Supporting information

SI Materials and Methods

Sequencing and variant calling. Genomic DNA was extracted from peripheral blood samples (Puregene, Qiagen) obtained from a series of genetically uncharacterized cases in the Dyskeratosis Congenita Registry (held at Barts and The London School of Medicine, London, UK), with written consent under the approval of our local research ethics committee (London – City and East). Exome data was processed and called jointly with a set of 2,500 WES internal control samples (UCL-ex consortium) using the recommendations from the Genome Analysis Toolkit (GATK v3.2) to minimize artefactual batch effects (1). All variants identified were validated by Sanger sequencing on a 3130xl Genetic Analyzer with a BigDye Terminator v.3.1 Cycle Sequencing Kit (Applied Biosystems). Using a next generation sequencing (NGS) assay we targeted the coding regions and some 5'UTRs from 90 genes associated with genetic bone marrow failure syndromes (*SI Appendix*, Table S1). We used the Illumina TruSeq custom amplicon kit for library preparation and capture according to the manufacturer's instructions. The resultant targeted fragments were indexed by dual barcodes and then sequenced on the Illumina MiSeq platform. Read alignment, variant calling and annotation was performed using an in-house pipeline involving the Burrows-Wheeler aligner, the Genome Analysis Toolkit and ANNOVAR, respectively.

Cell culture plasmids and transfection. 293T, HeLa and A549 cells were cultured in Dulbecco's modified Eagle's medium (DMEM) and lymphoblastoid lines (LCLs) acquired from ERCC6L2 mutated individuals and a FANCG mutated case were grown in RPMI. All culture media was supplemented with 10% (v/v) fetal bovine serum (FBS; HyClone), 100 IU/ml penicillin, and 100 mg/ml streptomycin (Invitrogen). Cells were maintained at 37°C in a humidified incubator with 5% CO_2. To achieve a non-cycling phase, cells were serum starved (0.1%) for 48 hours and labelled with BrdU (Sigma) at 32μM/ml concentration for 15 mins and subsequently harvested for genomic DNA extraction using QiaDNA genomic DNA prep kit. 1μg of extracted genomic DNA was slot blotted on to nitrocellulose (NC) membrane and subsequently immunoprobed with mouse monoclonal BrdU antibody (Sigma). To determine loading controls Sybr Gold staining was performed on stripped NC membranes. For ectopic expression studies, cells were transfected with Lipofectamine 2000 (Invitrogen) in optiMEM (Invitrogen) following manufacturer's protocol.

Cell viability analysis. LCLs were treated with serial doses of either mitomycin C, phleomycin, irofulven, actinomycin D (Act-D), 5,6 Dichlorobenzimidazole 1-β-D-ribofuranoside (DRB), and the DNA-PK inhibitor (NU7026) were dissolved in DMSO at indicated concentrations. Cell viability was assessed by neutral red dye uptake by live cells. All chemicals were obtained from Enzo Life Sciences and Sigma Aldrich. All readings were

normalized to the untreated sample. We calculated statistical significance by comparing the linear regression of the curves with GraphPad Prism 5 software (GraphPad).

Immunocytochemistry, imaging and analysis. In brief, after appropriate treatment, LCLs were subjected to cytospin on a poly-D lysine coated slides (Sigma) before fixation. After fixation with 4% PFA, cells were permeabilized with 0.1% Triton X-100 (TX100) in PBS, quenched in 50 mM NH_4Cl, and blocked in 10% goat serum and 1% BSA in PBS containing 0.05% TX100 for 1 hour. Cells were incubated in corresponding primary antibodies followed by Alexa Fluor 488/568 conjugated secondary antibodies (Invitrogen) in blocking solution. Cells were washed three times in PBS containing 0.05% TX100 between primary and secondary antibody incubations and mounted with vectashield containing DAPI (Vector Labs). For DNA-RNA hybrid (R-loop) S9.6 antibody was used at 1:200 dilution in blocking solution (Enzo Lifesciences). For RNase H treatment, cells were treated with RNase H (New England Biolabs) for 3 hours at 37°C post fixation, followed by incubation with S9.6 antibody. A Zeiss LSM700 confocal microscope with ZEN software was used and 63X captured images were acquired. Intensity per nucleus was determined using Image J (v 1.47). The DAPI signal was used to create a mask of the nucleus.

Transcription assay in intact cells. Inhibition of transcription after treatment with either irofulven, ActD and DRB was assessed with the Click-it RNA Imaging Kit and Alexa Fluor 488 azide (Invitrogen). LCLs were grown in 96 wells and following treatment with aforementioned compounds for 3 hours. Cells were then washed in PBS three times and incubated with 1mM 5-ethynlyl uridine (5-EU) for 15 min before fixation with 4% paraformaldehyde at specific time points. Click reactions were performed in accordance with the manufacturer's instructions. The fluorescence read out was measured with standard 480/520 nm excitation / emission wavelengths with the use of a FLUOROstar Optima plate reader (BMG Labtech).

Western Blots. For western blot analysis, whole cell protein extracts were prepared by lysing washed cells in denaturing buffer (9 M urea, 150 mM 2-mercaptoethanol, 50 mM Tris-HCl pH 7.3) and subsequently sonicated to shear genomic DNA. Total and phosphorylated forms of DNA-PK and RNA polymerase were separated on 3-8% Tris glycine mini gels (Life Technologies). For all other proteins, 4-12% NuPAGE Bis-Tris mini gels (Life Technologies) were used. Gels were transferred onto PVDF membrane (GE Healthcare). Blotting was performed using primary antibodies and the corresponding alkaline phosphatase conjugated secondary antibodies supplied in the WesternBreeze chemiluminescent kit (Thermo Fisher). β-actin antibody was used as a loading control. We performed semi-quantitative analysis of raw immunoblots by scanning the colorimetric blots at 600 dpi resolution to TIFF format files,

which we subjected to densitometry analysis software (GelPro) to identify quantitative changes in protein levels.

Cell cycle analysis. LCLs were treated with DMSO or irofulven for 3 hours, washed with PBS containing 2% FBS and left to grow in complete media. The following day, cells were re-suspended in 0.5 ml PBS with 2% FBS and fixed with ice cold 70% ethanol. Cells were then washed with PBS, incubated with 2 M HCl for 30 min, washed twice with PBS containing 0.5% Tween-20 and 1% BSA, and subsequently stained with propidium iodide solution (PBS, 0.5% Tween-20, 1% BSA, 20 mg/ml propidium iodide, 250 mg/ml RNase A) for 30 min at 37°C. Cell cycle profiles were analyzed using ACEA Novocyte (ACEA Biosciences) and NovoExpress™ software.

Mass spectrometry, statistical analysis and data visualization. Following co-immunoprecipitation with GFP-TRAP agarose beads, immobilized protein complexes were digested on agarose beads using trypsin (Sigma). Desalting and enrichment steps on the resultant peptides were conducted as previously described (2). LC-MS/MS analysis of peptides was performed in an Orbitrap mass spectrometer (Q-Exactive Plus). The Q-Exactive plus was operated in data dependent mode with one survey MS scan followed by 15 MS/MS scans. The full scans were acquired in the mass analyzer at 375- 1500 m/z with the resolution of 70000, and the MS/MS scans were obtained with a resolution of 17500. Peptides were identified using the Mascot search daemon (v.2.5.0), whereby sequences were searched against the SwissProt human protein database (mass windows were 10 ppm and 25 mmu for parent and fragment mass to charge values, respectively). Variable modifications included in searches were oxidation of methionine, pyro-glu (N-term) and phosphorylation of serine, threonine and tyrosine. Peptide quantification was carried out using PESCAL software, as described by Casado et al (3). Quantitative data were normalized in Excel, following which fold changes in peptide intensity and peptide numbers were calculated between the control GFP and GFP-ERCC6L2 pulldown experiments. The statistical significance of differences between conditions were assessed by Student's t-test. Network graphs depicting ERCC6L2 experimental interactome were constructed using the Cytoscape software package [v3.5.1(4)] and selected ontologies were derived from UniProt Knowledgebase lists using experimental evidence codes only (5). For generating the cytoscape map, the protein spheres (nodes) were weighted based on average max intensity and the font size of protein names was weighted on spectral counts. The thickness of the lines (edges) that linked proteins to ERCC6L2 were weighted on peptide numbers and the protein spheres (nodes) distance from ERCC6L2 were based on Pearson's correlation

scores. Subsequently proteins were grouped by ontologies and then by overlapping ontologies.

Chromatin Immunoprecipitation (ChIP) PCR and RTPCR. ChIP was preformed using Zymo-Spin™ ChIP kit. Briefly 5 x 10^6 293T cells were cross linked using 0.75% paraformaldehyde for 10 minutes and quenched by addition of 2.5M glycine. ChIP extracts were prepared following kits instructions. Sonication was performed on ice, using the Bioruptor instrument (Diagenode) that was optimized to produce DNA fragments ranging between 100 and 1,000 base pairs. ChIP grade antibodies against RNA Pol-II (Abcam), serine-2 RNA Pol-II CTD (Abcam), DNA-PK (Cell Signalling) and pDNA-PKcs S2056 (Abcam), were used at 3 µg/each IP. After IP, reverse cross-linking was performed and the ChIP DNA was purified and concentrated as instructed in the kit (Zymo research). Total sonicated genomic and ChIP DNA were PCR amplified using Q5 polymerase and specific primers for *MYC*, *FOS* and *JUN* gene bodies as mentioned previously (6). For RTPCR total RNAs from LCLs, A549 and $CD34^+$ hematopoietic progenitors were extracted using Trizol and further purified using the RNeasy kit (QIAGEN). Oligo $(dT)_{25}$ coated dyna beads were used specifically for pull down of poly(A) RNA. Equivalent amounts (~100 ng) of purified RNA were used as a template to synthesize cDNA using a mixture of random hexamers and Oligo $(dT)_{25}$ primers (1:1) and Superscript III reverse transcriptase (Invitrogen) according to the manufacturer's protocol. Specific primers were used for detection of *ERCC6L2* short (forward *TTGGGAACTGTGGAGGAAATC*; reverse *CTCTGAGATGGAGGTAGCAG*) and long isoform (forward *TTGGGAACTGTGGAGGAAAT*; reverse *TGATCCCTGCTTCTACTTGG*) transcripts in these cell types.

Telomere length measurement. Telomere lengths were measured using the monochrome multiplex quantitative PCR method modified from Cawthon (3) as described in Walne et al (7). Briefly, in each well, amplification of telomeric DNA (T) and a single copy gene (S) were quantified against standard curves obtained from the dilution of a reference DNA sample. The T/S ratio, obtained in triplicate for each sample, is proportional to the telomere length. This ratio was normalized to the T/S ratio of a second reference sample that was run on every plate to give a relative T/S ratio.

Statistics

Statistical analysis was performed with GraphPad Prism software (version 7), and a p value < 0.05 was considered statistically significant. In line graphs, for each experimental data set, a linear regression was conducted to determine the best-fit line describing the data from each independent experiment. The overall significance of cytotoxicity (Fig. 2*A-F* were

determined with one-way ANOVA with post Tukey's test on the slopes of the regression lines from each data set (n = 3 independent experiments performed in octuplicate). In bar graphs, a Mann-Whitney test was used to determine significant differences in between control and patient cell lines for both RNA Pol-II S2 CTD and DNA-PK serine 2056 phosphorylation level that were analyzed by GelPro densitometry software (Fig. S5*A and B*).

SI Figures

Fig. S1.

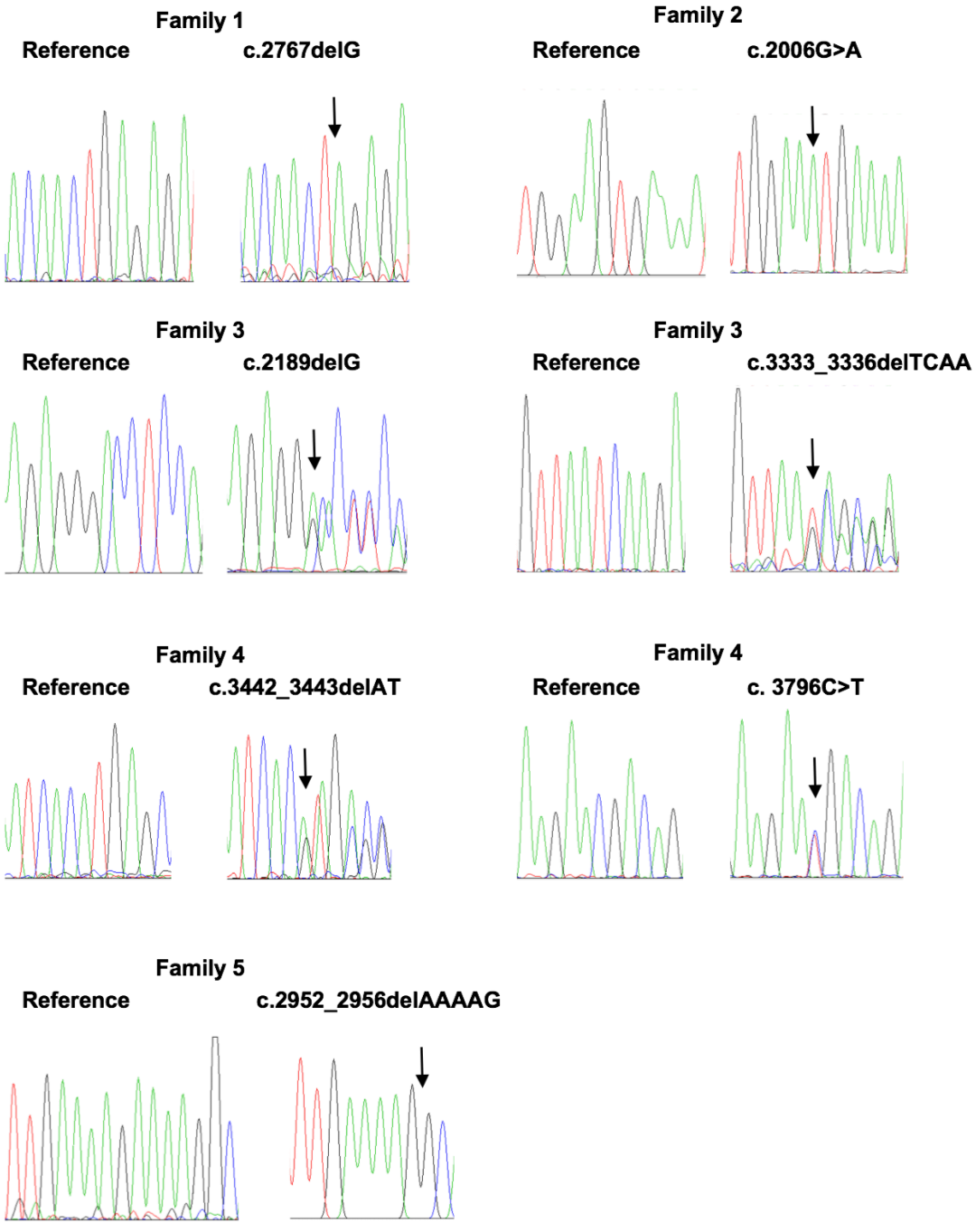

Fig. S1. Sanger sequencing traces of disease-causing variants in *ERCC6L2* (NM_020207.4) from each index case of the families, as indicated.

Fig. S2.

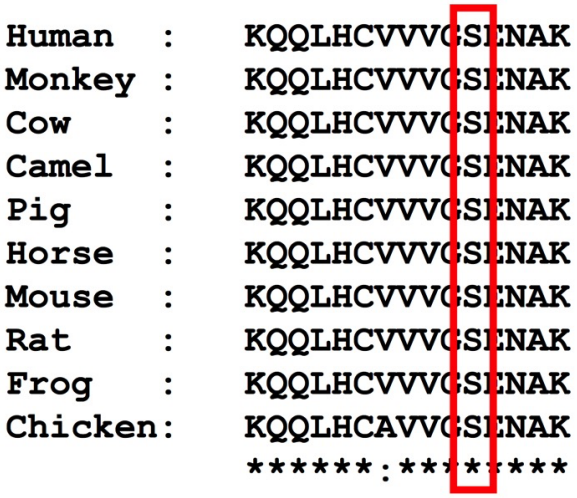

Fig. S2. Conservation of serine 669 residue in ERCC6L2. The alignment of serine 669 residue of human ERCC6L2 with indicated vertebrate species was generated with ClustalW. Asterisks indicate positions that have a single fully conserved residue and colons indicate conservation between groups of strongly similar properties

Fig. S3.

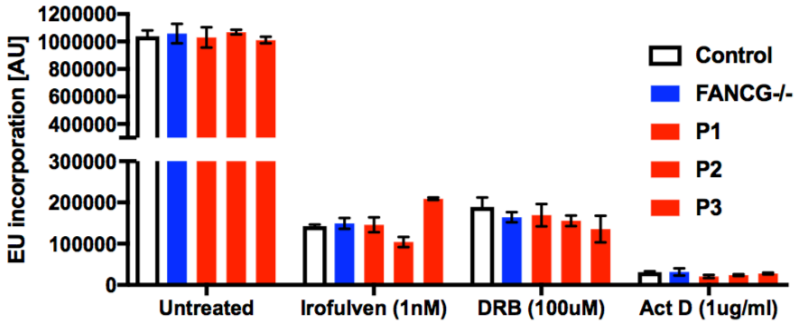

Fig. S3. Response to irofulven, ActD and DRB treatment in relevant LCLs using the Click-iT® RNA Assay. LCLs were treated with the indicated amounts of each drugs for 3 hours, followed by a 1hour incubation with 5-EU. Cells were then fixed and permeabilized and the 5-EU incorporated into newly synthesized RNA was detected using the green-fluorescent Alexa Fluor® 488 azide, using fluorescent plate reader at relevant exciting and emission filters. Error bars represent standard errors calculated between octuplets in each individual experiment (n=2 independent experiments).

Fig. S4.

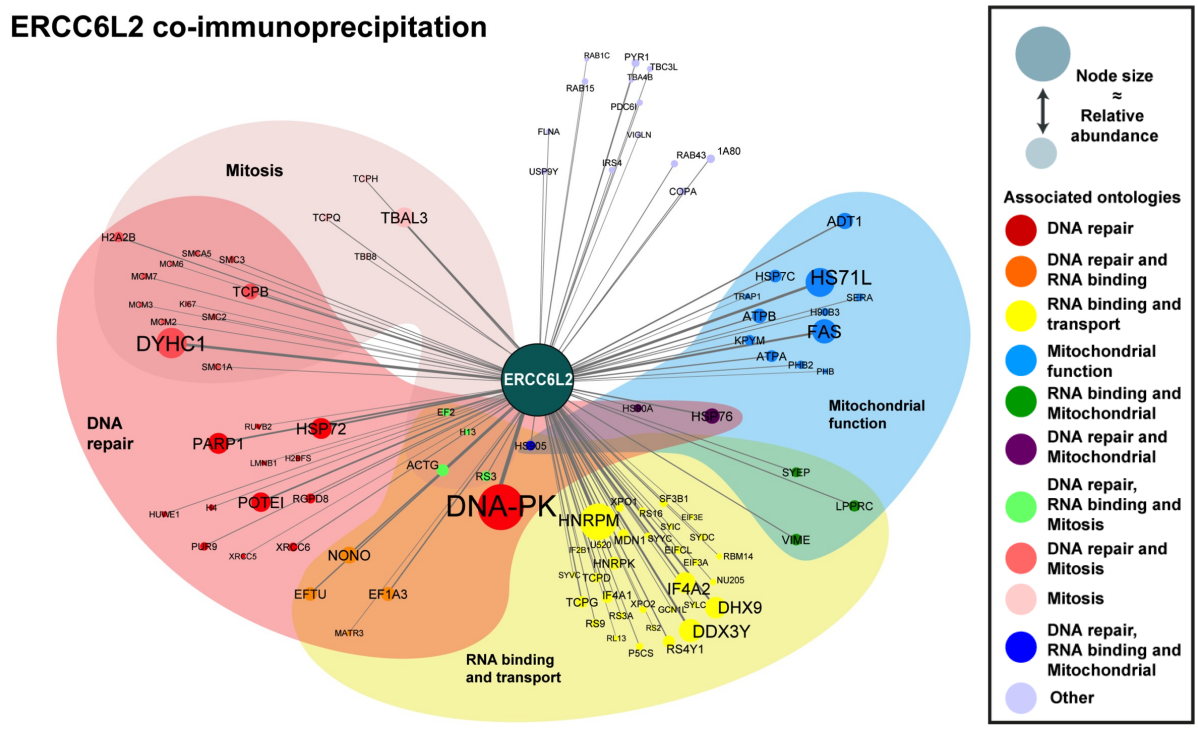

Fig. S4. Using the Cytoscape plug-in unit, the ERCC6L2 interaction landscape is over represented in clustered modules defining protein classes. The clusters display ERCC6L2 interaction with a network of proteins involved in DNA repair, RNA processing, mitosis and mitochondrial biogenesis.

Fig.S5.

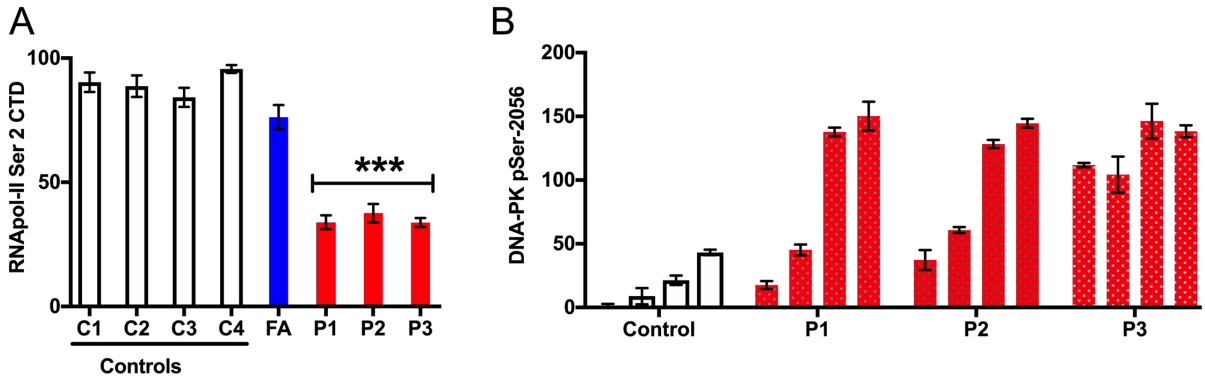

Fig. S5. (A) The bar graph represents quantified RNA Pol-II serine-2 phosphorylation levels, normalized to beta actin. Error bars represent the SEM (Mann-Whitney test), derived from data obtained from two independent experiments. (B) Densitometric analysis of the data in panel C shows the increase in serine-2065 phosphorylation of DNA-PK at the indicated time points after irofulven exposure in patients' cells compared control.

Fig. S6.

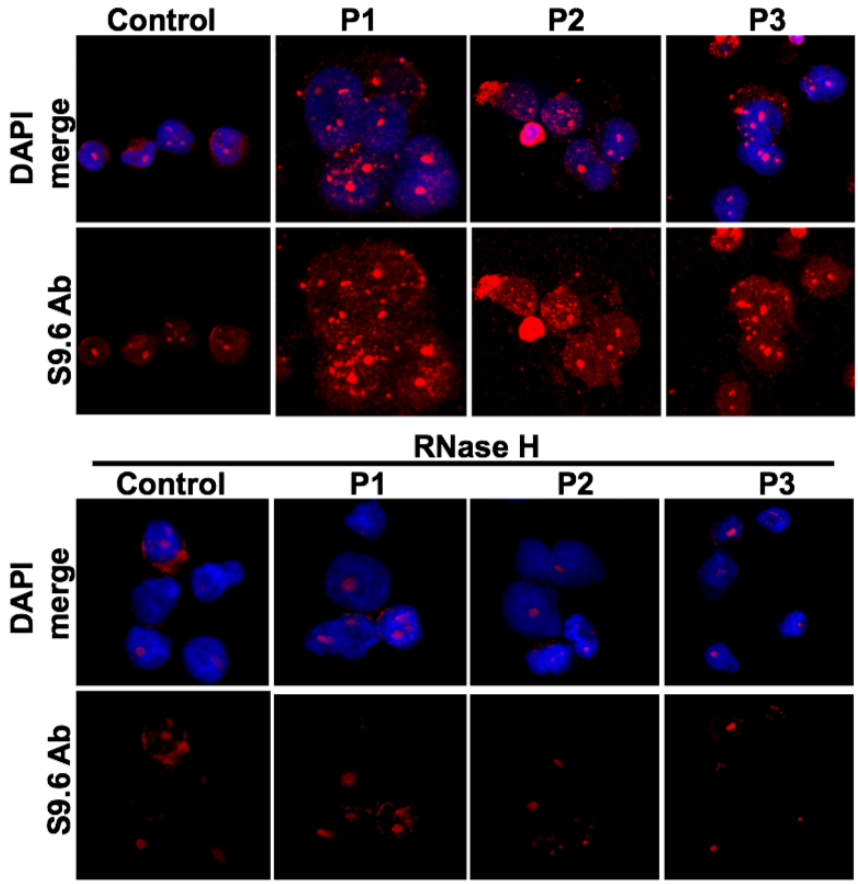

Fig.S6. Confocal images of the immunostaining normal control and patient cells with S9.6 antibody, with and without prior treatment with RNase H. DNA-RNA hybrids (R-loops) stained by S9.6 antibody are indicated in red and DAPI staining cell nuclei in blue. Images are representative of cells taken from 10 different fields of view (n=2 experiments, performed in triplicate).

Fig. S7.

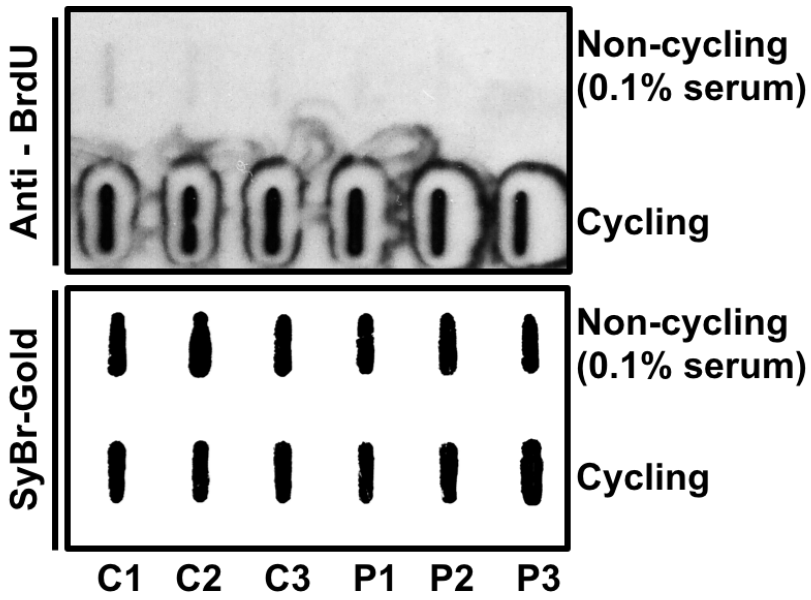

Fig. S7. Cells from three controls and three patients were maintained in 15% FBS (cycling) or grown to confluence and maintained in 0.1% FBS (non-cycling). DNA synthesis was determined by detecting the level of BrdU incorporation into genomic DNA (top panel). Immuno-slot blots were stained with Sybr Gold to ensure equal loading of genomic DNA.

Fig.S8.

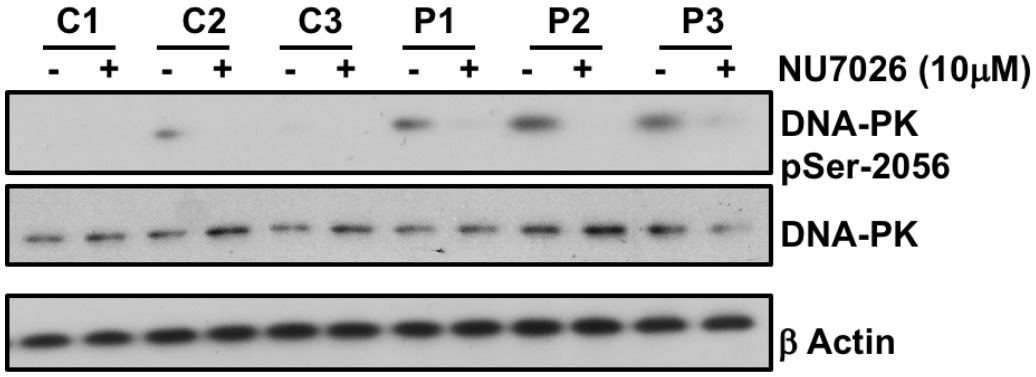

Fig. S8. Immunoblotting panels show levels of total and phosphorylated serine 2056 forms of DNA-PK and the loading control β-Actin in control and patients' cells before (-) and after (+) NU7026 treatment.

Fig.S9.

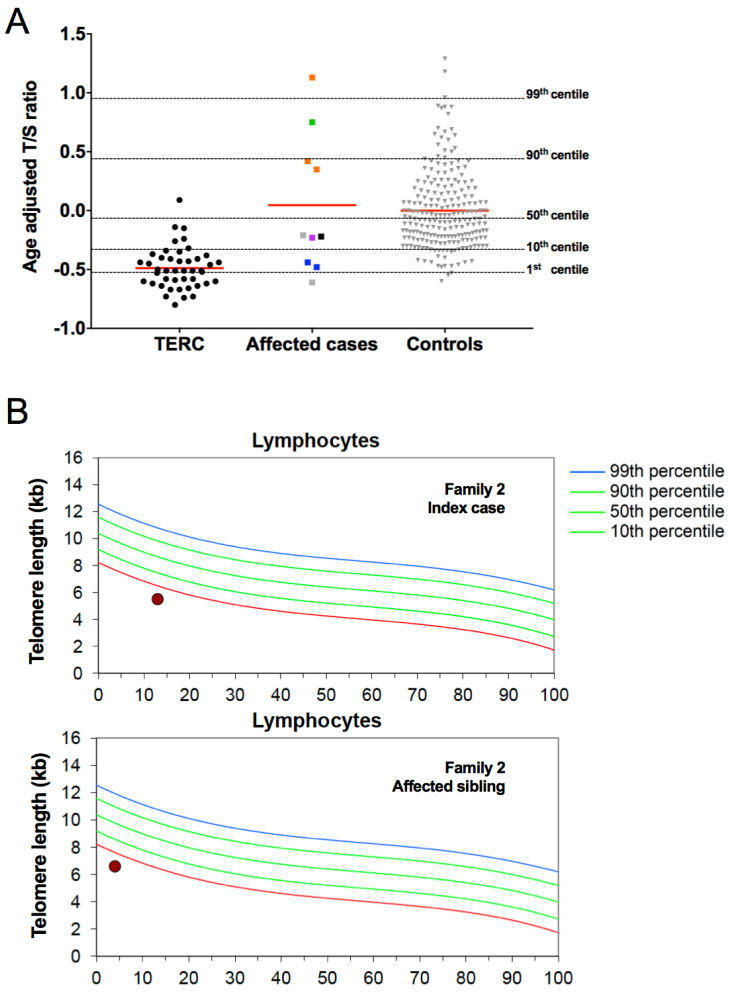

Fig. S9. (A) Age adjusted telomere length values (Δ-tel) were measured by subtracting the observed T/S ratio from the expected T/S ratio, using the equation derived from the line of best fit through the plot of T/S ratios from healthy control samples against age. Patients with TERC variants are included as a group with known short telomeres. Centiles were calculated from the control Δ-tel values as follows: 99th centile = 0.95, 90th centile = 0.44, 50th centile = -0.07, 10th centile = -0.33, 1st centile = -0.52. The different genotypes are represented as follows, TERC: black circles (n=44); ERCC6L2 cases: squares (n=10); controls: grey triangles (n=218). Colours indicate cases in green: family1; blue: family 2; purple: family3; orange: family 4; black: family 5; grey: cases from Tummala et al, 2016[2]. (B) Telomere lengths of affected cases from family 2 were measured by automated multicolor flow-FISH and depicted as percentiles by calculating a reference range for telomere length over age in lymphocytes from 400 healthy individuals (Repeat Diagnostics, Vancouver). Affected cases (indicated by red circles) show telomere lengths below first percentile for age in lymphocytes.

Fig. S10.

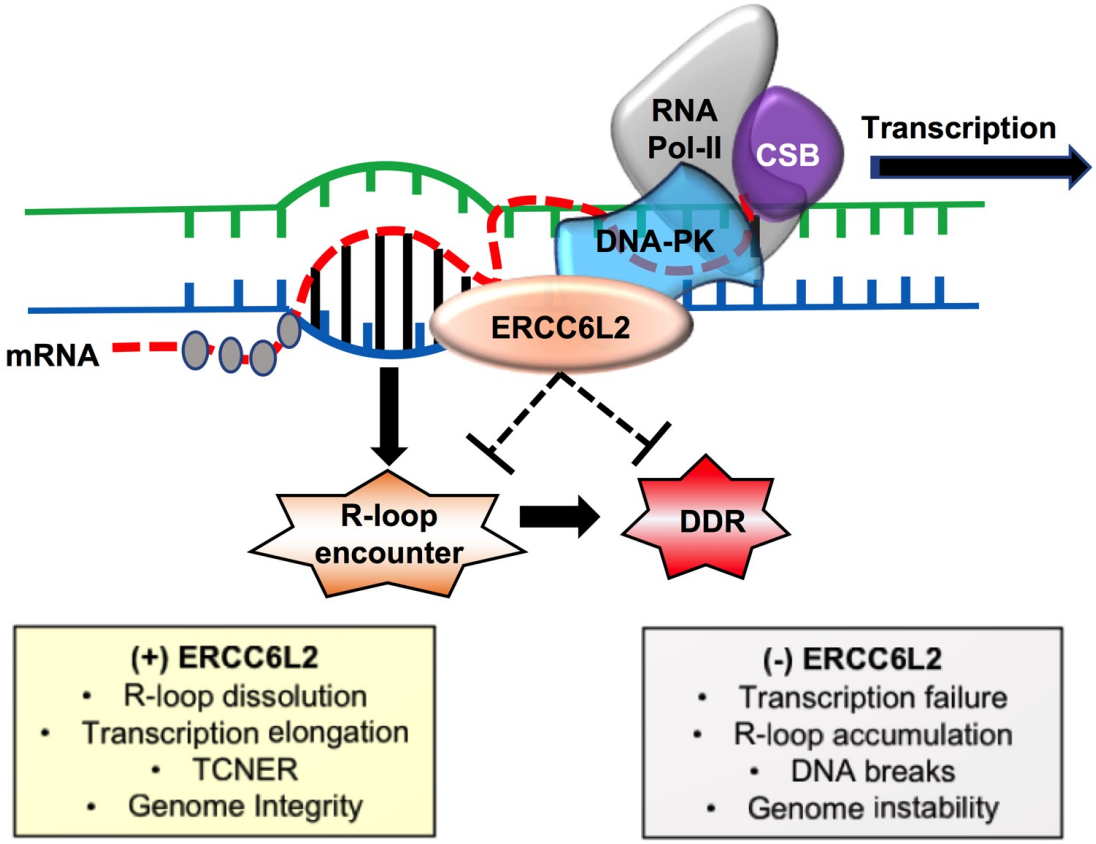

Fig. S10. ERCC6L2 at the nexus of transcription and the DNA damage response: RNA Pol-II proceeds along its DNA coding template, releasing the mRNA for processing. ERCC6L2:DNA-PK:RNA Pol-II complex regulates transcription elongation and accurate termination upon DNA damage. Loss of ERCC6L2 function leads to transcription elongation defects due to R-loop accumulation. Encountering R-loops stalls transcription and initiating DNA damage response (DDR). Presence of ERCC6L2 inhibits R loop formation and DNA damage response.

Table S1. Current gene list (n=90) that was used for the characterisation of new cases with bone marrow failure.

HUGO	Transcript ID	HUGO	Transcript ID	HUGO	Transcript ID
ACD	ENST00000393919	FANCM	ENST00000267430	RPL11	ENST00000374550
ANKRD26	ENST00000376087	G6PC	ENST00000253801	RPL15	ENST00000307839
BRCA1	ENST00000471181	G6PC3	ENST00000269097	RPL26	ENST00000584164
BRCA2	ENST00000544455	GATA1	ENST00000376670	RPL35A	ENST00000464167
BRIP1	ENST00000259008	GATA2	ENST00000341105	RPL5	ENST00000370321
C15ORF41	ENST00000566621	GFI1	ENST00000370332	RPS10	ENST00000326199
CDAN1	ENST00000356231	GRHL2	ENST00000251808	RPS17	ENST00000330244
CEBPA	ENST00000498907	HAX1	ENST00000328703	RPS19	ENST00000598742
CSF3R	ENST00000373103	HOXA11	ENST00000006015	RPS24	ENST00000440692
CTC1	ENST00000315684	JAGN1	ENST00000307768	RPS26	ENST00000356464
CXCR4	ENST00000409817	KIF23	ENST00000260363	RPS28	ENST00000600659
DDX41	ENST00000507955	KLF1	ENST00000264834	RPS29	ENST00000396020
DKC1	ENST00000369550	LIG4	ENST00000356922	RPS7	ENST00000304921
DNAJC21	ENST00000382021	MAD2L2	ENST00000235310	RTEL1	ENST00000508582
DNAJC3	ENST00000602402	MECOM	ENST00000264674	RUNX1	ENST00000300305
EGFR	ENST00000275493	MPL	ENST00000372470	SAMD9	ENST00000379958
ELANE	ENST00000590230	MYSM1	ENST00000472487	SAMD9L	ENST00000318238
ERCC4	ENST00000311895	NAF1	ENST00000274054	SBDS	ENST00000246868
ERCC6L2	**NM_020207.4***	NHP2	ENST00000274606	SEC23B	ENST00000336714
ETV6	ENST00000396373	NOP10	ENST00000328848	SLX4	ENST00000294008
FANCA	ENST00000389301	NPM1	ENST00000296930	SRP72	ENST00000342756
FANCB	ENST00000398334	PALB2	ENST00000261584	TAZ	ENST00000601016
FANCC	ENST00000289081	PARN	ENST00000437198	TERC	ENST00000363312
FANCD2	ENST00000287647	PAX5	ENST00000358127	TERT	ENST00000310581
FANCE	ENST00000229769	RAD51	ENST00000382643	TINF2	ENST00000267415
FANCF	ENST00000327470	RAD51C	ENST00000337432	TP53	ENST00000269305
FANCG	ENST00000378643	RBM8A	ENST00000583313	UBE2T	ENST00000367274
FANCI	ENST00000310775	RECQL4	ENST00000428558	USB1	ENST00000219281
FANCL	ENST00000402135	RMRP	ENST00000602361	VPS45	ENST00000369130
XRCC2	ENST00000359321	WRAP53	ENST00000316024	WAS	ENST00000376701

Table S2. *ERCC6L2* variants associated with bone marrow failure identified in this study.

Patient ID	**P1**	**P2**	**P3**	**P4**	**P5**
Nucleic acid change	c.2767delG homozygous	c.2006G>A homozygous	c.2189delG c.3333_3336delTCAA	c.3442_3443delAT c.3796C>T	c.2952_2956delAAAAG homozygous
Amino acid substitution	p.Glu923Argfs*8	p.Ser669Asn	p.Gly730Aspfs*50 p.Asn1111Lysfs*12	p.Met1148Glufs*7 p.Arg1266*	p.Lys985Hisfs*3
GNOMAD allele frequency	NR	NR	NR NR	4/242698 4/271282	NR
Previously reported in BMF	No	No	Yes[4] No	No Yes[4]	No
CADD score	1.4	20.7	20.3 11.3	35 38	14.4

GNOMAD, Genome Aggregation Database; CADD, Combined Annotation Dependent Depletion (PHRED score) that determines deleteriousness of single nucleotide variants as well as insertion/deletions; NR, not reported.

Table S3. Features of patients with biallelic *ERCC6L2* variants

Family	1	2	2	3	4	4	4	5
Index case	P1	P2		P3		P4		P5
Amino acid substitution	Glu923fs hom	Ser699Asn hom	Ser699Asn hom	Gly730fs Asn1111fs	Met1148fs Arg1266*	Met1148fs Arg1266*	Met1148fs Arg1266*	Lys985fs hom
Sex	F	M	F	F	F	M	F	M
Ethnic origin	Pk	Pk	Pk	UK	Ire	Ire	Ire	Sy
Age[a]	8	13	3	17	18	2	13	12
Parents first cousins	Yes	Yes	Yes	No	No	No	No	No
Bone marrow failure	Yes[b]	Yes[c]	Yes[d]	Yes[e]	Yes	Yes[f]	Yes	Yes[g]
MDS and/or AML	No	No	No	No	No	Yes	No	Yes
Hemoglobin (g/l)		70	112	120	119	97	127	81
Wbc (x10^9/l)		2.0	8.6	3.5	2.4	3.0	3.9	4.9
Platelets (x10^9/l)		58	131	90	166	12	85	102
Learning difficulties/DD	No	No	No	No	No	No	No	Yes
Microcephaly	No	No	No	Yes	No	No	No	No
Other features	No	Yes[h]	No	Yes[i]	No	Yes[j]	No	Yes[k]
Chromosomal breakage[l]	Normal	Normal	Normal	Normal	?	Normal	?	Normal
Telomere length[m]	Normal	Short[n]	Short[n]	Normal	Normal	Normal	Normal	Normal

[a]In years, at investigation; [b]anemia and hypocellular bone marrow; [c]hypocellular bone marrow; [d]investigated as a potential bone marrow donor for her older brother and was found to have a raised hemoglobin F at 2.6% and a hypocellular bone marrow for her age; [e]macrocytosis (mean corpuscular volume of 109 fentolitres); [f]this patient's bone marrow hypoplasia progressed to MDS and AML [complex karyotype including del(5) (q11.2q31), -18, +8] leading to fatal complications; [g]hypocellular bone marrow with tri-lineage dysplasia, monosomy 7 and trisomy 20; [h]failure to thrive, thin teeth, muscle pain; [i]delayed switch to adult teeth; [j]arterio-venous malformation; [k]café au lait pigmentation, leucoplakia, low birth weight, short stature; [l]after treatment of peripheral blood lymphocytes with diepoxybutane or mitomycin C; [m]age adjusted telomere length measurements showed a considerable range in the patient group (*SI Appendix*, Fig. S8*A*); [n]the index case and the affected sister from family 2 both have very short telomeres, as demonstrated by flow-FISH analysis (*SI Appendix*, Fig. S8*B*). Hom, homozygous; F, female; M, male; Pk, Pakistani; Ire, Irish; Sy, Syrian; MDS, myelodysplasia; AML, acute myeloid leukemia; Wbc, white blood cell count; DD, developmental delay; ?, unknown;

Table S4. ERCC6L2 interaction partners

Protein sorted by Max spectral counts	R - pearson correlation (sum of peptide intensity versus ERCC6L2 intensity)	Max spectral counts	Unique peptides	MASCOT score	Max intensity values	Log2 fold changes in intensity (ERCC6L2 Ab vs control)
ERCC6L2_HUMAN	1	423	138	22996.39	161195483.78	11.7900
PRKDC_HUMAN	0.9999	138	105	23361.62	9742436.36	5.4364
HNRPM_HUMAN	0.9974	112	60	11211.97	6054238.48	4.8041
DYHC1_HUMAN	0.9989	94.5	86	17672.04	4740097.01	4.6872
HS71L_HUMAN	0.9993	90	13	5020.98	14687658.12	4.2540
FAS_HUMAN	0.9995	77	61	13230.86	6933448.21	4.2745
DDX3Y_HUMAN	0.9999	74	18	4389.07	5432500.44	4.7960
DHX9_HUMAN	0.9956	69.5	47	8997.90	5579611.35	5.0394
IF4A2_HUMAN	0.9970	67.5	17	3703.57	3642035.71	4.0530
PARP1_HUMAN	0.9996	67	44	9659.29	5015744.27	3.8313
HSP72_HUMAN	0.9990	67	12	4308.64	13801594.45	3.9654
TBAL3_HUMAN	0.9916	64.5	5	1038.95	4753808.82	2.5872
POTEI_HUMAN	0.9990	63	2	2101.91	9007619.11	2.7111
NONO_HUMAN	0.9993	56	27	4624.00	5014815.63	4.1968
CLH1_HUMAN	0.9988	54	40	8845.20	4161084.87	4.8698
ADT1_HUMAN	0.9914	54	8	3157.71	6777985.08	3.7165
TCPB_HUMAN	0.9937	53	35	7097.82	2645043.70	3.7473
HSP76_HUMAN	0.9968	51.5	5	3658.75	14490604.64	4.4314
ATPB_HUMAN	0.9994	49	27	6835.88	4495716.91	3.8664
EF1A3_HUMAN	0.9962	48	8	4913.56	7480291.97	4.4300
EFTU_HUMAN	0.9982	46.5	31	6648.13	3568254.66	4.3308
HNRPK_HUMAN	0.9910	44.5	25	4907.33	3231566.73	3.3034
HSP7C_HUMAN	0.9987	44	28	8306.30	6129648.60	5.4232
RS4Y1_HUMAN	0.9988	43	7	1682.61	3972654.45	3.2803
ACTG_HUMAN	0.9973	42	7	7532.53	20447353.46	3.8615
ATPA_HUMAN	0.9995	41.5	27	7557.54	5379053.41	3.7949
TCPG_HUMAN	0.9944	41.5	32	6047.20	4306617.98	3.3484
VIME_HUMAN	0.9996	40.5	29	7359.83	3718021.60	5.1741
C1TC_HUMAN	0.9976	40	29	8488.16	3059957.29	3.5924
LPPRC_HUMAN	0.9998	40	32	7534.06	1344421.93	4.0167
IF4A1_HUMAN	0.9941	39.5	17	7001.82	3774115.16	5.0256
MDN1_HUMAN	0.9962	39.5	40	6966.04	1286640.50	11.2869
KPYM_HUMAN	0.9982	38	22	7088.50	7649329.76	3.7723
H2A2B_HUMAN	0.9944	37.5	1	453.02	22964759.70	3.7423
RS3_HUMAN	0.9996	37	20	4761.26	4860490.24	3.5947
SYEP_HUMAN	0.9954	36	30	6553.11	1862515.44	4.8846
RGPD8_HUMAN	0.9932	36	4	931.21	1961381.74	4.5805
TCPD_HUMAN	0.9965	35.5	24	5681.03	3189772.42	3.8360
HS905_HUMAN	0.9907	35	8	1002.09	3261098.54	4.2883

TCPQ_HUMAN	0.9921	32.5	26	7278.99	2908589.45	2.4269
RS9_HUMAN	0.9935	32.5	15	3707.51	4628802.59	4.2591
XPO1_HUMAN	0.9974	32	28	4282.20	1602533.82	3.4011
1A80_HUMAN	0.9954	32	1	210.03	1538194.88	11.5446
PHB2_HUMAN	0.9996	31.5	20	5297.90	1971810.53	3.9401
PUR9_HUMAN	0.9929	31.5	22	4299.21	1691924.58	4.3912
HS90A_HUMAN	0.9942	31	14	8757.32	3475356.80	4.1393
PYR1_HUMAN	0.9917	31	31	6479.46	1631566.75	4.2555
EF2_HUMAN	0.9969	31	23	6217.34	2299927.10	2.5384
SF3B1_HUMAN	0.9998	31	21	5055.54	1881935.20	6.5578
H90B3_HUMAN	0.9992	31	6	3714.42	2901292.32	3.3335
EIFCL_HUMAN	0.9954	31	15	1853.47	1997407.36	11.9214
COPA_HUMAN	0.9997	29.5	26	5755.42	1849889.02	5.4076
P5CS_HUMAN	0.9985	29.5	23	5060.39	1958550.11	5.0331
SERA_HUMAN	0.9935	29.5	17	4945.83	2762218.11	3.0590
SMC3_HUMAN	0.9963	29	28	5111.24	1314229.68	6.5629
TCPH_HUMAN	0.9938	29	21	4420.14	2369583.14	4.3817
XPO2_HUMAN	0.9934	29	19	4023.92	3057455.38	4.5438
RAB43_HUMAN	0.9958	28.5	1	273.72	8063020.50	3.2597
GBLP_HUMAN	0.9956	28	13	3906.20	2568293.04	3.6904
RS3A_HUMAN	0.9941	28	18	3861.90	3450186.35	2.6445
RS16_HUMAN	0.9997	28	10	2428.72	2985909.03	4.7129
EIF3A_HUMAN	0.9938	27	21	3912.72	1038788.79	2.4391
DDX17_HUMAN	0.9934	26.5	20	6728.10	2042982.74	3.9320
SYDC_HUMAN	0.9962	26.5	22	5806.36	1865722.15	4.3155
NU205_HUMAN	0.9987	26.5	19	4781.08	1159387.93	5.0363
MCM7_HUMAN	0.9968	26.5	23	4147.00	1577095.79	5.3619
RAB15_HUMAN	0.9986	26.5	3	895.53	2657796.78	3.5690
MCM2_HUMAN	0.9978	26	22	5276.20	2244969.76	4.9467
SMC1A_HUMAN	0.9969	26	22	4532.82	1564871.69	5.5565
IRS4_HUMAN	0.9977	26	18	4453.14	1062141.45	4.3266
RBM14_HUMAN	0.9949	26	20	4094.97	1269001.05	4.3006
H13_HUMAN	0.9908	26	5	2337.85	8484394.90	9.1535
HS902_HUMAN	0.9978	26	4	1728.89	4166611.10	4.1208
PDC6I_HUMAN	0.9958	25.5	26	5384.53	3140785.03	3.0418
SMC2_HUMAN	0.9995	25.5	26	4630.41	2619909.97	4.2220
SYYC_HUMAN	0.9961	25.5	21	4533.69	1629551.84	3.9763
U520_HUMAN	0.9992	25.5	23	4454.53	755540.34	5.0469
TBB8_HUMAN	0.9956	25.5	3	2763.72	3949762.63	4.2024
USP9Y_HUMAN	0.9991	25.5	10	1733.94	678750.49	10.3643
H4_HUMAN	0.9982	25	11	3020.03	2944325.31	2.5660
RL26_HUMAN	0.9979	25	8	2079.03	2749994.07	3.8152
TBC3L_HUMAN	1.0000	25	3	52.73	1342575.59	11.3483
HUWE1_HUMAN	0.9989	24.5	24	5178.44	685822.99	4.8118
GCN1L_HUMAN	0.9981	24.5	22	4564.17	614004.34	5.3828

SYIC_HUMAN	0.9990	24.5	23	4325.47	1701082.41	5.9659
FLNA_HUMAN	0.9910	24	23	7137.77	808133.29	2.2985
SMCA5_HUMAN	0.9945	24	20	4894.09	1009774.38	4.0065
SYLC_HUMAN	0.9980	24	22	4672.98	1294757.22	4.0111
TRAP1_HUMAN	0.9962	23.5	19	5319.41	1802614.82	3.9355
MATR3_HUMAN	0.9913	23.5	20	4218.15	1218129.43	3.6961
PHB_HUMAN	0.9966	23	17	4247.11	2523600.47	3.9495
RUVB2_HUMAN	0.9919	22.5	21	4468.58	766978.34	4.0745
H2BFS_HUMAN	0.9928	22.5	3	1882.00	2439421.24	4.2513
HS90B_HUMAN	0.9993	22	9	10483.35	1962966.30	3.9917
IF2B1_HUMAN	0.9971	22	14	4175.07	923732.63	3.3711
MCM3_HUMAN	0.9922	22	17	4053.08	816993.63	3.4820
VIGLN_HUMAN	0.9988	22	21	3779.35	871168.65	4.6355
RL4_HUMAN	0.9995	22	11	3324.31	2858824.94	4.2088
KI67_HUMAN	0.9996	22	23	3324.07	1593813.25	5.1718
RL13_HUMAN	0.9956	22	10	3118.53	1426783.90	2.8084
LMNB1_HUMAN	0.9982	21.5	17	5683.29	1008389.96	2.8124
MCM6_HUMAN	0.9998	21.5	19	4475.73	1422480.62	5.0089
SYVC_HUMAN	0.9983	21	17	3740.22	839646.92	4.3595
TBA4B_HUMAN	0.9940	21	3	510.60	2560793.64	5.2665
EIF3E_HUMAN	0.9999	20.5	14	3362.75	917753.71	4.4170
RS2_HUMAN	0.9938	20.5	14	2548.98	2932423.54	4.0580
RAB1C_HUMAN	0.9983	20.5	6	1448.90	2225511.53	7.6244

List of proteins identified by mass spectrometry after GFP-IP using soluble cells extracts from 293T cells overexpressing GFP-ERCC6L2. Result columns show the average of two experiments, with the number of unique peptides identified, maximum spectral counts and the calculated MASCOT scores (that relate to confidence in peptide identification) for each individual protein, sorted by at maximum spectral counts. ERCC6L2 and PRKDC are highlighted in green.

Table S5. Uniprot analysis of ERCC6L2 interactors.

DNA repair	RNA binding	Mitochondrial biogenesis	Mitosis	Other
ACTG_HUMAN	ACTG_HUMAN	ADT1_HUMAN	ACTG_HUMAN	1A80_HUMAN
DDX17_HUMAN	DDX17_HUMAN	ATPA_HUMAN	CLH1_HUMAN	COPA_HUMAN
DYHC1_HUMAN	DDX3Y_HUMAN	ATPB_HUMAN	DYHC1_HUMAN	FLNA_HUMAN
EF1A3_HUMAN	DHX9_HUMAN	C1TC_HUMAN	EF2_HUMAN	IRS4_HUMAN
EF2_HUMAN	EF1A3_HUMAN	FAS_HUMAN	H13_HUMAN	PDC6I_HUMAN
EFTU_HUMAN	EF2_HUMAN	H90B3_HUMAN ?	H2A2B_HUMAN	PYR1_HUMAN
H13_HUMAN	EFTU_HUMAN	HS71L_HUMAN	KI67_HUMAN	RAB15_HUMAN
H2A2B_HUMAN	EIF3A_HUMAN	HS905_HUMAN	MCM2_HUMAN	RAB1C_HUMAN
H2BFS_HUMAN	EIF3E_HUMAN	HS90A_HUMAN	MCM3_HUMAN	RAB43_HUMAN
H4_HUMAN	EIFCL_HUMAN	HSP72_HUMAN	MCM6_HUMAN	TBA4B_HUMAN
HS905_HUMAN	GCN1L_HUMAN	HSP76_HUMAN	MCM7_HUMAN	TBC3L_HUMAN
HS90A_HUMAN	H13_HUMAN	HSP7C_HUMAN	RS3_HUMAN	USP9Y_HUMAN
HSP72_HUMAN	HNRPK_HUMAN	KPYM_HUMAN	SMC1A_HUMAN	VIGLN_HUMAN
HSP76_HUMAN	HNRPM_HUMAN	LPPRC_HUMAN	SMC2_HUMAN	
HUWE1_HUMAN	HS905_HUMAN	PHB_HUMAN	SMC3_HUMAN	
KI67_HUMAN	IF2B1_HUMAN	PHB2_HUMAN	SMCA5_HUMAN	
LMNB1_HUMAN	IF4A1_HUMAN	SERA_HUMAN	TBAL3_HUMAN	
MATR3_HUMAN	IF4A2_HUMAN	SYEP_HUMAN	TBB8_HUMAN	
MCM2_HUMAN	LPPRC_HUMAN	TCPB_HUMAN	TCPB_HUMAN	
MCM3_HUMAN	MATR3_HUMAN	TRAP1_HUMAN	TCPH_HUMAN	
MCM6_HUMAN	MDN1_HUMAN	VIME_HUMAN	TCPQ_HUMAN	
MCM7_HUMAN	NONO_HUMAN			
NONO_HUMAN	NU205_HUMAN			
PARP1_HUMAN	P5CS_HUMAN			
POTEI_HUMAN	RBM14_HUMAN			
PRKDC_HUMAN	RL13_HUMAN			
PUR9_HUMAN	RS16_HUMAN			
RGPD8_HUMAN	RS2_HUMAN			
RS3_HUMAN	RS3_HUMAN			
RUVB2_HUMAN	RS3A_HUMAN			
SMC1A_HUMAN	RS3A_HUMAN			
SMC2_HUMAN	RS4Y1_HUMAN			
SMC3_HUMAN	RS9_HUMAN			
SMCA5_HUMAN	SF3B1_HUMAN			
	SYDC_HUMAN			
	SYEP_HUMAN			
	SYIC_HUMAN			
	SYLC_HUMAN			
	SYVC_HUMAN			
	SYYC_HUMAN			

	TCPD_HUMAN			
	TCPG_HUMAN			
	U520_HUMAN			
	VIME_HUMAN ?			
	XPO1_HUMAN			
	XPO2_HUMAN			

The indicated protein identified by mass spectrometry in the ERCC6L2 interactome reveal the proteins involved in DNA repair, RNA binding and transport, mitochondrial biogenesis, mitosis and other miscellaneous functions.

Table S6. Derivation of PRKDC peptides.

Protein_id	Peptide	MASCOT score	m/z	max expectancy	pFDR	MS/MS Ion intensity	Retention time (mins)
PRKDC_HUMAN	STVLTPMFVETQASQGTLQTR	143.45	1148.09	9.24E-14	0	94153.94	112.11
PRKDC_HUMAN	KEEENASVIDSAELQAYPALVVEK	120.7	878.12	1.74E-11	0	178713.97	109.29
PRKDC_HUMAN	LLLQGEADQSLLTFIDK	109.71	952.52	1.13E-10	0	12078.68	132.72
PRKDC_HUMAN	MEVQEQEEDISSLIR	108.47	903.44	3.69E-10	0	60258.16	108.55
PRKDC_HUMAN	LGASLAFNNIYR	105.27	669.86	3.97E-10	0	174160.18	102.84
PRKDC_HUMAN	STVLTPMFVETQASQGTLQTR	101.21	765.73	1.56E-09	0	128241.33	112.15
PRKDC_HUMAN	NLSSNEAISLEEIR	99.62	787.90	2.34E-09	0	351540.10	94.70
PRKDC_HUMAN	TVGALQVLGTEAQSSLLK	98.59	908.01	9.20E-10	0	25216.13	122.92
PRKDC_HUMAN	LTPLPEDNSMNVDQDGDPSDR	96.73	1158.01	6.47E-09	0	93946.71	84.07
PRKDC_HUMAN	SDPGLLTNTMDVFVK	96.14	818.92	5.36E-09	0	17183.13	120.30
PRKDC_HUMAN	KQNNFSLAMK	92.63	590.81	1.03E-08	0	14505.01	69.69
PRKDC_HUMAN	QMFLTQTDTGDDR	90	772.34	4.92E-08	0	31423.28	71.94
PRKDC_HUMAN	IMEFTTTLLNTSPEGWK	84.45	992.49	9.46E-08	0	25471.14	114.46
PRKDC_HUMAN	DVDFMYVELIQR	83.28	764.38	9.33E-08	0	20125.39	132.34
PRKDC_HUMAN	MSTSPEAFLALR	80.99	661.84	1.45E-07	0	76148.66	112.96
PRKDC_HUMAN	ATQMPEGGQGAPPMYQLYK	78.54	1033.99	4.11E-07	0	82123.20	93.51
PRKDC_HUMAN	IMEFTTTLLNTSPEGWK	77.95	984.49	3.68E-07	0	20307.78	128.66
PRKDC_HUMAN	LLALNSLYSPK	75.59	609.86	1.66E-07	0	84070.71	104.04
PRKDC_HUMAN	MVSAVLNGMLDQSFR	75.42	556.61	6.92E-07	0	6837.74	126.53
PRKDC_HUMAN	TVGALQVLGTEAQSSLLK	75.29	605.68	1.97E-07	0	19009.12	122.96
PRKDC_HUMAN	YNFPVEVEVPMER	74.16	804.89	9.75E-07	0	292221.10	110.29
PRKDC_HUMAN	DPESETDNDSQEIFK	71.72	877.37	3.32E-06	0	15076.83	79.07
PRKDC_HUMAN	MSTSPEAFLALR	69.93	669.84	1.98E-06	0	72375.50	106.76
PRKDC_HUMAN	MVSAVLNGMLDQSFR	69.19	834.42	2.73E-06	0	35592.90	126.52
PRKDC_HUMAN	QITQSALLAEAR	68.43	642.35	1.64E-06	0	124802.37	106.60
PRKDC_HUMAN	NLLTVTSSDEMMK	65.31	742.85	6.51E-06	0	55958.00	90.70
PRKDC_HUMAN	QGNLSSQVPLK	65.05	585.83	2.89E-06	0	37067.70	105.00
PRKDC_HUMAN	MEVQEQEEDISSLIR	65.05	602.62	8.66E-06	0	70710.13	67.60
PRKDC_HUMAN	ATQQQHDFTLTQTADGR	62.49	959.46	3.49E-05	0	10603.05	70.12
PRKDC_HUMAN	VVQMLGSLGGQINK	61.92	722.40	7.91E-06	0	174075.32	94.46

PRKDC_HUMAN	DVDFMYVELIQR	59.69	772.38	5.52E-05	0	8440.34	119.13
PRKDC_HUMAN	HGDLPDIQIK	58.75	568.31	1.75E-05	0	136324.36	83.82
PRKDC_HUMAN	LPLISGFYK	57.4	519.31	8.73E-06	0	77786.40	110.13
PRKDC_HUMAN	MAVLALLAK	56.64	465.30	1.29E-05	0	10934.87	122.13
PRKDC_HUMAN	DPTVHDDVLELEMDELNR	56.16	714.00	7.72E-05	0	127124.61	113.65
PRKDC_HUMAN	QNNFSLAMK	55.1	526.76	5.59E-05	0	12838.92	85.18
PRKDC_HUMAN	SMEEDPQTSR	53.91	590.25	5.04E-05	0	6345.26	45.23
PRKDC_HUMAN	IPALDLLIK	52.46	498.33	5.68E-06	0	10131.39	118.82
PRKDC_HUMAN	AALSALESFLK	52.37	575.33	5.27E-05	0	50321.55	131.02

Mass spectrometry data were processed using Mascot Daemon software. All 39 peptides that matched to human DNA-PK had MASCOT scores >50. m/z represents mass divided by charge number and the m/z value is often considered to be the mass of a given peptide.

Table S7. ERCC6L2 interactors identified by MS with known implication in R-loop biology

R-loop processing proteins	R - pearson correlation	References	Function
ERCC6L2	1.0000	This study	DNA excision repair protein, involved in transcription termination and R-loop suppression
PRKDC_HUMAN	0.9999	This study	DNA-dependent protein kinase catalytic subunit activity regulates transcription and supress R-loop accumulation
PARP1_HUMAN	0.9996	Cristina A et al., 2018	Poly [ADP-ribose] polymerase 1 involved in processing R-loops through its PARylation activity
DHX9_HUMAN	0.9956	Chakraborthy P et al., 2011 Cristina A et al., 2018	ATP-dependent RNA helicase A involved in processing R-loops
BRCA2_HUMAN	0.9935	Bhatia et al., 2014; Shivji MKK et al., 2018	Prevents R-loop accumulation
SETX_HUMAN	0.9932	Skourti-Stathaki, et al., 2011	Helicase involved in resolving R-loops during transcription termination
AQR_HUMAN	0.9931	Sollier J, et al., 2014	Prevents R-loop accumulation
PCID2_HUMAN	0.9777	Bhatia et al., 2014	PCI domain-containing protein 2 interacts with BRCA2 to suppress R-loops
XRN2_HUMAN	0.9465	Morales et al., 2016	5'-3' exoribonuclease 2 involved in transcription termination and R-loop suppression
FIP1_HUMAN	0.9001	Stirling et al., 2012	Pre-mRNA 3'-end-processing factor involved in R-loop suppression

SI References

1. Pontikos N, et al. (2017) Phenopolis: an open platform for harmonization and analysis of genetic and phenotypic data. Bioinformatics 33(15):2421-2423.
2. Wilkes EH, Terfve C, Gribben JG, Saez-Rodriguez J, Cutillas PR. (2015) Empirical inference of circuitry and plasticity in a kinase signaling network. *Proc NatlAcad Sci U S A* 112(25):7719-7724.
3. Casado P, et al. (2013) Kinase-substrate enrichment analysis provides insights into the heterogeneity of signaling pathway activation in leukemia cells. *Science signalling* 6(268):rs6.
4. Shannon P, et al. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 13(11):2498-2504.
5. The UniProt Consortium. (2017) UniProt: the universal protein knowledgebase. *Nucleic Acids Res* 45(D1):D158-D169.

6. Bunch H, Lawney BP, Lin YF, et al. (2015) Transcriptional elongation requires DNA break-induced signalling. *Nat Commun* 16(6):10191.
7. Cawthon RM. (2009) Telomere length measurement by a novel monochrome multiplex quantitative PCR method. *Nucleic Acids Res* 37:e21.
8. Chakraborty P, Grosse F. (2011) Human DHX9 helicase preferentially unwinds RNA-containing displacement loops (R-loops) and G-quadruplexes. *DNA Repair (Amst)*. 10(6):654-665.
9. Cristini A, Groh M, Kristiansen MS, Gromak N. (2018) RNA/DNA Hybrid Interactome Identifies DXH9 as a Molecular Player in Transcriptional Termination and R-Loop-Associated DNA Damage. *Cell Rep.* 23(6):1891-1905.
10. Shivji MKK, Renaudin X, Williams ÇH, Venkitaraman AR. (2018) BRCA2 Regulates Transcription Elongation by RNA Polymerase II to Prevent R-Loop Accumulation. *Cell Rep.* 22(4):1031-1039.
11. Skourti-Stathaki K, Proudfoot NJ, Gromak N. (2011) Human senataxin resolves RNA/DNA hybrids formed at transcriptional pause sites to promote Xrn2-dependent termination. *Mol Cell.* 42(6):794-805.
12. Sollier J, et al. (2014) Transcription-coupled nucleotide excision repair factors promote R-loop-induced genome instability. *Mol Cell* 56(6):777-785.
13. Bhatia V, et al. (2014) BRCA2 prevents R-loop accumulation and associates with TREX-2 mRNA export factor PCID2. *Nature*. 511(7509):362-365.
14. Stirling PC, et al (2012) R-loop-mediated genome instability in mRNA cleavage and polyadenylation mutants. *Genes Dev.*, 26, 163–175.
15. Morales JC, et al (2016). XRN2 links transcription termination to DNA dam- age and replication stress. PLoS Genet. *12*, e1006107.