

Supplementary Information

Supplementary Figure 1. Flowchart of the study design.

Supplementary Figure 2. QQ plot and genomic inflation factors (λ) observed for the combined (GERA+UKB) multiethnic meta-analysis of cataract.

Supplementary Figure 3. Correlation of effect sizes for cataract between GERA and UKB cohorts for the lead 54 cataract-associated lead SNPs identified in the combined (GERA+UKB) multiethnic meta-analysis.

Supplementary Figure 4. Correlation of effect sizes for cataract between the meta-analysis (GERA+UKB) results and 23andMe dataset for the lead 54 cataract-associated lead SNPs identified in the combined (GERA+UKB) multiethnic meta-analysis.

Supplementary Figure 5. Manhattan plots of the GWAS of cataract (GERA + UKB) stratified by ethnic groups. **a.** European ancestry; **b.** East Asian ancestry; and **c.** African ancestry.

Supplementary Figure 6. LocusZoom plots of regions identified in the combined (GERA + UKB) meta-analysis of cataract stratified by ethnicity. We identified three additional novel regions specific to individuals of European ancestry: **a.** *EPHA4*, **b.** *CD83-JARID2*, and **c.** *EXOC3L2*.

Supplementary Figure 7. LocusZoom plots of regions identified in the sex-specific (GERA+UKB) analyses that show differential association with cataract across women and men. The following regions were significant in women ($P < 5 \times 10^{-8}$) but not significant ($P > 0.05$) in men: **a.** *GSTM2*, **b.** *DNMBP-CPN1*, and **c.** *CASP7*; the following regions were genome-wide significant in men but not in women: **d.** *QKI*, **e.** *SEMA4D*, **f.** *RBFOX1*, and **g.** *JAG1*.

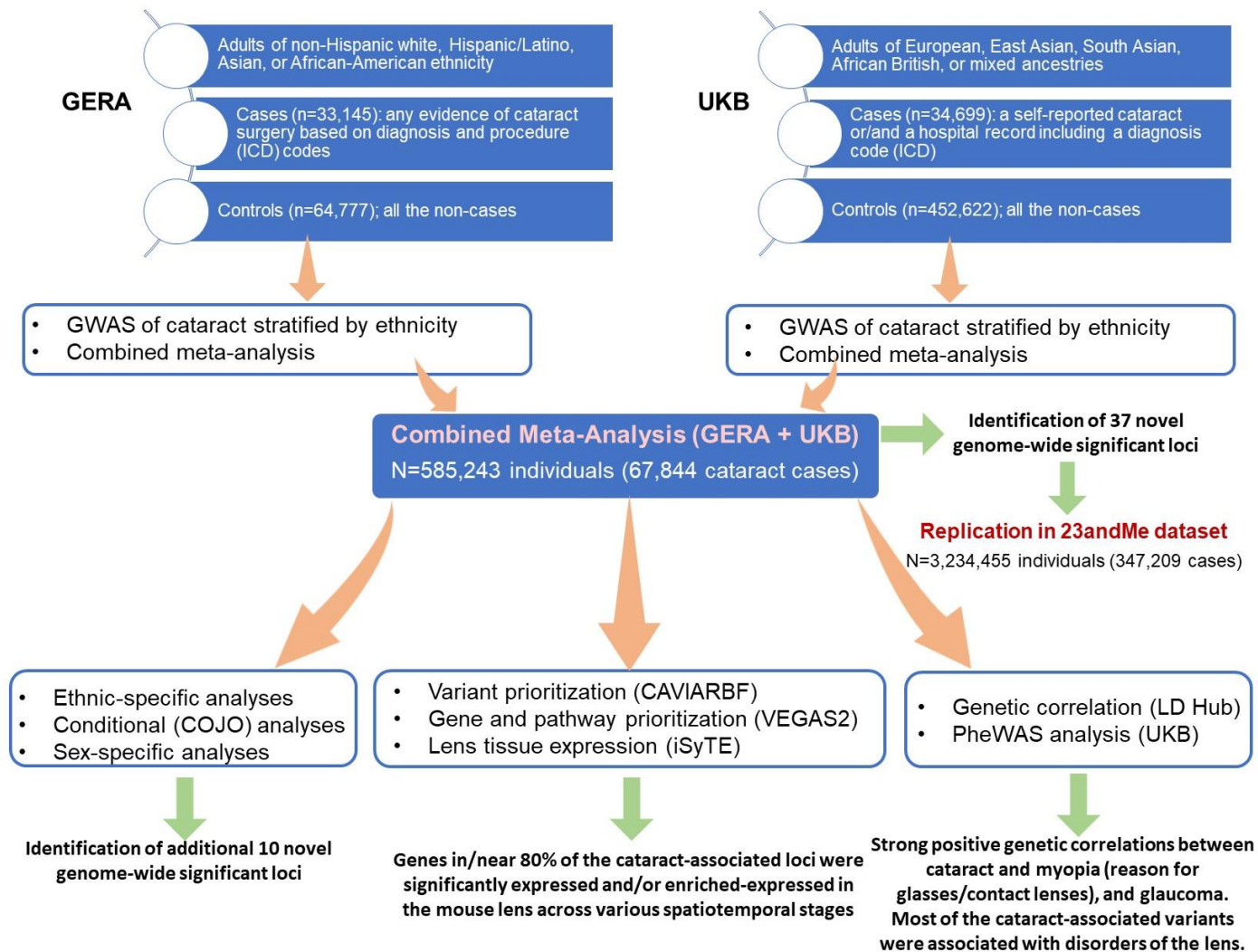
Supplementary Figure 8. Correlation of effect sizes for cataract between women and men for the lead 54 SNPs identified in the combined (GERA+UKB) GWAS multiethnic analysis and for the additional SNPs identified in the sex-specific analyses.

Supplementary Figure 9. Enriched expression of candidate genes in mouse lens.

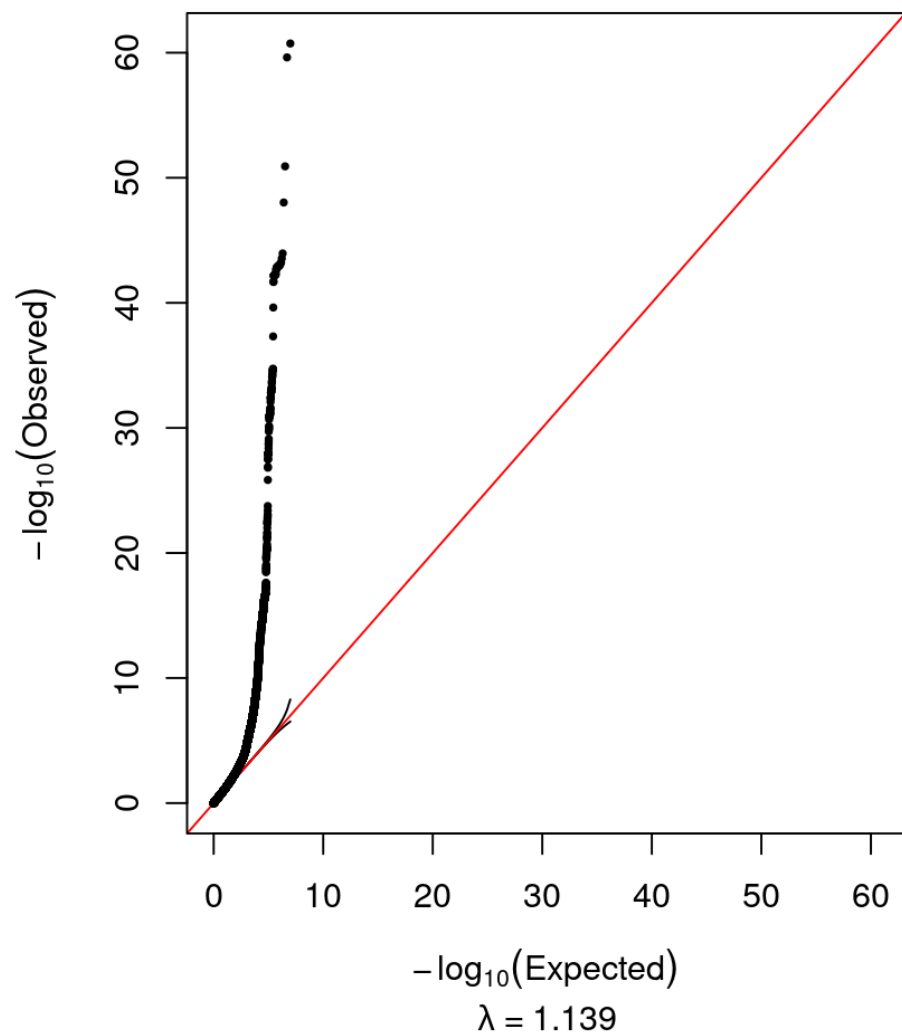
Supplementary Figure 10. Expression of candidate genes in lenses of gene perturbation mouse models with lens defects/cataract.

Supplementary Figure 11. RT-PCR based validation of candidate gene expression in the lens

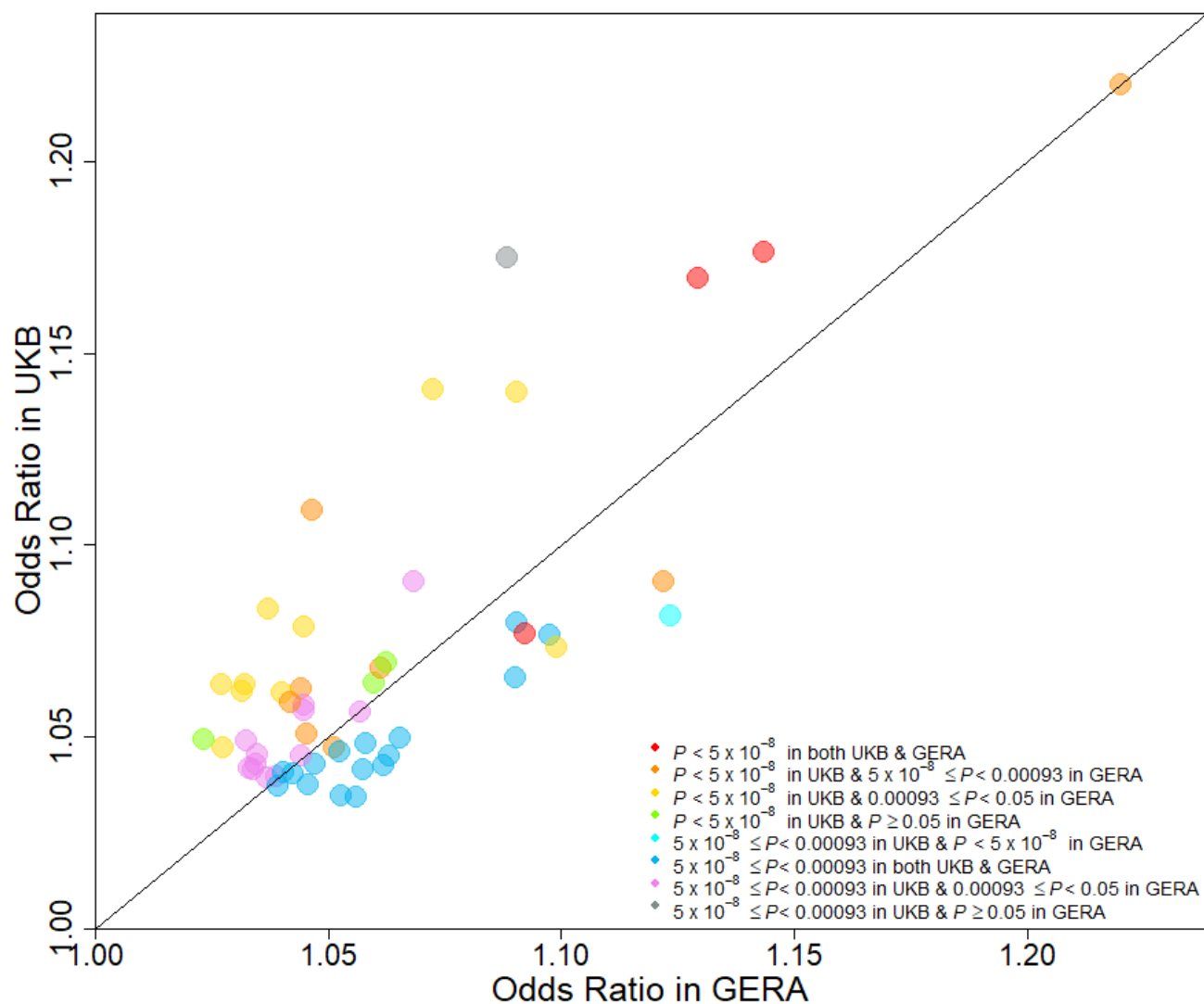
Supplementary Figure 1. Flowchart of the study design



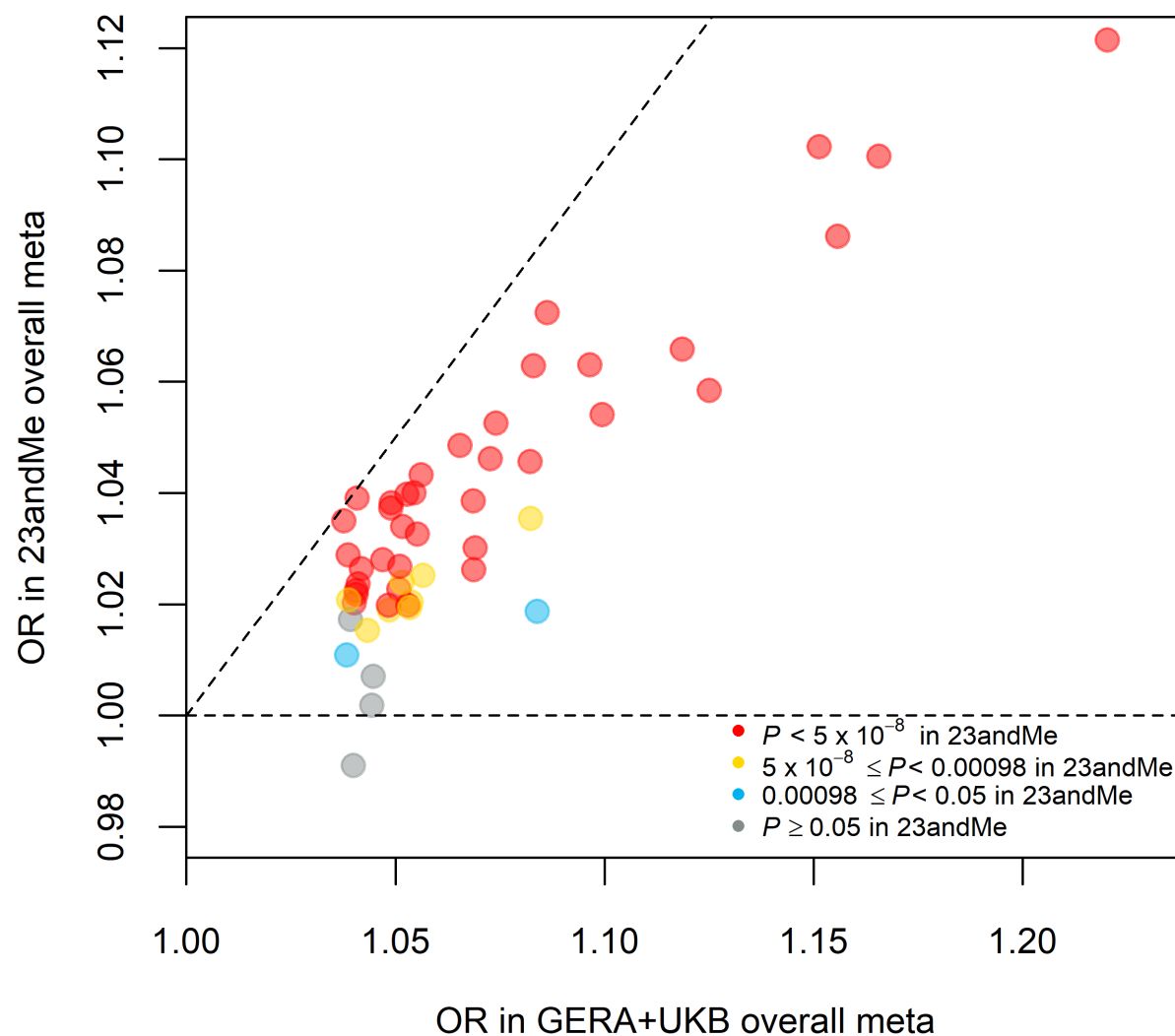
Supplementary Figure 2. QQ plot and genomic inflation factors (λ) observed for the combined (GERA+UKB) multiethnic meta-analysis of cataract



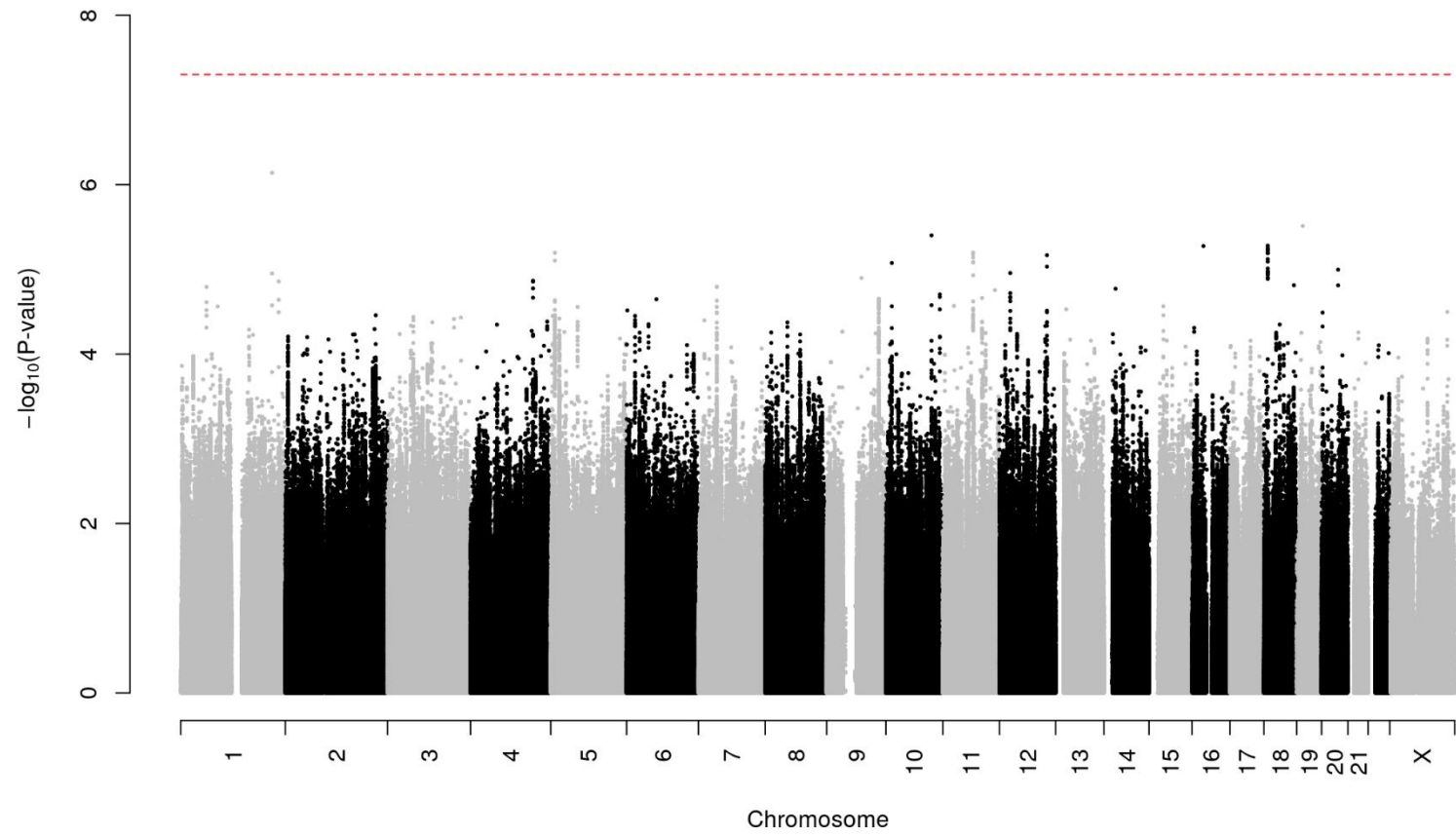
Supplementary Figure 3. Correlation of effect sizes for cataract between GERA and UKB cohorts for the lead 54 cataract-associated lead SNPs identified in the combined (GERA+UKB) multiethnic meta-analysis. The effect sizes were compared using a correlation test, two-sided; P -value for the correlation test is 4.1×10^{-12} .



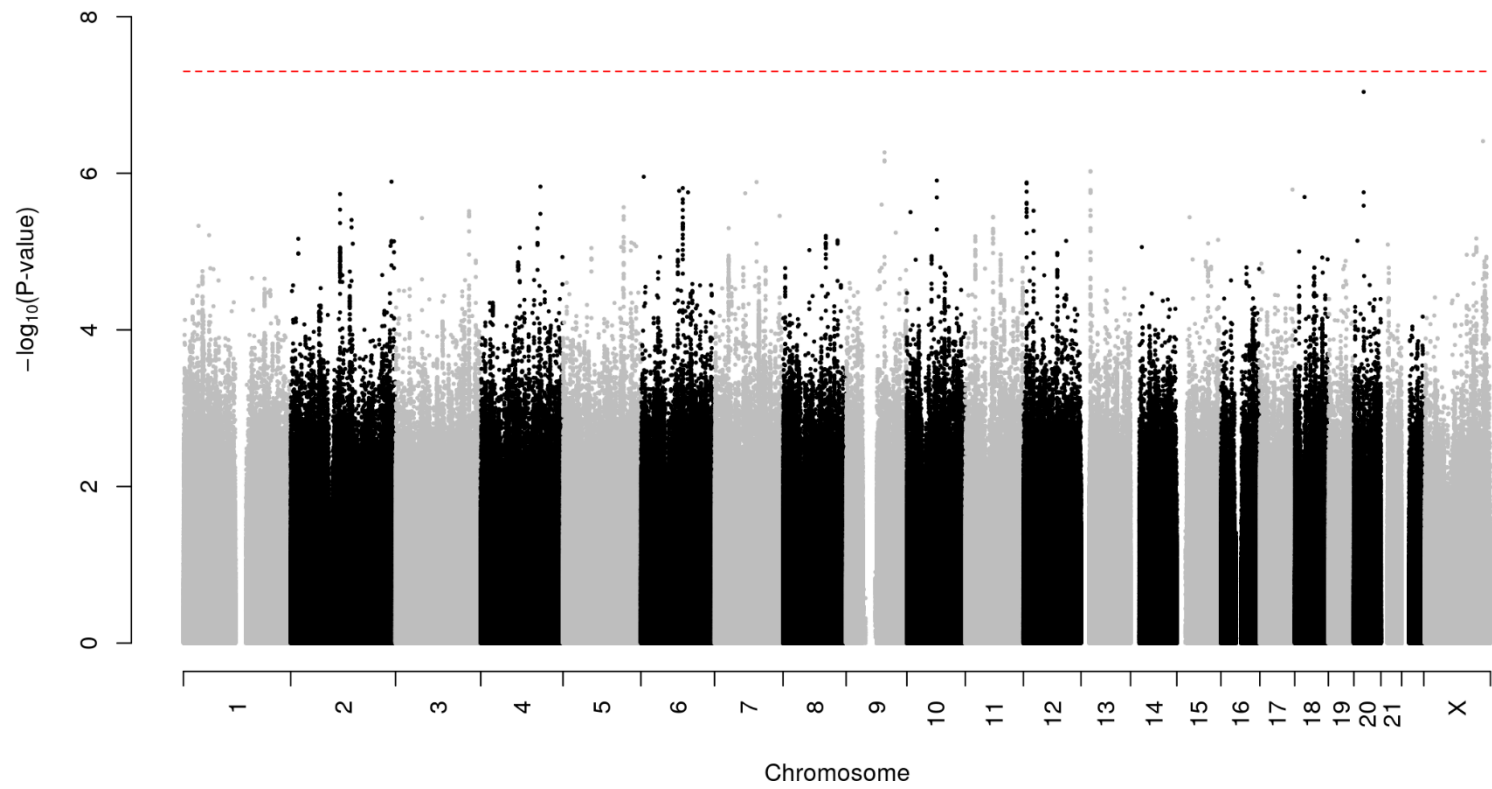
Supplementary Figure 4. Correlation of effect sizes for cataract between the meta-analysis (GERA+UKB) results and 23andMe dataset for the lead 54 cataract-associated lead SNPs identified in the combined (GERA+UKB) multiethnic meta-analysis. The effect sizes were compared using a correlation test, two-sided; P -value for the correlation test is 1.0×10^{-14} .



b.

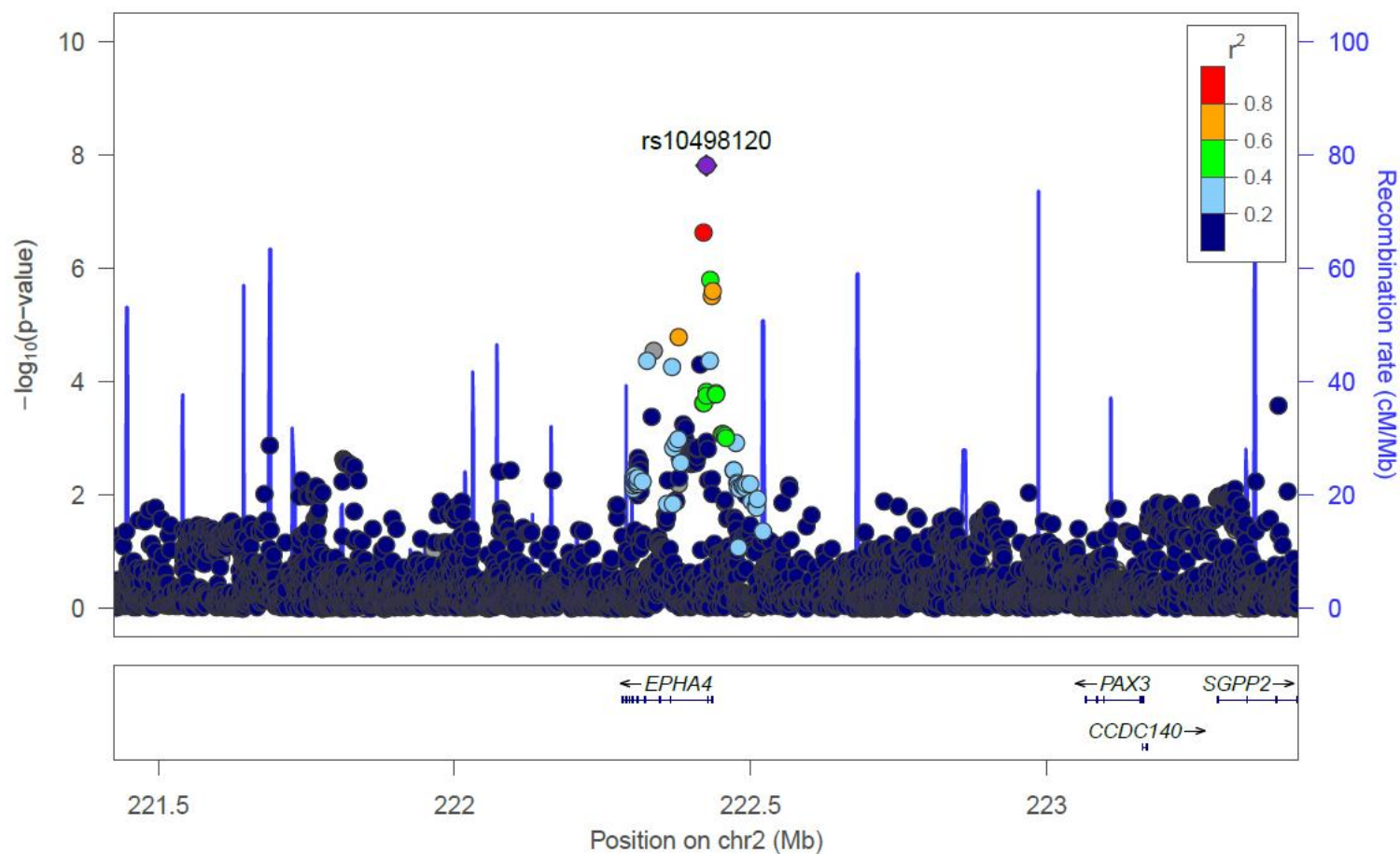


c.

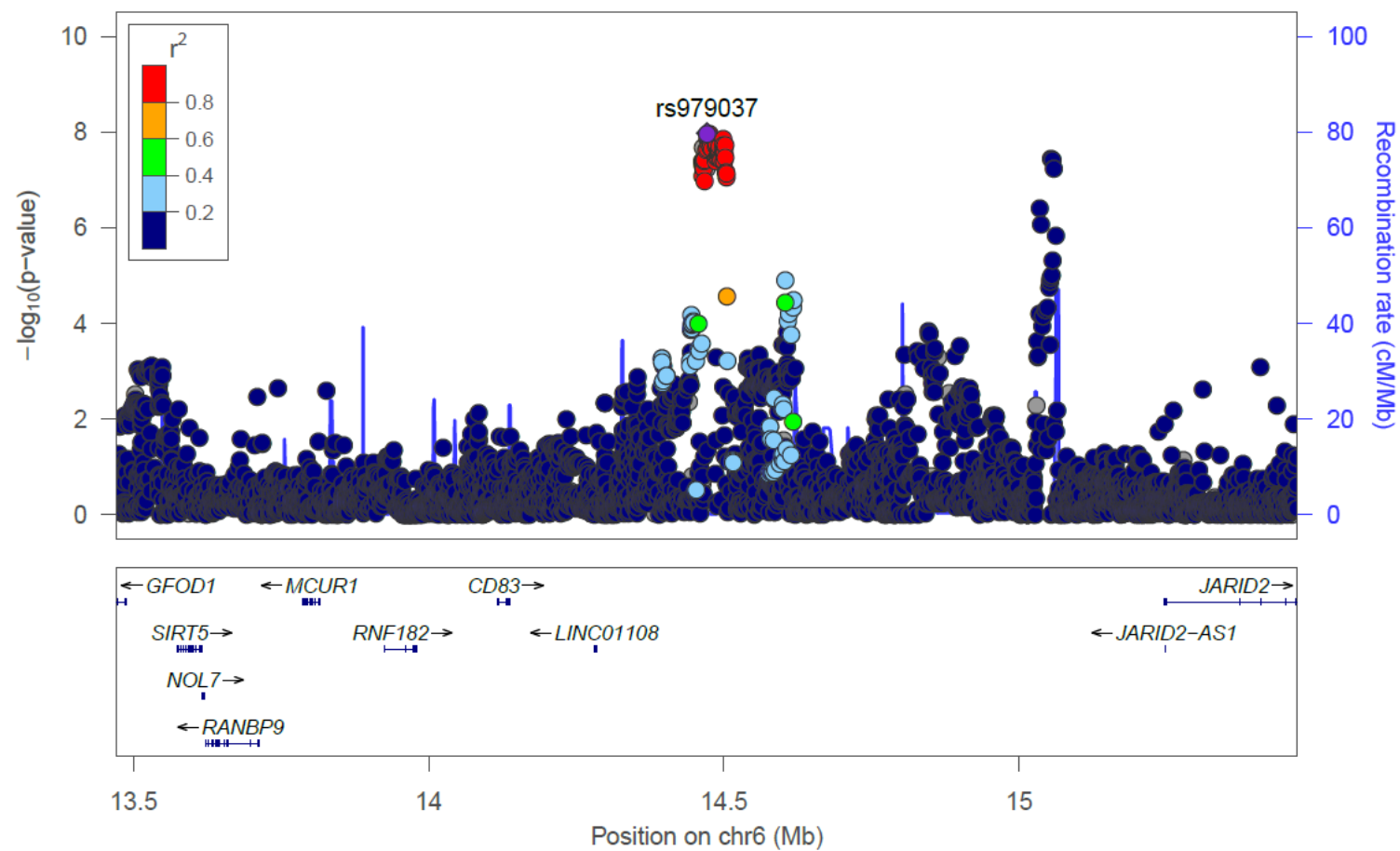


Supplementary Figure 6. LocusZoom plots of regions identified in the combined (GERA + UKB) meta-analysis of cataract stratified by ethnicity. We identified three additional novel regions specific to individuals of European ancestry: **a.** *EPHA4*, **b.** *CD83-JARID2*, and **c.** *EXOC3L2*

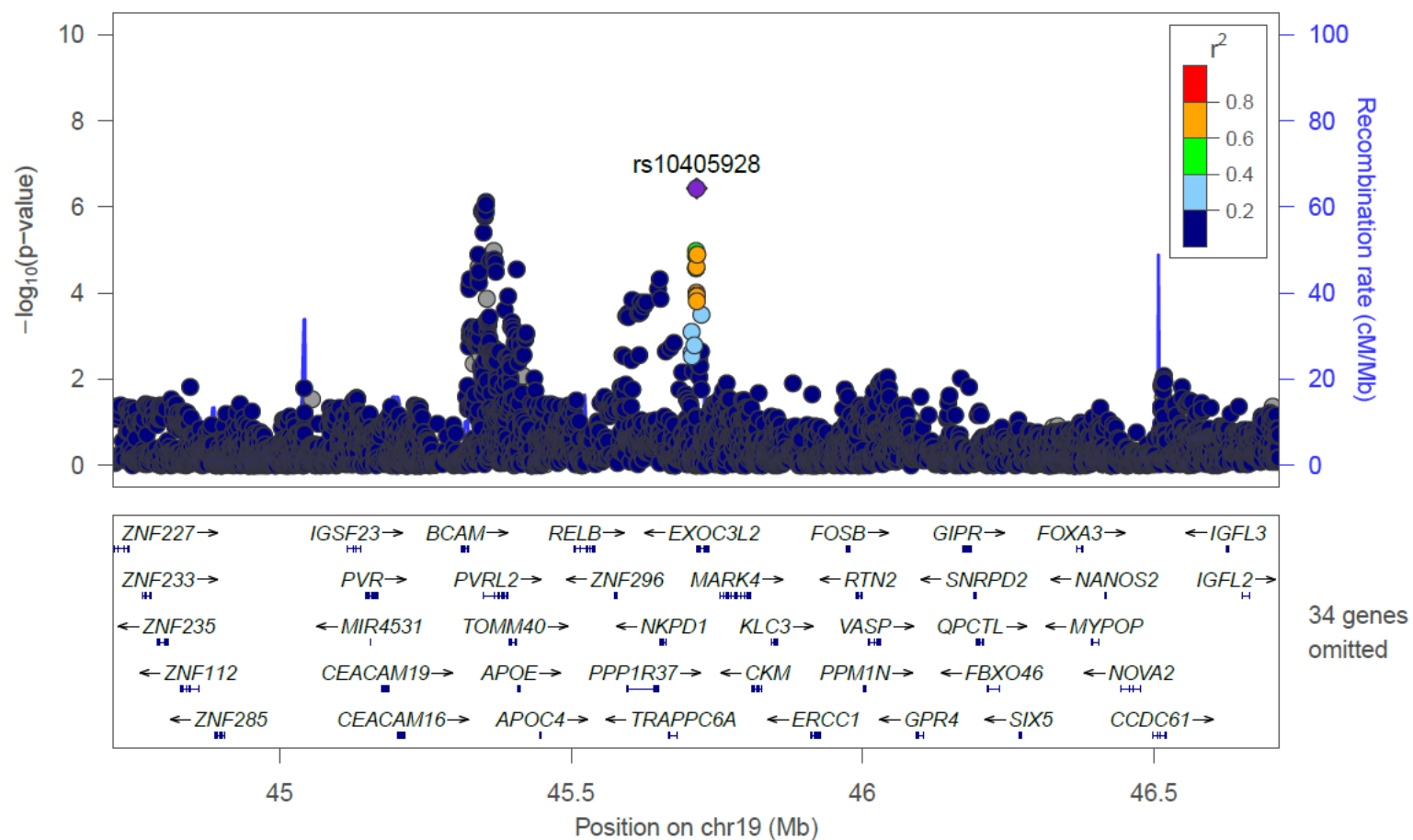
a.



b.

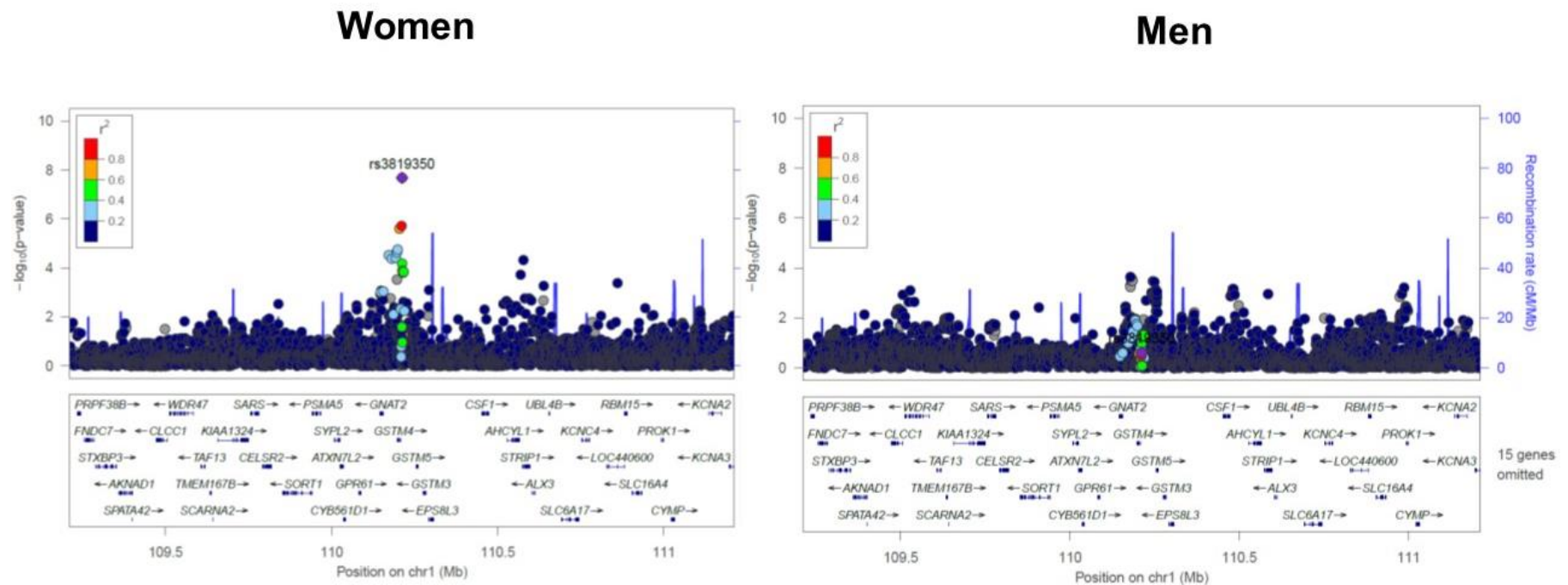


c.



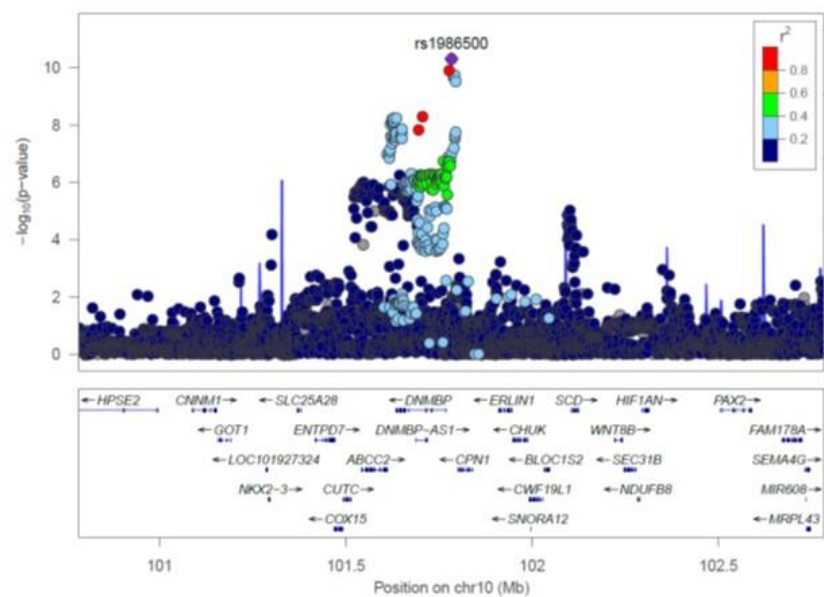
Supplementary Figure 7. LocusZoom plots of regions identified in the sex-specific (GERA+UKB) analyses that show differential association with cataract across women and men. The following regions were significant in women ($P < 5 \times 10^{-8}$) but not significant ($P > 0.05$) in men: **a. *GSTM2***, **b. *DNMBP-CPN1***, and **c. *CASP7***; the following regions were genome-wide significant in men but not in women: **d. *QKI***, **e. *SEMA4D***, **f. *RBFOX1***, and **g. *JAG1***. The y-axis represents the $-\log_{10}(P\text{-value})$; all P -values derived from logistic regression model are two-sided.

a.

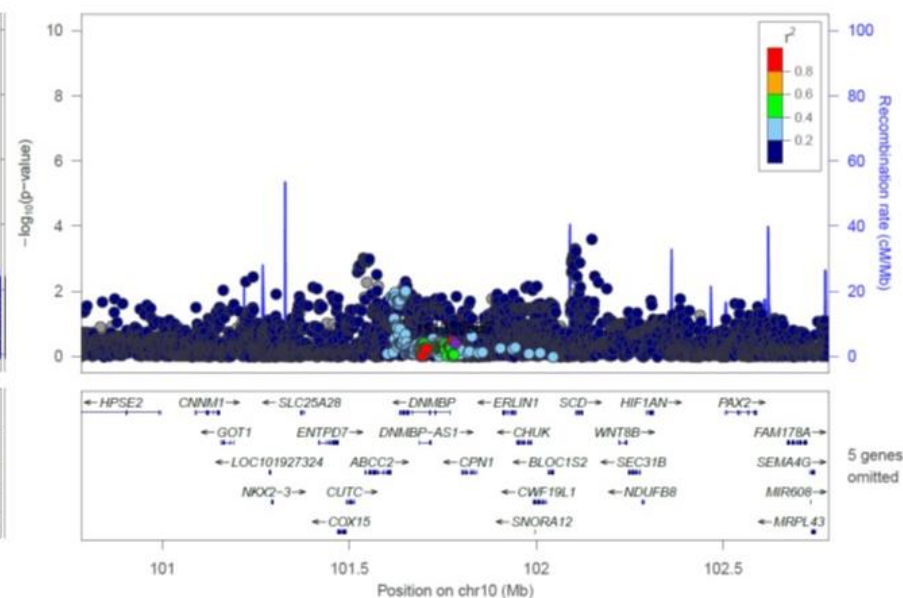


b.

Women

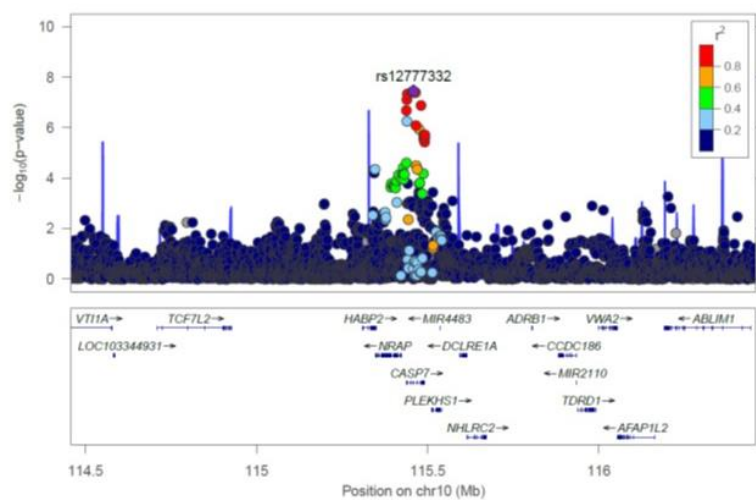


Men

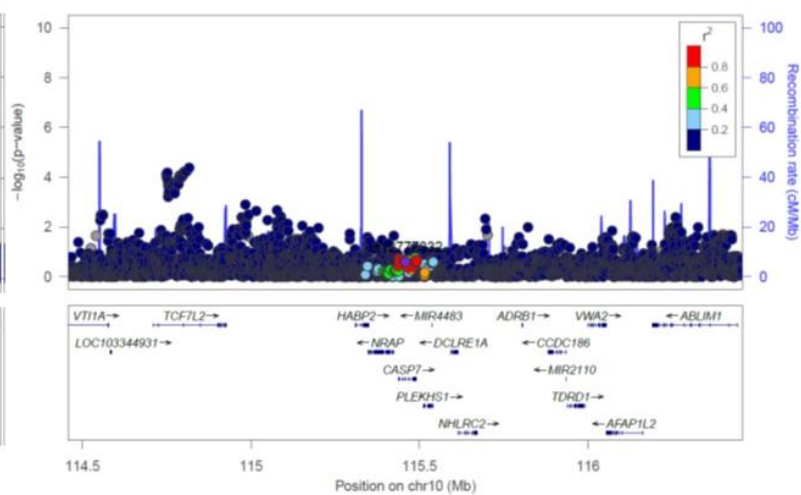


c.

Women

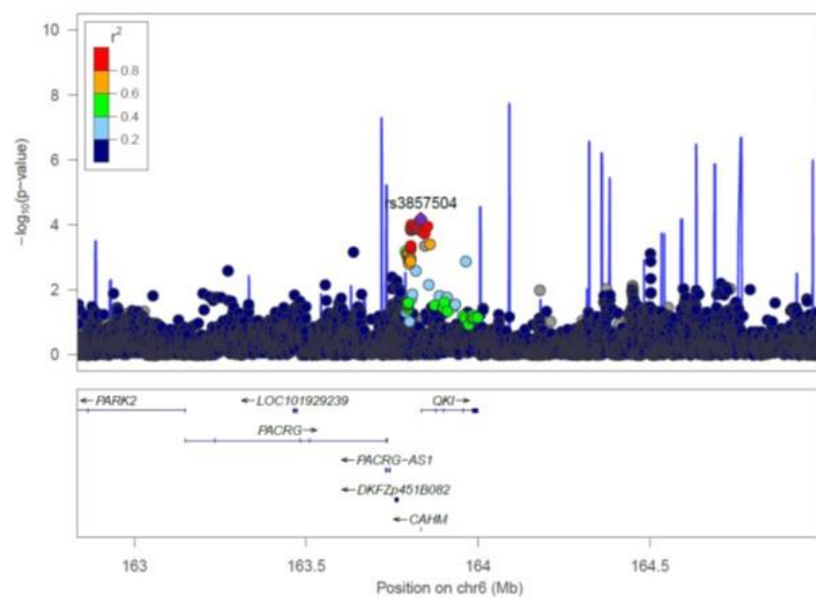


Men

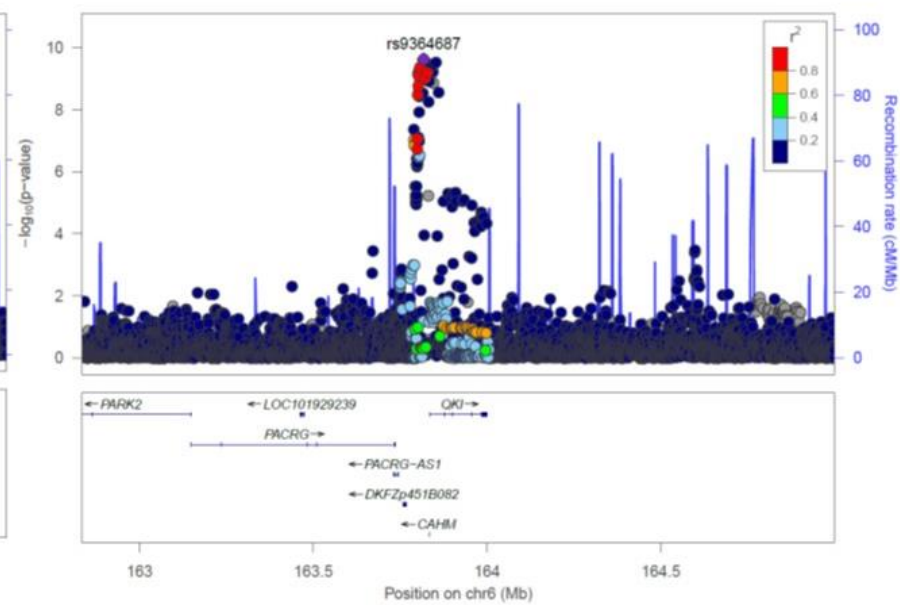


d.

Women

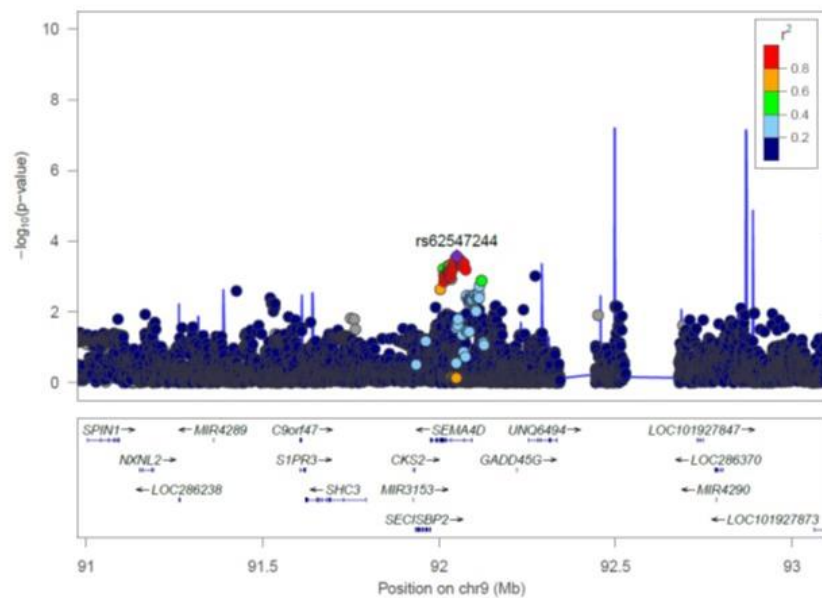


Men

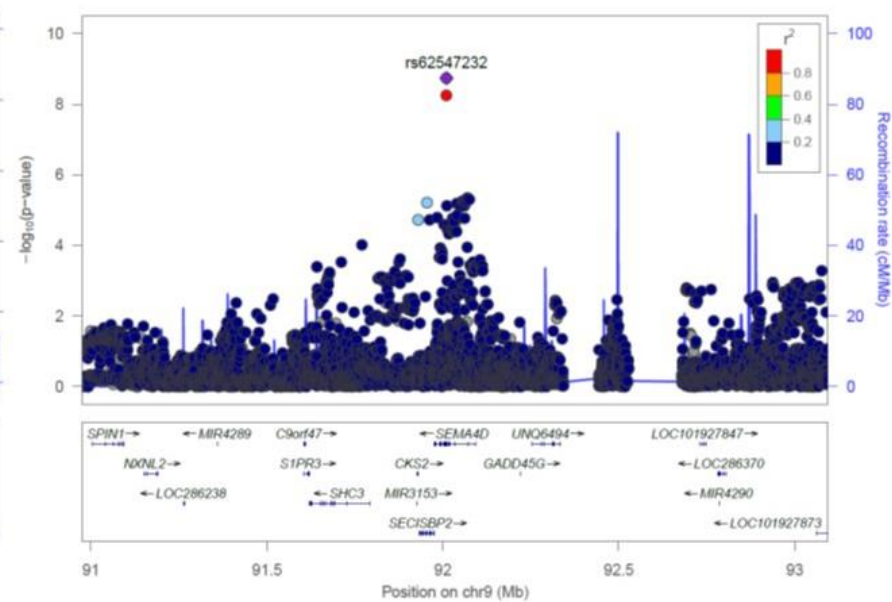


e.

Women

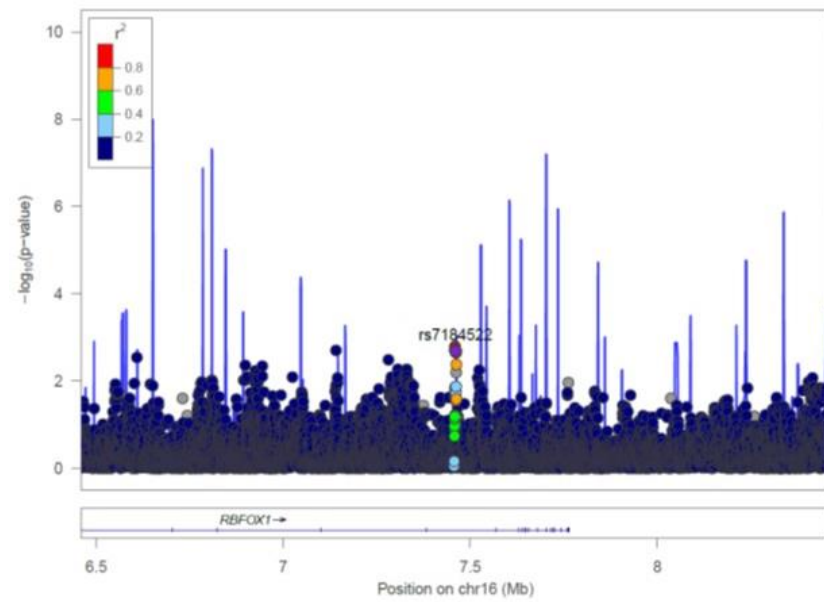


Men

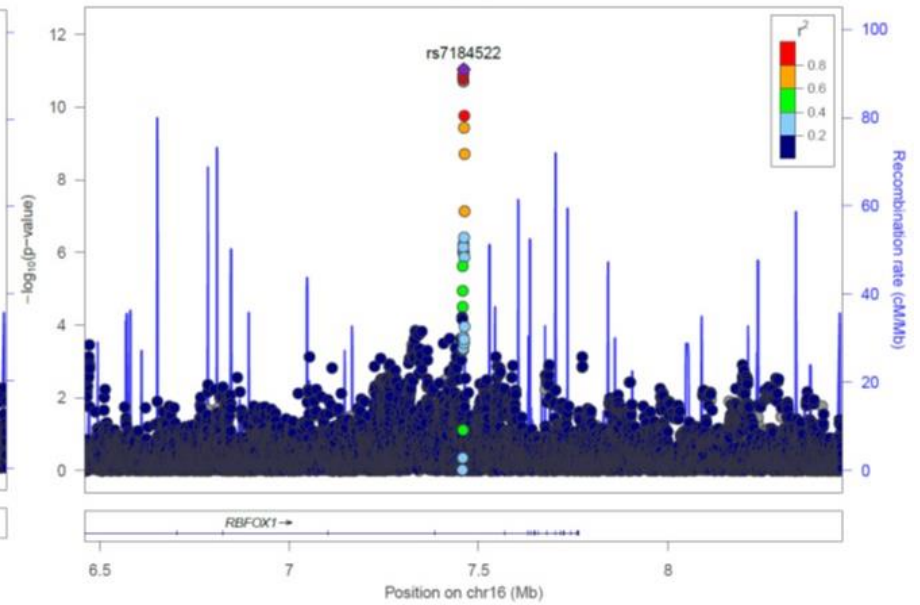


f.

Women

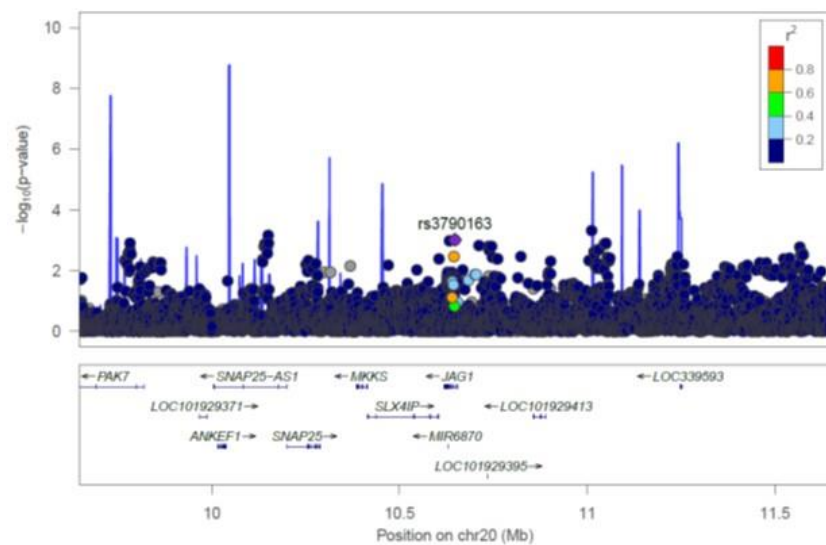


Men

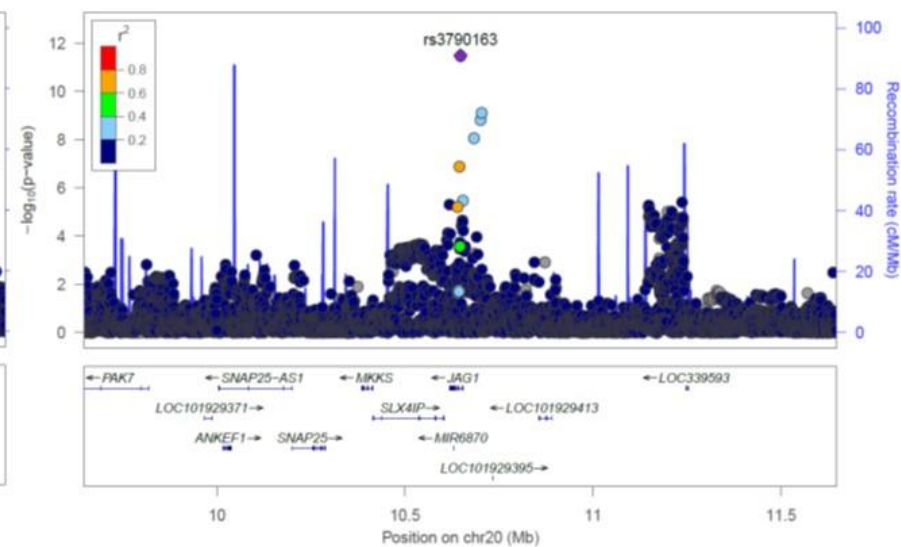


g.

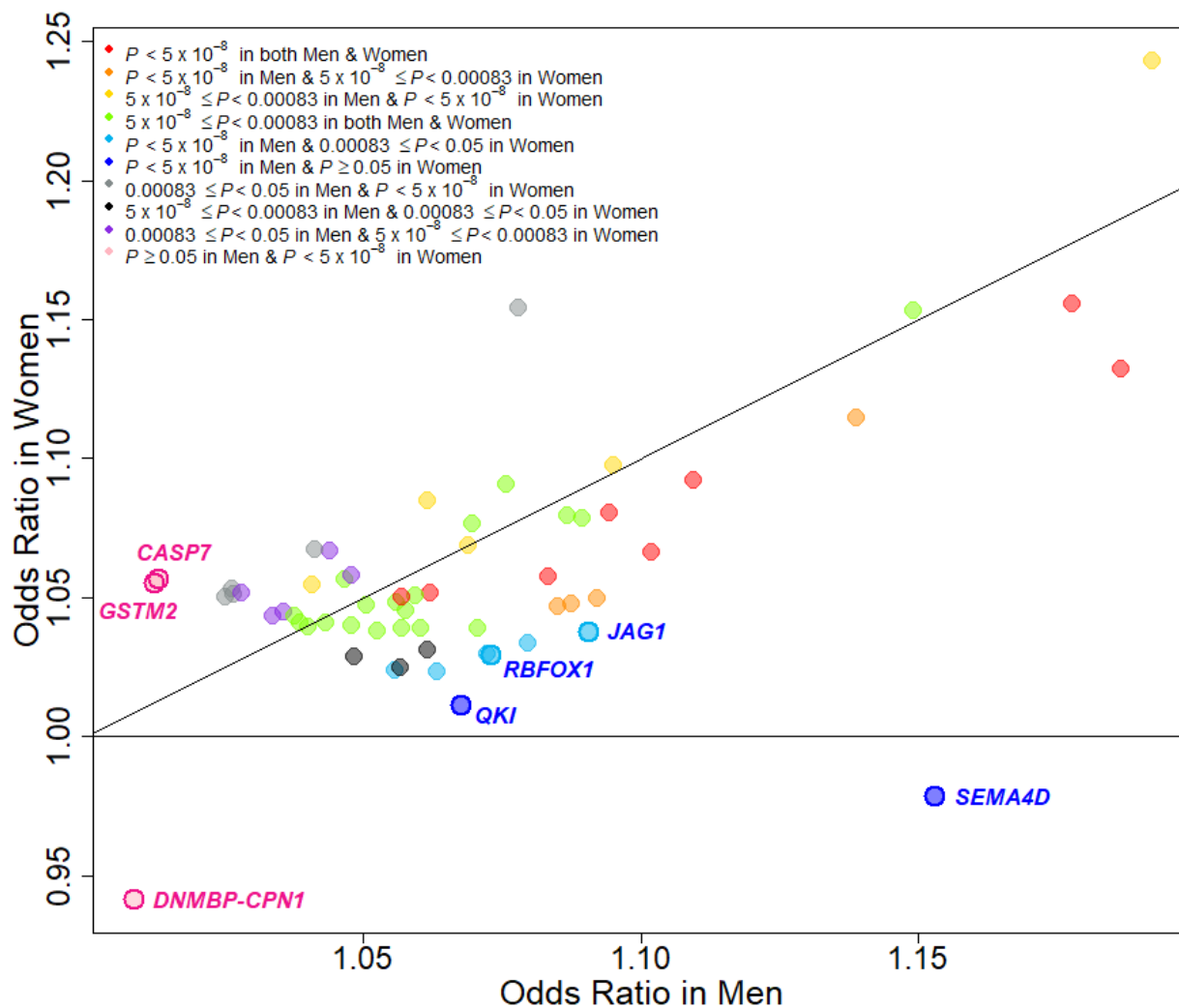
Women



Men

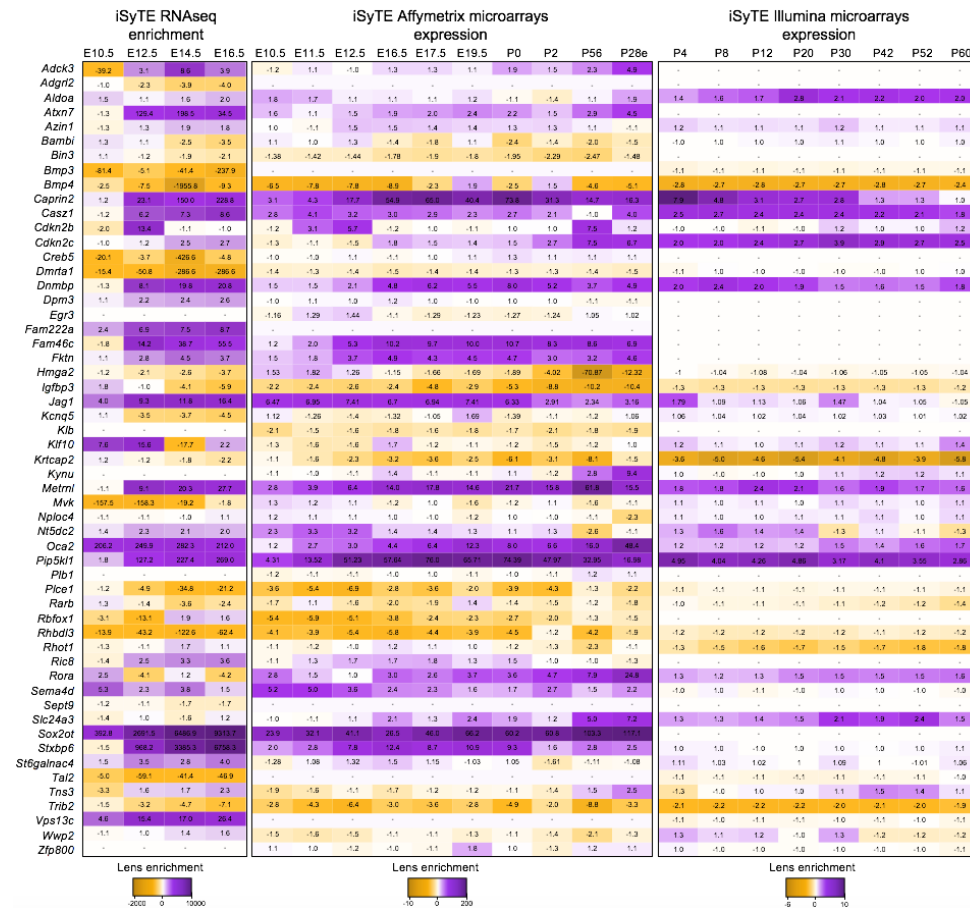


Supplementary Figure 8. Correlation of effect sizes for cataract between women and men for the lead 54 SNPs identified in the combined (GERA+UKB) GWAS multiethnic analysis and for the additional SNPs identified in the sex-specific analyses. The effect sizes were compared using a correlation test, two-sided; P -value for the correlation test is 5.8×10^{-8} .

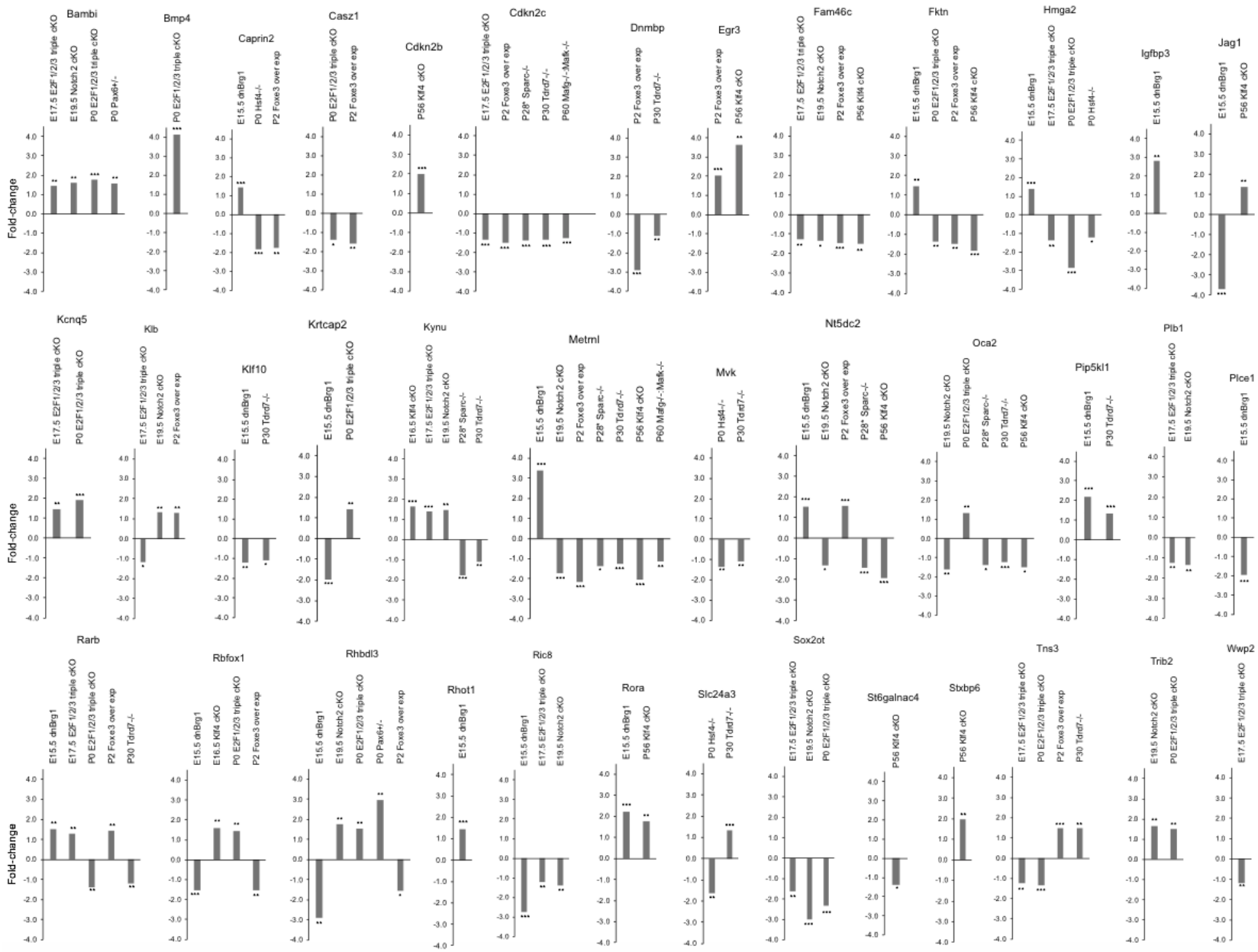


Here we report the locus names (in dark pink (or blue) color) for the lead SNPs that are specific to women (or men) and that have a significantly different OR between the women- and men-specific analyses.

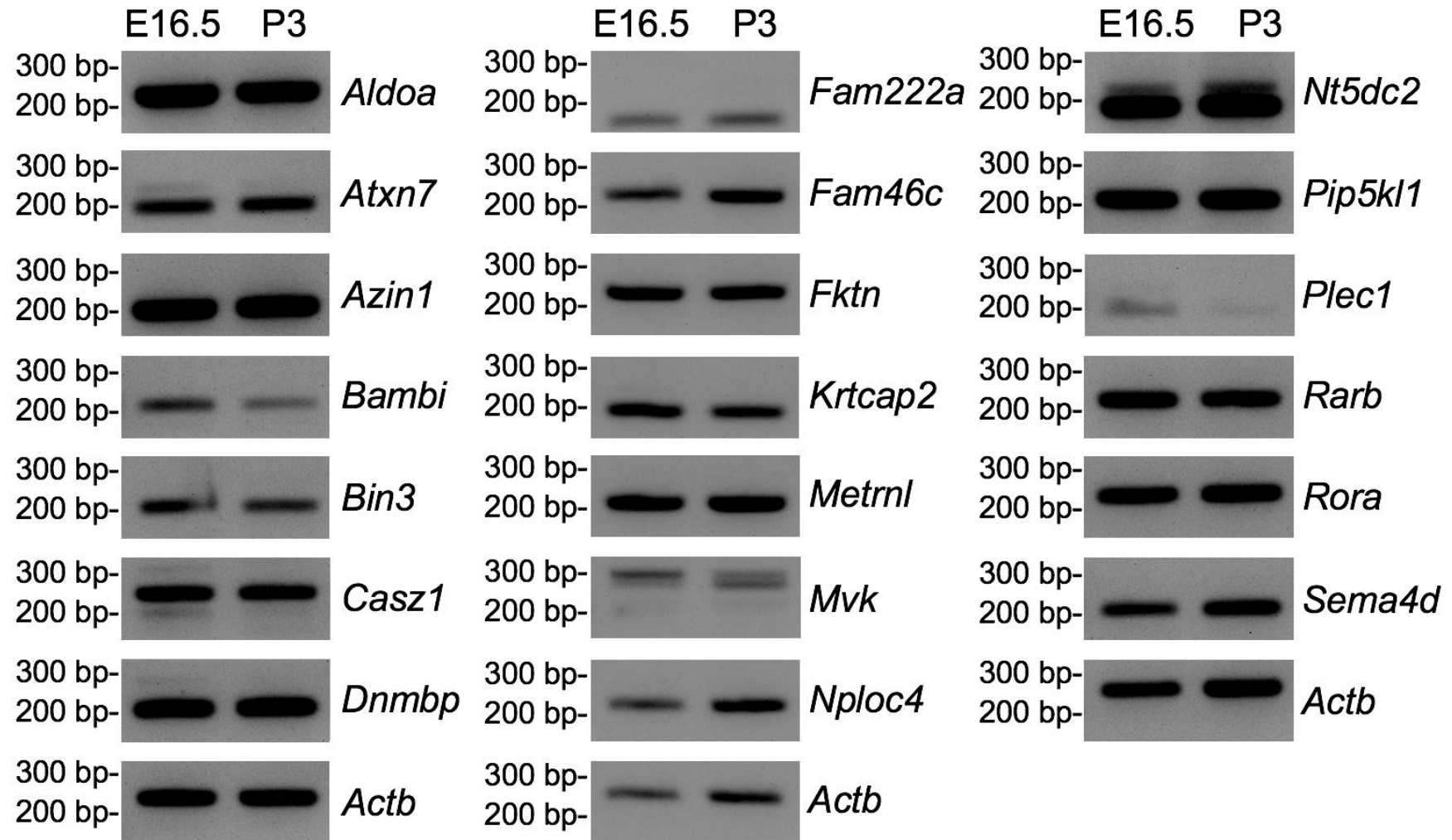
Supplementary Figure 9. Enriched expression of candidate genes in mouse lens. Mouse orthologs of the human candidate genes were examined for their lens enriched-expression in the iSyTE database. “Enriched expression” in the lens is estimated by analyzing candidate gene expression in the lens compared to that in whole-embryonic body (WB). Analysis of whole lens tissue data on various platforms, microarrays (Affymetrix, Illumina) and RNA-seq indicates lens enriched expression of genes at different stages indicated by embryonic (E) and postnatal (P) days and ranged from early lens development (*i.e.* E10.5) through adulthood (*i.e.* P60). Note: P28 in Affymetrix represents expression data on isolated lens epithelium. The range of expression on each platform is indicated by a specific heat-map. The numbers within individual tiles represent fold-change differences in the level of enriched expression in the lens compared to WB.



Supplementary Figure 10. Expression of candidate genes in lenses of gene perturbation mouse models with lens defects/cataract. Thirty-eight mouse orthologs of the human candidate genes in the 54 loci were examined for their expression in lens of ten different mouse models that exhibit lens defects and/or cataract. Analysis on the Affymetrix and Illumina platforms indicate expression of genes in the lens at different stages indicated by embryonic (E) and postnatal (P) days and ranged from mid lens development (*i.e.* E15.5) through adulthood (*i.e.* P56). Note: P28 in Affymetrix represents expression data on isolated lens epithelium. For individual candidate genes, the graphical representation is given. The y-axis indicates gene expression changes in fold-change ($*P \leq 0.05$, $**P \leq 0.01$, $***P \leq 0.001$) in the gene-perturbation mouse model lens compared to control and the x-axis indicates the different mouse models and their age. The statistical analysis used was limma using lmFit and makeContrasts functions. Microarray datasets were obtained from Gene Expression Omnibus (GEO) database. Details on the GEO accession number and the number of biological replicates for specific gene-perturbation models is provided in the Methods section.



Supplementary Figure 11. RT-PCR based validation of candidate gene expression in the lens. Several GWAS-identified candidates were independently validated by reverse transcriptase (RT)-polymerase chain reaction (PCR) assay for their expression in mouse lens at embryonic day (E)16.5 and postnatal day (P)3. Gel images are representative of RT-PCR assays performed in three biological replicates. The specific *Actb* loading control assigned for the samples below was run on the same gel.



Supplementary Figure 11. RT-PCR based validation of candidate gene expression in the lens. Original full gel scans.

