

Additional file 1

Supplementary Figures S1-S10 and Supplementary Tables S1-S6

A transcriptomic readout of chromosomal instability reveals genome-stress states beyond proliferation in colorectal cancer

File format: PDF. Figures S1-S10 appear on pages 2-11, each preceded by its legend. Tables S1-S6 appear from page 12, each with its title above and description below.

Fig. S1. Model development and cross-validation.

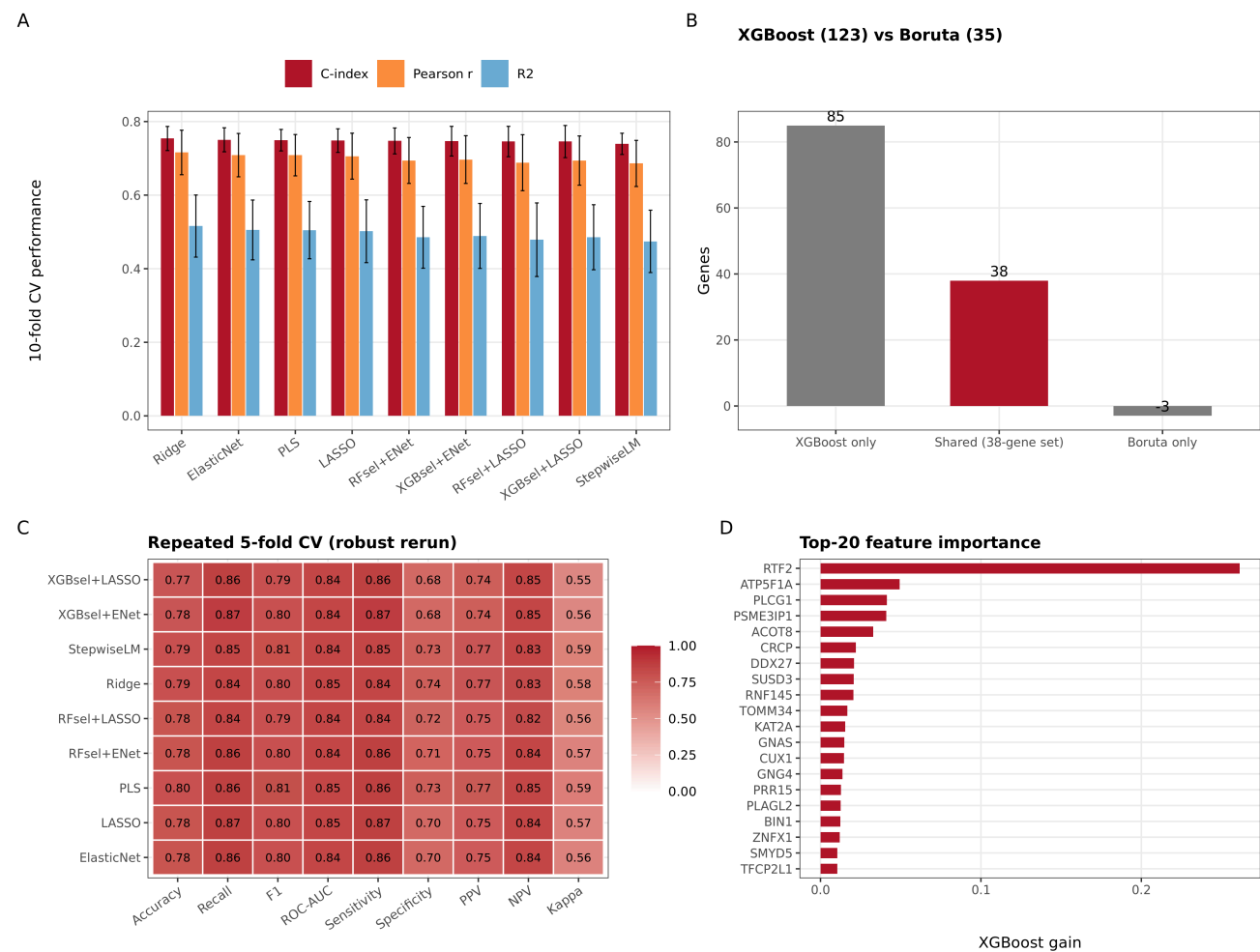


Fig. S2. Cross-cohort relationship between the mRNA-CIN score and comparator transcriptional signatures.

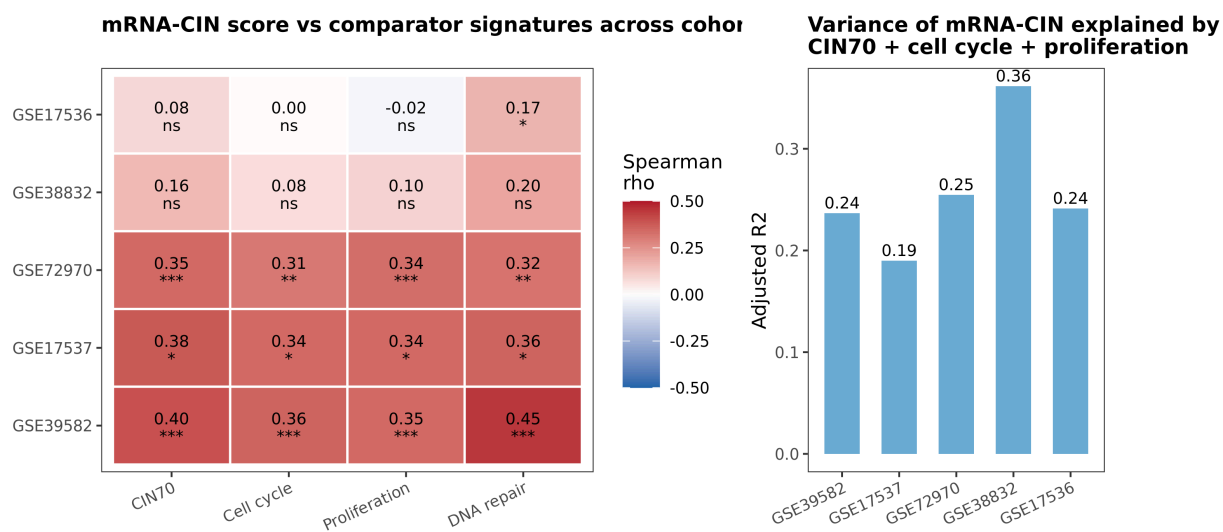


Fig. S3. Survival analysis and meta-analysis of the mRNA-CIN score.

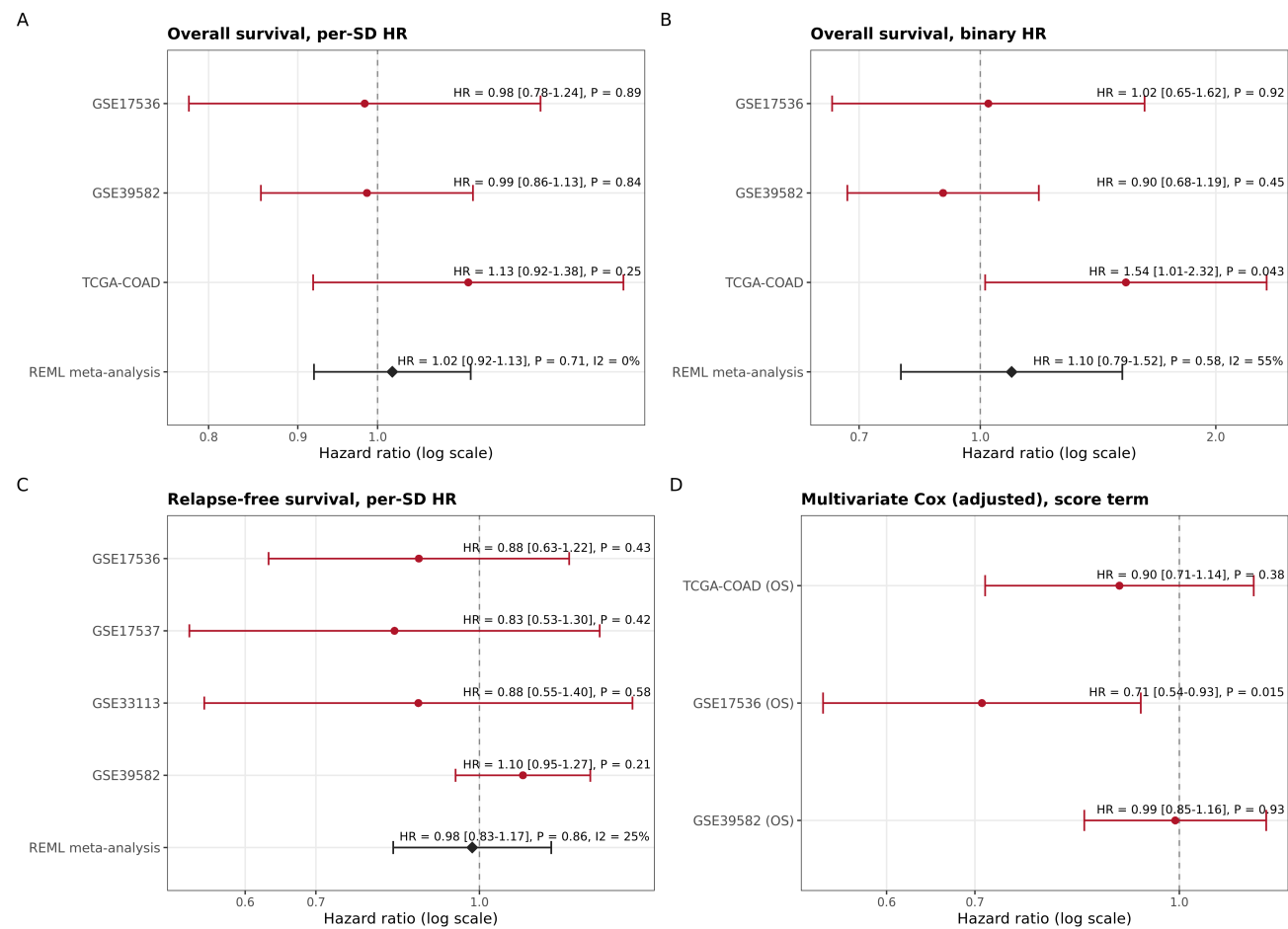


Fig. S4. Distribution of the mRNA-CIN score across CIT molecular subtypes in GSE39582 (tumor samples; Kruskal-Wallis test; points colored by MMR status).

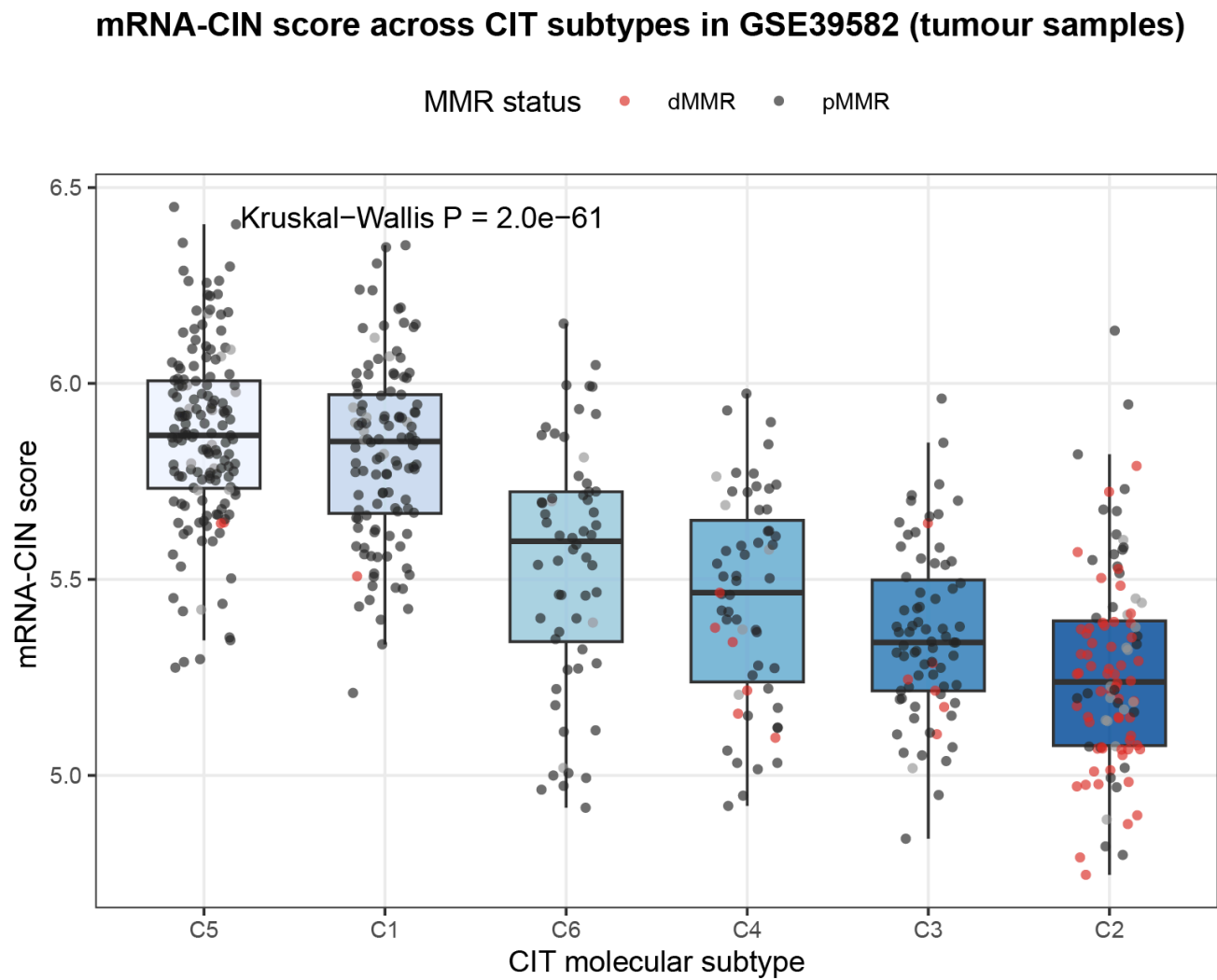


Fig. S5. Cross-cohort correlations between the mRNA-CIN score, CIN70 or proliferation scores and immune-related transcriptional activity (bubble size = |Spearman rho|; color = direction; asterisks = significance).

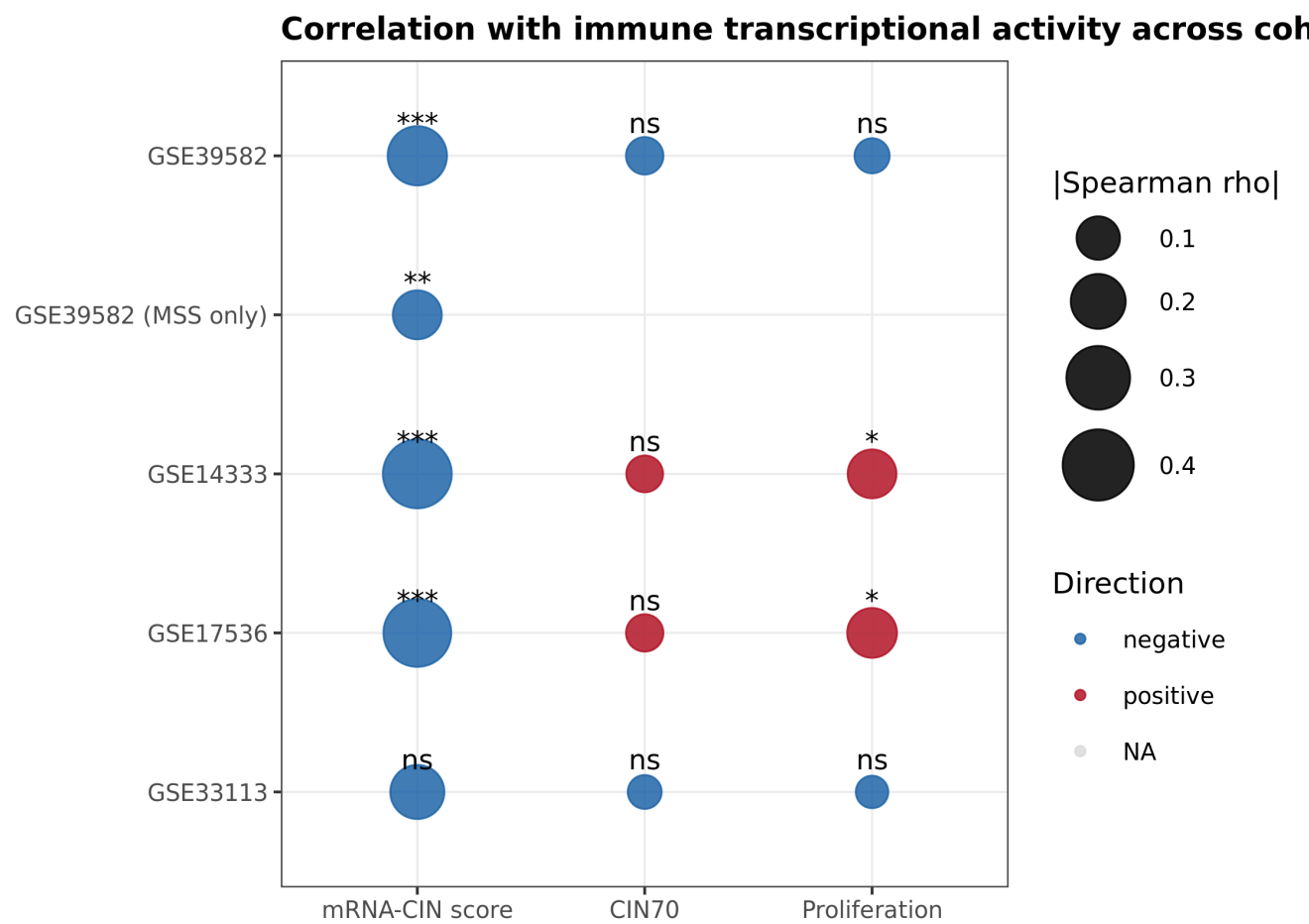


Fig. S6. Flow-cytometry summary of the Reversine perturbation.

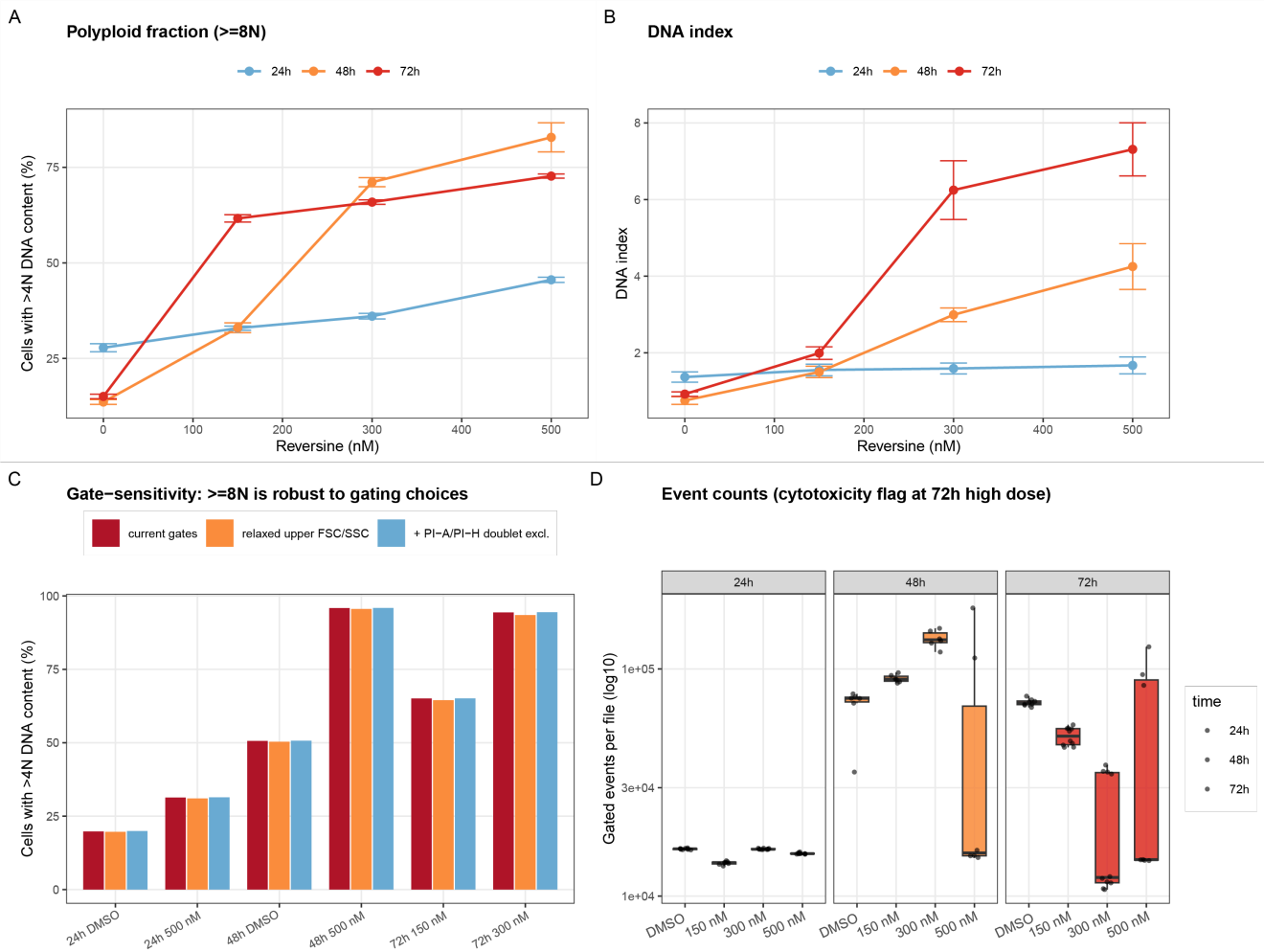


Fig. S7. Perturbation RNA-seq closed-loop analysis.

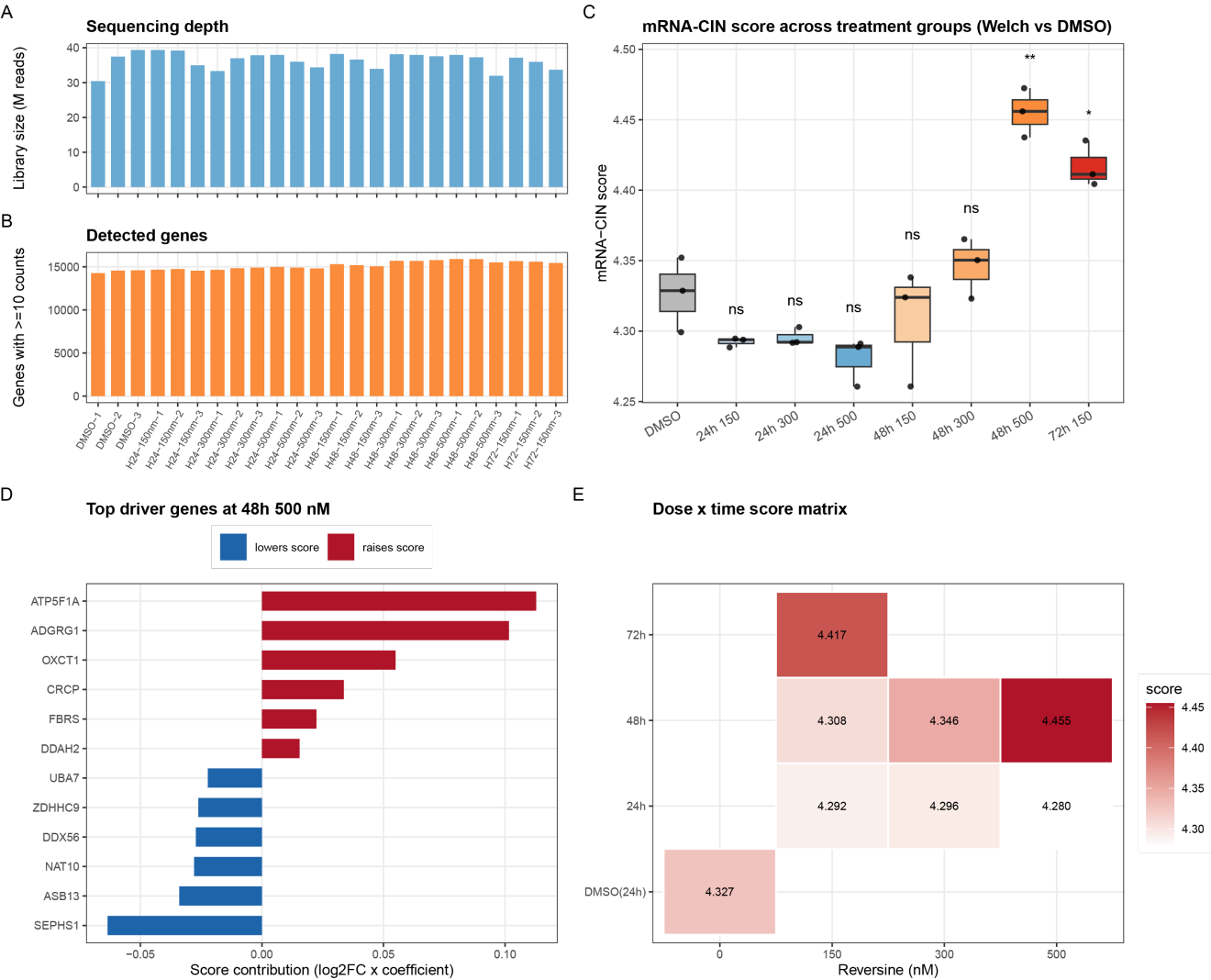


Fig. S8. DNA-content distributions of HCT116 cells following Reversine treatment. HCT116 cells were treated with DMSO or Reversine at 150, 300, or 500 nM for 24 or 48 h. Cellular DNA content was assessed by propidium iodide (PI) staining followed by flow cytometric analysis. Representative DNA-content histograms are shown for cells treated for 24 h (top row) and 48 h (bottom row). The x-axis represents PI fluorescence intensity (B572-A/PE-A), and the y-axis represents the relative number of cellular events.

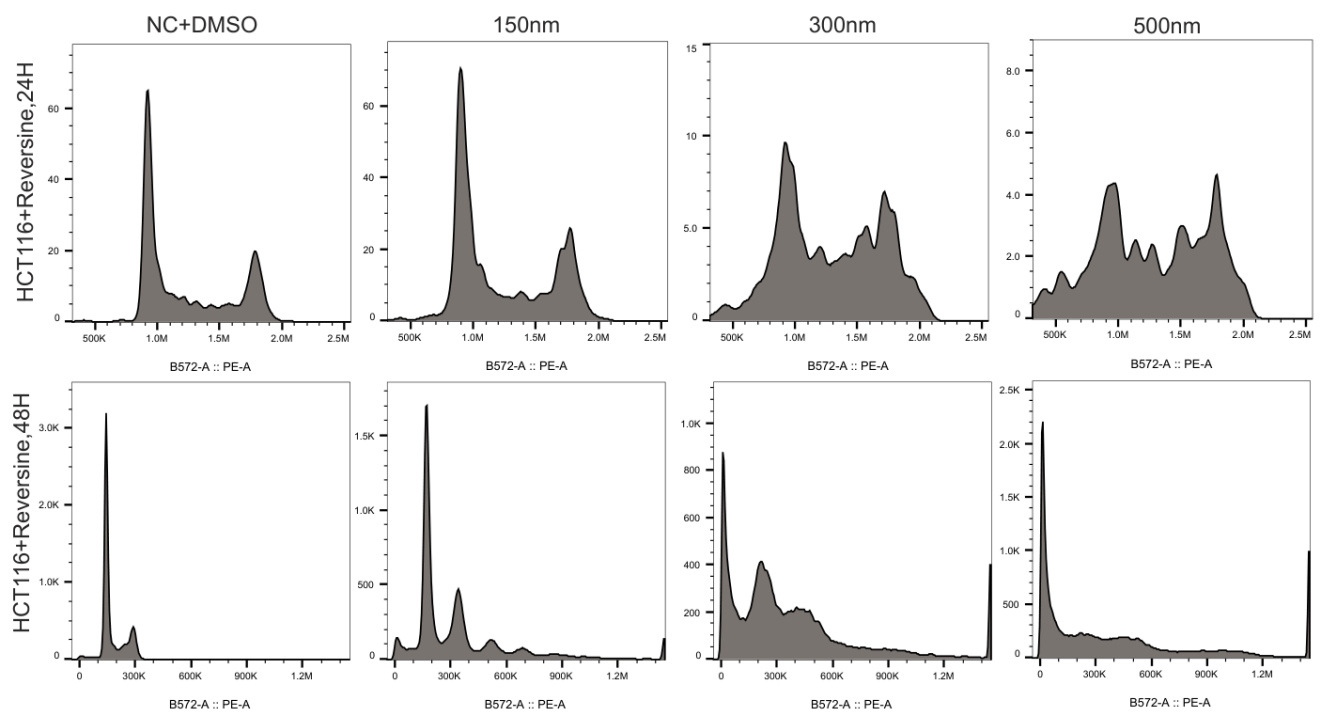


Fig. S9. Sensitivity analysis of reference-CIN discrimination in GSE39582 under the all-sample cohort definition (n = 481).

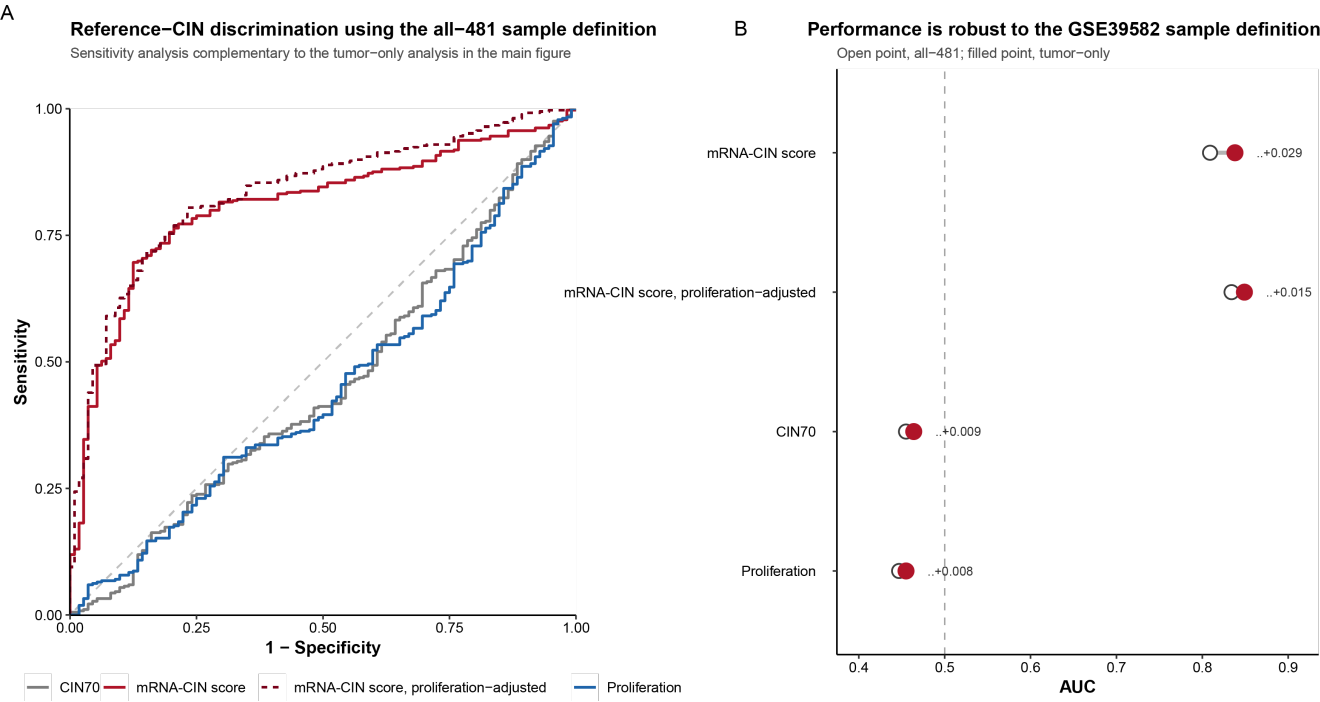


Fig. S10. Micronucleus formation after 24 h of Reversine treatment. (A) Representative immunofluorescence images of DAPI-stained HCT116 cells treated with DMSO or Reversine (150, 300 or 500 nM) for 24 h. (B) Quantification of micronucleated cells (%) across the 24-h dose series.

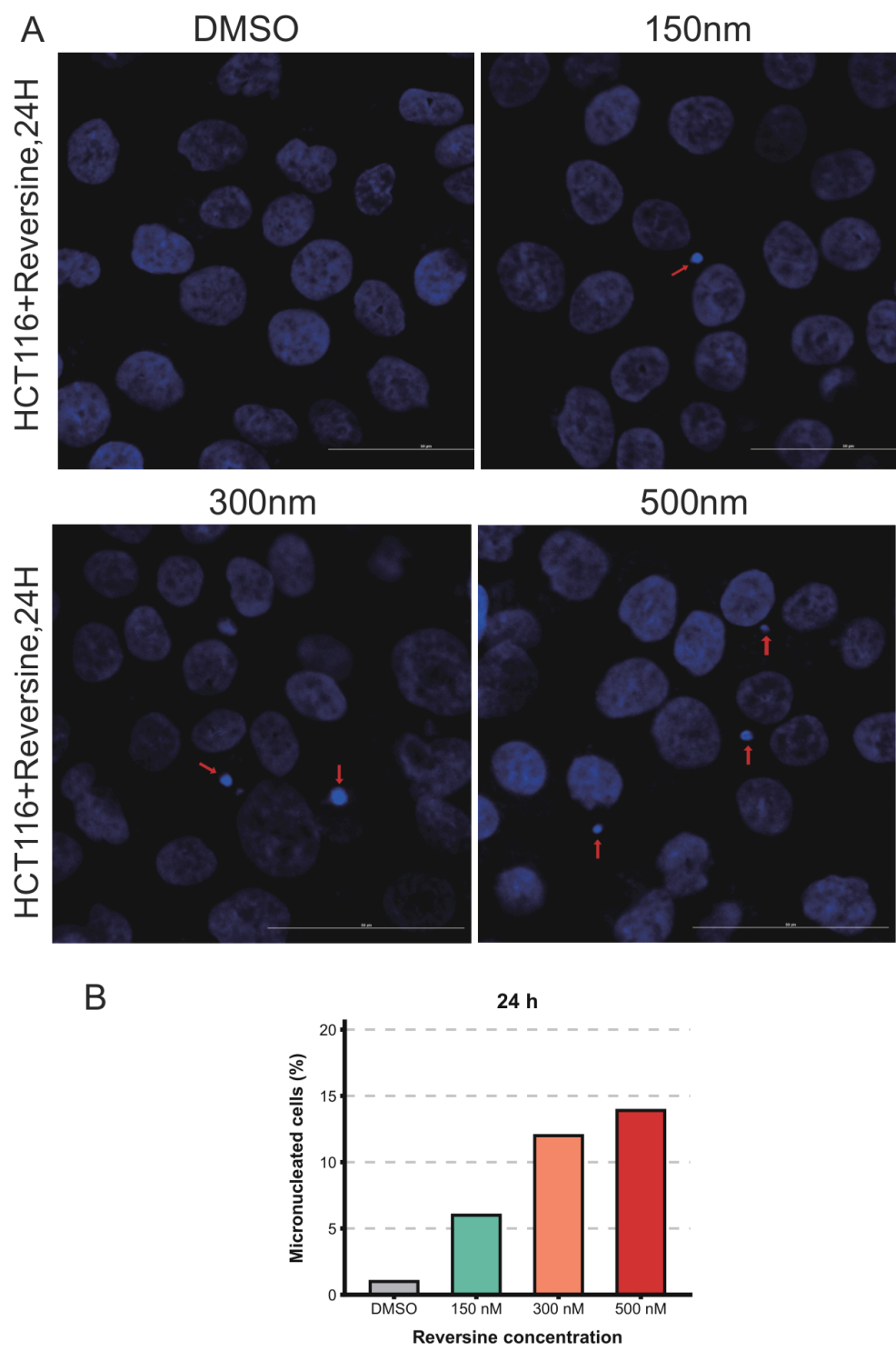


Table S1. The 38-gene mRNA-CIN model coefficients and perturbation-response directions.

Gene	Coef	logFC_RevP	score_cordirection_match
PCID2	0.092273	-0.06259 0.057774	-0.00578 FALSE
CRCP	0.090476	0.37199 0.001034	0.033656 TRUE
SEPHS1	0.076932	-0.82512 4.62E-05	-0.06348 FALSE
RHEB	0.075884	-0.03788 0.250634	-0.00287 FALSE
ASB13	0.075634	-0.44964 0.000967	-0.03401 FALSE
RTKN	0.061824	-0.2133 0.072369	-0.01319 FALSE
ADGRG1	0.059707	1.701439 1.3E-05	0.101587 TRUE
NAT10	0.059097	-0.47232 0.000139	-0.02791 FALSE
UBAC2	0.054402	0.280733 0.000819	0.015272 TRUE
NUDCD3	0.048028	-0.29553 0.000618	-0.01419 FALSE
DDAH2	0.045892	0.337288 0.004581	0.015479 TRUE
FBRS	0.045621	0.491642 0.003495	0.022429 TRUE
ACOT8	0.041157	0.333296 0.013718	0.013717 TRUE
TPD52L2	0.039651	-0.21494 0.002829	-0.00852 FALSE
RALY	0.039281	-0.26218 0.009563	-0.0103 FALSE
BLCAP	0.034313	0.323788 0.001049	0.01111 TRUE
RAE1	0.034222	-0.25484 0.000608	-0.00872 FALSE
DDX56	0.034215	-0.79419 0.000159	-0.02717 FALSE
ERGIC3	0.033837	0.086731 0.014315	0.002935 TRUE
RPRD1B	0.032921	0.033342 0.377211	0.001098 TRUE
NEU1	0.026056	0.58175 0.000293	0.015158 TRUE
GTF3A	0.021564	-0.27261 0.01809	-0.00588 FALSE
SAP18	0.02041	-0.10258 0.134553	-0.00209 FALSE
ABHD12	0.018484	-0.15415 0.126524	-0.00285 FALSE
POLR1D	0.009423	0.151495 0.026209	0.001428 TRUE
GNG4	0.008983	-0.41974 0.016586	-0.00377 FALSE
RTF2	0.008125	-0.1135 0.011016	-0.00092 FALSE
CLNS1A	-0.00466	-0.30999 0.012922	0.001445 TRUE
CKAP4	-0.00822	-0.59366 0.001732	0.004879 TRUE
DDX27	-0.01334	-0.16623 0.009231	0.002218 TRUE
CPNE1	-0.01473	-0.20133 0.000373	0.002965 TRUE
SLC35C2	-0.02884	0.328974 0.000129	-0.00949 FALSE
ZDHHC9	-0.03589	0.728241 0.004352	-0.02614 FALSE
LAMP1	-0.03718	-0.05662 0.018378	0.002105 TRUE
OXCT1	-0.05052	-1.08776 1.53E-05	0.054952 TRUE
TOMM34	-0.05685	-0.05166 0.338637	0.002937 TRUE
UBA7	-0.07392	0.301264 0.12255	-0.02227 FALSE
ATP5F1A	-0.13556	-0.83193 3.8E-05	0.112778 TRUE

Model coefficients of the 38-gene mRNA-CIN score, with each gene's perturbation-response direction in the Reversine RNA-seq experiment.

Table S2. Cross-validation performance of candidate model architectures.

method	Cindex_m	Cindex_s	Pearson_r	Pearson_s	R2_mean	R2_sd	RMSE_mean	RMSE_sd	CV_scheme	model	Accuracy	Recall	F1	ROC-AUC	Sensitivity	Specificity	PPV	NPV	Kappa
Ridge	0.753911	0.032757	0.715975	0.060529	0.515918	0.084401	0.699583	0.052297	stratified	10-fold									
ElasticNet	0.750434	0.032617	0.708715	0.058903	0.5054	0.081279	0.707027	0.053164	stratified	10-fold									
PLS	0.749344	0.029216	0.708562	0.056206	0.504903	0.077892	0.709791	0.048853	stratified	10-fold									
LASSO	0.748245	0.032047	0.705926	0.062438	0.501841	0.085368	0.70891	0.053385	stratified	10-fold									
RFsel+EN	0.747317	0.035222	0.694202	0.062385	0.485419	0.084041	0.721982	0.051547	stratified	10-fold									
XGBsel+EN	0.746611	0.040342	0.696644	0.064882	0.489101	0.088256	0.71871	0.056353	stratified	10-fold									
RFsel+LA	0.74569	0.041294	0.688164	0.07608	0.478779	0.100021	0.726658	0.058405	stratified	10-fold									
XGBsel+L	0.745638	0.043704	0.693939	0.066922	0.485582	0.088375	0.719703	0.050373	stratified	10-fold									
Stepwise	0.739402	0.028879	0.686128	0.062684	0.474308	0.084715	0.733469	0.066731	stratified	10-fold									
repeated ElasticNet	0.782077	0.864625	0.79961	0.844879	0.864625		0.7	0.747678	0.838865	0.564371									
repeated LASSO	0.784251	0.868972	0.802059	0.845068	0.868972		0.7	0.748675	0.843076	0.568719									
repeated PLS	0.79744	0.864625	0.810699	0.846632	0.864625	0.730435	0.766509	0.845727	0.594932										
repeated RFsel+EN	0.784348	0.860474	0.799985	0.837541	0.860474	0.708696	0.750749	0.837455	0.568873										
repeated RFsel+LA	0.78	0.843083	0.792701	0.837154	0.843083	0.717391	0.750566	0.823983	0.560177										
repeated Ridge	0.790918	0.842885	0.801166	0.853214	0.842885	0.73913	0.769921	0.829716	0.581888										
repeated Stepwise	0.792995	0.851581	0.80603	0.842619	0.851581	0.734783	0.771856	0.833519	0.586143										
repeated XGBsel+EN	0.777729	0.873518	0.79736	0.840505	0.873518	0.682609	0.741393	0.851691	0.555771										
repeated XGBsel+L	0.773382	0.864822	0.791884	0.840127	0.864822	0.682609	0.739602	0.84571	0.547075										

Performance of candidate model architectures under stratified 10-fold cross-validation and a repeated 5-fold robustness scheme.

Table S3. Cohort inventory: datasets, platforms, sample sizes, usage and sources.

cohort	platform	n	usage	in source
TCGA-COAI	RNA-seq	4	450 training;	UCSC Xena
GSE39582	Affymetri		585 referenc	cGEO
GSE30540	Affymetri		35 cross-pl	cGEO
GSE13294	Affymetri		155 MSS/MSI	cGEO
GSE18088	Affymetri		53 MSS/MSI	cGEO
GSE14333	Affymetri		290 immune r	cGEO
GSE17536	Affymetri		177 immune;	sGEO
GSE17537	Affymetri		55 survival	GEO
GSE33113	Affymetri		96 immune;	sGEO
GSE37892	Affymetri		130 CIN/MSI	rGEO
GSE38832/	Affymetri-		comparat	cGEO
DepMap/C	CRNA-seq	4	40 cell-lin	cDepMap Public
scRNA CR	10x scRN-		single-c	csee Methods

Datasets, platforms, sample sizes, analysis usage and sources for all cohorts analysed in this study.

Table S4. Univariate and multivariate Cox regression and meta-analysis summaries.

cohort	endpoint	model	HR	lo	hi	p	logHR	se	n	analysis term	events
TCGA-COAIOS		perSD	1.127252	0.918438	1.383543	0.251803	0.119783	0.104525		423 univariate_perSD	
TCGA-COAIOS		binary	1.535063	1.01447	2.322807	0.042566	0.428572	0.211333		423 univariate_perSD	
GSE17536 OS		perSD	0.983157	0.779461	1.240085	0.885975	-0.01699	0.118454		177 univariate_perSD	
GSE17536 OS		binary	1.024035	0.646608	1.621768	0.919352	0.023751	0.234579		177 univariate_perSD	
GSE17536 RFS		perSD	0.87629	0.63141	1.21614	0.429679	-0.13206	0.167218		145 univariate_perSD	
GSE17536 RFS		binary	0.793954	0.408851	1.541794	0.495626	-0.23073	0.338616		145 univariate_perSD	
GSE17537 RFS		perSD	0.83094	0.531215	1.299778	0.417179	-0.1852	0.228265		55 univariate_perSD	
GSE17537 RFS		binary	0.863333	0.357398	2.085472	0.743986	-0.14696	0.449983		55 univariate_perSD	
GSE33113 RFS		perSD	0.875323	0.548789	1.396145	0.576148	-0.13316	0.238207		89 univariate_perSD	
GSE33113 RFS		binary	1.171782	0.462181	2.970856	0.7384	0.158525	0.474664		89 univariate_perSD	
GSE39582 OS		perSD	0.986041	0.857217	1.134224	0.843992	-0.01406	0.071433		579 univariate_perSD	
GSE39582 OS		binary	0.8963	0.676221	1.188004	0.446314	-0.10948	0.143755		579 univariate_perSD	
GSE39582 RFS		perSD	1.099494	0.949305	1.273445	0.205613	0.09485	0.074938		574 univariate_perSD	
GSE39582 RFS		binary	0.935486	0.697849	1.254046	0.655594	-0.06669	0.149525		574 univariate_perSD	
TCGA-COAIOS			0.900746	0.712594	1.138577	0.381907				413 multivariscore	89
TCGA-COAIOS			1.578644	0.838295	2.972843	0.157427				413 multivaricin70	89
TCGA-COAIOS			0.637724	0.336735	1.207751	0.16739				413 multivariprolif	89
TCGA-COAIOS			2.224365	1.716998	2.881656	1.43E-09				413 multivaristage	89
GSE17536 OS			0.708609	0.537071	0.934936	0.014863				177 multivariscore	73
GSE17536 OS			2.727182	1.161781	6.401828	0.021201				177 multivaricin70	73
GSE17536 OS			0.469106	0.208443	1.055734	0.067412				177 multivariprolif	73
GSE17536 OS			3.185401	2.320079	4.373463	7.84E-13				177 multivaristage	73
GSE39582 OS			0.992878	0.847164	1.163656	0.929671				573 multivariscore	190
GSE39582 OS			0.899668	0.507477	1.594953	0.717411				573 multivaricin70	190
GSE39582 OS			0.981419	0.564477	1.706328	0.94701				573 multivariprolif	190
GSE39582 OS			1.954402	1.598801	2.389095	6.17E-11				573 multivaristage	190

Univariate and multivariate Cox regression results and random-effects meta-analysis summaries for overall and relapse-free survival.

Table S5. Closed-loop paired data and correlation statistics.

seq_group	time_h	dose_nM	n_seq	CIN38_mean	CIN38_sd	CIN38_se	CIN70z_mean	CIN70z_sd	flow_DI	flow_DI_se	flow_ge8_mean	flow_ge8_sd	n_flow	dCIN38	dCIN70z
NC(DMSO)	0	0	3	4.3267	0.02653	0.01532	0.95736	0.04435	1.3672	0.13482	27.79417	1.06074	9	0	0
H24-150nn	24	150	3	4.29237	0.00336	0.00194	0.94941	0.01902	1.55204	0.1518	32.92273	0.54633	7	-0.03433	-0.00795
H24-300nn	24	300	3	4.29556	0.00634	0.00366	0.89933	0.03242	1.59012	0.14216	36.04402	0.73529	10	-0.03114	-0.05803
H24-500nn	24	500	3	4.28017	0.01695	0.00979	0.73236	0.02333	1.67151	0.22104	45.56635	0.67561	10	-0.04653	-0.225
H48-150nn	48	150	3	4.30763	0.04121	0.02379	-0.06308	0.03846	1.501	0.14824	33.03563	1.27913	6	-0.01907	-1.02044
H48-300nn	48	300	3	4.34624	0.0214	0.01236	-0.58598	0.01111	2.99272	0.17859	71.10876	1.19396	6	0.01954	-1.54334
H48-500nn	48	500	3	4.45529	0.01745	0.01008	-1.14822	0.06109	4.25217	0.5968	82.84624	3.80427	6	0.12859	-2.10558
H72-150nn	72	150	3	4.41699	0.01618	0.00934	-1.74118	0.05532	1.99181	0.1644	61.63773	0.95137	10	0.09029	-2.69853

(a) Closed-loop paired data (mRNA-CIN and flow-cytometric metrics per condition). Part (b) correlation statistics appears on the next page.

Table S5 (continued). (b) Correlation statistics.

x	y	n	spearman_r	spearman_P	pearson_r	pearson_P
flow_gene8	CIN38_mez	8	0.57143	0.13896	0.80714	0.01544
flow_gene8	dCIN38	8	0.57143	0.13896	0.80714	0.01544
flow_gene8	CIN70z_mez	8	-0.85714	0.00653	-0.8066	0.01556
flow_gene8	dCIN70z	8	-0.85714	0.00653	-0.8066	0.01556
flow_DI	CIN38_mez	8	0.52381	0.18272	0.77941	0.02259
flow_DI	dCIN38	8	0.52381	0.18272	0.77941	0.02259
flow_DI	CIN70z_mez	8	-0.78571	0.02082	-0.63844	0.08844
flow_DI	dCIN70z	8	-0.78571	0.02082	-0.63844	0.08844

(b) Correlation statistics between mRNA-CIN and flow-cytometric CIN metrics across paired conditions.

Table S6. Flow-cytometry group-by-time summary statistics.

time	group	n	DI_mean	DI_sd	ge8N_mean	ge8N_sd	ge4N_mean	frac2N	frac4N	frac8N	frac16N	frac32N	subG1	apo	events
24h	nc	10	1.167076	0.191136	31.99261	2.07932	67.46022	24.17338	35.46761	23.82152	6.978255	1.192838	3.152284	0.996167	15768.1
24h	nc+DMSO	9	1.367204	0.13482	27.79417	1.060743	69.09974	24.65205	41.30558	22.45114	4.615044	0.72798	2.100116	1.334536	16125.22
24h	150 nM	7	1.552036	0.151798	32.92273	0.546331	76.95223	18.7173	44.0295	27.18097	5.073558	0.668199	1.624953	1.482216	13980.71
24h	300 nM	10	1.590117	0.142162	36.04402	0.735289	78.21393	18.21145	42.16991	28.51547	6.721946	0.80661	1.10407	1.516689	16116.5
24h	500 nM	10	1.671513	0.221036	45.56635	0.675613	82.43255	14.54392	36.8662	32.74368	11.41408	1.408588	0.878609	1.536946	15406.6
48h	nc	6	1.026939	0.092087	21.9421	1.490049	54.63193	32.85831	32.68982	15.83678	4.667955	1.43737	4.817498	1.020292	32275.17
48h	nc+DMSO	6	0.754557	0.101427	13.58905	0.625044	36.04738	38.66072	22.45834	9.343534	3.030294	1.215218	10.83699	0.786429	67889
48h	150 nM	6	1.500998	0.148242	33.03563	1.279134	73.06272	20.83365	40.02708	25.24662	6.237058	1.551959	1.944618	1.029205	90772.5
48h	300 nM	6	2.992724	0.178594	71.10876	1.193959	88.75284	4.585822	17.64408	32.44112	26.86237	11.80527	0.218422	1.117085	135842
48h	500 nM	6	4.252167	0.596804	82.84624	3.804273	93.2094	1.438444	10.36316	38.82267	33.7152	10.30837	0.118641	1.603688	59659.17
72h	nc	9	0.938363	0.074439	25.09295	0.663449	53.99036	30.77346	28.8974	15.77874	6.689702	2.624516	5.34908	0.994688	57290.44
72h	nc+DMSO	8	0.921826	0.058061	15.04116	0.585244	41.012	37.94774	25.97084	9.399145	3.794782	1.847235	8.435824	0.909536	71128.63
72h	150 nM	10	1.991808	0.164401	61.63773	0.951371	85.61881	6.214401	23.98107	31.01626	21.50818	9.113296	0.264063	1.351524	50564.2
72h	300 nM	10	6.247766	0.767052	65.89464	0.592828	69.34839	1.037059	3.453751	12.24833	25.76467	27.88163	0.194804	3.531581	21137.1
72h	500 nM	7	7.31133	0.693223	72.70573	0.547457	76.41087	1.205626	3.705138	12.02971	30.34734	30.32868	0.206738	2.739695	51685.43