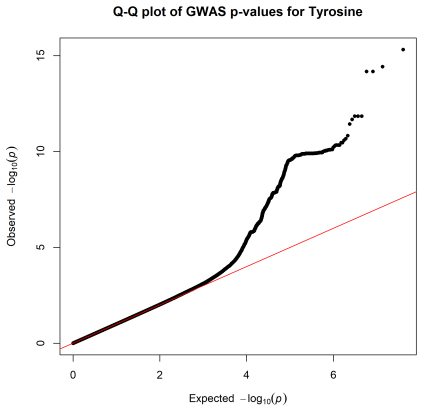


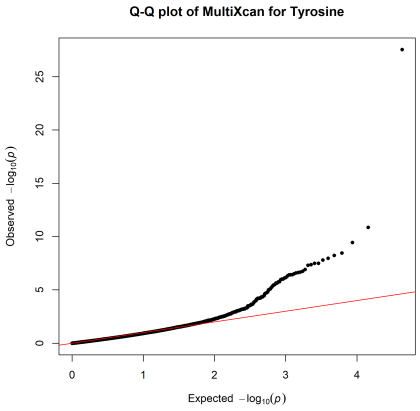
Chr7_55765998_56825777

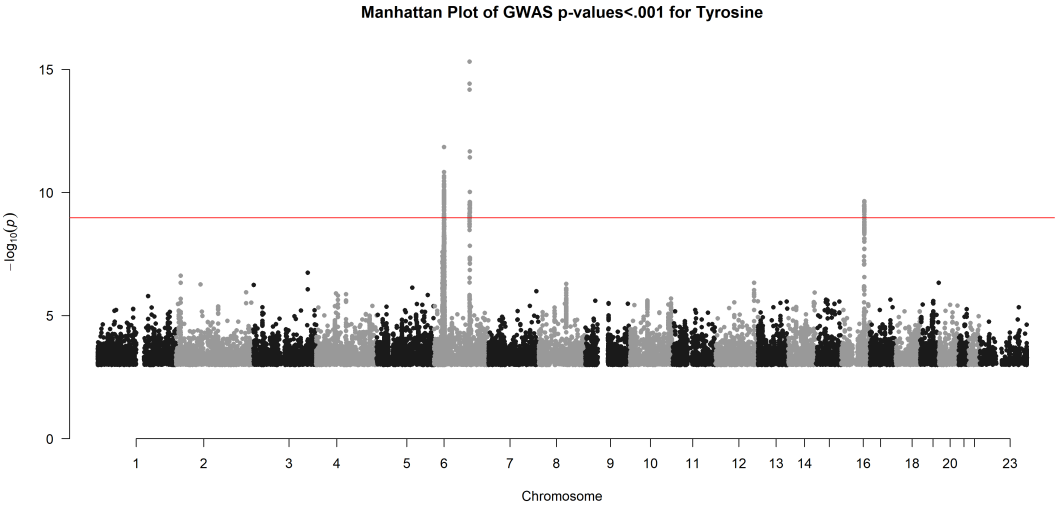


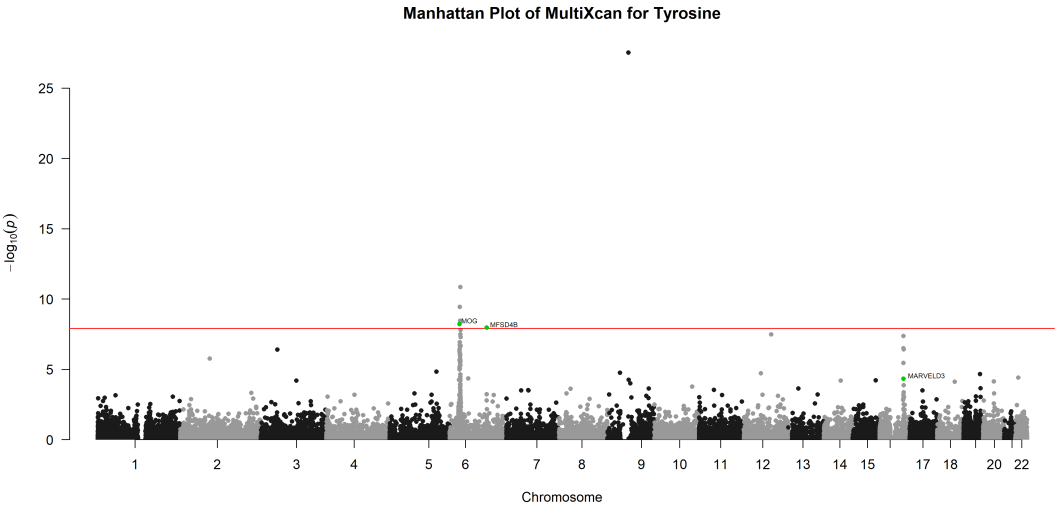
Tyrosine (mM)


Range: 0.372-26.88
Mean: 0.71
Median: 0.596
Skewness: 30.737
Count
Tyrosine (mM)
Sex
female
male

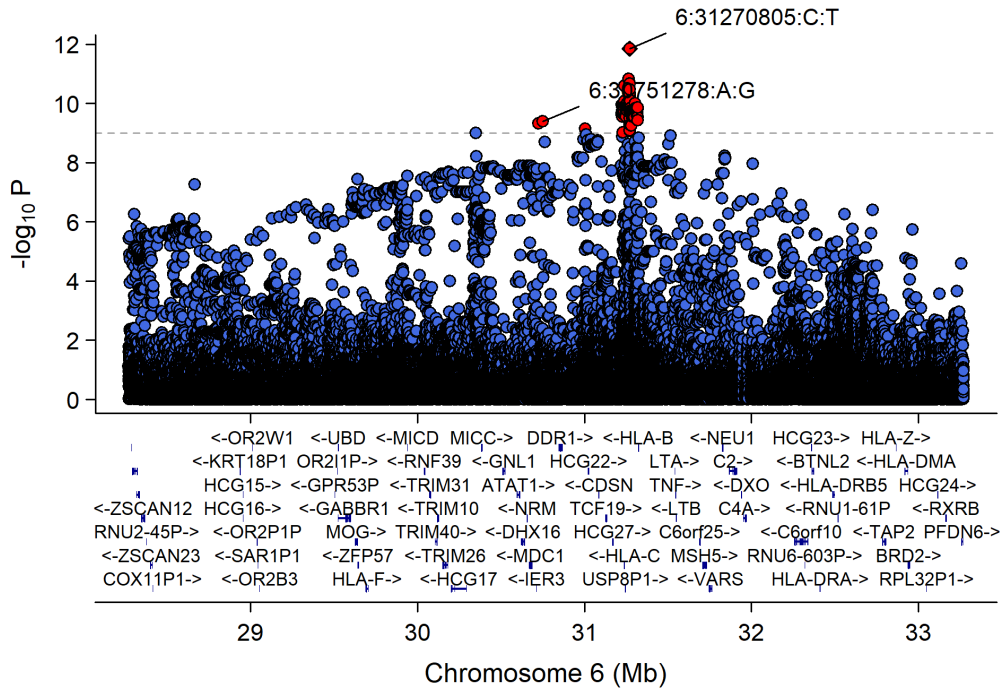

Q-Q plot of GWAS p-values for Tyrosine
Observed $-\log_{10}(p)$
Expected $-\log_{10}(p)$


Q-Q plot of MultiXcan for Tyrosine
Observed $-\log_{10}(p)$
Expected $-\log_{10}(p)$

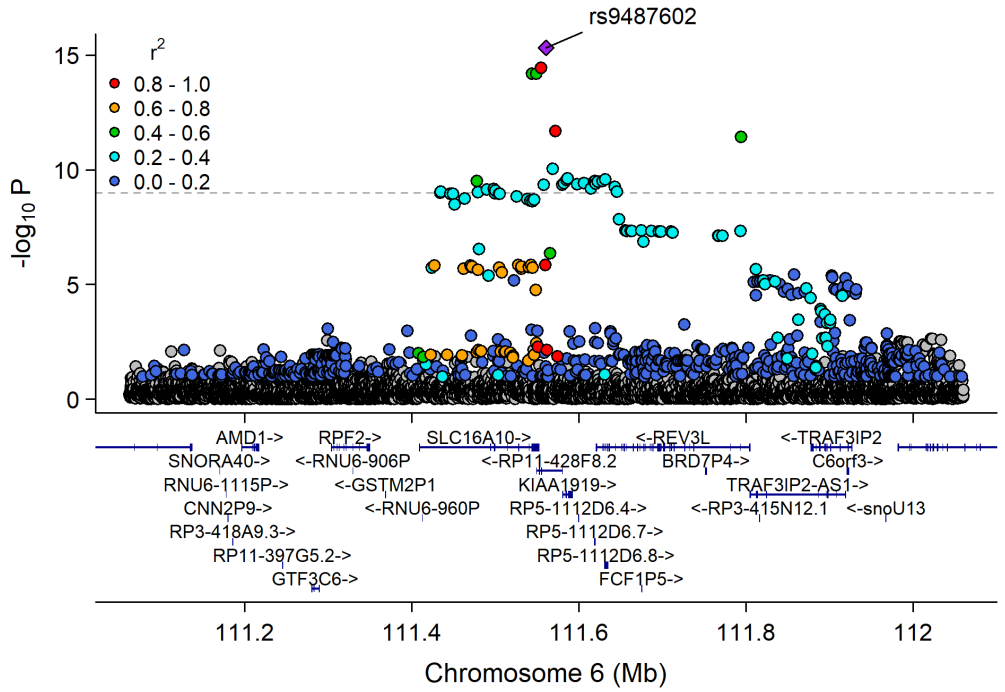

Manhattan Plot of GWAS p-values<.001 for Tyrosine
$-\log_{10}(p)$
Chromosome


Manhattan Plot of MultiXcan for Tyrosine
$-\log_{10}(p)$
MOG
MFSD4B
MARVELD3
Chromosome

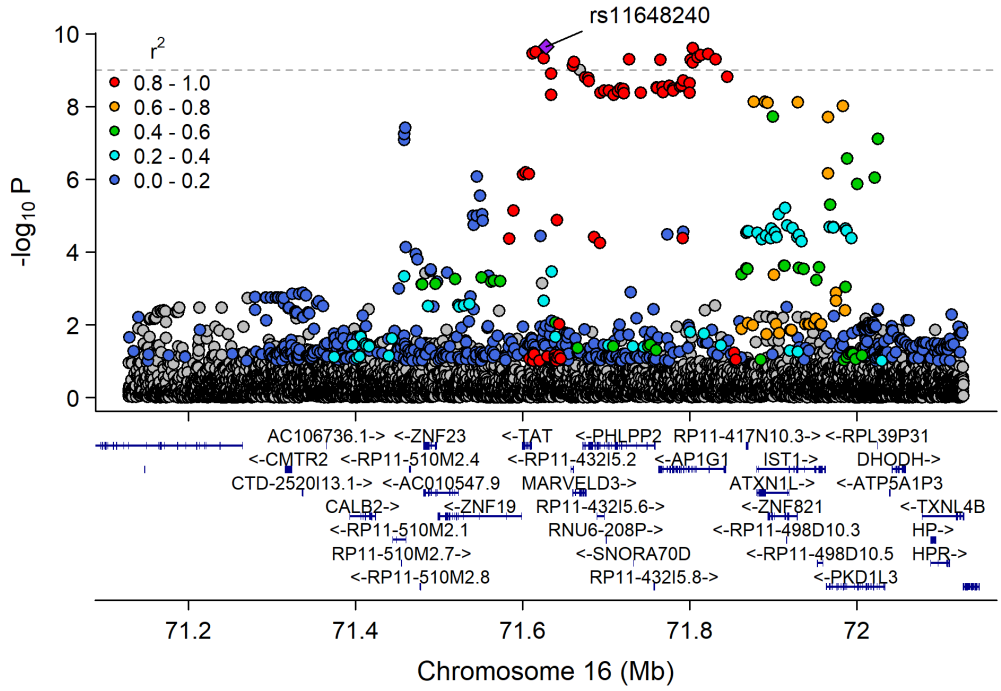
Chr6__24042708__34003583



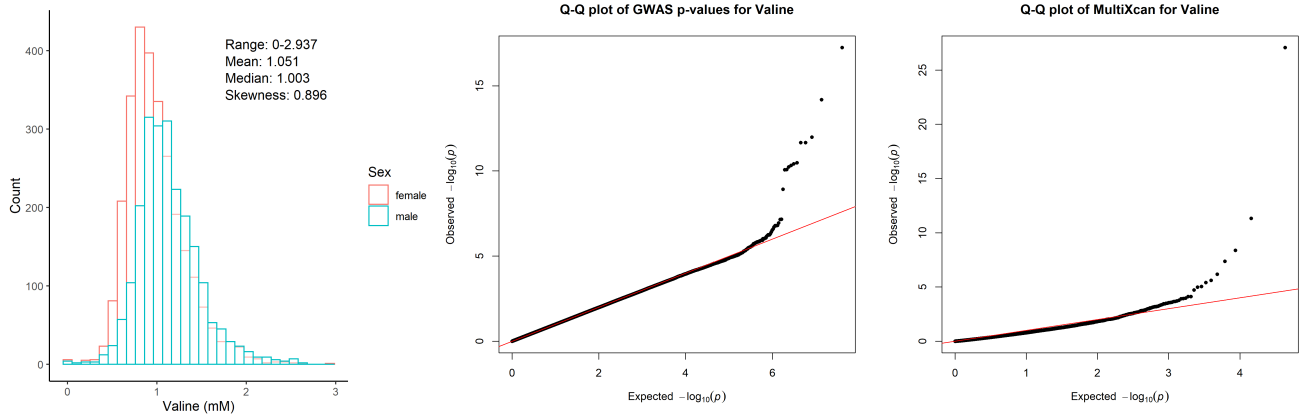
Chr6__111393318__111944979

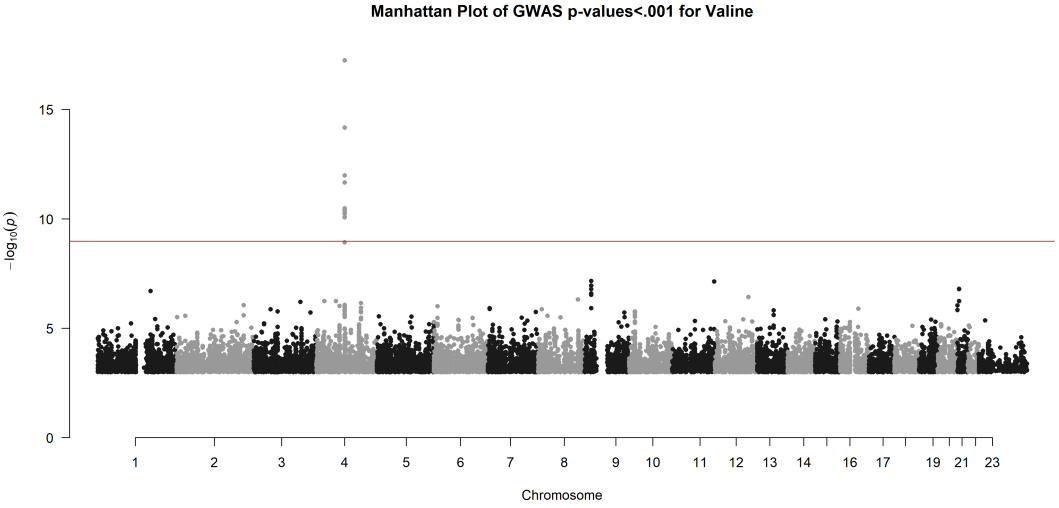


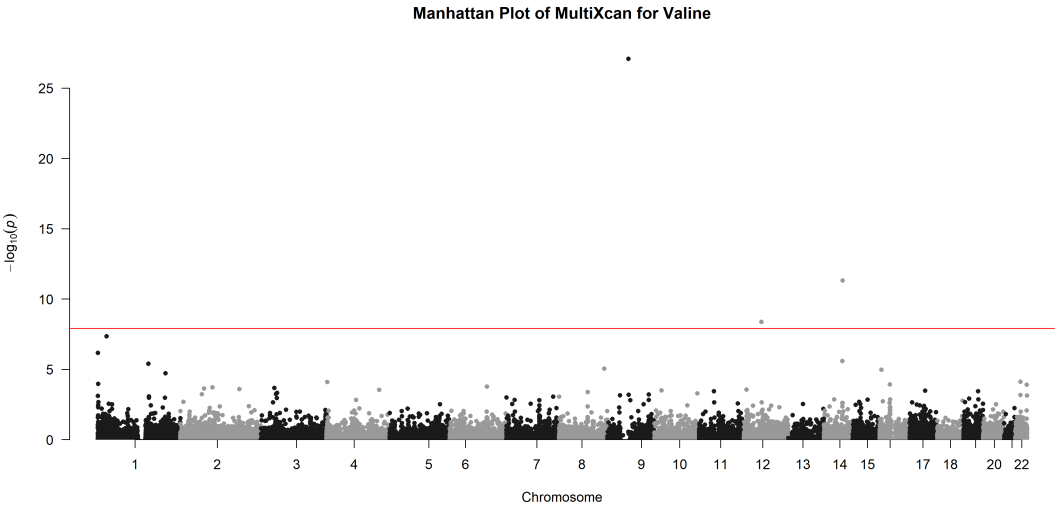
Chr16__70755610__72626132



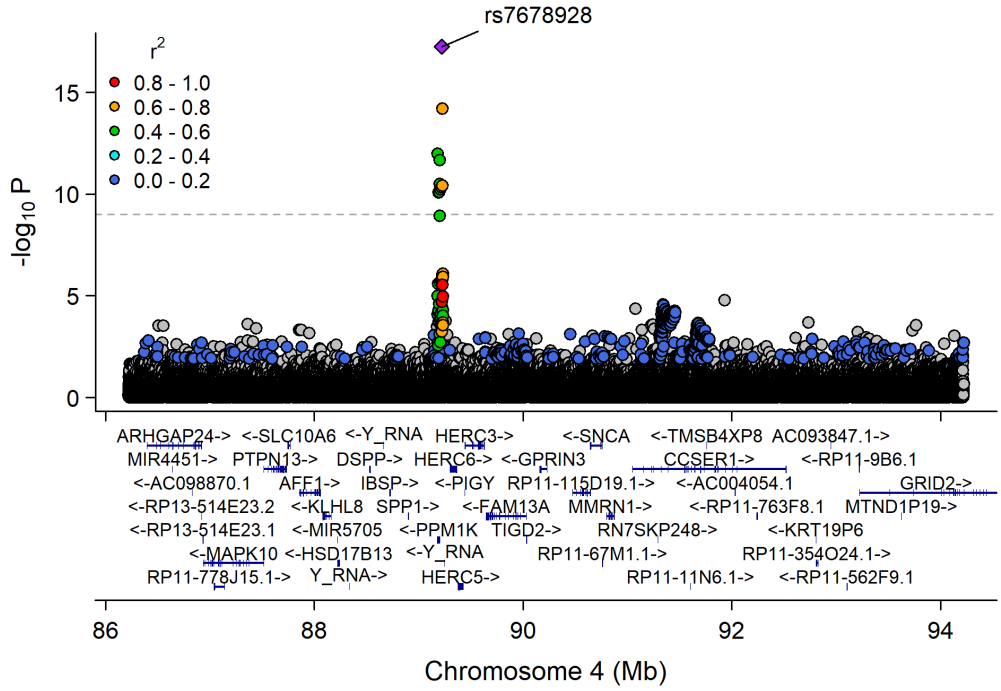
Valine (mM)


Range: 0-2.937
Mean: 1.051
Median: 1.003
Skewness: 0.896
Count
Valine (mM)
Sex
female
male
Q-Q plot of GWAS p-values for Valine
Observed $-\log_{10}(p)$
Expected $-\log_{10}(p)$
Q-Q plot of MultiXcan for Valine
Observed $-\log_{10}(p)$
Expected $-\log_{10}(p)$

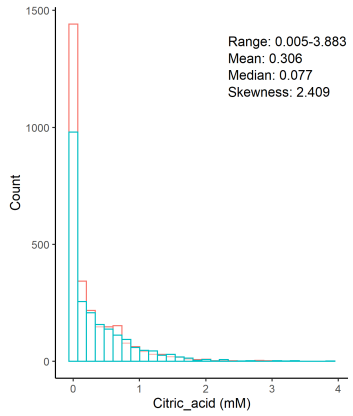

Manhattan Plot of GWAS p-values<.001 for Valine
$-\log_{10}(p)$
Chromosome

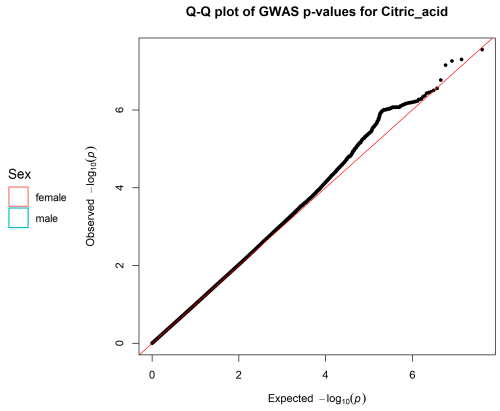

Manhattan Plot of MultiXcan for Valine
$-\log_{10}(p)$
Chromosome

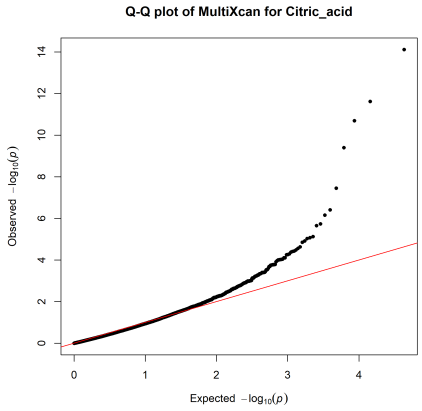
Chr4__89101913__89295163

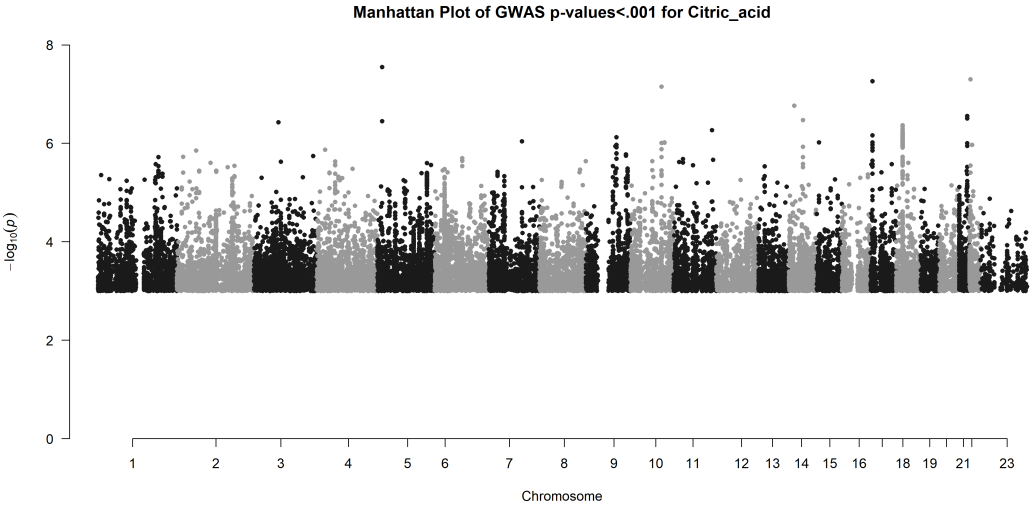


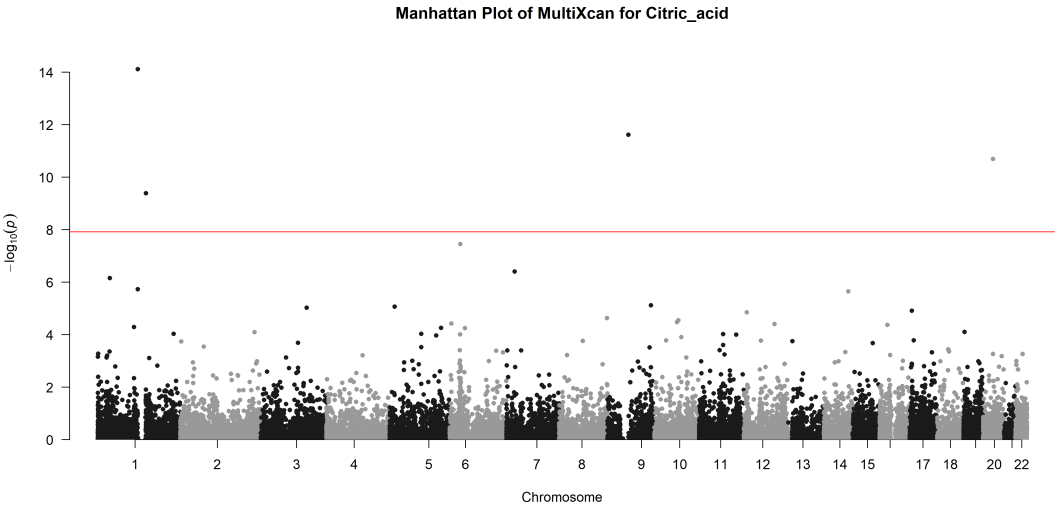
Citric acid (mM)



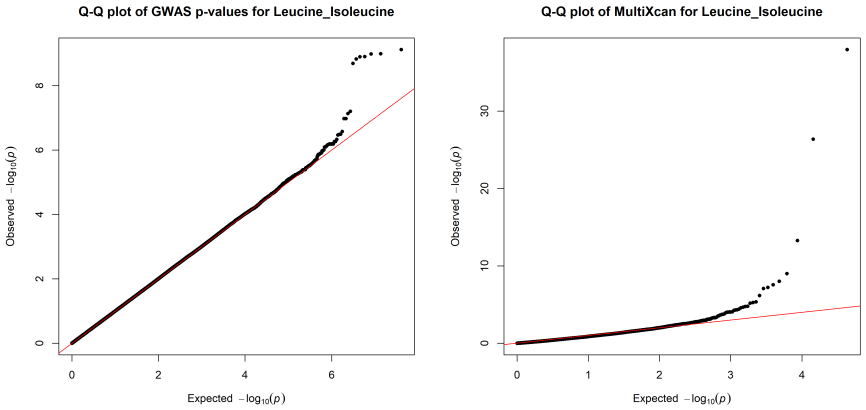


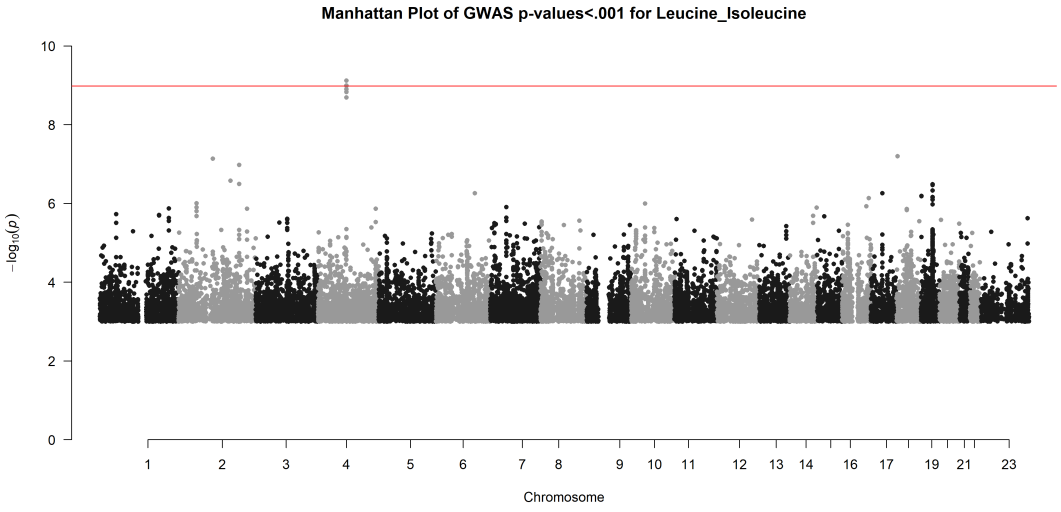


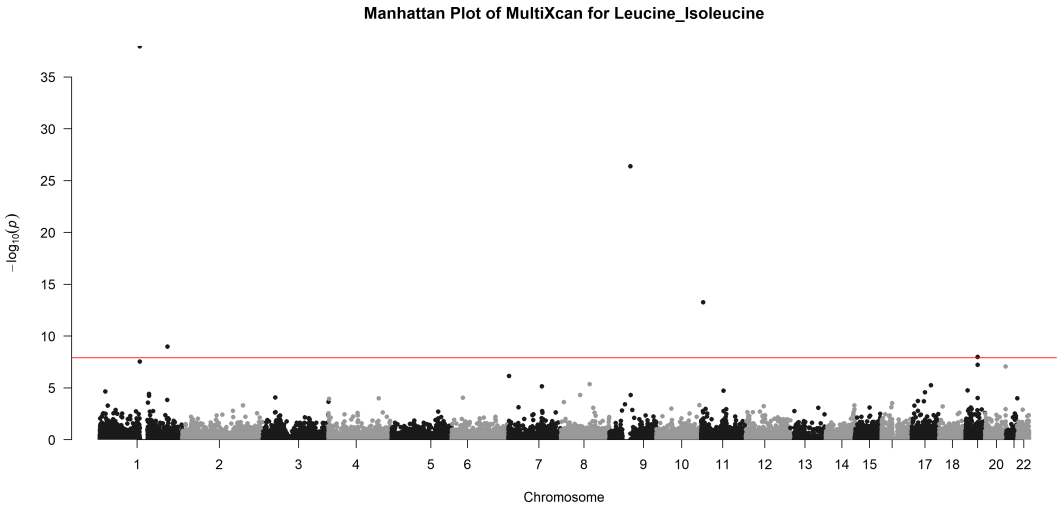




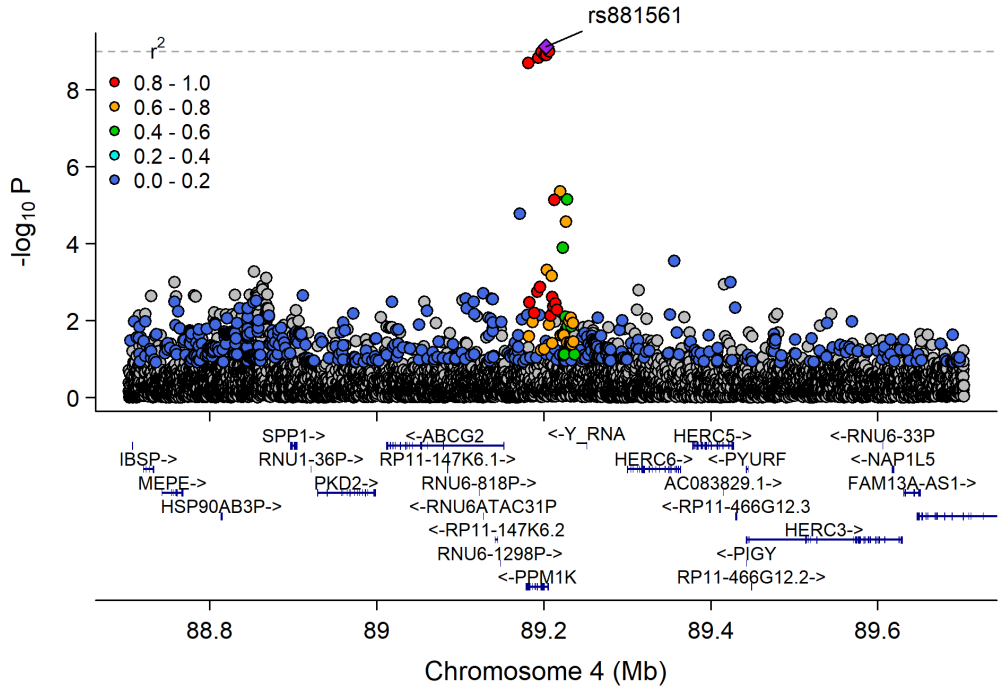
Ratio of leucine to isoleucine





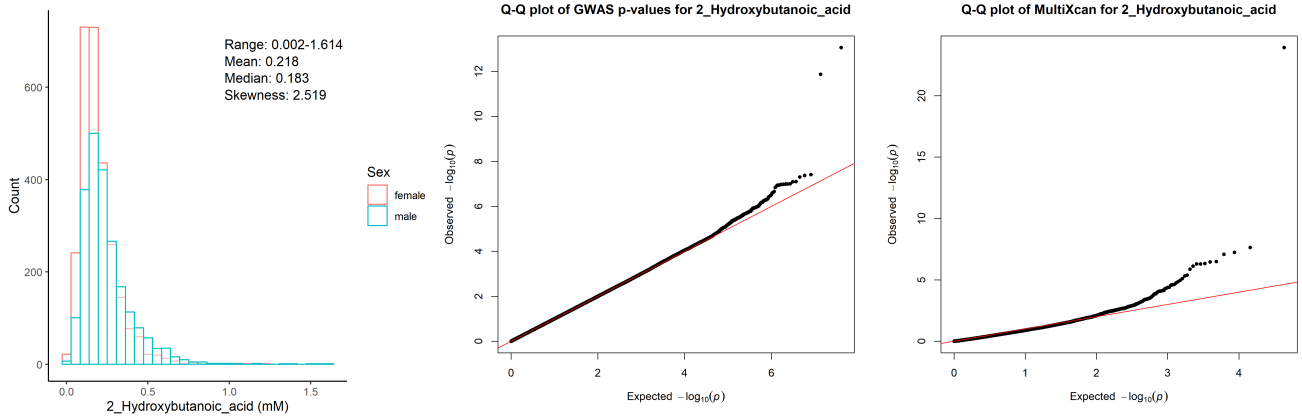


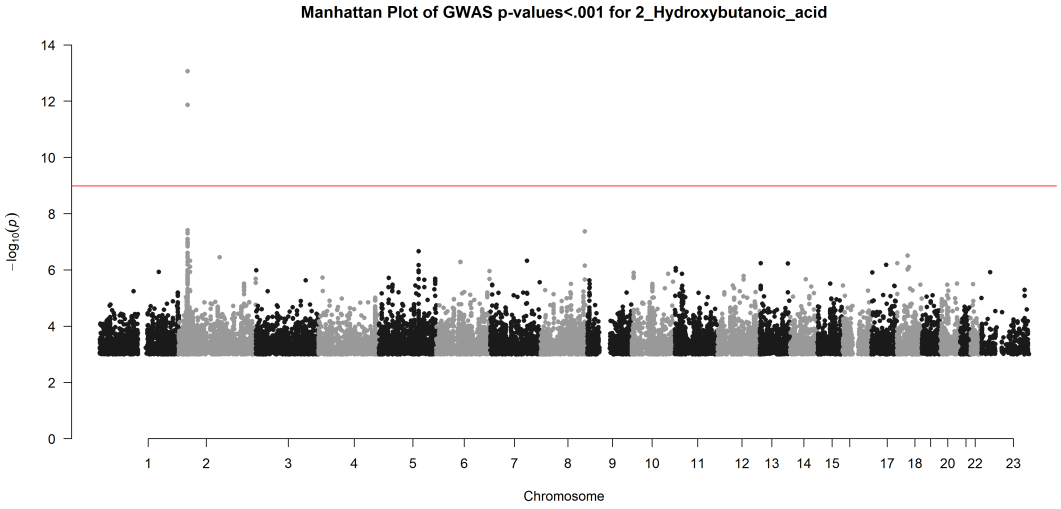
Chr4__89101913__89295163

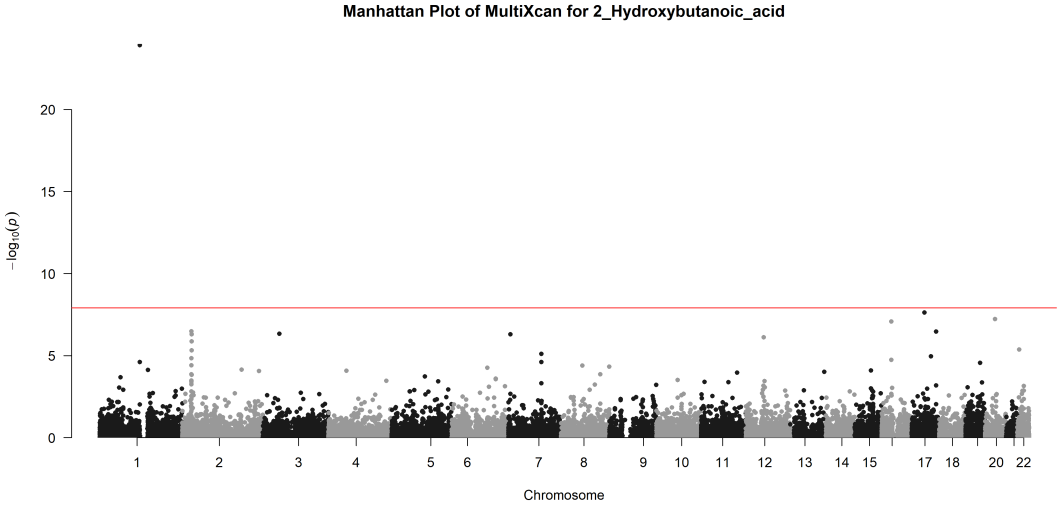


Hydroxy acids

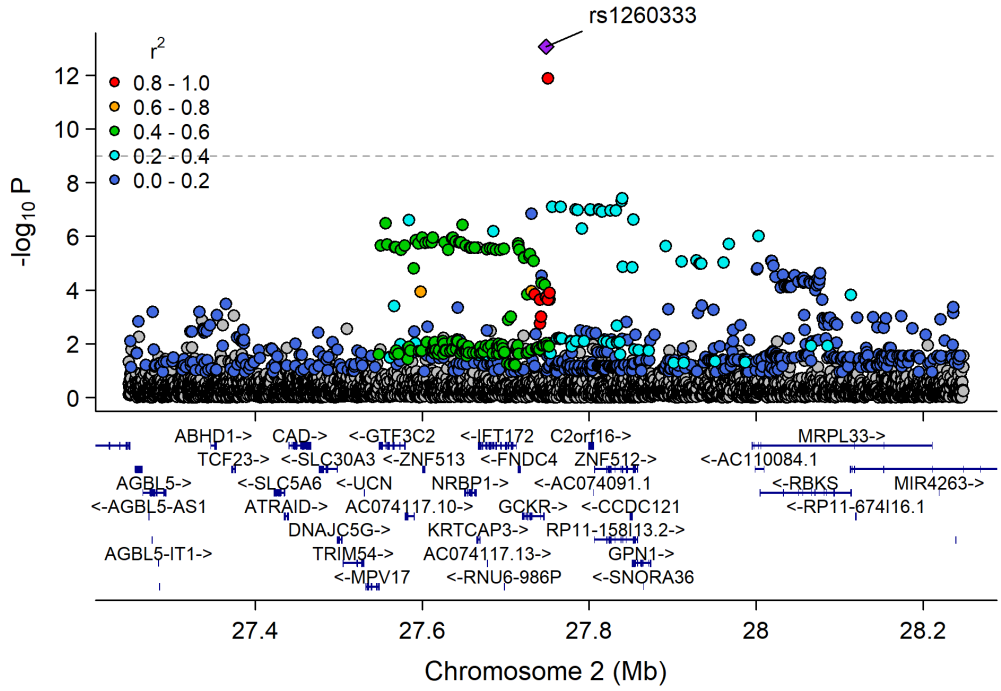
2-Hydroxybutanoic acid (mM)



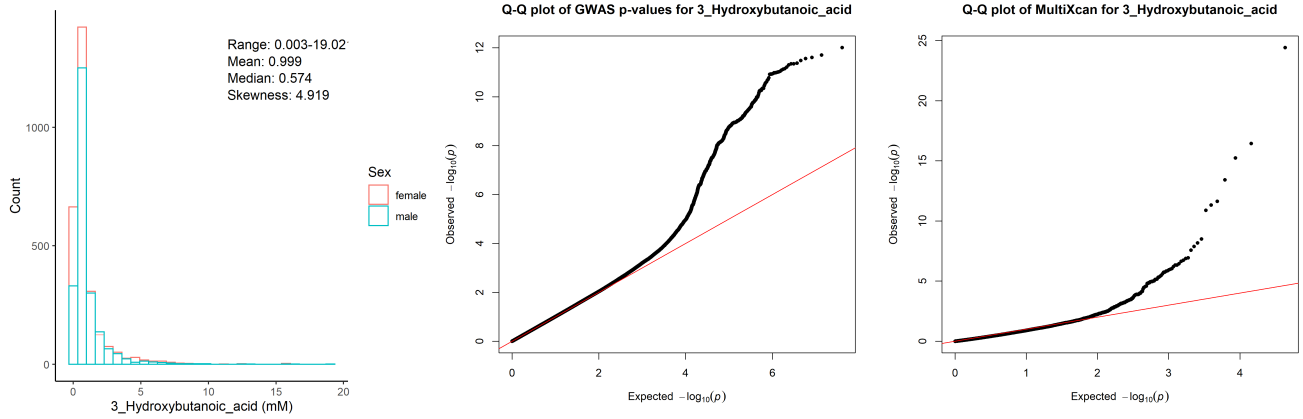


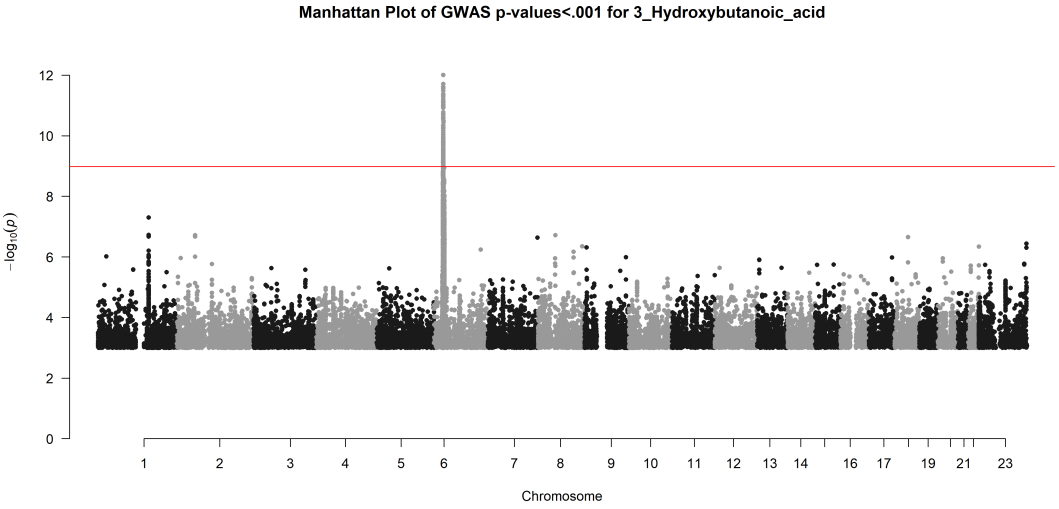


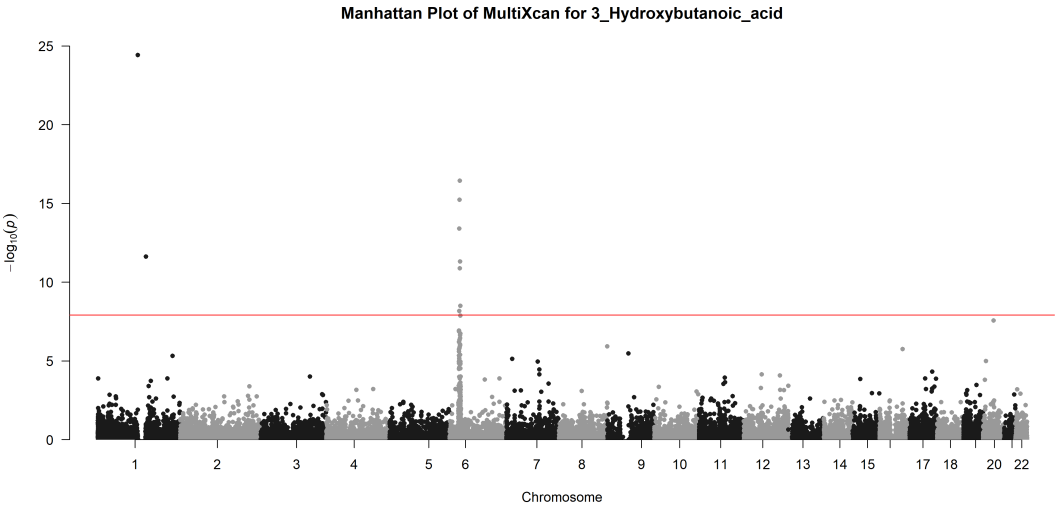
Chr2_26753815_28597624



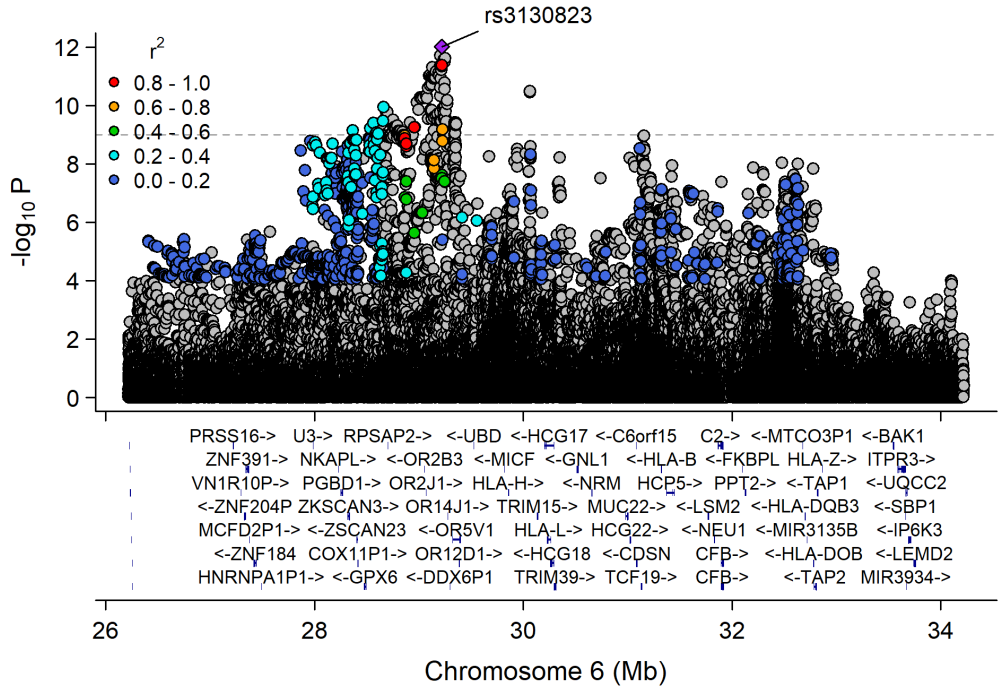
3-Hydroxybutanoic acid (mM)



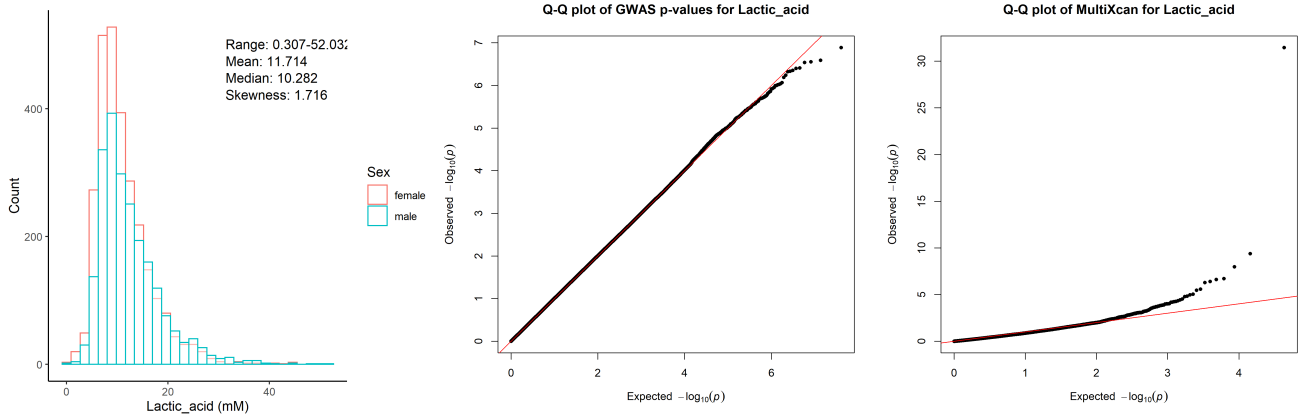


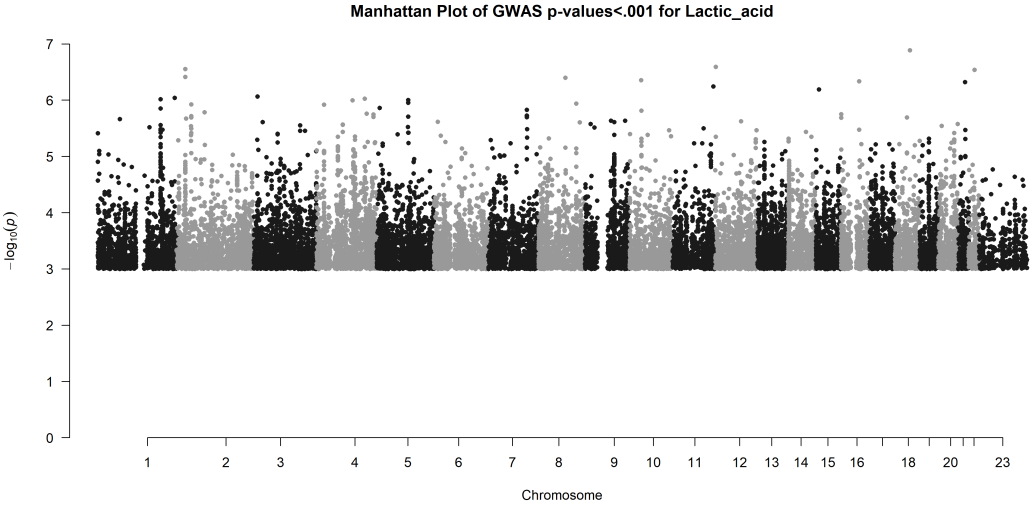


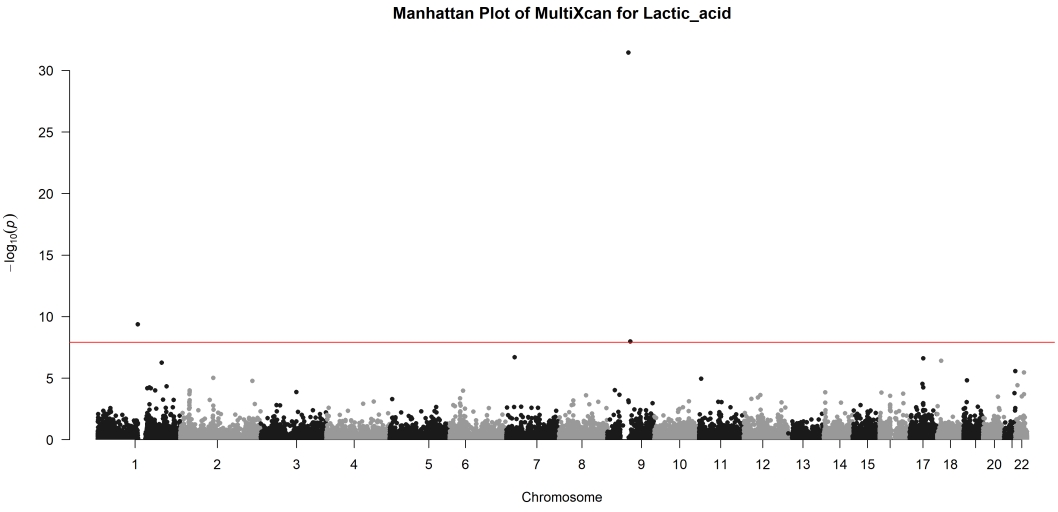
Chr6_24042708_34003583



Lactic acid (mM)

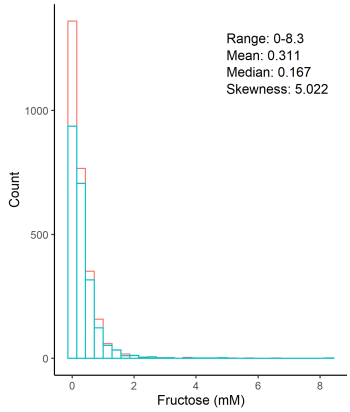


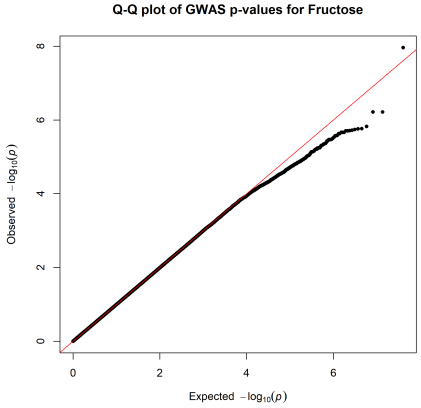


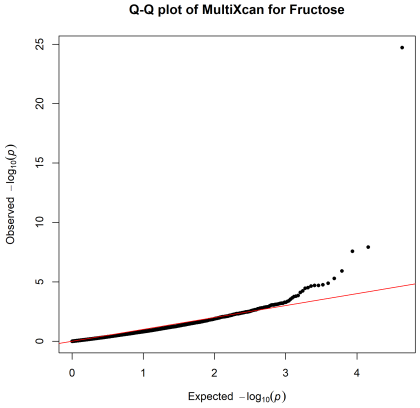


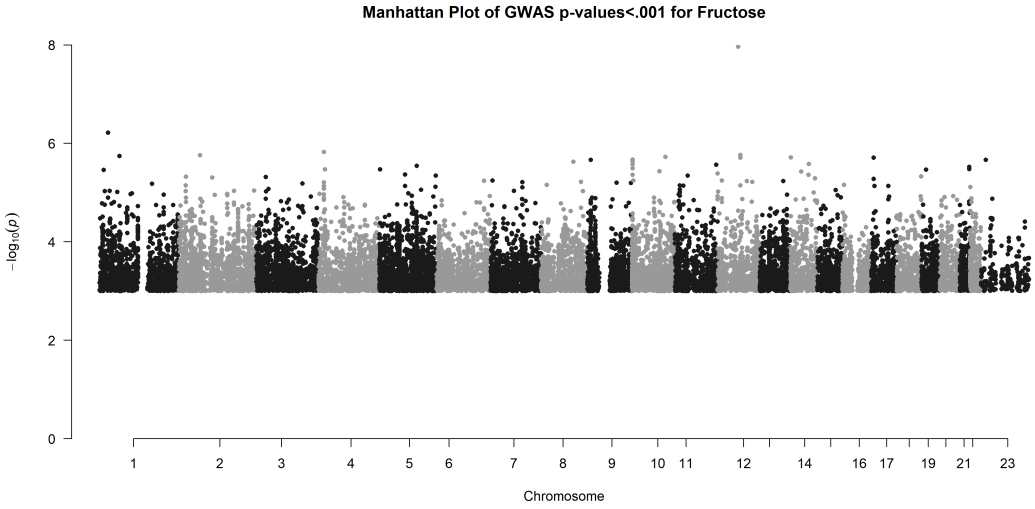
Organooxygen compounds

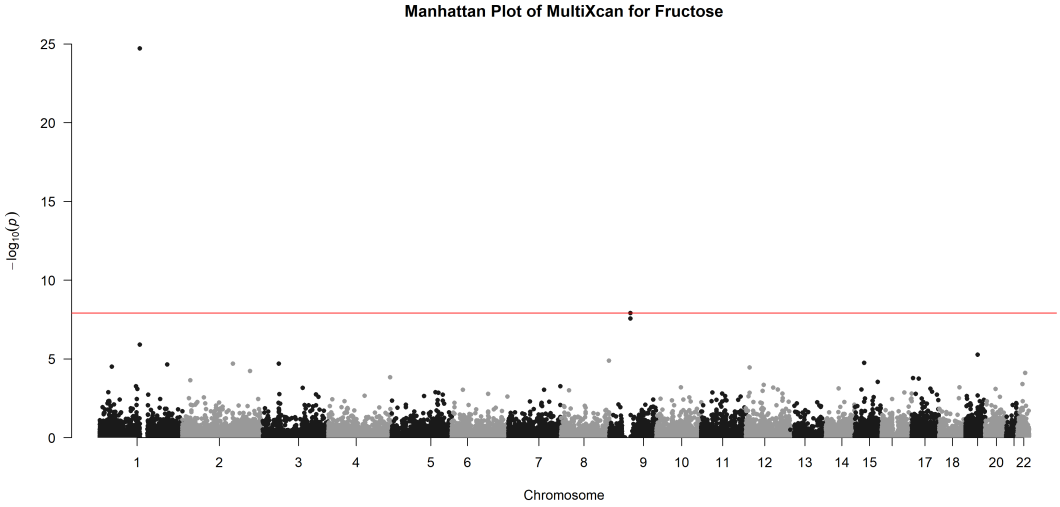
Fructose (mM)



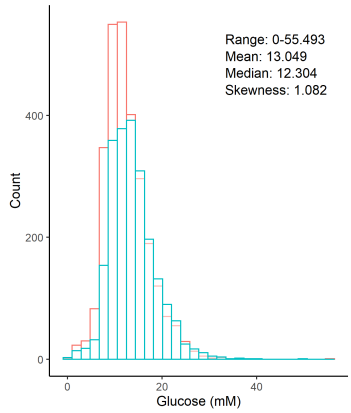


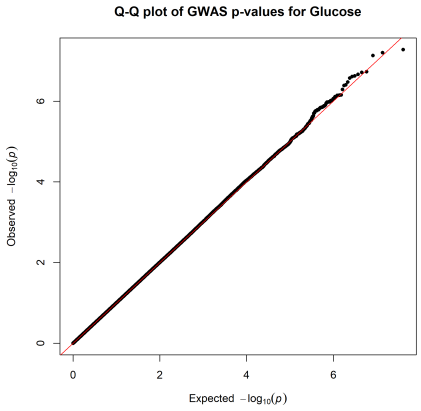




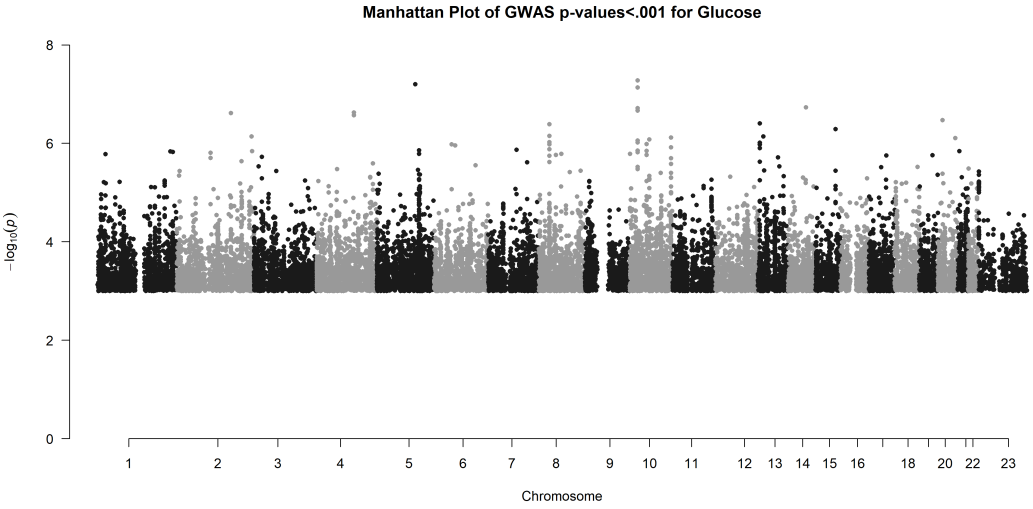


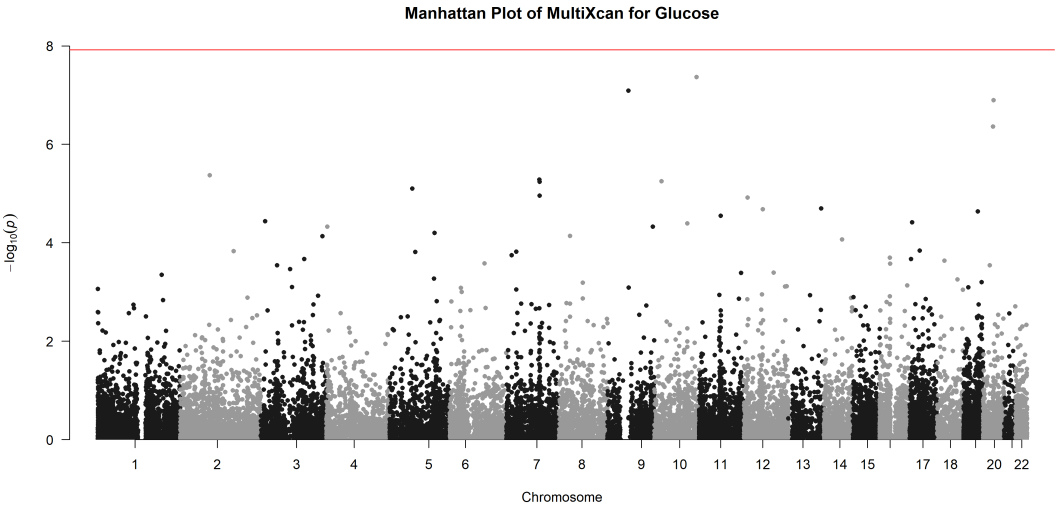
Glucose (mM)


Range: 0-55.493
Mean: 13.049
Median: 12.304
Skewness: 1.082
Count
Glucose (mM)
Sex
female
male

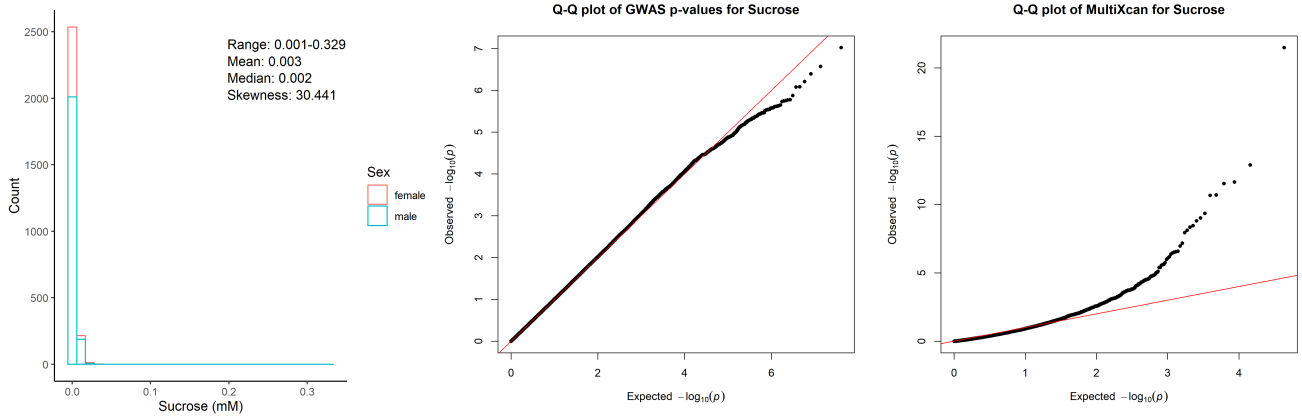

Q-Q plot of GWAS p-values for Glucose
Observed $-\log_{10}(p)$
Expected $-\log_{10}(p)$

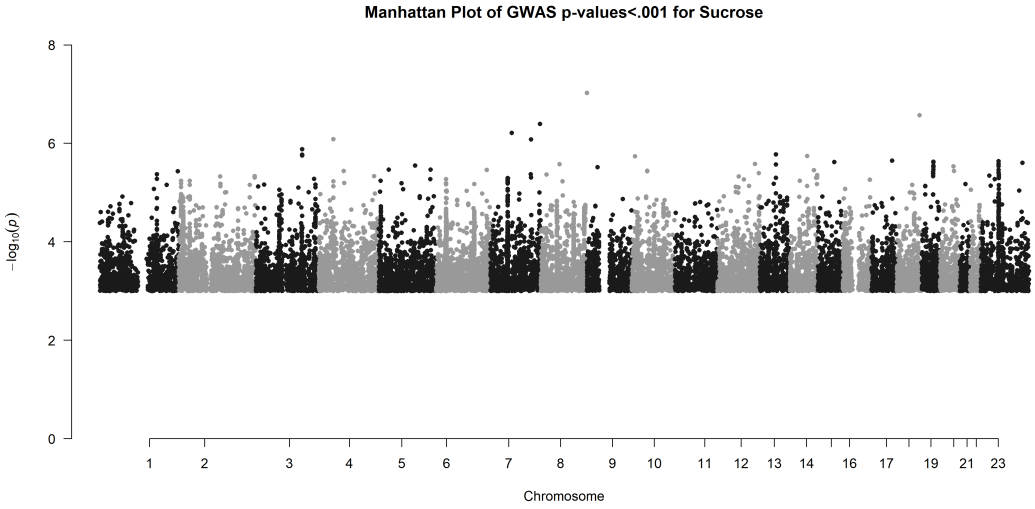

Q-Q plot of MultiXcan for Glucose
Observed $-\log_{10}(p)$
Expected $-\log_{10}(p)$

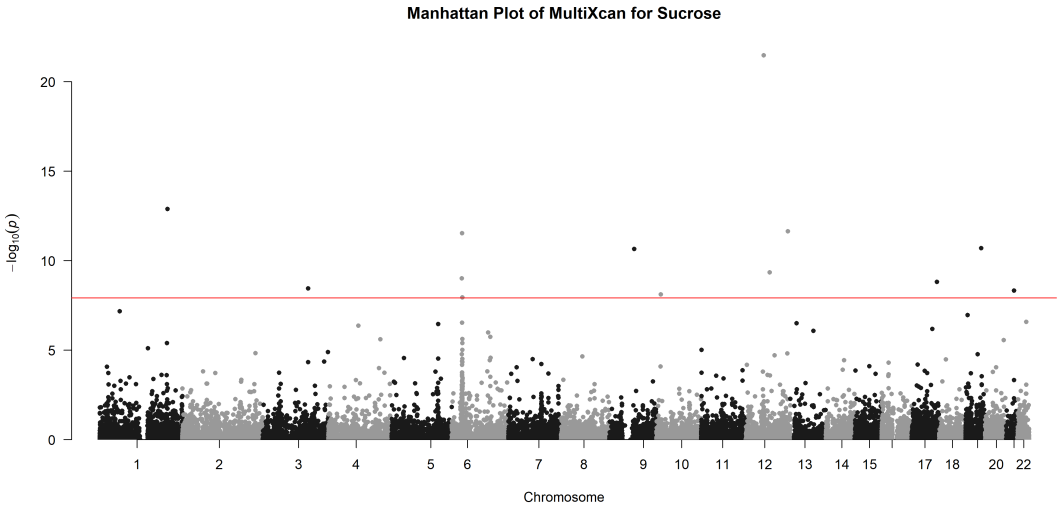

Manhattan Plot of GWAS p-values<.001 for Glucose
$-\log_{10}(p)$
Chromosome


Manhattan Plot of MultiXcan for Glucose
$-\log_{10}(p)$
Chromosome

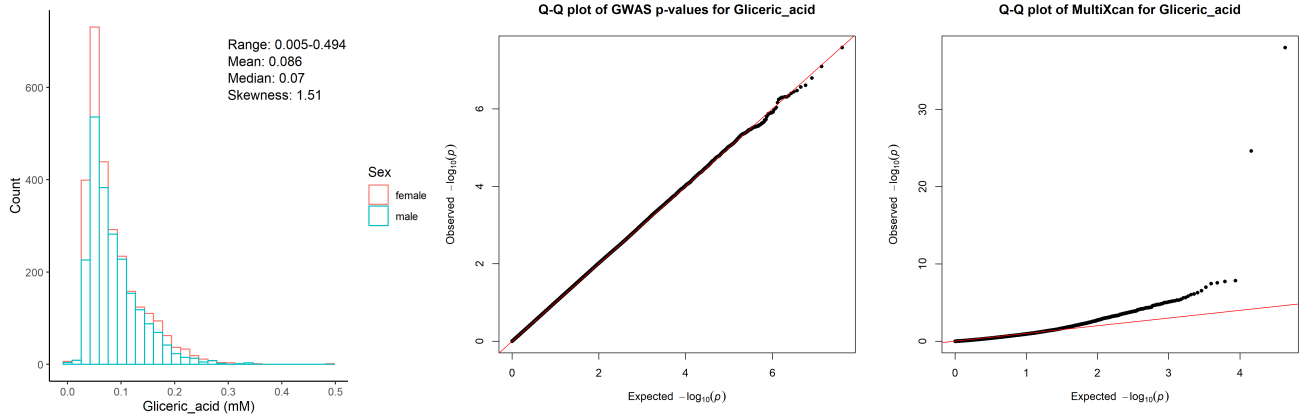
Sucrose (mM)

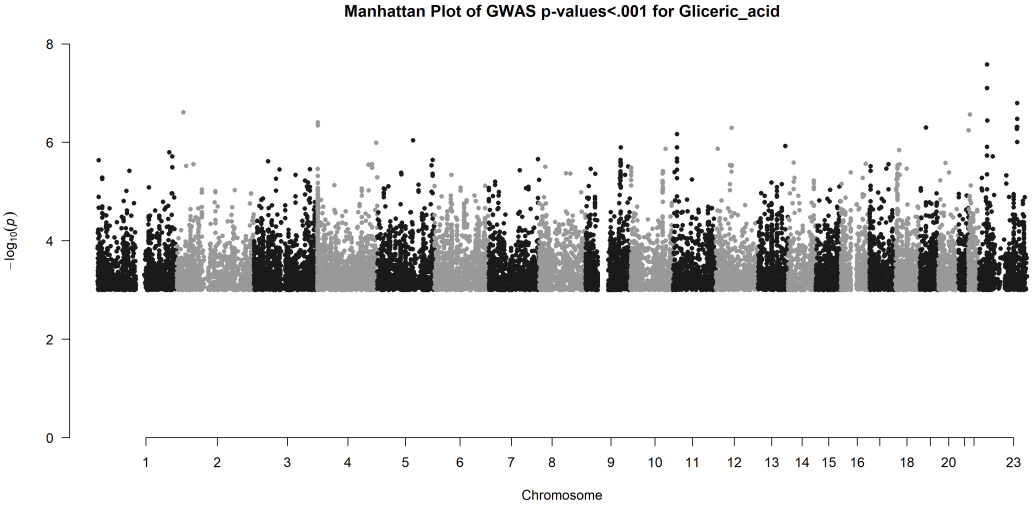

Range: 0.001-0.329
Mean: 0.003
Median: 0.002
Skewness: 30.441
Count
Sucrose (mM)
Sex
female
male
Q-Q plot of GWAS p-values for Sucrose
Observed $-\log_{10}(p)$
Expected $-\log_{10}(p)$
Q-Q plot of MultiXcan for Sucrose
Observed $-\log_{10}(p)$
Expected $-\log_{10}(p)$


Manhattan Plot of GWAS p-values<.001 for Sucrose
$-\log_{10}(p)$
Chromosome

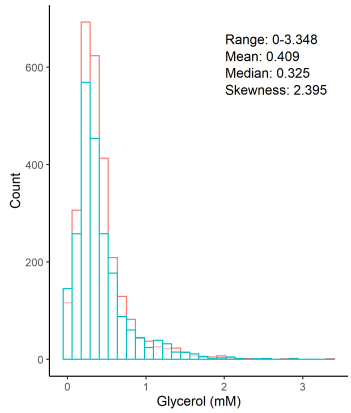

Manhattan Plot of MultiXcan for Sucrose
$-\log_{10}(p)$
Chromosome

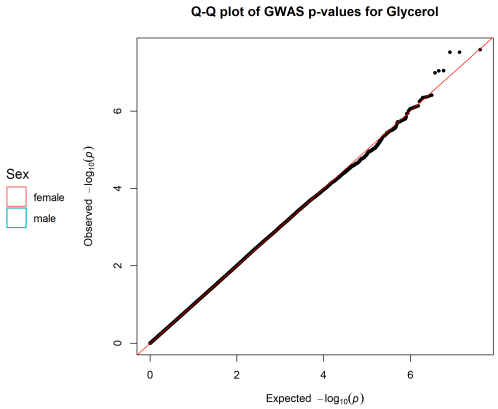
Gliceric acid (mM)



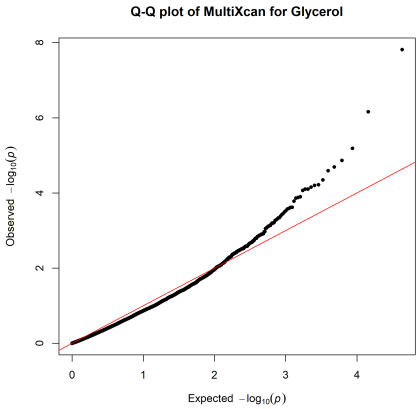


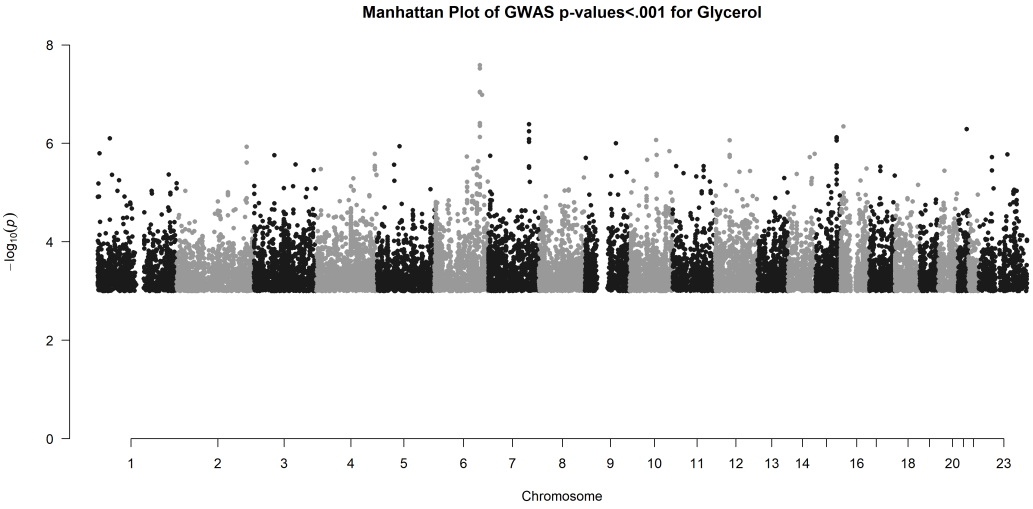
Glycerol (mM)

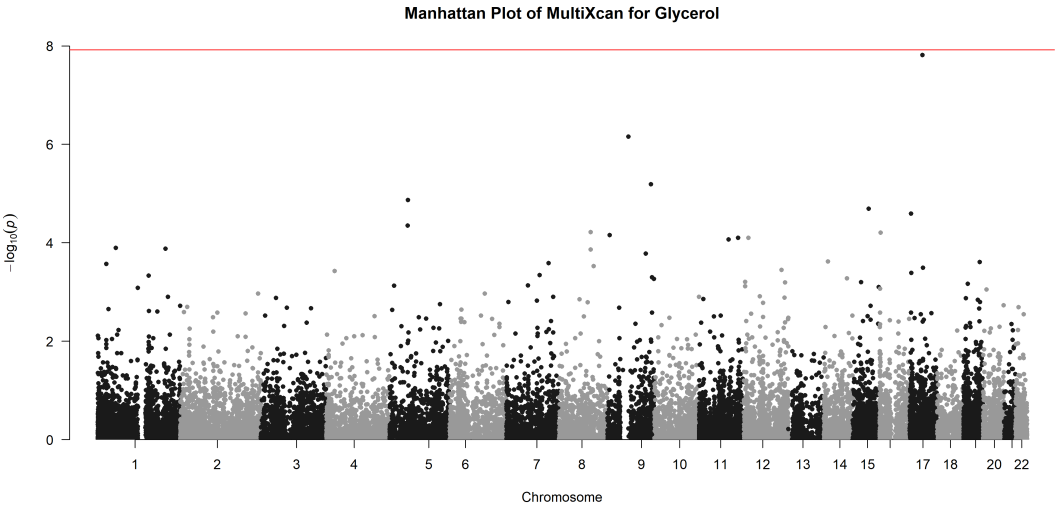

Range: 0-3.348
Mean: 0.409
Median: 0.325
Skewness: 2.395
Count
Glycerol (mM)


Sex
female
male

Q-Q plot of GWAS p-values for Glycerol
Observed $-\log_{10}(p)$
Expected $-\log_{10}(p)$

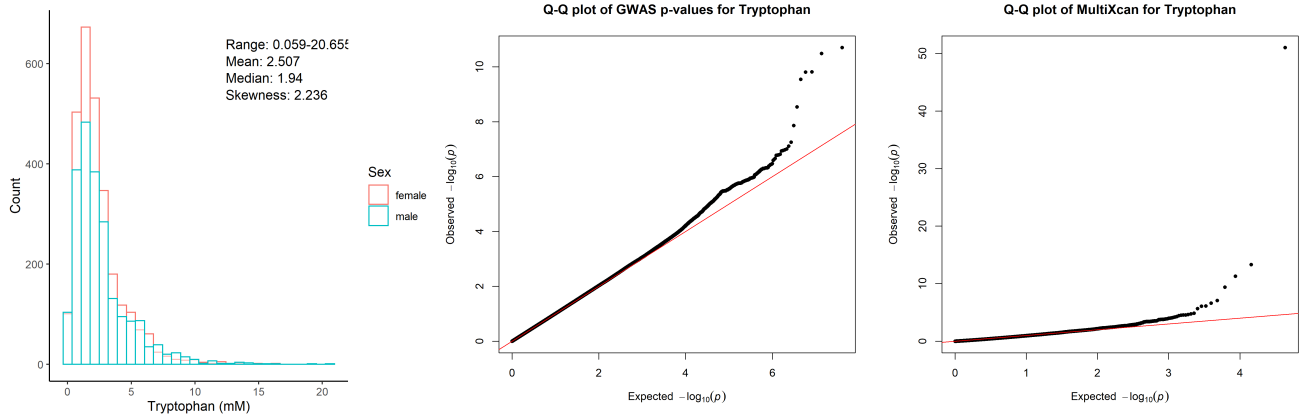

Q-Q plot of MultiXcan for Glycerol
Observed $-\log_{10}(p)$
Expected $-\log_{10}(p)$

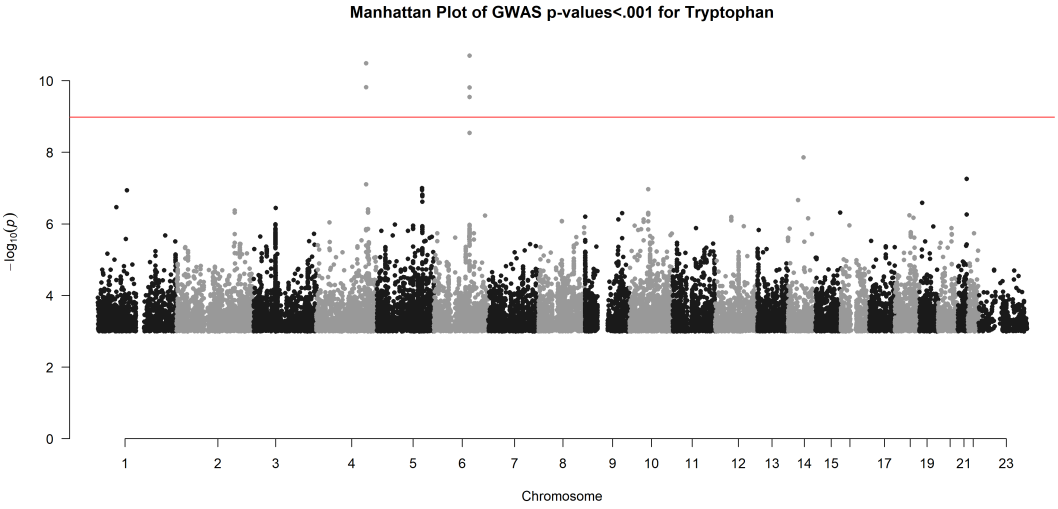

Manhattan Plot of GWAS p-values<.001 for Glycerol
$-\log_{10}(p)$
Chromosome

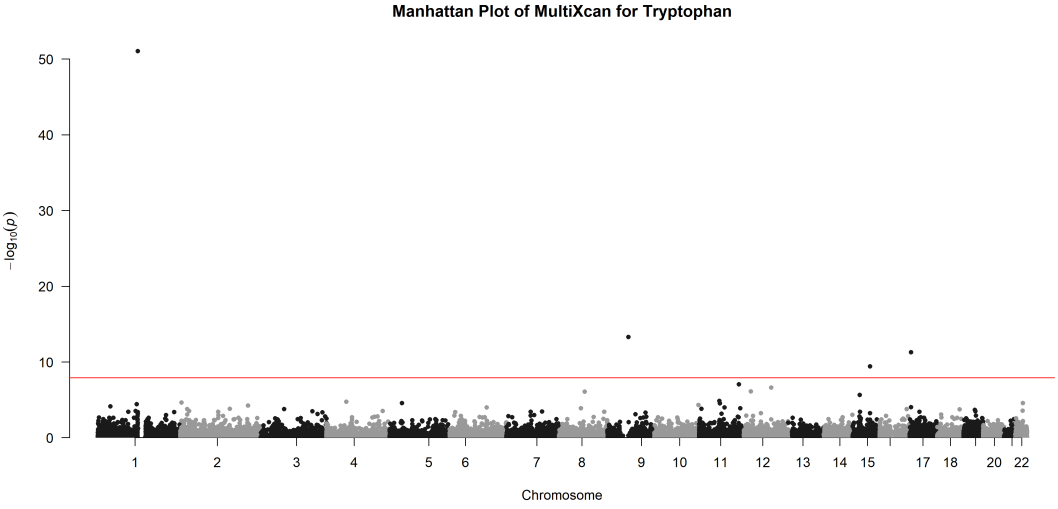

Manhattan Plot of MultiXcan for Glycerol
$-\log_{10}(p)$
Chromosome

Indoles

Tryptophan (mM)







Chr4_156167936_157083184