# Deficiency of global genome nucleotide excision repair explains mutational signature observed in cancer

Myrthe Jager[1,3]

Francis Blokzijl[1,3]

Ewart Kuijk[1]

Maria Vougioukalaki[2]

Roel Janssen[1]

Nicolle Besselink[1]

Sander Boymans[1]

Joep de Ligt[1]

Jan Hoeijmakers[2]

Joris Pothof[2]

Ruben van Boxtel[1,‡,*]

Edwin Cuppen[1,*]

[1]Center for Molecular Medicine, Cancer Genomics Netherlands, University Medical Center Utrecht, Utrecht University, Heidelberglaan 100, 3584 CX Utrecht, The Netherlands

[2]Erasmus Medical Center, Wytemaweg 80, 3015 CN Rotterdam, The Netherlands

[3]These authors contributed equally to this work.

[‡]Present address: Princess Máxima Center for Pediatric Oncology, 3584 CT Utrecht, The Netherlands

[*]Corresponding authors: R.vanBoxtel@prinsesmaximacentrum.nl, ecuppen@umcutrecht.nl

Running title: GG-NER-deficiency explains mutational signature

Keywords: ERCC1, XPC, nucleotide excision repair, NER, adult stem cell, liver, small intestine, organoids, mutational signatures, Signature 8, cancer, progeria, CRISPR-Cas

## ABSTRACT

Nucleotide excision repair (NER) is one of the main DNA repair pathways that protect cells against genomic damage. Deficiency in this pathway can contribute to the development of cancer and accelerate aging. In addition, NER-deficiency is an important determinant for cancer treatment outcome, as NER-deficient tumors are sensitive to cisplatin treatment. Non-silent mutations in the NER-gene *ERCC2* were previously associated with a specific mutational footprint (Signature 5), potentially providing a functional readout for NER-deficiency. However, not all NER-deficient tumors are characterized by a high Signature 5 contribution, illustrating the importance to further characterize the mutational consequences of NER-deficiency. Here, we analyzed the somatic mutational profiles of adult stem cells (ASCs) from NER-deficient progeroid *Ercc1*$^{-/\Delta}$ mice, using whole-genome sequencing analysis of clonally derived organoids. Our results indicate that NER-deficiency increases the point mutation load in liver, but not in small intestinal ASCs, which coincides with a tissue-specific aging-pathology observed in these mice. More specifically, this increase can be largely explained by Signature 8, a mutational footprint observed in human cancer with as yet unknown etiology. The genomic distribution of the acquired point mutations indicates that deficiency of global-genome NER (GG-NER), rather than transcription-coupled NER (TC-NER), is responsible for the accumulated mutations. We independently confirmed the link between Signature 8 and GG-NER-deficiency through mutational analysis of a human organoid culture that was deleted for *XPC* using CRISPR-Cas9 gene-editing. Elevated levels of Signature 8 may, therefore, serve as a novel biomarker for GG-NER-deficient tumors and could improve personalized cancer treatment strategies.

## INTRODUCTION

The genome is continuously exposed to DNA-damaging agents, which ultimately can result in accumulation of mutations or, alternatively, the induction of senescence or cell death. To counteract mutagenic processes, cells exploit multiple DNA repair pathways that each repair specific lesions. Deficiency of these DNA repair pathways can contribute to cancer,

accelerate aging, or both (Hoeijmakers 2009). To increase insight into the cellular processes that underlie mutation accumulation, including the activity of DNA repair pathways, mutation loads and types can be characterized (Alexandrov et al. 2013; Nik-Zainal et al. 2016). To date, systematic analyses of tumor genomes have revealed 30 signatures of point mutations and 6 rearrangement signatures of mutational processes in cancer genomes (Alexandrov et al. 2013; Nik-Zainal et al. 2016). These mutational signatures may have important diagnostic value. For example, several signatures are associated with BRCA1/2 inactivity and can consequently be predictive for a potentially beneficial response to PARP inhibition or cisplatin treatment (Waddell et al. 2015; Davies et al. 2017).

Although for some signatures the underlying molecular process (Kim et al. 2016; Alexandrov et al. 2013, 2016) and/or involved DNA repair pathway (Kim et al. 2016; Davies et al. 2017; Alexandrov et al. 2013) is known, in-depth mechanistic insight is still lacking for the majority of the mutational signatures. Efforts to link mutational processes to specific signatures have mainly focused on associating mutation data from tumors to mutagen exposure and DNA repair-deficiency. However, tumors are genomically highly unstable and typically multiple mutational processes are active simultaneously and/or consecutively (Alexandrov et al. 2013; Nik-Zainal et al. 2016), which hampers the identification of the underlying processes that cause specific mutational signatures. Tissue-specific adult stem cells (ASCs), on the other hand, maintain a highly stable genome during life (Blokzijl et al. 2016a). In addition, ASCs provide a relevant cell source to study mutational patterns, as these cells are the cell-of-origin for specific types of cancer (Barker et al. 2009; Zhu et al. 2016; Adams et al. 2015). Moreover, DNA damage in the genomes of ASCs is suggested to contribute to the depletion of ASCs in tissues, which is one of the hallmarks of aging (López-Otín et al. 2013; Rossi et al. 2008; Hoeijmakers 2009). We recently developed a protocol for studying the mutational consequences of mutagen exposure or inactivation of DNA repair components in non-cancerous ASCs, by combining organoid culturing technology with whole-genome sequencing (WGS) (M Jager, F Blokzijl, V Sasselli, S Boymans, R Janssen, N Besselink, H Clevers, R van Boxtel and E Cuppen, in press).

Nucleotide excision repair (NER) is one of the main cellular DNA repair pathways (Iyama and Wilson 2013). Within NER, two subpathways are recognized: global-genome NER (GG-NER), which repairs bulky helix-distorting lesions throughout the genome, and transcription-coupled NER (TC-NER), which resolves RNA polymerase blocking lesions during transcription (Iyama and Wilson 2013; Marteijn et al. 2014; Hoeijmakers 2009). Identification of NER-deficient tumors has great clinical potential, as these tumors show an increased response to chemotherapeutic agents inducing bulky chemical adducts, distorting helical structure and disrupting base-pairing, such as cisplatin treatment (Stubbert et al. 2010; Amable 2016; Van Allen et al. 2014; Olaussen et al. 2006; Li et al. 2000). Somatic mutations in *ERCC2*, a key factor of NER, were recently associated with Signature 5 in urothelial tumors (Kim et al. 2016). However, NER has been suggested to underlie multiple mutational signatures, based on large-scale tumor mutation analyses (Alexandrov et al. 2013). Indeed, not all NER-deficient tumors are characterized by a high Signature 5 contribution (Kim et al. 2016). It is, therefore, important to determine which other mutational signatures present in human cancer reflect NER-deficiency.

To determine the mutational consequences of NER-deficiency *in vivo*, we used *Ercc1*$^{-/\Delta}$ mice. ERCC1, which forms a complex with XPF, plays an essential role in the core NER-reaction involving both GG-NER and TC-NER (Kirschner and Melton 2010; Iyama and Wilson 2013; Sijbers et al. 1996a; Aboussekhra et al. 1995), crosslink repair (Rahn et al. 2010) and single strand annealing (SSA) of double strand breaks (Al-Minawi et al. 2008). *ERCC1* is mutated in ~4.5% of all human tumors, especially skin and liver cancer (http://dcc.icgc.org), and single nucleotide polymorphisms in *ERCC1* have been linked to an increased risk of developing colorectal cancer (Ni et al. 2014). Germline mutations in *ERCC1* are extremely rare and have only been observed sporadically in humans. These patients develop a severe form of Cockayne syndrome, a progeroid-like disease, called cerebro-oculo-facial-skeletal syndrome (COFS) (Jaspers et al. 2007; Kashiyama et al. 2013).

Similar to patients with deleterious mutations in *ERCC1*, *Ercc1*$^{-/\Delta}$ mice have a reduced lifespan as a result of progeroid-like symptoms. *Ercc1*$^{-/\Delta}$ mice can reach a maximum

age of approximately 26 weeks, which is approximately five times shorter than wild-type (WT) littermates (Dollé et al. 2011; Vermeij et al. 2016). *Ercc1*$^{-/\Delta}$ mice are hemizygous for a single truncated *Ercc1* allele, encoding a protein lacking the last seven amino acids at the C-terminus, which largely corrupts protein function (Dollé et al. 2011; Weeda et al. 1997). Disruption of this domain results in reduced NER-activity (Su et al. 2012). *Ercc1*$^{-/\Delta}$ livers display various aging-like characteristics and pathology such as liver lipofuscin, hepatocellular senescence and cell death, and steatosis (Dollé et al. 2011; Gregg et al. 2012; Niedernhofer et al. 2006; Weeda et al. 1997). In contrast, other organs, such as the small intestine, do not show an obvious pathological phenotype, indicating that the consequences of loss of ERCC1 differ considerably between tissues, although the reason for this remains unclear.

Here, we exploited the organoid culture system to study the mutational consequences of NER-deficiency in mouse and human ASCs. We first determined the mutations that are acquired *in vivo* in the genomes of single ASCs in the liver and small intestine of *Ercc1*$^{-/\Delta}$ mice and WT littermates. Our results indicated that mutation accumulation in *Ercc1*$^{-/\Delta}$ ASCs is predominantly a result of dysfunctional GG-NER. Next, we confirmed the observed mutational consequences of GG-NER-deficiency through targeted deletion of the essential GG-NER component *XPC* (Iyama and Wilson 2013) in single human ASCs using the CRISPR-Cas9 gene-editing technique in human organoids.

## RESULTS

### Loss of NER protein ERCC1 increases point mutation load in liver, but not in small intestinal mouse ASCs

To characterize the mutational consequences of NER-deficiency, we generated clonal organoid cultures from single liver and small intestinal ASCs of three 16-week-old female *Ercc1*$^{-/\Delta}$ mice and three female WT littermates (Fig. 1A). The tissues were harvested at the age of 16 weeks, which is the time point at which *Ercc1*$^{-/\Delta}$ mice generally start to die (Vermeij et al. 2016). WGS analysis of DNA isolated from the clonal organoid cultures allows for

reliable determination of the somatic mutations that were present in the original ASCs and accumulated during life, either in presence or absence of ERCC1 (Blokzijl et al. 2016a) (M Jager, F Blokzijl, V Sasselli, S Boymans, R Janssen, N Besselink, H Clevers, R van Boxtel and E Cuppen, in press). Subclonal mutations acquired after the single-cell-step will only be present in a subpopulation of the cells and are filtered out based on a low allele frequency. We also sequenced the genomes of polyclonal biopsies from the tail of each mouse, which served as control samples to exclude germline variants.

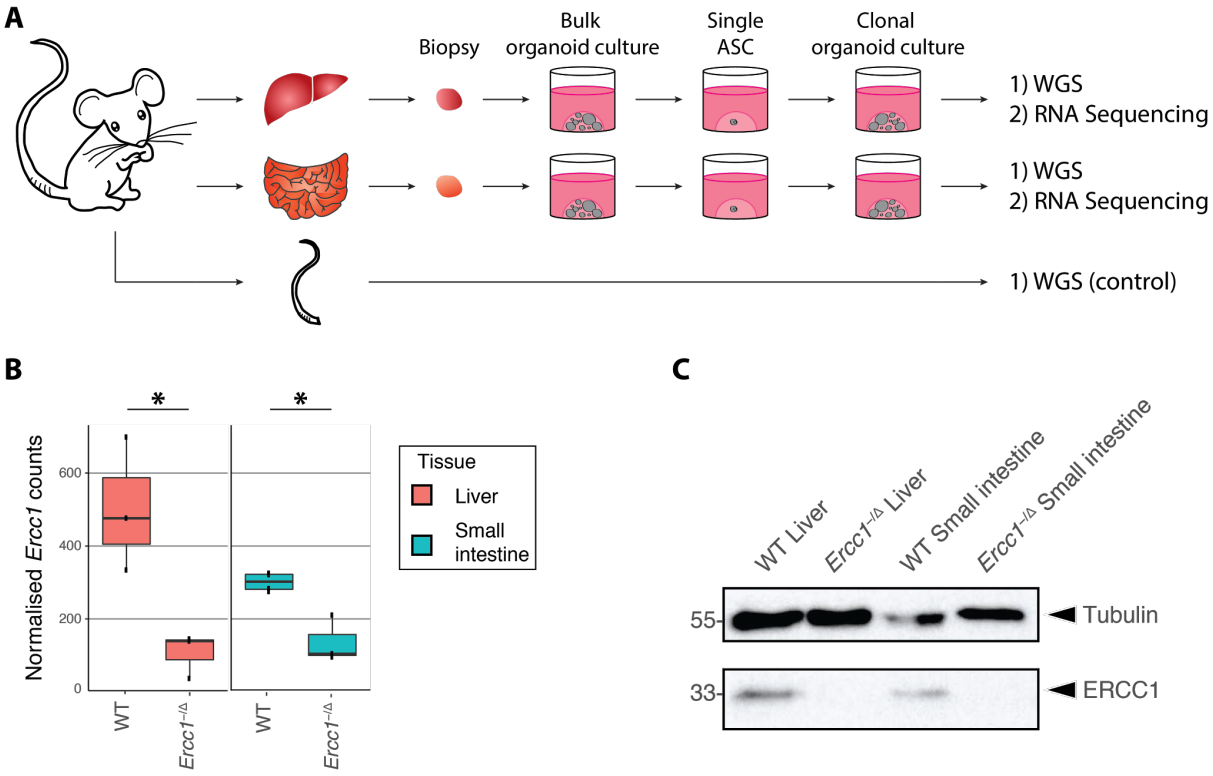


**Figure 1.** Experimental setup and tissue-specific expression of *Ercc1* in mouse ASCs. (A) Schematic overview of the experimental setup used to determine the mutational patterns in single ASCs from the liver and small intestine of mice. Biopsies from the liver and small intestine of six 16-week-old female mice (three *Ercc1*$^{-/\Delta}$ mice and three WT littermates) were cultured in bulk for ~1.5 week to enrich for ASCs. Subsequently, clonal organoids were derived from these bulk organoid cultures and expanded for approximately 1 month, until there were enough cells to perform both WGS and RNA sequencing. As a control sample for filtering germline variants, a biopsy of the tail of each mouse was also subjected to WGS. (B)

Boxplots depicting normalized mRNA counts of *Ercc1* in ASC organoid cultures from liver and small intestine of *Ercc1*$^{-/\Delta}$ mice (n = 3 and n = 3, respectively) and WT littermates (n = 3 and n = 4, respectively). Asterisks represent significant differences ($P < 0.05$, negative binomial test). (C) Western blot analysis of ERCC1 in *Ercc1*$^{-/\Delta}$ and WT small intestinal and liver mouse organoids.

To determine transcriptome profiles, we performed RNA sequencing on one clonal organoid culture from each tissue of each mouse. *Ercc1* is significantly differentially expressed ($P < 0.05$, negative binomial test) between WT and *Ercc1*$^{-/\Delta}$ in both liver and small intestinal ASCs (Fig. 1B), confirming the anticipated effects of the *Ercc1* mutations at the mRNA level. While there is some *Ercc1* expression in *Ercc1*$^{-/\Delta}$ ASCs, the C-terminal domain of ERCC1 is essential in ERCC1-XPF complex formation and disruption of this interaction reduces the stability of ERCC1 protein (Tripsianes et al. 2005; de Laat 1998; Sijbers et al. 1996b). Indeed, ERCC1 protein is not detectable by immunoblotting in *Ercc1*$^{-/\Delta}$ organoid cultures of both tissues (Fig. 1C). No other DNA repair genes were differentially expressed between WT and *Ercc1*$^{-/\Delta}$ ASCs (Supplemental File S1). Notably, the expression of 8 out of 9 core NER genes, including *Ercc1*, is higher in WT liver ASCs than WT small intestinal ASCs (Supplemental Fig. S1, Supplemental Table S1).

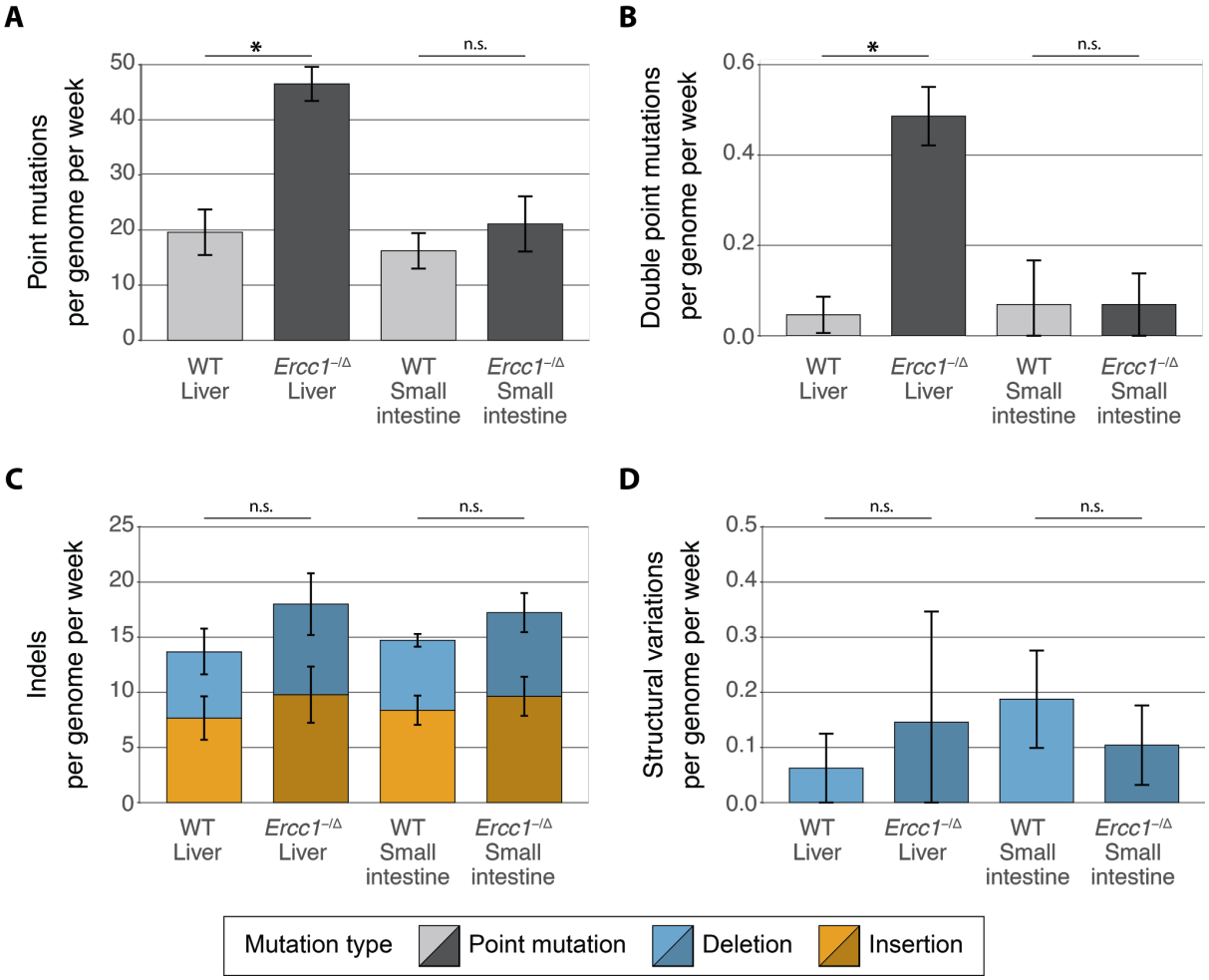


**Figure 2.** Somatic mutation rates in the genomes of ASCs from liver and small intestine of WT and *Ercc1*$^{-/\Delta}$ mice. (A) Point mutations, (B) double point mutations, (C) indels, and (D) SVs acquired per genome per week in ASCs of WT liver (n = 3), *Ercc1*$^{-/\Delta}$ liver (n = 3), WT small intestine (n = 2), and *Ercc1*$^{-/\Delta}$ small intestine (n = 3). Error bars represent standard deviations. Asterisks represent significant differences ($q$ < 0.05, two-sided $t$-test, FDR correction). n.s. : non-significant ($q$ ≥ 0.05, two-sided $t$-test, FDR correction).

Next, we performed WGS analysis on the clonally-expanded organoid cultures (Fig. 1A). We identified 4,238 somatic point mutations in the autosomal genome of 11 clonal ASC samples. Liver ASCs of WT mice acquire 19.6 ± 4.1 (mean ± standard deviation) mutations per week. This rate is similar in ASCs of the small intestine, at 16.2 ± 3.2 mutations per week, and is in line with the observation that human liver and intestinal ASCs have similar

mutation accumulation rates *in vivo* (Blokzijl et al. 2016a). Loss of ERCC1 induces a twofold increase (46.5 ± 3.1 point mutations per week) in the genome-wide number of point mutations in ASCs in the liver (Fig. 2A, Supplemental Fig. S2A). However, we did not observe a different mutation rate in small intestinal ASCs of *Ercc1*$^{-/\Delta}$ mice (21.1 ± 5.0 point mutations per week) compared with WT small intestinal ASCs (Fig. 2A, Supplemental Fig. S2A). We also observe a significant increase in the number of double point mutations in liver ASCs lacking ERCC1 ($q$ < 0.05, $t$-test, FDR correction; Fig. 2B, Supplemental Fig. S2B, Supplemental Table S2). *Ercc1*$^{-/\Delta}$ liver ASCs acquire 0.49 ± 0.06 double point mutations per week, while WT liver ASCs acquire only 0.05 ± 0.04 double point mutations per week. Again, we did not observe this difference between WT and mutant ASCs of the small intestine (0.07 ± 0.10 and 0.07 ± 0.07 per week, respectively). The increased number of double point mutations in the liver ASCs was still significant when we normalized for the total number of point mutations ($q$ < 0.05, $t$-test, FDR correction; Supplemental Fig. S2C), indicating a liver-specific enrichment of double point mutations in *Ercc1*$^{-/\Delta}$ ASCs compared with WT.

In addition to the 4,238 point mutations, we identified 2,543 small insertions and deletions (indels) and 21 larger deletions (≥100 bp) in the autosomal genome of the 11 clonal ASC samples. As opposed to the point mutations, we observed similar indel numbers for WT and *Ercc1*$^{-/\Delta}$ ASCs of both tissues (Fig. 2C, Supplemental Fig. S2D). ASCs in the small intestine and liver of the mice acquire approximately 13.7 - 18.0 indels per week, independent of *Ercc1* mutation status. Likewise, loss of ERCC1 does not influence the number or type of structural variations (SVs) in ASCs of the small intestine and the liver (Fig. 2D, Supplemental Fig. S2E, Supplemental Table S3). Each mouse ASCs carried 0 - 6 deletions (median length of 539 bp) after 16 weeks (Supplemental Table S3). Finally, a genome-wide copy-number profile was generated to identify chromosomal gains and losses. These profiles indicate that all WT and *Ercc1*$^{-/\Delta}$ ASCs were karyotypically stable during life (Supplemental Fig. S3). Nevertheless, some subclonal aneuploidies were detected in a WT

as well as an *Ercc1*$^{-/\Delta}$ liver organoid sample, which are most likely culturing artefacts that occurred *in vitro* after the clonal step irrespective of *Ercc1* mutation status.

**Loss of NER protein ERCC1 induces Signature 8 mutations in mouse ASCs**

To further dissect the mutational consequences of NER-deficiency, we characterized the mutation spectra in the mouse ASCs. Irrespective of tissue-type, the mutation spectra of all assessed ASCs are predominantly characterized by C:G > A:T mutations and C:G > T:A mutations (Fig. 3A, Supplemental Fig. S4). Although the mutation spectra are highly similar (Supplemental Fig. S5), we did observe some notable differences, such as an increased contribution of T:A > A:T mutations in *Ercc1*$^{-/\Delta}$ ASCs compared with WT ASCs (Fig. 3A). Furthermore, two subsequent nucleotides are more frequently substituted by TT or AA in NER-deficient mouse ASCs (Supplemental Table S2). These results suggest that different mutational processes are active in the *Ercc1*$^{-/\Delta}$ ASCs.

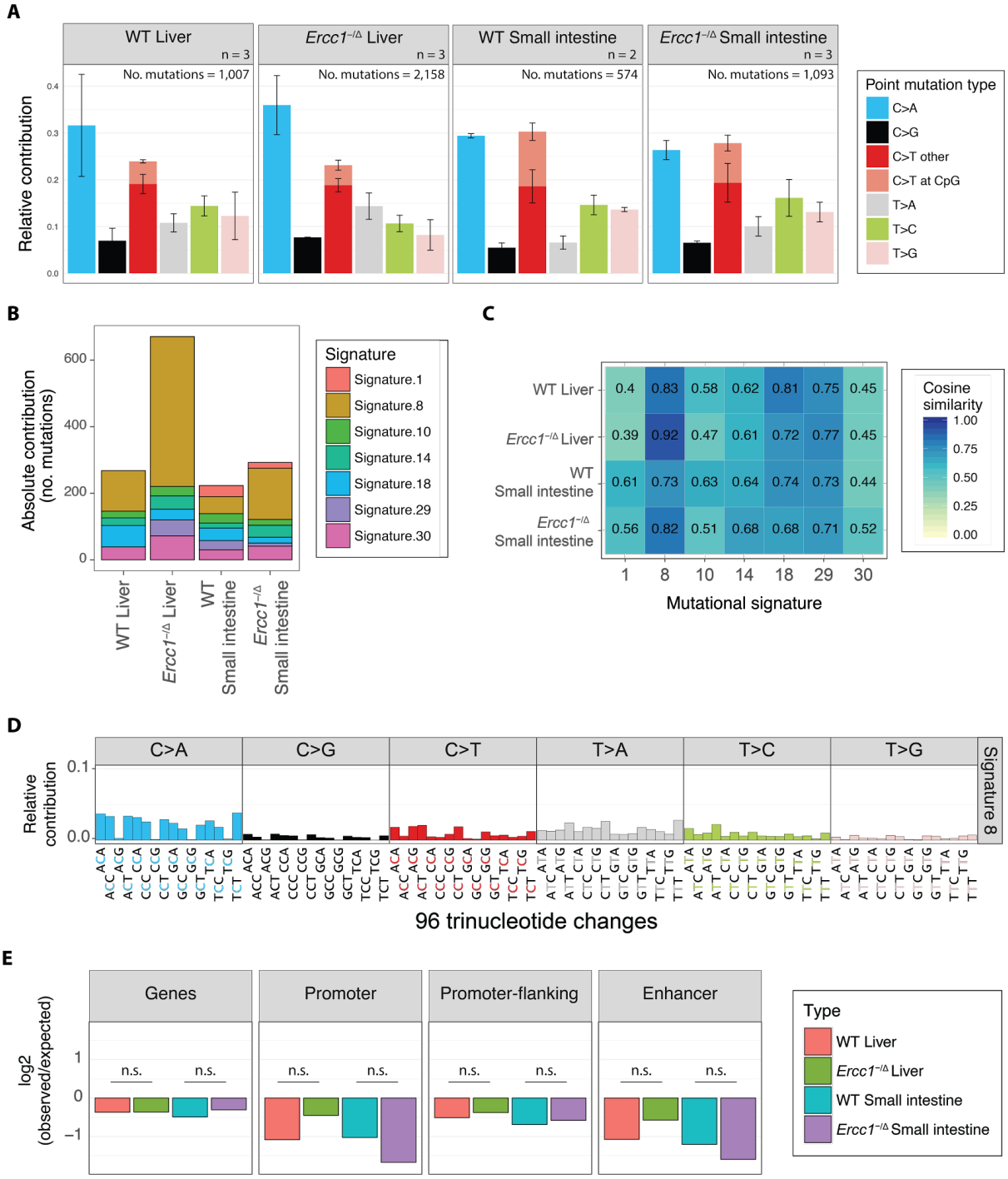


**Figure 3.** Mutational patterns of point mutations acquired in the genomes of ASCs from liver and small intestine of WT and *Ercc1*$^{-/\Delta}$ mice. (A) Mean relative contribution of the indicated mutation types to the mutational spectrum for each mouse ASC group. Error bars represent standard deviations. The total number of mutations, and total number of ASCs (n) per group is indicated. (B) Optimal absolute contribution of the indicated COSMIC mutational signatures per genome to reconstruct the mutational profile of each indicated mouse ASC

group. (C) Similarity of the mutational profiles to the COSMIC mutational signatures depicted in (B). (D) Relative contribution of each indicated context-dependent point mutation type to mutational Signature 8 (http://cancer.sanger.ac.uk/cosmic/signatures). (E) Depletion of somatic point mutations in genes, promoters, promoter-flanking regions, and enhancers for each indicated ASC group. The log2 ratio of the number of observed and expected point mutations indicates the effect size of the depletion in each region.

To determine if any of the known mutational signatures is associated to NER-deficiency, we assessed the contribution of each COSMIC mutational signature (http://cancer.sanger.ac.uk/cosmic/signatures) to the mutational profiles of all ASCs. For this analysis, we first inferred the contribution of each mutational signature and then selected those with a contribution of at least 10% to the mutational profile of at least one of the four ASC groups (*Ercc1*$^{-/\Delta}$ liver, *Ercc1*$^{-/\Delta}$ small intestine, WT liver, WT small intestine; Fig. 3B, Supplemental Fig. S6). The mutational profiles of NER-deficient liver ASCs not only closely resemble Signature 8 (cosine similarity of 0.92; Fig. 3C, Supplemental Fig. S7), but Signature 8 can almost fully explain the increase in point mutations in NER-deficient liver ASCs (Fig. 3B-D). The number of Signature 8 mutations is also increased in all small intestinal ASCs of *Ercc1*$^{-/\Delta}$ mice compared with WT small intestinal ASCs. This result suggests that NER-deficiency can cause Signature 8 mutations in ASCs, regardless of tissue-type.

Mutations are distributed non-randomly throughout the genome in cancer cells and in human ASCs (Blokzijl et al. 2016a; Schuster-Böckler and Lehner 2012). NER is one of the pathways that is suggested to underlie this non-random distribution of mutations (Perera et al. 2016; Zheng et al. 2014). Firstly, NER-activity has been linked to a local enrichment of mutations at gene promoters (Perera et al. 2016). However, we do not observe any significant differences in the depletion of mutations in promoters, promoter-flanking, and enhancer regions between NER-proficient and -deficient ASCs (Fig. 3E, Supplemental Fig. S8A). Secondly, TC-NER results in a depletion of mutations in expressed genes, as this

pathway repairs lesions on the transcribed strand during transcription (Pleasance et al. 2010). Mutations are indeed depleted in genic regions of NER-proficient WT mouse ASCs, but the depletion is not significantly different in NER-deficient ASCs (n.s., Poisson test, FDR correction; Fig. 3E, Supplemental Fig. S8A). Moreover, the average expression levels of genes in which the somatic mutations are located do not differ between *Ercc1*$^{-/\Delta}$ and WT ASCs (n.s., *t*-test, FDR correction; Supplemental Fig. S8B), suggesting that *Ercc1*$^{-/\Delta}$ ASCs do not accumulate more mutations in expressed genes. Finally, there are no obvious changes in transcriptional strand bias, although the mutation numbers are too low to be conclusive (Supplemental Fig. S8C). NER-deficiency thus influences both the mutation load and mutation type, but not the genomic distribution of the observed point mutations in mouse ASCs, suggesting that the contribution of TC-NER in the observed mutational consequences is minimal.

**Loss of GG-NER protein XPC induces Signature 8 mutations in human ASCs**

To confirm the causal relationship between GG-NER-deficiency and Signature 8 and double point mutations, we generated a human *XPC*-knockout (*XPC*$^{KO}$) using CRISPR-Cas9 gene-editing in a small intestinal human organoid culture (Fig. 4A). After confirming absence of XPC protein (Fig. 4B), we passaged the *XPC*$^{KO}$ clones for ± 2 months to allow the ASCs to accumulate sufficient mutations required for downstream analyses. Subsequently, we derived subclonal cultures of single ASCs and expanded these until sufficient DNA could be isolated for WGS. This approach allowed us to catalog the mutations that specifically accumulated between the two clonal expansion steps in the absence of XPC (Supplemental Fig. S9A) (Blokzijl et al. 2016a; Drost et al. 2017). As a control, WGS data of three previously-established *XPC*$^{WT}$ organoid cultures of the same human donor was used.

Similar to the *Ercc1*$^{-/\Delta}$ mouse ASCs, loss of XPC in human ASCs induced an increase in the genome-wide number of point mutations acquired per week (3-fold increase; Fig. 4C, Supplemental Fig. S9B). In addition, the number of double point mutations acquired per week was also increased (12-fold increase; Fig. 4D). Similar as for the *Ercc1*$^{-/\Delta}$ ASCs,

we did not observe a significant change in the genomic distribution of acquired mutations as a result of *XPC* deletion in human ASC, nor a change in transcriptional strand bias (Supplemental Fig. S9C-D). Importantly, ~39% of the increase in point mutations in the $XPC^{KO}$ ASCs can be explained by Signature 8 on average (Fig. 4E, Supplemental Fig. S9B), underscoring that NER-deficiency causes Signature 8, independent of tissue-type or species.

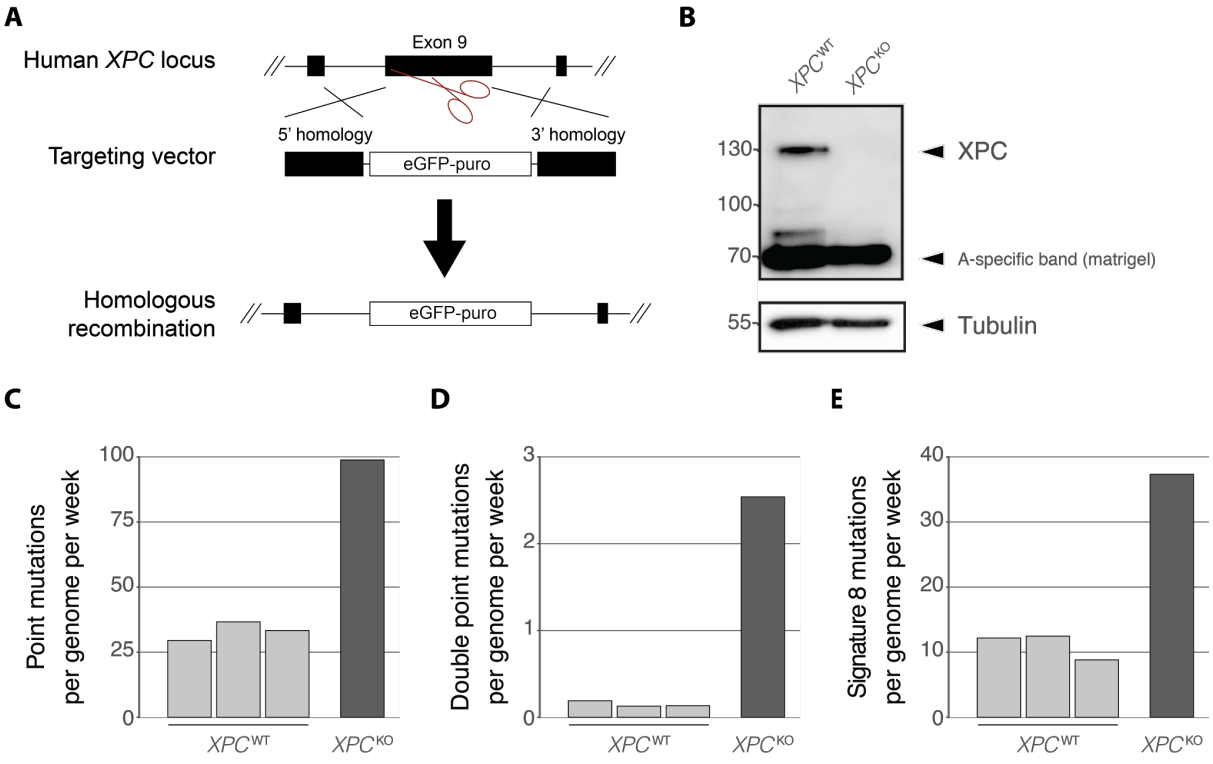


**Figure 4.** Mutational consequences of $XPC^{KO}$ in human intestinal organoid cultures *in vitro*. (A) Targeting strategy for the generation of $XPC^{KO}$ organoid cultures using CRISPR-Cas9 gene-editing. (B) Western blot analysis of XPC in human $XPC^{WT}$ and $XPC^{KO}$ organoids. (C) Number of point mutations, (D) double point mutations, and (E) Signature 8 mutations acquired per genome per week in human $XPC^{WT}$ ASCs (n = 3) and an $XPC^{KO}$ ASC (n = 1) *in vitro*.

## DISCUSSION

We exploited the organoid culturing system, CRISPR-Cas9, WGS, and mutational signature analyses to study the genome-wide mutational consequences of NER-deficiency in individual ASCs of human and mice. Our results show that loss of ERCC1 induces a significant increase in the number of both single and double point mutations in mouse liver ASCs, but not in mouse small intestinal ASCs. Similarly, a LacZ reporter assay showed increased point mutation frequency in livers of *Ercc1*$^{-/\Delta}$ mice (Dollé et al. 2006). Interestingly, this tissue-specific mutational effect coincides with the tissue-specific pathological aging phenotype observed in *Ercc1*$^{-/\Delta}$ mice (Dollé et al. 2011; Gregg et al. 2012). A possible explanation for this tissue-specific effect is that liver ASCs might be more dependent on DNA repair facilitated by ERCC1 compared with small intestinal ASCs, e.g. as a result of tissue-specific mutagen exposure. In line with this, WT liver ASCs show a higher basal expression of *Ercc1* and other NER genes than WT small intestinal ASCs. However, the transcription levels of DNA repair components do not necessarily reflect DNA repair-activity, due to post-transcriptional regulation (Naipal et al. 2015). Alternatively, liver and small intestinal ASCs might use different mechanisms to cope with unrepaired DNA damage as a result of the loss of ERCC1, such as the utilization of alternative DNA repair mechanisms, like translesion synthesis (TLS) polymerases to bypass polymerase-blocking lesions, or differential induction of apoptosis or senescence.

ERCC1 is involved in multiple DNA repair pathways, including TC-NER, GG-NER, SSA, and crosslink repair. Previously, it has been shown that SSA- and crosslink repair-deficiencies result in increased number of indels and SVs in mice, whereas NER-deficiency introduces point mutations (Dollé et al. 2006). Since we only observe an increase in point mutations, NER-deficiency is likely responsible for the mutational consequences of loss of ERCC1 in mouse liver ASCs. If TC-NER-deficiency underlies the mutation accumulation, this would be reflected by an increase in mutations in expressed genes in *Ercc1*$^{-/\Delta}$ mice. However, in both WT and *Ercc1*$^{-/\Delta}$ cells a depletion of mutations is observed in genes, and the expression levels of genes in which the somatic mutations are located do not differ between NER-proficient and -deficient ASCs, suggesting that the observed mutational

consequences of impaired ERCC1 is rather an effect of defective GG-NER. We independently confirmed that GG-NER-deficiency induces an increase in both single and double point mutations, by cataloguing mutations that are acquired in a human small intestinal organoid culture that is deleted for GG-NER component *XPC* by CRISPR-Cas9 gene-editing. More specifically, the majority of the increase in point mutations can be explained by an increase in Signature 8 in both systems, suggesting that perturbations of GG-NER result in the accumulation of Signature 8 in both mouse and human ASCs.

Until now, the etiology of Signature 8 was unknown (http://cancer.sanger.ac.uk/cosmic/signatures). Signature 8 is characterized by C:G > A:T point mutations and is associated with double point mutations (Alexandrov et al. 2013; Nik-Zainal et al. 2016). C:G > A:T mutations have been linked to several processes, including oxidative stress (Kamiya et al. 1995; Degtyareva et al. 2013). Interestingly, NER has been suggested to play a role in the repair of tandem DNA lesions that result from oxidative stress (Bergeron et al. 2010; Cadet et al. 2012). If left unrepaired, these lesions can block regular DNA polymerases, but can be bypassed by error-prone TLS polymerases, resulting in increased mutation incorporation (Cadet et al. 2012). Moreover, it has been shown that oxidative stress results in increased induction of double point mutations in NER-deficient human fibroblasts (Lee 2002). Similarly, UV- and smoking-induced tandem DNA lesions have been linked to NER and double point mutation rates, but these mutagens are unlikely to affect the genomes of ASCs in liver and small intestine (Helleday et al. 2014). Taken together, an increase in double point and Signature 8 mutations could thus reflect mutation incorporation as a result of oxidative damage in NER-deficient cells.

In human urothelial cancer, somatic mutations in *ERCC2*, a key factor in both TC-NER and GG-NER, have been associated with a high contribution of mutational Signature 5 (Kim et al. 2016; Iyama and Wilson 2013). However, we did not observe an increase Signature 5 contribution in ASCs without ERCC1 or XPC. This difference in mutational consequences could reflect various differences between these systems, such as different effects of the mutations on protein function, distinct roles of the proteins, or tumor- and/or

tissue-specific activity of mutagenic damage and/or DNA repair processes. In our study, we deleted specific NER components in an otherwise normal genetic background, providing us with the unique opportunity to characterize the direct mutational consequences of NER-deficiency.

Determination of the NER-capacity of tumors can be important for precision medicine, as it has been shown that tumors with mutations in NER genes (Stubbert et al. 2010; Van Allen et al. 2014; Amable 2016; Zhang et al. 2017), and tumors with low expression of *ERCC1* (Olaussen et al. 2006; Li et al. 2000; Amable 2016) are sensitive to cisplatin treatment. However, translation of these findings into the clinical setting has been challenging, because connecting tumor biopsy mRNA levels and immunohistochemistry measurements to NER-activity remains an unresolved issue (Bowden 2014), and interpreting the effects of mutations in DNA repair genes on NER-capacity is challenging. Rather than looking for the presence of causal events, mutational catalogs can be used as a functional readout of NER-capacity in tumors. Here, we show that GG-NER-deficiency specifically induces Signature 8 and double point mutations in both mouse and human ASCs. Signature 8 has an overall prevalence of 2% in sequenced human tumors, is found in medulloblastoma (Alexandrov et al. 2013) and contributes to the mutational profile of the majority of, but not all, breast cancer tumors (Nik-Zainal et al. 2016; Alexandrov et al. 2013). We propose that an elevated contribution of Signature 8 in combination with double point mutations might serve as a novel biomarker for GG-NER-deficiency and has the potential to guide treatment decision. Additional studies are required to demonstrate the predictive value of these mutations for treatment response.

## METHODS

### Mouse tissue material

*Ercc1*$^{-/\Delta}$ mice were generated and maintained as previously described (Vermeij et al., 2016). Briefly, by crossing *Ercc1*$^{\Delta/+}$ (C57BL6J or FVB background) with *Ercc1*$^{+/-}$ mice (FVB or

C57BL6J background), *Ercc1*$^{-/\Delta}$ mice were generated in a uniform F1 C57BL6J/FVB hybrid background. Wild type F1 littermates were used as controls. Animals were housed in individual ventilated cages under specific pathogen-free conditions in a controlled environment (20–22 °C, 12 h light : 12 h dark cycle). Experiments were performed in accordance with the Principles of Laboratory Animal Care and with the guidelines approved by the Dutch Ethical Committee in full accordance with European legislation.

We used three 16-week old female *Ercc1*$^{-/\Delta}$ mice and three female WT littermates for our experiments. Tails were harvested and stored at -20°C. Livers and small intestines were harvested and kept on ice in Adv+++ medium (Advanced DMEM/F-12 with 1% GlutaMAX, HEPES 10 mM and 1% penicillin/streptomycin) for a few hours until further processing.

**Human tissue material**

Endoscopic biopsies were performed at the University Medical Center Utrecht and the Wilhelmina Children's Hospital. The patients' informed consent was obtained and this study was approved by the ethical committee of University Medical Center Utrecht.

**Generation of clonal *Ercc1*$^{-/\Delta}$ and WT mouse organoid cultures**

Single liver ASCs were isolated from livers as described previously (Kuijk et al. 2016). Liver organoid cultures were initiated by culturing the liver ASCs in BME overlaid with mouse liver culture initiation medium (50% Adv+++ medium, 35% WNT3A conditioned medium (produced in house), 5% NOGGIN conditioned medium (produced in house), 5% RSPOI conditioned medium (produced in house), 1x B27 without retinoic acid, 1x N2, 1x Primocin, 10mM Nicotinamide, 0.625mM N-acetylcysteine, 100ng/ml FGF-10, 10µM ROCKi, 50 ng/ml HGF, 10nM Gastrin, and 50ng/ml hEGF). 1.5 week after culture initiation, clonal organoid liver cultures were generated and expanded according to protocol (M Jager, F Blokzijl, V Sasselli, S Boymans, R Janssen, N Besselink, H Clevers, R van Boxtel and E Cuppen, in press) in mouse liver expansion medium (90% Adv+++ medium, 5% RSPOI conditioned medium (produced in house), 1x B27 without retinoic acid, 1x N2, 1x Primocin, 10mM

Nicotinamide, 0.625mM N-acetylcysteine, 100ng/ml FGF-10, 50 ng/ml HGF, 10nM Gastrin, and 50ng/ml hEGF).

Crypts were isolated from small intestines as described previously (Sato et al. 2009). Small intestinal organoid cultures were initiated by culturing the small intestinal ASCs in matrigel overlaid with mouse small intestine medium (50% WNT3A conditioned medium (produced in house), 30% Adv+++ medium, 10% NOGGIN conditioned medium (produced in house), 10% RSPOI conditioned medium (produced in house), 1x B27, 1x hES Cell Cloning & Recovery Supplement, 1x Primocin, 10µM ROCKi, 1.25mM N-acetylcysteine, and 50ng/ml hEGF). Clonal small intestinal organoid cultures were generated by picking single organoids manually and clonally expanding these organoid cultures according to protocol in mouse small intestine medium (M Jager, F Blokzijl, V Sasselli, S Boymans, R Janssen, N Besselink, H Clevers, R van Boxtel and E Cuppen, in press). Culture expansion failed for the small intestine of mouse WT1.

**Generation of a clonal and subclonal *XPC*$^{KO}$ organoid culture**

Clonal *XPC*$^{KO}$ organoid cultures were generated froma small intestinal bulk organoid culture derived previously (Blokzijl et al. 2016a) using the CRISPR-Cas9 gene-editing technique as described in (Drost et al. 2017). One clonal human *XPC*$^{KO}$ organoid culture was obtained and cultured for 72 days in human small intestinal organoid medium (50% WNT3A conditioned medium (produced in house), 30% Adv+++ medium, 20% RSPOI conditioned medium (produced in house), 1x B27, 1x Primocin, 1.25mM N-acetylcysteine, 0.5µM A83-01, 10µM SB202190, 100ng/ml recombinant Noggin, and 50ng/ml hEGF). Subsequently, a subclonal culture was derived according to protocol (M Jager, F Blokzijl, V Sasselli, S Boymans, R Janssen, N Besselink, H Clevers, R van Boxtel and E Cuppen, in press).

**Western blot**

Protein samples from mouse organoid cultures were collected in Laemmli buffer and measured using the Qubit$^{tm}$ 3.0 Fluorometer (Thermo Fisher Scientific) with the Qubit$^{tm}$

Protein Assay Kit (Thermo Fisher Scientific, Q33211). Protein samples from human organoid cultures were collected in Laemmli buffer and measured using a Lowry protein assay. 30µg of protein per sample was run on a 10% SDS page gel. Subsequently, the proteins were transferred to a nitrocellulose membrane. After transfer, the membrane was blocked for 1 hour using 5% ELK (Campina) at room temperature and subsequently incubated overnight with the primary antibody (ERCC1: Abcam, ab129267; XPC: Cell Signaling Technology; #12701). Secondary antibody was incubated 1 hour at room temperature, and subsequently proteins were visualized using the Amersham ECL Western blotting analysis system (GE Healthcare, RPN2109) and the Amersham Imager 600 system (GE Healthcare).

**RNA sequencing and differential expression analysis of *Ercc1*$^{-/\Delta}$ and WT mouse organoid cultures**

For each mouse (three *Ercc1*$^{-/\Delta}$ mice and three WT littermates), we performed RNA sequencing on one clonal organoid culture from the liver and the small intestine. An additional small intestinal organoid clone was sequenced of mice WT2 and WT3 to increase the amount of replicates for differential expression analysis, as culture expansion failed for the small intestine of WT1. Total RNA was collected in TRIzol and purified from all organoid cultures using the Qiasymphony (Qiagen). RNA libraries for Illumina sequencing were generated from 50 ng of poly-A selected mRNA using the Neoprep (Illumina) and sequenced 2 x 75 bp paired-end to approximately 3300 Million base pairs per sample with the Illumina NextSeq 500 at the Utrecht Sequencing Facility.

RNA sequencing reads were mapped with STAR v.2.4.2a to the mouse reference genome GRCm38. The BAM files were sorted with Sambamba v0.5.8 and reads were counted with HTSeq-count version 0.6.1p1 (default settings) to exons as defined in GRCm38v70.gtf (Ensembl). Non-uniquely mapped reads were not counted. Subsequently, DESeq v1.28.0 was used to normalize counts. DESeq nbinomTest was used to test for differential expression (1) of *Ercc1* between *Ercc1*$^{-/\Delta}$ and WT liver ASCs, (2) of *Ercc1* between *Ercc1*$^{-/\Delta}$ and WT small intestinal ASCs, (3) of 83 other DNA repair genes (Casorelli

et al. 2006) between *Ercc1*$^{-/\Delta}$ and WT liver ASCs, and (4) between *Ercc1*$^{-/\Delta}$ and WT small intestinal ASCs, and (5) of 9 NER genes between the WT liver and WT small intestinal ASCs. Differentially expressed genes with an adjusted P-value $q < 0.05$ (Benjamini-Hochberg FDR multiple-testing correction) were considered significant.

**WGS and read alignment**

DNA was isolated from mouse liver organoid cultures and mouse control (tail) samples using the genomic tip 20-G kit (Qiagen) and from mouse small intestinal organoid samples and the human *XPC*$^{KO}$ sample using the Qiasymphony (Qiagen). DNA libraries for Illumina sequencing were generated from 200 ng genomic DNA using standard protocols (Illumina) and sequenced 2 x 100 bp paired-end to 30X base coverage with the Illumina HiSeq Xten at the Hartwig Medical Foundation. The sequence reads of *XPC*$^{KO}$ were mapped to the GRCh37 human reference genome using using the Burrows–Wheeler Aligner (BWA) v0.7.5a (Li and Durbin 2009), with settings '-t 4 -c 100 -M'. The mapped data of clonal *XPC*$^{WT}$ organoids was previously generated in the study ('donor_id' 6) (Blokzijl et al. 2016a). The sequence reads of the mouse ASCs were mapped to the GRCm38 mouse reference genome using using the Burrows–Wheeler Aligner (BWA) v0.7.5a (Li and Durbin 2009), with settings '-t 4 -c 100 -M'. The WGS data of the tails confirmed that the *Ercc1*$^{-/\Delta}$mice have compound heterozygous mutations in *Ercc1* and the WT littermates do not (Supplemental Fig. S10).

**Callable genome**

The callable genome was defined for all sequenced samples using the GATK CallableLoci tool v3.4.46 (Van der Auwera et al. 2013) with default settings and additional optional parameters 'minBaseQuality 10', 'minMappingQuality 10', 'maxFractionOfReadsWithLowMAPQ 20', and 'minDepth 20'. 'CALLABLE' regions were extracted from every output file. Subsequently, genomic regions that were callable (1) in the mouse organoid clone and the control (tail) sample, and (2) in the human organoid clone,

subclone, and control (blood) were intersected to define a genomic region that is surveyed in all samples that were compared. Approximately 90 ± 1% of the autosomal genome was surveyed in every mouse clone, and 73 - 88% of the autosomal genome was surveyed in each human subclone (Supplemental Table S4).

**Point mutation and indel calling**

For both human and mouse samples, point mutations and indels were multi-sample called with GATK HaplotypeCaller v3.4.46 with default settings and additional options '-stand_call_conf 30 -stand_emit_conf 15' and GATK Queue v3.4.46. For mouse samples the quality of the calls was assessed using GATK VariantFiltration v3.4.46 with options 'QD < 2.0, MQ < 40.0, FS > 60.0, HaplotypeScore > 13.0, MQRankSum < -12.5, ReadPosRankSum < -8.0' for point mutations and 'QD < 2.0, FS > 200.0, ReadPosRankSum < -20.0' for indels, with additional options 'clusterSize 3' and 'clusterWindowSize 35'. For human samples the quality of the calls was assessed using GATK VariantFiltration v3.4.46 with options 'QD < 2.0, MQ < 40.0, FS > 60.0, HaplotypeScore > 13.0, MQRankSum < -12.5, ReadPosRankSum < -8.0, MQ0 >= 4 && ((MQ0 / (1.0 * DP)) > 0.1), DP < 5, QUAL < 30, QUAL >= 30.0 && QUAL < 50.0, SOR > 4.0' for point mutations and 'QD < 2.0, FS > 200.0, ReadPosRankSum < -20.0, MQ0 >= 4 && ((MQ0 / (1.0 * DP)) > 0.1), DP < 5, QUAL < 30.0, QUAL >= 30.0 && QUAL < 50.0, SOR > 10.0' for indels, with additional options 'clusterSize 3' and 'clusterWindowSize 10.

**Point mutation filtering**

To obtain high-quality catalogs of somatic point mutations, we applied a comprehensive filtering procedure. For the mouse samples, we only considered positions on the autosomal genome that were callable (see "Callable genome") in both the organoid and control (tail) sample. We excluded positions at which indels were called, as these positions likely represent false-positive point mutation calls. Furthermore, we only included positions with a

'PASS' flag by GATK VariantFiltration, a GATK phred-scaled quality score ≥ 100, a sample-level genotype quality of 99 in the organoid culture and ≥ 10 in the control (tail) sample, and a coverage of ≥ 20X in the organoid and the tail sample. We subsequently excluded variants with any evidence in another organoid sample or control (tail) sample of the same mouse to remove germline variants. To exclude potentially missed germline events, we also removed positions that have any evidence in the organoid and/or control samples of the other mice. Finally, we excluded positions with a variant allele frequency (VAF) < 0.3 in the organoid sample to exclude mutations that were induced after the clonal step.

For the human samples, we only considered positions on the autosomal genome that were callable (see "Callable genome") in the control (blood) sample, clonal organoid and subclonal organoid culture. We considered mutations with a 'PASS' flag by GATK VariantFiltration and a GATK phred-scaled quality score ≥ 100. For both the clonal and subclonal organoid cultures, all variants with evidence in the control (blood) sample were excluded, to remove germline variants. To exclude potentially missed germline events, we removed positions that are in the Single Nucleotide Polymorphism Database v137.b3730, or in a blacklist with positions that are recurrent in unmatched individuals (BED-file available upon request). Subsequently, for both the clonal and subclonal cultures, all variants with a VAF < 0.3 were excluded. Finally, the resulting somatic point mutation catalogs of the clonal and subclonal cultures were compared and all events unique to the subclonal organoid were considered to be accumulated after the *XPC* deletion, that is: between the two sequential clonal expansion steps.

**Clonality of organoid cultures**

We validated whether the organoid samples were clonal based on the VAF of somatic point mutations, before the final filter step (VAF < 0.3). Each cell acquires its own set of somatic

552 mutations and the reads supporting a mutation will be diluted in the WGS data of non-clonal
553 samples, resulting in a low VAF. After extensive filtering of somatic point mutations, liver
554 organoid samples from WT1, WT2, and *Ercc1*$^{-/\Delta}$2 showed a shift in the VAK-peak away from
555 0.5 and therefore these samples were excluded from further analyses (Supplemental Fig.
556 S11). An additional liver organoid culture from these mice was sequenced and these
557 samples were confirmed to be clonal (Supplemental Fig. S11).
558
559 **Double point mutations**
560 We selected point mutations from the filtered variant call format (VCF) files that were called
561 on consecutive bases in the mouse or human reference genome. The double point
562 mutations were subsequently manually checked in the Integrative Genomics Viewer (IGV) to
563 exclude double point mutations present in the control sample, and/or with many point
564 mutations or indels in the region, as these are (likely) false positives.
565
566 **Indel filtration of *Ercc1*$^{-/\Delta}$ and WT mouse organoid cultures**
567 We only considered positions on the autosomal genome that were callable (see "Callable
568 genome") and had a sequencing depth of ≥ 20X in both the organoid sample and the control
569 (tail) sample. We excluded positions that overlap with a point mutation. Furthermore, we only
570 considered positions with a filter 'PASS' from VariantFiltration, a GATK phred-scaled quality
571 score > 250 and a sample-level genotype quality of 99 in both the organoid sample and the
572 control (tail) sample. We subsequently excluded Indels that are located within 50 base pairs
573 of an indel called in another organoid sample and indels with any evidence in another
574 organoid sample or a control (tail) sample. Finally, we excluded positions with a VAF < 0.3 in
575 the organoid sample.
576
577 **SV calling and filtration of *Ercc1*$^{-/\Delta}$ and WT mouse organoid cultures**

578 SVs were called with DELLY v0.7.2 with settings 'type DEL DUP INV TRA INS', 'map-qual
579 1', 'mad-cutoff 9', 'min-flank 13', and 'geno-qual 5' (Rausch et al. 2012). We only considered
580 SVs of at least 100 bp on the autosomal chromosomes that were called with a filter 'PASS',
581 and a sample-specific genotype quality of at least 90 in the organoid culture and the control
582 sample. We subsequently excluded positions with any evidence in the control (tail) sample.
583 The filtered SVs were finally checked manually in IGV to reduce false-positives and we
584 excluded SVs present in the tail sample, with no visible change in the read-depth (for
585 duplications and deletions), and/or with many point mutations in the region.
586
587 **Genome-wide copy number profiles of *Ercc1*$^{-/\Delta}$ and WT mouse organoid cultures**
588 To generate a virtual karyotype, genome-wide copy number states were determined using
589 FreeC v7.2 with settings 'ploidy 2', 'window 1000' and 'telocentromeric 50000' (Boeva et al.
590 2012). Subsequently, the average copy number across bins of 500,000 bp was calculated
591 and plotted to assess genome stability.
592
593 **Quantification of the contribution of COSMIC mutational signatures to the mutational**
594 **profiles**
595 We retrieved the point mutation types from all the filtered VCF files, converted them to the 6
596 types of point mutations that are distinguished by convention and generated a mutational
597 spectrum for four ASC groups (*Ercc1*$^{-/\Delta}$ liver, *Ercc1*$^{-/\Delta}$ small intestine, WT liver, and WT small
598 intestine), as well as *XPC*$^{KO}$, *XPC*$^{WT}$1, *XPC*$^{WT}$2, and *XPC*$^{WT}$3 ASCs. We retrieved the
599 sequence context for all point mutations to generate the mutational profiles with 96 point
600 mutation types. Subsequently, we estimated the contribution of the 30 COSMIC mutational
601 signatures (http://cancer.sanger.ac.uk/cosmic/signatures) to the retrieved mutational profiles.
602 This function determines the non-negative linear combination of signatures that best
603 reconstructs each mutational profile, by minimizing the Euclidean norm of the residual. Next,
604 we selected signatures with a contribution of at least 10% to the mutational profile of at least
605 one of the four ASC groups (Supplemental Fig. S6), and re-estimated the contribution of

606 these 7 signatures to the mutational profiles of the four ASC groups (Fig. 3B). Furthermore,
607 we calculated the cosine similarity of the 7 signatures to the mutational profiles of each of
608 the four ASC groups (Fig 3C). All mutational pattern analyses were performed with
609 MutationalPatterns (Blokzijl et al. 2016b).
610
611 **Enrichment or depletion of point mutations in genomic regions**
612 To test whether the point mutations appear more or less frequently than expected in genes,
613 promoters, promoter-flanking, and enhancer regions, we loaded the UCSC Known Genes
614 tables as TxDb objects for Mm10 (Team BC and Maintainer 2016) and Hg19 (Carlson and
615 Maintainer 2015), and the regulatory features for Mm10 and Hg19 from Ensembl using
616 biomaRt (Durinck et al. 2005, 2009). We tested for enrichment or depletion of point
617 mutations in the genomic regions per ASC group (*Ercc1*$^{-/\Delta}$ liver, *Ercc1*$^{-/\Delta}$ small intestine, WT
618 liver, WT small intestine, *XPC*$^{KO}$ and *XPC*$^{WT}$) using a one-sided Binomial test with
619 MutationalPatterns (Blokzijl et al. 2016b), which corrects for the surveyed genomic areas
620 (Supplemental Fig. S8A, Supplemental Fig. S9C). Two-sided Poisson tests were performed
621 to test for significant differences in the ratio of point mutations within a genomic region
622 divided by the total number of point mutations between (1) mouse WT and *Ercc1*$^{-/\Delta}$ liver
623 ASCs, (2) mouse WT and *Ercc1*$^{-/\Delta}$ small intestinal ASCs, and (3) human *XPC*$^{KO}$ and human
624 *XPC*$^{WT}$ ASCs (Fig. 3E, Supplemental Fig. S8A, Supplemental Fig. S9C). Within species,
625 differences in mutation rates with an adjusted P-value $q < 0.05$ (Benjamini-Hochberg FDR
626 multiple-testing correction) were considered significant.
627 To test whether point mutations occur more frequently in more highly expressed
628 genes in the NER-deficient mouse ASCs, we first selected point mutations that occurred
629 within genes in the mouse ASCs. Per ASC group, we next determined the average Reads
630 Per Kilobase per Million mapped reads (RPKM) of these genes. Two-sided *t*-tests were
631 performed to test for significant difference in the average expression of genes that carry a
632 somatic mutation between (1) mouse WT and *Ercc1*$^{-/\Delta}$ liver ASCs, and (2) mouse WT and
633 *Ercc1*$^{-/\Delta}$ small intestinal ASCs (Supplemental Fig. S8B). Differences in gene expression

634 distributions with an adjusted P-value $q < 0.05$ (Benjamini-Hochberg FDR multiple-testing
635 correction) were considered significant.
636
637 **Transcriptional strand bias of point mutations**
638 For the point mutations within genes we determined whether the mutations are located on
639 the transcribed or the non-transcribed strand. To this end, we determined whether the
640 mutated "C" or "T" base is on the same strand as the gene definition, which is untranscribed,
641 or the opposite strand, which is transcribed. We generated a mutational profile per ASC
642 group with the relative contribution of each mutation type with separate bars for the
643 mutations on the transcribed and untranscribed strand, and calculated the significance of the
644 strand bias using a two-sided Poisson test with MutationalPatterns (Supplemental Fig. S8C,
645 Supplemental Fig. S9D) (Blokzijl et al. 2016b). Furthermore, we performed two-sided
646 Poisson tests to test whether there is a significant difference in strand bias per mutation type
647 between (1) mouse WT and *Ercc1*$^{-/\Delta}$ liver ASCs, (2) mouse WT and *Ercc1*$^{-/\Delta}$ small intestinal
648 ASCs, and (3) human *XPC*$^{KO}$ and human *XPC*$^{WT}$ ASCs (Supplemental Fig. S8C,
649 Supplemental Fig. S9D). Within species, differences in strand bias with an adjusted P-value
650 $q < 0.05$ (Benjamini-Hochberg FDR multiple-testing correction) were considered significant.
651
652 **Calculation and comparison of mutation rates**
653 To calculate the mutation rates per genome per week, we quantified the number of somatic
654 point mutations, double nucleotide mutations, indels, and SVs for each mouse ASC.
655 Moreover, we quantified the number of point mutations, double nucleotide mutations, and
656 Signature 8 mutations for the human ASCs. All event counts were extrapolated to the entire
657 autosomal genome using the callable genome length (see" Callable genome") for both
658 mouse and human ASCs to correct for differences in the surveyed genome (Supplemental
659 Table S4). Subsequently, the mutation rates were calculated by dividing the extrapolated
660 number of mutations by the number of weeks in which the mutations were accumulated (WT
661 and *Ercc1*$^{-/\Delta}$ mouse organoids: 16 weeks; *XPC*$^{WT}$ human organoids: 20.6 weeks; *XPC*$^{KO}$

human organoids 10.3 weeks). Two-tailed *t*-tests were performed to determine whether the mutation rates differ significantly between (1) mouse WT and *Ercc1*$^{-/\Delta}$ liver ASCs, and (2) mouse WT and *Ercc1*$^{-/\Delta}$ small intestinal ASCs. Differences in mutation rates between *Ercc1*$^{-/\Delta}$ and WT mouse ASCs with an adjusted P-value $q < 0.05$ (Benjamini-Hochberg FDR multiple-testing correction) were considered significant. To determine the proportion of additionally accumulated mutations in the *XPC*$^{KO}$ culture that can be attributed to Signature 8 in human ASCs, we first calculated the increase in point mutations and the increase in Signature 8 mutations of *XPC*$^{KO}$ compared to *XPC*$^{WT}$1, *XPC*$^{WT}$2, and *XPC*$^{WT}$3 separately. We then divided the increase in Signature 8 mutations by the total increase in point mutations.

## DATA ACCESS

The sequencing data of the mouse samples have been deposited at the European Nucleotide Archive under accession number ERP021379. The sequencing data of the human samples have been deposited at the European Genome-Phenome archive under accession numbers EGAS00001001682 and EGAS00001002681.

## ACKNOWLEDGEMENTS

The authors would like to thank the the animal caretakers of the Erasmus MC for taking care of the mice and the Utrecht Sequencing Facility for providing the sequencing service and data. Utrecht Sequencing Facility is subsidized by the University Medical Center Utrecht, Hubrecht Institute and Utrecht University. This study was financially supported by the NWO Zwaartekracht program Cancer Genomics.nl.

## AUTHOR CONTRIBUTIONS

M.J., E.K., M.V., N.B., and R.B. performed organoid culture. N.B. and R.B. generated western blots and sequenced the organoid cultures. M.J., F.B., R.J., S.B., J.L., and R.B. performed bioinformatic analyses. M.J., F.B., E.K., J.H., J.P., R.B., and E.C. were involved in the conceptual design of this study. M.J., F.B., R.B., and E.C. wrote the manuscript.

690

691 **DISCLOSURE DECLARATION**

692 The authors have nothing to disclose.

693 **REFERENCES**

694 Aboussekhra A, Biggerstaff M, Shivji MK, Vilpo JA, Moncollin V, Podust VN, Protić M,
695 Hübscher U, Egly JM, Wood RD. 1995. Mammalian DNA nucleotide excision repair
696 reconstituted with purified protein components. *Cell* **80**: 859–868.

697 Adams PD, Jasper H, Lenhard Rudolph K. 2015. Aging-Induced Stem Cell Mutations as
698 Drivers for Disease and Cancer. *Cell Stem Cell* **16**: 601–612.

699 Alexandrov LB, Ju YS, Haase K, Van Loo P, Martincorena I, Nik-Zainal S, Totoki Y, Fujimoto
700 A, Nakagawa H, Shibata T, et al. 2016. Mutational signatures associated with tobacco
701 smoking in human cancer. *Science* **354**: 618–622.

702 Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SAJR, Behjati S, Biankin AV, Bignell GR,
703 Bolli N, Borg A, Børresen-Dale A-L, et al. 2013. Signatures of mutational processes in
704 human cancer. *Nature* **500**: 415–421.

705 Al-Minawi AZ, Saleh-Gohari N, Helleday T. 2008. The ERCC1/XPF endonuclease is
706 required for efficient single-strand annealing and gene conversion in mammalian cells.
707 *Nucleic Acids Res* **36**: 1–9.

708 Amable L. 2016. Cisplatin resistance and opportunities for precision medicine. *Pharmacol*
709 *Res* **106**: 27–36.

710 Barker N, Ridgway RA, van Es JH, van de Wetering M, Begthel H, van den Born M,
711 Danenberg E, Clarke AR, Sansom OJ, Clevers H. 2009. Crypt stem cells as the cells-of-
712 origin of intestinal cancer. *Nature* **457**: 608–611.

713 Bergeron F, Auvré F, Radicella JP, Ravanat J-L. 2010. HO* radicals induce an unexpected
714 high proportion of tandem base lesions refractory to repair by DNA glycosylases. *Proc*
715 *Natl Acad Sci U S A* **107**: 5528–5533.

716 Blokzijl F, de Ligt J, Jager M, Sasselli V, Roerink S, Sasaki N, Huch M, Boymans S, Kuijk E,
717 Prins P, et al. 2016a. Tissue-specific mutation accumulation in human adult stem cells
718 during life. *Nature* **538**: 260–264.

719 Blokzijl F, Janssen R, Van Boxtel R, Cuppen E. 2016b. MutationalPatterns: an integrative R
720 package for studying patterns in base substitution catalogues.
721 http://dx.doi.org/10.1101/071761.

722 Boeva V, Popova T, Bleakley K, Chiche P, Cappo J, Schleiermacher G, Janoueix-Lerosey I,
723 Delattre O, Barillot E. 2012. Control-FREEC: a tool for assessing copy number and
724 allelic content using next-generation sequencing data. *Bioinformatics* **28**: 423–425.

725 Bowden NA. 2014. Nucleotide excision repair: Why is it not used to predict response to
726 platinum-based chemotherapy? *Cancer Lett* **346**: 163–171.

727 Cadet J, Ravanat J-L, TavernaPorro M, Menoni H, Angelov D. 2012. Oxidatively generated
728 complex DNA damage: tandem and clustered lesions. *Cancer Lett* **327**: 5–15.

729 Carlson M, Maintainer BP. 2015. *TxDb.Hsapiens.UCSC.hg19.knownGene: Annotation*
730 *package for TxDb object(s)*.

731 Casorelli I, Tenedini E, Tagliafico E, Blasi MF, Giuliani A, Crescenzi M, Pelosi E, Testa U,
732 Peschle C, Mele L, et al. 2006. Identification of a molecular signature for leukemic
733 promyelocytes and their normal counterparts: Focus on DNA repair genes. *Leukemia*

**20**: 1978–1988.

Davies H, Glodzik D, Morganella S, Yates LR, Staaf J, Zou X, Ramakrishna M, Martin S, Boyault S, Sieuwerts AM, et al. 2017. HRDetect is a predictor of BRCA1 and BRCA2 deficiency based on mutational signatures. *Nat Med* **23**: 517–525.

Degtyareva NP, Heyburn L, Sterling J, Resnick MA, Gordenin DA, Doetsch PW. 2013. Oxidative stress-induced mutagenesis in single-strand DNA occurs primarily at cytosines and is DNA polymerase zeta-dependent only for adenines and guanines. *Nucleic Acids Res* **41**: 8995–9005.

de Laat W. 1998. Mapping of interaction domains between human repair proteins ERCC1 and XPF. *Nucleic Acids Res* **26**: 4146–4152.

Dollé MET, Busuttil RA, Garcia AM, Wijnhoven S, van Drunen E, Niedernhofer LJ, van der Horst G, Hoeijmakers JHJ, van Steeg H, Vijg J. 2006. Increased genomic instability is not a prerequisite for shortened lifespan in DNA repair deficient mice. *Mutat Res* **596**: 22–35.

Dollé MT, Kuiper R, Roodbergen M, Robinson J, de Vlugt S, Wijnhoven SP, Beems RB, de la Fonteyne L, de With P, van der Pluijm I, et al. 2011. Broad segmental progeroid changes in short-lived Ercc1 −/Δ7 mice. *Pathobiology of Aging & Age-related Diseases* **1**: 7219.

Drost J, van Boxtel R, Blokzijl F, Mizutani T, Sasaki N, Sasselli V, de Ligt J, Behjati S, Grolleman JE, van Wezel T, et al. 2017. Use of CRISPR-modified human stem cell organoids to study the origin of mutational signatures in cancer. *Science*. http://dx.doi.org/10.1126/science.aao3130.

Durinck S, Moreau Y, Kasprzyk A, Davis S, De Moor B, Brazma A, Huber W. 2005. BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis. *Bioinformatics* **21**: 3439–3440.

Durinck S, Spellman PT, Birney E, Huber W. 2009. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nat Protoc* **4**: 1184–1191.

Gregg SQ, Gutiérrez V, Robinson AR, Woodell T, Nakao A, Ross MA, Michalopoulos GK, Rigatti L, Rothermel CE, Kamileri I, et al. 2012. A mouse model of accelerated liver aging caused by a defect in DNA repair. *Hepatology* **55**: 609–621.

Helleday T, Eshtad S, Nik-Zainal S. 2014. Mechanisms underlying mutational signatures in human cancers. *Nat Rev Genet* **15**: 585–598.

Hoeijmakers JHJ. 2009. DNA damage, aging, and cancer. *N Engl J Med* **361**: 1475–1485.

Iyama T, Wilson DM 3rd. 2013. DNA repair mechanisms in dividing and non-dividing cells. *DNA Repair* **12**: 620–636.

Jaspers NGJ, Raams A, Silengo MC, Wijgers N, Niedernhofer LJ, Robinson AR, Giglia-Mari G, Hoogstraten D, Kleijer WJ, Hoeijmakers JHJ, et al. 2007. First reported patient with human ERCC1 deficiency has cerebro-oculo-facio-skeletal syndrome with a mild defect in nucleotide excision repair and severe developmental failure. *Am J Hum Genet* **80**: 457–466.

Kamiya H, Murata-Kamiya N, Koizume S, Inoue H, Nishimura S, Ohtsuka E. 1995. 8-Hydroxyguanine (7,8-dihydro-8-oxoguanine) in hot spots of the c-Ha-ras gene: effects of

776 sequence contexts on mutation spectra. *Carcinogenesis* **16**: 883–889.

777 Kashiyama K, Nakazawa Y, Pilz DT, Guo C, Shimada M, Sasaki K, Fawcett H, Wing JF,
778 Lewin SO, Carr L, et al. 2013. Malfunction of nuclease ERCC1-XPF results in diverse
779 clinical manifestations and causes Cockayne syndrome, xeroderma pigmentosum, and
780 Fanconi anemia. *Am J Hum Genet* **92**: 807–819.

781 Kim J, Mouw KW, Polak P, Braunstein LZ, Kamburov A, Kwiatkowski DJ, Rosenberg JE,
782 Van Allen EM, D'Andrea A, Getz G. 2016. Somatic ERCC2 mutations are associated
783 with a distinct genomic signature in urothelial tumors. *Nat Genet* **48**: 600–606.

784 Kirschner K, Melton DW. 2010. Multiple roles of the ERCC1-XPF endonuclease in DNA
785 repair and resistance to anticancer drugs. *Anticancer Res* **30**: 3223–3232.

786 Kuijk EW, Rasmussen S, Blokzijl F, Huch M, Gehart H, Toonen P, Begthel H, Clevers H,
787 Geurts AM, Cuppen E. 2016. Generation and characterization of rat liver stem cell lines
788 and their engraftment in a rat model of liver failure. *Sci Rep* **6**.
789 http://dx.doi.org/10.1038/srep22154.

790 Lee D-H. 2002. Oxidative DNA damage induced by copper and hydrogen peroxide promotes
791 CG->TT tandem mutations at methylated CpG dinucleotides in nucleotide excision
792 repair-deficient cells. *Nucleic Acids Res* **30**: 3566–3573.

793 Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler
794 transform. *Bioinformatics* **25**: 1754–1760.

795 Li Q, Yu JJ, Mu C, Yunmbam MK, Slavsky D, Cross CL, Bostick-Bruton F, Reed E. 2000.
796 Association between the level of ERCC-1 expression and the repair of cisplatin-induced
797 DNA damage in human ovarian cancer cells. *Anticancer Res* **20**: 645–652.

798 López-Otín C, Blasco MA, Partridge L, Serrano M, Kroemer G. 2013. The Hallmarks of
799 Aging. *Cell* **153**: 1194–1217.

800 Marteijn JA, Lans H, Vermeulen W, Hoeijmakers JHJ. 2014. Understanding nucleotide
801 excision repair and its roles in cancer and ageing. *Nat Rev Mol Cell Biol* **15**: 465–481.

802 Naipal KAT, Raams A, Bruens ST, Brandsma I, Verkaik NS, Jaspers NGJ, Hoeijmakers JHJ,
803 van Leenders GJLH, Pothof J, Kanaar R, et al. 2015. Attenuated XPC Expression Is Not
804 Associated with Impaired DNA Repair in Bladder Cancer. *PLoS One* **10**: e0126029.

805 Niedernhofer LJ, Garinis GA, Raams A, Lalai AS, Robinson AR, Appeldoorn E, Odijk H,
806 Oostendorp R, Ahmad A, van Leeuwen W, et al. 2006. A new progeroid syndrome
807 reveals that genotoxic stress suppresses the somatotroph axis. *Nature* **444**: 1038–1043.

808 Nik-Zainal S, Davies H, Staaf J, Ramakrishna M, Glodzik D, Zou X, Martincorena I,
809 Alexandrov LB, Martin S, Wedge DC, et al. 2016. Landscape of somatic mutations in
810 560 breast cancer whole genome sequences. *Nature* **534**: 47.

811 Ni M, Zhang W-Z, Qiu J-R, Liu F, Li M, Zhang Y-J, Liu Q, Bai J. 2014. Association of ERCC1
812 and ERCC2 polymorphisms with colorectal cancer risk in a Chinese population. *Sci Rep*
813 **4**. http://dx.doi.org/10.1038/srep04112.

814 Olaussen KA, Dunant A, Fouret P, Brambilla E, André F, Haddad V, Taranchon E, Filipits M,
815 Pirker R, Popper HH, et al. 2006. DNA repair by ERCC1 in non-small-cell lung cancer
816 and cisplatin-based adjuvant chemotherapy. *N Engl J Med* **355**: 983–991.

817 Perera D, Poulos RC, Shah A, Beck D, Pimanda JE, Wong JWH. 2016. Differential DNA

repair underlies mutation hotspots at active promoters in cancer genomes. *Nature* **532**: 259–263.

Pleasance ED, Cheetham RK, Stephens PJ, McBride DJ, Humphray SJ, Greenman CD, Varela I, Lin M-L, Ordóñez GR, Bignell GR, et al. 2010. A comprehensive catalogue of somatic mutations from a human cancer genome. *Nature* **463**: 191–196.

Rahn JJ, Adair GM, Nairn RS. 2010. Multiple roles of ERCC1-XPF in mammalian interstrand crosslink repair. *Environ Mol Mutagen* **51**: 567–581.

Rausch T, Zichner T, Schlattl A, Stütz AM, Benes V, Korbel JO. 2012. DELLY: structural variant discovery by integrated paired-end and split-read analysis. *Bioinformatics* **28**: i333–i339.

Rossi DJ, Jamieson CHM, Weissman IL. 2008. Stems cells and the pathways to aging and cancer. *Cell* **132**: 681–696.

Sato T, Vries RG, Snippert HJ, van de Wetering M, Barker N, Stange DE, van Es JH, Abo A, Kujala P, Peters PJ, et al. 2009. Single Lgr5 stem cells build crypt□villus structures in vitro without a mesenchymal niche. *Nature* **459**: 262–265.

Schuster-Böckler B, Lehner B. 2012. Chromatin organization is a major influence on regional mutation rates in human cancer cells. *Nature* **488**: 504–507.

Sijbers AM, de Laat WL, Ariza RR, Biggerstaff M, Wei Y-F, Moggs JG, Carter KC, Shell BK, Evans E, de Jong MC, et al. 1996a. Xeroderma Pigmentosum Group F Caused by a Defect in a Structure-Specific DNA Repair Endonuclease. *Cell* **86**: 811–822.

Sijbers AM, van der Spek PJ, Odijk H, van den Berg J, van Duin M, Westerveld A, Jaspers NG, Bootsma D, Hoeijmakers JH. 1996b. Mutational analysis of the human nucleotide excision repair gene ERCC1. *Nucleic Acids Res* **24**: 3370–3380.

Stubbert LJ, Smith JM, McKay BC. 2010. Decreased transcription-coupled nucleotide excision repair capacity is associated with increased p53- and MLH1-independent apoptosis in response to cisplatin. *BMC Cancer* **10**: 207.

Su Y, Orelli B, Madireddy A, Niedernhofer LJ, Scharer OD. 2012. Multiple DNA Binding Domains Mediate the Function of the ERCC1-XPF Protein in Nucleotide Excision Repair. *J Biol Chem* **287**: 21846–21855.

Team BC, Maintainer BP. 2016. *TxDb.Mmusculus.UCSC.mm10.knownGene: Annotation package for TxDb object(s)*.

Tripsianes K, Folkers G, Ab E, Das D, Odijk H, Jaspers NGJ, Hoeijmakers JHJ, Kaptein R, Boelens R. 2005. The structure of the human ERCC1/XPF interaction domains reveals a complementary role for the two proteins in nucleotide excision repair. *Structure* **13**: 1849–1858.

Van Allen EM, Mouw KW, Kim P, Iyer G, Wagle N, Al-Ahmadie H, Zhu C, Ostrovnaya I, Kryukov GV, O'Connor KW, et al. 2014. Somatic ERCC2 mutations correlate with cisplatin sensitivity in muscle-invasive urothelial carcinoma. *Cancer Discov* **4**: 1140–1153.

Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, Del Angel G, Levy-Moonshine A, Jordan T, Shakir K, Roazen D, Thibault J, et al. 2013. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr*

*Protoc Bioinformatics* **43**: 11.10.1–33.

Vermeij WP, Dollé MET, Reiling E, Jaarsma D, Payan-Gomez C, Bombardieri CR, Wu H, Roks AJM, Botter SM, van der Eerden BC, et al. 2016. Restricted diet delays accelerated ageing and genomic stress in DNA-repair-deficient mice. *Nature* **537**: 427–431.

Waddell N, Pajic M, Patch A-M, Chang DK, Kassahn KS, Bailey P, Johns AL, Miller D, Nones K, Quek K, et al. 2015. Whole genomes redefine the mutational landscape of pancreatic cancer. *Nature* **518**: 495–501.

Weeda G, Donker I, de Wit J, Morreau H, Janssens R, Vissers CJ, Nigg A, van Steeg H, Bootsma D, Hoeijmakers JHJ. 1997. Disruption of mouse ERCC1 results in a novel repair syndrome with growth failure, nuclear abnormalities and senescence. *Curr Biol* **7**: 427–439.

Zhang R, Jia M, Xue H, Xu Y, Wang M, Zhu M, Sun M, Chang J, Wei Q. 2017. Genetic variants in ERCC1 and XPC predict survival outcome of non-small cell lung cancer patients treated with platinum-based therapy. *Sci Rep* **7**: 10702.

Zheng CL, Wang NJ, Chung J, Moslehi H, Sanborn JZ, Hur JS, Collisson EA, Vemula SS, Naujokas A, Chiotti KE, et al. 2014. Transcription restores DNA repair to heterochromatin, determining regional mutation rates in cancer genomes. *Cell Rep* **9**: 1228–1234.

Zhu L, Finkelstein D, Gao C, Shi L, Wang Y, López-Terrada D, Wang K, Utley S, Pounds S, Neale G, et al. 2016. Multi-organ Mapping of Cancer Risk. *Cell* **166**: 1132–1146.e7.