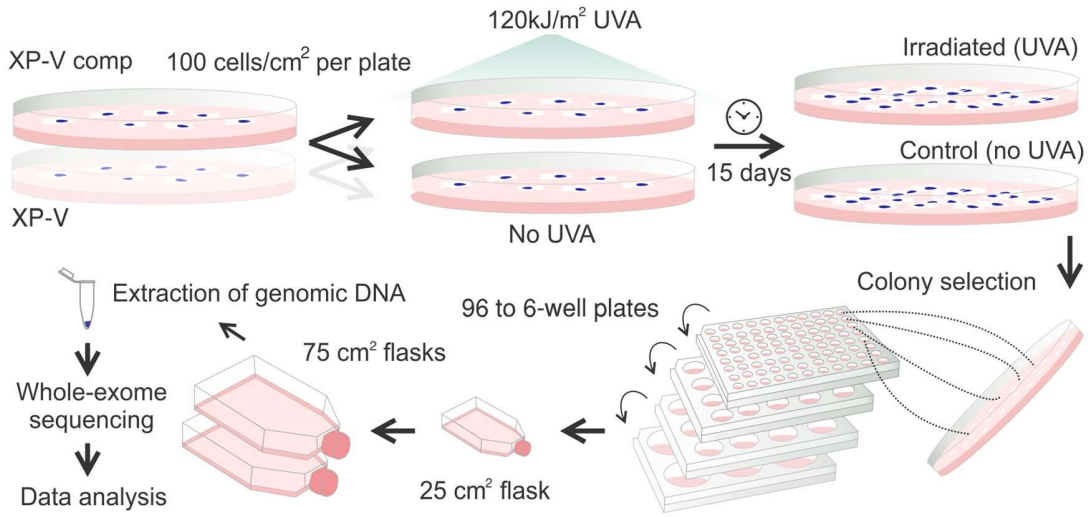
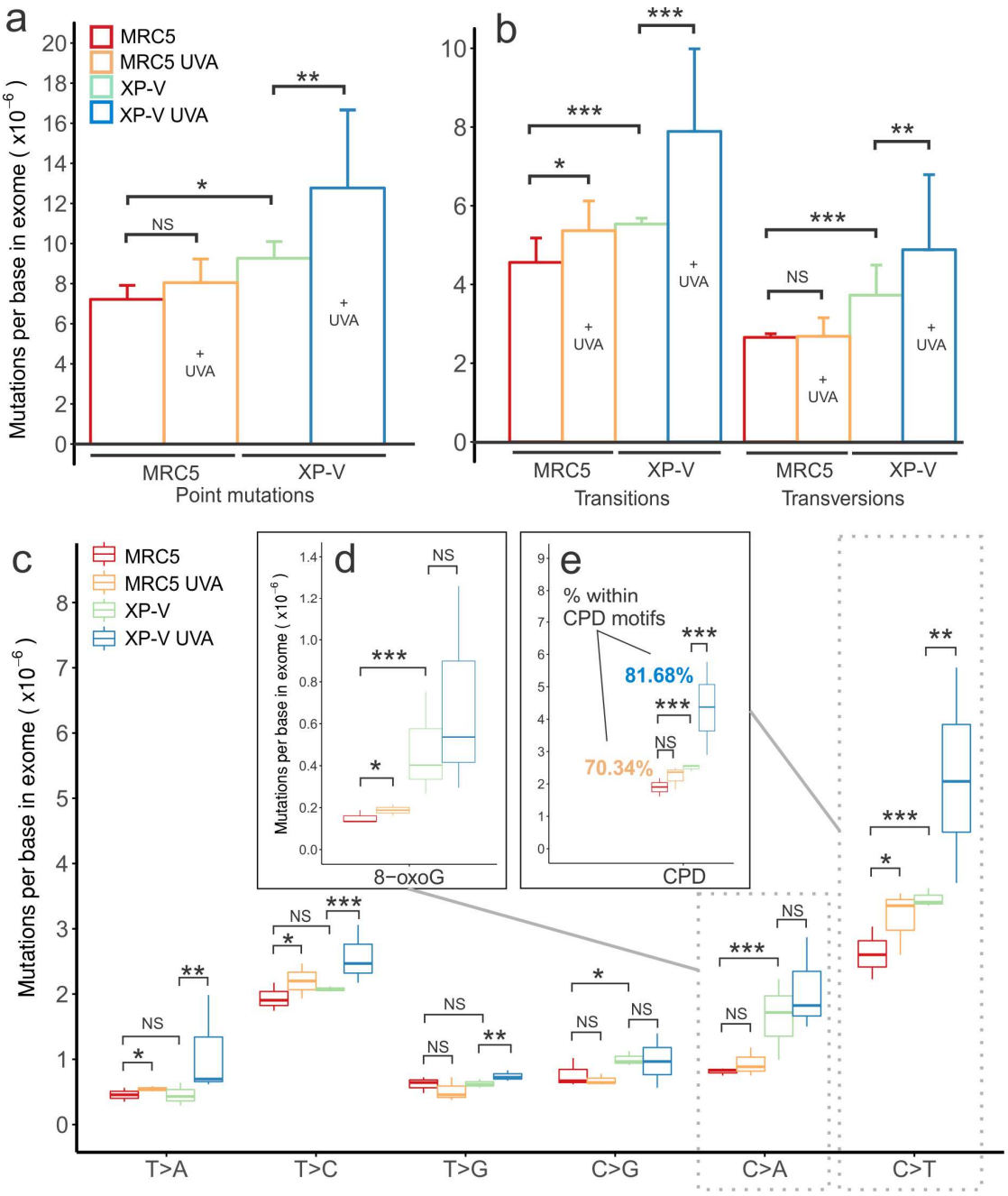
Whole-exome sequencing reveals the impact of UVA light mutagenesis in xeroderma pigmentosum variant human cells

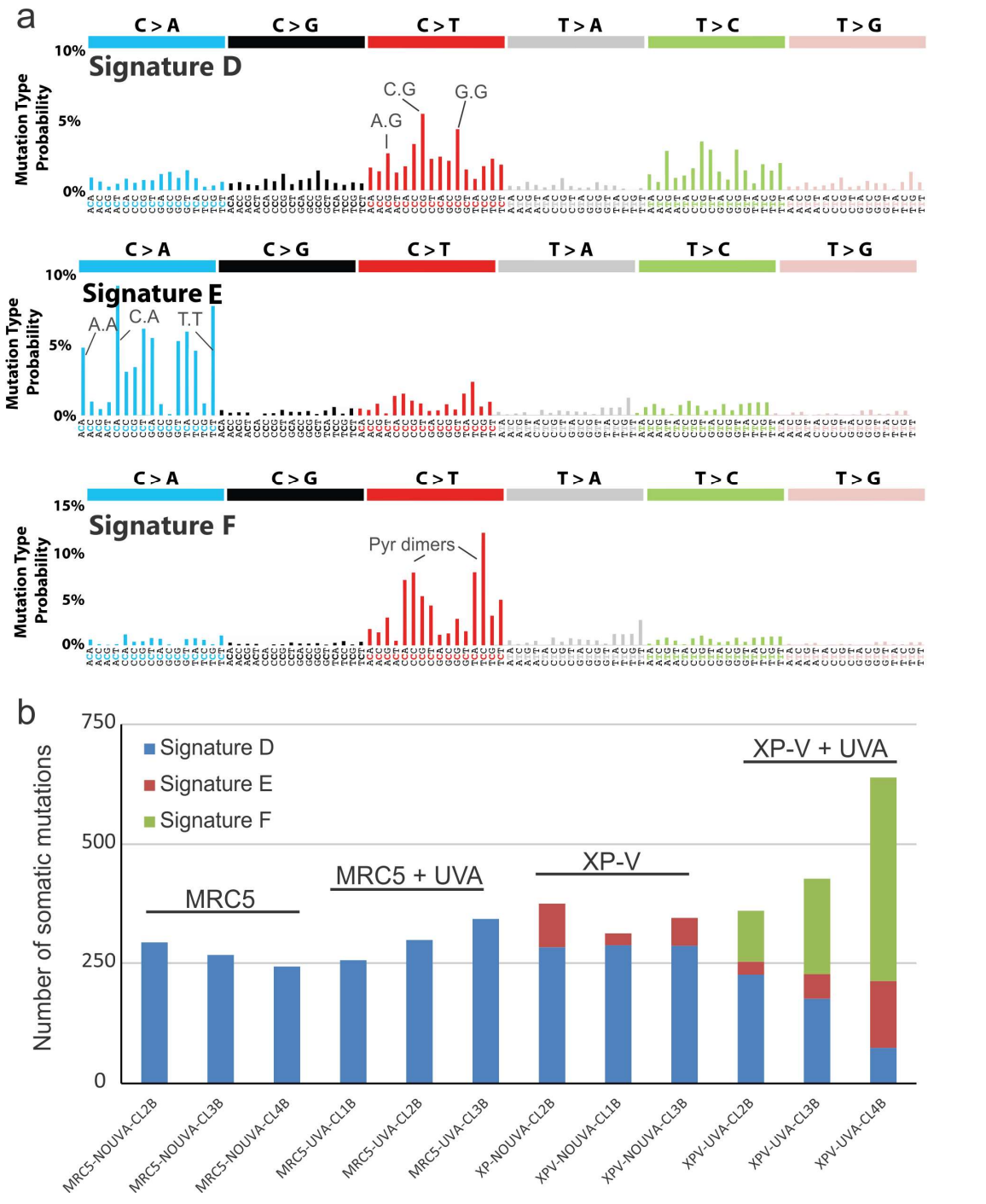
Moreno and Souza et al. Supplementary data



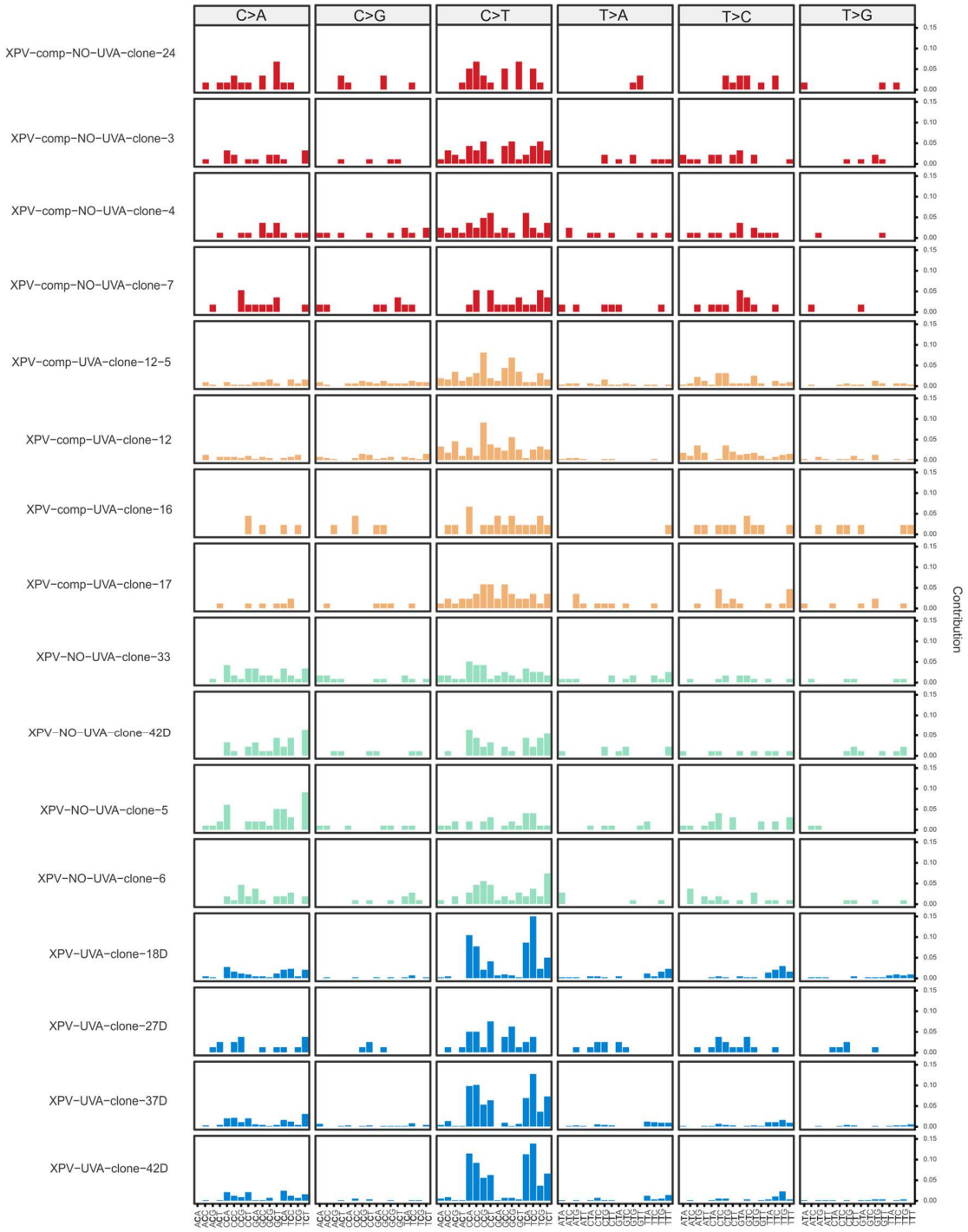
Supplementary Figure 1. **Experimental strategy to detect point mutations using whole-exome sequencing (WES)**. XP-V complemented (XP-V comp), and XP-V cells (XP-V) were irradiated with 120 kJ/m² of UVA light and then subjected to a 45-day cloning procedure based on colony selection and expansion. Genomic DNA was extracted from four clones of each cell type and treatment. Whole-exome enrichment was performed to construct libraries for sequencing on HiSeq Platform. Read mapping and SNP calling were done using BWA-GATK pipeline.



Supplementary Figure 2. Point mutation profiles of MRC5 and XP-V cell-lines with UVA using SOLiD data sequencing. (**a**) Total of point mutations per base in exome in all four groups of cell-lines sequenced. (**b**) Total of transitions and transversions per base in exome in the four groups of cell-lines. (**c**) T>A, T>C, T>G, C>G, C>A, and C>T mutations per base in exome. (**d**) Selection of all C>A mutations within RGR sequence motif beyond strand bias. (**e**) Selection of all C>T mutations within YY sequence (pyrimidine dimer) motif beyond strand bias. Statistical differences were calculated by non-parametric permutation tests: $P < 0.01$ (***), $0.01 < P < 0.025$ (**), $0.025 < P < 0.05$ (*), $P > 0.05$ (NS). Data represent mean ± SD. Y = IUPAC code for T or C nucleotides. R = IUPAC code for A or G nucleotides.



Supplementary Figure 3. Mutational signatures extracted using SOLiD data sequencing from MRC5 and XP-V cell-lines. Each clone was used in a signature analysis based on the framework based on the NMF method[1]. (**a**) Mutational signatures D, E, and F extracted using all clones of all cell-types. (**b**) Contribution of each of the three signatures extracted (D,E,F) in all samples sequenced.



Supplementary Figure 4. Somatic spectrum of mutations in each cell clone sequenced. Somatic spectrum of point mutations and its trinucleotide context sequence extracted in each cell line clone. Clones colored in red, orange, light green and blue corresponds to XP-V comp, XP-V comp UVA, XP-V, and XP-V UVA, respectively. Trinucleotide contribution was calculated by its frequency.

Supplementary Table 1. Sequencing, mapping, and SNP calling statistics of all clones sequenced using Illumina HiSeq platform and SureSelect exome enrichment kit.

Cell line	XP-V comp								XP-V							
Treatment	No UVA	No UVA	No UVA	No UVA	UVA	UVA	UVA	UVA	No UVA	No UVA	No UVA	No UVA	UVA	UVA	UVA	UVA
Name	Clone 4	Clone 24	Clone 3	Clone 7	Clone 12	Clone 12-5	Clone 16	Clone 17	Clone 5	Clone 6	Clone 42D	Clone 33	Clone 18D	Clone 27D	Clone 37D	Clone 42D
Total reads	54179838	59284764	25094523	29006248	47909807	25693404	25866712	62072506	24737424	57247274	69301008	54410125	52802182	48115905	53244891	68190455
Number of reads mapped	54021817	59075745	25074274	28986168	47758306	25671034	25834976	61820131	24713754	57038353	69074681	54206650	52640171	47959378	53080612	67940011
% reads mapped	99.71%	99.65%	99.92%	99.93%	99.68%	99.91%	99.88%	99.59%	99.90%	99.64%	99.67%	99.63%	99.69%	99.67%	99.69%	99.63%
Bases on target	5692399360	3675935895	2212490076	2642954163	4882986866	2513919062	2479714982	6783912997	2184064569	6104793803	7506421663	5870863777	5491025608	5038229178	5488468167	7142231902
Percent bases on target	87.43%	88.07%	82.53%	85.80%	88.81%	89.55%	86.77%	88.23%	81.74%	88.10%	87.69%	87.30%	86.21%	88.36%	86.98%	87.59%
Bases off target	1021283832	1073441438	606302689	574600044	794955625	376121139	482075536	1155015976	623970993	1029770069	1321069719	1066841659	1093607496	840368568	1033709571	1281838642
Percent off target	12.57%	11.19%	17.47%	14.20%	11.19%	10.45%	13.23%	11.77%	18.26%	11.90%	12.31%	12.70%	13.79%	11.65%	13.02%	12.41%
Percent of target bp not covered	2.72%	2.74%	2.94%	2.73%	2.76%	2.79%	2.87%	2.72%	2.91%	2.67%	2.67%	2.76%	2.73%	2.78%	2.90%	2.80%
Percent of target bp covered at ≥ 1X	96.94%	96.69%	96.61%	96.91%	95.95%	96.73%	96.68%	96.88%	96.62%	97.04%	96.99%	96.83%	96.92%	96.79%	96.76%	96.92%
Percent of target bp covered at ≥ 10X	93.14%	92.93%	82.64%	86.21%	89.87%	84.07%	84.72%	92.96%	80.52%	93.76%	94.06%	92.74%	93.00%	91.70%	92.76%	94.10%
Percent of target bp covered at ≥ 20X	86.15%	86,98%	55.59%	64.26%	80.59%	59.96%	60.07%	87.42%	52.80%	86.92%	89.54%	85.87%	85.58%	83.13%	85.47%	89.48%
Average depth of coverage	94.16	104.16	36.60	43.72	80.77	41.58	41.02	112.21	36.13	100.98	124.16	97.11	90.83	83.33	90.78	118.14
Raw variants called	49893	49823	48001	48582	47566	46658	48280	49895	48230	50858	51301	50905	51659	50485	51022	52309
Novel SNPs	1615	1550	1444	1452	1155	1090	1435	1672	1277	1386	1529	1517	2230	1416	2884	2554
Percent SNPs (dbSNP)	96.76%	96.89%	96.99%	97.01%	97.57%	97.66%	97.03%	96.65%	97.35%	97.27%	97.02%	97.02%	95.68%	97.20%	94.35%	95.12%
Raw Indels called	4573	4612	4079	4085	4480	4031	4097	4652	3965	4657	4779	4635	4660	4627	4521	4723
Novel Indels	369	368	317	301	342	288	294	383	298	348	393	348	358	370	361	348
Percent Indels (dbSNP)	91.93%	92.02%	92.23%	92.63%	92.37%	92.86%	92.82%	91.77%	92.48%	92.53%	91.78%	92.49%	92.32%	92.00%	92.02%	92.63%
Exclusive SNVs	210	165	257	159	1209	849	143	221	270	243	221	304	919	198	1536	1206
Exonic/splicing exclusive SNVs	83	59	93	57	396	321	45	86	99	108	93	119	441	80	754	578
Non-synonymous SNVs	55	40	61	35	196	180	33	57	78	76	61	80	280	54	461	362

Supplementary Table 2. Sequencing, mapping, and SNP calling statistics of all clones sequenced using SOLiD 5500XL platform and TargetSeq exome enrichment kit.

Cell line	MRC5						XP-V					
Treatment	UVA	UVA	UVA	No UVA	No UVA	No UVA	UVA	UVA	UVA	No UVA	No UVA	No UVA
Name	cl1b	cl2b	cl3b	cl2b	cl3b	cl4b	cl2b	cl3b	cl4b	cl1b	cl2b	cl3b
Total reads	138071194	132813002	137092092	122189552	124050652	133681560	117975188	116340286	122784478	127684938	124110282	99561778
Number of reads mapped	119918359	115714515	120234337	108131487	110276082	116316194	104944385	102131864	107606654	112424729	109632143	87613550
% reads mapped	86.85%	87.13%	87.70%	88.49%	88.90%	87.01%	88.95%	87.79%	87.64%	88.05%	88.33%	88.00%
Reads on target [a]	50924905	50145243	50803594	51945436	53965486	48933172	51784896	49284839	51730897	55335029	52671532	41680620
Percent reads on target [b]	51.97%	53.36%	51.92%	59.40%	60.48%	51.60%	60.27%	59.72%	58.94%	60.51%	58.77%	58.45%
Reads off target [c]	47069615	43834724	47051761	35500717	35256362	45897518	34132925	33244696	36038989	36119306	36947990	29623578
Percent off target [d]	48.03%	46.64%	48.08%	40.60%	39.52%	48.40%	39.73%	40.28%	41.06%	39.49%	41.23%	41.55%
Percent of target bp not covered: [e]	6.63%	6.43%	7.24%	6.55%	6.38%	6.93%	6.39%	6.04%	6.31%	5.90%	6.32%	7.02%
Percent of target bp covered at ≥ 1X:	93.37%	93.57%	92.76%	93.45%	93.62%	93.07%	93.61%	93.96%	93.69%	94.10%	93.68%	92.98%
Percent of target bp covered at ≥ 10X:	80.44%	81.08%	79.18%	80.71%	81.47%	79.34%	81.32%	82.02%	81.61%	82.67%	81.55%	78.92%
Percent of target bp covered at ≥ 20X:	71.24%	71.93%	70.17%	71.70%	72.65%	69.82%	72.46%	73.04%	72.98%	74.22%	72.82%	68.79%
Average depth of coverage [f]	61.08	60.26	60.88	62.33	64.89	58.66	62.19	59.26	62.13	66.82	63.43	50.02
Raw variants called	33363	33463	33226	33540	33784	32896	35796	35981	35648	36309	36022	34108
Exclusive SNVs	639	694	859	722	697	669	835	961	1349	703	846	736
Exonic/splicing exclusive SNVs	256	300	344	295	269	243	361	428	639	314	376	346
Total non-synonymous SNVs	155	165	205	157	162	163	229	257	427	177	242	212

Supplementary Table 3. Transcriptional strand bias score (SC) associated to mutagenic motifs.

Cell-line	CPD	8-oxoG	G to T
XP-V comp	0.97 ± 0.37	0.83 ± 0.33	0.81 ± 0.30
XP-V comp UVA	0.89 ± 0.35	0.63 ± 0.48	0.79 ± 0.58
XP-V	1.47 ± 0.34	1.31 ± 0.24	0.79 ± 0.12
XP-V UVA	1.48 ± 0.60	1.55 ± 1.66	1.04 ± 0.53

Transcriptional strand bias score (SC) is measured by concordance:discordance ratio for the number of mutational motifs detected in concordance (coding strand as annotated in RefSeq) and discordance (motifs found in the transcribed strand). A strand score of 1 means no transcriptional strand bias for a motif. Values > 1 means more motifs in concordance with coding strand but not subjected to TCR. Values < 1 means more motifs in the transcribed strand (discordance) and potentially subjected to TCR. CPD mutagenic motifs were defined as C>T changes in YY context sequences. 8-oxoG motifs were G>T changes in RGR motifs. Data represent mean ± SD. Y = IUPAC code for T or C nucleotides. R = IUPAC code for A or G nucleotides. Data represent mean ± SD.

Supplementary Table 4. Mutations at T:A and C:G in XP-V comp and XP-V cells irradiated with UVA.

Cell-line	T:A (10^{-7})	C:G (10^{-7})	C:G / T:A ratio	T:A at TT sites (10^{-7})
XP-V comp	3.6 ± 0.9	8.5 ± 2.1	0.58	2.2 ± 0.5
XP-V comp UVA	10.9 ± 8.3	24.1 ± 20.2	0.55	7.4 ± 5.3
XP-V	4.4 ± 0.4	12.9 ± 1.5	0.66	3.4 ± 0.6
XP-V UVA	13.1 ± 6.1	63.5 ± 41.8	0.80	11.9 ± 6.1

Data represent mean ± SD.

Supplementary Methods

Color-space reads produced by SOLiD were mapped to the GRCh37/hg19 genome using the LifeScope v2.1 software package (Life Technologies, Carlsbad, California, EUA). SNP calling was performed by using the DiBayes algorithm to detect two adjacent SNPs. SNPs were called only if the altered allele supporting reads were higher than 0.15 and covered at least 2X.

To identify induced point mutations and filter out polymorphisms, unique single nucleotide variants (SNVs) were selected, considering the position and nucleotide change in comparison to all other samples sequenced, despite genotyping assignment. Whenever a variant was detected in two or more independent clones, it was considered a polymorphism and was discarded. Comparisons were performed using VCFtools, and unique SNVs for each sample were annotated with ANNOVAR using RefSeq. Alternatively, an additional filtering step involving only SNVs annotated as “exonic” or “splicing” in each clone sample was applied to restrict the analysis only to regions enriched in exome sequencing.

Supplementary Reference

Alexandrov, L. B., Nik-Zainal, S., Wedge, D. C., Campbell, P. J. and Stratton, M. R. (2013) Deciphering signatures of mutational processes operative in human cancer. Cell Rep., **3**, 246–259.