

Supplementary Materials

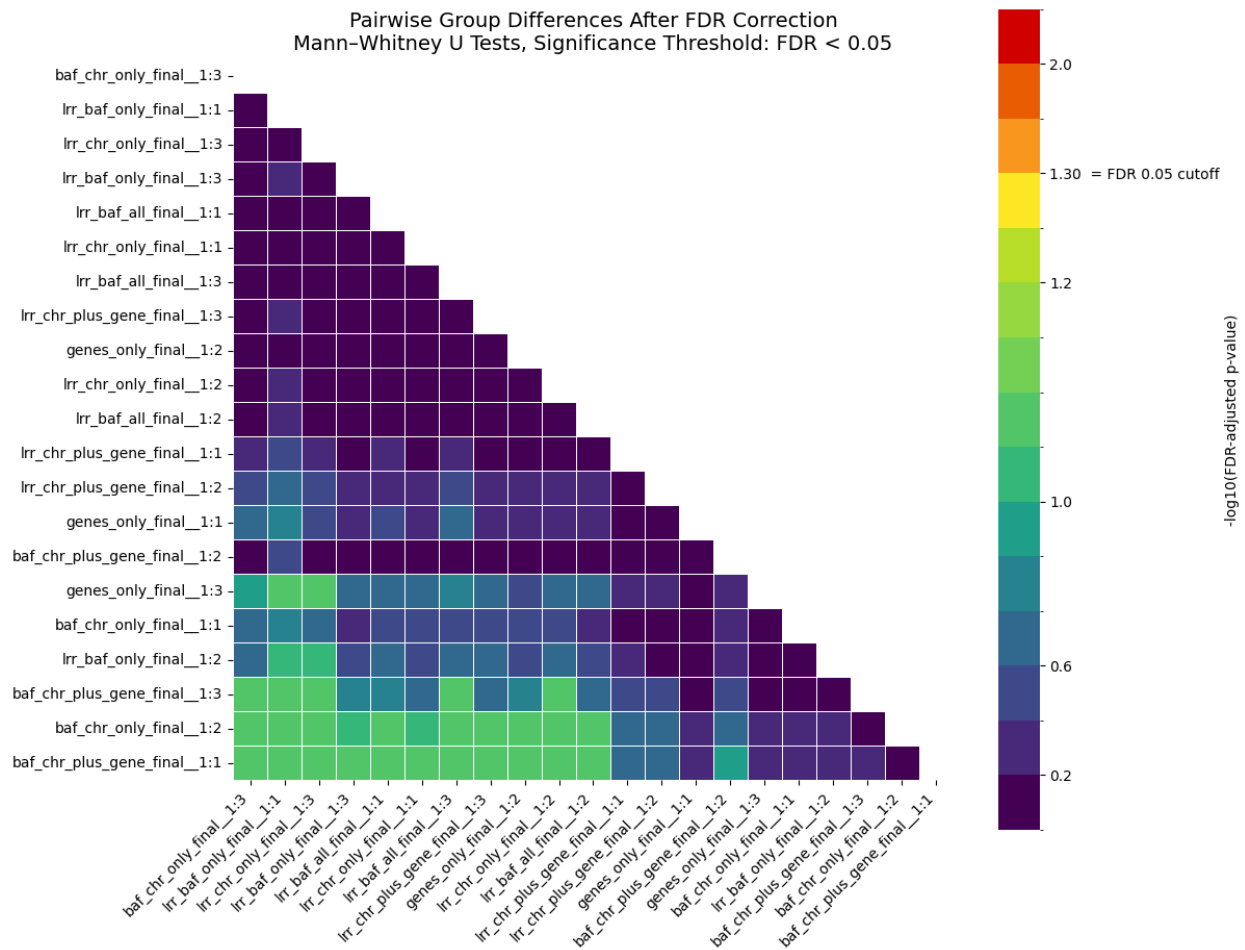
Supplementary Table 1. Ovarian cancer susceptibility genes used for gene-level feature construction. The table summarizes each gene's chromosomal location, Hail genomic coordinates used for feature extraction, and biological relevance to ovarian cancer risk, including homologous recombination repair, mismatch repair, tumor suppression, oncogenic signaling, and cell-cycle regulation.

Gene	Chr Loc	Hail Coordinates	Function / Notes
BRCA1	17q21	chr17:43044292-43170245	Homologous recombination DNA repair gene involved in double-strand break repair, maintenance of genomic stability, and tumor suppression.
BRCA2	13q12.3	chr13:32315086-32400268	Homologous recombination DNA repair gene required for RAD51-mediated repair of double-strand breaks and maintenance of chromosomal stability.
TP53	17p13.1	chr17:7661779-7687546	Tumor suppressor regulating DNA damage response, cell-cycle arrest, apoptosis, and genomic stability; commonly described as the guardian of the genome.
PALB2	16p12.2	chr16:23603160-23641321	Partner and localizer of BRCA2; supports homologous recombination repair through recruitment and stabilization of BRCA2 and RAD51 at sites of DNA damage.
RAD51	15q15.1	chr15:40694774-40732344	Central recombinase in homologous recombination repair; mediates strand invasion during double-strand break repair.
CHEK2	22q12.1	chr22:28687742-28741834	DNA damage checkpoint kinase involved in cell-cycle control, DNA repair signaling, and response to genomic stress.
BARD1	2q34	chr2:214725646-214809707	BRCA1-interacting protein that contributes to homologous recombination repair and tumor suppressor activity through the BRCA1-BARD1 complex.
BRIP1	17q23.2	chr17:61679139-61863528	DNA helicase involved in BRCA1-associated homologous recombination repair; pathogenic variants are associated with increased ovarian cancer susceptibility.
MRE11	11q21	chr11:94419729-94493908	Component of the MRN complex with RAD50 and NBN; participates in double-strand break detection, processing, and homologous recombination repair.
NBN	8q21.3	chr8:89924515-90003228	Member of the MRN complex involved in DNA double-strand break recognition, checkpoint activation, and repair pathway coordination.
RAD50	5q31	chr5:132556019-132646349	Structural component of the MRN complex involved in double-strand break repair, DNA end tethering, and homologous recombination signaling.

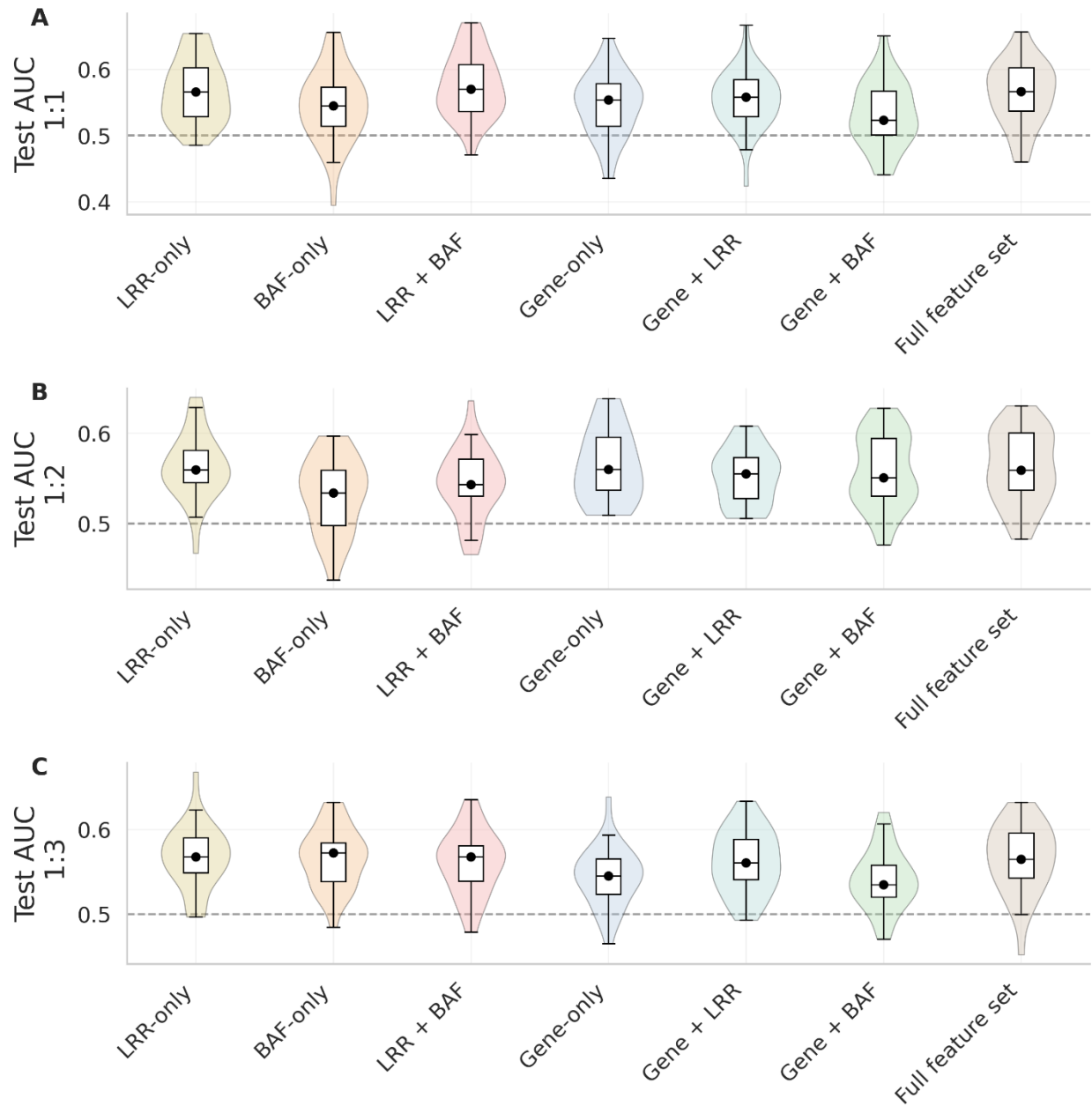
RAD51C	17q23	chr17:58692573-58735611	RAD51C paralog involved in homologous recombination repair; pathogenic variants are linked to homologous recombination deficiency and ovarian cancer risk.
MLH1	3p21.3	chr3:36993226-37050896	DNA mismatch repair gene associated with Lynch syndrome; loss of function can contribute to microsatellite instability and cancer susceptibility.
MSH2	2p22-p21	chr2:47403067-47663146	Core mismatch repair gene associated with Lynch syndrome; recognizes base-pair mismatches and insertion-deletion loops during DNA replication.
MSH6	2p16	chr2:47695530-47810109	Mismatch repair gene that partners with MSH2 to detect replication errors; germline alterations are associated with Lynch syndrome.
PMS2	7p22	chr7:5970925-6009130	Mismatch repair gene that partners with MLH1 during repair of DNA replication errors; pathogenic variants are associated with Lynch syndrome.
BRAF	7q34	chr7:140719327-140924976	MAPK pathway oncogene; recurrent activating mutations occur in selected ovarian cancer subtypes and may influence tumor signaling. BRAF germline structural variation was included for completeness given the locus's recurrent involvement in low-grade serous and mucinous ovarian cancers.

Supplementary Table 2. Summary of H2O AutoML leader model families across repeated runs. For each feature-set and case-control ratio combination, the most frequent and second most frequent leader model families are reported across 30 independent runs. Values in parentheses indicate the number and percentage of runs in which each model family was selected as the AutoML leader.

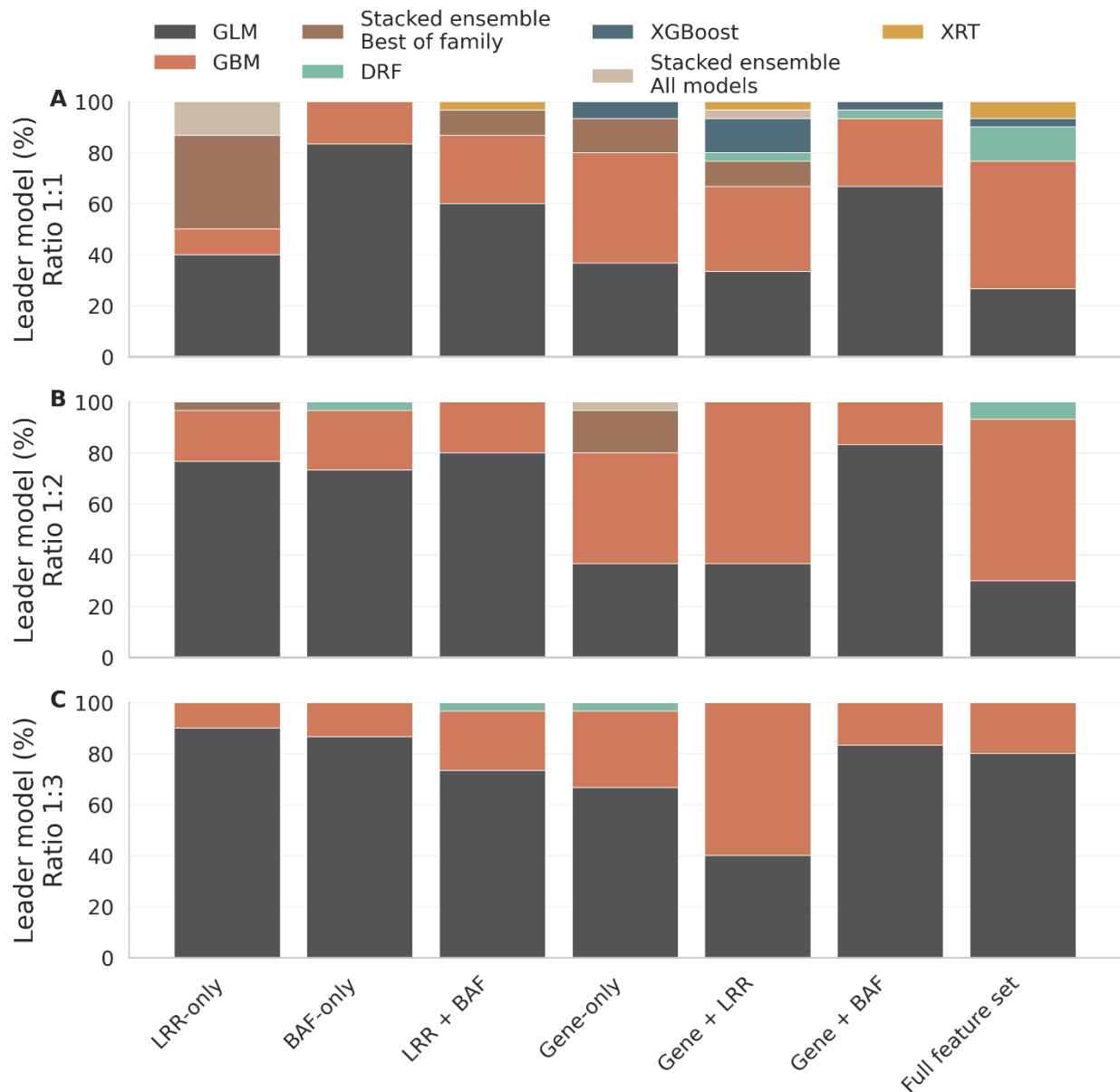
Feature Set	Ratio	Most frequent leader	Second most frequent leader
LRR-only	1:1	GLM (12/30, 40.0%)	Stacked ensemble (11/30, 36.7%)
BAF-only	1:1	GLM (25/30, 83.3%)	GBM (5/30, 16.7%)
LRR + BAF	1:1	GLM (18/30, 60.0%)	GBM (8/30, 26.7%)
Gene-only	1:1	GBM (13/30, 43.3%)	GLM (11/30, 36.7%)
Gene + LRR	1:1	GBM (10/30, 33.3%)	GLM (10/30, 33.3%)
Gene + BAF	1:1	GLM (20/30, 66.7%)	GBM (8/30, 26.7%)
Full feature set	1:1	GBM (15/30, 50.0%)	GLM (8/30, 26.7%)
LRR-only	1:2	GLM (23/30, 76.7%)	GBM (6/30, 20.0%)
BAF-only	1:2	GLM (22/30, 73.3%)	GBM (7/30, 23.3%)
LRR + BAF	1:2	GLM (24/30, 80.0%)	GBM (6/30, 20.0%)
Gene-only	1:2	GBM (13/30, 43.3%)	GLM (11/30, 36.7%)
Gene + LRR	1:2	GBM (19/30, 63.3%)	GLM (11/30, 36.7%)
Gene + BAF	1:2	GLM (25/30, 83.3%)	GBM (5/30, 16.7%)
Full feature set	1:2	GBM (19/30, 63.3%)	GLM (9/30, 30.0%)
LRR-only	1:3	GLM (27/30, 90.0%)	GBM (3/30, 10.0%)
BAF-only	1:3	GLM (26/30, 86.7%)	GBM (4/30, 13.3%)
LRR + BAF	1:3	GLM (22/30, 73.3%)	GBM (7/30, 23.3%)
Gene-only	1:3	GLM (20/30, 66.7%)	GBM (9/30, 30.0%)
Gene + LRR	1:3	GBM (18/30, 60.0%)	GLM (12/30, 40.0%)
Gene + BAF	1:3	GLM (25/30, 83.3%)	GBM (5/30, 16.7%)
Full feature set	1:3	GLM (24/30, 80.0%)	GBM (6/30, 20.0%)



Supplementary Figure 1. Pairwise FDR-corrected Mann-Whitney U test p-values across all 21 feature set and case-control ratio combinations. Pairwise comparisons were corrected using the Benjamini-Hochberg false-discovery-rate procedure. No pairwise comparison reached statistical significance after correction, indicating overlapping performance distributions across configurations.



Supplementary Figure 2. Test-set AUC distributions across feature sets and case-control ratios. Violin plots show the distribution of test-set AUC values across 30 independent runs for each of the seven feature sets and three case-control ratios (1:1, 1:2, 1:3). White boxplots indicate the interquartile range and black dots indicate the median. The x-axis represents feature-set and case-control ratio combinations, and the y-axis represents the test-set AUC across the 30 independent runs.



Supplementary Figure 3. Lead model distribution across feature sets and case-control ratios. Stacked bar plots show the frequency of each model family (GLM, GBM, Distributed Random Forest, Stacked Ensemble, and others) selected as the AutoML leader across 30 independent runs for each feature-set and ratio combination. The x-axis represents feature-set and ratio combinations, and the y-axis represents the count of runs in which each model family was selected as the AutoML leader.