Supplementary Online Content

Evans JR, Zhao SG, Chang SL, et al. Patient-level DNA damage and repair pathway profiles and prognosis after prostatectomy for high-risk prostate cancer. *JAMA Oncol*. Published online January 7, 2015. doi:10.1001/jamaoncol.2015.4955.

This supplementary material has been provided by the authors to give readers additional information about their work.

Title: Patient-level DNA Damage and Repair Pathway Profiles Are Prognostic after Prostatectomy for High-Risk Prostate Cancer

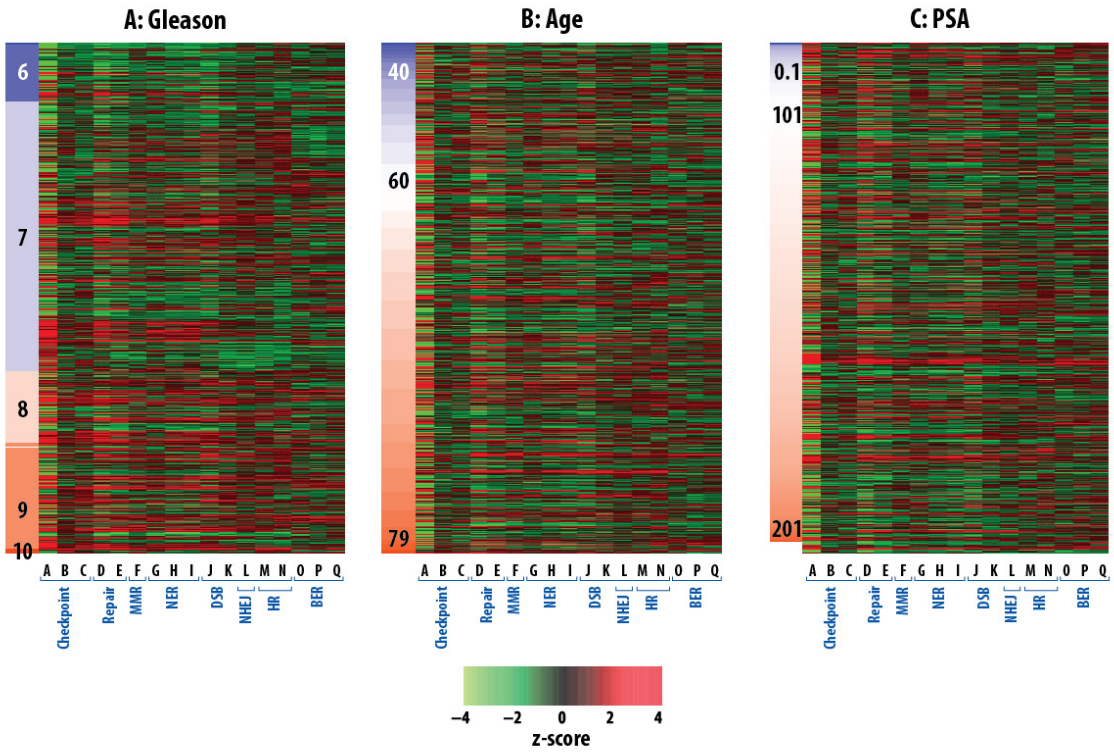
Evans, Zhao, et al.

Supplemental Contents

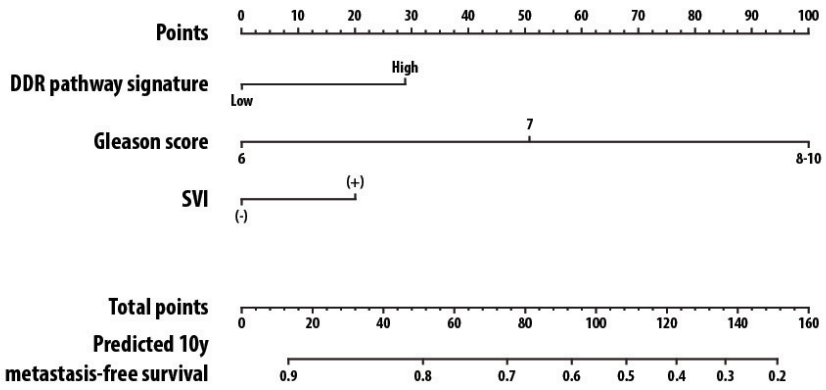
Supplemental Materials and Methods

DDR GSEA gene set member gene mutation analysis

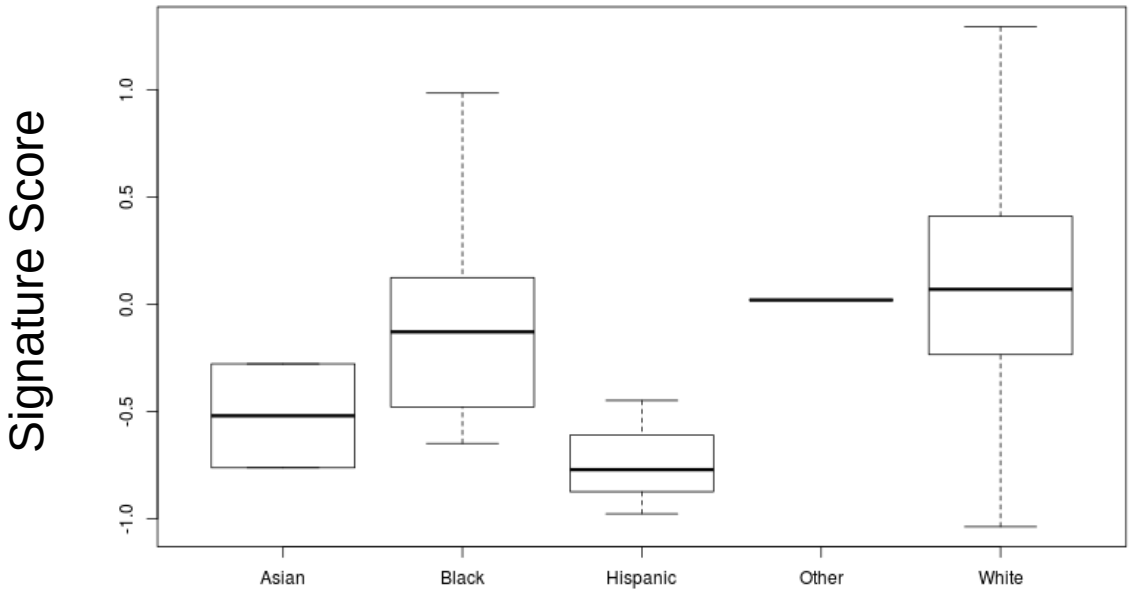
Primary non-metastatic prostate genomic profiles from published studies and The Cancer Genome Atlas unpublished data were analyzed using the R API from cBioPortal (cbioportal.org) with a custom script (Supplemental Materials), which was run on 6/5/2015 using R 3.1.2. The primary cases for the Broad/Cornell Cell 2013 and MSKCC Cancer Cell 2010 cohorts were retrieved manually from cBioPortal, and stored in the input files listed below. The DDR GSEA gene sets were retrieved manually from the GSEA mSigDB web site (www.broadinstitute.org/gsea/msigdb/index.jsp), and combined into the single input file below. The following genes were removed from the analysis as cBioPortal indicated they were not valid gene symbols for analysis: LOC731751, LOC389901, LOC648152, LOC651610, LOC651921, LOC652672, LOC652857, and LOC652826.



eFigure 1: DDR pathway expression heatmaps ordered by clinical variables Gleason (A), Age (B), and PSA (C). Color gradient sidebars at left are labeled with the maximum, average, and minimum values for each clinical variable. X-axis is labeled with DDR gene sets and pathways as described in eTable 1 in the Supplement.



eFigure 2: Nomogram for predicting 10 year metastasis-free survival using the significant factors from the multivariate Cox model in the pooled validation cohorts.



eFigure 3: Mean DDR pathway signature score by ethnicity. n=313.

eTable 1: Gene sets, heatmap positions, pathway groupings, and correlation between each DDR gene set and each clinical or molecular variable across all patients by Spearman's correlation.

Gene Set	Pos.		Spearman's rho statistic						p-value					
		Pathway	AR	ARp*	ERG	Age	Gleason	PSA	ERG	AR	ARp*	Age	Gleason	PSA
REACTOME_P53_INDEPENDENT_G1_S_DNA_DAMAGE_CHECKPOINT	A	Checkpoint	0.54	0.82	0.21	-0.07	0.09	-0.07	<1E-10	<1E-10	<1E-10	0.02	0.0023	0.02
REACTOME_G2_M_DNA_DAMAGE_CHECKPOINT	B	Checkpoint	0.19	0.35	0.03	0.02	0.14	-0.0005	3.37E-01	7.79E-10	<1E-10	0.58	2.06E-06	0.99
DNA_DAMAGE_CHECKPOINT	C	Checkpoint	0.46	0.53	0.05	-0.03	0.20	0.02	9.61E-02	<1E-10	<1E-10	0.33	<1E-10	0.61
DNA_REPAIR	D	Repair	0.48	0.78	0.25	-0.005	0.11	-0.07	<1E-10	<1E-10	<1E-10	0.88	2.14E-04	0.02
REACTOME_DNA_REPAIR	E	Repair	0.44	0.68	0.22	0.03	0.16	-0.02	<1E-10	<1E-10	<1E-10	0.35	1.69E-07	0.59
KEGG_MISMATCH_REPAIR	F	MMR	0.18	0.47	0.16	0.10	0.17	0.02	1.22E-07	3.17E-09	<1E-10	9.49E-04	4.21E-08	0.56
NUCLEOTIDE_EXCISION_REPAIR	G	NER	0.31	0.57	0.22	-0.002	0.03	-0.005	<1E-10	<1E-10	<1E-10	0.93	0.26	0.88
REACTOME_NUCLEOTIDE_EXCISION_REPAIR	H	NER	0.40	0.62	0.17	0.02	0.10	-0.02	8.95E-09	<1E-10	<1E-10	0.57	6.48E-04	0.48
KEGG_NUCLEOTIDE_EXCISION_REPAIR	I	NER	0.41	0.67	0.17	0.03	0.11	-0.04	1.36E-08	<1E-10	<1E-10	0.38	3.61E-04	0.24
DOUBLE_STRAND_BREAK_REPAIR	J	DSB	0.36	0.68	0.18	0.03	0.17	-0.02	2.21E-09	<1E-10	<1E-10	0.29	3.85E-08	0.58
REACTOME_DOUBLE_STRAND_BREAK_REPAIR	K	DSB	0.28	0.49	0.12	0.08	0.18	0.01	1.26E-04	<1E-10	<1E-10	0.01	3.03E-09	0.87
KEGG_NON_HOMOLOGOUS_END_JOINING	L	DSB, NHEJ	0.20	0.37	0.15	0.08	0.13	0.01	3.84E-07	<1E-10	<1E-10	0.01	1.35E-05	0.79
REACTOME_HOMOLOGOUS_RECOMBINATION_REPAIR_OF_REPLICATION_INDEPENDENT_DOUBLE_STRAND_BREAKS	M	HR1, DSB	0.12	0.27	0.04	0.12	0.16	0.03	1.47E-01	1.30E-04	<1E-10	6.79E-05	1.50E-07	0.31
KEGG_HOMOLOGOUS_RECOMBINATION	N	HR2, DSB	-0.22	-0.11	0.01	0.24	0.16	0.10	6.84E-01	<1E-10	2.02E-04	5.73E-16	6.78E-08	8.10E-04
BASE_EXCISION_REPAIR	O	BER	0.30	0.33	0.20	0.06	0.16	0.01	<1E-10	<1E-10	<1E-10	0.04	1.16E-07	0.84
REACTOME_BASE_EXCISION_REPAIR	P	BER	0.21	0.06	0.10	0.01	0.11	0.02	1.39E-03	<1E-10	0.040	0.65	3.05E-04	0.49

KEGG_BASE_EXCISION_REPAIR	Q	BER	0.18	0.11	0.09	0.02	0.15	0.05	1.87E-03	1.74E-09	3.87E-04	0.49	4.09E-07	0.11

Pos.=heatmap position. MMR=Mismatch Repair. NER=Nucleotide Excision Repair. DSB=double-strand break repair. NHEJ=Non-Homologous End-Joining. HR=Homologous Recombination. BER=Base Excision Repair. ARp*=AR pathway activity by GSEA.

eTable 2: Cohort clinical characteristics					
		Mayo Discovery (n=545)	**Mayo Validation (n=232)**	**Cleveland Clinic (n =183)**	**Thomas Jefferson University (n = 130)**
Age at diagnosis (Years ± SD)		65.3 ± 6.4	63.1 ± 7.4	61.6 ± 6.3	60.0 ± 7.0
Follow-up (Months ± SD)		160.7 ± 56.2	80.6 ± 30.1	116.6 ± 50.1	103.9 ± 58.9
Pre-operative PSA					
	<10	282 (52%)	126 (54%)	127 (69%)	83 (64%)
	10 to 20	117 (22%)	62 (27%)	41 (23%)	25 (19%)
	>20	131 (24%)	44 (19%)	12 (7%)	15 (11%)
	Not available	15 (3%)	0 (0%)	3 (1%)	7 (5%)
Gleason score					
	6	60 (11%)	17 (7%)	25 (17%)	17 (13%)
	7	271 (49%)	117 (50%)	113 (62%)	74 (57%)
	8	68 (13%)	39 (17%)	23 (13%)	22 (17%)
	9	134 (24%)	57 (25%)	22 (12%)	13 (10%)
	10	9 (2%)	1 (1%)	0 (0%)	2 (1.5%)
	Not available	3 (1%)	1 (1%)	0 (0%)	2 (1.5%)
Tumor stage					
	I	0 (0%)	0 (0%)	0 (0%)	0 (0%)
	II	219 (40%)	97 (42%)	0 (0%)	10 (8%)
	III	253 (46%)	102 (44%)	0 (0%)	113 (87%)
	IV	0 (0%)	0 (0%)	0 (0%)	7 (5%)
	Not available	73 (13%)	33 (14%)	183 (100%)	0 (0%)
Extracapsular extension					
	Negative	272 (50%)	136 (59%)	51 (28%)	23 (18%)
	Positive	273 (50%)	96 (41%)	132 (72%)	106 (81%)
	Not available	0 (0%)	0 (0%)	0 (0%)	1 (1%)
Seminal vesicle invasion					
	Negative	369 (68%)	149 (64%)	152 (83%)	82 (63%)
	Positive	176 (32%)	83 (36%)	31 (17%)	48 (37%)
Lymph node invasion					
	Negative	472 (87%)	199 (86%)	183 (100%)	128 (98%)
	Positive	73 (13%)	33 (14%)	0 (0%)	2 (2%)
Surgical margin status					
	Negative	279 (51%)	99 (43%)	92 (50%)	31 (24%)
	Positive	266 (49%)	133 (57%)	91 (50%)	99 (76%)
Metastatic progression					
	No	333 (61%)	157 (68%)	134 (73%)	120 (92%)
	Yes	212 (39%)	75 (32%)	49 (27%)	10 (8%)

eTable 3: Mutation frequencies of the genes in the DDR gene sets

The mutation frequencies in percent (%) for each gene in each of the GSEA DDR gene sets used was extracted from cBioPortal using a custom R script for the cBioPortal R-API as described in detail in the Supplemental Material and Methods in the Supplement.

BASE_EXCISION_REPAIR												
MSH6	MSH3	MSH2	POLG	TP53	MPG	ERCC6	OGG1	HMGB1P10	HMGB1	HMGB2	UNG	SMUG1
0.5	0.8	0.2	0.3	6.8	0.2	0.8	0.2	0.0	0.0	0.0	0.0	0.0
TDG	NTHL1	CCNO	APEX1									
0.0	0.0	0.0	0.0									
DNA_DAMAGE_CHECKPOINT												
NBN	HUS1	RINT1	PML	CHEK2	BRSK1	FOXN3	NAE1	PCBP4	ATRIP	RAD17	ZAK	RAD9A
0.3	0.2	0.2	0.2	0.8	0.6	0.2	0.2	0.3	0.2	0.2	0.0	0.0
CCNA2	CHEK1	ATR	NEK11	RAD1	TRIAP1	GML						
0.0	0.0	0.0	0.0	0.0	0.0	0.0						
DNA_REPAIR												
MMS19	XRCC6	SETX	IGHMBP2	MUTYH	RAD21	POLL	POLI	POLG	LIG1	POLE	LIG3	LIG4
0.2	0.5	0.6	0.3	0.2	0.2	0.2	0.6	0.3	0.8	0.8	0.2	0.3
HUS1	XAB2	POLQ	RECQL4	RECQL5	BRCA2	RAD54L	ATM	BRCA1	VCP	POLD1	RAD54B	ABL1
0.2	0.2	0.6	0.6	0.2	1.6	0.2	3.0	0.6	0.3	0.2	0.5	0.5
FANCC	REV1	TP53	PRKCG	RAD52	RAD51	RECQL	XPC	RUVBL2	WRNIP1	POLA1	RPA1	SUMO1
0.2	0.5	6.8	0.2	0.3	0.2	0.2	0.6	0.3	0.2	0.3	0.2	0.2
MSH6	UPF1	MSH3	MSH2	MSH5	ATRX	MPG	ATXN3	CSNK1E	ALKBH1	OGG1	XRCC4	RAD51C
0.5	0.5	0.8	0.2	0.2	0.6	0.2	0.2	0.2	0.3	0.2	0.0	0.0
POLH	RAD9A	RAD1	HMGB1	RAD23B	HMGB2	BLM	MRE11A	UBE2V1	UBE2V2	POLE2	RAD51B	NTHL1
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
SMC1A	PMS2P1	CCNO	UBE2A	DDB1	GTF2H4	APTX	UBE2B	RAD50	XRCC6BP1	GTF2H1	RBBP8	UBE2N
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
SMUG1	ERCC8	APEX1	ERCC1	EXO1	CEBPG	TNP1	SOD1	TP73	MNAT1	CSNK1D	KAT5	PNKP

0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.2	0.2
RFC3	RPAIN	RAD17	RAD23A	PARP1	NBN	UVRAG	MLH1	FANCG	FANCA	ERCC5	ERCC6	ERCC3
0.2	0.2	0.2	0.2	0.3	0.3	0.2	0.2	0.2	0.5	0.3	0.8	0.3
ERCC4	FEN1	ERCC2	XRCC3	XRCC2	CDKN2D	PMS2	PMS1	CIB1	ASF1A	HMGB1P10	TREX2	ATR
0.2	0.2	0.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
TDG	PARP3	BTG2	GADD45G	DDB2	NHEJ1	GADD45A	UNG					
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0					
DOUBLE_STRAND_BREAK_REPAIR												
NBN	XRCC6	POLA1	LIG4	RAD52	KAT5	BRCA1	RAD51	SETX	RAD21	VCP	RAD54B	ERCC4
0.3	0.5	0.3	0.3	0.3	0.2	0.6	0.2	0.6	0.2	0.3	0.5	0.2
XRCC6BP1	UBE2N	NHEJ1	CIB1	FEN1	XRCC4	MRE11A	UBE2V2	SOD1	RAD50			
0.0	0.0	0.0	0.0	0.2	0.0	0.0	0.0	0.0	0.0			
KEGG_BASE_EXCISION_REPAIR												
MPG	XRCC1	MBD4	OGG1	POLD3	PARP4	PARP2	POLB	POLL	POLD1	POLE	NEIL3	APEX2
0.2	0.2	0.2	0.2	0.2	0.3	0.2	0.2	0.2	0.2	0.8	0.2	0.2
SMUG1	POLE4	HMGB1	POLE3	POLD4	UNG	PCNA	NEIL1	POLE2	PARP3	APEX1	POLD2	TDG
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
LIG3	FEN1	LIG1	MUTYH	PARP1	NEIL2	HMGB1P1	NTHL1	HMGB1P40				
0.2	0.2	0.8	0.2	0.3	0.0	0.0	0.0	0.0				
KEGG_HOMOLOGOUS_RECOMBINATION												
RAD54L	NBN	RAD52	RAD51D	RPA1	RAD51	POLD3	RPA2	RPA4	BRCA2	POLD1	MUS81	TOP3A
0.2	0.3	0.3	0.2	0.2	0.2	0.2	0.2	0.3	1.6	0.2	0.2	0.3
POLD4	BLM	EME1	RAD50	RPA3	RAD51C	MRE11A	POLD2	SSBP1	RAD54B	SHFM1	TOP3B	XRCC3
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.5	0.2	0.3	0.0
RAD51B	XRCC2											
0.0	0.0											
KEGG_MISMATCH_REPAIR												
MLH3	POLD1	MLH1	RFC1	MSH2	RFC3	RFC2	MSH3	LIG1	RPA1	MSH6	POLD3	RPA2
0.6	0.2	0.2	0.3	0.2	0.2	0.2	0.8	0.8	0.2	0.5	0.2	0.2

RPA3	PCNA	SSBP1	EXO1	RPA4	POLD2	POLD4	PMS2	RFC4	RFC5			
0.0	0.0	0.0	0.0	0.3	0.0	0.0	0.0	0.0	0.0			
KEGG_NON_HOMOLOGOUS_END_JOINING												
POLL	POLM	LIG4	FEN1	DNTT	XRCC5	XRCC6	PRKDC	DCLRE1C	XRCC4	MRE11A	LOC731751	NHEJ1
0.2	0.2	0.3	0.2	0.3	0.3	0.5	1.4	0.3	0.0	0.0	0.0	0.0
RAD50												
0.0												
KEGG_NUCLEOTIDE_EXCISION_REPAIR												
ERCC4	ERCC3	ERCC6	ERCC5	ERCC2	CETN2	RPA1	POLD3	RPA2	RAD23A	RPA4	POLD1	POLE
0.2	0.3	0.8	0.3	0.5	0.2	0.2	0.2	0.2	0.2	0.3	0.2	0.8
CUL4A	CCNH	GTF2H3	MNAT1	POLE4	POLE3	GTF2H5	POLD4	RFC4	RFC5	RAD23B	RBX1	DDB2
0.2	0.2	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
GTF2H2	GTF2H1	CDK7	CUL4B	ERCC8	GTF2H4	RFC1	RFC3	RFC2	XPC	XPA	LIG1	RPA3
0.0	0.0	0.0	0.0	0.0	0.0	0.3	0.2	0.2	0.6	0.5	0.8	0.0
PCNA	DDB1	POLE2	ERCC1	POLD2								
0.0	0.0	0.0	0.0	0.0								
NUCLEOTIDE_EXCISION_REPAIR												
POLL	MMS19	RAD23A	TP53	BRCA2	LIG4	XAB2	ATXN3	ERCC5	ERCC6	XPC	ERCC3	ERCC4
0.2	0.2	0.2	6.8	1.6	0.3	0.2	0.2	0.3	0.8	0.6	0.3	0.2
ERCC1	ERCC2	RAD23B	DDB1	ERCC8	DDB2	NTHL1						
0.0	0.5	0.0	0.0	0.0	0.0	0.0						
REACTOME_BASE_EXCISION_REPAIR												
POLD3	FEN1	LIG1	LIG3	MPG	MUTYH	OGG1	POLB	POLD1	XRCC1	MBD4	CCNO	SMUG1
0.2	0.2	0.8	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.0	0.0
APEX1	NTHL1	PCNA	POLD2	POLD4	TDG							
0.0	0.0	0.0	0.0	0.0	0.0							
REACTOME_DNA_REPAIR												
MAD2L2	POLD3	ALKBH2	ERCC2	ERCC3	ERCC4	ERCC5	ERCC6	FANCA	FANCC	FANCD2	FANCB	FANCF

0.2	0.2	0.3	0.5	0.3	0.2	0.3	0.8	0.5	0.2	0.2	0.3	0.2
LIG3	LIG4	MPG	MUTYH	NBN	ATM	OGG1	REV1	POLB	POLD1	POLE	POLR2A	POLR2B
0.2	0.3	0.2	0.2	0.3	3.0	0.2	0.5	0.2	0.2	0.8	0.5	0.5
RAD52	REV3L	RFC2	RFC3	RPA1	RPA2	RPS27A	BRCA1	BRCA2	TCEA1	TP53BP1	USP1	XPA
0.3	0.3	0.2	0.2	0.2	0.2	0.3	0.6	1.6	0.2	1.1	0.3	0.5
MBD4	CCNH	C19orf40	MDC1	RAD50	CDK7	CCNO	ERCC8	DDB1	DDB2	ERCC1	FANCE	ALKBH3
0.2	0.2	0.2	0.6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
APEX1	LOC389901	MGMT	MNAT1	MRE11A	NTHL1	PCNA	POLD2	POLE2	POLH	POLR2C	POLR2D	POLR2E
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
ATR	FANCL	POLD4	RAD23B	RFC4	RFC5	RPA3	LOC648152	LOC651610	LOC651921	LOC652672	LOC652857	GTF2H2B
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
FANCG	FEN1	XRCC6	ZBTB32	GTF2H3	LIG1	POLR2K	TDP1	PRKDC	XAB2	FANCM	RAD51	XPC
0.2	0.2	0.5	0.5	0.2	0.8	0.2	0.2	1.4	0.2	0.3	0.2	0.6
XRCC1	XRCC5	PALB2	C17orf70	BRIP1	SMUG1	UBE2T	GTF2H1	GTF2H2	GTF2H4	H2AFX	POLR2F	POLR2G
0.2	0.3	0.6	0.3	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
POLR2H	POLR2I	POLR2J	POLR2L	TDG	RPS27AP11	UBA52	XRCC4					
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0					
REACTOME_DOUBLE_STRAND_BREAK_REPAIR												
XRCC6	LIG1	LIG4	NBN	ATM	TDP1	PRKDC	RAD51	RAD52	RPA1	RPA2	BRCA1	BRCA2
0.5	0.8	0.3	0.3	3.0	0.2	1.4	0.2	0.3	0.2	0.2	0.6	1.6
LOC389901	MRE11A	RPA3	LOC651610	XRCC4	TP53BP1	XRCC5	BRIP1	MDC1	RAD50	H2AFX		
0.0	0.0	0.0	0.0	0.0	1.1	0.3	0.2	0.6	0.0	0.0		
REACTOME_G2_M_DNA_DAMAGE_CHECKPOINT												
CHEK2	ATM	WEE1	ATRIP	CHEK1	ATR	LOC648152	LOC651610	LOC651921	CCNB1	CDK1	CDC25C	
0.8	3.0	0.2	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
REACTOME_HOMOLOGOUS_RECOMBINATION_REPAIR_OF_REPLICATION_INDEPENDENT_DOUBLE_STRAND_BREAKS												
LIG1	NBN	ATM	RAD51	RAD52	RPA1	RPA2	BRCA1	BRCA2	TP53BP1	BRIP1	MDC1	RAD50
0.8	0.3	3.0	0.2	0.3	0.2	0.2	0.6	1.6	1.1	0.2	0.6	0.0
H2AFX	MRE11A	RPA3	LOC651610									
0.0	0.0	0.0	0.0									

REACTOME_NUCLEOTIDE_EXCISION_REPAIR												
POLD3	ERCC2	ERCC3	ERCC4	ERCC5	ERCC6	GTF2H3	LIG1	POLD1	POLE	POLR2A	POLR2B	POLR2K
0.2	0.5	0.3	0.2	0.3	0.8	0.2	0.8	0.2	0.8	0.5	0.5	0.2
XPA	XPC	CCNH	CDK7	ERCC8	DDB1	DDB2	ERCC1	GTF2H1	GTF2H2	GTF2H4	MNAT1	PCNA
0.5	0.6	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
POLR2G	POLR2H	POLR2I	POLR2J	POLR2L	POLD4	RAD23B	RFC4	RFC5	RPA3	LOC652672	LOC652857	GTF2H2B
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
XAB2	RFC2	RFC3	RPA1	RPA2	TCEA1	POLD2	POLE2	POLR2C	POLR2D	POLR2E	POLR2F	
0.2	0.2	0.2	0.2	0.2	0.2	0.0	0.0	0.0	0.0	0.0	0.0	
REACTOME_P53_INDEPENDENT_G1_S_DNA_DAMAGE_CHECKPOINT												
PSMD14	CHEK2	PSMA8	PSME4	ATM	PSMA1	PSMA2	PSMA6	PSMB4	PSMB5	PSMB6	PSMC5	PSMC6
0.2	0.8	0.3	0.2	3.0	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.3
PSME2	RPS27A	PSMF1	PSMD6	CDC25A	CHEK1	PSMA3	PSMA4	PSMA5	PSMA7	PSMB1	PSMB2	PSMB3
0.2	0.3	0.2	0.2	0.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
PSMC3	PSMC4	PSMD7	PSMD8	PSMD9	PSMD10	PSMD11	PSMD12	PSMD13	LOC651610	LOC652826	RPS27AP11	UBA52
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
PSMD1	PSMD2	PSMD3	PSMD4	PSMD5	PSME1	PSMB7	PSMB8	PSMB9	PSMB10	PSMC1	PSMC2	
0.2	0.3	0.3	0.2	0.2	0.2	0.0	0.0	0.0	0.0	0.0	0.0	

eTable 4: Comparison of the gene members of the HR gene sets

REACTOME_HOMOLOGOUS_RECOMBINATION_REPAIR_OF_REPLICATION_INDEPENDENT_DOUBLE_STRAND_BREAKS (M)	**Common**	**KEGG_HOMOLOGOUS_RECOMBINATION (N)**
ATM	BRCA2	BLM
BRCA1	MRE11A	EME1
BRIP1	NBN	MUS81
H2AFX	RAD50	POLD1
LIG1	RAD51	POLD2
LOC651610	RAD52	POLD3
MDC1	RPA1	POLD4
TP53BP1	RPA2	RAD51B
	RPA3	RAD51C
		RAD51D
		RAD54B
		RAD54L
		RPA4
		SHFM1
		SSBP1
		TOP3A
		TOP3B
		XRCC2
		XRCC3

eTable 5: Univariate and multivariate Cox analysis of individual DDR pathways and metastasis-free survival in the training cohort

Checkpoint				
			HR (95% CI)	***P* Value**
UVA			1.44 (1.28-1.61)	1.18E-09
MVA				
	Pathway		1.24 (1.09-1.40)	0.00067
	Age		1.00 (0.97-1.02)	0.815
	PSA			
		<10 (Low Risk)	1 [Reference]	
		10-20 (Int. Risk)	0.81 (0.56-1.17)	0.25
		>20 (High Risk)	0.81 (0.56-1.17)	0.26
	Gleason			
		≤6 (Low Risk)	1 [Reference]	
		7 (Int. Risk)	2.42 (1.04-5.62)	0.039
		8-10 (High Risk)	6.59 (2.83-15.33)	1.23E-05
	SMS (+ vs. -)		1.03 (0.78-1.38)	0.82
	SVI		1.45 (1.04-2.02)	0.028
	ECE		1.13 (0.83-1.55)	0.42
	LNI		1.10 (0.75-1.61)	0.63
UVA=univariate. MVA=multivariate. PSA=prostate specific antigen. SMS=surgical margin status. SVI=seminal vesicle invasion. ECE=extra-capsular extension.				
Repair				
			HR (95% CI)	***P* Value**
UVA			1.25 (1.13-1.38)	1.25E-05
MVA				
	Pathway		1.13 (1.02-1.26)	0.018
	Age		1.00 (0.98-1.02)	0.89
	PSA			
		<10 (Low Risk) (Ref.)	1 [Reference]	
		10-20 (Int. Risk)	0.81 (0.56-1.16)	0.25
		>20 (High Risk)	0.83 (0.57-1.19)	0.31
	Gleason			
		≤6 (Low Risk) (Ref.)	1 [Reference]	
		7 (Int. Risk)	2.50 (1.08-5.79)	0.033
		8-10 (High Risk)	6.93 (2.98-16.10)	6.81E-06
	SMS (+ vs. -)		1.03 (0.77-1.37)	0.84
	SVI		1.46 (1.05-2.03)	0.024
	ECE		1.12 (0.82-1.53)	0.46

LNI		1.17 (0.81-1.71)	0.40
MMR			
		HR (95% CI)	***P* Value**
UVA		1.23 (1.09-1.40)	7.77E-04
MVA			
Pathway		1.13 (0.99-1.28)	0.061
Age		1.00 (0.98-1.02)	0.87
PSA			
	<10 (Low Risk)	1 [Reference]	
	10-20 (Int. Risk)	0.78 (0.54-1.13)	0.18
	>20 (High Risk)	0.81 (0.57-1.17)	0.27
Gleason			
	≤6 (Low Risk)	1 [Reference]	
	7 (Int. Risk)	2.57 (1.11-5.95)	0.028
	8-10 (High Risk)	7.17 (3.09-16.62)	4.42E-06
SMS (+ vs. -)		1.03 (0.77-1.37)	0.84
SVI		1.50 (1.08-2.08)	0.016
ECE		1.14 (0.83-1.55)	0.42
LNI		1.18 (0.81-1.72)	0.38
NER			
		HR (95% CI)	***P* Value**
UVA		1.27 (1.12-1.45)	2.11E-04
MVA			
Pathway		1.15 (1.01-1.32)	0.041
Age		1.00 (0.98-1.02)	0.88
PSA			
	<10 (Low Risk)	1 [Reference]	
	10-20 (Int. Risk)	0.79 (0.55-1.13)	0.20
	>20 (High Risk)	0.82 (0.57-1.18)	0.29
Gleason			
	≤6 (Low Risk)	1 [Reference]	
	7 (Int. Risk)	2.54 (1.10-5.88)	0.030
	8-10 (High Risk)	7.17 (3.09-16.62)	4.45E-06
SMS (+ vs. -)		1.04 (0.78-1.38)	0.80
SVI		1.48 (1.07-2.06)	0.019
ECE		1.12 (0.82-1.53)	0.46
LNI		1.15 (0.79-1.68)	0.47

DSB Repair				
			HR (95% CI)	***P* Value**
UVA			1.31 (1.16-1.47)	1.05E-05
MVA				
	Pathway		1.15 (1.02-1.30)	0.023
	Age		1.00 (0.98-1.02)	0.89
	PSA			
		<10 (Low Risk)	1 [Reference]	
		10-20 (Int. Risk)	0.80 (0.55-1.15)	0.22
		>20 (High Risk)	0.83 (0.57-1.19)	0.31
	Gleason			
		≤6 (Low Risk)	1 [Reference]	
		7 (Int. Risk)	2.48 (1.07-5.75)	0.035
		8-10 (High Risk)	6.86 (2.95-15.96)	7.73E-06
	SMS (+ vs. - margins)		1.03 (0.77-1.38)	0.83
	SVI		1.47 (1.06-2.04)	0.022
	ECE		1.13 (0.83-1.55)	0.42
	LNI		1.17 (0.80-1.70)	0.43
NHEJ				
			HR (95% CI)	***P* Value**
UVA			1.24 (1.08-1.42)	1.79E-03
MVA				
	Pathway		1.12 (0.98-1.29)	0.095
	Age		1.00 (0.98-1.02)	0.90
	PSA			
		<10 (Low Risk)	1 [Reference]	
		10-20 (Int. Risk)	0.78 (0.54-1.13)	0.19
		>20 (High Risk)	0.81 (0.56-1.17)	0.26
	Gleason			
		≤6 (Low Risk)	1 [Reference]	
		7 (Int. Risk)	2.59 (1.12-5.99)	0.027
		8-10 (High Risk)	7.26 (3.13-16.83)	3.78E-06
	SMS (+ vs. -)		1.03 (0.77-1.37)	0.87
	SVI		1.48 (1.07-2.06)	0.019
	ECE		1.13 (0.83-1.54)	0.43
	LNI		1.22 (0.84-1.77)	0.29

HR1				
			HR (95% CI)	***P* Value**
UVA			1.18 (1.03-1.35)	1.77E-02
MVA				
	Pathway		1.02 (0.89-1.17)	0.75
	Age		1.00 (0.98-1.02)	0.90
	PSA			
		<10 (Low Risk)	1 [Reference]	
		10-20 (Int. Risk)	0.79 (0.55-1.13)	0.20
		>20 (High Risk)	0.81 (0.56-1.17)	0.25
	Gleason			
		≤6 (Low Risk)		
		7 (Int. Risk)	2.64 (1.14-6.12)	0.024
		8-10 (High Risk)	7.47 (3.22-17.36)	2.9E-06
	SMS (+ vs. -)		1.03 (0.77-1.37)	0.84
	SVI		1.52 (1.10-2.11)	0.012
	ECE		1.13 (0.83-1.54)	0.45
	LNI		1.22 (0.84-1.77)	0.30
HR2				
			HR (95% CI)	***P* Value**
UVA			1.36 (1.13-1.64)	1.10E-03
MVA				
	Pathway		1.18 (0.98-1.43)	0.077
	Age		1.00 (0.97-1.02)	0.78
	PSA			
		<10 (Low Risk)	1 [Reference]	
		10-20 (Int. Risk)	0.79 (0.55-1.14)	0.20
		>20 (High Risk)	0.82 (0.57-1.18)	0.28
	Gleason			
		≤6 (Low Risk)	1 [Reference]	
		7 (Int. Risk)	2.59 (1.12-5.99)	0.026
		8-10 (High Risk)	7.25 (3.13-16.79)	3.8E-06
	SMS (+ vs. -)		1.03 (0.77-1.37)	0.85
	SVI		1.50 (1.08-2.08)	0.015
	ECE		1.13 (0.83-1.53)	0.45
	LNI		1.19 (0.82-1.73)	0.36

BER				
			HR (95% CI)	***P* Value**
UVA			1.11 (0.96-1.29)	1.58E-01
MVA				
	Pathway		0.98 (0.84-1.15)	0.82
	Age		1.00 (0.98-1.02)	0.93
	PSA			
		<10 (Low Risk)	1 [Reference]	
		10-20 (Int. Risk)	0.79 (0.54-1.13)	0.20
		>20 (High Risk)	0.80 (0.56-1.16)	0.24
	Gleason			
		≤6 (Low Risk)	1 [Reference]	
		7 (Int. Risk)	2.66 (1.15-6.17)	0.022
		8-10 (High Risk)	7.61 (3.29-17.63)	2.18E-06
	SMS (+ vs. -)		1.03 (0.77-1.37)	0.86
	SVI		1.53 (1.10-2.12)	0.011
	ECE		1.13 (0.83-1.54)	0.44
	LNI		1.22 (0.84-1.77)	0.30

eTable 6: Multivariate Cox analysis of the DDR pathway signature in the training cohort

			BCR-free survival		Metastasis-free survival		Overall survival	
			HR (95% CI)	***P* Value**	**HR (95% CI)**	***P* Value**	**HR (95% CI)**	***P* Value**
UVA			1.43 (1.16-1.75)	7.73E-04	2.49 (1.90-3.26)	3.82E-11	1.92 (1.51-2.44)	8.40E-08
MVA								
	DDR pathway signature (High vs. Low Risk)		1.23 (0.98-1.53)	0.069	1.73 (1.30-2.31)	0.00017	1.44 (1.12-1.86)	0.0048
	Age		1.00 (0.99-1.02)	0.67	1.00 (0.98-1.02)	0.86	1.04 (1.02-1.06)	0.00066
	PSA							
		<10 (Low Risk)	1 [Reference]		1 [Reference]		1 [Reference]	
		10-20 (Int. Risk)	1.07 (0.82-1.40)	0.60	0.78 (0.54-1.13)	0.189	0.71 (0.52-0.98)	0.038
		>20 (High Risk)	1.40 (1.07-1.85)	0.015	0.83 (0.57-1.19)	0.30	0.66 (0.48-0.92)	0.015
	Gleason							
		≤6 (Low Risk)	1 [Reference]		1 [Reference]		1 [Reference]	
		7 (Int. Risk)	1.06 (0.73-1.55)	0.76	2.49 (1.07-5.76)	0.034	2.13 (1.18-3.83)	0.012
		8-10 (High Risk)	1.99 (1.33-2.97)	0.00075	6.53 (2.80-15.20)	0.000014	4.68 (2.56-8.56)	0.00000051
	SMS (+ vs. -)		1.12 (0.90-1.39)	0.32	1.03 (0.78-1.38)	0.82	1.31 (1.01-1.69)	0.039
	SVI		1.51 (1.18-1.94)	0.0012	1.45 (1.04-2.01)	0.028	1.61 (1.20-2.16)	0.0016
	ECE		0.98 (0.78-1.23)	0.87	1.12 (0.83-1.53)	0.46	1.09 (0.83-1.42)	0.55
	LNI		0.48 (0.34-0.66)	0.000013	1.11 (0.76-1.62)	0.57	1.29 (0.92-1.81)	0.14

UVA=univariate. MVA=multivariate. DDR=DNA damage and repair. PSA=prostate specific antigen. SMS=surgical margin status. SVI=seminal vesicle invasion. ECE=extra-capsular extension. LNI=lymph node involvement.

eTable 7: C-indices from Cox proportional hazards models.

Cox model variable(s)	Cohort	c-index
DDR pathway signature	Training	0.61
	Pooled validation	0.60
DDR pathway signature + CPV	Training	0.72
	Pooled validation	0.71
Age	Training	0.51
	Pooled validation	0.50
PSA	Training	0.54
	Pooled validation	0.55
Gleason	Training	0.65
	Pooled validation	0.58
SMS	Training	0.54
	Pooled validation	0.56
SVI	Training	0.59
	Pooled validation	0.54
ECE	Training	0.58
	Pooled validation	0.47

DDR=DNA damage and repair. CPV=clinico-pathological variables. PSA=prostate specific antigen. SMS=surgical margin status. SVI=seminal vesicle invasion. ECE=extra-capsular extension.

	White	Black	Hispanic	Asian	Other
Low Risk	186	24	3	2	1
High Risk	90	7	0	0	0

eTable 8: DDR signature risk group distribution by ethnicity. n=313.

Custom R scripts and files for DDR mutation analysis

Contents of BroadCornellCell2013_primary_case_ids.rds

[[1]]
[1] "PR-STID0000003127" "PR-STID0000002872" "PR-STID0000002682" "PR-STID0000000415" "PR-3042" "PR-2741" "PR-2621" "PR-2525" "PR-09-601"
[10] "PR-09-3983" "PR-09-146" "PR-08-784" "PR-08-556" "PR-08-4154" "PR-08-2455" "PR-08-2153" "PR-07-4814" "PR-07-4610"
[19] "PR-07-360" "PR-07-3258" "PR-06-3199" "PR-06-1749" "PR-05-3595" "PR-0410" "PR-04-1243" "PR-03-728" "PR-03-3125"
[28] "PR-02-1431" "P09-730" "P09-628" "P09-396" "P09-37" "P09-1042" "P08-716" "P08-688" "P08-5852"
[37] "P08-501" "P08-492" "P08-217" "P08-1541" "P08-1042" "P07-837" "P07-5318" "P07-5037" "P07-5021"
[46] "P07-4941" "P07-144" "P05-620" "P05-3852" "P05-2709" "P05-1657" "P04-1084" "P03-2345" "P03-1426"
[55] "P01-28"

Contents of MSKCC_CancerCell2010_primary_case_ids.rds:

[[1]]
[1] "PCA0181" "PCA0180" "PCA0179" "PCA0178" "PCA0177" "PCA0176" "PCA0175" "PCA0174" "PCA0173" "PCA0172" "PCA0171" "PCA0170" "PCA0169" "PCA0168" "PCA0167" "PCA0166" "PCA0165" "PCA0164" "PCA0163"
[20] "PCA0162" "PCA0161" "PCA0160" "PCA0159" "PCA0158" "PCA0157" "PCA0156" "PCA0155" "PCA0154" "PCA0153" "PCA0152" "PCA0151" "PCA0150" "PCA0149" "PCA0148" "PCA0147" "PCA0146" "PCA0145" "PCA0144"
[39] "PCA0143" "PCA0142" "PCA0141" "PCA0140" "PCA0139" "PCA0138" "PCA0137" "PCA0136" "PCA0135" "PCA0134" "PCA0133" "PCA0132" "PCA0131" "PCA0130" "PCA0129" "PCA0128" "PCA0127" "PCA0126" "PCA0125"
[58] "PCA0124" "PCA0123" "PCA0122" "PCA0121" "PCA0120" "PCA0119" "PCA0118" "PCA0117" "PCA0116" "PCA0115" "PCA0114" "PCA0113" "PCA0112" "PCA0111" "PCA0110" "PCA0109" "PCA0108" "PCA0107" "PCA0106"
[77] "PCA0105" "PCA0104" "PCA0103" "PCA0102" "PCA0101" "PCA0100" "PCA0099" "PCA0098" "PCA0097" "PCA0096" "PCA0095" "PCA0094" "PCA0093" "PCA0092" "PCA0091" "PCA0090" "PCA0089" "PCA0088" "PCA0087"
[96] "PCA0086" "PCA0085" "PCA0084" "PCA0083" "PCA0082" "PCA0081" "PCA0080" "PCA0079" "PCA0078" "PCA0077" "PCA0076" "PCA0075" "PCA0074" "PCA0073" "PCA0072" "PCA0071" "PCA0070" "PCA0069" "PCA0068"
[115] "PCA0066" "PCA0065" "PCA0064" "PCA0063" "PCA0062" "PCA0061" "PCA0060" "PCA0059" "PCA0058" "PCA0057" "PCA0056" "PCA0055" "PCA0054" "PCA0053" "PCA0052" "PCA0050" "PCA0049" "PCA0048" "PCA0047"
[134] "PCA0046" "PCA0045" "PCA0044" "PCA0043" "PCA0042" "PCA0041" "PCA0040" "PCA0039" "PCA0038" "PCA0037" "PCA0036" "PCA0035" "PCA0034" "PCA0033" "PCA0032" "PCA0031" "PCA0030" "PCA0029" "PCA0028"
[153] "PCA0027" "PCA0026" "PCA0025" "PCA0024" "PCA0023" "PCA0022" "PCA0021" "PCA0020" "PCA0019" "PCA0018" "PCA0017" "PCA0016" "PCA0015" "PCA0014" "PCA0013" "PCA0012" "PCA0011" "PCA0010" "PCA0009"
[172] "PCA0008" "PCA0007" "PCA0006" "PCA0005" "PCA0004" "PCA0003" "PCA0002" "PCA0001"

Contents of DDR_GSEA_genesets.gmt:

BASE_EXCISION_REPAIR http://www.broadinstitute.org/gsea/msigdb/cards/BASE_EXCISION_REPAIR HMGB1P10 HMGB1 MSH6 HMGB2 MSH3 MSH2 POLG UNG TP53 SMUG1 MPG ERCC6 TDG OGG1 NTHL1 CCNO APEX1

DNA_DAMAGE_CHECKPOINT http://www.broadinstitute.org/gsea/msigdb/cards/DNA_DAMAGE_CHECKPOINT NBN ZAK HUS1 RINT1 PML RAD9A CHEK1 CHEK2 ATR BRSK1 NEK11 FOXN3 NAE1 RAD1 TRIAP1 PCBP4 GML CCNA2 ATRIP RAD17

DNA_REPAIR http://www.broadinstitute.org/gsea/msigdb/cards/DNA_REPAIR MMS19 XRCC4 RAD51C XRCC3 XRCC2 XRCC6 SETX IGHMBP2 MUTYH RAD21 CDKN2D PMS2 PMS1 CIB1 POLL POLI POLH POLG LIG1 POLE LIG3 RAD9A LIG4 KAT5 RAD1 PNKP RFC3 RPAIN RAD17 HMGB1 RAD23B HMGB2 BLM MRE11A RAD23A HUS1 UBE2V1 UBE2V2 XAB2 POLE2 RAD51B POLQ NTHL1 ASF1A RECQL4 HMGB1P10 RECQL5 TREX2 BRCA2 ATR RAD54L ATM BRCA1 VCP POLD1 TDG RAD54B PARP3 SMC1A ABL1 PARP1 NBN UVRAG MLH1 PMS2P1 FANCG CCNO FANCA FANCC UBE2A REV1 DDB1 GTF2H4 TP53 APTX PRKCG RAD52 UBE2B RAD50 XRCC6BP1 RAD51 GTF2H1 RBBP8 UBE2N RECQL XPC BTG2 GADD45G DDB2 RUVBL2 NHEJ1 GADD45A WRNIP1 UNG POLA1 SMUG1 RPA1 ERCC8 SUMO1 ERCC5 ERCC6 ERCC3 ERCC4 APEX1 FEN1 ERCC1 ERCC2 EXO1 MSH6 UPF1 MSH3 MSH2 MSH5 CEBPG TNP1 SOD1 TP73 ATRX MNAT1 MPG ATXN3 CSNK1D CSNK1E ALKBH1 OGG1

DOUBLE_STRAND_BREAK_REPAIR http://www.broadinstitute.org/gsea/msigdb/cards/DOUBLE_STRAND_BREAK_REPAIR XRCC4 NBN MRE11A XRCC6 POLA1 UBE2V2 LIG4 RAD52 SOD1 KAT5 BRCA1 RAD50 XRCC6BP1 RAD51 SETX UBE2N RAD21 VCP RAD54B NHEJ1 ERCC4 FEN1 CIB1

KEGG_BASE_EXCISION_REPAIR http://www.broadinstitute.org/gsea/msigdb/cards/KEGG_BASE_EXCISION_REPAIR NEIL2 MPG SMUG1 XRCC1 POLE4 HMGB1 POLE3 POLD4 MBD4 OGG1 UNG POLD3 PCNA NEIL1 POLE2 PARP4 PARP3 PARP2 POLB APEX1 POLL POLD1 POLD2 POLE NEIL3 TDG APEX2 LIG3 HMGB1P1 NTHL1 HMGB1P40 FEN1 LIG1 MUTYH PARP1

KEGG_HOMOLOGOUS_RECOMBINATION http://www.broadinstitute.org/gsea/msigdb/cards/KEGG_HOMOLOGOUS_RECOMBINATION RAD54L XRCC3 RAD51B NBN RAD52 RAD51D XRCC2 POLD4 BLM EME1 RPA1 RAD51 POLD3 RAD50 RPA3 RPA2 RPA4 RAD51C MRE11A BRCA2 POLD1 POLD2 MUS81 TOP3A RAD54B SHFM1 TOP3B SSBP1

KEGG_MISMATCH_REPAIR http://www.broadinstitute.org/gsea/msigdb/cards/KEGG_MISMATCH_REPAIR MLH3 POLD1 MLH1 POLD2 RFC1 MSH2 RFC3 RFC2 MSH3 POLD4 PMS2 RFC4 LIG1 RFC5 RPA1 MSH6 RPA3 POLD3 RPA2 PCNA SSBP1 RPA4 EXO1

KEGG_NON_HOMOLOGOUS_END_JOINING http://www.broadinstitute.org/gsea/msigdb/cards/KEGG_NON_HOMOLOGOUS_END_JOINING XRCC4 MRE11A POLL POLM LOC731751 NHEJ1 LIG4 FEN1 DNTT XRCC5 RAD50 XRCC6 PRKDC DCLRE1C

KEGG_NUCLEOTIDE_EXCISION_REPAIR http://www.broadinstitute.org/gsea/msigdb/cards/KEGG_NUCLEOTIDE_EXCISION_REPAIR MNAT1 POLE4 ERCC4 POLE3 ERCC3 ERCC6 ERCC5 GTF2H5 POLD4 ERCC2 RFC4 CETN2 RFC5 RPA1 RAD23B RBX1 DDB2 RPA3 POLD3 RPA2 RAD23A PCNA RPA4 DDB1 POLE2 ERCC1 POLD1 POLD2 POLE RFC1 RFC3 RFC2 XPC XPA GTF2H2 GTF2H1 CDK7 LIG1 CUL4A CUL4B ERCC8 CCNH GTF2H4 GTF2H3

NUCLEOTIDE_EXCISION_REPAIR http://www.broadinstitute.org/gsea/msigdb/cards/NUCLEOTIDE_EXCISION_REPAIR
POLL MMS19 RAD23B DDB1 RAD23A TP53 BRCA2 LIG4 XAB2 ERCC8 ATXN3 ERCC5
ERCC6 XPC DDB2 ERCC3 NTHL1 ERCC4 ERCC1 ERCC2
REACTOME_BASE_EXCISION_REPAIR
http://www.broadinstitute.org/gsea/msigdb/cards/REACTOME_BASE_EXCISION_REPAIR CCNO POLD3 FEN1
SMUG1 APEX1 LIG1 LIG3 MPG MUTYH NTHL1 OGG1 PCNA POLB POLD1 POLD2 POLD4 TDG
XRCC1 MBD4
REACTOME_DNA_REPAIR http://www.broadinstitute.org/gsea/msigdb/cards/REACTOME_DNA_REPAIR RAD50
CDK7 CCNO MAD2L2 POLD3 ERCC8 ALKBH2 DDB1 DDB2 ERCC1 ERCC2 ERCC3 ERCC4 ERCC5
ERCC6 FANCA FANCC FANCD2 FANCE FANCB FANCF FANCG ALKBH3 FEN1 SMUG1 XRCC6 ZBTB32
UBE2T GTF2H1 GTF2H2 GTF2H3 GTF2H4 H2AFX APEX1 LOC389901 LIG1 LIG3 LIG4 MGMT MNAT1
MPG MRE11A MUTYH NBN ATM NTHL1 OGG1 PCNA REV1 POLB POLD1 POLD2 POLE
POLE2 POLH POLR2A POLR2B POLR2C POLR2D POLR2E POLR2F POLR2G POLR2H POLR2I POLR2J POLR2K
POLR2L ATR FANCL TDP1 PRKDC XAB2 FANCM POLD4 RAD23B RAD51 RAD52 REV3L RFC2
RFC3 RFC4 RFC5 RPA1 RPA2 RPA3 RPS27A LOC648152 LOC651610 LOC651921
LOC652672 LOC652857 GTF2H2B BRCA1 BRCA2 TCEA1 TDG TP53BP1 RPS27AP11
UBA52 USP1 XPA XPC XRCC1 XRCC4 XRCC5 PALB2 C17orf70 BRIP1 MBD4 CCNH
C19orf40 MDC1
REACTOME_DOUBLE_STRAND_BREAK_REPAIR
http://www.broadinstitute.org/gsea/msigdb/cards/REACTOME_DOUBLE_STRAND_BREAK_REPAIR RAD50
XRCC6 H2AFX LOC389901 LIG1 LIG4 MRE11A NBN ATM TDP1 PRKDC RAD51 RAD52
RPA1 RPA2 RPA3 LOC651610 BRCA1 BRCA2 TP53BP1 XRCC4 XRCC5 BRIP1 MDC1
REACTOME_G2_M_DNA_DAMAGE_CHECKPOINT
http://www.broadinstitute.org/gsea/msigdb/cards/REACTOME_G2_M_DNA_DAMAGE_CHECKPOINT CHEK1
CHEK2 ATM ATR LOC648152 LOC651610 LOC651921 WEE1 ATRIP CCNB1 CDK1 CDC25C
REACTOME_HOMOLOGOUS_RECOMBINATION_REPAIR_OF_REPLICATION_INDEPENDENT_DOUBLE_STRAND_BREAKS
http://www.broadinstitute.org/gsea/msigdb/cards/REACTOME_HOMOLOGOUS_RECOMBINATION_REPAIR_OF_
REPLICATION_INDEPENDENT_DOUBLE_STRAND_BREAKS RAD50 H2AFX LIG1 MRE11A NBN ATM
RAD51 RAD52 RPA1 RPA2 RPA3 LOC651610 BRCA1 BRCA2 TP53BP1 BRIP1 MDC1
REACTOME_NUCLEOTIDE_EXCISION_REPAIR
http://www.broadinstitute.org/gsea/msigdb/cards/REACTOME_NUCLEOTIDE_EXCISION_REPAIR CDK7 POLD3
ERCC8 DDB1 DDB2 ERCC1 ERCC2 ERCC3 ERCC4 ERCC5 ERCC6 GTF2H1 GTF2H2 GTF2H3 GTF2H4 LIG1
MNAT1 PCNA POLD1 POLD2 POLE POLE2 POLR2A POLR2B POLR2C POLR2D POLR2E POLR2F POLR2G
POLR2H POLR2I POLR2J POLR2K POLR2L XAB2 POLD4 RAD23B RFC2 RFC3 RFC4 RFC5 RPA1
RPA2 RPA3 LOC652672 LOC652857 GTF2H2B TCEA1 XPA XPC CCNH
REACTOME_P53_INDEPENDENT_G1_S_DNA_DAMAGE_CHECKPOINT
http://www.broadinstitute.org/gsea/msigdb/cards/REACTOME_P53_INDEPENDENT_G1_S_DNA_DAMAGE_CHE
CKPOINT PSMD14 CHEK1 CHEK2 PSMA8 PSME4 ATM PSMA1 PSMA2 PSMA3 PSMA4 PSMA5 PSMA6
PSMA7 PSMB1 PSMB2 PSMB3 PSMB4 PSMB5 PSMB6 PSMB7 PSMB8 PSMB9 PSMB10 PSMC1 PSMC2
PSMC3 PSMC4 PSMC5 PSMC6 PSMD1 PSMD2 PSMD3 PSMD4 PSMD5 PSMD7 PSMD8 PSMD9 PSMD10
PSMD11 PSMD12 PSMD13 PSME1 PSME2 RPS27A LOC651610 LOC652826
RPS27AP11 UBA52 PSMF1 PSMD6 CDC25A

Custom R script for mutation analysis:

```
library(cgdsr)

date <- "06052015"

mycgds = CGDS("http://www.cbioportal.org/public-portal/")

#create data object for gene sets
file_conn <- file("DDR_GSEA_genesets.gmt")
DDR_GSEA_genesets <- strsplit(readLines(file_conn), "\t")
close(file_conn)
names(DDR_GSEA_genesets) <- sapply(DDR_GSEA_genesets, function(x) x[1])
DDR_GSEA_genesets <- lapply(DDR_GSEA_genesets, function(x) x[-c(1:2)])

#create CancerStudy list
study_list <- list("prad_broad", "prad_tcga", "prad_tcga_pub",
        "prad_broad_2013", "prad_mskcc")

#create overall case_id list
get_seq_case_list <- function (cancer_study) {
        seq_case_list_row <-
                getCaseLists(mycgds, cancer_study)[ , 2] == "Sequenced Tumors" |
                getCaseLists(mycgds, cancer_study)[ , 2] == "Sequenced Samples"
        initial_list <- unlist(strsplit(getCaseLists(mycgds, cancer_study)[
                seq_case_list_row, "case_ids"], split=" "))
        if (cancer_study == "prad_broad_2013") {
                return(initial_list[initial_list %in%
                        unlist(readRDS("BroadCornellCell2013_primary_case_ids.rds"))])
        } else if (cancer_study == "prad_mskcc") {
                return(initial_list[initial_list %in%
                        unlist(readRDS("MSKCC_CancerCell2010_primary_case_ids.rds"))])
        }
        else return(initial_list)
}
overall_case_id_list <- unique(unlist(lapply(study_list, get_seq_case_list)))

#generate list of genetic profiles and case lists
get_seq_info <- function (cancer_study) {
        mut_profile_list_row <-
                getGeneticProfiles(mycgds, cancer_study)[ , 2] == "Mutations"
        seq_case_list_row <-
                getCaseLists(mycgds, cancer_study)[ , 2] == "Sequenced Tumors" |
                getCaseLists(mycgds, cancer_study)[ , 2] == "Sequenced Samples"
```

```
        return(c(case_list=getCaseLists(mycgds, cancer_study)[seq_case_list_row, 1],
                profile=getGeneticProfiles(mycgds, cancer_study)[mut_profile_list_row, 1]))
}
seq_info <- lapply(study_list, get_seq_info)
names(seq_info) <- unlist(study_list)

get_sample_alteration <- function(get_mut_vector) {
        tryCatch({
                mut_data <- getMutationData(mycgds, get_mut_vector[1], get_mut_vector[2],
                        get_mut_vector[3])[ , c("case_id", "amino_acid_change")]
                data.frame(row.names=rownames(mut_data),
                        Sample=mut_data$case_id,
                        Gene=rep(gene, nrow(mut_data)),
                        Alteration=mut_data$amino_acid_change)
                },
                error = function (err){
                        #print(err)
                        print(get_mut_vector)
                        return(NULL)}
        )
}

oncoprint_list_to_mutation_call_matrix <- function(op_list) {

        genes <- unique(op_list$Gene[op_list$Gene != ""])
        mut_matrix <- matrix(
                rep(0, length(unique(op_list$Sample)) * length(genes)),
                nrow=length(unique(op_list$Sample)), ncol=length(genes),
                dimnames=list(unique(op_list$Sample), genes)
        )
        mut_calls_df <- op_list[op_list$Alteration != "", ]
        for(i in 1:nrow(mut_calls_df)) {
                mut_matrix[mut_calls_df[i, "Sample"], mut_calls_df[i, "Gene"]] <- 1
        }
        return(mut_matrix)
}

overall_mut_percentage_df <- data.frame()
mut_pct_list <- list()

#iterate over gene sets to produce mutation matrix
for(i in 1:length(DDR_GSEA_genesets)) {

        gene_set <- DDR_GSEA_genesets[[i]]
        gene_set_name <- names(DDR_GSEA_genesets)[i]
```

```
oncoprint_list <- data.frame()

#iterate over genes in gene set
for (gene in gene_set) {
	#check mutations in Broad/Cornell Nat Gen 2012 – all cases (112), and collect case_ids
	oncoprint_list <- rbind(oncoprint_list,
		get_sample_alteration(c(seq_info[[1]], gene)))
	#check mutations in TCGA in prep – all cases, and collect case_ids
	oncoprint_list <- rbind(oncoprint_list,
		get_sample_alteration(c(seq_info[[2]], gene)))
	#check mutations in TCGA provisional – all cases, and collect case_ids
	oncoprint_list <- rbind(oncoprint_list,
		get_sample_alteration(c(seq_info[[3]], gene)))
	#check mutations in Broad/Cornell Cell 2013, narrow to primary cases (55) and collect case_ids
	temp_oncoprint_list <- get_sample_alteration(c(seq_info[[4]], gene))
	if(!is.null(temp_oncoprint_list)) {
		temp_oncoprint_list <- temp_oncoprint_list[
			rownames(temp_oncoprint_list) %in%
			unlist(readRDS("BroadCornellCell2013_primary_case_ids.rds")), ]
		oncoprint_list <- rbind(oncoprint_list, temp_oncoprint_list)
	}
	#check mutations in MSKCC Cancer Cell 2010, narrow to primary cases (157 cases), and collect case_ids
	temp_oncoprint_list <- get_sample_alteration(c(seq_info[[5]], gene))
	if(!is.null(temp_oncoprint_list)) {
		temp_oncoprint_list <- temp_oncoprint_list[
			rownames(temp_oncoprint_list) %in%
			unlist(readRDS("BroadCornellCell2013_primary_case_ids.rds")), ]
		oncoprint_list <- rbind(oncoprint_list, temp_oncoprint_list)
	}

} #end gene iteration

#collect number of cases mutated for at least 1 gene in this gene set
num_mut_cases <- length(unique(oncoprint_list$Sample))

#collect non-mutated cases and genes
no_mut_case_ids <- overall_case_id_list[!overall_case_id_list %in%
	oncoprint_list$Sample]
no_mut_genes <- gene_set[!gene_set %in% oncoprint_list$Gene]
no_mut_Gene_column <- c(no_mut_genes, rep("",
	length(no_mut_case_ids) - length(no_mut_genes))
)

#build list of non-mutated cases and genes
no_mut_df <- data.frame(Sample=no_mut_case_ids,Gene=no_mut_Gene_column,
```

```
            Alteration=rep("", length(no_mut_case_ids))
        )

        #add non-mutated cases and genes to oncoprint_list
        oncoprint_list <- rbind(oncoprint_list, no_mut_df)

        #write to file
        filename <- paste(paste(gene_set_name, "oncoprint", date,
            sep="_"), ".txt", sep="")
        write.table(oncoprint_list, file=filename, sep="\t", row.names=F,
            col.names=T, quote=F)

        #generate mutation call matrix and write to file
        mut_call_matrix <- oncoprint_list_to_mutation_call_matrix(oncoprint_list)
        filename <- paste(paste(gene_set_name, "mutation_call_matrix", date,
            sep="_"), ".txt", sep="")
        write.table(mut_call_matrix, file=filename, sep="\t", row.names=T,
            col.names=T, quote=F)

        #tabulate mutation percentages by gene for each geneset
        mut_pct_list[[i]] <- apply(mut_call_matrix, 2, function(x)
            100 * (sum(x) / length(x))
            )
        names(mut_pct_list)[i] <- names(DDR_GSEA_genesets)[i]


        #tabulate overall mutation information
        num_nonmut_cases <- length(unique(no_mut_case_ids))
        percentage_mut_cases <- 100 * (num_mut_cases/(num_mut_cases +
            num_nonmut_cases))
        overall_mut_percentage_df <- rbind(
            overall_mut_percentage_df,
            data.frame(Gene_Set=gene_set_name,
                Num_mut=num_mut_cases,
                Num_nonmut=num_nonmut_cases,
                Mut_pct=percentage_mut_cases
            )
        )

} # end gene set iteration

filename <- paste("DDR_genesets_overall_mutation_info_", date, ".txt",
        sep="")
write.table(overall_mut_percentage_df, file=filename, row.names=F,
        col.names=T, quote=F)
```

```
filename <- paste("DDR_genesets_individual_gene_mut_pct_", date, ".txt",
          sep="")
sink(file=filename)
print(mut_pct_list)
sink()
```