

**Benchmark Integrity Checklist: 4/6 Passed**  
**Overall: FAIL (remediation recommended)**

<input type="checkbox"/>	Label Balance (3-20% positive)	0.04059646118721461
<input type="checkbox"/>	Sample Size (>100 positives)	{'n_pairs': 14016, 'n_positive
<input type="checkbox"/>	Source Diversity (≥2 studies)	1
<input type="checkbox"/>	Gene Coverage (>1000 genes)	4548
<input type="checkbox"/>	Leakage-Free (no training overlap)	['14016 ABC training examples'
<input type="checkbox"/>	CI Reporting (95% bootstrap)	1000 bootstrap

☒ Passed

☐ Failed