Supplementary Materials

Genetic landscape of chronic obstructive pulmonary disease identifies heterogeneous cell type and phenotype associations.

Contents

Supplementary Note

Supplementary Methods

Cohort descriptions and cohort-specific methods

ARIC: Atherosclerosis Risk in Communities (ARIC)[1,2] (NCT00005131), is a population based study of risk factors for atherosclerosis and its sequelae in adults from four U.S. centers. Subjects were aged 45-64 at recruitment in 1987-1989. Institutional Review Board (IRB) approval was obtained at all associated study centers and informed consent was obtained for all participants. ARIC spirometry measurements were performed with a Collins Survey II water-seal spirometer (Collins Medical, Inc.) and Pulmo-Screen II software (PDS Healthcare Products, Inc.). Genotyping was performed using the AffymetrixGeneChip SNP Array 6.0. The current analysis includes 7,224 Caucasian subjects with genotyping data, pulmonary function measures and complete covariate information. Imputation was performed using the 1000 Genomes[3] Integrated Phase 1 v3 reference panel (March 2012) in IMPUTE2[4]. Logistic regression was performed using FAST (https://bitbucket.org/baderlab/fast/wiki/Home) adjusting for age, sex, pack-years, current and ever smoking, and ancestry.

Cardiovascular Health Study (CHS): The Cardiovascular Health Study (CHS) is a population-based cohort study of risk factors for coronary heart disease and stroke in adults ≥65 years conducted across four centers[5] (NCT00005133 and NCT00149435). IRB approval was obtained at participating centers and written informed consent was obtained for all participants. 5,201 predominantly European ancestry persons were recruited in 1989-1990 from random samples of Medicare eligibility lists. An additional predominantly African-American cohort of 687 persons was subsequently enrolled in 1992-1993 for a total sample of 5,888. European ancestry participants were excluded from the GWAS study sample due to the presence at study baseline of coronary heart disease, congestive heart failure, peripheral vascular disease, valvular heart disease, stroke or transient ischemic attack. Genotyping was performed at the General Clinical Research Center's Phenotyping/Genotyping Laboratory at Cedars-Sinai among CHS participants who consented to genetic testing and had DNA available using the Illumina 370CNV BeadChip system (for European ancestry participants, in 2007) or the Illumina HumanOmni1-Quad_v1 BeadChip system (for African-American participants, in 2010). Additional genotypes were provided from the ITMAT-Broad-CARe (IBC) Illumina iSELECT chip (for European ancestry participants). Imputation was performed using 1000 Genomes[3] Phase 1 v3 haplotypes and minimac[4] (for European ancestry participants, 2012-11-16) or IMPUTE version 2.2.2 (for African-American participants). Logistic regression was performed in R, adjusting for age, sex, pack-years, current and ever smoking, CHS clinic (4 sites) and PCs 1-5.

COPDGene: Eligible subjects in COPDGene Study (NCT00608764, www.copdgene.org) were of non-Hispanic white (NHW) or African-American (AA) ancestry, aged 45-80 years old, with at least 10 pack-years of smoking and no diagnosed lung disease other than COPD or asthma[6,7 21,22]. IRB approval was obtained at all study centers, and all study participants provided written informed consent. Illumina (San Diego, CA) performed genotyping on the HumanOmniExpress array. Genotyping at the Z and S alleles was performed in all subjects. Subjects known or found to have severe alpha-1 antitrypsin deficiency were excluded. We performed imputation using MaCH[8,9 10,11] and minimac[4 4] (version 2012-10-09) and 1000 Genomes[3] Phase I v3 European (EUR) and cosmopolitan reference panels, for whites and African-Americans, respectively. We removed variants with an r^2 value of ≤ 0.3. We performed logistic regression on cases and controls defined based on pre-bronchodilator spirometry, adjusting for age, sex, pack-years, current smoking, and principal components of genetic ancestry, separately in non-Hispanic whites and African-Americans, using PLINK1.9[10] .

Evaluation of COPD Longitudinally to Identify Predictive Surrogate End-points (ECLIPSE; SCO104960, NCT00292552, www.eclipse-copd.com**):** Details of the ECLIPSE study and genome-wide association

analysis have been described previously[12]. The ECLIPSE study was approved by the relevant ethics and review boards at the participating clinical centers. All participants provided written informed consent. Cases and controls were aged 40-75 with at least a 10 pack-year smoking history without other respiratory diseases. Genotyping was performed using the Illumina HumanHap 550 V3 (Illumina, San Diego, CA). Subjects and markers with a call rate of < 95% were excluded. Imputation was performed using MaCH[8,9] and minimac[4] (version 2012-10-09) and the 1000 Genomes[3] Phase I v3 European (EUR) reference panel. Logistic regression was performed on cases and controls defined by pre-bronchodilator spirometry, adjusting for age, sex, pack-years, current smoking, and principal components of genetic ancestry using PLINK1.9[10].

Expression Quantitative Trait Loci – Lung (eQTL): The lung eQTL study has been described previously[13]. Briefly, patients who underwent thoracic surgery were recruited at three academic sites: Laval University, University of British Columbia (UBC), and University of Groningen, henceforth referred to as Laval, UBC, and Groningen, respectively. Patients from Laval were those undergoing lung cancer surgery, the majority of the UBC patients had lung resection for small peripheral lung lesions with some samples derived from autopsy or at the time of lung transplantation. At Groningen, patients were recruited from those having surgery for various lung diseases, including patients that underwent therapeutic resection for lung tumors and lung transplantation. All patients provided written informed consent, and the study was approved by the ethics committees of the Institut universitaire de cardiologie et de pneumologie de Québec and the UBC-Providence Health Care Research Institute Ethics Board for Laval and UBC, respectively. The study protocol was consistent with the Research Code of the University Medical Center Groningen and Dutch national ethical and professional guidelines. Patients whose lung function could have been influenced by lung diseases other than COPD and lung cancer were excluded. This includes patients with severe alpha-1 antitrypsin deficiency (n=11), amyloidosis (n=1), bronchiectasis (n=3), bronchiolitis obliterans (n=2), bronchopulmonary dysplasia (n=2), cystic fibrosis (n=14), idiopathic pulmonary fibrosis (n=13), langerhans cell histiocytosis (n=1), lymphangioleiomyomatosis (n=1), primary pulmonary hypertension (n=4), sarcoidosis (n=3) and vascular malformation (n=1).
Genotyping was carried out using the Illumina Human1M-Duo BeadChip. Standard genotyping quality controls were performed independently in the Laval, UBC and Groningen cohorts. Genotypes were then imputed with the Michigan Imputation Server[14] using the Haplotype Reference Consortium version 1 (HRC.r1-1) data as reference set. Variants with an r^2 value of ≤ 0.3 were removed from further analysis. Single-marker association tests were performed with PLINK v1.90[10,11] adjusting for age, sex, smoking status, pack-years, ever smoking status, clinical center and genetic ancestry PC1 to PC10.

Framingham Heart Study (FHS; NCT00005121**):** Details on pulmonary function in the FHS have been previously published[15,16]. FHS was IRB-approved at the relevant institutions, and all participants provided written informed consent. We analyzed data from the most recent exam for each of the three generations of families participating in the FHS were analyzed. Genotypes were from the Affymetrix 500K array supplemented by the Affymetrix MIPS 50K. From a total number of 549,781 genotyped SNPs, 412,053 were used with MaCH[8,9] for haplotype phasing, of which of 137,728 genotyped SNPs were removed by quality control. MaCH/minimac[4,8,9] were used in this genotype imputation process to impute the FHS sample using the November 2010 release of the 1000 Genomes[3] multi-ethnic panel. We used GEE implemented in the R package geepack with independent correlation matrix and clustering based on family, adjusted for sex, age, smoking status, pack years and genetic ancestry principal component 1 (to adjust for population stratification).

KARE: Details on the Korean Association Resource project (KARE) have been previously published[17,18]. KARE was initiated in 2007 to undertake genome-wide analyses among 10,038 participants in the rural-based Ansung and city-based Ansan South Korean cohorts. The study was approved at appropriate IRBs from participating institutions, and participants provided informed consent. KARE3 data were obtained from the third phenotype collection in 2008; lung function was collecting using the Vmax-2130 (Sensor

Medics, Yorba Linda, CA, USA). Genotyping was performed using the Affymetrix Genome-Wide Human array 5.0 (Affymetrix, Inc., Santa Clara, CA, USA). We performed genotype imputation using IMPUTE2 and the 1000 Genomes[3] Phase 3 cosmopolitan panel. Markers were converted to genotype from dosage with call rate ≥ 95%, minor allele frequency ≥ 1%, p for HWE ≥ $1.0x10^{-5}$, imputation quality score ≥ 0.9. Logistic regression, adjusting for age, sex, pack-years, principal components, current and ever smoking, was performed using PLINK[19].

LifeLines: The LifeLines Cohort Study is a population-based cohort study established as a resource for research on complex interactions between environmental, phenotypic and genomic factors in the development of chronic diseases and healthy aging[20–22]. Between 2006 and 2013, inhabitants of the northern part of The Netherlands and their families were invited to participate, thereby contributing to a three-generation design. Participants visited one of the LifeLines research sites for pre-bronchodilator spirometry following ATS guidelines. All participants signed an informed consent form before they received an invitation for the physical examination. The LifeLines Cohort Study is conducted according to the principles of the Declaration of Helsinki and in accordance with research code University Medical Center Groningen (UMCG), The Netherlands. The LifeLines study is approved by the medical ethical committee of the UMCG. Blood samples for a subset of individuals were genotyped using the Illumina CytoSNP-12v2 array. Independent Caucasian-ancestry samples (n = 13,436) have been imputed using the 1000 Genomes[3] phase1 v3 reference panels. Genotypes were pre-phased using SHAPEIT2[23] and aligned to the reference panels using Genotype Harmonizer (www.molgenis.org/systemsgenetics) in order to resolve strand issues. The samples were imputed using minimac[4] (version 2012-10-09), yielding 28,681,763 SNPs. Associations between genomic dosages with moderate/severe COPD were assessed with logistic regression models adjusted for age, smoking status (never/ever), current smoking (no/yes), pack years smoked and sex. All analysis were performed in software package PLINK version 1.07[19,24].

Lovelace: The Lovelace Smokers Cohort (LSC) has been actively enrolling smokers from the Albuquerque, NM metropolitan area since 2001[25]. All participants provided written informed consent, and the study was approved by the relevant IRB. Enrollment was restricted to current and former smokers age 40 to 74 years old with a minimum of 10 pack-years of smoking and no personal history of lung cancer. A detailed questionnaire written in English was used to collect information on demographics; medical, cigarette smoking, and exposure history; socioeconomic status; diet; and quality of life. Pulmonary function testing was performed at each visit. All participants signed a consent form, and the Western Institutional Review Board approved this project. The GWAS discovery set was comprised of 1200 Caucasian (self-reported) smokers. The HumanOmni2.5-4v1-H BeadChip (Illumina, San Diego, CA) was used to genotype 2,450,000 SNPs in 1200 Caucasian smokers from the LSC. After quality assessment, 1163 subjects with 1,599,980 SNPs remained in the genetic association analysis. Logistic regression was performed using PLINK on white subjects, adjusting for age, sex, pack-years, current and ever smoking, and ancestry.

Multi-Ethnic Study of Atherosclerosis (MESA): MESA is a longitudinal study of subclinical cardiovascular disease and risk factors that predict progression to clinically overt cardiovascular disease or progression of the subclinical disease[26]. Between 2000 and 2002, MESA recruited 6,814 men and women 45 to 84 years of age. Exclusion criteria were clinical cardiovascular disease, weight exceeding 136 kg (300 lb.), pregnancy, and impediment to long-term participation. The MESA Family Study recruited 1,595 African American and Hispanic participants, generally siblings of MESA participants, using the same inclusion and exclusion criteria as MESA except that clinical cardiovascular disease was permitted. The MESA Air Pollution Study recruited an additional 257 participants from Los Angeles and Riverside County, CA, and Rockland County, NY, using the same criteria as MESA, except that participants were ages 50 to 89 who lived in the area more than 50% of the year and had no plans to move in the next five years[27]. The MESA Lung Study performed spirometry following the 2005 ATS/ERS guidelines in a subset of the MESA and MESA Family Studies and all of the new recruits in the MESA Air Pollution Study[28]. All participants

provided informed consent and the protocols of MESA were approved by the IRBs of collaborating institutions and the National Heart, Lung and Blood Institute.

Participants who consented to genetic analyses were genotyped in 2009 using the Affymetrix Human SNP array 6.0 [29]. The cleaned genotypic data was deposited with MESA phenotypic data into dbGaP as the MESA SHARe project (study accession phs000209); 8,224 consenting individuals (2,685 White, 2,588 non-Hispanic African-American, 2,174 Hispanic, 777 Chinese) were included, with 897,981 SNPs passing study specific quality control (QC). For GWAS, IMPUTE version 2.2.2 was used to perform imputation for the MESA SHARe participants using the cosmopolitan 1,000 Genomes[3] Phase 1 v3 March 2012 reference set. Logistic regression was performed using SNPTEST v2.4.0[30], adjusting for age, sex, pack-years, current and ever smoking, and ancestry.

National Emphysema Treatment Trial (NETT; NCT00000606**,** www.nhlbi.nih.gov/health/prof/lung/nett/**) and Normative Aging Study (NAS)**: [31]. NETT was a multicenter clinical trial to evaluate lung volume reduction surgery. All participants provided written informed consent, and the study was approved by the IRB at all participating institutions. Enrolled subjects had severe airflow obstruction by post-bronchodilator spirometry (FEV_1 < 45% predicted) and evidence of emphysema on computed tomography (CT) chest imaging; exclusion criteria included, but not limited to, history of recurrent infections with significant sputum production or bronchiectasis. A subset of 382 self-reported white subjects without severe alpha-1 antitrypsin deficiency were enrolled in the NETT Genetics Ancillary Study. The Normative Aging Study is a longitudinal study of healthy men established in 1963 aged 21 to 80 years from the greater Boston area, free of known chronic medical conditions. The study was conducted by the Veterans Administration (VA)[32], and the local Institutional Review Board approved the study. Controls were of self-reported white ancestry and at least 10 pack-years of cigarette smoking with no evidence of airflow limitaiton on spirometry on their most recent visit. Genotyping for NETT-NAS was performed using the Illumina Quad 610 array (Illumina, San Diego, CA)[33,34]. Imputation was performed using MaCH[8,9] and minimac[4] (version 2012-10-09) and the 1000 Genomes Phase I v3 European (EUR) reference panel. Logistic regression was performed on cases and controls based on pre-bronchodilator spirometry, adjusting for age, sex, pack-years, current smoking, and principal components of ancestry using PLINK1.9[10,11].

GenKOLS (Norway): The Norwegian GenKOLS (Genetics of Chronic Obstructive Lung Disease, GSK code RES11080)[35] recruited subjects with > 2.5 pack years of smoking history from Bergen, Norway. Subjects with severe alpha-1 antitrypsin deficiency and other lung diseases (aside from asthma) were excluded. The Regional Committee for Medical Research Ethics (REK Vest), the Norwegian Data Inspectorate and the Norwegian Department of Health approved the case–control study. Written informed consent was obtained from all participants. Genotyping was performed using Illumina HumanHap 550 arrays (Illumina, San Diego, CA). Genotype imputation was performed using MaCH[8,9] and minimac[4] (version 2012-10-09) and the 1000 Genomes[3] Phase I v3 European (EUR) reference panel. Logistic regression was performed on cases and controls defined by pre-bronchodilator spirometry, adjusting for age, sex, pack-years, current smoking, and principal components of genetic ancestry using PLINK1.9[10,11].

The Rotterdam Study: The Rotterdam Study is a prospective population-based cohort study founded in 1990 in a suburb of Rotterdam, the Netherlands[36,37]. The first cohort (RS-I) consists of 7,983 participants, aged 55 years and over. The second cohort (RS-II) was recruited in 2000 with the same inclusion criteria. The third cohort (RS-III) consists of 3,932 participants, aged 45 years and over and was recruited in 2006. The Rotterdam Study was approved by the institutional review board (Medical Ethics Committee) of the Erasmus Medical Center and by the review board of The Netherlands Ministry of Health, Welfare and Sports. All participants provided written informed consent. Spirometry was performed using the Master Screen® PFT Pro (CareFusion, San Diego, CA). A total of 6,291 subjects for RS I, 2,157 for RS II and 3,048 for RS III passed genotyping quality control. Regression coefficients and their standard errors were

determined using the ProbABEL[38] program according to an additive model, adjusting for age, sex, pack-years, smoking status, and ancestry.

Subpopulations and intermediate outcome measures in COPD study (SPIROMICS; NCT01969344**):** Participants of the NHLBI SPIROMICS study were 40-80 years of age at baseline with a smoking history ≥20 pack-years. Recruitment included non-smokers, smokers without COPD, mild-moderate COPD, and severe COPD[39]. All participants provided written informed consent and the Institutional Review Boards/Ethics Committees of all the cooperating institutions approved the study protocols. Genome-wide genotyping was performed using the Illumina OmniExpress HumanExome BeadChip using standard techniques in the first 571 subjects with COPD and 175 smoking controls. Imputation was performed against 1000 Genomes reference panels using Impute-v2.30 using a quality cutoff of 0.9, and association analysis performed using PLINK adjusting for age, sex, pack-years, smoking status, and ancestry.

Studies with Custom Genotyping:

Boston Early-Onset COPD Study (BEOCOPD; ClinicalTrials.gov: NCT01177618). BEOCOPD is an extended pedigree study constructed based on probands under 53 years of age with severe COPD (defined as pre-bronchodilator forced expiratory volume in one second (FEV_1) < 40% predicted) and without severe alpha-1 antitrypsin deficiency[40,41]. IRB approval was obtained for the study, and all participants provided written informed consent.

International COPD Genetics Network (ICGN): ICGN recruited subjects (FEV_1 < 60% predicted and FEV_1/FVC < 90% predicted between ages 45-65) as probands and then enrolled available siblings and parents of the proband[42,43]. The study was IRB approved at all relevant institutions and participants provided written informed consent.

Transcontinental COPD Genetics Study (TCGS) – Korea and Poland: TCGS[44] and comprised of two case-control studies, based in Poland and in Korea. The study was approved by the appropriate IRBs, and all participants provided written informed consent. Both studies recruited individuals between 40 and 80 years of age, with at least 10 pack-years of cigarette smoking; where cases had severe COPD (FEV1 < 50% predicted) and controls had normal spirometry. Subjects with other lung diseases were excluded. TCGS-Poland enrolled white individuals, and TCGS-Korea enrolled Korean individuals.

Genotyping in BECOPD, ICGN, and TCGS:

We genotyped subjects using the Illumina HumanExome v1.2 array with custom content, including top results from prior genome-wide association studies, as previously described[45]. We performed single-variant association analysis adjusting for age, pack-years, sex, and current and ever smoking, together using a covariate additionally indicating study, via a logistic mixed model as implemented in GMMAT version 0.5 in R (version 3.2.0, http://www.R-project.org/)[46] in whites in our family-based studies, and using standard logistic regressions in each of the two TCGS studies.

Ethical statements for ICGC studies

ICGC studies obtained IRB or other relevant ethical body approval as follows:

ARIC: IRB approval obtained from: Wake Forest Baptist Medical Center, Winston-Salem, NC; University of Mississippi Medical Center, Jackson, MS; University of Minnesota, Minneapolis, MN; Johns Hopkins University, Baltimore, MD

CHS: IRB approval obtained from: Sacramento County, Sacramento, CA - University of California, Davis; Washington County, Hagerstown, MD - Johns Hopkins University; Forsyth County, Winston-Salem, NC - Wake Forest University School of Medicine; Pittsburgh, PA - University of Pittsburgh

COPDGene: IRB approval obtained from: Ann Arbor VA; Baylor College of Medicine, Houston, TX; Brigham and Women's Hospital, Boston, MA; Columbia University, New York, NY; Duke University

Medical Center, Durham, NC; Fallon Clinic, Worcester, MA; Health Partners Research Foundation, Minneapolis, MN; Johns Hopkins University, Baltimore, MD; Los Angeles Biomedical Research Institute at Harbor UCLA Medical Center, Los Angeles, CA; Michael E. DeBakey VAMC, Houston, TX; Minneapolis VA; Morehouse School of Medicine, Atlanta, GA; National Jewish Health, Denver; Temple University, Philadelphia, PA; University of Alabama, Birmingham, AL; University of California, San Diego, CA; University of Iowa, Iowa City; University of Michigan, Ann Arbor, MI; University of Minnesota, Minneapolis, MN; University of Pittsburgh, Pittsburgh, PA; University of Texas Health Science Center at San Antonio, San Antonio, TX

ECLIPSE: The ECLIPSE study was conducted in accordance with the Declaration of Helsinki and good clinical practice guidelines, and was approved by the relevant ethics and review boards at the participating centers. All participants provided written informed consent.

eQTL: All patients provided written informed consent and the study was approved by the ethics committees of the Institut universitaire de cardiologie et de pneumologie de Québec and the UBC-Providence Health Care Research Institute Ethics Board for Laval and UBC, respectively. The study protocol was consistent with the Research Code of the University Medical Center Groningen and Dutch national ethical and professional guidelines.

FHS: IRB approval obtained from: Winston-Salem, NC--University of North Carolina; Minneapolis MN—University of Minnesota; Framingham, MA—Boston University; Salt Lake City, UT—University of Utah

KARE: The study was approved at appropriate IRBs from participating institutions and participants provided informed consent.

LifeLines: All participants signed an informed consent form before they received an invitation for the physical examination. The LifeLines Cohort Study is conducted according to the principles of the Declaration of Helsinki and in accordance with research code University Medical Center Groningen (UMCG), The Netherlands. The LifeLines study is approved by the medical ethical committee of the UMCG.

Lovelace: All participants signed a consent form, and the Western Institutional Review Board (Olympia, WA) approved this project.

MESA: IRB approval obtained from: National Heart, Lung and Blood Institute and Forsyth County, North Carolina: Wake Forest University School of Medicine; St. Paul, Minnesota: University of Minnesota; Chicago, Illinois: Northwestern University, University of Illinois, Loyola University; New York, New York: Columbia University, St. Francis Hospital; Baltimore, Maryland: Johns Hopkins University; Los Angeles, California: University of California, Los Angeles.

NETT/NAS: IRB approval obtained from: Baylor College of Medicine, Houston, TX; Brigham and Women's Hospital, Boston, MA; Cedars-Sinai Medical Center, Los Angeles, CA; Cleveland Clinic Foundation, Cleveland, OH; Columbia University, New York, NY; Duke University Medical Center, Durham, NC; Mayo Foundation, Rochester, MN; National Jewish Medical and Research Center, Denver, CO; Ohio State University, Columbus, OH; Saint Louis University, Saint Louis, MO; Temple University, Philadelphia, PA; University of California, San Diego, San Diego, CA; University of Maryland at Baltimore, Baltimore, MD; University of Michigan, Ann Arbor, MI; University of Pennsylvania, Philadelphia, PA; University of

Pittsburgh, Pittsburgh, PA; University of Washington, Seattle, WA; the Human Studies Subcommittee of the Department of Veterans Affairs Medical Center

GenKOLS: The study was performed in accordance with the ethical standards laid down in the Helsinki Declaration. The Regional Committee for Medical Research Ethics (REK Vest), the Norwegian Data Inspectorate and the Norwegian Department of Health approved the case–control study.

The Rotterdam Study: The Rotterdam Study was approved by the institutional review board (Medical Ethics Committee) of the Erasmus Medical Center and by the review board of The Netherlands Ministry of Health, Welfare and Sports. All participants provided written informed consent.

SPIROMICS: IRB approval obtained from: Columbia University; the University of California at Los Angeles; the University of California at San Francisco; the University of Michigan; the University of Utah; Wake Forest University

BEOCOPD: IRB approval obtained from: The Human Research Committees of Partners Health Care (Brigham and Women's Hospital and Massachusetts General Hospital) and the Brockton/West Roxbury VA Hospital

ICGN: The study was IRB approved at all relevant institutions and participants provided written informed consent.

TCGS: IRB approval obtained from: Institute of Tuberculosis and Lung Diseases in Warsaw; Brigham and Women's Hospital in Boston

UK Biobank

Derivation and quality control of FEV_1 and FVC, COPD status, and smoking

The UK Biobank is a resource with approximately 500,000 persons age 40-69 for whom extensive baseline questionnaire data, physiologic measures, and biologic specimens (urine and blood samples) have been obtained[47]. The UK Biobank participants provided informed consent. The UK Biobank's Ethics and Governance Framework guides research ethics principles and polices, the adherence to which is overseen by the independent Ethics and Governance Council. Details of the derivation and quality control of spirometry, smoking, and case-control status were performed as described[48]. Briefly, of 502,682 individuals in UK Biobank, 445,754 individuals had at least two measures of FEV_1 and FVC and spirometry metrics, age, sex, height, and smoking. We used blow quality metrics from the Vitalograph spirometer and compared pre-defined measurements with the blow curve time series. We considered blows as 'acceptable' if their blow quality metrics were blank, ACCEPT, BELOW6SEC ACCEPT, or BELOW6SEC. We assessed back-extrapolated volumes derived from blow curve time series measurements as previously described[49] and excluded blows which their back-extrapolated volumes were greater than the larger of either 5% of FVC or 150 mL, leaving 776,927 blows from 387,277 participants. To confirm pre-defined FEV_1 and FVC, we independently derived FEV_1 and FVC from blow curve time series measurements. We further excluded blows which differences between pre-defined and newly derived FEV_1 and FVC were greater than 5%. Following above exclusion, 776,318 blows from 387,052 participants remained in the analysis. Additionally, we assessed if individuals' measures of FEV_1 and FVC were reproducible, allowing differences between the highest measure and others to be up to 250 mL. We defined COPD using modified GOLD criteria (as stated previously) based on highest values of reproducible measures of FEV_1 and FVC available in 324,299 individuals.

We assigned smoking status to individuals in UK Biobank based on their responses on questionnaires. Never-smokers were non-current smokers or smoked occasionally allowing up to 100 life-time cigarettes; ever-smokers were defined as either current, most days (current or all days in the past), or smoked occasionally with > 100 cigarettes. We defined pack-years as packs of cigarettes per day multiplied by years smoked. Pack-years was 0 for never smokers based on VariableID: 20116 smoking status. For ever smokers, we set pack-years to cigarettes per day (VariableID: 2887 in former smokers or VariableID: 3456 in current smokers) divided by 20 multiplied by duration of smoking (VariableID: 2897 minus VariableID: 2867 for former smokers and age at time of study minus VariableID: 3436 in current smokers).

Sample-based genotyping quality control

Details on procedure, imputation, and quality control of genotyping in UK Biobank were published earlier[50]. Briefly, we excluded 968 individuals with outlying heterozygosity or missingness. We further excluded: 1) 378 samples with sex mismatch (reported and genetic sex) 2) 188 samples with >10 3rd degree relatives 3) 652 samples with putative sex chromosome aneuploidy. This left 486,367 samples with genotypes for further analysis.

SNP-based Genotyping quality control

Besides the QC steps described in Bycroft et al.[50], we filtered out: 1) variants with minor allele frequency lower than 1%; 2) variants with imputation quality lower than 0.5; 3) variants not included into the Haplotype Reference Consortium (HRC) imputation panel, as recommended by UK Biobank at the time of analysis. This left us with 7,810,596 variants.

Identification of individuals of European ancestry

To identify individuals of European ancestry, we used K-means clustering on principal components (PCs) provided by UK Biobank. We performed K-means clustering using the first two PCs with number of clusters 3-8 in 486,367 samples passing sample-based quality control. Together with self-reported ethnicity data, we chose 6-cluster model as it reflects the expected broad ethnic groups. This left us 453,958 European individuals for further analysis (additional 45,865 putative European individuals in addition to 408,093 self-reported white British[50]).

Selecting individuals for genome-wide association analysis

We intersected the individuals passing spirometry QC and sample-based genotyping QC and selected individuals with European ancestry as described above. This left us 321,057 individuals in total. To identify individuals who were outliers in phenotype distribution, we plotted the distribution for each measure (FEV_1, FVC, and FEV_1/FVC) adjusting for sex, age, age^2, height, and smoking status (ever/never). We re-calculated phenotypic adjustment after excluding 10 European samples who were obvious outliers. As stated previously, we defined COPD using modified GOLD criteria for moderate to very severe airflow limitation[51]: FEV_1 less than 80% of predicted value and FEV_1/FVC less than 0.7. We excluded 59,358 individuals, who did not match the above criteria. To select completely unrelated individuals for genome-wide association analysis, we removed at least one individual from each related pair, giving preference to cases. Briefly, we created a graph from related pairs using genetic kinship information provided by UK Biobank. For each unconnected component, we removed control nodes (healthy individual), starting with the one with the highest degree. If all nodes were cases, we removed the one with the highest degree. We excluded additional 26,855 individuals based on kinship and further 34,035 individuals who had missing information on sex, ever smoking status, pack-years of

smoking, and age. We also excluded subjects who had withdrawn consent (updated as of May 2018). After excluding individuals, we were left with 200,792 individuals for genome-wide association analysis of COPD.

Sensitivity analyses for smoking, sex, and COPD phenotypic misclassification

To determine whether our results were driven by, or differed by, smoking status, we examined COPD-associated variants in relationship to ever-smoking status in UK Biobank. In addition, we separately examined these variants in ever- and never-smokers in UK Biobank. To test for sex-difference in genetic effects among our top variants, we performed a sex-stratified GWAS of COPD in UK Biobank and tested for differences between effects among males and females. For the latter we utilized the "difference test[52,53]". In addition, we also included a variant-by-sex interaction term and tested this term. We defined significant effect differences by Bonferroni correction and further, we attempted to replicate all nominally significant variants (P for the difference test < 0.05) from our top SNPs in a meta-analysis of sex-stratified COPD association analyses in COPDGene NHW, COPDGene AA, ECLIPSE, and GenKOLS. We also investigated variants not our top variants but reaching genome-wide significance ($P<5x10^{-8}$) only in one sex. We attempted to replicate these variants in the meta-analysis of COPD case-control cohorts as mentioned above. We used 5% Bonferroni corrected P value to determine significance in replication.

To simulate power for the difference test, we modified the power simulations in Winkler et al. [52] and made them applicable to a case-control dataset. We approximated the phenotypic variance attributable to the risk locus as in Peyrot et al.[54]:

$$R_i^2 = 2pq(RR-1)^2/m^2$$

where m^2 is the mean liability of case subjects and depends on the prevalence of COPD, p is the risk allele frequency of variant i, $q = 1 - p$ and RR is the relative risk, respectively. $R_i$can be derived from R_i^2 and the direction of the effect i.

The variance of a variant i ($se_i{}^2$) is defined as

$$se_i{}^2 = \left[\left(1 - R_i^2\right)\sigma_Y^2\right]/(n_i\sigma_G^2)$$

Where n_i is sample size in stratum i.

Hence, we can derive the power of difference test as[52]

$$\Phi\left(-z_{1-\frac{\alpha}{2}} - \sqrt{n_1}\frac{R_1 - R_2}{\sqrt{1 - R_1^2 + \frac{1}{f}(1 - R_2^2)}}\right) + \Phi\left(-z_{1-\frac{\alpha}{2}} + \sqrt{n_1}\frac{R_1 - R_2}{\sqrt{1 - R_1^2 + \frac{1}{f}(1 - R_2^2)}}\right)$$

As Φ denotes the cumulative standard normal distribution, z_q is the q-th quantile of Φ, α is the alpha level of the test, and $f = n_2/n_1$.

We performed an array of analyses to ascertain if a pre-bronchodilator spirometry definition of COPD (as opposed post-bronchodilator spirometry) and the inclusion of subjects with self-reported asthma subsequently impacted both effect size estimates and P values of genetic association. First, in the COPDGene study, we examined the difference in COPD diagnosis using pre- and post-bronchodilator spirometry. Then, we stratified individuals by self-reported asthma and compared the percent of

individuals in each stratum that were misclassified as having COPD based on a pre-bronchodilator spirometry definition of COPD. We also re-tested the association of COPD by excluding individuals in UK Biobank with self-reported asthma and the same control individuals as the primary analysis. Finally, in a subset of COPD case-control studies including COPDGene NHW and AA, ECLIPSE, and Norway/GenKOLS, we performed logistic regression (controlling for pack years, current smoking, sex, age, and PCs of genetic ancestry) and a meta-analysis to assess the relative COPD association effect sizes and P values of our top variants when using a pre- vs post-bronchodilator definition of COPD. For the comparison of effect sizes, we limited the variants being compared to those that were at least nominally significant ($P < 0.05$) in the meta-analysis of the COPD case-control studies.

Identification of cell types

Linkage equilibrium score regression (LDSC) requires specifically expressed gene sets to perform cell type-specific analysis[55]. For two human datasets[56,57], we computed t-statistics for each cell type on gene expression data in Transcripts Per Kilobase Million (TPM) as described previously[55]. We constructed gene sets for each cell type using genes in the top 10% of sorted t-statistics[55], and a control gene set (i.e., all genes available in a dataset). We annotated regions +/- 100-kb from each gene and computed LD scores as previously described[55]. For two mouse datasets[58], we used pre-computed Wilcoxon or Welch P-values for each cell to select genes in the top 10%. We mapped gene identifiers from mouse to human using biomaRt[59].

We tested for enrichment of lung cell type in sets of genome-wide significant variants using SNPsea[60,61]. Briefly, SNPsea performed three steps: identification of specifically expressed genes, assignment of gene in a locus, and significance testing. We first computed the Euclidean norm of gene expression values (TPM) in all cell types and divided expression values in each cell type using this value. The score was then converted to nonparametric percentiles. Second, we identified the most specifically expressed gene in a given locus by ranking the score. Finally, we then tested for significance using permutation using matched SNP sets.

Fine-mapping

We used biomaRt[59] to annotate variants in each credible set based on the Ensembl Variant Effect Predictor . We defined deleterious variants as those which resulted in non-synonymous, stop, or splice variants (terms: transcript_ablation, splice_acceptor_variant, splice_donor_variant, stop_gained, frameshift_variant, stop_lost, start_lost, transcript_amplification, inframe_insertion, inframe_deletion, missense_variant, and protein_altering_variant). Variants were annotated using Haploreg v4.1[62] and SNPnexus[63].

Target gene identification

Rare coding variants

We performed single-variant analyses using Firth and efficient resampling methods (SKAT R package[64]) for the COPDGene data (case-control) and generalized linear mixed models (GMMAT) for the BEOCOPD-ICGN data (using lung function) as previously published[65]. Gene-based analyses were conducted using burden, SKAT, and SKAT-O tests with asymptotic and efficient resampling methods (SKAT package) combined with Fisher's method for the COPDGene data, and using SKAT-O tests (MONSTER) for the BEOCOPD-ICGN data. Two variant-filtering criteria were considered: deleterious variants (predicted by FATHMM) with minor allele frequency (MAF) < 0.01, and functional variants (moderate effect predicted by SNPEff) with MAF < 0.05. We also applied a gene-based segregation test (GESE) to the ultra-rare

(MAF < 0.1%) and loss-of-function variants in the BEOCOPD-ICGN data on the severe COPD affection status. In gene-based analyses, we combined results from all methods above and retained only most significant P values for each gene.

DEPICT

DEPICT utilizes pairwise gene correlation data from ~77,000 microarrays to create data-driven "reconstituted" gene sets from a backbone of curated data from GO, KEGG, REACTOME, protein-protein interaction data, and murine phenotypic gene sets. These "reconstituted" gene sets are used to prioritize genes at each GWAS loci based on the similarity each gene's "reconstituted" gene set membership to the gene set memberships of genes at other GWAS loci[66].

Methylation quantitative trait loci (mQTL)

Of all genome-wide significant loci, we searched for overlapping methylation quantitative trait loci (mQTL) using previously published data[67]. Briefly, lung tissues from 90 severe COPD cases (FEV_1 < 50% predicted) and 36 control subjects undergoing lung transplantation, lung volume reduction surgery, or lung nodule resection[68]. All subjects were self-reported former smokers at least 1 month prior to the surgery. A cis-mQTL analysis was performed using the R/Bioconductor package Matrix eQTL (version 2.1.1)[69]. We tested associations of each CpG site and genetic variants within 500 kb upstream and downstream (from the CpG site) using a linear regression model adjusting for age, sex, smoking pack-years, two principal components of genetic ancestry, batch number, and principal components of methylation data as previous described[67]. To determine whether these signals co-localized (rather than being related due to linkage disequilibrium), we performed colocalization analysis between our genome-wide significant loci and mQTL using eCAVIAR[70] (eQTL and GWAS CAusal Variants Identification in Associated Regions). We tested variants that were significant in both datasets, P < 0.0027 in GWAS (equivalent to Z score > 3, as recommended by the author[70]) and P < 3.2×10^{-6} in mQTL[67]. We estimated the posterior probability of a variant being shared in both GWAS and mQTL, using a cut-off of 0.1 as previous demonstrated[70].

Drug repositioning using drug-gene expression signatures

We used a gene-based association method that utilizes GWAS summary statistics and gene expression reference databases to produce a gene list for drug-gene expression similarity analysis[71,72]. In brief, we tested for associations of the genetic component of gene expression and COPD using gene expression data in a lung[73]. This method gave us a rank gene list of up-regulated and down-regulated genes. We used the Query, an application in clue.io[74], to calculate connectivity score of our top 100 up-regulated and 100 down-regulated genes and drug-gene expression signatures. The connectivity score, ranging from -1 to 1, reflects the closeness between the expression profiles[75]. A negative connectivity score means that our down-regulated genes are at the top of the reference profile[75]. We included drug-gene expression signatures from 2,837 compounds in nine cell lines in the analysis[74]. To find candidate drugs for repositioning, we used negative connectivity score less than -90% as the threshold[74], hypothesizing that a drug candidate need to produce an opposing or reverse expression signature induced by the disease.

Determining numbers and features of clusters

To determine phenotypic clustering, we identified the optimal number of clusters using the Calinski index[76,77]. To identify features that independently predict cluster membership, we fitted a logistic

regression model via penalized maximum likelihood using the glmnet package[78]. We determined optimal regularization parameters using 10-fold cross validation.

Supplementary Results

Relationship of genome-wide significant variants to smoking

To assess the effect of smoking, we tested association of 82 genetic variants and smoking status (ever- and never- smokers) in the UK Biobank. Of the 82 identified loci, three were associated with cigarette smoking after Bonferroni correction ($P < 0.05/82$) including the strongest association at the known 15q25 locus, but also *IER3* ($P = 9{\cdot}2 \times 10^{-5}$) at 6p21 and *SPPL2C* ($P = 1{\cdot}4 \times 10^{-4}$) at 17q21. To test if smoking status could have confounded the association of these variants and COPD, we re-performed the analysis of COPD stratifying by smoking status in the UK Biobank. The lead variants at these latter two loci were highly significant for COPD in never-smokers ($P = 6{\cdot}7 \times 10^{-13}$ and $2{\cdot}8 \times 10^{-5}$ at *IER3* and *SPPL2C* loci, respectively). Seventy-eight of 82 genetic loci were nominally significant in never-smokers ($P < 0.05$) with the exception of four loci including *TESK2*, *RBMS3*, *RIN3*, and *CHRNA3* (**Supplementary Tables 14 and 15**).

Identification of sex-specific genetic effects

Of 82 genetic variants associated with COPD, we did not find significant evidence of difference in effect sizes of COPD among males and females after adjusting for multiple testing (**Supplementary Table 16**). The strongest evidence for different effect size was at *TGFB2* locus (rs3009947, $beta_{males}$=-0.06 and $beta_{females}$=-0.12, and P unadjusted =0.003). Twelfth of 82 COPD-associated variants were nominally significant for sex-difference ($P<0.05$); none of which was replicated in a meta-analysis of a subset of COPD case-control cohorts (**Supplementary Note** and **Supplementary Table 16**). Similarly, using a test for sex interactions, we did not identify any significant variants after correcting for multiple testing (top unadjusted P = 0.003, **Supplementary Table 16**). Power calculations indicate that we were well-powered to detect effect size different of > 0.04 for an allele frequency of ~0.27; for details, see the **Methods and Supplementary Figure 6**.

COPD misclassification sensitivity analyses

In 10,720 persons with both pre-bronchodilator and post-bronchodilator spirometry available in the COPDGene study, we defined ground truth for COPD status using post-bronchodilator spirometry. Using a pre-bronchodilator definition of COPD resulted in 113 out of 4,289 individuals (2.6%) without COPD being incorrectly assigned as a COPD cases and 18 out of 3,694 individuals (0.49%) with COPD being incorrectly assigned as controls (**Supplementary Table 17**). A self-reported doctor diagnosis of asthma had no significant impact on either the overall misclassification of individuals as being controls, COPD cases, or neither (16.6% without asthma and 18.3% with asthma, χ squared p value = 0.17); or specifically on the misclassification of individuals without COPD as having COPD (1.5% without asthma and 1.7% with asthma, χ squared p value = 0.77) when using pre-bronchodilator instead of post-bronchodilator spirometry.

When comparing the relative effect size of our top 164 variants for a pre- versus post-bronchodilator definition of COPD in the subset of our COPD case-control cohorts (COPDGene, ECLIPSE, GenKOLS), 49 variants were at least nominally significant for association with COPD in the meta-analysis (**Supplementary Table 18**). For these 49 variants, the odds ratio obtained when using a pre- versus post-bronchodilator definition of COPD was very similar with Pearson correlation coefficient of 0.92 (p value < 1e-10). Generally, the post-bronchodilator COPD definition yielded larger effect sizes; however, the 95% confidence intervals for all estimates crossed the line of identity (**Supplementary Figure 7**).

We performed additional GWAS by excluding 6,717 self-reported asthma from 21,081 COPD cases in the UK Biobank. While the decreased sample size resulted in less significant associations, we observed highly concordant estimates of effect size between the analysis including and excluding asthma (**Supplementary Figure 8**). One outlier was the variant rs2070600 in *AGER*, which demonstrated a higher OR after excluding individuals with asthma. Association statistics for all genome-wide significant loci are shown in **Supplementary Tables 19 and 20**.

To estimate the effect of asthma inclusion in the overlapping loci analysis, we re-ran gwas-pw using statistics from GWAS of COPD including individuals with asthma. The analysis including asthmatic individuals identified 24 shared genome segments between COPD and asthma (posterior probability of shared association > 0.7). All segments identified in the analysis excluding asthmatics were also identified in this analysis, with the exception of the segment in chr12 (near *STAT6*) (posterior probability of 0.81 excluding asthma, 0.29 including asthma). Results of all loci (with and without exclusion of asthma cases) are shown in **Supplementary Table 21**.

In addition to individual loci effect, we re-estimated the genetic correlation between COPD and asthma using asthma-excluded statistics of COPD. The genetic correlation decreased but remained highly significant, from 0.42 (s.e.=0.04, $p=3.1x10^{-26}$) to 0.26 (s.e.=0.05, $p=4.2x10^{-8}$).

Functional consequences of fine-mapped variants

We explored functional consequence of variants with high posterior probability of association (PPA) from the fine-mapping analysis. We used various non-coding variation scoring systems to suggest variants' functional status. An intronic variant at *NPNT* (rs34712979) appeared to reside in regulatory regions for several cell types, including enhancers for fetal lung (**Supplementary Table 6**), and was predicted to be functional: CADD phred score =15.18 (among 15% most deleterious substitutions), FATHMM non-coding score=0.98 (predicted deleterious if >0.5), DeepSEA functional significance score=0.004 (a range of 0-1; lower scores mean higher likelihood of functional significance of the variant), FunSeq2 non-coding score=0.96 (a range of 0-1; higer scores mean more likely to be functional), ReMM score=0.95 (higher scores indicate more likely to be deleterious variants).

Genetic correlation of COPD and related traits

To gain further insight of genetic contribution of COPD and other traits, we estimated genetic correlation among traits using linkage disequilibrium score regression (LDSC) in LD hub[79,80] (**Supplementary Table 22**). We again identified correlations between COPD and lung function (FEV_1 and FEV_1/FVC), asthma, height, and additionally identified correlation of COPD and lung cancer (r_g=0.18, P=0.003). However, we were not able to identify significant correlations with common COPD comorbidities, including bone mineral density, major depressive disorder, coronary artery disease, or type 2 diabetes mellitus. We found suggestive (P < 0.05) genetic correlation between COPD and hand grip, angina, weight, schizophrenia, and gastroesophageal reflux disease (**Supplementary Table 22**).

Identification of drug targets

The recent development of richer datasets of drug-induced gene expression signatures[74] and a statistical framework that utilizes genome-wide associations[71] allow us to utilize genome-wide data for drug repositioning. This approach considers the similarity between disease and drug-induced gene expression signatures in an opposing pattern (i.e., use drug to reverse diseased gene expression signature). We approximated COPD gene expression patterns by calculating transcriptome-wide associations in lung

tissue from genetic predictors[72]. We calculated standardized connectivity scores[74] from drug-gene expression profiles across 2,837 compounds in nine cell lines including A594 and A555. We identified seven compounds with an opposing connectivity score >= 90%: leu-enkephalin (an opioid receptor agonist), huperizine-a (an acetylcholinesterase inhibitor), periplocymarin (an apoptosis stimulant), PAC-1 (a caspase activator), TER-14687 (an inhibitor of translocation of PKCq in T cells), vincristine (a tubulin inhibitor), and terreic-acid (a Bruton's tyrosine kinase (BTK) inhibitor).

Supplementary Information for Select Candidate Target Genes

* empirical pattern of gene expression in tissues from Human Protein Atlas[81] (broad=expressed in multiple tissues, exclusive=expressed exclusively in lung, none=rarely expressed in lung).

Locus (SNP)	Genes	Expression*	Evidence
1p13.3	*DENND2D*	Broad	DENN domain containing 2D predominantly expresses in lymph node and appendix[81]. It is involved in Rab guanyl-nucleotide exchange factor activity[82].
	CEPT1	Broad	Choline/ethanolamine phosphotransferase 1 encodes an enzyme which functions in the synthesis of choline- or ethanolamine- containing phospholipids[83].
	DRAM2	Broad	DNA damage regulated autophagy modulator 2 encodes the protein that binds microtubule-associated protein 1 light chain 3 and is required for autophagy[84]. It is also associated with non-ST elevation myocardial infarction[85] and retinal dystrophy[86].
	CHIA	Exclusive	Chitinase, acidic encodes a protein that degrades chitin[83]. Its expression is specific to lung[81,87]. Genetic variants in or around this gene were associated with baseline FEV_1 and rate of FEV_1 decline[88], asthma[89–92], acid mammalian chitinase activity[91,93], and IgE[90] although there were also some null results[94]. The protein is involved in T helper-2 (Th2)-mediated diseases and protects pulmonary epithelial cells from growth factor withdrawal- and Fas ligand-induced apoptosis[95]. It was induced via a Th2-specific, interleukin-13 (IL-13)-mediated pathway in epithelial cells and macrophages in an aeroallergen asthma model[96], and ADAM17/EGFR-dependent pathway[97].
	OVGP1	Broad	Oviductal glycoprotein 1 encodes an epithelial glycoprotein. It is secreted from non-ciliated oviductal epithelial cells and associates with ovulated oocytes, blastomeres, and spermatozoan acrosomal regions[83].
	WDR77	Broad	WD repeat domain 77 encodes an androgen receptor coactivator[83]. It may be involved in the early stages of prostate cancer[83].
	ATP5F1 (*ATP5PB*)	Broad	ATP synthase peripheral stalk-membrane subunit b encodes a subunit of mitochondrial ATP synthase[83]. Its hypomethylation was involved in drug resistance in chronic myeloid leukemia[98].
	FAM212B (*INKA2*)	Broad	Inka box actin regulator 2 is broadly expressed in multiple tissues including lung[83].
	PROK1	None	Prokineticin 1 induces proliferation, migration, and fenestration in capillary endothelial cells derived from endocrine glands[83]. It is induced by hypoxia, and often is complementary to the expression of vascular endothelial

			growth factor (VEGF)[83]. It was involved in angiogenesis in cystic fibrosis[99].
	KCNC4	Broad	Potassium voltage-gated channel subfamily C member 4 encodes a protein that generates atypical voltage-dependent transient current that may be important for neuronal excitability[83]. It was related to subacute hypoxia in pulmonary arterial smooth muscle cells[100].
3p14.3	*ABHD6*	Broad	Abhydrolase domain containing 6 is involved in the control of macrophage activation and inflammation[101].
	ARF4	Broad	ADP ribosylation factor 4 encodes a protein that stimulates the ADP-ribosyltransferase activity of cholera toxin and plays a role in vesicular trafficking and as an activator of phospholipase D[83]. It is involved in Golgi stress and susceptibility to pathogens[102].
	IL17RD	Broad	Interleukin 17 receptor D encodes a membrane protein belonging to the interleukin-17 receptor (IL-17R) protein family[83]. The gene product affects fibroblast growth factor signaling, inhibiting or stimulating growth through MAPK/ERK signaling[83]. It also interacts with TNF receptor 2 (TNFR2) to activate NF-κB[103].
	ARHGEF3	Broad	Rho guanine nucleotide exchange factor 3 encodes a protein that activates RHOA and RHOB, which have a role in bone cell biology[83]. Its gene product also regulates a SPARC protein that participates in the assembly and turnover of the extracellular matrix[104]. It also has a role in iron uptake[105]. Genetic variants in or around this gene were associated with bone mineral density[83].
7p21.1	*ITGB8*	Broad	Integrin subunit beta 8 is a member of the integrin beta chain family and encodes a single-pass type I membrane protein that binds to an alpha subunit to form an integrin complex[83]. The complexes mediate cell-cell and cell-extracellular matrix interactions and this complex plays a role in human airway epithelial proliferation[83] and repair[106]. Its expression was increased in COPD[107–109]. It is involved in dendritic cell trafficking, and airway inflammation and fibrosis processes[108]. It mediates epithelial homeostasis through an MMP-dependent pathway[110] and TGF-B[107,111]. It is also regulated by SP3, AP-1, and the p38 pathway[112].
	ABCB5	None	ATP binding cassette subfamily B member 5 belongs to the ATP-binding cassette (ABC) transporter superfamily of integral membrane proteins[113].
	TMEM196	None	Transmembrane protein 196 is a novel functional tumor suppressor and is associated with lung cancer[114].
11p15.2	*BTBD10*	Broad	The gene product of BTB domain containing 10 is an Akt activator[115]. This gene was associated with neurologic diseases[116].

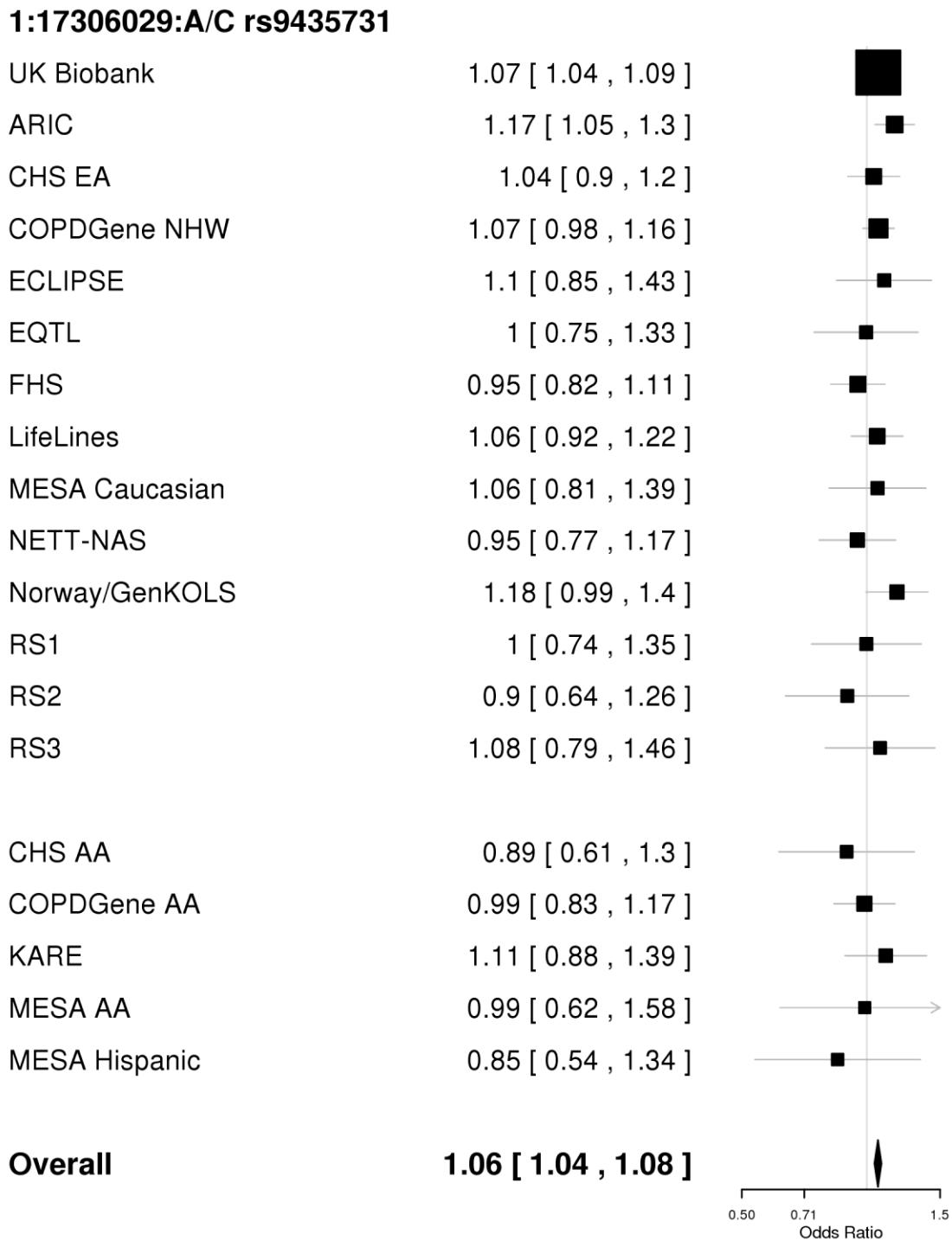
	PARVA	Broad	Parvin alpha encodes a member of the parvin family of actin-binding proteins[83]. The encoded protein is part of the integrin-linked kinase signaling complex and plays a role in cell adhesion, motility and survival[83].
	MICALCL	Broad	MICAL C-terminal like gene product is predominantly expressed in skin, testis, and lung[81].
15q25.2	*ADAMTSL3*	Broad	ADAMTS like 3 encodes a gene product that plays a role in cell-matrix interactions or in assembly of specific extracellular matrices[117]. It was associated with schizophrenia[118] and cardiac disorders in tetrasomy[119].
	BNC1	None	Basonuclin 1 encodes a zinc finger protein that is present in the basal cell layer of the epidermis and in hair follicles, and in the germ cells of testis and ovary[83,120]. Its gene product modulates epithelial plasticity and TGF-β1-induced loss of epithelial cell integrity[121]. It is a Pol I and Pol II transcription factor that is associated with epithelial expansion and proliferation[122,123].
	BTBD1	Broad	BTB domain containing 1 encodes a protein that binds topoisomerase I. It is a transcription factor[124] in the human histone deacetylase family[125]. Its gene product is involved in mesenchymal[126] and muscle cell differentiation[127].

Supplementary Figures

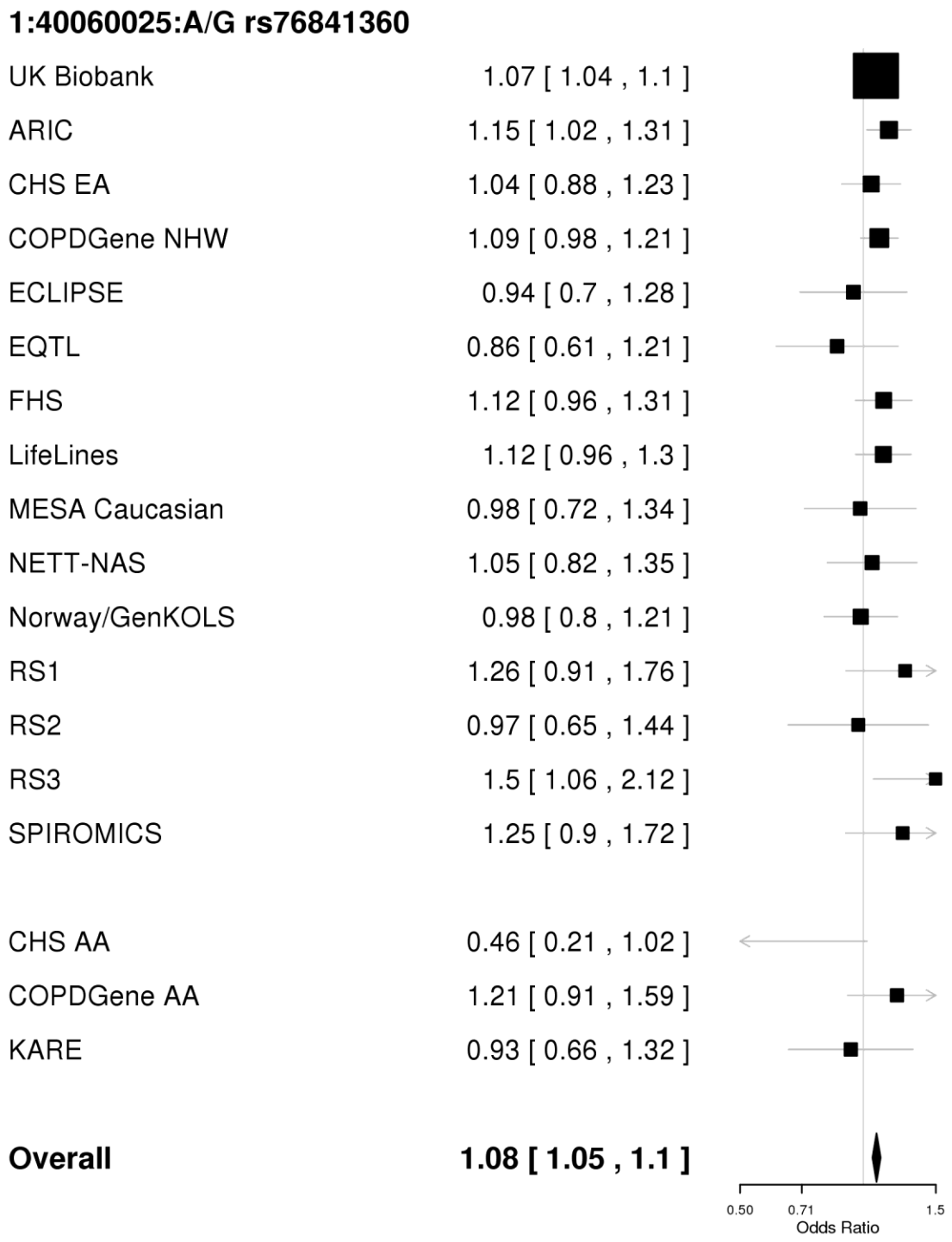
Supplementary Figure 1: Forest plots for 82 genome-wide significant associations

Association statistics are based on the overall meta-analysis of COPD (35,735 cases and 222,076 controls). Error bars indicate 95% confidence interval.

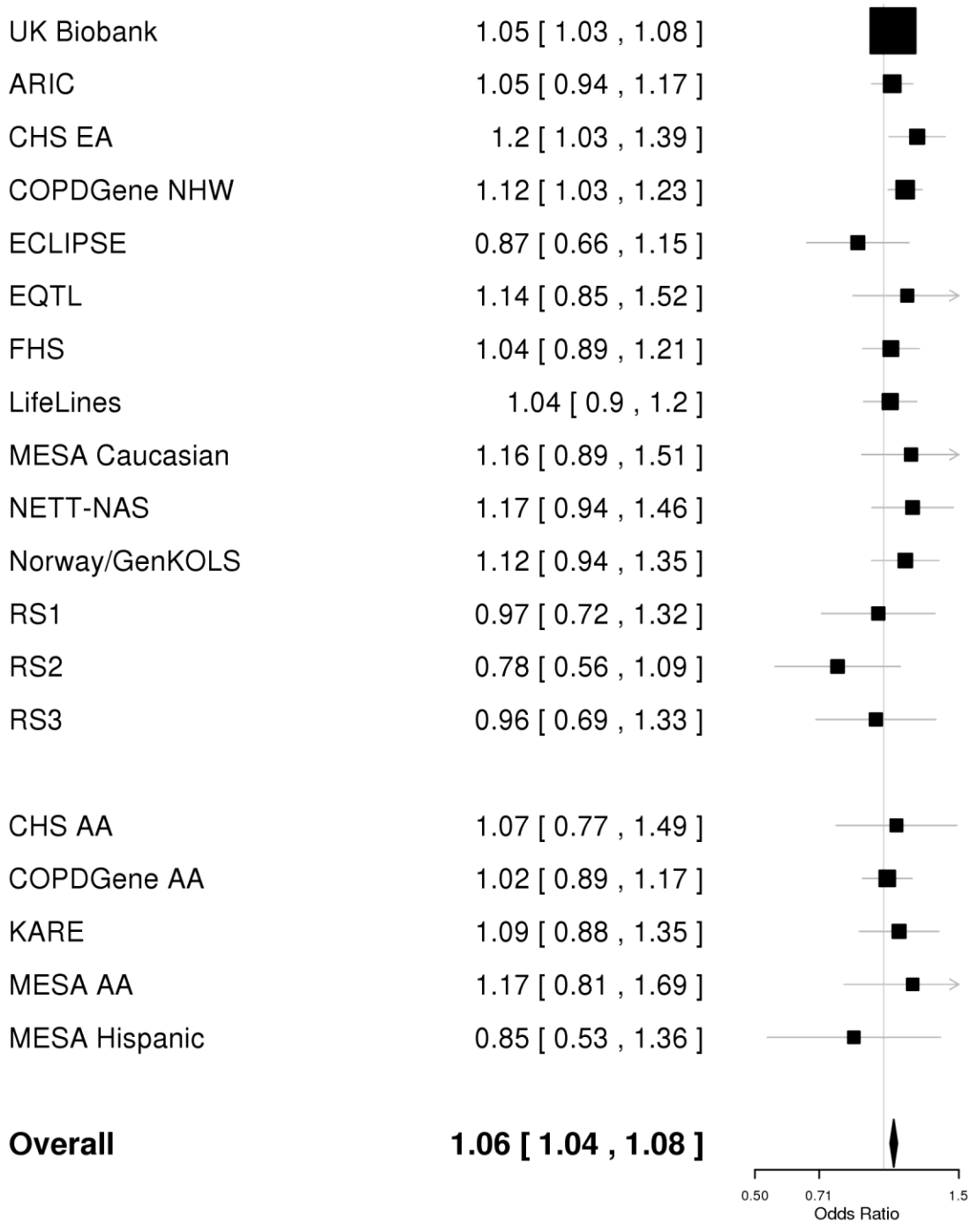
Supplementary Figure 1-1: Forest plot for rs9435731 (*MFAP2* locus at 1p36.13)



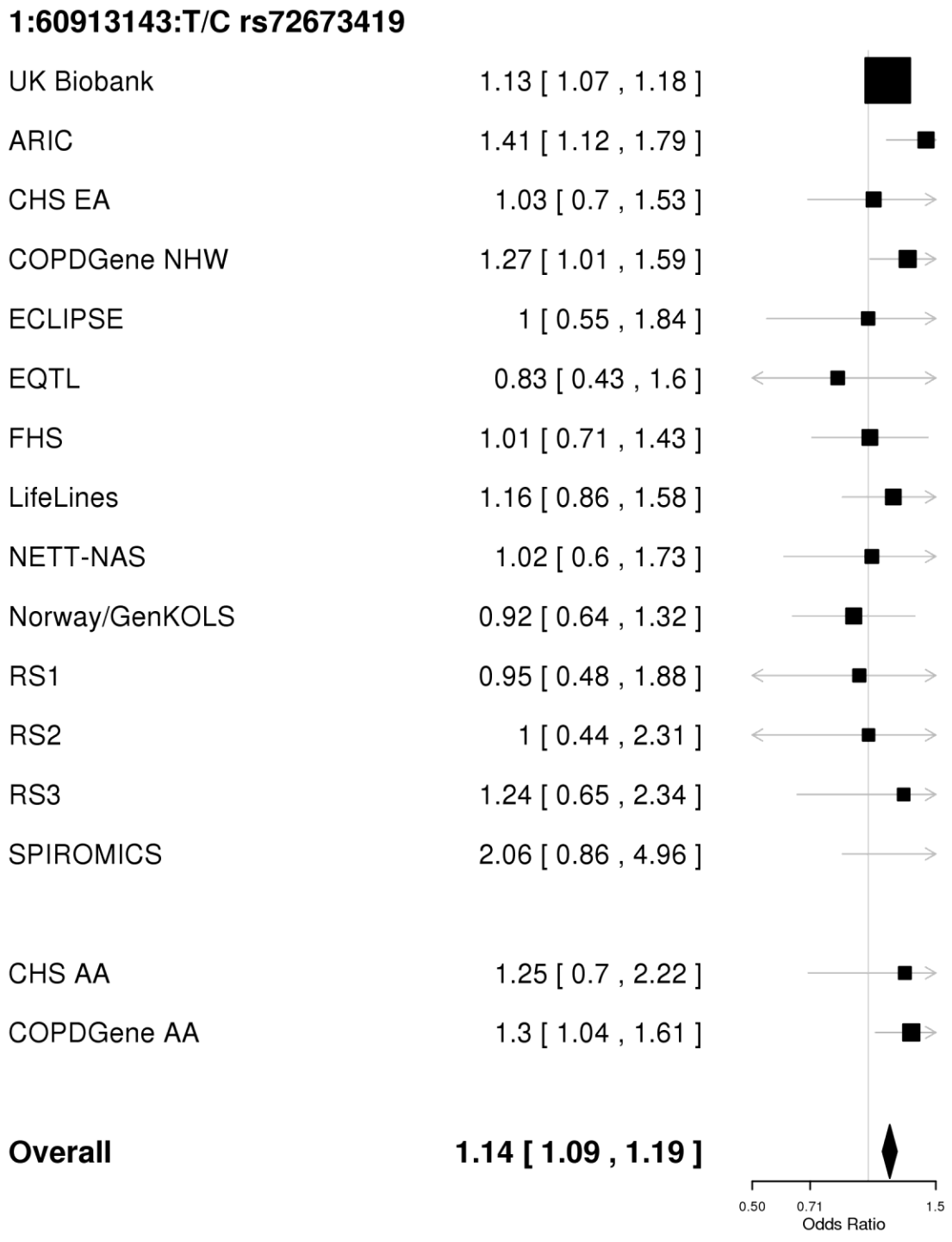
Supplementary Figure 1-2: Forest plot for rs76841360 (*PABPC4* locus at 1p34.3)



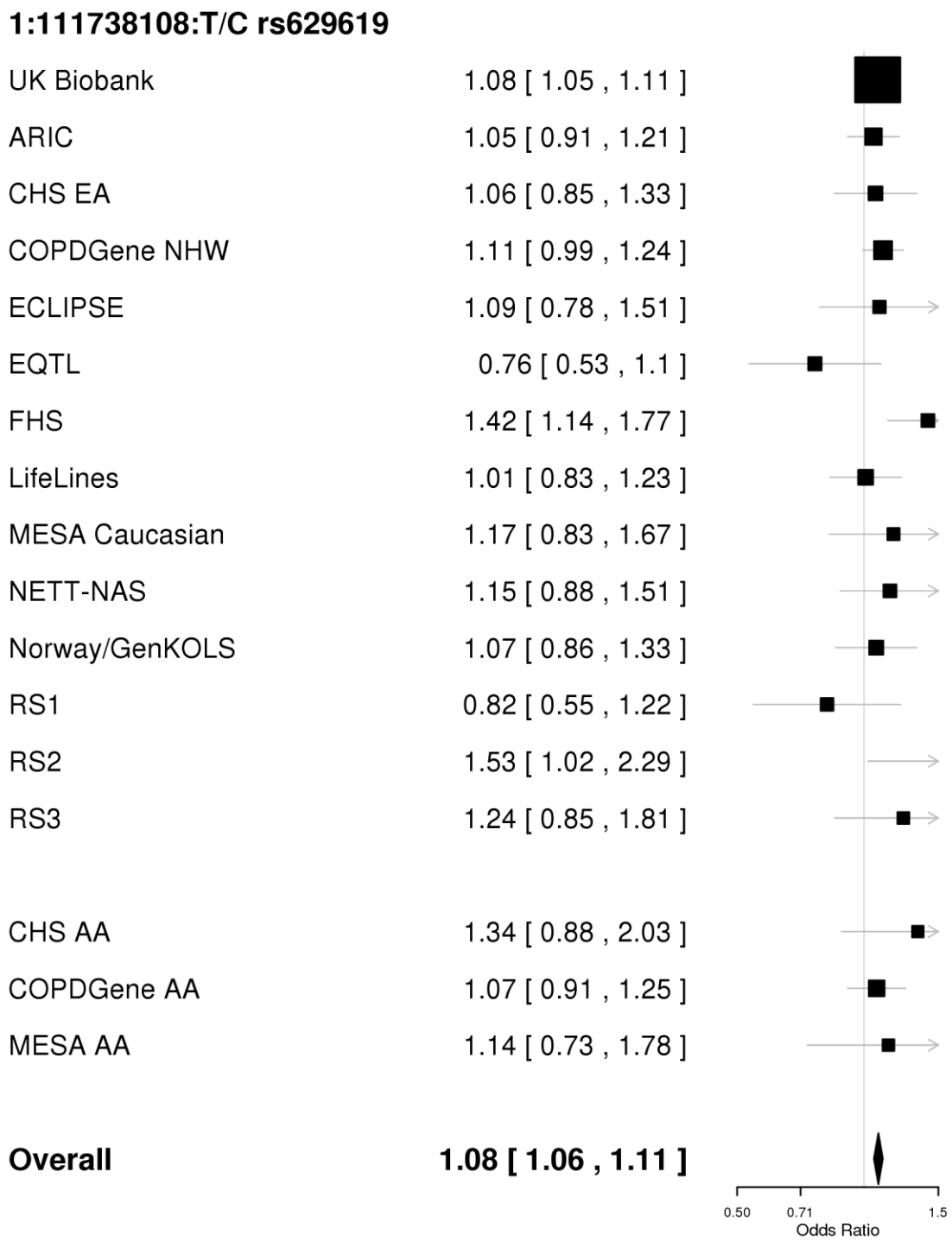
Supplementary Figure 1-3: Forest plot for rs4660861 (*TESK2* locus at 1p34.1)



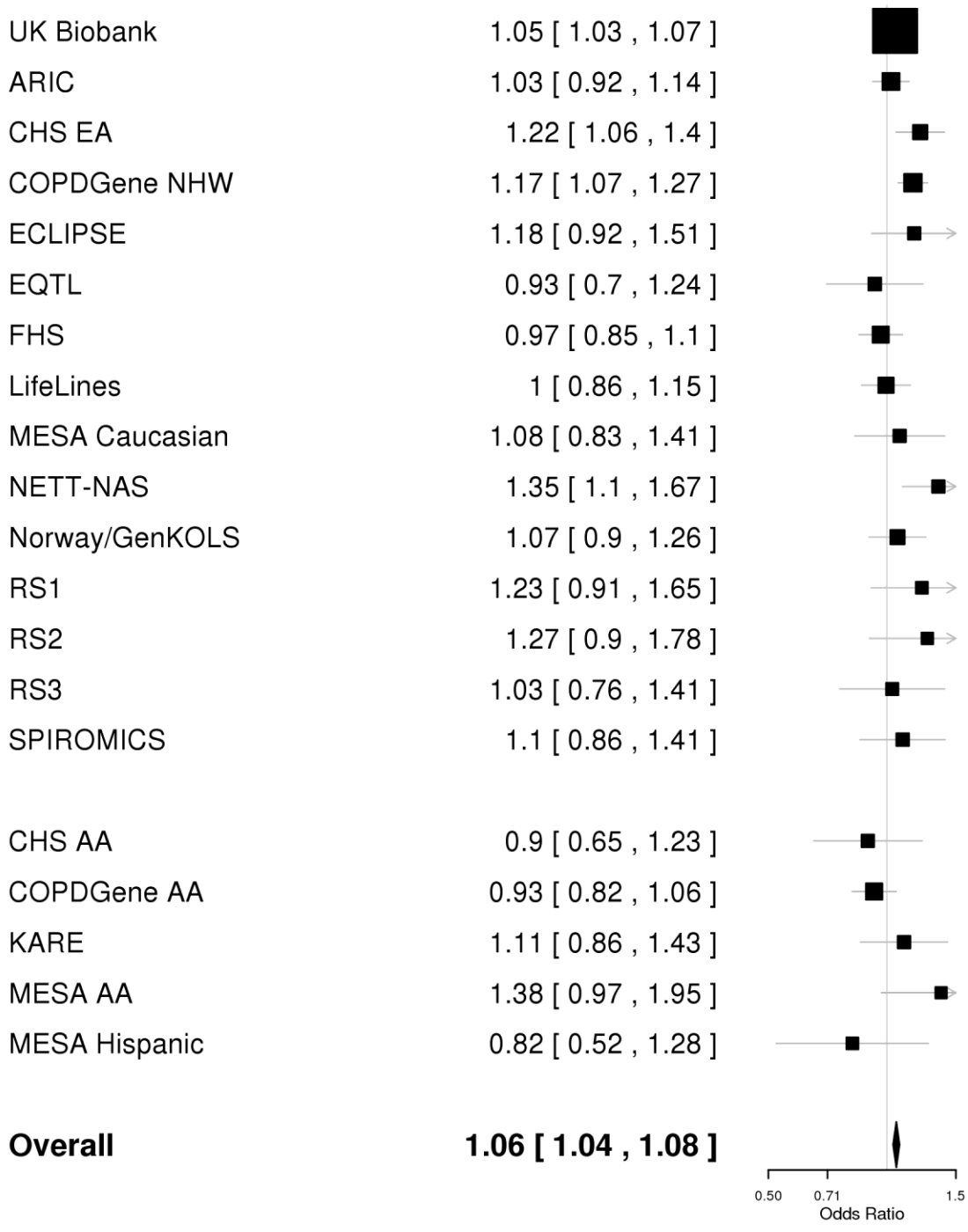
Supplementary Figure 1-4: Forest plot for rs72673419 (*C1orf87* locus at 1p32.1)



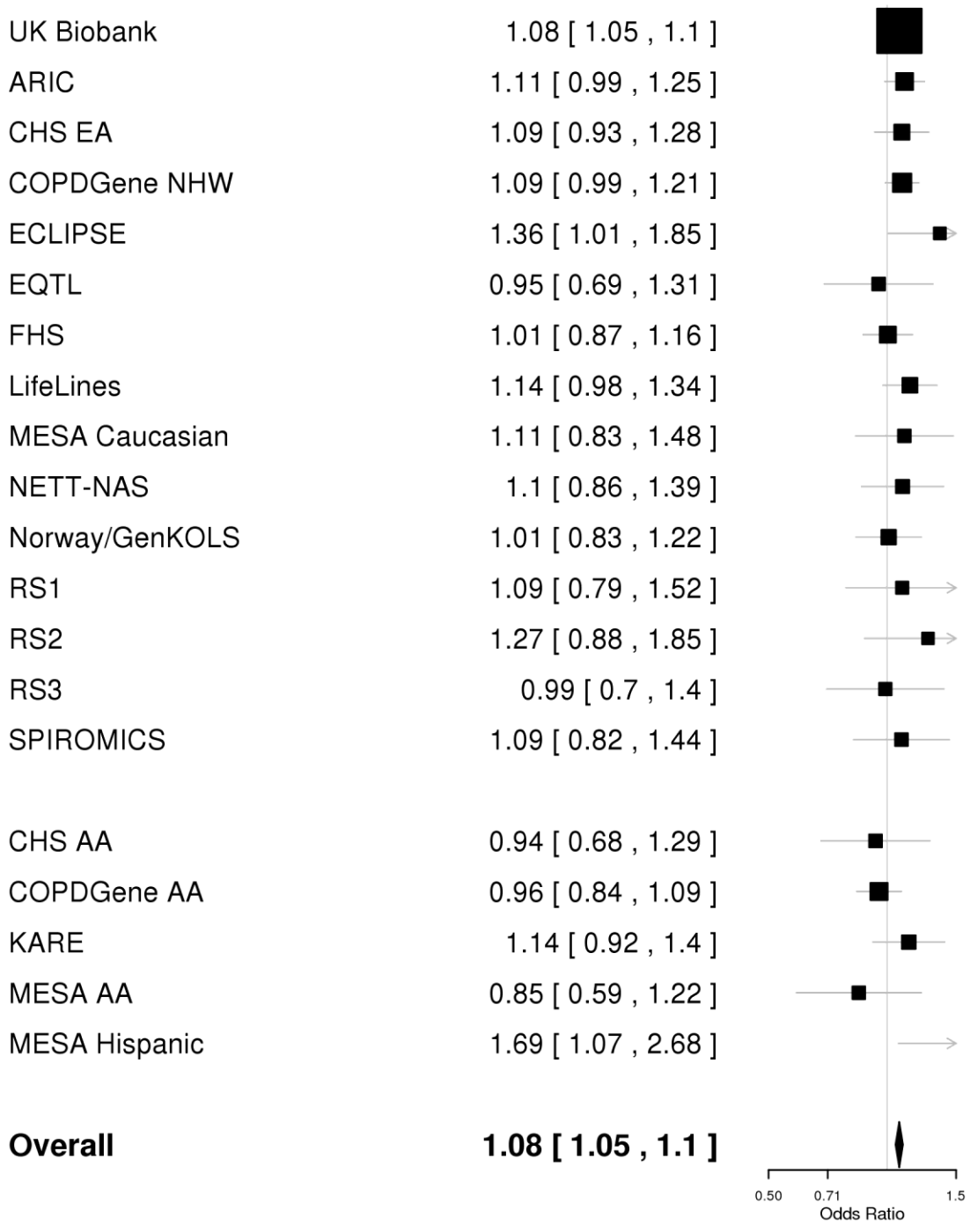
Supplementary Figure 1-5: Forest plot for rs629619 (*DENND2D* locus at 1p13.3)



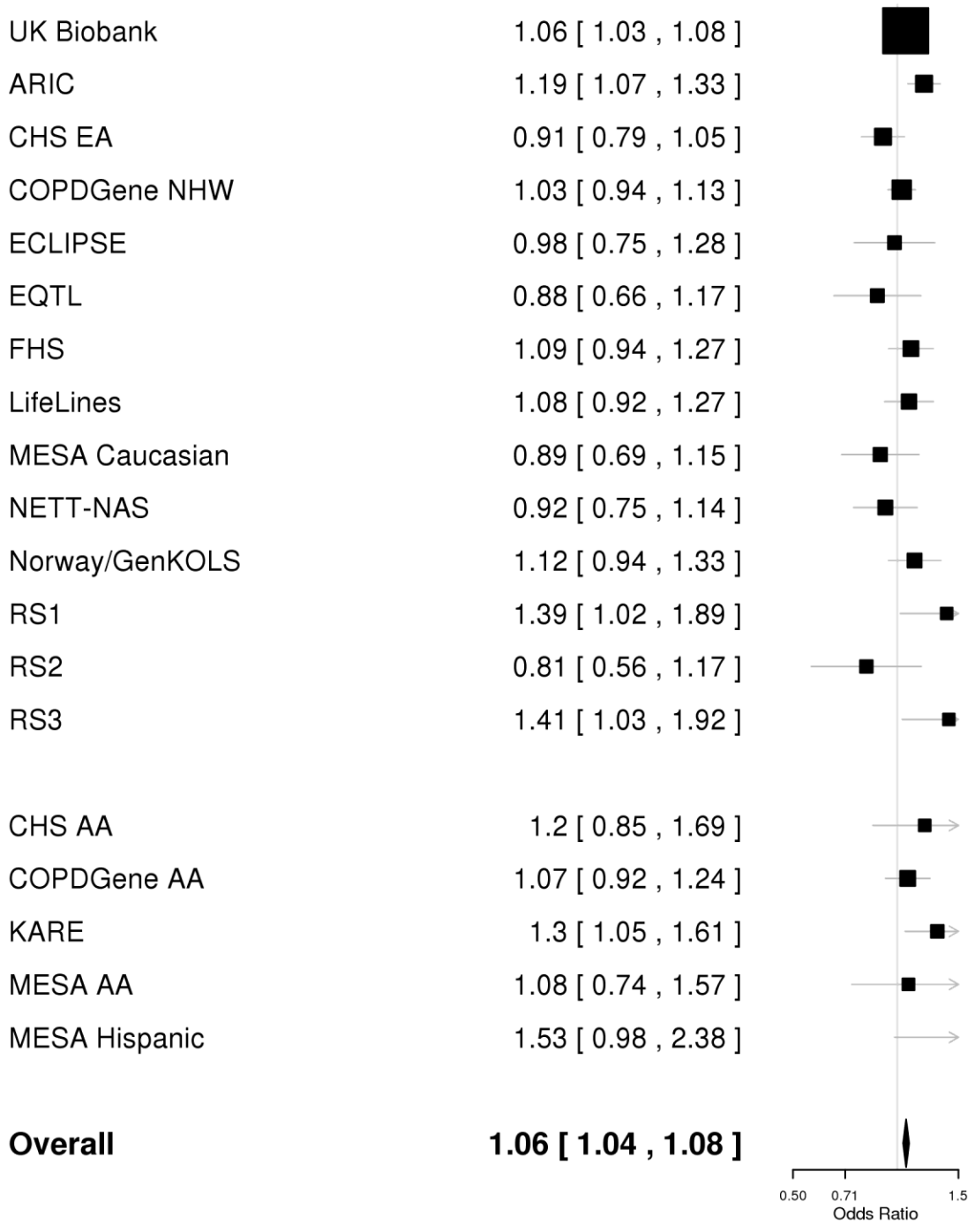
Supplementary Figure 1-6: Forest plot for rs3009947 (*TGFB2* locus at 1q41)



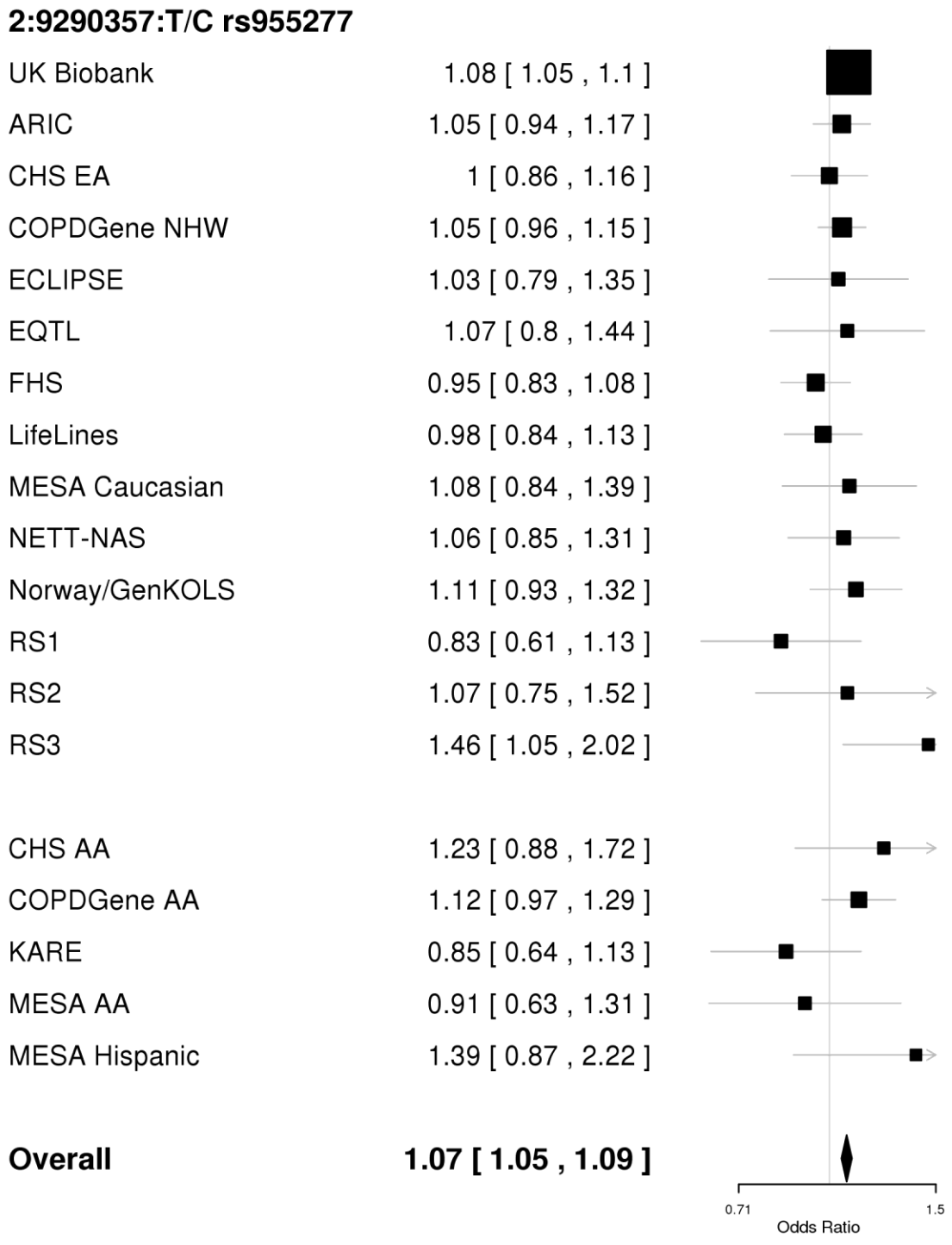
Supplementary Figure 1-7: Forest plot for rs11118406 (*SLC30A10* locus at 1q41)



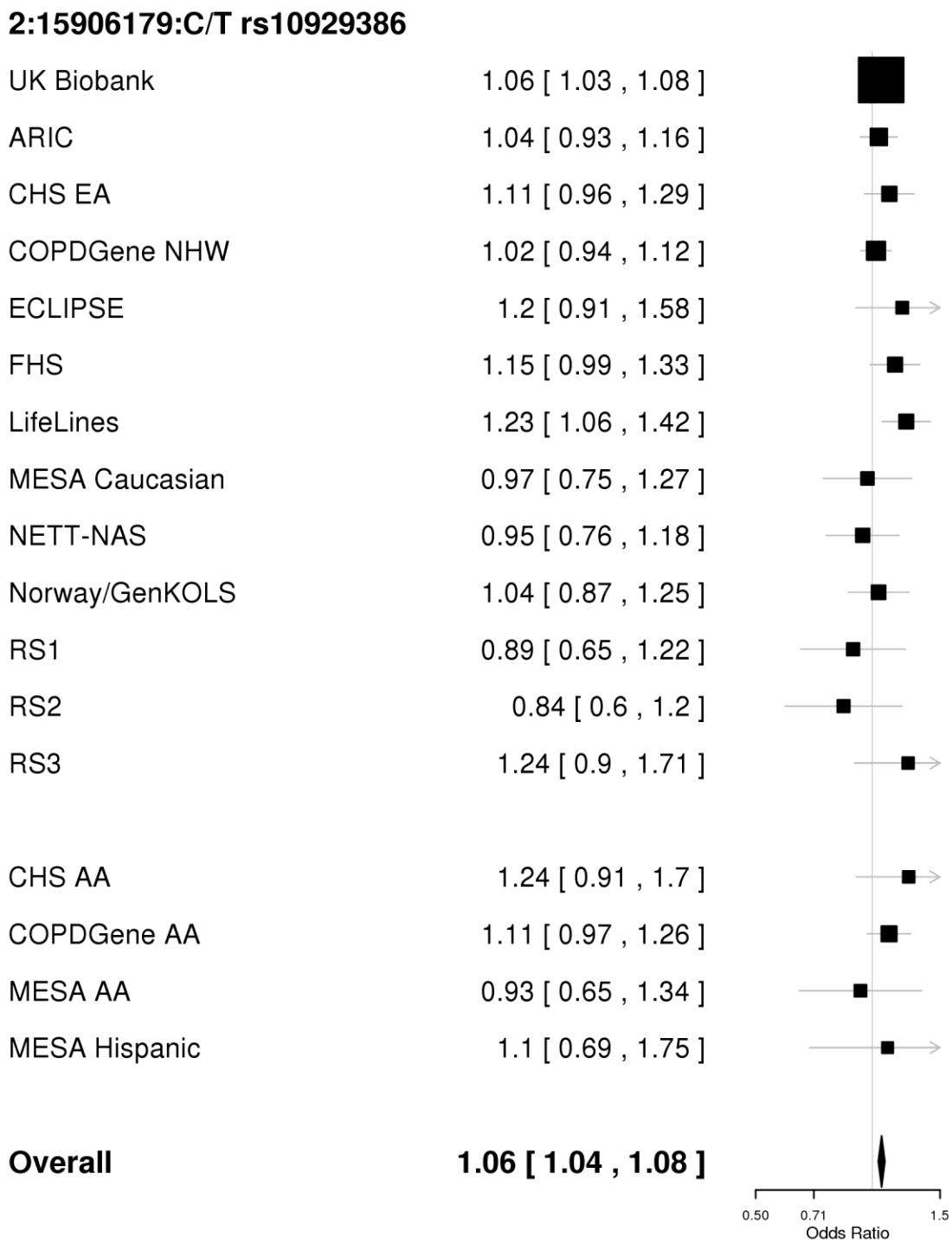
Supplementary Figure 1-8: Forest plot for rs11579382 (*CHRM3* locus at 1q43)



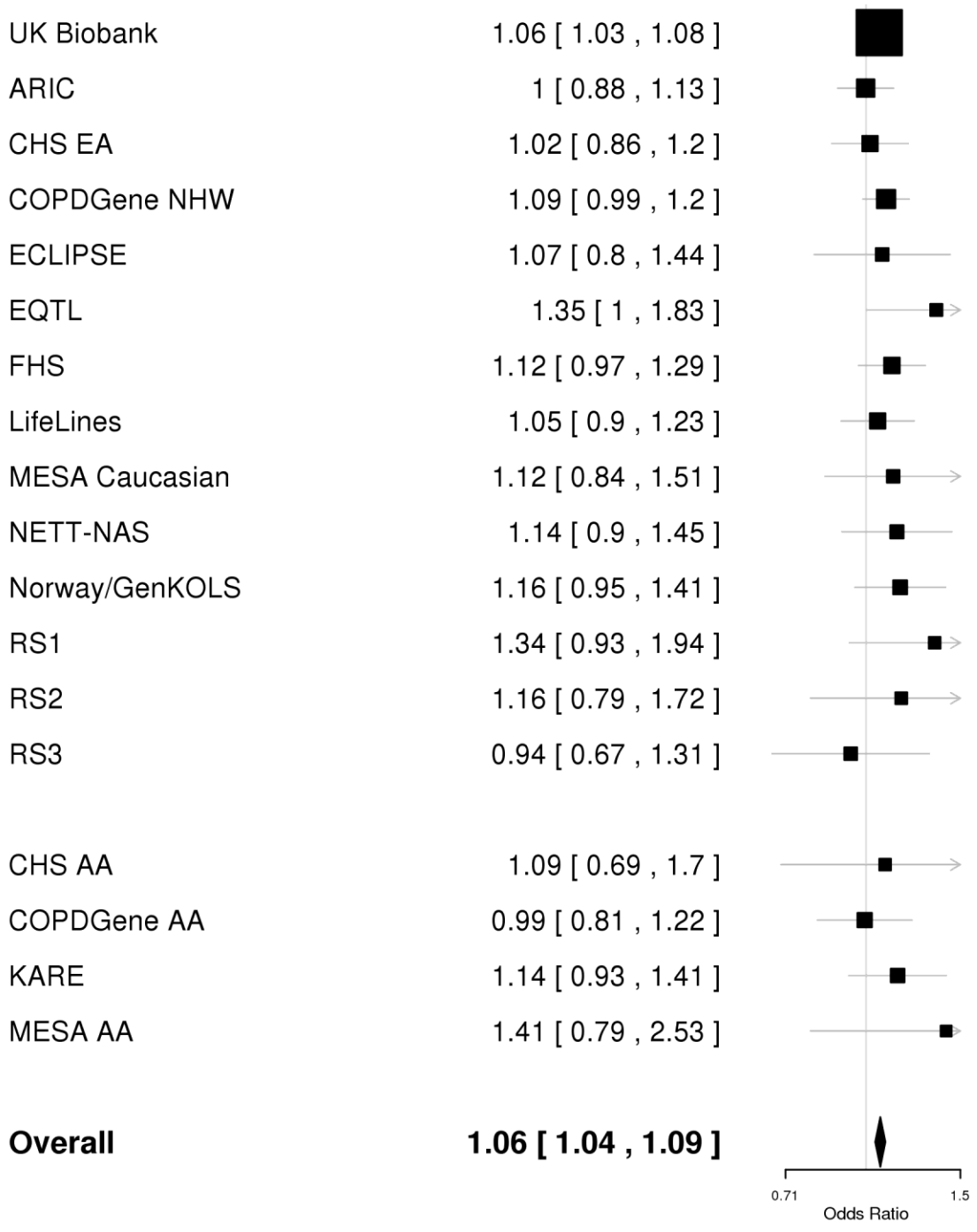
Supplementary Figure 1-9: Forest plot for rs955277 (*ASAP2* locus at 2p25.1)



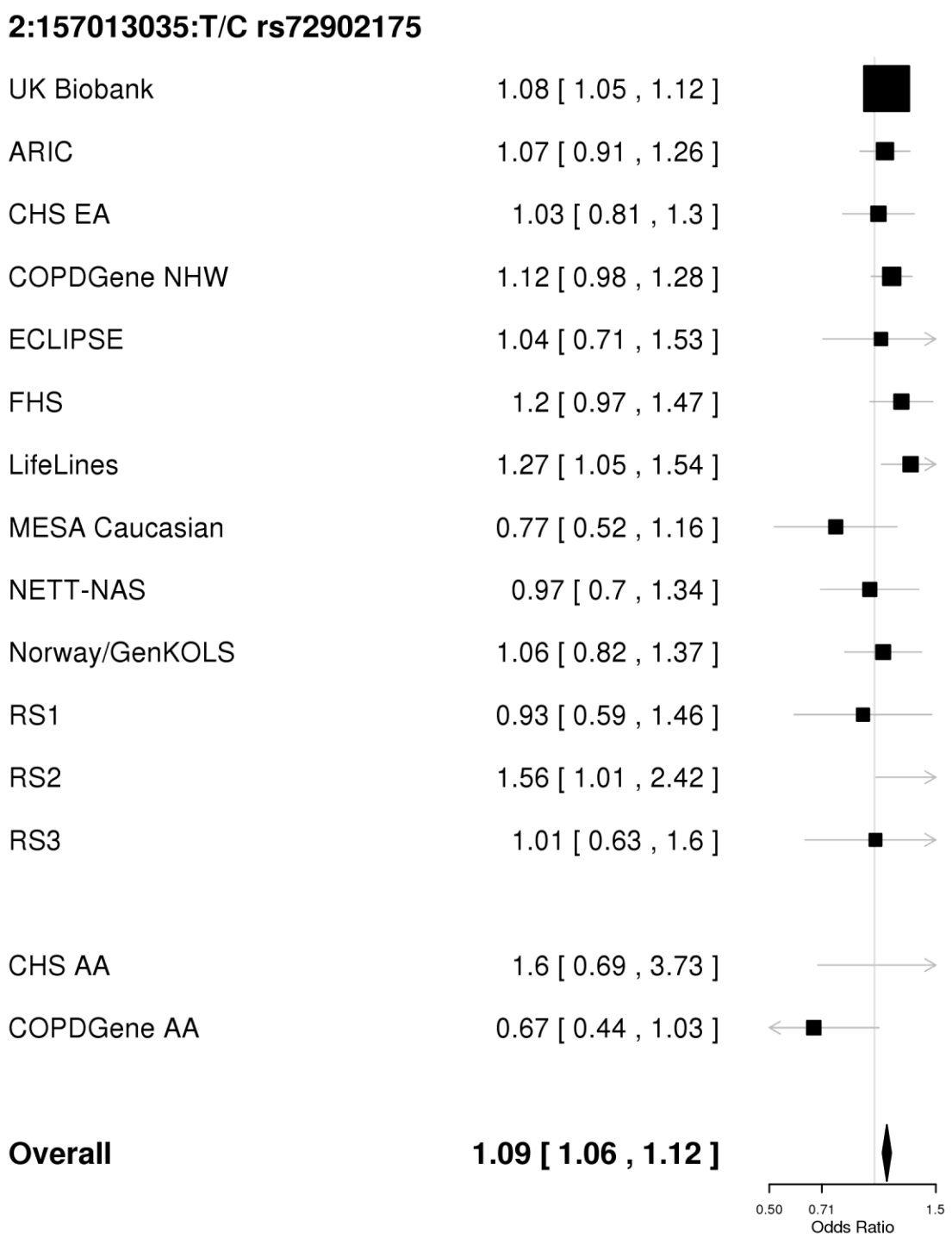
Supplementary Figure 1-10: Forest plot for rs10929386 (*DDX1* locus at 2p24.3)



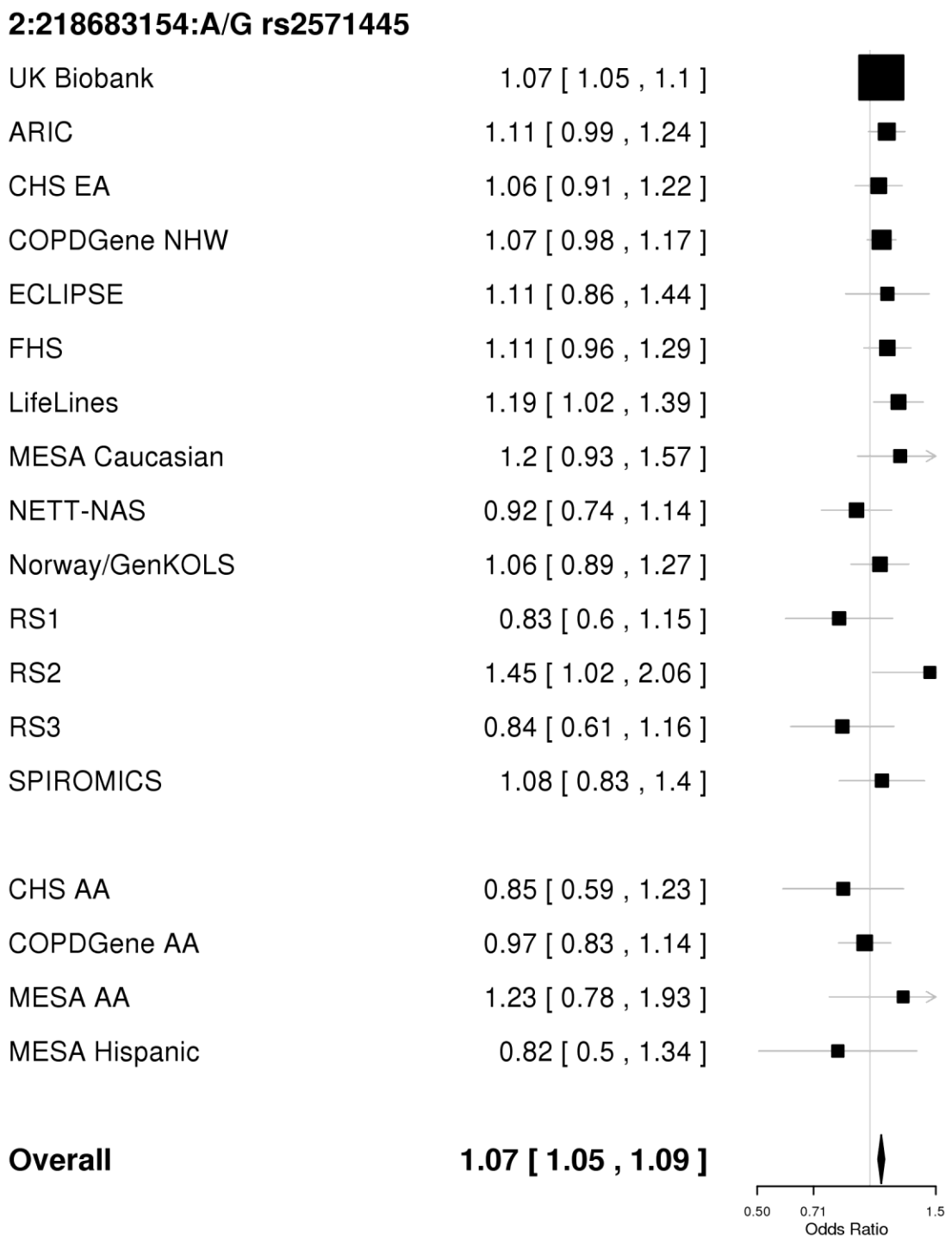
Supplementary Figure 1-11: Forest plot for rs12466981 (*EML4* locus at 2p21)



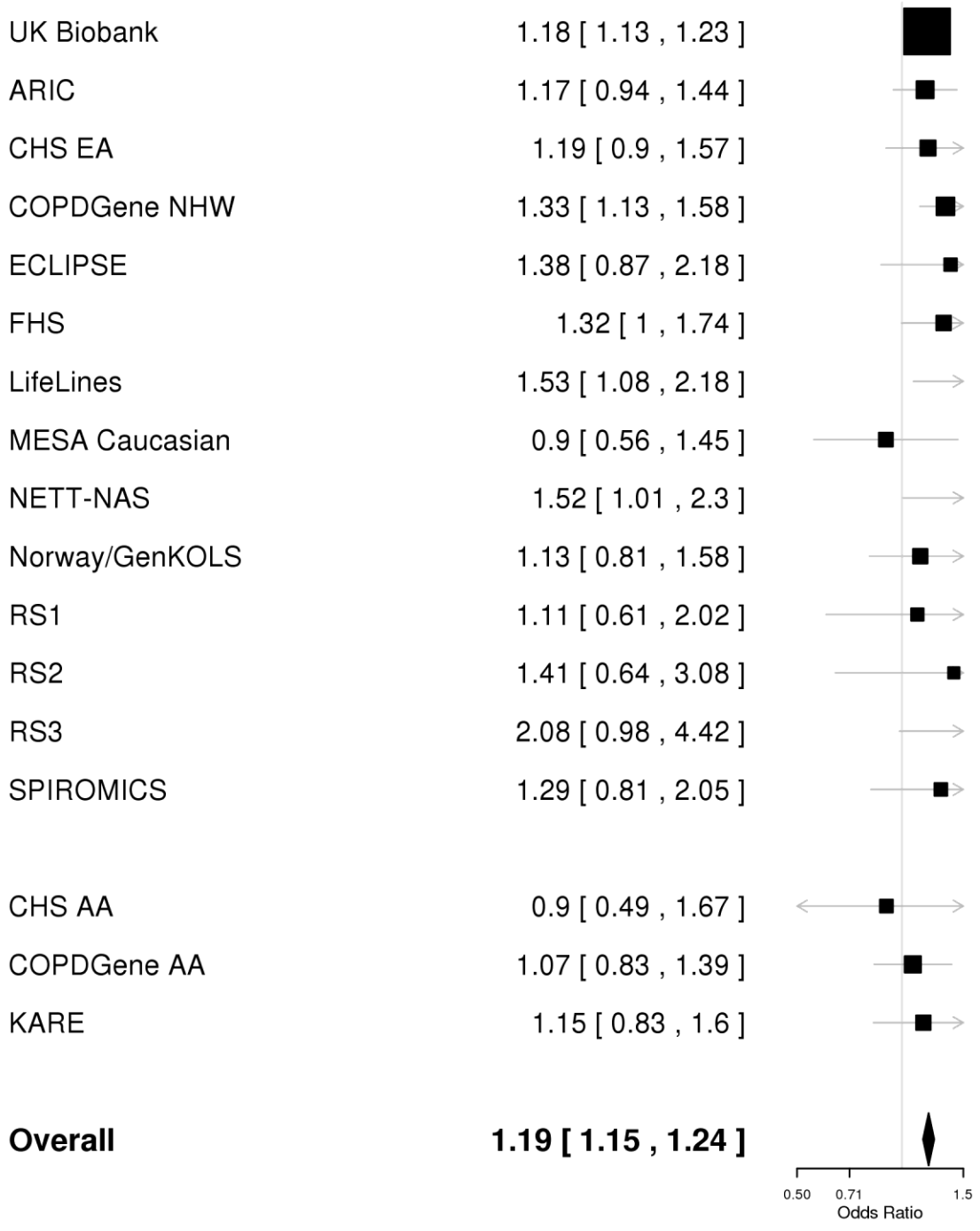
Supplementary Figure 1-12: Forest plot for rs72902175 (*NR4A2* locus at 2q24.1)



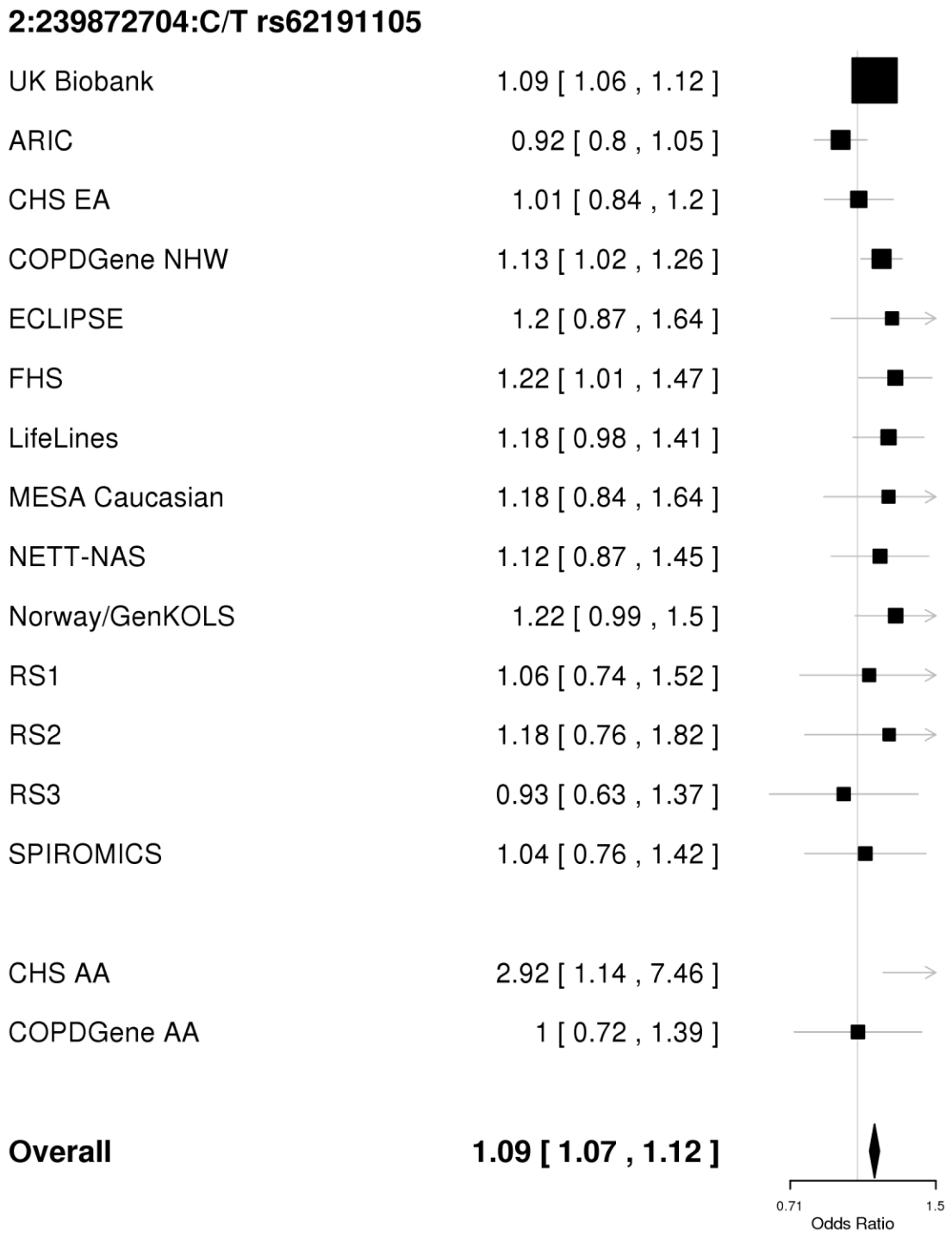
Supplementary Figure 1-13: Forest plot for rs2571445 (*TNS1* locus at 2q35)



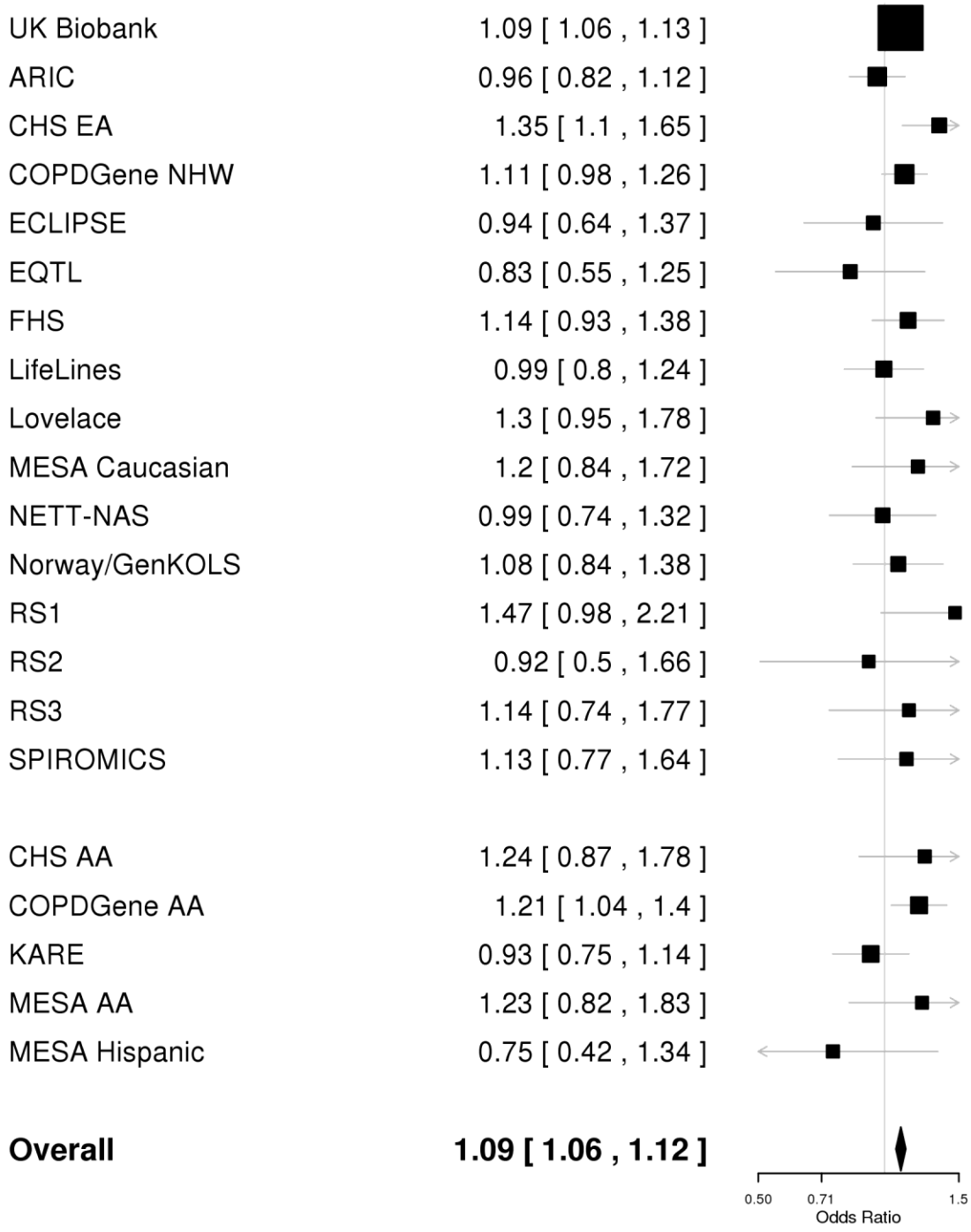
Supplementary Figure 1-14: Forest plot for rs16825267 (*PID1* locus at 2q36.3)



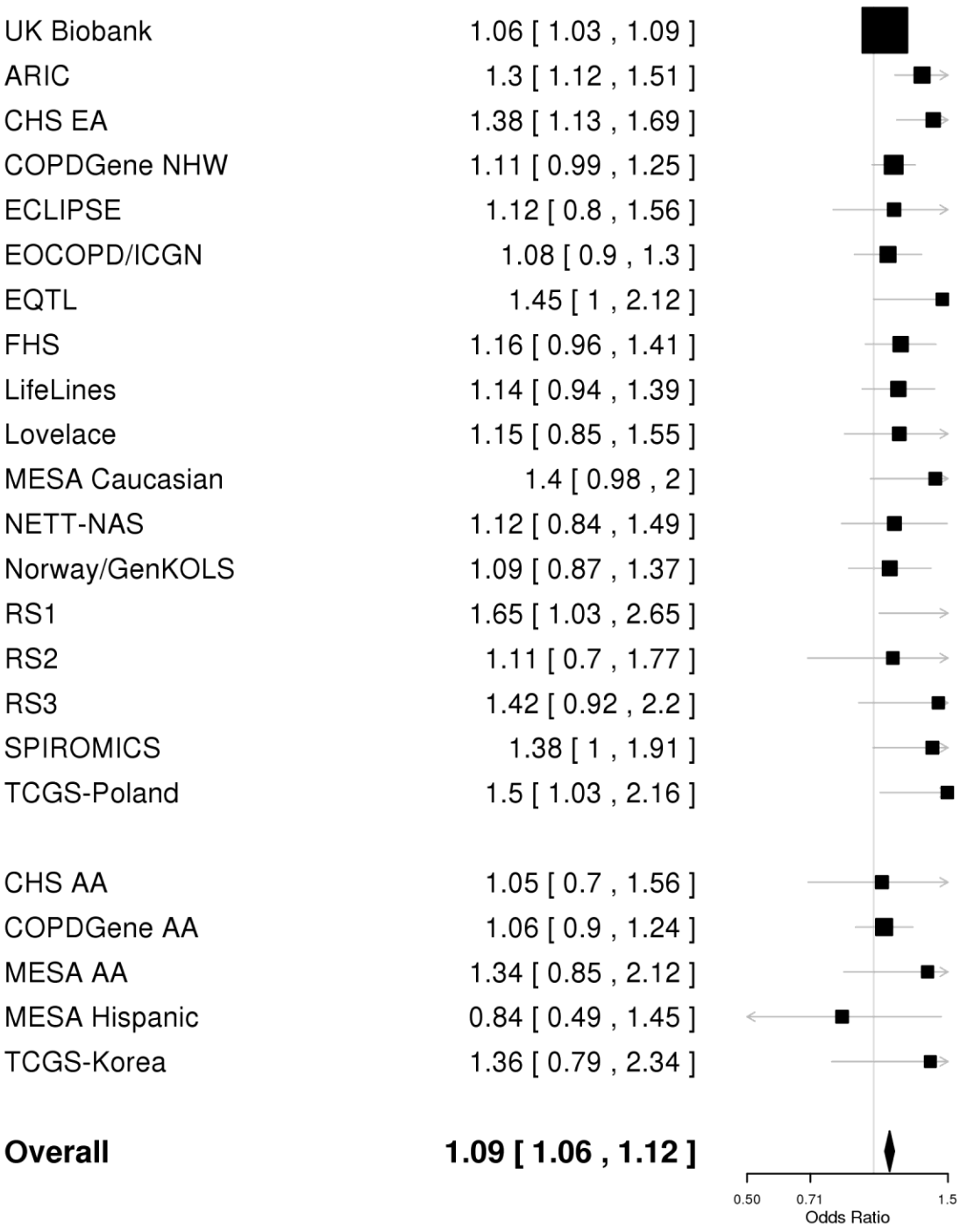
Supplementary Figure 1-15: Forest plot for rs62191105 (*TWIST2* locus at 2q37.3)



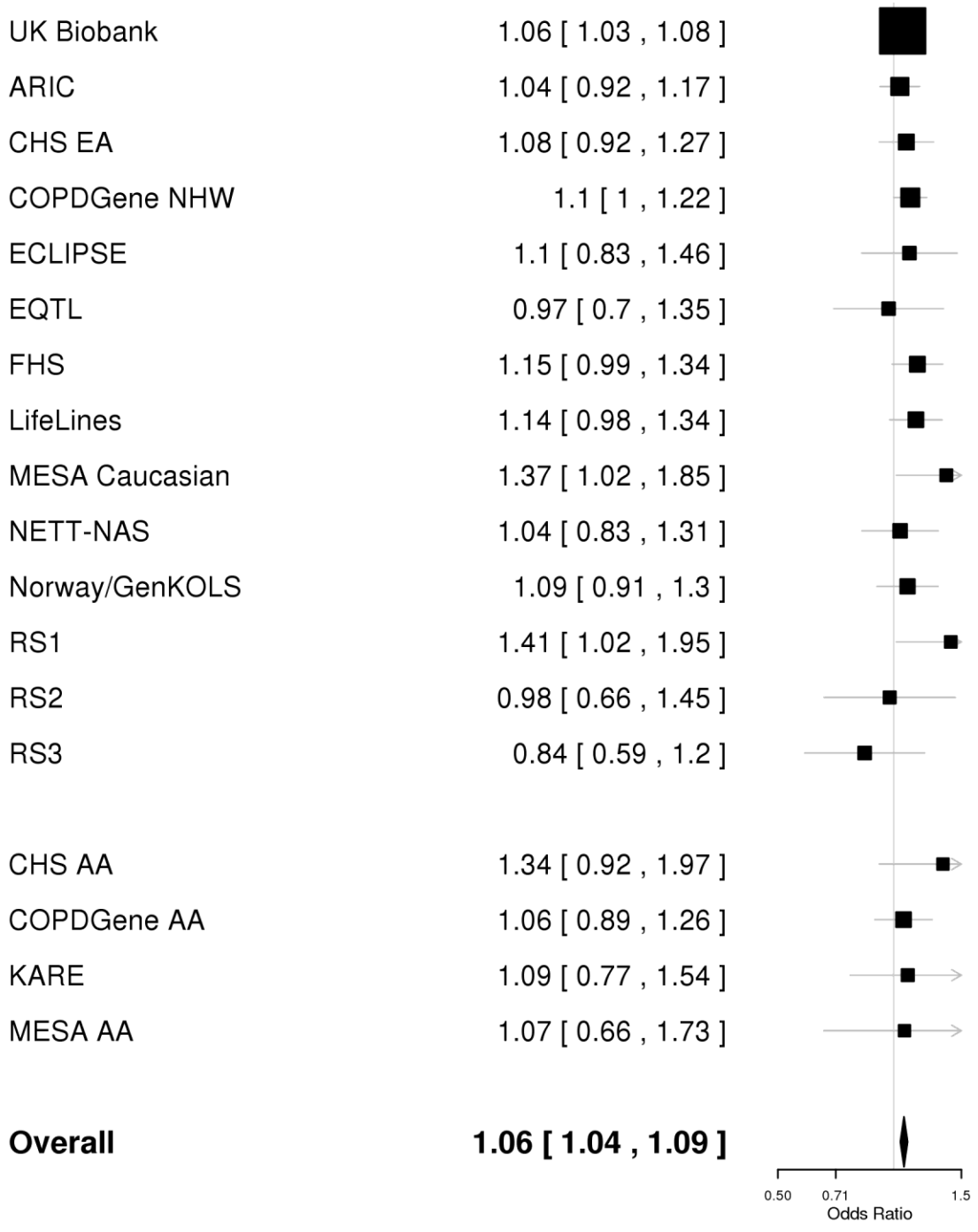
Supplementary Figure 1-16: Forest plot for rs2442776 (*VGLL4* locus at 3p25.3)



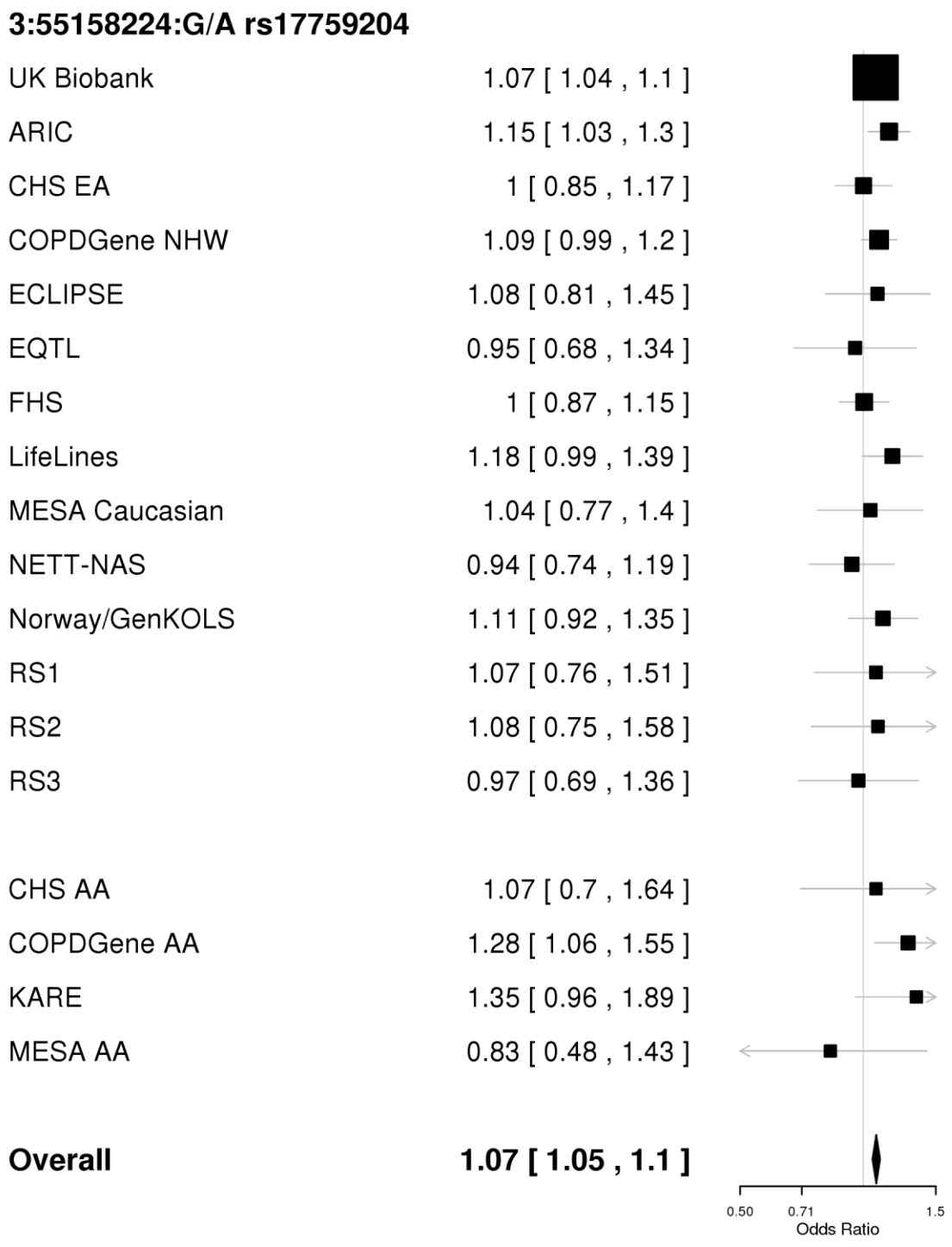
Supplementary Figure 1-17: Forest plot for rs1529672 (*RARB* locus at 3p24.2)



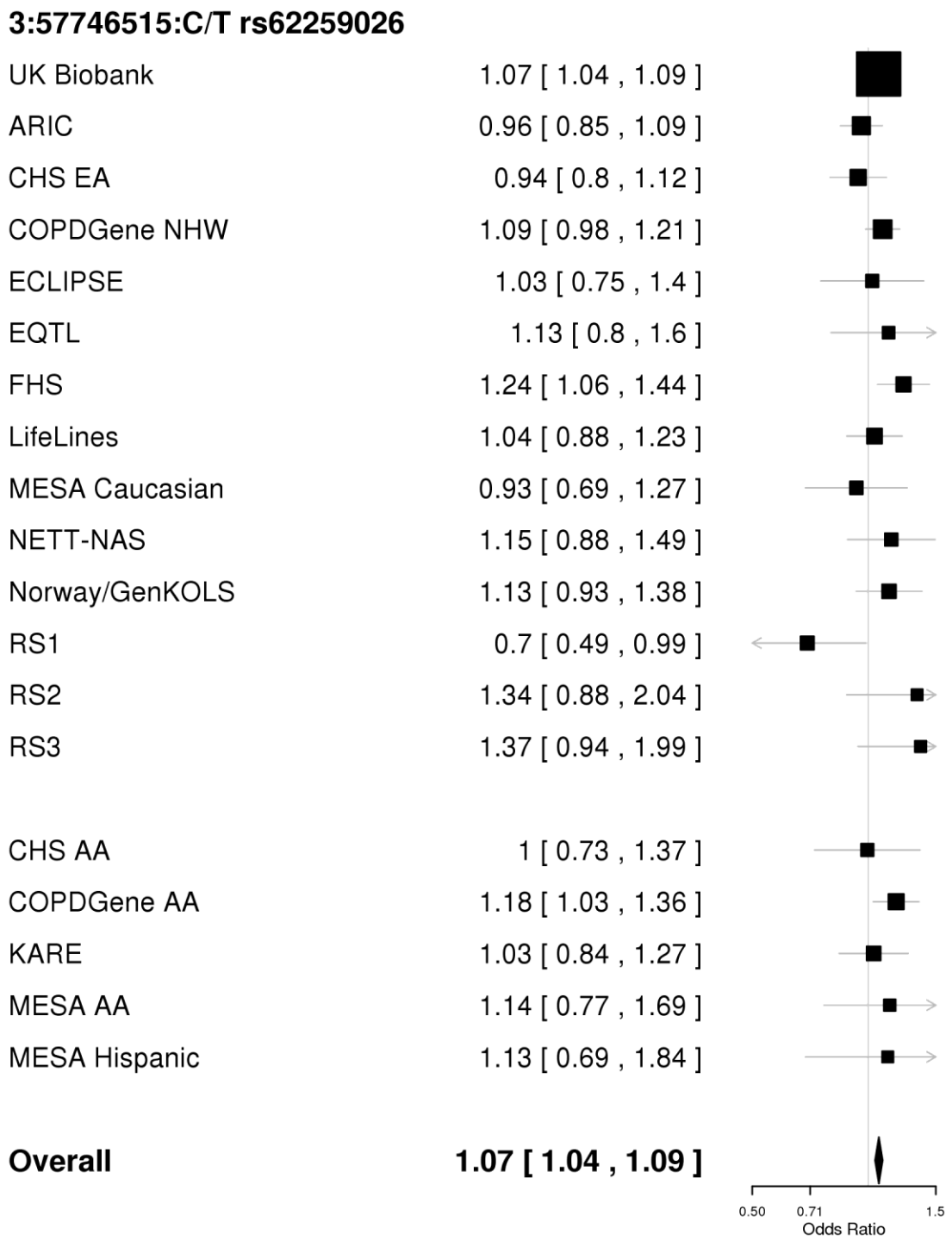
Supplementary Figure 1-18: Forest plot for rs13073544 (*RBMS3* locus at 3p24.1)



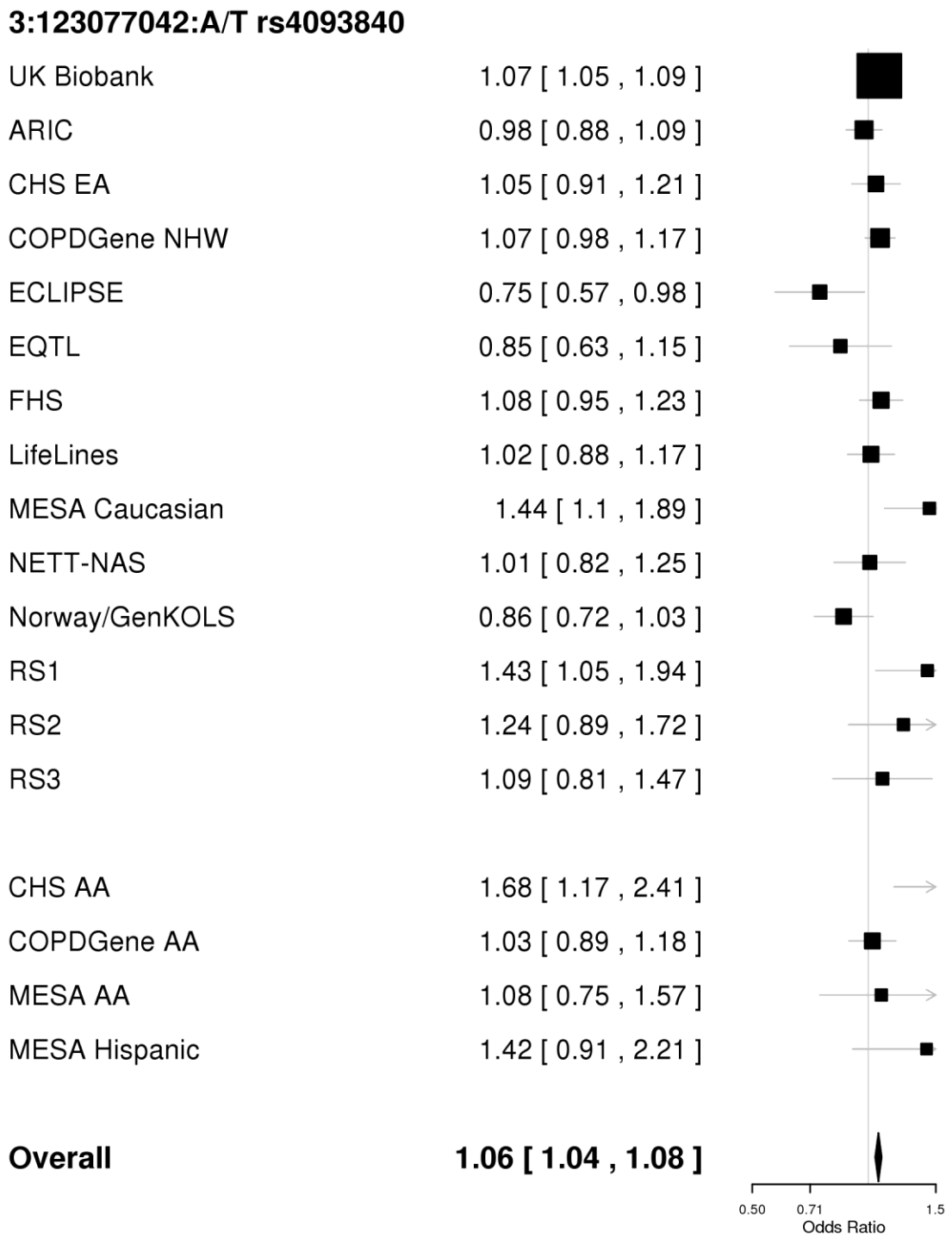
Supplementary Figure 1-19: Forest plot for rs17759204 (*CACNA2D3* locus at 3p14.3)



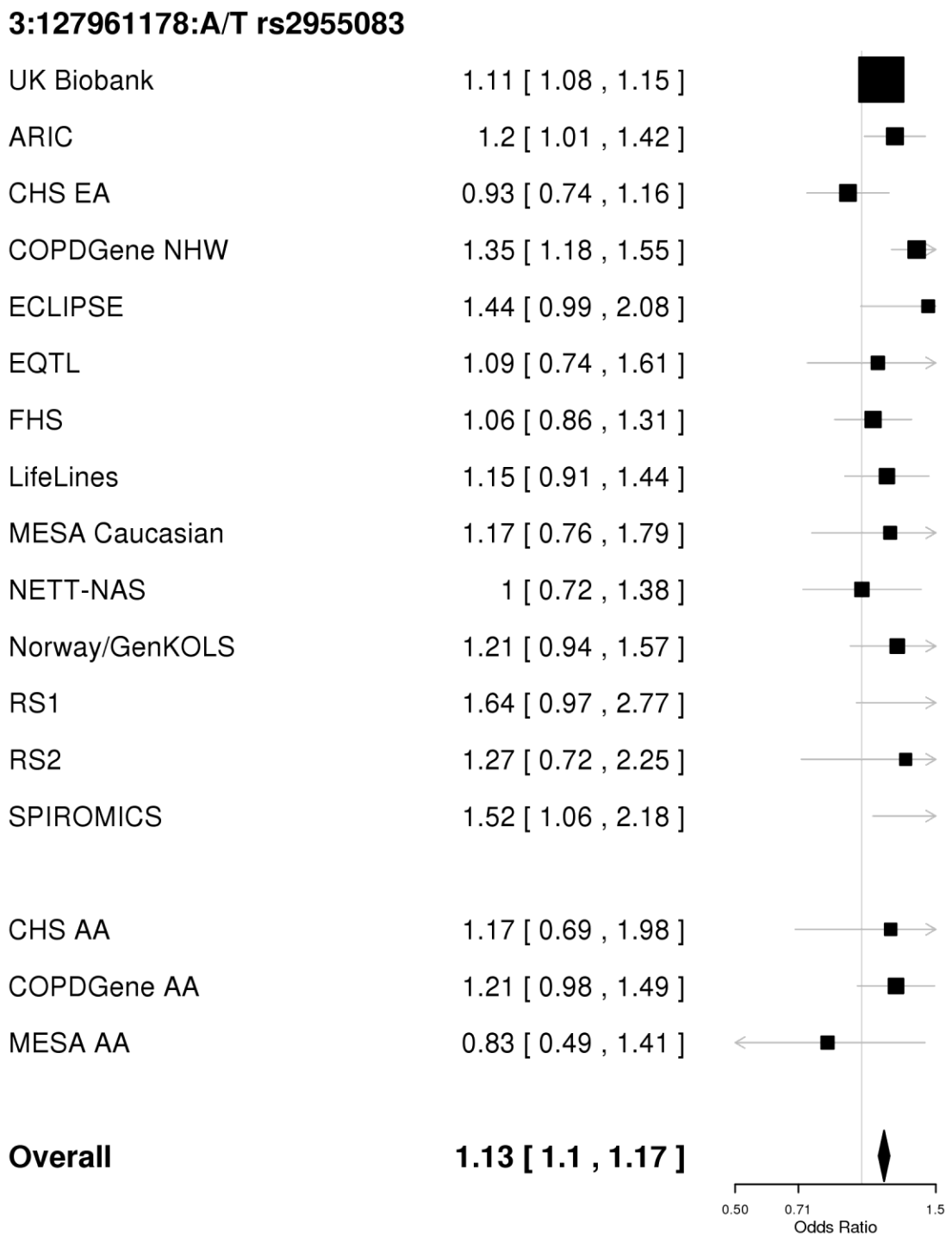
Supplementary Figure 1-20: Forest plot for rs62259026 (*SLMAP* locus at 3p14.3)



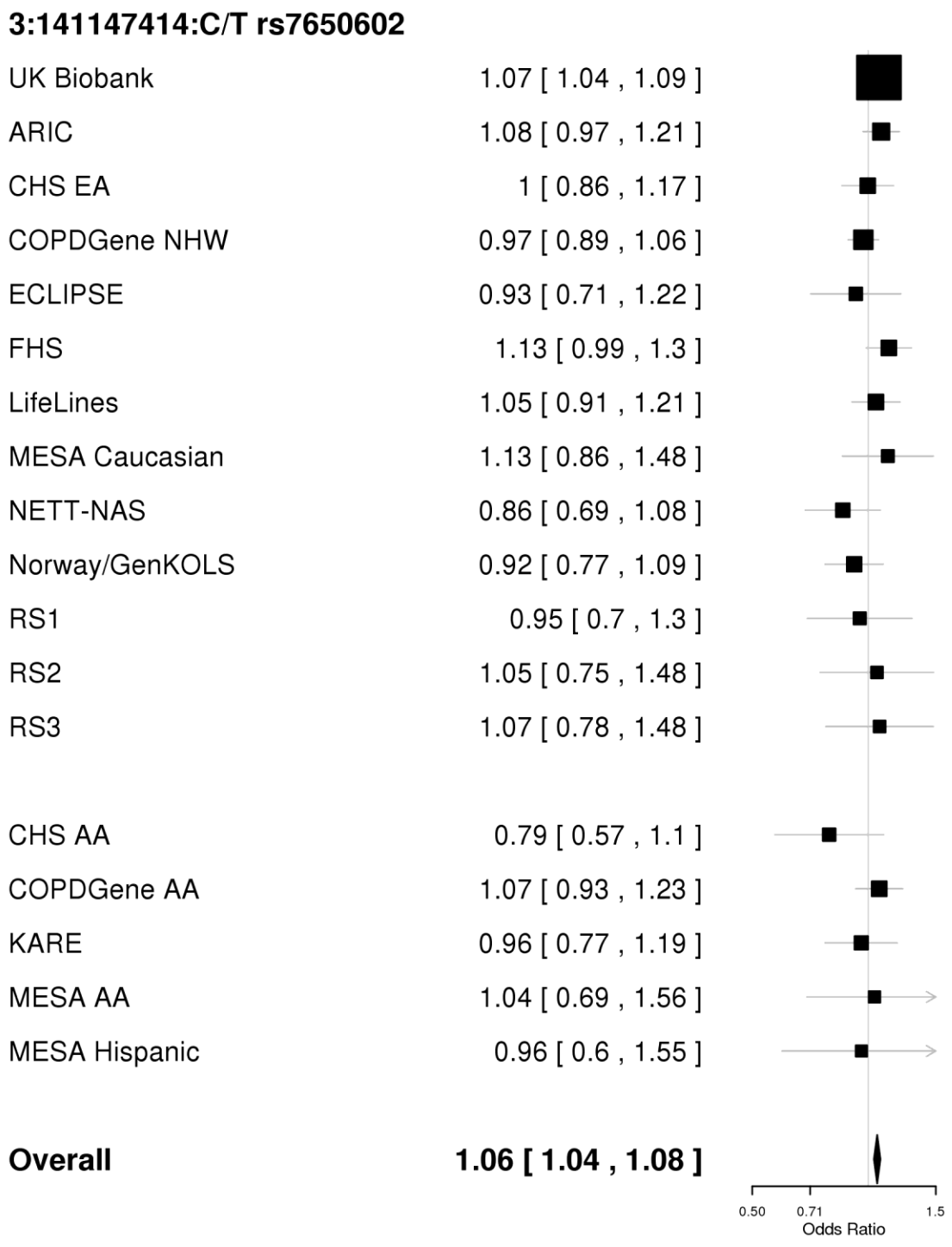
Supplementary Figure 1-21: Forest plot for rs4093840 (*ADCY5* locus at 3q21.1)



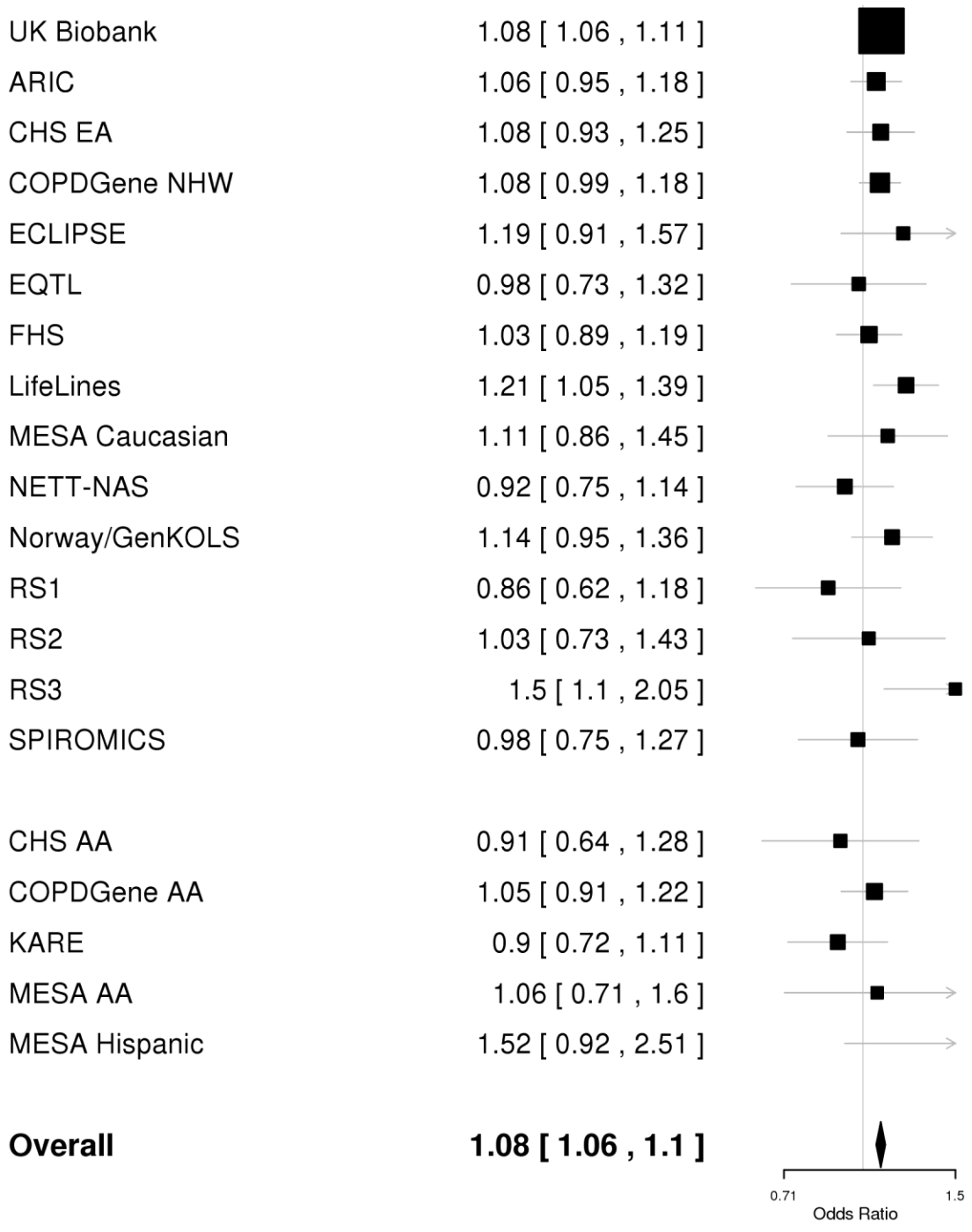
Supplementary Figure 1-22: Forest plot for rs2955083 (*EEFSEC* locus at 3q21.3)



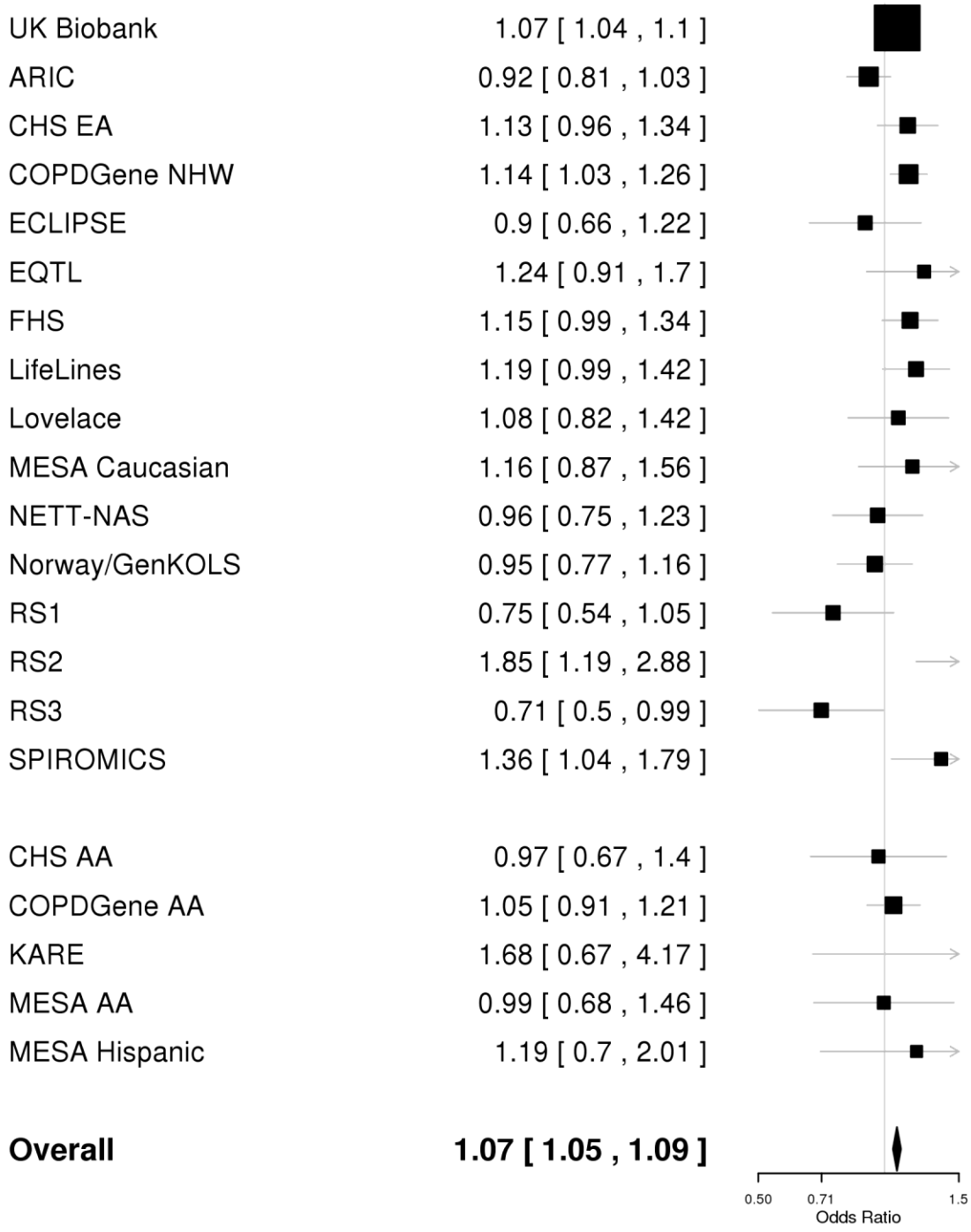
Supplementary Figure 1-23: Forest plot for rs7650602 (*ZBTB38* locus at 3q23)



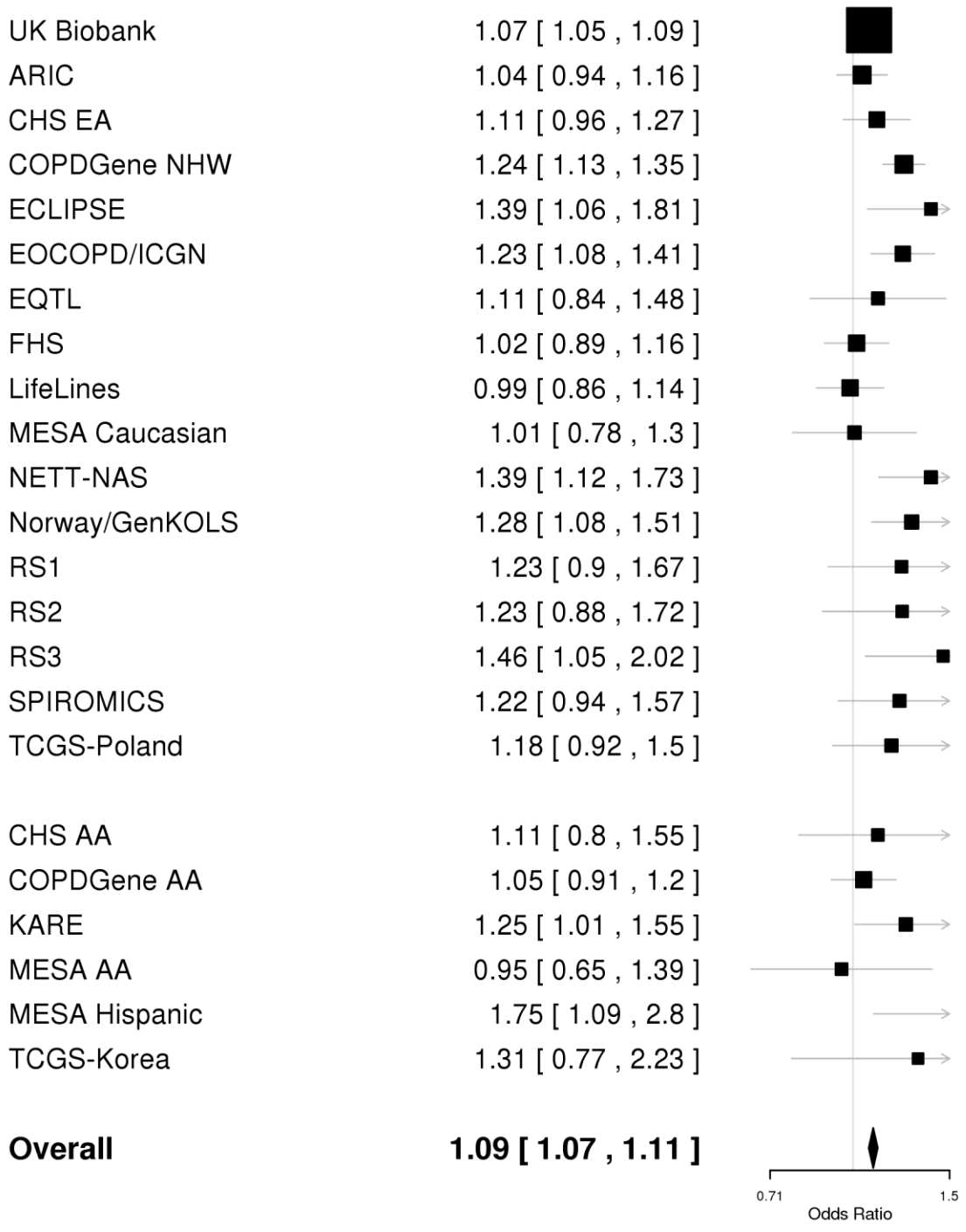
Supplementary Figure 1-24: Forest plot for rs7642001 (*MECOM* locus at 3q26.2)



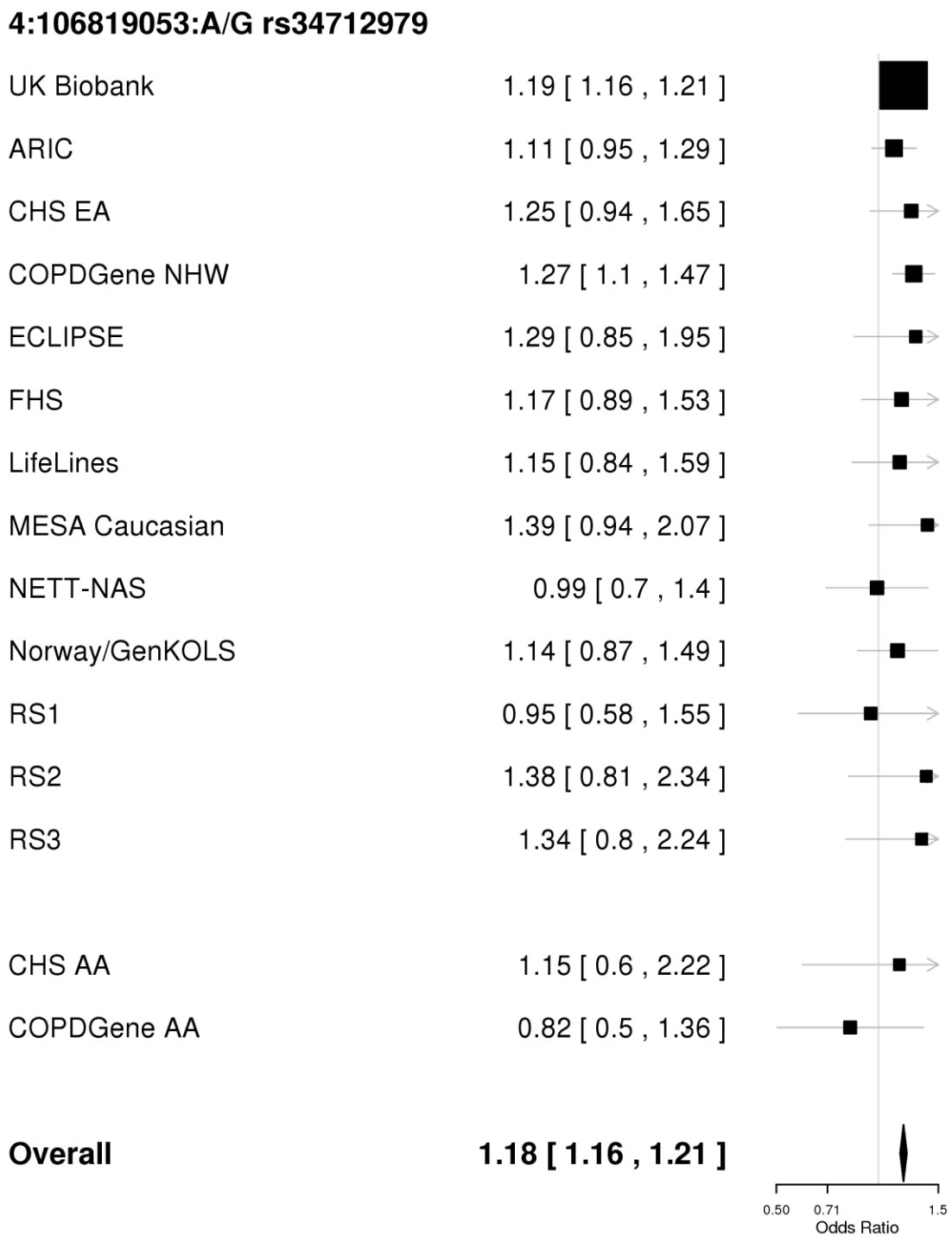
Supplementary Figure 1-25: Forest plot for rs4585380 (*BTC* locus at 4q13.3)



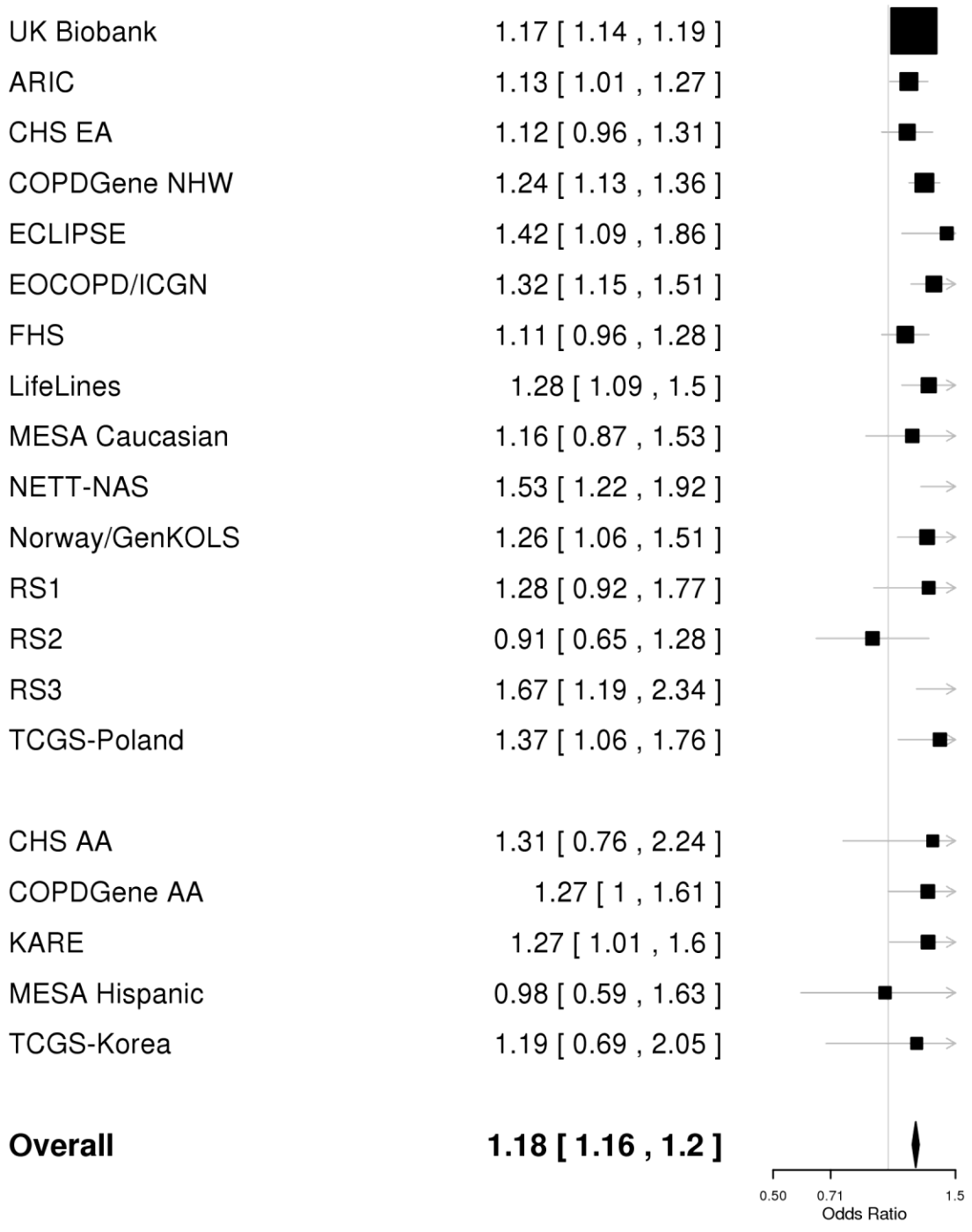
Supplementary Figure 1-26: Forest plot for rs7671261 (*FAM13A* locus at 4q22.1)



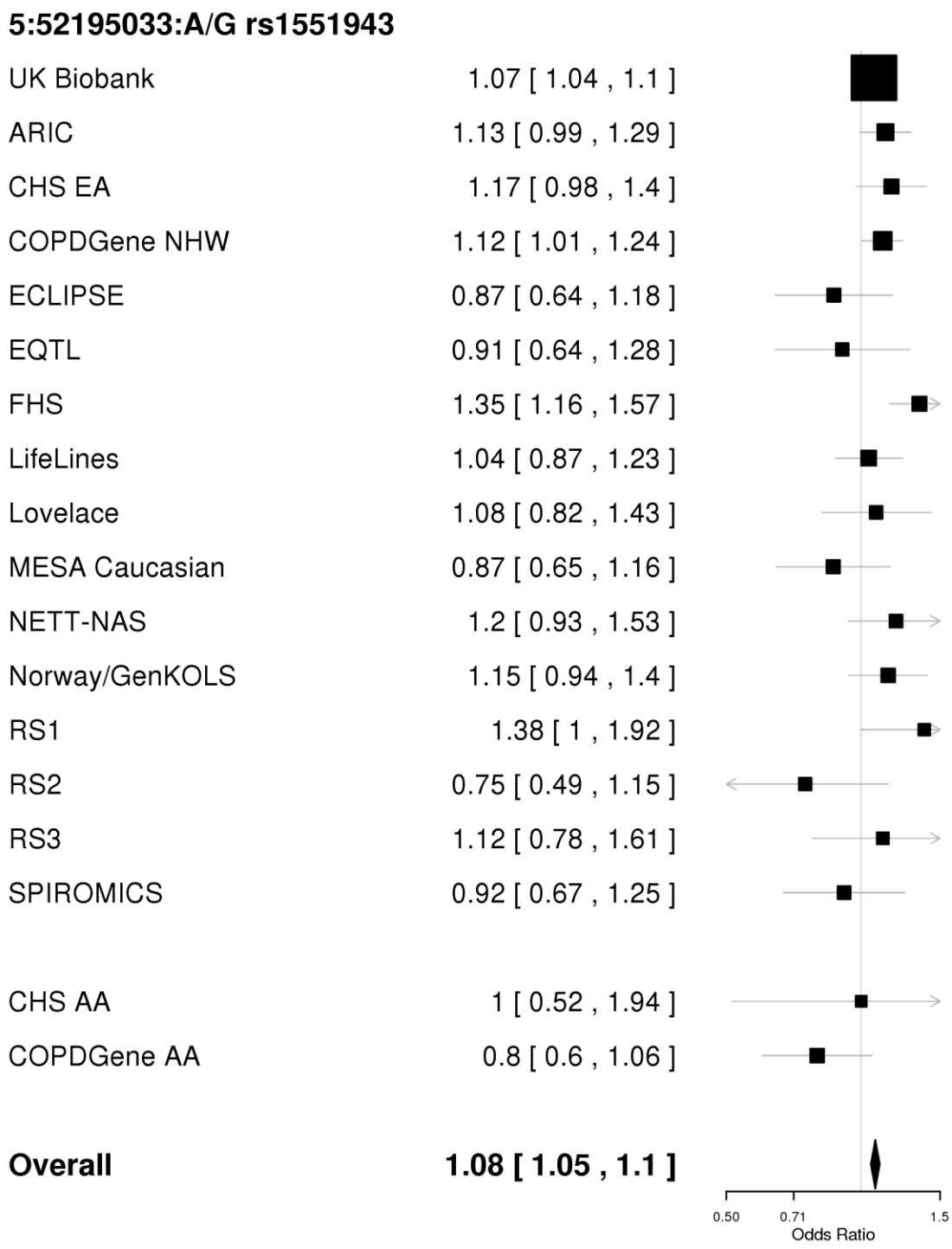
Supplementary Figure 1-27: Forest plot for rs34712979 (*NPNT* locus at 4q24)



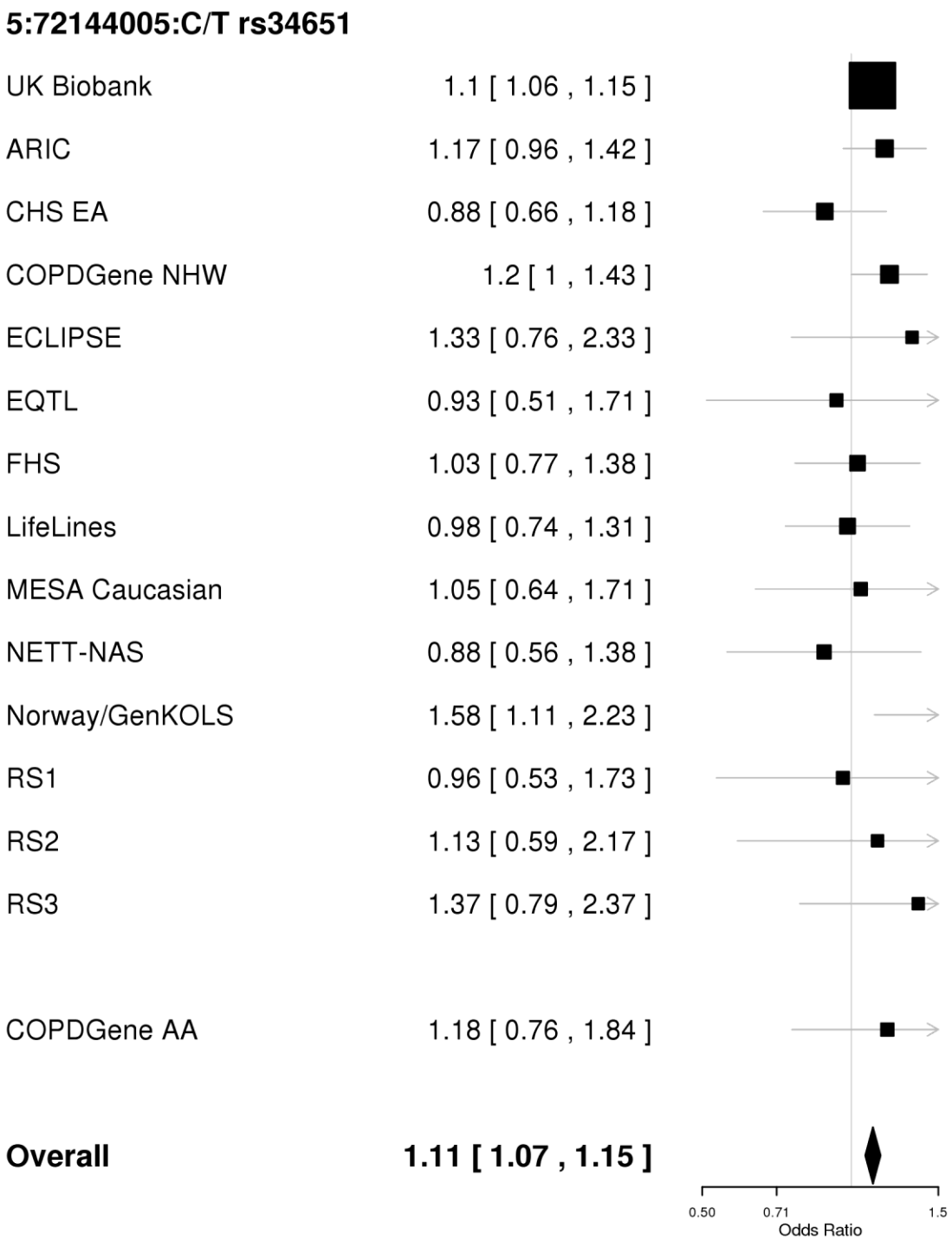
Supplementary Figure 1-28: Forest plot for rs13140176 (*HHIP* locus at 4q31.21)



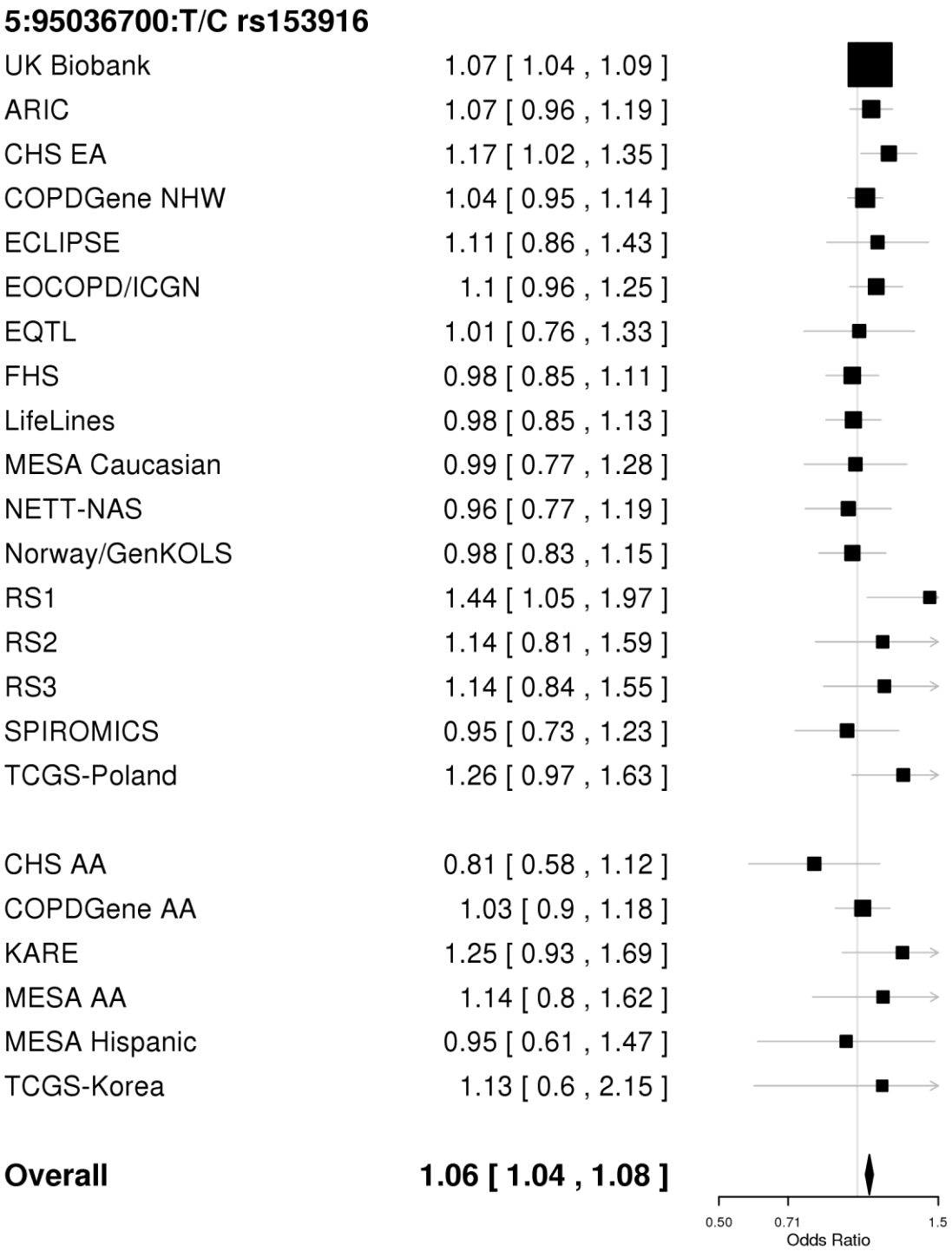
Supplementary Figure 1-29: Forest plot for rs1551943 (*ITGA1* locus at 5q11.2)



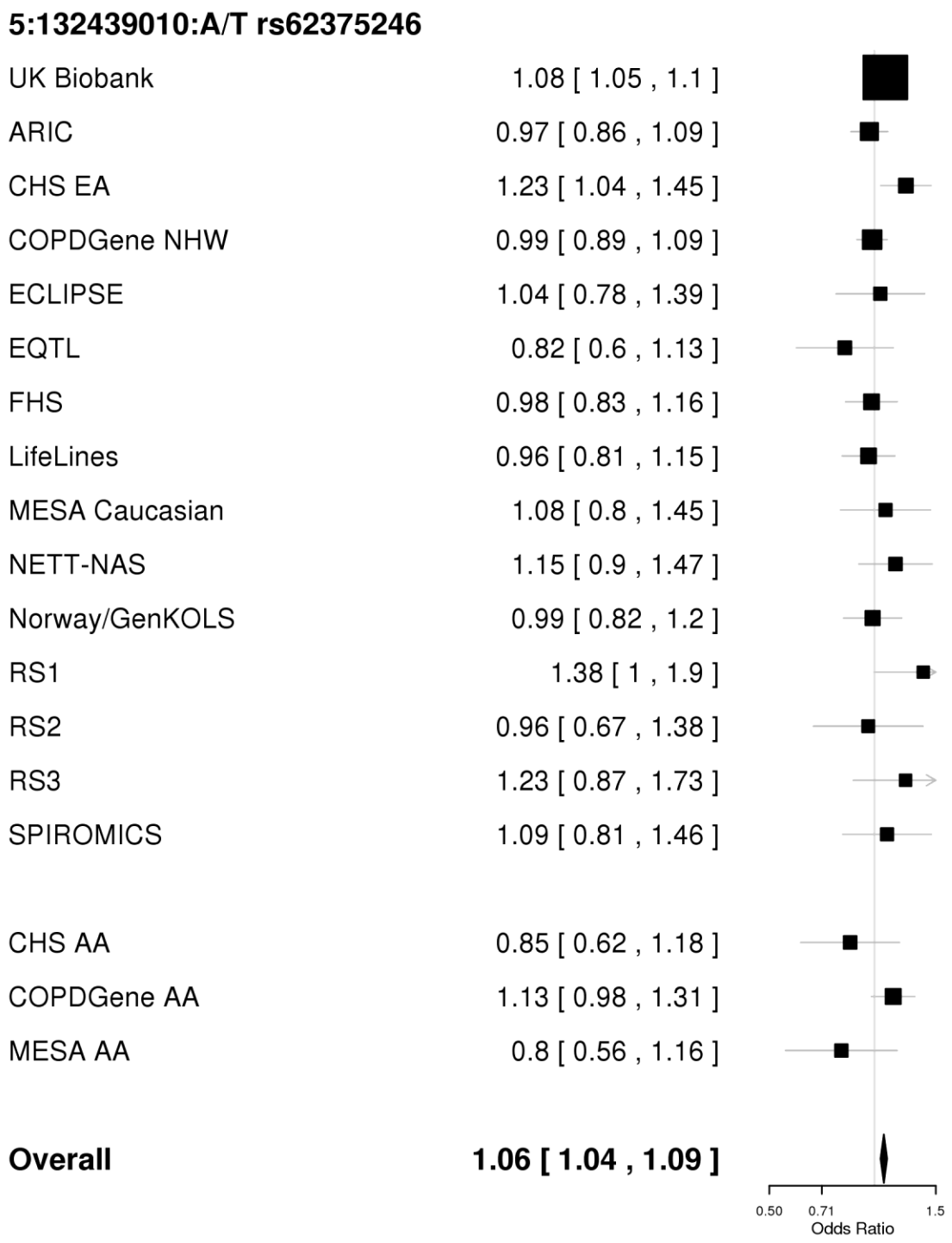
Supplementary Figure 1-30: Forest plot for rs34651 (*TNPO1* locus at 5q13.2)



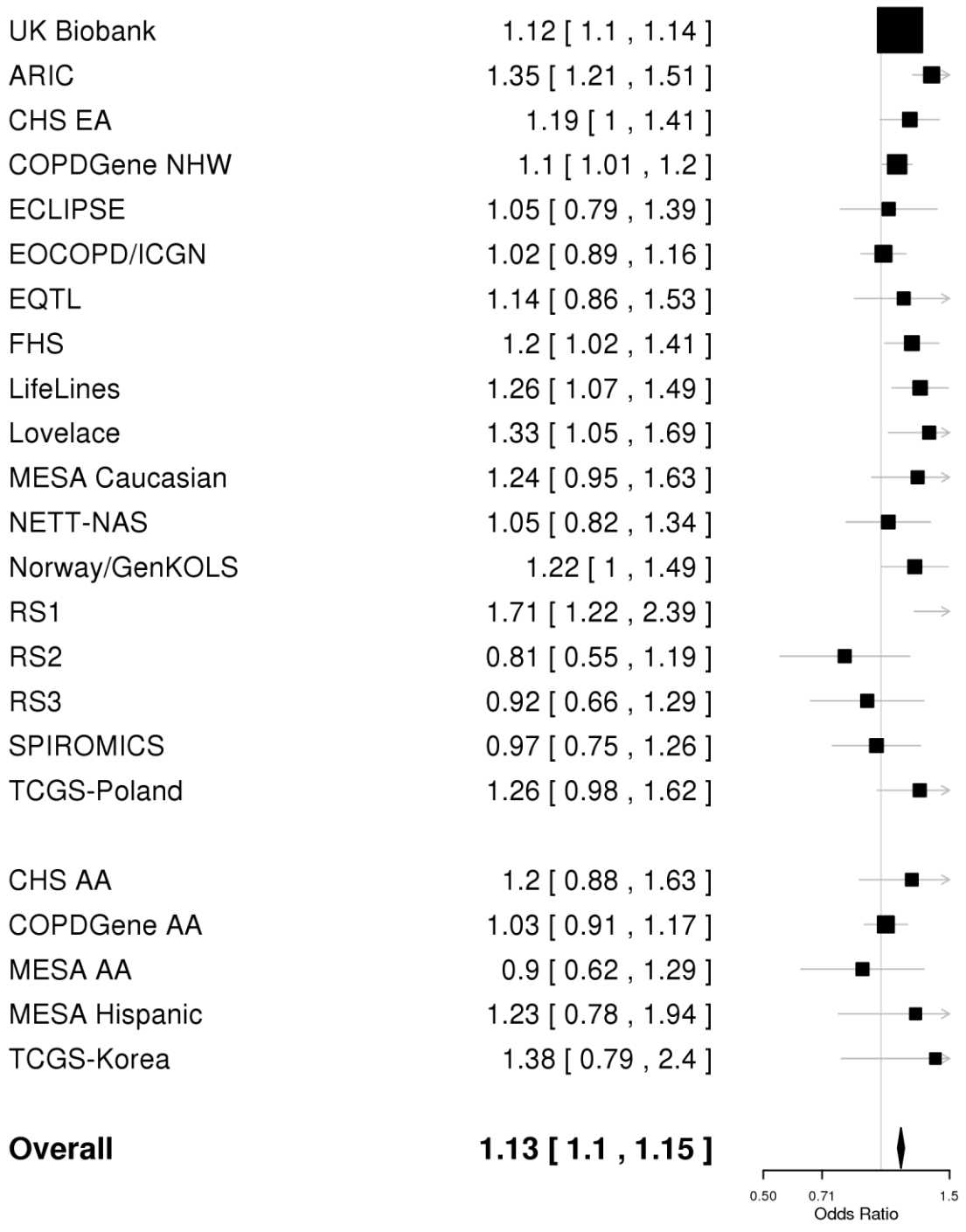
Supplementary Figure 1-31: Forest plot for rs153916 (*SPATA9* locus at 5q15)



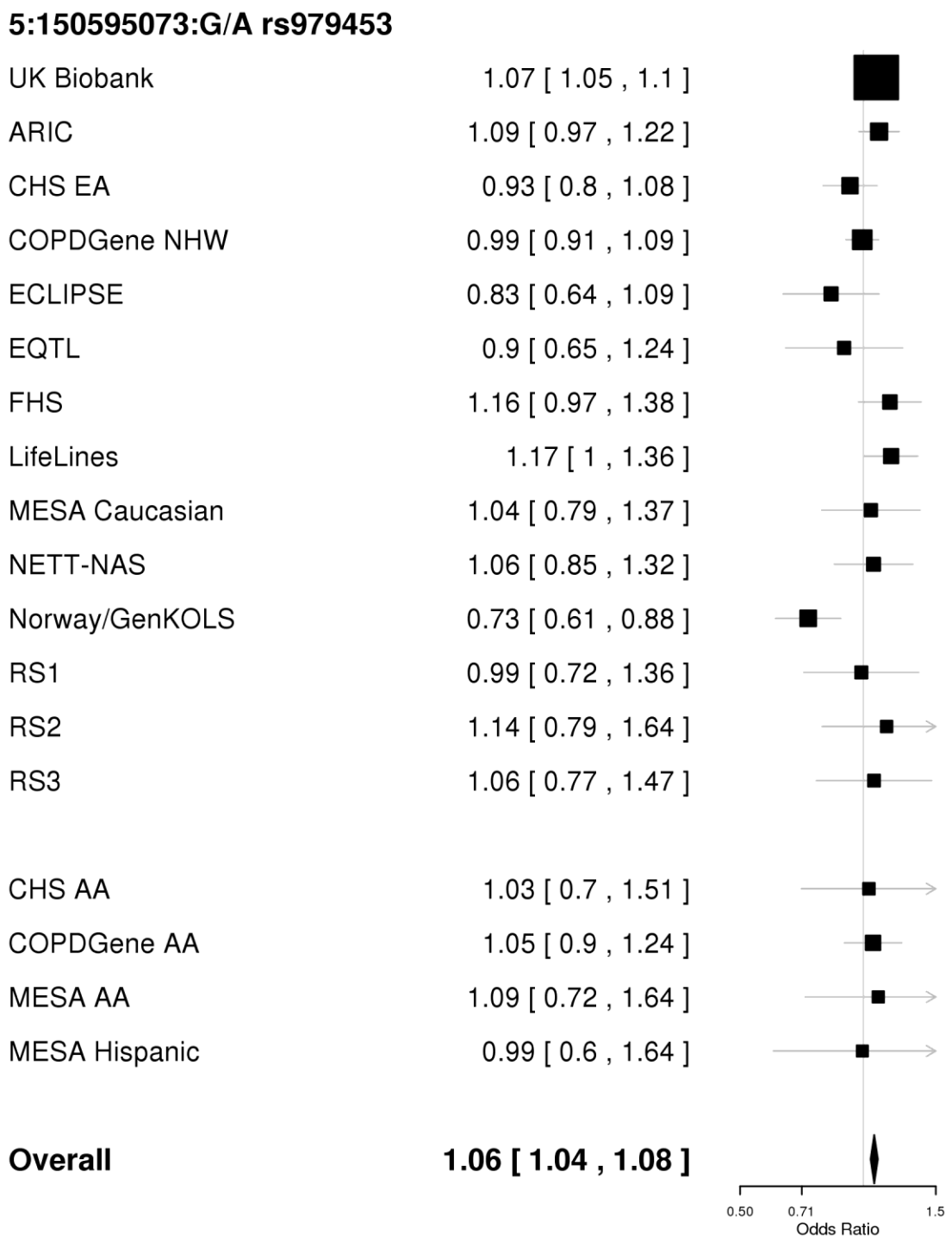
Supplementary Figure 1-32: Forest plot for rs62375246 (*HSPA4* locus at 5q31.1)



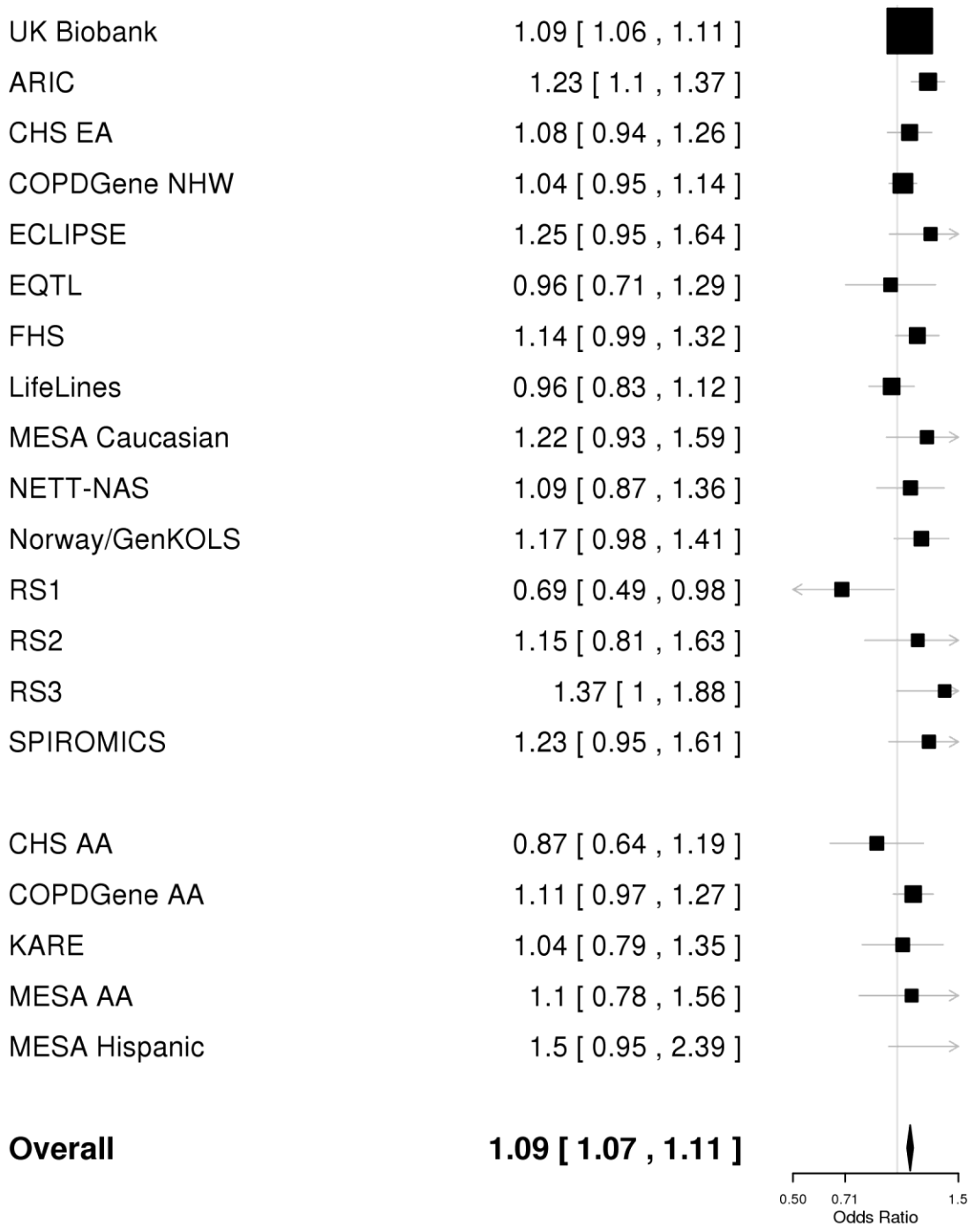
Supplementary Figure 1-33: Forest plot for rs10037493 (*HTR4* locus at 5q32)



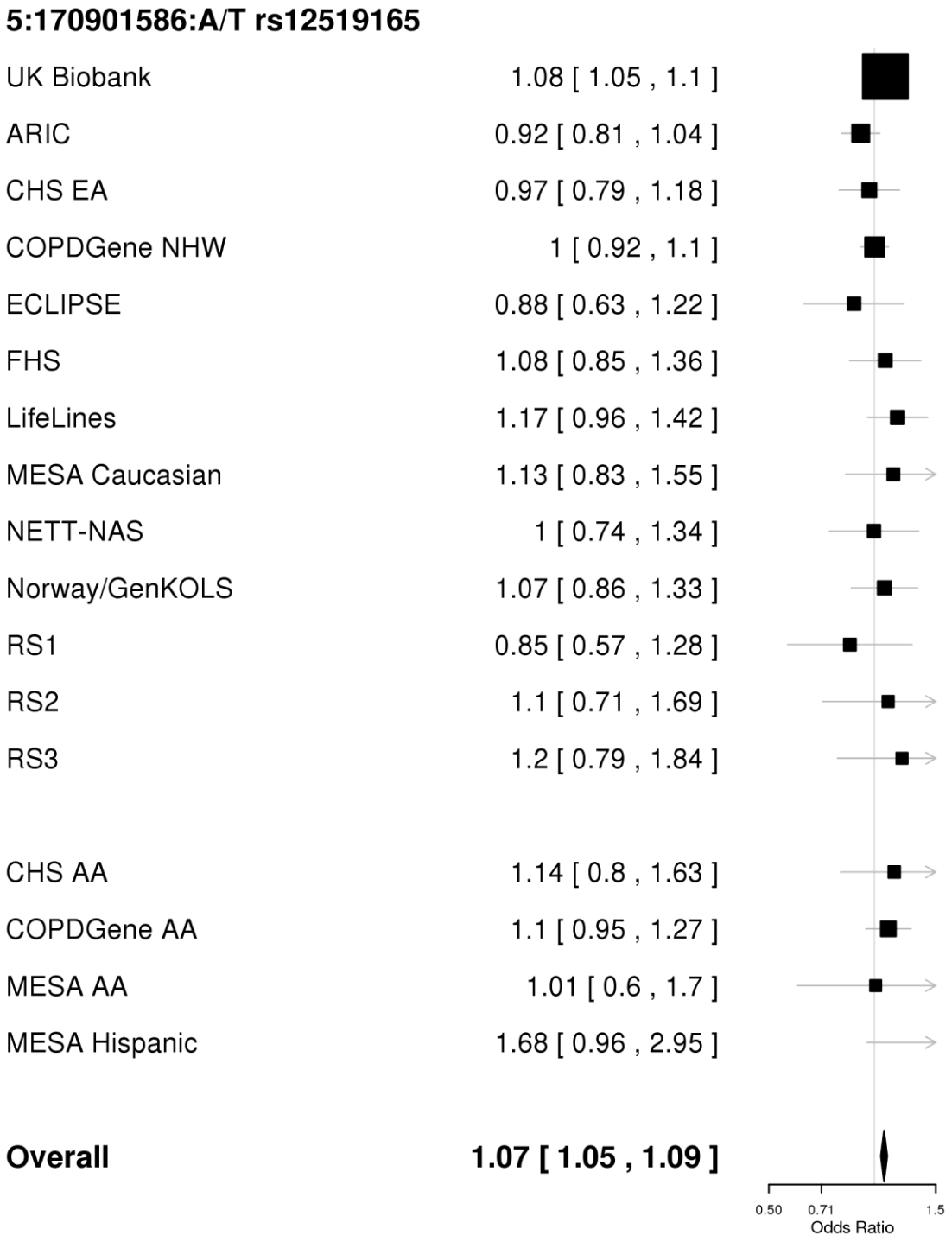
Supplementary Figure 1-34: Forest plot for rs979453 (*CCDC69* locus at 5q33.1)



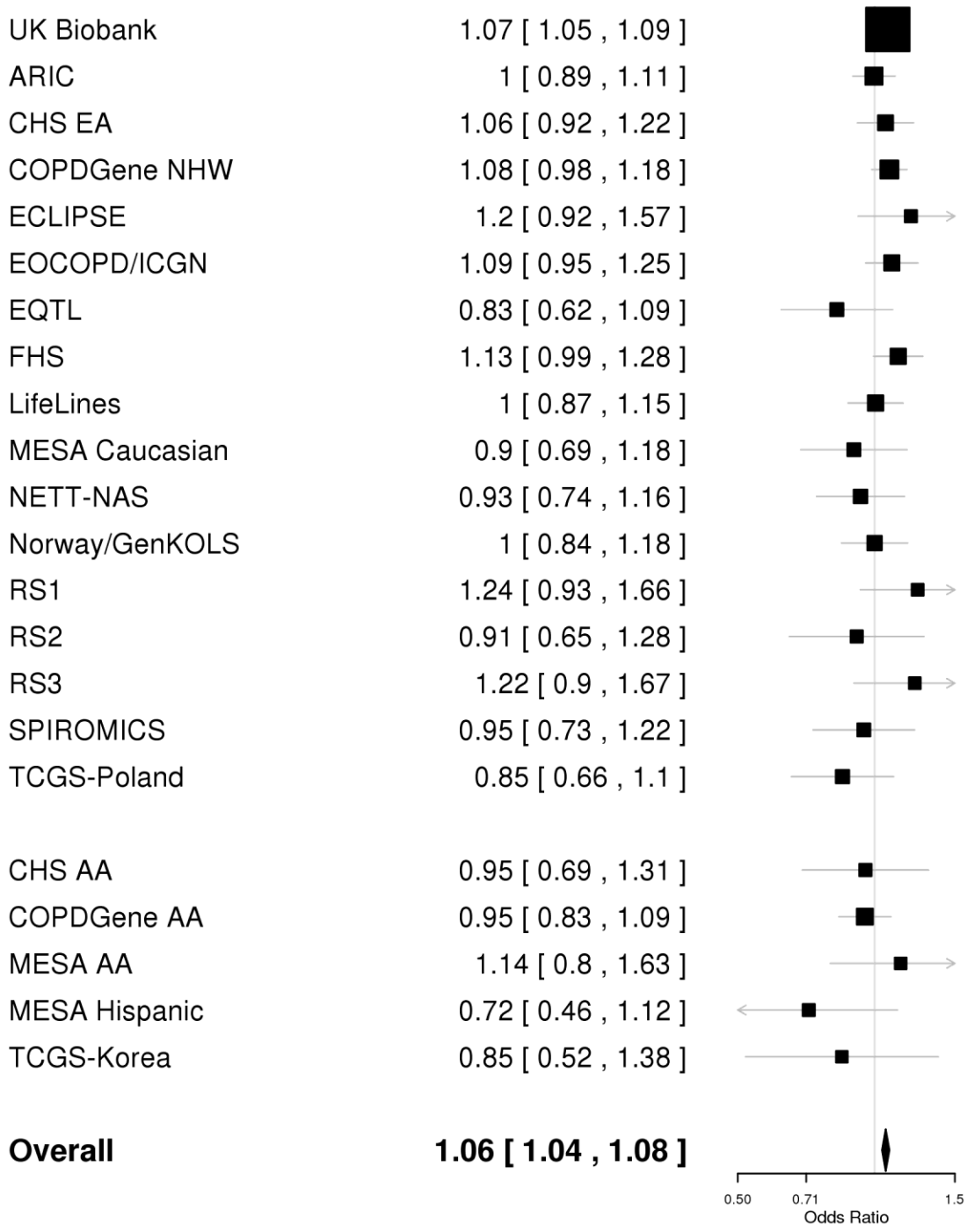
Supplementary Figure 1-35: Forest plot for rs10866659 (*ADAM19* locus at 5q33.3)



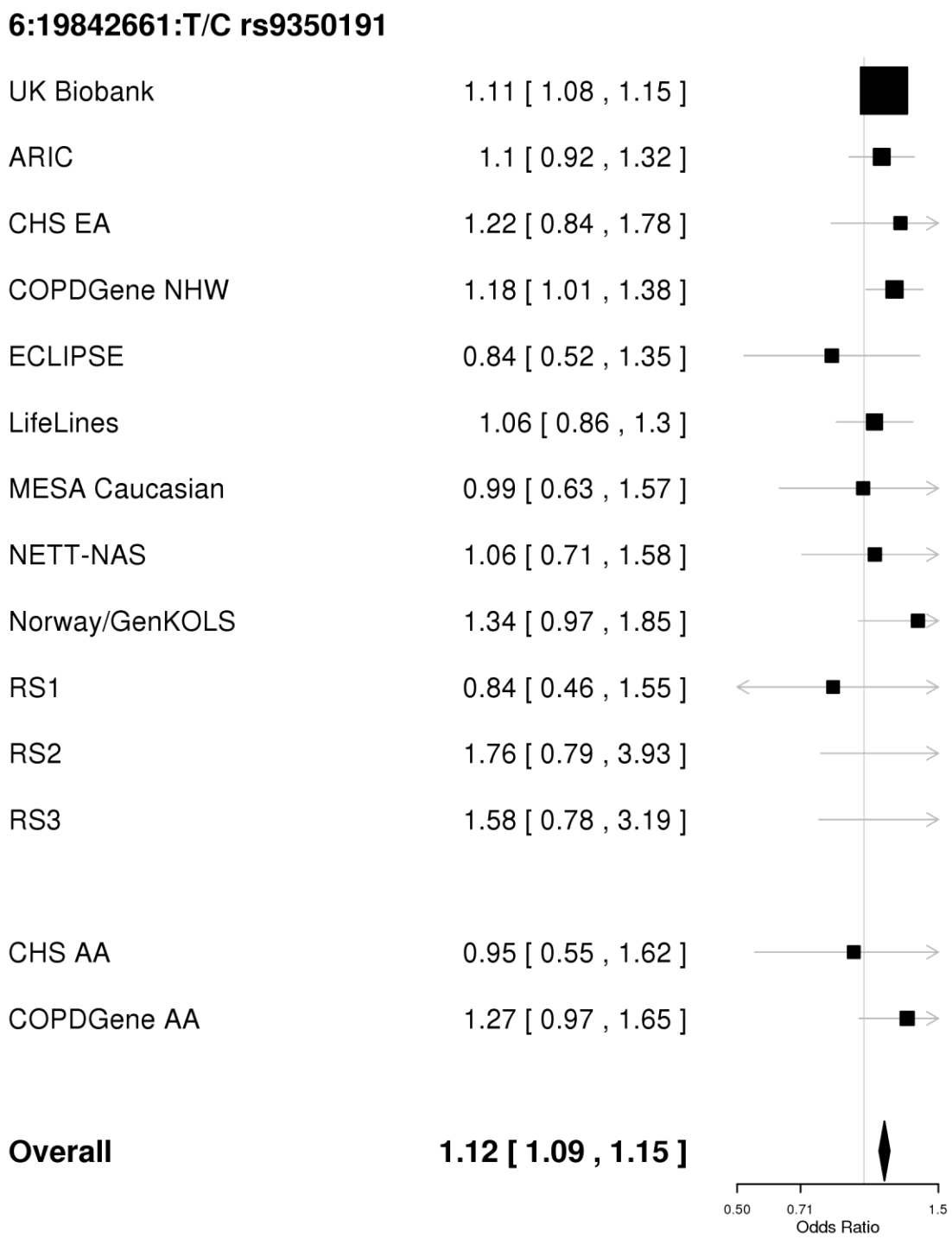
Supplementary Figure 1-36: Forest plot for rs12519165 (*FGF18* locus at 5q35.1)



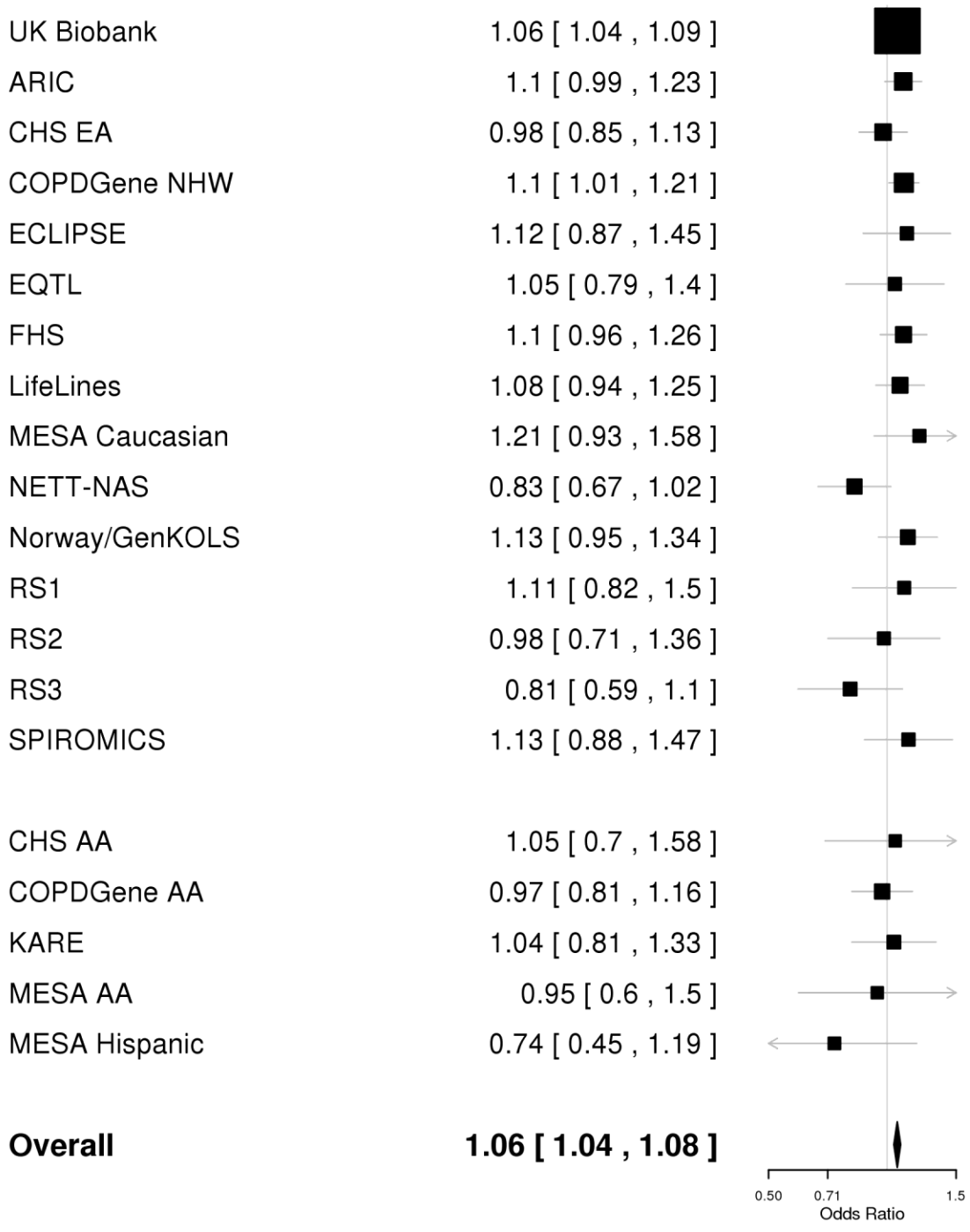
Supplementary Figure 1-37: Forest plot for rs1334576 (*RREB1* locus at 6p24.3)



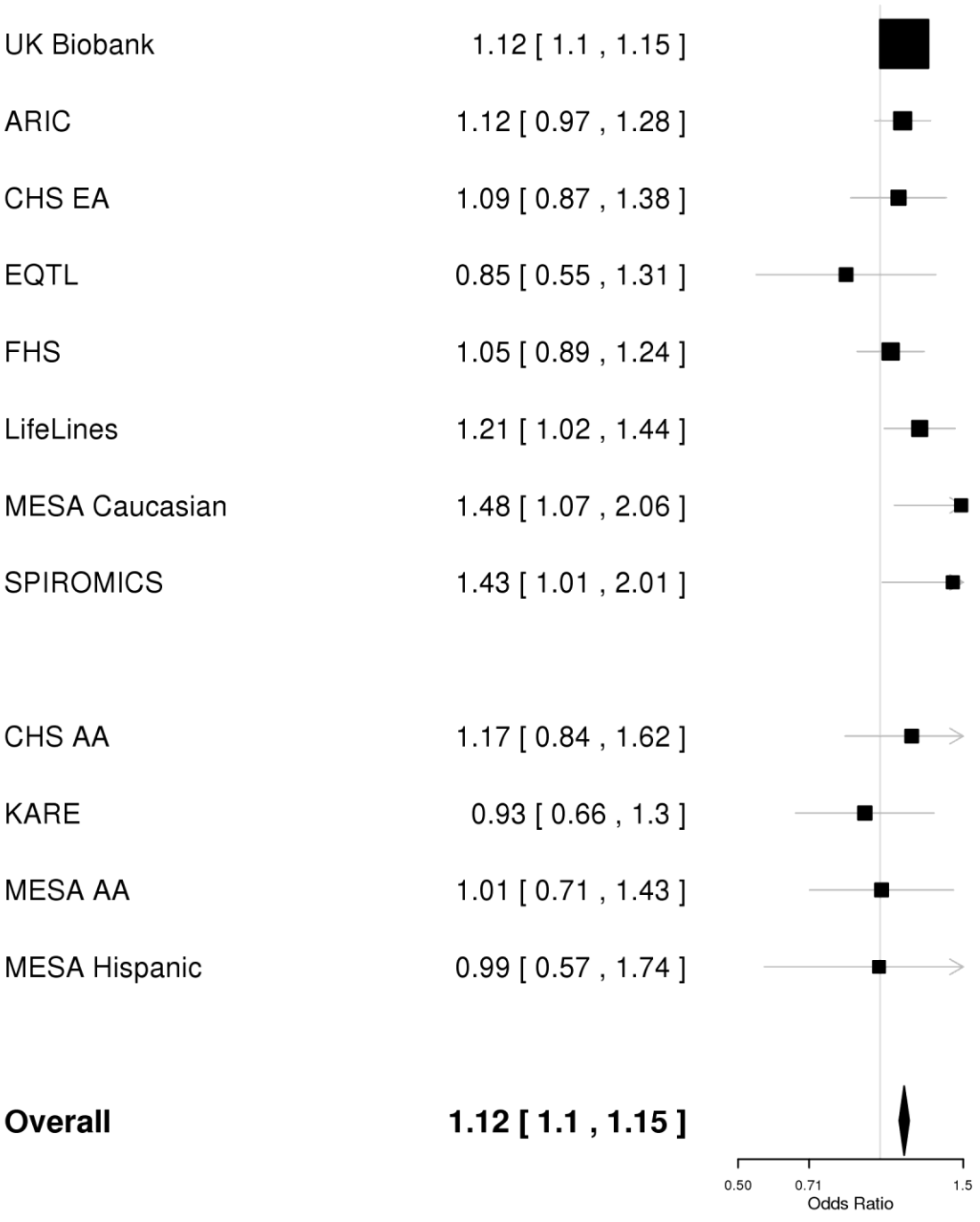
Supplementary Figure 1-38: Forest plot for rs9350191 (*ID4* locus at 6p22.3)



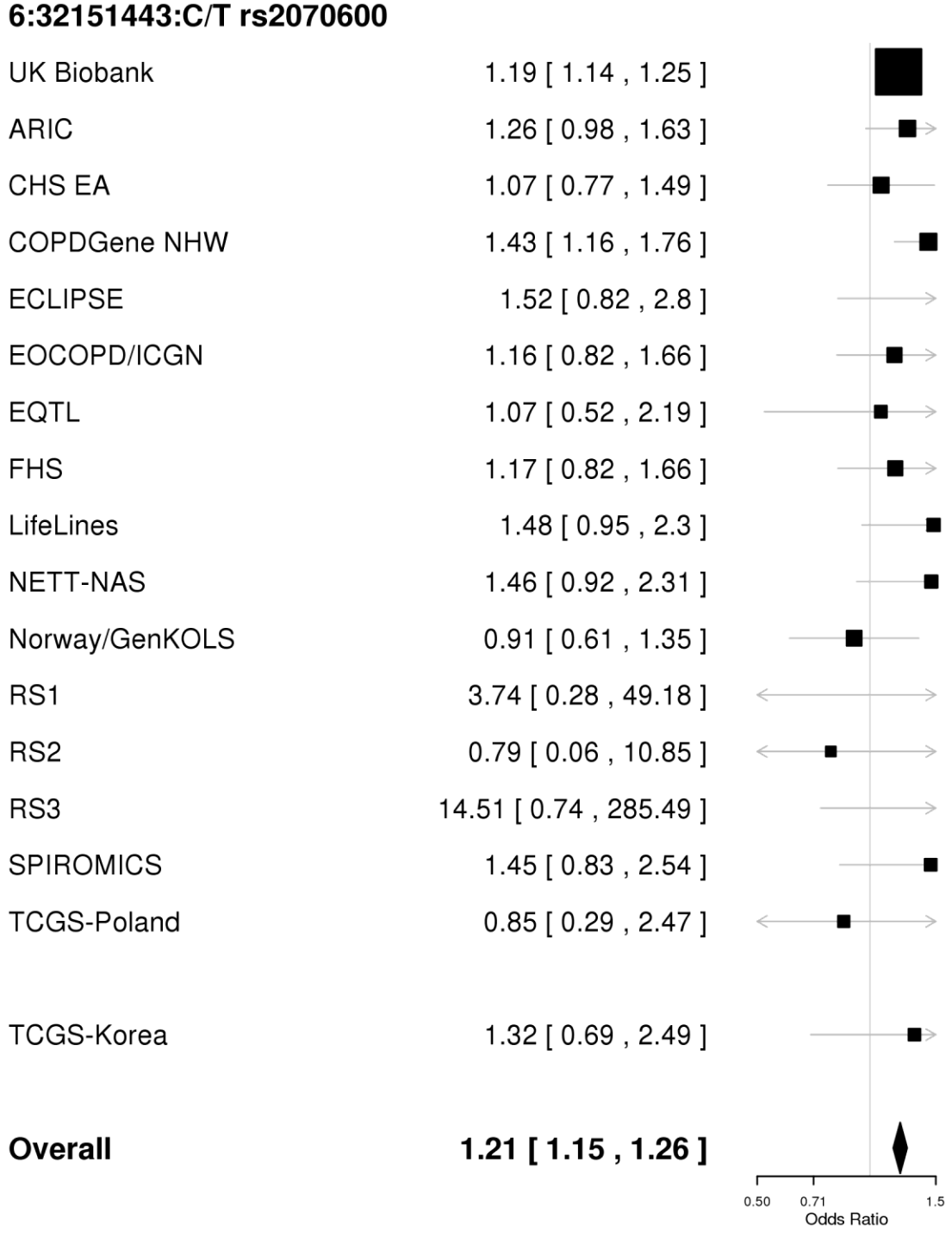
Supplementary Figure 1-39: Forest plot for rs13198656 (*PRL* locus at 6p22.3)



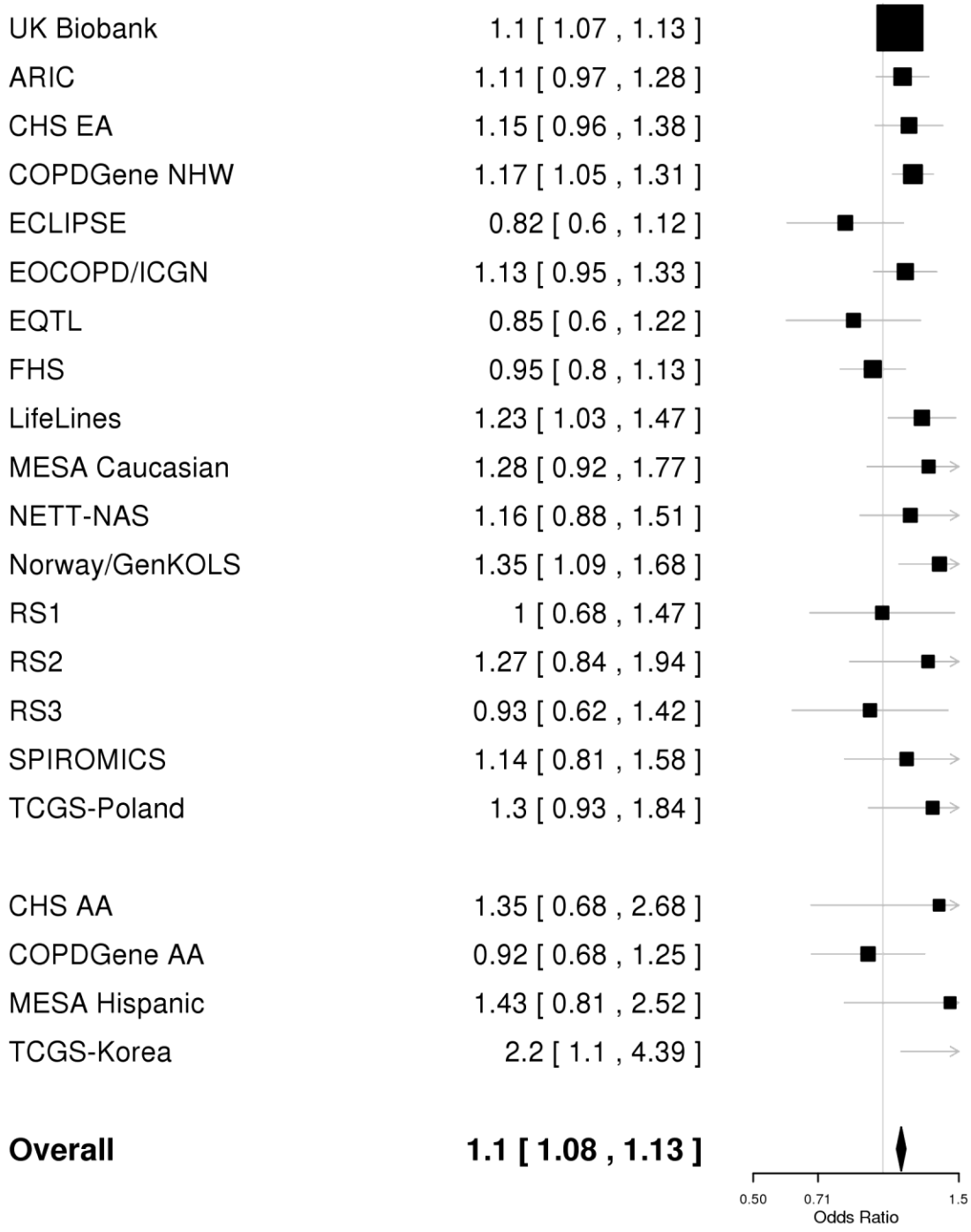
Supplementary Figure 1-40: Forest plot for rs2284174 (*IER3* locus at 6p21.33)



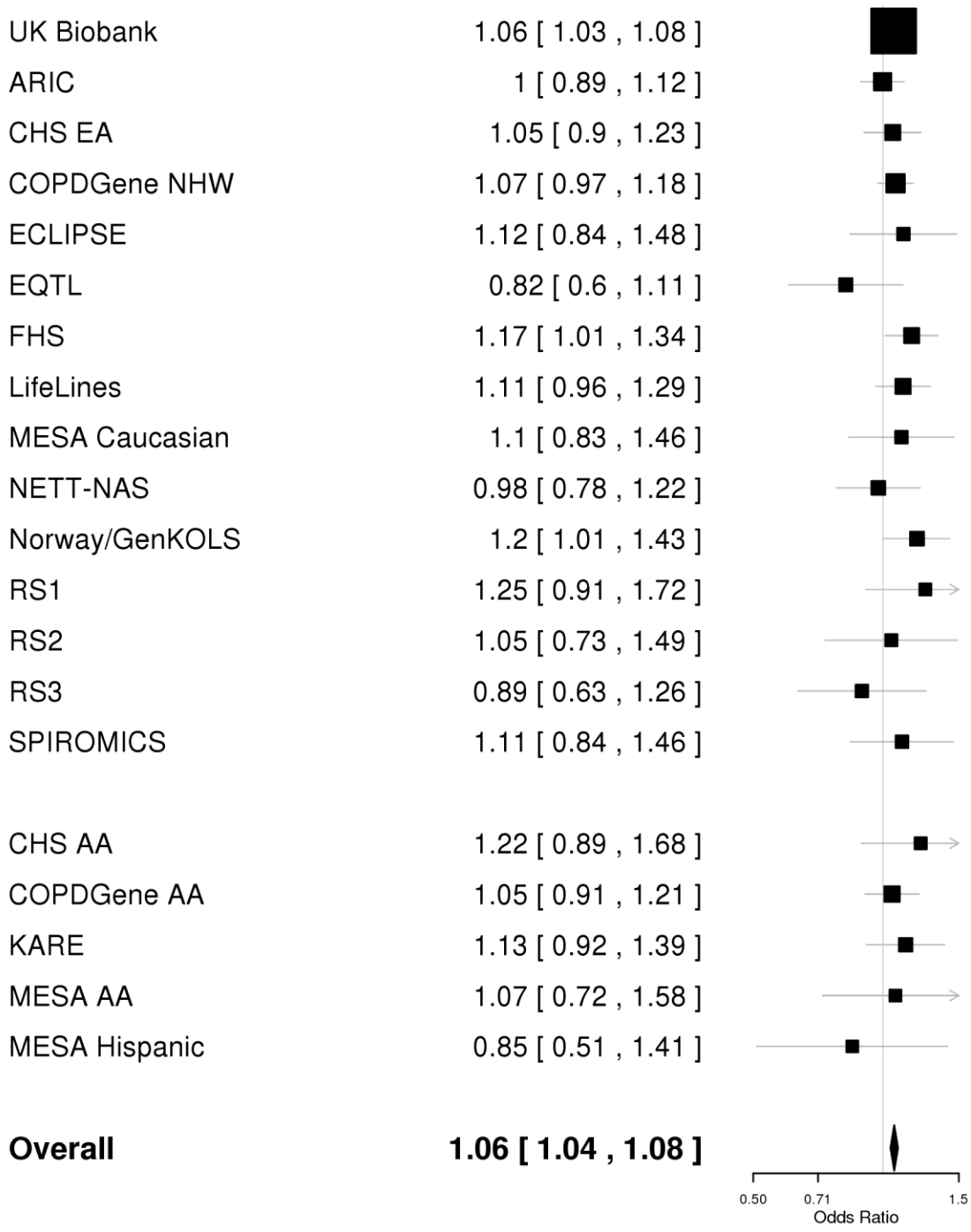
Supplementary Figure 1-41: Forest plot for rs2070600 (*AGER* locus at 6p21.32)



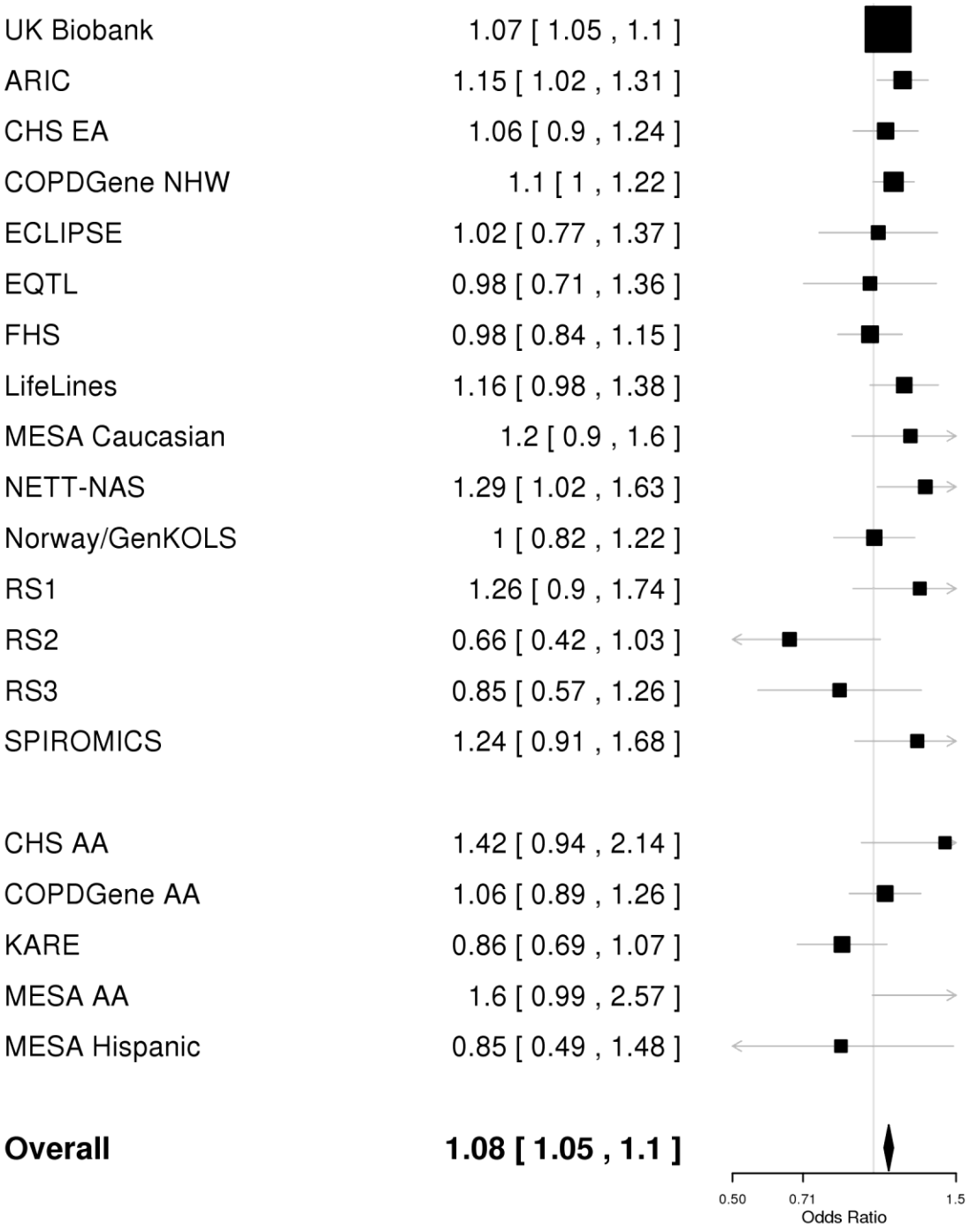
Supplementary Figure 1-42: Forest plot for rs2806356 (*ARMC2* locus at 6q21)



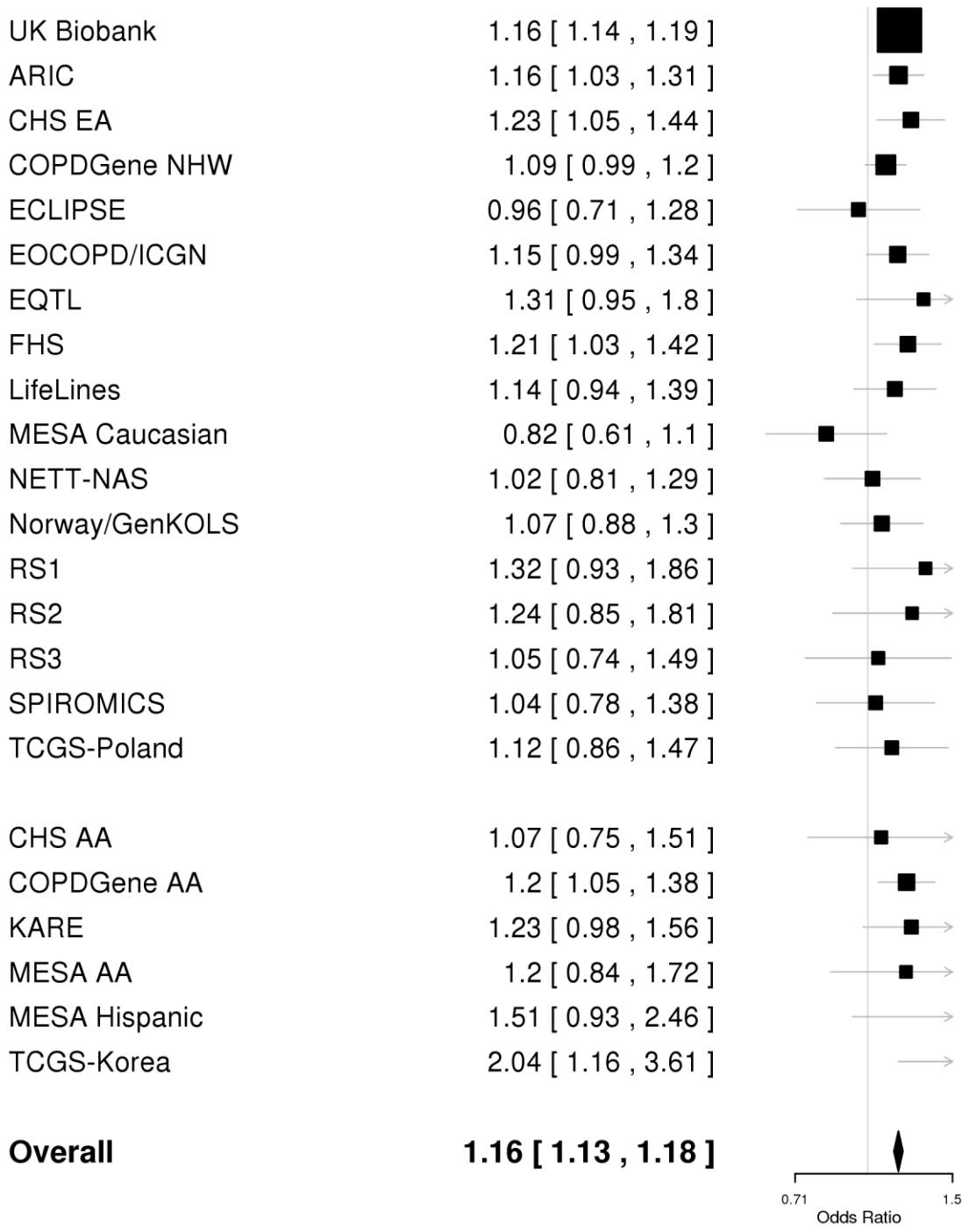
Supplementary Figure 1-43: Forest plot for rs674621 (*RFX6* locus at 6q22.1)



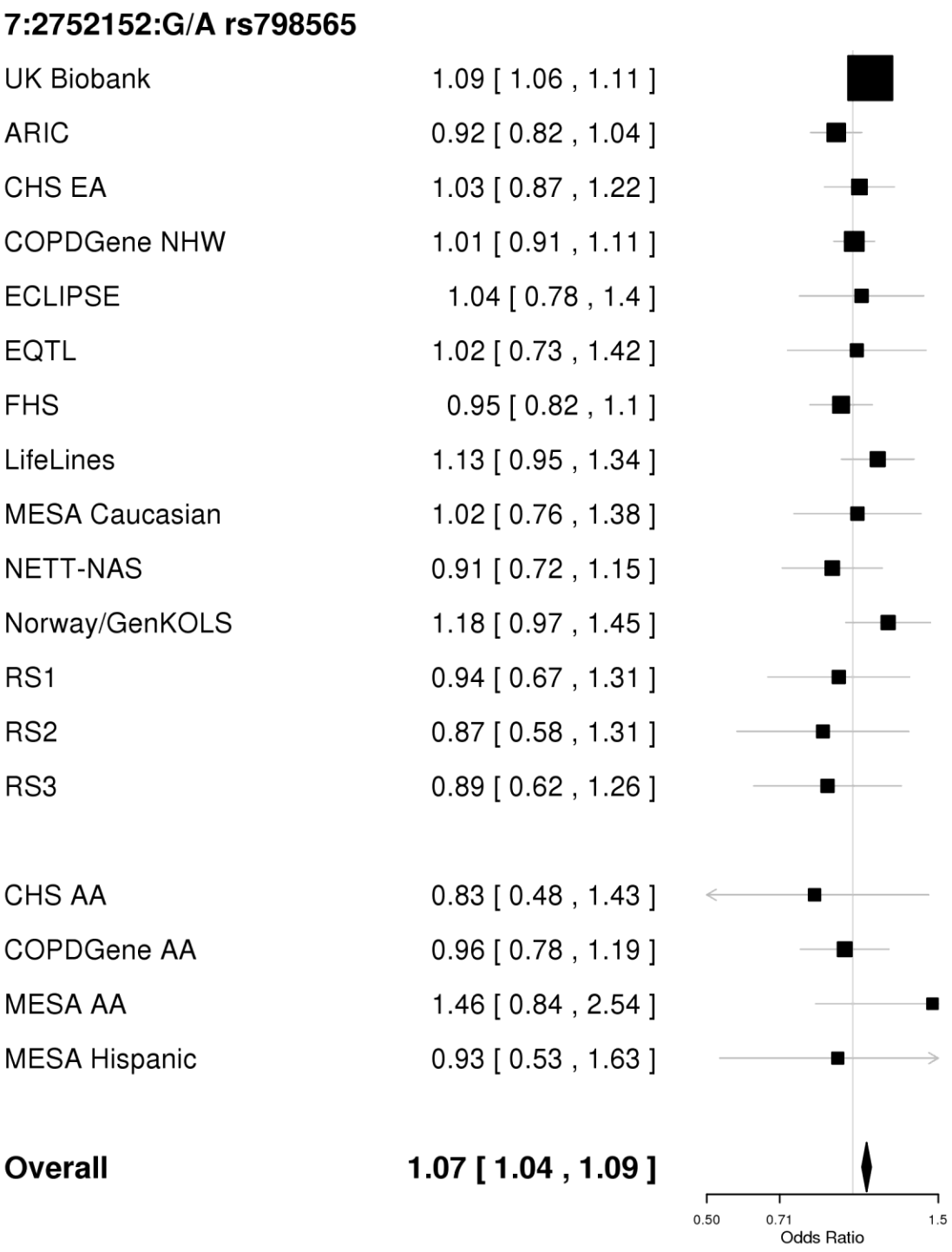
Supplementary Figure 1-44: Forest plot for rs646695 (*CITED2* locus at 6q24.1)



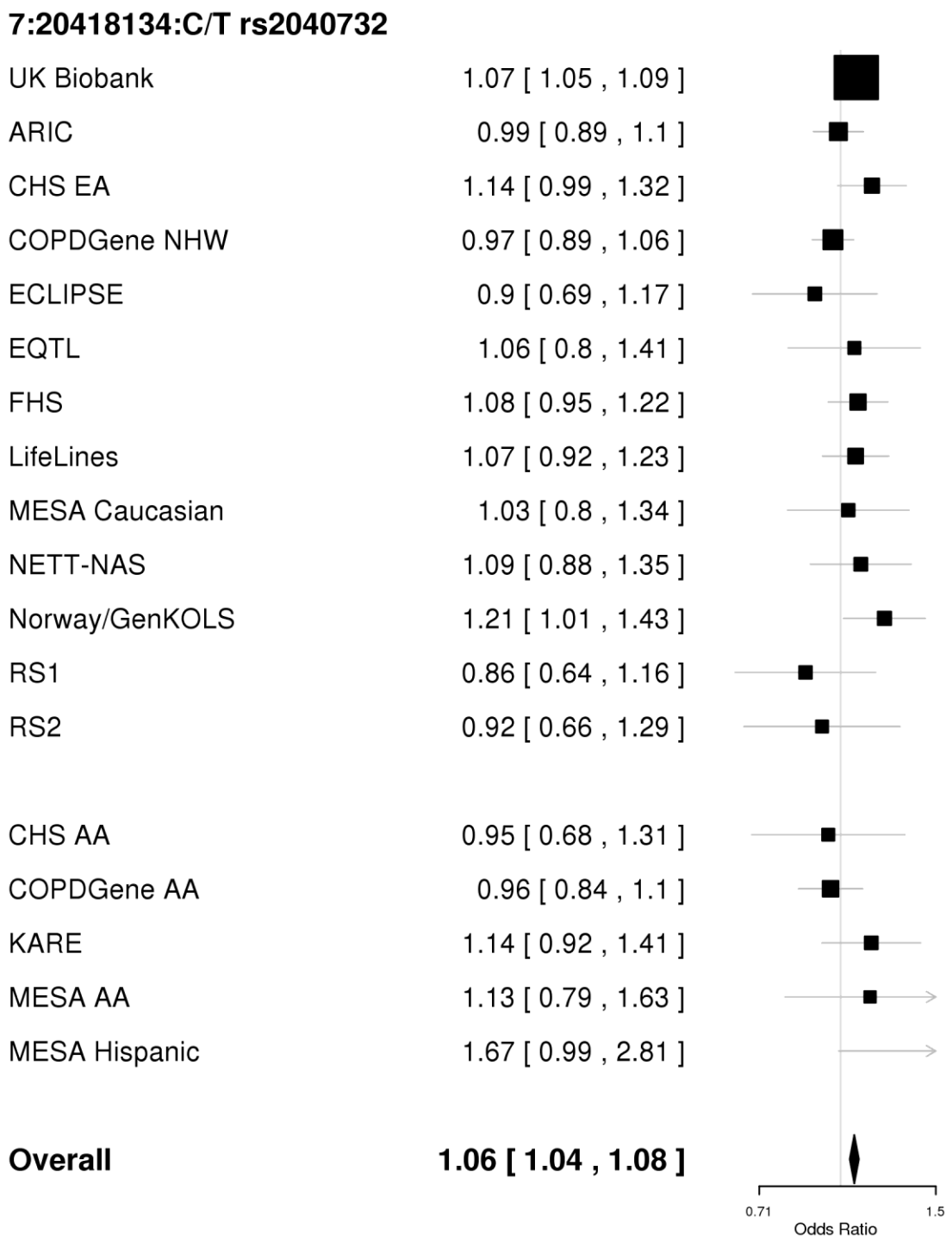
Supplementary Figure 1-45: Forest plot for rs9399401 (*ADGRG6* locus at 6q24.1)



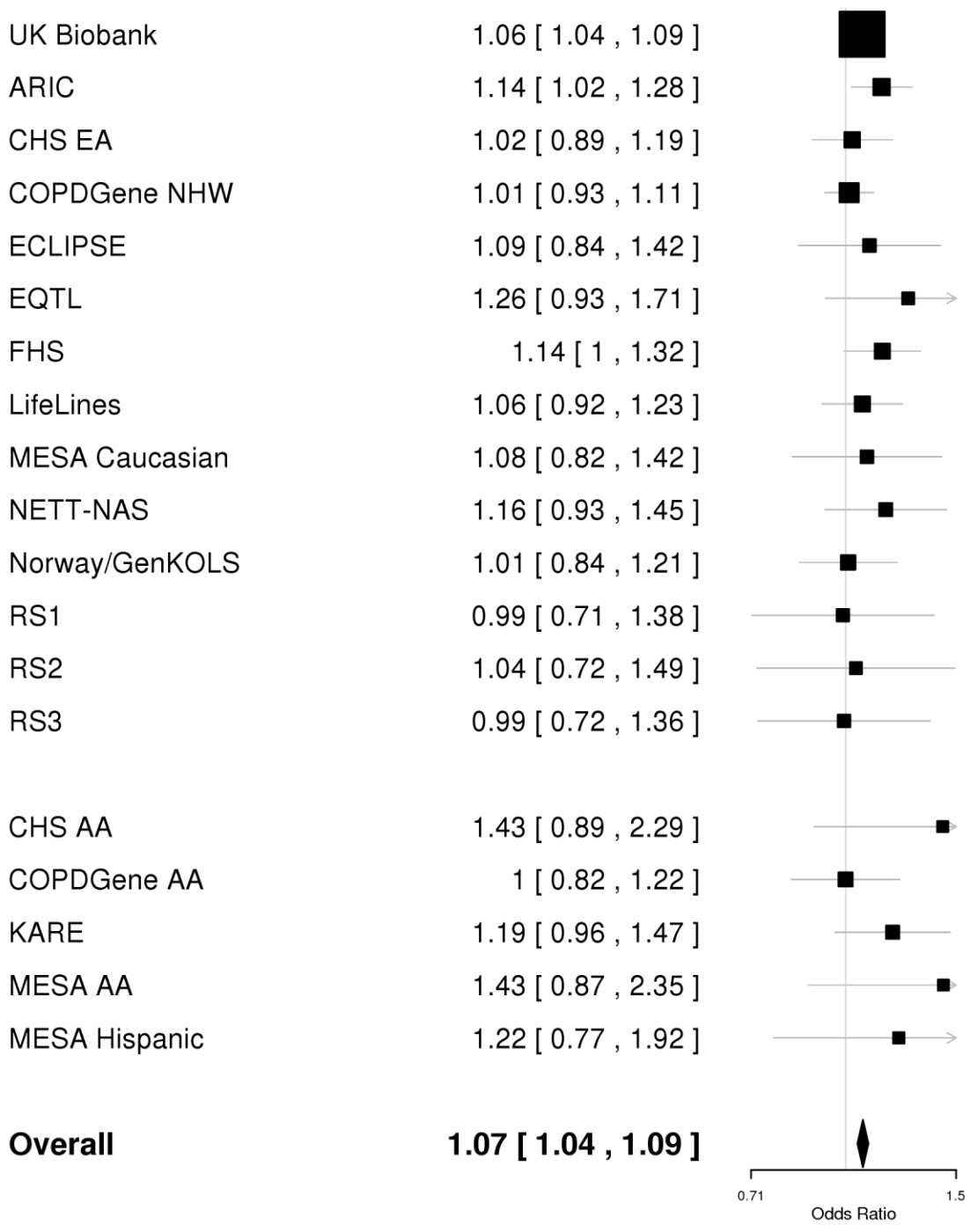
Supplementary Figure 1-46: Forest plot for rs798565 (*AMZ1* locus at 7p22.3)



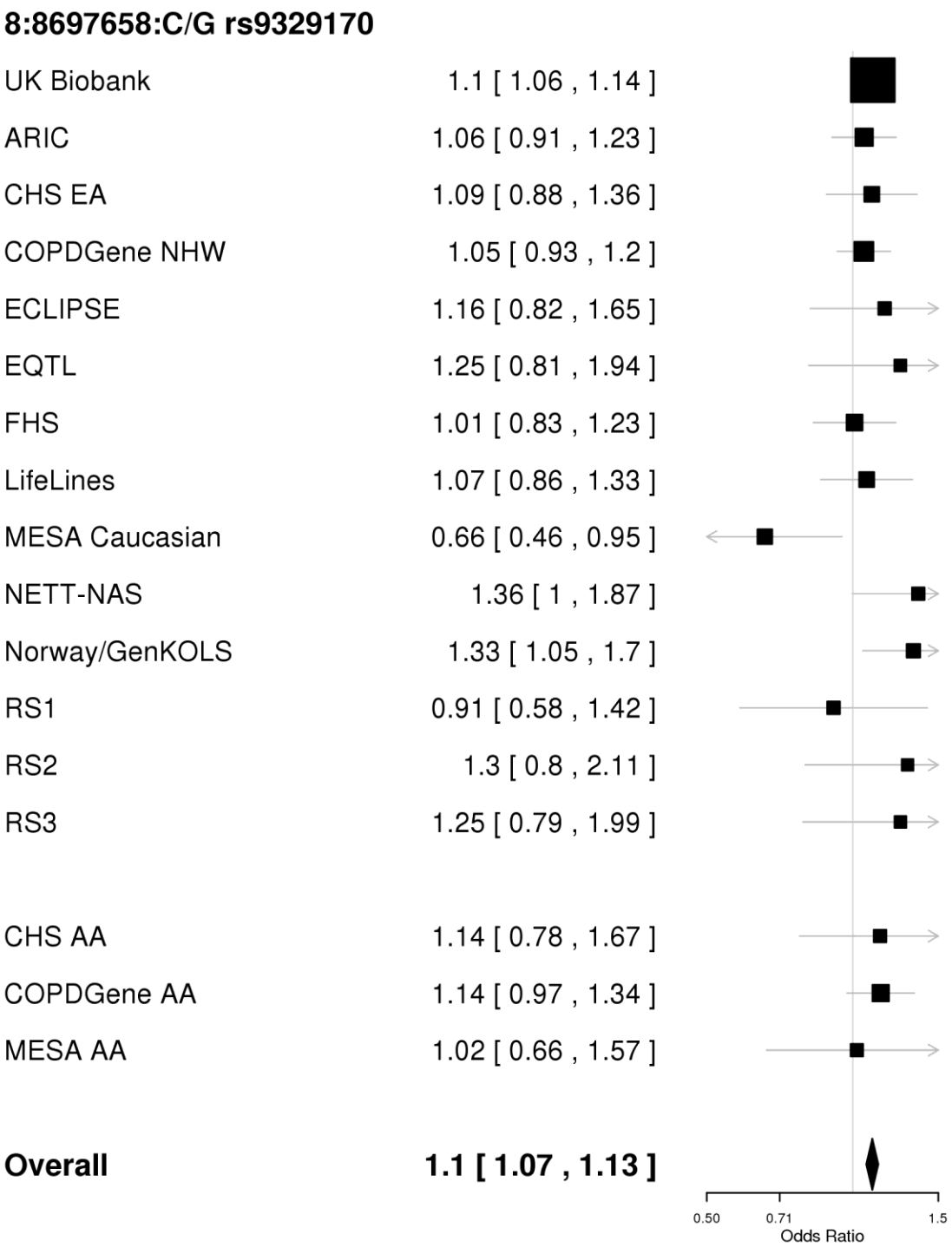
Supplementary Figure 1-47: Forest plot for rs2040732 (*ITGB8* locus at 7p21.1)



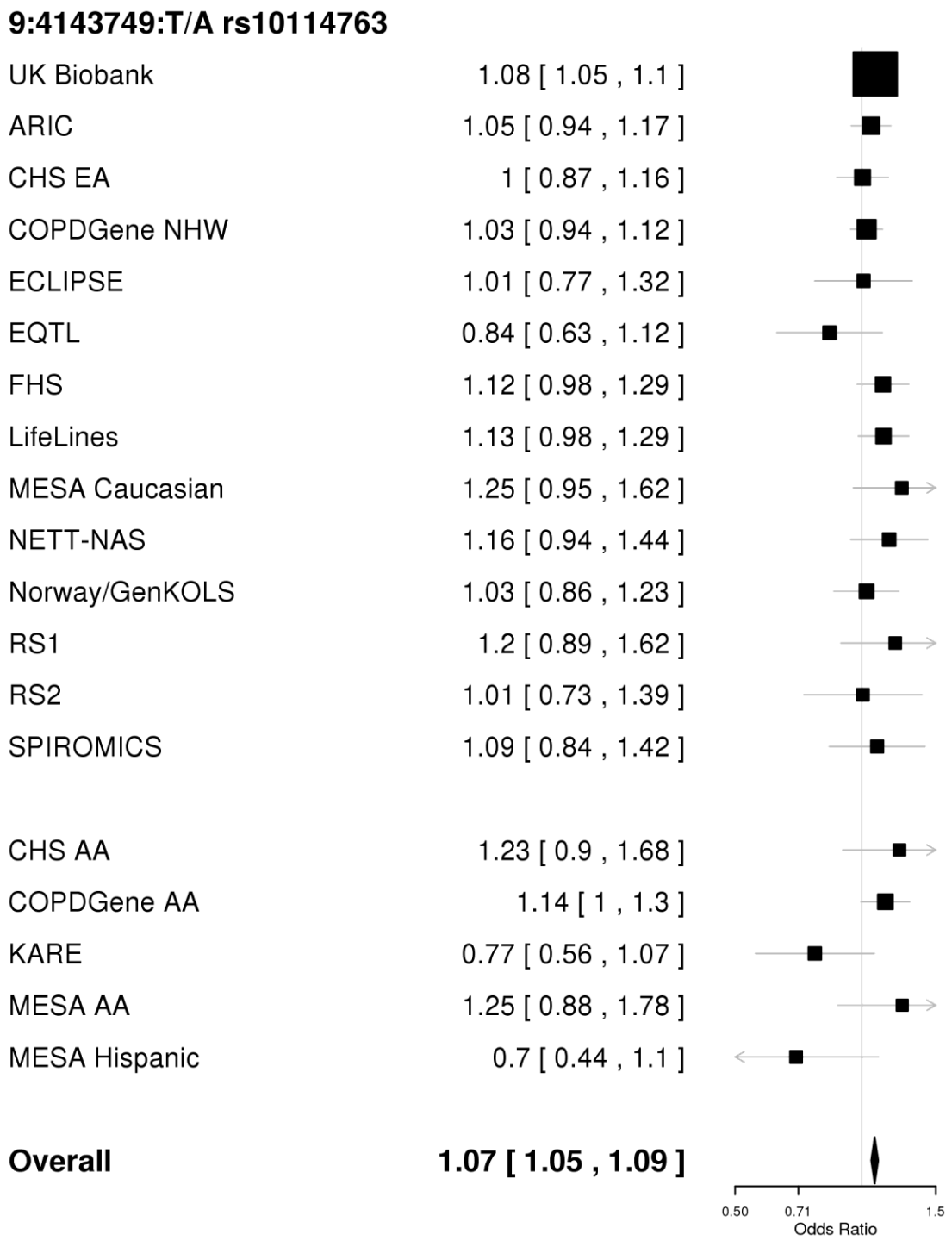
Supplementary Figure 1-48: Forest plot for rs2897075 (*ZKSCAN1* locus at 7q22.1)



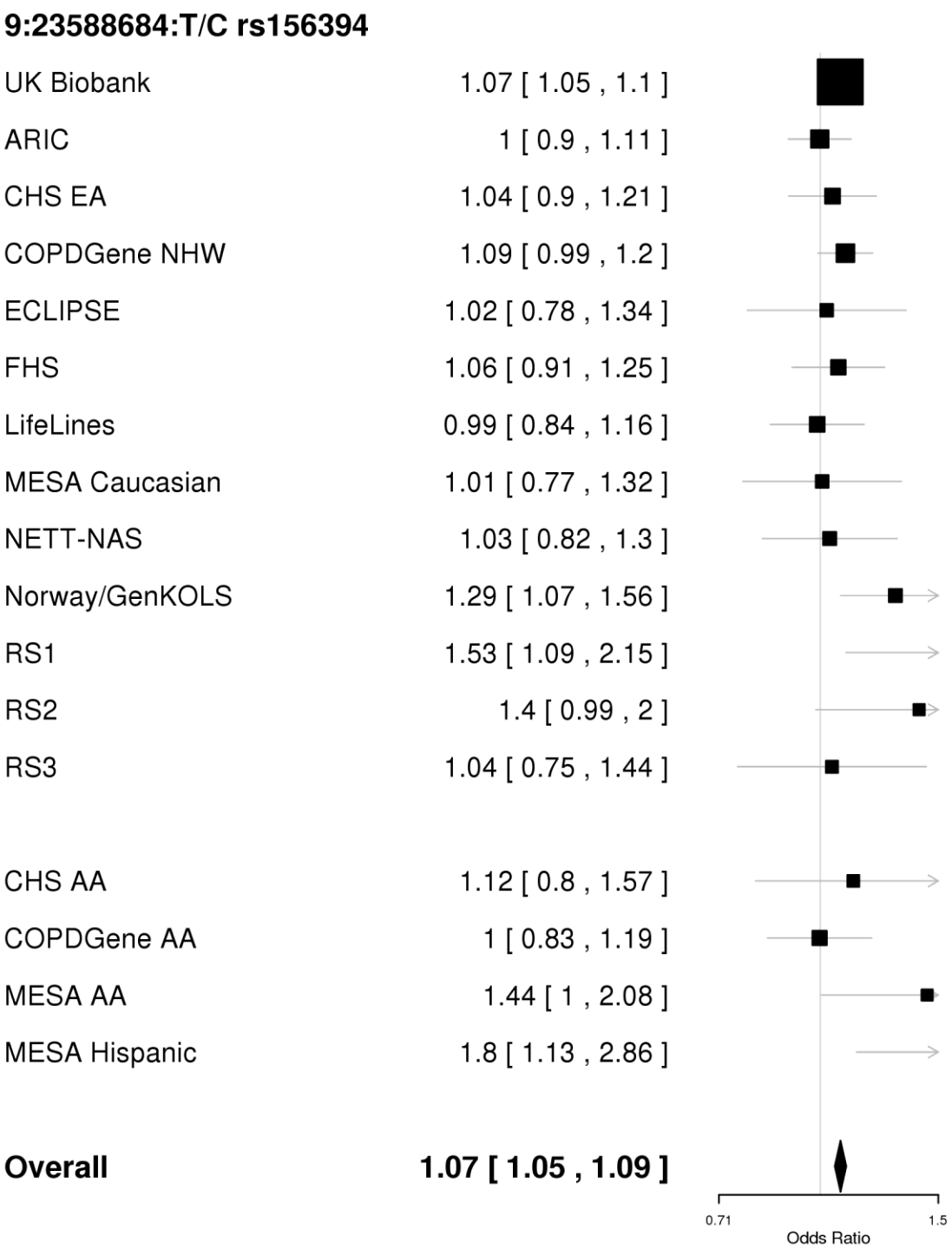
Supplementary Figure 1-49: Forest plot for rs9329170 (*MFHAS1* locus at 8p23.1)



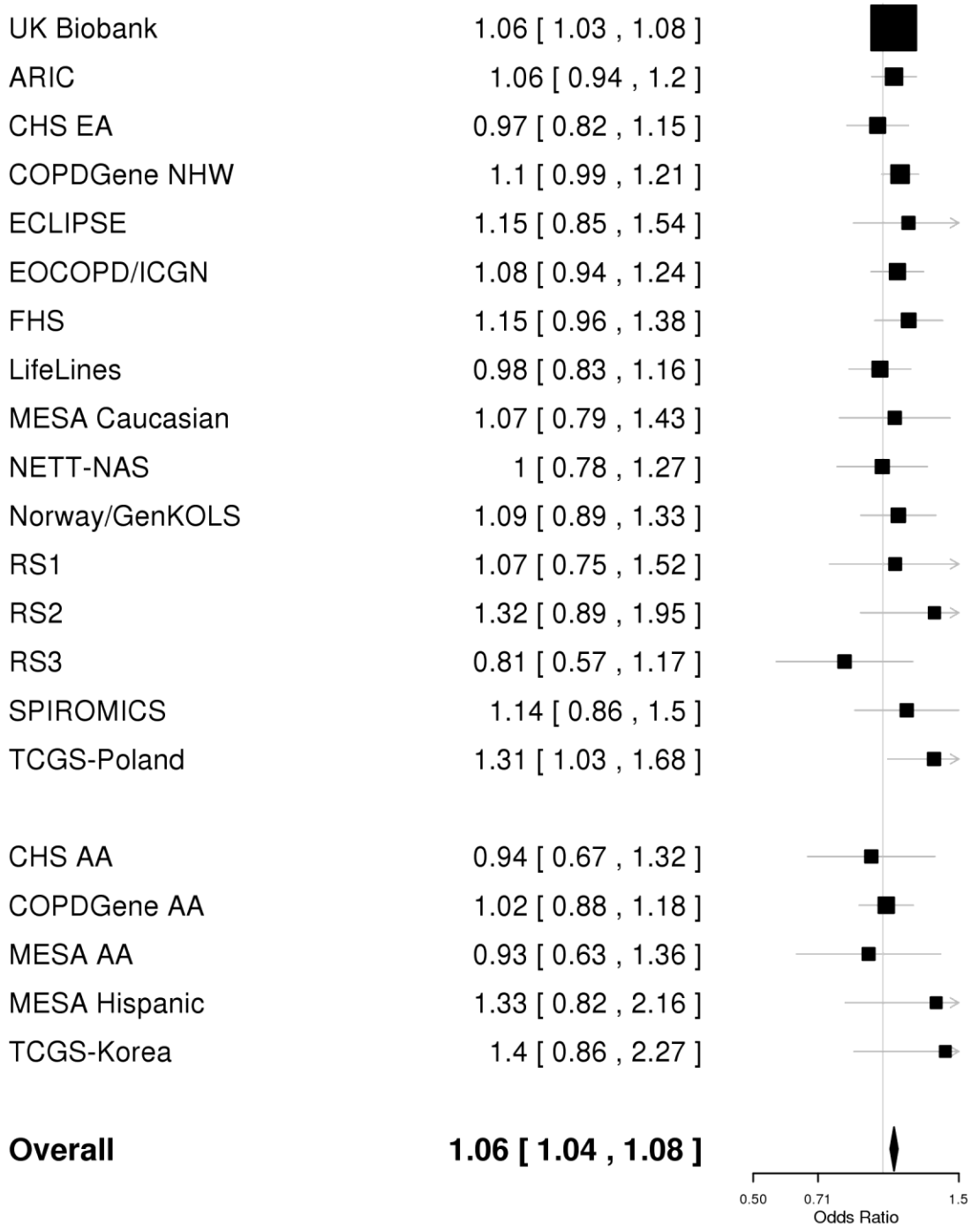
Supplementary Figure 1-50: Forest plot for rs10114763 (*GLIS3* locus at 9p24.2)



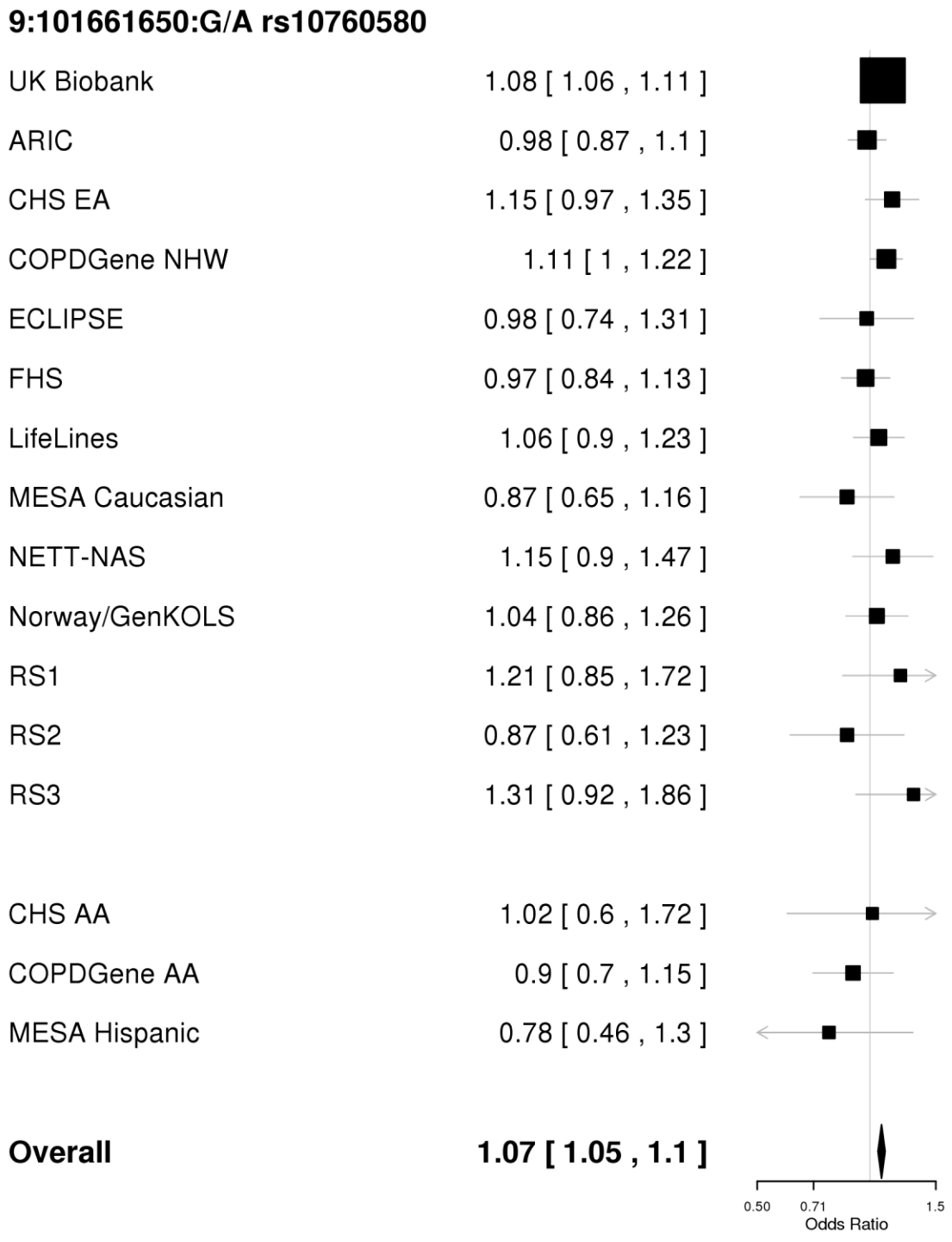
Supplementary Figure 1-51: Forest plot for rs156394 (*ELAVL2* locus at 9p21.3)



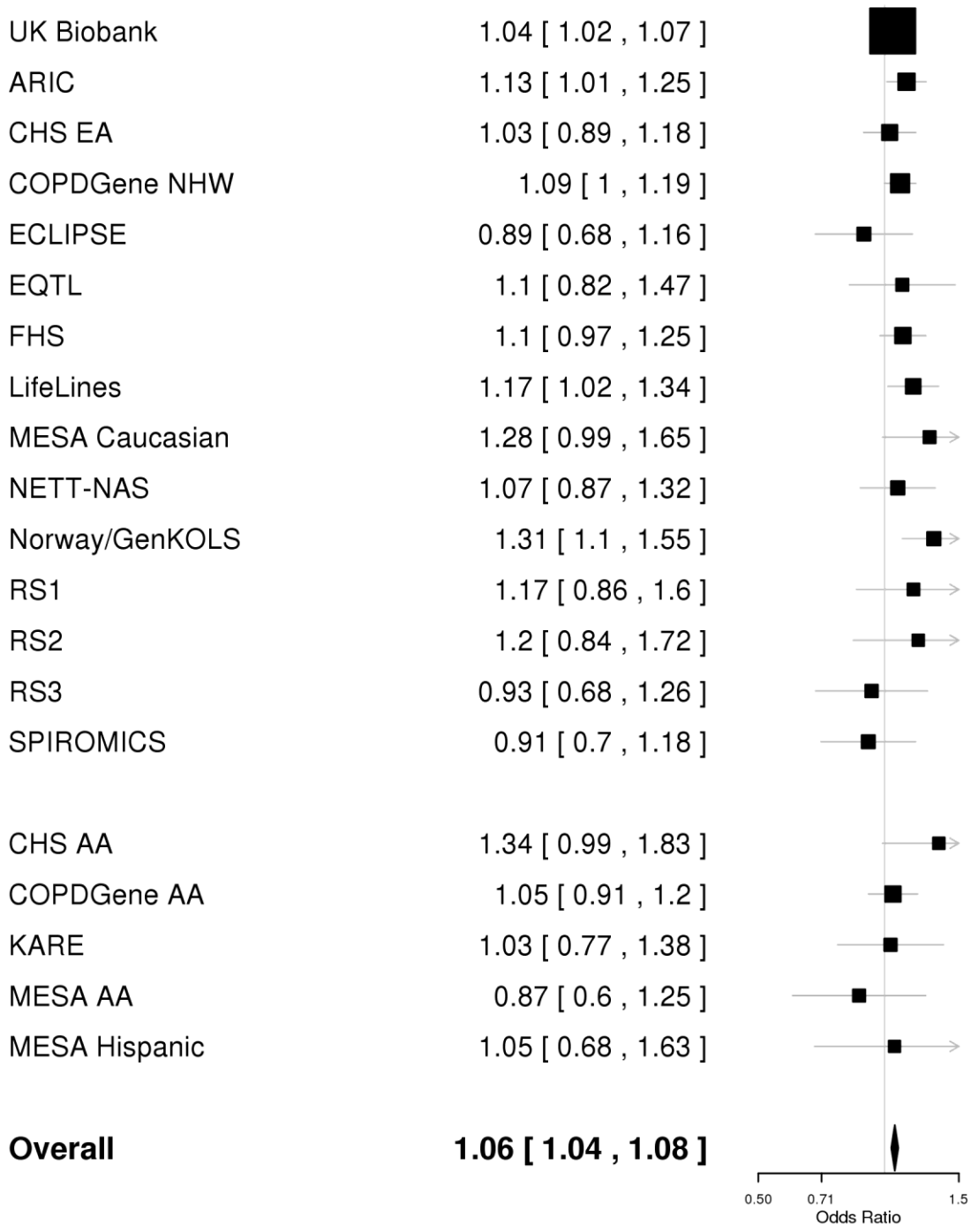
Supplementary Figure 1-52: Forest plot for rs7866939 (*RASEF* locus at 9q21.32)



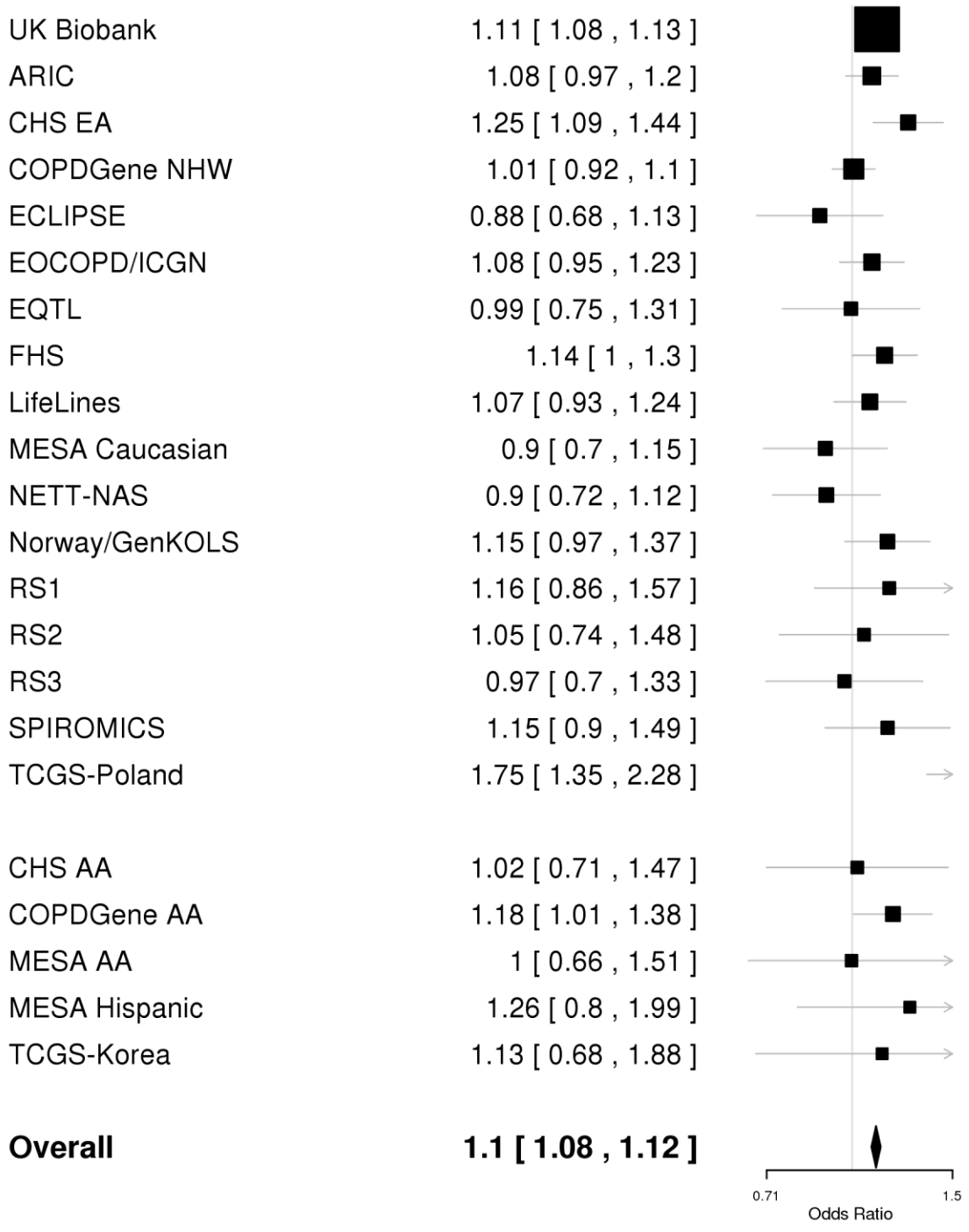
Supplementary Figure 1-53: Forest plot for rs10760580 (*COL15A1* locus at 9q22.33)



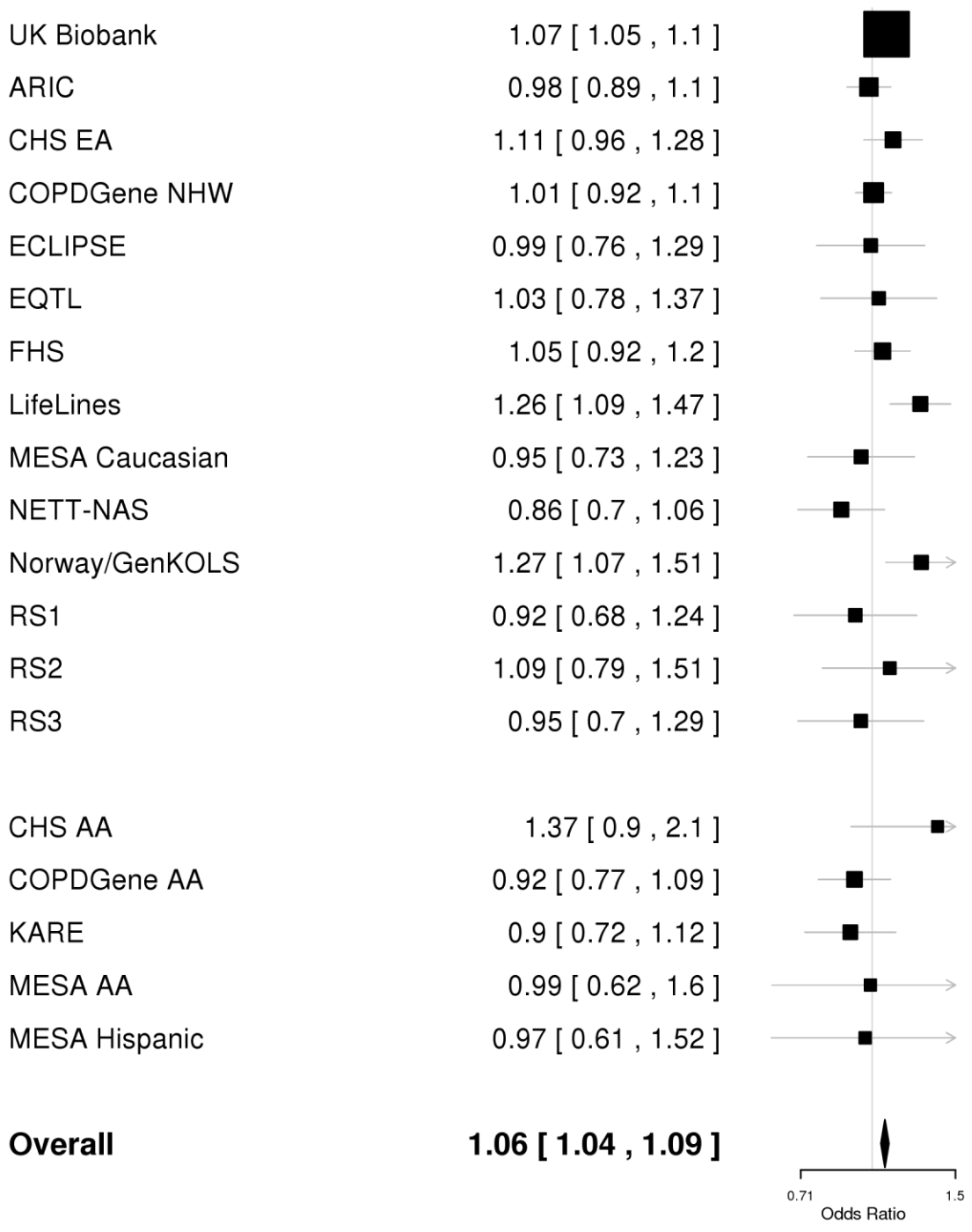
Supplementary Figure 1-54: Forest plot for rs803923 (*ASTN2* locus at 9q33.1)



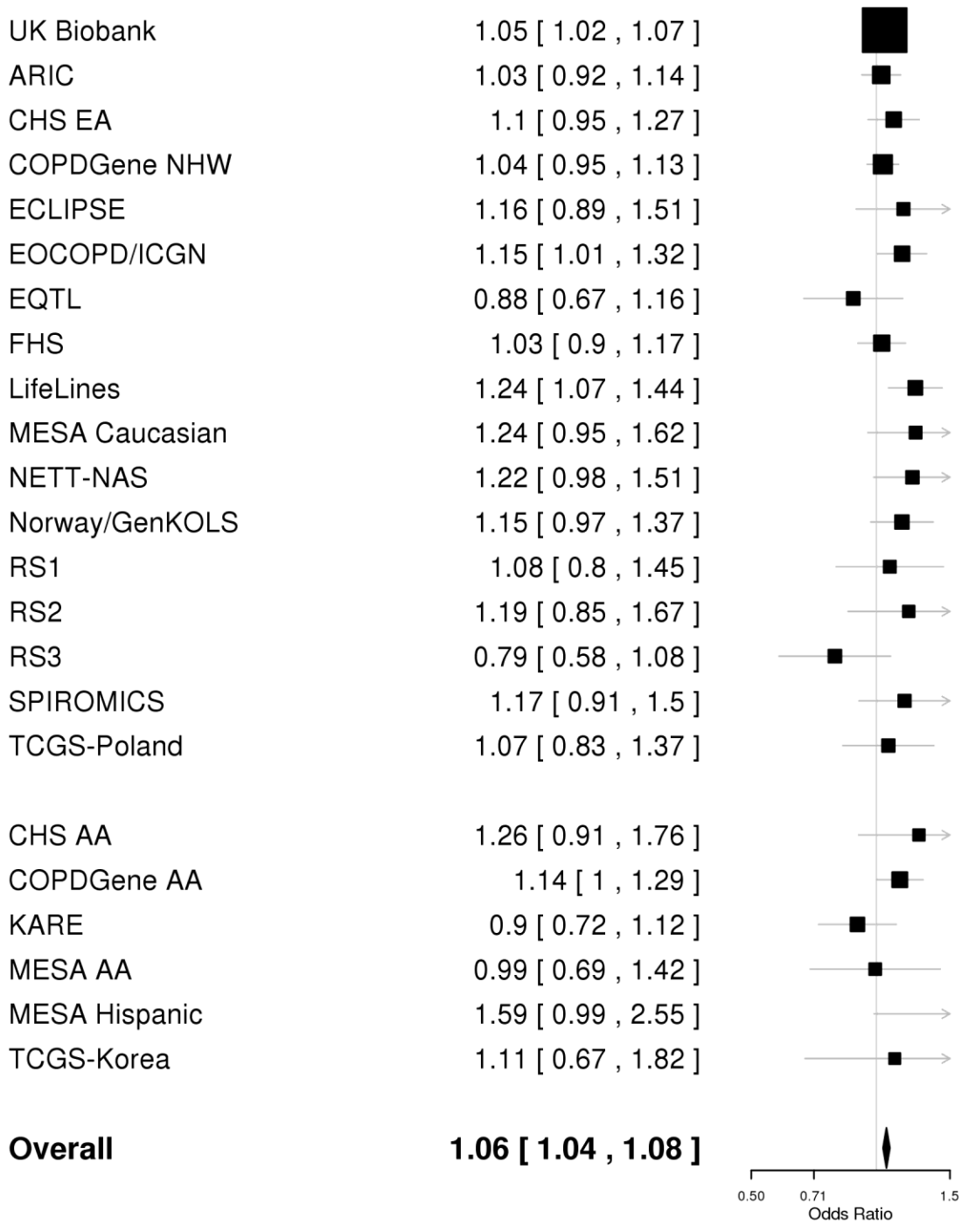
Supplementary Figure 1-55: Forest plot for rs7068966 (*CDC123* locus at 10p13)



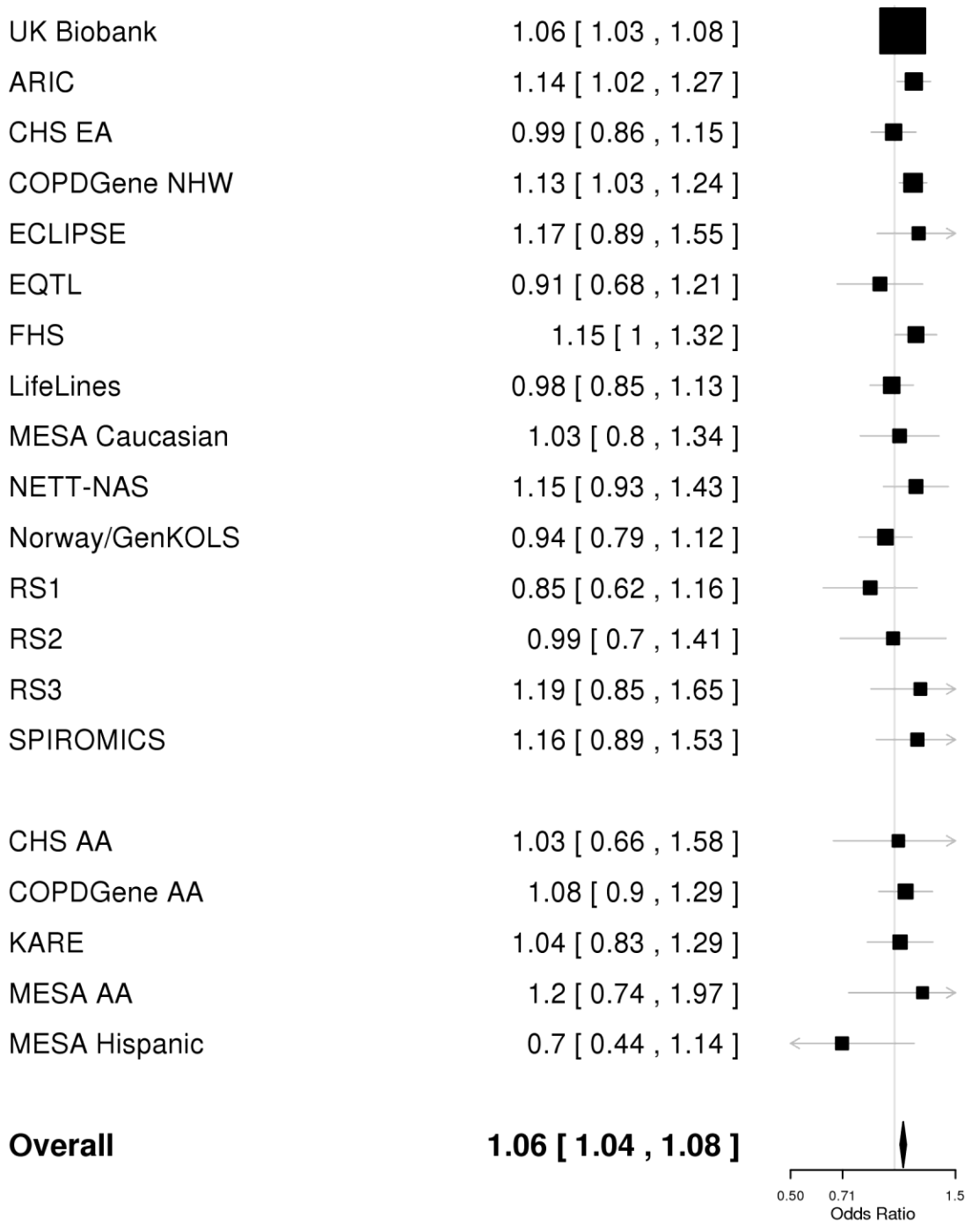
Supplementary Figure 1-56: Forest plot for rs2579762 (*LRMDA* locus at 10q22.3)



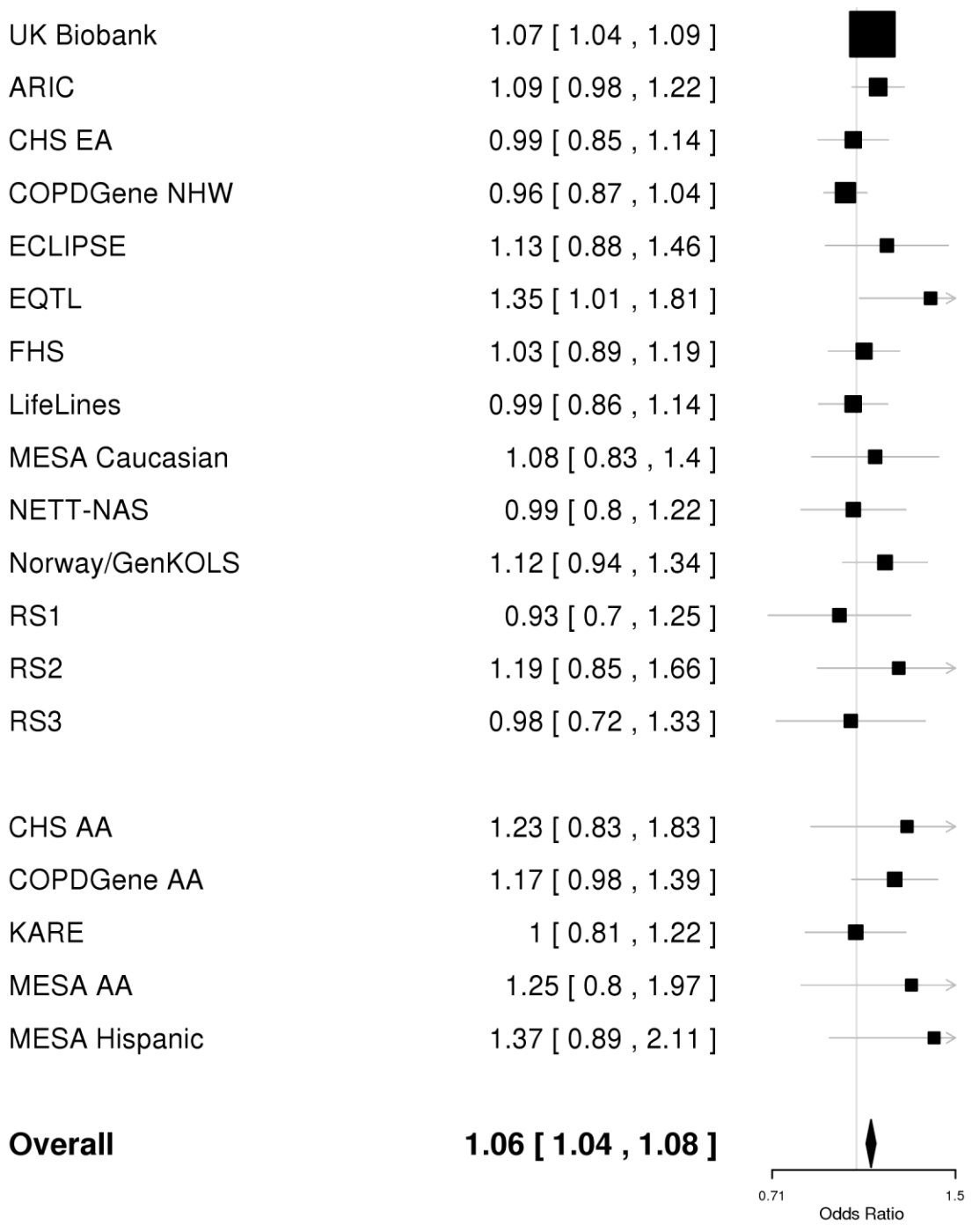
Supplementary Figure 1-57: Forest plot for rs721917 (*SFTPD* locus at 10q22.3)



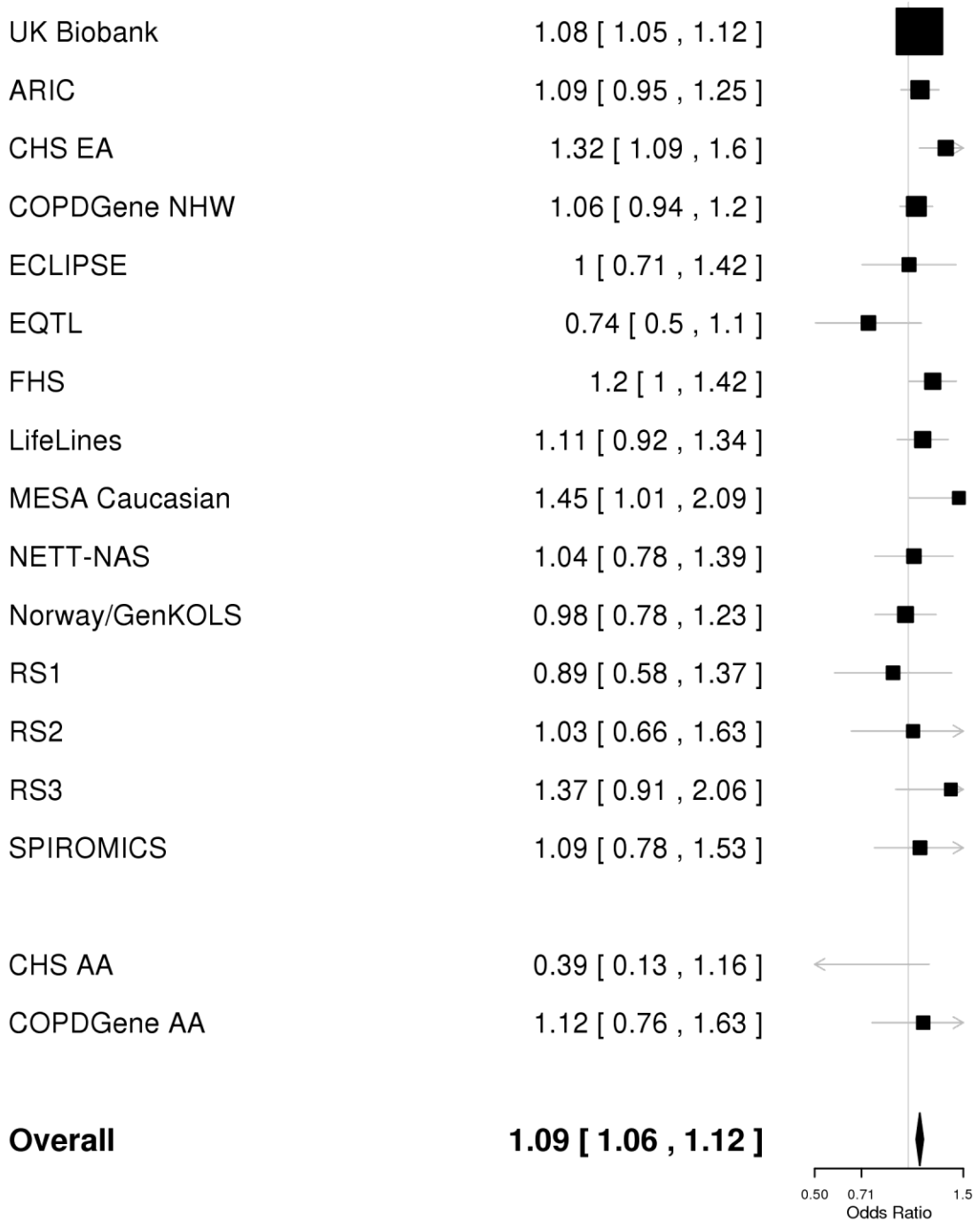
Supplementary Figure 1-58: Forest plot for rs1570221 (*STN1* locus at 10q24.33)



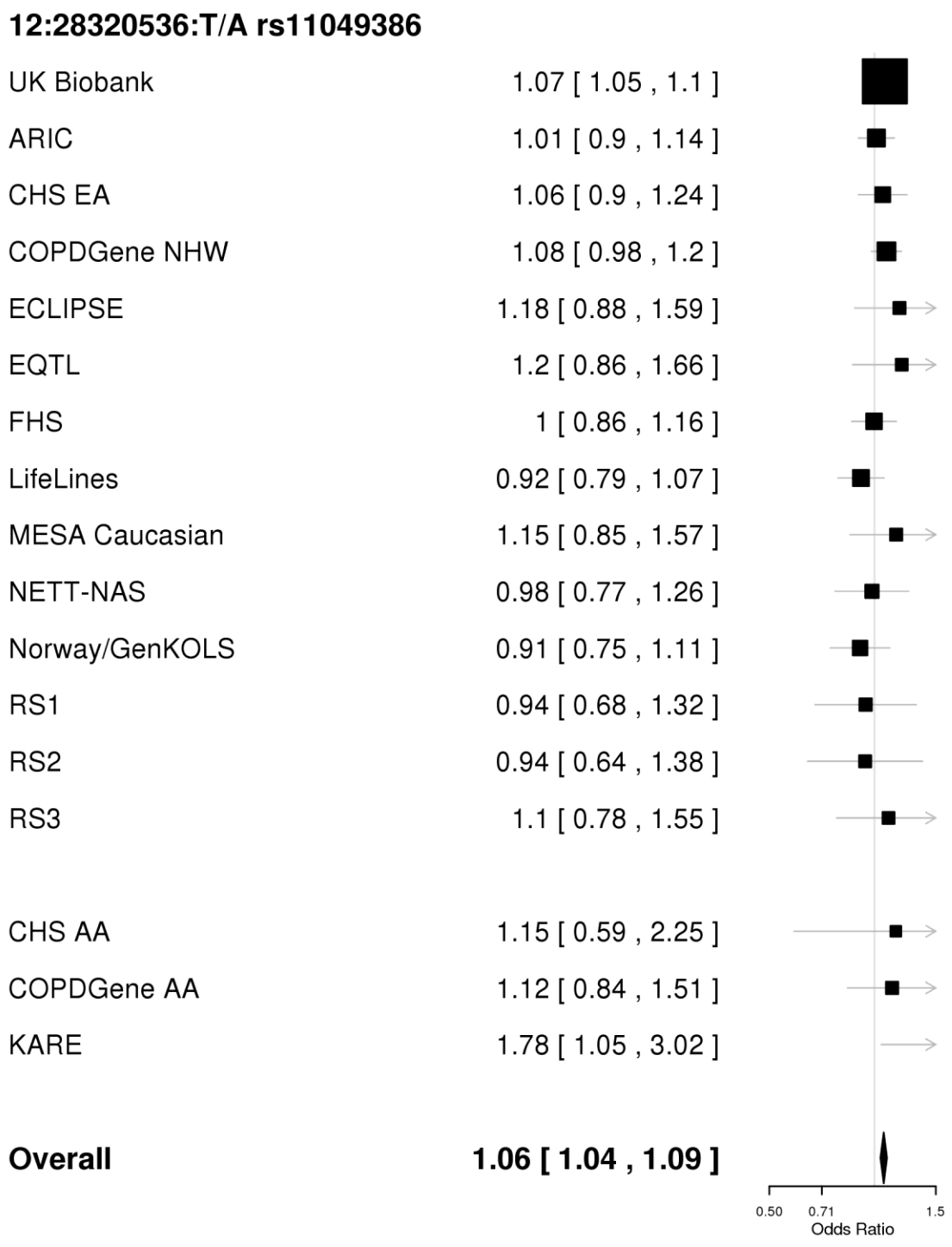
Supplementary Figure 1-59: Forest plot for rs4757118 (*ARNTL* locus at 11p15.2)



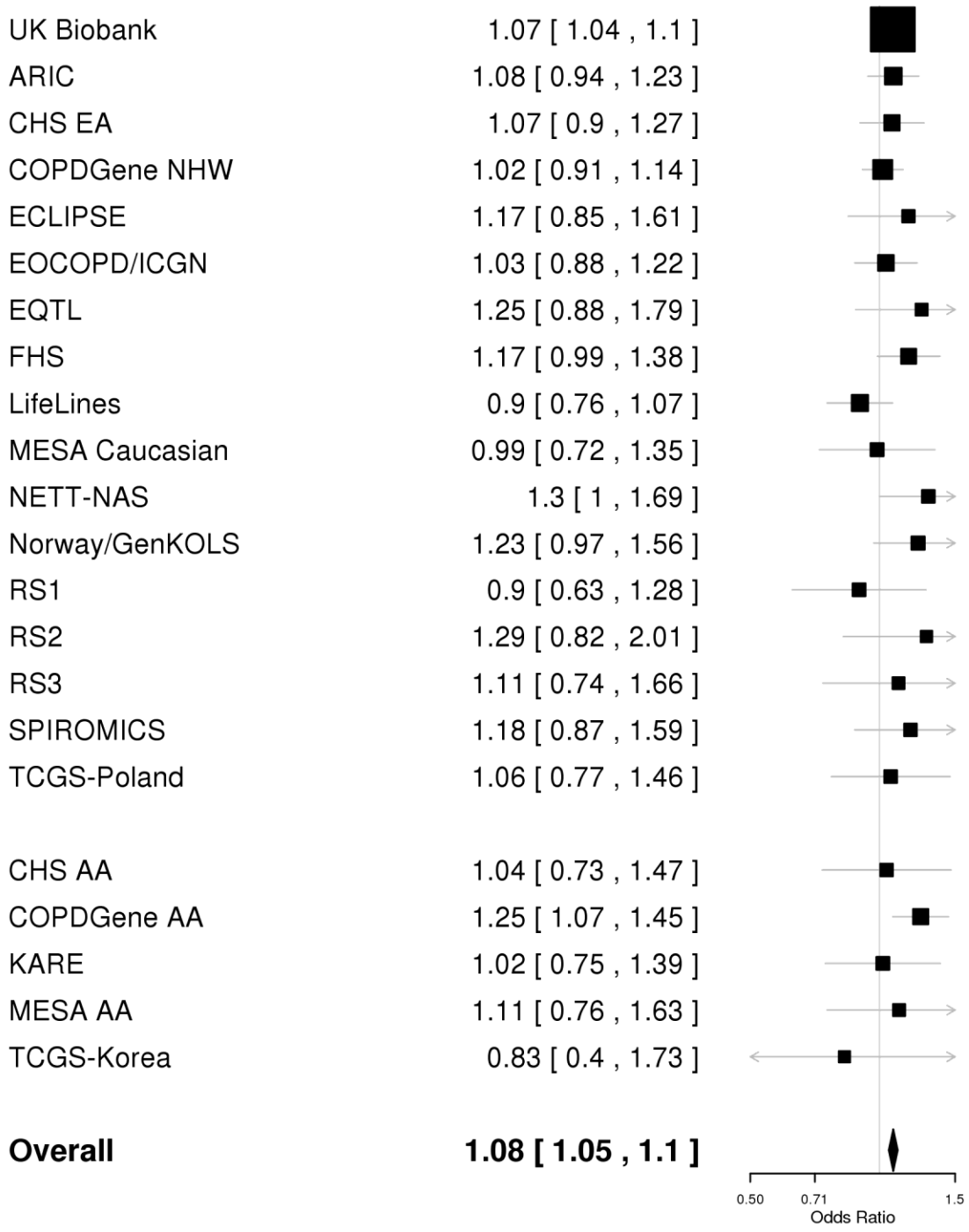
Supplementary Figure 1-60: Forest plot for rs117261012 (*PRSS23* locus at 11q14.2)



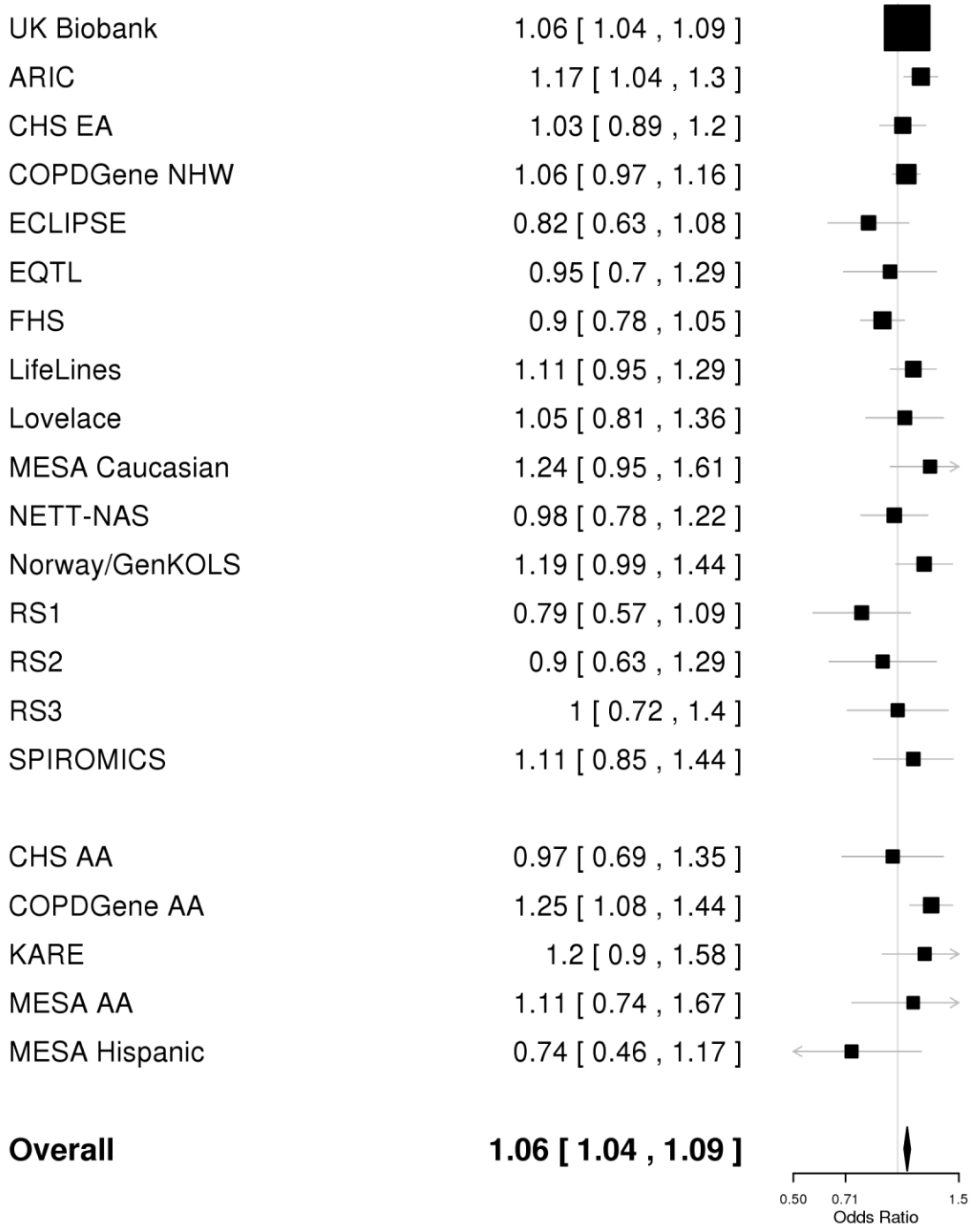
Supplementary Figure 1-61: Forest plot for rs11049386 (*CCDC91* locus at 12p11.22)



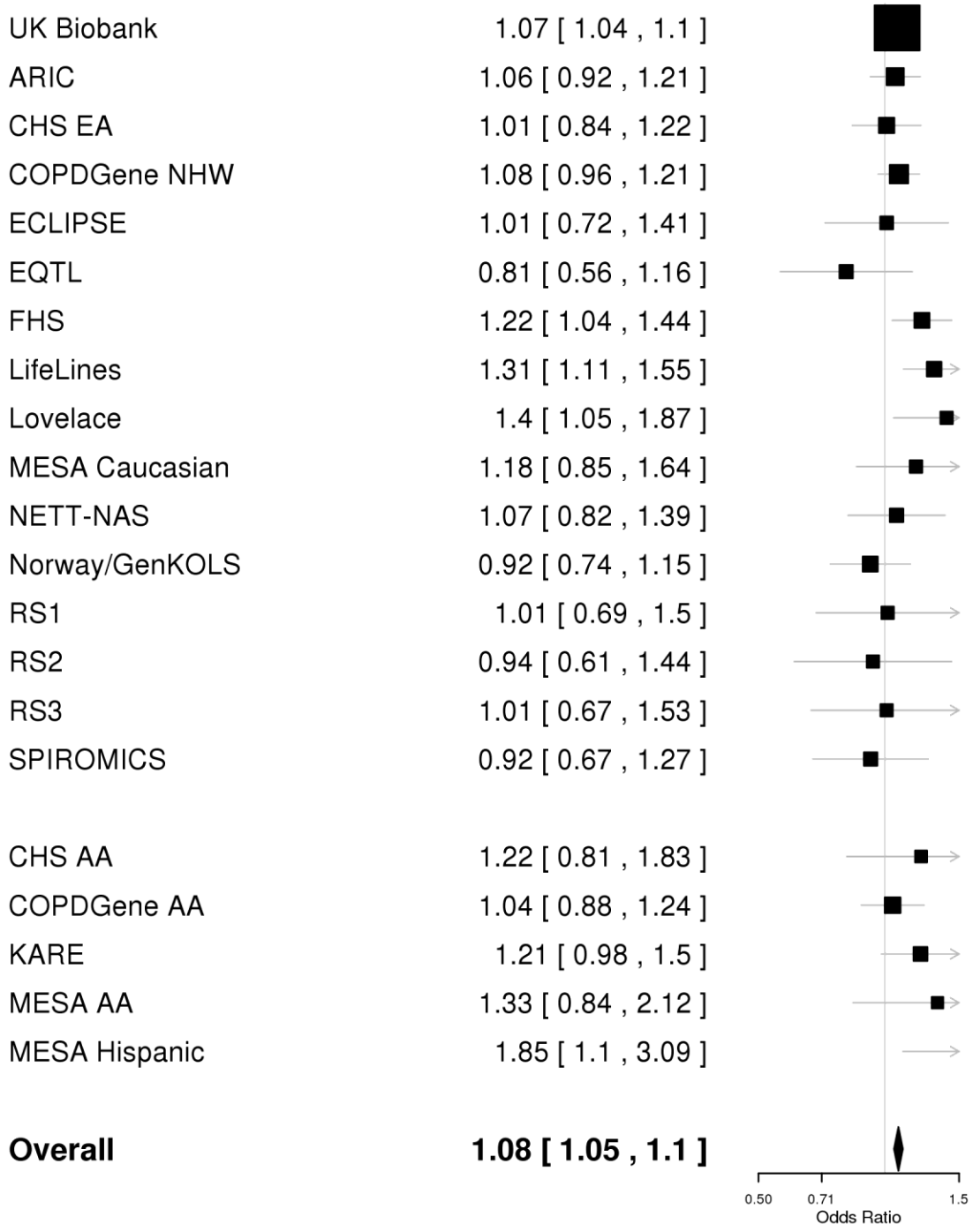
Supplementary Figure 1-62: Forest plot for rs7307510 (*SNRPF* locus at 12q23.1)



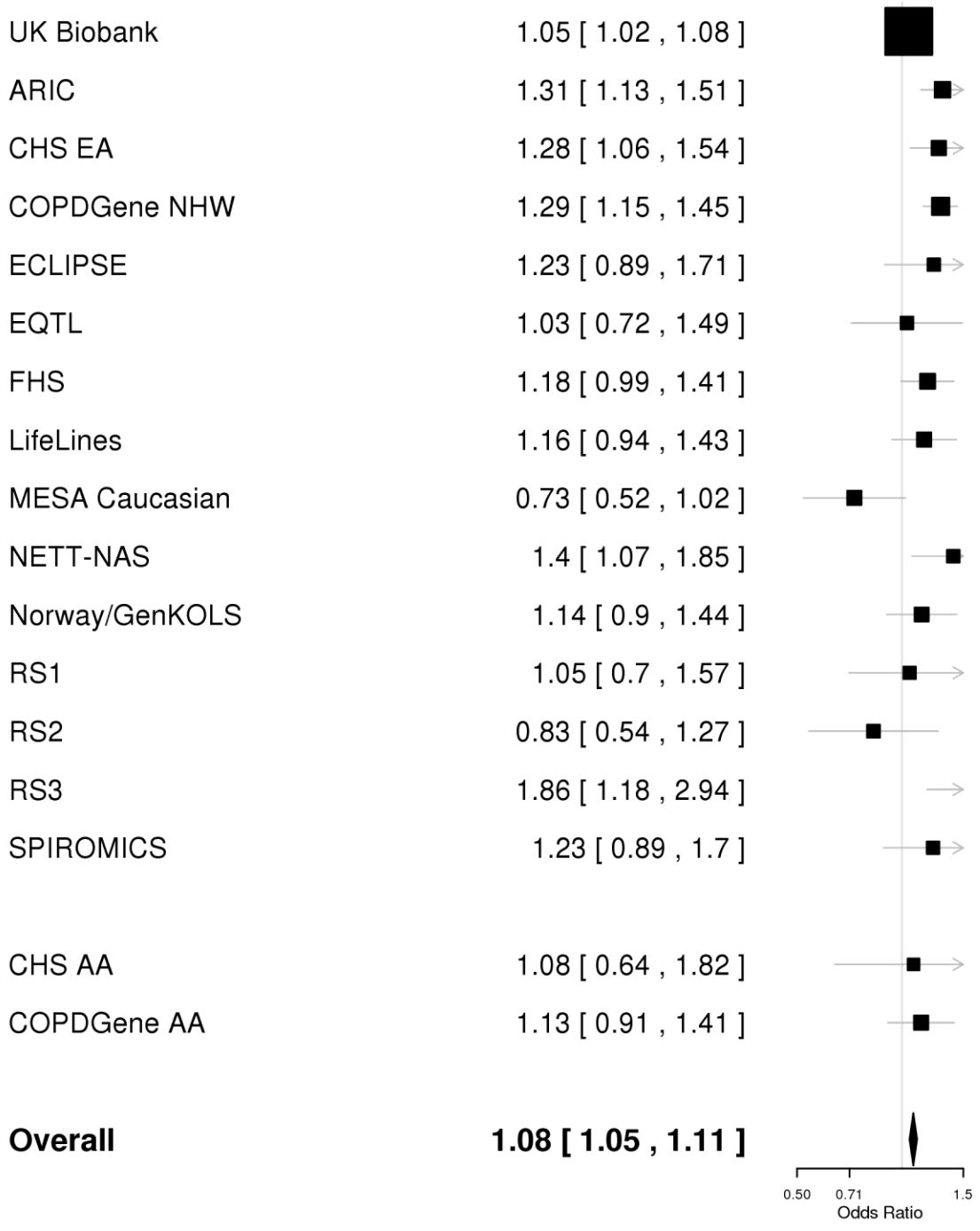
Supplementary Figure 1-63: Forest plot for rs7958945 (*MED13L* locus at 12q24.21)



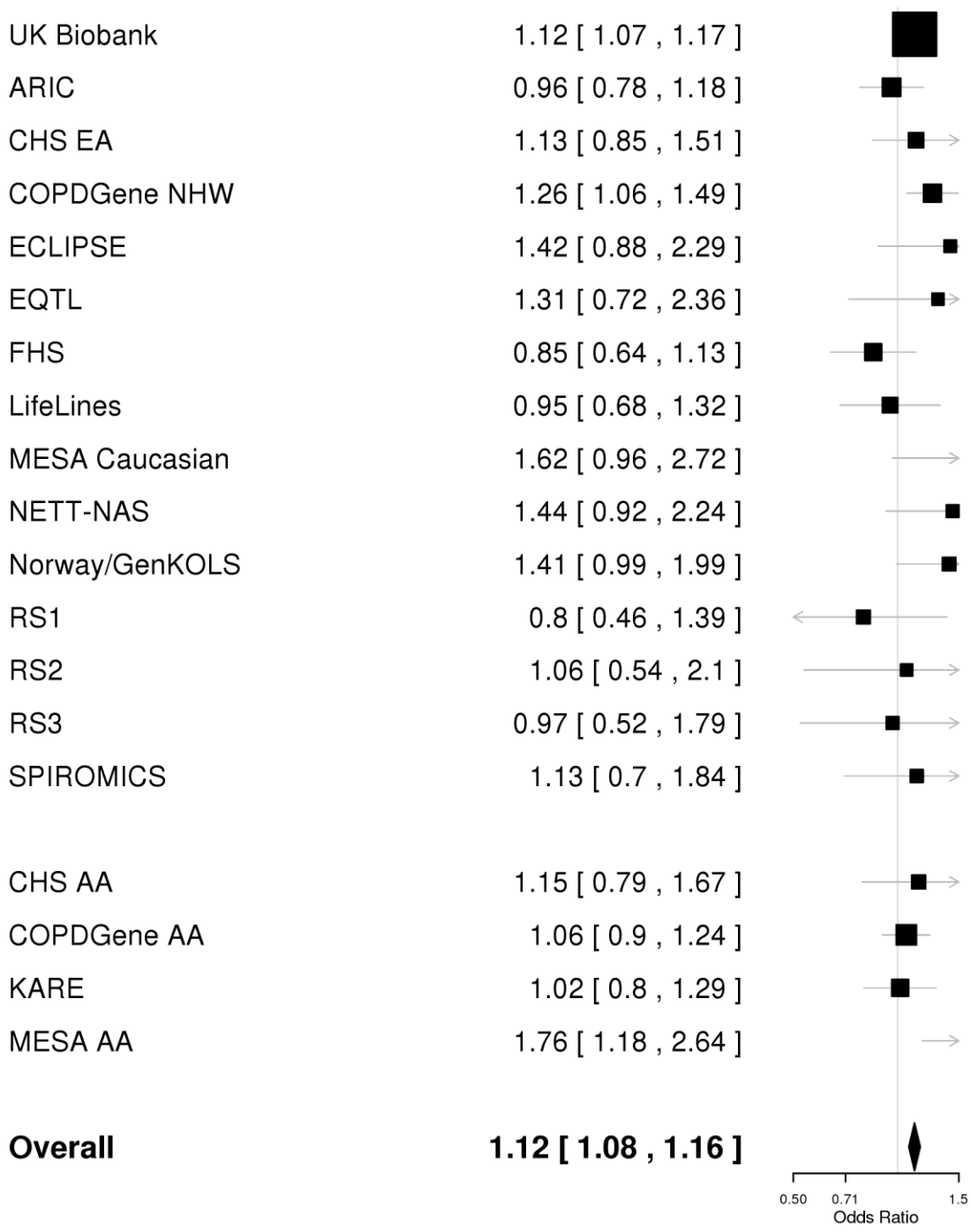
Supplementary Figure 1-64: Forest plot for rs9525927 (*SERP2* locus at 13q14.11)



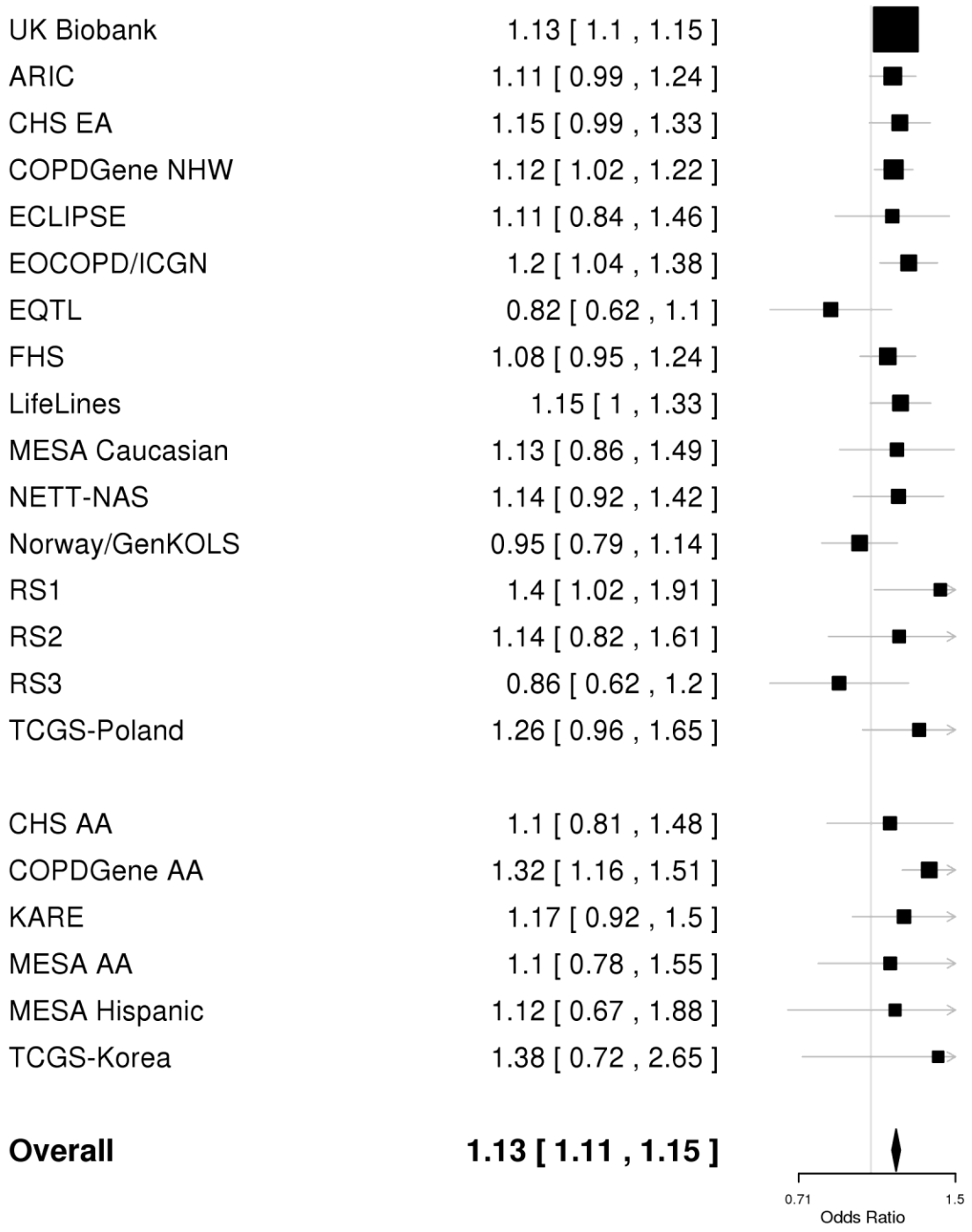
Supplementary Figure 1-65: Forest plot for rs72699855 (*RIN3* locus at 14q32.12)



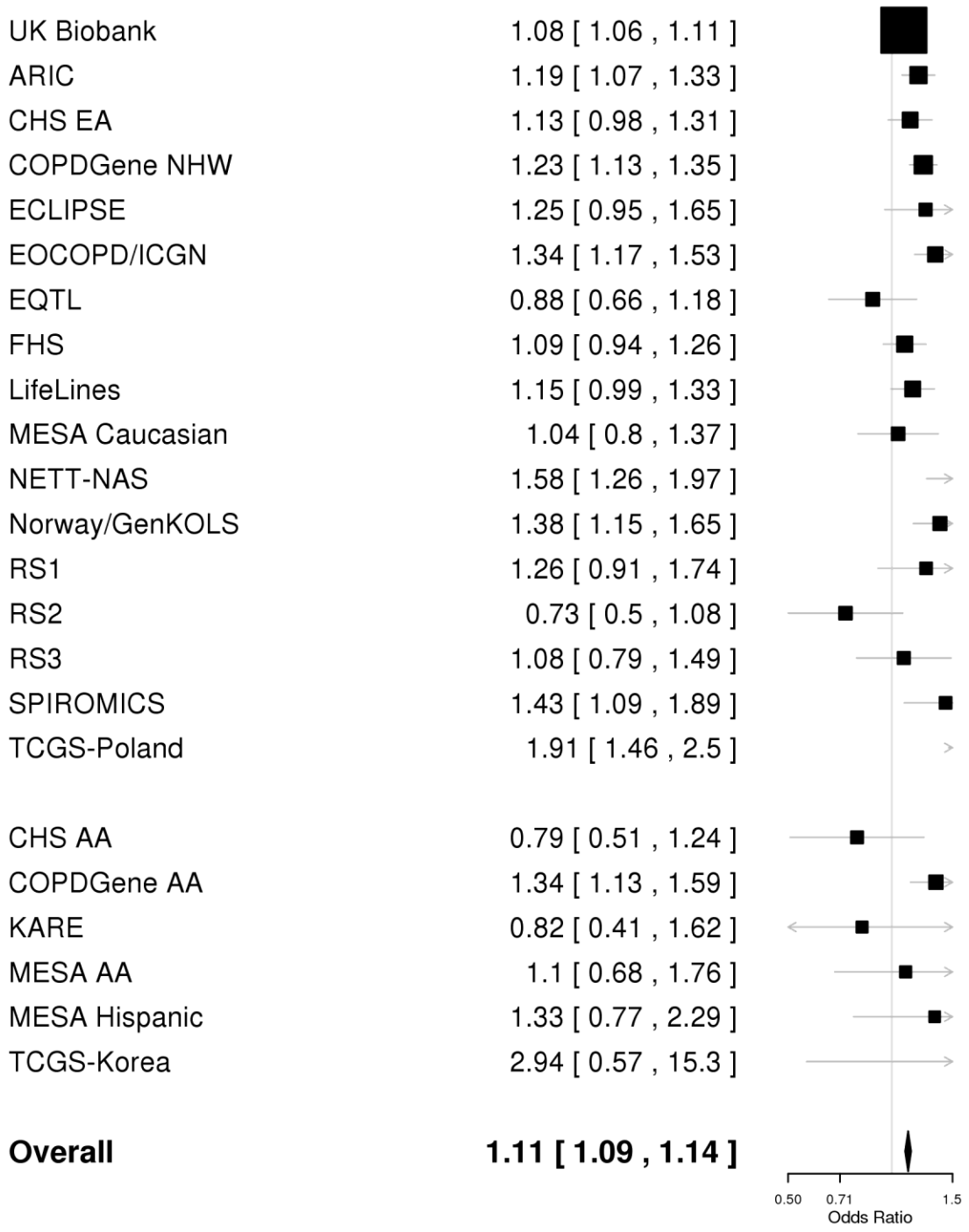
Supplementary Figure 1-66: Forest plot for rs72731149 (*DTWD1* locus at 15q21.2)



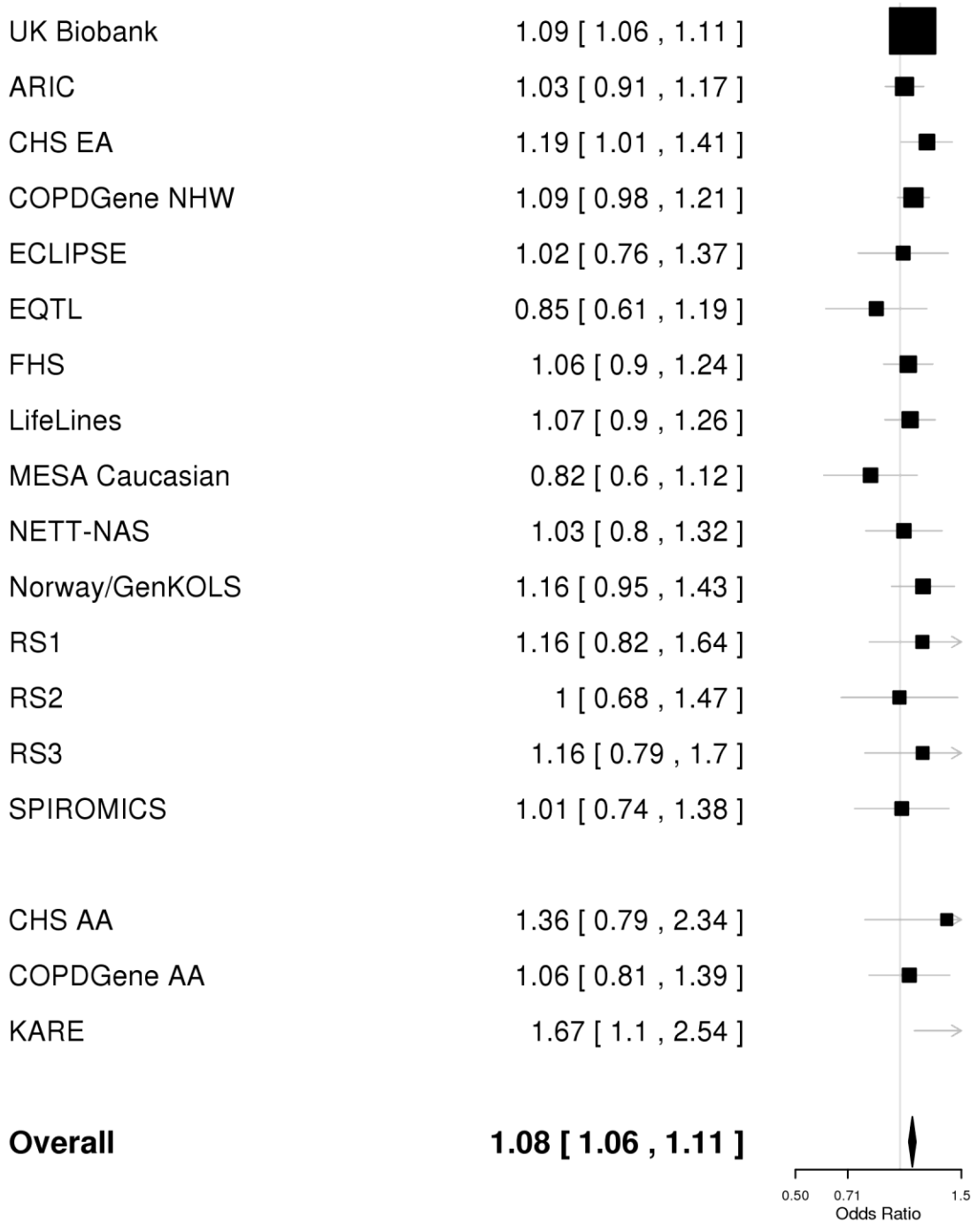
Supplementary Figure 1-67: Forest plot for rs1441358 (*THSD4* locus at 15q23)



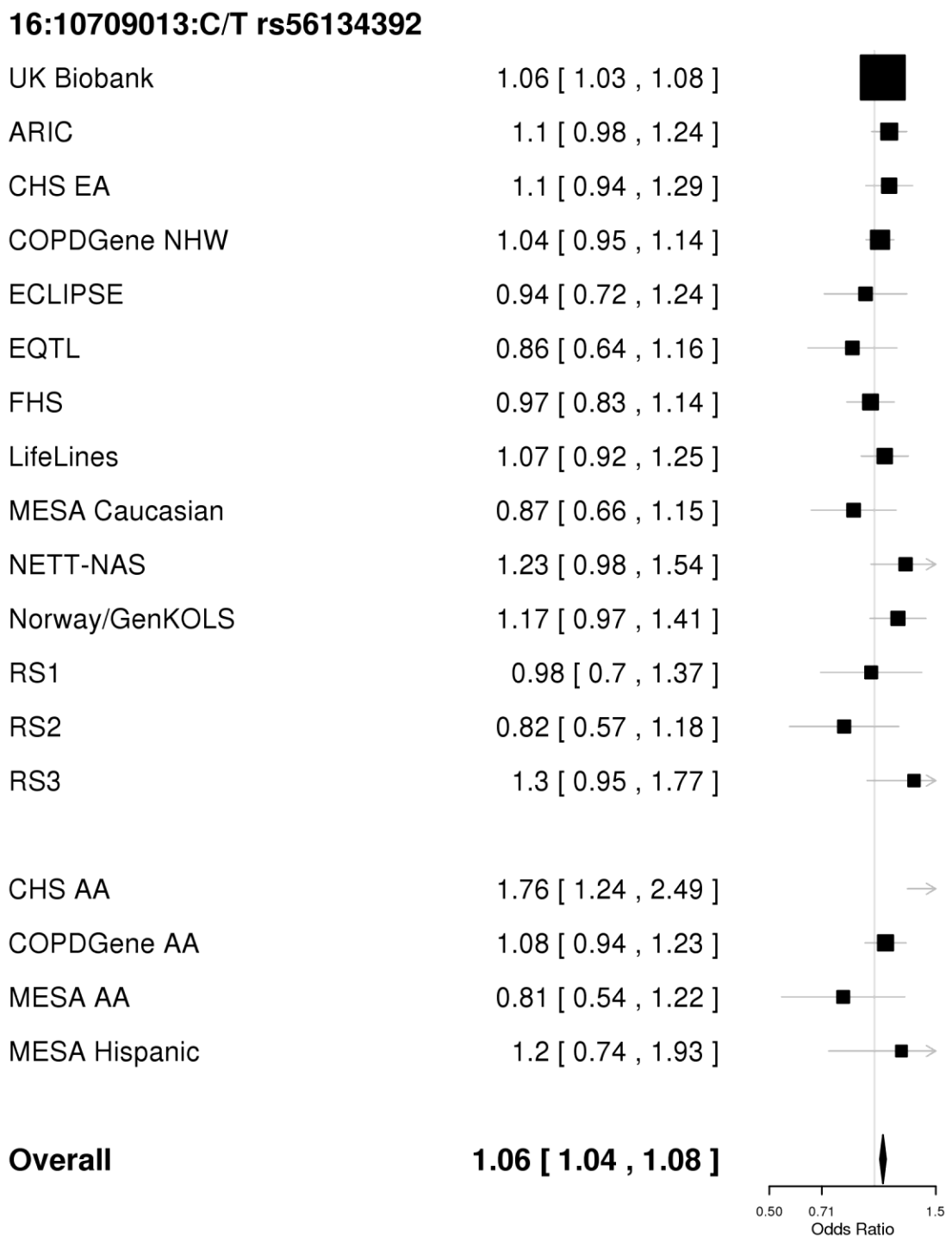
Supplementary Figure 1-68: Forest plot for rs55676755 (*CHRNA3* locus at 15q25.1)



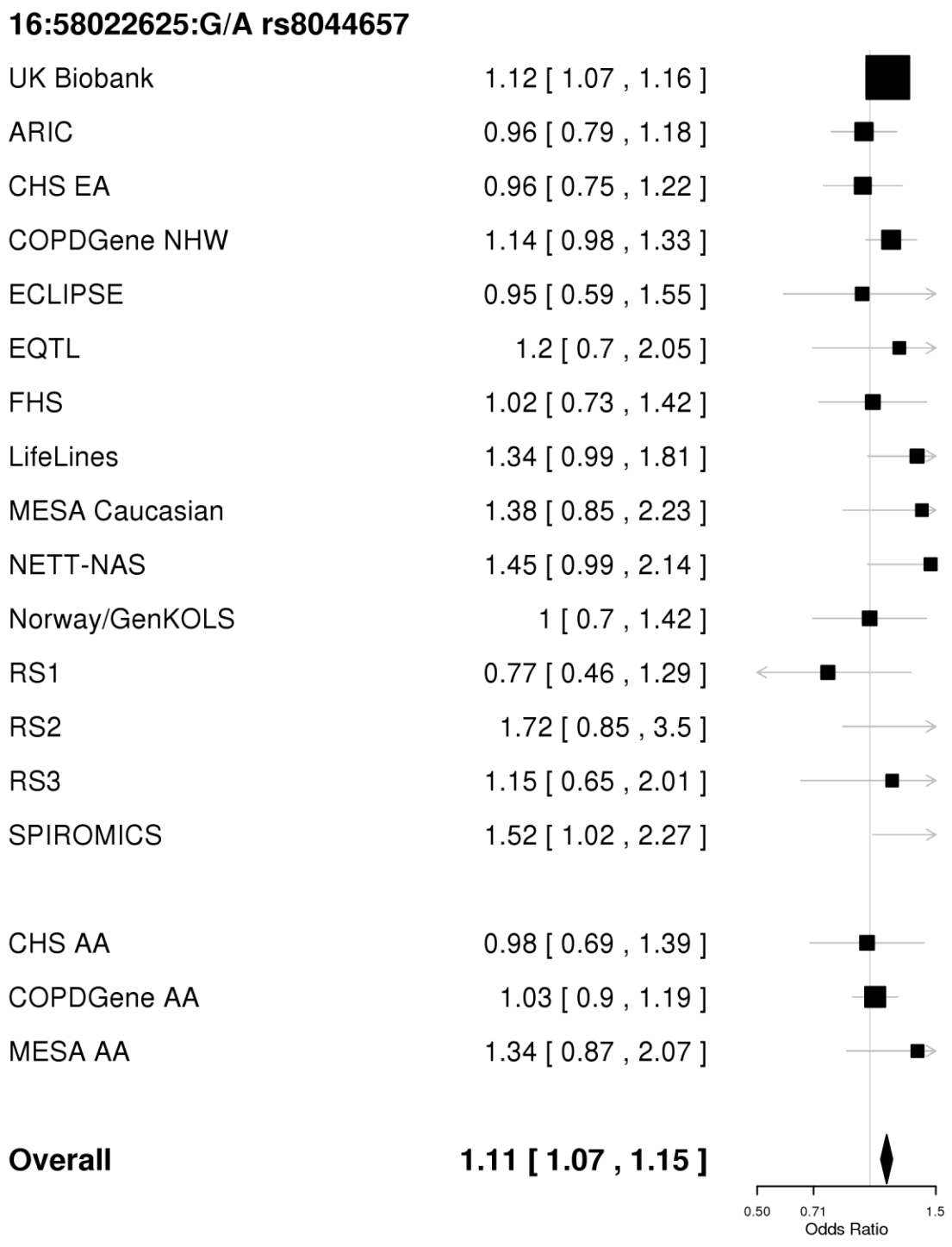
Supplementary Figure 1-69: Forest plot for rs10152300 (*ADAMTSL3* locus at 15q25.2)



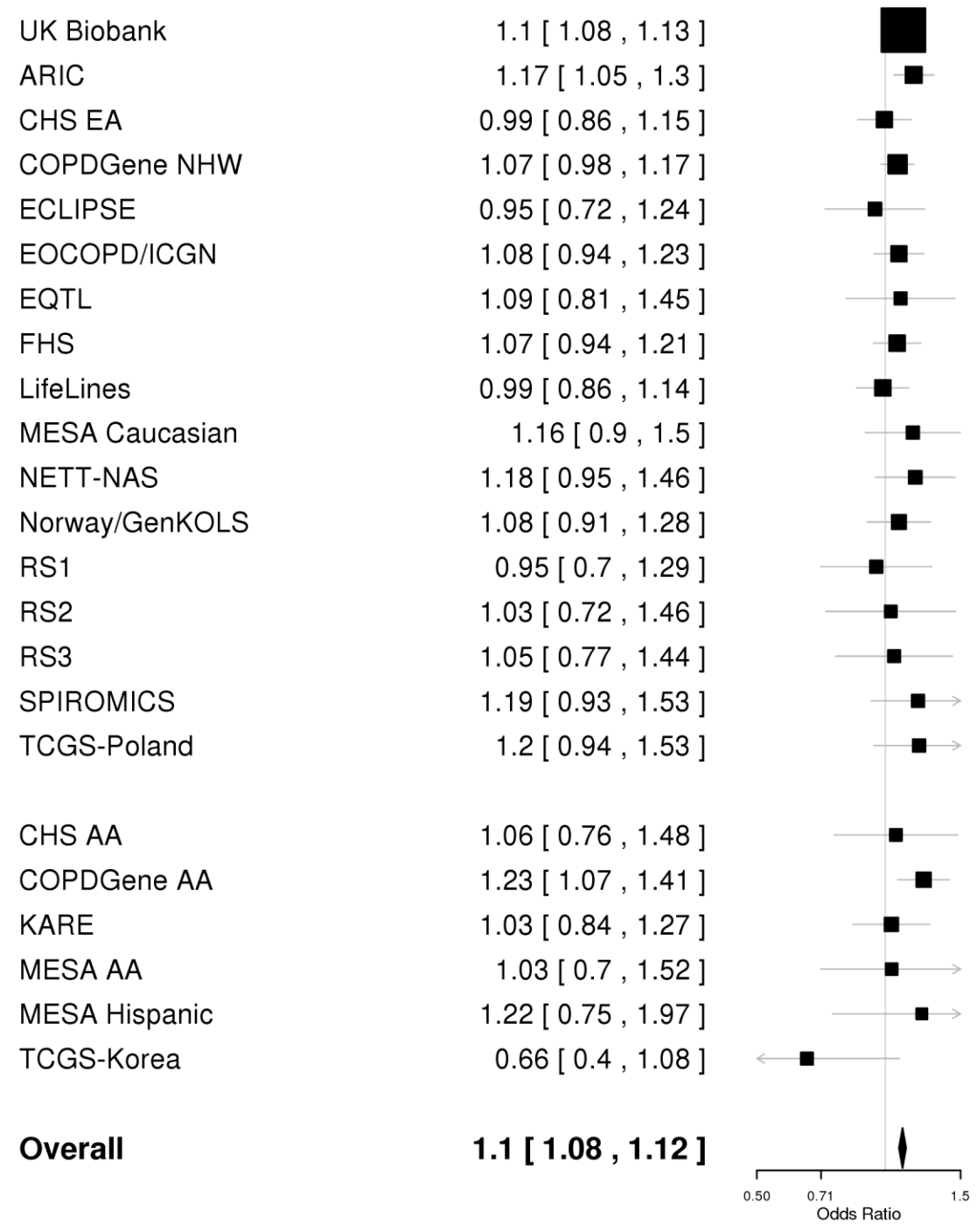
Supplementary Figure 1-70: Forest plot for rs56134392 (*TEKT5* locus at 16p13.13)



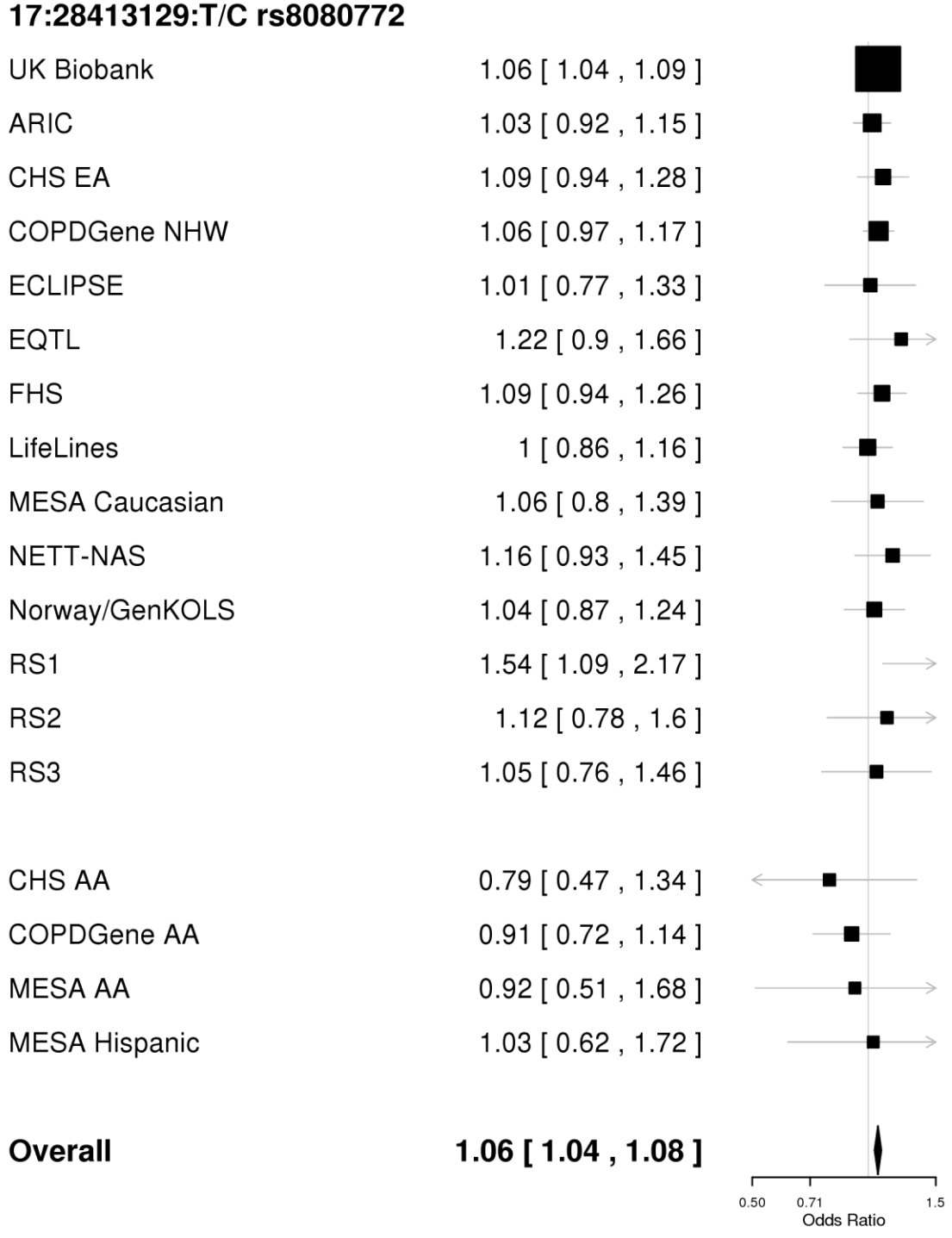
Supplementary Figure 1-71: Forest plot for rs8044657 (*TEPP* locus at 16q21)



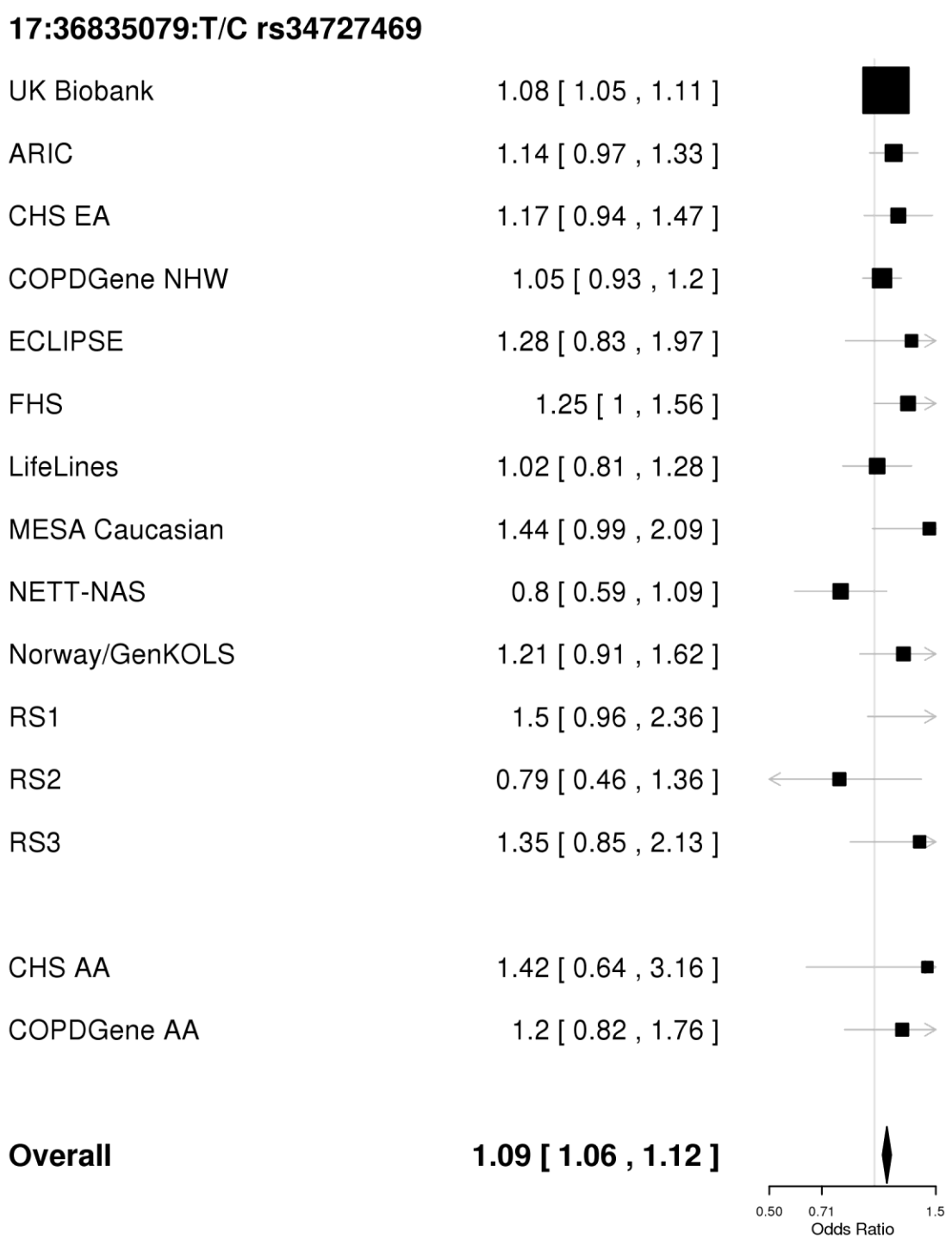
Supplementary Figure 1-72: Forest plot for rs4888379 (*CFDP1* locus at 16q23.1)



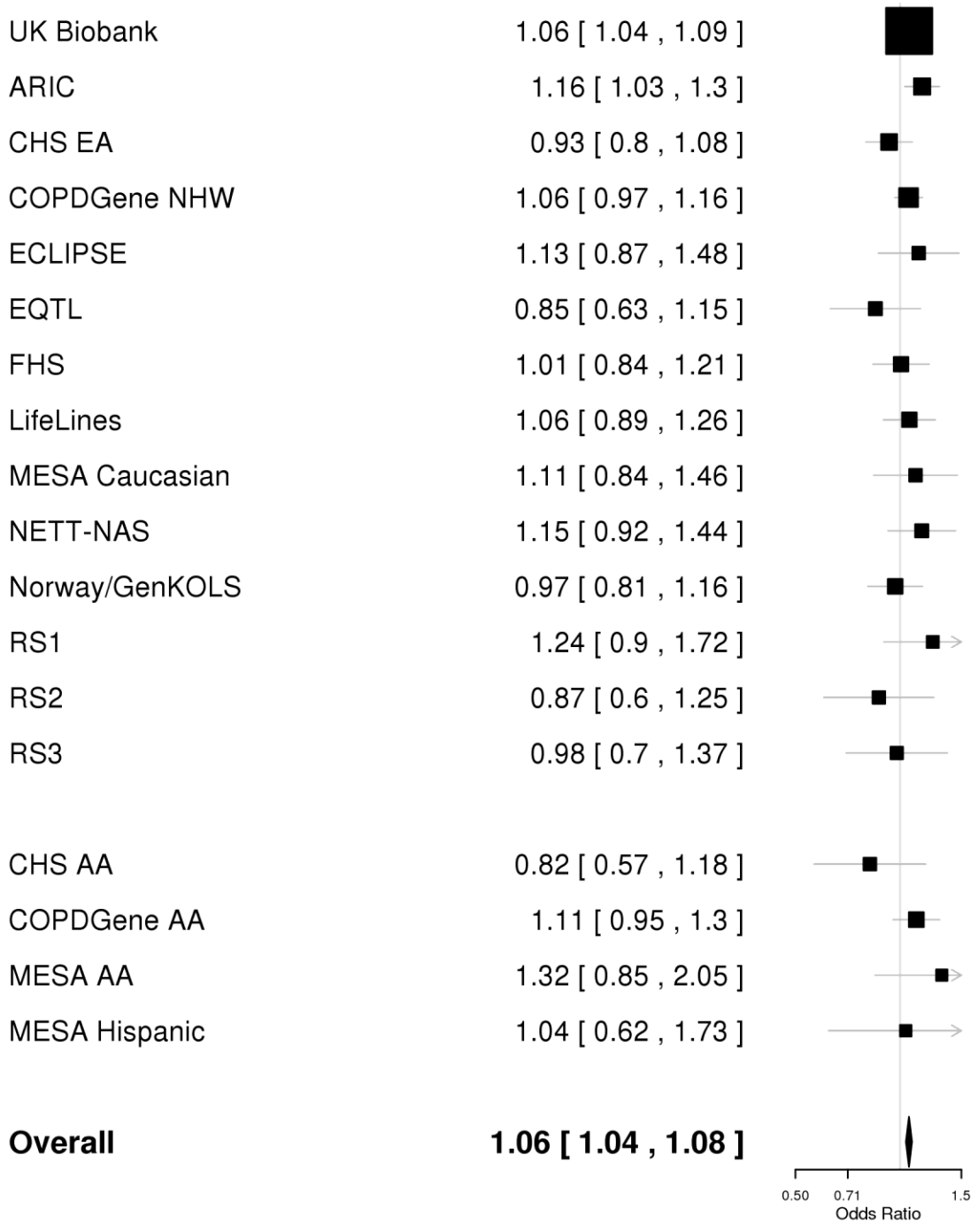
Supplementary Figure 1-73: Forest plot for rs8080772 (*EFCAB5* locus at 17q11.2)



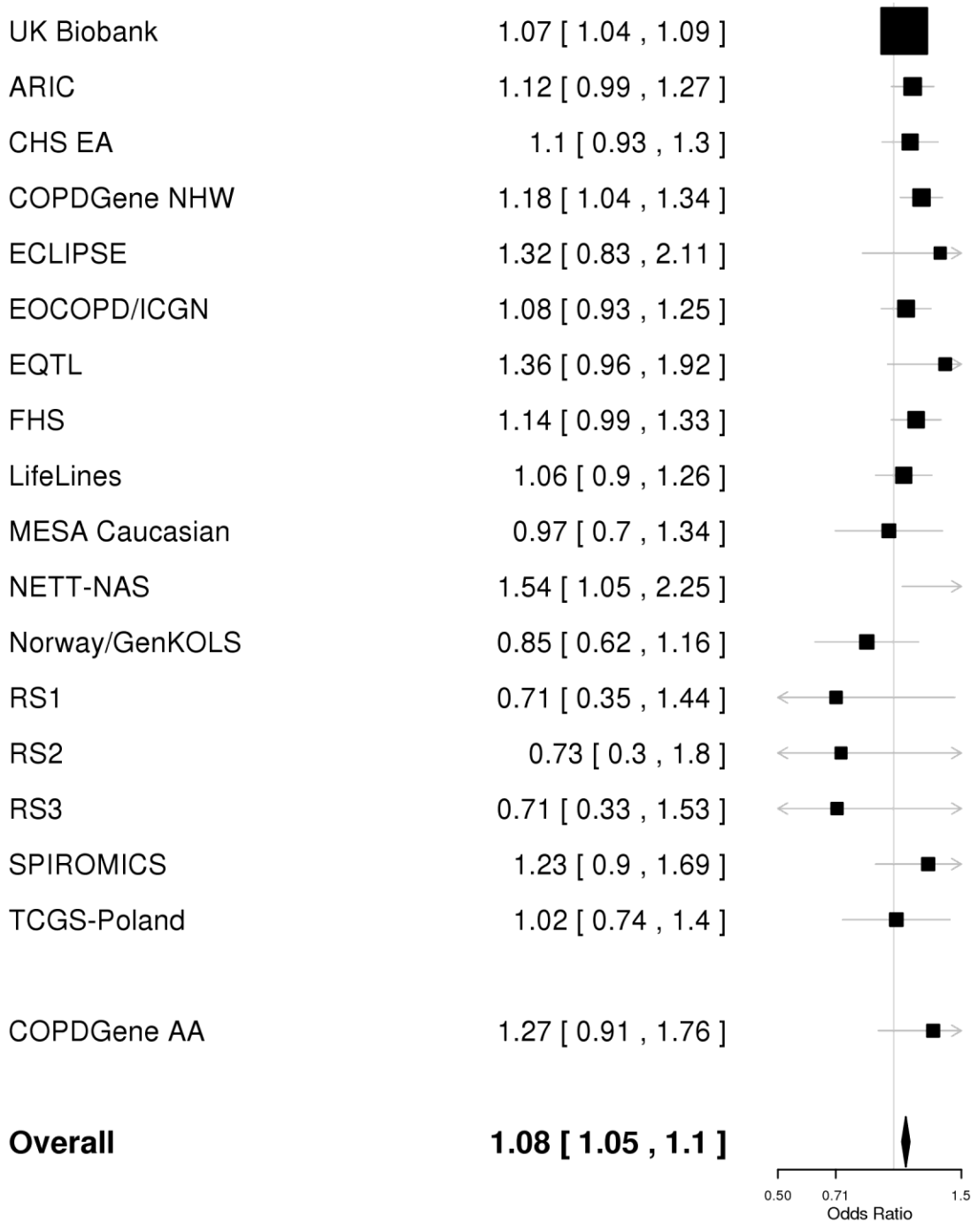
Supplementary Figure 1-74: Forest plot for rs34727469 (*RPL23* locus at 17q12)



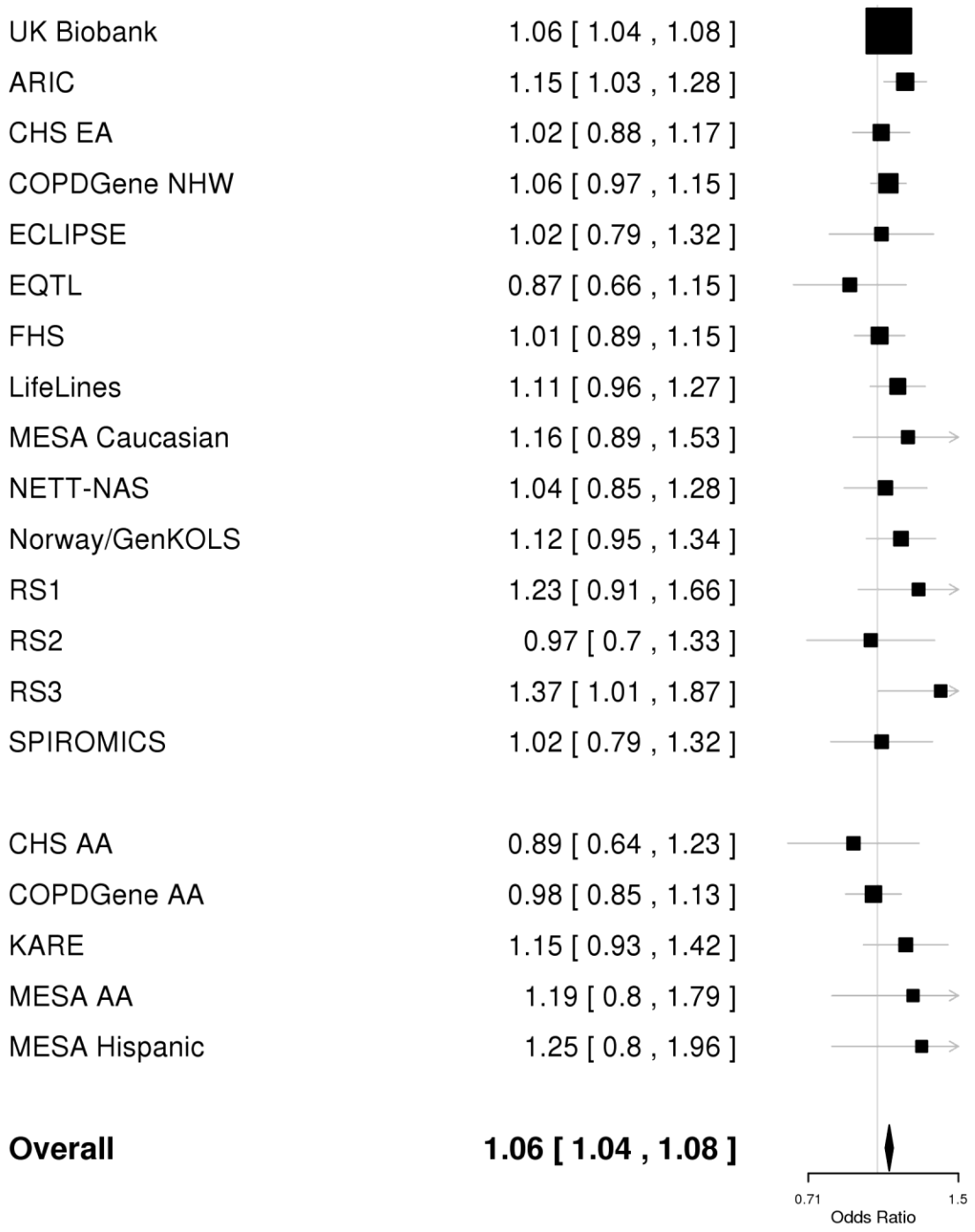
Supplementary Figure 1-75: Forest plot for rs62065216 (*THRA* locus at 17q21.1)



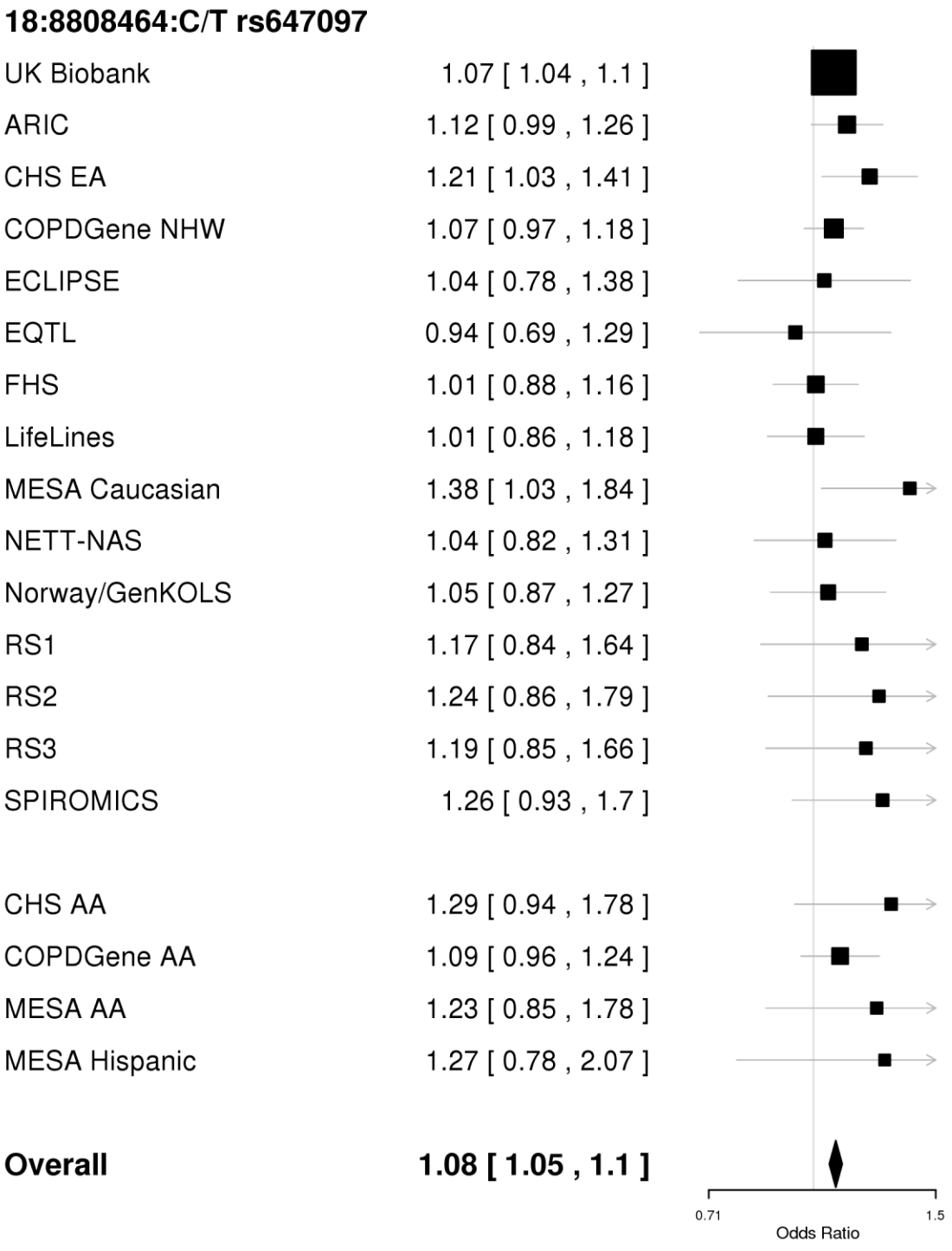
Supplementary Figure 1-76: Forest plot for rs12373142 (*SPPL2C* locus at 17q21.31)



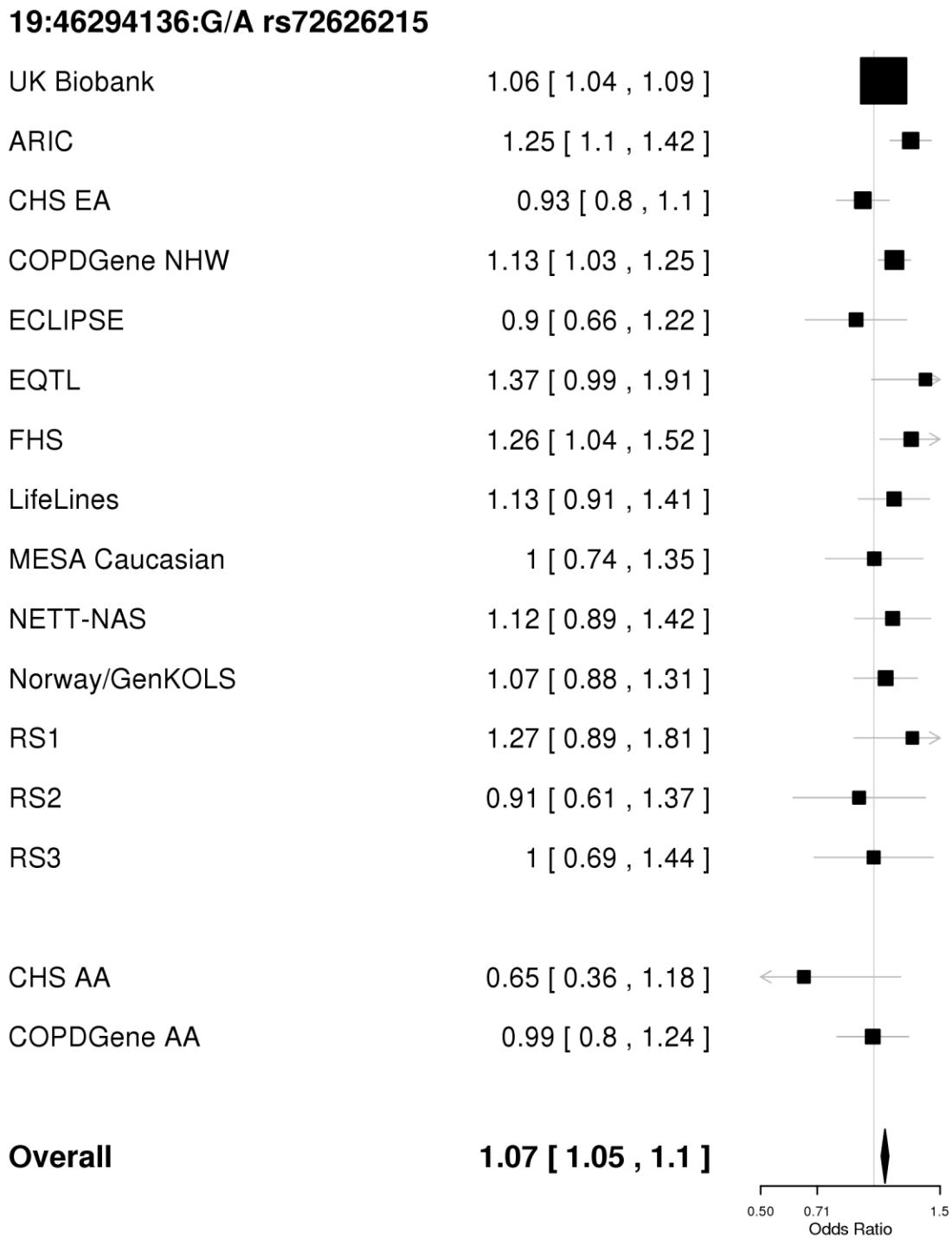
Supplementary Figure 1-77: Forest plot for rs11655567 (*SOX9* locus at 17q24.3)



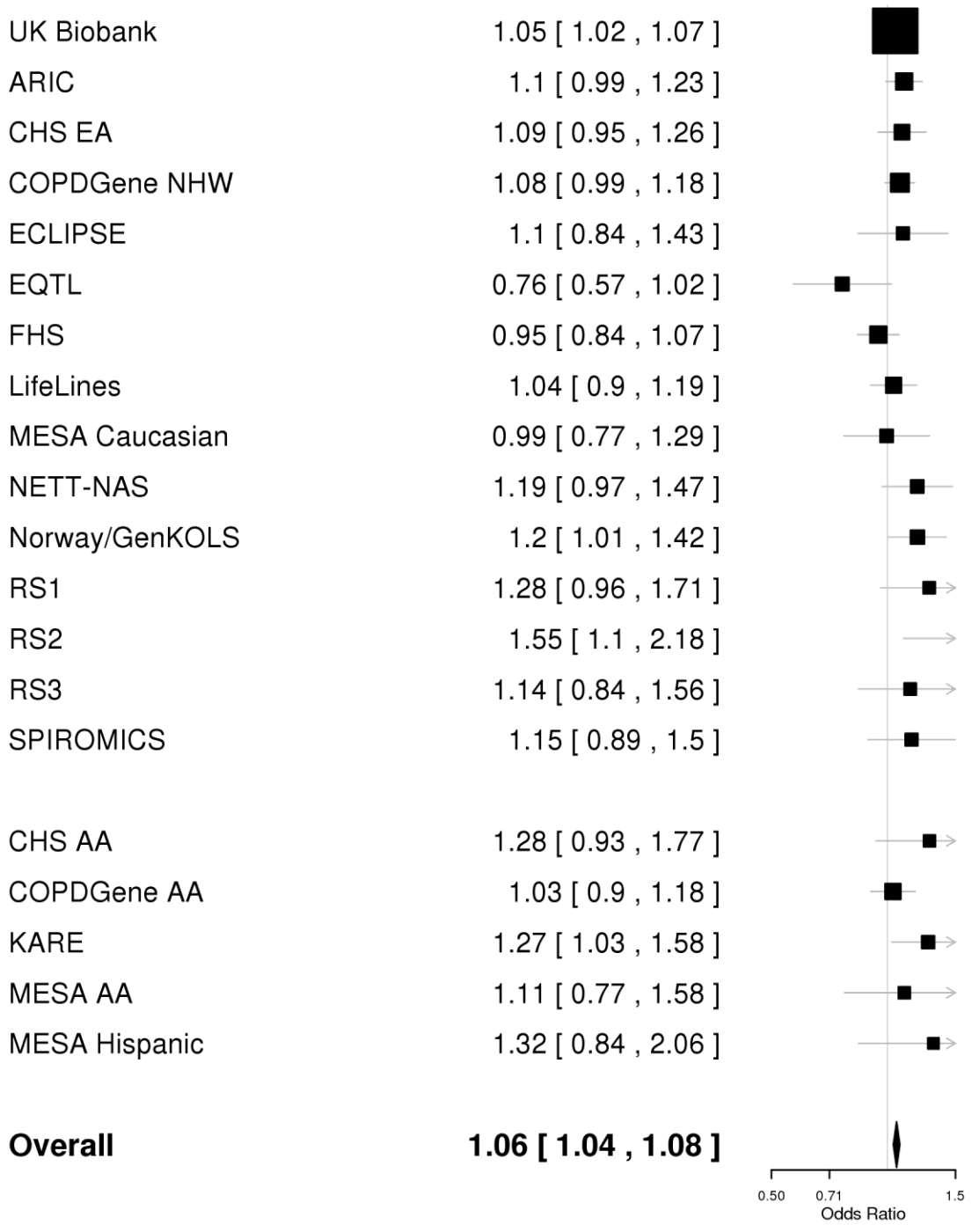
Supplementary Figure 1-78: Forest plot for rs647097 (*MTCL1* locus at 18p11.22)



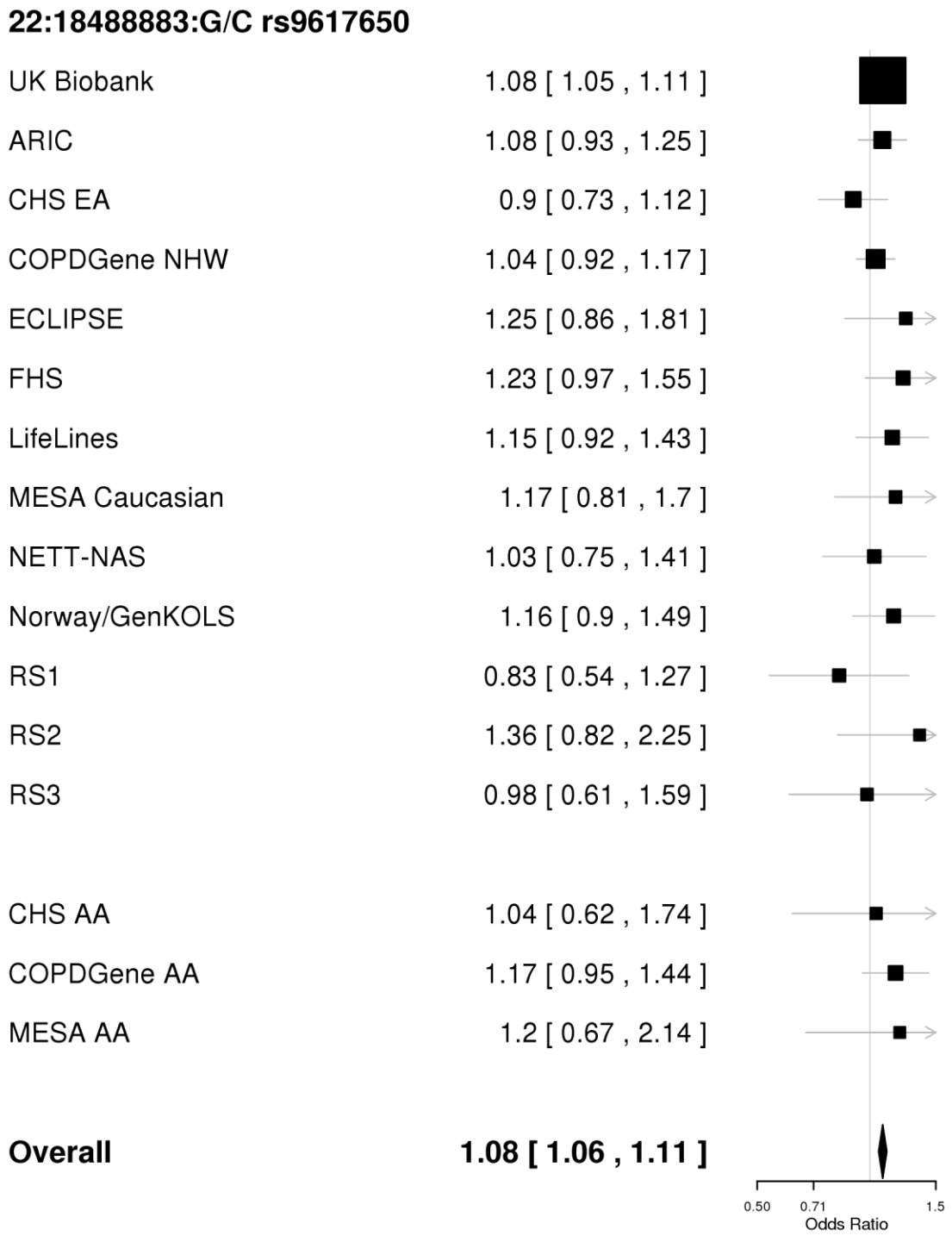
Supplementary Figure 1-79: Forest plot for rs72626215 (*DMWD* locus at 19q13.32)



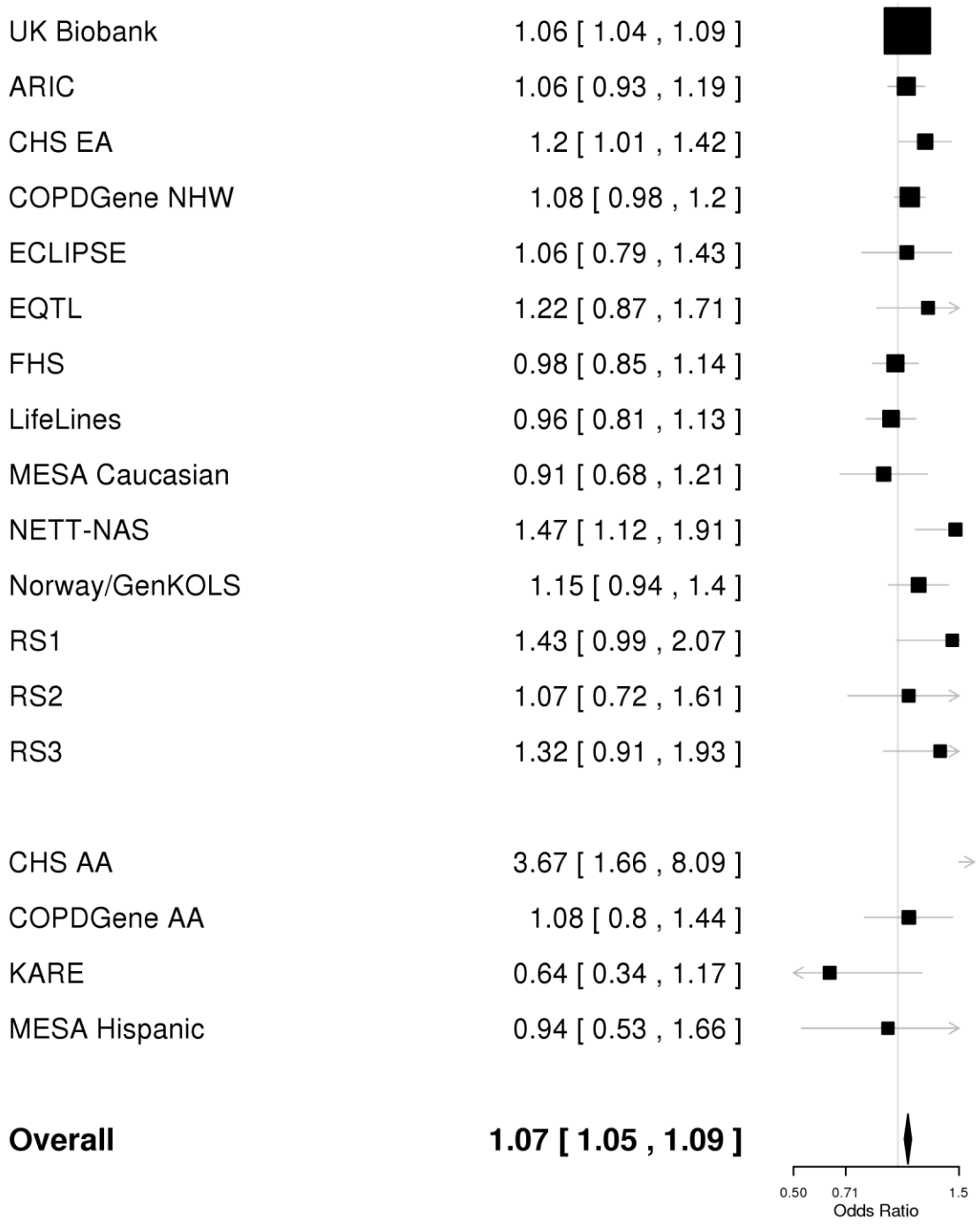
Supplementary Figure 1-80: Forest plot for rs2096468 (*KCNE2* locus at 21q22.11)



Supplementary Figure 1-81: Forest plot for rs9617650 (*MICAL3* locus at 22q11.21)



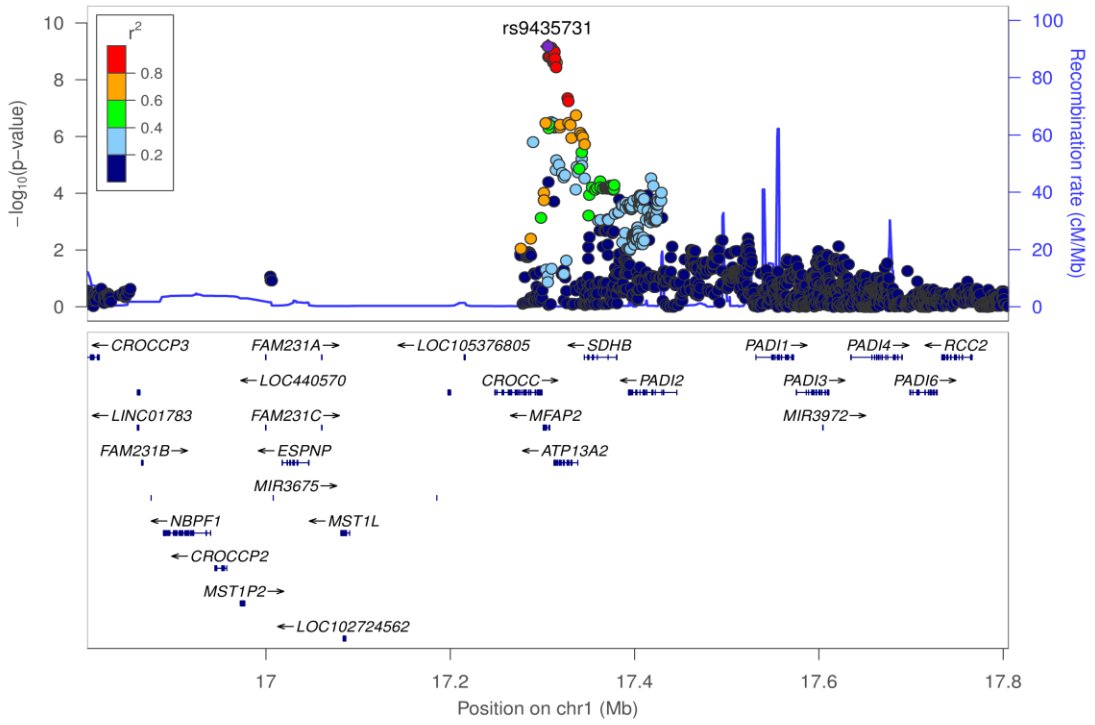
Supplementary Figure 1-82: Forest plot for rs73158393 (*SYN3* locus at 22q12.3)



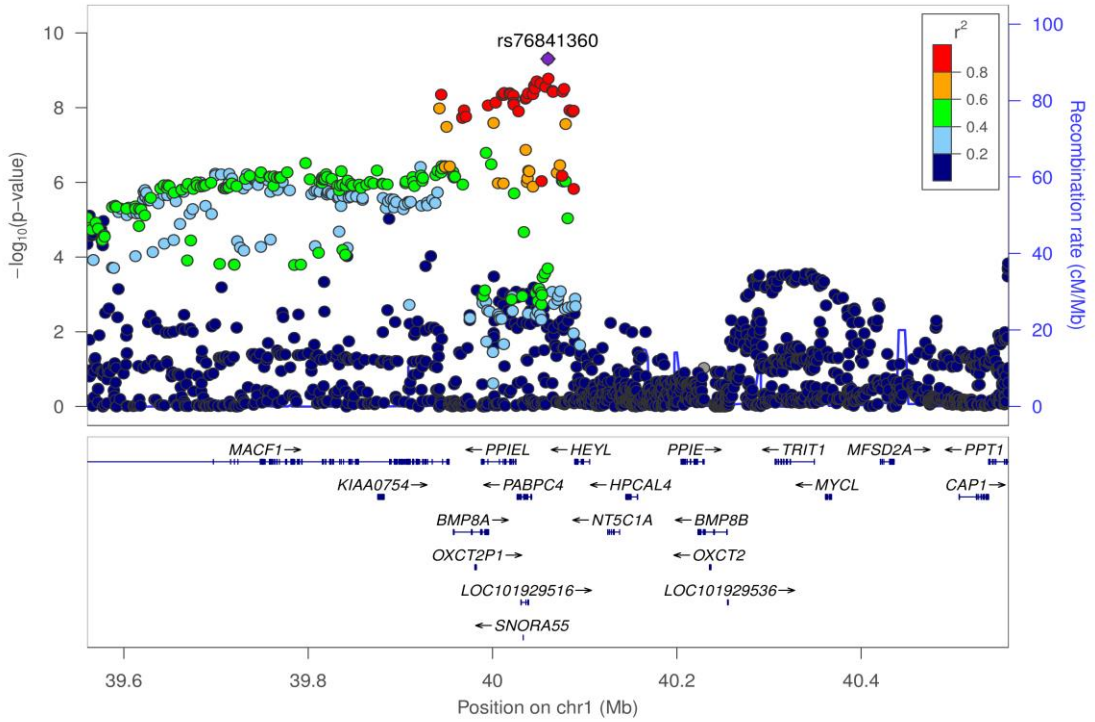
Supplementary Figure 2: Regional association plots for 82 genome-wide significant associations

Association statistics are based on the overall meta-analysis of COPD (35,735 cases and 222,076 controls). P values are two-sided based on Wald statistics without multiple comparison adjustment.

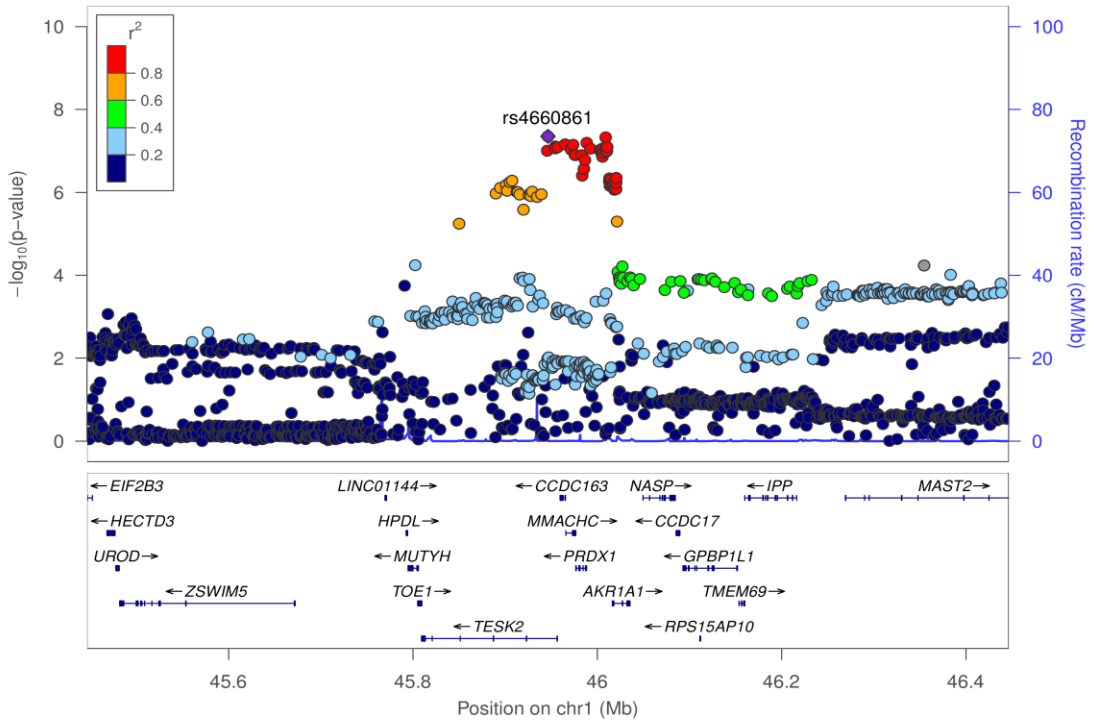
Supplementary Figure 2-1: Regional association plot for rs9435731 (*MFAP2* locus at 1p36.13)



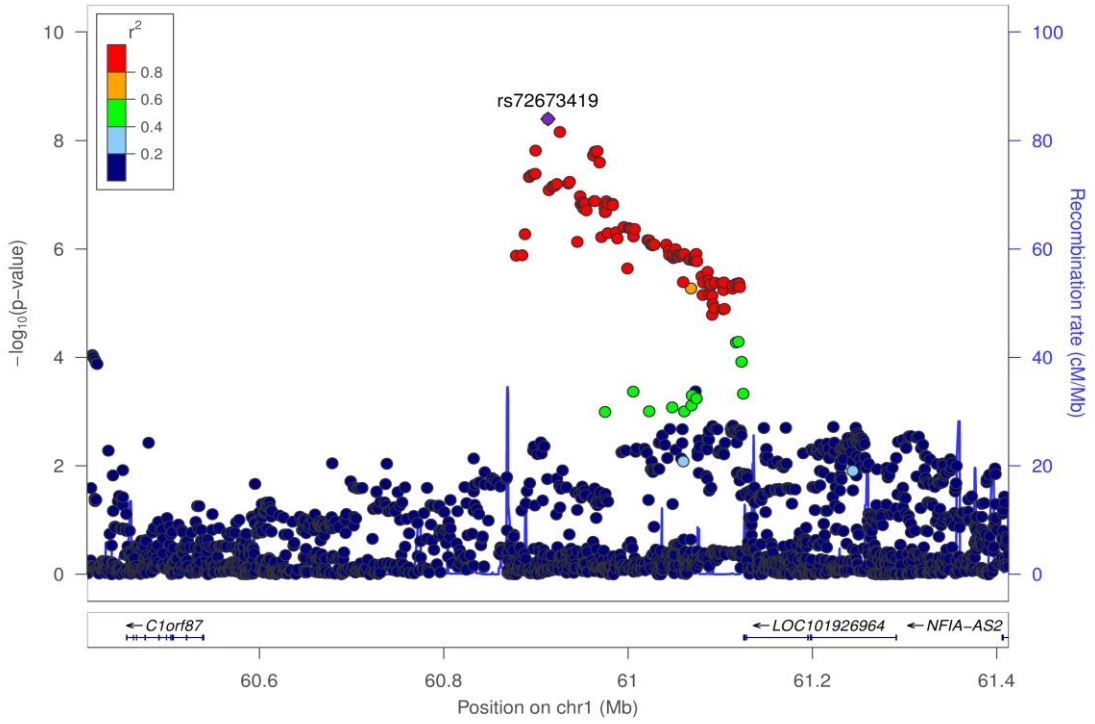
Supplementary Figure 2-2: Regional association plot for rs76841360 (*PABPC4* locus at 1p34.3)



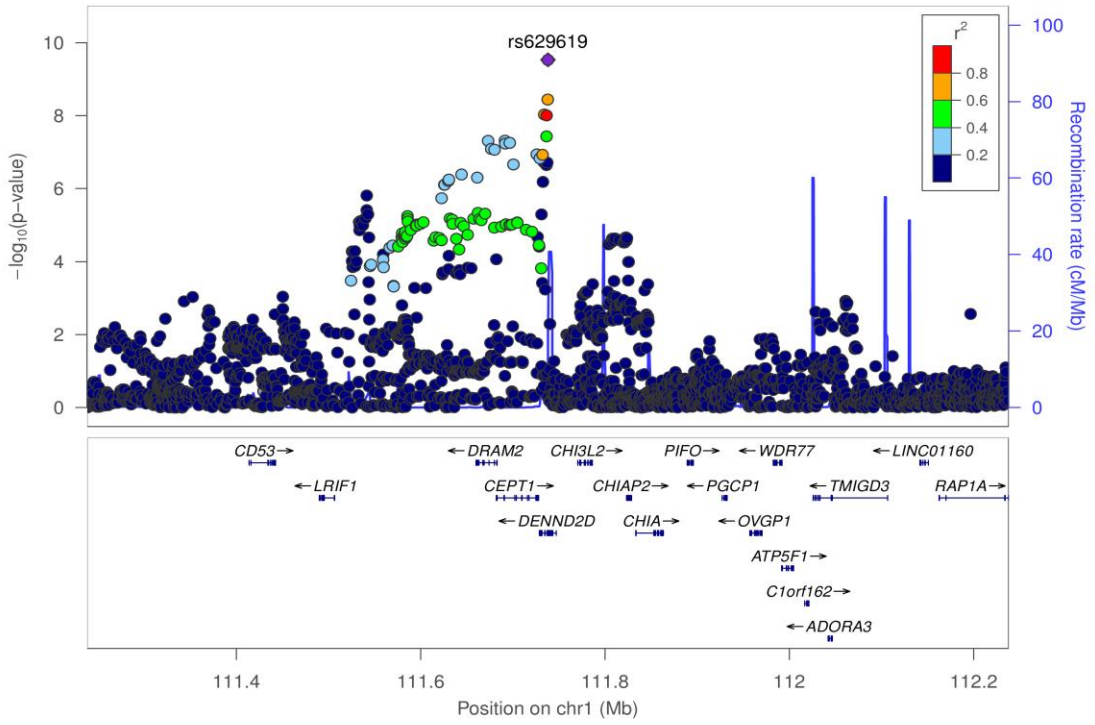
Supplementary Figure 2-3: Regional association plot for rs4660861 (*TESK2* locus at 1p34.1)



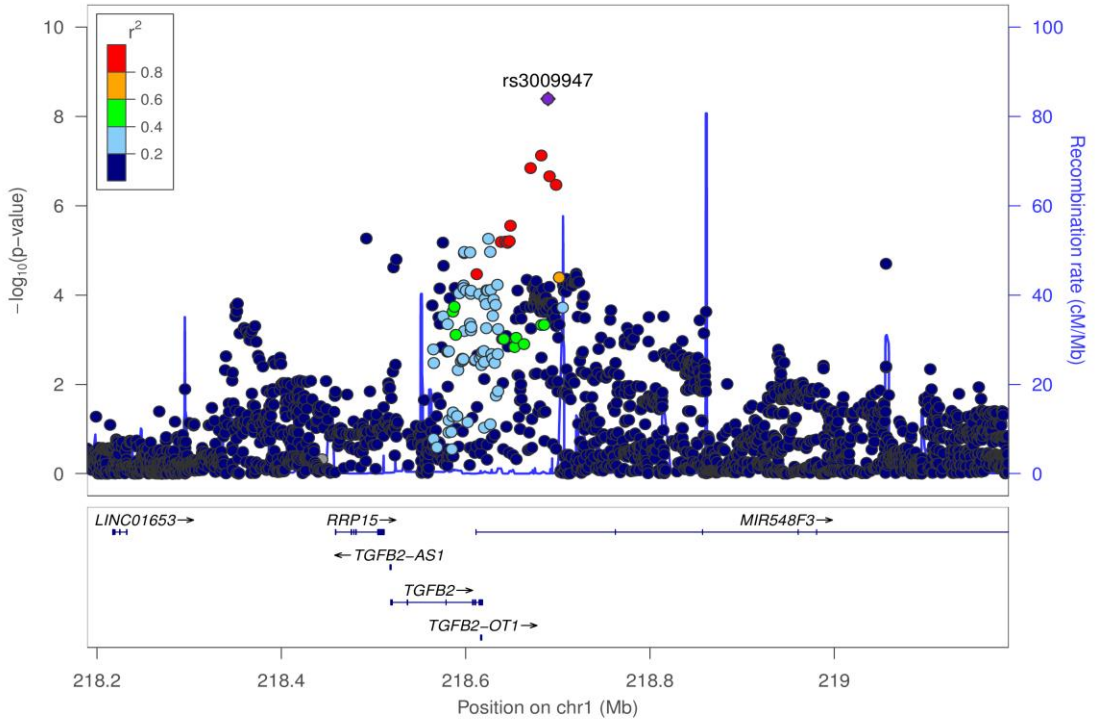
Supplementary Figure 2-4: Regional association plot for rs72673419 (*C1orf87* locus at 1p32.1)



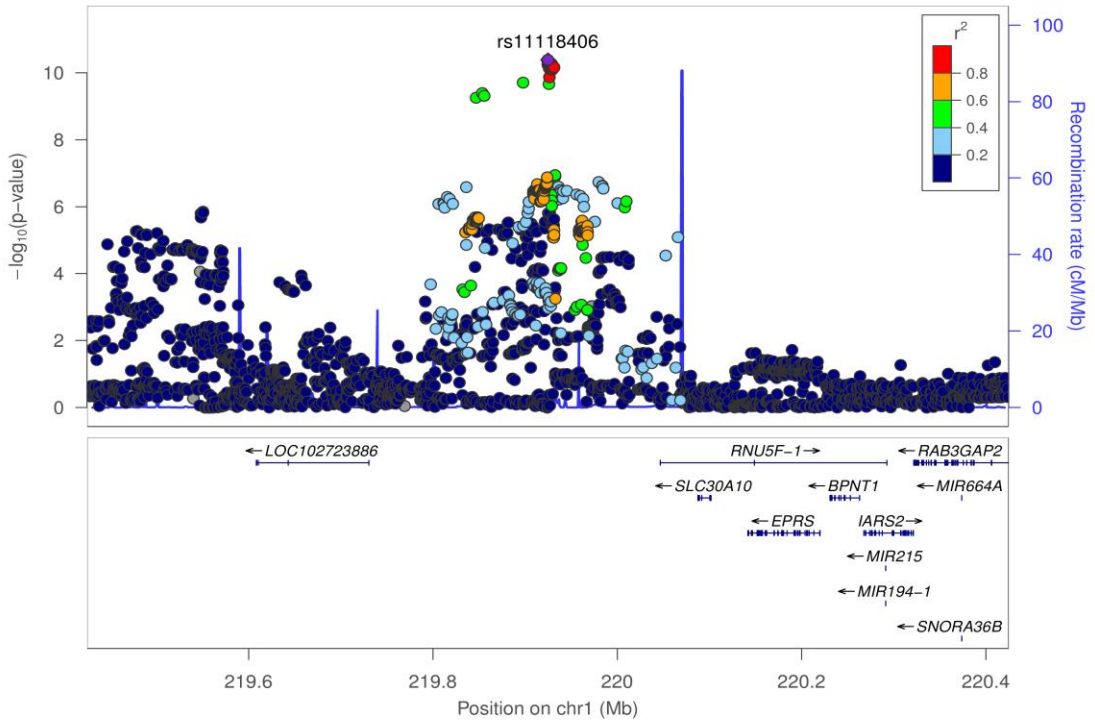
Supplementary Figure 2-5: Regional association plot for rs629619 (*DENND2D* locus at 1p13.3)



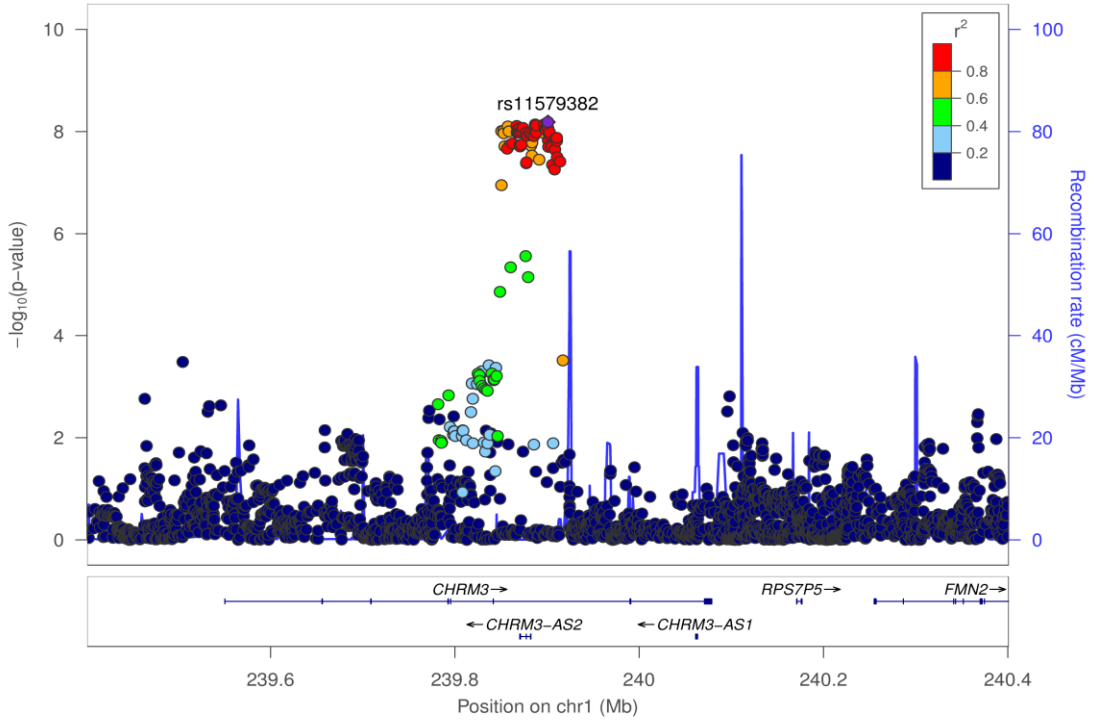
Supplementary Figure 2-6: Regional association plot for rs3009947 (*TGFB2* locus at 1q41)



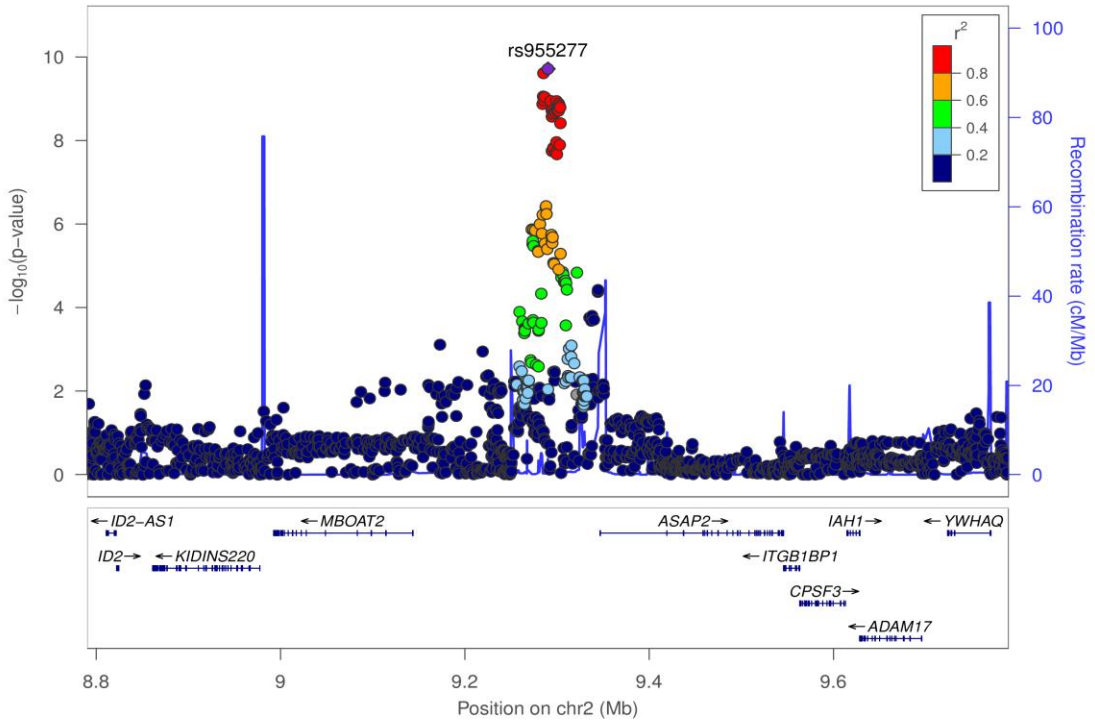
Supplementary Figure 2-7: Regional association plot for rs11118406 (*SLC30A10* locus at 1q41)



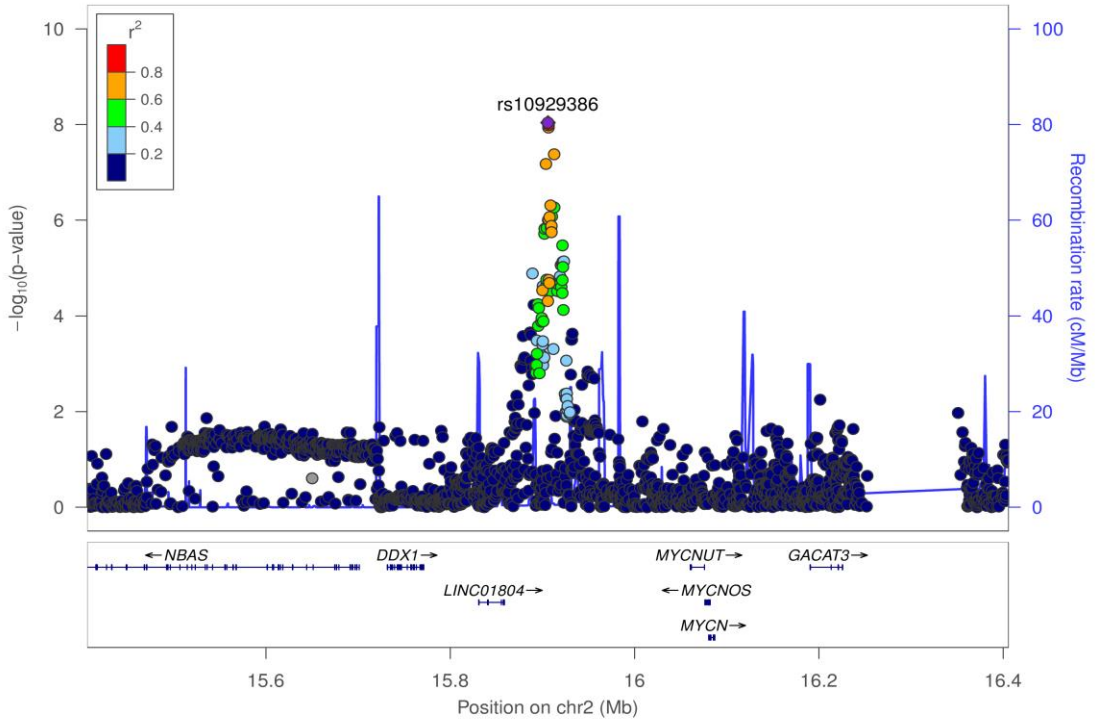
Supplementary Figure 2-8: Regional association plot for rs11579382 (*CHRM3* locus at 1q43)



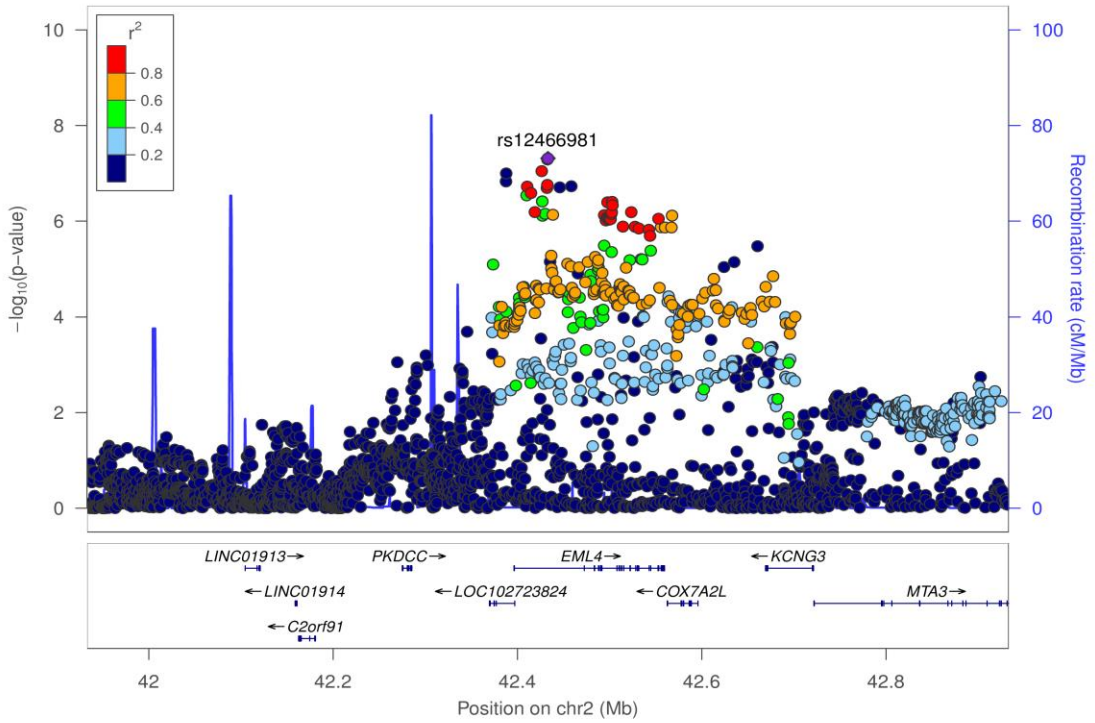
Supplementary Figure 2-9: Regional association plot for rs955277 (*ASAP2* locus at 2p25.1)



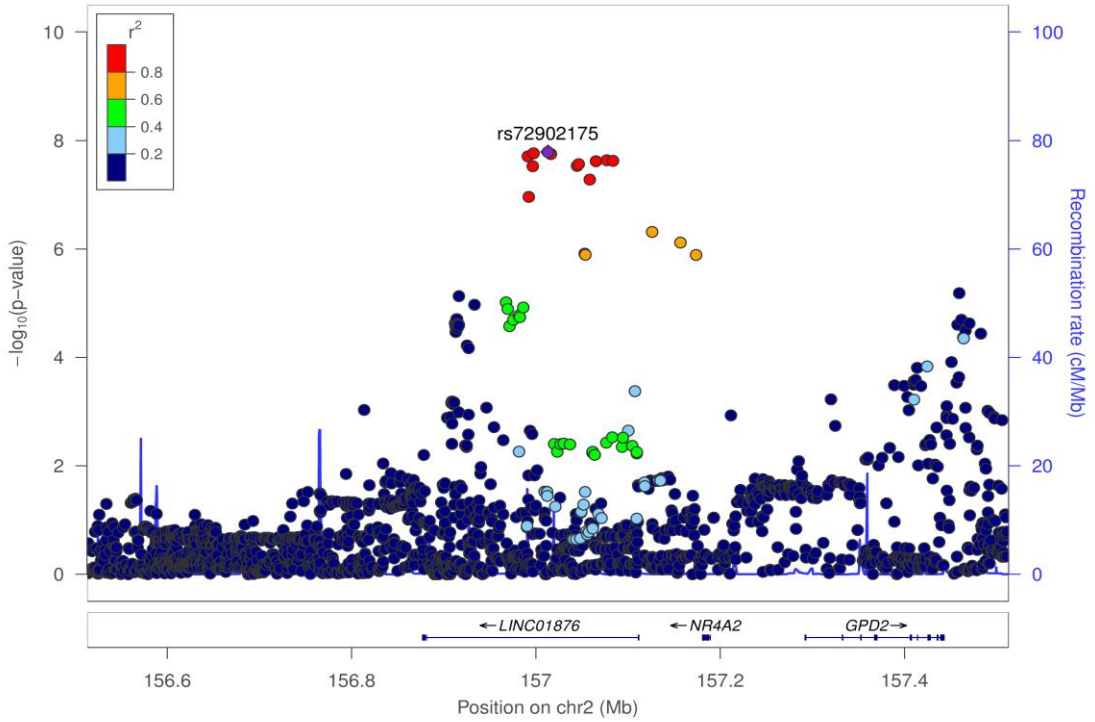
Supplementary Figure 2-10: Regional association plot for rs10929386 (*DDX1* locus at 2p24.3)



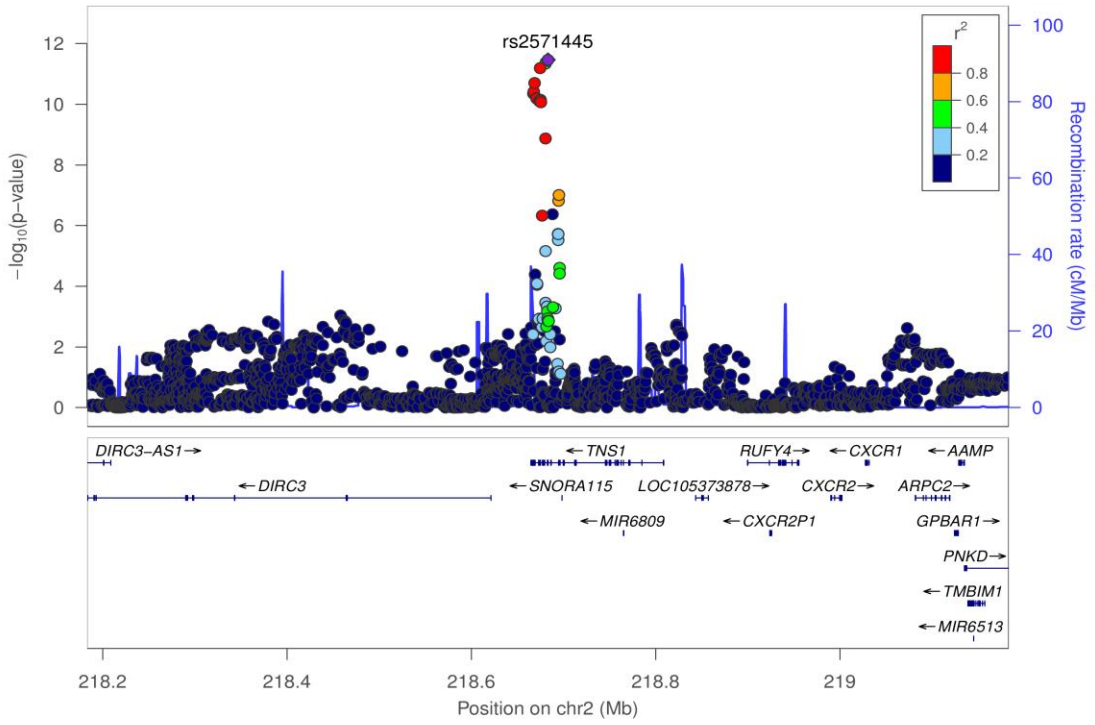
Supplementary Figure 2-11: Regional association plot for rs12466981 (*EML4* locus at 2p21)



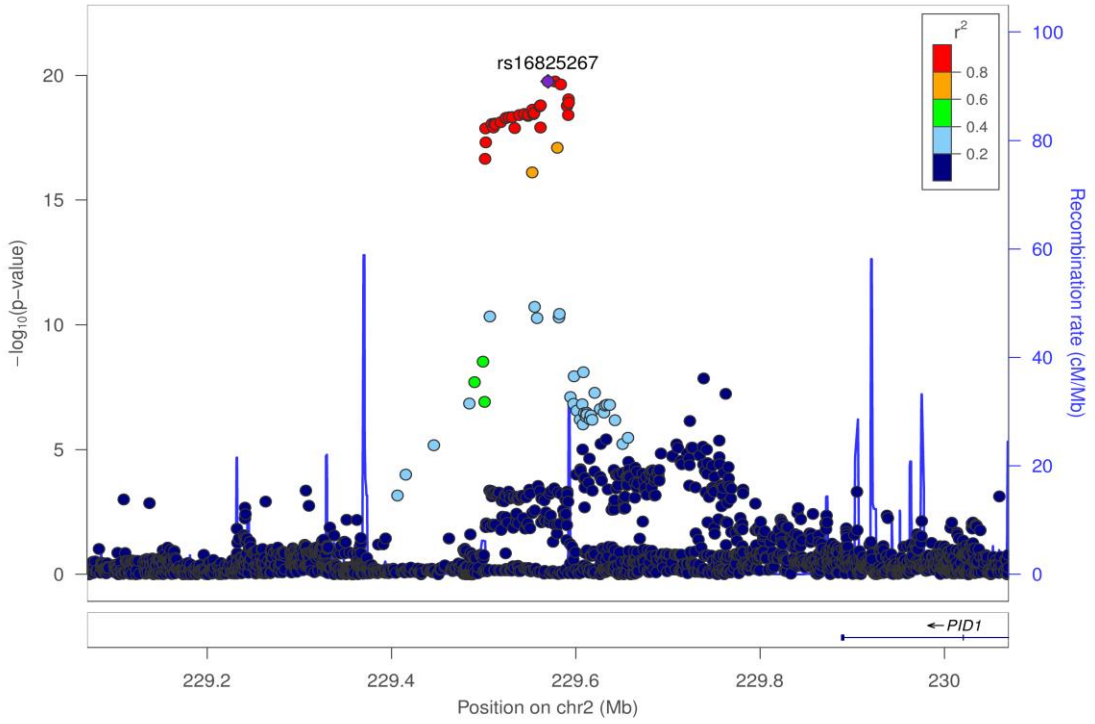
Supplementary Figure 2-12: Regional association plot for rs72902175 (*NR4A2* locus at 2q24.1)



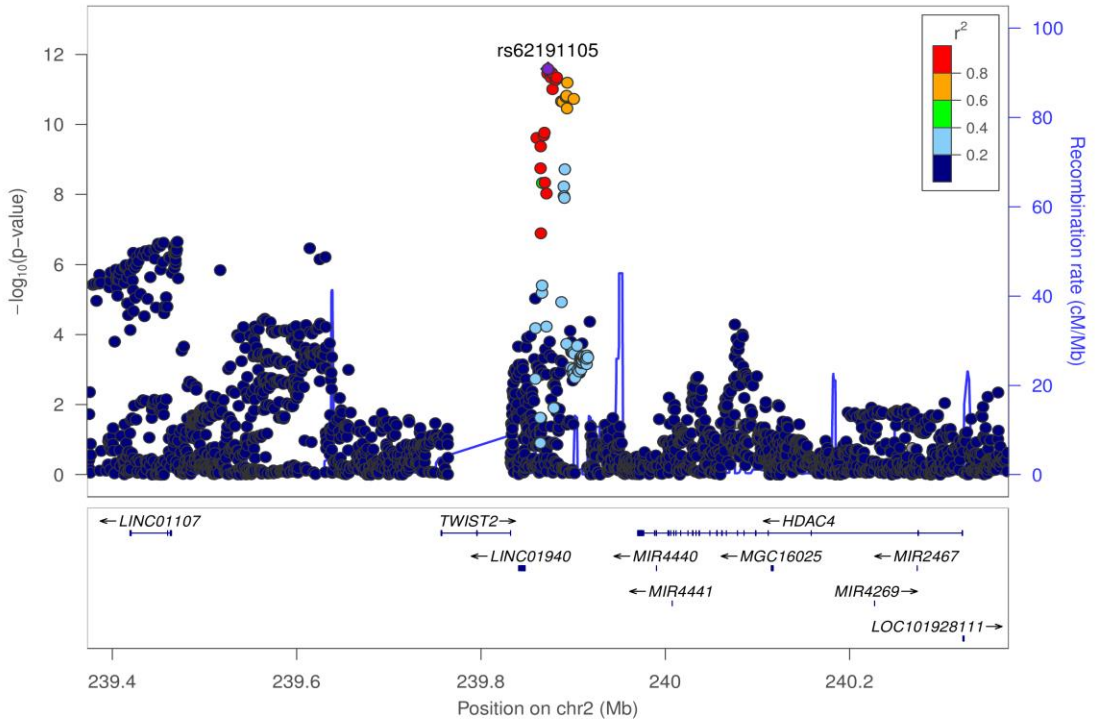
Supplementary Figure 2-13: Regional association plot for rs2571445 (*TNS1* locus at 2q35)



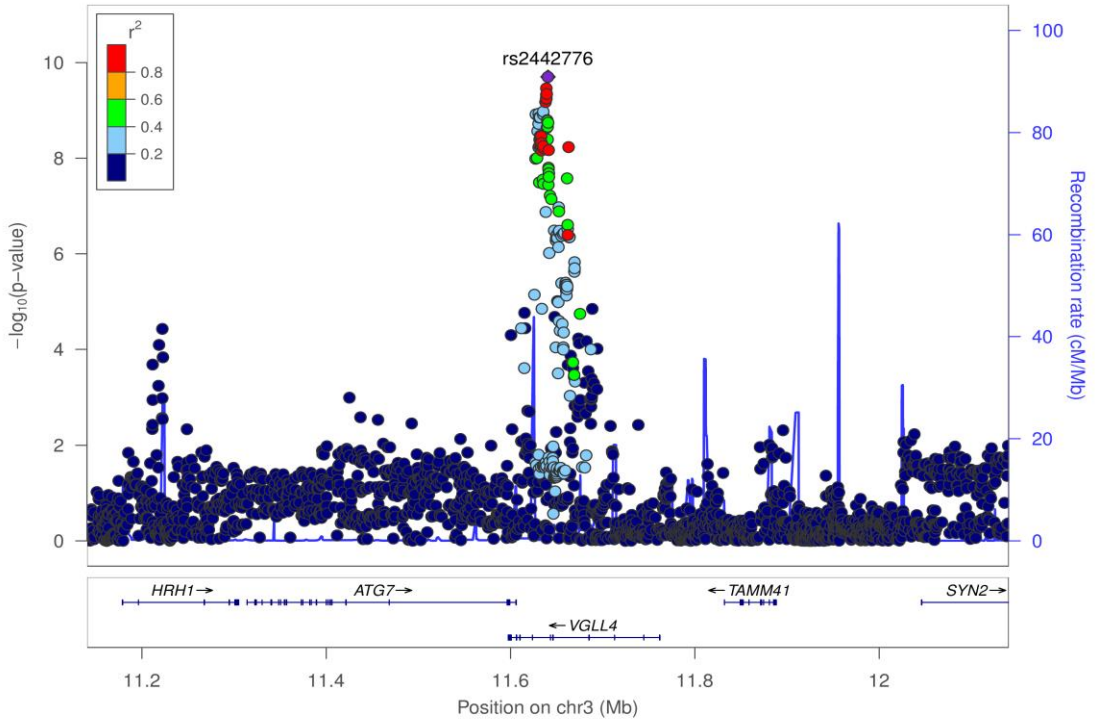
Supplementary Figure 2-14: Regional association plot for rs16825267 (*PID1* locus at 2q36.3)



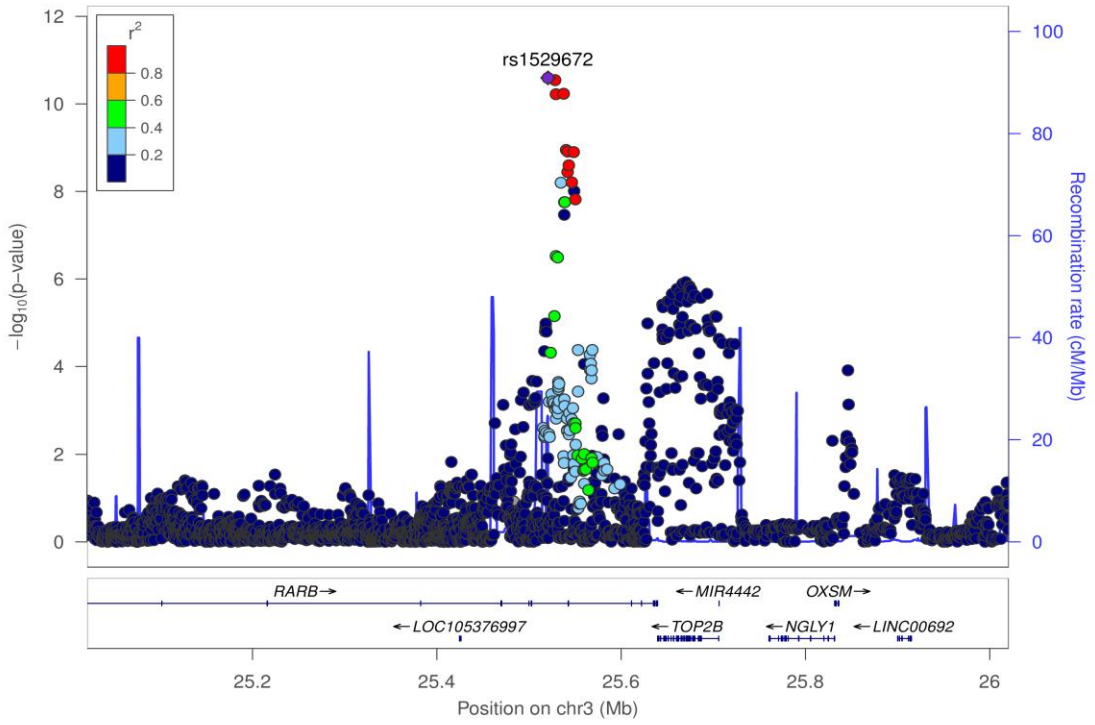
Supplementary Figure 2-15: Regional association plot for rs62191105 (*TWIST2* locus at 2q37.3)



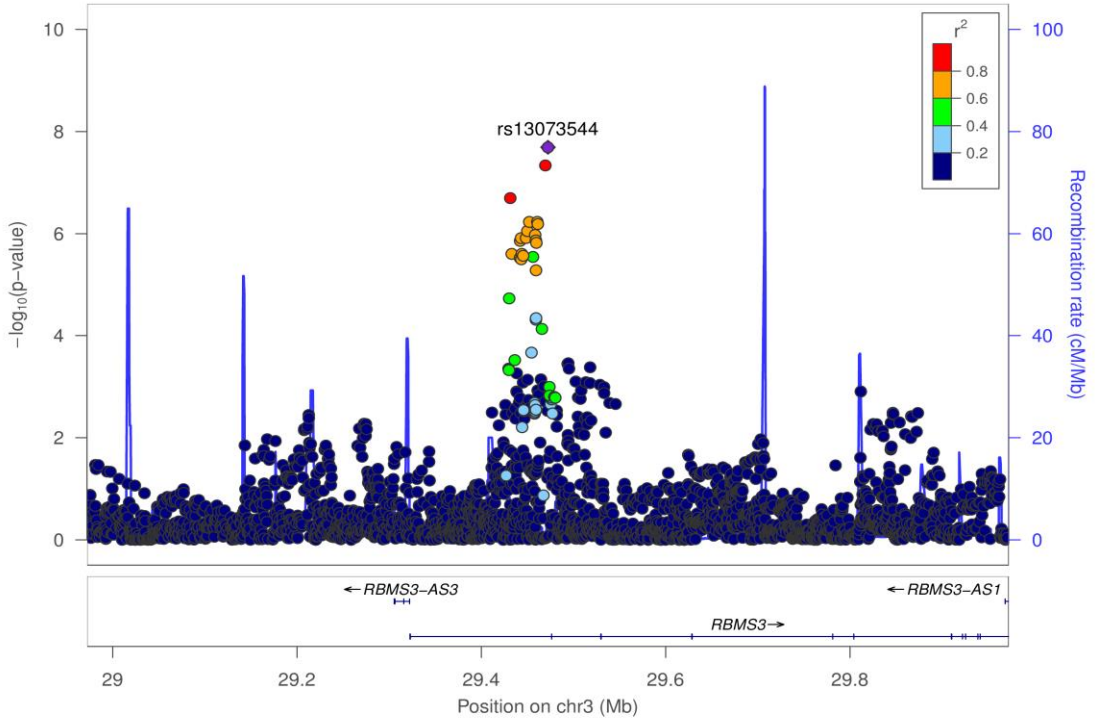
Supplementary Figure 2-16: Regional association plot for rs2442776 (*VGLL4* locus at 3p25.3)



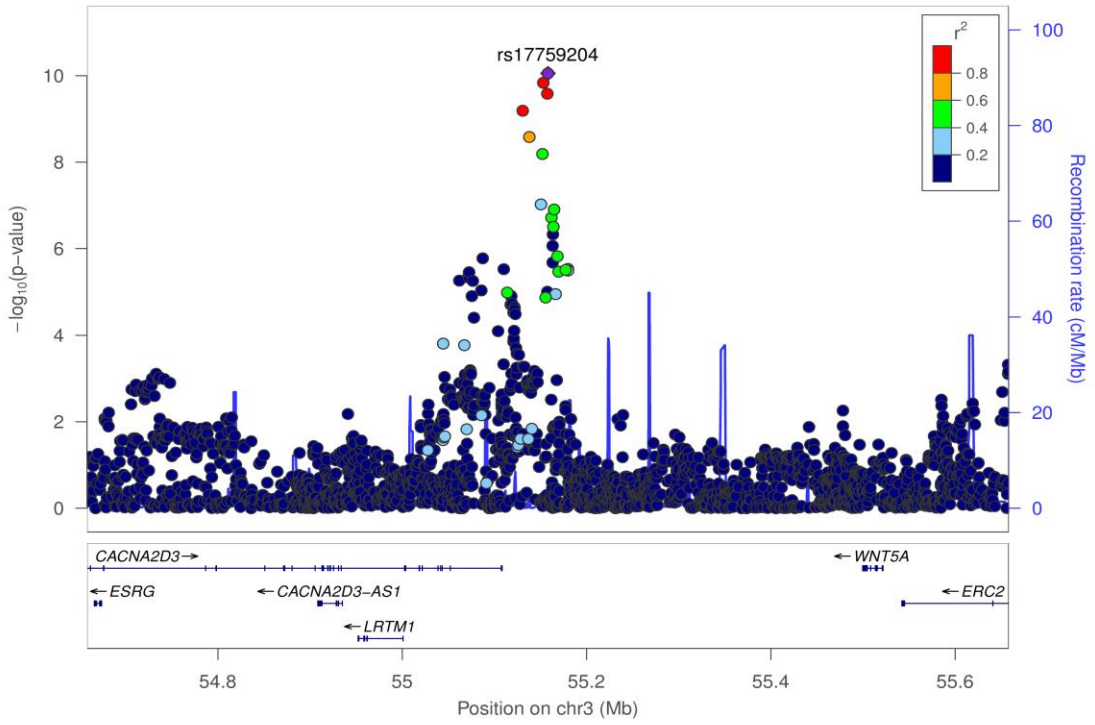
Supplementary Figure 2-17: Regional association plot for rs1529672 (*RARB* locus at 3p24.2)



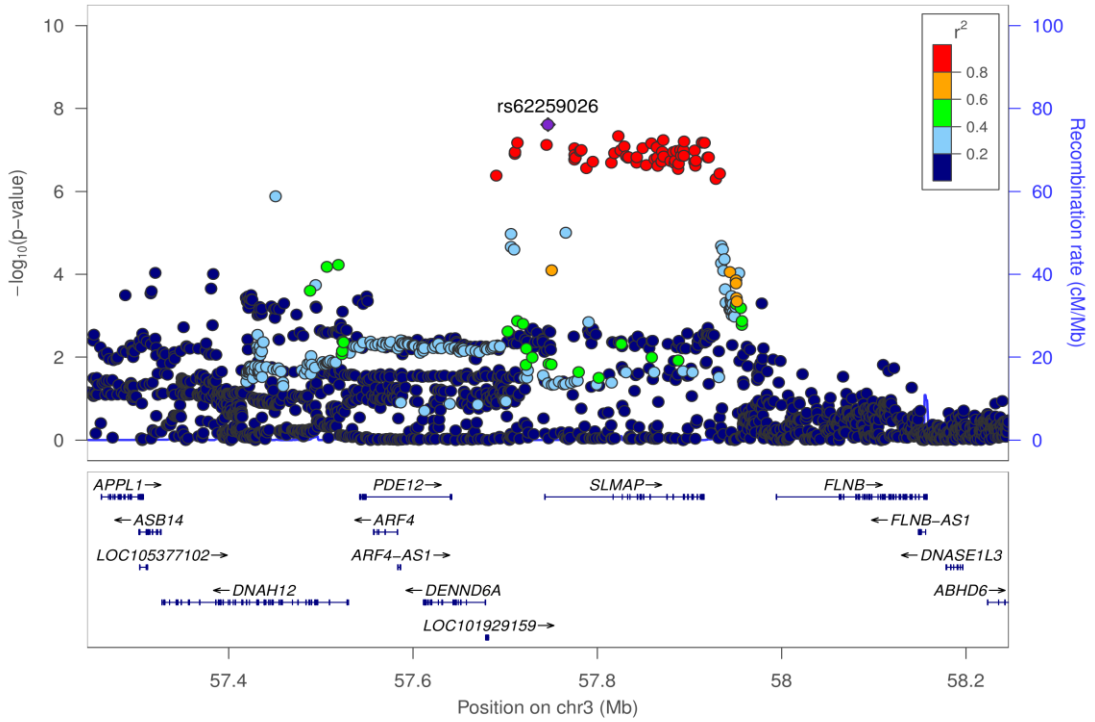
Supplementary Figure 2-18: Regional association plot for rs13073544 (*RBMS3* locus at 3p24.1)



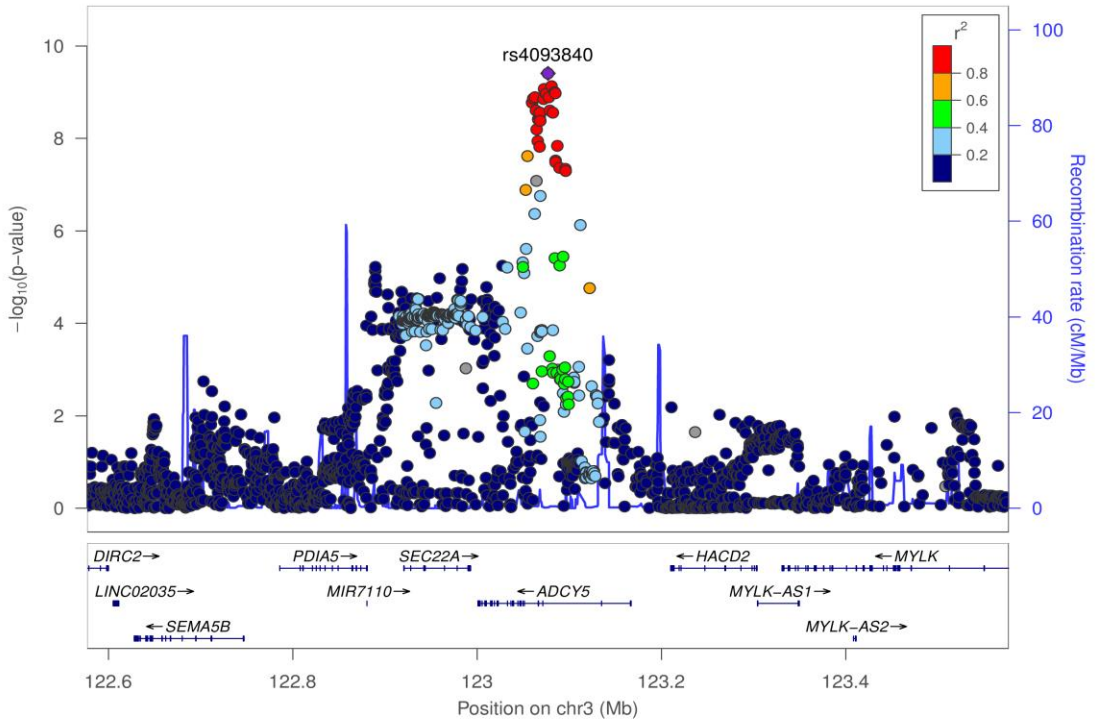
Supplementary Figure 2-19: Regional association plot for rs17759204 (*CACNA2D3* locus at 3p14.3)



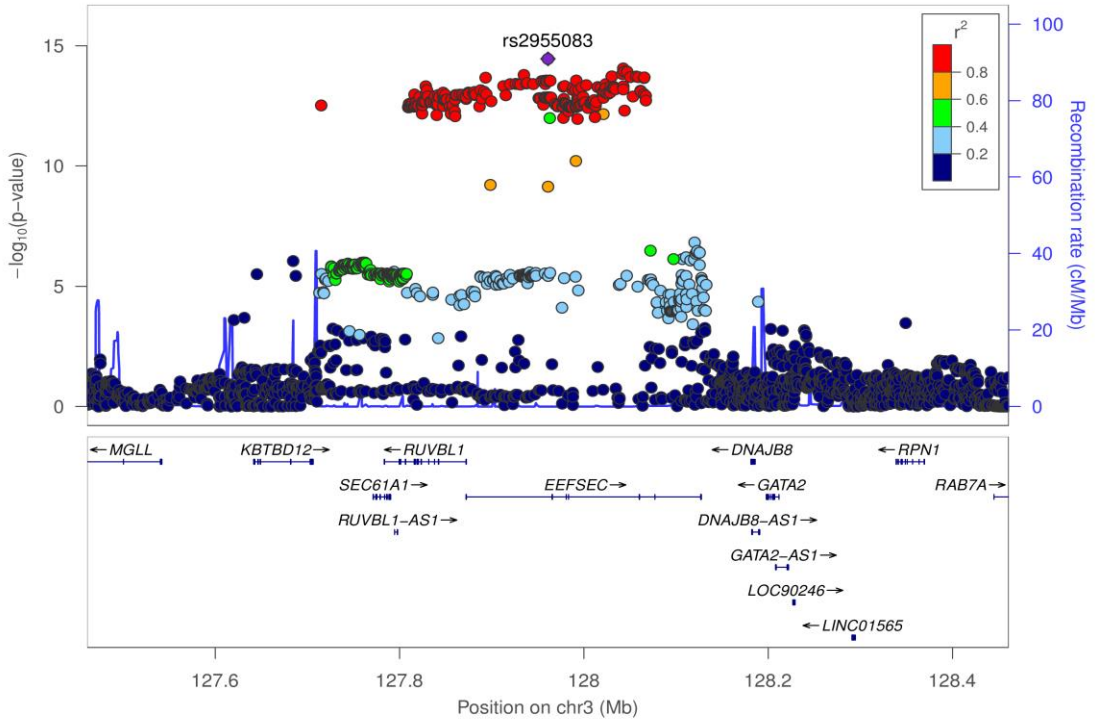
Supplementary Figure 2-20: Regional association plot for rs62259026 (*SLMAP* locus at 3p14.3)



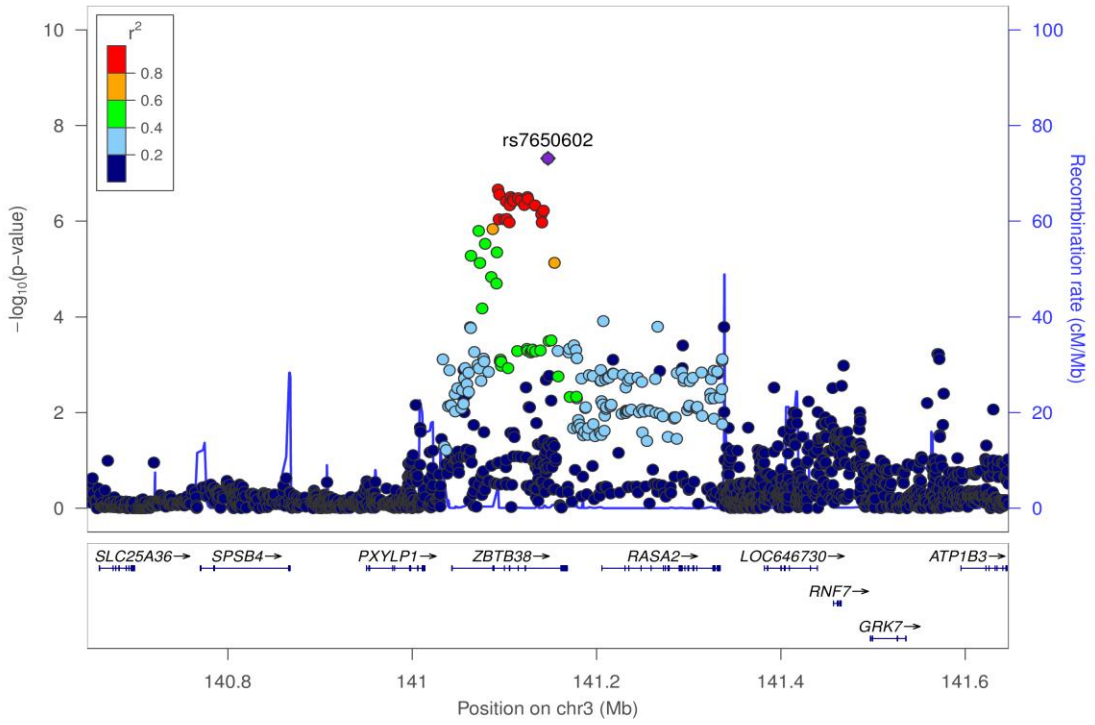
Supplementary Figure 2-21: Regional association plot for rs4093840 (*ADCY5* locus at 3q21.1)



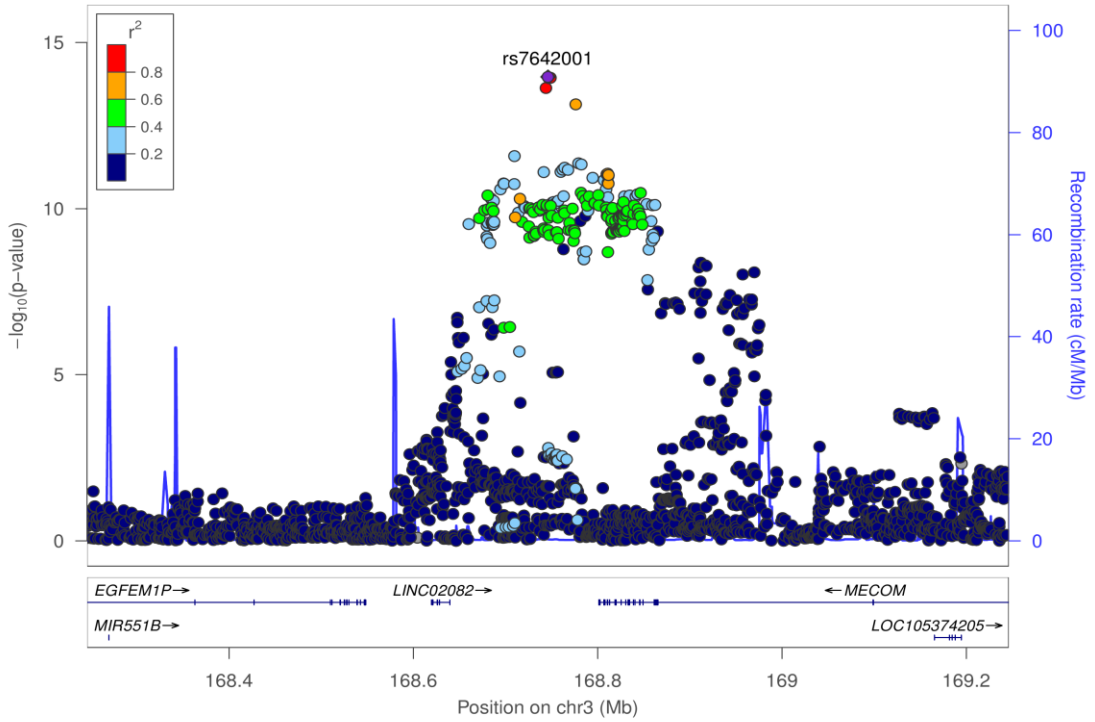
Supplementary Figure 2-22: Regional association plot for rs2955083 (*EEFSEC* locus at 3q21.3)



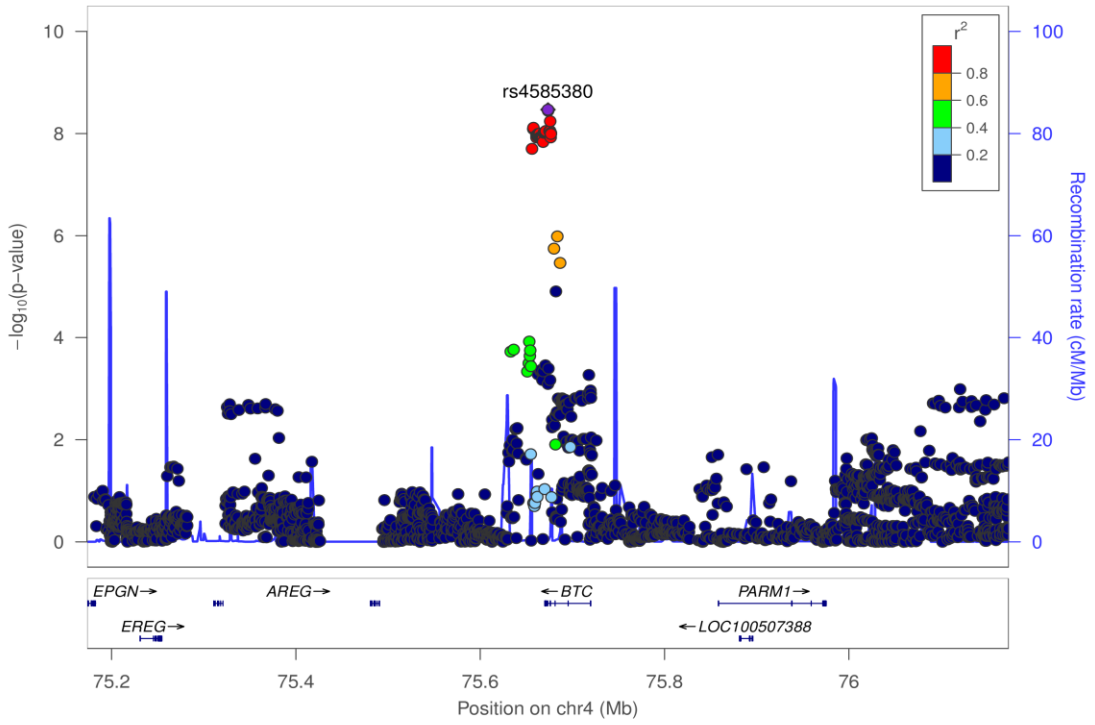
Supplementary Figure 2-23: Regional association plot for rs7650602 (*ZBTB38* locus at 3q23)



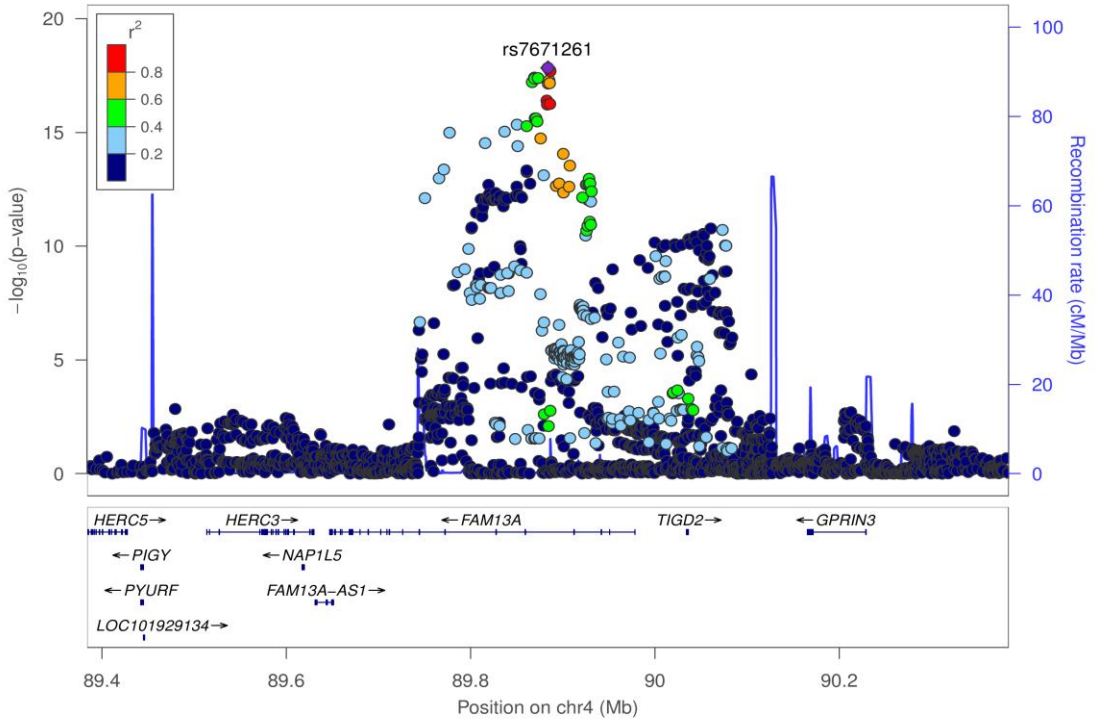
Supplementary Figure 2-24: Regional association plot for rs7642001 (*MECOM* locus at 3q26.2)



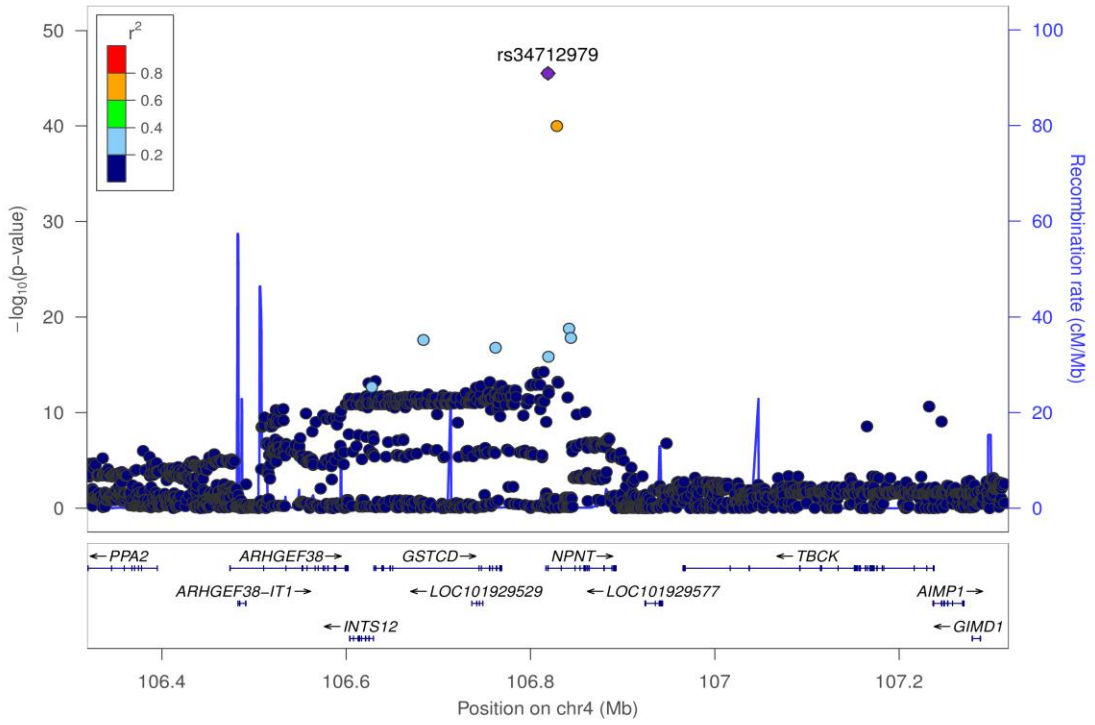
Supplementary Figure 2-25: Regional association plot for rs4585380 (*BTC* locus at 4q13.3)



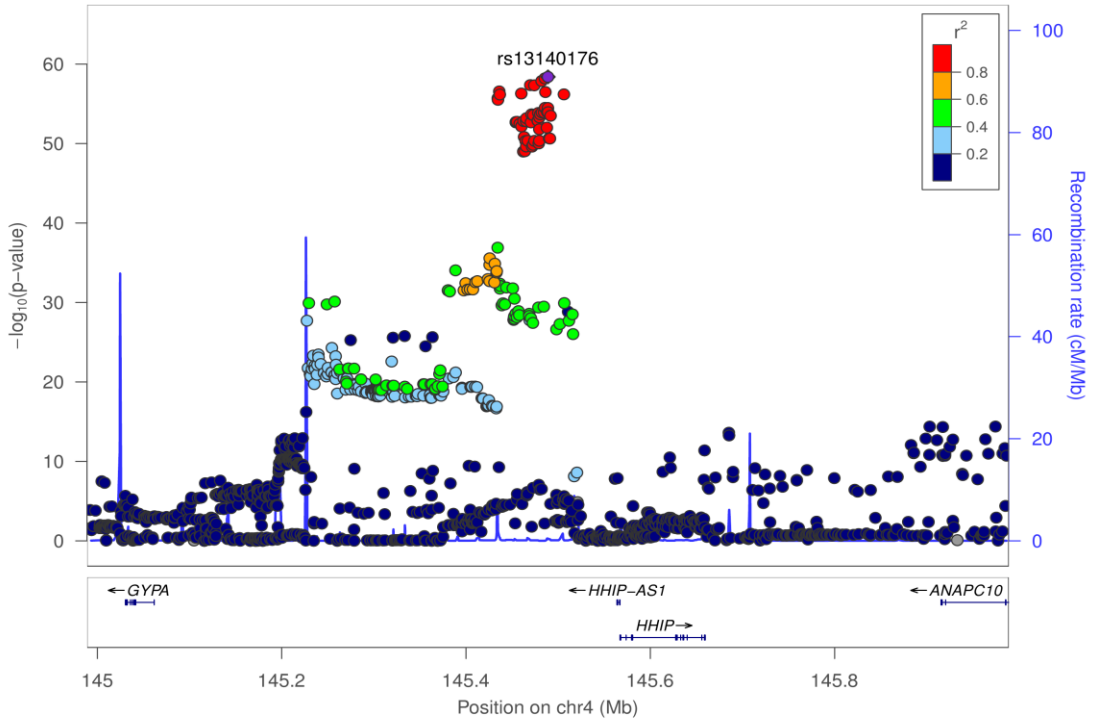
Supplementary Figure 2-26: Regional association plot for rs7671261 (*FAM13A* locus at 4q22.1)



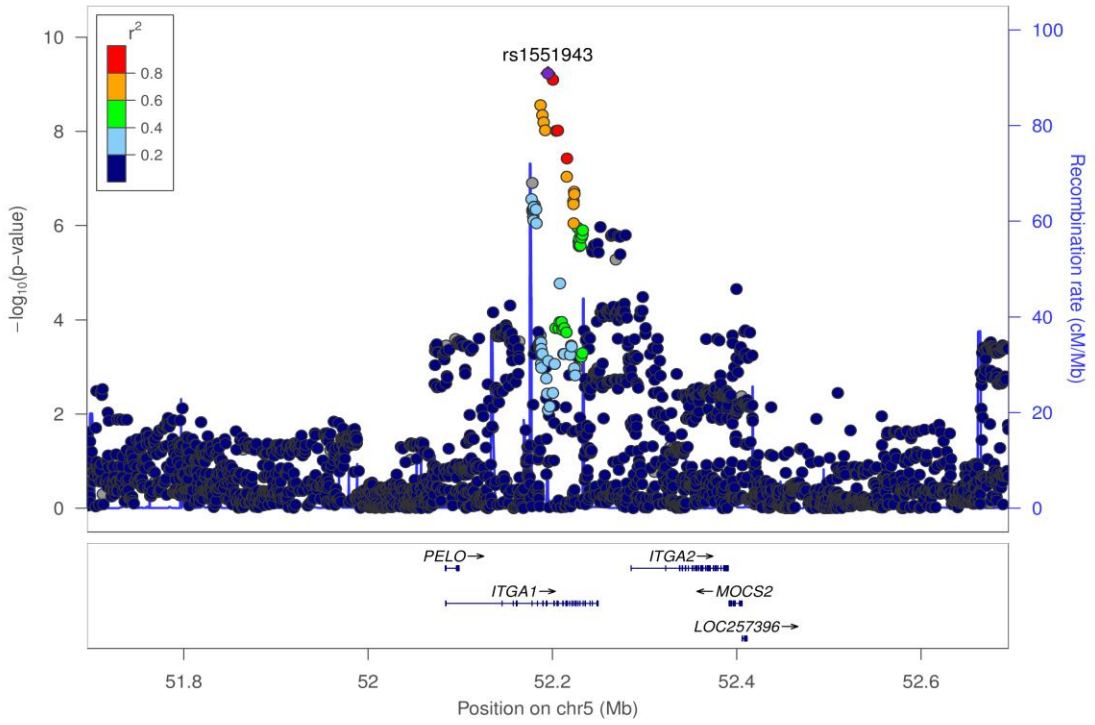
Supplementary Figure 2-27: Regional association plot for rs34712979 (*NPNT* locus at 4q24)



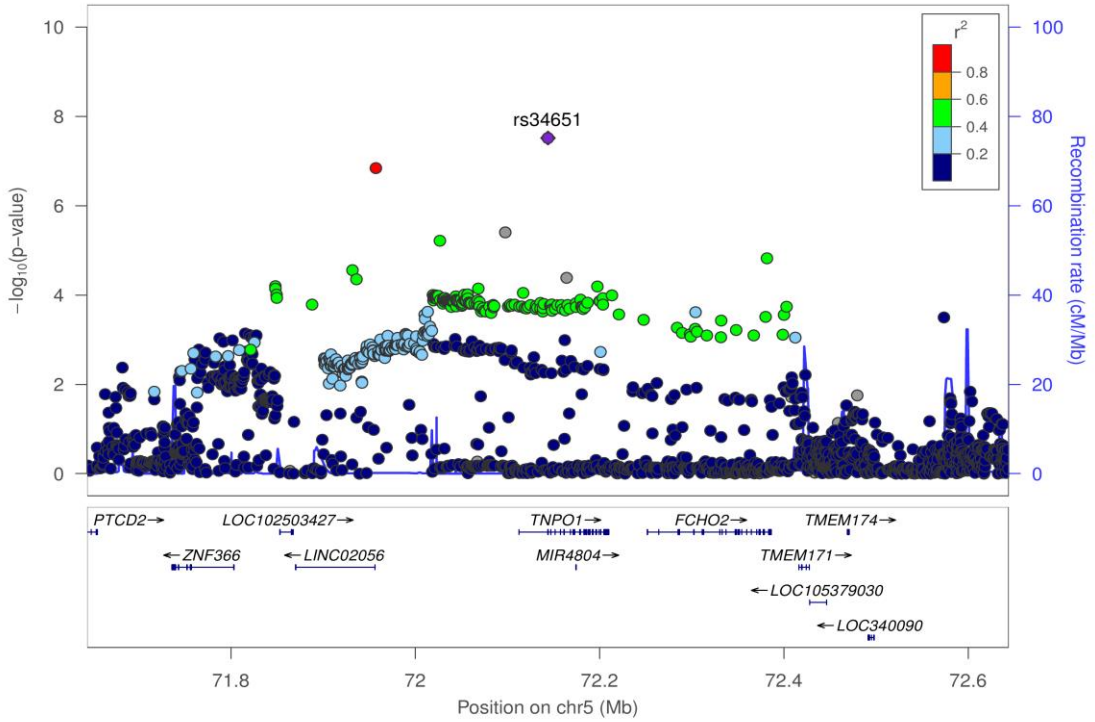
Supplementary Figure 2-28: Regional association plot for rs13140176 (*HHIP* locus at 4q31.21)



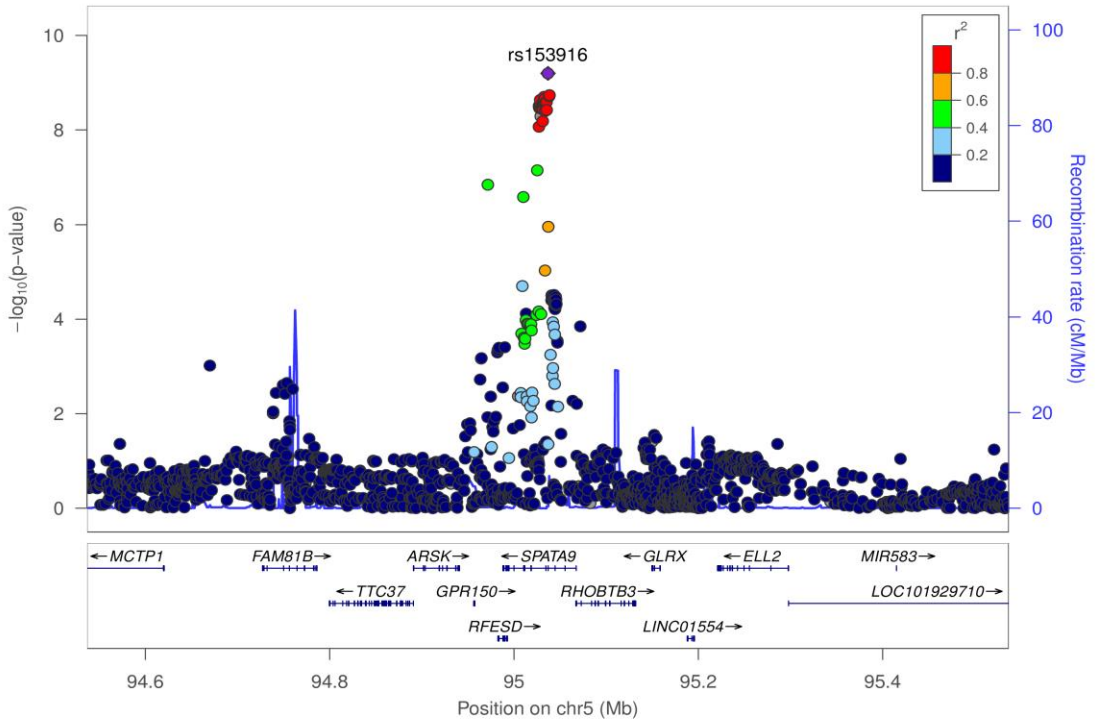
Supplementary Figure 2-29: Regional association plot for rs1551943 (*ITGA1* locus at 5q11.2)



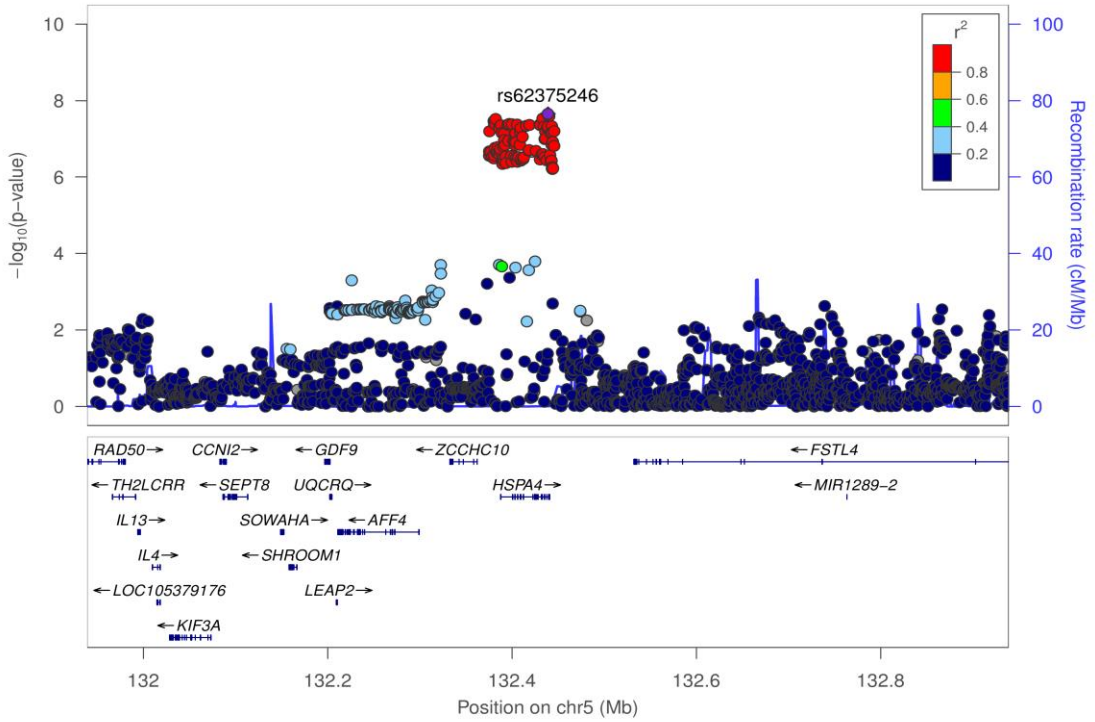
Supplementary Figure 2-30: Regional association plot for rs34651 (*TNPO1* locus at 5q13.2)



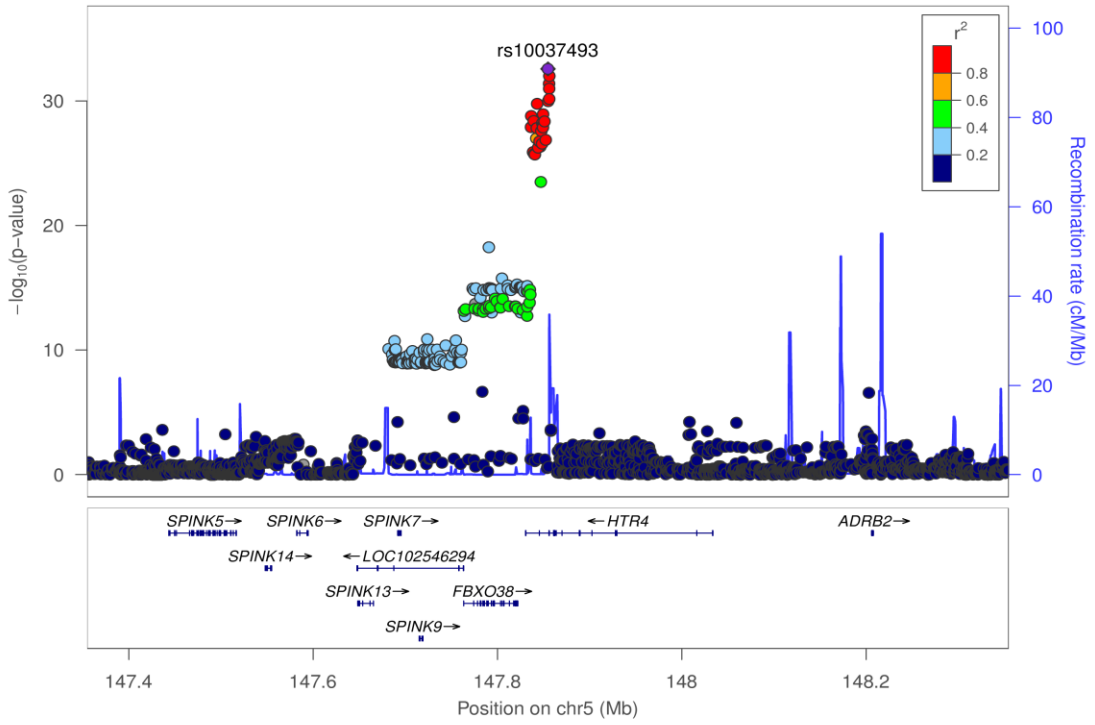
Supplementary Figure 2-31: Regional association plot for rs153916 (*SPATA9* locus at 5q15)



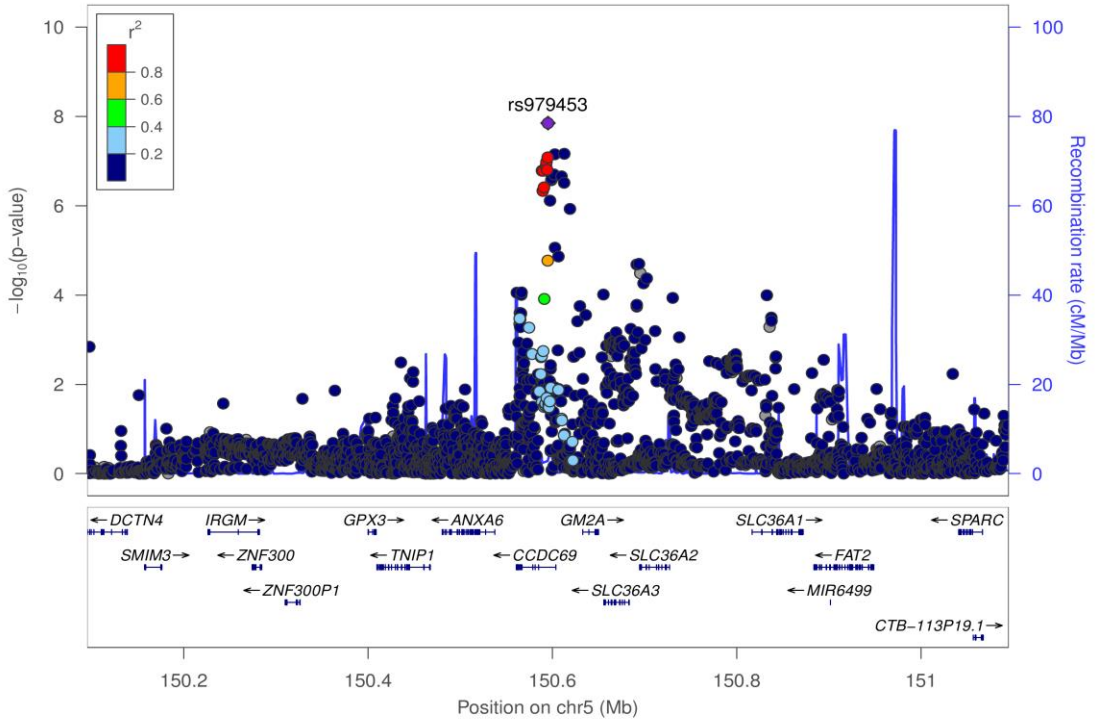
Supplementary Figure 2-32: Regional association plot for rs62375246 (*HSPA4* locus at 5q31.1)



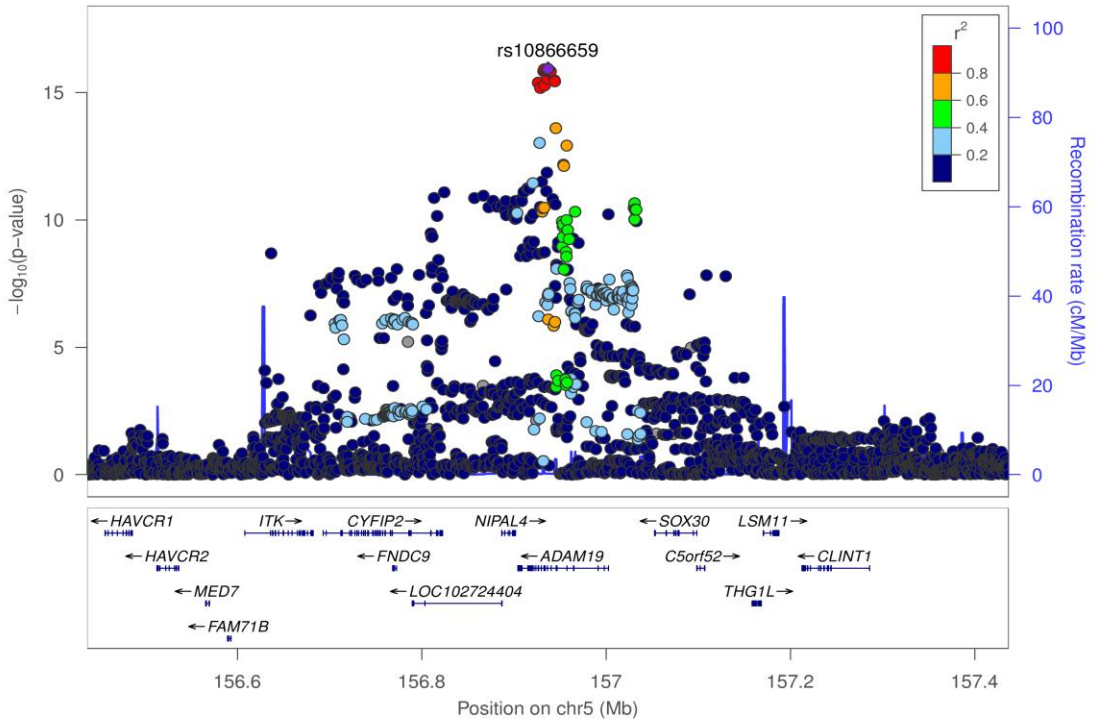
Supplementary Figure 2-33: Regional association plot for rs10037493 (*HTR4* locus at 5q32)



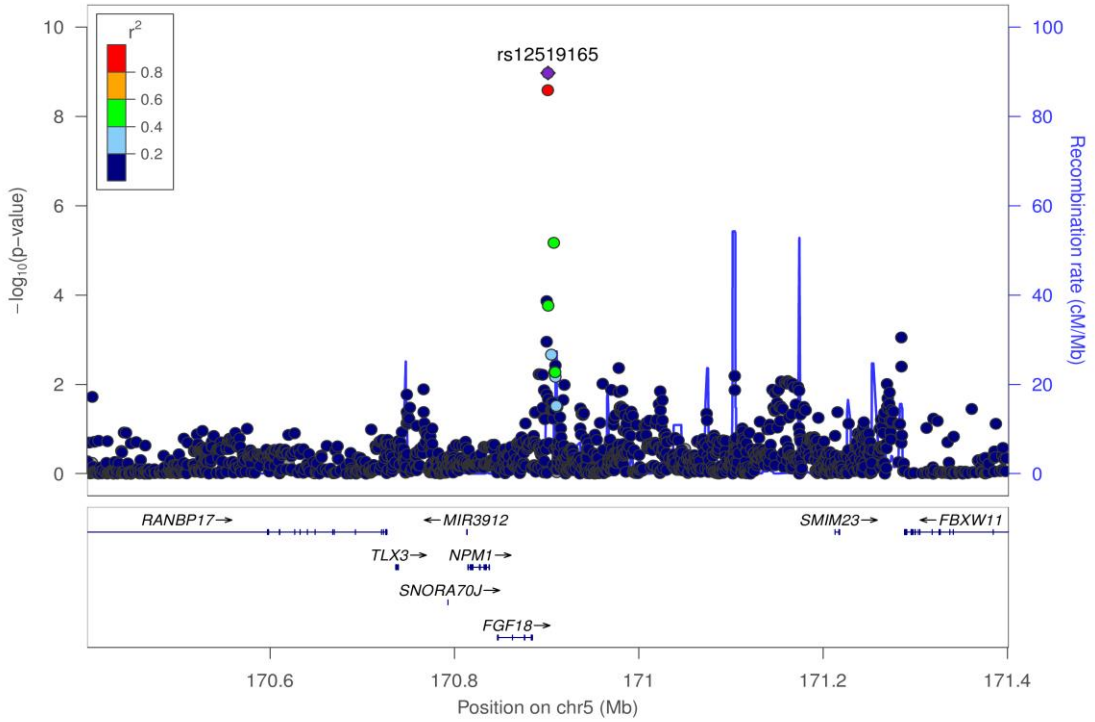
Supplementary Figure 2-34: Regional association plot for rs979453 (*CCDC69* locus at 5q33.1)



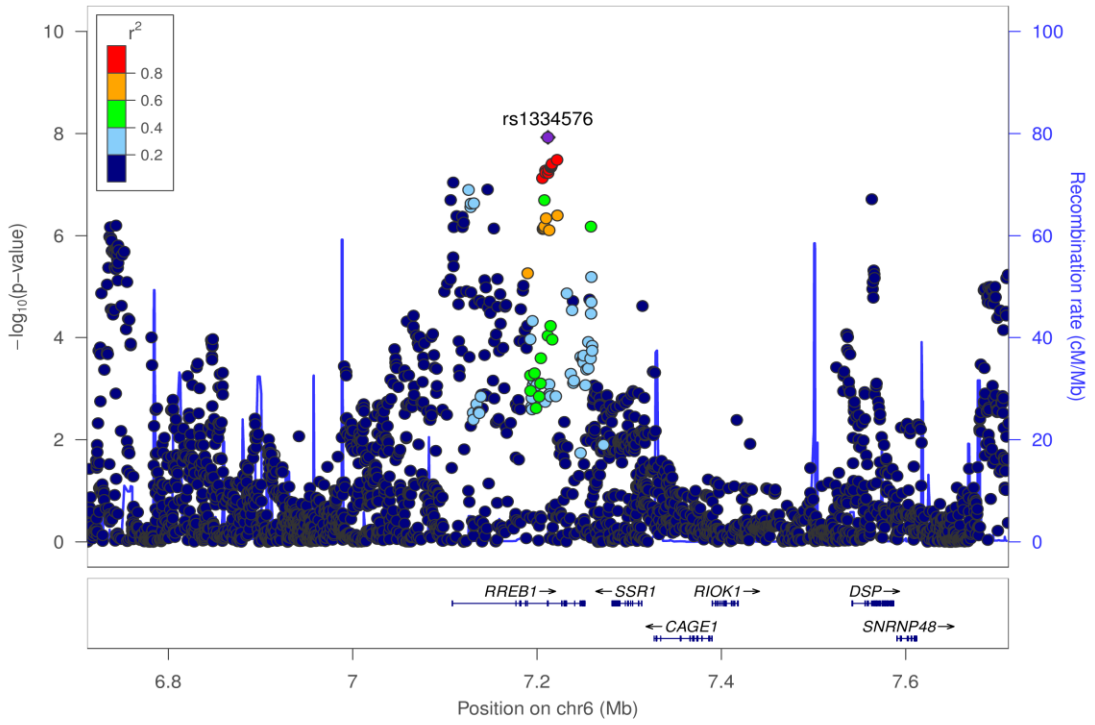
Supplementary Figure 2-35: Regional association plot for rs10866659 (*ADAM19* locus at 5q33.3)



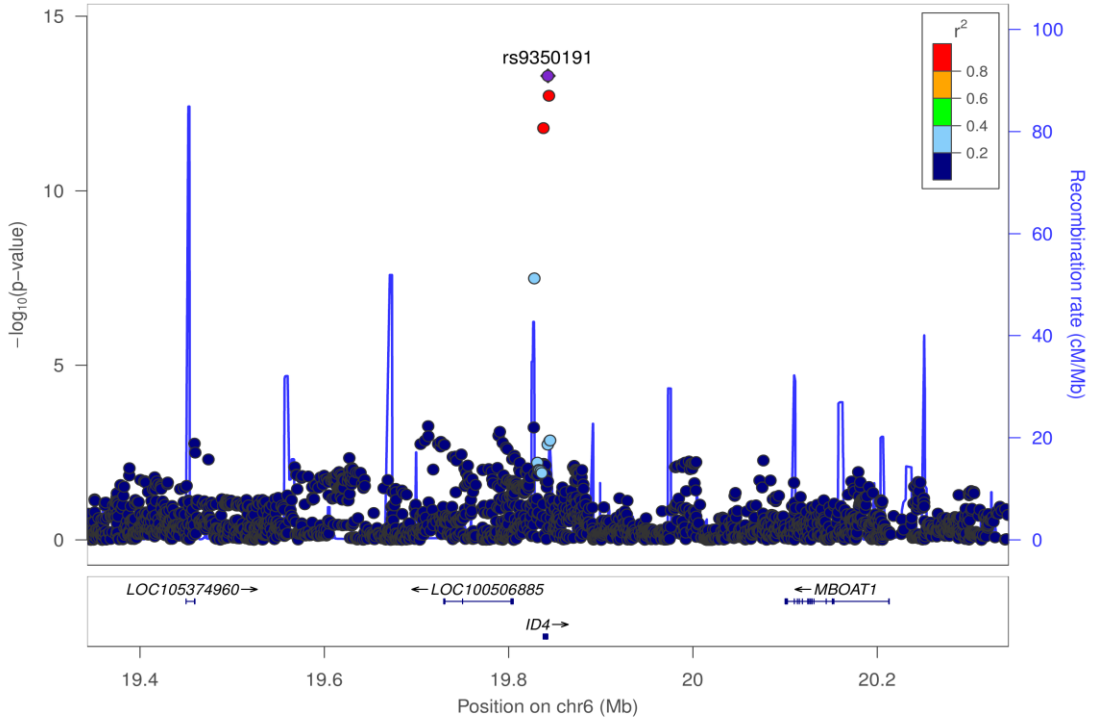
Supplementary Figure 2-36: Regional association plot for rs12519165 (*FGF18* locus at 5q35.1)



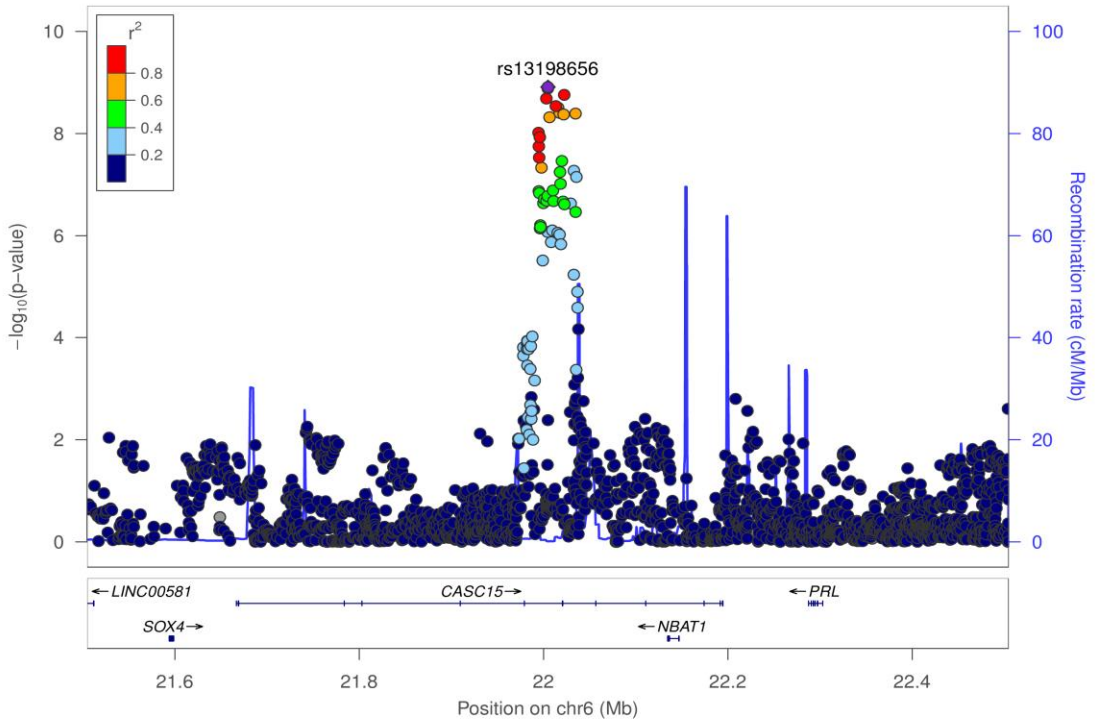
Supplementary Figure 2-37: Regional association plot for rs1334576 (*RREB1* locus at 6p24.3)



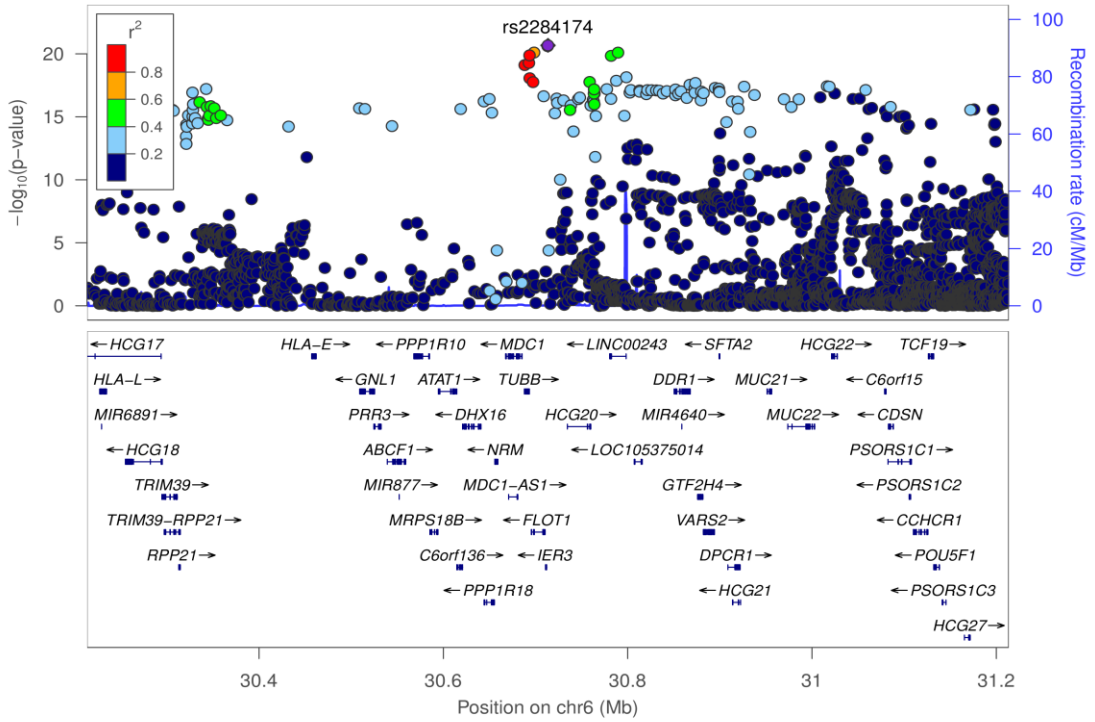
Supplementary Figure 2-38: Regional association plot for rs9350191 (*ID4* locus at 6p22.3)



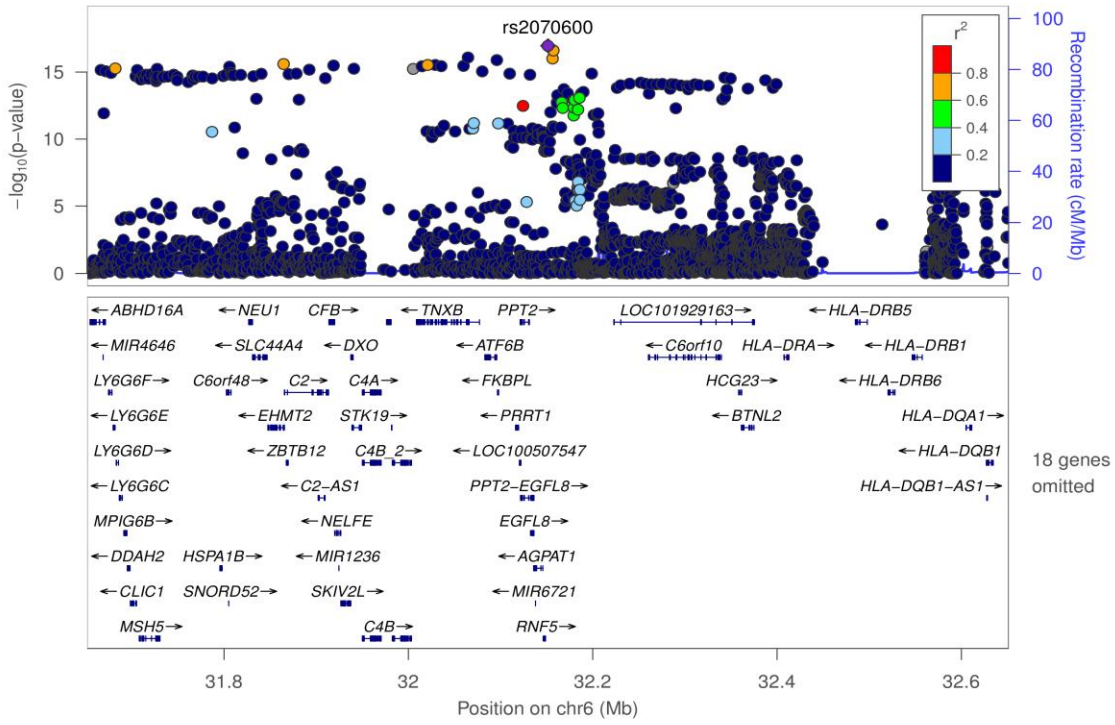
Supplementary Figure 2-39: Regional association plot for rs13198656 (*PRL* locus at 6p22.3)



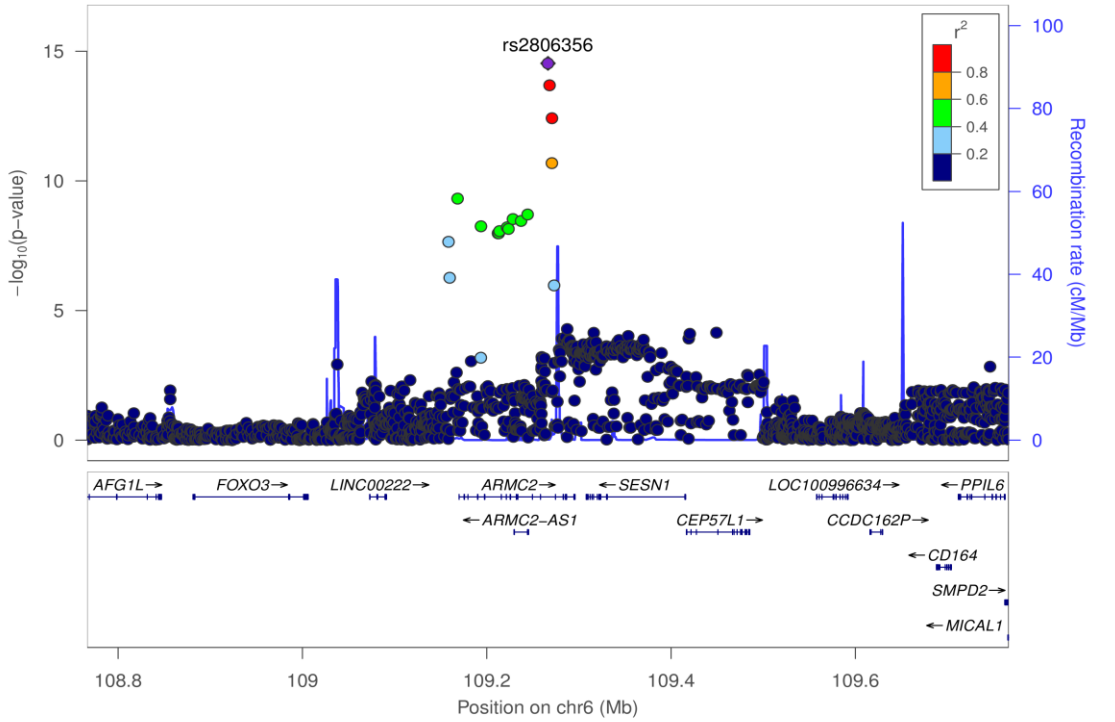
Supplementary Figure 2-40: Regional association plot for rs2284174 (*IER3* locus at 6p21.33)



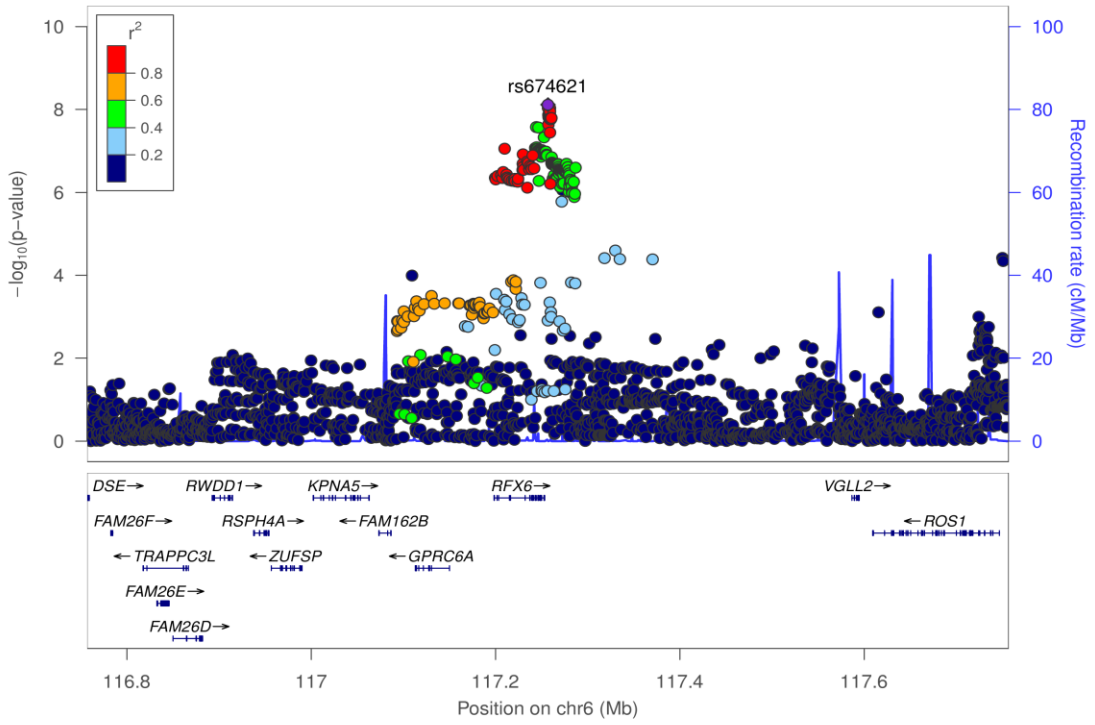
Supplementary Figure 2-41: Regional association plot for rs2070600 (*AGER* locus at 6p21.32)



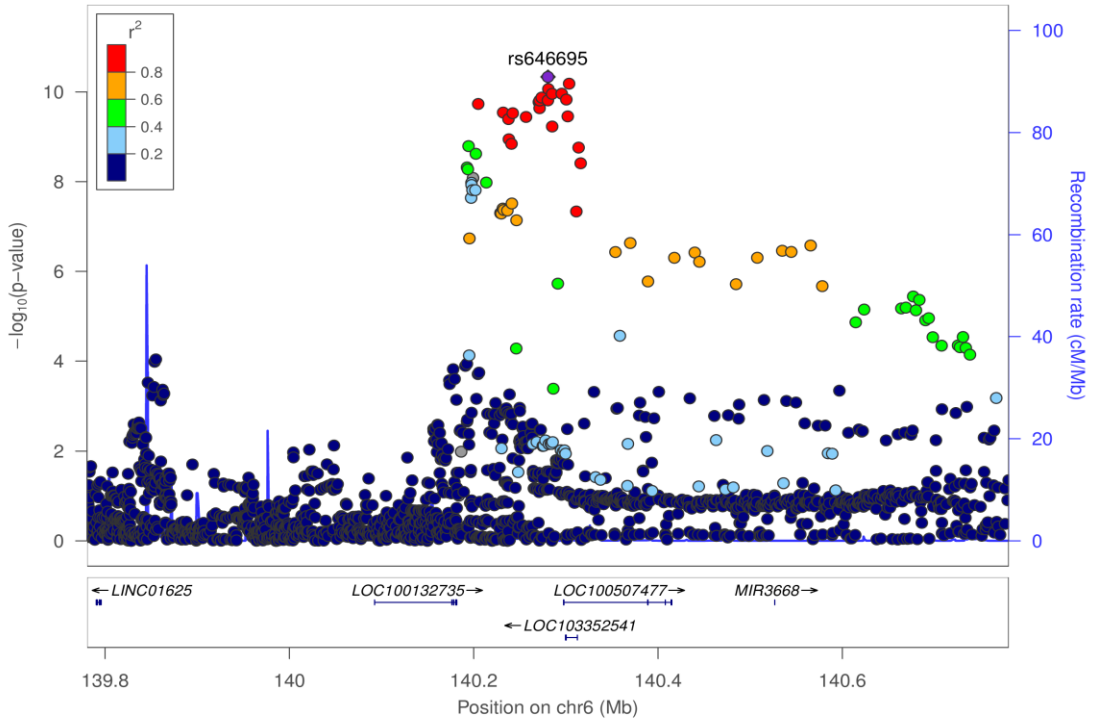
Supplementary Figure 2-42: Regional association plot for rs2806356 (*ARMC2* locus at 6q21)



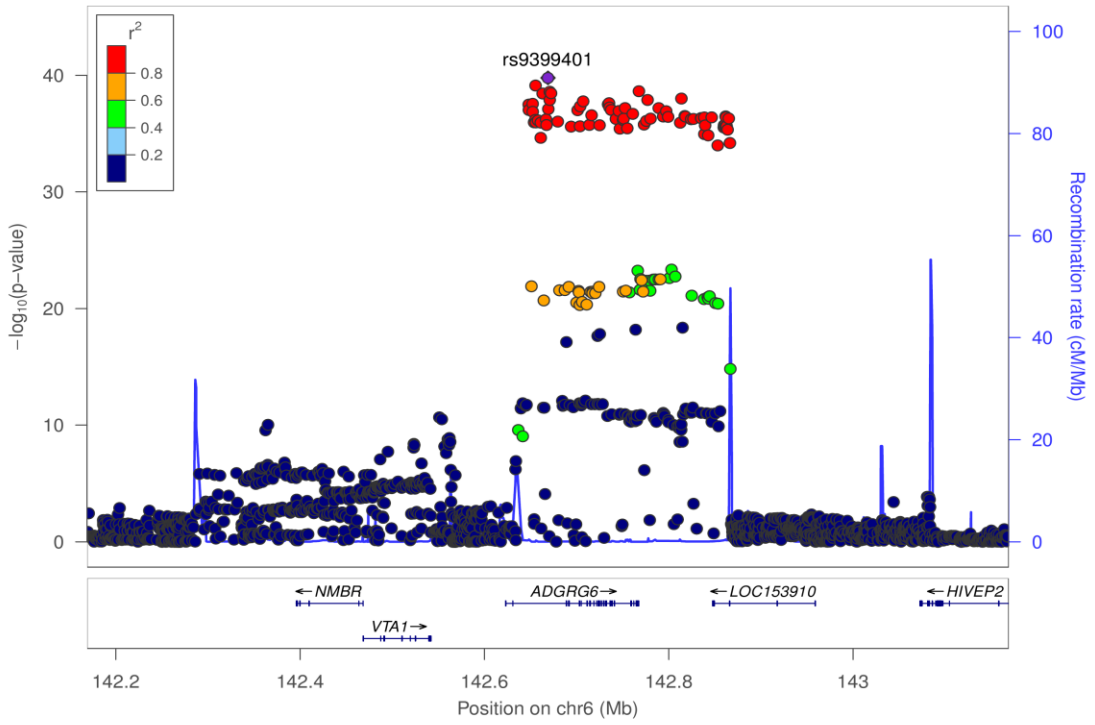
Supplementary Figure 2-43: Regional association plot for rs674621 (*RFX6* locus at 6q22.1)



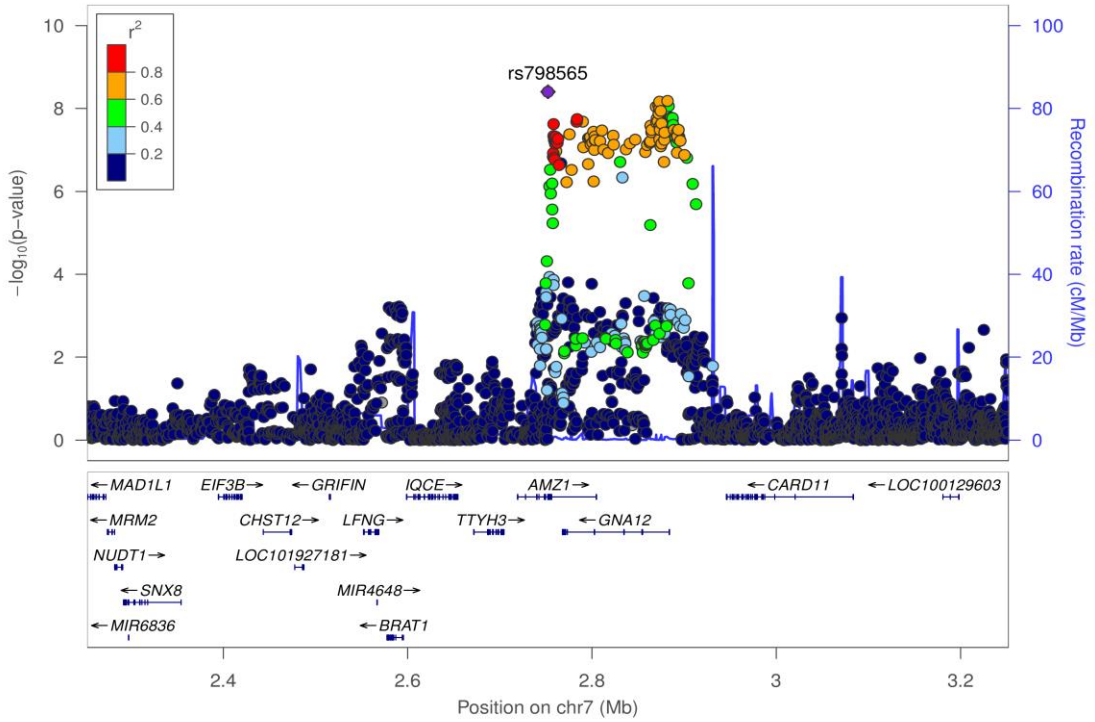
Supplementary Figure 2-44: Regional association plot for rs646695 (*CITED2* locus at 6q24.1)



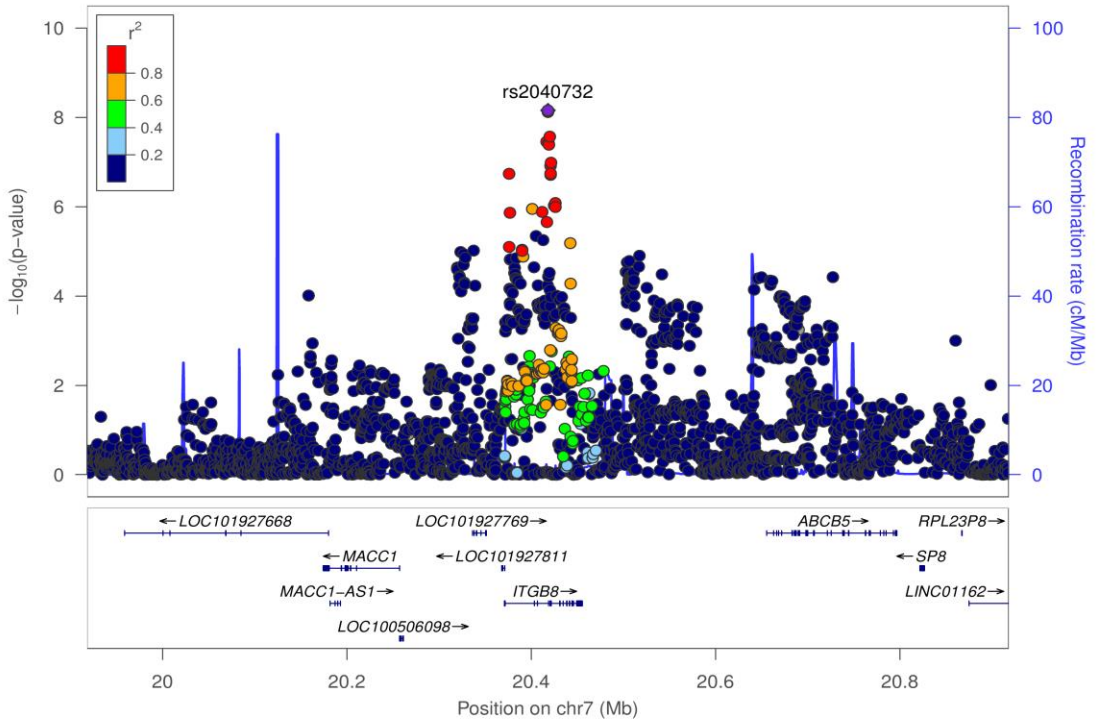
Supplementary Figure 2-45: Regional association plot for rs9399401 (*ADGRG6* locus at 6q24.1)



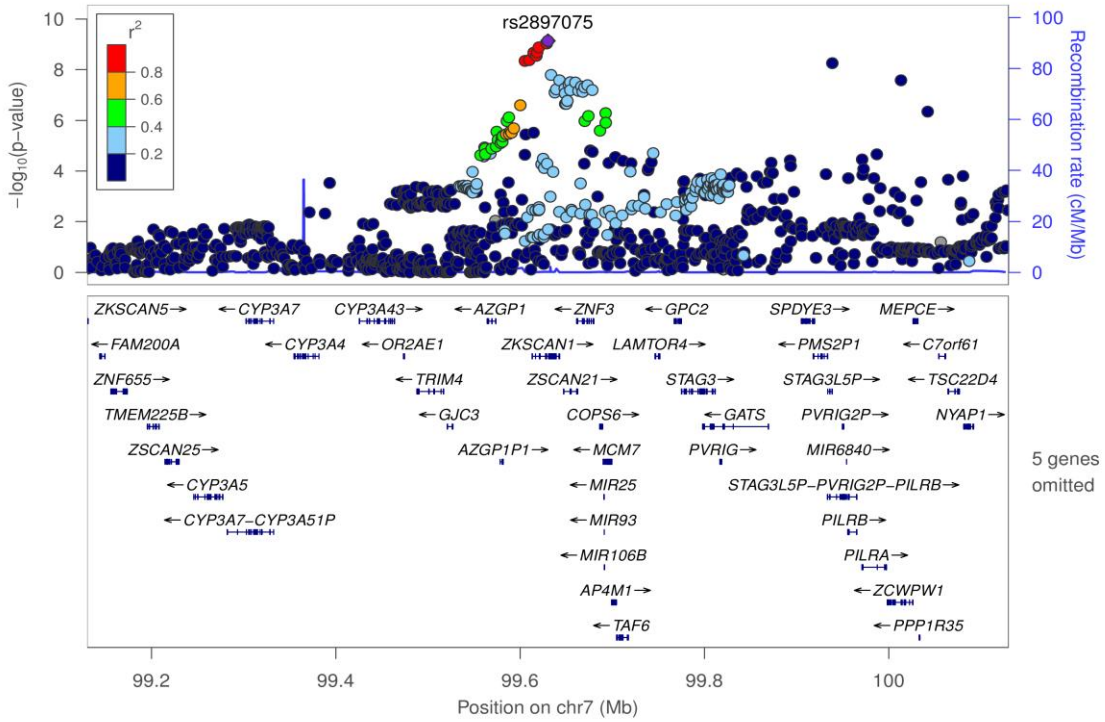
Supplementary Figure 2-46: Regional association plot for rs798565 (*AMZ1* locus at 7p22.3)



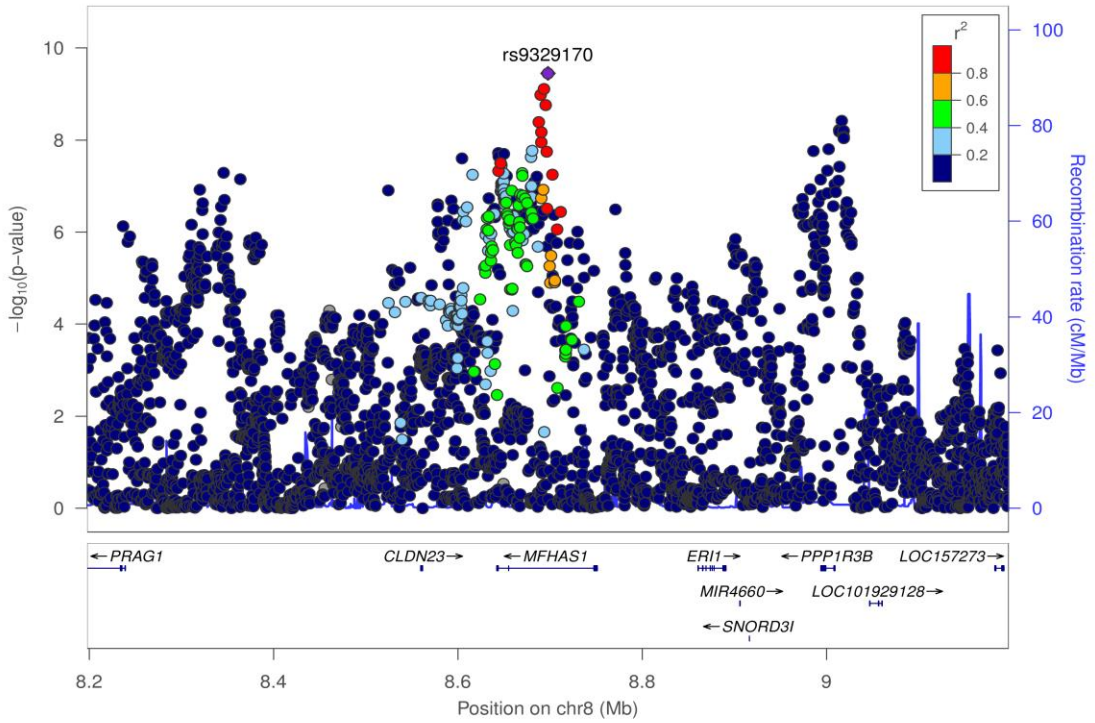
Supplementary Figure 2-47: Regional association plot for rs2040732 (*ITGB8* locus at 7p21.1)



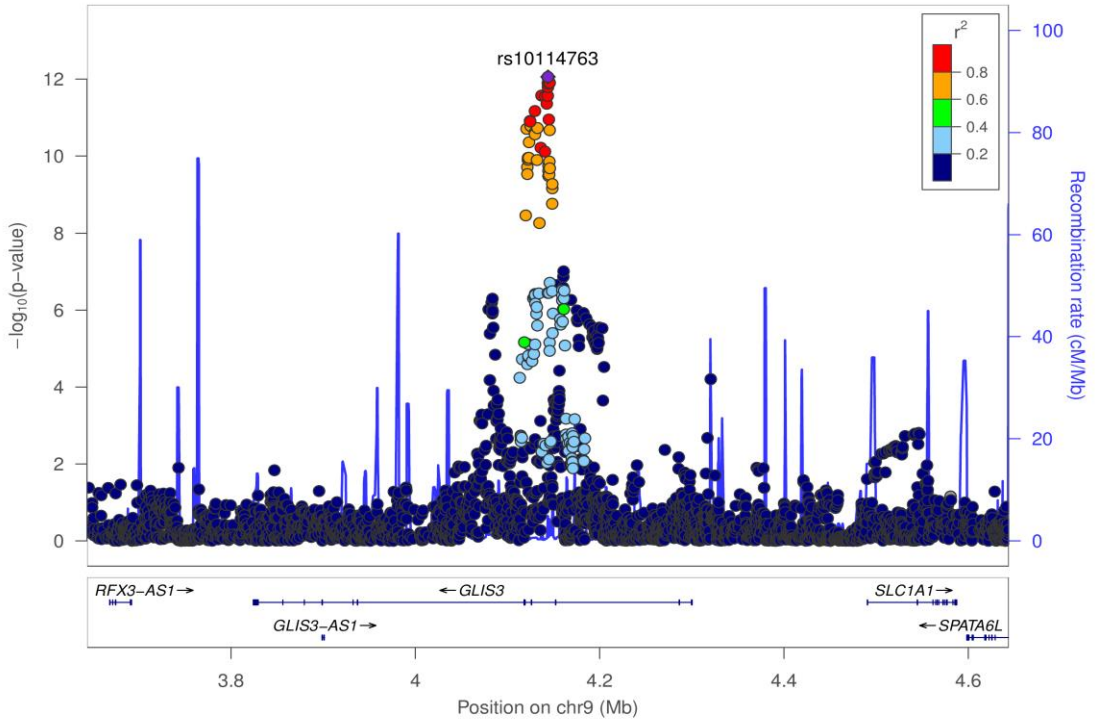
Supplementary Figure 2-48: Regional association plot for rs2897075 (*ZKSCAN1* locus at 7q22.1)



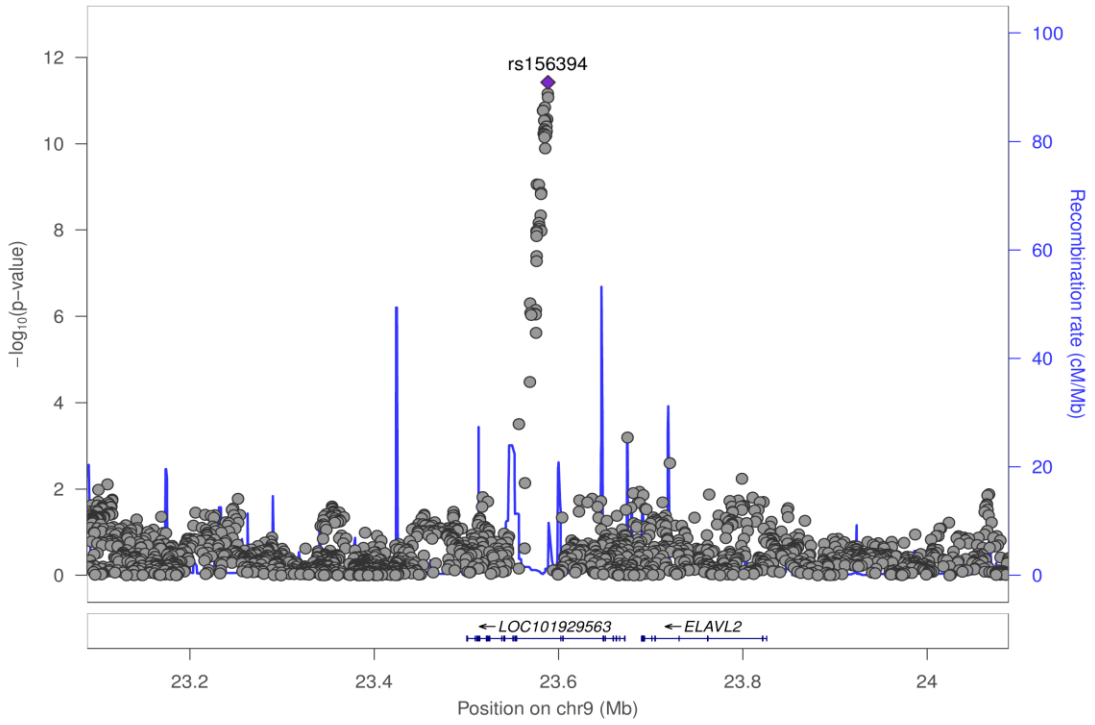
Supplementary Figure 2-49: Regional association plot for rs9329170 (*MFHAS1* locus at 8p23.1)



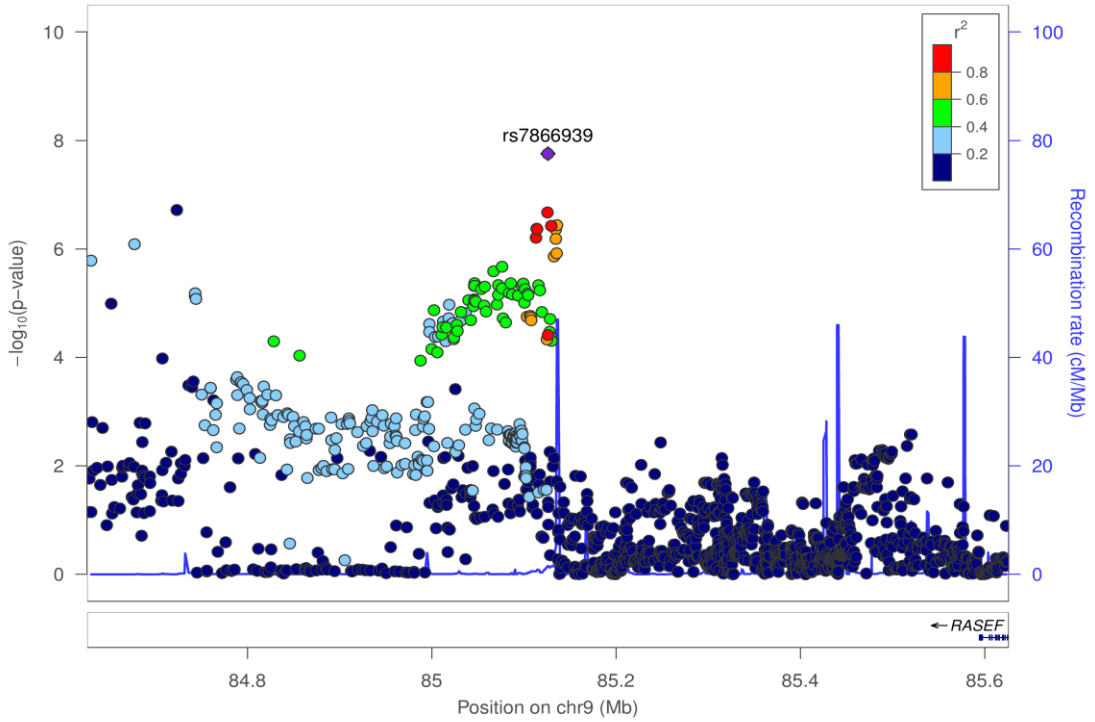
Supplementary Figure 2-50: Regional association plot for rs10114763 (*GLIS3* locus at 9p24.2)



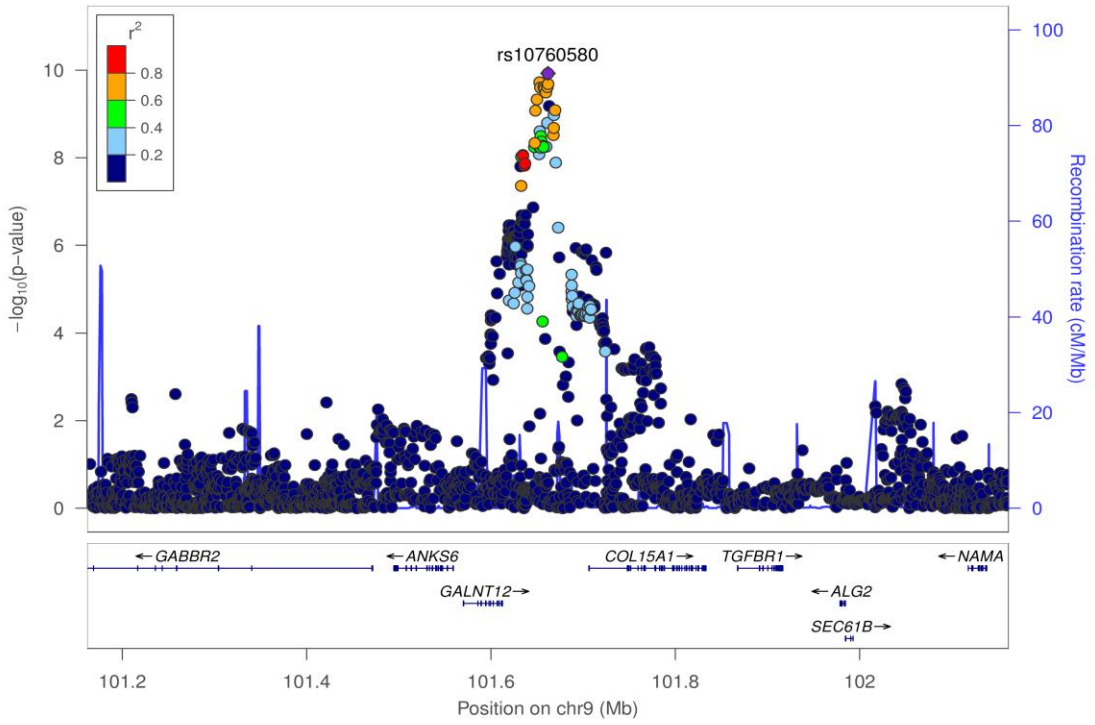
Supplementary Figure 2-51: Regional association plot for rs156394 (*ELAVL2* locus at 9p21.3)



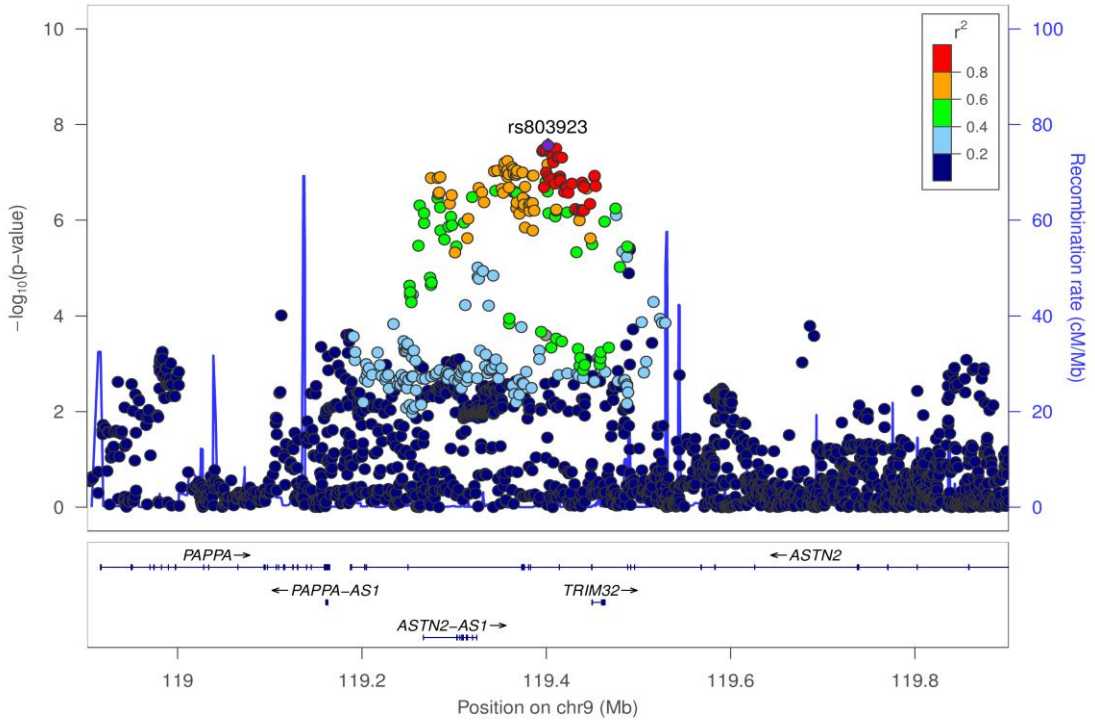
Supplementary Figure 2-52: Regional association plot for rs7866939 (*RASEF* locus at 9q21.32)



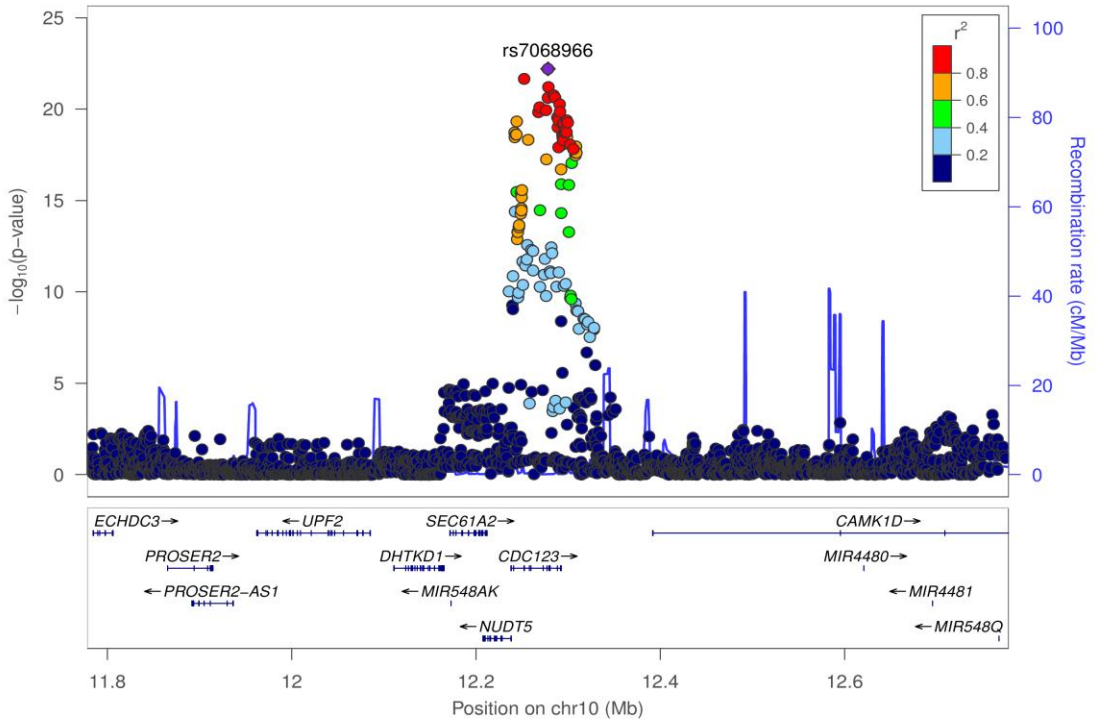
Supplementary Figure 2-53: Regional association plot for rs10760580 (*COL15A1* locus at 9q22.33)



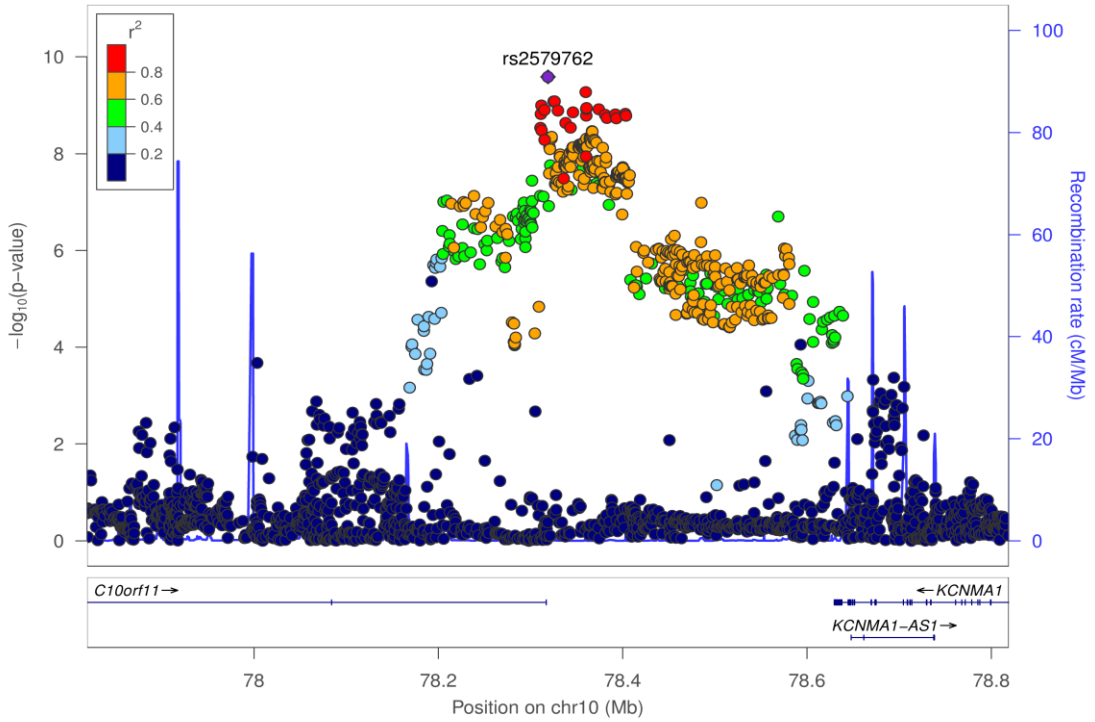
Supplementary Figure 2-54: Regional association plot for rs803923 (*ASTN2* locus at 9q33.1)



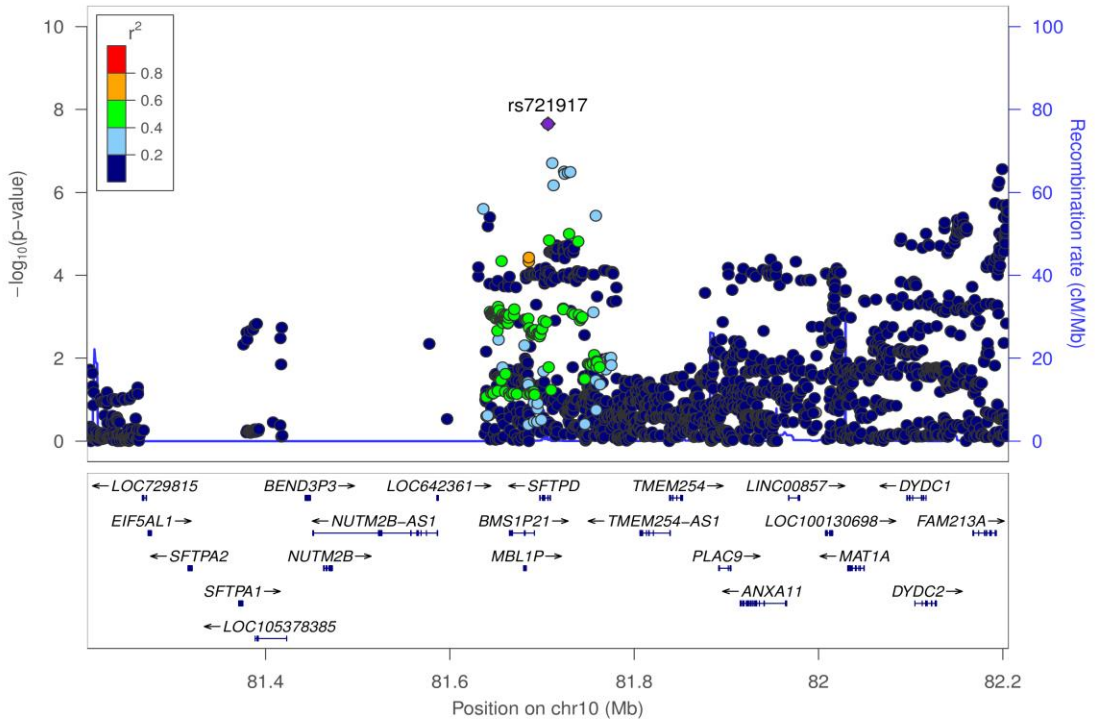
Supplementary Figure 2-55: Regional association plot for rs7068966 (*CDC123* locus at 10p13)



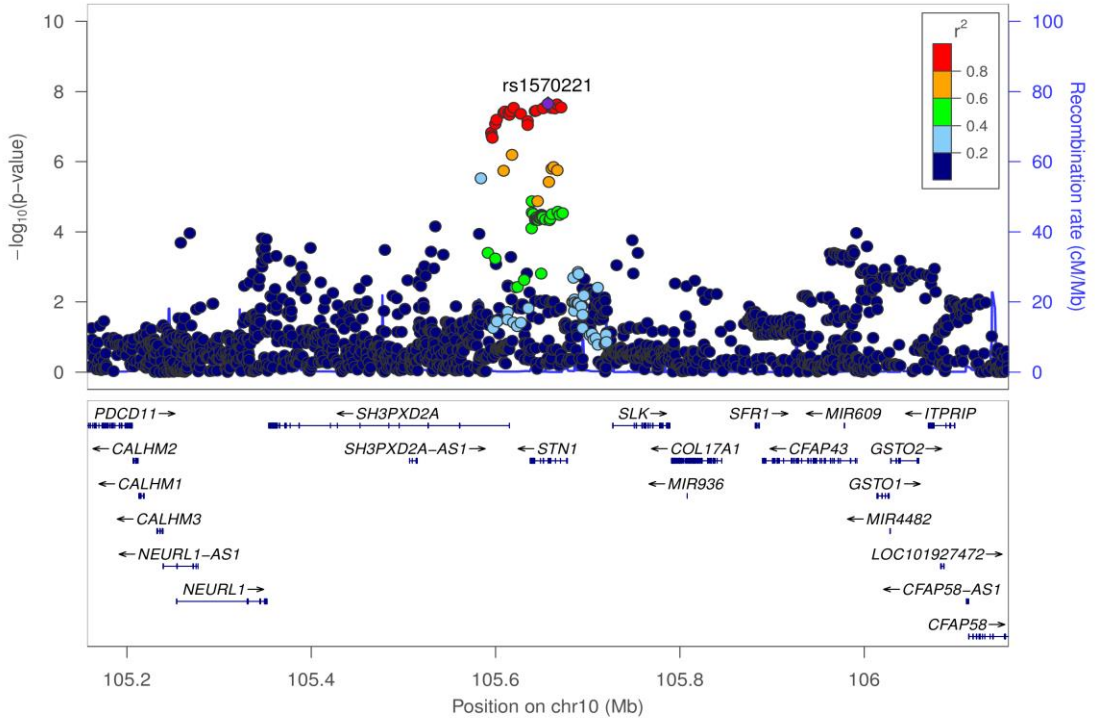
Supplementary Figure 2-56: Regional association plot for rs2579762 (*LRMDA* locus at 10q22.3)



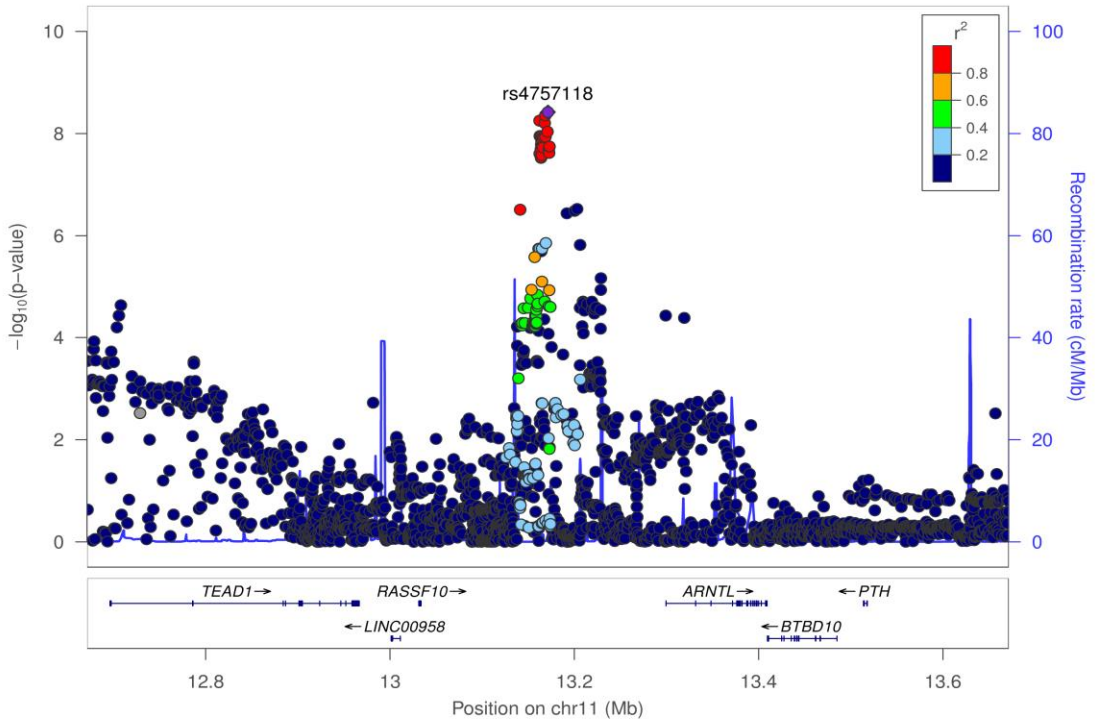
Supplementary Figure 2-57: Regional association plot for rs721917 (*SFTPD* locus at 10q22.3)



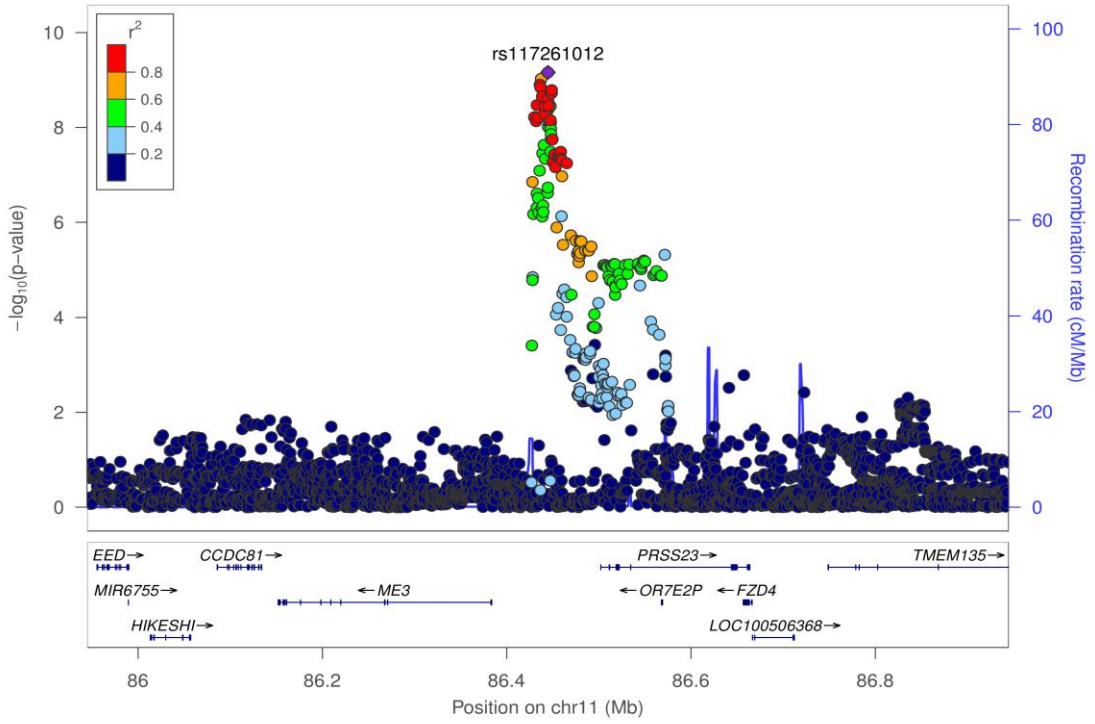
Supplementary Figure 2-58: Regional association plot for rs1570221 (*STN1* locus at 10q24.33)



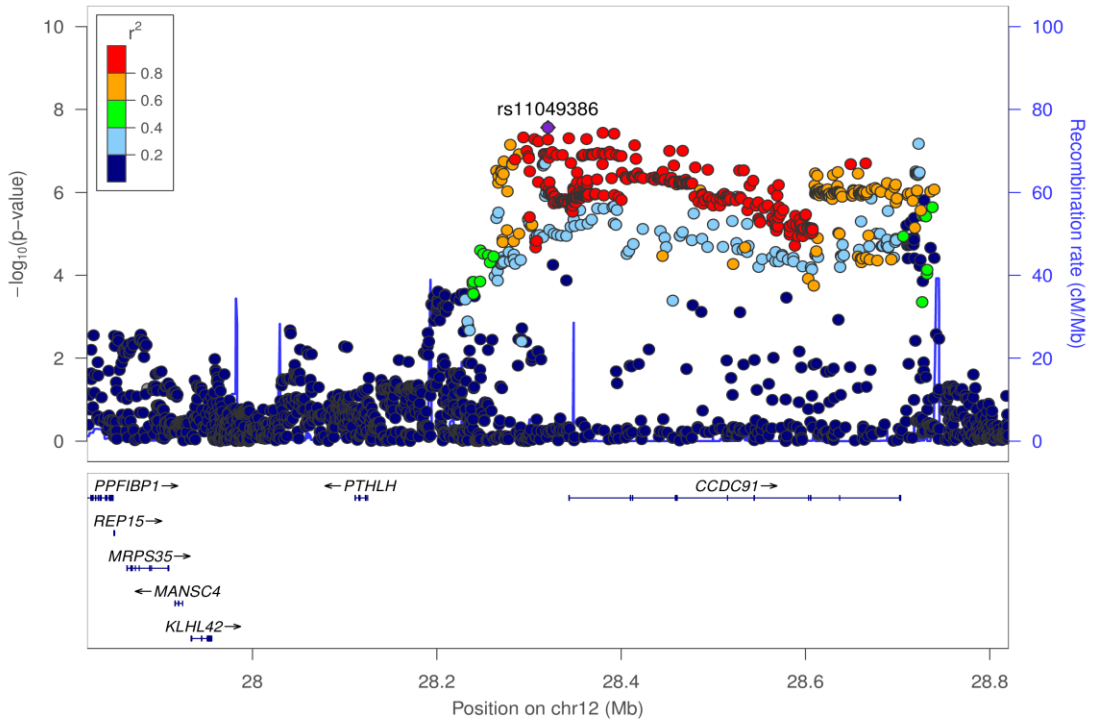
Supplementary Figure 2-59: Regional association plot for rs4757118 (*ARNTL* locus at 11p15.2)



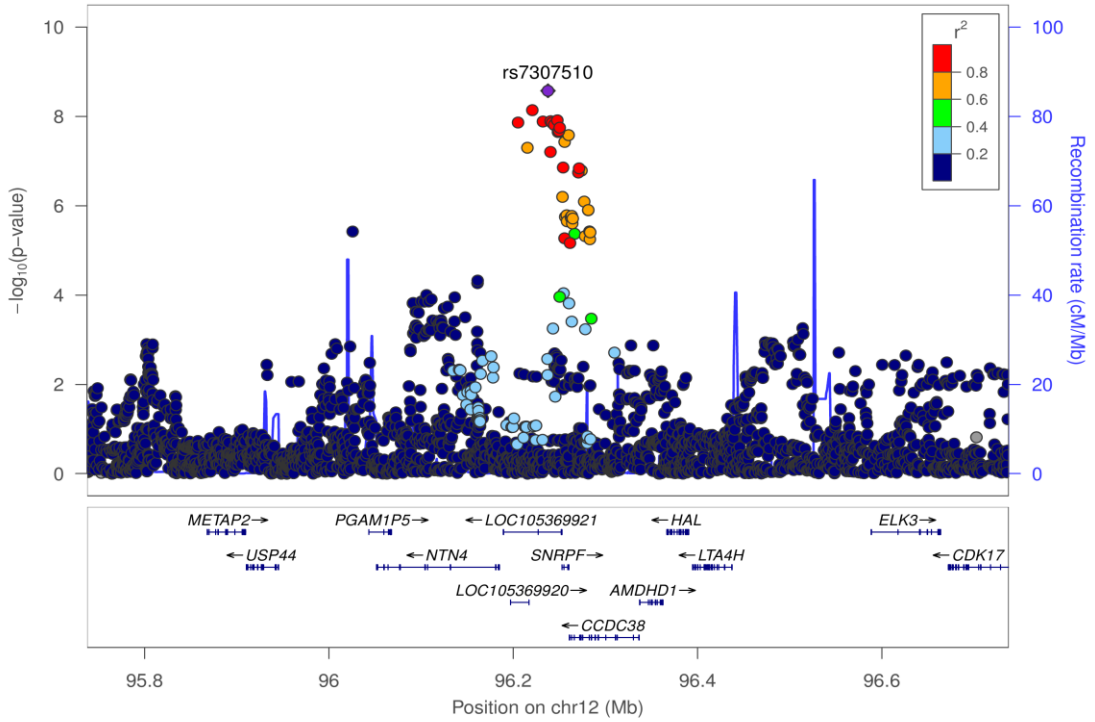
Supplementary Figure 2-60: Regional association plot for rs117261012 (*PRSS23* locus at 11q14.2)



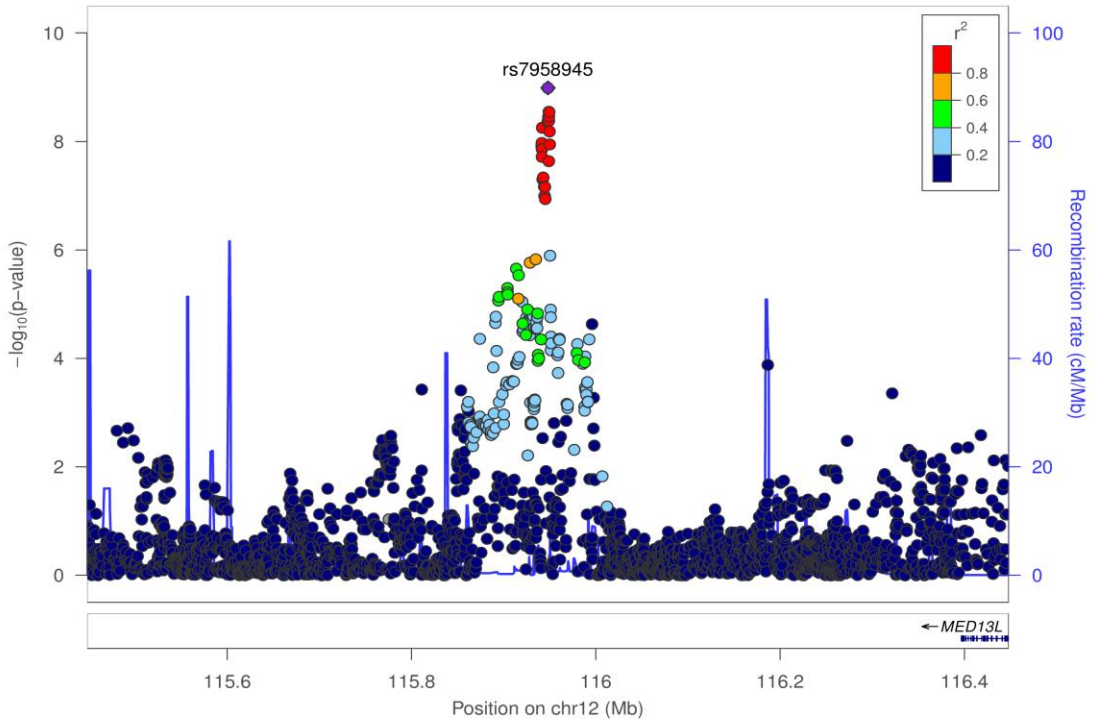
Supplementary Figure 2-61: Regional association plot for rs11049386 (*CCDC91* locus at 12p11.22)



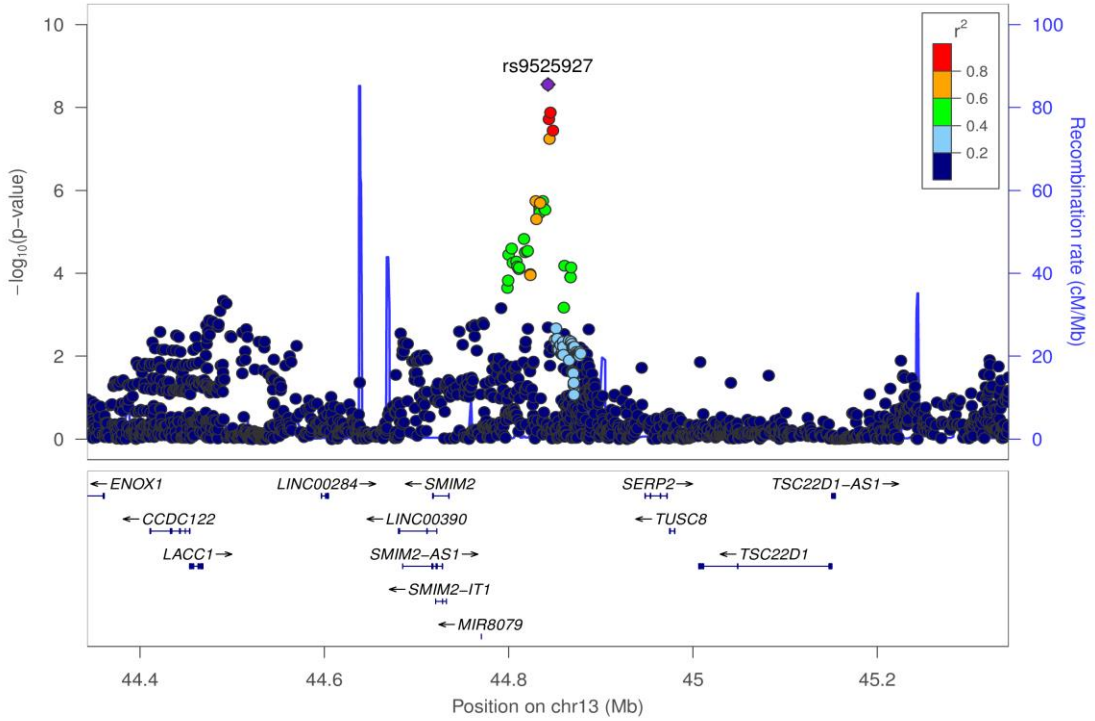
Supplementary Figure 2-62: Regional association plot for rs7307510 (*SNRPF* locus at 12q23.1)



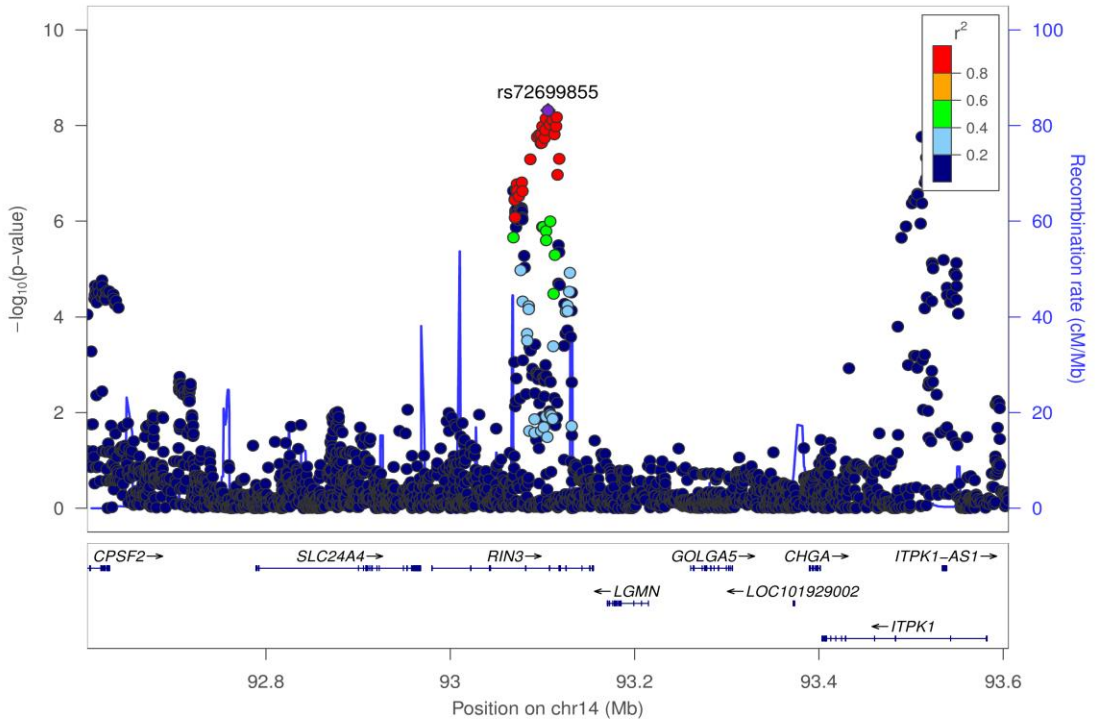
Supplementary Figure 2-63: Regional association plot for rs7958945 (*MED13L* locus at 12q24.21)



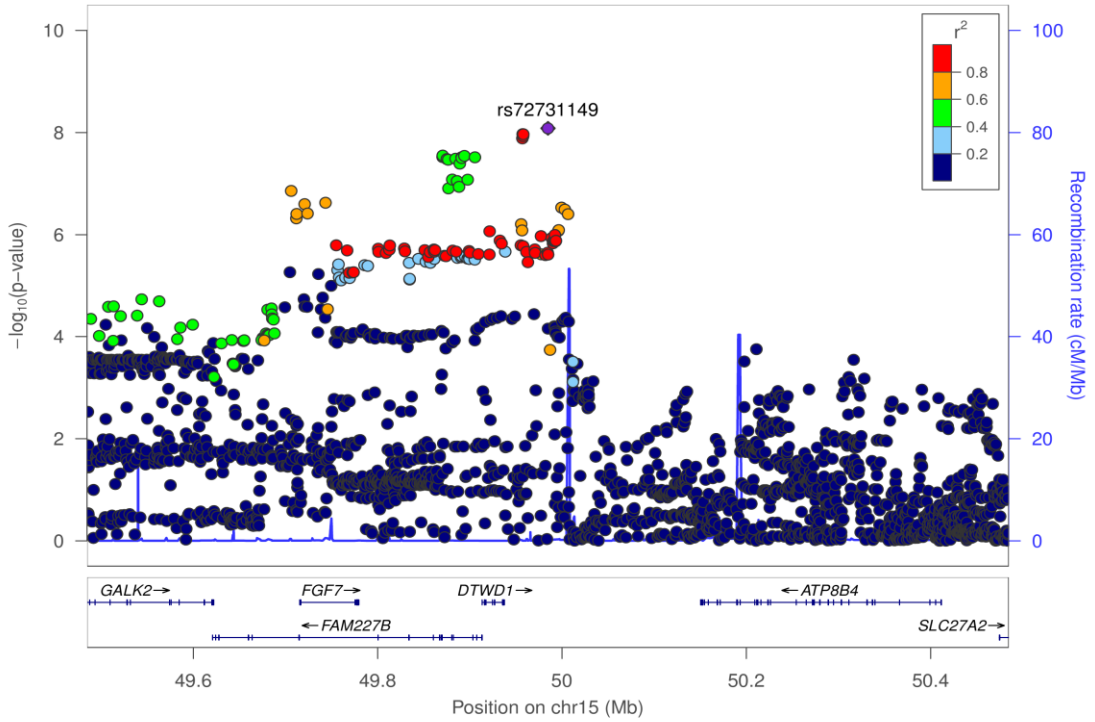
Supplementary Figure 2-64: Regional association plot for rs9525927 (*SERP2* locus at 13q14.11)



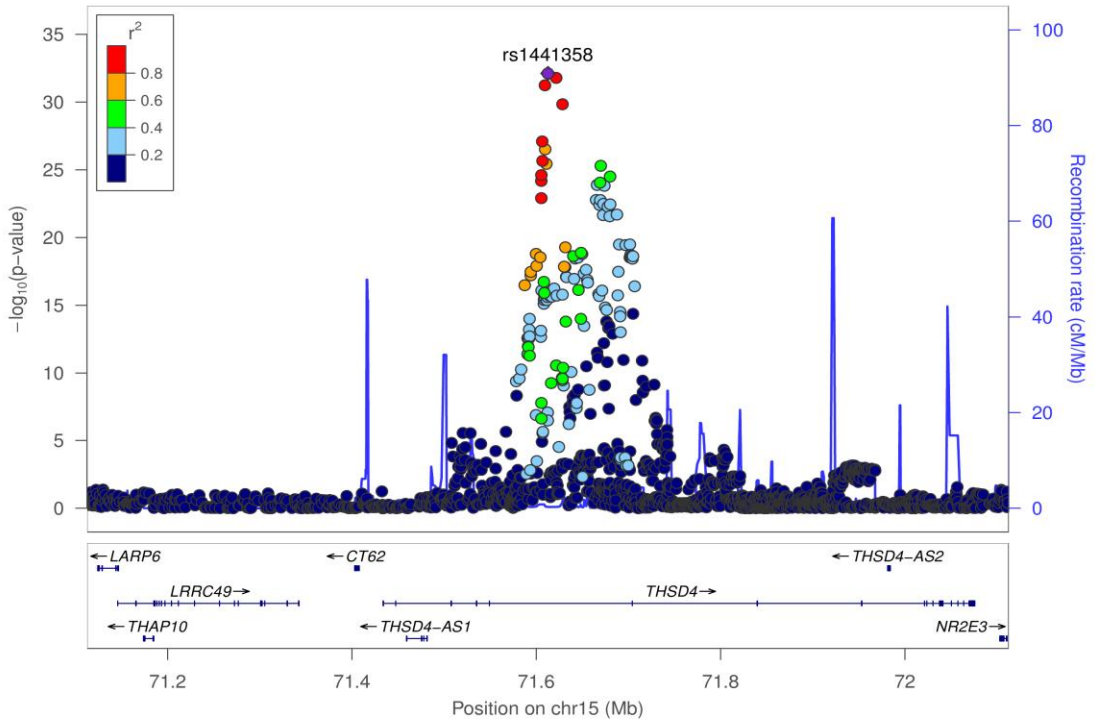
Supplementary Figure 2-65: Regional association plot for rs72699855 (*RIN3* locus at 14q32.12)



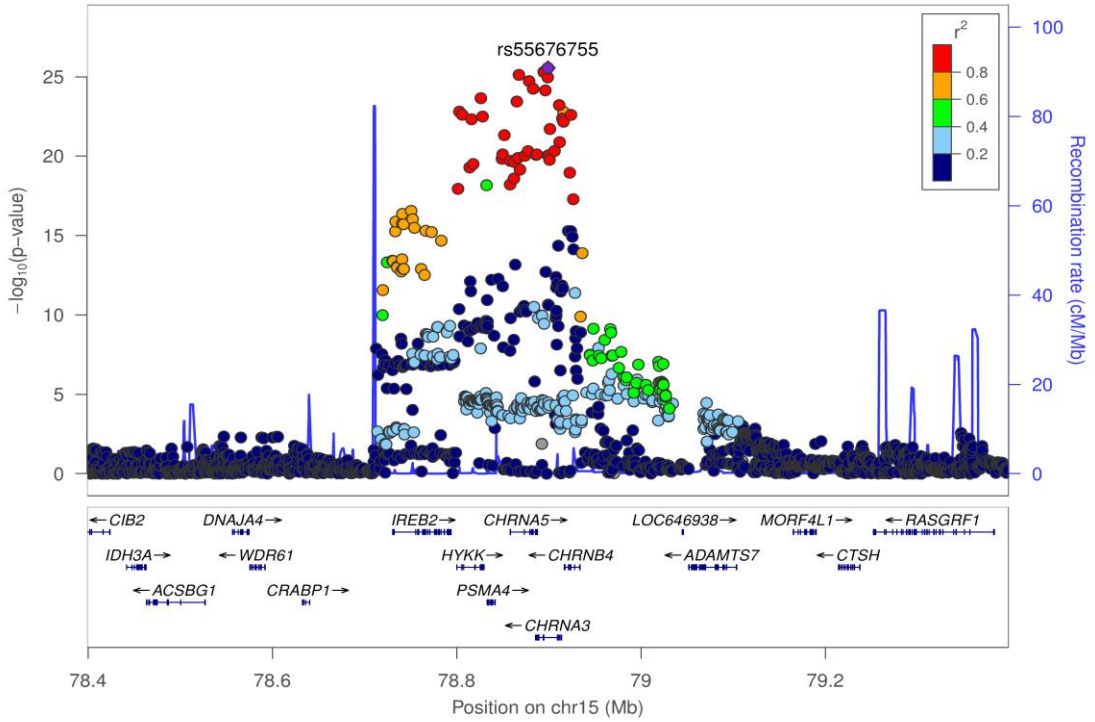
Supplementary Figure 2-66: Regional association plot for rs72731149 (*DTWD1* locus at 15q21.2)



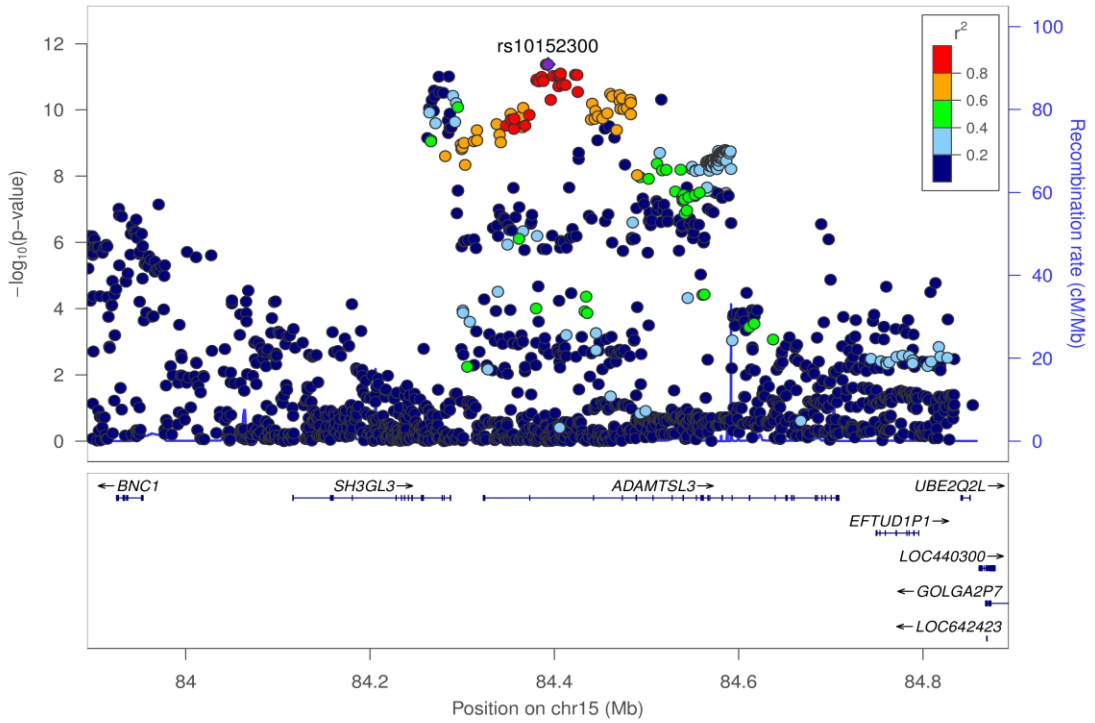
Supplementary Figure 2-67: Regional association plot for rs1441358 (*THSD4* locus at 15q23)



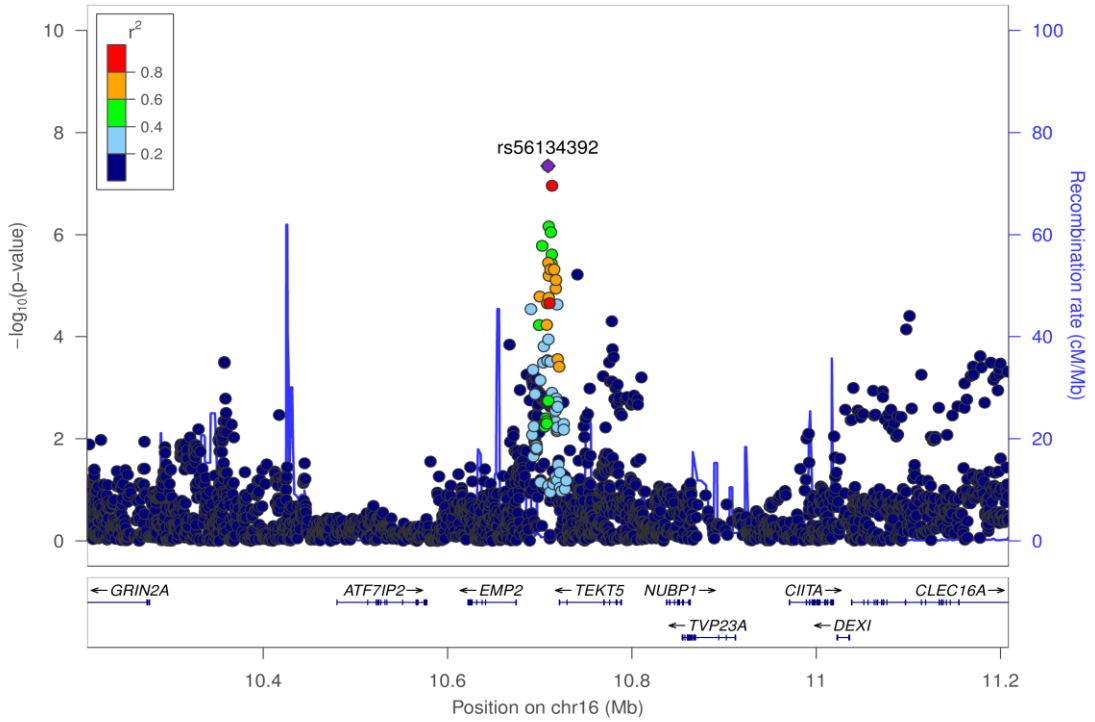
Supplementary Figure 2-68: Regional association plot for rs55676755 (*CHRNA3* locus at 15q25.1)



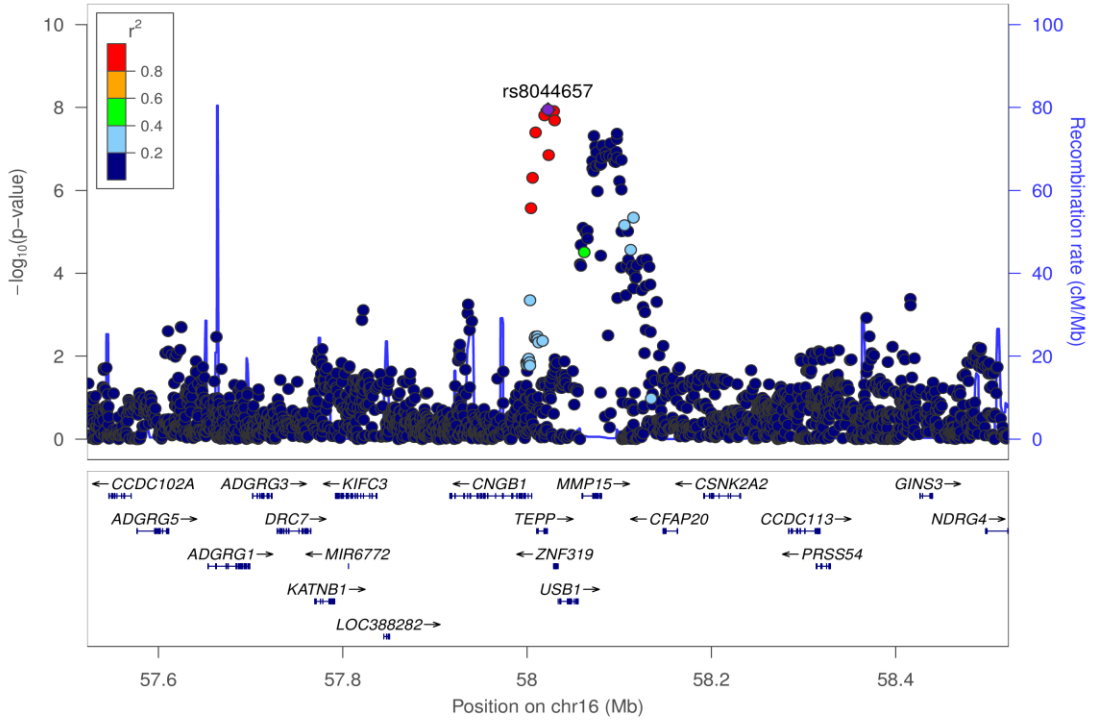
Supplementary Figure 2-69: Regional association plot for rs10152300 (*ADAMTSL3* locus at 15q25.2)



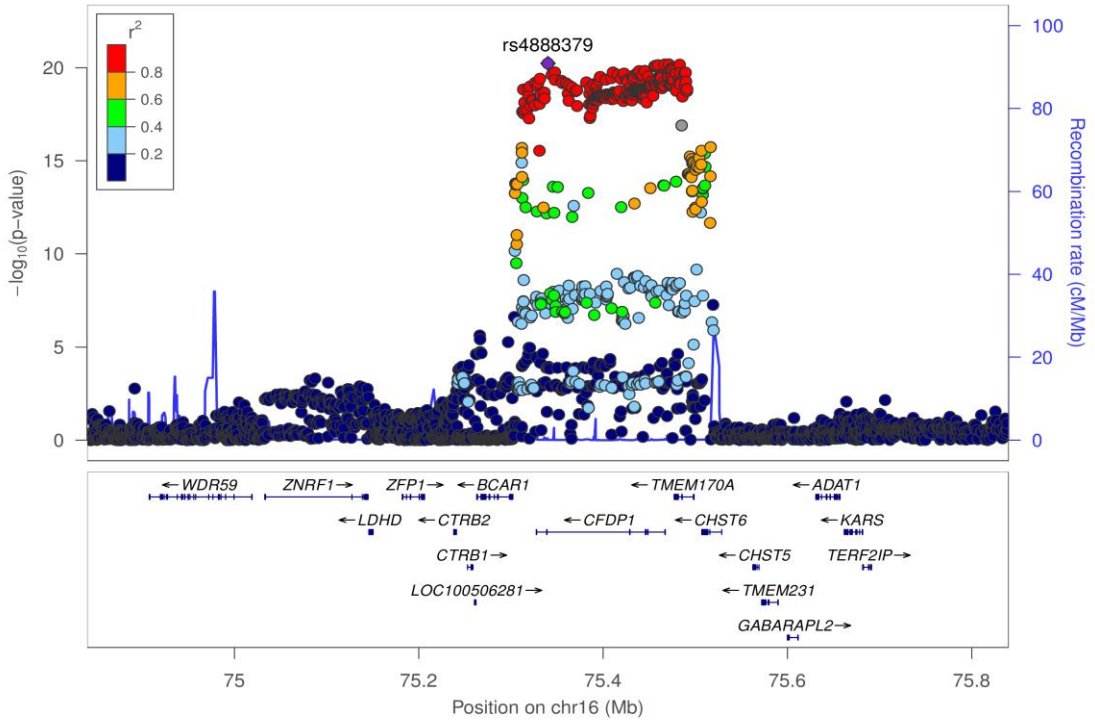
Supplementary Figure 2-70: Regional association plot for rs56134392 (*TEKT5* locus at 16p13.13)



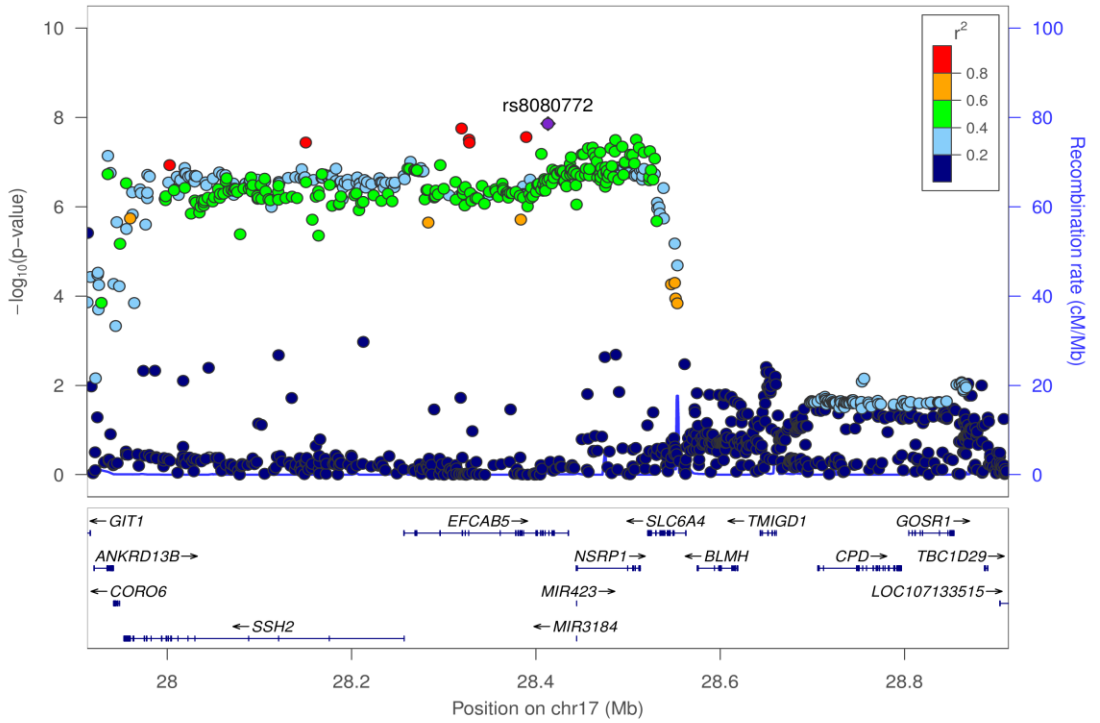
Supplementary Figure 2-71: Regional association plot for rs8044657 (*TEPP* locus at 16q21)



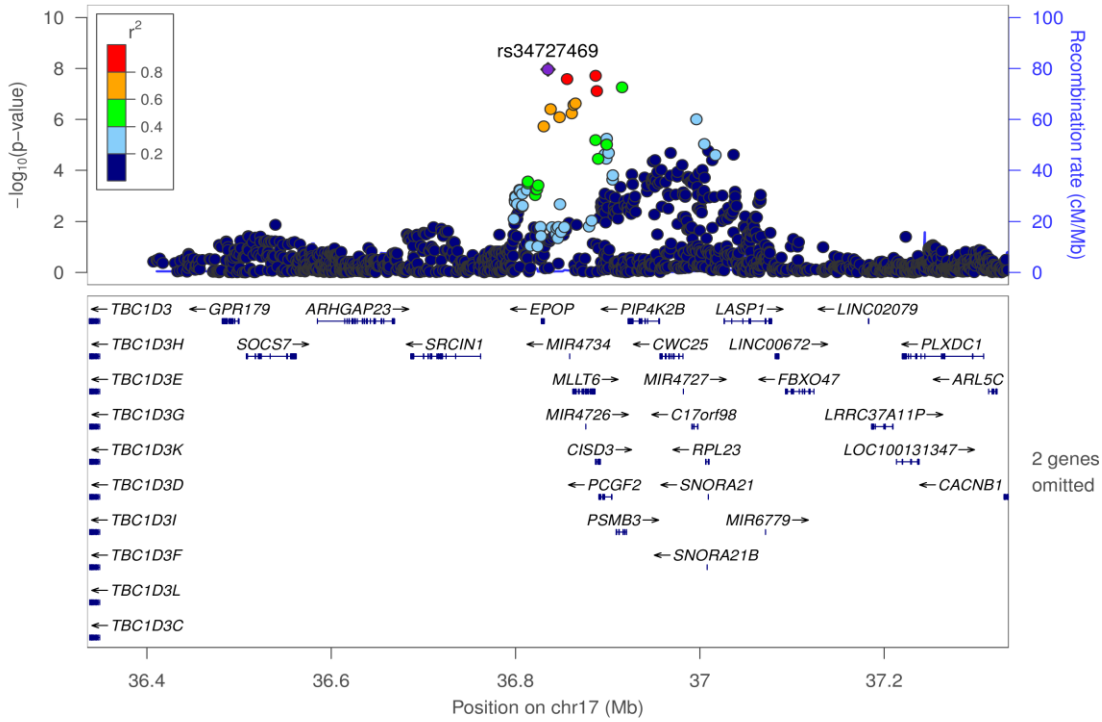
Supplementary Figure 2-72: Regional association plot for rs4888379 (*CFDP1* locus at 16q23.1)



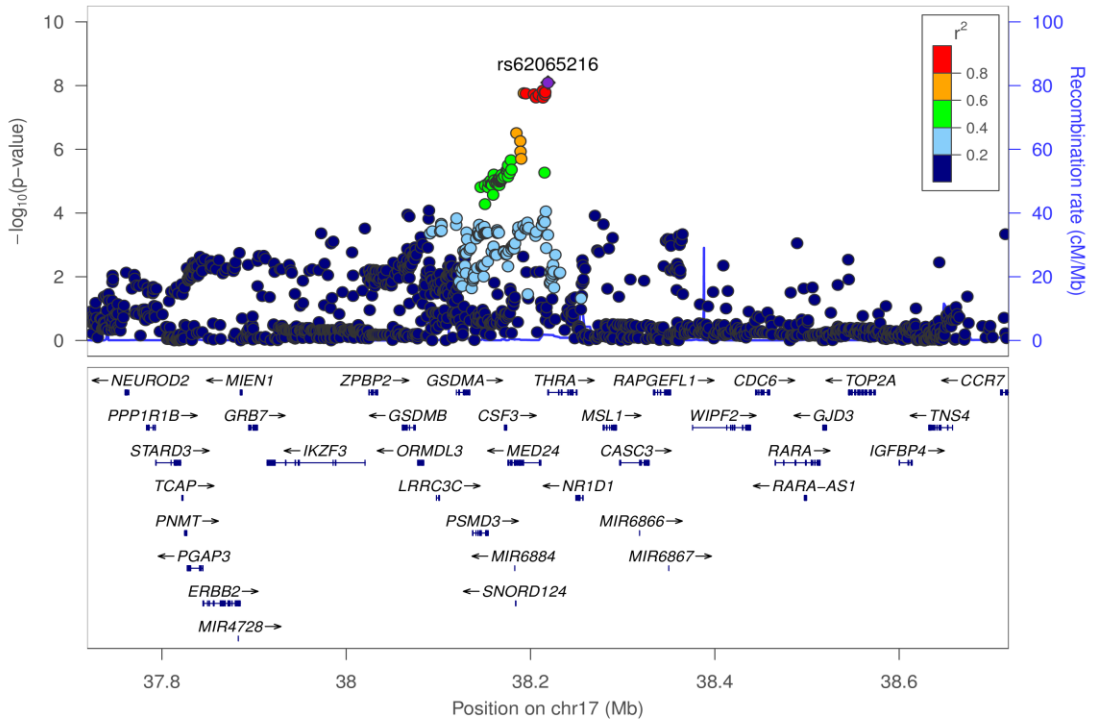
Supplementary Figure 2-73: Regional association plot for rs8080772 (*EFCAB5* locus at 17q11.2)



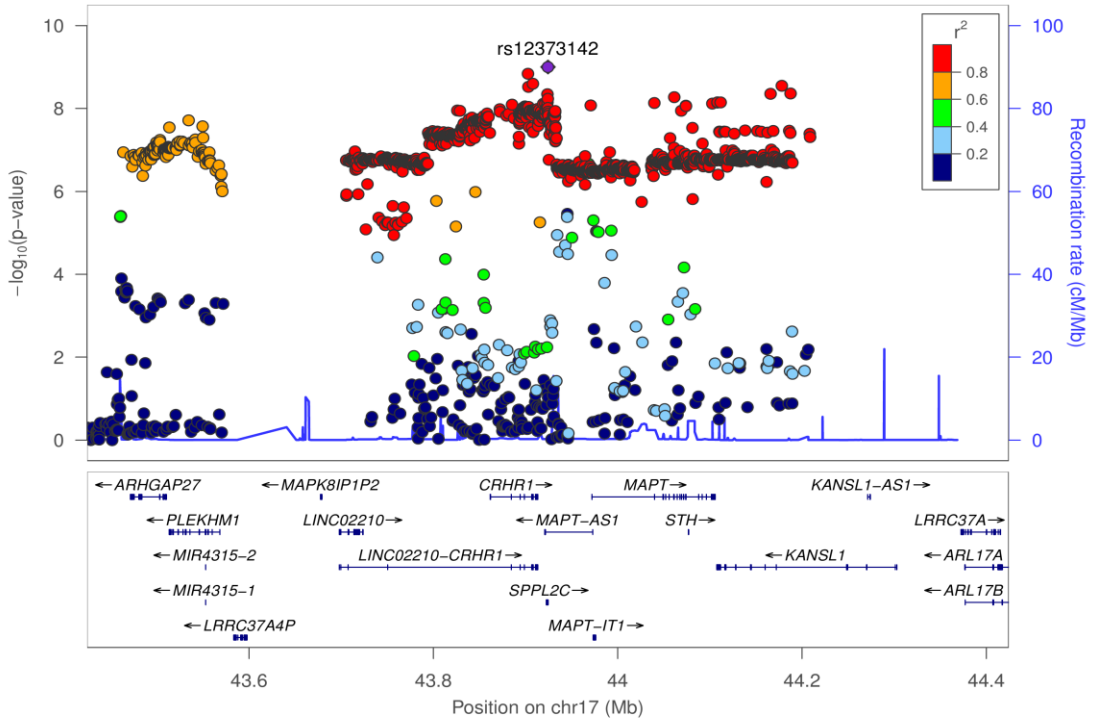
Supplementary Figure 2-74: Regional association plot for rs34727469 (*RPL23* locus at 17q12)



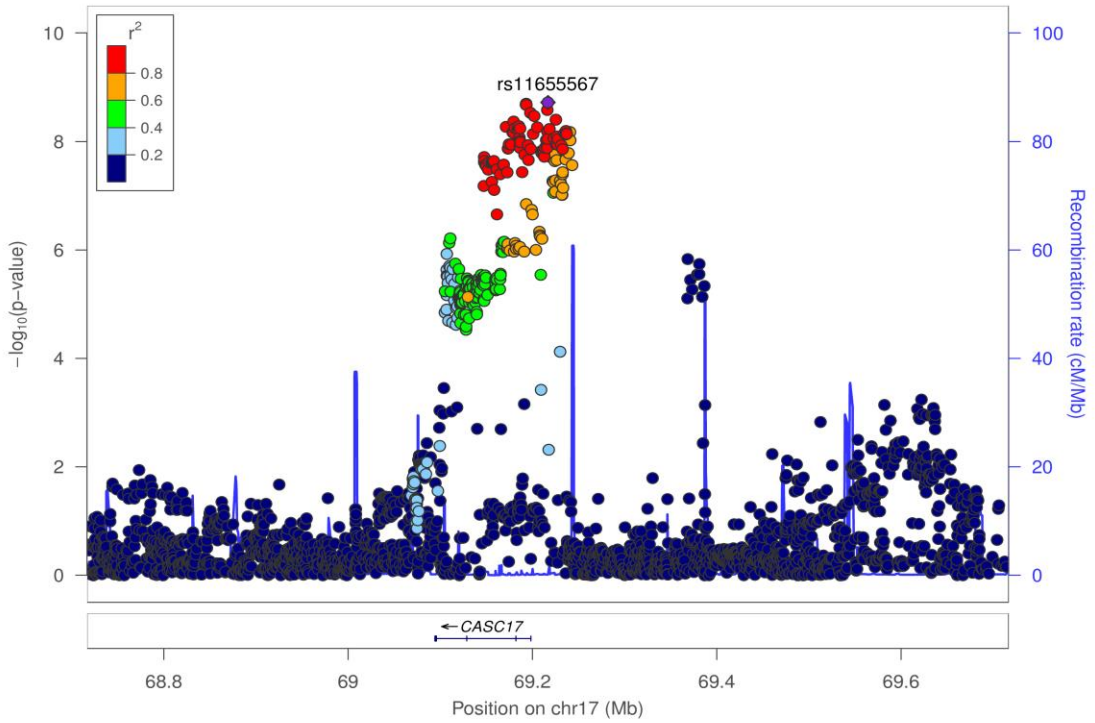
Supplementary Figure 2-75: Regional association plot for rs62065216 (*THRA* locus at 17q21.1)



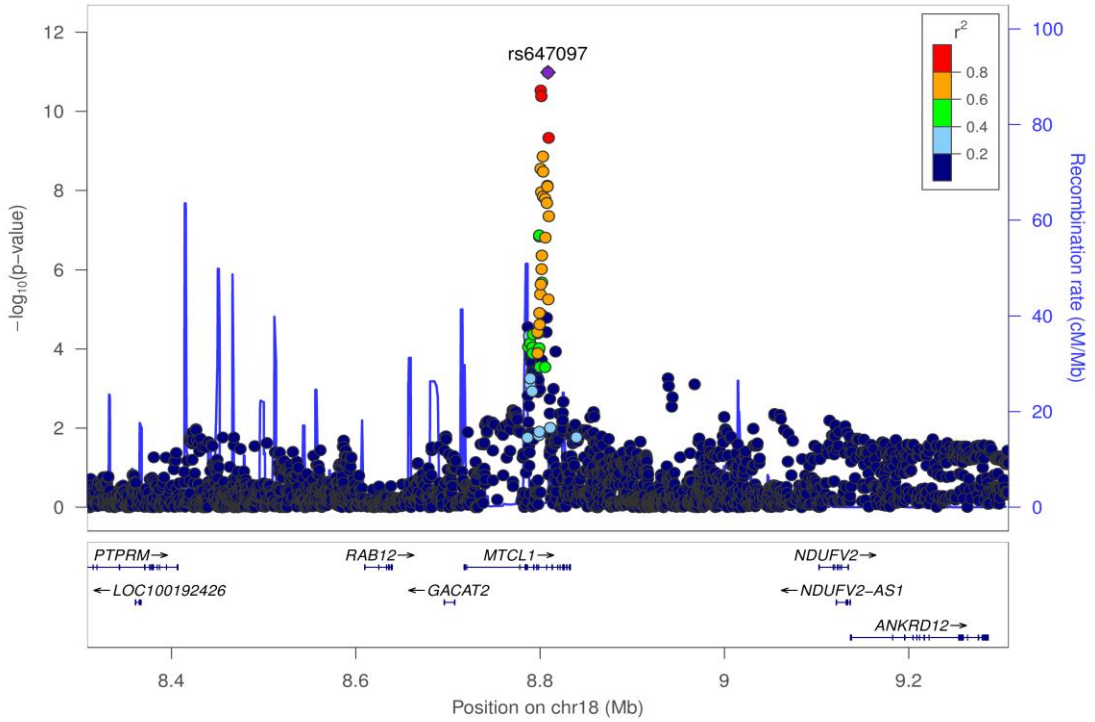
Supplementary Figure 2-76: Regional association plot for rs12373142 (*SPPL2C* locus at 17q21.31)



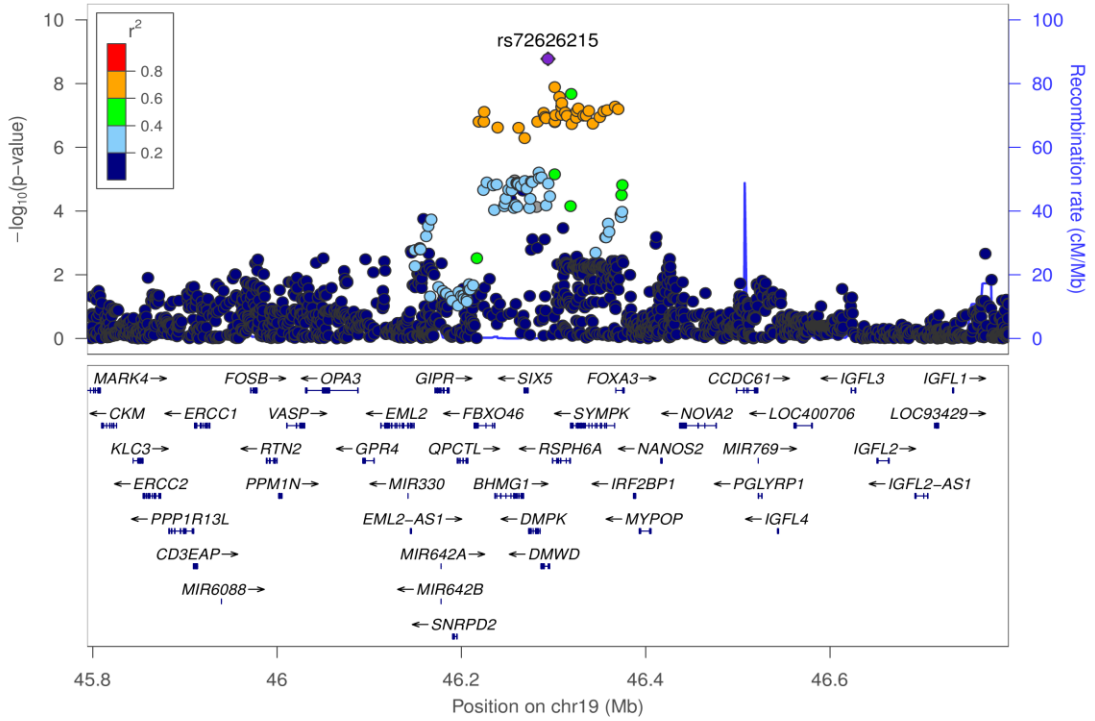
Supplementary Figure 2-77: Regional association plot for rs11655567 (*SOX9* locus at 17q24.3)



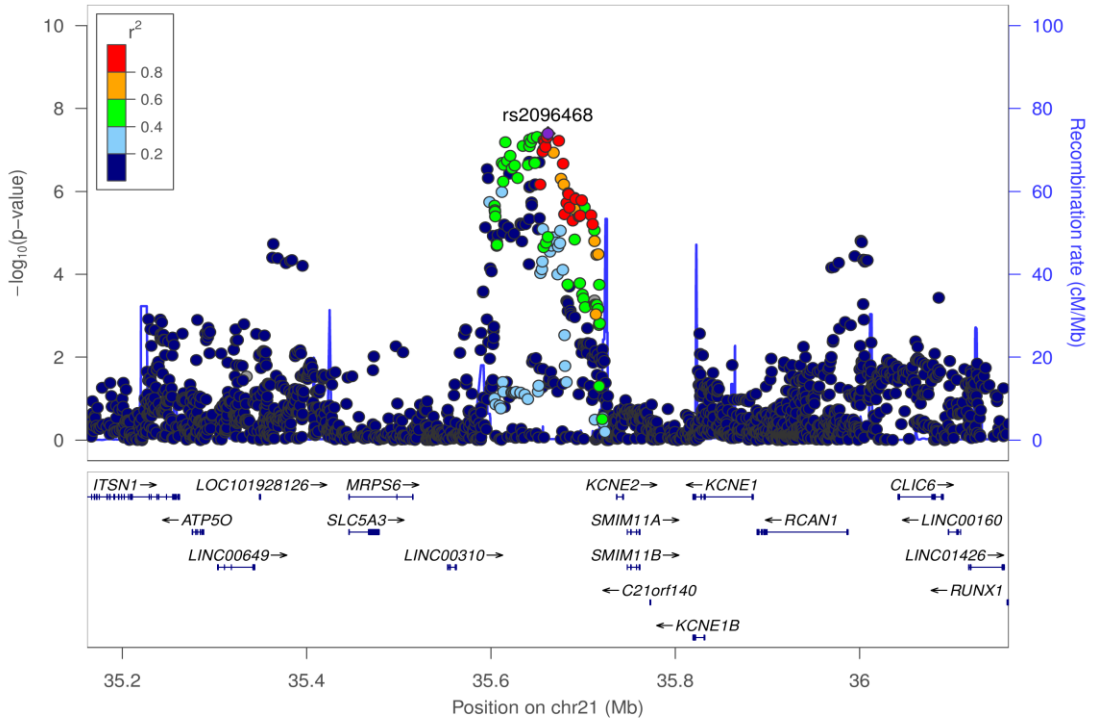
Supplementary Figure 2-78: Regional association plot for rs647097 (*MTCL1* locus at 18p11.22)



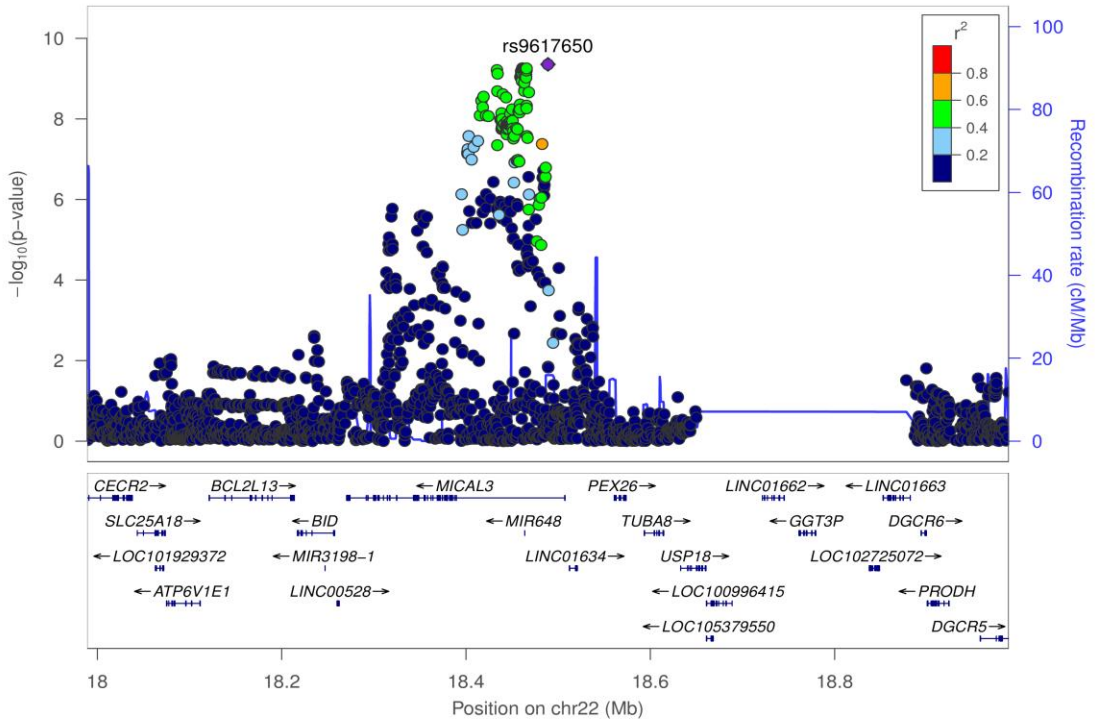
Supplementary Figure 2-79: Regional association plot for rs72626215 (*DMWD* locus at 19q13.32)



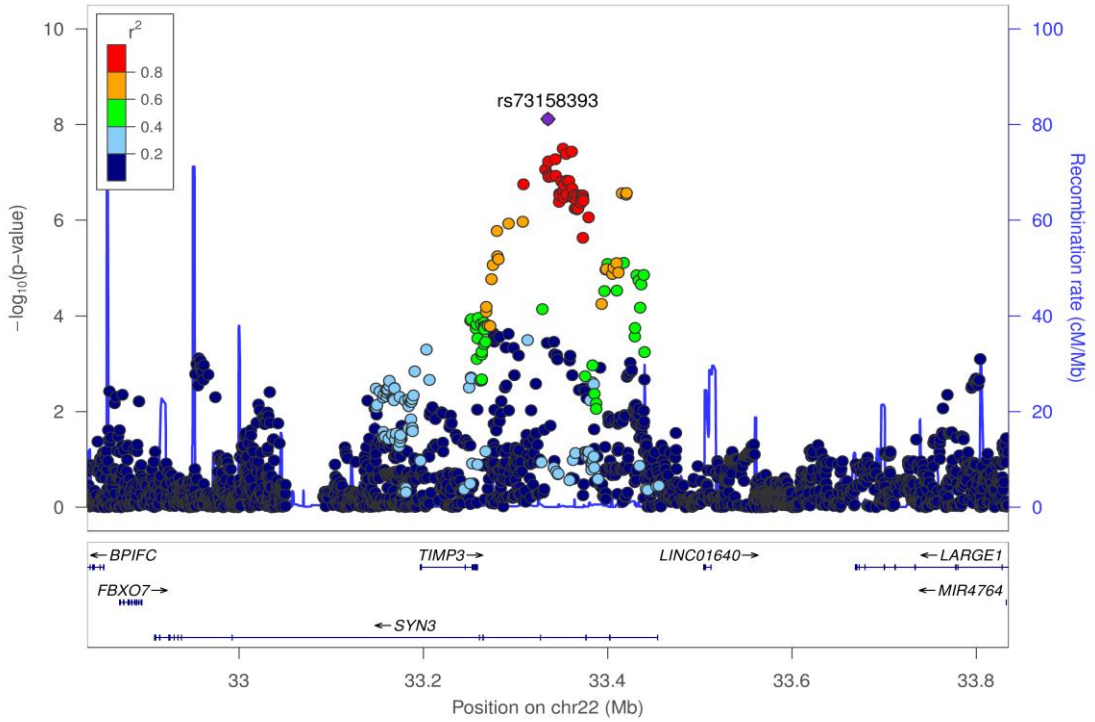
Supplementary Figure 2-80: Regional association plot for rs2096468 (*KCNE2* locus at 21q22.11)



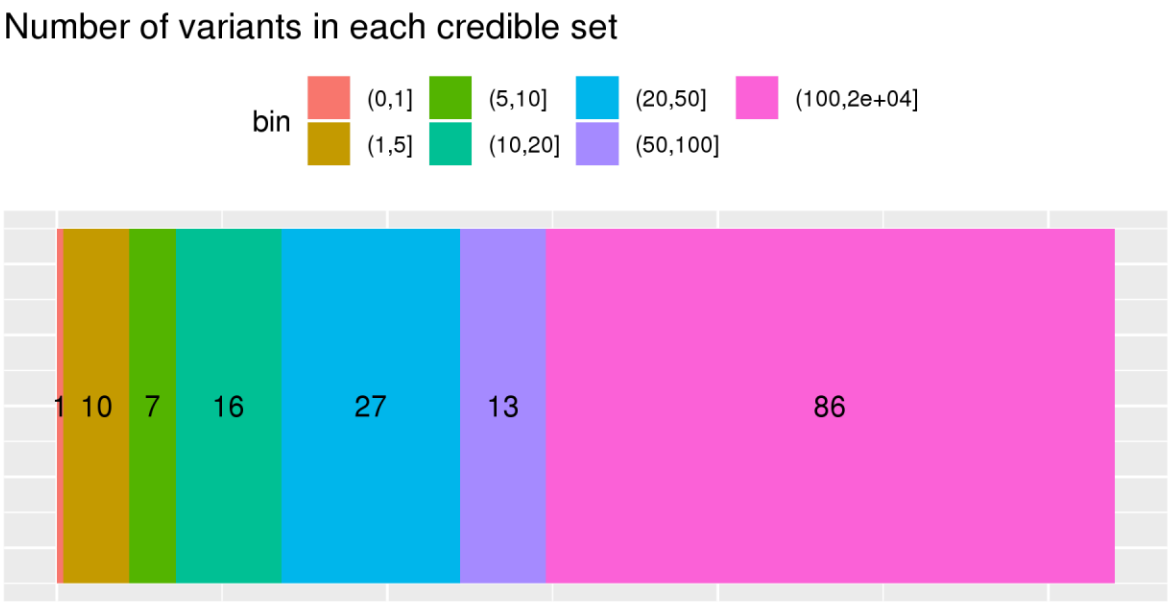
Supplementary Figure 2-81: Regional association plot for rs9617650 (*MICAL3* locus at 22q11.21)



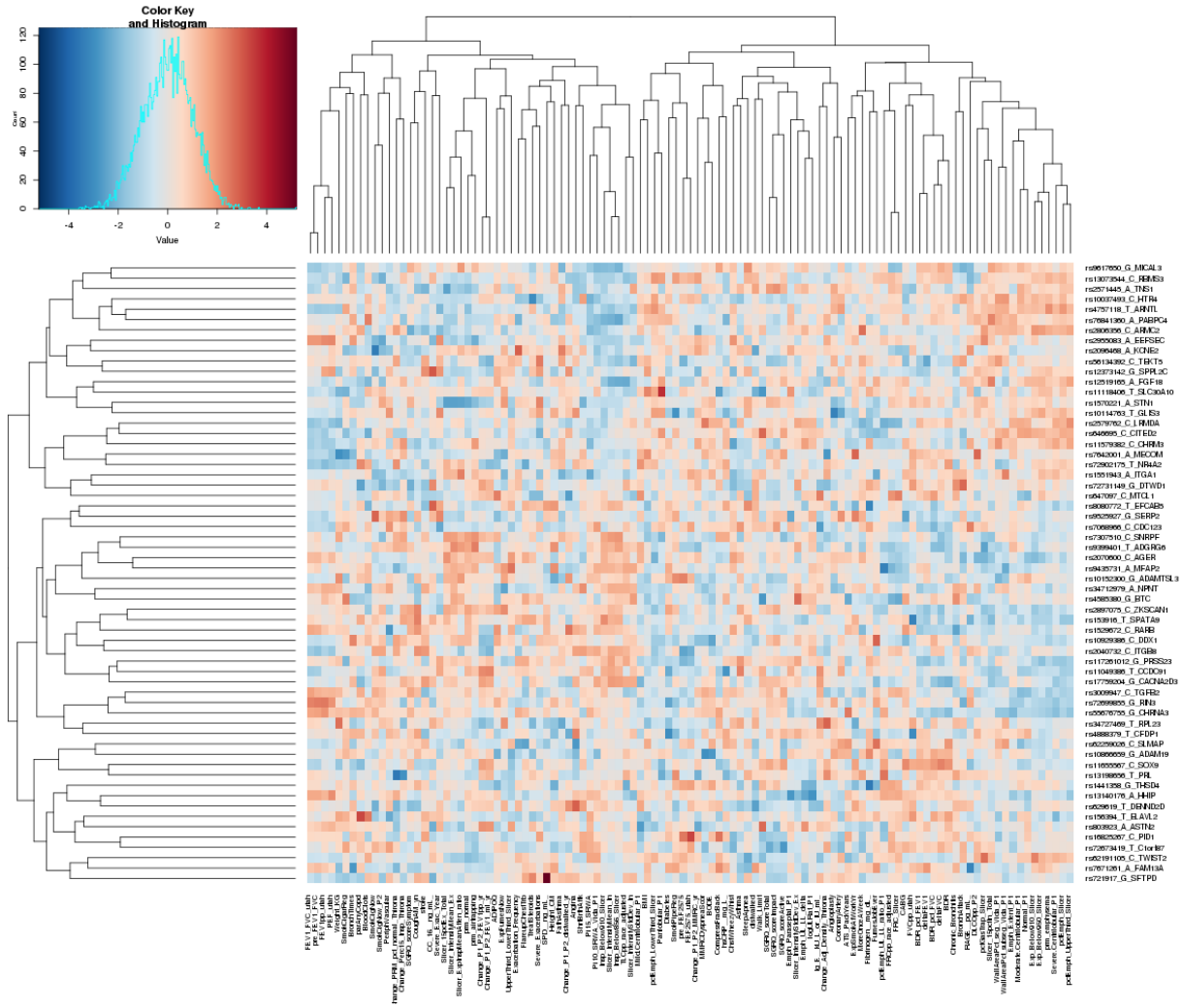
Supplementary Figure 2-82: Regional association plot for rs73158393 (*SYN3* locus at 22q12.3)



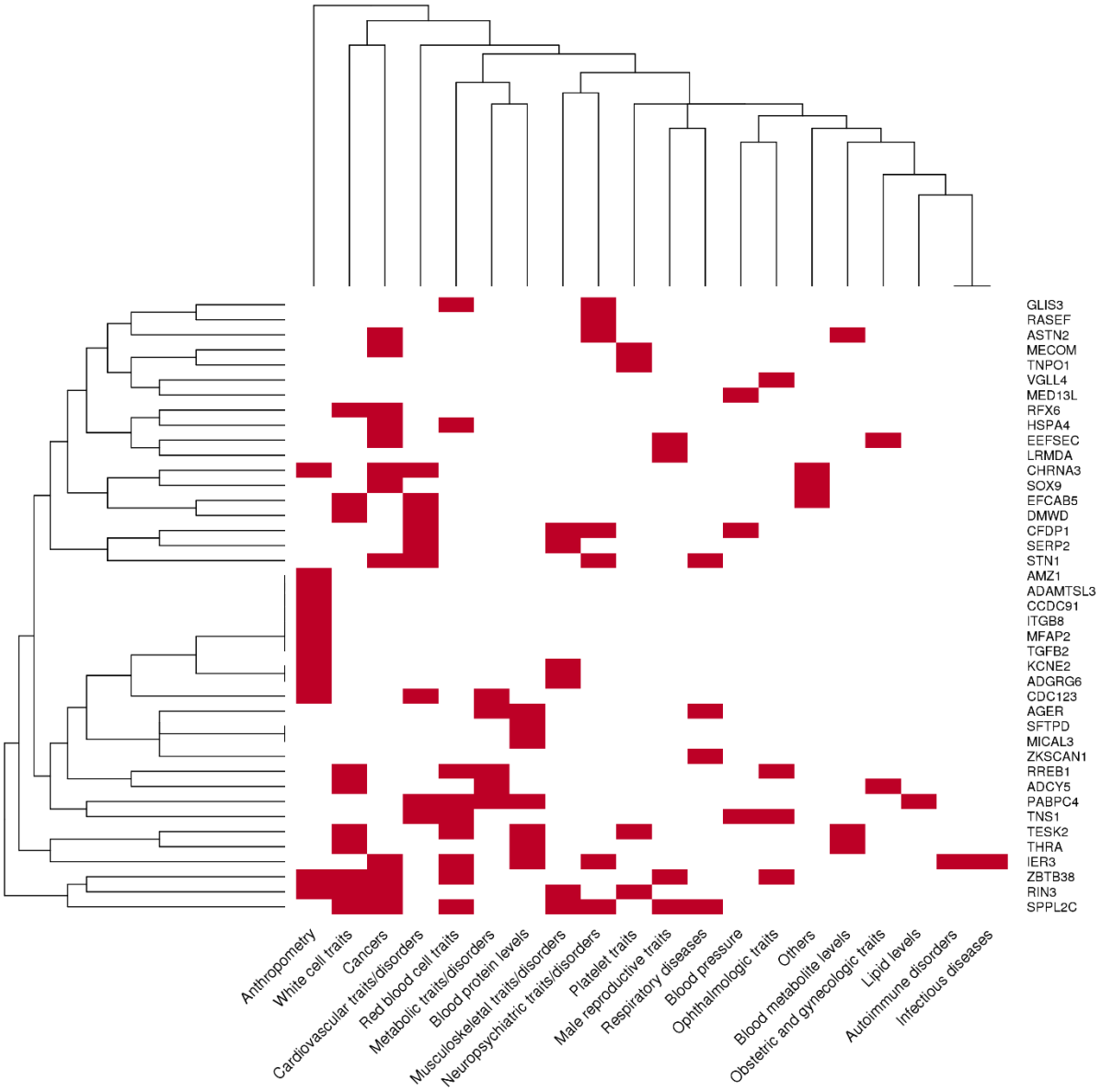
Supplementary Figure 3: Distribution of number of variants in 99% credible sets



Supplementary Figure 4: Heatmap of associations of phenotypes in COPDGene

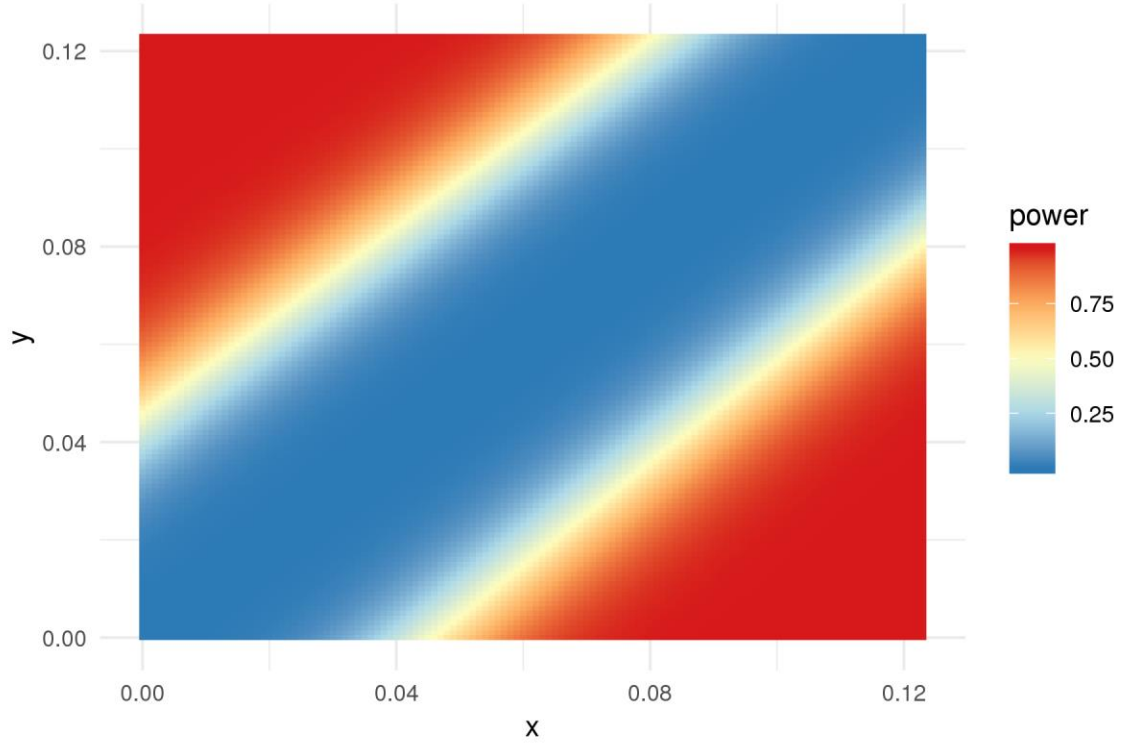


Supplementary Figure 5: Associations of index variants and traits in NHGRI-EBI GWAS Catalog



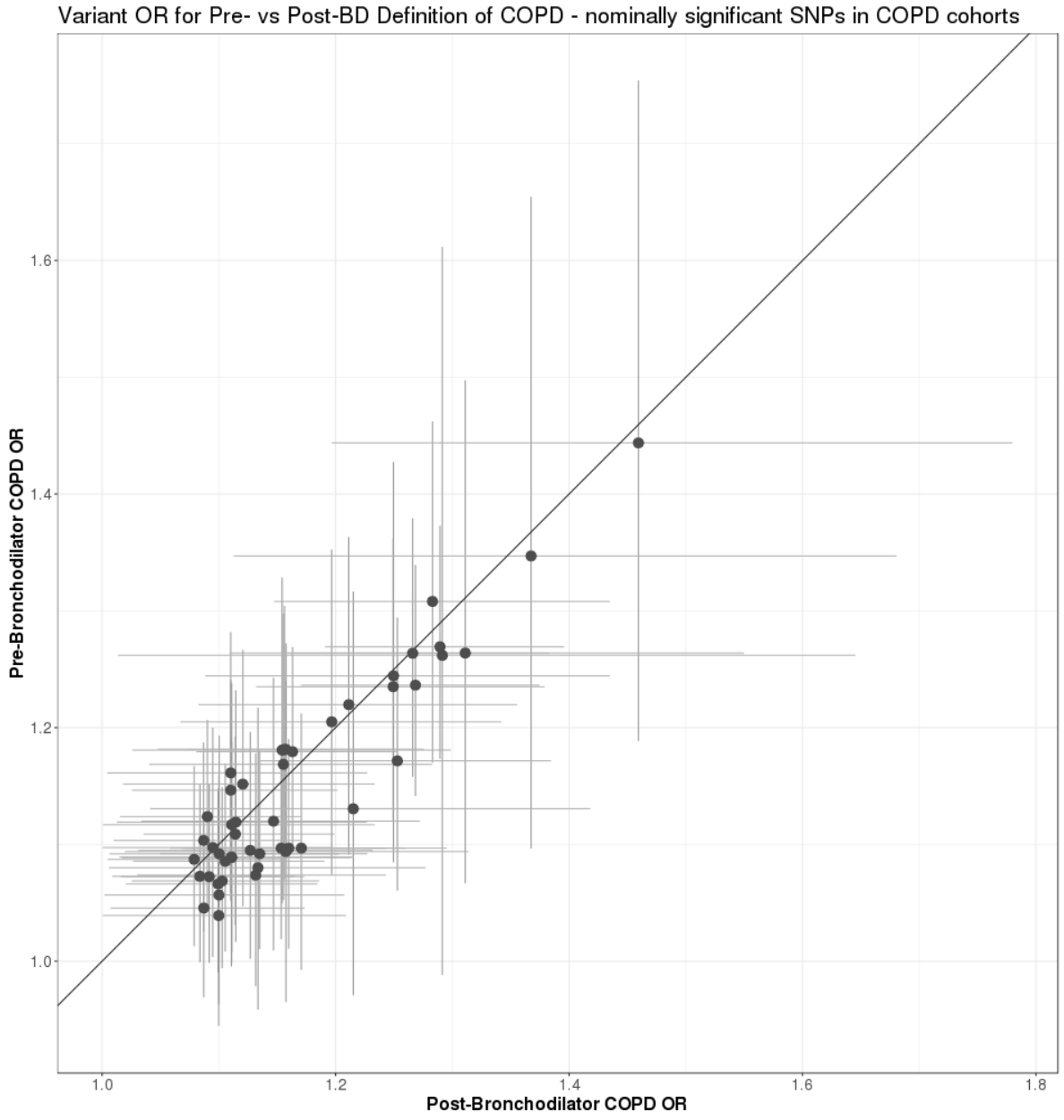
Supplementary Figure 6: Power analysis for sex-difference analysis

The power analysis was based on the effective sample size of 104,119 (male) and 130,707 (female). X-axis and Y-axis represent effect sizes in males and females, respectively.



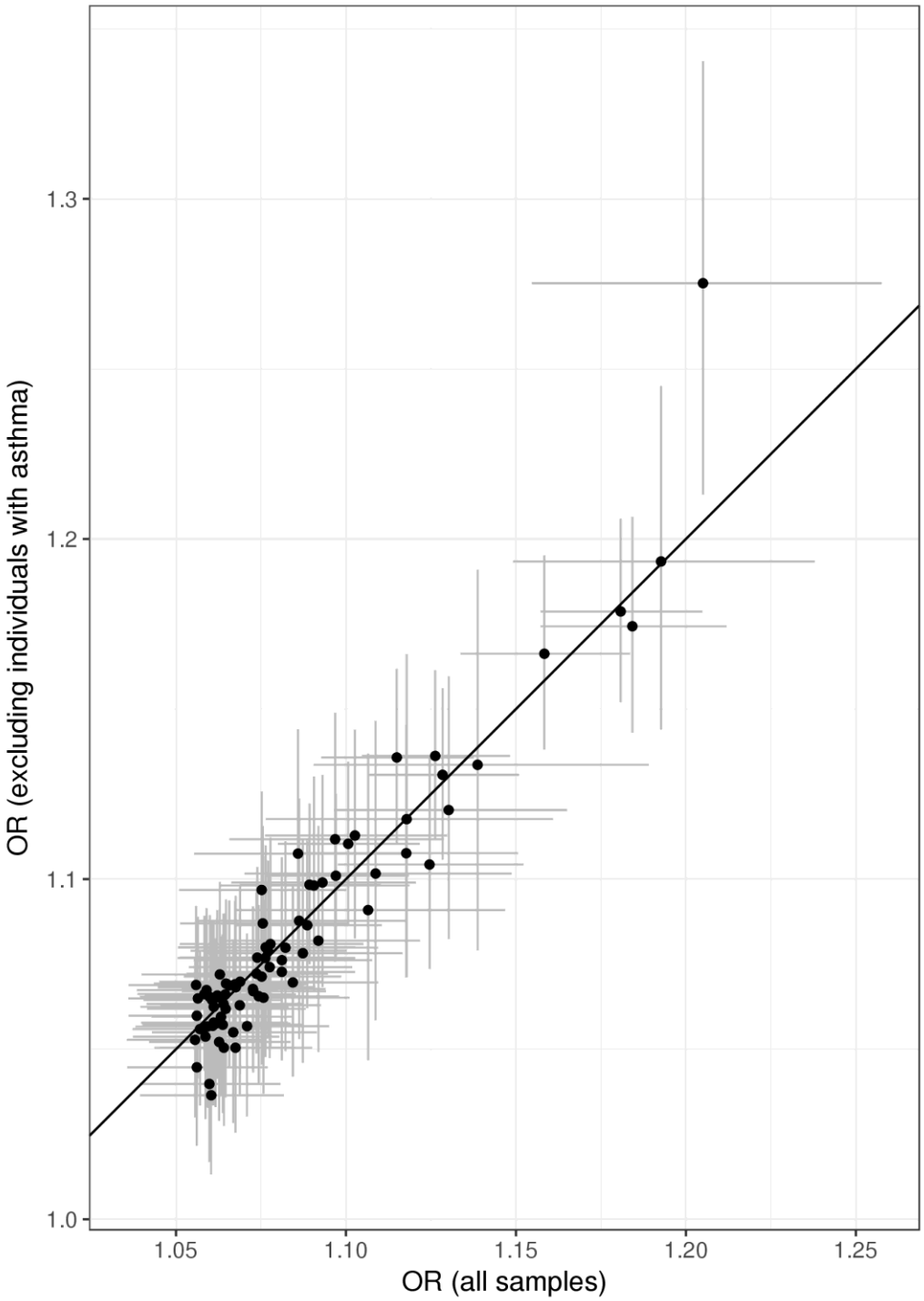
Supplementary Figure 7: Scatter plot of COPD odd ratio of nominally significant SNPs in meta-analysis of a subset of COPD case-control cohorts* using a pre- and post-bronchodilator definition of COPD

*COPD cohorts includes COPDGene AA, COPDGene NHW, ECLIPSE, and GenKOLS
The meta-analysis of a subset of COPD case-control controls included: 666 cases and 1,298 controls in COPDGene AA, 2,442 cases and 1,663 controls in COPDGene NHW, 1,721 cases and 130 controls in ECLIPSE, and 827 cases and 600 controls in GenKOLS. Dots represent odds ratio (OR). Error bars indicate 95% confidence interval.



Supplementary Figure 8: Comparison of odds ratios (OR) including and excluding individuals with asthma of 82 genome-wide significant variants

The association statistics are based on the analysis including (21,081 cases and 179,711 controls), excluding (14,364 cases and 179,711 controls) individuals with asthma. Dots represent odds ratio (OR). Error bars show 95% confidence interval for OR estimates.



Funding and acknowledgements

The content of this study is solely the responsibility of the authors and does not necessarily represent the official views of the National Heart, Lung, And Blood Institute or the National Institutes of Health.

This work was supported by R01 HL089897 (J.D.C.); P01 HL114501 (Augustine Choi); and a VA Research Career Scientist award (D.S.).

The research undertaken by M.D.T., N.F.R., A.L.G., V.E.J., L.V.W., N.S., and I.P.H. was partly funded by the National Institute for Health Research (NIHR). The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health. This was also funded by the Medical Research Council (MRC) and the Wellcome Trust. M.D.T. is supported by a Wellcome Trust Investigator Award (WT202849/Z/16/Z). M.D.T. and L.V.W. have been supported by the MRC (MR/N011317/1). L.V.W. holds a GSK/British Lung Foundation Chair in Respiratory Research.

The **Atherosclerosis Risk in Communities Study** is carried out as a collaborative study supported by National Heart, Lung, and Blood Institute contracts (HHSN268201100005C, HHSN268201100006C, HHSN268201100007C, HHSN268201100008C, HHSN268201100009C, HHSN268201100010C, HHSN268201100011C, and HHSN268201100012C), R01HL087641, R01HL59367 and R01HL086694; National Human Genome Research Institute contract U01HG004402; and National Institutes of Health contract HHSN268200625226C. Infrastructure was partly supported by Grant Number UL1RR025005, a component of the National Institutes of Health and NIH Roadmap for Medical Research.. This work was also supported in part by the Intramural Research Program of the NIH, National Institute of Environmental Health Sciences.

This **CHS** research was supported by NHLBI contracts HHSN268201200036C, HHSN268200800007C, HHSN268200960009C, HHSN268201800001C, N01HC55222, N01HC85079, N01HC85080, N01HC85081, N01HC85082, N01HC85083, N01HC85086; and NHLBI grants U01HL080295, R01HL087652, R01HL105756, R01HL103612, R01HL085251, R01HL120393, and U01HL130114 with additional contribution from the National Institute of Neurological Disorders and Stroke (NINDS). Additional support was provided through R01AG023629 from the National Institute on Aging (NIA). A full list of principal CHS investigators and institutions can be found at CHS-NHLBI.org.
The provision of genotyping data was supported in part by the National Center for Advancing Translational Sciences, CTSI grant UL1TR001881, and the National Institute of Diabetes and Digestive and Kidney Disease Diabetes Research Center (DRC) grant DK063491 to the Southern California Diabetes Endocrinology Research Center.

The **COPDGene** project (NCT00608764) was supported by Award Number U01 HL089897 and Award Number U01 HL089856 from the National Heart, Lung, and Blood Institute. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Heart, Lung, and Blood Institute or the National Institutes of Health.

COPD Foundation Funding

The COPDGene® project is also supported by the COPD Foundation through contributions made to an Industry Advisory Board comprised of AstraZeneca, Boehringer Ingelheim, GlaxoSmithKline, Novartis, Pfizer, Siemens and Sunovion.

COPDGene® Investigators – Core Units

Administrative Center: James D. Crapo, MD (PI); Edwin K. Silverman, MD, PhD (PI); Barry J. Make, MD; Elizabeth A. Regan, MD, PhD

Genetic Analysis Center: Terri Beaty, PhD; Ferdouse Begum, PhD; Peter J. Castaldi, MD, MSc; Michael Cho, MD; Dawn L. DeMeo, MD, MPH; Adel R. Boueiz, MD; Marilyn G. Foreman, MD, MS; Eitan Halper-Stromberg; Lystra P. Hayden, MD, MMSc; Craig P. Hersh, MD, MPH; Jacqueline Hetmanski, MS, MPH; Brian D. Hobbs, MD; John E. Hokanson, MPH, PhD; Nan Laird, PhD; Christoph Lange, PhD; Sharon M. Lutz, PhD; Merry-Lynn McDonald, PhD; Margaret M. Parker, PhD; Dandi Qiao, PhD; Elizabeth A. Regan, MD, PhD; Edwin K. Silverman, MD, PhD; Emily S. Wan, MD; Sungho Won, Ph.D.; Phuwanat Sakornsakolpat, M.D.; Dmitry Prokopenko, Ph.D.

Imaging Center: Mustafa Al Qaisi, MD; Harvey O. Coxson, PhD; Teresa Gray; MeiLan K. Han, MD, MS; Eric A. Hoffman, PhD; Stephen Humphries, PhD; Francine L. Jacobson, MD, MPH; Philip F. Judy, PhD; Ella A. Kazerooni, MD; Alex Kluiber; David A. Lynch, MB; John D. Newell, Jr., MD; Elizabeth A. Regan, MD, PhD; James C. Ross, PhD; Raul San Jose Estepar, PhD; Joyce Schroeder, MD; Jered Sieren; Douglas Stinson; Berend C. Stoel, PhD; Juerg Tschirren, PhD; Edwin Van Beek, MD, PhD; Bram van Ginneken, PhD; Eva van Rikxoort, PhD; George Washko, MD; Carla G. Wilson, MS;

PFT QA Center, Salt Lake City, UT: Robert Jensen, PhD

Data Coordinating Center and Biostatistics, *National Jewish Health, Denver, CO*: Douglas Everett, PhD; Jim Crooks, PhD; Camille Moore, PhD; Matt Strand, PhD; Carla G. Wilson, MS

Epidemiology Core, *University of Colorado Anschutz Medical Campus, Aurora, CO*: John E. Hokanson, MPH, PhD; John Hughes, PhD; Gregory Kinney, MPH, PhD; Sharon M. Lutz, PhD; Katherine Pratte, MSPH; Kendra A. Young, PhD

Mortality Adjudication Core: Surya Bhatt, MD; Jessica Bon, MD; MeiLan K. Han, MD, MS; Barry Make, MD; Carlos Martinez, MD, MS; Susan Murray, ScD; Elizabeth Regan, MD; Xavier Soler, MD; Carla G. Wilson, MS

Biomarker Core: Russell P. Bowler, MD, PhD; Katerina Kechris, PhD; Farnoush Banaei-Kashani, Ph.D.

COPDGene® Investigators – Clinical Centers

Ann Arbor VA: Jeffrey L. Curtis, MD; Carlos H. Martinez, MD, MPH; Perry G. Pernicano, MD.

Baylor College of Medicine, Houston, TX: Nicola Hanania, MD, MS; Philip Alapat, MD; Mustafa Atik, MD; Venkata Bandi, MD; Aladin Boriek, PhD; Kalpatha Guntupalli, MD; Elizabeth Guy, MD; Arun Nachiappan, MD; Amit Parulekar, MD;

Brigham and Women's Hospital, Boston, MA: Dawn L. DeMeo, MD, MPH; Craig Hersh, MD, MPH; Francine L. Jacobson, MD, MPH; George Washko, MD

Columbia University, New York, NY: R. Graham Barr, MD, DrPH; John Austin, MD; Belinda D'Souza, MD; Gregory D.N. Pearson, MD; Anna Rozenshtein, MD, MPH, FACR; Byron Thomashow, MD

Duke University Medical Center, Durham, NC: Neil MacIntyre, Jr., MD; H. Page McAdams, MD; Lacey Washington, MD

HealthPartners Research Institute, Minneapolis, MN: Charlene McEvoy, MD, MPH; Joseph Tashjian, MD

Johns Hopkins University, Baltimore, MD: Robert Wise, MD; Robert Brown, MD; Nadia N. Hansel, MD, MPH; Karen Horton, MD; Allison Lambert, MD, MHS; Nirupama Putcha, MD, MHS

Los Angeles Biomedical Research Institute at Harbor UCLA Medical Center, Torrance, CA: Richard Casaburi, PhD, MD; Alessandra Adami, PhD; Matthew Budoff, MD; Hans Fischer, MD; Janos Porszasz, MD, PhD; Harry Rossiter, PhD; William Stringer, MD

Michael E. DeBakey VAMC, Houston, *TX*: Amir Sharafkhaneh, MD, PhD; Charlie Lan, DO

Minneapolis VA: Christine Wendt, MD; Brian Bell, MD

Morehouse School of Medicine, Atlanta, GA: Marilyn G. Foreman, MD, MS; Eugene Berkowitz, MD, PhD; Gloria Westney, MD, MS.

National Jewish Health, Denver, CO: Russell Bowler, MD, PhD; David A. Lynch, MB

Reliant Medical Group, Worcester, MA: Richard Rosiello, MD; David Pace, MD
Temple University, Philadelphia, PA: Gerard Criner, MD; David Ciccolella, MD; Francis Cordova, MD; Chandra Dass, MD; Gilbert D'Alonzo, DO; Parag Desai, MD; Michael Jacobs, PharmD; Steven Kelsen, MD, PhD; Victor Kim, MD; A. James Mamary, MD; Nathaniel Marchetti, DO; Aditi Satti, MD; Kartik Shenoy, MD; Robert M. Steiner, MD; Alex Swift, MD; Irene Swift, MD; Maria Elena Vega-Sanchez, MD
University of Alabama, Birmingham, AL: Mark Dransfield, MD; William Bailey, MD; Surya Bhatt, MD; Anand Iyer, MD; Hrudaya Nath, MD; J. Michael Wells, MD
University of California, San Diego, CA: Joe Ramsdell, MD; Paul Friedman, MD; Xavier Soler, MD, PhD; Andrew Yen, MD
University of Iowa, Iowa City, IA: Alejandro P. Comellas, MD; Karin F. Hoth, PhD; John Newell, Jr., MD; Brad Thompson, MD
University of Michigan, Ann Arbor, MI: MeiLan K. Han, MD, MS; Ella Kazerooni, MD; Carlos H. Martinez, MD, MPH
University of Minnesota, Minneapolis, MN: Joanne Billings, MD; Abbie Begnaud, MD; Tadashi Allen, MD
University of Pittsburgh, Pittsburgh, PA: Frank Sciurba, MD; Jessica Bon, MD; Divay Chandra, MD, MSc; Carl Fuhrman, MD; Joel Weissfeld, MD, MPH
University of Texas Health Science Center at San Antonio, San Antonio, TX: Antonio Anzueto, MD; Sandra Adams, MD; Diego Maselli-Caceres, MD; Mario E. Ruiz, MD

The **ECLIPSE** study (NCT00292552; GSK code SCO104960) was funded by GlaxoSmithKline.
ECLIPSE Investigators — *Bulgaria:* Y. Ivanov, Pleven; K. Kostov, Sofia. *Canada:* J. Bourbeau, Montreal; M. Fitzgerald, Vancouver, BC; P. Hernandez, Halifax, NS; K. Killian, Hamilton, ON; R. Levy, Vancouver, BC; F. Maltais, Montreal; D. O'Donnell, Kingston, ON. *Czech Republic:* J. Krepelka, Prague. *Denmark:* J. Vestbo, Hvidovre. *The Netherlands:* E. Wouters, Horn-Maastricht. *New Zealand:* D. Quinn, Wellington. *Norway:* P. Bakke, Bergen. *Slovenia:* M. Kosnik, Golnik. *Spain:* A. Agusti, J. Sauleda, P. de Mallorca. *Ukraine:* Y. Feschenko, V. Gavrisyuk, L. Yashina, Kiev; N. Monogarova, Donetsk. *United Kingdom:* P. Calverley, Liverpool; D. Lomas, Cambridge; W. MacNee, Edinburgh; D. Singh, Manchester; J. Wedzicha, London. *United States:* A. Anzueto, San Antonio, TX; S. Braman, Providence, RI; R. Casaburi, Torrance CA; B. Celli, Boston; G. Giessel, Richmond, VA; M. Gotfried, Phoenix, AZ; G. Greenwald, Rancho Mirage, CA; N. Hanania, Houston; D. Mahler, Lebanon, NH; B. Make, Denver; S. Rennard, Omaha, NE; C. Rochester, New Haven, CT; P. Scanlon, Rochester, MN; D. Schuller, Omaha, NE; F. Sciurba, Pittsburgh; A. Sharafkhaneh, Houston; T. Siler, St. Charles, MO; E. Silverman, Boston; A. Wanner, Miami; R. Wise, Baltimore; R. ZuWallack, Hartford, CT.
ECLIPSE Steering Committee: H. Coxson (Canada), C. Crim (GlaxoSmithKline, USA), L. Edwards (GlaxoSmithKline, USA), D. Lomas (UK), W. MacNee (UK), E. Silverman (USA), R. Tal Singer (Co-chair, GlaxoSmithKline, USA), J. Vestbo (Co-chair, Denmark), J. Yates (GlaxoSmithKline, USA).
ECLIPSE Scientific Committee: A. Agusti (Spain), P. Calverley (UK), B. Celli (USA), C. Crim (GlaxoSmithKline, USA), B. Miller (GlaxoSmithKline, USA), W. MacNee (Chair, UK), S. Rennard (USA), R. Tal-Singer (GlaxoSmithKline, USA), E. Wouters (The Netherlands), J. Yates (GlaxoSmithKline, USA).

This work was partially supported by the National Heart, Lung and Blood Institute's **Framingham Heart Study** (contract number N01-HC-25195) and its contract with Affymetrix, Inc for genotyping services (contract number N02-HL-6-4278). Also supported by NIH P01 AI050516. J. Vestbo is supported by the National Institute of Health Manchester Biomedical Research Centre (NIHR Manchester BRC).

KARE was funded by the Consortium for Large Scale Genome Wide Association Study III (2011E7300400), which was supported by the genotyping data (the Korean Genome Analysis Project, 4845-301) and the phenotype data (the Korean Genome Epidemiology Study, 4851-302). This was also supported by the National Research Foundation of Korea (2017R1A2B4003790), the National Project for Personalized Genomic Medicine (A111218-11-GM02), Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT & Future Planning (2013R1A1A1057961) and the Ministry of Education, Science and Technology (NRF-355-2011-1-E00060, NRF-2012R1A6A3A01039450).

The **Lung eQTL** study at Laval University was supported by the Chaire de pneumologie de la Fondation JD Bégin de l'Université Laval, the Fondation de l'Institut universitaire de cardiologie et de pneumologie de Québec, the Respiratory Health Network of the FRQS, the Canadian Institutes of Health Research (MOP - 123369), and the Cancer Research Society and Read for the Cure. Y.B. holds a Canada Research Chair in Genomics of Heart and Lung Diseases.

The **Norway GenKOLS** study (Genetics of Chronic Obstructive Lung Disease, GSK code RES11080) was funded by GSK.

The **ICGN** study was funded by GSK.

The **LifeLines** cohort study was supported by the Dutch Ministry of Health, Welfare and Sport, the Ministry of Economic Affairs, Agriculture and Innovation, the province of Groningen, the European Union (regional development fund), the Northern Netherlands Provinces (SNN), the Netherlands Organisation for Scientific Research (NWO), University Medical Center Groningen (UMCG), University of Groningen, de Nierstichting (the Dutch Kidney Foundation), and the Diabetes Fonds (the Diabetic Foundation).

The **Lovelace** cohort and analysis was primarily supported by National Cancer Institute grant R01 CA097356 (SAB). The State of New Mexico as a direct appropriation from the Tobacco Settlement Fund to SAB. through collaboration with University of New Mexico provided initial support to establish the LSC. Additional support was provided by NIH/NCI P30 CA118100 (SAB), HL68111 (Y.T.), and HL107873-01 (YT and SB).

MESA and the MESA SHARe project are conducted and supported by the National Heart, Lung, and Blood Institute (NHLBI) in collaboration with MESA investigators. Support for MESA is provided by contracts N01-HC-95159, N01-HC-95160, N01-HC-95161, N01-HC-95162, N01-HC-95163, N01-HC-95164, N01-HC-95165, N01-HC-95166, N01-HC-95167, N01-HC-95168, N01-HC-95169, UL1-TR-001079, UL1-TR-000040, UL1-TR-001420, UL1-TR-001881, and DK063491. MESA Family is conducted and supported by the National Heart, Lung, and Blood Institute (NHLBI) in collaboration with MESA investigators. Support is provided by grants and contracts R01HL071051, R01HL071205, R01HL071250, R01HL071251, R01HL071258, and R01HL071259 by the National Center for Research Resources, Grant UL1RR033176, and the National Center for Advancing Translational Sciences, Grant UL1TR000124. The MESA Lung study was supported by grants R01 HL077612, R01 HL093081 and RC1 HL100543 from the NHLBI. This publication was developed under a STAR research assistance agreement, No. RD831697 (MESA Air),

awarded by the U.S Environmental protection Agency. It has not been formally reviewed by the EPA. The views expressed in this document are solely those of the authors and the EPA does not endorse any products or commercial services mentioned in this publication. Funding for SHARe genotyping was provided by NHLBI Contract N02-HL-64278. Genotyping was performed at Affymetrix (Santa Clara, California, USA) and the Broad Institute of Harvard and MIT (Boston, Massachusetts, USA) using the Affymetrix Genome-Wide Human SNP Array 6.0.

The **National Emphysema Treatment Trial** was supported by the NHLBI N01HR76101, N01HR76102, N01HR76103, N01HR76104, N01HR76105, N01HR76106, N01HR76107, N01HR76108, N01HR76109, N01HR76110, N01HR76111, N01HR76112, N01HR76113, N01HR76114, N01HR76115, N01HR76116, N01HR76118 and N01HR76119, the Centers for Medicare and Medicaid Services and the Agency for Healthcare Research and Quality. The **Normative Aging Study** is supported by the Cooperative Studies Program/ERIC of the US Department of Veterans Affairs and is a component of the Massachusetts Veterans Epidemiology Research and Information Center (MAVERIC). D.S. is supported by a VA Research Career Scientist award.

The **Rotterdam Study** is funded by Erasmus Medical Center and Erasmus University, Rotterdam, Netherlands Organization for the Health Research and Development (ZonMw), the Research Institute for Diseases in the Elderly (RIDE), the Ministry of Education, Culture and Science, the Ministry for Health, Welfare and Sports, the European Commission (DG XII), and the Municipality of Rotterdam.
The generation and management of GWAS genotype data for the **Rotterdam Study** (RS I, RS II, RS III) was executed by the Human Genotyping Facility of the Genetic Laboratory of the Department of Internal Medicine, Erasmus MC, Rotterdam, The Netherlands. The GWAS datasets are supported by the Netherlands Organisation of Scientific Research NWO Investments (nr. 175.010.2005.011, 911-03-012), the Genetic Laboratory of the Department of Internal Medicine, Erasmus MC, the Research Institute for Diseases in the Elderly (014-93-015; RIDE2), the Netherlands Genomics Initiative (NGI)/Netherlands Organisation for Scientific Research (NWO) Netherlands Consortium for Healthy Aging (NCHA), project nr. 050-060-810. The generation and management of spirometric data was supported by FWO project G035014N.

The Subpopulations and Intermediate Outcomes in COPD Study (SPIROMICS) was supported by contracts from the NIH/NHLBI (HHSN268200900013C, HHSN268200900014C, HHSN268200900015C, HHSN268200900016C, HHSN268200900017C, HHSN268200900018C, HHSN268200900019C, HHSN268200900020C), which were supplemented by contributions made through the Foundation for the NIH from AstraZeneca; Bellerophon Therapeutics; Boehringer-Ingelheim Pharmaceuticals, Inc; Chiesi Farmaceutici SpA; Forest Research Institute, Inc; GSK; Grifols Therapeutics, Inc; Ikaria, Inc; Nycomed GmbH; Takeda Pharmaceutical Company; Novartis Pharmaceuticals Corporation; Regeneron Pharmaceuticals, Inc; and Sanofi.

We would like to thank Anjali Jacob for providing the count matrix for a single-cell RNA-Seq dataset and Kamil Slowikowski for his helpful feedback on data normalization for SNPsea.

The results published here are in part based upon data generated by the LungMAP Consortium (U01HL122642) and downloaded from (www.lungmap.net), on September 27, 2019. The LungMAP

consortium and the LungMAP Data Coordinating Center (1U01HL122638) are funded by the National Heart, Lung, and Blood Institute (NHLBI).

We gratefully acknowledge all the studies and databases that made GWAS summary data available in LD Hub: **ADIPOGen** (Adiponectin genetics consortium), **C4D** (Coronary Artery Disease Genetics Consortium), **CARDIoGRAM** (Coronary ARtery DIsease Genome wide Replication and Meta-analysis), **CKDGen** (Chronic Kidney Disease Genetics consortium), **dbGAP** (database of Genotypes and Phenotypes), **DIAGRAM** (DIAbetes Genetics Replication And Meta-analysis), **ENIGMA** (Enhancing Neuro Imaging Genetics through Meta Analysis), **EAGLE** (EArly Genetics & Lifecourse Epidemiology Eczema Consortium, excluding 23andMe), **EGG** (Early Growth Genetics Consortium), **GABRIEL** (A Multidisciplinary Study to Identify the Genetic and Environmental Causes of Asthma in the European Community), **GCAN** (Genetic Consortium for Anorexia Nervosa), **GEFOS** (GEnetic Factors for OSteoporosis Consortium), **GIANT** (Genetic Investigation of ANthropometric Traits), **GIS** (Genetics of Iron Status consortium), **GLGC** (Global Lipids Genetics Consortium), **GPC** (Genetics of Personality Consortium), **GUGC** (Global Urate and Gout consortium), **HaemGen** (haemotological and platelet traits genetics consortium), **HRgene** (Heart Rate consortium), **IIBDGC** (International Inflammatory Bowel Disease Genetics Consortium), **ILCCO** (International Lung Cancer Consortium), **IMSGC** (International Multiple Sclerosis Genetic Consortium), **MAGIC** (Meta-Analyses of Glucose and Insulin-related traits Consortium), **MESA** (Multi-Ethnic Study of Atherosclerosis), **PGC** (Psychiatric Genomics Consortium), **Project MinE** consortium, **ReproGen** (Reproductive Genetics Consortium), **SSGAC (**Social Science Genetics Association Consortium) and **TAG** (Tobacco and Genetics Consortium), **TRICL** (Transdisciplinary Research in Cancer of the Lung consortium), **UK Biobank**. We gratefully acknowledge the contributions of Alkes Price (the systemic lupus erythematosus GWAS and primary biliary cirrhosis GWAS) and Johannes Kettunen (lipids metabolites GWAS).

The **International COPD Genetics Consortium**: *Executive Committee*: James D. Crapo, William MacNee, David Lynch, H. Marike Boezen, Edwin K. Silverman, Jørgen Vestbo. *Members*: Alvar Agusti, Wayne Anderson, Nawar Bakerly, Per Bakke, Robert Bals, Kathleen C. Barnes, R Graham Barr, Terri H. Beaty, Eugene R. Bleecker, Yohan Bossé, Russell Bowler, Christopher Brightling, Marleen de Bruijne, Peter J. Castaldi, Bartolome Celli, Michael H. Cho, Harvey O. Coxson, Ron Crystal, Pim de Jong, Asger Dirksen, Jennifer Dy, Marilyn Foreman, Judith Garcia-Aymerich, Pierre Gevenois, Soumitra Ghosh, Hester Gietema, Amund Gulsvik, Ian P. Hall, Nadia Hansel, Craig P. Hersh, Brian D. Hobbs, Eric Hoffman, Noor Kalsheker, Hans-Ulrich Kauczor, Woo Jin Kim, Deog Kyeom Kim, Tarja Laitinen, Diether Lambrechts, Sang-Do Lee, Augusto A. Litonjua, David A. Lomas, Stephanie J. London, Daan W. Loth, Sharon M. Lutz, Merry-Lynn McDonald, Deborah A. Meyers, John D. Newell, Borge G. Nordestgaard, George T. O'Connor, Ma'en Obeidat, Yeon-Mok Oh, Peter D. Paré, Massimo Pistolesi, Dirkje S. Postma, Milo Puhan, Elizabeth Regan, Stephen S. Rich, Joon Beom Seo, Andrea Short, David Sparrow, Berend Stoel, David P. Strachan, Nicola Sverzellati, Ruth Tal-Singer, Gerben ter Riet, Yohannes Tesfaigzi, Martin D. Tobin, Edwin J.R. Van Beek, Bram van Ginneken, Claus F. Vogelmeier, Louise V. Wain, Adam Wanner, George Washko, Els Wauters, Emiel FM Wouters, Robert P. Young, and Loems Zeigler-Heitbrock. The ICGC extends special thanks to Nora Franceschini, Kari North, Steve Rich, Xin-Qun Wang, Andre Uitterlinden, Bruno Stricker, Arfan Ikram, Megan Hardin, Gus Litonjua, Nick Locantore, Josée Dupuis, Elizabeth Ampleford, Eugene Bleecker, Yeon- Mok Oh, Shuguang Leng, Bruce Psaty, Susan Heckbert, and Jerry Rotter.

This research has been conducted using the UK Biobank Resource under application number 20915 (M.H.C.) and 648 (M.D.T.)

References

1. The Atherosclerosis Risk in Communities (ARIC) Study: design and objectives. The ARIC investigators. *Am J Epidemiol* **129,** 687–702 (1989).

2. Loth, D. W. *et al.* Genome-wide association analysis identifies six new loci associated with forced vital capacity. *Nat Genet* **46,** 669–677 (2014).

3. Genomes Project, C. *et al.* An integrated map of genetic variation from 1,092 human genomes. *Nature* **491,** 56–65 (2012).

4. Howie, B., Fuchsberger, C., Stephens, M., Marchini, J. & Abecasis, G. R. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nat Genet* **44,** 955–959 (2012).

5. Fried, L. P. *et al.* The Cardiovascular Health Study: design and rationale. *Ann Epidemiol* **1,** 263–276 (1991).

6. Regan, E. A. *et al.* Genetic epidemiology of COPD (COPDGene) study design. *COPD* **7,** 32–43 (2010).

7. Cho, M. H. *et al.* Risk loci for chronic obstructive pulmonary disease: a genome-wide association study and meta-analysis. *Lancet Respir Med* **2,** 214–225 (2014).

8. Li, Y., Willer, C., Sanna, S. & Abecasis, G. Genotype imputation. *Annu Rev Genomics Hum Genet* **10,** 387–406 (2009).

9. Li, Y., Willer, C. J., Ding, J., Scheet, P. & Abecasis, G. R. MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genet Epidemiol* **34,** 816–834 (2010).

10. Purcell, S. & Chang, C. PLINK [version 1.9]. URL: https://www.cog-genomics.org/plink2. (2015).

11. Chang, C. C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4,** 7 (2015).

12. Vestbo, J. *et al.* Evaluation of COPD Longitudinally to Identify Predictive Surrogate End-points (ECLIPSE). *Eur Respir J* **31,** 869–873 (2008).

13. Hao, K. *et al.* Lung eQTLs to help reveal the molecular underpinnings of asthma. *PLoS Genet.* **8,** e1003029 (2012).

14. Das, S. *et al.* Next-generation genotype imputation service and methods. *Nat. Genet.* **48,** 1284–1287 (2016).

15. Hancock, D. B. *et al.* Meta-analyses of genome-wide association studies identify multiple loci associated with pulmonary function. *Nat Genet* **42,** 45–52 (2010).

16. Wilk, J. B. *et al.* Genome-wide association studies identify CHRNA5/3 and HTR4 in the development of airflow obstruction. *Am J Respir Crit Care Med* **186,** 622–632 (2012).

17. Cho, Y. S. *et al.* A large-scale genome-wide association study of Asian populations uncovers genetic factors influencing eight quantitative traits. *Nat Genet* **41,** 527–534 (2009).

18. Kim, W. J. *et al.* Genome-wide association studies identify locus on 6p21 influencing lung function in the Korean population. *Respirology* **19,** 360–368 (2014).

19. Purcell, S. PLINK [version 1.07]. URL: http://pngu.mgh.harvard.edu/purcell/plink/. **http://png,**

20. de Jong, K. *et al.* NOS1: a susceptibility gene for reduced level of FEV1 in the setting of pesticide exposure. *Am J Respir Crit Care Med* **190,** 1188–1190 (2014).

21. de Jong, K. *et al.* Genome-wide interaction study of gene-by-occupational exposure and effects on FEV1 levels. *J Allergy Clin Immunol* **136,** 1664–1672 e14 (2015).

22. Scholtens, S. *et al.* Cohort Profile: LifeLines, a three-generation cohort study and biobank. *Int J Epidemiol* **44,** 1172–1180 (2015).

23. Delaneau, O., Zagury, J. F. & Marchini, J. Improved whole-chromosome phasing for disease and population genetic studies. *Nat Methods* **10,** 5–6 (2013).

24. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* **81,** 559–575 (2007).

25. Sood, A. *et al.* Difference in airflow obstruction between Hispanic and non-Hispanic White female smokers. *COPD* **5,** 274–281 (2008).

26. Bild, D. E. *et al.* Multi-Ethnic Study of Atherosclerosis: objectives and design. *Am J Epidemiol* **156,** 871–881 (2002).

27. Kaufman, J. D. *et al.* Prospective study of particulate air pollution exposures, subclinical atherosclerosis, and clinical cardiovascular disease: The Multi-Ethnic Study of Atherosclerosis and Air Pollution (MESA Air). *Am J Epidemiol* **176,** 825–837 (2012).

28. Hankinson, J. L. *et al.* Performance of American Thoracic Society-recommended spirometry reference values in a multiethnic sample of adults: the multi-ethnic study of atherosclerosis (MESA) lung study. *Chest* **137,** 138–145 (2010).

29. Manichaikul, A. *et al.* Association of SCARB1 variants with subclinical atherosclerosis and incident cardiovascular disease: the multi-ethnic study of atherosclerosis. *Arter. Thromb Vasc Biol* **32,** 1991–1999 (2012).

30. Marchini, J. & Howie, B. Genotype imputation for genome-wide association studies. *Nat Rev Genet* **11,** 499–511 (2010).

31. Fishman, A. *et al.* A randomized trial comparing lung-volume-reduction surgery with medical therapy for severe emphysema. *N Engl J Med* **348,** 2059–2073 (2003).

32. Bell, B., Rose, C. L. & Damon, D. The Normative Aging Study: an interdisciplinary and longitudinal study of health and aging. *Aging Hum Dev* **3,** 5–17 (1972).

33. Cho, M. H. *et al.* Variants in FAM13A are associated with chronic obstructive pulmonary disease. *Nat Genet* **42,** 200–202 (2010).

34. Cho, M. H. *et al.* A genome-wide association study of COPD identifies a susceptibility locus on chromosome 19q13. *Hum Mol Genet* **21,** 947–957 (2012).

35. Sorheim, I. C. *et al.* Case-control studies on risk factors for chronic obstructive pulmonary disease: how does the sampling of the cases and controls affect the results? *Clin Respir J* **4,** 89–96

(2010).

36. Hofman, A. *et al.* The Rotterdam Study: 2016 objectives and design update. *Eur J Epidemiol* **30,** 661–708 (2015).

37. Kreiner-Moller, E., Medina-Gomez, C., Uitterlinden, A. G., Rivadeneira, F. & Estrada, K. Improving accuracy of rare variant imputation with a two-step imputation approach. *Eur J Hum Genet* **23,** 395–400 (2015).

38. Aulchenko, Y. S., Struchalin, M. V & van Duijn, C. M. ProbABEL package for genome-wide association analysis of imputed data. *BMC Bioinformatics* **11,** 134 (2010).

39. Couper, D. *et al.* Design of the Subpopulations and Intermediate Outcomes in COPD Study (SPIROMICS). *Thorax* **69,** 491–494 (2014).

40. Hobbs, B. D. *et al.* Exome Array Analysis Identifies a Common Variant in IL27 Associated with Chronic Obstructive Pulmonary Disease. *Am J Respir Crit Care Med* **194,** 48–57 (2016).

41. Silverman, E. K. *et al.* Gender-related differences in severe, early-onset chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* **162,** 2152–2158 (2000).

42. Zhu, G. *et al.* The SERPINE2 gene is associated with chronic obstructive pulmonary disease in two large populations. *Am. J. Respir. Crit. Care Med.* **176,** 167–73 (2007).

43. Patel, B. D. *et al.* Airway wall thickening and emphysema show independent familial aggregation in chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* **178,** 500–505 (2008).

44. Zhou, X. *et al.* Identification of a chronic obstructive pulmonary disease genetic determinant that regulates HHIP. *Hum. Mol. Genet.* **21,** 1325–35 (2012).

45. Busch, R. *et al.* Genetic Association and Risk Scores in a Chronic Obstructive Pulmonary Disease Meta-analysis of 16,707 Subjects. *Am. J. Respir. Cell Mol. Biol.* **57,** 35–46 (2017).

46. R Core Team. R: A Language and Environment for Statistical Computing. (2018).

47. Sudlow, C. *et al.* UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med.* **12,** e1001779 (2015).

48. Shrine, N. *et al.* New genetic signals for lung function highlight pathways and pleiotropy, and chronic obstructive pulmonary disease associations across multiple ancestries. *bioRxiv* (2018).

49. Miller, M. R. *et al.* Standardisation of spirometry. *Eur. Respir. J.* **26,** 319–38 (2005).

50. Bycroft, C. *et al.* The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562,** 203–209 (2018).

51. Vogelmeier, C. F. *et al.* Global Strategy for the Diagnosis, Management, and Prevention of Chronic Obstructive Lung Disease 2017 Report. GOLD Executive Summary. *Am. J. Respir. Crit. Care Med.* **195,** 557–582 (2017).

52. Winkler, T. W. *et al.* Approaches to detect genetic effects that differ between two strata in genome-wide meta-analyses: Recommendations based on a systematic evaluation. *PLoS One* **12,** e0181038 (2017).

53. Randall, J. C. *et al.* Sex-stratified genome-wide association studies including 270,000 individuals

show sexual dimorphism in genetic loci for anthropometric traits. *PLoS Genet.* **9,** e1003500 (2013).

54. Peyrot, W. J., Boomsma, D. I., Penninx, B. W. J. H. & Wray, N. R. Disease and Polygenic Architecture: Avoid Trio Design and Appropriately Account for Unscreened Control Subjects for Common Disease. *Am. J. Hum. Genet.* **98,** 382–91 (2016).

55. Finucane, H. K. *et al.* Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. *Nat. Genet.* **50,** 621–629 (2018).

56. Xu, Y. *et al.* Single-cell RNA sequencing identifies diverse roles of epithelial cells in idiopathic pulmonary fibrosis. *JCI insight* **1,** e90558 (2016).

57. Jacob, A. *et al.* Differentiation of Human Pluripotent Stem Cells into Functional Lung Alveolar Epithelial Cells. *Cell Stem Cell* **21,** 472–488.e10 (2017).

58. Ardini-Poleske, M. E. *et al.* LungMAP: The Molecular Atlas of Lung Development Program. *Am. J. Physiol. Lung Cell. Mol. Physiol.* **313,** L733–L740 (2017).

59. Durinck, S., Spellman, P. T., Birney, E. & Huber, W. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nat. Protoc.* **4,** 1184–91 (2009).

60. Slowikowski, K., Hu, X. & Raychaudhuri, S. SNPsea: an algorithm to identify cell types, tissues and pathways affected by risk loci. *Bioinformatics* **30,** 2496–7 (2014).

61. Hu, X. *et al.* Integrating autoimmune risk loci with gene-expression data identifies specific pathogenic immune cell subsets. *Am. J. Hum. Genet.* **89,** 496–506 (2011).

62. Ward, L. D. & Kellis, M. HaploReg v4: systematic mining of putative causal variants, cell types, regulators and target genes for human complex traits and disease. *Nucleic Acids Res* (2015). doi:10.1093/nar/gkv1340

63. Dayem Ullah, A. Z. *et al.* SNPnexus: assessing the functional relevance of genetic variation to facilitate the promise of precision medicine. *Nucleic Acids Res.* **46,** W109–W113 (2018).

64. Lee, S., Fuchsberger, C., Kim, S. & Scott, L. An efficient resampling method for calibrating single and gene-based rare variant association analysis in case-control studies. *Biostatistics* **17,** 1–15 (2016).

65. Qiao, D. *et al.* Whole exome sequencing analysis in severe chronic obstructive pulmonary disease. *Hum. Mol. Genet.* (2018). doi:10.1093/hmg/ddy269

66. Pers, T. H. *et al.* Biological interpretation of genome-wide association studies using predicted gene functions. *Nat. Commun.* **6,** 5890 (2015).

67. Morrow, J. D. *et al.* Human Lung DNA Methylation Quantitative Trait Loci Colocalize with COPD Genome-wide Association Loci. *Am. J. Respir. Crit. Care Med.* (2018). doi:10.1164/rccm.201707-1434OC

68. Morrow, J. D. *et al.* DNA methylation profiling in human lung tissue identifies genes associated with COPD. *Epigenetics* 1–10 (2016). doi:10.1080/15592294.2016.1226451

69. Shabalin, A. A. Matrix eQTL: ultra fast eQTL analysis via large matrix operations. *Bioinformatics* **28,** 1353–8 (2012).

70. Hormozdiari, F. *et al.* Colocalization of GWAS and eQTL Signals Detects Target Genes. *Am. J. Hum. Genet.* **99,** 1245–1260 (2016).

71. So, H.-C. *et al.* Analysis of genome-wide association data highlights candidates for drug repositioning in psychiatry. *Nat. Neurosci.* **20,** 1342–1349 (2017).

72. Barbeira, A. N. *et al.* Exploring the phenotypic consequences of tissue specific gene expression variation inferred from GWAS summary statistics. *Nat. Commun.* **9,** 1825 (2018).

73. Lamontagne, M. *et al.* Leveraging lung tissue transcriptome to uncover candidate causal genes in COPD genetic associations. *Hum. Mol. Genet.* **27,** 1819–1829 (2018).

74. Subramanian, A. *et al.* A Next Generation Connectivity Map: L1000 Platform and the First 1,000,000 Profiles. *Cell* **171,** 1437–1452.e17 (2017).

75. Musa, A. *et al.* A review of connectivity map and computational approaches in pharmacogenomics. *Brief. Bioinform.* **19,** 506–523 (2018).

76. Caliński, T. & Harabasz, J. A dendrite method for cluster analysis. *Commun. Stat.* **3,** 1–27 (1974).

77. Dimas, A. S. *et al.* Impact of type 2 diabetes susceptibility variants on quantitative glycemic traits reveals mechanistic heterogeneity. *Diabetes* **63,** 2158–71 (2014).

78. Friedman, J., Hastie, T. & Tibshirani, R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J. Stat. Softw.* **33,** 1–22 (2010).

79. Zheng, J. *et al.* LD Hub: a centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. *Bioinformatics* **33,** 272–279 (2017).

80. Bulik-Sullivan, B. *et al.* An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* **47,** 1236–41 (2015).

81. Fagerberg, L. *et al.* Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics. *Mol. Cell. Proteomics* **13,** 397–406 (2014).

82. Yoshimura, S., Gerondopoulos, A., Linford, A., Rigden, D. J. & Barr, F. A. Family-wide characterization of the DENN domain Rab GDP-GTP exchange factors. *J. Cell Biol.* **191,** 367–81 (2010).

83. O'Leary, N. A. *et al.* Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* **44,** D733-45 (2016).

84. Yoon, J.-H., Her, S., Kim, M., Jang, I.-S. & Park, J. The expression of damage-regulated autophagy modulator 2 (DRAM2) contributes to autophagy induction. *Mol. Biol. Rep.* **39,** 1087–93 (2012).

85. Salo, P. P. *et al.* Genetic Variants on Chromosome 1p13.3 Are Associated with Non-ST Elevation Myocardial Infarction and the Expression of DRAM2 in the Finnish Population. *PLoS One* **10,** e0140576 (2015).

86. Sergouniotis, P. I. *et al.* Disease Expression in Autosomal Recessive Retinal Dystrophy Associated With Mutations in the DRAM2 Gene. *Invest. Ophthalmol. Vis. Sci.* **56,** 8083–90 (2015).

87. Saito, A., Ozaki, K., Fujiwara, T., Nakamura, Y. & Tanigami, A. Isolation and mapping of a human lung-specific gene, TSA1902, encoding a novel chitinase family member. *Gene* **239,** 325–31

(1999).

88. Aminuddin, F. *et al.* Genetic association between human chitinases and lung function in COPD. *Hum. Genet.* **131,** 1105–14 (2012).

89. Birben, E. *et al.* The effects of an insertion in the 5'UTR of the AMCase on gene expression and pulmonary functions. *Respir. Med.* **105,** 1160–9 (2011).

90. Chatterjee, R., Batra, J., Das, S., Sharma, S. K. & Ghosh, B. Genetic association of acidic mammalian chitinase with atopic asthma and serum total IgE levels. *J. Allergy Clin. Immunol.* **122,** 202–8, 208.e1–7 (2008).

91. Ober, C. & Chupp, G. L. The chitinase and chitinase-like proteins: a review of genetic and functional studies in asthma and immune-mediated diseases. *Curr. Opin. Allergy Clin. Immunol.* **9,** 401–8 (2009).

92. Heinzmann, A. *et al.* Joint influences of Acidic-Mammalian-Chitinase with Interleukin-4 and Toll-like receptor-10 with Interleukin-13 in the genetics of asthma. *Pediatr. Allergy Immunol.* **21,** e679-86 (2010).

93. Okawa, K. *et al.* Loss and Gain of Human Acidic Mammalian Chitinase Activity by Nonsynonymous SNPs. *Mol. Biol. Evol.* **33,** 3183–3193 (2016).

94. Wu, A. C., Lasky-Su, J., Rogers, C. A., Klanderman, B. J. & Litonjua, A. Polymorphisms of chitinases are not associated with asthma. *J. Allergy Clin. Immunol.* **125,** 754–7, 757.e1–757.e2 (2010).

95. Hartl, D. *et al.* Acidic mammalian chitinase regulates epithelial cell apoptosis via a chitinolytic-independent mechanism. *J. Immunol.* **182,** 5098–106 (2009).

96. Zhu, Z. *et al.* Acidic mammalian chitinase in asthmatic Th2 inflammation and IL-13 pathway activation. *Science* **304,** 1678–82 (2004).

97. Hartl, D. *et al.* Acidic mammalian chitinase is secreted via an ADAM17/epidermal growth factor receptor-dependent pathway and stimulates chemokine production by pulmonary epithelial cells. *J. Biol. Chem.* **283,** 33472–82 (2008).

98. Li, R. J. *et al.* Down-regulation of mitochondrial ATPase by hypermethylation mechanism in chronic myeloid leukemia is associated with multidrug resistance. *Ann. Oncol. Off. J. Eur. Soc. Med. Oncol.* **21,** 1506–14 (2010).

99. Chauvet, S. *et al.* EG-VEGF, BV8, and their receptor expression in human bronchi and their modification in cystic fibrosis: Impact of CFTR mutation (delF508). *Am. J. Physiol. Lung Cell. Mol. Physiol.* **309,** L314-22 (2015).

100. Guo, L. *et al.* Subacute hypoxia suppresses Kv3.4 channel expression and whole-cell K+ currents through endogenous 15-hydroxyeicosatetraenoic acid in pulmonary arterial smooth muscle cells. *Eur. J. Pharmacol.* **587,** 187–95 (2008).

101. Alhouayek, M., Masquelier, J., Cani, P. D., Lambert, D. M. & Muccioli, G. G. Implication of the anti-inflammatory bioactive lipid prostaglandin D2-glycerol ester in the control of macrophage activation and inflammation by ABHD6. *Proc. Natl. Acad. Sci. U. S. A.* **110,** 17558–63 (2013).

102. Reiling, J. H. *et al.* A CREB3-ARF4 signalling pathway mediates the response to Golgi stress and susceptibility to pathogens. *Nat. Cell Biol.* **15,** 1473–85 (2013).

103. Yang, S. *et al.* Tumor necrosis factor receptor 2 (TNFR2)·interleukin-17 receptor D (IL-17RD) heteromerization reveals a novel mechanism for NF-κB activation. *J. Biol. Chem.* **290,** 861–71 (2015).

104. Kamio, K. *et al.* XPLN is modulated by HDAC inhibitors and negatively regulates SPARC expression by targeting mTORC2 in human lung fibroblasts. *Pulm. Pharmacol. Ther.* **44,** 61–69 (2017).

105. Serbanovic-Canic, J. *et al.* Silencing of RhoA nucleotide exchange factor, ARHGEF3, reveals its unexpected role in iron uptake. *Blood* **118,** 4967–76 (2011).

106. Pilewski, J. M., Latoche, J. D., Arcasoy, S. M. & Albelda, S. M. Expression of integrin cell adhesion receptors during human airway epithelial repair in vivo. *Am. J. Physiol.* **273,** L256-63 (1997).

107. Markovics, J. A. *et al.* Interleukin-1beta induces increased transcriptional activation of the transforming growth factor-beta-activating integrin subunit beta8 through altering chromatin architecture. *J. Biol. Chem.* **286,** 36864–74 (2011).

108. Kitamura, H. *et al.* Mouse and human lung fibroblasts regulate dendritic cell trafficking, airway inflammation, and fibrosis through integrin αvβ8-mediated activation of TGF-β. *J. Clin. Invest.* **121,** 2863–75 (2011).

109. Araya, J. *et al.* Squamous metaplasia amplifies pathologic epithelial-mesenchymal interactions in COPD patients. *J. Clin. Invest.* **117,** 3551–62 (2007).

110. Mu, D. *et al.* The integrin alpha(v)beta8 mediates epithelial homeostasis through MT1-MMP-dependent activation of TGF-beta1. *J. Cell Biol.* **157,** 493–507 (2002).

111. Fjellbirkeland, L. *et al.* Integrin alphavbeta8-mediated activation of transforming growth factor-beta inhibits human airway epithelial proliferation in intact bronchial tissue. *Am. J. Pathol.* **163,** 533–42 (2003).

112. Markovics, J. A. *et al.* Transcription of the transforming growth factor beta activating integrin beta8 subunit is regulated by SP3, AP-1, and the p38 pathway. *J. Biol. Chem.* **285,** 24695–706 (2010).

113. Amberger, J. S., Bocchini, C. A., Schiettecatte, F., Scott, A. F. & Hamosh, A. OMIM.org: Online Mendelian Inheritance in Man (OMIM®), an online catalog of human genes and genetic disorders. *Nucleic Acids Res.* **43,** D789-98 (2015).

114. Liu, W. *et al.* TMEM196 acts as a novel functional tumour suppressor inactivated by DNA methylation and is a potential prognostic biomarker in lung cancer. *Oncotarget* **6,** 21225–39 (2015).

115. Nawa, M. *et al.* Reduced expression of BTBD10, an Akt activator, leads to motor neuron death. *Cell Death Differ.* **19,** 1398–407 (2012).

116. Furuta, N. *et al.* Reduced expression of BTBD10 in anterior horn cells with Golgi fragmentation and pTDP-43-positive inclusions in patients with sporadic amyotrophic lateral sclerosis. *Neuropathology* **33,** 397–404 (2013).

117. Hall, N. G., Klenotic, P., Anand-Apte, B. & Apte, S. S. ADAMTSL-3/punctin-2, a novel glycoprotein in extracellular matrix related to the ADAMTS family of metalloproteases. *Matrix Biol.* **22,** 501–10 (2003).

118. Dow, D. J. *et al.* ADAMTSL3 as a candidate gene for schizophrenia: gene sequencing and ultra-high density association analysis by imputation. *Schizophr. Res.* **127,** 28–34 (2011).

119. Xu, H. *et al.* Nonmosaic tetrasomy 15q25.2 → qter identified with SNP microarray in a patient with characteristic facial appearance and review of the literature. *Eur. J. Med. Genet.* **57,** 329–33 (2014).

120. Vanhoutteghem, A. & Djian, P. Basonuclins 1 and 2, whose genes share a common origin, are proteins with widely different properties and functions. *Proc. Natl. Acad. Sci. U. S. A.* **103,** 12423–8 (2006).

121. Feuerborn, A., Mathow, D., Srivastava, P. K., Gretz, N. & Gröne, H.-J. Basonuclin-1 modulates epithelial plasticity and TGF-β1-induced loss of epithelial cell integrity. *Oncogene* **34,** 1185–95 (2015).

122. Boldrup, L., Coates, P. J., Laurell, G. & Nylander, K. p63 Transcriptionally regulates BNC1, a Pol I and Pol II transcription factor that regulates ribosomal biogenesis and epithelial differentiation. *Eur. J. Cancer* **48,** 1401–6 (2012).

123. Tseng, H. Basonuclin, a zinc finger protein associated with epithelial expansion and proliferation. *Front. Biosci.* **3,** D985-8 (1998).

124. Vaquerizas, J. M., Kummerfeld, S. K., Teichmann, S. A. & Luscombe, N. M. A census of human transcription factors: function, expression and evolution. *Nat. Rev. Genet.* **10,** 252–63 (2009).

125. Joshi, P. *et al.* The functional interactome landscape of the human histone deacetylase family. *Mol. Syst. Biol.* **9,** 672 (2013).

126. Pisani, D. F. *et al.* Involvement of BTBD1 in mesenchymal differentiation. *Exp. Cell Res.* **313,** 2417–26 (2007).

127. Pisani, D. F., Cabane, C., Derijard, B. & Dechesne, C. A. The topoisomerase 1-interacting protein BTBD1 is essential for muscle cell differentiation. *Cell Death Differ.* **11,** 1157–65 (2004).