Recovering gene regulatory networks in single-cell multi-omics data with PRISM-GRN

Wenhao Zhang[1,2], Lan Cao[1,2], Xiaoxuan Gu[1], Yongyu Long[1,2], and Ying Wang[1,2,3,4,*]

[1] Department of Automation, Xiamen University, Xiamen, Fujian 361000, China

[2] National Institute for Data Science in Health and Medicine, Xiamen University, Xiamen, Fujian 361000, China

[3] State Key Laboratory of Mariculture Breeding, Xiamen University, Xiamen, Fujian 361000, China

[4] Xiamen Key Laboratory of Big Data Intelligent Analysis and Decision, Xiamen, Fujian 361005, China

* Correspondence to wangying@xmu.edu.cn

Contents

Table. S1. Details of the Baseline Methods.

	Omics Data	Supervising Category	Causal Detection	Prior GRN Integration
Grace	scRNA	Supervised	Directed	Integrated
GeneLink				
GENIE3		Unsupervised	Undirected	None
GRNBoost				
scAI	scRNA + scATAC			
CellOracle				
SupirFactor				Integrated

Table. S2. The effect of the hyperparameters on the performance of PRISM-GRN.

		Accuracy	Precision	Recall	F1	AUROC	AUPRC
learning rate	0.003	0.9907	0.5549	0.2384	0.3335	0.9891	0.4331
	0.001	0.9906	0.5453	0.2198	0.3133	0.9895	0.4461
	0.0005	0.9907	0.5499	0.2304	0.3248	0.9893	0.4443
	0.0003	0.9907	0.5435	0.257	0.349	0.9893	0.4423
	0.0001	0.9907	0.5575	0.2317	0.3274	0.989	0.4407
weight decay	0.01	0.9906	0.5377	0.2649	0.355	0.9892	0.4409
	0.001	0.9907	0.5435	0.257	0.349	0.9893	0.4423
	0.0001	0.9907	0.5414	0.2649	0.3558	0.9893	0.4417
dropout	**0.1**	0.9907	0.558	0.2171	0.3126	0.9893	0.4365
	0.2	0.9906	0.5336	0.2636	0.3529	0.9894	0.4389
	0.3	0.9907	0.5473	0.2729	0.3642	0.9893	0.4419
hidden layer size	**256**	0.9907	0.5473	0.2729	0.3642	0.9893	0.4419
	128	0.9907	0.5634	0.2005	0.2958	0.9894	0.441
	64	0.9909	0.5745	0.2278	0.3262	0.9893	0.4454
latent dimension size	256	0.9905	0.5421	0.158	0.2447	0.9893	0.4319
	128	0.9908	0.5729	0.2218	0.3198	0.9893	0.4501
	64	0.9909	0.5745	0.2278	0.3262	0.9893	0.4554
alpha	0.15	0.9909	0.5691	0.2543	0.3515	0.9892	0.4426
	0.35	0.9909	0.5745	0.2278	0.3262	0.9893	0.4454
	0.55	0.9908	0.5687	0.2145	0.3115	0.9893	0.445

Note: The selected default hyperparameters are highlighted with bold style.

Table. S3. Detailed Performance Metrics Comparison

	Accuracy	Precision	Recall	F1	AUROC	AUPRC
			A549			
PRISM-GRN	**0.9907**	**0.5481**	0.2537	**0.3468**	**0.9893**	**0.4485**
celloracle	0.9671	0.1367	0.4482	0.2095	0.7105	0.0716
scAI	0.9376	0.0028	0.0153	0.0047	0.4809	0.0096
SupirFactor	0.9896	0.0179	0.0013	0.0025	0.5533	0.0172
Grace	0.9812	0.2567	0.4934	0.3377	0.9770	0.2160
GeneLink	0.9899	0.4536	0.1720	0.2494	0.9833	0.3358
GENIE3	0.3933	0.0146	**0.9256**	0.0288	0.7411	0.0215
GRNBoost	0.9132	0.0227	0.1886	0.0405	0.5540	0.0117
			3T3			
PRISM-GRN	**0.9955**	**0.7333**	0.1199	0.2061	**0.9963**	**0.5321**
celloracle	0.9951	0.0000	0.0000	0.0000	0.5000	0.0049
scAI	0.9481	0.0025	0.0245	0.0046	0.4884	0.0048
SupirFactor	0.9946	0.0000	0.0000	0.0000	0.5543	0.0121
Grace	0.9906	0.3365	**0.9591**	**0.4982**	0.9955	0.3929
GeneLink	0.9948	0.4350	0.2371	0.3069	0.9783	0.2892
GENIE3	0.2932	0.0063	0.9292	0.0126	0.6670	0.0071
GRNBoost	0.8478	0.0077	0.2371	0.0149	0.5428	0.0054
			GM12878			
PRISM-GRN	**0.9814**	**0.6306**	0.1534	0.2468	**0.9804**	**0.4652**
celloracle	0.9783	0.2080	0.0320	0.0555	0.5148	0.0268
scAI	0.8841	0.0109	0.0538	0.0182	0.4774	0.0194
SupirFactor	0.9801	0.0000	0.0000	0.0000	0.6125	0.0667
Grace	0.9548	0.2563	**0.6671**	**0.3703**	0.9692	0.2901
GeneLink	0.9734	0.3685	0.4691	0.4128	0.9723	0.3714
GENIE3	0.4960	0.0353	0.9227	0.0679	0.8094	0.0661
GRNBoost	0.9114	0.0653	0.2594	0.1044	0.5930	0.0365
			K562			
PRISM-GRN	**0.9751**	**0.6228**	0.2053	0.3088	**0.9688**	**0.4577**
celloracle	0.9720	0.1596	0.0084	0.0160	0.5036	0.0283
scAI	0.9039	0.0106	0.0277	0.0154	0.4780	0.0266
SupirFactor	0.9728	0.3762	0.0076	0.0150	0.5611	0.0578
Grace	0.8855	0.1524	**0.7082**	0.2509	0.9186	0.1530
GeneLink	0.9567	0.3082	0.4807	**0.3756**	0.9466	0.2945
GENIE3	0.5385	0.0460	0.8129	0.0871	0.6955	0.0444
GRNBoost	0.9208	0.0439	0.0926	0.0595	0.5182	0.0284

Note: The best-performing results are clearly highlighted with bold style.

Table. S4. The weighted performance of PRISM-GRN and the baseline methods

	Accuracy	Precision	Recall	F1	AUROC	AUPRC
	A549					
PRISM-GRN	**0.985**	**0.983**	**0.985**	**0.9836**	**0.9868**	**0.9923**
Grace	0.8491	0.9825	0.8491	0.9028	0.9435	0.9824
GeneLink	0.8771	0.983	0.8771	0.9197	0.9527	0.9837
	3T3					
PRISM-GRN	**0.9906**	**0.9914**	**0.9906**	**0.9909**	**0.9878**	**0.9947**
Grace	0.983	0.9916	0.983	0.9864	0.9843	0.9933
GeneLink	0.9859	0.987	0.9859	0.9857	0.9865	0.9936
	GM12878					
PRISM-GRN	**0.9705**	**0.968**	**0.9705**	**0.969**	**0.9831**	**0.9849**
Grace	0.8932	0.9673	0.8932	0.9209	0.9597	0.9715
GeneLink	0.8477	0.968	0.8477	0.8918	0.954	0.9693
	K562					
PRISM-GRN	**0.9599**	**0.9564**	**0.9599**	**0.9577**	**0.9743**	**0.9796**
Grace	0.946	0.8949	0.946	0.9197	0.5069	0.8971
GeneLink	0.7479	0.9544	0.7479	0.8178	0.9358	0.9601

Note: The best-performing results are clearly highlighted with bold style.

Table. S5. The causality results evaluation of PRISM-GRN with SCENIC+

Dataset	Method	AUROC	AUPRC	Precision	Recall	F1
3T3	GRNBoost2	0.5128	0.0049	0.0063	0.1163	0.0120
	GENIE3	0.6831	0.0074	0.0064	**0.9114**	0.0128
	SCENIC+_GRNBoost2	0.5065	0.0119	0.1695	0.0136	0.0252
	SCENIC+_GENIE3	0.5062	0.011	0.101	0.0136	0.024
	PRISM-GRN	**0.9944**	**0.4945**	**0.5232**	0.6733	**0.5888**
A549	GRNBoost2	0.5178	0.0103	0.0152	0.0997	0.0264
	GENIE3	0.7335	0.0197	0.0144	**0.9096**	0.0284
	SCENIC+_GRNBoost2	0.5005	0.0197	0.3	0.001	0.002
	SCENIC+_GENIE3	0.517	0.0289	0.2842	0.0358	0.0637
	PRISM-GRN	**0.9938**	**0.6411**	**0.6948**	0.4598	**0.5534**
GM12878	GRNBoost2	0.5418	0.024	0.0538	0.1316	0.0764
	GENIE3	0.7938	0.0586	0.0369	**0.8929**	0.0709
	SCENIC+_GRNBoost3	0.504	0.0422	0.3318	0.0087	0.0169
	SCENIC+_GENIE3	0.5163	0.0506	0.3502	0.0352	0.064
	PRISM-GRN	**0.9903**	**0.6594**	**0.6684**	0.5286	**0.5903**
K562	GRNBoost2	0.5054	0.0279	0.0353	0.0492	0.0411
	GENIE3	0.6801	0.0433	0.0464	**0.772**	0.0875
	SCENIC+_GRNBoost2	0.5018	0.0554	0.4317	0.0038	0.0076
	SCENIC+_GENIE3	0.5004	0.0542	0.2763	0.001	0.002
	PRISM-GRN	**0.9848**	**0.6542**	**0.6617**	**0.5**146	**0.579**

Table. S6. The cross validation with CRISPR datasets.

Gene1	Gene2	Predicted by	Correlation	Entrez ID
		A549 (8/12)		
PHF8	*NOP10*	-	-0.122294	55505
CREB1	*SLC7A11*	1	0.15292	23657
SMARCA4	*JUN*	1	-0.167887	3725
CTCF	*NOP10*	-	0.118317	55505
CRTC2	*TOMM20L*	1	0.135353	387990
RAD21	*NPY1R*	1	0.2068	4886
BRD2	*NCL*	1	0.140719	4691
MAFK	*MXRA8*	-	0.179093	54587
EP300	*NAMPT*	-	0.28904	10135
SP1	*AHCY*	1	0.209686	10567
CTCF	*DEDD2*	1	0.222152	191
CRTC2	*COX7A2*	1	-0.15759	162989
		GM12878 (20/20)		
BCLAF1	*ZNF211*	1	-0.1866	10520
SP1	*BRD4*	1	0.216028	23476
SPI1	*MEF2C*	1	0.242078	4208
BATF	*LIPT1*	1	0.170407	51601
CBFB	*PTK2*	1	-0.232641	5747
YY1	*IQSEC3*	1	0.142477	440073
MLLT1	*EED*	1	0.213839	8726
E2F4	*WDR74*	1	0.197566	54663
MTA2	*CTBP1*	1	0.1753	1487
SP1	*LIPT1*	1	0.253064	51601
EP300	*EED*	1	0.337544	8726
NFKB1	*LIPT1*	1	0.233006	51601
IRF4	*LIPT1*	1	0.266799	51601
IKZF1	*ILK*	1	-0.294995	3611
MED1	*PTK2*	1	-0.160038	5747
SMC1A	*MST1*	1	-0.14236	4485
IKZF1	*PAX5*	1	0.271638	5079
CREB1	*MST1*	1	-0.134592	4485
POU2F2	*ETS1*	1	0.166122	2113
RELB	*CTBP1*	1	0.186326	1487
		K562 (36/38)		
CCNT2	*BECN1*	1	0.16882	8678
RUVBL2	*CDC5L*	1	0.168483	988
RBBP5	*FARS2*	1	0.220759	10667
ARID1B	*ERI2*	-	-0.154383	112479
ERG	*NUP153*	1	0.145037	9972
SMARCA4	*SSBP1*	1	0.185447	6742
MAX	*ZNF830*	1	-0.16264	91603
SP1	*TFRC*	1	0.231351	7037
ARID4B	*UBR4*	1	0.164488	23352
L3MBTL2	*MGA*	1	0.159437	23269
THAP1	*DHPS*	1	0.142219	1725

L3MBTL2	*MGA*	-	0.159437	23269
MAZ	*ZFYVE19*	1	0.195723	84936
XRCC5	*ZRANB2*	1	0.172594	9406
MAX	*RBM14*	1	0.162676	10432
CSDE1	*HNRNPH1*	1	-0.165508	3187
RBM22	*DDX46*	1	0.274178	9879
RBFOX2	*PTBP1*	1	-0.121711	5725
ETV1	*MTF2*	1	0.193292	22823
SUZ12	*MEN1*	1	0.28362	4221
BRD3	*ZNF507*	1	0.186243	22847
KDM4B	*C12orf57*	1	0.209236	113246
SMAD4	*E2F6*	1	0.145843	1876
SUPT5H	*PRMT5*	1	-0.141908	10419
ZZZ3	*NSMCE2*	1	0.159281	286053
SUPT5H	*PHB2*	1	-0.19096	11331
IRF1	*BATF2*	1	0.149695	116071
ZNF395	*ZNF143*	1	-0.12098	7702
NR2C2	*TARDBP*	1	0.147273	23435
CXXC5	*PSMD9*	1	0.16243	5715
BCOR	*MEN1*	1	0.163396	4221
GTF2F1	*DDX46*	1	0.233081	9879
MNT	*FOSL1*	1	-0.2057	8061
KDM1A	*DNMT1*	1	0.161553	1786
SMARCA4	*BRD9*	1	0.241961	65980
TBP	*GTF3C1*	1	0.164937	2975
BCOR	*DOT1L*	1	0.173447	84444
CBX3	*AGK*	1	0.172277	55750
	3T3 (12/13)			
Ctcfl	*Thsd7a*	1	0.168663	221981
Brd4	*Acly*	1	0.148405	47
Ctcf	*Dedd2*	-	-0.15759	162989
Smc1a	*Zeb2*	1	0.162241	9839
Ctcf	*Mcrs1*	1	0.167373	10445
Smc1a	*Zeb2*	1	0.162241	9839
Tbp	*Gle1*	1	0.175389	2733
Setd1a	*Inca1*	1	-0.167818	388324
Hcfc1	*Mcrs1*	1	0.194753	10445
Mxi1	*Pam*	1	-0.159817	5066
Smc1a	*Zeb2*	1	0.162241	9839
Smc1a	*Zeb2*	1	0.162241	9839
Hcfc1	*Rpl12*	1	-0.198794	6136

Note: The number of recovered edges and raw counts are given for each dataset.

Table. S7. The cross validation with TRRUST datasets.

Gene1	Gene2	Causality Label by PRISM-GRN	TRRUST Evidence
CSRP1	*CEBPB*	2	14522018
SP1	*EZR*	0	19164283
IGFBP3	*SP1*	2	12200149
YY1	*PDIA6*	0	1330541
POLD2	*CREB1*	2	10978529
CTCF	*WT1*	1	24534946
JUND	*GCLC*	1	16054171
CEBPB	*CSRP1*	1	14522018
PTGES	*CREB1*	2	20688046
H19	*CTCF*	2	11431321;14654216;16391843;18458536; 19209620;19584898;20966046;24725430'

Table. S8. Top 25 Genes with the Highest Degree in the inferred GRN.

Genes	Degree in Predicted GRN	Degree in Prior GRN	Betweenness in Predicted GRN	Betweenness in Prior GRN
ATF4	1083	230	0.0401	0.0887
CEBPB	1059	230	0.0432	0.0921
MED12	1044	195	0.0363	0.0372
ETS1	1006	177	0.0320	0.0383
TRIM28	893	471	0.0141	0.2758
BCL6	850	179	0.0244	0.0328
CTCF	847	508	0.0325	0.3440
BRD4	834	540	0.0257	0.2927
NELFA	769	112	0.0211	0.0207
CXXC1	326	85	0.0225	0.0152
REL	191	82	0.0314	0.0057
FOSL2	155	115	0.0231	0.0594
JUNB	58	81	0.0073	0.0076
IRF4	15	54	0.0055	0.0045
YY1	15	60	0.0036	0.0058
TP73	12	9	0.0028	0.0039
HLA-DRA	11	0	0.0020	0
TTC28	11	1	0.0023	0
ETV3L	11	0	0.0028	0
FADS3	11	1	0.0020	0
PRKCQ	11	5	0.0015	0.0027
HMGN1	11	19	0.0015	0.0002
EZH2	10	12	0.0024	0.0167
MYOF	10	0	0.0021	0.0000
NFATC2	10	43	0.0017	0.0155

Table. S9. Top 10 Gene Regulatory Relationships with the Highest Probability.

Gene1	Gene2	Probability	Supports
CTCF	*BCL6*	0.972349	[1, 2]
TRIM28	*TP73*	0.96763	-
TRIM28	*REL*	0.959507	-
TRIM28	*HOPX*	0.958511	-
TRIM28	*SLC11A1*	0.955651	[3]
TRIM28	*CEBPB*	0.954477	[4]
TRIM28	*SMAD3*	0.952957	[5, 6]
TRIM28	*CYP1B1-AS1*	0.952921	-
TRIM28	*ADGRG6*	0.952217	[7]
BRD4	*STON2*	0.952198	-

Note: "-" indicates that there are currently no studies directly proving the relationship.

Table. S10. Top 10 Genes with the Highest Degree in the inferred GRNs.

PBMC_ID_IFNB1			PBMC_ID_CTRL		
Genes	Degree	BC	Genes	Degree	BC
MBOAT1	25.8719	0.0047	*CREBBP*	1398.837	0.2196
ADAP2	25.8652	0.0045	*BATF*	1381.977	0.2255
PSD3	24.893	0.0042	*ETS1*	1278.961	0.1659
CYP1B1-AS1	24.8664	0.0044	*TRIM28*	992.9819	0.0874
SLC44A1	23.8944	0.004	*MED12*	960.9083	0.0884
SMC2	23.8831	0.0042	*CTCF*	956.8952	0.0843
MAPKAPK2	23.8781	0.004	*BRD4*	918.0545	0.0711
TBC1D32	23.8509	0.0041	*ATF4*	627.038	0.036
ETV6	22.9016	0.0034	*BCL6*	340.2386	0.0094
JAK2	22.8998	0.0037	*CEBPB*	252.9377	0.0053

Note: Bold style indicates that the gene has evidence to be potentially relevant with IFN-β stimulation.

Table. S11. The computational efficacy comparison of PRISM-GRN and the baseline methods.

	Tasks	Time (s)	Peak CPU memory (Mb)	Peak GPU memory (Mb)
		A549		
PRISM-GRN	GRN Reconstruction	205.63	255.61	4767.23
	Causality Prediction	209.99	256.02	4792.92
GRACE	GRN Reconstruction	436.92	193.44	10696.31
	Causality Prediction	522.57	193.27	11624.14
GeneLink	GRN Reconstruction	539.88	119.27	348.49
	Causality Prediction	366.37	103.74	716.75
		3T3		
PRISM-GRN	GRN Reconstruction	138.71	136.38	2344.58
	Causality Prediction	137.29	136.04	2356.31
GRACE	GRN Reconstruction	217.28	93.16	5232.44
	Causality Prediction	250.78	92.79	5683.78
GeneLink	GRN Reconstruction	257.13	57.56	196.35
	Causality Prediction	187.56	49.58	436.13

Table. S12. The performance of PRISM-GRN on GRN Reconstruction with different sizes of genomic window in Signac.

Dataset	Window(bp)	Accuracy	Precision	Recall	F1	AUROC	AUPRC
A549	2000	0.9906	0.5375	0.2377	0.3297	0.9892	0.4458
	1500	0.9906	0.5378	0.241	0.3329	0.9892	0.4458
	1000	0.9906	0.5353	0.2517	0.3424	0.9892	0.4456
GM12878	2000	0.9817	0.6343	0.1892	0.2915	0.981	0.4746
	1500	0.9817	0.6383	0.1871	0.2894	0.981	0.4745
	1000	0.9817	0.6293	0.1904	0.2924	0.981	0.4745
3T3	2000	0.9953	0.5379	0.1935	0.2846	0.9947	0.464
	1500	0.9953	0.5308	0.188	0.2777	0.9947	0.464
	1000	0.9953	0.5379	0.1935	0.2846	0.9947	0.464
K562	2000	0.9756	0.6477	0.2171	0.3252	0.9713	0.488
	1500	0.9756	0.6509	0.2158	0.3241	0.9713	0.4881
	1000	0.9756	0.6529	0.2118	0.3199	0.9713	0.4881

Table. S13. The performance of PRISM-GRN on causality prediction with different sizes of genomic window in Signac.

Dataset	Window(bp)	Accuracy	Precision	Recall	F1	AUROC	AUPRC
A549	2000	0.9852	0.786	0.6267	0.6889	0.9917	0.7467
	1500	0.9852	0.7864	0.6298	0.6913	0.9917	0.7466
	1000	0.9852	0.7927	0.6194	0.6855	0.9917	0.7468
GM12878	2000	0.9711	0.7596	0.6704	0.7095	0.9875	0.7709
	1500	0.9712	0.7742	0.6422	0.696	0.9875	0.7706
	1000	0.9712	0.7671	0.6558	0.7029	0.9875	0.7707
3T3	2000	0.9928	0.7844	0.6591	0.7109	0.9959	0.7553
	1500	0.9928	0.7835	0.6573	0.7094	0.9959	0.7553
	1000	0.9928	0.784	0.6583	0.7102	0.9959	0.7553
K562	2000	0.9611	0.7868	0.6381	0.6968	0.9803	0.7623
	1500	0.9612	0.7748	0.6614	0.7091	0.9803	0.7628
	1000	0.9613	0.7807	0.6518	0.7046	0.9803	0.7626

Figure S1

Figure S1. The performance comparison of PRISM-GRN with the baseline methods on the benchmark datasets with different peak-to-TSS distance thresholds (1500, 1000, 500) in ground truth GRN reconstruction. The AUROC and AUPRC metrics are used to evaluate the performance, and the results demonstrate that PRISM-GRN outperforms the baseline methods across all datasets.

Figure S2

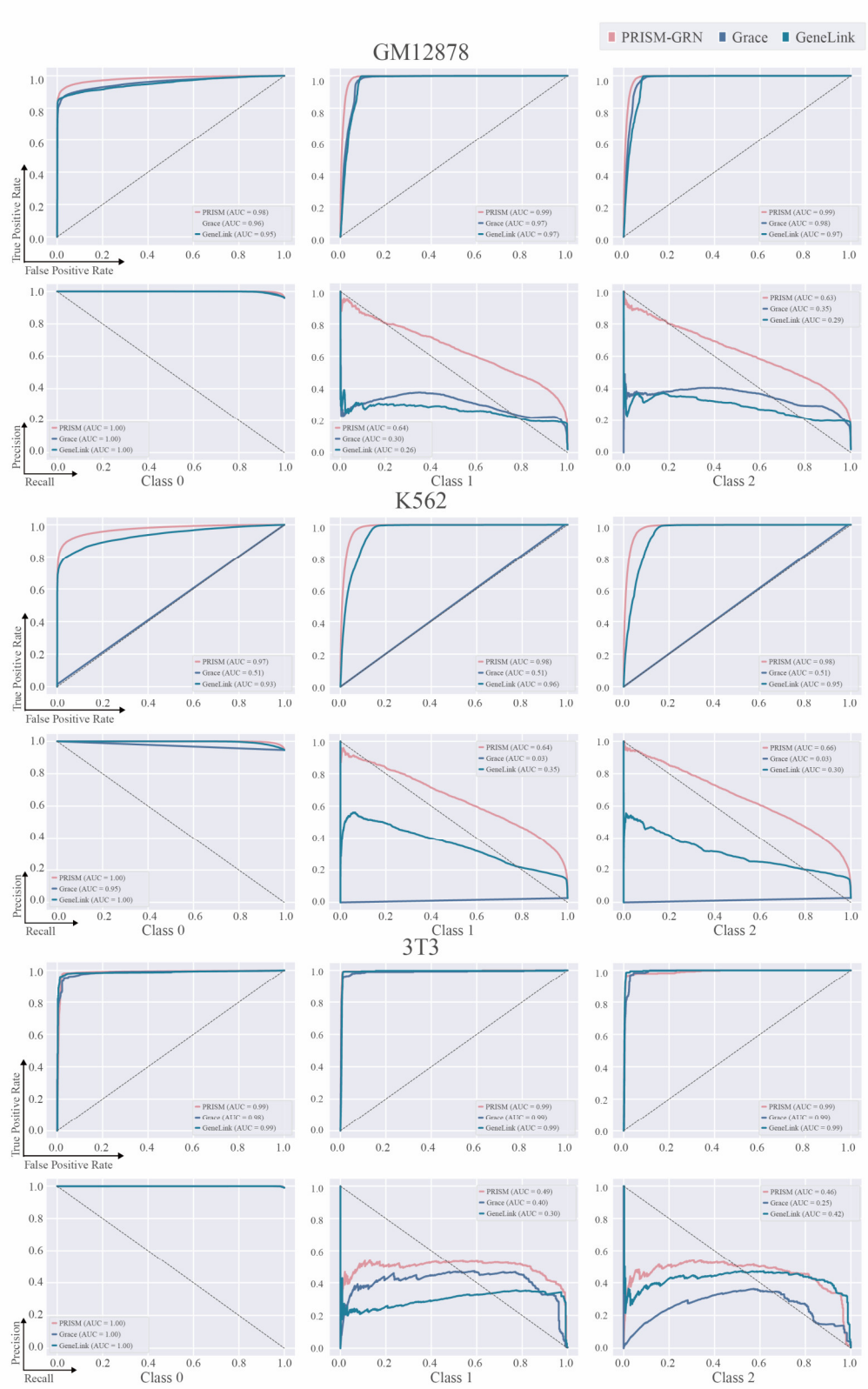

Figure S2. Receiver Operating Characteristic (ROC) curves and Precision-Recall (PRC) Curves for each class in causality prediction on the rest three benchmark datasets, including GM12878, K562, and 3T3 dataset, illustrating the detailed performance of PRISM-GRN and the baselines in causality prediction.

Figure S3

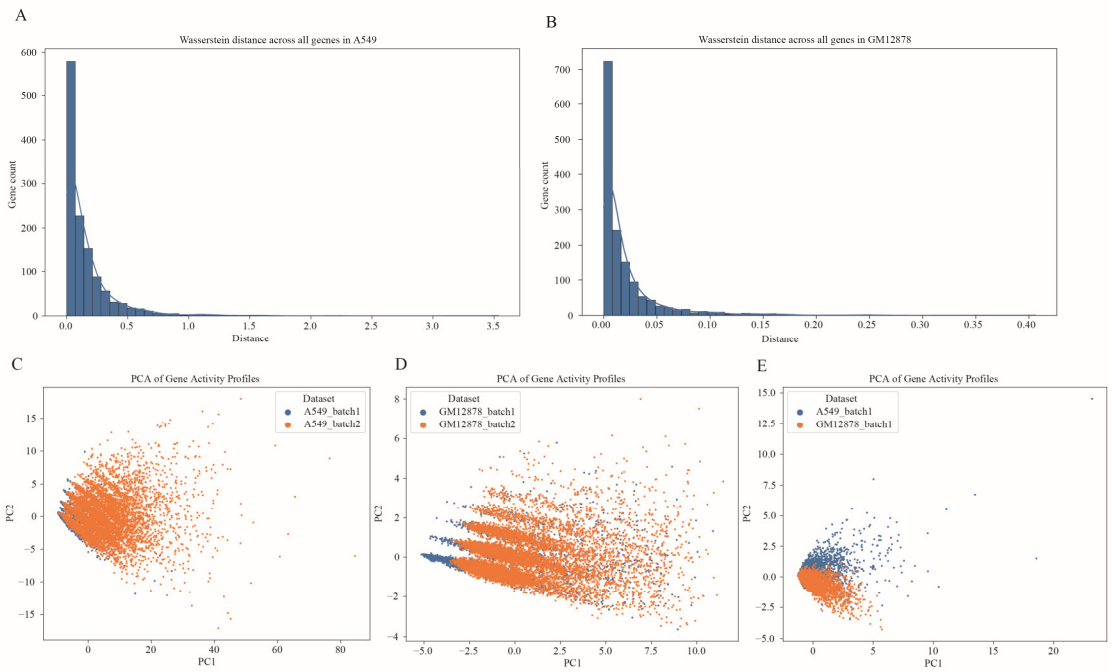

Figure S3. A-B. The distribution of the Wasserstein distance across all genes between different batches in the A549 and GM12878 datasets, respectively. C-E. The PCA plots of the concatenated data of the batches from different datasets.

Figure S4

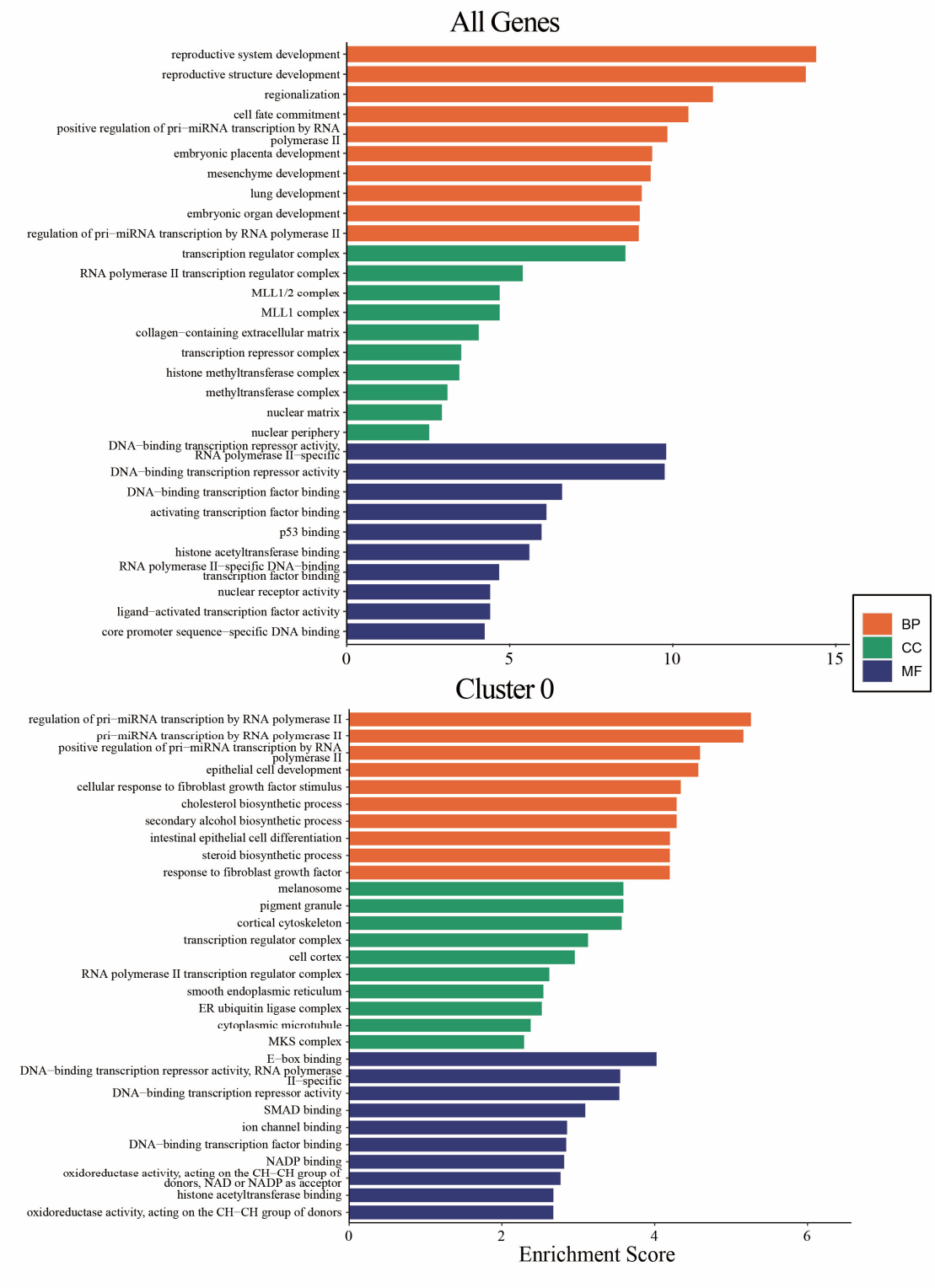

Figure S4. GO enrichment analysis results based on all genes and the genes in Cluster 0 of the A549 dataset. The results from the specific gene cluster reveal additional GO terms that are more relevant to A549 cells, thereby enhancing the understanding of gene functions.

Figure S5

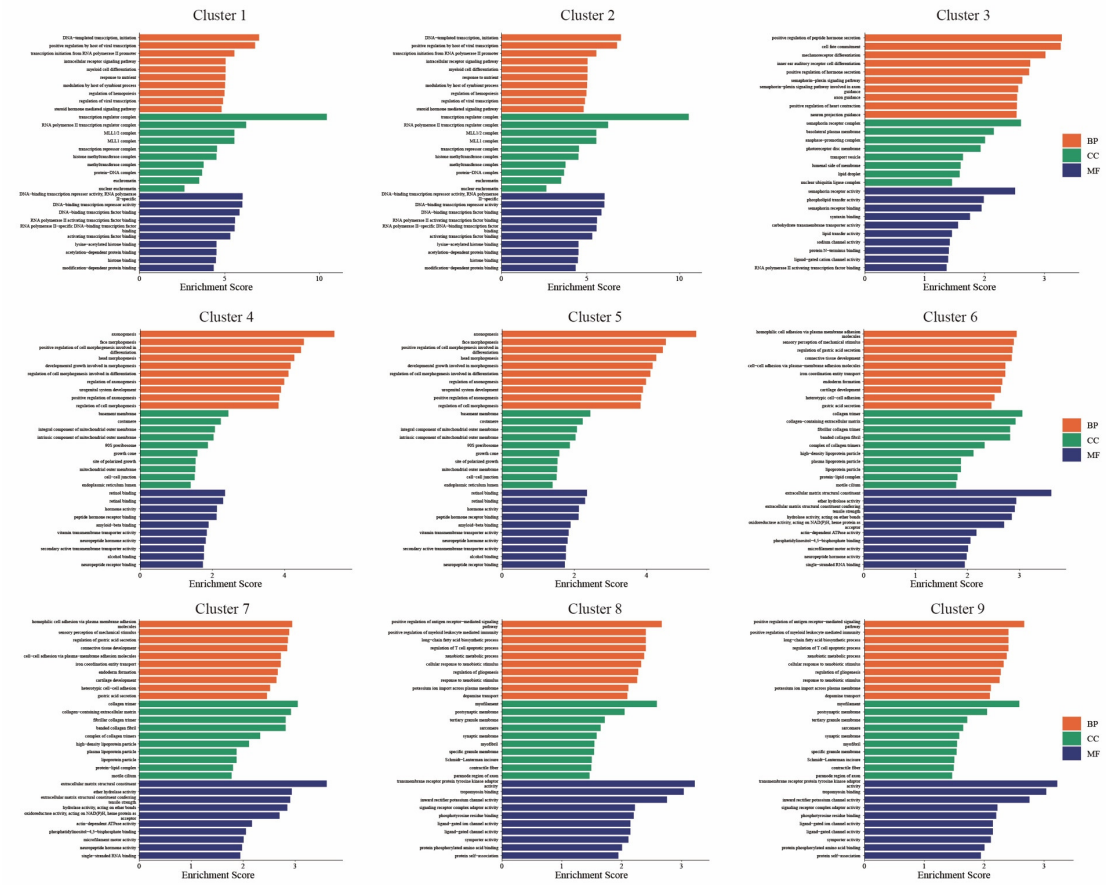

Figure S5. GO enrichment analysis results based on the genes from Cluster 1 to Cluster 9 of the A549 dataset.

Figure S6

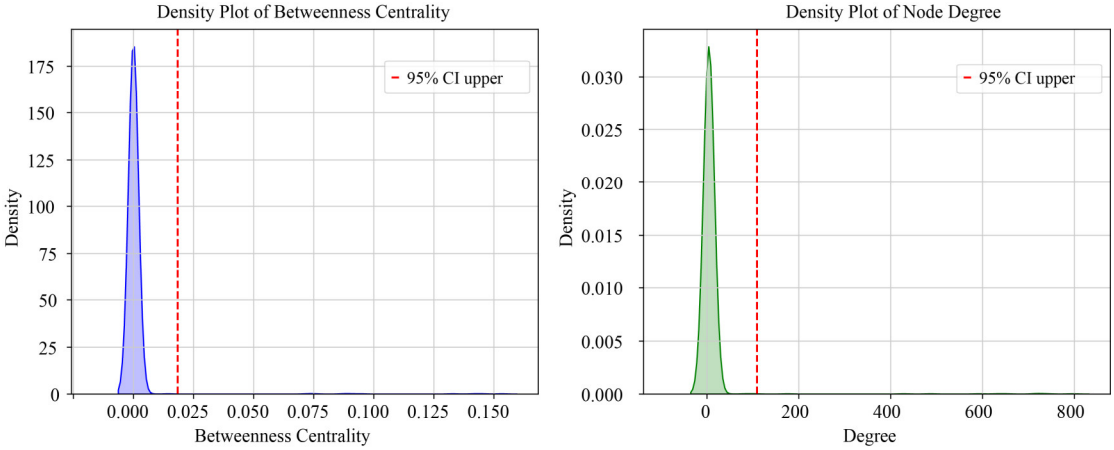

Figure S6. Density plots of betweenness centrality and degree for all genes calculated in the predicted GRN of CD4 Naïve cells. These plots illustrate the distribution of centrality and connectivity metrics for genes within the network.

Figure S7

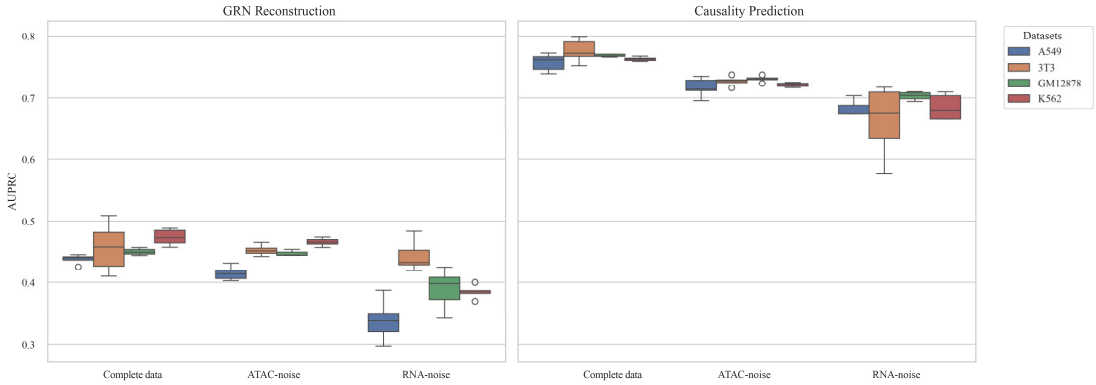

Figure S7. Ablation study comparing the impact of perturbations added to different modalities. In both GRN reconstruction and causality prediction tasks, introducing noise to either modality leads to a decline in PRISM-GRN's performance, highlighting the necessity of incorporating both modalities for optimal results. When noise was introduced into the scRNA-seq data, the average AUPRC across the four benchmark datasets reduced by about 14.54% in GRN reconstruction and 11.48% in causality prediction, for the scATAC-seq data, the perturbations led to a smaller average decline of approximately 2.88% in GRN reconstruction and 5.40% in causality prediction.

Figure S8

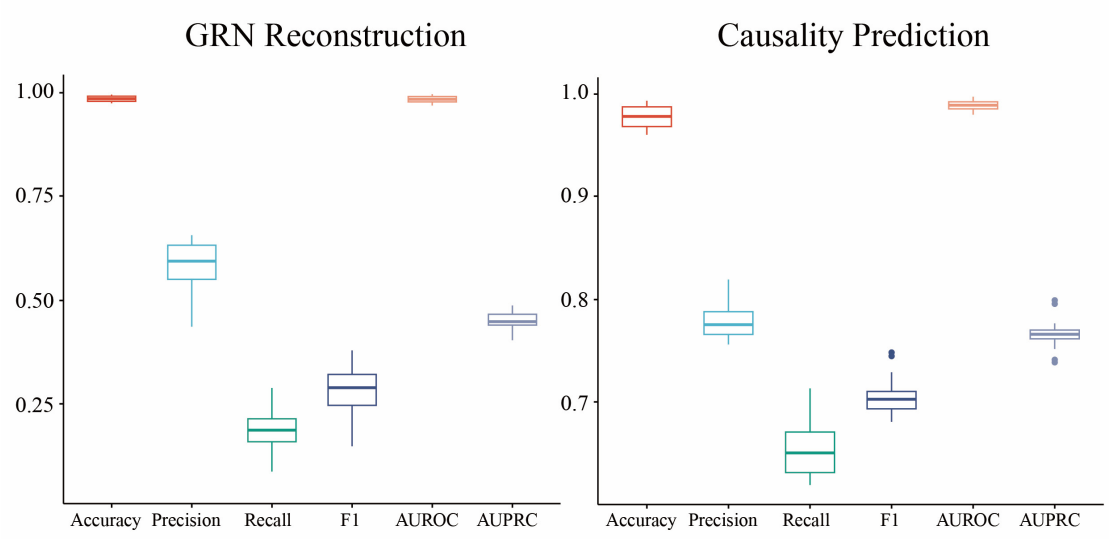

Figure S8. Boxplots of PRISM-GRN's performance under different random seeds. The results show that PRISM-GRN achieves consistently stable performance across both GRN reconstruction and causality prediction tasks, indicating robustness to training randomness.

Figure S9

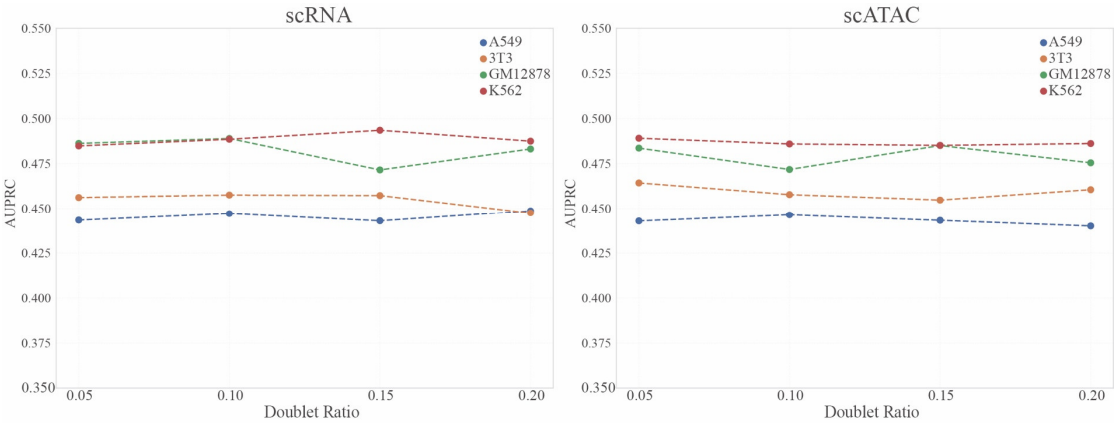

Figure S9. Line plots of changes in AUPRC under varying proportions of doublet noise introduced into the two modalities. PRISM-GRN demonstrates relatively stable performance across different noise levels.

Figure S10

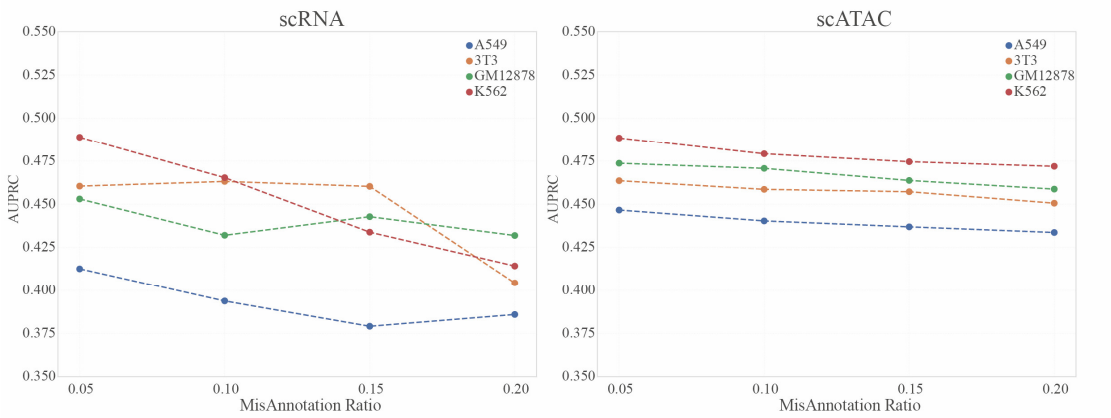

Figure S10. Line plots of changes in AUPRC under varying proportions of mis-annotation noise label introduced into the two modalities. PRISM-GRN demonstrates its dependency on the accuracy of the cell annotation of both modalities.

Figure S11

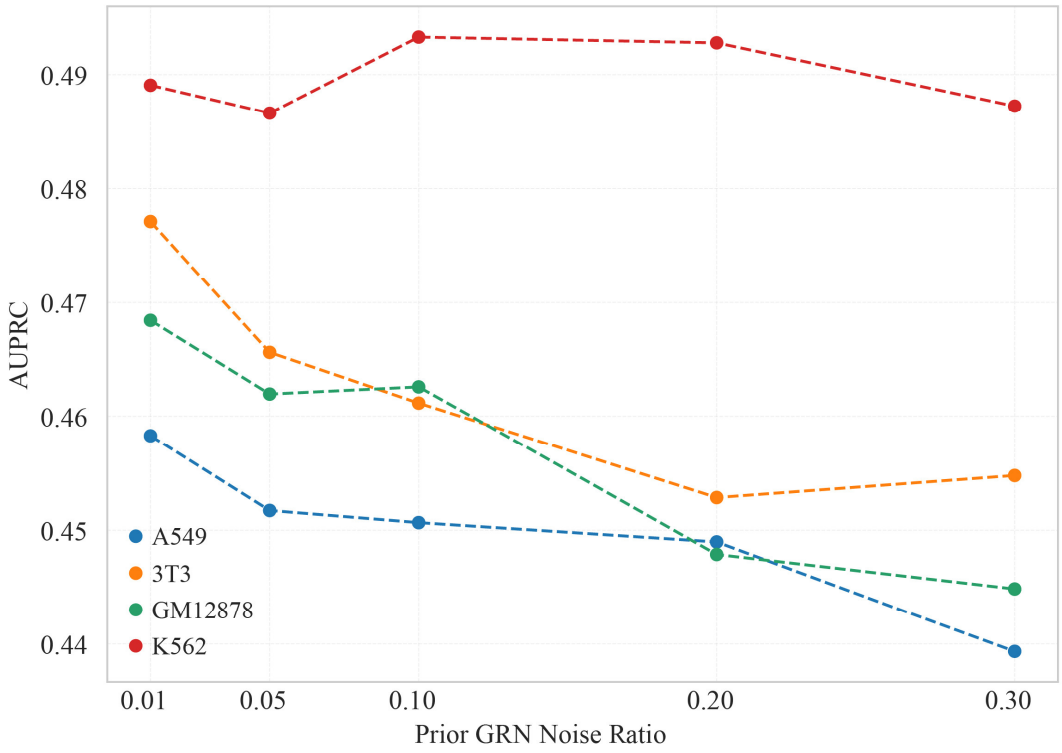

Figure S11. Line plots show the changes in AUPRC under varying proportions of incorrect prior regulatory relationships. As expected, AUPRC generally decreases as the proportion of incorrect priors increases. Introducing 30% noisy prior GRNs resulted in an average decrease of 2.76% in AUPRC across the four benchmark datasets, whereas performance remained relatively stable when the noise level was below 0.1. This indicates that PRISM exhibits a certain degree of robustness to prior knowledge noise.

Content S1. Data preprocessing pipeline

ChIP-seq Data

For ChIP-seq data derived from ChIP-Atlas, we used the bedtools [8] with the Linux command "bedtools window -a $A_Bed -b $B_Bed -w $width" to process the raw sequencing data. $A_Bed is a cell-type-specific TF ChIP-seq BED file, $B_Bed is a BED file containing the transcription start sites (TSS) of all genes, and the $width specifies the distance range from the transcription start site (TSS) of the gene within which the transcription factor (TF) binding is considered significant, which is set as 2,000. Accordingly, we can obtain the prior Gene Regulatory Network (GRN) $\boldsymbol{G}$, along with the TF set and the target gene set, denoted as $\mathrm{G(TF)}$ and $\mathrm{G(ChIP)}$.

scRNA-seq Data

For scRNA-seq data, following the standard pipeline using Scanpy [9], the cell-by-gene count matrix was performed with the quality control to filter out the low-quality cells and genes. First, the number of detected genes per cell is calculated, where cells with minimal gene expression are removed. Next, the total UMI count per cell is computed to assess sequencing depth, and cells with low UMI counts are excluded. Additionally, the percentage of mitochondrial gene expression is calculated, the cells whose mitochondrial gene expression exceeds 5% are then removed. Lastly, filtering criteria are applied to retain cells that meet specific thresholds for the minimum and maximum number of detected genes (150 and 5,000, respectively) and the minimum UMI count (100). The cells after quality control are denoted as $\mathrm{C(scRNA)}$.

Based on the quality-controlled cell-by-gene matrix, the top 3,000 highly variable genes (HVGs) are detected using function scanpy.pp.highly_variable_genes with default parameters. Specifically, it normalizes the dispersions based on the gene expression values for each gene, and sorts them to select the HVGs, denoted as $\mathrm{G(HVG)}$. We finally performed the log transformation on the filtered cell-by-gene count matrix to obtain the expression profile matrix.

scATAC-seq Data

For scATAC-seq data, the low-quality cells with fewer than 150 fragments or more than 5,000 fragments are filtered out, yielding the high-quality cells denoted as $\mathrm{C(scATAC)}$. Signac [10] was subsequently used to transform the cell-by-peak matrix to the cell-by-gene chromatin accessibility score matrix, with the CreateGeneActivityMatrix() function (default settings). Notably, the upstream and downstream parameters that define the genomic regions around each gene were also set as 2,000, consistent with the standard used in ChIP-seq data preprocessing.

Input Matrix Generation

For paired datasets, the final cells are the intersection of the quality-controlled cells from scRNA-seq data and scATAC-seq data, namely, $\mathrm{C(scRNA)} \cap \mathrm{C(scATAC)}$. For unpaired dataset, the final cells correspond to their quality control results. The final genes are the union of TFs derived from ChIP-seq data, and the intersection of the target genes and the HVGs, namely, $\mathrm{G(TF)} \cup (\mathrm{G(ChIP)} \cap \mathrm{G(HVG)})$, covering the known TFs and the significant genes among the cell-type-specific dataset. According to the final cells and final genes, the input gene expression matrix $\boldsymbol{E} \in \mathbb{R}^{M \times N}$, and the chromatin accessibility score matrix $\boldsymbol{R} \in \mathbb{R}^{M \times N}$, where M is the number of final genes and N is the number of final cells, can be generated by trimming the normalized cell-by-gene expression matrix and the cell-by-gene chromatin accessibility score matrix.

Content S2. Detailed descriptions of The performance metrics

We firstly evaluate the performance of PRISM-GRN and the baseline methods based on Area Under the Receiver Operating Characteristic (AUROC) and Area Under Precision-Recall Curve (AUPRC).

Accuracy, Precision, Recall, F1 are also used to show more explicit comparisons in some evaluation experiments.

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+FP+TN+FN}} , \tag{1}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} , \tag{2}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}} , \tag{3}$$

$$\text{F1} = \frac{2*\text{precision}*\text{recall}}{\text{precision+recall}} , \tag{4}$$

where TP, FP, TN, FN represents the true predicted gene regulatory interactions, false predicted gene regulatory interactions, true predicted non-interacted genes, and false predicted non-interacted genes, respectively. Notably, the macro averaging was used in all evaluations.

The betweenness centrality used to identify regulators in GRNs can be defined as follows:

$$BC_i = \sum_{i \neq j \neq k} \frac{\sigma(j,k \mid i)}{\sigma(j,k)} , \tag{5}$$

where BC_i is the betweenness centrality of gene i, $\sigma(j,k)$ is the total number of shortest paths between gene j and gene k, and $\sigma(j,k \mid i)$ is the number of regulatory paths passing through node gene i.

Content S3. The batch effect analysis when using unpaired scATAC-seq data

In fact, PRISM-GRN does not directly utilize raw peak-resolution scATAC-seq data. Instead, we preprocess the data using the Signac toolkit, which transforms the high-resolution peak matrix into gene-resolution gene accessibility scores. While this transformation may result in a certain degree of resolution loss of the fine-grained chromatin accessibility signal, it serves two important purposes: (1) it facilitates alignment with the scRNA-seq gene expression data, and (2) it mitigates certain batch-specific artifacts associated with peak calling and sequencing depth variability. By aggregating signal to the gene level, this preprocessing step acts as a form of normalization that smooths over some of the noise and variability across batches.

To further illustrate this, we provide examples from two datasets used in this study: A549 and GM12878. Specifically, the A549 dataset consists of two independently sequenced scATAC-seq datasets from separate experimental batches, while the GM12878 dataset includes two sequencing replicates derived from the same experimental protocol. For both datasets, we applied the Signac pipeline to convert raw peak-level scATAC-seq data into gene-level gene accessibility scores.

We used the Wasserstein distance to assess the similarity between the distribution of the transformed gene accessibility score across different batches. For each gene, we computed the Wasserstein distance between its gene accessibility score distributions in two batches. This yielded a gene-level measure of distributional divergence. Specifically, let $x = [x_1, x_2, \dots, x_M]$ and $y = [y_1, y_2, \dots, y_M]$ denote the gene accessibility score vectors of a given gene across M cells in two batches. The Wasserstein distance between these two distributions is defined as:

$$W(x, y) = \frac{1}{M}\sum_{i=1}^{M}\left|x_{(i)} - y_{(i)}\right| , \tag{6}$$

where $x_{(i)}$ and $y_{(i)}$ are the i-th elements of the sorted vectors x and y, respectively. The lower the value of $W(x, y)$ is, the greater the similarity between the distributions of x and y.

We then aggregated these values across all genes and visualized the global distribution of Wasserstein distances using a density plot. As shown in Supplemental Figures S6, the Wasserstein distances between gene accessibility scores across different batches for both A549 and GM12878 remain at relatively low levels, indicating high distributional similarity of gene accessibility scores across batches within the same dataset. Notably, the GM12878 dataset, which consists of two replicates from the same sequencing experiment, exhibits even lower Wasserstein distances compared to A549, whose two batches originate from independent sequencing experiments. This further supports the validity of using the Wasserstein distance as a quantitative measure of similarity across batches.

Furthermore, we extracted the gene score matrices from two batches of the A549 dataset and two batches of the GM12878 dataset, and also concatenated one batch from A549 with one batch from GM12878. PCA was performed separately on these three data groups. As shown in the figure, the two batches from the same dataset exhibit highly consistent distributions in the PCA space, indicating minimal batch effects. In contrast, the concatenated data from different cell lines (A549 and GM12878) show clearly distinct distributions, reflecting substantial differences in their gene accessibility score profiles.

Although PRISM-GRN itself does not incorporate explicit batch correction modules, the transformation by Signac and PRISM-GRN's Bayesian variational framework and integration at the gene level allow it to model shared regulatory patterns in a way that is relatively robust to residual

batch effects.

Content S4. The ablation experiments

PRISM-GRN is built upon a probabilistic framework that explicitly models the generation of gene expression as a downstream consequence of the expression levels of transcription factors (TFs) targeting it and the accessibility of its chromatin through the cell-type-specific GRN. As such, both modalities, expression (from scRNA-seq) and *cis*-regulatory accessibility (from scATAC-seq), along with the prior GRNs, play essential roles in the inference process. Removing one of the modalities would undermine the biological interpretability of the inferred regulatory relationships, as the model would no longer reflect the hierarchical nature of transcriptional regulation.

To empirically assess the impact of each modality, we conducted ablation experiments using structured noise injection into either the scRNA-seq or scATAC-seq data across several benchmark datasets. These experiments serve to approximate modality dropout while preserving the model's structure. Specifically, for each selected modality data, we added values drawn from a truncated normal distribution $\mathcal{N}^{+}(0, 0.1^2)$ to all features, ensuring the perturbation is strictly non-negative. The performance was evaluated on both tasks of GRN reconstruction and causality prediction across the four benchmark datasets by repeating each ablation setting with six independent random seeds to ensure statistical robustness.

As shown in Supplemental Figure S7, the values of AUPRC consistently decreased as noise intensity increased, confirming that perturbations in either modality degrade model performance. Notably, we observed that the degradation was more severe when noise was introduced into the scRNA-seq data, suggesting that expression information plays a more dominant role in PRISM-GRN's inference pipeline. In contrast, while noise in the scATAC-seq data also impaired performance, its effect was less pronounced.

This asymmetry likely stems from the preprocessing approach applied to the scATAC-seq data, where we employ Signac to convert high-resolution peak signals into gene-level chromatin accessibility scores. Although this facilitates alignment between chromatin features and gene expression, it may introduce a degree of information loss due to the smoothing and peak-to-gene linkage steps. Consequently, the contribution of the scATAC-seq modality, while biologically meaningful, its contribution may appear diminished relative to scRNA-derived features in PRISM-GRN's current implementation.

Content S5. The sensitivity analysis on GRN preprocessing variation

We used Signac to preprocess the scATAC-seq data. In this manuscript, no additional normalization was applied beyond the default procedures in Signac, which we used to convert peak-level signals into gene-level chromatin accessibility scores. This transformation enables alignment between modalities and helps mitigate batch effects to some extent; however, it leads to information loss due to resolution reduction and score aggregation.

Moreover, the peak-to-gene linkage in Signac requires setting a genomic window around the gene body to associate nearby peaks with each gene. In our default setting, we used a ±2000 bp window (2000 bp upstream and downstream of the gene body). To evaluate the robustness of our results to this parameter, we conducted additional experiments using alternative window sizes of ±1000 bp, ±1500 bp, and ±20000 bp. Specifically, we performed PRISM-GRN on the four benchmark datasets across the three alternative window sizes in both tasks of GRN reconstruction and causality prediction. As shown in Supplemental Table S12 and Table S13, while performance fluctuates slightly with different window sizes, the overall trends in GRN reconstruction and causality prediction remain consistent. This suggests that PRISM-GRN is relatively robust to the specific choice of peak-to-gene linkage parameters.

Content S6. The sensitivity analysis on cell annotation noise

We first evaluated the impact of doublets. Following the simulation strategy adopted by scIBD [11], we constructed datasets with varying doublet ratios ranging from 0.05 to 0.2 (with an interval of 0.05). Because we used cell-type-specific data as the benchmark, all simulated doublets were homotypic. For scRNA-seq data, each doublet was generated by randomly selecting two cells and averaging their gene expression profiles. In contrast, for scATAC-seq data, doublets were constructed by taking the union (element-wise maximum) of chromatin accessibility profiles from two randomly selected cells. As illustrated in Supplemental Figure S9, the performance of PRISM-GRN remains robust across different doublet ratios, demonstrating its resilience to homotypic doublet contamination.

Subsequently, to further assess the robustness of our method under imperfect cell annotations, we simulated mislabeled or noisy data by injecting perturbations into both scRNA-seq expression and scATAC-seq gene accessibility profiles. Instead of modifying the label space directly, we perturbed the input features to reflect biologically plausible annotation errors such as expression shifts or experimental noise. Similarly, we randomly selected a prosportion of cells (ranging from 0.05 to 0.2 with an interval of 0.05) and added noises. For each selected cell, we added values drawn from a truncated normal distribution $\mathcal{N}^{+}(0, 0.1^2)$ to all features, ensuring the perturbation is strictly non-negative.

As shown in Supplemental Figure S10, PRISM-GRN exhibits a gradual performance drop as noise increases, indicating its sensitivity to incorrect annotations in both modalities. Notably, in PRISM-GRN, the performance degradation is more pronounced when noise was injected into the scRNA-seq data compared to scATAC-seq, suggesting that accurate cell annotation for transcriptomic data plays a more critical role in model performance. In addition, although the use of Signac to convert peak-resolution scATAC-seq data into gene accessibility scores facilitates data alignment and mitigates batch effects, it may lead to a loss of regulatory information due to the transformation.

Content S7. The sensitivity analysis on prior GRN noise

To assess PRISM-GRN's robustness to imperfect prior knowledge, we conducted a sensitivity analysis by introducing increasing levels of noise into the prior GRNs. Specifically, to simulate the effect of noisy biological priors, we perturbed the training set by randomly introducing incorrect TF-target regulatory edges at five different noise levels: 1%, 5%, 10%, 20%, and 30% of the original number of positive edges.

As shown in Supplemental Figure S11, AUPRC exhibits an overall downward trend as the level of noise in the prior GRNs increases, reflecting PRISM-GRN's inherent dependence on high-quality prior networks. However, when the noise ratio remains below 10%, the performance of PRISM-GRN stays relatively stable. This indicates that the variational inference framework underlying PRISM-GRN is capable of effectively absorbing moderate levels of noise, thus demonstrating its robustness and capacity to denoise imperfect biological priors to some extent.

Reference

1. Lai AY, Fatemi M, Dhasarathy A, Malone C, Sobol SE, Geigerman C, Jaye DL, Mav D, Shah R, Li L: **DNA methylation prevents CTCF-mediated silencing of the oncogene BCL6 in B cell lymphomas.** *Journal of Experimental Medicine* 2010, **207:**1939–1950.
2. Batlle-López A, Cortiguera MG, Rosa-Garrido M, Blanco R, del Cerro E, Torrano V, Wagner SD, Delgado MD: **Novel CTCF binding at a site in exon1A of BCL6 is associated with active histone marks and a transcriptionally active locus.** *Oncogene* 2015, **34:**246–256.
3. Katiyar A, Sharma S, Singh TP, Kaur P: **Identification of shared molecular signatures indicate the susceptibility of endometriosis to multiple sclerosis.** *Frontiers in Genetics* 2018, **9:**42.
4. Shibata M, Blauvelt KE, Liem Jr KF, García-García MJ: **TRIM28 is required by the mouse KRAB domain protein ZFP568 to control convergent extension and morphogenesis of extra-embryonic tissues.** *Development* 2011, **138:**5333–5343.
5. Chen L, Munoz-Antonia T, Cress WD: **Trim28 contributes to EMT via regulation of E-cadherin and N-cadherin in lung cancer cell lines.** *PloS one* 2014, **9:**e101040.
6. Chikuma S, Suita N, Okazaki I-M, Shibayama S, Honjo T: **TRIM28 prevents autoinflammatory T cell development in vivo.** *Nature immunology* 2012, **13:**596–603.
7. Parker KA, Gooding AJ, Valadkhan S, Schiemann WP: **lncRNA BORG: TRIM28 complexes drive metastatic progression by inducing α6 integrin/CD49f expression in breast cancer stem cells.** *Molecular Cancer Research* 2021, **19:**2068–2080.
8. Quinlan AR, Hall IM: **BEDTools: a flexible suite of utilities for comparing genomic features.** *Bioinformatics* 2010, **26:**841–842.
9. Wolf FA, Angerer P, Theis FJ: **SCANPY: large-scale single-cell gene expression data analysis.** *Genome biology* 2018, **19:**1–5.
10. Stuart T, Srivastava A, Madad S, Lareau CA, Satija R: **Single-cell chromatin state analysis with Signac.** *Nature methods* 2021, **18:**1333–1341.
11. Zhang W, Jiang R, Chen S, Wang Y: **scIBD: a self-supervised iterative-optimizing model for boosting the detection of heterotypic doublets in single-cell chromatin accessibility data.** *Genome Biology* 2023, **24:**225.