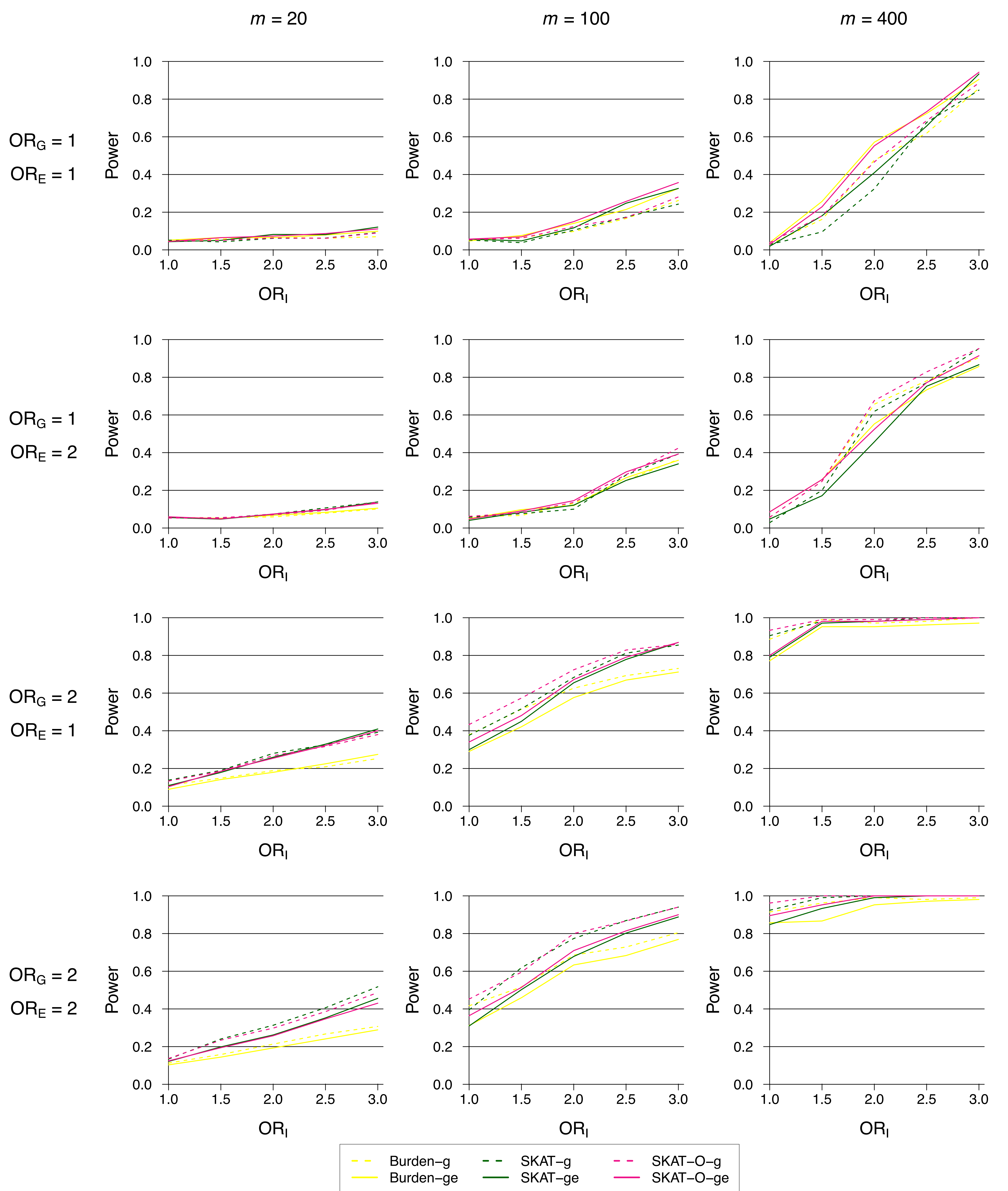


Supp. Fig. 2. Power of the different options of the Cohort Allelic Sums Test (CAST) to detect association with the genetic region as a function of the gene-environment interaction (OR_I) under scenario 2 at $\alpha = 0.05$

Bin: binary coding of S_i ; Add: additive coding of the score S_i ; T: allele frequency threshold for variant inclusion in S_i

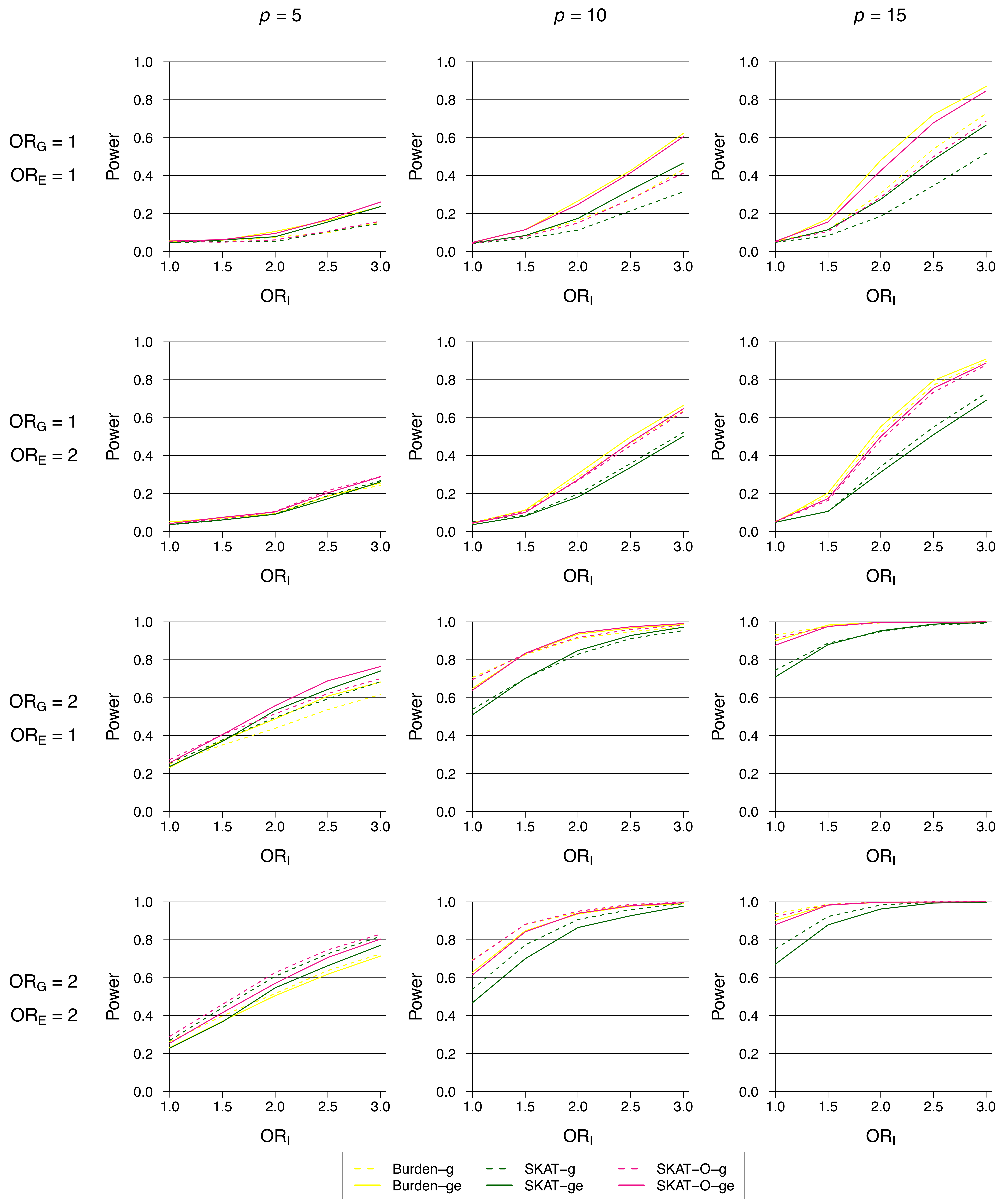
Figures are arranged in column by causal variant count p and in row by disease model, genetic effect (OR_G) and environmental effect (OR_E).

Fixed parameters: region size, $m = 20$ variants; baseline risk of disease, $B = 0.01$; and environmental exposure prevalence, $f_E = 0.3$



Supp. Fig. 3. Power of the different options of the Sequence Kernel Association Test (SKAT) to detect association with the genetic region as a function of the gene-environment interaction (OR_I) under scenario 1 at $\alpha = 0.05$

Burden: unweighted burden test; SKAT: linear kernel test; SKAT-O: optimal test – All test are applied with Beta(MAF, 1, 25) weights
 Figures are arranged in column by region size m and in row by disease model, genetic effect (OR_G) and environmental effect (OR_E).
 Fixed parameters: Proportion of causal variants, 10%; baseline risk of disease, $B = 0.01$; and environmental exposure prevalence, $f_E = 0.3$



Supp. Fig. 4. Power of the different options of the Sequence Kernel Association Test (SKAT) to detect association with the genetic region as a function of the gene–environment interaction (OR_I) under scenario 2 at $\alpha = 0.05$

Burden: unweighted burden test; SKAT: linear kernel test; SKAT-O: optimal test – All test are applied with Beta(MAF, 1, 25) weights
 Figures are arranged in column by causal variant count (p) and in row by disease model, genetic effect (OR_G) and environmental effect (OR_E).
 Fixed parameters: region size, $m = 20$ variants; baseline risk of disease, $B = 0.01$; and environmental exposure prevalence, $f_E = 0.3$