Article

Alcohol-derived DNA crosslinks are repaired by two distinct mechanisms

https://doi.org/10.1038/s41586-020-2059-5

Received: 17 August 2018

Accepted: 31 January 2020

Published online: 4 March 2020

Check for updates

Michael R. Hodskinson[1,8], Alice Bolner[2,8], Koichi Sato[2], Ashley N. Kamimae-Lanning[1], Koos Rooijers[2], Merlijn Witte[2], Mohan Mahesh[1,5], Jan Silhan[1,6], Maya Petek[1,7], David M. Williams[3], Jop Kind[2], Jason W. Chin[1], Ketan J. Patel[1,4]✉ & Puck Knipscheer[2]✉

Acetaldehyde is a highly reactive, DNA-damaging metabolite that is produced upon alcohol consumption[1]. Impaired detoxification of acetaldehyde is common in the Asian population, and is associated with alcohol-related cancers[1,2]. Cells are protected against acetaldehyde-induced damage by DNA crosslink repair, which when impaired causes Fanconi anaemia (FA), a disease resulting in failure to produce blood cells and a predisposition to cancer[3,4]. The combined inactivation of acetaldehyde detoxification and the FA pathway induces mutation, accelerates malignancies and causes the rapid attrition of blood stem cells[5–7]. However, the nature of the DNA damage induced by acetaldehyde and how this is repaired remains a key question. Here we generate acetaldehyde-induced DNA interstrand crosslinks and determine their repair mechanism in *Xenopus* egg extracts. We find that two replication-coupled pathways repair these lesions. The first is the FA pathway, which operates using excision—analogous to the mechanism used to repair the interstrand crosslinks caused by the chemotherapeutic agent cisplatin. However, the repair of acetaldehyde-induced crosslinks results in increased mutation frequency and an altered mutational spectrum compared with the repair of cisplatin-induced crosslinks. The second repair mechanism requires replication fork convergence, but does not involve DNA incisions—instead the acetaldehyde crosslink itself is broken. The Y-family DNA polymerase REV1 completes repair of the crosslink, culminating in a distinct mutational spectrum. These results define the repair pathways of DNA interstrand crosslinks caused by an endogenous and alcohol-derived metabolite, and identify an excision-independent mechanism.

To study the repair of alcohol-induced DNA damage, we generated an acetaldehyde-crosslinked DNA substrate. Acetaldehyde reacts with guanine to create a crosslink precursor, N^2-propanoguanine (PdG)[8] (Fig. 1a). In a 5′-CpG sequence, PdG reacts with the N^2-amine of guanine on the opposite strand to create an acetaldehyde interstrand crosslink (AA-ICL); this crosslink exists in equilibrium between three states[9]. We synthesized a site-specific native AA-ICL, denoted AA_{NAT}-ICL, within an oligonucleotide duplex (Extended Data Fig. 1a, b, d, Supplementary Fig. 1). A control reaction of PdG with deoxyinosine—which lacks an N^2-amine—did not crosslink, thus confirming the site-specificity of the AA_{NAT}-ICL (Extended Data Fig. 1c, for gel source data see Supplementary Fig. 2). AA_{NAT}-ICLs were stable at physiological pH and temperature, with less than 10% reversal after 72 h at 37 °C (Extended Data Fig. 1e). However, increased temperature (55 °C) or exposure to acidic conditions reversed AA_{NAT}-ICLs, which is consistent with Schiff-base hydrolysis and protonation of the deoxyguanine (dG) N^2-amine[10] (Extended Data Fig. 1e, f). DNA crosslink repair is conserved among vertebrates and has been comprehensively studied in *Xenopus* egg extracts[11]. To examine the repair of AA_{NAT}-ICLs using this system, the oligonucleotide was ligated into a plasmid (pICL-AA_{NAT}). We also generated plasmids containing a cisplatin interstrand crosslink (pICL-Pt) or PdG (pPdG), and unmodified control plasmids (pCon) (Extended Data Fig. 1g, h). Crosslinked vectors were stable in non-replicating *Xenopus* egg extracts (Extended Data Fig. 1i).

Acetaldehyde crosslinks are repaired by two routes

Cisplatin ICLs are repaired by the replication-dependent FA pathway, which involves unhooking of the ICL by endonucleases, translesion synthesis (TLS) to bypass the adduct, and homologous recombination to resolve double-strand breaks[12–14] (Fig. 1b, Extended Data Fig. 2a). Replication of pICL-Pt in *Xenopus* egg extracts generates a temporal pattern of repair intermediates, starting with converged forks ('figure 8' structure) and followed by low-mobility products that include

[1]MRC Laboratory of Molecular Biology, Cambridge, UK. [2]Oncode Institute, Hubrecht Institute–KNAW and University Medical Center Utrecht, Utrecht, The Netherlands. [3]Department of Chemistry, The University of Sheffield, Sheffield, UK. [4]MRC Weatherall Institute of Molecular Medicine, University of Oxford, John Radcliffe Hospital, Oxford, UK. [5]Present address: Department of Chemistry, Imperial College London, London, UK. [6]Present address: Institute of Organic Chemistry and Biochemistry of the Czech Academy of Sciences, Prague, Czech Republic. [7]Present address: Department of Biochemistry, University of Cambridge, Cambridge, UK. [8]These authors contributed equally: Michael R. Hodskinson, Alice Bolner. ✉e-mail: kjp@mrc-lmb.cam.ac.uk; p.knipscheer@hubrecht.eu

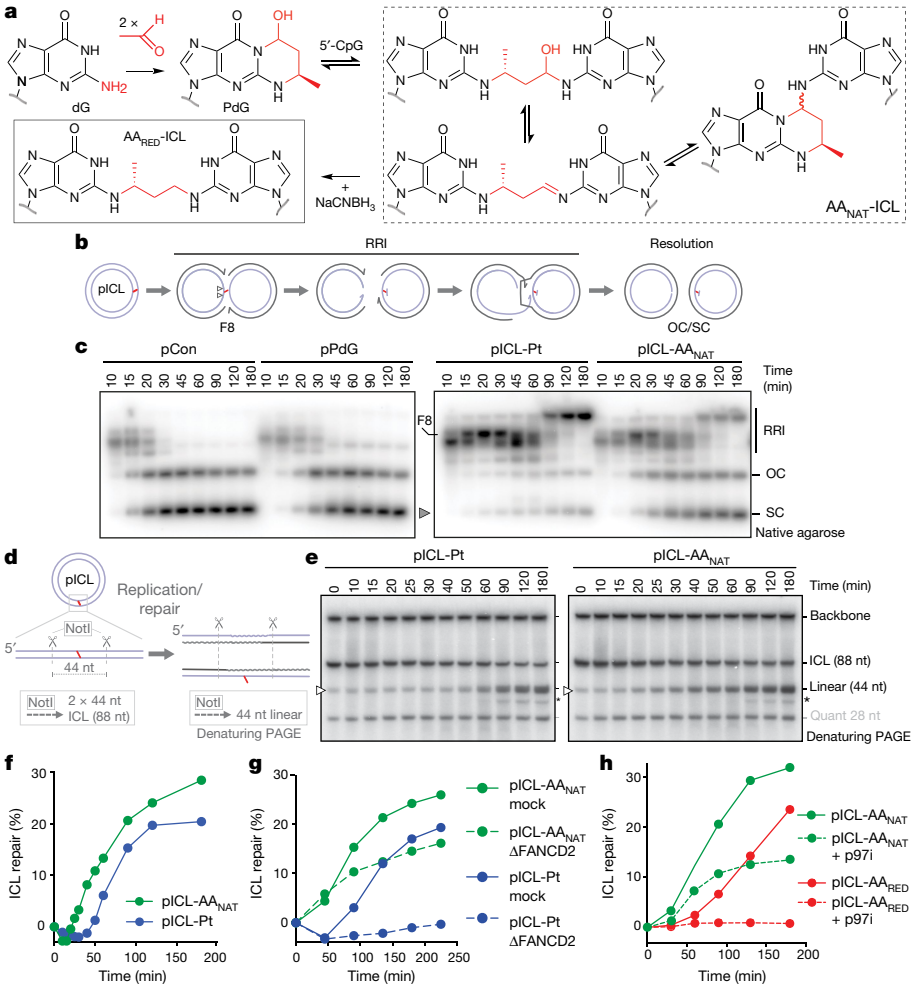


Fig. 1 | AA-ICL repair by Fanconi-dependent and Fanconi-independent mechanisms. a, Reaction scheme of the formation of AA_{NAT}-ICL. Two acetaldehyde molecules react with deoxyguanine (dG) to generate PdG; this reacts with a 5′-CpG guanine on the opposite strand. The AA_{NAT}-ICL exists in an equilibrium between three forms, shown in the dashed box. The crosslink can be chemically reduced with sodium cyanoborohydride to form ICL-AA_{RED} (solid box). **b**, Replication intermediates generated during ICL repair. **c**, Plasmids were replicated in *Xenopus* egg extracts, and the reaction products were resolved by electrophoresis and visualized by autoradiography. Figure 8 structures (F8), later RRIs, open circle (OC) and supercoiled (SC) products (grey arrow) are indicated. Six independent experiments were performed. **d**, Scheme for the NotI ICL repair assay. Wavy lines indicate the part of the strand that is synthesized during repair. **e**, Plasmids were replicated in *Xenopus* egg extracts, repair intermediates were isolated, digested by NotI, and resolved by PAGE. Accumulation of the 44-nt product (white arrow) indicates replication/repair. The asterisk marks a product that is probably generated from end-joining activity in some extracts. Ten independent experiments were performed. **f**, Quantification of ICL repair based on the gels in **e**, as described in the Supplementary Methods. Ten independent experiments were performed. **g**, Quantification of repair in mock-depleted or FANCD2-depleted (ΔFANCD2) extract, based on the gel in Extended Data Fig. 2e. Three independent experiments were performed. **h**, Quantification of repair in the presence or in the absence of p97i, based on the gel in Extended Data Fig. 3c. Three independent experiments were performed.

homologous recombination intermediates (termed replication/repair intermediates (RRI)) and resolved nicked and supercoiled products[14] (Fig. 1b, c). Because the structures of cisplatin crosslinks and acetaldehyde crosslinks differ substantially (Extended Data Fig. 1j), we asked whether they were repaired by similar mechanisms. We replicated pICL-AA_{NAT} and pICL-Pt in *Xenopus* egg extracts, along with non-crosslinked controls, and separated the products by electrophoresis (Fig. 1c). Nicked and supercoiled products accumulated rapidly in experiments with pCon and pPdG, indicating that there was little or no impediment to replication. Experiments with pICL-AA_{NAT} resulted in RRI products similar to those seen with pICL-Pt; however, an earlier accumulation of nicked and supercoiled products compared with pICL-Pt suggested that some pICL-AA_{NAT} were repaired quickly.

We developed an assay, termed the 'NotI assay', to determine whether the nicked and supercoiled products that we observed were indeed AA_{NAT}-ICL repair products. Repair intermediates were digested using NotI, labelled at the 3′ end and resolved by denaturing PAGE (Fig. 1d). Before DNA replication ($t = 0$), this resulted in fragments of 88 nucleotides (nt) (2 × 44 nt, crosslinked), 44 nt (low-level background of non-crosslinked plasmids) and the unresolved vector backbone. For pICL-AA_{NAT} and pICL-Pt, the proportion of the 44-nt fragment increased over time, confirming ICL repair (Fig. 1e). We quantified the repair products and found that they accumulated in greater quantities for pICL-AA_{NAT} compared with pICL-Pt (around 30% compared with around 20%, respectively, at 180 min) and, notably, pICL-AA_{NAT} displayed a faster rate of repair (11% compared with 1% at 50 min) (Fig. 1f, Extended Data Fig. 1k, l). These data suggest that a proportion of pICL-AA_{NAT} was processed in a similar manner to pICL-Pt, but that pICL-AA_{NAT} was additionally repaired by a second, faster mechanism.

Because cisplatin ICLs are repaired by the FA pathway (Extended Data Fig. 2a), we examined whether the same pathway also repaired AA_{NAT}-ICLs. In *Xenopus* egg extracts, both pICL-Pt and pICL-AA_{NAT} stimulated monoubiquitination of FANCD2—the activation step of the FA pathway (Extended Data Fig. 2b). FANCD2 depletion causes a defect in the

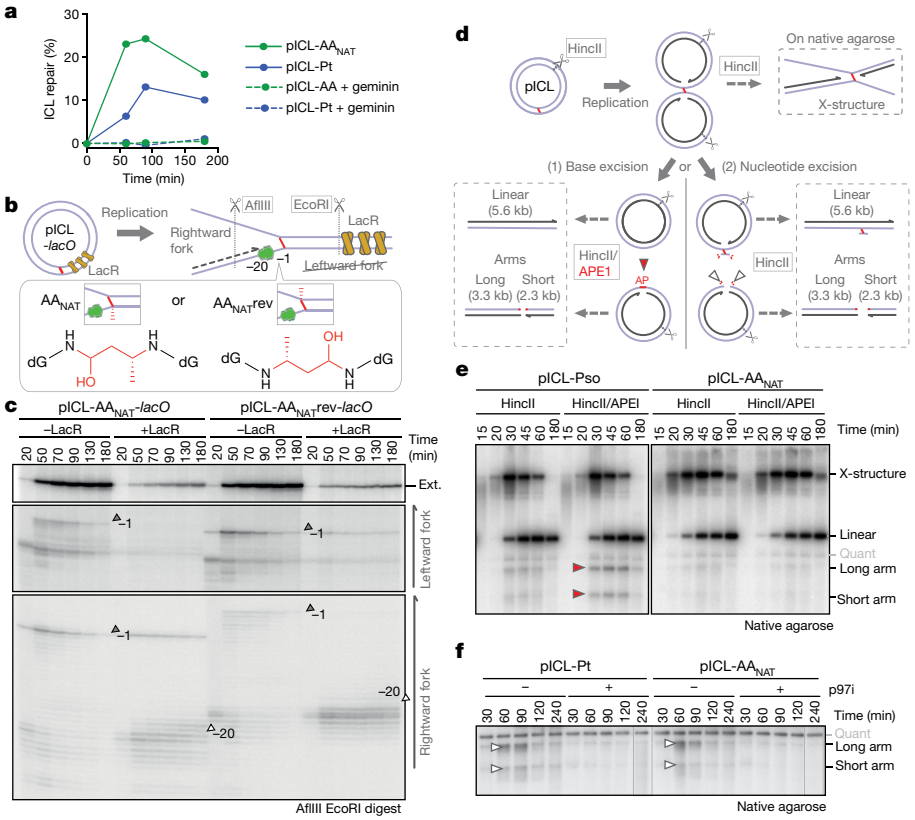


Fig. 2 | The alternative AA-ICL repair route requires DNA replication and fork convergence but not DNA excision. a, Quantification of repair in the presence or in the absence of geminin, based on the gel in Extended Data Fig. 4a. Three independent experiments were performed. **b**, Schematic representation of AA-ICL *lacO* plasmids, showing replication intermediate and digestion sites. **c**, Plasmids were replicated in *Xenopus* egg extracts in the presence or in the absence of LacR. Repair intermediates were digested and separated on a sequencing gel. Grey arrows, –1 product; white arrows, –20 product. Two independent experiments were performed. **d**, Scheme of the assays used to detect base excision (left) and nucleotide excision (right) pathways in **e** and **f**, respectively. DNA fragments formed after HincII/APE1 digestion (left box) or HincII digestion (top and right box). **e**, Plasmids were replicated in *Xenopus* egg extracts with p97i. Repair intermediates were digested and separated by electrophoresis. Red arrows indicate the arms from APE1 incisions. The quantification of these species is shown in Extended Data Fig. 5e. Four independent experiments were performed. **f**, Plasmids were replicated in *Xenopus* egg extracts with or without p97i. Repair intermediates were digested and separated by electrophoresis. White arrows indicate the arms from backbone incisions. The quantification of these species is shown in Extended Data Fig. 5h. Six independent experiments were performed.

unhooking of cisplatin ICLs, which results in the persistence of nascent leading strands one nucleotide before the ICL[12] (–1 position, Extended Data Fig. 2a). To visualize this, we replicated plasmids in mock and FANCD2-depleted extracts, digested the intermediates, and separated the products on a sequencing gel (Extended Data Fig. 2c, d). For both pICL-Pt and pICL-AA_{NAT}, FANCD2 depletion resulted in persisting –1 products (Extended Data Fig. 2d) and fewer extension products, indicating that the repair of both pICL-Pt and pICL-AA_{NAT} involves the FA pathway. We then questioned whether the second route of AA_{NAT}-ICL repair required the FA pathway. In NotI assays, FANCD2-depleted extracts did not support pICL–Pt repair, but pICL-AA_{NAT} was still partially repaired (Fig. 1g, Extended Data Fig. 2e–h). FA-dependent pICL-Pt repair requires unloading of the CMG helicase by the p97 segragase[15]. Consistent with published results, a p97 segregase inhibitor (referred to as p97i) halted pICL-Pt repair (Extended Data Fig. 2i–m). By contrast, p97i blocked only a proportion of AA_{NAT}-ICL repair, while the faster route of repair was unaffected by this treatment (Extended Data Fig. 2i–m). Together, these results indicate that pICL-AA_{NAT} repair proceeds through both an FA-dependent and an FA-independent mechanism.

Chemical reduction of AA_{NAT}-ICL by sodium cyanoborohydride results in a single, stable form of the acetaldehyde crosslink (AA_{RED}-ICL), as confirmed by high-performance liquid chromatography (HPLC) analysis (not shown), resistance to hydrolysis, and matrix-assisted laser desorption/ionization (MALDI) mass spectrometry (Fig. 1a, Extended Data Figs. 1f, 3a). Replication of a plasmid containing an AA_{RED}-ICL (pICL-AA_{RED}) in *Xenopus* egg extracts yielded RRIs that resembled those of the FA pathway, and there was very little accumulation of nicked and supercoiled products (Extended Data Fig. 3b). The NotI assay revealed that repair of pICL-AA_{RED} was slower than that of pICL-AA_{NAT}, and it was abolished by p97i (Fig. 1h, Extended Data Fig. 3c–e). AA_{RED}-ICL is therefore repaired exclusively by the FA pathway, indicating that the alternative repair route is restricted to the native AA-ICL.

Fork convergence is required for AA-ICL repair

To further characterize the second, faster mechanism of AA_{NAT}-ICL repair, we first asked whether it was replication-dependent. The addition of recombinant geminin, which inhibits DNA replication, blocked all repair of both pICL-Pt and pICL-AA_{NAT} (Fig. 2a, Extended Data Fig. 4a–c). Two replication forks must converge in order for pICL-Pt repair to initiate, promoting the ubiquitination and unloading of CMG helicase. Although CMG unloading is not required for the second route of AA_{NAT}-ICL repair, we questioned whether fork convergence was necessary. We generated an AA_{NAT}-ICL plasmid containing a *lacO* array that, when bound by Lac repressor (LacR), blocks the replication fork[16,17]. Because the AA_{NAT}-ICL is non-symmetrical we generated two versions of this plasmid, with the ICL in either orientation with respect to the rightward fork (pICL-AA-*lacO* and pICL-AAreverse-*lacO*, respectively, Fig. 2b).

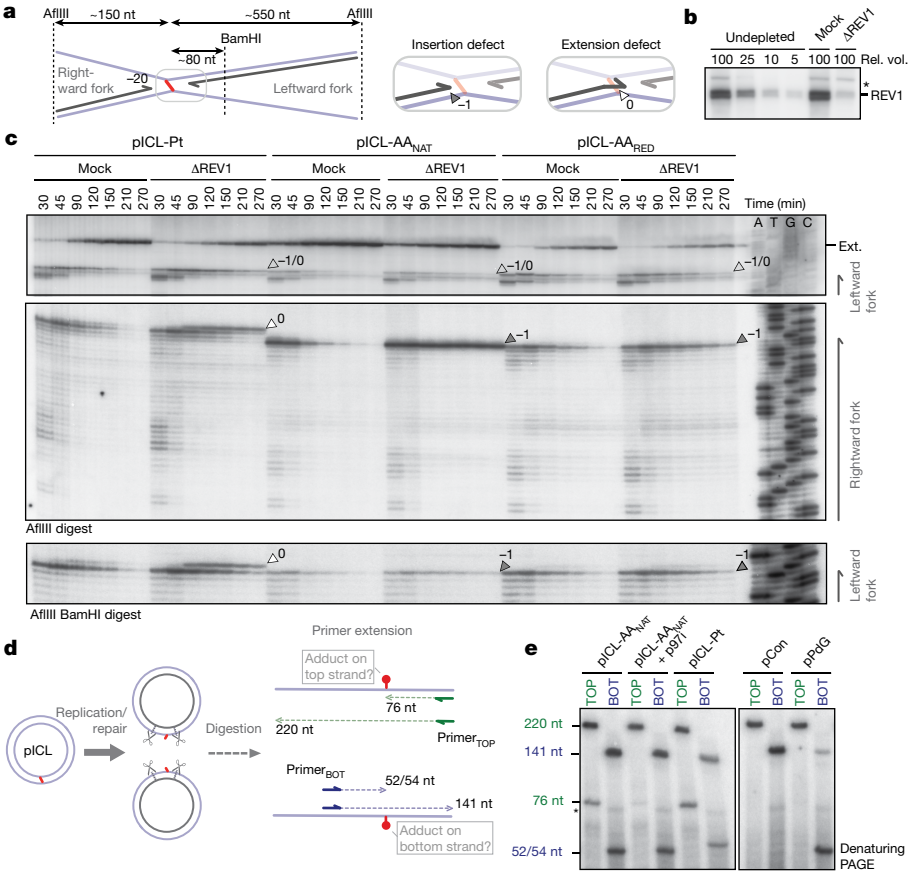


Fig. 3 | AA-ICL repair is mediated by REV1. a, Scheme for the formation of products detected on the sequencing gel in **c**. **b**, Western blot analysing REV1 in the mock-depleted and the REV1-depleted (ΔREV1) extract, compared with a titration of undepleted extract. The asterisk indicates a background band. Nine independent experiments were performed. **c**, Plasmids were replicated in mock-depleted or REV1-depleted extracts and repair intermediates were digested and separated on a sequencing gel alongside a sequencing ladder. White arrows, 0 products; grey arrows, −1 products; open arrows, 0/−1 products. Three independent experiments were performed. **d**, Scheme of the primer extension assay. Late repair products were digested and used as template for primer extension from labelled TOP or BOT (bottom) primers. **e**, Primer extension products were separated by denaturing PAGE. Four independent experiments were performed.

We replicated these plasmids and separated the digested repair intermediates on a sequencing gel (Fig. 2b, c). In the absence of LacR, both leftward and rightward leading strands arrived at the −20 position, approached the −1 position, and were extended past the lesion over time. In the presence of LacR, arrival of the leftward fork was inhibited (as indicated by the absence of −1 and −20 products), showing that the LacR block was functional. The rightward fork persisted at the −20 position, which suggests a failure in repair progression. Moreover, the formation of extension products was impeded upon incubation with LacR, and this was similar for both orientations of AA_{NAT}-ICL (Fig. 2c) and pICL-Pt-*lacO* (Extended Data Fig. 4d). This indicates that the unhooking of AA_{NAT}-ICL requires replication fork convergence. Consistent with this, in the presence of LacR, repair of pICL-AA-*lacO* and pICL-AAreverse-*lacO* was greatly reduced (Extended Data Fig. 4e–h). These data show that the collision of one replication fork with an AA_{NAT}-ICL is insufficient for repair. Although it is plausible that FA-independent AA_{NAT}-ICL unhooking is a consequence of two forks colliding, we think that the mechanism is more likely to be enzymatic, as is the case for all known ICL repair pathways.

New repair route does not involve DNA excision

Whereas cisplatin ICLs are unhooked by nucleolytic incisions, psoralen and abasic site ICLs are preferentially cleaved at a glycosidic bond by NEIL3 glycosylase[12,15,18]. To investigate whether NEIL3 has a role in AA_{NAT}-ICL repair, we complemented NEIL3-depleted extract with recombinant wild-type or catalytically inactive NEIL3 (Extended Data Fig. 5a). Unlike those of the psoralen-ICL plasmid (pICL-Pso)[15], pICL-AA_{NAT} repair intermediates were unaffected by the lack of NEIL3 activity (Extended Data Fig. 5b). Furthermore, human HAP1 cells in which *NEIL3* was knocked out (Δ*NEIL3*) were not hypersensitive to acetaldehyde, and NEIL3 deficiency did not further sensitize a FANCL-deficient strain (Extended Data Fig. 5c, d). It is plausible that another glycosylase unhooks the AA_{NAT}-ICL, so we examined the accumulation of abasic sites—the product of glycosylase cleavage. Recombinant APE1 cleaves abasic sites resulting in the generation of arm fragments upon linearization of replication intermediates. Such arms were generated during pICL-Pso repair, but they did not form during AA_{NAT}-ICL repair[15] (route 1, Fig. 2d, e, Extended Data Fig. 5e–g). This indicates that no abasic site is formed and that AA_{NAT}-ICL repair does not cut the *N*-glycosyl bond.

Next, we tested whether nucleotide excision products are formed when an AA_{NAT}-ICL is processed in *Xenopus* egg extract. Backbone incisions should generate arm fragments when repair intermediates are linearized, as shown for repair of pICL-Pt (route 2, Fig. 2d, f). These arms were also formed during pICL-AA_{NAT} repair, which is consistent with FA pathway activity. Notably, when p97i was added—blocking the FA repair route—no incision products formed, indicating that the second ICL-AA_{NAT} repair pathway does not create backbone incisions (Fig. 2f, Extended Data Fig. 5h–j). To further confirm this, we examined the formation of adducts. In the FA pathway, incisions of one strand

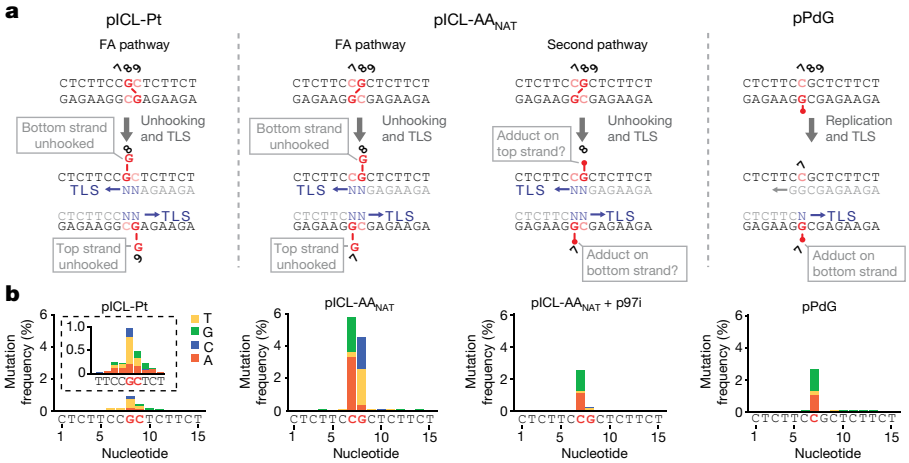


Fig. 4 | AA-ICL repair causes point mutations. a, Scheme showing the positions of DNA lesions and potential adducts (shown in red) generated during repair. Mutations are generated during translesion synthesis (blue Ns and arrow). **b**, Distribution and frequency of nucleotide misincorporation in a 15-bp region flanking the lesions. Strand specificity is lost because sample preparation involves PCR amplification; as such, only the top sequence is indicated below the graphs. Mutation frequency is corrected for mutations found in pCon. Two independent experiments were performed.

result in an unhooked adduct on the other (Extended Data Fig. 5k). Therefore, late repair intermediates of pICL-Pt contain adducts on the top or the bottom strand, because either can be incised. Repair of pICL-AA$_{NAT}$ also generated adducts, but they were largely restricted to the top stand—it is possible that adducts on the bottom strand either were present in amounts below the detection limit or were processed in the extract (Extended Data Fig. 5k, right). However, after the addition of p97i no adducts were detected, indicating that they depend on the FA pathway. In conclusion, the second, faster repair route for AA$_{NAT}$-ICLs does not involve a DNA excision step. This route must therefore operate by cutting within the crosslink itself.

Role of REV1 in acetaldehyde crosslink repair

Such a repair pathway would create an adduct on one or both strands, and should require TLS for nucleotide insertion opposite the adduct and extension beyond it. The TLS factor REV1 is critical for pICL-Pt repair, and so we tested whether it also operates in AA$_{NAT}$-ICL repair. Plasmids were replicated in mock and in REV1-depleted extracts, and digested intermediates were analysed on a sequencing gel (Fig. 3a–c). For pICL-Pt, REV1 depletion led to the accumulation of insertion products (0 products) and reduced amounts of extension products[19] (Fig. 3c). By contrast, for pICL-AA$_{NAT}$, REV1 depletion led to the accumulation of −1 products, which is indicative of REV1-mediated insertion opposite the adduct (Fig. 3c, Extended Data Fig. 6a). Notably, this was also the case for pICL-AA$_{RED}$ repair, indicating that there are different TLS mechanisms within the FA pathway. Furthermore, REV1 depletion caused more extensive leading-strand stalling at the rightward than at the leftward fork for AA-ICL$_{NAT}$ (Fig. 3c). This suggests that unhooking by the second pathway may create an adduct on the bottom strand, which is bypassed by REV1. To test this, we inhibited the FA pathway using p97i and examined lesion bypass in the second pathway of AA$_{NAT}$-ICL repair. As expected, repair of pICL-Pt in the presence of p97i resulted in persistent stalling at the −20 to −40 position, due to defective CMG unloading[15] (Extended Data Fig. 6a). Similar stalled products were observed for AA$_{NAT}$-ICL repair, indicating that the FA pathway repair was inhibited (Extended Data Fig. 6a). However, we observed an accumulation of rightward −1 products, indicating that REV1 depletion prevents lesion bypass by the second pathway. By contrast, the leftward fork was extended without hindrance as no stalled products accumulated and a considerable amount of extension products were formed (Extended Data Fig. 6a). Depletion of REV7 (the regulatory subunit of Pol ζ, which is encoded by *FANCV*, also known as *MAD2L2*), had a very similar effect on AA-ICL repair to that of REV1 depletion (Extended Data Fig. 6d). It also caused an insertion defect in the FA pathway during AA$_{RED}$-ICL repair, as well as persistent stalling—especially of the rightward fork—in AA$_{NAT}$-ICL repair. As Pol ζ is primarily known for its extension activity[20], this defect could be due to co-depletion of the REV1–Pol ζ complex[19] (Extended Data Fig. 6b, c). In summary, the second pathway of AA$_{NAT}$-ICL repair generates an adduct on the bottom strand that requires REV1 and Pol ζ for bypass. The top strand, however, is readily extended without these TLS factors.

To further examine adduct formation, we isolated late repair products and subjected them to primer extension reactions at the ICL region using a high-fidelity polymerase (Fig. 3d). As expected, pICL-Pt repair products—which contain adducts on either parental strand—induced the formation of stalled extension products on both strands (Fig. 3e). For AA$_{NAT}$-ICL, stalling also occurred on both strands but was more extensive on the bottom strand. Moreover, treatment with p97i—that is, blocking the FA pathway—almost entirely eliminated stalling on the top strand (Fig. 3e). These results suggest that the second repair pathway regenerates dG on the top strand, but creates a dG adduct on the bottom strand.

Acetaldehyde crosslink repair is mutagenic

Finally, we examined the fidelity of AA$_{NAT}$-ICL repair. To achieve this, we replicated plasmids in *Xenopus* egg extracts, recovered late repair products, and subjected them to high-throughput sequencing (Extended Data Table 1, Fig. 4a, b). Consistent with a previous report[19], we confirmed that pICL-Pt is repaired with a low error rate—less than 2% of the products carried a mutation at sites corresponding to crosslinked guanines (Fig. 4b, Extended Data Fig. 7a). The most common substitution was G>T transversion. Repaired pICL-AA$_{NAT}$ products show two differences: first, around 10% carry mutations at the crosslinked sites; and second, the mutational spectrum differs, with C>G, C>A, G>C and G>T transversions all observed (Fig. 4b, Extended Data Fig. 7a–c). The frequency of consecutive mutations (at least 2 nt) around the ICL was around 100-fold lower than that of single mutations, in agreement with a report suggesting that ICLs do not drive tandem mutations caused by acetaldehyde[21]. AA$_{NAT}$-ICL repair products generated in the presence of p97i—blocking the FA route—lost most of their mutations at position G8. This indicates that FA-dependent bypass of the top strand adduct is the predominant source of mutation at G8. Under these conditions,

mutations are almost entirely restricted to C7 and the mutational spectrum is almost identical to that obtained from pPdG repair (Fig. 4b, Extended Data Fig. 7b). These data strongly suggest that the second route of AA_{NAT}-ICL repair reverses this crosslink to yield a monoadduct that is similar or identical to the original propanoguanine. Consistent with this model, we found that the insertion step of TLS past a PdG adduct is mediated by REV1 (Extended Data Fig. 6e).

Conclusion

In summary, we have determined the repair mechanisms of an important class of endogenous DNA crosslinks—those caused by acetaldehyde. The central role of the FA repair pathway in removing such crosslinks agrees with the strong genetic evidence underpinning two-tier protection against this aldehyde[3,6,7]. However, we also uncovered a new pathway of DNA crosslink repair that operates without excision repair. This mechanism requires replication fork convergence and is unique in that it unhooks the ICL by cutting within the crosslink itself. Repair of AA_{NAT}-ICLs by both pathways is error-prone and requires the TLS polymerases REV1 and Pol ζ. However, this new repair modality has an obvious advantage in that it avoids the creation of DNA strand breaks or abasic sites, both of which can promote large-scale genome instability.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests are available at https://doi.org/10.1038/s41586-020-2059-5.

1. Brooks, P. J., Enoch, M. A., Goldman, D., Li, T. K. & Yokoyama, A. The alcohol flushing response: an unrecognized risk factor for esophageal cancer from alcohol consumption. *PLoS Med*. **6**, e1000050 (2009).
2. Lai, C. L. et al. Dominance of the inactive Asian variant over activity and protein contents of mitochondrial aldehyde dehydrogenase 2 in human liver. *Alcohol. Clin. Exp. Res*. **38**, 44–50 (2014).
3. Langevin, F., Crossan, G. P., Rosado, I. V., Arends, M. J. & Patel, K. J. Fancd2 counteracts the toxic effects of naturally produced aldehydes in mice. *Nature* **475**, 53–58 (2011).
4. Garaycoechea, J. I. & Patel, K. J. Why does the bone marrow fail in Fanconi anemia? *Blood* **123**, 26–34 (2014).
5. Hira, A. et al. Variant ALDH2 is associated with accelerated progression of bone marrow failure in Japanese Fanconi anemia patients. *Blood* **122**, 3206–3209 (2013).
6. Garaycoechea, J. I. et al. Genotoxic consequences of endogenous aldehydes on mouse haematopoietic stem cell function. *Nature* **489**, 571–575 (2012).
7. Garaycoechea, J. I. et al. Alcohol and endogenous aldehydes damage chromosomes and mutate stem cells. *Nature* **553**, 171–177 (2018).
8. Wang, M. et al. Identification of DNA adducts of acetaldehyde. *Chem. Res. Toxicol*. **13**, 1149–1157 (2000).
9. Cho, Y. J. et al. Stereospecific formation of interstrand carbinolamine DNA cross-links by crotonaldehyde- and acetaldehyde-derived α-CH_3-γ-OH-1,N^2-propano-2′-deoxyguanosine adducts in the 5′-CpG-3′ sequence. *Chem. Res. Toxicol*. **19**, 195–208 (2006).
10. Loudon, G. M. *Organic Chemistry* 874–878 (Oxford Univ. Press, 2002).
11. Duxin, J. P. & Walter, J. C. What is the DNA repair defect underlying Fanconi anemia? *Curr. Opin. Cell Biol*. **37**, 49–60 (2015).
12. Knipscheer, P. et al. The Fanconi anemia pathway promotes replication-dependent DNA interstrand cross-link repair. *Science* **326**, 1698–1701 (2009).
13. Long, D. T., Räschle, M., Joukov, V. & Walter, J. C. Mechanism of RAD51-dependent DNA interstrand cross-link repair. *Science* **333**, 84–87 (2011).
14. Räschle, M. et al. Mechanism of replication-coupled DNA interstrand crosslink repair. *Cell* **134**, 969–980 (2008).
15. Semlow, D. R., Zhang, J., Budzowska, M., Drohat, A. C. & Walter, J. C. Replication-dependent unhooking of DNA interstrand cross-links by the NEIL3 glycosylase. *Cell* **167**, 498–511 (2016).
16. Duxin, J. P., Dewar, J. M., Yardimci, H. & Walter, J. C. Repair of a DNA-protein crosslink by replication-coupled proteolysis. *Cell* **159**, 346–357 (2014).
17. Zhang, J. et al. DNA interstrand cross-link repair requires replication-fork convergence. *Nat. Struct. Mol. Biol*. **22**, 242–247 (2015).
18. Klein Douwel, D. et al. XPF-ERCC1 acts in unhooking DNA interstrand crosslinks in cooperation with FANCD2 and FANCP/SLX4. *Mol. Cell* **54**, 460–471 (2014).
19. Budzowska, M., Graham, T. G. W., Sobeck, A., Waga, S. & Walter, J. C. Regulation of the Rev1–Pol ζ complex during bypass of a DNA interstrand cross-link. *EMBO J*. **34**, 1971–1985 (2015).
20. Johnson, R. E., Washington, M. T., Haracska, L., Prakash, S. & Prakash, L. Eukaryotic polymerases ι and ζ act sequentially to bypass DNA lesions. *Nature* **406**, 1015–1019 (2000).
21. Matsuda, T., Kawanishi, M., Yagi, T., Matsui, S. & Takebe, H. Specific tandem GG to TT base substitutions induced by acetaldehyde are due to intra-strand crosslinks between adjacent guanine bases. *Nucleic Acids Res*. **26**, 1769–1774 (1998).

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

Datasets generated during the current study are available from the corresponding authors upon reasonable request.

Code availability

Code used in the current study is available from the corresponding authors upon reasonable request.

Acknowledgements We thank G. P. Crossan, J. T. P. Yeeles and members of the Knipscheer and Patel laboratories for critical reading of the manuscript; Hubrecht animal caretakers for animal support; J. C. Walter and D. R. Semlow for providing us with recombinant *X. laevis* NEIL3 proteins, the AP-ICL and *lacO* plasmids; and S. Y. Peak-Chew, S. Maslen and M. Skehel for mass spectrometry analysis. This work was supported by a project grant from the Dutch Cancer Society (KWF HUBR 2015-7736 to P.K.), the Gravitation program CancerGenomiCs.nl from the Netherlands Organisation for Scientific Research (NWO), part of the Oncode Institute, which is partly financed by the Dutch Cancer Society. K.S. was supported by the Uehara Memorial Foundation, the Mochida Memorial Foundation for Medical and Pharmaceutical Research, and the JSPS Postdoctoral Fellowship for Research Abroad. A.N.K.-L. was supported by the Wellcome Trust. J.S. was supported by Cancer Research UK. This work was supported by a European Research Council Starting Grant (ERC-STG 678423-EpiID) to J.K. and K.R. and an NWO Veni grant (016.Veni.181.013) to K.R.

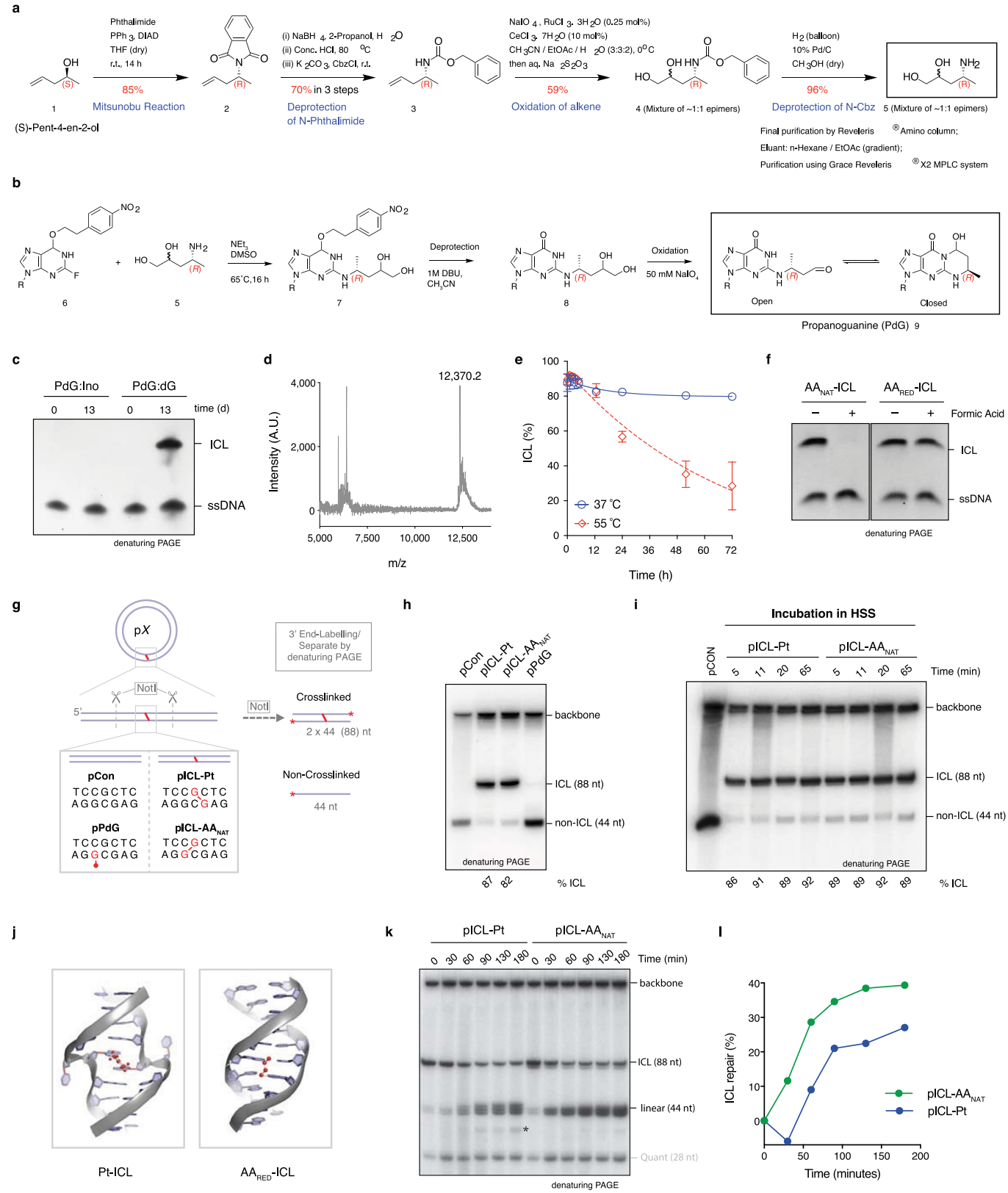
Author contributions K.J.P. and P.K. initiated and supervised the study. M.R.H. co-ordinated key aspects of the project. K.S., A.N.K.-L. and K.R. contributed equally. M.R.H., M.M., J.S., M.P., D.M.W. and J.W.C. designed the strategy and synthesized the AA-ICL and PdG adducts. A.N.K.-L. performed cell toxicity experiments. A.B., M.W. and P.K. designed biochemical assays in *Xenopus* egg extracts. A.B., K.S. and M.W. conducted *Xenopus* extract assays. K.R. and J.K. performed bioinformatic analysis. K.J.P. and P.K. wrote the manuscript and M.R.H. designed and created the figures.

Competing interests The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41586-020-2059-5.

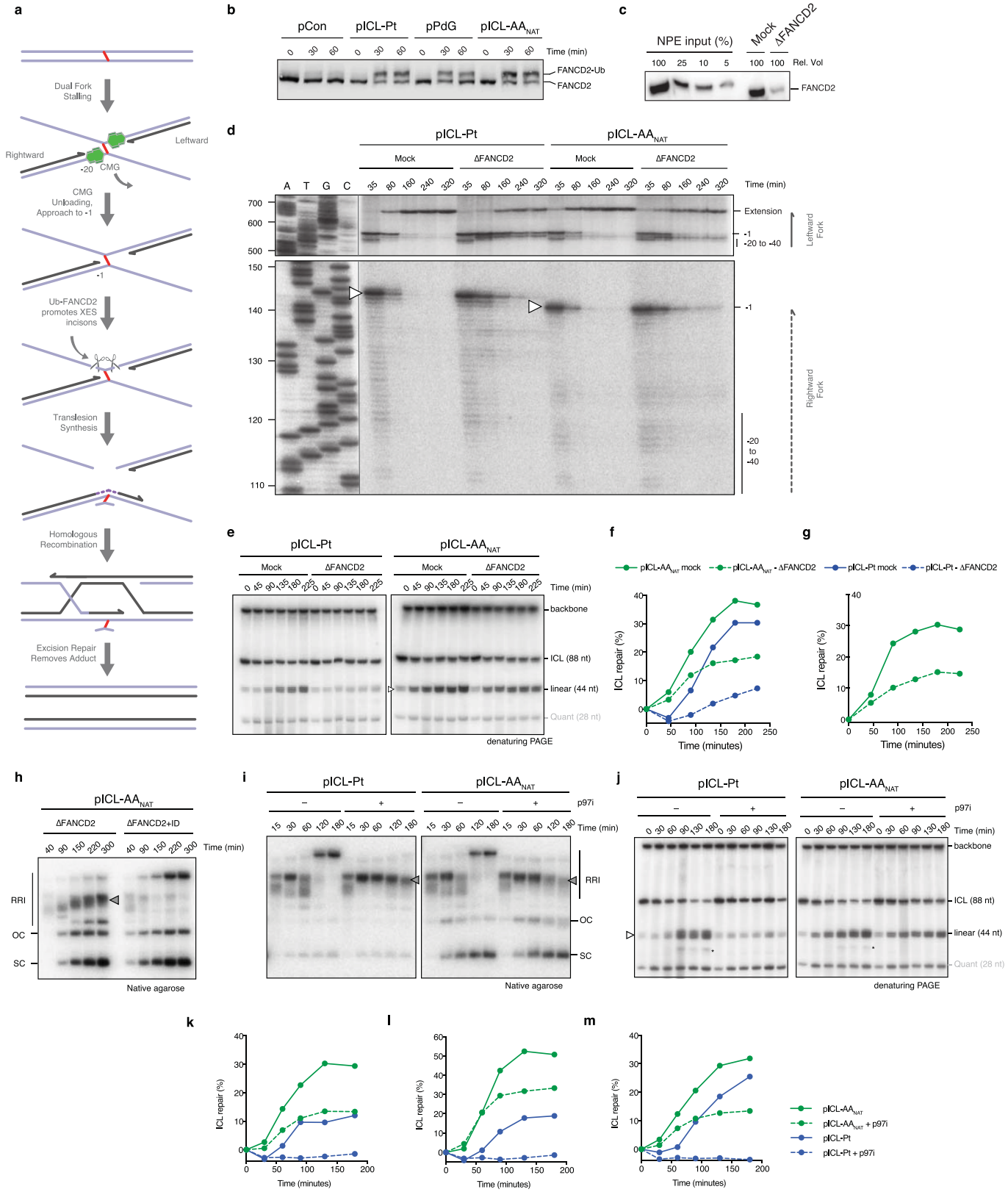
Correspondence and requests for materials should be addressed to K.J.P. or P.K.



Extended Data Fig. 1 | See next page for caption.

Extended Data Fig. 1 | Acetaldehyde ICLs are stable and are repaired in *Xenopus* egg extracts. a, Scheme for the synthesis of the precursor, 4-(*R*)-aminopentane-1,2-diol. **b**, Site-specific synthesis of a PdG adduct in a DNA oligonucleotide. **c**, Denaturing PAGE showing crosslink formation between dG and PdG, but not between dG and inosine (ino), which lacks an N^2 amine. Two independent experiments were performed. **d**, Confirmation of AA-ICL formation by MALDI MS. The peak at *m/z* 12,370.2 represents the imine or pyrimidopurinone form. Two further peaks at *m/z* 5,979.41 and 6,409.74 equate to masses for the two parent oligonucleotides, consistent with the mass of the carbinolamine form after dissociation back to PdG and dG under the desorption/ionization conditions. Three independent experiments were performed. **e**, Stability of AA_{NAT}-ICL as a function of temperature and time, as determined by radiolabelling ([α-^{32}P]dCTP) and resolution by denaturing PAGE. Error bars represent s.e.m. from three independent experiments. **f**, AA_{NAT}-ICL is susceptible to hydrolysis in aqueous acid, whereas AA_{RED}-ICL is stable. Pre-purification crosslink reactions were incubated with or without formic acid and products were resolved by denaturing PAGE. Three independent experiments were performed. **g**, Scheme depicting the type and position of the DNA lesions used in this study. Duplex DNA with or without the indicated lesions was annealed into a backbone vector to generate circular plasmids with or without damage. **h**, To determine the percentage of crosslinks, the ICL-containing plasmids were digested with NotI, labelled at the 3′-end by end filling with [α-^{32}P]dCTP, and separated by denaturing PAGE. Crosslinked DNA (88 nt) shows slower mobility compared with non-crosslinked DNA (44 nt). The percentage of crosslinks was calculated by comparing the 88-nt product with the 44-nt products. Two independent experiments were performed. **i**, AA_{NAT}-ICL and Pt-ICL are stable in *Xenopus* egg extracts. Plasmids were incubated in a high-speed supernatant (HSS) extract. DNA was extracted and analysed as described in **h**. Three independent experiments were performed. **j**, Solution structures of a cisplatin ICL and a reduced form of an acetaldehyde ICL (PDB: 1DDP[22] and 2HMD[23], cartoon representation generated in PyMOL). **k**, The indicated plasmids were replicated in *Xenopus* egg extracts and repair intermediates were digested with NotI, labelled at the 3′-end, and resolved by denaturing PAGE. The increase in intensity of the 44-nt band over time indicates ongoing replication and repair. A higher mobility band, probably generated from end-joining activity in some extracts, is indicated by an asterisk. This gel is the independent experimental duplicate of that in Fig. 1e. **l**, Quantification of repair based on the intensity of the 44-nt product on the gel in **k**, as described in the Supplementary Methods. This graph is the independent experimental duplicate of that in Fig. 1f. Additional replicates of these experiments are presented in Fig. 2a, Extended Data Figs. 2k–m, 4b.

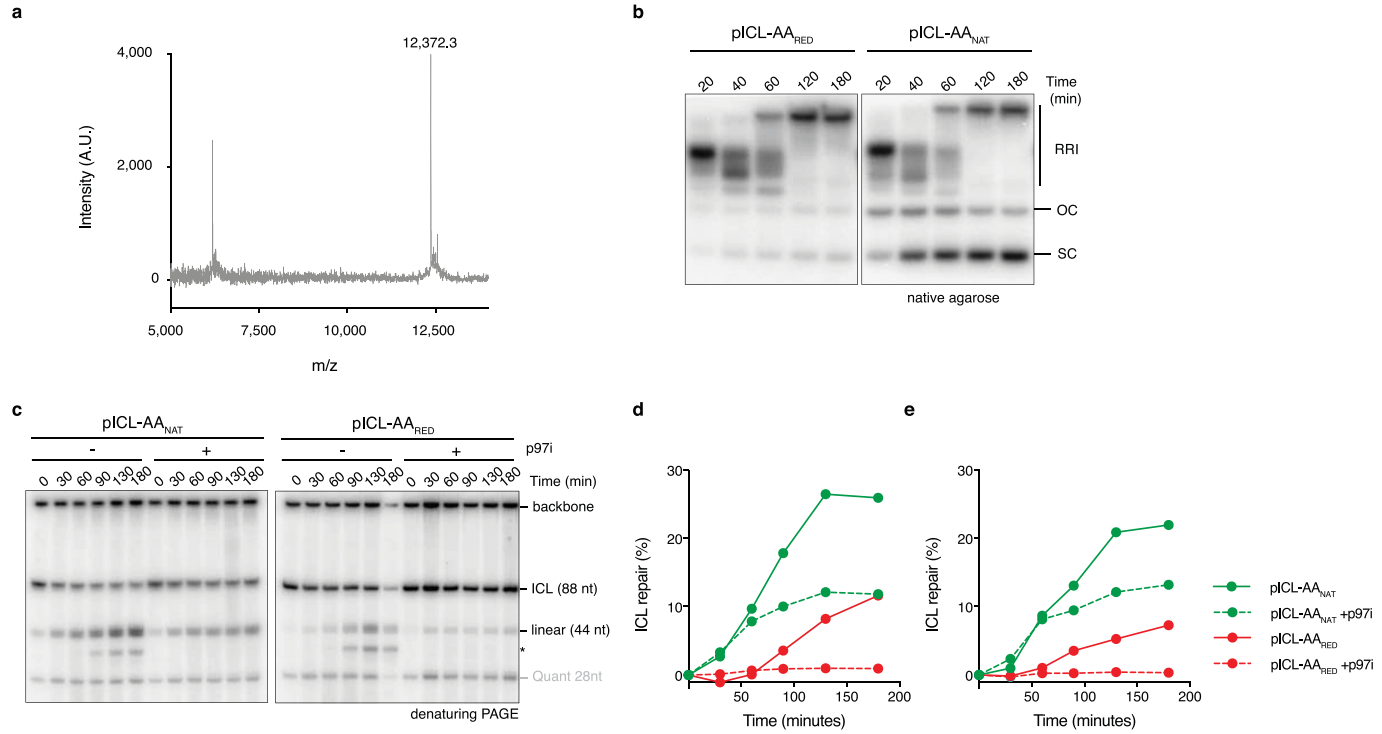
22. Huang, H., Zhu, L., Reid, B. R., Drobny, G. P. & Hopkins, P. B. Solution structure of a cisplatin-induced DNA interstrand cross-link. *Science* **270**, 1842–1845 (1995).
23. Cho, Y. J., Kozekov, I. D., Harris, T. M., Rizzo, C. J. & Stone, M. P. Stereochemistry modulates the stability of reduced interstrand cross-links arising from *R*- and *S*-α-CH_3-γ-OH-1,N^2-propano-2′-deoxyguanosine in the 5′-CpG-3′ DNA sequence. *Biochemistry* **46**, 2608–2621 (2007).



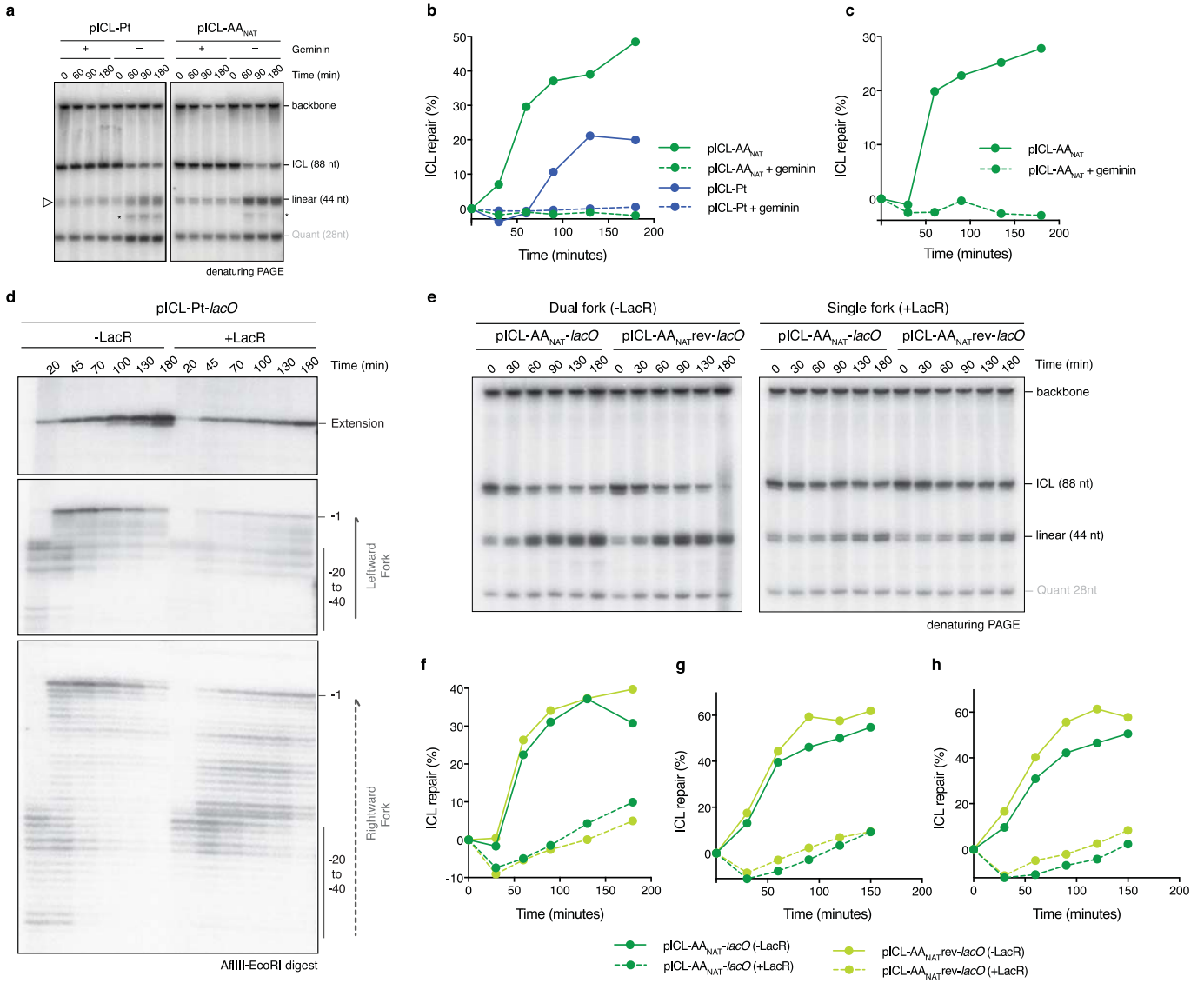
Extended Data Fig. 2 | See next page for caption.

Extended Data Fig. 2 | Acetaldehyde ICLs are repaired by Fanconi-dependent and Fanconi-independent mechanisms. a, Model for ICL repair by the FA pathway. Upon convergence of two replication forks at the crosslink, the CMG helicase is unloaded from the DNA to enable the approach of one replication fork to the −1 position. Ubiquitylation of FANCD2 promotes the recruitment of the XPF-ERCC1-SLX4 (XES) complex to the ICL, which enables nucleolytic incisions that unhook the crosslink. This step could be preceded by fork reversal of one of the stalled replication forks[24]. Incisions generate a broken strand and a strand with an adduct; the latter is bypassed by TLS whereas the broken strand is repaired by homologous recombination. In mammalian cells, it has been shown that a single fork can pass over the ICL without unhooking[25]; this 'traverse' gives rise to a structure that resembles the one generated after fork convergence and CMG unloading and could follow the same steps subsequently. **b**, The indicated plasmids were replicated in *Xenopus* egg extracts and reaction samples were analysed by western blot with FANCD2 antibody. Two independent experiments were performed. **c**, Western blot of FANCD2, showing a titration of *Xenopus* egg extracts compared to mock and FANCD2-depleted extracts. Two independent experiments were performed. **d**, The indicated plasmids were replicated in mock or in FANCD2-depleted extracts in the presence of [α-^{32}P]dCTP. Repair products were digested by AflIII, separated on a sequencing gel alongside a ladder derived from extension primer S, and visualized by autoradiography. The white arrow denotes the −1 product, which is 2 nt larger in pICL-Pt owing to the position of the ICL. Three independent experiments were performed. **e**, The indicated plasmids were replicated in mock or FANCD2-depleted extracts and repair intermediates were digested with NotI, labelled at the 3′-end, and resolved by denaturing PAGE. Quantification of repair based on the intensity of the 44-nt product is shown in Fig. 1g. **f**, The independent experimental duplicate of Fig. 1g. **g**, The independent experimental triplicate of Fig. 1g, but using only pICL-AA$_{NAT}$. **h**, Plasmid pICL-AA$_{NAT}$ was replicated in FANCD2-depleted extract, or FANCD2-depleted extract supplemented with a recombinant FANCI–FANCD2 complex (ID). Reaction samples were resolved by native agarose gel and visualized by autoradiography. RRIs, open circle (OC) and supercoiled (SC) products are indicated. The stalled repair product (grey arrow) is indicated. Two independent experiments were performed. **i**, The indicated plasmids were replicated in *Xenopus* egg extracts in the presence or in the absence of p97i and the intermediates were resolved by native agarose gel electrophoresis. The stalled repair products (grey arrow) are indicated. Seven independent experiments were performed. **j**, The indicated plasmids were replicated in *Xenopus* egg extracts in the presence or in the absence of p97i, and repair intermediates were digested with NotI, labelled at the 3′-end, and resolved by denaturing PAGE. The increase in intensity of the 44-nt band (white arrow) over time indicates ongoing replication and repair. A higher mobility band, probably generated from end-joining activity in some extracts, is indicated with an asterisk. **k**, Quantification of repair based on the intensity of the 44-nt product on the gel shown in **j**, as indicated in the Supplementary Methods. **l**, Quantified independent experimental duplicate of **j**. **m**, Quantified independent experimental triplicate of **j**.

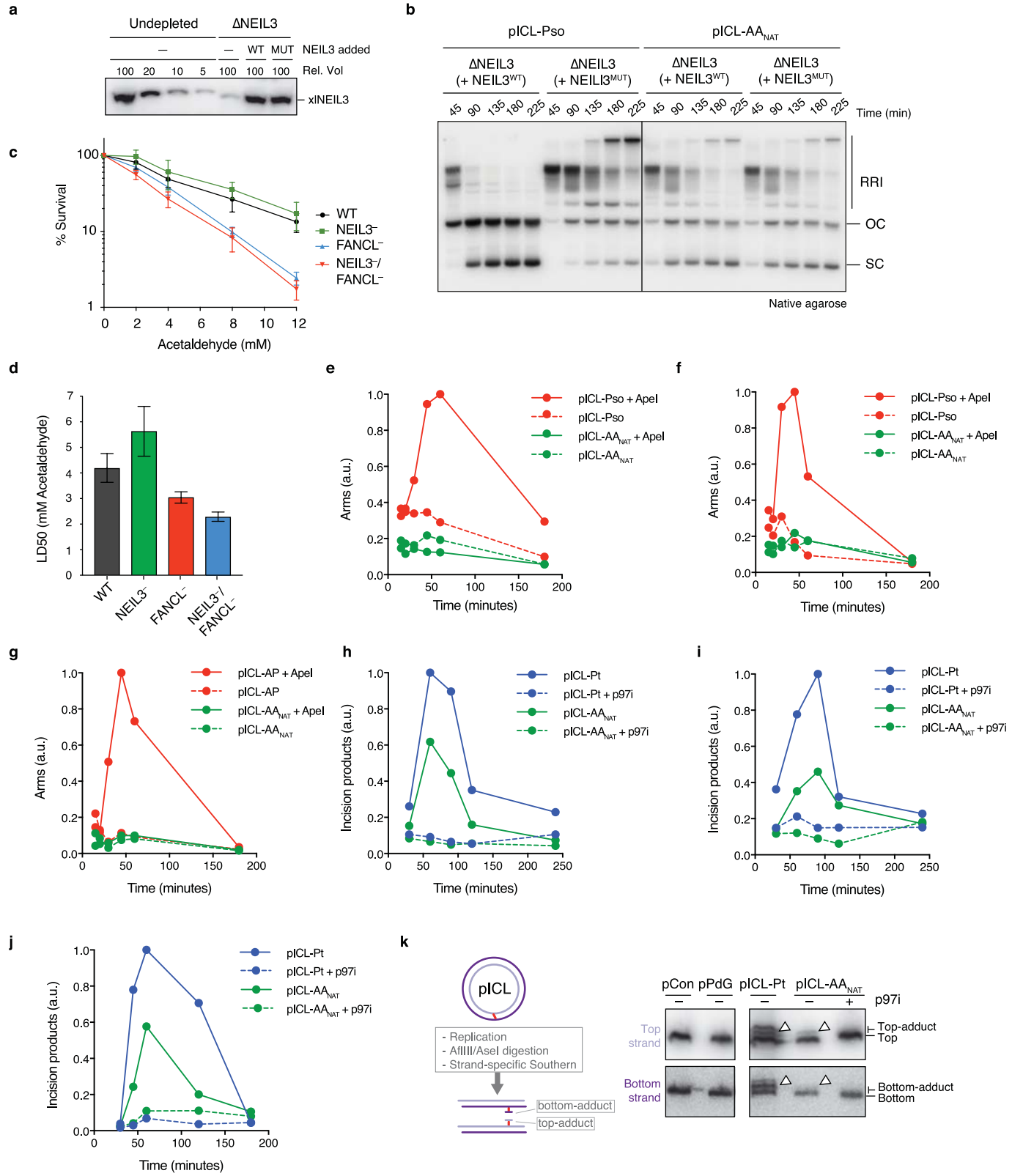
24. Amunugama, R. et al. Replication fork reversal during DNA interstrand crosslink repair requires CMG unloading. *Cell Rep.* **23**, 3419–3428 (2018).
25. Huang, J. et al. The DNA translocase FANCM/MHF promotes replication traverse of DNA interstrand crosslinks. *Mol. Cell* **52**, 434–446 (2013).



Extended Data Fig. 3 | Reduced acetaldehyde ICLs are repaired by the Fanconi pathway. a, MALDI MS data confirms the identity and the stability of the reduced ICL. Three independent experiments were performed. **b**, Plasmids were replicated in extract and products were resolved on a native agarose gel. Three independent experiments were performed. **c**, The indicated plasmids were replicated in *Xenopus* egg extracts in the presence or in the absence of p97i and the repair intermediates were analysed using the NotI digestion assay. The increase in intensity of the 44-nt band indicates ongoing replication and repair. A band with higher mobility, probably generated from end-joining activity in some extracts, is indicated with an asterisk. Quantification of repair based on this gel is shown in Fig. 1h. Three independent experiments were performed. **d**, The independent experimental duplicate of Fig. 1h. **e**, The independent experimental triplicate of Fig. 1h.

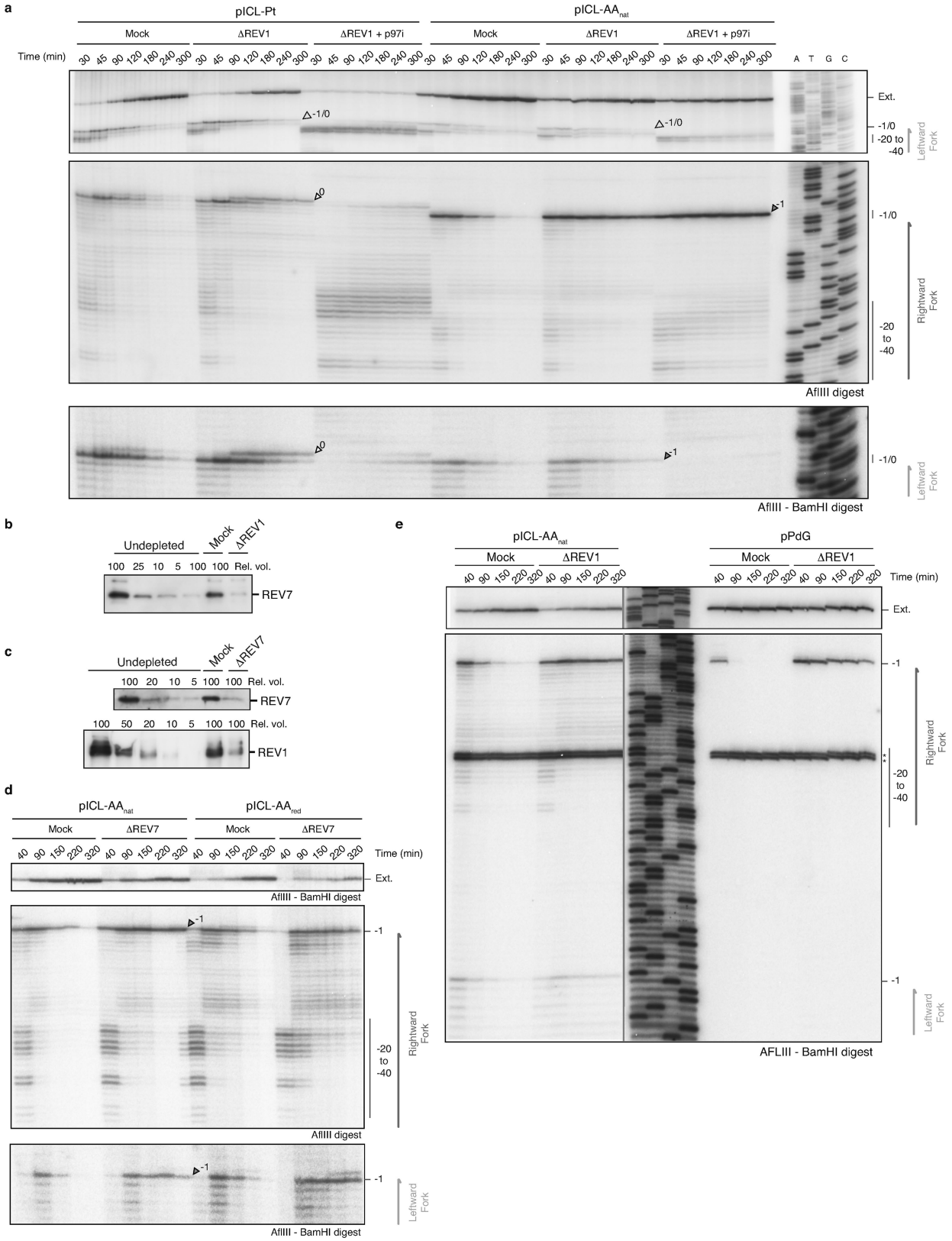


Extended Data Fig. 4 | Acetaldehyde ICL repair requires DNA replication and replication fork convergence. a, The indicated plasmids were replicated in *Xenopus* egg extracts in the presence or in the absence of geminin. Repair intermediates were digested with NotI, labelled at the 3′-end, and resolved by denaturing PAGE. Quantification of repair based on the intensity of the 44-nt product is shown in Fig. 2a. **b**, The independent experimental duplicate of Fig. 2a. **c**, The independent experimental triplicate of Fig. 2a but using only pICL-AA$_{NAT}$. **d**, pICL-Pt-*lacO* was replicated in *Xenopus* egg extracts containing [α-^{32}P]dCTP in the presence or in the absence of LacR. The repair intermediates were digested by AflIII and EcoRI, separated on a sequencing gel and visualized by autoradiography. Two independent experiments were performed. **e**, The indicated plasmids were replicated in extracts in the presence or in the absence of LacR. Repair products were digested with NotI, labelled at the 3′-end, and resolved by denaturing PAGE. Three independent experiments were performed. **f**, Quantification of repair based on the intensity of the 44-nt product on the gel shown in **e**, as described in the Supplementary Methods. Three independent experiments were performed. **g**, The independent experimental duplicate of **f**. **h**, The independent experimental triplicate of **f**.



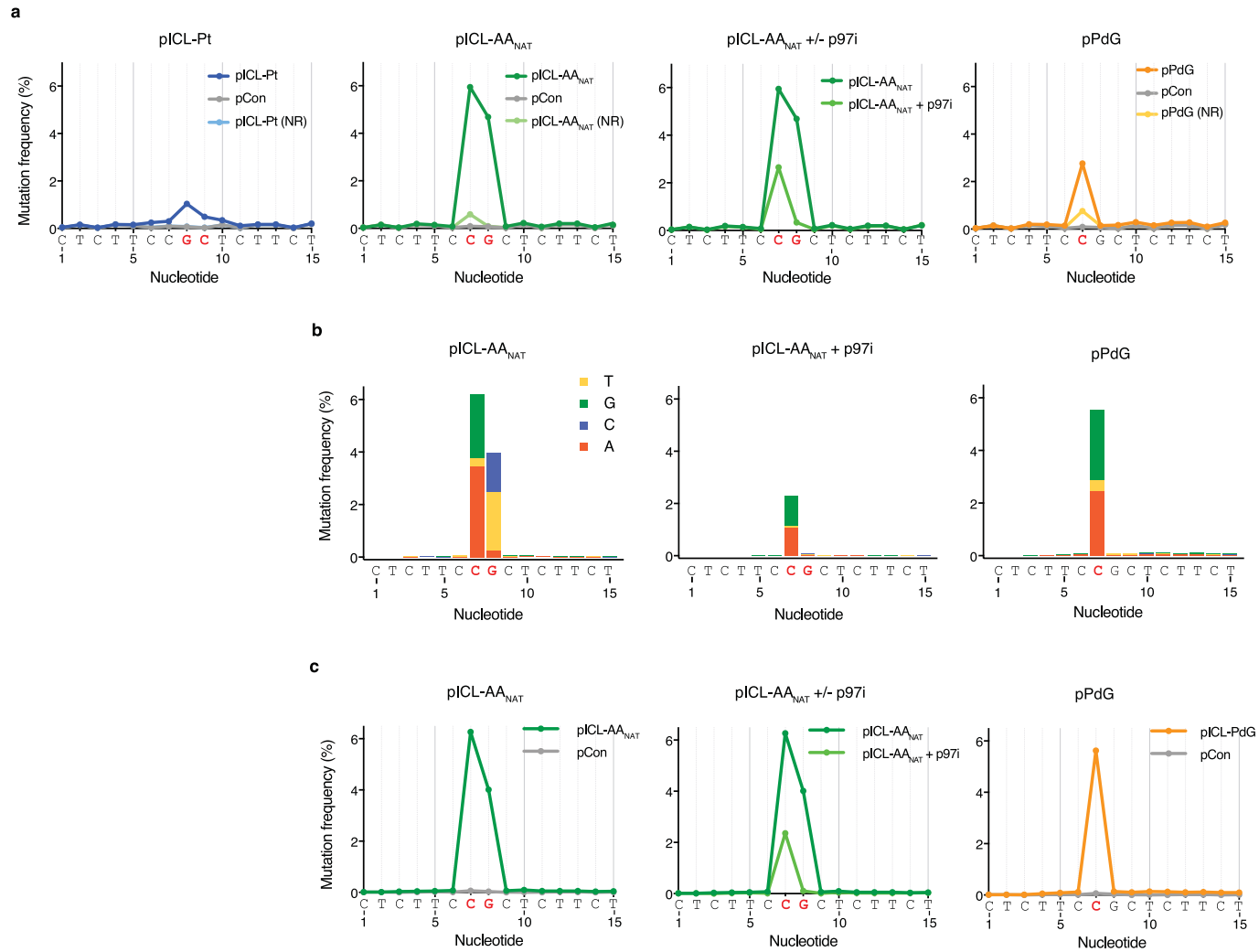
Extended Data Fig. 5 | See next page for caption.

Extended Data Fig. 5 | The alternative route of AA-ICL repair does not involve DNA excision. a, Western blot showing a titration of *Xenopus* egg extracts compared to the NEIL3-depleted (ΔNEIL3) extract and NEIL3-depleted extract supplemented with recombinant wild-type NEIL3 (WT) or catalytically inactive NEIL3 (MUT). Three independent experiments were performed. **b**, The indicated plasmids were replicated in NEIL3-depleted extract containing [α-^{32}P]dCTP, supplemented with wild-type (WT) or catalytically inactive (MUT) NEIL3. Replication intermediates were resolved by native agarose gel electrophoresis and visualized by autoradiography. Three independent experiments were performed. **c**, Clonogenic survival of wild-type, NEIL3-, FANCL- or NEIL3/FANCL-deficient human HAP1 cells after a 2-h exposure to acetaldehyde. Three independent experiments were performed. Data are mean ± s.e.m. **d**, The median lethal dose (LD_{50}) of acetaldehyde for the survival of wild-type and deficient HAP1 cells was calculated by regression analysis of the curves presented in **c**. Data are mean ± s.e.m. Three independent experiments were performed. **e**, Quantification of the arm fragments resulting from APE1 treatment (AP sites) from the gel in Fig. 2e. **f**, Quantification of the APE1 arms as in **e**, from an independent duplicate experiment without the addition of p97i. **g**, Quantification of the APE1 arms as in **e**, from an independent triplicate experiment without the addition of p97i. As a positive control we used a plasmid containing an abasic-site-induced interstrand crosslink (pICL-AP) that is also repaired via the glycosylase NEIL3. **h**, Quantification of the HincII arm fragments from the gel in Fig. 2f. **i**, Quantification of the HincII arms as in **h**, from an independent duplicate experiment. **j**, Quantification of the HincII arms as in **h**, from an independent triplicate experiment. **k**, Schematic representation of the formation of DNA adducts by unhooking incisions during ICL repair (left). These adducts are not removed during ICL repair in *Xenopus* egg extracts[14] and can therefore be visualized. Plasmids were replicated in *Xenopus* egg extracts in the presence or in the absence of p97i. Late reaction samples were digested with AflIII and AseI and separated on a sequencing gel. Adducts on either the top or the bottom strand (white arrowheads) were detected by strand-specific Southern blotting (right). Three independent experiments were performed.



Extended Data Fig. 6 | See next page for caption.

Extended Data Fig. 6 | Both routes of AA-ICL repair are mediated by REV1 and REV7. a, The indicated plasmids were replicated in mock or REV1-depleted extracts, in the presence or in the absence of p97i. Reaction intermediates were digested by either AflIII or AflIII and BamHI, separated on a sequencing gel alongside a ladder derived from extension of primer S, and visualized by autoradiography. White arrows denote 0 products, dark grey arrows indicate −1 products, and light grey arrows indicate 0/−1 products (not separated). Two independent experiments were performed. **b**, Western blot detection of REV7 in REV1- or mock-depleted *Xenopus* egg extracts compared to a titration of undepleted extract. Two independent experiments were performed. **c**, Western blot detection of REV7 and REV1 in REV7-depleted (ΔREV7) or mock-depleted *Xenopus* egg extracts compared to a titration of undepleted extract. Two independent experiments were performed. **d**, The indicated plasmids were replicated in mock- or REV7-depleted extracts and reaction intermediates were digested by either AflIII or AflIII and BamHI, separated on a sequencing gel and visualized by autoradiography. Grey arrows indicate −1 products. Two independent experiments were performed. **e**, Indicated plasmids were replicated in mock- or REV1-depleted extracts and reaction intermediates were digested by AflIII and BamHI, separated on a sequencing gel alongside a ladder derived from extension of primer S, and visualized by autoradiography. The asterisk indicates a 121-nt background fragment caused by a second BamHI restriction site in the leftward fork. Two independent experiments were performed.



Extended Data Fig. 7 | Mutagenic outcome of acetaldehyde crosslink repair. **a**, Frequency of nucleotide misincorporation in a 15-bp region flanking the lesions present in the indicated plasmids. The mutation frequencies for the same plasmid that has not been replicated in *Xenopus* egg extracts (NR) and for the control vector (pCon) are plotted together. Strand specificity is lost because the sample preparation involves PCR amplification; as such, only the top sequence is indicated below the graphs. See Fig. 4. **b**, Distribution and frequency of nucleotide misincorporations in a 15-bp region flanking the lesions present in the indicated plasmids. Independent duplicate sequencing experiment of Fig. 4b. The heights of the bars represent the mutation frequency minus the baseline mutations found in pCon. **c**, Frequency of nucleotide misincorporations in a 15-bp region flanking the lesions present in the indicated plasmids (data from the same sequencing experiment as in **b**). The mutation frequency for pCon is also plotted.

Extended Data Table 1 | Sequence and read numbers for high-throughput sequencing experiments

a

Amplified amplicon	
Experiment 1	AGAACCAATGCATGCGGCCGCGAAGACAGCCCTCTTCCGCTCTT CTTTCGTGCGCGGCCGCGATCCGCTGCATTAATGAAT
Experiment 2	CTCGAGCGGAAGTGCAGAACCAATGCATGCGGCCGCGAAGACA GCCCTCTTCCGCTCTTCTTTCGTGCGCG GCCGCGATCCGCTGCA TTAATGAATCGGCCAACGC GCGGGGAGAGGCGGTTTGCGTATT

b

Sample	Total reads	Pair-matched reads	Reads with indels	Reads with substitutions	Perfect match reads
pICL-Pt	8886804	7928127	2649263	249138	5029726
pICL-AA	14852974	13541980	2396503	1458825	9686652
pPdG	10122605	9331173	69747	565567	8695859
pCon	10449164	9644715	6466	240998	9397251
pICL-AA + p97i	32164058	29539112	491796	1600097	27447219
pICL-Pt-NR	2571879	2385136	102534	58870	2223732
pICL-AA-NR	7316392	6729595	170216	200250	6359129
pPdG-NR	8857493	8137089	52886	270870	7813333

c

Sample	Total reads	Pair-matched reads	Reads with indels	Reads with substitutions	Perfect match reads
pICL-AA	13877258	4566626	388256	538491	3639879
pPdG	15334790	6649408	32855	585454	6031099
pCon	15403988	6714207	7772	171840	6534595
pICL-AA + p97i	15455175	5803423	75922	284148	5443353

a, Amplicon sequence for sequencing experiments 1 and 2. **b**, Total and specific read numbers for sequencing experiment 1 (see Fig. 4). **c**, Total and specific read numbers for sequencing experiment 2 (see Extended Data Fig. 7).

nature research

Corresponding author(s): KJ Patel, Puck Knipscheer

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☒	☐	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated
☐	☒	Clearly defined error bars *State explicitly what error bars represent (e.g. SD, SE, CI)*

Our web collection on statistics for biologists may be useful.

Software and code

Policy information about availability of computer code

Data collection	Autoradiograms were scanned by Typhoon Trio+ using Typhoon scanner control v5.0 as interface. Immunoblots were imaged using Amersham Imager 600 and ImageQuant LAS400. For the clonogenic survival assay colonies were quantified using GelCount colony counter (Oxford Optronix, v.1.2.4.2). For MALDI experiments data was acquired using flexControl (v. 3.3).
Data analysis	The custom algorithms used for sequencing data analysis are described in the methods section and all codes are available upon request. For demultiplexing sequencing data, FASTX_toolkit v0.0.14 and cutadapt v1.9.1 were used (see Methods section). For aligning reads to the amplicon BLAST suite, v2.4.0+ was used. Autoradiograms and immunoblots were analyzed using ImageJ version 1.52 and ImageQuantTL version v8.1.0.0. LD50 values of cell survival curves were calculated in Prism 8. For MALDI experiments data was analysed using flexAnalysis (v. 3.3) and was plotted in Prism 8.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The datasets generated in the current study and the custom code used for analysis are available from the corresponding author on reasonable request.

Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/authors/policies/ReportingSummary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Sample sizes were chosen based on previous experience and published studies to evaluate reproducibility of assays. For quantified results, the number of replicates is indicated in the Figure legends.
Data exclusions	Data exclusion during sequencing data analysis is described in the method section
Replication	Cellular studies were performed with a minimum of three independent experiments. All experiments using Xenopus egg extracts were reproducible and repeated at least twice. For all experiments that show a quantification graph, triplicates are presented in the extended data figure.
Randomization	Randomization is not relevant to this study. For experiments in Xenopus egg extracts, we subjected common extracts to the treatments indicated in the text. Cellular studies were conducted on cells from a common pool that were treated with increasing amounts of DNA-damaging agent.
Blinding	Blinding is not relevant to the experiments performed in this study.

Reporting for specific materials, systems and methods

Materials & experimental systems

n/a	Involved in the study
☐	☒ Unique biological materials
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology
☐	☒ Animals and other organisms
☒	☐ Human research participants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Unique biological materials

Policy information about availability of materials

Obtaining unique materials	All plasmids generated in this study are available upon reasonable request.

Antibodies

Antibodies used	Antibodies against xlFANCD2, xlREV1, xlREV7 were previously described (12,14,19). The xlNEIL3 antibody was raised against a C-terminal peptide of Xenopus laevis NEIL3 and affinity purified (New England peptide).

Validation	The xlFANCD2, xlREV1, xlREV7 antibodies were previously validated in Xenopus egg extract (12,14,19). The Neil3 antibody was validated by western blot, immunodepletion, and rescue in Xenopus egg extract.

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	WT and NEIL3 deficient HAP1 cells were obtained directly from Horizon Discovery.
Authentication	The parental cell line was not authenticated. Gene disruptions were confirmed by PCR and/or western blotting
Mycoplasma contamination	All cell lines were confirmed mycoplasma negative.
Commonly misidentified lines (See ICLAC register)	No misidentified cell line was used in this study.

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals	Female Xenopus laevis frogs used in this study were older than 2 years and obtained from Nasco
Wild animals	The study did not involve wild animals.
Field-collected samples	This study did not involve field-collected samples.