

Supplemental Tables and Figures

Xiaoxiao Zhang^{1,2} and Maik Kschischo^{1*}

^{1*}Department of Mathematics and Technology, University of Applied Sciences Koblenz, Joseph-Rovan-Allee 2, 53424 Remagen, Germany.

²Department of Informatics, Technical University of Munich, Munich, 81675, Germany.

*Corresponding author(s). E-mail(s):

kschischo@rheinahrcampus.de;

Contributing authors: zhang@hs-koblenz.de;

Supplementary tables for the survival analysis

Table 1 Association between WCIN and overall survival across cancer types

cohort	sample_number	pvalue	low_surv5 ^a	high_surv5 ^b	low_surv5_n ^c	high_surv5_n ^d
LGG	509	0.00	0.76	0.48	38.00	27.00
UCEC	518	0.00	0.88	0.65	65.00	42.00
LIHC	366	0.00	0.54	0.44	24.00	17.00
BRCA	1066	0.00	0.87	0.77	133.00	116.00
OV	558	0.01	0.40	0.29	62.00	54.00
SARC	252	0.03	0.59	0.50	30.00	26.00
TGCT	133	0.04	1.00	0.95	28.00	23.00
HNSC	516	0.05	0.56	0.40	24.00	28.00
UVM*	80	0.07	0.94	0.92	33.00	31.00
THCA	497	0.09	0.97	0.89	51.00	45.00
THYM	122	0.09	0.98	0.88	18.00	15.00
KIRP	282	0.10	0.84	0.68	34.00	19.00
PCPG	161	0.14	0.98	0.95	16.00	12.00
GBM	571	0.17	0.07	0.06	10.00	9.00
ACC	89	0.18	0.69	0.56	14.00	14.00
DLBC	48	0.19	0.72	0.90	6.00	3.00
STAD	433	0.25	0.41	0.35	10.00	8.00
CESC	294	0.27	0.73	0.60	25.00	16.00
LAML	179	0.34	0.24	0.17	6.00	3.00
PRAD	489	0.41	0.99	0.97	35.00	49.00
KIRC	506	0.44	0.62	0.63	66.00	81.00
CHOL*	36	0.45	0.71	0.88	12.00	15.00
READ*	154	0.48	0.94	0.96	51.00	64.00
ESCA*	182	0.49	0.82	0.70	65.00	50.00
MESO*	86	0.59	0.65	0.71	28.00	28.00
LUAD	491	0.63	0.42	0.40	28.00	25.00
COAD	425	0.75	0.66	0.54	25.00	15.00
KICH	65	0.77	0.86	0.85	17.00	20.00
SKCM*	104	0.77	0.84	0.91	33.00	37.00
PAAD	183	0.85	0.16	0.33	3.00	5.00
BLCA	405	0.87	0.42	0.42	24.00	23.00
UCS*	56	0.88	0.79	0.81	22.00	20.00
LUSC	481	0.96	0.50	0.45	38.00	42.00

^a5-year overall survival probability in low WCIN group.

^b5-year overall survival probability in high WCIN group.

^cnumber of samples at risk in low WCIN group at 5th year.

^dnumber of samples at risk in high WCIN group at 5th year.

*1-year overall survival statistics was reported in these cancer types due to short survival.

Table 2 Association between SCIN and overall survival across cancer types

cohort	sample_number	pvalue	low_surv5 ^a	high_surv5 ^b	low_surv5_n ^c	high_surv5_n ^d
UCEC	518	0.00	0.90	0.60	69.00	38.00
ACC	89	0.00	0.81	0.39	21.00	7.00
THCA	497	0.00	0.96	0.72	86.00	10.00
SARC	252	0.00	0.62	0.47	31.00	25.00
KIRP	282	0.00	0.86	0.65	30.00	23.00
THYM	122	0.01	0.98	0.76	29.00	4.00
PCPG	161	0.01	1.00	0.92	15.00	13.00
LGG	509	0.01	0.73	0.52	32.00	33.00
KICH	65	0.02	0.97	0.70	23.00	14.00
COAD	425	0.02	0.71	0.52	23.00	17.00
OV	558	0.04	0.32	0.37	58.00	58.00
ESCA*	182	0.06	0.81	0.72	60.00	55.00
LUAD	491	0.08	0.48	0.35	28.00	25.00
TGCT	133	0.09	0.95	1.00	25.00	26.00
UCS*	56	0.09	0.75	0.85	20.00	22.00
READ*	154	0.10	0.96	0.93	64.00	51.00
LAML*	179	0.10	0.58	0.45	66.00	22.00
PAAD	183	0.11	0.25	0.26	5.00	3.00
KIRC	506	0.17	0.65	0.59	83.00	64.00
BRCA	1066	0.22	0.83	0.81	129.00	120.00
GBM	571	0.23	0.05	0.08	7.00	12.00
CHOL*	36	0.39	0.76	0.83	13.00	14.00
PRAD	489	0.45	0.99	0.97	40.00	44.00
BLCA	405	0.46	0.40	0.44	21.00	26.00
UVM*	80	0.50	0.93	0.93	38.00	26.00
HNLC	516	0.56	0.51	0.44	27.00	25.00
LUSC	481	0.57	0.46	0.49	36.00	44.00
SKCM*	104	0.63	0.84	0.93	36.00	34.00
MESO*	86	0.64	0.62	0.75	28.00	28.00
LIHC	366	0.73	0.48	0.50	23.00	18.00
CESC	294	0.77	0.68	0.66	20.00	21.00
STAD	433	0.90	0.35	0.43	11.00	7.00
DLBC*	48	0.93	0.96	0.89	23.00	15.00

^a5-year overall survival probability in low SCIN group.

^b5-year overall survival probability in high SCIN group.

^cnumber of samples at risk in low SCIN group at 5th year.

^dnumber of samples at risk in high SCIN group at 5th year.

*1-year overall survival statistics was reported in these cancer types due to short survival.

Table 3 Association between WCIN and disease specific survival across cancer types

cohort	sample_number	pvalue	low_surv5 ^a	high_surv5 ^b	low_surv5_n ^c	high_surv5_n ^d
UCEC	516	0.00	0.92	0.74	65.00	42.00
LGG	501	0.00	0.76	0.53	36.00	26.00
LIHC	357	0.01	0.63	0.58	24.00	16.00
OV	524	0.01	0.42	0.32	59.00	53.00
BRCA	1048	0.02	0.92	0.86	128.00	112.00
UVM*	80	0.03	0.94	0.92	33.00	31.00
DLBC	48	0.04	0.75	1.00	6.00	3.00
SARC	247	0.05	0.65	0.56	28.00	26.00
KIRP	278	0.06	0.89	0.80	31.00	19.00
PCPG	161	0.06	1.00	0.96	16.00	12.00
TGCT	133	0.08	1.00	0.95	28.00	23.00
BLCA	391	0.13	0.51	0.60	24.00	23.00
ACC	87	0.14	0.72	0.58	14.00	14.00
HNSC	492	0.19	0.65	0.59	23.00	26.00
THCA	491	0.21	0.98	0.95	50.00	44.00
KIRC	495	0.23	0.74	0.72	65.00	79.00
PRAD	488	0.26	1.00	0.98	35.00	49.00
GBM	533	0.26	0.07	0.06	9.00	8.00
THYM	122	0.32	1.00	0.95	18.00	15.00
CESC	290	0.36	0.77	0.70	24.00	16.00
STAD	406	0.37	0.53	0.54	10.00	8.00
ESCA*	180	0.38	0.90	0.79	64.00	49.00
LUSC	432	0.39	0.72	0.65	32.00	36.00
CHOL*	35	0.39	0.75	0.94	12.00	15.00
READ*	148	0.62	0.95	0.99	48.00	62.00
KICH	65	0.72	0.89	0.88	17.00	20.00
LUAD	456	0.78	0.59	0.55	25.00	23.00
PAAD	176	0.82	0.20	0.41	3.00	5.00
SKCM*	104	0.83	0.91	0.94	33.00	37.00
UCS*	55	0.94	0.78	0.81	21.00	20.00
COAD	410	0.94	0.76	0.77	23.00	14.00
MESO*	66	0.95	0.74	0.74	20.00	23.00

^a5-year disease specific survival probability in low WCIN group.

^b5-year disease specific survival probability in high WCIN group.

^cnumber of samples at risk in low WCIN group at 5th year.

^dnumber of samples at risk in high WCIN group at 5th year.

*1-year disease-specific survival statistics was reported in these cancer types due to short survival.

Table 4 Association between SCIN and disease specific survival across cancer types

cohort	sample_number	pvalue	low_surv5 ^a	high_surv5 ^b	low_surv5_n ^c	high_surv5_n ^d
UCEC	516	0.00	0.94	0.69	69.00	38.00
THCA	491	0.00	0.99	0.82	84.00	10.00
KIRP	278	0.00	0.97	0.72	27.00	23.00
ACC	87	0.00	0.84	0.41	21.00	7.00
SARC	247	0.00	0.69	0.51	30.00	24.00
LGG	501	0.00	0.79	0.52	32.00	30.00
OV	524	0.01	0.34	0.41	54.00	58.00
KICH	65	0.01	1.00	0.73	23.00	14.00
ESCA*	180	0.02	0.89	0.81	58.00	55.00
READ*	148	0.04	1.00	0.93	64.00	46.00
PCPG	161	0.05	1.00	0.96	15.00	13.00
LUAD	456	0.06	0.64	0.50	26.00	22.00
BRCA	1048	0.06	0.91	0.87	124.00	116.00
COAD	410	0.08	0.81	0.72	21.00	16.00
TGCT	133	0.11	0.95	1.00	25.00	26.00
THYM	122	0.12	1.00	0.91	29.00	4.00
PAAD	176	0.13	0.30	0.32	5.00	3.00
KIRC	495	0.17	0.75	0.70	82.00	62.00
UCS*	55	0.19	0.75	0.84	20.00	21.00
PRAD	488	0.25	1.00	0.98	40.00	44.00
LIHC	357	0.27	0.60	0.61	22.00	18.00
CHOL*	35	0.34	0.81	0.88	13.00	14.00
STAD	406	0.37	0.53	0.55	11.00	7.00
GBM	533	0.44	0.06	0.08	7.00	10.00
LUSC	432	0.53	0.72	0.66	29.00	39.00
MESO*	66	0.54	0.71	0.78	24.00	19.00
DLBC*	48	0.55	1.00	0.94	23.00	15.00
UVM*	80	0.72	0.93	0.93	38.00	26.00
SKCM*	104	0.89	0.92	0.93	36.00	34.00
BLCA	391	0.91	0.59	0.53	21.00	26.00
HNSC	492	0.94	0.63	0.62	25.00	24.00
CESC	290	0.98	0.71	0.75	19.00	21.00

^a5-year disease specific survival probability in low SCIN group.

^b5-year disease specific survival probability in high SCIN group.

^cnumber of samples at risk in low SCIN group at 5th year.

^dnumber of samples at risk in high SCIN group at 5th year.

*1-year disease-specific survival statistics was reported in these cancer types due to short survival.

Table 5 Association between WCIN and disease free survival across cancer types

cohort	sample_number	pvalue	low_surv5 ^a	high_surv5 ^b	low_surv5_n ^c	high_surv5_n ^d
UCEC	406	0.00	0.90	0.72	55.00	28.00
OV	279	0.00	0.25	0.10	17.00	8.00
PRAD	332	0.02	0.87	0.78	24.00	27.00
LIHC	315	0.03	0.41	0.22	14.00	4.00
COAD	175	0.05	0.83	0.62	11.00	3.00
LGG*	130	0.06	0.97	0.98	61.00	41.00
LUSC	295	0.08	0.73	0.63	24.00	24.00
MESO*	15	0.08	0.67	1.00	6.00	2.00
CESC	170	0.11	0.85	0.76	17.00	9.00
KICH	29	0.16	0.91	1.00	5.00	12.00
BRCA	927	0.25	0.86	0.82	105.00	85.00
CHOL*	24	0.26	0.71	0.50	10.00	4.00
DLBC	28	0.28	1.00	0.92	5.00	3.00
UCS*	26	0.28	1.00	0.77	11.00	9.00
PAAD*	68	0.35	0.86	0.80	21.00	21.00
KIRC	107	0.38	0.92	0.76	19.00	19.00
TGCT	104	0.39	0.68	0.82	9.00	15.00
SARC	148	0.44	0.56	0.45	16.00	11.00
GBM*	3	0.48	1.00	1.00	1.00	2.00
KIRP	180	0.52	0.74	0.89	17.00	14.00
READ*	42	0.53	0.90	1.00	15.00	20.00
LUAD	291	0.59	0.63	0.54	19.00	18.00
THCA	352	0.76	0.91	0.90	36.00	33.00
ESCA*	87	0.80	0.75	0.82	28.00	22.00
ACC	52	0.83	0.68	0.72	11.00	10.00
BLCA	187	0.88	0.72	0.71	11.00	15.00
HNSC	130	0.94	0.69	0.54	7.00	4.00
STAD	255	0.96	0.59	0.70	7.00	6.00
PCPG	144	0.97	0.95	0.97	12.00	10.00

^a5-year disease free survival probability in low WCIN group.

^b5-year disease free survival probability in high WCIN group.

^cnumber of samples at risk in low WCIN group at 5th year.

^dnumber of samples at risk in high WCIN group at 5th year.

*1-year disease-free survival statistics was reported in these cancer types due to short survival.

Table 6 Association between SCIN and disease free survival across cancer types

cohort	sample_number	pvalue	low_surv5 ^a	high_surv5 ^b	low_surv5_n ^c	high_surv5_n ^d
UCEC	406	0.00	0.92	0.65	61.00	22.00
ACC	52	0.00	0.81	0.48	16.00	5.00
PRAD	332	0.00	0.88	0.75	27.00	24.00
KIRP	180	0.02	0.88	0.72	18.00	13.00
LUSC	295	0.03	0.71	0.64	25.00	23.00
THCA	352	0.03	0.92	0.81	64.00	5.00
KIRC	107	0.04	0.94	0.70	24.00	14.00
READ*	42	0.09	1.00	0.89	20.00	15.00
BRCA	927	0.10	0.88	0.80	100.00	90.00
PAAD*	68	0.14	0.91	0.74	23.00	19.00
LIHC	315	0.15	0.33	0.31	12.00	6.00
GBM*	3	0.16	1.00	1.00	1.00	2.00
KICH	29	0.18	0.94	1.00	11.00	6.00
COAD	175	0.27	0.79	0.71	9.00	5.00
CESC	170	0.30	0.86	0.75	15.00	11.00
HNSC	130	0.31	0.66	0.59	8.00	3.00
LUAD	291	0.37	0.63	0.54	19.00	18.00
DLBC*	28	0.45	1.00	1.00	14.00	11.00
MESO*	15	0.46	0.69	1.00	6.00	2.00
ESCA*	87	0.49	0.80	0.76	26.00	24.00
PCCPG	144	0.49	0.93	1.00	12.00	10.00
LGG	130	0.61	0.76	0.65	3.00	3.00
OV	279	0.64	0.22	0.14	17.00	8.00
CHOL*	24	0.65	0.67	0.57	10.00	4.00
BLCA	187	0.69	0.73	0.69	14.00	12.00
TGCT	104	0.70	0.75	0.75	14.00	10.00
STAD	255	0.76	0.67	0.64	8.00	5.00
UCS*	26	0.76	0.92	0.83	10.00	10.00
SARC	148	0.79	0.53	0.48	14.00	13.00

^a5-year disease free survival probability in low SCIN group.

^b5-year disease free survival probability in high SCIN group.

^cnumber of samples at risk in low SCIN group at 5th year.

^dnumber of samples at risk in high SCIN group at 5th year.

*1-year disease-free survival statistics was reported in these cancer types due to short survival.

Table 7 Association between WCIN and progression free survival across cancer types

Cohort	sample_number	pvalue	low_surv5	high_surv5	low_surv5_n	high_surv5_n
UCEC	518	0.00	0.85	0.56	60.00	34.00
LGG	509	0.00	0.55	0.28	26.00	10.00
PRAD	489	0.00	0.75	0.66	27.00	36.00
OV	558	0.00	0.18	0.08	24.00	11.00
LIHC	366	0.00	0.35	0.19	13.00	5.00
UVM*	79	0.01	0.94	0.69	31.00	24.00
THYM	122	0.01	0.91	0.65	17.00	12.00
SKCM*	104	0.03	0.78	0.64	28.00	24.00
CHOL*	36	0.06	0.63	0.34	10.00	5.00
KIRC	504	0.06	0.68	0.59	55.00	63.00
CESC	294	0.06	0.71	0.62	22.00	14.00
ESCA*	182	0.07	0.66	0.58	47.00	37.00
ACC	89	0.08	0.50	0.39	11.00	10.00
SARC	252	0.14	0.44	0.34	20.00	15.00
HNSC	516	0.20	0.51	0.46	20.00	24.00
BRCA	1066	0.22	0.80	0.76	119.00	102.00
BLCA	406	0.29	0.38	0.44	19.00	17.00
READ*	154	0.32	0.90	0.88	47.00	57.00
GBM	571	0.33	0.04	0.02	5.00	3.00
TGCT	133	0.43	0.69	0.80	17.00	20.00
COAD	425	0.43	0.63	0.55	21.00	9.00
DLBC	48	0.43	0.70	0.77	6.00	3.00
KIRP	281	0.48	0.72	0.79	28.00	17.00
LUSC	482	0.53	0.55	0.54	31.00	34.00
LUAD	491	0.68	0.39	0.39	20.00	18.00
THCA	497	0.74	0.85	0.83	42.00	40.00
KICH	65	0.75	0.87	0.87	17.00	20.00
STAD	435	0.82	0.40	0.48	10.00	8.00
PCPG	161	0.83	0.80	0.87	11.00	11.00
PAAD*	183	0.85	0.62	0.64	42.00	50.00
UCS*	56	0.92	0.50	0.58	14.00	14.00
MESO*	84	0.97	0.59	0.52	22.00	19.00

^a5-year progression free survival probability in low WCIN group.

^b5-year progression free survival probability in high WCIN group.

^cnumber of samples at risk in low WCIN group at 5th year.

^dnumber of samples at risk in high WCIN group at 5th year.

*1-year progression-free survival statistics was reported in these cancer types due to short survival.

Table 8 Association between SCIN and progression free survival across cancer types

cohort	sample_number	pvalue	low_surv5 ^a	high_surv5 ^b	low_surv5_n ^c	high_surv5_n ^d
UCEC	518	0.00	0.86	0.50	65.00	29.00
ACC	89	0.00	0.66	0.22	16.00	5.00
KIRP	281	0.00	0.86	0.61	27.00	18.00
THCA	497	0.00	0.87	0.64	75.00	7.00
PRAD	489	0.00	0.76	0.65	32.00	31.00
LGG	509	0.00	0.50	0.33	18.00	18.00
CHOL*	36	0.02	0.69	0.27	11.00	4.00
THYM	122	0.03	0.85	0.60	26.00	3.00
GBM*	571	0.06	0.29	0.32	73.00	70.00
COAD	425	0.07	0.65	0.53	19.00	11.00
SARC	252	0.08	0.44	0.34	19.00	16.00
PAAD*	183	0.09	0.67	0.59	46.00	46.00
KICH	65	0.09	0.97	0.72	23.00	14.00
KIRC	504	0.11	0.66	0.60	67.00	51.00
ESCA*	182	0.12	0.61	0.63	41.00	43.00
LIHC	366	0.12	0.26	0.27	11.00	7.00
BRCA	1066	0.12	0.81	0.75	118.00	103.00
READ*	154	0.12	0.93	0.84	59.00	45.00
UVM*	79	0.21	0.82	0.80	32.00	23.00
OV	558	0.25	0.14	0.13	22.00	13.00
BLCA	406	0.31	0.44	0.39	16.00	20.00
SKCM*	104	0.31	0.74	0.67	28.00	24.00
LUAD	491	0.35	0.40	0.38	19.00	19.00
UCS*	56	0.41	0.54	0.54	14.00	14.00
CESC	294	0.41	0.70	0.63	19.00	17.00
STAD	435	0.41	0.45	0.44	11.00	7.00
PCPG	161	0.45	0.84	0.84	13.00	9.00
DLBC*	48	0.51	0.83	0.84	19.00	15.00
MESO*	84	0.73	0.63	0.47	24.00	17.00
TGCT	133	0.77	0.72	0.77	19.00	18.00
LUSC	482	0.79	0.51	0.57	29.00	36.00
HNSC	516	0.93	0.49	0.49	23.00	21.00

^a5-year progression free survival probability in low SCIN group.

^b5-year progression free survival probability in high SCIN group.

^cnumber of samples at risk in low SCIN group at 5th year.

^dnumber of samples at risk in high SCIN group at 5th year.

*1-year progression-free survival statistics was reported in these cancer types due to short survival.

Supplemental Figures

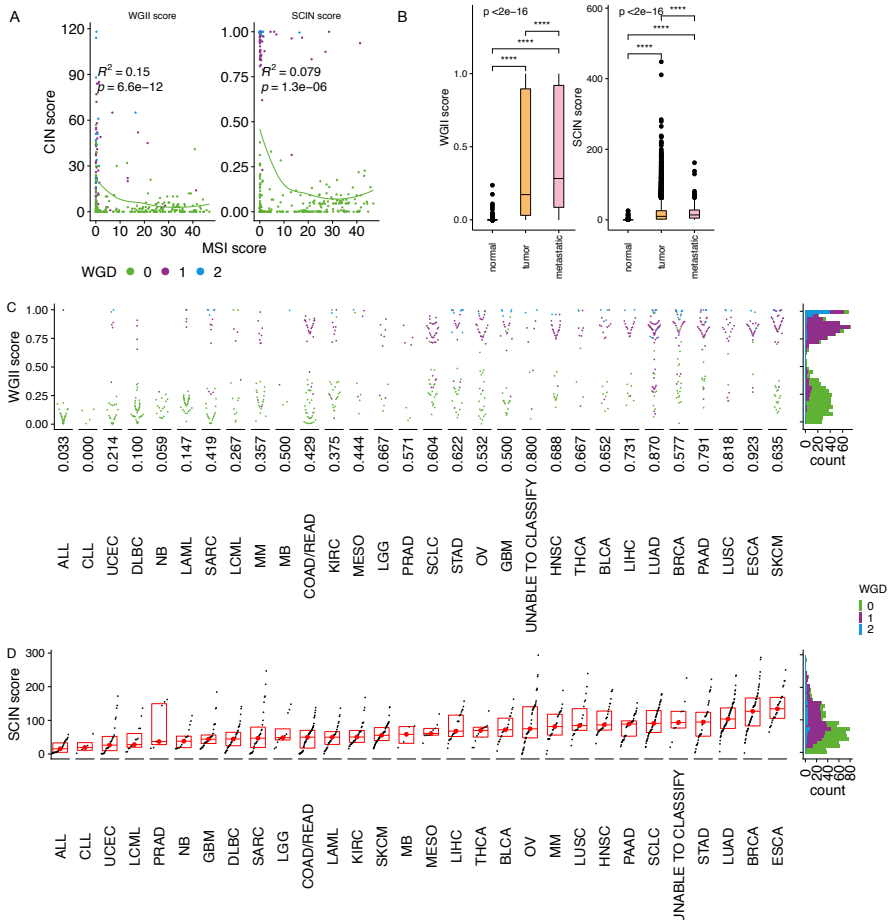


Fig. 1 **(A)** Comparison of the WGI and the SCIN score between samples from normal tissue with primary and metastatic samples. significant differences between high and low WGI were observed. **(B)** The relationship between microsatellite instability (MSI) scores and the the WGI and the SCIN score. **(C)** Cancer-type-wise WGI score distribution in CCLE cell lines, cancer types are ordered by the median WGI scores; whole genome doubling (WGD) status is encoded by colours. The number reported on x axis is the proportion of samples that underwent WGD. **(D)** SCIN score distribution in CCLE cell lines, cancer types are ordered by their median SCIN scores

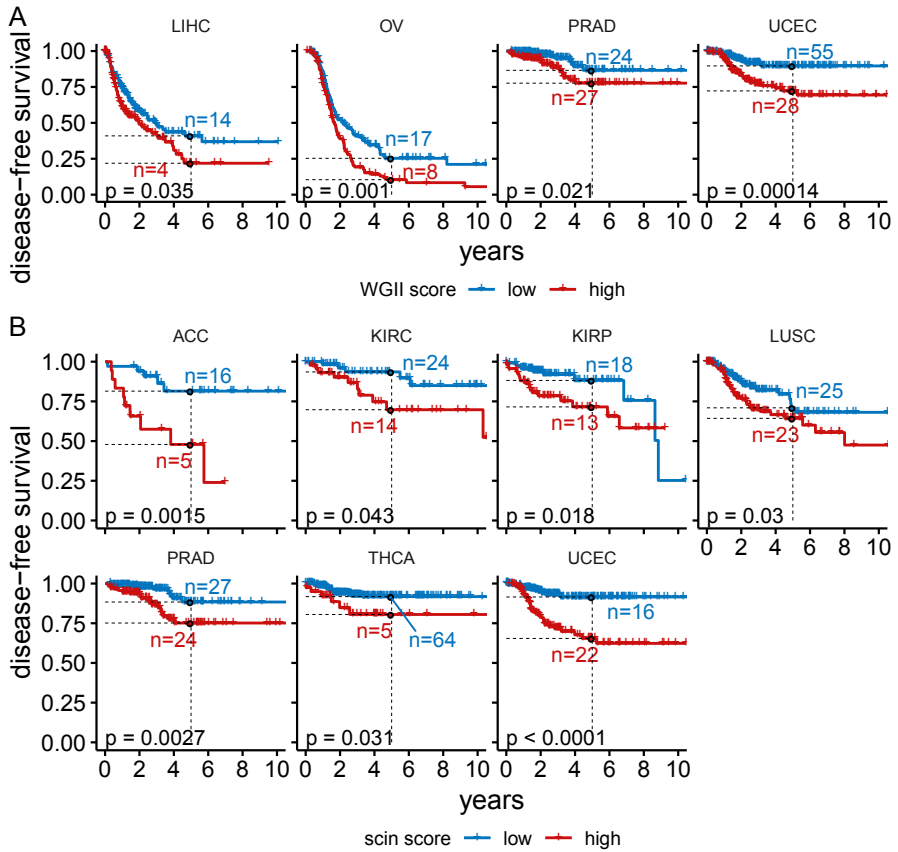


Fig. 2 (A) Disease free survival in four cancer types where significant differences between high and low WGII were observed. (B) Disease free survival in four cancer types where significant differences between high and low SCIN scores were observed.

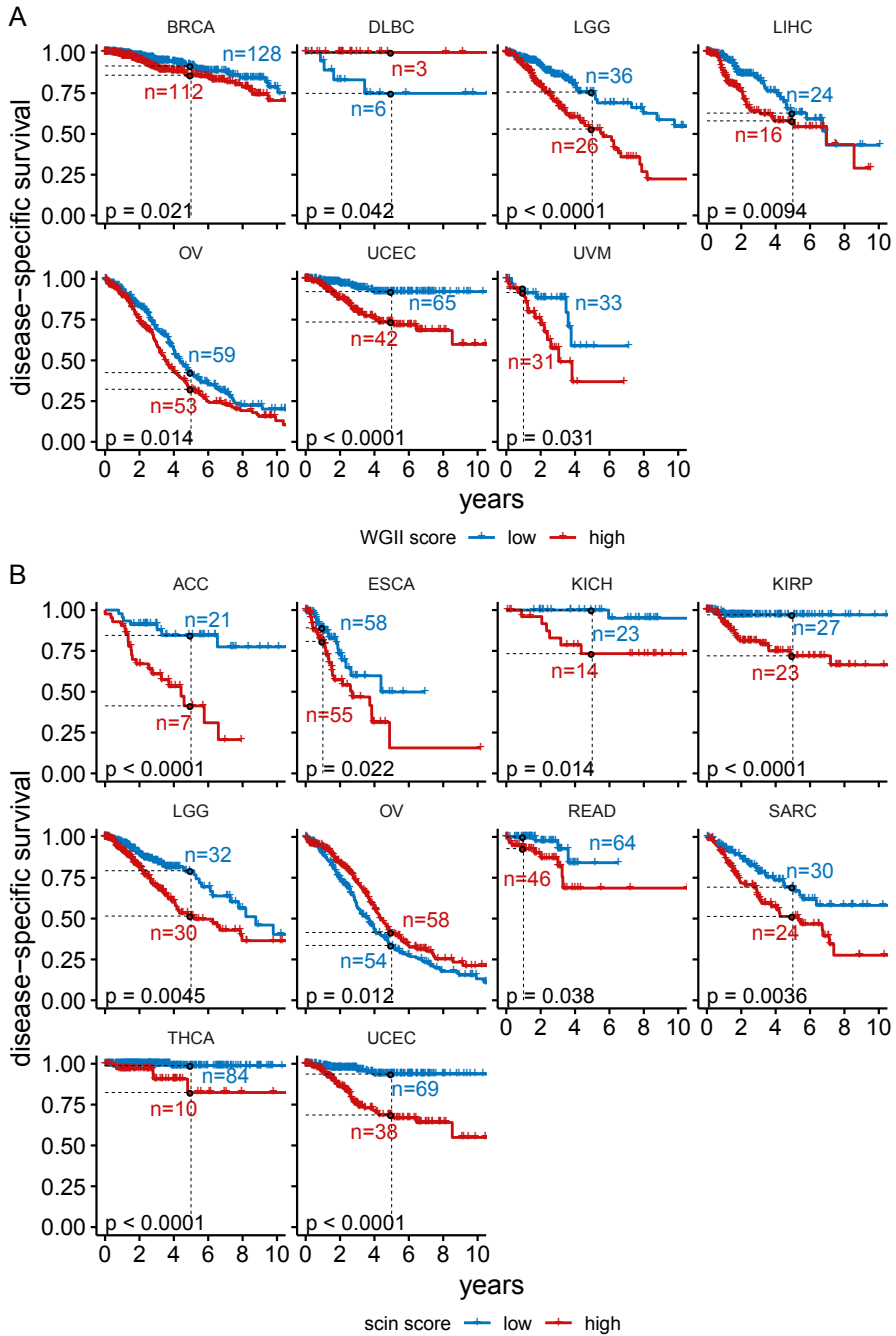


Fig. 3 (A) Disease specific survival in seven cancer types where significant differences between high and low WGII were observed. (B) Disease specific survival where significant differences between high and low SCIN score were observed.

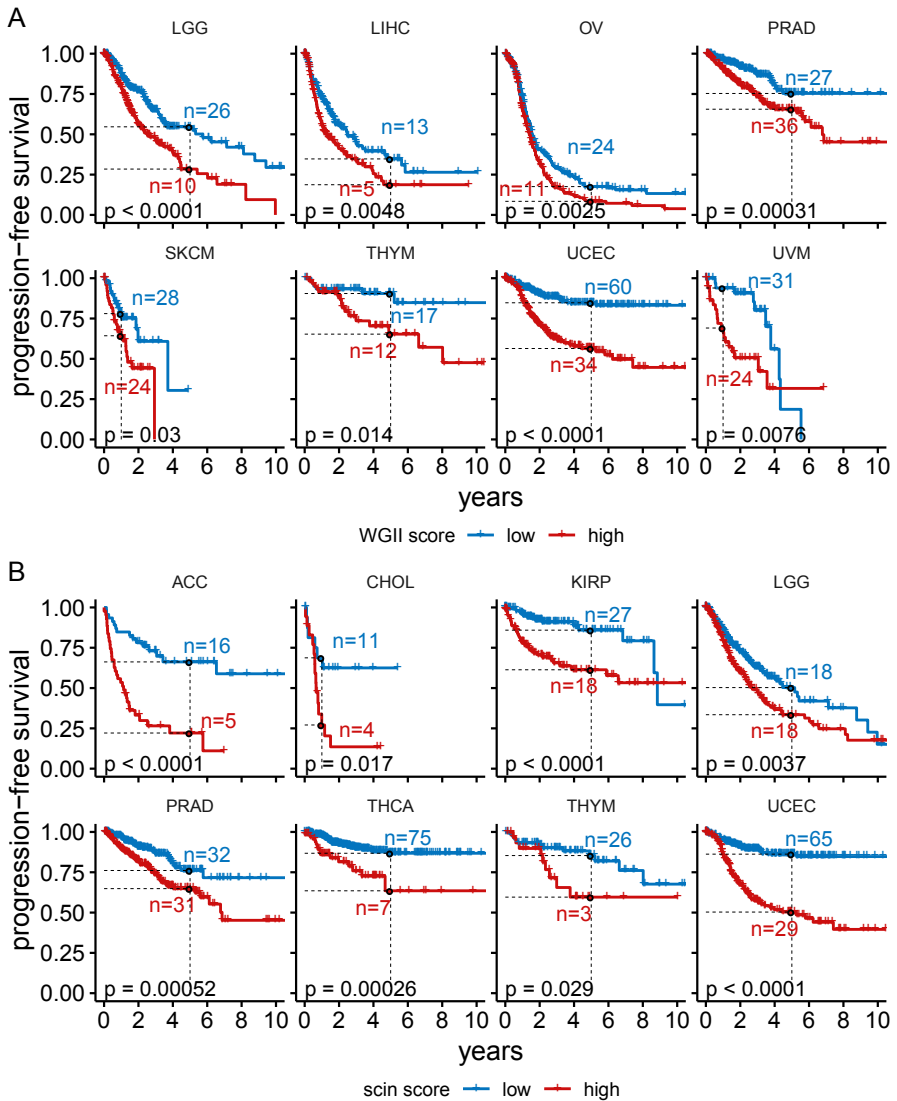


Fig. 4 (A) Progression free survival in eight cancer types where significant differences between high and low WGII were observed. (B) Progression free survival where significant differences between high and low SCIN scores were observed.

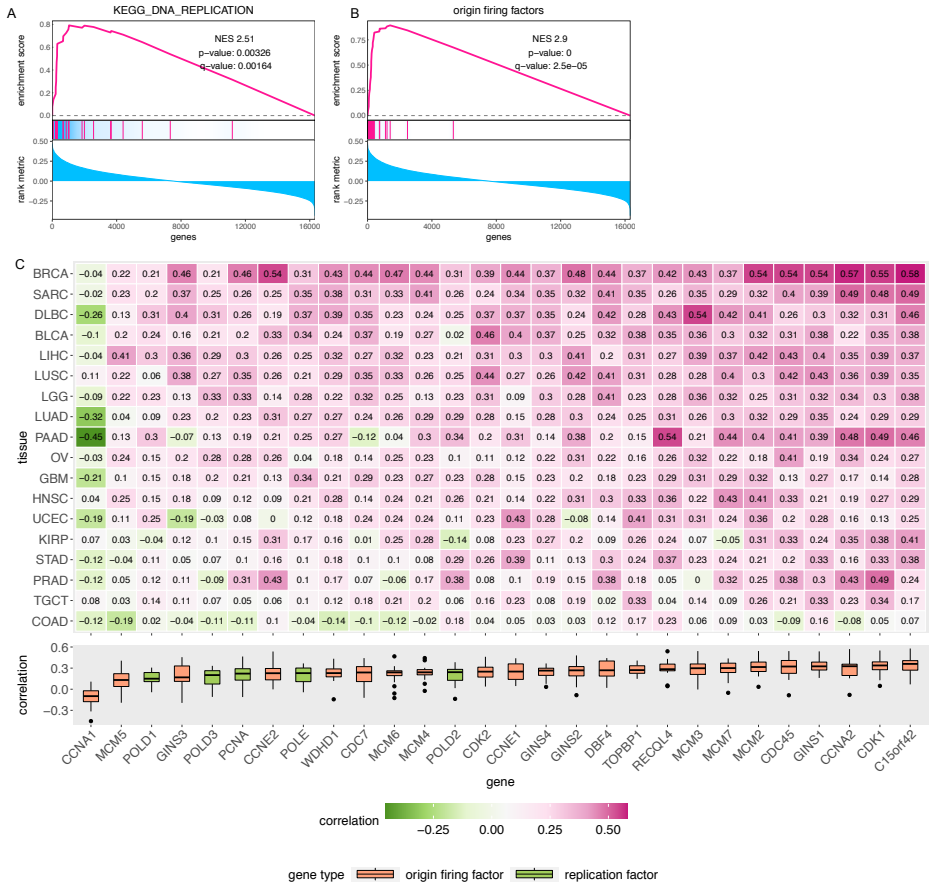
14 *Supplemental Tables and Figures*

Fig. 5 (A) Gene set enrichment analysis (GSEA) for KEGG DNA replication gene set. All genes were ordered according to the correlation of their expression with the SCIN score and enrichment of the replication gene set was tested using resampling. (B) GSEA analysis for manually curated origin firing factor gene set. (C) Gene expression of many origin firing factors is positively correlated with SCIN scores in many cancer types. Rows and columns of the heatmap represent cancer types and origin firing factor genes, respectively. Cancer types are clustered based on their correlation coefficients with origin firing factors. Genes are ordered based on the median correlation coefficient. Colour and values encoded in the heatmap represent the Spearman correlation coefficient.

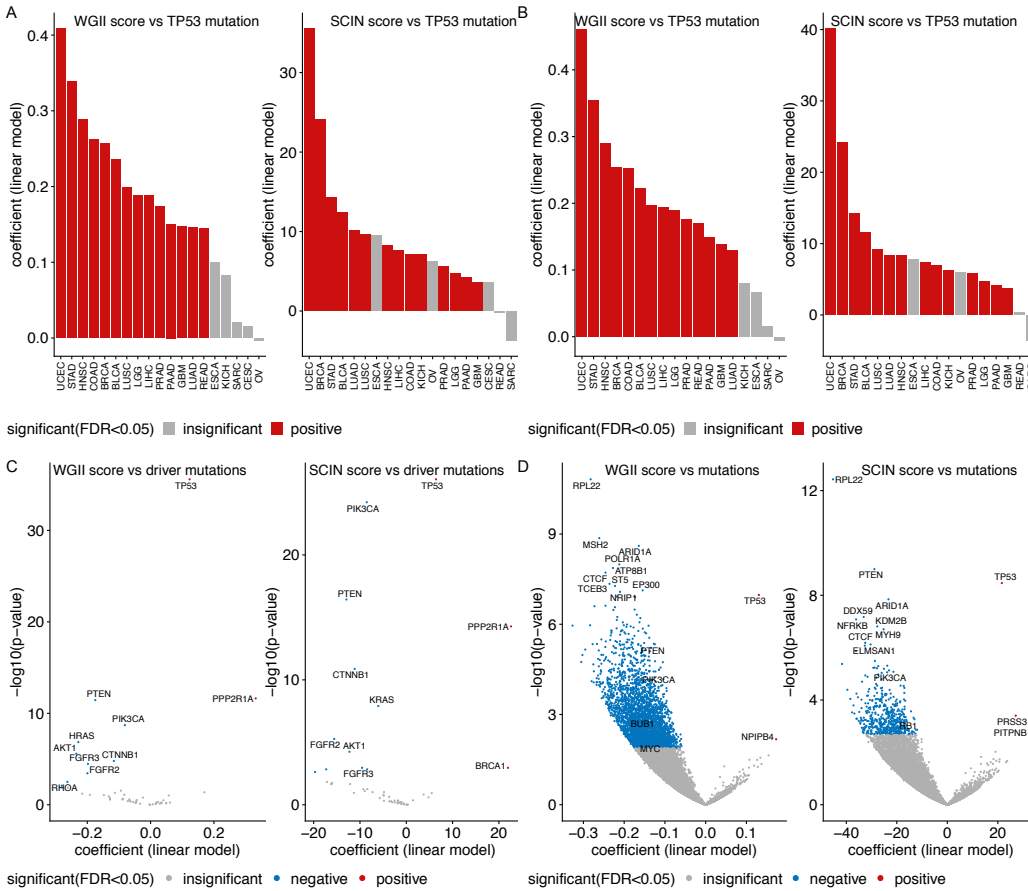


Fig. 6 (A) TP53 mutation is positively associated with high SCIN in multiple cancer types. The bar shows the linear regression model coefficient using SCIN score as dependent variable and TP53 mutation as explanatory variable. Only cancer types with both TP53 mutant and TP53 wild type in more than 20 samples are considered. (B) TP53 mutation is positively associated with SCIN score, the association analysis is performed as in (A), except excluding microsatellite instable (MSI) samples. (C) The volcano plot shows the association between CIN and validated driver mutations, association analysis is performed using non-MSI samples only. (D) The volcano plot shows the correlation between CIN score and somatic mutation in CCLE cell line samples. For all association in (A)-(D), red, grey and blue colors encode positive, insignificant and negative associations. $FDR > 0.05$ is considered as insignificant. Gene names of known important oncogenes and CIN driver genes are annotated in the volcano plot, if significantly associated with CIN.