Stony Brook University

MULTIPLE DNA BINDING DOMAINS MEDIATE THE FUNCTION OF ERCC1-XPF IN NUCLEOTIDE EXCISION REPAIR AND INTERSTRAND CROSSLINK REPAIR

A Dissertation Presented

by

Yan Su

to

The Graduate School

in Partial Fulfillment of the

Requirements

for the Degree of

Doctor of Philosophy

in

Chemistry

Stony Brook University

December 2012

Stony Brook University

The Graduate School

Yan Su

We, the dissertation committee for the above candidate for the

Doctor of Philosophy degree, hereby recommend

acceptance of this dissertation.

Orlando D. Schärer, PhD – Dissertation Advisor
Professor, Department of Pharmacological Sciences

Isaac Carrico, PhD - Chairperson of Defense
Assistant Professor, Department of Chemistry

Carlos Simmerling, PhD
Professor, Department of Chemistry

Bruce Demple, PhD
Professor, Department of Pharmacological Sciences, Stony Brook University

This dissertation is accepted by the Graduate School

Charles Taber
Interim Dean of the Graduate School

Abstract of the Dissertation

MULTIPLE DNA BINDING DOMAINS MEDIATE THE FUNCTION OF ERCC1-XPF IN NUCLEOTIDE EXCISION REPAIR AND INTERSTRAND CROSSLINK REPAIR

by

Yan Su

Doctor of Philosophy

in

Chemistry

Stony Brook University

2012

Structure-specific endonucleases are widespread enzymes that incise phosphodiester bonds in DNA and play key roles in DNA transactions including the maintenance of genome stability. One such enzyme is ERCC1-XPF, which cleaves single-stranded/double-stranded DNA junctions, and is involved in nucleotide excision repair (NER), interstrand crosslink (ICL) repair, homologous recombination, and other pathways. Defects in ERCC1-XPF lead to several hereditary diseases, like xeroderma pigmentosum, XFE progeroid syndrome, cranio-occulo-facio-skeletal syndrome, and Fanconi anemia.

In NER, ERCC1-XPF is recruited to DNA lesions by interaction with XPA and incises the DNA 5' to the lesion. In this dissertation, the roles of the four C-terminal DNA binding domains in mediating NER activity and cleavage of model substrates were studied. It was found that mutations in the helix-hairpin-helix domain of ERCC1 and the nuclease domain of XPF abolished cleavage activity on model substrates. Interestingly, mutations in multiple DNA binding domains were required to significantly diminish NER activity, suggesting that interactions with proteins in the NER incision complex can compensate for some defects in DNA binding. These studies demonstrate that multiple domains of ERCC1-XPF contribute to substrate binding, and are consistent with the hypothesis that multiple weak protein-DNA and protein-protein interactions drive progression through the NER pathway.

In addition, DNA binding residues of ERCC1-XPF were found to be more important for ICL repair than for NER. Impairment of DNA binding of ERCC1-XPF rendered cells more sensitive to exposure to crosslinking agents than to UV, suggesting tighter substrate binding by

ERCC1-XPF is needed for ICL repair than for NER. Intriguingly, subsequent to our studies, mutations in DNA binding residues of XPF were found in Fanconi anemia patients, who have defects in ICL repair, showing that mutations in DNA binding domains of ERCC1-XPF can have pathway-specific pathogenic consequences in humans.

To My Parents

Table of Contents

LIST OF FIGURES/TABLES VIII

LIST OF ABBREVIATIONS XI

ACKNOWLEDGMENTS XIII

CHAPTER 1 GENERAL INTRODUCTION: AN OVERVIEW OF ERCC1-XPF 1

INTRODUCTION 2

THE HOMOLOGS OF ERCC1-XPF 3

STRUCTURES AND BIOCHEMICAL PROPERTIES OF THE DOMAINS OF ERCC1-XPF 5

BIOLOGICAL FUNCTIONS OF ERCC1-XPF AND ITS HOMOLOGS 8

PHYSIOLOGICAL FUNCTIONS OF ERCC1-XPF 17

PREVIEW 25

CHAPTER 2 MULTIPLE DNA BINDING DOMAINS MEDIATE THE FUNCTION OF ERCC1-XPF IN NUCLEOTIDE EXCISION REPAIR 27

ABSTRACT 28

INTRODUCTION 29

EXPERIMENTAL PROCEDURES 32

RESULTS 37

DISCUSSION 49

SUPPLEMENTARY METHODS AND MATERIALS 54

ACKNOWLEDGEMENTS 57

CHAPTER 3 XPF MUTATIONS SEVERELY DISRUPTING DNA INTERSTRAND CROSS-LINK REPAIR CAUSE FANCONI ANEMIA 58

ABSTRACT 59

SUPPLEMENTARY 76

ACKNOWLEDGEMENTS 99

CHAPTER 4 DNA BINDING RESIDUES NEAR THE ACTIVE SITE OF ERCC1-XPF MEDIATE ITS FUNCTIONS IN NER AND ICL REPAIR 100

ABSTRACT 101

INTRODUCTION 102

EXPERIMENTAL PROCEDURES 103

RESULTS 106

DISCUSSION 111

ACKNOWLEDGEMENTS 112

REFERENCE 113

LIST OF FIGURES/TABLES

CHAPTER 1

FIGURE 1. REPRESENTATIVES OF THE XPF/MUS81 FAMILY 4

FIGURE 2. STRUCTURES OF DOMAINS IN ERCC1-XPF OR ITS HOMOLOGS 7

FIGURE 3. A MODEL OF NER REPAIR PATHWAY 10

FIGURE 4. A MODEL OF ICL REPAIR PATHWAY 13

FIGURE 5. A MODEL OF SSA PATHWAY IN YEAST 15

CHAPTER 2

FIGURE 1. STRUCTURE OF THE ERCC1-XPF HHH DOMAINS AND SCHEME OF THE DNA BINDING DOMAINS IN ERCC1-XPF. 30

FIGURE 2. MUTATIONS IN THE HHH DOMAIN OF ERCC1 AND THE NUCLEASE DOMAIN OF XPF IMPACT NUCLEASE ACTIVITY ON A STEM-LOOP SUBSTRATE. ... 39

FIGURE 3. MUTATIONS IN THE HHH DOMAIN OF ERCC1 LEAD TO REDUCED DNA BINDING AFFINITY. 41

FIGURE 4. MUTATIONS IN MULTIPLE DOMAINS ARE NEEDED TO AFFECT NER IN VITRO. 43

FIGURE 5. MUTATIONS IN DNA BINDING DOMAINS OF ERCC1-XPF DO NOT AFFECT THE RECRUITMENT TO SITES OF UV DAMAGE. 45

FIGURE 6. MUTATIONS IN DNA BINDING DOMAINS OF ERCC1-XPF AFFECT REPAIR KINETICS OF (6-4)PPS IN VIVO. 47

FIGURE 7. MUTATIONS IN DNA BINDING RESIDUES RENDER CELLS MORE SENSITIVE TO MMC THAN TO UV. 49

FIGURE S1. ERCC1-XPF PROTEINS WITH MUTATIONS IN THE HHH, NUCLEASE AND CENTRAL DOMAINS ARE PROPERLY FOLDED. 55
FIGURE S2. EXPRESSION LEVEL OF ERCC1 AND XPF PROTEIN IN TRANSDUCED UV20 AND XP2YO CELLS. 56

CHAPTER 3

FIGURE 1. XPF MUTATIONS AND XPF-DEFICIENCY IN FANCONI ANEMIA PATIENTS. 63
FIGURE 2. UV AND MMC SENSITIVITIES OF XPF MUTANTS LEADING TO FANCONI ANEMIA. 66
FIGURE 3. NER ANALYSIS OF XPF MUTANTS IN VIVO 68
FIGURE 4. EXCISION, NUCLEASE AND UNHOOKING ACTIVITIES OF XPF MUTANTS. 72
FIGURE 5. XPF MUTATIONS LEAD TO THREE RARE DNA DAMAGE RESPONSE SYNDROMES. 75
SUPPLEMENTARY FIG 1. DIAGNOSTIC CONFIRMATION OF FA BY THE CHROMOSOME FRAGILITY TEST. 89
SUPPLEMENTARY FIG 2. EVOLUTIONARY CONSERVATION OF XPF RESIDUES WITH MISSENSE MUTATIONS IN FA PATIENTS. 90
SUPPLEMENTARY FIG 3. ANALYSIS OF PARENTAL DNAS AND REVERTED CELL LINES. 92
SUPPLEMENTARY FIG 4. XPF CELLULAR LOCALIZATION 93

SUPPLEMENTARY FIG 5. SLX4 AND ERCC1 INTERACTIONS IN XPF-DEFICIENT FA PATIENTS 94
SUPPLEMENTARY TABLE 1. LIST OF CANDIDATE GENES AFTER WHOLE EXOME SEQUENCING WHEN ASSUMING A RECESSIVE INHERITANCE MODEL. 95
SUPPLEMENTARY TABLE 2. COMPARATIVE SUMMARY OF CLINICAL AND CELLULAR/MOLECULAR FEATURES OF XP, XFE AND FA PATIENTS WITH MUTATIONS IN XPF. 97
SUPPLEMENTARY TABLE 3. PCR PRIMERS USED FOR SANGER SEQUENCING OF XPF 98

CHAPTER 4

FIGURE 1. TWO WIDELY USED WIDE TYPES OF XPF HAVE THE SAME ACTIVITIES IN NER. 107
FIGURE 2. THE RECRUITMENT OF ERCC1-XPF DOES NOT DEPEND ON ITS DNA BINDING ABILITY. 108
FIGURE 3. MULTIPLE DNA BINDING RESIDUES NEAR THE ACTIVE SITE SYNERGISTICALLY AFFECT NER ACTIVITY 109
FIGURE 4. DETECTION OF THE UNHOOKING STEP OF THE ICL REPAIR BY AKLINE COMET ASSAY. 110

LIST OF ABBREVIATIONS

ERCC: excision repair cross-complementing

XP: xeroderma pigmentosum

NER: nucleotide excision repair

ICL, interstrand crosslink

HR: homologous recombination

FA: Fanconi anemia

HhH, helix-hairpin-helix

ID: FANCD2-FANCI

SF2: superfamily II

HJ: Holiday junction

HLD, helicase like domain

dsDNA: double-stranded DNA

ssDNA: single-stranded DNA

NLS: nuclear localization signal

GG-NER: global genome nucleotide excision repair

TC-NER: transcription coupled nucleotide excision repair

NHEJ: non-homologous end joining

DSB: Double strand break repair

MMEJ: microhomology-mediated end joining

SSA: single strand annealing

Top1: Topoisomerase I

XFE: XPF-ERCC1

UDS: unscheduled DNA synthesis

COFS: Cranio-occulo-facio-skeletal syndrome

MEF: mouse embryonic fibroblast

SNP: single nucleotide polymorphisms

NSCLC: non-small cell lung carcinoma

mRNA: messenger RNA

IHC: immunohistochemistry

HA: hemaglutinin

MMC: mitomycin C

cis-Pt: cisplatin

AAF: acetyl aminofluorene

(6-4)PP: 6-4 photoproduct

CPD: cyclobutane pyrimidine dimmer

BMF: bone marrow failure

ACKNOWLEDGMENTS

First of all, I would like to thank my advisor Dr. Orlando D. Schärer for allowing me to start this exciting project and sharing his wisdom in science with me over the years. His support, guidance, and encouragement largely helped me to accomplish my work in this dissertation. He also provided me many opportunities to develop professionally and to collaborate with other scientists. I sincerely appreciate his help, advice and patience, especially in correcting my writing.

I would like to thank my thesis committee members Dr. Isaac Carrico and Dr. Carlos Simmerling for their time and help during the years. I would also like to thank Dr. Bruce Demple for being my outside member.

To accomplish my work, I spent a couple of weeks with Dr. Jill O. Fuss in Lawrence Berkeley National Laboratory; I also conducted experiments in Dr. Bevin Engelward's lab in MIT, mentored by Patrizia Mazzucato. I would like to thank them for all their help. I learned a lot from these talented and smart people.

I would like to thank our collaborators: Dr. Advaitha Madireddy and Dr. Laura J. Niedernhofer for their collaboration for the work in Chapter 2; Dr. Jordi Surralles for inviting our group to participate in the XPF- FA project in Chapter 3.

Thank you to all Schärer group members for creating a great working and academic atmosphere in the lab. In particular, Barbara kindly mentored me when I first joined the lab. Vinh and Banke gave me advices in science and bench techniques during my beginning time in the lab. Thanks to Jérôme, Jung-Eun, AJ, Burak, Shivam, Andy, Alejandra, Julie and all the high school and rotation students for all the help and friendship.

At last, I would like to say thank you to my beloved parents for their unconditional love and support. I love you both!

CHAPTER 1

GENERAL INTRODUCTION:

AN OVERVIEW OF ERCC1-XPF

Adapted from the manuscript by Yan Su, Nikhil R. Bhagwat, Laura J. Niedernhofer and Orlando D. Schärer, submitted in the near future.

INTRODUCTION

ERCC1-XPF is a structure-specific endonuclease initially discovered for its role in nucleotide excision repair (NER), a DNA repair pathway that removes helix distorting lesions formed by UV light and environmental mutagens from the nuclear genome (1,2). Consistent with this finding, people with mutations in the XPF gene suffer from xeroderma pigmentosum (XP), an inherited DNA repair disorder characterized by extreme sensitivity to UV light and more than 2000-fold increased incidence of skin cancer. Subsequently, ERCC1-XPF was shown to function in a number of additional pathways, including the repair of interstrand crosslinks (ICLs), homologous recombination (HR) and telomere maintenance. Consequently, mice deficient in ERCC1-XPF and two identified human patients not only suffer from cancer predisposition, but also show developmental abnormalities and symptoms of accelerated aging (3,4). Recently, two unrelated patients were reported to show the typical Fanconi anemia (FA) characteristics: bone marrow failure and hypersensitivity to interstrand cross-linking agents. Here I review the biochemical, structural, cell biological and genetic basis of the various functions of ERCC1-XPF, and I aim to connect these functions to diverse phenotypes caused by ERCC1-XPF deficiency.

History of ERCC1-XPF

In order to find the genetic defects of XP patients, cloned human genes were complementarily transfected into mutant rodent cell lines (1,5). Excision repair cross-complementing group 1 (ERCC1) is the first cloned human DNA repair gene (5), although no XP patient has been found to have defects in ERCC1 so far. The XPF gene was discovered by

the same method, and the expressed protein was shown to form a heterodimer with ERCC1 (1,6-8).

THE HOMOLOGS OF ERCC1-XPF

ERCC1-XPF is a member of the XPF/Mus81 family. Proteins in this family share several key features: They form dimers though helix-hairpin-helix (HhH) domains, contain a conserved (V/I)ERKX_3D nuclease motif in the active site, and have a domain resembling the superfamily 2 (SF2) RNA helicases. In the case of specific proteins, some of these features may be physically or functionally lost during the evolutionary process (9) (**Figure 1**).

In higher eukaryotes, the XPF/MUS81 family has multiple members that specialize in different DNA repair pathways. In humans, ERCC1-XPF has three homologs: endonucleases EME1-MUS81 and EME2-MUS81, and translocase FAAP24-FANCM. The two nucleases show specificity for flap structures, replication forks, and nicked Holliday junctions (HJs) with the active site located in MUS81 (9-12). EME1 and EME2 share the similar sequences, and they do not have the crucial residues at the nuclease active site (13-15). EME1-MUS81 is a HJ resolvase, involved in chromosome segregation during meiosis and resolution of HJ in DSB repair. It shares functions with another HJ resolvase, GEN1, in chromosome segregation (16,17). FANCM contains the ATP binding site in the helicase-like domain but lacks the key residues for nuclease activity. Therefore, in contrast to XPF, it contains ATPase and DNA translocate activity, but no nuclease activity (15,18). FAAP24-FANCM is a member of Fanconi anemia core complex, and

it is therefore crucial for monoubiquitination of FANCD2-FANCI (ID) complex. The translocase is also associated with the Bloom helicase complex in ICL repair (11,18-20) (**Figure 1A**).

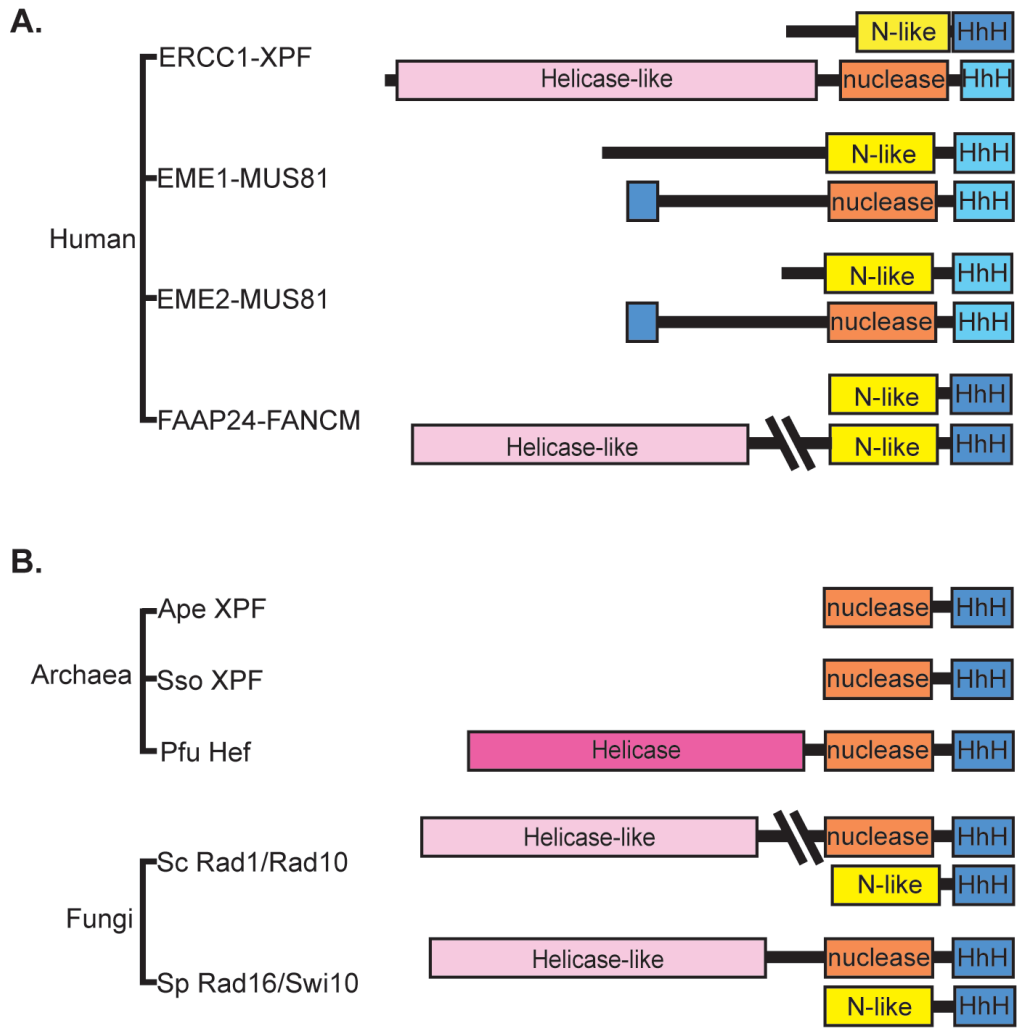


Figure 1. Representatives of the XPF/MUS81 family A. ERCC1-XPF and its homologs in human. **B.** Some homologs of ERCC1-XPF in archaea and yeast. The homologs in archaea are homodimers, while most of the family members in yeast or human are heterodimers. The conserved (V/I)ERKX3D nuclease motif in the active site is located in nuclease domains (orange). Nuclease-like (N-like) domains (yellow) share structural similarities with nuclease domains but have lost the conserved active residues. Helix-hairpin-helix (HhH) motifs are all canonical in domains colored by dark blue, while the domains in light blue have some interrupted HhH motifs. The helicase domain (dark pink) of PfuHef belongs to the helicase superfamily II (SF2). However, most helicase like domains (HLDs) in the XPF/MUS81 family

lack both ATP binding residues and helicase activities. One exception is FAAP24-FANCM, which has ATPase activity but acts as a translocase instead of a helicase.

Studies of ERCC1-XPF homologs in archaea or lower eukaryotes provide large amounts of valuable information for understanding the biochemical and structural properties of the human counterpart. Although there are no known homologs in bacteria (21), XPF is well represented in archaea by three homodimers: PfuHef (in *Pyrococcus furiosus*), ApeXPF (in *Aeropyrum pernix*) and SsoXPF (in *Sulfolobus solfataricus*). All of them have nuclease motifs and HhH domains. PhuHef contains a helicase like domain (HLD) with functionally helicase activity, while the later two do not have HLD. These three proteins have been studied in great detail (Figure 1B) (22-25). Heterodimer Rad10-Rad1 in *Saccharomyces cerevisiae* and Swi10-Rad16 in *Schizosaccharomyces pombe* are the widely studied homologs of ERCC1-XPF in yeast (26-28). Like XPF, Rad1 and Rad16 have the nuclease motif, HhH domain, and the HLD without ATPase acitivity, while Rad10 and Swi10 are similar to ERCC1, containing only the nuclease-like domain lacking the active site and HhH domain (**Figure 1B**). In the section below, we discuss how structural and functional studies of homologs helped in understanding the function of human ERCC1-XPF.

STRUCTURES AND BIOCHEMICAL PROPERTIES OF THE DOMAINS OF ERCC1-XPF

The C-terminal helix-hairpin-helix domains are involved in dimerization and DNA binding

ERCC1 shares a common ancestor with XPF (29). Therefore, the two exhibit structural similarities at the C-terminus despite diverse functions (**Figure 2**). They both have HhH motifs at C-terminus. Structural studies show that the two HhH domains contribute to the dimerization of ERCC1-XPF through hydrophobic and hydrogen bonding interactions(30,31) (Figure 2A). ERCC1 contains canonical double HhH motifs, while the hairpin in the second HhH motif of XPF is replaced by a beta-turn (30,31). In the arecheal homolog of XPF, the two HhH motifs interact with the minor groove of double-stranded DNA (dsDNA) (32,33), consistent with the canonical role of HhH domain in DNA binding (34). The HhH domain of human XPF can form a homodimer and bind single-stranded DNA (ssDNA) (30,35,36). Furthermore, in a recent study of the full-length ERCC1-XPF, ERCC1 HhH domain plays a more important role in DNA binding than XPF HhH domain (37). Therefore, it is proposed that the HhH domain of ERCC1 binds to the dsDNA part of a ss/dsDNA junction, while XPF HhH domain interacts with the ssDNA (**Figure 2**).

The nuclease active site is located in XPF

ERCC1-XPF belongs to the type II restriction enzyme family, and the active site is located in nuclease domain of XPF (38). Structural studies of archeal XPF homologs show that the active site is made up of a V-shape groove containing the two conserved motifs, VDXRE and ERKX$_3$D (22,38). These residues are involved in metal ion coordination (24). A Mn^{2+} or Mg^{2+} ion required for cleavage (1,38,39) (**Figure 2**).

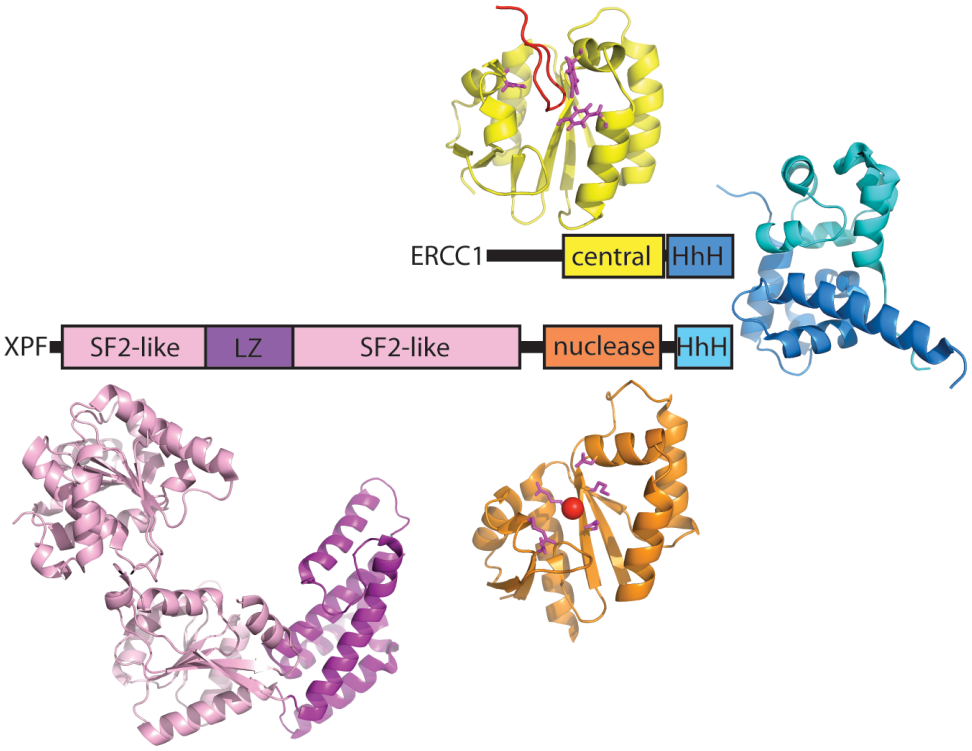


Figure 2. Structures of domains in ERCC1-XPF or its homologs HhH domains of ERCC1 (dark blue) and XPF (light blue) contribute to dimerization of the protein. The active site of ERCC1-XPF is located in the nuclease domain (orange) of XPF. This particular nuclease domain structure is from PfuHef, a XPF homolog. The active site residues (purple) can coordinate with a metal ion (red). The central domain of ERCC1 (yellow) is structurally similar to the nuclease domain and interacts with one peptide (red) of XPA in NER. Key residues involved in this interaction are shown in purple. The structure of HLD is from PfuHef and contains two RecA subdomains (pink) and a thumb-like subdomain (purple).

The central domain of ERCC1 is structurally related to the nuclease domain of XPF

The central domain, located in ERCC1, is structurally similar to nuclease domain of XPF (29,31,40). However, instead of the metal binding motif, a V-shape groove of central domain is populated with aromatic and basic residues that interact with XPA (31,40,41). This XPA-central domain interaction is crucial in guiding ERCC1 to the pre-incision complex in NER. Deficiency of this interaction leads to a defect in NER (41,42). The central domain of ERCC1 has also been

implicated in DNA binding, but the exact location of this binding domain and its importance in mediating ERCC1-XPF function remains to be established (**Figure 2**).

The helicase-like domain of XPF has lost the ability to bind ATP

The N-terminal HLD of XPF has no counterpart in ERCC1, and is likely involved in stabilizing the whole protein complex, DNA binding, nuclear localization and interactions with other proteins. Without the HLD, the other four domains of XPF-ERCC1 become fragile, but can still cut ss/ds DNA at a lower efficiency (31). The analogous domain in Hef belongs to the superfamily II (SF2) helicase, and contains two RecA subdomains and a thumb-like subdomain (25). However, HLD of ERCC1-XPF lacks the ATP binding site, and is therefore devoid of helicase activity (23). The RecA-like subdomains are likely to interact with DNA as has been shown for other SF2 family helicases. Interestingly, crystallographic data indicate that a SF2 helicase, Hel308, can separate 2 base pairs by the energy released from the protein-DNA interaction without ATP hydrolysis (43). Since ERCC1-XPF incises the ss/ds DNA junction in the double stranded part (1,39), it is possible that the HLD opens up 2 base pairs before the nuclease domain conducts the incision. HLD also contains nuclear localization signals (NLSs) to transport ERCC1-XPF into cell nuclei (44). Recently, the HLD has been shown to be involved in the interaction with SLX4, a scaffold protein for nucleases, which is believed to drive ERCC1-XPF to the ICL repair and HR (45-48) (**Figure 2**).

BIOLOGICAL FUNCTIONS OF ERCC1-XPF AND ITS HOMOLOGS

Nucleotide excision repair

NER repairs a wide variety of helix-distorting DNA lesions formed by UV light, environmental mutagens and cancer chemotherapeutic agents. It involves over 30 proteins that perform a highly coordinated excision of a single-stranded oligonucleotide containing the damage and restoration of the original DNA sequence using the non-damaged strand as a template (2,49).

NER consists of two pathways: global genome NER (GG-NER) and transcription-coupled NER (TC-NER)(50). As the names suggest, GG-NER deals with DNA damage throughout the genome, whereas TC-NER allows the cell to preferentially repair the transcribed strand of genes. TC-NER is initiated by a stalled RNA PolII at a lesion (51,52) and also requires the CSA, CSB and XAB2 proteins (53-55). GG-NER by contrast is initiated by XPC-RAD23B, which detects the helical distortion induced by lesions in DNA (56-58), in some cases with the help of the UV-DDB protein complex (59). Following damage recognition, the two pathways are believed to operate by the sequential recruitment and assembly of the same downstream proteins (57,60,61). After damage recognition, the transcription and NER factor TFIIH is recruited to the site of damage to open up DNA and verify the presence of the lesion by one of its helicase subunits, XPD (62-64). Engagement of TFIIH with the lesion provides a landing platform for the next set of NER factors, XPA, RPA and XPG (65,66). RPA, the ssDNA binding protein, is believed to bind the non-damaged strand of DNA to help position the two endonucleases ERCC1-XPF and XPG, and protect the gap during repair synthesis (67,68). XPG has a structural role in the assembly of the NER pre-incision complex and later makes the incision 3' to the lesion (69,70). XPA interacts with TFIIH, RPA and DNA, and recruits ERCC1-XPF through interaction with the central domain of ERCC1 (71-74). In addition to a core folded region, XPA

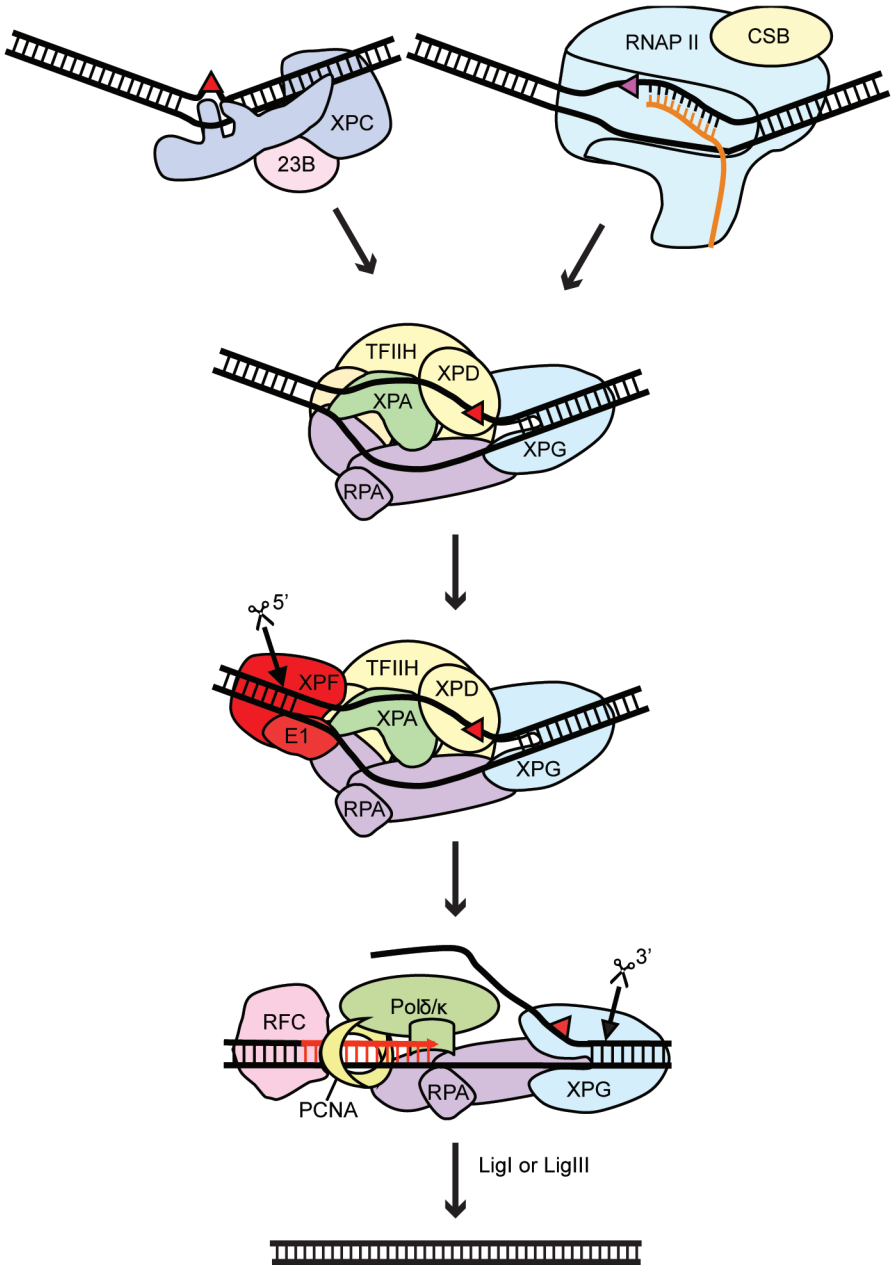


Figure 3. A model of NER repair pathway NER is initiated by a stalled RNA PolII at a DNA lesion during transcription (TC NER, on the right), or, alternatively, by the recognition of a DNA distortion by XPC-RAD23B (GG NER, on the left). TFIIH is subsequently recruited to open up the DNA, verify the presence of the lesion and further recruit XPG, RPA, and XPA. ERCC1-XPF is the last factor to form the pre-incision complex. ERCC1-XPF then cuts 5' to the lesion, which is followed by initiation of DNA synthesis, 3' incision by XPG, completion of DNA synthesis, and ligation.

has unstructured regions. One of these, encompassing a GGGF motif, undergoes a disordered to ordered transition upon binding to the central domain of ERCC1. This interaction is necessary and sufficient for the recruitment of ERCC1-XPF to damaged sites in DNA (40-42). Dual incision and repair synthesis are not a simple matter of cutting and pasting, but a tightly coordinated process starting with 5' incision by ERCC1-XPF, followed by initiation of repair synthesis, then 3'incision by XPG, and at last completion of repair synthesis and ligation (75). Repair synthesis involves at least three polymerases (Pol δ, Pol ε and Pol κ) and accessory factors. The nick may be sealed by either DNA ligase I or ligase III/XRCC1 (76,77) (**Figure 3**).

Interstrand crosslink repair

ICLs covalently link two bases on opposite strands of DNA duplex, thereby preventing strand separation for replication or transcription (78). An important role for ERCC1-XPF in ICL repair is suggested by the exquisite sensitivity of ERCC1- or XPF-deficient cells to cross-linking agents, far exceeding that of other cell lines with defects in XP/NER genes (79). The main pathway for ICL repair in humans occurs in S-Phase and is triggered by the stalling of a replication fork at the ICL site. It involves components of many different repair pathways, and one of the earliest discernible steps is the activation of the Fanconi anemia pathway. The FA core complex, consisting FANCA, B, C, E, F, G, L, M and FAAP24, monoubiquitinates FANCD2 and FANCI (ID complex). FANCL serves as the ubiquitin ligase, while the XPF homolog FANCM is a potential ICL sensor via its DNA translocation activity (80-82). Following ubiquitination of ID complex, the cross-linked DNA is unhooked by incisions on either side of the ICL on one strand (83). A number of nucleases, including ERCC1-XPF, EME1-MUS81, FAN1, SLX4-SLX1 and SNM1B, have been implicated in this step, based on sensitivities of

deficient cell lines to cross-linking agents, interactions with ubiquitinated ID complex, or their involvements in DSB formation during ICL repair. Of the four endonucleases, FAN1 and SLX1 share the same polarity that would enable them to cut 5' to the lesions on the lagging strand (**Figure 4**), while MUS81 and XPF share the polarity to cut on the other side of the ICL. It was initially thought that MUS81 makes the 3' incision, as the absence of MUS81, rather than ERCC1, prevents the formation of ICL-induced DSB. However, a recent work suggests that MUS81 forms DSBs only when the ERCC1-XPF pathway fails (84), indicating ERCC1-XPF having a primary role in ICL repair. Along with recent findings that XPF is a FA gene, it is likely that ERCC1-XPF is the primary nuclease for the 3' incision to the lesion at a stalled replication fork (unpublished data, Chapter 3). The nuclease responsible for 5' incision is currently unknown, although evidence points to FAN1 (85-88) or SLX1 (unpublished data). It is possible that ERCC1-XPF acts in cooperation with the exonuclease SNM1A that can trim DNA around an ICL (84,89). Following unhooking of the ICL, the leading strand is extended beyond the ICL remnant by a translesion polymerase, most likely Rev1 and Polζ, leading to the restoration of the leading strand (90). It has been suggested, but not experimentally shown, that this restored leading strand subsequently serves as an intact strand in NER to get rid of the ICL remnant, again involving ERCC1-XPF. Afterwards, HR is conducted to remove the DSB generated during the unhooking step. It has also been suggested that ERCC1-XPF is involved in the HR step, although the details remain to be determined (91,92) (**Figure 4**).

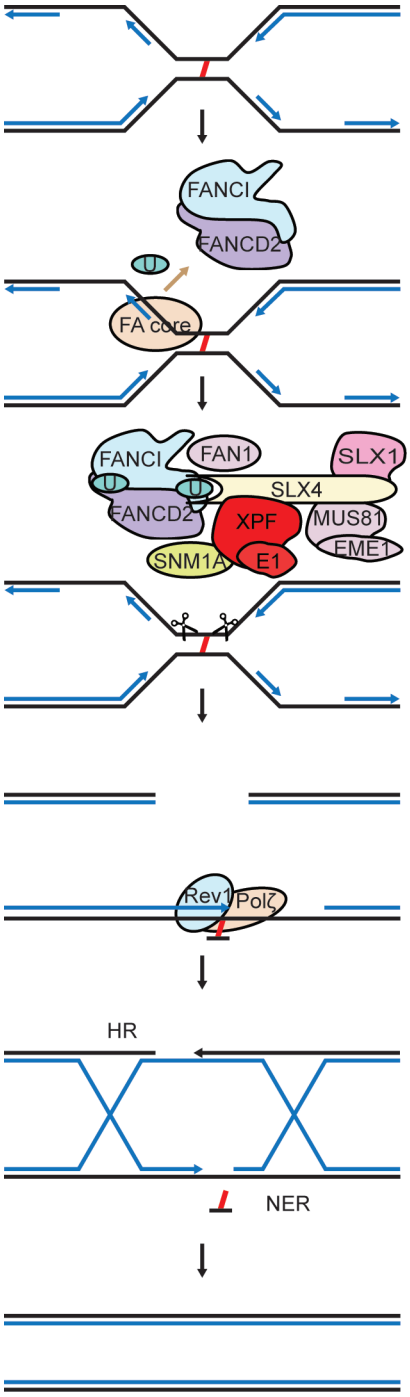


Figure 4. A model of ICL repair pathway The ICL repair pathway is initiatesd by a stalled replication fork at an ICL lesion, thereby triggering the FA pathway in which FANCD2/FANCI complex is monoubiquitinated by the FA core complex. Several endonucleases, including ERCC1-XPF, SLX1, EME1-MUS81, and FAN1, have been proposed to be involved in the unhooking of the lesion from one strand, allowing Rev1 and Polζ to bypass the lesion. Then, NER and HR processes are implicated in completing the repair.

Double strand break repair (DSB)

Like ICLs, DSBs are extremely cytotoxic lesions that are addressed mainly by homologous recombination (HR) and non-homologous end joining (NHEJ) (93,94). The yeast Rad10-Rad1 protein has been shown to participate in DSB repair via single strand annealing (SSA), a subpathway of HR, and microhomology-mediated end joining (MMEJ), a subpathway of NHEJ (93,94). SSA (**Figure 5**) is a Rad52 dependent process (95), and operates on DSBs flanked by pairs of homologous sequences of more than 29bp (96). After the nucleolytic processing, the homologous sequences on both side of the DSB present as ssDNA, and the two repeats can anneal to form 3' overhangs. The overhangs are subsequently cut by Rad10-Rad1 to remove extra nonhomologous sequences (97,98). Interestingly, Msh2 and Msh3 proteins are needed for SSA only when regions of homology are less than 1kb, but not for longer repeats (99)(Lee, EMBO J., in press). Recently, a genetic screen revealed that the Saw1 plays a role in SSA in conjunction with Rad10-Rad1 (100). Targeting of Rad10-Rad1 to sites of SSA requires SAW1, but not SLX4 (Lee, EMBO J., in press), although SLX4 has been implicated in the regulation of Rad10-Rad1 in SSA (100-102). What the relationship of SLX4 and SAW1 is and whether there is a functional equivalent of SAW1 in humans remains to be determined (**Figure 5**). In mammals, the SSA pathway is severely affected in ERCC1 deficient CHO cells, suggesting the involvement of ERCC1-XPF in SSA (103-105). In support of this notion, ERCC1-XPF has been shown to interact directly with RAD52 in human (106). In addition, in mouse ES cells, ERCC1-XPF has been shown to be crucial in special HR class: the targeted gene replacement (107).

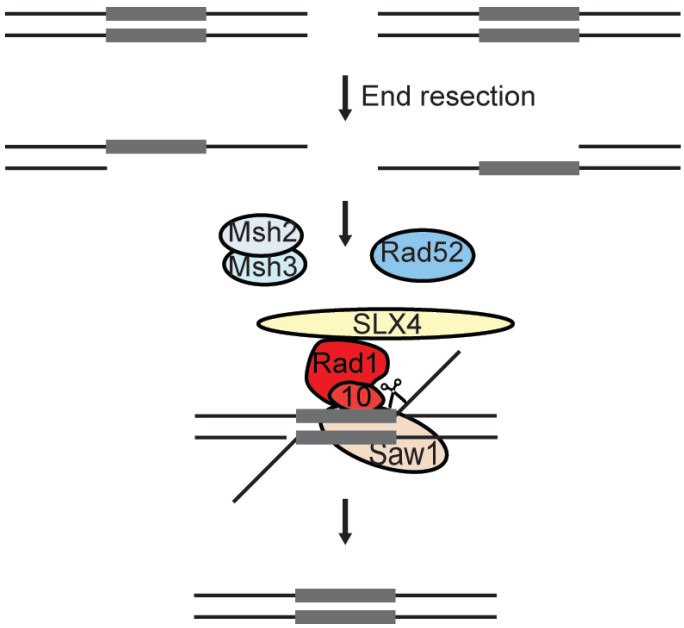


Figure 5. A model of SSA pathway in yeast SSA occurs when two homologous sequences flank a DSB. With the help of Rad52 and sometimes Msh2/Msh3, single-strand resection generates the two complementary strands of DNA that can anneal to each other. Saw1 has been recently discovered to recruit Rad10-Rad1 to the complex in order to remove 3' flapped fragments. SLX4 may also play a role in this process.

The principle of MMEJ is similar to SSA pathway, but with much shorter homologous sequences (108,109). The Mre11-Rad50-Xrs2 complex works as exo/ss nuclease to generate the ssDNA for annealing. This process is independent of Ku86 and Rad52. Rad10-Rad1 is implicated to incise 3' overhang DNA in MMEJ (109). Results of studies of ERCC1-deficient MEFs support a role for ERCC1-XPF in MMEJ in mammals (110,111).

Meiotic recombination

Meiotic recombination is a specialized HR mechanism between homologous chromosomes with divergent sequences, allowing for the generation of genetic diversity. Studies show that mutations in MEI-9, the XPF homolog in *Drosophila*, cause decrease in crossovers,

indicating a role for MEI-9 in the resolution of Holliday junctions during meiosis (112,113). Interestingly, this function of MEI-9 is dependent on an interaction with MUS312, the homolog of ICL repair protein SLX4 (46,48,114).

A fertility defect of both male and female Ercc1-mutant mice also indicates a role of ERCC1-XPF in meiosis. Since ERCC1-XPF is important at all stages of development and maturation of both male and female gametes, it is unclear whether this is due to a specific defect in resolving meiotic crossovers (115).

Telomere maintenance

ERCC1-XPF is a component of the TRF2 complex, a key regulator of telomere maintenance (116). ERCC1-XPF does not appear to have a constitutive role at telomeres, and absence of it alone does not lead to telomere shortening. The function of ERCC1-XPF is revealed when TRF2 levels are abnormal. In the absence of TRF2, ERCC1-XPF cleaves 3' overhangs with G-rich sequences present at telomeres, resulting in telomere fusions and the formation of dicentric chromosomes by NHEJ (116). By contrast, overexpression of TRF2 leads to loss of telomeres. This process is dependent on the physical presence, but not the nuclease activity of ERCC1-XPF (117,118). The mechanisms by which ERCC1-XPF is involved in telomere maintenance remain to be further established.

Repair of by Topoisomerase I induced damages and abasic sites

In addition to its role in NER, ICL repair, DSB repair and telomere maintenance, ERCC1-XPF has been implicated in other pathways via its ability to cleave ss/ds DNA junctions. In *S. cerevisiae*, the inhibition of Topoisomerase I (Top1) by camptothecin results in the

blockade of the Top1-DNA intermediate (119), and the subsequent recruitment of the Tdp1 dependent pathway (120). In the absence of Tdp1, Rad10-Rad1 functions as a backup pathway (121,122), likely removing the flapped Top1-DNA intermediate and initiating Rad52-dependent homologous recombination pathway (122).

Rad10-Rad1 also likely plays a back-up role in the repair of abasic sites. This is supported by the observation that deletion of Rad1, together with the two AP endonucleases *apn1* and *apn2*, causes synthetic lethality. This particular function of Rad10-Rad1 is independent of NER, as deletion of Rad14 (the *S. cerevisiae* XPA homolog) does not have the same effect (123). These observations suggest that Rad10-Rad1 has a role in removing bases with oxidized sugars from the 3' end of ssDNA breaks in the absence of *apn1* and *apn2*. In line with these findings, XPF-deficient and ERCC1-deficient CHO cells are sensitive to H_2O_2, and ERCC1-XPF has the ability to make incisions at 3' blocked ends (124). In addition, ERCC1-XPF/Rad10-Rad1 may have a back-up function in the repair of abasic sites through NER, although yeast genetic studies suggest that this role is less important than the trimming of the 3' blocked ends (125-127).

PHYSIOLOGICAL FUNCTIONS OF ERCC1-XPF

a. HUMAN DISEASES ASSOCIATED WITH DEFECTS IN ERCC1-XPF

ERCC1-XPF is involved in multiple genome maintenance mechanisms, and its defects are associated with human diseases. The severity and manifestations vary significantly, depending on the specific mutant alleles of ERCC1-XPF.

Xeroderma pigmentosum (XP)

Xeroderma pigmentosum (XP) is a hereditary disorder associated with defects in DNA repair, and was linked to NER in 1968 (128). XP is an autosomal recessive syndrome, characterized by extreme sun sensitivity and a 2000-fold increase in skin cancers in sun-exposed areas (129). XP has eight complementation groups, XP-A through G and a variant form, XP-V. XP-F, first reported in Japan, usually presents with a mild phenotype with late onset compared to the other XP groups (130). In some cases, a late onset neurodegenerative phenotype is observed. The XPF protein in XP-F patients fails to localize properly to the nucleus, and mislocalization of ERCC1-XPF to the cytoplasm is believed to be responsible for the NER defect (44). No XP patient with a defect in ERCC1 has been found so far.

XFE progeroid syndrome

A second disease associated with mutations in XPF is XFE (XPF-ERCC1) progeroid syndrome. Only one patient with XFE syndrome has been described so far. The patient presented with a progeriod phenoptype in addition to sun sensitivity (3). Despite photosensitivity and abnormal skin pigmentation, this patient did not develop skin cancer, but presented with an array of aging-related degenerative changes and growth retardation including short stature, microcephaly and developmental delay. His birth-weight was normal, but his skin was dry, atrophic and short of subcutaneous fat. He had an aged appearance with bird-like facies. Neurodegeneration was evident from cerebral atrophy, tremors and ataxia with loss of coordination, impaired hearing and vision loss with optic atrophy. Scoliosis, muscle wasting and hypotonia suggested musculoskeletal degeneration. In addition, hypertension, renal and hepatic insufficiency, and anemia distinguished this case from the DeSanctis-Cacchione and Cockayne

syndromes. The patient died at the age of 16 years from multi-organ failure following pneumonia. The symptoms of this patient were remarkably similar to the phenotypes of the *Ercc1*$^{-/-}$ (3) and the *Xpf* $^{m/m}$ (131) mouse models which also show accelerated aging characteristics.

The causative allele of this patient contained a mutation in a conserved residue in the HLD of XPF (R153P), leading to aggregation of the protein and mislocalization to the cytoplasm to a higher degree than in other XP-F patients (3,44). Fibroblasts obtained from skin biopsy of the patient showed markedly decreased UV induced unscheduled DNA synthesis (UV-UDS) (~5% of wild-type), a lower value than that observed in typical XP-F patient fibroblasts, although these cells were less sensitive to UV than XP-A fibroblasts. The hypersensitivity to cross-linking agent MMC indicates that a deficiency in ICL repair may contribute significantly to this phenotype.

Cranio-occulo-facio-skeletal syndrome (COFS)

Mutations in ERCC1 were linked to cranio-occulo-facio-skeletal syndrome (COFS), a disease previously diagnosed in patients with mutations in either CSB (132) or XPD (133). The patient was born with low birth-weight, microcephaly with premature closure of the fontanels, bilateral microphthalmia, blepharophimosis, high nasal bridge, short filtrum, micrognathia, low-set and posterior-rotated ears, arthrogryposis with rocker-bottom feet, flexion contractures of the hands, and bilateral congenital hip dislocation (4). MRI showed a simplified gyral pattern and cerebellar hypoplasia. The child did not pass any developmental milestones and died at 14 months (4). The phenotype of the patient matches that of *Ercc1*$^{-/-}$ mice. Together, these observations show that ERCC1 is essential for human development and viability.

Fanconi anemia

Fanconi anemia (FA) is genetic disorder, named after the Swiss paediatrician, Guido Fanconi, who described three brothers with this disease in 1927 (134). To this date, mutations in sixteen genes have been discovered to cause FA, and all of these genes are involved in repairing ICLs (92,135). XPF is a newly identified member in the FA family. Mutations in XPF have been identified to cause FA disorder in two patients (Chapter 3, unpublished data). Both patients displayed bone marrow failure and other FA-related malformations. In addition, cells derived from them were extremely sensitive to cross-linking agents and showed positive results in the chromosome fragility test. Sequencing analysis showed that one patient (from Spain) carried a XPF allele with a premature stop codon due to a 5 base pair (bp) deletion in exon 8 resulting in a protein of only 495 amino acids, while the other allele had a point mutation at a highly conserved residue near the active site (R689S) (38). Only the protein encoded by the second allele was detected. The other patient (from Germany) expressed two unstable variants of XPF: one lacked the HhH domain at C-terminus due to a 28 bp duplication in exon 11, and the other contained a point mutation in the HLD, L230P. Interestingly, cellular experiments showed mild sensitivity to UV for both patients. They had no signs of photosensitivity, atrophic epidermis or skin cancer, distinguishing them from XP or XFE patients (Chapter 3, unpublished data)(3,44).

b. MOUSE MODELS OF ERCC1-XPF DEFICIENCY

***Ercc1*$^{-/-}$ mouse**

Ercc1$^{-/-}$ mice were made independently by two research groups, by inserting a *neo* either between exons 4 and 6 or in exon 7 (136-138). The mice were runted at birth, showed abnormalities of liver with large polyploid nuclei, and died of hepatic failure with increased levels of p53 in liver, kidneys and brain. They had a shortened lifespan of about 3 - 4 weeks (136). The primary mouse embryonic fibroblasts (MEFs) grew slower than their wild-type counterparts (3,137). These MEFs and mouse livers had elevated p21 levels, suggesting that they were undergoing senescent changes (137,139). In contrast to normal cerebellar development in the NER defective *Xpa*$^{-/-}$ mice (140), the cerebellum in *Ercc1*$^{-/-}$ mice exhibited delayed development (4). They also had reduced hematopoietic reserves and showed a multilineage cytopenia with reduction in stress induced erythropoiesis and fatty replacement of the bone marrow (141), resembling Fanconi anemia (142). This result indicates that at least a part of the *Ercc1*$^{-/-}$ mouse phenotype is attributed to the defect in ICL repair. Additionally, these mice showed neurodegenerative changes including dystonia and ataxia, musculoskeletal degeneration evident from sarcopenia, kyphosis and renal failure. Combined with the cellular observations, these findings suggest an accelerated aging process in *Ercc1*$^{-/-}$ mice, consistent with the phenotype of XFE patient (3).

Xpf$^{m/m}$ mouse

Mouse Xpf was mutated with a *neo* insert between exons 7 and 8 and a stop codon at position 445 (131). Consistent with the obligate heterodimeric nature of ERCC1 and XPF, this mouse was a phenocopy of the *Ercc1*$^{-/-}$ mouse. The mice were runted, contained enlarged and polyploid hepatocytes and died within 3 weeks after birth (131).

Ercc1 hypomorphs

The *Ercc1*$^{\Delta/\Delta}$ mice were generated by inserting a stop codon at position 292 of *Ercc1* (137), resulting in a truncated transcript and expressing a protein missing seven amino acids at the C-terminus of ERCC1. The expression from the *Ercc1*$^{\Delta/\Delta}$ cells was reduced to about 15% of the wild-type cells (137), possibly due to an interrupted interaction with XPF. The *Ercc1*$^{\Delta/\Delta}$ mice lived for about 6 months from birth, longer than the *Ercc1*$^{-/-}$ mice but shorter than wild type mice. In addition, they displayed ferritin deposition in the spleens, thin skin lacking subcutaneous fat, and renal tubular dilatation with abnormal nuclei and proteinuria. Unlike the *Ercc1*$^{-/-}$ mice, no evidence of increased apoptosis was shown in tissues of *Ercc1*$^{\Delta/\Delta}$ mice. Both *Ercc1*$^{-/-}$ and *Ercc1*$^{\Delta/\Delta}$ mice were infertile (137). In addition, the phenotype of the ERCC1$^{-/\Delta}$ mice (with one null allele and one hypomorphic allele of ERCC1) was shown to be similar to *Ercc1*$^{\Delta/\Delta}$ mice (143,144).

Ercc1 conditional mutants

In order to correct the hepatic failure of *Ercc1*$^{-/-}$ mice, transgenic Ercc1 was expressed under a liver specific promoter (145). This prevented the polyploidy of hepatocytes and liver failure, and increased lifespan up to 61 – 88 days. These mice eventually died of renal failure with a similar polyploid phenotype in the proximal tubule cells. They had elevated plasma lactate levels, pointing towards possible mitochondrial consequences of ERCC1 deficiency. Neurodegeneration was evident from lack of coordination, ataxia and loss of visual acuity, but no histological changes in the brain were detected (146).

Ercc1$^{-/-}$ mice had extremely short lifespan, making it difficult to study the UV sensitivity of their skin. To overcome this problem, mice were made with a floxed Ercc1 allele under bovine

K5 promoter. These mice showed a 20-fold increase in sensitivity to the short-term erythematous changes in epidermis with UV-B irradiation. They also showed rapid onset of tumors, mainly squamous cell carcinomas (147).

c. ALTERED EXPRESSION OF ERCC1-XPF IN CANCERS

ERCC1-XPF functions in multiple repair pathways to maintain genome stability via NER (1), ICL repair (148), DSB repair (149,150) and telomere maintenance (116,151). On the other hand, chemotherapeutic agents and radiotherapy kill cancer cells by introducing DNA lesions. Therefore, ERCC1-XPF has the potential to serve as a biomarker in cancer cells, providing valuable information for cancer diagnosis or drug development. Studies used to evaluate potential cancer-specific trends in ERCC1 and XPF include single nucleotide polymorphisms (SNPs), mRNA levels, and protein levels.

SNPs

The effect of SNPs on protein expression and function depends on their specific natures and locations. There is no direct evidence that polymorphisms of ERCC1 or XPF affect protein levels. A survey of polymorphisms in NER genes shows that ERCC1 contains six SNPs and XPF contain at least seven SNPs. SNPs in ERCC1 do not result in amino acid substitutions, while one in XPF leads to P379S substitution (152). For instance, a commonly studied silent polymorphism of ERCC1 contains a change in codon 118 from a commonly used AAC to an infrequently used AAT (153).

Despite the lack of evidence that mRNA or protein levels are affected, ERCC1 SNPs are correlated with many different types of tumors. In this context, the association with non-small

cell lung carcinoma (NSCLC) has been the most extensively studied. Risk, survival and treatment response in NSCLC are correlated with either or both codon 118 and C8092A polymorphisms of ERCC1 (154-159). However, a large meta-analysis of multiple studies failed to find any correlation (160). Several other cancers, including esophageal cancer (161,162), ovarian cancer (163), squamous cell carcinoma of head and neck (164), basal cell carcinoma (165) and bladder cancer (166) have shown correlation of outcome or chemotherapy resistance to ERCC1 genotypes. However, pancreatic adenocarcinoma (167), carcinoma of prostate (168), colorectal carcinoma (169,170), glioma (171), renal cell carcinoma (172) and cervical carcinoma (173) do not have any link with ERCC1 SNPs.

mRNA

Messenger RNA (mRNA) levels of ERCC1 and XPF have absolutely no correlation to their respective protein levels (Niedernhofer *et al.*, unpublished data). Multiple studies have suggested a correlation between treatment with cisplatin and the level of ERCC1 mRNA in tumors or tumor cell lines from gastric, ovarian, colorectal, esophageal and non-small cell lung cancer (174-184). Similar to SNPs, mRNA levels of ERCC1 are most extensively studied in NSCLC. Importantly, clinical evidence indicates that patients with low *ERCC1* mRNA expression in tumors respond better to cisplatin-based therapy (179), while high mRNA levels of ERCC1 lead to resistant to cisplatin in ovarian cancer and gastric cancer (176,180). Patients with tailored chemotherapy based on the ERCC1 mRNA levels perform better than those receiving a standard control therapy of docetaxel and cisplatin (185).

Protein levels in tumors

Measurement of ERCC1 protein levels of paraffin-embedded tumor samples by immunohistochemistry (IHC) has several advantages when used to study the effect of the ERCC1 levels in response to various chemotherapeutic agents. Since proteins are more stable than RNA, samples can be obtained much more easily than mRNA. NSCLC remains the most studied tumor type (186-189). High ERCC1 protein expression correlated with poor response to cisplatin chemotherapy and poor survival outcome (187), consistent with the results observed from mRNA. On the other hand, high ERCC1 was associated with better survival following surgery (186,189). Unfortunately, the antibody (8F1) used in almost all studies is not specific for ERCC1 (190). It recognizes human ERCC1, but also a second unknown nuclear antigen(190). Recently, antibodies specifically targeting ERCC1 and XPF were found, which are better suited for clinical measurements of ERCC1-XPF (191).

PREVIEW

This thesis is focused on elucidating the roles of DNA binding domains of ERCC1-XPF in NER and ICL repair pathways. The results of Chapter 2 show that domains of ERCC1-XPF are of hierarchical importance in DNA binding and act synergistically in NER. Subsequent to our finding that mutations in DNA binding residues render cells more sensitive to cross-linking agents than to UV, mutations in XPF were found to cause FA disorder for the first time (Chapter 3). A FA patient coincidently possessed a mutation in one of our proposed DNA binding residues in Chapter 2. Following the studies on the DNA binding-FA mutant, we further investigated

how mutations in DNA binding residues near the active site affect the functions of ERCC1-XPF in NER and ICL repair (Chapter 4).

CHAPTER 2

MULTIPLE DNA BINDING DOMAINS MEDIATE THE FUNCTION OF ERCC1-XPF IN NUCLEOTIDE EXCISION REPAIR

Adapted from the manuscript by Yan Su, Barbara Orelli, Advaitha Madireddy, Laura J. Niedernhofer and Orlando D. Schärer, published in "Journal of Biological Chemistry" in 2012, volume 287, pages 21846 –21855.

ABSTRACT

ERCC1-XPF is a heterodimeric, structure-specific endonuclease that cleaves single-stranded/double-stranded DNA junctions and has roles in nucleotide excision repair (NER), interstrand crosslink (ICL) repair, homologous recombination, and possibly other pathways. In NER, ERCC1-XPF is recruited to DNA lesions by interaction with XPA and incises the DNA 5' to the lesion. We studied the role of the four C-terminal DNA binding domains in mediating NER activity and cleavage of model substrates. We found that mutations in the helix-hairpin-helix domain of ERCC1 and the nuclease domain of XPF abolished cleavage activity on model substrates. Interestingly, mutations in multiple DNA binding domains were needed to significantly diminish NER activity *in vitro* and *in vivo*, suggesting that interactions with proteins in the NER incision complex can compensate for some defects in DNA binding. Mutations in DNA binding domains of ERCC1-XPF render cells more sensitive to the crosslinking agent mitomycin C than to Ultraviolet radiation, suggesting that the ICL repair function of ERCC1-XPF requires tighter substrate binding than NER. Our studies show that multiple domains of ERCC1-XPF contribute to substrate binding, and are consistent with models of NER suggesting that multiple weak protein-DNA and protein-protein interactions drive progression through the pathway. Our findings are discussed in the context of structural studies of individual domains of ERCC1-XPF and of its role in multiple DNA repair pathways.

INTRODUCTION

Structure-specific endonucleases are widespread enzymes that incise DNA as components of most DNA repair and recombination pathways. The activity of these enzymes needs to be tightly regulated since they might otherwise inadvertently fragment DNA (192). One of the most important pathways depending on the action of endonucleases is nucleotide excision repair (NER), which addresses lesions induced by UV light, environmental mutagens and certain cancer chemotherapeutic agents. In NER, an oligonucleotide of 24-32 nucleotides in length containing the damage is removed and the original DNA sequence restored using the non-damaged strand as a template (2). NER can be initiated in two ways: Transcription coupled NER (TC-NER) is triggered when RNA polymerase II is stalled by a bulky DNA lesion during transcription (193); Global genome NER (GG-NER) occurs anywhere in the genome and is initiated by the damage sensor XPC-RAD23B, in some cases with the help of the UV-DDB-ubiquitin ligase complex (56,58,194). Subsequently, TFIIH verifies the presence of the lesion and opens up the DNA helix, allowing the formation of a pre-incision complex containing the endonuclease XPG, the single stranded DNA binding protein RPA and the architectural protein XPA (65,66,68). Finally, the second endonuclease, ERCC1-XPF is recruited to the pre-incision complex (41,57) and incises the DNA 5' to the lesion, triggering initiation of repair synthesis, incision 3' to the lesion by XPG, completion of repair synthesis and ligation (77,195,196).

The two structure-specific endonucleases involved in NER, XPG and ERCC1-XPF, are multifunctional proteins, with diverse roles in NER and other pathways. XPG is a latent endonuclease, fulfilling first a structural and subsequently a catalytic role in NER. It has additional roles in transcription in conjunction with TFIIH (192). The roles of XPG outside of NER are manifest in the severe phenotypes of many XP-G patients (197,198). By contrast, most

of the known XP-F patients present with a mild XP phenotype and have significant residual NER activity due to the presence of low levels of active XPF protein (44). However, a subset of patients and mice with deficiencies in XPF or ERCC1 are much more severely affected and suffer symptoms not caused by NER deficiency alone including developmental abnormalities, premature aging and early death, (3,4,131,137,199). It is believed that these additional symptoms are due to the roles of ERCC1-XPF in interstrand crosslink (ICL) repair, homologous recombination and possibly telomere maintenance (36,110,116,200,201).

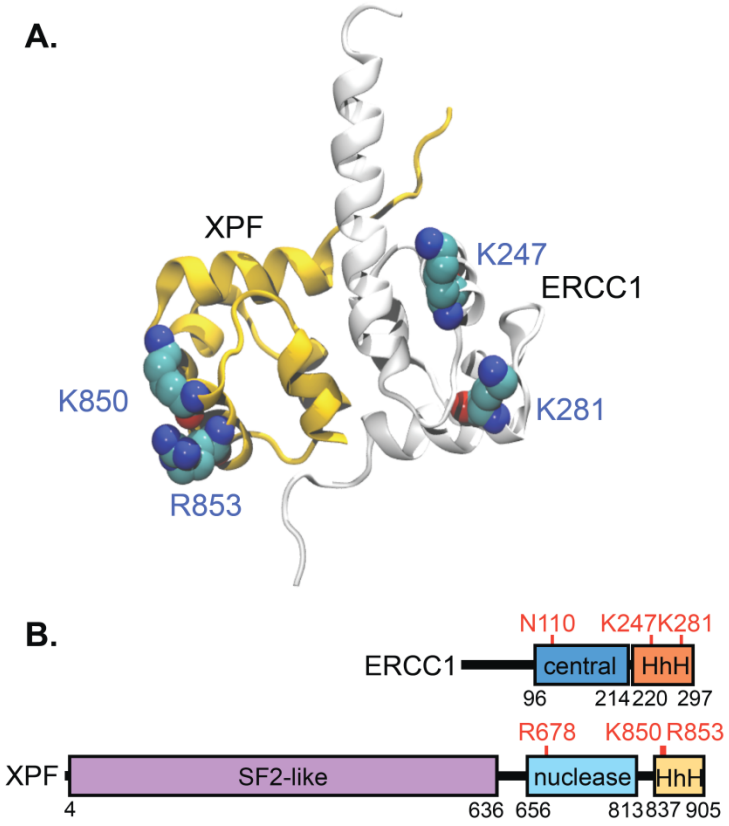


Figure 1. Structure of the ERCC1-XPF HhH domains and scheme of the DNA binding domains in ERCC1-XPF. A. The HhH domains of ERCC1 (gray) and XPF (yellow) contribute to dimerization. The residues shown in atom color are believed to contribute to DNA binding and were mutated in our study. The figure is adapted from (31). B. Five domains of ERCC1-XPF that may contribute to DNA binding are shown. Residues in the central and HhH domains of ERCC1

and in the nuclease and HhH domains of XPF mutated to alanine in our study are highlighted in red.

ERCC1-XPF forms an obligate heterodimer and contains a number of distinct domains that contribute to various aspects of its function (**Figure 1**). Both proteins contain helix-hairpin-helix (HhH) domains at their C-termini that are required for heterodimer formation (31,202). The active site with the conserved nuclease motif is located adjacent to the HhH domain in XPF (38). The central domain of ERCC1 is structurally homologous to the XPF nuclease domain (24,31), however, instead of the active site rich in acidic residues, it contains a groove lined with basic and aromatic residues that interact with the XPA protein, connecting ERCC1-XPF to the NER machinery (40-42,72). XPF, the larger of the two proteins, contains an N-terminal SF2-helicase like domain (HLD), which has lost the ability to bind ATP and to unwind duplex DNA (8). This domain has been implicated in DNA binding and protein-protein interactions, possibly mediating an interaction with SLX4 in ICL repair and other pathways (45-48).

Interestingly, these five domains have been implicated in DNA binding, but evidence to date has been based on analysis of individual domains or on studies of archaeal XPF proteins (25,30-33,40,41). We investigated the role of four C-terminal DNA binding domains by mutational analysis in the context of full-length ERCC1-XPF. Our studies show that mutations in any single domain are insufficient to abolish NER *in vitro* and *in vivo*. Instead, we report that mutations in multiple domains are necessary to disrupt NER and that there is a hierarchy of importance of the individual domains. Our studies are consistent with the notion that multiple weak interactions among proteins and DNA substrates drive progress through the NER reaction (2,203).

EXPERIMENTAL PROCEDURES

Protein expression and purification

Site directed mutagenesis of pFastBac-XPF and pFastBac-ERCC1-His (38) was performed using the QuikChange site-directed mutagenesis kit (Stratagene) to generate the following mutations: XPF^{R678A}, $XPF^{K850A/R853A}$, $XPF^{R678A/K850A/K853A}$, $ERCC1^{K247A/K281A}$, $ERCC1^{N110A}$, and $ERCC1^{N110A/K247A/K281A}$. Bacmid DNA was generated in DH10Bac cells and transfected into the Sf9 insect cells to obtain baculoviruses according to a standard protocol (Bac-to-Bac, Invitrogen). The following combination of ERCC1 and XPF proteins were co-expressed in Sf9 cells for 60 to 65 hours with an MOI of 5: ERCC1-XPF, ERCC1-XPF^{R678A}, ERCC1-$XPF^{K850A/R853A}$, ERCC1-$XPF^{R678A/K850A/R853A}$,$ERCC1^{K247A/K281A}$-XPF, $ERCC1^{N110A}$-XPF, $ERCC1^{N110A/K247A/K281A}$-XPF, $ERCC1^{K247A/K281A}$-XPF^{R678A}, $ERCC1^{N110A}$–XPF^{R678A}, and $ERCC1^{K247A/K281A}$-$XPF^{K850A/K853A}$. Cells were lysed and proteins purified over a 1 ml Nickel (His Trap) column (Amersham Biosciences), a HiLoad 16/60 Superdex 200 column (Amersham Biosciences) and a 1 ml Hitrap Heparin column (Amersham Biosciences) as described in (38). The proteins eluted at 650 mM NaCl from the Heparin column and were in some cases concentrated on an Amicon Ultra-4 Centrifugal Filter (Millipore). Proteins were frozen in aliquots in liquid N_2 and stored at -80°C.

Endonuclease Assays

10 pmol of a stem-loop oligonucleotide (GCCAGCGCTCGG$(T)_{22}$CCGAGCGCTGGC) labeled with fluorescent dye Cy5 at the 3' end (IDT) were annealed in 200 µl solution (10 mM Tris pH 8.0, 50 mM NaCl, 1 mM $MgCl_2$) by heating at 90°C for 10 min and allowing to cool to RT for 2 h. 100 fmol of the substrate were incubated in 25 mM Tris pH 8.0, 2 mM $MgCl_2$, 10%

glycerol, 0.5 mM β-mercaptoethanol, 0.1 mg/ml BSA, 40 mM NaCl with various amounts of protein in a volume of 15 µl. The reactions were incubated at 30°C for 30 min and quenched by adding 10 µl of 80% formamide/10 mM EDTA. After heating at 95°C for 5 min and cooling on ice, 3µl of each sample were analyzed on a 12% denaturing polyacrylamide gel. Bands were visualized by fluorescence imaging on a Typhoon 9400 imaging system (Amersham Biosciences).

In vitro NER Assay

Extracts derived from XPF-deficient XP2YO cells and plasmids containing dG-acetylaminofluorene (dG-AAF) or 1,3-intrastrand cisplatin (cis-Pt) lesions were prepared as described previously (204-206). For each reaction, 2 µl of repair buffer (200 mM Hepes-KOH (pH 7.8), 25 mM $MgCl_2$, 2.5 mM DTT, 10 mM ATP, 110 mM phosphocreatine (di-Tris salt, Sigma), and 1.8 mg/ml BSA), 0.2 µl of creatine phosphokinase (2.5 mg/ml, sigma), 3 µl of XPF-deficient cell extract (about 10 mg/ml), NaCl (to a final concentration of 70 mM), and different amounts of purified ERCC1-XPF in a total volume of 9 µl were pre-warmed at 30°C for 10 min. 1 µl plasmid containing dG-AAF or cis-Pt (50 ng/µl) was added to each reaction and the samples were incubated at 30°C for 45 min. After adding 0.5 µl of 1 µM of an oligonucleotide complementary to the excision product with a 4G overhang for either dG-AAF (5' GGGGCATGTGGCGCCGGTAATAGCTACGTAGCTCp-3') or cis-Pt (5'GGGGGAAGAGTGCACAGAAGAAGACCTGGTCGACCp-3'), the mixtures were denatured by heating at 95°C for 5 min. After cooling to RT for 15 min, 1 µl sequenase mixture (0.13 units of sequenase (USB), and 2.0 µCi [α-^{32}P] dCTP for each reaction) was added and incubated at 37°C for 3 min, followed by addition of 1.2 µl dNTP mixture (50 µM dCTP, 100

μM dTTP, 100 μM dATP and 100 μM dGTP). The reactions were incubated at 37°C for 12 min and stopped by addition of 8 μl of 80% formamide/10 mM EDTA. After heating to 95°C for 5 min, samples were placed on ice and analyzed on a 14% denaturing polyacrylamide gel. Gels were exposed to a phosphor screen and visualized by PhosphorImager (Typhoon 9400, Amersham Biosceinces).

DNA binding measurements by fluorescent anisotropy

For the anisotropy experiments, the protein storage buffer was exchanged to 25 mM KH_2PO_4 (7.6), 5 mM b-mercaptoethanol, 10% glycerol, 150 mM NaCl. Increasing concentrations of protein were incubated with the splayed arm substrate (10 nM) which was made by annealing the following oligonucleotides: 5'-CTTTCGAACATCCAGGAGAGCACGGCC TTTTTTTTTTTTTTTTTTTT with an FAM label at the 3' end and 5'-TTTTTTTTTTTTTTTTTTTTGGCCGTGCTCTCCTGGATGTTCGAAAG. 100nM of competitor double stranded DNA (5'-TCAAAGTCACGACCTAGACACTGCGAGCTCGAATTCACTGGAGTGACCTC and 5'-GAGGTCACTCCAGTGAATTCGAGCTCGCAGTGTCTAGGTCGTGACTTTGA) were used in each reaction. Each reaction was incubated in 20 μl of buffer (25mM Hepes-KOH (pH 8.0), 15% glycerol, 0.1mg/ml BSA, 2 mM $CaCl_2$, 0.5 mM b -mercaptoethanol, and 40 mM NaCl) at 30°C for 5 min. The experiments were conducted at least 4 times and measured on Infinite M1000 plate reader (Tecan). The data were fitted using the grafit4 program to the equation f=y+a*x^b/(c^b+x^b), where x is the protein concentration, f is the fluorescent anisotropy and c is the Kd value.

Generation of mutant cell lines using lentiviral transduction

Human XP-F mutant fibroblasts XP2YO, Chinese hamster ovary cells UV20 (with a mutation in ERCC1), and 293T cells were cultured in Dulbecco's modified Eagle's medium high glucose 1 x (GIBCO), 10% fetal bovine serum, 100 units/ml penicillin, and 0.1 mg/ml streptomycin at 37°C in a 5% CO_2 humid incubator. Wild-type or mutant XPF cDNA with an hemaglutinin (HA) tag at C-terminal was inserted into the pWPXL vector, which was co-transfected with the packaging plasmid psPAX2 and the envelop plasmid pMD2G into 293T cells to generate lentiviral particles. The particles were transduced into XP2YO cells to produce cell lines stably expressing wild-type or mutant XPF proteins according to established procedures (http://www.lentiweb.com) (207,208). The same procedure was applied to generate cell lines expressing wild-type or mutant ERCC1 in UV20 cells.

Local UV-irradiation and immunofluorescence

About 50,000 cells were plated on a coverslip in 6-well plates, grown for 2-3 days and irradiated through a polycarbonate membrane with 5 μm pores (Millipore) with UV light (254 nm) with a dose of 150 J/m^2 (XP2YO cells) or 120 J/m^2 (UV20 cells). Cells were incubated at 37°C, 5% CO_2 for 30 min to 24 hr. They were washed first with PBS and then with PBS plus 0.05% triton X-100 and fixed with 3% paraformaldehyde plus 0.1% triton X-100 (XP2YO cells) or washed first with PBS and then with PBS plus 0.1% triton X-100, and fixed by 3% paraformaldehyde plus 0.2% triton X-100 (UV20 cells). After fixation, cells were washed with PBS containing 0.2% triton X-100. To stain for (6-4)PPs, cells were treated with 0.07 M NaOH in PBS for 5 min, followed by washing with PBS plus 0.2% triton X-100. After blocking with PBS plus 5 mg/ml BSA and 1.5 mg/ml glycine, cells were stained with mouse monoclonal anti-

(6-4)PP antibody (Cosmo Bio) 1:400, rabbit polyclonal anti-ERCC1 antibody (FL-297, Santa Cruz) 1:100, or rabbit polyclonal anti-HA antibody (ChIP grade, Abcam) 1:2000 for 1.5 h and washed with PBS containing 0.2% triton X-100. Cells were then incubated with secondary antibodies: Cy3-conjugated affinipure goat anti-mouse IgG(H+L) (Jackson ImmunoResearch) 1:1000 and Alexa Fluor 488-labeled $F(ab')_2$ fragment of goat anti-rabbit IgG (H+L) (Invitrogen) 1:800 for 1 h, followed by washing with PBS with 0.2% triton X-100. Samples were washed with PBS, embedded in Vectashield Mounting Medium with 1.5 μg/ml of DAPI (Vector Laboratories) and analyzed using a confocal microscope (Zeiss LSM 510). About 100 cells were counted in at least three independent experiments for quantification.

Clonogenic Survival Assay

Exponentially growing cells were plated in 6 cm dishes in triplicate at a density of 1-20 x 10^3 cells/plate for human cells or 250-5,000 cells/plate for Chinese hamster ovary cells, depending on the dose of genotoxin used and the predicted phenotype of the cells. For the hamster cells, wild-type cells were plated as follows: 3 plates at 250 cells/plate (untreated) 3 plates at 250 cells/plate (dose 1), 3 plates at 250 cells/plate (dose 2), 3 plates at 500 cells/plate (dose 3), and 3 plates at 500 cells/plate (dose 4). For hamster cells, knock-out cells were plated as follows: 3 plates at 250 cells/plate (untreated), 3 plates at 250 cells/plate (dose 1), 3 plates at 500 cells/plate (dose 2), 3 plates at 1000 cells/plate (dose 3), and 3 plates at 5,000 cells/plate (dose 4). The next day, the cells were transiently exposed to genotoxins. For UV treatment, the media was aspirated from the plates; the cells were washed with PBS; irradiated with UV (254 nm, Spectroline X-15) and the cells were replenished with fresh medium. For mitomycin C (MMC) treatment, the cells were treated with medium containing MMC (Fisher) at 37°C for 1

hr. The plates were then washed twice with PBS and replenished with fresh medium. Seven to ten days after exposure to genotoxins, when colonies are visible to the naked eye, the cells were fixed and stained with 50% methanol, 7% acetic acid and 0.1% Coomassie brilliant blue. The colonies (defined as containing >10 cells) were counted using an Olympus SZ61 stereomicroscope with a 10X eyepiece. The data was plotted as the number of colonies on treated plates divided by the number of colonies on untreated plates ± S.E. for a minimum of 2 independent experiments, both done in triplicate.

RESULTS

Generation of ERCC1-XPF with mutations in four DNA binding domains

We used information from structural and functional studies to design ERCC1-XPF proteins with mutations in four DNA binding domains in the C-terminal halves of the two proteins. Previous studies have shown that a protein containing only the two HhH, XPF nuclease and ERCC1 central domains can still cleave ss/dsDNA junctions, albeit with much reduced efficiency (31). HhH domains have been shown to contribute to DNA binding in ERCC1-XPF (30,31) and other proteins (209). Based on structural studies, we chose two basic residues in each HhH domain, K247 and K281 in ERCC1, and K850 and R853 in XPF, which have been suggested to play a role in DNA binding (**Figure 1**) (30,31). NMR studies revealed that the HhH domain in ERCC1 appears to be more important for DNA binding (30). We expected the nuclease active site in XPF to interact with DNA directly and chose to include R678 (**Figure 1B**), as we have previously shown that mutating this residue to alanine resulted in a protein with

only marginal activity on model substrate, while being active in NER in the presence of the other NER proteins. We suggest that this residue contributes to positioning of the active site on the substrate rather than the catalytic mechanism (38,196). Although the central domain of ERCC1 can bind DNA (31,40), its main function is likely to interact with XPA. We included the N110A mutant, which has a weakened interaction with XPA (**Figure 1B**) (40,41), to determine to what extent defects in the interaction with XPA synergize with defects in DNA binding. We mutated all of these residues to alanine and combined the different mutations to study how these four domains contribute to the DNA binding and NER activity of ERCC1-XPF.

Mutations in the HhH domain of ERCC1 and the nuclease domain of XPF lead to the loss of incision activity on stem loop substrates

Heterodimeric ERCC1-XPF containing mutations in the HhH, nuclease and central domains were expressed in insect cells and purified using our established protocol (38). All the proteins, including those containing mutations in two domains formed heterodimers similar to the wild-type protein. Only fractions eluting at 60-70ml, representing heterodimeric ERCC1-XPF were collected during the gel filtration chromatography step (**Figure S1A**). We have previously shown that these fractions contain the active protein (30). Furthermore, the CD spectrums of selected mutants were identical to that of wild-type ERCC1-XPF suggesting that they were properly folded (**Figure S1B**). The endonuclease activity of ERCC1-XPF in wild-type and mutant form was tested on a 3' fluorescently labeled stem-loop model substrate, in which the protein cut at the ss and dsDNA junction in the presence of Mg^{2+} (39). ERCC1-XPF in wild-type form and with mutations in the HhH domain of XPF or the central domain of ERCC1, cleaved the stem-loop substrate with similar efficiencies as the wild-type protein in a concentration-

dependent manner (**Figure 2**). By contrast, no such incision products were observed for proteins containing mutations in the HhH domain of ERCC1 (ERCC1$^{K247A/K281A}$, **Figure 2A**, lanes 4-5) or in the nuclease domain of XPF (XPF^{R678A}, **Figure 2B**, lanes 4-5). From these results we conclude that the HhH domain of ERCC1 and the nuclease domain of XPF are more important than the ERCC1 central and XPF HhH domains in mediating nuclease activity.

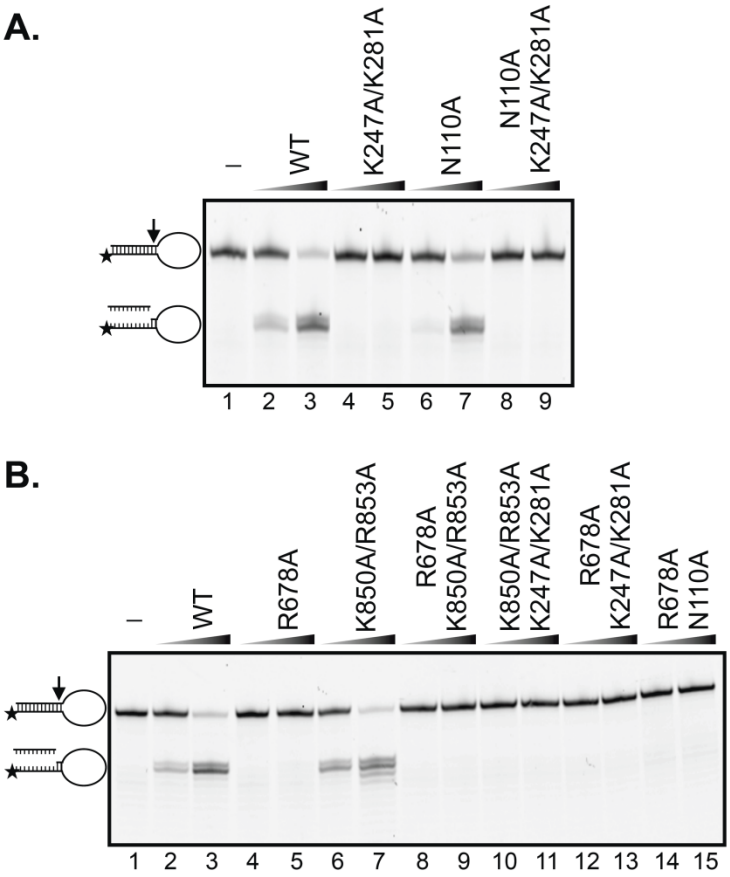


Figure 2. Mutations in the HhH domain of ERCC1 and the nuclease domain of XPF impact nuclease activity on a stem-loop substrate. ERCC1-XPF in wild-type form or with mutations in ERCC1 (**A.**), XPF or both (**B.**) were incubated with 3'-Cy5 labeled stem-loop DNA substrate (6.7 nM) at 37 °C for 30 min, and products were visualized on 12% denaturing polyacrylamide gels. Concentrations of ERCC1-XPF: lanes 1: no protein; Lanes 2, 4, 6, 8 and 10, 12 and 14: 3.3 nM; Lanes 3, 5, 7, 9 and 11: 26.7 nM. The position of the substrate and product are indicated on the left side of the gel.

The HhH domain of ERCC1 is a key contributor for DNA binding affinity

We wanted to gain insight into whether the reduced nuclease activity caused by mutations in the ERCC1 HhH and XPF nuclease domains correlated with reduced DNA binding activity. We used fluorescence anisotropy to measure the binding affinity of ERCC1-XPF to model DNA substrates. We titrated 3' labeled splayed arm, stem loop or duplex DNA substrates with wild-type ERCC1-XPF and measured the increase in anisotropy as a function of the protein concentration. Surprisingly, ERCC1-XPF bound dsDNA with similar affinity as the splayed arm or stem-loop substrates (data not shown), although ERCC1-XPF did not incise dsDNA (39). However, addition of a tenfold excess of non-labeled dsDNA abolished the binding of ERCC1-XPF to dsDNA, but not to the splayed arm and stem loop substrates, confirming that ERCC1-XPF specifically bound ss/dsDNA junctions.

To determine the effect of mutations in the four domains on DNA binding quantitatively, 3'-FAM-labeled spayed arm substrates were incubated with increasing amounts of protein in the presence of a tenfold excess of non-labeled dsDNA competitor. The anisotropy binding curves for the WT, XPF^{R678A}, $ERCC1^{N110A}$ and $XPF^{K850A/K853A}$ proteins displayed a similar shape, while the anisotropy signal of $ERCC1^{K247A/K281A}$ was about 40% lower, suggesting that the architecture of this protein-DNA complex was different from that of the wild-type protein (**Figure 3**). Quantification of the data revealed that the $ERCC1^{K247A/K281A}$ also displayed the lowest binding affinity for the splayed arm substrate (Kd = 219 ± 9 nM), about two fold lower than the WT, XPF^{R678A} and $ERCC1^{N110A}$ proteins. The affinity of $XPF^{K850A/K853A}$ was about 1.6 fold lower than that of the wild-type protein. The results suggest that the HhH domains of both ERCC1 and XPF contribute to substrate binding by ERCC1-XPF, with a more prominent contribution from the

ERCC1 HhH domain. Our data also show that mutations in a single domain alone are not sufficient to completely abolish DNA binding by ERCC1-XPF.

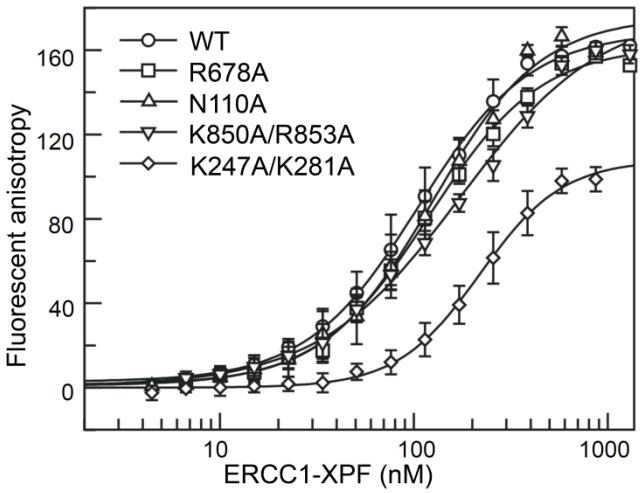


Figure 3. Mutations in the HhH domain of ERCC1 lead to reduced DNA binding affinity. Fluorescein-labeled splayed arm DNA substrate (10nM) in the presence of 100 nM non-labeled competitor double strand DNA was titrated with wild-type and mutant ERCC1-XPF. DNA affinities were measured by fluorescence anisotropy. Error bars represent the standard deviation from a minimum of 4 experiments.

Mutations in multiple DNA binding domains synergistically affect NER activity in vitro

We then investigated the effect of these mutations on the overall NER reaction in the context of cell-free extracts. Plasmids containing site-specific bulky DNA substrates, either 1,3-intrastrand cisplatin crosslinks (cis-Pt) or acetylaminofluorene adducts of dG (dG-AAF), were incubated with XPF-deficient cell-free extract and wild-type or mutant ERCC1-XPF to observe the ability of the protein to excise DNA in the context of other NER proteins (204,205,210). XP-F extracts are devoid of ERCC1 and XPF and did not display any NER activity, but addition of wild-type ERCC1-XPF restored NER activity as shown previously (**Figure 4A/B**, lanes 1-3) (42,196). Mutations in the HhH domain of ERCC1 ($ERCC1^{K247A/K281A}$) led to a mild reduction in

NER activity, at least on the cisplatin substrates, while the analogous mutation in XPF ($XPF^{K850A/R853A}$) had no effect (**Figure 4A/B**, lanes 4-5, and 12-13). Therefore, a defect in the intrinsic DNA binding activity can be overcome at least partially in the context of the full NER complex. XPF^{R678A} was similarly active in NER, although with a slightly shifted incision pattern (**Figure 4A**, compare lanes 10-11 to lanes 2-3). This suggests that R678 contributes to the proper positioning of active site residues to carry out the incision. As observed previously, $ERCC1^{N110A}$ displayed reduced NER activity, due to a weakened interaction with XPA (Figure 4A/B, lanes 6-7 (42).

Since mutations in the individual domains led to only partial inhibition of NER activity, we tested if a combination of multiple mutations would lead to a more severe defect. We found that, in all cases, a combination of mutations in multiple domains led to reduced NER activity compared to the respective mutations in any single domain. Combining the $ERCC1^{K247A/K281A}$ mutation with $ERCC1^{N110A}$ (**Figure 4A**, lanes 8-9), $XPF^{K850A/R853A}$ (**Figure 4A**, lanes 16-17) or XPF^{R678A} (**Figure 4A**, lanes 18-19) led to a loss of detectable NER activity on the cis-Pt substrate, while some activity was preserved for the dG-AAF substrate for the combination of $ERCC1^{K247A/K281A}$ with $ERCC1^{N110A}$ (**Figure 4B**, lanes 8-9) and $XPF^{K850A/R853A}$ (**Figure 4B**, lanes 16-17). Similarly, a combination of XPF^{R678A} with $XPF^{K850A/R853A}$ (**Figure 4A/B**, lanes 14-15) or $ERCC1^{N110A}$ (**Figure 4A/B**, lanes 20-21) led to decreased NER activity with both the cis-Pt and dG-AAF substrates. These studies show that combined mutations in multiple DNA binding domains of ERCC1-XPF synergistically affected NER activity and that many domains simultaneously contributed to the DNA binding and cleavage activity of ERCC1-XPF. Interestingly, the reactions of cis-Pt substrates were generally more severely affected than those of the dG-AAF substrates. This could be due to differences in the structure of the substrates in

the NER complex and/or because dG-AAF were more efficiently processed by NER under our reaction conditions.

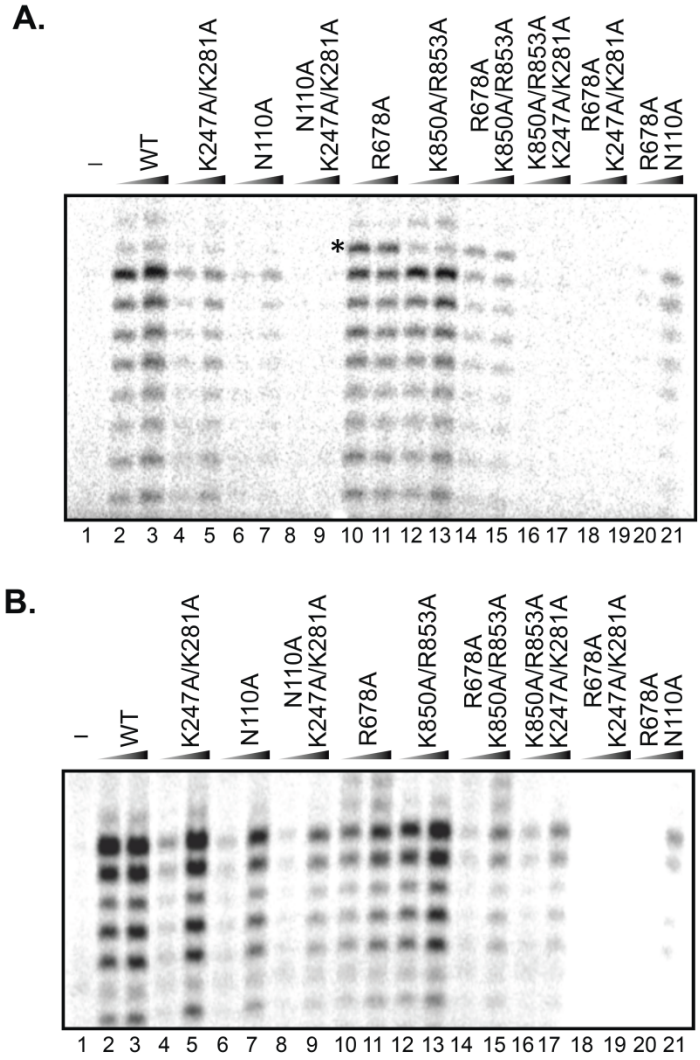


Figure 4. Mutations in multiple domains are needed to affect NER in vitro. Plasmids containing site-specific cis-Pt (**A.**) or dG-AAF (**B.**) lesions were incubated with XPF-deficient cell extracts and recombinant purified ERCC1-XPF in wt or mutant form. The excised 25-32mer oligonucleotide was annealed to a complementary strand with a 4G overhang followed by a fill-in reaction with a-^{32}P-dCTP. Products were analyzed on 14% denaturing polyacrylamide gels and visulized by phosphorimaging. ERCC1-XPF concentrations: lane 1: no protein; even lanes: 5 nM; odd lanes except lane 1: 40 nM. An asterisk denotes the unique band observed with XPF^{R678A}.

ERCC1-XPF with impaired DNA binding ability associates with NER complexes in cells

We wished to define how the DNA binding mutations of ERCC1-XPF affected NER activity *in vivo*, by studying the recruitment of the protein to sites of UV damage and the repair kinetics of 6-4 photoproducts ((6-4)PPs). ERCC1-deficient UV20 and XPF-deficient XP2YO cells were transduced with lentiviral vectors expressing wild-type or mutant ERCC1 and XPF proteins, respectively (42,196). Western blot analysis of the transduced cells revealed that wild-type ($ERCC1^{WT}$, XPF^{WT}) and mutant proteins ($ERCC1^{K247A/R281A}$, $ERCC1^{N110A/K247A/K281A}$, XPF^{R678A}, $XPF^{K850A/R853A}$ and $XPF^{R678A/K850A/R853A}$) were expressed at or above endogenous level (**Figure S2**). As it has been shown that only about 30% of the cellular ERCC1-XPF is engaged in NER following UV irradiation (53), we believe that slight differences in expression levels do not influence the outcome of the experiments described here. Cells were irradiated with UV light (254 nm) through a filter with 5 μm pores to generate sites of local UV damage in the cell nuclei. After incubation for 30 min following UV irradiation, cells were fixed and stained with antibodies to detect (6-4)PPs and ERCC1-XPF. As previously shown (42), around 90% of wild-type protein co-localized with (6-4)PPs (**Figure 5**). Interestingly, similar amounts of co-localization were observed for the proteins with mutations affecting DNA binding ($ERCC1^{K247A/R281A}$, XPF^{R678A}, $XPF^{K850A/R853A}$ and $XPF^{R678A/K850A/R853A}$), indicating that the DNA binding activity of ERCC1-XPF was not required for the stable association of ERCC1-XPF with NER complexes. The only variant that showed decreased co-localization with (6-4)PPs was $ERCC1^{N110A/K247A/K281A}$ (**Figure 5**), but this was expected, as we have shown previously that the N110A mutation affected binding to XPA (42).

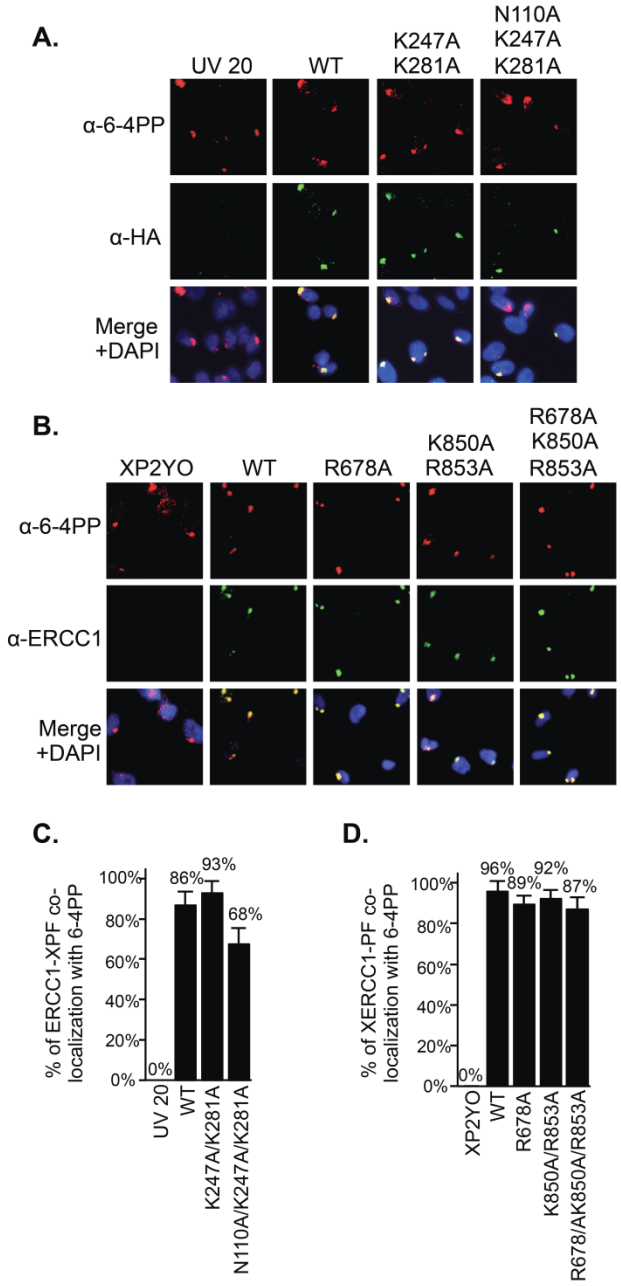


Figure 5. Mutations in DNA binding domains of ERCC1-XPF do not affect the recruitment to sites of UV damage. ERCC1-deficient UV20 cells were transduced with wild-type or mutant ERCC1, and XPF-deficient XP2YO with wild-type or mutant XPF. Cells were irradiated with UV light (254nm, 120 J/m^2 for UV20 cells; 150 J/m^2 for XP2YO cells) through a polycarbonate filter with 5 μm pores and incubated at 37°C for 30 min. Cells were fixed and stained with antibodies against (6-4)PPs (red), ERCC1-XPF (green, using an anti-HA antibody in UV20 cells (**A.**), and an anti-ERCC1 antibody in XP2YO cells (**B.**)). Nuclear DNA was visualized by DAPI staining. **C.** Graphical representation of the percentage of co-localization of ERCC1-XPF with

(6-4)PPs in UV20 cells. **D.** Graphical representation of the percentage of co-localization of ERCC1-XPF with (6-4)PPs in XP2YO cells. The quantification is based on at least three independent experiments and error bars represent the standard deviation.

The repair of (6-4)PPs is delayed by DNA binding mutations

We next measured how the mutations in ERCC1 and XPF affected the rate of repair of UV lesions. Cells were incubated for various times after local UV irradiation and fixed. The percentage of cells containing (6-4)PPs was determined. 30 minutes after UV irradiation, about 40% of all UV 20 cells contained sites of (6-4)PPs. At 24 h 17% of the non-transduced UV20 cells contained 6-4PPs, establishing the rate of removal rate in the absence of NER. (**Figure 6A** and **6C**). In $ERCC1^{WT}$-expressing cells, the damage was almost all gone at 2h and detectable at only marginal levels at later time points. The rate of repair was significantly reduced in $ERCC1^{K247A/R281A}$-expressing cells with 24% and 18% of the cells containing (6-4)PPs at 2 h and 4 h, respectively. Repair appeared to be mostly completed however at later time points (8 h, 24 h), indicating that this mutation in the HhH domain of ERCC1 slowed down, but did not abrogate NER. Adding the XPA-interacting mutation N110A to K247A/K281A led to further delay in repair, although repair was still completed at the 24 h time point (**Figure 6A** and **6C**).

In the XP-F deficient XP2YO cells, significant level (~ 16%) of (6-4)PPs also persisted at 24 h, and the observed damage levels at 30min was slightly higher than in the UV20 cells. Transduction of XPF^{WT} restored the complete removal of (6-4)PPs and the repair rate in $XPF^{K850A/R853A}$-expressing cells was only marginally slower than in wild-type cells. However, the damage removal was impaired in XPF^{R678A}-expressing cells, and a combination of the R678A and K850A/R853A mutations led to a further reduction in the repair rate (**Figure 6B** and **6D**). The cellular studies are thus consistent with the *in vitro* results showing that DNA binding

mutations in the nuclease domain of XPF and the HhH domain of ERCC1 synergistically affected NER.

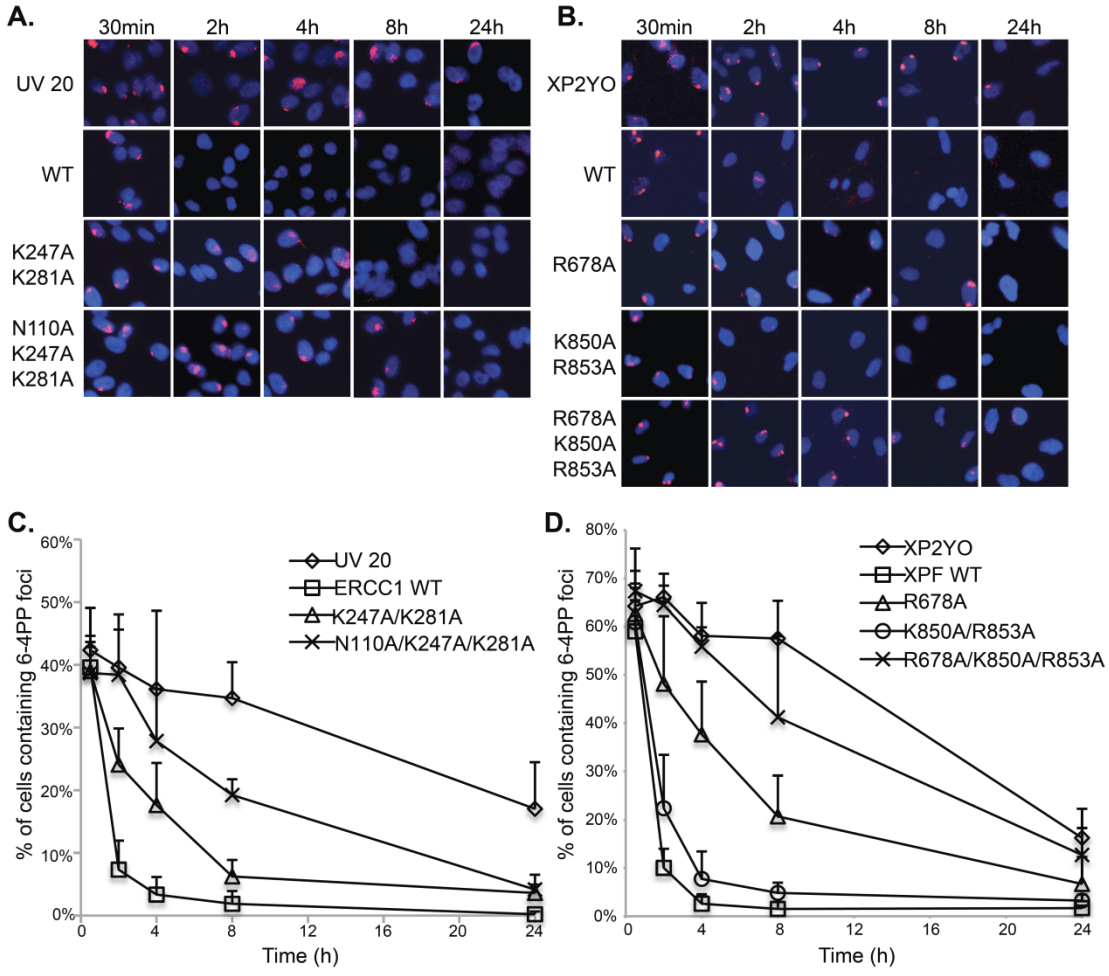


Figure 6. Mutations in DNA binding domains of ERCC1-XPF affect repair kinetics of (6-4)PPs in vivo. ERCC1-deficient UV20 cells were transduced with wild-type or mutant ERCC1 (**A.**), XPF-deficient XP2YO cells with wild-type or mutant XPF (**B.**). Cells are irradiated with UV light (254nm, 120 J/m^2 for UV20; 150 J/m^2 for XP2YO) through a polycarbonate filter with 5 μm pores, incubated at 37°C for 30 min, 2 h, 4 h, 8 h, 24 h, fixed and stained with antibody against (6-4)PPs (red) and with DAPI (blue). **C.** Graphical representation of the percentage of cells containing (6-4)PPs in (transduced) UV20 cells. **D.** Graphical representation of the percentage of cells containing (6-4)PPs in (transduced) XP2YO. Quantification is based on at least three independent experiments and the standard deviations are shown by error bars.

Mutations in multiple DNA binding domains render cells hypersensitive to UV and mitomycin C

Having shown that mutations in DNA binding domains in ERCC1-XPF synergistically affected NER *in vitro* and *in vivo*, we wanted to test whether these mutations also render cells hypersensitive to UV irradiation as well as mitomycin C (MMC), an agent that causes DNA interstrand crosslinks (ICLs). Cells were first treated with different doses of UV light (254 nm) and the sensitivity was assessed using clonogenical survival assays. Consistent with the results from the NER studies above, mutations in single domains (XPF^{R678A}, $XPF^{K850A/R853A}$ and $ERCC1^{K247A/K281A}$) led to mild UV sensitivity, while mutations in two domains ($XPF^{R678A/K850A/R853A}$ or $ERCC1^{N110A/K247A/K281A}$) led to a more pronounced effect. The hypersensitivity to MMC was even more prominent: Cells expressing ERCC1-XPF with mutations in single domains (XPF^{R678A}, $XPF^{K850A/R853A}$ and $ERCC1^{K247A/K281A}$) were already associated with dramatic sensitivity, especially XPF^{R678A}. Cells expressing $XPF^{R678A/K850A/R853A}$ or $ERCC1^{N110A/K247A/K281A}$ were as sensitive as non-transfected cells to MMC. These results demonstrate that proper DNA binding is even more important in the context of ICL repair.

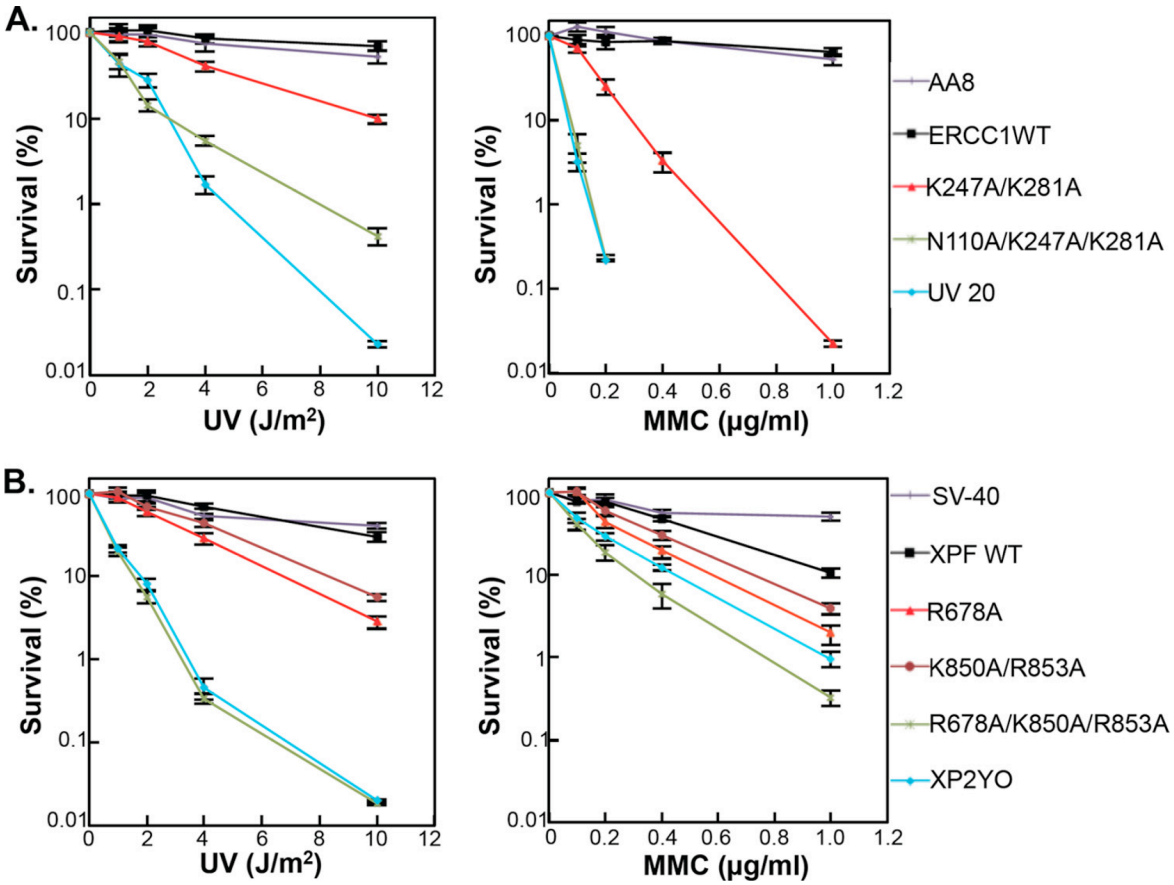


Figure 7. Mutations in DNA binding residues render cells more sensitive to MMC than to UV. Cells were challenged with UV (254nm) to determine the impact of the mutations on NER or with MMC to determine the impact on ICL repair. Top row: WT (AA8), ERCC1-deficient (UV20) and complement Chinese hamster ovary cell lines. Bottom row: WT (SV-40, SV40-transformed human fibroblasts), XPF-deficient (XP2YO) and complemented cell lines. The data were plotted as the percentage of colonies that were formed on the treated plates relative to the untreated plates ± S.E (error bars) for a minimum of 2 independent experiments each done in triplicate.

DISCUSSION

Multiple DNA- and protein-binding domains of ERCC1-XPF control its function in NER

In our study, we investigated the role of four C-termial DNA binding domains in ERCC1-XPF (**Figure 1**) and show that there is a hierarchy of importance of these domains in mediating the interaction with its DNA substrates. Mutations of DNA binding residues located in the HhH domain of ERCC1 and the nuclease domain of XPF have a more deleterious effect on nuclease and NER activity than mutations in the HhH domain of XPF or the central domain of ERCC1. Importantly, mutations in multiple DNA binding domains are required to significantly suppress NER activity in cultured cells and cell-free extracts, whereas mutations in the ERCC1 HhH or XPF nuclease domain alone are enough to deplete nuclease activity of the purified heterodimer on model substrates. These findings are consistent with current models of NER positing that this process is driven by the dynamic assembly and disassembly of the factors involved (43,68). Accordingly, multiple protein-protein and protein-DNA interactions drive NER and enable the ordered hand-off of the substrates in damage recognition, damage verification, pre-incision complex assembly, dual incision, repair synthesis and ligation (2,203). Our studies allow us to propose how multiple interactions between ERCC1-XPF, XPA and DNA control incision 5' to the damage during NER. ERCC1-XPF is recruited to NER complexes by interaction of the central domain of ERCC1 with XPA, in the context of a pre-incision complex consisting of XPA, RPA, TFIIH and XPG (41,42). Our cellular studies suggest that DNA binding of ERCC1-XPF is not required to localize the endonuclease to NER complexes (**Figure 5**). DNA binding domains therefore anchor the protein at the junction following recruitment, allowing the nuclease domain of XPF to be positioned at the incision site and to cleave the DNA, triggering dual incision and repair synthesis (196). It is likely that interactions with other NER proteins, for example RPA and TFIIH (22,67), contribute to the fine-tuning of the incision and coordination with other steps in the NER pathway.

The HhH domains of ERCC1 and XPF have distinct roles in NER

In accordance with NMR studies of individual domains of ERCC1-XPF (30), our results show that mutations in the HhH domain of ERCC1 affect DNA binding, nuclease and NER activity more severely than analogous mutations in the HhH domain of XPF. XPF and ERCC1 are believed to have evolved from a common ancestor (23), which has been studied in the form of archaeal XPF homologs (1). During divergent evolution, XPF retained the nuclease active site and ERCC1 conserved a typical double HhH domain. ERCC1 now harbors an XPA-interacting domain instead of a nuclease domain, and the HhH in XPF has lost some typical features. In accordance with this observation and with structural studies, our work demonstrates that the HhH domain of ERCC1 is more important for substrate binding and NER activity than the analogous domain in XPF (30). We know less about how the HLD of XPF contributes to the interaction with DNA. Based on studies of the homologous domain in the Hef helicase, the HLD is expected to contribute to the interaction with ss/dsDNA junctions. Consistent with this observation, human ERCC1-XPF lacking the HLD has been shown to cut model substrates only with marginal efficiency (31).

How does ERCC1-XPF bind substrates in different pathways?

ERCC1-XPF has been known for a long time to have roles outside of NER in the repair of interstrand crosslinks and in homologous recombination (5,110,200,211). Deficiencies in these processes have been implicated to cause the severe phenotypes of select ERCC1- or XPF-mutant patients and mice that cannot be explained by NER deficiency (3,4,131,137,199). Recent

evidence suggests that an interaction between XPF and the Fanconi anemia protein SLX4/FANCP, rather than with XPA, is crucial for ICL repair (45-48,212-214).

With respect to our present study, it will be of great interest to determine whether DNA binding by ERCC1-XPF in ICL repair is similar as in NER or whether it follows a distinct pattern. We have shown that even within the NER pathway the repair efficiency of cis-Pt and dG-AAF adducts is affected to different extents by mutations in DNA binding domains of ERCC1-XPF (compare **Figures 4A & B**, lanes 8-9, 16-17), perhaps reflecting a different conformation of the substrates in the NER complex. These differences may be even bigger between substrates for NER and ICL repair, as reflected by the sensitivities to UV light and MMC: cells expressing XPF^{R678A} or $ERCC1^{K247A/K281A}$ are more sensitive to MMC than to UV (Figure 7).

High resolution structures of other structure-specific endonucleases such as FEN1 or EXO1 in complex with DNA substrates have shown that these proteins firmly bind the double-stranded portion of the DNA and induce a kink of about 90° in the non-cleaved DNA strand at the junction (6,7). The cleaved strand undergoes local unpairing of the dsDNA at the incision site, and the ssDNA overhang of the cleaved strand may be bound at different sites of the protein. Based on the structures of archaeal homologs of XPF (32,33) a similar arrangement can be envisioned for ERCC1-XPF. Interestingly, the residue equivalent to R678 in XPF in the *Aeropyrum pernix* XPF homolog (R26) is located at a site that possibly binds the ssDNA overhang. Mutation of R678 of XPF affects ICL repair more severely than NER, showing that this residue is more crucial for the positioning of the endonuclease in the context of ICL repair.

We therefore believe that further studies of the five DNA binding domains in ERCC1-XPF will provide valuable insight into how this protein operates in multiple DNA repair

pathways. Furthermore, introduction of specific mutations in those domains may lead to pathway-specific defects, enabling new approaches to study the importance of ERCC1-XPF in maintaining genome stability.

SUPPLEMENTARY METHODS AND MATERIALS

Western Blot

Cells were lysed in NETN buffer (50 mM Tris pH8.0, 150mM NaCl, 1mM EDTA, 0.5% NP40, and protease inhibitor (Roche)). 30 µg of protein (60 µg for HeLa cell extracts) were separated on a 10% SDS gel and transferred to a Hybond-LFP membrane (low fluorescent PVDF membrane, Amersham Biosciences) at 350 mA for 2 h. After blocking with a 5% BSA solution, the membranes were incubated with primary antibodies: rabbit polyclonal anti-ERCC1 (FL-297, Santa Cruz) 1:200 (for UV20 cells), mouse monoclonal anti-XPF (3F2/3, Santa Cruz) 1:200 (for XP2YO cells), and mouse anti-tubulin (Sigma) 1:1000 for 1.5 h, washed by PBS plus 0.1% tween 20. Membranes were incubated with secondary antibodies: Cy3-conjugated affinipure goat anti-mouse IgG(H+L) (Jackson ImmunoResearch) 1:1000 and Alexa Fluor 488 F(ab')$_2$ fragment of goat anti-rabbit IgG (H+L) (Invitrogen) 1:1000 for 1 h, and washed with PBS plus 0.1% tween 20. At last, membranes were washed with PBS, dried at room temperature and visualized by fluorescence on a Typhoon imaging system (Amersham Biosciences).

CD experiments

Protein (ERCC1-XPFWT, ERCC1-XPF^{R678A}, ERCC1$^{k247A/K281A}$-XPF$^{R850/K853A}$) was diluted to 680 nM (dilution buffer containing 25 mM K-phosphate (pH=7.6), 10% glycerol, 5 mM beta-mercaptoethanol, and 650 mM NaCl), and the CD spectrum measured on an Applied Photophysics Chirascan CD spectrometer. The scan was a 1-nm/s intervals over 0.1 cm path length at 4°C with a wavelength range of 200 nm to 260 nm. The background signal for the buffer was subtracted and two experiments were conducted twice.

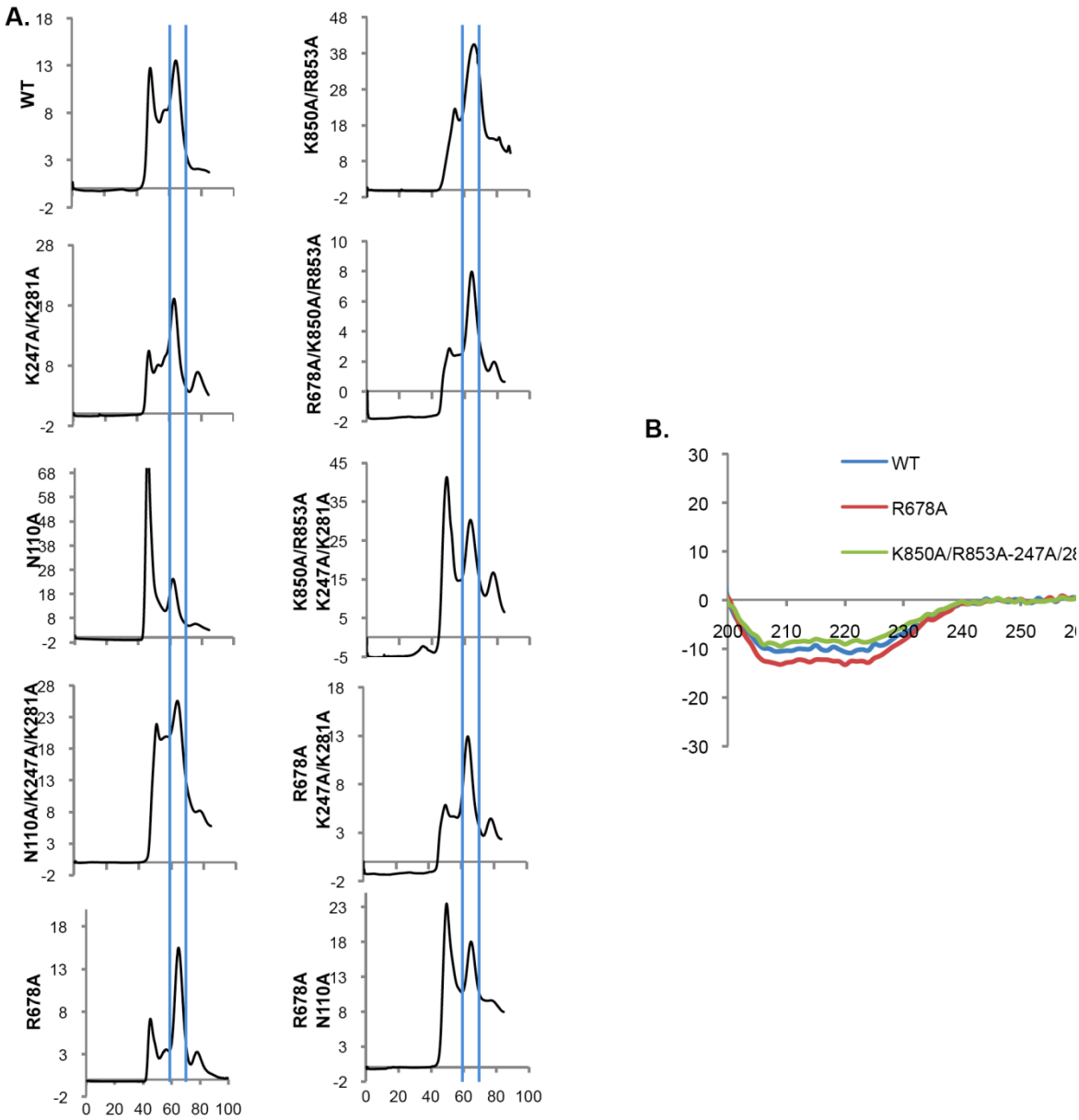


Figure S1. ERCC1-XPF proteins with mutations in the HhH, nuclease and central domains are properly folded. Wild type and mutant ERCC1-XPF are purified through Nickel column, HiLoad 16/60 Superdex 200 column and Hitrap Heparin column. On the HiLoad 16/60 Superdex 200 column, heterodimeric, active wild type and mutant ERCC1-XPF were eluted at 60-70 ml and fractions in that range were collected around the peaks and used for further purification (**A.**). Wild type, XPFR678A, and ERCC1K247A/K281A-XPFK850A/R853A proteins were test on the CD spectrum. They had the similar peaks (**B.**).

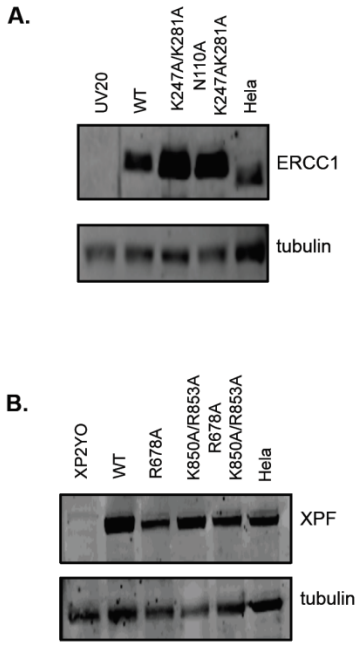


Figure S2. Expression level of ERCC1 and XPF protein in transduced UV20 and XP2YO cells. UV20 cells transduced with wild-type or mutant ERCC1 (**A.**) and XP2YO cells expressing wild-type or mutant XPF (**B.**) were lysed and 30 µg of protein from the extracts (60 µg for HeLa cell extracts) were separated on a 10% SDS gels and transferred to PVDF membrane. ERCC1 was detected using antibody FL-297 (**A.**), XPF using antibody 3F2/3. (**B.**) Tubulin was used as a loading control.

ACKNOWLEDGEMENTS

We thank Drs. Jill Fuss and John Tainer for help with the fluorescence anisotropy experiments, Dr. Jung-Eun Yeo for the gift of dG-AAF-containing plasmid, Dr. Arthur J. Campbell for help with Figure 1 and Drs. Tom Ellenberger and Oleg Tsodikov for a critical reading of the manuscript and helpful discussions.

This work was supported by National Institute of Health grants GM080454 and CA092584.

CHAPTER 3

XPF MUTATIONS SEVERELY DISRUPTING DNA INTERSTRAND CROSS-LINK REPAIR CAUSE FANCONI ANEMIA

Adapted from the manuscript by Massimo Bogliolo, Beatrice Schuster, Chantal Stoepker, Burak Derkunt, Yan Su, Anja Raams, Juan P. Trujillo, Jordi Minguillón, María J. Ramírez, Roser Pujol, José A. Casado, Rocío Baños, Paula Rio, Kerstin Knies, Sheila Zúñiga, Javier Benítez, Juan A. Bueren, Nicolaas G.J. Jaspers, Orlando D. Schärer, Johan P. de Winter, Detlev Schindler & Jordi Surrallés, submitted.

ABSTRACT

Fanconi anemia (FA) is a rare genomic instability syndrome and FA gene products are involved in the repair of DNA interstrand cross-links (ICL). Fifteen FA genes have been identified, but the genetic basis in some FA patients still remains unresolved. Here we used whole exome and Sanger sequencing on DNA of unclassified FA patients and discovered biallelic germline mutations of *XPF*, a structure-specific nuclease previously connected to xeroderma pigmentosum (XP) and segmental progeria (XFE). Further genetic, biochemical and functional analysis demonstrates that the newly identified FA-causing *XPF* mutations strongly disrupt the function of XPF in ICL repair without severely compromising nucleotide excision repair (NER). Our data show that depending on the type of *XPF* mutation and the balance between NER and ICL repair activities, patients present with one of three clinically distinct disorders, highlighting the multifunctional nature of XPF in genome stability and human disease.

FA is characterized by bone marrow failure (BMF), congenital malformations, hypersensitivity to DNA cross-linking agents, and a high susceptibility to cancer. FA genes are involved in the response of tumor cells to widely used anti-tumor drugs that induce ICLs and are inactivated in a number of cancers (135). The discovery of new FA genes improves our understanding of the FA/BRCA pathway, and may give further insight in the process that tumor cells use to respond to anti-cancer therapy. To identify additional FA genes, we performed whole exome sequencing on peripheral blood DNA from a Spanish patient (individual FA104) who was previously excluded from all known FA complementation groups. FA104 was a typical FA patient with BMF, absent thumbs and other FA-related malformations, and a positive chromosome fragility test (Supplementary text and Supplementary Fig. 1a, b on line). Lymphoblasts from this patient were hypersensitive to mitomycin C (MMC) and melphalan, but were normal in FANCD2 monoubiquitination and RAD51 focus formation (215). FA104 lymphoblasts were insensitive to the topoisomerase I inhibitor camptothecin and to the PARP inhibitor KU58948 (data not shown), suggesting a defect downstream in the FA/BRCA pathway but not involving BRCA2 interactions or homologous recombination. Based on a recessive mode of inheritance, exome sequencing identified 17 candidate disease genes for FA104 (Supplementary text and Supplementary Table 1 on line), of which *XPF/ERCC4* immediately caught our attention given the involvement of the XPF-ERCC1 structure-specific nuclease in ICL repair (94,135). Both *XPF* mutations were predicted to be pathogenic: a 5 base pairs (bp) deletion in exon 8 (c.1484_1488delCTCAA) leading to frameshift and premature termination of translation (p.T495NfsX5), and a missense mutation in exon 11 (c.2065C>A; p.R689S) generating an amino acid change in a highly conserved arginine residue within the nuclease active site of XPF (Supplementary Fig. 2a on line). Sanger sequencing on blood DNA confirmed

these mutations (Fig. 1a), and showed that the deletion was inherited from the mother, while the missense mutation is of paternal origin (Supplementary Fig. 3a on line). None of these mutations were found in 400 control alleles from healthy Spanish individuals (data not shown). In MMC-resistant FA104 lymphoblasts (FA104R) obtained after long-term exposure to low dose of MMC, the allele with the 5 bp deletion had gained an additional 7 bp deletion *in cis*, which restored the *XPF* reading frame (Supplementary Fig. 3b). In FA104 lymphoblasts, XPF levels were reduced, while in the reverted FA104R lymphoblasts, XPF levels were back to normal (Supplementary Fig. 3c on line). Non-sense mediated RNA decay was not observed in the allele carrying the premature stop codon, while no truncated XPF protein was detected by Western blotting with two independent N-terminal XPF antibodies (data not shown), suggesting that the short truncated protein may not be stable and/or is degraded and, therefore, that FA104 only expresses the XPF-R689S missense mutant. These data strongly suggest that the XPF defect is responsible for MMC sensitivity in FA104.

Sanger sequencing was then performed in a total of 18 unclassified German FA patients. Biallelic *XPF* mutations were found in another FA patient (individual 1333), unambiguously diagnosed with FA due to multiple FA-related malformations, BMF and a positive chromosome fragility test (Supplementary text and Supplementary Fig. 1c on line). Similar to FA104, lymphoblasts from patient 1333 were normal in FANCD2 monoubiquitination and RAD51 focus formation, were sensitive to MMC and melphalan but insensitive to the topoisomerase I inhibitor camptothecin and to the PARP inhibitor KU58948 (data not shown). Individual 1333 carried a 28 bp duplication in exon 11 (c.2371_2398dup28; p.I800TfsX23; Fig. 1b), which is predicted to result in a truncated XPF protein that lacks the double helix-hairpin-helix (HhH_2) domain involved in heterodimerization with ERCC1 and DNA binding (135). The other allele contained

a missense mutation of a highly conserved amino acid residue within the helicase-like domain (c.689T>C; p.L230P; Fig. 1b and Supplementary Fig. 2b on line). The missense mutation was inherited from the father while the 28 bp duplication is of maternal origin (Supplementary Fig. 3d on line). Western blot analysis on whole cell extracts of 1333 lymphoblasts showed that a missense mutant with normal size and a truncated XPF protein of approximately 90-95 kDa, compatible with the predicted downscaled size of the allele with premature termination of translation due to the duplication, are both expressed at very low levels of about 5% of normal (Fig. 1c). As expected, the truncated XPF protein was undetectable with an antibody against the very C-terminal HhH_2 domain of XPF, amino acids 866-916 (data not shown). Interestingly, the truncated XPF protein was absent in an MMC-resistant 1333 lymphoblastoid cell line (1333R) generated by long-term exposure to MMC, with near normal XPF levels (Supplementary Fig. 3e on line). PCR amplification and sequence analysis revealed that the 28 bp duplication had disappeared in 1333R (Supplementary Fig. 3f on line). Both, inherited duplication and somatic reversion may have been triggered by an inverted 5 bp repeat flanking the region. In conclusion, while FA104 only makes a single and rather stable mutant form of XPF with a mutation in the nuclease domain (XPF-R689S), 1333 expresses low levels of two XPF forms, a truncated protein that lacks the ERCC1-interacting and DNA-binding domain (XPF-I800TfsX23) and a full-length protein with a mutation in the helicase-like domain (XPF-L230P).

We then investigated genetic complementation of MMC sensitivity of both FA patients' cells by lentiviral transduction. Wild type *XPF* cDNA, but not an empty vector, the *XPF* mutant R689S or a *FANCA* cDNA containing vector complemented MMC sensitivity of both lymphoblast FA cell lines (Fig. 1d,e). We finally expressed human *XPF* cDNAs (XPF-WT, XPF-L230P and XPF-R689S) in embryonic fibroblasts (MEFs) from *Xpf* null mice and exposed them

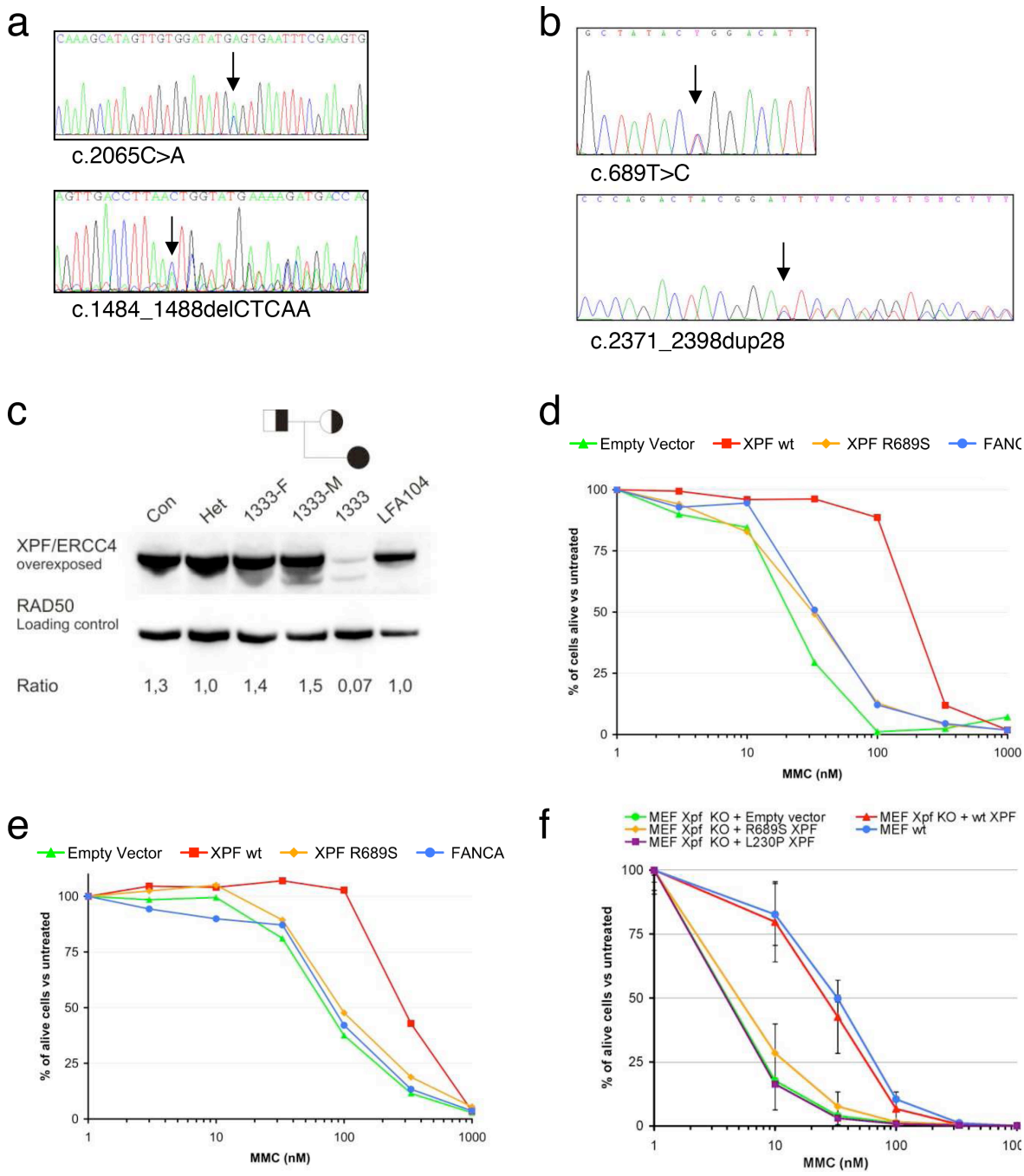


Figure 1. XPF mutations and XPF-deficiency in Fanconi anemia patients. (**a**) Sequence analysis of blood DNA from FA104 revealed a missense mutation in exon 11 (c.2065C>A; p.R689S) (upper panel) and a 5 bp deletion in exon 8 leading to a frameshift and premature termination of translation (c.1484_1488delCTCAA; p.T495NfsX5) (lower panel). (**b**) Sequence analysis of blood DNA from 1333 revealed a missense mutation in exon 4 (c.689T>C; p.L230P;

upper panel) and a 28 bp duplication in exon 11 (lower panel) leading to a frameshift and a premature stop codon (c.2371_2398dup28; p.I800TfsX23). (**c**) Western blot analysis showing XPF expression in lymphoblasts from 1333 and FA104. Lymphoblasts from a healthy individual (Con), the parents of 1333 (1333-F, 1333-M) and an unrelated *XPF* mutation carrier (Het) were used as controls. XPF levels are expressed as ratio relative to the loading control (Rad50). (**d**) Genetic complementation of MMC sensitivity of FA104 lymphoblasts by wild type XPF but not by XPF mutant R689S. Cells transduced with lentiviral particles carrying GFP (empty vector), wild type XPF, XPF mutant R689S or FANCA were grown for 10 days in presence of the indicated concentrations of MMC. Data represent a typical result of at least three independent experiments. (**e**) Genetic complementation of MMC sensitivity of 1333 lymphoblasts by wild type XPF (experiments were performed as in **d**). (**f**) MMC-induced growth inhibition of *Xpf* KO MEFs transduced with lentiviral particles carrying GFP (negative control vector), wild type XPF, XPF-R689S and XPF-L230P. Data represent means and SD of at least three independent experiments.

to MMC. Unlike wild type XPF, XPF-L230P and XPF-R689S mutants did not complement MMC sensitivity of *Xpf* MEFs (Fig. 1f), providing additional evidence that the *XPF* missense mutations of patients FA104 and 1333 are inactivating XPF. Since *XPF* mutations are responsible for a novel FA subtype (FA-Q) we propose *FANCQ* as an alias for *XPF*.

We then tried to understand why the newly identified *XPF* variants lead to FA but not to xeroderma pigmentosum (XP) or progeria (XFE). We hypothesized that these mutants are strongly deficient in ICL repair to cause an FA phenotype but have sufficient NER activity to prevent clinically relevant skin photosensitivity and other XP-related features. To test this hypothesis, we investigated UVC sensitivity of FA104 and 1333 lymphoblasts and observed that FA104 is not and 1333 is only modestly sensitive to UV irradiation when compared to an XPA lymphoblast (Fig. 2a). Since UV survival experiments are challenging in lymphoblastoid cell lines and no such cells are available from XFE, we studied primary and transformed skin fibroblasts from patient 1333 against XPF-mutated XP and XFE patients (no fibroblast were available from FA104). The UV sensitivity of 1333 proved similar to that of XPF patients with

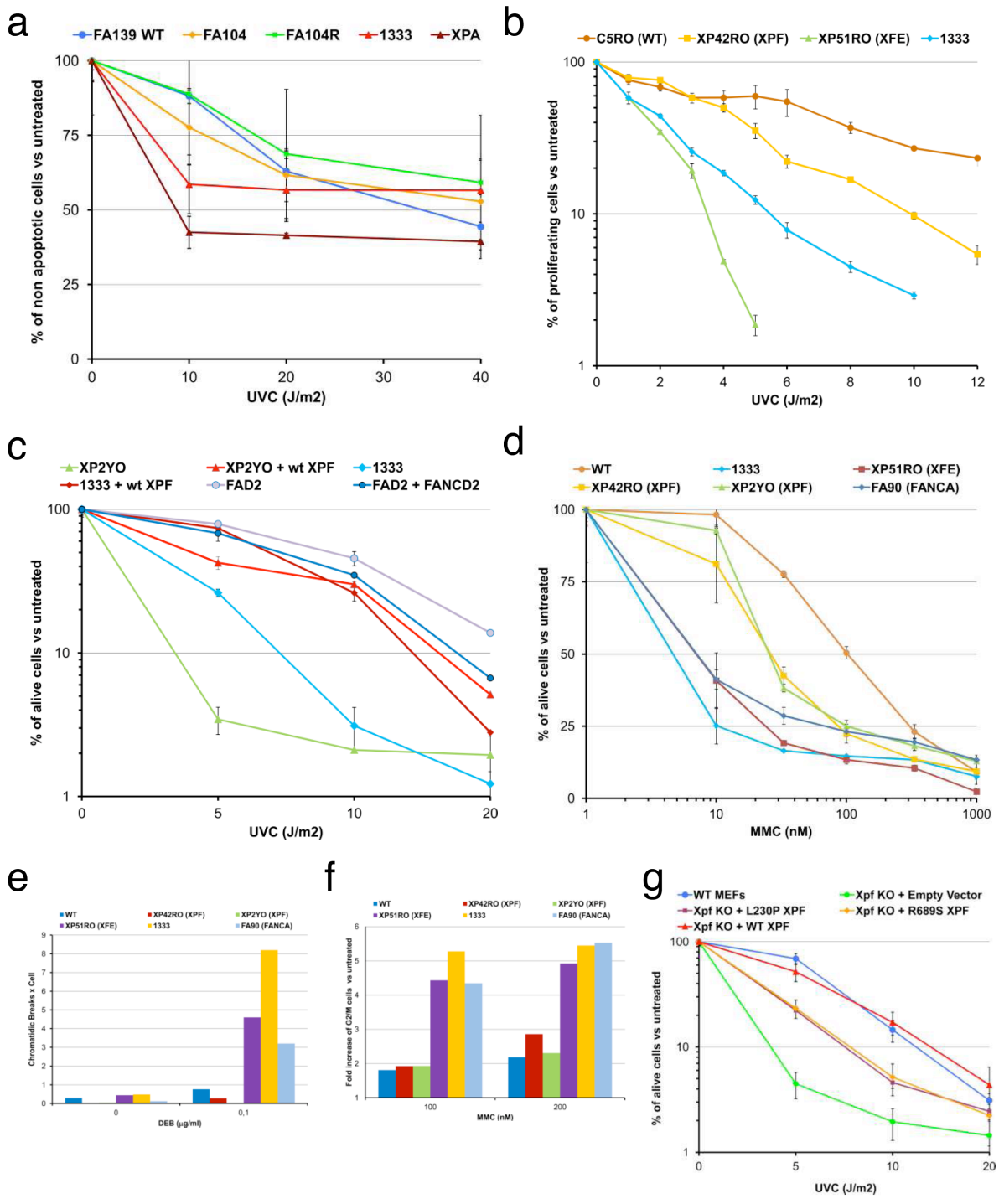

a
FA139 WT
FA104
FA104R
1333
XPA
% of non apoptotic cells vs untreated
UVC (J/m2)
b
C5RO (WT)
XP42RO (XPF)
XP51RO (XFE)
1333
% of proliferating cells vs untreated
UVC (J/m2)
c
XP2YO
XP2YO + wt XPF
1333
1333 + wt XPF
FAD2
FAD2 + FANCD2
% of alive cells vs untreated
UVC (J/m2)
d
WT
1333
XP51RO (XFE)
XP42RO (XPF)
XP2YO (XPF)
FA90 (FANCA)
% of alive cells vs untreated
MMC (nM)
e
WT
XP42RO (XPF)
XP2YO (XPF)
XP51RO (XFE)
1333
FA90 (FANCA)
Chromatidic Breaks x Cell
DEB (µg/ml)
f
WT
XP42RO (XPF)
XP2YO (XPF)
XP51RO (XFE)
1333
FA90 (FANCA)
Fold increase of G2/M cells vs untreated
MMC (nM)
g
WT MEFs
Xpf KO + Empty Vector
Xpf KO + L230P XPF
Xpf KO + R689S XPF
Xpf KO + WT XPF
% of alive cells vs untreated
UVC (J/m2)

Figure 2. UV and MMC sensitivities of XPF mutants leading to Fanconi anemia. (**a**) UVC-induced apoptosis in FA104 and 1333 lymphoblasts. Cells were analyzed for UVC-induced apoptosis 24h post irradiation. Cell lines FA139 (wt) and GM02344 (XPA) were included as positive and negative controls, respectively. Data represent means and SD of at least three independent experiments. (**b**) UV sensitivity of 1333 primary fibroblasts, according to a fast method equivalent to clonogenic survival. XPF (XP42RO) and XFE (XP51RO) primary fibroblasts were included for comparison. (**c**) UVC-induced growth inhibition of human *XPF*-deficient immortal cell lines from XP and FA patients (XP2YO and 1333, respectively) transduced with lentiviral particles carrying XPF-WT. FANCD2 deficient cell lines PD20 (FAD2) and their isogenic corrected counterpart PD20-3-15 (FAD2 + FANCD2) were used as a control. Data represent means and SD of two independent experiments. (**d**) MMC sensitivity of human *XPF*-deficient primary fibroblasts from XP, FA and XFE patients (XP42RO, 1333 and XP51RO, respectively). WT and FAA (FA90) primary fibroblasts were used as controls. Data represent means and SD of two independent experiments. (**e**) DEB-induced chromosome fragility test in human *XPF*-deficient primary fibroblasts from XP, FA and XFE patients (XP42RO, 1333 and XP51RO, respectively). WT and FAA (FA90) primary fibroblasts were used as negative and positive controls, respectively. (**f**) MMC-induced G2/M cell cycle arrest in human *XPF*-deficient primary fibroblasts from XP, FA and XFE patients (XP42RO, 1333 and XP51RO, respectively). WT and FAA (FA90) primary fibroblasts were used as negative and positive controls, respectively. (**g**) UV-induced growth inhibition of *Xpf* KO MEFs transduced with lentiviral particles carrying GFP (negative control vector), wild type XPF, XPF-R689S and XPF-L230P. Data represent means and SD of at least four independent experiments.

mild UV reactivity but less than that of XP51RO fibroblasts from the XFE patient with severe phenotype (Fig 2b,c). On the other hand, 1333 and XP51RO fibroblasts reacted like typical FA cells and were far more sensitive to MMC than XPF fibroblasts (Fig 2d). Sensitivity to ICL agents was also assessed by DEB-induced chromosome breakage assay where 1333 and XFE fibroblasts scored clearly positive but two XPF fibroblast lines negative (Fig 2e). The number of breaks per cell found in 1333 and XFE fibroblasts (4 to 8 breaks per cell after 0,1μg/ml DEB) is in the range of chromosome fragility reported in blood lymphocytes in a cohort of 66 FA patients using the same DEB concentration (range 1.4 to 10 breaks per cell; mean 4.3 breaks per cell) (216). Similar results were obtained measuring MMC-induced G2 phase arrest, where 1333 and XFE fibroblasts but not the XPF cells showed marked G2 phase accumulation in the cell cycle (Fig. 2f). These results demonstrate that XP, XFE and FA cells with XPF mutations have clearly

distinct phenotypes. This is further supported by the fact that the FA-specific XPF mutants XPF-L230P and XPF-R689S rescued approximately 80% of the UVC sensitivity of *Xpf* null MEFs (Fig. 2g), while both mutants were unable to complement MMC sensitivity as shown before in Figure 1f. Similar experiments in XPF-deficient CHO cells demonstrated that the XFE-specific XPF mutant R153P does not rescue either MMC or UV sensitivity (44).

To further investigate the extent of NER deficiency in the FA patients with XPF mutations, we measured UV-induced unscheduled DNA synthesis in primary skin fibroblasts from FA patient 1333 and from an XPF patient with mild clinical UV sensitivity (XP42RO). The results show 24±4% and 21±3% residual UDS activity, respectively (Fig. 3a). We also determined UDS in *Xpf* null MEFs expressing either FA-specific XPF missense mutants, XPF-L230P or XPF-R689S. The corresponding levels of activity were 39% and 44% (Fig. 3b). These *ex vivo* studies suggest that *XPF*-mutant FA cells retain a certain level of NER activity similar or even higher to that of XP patients with mild XPF mutations. To support this view, we expressed XPF-L230P and XPF-R689S in XPF-deficient human XP2YO fibroblasts and investigated their sensitivity to UVC-light and the repair kinetics of UV-induced (6-4)PP at sites of local UV damage. Both FA-specific XPF mutants rather efficiently corrected the defective removal of 6-4PP (Fig. 3c) and strongly improved UV-induced survival rates (Fig. 3d) of XP2YO cells. In contrast, XPF with the 28 bp duplication was completely deficient of NER activity (Fig. 3c) as predicted from the suggested disruption of the ERCC1 and DNA-binding domain of this truncated protein (XPF-I800TfsX23).

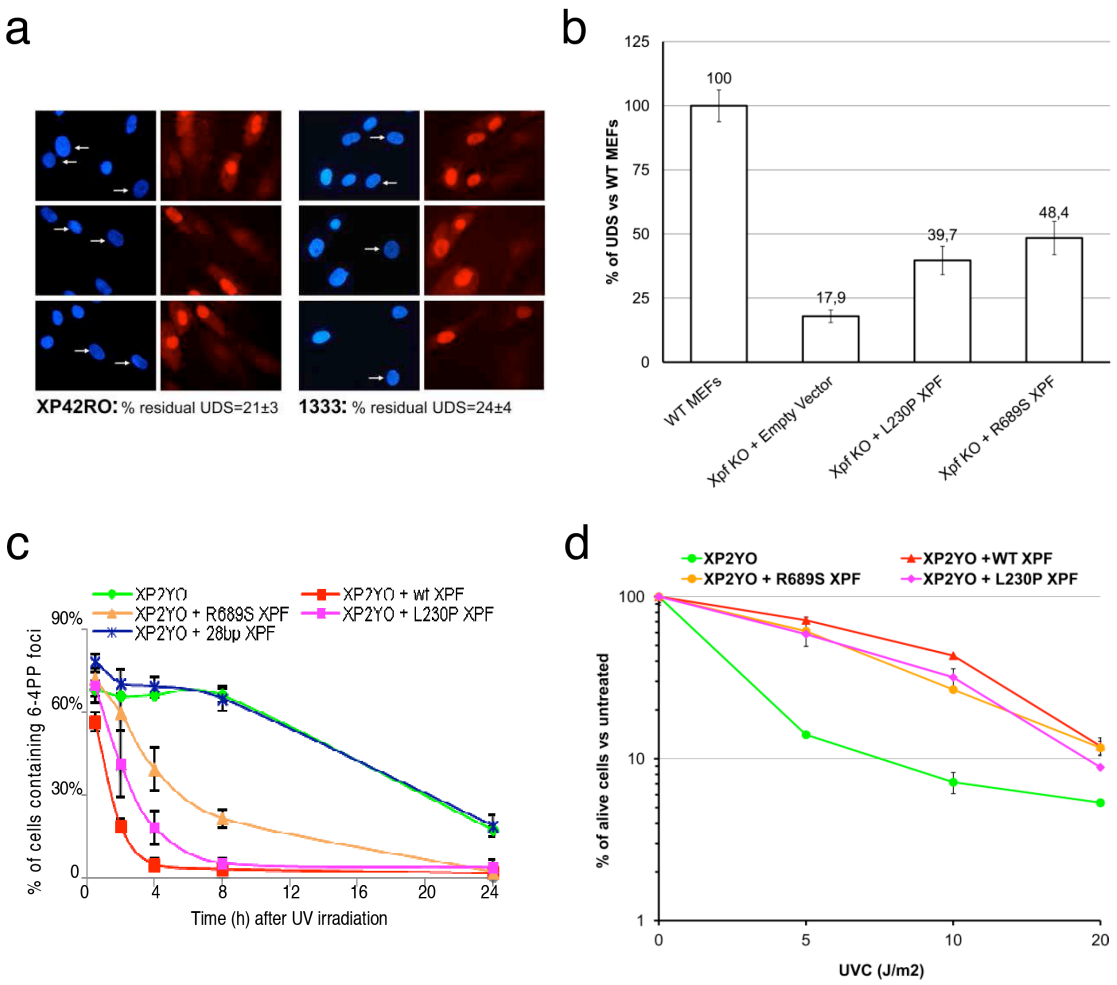


Figure 3. NER analysis of XPF mutants in vivo (**a**) Unscheduled DNA Synthesis (UDS) in primary fibroblasts representing global NER activity and measured as described in supplement. Cells (arrows) were compared to mixed-in normal fibroblasts preloaded with polystyrene microbeads (no arrows), used as an internal control. (**b**) UDS signals in *Xpf*-/- MEFs measured as in **a**, quantified from 20-40 random G1/G2 nuclei and expressed as a percentage of control wt MEFs. *Xpf*-/- cells were stably expressing an empty vector or one of various XPF cDNAs (wild type, L230P and R689S). (**c**) Repair kinetics of UV-induced DNA damage by FA-specific XPF mutants in *XPF*- and NER-deficient human cells (XP2YO). Cells expressing wild type XPF, XPF-R689S, XPF-L230P or XPF-28 bp duplication were locally UV-irradiated, cultured for 0.5, 2, 4, 8 or 24 h following irradiation, fixed and stained for (6-4)PPs. Data represent the percentage of cells with (6-4)PP spots at various time points; means and SD of at least two independent experiments are shown. For each experiment 100 cells were counted. (**d**) UVC-induced growth inhibition of human *XPF*-deficient cell line (XP2YO) transduced with lentiviral particles

carrying XPF-WT, XPF-R689S or XPF-L230P. Data represent means and SD of two independent experiments.

Cell lines from XPF patients show a characteristic failure of the mutant XPF protein to properly translocate to the nucleus, likely through aggregation of the protein in the cytoplasm (44). This feature is evident for XP-causing mutations and accentuated in cells from the patient with the XFE syndrome (3). We tested whether the FA-causing XPF-mutant proteins would localize normally. All three of the FA-associated XPF-mutant proteins were detected in the nucleus and bound to chromatin (Supplementary Fig. 4a on line). In line with this observation, endogenous (Supplementary Fig. 4b on line) or ectopically expressed (Supplementary Fig. 4c on line) mutant XPF-R689S relocated to sites of UVC-damaged DNA in FA104 lymphoblasts or XPF-deficient XP2YO fibroblasts, respectively. Proper co-localization with 6-4 PP also held, to a lesser extent, for ectopically expressed XPF-L230P but not for XPF protein that was truncated due to the 28 bp duplication. These observations show that the FA-causing XPF variants translocate to the nucleus, where they are recruited to sites of active NER.

A known ICL repair-specific feature of XPF is its interaction with SLX4/FANCP, a scaffold protein for several endonucleases which is mutated in a subset of FA patients (213,214). We investigated the interaction between SLX4 and XPF in our patients cell lines by co-immunoprecipitation and Western blotting and found that the XPF mutants in FA104 and 1333 still interact with SLX4 (Supplementary Fig. 5a on line). SLX4-MUS81 interactions were unaffected in FA104 and 1333. The SLX4-ERCC1 interaction was normal in FA104, while difficult to detect in 1333 due to the low levels of both XPF and ERCC1 in this patient cell line (Supplementary Fig. 5a) which is in keeping with the fact that XPF and ERCC1 stabilize each other (214). Quantification of ERCC1 immunolabeling signal indeed showed values of 11.2 ±

2.3 %, 21.5 ± 4.5 %, 17.1 ± 2.5 % for XFE (XP51RO), XPF (XP42RO) and 1333 fibroblast, respectively, when compared to the levels measured in wild type cells (C5RO) in the same slides (Supplementary 5b on line). Coimmunoprecipitation with ERCC1 showed more directly that the ERCC1 dimerization with the XPF-R689S and also with the XPF-L230P mutant proteins is preserved whereas no interaction was demonstrated with XPF truncated due to the 28 bp duplication, consistent with the absence of the ERCC1-interacting and DNA-binding domain (Supplementary Fig. 5c on line). These results, suggest that the XPF mutants do not cause FA through a defect in the interaction with SLX4 or ERCC1, but we cannot rule out that other, less well-characterized ICL repair-specific protein-protein interactions are affected.

We then investigated the ability of the FA-specific mutant XPF protein to cleave DNA. For this aim, we first overexpressed and purified XPF-R689S as a heterodimer with ERCC1 in insect cells. We performed NER reactions with the purified mutant protein and extracts from XPF-deficient cells using a plasmid containing an intrastrand cross-link, in order to visualize the excision fragments. Consistent with the other functional data, purified ERCC1-XPF-R689S is proficient in the excision step of NER similar to wild type XPF, as it restored the ability to cleave and remove a site-specific 1,3-cisplatin intrastrand cross-link from the plasmid in XP2YO cell extracts (Fig. 4a). Nevertheless, the excision reaction is not perfect as the excised fragment is, on average, 1 nucleotide longer than expected from a normal reaction with wild-type XPF-ERCC1 dimer (Fig.4a). We also performed *in vitro* nuclease activity assays with purified ERCC1-XPF-R689S on a stem-loop model DNA substrate. Contrary to global NER, XPF-R689S is unable to cleave such a substrate (Fig. 4b). This is unlike wild type XPF and XPF mutants causing XP (R799W) or XFE progeria (R153P) (44), indicating that the nuclease-type activity of XPF-R689S is grossly abnormal. Notably, the NER excision reaction and nuclease activity on

stem loop DNA substrate of XPF-R689S are altered in a way identical to that of the XPF artificial mutant R689A (Fig 4a,b), which was investigated in detail in biochemical studies on conserved amino acids in the XPF nuclease motif (37,38,196). Since the nuclease activity of XPF-R689S is strongly but the NER reaction only slightly affected, we hypothesize that the exact positioning of XPF on some of its substrates may be influenced by the mutation of FA104 rather than the catalytic activity of XPF itself. Apparently, this defect can be partially compensated by additional factors in the framework of NER, but not in an ICL repair complex. Unfortunately, we could not perform these biochemical experiments with the XPF-L230P mutant, as we were unable to express and purify ERCC1-XPF-L230P due to its low stability and tendency to aggregate, which is consistent with the lower levels of this mutant protein in 1333. We finally checked whether the FA-specific mutant XPF proteins ectopically expressed in *Xpf* null MEFs can perform the incision step of ICL repair needed for ICL unhooking. Cells were treated with melphalan and let recover for various times and the incision reaction was measured by the COMET assay as previously described (217). Both XPF-L230P and XPF-R689S completely restored the incision/unhooking defect of *Xpf* null MEFs (Fig. 4c), while the cells remained hypersensitive to ICL as shown before (Fig 1f). This indicates that ICL-sensitivity of FA cells with XPF mutations is not directly linked to an unhooking defect, which is in agreement with the results reported in other FA subtypes (218).

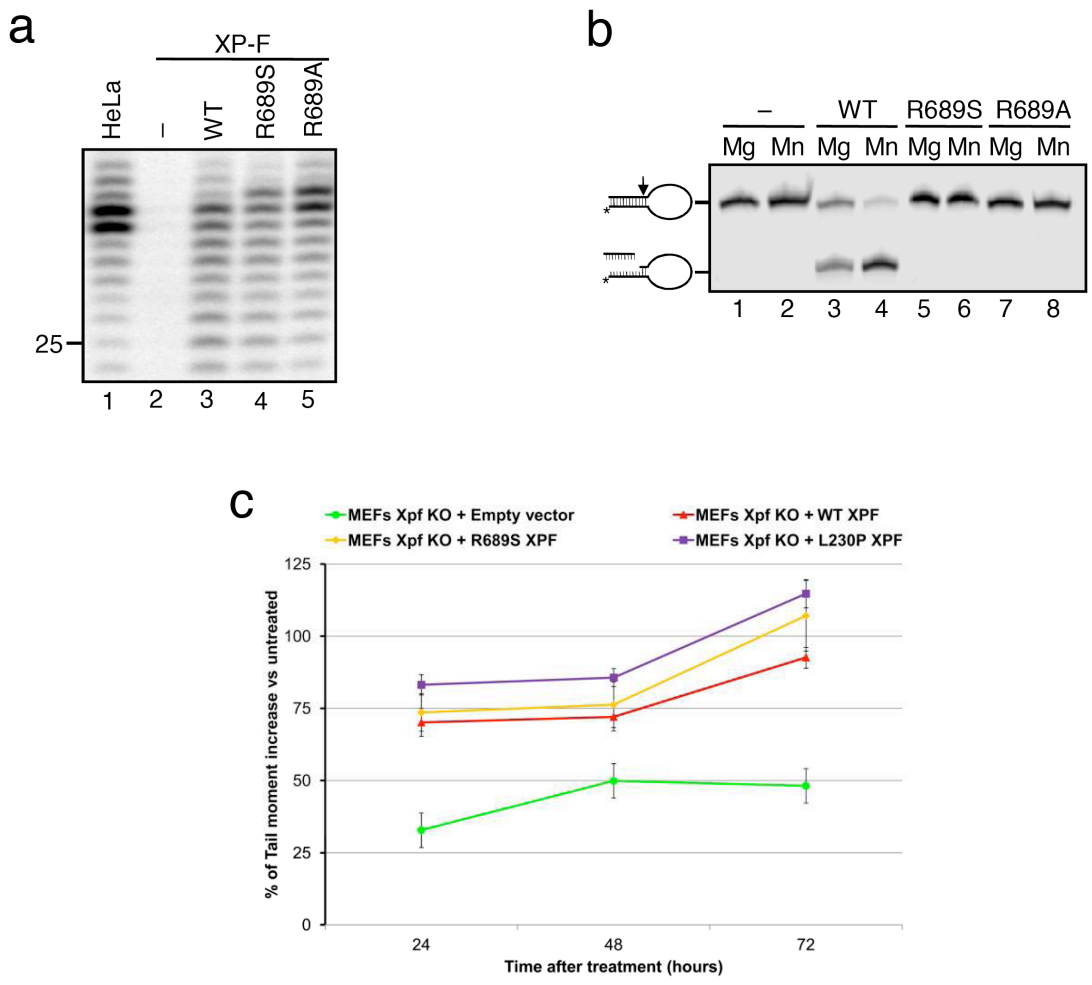


Figure 4. Excision nuclease and unhooking activities of XPF mutants. (**a**) NER activity of wild type and mutant ERCC1-XPF dimer. A plasmid containing a site-specific 1,3-intrastrand cis-Pt DNA cross-link was incubated with whole cell extracts from HeLa cells or XPF-deficient cells (XP2YO) complemented with recombinant ERCC1-XPF. The excised DNA fragments of 24-32 nucleotides are shown. The position of a 25mer is indicated. (**b**) Incision of a stem-loop substrate with wild type and mutant *XPF*. The 3' Cy5-labeled substrate was incubated with recombinant ERCC1-XPF in the presence of 2 mM $MgCl_2$ or 0.4 mM $MnCl_2$, and the products analyzed by denaturing PAGE. (**c**) Unhooking/incision activity of *Xpf* KO MEFs transduced with lentiviral particles carrying GFP (negative control vector), wild type XPF, XPF-R689S and XPF-L230P after melphalan-induced ICLs. Tail moment from 200 random nuclei was scored and expressed as a percentage of the untreated control. Error bars are SEM.

The role of nucleases in ICL repair is not well understood yet and an increasing number of nucleases have been linked to the FA pathway. The recently reported FANCD2 associated nuclease FAN1 (85-88) seems to have a minor role in ICL repair and lack of FAN1 does not cause FA in humans (219,220). Recent data suggest that XPF-ERCC1 endonuclease and the hSNM1A exonuclease act in the same pathway, together with SLX4, to initiate ICL repair, with the MUS81-EME1 fork incision activity becoming important in the absence of the XPF-SNM1A-SLX4-dependent pathway (84). The nuclease hSNM1B/Apollo has also been linked to the Fanconi anemia pathway via its interaction with FANCP/SLX4 (221). While the defect in FA104 and 1333 is not in the incision step, it seems unlikely that the defect is a downstream step of homologous recombination since FA104 and 1333 cells are not sensitive to PARP inhibitors and normal in Rad51 focus formation. Given that the nuclease activity of the FA-specific XPF mutants is abnormal, it is tempting to speculate that the ICL-unhooking step in FA cells leaves an intermediate aberrant substrate, which is unrepairable by subsequent ICL repair factors.

Our genetic, biochemical and functional studies, along with the characterization of previous XPF and XFE patient alleles, provide a model for the mechanistic understanding of how mutations in *XPF* lead to three distinct diseases (Fig. 5; supplementary Table 2). Most of the presently known XPF patients suffer from a relatively mild form of XP (94). Cells from these patients have a reduced level of XPF protein in the nucleus as the mutant XPF protein has the tendency to aggregate in the cytoplasm (44). This reduced level of nuclear XPF is insufficient to mediate complete NER functions, while it still has enough ICL repair-specific functions in preventing chromosome fragility, cell cycle arrest and subsequent FA clinical manifestations. A second set of *XPF* mutations, characterized in our study, allow localization of the protein to the nucleus, where they exert a certain level of NER activity but are fully deficient in ICL repair.

XPF-R689S is a stable and NER-proficient protein with an active site structure that prevents it from properly processing ICL repair intermediates. XPF-L230P is more similar to previously described XPF mutations in that it is less stable and might have a tendency to aggregate in the cytoplasm. However, sufficient amounts of the protein are properly folded and reach chromatin where it appears to have some activity in the removal of (6-4)PPs. Residual NER activity in the skin tissue of 1333 patient *in vivo* may explain why this patient has no clinically relevant skin photosensitivity although we cannot exclude that dermatological problems will arise later in life. A final category of *XPF* mutations is associated with the progeroid syndrome XFE (44), which is characterized by very low levels of nuclear XPF, apparently insufficient to support either NER or ICL repair. Importantly, the only XFE patient described suffered from both skin photosensitivity and anemia (44,94), and shared some cellular features with XP (NER defect, UV-sensitivity) and FA (extreme ICL-sensitivity, DEB-induced chromosome fragility and MMC-induced cell cycle arrest) suggesting that XFE is characterized by a combination of XP and FA manifestations (Supplementary Table 2). Exhaustion of hematopoietic stem cells is also an attribute of ERCC1-XPF hypomorphic mice that mimic XFE (Laura Niedernhofer, personal communication). Microsomy, microcephaly and liver fibrosis are likewise observed in FA patient 1333, in *Ercc1*- and *Xpf*-deficient mice and in the unique ERCC1-deficient patient, which all lack NER and ICL repair functions (4,131,137,199,222).

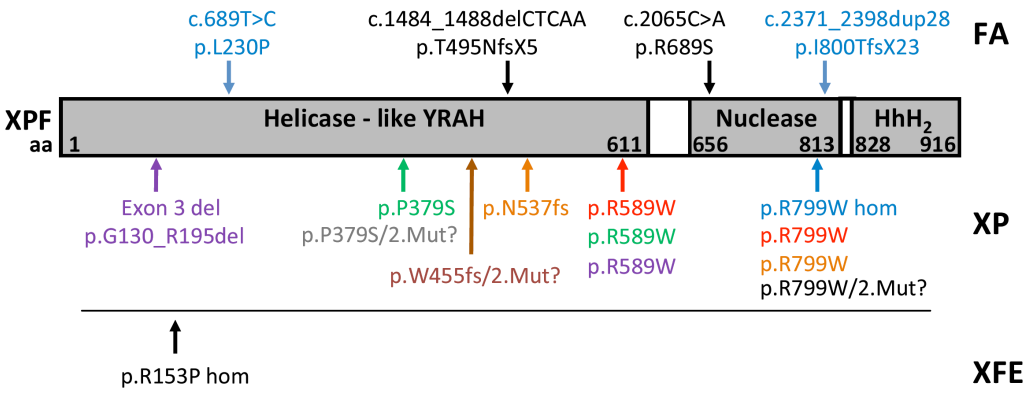


Figure 5. XPF mutations lead to three rare DNA damage response syndromes. XPF mutations previously reported in XPF and XFE patients are shown together with the mutations in two FA patients described in this study. Mutations found in the same patient are shown in the same color. The approximate position of the amino acid change within the *XPF* locus is indicated with arrows.

In a broader sense, this study demonstrates that depending on the type of *XPF* mutation and the balance between NER and ICL repair activities, patients present with one of three clinically distinct disorders. This resembles the *XPD* gene, which is involved in XP subtype D, trichothiodystrophy or Cockayne syndrome depending on the type of mutation, and highlights the value of characterizing rare genetic disorders to gain insight into the mechanisms of genome maintenance and human disease. XPF/FANCQ has a central role in preventing genome instability, cancer, BMF, developmental abnormalities and premature ageing. Like other breast and ovarian cancer susceptibility genes mutated in FA (*BRCA2/FANCD1*, *BRIP1/FANCJ* and *RAD51C/FANCO)* (223,224), XPF/FANCQ also acts downstream FANCD2 monoubiquitination. Therefore, it is important to study *XPF/FANCQ* as a candidate gene in hereditary breast and ovarian cancer.

SUPPLEMENTARY

MATERIALS AND METHODS ON LINE

Cell lines

The human EBV-transformed lymphoblastoid cell lines FA139 (wt), HSC93 (wt), FA104, 1333, HSC536 (FA-C), EUFA1354 (FA-P), 475/3 (FA-P) and GM02344 (XPA) were cultured in RPMI medium supplemented with 15% heat inactivated FCS and plasmocin (2,5 µg/ml, Invivogen,). XPA EBV-transformed lymphoblastoid cell line GM02344 was purchased from Coriell Cell Repositories. Immortal Xpf-KO mouse embryonic fibroblasts, human GM08437 (XP2YO), GM16633 (PD20), GM16634 (PD20-3-15) and 1333 SV40 immortalized fibroblasts and stable transfectants were cultured in DMEM supplemented with 10% FCS and plasmocin (2,5 µg ml-1, Invivogen).. Primary fibroblasts XP42RO, XP51RO, GM04313 (XP2YO) and 1333 were cultured in DMEM supplemented with 15% FCS and plasmocin (2,5 µg ml-1, Invivogen).GM08437 (XP2YO), GM16633 (PD20), GM16634 (PD20-3-15) GM04313 (XP2YO primary) human fibroblast were purchased from Coriell Cell Repositories Xpf-KO MEFs were a kind gift from Dr Maria A. Blasco (Centro Nacional de Investigaciones Oncológicas, Madrid).

Chromosome fragility assays

Chromosome fragility assays were performed on peripheral blood with DEB or MMC using standard cytogenetic procedures (216).

Sequence analysis

For direct sequencing on genomic DNA, PCR products were generated by XPF specific primers (Supplementary Table 3). PCR fragments were treated with Shrimp Alkaline Phosphatase and Exonuclease I for 30 min at 37°C and 15 min at 80°C according to the manufacturer's instructions (Amersham Biosciences). Sequencing reactions were carried out using 10 pmol of primer and the Big Dye terminator cycle sequencing kit (Applied Biosystems) in the Gene Amp PCR system 9700 (Applied Biosystems). Samples were analyzed in an ABI 3730 DNA analyzer (Applied Biosystems).

Generation of reverted cell lines FA104R and 1333R

Lymphoblastoid cells were cultured under constant selection with 15 nM MMC (FA104) or 40 nM MMC (1333). The MMC resistant phenotype was confirmed by a growth inhibition test with different concentrations of MMC.

Molecular cloning of single alleles in FA104R

Exon 8 and 11 of XPF were amplified from DNA of cell line FA104R, PCR products were cloned with the Topo TA Cloning kit (Invitrogen) and transformed in Library efficient DH5alpha bacterial cells (Invitrogen). The plasmids from single bacterial colonies were prepared with the NucleoSpin® Plasmid QuickPure Kit (MACHEREY-NAGEL) and sequenced as described before.

Lentiviral vectors and subtyping by complementation analysis

Lentiviral supernatants were produced as previously described (225,226). Subtyping of lymphoblastoid cell lines (LCL) from FA104, 1333 and an FA-A patient was based on the

correction of the MMC-hypersensitivity of cells transduced with lentiviral vectors expressing GFP, FANCA, XPF-wt or XPF-R689S. Briefly, LCLs were placed into fibronectin-coated wells (2 µg cm-2, Takara Shuzo). Two rounds of viral infection for 24 h were performed to achieve transduction levels higher than 50%. After the last infection, cells were collected and exposed to increasing concentrations of MMC (0 to 1,000 nM) in fresh medium. Ten days later, cell viability was determined by flow cytometry (EPICS, Coulter Elecronics) as previously described (215). MEFs and SV40 immortalised human fibroblasts were seeded in six well plates infections for 24 h were performed in presence of 1µg ml-1 of polybrene (Millipore) with lentiviral particles carrying GFP, XPF-wt, XPF-R689S and XPF L230P cDNAs.

UVC sensitivity assay in lymphoblasts

For UVC sensitivity, 1.5 million cells were washed twice in PBS and resuspended in 3 ml of PBS with 3% FCS. Cell suspension was transferred to 10 mm Petri dishes and distributed uniformly by gentle agitation. Cells were then irradiated at the indicated doses (each point in duplicate). Cells were recovered after the addition of 3 ml of complete RPMI and centrifuged. Cellular pellets were resuspended in 5 ml of complete RPMI and left to recover at 37 ºC in 5% CO2. After 24 h, apoptotic cells were stained with the Annexin-V-FLUOS Staining Kit (Roche) according to the manufacturer's instructions and 20,000 cells were evaluated by flow cytometry (FACS Canto, BD Biosciences). Results were analyzed with the Flowjo software (Tree Star, Inc.) and expressed as percentage of non-apoptotic cells (negative for both green (annexin) and red (Propidium iodide) signals) with respect to the untreated controls.

UVC and MMC survival assays in fibroblasts

For UVC and MMC sensitivity assays cells were seeded in duplicate, allowed to attach for 24 h, rinsed with PBS, and UVC irradiated with a Philips 15W UVC lamp G15-T8 or treated with MMC (Sigma) containing medium at the indicated doses. After 3 (primary fibroblasts) or 6 (MEFs and immortalized human fibroblasts) population doublings of the untreated controls, the cells were washed twice with PBS and harvested by trypsinization. The number of cells in the resulting cell suspension was counted in a Beckman-Coulter cell counter. Results are expressed as percentage of viable cells with respect to the untreated controls.

Unschedule DNA synthesis (UDS) assays

UDS was measured using 5-ethynyl-deoxyuridine, grossly as previously described (227). In short, mouse embryonic or human fibroblast cultures were exposed to 16 J/m2 of 254nm UV-light and incubated in medium with 20 μM 5'-Ethynyl-deoxyuridine (EdU) (ClickIt kit, Invitrogen), 10% dialyzed fetal bovine serum and 1μM 5-fluorodeoxyuridine for the next 3 hrs. After a 15min chase in medium with 10μM thymidine, the cells were fixed with 3.7% formaldehyde and permeabilised with 0.5% triton-X100 before chemical coupling of Azide-Alexa Fluor 594 to the ehtynyl-groups using monovalent Cu according to the manufacturer's prescription (Invitrogen). Fluorescence was quantified in 20-40 random non-mitotic nuclei with S-phases excluded.

Cell fractionation

Separation of cytoplasmic, nuclear and chromatin fractions was achieved with the Subcellular Protein Fractionation Kit from Pierce (Thermo Scientific) following the manufacturer’s instructions. Aliquots equalling 25 μg protein of each extract were loaded on 4-

12% BisTris gels. The following antibodies were used: antibody to XPF, ab17798; antibody to tubulin, ab44928; antibody to p300, ab3164; antibody to histone H3, ab1791, all from Abcam.

Local UVC irradiation of lymphoblastoid cells

Lymphoblasts were resuspended in 50% FBS/RPMI, allowed to attach to polylysine coated slides at 37 ºC for 1 h and washed with PBS. After removing the PBS, cells were covered with an isopore polycarbonate filter with pores of 5 μm diameter (Millipore), exposed to 60 J m-2 of UVC and processed for pyrimidine dimers immunodetection essentially as previously reported (227) using a mouse primary antibody to cyclobutane pyrimidine dimers (CPD) (Kamiya biomedical company) and rabbit antibody to XPF (ab76948, abcam) overnight at 4 ºC. The working concentration of these antibodies was 1:500 and 1:50, respectively.

Immunoprecipitation and immunoblotting

For immunoprecipitation, lymphoblasts (10 million) were lysed on ice with lysis buffer ((50 mM Tris HCL (pH 7.4), 150 mM NaCl and 1% Triton X-100, supplemented with Complete, EDTA-free Protease Inhibitor Cocktail tablets (Roche) and PhosSTOP Phosphatase Inhibitor Cocktail tablets (Roche)). Immunoprecipitations with antibodies to SLX4 or ERCC1 were performed for 2 h at 4 °C, followed by incubation for 30 min with Protein A/G Plus Agarose (Santa Cruz Biotechnology SC-2003). Precipitated proteins were separated on a NuPAGE® 3-8% Tris-Acetate precast gel (Invitrogen) and transferred to PVDF membrane Immobilon P (Millipore). The primary antibodies used for immunoprecipitation and protein blotting were as follows: polyclonal antibody raised against the first 300 amino acids of SLX4 (gift from J. Rouse, Dundee), antibody to XPF (Thermo MS-1381), antibody to ERCC1 (Santa Cruz FL297

and Acris-AP17002PU-N) and antibody to MUS81 (Abcam ab14387). For XPF immunoblots, cells were harvested by centrifugation and lysed in RIPA lysis buffer (Millipore, 0.5 M Tris-HCl, pH 7.4, 1.5 M NaCl, 2.5% deoxycholic acid, 10% NP-40, 10 mM EDTA) supplemented with benzonase (8 U ml-1, Novagen) and Complete, Mini, EDTA-free Protease Inhibitor Cocktail (Roche). Total protein concentration was measured by spectrophotometry with the Bio-Rad Protein Assay (Biorad). Forty μg of total protein was separated on a 6% SDS/PAGE gel, blotted onto nitrocellulose membranes (Biorad) and incubated overnight with the appropriate primary antibody. The following antibodies were used in 5% nonfat dried milk in TTBS: rabbit antibody to XPF (ab76948, abcam) 1:2,000; mouse antibody to XPF Ab-1 (Clone: 219, Thermo Scientific-Lab Vision), 1:500; antibody to XPF (Abcam ab17798), 1:200; antibody to P84 (Abcam ab487), 1:300; and antibody to Vinculin (ab18058, abcam), 1:2,500.

Local UV irradiation and (6-4)PP repair

Cells were plated on a coverslip, grown for 2-3 days and irradiated through polycarbonate membrane with 5 μm pores (millipore) by UV light (254 nm) with a dose of 150 J m-2 (XP2YO cells). Cells were incubated at 37 °C for 30 min to 24 h, washed with PBS and PBS plus 0.1% triton X-100, and fixed by 3% paraformaldehyde plus 0.2% triton X-100. After fixation, cells were washed with PBS containing 0.2% triton X-100. To stain (6-4)PPs, cells were treated with 0.07 M NaOH in PBS for 5 min, followed by washing with PBS plus 0.2% triton X-100. After blocked with PBS plus 5 mg ml-1 BSA and 1.5 mg ml-1 glycine, cells were stained with mouse monoclonal antibody to (6-4)PPs (Cosmo Bio) 1:400, and rabbit polyclonal antibody to HA-ChIP grade (abcam) 1:3,000 for 1.5 h and washed with PBS with 0.2% triton X-100. Cells were then incubated with secondary antibodies Cy3-conjugated affynipure goat antibody to mouse

lgG(H+L) (Jackson ImmunoResearch) 1:1,000 and Alexa Fluor 488 F(ab')2 fragment of goat antibody to rabbit IgG (H+L) (Invitrogen) 1:800 for 1 h, followed by washing with PBS with 0.2% triton X-100. Samples were embedded in Vectashield Mounting Medium with 1.5 μg ml^{-1} of DAPI (Vector Laboratories) and analyzed using confocal microscope (Zeiss LSM 510). About 100 cells were counted in at least two independent experiments for quantification.

Interstrand crosslink unhooking by the COMET assay

For each time point, cells were treated in duplicate with Melphalan at the indicated concentration in complete culture medium. After one hour, Melphalan containing medium was removed, cells were washed twice with PBS and allowed to recover for the indicated times in fresh growth medium at 37ºC, 5%, CO2. Cells were then harvested washed once in PBS and resuspended in freezing medium (90% FCS, 10% DMSO) at a concentration of 720.000 cells ml-1 and kept frozen at -80ºC until the moment of the Comet assay. Immediately before the electrophoresis, thawed cells were irradiated with 30 Gy and 25 ml of the cell suspension was added to 225 ml of LMP agarose (0,75% in PBS without Ca++ and Mg++, 10mM EDTA). For each time point replica, three 7 μl drops of the cell suspension in LMP agarose (corresponding to about 500 cells) were pipetted on a GelbondTM membrane (Lonza) where they were allowed to solidify. The membrane were then submerged in lysys buffer (1% Triton-X, 20% DMSO, 1% Sodium Lauryl Sarcosinate, 2,5 M NaCL, 0,1 M EDTA, 10 mM Tris, 0,2 M NaOH, pH 10) for 1 hour at 4 ºC. Membranes were than washed in electrophoresis buffer (0,3 M NaOH, 1 mM EDTA, pH 13.2) at 4ºC for 5 minutes and then left in the electrophoretic chamber in fresh electrophoresis buffer for 35 minutes at 4ºC (unwinding). Electroforesis was then performed for 20’ at 20 V and 300 mAmp. Membranes were washed once in ethanol and then left overnight in

fresh absolute ethanol at 4ºC (fixation). Membranes were allowed to dry for 2 hours at RT, rehydrated for 20 minutes in TE buffer pH 7.5 in presence of the SYBRGold fluorochrome (1:10000, Invitrogen). Membranes were washed in distilled water to remove the fluorochrome in excess, allowed to dry and mounted to be analysed by fluorescence microscopy. For each time point replica, the Olive Tail Moment of 100 cells (200 cells in total) was measured with the Komet 5.5 software (Andor Technology). The percentage of unhooking activity was measured according to the formula: (TMmi − TMc)/(TMci − TMc) × 100, where TMmi is the mean tail moment at a given time of the Melphalan-treated, irradiated sample, TMc is the mean tail moment of the cells without any treatment and TMci is the mean tail moment of the irradiated-only control sample.

XPF plasmids and site-directed mutagenesis

Site-directed mutagenesis was used to introduce point mutations in pFastBac1-XPF using the QuickChange method (Stratagene) as described (38) using the following primers:

F-XPF(R689S): GCATAGTTGTGGATATGAGTGAATTTCGAAGTGAGCTTCC

R-XPF(R689S): GGAAGCTCACTTCGAAATTCACTCATATCCACAACTATGC

XPF(L230P)(sNheI)-F:

ATGCTAGCTATACAGACTGCTATACCGGACATTTTAAATGCATGTCTAAAG

XPF(L230P)(sNheI)-R:

CTTTAGACATGCATTTAAAATGTCCGGTATAGCAGTCTGTATAGCTAGCAT

Note that all pFastBac1-XPF constructs were the 905 amino acid form of XPF (1) lack the N-terminal 11aa of XPF, which is fully functional in NER in vitro and in vivo. To generate 28 bp duplication allele, two PCR reactions were conducted using the pWPXL-XPF vector as the

template. CGATAATCCGGACTTACACTTCACTTCCCCAGACTACGG and GAAGTACTCCGGATGCAGCTCTCGGGC were as two primers for one reaction; the other reaction used the primers GAGCTGCATCCGGAGTACTTCAAG and GTTAACTCCGGAGTCTGGGGAAGTGAAGTGTAAGAAG. Two pieces of the DNA fragment were digested with restriction enzyme BmpE1 to generate the sticky ends. After ligation the two digested fragments, the products were used to transform XL1-Blue cells. To add the HA tag at the C-terminal of the protein, the newly generated pWPXL-XPF 28 bp was used as a template and TAATATACGCGTATGGAGTCAGGGCAGCCGGCTCGACGGATTGCCATGGCGCCGCT GCTGGAGTAC and AATCTAACTAGTTCAAGCGTAATCTGGAACATCGTATGGGTAAACAACTCCGCCGTT GCATGAGG as primers for a second PCR reaction. Afterwards, the product and pWPXL empty vector were digested with restriction enzyme MluI and SpeI separately and ligated together, generating pWPXL-XPF-28bp-HA. The coding sequence of all new plasmids was confirmed by sequencing.

Purification of ERCC1-XPF

Baculovirus production was done as described (38) using the pFastBac1-XPF constructs generated. XPF wild-type, XPF-R689A, and XPF-R689S were co-expressed with wild-type ERCC1 in Sf9 insect cells. The heterodimers were purified over nickel affinity, size-exclusion and heparin chromatography as described (38). Protein concentrations ranged from 0.1 to 0.3 mg ml-1, XPF-R689S was eluted from the heparin column with 600 mM NaCl, and the final protein concentration of XPF-R689S was 0.15 mg ml^{-1}.

In vitro NER Assay

XPF-deficient cell extract and the plasmid containing site specific 1,3-intrastrand d(GpTpG) cisplatin (cis-Pt) lesion were made as previously reported (228,229). For each reaction, 2 µl of repair buffer (200 mM Hepes-KOH, 25 mM MgCl2, 2.5 mM DTT, 10 mM ATP, 110 mM phosphocreatine, 1.8 mg ml-1 BSA, making the final pH 7.8), 0.2 µl of creatine phosphokinase buffer (2.5 mg ml-1 creatine phosphokinase, rabbit muscle, sigma, 10 mM glycine, pH 9.0, 50% glycerol), 3 µl of XPF-deficient cell extract, NaCl (to a final concentration of 70 mM), and purified proteins ((33.5 nM) in a total volume of 9 µl were pre-warmed at 30 °C for 10 min. 1 µl plasmid containing Cis-Pt (50 ng ml-1) was added to each reaction and the reactions were incubated at 30 oC for 45 min. The excised DNA fragments of 24-32 nucleotides were detected by annealing a complementary oligonucleotide containing a non-complementary 4G overhang and filling in with α-32P-dCTP. For this, the reaction mixture was then cooled on ice for 5 min, followed by addition of 0.5 µl of 1 µM complementary strand (5'GGGGGAAGAGTGCACAGAAGAAGACCTGGTCGACCp-3'), the reaction mixture was then denatured by heating at 95 oC for 5min. Following 15 min of annealing at room tempereature, 1 µl sequenase mixture (containing 0.13 units of sequenase, and 2.0 µCi [α-32P] dCTP for each reaction) was added. After pre-incubation at 37 oC for 3 min, 1.2 µl dNTP mixture (50 µM dCTP, 100 µM dTTP, 100 µM dATP and 100 µM dGTP) was added. The reaction mixture was incubated at 37oC for 12 min and the reaction was stopped by adding 8 µl loading dye (80% formamide and 10 mM EDTA). After 5 min of heating at 95oC and cooling down on ice, samples were loaded onto 14% denaturing polyacrylamide gel. Gels were run at 45 W for 1 h and visualized by PhosphorImager (Typhoon 9400, Amersham Biosciences).

Nuclease activity assay

10 pmol oligonucleotide (GCCAGCGCTCGG(T)22CCGAGCGCTGGC) labeled by fluorescent dye Cy5 at 3' site (IDT) were annealed in 200 µl solution (10mM Tris pH 8.0, 50 mM NaCl, 1 mM MgCl2) by heating at 90 oC for 10 min and slow cooling to room temperature over 2 h. 100 fmol of annealed substrate were used in each reaction. The reaction contained 25 mM Tris pH 8.0, 2 mM MgCl2 or 0.4 mM MgCl2, 10% glycerol, 0.5 mM β-mercaptoethanol, 0.1 mg ml-1 BSA, 40 mM NaCl and 100 fmol of various proteins, in a total volume of 15 µl. The reactions were incubated at 30 oC for 30 min and stopped by adding 10 µl 80% formamide/10mM EDTA. After heating at 95 oC for 5 min and cooling on ice, 3 µl of each sample were loaded on to 12% denaturing polyacrylamide gels. Gels were run at 50 oC for 40 min and visualized by fluorescence by Typhoon 9400 (Amersham Biosciences).

SUPPLEMENTARY TEXT

FA patients and controls

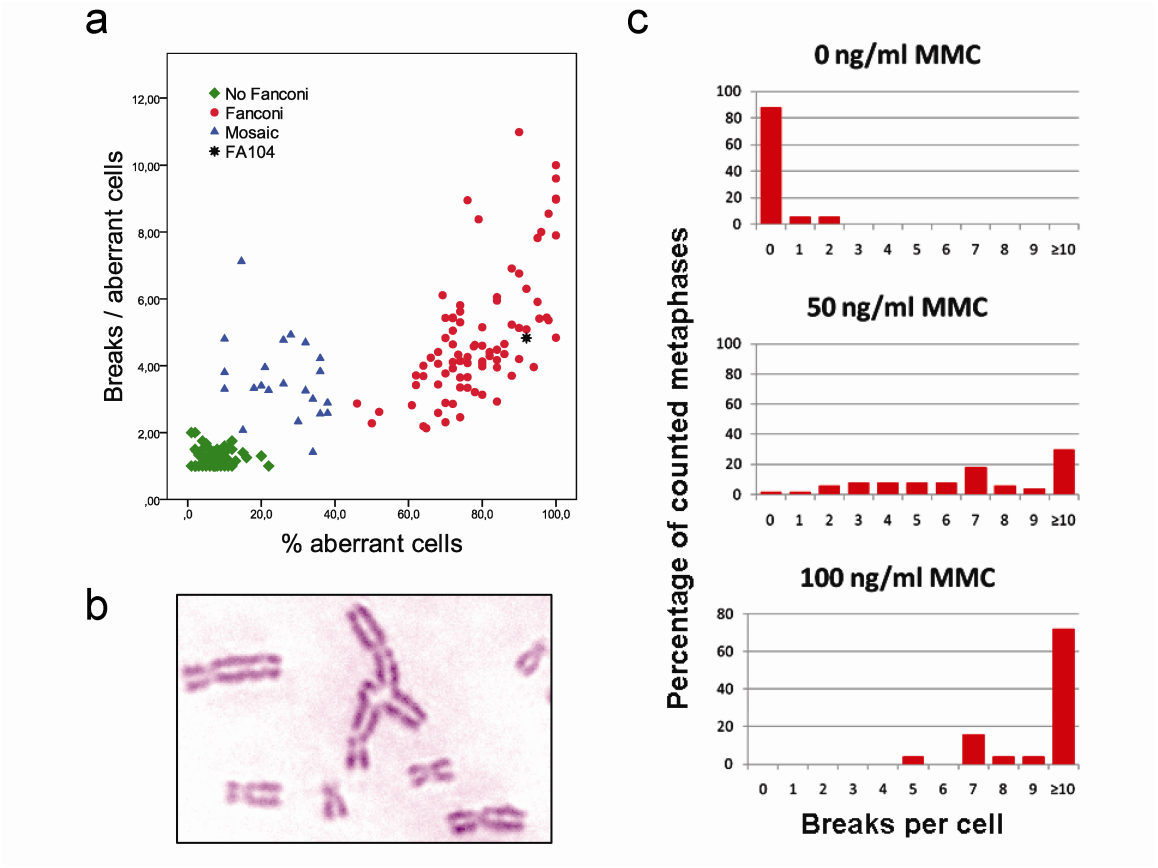
FA104 was born in 2000 from unrelated parents and diagnosed neonatally due to a malformation syndrome suggestive for FA, including absence of both thumbs, microsomy, oesophagus atresia, anterior anus, hypothyroidism, and dysmorphic and low implanted ears. The patient did not present with dermatological abnormalities such as skin hyperpigmentation,

photosensitivity or atrophic epidermis. The onset of the haematological disease was at the age of 2 years, with aplastic anaemia affecting the three blood lineages leading to bone marrow failure. The patient underwent bone marrow transplantation at the age of four years and died due to a hemorrhagic shock after the transplant. The chromosome fragility test was clearly positive as 92% of cells showed DEB-induced breaks, with an average of 4.44 breaks per cell, which confirmed the FA diagnosis. Chromosome fragility data of this patient are shown in supplementary Fig.1 in the frame of a Spanish chromosome fragility dataset1. Patient 1333 was born in 2002 and was diagnosed with FA at the age of 5 years. The clinical symptoms include low birth weight and postnatal growth retardation leading to short stature, pronounced microcephaly, café-au-lait spots, ostium primum defect, bilary tract hypoplasia with fibrosis of the liver and bone marrow failure. No skin photosensitivity or atrophic epidermis was reported in this patient. The chromosome fragility test was clearly positive with 0.2, 6.7 and 9.4 breaks per cell at 0, 50 and 100 ng ml^{-1} MMC, respectively (Supplementary Fig. 1c). Controls consisted of 200 healthy representatives of the Spanish population, mainly recruited from the Menopause Research Centre at the Instituto Palacios (Madrid, Spain) and from the College of Lawyers (Madrid, Spain). Details of this control series have been previously published (230). The necessary ethics committee approval was obtained as well as informed consent from all participants in the study.

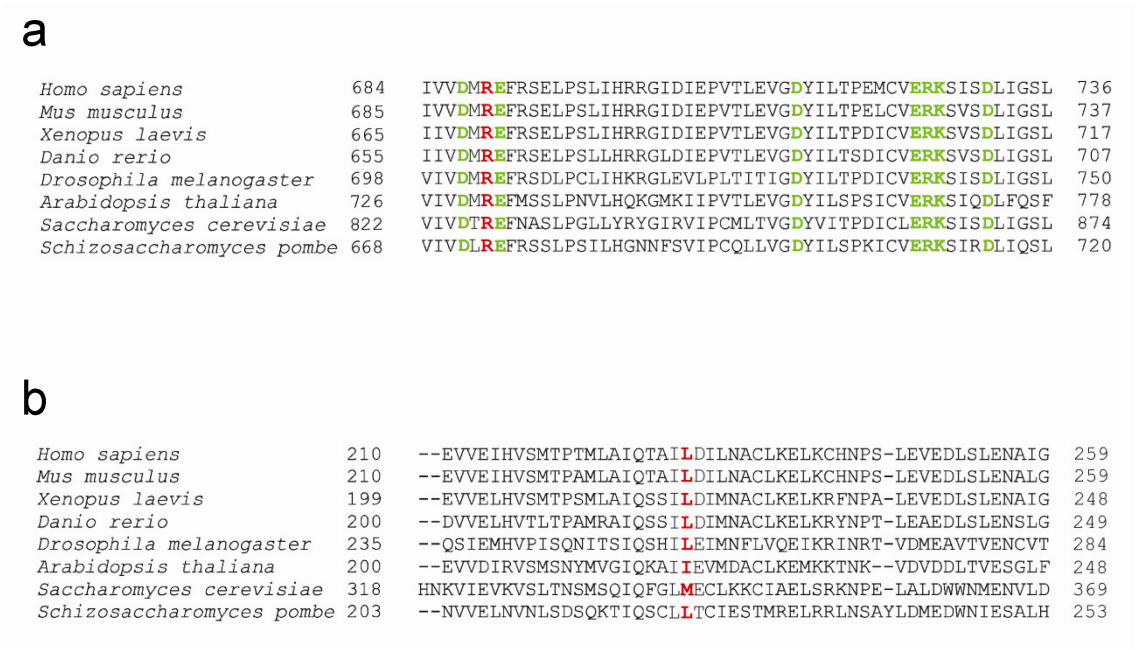
Whole exome capture, next-generation sequencing, and bioinformatics analysis pipeline

Genomic DNA was isolated from peripheral blood. Fragment sequencing libraries were prepared from 3 μg of DNA according to the SOLiDTM 4 library preparation protocol. Exon targeted enrichment was performed using Agilent's SureSelect Human All Exon Target

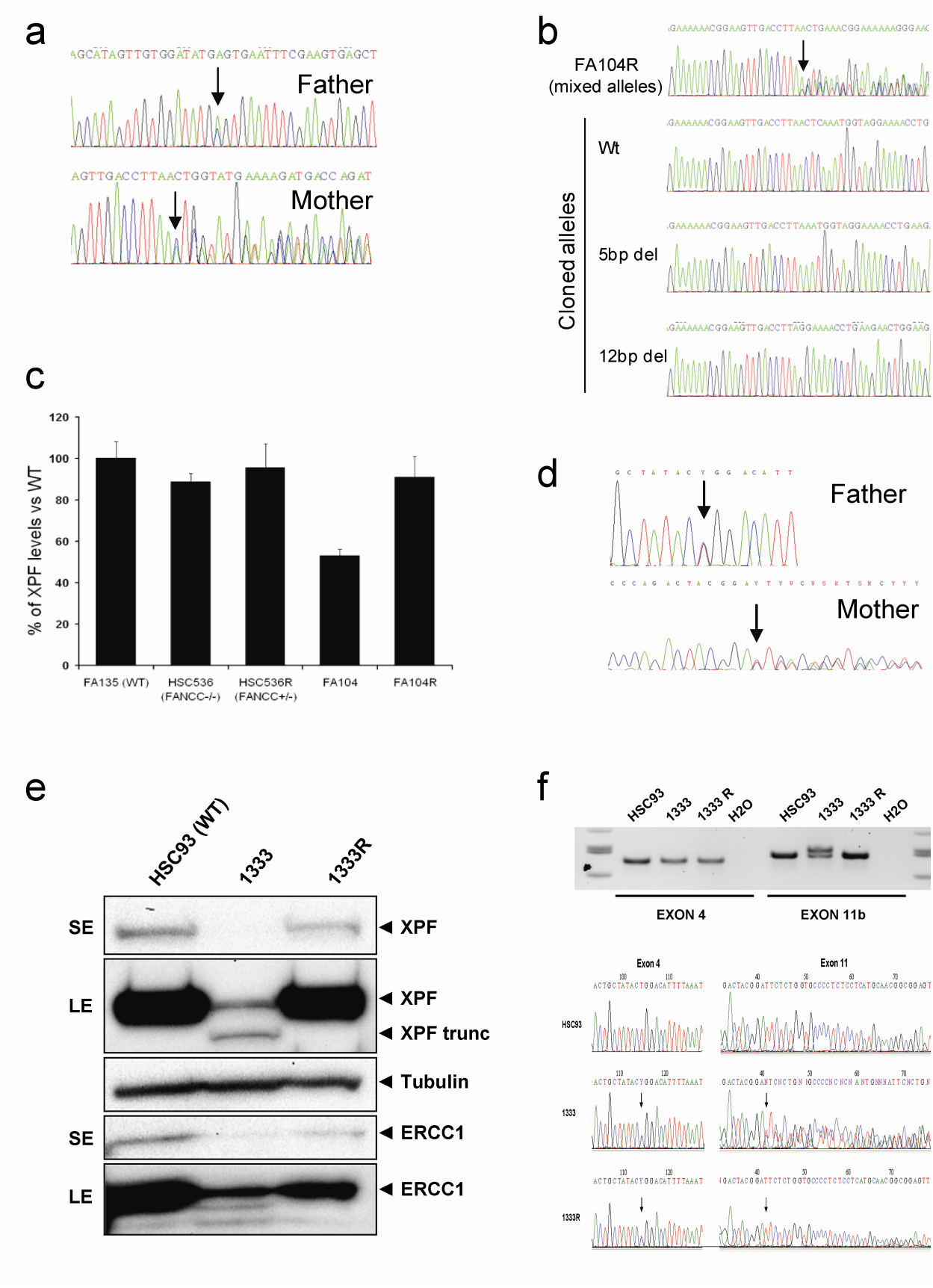
Enrichment System for 38 Mb followed by emulsion PCR amplification and sequencing according to the standards for SOLiDTM 4. 107.8 million beads were deposited in a quarter of a sequencing slide, which produced a total of 208.265.974 asymmetric paired-end reads (50nt and 25nt per pair). Color-space reads were aligned against the Human reference genome (NCBI36/hg18) using Bioscope v1.2 [http://solidsoftwaretools.com] with default settings. 74.25% of the total generated reads (5,7 Gb) mapped along the genome with 63.16% of the mapped reads being on target (3,6 Gb). SNVs were called with Samtools (231) with the following filtering criteria: SNV quality of 20 (Phred-like score), genotype quality of 30 (Phred-like value) and minimum coverage of 9x. Indel calling was performed with the Small Indel tool, part of the Bioscope v1.2 analysis suite. A unique filter of at least 9 non-redundant reads was applied for indel detection. For prioritization, identified variants were annotated and classified according to their position or effect on transcripts using the Application Programming Interfaces (APIs) from Ensembl (232) and scripts 'in house' written in Perl to query the Ensembl database (release 59) which includes data from dbSNP, the 1000Genome project and the HapMap project along with other sources. Novel variants that resulted in a non-synonymous coding (NSC), a frameshift coding (FC) or a non-frameshift coding (NFC) change on transcripts and variants located within splice sites (SS) were considered in order to identify novel candidate genes as described (233). Based on a recessive model of inheritance, a total of 17 candidate genes were identified (Supplementary Table 1).



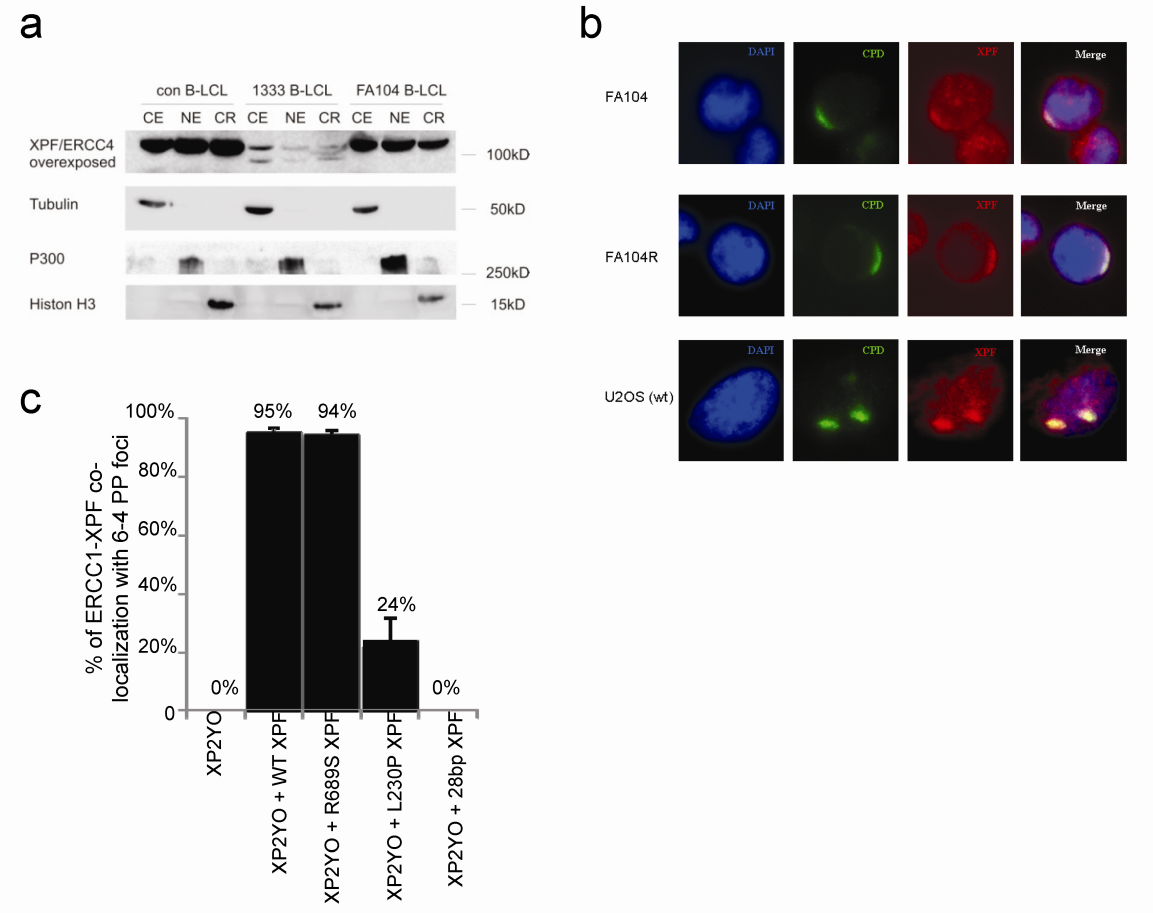
Supplementary Fig 1. Diagnostic confirmation of FA by the chromosome fragility test. (**a**) chromosome fragility after 0,1 μg ml^{-1} DEB in the affected patient FA104 in the frame of the Spanish FA chromosome fragility historical database as previously reported[1] (**b**) A tetraradial chromatid-type exchange found in FA104 upon DEB exposure. (**c**) Spontaneous and MMC induced chromosome fragility in patient 1333.



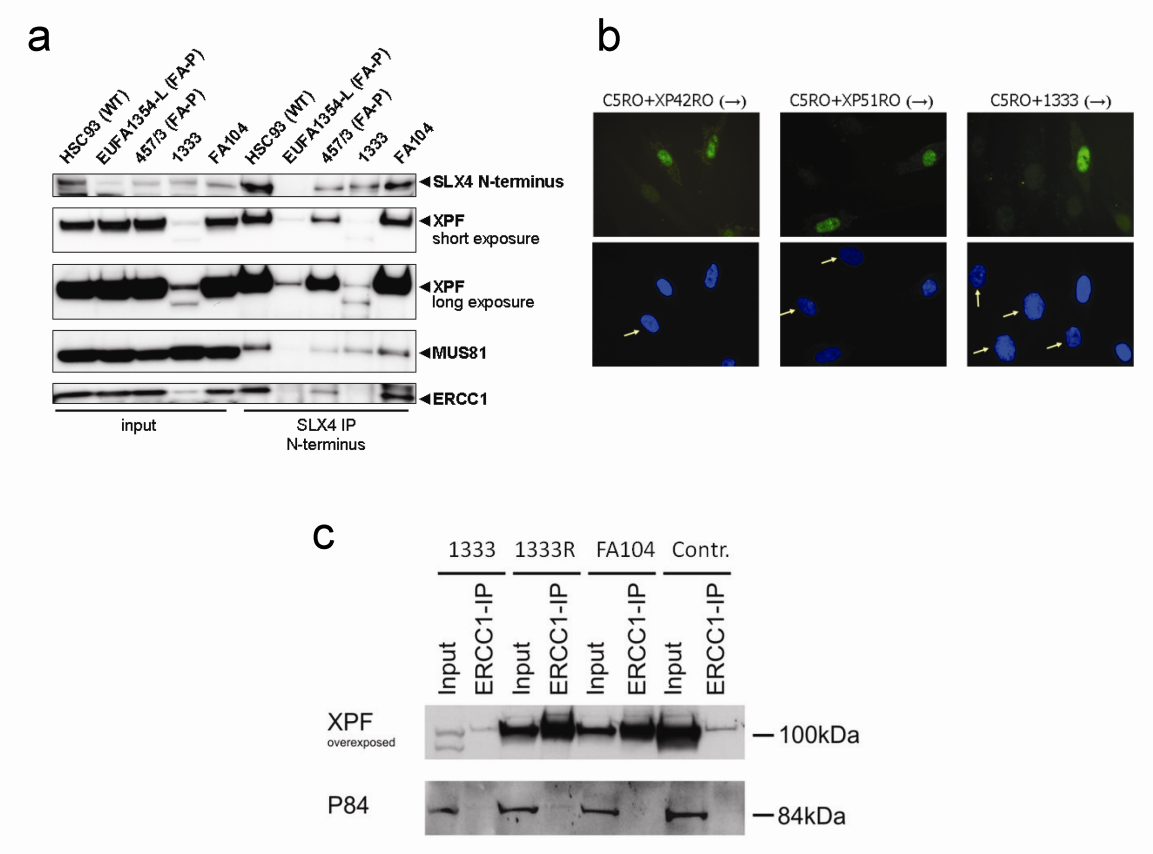
Supplementary Fig 2. Evolutionary conservation of XPF residues with missense mutations in FA patients. R689 (**a**) and L230 (**b**) are highly conserved from human to yeast. Amino acids that, along with R689, constitute the signature for the XPF nuclease motif are shown in green in (**a**).


a
Father
Mother
b
FA104R
(mixed alleles)
Cloned alleles
Wt
5bp del
12bp del
c
% of XPF levels vs WT
120
100
80
60
40
20
0
FA135 (WT)
HSC536
(FANCC-/-)
HSC536R
(FANCC+/-)
FA104
FA104R
d
Father
Mother
e
HSC93 (WT)
1333
1333R
SE
XPF
LE
XPF
XPF trunc
Tubulin
SE
ERCC1
LE
ERCC1
f
HSC93
1333
1333 R
H2O
HSC93
1333
1333 R
H2O
EXON 4
EXON 11b
Exon 4
Exon 11
HSC93
1333
1333R

Supplementary Fig 3. Analysis of parental DNAs and reverted cell lines. (**a**) Segregation of mutant alleles in FA104. Sequence analysis of genomic blood DNA of FA104's parents revealed that the allele with the missense mutation in exon 11 (c.2065C>A; p.R689S) was inherited from the father and that the allele bearing the 5 bp deletion in exon 8 (c.1484_1488delCTCAA; p.T495NfsX5) was inherited from the mother. (**b**) Sequence analysis of individual exon 8 alleles cloned from the FA104R cell line. Sequencing of single bacterial clones revealed the presence of a 12 bp deletion in exon 8 encompassing the pathogenic 5 bp deletion and restoring the reading frame of the *XPF g*ene. (**c**) Quantification of XPF expression by Western blot in lymphoblasts from FA104, FA104R, HSC536 (FA-C), HSC536R (HSC536 reverted to wt) and FA139 (wt). XPF levels are expressed as a ratio of the loading control (vinculin). The histogram represents XPF levels in the different cell lines normalized to the levels of the loading control. Means and SEM of at least three independent experiments are shown. (**d**) Segregation of mutant alleles in 1333. Sequence analysis of genomic blood DNA of 1333's parents revealed that the allele with the missense mutation in exon 4 of *XPF* (c.689T>C; p.L230P) was of paternal origin and that the 28 bp duplication (c.2371_2398dup28; p.I800TfsX23) was inherited from the mother. (**e**) Western blot analysis showing low levels of two XPF proteins in 1333 and a normal size XPF protein in the reverted cell line 1333R. (**f**) Absence of the 28 bp duplication in *XPF* exon 11 in 1333R eliminating the longer XPF mutant allele with the 28 bp duplication (upper panel) and restoring the wt sequence in exon 11 (lower panel).



Supplementary Fig 4. XPF cellular localization (**a**) FA104 shows an abundance and a distribution of XPF between the cytoplasmic, nuclear and chromatin compartments comparable to a normal control, whereas 1333 reveals reduced abundance and two species of that protein with sizes predicted by its mutations but, of note, XPF is still detected in the nucleus and on chromatin with grossly unaffected ratios to the cytoplasmic fraction. Specificity of the separation of extracts from lymphoblasts is confirmed by the compartment-specific marker proteins tubulin, p300 and histone H3. (**b**) Co-localization of XPF with cyclobutane pyrimidine dimers (CPDs) in FA104 and FA104R cell lines. LCLs were seeded on polylysine-treated coverslips, irradiated with UVC (60 J m^{-2}) through a polycarbonate filter with 5 µm pores, incubated for 0.5 h, fixed and stained for CPDs and XPF. U2OS cells were used as a positive control. (**c**) Graphical representation of the percent co-localization of XPF with (6-4)PP in XP2YO cells expressing various forms of XPF. XP2YO cells were transduced with wild type XPF, XPF-R689S, XPF-L230P, or XPF-28bp dup, irradiated with UVC (120 J m^{-2}) through a polycarbonate filter with 5 µm pores, incubated for 0.5 h, fixed and stained with antibodies to (6-4)PP and antibodies to HA antibodies. Data represent the average of at least 3 independent experiments ± the SD. For each experiment 100 cells were counted.



Supplementary Fig 5. SLX4 and ERCC1 interactions in XPF-deficient FA patients (**a**) SLX4 interactions in XPF-deficient FA patients. SLX4 was immunoprecipitated with a polyclonal antibody raised against the first 300 amino acids of SLX4 (SLX4 N-terminus). Precipitated proteins were visualized by Western blotting with antibodies to SLX4 N-terminus, XPF, ERCC1 and MUS81. Reduced XPF and ERCC1 expression was found in lymphoblasts of individual 1333. In these cells, full-length and truncated XPF and MUS81 were coprecipitated with SLX4, whereas ERCC1 is barely detectable. In lymphoblasts of individual FA104, the interaction between SLX4 and its binding partners XPF-ERCC1 and MUS81 is normal. Wild type lymphoblasts (HSC93) and lymphoblasts of FA-P patients (EUFA1354-L and 457/3) were used as controls. (**b**) Reduced ERCC1 nuclear levels in 1333 fibroblasts. Cells (arrows) were compared to mixed-in normal fibroblasts (C5RO) preloaded with polystyrene microbeads (no arrows), used as an internal control. (**c**) ERCC1-XPF interactions in FA104 and 1333 lymphoblast cell lines. ERCC1 was immunoprecipitated with a polyclonal antibody against ERCC1 and the precipitated proteins were visualized by Western bloting with antibodies against XPF and P83 as internal control.

Supplementary Table 1. List of candidate genes after whole exome sequencing when assuming a recessive inheritance model.

Chrom	Pos	Ref	Alt	Ensembl pred	AA change	GN	NRR	SNV Q	GT Q
1	169489751	A	W	SS	-	F5	42	171	171
1	169525877	T	Y	SS	-	F5	52	36	36
2	73675227	-	CTC	NFC	S/SP	ALMS1	16	N/A	N/A
2	73678183	G	R	NSC	G1509D	ALMS1	156	120	120
3	49094490	G	S	NSC	N381K	QRICH1	122	228	228
3	49095011	C	S	NSC	G208R	QRICH1	109	43	43
4	126238305	C	M	NSC	P247T	FAT4	52	178	178
4	126355484	C	M	NSC	A2368E	FAT4	56	190	190
5	156479444	TTG	-	NFC	TS/S	HAVCR1	61	N/A	N/A
5	156479568	-	GTT	NFC	T/TT	HAVCR1	106	N/A	N/A
6	31238942	G	W	NSC	A176V	HLA-C	23	61	39
6	31239577	A	C	NSC	S48A	HLA-C	21	90	90
6	32709309	A	R	SS	-	HLA-DQA2	29	84	84
6	32713044	C	Y	NSC	T64M	HLA-DQA2	192	228	228
6	32713188	C	Y	SS	-	HLA-DQA2	126	228	228
6	38840915	A	R	NSC	I2479V	DNAH8	72	216	216
6	38879340	A	T	NSC	E3267D	DNAH8	12	34	34
7	100686777	C	Y	NSC	T4027M	MUC17	323	228	228
7	100687107	G	R	SS	-	MUC17	66	79	79
8	30700598	T	Y	NSC	N1979S	TEX15	33	97	97
8	30701995	A	M	NSC	D1513E	TEX15	141	228	228
10	69682773	T	Y	NSC	D920G	HERC4	64	69	69
10	69785435	-	A	SS	-	HERC4	9	N/A	N/A
16	14029271	AACTC	-	FC	-	ERCC4	22	N/A	N/A
16	14041518	C	M	NSC	R689S	ERCC4	121	228	228
16	72137553	C	S	NSC	Q564E	DHX38	56	85	85
16	72142141	A	R	NSC	S994G	DHX38	52	106	106
17	74272839	C	Y	NSC	V1593M	QRICH2	54	33	33

17	74277009	T	Y	NSC	Q1264R	QRICH2	23	81	81
18	14105016	C	M	NSC	R508I	ZNF519	136	228	228
18	14105853	C	M	NSC	R229I	ZNF519	23	51	51
19	51918360	A	R	NSC	S445P	SIGLEC12	43	39	39
19	52004795	G	CT	FC	-	SIGLEC12	19	N/A	N/A
X	53561632	A	W	NSC	F4226I	HUWE1	42	53	53
X	53642759	C	M	NSC	E665D	HUWE1	16	33	33

Chrom: chromosome number; Pos: genomic position (GRCh37/hg19);Ref: reference allele; Alt: sample allele; Ensembl pred: consequence prediction of variants on transcript according to Ensembl v59. This column contains one of the following values: SS=splice site, NSC=non-synonymous coding, FC=frameshift coding, NFC=non-frameshift coding; AA change: amino acid change in the affected protein; GN: Gene name; NRR: Number of non-redundant reads; SNV Q: the phred-scaled likelihood that the genotype is identical to the reference; GT Q: Phred-scaled likelihood that the genotype is wrong.

Supplementary Table 2. Comparative summary of clinical and cellular/molecular features of XP, XFE and FA patients with mutations in XPF.

Clinical/cellular features	XPF	XFE	FA
Skin phosensitivity	mild	severe	no
Atrophic epidermis	variable	yes	no
Neurologic features	rare	yes	no
Hematology	normal	anemia[a]	anemia, BMF
Growth retardation[b]	no	yes	yes
Premature death	-	16yo	4yo (FA104). 1333 alive at age 10
UV sensitivity	mild	severe[c]	none (FA104)[d], mild (1333)
UDS defect	mild	severe[c]	mild (1333), ND in (FA104)[d]
MMC sensitivity	mild	severe	severe
DEB-test	negative	positive	positive
MMC induced G2/M arrest	negative	positive	positive
Nuclease in vitro	yes[e]	yes[e]	no

[a]It is not known whether anemia evolved to BMF in the XFE patient (Laura Niedernhofer, personal communication). [b]Include microsomy in 1333, FA104 and XFE and microcephaly in XFE and 1333. [c]Reported in Niedernhofer et al., 2005. [d]UDS assay was not done in FA104 due to the lack of skin fibroblasts but FA104 lymphoblasts were resistant to UV. [e]Reported in Ahmad et al., 2010 for XPF mutation R799W and XFE mutation R153P. Typically XP and FA features are marked in yellow and green, respectively.

Supplementary Table 3. PCR primers used for Sanger sequencing of XPF

Exon number	Primer sequence
Exon 1 for	ACTCGGCTCTCTTCGGTTGAGTT
Exon 1 rev	CGCTCAGGAGGCCCTCAACA
Exon 2 for	AACTGCCCTGTATTAAATAGCCTACTAA
Exon 2 rev	GTACAAATTTACATACTAATAATAAGATTTCA
Exon 3 for	AATGTGATGAATGAATGGCAATTACCTAC
Exon 3 rev seq	TTACATCAAAGTGCAACTAAGTTACAGTAG
Exon 3 rev	GTTTGTTATAAAACTGACTAGGTTCTATCA
Exon 4 for	CATAGCTGCTGAAACTCTAGAAAATTGTTGAAA
Exon 4 rev	AGTCAGAGTGATGCTTATATGCCAATCCAC
Exon 5 for	TACACAGGAAATAATCCTTTTGAAAGTATG
Exon 5 rev	CAGTATAACATATAGTTGAATATAGCACTT
Exon 6 for	ACAGGATGACAGCCAGTTACGTATGTAG
Exon 6 for seq	TAGGTCATGTGACCATCAGAGACTGTT
Exon 6 rev	ACACATTTAAAGACTTAACCCCACAAG
Exon 7 for	TATGTACTGATGCTCGTGTTATCTGTTGTT
Exon 7 for seq	GTTTTAAAAGCCTTTGGAAGACTTTATGG
Exon 7 rev	CACTAGGATCTCAGTGTTCATTTGCCAT
Exon 8 for	TAATACCAAAGGGTAAGATGTCTTCCCTT
Exon 8 for seq	GGTGAAGGAATAAGGGGGCAC
Exon 8 rev	CATAAACATAAGCAGCATCGTAACGGATAT
Exon 9 for	AATATTTGTTATTTGAGCGCTCTAGGTTGC
Exon 9 rev	CCTGAGCAGGACTATCTGATATTCC
Exon 10 for	ACTGCTATCATCATGTAGATCATTTTCAATAC
Exon 10 rev	TACAGACCAAGCCTTGGCAGAGTAACT
Exon 11a for	AGAGTTAACAACAGAAACATCGATTTTTAG
Exon 11a rev	CTCAAACAACTCCGCCGTTGCATGA
Exon 11b for	CTTGTTTCAGGAGATCTCCAGCAATG
Exon 11b rev	TATGATGTCTGGCAAGGAGCCGCT

ACKNOWLEDGEMENTS

We would like to thank Fanconi anemia families and their clinicians for their continuous support and for providing samples and clinical data and Dr María A. Blasco (Centro Nacional de Investigaciones Oncológicas, Madrid) for providing Xpf deficient MEFs. JAB laboratory is funded by grants from the European Program "7FWP, Health" (PERSIST; Agreement no: 222878), the Spanish Ministry of Science and Innovation (Refs110-90.1 and SAF 2009-07164), Programa RETICS-RD06/0010/0015 ISCIII and Fundación Botín. ODS acknowledges funding from the National Institutes of Health (GM080454 and CA092584). CS is funded by CCA/V-ICI Amsterdam. DS and BS received grants from the Deutsche Fanconi-Anaemie-Hilfe, Aktionskreis Fanconi-Anaemie and the Schroeder-Kurth Fund. JS's laboratory is funded by the Generalitat de Catalunya (SGR0489-2009), the ICREA-Academia award, the Spanish Ministry of Science and Innovation (CB06/07/0023, and SAF2009-11936), and the European Regional Development FEDER Funds. CIBERER is an initiative of the Instituto de Salud Carlos III, Spain.

CHAPTER 4

DNA BINDING RESIDUES NEAR THE ACTIVE SITE OF ERCC1-XPF MEDIATE ITS FUNCTIONS IN NER AND ICL REPAIR

ABSTRACT

ERCC1-XPF is a structural-specific endonuclease that incises ss/ds DNA junctions. Multiple protein-DNA interactions synergistically contribute to the role of ERCC1-XPF in NER. Previous results implicate that mutations in DNA binding residues of ERCC1-XPF affect ICL repair more severely than NER. In this chapter, we examined the roles of DNA binding residues of ERCC1-XPF in mediating its functions in NER and ICL repair. We proposed three arginine residues (Arg689, Arg715, Arg740) near the active site are involved in DNA binding, and mutated them to alanines. Our results confirmed the synergistic DNA binding effects of ERCC1-XPF in NER. However, the activities of these mutants in ICL repair remains to be determined.

INTRODUCTION

A number of anticancer drugs target DNA to generate ICLs that interfere with replication and proliferation of cancer cells. Therefore, it is important to understand how tumors become resistant to the therapeutic treatment of cross-linking agents, and how these lesions have been repaired (135). The repair of ICLs is an orchestrated process that requires many proteins to cooperate. A genetic disorder Fanconi anemia (FA) is associated with defects in ICL repair. Thus, the phenotypes of FA patients have been invaluable for studies of the mechanisms of ICL repair (135). The FA patients are divided into 15 subtypes (FANCA-FANCQ), depending on which gene harbors mutations (135). XPF is the most recently discovered FA gene (Chapter 3), and two unrelated patients with mutations in XPF have been indentified. Therefore, XPF is associated with at least three disorders: xerdoderma pigmentosum (XP), the XFE progeroid syndrome and FA.

ERCC1-XPF is a structural specific endonuclease (1) and plays roles in the maintenance of genome stability by participating multiple DNA processes including nucleotide excision repair (NER) (1) and ICL repair (200,217). It has been shown that the functions of ERCC1-XPF in NER and ICL repair can be compromised separately by introducing specific point mutations (37,42). In NER, ERCC1-XPF is recruited to sites of UV damage via interaction with another NER protein, XPA. A peptide of 14 amino acids of XPA binds a V-shape groove located in ERCC1 (31,40,41). Interrupting this interaction leads to a defect in NER without affecting the ICL repair (41,42). On the other hand, the mutations in DNA binding residues of XPF specifically impair ICL repair but only mildly affect NER (37). Interestingly, one of these proposed mutations has been found to cause FA disease (Chapter 3). This particular DNA binding residue is R689, located near the active site.

In this chapter I aimed to further explore the functions of ERCC1-XPF-DNA binding in NER and ICL repair. Crystallographic data of XPF homolog in *Aeropyrum pernix* indicate that residues R26 and R77 (equivalent to R740 and R689 in humans) are at similar positions for DNA binding (32). Biochemical studies have shown that XPF^{R726A} has no cleavage activity on model substrates, but cuts DNA in presence of the other NER proteins, similar to XPF^{R689A} (38,196). In addition, residues R689, R726 and R740 are highly conserved. Thus, I hypothesize that all of these three arginine residues interact with DNA near the active site, and I set out to test how mutations in them affect NER and ICL repair.

EXPERIMENTAL PROCEDURES

Generation of mutant cell lines using lentiviral transduction

Site directed mutagenesis of pFastBac-XPF and pFastBac-ERCC1-His was performed using the QuikChange site-directed mutagenesis kit (Stratagene) to generate the following mutations: XPF^{WT905}, XPF^{WT916}, XPF^{D703G}, $XPF^{R726A/D703G}$, $XPF^{R740A/D703G}$, and $XPF^{R689A/R726A}$. Human XP-F mutant fibroblasts XP2YO, and 293T cells were cultured in Dulbecco's modified Eagle's medium high glucose 1 x (GIBCO), 10% fetal bovine serum, 100 units/ml penicillin, and 0.1 mg/ml streptomycin at 37°C in a 5% CO_2 humid incubator. Wild-type or mutant XPF cDNA with a hemaglutinin (HA) tag at C-terminus was inserted into the pWPXL vector, which was co-transfected with the packaging plasmid psPAX2 and the envelop plasmid pMD2G into 293T cells to generate lentiviral particles. The particles were transduced into XP2YO cells to produce

cell lines stably expressing wild-type or mutant XPF proteins according to established procedures (http://www.lentiweb.com)(207,234).

Local UV-irradiation and immunofluorescence

About 50,000 cells were plated on a coverslip in 6-well plates, grown for 2-3 days and irradiated through a polycarbonate membrane with 5 μm pores (Millipore) with UV light (254 nm) with a dose of 150 J/m^2. Cells were incubated at 37°C, 5% CO_2 for 30 min to 24 hr. They were washed first with PBS and then with PBS plus 0.05% triton X-100 and fixed with 3% paraformaldehyde plus 0.1% triton X-100. After fixation, cells were washed with PBS containing 0.2% triton X-100. To stain for (6-4)PPs, cells were treated with 0.07 M NaOH in PBS for 5 min, followed by washing with PBS plus 0.2% triton X-100. After blocking with PBS plus 5 mg/ml BSA and 1.5 mg/ml glycine, cells were stained with mouse monoclonal anti-(6-4)PP antibody (Cosmo Bio) 1:400, and rabbit polyclonal anti-HA antibody (ChIP grade, Abcam) 1:3000 for 1.5 h and washed with PBS containing 0.2% triton X-100. Cells were then incubated with secondary antibodies: Cy3-conjugated affinipure goat anti-mouse IgG(H+L) (Jackson ImmunoResearch) 1:1000 and Alexa Fluor 488-labeled $F(ab')_2$ fragment of goat anti-rabbit IgG (H+L) (Invitrogen) 1:800 for 1 h, followed by washing with PBS with 0.2% triton X-100. Samples were washed with PBS, embedded in Vectashield Mounting Medium with 1.5 μg/ml of DAPI (Vector Laboratories) and analyzed using a confocal microscope (Zeiss LSM 510). About 100 cells were counted in at least two independent experiments for quantification.

Comet assay

The CometChips were made according to reference (235). Molten 1% agarose was loaded on GelBond film (Lonza), and then the mold was placed on the top until the gel was solidified. After removing the mold, the CometChip with microwells was generated.

Cells were treated with 3 ug/ml (15.6 μM) Mechlorethamine hydrochloride (NM) for 1h in serum-free DMEM medium. After washing two times and incubating in the growth medium for 0-72 h at 37°C, cells were trypsinized and loaded into the CometChip. Excess cells were washed away, and the CometChip was covered with 1% low melting temperature agarose. Then, the gel was submerged in alkaline lysis buffer (2.5 M NaCl, 100 mM Na_2EDTA, 10 mM Tris, and 1% Triton X-100, pH 10) overnight at 4°C to lysate the cells, followed by treatment of 12 Gy of gamma-radiation from a cesium-137 source. The DNA was unwound in alkaline electrophoresis buffer (0.3 mM NaOH, 1 mM Na_2EDTA) for 40 min at 4°C. The gels were run at 1 V/cm and 300 mA for 30 min at 4°C and neutralized with neutral buffer (2 mM Na_2EDTA, 90 mM Tris, and 90 mM Boric acid, pH =8.5) twice. After staining the gels with SYBR Gold, the results were visualized by an epifluorescent microscope with10 time objective lens. The unhooking percentage is calculated using the equation

% unhooking at T1=[(% DTM at T_0 - %DTM at T_1)/(%DTM at T_0)]*100

In which T_0 is the time immediately after the drug treatment and T_1 is the incubation time after removing the drug. While DTM is decrease in tail moment, and the %DTM can be calculated by the function

(%DTM)=[1-(TMdi-TMcu)/(TMci-TMcu)]*100

TMdi is the tail moment of drug-treated and irradiated cells, TMcu is the tail moment of undruged and unirradiated cells, and TMci is the tail moment of undruged and irradiated cells (217,235,236).

RESULTS

Two widely used wide types of XPF have the same activities in NER

Currently, two cDNA forms of XPF are widely used: The first one was isolated by the Hoeijmakers' laboratory (1) with a length of 905 amino acids and a aspartic acid at residue 703. The second one is from Thompson's laboratory (7). It is 11 amino acids longer at N-terminus and has the conserved glycine at position 703. Both of them restore full UV resistance to XPF-mutant cell lines. To examine the roles of the 11 amino acids and residue 703, we transfected XPF^{905} (905aa/D703), XPF^{916} (916aa/D703), XPF^{G703D} (916aa/G703) into XPF deficient cells (XP2YO), and generated the cell lines. In this chapter, unless specially labeled, all XPF are with 916aa/D703. To compare their NER activities, cells were irradiated by UV light (254nm) though polycarbonate filter with 5 μm pores to generate the local UV induced damages that are repaired by NER. After irradiation, cells were incubated at 37°C for certain time, and UV lesion (6-4) pothoproducts ((6-4)PPs) were detected. The repair efficiency has no differences in all these three cell lines, indicating the 11 amino acids and conserved residue 703 do not play roles in NER (**Figure 1**).

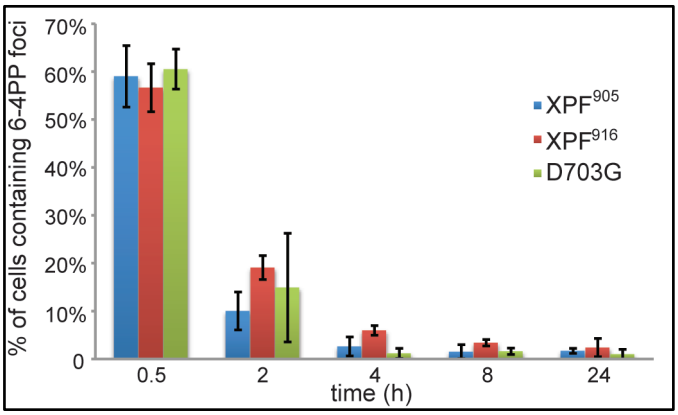


Figure 1. Two widely used wide types of XPF have the same activities in NER. Cells were irradiated with UV light (254nm) though polycarbonate filter with 5 μm pores and incubated at 37°C for 0.5 h, 2 h, 4 h, 8 h, and 24 h before fixed and stained with antibodies for UV lesion (6-4)PPs. Cells containing (6-4)PPs were counted. Quantification is based on at least two independent experiments, and about 100 cells were counted in each experiment. The standard deviations were shown in the error bar.

The recruitment of ERCC1-XPF does not depend on its DNA binding ability

The residues R689, R726 and R740 are proposed to interact with DNA near the active site. With the mutation R689A, the protein has a mild defect in NER, but a severe one in ICL repair (37). To investigate the effects of the other two residues, $XPF^{R726A/D703G}$ and $XPF^{R740A/D703G}$ have been transfected in to XP2YO cells and the corresponding cell lines have been generated. In addition, $XPF^{R689A/R726A}$ is also included in this study to observe the synergetic effects.

Cells were irradiated with UV light (254 nm) through a filter with 5 μm pores and incubated at 37°C for 30 min. The co-localization of (6-4)PPs and ERCC1-XPF was detected, to observe the recruitment of ERCC1-XPF to the NER complex. Interestingly, the co-localization percentages are all above 90%, indicating that mutations of the DNA binding residues near the active site do not affect the recruitment of the protein to the NER complex (**Figure 2**).

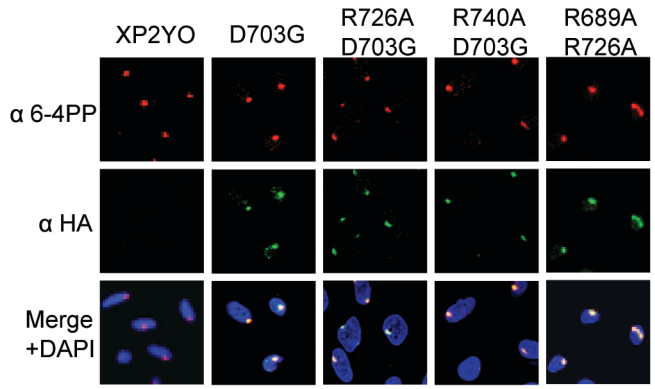


Figure 2. The recruitment of ERCC1-XPF does not depend on its DNA binding ability. XPF gene with HA tag at the C terminal was transfected into the XP2YO cells. Cells were irradiated with UV (254nm) though polycarbonate filter with 5 μm pores and incubated at 37°C for 0.5 h before fixed. The co-localization was observed by staining with antibodies for UV lesion (6-4)PPs and HA.

Multiple DNA binding residues near the active site synergistically affect the NER activity of ERCC1-XPF

After observing the co-localization of ERCC1-XPF with the UV lesions, the repair rates were also determined. We incubated from 30min to 24h, after local UV irradiation. Cells containing (6-4)PP foci were counted. After 30 min incubation, about 65% of cells contains UV lesions in all cell lines. In XPF^{D703G} cells (wide type), the percentage of lesion-containing cells was reduced to 2% at 4h. However, in the XPF deficient (XP2YO) cells, 18% of cells still have (6-4)PPs at 24h, indicating the lesion repair efficiency in absence of NER. $XPF^{R726A/D703G}$ and $XPF^{R740A/D703G}$ cells showed mild defects in NER, with 14% and 12% lesion-containing cells at 4h, and 10% and 4% at 8h respectively, similar to XPF^{R689A} as previously reported (37). According to **Figure 3**, the repair rate of $XPF^{R689A/R726A}$ was very close to XP2YO, in consistent with the previous results that multiple weak protein-DNA interactions synergetic affect the NER activity of ERCC1-XPF(37).

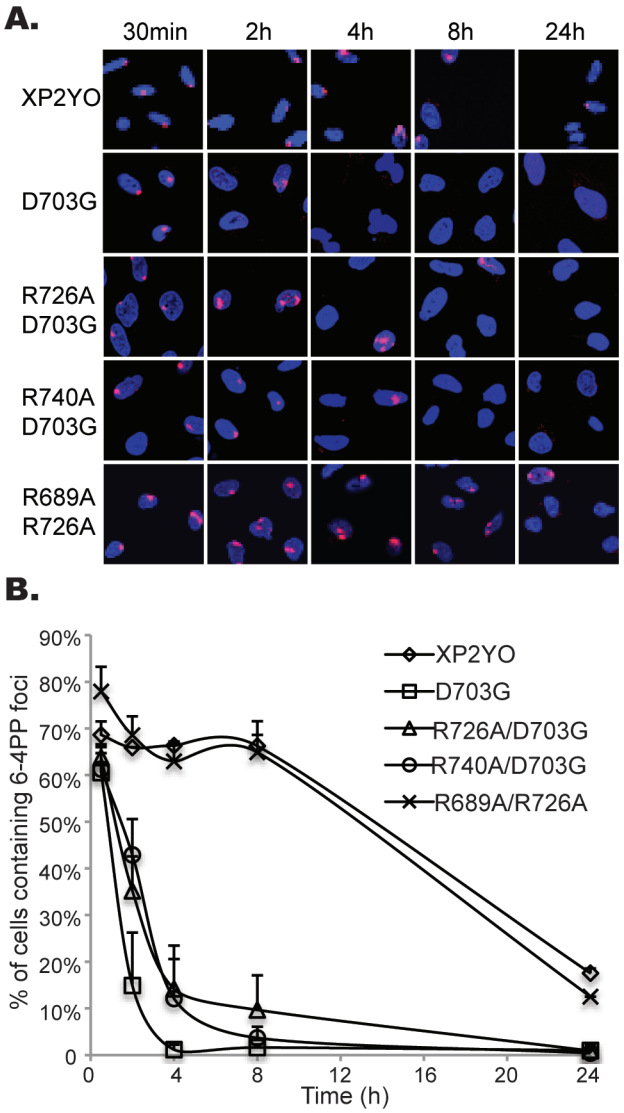


Figure 3. Multiple DNA binding residues near the active site synergistically affect NER activity. Wild type or mutant XPF genes were transfected into the XP2YO cells. After irradiated with UV (254nm) though polycarbonate filter with 5 μm pores, cells were incubated at 37°C for 0.5 h to 24 h, fixed and detected. The UV lesions, (6-4)PPs, were detected by antibodies and cells having (6-4)PP foci were counted. Each experiment was done at least twice and about 100 cells were counted in each experiment. The standard deviations were shown in the error bar.

The roles of DNA binding residues in the unhooking step of ICL repair still need to be examined

To study the role of ERCC1-XPF in ICL repair, the unhooking step was tested by single cell electrophoresis assay (comet assay). Cells were treated with cross-linking agent, nitrogen mustard (NM), and incubated at 37°C for 0-48 h. After incubation, cells were lysed and irradiated by 12 Gy of gamma-radiation, followed by neutralizing and electrophoresis running step. Big tails were observed under the control conditions, without treatment of the cross-linking agent. On the other hand, after treatment of NM, the DNA concentrated in the head at 0h time point, and the tail fraction was increased after longer incubation due to the unhooking of ICL lesions (217,236). In XP2YO cells, 70% DNA was unhooked after 48 h incubation, while 130% was observed in XPF^{D703G}. The reason of more than 100% of unhooking may be due to cell death, corresponding to the damaging agent. In **Figure 4**, the unhooking percentages of $XPF^{R726A/D703G}$, $XPF^{R740A/D703G}$ and FA patient mutant $XPF^{R689S/D703G}$ were lower than wild type, and higher than deficient cells, indicating mutations in these residues partially affect the unhooking step of ICL repair. However, since the study is still in the preliminary step and the difference between the wild type and deficient cells is not significant, the further investigation still need to be continued.

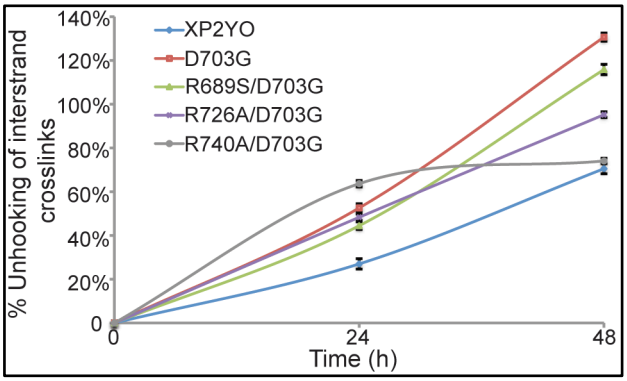


Figure 4. Detection of the unhooking step of the ICL repair by akline comet assay. Cells were treated with cross-linking agent NM (15.6 μM) for 1h, incubated in the growth medium for 0-48 h, and loaded into CometChip. Subsequently, lysation, gamma irradiation (12 Gy) and

electrophoresis running were conducted. The sample were stained by SYBR-Gold and visualized by microscope. The standard deviations were shown in error bars.

DISCUSSION

The results of this chapter confirm previous conclusions that the multiple protein-DNA interactions synergistically affect the NER activity of ERCC1-XPF (37), however its roles in ICL repair pathway are still to be determined. For example, SLX4 is involved in ICL repair and has been proved to interact with ERCC1-XPF (46-48). However, it is still unknown whether SLX4-XPF interaction is strong enough to recruit ERCC1-XPF to the ICL repair complex, like the XPA-ERCC1 interaction in NER.

ACKNOWLEDGEMENTS

I would like to thank Patrizia Mazzucato and Dr. Bevin Engelward's in MIT for help with comet assay.

REFERENCE

1. Sijbers, A. M., de Laat, W. L., Ariza, R. R., Biggerstaff, M., Wei, Y. F., Moggs, J. G., Carter, K. C., Shell, B. K., Evans, E., de Jong, M. C., Rademakers, S., de Rooij, J., Jaspers, N. G., Hoeijmakers, J. H., and Wood, R. D. (1996) *Cell* **86**, 811-822
2. Gillet, L. C., and Schärer, O. D. (2006) *Chem Rev* **106**, 253-276
3. Niedernhofer, L. J., Garinis, G. A., Raams, A., Lalai, A. S., Robinson, A. R., Appeldoorn, E., Odijk, H., Oostendorp, R., Ahmad, A., van Leeuwen, W., Theil, A. F., Vermeulen, W., van der Horst, G. T., Meinecke, P., Kleijer, W. J., Vijg, J., Jaspers, N. G., and Hoeijmakers, J. H. (2006) *Nature* **444**, 1038-1043
4. Jaspers, N. G., Raams, A., Silengo, M. C., Wijgers, N., Niedernhofer, L. J., Robinson, A. R., Giglia-Mari, G., Hoogstraten, D., Kleijer, W. J., Hoeijmakers, J. H., and Vermeulen, W. (2007) *Am J Hum Genet* **80**, 457-466
5. Westerveld, A., Hoeijmakers, J. H., van Duin, M., de Wit, J., Odijk, H., Pastink, A., Wood, R. D., and Bootsma, D. (1984) *Nature* **310**, 425-429
6. van Vuuren, A. J., Appeldoorn, E., Odijk, H., Yasui, A., Jaspers, N. G., Bootsma, D., and Hoeijmakers, J. H. (1993) *EMBO J* **12**, 3693-3701
7. Brookman, K. W., Lamerdin, J. E., Thelen, M. P., Hwang, M., Reardon, J. T., Sancar, A., Zhou, Z. Q., Walter, C. A., Parris, C. N., and Thompson, L. H. (1996) *Molecular and Cellular Biology* **16**, 6553-6562
8. Yagi, T., Wood, R. D., and Takebe, H. (1997) *Mutagenesis* **12**, 41-44
9. Ciccia, A., McDonald, N., and West, S. C. (2008) *Annu Rev Biochem* **77**, 259-287
10. Chen, X. B., Melchionna, R., Denis, C. M., Gaillard, P. H., Blasina, A., Van de Weyer, I., Boddy, M. N., Russell, P., Vialard, J., and McGowan, C. H. (2001) *Molecular Cell* **8**, 1117-1127
11. Ciccia, A., Ling, C., Coulthard, R., Yan, Z., Xue, Y., Meetei, A. R., Laghmani, E. H., Joenje, H., McDonald, N., de Winter, J. P., Wang, W., and West, S. C. (2007) *Molecular Cell* **25**, 331-343
12. Taylor, E. R., and McGowan, C. H. (2008) *Proc Natl Acad Sci USA* **105**, 3757-3762
13. Ciccia, A., Constantinou, A., and West, S. C. (2003) *The Journal of biological chemistry* **278**, 25172-25178
14. Bastin-Shanower, S. A., Fricke, W. M., Mullen, J. R., and Brill, S. J. (2003) *Mol Cell Biol* **23**, 3487-3496
15. Ciccia, A., Ling, C., Coulthard, R., Yan, Z., Xue, Y., Meetei, A. R., Laghmani el, H., Joenje, H., McDonald, N., de Winter, J. P., Wang, W., and West, S. C. (2007) *Molecular cell* **25**, 331-343
16. Matos, J., Blanco, M. G., Maslen, S., Skehel, J. M., and West, S. C. (2011) *Cell* **147**, 158-172
17. Wechsler, T., Newman, S., and West, S. C. (2011) *Nature* **471**, 642-646
18. Meetei, A. R., Medhurst, A. L., Ling, C., Xue, Y., Singh, T. R., Bier, P., Steltenpool, J., Stone, S., Dokal, I., Mathew, C. G., Hoatlin, M., Joenje, H., de Winter, J. P., and Wang, W. (2005) *Nat Genet* **37**, 958-963
19. Kim, J. M., Kee, Y., Gurtan, A., and D'Andrea, A. D. (2008) *Blood* **111**, 5215-5222
20. Deans, A. J., and West, S. C. (2009) *Molecular Cell* **36**, 943-953

21. Aravind, L., Walker, D. R., and Koonin, E. V. (1999) *Nucleic acids research* **27**, 1223-1242
22. Komori, K., Fujikane, R., Shinagawa, H., and Ishino, Y. (2002) *Genes Genet Syst* **77**, 227-241
23. Sgouros, J., Gaillard, P. H., and Wood, R. D. (1999) *Trends Biochem Sci* **24**, 95-97
24. Nishino, T., Komori, K., Ishino, Y., and Morikawa, K. (2003) *Structure* **11**, 445-457
25. Nishino, T., Komori, K., Tsuchiya, D., Ishino, Y., and Morikawa, K. (2005) *Structure* **13**, 143-153
26. Bailly, V., Sommers, C. H., Sung, P., Prakash, L., and Prakash, S. (1992) *Proc Natl Acad Sci USA* **89**, 8273-8277
27. Bardwell, L., Cooper, A. J., and Friedberg, E. C. (1992) *Molecular and Cellular Biology* **12**, 3041-3049
28. Carr, A. M., Schmidt, H., Kirchhoff, S., Muriel, W. J., Sheldrick, K. S., Griffiths, D. J., Basmacioglu, C. N., Subramani, S., Clegg, M., and Nasim, A. (1994) *Molecular and Cellular Biology* **14**, 2029-2040
29. Gaillard, P. H., and Wood, R. D. (2001) *Nucleic acids research* **29**, 872-879
30. Tripsianes, K., Folkers, G., Ab, E., Das, D., Odijk, H., Jaspers, N. G. J., Hoeijmakers, J. H. J., Kaptein, R., and Boelens, R. (2005) *Structure* **13**, 1849-1858
31. Tsodikov, O. V., Enzlin, J. H., Scharer, O. D., and Ellenberger, T. (2005) *Proc Natl Acad Sci U S A* **102**, 11236-11241
32. Newman, M., Murray-Rust, J., Lally, J., Rudolf, J., Fadden, A., Knowles, P. P., White, M. F., and McDonald, N. Q. (2005) *Embo J* **24**, 895-905
33. Nishino, T., Komori, K., Ishino, Y., and Morikawa, K. (2005) *Structure* **13**, 1183-1192
34. Doherty, A. J., Serpell, L. C., and Ponting, C. P. (1996) *Nucleic acids research* **24**, 2488-2497
35. Das, D., Folkers, G. E., van Dijk, M., Jaspers, N. G. J., Hoeijmakers, J. H. J., Kaptein, R., and Boelens, R. (2012) *Structure* **20**, 667-675
36. Das, D., Tripsianes, K., Jaspers, N. G. J., Hoeijmakers, J. H. J., Kaptein, R., Boelens, R., and Folkers, G. E. (2008) *Proteins* **70**, 1551-1563
37. Su, Y., Orelli, B., Madireddy, A., Niedernhofer, L. J., and Schärer, O. D. (2012) *J Biol Chem* **287**, 21846-21855
38. Enzlin, J. H., and Schärer, O. D. (2002) *Embo J* **21**, 2045-2053
39. de Laat, W. L., Appeldoorn, E., Jaspers, N. G. J., and Hoeijmakers, J. H. J. (1998) *J Biol Chem* **273**, 7835-7842
40. Tripsianes, K., Folkers, G. E., Zheng, C., Das, D., Grinstead, J. S., Kaptein, R., and Boelens, R. (2007) *Nucleic acids research* **35**, 5789-5798
41. Tsodikov, O. V., Ivanov, D., Orelli, B., Staresincic, L., Shoshani, I., Oberman, R., Scharer, O. D., Wagner, G., and Ellenberger, T. (2007) *EMBO J* **26**, 4768-4776
42. Orelli, B., McClendon, T. B., Tsodikov, O. V., Ellenberger, T., Niedernhofer, L. J., and Schärer, O. D. (2010) *J Biol Chem* **285**, 3705-3712
43. Büttner, K., Nehring, S., and Hopfner, K.-P. (2007) *Nat Struct Mol Biol* **14**, 647-652
44. Ahmad, A., Enzlin, J. H., Bhagwat, N. R., Wijgers, N., Raams, A., Appledoorn, E., Theil, A. F., JH, J. H., Vermeulen, W., NG, J. J., Scharer, O. D., and Niedernhofer, L. J. (2010) *PLoS Genet* **6**, e1000871
45. Andersen, S. L., Bergstralh, D. T., Kohl, K. P., LaRocque, J. R., Moore, C. B., and Sekelsky, J. (2009) *Mol Cell* **35**, 128-135

46. Fekairi, S., Scaglione, S., Chahwan, C., Taylor, E. R., Tissier, A., Coulon, S., Dong, M. Q., Ruse, C., Yates, J. R., 3rd, Russell, P., Fuchs, R. P., McGowan, C. H., and Gaillard, P. H. (2009) *Cell* **138**, 78-89
47. Munoz, I. M., Hain, K., Declais, A. C., Gardiner, M., Toh, G. W., Sanchez-Pulido, L., Heuckmann, J. M., Toth, R., Macartney, T., Eppink, B., Kanaar, R., Ponting, C. P., Lilley, D. M., and Rouse, J. (2009) *Mol Cell* **35**, 116-127
48. Svendsen, J. M., Smogorzewska, A., Sowa, M. E., O'Connell, B. C., Gygi, S. P., Elledge, S. J., and Harper, J. W. (2009) *Cell* **138**, 63-77
49. de Laat, W. L., Jaspers, N. G., and Hoeijmakers, J. H. (1999) *Genes & development* **13**, 768-785
50. Hanawalt, P. C., and Spivak, G. (2008) *Nature Reviews Molecular Cell Biology* **9**, 958-970
51. Venema, J., Bartosova, Z., Natarajan, A. T., van Zeeland, A. A., and Mullenders, L. H. (1992) *The Journal of biological chemistry* **267**, 8852-8856
52. Leadon, S. A., and Lawrence, D. A. (1991) *Mutation research* **255**, 67-78
53. Henning, K. A., Li, L., Iyer, N., McDaniel, L. D., Reagan, M. S., Legerski, R., Schultz, R. A., Stefanini, M., Lehmann, A. R., Mayne, L. V., and Friedberg, E. C. (1995) *Cell* **82**, 555-564
54. Troelstra, C., van Gool, A., de Wit, J., Vermeulen, W., Bootsma, D., and Hoeijmakers, J. H. (1992) *Cell* **71**, 939-953
55. Nakatsu, Y., Asahina, H., Citterio, E., Rademakers, S., Vermeulen, W., Kamiuchi, S., Yeo, J. P., Khaw, M. C., Saijo, M., Kodo, N., Matsuda, T., Hoeijmakers, J. H., and Tanaka, K. (2000) *The Journal of biological chemistry* **275**, 34931-34937
56. Sugasawa, K., Ng, J. M., Masutani, C., Iwai, S., van der Spek, P. J., Eker, A. P., Hanaoka, F., Bootsma, D., and Hoeijmakers, J. H. (1998) *Mol Cell* **2**, 223-232
57. Volker, M., Mone, M. J., Karmakar, P., van Hoffen, A., Schul, W., Vermeulen, W., Hoeijmakers, J. H., van Driel, R., van Zeeland, A. A., and Mullenders, L. H. (2001) *Mol Cell* **8**, 213-224.
58. Min, J. H., and Pavletich, N. P. (2007) *Nature* **449**, 570-575
59. Scrima, A., Konícková, R., Czyzewski, B. K., Kawasaki, Y., Jeffrey, P. D., Groisman, R., Nakatani, Y., Iwai, S., Pavletich, N. P., and Thomä, N. H. (2008) *Cell* **135**, 1213-1223
60. Houtsmuller, A. B., Rademakers, S., Nigg, A. L., Hoogstraten, D., Hoeijmakers, J. H., and Vermeulen, W. (1999) *Science (New York, N.Y* **284**, 958-961
61. Fousteri, M., Vermeulen, W., van Zeeland, A. A., and Mullenders, L. H. F. (2006) *Molecular Cell* **23**, 471-482
62. Schaeffer, L., Roy, R., Humbert, S., Moncollin, V., Vermeulen, W., Hoeijmakers, J. H., Chambon, P., and Egly, J. M. (1993) *Science (New York, N.Y* **260**, 58-63
63. Schaeffer, L., Moncollin, V., Roy, R., Staub, A., Mezzina, M., Sarasin, A., Weeda, G., Hoeijmakers, J. H., and Egly, J. M. (1994) *The EMBO journal* **13**, 2388-2392
64. Sugasawa, K., Akagi, J.-i., Nishi, R., Iwai, S., and Hanaoka, F. (2009) *Molecular Cell* **36**, 642-653
65. Wakasugi, M., and Sancar, A. (1998) *Proc Natl Acad Sci U S A* **95**, 6669-6674
66. Tapias, A., Auriol, J., Forget, D., Enzlin, J. H., Scharer, O. D., Coin, F., Coulombe, B., and Egly, J. M. (2004) *J Biol Chem* **279**, 19074-19083
67. de Laat, W. L., Appeldoorn, E., Sugasawa, K., Weterings, E., Jaspers, N. G., and Hoeijmakers, J. H. (1998) *Genes Dev* **12**, 2598-2609

68. Riedl, T., Hanaoka, F., and Egly, J. M. (2003) *Embo J* **22**, 5293-5303
69. Mu, D., Wakasugi, M., Hsu, D. S., and Sancar, A. (1997) *The Journal of biological chemistry* **272**, 28971-28979
70. Wakasugi, M., Reardon, J. T., and Sancar, A. (1997) *The Journal of biological chemistry* **272**, 16030-16034
71. Li, L., Elledge, S. J., Peterson, C. A., Bales, E. S., and Legerski, R. J. (1994) *Proceedings of the National Academy of Sciences of the United States of America* **91**, 5012-5016
72. Li, L., Peterson, C. A., Lu, X., and Legerski, R. J. (1995) *Mol Cell Biol* **15**, 1993-1998
73. Park, C. H., and Sancar, A. (1994) *Proceedings of the National Academy of Sciences of the United States of America* **91**, 5017-5021
74. Saijo, M., Kuraoka, I., Masutani, C., Hanaoka, F., and Tanaka, K. (1996) *Nucleic acids research* **24**, 4719-4724
75. Staresincic, L., Fagbemi, A. F., Enzlin, J. H., Gourdin, A. M., Wijgers, N., Dunand-Sauthier, I., Giglia-Mari, G., Clarkson, S. G., Vermeulen, W., and Scharer, O. D. (2009) *The EMBO journal*
76. Shivji, M. K., Podust, V. N., Hubscher, U., and Wood, R. D. (1995) *Biochemistry* **34**, 5011-5017
77. Moser, J., Kool, H., Giakzidis, I., Caldecott, K., Mullenders, L. H., and Fousteri, M. I. (2007) *Mol Cell* **27**, 311-323
78. McHugh, P. J., Spanswick, V. J., and Hartley, J. A. (2001) *The lancet oncology* **2**, 483-490
79. Hoy, C. A., Thompson, L. H., Mooney, C. L., and Salazar, E. P. (1985) *Cancer Res* **45**, 1737-1743
80. Meetei, A. R., de Winter, J. P., Medhurst, A. L., Wallisch, M., Waisfisz, Q., van de Vrugt, H. J., Oostra, A. B., Yan, Z., Ling, C., Bishop, C. E., Hoatlin, M. E., Joenje, H., and Wang, W. (2003) *Nature genetics* **35**, 165-170
81. Smogorzewska, A., Matsuoka, S., Vinciguerra, P., McDonald, E. R., 3rd, Hurov, K. E., Luo, J., Ballif, B. A., Gygi, S. P., Hofmann, K., D'Andrea, A. D., and Elledge, S. J. (2007) *Cell*
82. Moldovan, G.-L., and D'Andrea, A. D. (2009) *Annu Rev Genet* **43**, 223-249
83. Knipscheer, P., Räschle, M., Smogorzewska, A., Enoiu, M., Ho, T. V., Schärer, O. D., Elledge, S. J., and Walter, J. C. (2009) *Science* **326**, 1698-1701
84. Wang, A. T., Sengerová, B., Cattell, E., Inagawa, T., Hartley, J. M., Kiakos, K., Burgess-Brown, N. A., Swift, L. P., Enzlin, J. H., Schofield, C. J., Gileadi, O., Hartley, J. A., and McHugh, P. J. (2011) *Genes Dev* **25**, 1859-1870
85. Kratz, K., Schöpf, B., Kaden, S., Sendoel, A., Eberhard, R., Lademann, C., Cannavó, E., Sartori, A. A., Hengartner, M. O., and Jiricny, J. (2010) *Cell* **142**, 77-88
86. Liu, T., Ghosal, G., Yuan, J., Chen, J., and Huang, J. (2010) *Science* **329**, 693-696
87. MacKay, C., Déclais, A.-C., Lundin, C., Agostinho, A., Deans, A. J., MacArtney, T. J., Hofmann, K., Gartner, A., West, S. C., Helleday, T., Lilley, D. M. J., and Rouse, J. (2010) *Cell* **142**, 65-76
88. Smogorzewska, A., Desetty, R., Saito, T. T., Schlabach, M., Lach, F. P., Sowa, M. E., Clark, A. B., Kunkel, T. A., Harper, J. W., Colaiácovo, M. P., and Elledge, S. J. (2010) *Molecular Cell* **39**, 36-47
89. Bhagwat, N., Olsen, A. L., Wang, A. T., Hanada, K., Stuckert, P., Kanaar, R., D'Andrea, A., Niedernhofer, L. J., and McHugh, P. J. (2009) *Mol. Cell. Biol.*, MCB.00086-00009

90. Raschle, M., Knipscheer, P., Enoiu, M., Angelov, T., Sun, J., Griffith, J. D., Ellenberger, T. E., Scharer, O. D., and Walter, J. C. (2008) *Cell* **134**, 969-980
91. Bergstralh, D. T., and Sekelsky, J. (2008) *Trends Genet* **24**, 70-76
92. Kim, H., and D'Andrea, A. D. (2012) *Genes Dev* **26**, 1393-1408
93. Pâques, F., and Haber, J. E. (1999) *Microbiol Mol Biol Rev* **63**, 349-404
94. Gregg, S. Q., Robinson, A. R., and Niedernhofer, L. J. (2011) *DNA Repair* **10**, 781-791
95. Mortensen, U. H., Bendixen, C., Sunjevaric, I., and Rothstein, R. (1996) *Proceedings of the National Academy of Sciences of the United States of America* **93**, 10729-10734
96. Sugawara, N., Ira, G., and Haber, J. E. (2000) *Mol Cell Biol* **20**, 5300-5309
97. Fishman-Lobell, J., and Haber, J. E. (1992) *Science (New York, N.Y* **258**, 480-484
98. Ivanov, E. L., and Haber, J. E. (1995) *Mol Cell Biol* **15**, 2245-2251
99. Sugawara, N., Paques, F., Colaiacovo, M., and Haber, J. E. (1997) *Proceedings of the National Academy of Sciences of the United States of America* **94**, 9214-9219
100. Li, F., Dong, J., Pan, X., Oum, J.-H., Boeke, J. D., and Lee, S. E. (2008) *Molecular Cell* **30**, 325-335
101. Flott, S., Alabert, C., Toh, G. W., Toth, R., Sugawara, N., Campbell, D. G., Haber, J. E., Pasero, P., and Rouse, J. (2007) *Molecular and Cellular Biology* **27**, 6433-6445
102. Lyndaker, A. M., Goldfarb, T., and Alani, E. (2008) *Genetics* **179**, 1807-1821
103. Al-Minawi, A. Z., Saleh-Gohari, N., and Helleday, T. (2008) *Nucleic acids research* **36**, 1-9
104. Sargent, R. G., Rolig, R. L., Kilburn, A. E., Adair, G. M., Wilson, J. H., and Nairn, R. S. (1997) *Proceedings of the National Academy of Sciences of the United States of America* **94**, 13122-13127
105. Sargent, R. G., Meservy, J. L., Perkins, B. D., Kilburn, A. E., Intody, Z., Adair, G. M., Nairn, R. S., and Wilson, J. H. (2000) *Nucleic acids research* **28**, 3771-3778
106. Motycka, T. A., Bessho, T., Post, S. M., Sung, P., and Tomkinson, A. E. (2004) *The Journal of biological chemistry* **279**, 13634-13639
107. Niedernhofer, L. J., Essers, J., Weeda, G., Beverloo, B., de Wit, J., Muijtjens, M., Odijk, H., Hoeijmakers, J. H., and Kanaar, R. (2001) *The EMBO journal* **20**, 6540-6549
108. Roth, D. B., and Wilson, J. H. (1986) *Mol Cell Biol* **6**, 4295-4304
109. Ma, J.-L., Kim, E. M., Haber, J. E., and Lee, S. E. (2003) *Molecular and Cellular Biology* **23**, 8820-8828
110. Ahmad, A., Robinson, A. R., Duensing, A., van Drunen, E., Beverloo, H. B., Weisberg, D. B., Hasty, P., Hoeijmakers, J. H., and Niedernhofer, L. J. (2008) *Mol Cell Biol* **28**, 5082-5092
111. Gao, Y., Chaudhuri, J., Zhu, C., Davidson, L., Weaver, D. T., and Alt, F. W. (1998) *Immunity* **9**, 367-376
112. Yildiz, O., Kearney, H., Kramer, B. C., and Sekelsky, J. J. (2004) *Genetics* **167**, 263-273
113. Carpenter, A. T., and Baker, B. S. (1982) *Genetics* **101**, 81-89
114. Yildiz, O., Majumder, S., Kramer, B., and Sekelsky, J. J. (2002) *Mol Cell* **10**, 1503-1509
115. Hsia, K. T., Millar, M. R., King, S., Selfridge, J., Redhead, N. J., Melton, D. W., and Saunders, P. T. (2003) *Development (Cambridge, England)* **130**, 369-378
116. Zhu, X. D., Niedernhofer, L., Kuster, B., Mann, M., Hoeijmakers, J. H., and de Lange, T. (2003) *Mol Cell* **12**, 1489-1498
117. Munoz, P., Blanco, R., Flores, J. M., and Blasco, M. A. (2005) *Nature genetics* **37**, 1063-1071

118. Wu, Y., Zacal, N. J., Rainbow, A. J., and Zhu, X. D. (2007) *DNA repair* **6**, 157-166
119. Champoux, J. J. (2001) *Annu Rev Biochem* **70**, 369-413
120. Vance, J. R., and Wilson, T. E. (2001) *Mol Cell Biol* **21**, 7191-7198
121. Liu, C., Pouliot, J. J., and Nash, H. A. (2002) *Proceedings of the National Academy of Sciences of the United States of America* **99**, 14970-14975
122. Vance, J. R., and Wilson, T. E. (2002) *Proceedings of the National Academy of Sciences of the United States of America* **99**, 13669-13674
123. Guillet, M., and Boiteux, S. (2002) *The EMBO journal* **21**, 2833-2841
124. Fisher, L. A., Samson, L., and Bessho, T. (2011) *Chem Res Toxicol* **24**, 1876-1881
125. Boiteux, S., and Guillet, M. (2004) *DNA repair* **3**, 1-12
126. Torres-Ramos, C. A., Johnson, R. E., Prakash, L., and Prakash, S. (2000) *Mol Cell Biol* **20**, 3522-3528
127. Swanson, R. L., Morey, N. J., Doetsch, P. W., and Jinks-Robertson, S. (1999) *Mol Cell Biol* **19**, 2929-2935
128. Cleaver, J. E. (1968) *Nature* **218**, 652-656
129. Kraemer, K. H., Lee, M. M., and Scotto, J. (1987) *Arch Dermatol* **123**, 241-250
130. Matsumura, Y., Nishigori, C., Yagi, T., Imamura, S., and Takebe, H. (1998) *Human molecular genetics* **7**, 969-974
131. Tian, M., Shinkura, R., Shinkura, N., and Alt, F. W. (2004) *Mol Cell Biol* **24**, 1200-1205
132. Meira, L. B., Graham, J. M., Jr., Greenberg, C. R., Busch, D. B., Doughty, A. T., Ziffer, D. W., Coleman, D. M., Savre-Train, I., and Friedberg, E. C. (2000) *American journal of human genetics* **66**, 1221-1228
133. Graham, J. M., Jr., Anyane-Yeboa, K., Raams, A., Appeldoorn, E., Kleijer, W. J., Garritsen, V. H., Busch, D., Edersheim, T. G., and Jaspers, N. G. (2001) *American journal of human genetics* **69**, 291-300
134. Lobitz, S., and Velleuer, E. (2006) *Nat Rev Cancer* **6**, 893-898
135. Deans, A. J., and West, S. C. (2011) *Nat Rev Cancer* **11**, 467-480
136. McWhir, J., Selfridge, J., Harrison, D. J., Squires, S., and Melton, D. W. (1993) *Nature genetics* **5**, 217-224
137. Weeda, G., Donker, I., de Wit, J., Morreau, H., Janssens, R., Vissers, C. J., Nigg, A., van Steeg, H., Bootsma, D., and Hoeijmakers, J. H. J. (1997) *Curr Biol* **7**, 427-439
138. Selfridge, J., Pow, A. M., McWhir, J., Magin, T. M., and Melton, D. W. (1992) *Somatic cell and molecular genetics* **18**, 325-336
139. Nunez, F., Chipchase, M. D., Clarke, A. R., and Melton, D. W. (2000) *Faseb J* **14**, 1073-1082
140. de Vries, A., van Oostrom, C. T., Hofhuis, F. M., Dortant, P. M., Berg, R. J., de Gruijl, F. R., Wester, P. W., van Kreijl, C. F., Capel, P. J., van Steeg, H., and et al. (1995) *Nature* **377**, 169-173
141. Prasher, J. M., Lalai, A. S., Heijmans-Antonissen, C., Ploemacher, R. E., Hoeijmakers, J. H., Touw, I. P., and Niedernhofer, L. J. (2005) *The EMBO journal* **24**, 861-871
142. Parmar, K., D'Andrea, A., and Niedernhofer, L. J. (2009) *Mutation research* **668**, 133-140
143. Dollé, M. E. T., Busuttil, R. A., Garcia, A. M., Wijnhoven, S., van Drunen, E., Niedernhofer, L. J., van der Horst, G., Hoeijmakers, J. H. J., van Steeg, H., and Vijg, J. (2006) *Mutat Res* **596**, 22-35

144. Nevedomskaya, E., Meissner, A., Goraler, S., de Waard, M., Ridwan, Y., Zondag, G., van der Pluijm, I., Deelder, A. M., and Mayboroda, O. A. (2010) *J Proteome Res* **9**, 3680-3687
145. Selfridge, J., Hsia, K. T., Redhead, N. J., and Melton, D. W. (2001) *Nucleic acids research* **29**, 4541-4550
146. Lawrence, N. J., Sacco, J. J., Brownstein, D. G., Gillingwater, T. H., and Melton, D. W. (2008) *DNA repair* **7**, 281-291
147. Doig, J., Anderson, C., Lawrence, N. J., Selfridge, J., Brownstein, D. G., and Melton, D. W. (2006) *Oncogene* **25**, 6229-6238
148. De Silva, I. U., McHugh, P. J., Clingen, P. H., and Hartley, J. A. (2000) *Molecular and cellular biology* **20**, 7980-7990
149. Murray, D., Macann, A., Hanson, J., and Rosenberg, E. (1996) *International journal of radiation biology* **69**, 319-327
150. Murray, D., and Rosenberg, E. (1996) *Mutation research* **364**, 217-226
151. Wu, Y., Mitchell, T. R., and Zhu, X. D. (2008) *Mechanisms of ageing and development* **129**, 602-610
152. Shen, M. R., Jones, I. M., and Mohrenweiser, H. (1998) *Cancer research* **58**, 604-608
153. Yu, J. J., Mu, C., Lee, K. B., Okamoto, A., Reed, E. L., Bostick-Bruton, F., Mitchell, K. C., and Reed, E. (1997) *Mutation research* **382**, 13-20
154. Takenaka, T., Yano, T., Kiyohara, C., Miura, N., Kouso, H., Ohba, T., Kometani, T., Shoji, F., Yoshino, I., and Maehara, Y. (2009) *Lung Cancer*
155. Kalikaki, A., Kanaki, M., Vassalou, H., Souglakos, J., Voutsina, A., Georgoulias, V., and Mavroudis, D. (2009) *Clinical lung cancer* **10**, 118-123
156. Su, D., Ma, S., Liu, P., Jiang, Z., Lv, W., Zhang, Y., Deng, Q., Smith, S., and Yu, H. (2007) *Lung Cancer* **56**, 281-288
157. Zienolddiny, S., Campa, D., Lind, H., Ryberg, D., Skaug, V., Stangeland, L., Phillips, D. H., Canzian, F., and Haugen, A. (2006) *Carcinogenesis* **27**, 560-567
158. Park, S. Y., Hong, Y. C., Kim, J. H., Kwak, S. M., Cho, J. H., Lee, H. L., and Ryu, J. S. (2006) *Medical oncology (Northwood, London, England)* **23**, 489-498
159. Isla, D., Sarries, C., Rosell, R., Alonso, G., Domine, M., Taron, M., Lopez-Vivanco, G., Camps, C., Botia, M., Nunez, L., Sanchez-Ronco, M., Sanchez, J. J., Lopez-Brea, M., Barneto, I., Paredes, A., Medina, B., Artal, A., and Lianes, P. (2004) *Ann Oncol* **15**, 1194-1203
160. Kiyohara, C., and Yoshimasu, K. (2007) *International journal of medical sciences* **4**, 59-71
161. Pan, J., Lin, J., Izzo, J., Liu, Y., Xing, J., Huang, M., Ajani, J., and Wu, X. (2009) *Carcinogenesis*
162. Doecke, J., Zhao, Z. Z., Pandeya, N., Sadeghi, S., Stark, M., Green, A. C., Hayward, N. K., Webb, P. M., and Whiteman, D. C. (2008) *International journal of cancer* **123**, 174-180
163. Steffensen, K. D., Waldstrom, M., Jeppesen, U., Brandslund, I., and Jakobsen, A. (2008) *Int J Gynecol Cancer* **18**, 702-710
164. Quintela-Fandino, M., Hitt, R., Medina, P. P., Gamarra, S., Manso, L., Cortes-Funes, H., and Sanchez-Cespedes, M. (2006) *J Clin Oncol* **24**, 4333-4339
165. Yin, J., Vogel, U., Gerdes, L. U., Dybdahl, M., Bolund, L., and Nexo, B. A. (2003) *Biochemical genetics* **41**, 27-37

166. Garcia-Closas, M., Malats, N., Real, F. X., Welch, R., Kogevinas, M., Chatterjee, N., Pfeiffer, R., Silverman, D., Dosemeci, M., Tardon, A., Serra, C., Carrato, A., Garcia-Closas, R., Castano-Vinyals, G., Chanock, S., Yeager, M., and Rothman, N. (2006) *Cancer Epidemiol Biomarkers Prev* **15**, 536-542
167. McWilliams, R. R., Bamlet, W. R., Cunningham, J. M., Goode, E. L., de Andrade, M., Boardman, L. A., and Petersen, G. M. (2008) *Cancer research* **68**, 4928-4935
168. Hooker, S., Bonilla, C., Akereyeni, F., Ahaghotu, C., and Kittles, R. A. (2008) *Prostate Cancer Prostatic Dis* **11**, 349-356
169. Hansen, R. D., Sorensen, M., Tjonneland, A., Overvad, K., Wallin, H., Raaschou-Nielsen, O., and Vogel, U. (2008) *BMC cancer* **8**, 54
170. Monzo, M., Moreno, I., Navarro, A., Ibeas, R., Artells, R., Gel, B., Martinez, F., Moreno, J., Hernandez, R., and Navarro-Vigo, M. (2007) *Oncology* **72**, 364-370
171. Chen, H., Shao, C., Shi, H., Mu, Y., Sai, K., and Chen, Z. (2007) *Journal of neuro-oncology* **82**, 257-262
172. Hirata, H., Hinoda, Y., Matsuyama, H., Tanaka, Y., Okayama, N., Suehiro, Y., Zhao, H., Urakami, S., Kawamoto, K., Kawakami, T., Igawa, M., Naito, K., and Dahiya, R. (2006) *Biochemical and biophysical research communications* **342**, 1058-1062
173. Chung, H. H., Kim, M. K., Kim, J. W., Park, N. H., Song, Y. S., Kang, S. B., and Lee, H. P. (2006) *Gynecologic oncology* **103**, 1031-1037
174. Altaha, R., Liang, X., Yu, J. J., and Reed, E. (2004) *International journal of molecular medicine* **14**, 959-970
175. Dabholkar, M., Bostick-Bruton, F., Weber, C., Bohr, V. A., Egwuagu, C., and Reed, E. (1992) *Journal of the National Cancer Institute* **84**, 1512-1517
176. Dabholkar, M., Vionnet, J., Bostick-Bruton, F., Yu, J. J., and Reed, E. (1994) *J Clin Invest* **94**, 703-708
177. Joshi, M. B., Shirota, Y., Danenberg, K. D., Conlon, D. H., Salonga, D. S., Herndon, J. E., 2nd, Danenberg, P. V., and Harpole, D. H., Jr. (2005) *Clin Cancer Res* **11**, 2215-2221
178. Langer, R., Specht, K., Becker, K., Ewald, P., Bekesch, M., Sarbia, M., Busch, R., Feith, M., Stein, H. J., Siewert, J. R., and Hofler, H. (2005) *Clin Cancer Res* **11**, 7462-7469
179. Lord, R. V., Brabender, J., Gandara, D., Alberola, V., Camps, C., Domine, M., Cardenal, F., Sanchez, J. M., Gumerlock, P. H., Taron, M., Sanchez, J. J., Danenberg, K. D., Danenberg, P. V., and Rosell, R. (2002) *Clin Cancer Res* **8**, 2286-2291
180. Metzger, R., Leichman, C. G., Danenberg, K. D., Danenberg, P. V., Lenz, H. J., Hayashi, K., Groshen, S., Salonga, D., Cohen, H., Laine, L., Crookes, P., Silberman, H., Baranda, J., Konda, B., and Leichman, L. (1998) *J Clin Oncol* **16**, 309-316
181. Reed, E., Dabholkar, M., Thornton, K., Thompson, C., Yu, J. J., and Bostick-Bruton, F. (2000) *Oncology reports* **7**, 1123-1128
182. Shirota, Y., Stoehlmacher, J., Brabender, J., Xiong, Y. P., Uetake, H., Danenberg, K. D., Groshen, S., Tsao-Wei, D. D., Danenberg, P. V., and Lenz, H. J. (2001) *J Clin Oncol* **19**, 4298-4304
183. Rosell, R., Lord, R. V., Taron, M., and Reguart, N. (2002) *Lung Cancer* **38**, 217-227
184. Warnecke-Eberz, U., Metzger, R., Miyazono, F., Baldus, S. E., Neiss, S., Brabender, J., Schaefer, H., Doerfler, W., Bollschweiler, E., Dienes, H. P., Mueller, R. P., Danenberg, P. V., Hoelscher, A. H., and Schneider, P. M. (2004) *Clin Cancer Res* **10**, 3794-3799
185. Cobo, M., Isla, D., Massuti, B., Montes, A., Sanchez, J. M., Provencio, M., Vinolas, N., Paz-Ares, L., Lopez-Vivanco, G., Munoz, M. A., Felip, E., Alberola, V., Camps, C.,

Domine, M., Sanchez, J. J., Sanchez-Ronco, M., Danenberg, K., Taron, M., Gandara, D., and Rosell, R. (2007) *J Clin Oncol* **25**, 2747-2754
186. Lee, K. H., Min, H. S., Han, S. W., Oh, D. Y., Lee, S. H., Kim, D. W., Im, S. A., Chung, D. H., Kim, Y. T., Kim, T. Y., Heo, D. S., Bang, Y. J., Sung, S. W., and Kim, J. H. (2008) *Lung Cancer* **60**, 401-407
187. Olaussen, K. A., Dunant, A., Fouret, P., Brambilla, E., Andre, F., Haddad, V., Taranchon, E., Filipits, M., Pirker, R., Popper, H. H., Stahel, R., Sabatier, L., Pignon, J. P., Tursz, T., Le Chevalier, T., and Soria, J. C. (2006) *The New England journal of medicine* **355**, 983-991
188. Hwang, I. G., Ahn, M. J., Park, B. B., Ahn, Y. C., Han, J., Lee, S., Kim, J., Shim, Y. M., Ahn, J. S., and Park, K. (2008) *Cancer*
189. Zheng, Z., Chen, T., Li, X., Haura, E., Sharma, A., and Bepler, G. (2007) *The New England journal of medicine* **356**, 800-808
190. Niedernhofer, L. J., Bhagwat, N., and Wood, R. D. (2007) *The New England journal of medicine* **356**, 2538-2540; author reply 2540-2531
191. Bhagwat, N. R., Roginskaya, V. Y., Acquafondata, M. B., Dhir, R., Wood, R. D., and Niedernhofer, L. J. (2009) *Cancer research* **69**, 6831-6838
192. Fagbemi, A. F., Orelli, B., and Scharer, O. D. (2011) *DNA Repair (Amst)* **10**, 722-729
193. Hanawalt, P. C., and Spivak, G. (2008) *Nat Rev Mol Cell Biol* **9**, 958-970
194. Sugasawa, K., Okuda, Y., Saijo, M., Nishi, R., Matsuda, N., Chu, G., Mori, T., Iwai, S., Tanaka, K., Tanaka, K., and Hanaoka, F. (2005) *Cell* **121**, 387-400
195. Ogi, T., Limsirichaikul, S., Overmeer, R. M., Volker, M., Takenaka, K., Cloney, R., Nakazawa, Y., Niimi, A., Miki, Y., Jaspers, N. G., Mullenders, L. H., Yamashita, S., Fousteri, M. I., and Lehmann, A. R. (2010) *Mol Cell* **37**, 714-727
196. Staresincic, L., Fagbemi, A. F., Enzlin, J. H., Gourdin, A. M., Wijgers, N., Dunand-Sauthier, I., Giglia-Mari, G., Clarkson, S. G., Vermeulen, W., and Schärer, O. D. (2009) *EMBO J* **28**, 1111-1120
197. Lehmann, A. R. (2003) *Biochimie* **85**, 1101-1111
198. Schärer, O. D. (2008) *Adv Exp Med Biol* **637**, 83-92
199. McWhir, J., Selfridge, J., Harrison, D. J., Squires, S., and Melton, D. W. (1993) *Nat Genet* **5**, 217-224
200. Niedernhofer, L. J., Odijk, H., Budzowska, M., van Drunen, E., Maas, A., Theil, A. F., de Wit, J., Jaspers, N. G., Beverloo, H. B., Hoeijmakers, J. H., and Kanaar, R. (2004) *Mol Cell Biol* **24**, 5776-5787
201. Bhagwat, N., Olsen, A. L., Wang, A. T., Hanada, K., Stuckert, P., Kanaar, R., D'Andrea, A., Niedernhofer, L. J., and McHugh, P. J. (2009) *Mol Cell Biol* **29**, 6427-6437
202. de Laat, W. L., Sijbers, A. M., Odijk, H., Jaspers, N. G., and Hoeijmakers, J. H. (1998) *Nucleic acids research* **26**, 4146-4152
203. Stauffer, M. E., and Chazin, W. J. (2004) *J Biol Chem* **279**, 30915-30918
204. Shivji, M. K., Moggs, J. G., Kuraoka, I., and Wood, R. D. (1999) *Methods Mol Biol* **113**, 373-392
205. Gillet, L. C., Alzeer, J., and Schärer, O. D. (2005) *Nucleic acids research* **33**, 1961-1969
206. Biggerstaff, M., and Wood, R. D. (1999) *Methods Mol Biol* **113**, 357-372
207. Salmon, P., Kindler, V., Ducrey, O., Chapuis, B., Zubler, R. H., and Trono, D. (2000) *Blood* **96**, 3392-3398
208. Salmon, P., and Trono, D. (2006) *Curr Protoc Neurosci* **Chapter 4**, Unit 4 21

209. Doherty, A. J., Serpell, L. C., and Ponting, C. P. (1996) *Nucleic acids research* **24**, 2488-2497.
210. Dunand-Sauthier, I., Hohl, M., Thorel, F., Jaquier-Gubler, P., Clarkson, S. G., and Scharer, O. D. (2005) *J Biol Chem* **280**, 7030-7037
211. Pingoud, A., Fuxreiter, M., Pingoud, V., and Wende, W. (2005) *Cell Mol Life Sci* **62**, 685-707
212. Crossan, G. P., van der Weyden, L., Rosado, I. V., Langevin, F., Gaillard, P.-H. L., McIntyre, R. E., Project, S. M. G., Gallagher, F., Kettunen, M. I., Lewis, D. Y., Brindle, K., Arends, M. J., Adams, D. J., and Patel, K. J. (2011) *Nat Genet* **43**, 147-152
213. Kim, Y., Lach, F. P., Desetty, R., Hanenberg, H., Auerbach, A. D., and Smogorzewska, A. (2011) *Nat Genet* **43**, 142-146
214. Stoepker, C., Hain, K., Schuster, B., Hilhorst-Hofstee, Y., Rooimans, M. A., Steltenpool, J., Oostra, A. B., Eirich, K., Korthof, E. T., Nieuwint, A. W. M., Jaspers, N. G. J., Bettecken, T., Joenje, H., Schindler, D., Rouse, J., and de Winter, J. P. (2011) *Nat Genet* **43**, 138-141
215. Antonio Casado, J., Callén, E., Jacome, A., Río, P., Castella, M., Lobitz, S., Ferro, T., Muñoz, A., Sevilla, J., Cantalejo, A., Cela, E., Cervera, J., Sánchez-Calero, J., Badell, I., Estella, J., Dasí, A., Olivé, T., José Ortega, J., Rodriguez-Villa, A., Tapia, M., Molinés, A., Madero, L., Segovia, J. C., Neveling, K., Kalb, R., Schindler, D., Hanenberg, H., Surrallés, J., and Bueren, J. A. (2007) *J Med Genet* **44**, 241-249
216. Castella, M., Pujol, R., Callén, E., Ramírez, M. J., Casado, J. A., Talavera, M., Ferro, T., Muñoz, A., Sevilla, J., Madero, L., Cela, E., Beléndez, C., de Heredia, C. D., Olivé, T., de Toledo, J. S., Badell, I., Estella, J., Dasí, Á., Rodríguez-Villa, A., Gómez, P., Tapia, M., Molinés, A., Figuera, Á., Bueren, J. A., and Surrallés, J. (2011) *J Med Genet* **48**, 242-250
217. De Silva, I. U., McHugh, P. J., Clingen, P. H., and Hartley, J. A. (2000) *Molecular and Cellular Biology* **20**, 7980-7990
218. Rothfuss, A., and Grompe, M. (2004) *Molecular and Cellular Biology* **24**, 123-134
219. Trujillo, J. P., Mina, L. B., Pujol, R., Bogliolo, M., Andrieux, J., Holder, M., Schuster, B., Schindler, D., and Surrallés, J. (2012) *Blood* **120**, 86-89
220. Zhou, W., Otto, E. A., Cluckey, A., Airik, R., Hurd, T. W., Chaki, M., Diaz, K., Lach, F. P., Bennett, G. R., Gee, H. Y., Ghosh, A. K., Natarajan, S., Thongthip, S., Veturi, U., Allen, S. J., Janssen, S., Ramaswami, G., Dixon, J., Burkhalter, F., Spoendlin, M., Moch, H., Mihatsch, M. J., Verine, J., Reade, R., Soliman, H., Godin, M., Kiss, D., Monga, G., Mazzucco, G., Amann, K., Artunc, F., Newland, R. C., Wiech, T., Zschiedrich, S., Huber, T. B., Friedl, A., Slaats, G. G., Joles, J. A., Goldschmeding, R., Washburn, J., Giles, R. H., Levy, S., Smogorzewska, A., and Hildebrandt, F. (2012) *Nat Genet* **44**, 910-915
221. Salewsky, B., Schmiester, M., Schindler, D., Digweed, M., and Demuth, I. (2012) *Hum Mol Genet*
222. Gregg, S. Q., Gutiérrez, V., Robinson, A. R., Woodell, T., Nakao, A., Ross, M. A., Michalopoulos, G. K., Rigatti, L., Rothermel, C. E., Kamileri, I., Garinis, G., Stolz, D. B., and Niedernhofer, L. J. (2011) *Hepatology (Baltimore, Md)*
223. Levy-Lahad, E. (2010) *Nat Genet* **42**, 368-369
224. Meindl, A., Hellebrand, H., Wiek, C., Erven, V., Wappenschmidt, B., Niederacher, D., Freund, M., Lichtner, P., Hartmann, L., Schaal, H., Ramser, J., Honisch, E., Kubisch, C.,

Wichmann, H. E., Kast, K., Deissler, H., Engel, C., Müller-Myhsok, B., Neveling, K., Kiechle, M., Mathew, C. G., Schindler, D., Schmutzler, R. K., and Hanenberg, H. (2010) *Nat Genet* **42**, 410-414
225. Dull, T., Zufferey, R., Kelly, M., Mandel, R. J., Nguyen, M., Trono, D., and Naldini, L. (1998) *J Virol* **72**, 8463-8471
226. Almarza, E., Río, P., Meza, N. W., Aldea, M., Agirre, X., Guenechea, G., Segovia, J. C., and Bueren, J. A. (2007) *Mol Ther* **15**, 1487-1494
227. Limsirichaikul, S., Niimi, A., Fawcett, H., Lehmann, A., Yamashita, S., and Ogi, T. (2009) *Nucleic acids research* **37**, e31
228. Wood, R. D., Robins, P., and Lindahl, T. (1988) *Cell* **53**, 97-106
229. Moggs, J. G., Yarema, K. J., Essigmann, J. M., and Wood, R. D. (1996) *J Biol Chem* **271**, 7177-7186
230. Milne, R. L., Ribas, G., González-Neira, A., Fagerholm, R., Salas, A., González, E., Dopazo, J., Nevanlinna, H., Robledo, M., and Benítez, J. (2006) *Cancer Res* **66**, 9420-9427
231. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and Subgroup, G. P. D. P. (2009) *Bioinformatics* **25**, 2078-2079
232. Rios, D., McLaren, W. M., Chen, Y., Birney, E., Stabenau, A., Flicek, P., and Cunningham, F. (2010) *BMC Bioinformatics* **11**, 238
233. Ng, S. B., Buckingham, K. J., Lee, C., Bigham, A. W., Tabor, H. K., Dent, K. M., Huff, C. D., Shannon, P. T., Jabs, E. W., Nickerson, D. A., Shendure, J., and Bamshad, M. J. (2010) *Nat Genet* **42**, 30-35
234. Salmon, P., and Trono, D. (2006) *Curr Protoc Neurosci* **Chapter 4**, Unit 4.21
235. Wood, D. K., Weingeist, D. M., Bhatia, S. N., and Engelward, B. P. (2010) *Proc Natl Acad Sci USA* **107**, 10008-10013
236. Spanswick, V. J., Hartley, J. M., and Hartley, J. A. (2010) *Methods Mol Biol* **613**, 267-282