

Additional file 1:

Supplementary materials

Application of RNAi-induced gene expression profiles for prognostic prediction in breast cancer

Yue Wang^{1,2}, Kenneth M.K. Mark², Matthew H. Ung², Arminja Kettenbach^{3,4}, Todd Miller^{2,3}, Wei Xu¹, Wenqing Cheng¹, Tian Xia^{1,*}, and Chao Cheng^{2,3,5,*}

¹School of Electronic Information and Communications at Huazhong University of Science and Technology, Wuhan, Hubei, 430074, China.

²Department of Molecular and Systems Biology, Geisel School of Medicine at Dartmouth, Hanover, NH 03755, USA.

³Norris Cotton Cancer Center, Geisel School of Medicine at Dartmouth, Lebanon, NH 03766, USA.

⁴Department of Biochemistry, Geisel School of Medicine at Dartmouth, Hanover, NH 03755, USA.

⁵Department of Biomedical Data Sciences, Geisel School of Medicine at Dartmouth, Lebanon, NH 03766, USA.

Contents

Figure S1. The gene consistency across 6 different breast cancer datasets.

Figure S2. Distinguishing hereditary and sporadic breast cancer samples using RIPS.

Figure S3. *TP53* mutation rate in different breast cancer datasets.

Figure S4. Prognosis of Hatzis validation dataset using RIPS_{RAD51}.

Figure S5. Mutation rate of significant genes in different cell proliferation patient groups.

Figure S6. Gene expression of *BRCA1*, *BRCA2* and *RAD51* in different breast cancer mutations.

Figure S7. Gene expression of *XRCC2* in different cell proliferation patient groups.

Figure S8. Prediction of *BRIT1* in METABRIC breast cancer dataset.

Table S1. Breast cancer datasets used in this analysis.

Table S2. Regulated gene lists for *BRCA1* and *RAD51*.

Table S3. Pathway enrichment of genes regulated by *BRCA1* and *RAD51*.

Table S4. HR-related genes and their DNA methylation level in RIPS-low, -intermediate and -high groups.

Table S5. Survival analysis results for 11 significant HR genes using their expression level.

Figure S1. The gene consistency across 6 different breast cancer datasets. The size of each circle imply the Jaccard score of two datasets. The bigger circle, the higher Jaccard score. Purple infers high consistency and orange infers no consistency.

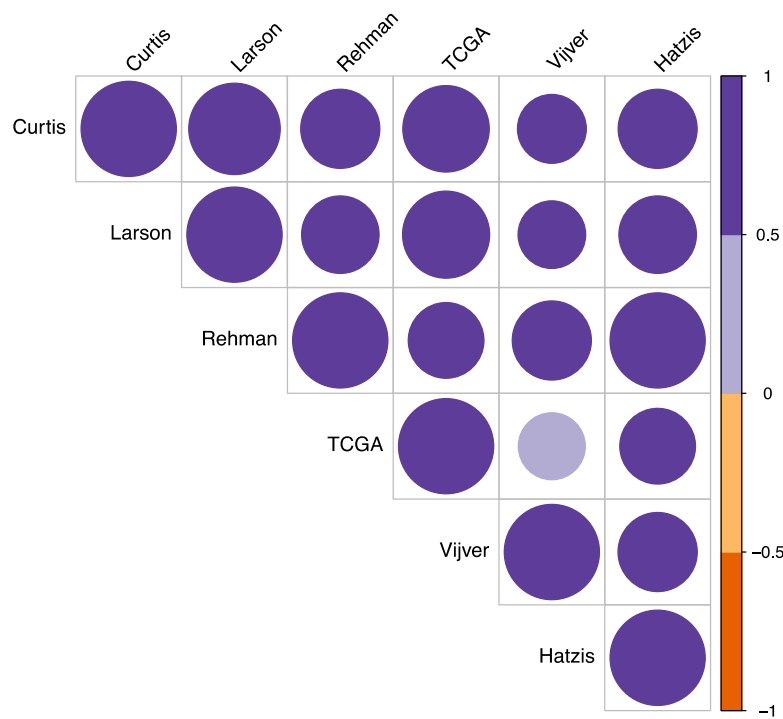


Figure S2. Distinguishing hereditary and sporadic breast cancer samples using RIPS. (A) Boxplot of inherited (both *BRCA1*- and *BRCA2*-related) cancers and nonfamilies sporadic cancer patient based on $RIPS_{BRCA1}$. The width of each box is proportional to the sample number. The p-value is a calculation of one-way analysis of variance (ANOVA). (B) Same as (A) but utilizing $RIPS_{RAD51}$.

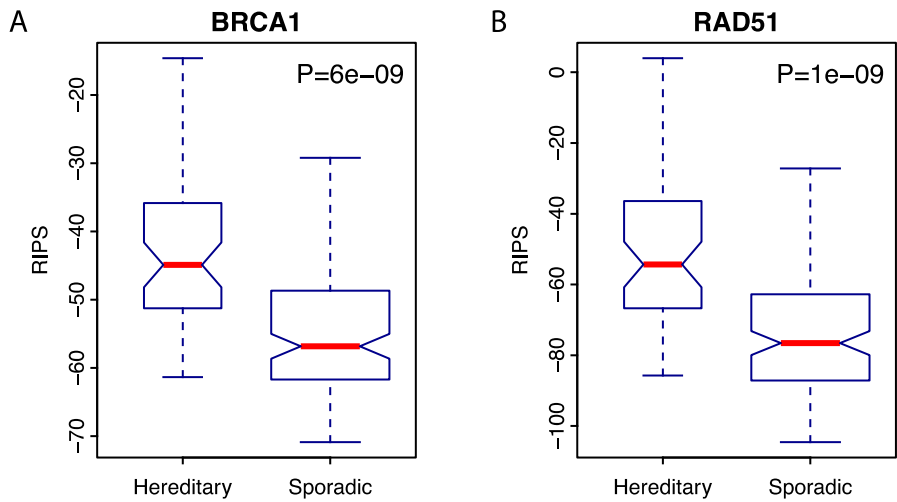


Figure S3. *TP53* mutation rate in different breast cancer datasets. (A) Using METABRIC breast cancer dataset. The grey dashed line represents the *TP53* mutation rate across all samples within the dataset. The x axis represents the average RIPS of each sliding window (here, the window size was 200). The Y axis demonstrates the *TP53* mutation rate for each window. (B) same as (A) but utilizing TCGA breast cancer dataset.

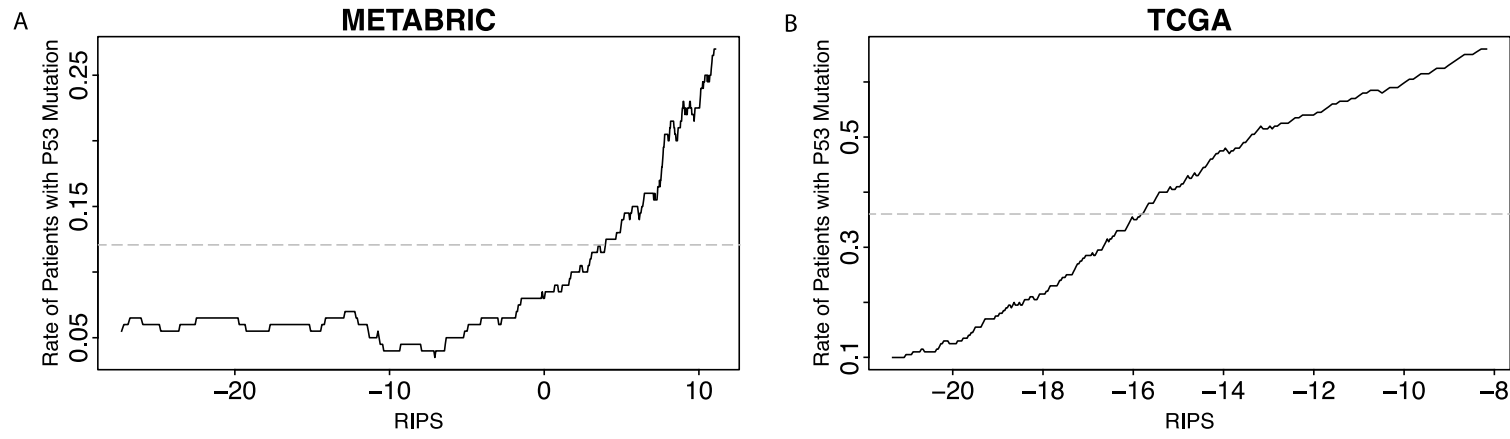


Figure S4. Prognosis of Hatzis validation dataset using RIPS_{RAD51}. (A) Kaplan-Meier plot of patients in Hatzis validation dataset. The green curve is patients with low RIPS and the red curve is patients with high RIPS. The difference between the two curve is significant (P=0.02). (B) Barplot of pCR% of samples within low, intermediate and high RIPS groups. The grey bar: the number of patients with RD status. The white bar: the number of patients with pCR status. The pCR% is shown above each bar. (C) ROC curve of patients with pCR status. Black curve: ROC for all patients in Hatzis dataset (AUC=0.653). Magenta curve: ROC for patients with ER+ status (AUC=0.66). Cyan curve: ROC for patients with ER- status (AUC=0.581). (D) Barplot comparison for prediction of pCR using either clinical information or clinical information plus RIPS_{RAD51}. Average of AUC calculated from 10-fold cross validation is shown above the bar. The vertical line in each bar represents the standard deviation of the 10 AUCs.

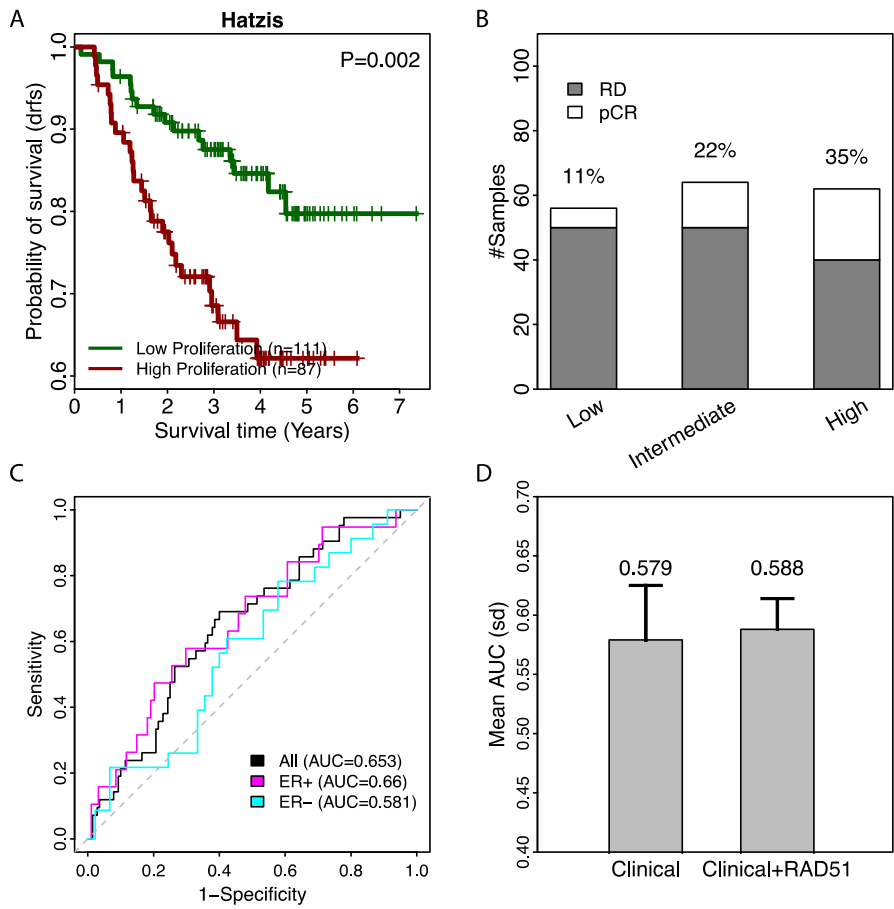


Figure S5. Mutation rate of significant genes in different cell proliferation patient groups. Red, yellow and blue bars represent low, intermediate and high RIPS tumor groups, respectively. Chi-squared test P values were presented with stars to show the difference between three groups. * indicated p value < 0.01 and ** indicated p value < 1e-10.

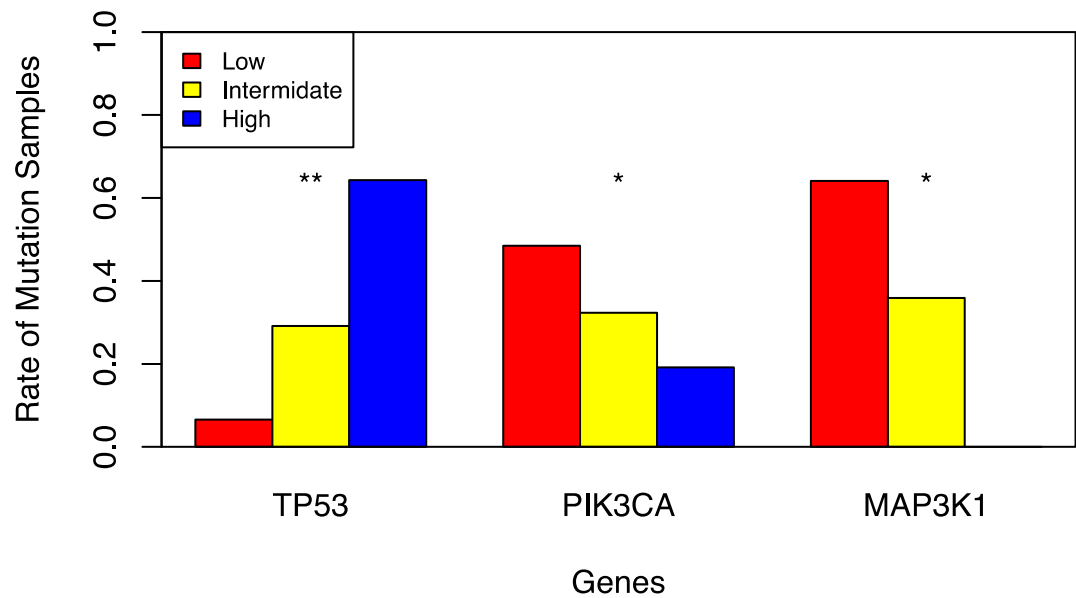


Figure S6. Gene expression of *BRCA1*, *BRCA2* and *RAD51* in different breast cancer mutations. (A) Boxplot of *BRCA1* expression in *BRCA1*-mutant, *BRCA2*-mutant and sporadic breast cancer samples. Each grey spot indicated a *BRCA1* expression score. The width of each box was proportional to the sample numbers. (B) Same as (A) for *BRCA2* expression data. (C) Same as (A) for *RAD51* expression data. T-test P values were presented using stars to show the difference between groups. ** indicated p value < 5e-04.

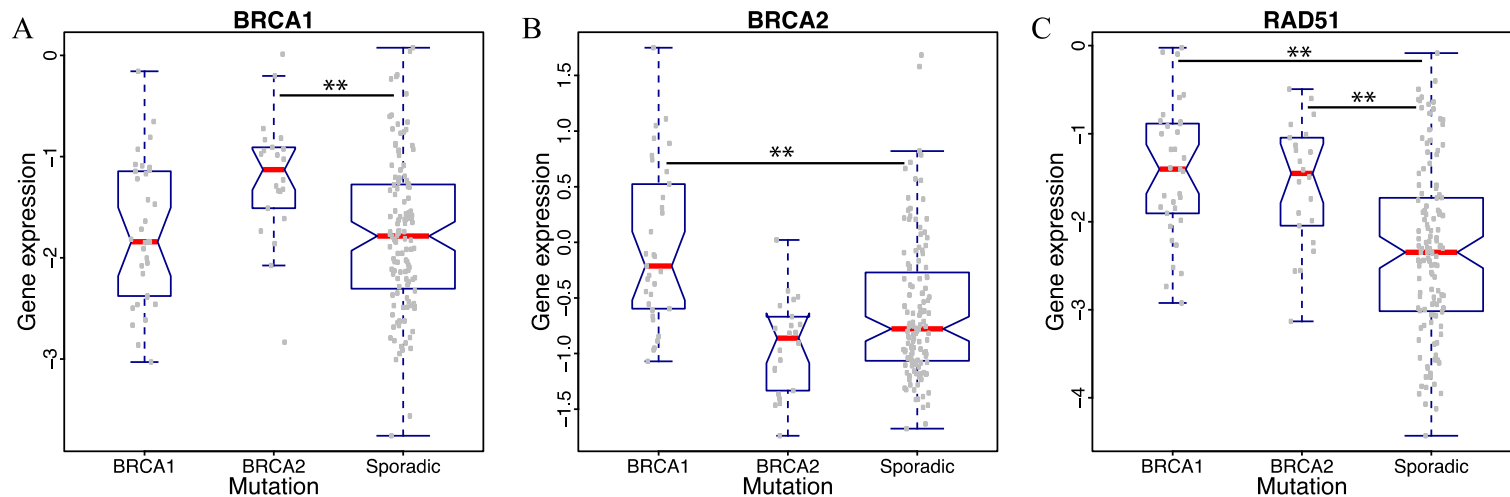


Figure S7. Gene expression of *XRCC2* in different cell proliferation patient groups. Boxplot of *XRCC2* expression in breast cancer samples with different cell proliferation. Each grey spot indicated a *XRCC2* expression score. The width of each box was proportional to the sample numbers. The p value was calculated by one-way analysis of variance (ANOVA).

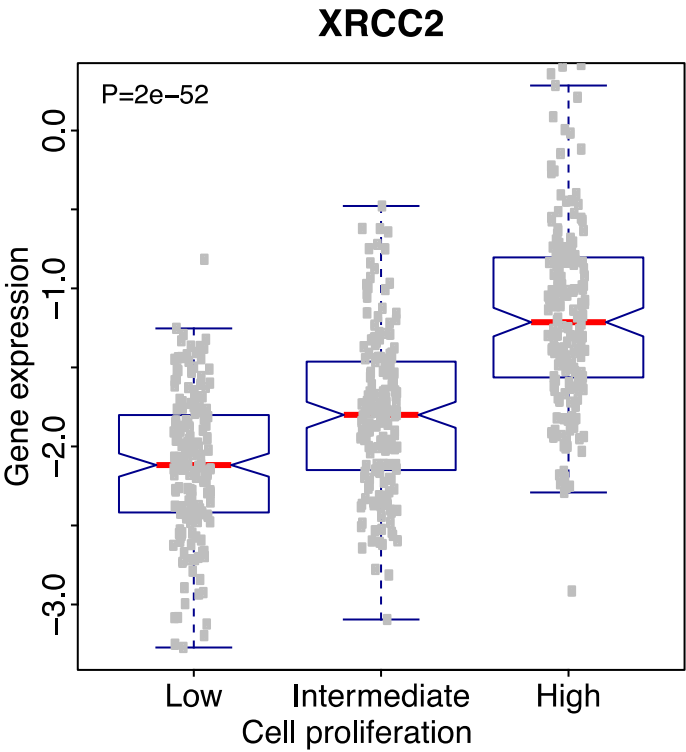


Figure S8. Prediction of *BRIT1* in METABRIC breast cancer dataset. Kaplan-Meier plot comparing survival of high to low cell proliferation patients. Patients with low RIPS (green curve) have significantly higher survival than patients with high RIPS (red curve).

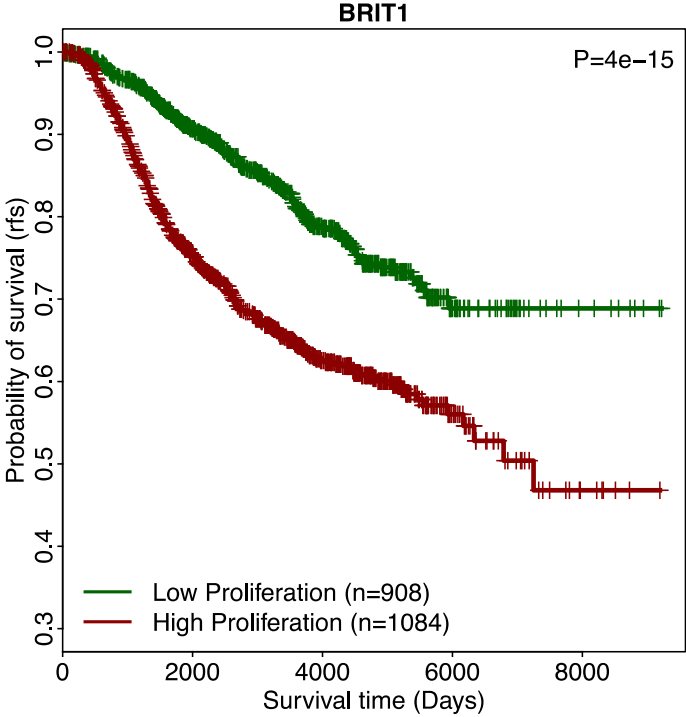


Table S1. Breast cancer datasets used in this analysis.

Source	PubMed ID	Accession ID	Number of samples	Outcome	Platform	Mutation Types
Larsen et.al	PMID:23704984	GSE40115	183		GPL15931 & two channel	Somatic & Germline
Ur-Rehman et.al	PMID:23756628	GSE47561	910	RFS	one channel	
Vijver et.al	PMID:12490681		260	OS	two channel	
Curtis et.al	PMID:22522925	EGAS00000000083	1992		one channel	
Hatzis et.al	PMID:21558518	GSE25066	310	DMFS	GPL96 & one channel	
Cancer Genome Atlas Network	PMID:23000897		972		two channel	

Table S3. Pathway enrichment of genes regulated by *BRCA1* and *RAD51*.

Gene Group	Term	Count	PValue	List.Total	Pop.Hits	Pop.Total	Fold.Enrichment	Bonferroni	Benjamini	FDR
BRCA1_up	Androgen and estrogen metabolism	4	2.70E-02	91	37	5085	6.04	9.51E-01	9.51E-01	2.66E+01
	p53 signaling pathway	5	3.16E-02	91	68	5085	4.11	9.71E-01	8.29E-01	3.05E+01
	Steroid hormone biosynthesis	4	4.71E-02	91	46	5085	4.86	9.95E-01	8.30E-01	4.20E+01
	Metabolism of xenobiotics by cytochrome P450	4	8.93E-02	91	60	5085	3.73	1.00E+00	9.24E-01	6.52E+01
	Systemic lupus erythematosus	5	9.75E-02	91	99	5085	2.82	1.00E+00	8.95E-01	6.86E+01
BRCA1_down	DNA replication	14	2.71E-13	115	36	5085	17.20	3.36E-11	3.36E-11	3.13E-10
	Mismatch repair	9	1.85E-08	115	23	5085	17.30	2.29E-06	1.15E-06	2.13E-05
	Cell cycle	14	3.38E-06	115	125	5085	4.95	4.19E-04	1.40E-04	3.91E-03
	Base excision repair	8	9.47E-06	115	35	5085	10.11	1.17E-03	2.93E-04	1.09E-02
	Nucleotide excision repair	7	3.92E-04	115	44	5085	7.03	4.75E-02	9.68E-03	4.52E-01
	Pyrimidine metabolism	9	1.20E-03	115	95	5085	4.19	1.38E-01	2.44E-02	1.37E+00
	Pathogenic Escherichia coli infection	6	8.59E-03	115	57	5085	4.65	6.57E-01	1.42E-01	9.48E+00
	Homologous recombination	4	2.39E-02	115	28	5085	6.32	9.50E-01	3.13E-01	2.44E+01
	Folate biosynthesis	3	2.40E-02	115	11	5085	12.06	9.51E-01	2.85E-01	2.45E+01
	Terpenoid backbone biosynthesis	3	4.33E-02	115	15	5085	8.84	9.96E-01	4.22E-01	4.00E+01
	Gap junction	6	4.86E-02	115	89	5085	2.98	9.98E-01	4.30E-01	4.38E+01
	One carbon pool by folate	3	4.87E-02	115	16	5085	8.29	9.98E-01	4.03E-01	4.38E+01
	Primary bile acid biosynthesis	3	4.87E-02	115	16	5085	8.29	9.98E-01	4.03E-01	4.38E+01
	PPAR signaling pathway	5	6.82E-02	115	69	5085	3.20	1.00E+00	4.90E-01	5.58E+01
	Steroid hormone biosynthesis	4	8.32E-02	115	46	5085	3.84	1.00E+00	5.37E-01	6.33E+01
RAD51_up	Metabolism of xenobiotics by cytochrome P450	5	2.11E-02	91	60	5085	4.66	8.91E-01	8.91E-01	2.12E+01
	Drug metabolism	5	2.35E-02	91	62	5085	4.51	9.15E-01	7.09E-01	2.33E+01
	Steroid hormone biosynthesis	4	4.71E-02	91	46	5085	4.86	9.93E-01	8.12E-01	4.17E+01
	Focal adhesion	8	6.41E-02	91	201	5085	2.22	9.99E-01	8.22E-01	5.23E+01
	Arginine and proline metabolism	4	6.67E-02	91	53	5085	4.22	9.99E-01	7.62E-01	5.38E+01
	Axon guidance	6	7.75E-02	91	129	5085	2.60	1.00E+00	7.53E-01	5.94E+01
	Systemic lupus erythematosus	5	9.75E-02	91	99	5085	2.82	1.00E+00	7.82E-01	6.82E+01
RAD51_down	DNA replication	15	5.37E-15	111	36	5085	19.09	6.02E-13	6.02E-13	6.05E-12
	Cell cycle	17	6.76E-09	111	125	5085	6.23	7.64E-07	3.82E-07	7.67E-06
	Mismatch repair	8	3.37E-07	111	23	5085	15.93	3.81E-05	1.27E-05	3.83E-04
	Nucleotide excision repair	7	3.23E-04	111	44	5085	7.29	3.59E-02	9.09E-03	3.66E-01
	NOD-like receptor signaling pathway	7	2.05E-03	111	62	5085	5.17	2.07E-01	4.54E-02	2.31E+00
	Pyrimidine metabolism	8	4.23E-03	111	95	5085	3.86	3.81E-01	7.67E-02	4.70E+00
	Base excision repair	5	6.48E-03	111	35	5085	6.54	5.20E-01	9.96E-02	7.11E+00
	Oocyte meiosis	8	9.35E-03	111	110	5085	3.33	6.54E-01	1.24E-01	1.01E+01
	Progesterone-mediated oocyte maturation	7	1.03E-02	111	86	5085	3.73	6.91E-01	1.22E-01	1.11E+01
	p53 signaling pathway	6	1.53E-02	111	68	5085	4.04	8.25E-01	1.60E-01	1.61E+01
	PPAR signaling pathway	6	1.62E-02	111	69	5085	3.98	8.42E-01	1.55E-01	1.69E+01
	Homologous recombination	4	2.18E-02	111	28	5085	6.54	9.17E-01	1.87E-01	2.21E+01
	Pathogenic Escherichia coli infection	5	3.41E-02	111	57	5085	4.02	9.80E-01	2.60E-01	3.25E+01
	Epithelial cell signaling in Helicobacter pylori infection	5	5.88E-02	111	68	5085	3.37	9.99E-01	3.87E-01	4.97E+01

Table S4. HR-related genes and their DNA methylation level in RIPS-low, -intermediate and -high groups.

CpG sites	Closest_TSS_gene_name	Distance_closest_TSS	Mean_Meth_RIPS_Low	Mean_Meth_RIPS_Inter	Mean_Meth_RIPS_High	T-test_Pvalue_Low_High	T-test_Pvalue_Low_Inter	T-test_Pvalue_Inter_High
cg26458617	BRCA2	-403	0.49	0.44	0.30	3.45E-07	1.86E-01	2.49E-04
cg12836863	BRCA2	-593	0.92	0.90	0.80	1.27E-06	7.25E-02	4.28E-05
cg01430588	RAD51C	-195	0.61	0.59	0.48	1.30E-06	4.92E-01	6.10E-05
cg25339112	RAD51C	-202	0.74	0.72	0.63	2.08E-06	1.75E-01	2.71E-04
cg00453717	POLD3	41001	0.92	0.91	0.89	6.55E-06	8.43E-03	1.29E-02
cg13042487	XRCC3	430	0.21	0.23	0.10	7.28E-06	5.38E-01	1.53E-05
cg16270633	BLM	-3748	0.79	0.77	0.72	8.50E-06	1.35E-01	1.50E-03
cg20864326	XRCC2	1828	0.84	0.82	0.74	8.89E-06	6.64E-02	1.59E-03
cg01605516	XRCC2	3527	0.63	0.66	0.78	9.93E-06	2.73E-01	1.42E-03
cg02194129	XRCC3	10510	0.82	0.78	0.74	1.37E-05	2.84E-02	4.29E-02
cg08196608	RPA1	74568	0.89	0.87	0.81	1.38E-05	8.03E-03	3.13E-03
cg14172849	XRCC3	10564	0.80	0.77	0.73	1.78E-05	1.62E-02	6.02E-02
cg01663440	RAD52	2735	0.83	0.82	0.76	2.43E-05	2.07E-01	5.11E-03
cg26693701	RPA1	88536	0.92	0.92	0.89	2.92E-05	5.47E-01	2.34E-04
cg18211249	RPA1	77187	0.86	0.85	0.77	3.21E-05	5.78E-01	2.32E-04
cg26377508	RPA1	77354	0.87	0.87	0.79	3.45E-05	8.20E-01	8.55E-05
cg17258364	RPA1	77270	0.92	0.91	0.84	3.88E-05	8.35E-02	1.08E-03
cg06788856	RPA1	88697	0.95	0.95	0.93	4.22E-05	2.43E-01	6.49E-04
cg22635541	RPA1	75301	0.64	0.66	0.51	5.28E-05	5.44E-01	1.48E-05
cg17724365	NBN	-109	0.06	0.05	0.05	7.17E-05	8.56E-02	1.11E-02
cg09741519	RPA1	77503	0.86	0.87	0.75	1.28E-04	7.12E-01	7.25E-05
cg01173485	POLD3	-735	0.88	0.87	0.82	1.67E-04	1.76E-01	4.84E-03
cg07918226	RAD54B	-17	0.06	0.05	0.05	1.89E-04	6.30E-01	1.37E-03
cg03719475	RAD52	5940	0.96	0.96	0.95	2.01E-04	3.05E-01	3.35E-03
cg23193616	XRCC3	4293	0.45	0.47	0.54	3.17E-04	5.39E-01	4.23E-03
cg16291854	POLD1	8157	0.94	0.93	0.93	5.57E-04	4.49E-01	3.05E-03
cg19197440	RPA1	77997	0.79	0.78	0.70	6.14E-04	6.59E-01	2.60E-03
cg02110529	RAD51C	-266	0.77	0.76	0.70	6.62E-04	6.99E-01	3.85E-03
cg13257157	RAD54B	329	0.03	0.03	0.03	6.62E-04	8.63E-01	2.80E-03
cg10443159	RPA3	-81020	0.75	0.79	0.83	6.96E-04	1.37E-01	6.38E-02
cg25858132	RPA1	75399	0.50	0.55	0.36	9.32E-04	2.37E-01	4.35E-05
cg09706243	POLD4	-582	0.13	0.16	0.18	1.00E-03	1.83E-02	1.59E-01
cg24362026	SHFM1	-29426	0.58	0.60	0.69	1.08E-03	6.63E-01	6.02E-03
cg10643407	RAD50	-92	0.05	0.05	0.05	1.30E-03	6.99E-03	2.31E-01
cg19146577	POLD2	-8661	0.75	0.78	0.80	1.35E-03	2.80E-02	2.06E-01
cg26323655	RAD54B	495	0.78	0.78	0.69	1.47E-03	9.44E-01	1.98E-03
cg01381637	POLD1	735	0.86	0.86	0.85	1.56E-03	1.68E-01	8.43E-02
cg26885816	POLD3	-616	0.84	0.83	0.77	1.70E-03	5.99E-01	7.62E-03
cg05488279	RPA1	86498	0.66	0.65	0.75	1.95E-03	6.03E-01	8.44E-04
cg27170268	XRCC3	10128	0.87	0.83	0.77	1.96E-03	1.05E-01	1.11E-01
cg25305774	RAD54B	-10677	0.51	0.49	0.64	2.39E-03	6.44E-01	1.04E-03
cg18433268	RAD54B	304	0.05	0.04	0.04	2.40E-03	5.09E-01	1.23E-02
cg27091703	RAD54B	-85	0.05	0.05	0.05	2.55E-03	4.97E-01	2.90E-02
cg23436779	RAD51C	-132	0.09	0.10	0.07	2.55E-03	7.54E-01	1.09E-03
cg18263559	TOP3B	6445	0.94	0.93	0.90	2.68E-03	3.49E-01	2.99E-02
cg26262057	MRE11A	2	0.04	0.04	0.04	2.72E-03	8.12E-01	4.56E-03
cg10104687	RAD52	221	0.65	0.67	0.70	3.49E-03	4.01E-01	3.76E-02
cg10347949	XRCC3	6075	0.88	0.88	0.91	3.51E-03	8.91E-01	1.03E-02
cg07997877	RPA1	77581	0.84	0.84	0.77	3.80E-03	7.21E-01	1.04E-02
cg18597188	XRCC3	3153	0.37	0.44	0.44	3.86E-03	2.92E-03	9.69E-01
cg04496177	RAD54B	589	0.08	0.08	0.07	4.46E-03	5.24E-01	2.90E-02
cg10656845	POLD2	3327	0.93	0.93	0.92	4.56E-03	7.54E-01	5.89E-03
cg23639841	RAD54B	-6	0.02	0.02	0.02	5.11E-03	3.90E-01	6.71E-02
ch.7.313144R	RPA3	-69223	0.07	0.07	0.06	5.65E-03	4.62E-01	4.35E-02
cg19366479	POLD1	-3940	0.74	0.77	0.80	6.62E-03	2.50E-01	1.47E-01
cg01086369	POLD1	-2219	0.99	0.99	0.99	6.89E-03	1.58E-01	2.97E-01
cg07154931	RAD54B	111	0.05	0.05	0.04	7.66E-03	3.54E-01	7.20E-02
cg15612927	RAD52	621	0.04	0.04	0.04	7.72E-03	4.65E-01	4.62E-03
cg23369670	XRCC3	9879	0.78	0.74	0.70	7.79E-03	1.40E-01	2.29E-01
cg08157964	BRCA2	245	0.01	0.01	0.01	8.59E-03	7.36E-03	8.08E-01
cg12385370	RAD52	5961	0.95	0.94	0.94	9.17E-03	2.74E-02	6.37E-01
cg04600982	RPA1	58736	0.95	0.95	0.95	9.21E-03	1.23E-01	2.88E-01
cg27017251	TOP3B	2594	0.88	0.90	0.91	9.48E-03	7.67E-02	3.79E-01
cg12383621	RPA3	22627	0.82	0.85	0.85	9.98E-03	2.25E-02	6.60E-01

Table S4. Survival analysis results for 11 significant HR genes using their expression level.

Gene	Survival analysis		Cox regression model	
	P vaule	Q value	P vaule	Q value
BLM	6.25E-14	1.62E-12	1.11E-16	2.89E-15
RAD54L	3.84E-11	4.99E-10	1.49E-13	1.94E-12
POLD1	1.87E-08	1.62E-07	3.91E-11	3.39E-10
RAD54B	7.26E-08	4.72E-07	7.52E-08	4.89E-07
EME1	1.49E-06	7.73E-06	9.22E-07	4.79E-06
XRCC3	7.84E-06	3.40E-05	1.37E-06	5.94E-06
RPA2	1.15E-04	4.27E-04	1.94E-04	6.31E-04
POLD3	6.21E-04	2.02E-03	8.45E-04	2.20E-03
TOP3B	3.52E-03	1.02E-02	3.55E-04	1.03E-03
RAD51	4.33E-03	1.13E-02	1.52E-04	5.63E-04
MUS81	1.78E-02	3.85E-02	3.99E-03	9.44E-03