

A Reproducible and Unified Benchmark of Deep Learning Feature Selection Across Simulations and Multi-Omics datasets

Qingyu Meng^{1 †}, Xiaoyan Sun^{1 †}, Ning Li³, Long Liu^{2,*}, Deqiang Zheng^{3,*}, Yalu Wen^{1,*}

1 Department of Statistics, University of Auckland, Auckland, New Zealand

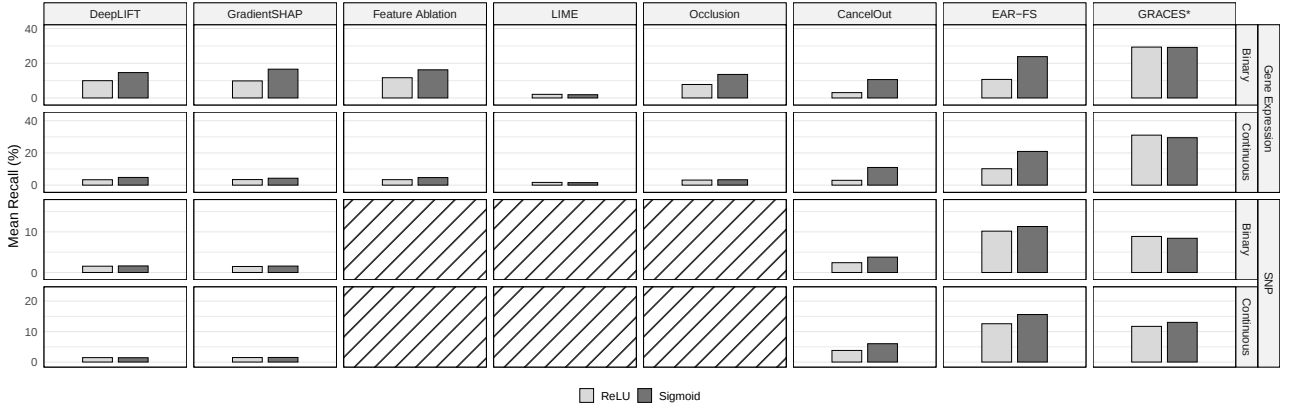
2 Department of Health Statistics, School of Public Health, Binzhou Medical University, Yantai, Shandong, China

3 Department of Epidemiology and Health Statistics, School of Public Health, Capital Medical University, Beijing, China

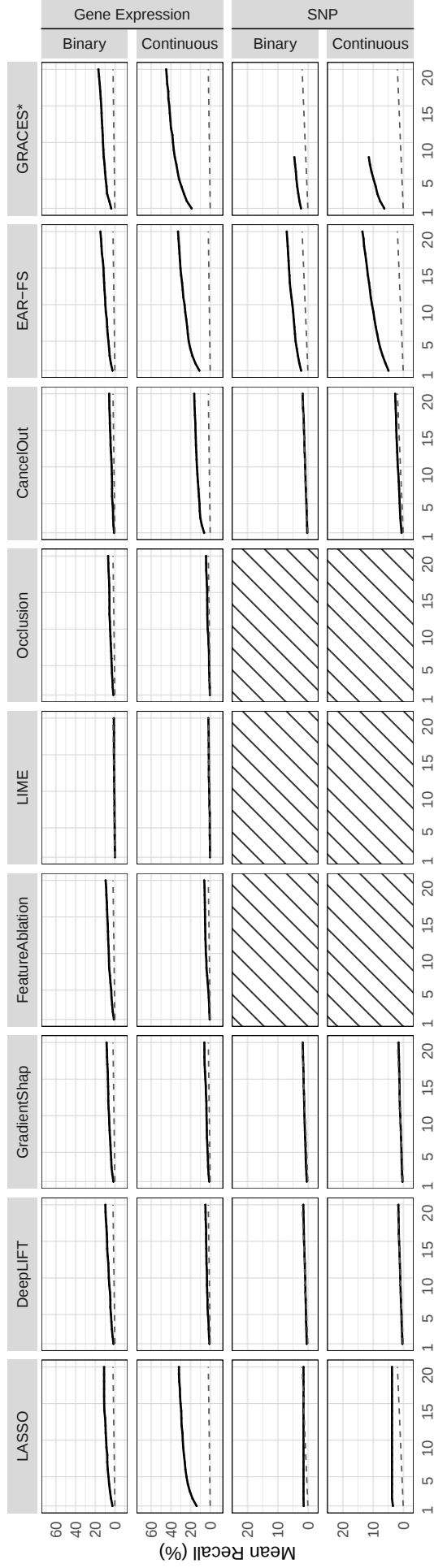
† These authors contributed equally to this work.

* Correspondence: y.wen@auckland.ac.nz; longliu@bzmc.edu.cn; dqzheng@ccmu.edu.cn

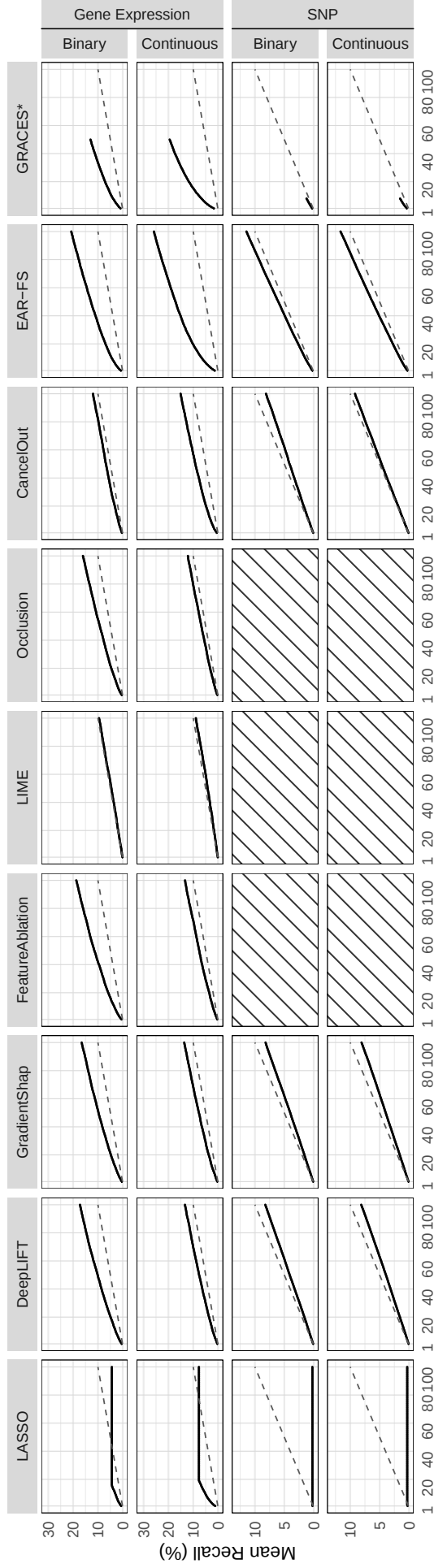
Supplementary Figures



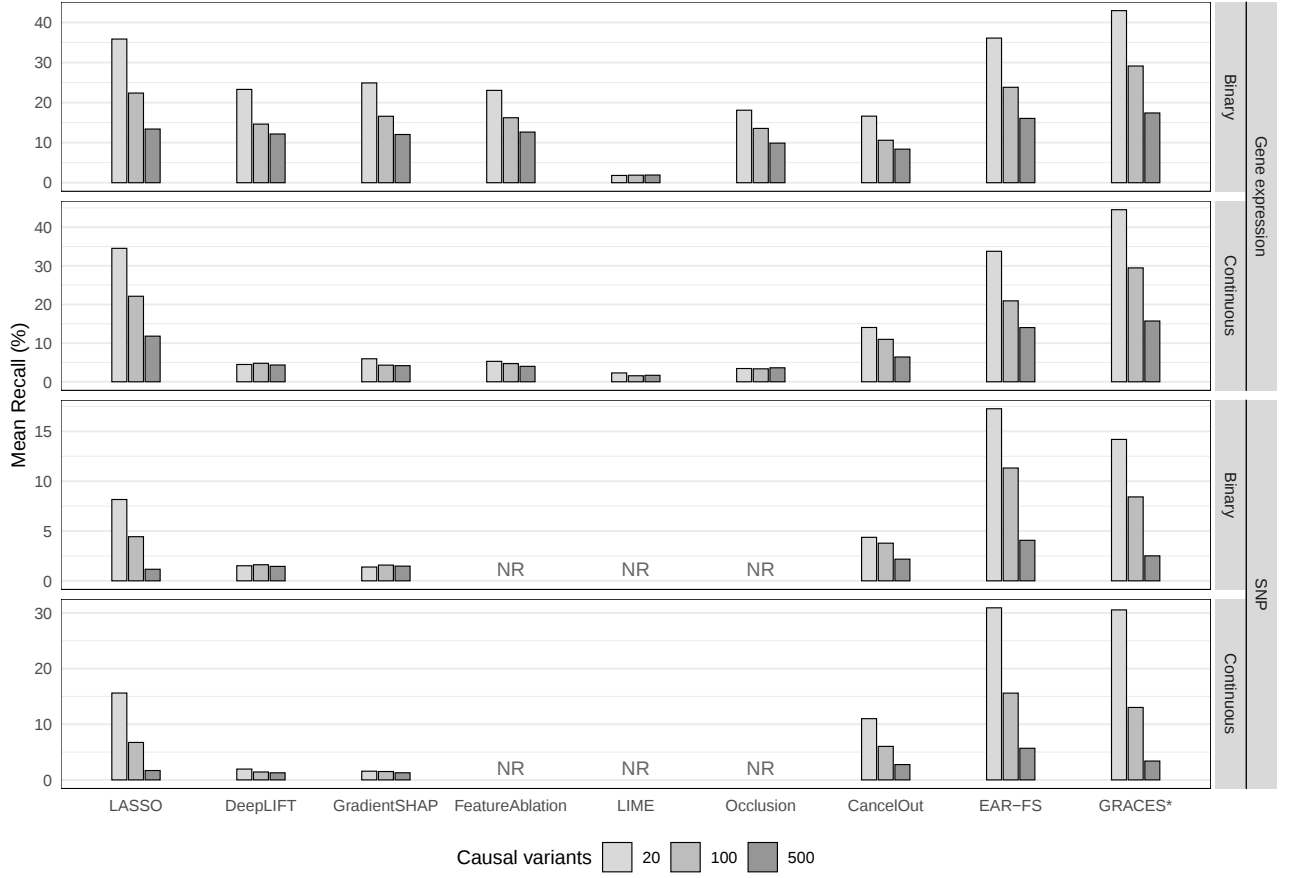
Supplementary Fig. S1: **Comparison of activation functions under a setting with 100 causal variables and a sample size of 5,000.** Mean recall is evaluated under ReLU and Sigmoid activations at a fixed top-20% selection threshold which is defined relative to the total feature dimensionality of each modality (26,000 for gene expression and 350,000 for SNP). *LIME*, *Feature Ablation*, and *Occlusion* are omitted for SNP data due to computational infeasibility. The asterisk(*) on GRACES indicates that, due to computational constraints, GRACES was evaluated only up to a truncated feature budget for SNP data; results beyond this limit were not computed.



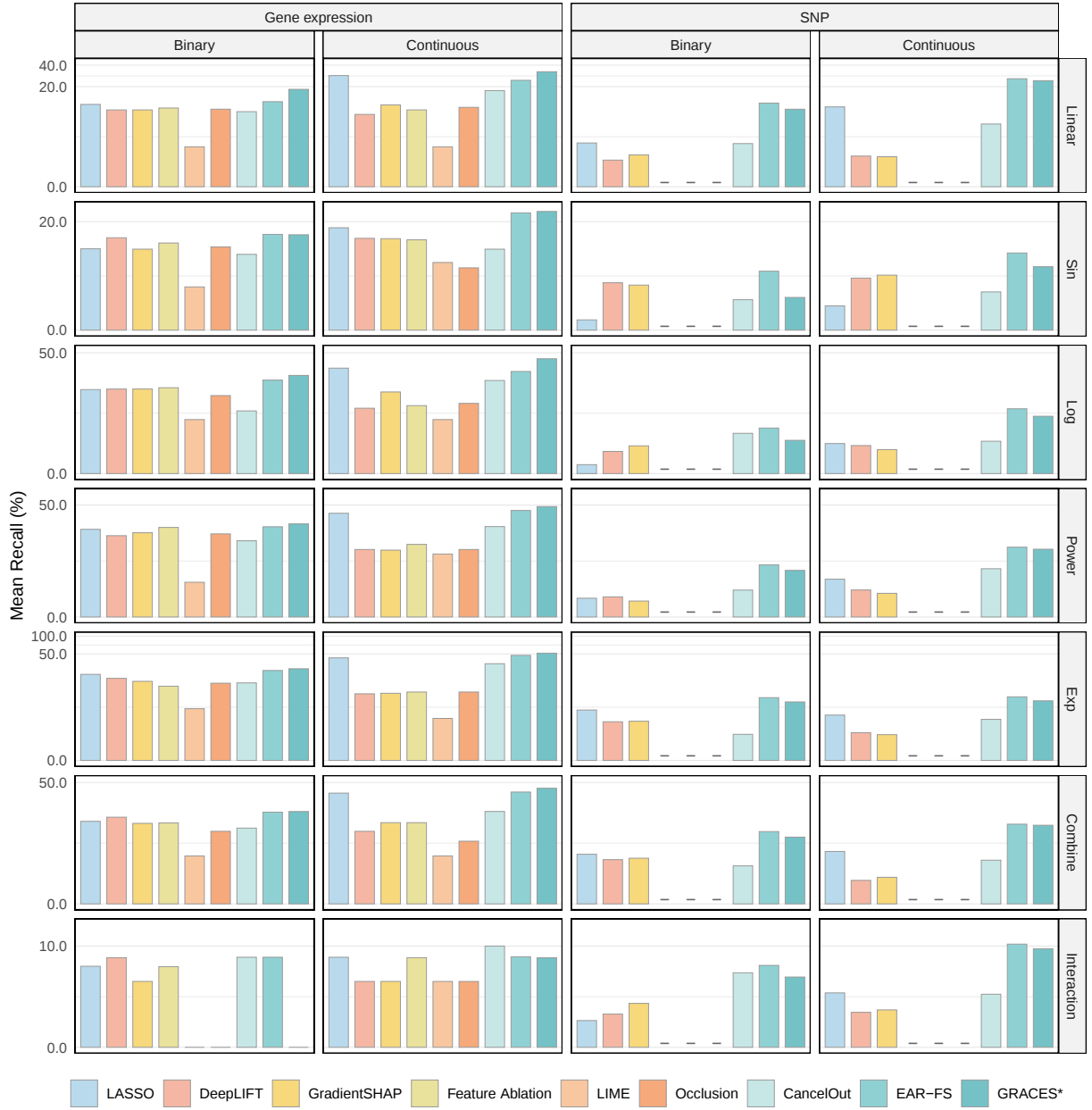
Supplementary Fig. S2: Recall of feature selection methods under a setting with 20 causal variables and a sample size of 1,000, evaluated across different top- K thresholds under Sigmoid activation. Top- K thresholds range from 1% to 20%, defined relative to the total number of features in each modality (26,000 for gene expression and 350,000 for SNP). The dashed line represents the chance-level baseline, corresponding to the expected performance under random selection. *LIME*, *Feature Ablation*, and *Occlusion* were not evaluated on SNP data due to computational infeasibility. The asterisk(*) on GRACES indicates that, due to computational constraints, GRACES was evaluated only up to a truncated feature budget for SNP data; results beyond this limit were not computed.



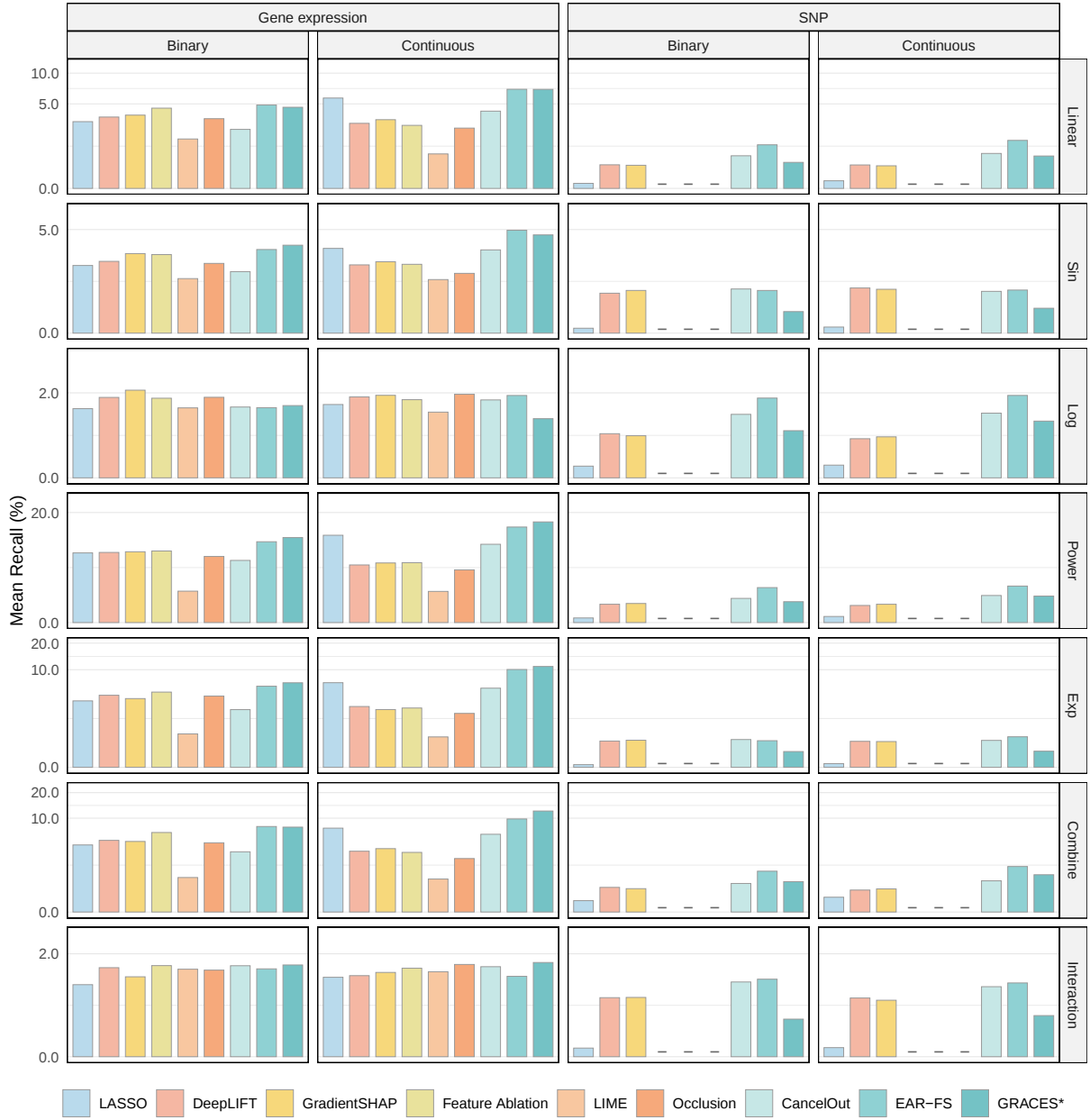
Supplementary Fig. S3: Recall of feature selection methods under a setting with 500 causal variables and a sample size of 1,000, evaluated across different top- K thresholds under Sigmoid activation. Top- K thresholds range from 1% to 100%, defined relative to the total number of features in each modality (26,000 for gene expression and 350,000 for SNP). The dashed line represents the chance-level baseline, corresponding to the expected performance under random selection. LIME, Feature Ablation, and Occlusion were not evaluated on SNP data due to computational infeasibility. The asterisk(*) on GRACES indicates that, due to computational constraints, GRACES was evaluated only up to a truncated feature budget for SNP data; results beyond this limit were not computed.



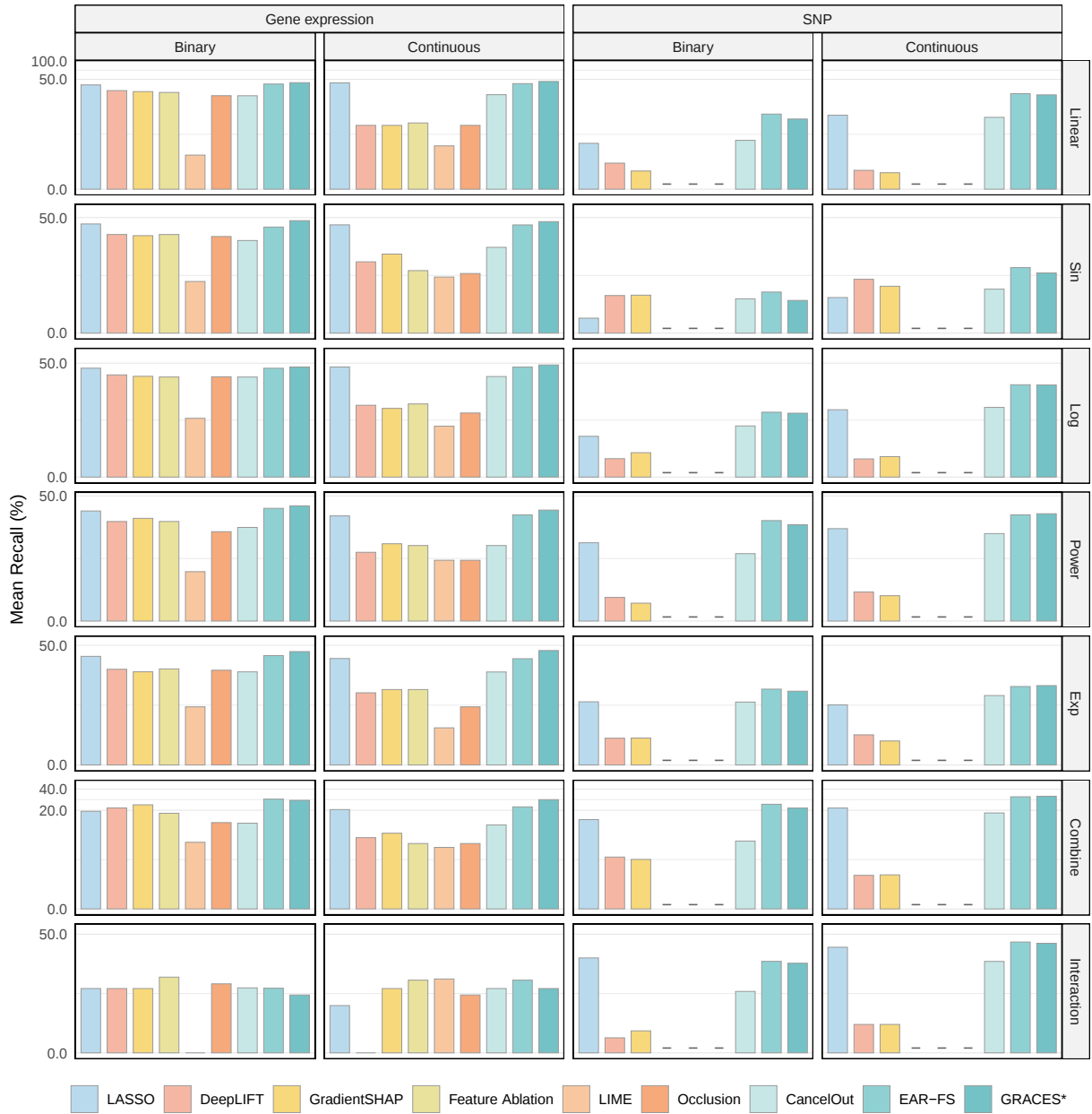
Supplementary Fig. S4: **Recall for different numbers of causal features (20, 100, and 500) at a fixed top-20% threshold for simulated gene expression and SNP data under Sigmoid activation (sample size $n = 5000$).** The top-20% threshold is defined relative to the total feature dimensionality of each modality (26,000 for gene expression and 350,000 for SNP). *LIME*, *Feature Ablation*, and *Occlusion* were not evaluated on SNP data due to computational infeasibility, marked as **NR**. The asterisk(*) on GRACES indicates that, due to computational constraints, GRACES was evaluated only up to a truncated feature budget for SNP data; results beyond this limit were not computed.



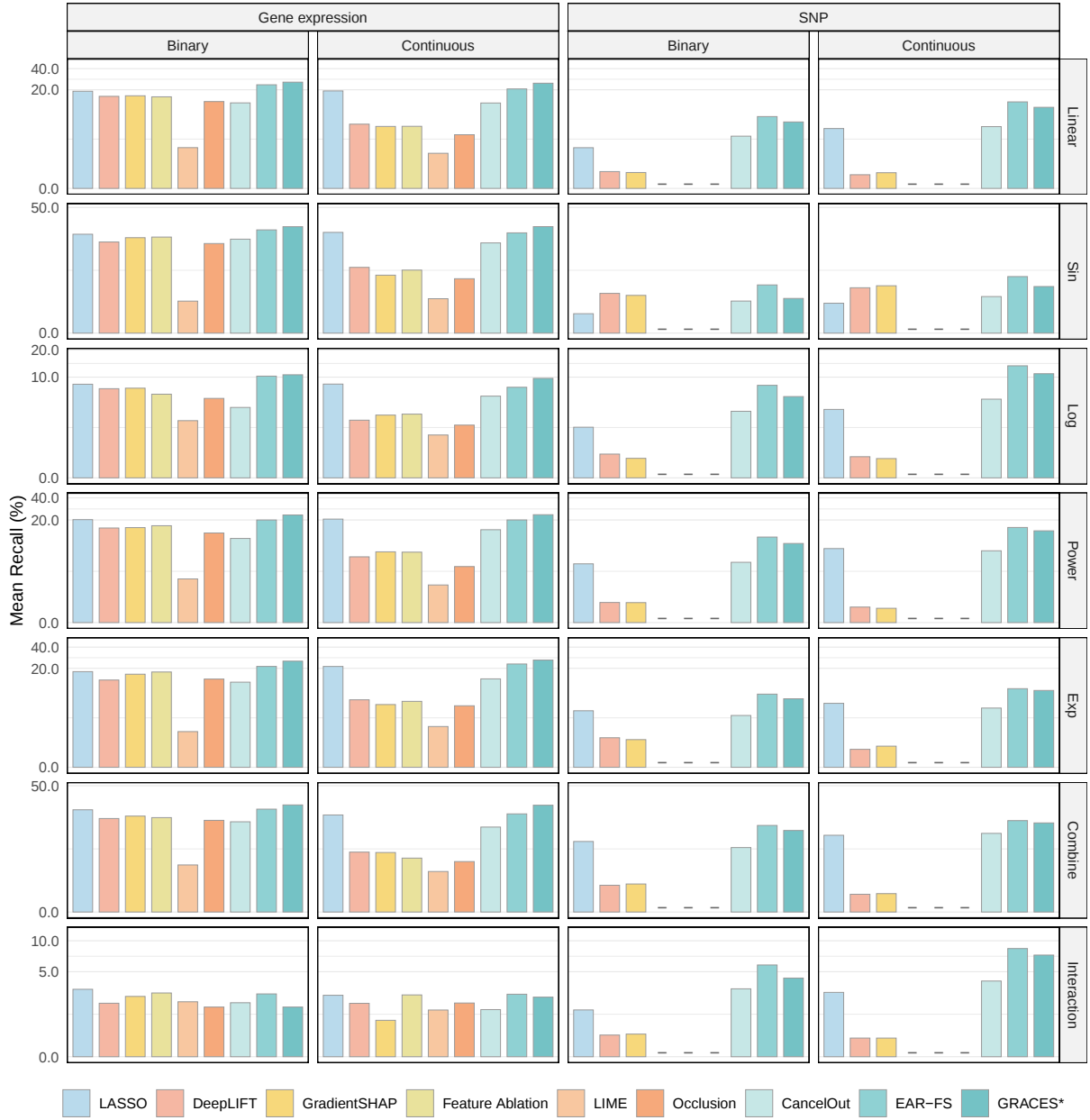
Supplementary Fig. S5: **Recall across feature selection methods under a setting with 20 causal variables and a sample size of 1,000, evaluated under seven functional relationships (linear, sine, logarithmic, power, exponential, combined that is defined as a composite per-predictor application of these transformations without interaction terms, and interaction).** The top-20% threshold is defined relative to the total feature dimensionality of each modality (26,000 for gene expression and 350,000 for SNP). *LIME*, *Feature Ablation*, and *Occlusion* were not evaluated for SNP data due to computational infeasibility. The asterisk(*) on GRACES indicates that, due to computational constraints, GRACES was evaluated only up to a truncated feature budget for SNP data; results beyond this limit were not computed.



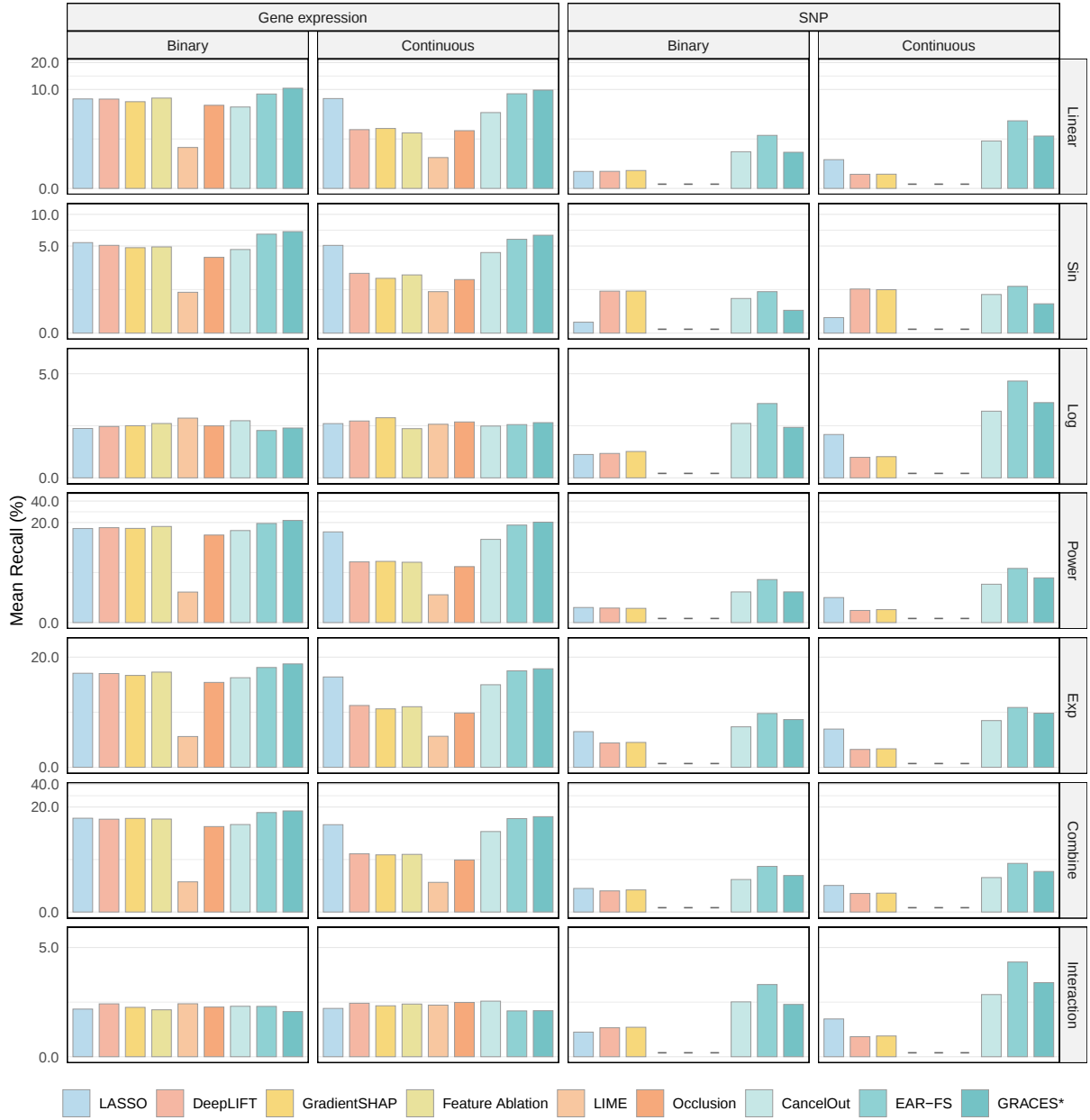
Supplementary Fig. S6: **Recall across feature selection methods under a setting with 500 causal variables and a sample size of 1,000, evaluated under seven functional relationships (linear, sine, logarithmic, power, exponential, combined that is defined as a composite per-predictor application of these transformations without interaction terms, and interaction).** The top-20% threshold is defined relative to the total feature dimensionality of each modality (26,000 for gene expression and 350,000 for SNP). *LIME*, *Feature Ablation*, and *Occlusion* were not evaluated for SNP data due to computational infeasibility. The asterisk(*) on GRACES indicates that, due to computational constraints, GRACES was evaluated only up to a truncated feature budget for SNP data; results beyond this limit were not computed.



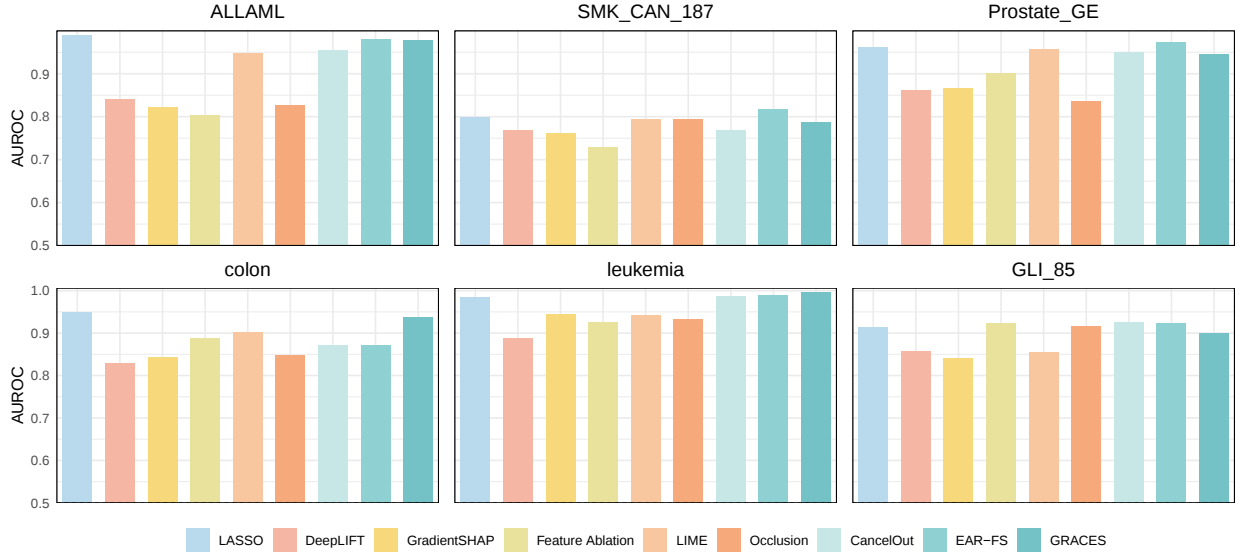
Supplementary Fig. S7: **Recall across feature selection methods under a setting with 20 causal variables and a sample size of 5,000, evaluated under seven functional relationships (linear, sine, logarithmic, power, exponential, combined that is defined as a composite per-predictor application of these transformations without interaction terms, and interaction).** The top-20% threshold is defined relative to the total feature dimensionality of each modality (26,000 for gene expression and 350,000 for SNP). *LIME*, *Feature Ablation*, and *Occlusion* were not evaluated for SNP data due to computational infeasibility. The asterisk(*) on GRACES indicates that, due to computational constraints, GRACES was evaluated only up to a truncated feature budget for SNP data; results beyond this limit were not computed.



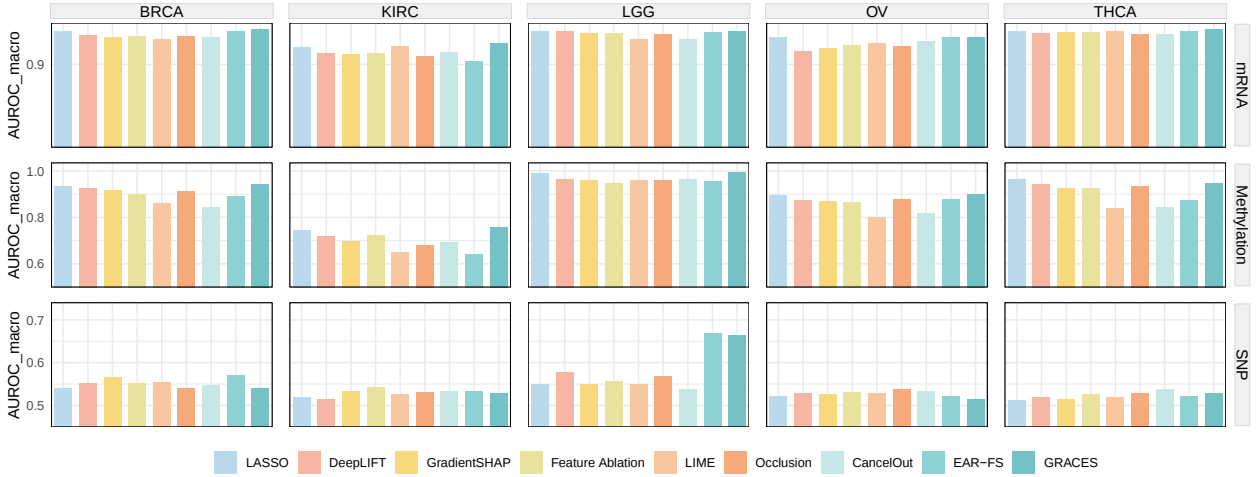
Supplementary Fig. S8: **Recall across feature selection methods under a setting with 100 causal variables and a sample size of 5,000, evaluated under seven functional relationships (linear, sine, logarithmic, power, exponential, combined that is defined as a composite per-predictor application of these transformations without interaction terms, and interaction).** The top-20% threshold is defined relative to the total feature dimensionality of each modality (26,000 for gene expression and 350,000 for SNP). *LIME*, *Feature Ablation*, and *Occlusion* were not evaluated for SNP data due to computational infeasibility. The asterisk(*) on GRACES indicates that, due to computational constraints, GRACES was evaluated only up to a truncated feature budget for SNP data; results beyond this limit were not computed.



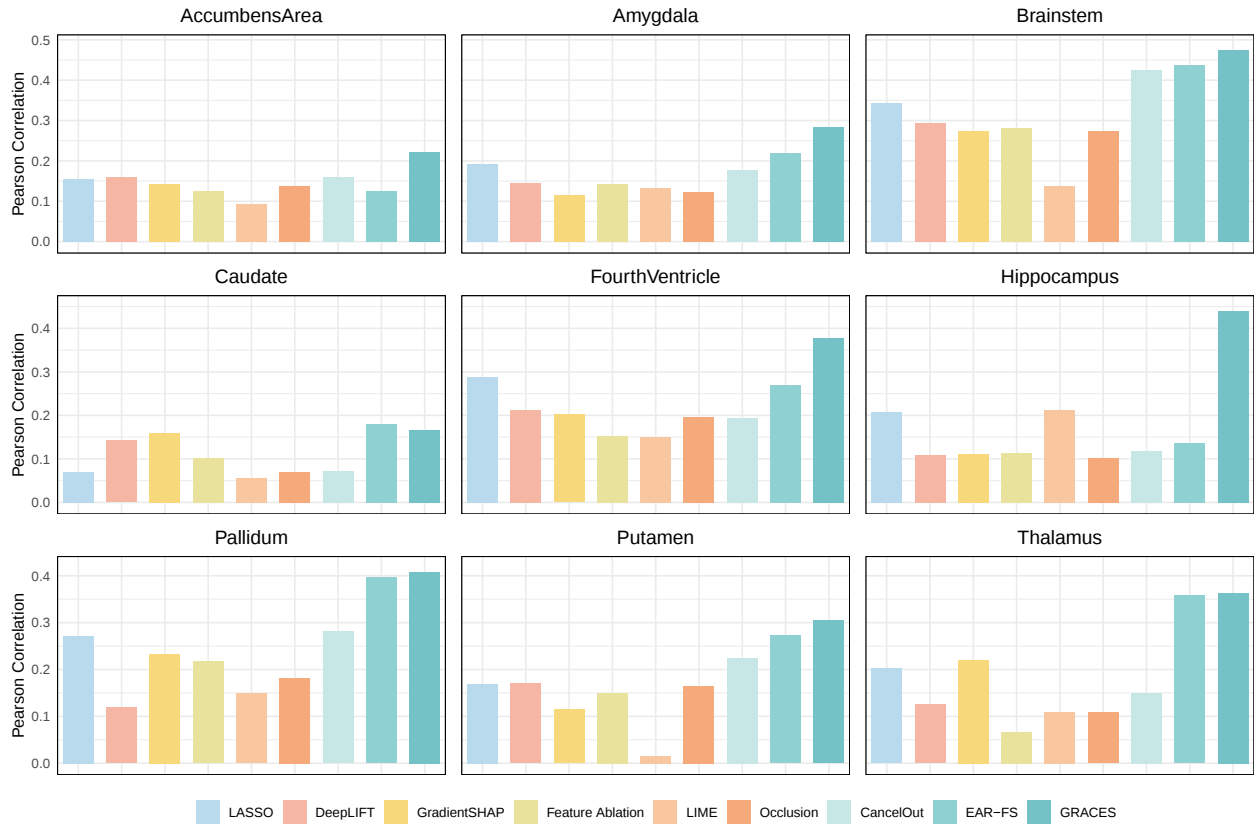
Supplementary Fig. S9: **Recall across feature selection methods under a setting with 500 causal variables and a sample size of 5,000, evaluated under seven functional relationships (linear, sine, logarithmic, power, exponential, combined that is defined as a composite per-predictor application of these transformations without interaction terms, and interaction).** The top-20% threshold is defined relative to the total feature dimensionality of each modality (26,000 for gene expression and 350,000 for SNP). *LIME*, *Feature Ablation*, and *Occlusion* were not evaluated for SNP data due to computational infeasibility. The asterisk(*) on GRACES indicates that, due to computational constraints, GRACES was evaluated only up to a truncated feature budget for SNP data; results beyond this limit were not computed.



Supplementary Fig. S10: **Performance comparison of feature selection methods on six public datasets under the Sigmoid activation.** The figure shows the average test AUROC achieved when selecting 100 features on datasets ALLAML, SMK_CAN_187, Prostate_GE, Colon, Leukemia, and GLI_85. Error bars represent the standard deviation across 10 runs.



Supplementary Fig. S11: **Performance comparison of feature selection methods on TCGA datasets under the Sigmoid activation.** The figure shows the macro-averaged test AUROC achieved when selecting 100 features across five cancer types (BRCA, KIRC, LGG, OV, THCA) and three molecular modalities (mRNA, DNA methylation, SNP). Error bars represent the standard deviation across 10 runs.



Supplementary Fig. S12: **Performance comparison of feature-selection methods on nine ADNI brain regions under the Sigmoid activation.** The figure shows the average test Pearson correlation achieved when selecting 100 features across 9 brain regions (Accumbens Area, Amygdala, Brainstem, Caudate, Fourth Ventricle, Hippocampus, Pallidum, Putamen, Thalamus). Error bars represent the standard deviation across 10 runs.

Supplementart Tables

Supplementary Table S1: Mean peak memory (GB) across feature selection methods for all real datasets under the Sigmoid activation.

Project	Modality	Dataset	Method							
			DeepLIFT	GradientShap	FeatureAblation	LIME	Occlusion	CANCELOUT	EAR-FS	GRACES
Public Datasets	mRNA	ALLAML	1.52	1.67	1.51	1.67	1.46	1.34	1.37	1.56
		SMK_CAN_187	3.19	3.06	3.06	3.15	3.17	3.17	2.95	3.22
		Prostate_GE	1.37	1.58	1.35	1.64	1.52	1.50	1.41	1.85
		colon	0.56	0.61	0.56	0.60	0.54	0.60	0.56	0.65
		leukemia	1.70	1.43	1.44	1.65	1.59	1.49	1.52	1.60
		GLI85	3.43	3.24	3.43	3.13	3.21	3.14	3.39	3.00
TCGA	mRNA	BRCA	2.19	2.20	2.30	2.44	2.26	2.22	2.18	3.28
		KIRC	1.89	2.01	1.98	1.92	1.96	1.96	2.07	2.45
		LGG	2.11	2.05	2.10	2.12	2.07	2.14	2.07	2.67
		OV	1.78	2.04	1.94	2.05	1.92	1.95	1.91	2.50
		THCA	1.90	2.11	1.95	2.12	1.90	1.78	2.05	2.49
	methy	BRCA	17.41	17.36	16.54	17.01	16.74	17.25	16.96	17.50
		KIRC	15.54	16.25	15.61	15.82	16.01	16.72	16.33	14.40
		LGG	16.14	17.70	16.39	17.11	17.40	17.47	15.48	17.33
		OV	1.53	1.49	1.52	1.47	1.47	1.47	1.48	1.80
		THCA	17.64	15.38	18.22	17.30	16.73	17.84	18.28	17.09
	SNP	BRCA	42.01	40.82	40.53	39.76	39.08	41.19	41.75	33.30
		KIRC	27.36	26.45	27.34	26.79	27.51	27.74	29.12	23.08
		LGG	29.74	29.21	28.46	28.80	30.75	30.79	30.13	24.87
		OV	19.37	19.53	18.50	20.11	19.21	19.37	20.07	15.77
		THCA	29.12	29.59	29.52	28.37	28.82	30.10	29.75	24.72
ADNI	mRNA	AccumbensArea	8.59	8.89	8.70	8.77	8.48	8.78	8.87	9.22
		Amygdala	8.31	8.57	8.03	5.59	8.16	8.66	8.98	9.05
		Brainstem	8.22	8.56	8.29	7.92	8.54	8.93	8.81	9.22
		Caudate	8.36	8.52	8.47	7.89	8.41	8.41	8.56	9.10
		FourthVentricle	8.76	8.77	9.02	8.08	8.60	8.39	8.67	9.06
		Hippocampus	8.68	8.49	8.46	8.62	8.19	8.32	8.70	8.12
		Pallidum	8.49	8.42	8.49	8.61	8.41	8.68	8.41	8.67
		Putamen	8.47	8.25	8.32	8.59	8.34	8.34	8.26	8.79
		Thalamus	8.11	8.44	8.45	8.69	8.08	8.50	8.60	9.31

Supplementary Table S2: Hyperparameter tuning frequencies across datasets.

Project	Modality	Dataset	Method	Hidden size scaling factor*					Batch size		
				100	200	500	1000	2000	8	16	32
Public Dataset	mRNA	ALLAML	DeepLIFT	0.447	0.23	0.187	0.136	-	0.32	0.394	0.287
			GradientShap	0.427	0.289	0.139	0.145	-	0.357	0.342	0.302
			FeatureAblation	0.396	0.267	0.193	0.144	-	0.249	0.49	0.261
			LIME	0.328	0.317	0.195	0.16	-	0.278	0.474	0.248
			Occlusion	0.364	0.333	0.164	0.138	-	0.316	0.47	0.214
			CANCELOUT	0.414	0.301	0.156	0.129	-	0.362	0.305	0.333
			EAR-FS	0.304	0.317	0.216	0.163	-	0.277	0.404	0.319
			GRACES	0.267	0.284	0.286	0.164	-	0.174	0.549	0.277
		GLL85	DeepLIFT	0.403	0.259	0.169	0.169	-	0.333	0.267	0.4
			GradientShap	0.333	0.288	0.237	0.142	-	0.367	0.202	0.431
			FeatureAblation	0.345	0.255	0.235	0.165	-	0.446	0.294	0.26
			LIME	0.223	0.334	0.222	0.221	-	0.318	0.237	0.445
			Occlusion	0.364	0.296	0.225	0.115	-	0.268	0.307	0.425
			CANCELOUT	0.379	0.297	0.173	0.151	-	0.305	0.208	0.487
			EAR-FS	0.332	0.28	0.232	0.156	-	0.382	0.267	0.351
			GRACES	0.223	0.301	0.238	0.238	-	0.288	0.24	0.472
		Prostate_GE	DeepLIFT	0.367	0.331	0.198	0.104	-	0.212	0.595	0.193
			GradientShap	0.379	0.292	0.152	0.177	-	0.297	0.471	0.232
			FeatureAblation	0.349	0.323	0.194	0.134	-	0.319	0.43	0.251
			LIME	0.478	0.248	0.162	0.112	-	0.283	0.546	0.171
			Occlusion	0.405	0.255	0.208	0.132	-	0.297	0.551	0.152
			CANCELOUT	0.485	0.233	0.156	0.126	-	0.412	0.401	0.187
			EAR-FS	0.403	0.332	0.159	0.106	-	0.326	0.478	0.196
			GRACES	0.363	0.277	0.215	0.145	-	0.246	0.302	0.451
		SMK_CAN_187	DeepLIFT	0.463	0.268	0.132	0.137	-	0.27	0.368	0.362
			GradientShap	0.45	0.263	0.153	0.134	-	0.232	0.452	0.316
			FeatureAblation	0.404	0.271	0.17	0.155	-	0.289	0.312	0.399
			LIME	0.436	0.291	0.173	0.1	-	0.268	0.31	0.422
			Occlusion	0.459	0.282	0.149	0.11	-	0.274	0.301	0.425
			CANCELOUT	0.443	0.212	0.205	0.14	-	0.334	0.298	0.368
			EAR-FS	0.495	0.253	0.125	0.127	-	0.206	0.322	0.472
			GRACES	0.352	0.231	0.217	0.2	-	0.278	0.276	0.446
		colon	DeepLIFT	0.493	0.213	0.154	0.14	-	0.597	0.261	0.142
			GradientShap	0.459	0.27	0.143	0.128	-	0.573	0.241	0.186
			FeatureAblation	0.463	0.248	0.152	0.137	-	0.655	0.191	0.154
			LIME	0.354	0.309	0.209	0.128	-	0.615	0.228	0.157
			Occlusion	0.43	0.289	0.153	0.128	-	0.561	0.25	0.189
			CANCELOUT	0.412	0.313	0.144	0.131	-	0.514	0.279	0.207
			EAR-FS	0.511	0.235	0.134	0.12	-	0.46	0.32	0.22
			GRACES	0.341	0.311	0.204	0.144	-	0.477	0.371	0.152
		leukemia	DeepLIFT	0.369	0.294	0.198	0.139	-	0.343	0.416	0.241
			GradientShap	0.401	0.276	0.196	0.126	-	0.361	0.337	0.302
			FeatureAblation	0.45	0.25	0.148	0.151	-	0.241	0.454	0.305
			LIME	0.257	0.356	0.182	0.204	-	0.278	0.416	0.306
			Occlusion	0.337	0.345	0.205	0.113	-	0.386	0.329	0.285
			CANCELOUT	0.345	0.277	0.243	0.135	-	0.323	0.467	0.21
			EAR-FS	0.335	0.309	0.194	0.162	-	0.387	0.372	0.241
			GRACES	0.356	0.279	0.231	0.134	-	0.196	0.496	0.308
TCGA	mRNA	BRCA	DeepLIFT	0.327	0.34	0.208	0.125	-	0.417	0.32	0.263
			GradientShap	0.335	0.37	0.171	0.124	-	0.327	0.283	0.39
			FeatureAblation	0.406	0.294	0.159	0.141	-	0.253	0.435	0.312
			LIME	0.38	0.294	0.185	0.141	-	0.244	0.363	0.393

Project	Modality	Dataset	Method	Hidden size scaling factor*					Batch size		
				100	200	500	1000	2000	8	16	32
			Occlusion	0.398	0.267	0.195	0.14	-	0.36	0.295	0.345
			CANCELOUT	0.412	0.309	0.135	0.144	-	0.341	0.347	0.312
			EAR-FS	0.309	0.311	0.241	0.139	-	0.343	0.316	0.341
			GRACES	0.437	0.264	0.138	0.161	-	0.24	0.188	0.572
			DeepLIFT	0.278	0.363	0.186	0.173	-	0.268	0.242	0.49
			GradientShap	0.411	0.246	0.23	0.113	-	0.225	0.233	0.542
			FeatureAblation	0.301	0.378	0.186	0.135	-	0.207	0.203	0.59
			LIME	0.422	0.31	0.136	0.132	-	0.255	0.282	0.463
			Occlusion	0.276	0.289	0.283	0.152	-	0.333	0.234	0.433
			CANCELOUT	0.328	0.327	0.201	0.144	-	0.215	0.261	0.524
			EAR-FS	0.451	0.193	0.219	0.137	-	0.265	0.297	0.438
			GRACES	0.36	0.215	0.244	0.181	-	0.257	0.347	0.396
			DeepLIFT	0.339	0.375	0.143	0.143	-	0.245	0.461	0.294
			GradientShap	0.381	0.317	0.161	0.141	-	0.333	0.272	0.395
			FeatureAblation	0.354	0.347	0.155	0.144	-	0.419	0.343	0.238
			LIME	0.367	0.323	0.172	0.138	-	0.266	0.505	0.229
			Occlusion	0.403	0.229	0.237	0.131	-	0.27	0.428	0.302
			CANCELOUT	0.448	0.268	0.142	0.142	-	0.31	0.307	0.383
			EAR-FS	0.314	0.379	0.152	0.155	-	0.258	0.539	0.203
			GRACES	0.351	0.236	0.215	0.198	-	0.211	0.376	0.413
			DeepLIFT	0.416	0.249	0.203	0.131	-	0.28	0.266	0.454
			GradientShap	0.412	0.261	0.201	0.126	-	0.264	0.388	0.348
			FeatureAblation	0.293	0.393	0.185	0.129	-	0.257	0.274	0.469
			LIME	0.347	0.31	0.207	0.136	-	0.208	0.341	0.451
			Occlusion	0.471	0.2	0.189	0.139	-	0.241	0.318	0.441
			CANCELOUT	0.384	0.354	0.14	0.122	-	0.2	0.278	0.522
			EAR-FS	0.297	0.293	0.219	0.191	-	0.269	0.291	0.44
			GRACES	0.335	0.314	0.175	0.176	-	0.265	0.392	0.343
			DeepLIFT	0.439	0.287	0.142	0.132	-	0.318	0.312	0.37
			GradientShap	0.359	0.318	0.178	0.145	-	0.339	0.4	0.261
			FeatureAblation	0.346	0.366	0.162	0.126	-	0.291	0.357	0.353
			LIME	0.399	0.286	0.185	0.13	-	0.308	0.327	0.365
			Occlusion	0.296	0.431	0.154	0.12	-	0.232	0.434	0.334
			CANCELOUT	0.494	0.235	0.152	0.119	-	0.342	0.385	0.273
			EAR-FS	0.397	0.288	0.193	0.123	-	0.243	0.499	0.258
			GRACES	0.329	0.313	0.193	0.165	-	0.184	0.218	0.598
Methy			DeepLIFT	0.404	0.227	0.192	0.177	-	0.242	0.314	0.444
			GradientShap	0.406	0.211	0.216	0.167	-	0.264	0.267	0.469
			FeatureAblation	0.323	0.268	0.192	0.217	-	0.241	0.348	0.411
			LIME	0.352	0.279	0.223	0.146	-	0.298	0.309	0.393
			Occlusion	0.294	0.205	0.258	0.243	-	0.274	0.361	0.365
			CANCELOUT	0.45	0.236	0.172	0.142	-	0.28	0.258	0.462
			EAR-FS	0.395	0.253	0.213	0.139	-	0.222	0.304	0.474
			GRACES	0.424	0.22	0.193	0.163	-	0.169	0.162	0.669
			DeepLIFT	0.455	0.275	0.147	0.123	-	0.195	0.286	0.519
			GradientShap	0.467	0.238	0.162	0.133	-	0.263	0.218	0.519
			FeatureAblation	0.382	0.308	0.161	0.149	-	0.239	0.258	0.503
			LIME	0.462	0.197	0.207	0.134	-	0.248	0.202	0.55
			Occlusion	0.449	0.258	0.143	0.15	-	0.288	0.225	0.487
			CANCELOUT	0.465	0.253	0.157	0.125	-	0.268	0.283	0.449
			EAR-FS	0.454	0.261	0.165	0.12	-	0.282	0.345	0.373
			GRACES	0.386	0.235	0.189	0.19	-	0.186	0.19	0.624
			DeepLIFT	0.18	0.344	0.201	0.275	-	0.206	0.408	0.386
			GradientShap	0.258	0.339	0.243	0.16	-	0.256	0.533	0.211
			FeatureAblation	0.185	0.281	0.355	0.179	-	0.311	0.406	0.283
			DeepLIFT	0.18	0.344	0.201	0.275	-	0.206	0.408	0.386
			GradientShap	0.258	0.339	0.243	0.16	-	0.256	0.533	0.211
			FeatureAblation	0.185	0.281	0.355	0.179	-	0.311	0.406	0.283
			DeepLIFT	0.18	0.344	0.201	0.275	-	0.206	0.408	0.386
			GradientShap	0.258	0.339	0.243	0.16	-	0.256	0.533	0.211
			FeatureAblation	0.185	0.281	0.355	0.179	-	0.311	0.406	0.283
			DeepLIFT	0.18	0.344	0.201	0.275	-	0.206	0.408	0.386
			GradientShap	0.258	0.339	0.243	0.16	-	0.256	0.533	0.211
			FeatureAblation	0.185	0.281	0.355	0.179	-	0.311	0.406	0.283

LGG

Project	Modality	Dataset	Method	Hidden size scaling factor*					Batch size		
				100	200	500	1000	2000	8	16	32
THCA	OV	LIME	0.26	0.311	0.241	0.188	-	0.27	0.358	0.372	
		Occlusion	0.246	0.269	0.323	0.162	-	0.197	0.467	0.336	
		CANCELOUT	0.278	0.353	0.241	0.128	-	0.278	0.322	0.4	
		EAR-FS	0.197	0.316	0.257	0.23	-	0.284	0.412	0.304	
		GRACES	0.296	0.236	0.281	0.187	-	0.203	0.433	0.364	
		DeepLIFT	0.484	0.205	0.173	0.138	-	0.345	0.311	0.344	
		GradientShap	0.452	0.274	0.141	0.133	-	0.382	0.328	0.29	
		FeatureAblation	0.49	0.211	0.17	0.129	-	0.324	0.296	0.38	
		LIME	0.511	0.223	0.128	0.138	-	0.26	0.402	0.338	
		Occlusion	0.521	0.237	0.131	0.111	-	0.342	0.376	0.282	
		CANCELOUT	0.386	0.324	0.156	0.134	-	0.296	0.37	0.334	
		EAR-FS	0.422	0.305	0.146	0.127	-	0.259	0.324	0.417	
		GRACES	0.405	0.293	0.188	0.114	-	0.166	0.262	0.572	
		DeepLIFT	0.291	0.283	0.226	0.2	-	0.32	0.213	0.467	
		GradientShap	0.228	0.36	0.243	0.169	-	0.219	0.222	0.559	
	FeatureAblation	0.437	0.208	0.19	0.165	-	0.235	0.241	0.524		
	LIME	0.315	0.267	0.273	0.145	-	0.207	0.214	0.58		
	Occlusion	0.316	0.29	0.154	0.24	-	0.281	0.216	0.503		
	CANCELOUT	0.36	0.288	0.157	0.194	-	0.168	0.357	0.476		
	EAR-FS	0.355	0.243	0.212	0.19	-	0.188	0.29	0.522		
	GRACES	0.276	0.3	0.236	0.188	-	0.171	0.148	0.681		
SNP	BRCA	DeepLIFT	-	0.372	0.269	0.234	0.125	0.221	0.351	0.428	
		GradientShap	-	0.314	0.287	0.217	0.182	0.217	0.33	0.453	
		FeatureAblation	-	0.329	0.335	0.219	0.117	0.205	0.385	0.41	
		LIME	-	0.369	0.258	0.228	0.145	0.186	0.27	0.544	
		Occlusion	-	0.324	0.327	0.195	0.154	0.202	0.275	0.523	
		CANCELOUT	-	0.391	0.289	0.174	0.146	0.201	0.264	0.535	
		EAR-FS	-	0.348	0.318	0.17	0.164	0.266	0.396	0.338	
		GRACES	-	0.197	0.301	0.283	0.219	0.194	0.288	0.518	
	KIRC	DeepLIFT	-	0.238	0.287	0.334	0.14	0.207	0.326	0.467	
		GradientShap	-	0.234	0.303	0.29	0.173	0.242	0.22	0.538	
		FeatureAblation	-	0.256	0.192	0.305	0.247	0.303	0.256	0.441	
		LIME	-	0.301	0.317	0.188	0.194	0.18	0.208	0.612	
		Occlusion	-	0.278	0.338	0.194	0.19	0.273	0.203	0.524	
		CANCELOUT	-	0.243	0.318	0.27	0.169	0.193	0.299	0.508	
		EAR-FS	-	0.283	0.343	0.226	0.148	0.31	0.245	0.445	
		GRACES	-	0.163	0.211	0.327	0.299	0.264	0.306	0.43	
	LGG	DeepLIFT	-	0.483	0.222	0.139	0.156	0.257	0.273	0.47	
		GradientShap	-	0.527	0.167	0.175	0.131	0.231	0.392	0.377	
		FeatureAblation	-	0.44	0.2	0.195	0.165	0.343	0.419	0.238	
		LIME	-	0.385	0.262	0.197	0.156	0.212	0.405	0.383	
		Occlusion	-	0.512	0.198	0.14	0.15	0.27	0.341	0.389	
CANCELOUT		-	0.502	0.207	0.14	0.151	0.183	0.511	0.306		
EAR-FS		-	0.474	0.222	0.168	0.136	0.224	0.469	0.307		
GRACES		-	0.423	0.282	0.147	0.148	0.263	0.313	0.424		
OV	DeepLIFT	-	0.295	0.236	0.313	0.156	0.245	0.335	0.42		
	GradientShap	-	0.224	0.291	0.234	0.251	0.285	0.295	0.42		
	FeatureAblation	-	0.207	0.301	0.278	0.214	0.291	0.26	0.449		
	LIME	-	0.332	0.222	0.26	0.186	0.396	0.324	0.28		
	Occlusion	-	0.216	0.296	0.202	0.286	0.319	0.311	0.37		
	CANCELOUT	-	0.267	0.307	0.228	0.198	0.28	0.299	0.421		
	EAR-FS	-	0.27	0.223	0.316	0.191	0.31	0.339	0.351		
	GRACES	-	0.156	0.229	0.372	0.243	0.186	0.212	0.602		
	DeepLIFT	-	0.295	0.357	0.15	0.198	0.27	0.325	0.405		
	GradientShap	-	0.257	0.234	0.286	0.223	0.321	0.314	0.365		

Project	Modality	Dataset	Method	Hidden size scaling factor*					Batch size		
				100	200	500	1000	2000	8	16	32
ADNI	mRNA		FeatureAblation	-	0.217	0.323	0.225	0.235	0.314	0.34	0.346
			LIME	-	0.211	0.321	0.283	0.185	0.331	0.348	0.321
			Occlusion	-	0.259	0.292	0.273	0.176	0.309	0.331	0.36
			CANCELOUT	-	0.229	0.387	0.216	0.168	0.345	0.322	0.333
			EAR-FS	-	0.215	0.275	0.31	0.2	0.299	0.348	0.353
			GRACES	-	0.162	0.328	0.298	0.212	0.151	0.232	0.617
			DeepLIFT	0.187	0.167	0.323	0.323	-	0.487	0.29	0.223
			GradientShap	0.197	0.205	0.351	0.247	-	0.335	0.19	0.475
			FeatureAblation	0.168	0.181	0.257	0.394	-	0.39	0.214	0.396
			LIME	0.186	0.234	0.25	0.33	-	0.452	0.203	0.345
		AccumbensArea	Occlusion	0.174	0.243	0.231	0.352	-	0.365	0.208	0.427
			CANCELOUT	0.144	0.152	0.319	0.385	-	0.451	0.209	0.34
			EAR-FS	0.141	0.237	0.377	0.245	-	0.547	0.208	0.245
			GRACES	0.201	0.163	0.277	0.359	-	0.602	0.238	0.16
			DeepLIFT	0.132	0.175	0.33	0.363	-	0.601	0.177	0.222
			GradientShap	0.207	0.194	0.315	0.284	-	0.567	0.176	0.256
			FeatureAblation	0.2	0.215	0.357	0.228	-	0.379	0.251	0.37
			LIME	0.127	0.236	0.399	0.238	-	0.408	0.229	0.364
		Amygdala	Occlusion	0.129	0.184	0.371	0.316	-	0.422	0.145	0.433
			CANCELOUT	0.215	0.177	0.224	0.384	-	0.396	0.196	0.408
			EAR-FS	0.213	0.209	0.401	0.177	-	0.487	0.191	0.322
			GRACES	0.228	0.187	0.202	0.382	-	0.402	0.28	0.318
			DeepLIFT	0.171	0.288	0.316	0.225	-	0.428	0.188	0.384
			GradientShap	0.21	0.206	0.364	0.22	-	0.346	0.229	0.425
			FeatureAblation	0.218	0.251	0.245	0.286	-	0.479	0.159	0.363
			LIME	0.183	0.254	0.282	0.281	-	0.362	0.249	0.39
		Brainstem	Occlusion	0.217	0.223	0.258	0.302	-	0.48	0.178	0.342
			CANCELOUT	0.193	0.239	0.227	0.341	-	0.204	0.273	0.523
			EAR-FS	0.271	0.189	0.348	0.192	-	0.365	0.225	0.41
			GRACES	0.223	0.256	0.198	0.323	-	0.565	0.292	0.143
			DeepLIFT	0.148	0.251	0.266	0.335	-	0.279	0.168	0.553
			GradientShap	0.166	0.241	0.287	0.306	-	0.275	0.209	0.516
			FeatureAblation	0.158	0.164	0.295	0.383	-	0.441	0.149	0.41
			LIME	0.198	0.18	0.306	0.316	-	0.335	0.153	0.512
		Caudate	Occlusion	0.2	0.164	0.332	0.304	-	0.306	0.196	0.498
			CANCELOUT	0.153	0.176	0.293	0.378	-	0.399	0.148	0.453
			EAR-FS	0.182	0.178	0.243	0.397	-	0.41	0.257	0.333
			GRACES	0.164	0.164	0.218	0.454	-	0.602	0.244	0.154
			DeepLIFT	0.218	0.338	0.259	0.185	-	0.527	0.184	0.289
			GradientShap	0.194	0.376	0.288	0.142	-	0.319	0.198	0.483
			FeatureAblation	0.202	0.3	0.27	0.228	-	0.332	0.196	0.472
			LIME	0.184	0.24	0.288	0.288	-	0.196	0.191	0.613
		FourthVentricle	Occlusion	0.194	0.207	0.268	0.331	-	0.407	0.15	0.443
			CANCELOUT	0.145	0.205	0.382	0.267	-	0.379	0.188	0.433
			EAR-FS	0.218	0.328	0.272	0.181	-	0.277	0.151	0.572
			GRACES	0.176	0.199	0.254	0.371	-	0.495	0.336	0.169
			DeepLIFT	0.209	0.144	0.244	0.402	-	0.333	0.271	0.396
			GradientShap	0.15	0.155	0.343	0.353	-	0.491	0.159	0.35
			FeatureAblation	0.14	0.155	0.291	0.414	-	0.47	0.152	0.378
			LIME	0.157	0.154	0.263	0.426	-	0.403	0.217	0.38
		Hippocampus	Occlusion	0.132	0.237	0.335	0.296	-	0.41	0.182	0.408
			CANCELOUT	0.163	0.244	0.313	0.28	-	0.297	0.18	0.523
			EAR-FS	0.137	0.19	0.287	0.385	-	0.486	0.177	0.337
			GRACES	0.188	0.201	0.167	0.444	-	0.651	0.176	0.173
			DeepLIFT	0.182	0.297	0.264	0.257	-	0.333	0.16	0.507

Project	Modality	Dataset	Method	Hidden size scaling factor*					Batch size		
				100	200	500	1000	2000	8	16	32
			GradientShap	0.233	0.205	0.236	0.326	-	0.302	0.225	0.473
			FeatureAblation	0.159	0.266	0.296	0.279	-	0.505	0.16	0.335
			LIME	0.169	0.235	0.303	0.293	-	0.395	0.242	0.363
			Occlusion	0.175	0.232	0.315	0.278	-	0.346	0.206	0.448
			CANCELOUT	0.193	0.29	0.214	0.303	-	0.301	0.188	0.511
			EAR-FS	0.203	0.254	0.224	0.319	-	0.408	0.158	0.434
			GRACES	0.138	0.173	0.247	0.442	-	0.589	0.263	0.147
			DeepLIFT	0.135	0.21	0.27	0.384	-	0.308	0.175	0.517
			GradientShap	0.169	0.256	0.22	0.355	-	0.257	0.224	0.519
			FeatureAblation	0.176	0.197	0.298	0.329	-	0.359	0.161	0.48
			LIME	0.182	0.226	0.301	0.291	-	0.392	0.155	0.453
			Occlusion	0.138	0.182	0.276	0.404	-	0.347	0.187	0.466
			CANCELOUT	0.189	0.206	0.218	0.387	-	0.28	0.176	0.544
			EAR-FS	0.155	0.164	0.207	0.474	-	0.493	0.161	0.346
			GRACES	0.158	0.157	0.188	0.497	-	0.524	0.305	0.171
			DeepLIFT	0.139	0.171	0.32	0.37	-	0.477	0.213	0.31
			GradientShap	0.15	0.17	0.355	0.326	-	0.442	0.17	0.388
			FeatureAblation	0.177	0.277	0.24	0.306	-	0.305	0.194	0.501
			LIME	0.221	0.222	0.189	0.368	-	0.403	0.192	0.405
			Occlusion	0.14	0.251	0.342	0.267	-	0.399	0.218	0.383
			CANCELOUT	0.156	0.256	0.249	0.339	-	0.379	0.195	0.426
			EAR-FS	0.142	0.178	0.368	0.313	-	0.339	0.189	0.473
			GRACES	0.158	0.184	0.309	0.349	-	0.572	0.266	0.162

* Hidden size scaling factor is defined as (input feature dimension) divided by this value, determining the size of the hidden layers in the deep learning models.

Supplementary Table S3: Hyperparameter value ranges used in simulation and real data experiments.

Parameter Type	Name	Simulation	Real Data
General	Dropout Rate	0.2–0.8	0.2–0.8
	Learning Rate	1e-5–1e-1	1e-5–1e-1
	Batch Size	32	8, 16, 32
	Hidden Size scaling factor	200	100, 200, 500, 1000, 2000
	L2 Regularization	0.0–1e-1	0.0–1e-1
CancelOut	λ_1	1e-5–1e-1	1e-5–1e-1
	λ_2	1e-5–1e-1	1e-5–1e-1
GRACES	f_{correct}	0, 0.1, 0.5, 0.9	0, 0.1, 0.5, 0.9
	α	0.8–0.99	0.8–0.99

Supplementary Table S4: Causal variant allocation across effect types for gene expression and SNP modalities.

Effect type	Gene expression			SNP		
	20	100	500	20	100	500
Linear	3	14	67	3 (30 loci)	14 (140 loci)	67 (670 loci)
Sin	3	14	67	3 (30 loci)	14 (140 loci)	67 (670 loci)
Logarithm	3	13	67	3 (30 loci)	13 (130 loci)	67 (670 loci)
Power	3	13	67	3 (30 loci)	13 (130 loci)	67 (670 loci)
Exponential	3	13	66	3 (30 loci)	13 (130 loci)	66 (660 loci)
Combined	3	13	66	3 (30 loci)	13 (130 loci)	66 (660 loci)
Interaction	2	20	100	2 (20 loci)	20 (200 loci)	100 (1000 loci)

Supplementary Table S5: Effect sizes used for different functional forms and sample sizes in Gene expression and SNP data.

Function	Gene expression						SNP					
	1000			5000			1000			5000		
Type	20	100	500	20	100	500	20	100	500	20	100	500
Linear	0.1	0.3	4.0	0.1	0.15	1.0	0.2	0.3	1.0	0.05	0.15	0.5
Cosine	0.15	0.5	5.0	0.15	0.25	1.25	0.4	0.5	2.0	0.1	0.5	1.5
Logarithm	1.5	2.0	5.0	1.5	1.0	1.25	0.4	1	5.0	0.2	0.7	2.5
Power	0.05	0.1	2.0	0.025	0.05	0.5	0.001	0.002	0.01	0.0015	0.004	0.01
Exponential	0.1	0.2	3.0	0.05	0.1	0.75	0.0001	0.0002	0.0001	0.00008	0.0004	0.001
Combined	0.025	0.06	1.0	0.0125	0.03	0.25	0.0001	0.0001	0.0004	0.00025	0.0004	0.003
Interaction	0.25	0.8	1.5	0.25	0.4	0.375	0.02	0.05	1.5	0.01	0.025	0.1

Supplementary Table S6: Summary of real-world datasets.

Dataset	Disease area	Outcome	Endpoint	Modality	Samples ^c	Features
Colon	Colorectal cancer	Binary	Tumor vs. Control	mRNA	62 (40, 22)	2,000
Leukemia	Hematologic malignancy	Binary	Acute lymphoblastic leukemia vs. Control	mRNA	72 (47, 25)	7,070
ALLAML	Hematologic malignancy	Binary	Acute lymphoblastic leukemia vs. Acute myeloid leukemia	mRNA	72 (47, 25)	7,129
GLI85	Neuro-oncology	Binary	Glioma vs. Control	mRNA	85 (26, 59)	22,283
Prostate_GE	Prostate cancer	Binary	Tumor vs. Control	mRNA	102 (50, 52)	5,966
SMK_CAN_187	Lung cancer (smokers)	Binary	Cancer vs. non-Cancer	mRNA	187 (90, 97)	19,993
TCGA-BRCA	Breast cancer	Multi-class	Basal / HER2+ / LumA / LumB / Normal	mRNA	1,094 (190, 82, 565, 217, 40)	14,133
				Methylation	781 (135, 46, 419, 147, 34)	68,521
				Genotyping array	1,094 (190, 81, 566, 217, 40)	163,135
				mRNA	412 (108, 90, 98, 116)	14,488
TCGA-OV	Ovarian cancer	Multi-class	Differentiated / Immunoreactive / Mesenchymal / Proliferative	Methylation	488 (135, 107, 108, 138)	8,085
				Genotyping array	484 (135, 107, 106, 136)	110,068
TCGA-THCA	Thyroid cancer	Multi-class	Stage1-5	mRNA	482 (134, 67, 82, 106, 93)	14,062
				Methylation	482 (134, 67, 82, 106, 93)	72,413
				Genotyping array	480 (134, 67, 82, 105, 92)	144,658
TCGA-KIRC	Kidney clear cell	Multi-class	Grade1-4	mRNA	417 (147, 90, 94, 86)	14,220
				Methylation	234 (91, 36, 53, 54)	70,090
				Genotyping array	415 (146, 90, 93, 86)	139,496
TCGA-LGG	Lower-grade glioma	Multi-class	Classic-like / Codel / G-CIMP-high / G-CIMP-low / Mesenchymal-like / PA-like	mRNA	513 (23, 173, 234, 12, 45, 26)	14,340
				Methylation	513 (23, 173, 234, 12, 45, 26)	71,637
ADNI ^a	Neurodegeneration	Regression	Volumetric phenotypes ^b	Genotyping array	512 (23, 173, 233, 12, 45, 26)	138,628
				mRNA	449	49,386

^a Abbreviations: ADNI = Alzheimer’s Disease Neuroimaging Initiative.

^b Vol. phenotypes include: brainstem, fourth ventricle, hippocampus, accumbens area, amygdala, caudate, pallidum, putamen, thalamus.

^c Binary datasets: parentheses show (cases, controls). Multi-class datasets: parentheses list per-class sample counts in the order of the “Endpoint” column.