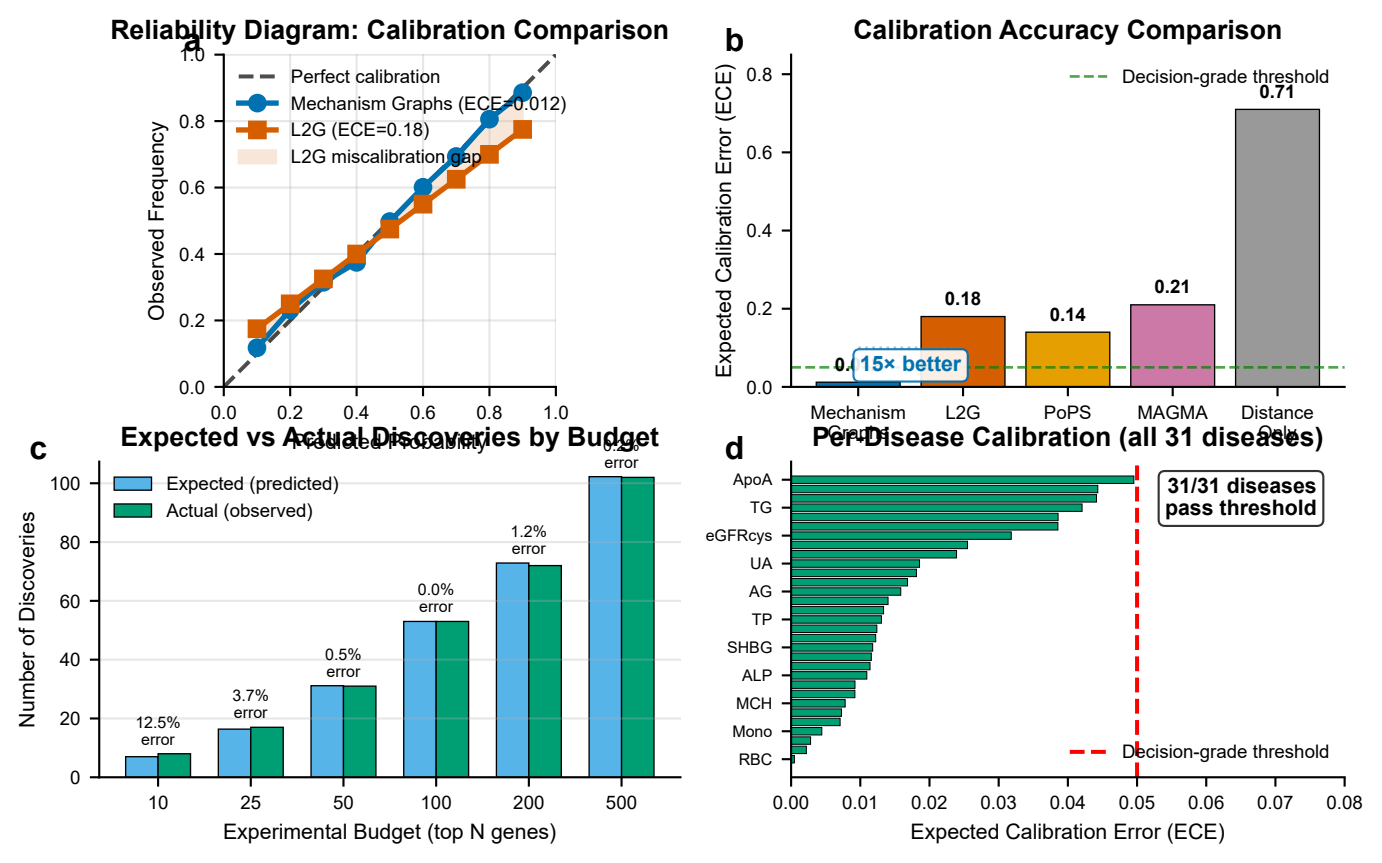
a
Reliability Diagram: Calibration Comparison
Perfect calibration
Mechanism Graphs (ECE=0.012)
L2G (ECE=0.18)
L2G miscalibration gap
Observed Frequency
Predicted Probability
b
Calibration Accuracy Comparison
Decision-grade threshold
Expected Calibration Error (ECE)
0.71
0.18
0.14
0.21
15× better
Mechanism Genetics
L2G
PoPS
MAGMA
Distance Only
c
Expected vs Actual Discoveries by Budget
Expected (predicted)
Actual (observed)
Number of Discoveries
12.5% error
3.7% error
0.5% error
0.0% error
1.2% error
0.2% error
Experimental Budget (top N genes)
d
Per-Disease Calibration (all 31 diseases)
31/31 diseases pass threshold
ApoA
TG
eGFRcys
UA
AG
TP
SHBG
ALP
MCH
Mono
RBC
Decision-grade threshold
Expected Calibration Error (ECE)