The genetic basis of endometriosis and comorbidity with other pain and inflammatory conditions

Supplementary Information

Supplementary Text

Results

Heterogeneity of signals at genome-wide significant loci and sensitivity analyses

Of the 42 genome-wide significant loci, we observed nominally significant between-study heterogeneity for 11 loci (p-overall heterogeneity <0.05, 2.1 loci expected) in the overall endometriosis meta-analysis, six of which were explained by case ascertainment and two by ancestry. For lead SNPs at these 11 loci, we used meta-regression to partition overall heterogeneity into components due to ancestry, case ascertainment and residual (reflecting other differences in study design). Heterogeneity at six loci was explained by case ascertainment (surgical/medical records vs. mixed self-report and medically confirmed vs. self-reported only) and at two loci by ancestry (European vs. East Asian) (Supplementary Tables 4 and 5; Supplementary Figs. 3 and 4). For the remaining three loci, the heterogeneity could not be explained by ancestry or case ascertainment. Between-study heterogeneity also impacted the two previously reported loci [1] that did not attain genome-wide significance in the overall endometriosis meta-analysis: one mapping to *IL1A*/2p13 (rs3783521, $p=6.38x10^{-8}$) and one in *FN1*/2q35 (rs1250242, $p=3.25x10^{-4}$) (Supplementary Table 6, Supplementary Fig. 5). For these two loci, heterogeneity was explained by case ascertainment (Supplementary Table 4); the effect sizes of both loci were largest for stage III/IV disease, consistent with Sapkota et al (2017)[1].

Sensitivity analyses were conducted to test the impact of small datasets (9/24 with <300 cases) and the inclusion of male controls in sex-combined autosomal analysis (4/24 datasets; Supplementary Table 1). Meta-analysis restricted to 15 studies with >300 cases (59,101 cases/686,278 controls) retained genome-wide significant associations for 40 of 42 loci, with the remaining two signals still showing similar effect sizes and borderline significant associations (Supplementary Table 7, Extended Data Fig. 1e). We also observed no significant heterogeneity ($p>1.19x10^{-3}$) in effect sizes for the 42 lead SNPs between studies including female only controls vs. sex-combined controls (Extended Data Fig. 3).

Enrichment of gene expression across tissues and pathways

We sought to characterise the regulatory role of associated non-coding variants across tissues and pathways. First, all genes (n=415) within ±200kb of each of the 49 index SNPs were tested for enrichment in expression across tissues available in Human Protein Atlas (HPA) and GTEx RNA-seq datasets (See Methods). Across 35 human tissues available in the HPA dataset, endometrium and smooth muscle were most highly enriched for expression of genes signposted by endometriosis risk loci (Extended Data Fig. 4a and b). Similarly, across 29 tissues in the GTEx dataset, uterus (endometrium and myometrium combined) was most highly enriched (Extended Data Fig. 4c and d). These enrichment analyses based on genes in close proximity to endometriosis risk variants highlighted tissues most relevant to endometriosis.

Genetic analyses of endometriosis vs. adenomyosis

We tested for differences in effect sizes between endometriosis and adenomyosis (the growth of endometrium into the myometrium) for each of the 42 endometriosis lead SNPs using two independent GWAS analyses in the UKBB (see Supplementary Text: Methods): 1) 1,764 adenomyosis cases (without evidence of endometriosis) and 2) 2,729 endometriosis cases (without evidence of adenomyosis) vs. independent control sets. No significant differences in effect sizes were observed ($p<1.19x10^{-3}$, Supplementary Table 22). We then compared the effects of the 42 SNPs between adenomyosis (3,531 cases from UKBB and FinnGen; see Methods, Supplementary Table 22) and endometriosis rASRM and surgical subtypes. The strongest correlation was observed between adenomyosis and rASRM stage I/II (r=0.52), followed by superficial lesions (r=0.40) and deep lesions (r=0.31) (Extended Data Fig. 8). Low correlations were observed with rASRM stage III/IV (r=0.02) and endometrioma (r^2=0.18). Of the 42 endometriosis lead SNPs, 6 (rs71575922 in *SYNE1*/6q25.1, rs17053711 in *KCTD9*/8p21.2, rs10090060 in *GDAP1*/8q21.11, rs507666 in *ABO*/9q34.2, rs3858429 in *FSHB*/11p14.1, rs13441059 in *TEX11*/Xq13.1) were significantly associated with adenomyosis ($p<1.19x10^{-3}$; Supplementary Table 21, Extended Data Fig. 8). These results suggest shared genetic susceptibility between adenomyosis and stage I/II and superficial peritoneal endometriosis, or associated symptoms.

Genetic regulation of expression and methylation at endometriosis risk loci

To identify specific genes regulated by the 49 distinct endometriosis association signals, we analysed four expression quantitative trait loci (eQTL) datasets: 1) a gene expression microarray study of 229 endometrial samples [2]; 2) a novel meta-analysis of RNAseq-based eQTL datasets including 368 endometrial samples ([3] and unpublished data; see Methods); 3) RNAseq expression data from 129 uterus tissue samples from GTEx[4]; and 4) data from 31,684 blood samples from the eQTLGen Consortium [5].

Summary data-based Mendelian Randomisation (SMR)[6] was used to identify genes whose expression levels are associated with endometriosis due to the effects of a common genetic variant (either by direct causal or pleiotropic effects) rather than due to linkage disequilibrium (LD). Table 1 summarises the significant eQTL/mQTL SMR results across endometriosis GWAS loci, together with evidence from chromatin interactions.

Analysis using the smallest and most heterogeneous 'uterus tissue' eQTL data from GTEx identified only a single gene (*HCG4P7* in *ID4*/6p22.3 locus) but the results did not pass the HEIDI test of heterogeneity (Supplementary Table 14). Using the endometrium datasets, we identified significant associations between endometriosis risk variants and expression of five genes (*SRP14 (SRP14-AS1/15q15.1), LNC-LBCS (ID4/6p22.3), TRA2A (7p15.2), VEZT (VEZT/12q22), HOXB9 (SKAP1/17q21.32))*, with a sixth (*LINC00339*) not passing the heterogeneity test (Table 1, Supplementary Table 15). In the blood SMR analysis, seven genes were significantly associated with endometriosis risk passing both SMR and HEIDI tests (Table 1, Supplementary Table 16).

We also associated SNPs in the endometriosis risk regions with DNA methylation of nearby CpG sites in endometrium and blood using previously published mQTL datasets (Supplementary Table 17).[7,8] In endometrium and blood, numerous methylation probes related to *GREB1* passed the SMR and HEIDI tests for association with risk variants, consistent with previous reports (Mortlock et al. 2019). In blood, an additional 18 probes passed the SMR and HEIDI tests (Supplementary Table 18) close to 8 gene (clusters) of interest

including *ARL14EP/FSHB* (*FSHB*/11p14.1)*, BMF/SRP14-AS1* (*SRP14-AS1*/15q15.1)*, ESR1/SYNE1* (*SYNE1*/6q25.1), *GDAP1* (*GDAP1*/8q21.11), *MLLT10 (MLLT10*/10p12.31)*, WNT4* (*WNT4*/1p36.12), *CD109* (*CD109*/6q13), and *HOXC-AS2* (*HOXC10*/12p13.13) (Table 1), with 10 regulated by SNPs in the 99% credible set.

Results showing that endometriosis risk variants on chromosomes 2p25.1 and 12q22 may function through changes in expression of *GREB1*, and *VEZT* and/or *FGD6*, respectively, have been reported previously [2,3,9]. Many of the other potentially causal genes are novel and have strong biological support. Notable was the signal for *GDAP1* (*GDAP1*/8q21.11), previously associated with dysmenorrhea severity and neuronal development (Fig. 2) [10,11]. In the *GDAP1*/8q21.11 region, the rs4567029 variant, which regulates methylation of probes near *GDAP1,* is in perfect LD with rs10283076 that was identified as the variant regulating *GDAP1* expression in blood tissue (Fig. 2). The *SRP14-AS1*/15q15.1 locus harboured multiple distinct association signals in endometrium and blood, and had a chromatin interaction at this locus with *BMF* (Bcl2 modifying factor) (Fig. 3; Supplementary Table 15,16 and 18). *BMF* encodes for a glycoprotein associated with sex hormone binding globulin and regulating bioavailability of oestrogen and testosterone[12]. *SRP14* is a constituent of the signal recognition particle (SRP) with functions including targeting secretory proteins to the rough endoplasmic reticulum membrane. *SRP14* variants may affect endometriosis-associated pain genesis and maintenance. Association of variants at this locus with DHEA-sulfate (DHEA-S) levels has been reported [13]. DHEA-S is a neurosteroid functioning as a neurotropin, that can bind and activate nerve growth factor (*NGF*) and brain-derived neurotrophic factor (*BDNF*) [14,15]. *BDNF* has been shown to regulate the maintenance of chronic pain in various chronic disorders [16], and its expression appears increased in the eutopic endometrium of women with endometriosis compared to controls [17,18]. The expression of *NGF,* one of our other GWAS loci, has been suggested to partly mediate local nerve density around endometriosis lesions, associated with dyspareunia [19].

SMR analysis in endometrium also showed evidence that the endometriosis risk variant at *SKAP1*/17q21.32 likely impacts the expression of *HOXB9,* further supported by chromatin interactions of the endometriosis risk variant with this gene (Extended Data Fig. 5). *HOXB9* is a proangiogenic transcription factor [20], upregulated in

various cancers including breast and endometrial cancers [21,22]. And additional gene of interest from blood eQTL analyses was *ABO* (Table 1). *ABO* encodes for histo-blood group antigens. Lead variant rs507666, intronic to *ABO,* has been associated with a wide range of other traits and conditions [23] but in particular with soluble Intercellular Adhesion Molecule 1 (sICAM-1) concentrations in women, suggesting a potential regulatory role of inflammatory adhesion processes[24]. We tested the association between sICAM-1 levels in serum from 136 endometriosis cases (85 rASRM stage I/II, 51 rASRM stage III/IV) and 54 endometriosis-free controls (see Supplementary Note: Methods). We did not find a significant association with disease status or stage (mean log-sICAM1 in cases=2.40, mean log-sICAM1 in controls=2.44, p=0.16; Extended Data Fig. 6, Supplementary Table 19), however, further investigation in a larger sample size is warranted.

Genome-wide significant loci shared between endometriosis and 11 traits/conditions

We investigated if endometriosis-associated variants in the 99% credible sets for each distinct endometriosis signal were associated with any of the genetically correlated traits and conditions at genome-wide significance, using Phenoscanner and GWAS Catalog. This revealed 10 genome-wide significant variants shared with 11 different traits and conditions (Supplementary Table 29). Three variants were shared with pain traits: two with multi-site chronic pain (rs1352889 at *BSN*/3p21.31 and rs10828249 at *MLLT10*/10p12.31), and one with migraine and dysmenorrhea (rs12030576 at *NGF*/1p13.2). Loci shared with other reproductive traits and conditions included three with uterine fibroids (rs10917151 at *WNT4*/1p36.12, rs2510770 at *PDLIM5*/4q22.3, rs71575922 at *SYNE1*/6q25.1); two with menstrual cycle length (rs17053711 at *KCTD9*/8p21.2, rs3858429 at *FSHB*/11p14.1); one with age at menarche (rs1352889 at *BSN*/3p21.31) and one with age at menopause (rs3858429 at *FSHB*/11p14.1). Loci shared with metabolic traits included two with BMI (rs1352889 at *BSN*/3p21.31; rs108282249 at *MLLT10*/10p12.31), and one with type 2 diabetes (rs507666 at *ABO*/9q34.2). One genome-wide significant locus, rs2967684 at *ACTL9* at 19p13.2 was shared with asthma.

Supplementary Methods

sICAM-1 analysis

Soluble intercellular Adhesion Molecule-1 (sICAM-1) levels were measured in serum samples of 190 women from ENDOX study at Nuffield Department of Women's & Reproductive Health, John Radcliffe Hospital, University of Oxford. Of these 190 women, 136 were laparoscopically diagnosed endometriosis case (85 rASRM stage I/II, 51 rASRM stage III/IV) and 54 were laparoscopically confirmed endometriosis free female controls. Participants were not on any hormones for the past 3 months and had no previous endometriosis diagnosis. s-ICAM1 levels were measured using Quantikine ELISA kit (R&D systems) that is human sICAM-1/CD54 allele-specific in serum samples located prior to their laparoscopic surgery. Each serum sample was analysed in duplicates and samples were randomised across 4 plates of 96 wells. The mean of the duplicate samples was used as the raw sICAM-1 level. The sICAM-1 levels were log-transform to achieve a normal distribution (Shapiro normality p=0.52). There was not significant difference between age, BMI, menstrual phase, smoking status, alcohol use, education level, work status, marital status, ethnicity and blood-group between the cases and controls ($p>0.05$). Logistic regression analysis was conducted to test for difference log-sICAM1 levels between endometriosis cases vs. controls in both an adjusted model and also a model where BMI, education, work-status and blood-group was included as covariates. The logistic regression analysis was also conducted comparing log-sICAM1 levels in rASRM stage I/II vs. controls and rASRM stage III/IV vs. controls.

Sub-phenotype analysis

Sub-phenotype definitions. Endometriosis definitions within each population contributing to these sub-phenotype analyses remained consistent with those described above that were utilized for the overall endometriosis discovery. Those included in these sub-phenotype analyses are described again in brief below. *Surgical sub-phenotypes* were defined as presence of: a) superficial peritoneal lesions (shallow lesions found on peritoneal surface of the pelvic cavity), b) deep peritoneal lesions (lesions with >5mm depth that can infiltrate bowel, ureters, bladder, etc., including rectovaginal lesions), c) endometrioma (cystic lesions implanted on the surface or infiltrating one or both ovaries).

Symptom sub-phenotypes were defined as report of: a) dysmenorrhea (pelvic pain / pelvic cramping occurring during menstruation; dysmenorrhea was categorized as ever having experienced dysmenorrhea and also by reported dysmenorrhea-specific pain severity experienced within the last three months), b) dyspareunia (pelvic pain during or within 24 hours after sexual intercourse; dyspareunia was categorized as ever having experienced dyspareunia and also reported worst ever dyspareunia-specific pain severity), c) acyclic pelvic pain (chronic recurrent pelvic pain experienced at any time throughout the menstrual cycle; acyclic pelvic pain was categorized as ever having experienced acyclic pelvic pain and also reported current acyclic pelvic pain severity), d) gastrointestinal pain/irritable bowel syndrome symptoms (present at any time), e) bladder pain (present at any time). Severity for dysmenorrhea, dyspareunia, and acyclic pelvic pain symptoms was determined using the numerical rating scale with range 0-10 – with 0 reflecting no pain, and 7-10 defined as severe pain[25].

Common morbidity sub-phenotype included adenomyosis, endometriosis of the uterus, where endometrium infiltrates myometrium. Adenomyosis is commonly observed in endometriosis patients, where prevalence of adenomyosis in endometriosis has been estimated as high as 91%[26].

Data sources. Nine studies, consisting of 18,867 endometriosis cases and 301,088 controls, contributed to the sub-phenotype analyses (Supplementary Table 21). Surgical and symptom sub-phenotype data were compliant with the World Endometriosis Research Foundations' Endometriosis Phenome and Biobanking Harmonization Project (WERF-EPHect)[27-29]. As defined for each contributing cohort above, cases were confirmed via surgery for five clinic populations (A2A, Oxford-P1&P2, Leuven, UCSF) and via self-report, medical records, or claims databases for four national populations (FinnGen, NHSII, EGCUT, UKBB). Controls from all nine cohorts were restricted to those who had never been diagnosed with or suspected of having endometriosis. Three studies (Oxford-P1&P2, Leuven) further restricted controls to those with surgery during which endometriosis was not visualized. Harmonized genomic data for each site underwent standardized preparation and quality control. For adenomyosis, two datasets, namely UKBB and FinnGen (3,531 cases, 258,562 controls) were utilised that define adenomyosis from medical records (ICD10 code: N800). Quality control included removing duplicates, removing samples and SNPs with low call rates, removing subjects with

extreme heterozygosity, removing SNPs showing extreme deviations from Hardy-Weinberg equilibrium, removing samples showing a mis-match between self-reported and genetic sex, removal of close relative pairs per site-specific cut-offs. GWAS scaffold was mapped to the human reference genome assembly build 37, and genotypes phased and imputed up to Haplotype Reference Consortium (HRC r1.1 2016), 1000 Genomes (1000G P3v5), or population-specific whole genome sequence data[30].

Statistical analyses. Sites contributing to the sub-phenotype analyses conducted association statistical analyses per the standardized protocol provided to them. Phenotype measurement was operationalized as a combination of counts, presence/absence, or numerical rating pain scales as described above (Supplementary Table 20). Study sites conducted case-control association tests using regression methods implemented in SNPTEST software[31], EPACTS software[32], or other equivalent software. Selection of co-variates was site-specific based on variables deemed most relevant in each setting. A log-additive model was assumed, and each site independently conducted the identical statistical analyses in-house and provided summary statistics for their respective study with descriptive quality control data via FTP to a cloud-based server for meta-analysis. Summary results from study sites underwent quality control; thresholds for data removal had been defined previously by the analysis plan and conducted by study sites. All summary statistics provided were again quality checked, and files cleaned including variable renaming for meta-analysis.

Independent GWAS analysis of endometriosis and adenomyosis in UK Biobank

The largest dataset where we had adenomyosis diagnosis in our case/control groups is UK Biobank. In UK Biobank, adenomyosis cases are coded in the ICD10 and ICD 9 codes of N800 and 1670. We have identified a set of 1764 adenomyosis cases that did not report endometriosis diagnosis in either medical record data or self-reported data. We have also identified 2729 endometriosis cases that did not have an adenomyosis diagnosis. For this we have restricted the endometriosis cases to only the ones with ICD9/10 codes as for self-reported endometriosis cases there are insufficient self-reported adenomyosis diagnosis reported. We also partitioned the female endometriosis and adenomyosis free controls into two equal groups randomly. In total, adenomyosis only GWAS was conducted in 1,764 cases and 106,763 controls and endometriosis only GWAS

was conducted in 2,729 cases vs. 106,979 controls using lmm model in BOLT[33]. From both GWAS results, we extracted the 42 endometriosis associated lead SNPs, converted the lmm-beta and standard error to the log-odds scale. The z-score and p-value were then calculated to test whether there is statistical difference between two odds ratio estimates to determine if any of the 42 lead SNP associations has statistically different effects in endometriosis GWAS vs. adenomyosis GWAS.

Calculation of the z-score and p-value:

(1) Take the absolute value of the difference between the two log odds ratios (δ).

(2) Calculate the standard error for δ, SE(δ), using the formula:

$$\sqrt{(SE1^2+SE2^2)}$$

(3) Calculate the Z score for the test:

$$z=\delta/SE(\delta)$$

(4) Calculate the p-value from the z score in R:

P-value=2*(1-pnorm(Z))

Contributors to FinnGen

Aarno Palotie[1], Adam Ziemann[2], Adele Mitchell[3], Adriana Huertas-Vazquez[4], Aino Salminen[5], Airi Jussila[6], Aki Havulinna[1], Alex Mackay[7], Ali Abbasi[2], Amanda Elliott[1,8], Amy Cole[9], Anastasia Shcherban[1], Anders Mälarstig[10], Andrea Ganna[1], Andrey Loboda[4], Anna Podgornaia[4], Anne Lehtonen[2], Anne Pitkäranta[11], Anne Remes[12], Annika Auranen[6], Antti Hakanen[13], Antti Palomäki[14], Anu Jalanko[1], Anu Loukola[11], Aparna Chhibber[4], Apinya Lertratanakul[2], Arto Lehisto[1], Arto Mannermaa[15], Åsa Hedman[10], Audrey Chu[16], Aviv Madar[9], Awaisa Ghazal[1], Benjamin Challis[7], Benjamin Sun[3], Beryl Cummings[17], Bridget Riley-Gillis[2], Caroline Fox[4], Chia-Yen Chen[3], Clarence Wang[18], Clement Chatelain[18], Daniel Gordin[5], Danjuma Quarless[2], Danny Oh[19], David Choy[19], David Close[7], David Pulford[16], David Rice[5], Dawn Waterworth[20], Deepak Raipal[18], Denis Baird[3], Dhanaprakash Jambulingam[21], Diana Chang[19], Diptee Kulkarni[16], Dirk Paul[7], Dongyu Liu[18], Edmond Teng[19], Eero Punkka[11], Eeva Ekholm[14], Eeva Kangasniemi[22], Eija Laakkonen[23], Eleonor Wigmore[7], Elina Järvensivu[24], Elina Kilpeläinen[1], Elisabeth Widen[1], Ellen Tsai[3], Elmutaz Mohammed[25], Erich Strauss[19], Erika Kvikstad[25], Esa Pitkänen[1], Essi Kaiharju[24], Ethan Xu[18], Fanli Xu[16], Fedik Rahimov[2], Felix Vaura[26], Franck Auge[18], Georg Brein[1], Glenda Lassi[7], Graham Heap[2], Hannele Laivuori[1], Hannele Mattsson[24], Hannele Uusitalo-Järvinen[6], Hannu Kankaanranta[6], Hannu Uusitalo[6], Hao Chen[19], Harri Siirtola[27], Heikki Joensuu[5], Heiko Runz[3], Heli Lehtonen[10], Henrike Heyne[1], Hilkka Soininen[28], Howard Jacob[2], Hubert Chen[19], Huei-Yi Shen[1], Huilei Xu[9], Ilkka Kalliala[5], Ioanna Tachmazidou[7], Jaakko Kaprio[1], Jaakko Parkkinen[10], Jaison Jacob[9], Janet Kumar[16], Janet van Adelsberg[25], Jari Laukkanen[29], Jarmo Ritari[30], Javier Garcia-Tabuenca[27], Jeffrey Waring[2], Jennifer Schutzman[19], Jimmy Liu[16], Jiwoo Lee[1,8], Joanna Betts[16], Joel Rämö[1], Johanna Huhtakangas[12], Johanna Mäkelä[22], Johanna Mattson[5], Johanna Schleutker[13], Johannes Kettunen[31], John Eicher[16], Jonas Zierer[9], Jonathan Chung[9], Joni A. Turunen[5], Jorge Esparza Gordillo[16], Joseph Maranville[25], Juha Karjalainen[1,8], Juha Mehtonen[1], Juha Rinne[14], Juha Sinisalo[5], Juhani Junttila[31], Jukka Koskela[5], Jukka Partanen[30,32], Jukka Peltola[6], Julie Hunkapiller[19], Jussi Pihlajamäki[28], Justin Wade Davis[2], Juulia Partanen[1], Kaarin Mäkikallio[14], Kai Kaarniranta[28], Kaisa Tasanen[12], Kaj Metsärinne[14], Kalle Pärn[1], Karen S. King[16], Kari Eklund[5], Kari Linden[10], Kari Nieminen[6], Katariina Hannula-Jouppi[5], Katherine Call[18], Katherine Klinger[18], Kati Donner[1], Kati Hyvärinen[30], Kati Kristiansson[24], Katja Kivinen[1], Katri Kaukinen[6], Katri Pylkäs[32], Katrina de Lange[9], Keith Usiskin[25], Kimmo Palin[33], Kirill Shkura[4], Kirsi Auro[16], Kirsi Kalpala[10], Kirsi Sipilä[12], Klaus Elenius[14], Kristin Tsuo[1,8], Elisa L. Lahtela[1], Laura Addis[16], Laura Huilaja[12], Laura Kotaniemi-Talonen[6], Laura Mustaniemi[35], Laura Pirilä[14], Laure Morin-Papunen[12], Lauri Aaltonen[5], Leena Koulu[14], Liisa Suominen[28], Lila Kallio[13], Linda McCarthy[16], Liu Aoxing[1], Lotta Männikkö[24], Maen Obeidat[9], Manuel Rivas[35], Marco Hautalahti[35], Margit Pelkonen[28], Mari E. Niemi[1], Mari Kaunisto[1], Maria Siponen[28], Marika Crohns[18], Marita Kalaoja[32], Marja Luodonpää[12], Marja Vääräsmäki[12], Marja-Riitta Taskinen[5], Marjo Tuppurainen[28], Mark Daly[1], Mark McCarthy[19], Markku Laakso[28], Markku Laukkanen[24], Markku Voutilainen[14], Markus Juonala[14], Markus Perola[24], Marla Hochfeld[25], Martti Färkkilä[5], Mary Pat Reeve[1], Masahiro Kanai[8], Matt Brauer[17], Matthias Gossel[18], Matti Peura[1], Meg Ehm[16], Melissa Miller[10], Mengzhen Liu[2], Mervi Aavikko[1], Miika Koskinen[11], Mika Kähönen[6], Mikko Arvas[30,32], Mikko Hiltunen[28], Mikko Kiviniemi[28], Minal Caliskan[25], Minna Karjalainen[32], Minna Raivio[5], Mirkka Koivusalo[35], Mitja Kurki[1,8], Mutaamba Maasha[8], Nan Bing[10], Natalie Bowers[19], Neha Raghavan[4], Nicole Renaud[9], Niko Välimäki[33], Nina Hautala[12], Nina Mars[1], Nina Pitkänen[13], Nizar Smaoui[2], Oili Kaipiainen-Seppänen[28], Olli Carpén[11], Oluwaseun A. Dada[1], Onuralp Soylemez[4], Oskari Heikinheimo[5], Outi Tuovila[36], Outi Uimari[12], Padhraig Gormley[16], Päivi Auvinen[28], Päivi Laiho[24], Päivi Mäntylä[28], Päivi Polo[14], Paola Bronson[3], Paula Kauppi[5], Peeter Karihtala[12], Pekka Nieminen[5], Pentti Tienari[5], Petri Virolainen[13], Pia Isomäki[6], Pietro Della Briotta Parolo[1], Pirkko Pussinen[5], Priit Palta[1], Raimo Pakkanen[36], Raisa Serpi[31], Rajashree Mishra[16], Reetta Hinttala[31], Reetta Kälviäinen[28], Regis Wong[24], Relja Popovic[2], Richard Siegel[9], Riitta Lahesmaa[14], Risto Kajanne[1], Robert Graham[17], Robert Plenge[25], Robert Yang[20], Roosa Kallionpää[14], Ruoyu Tian[3], Russell Miller[10], Sahar Esmaeeli[2], Saila Kauppila[12], Sally John[3], Sami Heikkinen[37], Sami Koskelainen[24], Samir Wadhawan[25], Sampsa Pikkarainen[5], Samuel Heron[21], Samuli Ripatti[1], Sanna Seitsonen[5], Sanni Lahdenperä[3], Sanni Ruotsalainen[1], Sarah Pendergrass[19], Sarah Smith[34], Sauli Vuoti[18], Shabbeer Hassan[1], Shameek Biswas[25], Shuang Luo[1], Sina Rüeger[1], Sini Lähteenmäki[24], Sirkku Peltonen[14], Sirpa Soini[24], Slavé Petrovski[7], Soumitra Ghosh[16], Stefan McDonough[10], Stephanie Loomis[3], Steven Greenberg[25], Susan Eaton[3], Susanna Lemmelä[1] ,Tai-He Xia[18] ,Tarja Laitinen[22], Taru Tukiainen[1], Teea Salmi[6], Teemu Niiranen[26], Teijo Kuopio[29], Terhi Kilpi[24], Terhi Ollila[5], Tero Hiekkalinna[24], Tero Jyrhämä[1], Terttu Harju[12], Tiinamaija Tuomi[5], Tim Behrens[17], Tim Lu[19], Timo Blomster[12], Timo P. Sipilä[1], Tom Southerington[34], Tomi Mäkelä[1], Tuomo Kiiskinen[1], Tuomo Mantere[31], Tushar Bhangale[19], Tuula Salo[5], Tuuli Sistonen[24], Ulla Palotie[5],

Ulvi Gursoy[14], Urho Kujala[29], Valtteri Julkunen[28], Veikko Salomaa[26], Veli-Matti Kosma[15], Venkat Subramaniam Rathinakannan[21], Venla Kurra[6], Vesa Aaltonen[14], Victor Neduva[4], Vincent Llorens[1], Vishal Sinha[1], Vuokko Anttonen[12], Wei Zhou[8], Wilco Fleuren[20], Xing Chen[10], Xinli Hu[10], Ying Wu[10], Yunfeng Huang[3].

FinnGen Endometriosis Taskforce
Outi Uimari[12], Johannes Kettunen[31], Kirsi Auro[16], Päivi Härkki[5], Oskari Heikinheimo[5], Liisu Saavalainen[5], Hannele Laivuori[1], Venla Kurra[6], Bridget Riley-Gillis[2], Zhen Liu[2].

[1] Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Finland.
[2] Abbvie, Chicago, IL, United States.
[3] Biogen, Cambridge, MA, United States.
[4] Merck, Kenilworth, NJ, United States.
[5] Hospital District of Helsinki and Uusimaa, Helsinki, Finland.
[6] Pirkanmaa Hospital District, Tampere, Finland.
[7] Astra Zeneca, Cambridge, United Kingdom.
[8] Broad Institute, Cambridge, MA, United States.
[9] Novartis, Basel, Switzerland.
[10] Pfizer, New York, NY, United States.
[11] Helsinki Biobank, Helsinki University and Hospital District of Helsinki and Uusimaa, Helsinki, Finland.
[12] Northern Ostrobothnia Hospital District, Oulu, Finland.
[13] Auria Biobank, University of Turku, Hospital District of Southwest Finland, Turku, Finland.
[14] Hospital District of Southwest Finland, Turku, Finland.
[15] Biobank of Eastern Finland, University of Eastern Finland, Northern Savo Hospital District, Kuopio, Finland.
[16] GlaxoSmithKline, Brentford, United Kingdom.
[17] Maze Therapeutics, San Francisco, CA, United States.
[18] Sanofi, Paris, France.
[19] Genentech, San Francisco, CA, United States.
[20] Janssen Biotech, Beerse, Belgium.
[21] University of Turku, Turku, Finland.
[22] Finnish Clinical Biobank Tampere, University of Tampere, Pirkanmaa Hospital District, Tampere, Finland.
[23] University of Jyväskylä, Jyväskylä, Finland.
[24] THL Biobank, The National Institute of Health and Welfare Helsinki, Finland.
[25] Celgene, Summit, NJ, United States/ Bristol Myers Squibb, New York, NY, United States.
[26] The National Institute of Health and Welfare Helsinki, Finland
[27] University of Tampere, Tampere, Finland.
[28] Northern Savo Hospital District, Kuopio, Finland.
[29] Central Finland Biobank, University of Jyväskylä, Central Finland Health Care District, Jyväskylä, Finland.
[30] Finnish Red Cross Blood Service, Helsinki, Finland.
[31] Northern Finland Biobank Borealis, University of Oulu, Northern Ostrobothnia Hospital District, Oulu, Finland.
[32] University of Oulu, Oulu, Finland.
[33] University of Helsinki, Helsinki, Finland.
[34] FinBB - Finnish biobank cooperative.
[35] University of Stanford, Stanford, CA, United States.
[36] Business Finland, Helsinki, Finland.
[37] University of Eastern Finland, Kuopio, Finland.

Contributors to DBDS Genomic Consortium

Steffen Andersen[1], Karina Banasik[2], Søren Brunak[2], Kristoffer Burgdorf[3], Christian Erikstrup[4], Thomas Folkmann Hansen[5], Henrik Hjalgrim[6], Gregor Jemec[7], Poul Jennum[8], Per Ingemar Johansson[3], Kasper Rene Nielsen[9], Mette Nyegaard[10], Topholm Bruun Mie[11], Ole Birger Pedersen[12], Mikkel Petersen[4], Erik Sørensen[3], Henrik Ullum[3], Sisse Ostrowski[3], Thomas Werge[13], Pär Ingemar Johansson[3], Daniel Gudbjartsson[14], Kari Stefansson[14], Hreinn Stefánsson[14], Unnur Þorsteinsdóttir[14].

[1] Department of Finance, Copenhagen Business School, Copenhagen, Denmark.
[2] Novo Nordisk Foundation Center for Protein Research Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark.
[3] Department of Clinical Immunology, Copenhagen University, Hospital Copenhagen, Copenhagen, Denmark.
[4] Department of Clinical Immunology, Aarhus University Hospital, Aarhus, Denmark.
[5] Danish Headache Center, Department of Neurology Rigshospitalet, Glostrup, Denmark.
[6] Department of Epidemiology, Research Statens Serum Institut, Copenhagen, Denmark.
[7] Department of Clinical Medicine, Sealand University Hospital Roskilde, Denmark.
[8] Department of Clinical Neurophysiology, University of Copenhagen, Copenhagen, Denmark.
[9] Department of Clinical Immunology, Aalborg University Hospital, Aalborg, Denmark.
[10] Department of Health, Science and Technology, Aalborg University, Aalborg, Denmark.
[11] Department of Clinical Immunology, Odense University Hospital, Odense, Denmark
[12] Department of Clinical Immunology, Naestved Hospital, Naestved, Denmark.
[13] Institute of Biological Psychiatry Mental Health Centre Sct. Hans Copenhagen University Hospital Roskilde, Denmark.
[14] deCODE genetics Reykjavik, Iceland.

Ethics Statements

The Women's Health Study: From Adolescence to Adulthood (A2A): Approved by the Institutional Review Board of Boston Children's Hospital, with IRB reliance agreements approved at Brigham and Women's Hospital and at Michigan State University College of Human Medicine.

Crete dataset (CRETE): Approved by the Ethics Research Committee of the Venizeleion Hospital of Heraklion, Crete, Greece (Ref no. ECHR 47/773).

DeCODE Genetics (DECODE): Approved by the National Bioethics Committee, Iceland (Approval number VSN-17-188).

The ENDOX study and Liverpool datasets (Oxford-P1): Approved by National Research Ethics Service (NRES) Committee South Central Oxford (09/H0604/58b). Liverpool study was approved by the Liverpool Adult Research Ethics Committee (LREC 09/H1005/55).

The ENDOX study part 2 and Liverpool and Edinburgh datasets (Oxford-P2): Approved by National Research Ethics Service (NRES) Committee South Central Oxford (09/H0604/58b). Understanding and Managing Endometriosis Study was approved by Scotland A research committee (Rec reference 11/AL/0376).

Leuven dataset (LEUVEN): Approved by the Commission of Medical Ethics of the Leuven University Hospital, Belgium and QIMR Berghofer Human Ethics Research Committee, Australia.

Lodz dataset (LODZ): Approved by the bioethical Committee of the Institute of the Polish Mother's Memorial Hospital in Lodz (Approval number, 98/2015) and approved by the University of Lodz Research Ethics Committee (approval number, 7/KBBN-UL/II/2014).

Melbourne dataset (MELBOURNE): Approved by the Human Research Ethics Committee of the Royal Women's Hospital, Melbourne, Australia (Projects 11–24 and 16–43).

Oxford Endometriosis Gene Study (OXEGENE): Approved by the Oxford regional multi-center and local research ethics committees.

Queensland Institute of Medical Research and Hunter Community Study (QIMRHCS): Approved by the QIMR Human Ethics Research Committee, the University of Newcastle and Hunter New England Population Health Human Research Ethics Committees.

University of California, San Francisco dataset (UCSF): Approved by the Institutional Review Board at the University of California San Francisco.

Vanderbilt Biorepository (BIOVU): Approved by the Institutional Review Board of Vanderbilt University Medical Center.

Danish Blood Donor Study (DBDS): Approved by the Danish Data Protection Agency under the combined approval for health care research at The Capital Region of Denmark (P-2019-99), and the national ethics committee under "Genetics of healthy ageing and specific diseases among blood donors – a GWA study under the Danish Blood Donor Study (NVK-1700407).

Generation Scotland: Scottish Family Health Study (Gen Scotland): Approved by the Tayside Committee on Medical Research Ethics (on behalf of the National Health Service).

The Estonian Biobank Cohort (EGCUT): Approved by Estonian Bioethics and Human Research Council (EBIN). Estonian Biobank data release N05.

Northern Finland Birth Cohort (NFBC): Approved by the Ethical Committee of the Northern Ostrobothnia Hospital District (ETTMK: 94/2011).

Nurses' Health Study II (NHS2): Approved by the Human Subject Committee of Harvard T.H. Chan School of Public Health and by the Institutional Review Board of Brigham and Women's Hospital.

UK Biobank dataset (UKBiobank): Approved by the North West Multi-Centre Research Ethics Committee (MREC).

23andMe dataset (23andMe): Approved by Ethical & Independent Review Services, an independent institutional review board (http://www.eandireview.com) accredited by the Association for the Accreditation of Human Research Protection Programs, Inc.

QSkin Sun and Health Study (QSKIN): Approved by QIMR Berghofer Medical Research Institute (P1309, P2034).

Twins UK dataset (TwinsUK): Approved by NRES Committee London–Westminster (REC ref: EC04/015).

The Women's Genome Health Study (WGHS): Approved by the Institutional Review Board of Brigham and Women's Hospital and Mass General Brigham.

Adachi dataset (ADACHI): Approved by the ethical committees of the Niigata University.

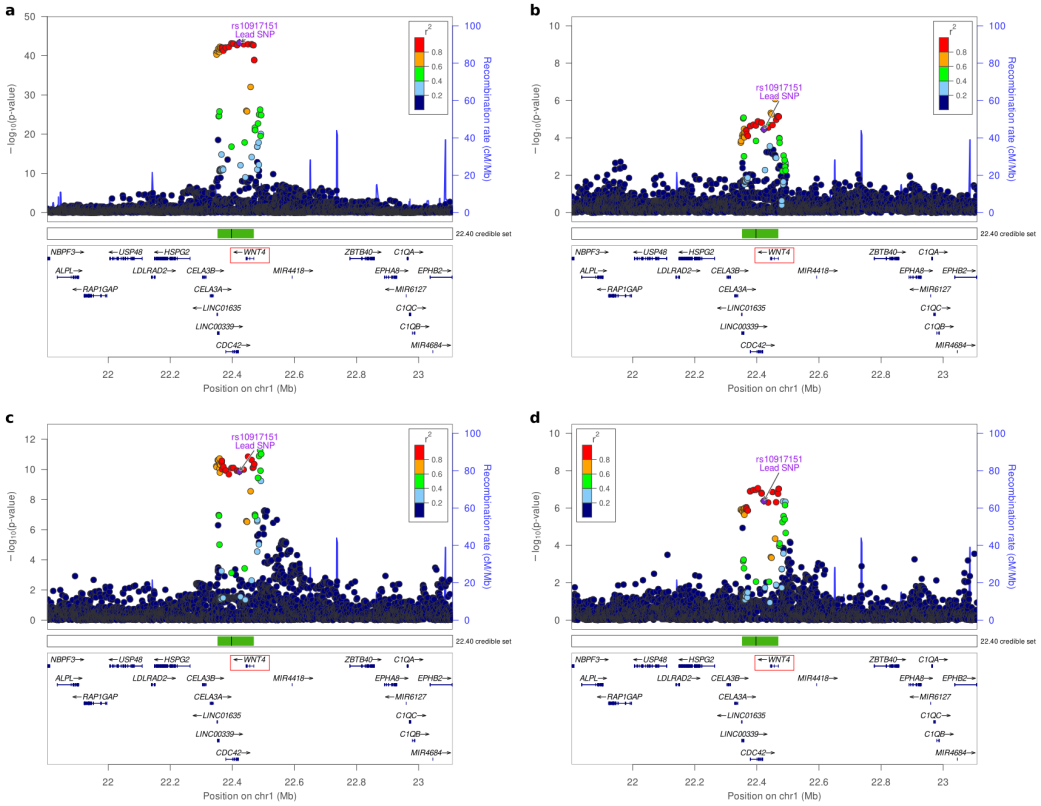
BioBank Japan dataset (BBJ): Approved by the ethical committees at the Institute of Medical Science at the University of Tokyo and the Center for Genomic Medicine at the RIKEN Yokohama Institute.

FinnGen dataset (FinnGen): Approved by the Coordinating Ethics Committee of the Hospital District of Helsinki and Uusimaa (HUS/990/2017).

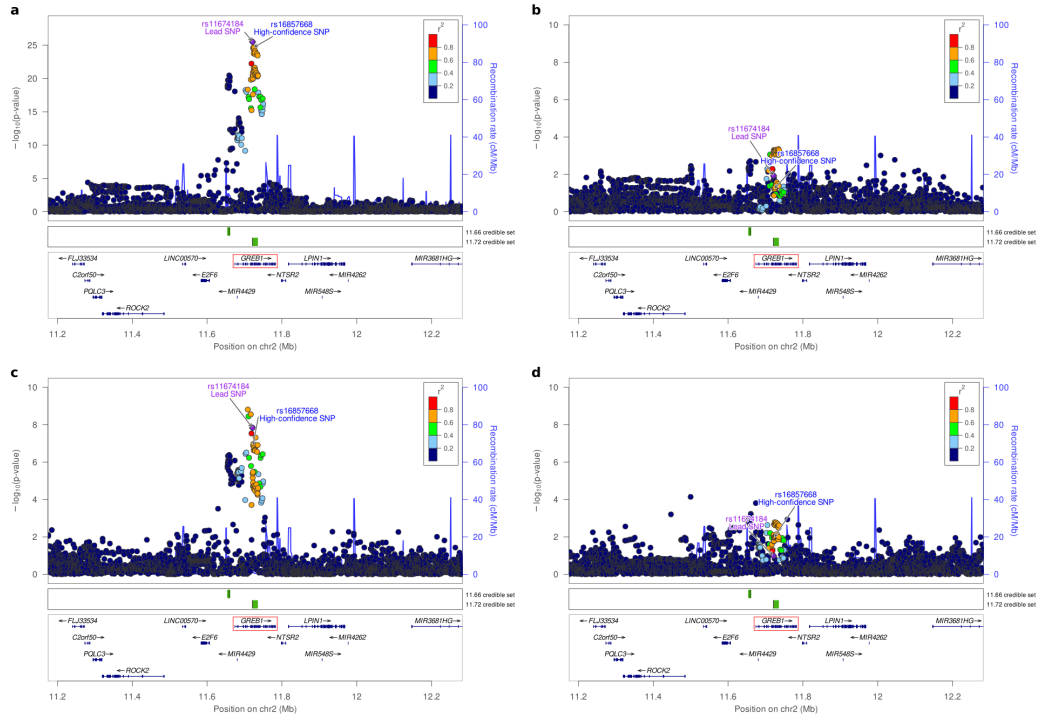
Supplementary Figures

Supplementary Figure 1. Regional association plots for 11 previously established genome-wide significant loci: loci: (i) WNT4/1p36.12, (ii) GREB1/2p25.1, (iii) ETAA1/2p14, (iv) KDR/4q12, (v) ID4/6p22.3, (vi) SYNE1/6q25.1, (vii) 7p15.2/7p15.2, (viii) 7p12.3/7p12.3, (ix) CDKN2-BAS1/9p21.3, (x) FSHB/11p14.1 (xi) VEZT/12q22 for (a) endometriosis, (b) rASRM stage III/IV endometriosis, (c) rASRM stage I/II endometriosis, (d) endometriosis associated infertility. The association results are shown on the y-axis as $-\log_{10}$(P-value) and on the x-axis is the genomic location (hg 19). The top associated SNP is coloured purple and the other SNPs are coloured according to the strength of LD with the top SNP by r^2 in the European 1000 Genomes dataset.

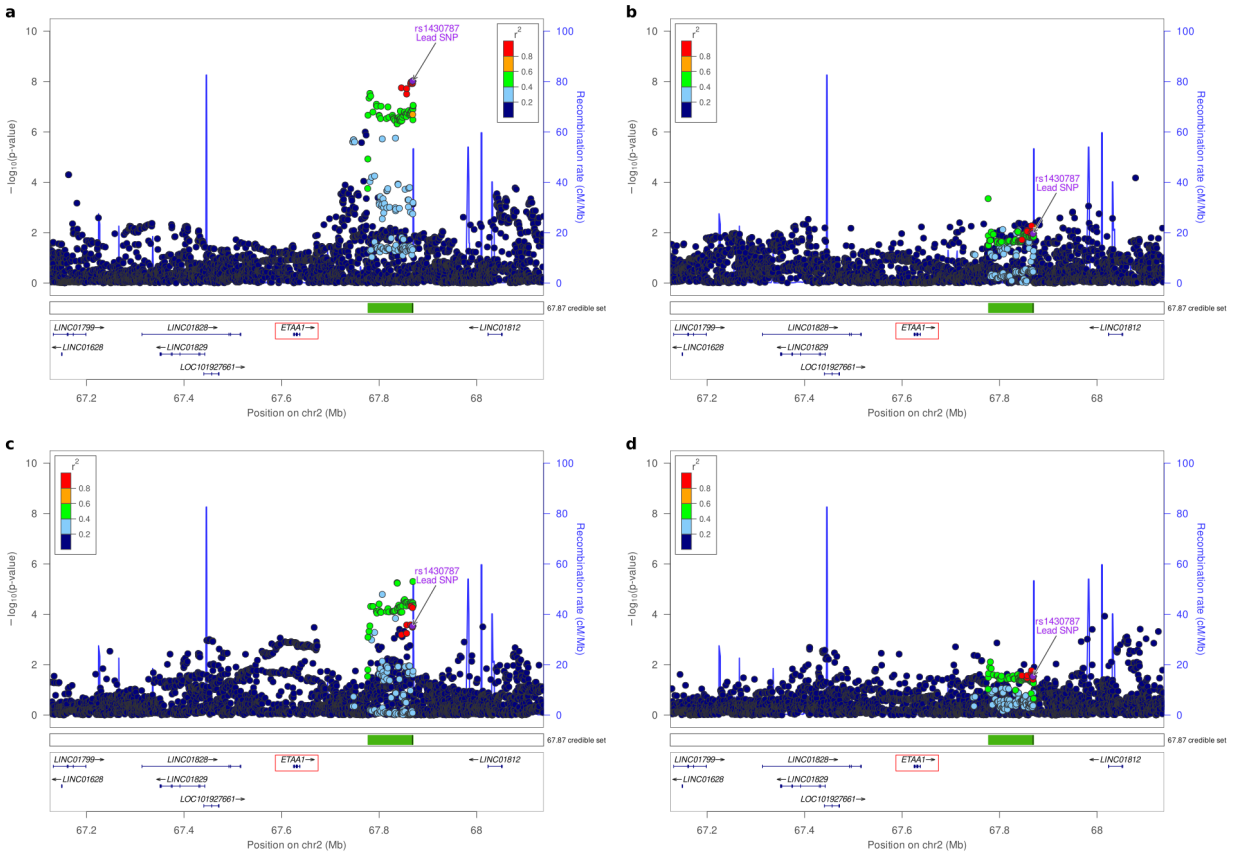
(i) WNT4/1p36.12



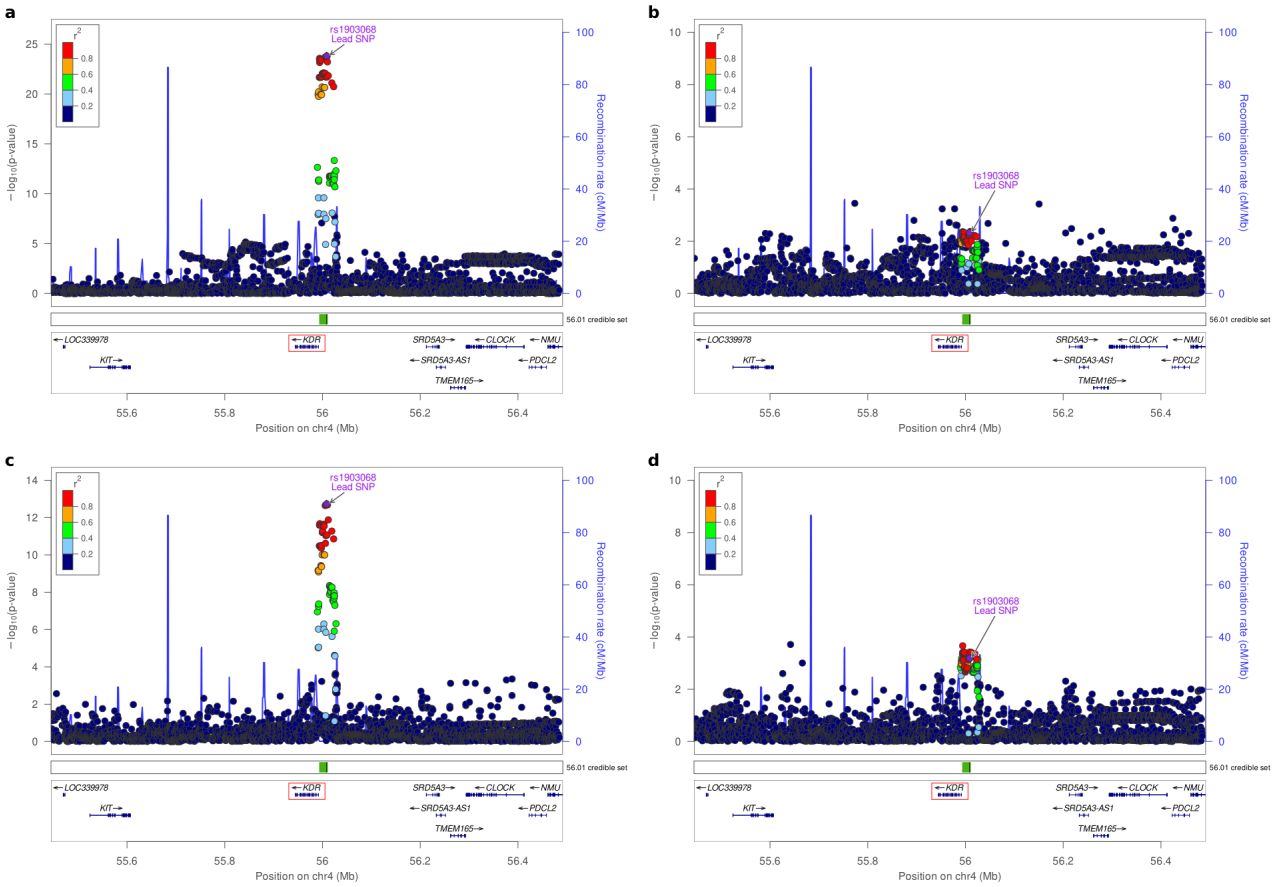
(ii) GREB1/2p25.1



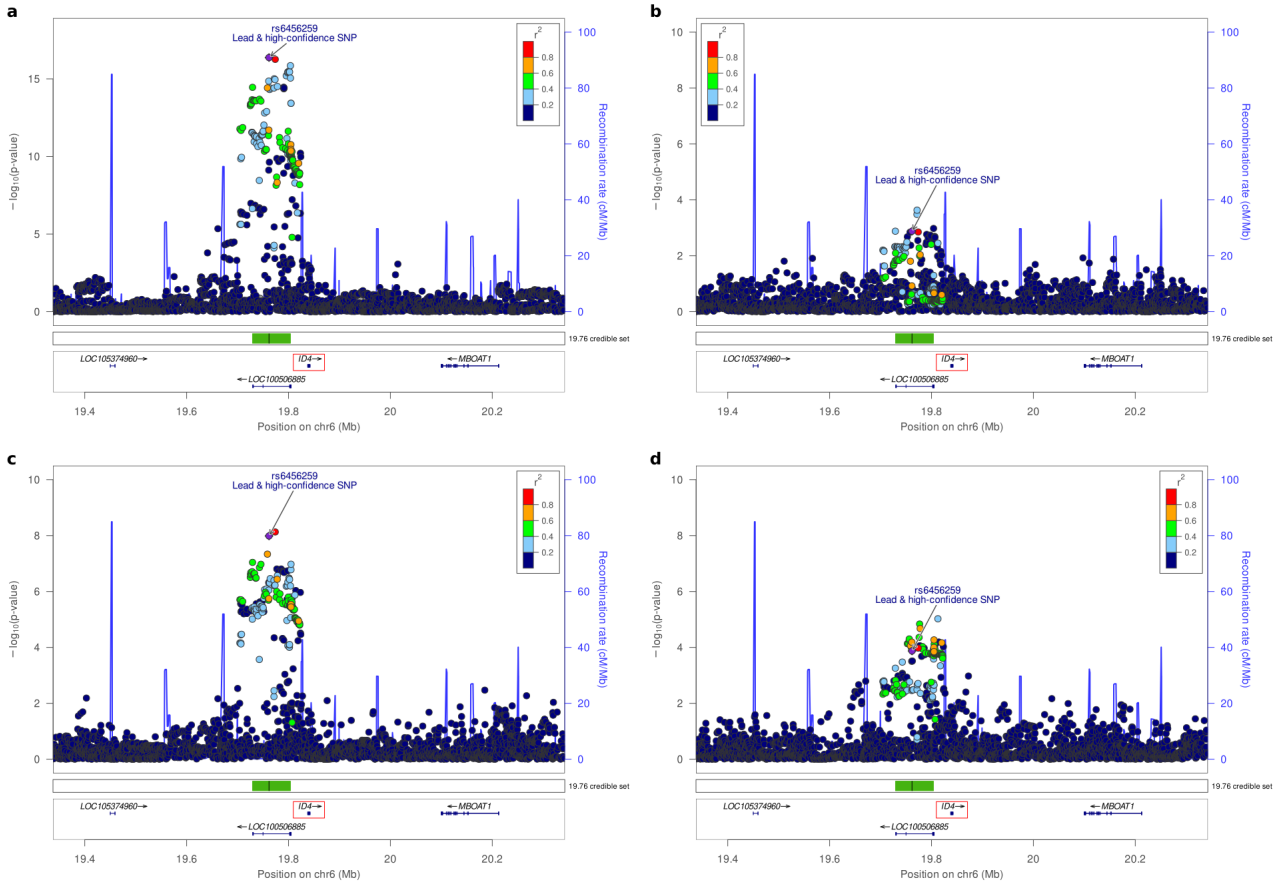
(iii) ETAA1/2p14



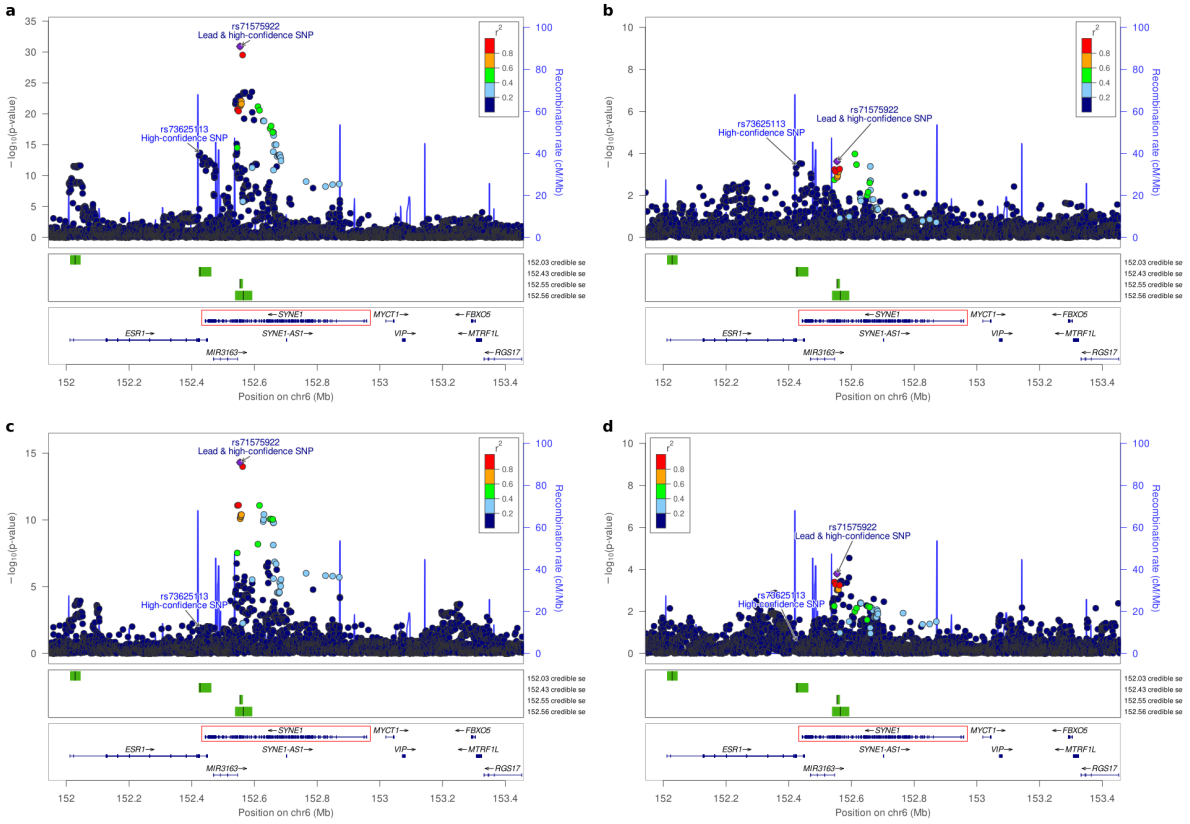
(iv) KDR/4q12



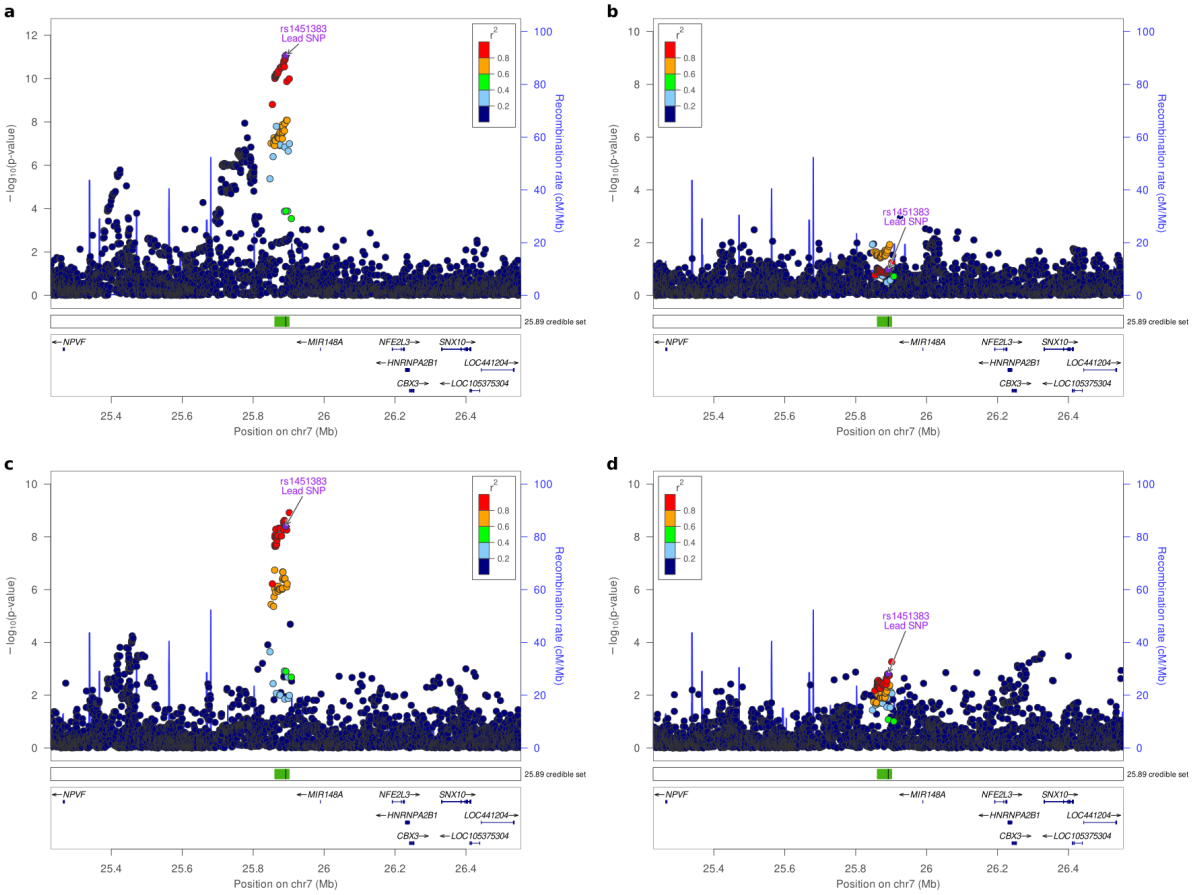
(v) ID4/6p22.3



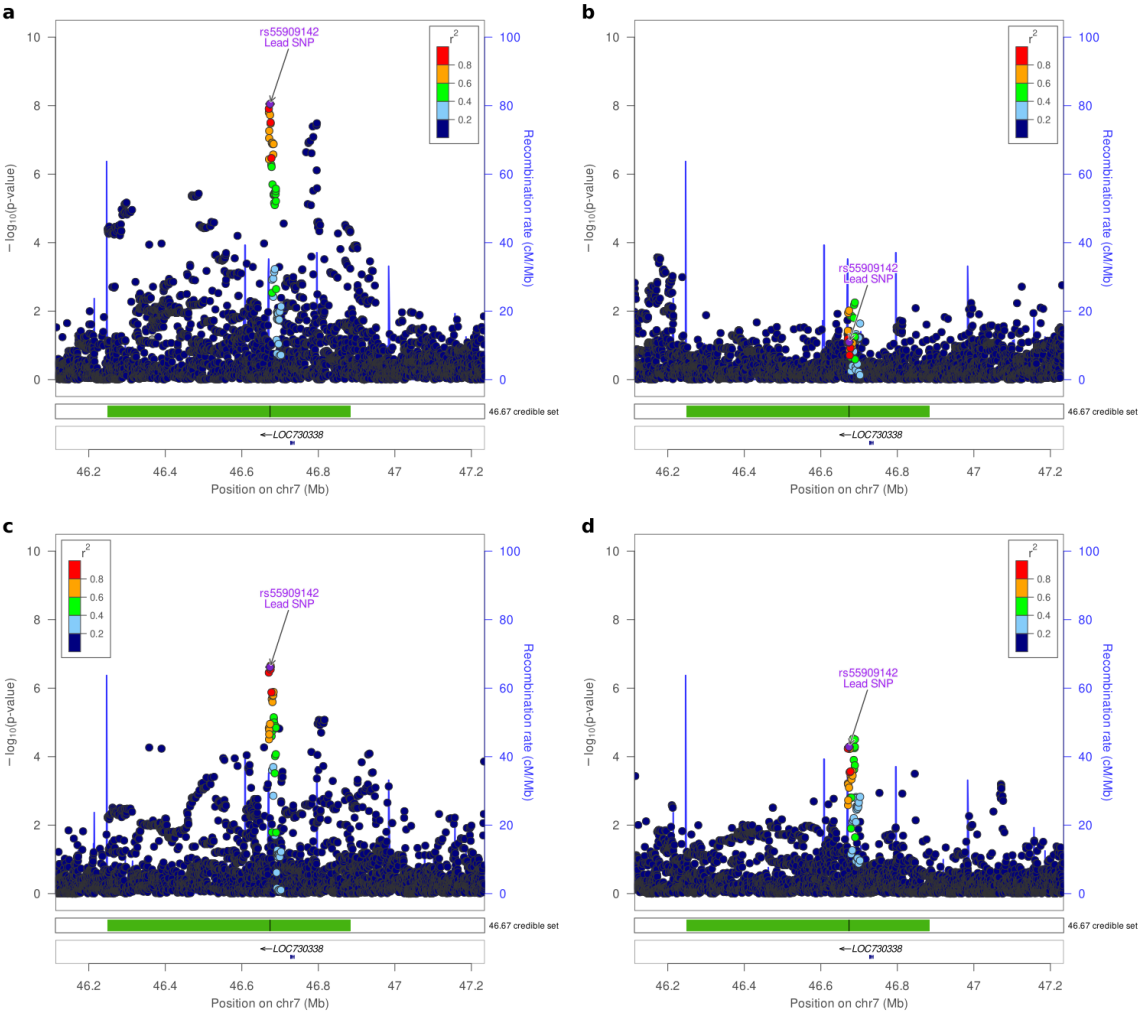
(vi) SYNE1/6q25.1



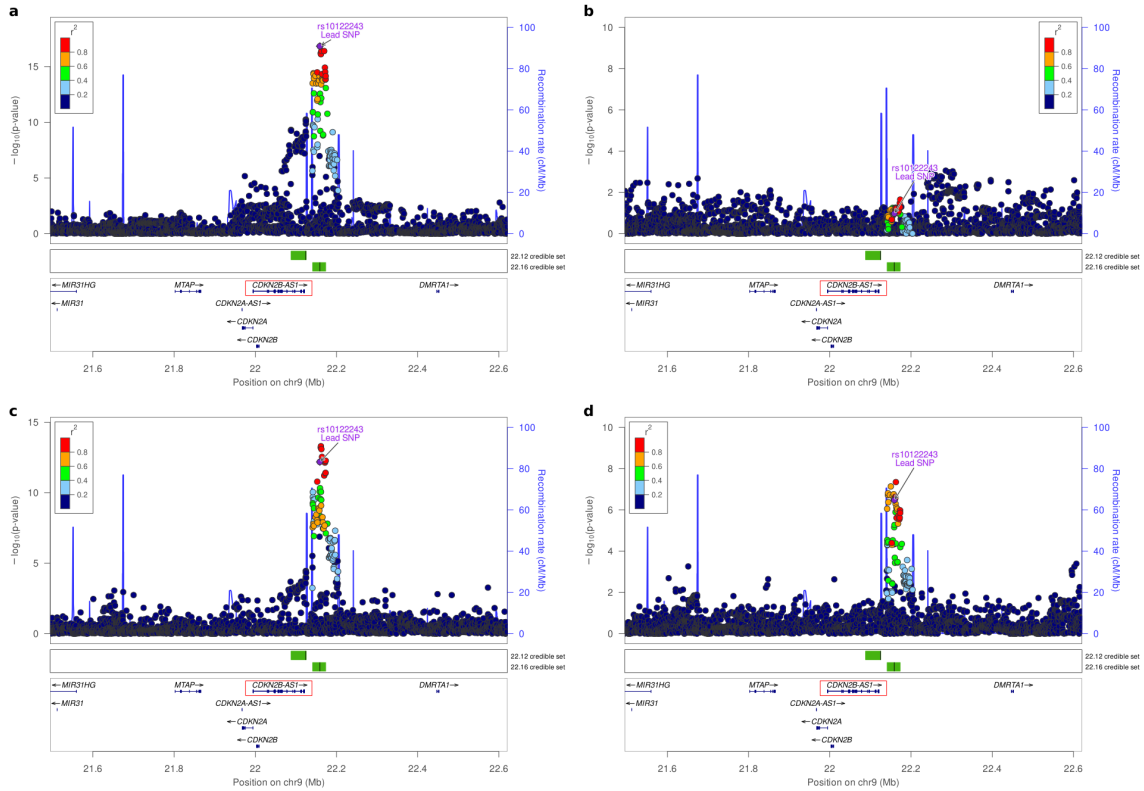
(vii) 7p15.2/7p15.2



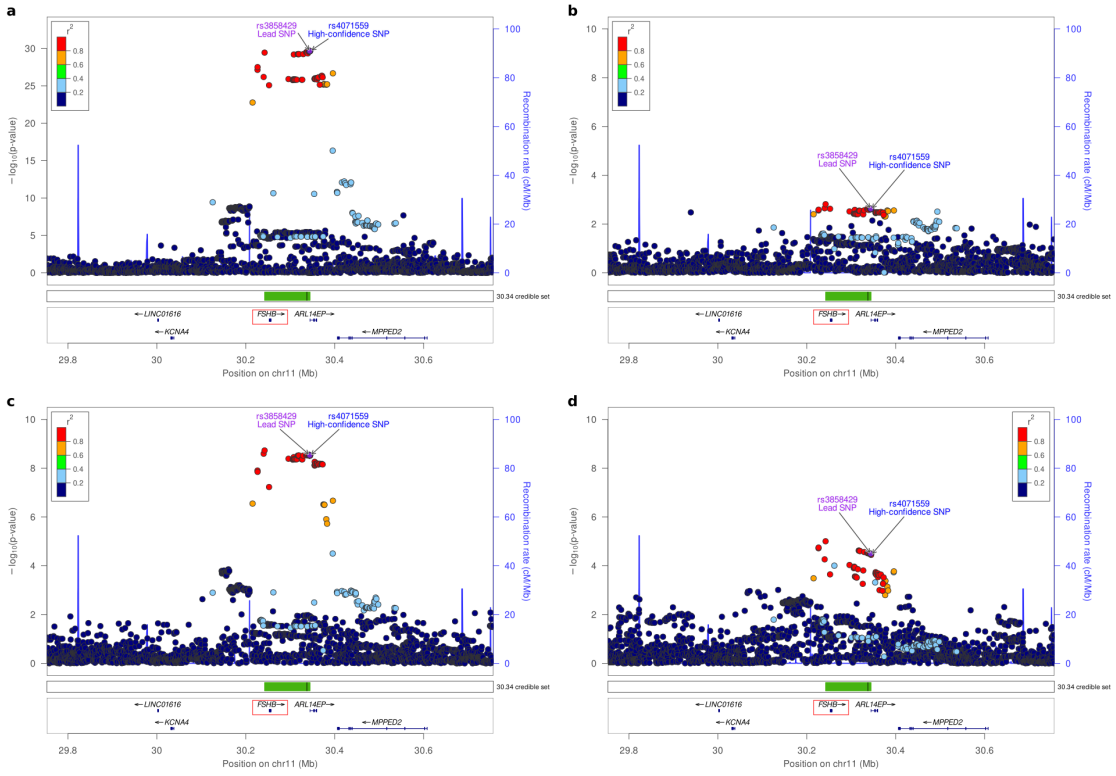
(viii) 7p12.3/7p12.3



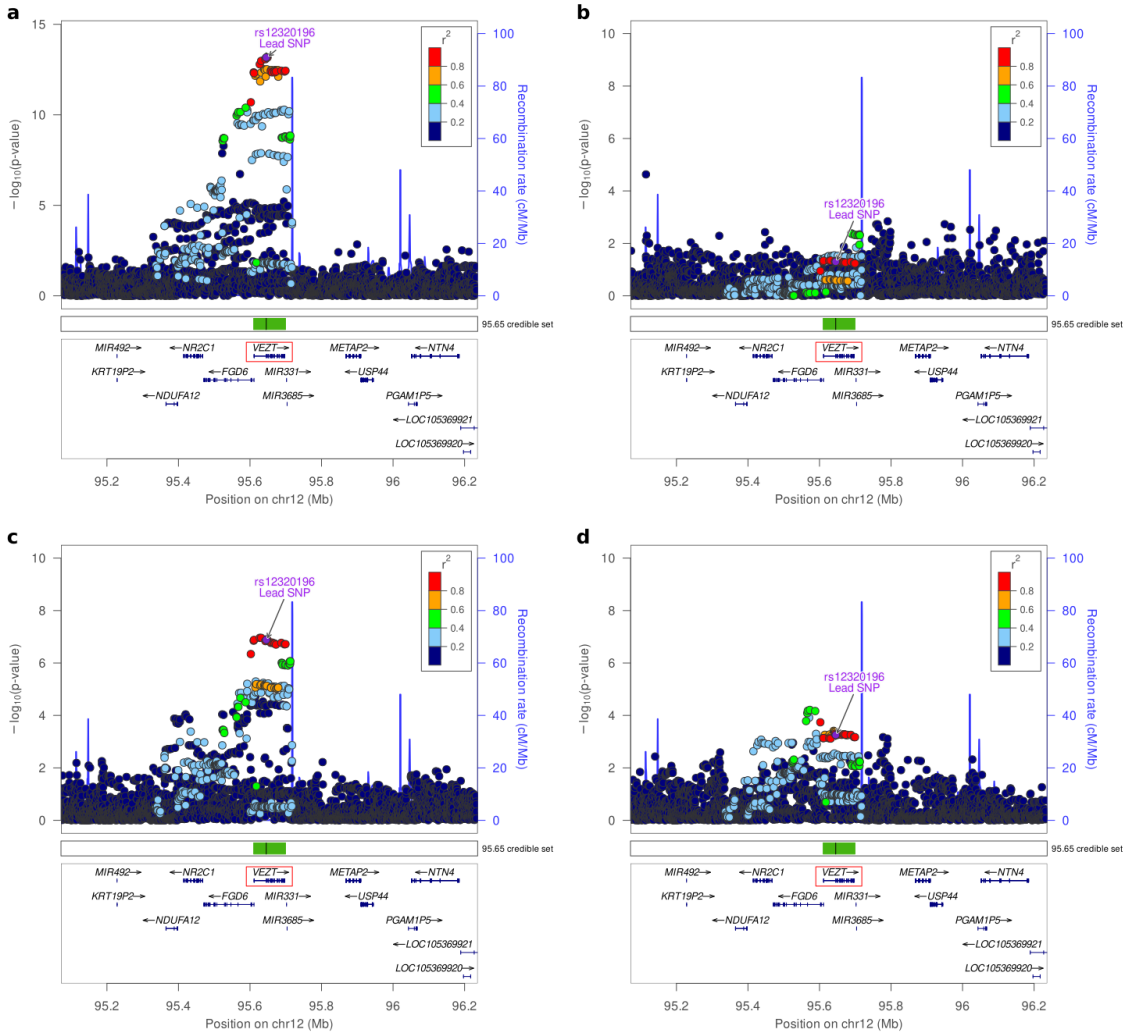
(ix) CDKN2B-AS1/9p21.3



(x) FSHB/11p14.1

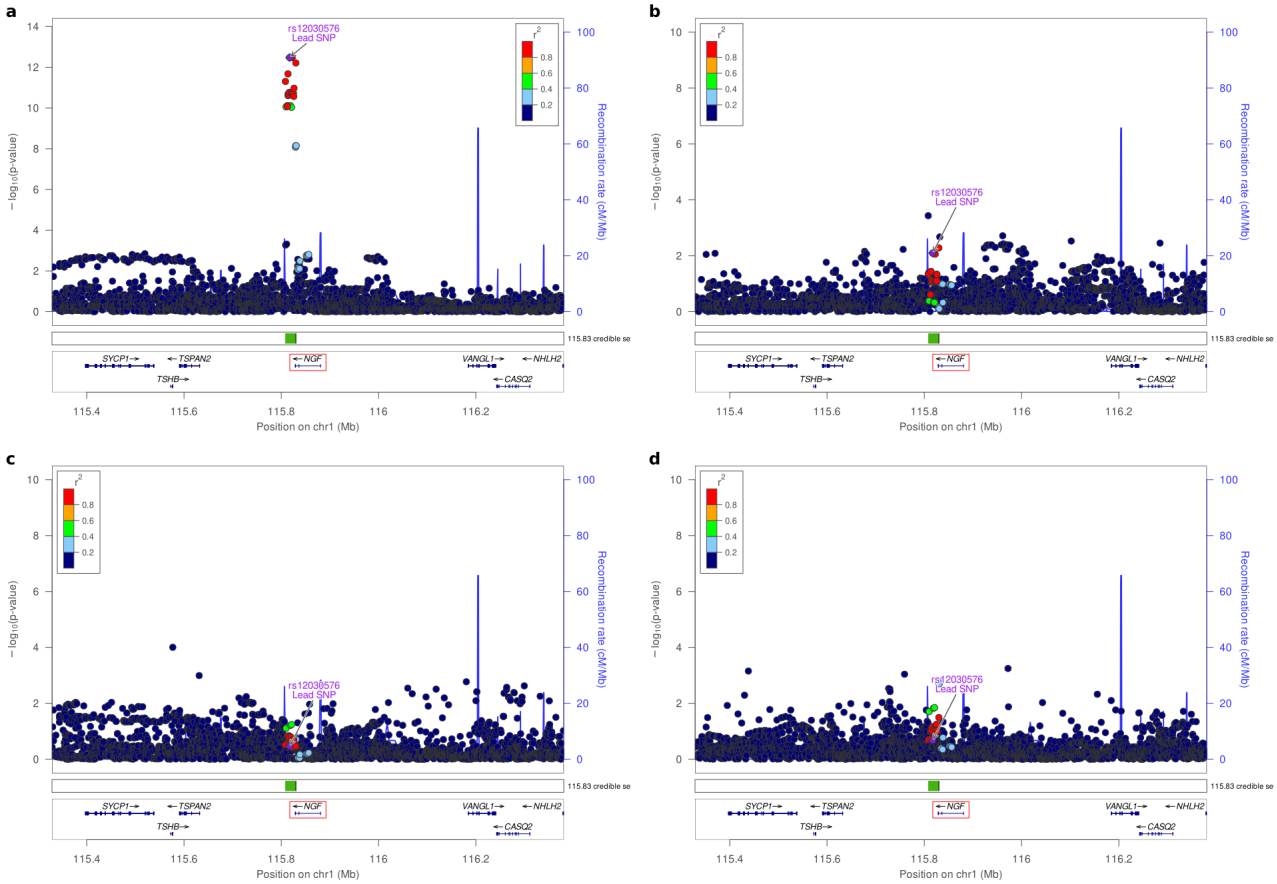


(xi) VEZT/12q22

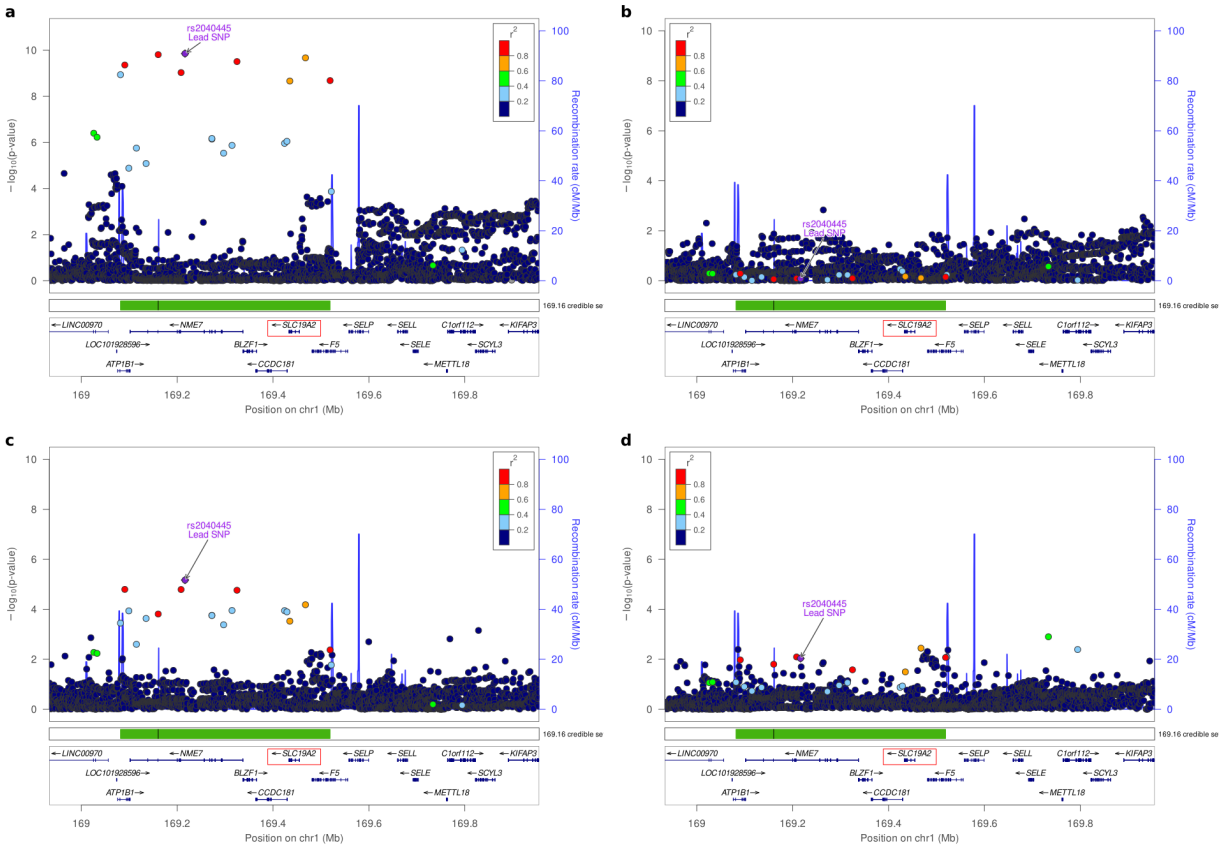


Supplementary Figure 2. Regional association plots for 31 novel genome-wide significant loci: (i) NGF/1p13.2, (ii) SLC19A2/1q24.2, (iii) DNM3/1q24.3, (iv) BMPR2/2q33.1, (v) BSN/3p21.31, (vi) PDLIM5/4q22.3, (vii) EBF1/5q33.3, (viii) CD109/6q13, (ix) HEY2/6q22.31, (x) FAM120B/6q27, (xi) HOXA10/7p15.2, (xii) KCTD9/8p21.2, (xiii) GDAP1/8q22.2, (xiv) VPS13B/8q22.2, (xv) ASTN2/9q33.1, (xvi) ABO/9q34.2, (xvii) MLLT10/10p12.31, (xviii) RNLS/10q23.31, (xix) WT1/11p14.1, (xx) PTPRO/12p12.3, (xxi) HOXC10/12p13.13, (xxii) IGF1/12q23.2, (xxiii) DLEU1/13q14.2, (xxiv) RIN3/14q32.12, (xxv) SRP14-AS1/14q32.12, (xxvi) SKAP1/17q21.32, (xxvii) CEP112/17q24.1, (xxviii) ACTL9/19p13.2, (xxix) TEX11/Xq13.1, (xxx) FRMD7/Xq26.2, (xxxi) LINC00629/Xq26.3 for (a) endometriosis, (b) rASRM stage I/II endometriosis, (c) rASRM stage III/IV endometriosis, (d) endometriosis associated infertility. The association results are shown on the y-axis as $-\log_{10}$(P-value) and on the x-axis is the genomic location (hg 19). The top associated SNP is coloured purple and the other SNPs are coloured according to the strength of LD with the top SNP by r^2 in the European 1000 Genomes dataset.

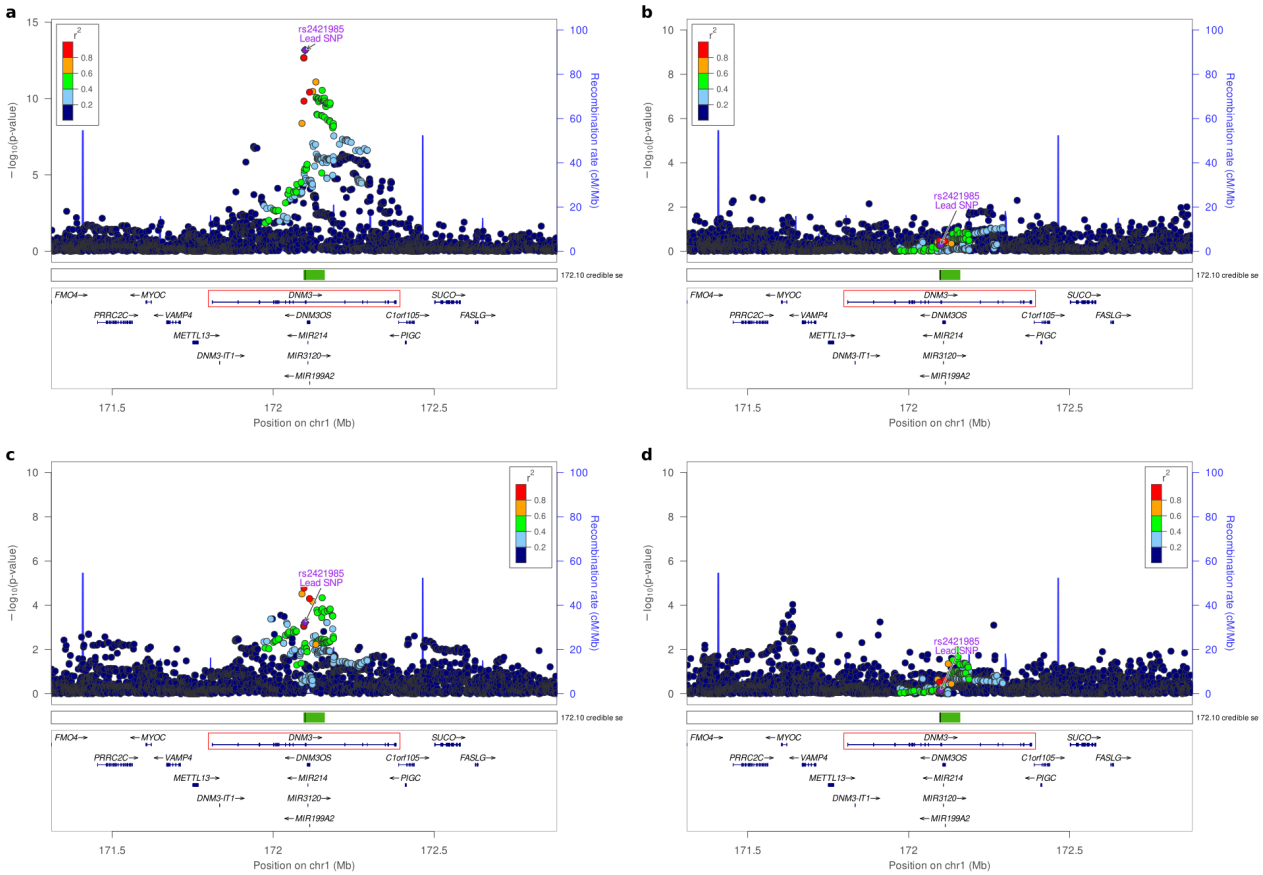
(i) NGF/1p13.2



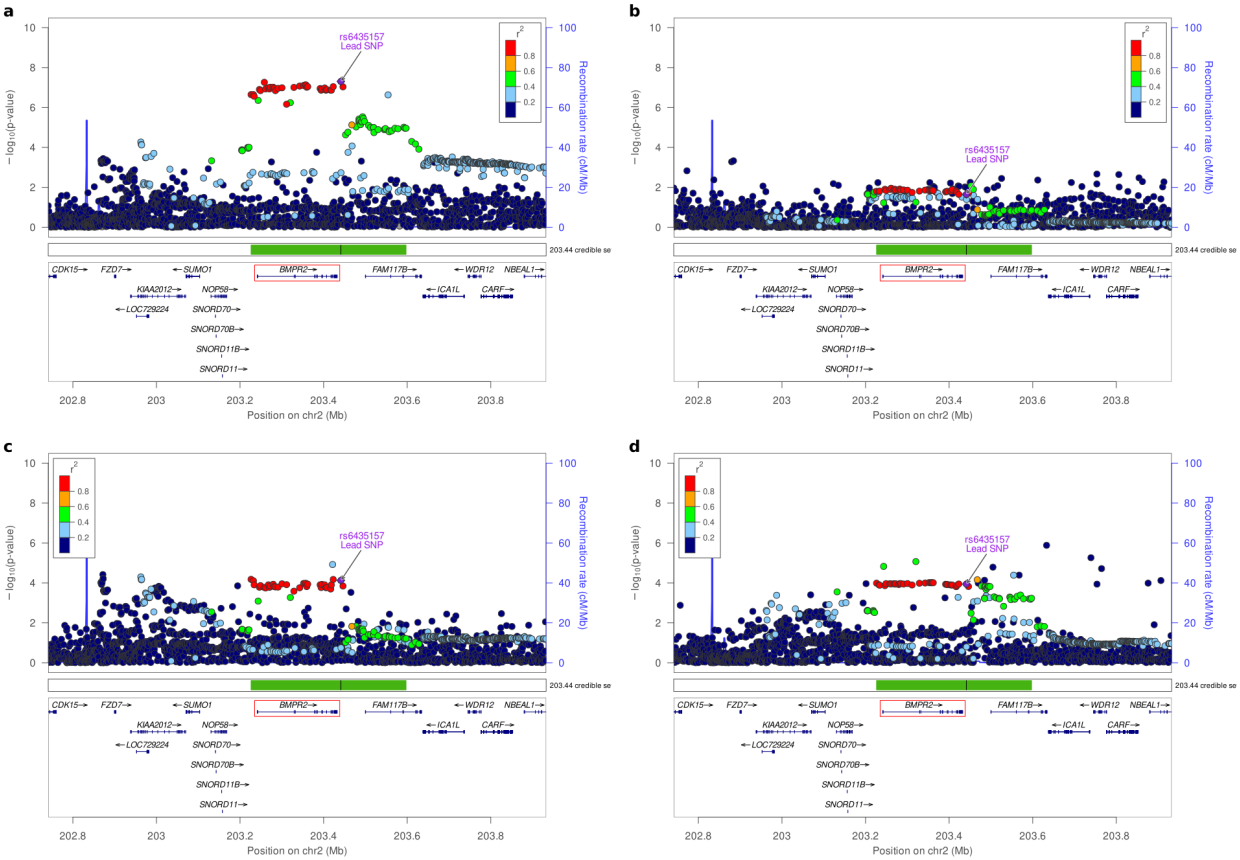
(ii) SLC19A2/1q24.2



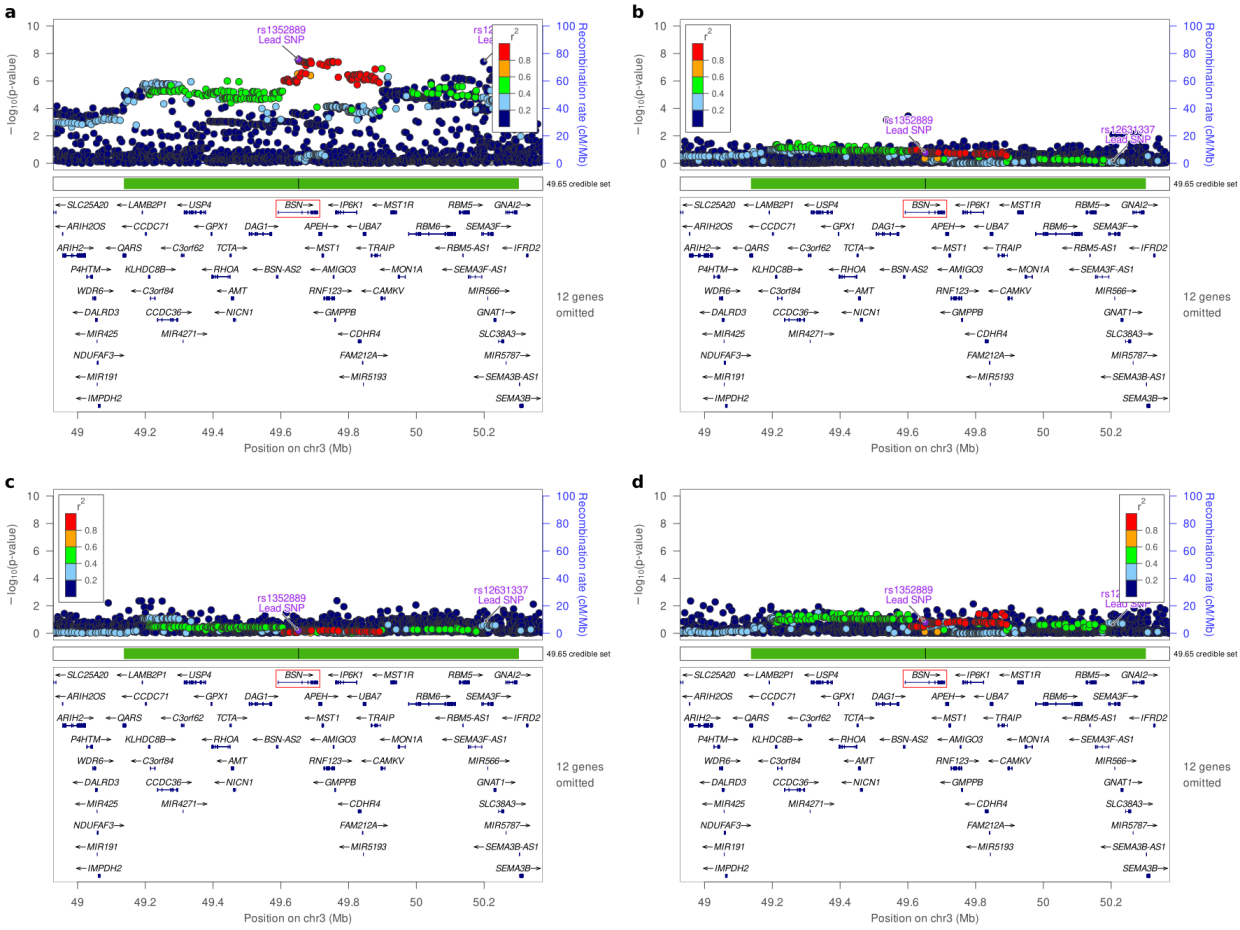
(iii) DNM3/1q24.3



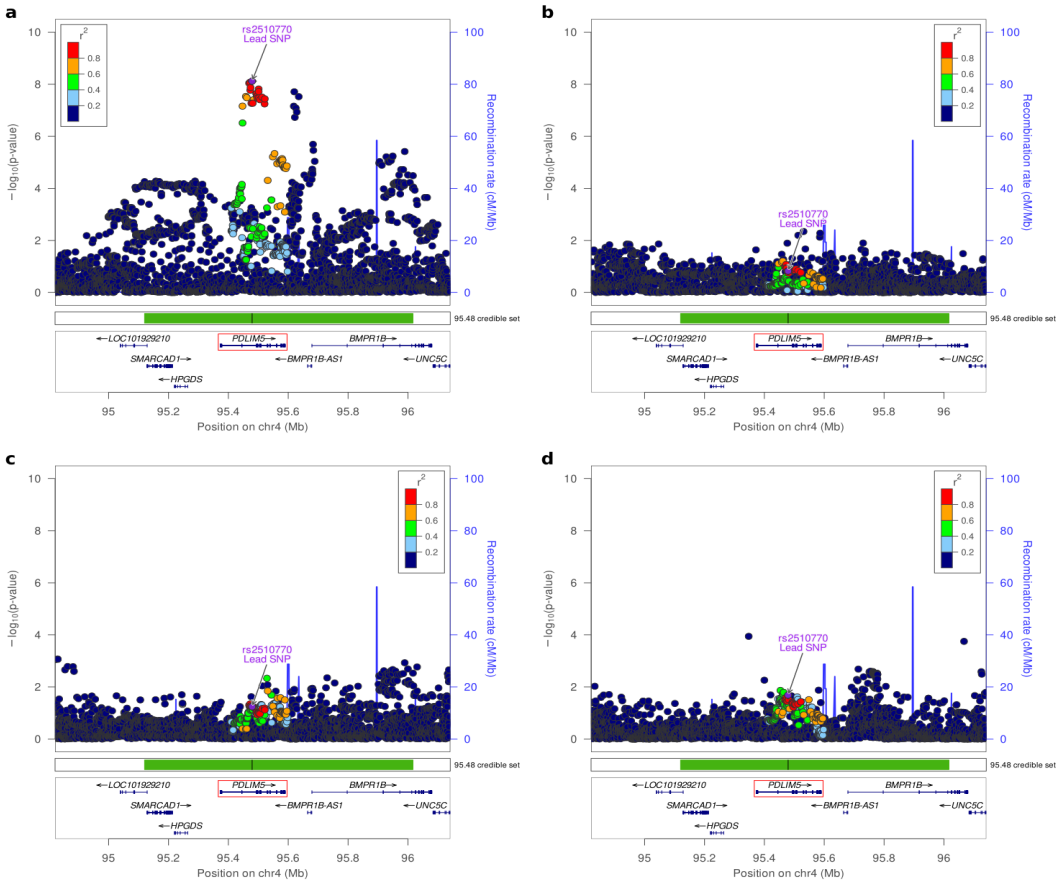
(iv) BMPR2/2q33.1



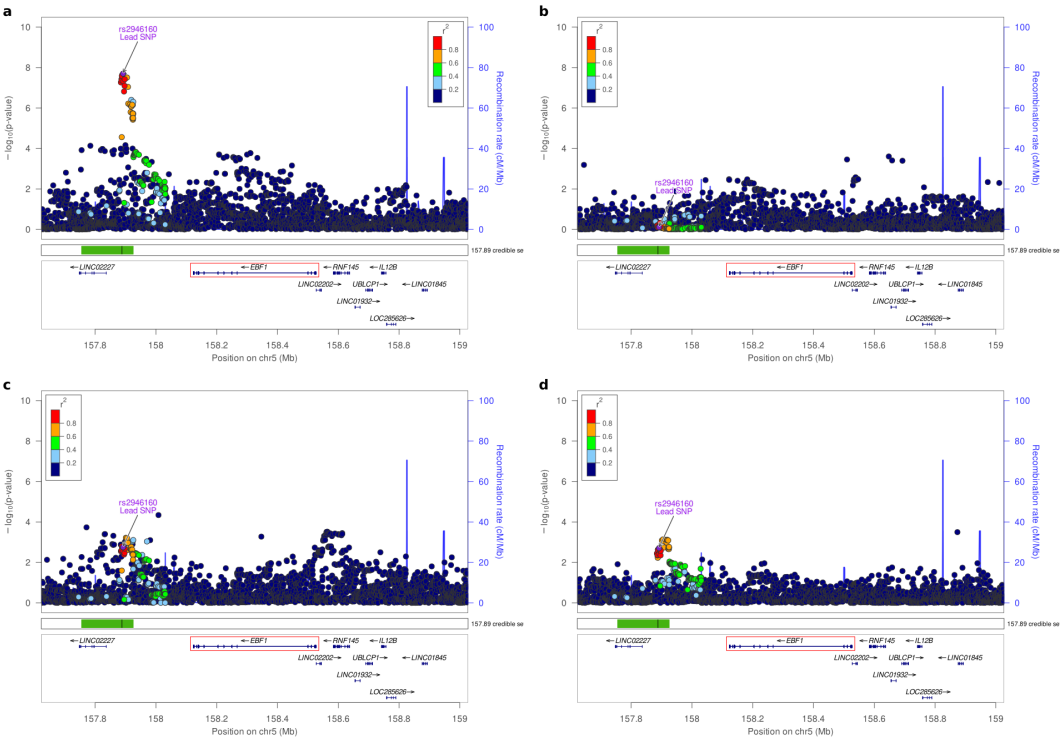
(v) BSN/3p21.31



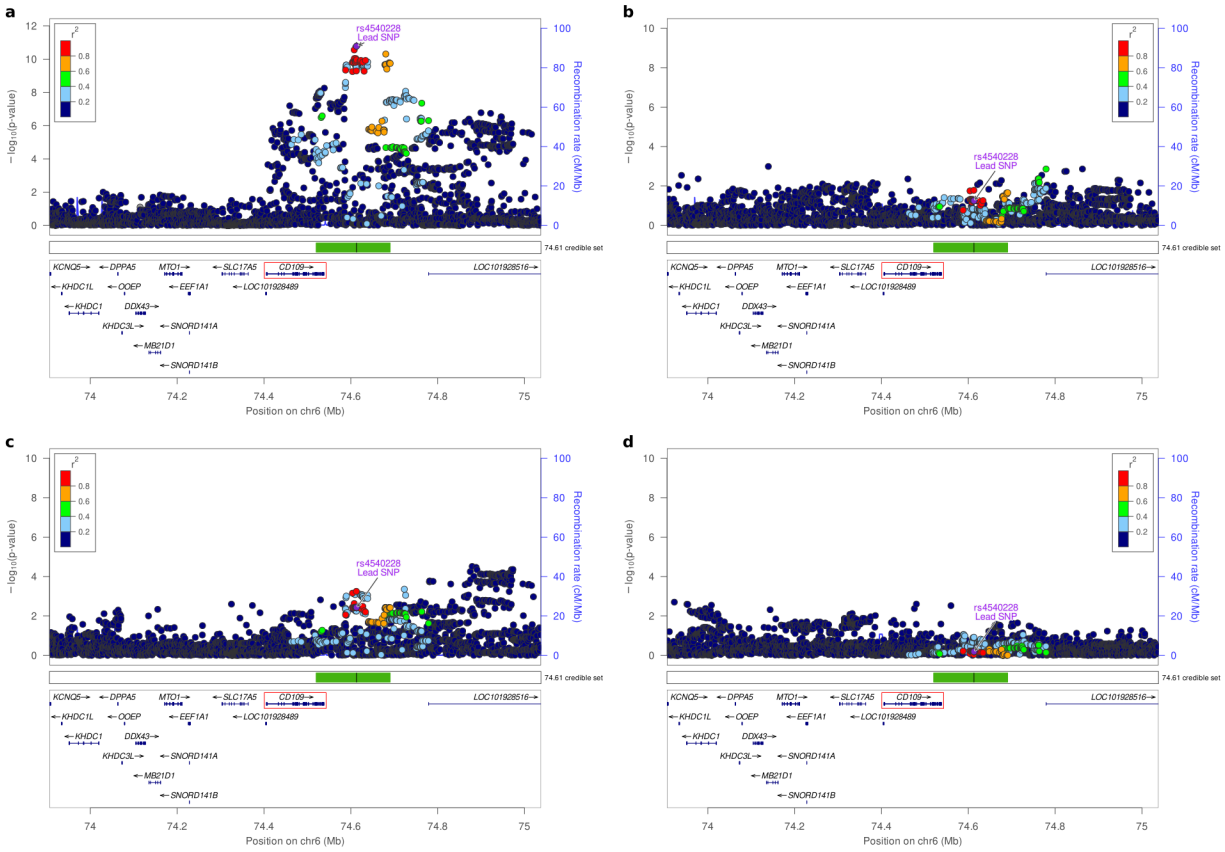
(vi) PDLIM5/4q22.3



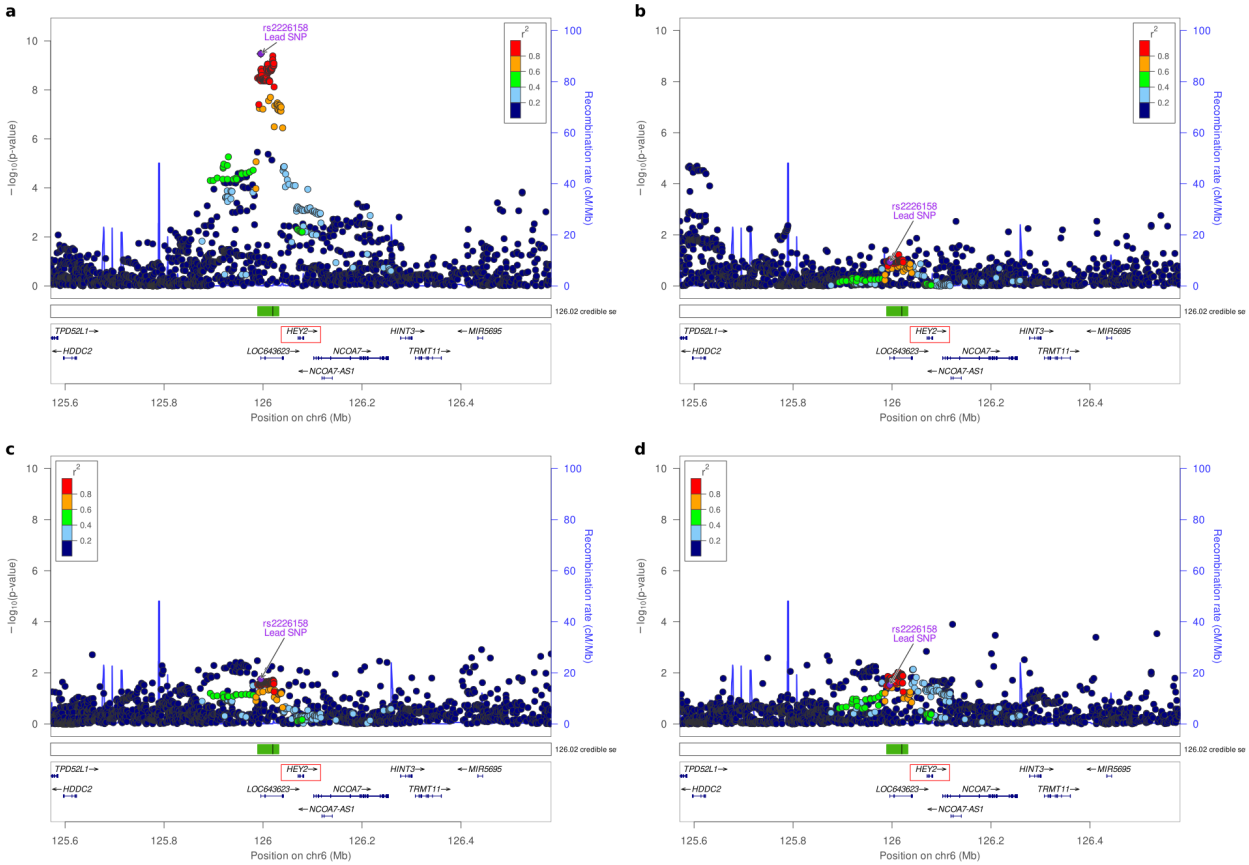
(vii) EBF1/5q33.3



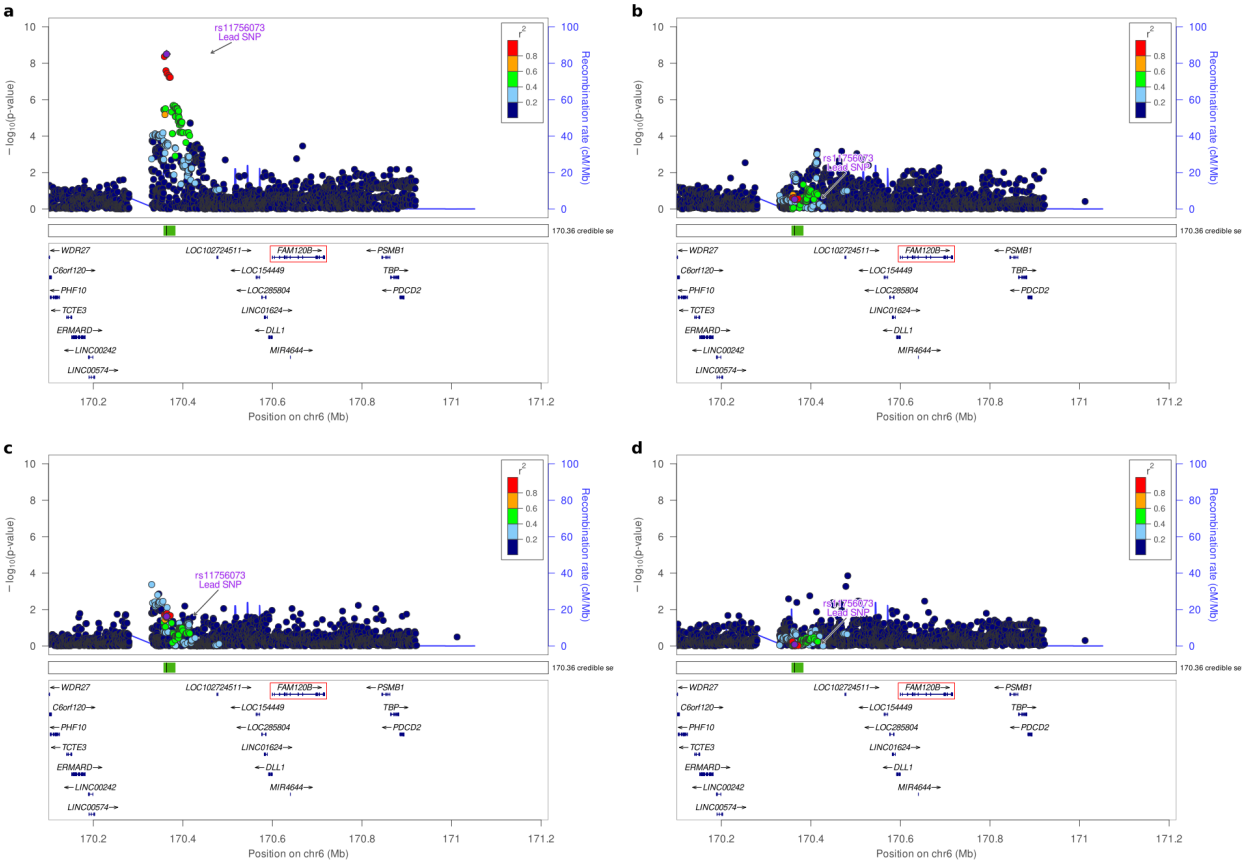
(viii) CD109/6q13



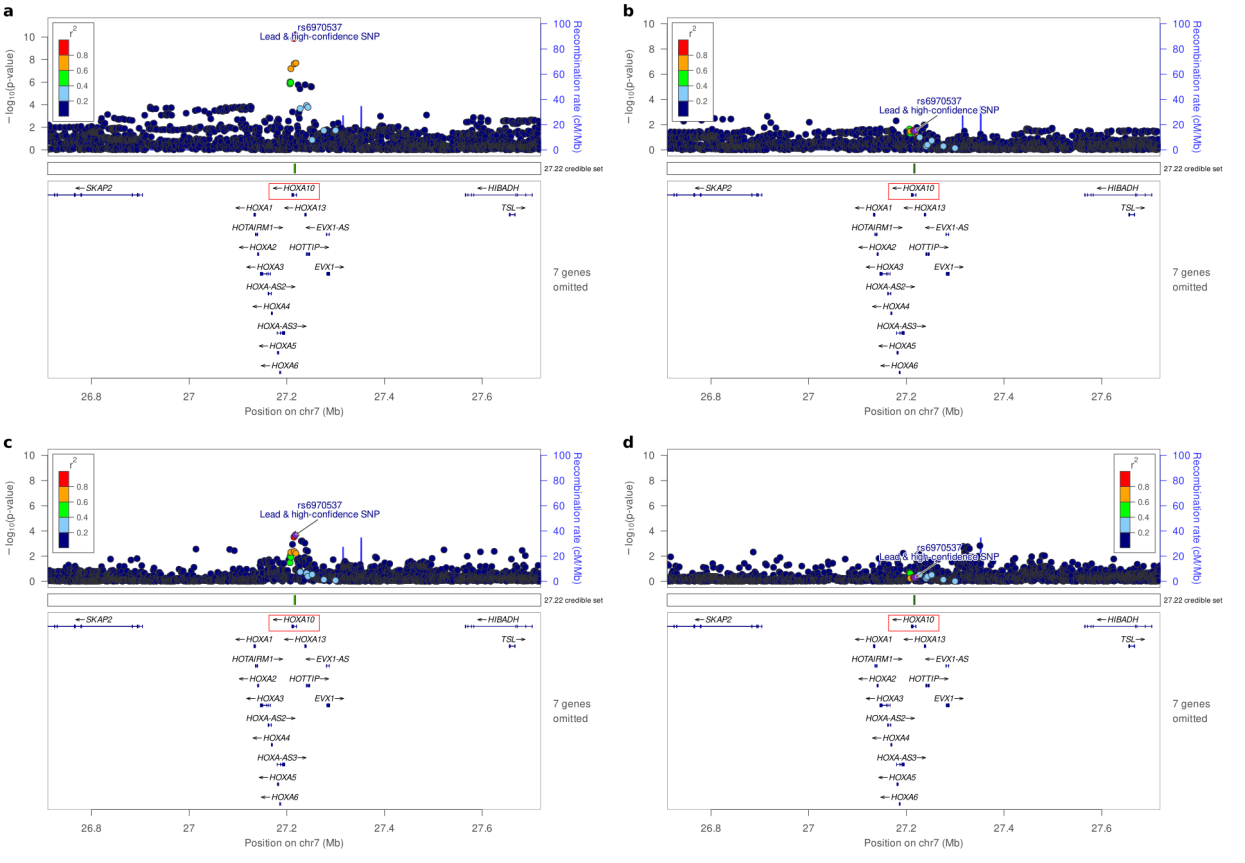
(ix) HEY2/6q22.31



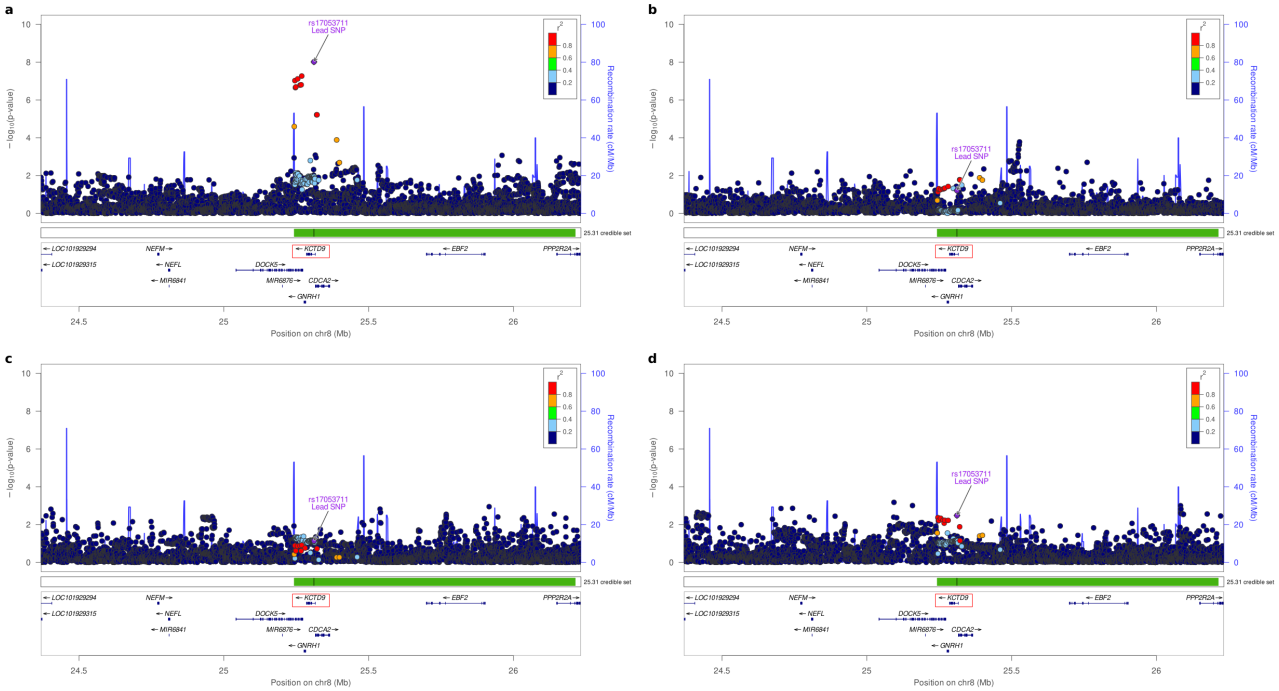
(x) FAM120B/6q27



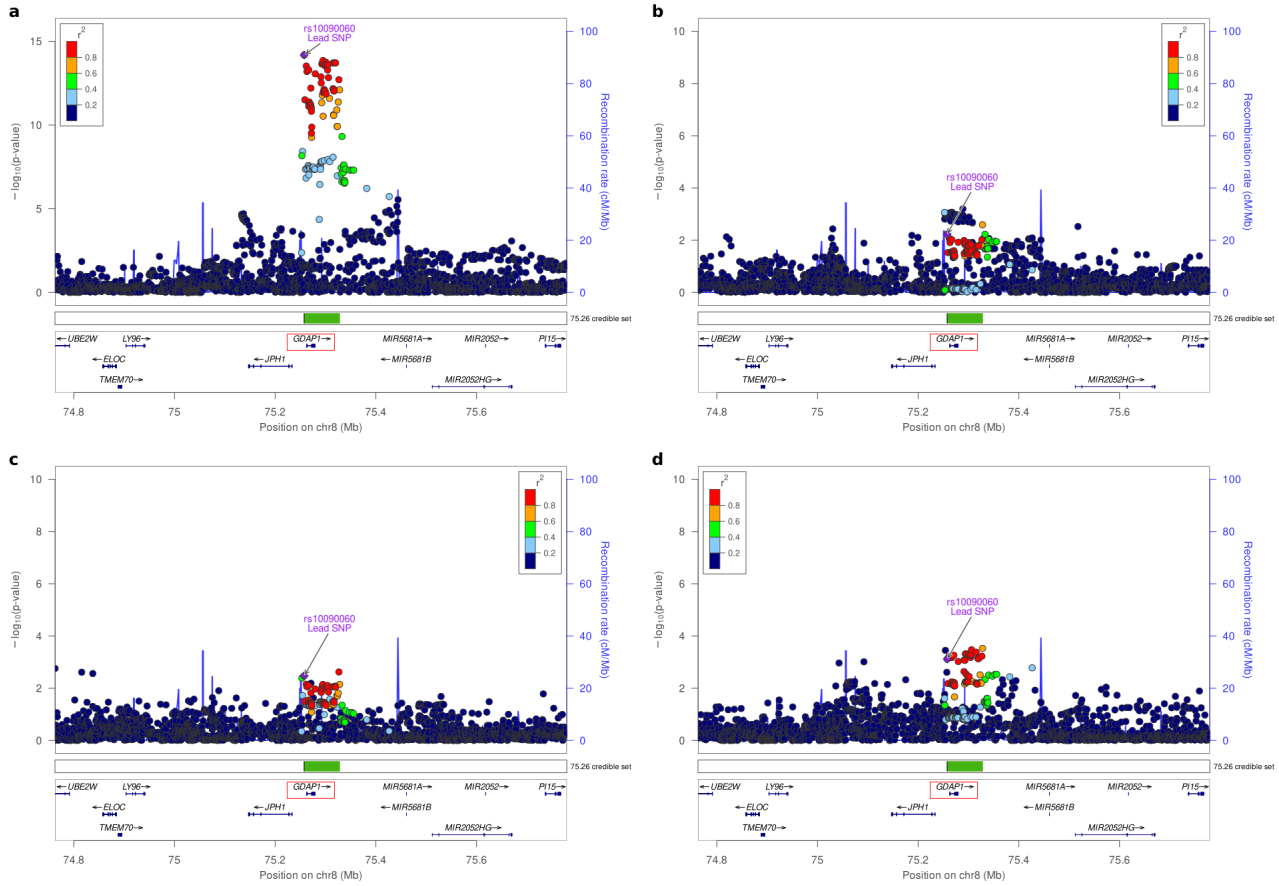
(xi) HOXA10/7p15.2



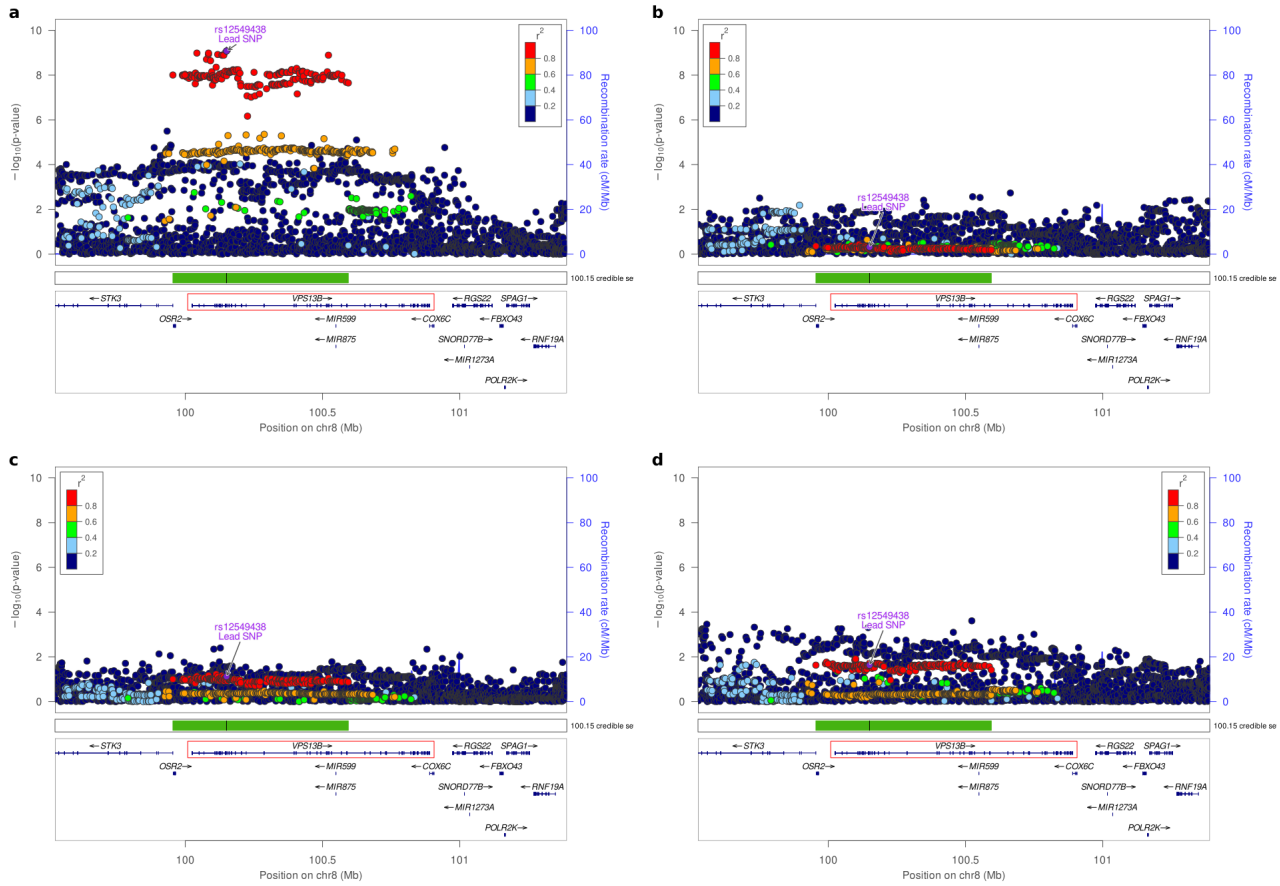
(xii) KCTD9/8p21.2



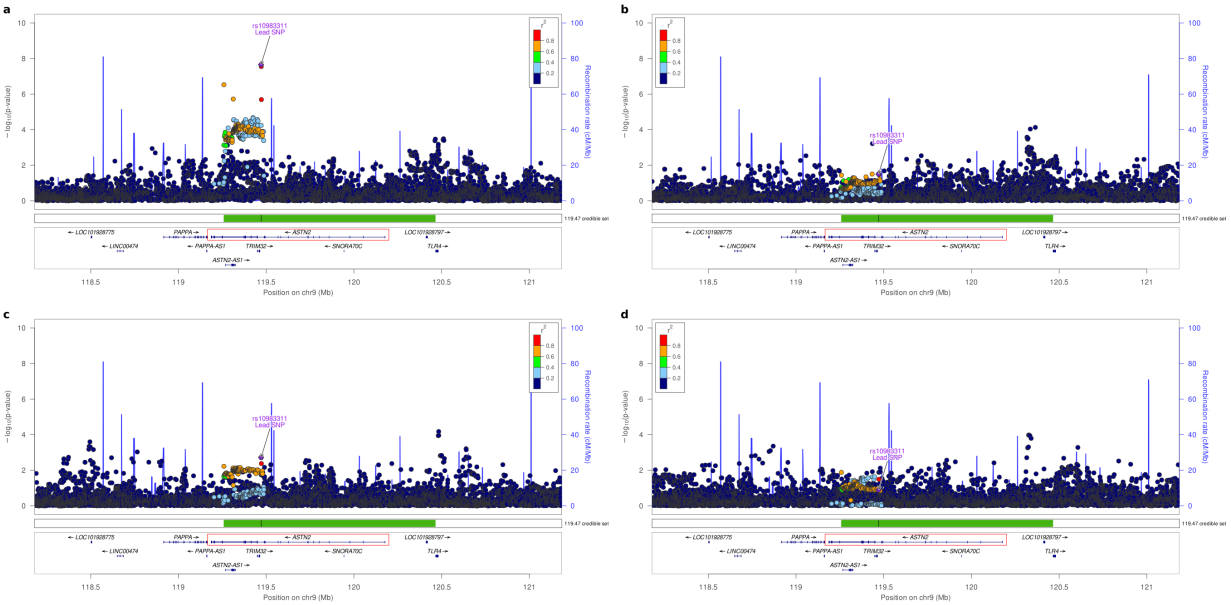
(xiii) GDAP1/8q22.2



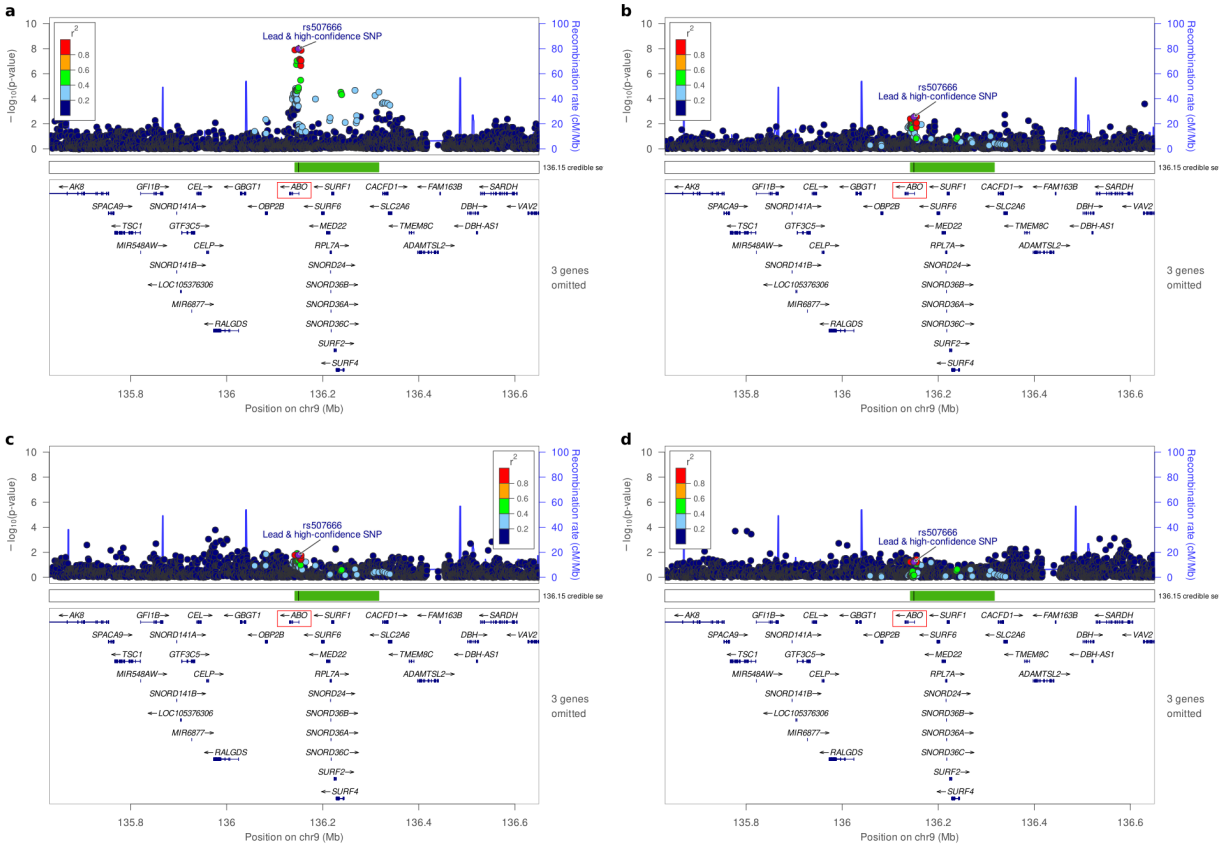
(xiv) VPS13B/8q22.2



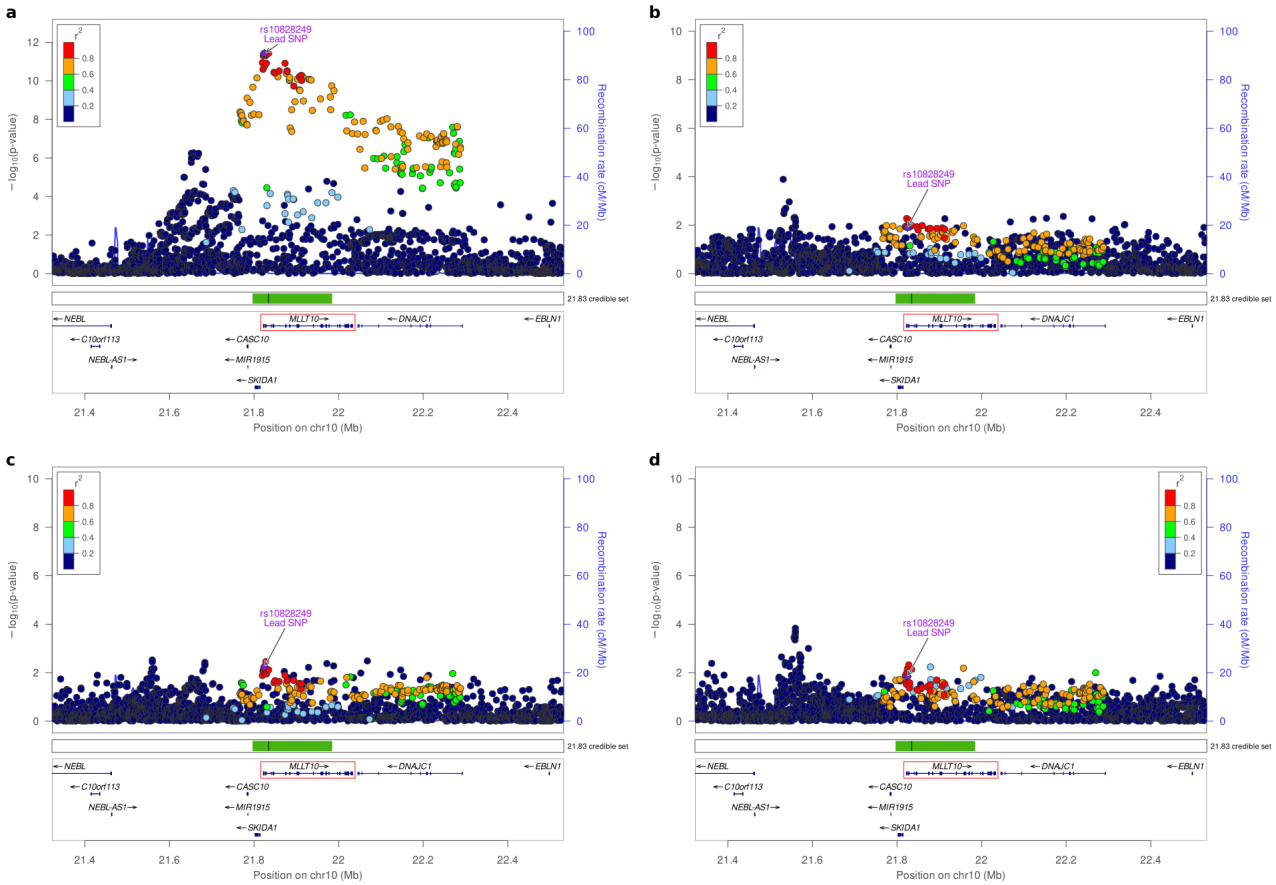
(xv) ASTN2/9q33.1



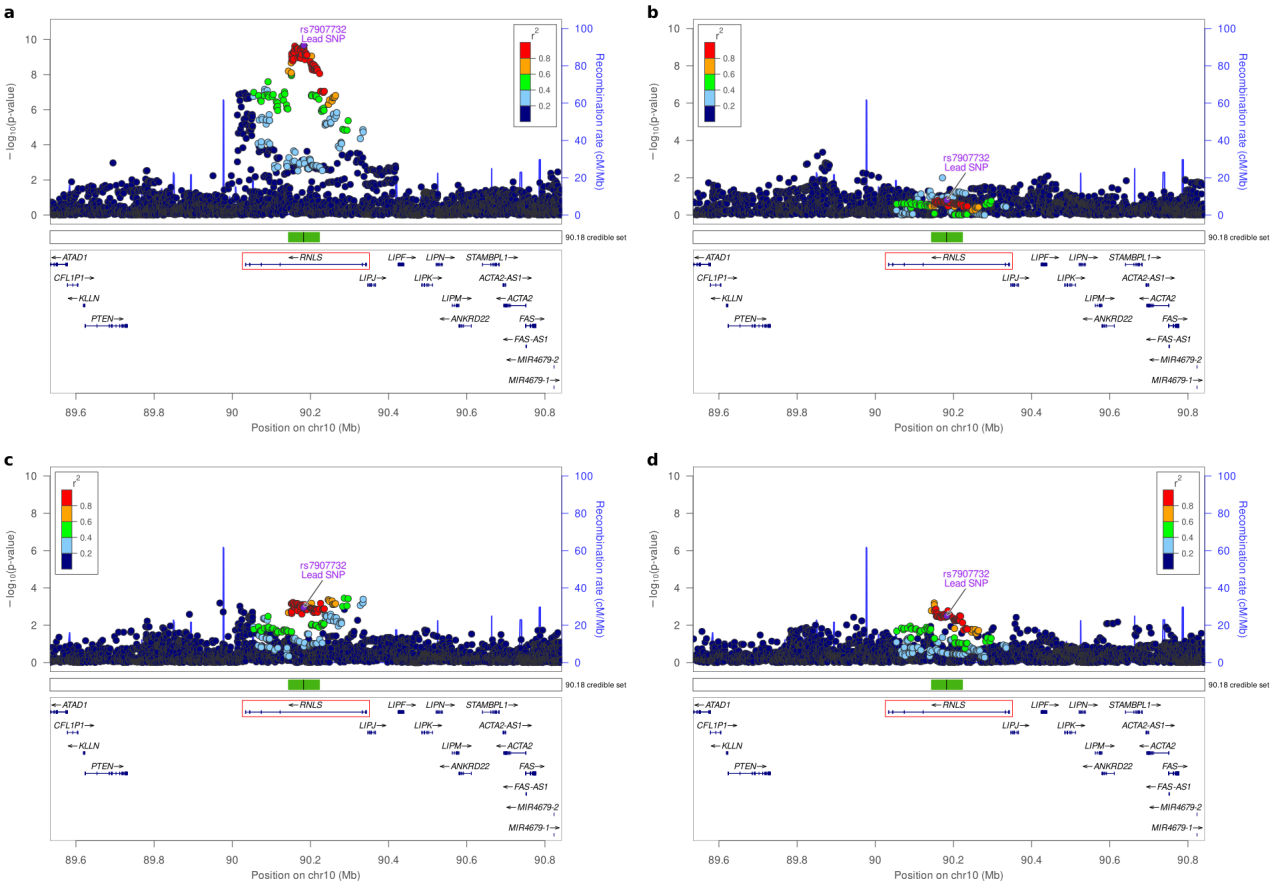
(xvi) ABO/9q34.2



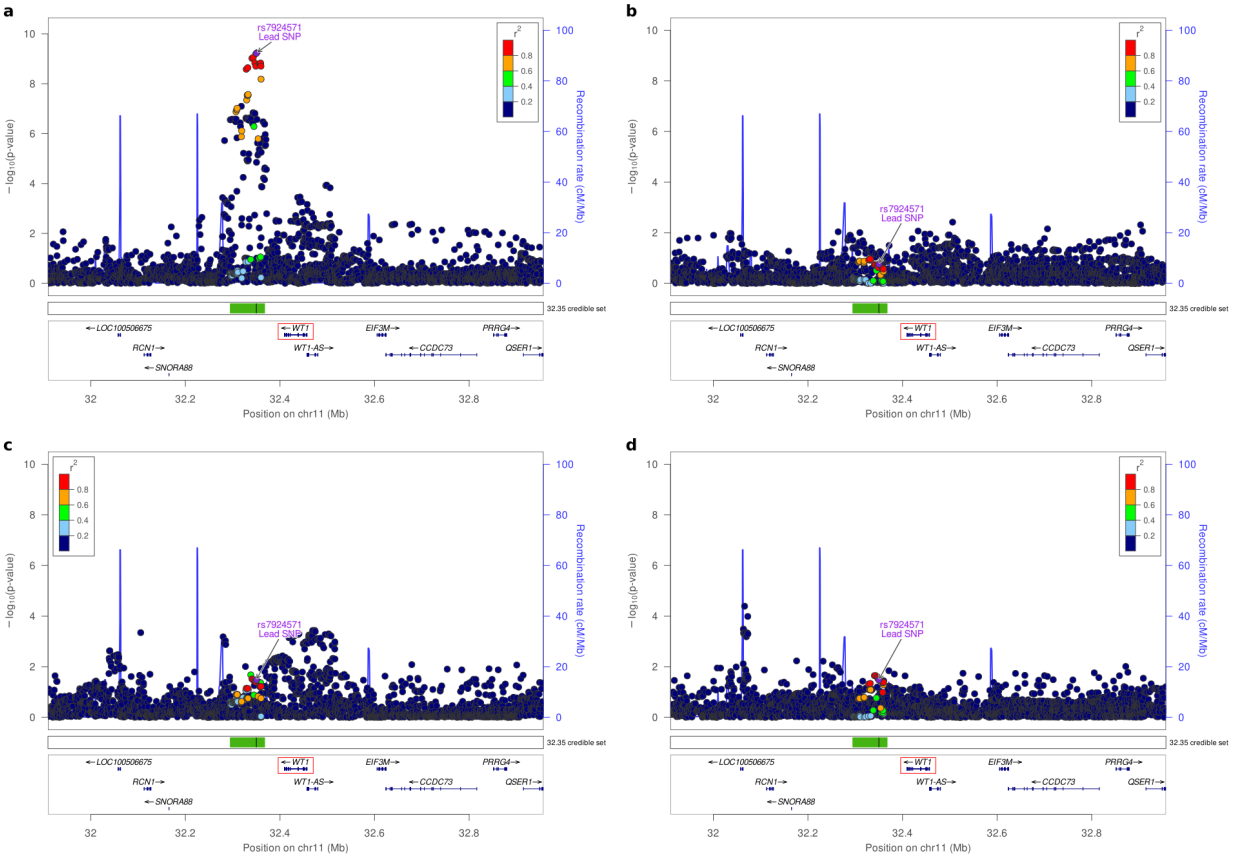
(xvii) MLLT10/10p12.31



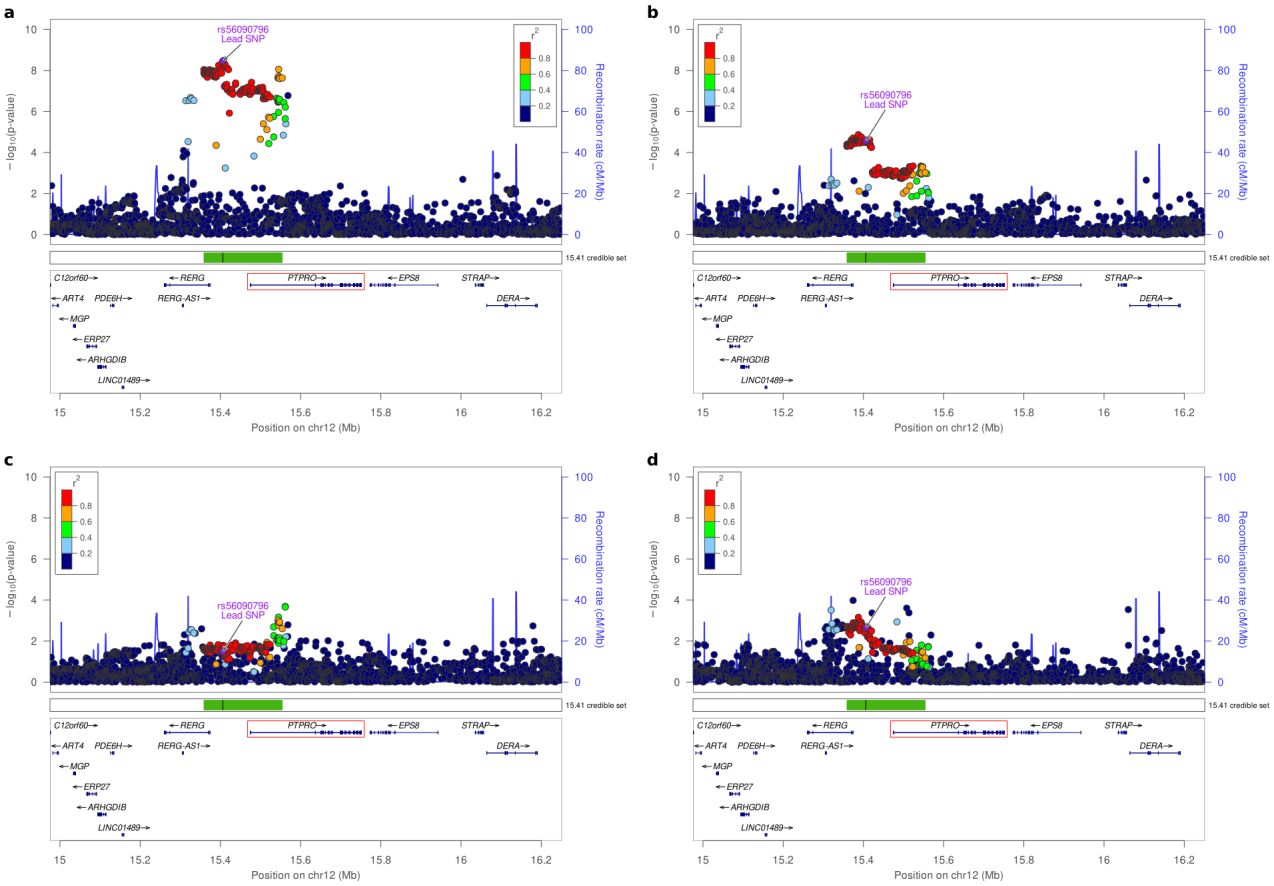
(xviii) RNLS/10q23.31



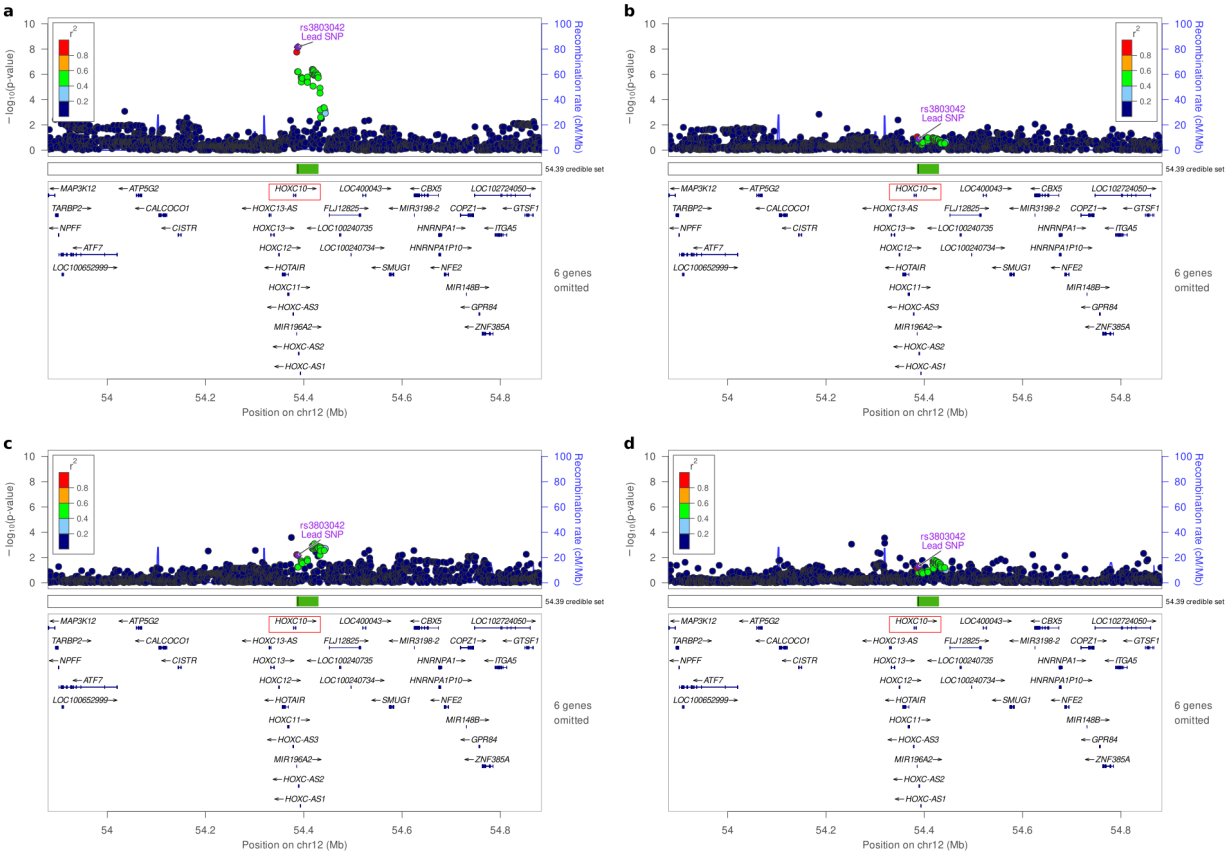
(xix) WT1/11p14.1



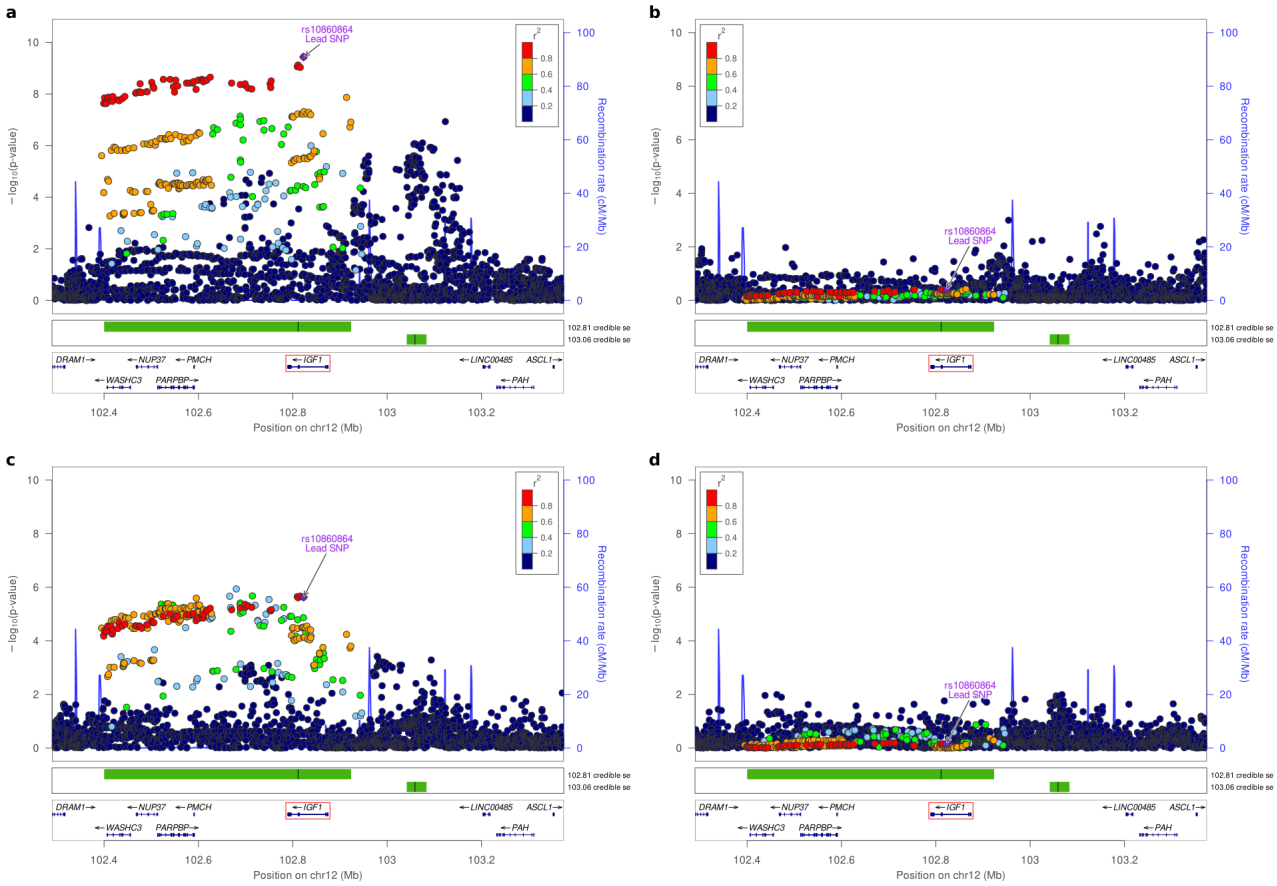
(xx) PTPRO/12p12.3



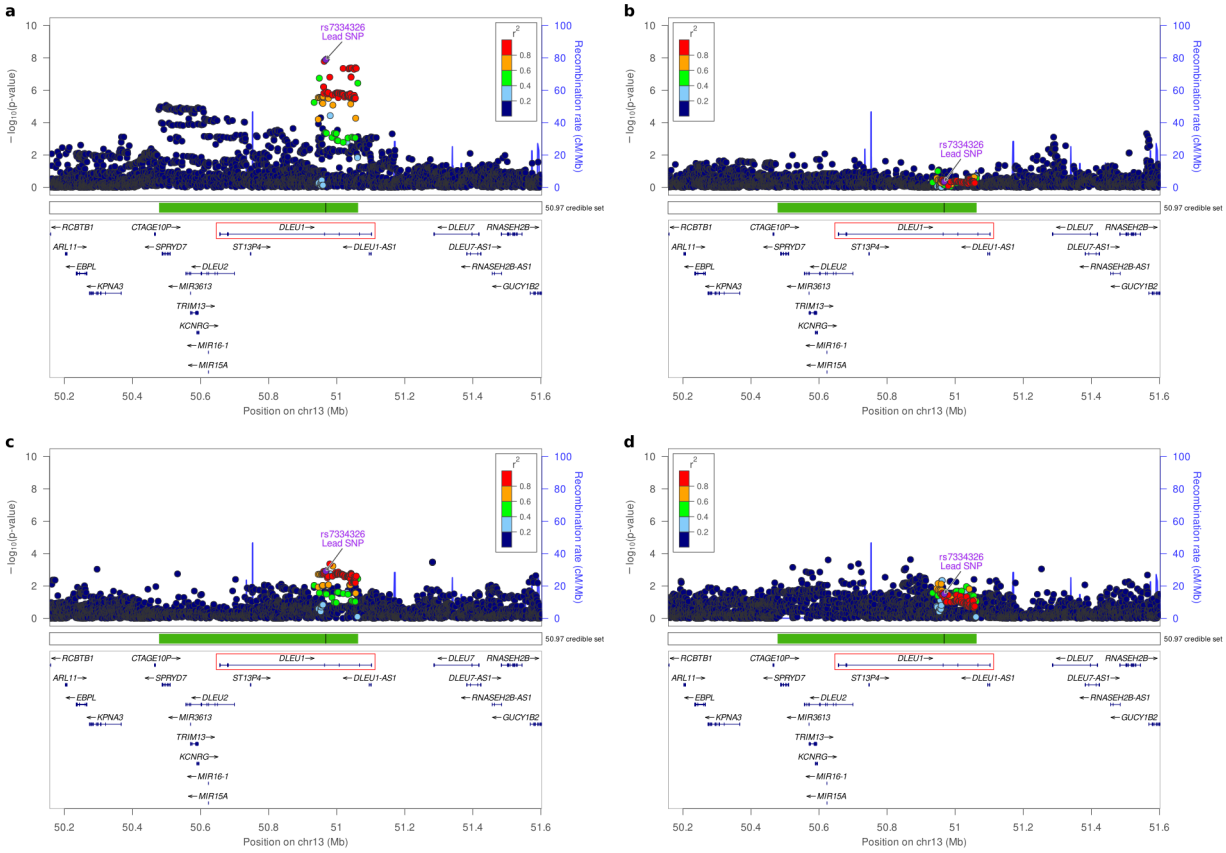
(xxi) HOXC10/12p13.13



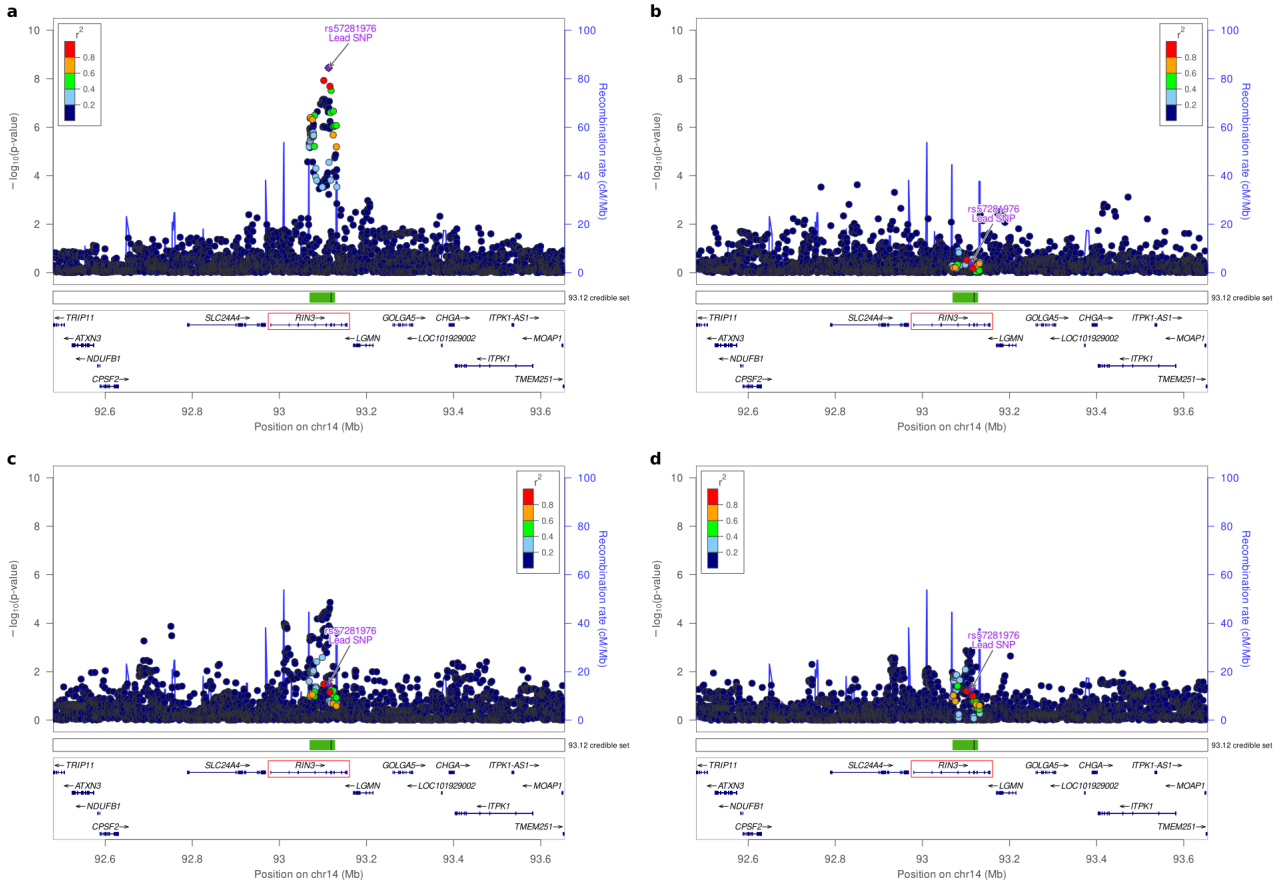
(xxii) IGF1/12q23.2



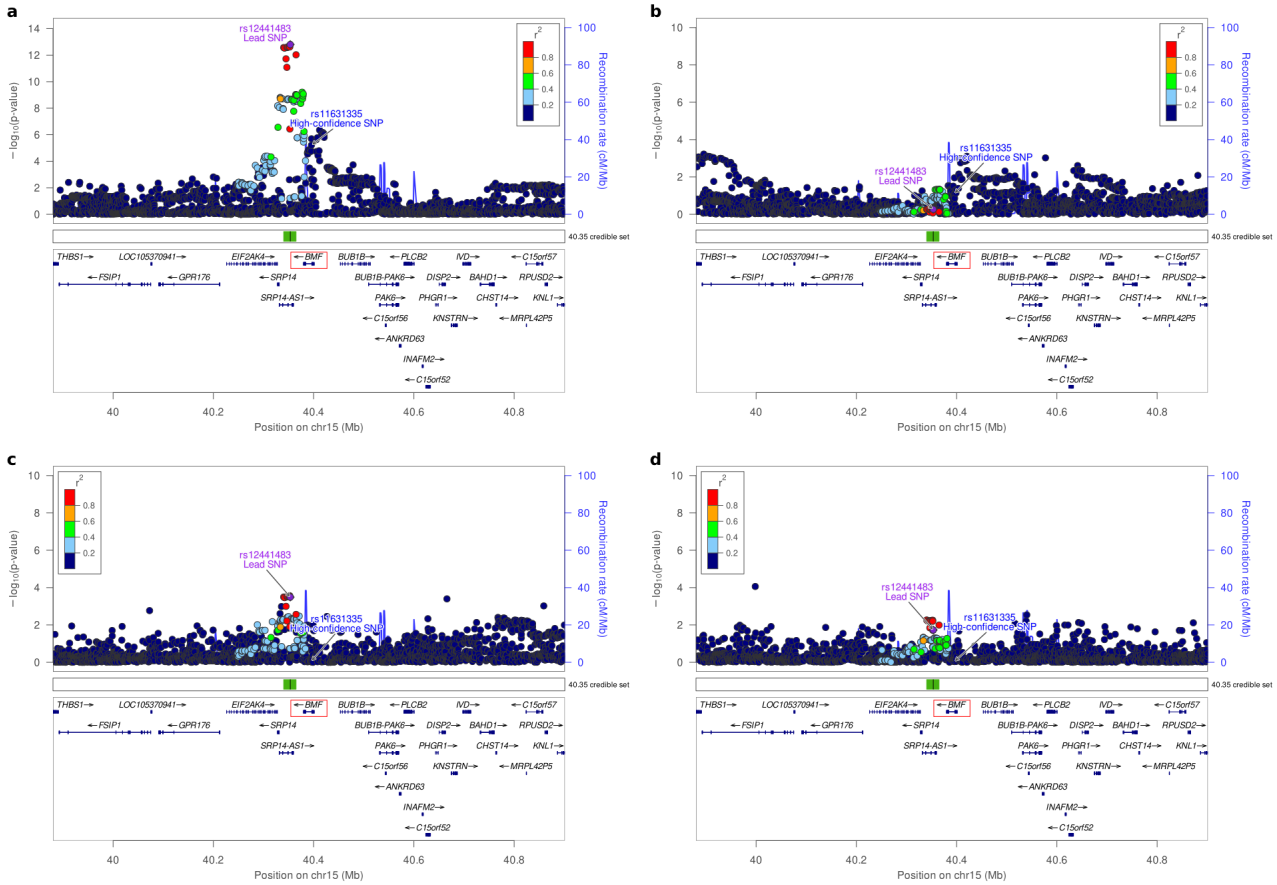
(xxiii) DLEU1/13q14.2



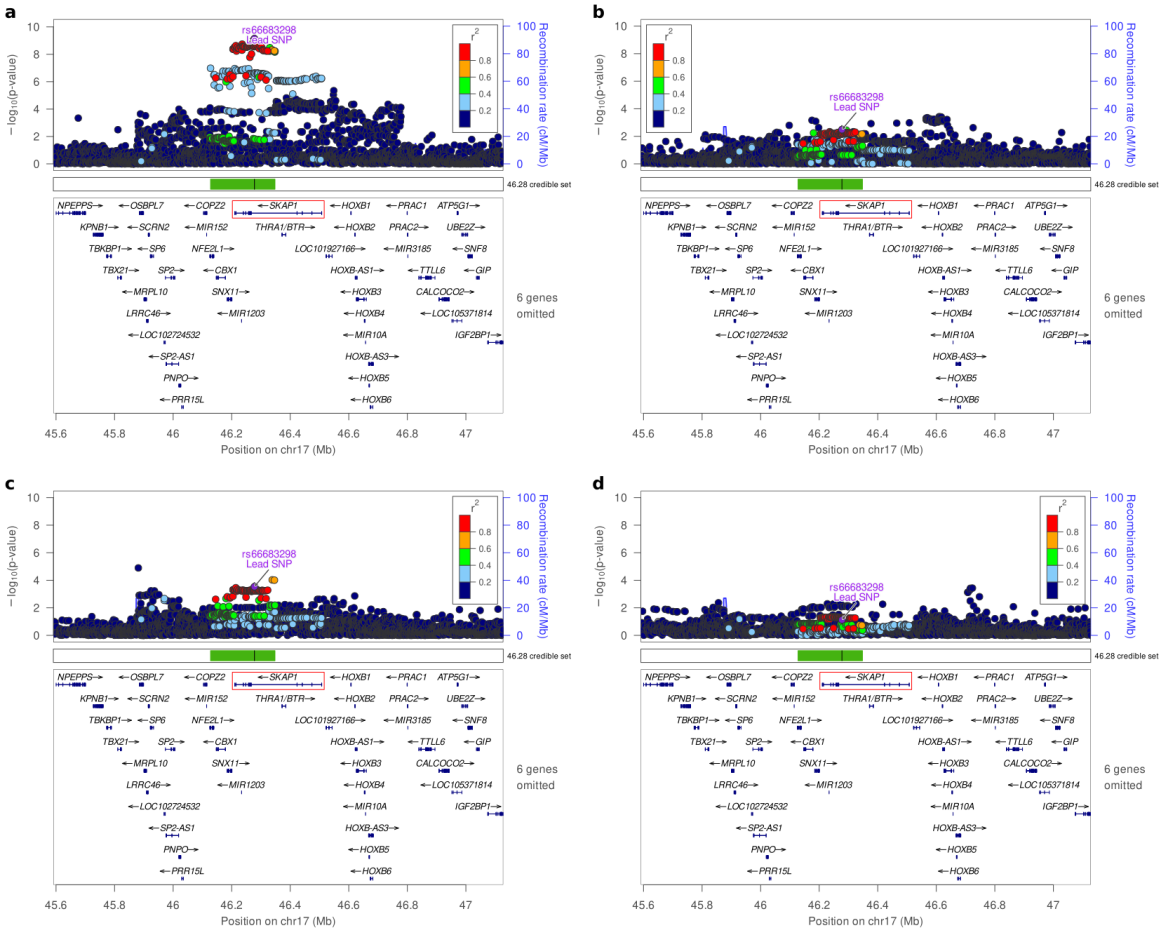
(xxiv) RIN3/14q32.12



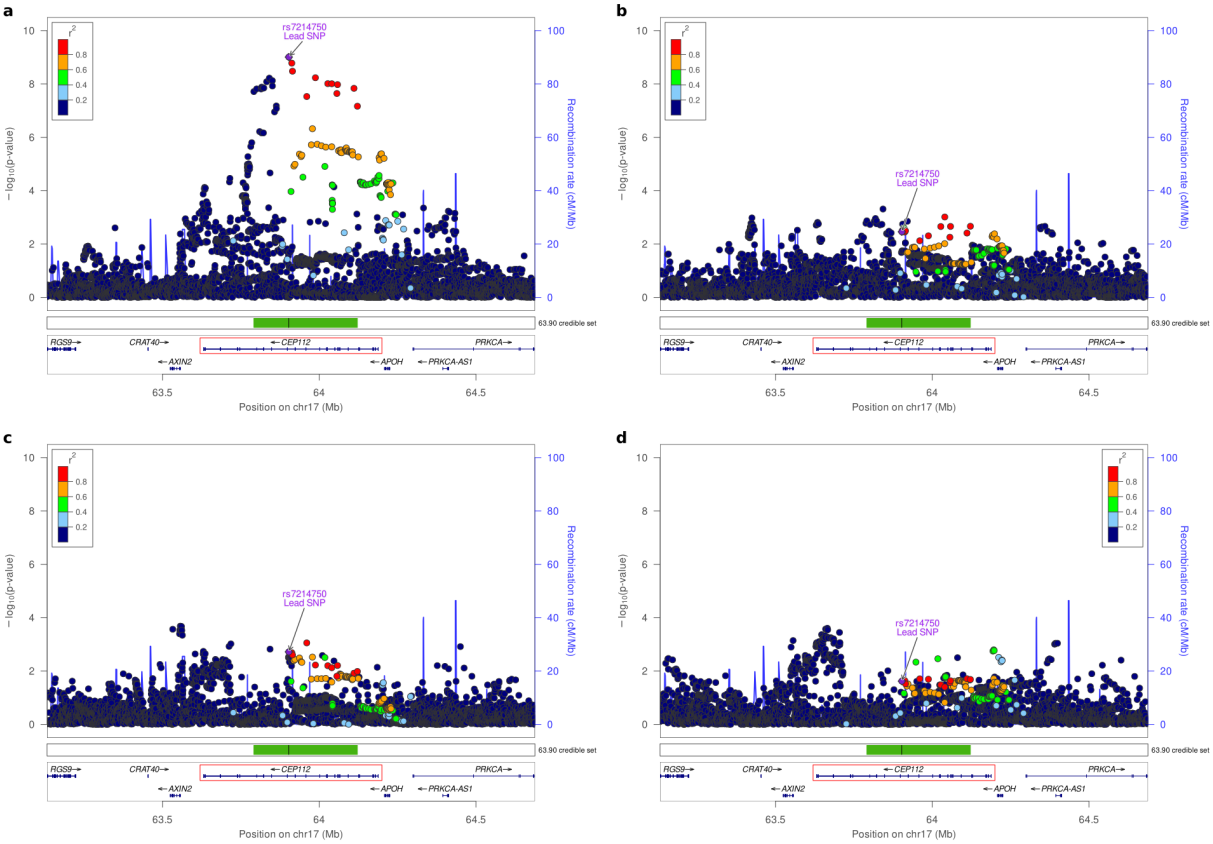
(xxv) SRP14-AS1/14q32.12



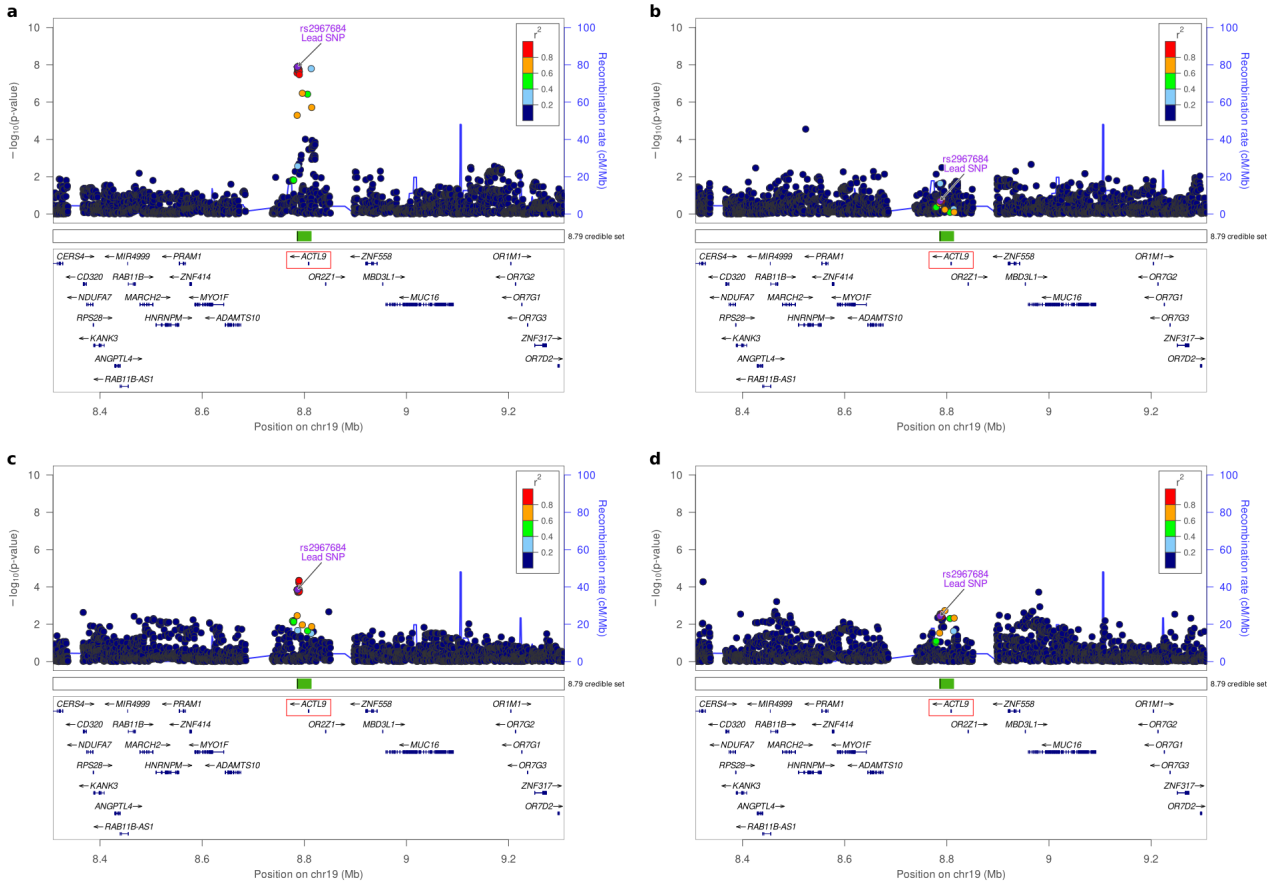
(xxvi) SKAP1/17q21.32



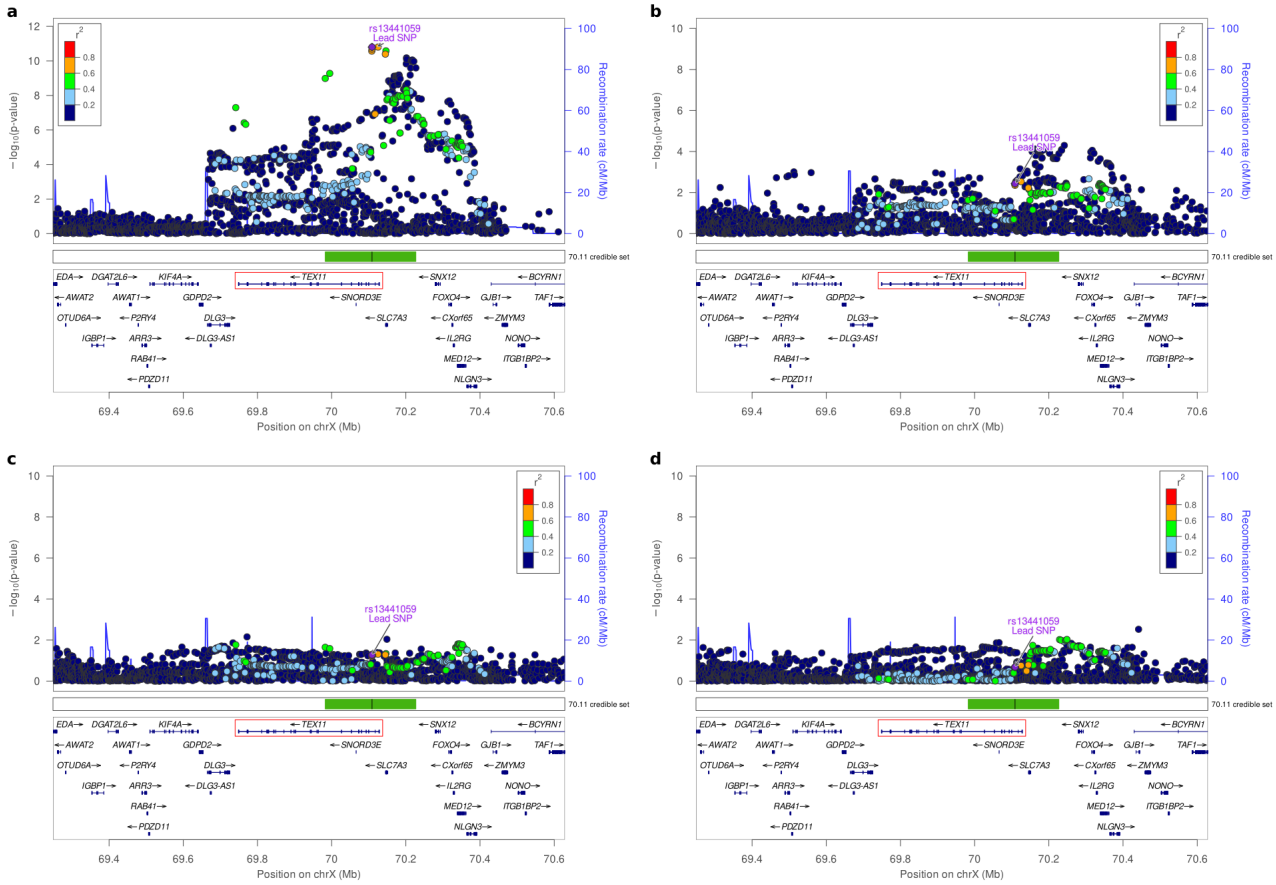
(xxvii) CEP112/17q24.1



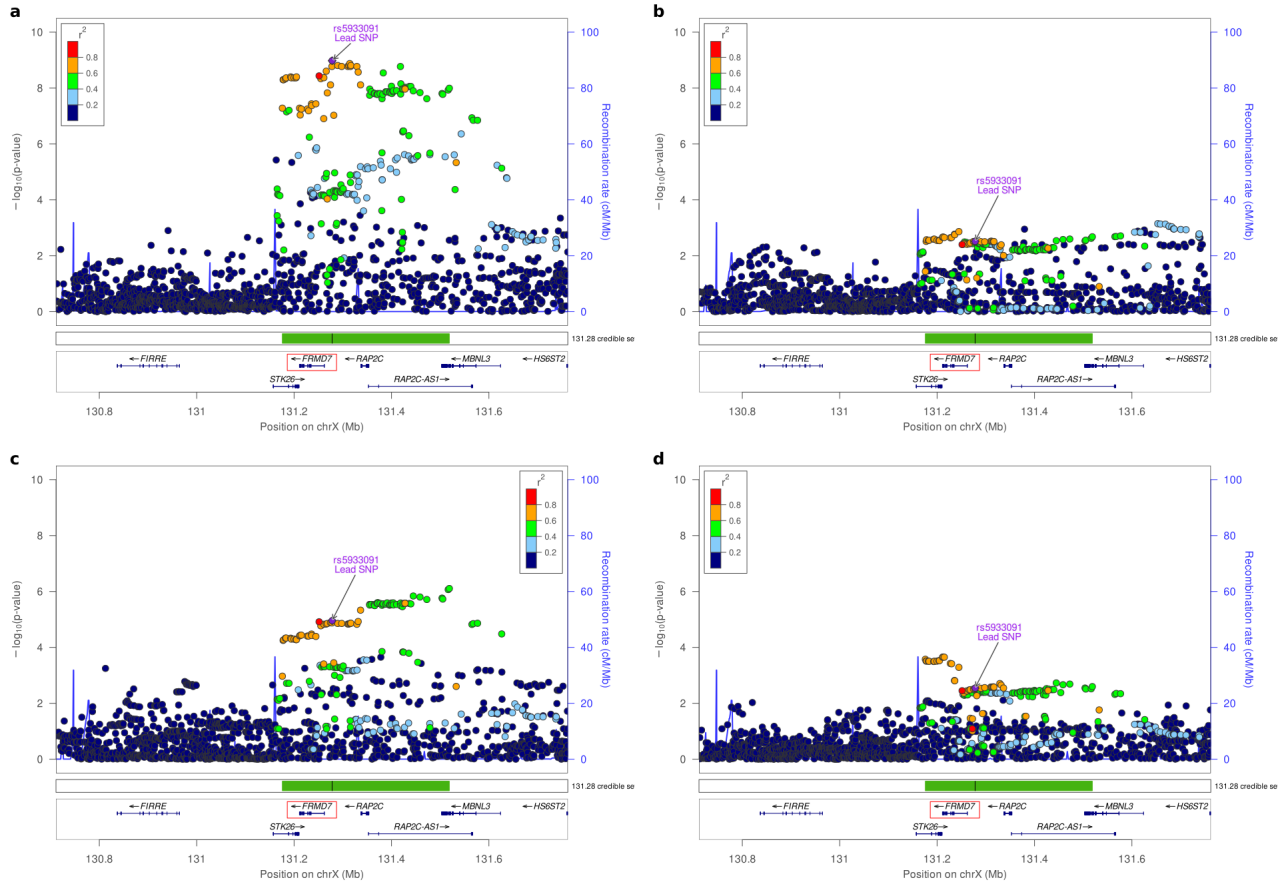
(xxviii) ACTL9/19p13.2



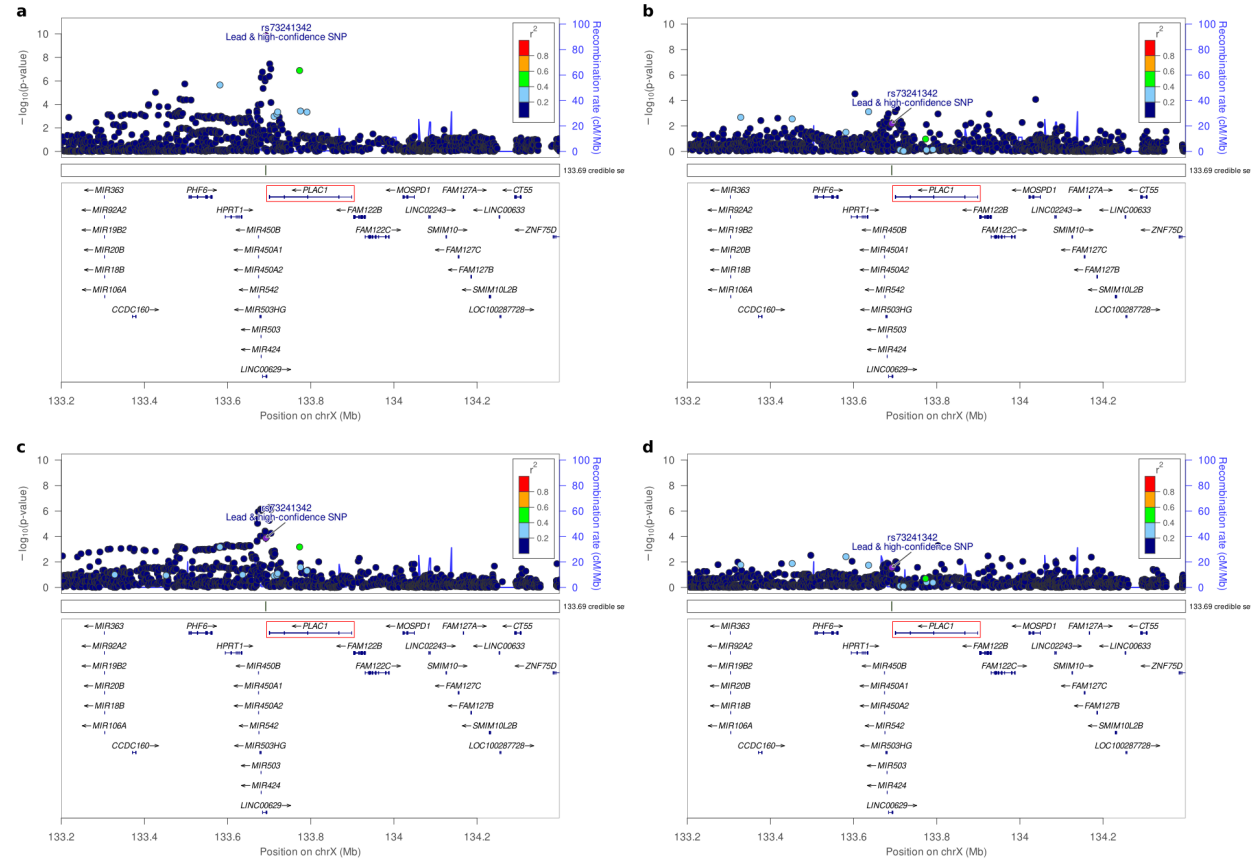
(xxix) TEX11/Xq13.1



(xxx) FRMD7/Xq26.2

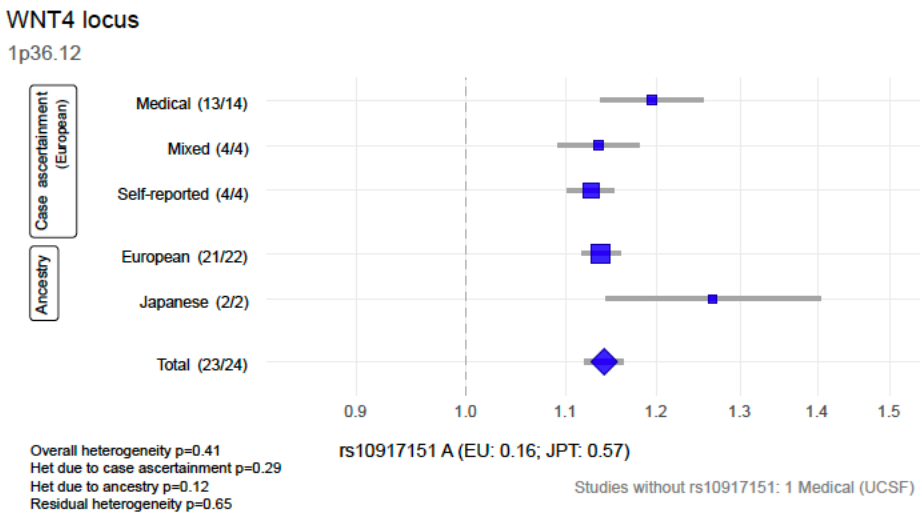


(xxxi) *LINC00629*/Xq26.3

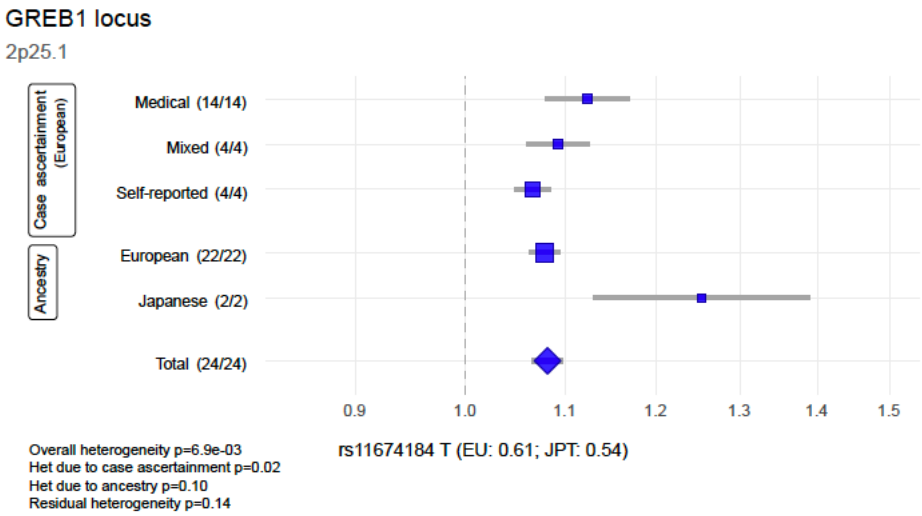


Supplementary Figure 3. Forest-plots for the lead SNPs of 11 previously established genome-wide significant endometriosis loci: a. WNT4/1p36.12, b. GREB1/2p25.1, c. ETAA1/2p14, d. KDR/4q12, e. ID4/6p22.3, f. SYNE1/6q25.1, g. 7p15.2/7p15.2, h. 7p12.3/7p12.3, i. CDKN2-BAS1/9p21.3, j. FSHB/11p14.1 k. VEZT/12q22. The rsid and the effective allele along with effective allele frequency are given under each plot. The x-axis displays the association log odds-ratio scale, and meta-analysis results by three different case ascertainment categories including medical, missed and self-reported and two different ancestries including European and Japanese ancestry results.

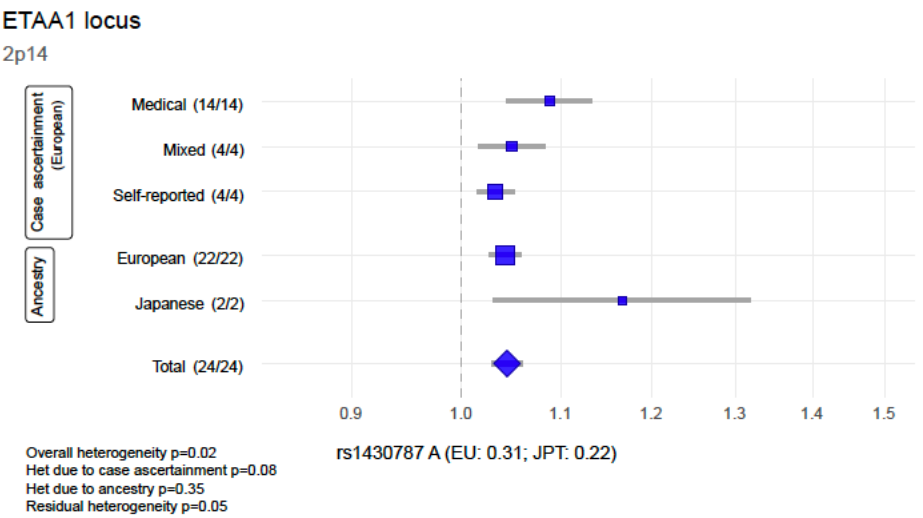
a.



b.



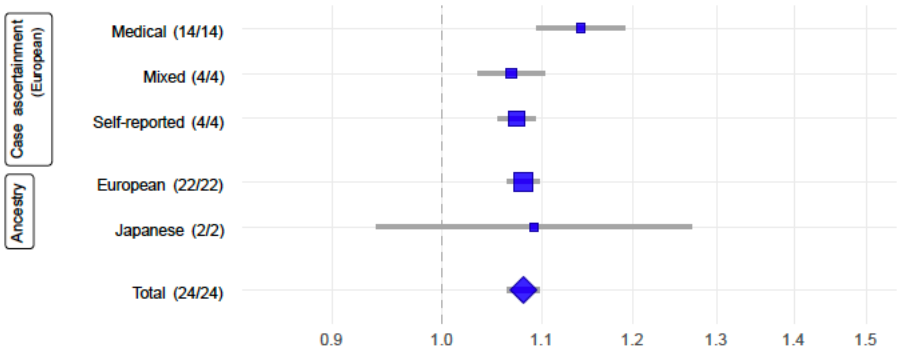
c.



d.

KDR locus

4q12



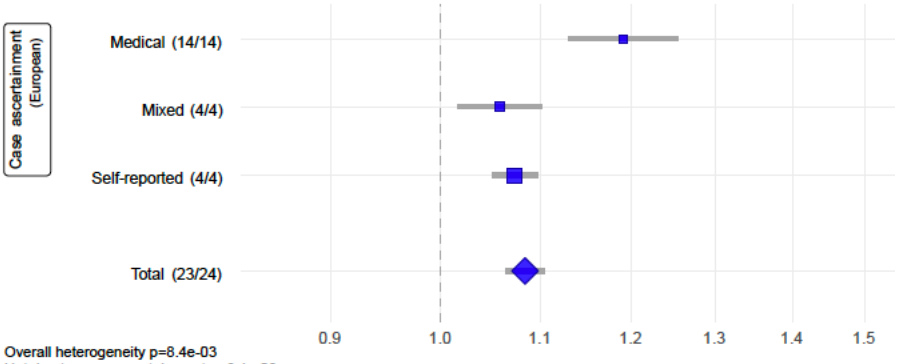
Overall heterogeneity p=0.04
Het due to case ascertainment p=0.17
Het due to ancestry p=0.84
Residual heterogeneity p=0.04

rs1903068 A (EU: 0.68; JPT: 0.88)

e.

ID4 locus

6p22.3



Overall heterogeneity p=8.4e-03
Het due to case ascertainment p=6.1e-03
Het due to ancestry p=0.31
Residual heterogeneity p=0.07

rs6456259 G (0.16)

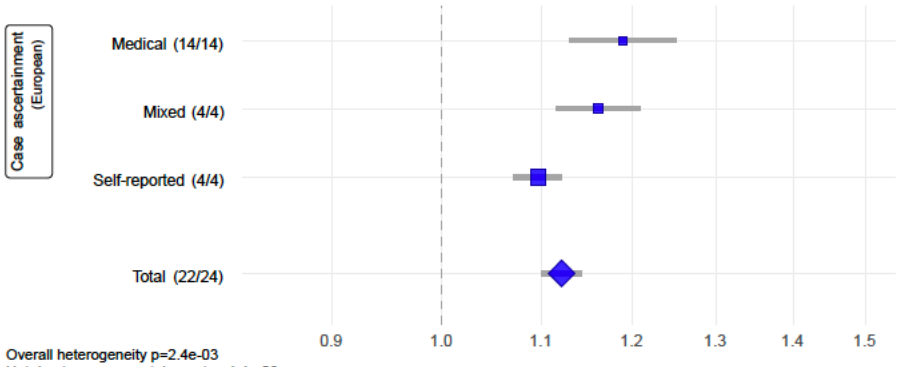
Studies without rs6456259: 1 Japanese (ADACHI)

rs6456259 was only available in one Japanese study: BBJ, OR=2.1 (0.84-5.27) and did not allow ancestry-based analyses.

f.

SYNE1 locus

6q25.1



Overall heterogeneity p=2.4e-03
Het due to case ascertainment p=1.4e-03
Het due to ancestry NA
Residual heterogeneity p=0.04

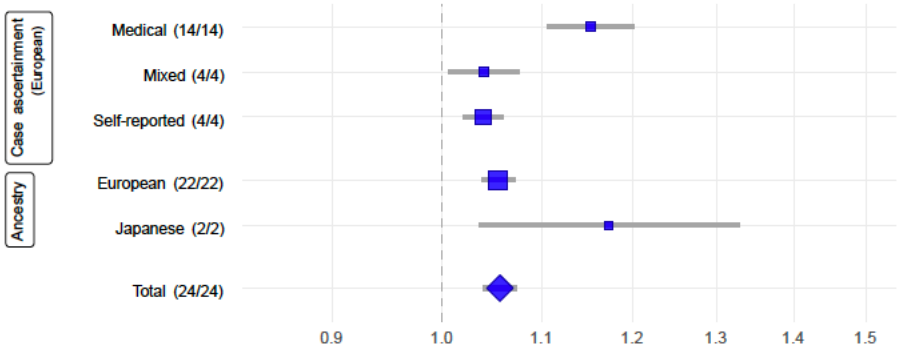
rs71575922 G (0.15)

rs71575922 was not available in Japanese studies, and therefore did not allow ancestry-based stratified analyses.

g.

7p15.2 locus

7p15.2



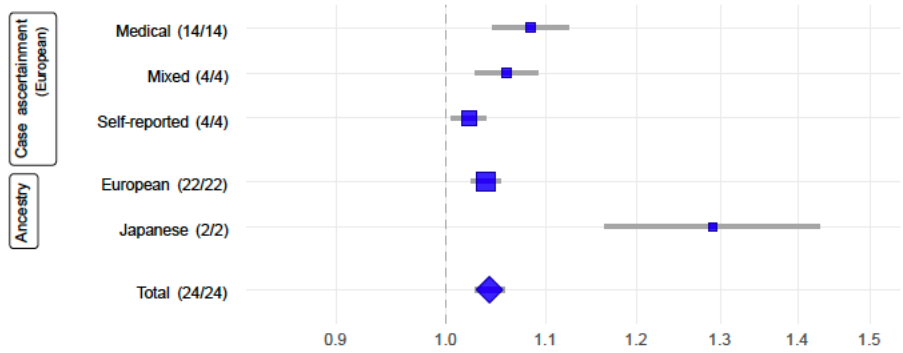
Overall heterogeneity p=8.7e-03
Het due to case ascertainment p=2.8e-04
Het due to ancestry p=0.58
Residual heterogeneity p=0.29

rs1451383 G (EU: 0.75; JPT: 0.2)

h.

7p12.3 locus

7p12.3



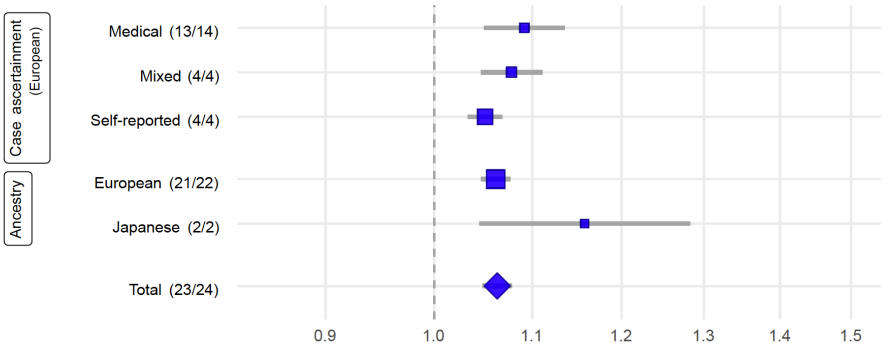
Overall heterogeneity p=4e-03
Het due to case ascertainment p=4.3e-03
Het due to ancestry p=1e-03
Residual heterogeneity p=0.68

rs55909142 C (EU: 0.59; JPT: 0.5)

i.

CDKN2B-AS1 locus

9p21.3



Overall heterogeneity p=0.02
Het due to case ascertainment p=0.08
Het due to ancestry p=0.31
Residual heterogeneity p=0.06

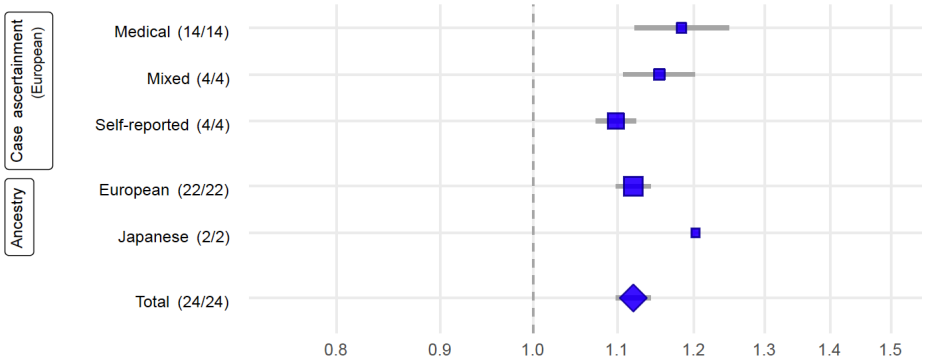
rs10122243 T (EU: 0.41; JPT: 0.39)

Studies without rs10122243: 1 Medical (GREEK)

j.

FSHB locus

11p14.1



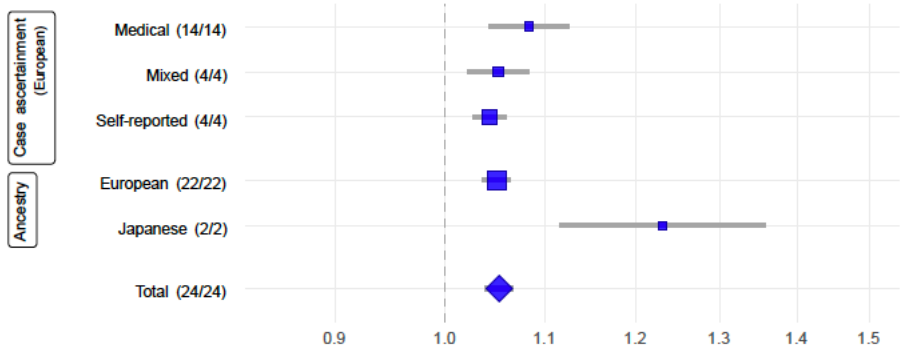
Overall heterogeneity p=0.20
Het due to case ascertainment p=4.9e-03
Het due to ancestry p=0.94
Residual heterogeneity p=0.61

rs3858429 C (EU: 0.84; JPT: 0.98)

k.

VEZT locus

12q22

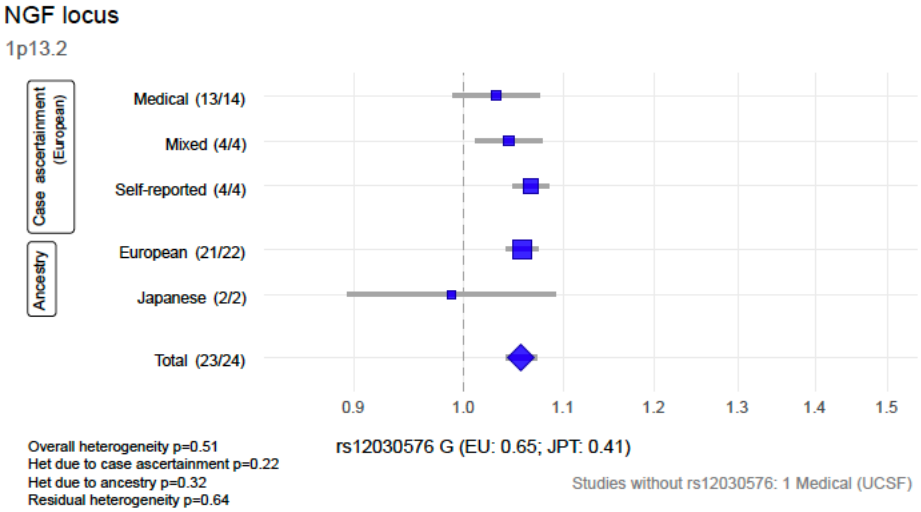


Overall heterogeneity p=0.08
Het due to case ascertainment p=0.20
Het due to ancestry p=0.02
Residual heterogeneity p=0.47

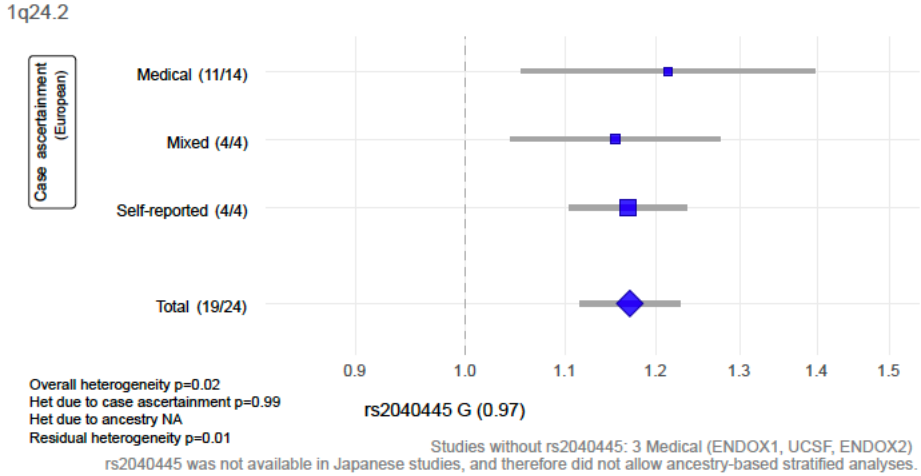
rs12320196 G (EU: 0.47; JPT: 0.49)

Supplementary Figure 4. Forest-plots for the lead SNPs of 31 novel genome-wide significant endometriosis loci: a. NGF/1p13.2, b. SLC19A2/1q24.2, c. DNM3/1q24.3, d. BMPR2/2q33.1, e. BSN/3p21.31, f. PDLIM5/4q22.3, g. EBF1/5q33.3, h. CD109/6q13, i. HEY2/6q22.31, j. FAM120B/6q27, k. HOXA10/7p15.2, l. KCTD9/8p21.2, m. GDAP1/8q22.2, n. VPS13B/8q22.2, o. ASTN2/9q33.1, p. ABO/9q34.2, q. MLLT10/10p12.31, r. RNLS/10q23.31, s. WT1/11p14.1, t. PTPRO/12p12.3, u. HOXC10/12p13.13, w. IGF1/12q23.2, v. DLEU1/13q14.2, x. RIN3/14q32.12, y. SRP14-AS1/14q32.12, z. SKAP1/17q21.32, a2. CEP112/17q24.1, b2. ACTL9/19p13.2, c2. TEX11/Xq13.1, d2. FRMD7/Xq26.2, e2. LINC00629/Xq26.3. The rsid and the effective allele along with effective allele frequency are given under each plot. The x-axis displays the association log odds-ratio scale, and meta-analysis results by three different case ascertainment categories including medical, missed and self-reported and two different ancestries including European and Japanese ancestry results.

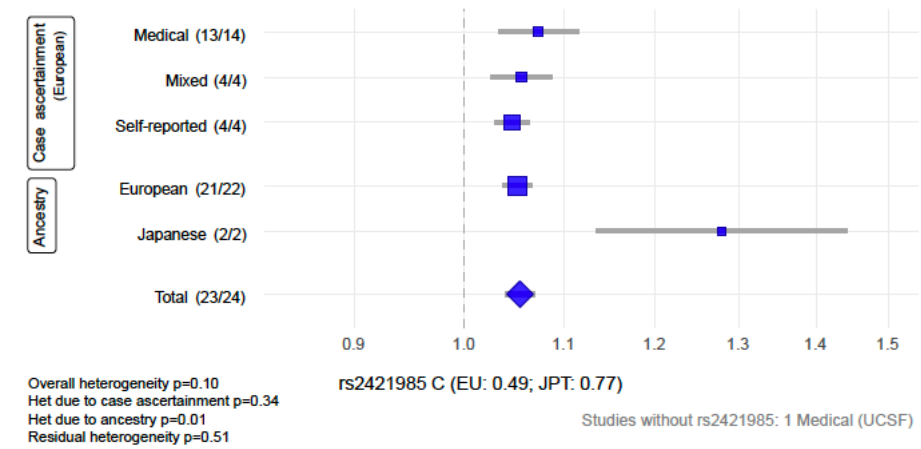
a.



b.



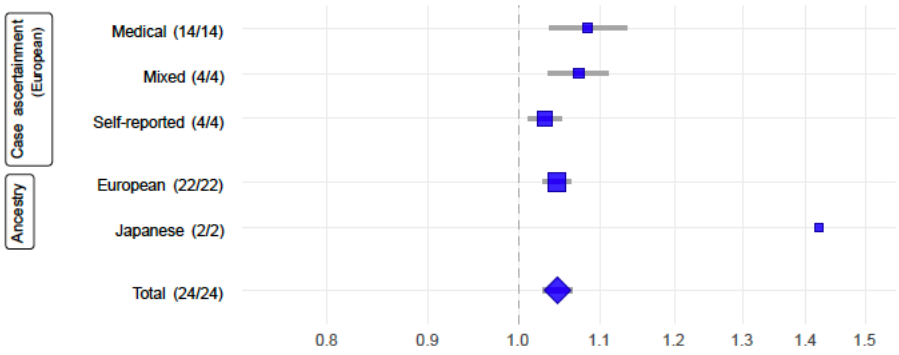
c.



d.

BMPR2 locus

2q33.1



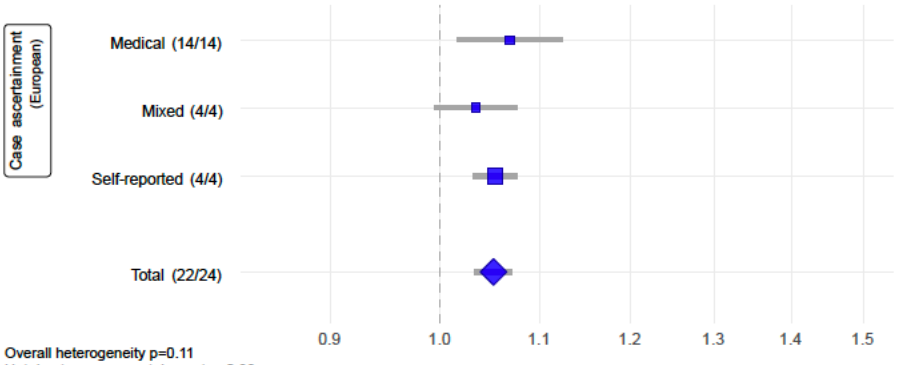
Overall heterogeneity p=0.55
Het due to case ascertainment p=0.02
Het due to ancestry p=0.04
Residual heterogeneity p=0.96

rs6435157 T (EU: 0.76; JPT: 0.97)

e.

BSN locus

3p21.31



Overall heterogeneity p=0.11
Het due to case ascertainment p=0.93
Het due to ancestry NA
Residual heterogeneity p=0.07

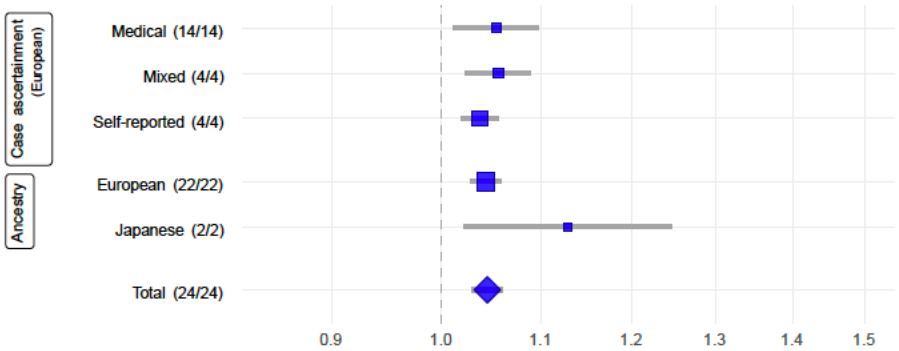
rs1352889 T (0.17)

rs1352889 was not available in Japanese studies, and therefore did not allow ancestry-based stratified analyses.

f.

PDLIM5 locus

4q22.3



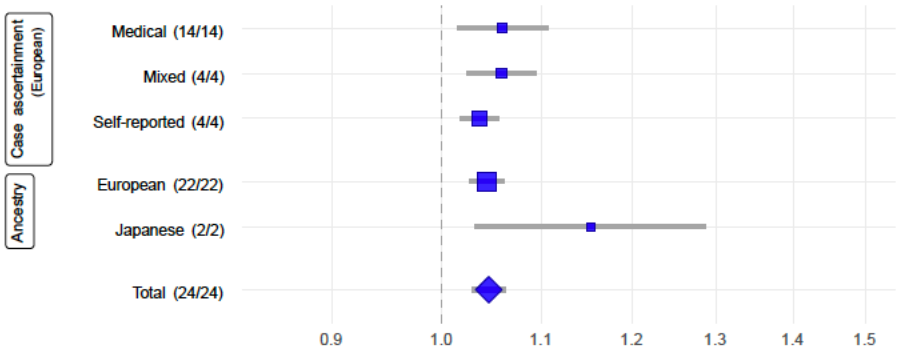
Overall heterogeneity p=0.64
Het due to case ascertainment p=0.36
Het due to ancestry p=0.11
Residual heterogeneity p=0.74

rs2510770 A (EU: 0.32; JPT: 0.49)

g.

EBF1 locus

5q33.3



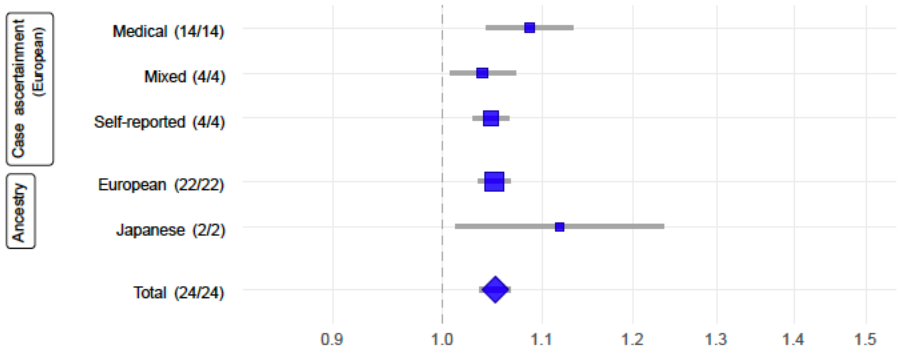
Overall heterogeneity p=0.73
Het due to case ascertainment p=0.39
Het due to ancestry p=0.10
Residual heterogeneity p=0.86

rs2946160 A (EU: 0.74; JPT: 0.69)

h.

CD109 locus

6q13



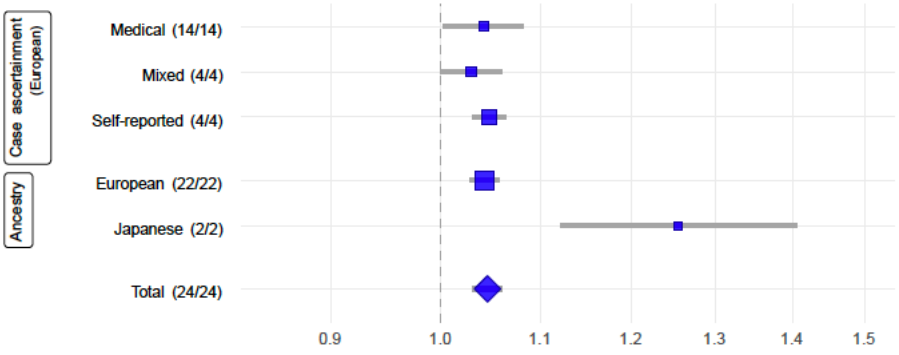
Overall heterogeneity p=0.53
Het due to case ascertainment p=0.75
Het due to ancestry p=0.36
Residual heterogeneity p=0.47

rs4540228 C (EU: 0.67; JPT: 0.58)

i.

HEY2 locus

6q22.31



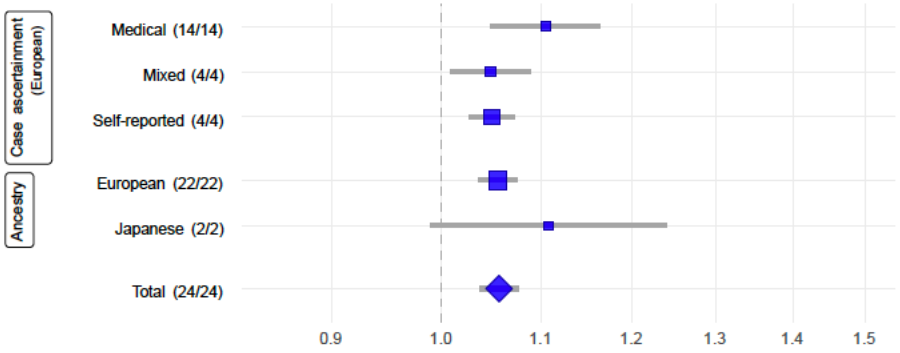
Overall heterogeneity p=0.04
Het due to case ascertainment p=0.62
Het due to ancestry p=9.4e-03
Residual heterogeneity p=0.19

rs2226158 G (EU: 0.46; JPT: 0.27)

j.

FAM120B locus

6q27



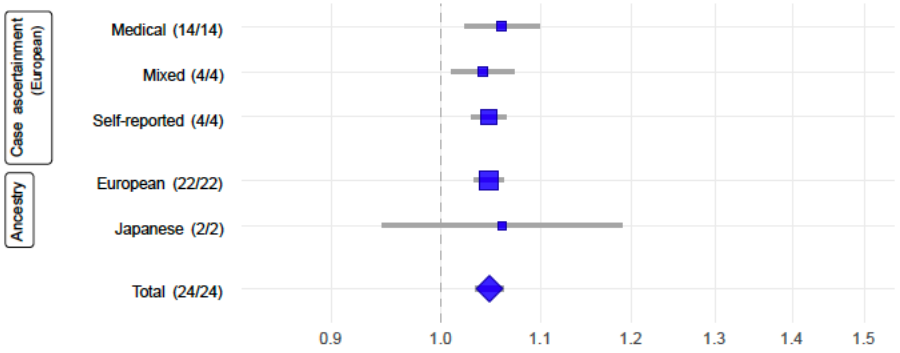
Overall heterogeneity p=0.89
Het due to case ascertainment p=0.16
Het due to ancestry p=0.82
Residual heterogeneity p=0.95

rs11756073 A (EU: 0.17; JPT: 0.26)

k.

HOXA10 locus

7p15.2



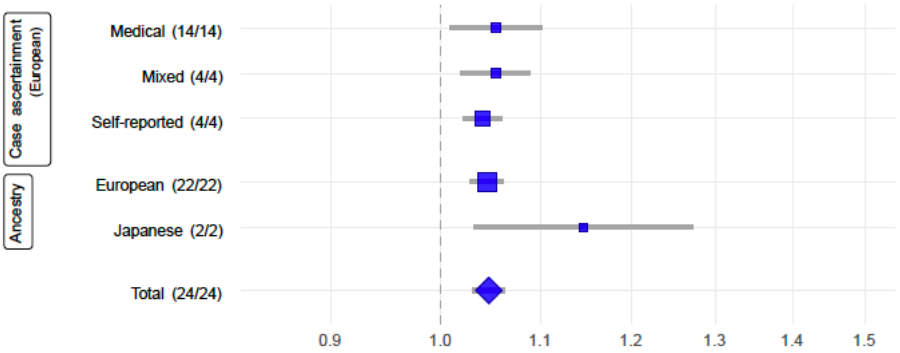
Overall heterogeneity p=0.33
Het due to case ascertainment p=0.94
Het due to ancestry p=0.84
Residual heterogeneity p=0.19

rs6970537 G (EU: 0.57; JPT: 0.75)

l.

KCTD9 locus

8p21.2



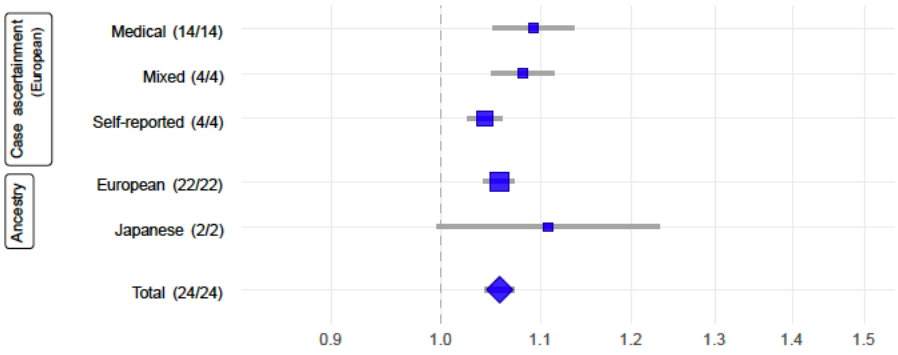
Overall heterogeneity p=0.36
Het due to case ascertainment p=0.71
Het due to ancestry p=0.17
Residual heterogeneity p=0.40

rs17053711 G (EU: 0.73; JPT: 0.54)

m.

GDAP1 locus

8q21.11



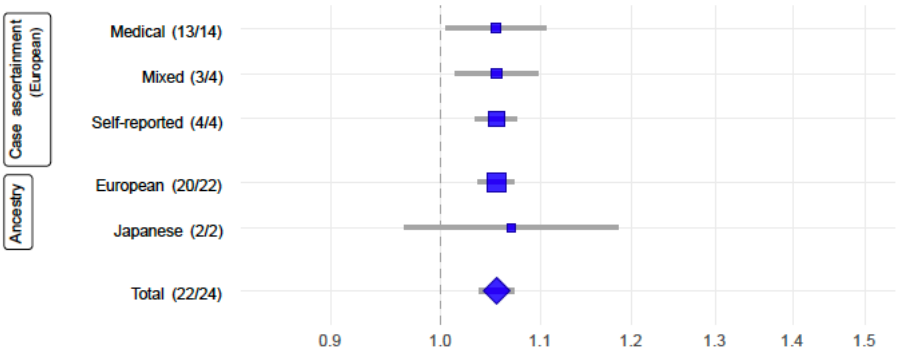
Overall heterogeneity p=0.35
Het due to case ascertainment p=0.02
Het due to ancestry p=0.63
Residual heterogeneity p=0.73

rs10090060 A (EU: 0.57; JPT: 0.68)

n.

VPS13B locus

8q22.2



Overall heterogeneity p=0.65
Het due to case ascertainment p=1.00
Het due to ancestry p=0.79
Residual heterogeneity p=0.46

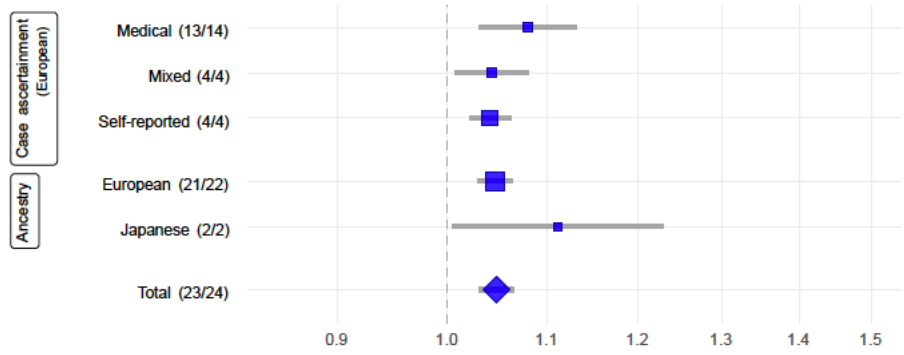
rs12549438 T (EU: 0.2; JPT: 0.39)

Studies without rs12549438: 1 Medical (GREEK), 1 Mixed (EGCUT)

o.

ASTN2 locus

9q33.1



Overall heterogeneity p=0.11
Het due to case ascertainment p=0.69
Het due to ancestry p=0.49
Residual heterogeneity p=0.08

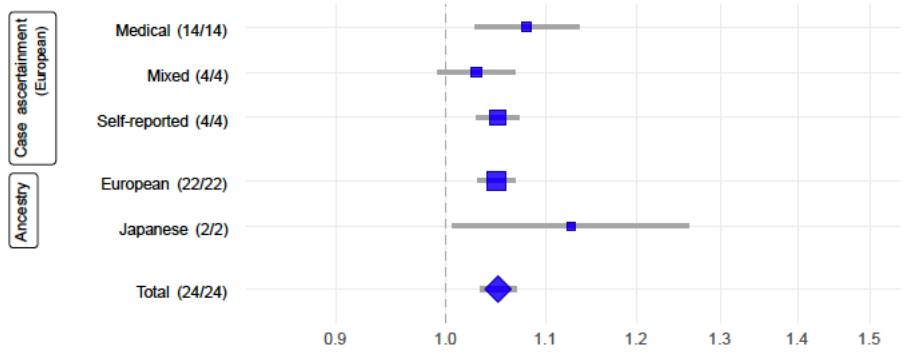
rs10983311 T (EU: 0.78; JPT: 0.61)

Studies without rs10983311: 1 Medical (GREEK)

p.

ABO locus

9q34.2



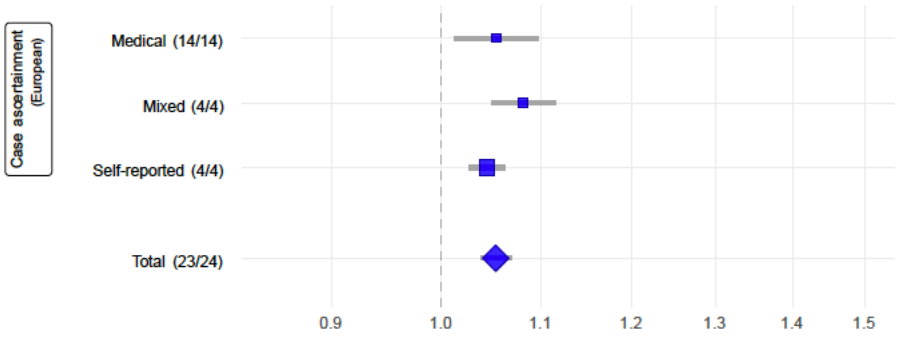
Overall heterogeneity p=0.47
Het due to case ascertainment p=0.92
Het due to ancestry p=0.30
Residual heterogeneity p=0.39

rs507666 A (EU: 0.19; JPT: 0.27)

q.

MLLT10 locus

10p12.31



rs10828249 A (0.35)

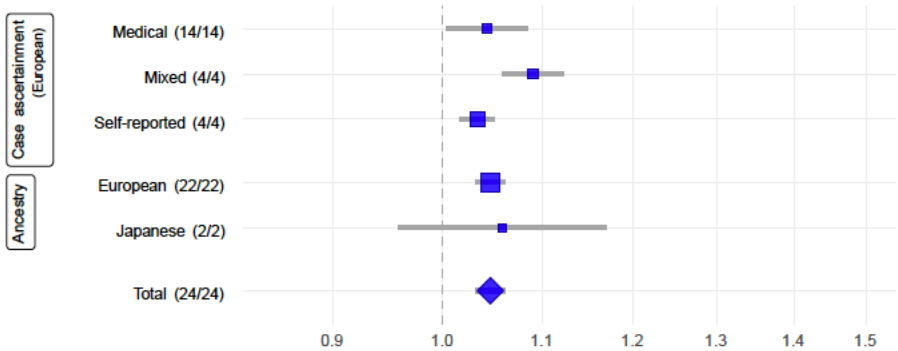
Overall heterogeneity p=0.99
Het due to case ascertainment p=0.15
Het due to ancestry p=0.23
Residual heterogeneity p=1.00

Studies without rs10828249: 1 Japanese (ADACHI)
rs10828249 was only available in one Japanese study: BBJ,
OR=1.2 (0.84-1.73) and did not allow ancestry-based analyses.

r.

RNLS locus

10q23.31



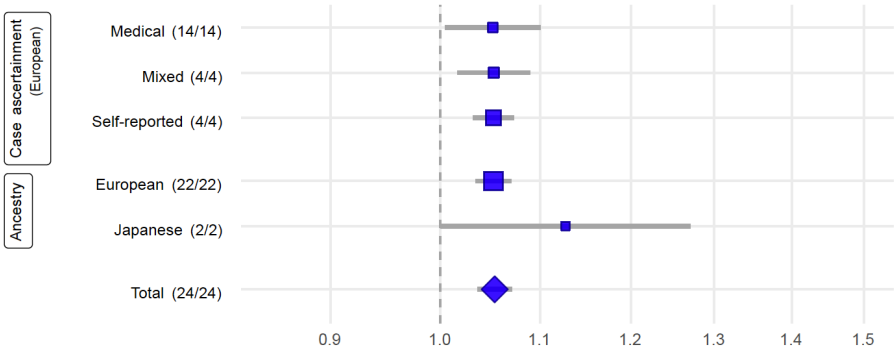
Overall heterogeneity p=0.05
Het due to case ascertainment p=0.01
Het due to ancestry p=0.94
Residual heterogeneity p=0.17

rs7907732 T (EU: 0.38; JPT: 0.51)

s.

WT1 locus

11p14.1



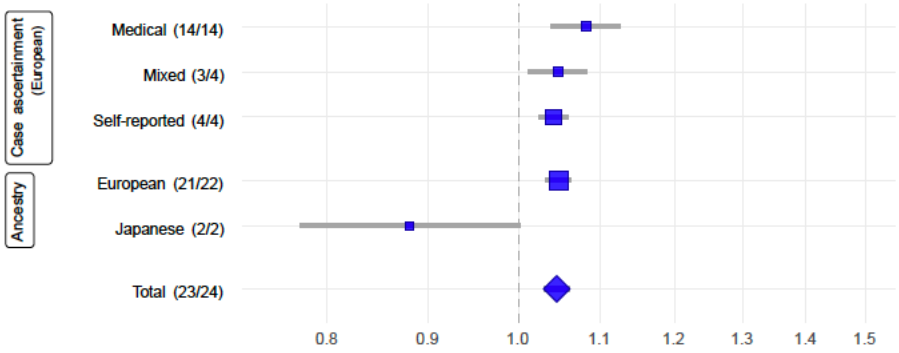
Overall heterogeneity p=0.86
Het due to case ascertainment p=0.86
Het due to ancestry p=0.21
Residual heterogeneity p=0.59

rs7924571 C (EU: 0.77; JPT: 0.23)

t.

PTPRO locus

12p12.3



Overall heterogeneity p=0.06
Het due to case ascertainment p=0.21
Het due to ancestry p=0.01
Residual heterogeneity p=0.24

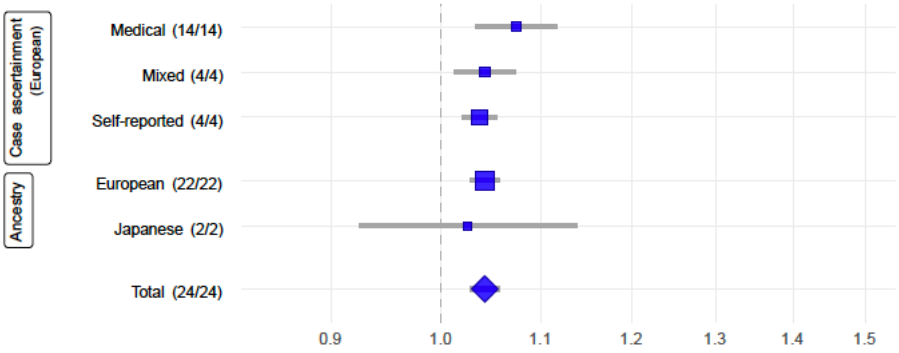
rs56090796 A (EU: 0.33; JPT: 0.18)

Studies without rs56090796: 1 Mixed (EGCUT)

u.

HOXC10 locus

12p13.13



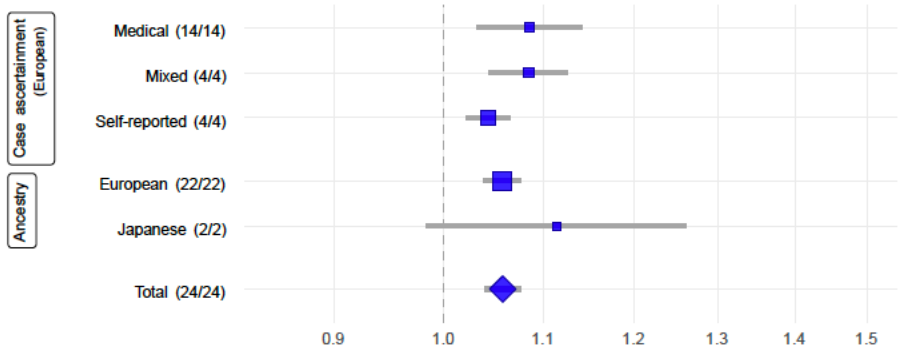
Overall heterogeneity p=0.17
Het due to case ascertainment p=0.39
Het due to ancestry p=0.59
Residual heterogeneity p=0.13

rs3803042 A (EU: 0.42; JPT: 0.55)

w.

IGF1 locus

12q23.2



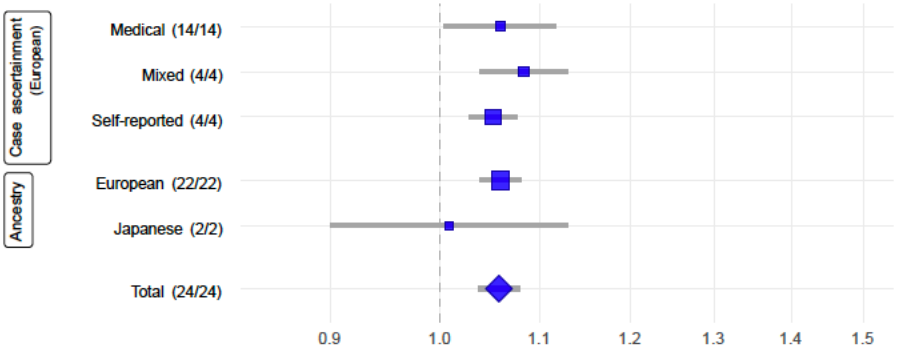
Overall heterogeneity p=0.30
Het due to case ascertainment p=0.08
Het due to ancestry p=0.65
Residual heterogeneity p=0.43

rs10860864 C (EU: 0.82; JPT: 0.8)

v.

DLEU1 locus

13q14.2



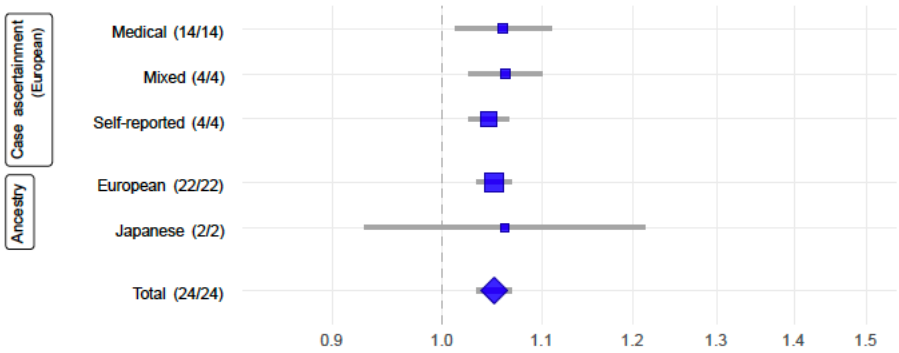
Overall heterogeneity p=0.86
Het due to case ascertainment p=0.47
Het due to ancestry p=0.29
Residual heterogeneity p=0.84

rs7334326 C (EU: 0.14; JPT: 0.77)

x.

RIN3 locus

14q32.12



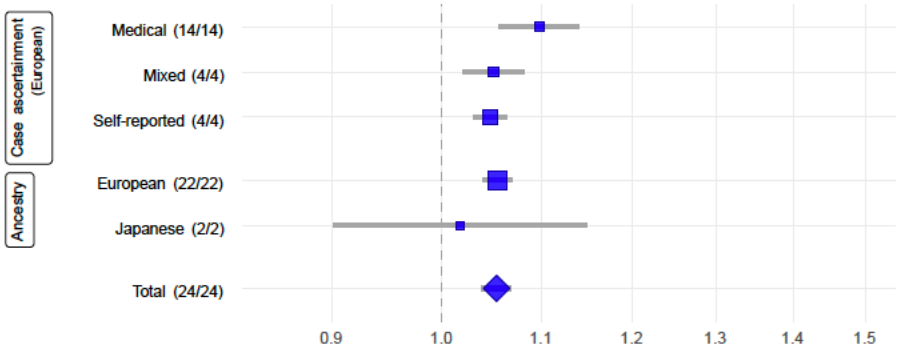
Overall heterogeneity p=0.46
Het due to case ascertainment p=0.17
Het due to ancestry p=0.74
Residual heterogeneity p=0.49

rs57281976 G (EU: 0.77; JPT: 0.81)

y.

BMF locus

15q15.1



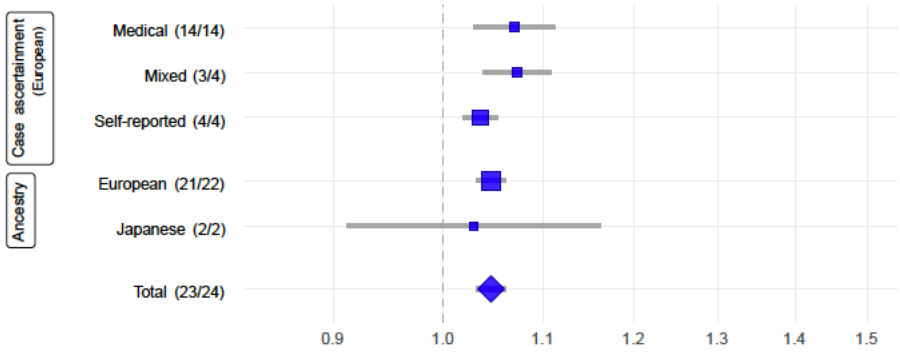
Overall heterogeneity p=0.28
Het due to case ascertainment p=0.22
Het due to ancestry p=0.39
Residual heterogeneity p=0.28

rs12441483 C (EU: 0.47; JPT: 0.78)

z.

SKAP1 locus

17q21.32



Overall heterogeneity p=0.14
Het due to case ascertainment p=0.06
Het due to ancestry p=0.59
Residual heterogeneity p=0.22

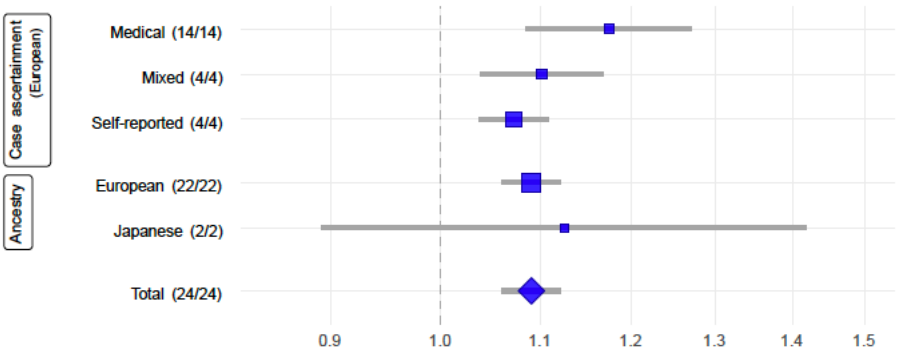
rs66683298 C (EU: 0.6; JPT: 0.78)

Studies without rs66683298: 1 Mixed (EGCUT)

a2.

CEP112 locus

17q24.1



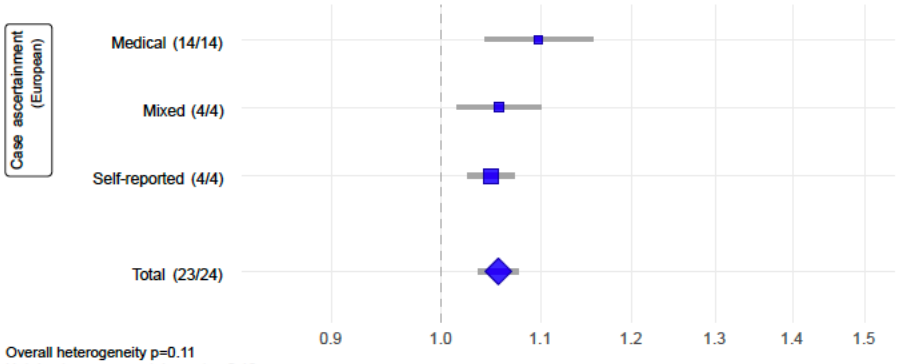
Overall heterogeneity p=0.63
Het due to case ascertainment p=0.11
Het due to ancestry p=0.79
Residual heterogeneity p=0.73

rs7214750 C (EU: 0.93; JPT: 0.94)

b2.

ACTL9 locus

19p13.2



Overall heterogeneity p=0.11
Het due to case ascertainment p=0.18
Het due to ancestry p=0.51
Residual heterogeneity p=0.11

rs2967684 A (0.16)

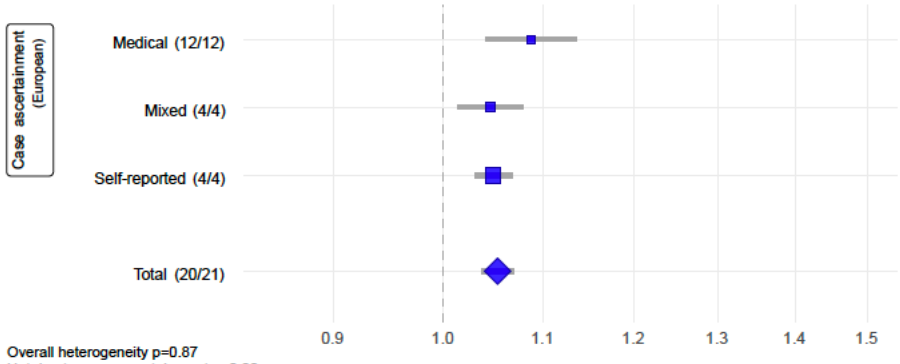
Studies without rs2967684: 1 Japanese (ADACHI)

rs2967684 was only available in one Japanese study: BBJ, OR=1.03 (0.89-1.19) and did not allow ancestry-based analyses.

c2.

TEX11 locus

Xq13.1



Overall heterogeneity p=0.87
Het due to case ascertainment p=0.20
Het due to ancestry NA
Residual heterogeneity p=0.94

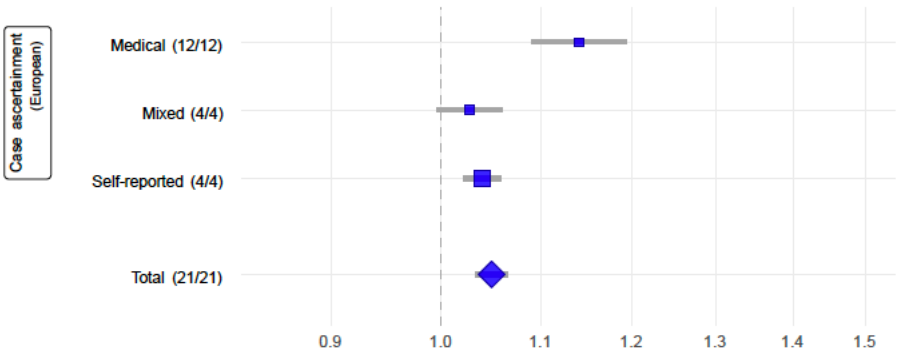
rs13441059 A (0.35)

rs13441059 was not available in Japanese studies, and therefore did not allow ancestry-based stratified analyses.

d2.

FRMD7 locus

Xq26.2



Overall heterogeneity p=7.7e-03
Het due to case ascertainment p=0.02
Het due to ancestry NA
Residual heterogeneity p=0.03

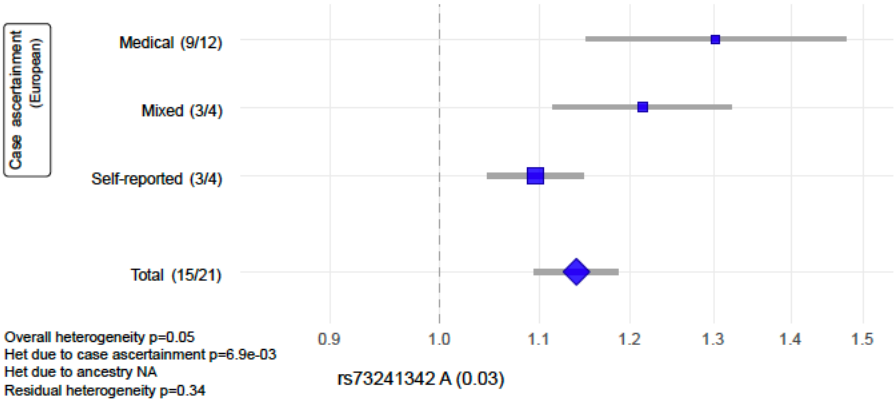
rs5933091 T (0.69)

rs5933091 was only available in one Japanese study: ADACHI, OR=3.65 (1.99-6.68) and did not allow ancestry-based analyses.

e2.

PLAC1 locus

Xq26.3

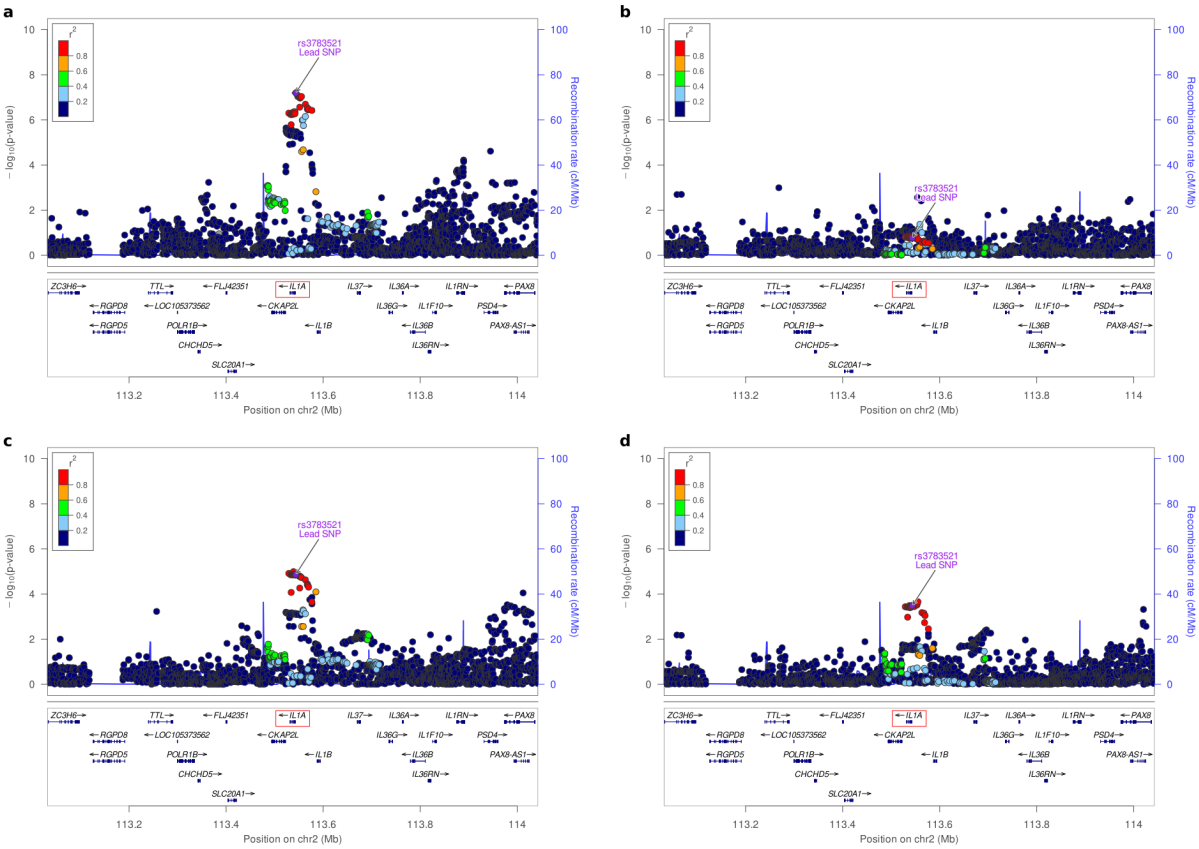


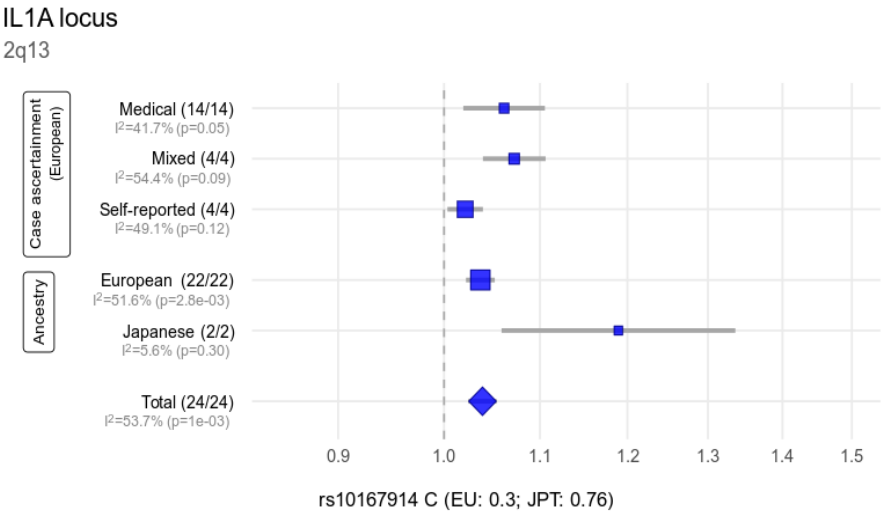
Overall heterogeneity p=0.05
Het due to case ascertainment p=6.9e-03
Het due to ancestry NA
Residual heterogeneity p=0.34

Studies without rs73241342: 3 Medical (ENDOX1, UCSF, DBDS), 1 Mixed (NFBC), 1 Self-reported (TUK)
rs73241342 was not available in Japanese studies, and therefore did not allow ancestry-based stratified analyses.

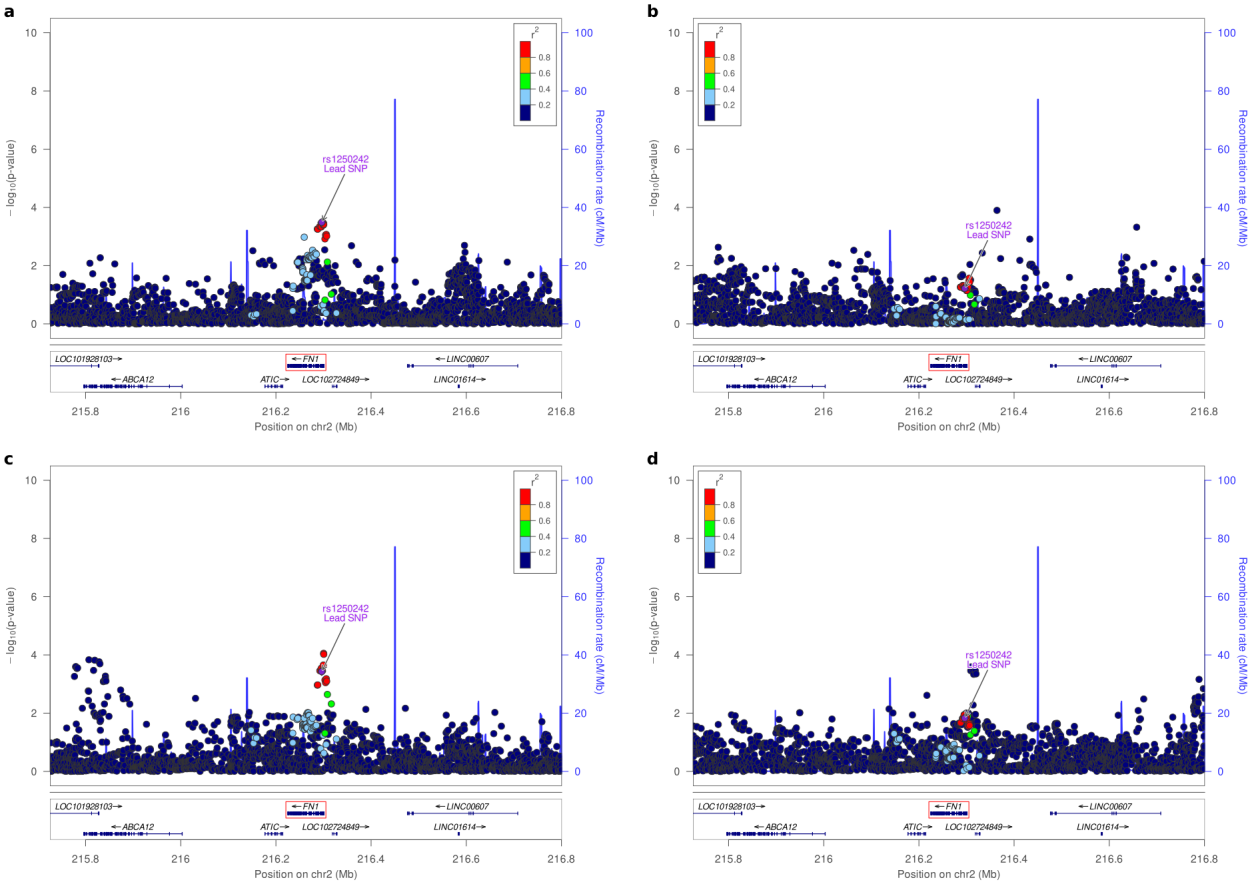
Supplementary Figure 5. Regional association plots and forest plots for (a) IL1A/2q13 and (b) FN1/2q35. The regional association plots include results for (i) endometriosis, (ii) rASRM stage I/II endometriosis, (iii) rASRM stage III/IV endometriosis, (iv) endometriosis-associated infertility. The association results are shown on the y-axis as $-\log_{10}$(P-value) and on the x-axis is the genomic location (hg 19). The top associated SNP is coloured purple and the other SNPs are coloured according to the strength of LD with the top SNP by r^2 in the European 1000 Genomes dataset. In the forest plots the rsid and the effective allele along with effective allele frequency are given under each plot. The x-axis displays the association log odds-ratio scale, and meta-analysis results by three different case ascertainment categories including medical, missed and self-reported and two different ancestries including European and Japanese ancestry results.

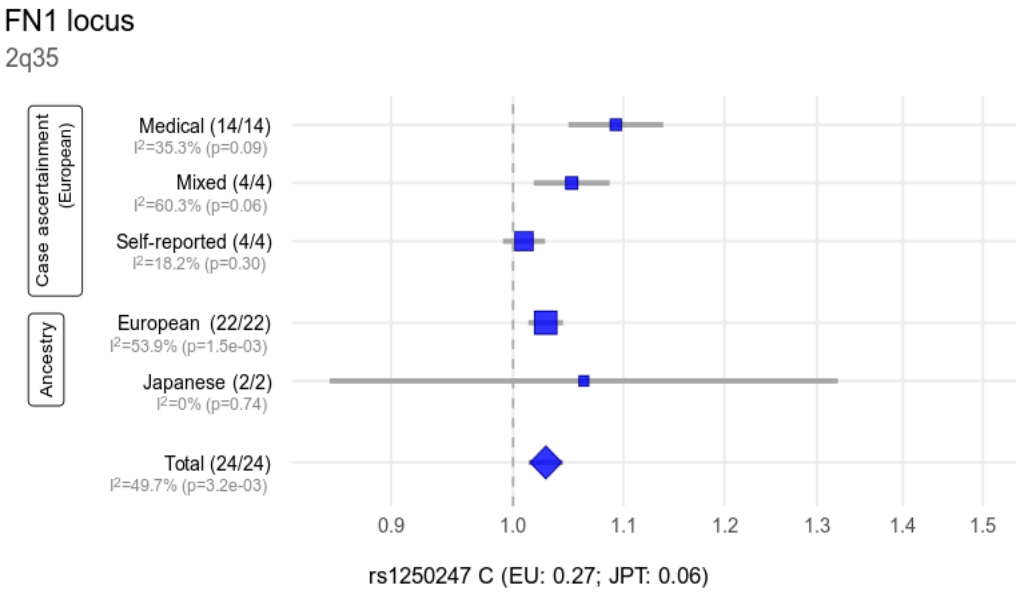
(i) IL1A/2q13





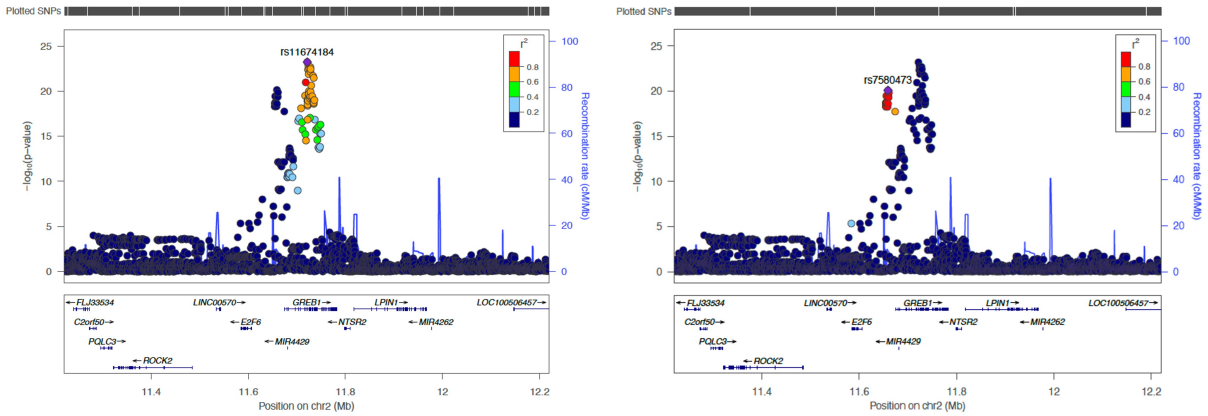
(ii) FN1/2q35


a
b
c
d
rs1250242
Lead SNP
Position on chr2 (Mb)
Recombination rate (cM/Mb)
LOC101928103
FN1
LINC00607
ABCA12
ATIC
LOC102724849
LINC01614

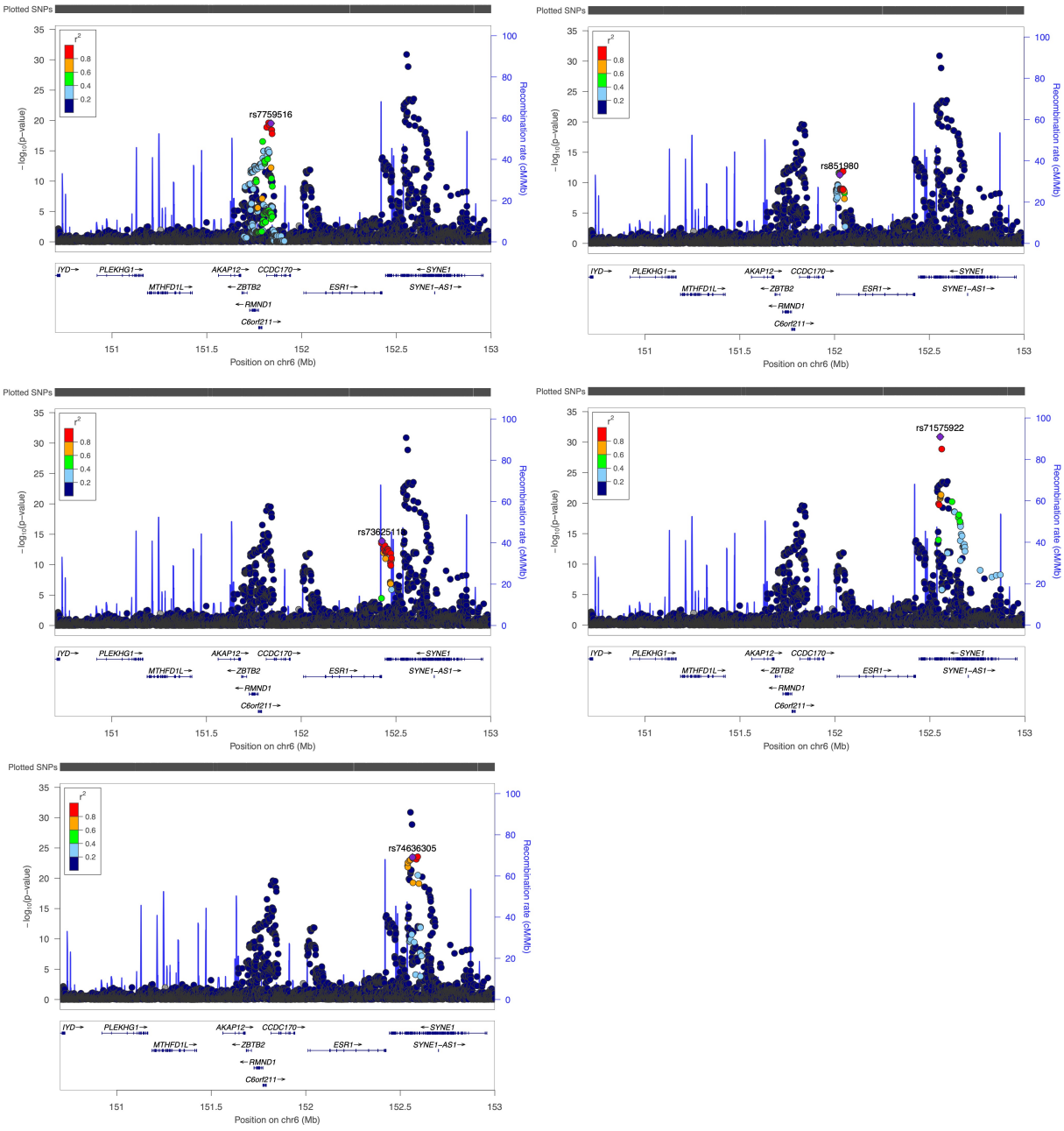

FN1 locus
2q35
Case ascertainment (European)
Medical (14/14)
I²=35.3% (p=0.09)
Mixed (4/4)
I²=60.3% (p=0.06)
Self-reported (4/4)
I²=18.2% (p=0.30)
Ancestry
European (22/22)
I²=53.9% (p=1.5e-03)
Japanese (2/2)
I²=0% (p=0.74)
Total (24/24)
I²=49.7% (p=3.2e-03)
0.9
1.0
1.1
1.2
1.3
1.4
1.5
rs1250247 C (EU: 0.27; JPT: 0.06)

Supplementary Figure 6. Regional association plots for 4 endometriosis risk loci with multiple independent signals: (a) *GREB1*/2p25.1, (b) *SYNE1*/6q25.1, (c) *CDKN2B-AS1*/9p21.3, (d) *IGF1*/12q23.2. The association results are shown on the y-axis as $-\log_{10}$(P-value) and the x-axis displays the genomic location (hg 19). The top associated SNP is coloured purple and the other SNPs are coloured according to the strength of LD (r^2) with the top SNP in the European 1000 Genomes dataset.

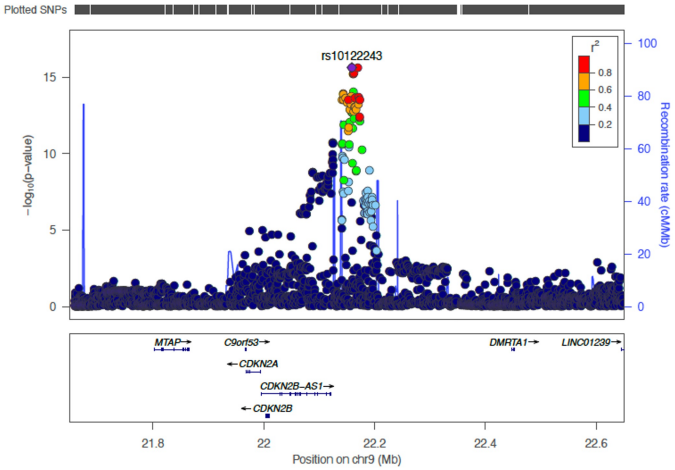
(a) *GREB1*/2p25.1

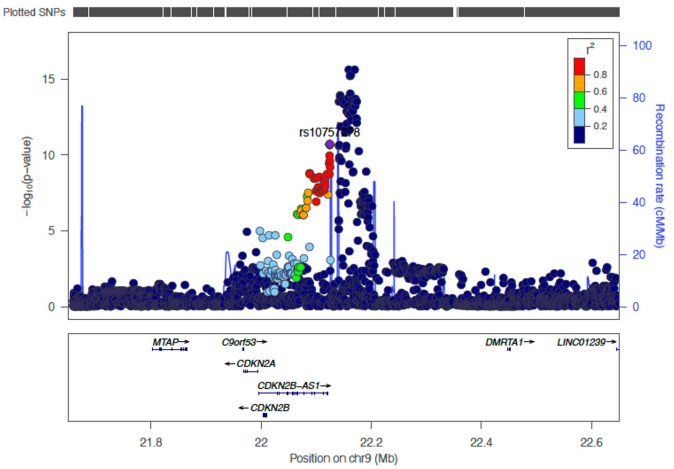


(b) *SYNE1*/6q25.1

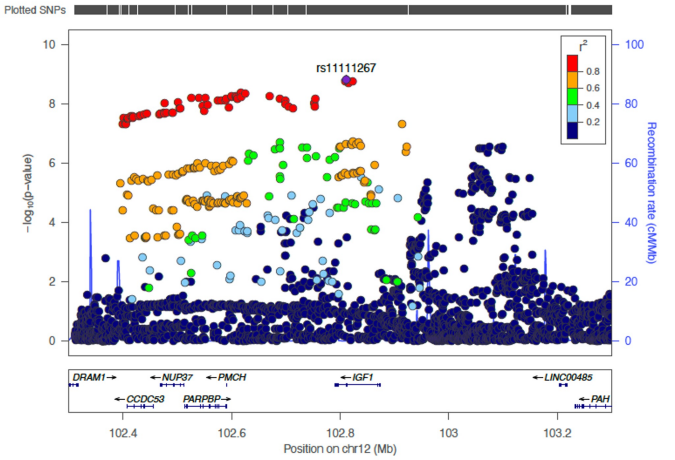


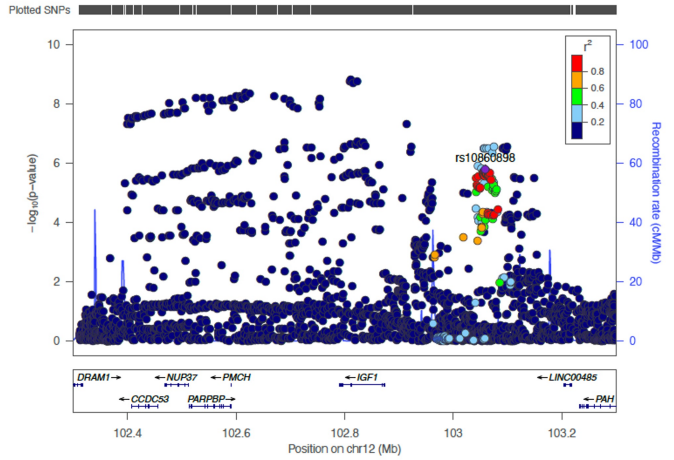
(c) *CDKN2B-AS1*/9p21.3





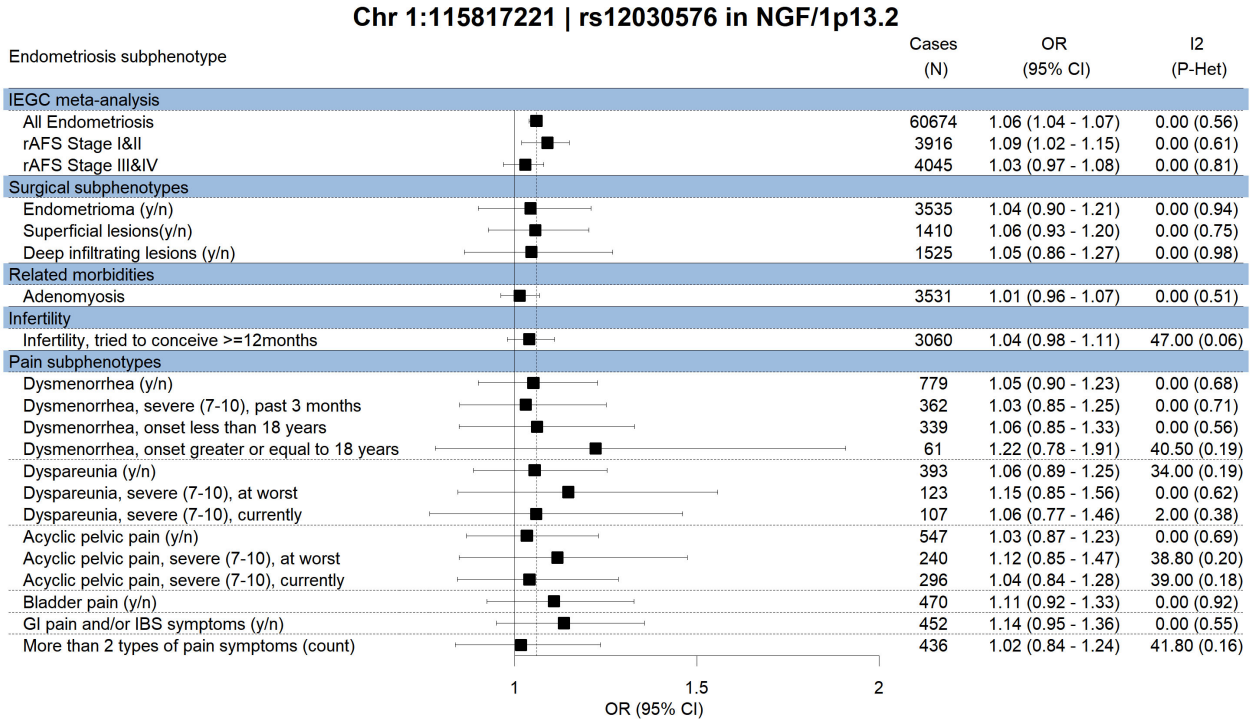
(d) *IGF1*/12q23.2



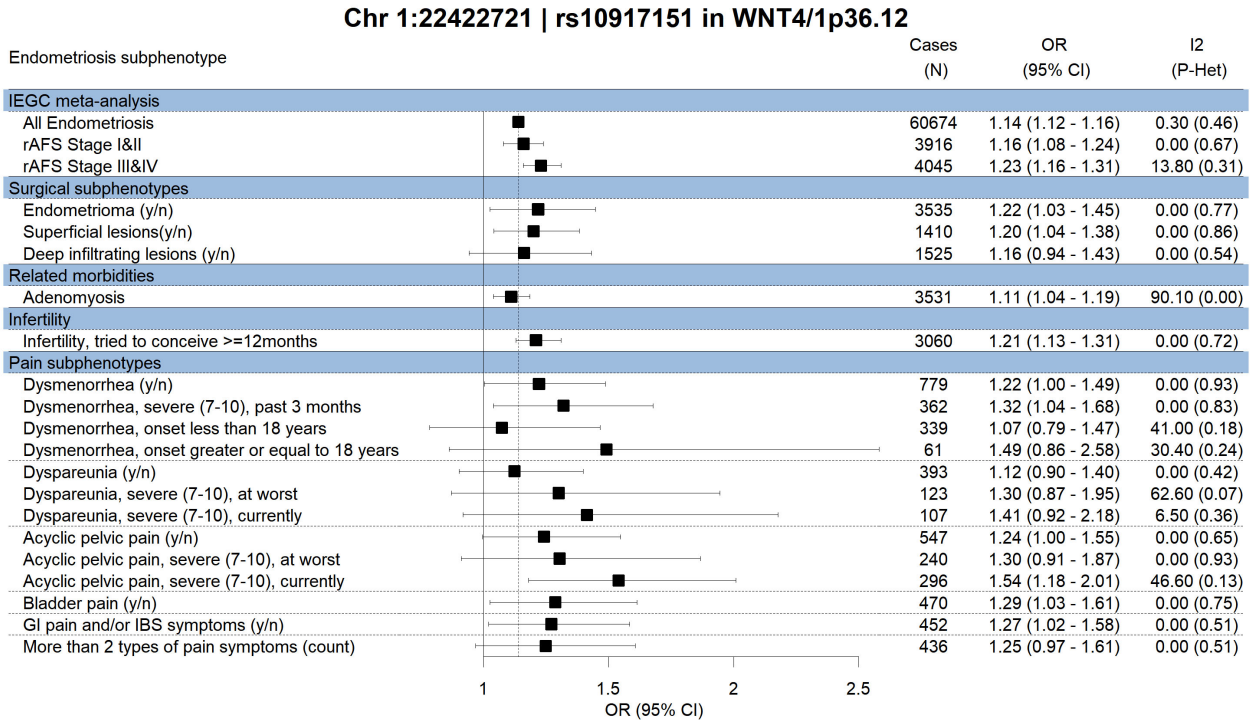


Supplementary Figure 7. Forest plots for the 42 lead SNPs showing the sub-phenotype analysis results: (i) NGF/1p13.2, (ii) WNT4/1p36.12 rs10917151, (iii) SLC19A2/1q24.2, (iv) DNM3/1q24.3, (v)GREB1/2p25.1, (vi) BMPR2/2q33.1 rs6435157, (vii) ETAA1/2p14 rs1430787, (viii) BSN/3p21.31, (ix) PDLIM5/4q22.3, (x) KDR/4q12 rs1903068, (xi) EBF1/5q33.3 rs2946160, (xii) FAM120B/6q27 rs11756073, (xiii) SYNE1/6q25.1 rs71575922, (xiv) ID4/6p22.3 rs6456259, (xv) CD109/6q13, (xvi) HEY2/6q22.31, (xvii) HOXA10/7p15.2, (xviii) 7p15.2/7p15.2, (ixx) 7p12.3/7p12.3,(xx) VPS13B/8q22.2 rs12549438, (xxi) KCTD9/8p21.2, (xxii) GDAP1/8q22.2, (xxiii) CDKN2-BAS1/9p21.3, (xxiv) ABO/9q34.2, (xxv) ASTN2/9q33.1 rs10983311,(xxvi) MLLT10/10p12.31 rs10828249, (xxvii) RNLS/10q23.31,(xxviii) WT1/11p14.1 rs7924571, (xxix) FSHB/11p14.1 rs3858429, (xxx) VEZT/12q22 rs12320196, (xxxi) PTPRO/12p12.3, (xxxii) HOXC10/12p13.13, (xxxiii) IGF1/12q23.2, (xxxiv) DLEU1/13q14.2 rs7334326, (xxxv) RIN3/14q32.12, (xxxvi) SRP14-AS1/15q15.1 rs12441483, (xxxvii) SKAP1/17q21.32, (xxxviii) CEP112/17q24.1, (xxxix) ACTL9/19p13.2, (xxxx) TEX11/Xq13.1, (xxxxi) FRMD7/Xq26.2,(xxxxii) LINC00629/Xq26.3 rs73241342.

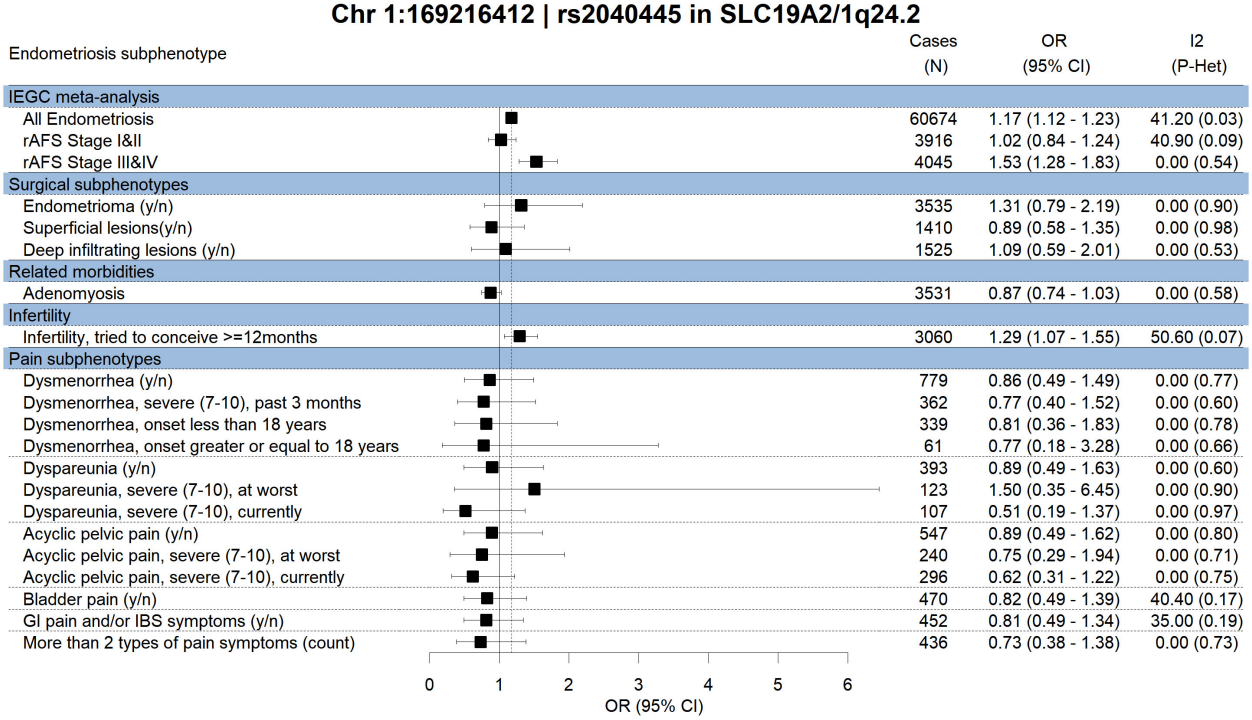
(i) NGF/1p13.2



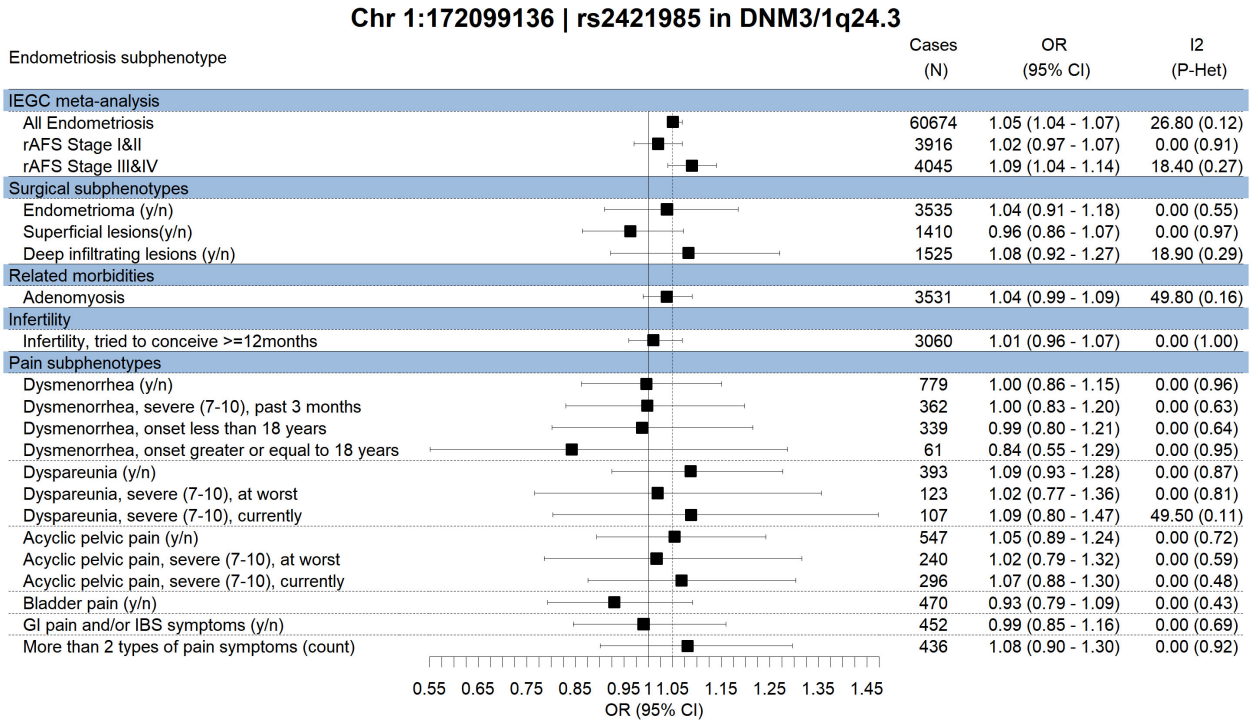
(ii) WNT4/1p36.12 rs10917151



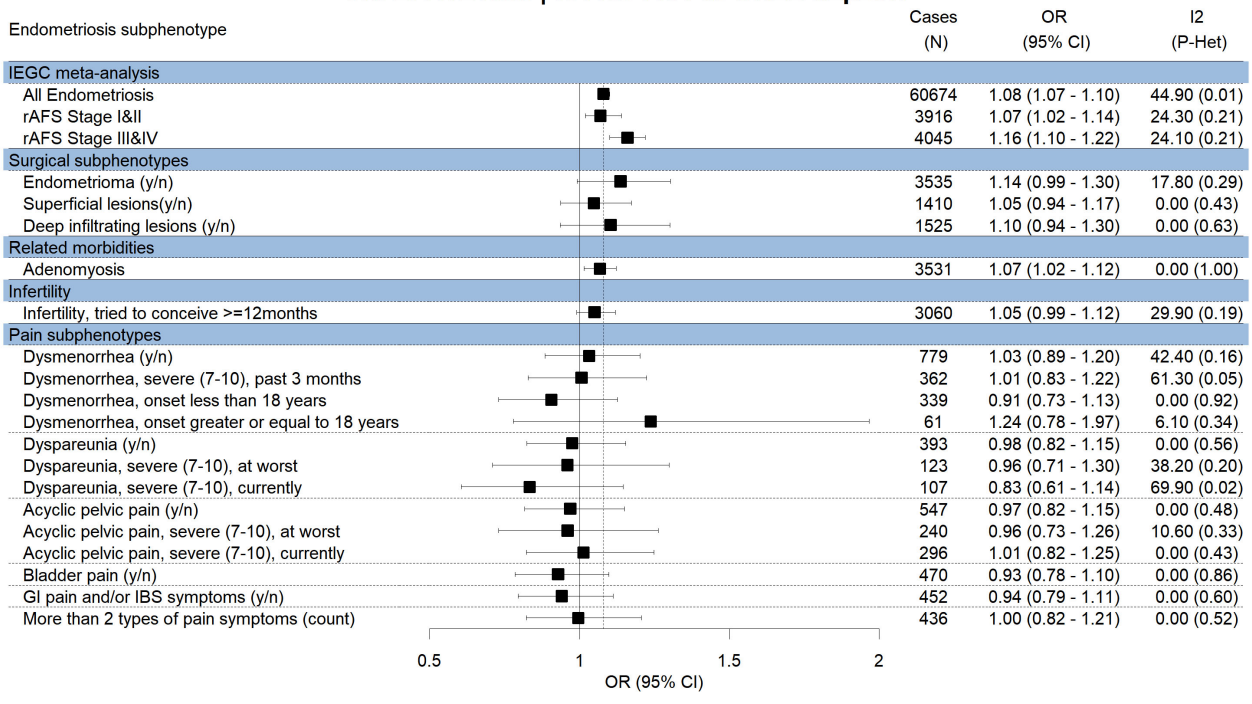
(iii) SLC19A2/1q24.2



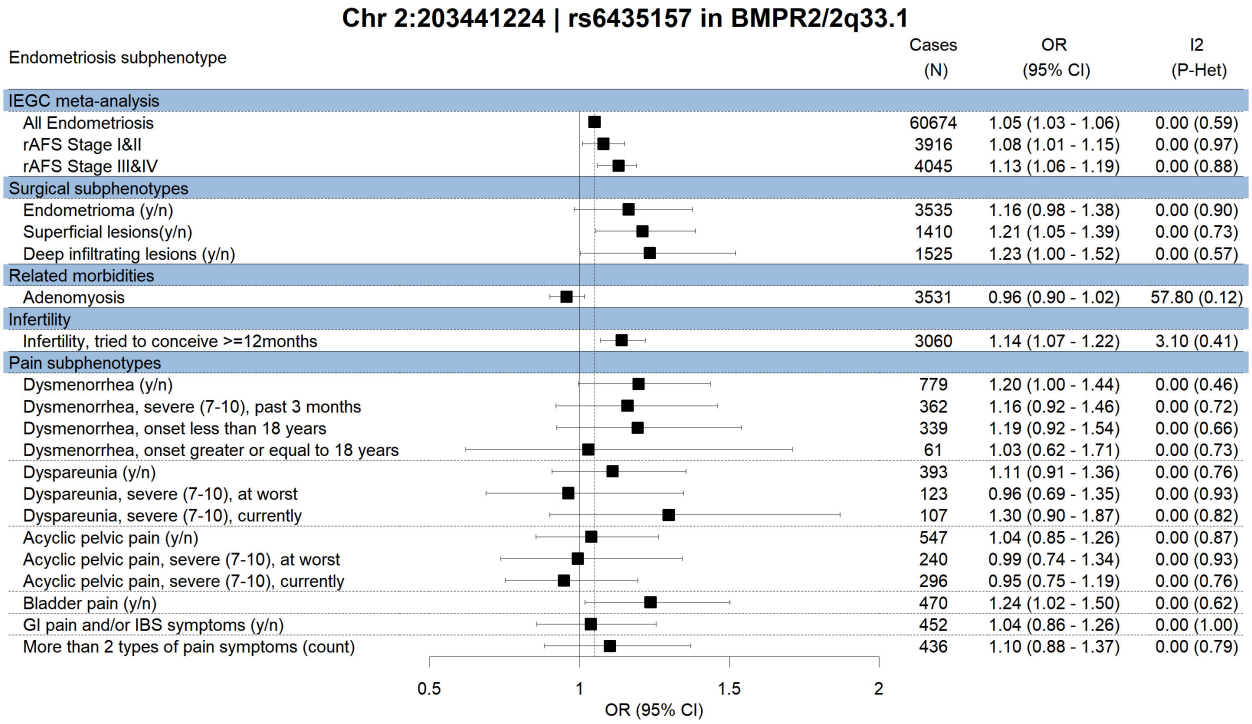
(iv) DNM3/1q24.3



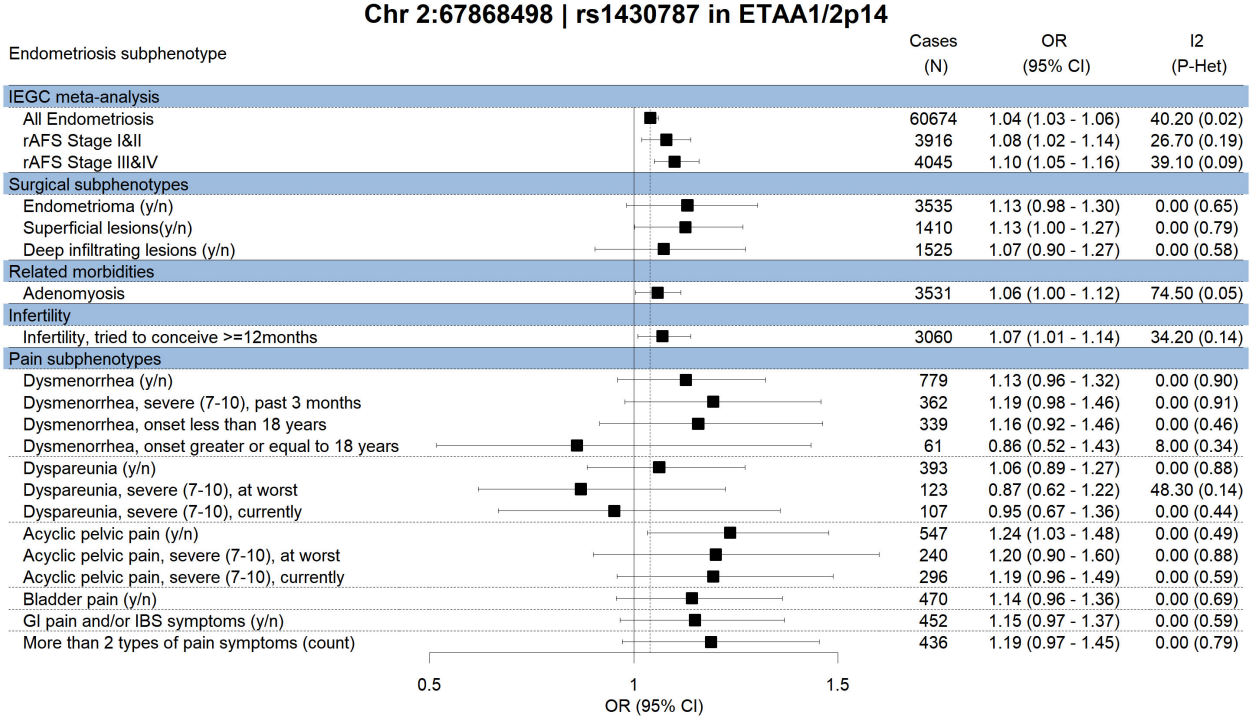
(v) GREB1/2p25.1



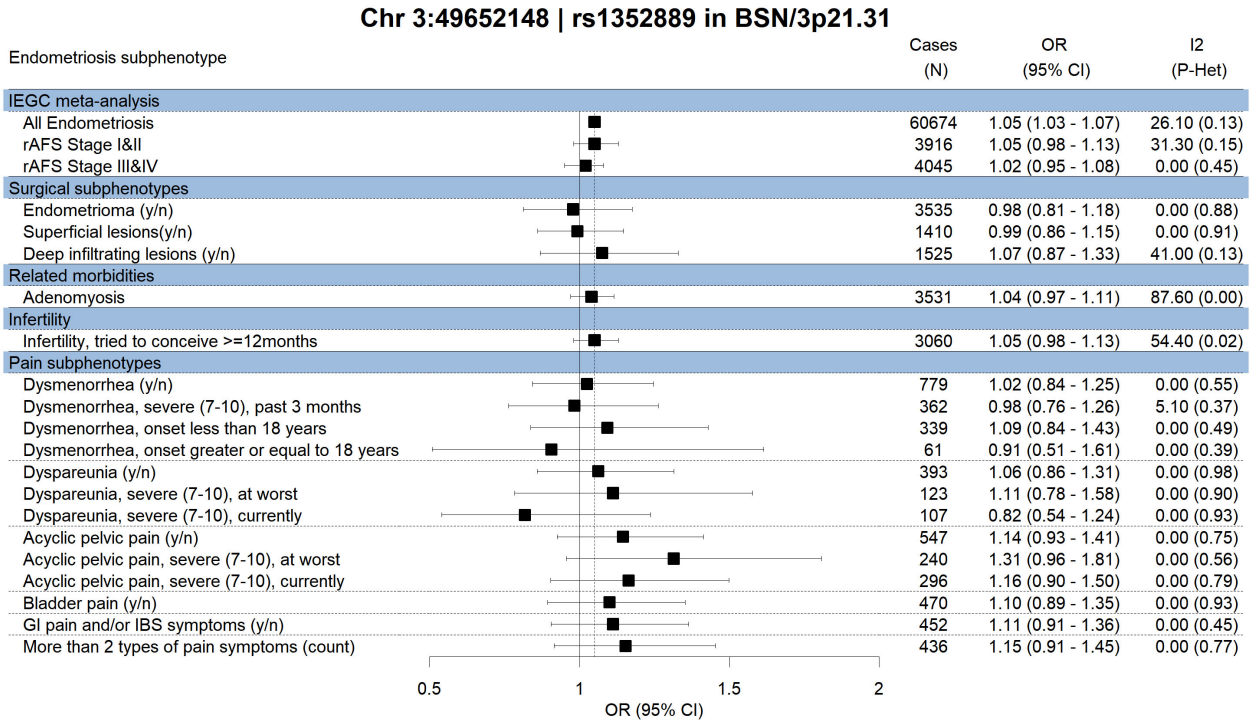
(vi) BMPR2/2q33.1 rs6435157



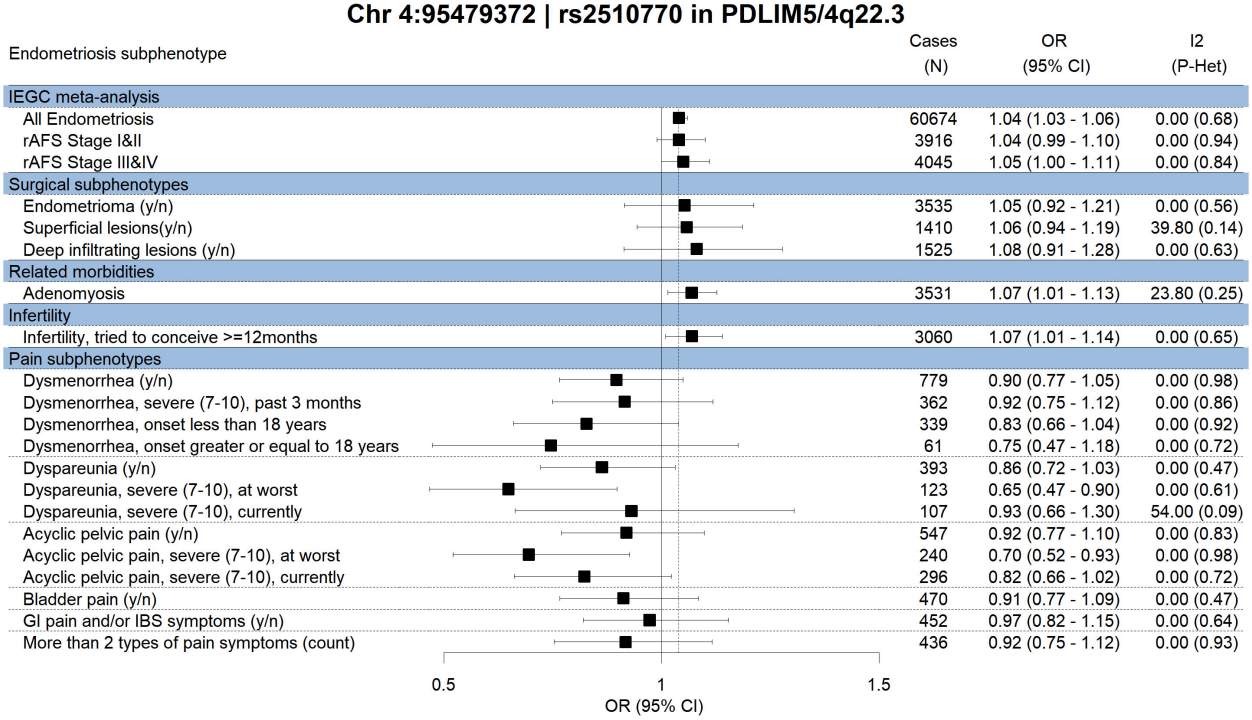
(vii) ETAA1/2p14 rs1430787



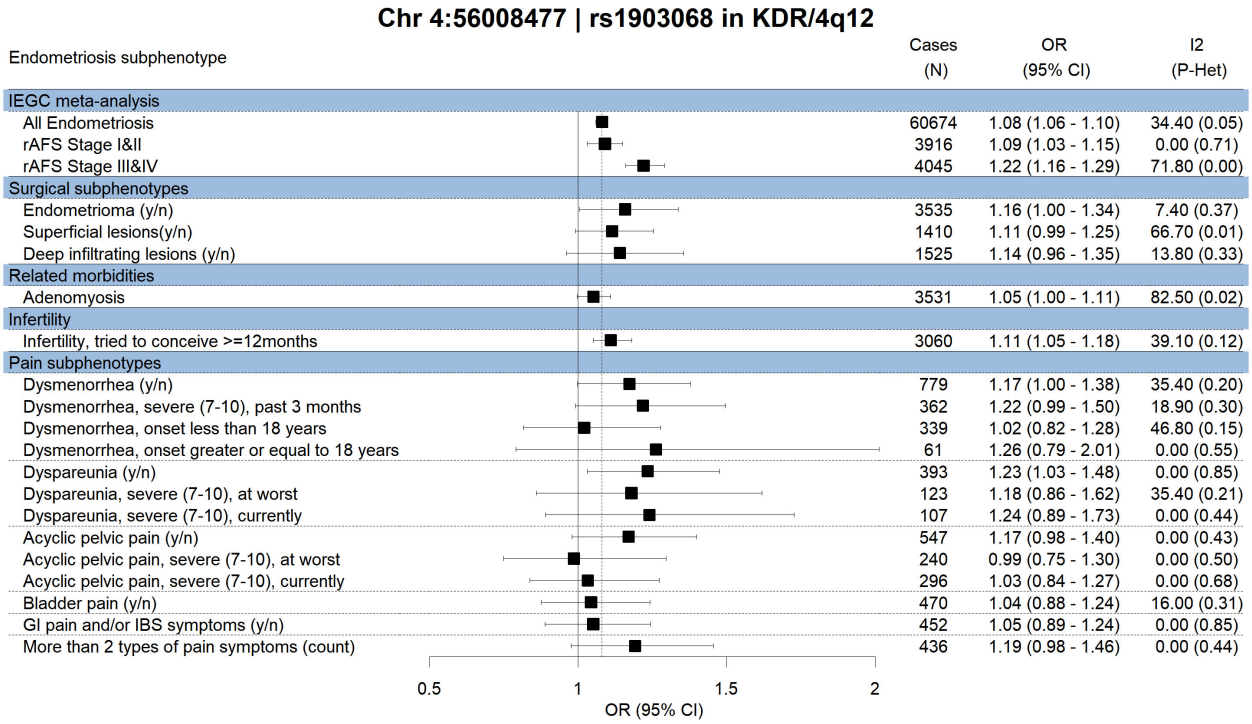
(viii) BSN/3p21.31



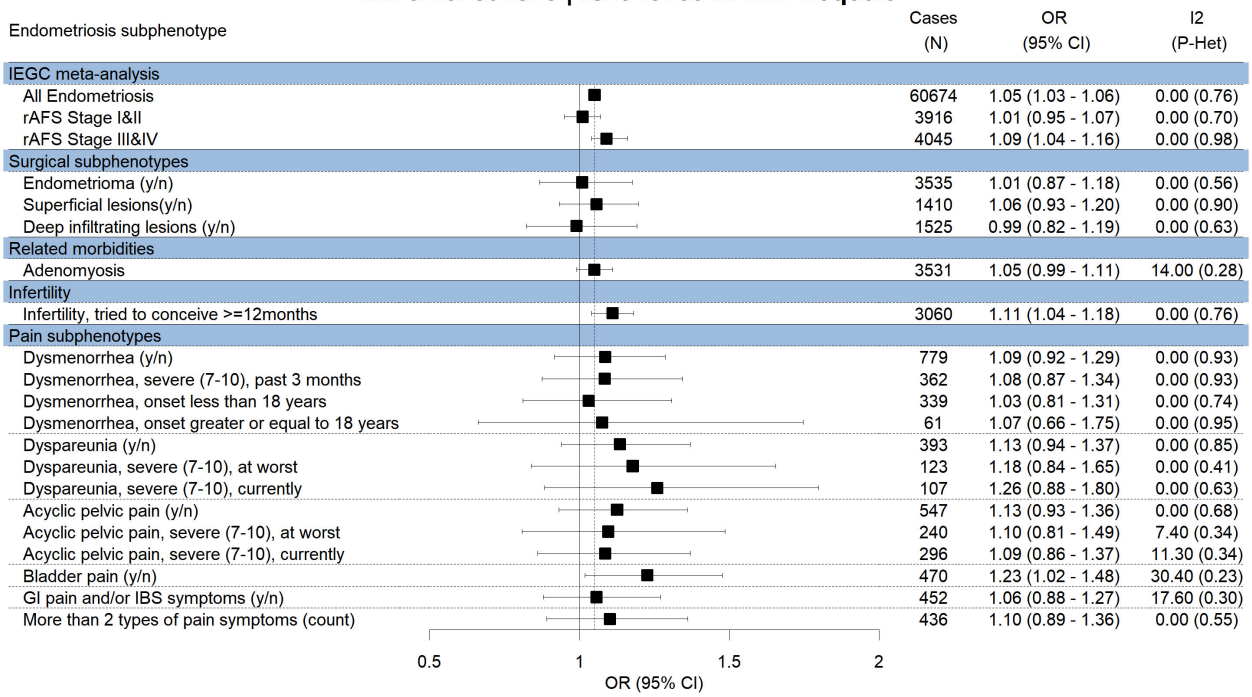
(ix) PDLIM5/4q22.3



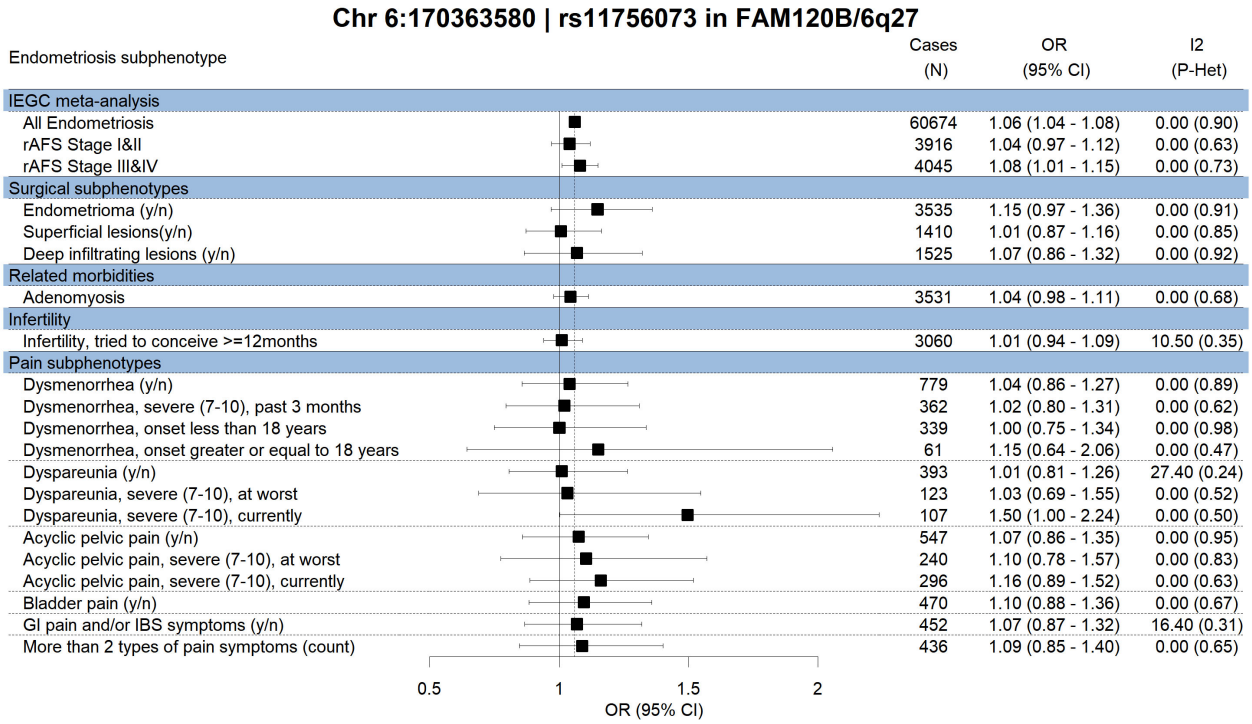
(x) KDR/4q12 rs1903068



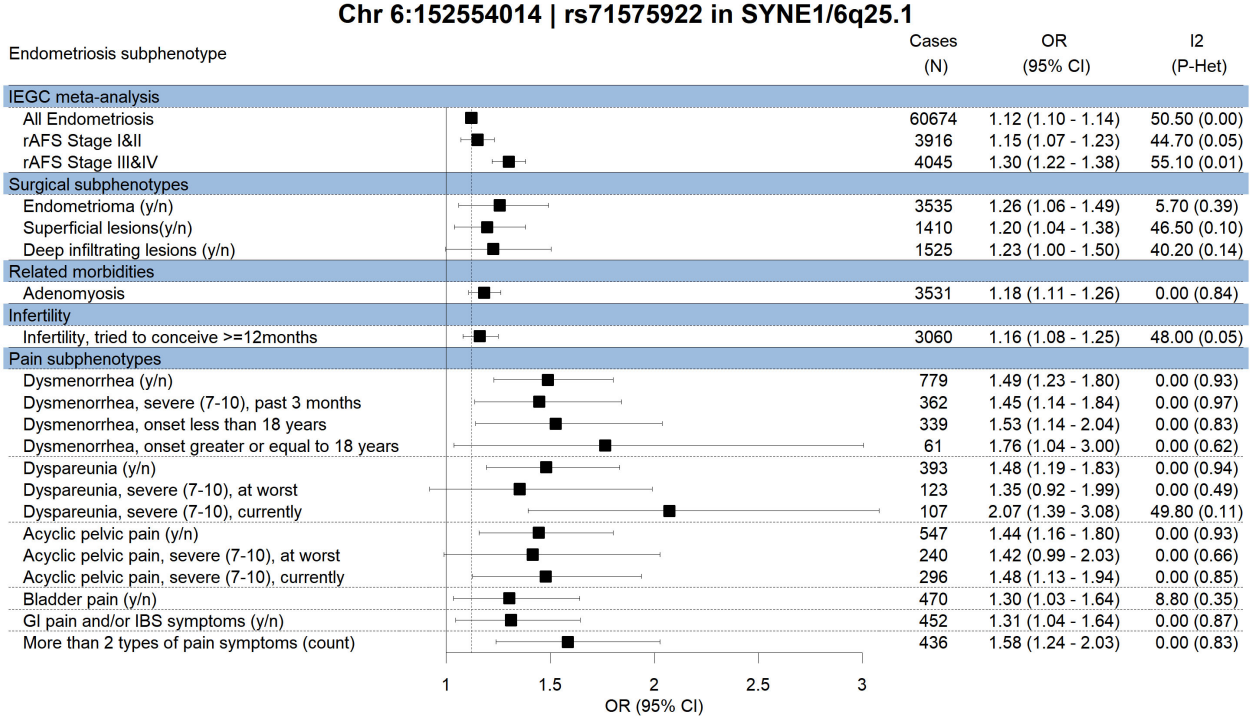
(xi) EBF1/5q33.3 rs2946160



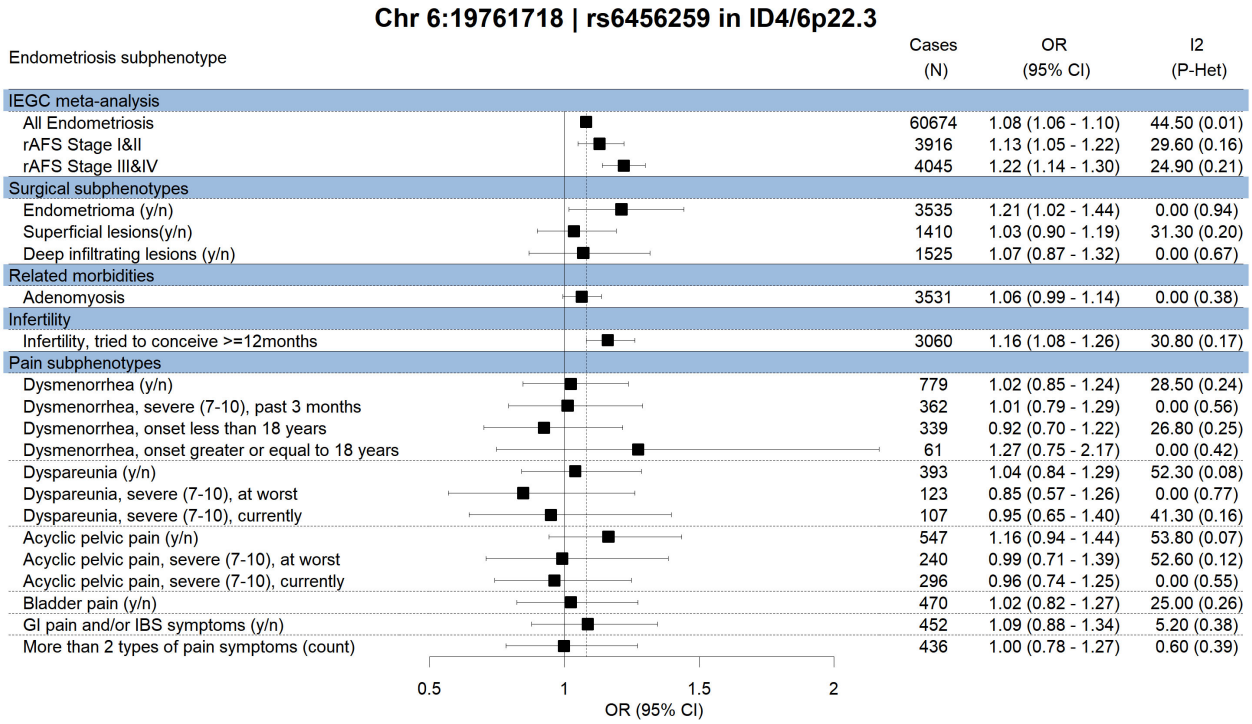
(xii) FAM120B/6q27 rs11756073



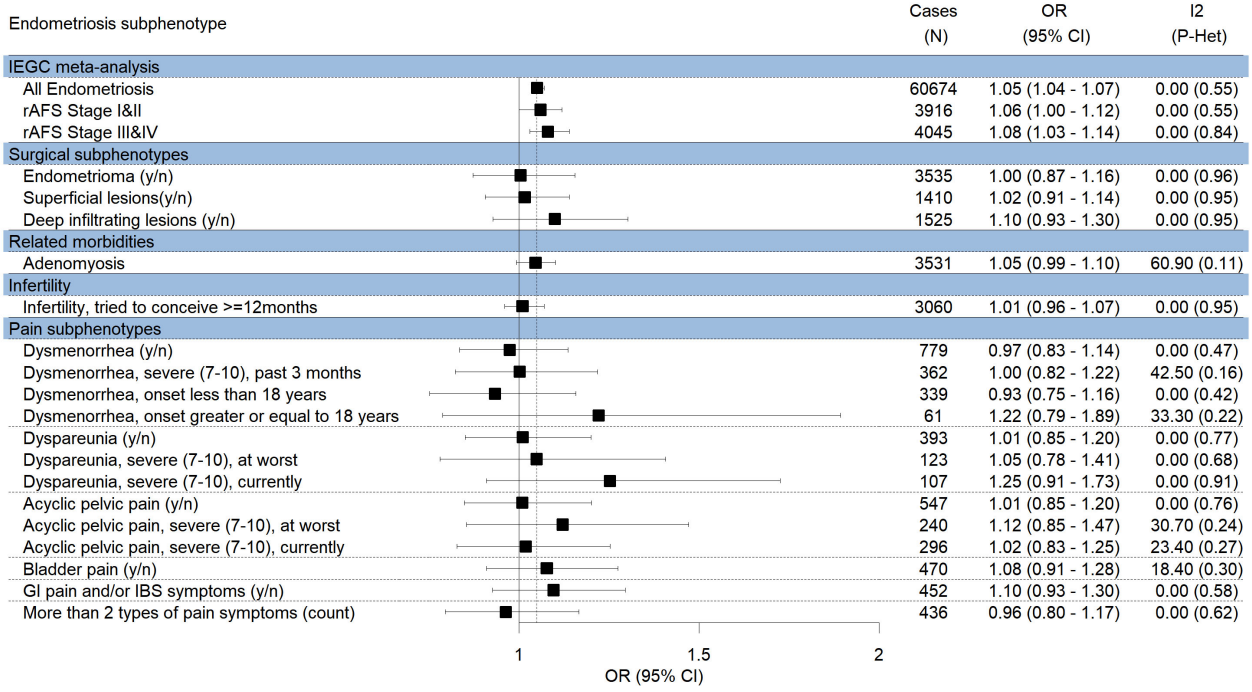
(xiii) SYNE1/6q25.1 rs71575922



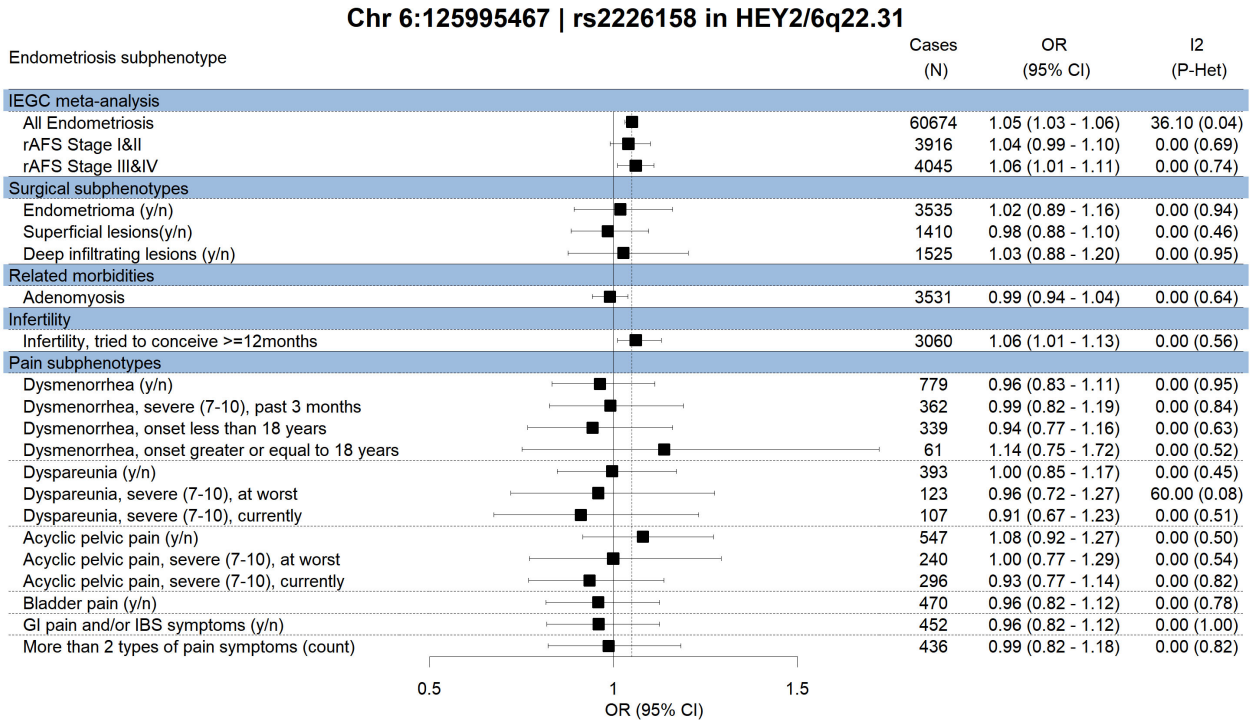
(xiv) ID4/6p22.3 rs6456259



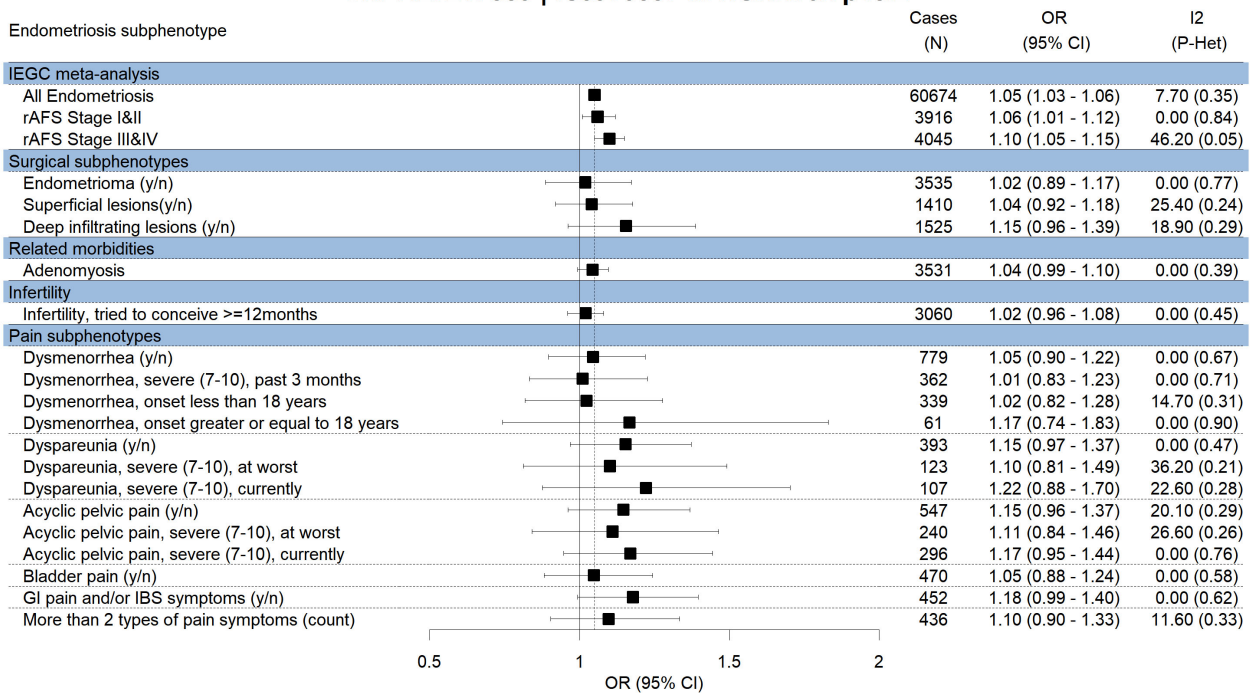
(xv) CD109/6q13



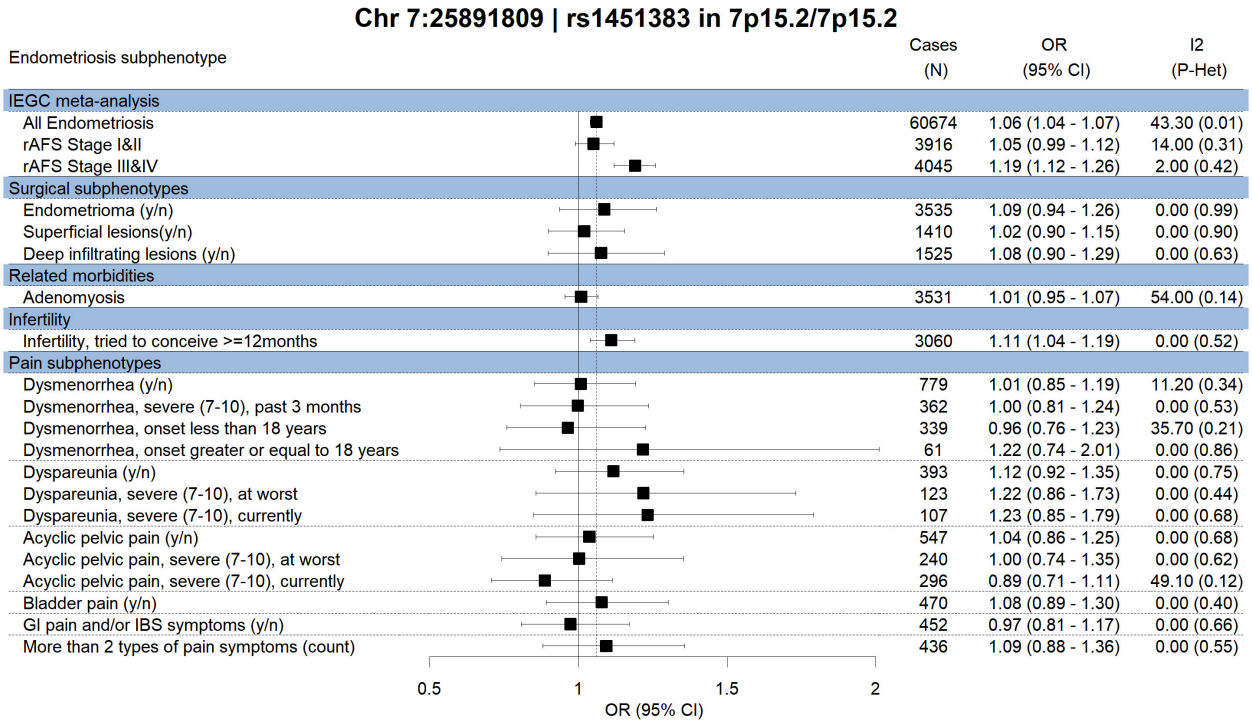
(xvi) HEY2/6q22.31



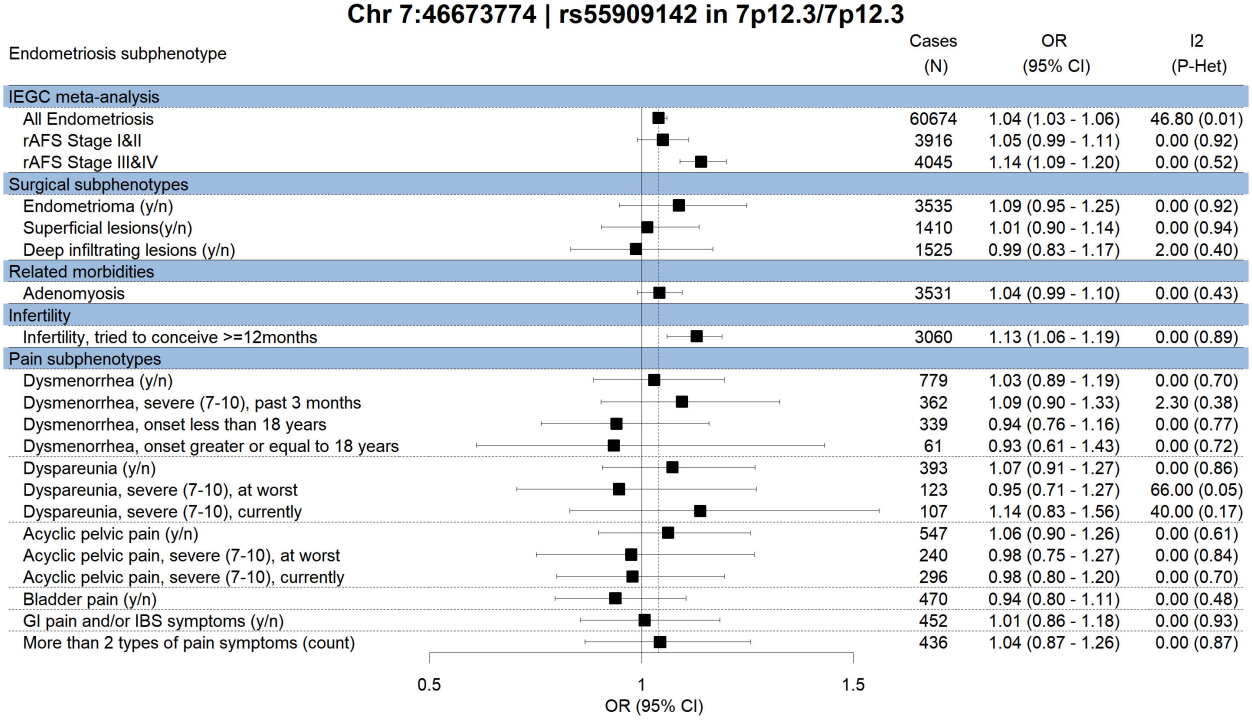
(xvii) HOXA10/7p15.2



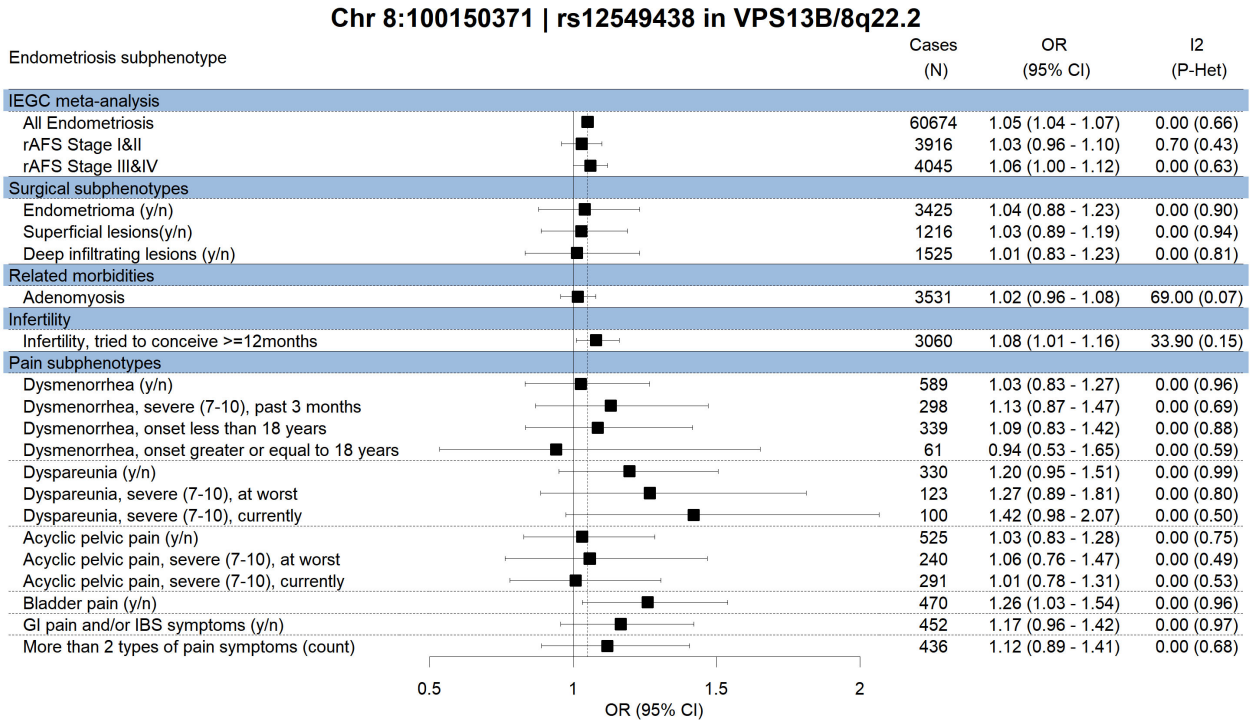
(xviii) 7p15.2/7p15.2



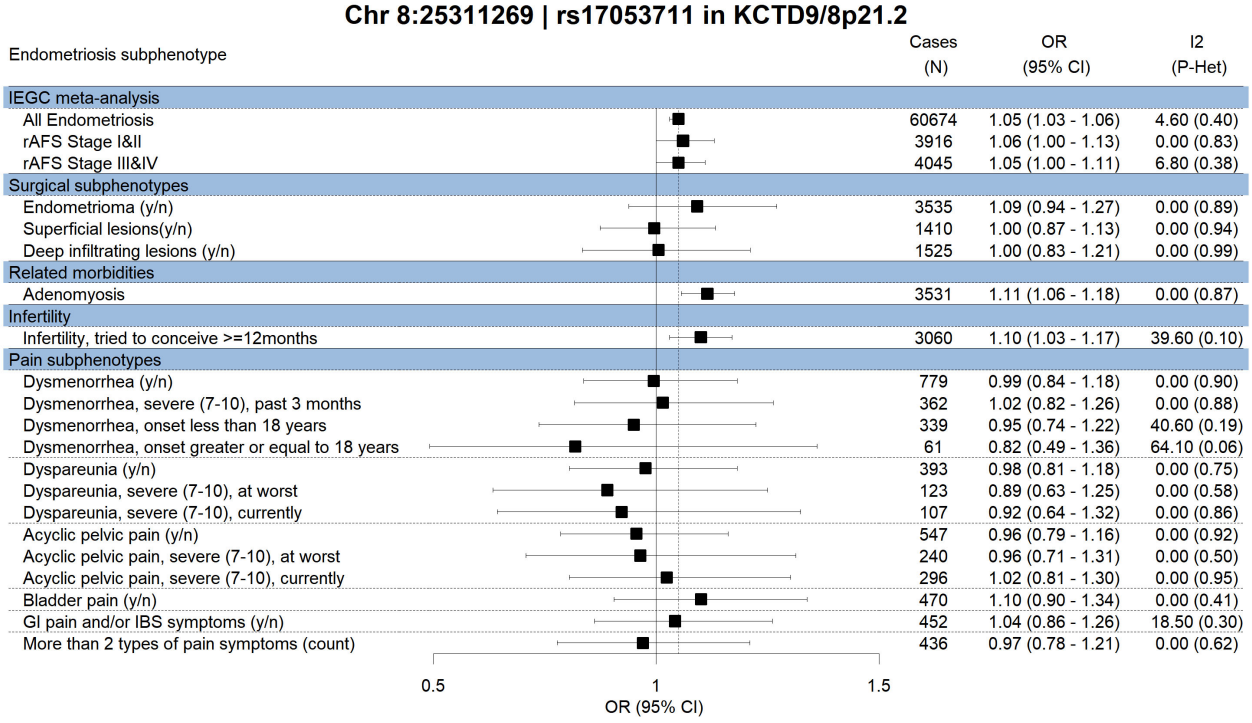
(ixx) 7p12.3/7p12.3



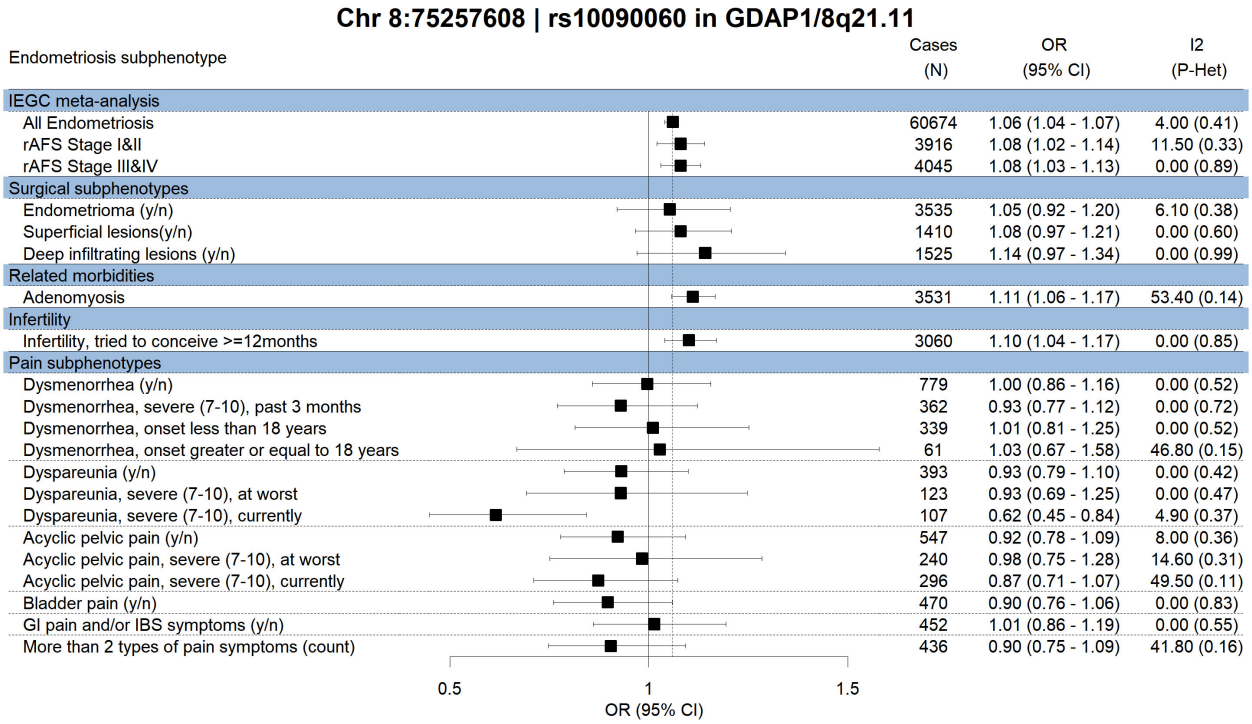
(xx) VPS13B/8q22.2 rs12549438



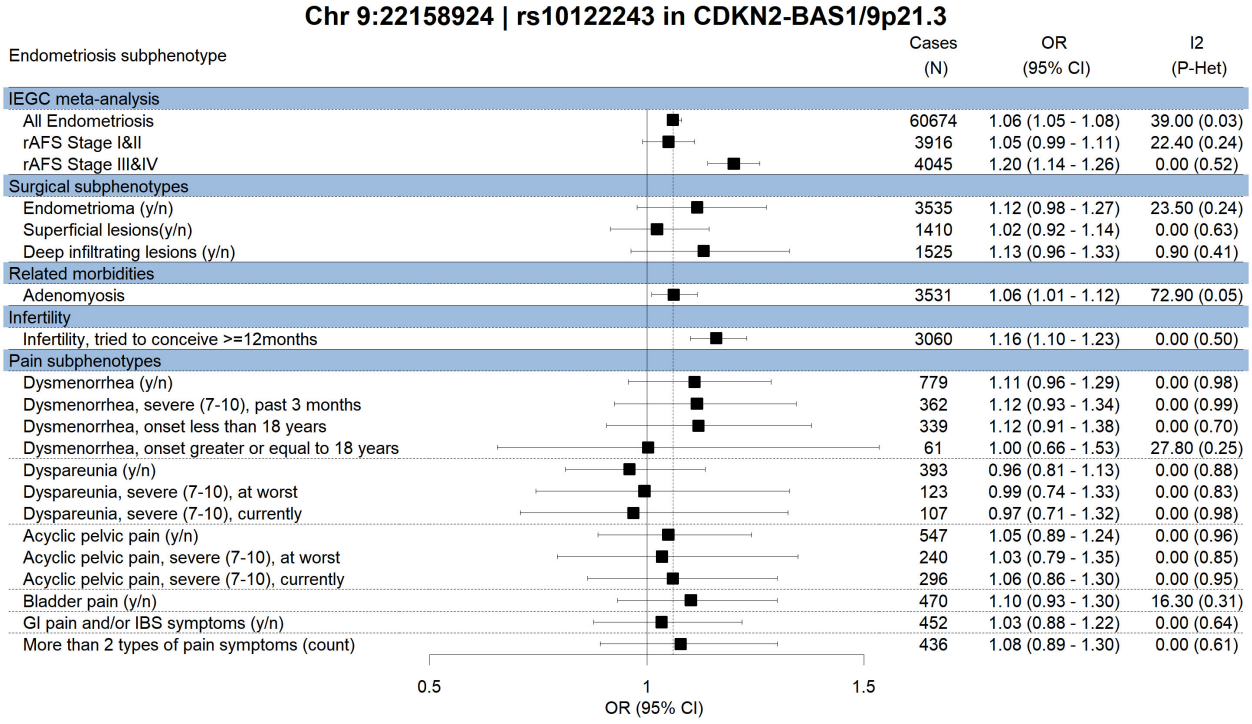
(xxi) KCTD9/8p21.2



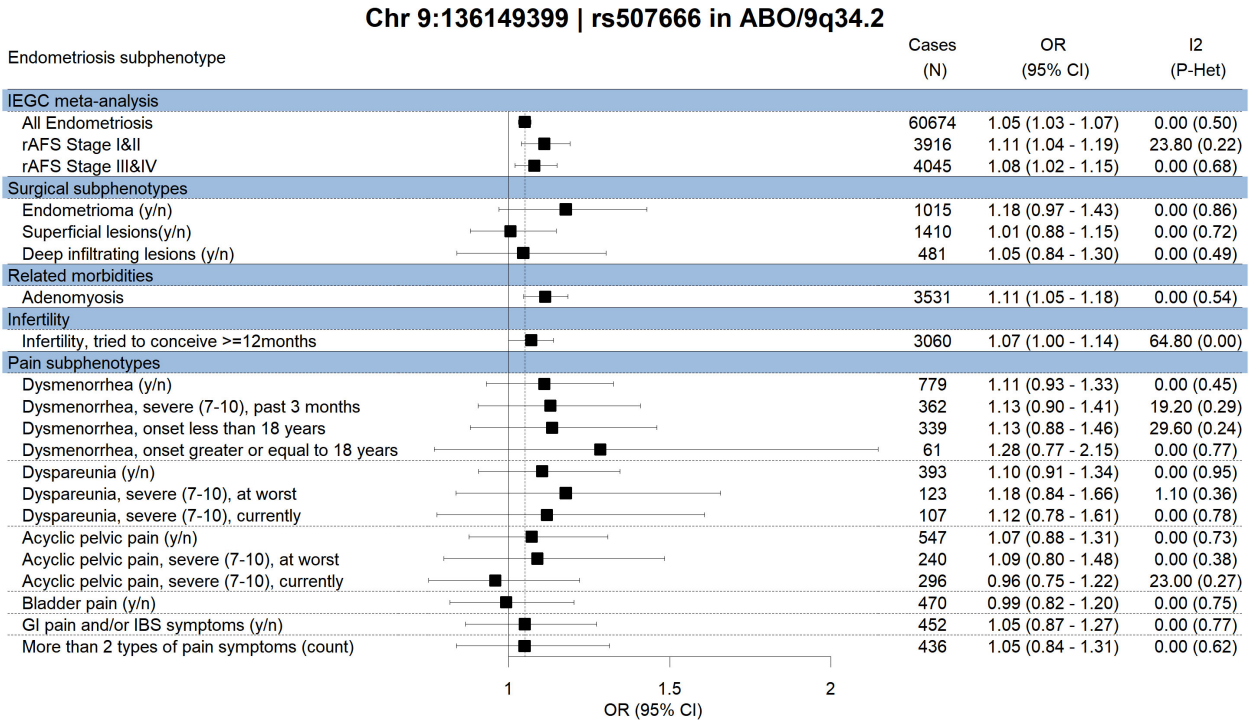
(xxii) GDAP1/8q22.2



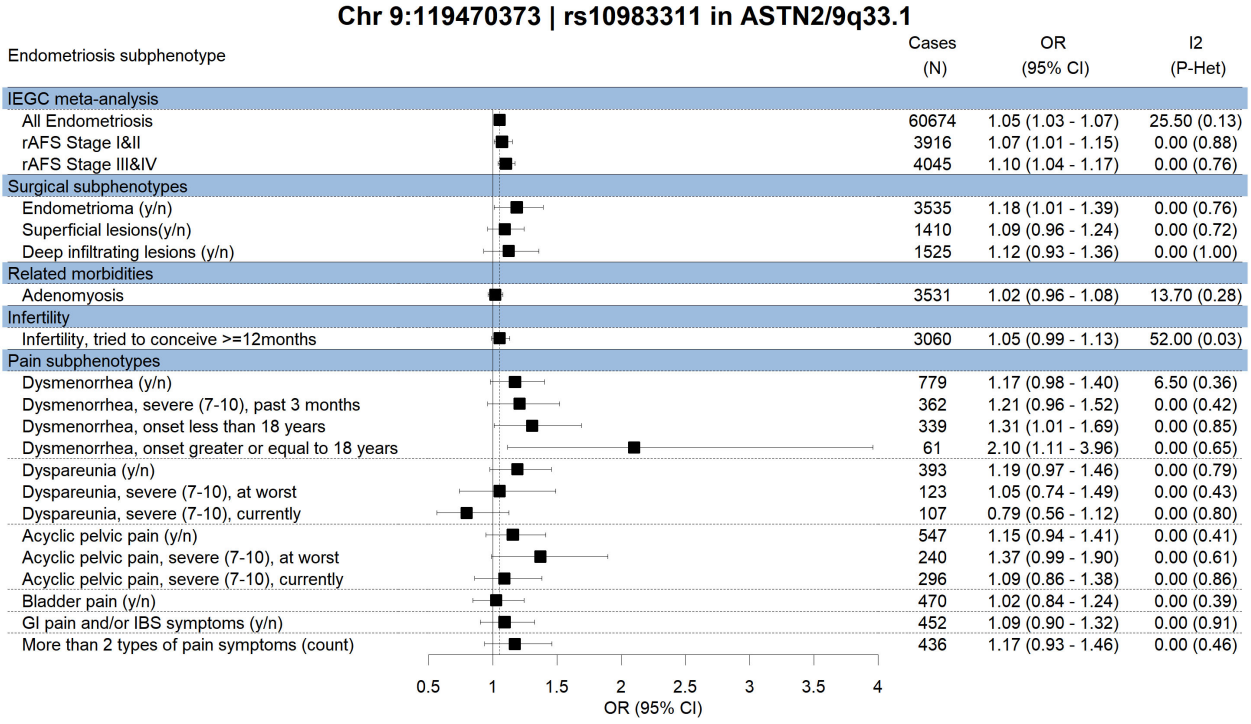
(xxiii) CDKN2-BAS1/9p21.3



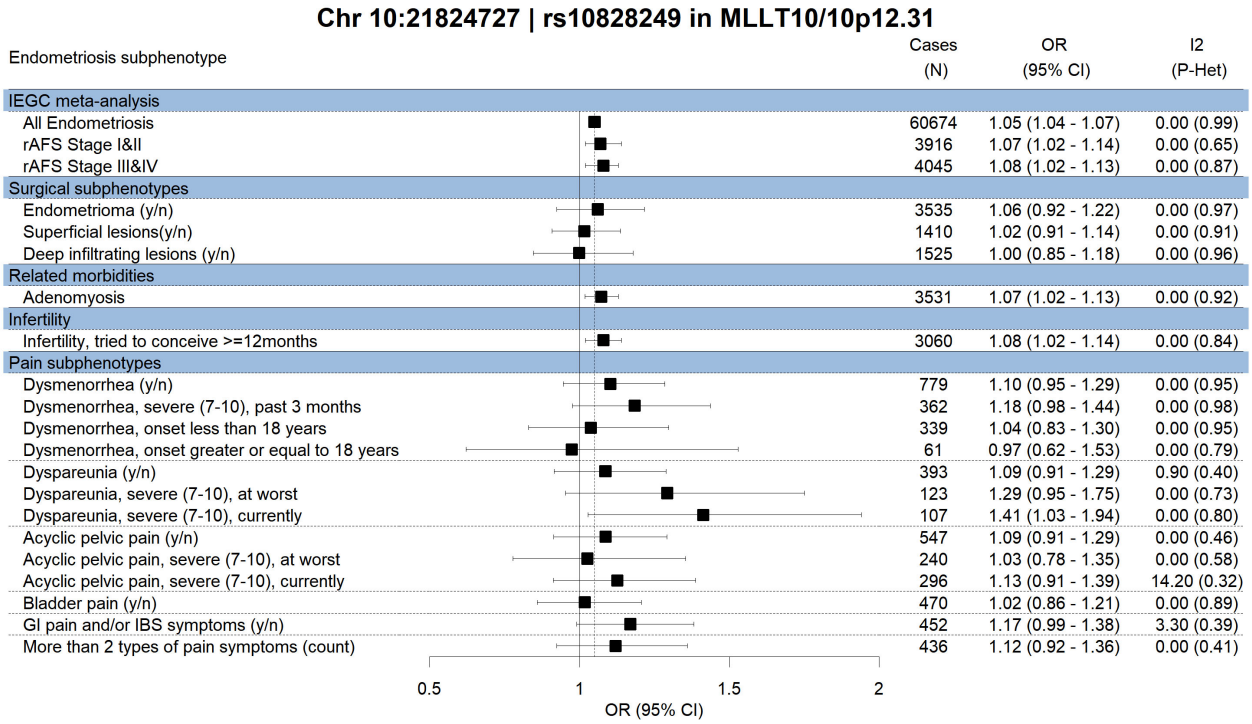
(xxiv) ABO/9q34.2



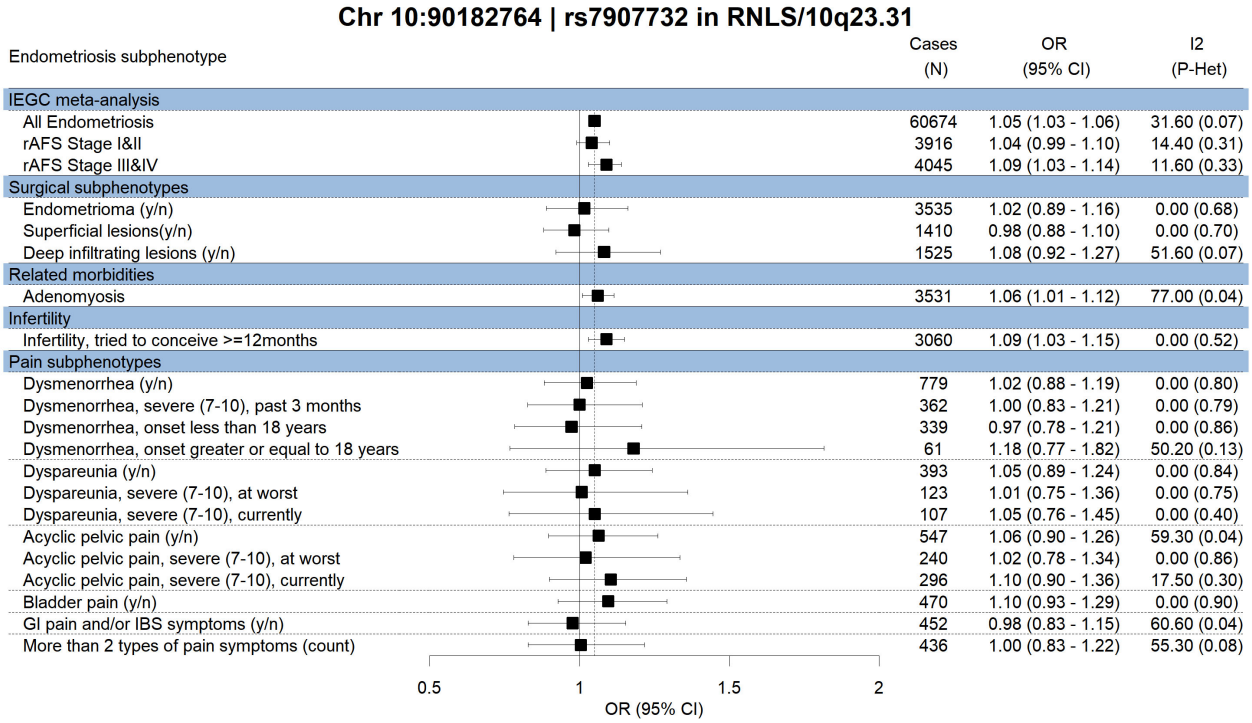
(xxv) ASTN2/9q33.1 rs10983311



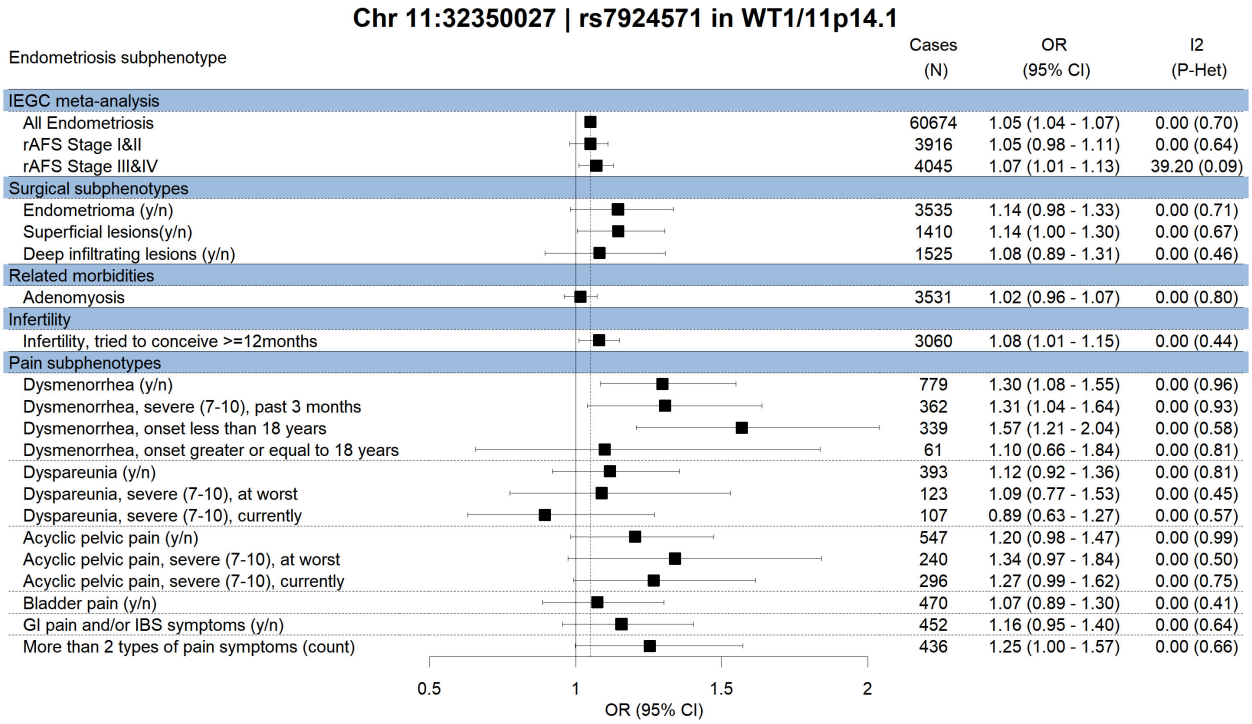
(xxvi) MLLT10/10p12.31 rs10828249



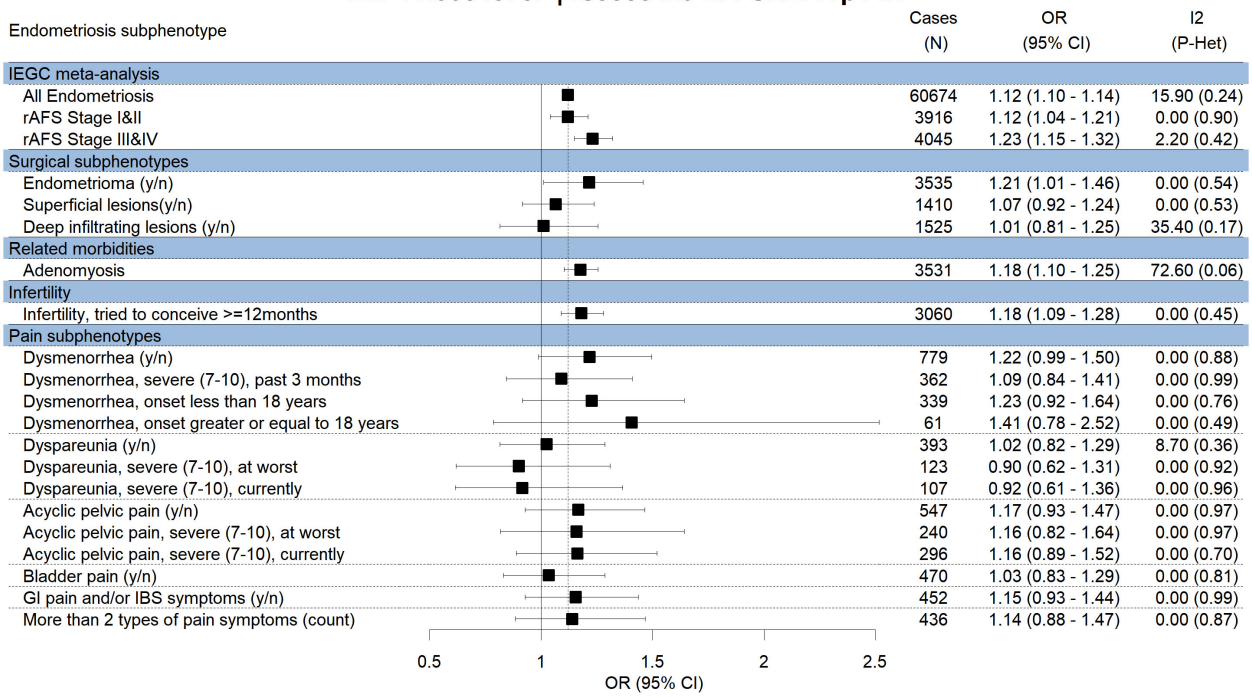
(xxvii) RNLS/10q23.31



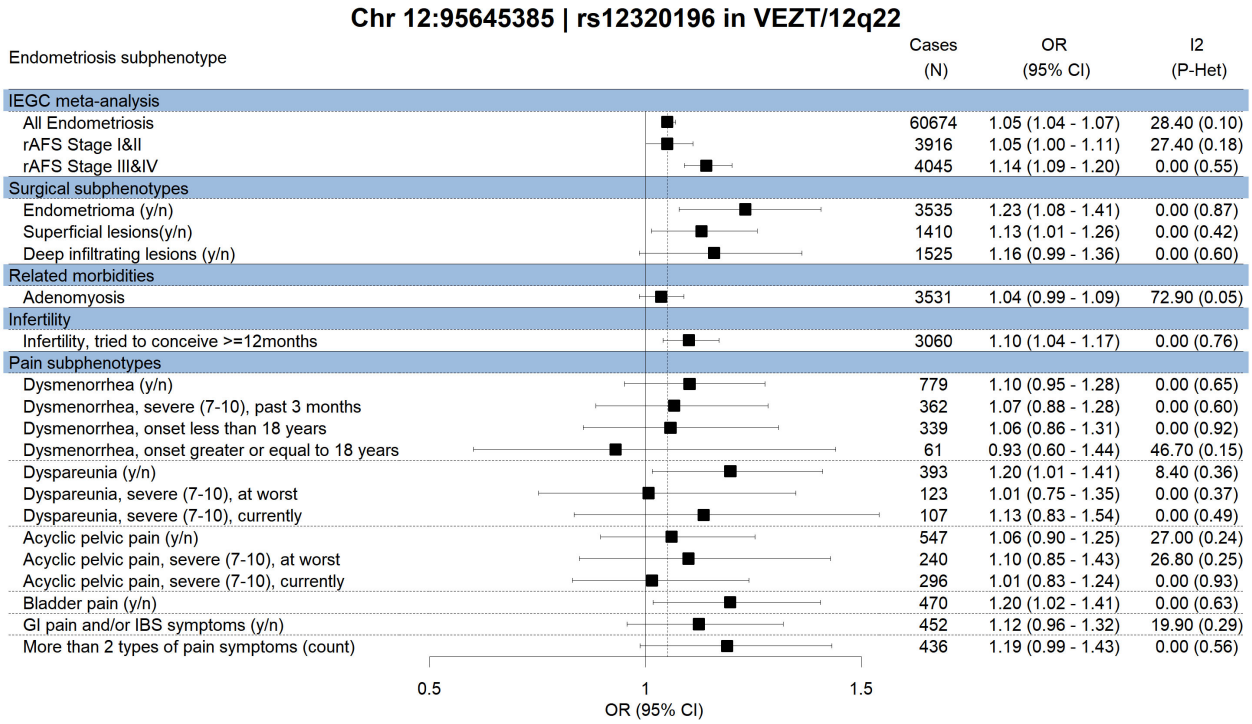
(xxviii) WT1/11p14.1 rs7924571



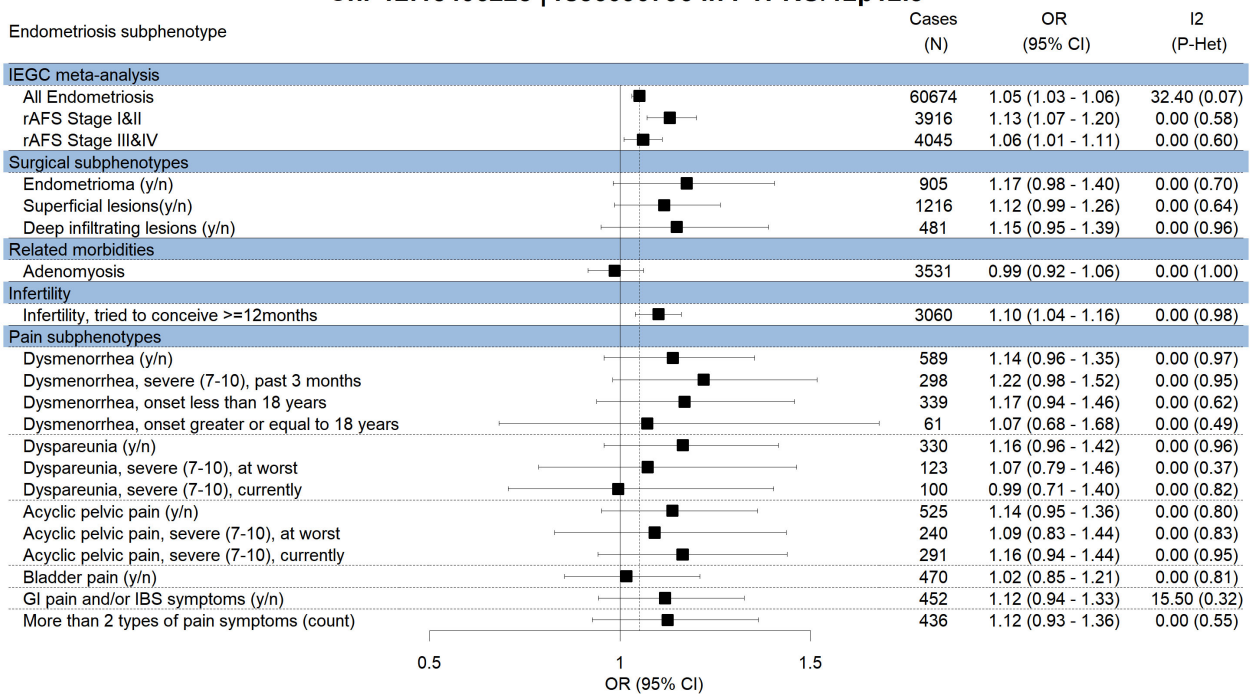
(xxix) FSHB/11p14.1 rs3858429



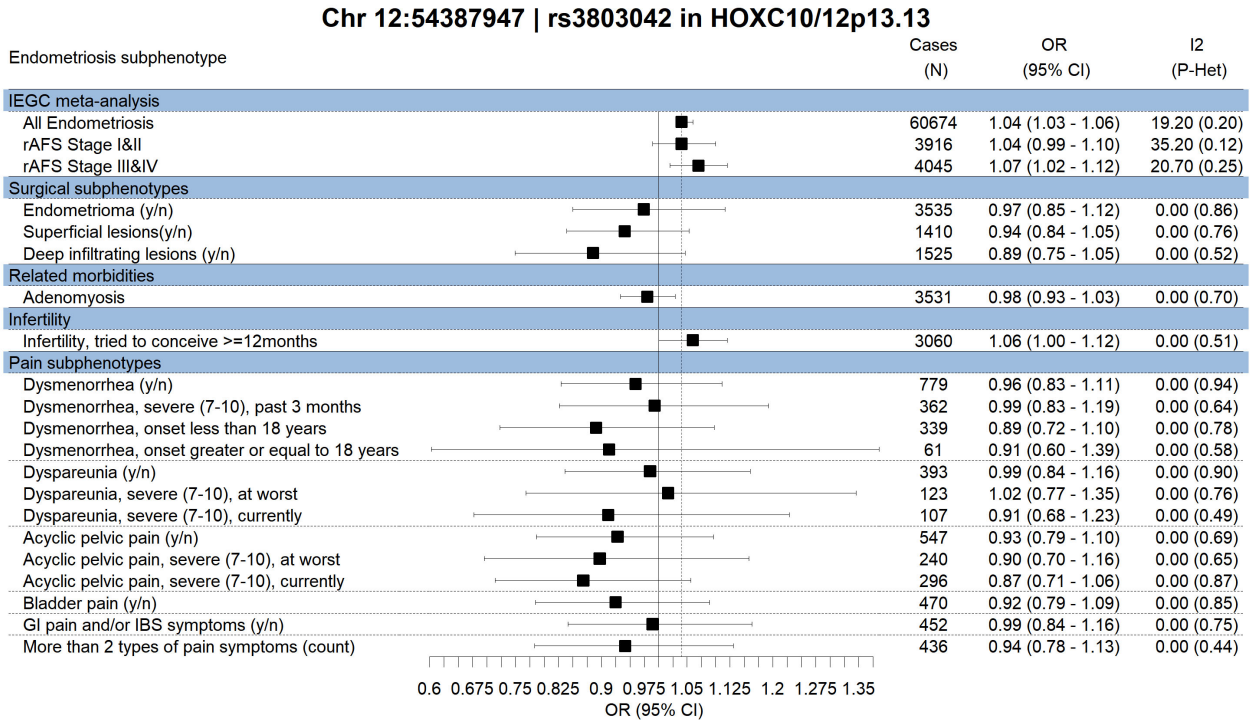
(xxx) VEZT/12q22 rs12320196



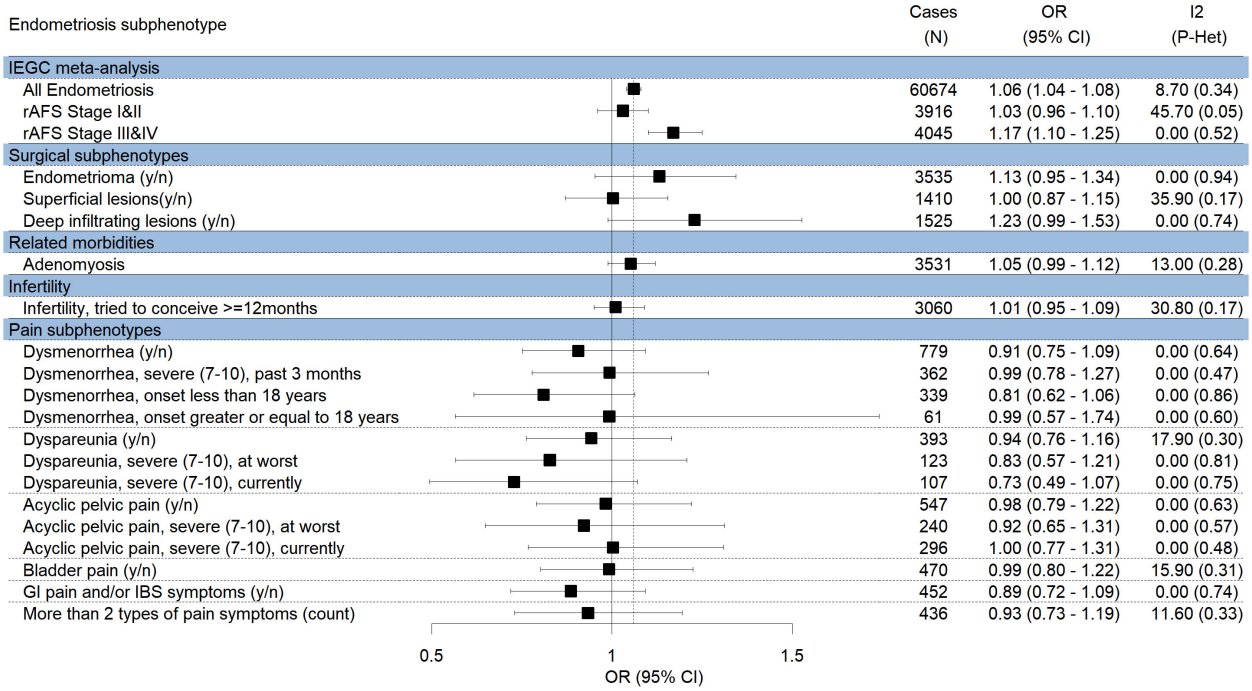
(xxxi) PTPRO/12p12.3



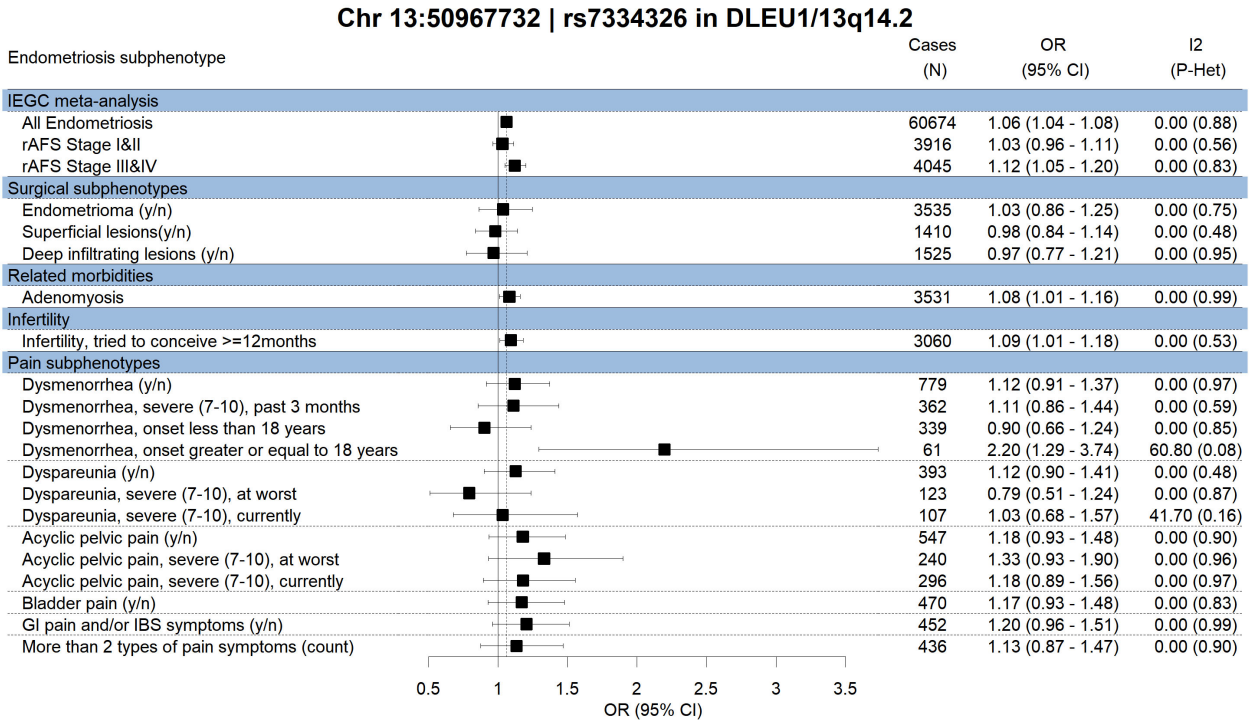
(xxxii) HOXC10/12p13.13



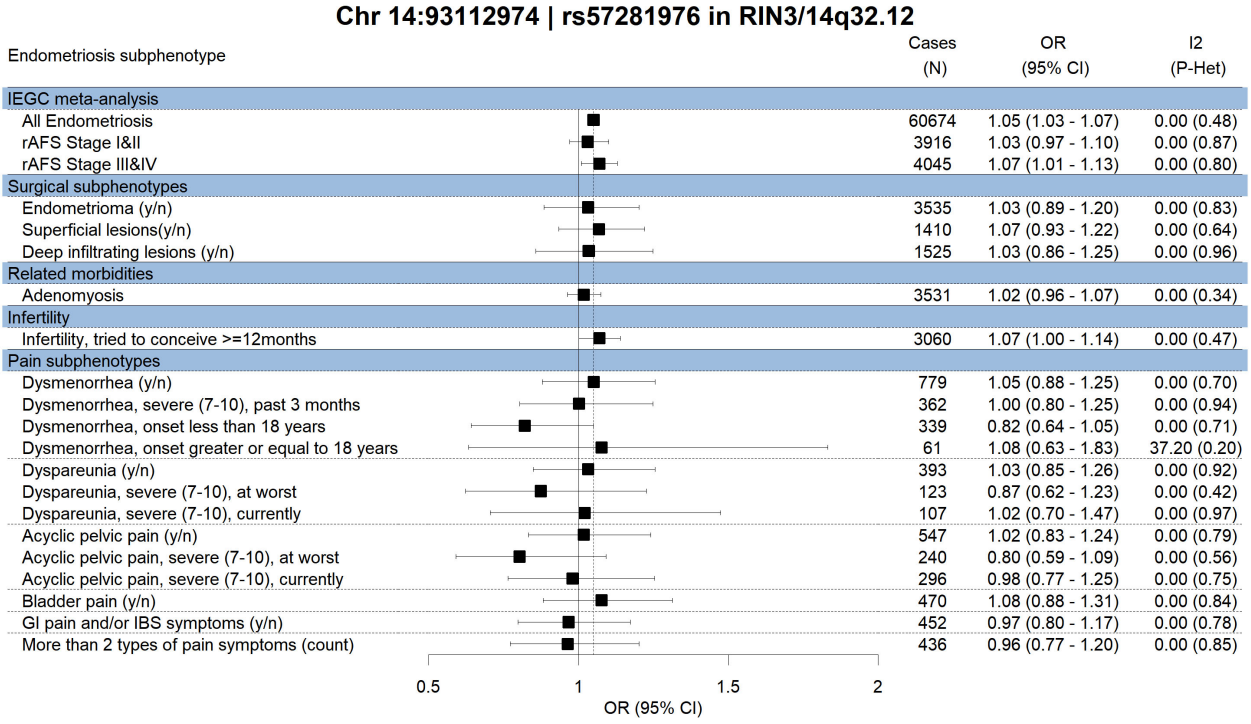
(xxxiii) IGF1/12q23.2



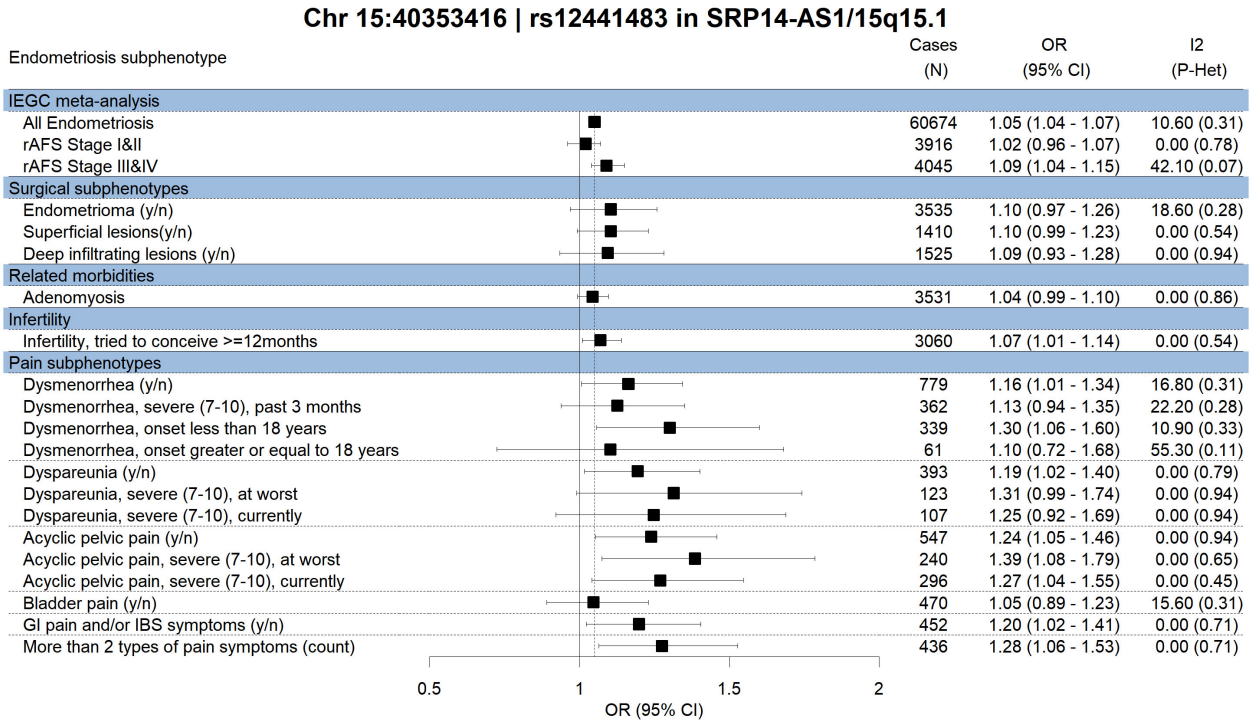
(xxxiv) DLEU1/13q14.2 rs7334326



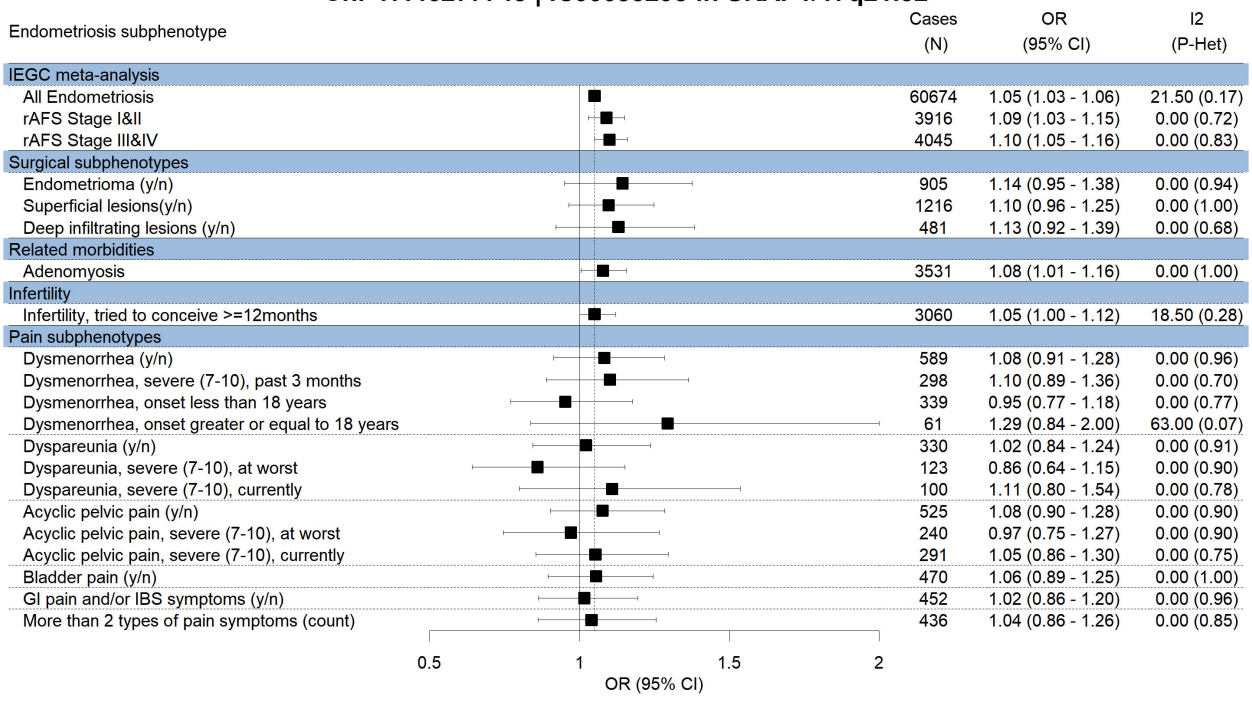
(xxxv) RIN3/14q32.12



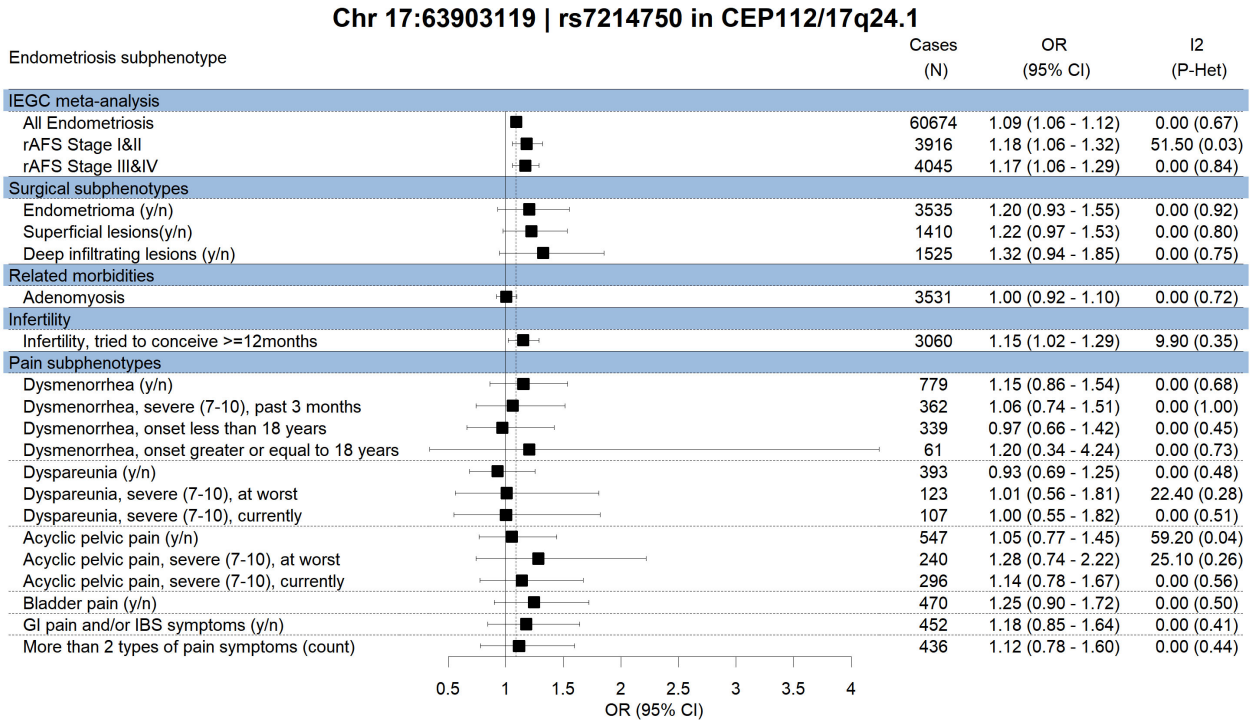
(xxxvi) SRP14-AS1/15q15.1 rs12441483



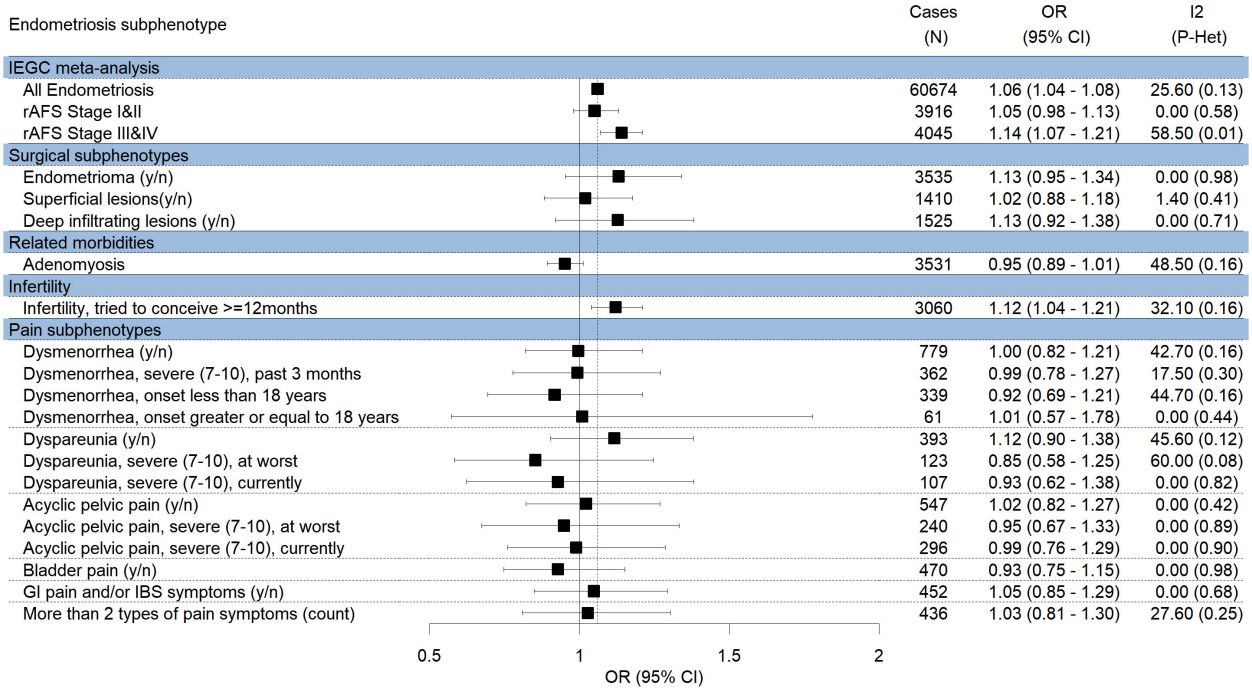
(xxxvii) SKAP1/17q21.32



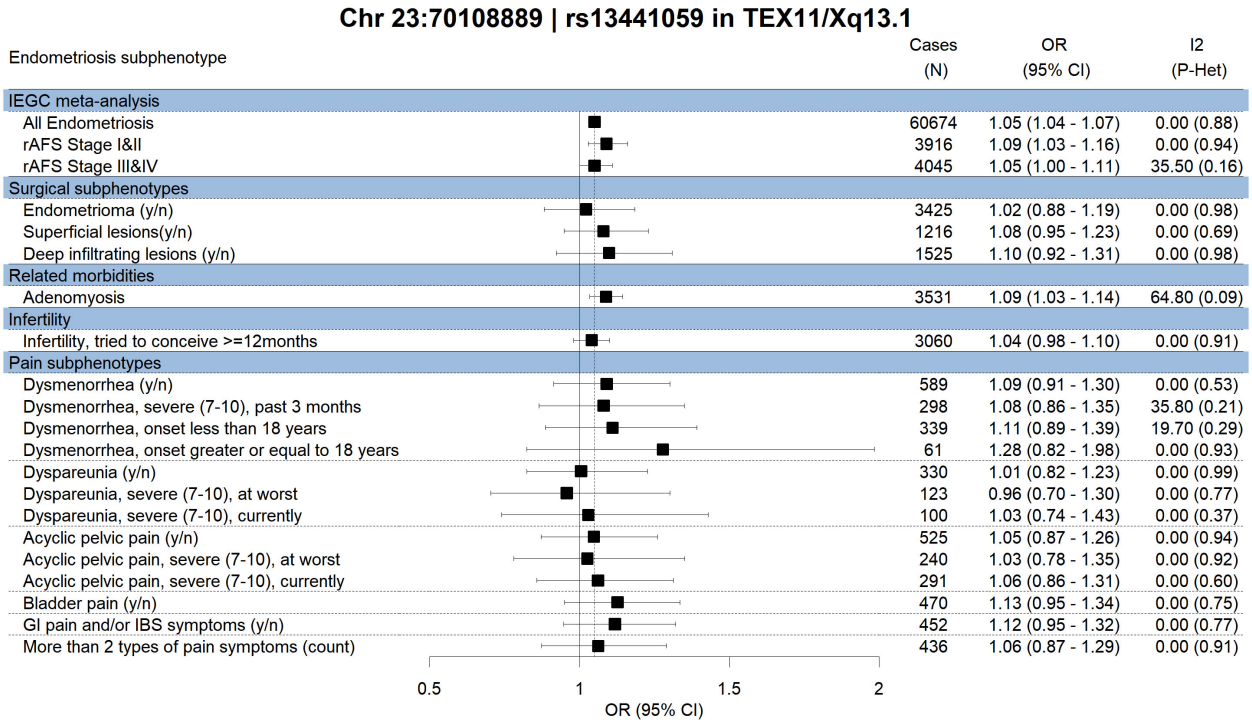
(xxxviii) CEP112/17q24.1



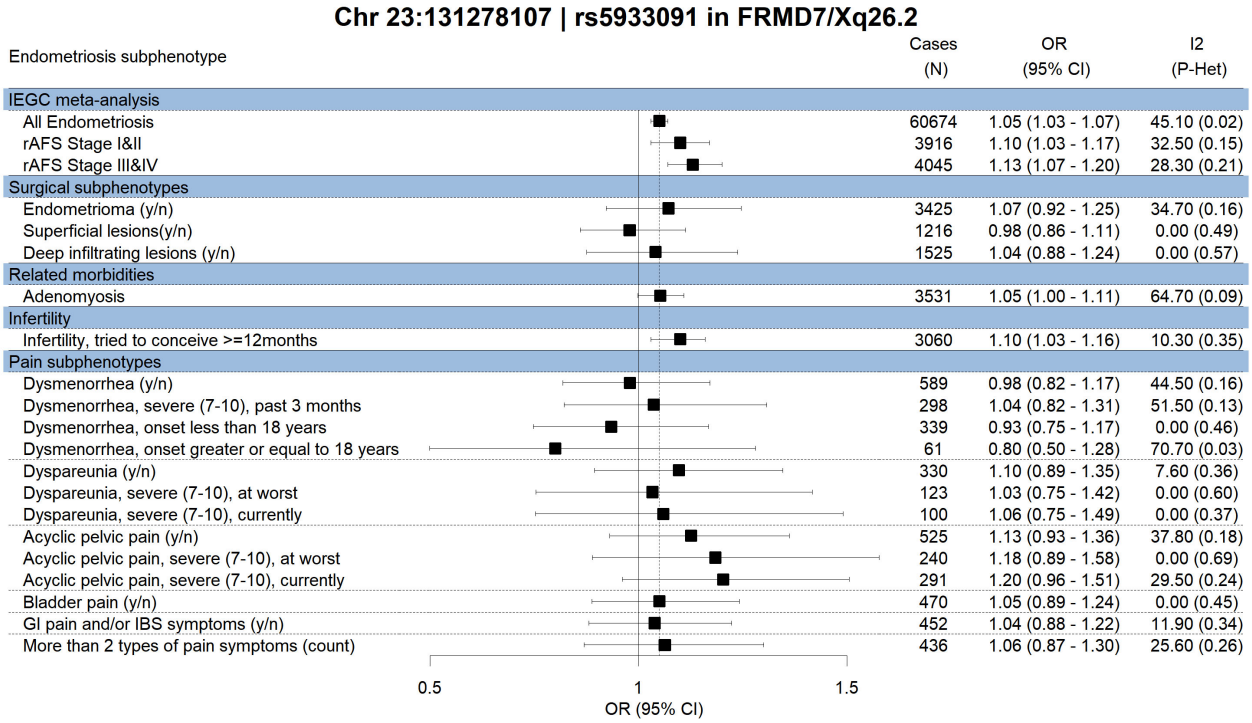
(xxxix) ACTL9/19p13.2



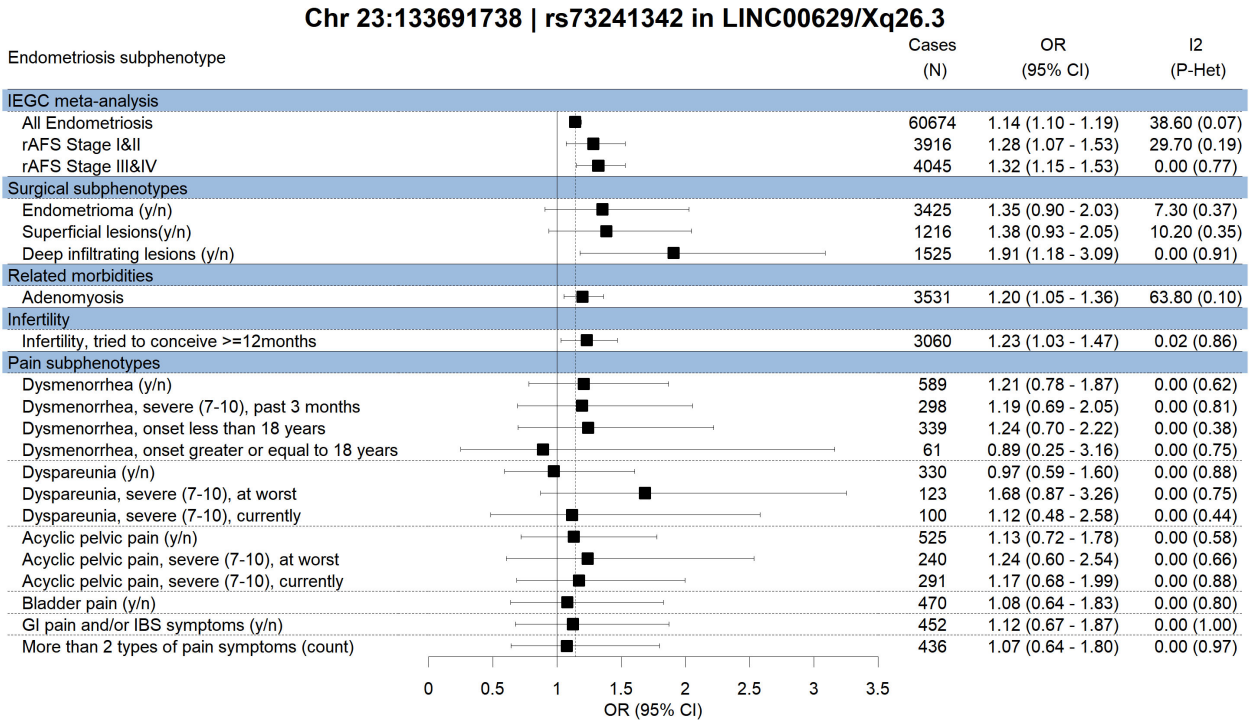
(xxxx) TEX11/Xq13.1



(xxxxi) FRMD7/Xq26.2

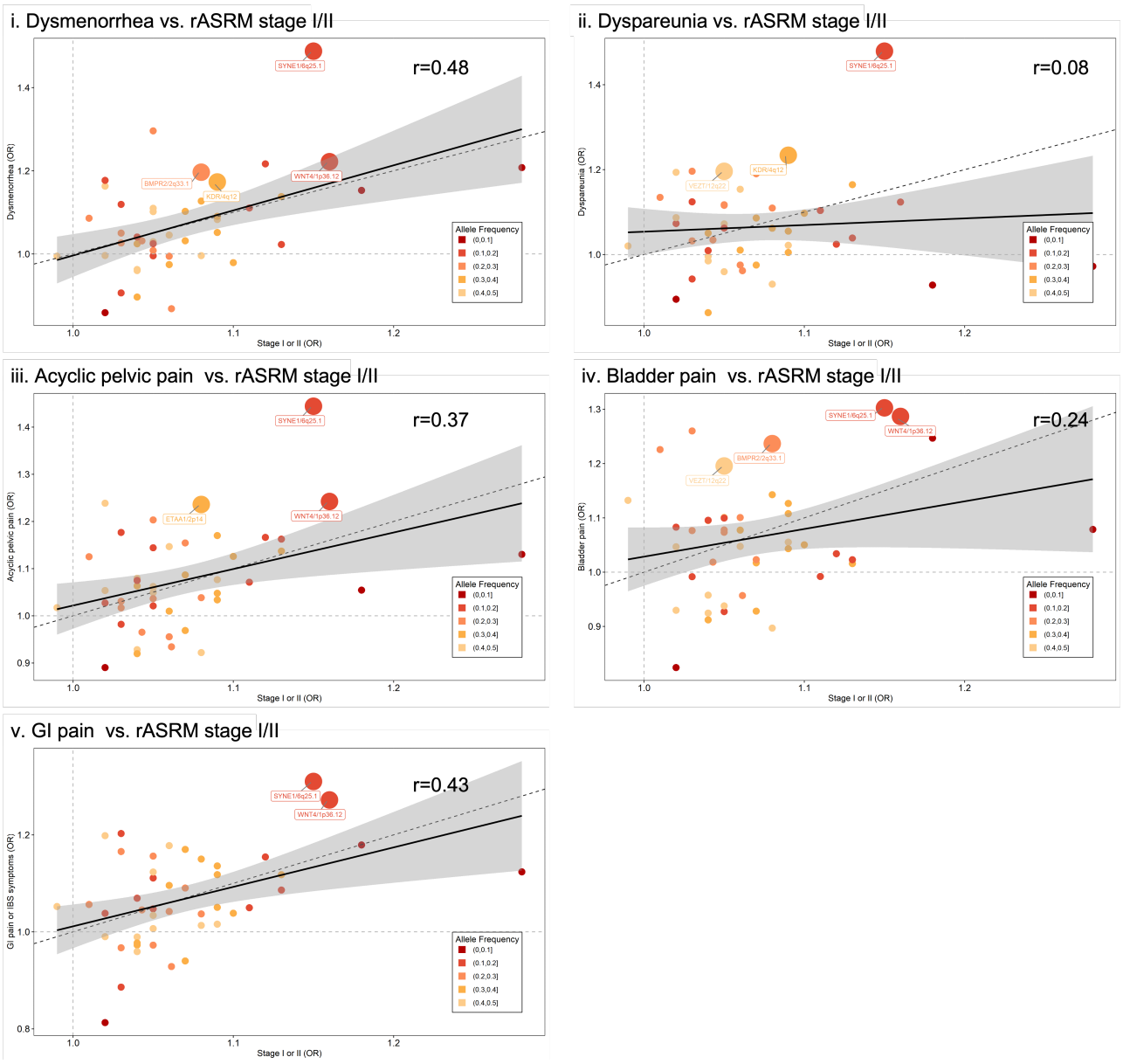


(xxxxii) LINC00629/Xq26.3 rs73241342

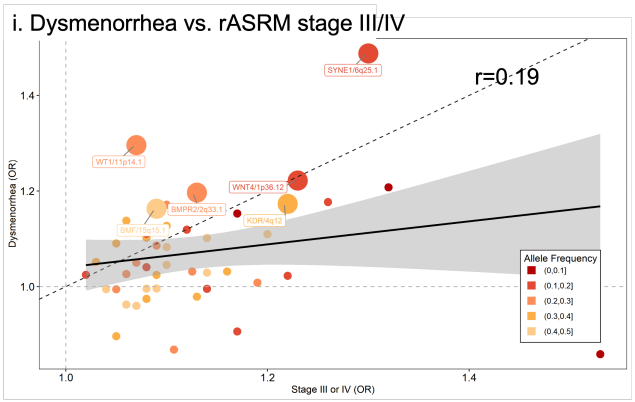


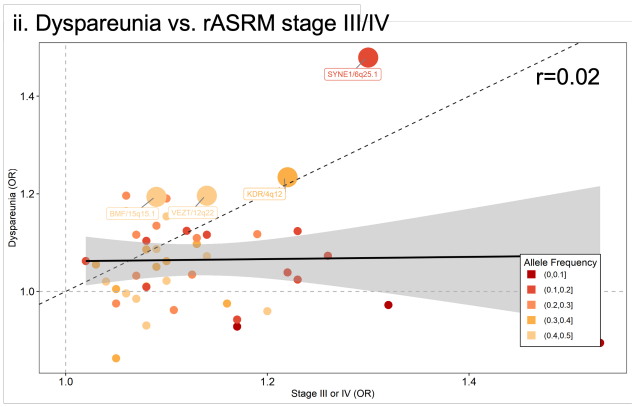
Supplementary Figure 8. Correlation between the effect sizes of 42 endometriosis associated loci between having surgical sub-types and experiencing pain symptomatology: a. rASRM stage I/II vs. 5 main pain symptomatology i. Dysmenorrhea, ii. dyspareunia, iii. non-cylic pain, iv. bladder pain, v. GI pain; b. rASRM stage III/IV vs. 5 main pain symptomatology i. Dysmenorrhea, ii. dyspareunia, iii. non-cylic pain, iv. bladder pain, v. GI pain; c. Deep lesions vs. 5 main pain symptomatology i. Dysmenorrhea, ii. dyspareunia, iii. non-cylic pain, iv. bladder pain, v. GI pain. Minor allele frequency for each of the 42 variants is given by shade of the red to yellow: Darker shade of red is smaller MAF, lighter shade of yellow is larger MAF. Nominal associations (p<0.05) are annotated with locus name and larger circles. Solid black line represents the linear regression line and dotted black line is the x=y with a slope of 1 for reference of change in ORs. The grey error band represents the 95% confidence interval. Test statistics including p-values for all the associations are provided in Supplementary Table 21.

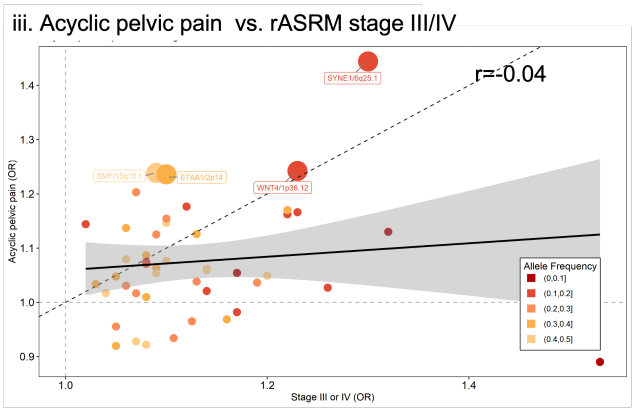
a. rASRM stage I/II vs. 5 main pain symptomatology

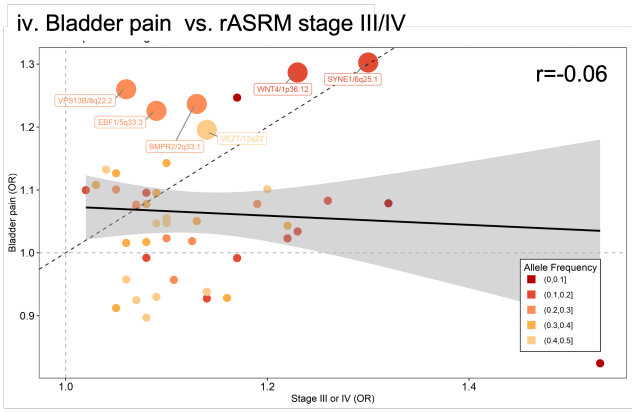


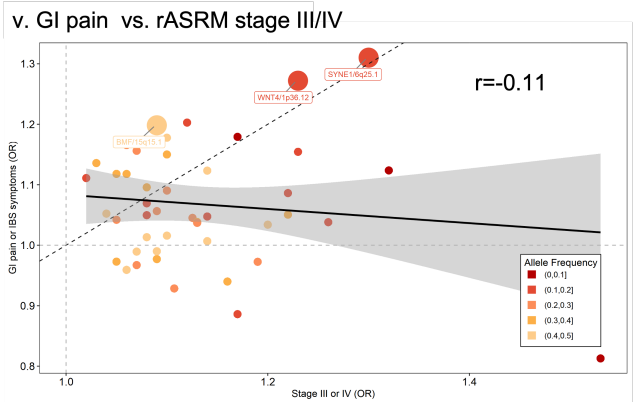
b. rASRM stage III/IV vs. 5 main pain symptomatology





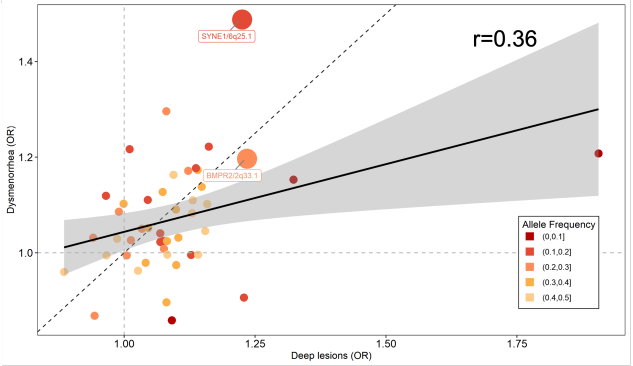




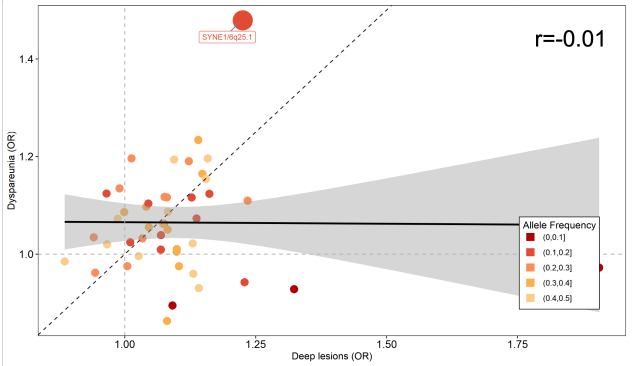


c. Deep lesions vs. 5 main pain symptomatology

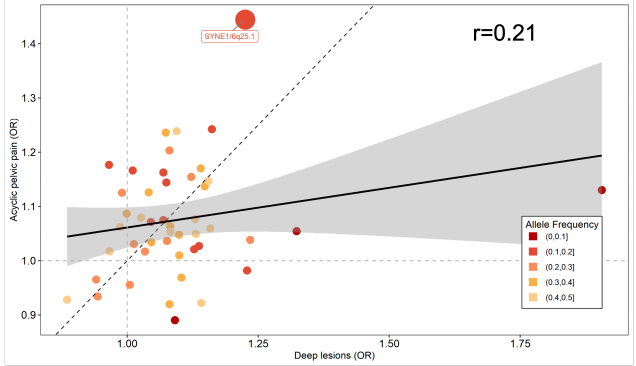
i. Dysmenorrhea vs. Deep lesions



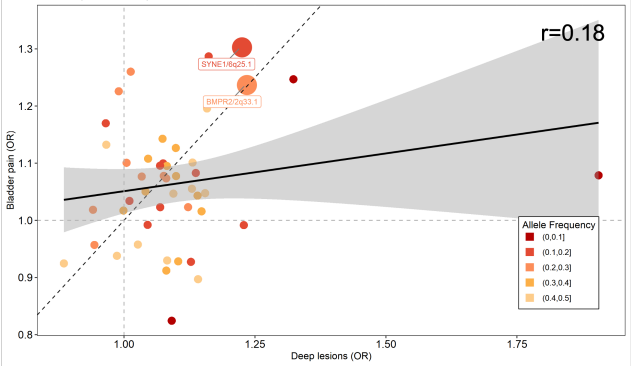
ii. Dyspareunia vs. Deep lesions



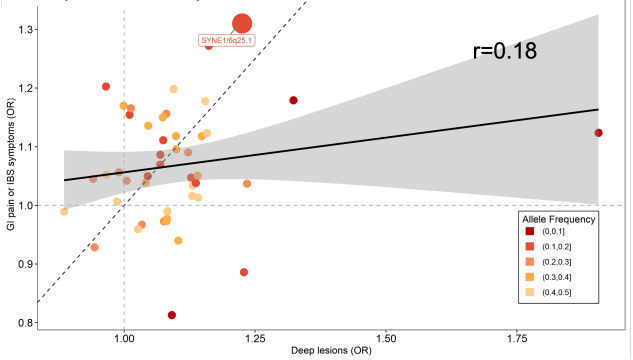
iii. Acyclic pelvic pain vs. Deep lesions



iv. Bladder pain vs. Deep lesions



v. GI pain vs. Deep lesions



References

1. Sapkota, Y. *et al.* Meta-analysis identifies five novel loci associated with endometriosis highlighting key genes involved in hormone metabolism. *Nat Commun* **8**, 15539 (2017).
2. Fung, J.N. *et al.* Genetic regulation of disease risk and endometrial gene expression highlights potential target genes for endometriosis and polycystic ovarian syndrome. *Sci Rep* **8**, 11424 (2018).
3. Mortlock, S. *et al.* Tissue specific regulation of transcription in endometrium and association with disease. *Hum Reprod* **35**, 377-393 (2020).
4. Consortium, G.T. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* **348**, 648-60 (2015).
5. Võsa, U. Unraveling the polygenic architecture of complex traits using blood eQTL meta-analysis. *BioRxiv* (2018).
6. Zhu, Z. *et al.* Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat Genet* **48**, 481-7 (2016).
7. McRae, A.F. *et al.* Identification of 55,000 Replicated DNA Methylation QTL. *Sci Rep* **8**, 17605 (2018).
8. Mortlock, S. *et al.* Genetic regulation of methylation in human endometrium and blood and gene targets for reproductive diseases. *Clin Epigenetics* **11**, 49 (2019).
9. Fung, J.N. *et al.* Functional evaluation of genetic variants associated with endometriosis near GREB1. *Hum Reprod* **30**, 1263-75 (2015).
10. Jones, A.V. *et al.* Genome-wide association analysis of pain severity in dysmenorrhea identifies association at chromosome 1p13.2, near the nerve growth factor locus. *Pain* **157**, 2571-2581 (2016).
11. Barneo-Munoz, M. *et al.* Lack of GDAP1 induces neuronal calcium and mitochondrial defects in a knockout mouse model of charcot-marie-tooth neuropathy. *PLoS Genet* **11**, e1005115 (2015).
12. Ruth, K.S. *et al.* Using human genetics to understand the disease impacts of testosterone in men and women. *Nat Med* **26**, 252-258 (2020).
13. Zhai, G. *et al.* Eight common genetic variants associated with serum DHEAS levels suggest a key role in ageing mechanisms. *PLoS Genet* **7**, e1002025 (2011).
14. Rahmani, A., Shoae-Hassani, A., Keyhanvar, P., Kheradmand, D. & Darbandi-Azar, A. Dehydroepiandrosterone stimulates nerve growth factor and brain derived neurotrophic factor in cortical neurons. *Adv Pharmacol Sci* **2013**, 506191 (2013).
15. Maninger, N., Wolkowitz, O.M., Reus, V.I., Epel, E.S. & Mellon, S.H. Neurobiological and neuropsychiatric effects of dehydroepiandrosterone (DHEA) and DHEA sulfate (DHEAS). *Front Neuroendocrinol* **30**, 65-91 (2009).
16. Obata, K. & Noguchi, K. BDNF in sensory neurons and chronic pain. *Neurosci Res* **55**, 1-10 (2006).
17. Browne, A.S. *et al.* Proteomic identification of neurotrophins in the eutopic endometrium of women with endometriosis. *Fertil Steril* **98**, 713-9 (2012).
18. Wang, S. *et al.* BDNF and TrKB expression levels in patients with endometriosis and their associations with dysmenorrhoea. *J Ovarian Res* **15**, 35 (2022).
19. Peng, B., Alotaibi, F.T., Sediqi, S., Bedaiwy, M.A. & Yong, P.J. Role of interleukin-1beta in nerve growth factor expression, neurogenesis and deep dyspareunia in endometriosis. *Hum Reprod* **35**, 901-912 (2020).
20. Contarelli, S., Fedele, V. & Melisi, D. HOX Genes Family and Cancer: A Novel Role for Homeobox B9 in the Resistance to Anti-Angiogenic Therapies. *Cancers (Basel)* **12**(2020).
21. Hayashida, T. *et al.* HOXB9, a gene overexpressed in breast cancer, promotes tumorigenicity and lung metastasis. *Proc Natl Acad Sci U S A* **107**, 1100-5 (2010).
22. Wan, J., Liu, H., Feng, Q., Liu, J. & Ming, L. HOXB9 promotes endometrial cancer progression by targeting E2F3. *Cell Death Dis* **9**, 509 (2018).
23. http://www.nealelab.is/uk-biobank/.
24. Pare, G. *et al.* Novel association of ABO histo-blood group antigen with soluble ICAM-1: results of a genome-wide association study of 6,578 women. *PLoS Genet* **4**, e1000118 (2008).
25. Bourdel, N. *et al.* Systematic review of endometriosis pain assessment: how to choose a scale? *Hum Reprod Update* **21**, 136-52 (2015).

26. Choi, E.J. *et al.* Comorbidity of gynecological and non-gynecological diseases with adenomyosis and endometriosis. *Obstet Gynecol Sci* **60**, 579-586 (2017).
27. Becker, C.M. *et al.* World Endometriosis Research Foundation Endometriosis Phenome and Biobanking Harmonisation Project: I. Surgical phenotype data collection in endometriosis research. *Fertil Steril* **102**, 1213-22 (2014).
28. Fassbender, A. *et al.* World Endometriosis Research Foundation Endometriosis Phenome and Biobanking Harmonisation Project: IV. Tissue collection, processing, and storage in endometriosis research. *Fertil Steril* **102**, 1244-53 (2014).
29. Rahmioglu, N. *et al.* World Endometriosis Research Foundation Endometriosis Phenome and Biobanking Harmonization Project: III. Fluid biospecimen collection, processing, and storage in endometriosis research. *Fertil Steril* **102**, 1233-43 (2014).
30. Das, S. *et al.* Next-generation genotype imputation service and methods. *Nat Genet* **48**, 1284-1287 (2016).
31. Marchini, J., Howie, B., Myers, S., McVean, G. & Donnelly, P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat Genet* **39**, 906-13 (2007).
32. Kang, H.M. Efficient and parallelizable association container toolbox (EPACTS). *University of Michigan Center for Statistical Genetics Accessed. 2014;6:16* (2014).
33. Loh, P.R. *et al.* Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nat Genet* **47**, 284-90 (2015).