SUPPLEMENTARY MATERIAL

SUPPLEMENTARY FIGURE LEGENDS

Figure S1. Elevated γH2AX in T-cells from patients with uFCRC. (**a**) Primary T-cells from pretreatment blood of 10 uFCRC cases, 10 sporadic CRC cases and their respective age- and sex-matched controls were stimulated by PHA and IL-2, and stained for nuclear γH2AX foci. Relative levels of γH2AX foci/cell are depicted. Cases: filled circles, sporadic CRC: filled triangle, uFCRC. Controls: open circles, sporadic CRC matched controls, open triangles uFCRC matched controls. Differences are not significant between matched cases and controls. (**b**) Comparison of γH2AX foci/cell when cells in (a) are treated with vehicle, aphidicolin or UV, represented as levels observed uFCRC or sporadic CRC patients normalized to their control groups. For uFCRC patients, a statistically significant increase in accumulation of γH2AX relative to controls was observed with aphidicolin ($p=0.03$) or UV ($p=0.04$); for sporadic CRC patients, only aphidicolin produced a significant increase relevant to controls ($p=0.04$); for UV, difference between uFCRC and sCRC was highly significant ($P = 0.0002$). P values were calculated using Wilcoxon rank-sum tests.

Figure S2. Molecular model of SHPRH variant in patients 118294 and 130924. The SHPRH structure (left, green) was modeled after the homologous 1Z63 Sulfolobus solfataricus SWI2/SNF2 ATPase core. The human missense R1184C substitution (dark blue; corresponding to residue R1187 in Sulfolobus) was mapped on this model adjacent to the DNA ligand found in this template structure (right, light brown) and the nearby C942 (yellow), predicted to compromise DNA binding directly and potentially via an aberrant disulfide bond to C942.

Figure S3. Partial knockdown of variant proteins to simulate haploinsufficiency in HCT116 CRC cells induces γH2AX. Two different siRNAs, (used at 5nM concentration) per gene noted or GL2 negative control siRNA were used for transfection of HCT 116 cells (**a, b**) Total nuclear γH2AX was scored by automated, computer-assisted microscopy in cells that were untreated (**a**) or treated with the DNA damaging agent irinotecan (**b**). Results are depicted as the mean of percent positive cells normalized to GL2 control. *, P values are $p<0.05$; ns, not significant, calculated using a Wilcoxon rank-sum test. (**c**) Western analysis was used to confirm degree of knockdown ranged from ~35-60%, normalized to GL2 control.

Figure S4. Knockdown of WRN and ERCC6 in HCT 116 CRC cells induces γH2AX. Using two different siRNAs per gene indicated or GL2 negative control at a 5nM concentration, HCT116 cells were transfected and degree of protein knockdown confirmed by Western analysis. Total nuclear γH2AX was scored by automated, computer-assisted microscopy. Results are depicted as the mean of percent positive cells (y-axis). Knockdown of WRN and ERCC6 markedly increased γH2AX ($p<0.05$ for each compared to GL2) calculated using a Wilcoxon rank-sum test.

Figure S5. Consurf program analysis of the ERCC6 amino-terminal domain region predicts that the Pt1 variant N190 is functionally significant. Sequence conservation is depicted by red shading (see key below). Residues predicted to be exposed are marked below by an ‘e’. Residues predicted by these criteria to be functionally important are marked below by an ‘f’. The variant residue (arrow) is in the center of a run of residues predicted to be functionally important.

Figure S6. **Evidence for genomic instability in Pt1 primary T-cells.** (**a**) A single Pt1 primary T-cell metaphase spread is shown (left) that exhibited a gain of chromosome 9 (right, arrow). (**b**) T-cells from Pt1 and her matched control were left untreated (No Rx) or treated with aphidicolin (Aph), camptothecin (Campto), etoposide (Etop), or UV. Pt1 cells exhibited consistently higher fractions of cells with > 10 nuclear γH2AX foci (each $P < 0.001$, except UV: $P = 0.006$). Cell numbers allowed an independent experiment for the no treatment and aphidicolin conditions, with good agreement (means +/- ranges). Inserted image: γH2AX foci in Pt1 cells treated with aphidicolin. A total of 2,082 cells were counted, a minimum of 124 and mean of 149/ condition.

Figure S7. Increased γH2AX in Pt1 primary T-cells by confocal microscopy. Pt1 cells and those from an age- and sex-matched control were processed as per Fig 1 without or with treatment with UV. γH2AX foci in the nucleus were identified by co-staining with DAPI and scored by confocal microscopy. Depicted is the percentage of cells with >10 nuclear foci. A total of 183 cells were scored.

Figure S8. Lower WRN and ERCC6 levels in Pt1-derived lymphocyte cell lines. Two independent EBV-transformed B-cell lines (A, B) derived from each of Pt1 and a matched control were subjected to IB for WRN, ERCC6, and a loading controls (α-tubulin). The ERCC6 experiment is shown in duplicate. Note the lower levels of WRN and ERCC6 in the patient-derived lines. Some degradation occurred in one Pt1A ERCC6 sample (left).

Figure S9. Higher γH2AX, p53, and phospho(P)-ATM levels in Pt1 cell lines A and B than in matched control lines. The lines were cultured without DNA damaging treatment (**a** and **b**)

or following treatment with aphidicolin **(a)**. (**a**) γH2AX levels detected by immunoblotting. H2AX: loading control. (**b**) P-ATM and p53 levels. α-tubulin: loading control. Similar results were seen in two independent experiments.

SUPPLEMENTARY TABLES

Table S1: Detailed Summary of HQVs in uFCRC Patients

Table S2: HQVs in uFCRC patients, based on predictions of 4 programs

Table S3: Detailed Summary of HQVs in Polyposis Patients

Table S4: Demographics of uFCRC Patients

Table S5: Representation of Variants in EVS

Table S6: Representation of Variants in ExAC

Table S7: Representation of variants in the ITMI cohort

SUPPLEMENTARY METHODS.

Clinical data. Clinical information was obtained from medical records of the Fox Chase Cancer Center Risk Assessment Program. Documentation of colon neoplasia was typically from records of colonoscopy and/or pathology of resected colon, as noted. Extensive polyposis was ruled in or out by examination of records of colonoscopy and/or pathological analysis of the resected colon. If polyposis was observed, patients received genetic testing for FAP and MutYH polyposis. CRCs were tested for microsatellite instability and/or immunohistochemistry for mismatch repair proteins. In addition, exome sequencing was performed, which excluded known pathogenic mutations in FAP, MutYH, and mismatch repair proteins. Family histories were obtained by trained genetics counselors and verified by attending physicians. Blood samples were banked under broad consent for research and de-identified. Patient identifying information was maintained separately and provided to the principal investigator in de-identified fashion. All work was approved by the Institutional Review Board.

Lymphocyte cell preservation, culture, drug treatments, and metaphase spreads. Peripheral blood was collected by venipuncture in 10 ml acid-citrate-dextrose vacutainor tubes at room temperature and processed the same day. The tubes were centrifuged at 800g for 10 min. The buffy coat was extracted, diluted with an equal volume of phosphate buffered saline (PBS), layered over 7 ml of Ficoll-Paque and centrifuged at 450g for 15 min. The cloudy band below the red platelet/plasma layer and above the Ficoll/granulocyte layer was extracted extracted, diluted in an equal volume of PBS, and centrifuged at 300g for 12 min. The cell pellet was resuspended in 5mL of PBS and centrifuged again at 300g for 12 min. The cell pellet as re-suspended in 90% fetal bovine serum and 10% dimethylsulfoxide. The cells derived from 10 ml

of blood were aliquotted into one 2 ml tube, placed into a controlled-rate freezing container, stored at -80˚C for a minimum of 4 h, and transferred to the vapor phase of a liquid nitrogen tank.

The samples were cultured at $1x10^6$ cells per ml in RPMI-1640 containing 15% fetal bovine serum (HyClone), 2 mM L-glutamine (Life Technologies), 50 mM 2-mercaptoethanol (Sigma), 0.2 units human recombinant insulin (Sigma) per ml, 50 units penicillin and 50 mg streptomycin per ml (complete RPMI) with phytohemaglutinin (PHA), M form, (Life Technologies) at a final concentration of 1.5%. All incubations were done in a humidified, 37°C incubator with 5% CO2. Proliferation of resting T cells, previously stimulated with PHA, was induced by culturing them in 12-well plates with the following immobilized monoclonal antibodies (mAb): anti-human CD28 (clone CD28.2, BioLegend) and anti-human CD3 (clone OKT3, prepared in this laboratory). To prepare the plate-bound mAb, 0.5 ml of PBS containing both mAb at 1 mg/ml was added per well and the plate incubated overnight at 4°C. After washing each well twice with PBS, 1×10^6 T cells/well were added in 2 ml complete RPMI containing 50 units/ml recombinant human interleukin 2 (IL-2) (NCI Preclinical Repository). After a 48 h incubation, the cells were adjusted to 5×10^5/ml by adding additional complete RPMI with IL-2 and returned to the incubator.

To generate metaphase spreads, 66h after stimulation KaryoMax colcemid (Life Technologies) was added (0.03 mg/ml final concentration) and cultures were incubated an additional 2 h. Cultured cells arrested in metaphase were swollen in a hypotonic solution, then fixed according to standard protocols. Briefly, cells were centrifuged and resuspended in a hypotonic solution of 0.075 M KCl. After 15 min at 37°C, a few drops of cold, freshly prepared Carnoy's Fixative (3 parts methanol: 1 part glacial acetic acid) was added, the cells centrifuged,

then treated for 30 min at 4°C in 10 ml fixative. Cells were washed three more times in cold fixative, then resuspended in fixative to give a cloudy suspension, and dropped onto slides, according to classical methods. 50 well-separated spreads were scored per sample.

Variant frequency analysis in controls. Variants in the DNA DSB repair genes were extracted using genome coordinates retrieved from the UCSC hg19 human genome browser (http://genome.ucsc.edu/) then filtered for call quality and predicted impact on the encoded protein sequence as previously described for the ITMI cohort [1]. Allele frequencies were computed with VCFtools v0.1.12b [2] and annotations with ANNOVAR version 2014-07-22 [3] using refSeq transcripts [4].

Mutation frequency analysis. The frequency of somatic mutations in genes of interest was extracted from TCGA studies (http://cancergenome.nih.gov/) using cBioPortal [5]. Information about most common transcripts for each gene was extracted from the Ensemble database using Biomart tool (http://www.ensembl.org/biomart/martview/), and the somatic mutation frequency was normalized to the length of longest mRNA for each gene. Heatmaps visualizing the frequencies of somatic mutations in the selected TCGA sets were produced using the software MeV (MultiExperiment Viewer), version 4.8 (http://www.tm4.org/mev.html**).**

Generation and transfection of B-cell lines. B-cells from patients and normal controls were immortalized with Epstein-Barr virus (EBV) according to Tosato and Cohen [6]. Briefly, lymphocytes isolated from 10 ml blood by centrifugation over Ficoll-Paque were resuspended in 2.5 ml complete RPMI-1640 without insulin. Immortalization was initiated by the addition of

EBV strain B95-8 as 1 ml of filtered supernatant from the marmoset B-cell line GM 7404. After incubation for 2 h at 37°C, 6.5 ml of complete RPMI containing 1 mg/ml cyclosporine A (Sigma) was added and the suspension cultured at 37°C, 5% CO2 in an up-right flask. After 3 weeks, the culture was split and 5 ml fresh medium added to each flask. After an additional week of incubation, cells in one of the flask pairs were cryopreserved and cells in the second flask were subcultured for eventual harvest. For WRN and ERCC6 'rescue' experiments, cell lines were transfected with empty or control vector (hWRN-pDsFLAG, and pGK-lenti-Neo-ERCC6, respectively) using lipofectamine and selected in neomycin.

Exome sequencing and variant calling. Exome libraries enriched for coding exon sequences were captured using the SureSelect Version V Target Enrichment System (Agilent Technologies, Wilmington, DE). The libraries were amplified by PCR using the supplied paired-end PCR primers. Sequence reads were mapped to the human reference genome (hg19) using the SamTools package and BWA. Duplicated reads were removed with the Picard software. Recalibration of base quality and indel realignment was performed with the GATK package from Broad Institute (Cambridge, MA). Single nucleotide variants and indel variants were identified using the Unified Genotyper caller of the GATK package, using external and high quality internal exome references. Rare variants (frequency less than 1/100 in public exome data bases (Exome Variant server, www. http://evs.gs.washington.edu/EVS/)) were annotated with SeattleSeq Annotation (http://gvs.gs.Washington.edu/SeattleSeqAnnotation/). A SQL database was created from the annotated dataset. Manual examination was conducted with TViewer of SamTools and the Integrated Genomic Viewer, to identify high confidence variants from the raw sequence data. Missense variants that had at least 15 sequencing reads covering the site, of which

at least 6 were variant, were validated by direct Sanger sequencing. All 23 validated, as noted in Supplementary Table 1. Thereafter, missense variants were considered validated if the encompassing sequence was present in at least 30 reads, at least 10 of which were variant, and surrounding sequences contained few variants. To minimize type 1 error, a few variants were excluded based on reported highly restricted tissue or subcellular expression (e.g. not detected in colon; GeneCards website (Weizman Institute of Science, http://www.genecards.org/)). These variants were included in the statistical analysis of potential enrichment of HQVs within pathways, to avoid bias.

Analysis of somatic mutation frequencies. Mutation frequencies were retrieved using cBioPortal (http://www.cbioportal.org/public-portal/index.do. Data were extracted from the datasets with >100 sequenced samples. The most recent (provisional) TCGA datasets were used, when available. Data were visualized using MeV: MultiExperiment Viewer (http://www.tm4.org/mev.html). Transcripts lengths were obtained from the Ensemble Genome Browser (www.ensembl.org/), and the frequencies of mutations in CRCs were normalized to the length of the longest transcript (per 1000 nucleotides).

Molecular modeling. A homology model of SHPRH was built using the Biological Assembly Modeller (BAM) software [6] based on the template Sulfolobus solfataricus SWI2/SNF2 ATPase core in complex with dsDNA (PDB code 1Z63, 14% identity homolog). Side-chain rotamers were optimized with a backbone dependent library and the SCWRL 4 software [7]. Modeling using a Zebrafish protein (Uniprot accession Q7ZV09) template (PDB code 1Z3I), placed the R1187 in the exact same position.

WRN helicase assays. The WRN vectors were introduced into the bacterial strain ArcticExpress (ARX) by transformation and the expression of the fusion proteins was induced with 1mM IPTG at 12°C for 24 h. The fusion proteins were purified with glutathione affinity beads (Sigma, MO) following standard procedures [8]. The purified fusion proteins were dialyzed against ELB (10mM HEPES (pH 7.5)/50mM KCl/2.5mM MgCl2/250mM sucrose/1mM DTT) at 4°C and then stored at –80°C in 5µl aliquots.

ERCC6 chromatin remodeling assay. For nucleosome assembly and remodeling assays**,** DNA fragments of ~240 bp containing the 601 nucleosome phasing sequence was assembled into mononucleosome with purified HeLa histones using step-gradient salt dialysis [9]. DNA fragments used for assembly were generated by PCR and radio-labeled with [^{32}P] α-dATP. Remodeling assays were carried out in 12 mM Hepes (pH7.9), 10 mM Tris·HCl (pH 7.5), 60 mM KCl, 8% glycerol, 4 mM $MgCl_2$, 2 mM ATP·Mg and 0.02% NP40 at 30°C for 15 min, as described previously [9]. Nucleosomes were used at 1 nM, ERCC6 and the variant were used at 25 nM.

Comet assays. Briefly, cells were washed with PBS and embedded in 1% low melting agarose on slides, lysed and then electrophoresed under neutral conditions following the protocol from Trevigen (http://www.trevigen.com/cat/1/3/0/CometAssay). Slides were placed in a DNA precipitation solution, dried, and then stained with DAPI and imaged on a Nikon Eclipse E800 epifluorescence microscope. At least 50 cells were analyzed per condition using the CometScore software (TriTek Corp., Sumerduck, VA). The 'Olive tail moment' is computed as the summation of each tail intensity integral value, multiplied by its relative distance from the center

of the head (the point at which the head integral was mirrored), and divided by the total comet intensity.

siRNA studies in HCT 116 cells. Total nuclear γH2AX in HCT 116 CRC cells was scored in automated fashion, by computer-assisted microscopy. Images in each wavelength (TRITC, DAPI) were acquired per well using the ImageXpress micro automated microscope (Molecular Devices) driven by MetaXpress software. Images were analyzed in the Multiwavelength Scoring module of MetaXpress and results were displayed and exported utilizing the AcuityXpress software package (Molecular Devices). HCT 116 CRC cells were transfected with GL2 or two independent siRNAs (5nM) for each protein. Western analysis was performed for the respective proteins and α-tubulin loading control. Total protein levels were quantified and normalized to the loading control. The data are plotted as relative induction of γ-H2AX to GL2 control from 2 independent experiments.

Statistics. We summarized continuous and count data using means and medians, and tested for differences between cell cultures (case versus control) using Wilcoxon rank-sum tests. For binary and categorical outcomes, we created frequency tables and used Fisher's exact test to assess the relationship with case/control status. For γH2AX staining in the validation set patients, we used ROC curves to assess discrimination between cases and controls. We found optimal cutpoints using Youden's index [10], and used the best thresholds to determine whether samples were γH2AX positive or negative. To find a threshold using combined UV and aphidicolin, we created an ROC curve using regression-based predictors. We then used Fisher's exact test to assess whether the dichotomized outcomes were significantly associated with

case/control status. To assess the relationship between olive tail moments and case status, we used Poisson models for ‘No treatment’ and Aph conditions, and a negative binomial model for UV due to evidence of overdispersion. We accounted for within-patient correlation using Generalized Estimating Equations. Variant frequencies in patients and controls were compared using a Poisson ratio test. The fraction of variants in a specific pathway that were high quality was compared to the fraction of HQVs in all GO term-flagged variants by Fisher’s exact test. A Wilcoxon rank-sum test was used in siRNA experiments to compare γH2AX levels in gene-targeting conditions to the GL2 control. For comparing protein expression post knockdown, a generalized linear model was used assuming gamma family and log link with robust standard errors.

SUPPLEMENTARY REFERENCES

1. Bodian DL, McCutcheon JN, Kothiyal P, et al. Germline variation in cancer-susceptibility genes in a healthy, ancestrally diverse cohort: implications for individual genome sequencing. PLoS One 2014;9:e94554.
2. Danecek P, Auton A, Abecasis G, et al. The variant call format and VCFtools. Bioinformatics 2011;27:2156-8.
3. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res 2010;38:e164.
4. Pruitt KD, Brown GR, Hiatt SM, et al. RefSeq: an update on mammalian reference sequences. Nucleic Acids Res 2014;42:D756-63.
5. Cerami E, Gao J, Dogrusoz U, et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. Cancer Discov 2012;2:401-4.
6. Tosato, G. and J.L. Cohen, Current Protocols in Immunol 7.22.1 – 7.22-4, 2007.
7. Shapovalov MV, Wang Q, Xu Q, et al. BioAssemblyModeler (BAM): User-Friendly Homology Modeling of Protein Homo- and Heterooligomers. PLoS One 2014;9:e98309.
8. Ausubel FM, Brent R, Kington RE, et al. Current Protocols in Molecular Biology. Greene Publishing Associates & Wiley-Interscience. 1988.
9. Cho I, Tsai PF, Lake RJ, et al. ATP-dependent chromatin remodeling by Cockayne syndrome protein B and NAP1-like histone chaperones is required for efficient transcription-coupled DNA repair. PLoS Genet 2013;9:e1003407.
10. Youden WJ. Index for rating diagnostic tests. Cancer 1950;3:32-5.

Supplementary Table 1. Detailed Summary of High Quality Variants in uFCRC Patients

Patient ID No.	HQ Variant	Chr Band#	Position	DNA Sequence	AA Sequence	Freq	Rs No.	SIFT	PP2
111797	CHAF1B	21q22.12-3	37758481	c.47T>C	p.V16A	0.004	rs143543321	D	1
118294	SHPRH	6q24.3	146243968	c.3550C>T	p.R1184C	0.007	rs148401398	D	0.99
120713	WRN	8p12	30969156	c.2114C>T	p.T705I	0.0004	rs141563618	D	1
(Pt1)	ERCC6	10q11.23	50738771	c.538A>T	p.N180Y	0	NA	D	1
122517	POLD1	**19q13.33**	50909713	c.1433G>A	p.S478N	0 (c.f. Palles)	NA	D	1
	MCM10	10p13	13240791	c.2225C>A	p.T742K	0.003	rs41291311	D	0.99
	MPG	16p13.3	135643	c.765_771del	p.A255NA	0	NA	NA	NA
	SKA1	**18q21.1**	47917649	c.467C>T	p.T156M	0	NA	D	1
123000	SPAG5	17q11.2	26918796	c.1357C>T	p.R453W	0	NA	D	1
	BRD4	19p13.12	15376259	c.755C>G	p.P252R	0.002	rs200633797	D	0.97
123006	MASTL	10p12.1	27454325	c.668C>A	p.P223Q	0	NA	D	1
123163	DYNC1H1	14q32.1	102507911	c.11942C>G	p.T3981R	0.002	rs138428684	T	0.98
	CHAF1A	19p13.3	4431979	c.1978C>T	p.R660C	0	NA	D	1
123320	MSH2	2p21	47630468	c.138C>G	p.H46Q	0.002	rs33946261	D	0.99
124744	FANCL	2p16.1	58392874	c.676C>T	p.R226C	0	NA	D	1
125380	None								
125659	EEPD1	7p14.2	36327344	c.1273C>T	p.R425W	0.0002	rs202090314	D	1
126780	ERCC2	19q13.3	45856060	c.1846C>T	p.R616W	0.00008	rs121913024	D	1
126875	ANAPC10	**4q31.21**	145916666	c.417T>G	p.I139M	0.0005	rs199788158	D	0.99
127581	CTC1	17p13.1	8141897	c.248G>C	p.S83T	0.006	rs78870822	D	1
129338	None								
132231	C19Orf40	19q13.11	33467427	c.487G>A	p.V163M	0.0004	rs141801484	D	1
132406	DDX11	12p11.21	31244665	c.1102C>T	p.P368S	0	NA	D	1
	LIG1	19q13:33	48619195	c.2611C>T	p.R871C	0	NA	D	1
132667	LIG3	17q11.2-12	33318982	c.1214C>T	p.T405I	0	NA	D	1
133012	POLD3	**11q13.9**	74345739	c.1118A>C	p.K373T	0.001	rs181857010	T	0.96
134321	none								

Chr band: chromosome band, bold: loci linked by GWAS to CRC risk; Position: chromosomal position; DNA seq: position in the cDNA Freq: frequency of HQVs in on-line exome databases (Exome Variant Server); Rs No: reference single nucleotide polymorphism (SNP) number; SIFT: scale-invariant feature transform algorithm score; PP2: PolyPhen2 algorithm score.

Supplementary Table 2
HQVs in uFCRC patients based on predictions of 4 programs

patient ID	Gene	INPUT (chromosome; position; ref. nt; detected nt)	Amino acid change	Polyphen-2 prediction	MutationAssessor prediction	SIFT prediction	PROVEAN prediction
111797	CHAF1B	21,37758481,T,C	V16A	probably damaging	medium	Damaging	Deleterious
118294	SHPRH	6,146243968,G,A	R1184C	probably damaging	medium	Damaging	Deleterious
120713	WRN	8,30969156,C,T	T705I	probably damaging	high	Damaging	Deleterious
	ERCC6	10,50738771,T,A	N180Y	probably damaging	medium	Damaging	Deleterious
122517	POLD1	19,50909713,G,A	S478N	probably damaging	high	Damaging	Deleterious
	MCM10	10,13240791,C,A	T742K	probably damaging	low	Damaging	Deleterious
	SKA1	18,47917649,C,T	T202M	probably damaging	low	Damaging	Deleterious
123000	SPAG5	17,26918796,G,A	R453W	probably damaging	low	Damaging	Deleterious
	BRD4	19,15376259,G,C	P252R	probably damaging	low	Damaging	Deleterious
123006	MASTL	10,27454325,C,A	P223Q	probably damaging	medium	Damaging	Deleterious
123163	DYNC1H1	14,102507911,C,G	T3981R	probably damaging	medium	Tolerated	Deleterious
	CHAF1A	19,4431979,C,T	R660C	probably damaging	low	Damaging	Deleterious
123320	MSH2	2,47630468,C,G	H46Q	probably damaging	medium	Damaging	Deleterious
124744	FANCL	2,58392874,G,A	R226C	probably damaging	medium	Damaging	Deleterious
125659	EEPD1	7,36327344,C,T	R425W	probably damaging	low	Damaging	Deleterious
126780	ERCC2	19,45856060,G,A	R616W	probably damaging	high	Damaging	Deleterious
126875	ANAPC10	4,145916666,A,C	I139M	probably damaging	medium	Damaging	Neutral
127581	CTC1	17,8141897,C,G	S83T	probably damaging	medium	Damaging	Neutral
132231	C19orf40	19,33467427,G,A	V163M	probably damaging	medium	Damaging	Neutral
132406	DDX11	12,31244665,C,T	P387S	probably damaging	medium	Damaging	Deleterious
	LIG1	19,48619195,G,A	R871C	probably damaging	high	Damaging	Deleterious
133012	POLD3	11,74345739,A,C	K373T	probably damaging	low	Damaging	Deleterious
130170	DCLRE1A	10,115612595,G,A	P116L	probably damaging	medium	Tolerated	Deleterious
	ERCC4	16,14028081,C,T	P379S	probably damaging	high	Damaging	Deleterious
	SPC24	19,11257962,T,C	H164R	probably damaging	medium	Damaging	Deleterious
	BMP7	20,55758829,G,A	R303C	probably damaging	low	Damaging	Deleterious
130924	SHPRH	6,146243968,G,A	R1184C	probably damaging	medium	Damaging	Deleterious
	NCAPD3	11,134090560,A,G	I42T	probably damaging	medium	Damaging	Neutral
	TEX14	17,56729334,G,A	P10L	probably damaging	neutral	Damaging	Deleterious
132667	LIG3	17,33318982,C,T	T405I	probably damaging	medium	Damaging	Deleterious
133486	RFC2	7,73654270,C,G	Q231E	probably damaging	N/A	Damaging	*Deleterious*
	MIS18BP1	14,45686268,T,A	H986L	probably damaging	medium	Damaging	Deleterious

Supplementary Table 3. Detailed Summary of High Quality Variants in Polyposis Patients

Patient ID No.	HQ Variant	Chr band	Position	DNA Sequence	AA Sequence	Freq	Rs No.	SIFT	PP2
120200 >20 adenos @50	ERCC3	2q14-21	128051205	c.117_118insC	p. S40Lfs*9	0	none	NA	early truncation
126784 10 adenos in 20s+	None								
130170 >20 adenos @47	DCLRE1A	**10q25.1-3**	115612595	c.347C>T	p.P116L	0	none	D	0.96
	ERCC4	16p13.12-.3	14028081	c.1135C>T	p.P379S	0.004	rs1799802	D	1
	SPC24	19p13.2	11257962	c.491A>G	p.H164R	0	none	D	1
	BMP7	20q13.31	55758829	c.907C>T	p.R303C	0	none	D	1
130924 50-100 adenos @50	SHPRH	6q24.3	146243968	c.3550C>T	p.R1184C	0.007	rs148401398	D	0.99
	NCAPD3	11q25	134090560	c.125T>C	p.I42T	0.002	rs112183153	D	0.95
	ESPL1	12q13.13	53682043	c.4464C>G	p.I1488M	0.005	rs61737629	D	1
	TEX14	17q22	56729334	c.29C>T	p.P10L	0	none	D	1
133486 11 adenos @33+	RFC2	7q11.23	73654270	c.691C>G	p.Q231E	0	none	D	1
	MIS18BP1	14q21.2	45686268	c.2957A>T	p.H986L	0.001	rs139950206	D	0.99

Supplementary Table 4. Demographics of uFCRC Patients

Patient ID #	Age at Diagnosis, Sex	CRC Location	Chemotherapy	Smoking	Smoke Exposure
111797	45, unkwn	unknown	unknown	unknown	unknown
118394	49, F	sigmoid	N	N	N
120713 (Pt1)	48, F	left	N	Y	distant
122517	<50, F	left	N	N	N
123000	>50, F	right	N	Y	<1ppd
123006	<50, M	rectum	Y	Y	distant, 2hd
123163	<50, F	rectum	Y	unknown	unknown
123320	<50, F	unknown	Y	N	2hnd
124744	<50, F	unknown	Y	N	2hnd
125380	<50, M	rectum	Y	Y	occ cigar
125659	<50, M	rectum	Y	N	2hnd
129338	>50, F	right	Y	Y	10pk-ys
132231	40, F	rectum	N	Y	20pk-ys
132406	>50, M	transverse	Y	Y	30pk-ys
132667	<50, M	transverse	N	Y	30pk-ys
133012	<50, M	sigmoid	Y	Y	40pk-ys
134321	<50, M	recto-sigmoid	Y	N	2hnd
133486	>50, F	unknown	N	unknown	unknown

Smoke exposure legends: ppd- pack per day; occ- occasional; pk-ys- pack years, 2hnd- second hand smoke.

Supplementary table 5

Representation of the variants in the Exome Variant Server (EVS)

Variants from Supplementary Table 1

		EA Allele #	AA Allele #	All Allele #	EA Genotype #	AA Genotype #	All Genotype #
CHAF1B	21: 37758481	C=46/T=8554	C=8/T=4398	C=54/T=12952	CC=0/CT=46/TT=4254	CC=0/CT=8/TT=2195	CC=0/CT=54/TT=6449
SHPRH	6:146243968	A=51/G=8175	A=4/G=3794	A=55/G=11969	AA=0/AG=51/GG=4062	AA=0/AG=4/GG=1895	AA=0/AG=55/GG=5957
WRN	8:30969156	T=5/C=8595	T=0/C=4406	T=5/C=13001	TT=0/TC=5/CC=4295	TT=0/TC=0/CC=2203	TT=0/TC=5/CC=6498
ERCC6	10:50738771						
POLD1	19:50909713						
MCM10	10:13240791	A=42/C=8558	A=6/C=4400	A=48/C=12958	AA=0/AC=42/CC=4258	AA=0/AC=6/CC=2197	AA=0/AC=48/CC=6455
MPG	16:135643						
SKA1	18:47917649						
SPAG5	17:26918796						
BRD4	19:15376259						
MASTL	10:27454325						
DYNCH1	14:102507911	G=38/C=8562	G=4/C=4402	G=42/C=12964	GG=0/GC=38/CC=4262	GG=0/GC=4/CC=2199	GG=0/GC=42/CC=6461
CHAF1A	19:4431979						
MSH2	2:47630468	G=2/C=8584	G=1/C=4379	G=3/C=12963	GG=0/GC=2/CC=4291	GG=0/GC=1/CC=2189	GG=0/GC=3/CC=6480
FANCL	2:58392874						
EEPD1	7:36327344	T=1/C=8599	T=1/C=4405	T=2/C=13004	TT=0/TC=1/CC=4299	TT=0/TC=1/CC=2202	TT=0/TC=2/CC=6501
ERCC2	19:45856060	A=1/G=8599	A=0/G=4406	A=1/G=13005	AA=0/AG=1/GG=4299	AA=0/AG=0/GG=2203	AA=0/AG=1/GG=6502
ANAPC10	4:145916666	C=6/A=8196	C=0/A=3738	C=6/A=11934	CC=0/CA=6/AA=4095	CC=0/CA=0/AA=1869	CC=0/CA=6/AA=5964
CTC1	17:8141897	G=76/C=8336	G=4/C=4114	G=80/C=12450	GG=1/GC=74/CC=4131	GG=0/GC=4/CC=2055	GG=1/GC=78/CC=6186
C19orf40	19:33467427	A=2/G=8598	A=3/G=4403	A=5/G=13001	AA=0/AG=2/GG=4298	AA=0/AG=3/GG=2200	AA=0/AG=5/GG=6498
DDX11	12:31244665						
LIG1	19:48619195						
LIG3	17:33318982						
POLD3	11:74345739	C=2/A=8584	C=2/A=4398	C=4/A=12982	CC=0/CA=2/AA=4291	CC=0/CA=2/AA=2198	CC=0/CA=4/AA=6489

Variants from Supplementary Table 3

ERCC3	2:128051205						
DCLRE1A	10:115612595						
ERCC4	16:14028081	T=50/C=8550	T=7/C=4387	T=57/C=12937	TT=0/TC=50/CC=4250	TT=0/TC=7/CC=2190	TT=0/TC=57/CC=6440
SPC24	19:11257962						
BMP7	20:55758829						
SHPRH	6:146243968	A=51/G=8175	A=4/G=3794	A=55/G=11969	AA=0/AG=51/GG=4062	AA=0/AG=4/GG=1895	AA=0/AG=55/GG=5957
NCAPD3	11:134090560	G=33/A=8561	G=4/A=4398	G=37/A=12959	GG=0/GA=33/AA=4264	GG=0/GA=4/AA=2197	GG=0/GA=37/AA=6461
ESPL1	12:53682043	G=105/C=8495	G=15/C=4391	G=120/C=12886	GG=1/GC=103/CC=4196	GG=0/GC=15/CC=2188	GG=1/GC=118/CC=6384
TEX14	17:56729334						
RFC2	7:73654270						
MIS18BP1	14:45686268	A=12/T=8588	A=2/T=4404	A=14/T=12992	AA=0/AT=12/TT=4288	AA=0/AT=2/TT=2201	AA=0/AT=14/TT=6489

Supplementary table 6

Representation of the variants in Exome Aggregation Consortium (ExAC)

Variants from Supplementary Table 1

		in total population					in European non Finnish			
		allele count	allele number	number of homozygous	allele frequency		allele count	allele number	number of homozygous	allele frequency
CHAF1B	21: 37758481	335	121332	2	0.002761		261	66668	2	0.003915
SHPRH	6:146243968	356	118900	2	0.002994		266	65964	0	0.004033
WRN	8:30969156	66	121302	0	0.0005441		63	66676	0	0.0009449
ERCC6	10:50738771	1	121396	0	0.000008238		0	66732	0	0
POLD1	19:50909713									
MCM10	10:13240791	482	120748	2	0.003992		98	66438	2	0.005991
MPG	16:135643									
SKA1	18:47917649									
SPAG5	17:26918796	2	121300	0	0.00001649		0	66682	0	0
BRD4	19:15376259	83	89582	1	0.0009265		63	51154	0	0.001232
MASTL	10:27454325									
DYNCH1	14:102507911	414	121362	2	0.003411		378	66710	2	0.005666
CHAF1A	19:4431979	2	120598	0	0.00001658		0	66432	0	0
MSH2	2:47630468	23	74776	0	0.0003076		23	43494	0	0.0005288
FANCL	2:58392874	2	121404	0	0.00001647		0	66738	0	0
EEPD1	7:36327344	41	121376	0	0.0003378		36	66710	0	0.0005396
ERCC2	19:45856060	6	120726	0	0.0000497		4	66302	0	6.03E-05
ANAPC10	4:145916666	88	120700	0	0.0007291		78	66726	0	0.001169
CTC1	17:8141897	1090	119254	12	0.00914		804	66256	6	0.01213
C19orf40	19:33467427	76	121412	0	0.000626		39	66740	0	0.0005844
DDX11	12:31244665	18	81074	0	0.000222		11	45444	0	2.421E-05
LIG1	19:48619195									
LIG3	17:33318982									
POLD3	11:74345739	49	121286	0	0.000404		27	66678	0	0.0004049

Variants from Supplementary Table 3

ERCC3	2:128051205									
DCLRE1A	10:115612595	8	121412	0	0.00006589		8	66740	0	0.0001199
ERCC4	16:14028081	452	119924	1	0.003769		347	66002	1	0.005257
SPC24	19:11257962									
BMP7	20:55758829	2	121306	0	0.00001649		2	66670	0	3.00E-05
SHPRH	6:146243968	356	118900	2	0.002994		266	65964	0	0.004033
NCAPD3	11:134090560	364	121410	0	0.002998		255	66738	0	0.003821

ESPL1	12:53682043	719	84786	10	0.00848		585	48156	5	0.01215
TEX14	17:56729334									
RFC2	7:73654270	1	54622	0	0.00001831		1	30422	0	3.29E-05
MIS18BP1	14:45686268	114	121378	1	0.0009392		98	66718	1	0.001469

Supplementary table 7

Filteration of HQVs in 1508 normal control population (ITMI cohort) by prediction programs

Gene	INPUT (chromosome; position; ref. nt; detected nt)	Amino acid change	MutationAssessor prediction	SIFT prediction	PROVEAN prediction	Count of non-neutral impact
ANAPC10	4,145916666,A,C	I139M	medium	Damaging	-2.32	2
BRD4	19,15367881,G,A	P482L	medium	Damaging	-4.13	3
BRD4	19,15374335,G,A	R413C	low	Damaging	-5.66	2
BRD4	19,15376259,G,C	P252R	low	Damaging	-2.96	2
BRD4	19,15383849,A,G	L21P	low	Damaging	-3.55	2
C19orf40	19,33465062,A,G	M114V	medium	Damaging	-0.93	2
C19orf40	19,33467427,G,A	V163M	medium	Damaging	-1.35	2
C19orf40	19,33467448,G,C	G170R	high	Damaging	-7.1	3
CCNH	5,86703934,G,T	N128K	low	Damaging	-4.81	2
CCNH	5,86703939,C,T	G127R	low	Damaging	-2.64	2
CCNH	5,86703977,T,G	K114T	high	Damaging	-5.92	3
CDK7	5,68550456,T,G	I63R	medium	Damaging	-6.42	3
CDK7	5,68550486,A,G	N73S	medium	Damaging	-4.19	3
CDK7	5,68553977,T,G	H135Q	high	Damaging	-7.58	3
CDK7	5,68558096,G,A	A198T	medium	Damaging	-3.57	3
CDK7	5,68565119,C,T	P238L	low	Damaging	-8.41	2
CHAF1B	21,37759978,A,G	N70S	neutral	Damaging	-4.83	2
CHAF1B	21,37785217,C,T	T366M	medium	Damaging	-4.23	3
CHAF1B	21,37785612,C,T	R498W	low	Damaging	-2.73	2
CTC1	17,8138536,A,G	L425P	medium	Damaging	-3.96	3
DCLRE1A	10,115594988,C,T	V1016I	high	Damaging	-0.97	2
DCLRE1A	10,115607074,C,T	A716T	medium	Damaging	-3.95	3
DCLRE1A	10,115612553,G,A	S130F	medium	Damaging	-2.51	3
DDB1	11,61077336,G,A	T833I	medium	Damaging	-5.71	3
DDX11	12,31249877,C,T	A591V	medium	Damaging	-2.75	3
DYNC1H1	14,102516823,G,A	V4622M	medium	Damaging	-1.6	2
EEPD1	7,36194388,G,A	G152D	high	Damaging	-5.22	3
EEPD1	7,36320746,C,T	P318L	low	Damaging	-4.35	2
EEPD1	7,36327344,C,T	R425W	low	Damaging	-2.5	2
EEPD1	7,36338681,G,A	G526R	medium	Damaging	-4.92	3
EEPD1	7,36338721,A,G	Y539C	low	Damaging	-3.67	2
ERCC1	19,45917271,C,T	V242M	medium	Damaging	-2.48	2
ERCC1	19,45923658,G,A	R117C	medium	Damaging	-4.94	3
ERCC2	19,45854939,A,G	L744P	medium	Damaging	-4.48	3
ERCC2	19,45855574,G,A	R695C	medium	Damaging	-4.77	3
ERCC2	19,45855588,C,T	R690Q	low	Damaging	-3.04	2
ERCC2	19,45855889,G,A	R641W	medium	Damaging	-7.07	3
ERCC2	19,45855906,G,A	A635V	medium	Damaging	-3.76	3
ERCC2	19,45856370,C,T	R601Q	high	Damaging	-3.81	3
ERCC2	19,45856397,C,T	R592H	medium	Damaging	-4.69	3
ERCC2	19,45864824,G,A	L399F	medium	Damaging	-2.28	2
ERCC2	19,45867291,G,A	T301M	medium	Damaging	-3.57	3
ERCC2	19,45867721,G,A	R227C	medium	Damaging	-5.21	3
ERCC2	19,45867790,C,T	V204M	medium	Damaging	-2.28	2

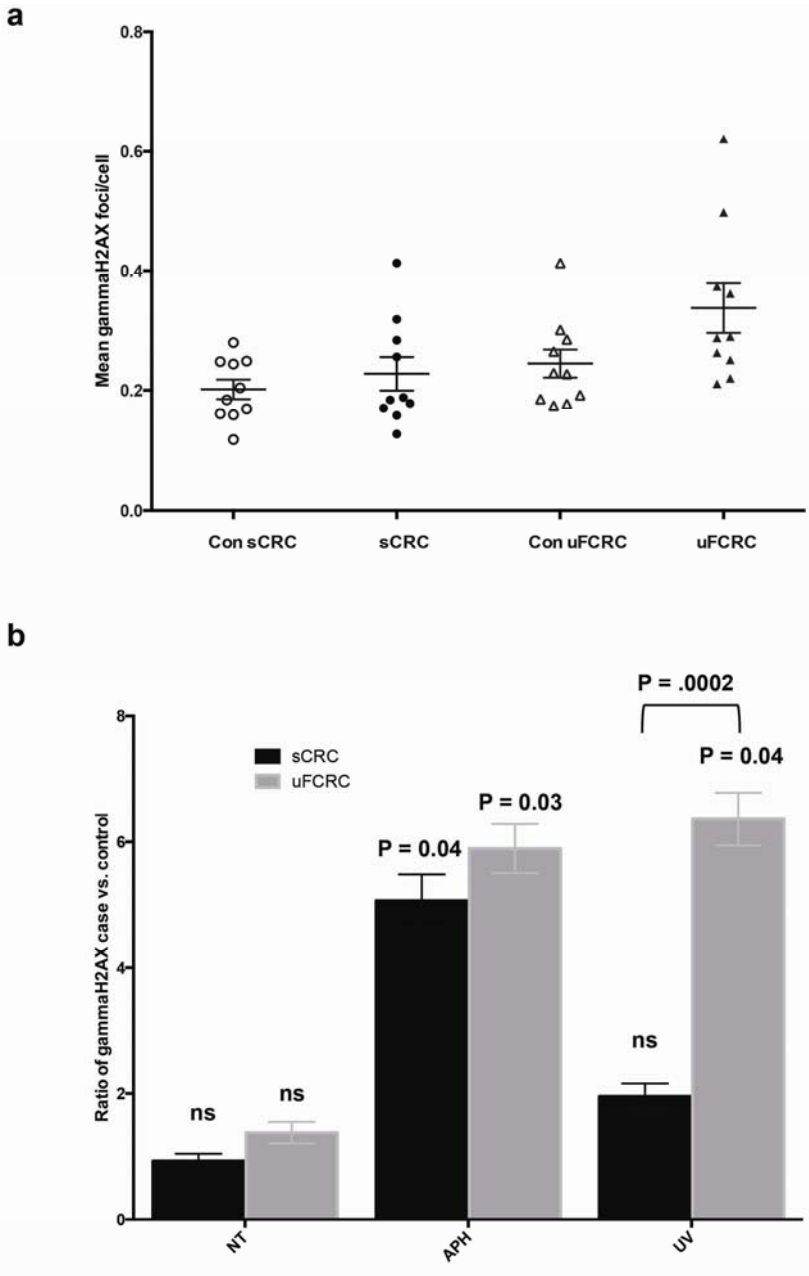
ERCC2	19,45868103,C,T	R196Q	high	Damaging	-3.31	3
ERCC2	19,45868404,G,A	R125C	medium	Damaging	-5.47	3
ERCC2	19,45872347,A,G	L55P	medium	Damaging	-6.32	3
ERCC2	19,45873449,T,C	Y16C	medium	Tolerated	-5.11	2
ERCC3	2,128018872,C,T	D666N	medium	Damaging	-4.84	3
ERCC3	2,128038147,G,A	A468V	high	Damaging	-3.83	3
ERCC3	2,128047345,C,T	V193M	medium	Damaging	-2.25	2
ERCC4	16,14024712,T,C	L313P	medium	Damaging	-4.78	3
ERCC4	16,14026028,G,C	D330H	high	Damaging	-5.94	3
ERCC4	16,14031663,C,T	R618C	medium	Damaging	-3.99	3
ERCC4	16,14031664,G,A	R618H	medium	Damaging	-2.32	2
ERCC4	16,14041570,T,C	I706T	low	Damaging	-4.1	2
ERCC4	16,14041701,C,T	R750C	high	Damaging	-6.35	3
ERCC4	16,14041848,C,T	R799W	medium	Damaging	-6.12	3
BIVM-ERCC5	13,103504591,G,A	R71H	medium	Damaging	-2.89	3
ERCC5	13,103514032,C,T	S283F	medium	Damaging	-3.93	3
ERCC5	13,103514049,T,C	Y289H	medium	Damaging	-3.2	3
ERCC5	13,103514060,A,G	I292M	medium	Damaging	-1.97	2
BIVM-ERCC5	13,103520471,C,T	R848W	medium	Damaging	-4.42	3
BIVM-ERCC5	13,103524556,G,C	W896S	medium	Damaging	-7.73	3
BIVM-ERCC5	13,103525619,C,T	R964W	low	Damaging	-2.82	2
BIVM-ERCC5	13,103527733,G,A	R1014K	medium	Damaging	-1.43	2
BIVM-ERCC5	13,103527791,A,T	E1033D	medium	Damaging	-1.21	2
ERCC6	10,50666943,C,T	R1467Q	medium	Damaging	-2.49	2
ERCC6	10,50669416,C,A	G1322V	medium	Damaging	-4.57	3
ERCC6	10,50678245,T,C	K1254R	medium	Damaging	-2.82	3
ERCC6	10,50680998,A,G	V929A	medium	Damaging	-3.8	3
ERCC6	10,50690906,G,A	R666C	medium	Damaging	-5.42	3
ERCC6	10,50732202,T,G	D425A	medium	Damaging	-3.65	3
ERCC6	10,50740716,C,T	V99I	medium	Damaging	-0.51	2
ERCC8	5,60194136,C,A	R270S	medium	Damaging	-4.98	3
ERCC8	5,60199475,C,T	G184S	medium	Tolerated	-4.13	2
ERCC8	5,60224715,T,C	D50G	low	Damaging	-4	2
ESPL1	12,53676058,G,A	G877D	medium	Damaging	-3.77	3
ESPL1	12,53680689,G,A	R1390H	medium	Damaging	-2.04	2
FANCL	2,58390062,A,G	L281S	medium	Damaging	-2.14	2
FANCL	2,58392928,C,T	D208N	medium	Damaging	-4.45	3
FANCL	2,58459236,G,C	F36L	medium	Damaging	-3.69	3
GTF2H1	11,18387397,G,A	R543H	low	Damaging	-2.5	2
GTF2H3	12,124144129,C,G	L228V	medium	Damaging	-1.97	2
GTF2H3	12,124144450,G,T	G265C	medium	Damaging	-7.95	3
GTF2H3	12,124144623,C,T	T284M	medium	Damaging	-5.4	3
LIG1	19,48620943,C,A	K845N	medium	Damaging	-4.41	3
LIG1	19,48620981,C,G	D833H	high	Damaging	-6.63	3
LIG1	19,48624522,C,T	A764T	medium	Damaging	-2.53	3
LIG1	19,48626566,G,A	R672C	medium	Damaging	-5.92	3
LIG1	19,48631177,C,A	R641L	high	Damaging	-6.99	3
LIG1	19,48639372,C,T	G448R	medium	Damaging	-7.99	3
LIG1	19,48643312,G,A	L335F	medium	Damaging	-3.66	3
LIG1	19,48646806,G,A	S303F	medium	Damaging	-5.36	3

LIG3	17,33310476,G,A	R151Q	medium	Damaging	-3.79	3
LIG3	17,33318086,C,T	R332C	medium	Damaging	-5.79	3
LIG3	17,33326328,G,A	G706S	medium	Damaging	-5.79	3
LIG3	17,33326434,T,C	L741P	medium	Damaging	-6.12	3
MASTL	10,27450076,C,T	T173M	low	Damaging	-5.67	2
MASTL	10,27456090,C,T	L291F	medium	Damaging	-1.77	2
MASTL	10,27456142,C,A	T308N	medium	Damaging	-1.42	2
MASTL	10,27459243,C,A	S452Y	medium	Damaging	-2.3	2
MASTL	10,27459497,G,A	D537N	medium	Damaging	-1.68	2
MASTL	10,27459845,G,C	V653L	medium	Damaging	-1.32	2
MASTL	10,27459855,G,C	R656P	medium	Damaging	-2.28	2
MCM10	10,13228184,T,A	H374Q	medium	Tolerated	-2.93	2
MCM10	10,13228213,C,G	A384G	medium	Damaging	-2.3	2
MCM10	10,13230942,G,A	R427H	medium	Damaging	-3.85	3
MCM10	10,13231019,C,T	R453W	low	Damaging	-3.88	2
MCM10	10,13231065,C,G	S468W	neutral	Damaging	-2.81	2
MCM10	10,13246349,C,T	P829L	medium	Damaging	-9.11	3
MCM10	10,13246360,T,C	C833R	medium	Damaging	-11	3
MCM10	10,13251099,G,A	G837D	medium	Damaging	-4.18	3
MCM10	10,13251117,G,A	R843Q	medium	Damaging	-3.3	3
MCM10	10,13251278,G,A	E866K	medium	Damaging	-3.63	3
MIS18BP1	14,45686268,T,A	H986L	medium	Damaging	-6.22	3
MIS18BP1	14,45687647,A,G	S894P	medium	Damaging	-3.02	3
MIS18BP1	14,45693152,A,G	W880R	medium	Damaging	-10.18	3
MIS18BP1	14,45693698,T,G	T698P	low	Damaging	-2.96	2
MMS19	10,99218991,G,A	P985S	medium	Damaging	-4.04	3
MMS19	10,99219865,C,T	R865H	medium	Damaging	-2.94	3
MMS19	10,99219890,G,A	H857Y	medium	Damaging	-4.3	3
MMS19	10,99219980,G,A	L827F	medium	Damaging	-2.58	3
MMS19	10,99226333,G,T	S478Y	low	Damaging	-2.93	2
MMS19	10,99228103,A,C	L375W	medium	Damaging	-5.45	3
MMS19	10,99229410,C,T	R306H	medium	Damaging	-3.68	3
MNAT1	14,61201615,C,T	T12I	low	Damaging	-5.26	2
MPG	16,133061,G,A	R109Q	high	Damaging	-3.43	3
MPG	16,133099,G,A	V122M	high	Damaging	-2.78	3
MPG	16,133143,C,A	H136Q	high	Tolerated	-6.54	2
MPG	16,133199,C,A	T155N	medium	Damaging	-4.59	3
MPG	16,133222,G,A	G163S	high	Damaging	-5.83	3
MPG	16,135498,C,T	R207C	medium	Damaging	-4.39	3
MPG	16,135717,C,T	R280W	medium	Damaging	-3.66	3
MPG	16,135720,G,A	G281S	medium	Damaging	-4.28	3
MSH2	2,47630468,C,G	H46Q	medium	Damaging	-6.71	3
MSH2	2,47630523,A,G	K65E	medium	Damaging	-3.18	3
MSH2	2,47635588,C,A	S87Y	medium	Damaging	-3.7	3
MSH2	2,47641430,C,T	A272V	medium	Damaging	-3.5	3
MSH2	2,47641520,T,C	L302S	medium	Damaging	-5.57	3
MSH2	2,47643513,C,G	L341V	high	Damaging	-2.93	3
MSH2	2,47693886,C,T	R534C	medium	Damaging	-7.68	3
MSH2	2,47693887,G,A	R534H	medium	Damaging	-4.72	3
MSH2	2,47702202,G,T	A600S	medium	Damaging	-2.72	3

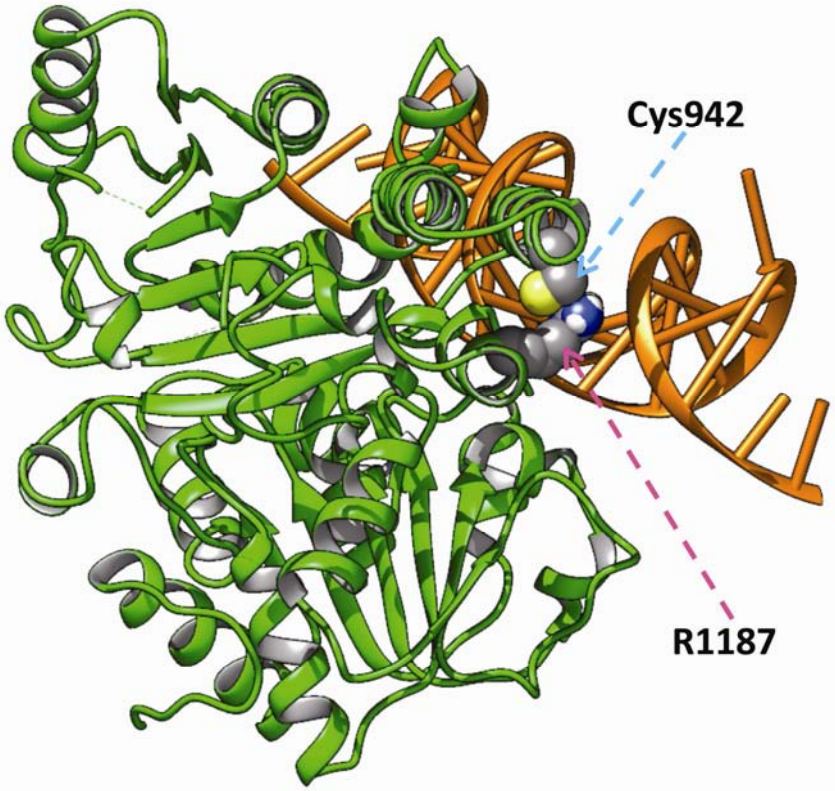
MSH2	2,47710015,T,G	L911R	low	Damaging	-4.26	2
NCAPD3	11,134046212,C,A	D1031Y	low	Damaging	-4.43	2
NCAPD3	11,134047319,T,C	K939R	medium	Damaging	-2.69	3
NCAPD3	11,134062723,G,C	Q636E	medium	Damaging	-1.91	2
NCAPD3	11,134062766,C,G	L621F	medium	Damaging	-3.07	3
POLD1	19,50905504,G,A	R211H	low	Damaging	-3.35	2
POLD1	19,50905989,G,A	G321S	low	Damaging	-5.36	2
POLD1	19,50905992,C,T	R322C	medium	Damaging	-7.57	3
POLD1	19,50909683,A,G	K468R	medium	Damaging	-2.91	3
POLD1	19,50910297,C,A	L518M	high	Damaging	-1.97	2
POLD1	19,50910352,C,T	A536V	high	Damaging	-3.88	3
POLD1	19,50912813,C,T	R682W	medium	Damaging	-5.8	3
POLD1	19,50917074,C,T	R776W	high	Damaging	-6.69	3
POLD1	19,50919693,C,G	T954R	medium	Damaging	-3.75	3
POLD1	19,50919693,C,T	T954M	medium	Damaging	-3.31	3
POLD1	19,50920465,C,T	R1053C	high	Damaging	-6.77	3
POLD3	11,74340350,G,A	D330N	medium	Damaging	-2.72	3
POLD3	11,74345670,C,T	S350F	medium	Damaging	-1.98	2
POLD3	11,74345739,A,C	K373T	low	Damaging	-2.76	2
RAD23B	9,110086255,C,T	A301V	medium	Damaging	-3.05	3
RPA1	17,1747233,C,T	T35M	medium	Damaging	-2	2
RPA1	17,1782352,C,G	N252K	low	Damaging	-3.57	2
RPA1	17,1782542,G,T	A265S	medium	Damaging	-2.97	3
RPA1	17,1782605,G,T	V286F	medium	Damaging	-3.83	3
RPA1	17,1783909,C,T	R389W	medium	Damaging	-5.53	3
RPA1	17,1792054,A,G	N487S	medium	Damaging	-4.03	3
RPA3	7,7676650,C,A	G116V	low	Damaging	-3.35	2
RPA3	7,7677536,C,T	C81Y	low	Damaging	-5.95	2
SKA1	18,47902248,G,T	K16N	medium	Damaging	-2.56	3
SPAG5	17,26910923,G,A	R832W	low	Damaging	-4.89	2
TEX14	17,56676322,T,C	Y801C	medium	Damaging	-4.85	3
TEX14	17,56676493,T,C	Q744R	medium	Damaging	-2.63	3
TEX14	17,56676845,C,T	E627K	medium	Damaging	-2.29	2
TEX14	17,56682353,T,C	Q447R	medium	Damaging	-3.07	3
TEX14	17,56688585,T,C	Y380C	low	Damaging	-7.06	2
TEX14	17,56688600,C,T	R375H	high	Damaging	-2.87	3
TEX14	17,56688648,T,C	Q359R	high	Damaging	-3.4	3
TEX14	17,56688700,G,A	R342W	medium	Damaging	-5.53	3
TEX14	17,56690823,G,A	R328C	low	Damaging	-5.27	2
TEX14	17,56692633,G,A	R287W	medium	Damaging	-4.9	3
TEX14	17,56700278,C,T	R116Q	neutral	Damaging	-3.43	2
TEX14	17,56700329,G,A	S99L	low	Damaging	-4.88	2
WRN	8,30916020,G,A	M19I	medium	Damaging	-2.19	2
WRN	8,30916717,T,A	F49I	medium	Damaging	-3.42	3
WRN	8,30922518,T,A	L148Q	low	Damaging	-2.88	2
WRN	8,30924631,G,A	R196H	high	Damaging	-3.08	3
WRN	8,30924661,C,T	T206I	medium	Damaging	-3.09	3
WRN	8,30948045,A,G	T573A	high	Damaging	-4.71	3
WRN	8,30949351,C,G	S612C	low	Damaging	-3.38	2
WRN	8,30954294,C,T	R637W	medium	Damaging	-6.76	3

WRN	8,30969150,C,T	T703I	high	Damaging	-5.63	3
WRN	8,30998956,G,A	R993H	medium	Damaging	-3.98	3
WRN	8,31004622,C,T	S1146L	medium	Tolerated	-3.23	2
WRN	8,31012237,C,G	T1262R	medium	Damaging	-3.79	3
XAB2	19,7685473,C,T	R685Q	high	Damaging	-3.91	3
XAB2	19,7685492,C,T	G679R	medium	Damaging	-7.84	3
XAB2	19,7685860,C,T	V615M	medium	Damaging	-2.63	3
XAB2	19,7687309,G,A	R509C	medium	Damaging	-6.94	3
XAB2	19,7687766,A,G	L418P	low	Damaging	-4.17	2
XAB2	19,7687768,G,C	I417M	medium	Damaging	-2.36	2
XAB2	19,7687776,G,A	R415C	medium	Damaging	-7.8	3
XAB2	19,7688667,C,T	V357M	medium	Damaging	-2.9	3
XAB2	19,7690786,G,A	R268C	medium	Damaging	-7.25	3
XAB2	19,7690839,C,T	R250H	medium	Damaging	-3.98	3
XAB2	19,7690855,G,A	R245C	medium	Damaging	-7.67	3
XPC	3,14190078,C,T	G802S	medium	Damaging	-4.52	3
XPC	3,14190117,C,T	D789N	medium	Damaging	-4.63	3
XPC	3,14190421,G,A	R715W	medium	Damaging	-4.24	3
XPC	3,14197947,A,G	Y641H	medium	Damaging	-4.13	3
XPC	3,14206351,C,G	A288P	low	Damaging	-2.98	2

Arora et al, Supp Figure S1

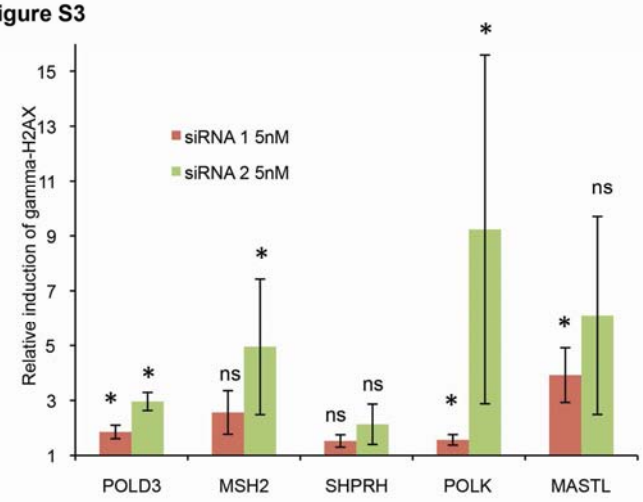


Arora et al, Supp Figure S2

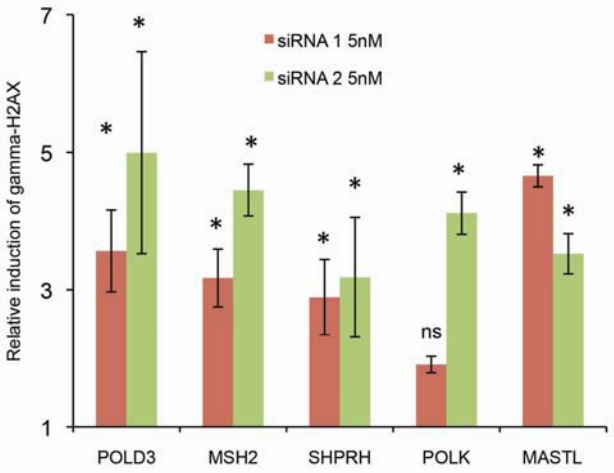


Arora et al, Supp Figure S3

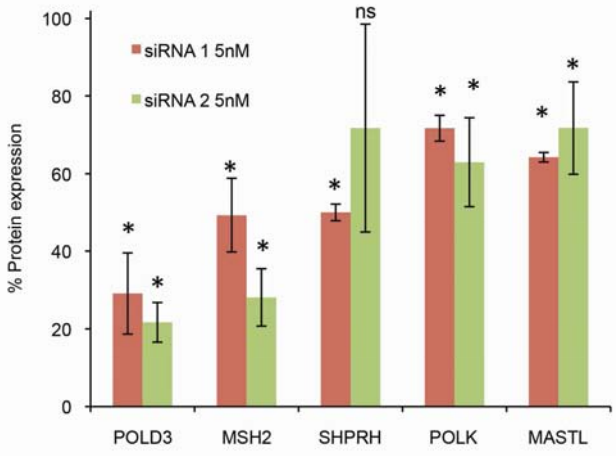
a



b

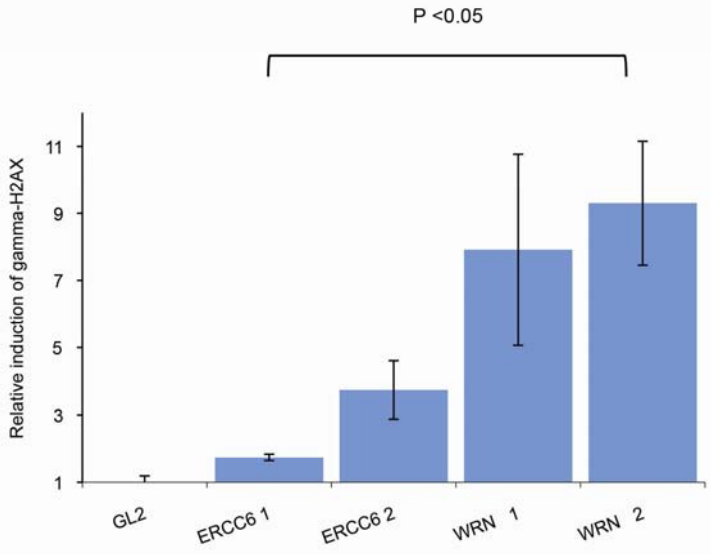


c

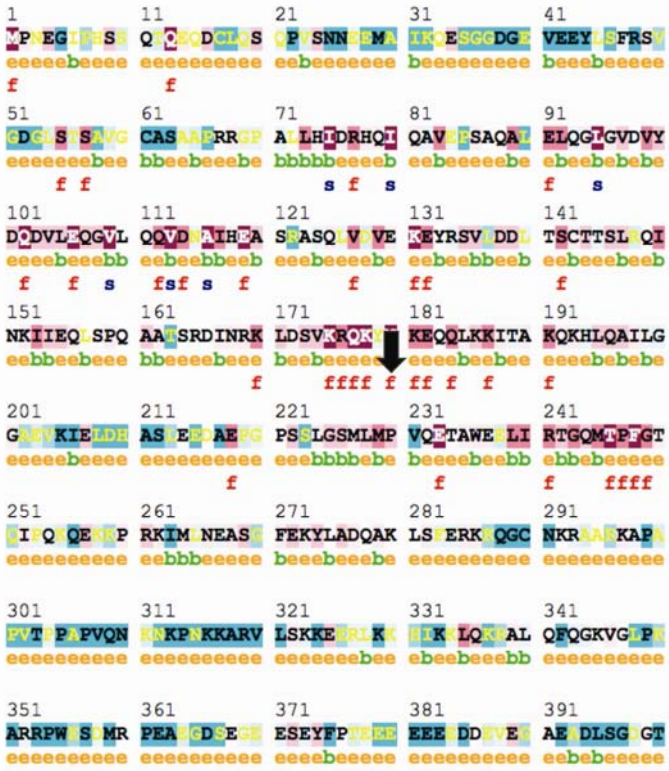


Arora et al, Supp Figure S4

a



Arora et al, Supp Figure S5



The conservation scale:

1	2	3	4	5	6	7	8	9
Variable			Average				Conserved	

e - An exposed residue according to the neural-network algorithm.

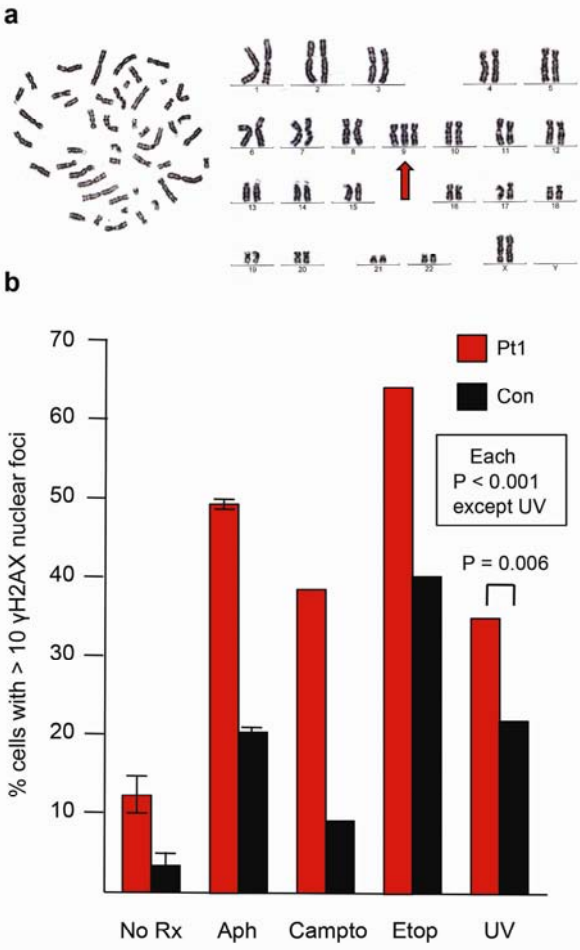
b - A buried residue according to the neural-network algorithm.

f - A predicted functional residue (highly conserved and exposed).

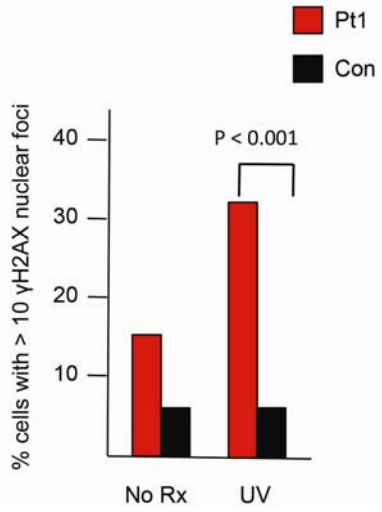
s - A predicted structural residue (highly conserved and buried).

X - Insufficient data - the calculation for this site was performed on less than 10% of the sequences.

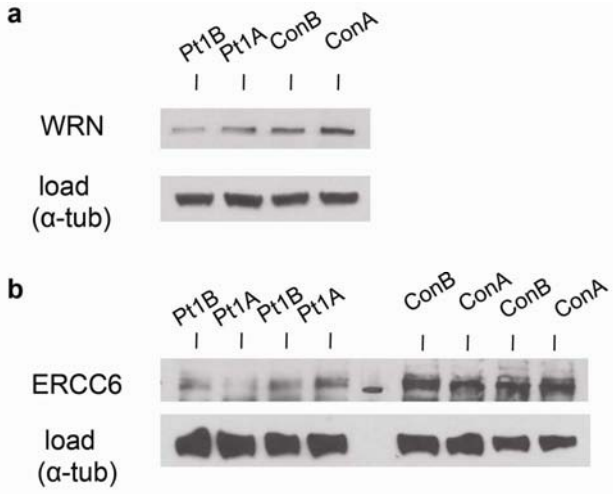
Arora et al, Supp Figure S6



Arora et al, Supp Figure S7



Arora et al, Supp Figure S8



Arora et al, Supp Figure S9