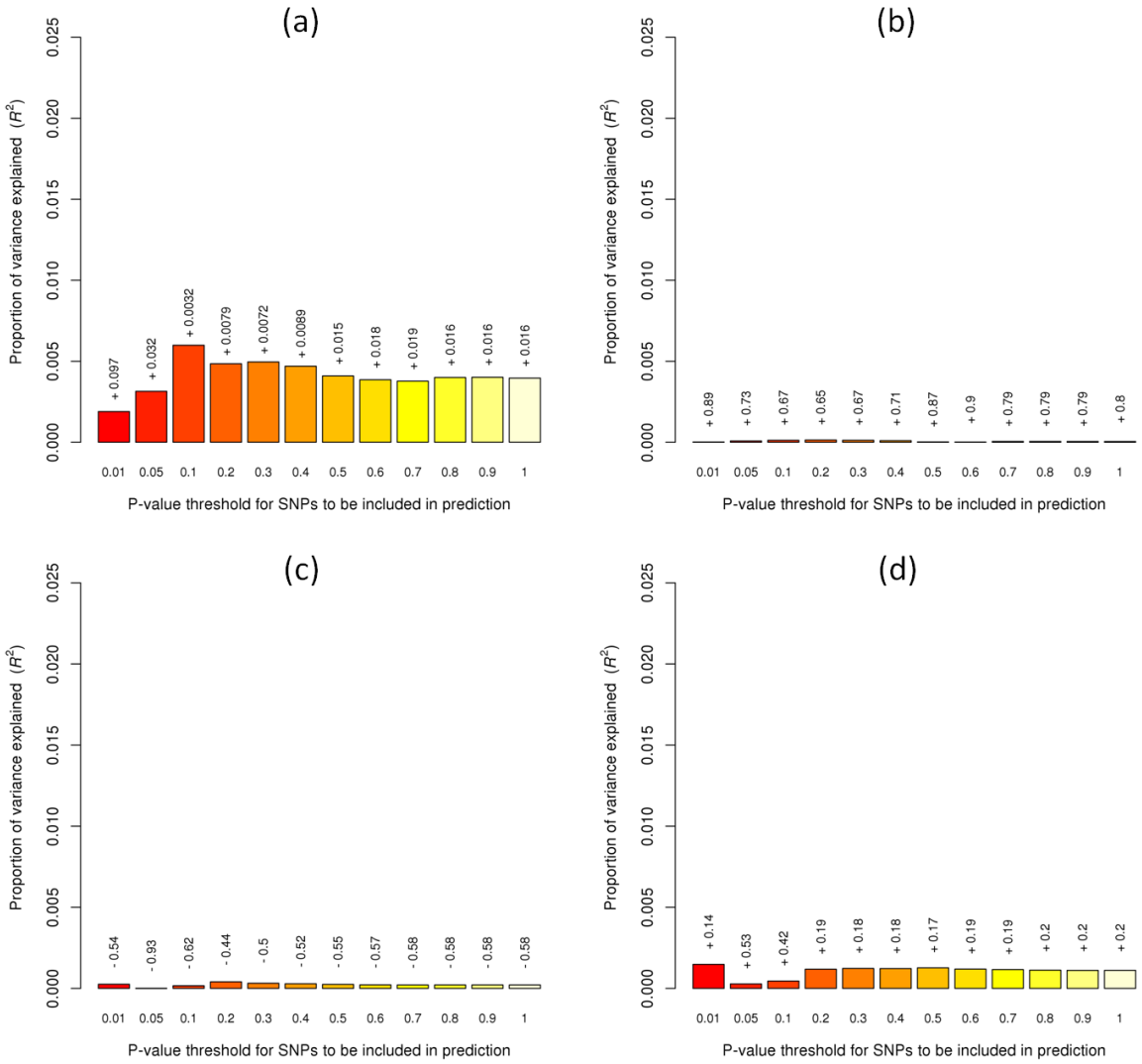
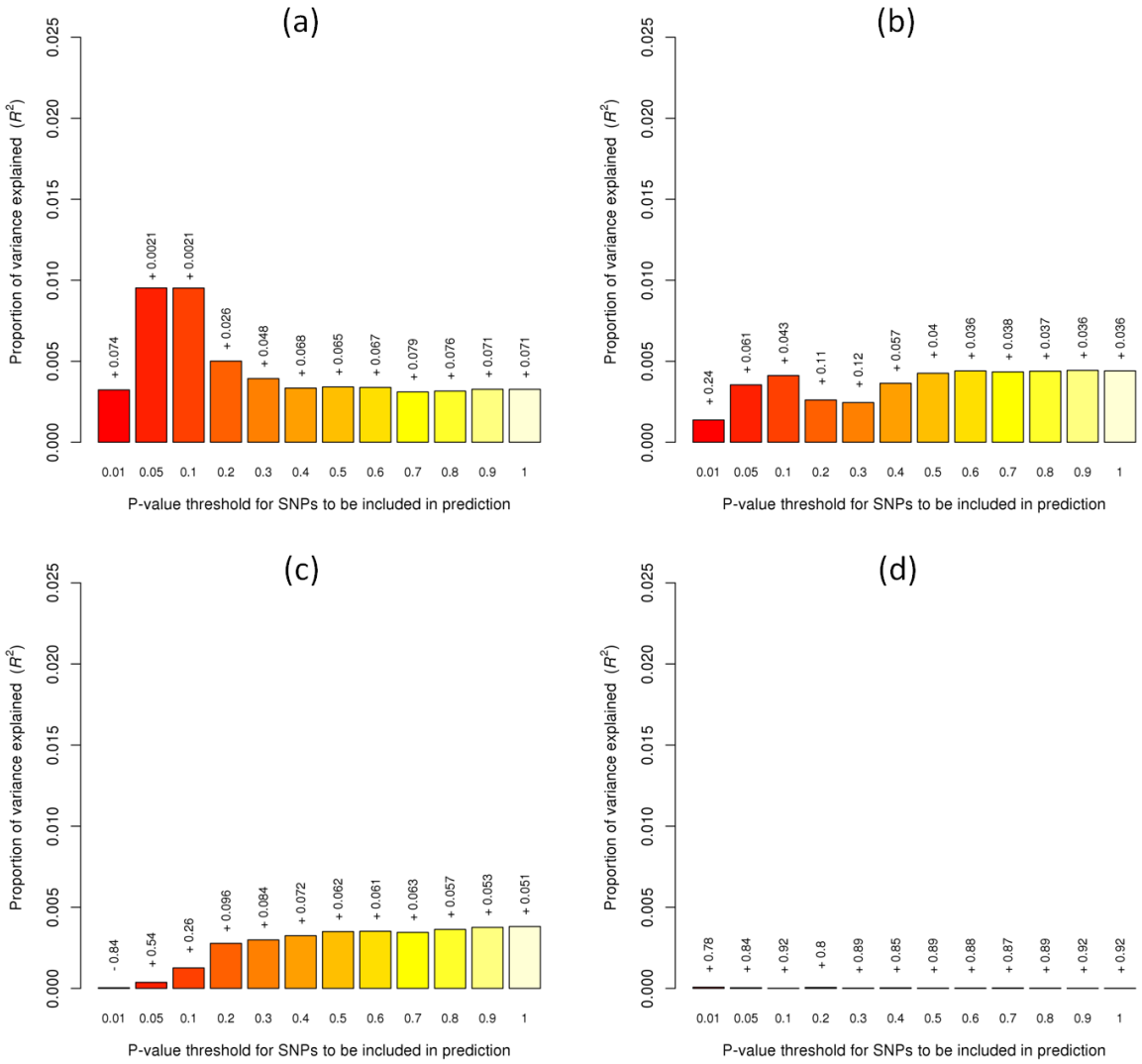
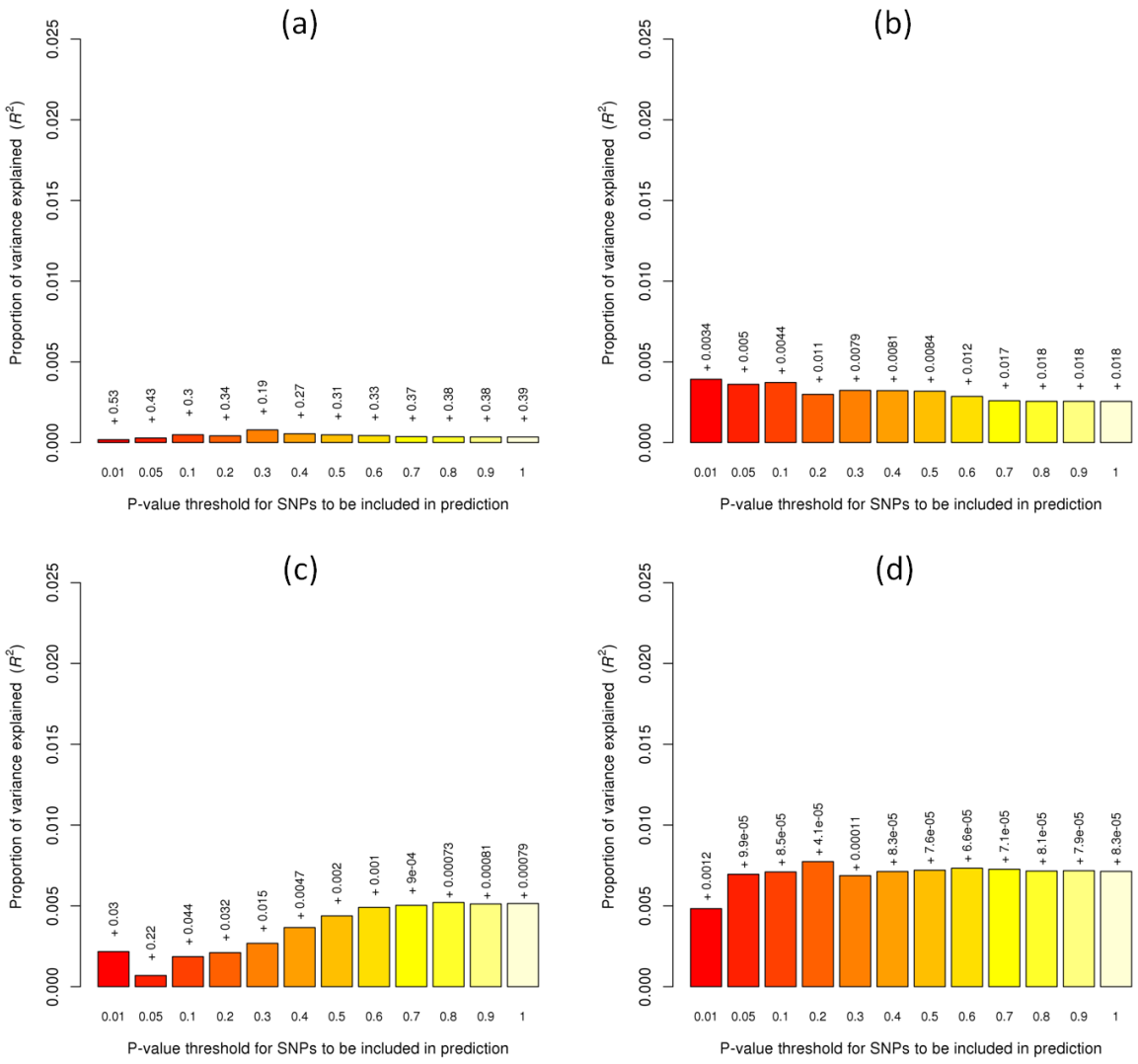


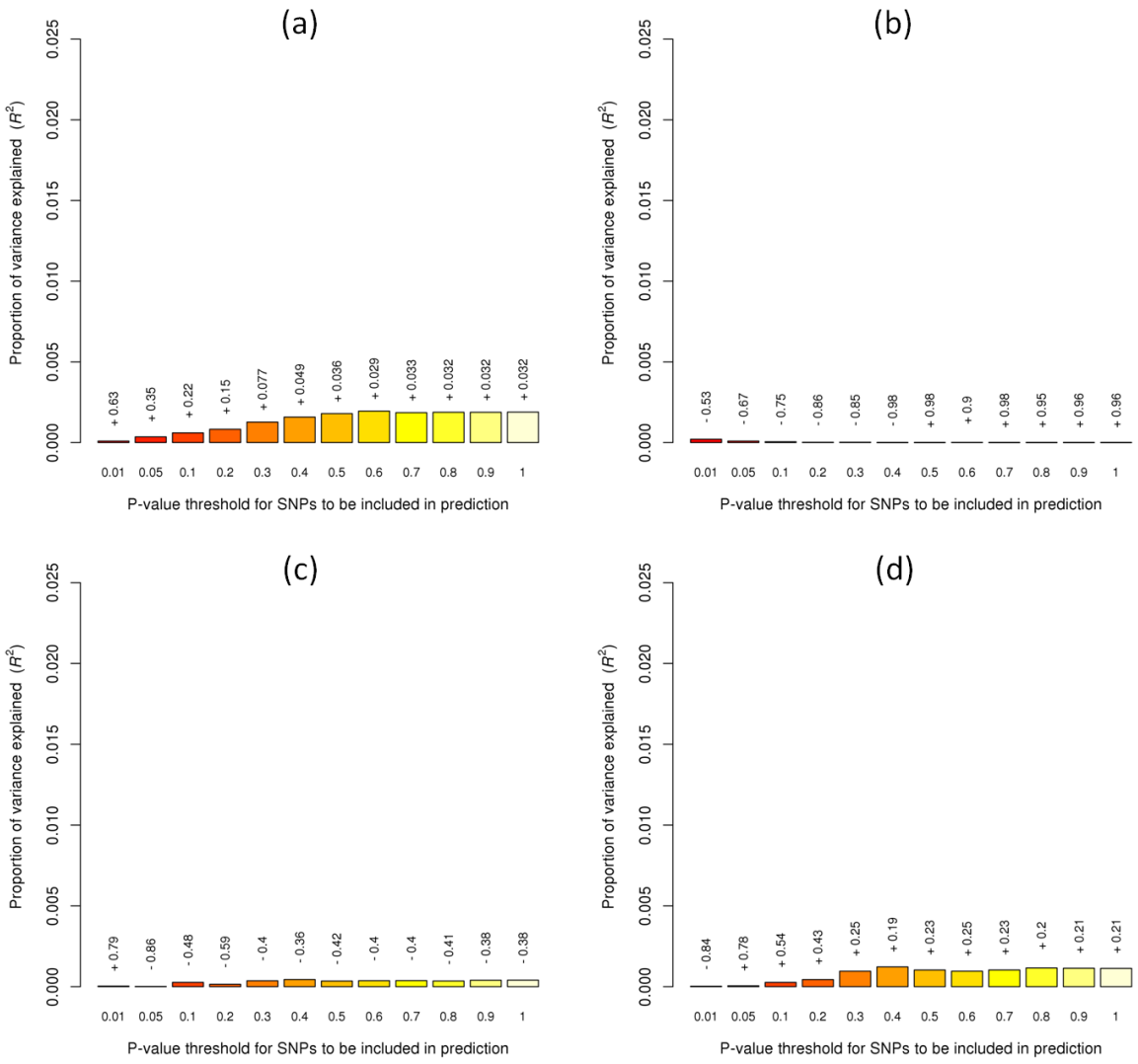
Supplementary Fig 1 (S1): Allele-specific score prediction for endometriosis, using Stage A endometriosis cases in the OX case-control set as the target population and the QIMR-HCS case-control set as the discovery population. Results for the QIMR-HCS rAFS Stage I, II, III and IV endometriosis as the discovery datasets are shown in **(a)**, **(b)**, **(c)** and **(d)** , respectively. The proportion of variance (represented by Nagelkerke's pseudo R^2) explained in the target dataset on the basis of allele-specific scores derived from the discovery dataset for the 12 significance thresholds is plotted in y-axis. The number above each bar is the p-value for the target dataset prediction analysis (R^2 significance).



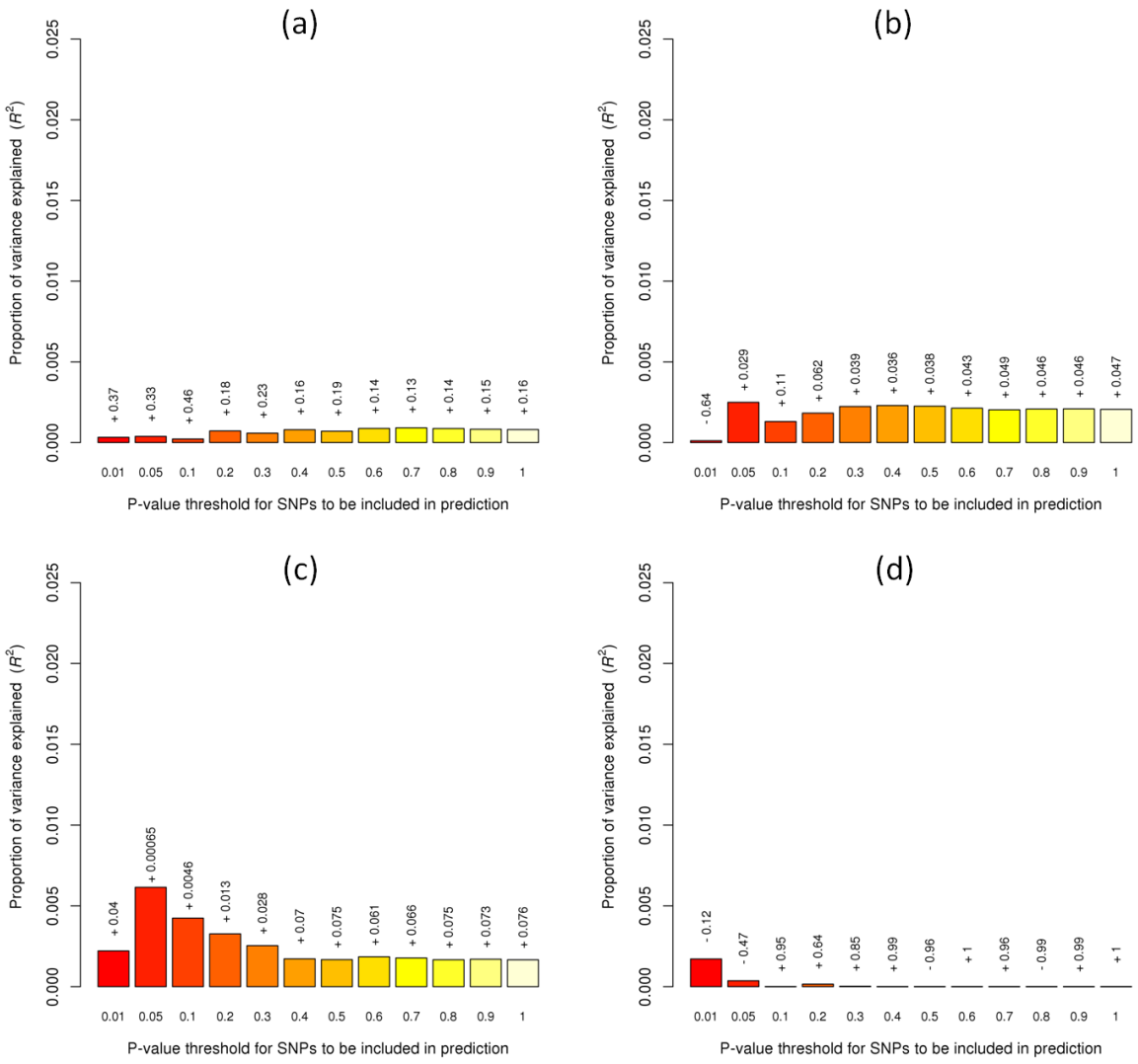
Supplementary Fig 2 (S2): Allele-specific score prediction for endometriosis, using Stage A+ endometriosis cases in the OX case-control set as the target population and the QIMR-HCS case-control set as the discovery population. Results for the QIMR-HCS rAFS Stage I, II, III and IV endometriosis as the discovery datasets are shown in **(a)**, **(b)**, **(c)** and **(d)** , respectively. The proportion of variance (represented by Nagelkerke's pseudo R^2) explained in the target dataset on the basis of allele-specific scores derived from the discovery dataset for the 12 significance thresholds is plotted in y-axis. The number above each bar is the p-value for the target dataset prediction analysis (R^2 significance).



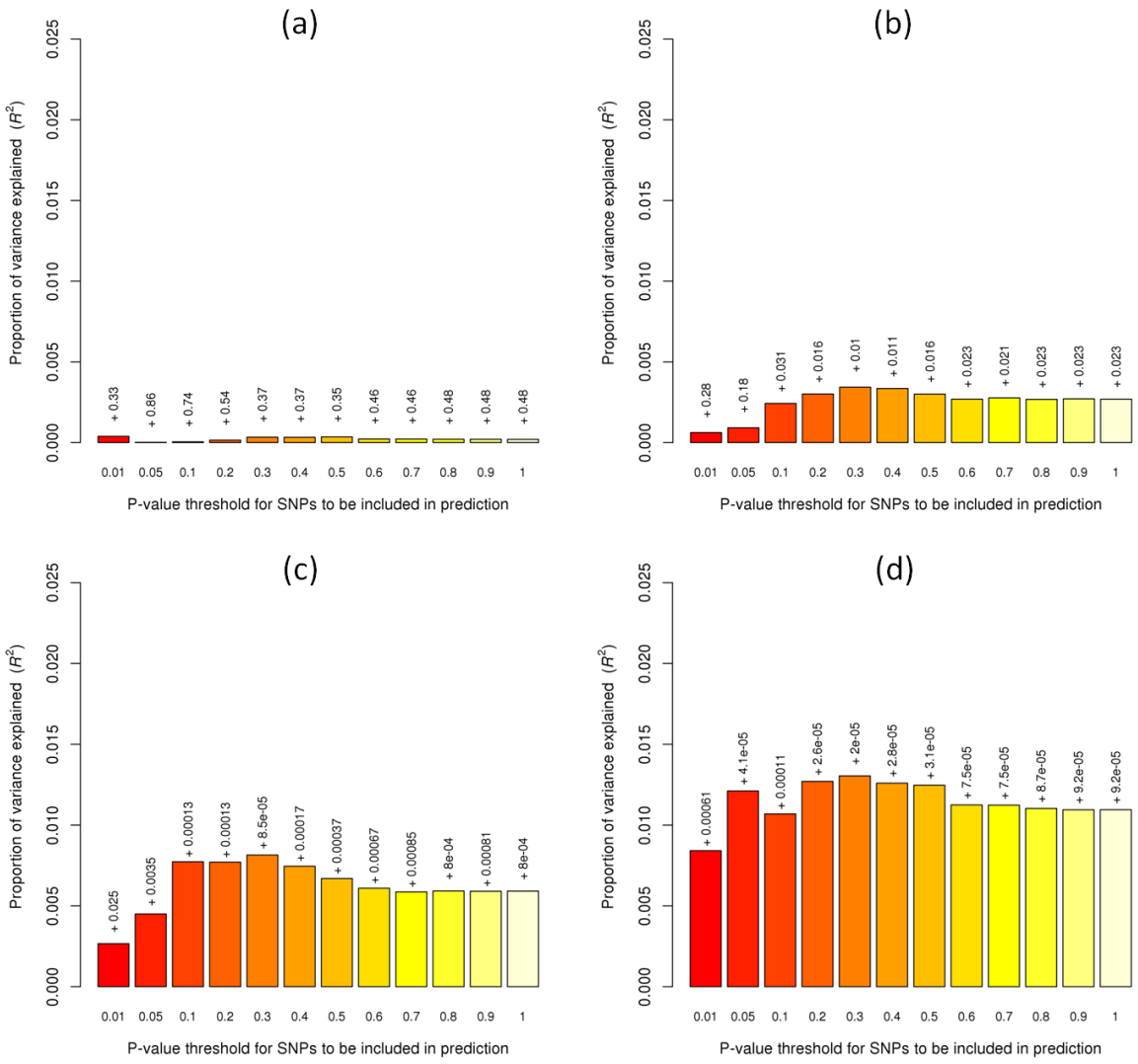
Supplementary Fig 3 (S3): Allele-specific score prediction for endometriosis, using Stage B endometriosis cases in the OX case-control set as the target population and the QIMR-HCS case-control set as the discovery population. Results for the QIMR-HCS rAFS Stage I, II, III and IV endometriosis as the discovery datasets are shown in **(a)**, **(b)**, **(c)** and **(d)** , respectively. The proportion of variance (represented by Nagelkerke's pseudo R^2) explained in the target dataset on the basis of allele-specific scores derived from the discovery dataset for the 12 significance thresholds is plotted in y-axis. The number above each bar is the p-value for the target dataset prediction analysis (R^2 significance).



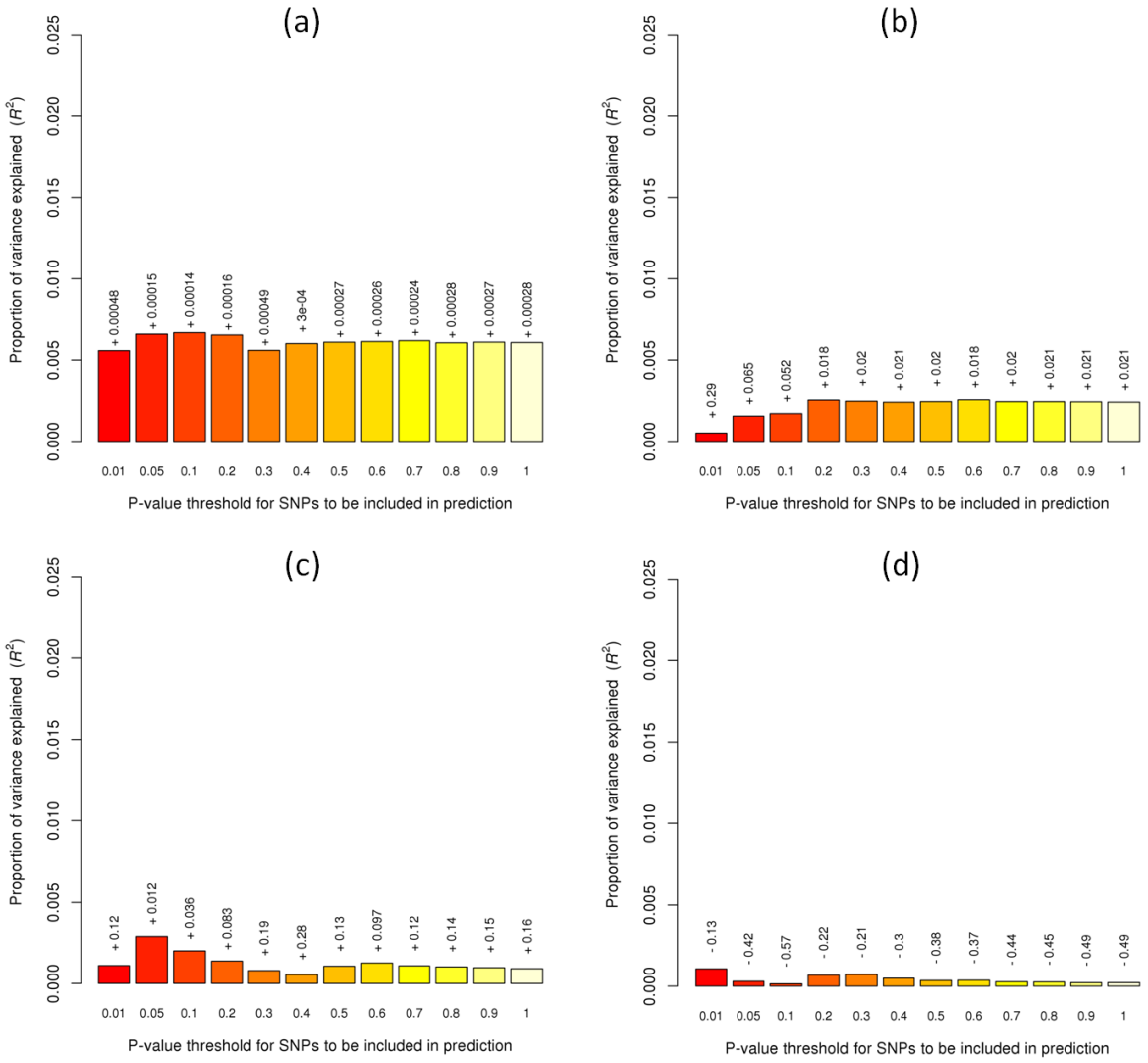
Supplementary Fig 4 (S4): Allele-specific score prediction for endometriosis, using Stage A endometriosis cases in the OX case-control set as the discovery population and the QIMR-HCS case-control set as the target population. Results for the QIMR-HCS rAFS Stage I, II, III and IV endometriosis as the target datasets are shown in **(a)**, **(b)**, **(c)** and **(d)** , respectively. The proportion of variance (represented by Nagelkerke's pseudo R^2) explained in the target dataset on the basis of allele-specific scores derived from the discovery dataset for the 12 significance thresholds is plotted in y-axis. The number above each bar is the p-value for the target dataset prediction analysis (R^2 significance).



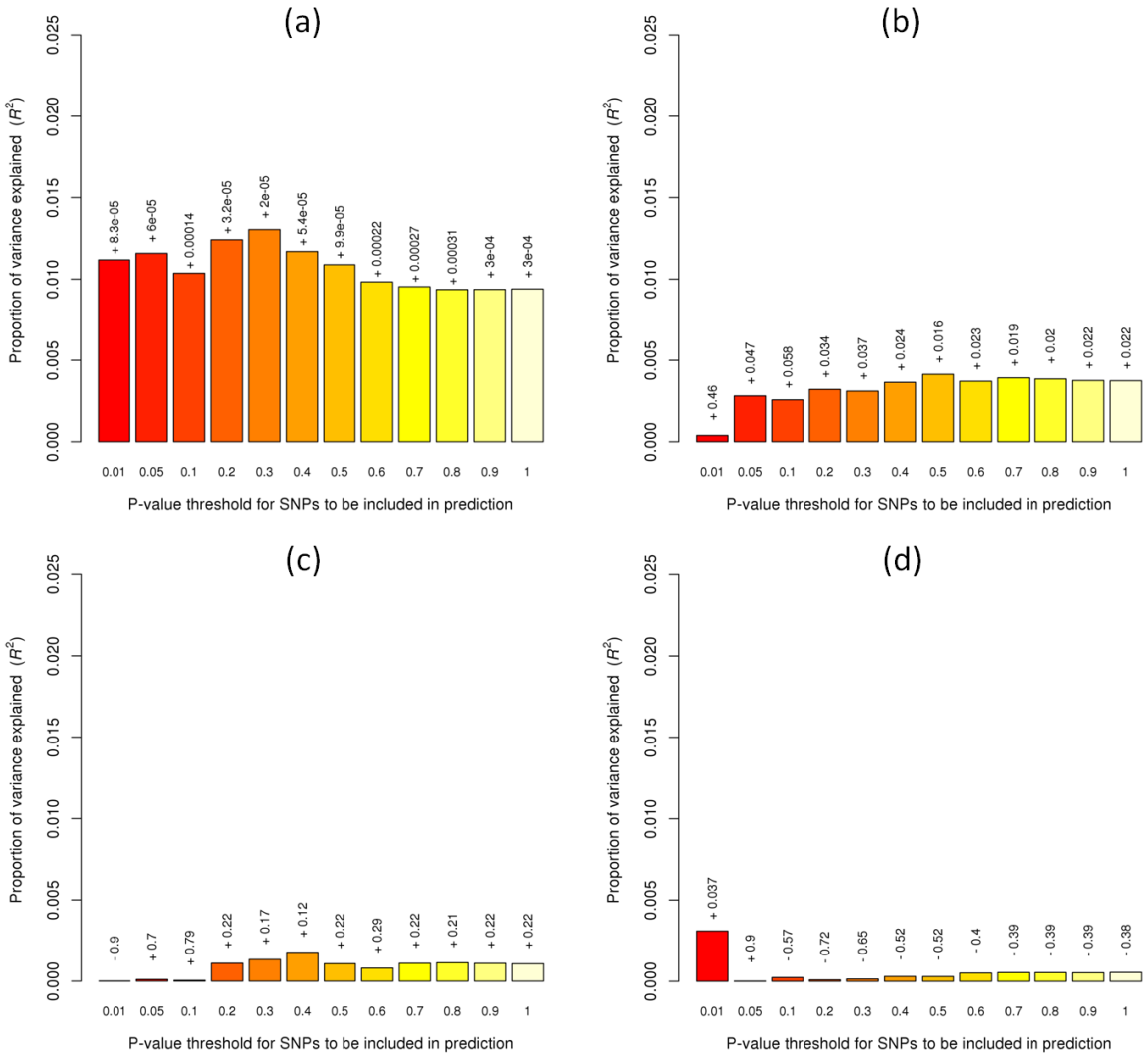
Supplementary Fig 5 (S5): Allele-specific score prediction for endometriosis, using Stage A+ endometriosis cases in the OX case-control set as the discovery population and the QIMR-HCS case-control set as the target population. Results for the QIMR-HCS rAFS Stage I, II, III and IV endometriosis as the target datasets are shown in **(a)**, **(b)**, **(c)** and **(d)** , respectively. The proportion of variance (represented by Nagelkerke's pseudo R^2) explained in the target dataset on the basis of allele-specific scores derived from the discovery dataset for the 12 significance thresholds is plotted in y-axis. The number above each bar is the p-value for the target dataset prediction analysis (R^2 significance).



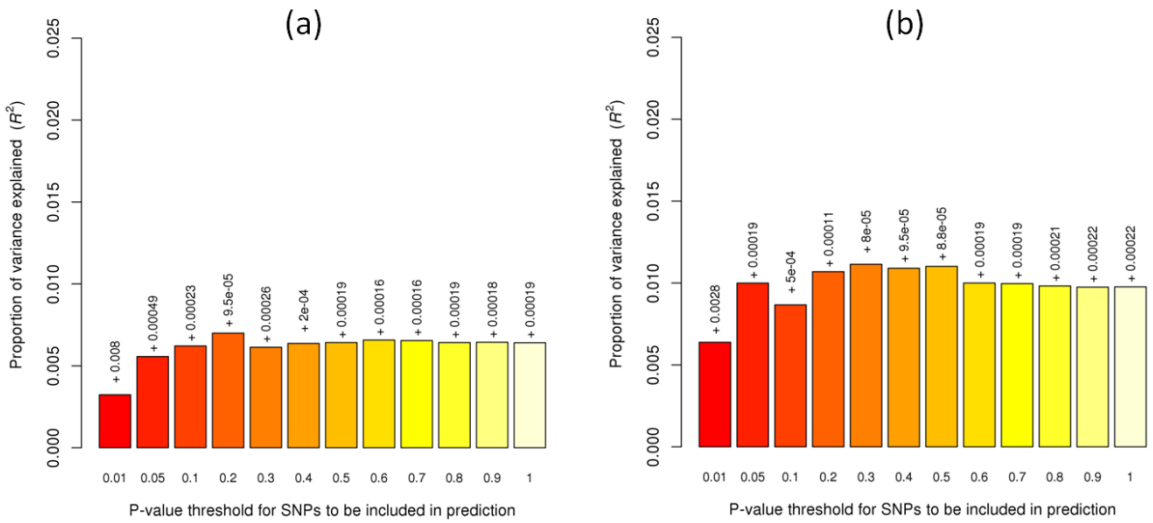
Supplementary Fig 6 (S6): Allele-specific score prediction for endometriosis, using Stage B endometriosis cases in the OX case-control set as the discovery population and the QIMR-HCS case-control set as the target population. Results for the QIMR-HCS rAFS Stage I, II, III and IV endometriosis as the target datasets are shown in **(a)**, **(b)**, **(c)** and **(d)** , respectively. The proportion of variance (represented by Nagelkerke's pseudo R^2) explained in the target dataset on the basis of allele-specific scores derived from the discovery dataset for the 12 significance thresholds is plotted in y-axis. The number above each bar is the p-value for the target dataset prediction analysis (R^2 significance).



Supplementary Fig 7 (S7): Allele-specific score prediction for endometriosis, using Stage IV endometriosis cases in the QIMR-HCS case-control set as the discovery population and Stage B in the QIMR case-control set as the target population. Results are shown for **(a)** genic SNPs (within 20kb from the transcript start and stop position), **(b)** SNPs in gene desert (excluding genic SNPs), **(c)** non-synonymous SNPs (based on snp138 functional annotation) and **(d)** rare variants. The proportion of variance (represented by Nagelkerke's pseudo R^2) explained in the target dataset on the basis of allele-specific scores derived from the discovery dataset for the 12 significance thresholds is plotted in y-axis. The number above each bar is the p-value for the target dataset prediction analysis (R^2 significance).



Supplementary Fig 8 (S8): Allele-specific score prediction for endometriosis, using Stage B endometriosis cases in the OX case-control set as the discovery population and Stage IV in the QIMR-HCS case-control set as the target population. Results are shown for **(a)** genic SNPs (within 20kb from the transcript start and stop position), **(b)** SNPs in gene desert (excluding genic SNPs), **(c)** non-synonymous SNPs (based on snp138 functional annotation) and **(d)** rare variants. The proportion of variance (represented by Nagelkerke's pseudo R^2) explained in the target dataset on the basis of allele-specific scores derived from the discovery dataset for the 12 significance thresholds is plotted in y-axis. The number above each bar is the p-value for the target dataset prediction analysis (R^2 significance).



Supplementary Fig 9 (S9): Allele-specific score prediction for endometriosis after excluding SNPs within the 8 endometriosis-associated regions. Results for Stage IV endometriosis cases in the QIMR-HCS case-control set as the discovery population and Stage B in the QIMR case-control set as the target population, and Stage B endometriosis cases in the OX case-control set as the discovery population and Stage IV in the QIMR-HCS case-control set as the target population are shown in **(a)** and **(b)**, respectively. The proportion of variance (represented by Nagelkerke's pseudo R^2) explained in the target dataset on the basis of allele-specific scores derived from the discovery dataset for the 12 significance thresholds is plotted in y-axis. The number above each bar is the p-value for the target dataset prediction analysis (R^2 significance).