Supplementary Information.

Mutations in *PNKP* cause microcephaly, seizures and defects in DNA repair.

Jun Shen[1*], Edward C. Gilmore[2,3*], Christine A. Marshall[1], Mary Haddadin[4], John J. Reynolds[5], Wafaa Eyaid[6], Adria Bodell[1], Brenda Barry[1], Danielle Gleason[2], Kathryn Allen[1], Vijay Ganesh [1], Bernard S. Chang[1], Arthur Grix[7], R. Sean Hill[2], Meral Topcu[8], Keith W. Caldecott[5], A. James Barkovich[9], and Christopher A. Walsh[1,2,10]

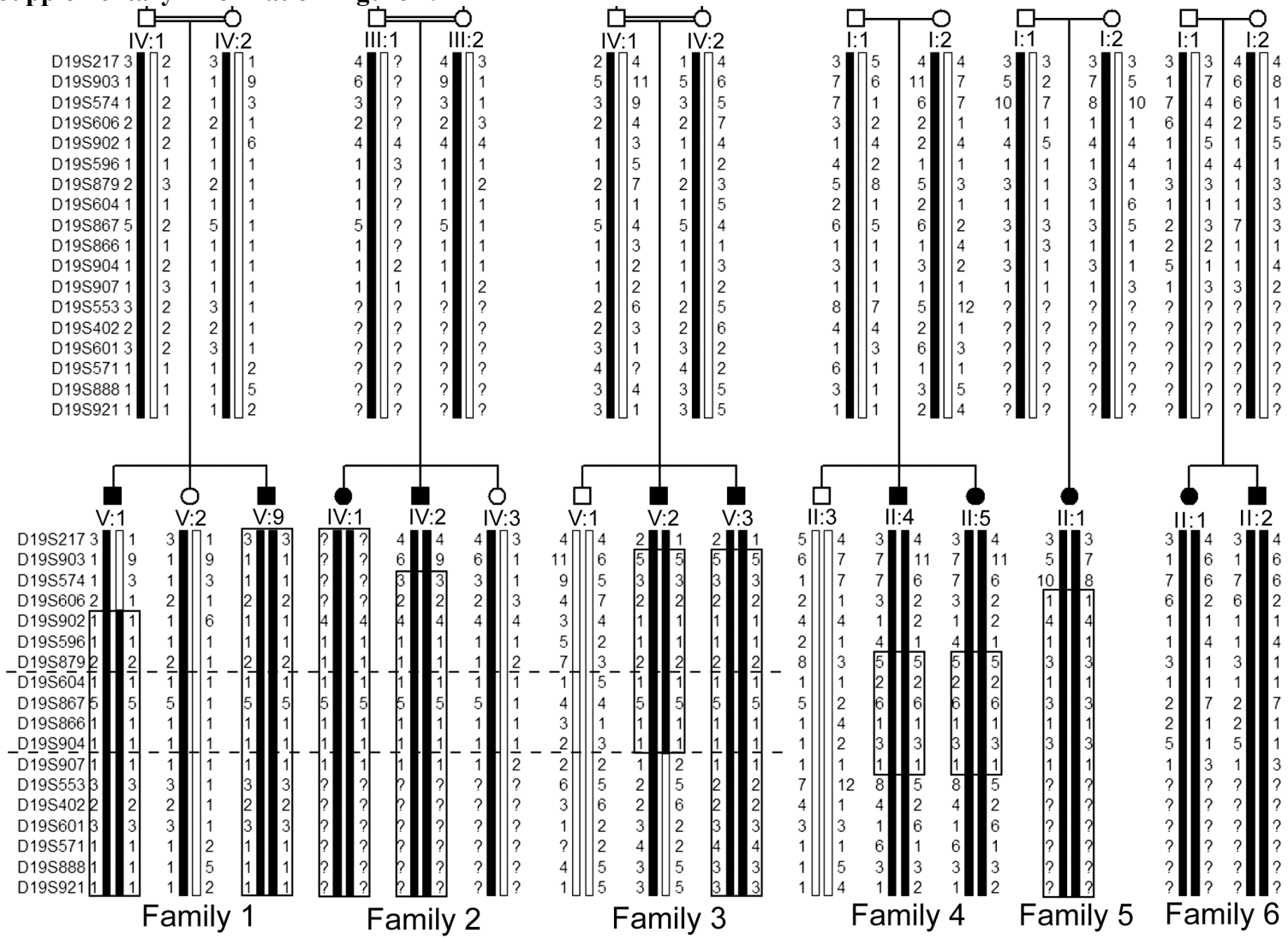
[1]Howard Hughes Medical Institute, Department of Neurology, Beth Israel Deaconess Medical Center and Program in Neuroscience, Harvard Medical School, Boston, MA 02115, USA.
[2]Division of Genetics and The Manton Center for Orphan Disease Research, Department of Medicine, Children's Hospital Boston, Harvard Medical School, Boston, MA 02115, USA.
[3]Division of Child Neurology, Department of Neurology, Massachusetts General Hospital, Harvard Medical School, Boston, MA 02114, USA.
[4]Department of Pathology, Cytogenetics Laboratory, Al-Bashir Hospital, Ministry of Health, Amman, Jordan (Current address: Quest Diagnostics, Nichols Institute, San Juan Capistrano, CA, USA).
[5]Genome Damage and Stability Centre, University of Sussex, Falmer, Brighton, BN1 9RQ, UK
[6]Genetics & Endocrinology, Department of Pediatrics Mail Code 1510, King Fahad National Guard Hospital, King Abdul Aziz Medical City, Saudi Arabia.
[7]Department of Medical Genetics, Kaiser-Permanente Point West Medical Offices, Sacramento, USA.
[8]Hacettepe University, Medical Faculty, Ihsan Dogramaci Children's Hospital, Department of Pediatrics, Section of Pediatric Neurology, Sihhiye 06100, Ankara, Turkey.
[9]Department of Radiology, Department of Neurology, and Department of Pediatrics, University of California at San Francisco, San Francisco, CA 94143-0628, USA
[10]Broad Institute of Massachusetts Institute of Technology and Harvard, Cambridge, MA 02142, USA.

* These authors contributed equally to the work.

Corresponding author: Dr. Christopher A. Walsh
Mailing address: Division of Genetics, Children's Hospital Boston, Center for Life Sciences 14047.1, 3 Blackfan Circle, Boston, MA 02115
Telephone number: 617-919-2923
Fax number: 617-919-2010
Email: Christopher.Walsh@childrens.harvard.edu

Supplemental Information:

Supplementary Information Figure 1.



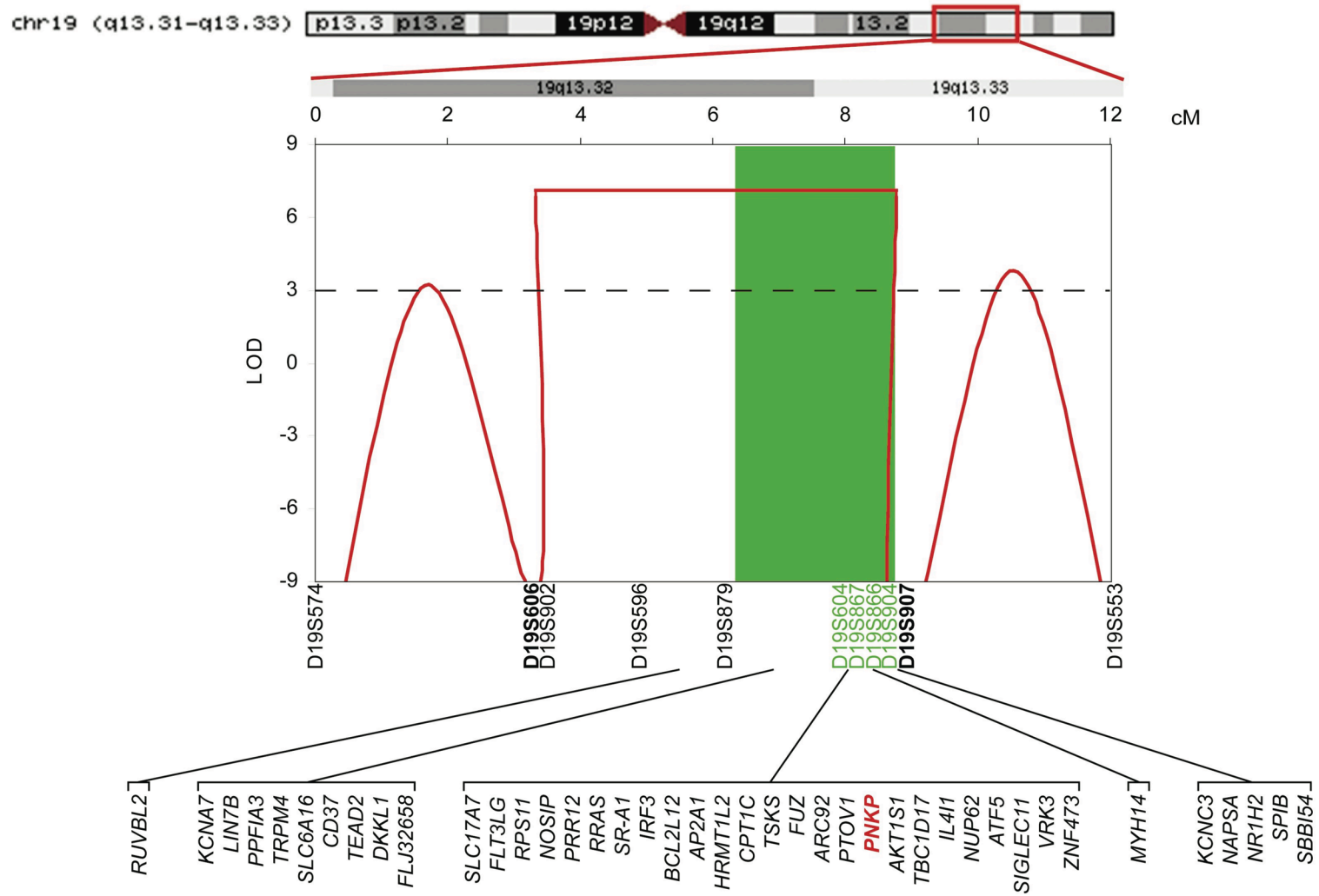
Supplementary Information Figure 1 Legend.

Genotypes of individuals from five MCSZ families at chromosome 19q.

The names of the markers are listed on the left. The marker orders are: cen-D19S217-…-D19S921-tel. The numbers represent the two alleles for each marker. The black bars represent the haplotypes of the disease alleles. The white bars represent the haplotypes of the normal alleles. The boxed regions highlight the consecutive homozygous markers suggesting identity by descent. The region between D19S879 and D19S907 as indicated by the two dashed lines shows linkage and allele sharing among the patients of the three Palestinian families. The common haplotype suggested the presence of a founder mutation, further confirming linkage. The two patients in Family 4, from the Kingdom of Saudi Arabia, shared a homozygous haplotype between D19S596 and D19S553. Therefore, the minimal overlapping homozygous region for the first four families was bounded by D19S596 and D19S907. The fifth family was not reported to be consanguineous and only had one affected child. Nonetheless, the affected child was homozygous for the entire locus defined by the other families, suggesting the possibility that the parents share recent common ancestry. Family 6 was non-consanguineous with two affected children. This family showed linkage under a recessive mode of inheritance for the 19q13.33 locus that was homozygous in Families 1 - 5.

Supplementary Information Figure 2.

Combined Multipoint LOD score at Chromosome 19q13.



Supplementary Information Figure 2 Legend.

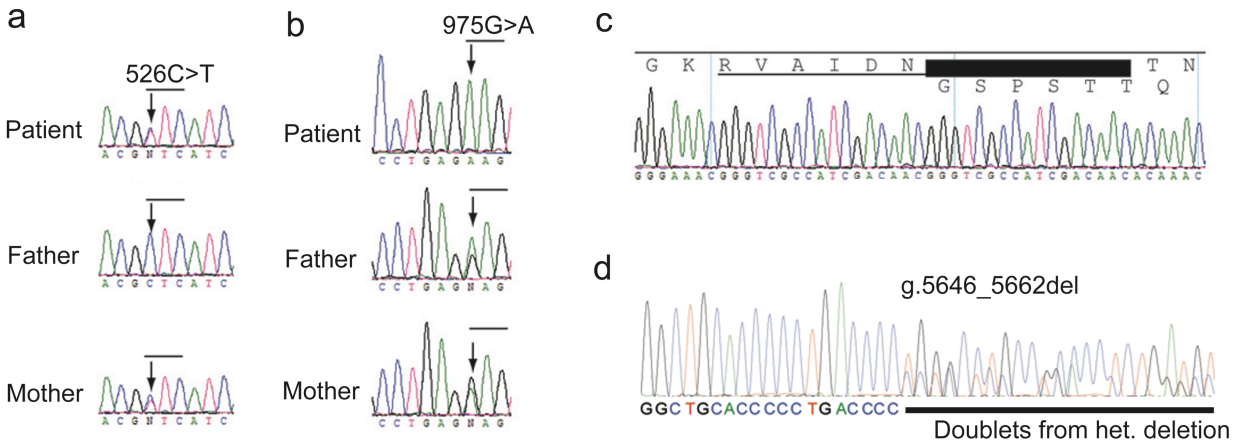
Multipoint LOD score on chromosome 19q (chromosome region shown from UCSC Genome Browser). Linkage analysis was performed on all the families except Family 5 because it was not informative and Family 7 which was ascertained after *PNKP* mutations first were identified. A combined maximum multipoint LOD score of 7.12 was obtained between markers D19S606 and D19S907 shown in bold (calculated under the assumption that parents in Family 4 are distantly related). All patients in the three Palestinian Families 1-3 shared the common homozygous haplotypes of the markers shown in green. The green shaded area between markers D19S879 and D19S907 in the plot indicates the region of allele sharing, which corresponds to the area between the two dashed lines shown in Supplementary Information Figure 1. We chose candidate genes mainly from this region. The candidate genes we screened are listed in the order of their position on the physical map of human chromosome 19 relative to the markers. *PNKP*, is highlighted in red. We sequenced the coding regions and exon-intron boundaries of the following forty-one candidate genes *AKT1S1*, *AP2A1*, *ARC92*, *ATF5*, *BCL2L12*, *CD37*, *CPT1C*, *DKKL1*, *FLJ32658*, *FLT3LG*, *FUZ*, *HRMT1L2*, *IL4I1*, *IRF3*, *KCNA7*, *KCNC3*, *LIN7B*, *NAPSA*, *NOSIP*, *NR1H2*, *NUP62*, *MYH14*, *PNKP*, *POTV1*, *PPFIA3*, *PRR12*, *RPS11*, *RRAS*, *RUVBL2*, *SBBI54*, *SIGLEC11*, *SLC17A7*, *SLC6A16*, *SPIB*, *SR-A1 (SCAF1)*, *TBC1D17*, *TEAD2*, *TRPM4*, *TSKS*, *VRK3* and *ZNF473* for mutations by sequencing the genomic DNA in the MCSZ families. Mutations were only found in *PNKP*.

Table 1. Two point LOD scores (θ =0) for markers on chromosome 19q13.32-q13.33.

Microsatellite Marker [a]	Genetic Map (cM) [b]	Physical Map (Mb) [c]	Family 1	Family 2	Family 3	Family 4	Family 6	Combined
D19S217	68.08	48.97	-∞	0.781	-0.968	0.426	-∞	-∞
D19S903	69.50	49.74	-∞	-1.799	1.680	0.727	0.602	-∞
D19S574	69.50	49.82	-∞	0.921	1.653	0.727	0.602	-∞
D19S606	72.72	52.67	-∞	0.892	1.545	0.125	0.602	-∞
D19S902	72.72	53.02	1.331	0.444	1.545	0.727	0.602	4.65
D19S596	74.07	53.94	0.108	0.682	1.434	0.426	0.301	2.95
D19S879	75.41	54.21	1.165	1.371	1.587	0.727	0.301	5.15
D19S604	75.41	54.59	0.139	0.585	0.982	0.727	0.301	2.73
D19S867	77.54	55.23	1.445	1.327	1.497	0.727	0.602	5.60
D19S866	77.54	55.45	0.100	0.471	0.935	0.426	0.000	1.93
D19S904	78.08	55.47	0.447	0.682	1.434	0.727	0.602	3.89
D19S907	78.08	55.75	0.330	0.928	-∞	0.000	0.602	-∞
D19S553	81.51	56.24	1.574	-∞	-∞	0.727	-∞	-∞
D19S402	83.19	56.86	0.481	-∞	-∞	0.426	-∞	-∞
D19S601	83.19	57.29	1.181	-∞	-∞	0.727	-∞	-∞
D19S571	84.08	57.99	0.487	-∞	-∞	0.426	-∞	-∞
D19S888	85.87	58.35	0.676	-∞	-∞	0.727	-∞	-∞
D19S921	87.66	58.46	0.439	-∞	-∞	0.426	-∞	-∞

[a] The marker orders is cen-D19S217-..-D19S921-tel.
[b] The sex averaged genetic map distances from the Marshfield map were used [1].
[c] The physical map positions were from the human reference sequence at UCSC human genome database hg17 based on NCBI Human Build 35.

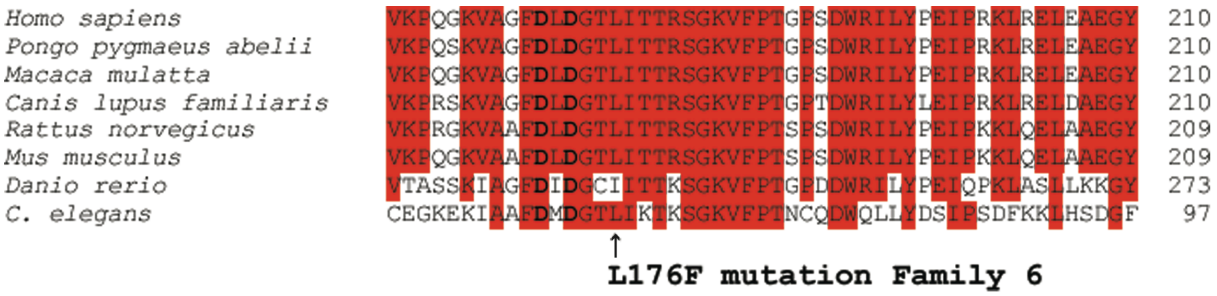
Supplementary Information Figure 3a-d.



Supplementary Information Figure 3a-d Legend.

Representative sequence traces of *PNKP* mutations. (a) Representative sequencing traces show that the patient and the phenotypically normal mother from Family 6 are heterozygous for the L176F (526C>T) missense mutation. The father and the patient are also heterozygous for the 17 bp dup (not shown), making the patient compound heterozygous for the two mutations. (b) Representative sequencing traces show that the patient from Family 3 is homozygous for the E326K (976G>A) missense mutation and that the parents are heterozygous carriers of the mutation. (c) Representative sequencing traces show the homozygous 17 bp dup in a patient from Family 4. The normal 17 bp sequence is marked by the thin line and the duplication by the thick line. The normal translation is shown above the line and the frameshifted translation shown below. (d) Sequencing traces of the heterozygous intronic deletion found in Family 7.

Supplementary Information Figure 3e.



Supplementary Information Figure 3e Legend.

Analysis of *PNKP* L176F mutation. Partial protein sequence alignment of the phosphatase active site from multiple species of PNKP protein from *Homo sapiens* (human) (Entrez clone ID BC033822), *Pongo pygmaeus abelii* (orangutan) (UCSC), *Macaca mulatta* (rhesus) (Ensembl Gene ID ENSMMUG0000001162), *Canis lupus familiaris* (dog) (UCSC), *Rattus norvegicus* (rat) (Entrez clone ID BC079307), *Mus musculus* (mouse) (UCSC), *Danio rerio* (zebra fish)

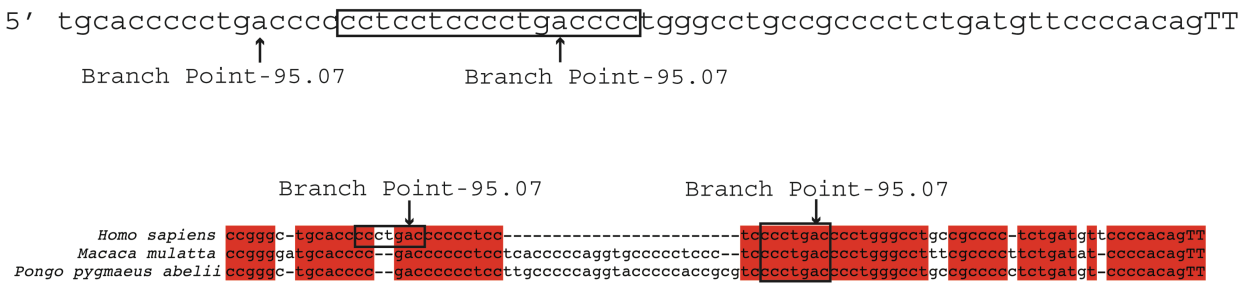
(UCSC) and *C. elegans* (UCSC). All UCSC sequences obtained from N-Scan gene prediction from the genomic sequence. Regions highlighted in red indicate identity to consensus sequence. The region of the phosphatase that contains the L176F mutation contains two aspartates (D171 and D173 in bold) that participate within the active site.

Supplementary Information Figure 3f.

Supplementary Information Figure 3f Legend.

Analysis of *PNKP* E326K mutation. Partial protein alignment of PNKP phosphatase domain region including the 7th alpha helix from the same species shown above. E326K mutation occurs in a conserved region that is part the 7th alpha helix of the protein (boxed region) [2]. The Glu326 (that is mutated) may stabilize this alpha helix via its bonds to Thr323. In addition, while this mutation is near the end of the phosphatase domain in the primary sequence, the crystal structure from mouse shows the 7th alpha helix is adjacent to the active site with Pro324 being within 4 angstroms of Asp 289 that is part of the active site [2]. Therefore, this mutation likely disrupts the 7th alpha helix that potentially perturbs the active site.

Supplementary Information Figure 3g.

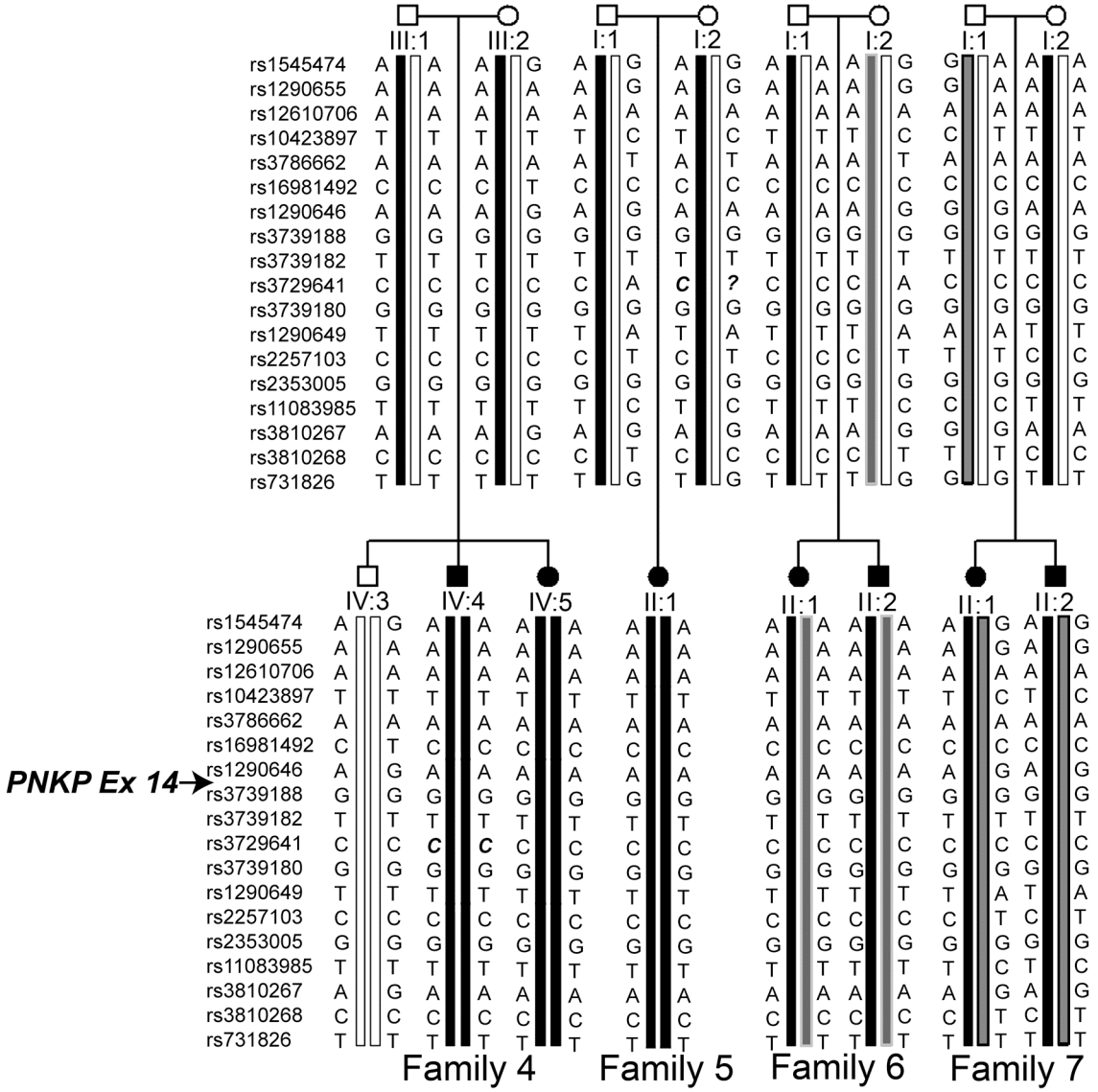


Supplementary Information Figure 3g Legend.

Analysis of *PNKP* intron 15 deletion. The 17 bp deletion within intron 15 (lower case) in boxed area (top sequence) contains an adenine that is the predicted normal branch point involved in RNA splicing to the exon acceptor of exon 16 (upper case TT to the right). This A is a perfect match for a branch point consensus sequence and is within 50 nucleotides of the exonic splice site [3]. This adenine is also the predicted branch point using splicing prediction software (Human Splicing Finder Version 2.3)[4]. The branch point score shown below the sequence is based upon the heptamers containing adenines with scores greater than 70 considered being good candidates and 100 the maximal score (Human Splicing Finder Version 2.3). The deletion brings a different adenine within 50 nucleotides of the exon 16 (middle sequence) that has an equally good score. However, the efficiency of a site for splicing has many other influences including splicing enhancers [3] many of which are potentially lost after the deletion (not shown). To further clarify which site may act as a branch point we compared human, rhesus and orangutan sequences within this region (bottom sequences). The two heptamers (boxed areas) that are used to calculate branch point strength (based upon consensus sequence and calculated for human only) are shown. The heptamer that is disrupted by the deletion in Family 7 (on the right) and surrounding region is conserved. However, the alternative branch point further 5' (on the left) is

not conserved in these primates, further supporting that the deleted adenine is the normal branch point. Abnormal splicing was confirmed in Figure 3c.

Supplementary Information Figure 4a



Supplementary Information Figure 4a Legend.

SNP haplotypes associated with Exon 14 17 bp dup. Pedigrees from Families 4, 5, 6, and 7 with black bars representing Exon 14 17 bp dup mutations, gray bars other disease causing alleles, and white bars non-disease causing alleles. The SNPs span a distance of ~35 kb. Alleles associated with Exon 14 17 dup mutations are all the same. 3 nucleotide calls that were not directly determined, but deduced from inheritance patterns, are shown in Italics and Bold. One nucleotide (marked with ?) could not be determined but was not on a disease-carrying allele.

Supplementary Information Figure 4b.

rs1545474	A	A	G	G	A	G	A	A	G	G	G	G	A	G
rs1290655	A	A	G	G	G	G	A	A	G	A	G	A	G	G
rs12610706	A	A	A	G	A	A	A	A	A	A	G	A	G	A
rs10423897	T	T	C	C	T	C	T	T	C	T	C	C	C	C
rs3786662	A	A	A	A	A	T	T	A	T	A	A	A	A	A
rs16981492	C	C	C	C	C	C	C	C	C	C	C	T	C	C
rs1290646	A	G	G	G	A	G	A	G	G	A	G	G	G	G
rs3739188	G	G	G	G	G	G	G	G	G	G	G	G	G	G
rs3739182	T	T	T	T	T	T	T	T	T	T	T	C	T	T
rs3729641	C	C	C	A	C	A	C	C	C	C	C	C	A	A
rs3739180	G	G	G	G	G	G	G	G	G	G	G	G	G	G
rs1290649	T	A	A	A	T	A	T	T	T	T	T	T	A	A
rs2257103	C	T	T	T	C	T	C	C	C	C	C	C	T	T
rs2353005	G	G	G	A	G	G	G	G	G	G	G	G	A	A
rs11083985	T	C	C	C	T	C	T	T	T	T	T	T	C	C
rs3810267	A	G	G	G	A	G	A	A	A	A	A	G	G	G
rs3810268	C	T	T	T	C	T	C	C	C	C	C	C	T	T
rs731826	T	G	G	G	T	G	T	T	T	T	T	T	G	G
Haplotypes	144	71	66	62	55	54	39	24	19	16	16	14	9	8
Freq	0.2188	0.1079	0.1003	0.0942	0.0836	0.0821	0.0593	0.0365	0.0289	0.0243	0.0243	0.0213	0.0137	0.0122
rs1545474	A	G	G	G	G	G	A	A	A	A	G	G	G	G
rs1290655	A	A	G	A	A	G	A	A	A	A	A	A	A	A
rs12610706	A	A	A	A	A	G	A	A	A	A	A	A	A	A
rs10423897	T	C	C	C	C	C	C	T	T	T	C	C	C	C
rs3786662	A	A	A	A	A	T	A	A	A	T	A	A	A	A
rs16981492	C	T	C	C	C	C	C	C	C	C	C	T	T	T
rs1290646	G	G	G	G	G	G	G	A	G	G	G	G	G	G
rs3739188	G	G	G	A	G	G	G	G	G	G	G	G	G	G
rs3739182	T	T	T	T	T	T	T	T	C	T	T	T	T	T
rs3729641	C	C	A	C	C	A	C	A	C	C	C	A	C	C
rs3739180	A	G	G	A	G	G	G	G	G	G	G	G	G	G
rs1290649	T	T	A	T	T	A	A	A	T	A	T	A	A	T
rs2257103	C	C	T	C	C	T	T	T	C	T	C	T	T	C
rs2353005	G	G	G	G	G	A	G	A	G	G	G	A	G	G
rs11083985	T	T	C	T	T	C	C	C	T	C	T	C	C	T
rs3810267	G	A	G	G	G	G	G	G	G	G	G	G	G	G
rs3810268	T	C	T	C	T	T	T	T	C	T	C	T	T	C
rs731826	T	T	G	T	T	G	G	G	T	G	T	G	G	T
Haplotypes	7	7	6	3	3	3	2	2	2	2	2	2	2	2
Freq	0.0106	0.0106	0.0091	0.0046	0.0046	0.0046	0.0030	0.0030	0.0030	0.0030	0.0030	0.0030	0.0030	0.0030
rs1545474	G	G	A	A	A	G	G	G	G	G	G	G	G	G
rs1290655	G	G	A	A	A	A	A	A	A	G	G	G	G	G
rs12610706	A	G	A	A	A	A	A	A	A	A	A	A	A	G
rs10423897	T	C	T	T	T	C	C	T	T	C	C	C	T	C
rs3786662	A	A	A	A	A	A	A	A	A	A	A	T	A	A
rs16981492	C	C	C	C	T	C	C	C	C	C	C	C	C	C
rs1290646	G	G	G	G	A	A	G	G	G	G	G	A	G	G
rs3739188	G	G	G	G	G	G	G	G	G	G	G	G	G	G
rs3739182	T	T	T	T	T	T	T	T	T	T	T	T	T	C
rs3729641	C	C	A	C	C	C	C	A	C	C	C	C	C	C
rs3739180	G	G	G	G	G	G	G	G	G	A	G	G	A	G
rs1290649	T	A	A	T	T	T	T	A	T	T	T	T	T	T
rs2257103	C	T	T	C	C	T	C	T	C	C	C	C	C	C
rs2353005	G	G	G	G	G	G	G	G	G	G	G	G	G	G
rs11083985	T	C	C	C	T	C	T	C	T	T	T	T	T	T
rs3810267	A	G	G	G	A	G	A	G	A	G	G	A	G	G
rs3810268	C	T	T	T	C	T	C	T	C	T	C	C	T	C
rs731826	T	G	G	G	T	G	T	G	T	T	T	T	T	T
Haplotypes	2	2	1	1	1	1	1	1	1	1	1	1	1	1
Freq	0.0030	0.0030	0.0015	0.0015	0.0015	0.0015	0.0015	0.0015	0.0015	0.0015	0.0015	0.0015	0.0015	0.0015

Supplementary Information Figure 4b Legend.

Haplotype Frequencies from Normal Controls. Haplotypes frequencies determined from 329 normal controls (171 European and 158 Middle Eastern). Haplotype frequencies were determined via their genotypes using Phase [5-7]. The haplotype associated with the Exon 14 17 bp dup was the most common, frequency=0.2188. However, the chance that the 17 bp duplication occurred independently on four different occasions on that same haplotype is low, $1*(0.2188)^3=0.0105$, or approximately 1%. This is calculated by assuming that the first mutation could occur on any haplotype, therefore 1. The chance of each additional mutation occurring on that haplotype is 0.2188 for the three additional mutations, therefore $(0.2188)^3$. Determining European and Middle Eastern haplotype frequency separately did not significantly alter the calculation of this haplotype frequency (data not shown). This supports the conclusion that the Exon 14 17 bp dup mutation more likely occurred once in the distant past (as estimated from the lack of concordance of the microsatellite markers) rather than on multiple independent occasions.

Supplementary Information Figure 5.

% Tail DNA before and after hydrogen peroxide treatment.

	Sample	II:1	II:1	II:1	II:1	II:4	II:4	II:4	II:4	I:1	I:1	I:1	I:1	II:3	II:3	II:3	II:3
	PNKP Genotype	-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-	+/-	+/-	+/-	+/-	+/+	+/+	+/+	+/+
		Well 1	Well 2	Well 3	Combined	Well 1	Well 2	Well 3	Combined	Well 1	Well 2	Well 3	Combined	Well 1	Well 2	Well 3	Combined
No Tx	Average	11.67	17.05	16.49	15.07	7.20	5.58	4.61	5.84	7.17	13.67	10.77	10.50	9.35	8.21	3.54	7.05
	Median	1.81	7.64	7.64	6.58	0.14	0.84	0.66	0.56	0.16	1.29	0.63	0.58	0.44	0.38	0.76	0.52
	Stand Dev	17.64	22.59	21.14	20.63	20.80	15.82	14.98	17.50	20.74	30.00	27.41	26.32	22.44	22.98	6.20	19.09
	Stand Error	1.98	2.51	2.42	2.34	2.17	1.77	1.63	1.83	2.29	3.38	3.03	2.92	2.56	2.41	0.69	2.19
	Count	79	81	76	236	92	80	84	256	82	79	82	243	77	91	81	249
T-Test					**0.03074210**				0.19850799								
0 min.	Average	55.94	55.01	57.45	56.13	77.31	84.26	75.13	78.75	34.84	31.82	39.12	35.34	50.13	46.37	52.19	49.59
	Median	50.02	57.35	65.32	58.45	97.89	96.20	91.12	95.06	26.48	15.76	31.92	25.69	46.09	40.49	51.22	46.42
	Stand Dev	41.37	41.29	40.83	41.01	32.44	24.75	30.35	29.56	33.15	33.39	32.11	32.87	34.68	33.26	31.48	33.08
	Stand Error	4.63	4.65	4.59	4.61	3.67	2.86	3.33	3.37	3.85	3.71	3.48	3.85	3.93	3.70	3.43	3.77
	Count	80	79	79	238	78	75	83	236	74	81	85	240	78	81	84	243
T-Test					**0.02944037**				**0.00024095**								
15 min.	Average	46.84	56.75	46.84	50.26	71.85	70.95	76.09	72.82	27.19	33.12	25.93	28.71	30.85	33.22	43.26	35.39
	Median	32.83	77.45	32.83	44.26	93.69	85.15	93.50	91.89	10.88	21.51	11.60	15.01	17.31	22.48	35.54	23.99
	Stand Dev	39.98	42.09	39.98	40.81	35.32	30.73	31.00	32.47	33.45	33.59	29.78	32.32	31.94	33.30	36.54	34.11
	Stand Error	4.65	4.77	4.65	4.78	3.83	3.44	3.68	3.54	3.76	3.83	3.35	3.66	3.21	3.84	4.14	3.45
	Count	74	78	74	226	85	80	71	236	79	77	79	235	99	75	78	252
T-Test					**0.00412241**				**0.00001433**								
30 min.	Average	57.87	53.83	43.31	51.65	64.24	70.87	69.76	68.13	28.87	33.93	27.48	30.07	27.77	33.18	28.78	29.83
	Median	67.40	53.61	36.22	47.48	74.40	88.01	79.65	79.54	12.78	22.95	19.03	17.58	19.39	26.28	21.53	21.74
	Stand Dev	40.25	36.07	36.92	38.14	34.59	31.59	30.54	32.37	34.68	33.06	29.22	32.39	29.37	31.44	26.83	29.23
	Stand Error	4.53	4.11	4.15	4.32	3.73	3.65	3.44	3.51	3.93	3.79	3.33	3.69	3.20	3.61	3.04	3.21
	Count	79	77	79	235	86	75	79	240	78	76	77	231	84	76	78	238
T-Test					**0.00031443**				**0.00000044**								
60 min.	Average	42.02	42.52	40.93	41.82	55.90	56.40	60.51	57.57	16.48	19.89	14.28	16.91	12.54	20.75	21.85	18.48
	Median	31.03	35.48	33.04	33.08	48.05	55.57	59.74	58.04	3.86	9.48	7.76	6.39	6.41	11.71	8.07	8.13
	Stand Dev	34.87	34.91	35.70	35.02	35.12	35.89	33.54	34.78	25.75	26.68	19.38	24.20	15.47	24.12	28.22	23.73
	Stand Error	4.03	3.95	4.02	4.07	3.72	4.06	3.70	3.71	2.90	3.00	2.21	2.74	1.70	2.77	2.90	2.62
	Count	75	78	79	232	89	78	82	249	79	79	77	235	83	76	95	254
T-Test					**0.00001422**				**0.00000079**								
90 min.	Average	26.48	31.03	35.45	30.89	49.79	42.31	48.77	46.94	11.80	12.71	10.26	11.53	11.07	9.96	12.72	11.21
	Median	11.19	20.00	28.63	19.75	54.60	28.17	41.31	42.09	3.88	2.62	2.12	2.78	2.74	3.98	4.00	3.59
	Stand Dev	31.54	32.71	34.48	32.97	37.24	38.13	37.60	37.65	18.31	22.67	19.38	20.05	19.85	16.91	21.11	19.27
	Stand Error	3.48	3.70	3.93	3.66	4.22	4.29	4.26	4.29	1.97	2.57	2.03	2.17	2.19	1.87	2.42	2.14
	Count	82	78	77	237	78	79	78	235	86	78	91	255	82	82	76	240
T-Test					**0.00001423**				**0.00000014**								

% Tail DNA before and after camptothecin treatment.

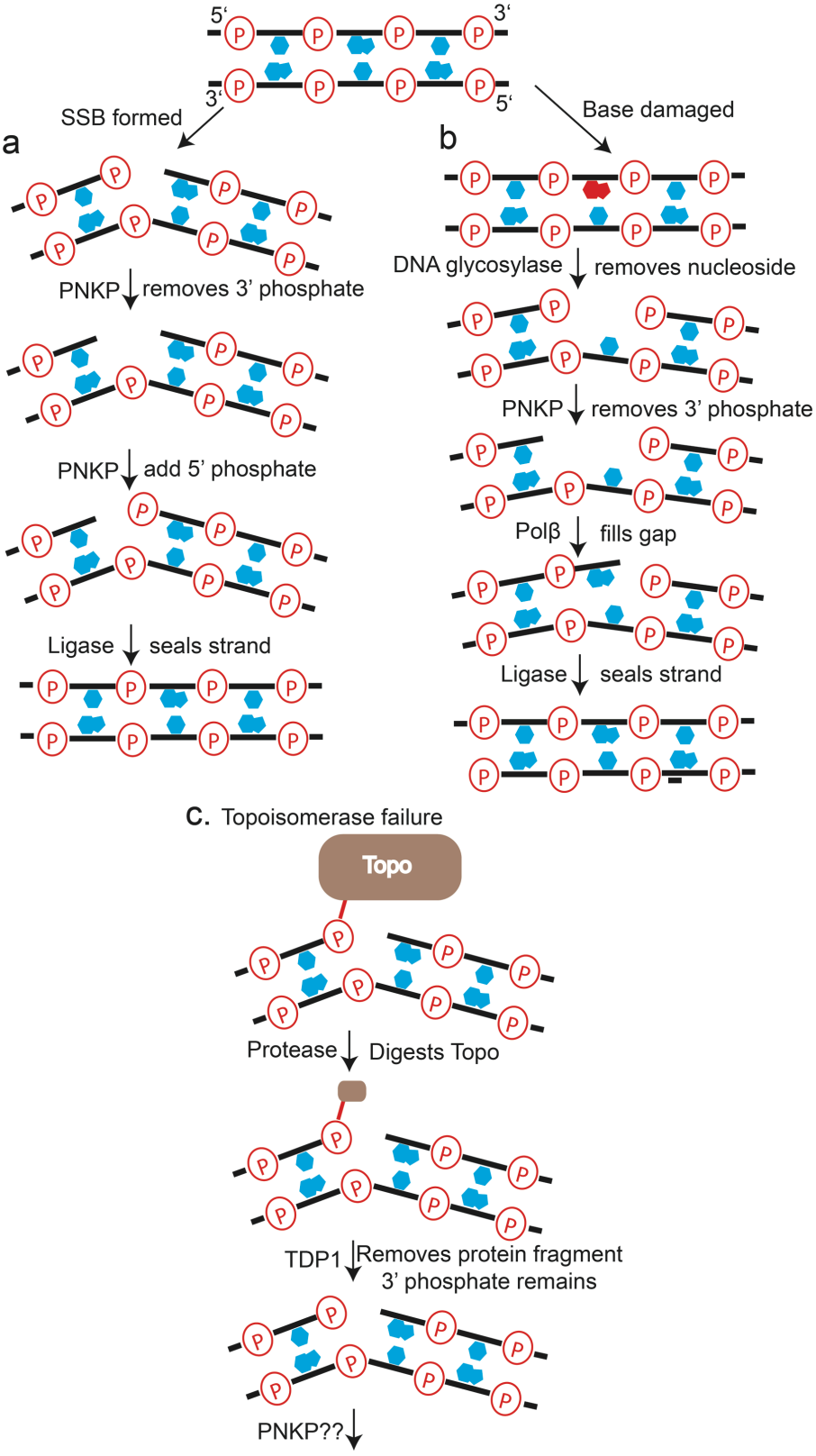
		II:1	II:1	II:1	II:1	II:4	II:4	II:4	II:4	I:1	I:1	I:1	I:1	II:3	II:3	II:3	II:3
	PNKP Genotype	-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-	+/-	+/-	+/-	+/-	+/+	+/+	+/+	+/+
		Well 1	Well 2	Well 3	Combined	Well 1	Well 2	Well 3	Combined	Well 1	Well 2	Well 3	Combined	Well 1	Well 2	Well 3	Combined
No Tx	Average	12.02	10.86	15.18	12.64	8.80	13.57	7.07	9.80	3.64	7.41	10.57	7.22	8.46	5.51	5.09	6.46
	Median	0.47	1.28	3.76	1.51	0.82	0.87	0.80	0.81	0.13	0.41	0.43	0.30	0.54	0.55	0.24	0.46
	Stand Dev	26.11	20.96	25.22	24.26	19.61	25.05	17.67	21.08	8.44	17.44	24.00	18.00	20.17	14.90	13.54	16.67
	Stand Error	2.75	2.37	2.87	2.57	2.17	2.82	1.99	2.34	0.91	1.95	2.57	1.95	2.06	1.68	1.47	1.71
	Count	90	78	77	245	82	79	79	240	86	80	87	253	96	79	85	260
T-Test					**0.011256236**			0.091424219									
0 min.	Average	46.87	51.86	48.18	48.90	51.76	53.42	51.48	52.25	43.46	36.96	36.84	39.15	37.06	37.74	39.53	38.11
	Median	44.83	52.02	43.41	47.08	50.98	52.52	50.79	51.18	42.94	39.05	38.56	40.10	38.48	35.70	37.68	37.57
	Stand Dev	17.75	17.37	22.55	19.45	18.48	22.00	19.39	19.99	20.95	17.24	18.16	19.06	17.25	17.73	18.18	17.68
	Stand Error	1.97	2.01	2.51	2.17	2.04	2.37	2.20	2.22	2.29	1.92	2.03	2.09	1.97	2.03	2.07	2.03
	Count	81	75	81	237	82	86	78	246	84	81	80	245	77	76	77	230
T-Test					**0.004485079**			**0.000257789**									
7.5 min.	Average	43.93	39.86	35.46	39.74	41.01	42.03	38.95	40.68	16.33	18.86	17.84	17.63	14.91	17.85	13.98	15.56
	Median	45.73	40.73	34.96	40.20	40.77	41.27	40.63	40.77	12.43	18.34	14.15	14.86	9.59	12.11	9.59	9.86
	Stand Dev	23.63	20.21	17.08	20.65	19.74	21.81	21.59	20.99	13.17	12.00	13.36	12.88	15.00	17.91	13.79	15.67
	Stand Error	2.61	2.02	1.87	2.29	2.21	2.49	2.49	2.36	1.42	1.39	1.44	1.40	1.55	1.97	1.53	1.62
	Count	82	100	83	265	80	77	75	232	86	75	86	247	94	83	81	258
T-Test					**7.18852E-06**			**2.35175E-06**									
15 min.	Average	24.12	26.53	29.89	26.69	34.36	34.60	32.44	33.83	5.49	11.08	9.92	8.87	6.34	8.15	14.32	9.53
	Median	19.58	25.12	25.91	23.92	32.52	34.39	26.50	28.69	1.43	3.09	2.85	2.29	1.35	1.65	5.57	2.21
	Stand Dev	21.44	19.74	23.34	21.60	23.00	26.73	26.69	25.49	13.27	19.01	17.24	16.82	11.55	12.54	19.29	15.12
	Stand Error	2.25	2.26	2.66	2.28	2.66	2.92	3.08	2.96	1.47	2.06	1.92	1.88	1.30	1.43	2.23	1.71
	Count	91	76	77	244	75	84	75	234	81	85	81	247	79	77	75	231
T-Test					**0.000100485**			**4.6809E-05**									
30 min.	Average	18.58	18.79	17.76	18.39	23.09	22.18	16.06	20.50	7.90	7.68	6.25	7.32	8.47	8.09	9.77	8.79
	Median	8.24	6.81	9.63	8.66	13.87	11.58	6.67	10.51	1.82	1.48	1.15	1.48	1.25	1.68	1.79	1.52
	Stand Dev	20.91	26.04	23.01	23.14	25.63	25.16	19.07	23.64	16.12	17.84	12.90	15.72	16.95	12.46	18.45	16.18
	Stand Error	2.11	2.95	2.61	2.35	2.87	2.85	2.19	2.66	1.57	1.99	1.42	1.54	1.87	1.42	2.02	1.80
	Count	98	78	78	254	80	78	76	234	105	80	83	268	82	77	83	242
T-Test					**1.75644E-06**			**3.01488E-06**									
45 min.	Average	9.30	6.68	11.50	9.10	10.48	11.68	15.34	12.50	6.58	8.53	4.98	6.66	8.45	8.80	4.37	7.23
	Median	2.00	1.53	4.54	2.38	2.75	4.57	4.89	4.13	1.44	0.44	0.46	0.63	1.40	0.41	0.52	0.60
	Stand Dev	17.34	10.07	17.56	15.31	16.86	16.51	22.22	18.76	9.96	20.15	10.42	13.99	12.97	19.61	9.19	14.72
	Stand Error	2.00	1.10	1.99	1.78	1.87	1.87	2.48	2.10	1.02	2.28	1.14	1.44	1.44	2.14	1.02	1.65
	Count	75	84	78	237	81	78	80	239	96	78	83	257	81	84	81	246
T-Test					0.178236809			**0.031341402**									

Supplementary Information Figure 5 Legend

Data from comet assays. Data from EBV transformed lymphocytes derived from Family 7. Cell lines derived from II:1 and II:4 (listed across the top of the table) are MCSZ patients and share identical compound heterozygous mutations while cell lines derived from I:1 and II:3 are unaffected being

heterozygous for the intron 15 17 bp deletion (splicing mutation) or no mutation (see Figure 3 for more detailed description of mutations). Cells were collected prior to drug exposure or treated with 100 μM hydrogen peroxide for 30 minutes or 10 μM camptothecin 60 minutes. After treatment, cells were collected for the 0 minute recovery time point or washed and resuspended in media and incubated at 37^{o} C. Cells were collected at time points listed in left column of the table. The % Tail DNA was measured in 3 separate wells per time point per cell line. The average, median, standard deviation, standard error and number of cells measured were calculated for each well and for all three wells combined. Samples derived from MCSZ patients consistently had statistically significantly (significant P values in bold) higher % Tail DNA after DNA toxins were removed indicating that the damage was slower to be repaired. The MCSZ patient derived cells were eventually able to recover from camptothecin but not from hydrogen peroxide treatment. P-values were determined by comparing the average % Tail DNA from 3 separate wells compared to 6 wells of control cell lines using T-Test with two-tailed distribution and homoscedastic test.

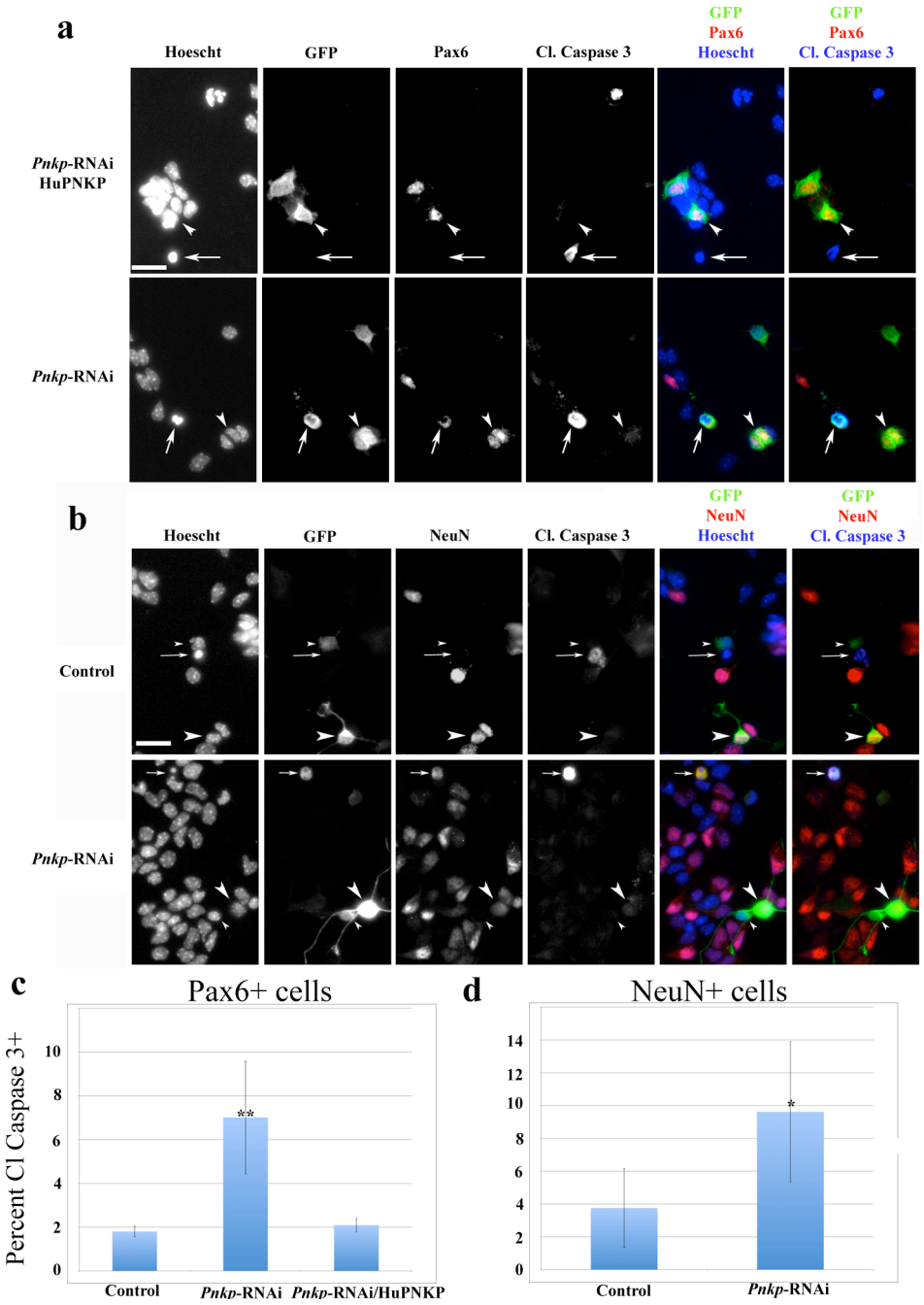
Supplementary Information Figure 6.



Supplementary Information Figure 6 Legend.

Model of PNKP function. At the top is a diagram of normal DNA. 6a shows the most basic SSB pathway requiring PNKP. DNA damage causes a single strand break that splits the phosphodiester bond leaving the phosphate on the 3' hydroxyl group. PNKP removes the 3' phosphate and phosphorylates the 5' phosphate, allowing ligase to re-seal the phosphodiester backbone. 6b is an example of BER after damage to a base. Some DNA glycosylases repair this type of damage by removing the damaged nucleoside (both base and ribose) leaving both 3' and 5' phosphates. The 3' phosphate is then removed by PNKP, the empty nucleoside is filled by DNA polymerase β and ligated via ligase III all of which are coordinated by XRCC1[8-12]. PNKP performs the same modification when required for DSB repair by NHEJ via its interaction with XRCC4. 6c is the current model of the defect in SCAN1 patients [13]. Topoisomerase I (Topo) relaxes DNA by forming a covalent bond with the 3' phosphate and inducing a DNA break in one strand allowing tension to unwind around the other strand and then reconnects the 3' phosphate to the 5' hydroxyl group as Topo separates from DNA. If DNA or RNA polymerase is near the Topo, excessive DNA damage is present or the cells are treated with the DNA toxin (Topo inhibitor) camptothecin, the Topo molecule becomes stuck upon the phosphate group and cannot complete the re-ligation step. Topo is then is proteolyzed to a small covalently linked peptide that inhibits further DNA repair. TDP1 removes the proteolyzed fragment. PNKP is thought to be required for further processing.

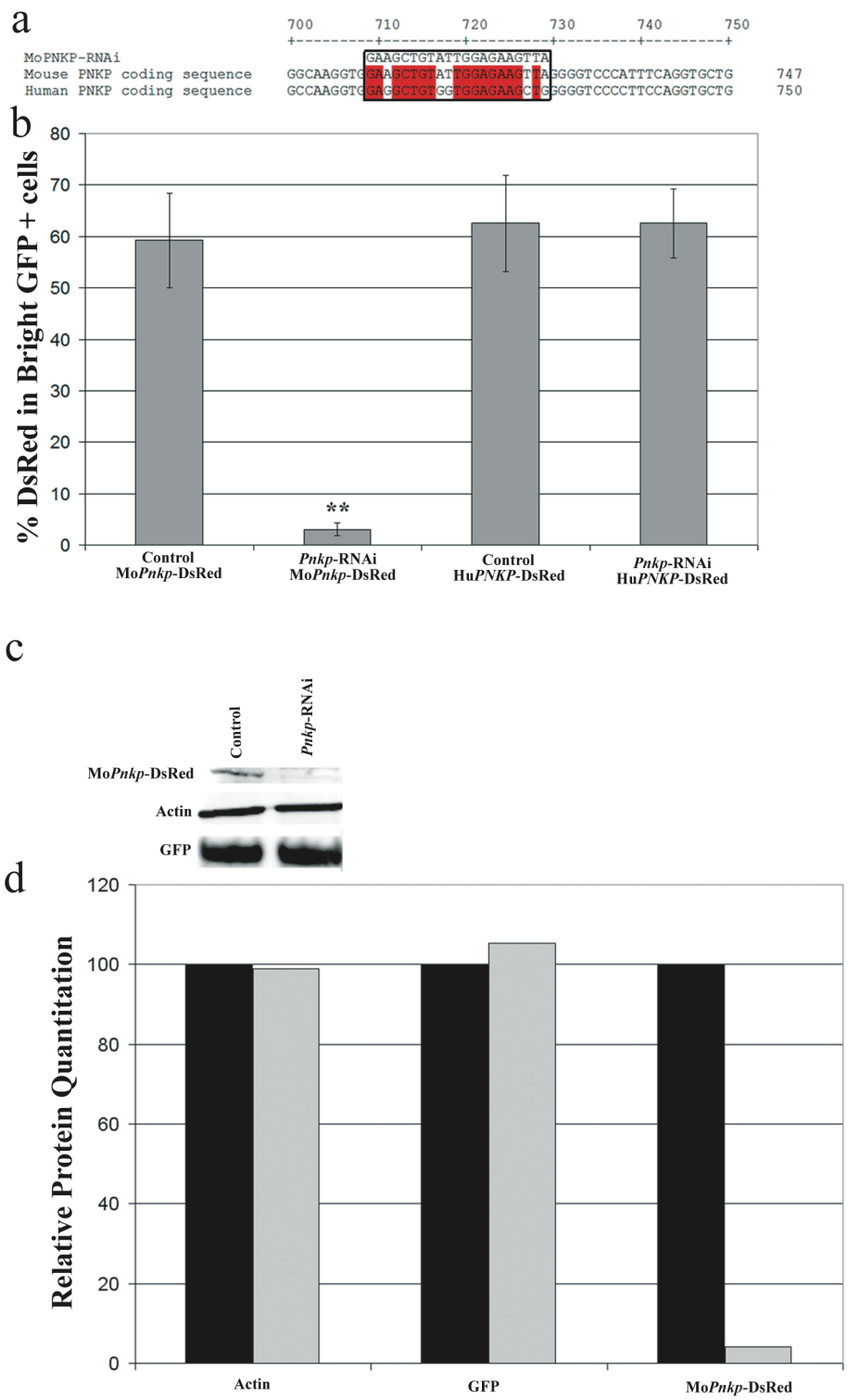
Figure 7.



Supplementary Information Figure 7 Legend. *Pnkp* RNAi induces apoptosis of neuronal progenitors and post-mitotic neurons. Mouse E13.5 cerebral cortical cells were transfected with pSilencer-GFP[14] that generates RNAi directed to mouse (mRNA) *Pnkp* or control vector (see Supplementary Information Figure 8 for details of *Pnkp* RNAi specificity) with or without

human *PNKP* cDNA. Cells were cultured for one day. Cells were fixed and stained with Hoechst (DNA stain for nuclei), GFP (transfected cells), Pax6 (neuronal precursors[15]) or NeuN (differentiated neurons[16]), and cleaved caspase 3 (CC3, apoptotic marker[17]). Examples of staining can be seen in Figure 7a and 7b. First, cells were examined for GFP status, if positive, then cell marker status, either Pax6 or NeuN, and finally for immunoreactivity with CC3. The Hoechst nuclear staining was used to distinguish adjacent cells. For example, in 7a, lower panel of *Pnkp*-RNAi, the arrow points to a GFP+, Pax6+, CC3+ cell, while the arrowhead points to two adjacent GFP+, Pax6+, CC3- cells. 7c shows quantification of the percentage of GFP+/Pax6+ cells that are positive for CC3 when transfected with RNAi-control, Mo*Pnkp*-RNAi, or Mo-*Pnkp*-RNAi and Hu*PNKP*. There is statistically significantly more death when Mo*Pnkp*-RNAi is present and that death is rescued by human *PNKP* demonstrating the specificity of the Mo*Pnkp*-RNAi effect. 7d shows the percentage of GFP+/NeuN+ cells that are CC3+, again demonstrating that when RNAi lowers endogenous *Pnkp* levels in neurons, apoptosis is more likely to occur. Most CC3+ cells contain no markers because of protein degradation during apoptosis, therefore apoptotic rates are likely underestimated. Since Pax6 and NeuN antigens may have different half-lives during apoptosis, rates of their detection in our system could vary between these markers. Therefore, we did not attempt to compare rates of apoptosis between neuronal precursors and differentiated neurons. Error bars are standard error. ** ($P<0.00001$) and *($P=0.004$) (G-test of goodness-of-fit)[18]. Bar=40 μm. See Supplementary Table 2 for raw cell counts.

Supplementary Information Figure 8.



Supplementary Information Figure 8 Legend.

***Pnkp* RNAi specificity.** (a) shows the region of targeting for the RNAi construct. It is targeted to mouse *Pnkp* with 21 of 21 nucleotides being identical to mouse. The human *PNKP* cDNA sequence shows 5 nucleotides that are different between these two species in this region. Targeting of the mouse *Pnkp* RNAi was demonstrated by reduction in expression of a Pnkp-DsRed fusion protein. The graph in b shows results of cell counts with double transfection of control RNAi (empty vector)/mouse (Mo)-*Pnkp*-DsRed fusion cDNA, Mo*Pnkp*-RNAi/Mo*Pnkp*-DsRed, RNAi-control/human (Hu)-*PNKP*-DsRed fusion, and Mo*Pnkp*-RNAi/Hu*PNKP*-DsRed. Cells that were strongly positive for GFP were scored as positive or negative for DsRed (Mo*Pnkp*-DsRed or Hu*PNKP*-DsRed). While the Mo*Pnkp*-RNAi dramatically reduced the levels of Mo*Pnkp*-DsRed (** P-values<<0.0001), it had little to no effect upon Hu*PNKP*-DsRed (P-value=0.39). (c) To further confirm protein knockdown, a Western blot was performed with protein derived from NIH 3T3 cells transfected with Mo*Pnkp*-DsRed and either RNAi-control or Mo*Pnkp*-RNAi. Proteins were detected with anti-DsRed (size expected with PNKP and DsRed fusion was seen), anti-actin and anti-GFP antibodies. The levels were quantified using Li-Cor imaging system. The relative levels of Actin and GFP were not different, but the amount of DsRed-Mo*Pnkp* was much lower in the presence of Mo*Pnkp*-RNAi reduced further documenting the ability of Mo-*Pnkp*-RNAi to lower PNKP levels.

Supplementary Table 2.

Experiment 1			
	GFP+/Pax6+/CC3-	GFP+/Pax6+/CC3+	% CC3+
Mo*Pnkp*-RNAi	80	11	12.09
RNAi-Control	90	2	2.17
Mo*Pnkp*-RNAi+Hu*PNKP*	108	3	2.70
Experiment 2			
	GFP+/Pax6+/CC3-	GFP+/Pax6+/CC3+	% CC3+
Mo*Pnkp*-RNAi	92	5	5.15
RNAi-Control	104	2	1.89
Mo*Pnkp*-RNAi+Hu*PNKP*	106	2	1.85
Experiment 3			
	GFP+/Pax6+/CC3-	GFP+/Pax6+/CC3+	% CC3+
Mo*Pnkp*-RNAi	203	8	3.79
RNAi-Control	146	2	1.35
Mo*Pnkp*-RNAi+Hu*PNKP*	228	4	1.72
Experiment 1			
	GFP+/NeuN+/CC3-	GFP+/NeuN+/CC3+	% CC3+
Mo*Pnkp*-RNAi	122	10	7.58
RNAi-Control	128	4	3.03
Experiment 2			
	GFP+/NeuN+/CC3-	GFP+/NeuN+/CC3+	% CC3+
Mo*Pnkp*-RNAi	56	2	3.45
RNAi-Control	44	0	0.00
Experiment 3			
	GFP+/NeuN+/CC3-	GFP+/NeuN+/CC3+	% CC3+
Mo*Pnkp*-RNAi	83	18	17.82
RNAi-Control	78	7	8.24

Supplementary Table 2.

Raw cell counts from experiments shown in Figure 7.

Supplementary Table 3

Primers.

	Forward	Reverse
Sequencing Primers		
Exon 2	GAGCAGTTAATGGTGGGGAA	ACATCCTCCCTAAATCGGCT
Exon 3	TGGCTACTGCATGTCAAGGA	ACCTACTGGGCTCAAGGGAT
Exon 4	GGAGAAGGGCTGATTTTTCC	GCTAAAAAGTTGAGAGCACGC
Exons 5 and 6	TGAGATGGAGGTGCTCTGTG	AGCAAAGGGTGTCACCAGAC
Exon 7	GCCGTAGGATCTTGTACCCA	AGCTTTAGCTCCCCCTCAAG
Exon 8	GAGGGCTTTAGACTCCAGGG	TTCAGCTCAAGAATGCGCTA
Exons 9 and 10	AAGTGGCTGGAGAGTGAGGA	CCAAGGTTGAGGGCAAACTA
Exon 11	GAAAGACTTCTCCTGCGCC	GACACTTACCCCCAGGGAAT
Exon 12	CACGCCTGAGGAGTTCTTTC	GTTGAGAGGTGGAGATGGGA
Exon 13	GATTCCCTGGGGGTAAGTGT	TCTGGGTTTGTGTTGTCGAT
Exon 14	TCCCATCTCCACCTCTCAAC	CTCGAAACTGTGGGGAACAT
Exons 15 and 16	CATCGACAACACAAACCCAG	ATCTCCAGGATGGCAGAGAA
Exon 17	ATGTTCCCCACAGTTTCGAG	TGTCCTTAAAGCCTCCTCCA
Other Primers		
Exon 10 and 17 RT-PCR	CGGGGCGGAAGAAGAAAGACT	GCCTCGGCTCCACCCATAG
Hu *PNKP* cDNA *In situ* probe	GACTGGAGGATCTTGTACCCAGAGA	ATCTCCAGGATGGCAGAGAAGCCTTCA
Mo *Pnkp* cDNA *In situ* probe	GGATCCGGAGGAGGGAGAGGATAC	GCCCTCGGAAAACTGGCGGTACA
Mo-Pnkp RNAi	AGCTGTATTGGAGAAGTTACTCAAGA-GATAACTTCTCCAATACAGCTTCTTTTTT	AATTAAAAAAGAAGCTGTATTGGAGAAGTTA-TCTCTTGAGTAACTTCTCCAATACAGCTGGCC

Supplementary Note.

Microcephaly with Seizure Patient Clinical Information:

The two patients examined in the first pedigree (V:1 and V:9) were 21 and 6 years of age, respectively, at the time of the last clinical examination. The pregnancies were uneventful. They were both born at full term with induced, vacuum delivery. The birth weights were 2.5 kg. The head circumferences at birth were 30 cm (-2 standard deviations (SD) below the mean and 31 cm (-2 SDs) [19]. The most recently measured head circumferences were 45 cm (-6 SDs) for the 21 year old V:1 and 42.3 cm (-6 SDs) for the 6 year old V:9. Both patients had frequent seizures, with first onset at 3 and 6 months of age, respectively. V:9 had surgery to control the seizures. By report of the mother, the EEGs were abnormal, but the reports were not available. Both patients had severe intellectual disability and delayed motor milestones. They could only communicate with a few words. The behavior of V:1 was very difficult to control because of hyperactivity. No chromosomal abnormality was detected by cytological analysis in either patient, and they did not have any abnormalities of other organ systems. The results of biochemical tests of the body fluids were unremarkable.

In the second family, the affected boy and girl had severe microcephaly, intractable seizures, speech impairment, and hyperactivity. The clinical features were similar to the first pedigree by report of the parents; however, more detailed medical reports were not available.

Two affected boys in Family 3 (VI:2 and VI:3) were last examined at the ages of 4.5 years and 21 months, respectively. The pregnancies were normal. Birth occurred at 42 and 39 gestational weeks respectively. Deliveries were induced with vertex presentations. The birth weights were 3.3 kg for both. The head circumferences at birth were 31 cm (-2 SDs) and 32 cm (-2 SDs). The most recent measurements of the head circumferences were 44 cm (-4 SDs) for

the 4.5 year old VI:2 and 38.5 cm (-7 SDs) for the 21 month old VI:3. The mother had ultrasound scans while pregnant with VI:3 and microcephaly was noticed at 37 weeks gestation, but not at 24 weeks. VI:2 and VI:3 began to have seizures at 3 and 6 months of age, respectively. Their seizures were refractory to medication. VI:2 had a left perisylvian resection at 9 months, but the seizures recurred after the surgery. Both children were developmentally delayed, only speaking a few words. Both were hyperactive and their behaviors were difficult to control. Brain MRIs of the patients showed grossly normal cerebral structures (pre-operatively) except for the small sizes. The gyral patterns were preserved and there were no heterotopia. Slight thinning of the corpus callosum and reduction of white matter volume were observed in both patients, which were consistent with the small brain sizes. The ventricles were somewhat enlarged. The cerebellar vermis was hypoplastic. Neither chromosomal abnormalities nor metabolic defects were detected in the patients.

The clinical presentation of the affected individuals in the Family 4 (II:4 and II:5) was very similar to the first three pedigrees. The birth weights were 2.5 kg for both children. The head circumference of II:4 was 38 cm at 25 months (-8 SDs). The head circumference of II:5 was 29 cm at birth (-3 SDs) and 37 cm at 11 months (-6 SDs). Seizures began at 3 and 12 months, respectively. Both patients were developmentally delayed. II:4 was not able to walk independently at 21 months. II:5 could not walk at 1 year of age when last examined. II:4 had no speech beyond “dada”, and II:5 could only babble. The brain MRIs of II:4 and II:5 showed no structural malformations except the small brain sizes. The cerebral cortical gyral pattern appeared normal, the volume of the white matter was slightly reduced, consistent with the small size of the brain, and the fiber tracts showed normal development and myelination. There were

no other abnormalities or heterotopia. The reduction in brain size appeared to be fairly uniform without preferential involvement of particular brain regions.

The patient in the fifth family was a girl. Her head circumference was 31cm at 1 month (-4 SDs), when she had her first seizure. The brain MRI showed a slightly simplified gyral pattern. The corpus callosum was fully formed but had reduced thickness. The volume of the white matter was also reduced. The frontal lobes appeared hypoplastic compared to other lobes.

The boy from the sixth family was born at 41 weeks gestation but concerns of microcephaly were noted at 29 weeks gestation. He was born via spontaneous vaginal delivery with a head circumference of 31 cm (-2 SD), length 51 cm (50 %ile) and 3385 g (25 %ile). Workup included normal CMV and toxoplasma titer, chromosome studies and serum/urine amino acids. Seizures started at 2 months of age. He had a vagal nerve stimulator placed at 3.5 years of age. At 7 years old he is described as severely disabled; though ambulatory, he is a toe walker. The young girl had good seizure control on two medications and a vagal nerve stimulator. The brain MRIs of II:1 in this family showed a slightly simplified gyral pattern, fully formed but thin corpus callosum, enlarged lateral ventricles, and hypoplasia of the cerebellar vermis.

The two patients in Family 7 were 8 years and 18 months old at the time of the last examination. The family was of mixed European heritage: Swedish, Italian, Irish and English. The older girl was born with a H.C. of 34.5 cm (~10 %ile) and 47.5 cm (-3 SDs) at 8 years and her brother was born with a H.C. of 33.7 cm (28 %ile) and now is at 41.5 cm (-4 SDs) at 18 months. The young girl has less severe developmental delay than the other patients described; she sat without support at 8 months, crawled at 9 months, but walked at 14 months (which is within normal limits). She spoke her first words at 18 months and could name colors at 5 years

old. She receives significant help with her schooling and converses at the equivalent of a ~3.5 year old level. She was slow with repetitive movements using her dominant (right) hand. Her first seizures were at 4 months old consisting of a staring seizure, the first being 20 minutes long. She has had moderate seizure control with seizures occurring 2 times per year from 1-6 years old but an increase in frequency since then. Her lymphocytes were tested for radiation sensitivity via clinical colony survival assay and were demonstrated to show radiation sensitivity from 1 Gy of 7% (normal range 50%+/-13%) [20,21]. Her brother showed worse delay as he sat at 11 months, crawled at 16 months and pulled to stand at 18 months. He had moderate hypotonia. He had more difficult to control seizures that started at 4 months old. The girl's MRI showed microcephaly with a question of but not definitive cortical dysplasia and significantly less white matter volume. Her brother had microcephaly with normal to thin cortex with moderate to severe reduction in white matter volume diffusely with thin corpus callosum. The girl's clinical evaluation included MECP2, Angelman, karyotype genetic studies and serum amino acids, urine organic acids, acylcarnitine, biotinidase, ammonia, along with lumbar puncture for CSF glucose and neurotransmitters, all of which are negative.

In summary, MCSZ patients have microcephaly, seizures, developmental delay, and variable behavioral problems particularly hyperactivity. The level of the severity of the disease appears to have a significant range, from severe microcephaly and very difficult seizure control (Families 1-6) to moderate microcephaly and mildly drug resistant seizures (Family 7). The patients likely have embryonic deficiencies in brain development and not progressive disease for multiple reasons. First, the MRIs show no evidence of atrophy or wasting that would be indicated by large spaces within the sulci between gyri. Second, there is no evidence of glial scaring on the MRI that would be demonstrated by T2 hyperintensity. Third, clinically the

patients have no evidence of significant developmental regression or loss of milestones that would indicate a progressive disease. The increases in relative microcephaly below the mean as the patients grow older likely reflect a lack of brain growth compared to normal children not brain shrinkage. This phenomenon is seen in other microcephaly syndromes. The MRI's consistently show proportional loss of cerebellum to cerebrum and subpallium (basal ganglia or ventral cerebrum) to pallium (dorsal cerebrum). This may indicate *PNKP* is required in most areas of neurogenesis during development. Although the patients have abnormalities in DNA repair, no patients have clear evidence of immunodeficiency or cancer, but clinical information is inadequate to completely rule them out.

Supplementary Table 4. The clinical features of MCSZ patients.

	Family 1		Family 3		Family 4		Family 5	Family 6		Family 7	
Patient	V:1	V:2	VI:2	VI:3	II:4	II:5	II:1	II:1	II:2	II:1	II:4
Sex	M	M	M	M	F	F	F	F	M	F	M
Microcephaly	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
HC at Birth	30 cm (-3SDs)	31 cm (-2SDs)	31 cm (-2SDs)	32 cm (-2SDs)	30.5 cm (-3SDs)	30.5 cm (-3SDs)	N/A	30.5 cm (-3SDs)	31 cm (-2SDs)	34.5 cm (10 %ile)	33.7 cm (28 %ile)
Most recent HC	45 cm @21yr (-6SDs)	42.5 cm @6yr (-6SDs)	44 cm @4.5yr (-4SDs)	38.5 cm @21mo (-7SDs)	44.2 cm @12yr (-6SDs)	44.2 cm @12yr (-6SDs)	31 cm @1mo (-4SDs)	43 cm @14yr (-7SDs)	40.5 cm @3.5yr (-4SDs)	47.5 cm @ 8yr (-3 SDs)	41.5 cm @ 18mo (-4 SDs)
Dev. Delay	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No
Sitting	1 yr	8 mo									11 mo
Walking	4 yr	2 yr	2 yr	not by 3 yr	18 mo	18 mo		18 mo		14 mo	
Speech	4-5 words	few words	10 words	few words	no	no		no		18 mo	no
Seizures	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
First onset	3 mo	6 mo	3 mo	6 mo	3 mo	3 mo	1 mo	3 mo	2.7 mo	4 mo	4 mo
Surgery		craniotomy	Perisylvian resection		Vagal stimulator	Vagal stimulator		Vagal stimulator	Vagal stimulator		
Type			Refractory	Lennox-Gastaut type	Complex partial	Complex partial		Complex partial	Complex partial		
Hyperactivity	Yes		Yes	Yes	Yes	Yes		Yes	Yes		
MRI	N/A	N/A							N		
Gyral pattern			Preserved	Preserved	Slightly simplified	Slightly simplified	Slightly simplified	Slightly simplified	Slightly simplified	Preserved	Normal to thin
CC			Thin	Thin	Thin	Thin	Thin	Thin			Thin
White matter			Reduced	Reduced	Reduced	Reduced	Reduced	Reduced		Reduced	
Ventricles			Enlarged	Enlarged	Enlarged	Enlarged		Enlarged			Enlarged
Cbl. Vermis			Reduced	Reduced	Reduced	Reduced	Reduced	Reduced			Normal

Abbreviations: HC, head circumference; CC, corpus callosum; Cbl., cerebellar; N/A, not available.

ELECTRONIC-DATABASE INFORMATION

Human Splicing Finder Version 2.3:

http://www.umd.be/HSF/

Online Mendelian Inheritance in Man (OMIM):

http://www.ncbi.nlm.nih.gov/Omim/

Supplementary Information References:

1. Broman, K.W., Murray, J.C., Sheffield, V.C., White, R.L. & Weber, J.L. Comprehensive human genetic maps: individual and sex-specific variation in recombination. *Am J Hum Genet* **63**, 861-9 (1998).
2. Bernstein, N.K. et al. The molecular architecture of the mammalian DNA repair enzyme, polynucleotide kinase. *Mol Cell* **17**, 657-70 (2005).
3. Cartegni, L., Chew, S.L. & Krainer, A.R. Listening to silence and understanding nonsense: exonic mutations that affect splicing. *Nat Rev Genet* **3**, 285-98 (2002).
4. Hubbard, T.J. et al. Ensembl 2007. *Nucleic Acids Res* **35**, D610-7 (2007).
5. Stephens, M., Smith, N.J. & Donnelly, P. A new statistical method for haplotype reconstruction from population data. *Am J Hum Genet* **68**, 978-89 (2001).
6. Stephens, M. & Donnelly, P. A comparison of bayesian methods for haplotype reconstruction from population genotype data. *Am J Hum Genet* **73**, 1162-9 (2003).
7. Stephens, M. & Scheet, P. Accounting for decay of linkage disequilibrium in haplotype inference and missing-data imputation. *Am J Hum Genet* **76**, 449-62 (2005).
8. Whitehouse, C.J. et al. XRCC1 stimulates human polynucleotide kinase activity at damaged DNA termini and accelerates DNA single-strand break repair. *Cell* **104**, 107-17 (2001).
9. Caldecott, K.W. XRCC1 and DNA strand break repair. *DNA Repair (Amst)* **2**, 955-69 (2003).
10. Wiederhold, L. AP Endonuclease-Independent DNA Base Excision Repair in Human Cells. *Molecular Cell* **15**, 209-220 (2004).
11. Rass, U., Ahel, I. & West, S.C. Actions of aprataxin in multiple DNA repair pathways. *J Biol Chem* (2007).
12. Zharkov, D.O. Base excision DNA repair. *Cell Mol Life Sci* **65**, 1544-65 (2008).
13. Caldecott, K.W. Single-strand break repair and genetic disease. *Nat Rev Genet* **9**, 619-31 (2008).
14. Konishi, Y. Cdh1-APC Controls Axonal Growth and Patterning in the Mammalian Brain. *Science* **303**, 1026-1030 (2004).
15. Götz, M., Stoykova, A. & Gruss, P. Pax6 controls radial glia differentiation in the cerebral cortex. *Neuron* **21**, 1031-44 (1998).
16. Mullen, R., Buck, C. & Smith, A. NeuN, a neuronal specific nuclear protein in vertebrates. *Development* **116**, 201 (1992).

17. Srinivasan, A. et al. In situ immunodetection of activated caspase-3 in apoptotic neurons in the developing nervous system. *Cell Death Differ* **5**, 1004-16 (1998).
18. McDonald, J.H. *Handbook of biological statistics*, 293 (Sparky House Publishing, Baltimore, 2008).
19. Roche, A.F., Mukherjee, D., Guo, S.M. & Moore, W.M. Head circumference reference data: birth to 18 years. *Pediatrics* **79**, 706-12 (1987).
20. Huo, Y.K. et al. Radiosensitivity of ataxia-telangiectasia, X-linked agammaglobulinemia, and related syndromes using a modified colony survival assay. *Cancer Res* **54**, 2544-7 (1994).
21. Sun, X. et al. Early diagnosis of ataxia-telangiectasia using radiosensitivity testing. *J Pediatr* **140**, 724-31 (2002).