DISSERTATION
ZUR ERLANGUNG DES NATURWISSENSCHAFTLICHEN DOKTORGRADES
DER JULIUS-MAXIMILIANS-UNIVERSITÄT WÜRZBURG

GENOTYPING FANCONI ANEMIA

- FROM KNOWN TO NOVEL GENES -
- FROM CLASSICAL GENETIC APPROACHES TO NEXT GENERATION SEQUENCING -

VORGELEGT VON
BEATRICE SCHUSTER
GEBOREN IN LICHTENFELS

2012
INSTITUT FÜR HUMANGENETIK

Erklärung gemäss § 4 Absatz 3 der Promotionsordnung der Fakultät für Biologie der Julius-Maximilians-Universität Würzburg

1. Hiermit erkläre ich, dass ich die vorliegende Dissertation selbstständig und nur unter Verwendung der angegebenen Quellen und Hilfsmittel angefertigt habe.

2. Desweiteren bestätige ich, dass diese Arbeit weder in gleicher noch in ähnlicher Form einem anderen Prüfungsverfahren vorgelegen hat.

3. Ich erkläre ausserdem, dass ich ausser den mit dem Zulassungsantrag urkundlich vorgelegten Graden keine anderen akademischen Grade erworben habe oder zu erwerben versucht habe.

Beatrice Schuster

INDEX

Zusammenfassung

Die Fanconi Anämie (FA) ist eine autosomal rezessiv oder X-chromosomal vererbte Erkrankung, deren charakteristisches diagnostisches Merkmal das progressive Versagen des Knochenmarks darstellt. Viele, jedoch nicht alle Patienten leiden zusätzlich an kongenitalen Fehlbildungen, wie Radialstrahl-Anomalien oder Minderwuchs. Im Vergleich zur normalen Bevölkerung steigt zudem im jungen Erwachsenenalter das Risiko für hämatologische und auch solide Tumoren um ein Vielfaches. Verantwortlich hierfür ist sehr wahrscheinlich der zugrunde liegende Defekt in der Reparatur von DNA-Interstrang-Quervernetzungen. Diese Art der Läsion blockiert wichtige zelluläre Prozesse wie Transkription und Replikation, und kann daher nicht nur zur Entartung oder vorzeitigen Alterung der Zellen, sondern auch zu stark erhöhten Apoptose-Raten führen. Zur Entfernung dieser Quervernetzungen hat die Evolution ein komplexes Netzwerk an verschiedenen Reparaturwegen hervorgebracht, das nur in Vertebraten vollständig konserviert ist. Der sogenannte FA/BRCA-Reparaturweg ist in der Lage Quervernetzungen an stagnierten Replikationsgabeln zu erkennen und zu entfernen. Heute kennen wir 15 Gene (*FANCA, -B, -C, -D1, -D2, -E, -F, -G, -I, -J, -L, -M, -N, -O* und *-P*), deren Produkte in diesem Weg involviert sind und deren pathogene Veränderung zur Ausprägung des FA-Phänotyps führen. Rund 80% aller Fälle können durch biallelische Mutationen in *FANCA*, *FANCC* und *FANCG* erklärt werden. Pathogene Varianten in anderen Genen werden weitaus seltener gefunden und ein kleiner Anteil der Patienten kann keiner der bekannten Komplementationsgruppen zugeordnet werden. Das Ziel dieser Arbeit war es, den ursächlichen genetischen Defekt in diesen Patienten aufzudecken. Untersuchungen an den Kandidatengenen *FAN1*, *MHF1* und *MHF2* konnten keine pathogenen Veränderungen identifizieren. *FAN1* konnte darüber hinaus gänzlich als Kandidatengen ausgeschlossen werden, da Patienten mit einer homozygoten *FAN1*-Deletion keinen FA-Phänotyp zeigten. Im Fall von *MHF1* und *MHF2* sind Mutationsträger wahrscheinlich sehr selten oder unterscheiden sich in ihrem Phänotyp von den bisher bekannten FA Patienten. Nichtsdestotrotz trug diese Arbeit maßgeblich zur Aufklärung der genetischen Ursache in den Untergruppen FA-O, FA-P und FA-Q bei. Ursächlich für den Subtyp FA-O sind biallelische Mutationen in *RAD51C*, einem Paralog der Rekombinase RAD51, mit offenbar entscheidender Funktion in der homologen Rekombinationsreparatur. Da der einzige bislang beschriebene Patient zum Zeitpunkt der Veröffentlichung zwar charakteristische Fehlbildungen, aber weder hämatologische Auffälligkeiten, noch maligne Veränderungen zeigte, wird *RAD51C* (*FANCO*) bisher als zugrunde liegendes Gen einer FA-ähnlichen Krankheit bezeichnet. Bei der Identifizierung von *SLX4* als ursächliches Gen der Untergruppe FA-P gab es hingegen keine Zweifel; alle Patienten zeigten einen sehr typischen Phänotyp. SLX4 (FANCP) scheint eine entscheidende Rolle bei der Exzision von DNA-Quervernetzungen zu spielen, indem es die Funktion oder richtige Positionierung von Struktur-

spezifischen Nukleasen koordiniert. Eine dieser Nukleasen ist das Heterodimer XPF/ERCC1. *XPF* liegt wahrscheinlich der Komplementationsgruppe FA-Q zugrunde und ist das erste FA-Gen, das mittels Next Generation Sequencing (NGS) identifiziert wurde. Interessanterweise wurde es zuvor bereits als genetische Ursache von Xeroderma pigmentosum und segmentärer Progerie beschrieben. Diese Studie konnte jedoch belegen, dass die jeweiligen Mutationen die Proteinfunktion derart unterschiedlich beeinflussen, dass es tatsächlich zur Ausprägung von drei divergenten Phänotypen kommen kann.

Neben der Kandidatengensuche war ein weiteres Ziel dieser Arbeit die Implementierung neuer Techniken für die FA-Genotypisierung. Klassische Methoden der Molekulargenetik sollten hierfür durch Anwendungen des NGS ergänzt oder gänzlich ersetzt werden. Die Hochdurchsatz-Sequenzierung des gesamten Exoms erwies sich als geeignet und kann Komplementationsgruppen-Zuordnung und Mutationsanalyse in einem Schritt vereinen. Durch die Etablierung einer FA-spezifischen bioinformatischen Datenanalyse konnte im Rahmen dieser Arbeit der genetische Defekt bereits mehrerer Patienten aufgeklärt werden. Im Besonderen konnte ein weiterer Patient der neuen, noch wenig charakterisierten Untergruppe FA-P zugeordnet werden.

Insgesamt trug diese Arbeit also nicht nur zur weiteren Vervollständigung des FA/BRCA-Reparaturweges bei, indem drei neue FA-Gene hinzugefügt wurden; sie zeigte außerdem, dass klassische Methoden der Molekulargenetik sowohl in Forschung als auch Diagnostik künftig durch das NGS ersetzt werden könnten.

Summary

Fanconi anemia (FA) is an autosomal recessive or X-chromosomal inherited disorder, which is not only phenotypically but also genotypically very heterogeneous. While its hallmark feature is progressive bone marrow failure, many yet not all patients suffer additionally from typical congenital malformations like radial ray defects and growth retardation. In young adulthood the cumulative risk for developing hematological or other malignancies is compared to the general population several hundred-fold increased. The underlying molecular defect is the deficiency of DNA interstrand crosslink (ICL) repair. ICLs are deleterious lesions, which interfere with crucial cellular processes like transcription and replication and thereby can lead to malignant transformation, premature senescence or cell death. To overcome this threat evolution developed a highly complex network of interacting DNA repair pathways, which is conserved completely only in vertebrates. The so called FA/BRCA DNA damage response pathway is able to recognize ICLs on stalled replication forks and promotes their repair through homologous recombination (HR). Today we know 15 FA genes (*FANCA, -B, -C, -D1, -D2, -E, -F, -G, -I, -J, -L, -M, -N, -O* and *-P*) whose products are involved in this pathway. Although more than 80% of FA patients carry biallelic mutations in either *FANCA, FANCC* or *FANCG*, there are still some who cannot be assigned to any of the known complementation groups. This work aimed to indentify the disease causing mutations in a cohort of those unassigned patients.

Initial screens of the candidate genes *FAN1*, *MHF1* and *MHF2* did not reveal any pathogenic alterations. Moreover, *FAN1* could be excluded as FA candidate gene because patients carrying a homozygous microdeletion including the *FAN1* locus did not show a phenotype comparable to FA patients. In the case of *MHF1* and *MHF2* the reason for the negative screening result is not clear. Mutation carriers might be rare or, regarding the diverse and also FA pathway independent protein functions, phenotypically not comparable to FA patients. Nevertheless, this study contributed to the identification and characterization of the most recent members of the FA pathway - *RAD51C* (*FANCO*), *SLX4* (*FANCP*) and *XPF* (*FANCQ*). *FANCO* is one of the RAD51 paralogs and is involved in crucial steps of HR. But since the only reported FA-O patient has so far not developed any hematological anomalies, *FANCO* is tentatively designated as gene underlying an FA-like disorder. In contrast, patients carrying biallelic mutations in *FANCP* do not only show hematological anomalies, but as well congenital malformations typical for FA. The distinct role of FANCP in the FA pathway could not be determined, but it is most likely the coordination of structure-specific nucleases during ICL excision. One of these nucleases is the heterodimer XPF/ERCC1. *XPF* is probably disease causing in the complementation group FA-Q and is the first FA gene, which was identified by Next Generation Sequencing (NGS). Extraordinarily is that mutations in this gene had previously been reported to cause two other disorders, xeroderma

pigmentosum and segmental progeria. Despite some overlaps, it was shown that the divergent phenotypes could clearly be distinguished and are caused by distinct functional defects of XPF. Additionally, this work aimed to improve and accelerate the genotyping process of FA patients in general. Therefore, classical approaches should be complemented or fully replaced by approaches using NGS. Massively parallel sequencing of the whole exome proved to be most appropriate and the establishment of an FA-specific analysis pipeline facilitated improved molecular diagnostics by combining complementation group assignment and mutation analysis in one step. Consequently two NGS studies revealed the pathogenic defect in several previously unassigned FA patients and thereby added another patient to one of the most recent subtypes, FA-P.

In summary, this work contributed not only to further completion of the FA/BRCA DNA repair network by adding three novel genes, it also showed that classical molecular approaches for research as well as for diagnostics could be replaced by NGS.

1. Introduction

1.1 Fanconi Anemia

1.1.1 First Description and Clinical Phenotype

Fanconi Anemia (FA) was first described in 1927 by the Swiss pediatrician Guido Fanconi (Fig.1A). In his original publication he reported in German language on three brothers (Fig.1B), suffering from familial pernicious anemia. Later on he realized that not only erythropoiesis but the formation of all hematological cell lineages are affected, and that bone marrow failure might be the late manifestation of a highly complex disorder associated with variable congenital defects like microcephaly and hyperpigmentation of the skin (Fanconi 1927; Lobitz and Velleuer 2006). Today we know that FA is a very rare inherited disease with a prevalence of 4-7 per million live births. The estimated carrier frequency is about 1:200 in the general population, but in isolated ethnical groups, e.g. in the Ashkenazi Jews or Spanish gypsies, it might be far below 1:100 (Callen, Casado et al. 2005; Gulbis, Eleftheriou et al. 2010; Rosenberg, Tamary et al. 2011).

The clinical presentation of FA is heterogeneous and in 25-30% percent of patients no obvious congenital abnormalities can be found. Nevertheless there are very typical FA features. Most frequently reported are radial ray defects, especially hypolasia of one or both thumbs and the radius (Fig.2A). Not less prevalent are hypo- and hyperpigmentation of the skin, often in form of café-au-lait and vitilgo spots (Fig.2B). In more than 60% of all cases, FA is associated with pre- and postnatal growth retardation (Fig.2C), which sometimes but not always can be related to hypothyroidism or growth hormone deficiency. Microcephaly, microphthalmia and hearing loss with or without malformations of the external ear are found in 10-20% of FA patients. Interestingly, in males the incidence of genital abnormalities is much higher than in females. Undescended or absent testes, micropenis and hypospadias
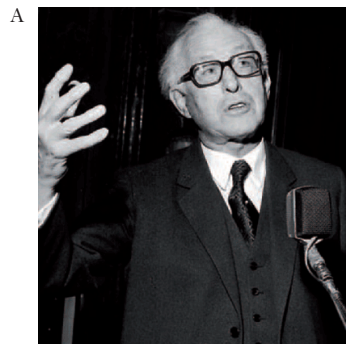
Fig. 1 Guido Fanconi and his first description of Fanconi anemia (FA). (A) G. Fanconi at the age of 83 years. The Swiss pediatrician first described the clinical phenotype of FA. *Kindly provided by Andreas Fanconi (Zurich, Switzerland)* (B) Original slide presented by Guido Fanconi in German language. It shows two of the first three FA patients and their pedigree. *Kindly provided by Felix H Sennhauser (Zurich, Switzerland). Both pictures have previously been published in Lobitz and Velleuer (2006).*
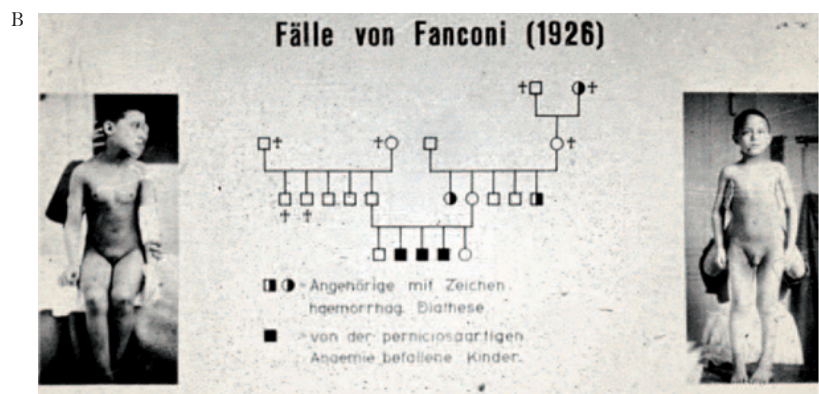
are common and regulary associated with male infertility, while pregnancy in females is possible in 98%. Regarding the inner organs, renal anomalies are most common. They can be uni- or bilateral and include hypo- or aplasia as well as horseshoe kidneys and double ureters. Less frequent are cardiopulmonary and gastrointestinal defects. Few cases have been reported with malformations of the central nervous system, like agenesis of the corpus callosum. However, intellectual disability is a less common FA feature (Alter and Kupfer 1993; Tischkowitz and Hodgson 2003; Auerbach 2009).

The hematological abnormalities (Fig.2D) are eponymous and very important in FA. Though blood counts are usually normal at birth, the cumulative risk of any hematological finding is more than 90% by age 50 years. Often FA is not diagnosed before evidence suggests bone marrow failure (BMF). Starting with macrocytosis and thrombocytopenia, pancytopenia typically becomes manifested in the first or second decade of life with a median age of onset of 7 years.
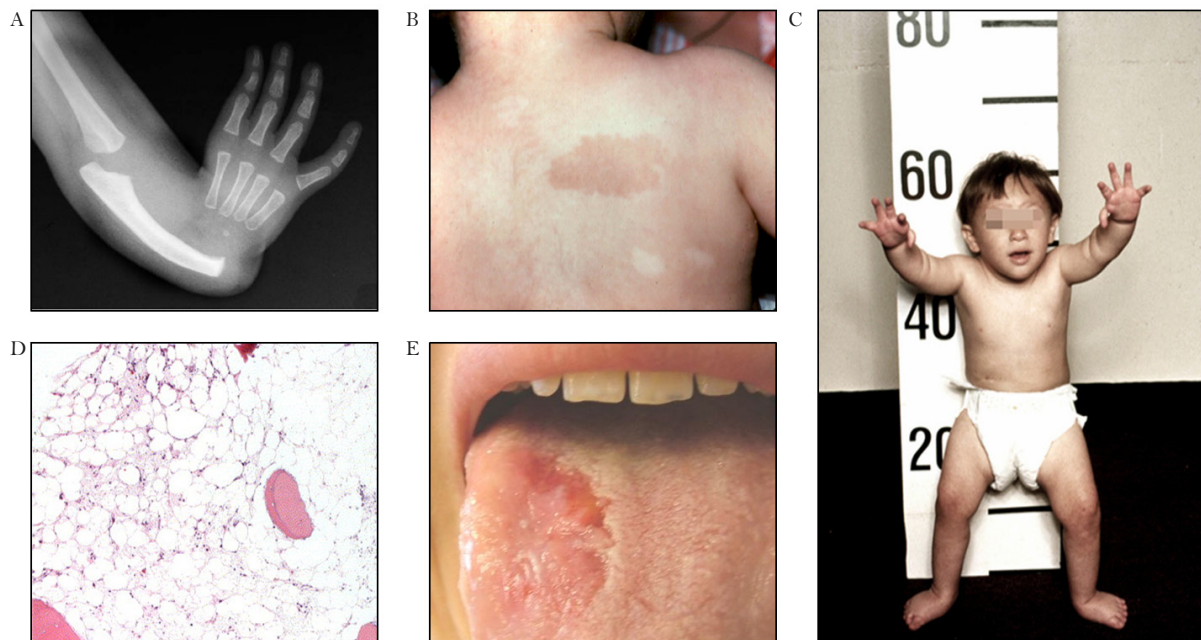
Fig. 2 Clinical features of Fanconi anemia (FA). (A) X-ray of a child with complete absence of the left radius and thumb. *Kindly provided by Shriners Hospital for Children, Philadelphia, USA. Previously published in: Fanconi Anemia. Guidelines for Diagnosis and Management. Third Edition, 2008.* (B) Café-au-lait spots and hypopigmented regions on a child´s shoulder. (C) Short stature (scaling in cm; below 3rd percentile) of a three year old boy. *B and C are reprinted from Alter and Young (1993) with permission of Elsevier.* (D) Hypocellular bonemarrow of a patient presented with pancytopenia. *Reprinted from Konoplev and Bueso-Ramos (2006) with permission of Elsevier.* (E) Squamous cell carcinoma of the tongue. *Kindly provided by E. Velleuer (Duesseldorf, Germany).*

FA patients additionally suffer from a high risk of developing myelodysplastic syndrome (MDS) and acute myeloid leukemia (AML), which could be associated with typical clonal cytogenetic abnormalities in bone marrow cells. Specific deletions and rearrangements involve especially chromosomes 1q, 3q and 7 (Alter and Kupfer 1993; Tischkowitz and Hodgson 2003; Neitzel, Kühl et al. 2007; Meyer, Neitzel et al. 2012). Next to the developing of AML or MDS, patients who survive into early adulthood have, compared to the general population, an approximately 50-fold higher risk to suffer from solid tumors. For squamous cell carcinoma (SCC) of the head and neck

(Fig.2E) and anogenital regions the cumulative risk for FA patients is even up to 700-fold increased. Liver tumors in form of adenomas and hepatocelluar carcinomas occur frequently in FA patients, mainly in the context of androgen therapy for aplastic anemia. More seldom, other solid tumors e.g. of the brain or kidneys are reported and often seem to be associated with distinct FA subtypes (FA-D1 and -N) which also predispose to leukemia in infancy. (Rosenberg, Greene et al. 2003; Neveling, Kalb et al. 2007; Auerbach 2009).

1.1.2   Cellular Phenotype and Diagnostics

Since the 1960s FA has been known to be a chromosomal instability disorder, a finding which could explain the high cancer incidence observed in FA patients. FA cells show chromatid aberrations (Fig.3A) including gaps, breaks and complex tri- or quadriradial figures, which can be observed spontaneously and especially after exposure of cultured cells to DNA interstrand crosslinking (ICL) agents like mitomycin C (MMC), diepoxybutane (DEB) or cisplatin (Schroeder, Anschutz et al. 1964; Auerbach 2009; Deans and West 2011). This specific hypersensitivity clearly distinguishes FA from other chromosomal breakage syndromes like Bloom syndrome and ataxia telangiectasia, but also sets it apart from Diamond-Blackfan anemia or VATER/VACTERL association (Tischkowitz and Hodgson 2003). The accumulation of unrepaired ICLs in proliferating FA cells induces replication fork stalling, which in turn activates the S/G2 cell cycle check point control and finally leads to cell cycle arrest in the G2 phase (Fig.3B) and increased rates of apoptosis. As this cellular feature is also dosage-dependent

A

B
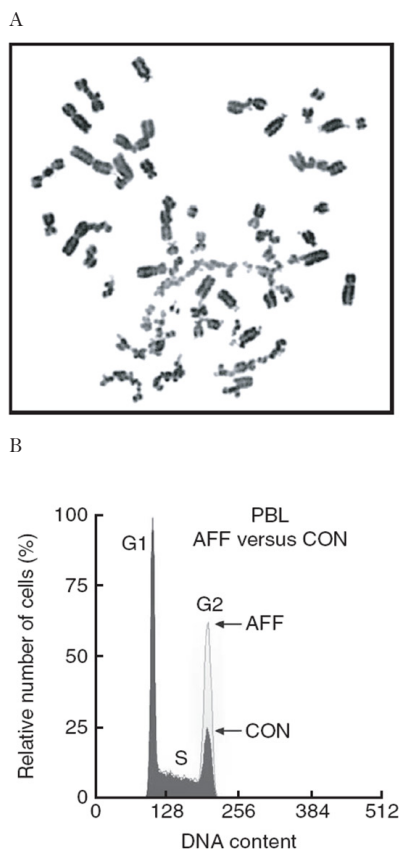
Fig.3 Cellular phenotype of Fanconi anemia (FA). (A) Chromosomal breakage after exposure to mitomycin C (MMC) in patient derived peripheral blood lymphocytes (left panel) as well as fibroblast (right panel). *Reprinted from Reid et al. (2007) with permission of Macmillan Publishers Ltd.* (B) Cell cycle distribution of patient (AFF) derived peripheral blood lymphocytes (PBL) after exposure to MMC shows typical G2 phase arrest compared to a control (CON). *Reprinted from Vaz et al. (2010), publication included.*

on MMC, it can likewise be used to discriminate between FA and non-FA cells via flow cytometry (Schindler, Friedl et al. 2007). Cells of some subtypes (FA-M, -D1, -N, -O and -P) have additionally been shown to be hypersensitive to camptothecin, a topoisomerase I inhibitor, what is probably associated with defects in homologous recombination (HR) repair (Stoepker, Hain et

al. 2011). Next to artificial chemotherapeutical DNA crosslinking drugs, there are environmental and endogenous sources of DNA damage, which are toxic for FA cells. For example, there is strong evidence that FA cells cannot tolerate oxidative stress. Early studies in the 1980s showed hypersensitivity of FA lymphocytes and fibroblasts towards ambient oxygen. Later on this was also described for hematopoietic stem cells carrying mutations in an FA gene. Similar to MMC or DEB, reactive oxygen species increase DNA damage and lead to cell cycle arrest in the G2 phase, premature senescence and apoptosis (Joenje, Arwert et al. 1981; Schindler and Hoehn 1988; Du, Adam et al. 2008). Moreover, there are reactive aldehydes, which are common environmental mutagens as well as products of our own metabolism. Aldehydes and especially acetaldehyde can induce different kinds of lesions, including ICLs, and have been shown to induce an FA-like phenotype in mice (Garaycoechea, Crossan et al. 2012). Many more substances like natural psoralens or oestrogens might be additional sources of ICLs and the driving force behind the evolution of the FA DNA damage response network, which shall be discussed below (Deans and West 2011).

1.2   GENETIC AND MOLECULAR BACKGROUND OF FA

1.2.1   THE FA GENES

The genetic background of FA is, similar to the clinical phenotype, very heterogeneous. Today we know 15 genes, designated as *FANCA, -B, -C, -D1, -D2, -E, -F, -G, -I, -J, -L, -M, -N, -O* and *-P*, any of which can be mutated in FA or FA-like patients. With the exception of the X-chromosomal recessive subtype FA-B, FA is inherited autosomal recessive. The underlying genes *FANCA* to *FANCP* do not show any accumulation in distinct genomic regions as it is summarized in Tab.1. Among FA patients about 60% carry disease causing mutations in *FANCA*, followed by *FANCG* and *FANCC* with more than 10%, respectively. Patients with mutations in other FA genes are less common and there are even some patients who cannot be assigned to any of the known complementation groups (Garner and Smogorzewska 2011; Kitao and Takata 2011). While biallelic mutations always lead to the manifestion of FA, monoallelic mutations in *FANCD1* (*BRCA2*), *FANCN* (*PALP2*), *FANCJ* (*BRIP1*), and *FANCO* (*RAD51C*) increase the risk for breast and ovarian cancer (Wong, Nordfors et al. 2011). FA genes with known orthologs in lower organisms down to yeast are *FANCD1, FANCD2, FANCJ, FANCL, FANCM* and *FANCP*, leading to the conclusion that the pathway in which they are involved has some evolutionary fundamental functions. However, the remaining FA genes have less well characterized functional domains and seem to be conserved only among vertebrates (Patel and Joenje 2007; McDaniel and Schultz 2008; Svendsen, Smogorzewska et al. 2009).

Tab.1 Summary of all known genes underlying Fanconi anemia. Besides gene symbols, other aliases are included as well as the chromosomal locations. The approximate proportion of patients carrying mutations in each particular gene are given in percent slightly altered, yet following Alter and Kupfer (1993, updated 2011).

Gene symbol	Other aliases	Chromosomal localization	Approximate portion of affected patients
FANCA		16q24.3	60-70%
FANCB	FAAP95	Xp22.3	2%
FANCC		9q22.3	14%
FANCD1	BRCA2	13q12.3	3%
FANCD2		3p25.3	3%
FANCE		6p22-p21	3%
FANCF		11p15	2%
FANCG	XRCC9	9p13	10%
FANCI	KIAA1794	15q25-q26	1%
FANCJ	BRIP1, BACH1	17q22	2%
FANCL	FAAP43	2p16.1	0.2%
FANCM	FAAP250	14q21.3	0.1%
FANCN	PALB2	16p12	0.7%
FANCO	RAD51C	17q22	0.1%
FANCP	SLX4, BTBD12	16p13.3	0.2%

1.2.2 THE FA PROTEINS IN DNA REPAIR

FA is a disorder defective in ICL-specific DNA damage response. The FA pathway of ICL repair is involved in a complex network consisting of nucleotide excision repair (NER), translesion synthesis (TLS) and HR (Deans and West 2011). In G0/G1 phase cells ICLs can be repaired by an error-prone mechanism. It combines NER, which promotes excision and unhooking of the crosslink with subsequent DNA repair synthesis by translesion DNA polymerases. In the S/G2 phase of cell cycle ICL repair is replication-coupled and depends on FA proteins and the HR machinery (Hlavin, Smeaton et al. 2010; Constantinou 2012). The FA pathway (Fig.4) is activated when two replication forks converge upon each site of an ICL. Forming a DNA remodelling complex together with the histone-fold heterodimer MHF1/2 and the associated protein FAAP24, the DNA translocase FANCM is not only able to recognize stalled replication forks and to facilitate DNA damage signaling mediated by ATR, but through direct interaction with FANCF it additionally recruits other members of the so called FA core complex (Kim, Kee et al. 2008; Ciccia and Elledge 2010). The entire functional core complex consists of at least eight FA proteins, including FANCA, -B, -C, -E, -F, -G, -L and -M, and the associated proteins FAAP100, FAAP 24, FAAP20, MHF1 and MHF2 (Leung, Wang et al. 2012). Its main function is monoubiquitination of FANCD2 and FANCI (ID complex) by the E3 ubiquitin ligase FANCL.
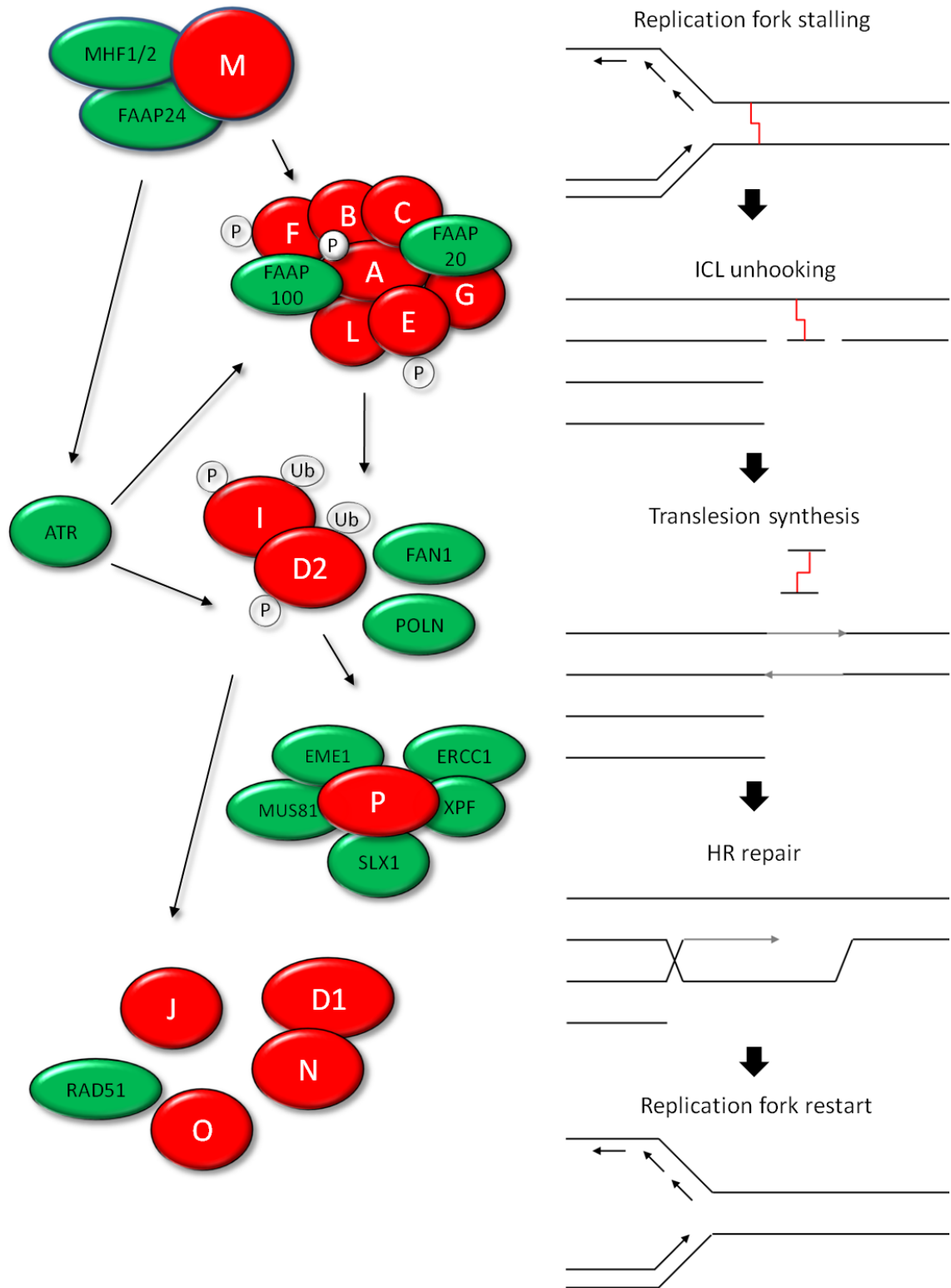
Fig. 4 The Fanconi anemia (FA) pathway of interstrand crosslink (ICL) repair. FANCM and its interacting proteins FAAP24, MHF1 and MHF2 are able to recognize stalled replication forks (displayed single-sided for the purpose of graphical simplification) and promote DNA damage signaling via ATR. Phosphorylation (P) of different FA proteins (red) by ATR is necessary for activation and progression of ICL repair. Following ICL recognition, FANCM recruits additional members (FANCA, -B, -C, -E, -F, -G, -L, FAAP20 and FAAP100) of the FA core complex, which in turn monoubiquitinates (Ub) the ID complex (FANCD2 and -I) via the E3-ubiquitin ligase FANCL. Monoubiquitination of FANCD2 recruits different downstream acting FA proteins as well as interacting non-FA proteins (green). FAN1 and the FANCP associated nucleases EME1/MUS81, XPF/ERCC1 and SLX1 are potential candidates for ICL unhooking and resolving intermediates of DNA repair by homologous recombination (HR). POLN might be responsible for translesion synthesis. FANCD1, -N and O interact directly with RAD51 in HR, while the precise function FANCJ helicase is unclear.

Although most of the FA core complex members have so far no characterized functional domains, they are without exception critical for complex integrity and ubiquitin ligase activity (Wang 2007; Yan, Delannoy et al. 2010; Kitao and Takata 2011). Phosphorylation of FANCI by ATR is also indispensable for FANCD2 monoubiquitination (Ishiai, Kitao et al. 2008). FANCD2 itself and some members of the core complex (FANCA, FANCG, and FANCF) are also phosphorylated in the course of DNA damage signaling, either by ATR or CHK1 (Enders 2008; Constantinou 2012). Monoubiquitination stabilizes the localization of the ID complex at sites of DNA damage and is necessary for recruitment of downstream effector proteins like the FANCD2-interacting nuclease FAN1 (Garner and Smogorzewska 2011). This protein and the FANCP-interacting nucleases SLX1, XPF/ERCC1 and MUS81/EME1 have hypothetical or proven functions in ICL unhooking and resolution of recombination intermediates (Deans and West 2011; Kitao and Takata 2011). Other proteins that are recruited by monoubiquitinated FANCD2 are the polymerases POLN and REV1, which are necessary for TLS after ICL unhooking, as well as the HR-promoting FA proteins FANCJ, FANCD1, FANCN and FANCO. While FANCD1 and FANCO are cofactors of the recombinase RAD51, FANCN is responsible for stabilization and localization of FANCD1 (Ciccia and Elledge 2010; Kitao and Takata 2011). FANCJ interacts with BRCA1, another breast cancer gene involved in DNA double strand break (DSB) repair by HR, but this interaction is probably not required for ICL repair. Though the precise function of FANCJ remains to be fully characterized, it seems likely to depend on BLM, a protein mutated in Bloom syndrome and known to interact with FANCM and the FA core complex. However, an intact helicase domain of FANCJ seems to be crucial for the FA pathway (Hiom 2010; Suhasini and Brosh 2012). Once the repair process is completed, FANCD2 is deubiquitinated by USP1, a process which is probably necessary for replication fork progression and further DNA repair processes. USP1, together with UAF1, additionally might be a positive regulator of HR repair by suppression of the competing, but error-prone DSB or ICL repair process of non-homologous end-joining (Murai, Yang et al. 2011).

1.3   GENOTYPING FA

1.3.1   CLASSICAL GENETIC APPROACHES – A HISTORICAL OVERVIEW

Genotyping of FA cells has always been and still is challenging, because of the immense heterogeneity. In 1992, the first FA gene to be identified was *FANCC*. Somatic cell hybridization was used in order to classify FA patients into different complementation groups. Thereafter, cells isolated from a subtype C patient were transfected with a cDNA library and selected in the presence

of MMC. Discovery of *FANCC* was possible because only cells that received the correspon-
ding gene construct were no longer hypersensitive for MMC and survived the selection process
(Strathdee, Duncan et al. 1992; Wang 2007). This technique also led to the identification of the
most prevalent FA gene *FANCA* in 1996, and few years later of three more core complex mem-
bers designated as *FANCE, -F* and *-G* (Lo Ten Foe, Rooimans et al. 1996; de Winter, Waisfisz et
al. 1998; de Winter, Leveille et al. 2000; de Winter, Rooimans et al. 2000). *FANCA* was addition-
ally and simultaneously published as FA-causing gene by *The Fanconi anemia / Breast cancer con-
sortium.* This study used linkage analysis followed by positional cloning (Apostolou S, Whitmore
S A et al. 1996). Since 2000 nearly each year a novel FA gene has been identified. The *FANCD2*
locus was narrowed down by microcell-mediated chromosome transfer into a FA cell line and
the underlying gene was finally confirmed by positional cloning and sequencing (Timmers, Ta-
niguchi et al. 2001). The identification of *FANCD2* was a major breakthrough that established
a functional link of the previously described FA core complex proteins to FANCD2 in a linear
pathway (Wang 2007). After findings of Garcia-Higuera et al. (2001) connecting *FANCD2* to
BRCA1 dependent DNA repair, *FANCD1, FANCJ* and *FANCN* were identified in candidate gene
approaches confirming the hypothesis of an FA/BRCA DNA repair network (Garcia-Higuera,
Taniguchi et al. 2001; Howlett, Taniguchi et al. 2002; Litman, Peng et al. 2005; Reid, Schindler
et al. 2007; Xia, Dorsman et al. 2007). *FANCJ* was found in two more independent studies by po-
sitional cloning and as a new approach by homozygosity mapping (Levitus, Waisfisz et al. 2005;
Levran, Attwooll et al. 2005). The importance of consanguineous families for the identification
of disease causing genes was again demonstrated by the discovery of *FANCO* (Vaz, Hanenberg
et al. 2010). In 2003, Meetei et al. isolated the so called BRAFT complex, consisting of the
hitherto known upstream FA proteins FANCA, -C, -E, -F, -G, the BLM helicase and other inter-
acting proteins (Meetei, Sechi et al. 2003). Since that time co-immunoprecipitation has become
a common tool for the detection of additional members of the FA core complex and has led to
identification of the ubiquitin ligase *FANCL*, the only X-chromosomal FA gene, *FANCB*, and
the FA gene with the highest evolutionary conservation, *FANCM* (Meetei, de Winter et al. 2003;
Meetei, Levitus et al. 2004; Meetei, Medhurst et al. 2005). Additionally, this approach discovered
at least five core complex associated genes, which so far have not been found to be mutated in FA
patients. Those genes include *FAAP20, FAAP24, FAAP100, MHF1* and *MHF2* (Ling, Ishiai et al.
2007; Yan, Delannoy et al. 2010; Leung, Wang et al. 2012; Yan, Guo et al. 2012).

1.3.2   NEXT GENERATION SEQUENCING – A NOVEL APPROACH

More than 30 years after Sanger et al. (1977) described DNA sequencing by chain-termination,
the first Next Generation Sequencing (NGS) platform was commercially launched in 2005 (Ma-

jewski, Schwartzentruber et al. 2011). The impact on molecular genetics was immense and is still ongoing. Today there are three leading NGS platforms from Roche Applied Science®, Applied Biosystems® and Illumina®. Depending on the provider, NGS performs massively parallel sequencing either by repeated cycles of polymerase-mediated nucleotide extension, or by oligonucleotide ligation. In contrast to conventional Sanger technique, several gigabases of data are produced in one run, but the average read length is only approximately 100-500 base pairs. These short reads have to be aligned to a reference sequence and can then be filtered for sequence variations (Shendure, Porreca et al. 2008; Pareek, Smoczynski et al. 2011). The subsequent development and usage of sequence capture methods in combination with high-throughput sequencing has made it possible to cost-effectively determine variants in a specific region of interest (Bamshad, Ng et al. 2011). Since 2010, Whole Exome Sequencing (WES) revealed the genetic basis of Miller and Kapuki syndrome, it has become the outstanding tool for the discovery of Mendelian disease genes (Ng, Bigham et al. 2010; Ng, Buckingham et al. 2010). For this technology all exonic regions are enriched out of a mechanically fragmented DNA sample, clonally amplified and subjected to NGS.

Not only for research but as well for diagnostic purposes, WES is getting more and more important. Especially for genetically very heterogenous disorders, like Charcot-Marie-Tooth disease or retinitis pigmentosa, genotyping via WES has already been reported to be successful and is an improvement to conventional techniques (Bamshad, Ng et al. 2011; Majewski, Schwartzentruber et al. 2011; Montenegro, Powell et al. 2011; Schrader, Heravi-Moussavi et al. 2011). The lack of distinct phenotype-genotype correlations, the high number of underlying genes and last but not least the fact that there are patients who cannot be assigned to any of the known complementation groups, made WES a promising tool also for genotyping FA patients. In 2012, after several successful applications in FA diagnostics, WES revealed pathogenic *XPF* mutations in two previously unassigned FA patients and thereby probably added the most recent member of the FA gene family, which could be designated *FANCQ* (Bogliolo, Schuster et al. 2012, submitted).

1.4   AIM OF THE STUDY

The present work had two major aims. One ambition was the identification and characterization of novel FA genes. Necessary for that end was the pre-classification of FA patients by exclusion of all known complementations groups. They also had to be assigned to two different groups depending on the functional defect lying up- or downstream of FANCD2 monoubiquitination. On the basis of this rough classification, patient derived cell lines should be screened for mutations in candidate genes on protein and DNA level. The candidate genes were mainly selected by literature research or based on preceding experiments, which unveiled a functional connection or similarity to previously reported FA genes. Next to the classical candidate gene approaches also linkage analysis and homozygosity mapping should be integrated in this study.

A second aim of the present work was the establishment of NGS in FA research and diagnostics. By testing different strategies like target enrichment and WES, and by developing an adequate bioinformatic analysis pipeline, the new technology should be integrated in the search for unknown genes and in genotyping of FA cells, in general.

2.   MATERIALS AND METHODS

Applied materials and methods are described explicitly in the enclosed publications and their supplementary information.

- Concerning the FA candidate genes *MHF1*, *MHF2* (*3.1.1 A Histone-Fold Complex and FANCM Form a Conserved DNA-Remodeling Complex to Maintain Genome Stability*) and *FAN1* (*3.1.2 On the Role of FAN1 in Fanconi Anemia*), screenings for protein deficiency or truncation were performed on whole protein extracts (page 39) with standard Western Blot procedures using antibodies listed on pages 37 and 60.

- Materials and methods used for the *RAD51C* study (*3.2.1 Mutation of the RAD51C Gene in a Fanconi Anemia-like disorder*) are explicitly described on pages 70-71 and 77. Amongst others, they include details on primers and PCR conditions used for mutation screening and information on immunoblotting experiments used for the exclusion of known complementation groups.

- Details on experimental procedures of the *SLX4* study (*3.2.2 SLX4, a Coordinator of Structure-Specific Endonucleases, is Mutated in a New Fanconi Anemia Subtype*) are given on pages 83-84 and 86-87. Included are primers for sequencing *SLX4* on genomic and transcript level, as well as information on the linkage analysis.

- An explicit summary of applied materials and methods for cytogenetic and molecular characterization of patient derived cells in the *XPF* study (*3.2.3 XPF Mutations Severely Disrupting DNA Interstrand Crosslink Repair Cause Fanconi Anemia*) is given on pages 126-133. Additional details on primer sequences are given on page 143-144.

- Details on WES analysis and validating experiments using classical genetic approaches for the FA-P case report (*3.3.1 Whole Exome Sequencing Reveals Novel Mutations in the Recently Identified Fanconi Anemia Gene SLX4/FANCP*) are included in the consecutive text (pages 151-152). Additional information on applied *in silico* tools is given on pages 152-153. Experimental procedures for cell cycle analysis are described in 3.2.1, whereas details for immoprecipitation, cell fractionation and immunfluorescence assays are decribed in 3.2.2.

- The study on WES in the molecular diagnosis of FA (*3.3.2 Genotyping Fanconi Anemia by Whole Exome Sequencing: Advantages and Challenges*) contains a section on applied materials and methods, including sequencing techniques, analysis tools and details on validating experiments, on pages 168-171.

3. RESULTS

3.1 FA CANDIDATE GENES

3.1.1 A HISTONE-FOLD COMPLEX AND FANCM FORM A CONSERVED DNA-REMODELING COMPLEX TO MAINTAIN GENOME STABILITY

3.1.2 ON THE ROLE OF FAN1 IN FANCONI ANEMIA

3.2 IDENTIFICATION OF NOVEL FA GENES

3.2.1 MUTATION OF THE RAD51C GENE IN A FANCONI ANEMIA-LIKE DISORDER

3.2.2 SLX4, A COORDINATOR OF STRUCTURE-SPECIFIC ENDONUCLEASES, IS MUTATED IN A NEW FANCONI ANEMIA SUBTYPE

3.2.3 XPF MUTATIONS SEVERELY DISRUPTING DNA INTERSTRAND CROSSLINK REPAIR CAUSE FANCONI ANEMIA

3.3 GENOTYPING FANCONI ANEMIA BY NEXT GENERATION SEQUENCING

3.3.1 WHOLE EXOME SEQUENCING REVEALS NOVEL MUTATIONS IN THE RECENTLY IDENTIFIED FANCONI ANEMIA GENE SLX4/FANCP

3.3.2 GENOTYPING FANCONI ANEMIA BY WHOLE EXOME SEQUENCING: ADVANTAGES AND CHALLENGES

Article

A Histone-Fold Complex and FANCM Form a Conserved DNA-Remodeling Complex to Maintain Genome Stability

Zhijiang Yan,[1] Mathieu Delannoy,[4,13] Chen Ling,[1,13] Danielle Daee,[6,13] Fekret Osman,[7,13] Parameswary A. Muniandy,[2] Xi Shen,[8] Anneke B. Oostra,[9] Hansen Du,[3] Jurgen Steltenpool,[9] Ti Lin,[1] Beatrice Schuster,[10] Chantal Décaillet,[4] Andrzej Stasiak,[5] Alicja Z. Stasiak,[5] Stacie Stone,[11] Maureen E. Hoatlin,[11] Detlev Schindler,[10] Christopher L. Woodcock,[12] Hans Joenje,[9] Ranjan Sen,[3] Johan P. de Winter,[9] Lei Li,[8] Michael M. Seidman,[2] Matthew C. Whitby,[7] Kyungjae Myung,[6] Angelos Constantinou,[4,*] and Weidong Wang[1,*]

[1]Laboratory of Genetics
[2]Laboratory of Molecular Gerontology
[3]Laboratory of Cellular and Molecular Biology
National Institute of Aging, National Institutes of Health, Baltimore, MD 21224, USA
[4]Department of Biochemistry
[5]Center for Integrative Genomics
University of Lausanne, Lausanne, Switzerland
[6]Genome Instability Section, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD 20892, USA
[7]Department of Biochemistry, University of Oxford, Oxford OX 3QU, UK
[8]Department of Experimental Radiation Oncology, University of Texas M.D. Anderson Cancer Center, Houston, TX 77030, USA
[9]Department of Clinical Genetics, VU University Medical Center, 1007 MB Amsterdam, The Netherlands
[10]Department of Human Genetics, University of Wurzburg, 97070 Wurzburg, Germany
[11]Department of Biochemistry and Molecular Biology, Oregon Health and Science University, Portland, OR 97239, USA
[12]Department of Biology, University of Massachusetts, Amherst, MA 01003, USA
[13]These authors contributed equally to this work
*Correspondence: angelos.constantinou@unil.ch (A.C.), wangw@grc.nia.nih.gov (W.W.)
DOI 10.1016/j.molcel.2010.01.039

SUMMARY

FANCM remodels branched DNA structures and plays essential roles in the cellular response to DNA replication stress. Here, we show that FANCM forms a conserved DNA-remodeling complex with a histone-fold heterodimer, MHF. We find that MHF stimulates DNA binding and replication fork remodeling by FANCM. In the cell, FANCM and MHF are rapidly recruited to forks stalled by DNA interstrand crosslinks, and both are required for cellular resistance to such lesions. In vertebrates, FANCM-MHF associates with the Fanconi anemia (FA) core complex, promotes FANCD2 monoubiquitination in response to DNA damage, and suppresses sister-chromatid exchanges. Yeast orthologs of these proteins function together to resist MMS-induced DNA damage and promote gene conversion at blocked replication forks. Thus, FANCM-MHF is an essential DNA-remodeling complex that protects replication forks from yeast to human.

INTRODUCTION

Fanconi anemia (FA) is a genetically heterogeneous syndrome characterized by genomic instability, congenital abnormalities, bone marrow failure, and cancer predisposition. FA can be caused by mutation in any of 13 different FANC genes. The hallmark of FA cells is their hypersensitivity to drugs that induce DNA interstrand crosslinks (ICLs), which prevent separation of the complementary strands of DNA and completely block replication and other essential processes. Increasing evidence suggests that FANC proteins act both as signal transducers and as DNA-processing molecules in a DNA damage response network, the FA-BRCA network, which is essential to repair or bypass ICLs during DNA replication (Thompson and Hinz, 2009; Wang, 2007).

The FANC proteins can be divided into three groups on the basis of their participation at different stages of the FA-BRCA network (Wang, 2007). Group I acts at the early stage of the pathway and consists of eight FANC proteins, FANCA, FANCB, FANCC, FANCE, FANCF, FANCG, FANCL, and FANCM. These proteins, together with FAAP100 and FAAP24, constitute the FA core complex. The FA core complex has two main functions. First, it monoubiquitinates group II FA proteins, FANCD2 and FANCI, in response to DNA damage or during the S phase of the cell cycle; second, it directly participates in DNA repair through a DNA-remodeling enzyme, FANCM (Xue et al., 2008). Group II FANC proteins include FANCD2 and FANCI, which form the FANCI-FANCD2 complex. Once ubiquitinated, this complex associates with chromatin and colocalizes with the Group III FANC proteins, FANCD1/BRCA2, FANCN/PALB2, and FANCJ/BRIP1. The group III proteins all interact with BRCA1 and are thought to mediate HR-dependent DNA repair.

FANC proteins and FAAPs usually have orthologs in vertebrates but not in yeast. A single exception is FANCM, which not only has an ortholog in yeast, but also in archea (Wang, 2007). This implies that the DNA-remodeling function of FANCM
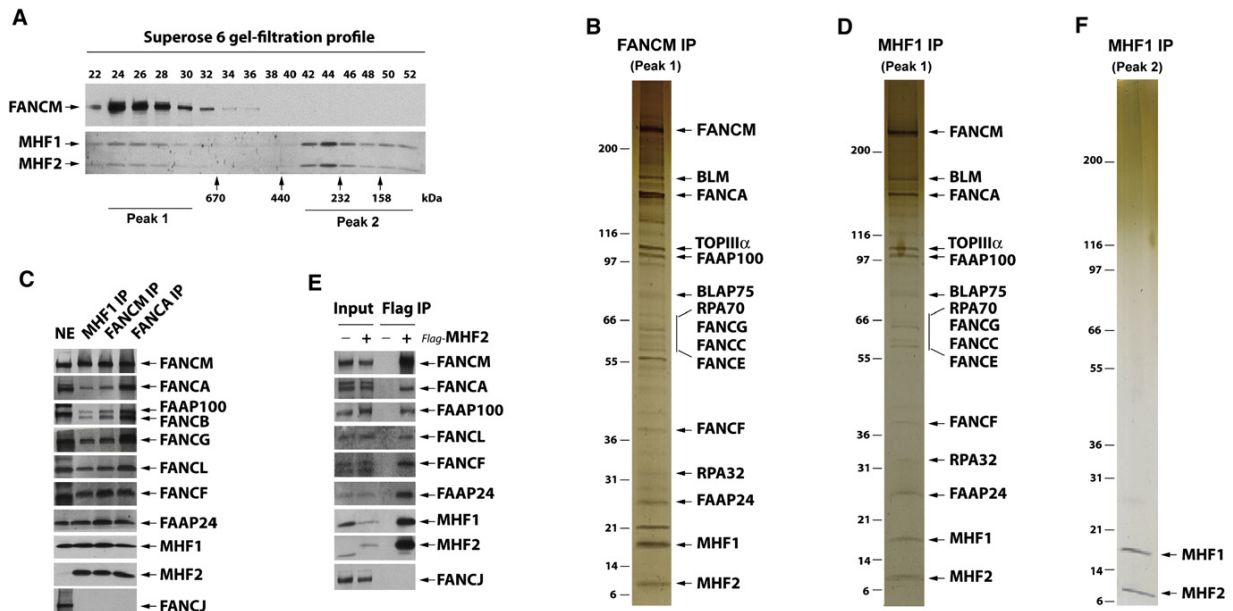
Figure 1. Identification of MHF1 and MHF2 as Integral Components of the FA Core Complex

(A) Immunoblotting showing Superose 6 gel-filtration profiles of FANCM, MHF1, and MHF2 in HeLa nuclear extract. The peak fractions and those corresponding to marker proteins are indicated at the bottom.

(B) A silver-stained gel showing that the complex purified by a FANCM antibody from the pooled peak 1 fractions in (A) contained MHF1, MHF2, and other components of FA core and BLM complexes. IP indicates immunoprecipitation.

(C) Immunoblotting shows that MHF1, MHF2, and other FA core complex components were present in the immunoprecipitate isolated from peak 1 fractions by MHF1, FANCM, or FANCA antibodies. Nuclear extract (NE) was used as a loading control.

(D) As described in (B), except a MHF1 antibody was used in IP.

(E) Immunoblotting shows that MHF2 coimmunoprecipitated with MHF1 and other FA core complex components from HeLa cells stably expressing Flag-tagged MHF2, but not from control HeLa cells. A Flag antibody was used in IP.

(F) As described in (B), except that peak 2 fractions in (A) were used for IP by a MHF1 antibody (see also Figures S1 and S2.)

is important enough to be conserved through evolution. The yeast orthologs of FANCM, Mph1 in *Saccharomyces cerevisiae* and Fml1 in *Schizosaccharomyces pombe*, possess a helicase motif and are capable of dissociating D-loops (an HR intermediate) and suppressing chromosome crossover recombination induced by double-strand breaks (Banerjee et al., 2008; Prakash et al., 2009; Sun et al., 2008). Fml1 can also stimulate replication fork reversal in vitro, and promote gene conversion at stalled replication forks in vivo (Sun et al., 2008).

Human FANCM contains an ATP-dependent branch-point translocase activity, which promotes migration of Holliday junctions, replication fork reversal, and dissociation of D-loops (Gari et al., 2008a; Gari et al., 2008b; Xue et al., 2008). The ATP-dependent activity of FANCM is required for cellular resistance to DNA crosslinking agents but is dispensable for the monoubiquitination function of the FA core complex (Rosado et al., 2009; Singh et al., 2009; Xue et al., 2008). FANCM binds specifically and with high affinity to Holliday junctions and replication forks (Gari et al., 2008b; Xue et al., 2008), and this DNA-binding activity seems to be required for efficient monoubiquitination of FANCD2 (Xue et al., 2008). FANCM has a DNA-binding partner, FAAP24, which can target FANCM to single-strand DNA, an intermediate of both DNA replication

and repair (Ciccia et al., 2007). The FANCM/FAAP24 dimer is needed to tolerate damage induced by UV and camptothecin and for suppression of crossover recombination (Rosado et al., 2009; Singh et al., 2009).

Here, we describe a histone-fold protein complex, named MHF, as a component of the FA core complex and an essential cofactor of FANCM. We present evidence that MHF and FANCM form a conserved DNA-remodeling complex that maintains genomic stability from yeast to human.

RESULTS

Identification of MHF1 and MHF2 as Integral Components of the FA Core Complex

To search for additional components of the FA core complex, we fractionated HeLa nuclear extract by gel-filtration chromatography, collected peak fractions containing FANCM (peak 1 in Figure 1A), and immunoprecipitated with a FANCM antibody. Silver-staining (Figure 1B), mass spectrometry analyses (data not shown), and immunoblotting (Figure 1C) revealed the presence of all known components of the FA core complex (FANC-A,-B, -C, -E, -F, -G -L, and -M, FAAP100, and FAAP24) in the FANCM immunoprecipitate. Most of the Bloom syndrome

Molecular Cell

A Conserved DNA-Remodeling Complex in DNA Repair

complex components (BLM, TOPIIIα, RMI1/BLAP75, RPA70, and RPA32) were also identified by mass spectrometry (data not shown), supporting our previous findings that the FA core and BLM complexes associate in a super-complex, BRAFT (Wang, 2007).

Silver-staining detected two polypeptides of about 16 and 10 kDa that were not previously identified (Figure 1B). Mass spectrometry revealed the 16 kDa protein as CENP-S or APITD1 (gene accession ID: NP_954988) (Foltz et al., 2006; Krona et al., 2004), and the 10 kDa polypeptide as a protein with accession ID of A8MT69 or CENP-X (Amano et al., 2009). A8MT69 has a confusing name of STRA13, which refers to two distinct proteins: one is A8MT69, and the other is an unrelated helix-loop-helix transcription factor. Of 56 published articles on STRA13, only one is on A8MT69, which shows that deletion of the A8MT69 ortholog in fission yeast may result in sensitivity to several genotoxins (Deshpande et al., 2009). To avoid the confusion, we renamed the 16 and 10 kDa polypeptides as MHF1 and MHF2, respectively (for FANCM [or Mph1]–associated Histone-Fold protein 1 or 2).

The following evidence indicates that MHF1 and MHF2 are components of the FA core complex. First, immunoblotting showed that antibody against MHF1 or MHF2 recognized the corresponding polypeptide in the immunoprecipitate obtained with either FANCM or FANCA antibody (Figure 1C). Second, one peak of MHF1 or MHF2 on Superose 6 column is coincident with that of FANCM, supporting the notion that they are present in the same FA core complex (Figure 1A, peak 1). Third, reciprocal immunoprecipitation using a MHF1 antibody from the pooled Superose fractions of peak 1 obtained the same set of polypeptides isolated by the FANCM antibody, including components of both FA core and BLM complexes, as evidenced by silver-staining (Figure 1D), mass spectrometry (data not shown), and immunoblotting (Figure 1C). Finally, MHF2-associated polypeptides isolated by a Flag antibody from the extract of HeLa cells stably expressing Flag-tagged MHF2 also contained MHF1, FANCM, and other FA core complex components (Figure 1E).

We noticed that MHF1 and MHF2 cofractionate in two peaks on a Superose 6 column (Figure 1A). Although peak 1 corresponds to the FA core complex of about 1 MDa, peak 2 corresponds to a complex of much smaller size. When peak 2 fractions were immunoprecipitated with a MHF1 antibody, only MHF1 and MHF2 were isolated (Figure 1F), as revealed by mass spectrometry (data not shown), indicating that these two proteins could comprise a complex distinct from the FA core complex. The two MHF proteins appear to be approximately equimolar amounts on the silver-stained gel, suggesting that they are likely obligate partners. Direct interaction between MHF1 and MHF2 was observed by mammalian two-hybrid analyses (see Figure S1 available with this article online).

MHF1 and MHF2 Are Conserved from Human to Yeast and Form a Histone-Fold Complex

Bioinformatics revealed that both MHF proteins contain a histone-fold, which can mediate both protein-protein and protein-DNA interactions (Figures S2A and S2B) (Arents and Moudrianakis, 1993). Proteins containing this motif often associate to form heterodimeric or heterotetrameric complexes that bind to bent DNA.

Our purification of MHF1 and MHF2 as a stoichiometric complex from HeLa extract (Figure 1F) fits well with the bioinformatic prediction that they form a histone-fold complex.

We coexpressed MHF1 with HIS-tagged MHF2 in Escherichia coli and purified the complex using Talon metal affinity chromatography. Coomassie-staining revealed that the two proteins were present in approximately equimolar amounts (Figure 2A, lane 3), indicating that they indeed form a heterodimeric complex. We named this complex MHF.

MHF Possesses DNA-Binding Properties Distinct from FANCM and FAAP24

Like other histone-fold complexes, MHF was found to bind double-strand DNA (dsDNA), but not single-strand DNA (ssDNA) (Figure 2B, lanes 1–10). This activity requires the MHF complex because individual MHF subunit lacked the activity.

MHF was also found to bind several structured DNAs containing different branch points (Figure 2B, lanes 11–30). Notably, the affinity of MHF to these structures was either similar or somewhat reduced compared to that of dsDNA of the same length (compare lanes 10, 15, 20, 25, and 30), indicating that MHF has no increased affinity for branched DNA, and its binding may even be precluded by such structures. The binding of MHF to dsDNA was further visualized by electron microscopy (Figures S3A and S3B). MHF appeared to form clusters, which resulted in compaction of DNA, suggesting self-association between MHF proteins.

The DNA-binding characteristics of MHF differ from those of FANCM and FAAP24, which specifically recognize branched and single-stranded DNA, respectively (Ciccia et al., 2007; Gari et al., 2008b; Xue et al., 2008). We propose that these proteins constitute a molecular machine that binds cooperatively to different parts of a stalled replication fork: FANCM binds the branch point, whereas MHF and FAAP24 associate with dsDNA and ssDNA regions, respectively (see Discussion). Moreover, we have also shown that MHF can bind chromatin (Figure S3C) and cooperate with histone octamers to assemble into nucleosome structures in vitro (Figures S3D and S3E), which is consistent with MHF aiding FANCM association with DNA in vivo. Furthermore, MHF can efficiently anneal complementary single-stranded DNAs (albeit not when they are prebound by RPA) (Figures S3F and S3G), which could assist the catalysis of branch point migration by FANCM.

MHF and FANCM Coevolved and Form an Independent Complex

Bioinformatics analyses revealed that MHF1 and MHF2 orthologs are present in all eukaryotes, including yeast (Figures S2C and S2D). This feature is shared by FANCM but not by other FANC proteins or FAAPs, most of which have orthologs only in vertebrates (Wang, 2007). The data suggest that FANCM, MHF1 and MHF2 may perform functions important to all eukaryotes that favor their coevolution.

The coevolution of FANCM and the two MHF proteins implies that they may constitute a complex that lacks other FANC proteins. To distinguish this complex from the FA core complex in HeLa cells, we omitted the Superose 6 fractionation step (Figure 1A) that was used for enrichment and purification of the FA
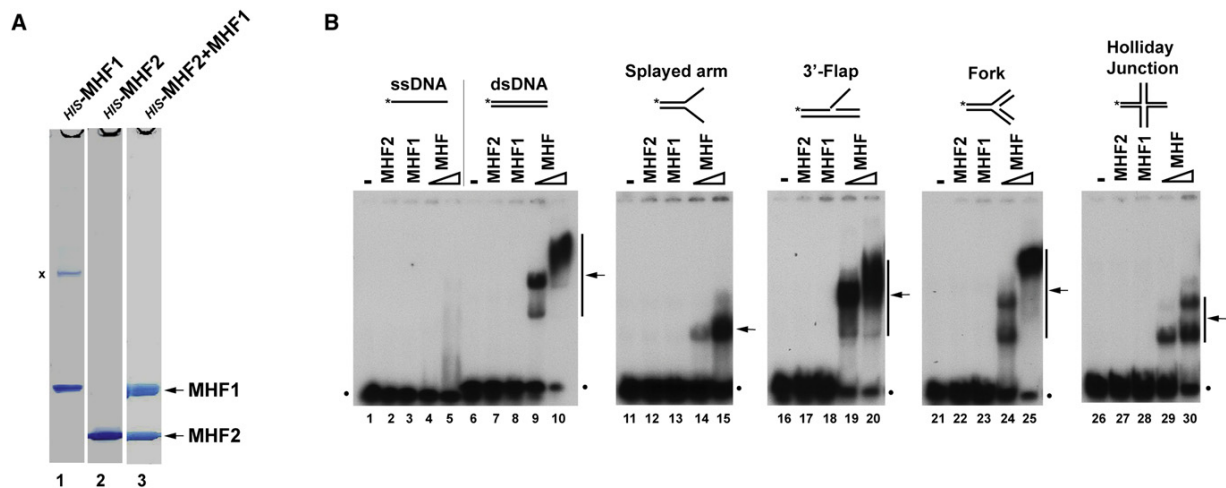
Figure 2. MHF Binds Double-Strand and Branch-Structured DNAs, but Not Single-Strand DNA

(A) Coomassie blue-stained SDS-PAGE gels showing *HIS*-tagged recombinant proteins MHF1, MHF2, and MHF complex purified from *E. coli*. A nonspecific band was marked with "x."

(B) Electrophoretic mobility shift assay (EMSA) for testing DNA-binding activity of MHF1, MHF2, and MHF complex. A variety of DNA substrates (0.5 nM) illustrated at the top were incubated with 0.6 μM MHF1, 0.6 μM MHF2, and increasing amounts (0.3 and 0.6 μM) of MHF complex, respectively. Asterisks denote ^{32}P label at the DNA 5′ end. The arrows indicate the shift bands of MHF-DNA complex. The dots represent free DNA probe (see also Figure S3).

core complex (Figure 1D). We performed immunoprecipitation directly from HeLa extract with the same MHF1 antibody and obtained FANCM, MHF1, and MHF2 as the only three major polypeptides (Figure 3A). Other FANC proteins can be detected only by immunoblotting because of their lower levels (data not shown). The data suggest that significant amounts of MHF and FANCM are present in a complex largely free of other FANC proteins. We named this complex FANCM-MHF.

Most MHF and FANCM Do Not Associate with the FA Core Complex

We quantitatively immunodepleted the FA core complex from HeLa extract with the FANCA antibody and found that less than 30% of FANCM, MHF, and FAAP24 were codepleted (Figure S4A). Similarly, depletion of FAAP24 or MHF from the extract codepleted FANCA by less than 25% (Figures S4B and S4C). These data suggest that the majority of FANCM and its two partners do not associate with the FA core complex. Conversely, immunodepletion of either FAAP24 or MHF codepleted FANCM by about 70% and 85% (Figures S4B and S4C), indicating that most of FANCM in cells associates with either or both of its partners.

MHF and FAAP24 Can Form Separate Complexes with FANCM

We immunoprecipitated MHF from the FAAP24-depleted extract and obtained FANCM but no FAAP24 (Figure S4B, lane 4), suggesting that MHF and FANCM can form a complex without FAAP24. Similarly, we immunoprecipitated FAAP24 from the MHF-depleted extract and obtained FANCM but no MHF (Figure S4C, lane 4), indicating FAAP24 and FANCM can also form a complex without MHF. Together with the results in

Figure 1 which showed that FANCM and both partners coimmunoprecipitate, our data suggest that FANCM can associate with its partners in a combinatorial manner to form distinct complexes: FANCM-MHF, FANCM-FAAP24, and FANCM-MHF-FAAP24.

MHF Interacts with FANCM and Promotes Its DNA-Binding Activity

We mapped the MHF-interaction domain within FANCM to a region near the helicase domain of FANCM (aa 661–800) by immunoprecipitation and Western analyses of a series of FANCM deletion mutants (Figure 3B and Figure S4D). We have previously shown that both the full-length FANCM protein and its N-terminal fragment (1–754 aa) encompassing the helicase domain (FANCM$_{754}$) have high affinity for branched DNA structures, but not for dsDNA (Gari et al., 2008b; Xue et al., 2008). The current findings that MHF interacts with both FANCM and dsDNA predict that MHF may recruit FANCM to dsDNA. Consistent with this prediction, FANCM$_{754}$ alone exhibited little dsDNA-binding activity (Figure 3C, lanes 2–4), whereas MHF and FANCM$_{754}$ together displayed strong binding activity (see the supershifted band in Figure 3C, lanes 6 to 8 versus 5).

FANCM and MHF Bind DNA Synergistically

The fact that FANCM and the two MHF proteins can form a complex raised a possibility that they may bind DNA cooperatively. Indeed, not only did MHF enhance the DNA binding of FANCM$_{754}$, FANCM$_{754}$ also stimulated the DNA-binding activity of MHF (Figure 3C; the unbound DNA was reduced in lane 6 compared to lane 5). Moreover, FANCM$_{754}$ and MHF bound DNA synergistically at low protein concentrations: although either protein alone showed little binding activity, both proteins

Molecular Cell

A Conserved DNA-Remodeling Complex in DNA Repair

together displayed an activity much higher than the sum of individuals (Figure 3C; less than 1% of dsDNA was shifted in lanes 10–13, whereas 50%–90% of DNA was shifted in lanes 14–16). This synergy was also observed for fork and Holliday junction (HJ) substrates (Figure 3C, lanes 17–32).

We reconstituted the FANCM-MHF complex by coexpressing full-length recombinant FANCM, MHF1, and MHF2 proteins in insect cells, and purified the trimeric complex (Figure 3D). Incubation of FANCM-MHF with synthetic replication forks led to formation of a defined protein-DNA complex whose mobility was reduced, compared with that of FANCM-fork complex, suggesting that FANCM, MHF1, and MHF2 bind together to branched DNA molecules (Figure 3E). Moreover, the FANCM-MHF complex had a stronger fork binding activity than FANCM alone (Figure 3E), providing further evidence that FANCM-MHF binds DNA cooperatively.

MHF Stimulates Replication Fork Reversal Activity of FANCM

FANCM exhibits an ATP-dependent replication fork reversal activity, which may stabilize stalled forks and facilitate assembly of DNA damage signaling and repair complexes (Gari et al., 2008a). We found that recombinant FANCM-MHF had a fork reversal activity stronger than that of FANCM (Figures 3F, 3G, and 3H). Both FANCM and FANCM-MHF catalyzed reversal of a model replication fork into a four-way junction intermediate and led to the formation of a labeled linear DNA duplex, the end product of complete fork reversal (Figures 3G and 3H). At low protein concentrations, however, FANCM-MHF produced higher levels of four-way junction intermediate (about 5–7-fold) and linear regression product (about 2–3-fold) than FANCM alone (Figures 3G and 3H). These data establish FANCM-MHF as a DNA-remodeling machine and suggest that MHF is a crucial cofactor for FANCM in both binding and ATP-dependent remodeling of DNA.

MHF Is Required for Stability of FANCM and for Activation of the FA Pathway

We next examined whether MHF is required for FANCM function in vivo. SiRNA depletion of either MHF1 or MHF2 in HeLa cells reduced the level of FANCM in whole cell lysates (Figure 4A) and the chromatin fractions (Figure 4B), suggesting that MHF is required for stability of FANCM and may also be important for chromatin association of FANCM. Depletion of one MHF protein also reduced the protein level of the other (Figure 4A), providing additional evidence that MHF1 and MHF2 are direct interacting partners and depend on each other for stability.

HeLa cells depleted of either MHF protein displayed reduced levels of monoubiquitinated FANCD2 and FANCI in response to DNA crosslinking drugs, mitomycin C (MMC), and cisplatin (Figures 4C and 4D). The cells also exhibited hypersensitivity to these drugs (Figures 4E and S5A) and increased chromosomal breaks in response to MMC (Figure S5B). All these phenotypes are characteristics for cells defective in FA core complex components, indicating that MHF is important for normal functions of the core complex and the FA pathway.

We noticed that HeLa cells depleted of FANCM or MHF were sensitive to methyl methanesulfonate (MMS), a DNA-alkylating agent (Figure 4F). This feature has been found in hamster cells mutated of FANCG (Tebbs et al., 2005) and in yeast FANCM (Banerjee et al., 2008; Sun et al., 2008) and MHF mutants (see below), suggesting that some DNA repair functions of FANCM-MHF are conserved in lower eukaryotes.

Cells depleted of MHF also exhibited sensitivity to camptothecin (CPT), a topoisomerase I inhibitor (Figure 4G). This feature was observed in cells depleted of FANCM but not in those lacking other components of the FA core complex (Rosado et al., 2009; Singh et al., 2009). The results suggest that FANCM-MHF may have functions independent of the FA core complex.

MHF and FANCM Are Rapidly Recruited to DNA Interstrand Crosslinks that Block Replication

Because cells depleted of FANCM or MHF are hypersensitive to drugs that induce interstrand crosslinks (ICLs), we examined whether FANCM-MHF localizes at ICLs using two independent techniques. First, we utilized laser-activated psoralen conjugates to generate ICLs within a localized region in nucleus and then to visualize proteins recruited to this region by indirect immunofluorescence (Thazhathveetil et al., 2007). Both FANCM and MHF1 were recruited to the ICLs within 15 min after photo-activation in a subgroup of cells in random culture (about 15%) (Figure 5A and data not shown), whereas they were not recruited to DNA monoadducts generated by laser-activated angelicin (Thazhathveetil et al., 2007). Costaining with a cell-cycle marker (NPAT) revealed that the recruitment occurred only in S phase cells (Figures S6A, S6B, and S6C and data not shown). When cells were synchronized in S phase, the recruitment was increased to approximately 50% of the cell population (Figure S6D). The fact that the recruitment of FANCM and MHF to ICLs occurs only during S phase provides in vivo evidence for FANCM-MHF to remodel replication forks blocked by cross-linked DNA. Notably, the recruitment of FANCM (but not γ-H2AX) was strongly diminished in cells depleted of either MHF1 or MHF2 (Figures 5B and 5C). This may be due to reduced stability (Figure 4A) and/or impaired targeting of FANCM in the absence of MHF.

We also used eChIP, a chromatin-IP–based method that detects proteins at a site-specific psoralen ICL on an episomal plasmid transfected into cells (Figure 5D) (Shen et al., 2009). This plasmid contains the replication origin of Epstein-Barr virus (EBV), so that the plasmid without the ICL can undergo replication unidirectionally in EBNA-293 cells that express the Epstein-Barr nuclear antigen 1 (EBNA). For the plasmid carrying the ICL (which was positioned 488 bp downstream of the replication origin), the replication fork is stalled by the crosslink, and proteins accumulated at the stalled fork can be detected by ChIP-PCR using a primer set that amplifies a DNA fragment near the crosslink. The same plasmid can be introduced into standard 293 cells that lack EBNA for detection of proteins recruited to ICL under nonreplicating conditions. We found that MHF1 was enriched about 5-fold at the ICL when the episomal vector was allowed to replicate in EBNA-293 cells (Figure 5E). However, the enrichment was reduced to less than 2-fold in the 293 cells that do not support vector replication. In comparison, FAAP24 was enriched at the ICL under both replicating
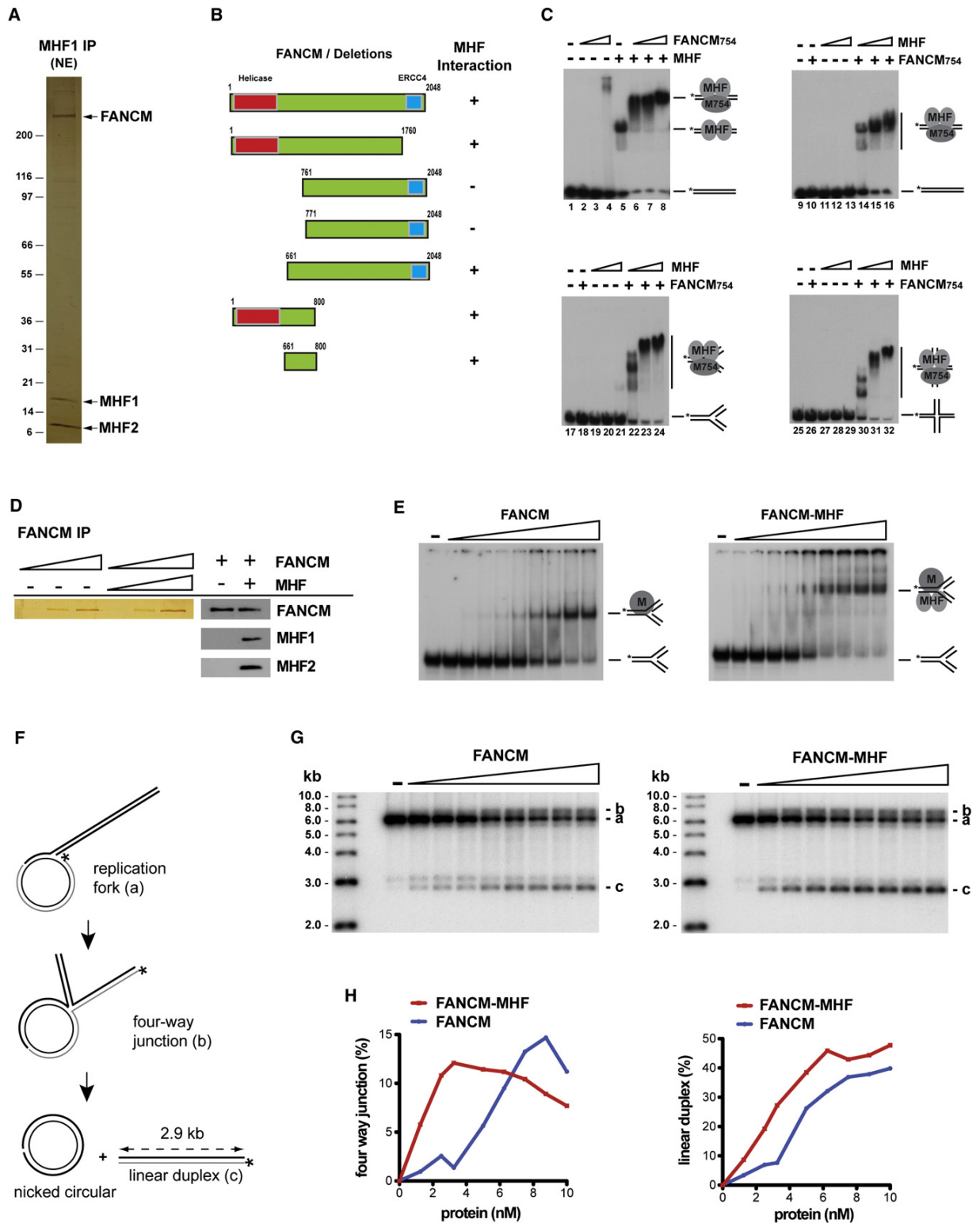
Molecular Cell

A Conserved DNA-Remodeling Complex in DNA Repair
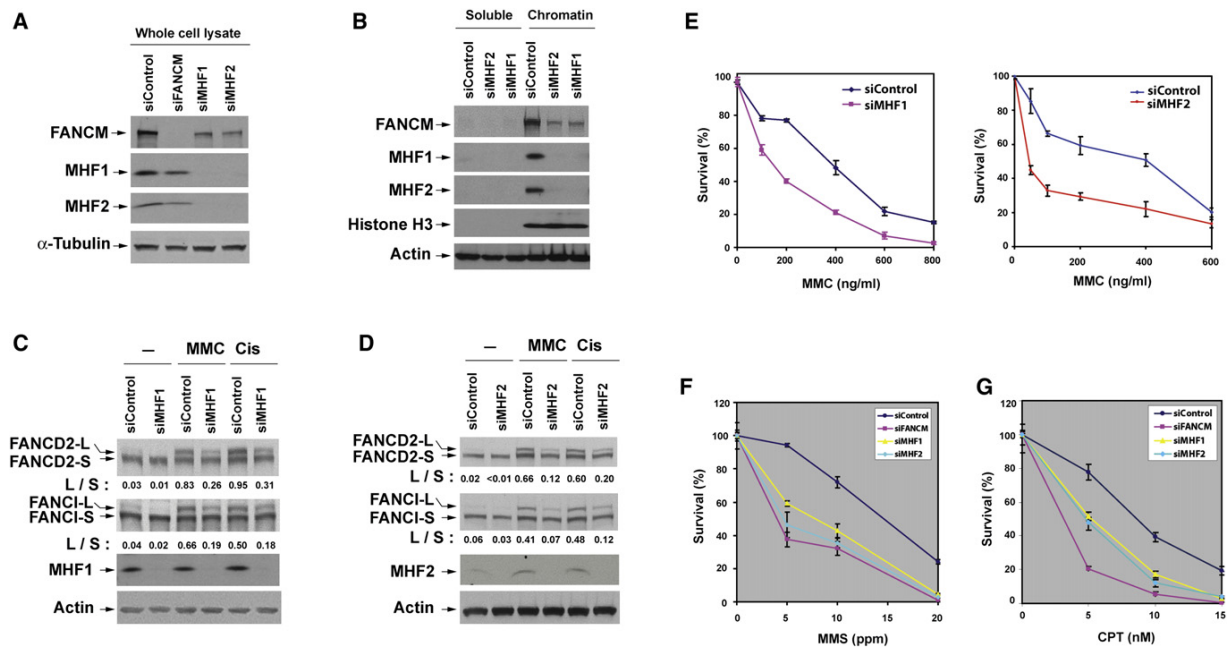
Figure 4. MHF Is Required for Stability of FANCM, Activation of the FA Pathway, and Cellular Resistance to DNA-Damaging Agents

(A and B) Immunoblotting shows that depletion of MHF1 or MHF2 reduces the level of FANCM in whole-cell lysates (A) and chromatin fractions (B) of HeLa cells. Nontargeting siRNA oligos were used as a control. α-Tubulin, Histone H3, and Actin were included as loading controls.

(C and D) Immunoblotting shows that HeLa cells depleted of MHF1 (C) or MHF2 (D) have reduced levels of monoubiquitinated FANCD2 and FANCI in the presence of MMC (60 ng/ml) or cisplatin (5 μM) (*Cis*). "L" (long) and "S" (short) represent ubiquitinated and nonubiquitinated forms, respectively. The ratio between long and short forms was obtained by using KODAK Molecular Imaging Software and shown below the blots.

(E, F, and G) Clonogenic survival assays of HeLa cells depleted of MHF1, MHF2, or FANCM by siRNA following the treatment with MMC, MMS, or CPT at the indicated concentrations. Three independent experiments were performed. The results were reproducible, and a representative data of mean surviving percentage with standard error of the mean (SEM) from triplicate cultures are shown (see also Figure S5).

and nonreplicating conditions (Figure 5E), in agreement with previous findings (Shen et al., 2009). These data are congruent with the notion that MHF functions at the replication forks blocked by ICLs, most likely in the form of the FANCM-MHF complex.

MHF and FANCM Act in the Same Pathway for FANCD2 Monoubiquitination and Suppression of Sister-Chromatid Exchange

We have screened FA patients for mutations in MHF1 and MHF2 but have failed to identify such individuals (data not shown). To

Figure 3. MHF Stimulates DNA-Binding and Fork-Reversal Activities of FANCM

(A) A silver-stained gel showing FANCM-MHF complex immunoprecipitated directly by a MHF1 antibody from HeLa nuclear extract (NE). Three major polypeptides were identified as FANCM, MHF1, and MHF2 by mass spectrometric analyses.

(B) A diagram shows the mapping of MHF-interaction domain within FANCM. Left panels show FLAG-tagged wild-type and various deletion mutants of FANCM used in Figure S4D. Right panels show the presence or absence of interactions between various FANCM constructs and MHF.

(C) EMSA showing the DNA-binding activity of MHF and FANCM$_{754}$. The reaction contained the DNA substrate (0.5 nM) with or without recombinant FANCM$_{754}$ protein or MHF complex as indicated. The protein concentrations are: FANCM$_{754}$: 22.5, 45, and 90 nM in lanes 2 to 4 and 6 to 8, respectively; MHF: 300 nM in lanes 5 to 8; FANCM$_{754}$: 15 nM in lanes 10, 14, 15, 16, 18, 22, 23, 24, 26, 30, 31, and 32; MHF: 50, 100 and 150 nM in lanes 11 to 13, 14 to 16, 19 to 21, 22 to 24, 27 to 29, and 30 to 32, respectively. The shifted bands of indicated protein-DNA complex and free DNA probe are illustrated.

(D) A silver-stained gel showing Flag-FANCM purified from baculovirus-infected insect cells (Sf21), either alone or in association with the MHF complex (*left panels*). At these protein concentrations, MHF proteins were hardly detectable by silver-staining because of their low molecular mass. Western blots confirmed that MHF1 and MHF2 were coimmunoprecipitated with Flag-FANCM (*right panels*).

(E) Comparison of the DNA-binding activity of purified FANCM and FANCM-MHF complex on fork DNA substrate by EMSA. 5'-^{32}P-labeled synthetic replication forks (0.5 nM) were incubated for 30 min at room temperature with increasing concentrations of FANCM or FANCM-MHF (0, 0.125, 0.25, 0.5, 1, 2, 4, 6, 8, and 10 nM). MHF does not bind this substrate at 10 nM.

(F) Experimental scheme. Fork reversal converts the replication fork into a four-way junction (b). Complete fork reversal dismantles the joint molecules into nicked circular and labeled linear duplex (c).

(G) Increasing concentrations of Flag-FANCM or Flag-FANCM-MHF (0, 0.125, 0.25, 0.5, 1, 2, 4, 6, 8, and 10 nM) were incubated with the replication fork substrate (0.5 nM) for one hour at 37°C. The different labeled species are (top down) the four-way junction intermediate (b), the original replication fork substrate (a), and the linear duplex end product of fork reversal (c).

(H) Product formation shown in (G) was quantified by PhosphorImaging and expressed as percentage of total DNA (see also Figure S4).
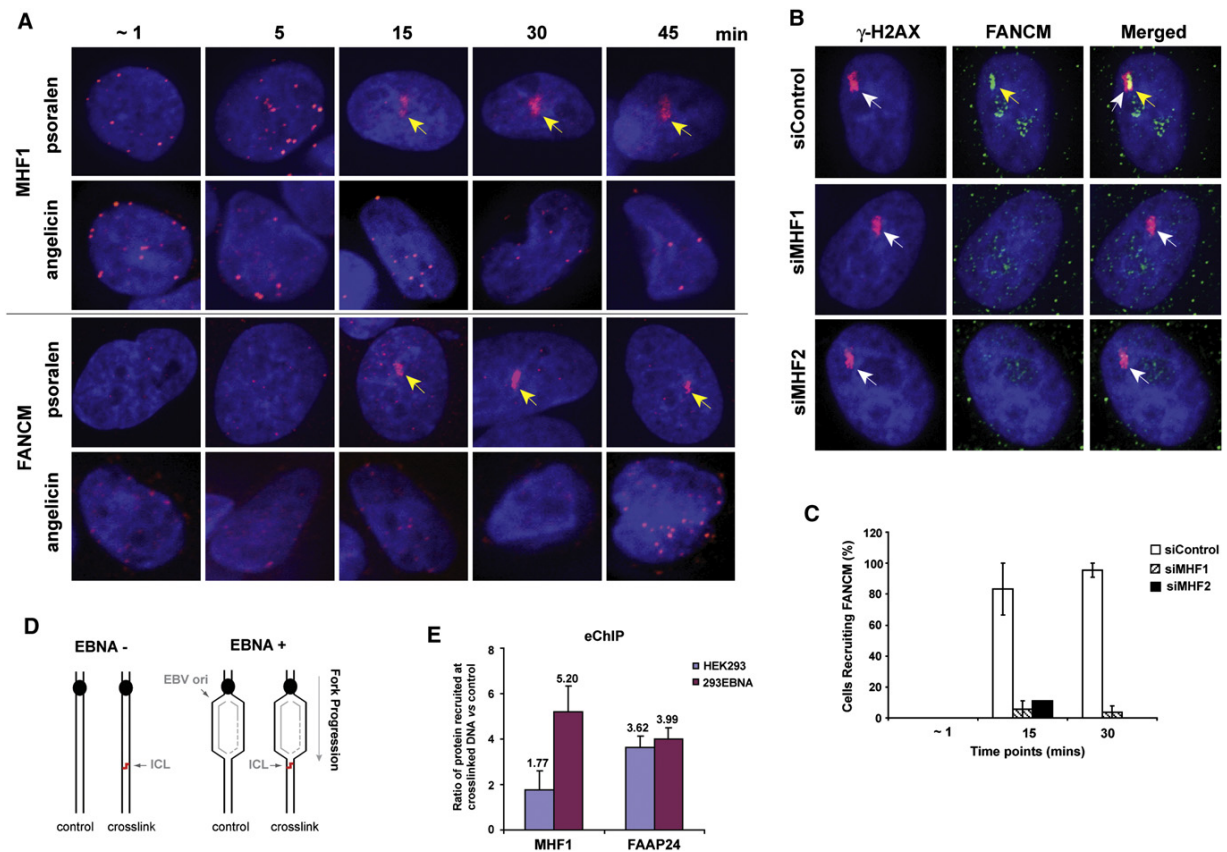
Figure 5. MHF Is Rapidly Recruited to ICL Sites in S Phase Cells and Is Required for FANCM Recruitment

(A) Images showing that MHF1 and FANCM were recruited to psoralen-induced ICLs beginning 15 min after laser photoactivation in S phase cells. The yellow arrow indicates the position of the laser-targeted region.

(B) Images showing that HeLa cells depleted of MHF1 or MHF2 by siRNA have deficient recruitment of FANCM to psoralen-induced ICLs. Nontargeting siRNA oligos and γ-H2AX were used as controls. Images represent 15 min after treatment time point. The recruitment of FANCM and γ-H2AX to ICL sites was indicated by yellow and white arrows, respectively.

(C) Graph showing the percentages of cells in which FANCM was recruited to psoralen-induced ICLs at various time points after photoactivation. No recruitment of FANCM to ICLs was observed in MHF2-depleted cells at 30 min after treatment. HeLa cells were either transfected with siRNA against MHF1, MHF2 or control siRNA and synchronized in S phase prior to the experiment. Values are averages from two independent experiments with error bars representing standard error of the mean.

(D) Schematic representation of the plasmid substrates used in the eChIP assay. The presence of psoralen-ICL is indicated. Cells expressing EBNA (+) support replication of the plasmid, whereas those lacking it (−) do not.

(E) Recruitment of MHF1 to the site of ICL was detected by eChIP. The enrichment of MHF1 and FAAP24 at the ICL in replicating (293EBNA) and nonreplicating (HEK293) DNA substrates was reflected by their relative recruitment ratio compared to the control substrate without ICL. The relative recruitment was derived by normalizing the comparative concentration (from real-time PCR) of crosslinked substrate to that of the unmodified control substrate. Error bars represent standard deviation from three independent experiments (see also Figure S6).

study the functions of MHF genetically, we inactivated *MHF1* in chicken DT40 cells (Figures S7A, S7B, and S7C). Compared with wild-type cells, *MHF1*$^{−/−}$ cells exhibited a lower level of FANCM and MHF2, a reduced level of monoubiquitinated FANCD2 (Figure 6A, lanes 1–2), and a decreased number of FANCD2 nuclear foci (Figures S7D and S7E). Introduction of human MHF1 into *MHF1*$^{−/−}$ cells restored FANCD2 monoubiquitination and also resulted in overexpression of FANCM (2-fold) and MHF2 (11-fold), compared with wild-type cells (Figure 6A, lanes 1–3). These findings are consistent with siRNA data from

HeLa cells that MHF1 is required for normal FANCD2 monoubiquitination and for stability of FANCM and MHF2.

We generated *FANCM*$^{−/−}$/*MHF1*$^{−/−}$ double-mutant cells (data not shown) and found that they contained the levels of monoubiquitinated FANCD2 comparable to that of *FANCM*$^{−/−}$ cells in the presence of MMC (Figure 6B). These results suggest that MHF and FANCM act in a common pathway to promote efficient monoubiquitination of FANCD2.

DT40 cells inactivated of FANC genes exhibit higher levels of sister-chromatid exchanges (SCEs) (Rosado et al., 2009). The

Molecular Cell
A Conserved DNA-Remodeling Complex in DNA Repair

level of SCEs in $MHF1^{-/-}$ cells was about 3–4-fold higher than that of wild-type cells (9.7 versus 2.5), and this elevated SCE level could be corrected by expression of human MHF1 (Figure 6C). The data suggest that MHF participates in suppression of SCEs in DT40 cells. The SCE level in $MHF1^{-/-}$ cells is lower than that of $FANCM^{-/-}$ cells (9.7 versus 18.3), suggesting that, without MHF, FANCM remains partially active in maintaining genome integrity (Figure 6C). The SCE level of $FANCM^{-/-}$/$MHF1^{-/-}$ cells is comparable to that of $FANCM^{-/-}$ cells (17.9 versus 18.3), indicating that MHF and FANCM act through the same pathway to suppress SCEs.

$MHF1^{-/-}$ DT40 cells lacked cellular sensitivity and chromosomal breakage in response to DNA ICL drugs (data not shown), which are phenotypes of DT40 cells inactivated of FANC genes. These results differ from those of siRNA studies in HeLa cells. Possibly, the balance between DNA repair and cell death pathways may be different between these cells.

The DNA-Binding Activity of MHF Is Required for Normal FANCD2 Monoubiquitination and Full Suppression of SCE

To study whether the DNA-binding activity of MHF is required for its function in vivo, we generated three MHF1 point mutants by substituting two clusters of conserved positively charged amino acid residues with alanine: mutant A (K73A/R74A), B (R87A/R88A), and AB (K73A/R74A/R87A/R88A) (Figure 6D). Mutagenesis of similar residues in other histone-fold proteins can disrupt protein-DNA interactions (Hori et al., 2008). We coexpressed these mutants with MHF2 in E. coli and found that only mutant A can be copurified with MHF2 in a stable complex (Figure 6E and data not shown). Notably, the recombinant complex containing mutant A lacked dsDNA-binding activity (Figure 6F) and also failed to recruit FANCM to fork DNA (Figure 6G, compare lanes 4 and 6), indicating that MHF requires its DNA-binding activity to recruit FANCM to DNA. Coimmunoprecipitation analyses in HEK293 cells showed that mutant A retained normal association with MHF2 and FANCM (Figure 6H). Importantly, $MHF1^{-/-}$ cells stably expressing mutant A had a lower level of monoubiquitinated FANCD2 (Figure 6A, lanes 3–4), and a higher SCE frequency (Figure 6C) than cells transfected with wild-type MHF1, suggesting that the DNA-binding activity of MHF is required for normal FANCD2 monoubiquitination and full suppression of SCE. Compared to $MHF1^{-/-}$ null cells, the mutant A-expressing $MHF1^{-/-}$ cells reproducibly exhibited a higher level of monoubiquitinated FANCD2 and a lower SCE frequency (Figure 6A, lanes 2 and 4; Figure 6C), indicating that this mutant remains partially functional even though it lacks the ability to bind DNA. This partial function might be due to the ability of mutant A to stabilize FANCM and MHF2, because the latter two proteins were recovered to levels higher than not only those of $MHF1^{-/-}$ null cells, but also those of wild-type cells (Figure 6A, lanes 1, 2 and 4). The findings that mutant A–complemented cells had a higher than normal amount of FANCM but still exhibited abnormal FANCD2 monoubiquitination and SCE frequency suggest that the stabilization and overexpression of FANCM cannot substitute the function of MHF in vivo.

The Interaction between MHF and FANCM Is Essential for FANCM Stability

We also analyzed MHF1 mutant B and AB using the same assays. These mutants cannot be copurified with MHF2 in a stable complex from E. coli (data not shown), indicating that protein interactions between MHF1 and MHF2 were impaired. Consistent with such an impairment, coimmunoprecipitation in HEK293 cells showed that mutant B was partially defective in association with MHF2 and FANCM, whereas mutant AB was completely deficient (Figure 6H). Notably, the levels of FANCM and MHF2 in the mutant B-complemented $MHF1^{-/-}$ DT40 cells were restored to near those of cells expressing wild-type MHF1 or mutant A (Figure 6A). This is in contrast to mutant AB–complemented $MHF1^{-/-}$ cells, in which the levels of FANCM and MHF2 were not restored and similar to those of $MHF1^{-/-}$ null cells, suggesting that the stability of these two proteins strongly depends on their interactions with MHF1.

The levels of monoubiquitinated FANCD2 and SCE in mutant B–complemented $MHF1^{-/-}$ cells were found to be intermediate between $MHF1^{-/-}$ null cells and those expressing wild-type MHF1 (Figures 6A and 6C). In contrast, the levels in the mutant AB–complemented $MHF1^{-/-}$ cells were indistinguishable to those of the $MHF1^{-/-}$ null cells. The degree of defects in FANCD2 monoubiquitination and SCE by the two mutants appears to correlate with their ability to stabilize FANCM. Together with the data of mutant A, the results suggest that MHF has two important activities: binding to DNA and stabilizing FANCM by protein-protein interactions. Only when both activities are inactivated (as in mutant AB), MHF completely loses its ability to promote FANCD2 monoubiquitination and suppress SCEs.

Budding Yeast MHF and FANCM Orthologs Work in the Same Pathway in Resistance to MMS-Induced DNA Damage

We investigated whether yeast orthologs of MHF (Figures S2C and S2D) and FANCM work together in DNA repair like their vertebrate counterparts. We disrupted budding yeast orthologs of MHF1 and MHF2 in a srs2 mutant background and found that both mutants displayed hypersensitivity to MMS (Figure 7A, middle panels, compare mhfΔ srs2Δ doubles with srs2Δ single mutant). This feature resembles that of an mph1Δ mutant, which was previously shown to have MMS hypersensitivity in combination with a srs2Δ mutant (Banerjee et al., 2008) (Figure 7A). The mhf1Δ mph1Δ or mhf2Δ mph1Δ double mutants displayed MMS hypersensitivity similar to that of the mph1Δ single mutant in the srs2Δ background (Figure 7A, bottom panels). Furthermore, the survival frequencies of the triple mutant strains at multiple MMS concentrations were statistically indistinguishable from srs2Δ mph1Δ (Figure 7B). Collectively, these results indicate that both MHF proteins are epistatic with Mph1 in tolerating MMS-induced DNA damage. In addition, a mhf1Δ mhf2Δ double mutant exhibited MMS hypersensitivity indistinguishable from each mhfΔ single mutant in the srs2Δ background (Figure 7A, middle panels), consistent with the evidence in mammalian cells that the two proteins work in the same complex. The MMS resistance of the mhf1Δ or mhf2Δ
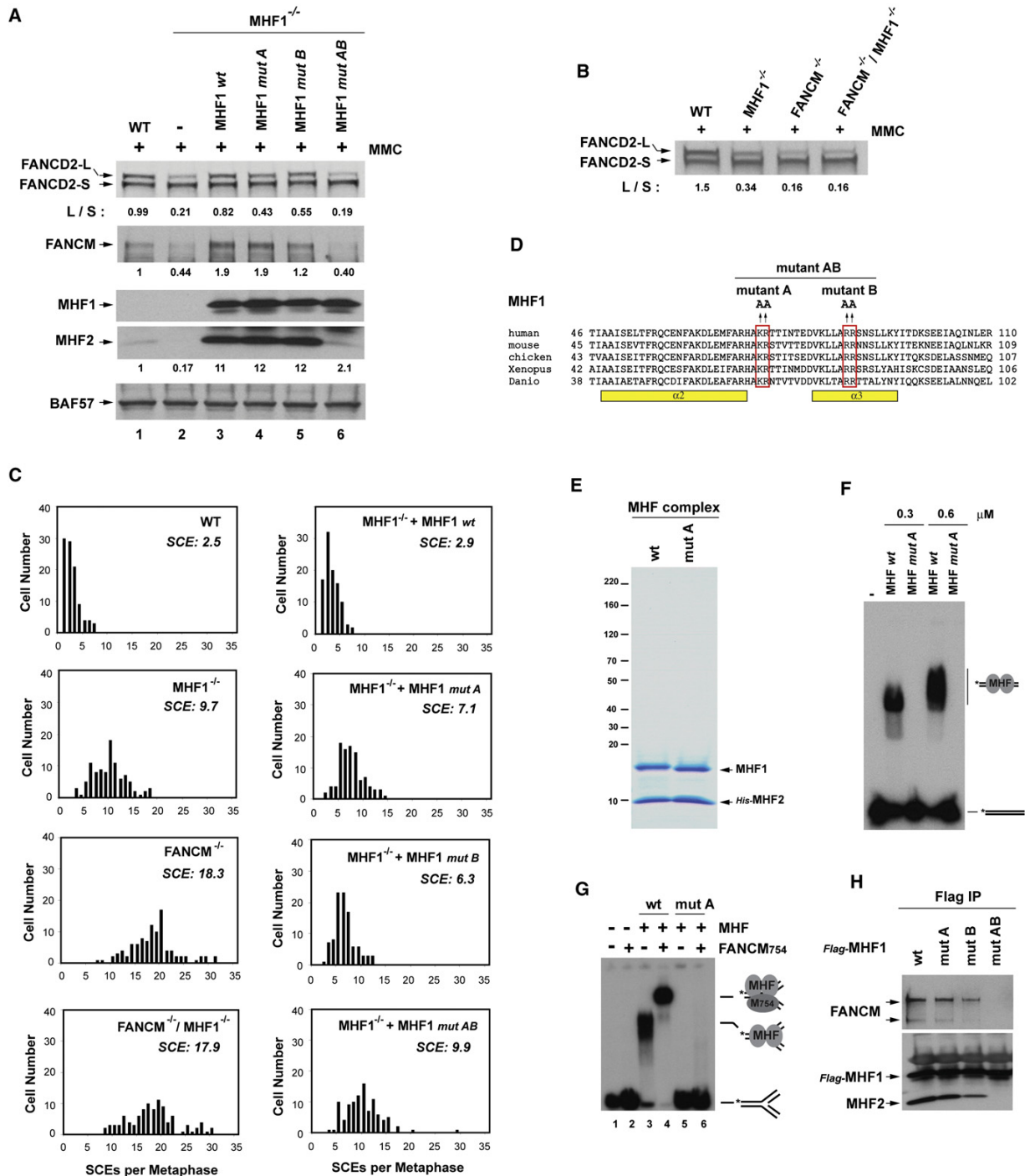
Figure 6. MHF and FANCM Act in the Same Pathway for FANCD2 Monoubiquitination and Suppression of SCE in Chicken DT40 Cells

(A) Immunoblotting shows levels of FANCD2, FANCM, and MHF2 in whole-cell lysates from various DT40 cells (wild-type [WT], $MHF1^{-/-}$ cells, and $MHF1^{-/-}$ cells complemented with human WT or mutant MHF1 as described in D). The ratio between the monoubiquitinated and unubiquitinated FANCD2 (L/S) was shown. Cells were treated with MMC (50 ng/ml) for 18 hr. BAF57 was used as a loading control.

(B) Immunoblotting shows levels of monoubiquitinated and unubiquitinated FANCD2 in whole-cell lysates from various DT40 cells as indicated on the top.

Molecular Cell

A Conserved DNA-Remodeling Complex in DNA Repair

mutant is weaker than that of the *mph1Δ* mutant, reminiscent of results in DT40 cells where the SCE level of the *MHF1$^{-/-}$* mutant is lower than that of the *FANCM$^{-/-}$* mutant (Figure 6C). FANCM or Mph1 may therefore be partially functional without MHF. Taken together, these data support the findings in mammalian cells and indicate that the MHF proteins play an evolutionarily conserved role—cooperating with Mph1/FANCM to protect genome integrity.

MHF2 and FANCM Act in the Same Pathway to Promote Gene Conversion at Stalled Replication Forks in Fission Yeast

The FANCM ortholog in *S. pombe*, Fml1, promotes Rad51-dependent gene conversion at blocked replication forks (Sun et al., 2008). To investigate whether MHF is also required, we deleted the *mhf2* gene and assessed what impact this had on the frequency of Ade$^+$ recombinants induced by a polar replication fork barrier (*RTS1*) positioned between a direct repeat of *ade6$^-$* heteroalleles on chromosome 3 of *S. pombe* (Figure 7C). This region is replicated unidirectionally, and, consequently, only orientation 2 of *RTS1* blocks replication between the repeats. The block of replication results in a strong induction of Ade$^+$ recombinants, which arise both from Rad51-dependent gene conversion, and Rad51-dependent and independent deletions (Sun et al., 2008). A His3$^+$ gene positioned between the repeats enables these two types of recombinants to be distinguished: Ade$^+$His3$^+$ are conversion-types, whereas Ade$^+$His3$^-$ are deletion-types. In a *mhf2Δ* mutant, the frequency of conversion-types in the absence of replication fork blockage at *RTS1* (i.e., when *RTS1* is in orientation 1) is not significantly different from wild-type (data not shown). However, when *RTS1* is in orientation 2 and replication forks are blocked, the frequency of conversion-types is reduced approximately two-fold in a *mhf2Δ* mutant, compared with wild-type (Figure 7D, left panel). This reduction is statistically significant ($p \leq 0.0001$), but is less than with the *fml1Δ* mutant, which shows a 7-fold reduction. Possibly, Fml1 retains some ability to promote gene conversion without MHF. Importantly the frequency of conversion-types in a *fml1Δ mhf2Δ* double mutant is the same as in a *fml1Δ* single mutant, indicating that MHF2 acts in the same pathway to promote gene conversion at blocked replication forks as Fml1 does.

Intriguingly, the loss of conversion-types in the *mhf2Δ* mutant is accompanied by an approximately two-fold increase in deletion-types (Figure 7D, right panel), which again is statistically significant ($p \leq 0.0001$). This increase is dependent on Fml1 as it is suppressed in a *fml1Δ mhf2Δ* double mutant ($p \leq$ 0.01). These data suggest that MHF may act to prevent Fml1-mediated Rad51-dependent strand invasion events from giving rise to deletions.

DISCUSSION

A network of DNA caretaker proteins keeps replication forks under surveillance to prevent gross chromosomal rearrangements in S phase. Here we have identified a conserved DNA-remodeling complex that protects replication forks in eukaryotes. This complex contains FANCM, a branched DNA-binding and -remodeling enzyme, as well as two histone-fold proteins, MHF1 and MHF2. In vertebrates, MHF1 and MHF2 are bona fide components of the FA core complex and are required for normal activation of the FA pathway, including optimal monoubiquitination of the FANCI-FANCD2 complex in response to DNA damage, cellular resistance to DNA crosslinking drugs, and prevention of chromosomal breakage. MHF and FANCM also constitute a FANC-independent complex, FANCM-MHF, which is rapidly recruited to blocked forks in vivo. Our genetic analyses in yeast suggest that FANCM-MHF is a conserved complex that promotes gene conversion at blocked replication forks, in contrast to the FA core complex, which exists only in vertebrates (Figure 7E).

MHF Is a Histone-Fold Complex that Functions in DNA Damage Response and at Centromeres

Histone-fold proteins can facilitate transcription (TAF octamer in TFIID), chromatin remodeling (Chrac14/Chrac16 in the CHRAC complex), and stabilization of centromeres (CENP-T/CENP-W). Our finding that MHF functions with FANCM extends the participation of histone-fold complexes to the DNA damage response. While our article was under review, Amano and colleagues described a CENP-S complex that is identical to MHF, with MHF1 and MHF2 being CENP-S and CENP-X, respectively (Amano et al., 2009). They showed that this complex is essential for the stable assembly of kinetochore structure and normal mitosis. Centromeric DNA is intrinsically difficult to replicate and rich in replication pause sites. MHF may target FANCM at centromeres to forestall the disastrous encounter of replication forks with secondary DNA structures and/or tightly DNA-bound protein complexes.

FANCM and MHF Constitute a Conserved DNA-Remodeling Complex

MHF1 and MHF2 have yeast orthologs that are genetically epistatic with the FANCM orthologs in resistance to MMS and

(C) Histograms showing spontaneous SCE levels of various DT40 cells as indicated within each graph. *MHF1$^{-/-}$* cells transfected with human MHF1 wild-type (wt) or mutants (described in D) were also shown. The mean number of SCEs per metaphase is listed.

(D) Illustration of three human MHF1 point mutants generated by substituting 2 clusters of positively charged amino acid residues (KR and RR) to alanine as indicated.

(E) A Coomassie blue–stained SDS-PAGE gel shows the purified recombinant MHF complex containing MHF1 of either wild-type (wt) or mutant A (mut A).

(F) Comparison of the DNA-binding activity of MHF complexes containing either wild-type or mutant A of MHF1 by EMSA. Reactions contain ^{32}P-labeled dsDNA substrate (0.5 nM) and 0.3 or 0.6 μM of the MHF complex as indicated.

(G) Comparison of the DNA-binding activity of FANCM$_{754}$ with either wild-type (wt) or mutant A MHF by EMSA. The reaction contained the ^{32}P-labeled fork DNA substrate (0.5 nM) with or without FANCM$_{754}$ protein or MHF complex (wt or mut A). The protein concentrations are: FANCM$_{754}$: 22.5 nM (lanes 2, 4, and 6); MHF: 450 nM (lanes 3 and 5) and 300 nM (lanes 4 and 6).

(H) IP-Western analyses show association between FANCM, MHF2, and various Flag-tagged MHF1 point mutants in HEK293 cells (see also Figure S7).
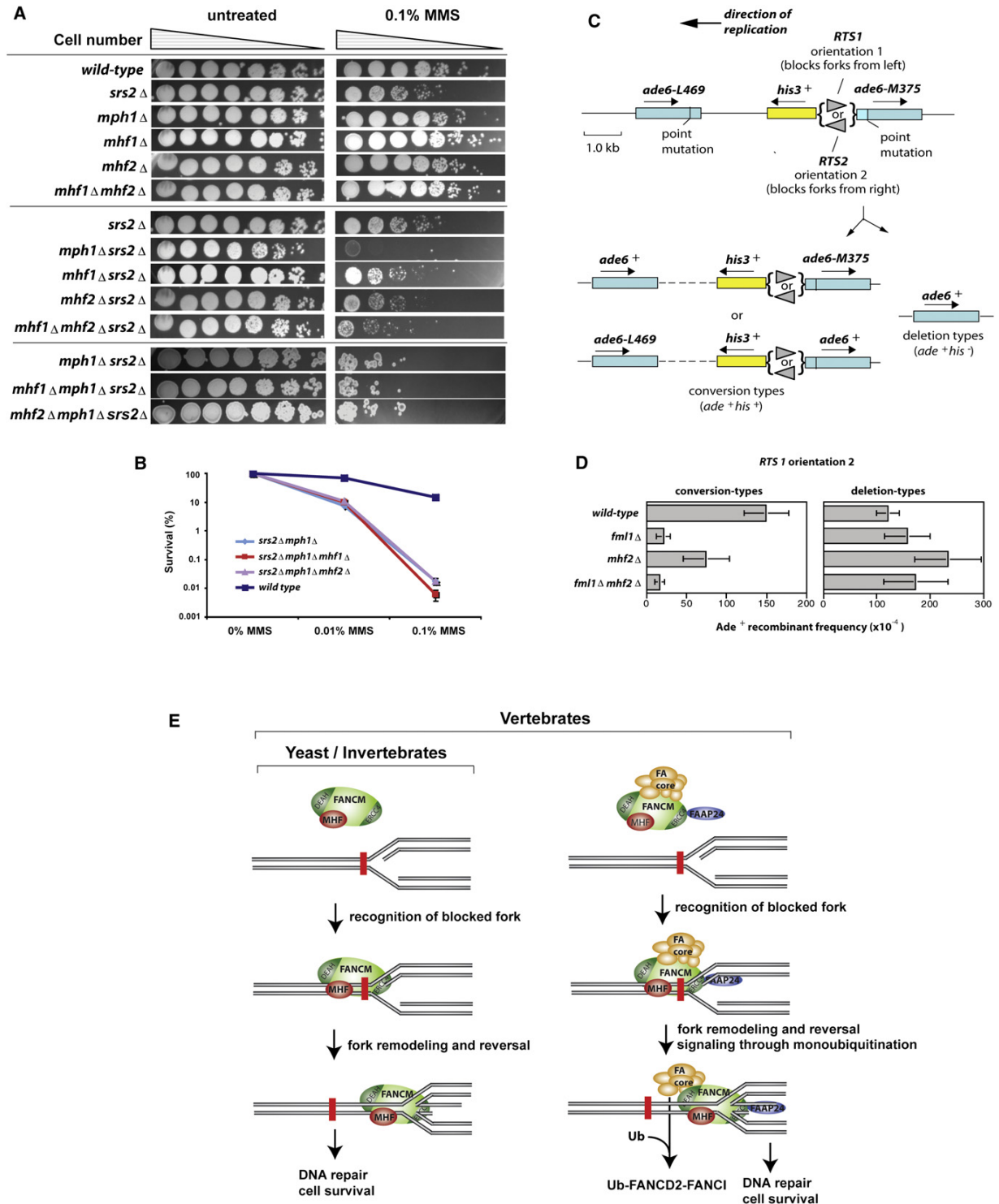
Figure 7. Yeast MHF and FANCM Orthologs Work in the Same Pathway to Resist MMS-Induced DNA Damage and Promote Gene Conversion at Stalled Replication Forks

(A) Qualitative spot tests for strain sensitivities to MMS reveal a common role and shared pathway for Mph1, MHF1, and MHF2 in DNA damage resistance.

Molecular Cell

A Conserved DNA-Remodeling Complex in DNA Repair

promotion of gene conversion at stalled forks. This leads us to propose that FANCM and the two MHF proteins constitute a core complex that remodels DNA structure in yeast and higher eukaryotes. In accord with this hypothesis, human MHF and FANCM can be coimmunoprecipitated as a complex that is largely free of the other FA core complex components. The FANCM-MHF complex can be reconstituted by baculovirus-expressed proteins, and the resulting complex possesses DNA-binding and fork-reversal activities stronger than FANCM alone. Both FANCM and MHF are rapidly recruited to ICLs specifically in S phase, and the recruitment of MHF1 to ICLs is stimulated by DNA replication. Moreover, depletion of either FANCM or MHF in HeLa cells results in cellular sensitivity to DNA crosslinking drugs, which blocks progression of replication forks. Furthermore, Fml1 and MHF mutants in *S. pombe* are defective in promoting gene conversion at replication fork barriers. These findings suggest that FANCM-MHF may stabilize and remodel blocked replication forks to facilitate DNA damage signaling and repair.

How might MHF contribute to the remodeling activity of FANCM-MHF? MHF may anchor the complex to DNA damage sites in chromatin. Consistent with this, MHF can bind in vitro assembled chromatin template, and its depletion reduces the level of FANCM in the chromatin fraction and abrogates recruitment of FANCM to laser-induced ICLs. A MHF point mutant unable to bind dsDNA is defective in recruiting FANCM to DNA in vitro and fails to promote normal FANCD2 monoubiquitination and fully suppress SCE in DT40 cells. Our biochemical studies suggest that, at stalled forks, FANCM most likely binds the DNA branch point of replication intermediates (Gari et al., 2008b; Xue et al., 2008), whereas MHF would associate with surrounding dsDNA. Binding of MHF to dsDNA may help to recruit and orient FANCM at the branch point. In the FA core complex, FAAP24 may further facilitate this process by anchoring FANCM to the surrounding ssDNA region of stalled forks. After FANCM is recruited to stalled forks, MHF may facilitate repair or bypass of the stalled forks by stimulating fork reversal activity of FANCM. This stimulation may be independent of the role of MHF in FANCD2 monoubiquitination, which has been shown not to require remodeling activity of FANCM (Rosado et al., 2009; Singh et al., 2009; Xue et al., 2008).

In a cellular context, MHF may play an additional role in stabilizing FANCM via protein-protein interactions. This was revealed because the MHF mutant A is only partially defective in FANCD2 monoubiquitination and SCE suppression despite inactivation of its DNA-binding activity. We found a strong correlation between FANCM stability and its ability to associate with various MHF point mutants. MHF mutant A has normal association with FANCM and can fully stabilize FANCM as the wild-type protein. Mutant B has reduced association with FANCM and cannot restore FANCM to the level by the wild-type protein. Mutant AB is completely defective in FANCM association so that it fails to stabilize FANCM. Notably, mutant AB, which lacks both DNA-binding activity and FANCM association, is completely deficient in promoting normal FANCD2 monoubiquitination and suppressing SCE. These data suggest that MHF requires both DNA binding and FANCM association to function in vivo.

FANCM-MHF and FA Core Complex May Have Independent Functions

In vertebrates, MHF also plays a signaling role as part of the FA core complex, because its depletion or inactivation of its DNA-binding activity reduces FANCD2 monoubiquitination in response to DNA damage. A recent report shows that FANCD2 monoubiquitination in vitro requires only FANCL but not other subunits of the FA core complex (Alpi et al., 2008), suggesting that the role of MHF in this reaction is indirect. A previous study suggested that the DNA-binding activity of FANCM is required for FANCD2 monoubiquitination (Xue et al., 2008), possibly by anchoring FA core complex to chromatin (Kim et al., 2008). Our findings that MHF stabilizes FANCM and cooperates with FANCM to bind DNA suggest that MHF indirectly contributes to this process by stimulating recruitment of FANCM and associated FA core complex to chromatin.

In addition to their differential roles in FANCD2 monoubiquitination, FA core complex and FANCM-MHF appear to have other independent functions. For example, HeLa cells lacking MHF or FANCM display MMS and camptothecin hypersensitivity, a cellular feature absent in human cells deficient of other FA core complex components. Moreover, FANCM and MHF are epistatic in suppression of SCEs in DT40 cells, whereas FANCM and FANCC are not (Rosado et al., 2009).

In vertebrates, FANCM-MHF associates not only with components of the FA core complex but also with subunits of the BLM complex (BLM, Topo IIIa, RMI, and RPA). FANCM and BLM suppress SCE through the same pathway in DT40 cells (Rosado et al., 2009). These results reinforce our earlier findings that FANC and BLM complexes work together in the BRAFT supercomplex to protect genome integrity (Wang, 2007). Elucidation of the structure and function of this large complex should shed light on the mechanism of replication fork surveillance.

EXPERIMENTAL PROCEDURES

Gel Filtration, Immunoprecipitation, and Protein Identification

Superose 6 gel filtration and immunoprecipitation were performed as described elsewhere (Ciccia et al., 2007). The eluted immunoprecipitates

---

(B) Quantitative analysis of MMS sensitivities shows that the triple mutants are not statistically distinguishable from *srs2Δ mph1Δ* at two different MMS concentrations. Error bars are standard deviations. *P* values for the *mhf1Δ* and *mhf2Δ* triple mutants, respectively, versus *srs2Δ mph1Δ* are as follows: 0.35, 0.23 (0.01% MMS); 0.36, 0.97 (0.1% MMS).

(C) Schematic showing the direct repeat recombination substrate on chromosome 3 of *S. pombe* plus the two types of Ade[+] recombinants (Sun et al., 2008).

(D) Ade[+] recombinant frequencies for various strains as indicated. Error bars are the standard deviations.

(E) A model describes common and unique functions of FANCM-MHF and FA core complex in signaling and repairing of blocked replication forks. The former acts in yeast, invertebrates and vertebrates, whereas the latter functions only in vertebrates. Although both complexes can promote reversal of blocked forks (indicated by the red brick) to allow subsequent repair, only the FA core complex has a signaling role in monoubiquitinating FANCD2 and FANCI.

were resolved by 8%–16%Tris-Glycine gel (InVitrogen), visualized by silver-staining, and subjected to mass spectrometry analyses.

Fork Regression Assay

The plasmid-based replication fork reversal assay was performed as described elsewhere (Gari et al., 2008a).

Protein Recruitment to Laser-Induced Localized ICLs

We followed a previously published protocol to detect proteins recruited at laser-induced localized ICLs (Thazhathveetil et al., 2007).

Detecting Proteins Recruited to a Site-Specific Psoralen-ICL by eChIP

The eChIP was performed as described elsewhere (Shen et al., 2009).

Budding Yeast Strains and MMS Sensitivity Assay

We followed previously published protocols used for Mph1 study to generate *mhf* mutants and performed MMS sensitivity assay (Banerjee et al., 2008).

S. pombe Strains and Recombination Assay

Strain construction and the direct repeat recombination assay were performed in accordance with previously published protocols (Sun et al., 2008).

REFERENCES

Alpi, A.F., Pace, P.E., Babu, M.M., and Patel, K.J. (2008). Mechanistic insight into site-restricted monoubiquitination of FANCD2 by Ube2t, FANCL, and FANCI. Mol. Cell *32*, 767–777.

Amano, M., Suzuki, A., Hori, T., Backer, C., Okawa, K., Cheeseman, I.M., and Fukagawa, T. (2009). The CENP-S complex is essential for the stable assembly of outer kinetochore structure. J. Cell Biol. *186*, 173–182.

Arents, G., and Moudrianakis, E.N. (1993). Topography of the histone octamer surface: repeating structural motifs utilized in the docking of nucleosomal DNA. Proc. Natl. Acad. Sci. USA *90*, 10489–10493.

Banerjee, S., Smith, S., Oum, J.H., Liaw, H.J., Hwang, J.Y., Sikdar, N., Motegi, A., Lee, S.E., and Myung, K. (2008). Mph1p promotes gross chromosomal rearrangement through partial inhibition of homologous recombination. J. Cell Biol. *181*, 1083–1093.

Ciccia, A., Ling, C., Coulthard, R., Yan, Z., Xue, Y., Meetei, A.R., Laghmani el, H., Joenje, H., McDonald, N., de Winter, J.P., et al. (2007). Identification of FAAP24, a Fanconi anemia core complex protein that interacts with FANCM. Mol. Cell *25*, 331–343.

Deshpande, G.P., Hayles, J., Hoe, K.L., Kim, D.U., Park, H.O., and Hartsuiker, E. (2009). Screening a genome-wide S. pombe deletion library identifies novel genes and pathways involved in genome stability maintenance. DNA Repair (Amst.) *8*, 672–679.

Foltz, D.R., Jansen, L.E., Black, B.E., Bailey, A.O., Yates, J.R., III, and Cleveland, D.W. (2006). The human CENP-A centromeric nucleosome-associated complex. Nat. Cell Biol. *8*, 458–469.

Gari, K., Decaillet, C., Delannoy, M., Wu, L., and Constantinou, A. (2008a). Remodeling of DNA replication structures by the branch point translocase FANCM. Proc. Natl. Acad. Sci. USA *105*, 16107–16112.

Gari, K., Decaillet, C., Stasiak, A.Z., Stasiak, A., and Constantinou, A. (2008b). The Fanconi anemia protein FANCM can promote branch migration of Holliday junctions and replication forks. Mol. Cell *29*, 141–148.

Hori, T., Amano, M., Suzuki, A., Backer, C.B., Welburn, J.P., Dong, Y., McEwen, B.F., Shang, W.H., Suzuki, E., Okawa, K., et al. (2008). CCAN makes multiple contacts with centromeric DNA to provide distinct pathways to the outer kinetochore. Cell *135*, 1039–1052.

Kim, J.M., Kee, Y., Gurtan, A., and D'Andrea, A.D. (2008). Cell cycle-dependent chromatin loading of the Fanconi anemia core complex by FANCM/FAAP24. Blood *111*, 5215–5222.

Krona, C., Ejeskar, K., Caren, H., Abel, F., Sjoberg, R.M., and Martinsson, T. (2004). A novel 1p36.2 located gene, APITD1, with tumour-suppressive properties and a putative p53-binding domain, shows low expression in neuroblastoma tumours. Br. J. Cancer *91*, 1119–1130.

Prakash, R., Satory, D., Dray, E., Papusha, A., Scheller, J., Kramer, W., Krejci, L., Klein, H., Haber, J.E., Sung, P., and Ira, G. (2009). Yeast Mph1 helicase dissociates Rad51-made D-loops: implications for crossover control in mitotic recombination. Genes Dev. *23*, 67–79.

Rosado, I.V., Niedzwiedz, W., Alpi, A.F., and Patel, K.J. (2009). The Walker B motif in avian FANCM is required to limit sister chromatid exchanges but is dispensable for DNA crosslink repair. Nucleic Acids Res. *37*, 4360–4370.

Shen, X., Do, H., Li, Y., Chung, W.-Y., Tomasz, M., de Winter, J.P., Xia, B., Elledge, S.J., Wang, W., and Li, L. (2009). Recruitment of Fanconi Anemia and breast cancer proteins to DNA damage sites is differentially governed by replication. Mol. Cell *35*, 716–723.

Singh, T.R., Bakker, S.T., Agarwal, S., Jansen, M., Grassman, E., Godthelp, B.C., Ali, A.M., Du, C.H., Rooimans, M.A., Fan, Q., et al. (2009). Impaired FANCD2 monoubiquitination and hypersensitivity to camptothecin uniquely characterize Fanconi anemia complementation group M. Blood *114*, 174–180.

Sun, W., Nandi, S., Osman, F., Ahn, J.S., Jakovleska, J., Lorenz, A., and Whitby, M.C. (2008). The FANCM ortholog Fml1 promotes recombination at stalled replication forks and limits crossing over during DNA double-strand break repair. Mol. Cell *32*, 118–128.

Tebbs, R.S., Hinz, J.M., Yamada, N.A., Wilson, J.B., Salazar, E.P., Thomas, C.B., Jones, I.M., Jones, N.J., and Thompson, L.H. (2005). New insights into the Fanconi anemia pathway from an isogenic FancG hamster CHO mutant. DNA Repair (Amst.) *4*, 11–22.

Thazhathveetil, A.K., Liu, S.T., Indig, F.E., and Seidman, M.M. (2007). Psoralen conjugates for visualization of genomic interstrand cross-links localized by laser photoactivation. Bioconjug. Chem. *18*, 431–437.

Thompson, L.H., and Hinz, J.M. (2009). Cellular and molecular consequences of defective Fanconi anemia proteins in replication-coupled DNA repair: mechanistic insights. Mutat. Res. *668*, 54–72.

Wang, W. (2007). Emergence of a DNA-damage response network consisting of Fanconi anaemia and BRCA proteins. Nat. Rev. Genet. *8*, 735–748.

Xue, Y., Li, Y., Guo, R., Ling, C., and Wang, W. (2008). FANCM of the Fanconi anemia core complex is required for both monoubiquitination and DNA repair. Hum. Mol. Genet. *17*, 1641–1652.

Supplemental Information

A Histone-Fold Complex and FANCM Form a Conserved DNA-Remodeling Complex to Maintain Genome Stability

**Zhijiang Yan, Mathieu Delannoy, Chen Ling, Danielle Daee,
Fekret Osman, Parameswary A. Muniandy, Xi Shen, Anneke B. Oostra,
Hansen Du, Jurgen Steltenpool, Ti Lin, Beatrice Schuster,
Chantal Décaillet, Andrzej Stasiak, Alicja Z. Stasiak, Stacie Stone,
Maureen E. Hoatlin, Detlev Schindler, Christopher L. Woodcock,
Hans Joenje, Ranjan Sen, Johan P. de Winter, Lei Li, Michael M. Seidman,
Matthew C. Whitby, Kyungjae Myung, Angelos Constantinou,
and Weidong Wang**

Supplemental Experimental Procedures

Cell lines, antibodies and DNA-damaging drugs

HeLa and HEK293 cells were grown in DMEM medium supplemented with 10% fetal bovine serum (FBS). Wild-type and mutant chicken DT40 cells were grown in RPMI-1640 medium supplemented with 10% FBS, 1% chicken serum, 1% Hepes, in a 5% CO_2 incubator at 39.5^0C. *FANCM$^{-/-}$* DT40 cell line was kindly provided by Dr. K.J. Patel.

The antibodies against various FANC proteins have been previously described (Ciccia et al., 2007; Ling et al., 2007; Shen et al., 2009; Xue et al., 2008). Anti-MHF1 and MHF2 antibodies were raised against chimeric proteins containing full length MHF1 or MHF2 fused to maltose-binding protein (New England Biolabs). Anti-chicken FANCD2 antibody was kindly provided by Dr. M. Takata. Anti-chicken FANCM antibody was kindly provided by Dr. K.J. Patel. Anti-FANCA antibody for Immunoblotting and anti-actin were purchased from Bethyl Laboratories, Inc. Anti-Flag antibody and anti-Flag M2 agarose beads were purchased from Sigma. Anti-histone H3 was purchased from Upstate. Anti-α-tubulin was purchased from Santa Cruz Biotecnology, Inc. Anti-BAF57 antibody has been previously described (Yan et al., 2008). Anti-γH2AX antibody was purchased from Upstate Biotechnologies.

The drugs used for induction of DNA damage, including mitomycin C (MMC), cisplatin, methyl methanesulfonate (MMS), and camptothecin (CPT) were purchased from Sigma.

Plasmid construction, mutagenesis and transfection

Human MHF1 cDNA clone was purchased from Invitrogen. Human MHF2 cDNA was generated by RT-PCR using HeLa mRNA as a source. The expression plasmid pIRES-Flag-MHF2 used for generation of HeLa stable cell lines was constructed by inserting MHF2 cDNA with a FLAG tag at its N-terminus into the *EcoRI* and *NotI* sites of pIRES-Neo3 (Clontech). pcDNA3.1-Flag-MHF1 expression plasmid was created by inserting human MHF1 cDNA with a FLAG tag at the N-terminus into the *NheI* and *BamHI* sites of pcDNA3.1-Zeocin (Invitrogen). To generate point mutations in MHF1, site-directed mutagenesis was performed using the QuikChange kit from Stratagene. pcDNA-Flag-FANCM wildtype and deletion mutants were constructed by inserting FANCM full length cDNA or deletion fragment with a FLAG tag at its N-terminus into the *NheI* and *BamHI* sites of pcDNA3.1. Lipofectamine 2000 (Invitrogen) was used for transfection in HeLa and HEK293 cells.

Mammalian two-hybrid analysis

Human embryonic kidney HEK293 cells were plated onto six well plates. After 48h, the cells were transiently transfected with MHF1, MHF2 or FAAP24 cDNA fused to either the GAL4 activation domain (pVP16; Clontech) or to the GAL4 DNA-binding domain (pM; Clontech) (1 μg of each), together with a GAL4 driven reporter plasmid (G5E1bLUC, 0.2 μg). The luciferase activity was monitored after 24 h using a Dual-Luciferase Reporter Assay System (Promega) and a single tube luminometer (DLReady, Berthold Detection Systems), according to the manufacturer's instructions. All GAL4 constructs were sequenced to confirm the correct reading frame and each experimental data set was performed in triplicate to overcome the variability inherent to transfections.

siRNA experiments

siRNA pool oligos (ON-TARGETplus SMARTpool L-032895-01 against human MHF1, ON-TARGETplus SMARTpool L-016829-01 against MHF2, ON-TARGETplus SMARTpool L-021955-00 against FANCM and ON-TARGETplus siCONTROL Non-targeting pool D-001810-10) were purchased from Dharmacon. HeLa cells were

transfected with these siRNA pool oligos using Lipofectamine RNAi MAX (Invitrogen) according to manufacturer's protocol. Two to three days post-transfection, assays for FANCD2/FANCI monoubiquitination and chromosomal aberrations were performed as described (Ciccia et al., 2007; Meetei et al., 2003). Clonogenic survival assays for MMC, MMS and cisplatin were carried out as described (Meetei et al., 2005). For CPT survival assay, cells were treated with CPT for 48 hours.

Cell extraction and fractionation

The lysis buffer (10 mM Tris-HCl at pH 7.5, 1% Triton X-100, 0.1% SDS, 0.1% sodium deoxycholate, 140mM NaCl, 1 mM EDTA) was used for the preparation of whole cell extract. All buffers were supplemented with complete protease inhibitor cocktail (Roche). For cellular fractionation, the cells were washed with PBS twice and were lysed in low salt buffer (20mM Tris-HCL, pH 7.5, 100mM NaCl, 1mM EDTA, 0.5% Triton X-100) for 5 min on ice. Following centrifugation, the supernatant was collected for soluble fraction (soluble cytoplasmic and nucleoplasmic proteins). The pellet was washed once with low salt buffer, once with micrococcal nuclease buffer (20mM HEPES, pH 7.9, 300 mM NaCl, 20mM KCl, 3mM $CaCl_2$, 0.5mM DTT, 0.5mM PMSF) and then digested with micrococcal nuclease ($3U/\mu l$) for 20 min at room temperature. Following digestion, 5mM EGTA was added to the extract. After centrifugation, the supernatant was used as chromatin fraction (chromatin-bound proteins).

Expression and purification of *HIS*-tagged recombinant proteins from *E.coli.*

E.coli expression plasmid pET-MHF1 or pET-MHF2 was constructed by cloning human MHF1 or MHF2 cDNA into the *NdeI* and *NotI* sites of pET28a (Novagen). To generate bicistronic MHF1/MHF2 coexpression construct, MHF1 cDNA with a StrepII tag at the N-terminus and an internal ribosomal entry site in the upstream was cloned into the *NotI* and *XhoI* sites of pET-MHF2 plasmid.

E. coli Rosetta (Novagen) cells carrying pET-MHF1, pET-MHF2, pET-MHF2/MHF1 construct or mutant version were grown at 37^0C to OD_{600} of 0.8 and induced with 0.4mM IPTG at 35^0C for 3 h. The cell pellet of 1 liter culture was lysed in 15 ml of 1X BugBuster (Novagen) containing 1mM β-mercaptoethanol, 5 mM imidazole and complete EDTA-free protease inhibitor cocktail (Roche). Following incubation at

room temperature for 30 min under constant rotation, the mixture was centrifuged twice (each for 30 min at 15,000rpm in 4^0C). The supernatant was then incubated for 2 hrs with 1ml Talon metal affinity resin (BD Bioscience). The Talon resin was poured into a column, washed with 50 ml of buffer (50mM NaH_2PO_4, 500mM NaCl, 0.1%NP-40, pH 7.8) supplemented with 30 mM imidazole, and eluted (buffer: 50mM NaH_2PO_4, 500mM NaCl , 500mM imidazole, pH 7.8). Peak fractions were pooled and dialyzed against 2 liters of PBS containing 5% glycerol in cold room for 6 hours.

Electrophoretic mobility shift assay

The indicated amounts of proteins were incubated with 0.5nM of 5'-^{32}P-labeled DNA substrates in PBS for 30 min at room temperature. The reactions were loaded on 6% native polyacrylamide gel in TBE and run in 0.5xTBE at 4^0C. Labeled DNA products were visualized by autoradiography. The oligos used for generating DNA substrates have been previously described (Ciccia et al., 2007), unless indicated otherwise.

Cloning of MHF1 and MHF2 cDNAs into baculovirus vectors

The cDNAs were amplified from pIRES-MHF1 and pIRES-MHF2 by PCR with primers containing Gateway attB-sites at their 5'-ends, and subsequently cloned into pDONR221 via gateway BP-reactions according to the manufacturer's instructions (Invitrogen). MHF1 and MHF2 cDNAs were sequence verified and transferred via LR-reactions into the Gateway destination vector pDEST10 (N-terminal 6xHis tag).

Gateway primers:

MHF1 forward:

5'GGGGACAAGTTTGTACAAAAAAGCAGGCTTCGCCACCATGGAGGAGGAGG CGGAGACC3';

MHF1 reverse:

5'GGGGACCACTTTGTACAAGAAAGCTGGGTCTTAATTCTCACTTTCCACCACT CC3'.

MHF2 forward:

5'GGGGACAAGTTTGTACAAAAAAGCAGGCTTCGCCACCATGGAGGGAGCAG GAGCTGGA3';

MHF2 reverse:

5'GGGGACCACTTTGTACAAGAAAGCTGGGTCCGCCTAGAAGTCCAGGA GCAGC3'.

Expression and purification of baculoviral recombinant proteins from insect cells

Baculoviruses were prepared according to the manufacturer's instructions (InVitrogen). Insect cells at a concentration of $2x10^6$ cells/ml were infected at an MOI of 1 with recombinant baculoviral particles coding for Flag-FANCM, His-MHF1 and His-MHF2. Cells were collected 48 h post infection and resuspended in 5 PCV of buffer A containing 50 mM Na_2HPO_4/NaH_2PO_4 (pH 7.0), 150 mM NaCl, 10 % glycerol, 0.1 % NP-40, 0.5 mM EDTA, 2 mM TCEP and protease inhibitors. The suspension was incubated on ice for 30 min, lysed using a Dounce homogeniser (25 strokes with a tight pestle), and centrifuged for 1 h at 100000xg. The supernatant was then incubated with 100μl of anti-Flag M2-agarose bead, and rotated overnight at 4°C. The beads were washed twice for 30 minutes with buffer A supplemented with 100 mM Arginine (pH 7.0), and then once with buffer A. Recombinant Flag-FANCM or Flag-FANCM/His-MHF1/His-MHF2 was eluted with 120 μl of buffer A containing 100 μg/ml 3x-Flag-peptide (Sigma). The eluate was aliquoted and stored at -80°C. Protein concentration was estimated by silver staining (20 nM).

Annealing assay

Reactions (10 μl) were performed in 25 mM Na_2HPO_4/NaH_2PO_4 (pH 7.0), 75 mM NaCl, 5 % glycerol, 0.005 % NP-40, 0.25 mM EDTA, 1 mM DTT, 100 μg/ml BSA and contained 3.5 nM 5'-^{32}P labeled oligo 1 (TGGGTGAACCTGCAGGTGGGCAAAGATGTCCTAGCAATCCATTGTCTATGAC G), 3.5 nM oligo 2 (CGTCATAGACAATGGATTGCTAGGACATCTTTGCCCACCTGCAGG TTCACCCA) and MHF (100 nM) or FAAP24 (100 nM). At the indicated time, annealing reactions were stopped with 20 mM EDTA, 0.4 % SDS, 2 mg/ml Proteinase K and a 100-fold excess of unlabeled oligo 1. Reaction products were resolved by PAGE through 10% polyacrylamide gels in TBE.

Fork regression assay

The plasmid-based replication fork structure is described (Gari et al., 2008a; Ralf et al., 2006). Reactions (10μl) were performed in 25 mM Na_2HPO_4/NaH_2PO_4 (pH 7.0), 75 mM NaCl, 5 % glycerol, 0.005 % NP-40, 0.25 mM EDTA, 1 mM DTT, 100 μg/ml BSA and contained 0.5 nM 5'-^{32}P labeled DNA substrate, 10 nM oligonucleotides (as competitor), 0.5 mM $MgCl_2$, 1 mM ATP, and recombinant proteins at the indicated concentrations. Reactions were carried out for one hour at 37°C, deproteinized for 15 minutes at 37°C with 2mg/ml Proteinase K and 0.4% SDS and resolved on a 0.8% agarose gel containing 0.5 μg/μl ethidium bromide.

Electron microscopy

MHF (1600 nM) was incubated with either linear or supercoiled PUC19 (35 nM) in 25 mM Na_2HPO_4/NaH_2PO_4 (pH 7.0), 75 mM NaCl, 5 % glycerol, 0.005 % NP-40, 0.25 mM EDTA, 1 mM TCEP for 15 min at 4^0C and then for 15 min at 37^0C . Samples were diluted and washed in 5 mM magnesium acetate, and stained with uranyl acetate, as described (Sogo et al., 1987). Protein-DNA complexes were visualized using a Phillips CM12 electron microscope.

Chromatin array preparation, binding reaction and electron microscopy

Nucleosomal arrays were reconstituted by salt dialysis onto DNA consisting of 12 tandem units of a 208 bp segment containing the '601' nucleosome positioning unit as described (Nikitina et al., 2007). Arrays (final DNA concentration 10 μg/ml) were mixed with the desired amount of protein in 50 mM NaCl, 10 mM HEPES pH 7.8, 0.2 mM EDTA, held at room temperature for 20 min, then placed on ice and fixed with 1.2% glutaraldehyde for 15 min before EM preparation. EM samples were prepared essentially as described (Nikitina et al., 2007). The fixed arrays were diluted with buffer as needed, applied to thin, glow-discharged carbon films for 5 min, rinsed with 5 mM MgAcetate, stained for 10 sec with 0.01% aqueous uranyl acetate, rinsed extensively with water, and air dried. Grids were examined in a Tecnai 12 TEM (FEI Corp, Hillsboro, OR) operated at 100 KV using tilted darkfield illumination, and images recorded on a 2048x2048 CCD camera (TVIPS, Gauting, Germany).

Mononucleosome assembly and binding assay

147bp substrate was generated by PCR amplification of pGEM-3Z-601 using forward primer (ACAGGATGTATATATCTGAC) and reverse primer (CTGGAGAATCCCGGTGCCGA). The DNA product was then end-labeled using radioactive γ- ^{32}P-ATP. 100,000cpm DNA mixed with and without 30 ng recombinant histone octamer in the presence of 1M NaCl, incubated at 30 ^{0}C for 30 min. The mixture was serially diluted with increasing volumes of Buffer A (50mM HEPES pH7.9, 1mM EDTA, 5mM DTT and 0.5mM PMSF) of 1.8μl, 3.5μl, 4.7μl, 13μl, incubated at 30 ^{0}C for 30 min for each dilution. Increasing amounts of MHF complex were added to the final dilution and incubated for another 30 min at 30 ^{0}C. To stop the reaction, 67μl of Buffer B (10mM Tris.HCl, pH 7.5, 1mM EDTA, 0.1% NP40, 5mM DTT, 0.5mM PMSF, 20% Glycerol, and 100μg/ml BSA) was added. 2μl was loaded on 6% native polyacrylamide gel.

Protein Recruitment to laser-induced localized ICLs

We followed a previous protocol to detect proteins recruited at laser-induced localized ICLs (Thazhathveetil et al., 2007). Cells seeded in a 35mm glass bottom culture dish (MatTek™) were incubated with 5 μM psoralen (which forms crosslinks) or 40 μM angelicin (which can form only monoadducts) at 37°C for 20 minutes prior to laser photoactivation. In some experiments, cells were synchronized in S phase using a double thymidine block and released 2 hours prior to an experiment. Localized irradiation was performed using the Nikon Eclipse TE2000 confocal microscope equipped with an SRS NL100 nitrogen laser-pumped dye laser (Photonics Instruments, St Charles, IL) that fires 5 ns pulses with a repetition rate of 10 Hz at 365 nm, with a power of 0.7 nW, measured at the back aperture of the 60X objective. The diffraction limited spot size was approximately 300 nm. The laser, controlled by Volocity-5 software (Improvision, Perkin Elmer), was directed to deliver pulses to a specified rectangular region of interest (ROI) within the nucleus of a cell (4x20 pixel, 0.16 micron/pixel) visualized with a Plan Fluor 60X/NA1.25 oil objective. The laser beam fired randomly throughout the ROI until the entire region was exposed, after which the photoactivation of the ROI was repeated. Throughout an experiment, cells were maintained at 37 ^{o}C, 5% CO_2 and 80% humidity using Live Cell™ environmental chamber. At different time intervals, cells from different

areas of the dish were treated with the laser to generate a time-course on a single plate. After the final time point, cells were fixed immediately in freshly prepared 4% formaldehyde in PBS for 10 minutes at room temperature. Cells were then permeabilized with 0.5% Triton X-100, 1% BSA, 100 mM Glycine and 0.2 mg/mL EDTA in PBS on ice for 10 minutes. The cells were subsequently digested with RNAse A in PBS-EDTA (5 mM) solution for 30 minutes at 37 ^{o}C and blocked in 10% goat serum in PBS containing 0.01% sodium azide for 1 hour at 37 ^{o}C or overnight at 4 ^{o}C. They were then incubated with appropriate primary antibody (FANCM, MHF1, or γH2AX) diluted in blocking solution for 1 hour at 37 ^{o}C. After three 10 minute washes using 0.05% Tween-20 in PBS, cells were incubated with a corresponding fluorescent tagged secondary antibody (Invitrogen). After another three 10 minute washes cells were mounted with ProLong Gold antifade reagent with DAPI from Molecular Probes. Stained cells were visualized and imaged using Hamamatsu EM-CCD digital camera attached to the Nikon Eclipse TE2000 confocal microscope.

eChIP DNA substrate construction, preparation and chromatin immunoprecipitation

 The eChIP substrate contains two essential components, the EBV OriP and site-specific DNA crosslinks downstream of the replication origin (approximately 500 bp). Preparation and insertion of psoralen-crosslinked oligonucleotide into plasmid vectors have been described previously (Li et al., 1999; Wang et al., 2001). Crosslinked or uncrosslinked control substrates were electroporated into cells (1.5 x10^{7} HEK293 or HEK293-EBNA cells) with an Amaxa device (Amaxa Biosystems, Germany) and conditions recommended by the manufacture. Chromatin immunoprecipitation was carried out as described (Shen et al., 2009).

Generation of MHF1-deficient DT40 cells, SCE analysis and FANCD2 focus formation assay

 The chicken MHF1 gene locus was identified by searching chicken genome sequence database with human MHF1 protein sequence. The MHF1 targeting vectors for replacement of all five MHF1 gene exons with selection markers (neomycin and puromycin) were constructed by using MultiSite Gateway System kit (Invitrogen)

according to the manufacturer's protocol. The following two pairs of Gateway PCR primers were used for amplifying genomic fragments to generate 5'- and 3'-arms:

5'-arm FW:

GGGGACAACTTTGTATAGAAAAGTTGGAGGACATACGCAAGGTGAGTCATTG;

5'-arm REV:

GGGGACTGCTTTTTTGTACAAACTTGGATGCAGGCTTCTCTATCTTGATTCTATGAG;

3'-arm FW:

GGGGACAGCTTTCTTGTACAAAGTGGGTTAGCCTGCCAGCAATTGCT

3'-arm REV:

GGGGACAACTTTGTATAATAAAGTTGGCTAGCTCAGCTTAGCAATGGCAA.

 Electroporation transfection in DT40 cells was carried out using Nucleofector kit (Amaxa) according to the manufacturer's manual. Neomycin (2 mg/ml) or puromycin (0.5 μg/mL) was used for selection of targeted clones. Southern blotting analysis and RT-PCR were performed as described (Yan et al., 2008). To generate $FANCM^{-/-}$ / $MHF1^{-/-}$-double mutant cells, the same targeting vectors were transfected into $FANCM^{-/-}$ DT40 cells to disrupt MHF1 gene.

To perform complementation analysis, pcDNA3.1-Zeocin expression plasmid carrying Flag-tagged human MHF1 wildtype or its mutant form was transfected into $MHF1^{-/-}$ DT40 cells. 0.5 mg/ml Zeocin was used for selection. Western blotting was performed for selecting the clones that stably express protein.

To measure the SCE levels, DT40 cells (10^5/ml) were incubated in 10 μM BrdU for 2 cell cycles (18-22 hr) and added with colcemid (0.1μg/ml) for 2 hr as described previously(Sonoda et al., 1999).

To analyze FANCD2 focus formation, cells ($3x10^5$/ml) treated with or without MMC(50ng/ml, 20 hr) were fixed with 4% PFA, stained with chicken FANCD2 antibody and followed by Alexa Fluor 488-conjugated secondary antibody (Invitrogen) with DAPI counterstaining as described (Matsushita et al., 2005). Images were captured under Zeiss microscope with AXioCam digital camera and Axiovision software.

Budding yeast strains and MMS sensitivity assay

All strains are isogenic to the S288c background strain YKJM1 (*ura3-52, leu2Δ1, trp1Δ63, his3Δ200, lys2ΔBgl, hom3-10, ade2Δ1, ade8, YEL069::URA3*). Null yeast mutants were generated using the conventional PCR-based gene disruption cassette strategies. Yeast transformations were performed as described previously (Myung et al., 2001; Smith et al., 2004). Yeast extract peptone-dextrose (YPD) was purchased from KD medical and consists of the following ingredients per liter of water: 20g peptone, 10g yeast extract, 20g dextrose, 17g agar. For the qualitative MMS sensitivity assay, 10 mL of YPD was innoculated with 200 μL of a 2 mL overnight culture. Yeast cells were grown to mid-logarithmic phase at 30°C (~3-4 hours), harvested, washed once in water, and resuspended in 100 μL water. 50 μL of each strain was added to either 1 mL of water or 1 mL of freshly made 0.1% MMS. Cells were incubated at 30°C for 2 hours with shaking. Following incubation, cells were harvested, washed once in water, resuspended in 100 μL of water, and serially diluted 1:5 in water. 2 μL of each dilution were spotted onto non-selective YPD. Cell viability was visualized after 2 days (no treatment) or 3 days (MMS) of growth at 30°C. The quantitative MMS sensitivity assay was similar to the qualitative assay, however cells were harvested, diluted, and single colonies were plated onto YPD following MMS treatment. Colonies were counted after 3 days of growth at 30°C. Percent survival was calculated as the number colonies that survived MMS treatment versus untreated controls. Results from three independent clones were averaged and standard deviations were calculated. *P* values were calculated with Microsoft Excel using the two-tailed, two-sample, equal variance Student's t-Test.

S. pombe strains and recombination assay

The *S. pombe mhf2Δ* mutant contains a replacement of the *mhf2* open reading frame with the kanMX4 selectable marker, and was obtained from the Bioneer collection of deletion mutants. The wild-type (MCW1433), *fml1Δ::natMX4* (MCW3061), *mhf2Δ::kanMX4* (MCW4515) and *fml1Δ::natMX4 mhf2Δ::kanMX4* (MCW4517) strains share the following genotype: *ura4-D18 his3-D1 leu1-32 arg3-D4 ade6-M375 int::pUC8/his3$^+$/RTS1 site A orientation 2/ade6-L469*. The direct repeat recombination assay has been described (Ahn et al., 2005). Recombinant frequencies are mean values from at least 15 colonies, and two-sample *t*-tests were used to assess whether differences were significant.

Supplemental Figures and Legends
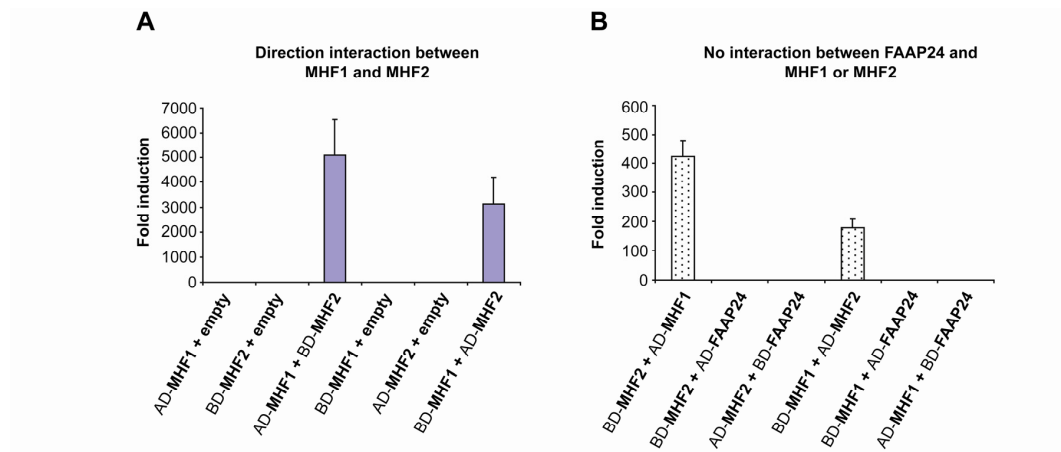
Figure S1 (related to Figure 1). Mammalian two-hybrid assay shows that MHF1 and MHF2 directly interact with each other but not with FAAP24. (A) Mammalian two-hybrid assay showing direct interaction between MHF1 and MHF2. **(B)** The same assay showing no detectable interaction between FAAP24 and MHF1 or MHF2. Indicated proteins are co-expressed in 293 cells as fusion with GAL4 DNA binding domain (BD) or GAL4 activation domain (AD). Fold induction is expressed relative to the luciferase activity obtained with empty vectors (pM and pVP16). Each experimental data set was performed in triplicate. Error bars are the standard deviations.

A

```
                              α1                    α2                      α3
human MHF1          ETEEQQRFSYQQRLKAAVHYTVGCLCEEVALDKEMQFSKQTIAAISELTFRQCENFAKDLEMFARHAKRTTINTEDVKLLARRSNSLLKYITDKSEEIAQ   85
mouse MHF1          EAEEPQEFSHRQRLKAAVHYTVGCLCQEVTLNKQVNFSKQTIAAISEVTFRQCENFAKDLEMFARHAKRSTVTTEDVKLLARRNNSLLKYITEKNEEIAQ  104
chicken MHF1        GGEQRE--LLIQRLRAAVHYTTGCLCQDVAEDKGVLFSKQTVAAISEITFRQCENFARDLEMFARHAKRSTITSEDVKLLARRSNSLLKYITQKSDELAS  107
Xenopus MHF1        EGQEEH-FSRTQRLKAAVHYVVGSLCQEVADDKEIDFSKQAIAAISEITFRQCESFAKDLEIFARHAKRTTINMDDVKLLARRSRSLYAHISKCSDEIAA  101
gallas histone H3   EIRRYQKSTELLIRKLPFQRLVREIAQDF--KTDLRFQSSAVMALQEASEAYLVGLFEDTNLCAIHAKRVTIMPKDIQL-ARRIRG                 127
Drosophila Chrac14           VIGRLIKE-ALPESASVSKEARAAIARAASVFAIFVTSSSTALAHKQNHKTITAKDILQTLTELD-FESFVPSLTQDLEV                      91
Human NC2beta                AINKMIKETLP--NVRVANDARELVVNCCTEFIHLISSEANEICNKSEKKTISPEHVIQ-ALESLGFGSYISEVKEVLQE                     96
Archaeal HMFA                PIGRIIKN---AGAERVSDDARIALAKVLEEMGEEIASEAVKLAKHAGRKTIKAEDIELARKMFK                                    68
yeast histone H4             AIRRLARRGGVK---RISGLIYEEVRAVLKSFLESVIRDSVTVTEHAKRKTVTSLDVVYALKRQ                                     93
Human CENP-T                 GLSHVKLFSFYAKMPMERKALEMVEKCLDKYFQHLCDDLEVFAAHAGRKTVKPEDLELLMRRQG                                    148
```

```
                    Charged tail
human MHF1          INLERKAQKKKKSEDGSKNSRQPAEAGVVESEN                138
mouse MHF1          LNLKRKAKKKRKPEDESRSSRESMAEELDGAEELQSES           142
chicken MHF1        SNMEQKEKKKKKSSAAKGRKTEENETPVTESEDSNMA            139
Xenopus MHF1        NSLEQKEKKNKKSVSGGNVSRNS----DMDTVVPESKD           135
Drosophila chrac14  YRKVVKEKKESKASKKDSNTAENANASATATAE                124
Human NC2beta       CKTVALKRRKASSR                                   110
```
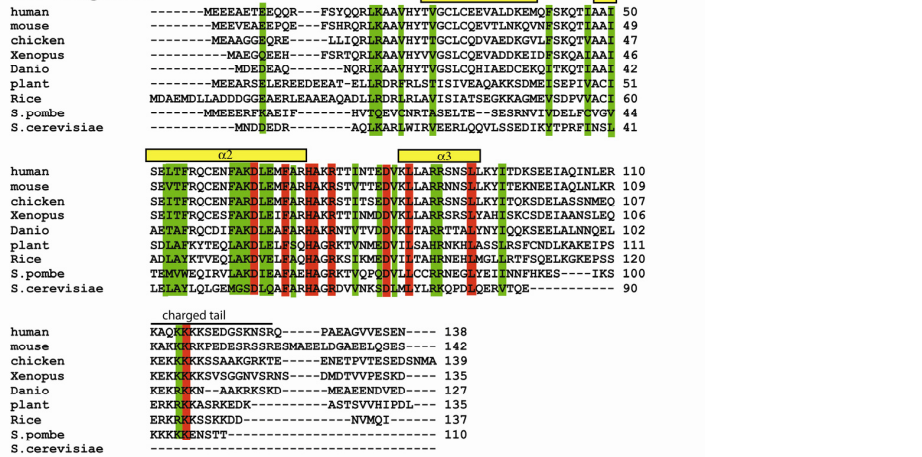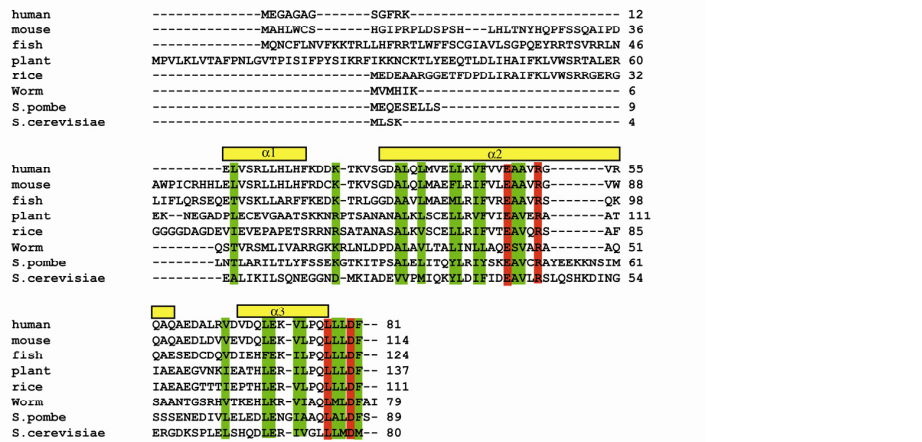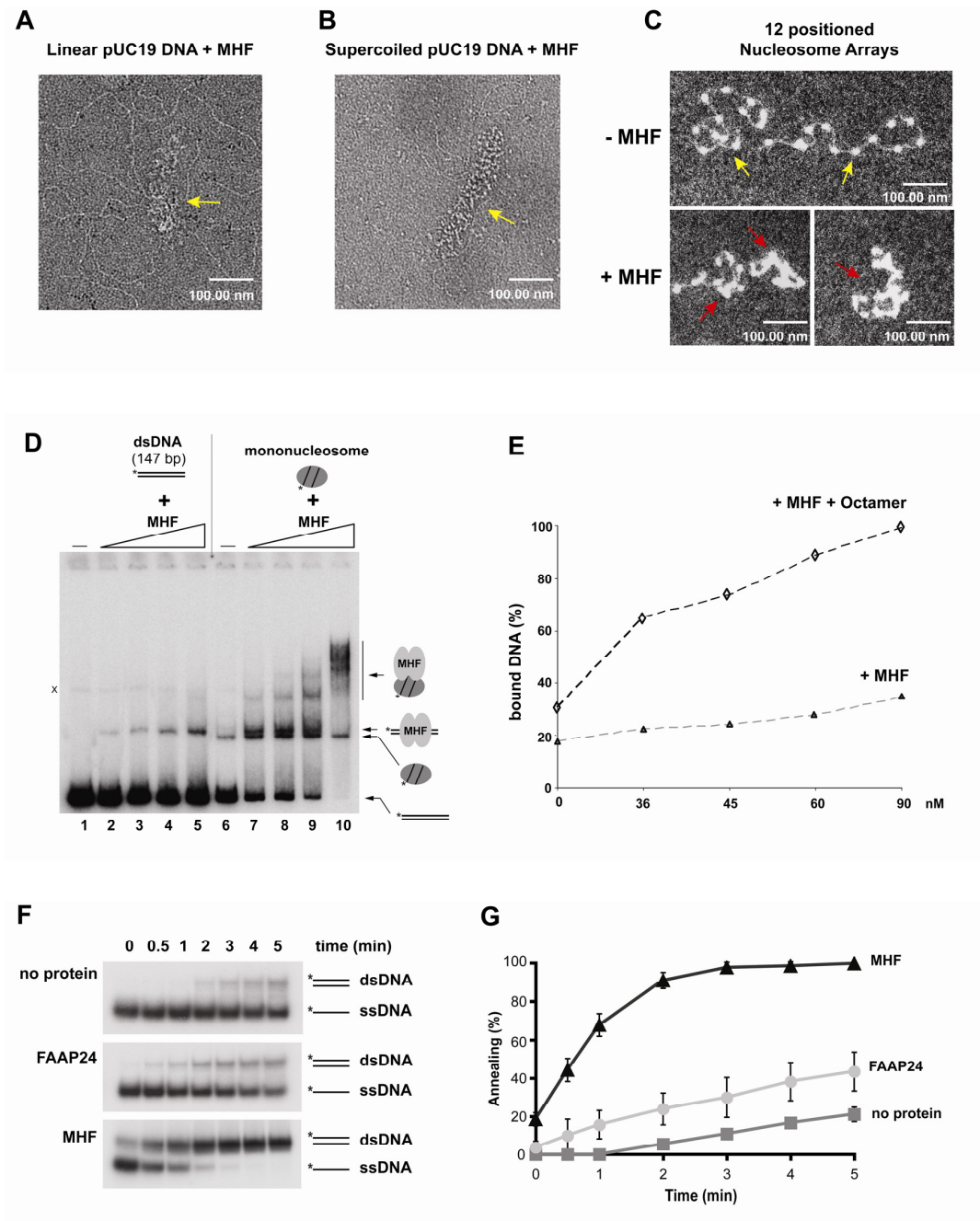
B

```
                              α1                    α2                  α3
Human MHF2    10  FRKELVSRLLHLHFKDDKTKVSGDALQLMVELLKVFVVEAAVRGVRQAQAEDALRVDVDQLEKVLPQL   77 (81)
Mouse MHF2    43  HHLELVSRLLHLHFRDCKTKVSGDALQLMAEFLRIFVLEAAVRGVWQAQAEDLDVVEVDQLEKVLPQL  110 (114)
Fish MHF2     53  SEQETVSKLLARFFKEDKTRLGGDAAVLMAEMLRIFVREAAVRSQKQAESEDCDQVDIEHFEKILPQL  120 (124)
Human NC2beta 16  IPRAAINKMIKETL--PNVRVANDARELVVNCCTEFIHLISSEANEICNKSEKKTISPEHVIQALESL   81 (179)
NF-YB          9  TPTANVARTMKNAI-PQTGKTAKDAKECVQECVSEFISFITSEASERCHQEKRKTTNGEDILFAMSTT   75 (93)
Chrac14       10  LPNAVIGRLIKEAL-PESASVSKEARAAIARAASVFAIFVTSSSTALAHKQNHKTITAKDILQTLTEL   76 (128)
```

C   MHF1 Alignment

```
                                                   α1
human         -------MEEEAETEEQQR---FSYQQRLKAAVHYTVGCLCEEVALDKEMQFSKQTIAAI   50
mouse         --------MEEVEAEEPQE---FSHRQRLKAAVHYTVGCLCQEVTLNKQVNFSKQTIAAI   49
chicken       --------MEAAGGEQRE----LLIQRLRAAVHYTTGCLCQDVAEDKGVLFSKQTVAAI   47
Xenopus       ----------MAEGQEEH----FSRTQRLKAAVHYVVGSLCQEVADDKEIDFSKQAIAAI   46
Danio         -----------MDEDEAQ-------NQRLKAAVHYTVGSLCQHIAEDCEKQITKQTIAAI   42
plant         -------MEEARSELEREEDEEAT-ELLRDRFRLSTISIVEAQAKKSDMEISEPIVACI   51
Rice          MDAEMDLLADDDGGEAERLEAAEAQADLLRDRLRLAVISIATSEGKKAGMEVSDPVVACI   60
S.pombe       -------MMEEERFKAEIF-------HVTQEVCNRTASELTE--SESRNVIVDELFCVGV   44
S.cerevisiae  -----------MNDDEDR--------AQLKARLWIRVEERLQQVLSSEDIKYTPRFINSL   41
```

```
                        α2                             α3
human         SELTFRQCENFAKDLEMFARHAKRTTINTEDVKLLARRSNSLLKYITDKSEEIAQINLER  110
mouse         SEVTFRQCENFAKDLEMFARHAKRSTVTTEDVKLLARRNNSLLKYITEKNEEIAQLNLKR  109
chicken       SEITFRQCENFARDLEMFARHAKRSTITSEDVKLLARRSNSLLKYITQKSDELASSNMEQ  107
Xenopus       SEITFRQCESFAKDLEIFARHAKRTTINMDDVKLLARRSRSLYAHISKCSDEIAANSLEQ  106
Danio         AETAFRQCDIFAKDLEARARHAKRNTVTVDDVKLTARRTTALYNYIQQKSEELALNNQEL  102
plant         SDLAFKYTEQLAKDLELFSQHAGRKTVNMEDYILSAHRNKHIASSLRSFCNDLKAKEIPS  111
Rice          ADLAYKYTEQLAKDLFAQHAGRKSIKMEDYILTAHRNEHIMGLLRTFSQELKGKEPSS  137
S.pombe       TEMVWEQIRVLAKDIEAFAEHAGRKTVQPQDVLLCCRRNEGLYEINNFHKES----IKS  100
S.cerevisiae  LELAYLQLGEMGSDLQAFARHAGRDVVNKSDLMIYLRKQPDLQERVTQE-----------   90
```

```
              charged tail
human         KAQKKKKSEDGSKNSRQ-----PAEAGVVESEN----  138
mouse         KAKKKRKPEDESRSSRESMAEELDGAEELQSES----  142
chicken       KEKKKKKSSAAKGRKTE-----ENETPVTESEDSNMA  139
Xenopus       KEKKKKKSVSGGNVSRNS----DMDTVVPESKD----  135
Danio         KEKRKKN--AAKRKSKD------MEAEENDVED----  127
plant         ERKRKKASRKEDK----------ASTSVVHIPDL---  135
Rice          ERKRKKSSSKKDD--------------NVMQI------  137
S.pombe       KKKKKENSTT-------------------------  110
S.cerevisiae  -----------------------------------
```

D   MHF2 Alignment

```
human         --------------MEGAGAG-------SGFRK--------------------------   12
mouse         --------------MAHLWCS-------HGIPRPLDSPSH---LHLTNYHQPFSSQAIPD   36
fish          --------------MQNCFLNVFKKTRLLHFRRTLWFFSCGIAVLSGPQEYRRTSVRRLN   46
plant         MPVLKLVTAFPNLGVTPISIFPYSIKRFIKKNCKTLYEEQTLDLIHAIFFKLVWSRTALER   60
rice          ---------------------------MEDEAARGGETFDPDLIRAIFKLVWSRRGERG   32
Worm          --------------------------MVMHIK--------------------------    6
S.pombe       ---------------------------MEQESELLS----------------------    9
S.cerevisiae  ---------------------------MLSK----------------------------    4
```

```
                        α1                     α2
human         ---------ELVSRLLHLHFKDDK-TKVSGDALQLMVELLKVFVVEAAVRG-------VR   55
mouse         AWPICRHHLELVSRLLHLHFRDCK-TKVSGDALQLMAEFLRIFVLEAAVRG-------VW   88
fish          LIFLQRSEQETVSKLLARFFKEDK-TRLGGDAAVLMAEMLRIFVREAAVRS-------QK   98
plant         EK--NEGADPLECEVGAATSKKNRPTSANANALKLSCELLRVFVIEAVERA-------AT  111
rice          GGGGDAGDEVIEVEPAPETSRRNRSATANASALKVSCELLRIFVTEAVQRS-------AF   85
Worm          --------QSTVRSMLIVARRGKKRLNLDPDALAVLTALINLLAQESVARA-------AQ   51
S.pombe       --------LNTLARILTLYFSSEKGTKITPSALELITQYLRIYSKEAVCRAYEEKKNSIM   61
S.cerevisiae  ---------EALIKILSQNEGGND-MKIADEVVPMIQKYLDLFIDEAVLRSLQSHKDING   54
```

```
                   α3
human         QAQAEDALRVDVDQLEK-VLPQLLLDF--  81
mouse         QAQAEDLDVVEVDQLEK-VLPQLLLDF--  114
fish          QAESEDCDQVDIEHFEK-ILPQLLLDF--  124
plant         IAEAEGVNKIEATHLER-ILPQLLLDF--  137
rice          IAEAEGTTTIEPTHLER-VLPQLLLDF--  111
Worm          SAANTGSRHVTKEHLER-VIAQLMLDFAI   79
S.pombe       SSSENEDIVLELEDLENGIAAQLALDFS-   89
S.cerevisiae  ERGDKSPLELSHQDLER-IVGLLLMDM--   80
```

Figure S2 (related to Figure 1). MHF1 and MHF2 are histone-fold proteins conserved from human to yeast. (A) Protein sequence alignment of human MHF1 with mouse, chicken and xenopus orthologs as well as some histones and histone-fold proteins as indicated. A charged tail of MHF1 is illustrated by a line. **(B)** Alignment of human MHF2 protein with mouse and fish orthologs. Several known histone-fold proteins were included in the alignment. The identical and similar residues are depicted with red and green backgrounds, respectively. The predicted secondary structures comprising three α helices (helix α1, α2, and α3) are indicated by yellow boxes above the sequences. **(C)** Sequence alignment of human MHF1 (NCBI: NP_954988.1) and its orthologs from various species, including mouse (NCBI: EDL14851.1), chicken (NCBI: XP_001234691.1), *Xenopus laevis* (NCBI: NP_001084907.1), *Danio rerio* (NCBI: NP_001122221.1), plant (NCBI: CAN67893.1), rice (NCBI: CAE05152.2), *S. pombe* (NCBI: NP_596235.1) and *S. cerevisiae* (NCBI: EDN63789.1). **(D)** Alignment of human MHF2 (NCBI: A8MT69.1) and its orthologs from several species, including mouse (NCBI: NP_057874.2), fish (NCBI: CAG00206.1), plant (NCBI: CA064206.1), rice (NCBI: NP_001053111.1), worm (NCBI: NP_501398.1), *S. pombe* (NCBI: NP_588439.1) and *S. cerevisiae* (NCBI: NP_878060.1).

A Linear pUC19 DNA + MHF

B Supercoiled pUC19 DNA + MHF

C 12 positioned Nucleosome Arrays

- MHF

+ MHF

D

dsDNA (147 bp)

mononucleosome

MHF

MHF

lanes 1 2 3 4 5 6 7 8 9 10

E

+ MHF + Octamer

+ MHF

F

no protein

time (min) 0 0.5 1 2 3 4 5

dsDNA
ssDNA

FAAP24

dsDNA
ssDNA

MHF

dsDNA
ssDNA

G

MHF

FAAP24

no protein

Figure S3 (related to Figure 2). MHF binds dsDNA and chromatin, cooperates with histone octamer to assemble into nucleosome structures, and promotes the annealing of complementary single-stranded DNAs. (A and **B)** Electron microscopy (EM) images of MHF complex bound to long dsDNA structure: linear pUC19 DNA in (A), and supercoiled pUC19 DNA in (B). The yellow arrows indicate MHF clusters on dsDNA. The scale bar represents 100 nm. **(C)** EM images of the chromatin arrays with or without MHF. The upper image shows two in vitro assembled chromatin arrays with 12 positioned nucleosomes as indicated by yellow arrows. The lower two images show MHF protein-bound chromatin arrays indicated by red arrows. The ratio of protein to DNA is 5 molecules of MHF / 20bp DNA. The scale bar represents 100 nm. In (C), MHF coated the linker DNA, in agreement with the data that MHF binds dsDNA (Figure 2). The presence of MHF increased intra- and inter-array associations, consistent with EM data of naked DNA (A and B) that MHF can self-associate to form higher order structures.

(D) An autoradiograph of electrophoretic mobility shift assay showing that MHF promotes mononucleosome assembly. Increased amounts of MHF complex (36 nM, lanes 2 and 7; 45 nM, lanes 3 and 8; 60 nM, lanes 4 and 9; 90 nM, lanes 5 and 10) were used. A nonspecific band was marked by "x". **(E)** Quantitative analysis of MHF-bound bands shown in (D).The results were expressed as percentage of bound DNA.

We found that MHF did not bind dsDNA within a pre-assembled mononucleosome (data not shown). However, when MHF was incubated with histones and free DNA in the mononucleosome assembly reaction, we observed formation of higher order complexes containing MHF, histones, and DNA (D, compare lanes 7-10 with lanes 2-5). The presence of histones increased the association between MHF and dsDNA (E), implying that MHF may interact with histones so that they can be co-assembled into higher order protein-DNA complexes during post-replication chromatin assembly.

(F) Autoradiographs showing the strand-annealing activity of the MHF complex. Annealing reactions were supplemented with either FAAP24 (100nM) or MHF (100nM), as indicated. The protein FAAP24 has been described before (Gari et al., 2008b). **(G)** Product formation shown in (F) was quantified by PhosphorImaging and expressed as percentage of dsDNA. Error bars are standard deviations.

MHF promoted the annealing of complementary single-stranded oligonucleotides, which is the activity of several DNA repair proteins (Cheok et al., 2005; Mortensen et al., 1996). FAAP24 had no significant activity in this assay, suggesting that MHF does not simply function as a crowding agent in the annealing reaction. This shows that MHF efficiently overcomes electrostatic repulsions of the negatively charged oligonucleotides, which may play a role in processing of DNA intermediates generated during replication and repair. The annealing activity of MHF was inhibited when the ssDNA was pre-bound by RPA (data not shown). This feature differs from that of RAD52 (Sugiyama et al., 1998), implying that MHF functions differently from RAD52.
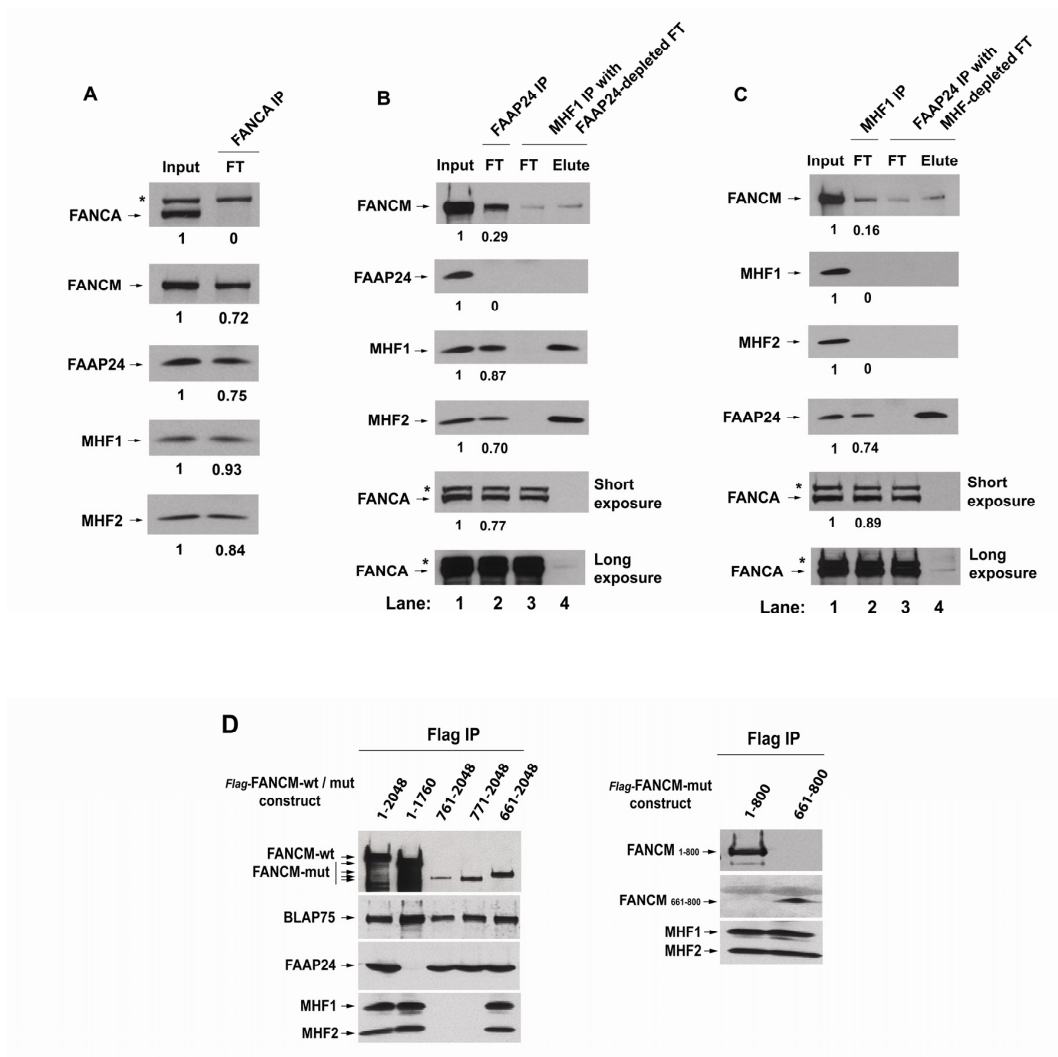
Figure S4 (related to Figure 3). Most of FANCM, MHF and FAAP24 are not associated with the FA core complex; and FANCM can form separate complexes with MHF and FAAP24. MHF interacts with a region near the helicase domain of FANCM. (A) Immunoblotting shows that the majority of FANCM, MHF and FAAP24 retained in the flowthrough (FT) fraction after FA core complex was completely immunodepleted by a FANCA antibody from HeLa nuclear extract. Asterisk indicates a crossreactive polypeptide that can be used as a loading control. The equal amount of nuclear extract was used as an input. The relative levels of indicated proteins on images were obtained by using KODAK Molecular Imaging Software and shown below the blots. The protein level of input was set as "1". **(B)** Immunoblotting shows that a significant fraction of FANCM was co-depleted when FAAP24-associated complexes were immunodepleted. IP-Western (lanes 3 and 4) illustrates co-immunoprecipitation of MHF and FANCM from FAAP24-depleted extract. **(C)** Immunoblotting shows that

majority of FANCM was co-depleted when MHF was depleted by a MHF1 antibody. IP-Western (lanes 3 and 4) shows co-immunoprecipitation of FANCM and FAAP24 in MHF-depleted extract. **(D)** IP-Western to map the MHF-interacting domain in FANCM. Flag-tagged FANCM wildtype (wt) and deletion mutants (mut) were transiently expressed in HEK293 cells, respectively. Immunoprecipitation was carried out using anti-Flag M2 agarose beads. The presence of various FANCM fragments, MHF1, MHF2, FAAP24, and BLAP75 (a control), were detected by Western blotting. A diagram summarizing the FANCM-MHF interaction data is shown in Figure 3B.
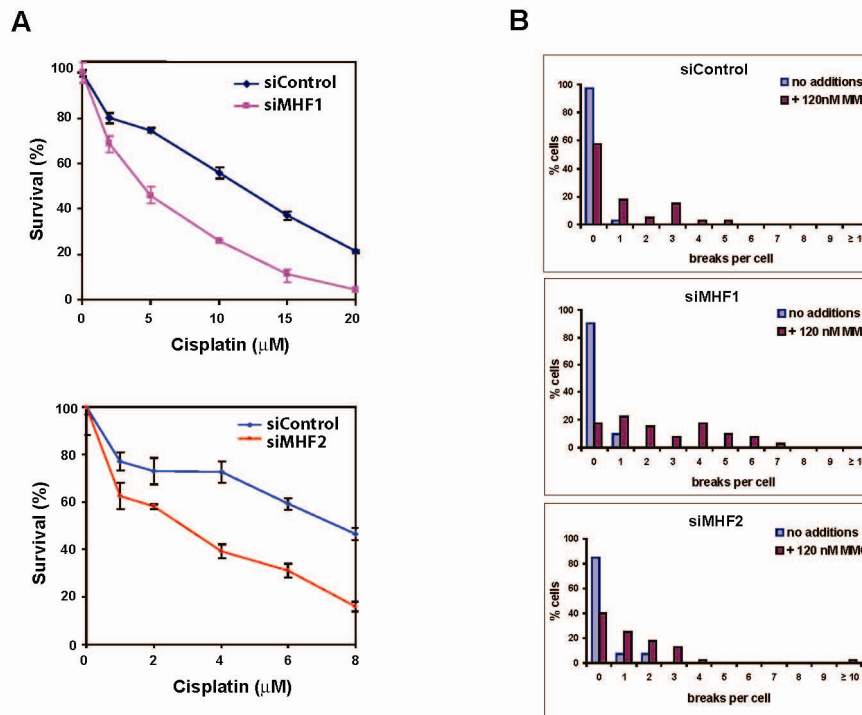
Figure S5 (related to Figure 4). MHF is required for cellular resistance to DNA damage agent and for chromosomal stability. (A) Clonogenic survival assays of HeLa cells depleted of MHF1 (top panel) or MHF2 (bottom panel) by siRNA following the treatment with cisplatin at indicated concentrations. HeLa cells transfected with non-targeting siRNA oligos were used as a control. Three independent experiments were performed, and the results were reproducible. A representative set of data of mean surviving percentage with S.E.M. from triplicate cultures is shown. **(B)** MMC-induced chromosomal aberrations in HeLa cells depleted of MHF1 or MHF2 by siRNA. Non-targeting siRNA oligos were used as a control. At 48hr after siRNA transfection, cells were treated with MMC at indicated concentration for 24 hours or left untreated. Metaphase preparations and chromosomal aberration counting were performed as described previously (Ciccia et al, 2007; Meetei et al, 2003). Percentages of metaphases with and without aberrations after MMC treatment were compared between siControl and siMHF1 or siMHF2 –treated cells using a two-sample Chi2 test. P-values for the differences between siMHF1 and siMHF2 cells were <0.0001 and 0.0587, respectively. When compared to mock transfected cells, p-values were <0.0001 and 0.0068 for siMHF1 and siMHF2 cells, respectively. SiControl and mock transfected cells were not significantly different (p=0.1779).
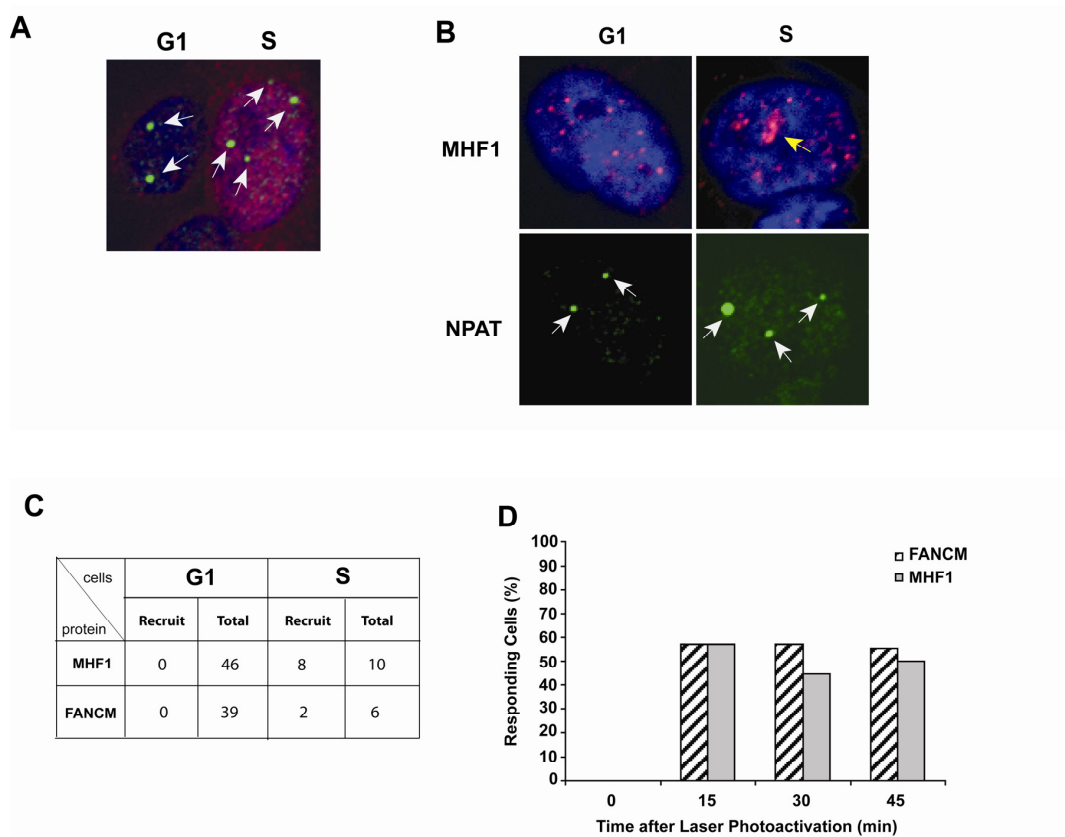
Figure S6 (related to Figure 5). Recruitment of MHF and FANCM to ICLs occurs only in S phase cells. (A) Cell cycle phases were identified by either staining for NPAT (Zhao et al., 2000) (2 spots in G1 or 4 spots in S/G2 cells as indicated by white arrows) or by staining for cyclin A (in red) which is highly expressed in S phase. **(B)** Recruitment of MHF1 was observed in S phase cells, but not in G1 as identified by NPAT staining (green spots as indicated by white arrows. The fourth spot appeared on a different plane). The yellow arrow indicates the laser-targeted region which MHF1 was recruited to. **(C)** In unsynchronized cell population treated randomly with psoralen and laser, recruitment of MHF1 and FANCM was observed only in a fraction of S phase cells and never in G1. **(D)** Bar graph showing the percentage of S phase cells in which MHF or FANCM is recruited to the crosslink at the indicated time points. Cells were synchronized in S phase by double thymidine block and released 2 hr prior to experiments.
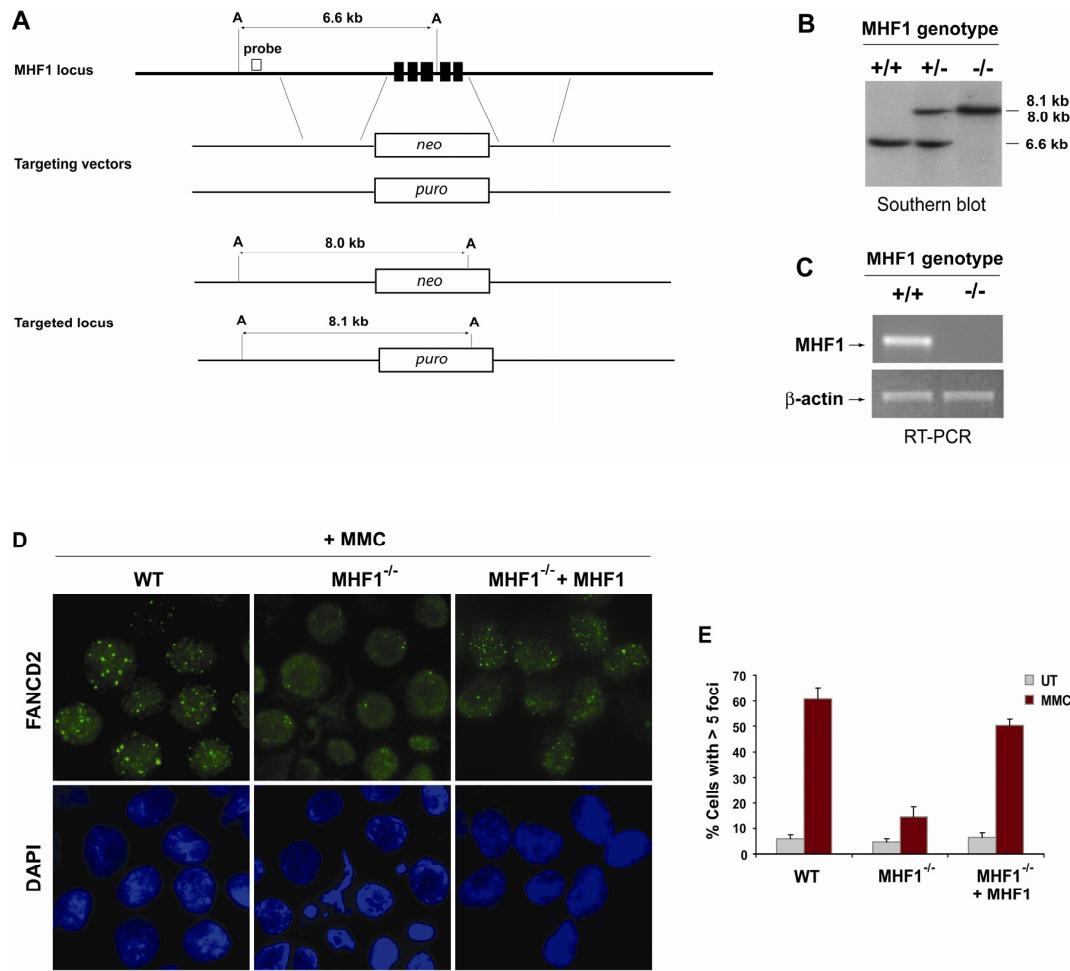
Figure S7 (related to Figure 6). Disruption of MHF1 gene in chicken DT40 cells. MHF is required for formation of FANCD2 foci. (A) A schematic diagram showing chicken MHF1 locus, the gene targeting vectors and targeted alleles. The positions of five exons denoted by black boxes and the probe for Southern blotting are shown. "A" indicates an AflII restriction enzyme site. **(B)** Southern blotting analysis of AflII-digested genomic DNA from the indicated genotypes. **(C)** The expression of MHF1 mRNA in wildtype and MHF1-inactivated DT40 cells was analyzed by RT-PCR. β-actin was used as a control. **(D)** Indirect immunofluorescence shows FANCD2 nuclear foci in DT40 wildtype (WT) cells, *MHF1*$^{-/-}$ cells and *MHF1*$^{-/-}$ cells complemented with human MHF1. Cells were treated with MMC (50 ng/ml) for 18 hr. The images represent immunostaining of FANCD2 in the indicated cells after MMC treatment. **(E)** The bar graphs show the mean values of the percentage of FANCD2-foci-positive cells in untreated (UT) and MMC-treated cells from three independent experiments with standard deviations. A cell containing more than five foci was considered as foci-positive. At least 600 nuclei were counted for each cell line.

Supplemental References

Ahn, J.S., Osman, F., and Whitby, M.C. (2005). Replication fork blockage by RTS1 at an ectopic site promotes recombination in fission yeast. EMBO J *24*, 2011-2023.

Cheok, C. F., Wu, L., Garcia, P. L., Janscak, P., and Hickson, I. D. (2005). The Bloom's syndrome helicase promotes the annealing of complementary single-stranded DNA. Nucleic Acids Res *33*, 3932-3941.

Li, L., Peterson, C.A., Lu, X., Wei, P., and Legerski, R.J. (1999). Interstrand cross-links induce DNA synthesis in damaged and undamaged plasmids in mammalian cell extracts. Mol Cell Biol *19*, 5619-5630.

Ling, C., Ishiai, M., Ali, A. M., Medhurst, A. L., Neveling, K., Kalb, R., Yan, Z., Xue, Y., Oostra, A. B., Auerbach, A. D*., et al.* (2007). FAAP100 is essential for activation of the Fanconi anemia-associated DNA damage response pathway. EMBO J *26*, 2104-2114.

Matsushita, N., kitao, H., Ishiai, M., Nagashima, N., Hirano, S., Okawa, K., Ohta, T., Yu, D.S., McHugh, P.J., Hickson, I.D., *et al.* (2005). A FancD2-Monoubiquitin Fusion Reveals Hidden Functions of Fanconi Anemia Core Complex in DNA Repair. Mol Cell *19*, 841-847.

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*., et al.* (2003). A novel ubiquitin ligase is deficient in Fanconi anemia. Nat Genet *35*, 165-170.

Meetei, A. R., Medhurst, A. L., Ling, C., Xue, Y., Singh, T. R., Bier, P., Steltenpool, J., Stone, S., Dokal, I., Mathew, C. G*., et al.* (2005). A human ortholog of archaeal DNA repair protein Hef is defective in Fanconi anemia complementation group M. Nat Genet *37*, 958-963.

Mortensen, U. H., Bendixen, C., Sunjevaric, I., and Rothstein, R. (1996). DNA strand annealing is promoted by the yeast Rad52 protein. Proc Natl Acad Sci U S A *93*, 10729-10734.

Myung, K., Datta, A., Chen, C., and Kolodner, R.D. (2001). SGS1, the Saccharomyces cerevisiae homologue of BLM and WRN, suppresses genome instability and homeologous recombination. Nat Genet *27*, 113-116.

Nikitina, T., Shi, X., Ghosh, R.P., Horowitz-Scherer, R.A., Hansen, J.C., and Woodcock, C.L. (2007). Multiple modes of interaction between the methylated DNA binding protein MeCP2 and chromatin. Mol Cell Biol *27*, 864-877.

Ralf, C., Hickson, I.D., and Wu, L. (2006). The Bloom's syndrome helicase can promote the regression of a model replication fork. J Biol Chem *281*, 22839-22846.

Smith, S., Hwang, J.Y., Banerjee, S., Majeed, A., Gupta, A., and Myung, K. (2004). Mutator genes for suppression of gross chromosomal rearrangements identified by a genome-wide screening in Saccharomyces cerevisiae. Proc Natl Acad Sci U S A *101*, 9039-9044.

Sogo, J., Stasiak, A., De Bernadin, W., Losa, R., and Koller, T. (1987). Binding proteins to nucleic acids as studies by electron microscopy. In Electron Miscroscopy in Molecular Biology, J.S.a.U. Scheer, ed. (IRL Press, Oxford), pp. 61-79.

Sonoda, E., Sasaki, M.S., Morrison, C., Yamaguchi-Iwai, Y., Takata, M., and Takeda, S. (1999). Sister chromatid exchanges are mediated by homologous recombination in vertebrate cells. Mol Cell Biol *19*, 5166-5169.

Sugiyama, T., New, J. H., and Kowalczykowski, S. C. (1998). DNA annealing by RAD52 protein is stimulated by specific interaction with the complex of replication protein A and single-stranded DNA. Proc Natl Acad Sci U S A *95*, 6049-6054.

Wang, X., Peterson, C.A., Zheng, H., Nairn, R.S., Legerski, R.J., and Li, L. (2001). Involvement of nucleotide excision repair in a recombination- independent and error-prone pathway of DNA interstrand cross-link repair. Mol Cell Biol *21*, 713-720.

Yan, Z., Wang, Z., Sharova, L., Sharov, A.A., Ling, C., Piao, Y., Aiba, K., Matoba, R., Wang, W., and Ko, M.S. (2008). BAF250B-associated SWI/SNF chromatin-remodeling complex is required to maintain undifferentiated mouse embryonic stem cells. Stem Cells *26*, 1155-1165.

Zhao, J., Kennedy, B. K., Lawrence, B. D., Barbie, D. A., Matera, A. G., Fletcher, J. A., and Harlow, E. (2000). NPAT links cyclin E-Cdk2 to the regulation of replication-dependent histone gene transcription. Genes Dev *14*, 2283-2297.

3.   Results

3.1   FA Candidate Genes

3.1.1   A Histone-Fold Complex and FANCM Form a Conserved DNA-Remodeling Complex to Maintain Genome Stability

3.1.2   On the Role of FAN1 in Fanconi Anemia

3.2   Identification of Novel FA Genes

3.2.1   Mutation of the RAD51C Gene in a Fanconi Anemia-like disorder

3.2.2   SLX4, a Coordinator of Structure-Specific Endonucleases, is Mutated in a New Fanconi Anemia Subtype

3.2.3   XPF Mutations Severely Disrupting DNA Interstrand Crosslink Repair Cause Fanconi Anemia

3.3   Genotyping Fanconi Anemia by Next Generation Sequencing

3.3.1   Whole Exome Sequencing Reveals Novel Mutations in the Recently Identified Fanconi Anemia Gene SLX4/FANCP

3.3.2   Genotyping Fanconi Anemia by Whole Exome Sequencing: Advantages and Challenges

Brief report

On the role of FAN1 in Fanconi anemia

*Juan P. Trujillo,[1,2] *Leonardo B. Mina,[1,2] Roser Pujol,[1,2] Massimo Bogliolo,[1,2] Joris Andrieux,[3] Muriel Holder,[4] Beatrice Schuster,[5] Detlev Schindler,[5] and Jordi Surrallés[1,2]

[1]Genome Instability and DNA Repair Group, Department of Genetics and Microbiology, Universitat Autònoma de Barcelona, Bellaterra, Barcelona, Spain; [2]Centre for Biomedical Network Research on Rare Diseases, Instituto de Salud Carlos III, Bellaterra, Barcelona, Spain; [3]Institut de Génétique Médicale, Hopital Jeanne de Flandre, Centre Hospitalier Regional Universitaire de Lille, Lille, France; [4]Génétique Clinique, Hôpital Jeanne de Flandre, Centre Hospitalier Regional Universitaire de Lille, Lille, France; and [5]Department of Human Genetics, University of Wurzburg, Wurzburg, Germany

Fanconi anemia (FA) is a rare bone marrow failure disorder with defective DNA interstrand crosslink repair. Still, there are FA patients without mutations in any of the 15 genes individually underlying the disease. A candidate protein for those patients, FA nuclease 1 (FAN1), whose gene is located at chromosome 15q13.3, is recruited to stalled replication forks by binding to monoubiquitinated FANCD2 and is required for interstrand crosslink repair, suggesting that mutation of *FAN1* may cause FA. Here we studied clinical, cellular, and genetic features in 4 patients carrying a homozygous 15q13.3 microdeletion, including *FAN1* and 6 additional genes. Biallelic deletion of the entire *FAN1* gene was confirmed by failure of 3′- and 5′-PCR amplification. Western blot analysis failed to show FAN1 protein in the patients' cell lines. Chromosome fragility was normal in all 4 FAN1-deficient patients, although their cells showed mild sensitivity to mitomycin C in terms of cell survival and G_2 phase arrest, dissimilar in degree to FA cells. Clinically, there were no symptoms pointing the way to FA. Our results suggest that FAN1 has a minor role in interstrand crosslink repair compared with true FA genes and exclude FAN1 as a novel FA gene. (*Blood*. 2012;120(1):86-89)

Introduction

Fanconi anemia (FA) is characterized by chromosome breakage, congenital malformations, pancytopenia, and cancer susceptibility.[1] FA is a rare disease with a carrier frequency of 1:65 to 1:209.[2,3] FA cells are hypersensitive to DNA interstrand crosslinking (ICL) drugs, such as mitomycin C (MMC) and diepoxybutane (DEB), and the diagnostics relies on an excess chromosome fragility after in vitro exposing patients' cells to these agents. There are at least 15 independent FA subtypes, each resulting from mutation of a distinct FA gene.[4-7] However, a minority of FA patients remain unassigned, suggesting the existence of additional FA genes. Recently, 4 groups reported that FA nuclease 1 (FAN1) is a good candidate for a novel FA gene.[8-11] The reason is that FAN1 is recruited to stalled replication forks by binding to monoubiquitinated FANCD2, and its nuclease activity is required for ICL repair. Transient depletion of FAN1 in human transformed fibroblasts led to increased MMC-induced chromosome breakage rates. Consequently, all 4 groups suggested that *FAN1* mutations may cause FA.[8-11]

FAN1 maps to 15q13.3. Heterozygous 15q13.3 microdeletion has been associated with a variety of symptoms, including mental retardation, epilepsy, psychiatric disease, autism spectrum disorders, muscular hypotonia, and dysmorphic facial features. Penetrance of the microdeletion disorder is variable and encompasses severely affected patients to normal persons.[12] Apart from *FAN1*, 6 additional genes are located in 15q13.3 (ARHGAP11B, MTMR10, TRPM1, KLF13, OTUD7A, and CHRNA7). Here we studied 4 patients with homozygous 15q13.3 microdeletion[12,13] to clarify whether lack of FAN1 may lead to FA.

Methods

Clinical features and blood samples were obtained from 4 homozygous 15q13.3 microdeletion patients (MD1-MD4) all previously diagnosed by array comparative genomic hybridization and quantitative PCR. Two of these patients (MD1 and MD2) have been mentioned before.[12,13] Lack of FAN1 was confirmed at the gene level by PCR and at the protein level by Western blotting. The PCR primers used to amplify the 3′ and 5′ flanking regions of FAN1 were as follows: ex1 forward, 5′AGGGTTGTCTCCTCGT-TACAGGA3′; ex1 reverse, 5′GCTGAATCACTTTGGCCAGG3′; ex15 forward, 5′CTTCCTAAAACCTGCTGGAGG3′; and ex15 reverse, 5′AAT-GTACTGACCGTGTGCTCA3′. PCR, Western blot analysis, survival assays, and chromosome breakage assays were performed as described elsewhere.[3,14-18] FAN1-monospecific antibody was kindly provided by Dr John Rouse (Dundee, United Kingdom) and used at 1:500 dilution. A total of 27 genetically unassigned FA cell lines had previously been excluded from belonging to any of the reported 15 FA complementation groups. This study was ethically approved by the Universitat Autònoma de Barcelona Institutional Review Board. Informed consent was obtained from all families in accordance with the Declaration of Helsinki.

Results and discussion

Study of FA candidate genes may enable the final classification of unassigned FA patients. Four recent studies have proposed *FAN1* as a putative FA gene.[8-11] Here we studied 4 patients (MD1-MD4) with homozygous 15q13.3 microdeletion to clarify whether FAN1 deficiency leads to features consistent with FA. Two of these
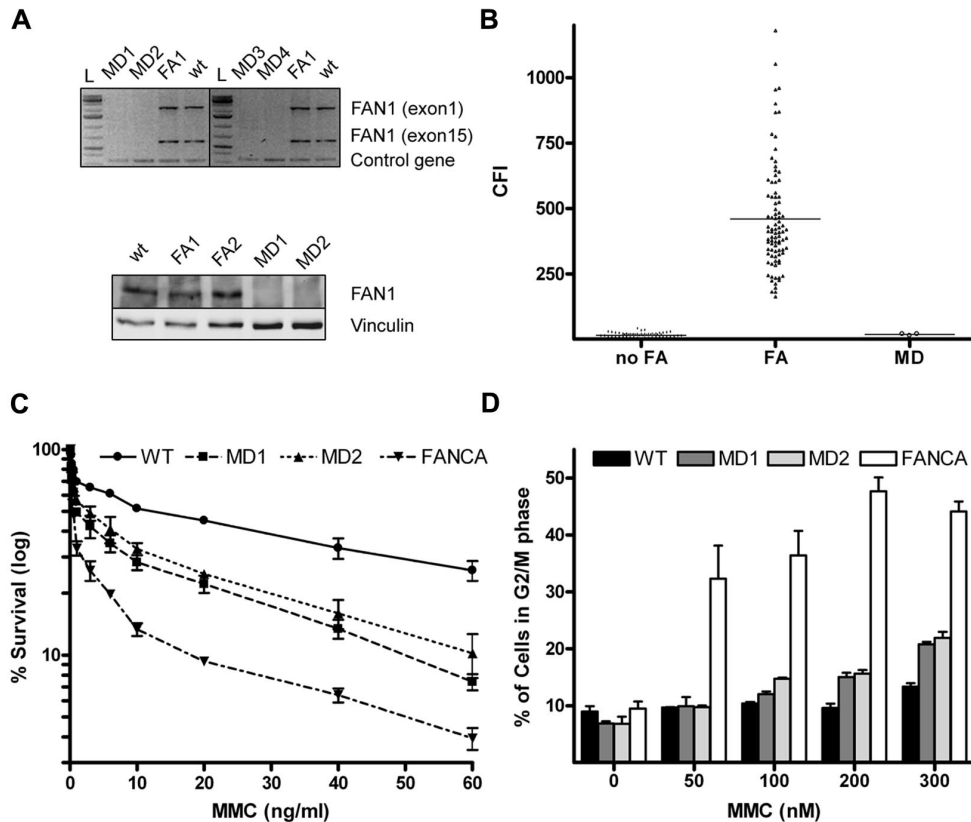
Figure 1. Absence of FAN1 and FA cellular phenotype in MD patients. (A) Fragments amplified by PCR corresponding to exons 1 and 15 of the *FAN1* gene were observed using DNA templates from a healthy person (WT) and a FANCA-deficient FA patient included as controls, whereas they were absent when templates from microdeletion patients (MD1 to MD4) were used, confirming the biallelic deletion of *FAN1* in the MD patients (top panel). Immunoblotting against FAN1 protein revealed a FAN1-specific band in WT and FA (FA1 and FA2) LCL that failed to be detected in MD1 and MD2. (bottom panel). (B) Dot plot of CFI showing individual values and average (solid line) of DEB-induced chromosome breakage from non-FA (n = 56), FA (n = 90, excluding mosaics) and MD (n = 4) persons. The CFI values of all MD patients ranged within the non-FA population. (C) Mild sensitivity of MD LCL to MMC on survival assay. The graph shows intermediate sensitivity to MMC of MD1 and MD2 compared with the highly sensitive FA (FANCA) cell line. (D) Near-normal sensitivity of MD LCL to MMC on cell cycle analysis. The graph plots the percentage of cells in G2/M phase after exposure to increasing concentrations of MMC for 72h. A WT and an FANCA cell lines were included as controls.

patients (MD1 and MD2) have previously been mentioned in unrelated reports[12,13] and the other 2 are newly recognized siblings, detected by array comparative genomic hybridization (supplemental Figure 1, available on the *Blood* Web site; see the Supplemental Materials link at the top of the online article). Considering that 15q13.3 microdeletion may have 3 different extensions,[19] we aimed to corroborate homozygous absence of the *FAN1* gene by failure to PCR amplify its first (exon 1) and last (exon 15) exon from genomic DNA. DNA of 2 additional lymphoblastoid cell lines (LCLs), from a normal person and from a FANCA-deficient patient, served as controls. As shown in Figure 1A top panel, FAN1 PCR products are absent in all MD patients, confirming that all 4 MD patients have biallelic deletion of the entire *FAN1* gene. Moreover, lack of FAN1 expression was confirmed by FAN1 immunoblotting. Clearly, the FAN1-specific band was missing in the 2 available LCLs from MD patients (MD1 and MD2), whereas FAN1 was readily detected in the control LCLs (Figure 1A bottom panel).

To check whether FAN1 deficiency leads to DEB-induced chromosome fragility,[20,21] we performed DEB tests on an LCL from patient MD1 and on blood T cells from patients MD2 to MD4. Chromosome breakage rates were quantified with the recently described chromosome fragility index (CFI)[17] and the results compared with our historical database.[17] Clearly, the CFI of all MD patients fell into the range of the non-FA group (Figure 1B). Similar results were obtained with MMC (data not shown).

We next tested the survival of the 2 available MD LCLs in response to MMC. MD1, MD2, a wild-type and a FANCA LCL were challenged with 0 to 100 ng/mL of MMC. Based on LD_{50} values, the MD cell lines showed mild sensitivity to MMC: whereas FANCA-deficient cells were more than 30-fold more sensitive to MMC than WT cells, MD1 and MD2 cells were, on average, 5-fold more sensitive to MMC than WT cells (Figure 1C). Silencing of the *FAN1* gene by siRNA was previously shown to impair ICL repair, leading to hypersensitivity of cells to ICL. However, this hypersensitivity was also intermediate compared with mRNA depletion of authentic FA genes, such as *FANCA*, *FANCD2*, or *FANCJ*.[8-10] This set of data suggests that the cellular response of FAN1-deficient cells to MMC is not fully functional but not impaired as in FA.

To further study the FA pathway in the absence of FAN1, cell cycle distributions of FAN1-deficient cell lines were analyzed by flow cytometry.[22] Exposure to increasing concentrations of MMC for 72 hours resulted in G_2 arrest at low MMC concentrations in FA-A LCL, whereas G_2 arrest was very mild in the MD samples (Figure 1D), compatible with the mild sensitivity to MMC shown

Table 1. Patient characteristics

Patient no.	Age, y	Nationality	Clinical features	Hematology	Chromosome fragility	Sensitivity to MMC	G₂/M block	Reference
MD1	11	United States	Visual impairment, hypotonia, areflexia, absent language, epilepsy, microsomy, and microcephaly	Normal	Negative	Mild	Mild	13
MD2	6	France	Hypotonia, severe developmental delay; rod–cone dystrophy, epilepsy, and autistic features	Normal	Negative	Mild	Mild	12
MD3	1	France	Severe developmental delay, visual impairment, microsomy, and microcephaly	Normal	Negative	ND	ND	Present study
MD4	3	France	Severe developmental delay, absent language, visual impairment, microsomy, and microcephaly	Normal	Negative	ND	ND	Present study

MMC indicates mitomycin C; and MD, microdeletion.

before. These results are consistent with a recent report on ΔFAN1-DT40 cells showing that FAN1 protects cells against ICL agents in a pathway, which is not epistatic with the FA pathway and that FAN1 assumes in the processing of ICL only a secondary role or functions independently of the FA pathway.[23] We finally analyzed FAN1 protein expression levels in 27 cell line from unassigned FA patients by Western blotting. All of the unassigned FA cell lines expressed FAN1 protein at control levels, suggesting that none of these patients had major deficiency of this protein (supplemental Figure 2).

To assess the hematologic impact of FAN1 deficiency, we obtained clinical data and hemograms of MD2, MD3, and MD4. Normal hematology had earlier been reported for MD1.[12] As shown in Table 1, MD patients do not present with anemia, bone marrow failure, skin pigmentation anomalies, or FA-typical malformations, such as skeletal abnormalities of the upper limbs. Three of the MD patients (MD1, MD3, and MD4) showed microsomy and microcephaly, which is often seen in FA patients but also in other syndromes with defective processing of stalled replication forks, such as Seckle and Bloom syndromes and can be regarded as common symptoms of patients with DNA repair defects.[24] Yet we cannot conclude for certain that microcephaly and microsomy found in MD patients are caused by FAN1 deficiency because 6 additional genes are included in the 15q13.3 region. However, it is tempting to speculate that this is the case as FAN1 directly interacts with FANCD2, and 90% of patients with FANCD2 mutations have microcephaly.[15]

Even though LCLs with total FAN1 deficiency reveal mild sensitivity to MMC on some assays, normal expression of FAN1 in 27 unassigned FA cell lines, the lack of DEB- or MMC-induced chromosome fragility, and the absence of hematologic defects or FA-archetypal malformations exclude FAN1 as being an FA gene.

Authorship

Contribution: J.P.T., L.B.M., R.P., M.B., and B.S. performed experiments and helped write the manuscript; J.A. and M.H. provided essential research materials and clinical data and performed experiments; D.S. designed experiments and provided essential research materials; and J.S. coordinated and supervised the study, designed experiments, and wrote the paper with the help of L.B.M.

Conflict-of-interest disclosure: The authors declare no competing financial interests.

Correspondence: Jordi Surrallés, Genome Instability and DNA Repair Group, Department of Genetics and Microbiology, Universitat Autònoma de Barcelona, Campus de Bellaterra S/N, 08193, Bellaterra, Barcelona, Spain; e-mail: jordi.surralles@uab.es.

References

1. Kutler DI, Singh B, Satagopan J, et al. A 20-year perspective on the International Fanconi Anemia Registry (IFAR). *Blood.* 2003;101(4):1249-1256.

2. Rosenberg PS, Tamary H, Alter BP. How high are carrier frequencies of rare recessive syndromes? Contemporary estimates for Fanconi anemia in the United States and Israel. *Am J Med Genet A.* 2011;155A(8):1877-1883.

3. Callén E, Casado JA, Tischkowitz MD, et al. A common founder mutation in FANCA underlies the world's highest prevalence of Fanconi anemia in Gypsy families from Spain. *Blood.* 2005;105(5):1946-1949.

4. Wang W. Emergence of a DNA-damage response network consisting of Fanconi anaemia and BRCA proteins. *Nat Rev Genet.* 2007;8(10):735-748.

5. Meindl A, Hellebrand H, Wiek C, et al. Germline mutations in breast and ovarian cancer pedigrees establish RAD51C as a human cancer susceptibility gene. *Nat Genet.* 2010;42(5):410-414.

6. Meetei AR, Medhurst AL, Ling C, et al. A human ortholog of archaeal DNA repair protein Hef is defective in Fanconi anemia complementation group M. *Nat Genet.* 2005;37(9):958-963.

7. Kim Y, Lach FP, Desetty R, Hanenberg H, Auerbach AD, Smogorzewska A. Mutations of the SLX4 gene in Fanconi anemia. *Nat Genet.* 2011;43(2):142-146.

8. Kratz K, Schopf B, Kaden S, et al. Deficiency of FANCD2-associated nuclease KIAA1018/FAN1 sensitizes cells to interstrand crosslinking agents. *Cell.* 2010;142(1):77-88.

9. Liu T, Ghosal G, Yuan J, Chen J, Huang J. FAN1 acts with FANCI-FANCD2 to promote DNA interstrand cross-link repair. *Science.* 2010;329(5992):693-696.

10. MacKay C, Declais AC, Lundin C, et al. Identification of KIAA1018/FAN1, a DNA repair nuclease recruited to DNA damage by monoubiquitinated FANCD2. *Cell.* 2010;142(1):65-76.

11. Smogorzewska A, Desetty R, Saito TT, et al. A genetic screen identifies FAN1, a Fanconi anemia-associated nuclease necessary for DNA

interstrand crosslink repair. *Mol Cell.* 2010;39(1): 36-47.

12. Masurel-Paulet A, Andrieux J, Callier P, et al. Delineation of 15q13.3 microdeletions. *Clin Genet.* 2010;78(2):149-161.

13. Lepichon JB, Bittel DC, Graf WD, Yu S. A 15q13.3 homozygous microdeletion associated with a severe neurodevelopmental disorder suggests putative functions of the TRPM1, CHRNA7, and other homozygously deleted genes. *Am J Med Genet A.* 2010;152A(5):1300-1304.

14. Bogliolo M, Lyakhovich A, Callen E, et al. Histone H2AX and Fanconi anemia FANCD2 function in the same pathway to maintain chromosome stability. *EMBO J.* 2007;26(5):1340-1351.

15. Kalb R, Neveling K, Hoehn H, et al. Hypomorphic mutations in the gene encoding a key Fanconi anemia protein, FANCD2, sustain a significant group of FA-D2 patients with severe phenotype. *Am J Hum Genet.* 2007;80(5):895-910.

16. Raya A, Rodriguez-Piza I, Guenechea G, et al. Disease-corrected haematopoietic progenitors from Fanconi anaemia induced pluripotent stem cells. *Nature.* 2009;460(7251):53-59.

17. Castella M, Pujol R, Callen E, et al. Chromosome fragility in patients with Fanconi anaemia: diagnostic implications and clinical impact. *J Med Genet.* 2011;48(4):242-250.

18. Castella M, Pujol R, Callen E, et al. Origin, functional role, and clinical impact of Fanconi anemia FANCA mutations. *Blood.* 2011;117(14):3759-3769.

19. Liao J, DeWard SJ, Madan-Khetarpal S, Surti U, Hu J. A small homozygous microdeletion of 15q13.3 including the CHRNA7 gene in a girl with a spectrum of severe neurodevelopmental features. *Am J Med Genet A.* 2011;155A(11):2795-2800.

20. Antonio Casado J, Callen E, Jacome A, et al. A comprehensive strategy for the subtyping of patients with Fanconi anaemia: conclusions from the Spanish Fanconi Anemia Research Network. *J Med Genet.* 2007;44(4):241-249.

21. Auerbach AD. Diagnosis of Fanconi anemia by diepoxybutane analysis. *Curr Protoc Hum Genet.* 2003;Chapter 8:Unit 8.7.

22. Ceccaldi R, Briot D, Larghero J, et al. Spontaneous abrogation of the GDNA damage checkpoint has clinical benefits but promotes leukemogenesis in Fanconi anemia patients. *J Clin Invest.* 2011;121(1):184-194.

23. Yoshikiyo K, Kratz K, Hirota K, et al. KIAA1018/ FAN1 nuclease protects cells against genomic instability induced by interstrand cross-linking agents. *Proc Natl Acad Sci U S A.* 2010;107(50): 21553-21557.

24. O'Driscoll M, Jeggo PA. The role of the DNA damage response pathways in brain development and microcephaly: insight from human disorders. *DNA Repair (Amst).* 2008;7(7):1039-1050.

Supplemental File

On the role of FAN1 in Fanconi anemia

Juan P. Trujillo, Leonardo B. Mina, Roser Pujol, Massimo Bogliolo, Joris Andrieux, Muriel Holder, Beatrice Schuster, Detlev Schindler, and Jordi Surralles

Figure S1. Homozygous 15q13.3 of siblings MD3 and MD4 siblings. 44K array-CGH profile, showing the extent of homozygous 15qq13.3 deletion (left panel) involving chr15:28,725,507-30,298,155 (hg18-build36) and genes included in the deleted region (right panel). FAN1 is identical with KIAA1018.

3. RESULTS

3.1 FA CANDIDATE GENES

3.1.1 A HISTONE-FOLD COMPLEX AND FANCM FORM A CONSERVED DNA-REMODELING COMPLEX TO MAINTAIN GENOME STABILITY

3.1.2 ON THE ROLE OF FAN1 IN FANCONI ANEMIA

3.2 IDENTIFICATION OF NOVEL FA GENES

3.2.1 MUTATION OF THE RAD51C GENE IN A FANCONI ANEMIA-LIKE DISORDER

3.2.2 SLX4, A COORDINATOR OF STRUCTURE-SPECIFIC ENDONUCLEASES, IS MUTATED IN A NEW FANCONI ANEMIA SUBTYPE

3.2.3 XPF MUTATIONS SEVERELY DISRUPTING DNA INTERSTRAND CROSSLINK REPAIR CAUSE FANCONI ANEMIA

3.3 GENOTYPING FANCONI ANEMIA BY NEXT GENERATION SEQUENCING

3.3.1 WHOLE EXOME SEQUENCING REVEALS NOVEL MUTATIONS IN THE RECENTLY IDENTIFIED FANCONI ANEMIA GENE SLX4/FANCP

3.3.2 GENOTYPING FANCONI ANEMIA BY WHOLE EXOME SEQUENCING: ADVANTAGES AND CHALLENGES

nature genetics

Mutation of the *RAD51C* gene in a Fanconi anemia–like disorder

Fiona Vaz[1,10], Helmut Hanenberg[2,3,10], Beatrice Schuster[4], Karen Barker[5], Constanze Wiek[2], Verena Erven[2], Kornelia Neveling[4], Daniela Endt[4], Ian Kesterton[6], Flavia Autore[7], Franca Fraternali[7], Marcel Freund[2], Linda Hartmann[8], David Grimwade[1], Roland G Roberts[1], Heiner Schaal[8], Shehla Mohammed[9], Nazneen Rahman[5], Detlev Schindler[4,11] & Christopher G Mathew[1,11]

Fanconi anemia (FA) is a rare chromosomal-instability disorder associated with a variety of developmental abnormalities, bone marrow failure and predisposition to leukemia and other cancers[1]. We have identified a homozygous missense mutation in the *RAD51C* gene in a consanguineous family with multiple severe congenital abnormalities characteristic of FA. *RAD51C* is a member of the RAD51-like gene family involved in homologous recombination–mediated DNA repair. The mutation results in loss of RAD51 focus formation in response to DNA damage and in increased cellular sensitivity to the DNA interstrand cross-linking agent mitomycin C and the topoisomerase-1 inhibitor camptothecin. Thus, biallelic germline mutations in a RAD51 paralog are associated with an FA-like syndrome.

FA is a highly heterogeneous disorder, arising from biallelic mutations in one of at least 13 different genes (*FANCA, FANCB, FANCC, BRCA2 (FANCD1), FANCD2, FANCE, FANCF, FANCG, FANCI, BRIP1 (FANCJ), FANCL, FANCM* and *PALB2 (FANCN)*; ref. 1). A key step in the FA pathway is monoubiquitination of FANCD2 and FANCI, which requires the presence of a core complex of FA and FA-associated proteins. This post-translational modification is intact in FA groups D1, J and N, which therefore appear to function downstream of the core and I-D2 complexes[1]. Most individuals diagnosed with FA have germline defects in one of the known FA genes. However, identification of the causal genetic defects in the small minority of individuals with unclassified FA may offer important insights into the function of the FA pathway in inherited disorders and cancer.

The pedigree and genetic analysis of the family investigated in this study is shown in **Figure 1** and clinical and laboratory details are summarized in **Table 1**. The parents of the affected children are first cousins of Pakistani origin (**Fig. 1a**). A daughter (IV-3) died at

2 months of age with multiple congenital abnormalities, including one absent and one vestigial thumb, a congenital heart defect, imperforate anus and hydronephrosis. Her lymphocytes showed elevated chromosome breakage after treatment with the DNA interstrand cross-linking (ICL) agent mitomycin C (MMC), indicating a diagnosis of FA. A son (IV-2) died 2 d after birth with congenital abnormalities similar
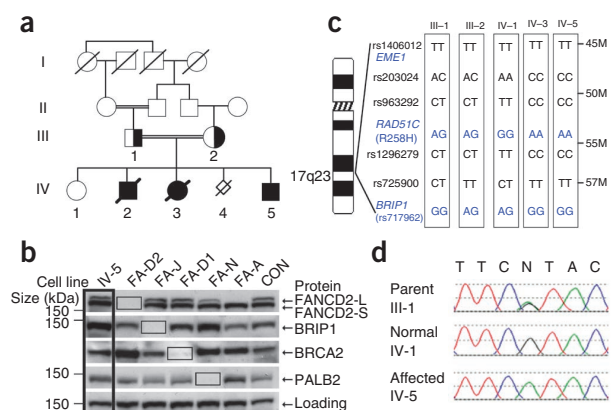
Figure 1 Genetic analysis of the family carrying an FA-like disorder. (**a**) Pedigree of the family. (**b**) Protein blot of fibroblasts from subject IV-5 and of other FA cell lines (FA-D2, FA-J, FA-D1, FA-N and FA-A denote lines with mutations in *FANCD2, BRIP1, BRCA2, PALB2* and *FANCA*, respectively); ubiquitinated FANCD2 (FANCD2-L) and nonubiquitinated FANCD2 (FANCD2-S) and other downstream FANC proteins are present in cells from IV-5. CON, unaffected control; loading control is RAD50. (**c**) Linkage analysis with SNPs on chromosome 17q23. M, megabases. The first fully informative SNP rs203024 is distal to *EME1*. (**d**) Sequencing of the mutation in *RAD51C* (G773A) in family members.

[1]Department of Medical and Molecular Genetics, King's College London School of Medicine, Guy's Hospital, London, UK. [2]Department of Pediatric Hematology, Oncology and Clinical Immunology, Children's Hospital, Heinrich Heine University, Düsseldorf, Germany. [3]Department of Pediatrics, Wells Center for Pediatric Research, Riley Hospital, Indiana University School of Medicine, Indianapolis, Indiana, USA. [4]Department of Human Genetics, University of Würzburg, Würzburg, Germany. [5]Section of Cancer Genetics, Institute of Cancer Research, Sutton, Surrey, UK. [6]Cytogenetics Laboratory, Genetics Centre, Guy's and St. Thomas' NHS Foundation Trust, Guy's Hospital, London, UK. [7]Randall Division of Cell and Molecular Biophysics, King's College London, New Hunt's House, Guy's Hospital, London, UK. [8]Institute of Virology, Heinrich Heine University, Düsseldorf, Germany. [9]Department of Clinical Genetics, Guy's and St. Thomas' NHS Foundation Trust, Guy's Hospital, London, UK. [10]These authors contributed equally to this work. [11]These authors contributed equally to the direction of this work. Correspondence should be addressed to C.G.M. (christopher.mathew@genetics.kcl.ac.uk), H.H. (helmut.hanenberg@uni-duesseldorf.de) or D.S. (schindler@biozentrum.uni-wuerzburg.de).

LETTERS

Table 1 Clinical and laboratory data for siblings family

Sibling gender	Clinical phenotype	Status	Chromosome breakage[a]
IV-1 Female	No features of FA	Well, aged 22 years	Blood lymphocytes IV-1: 0.01 (Sp); 0.03 (MMC) Control: 0.02 (Sp); 0.06 (MMC) Conclusion: normal
IV-2 Male	Intestinal, anal and respiratory abnormalities	Died age 2 d	Not done
IV-3 Female	Absent and vestigial thumb Severe congenital heart disease Hydronephrosis Imperforate anus	Died age 2 months	Blood lymphocytes IV-3: 0.07 (Sp); 0.90 (MMC); 17 multiradials in 80 metaphases Control: 0.02 (sp); 0.06 (MMC) Conclusion: elevated
IV-4	Unknown	Miscarriage at 11 weeks	Not done
IV-5 Male	Bilateral radial hypoplasia Long, slim fingers with proximally placed thumbs Hypoplastic thenar eminences Bilateral cystic kidneys Duodenal web Anal/rectal atresia Mild hypothyroidism Undescended testes Small genitalia	10 years old	Blood lymphocytes 0.12 (Sp); 0.38 (DEB); 5 multiradials in 100 metaphases Control: 0.0 (Sp), 0.04 (DEB) Fibroblasts IV-5: 0.15 (Sp); 1.24 (MMC) Control: 0.03 (Sp); 0.09 (MMC) Conclusion: elevated

[a]Mean breaks per cell occurring either spontaneously (Sp) or induced by MMC or DEB. Rates of multiradials were observed in stressed cultures.

to IV-3, and a fourth pregnancy (IV-4) miscarried at 11 weeks. The youngest child (IV-5), now aged 10 years, has extensive congenital abnormalities including short stature, bilateral radial hypoplasia, anal atresia, bilateral cryptorchidism, small genitalia, bilateral cystic dysplasia of the kidneys and chronic renal failure. Chromosome breakage testing of primary cultured fibroblasts showed clearly elevated breakage after exposure to MMC and diepoxybutane (DEB; **Fig. 2a** and **Table 1**), and cell cycle analysis of primary lymphocytes and cultured fibroblasts showed pronounced arrest in G2 after MMC treatment (**Fig. 2b,c**). The diagnosis of FA in this family was based on the presence of characteristic congenital abnormalities and the elevated sensitivity of cells from the siblings IV-3 and IV-5 to ICL agents. The only surviving affected child (IV-5) has not developed hematological abnormalities or cancer by age 10 years. In the absence of hematological symptoms thus far, we will refer to the clinical phenotype in this family as a Fanconi anemia–like disorder. However, the age of onset of these features in FA is variable, and cumulative incidence data from North American and German Registries indicate that by age 10 only 30–35% of individuals with FA have bone marrow failure, and less than 3% have developed leukemia or solid tumors[2].

Transduction of primary fibroblasts from IV-5 (SH2038-F) with retroviral vectors containing the *FANCA*, *FANCB*, *FANCC*, *FANCE*, *FANCF*, *FANCG* or *FANCL* complementary DNAs did not complement G2 arrest (data not shown). Protein blotting showed that FANCD2 and its monoubiquitinated form were present, suggesting a downstream defect in the FA pathway, and also confirmed the presence of the BRCA2, BRIP1 and PALB2 proteins (**Fig. 1b**). Sequencing of the exons and adjacent splice sites of the *BRIP1*, *BRCA2* and *PALB2* genes detected only known noncoding or synonymous polymorphisms. These data indicated that the family was likely to have a mutation in the FA/BRCA pathway in a previously unrecognized gene that functions downstream of the FA core and I-D2 complexes.

We genotyped DNA from the parents, the unaffected daughter (IV-1), and the two affected children (IV-3 and IV-5) on the Affymetrix 10K SNP array to search for candidate regions containing the causal gene by autozygosity mapping (Online Methods). The affected siblings showed seven substantial regions of shared homozygosity (>3 Mb) for

which the unaffected sibling either was heterozygous or was homozygous for the opposite allele (**Supplementary Table 1**). The largest of these was a 14.6-Mb region on chromosome 17q21–q24 that was of particular interest (**Fig. 1c**), as it contained three genes involved in DNA repair (*EME1*, *RAD51C* (*RAD51L2*) and *BRIP1*) for which loss of function is associated with hypersensitivity to DNA interstrand cross-linking agents[3,4]. Sequencing of the coding regions of *EME1* in the index case (IV-5) revealed heterozygosity for multiple common intragenic polymorphisms, consistent with the fact that *EME1* is actually located 1.6 Mb proximal to the first fully informative SNP in this region (rs203024; see **Fig. 1c**). The region of linkage includes *BRIP1* (haplotype formed by SNPs rs725900 and rs717962), but this gene had already been excluded through protein blotting and sequencing (see above). Sequencing of the nine exons and splice sites of the *RAD51* paralog *RAD51C* in IV-5 revealed homozygosity for a mutation (G773A) in exon 5, which results in the amino acid substitution R258H (**Fig. 1d**). The other affected sibling (IV-3) was also homozygous for this mutation; the parents were heterozygous, and the child without FA-like abnormalities and with normal ICL sensitivity (IV-1) was homozygous for the wild-type sequence. The mutation was not present in 47 regionally and ethnically matched controls from the Lahore region of Pakistan. Sequencing of *RAD51C* coding regions in 21 subjects with FA excluded from the known complementation groups with intact FANCD2 monoubiquitination did not detect any additional mutations. The potential functional importance of the R258H mutation is supported by the fact that the arginine residue is conserved in RAD51C proteins from a wide range of species, including chicken, zebrafish, sea urchin and thale cress, and is also conserved in RAD51 itself and in two of the RAD51-like proteins, RAD51B (RAD51L1) and XRCC3 (**Supplementary Fig. 1**).

To determine whether R258H was the causal mutation in this family, we transduced primary fibroblasts from subject IV-5 (SH2038-F) with a retroviral vector containing wild-type *RAD51C* cDNA, the *RAD51C*$_{G773A}$ mutant, or a control vector transferring either the neomycin phosphotransferase II (*nptII*) or the puromycin N-acetyl-transferase (*pac*) gene, selected the cells in G418 or puromycin and analyzed cell cycle distributions after exposure of cells to 36 nM MMC for 48 h (Online Methods and **Supplementary Fig. 2**). The results showed that G2 arrest of the affected individual's fibroblasts was rescued specifically by expression of the wild-type *RAD51C* (**Fig. 2c,d**) but not by vector-mediated overexpression of the *RAD51C*$_{G773A}$ mutant cDNA (**Fig. 2e**). We also tested the effect of the mutation by expressing the wild-type or mutant RAD51C in two other eukaryotic cell lines that are deficient in the RAD51C protein. The hamster cell line irs3 has a splice-site mutation in *Rad51c* that causes skipping of exon 6 (ref. 5). We performed cell cycle analysis after MMC challenge of irs3 cells that had been transduced with retroviral vectors containing either the human wild-type or the mutant *RAD51C*$_{G773A}$ cDNA. The results (**Fig. 2f–h**) demonstrated that whereas the wild-type protein restored MMC resistance in irs3 cells, expression of the mutant protein resulted in only a modest degree of correction of cross-linker sensitivity when compared to transduction with the control vector. This suggests that R258H is a hypomorphic mutant, at least in the context of this cell line
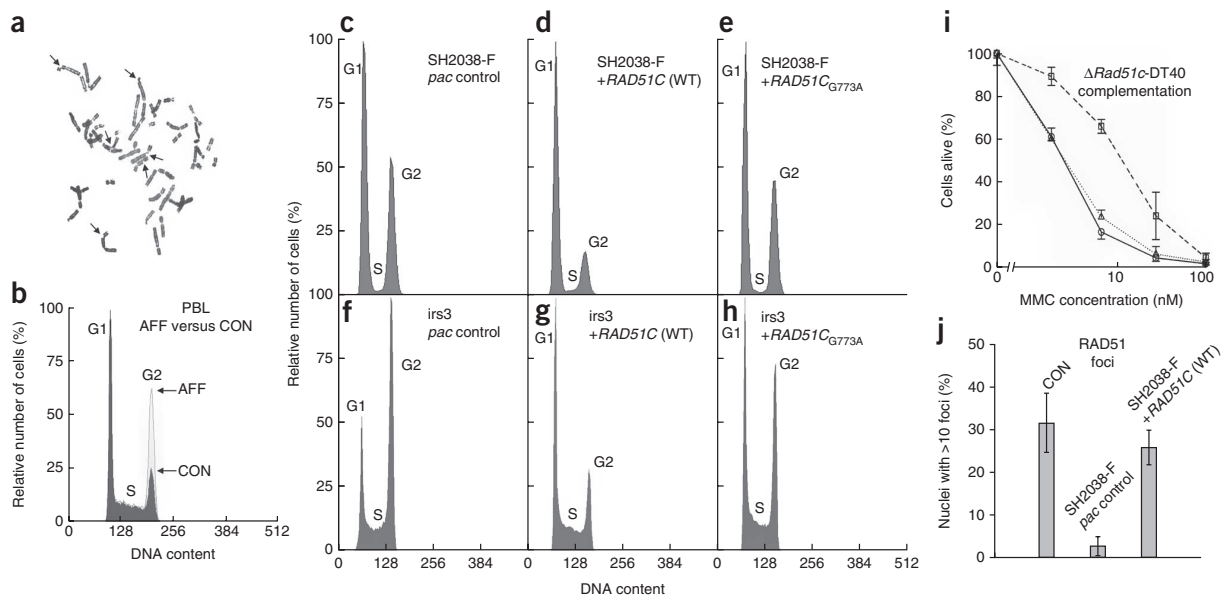
Figure 2 Functional analysis of the RAD51C alleles. (**a**) Fibroblast metaphase after exposure of the culture to 10 ng/ml MMC shows chromatid-type breakage and a radial rejoining figure (arrows). (**b**) Cell cycle distribution of peripheral blood lymphocyte culture from subject IV-5 shows increased G2 arrest (37.4% of cells in G2) in response to 45 nM MMC (AFF, light gray) compared to an unaffected control (18.4%; CON, dark gray overlay). (**c**) Untransduced (not shown) or *pac* (mock)-transduced fibroblasts from subject IV-5 (SH2038-F) show elevated G2 phase arrest ($47.3 \pm 8.5\%$, $n = 4$ experiments) after exposure to 36 nM MMC. (**d**) Transduction of SH2038-F cells with wild-type (WT) *RAD51C* rescues G2 arrest ($22.6 \pm 0.6\%$, $n = 4$, $P < 0.005$) under the same conditions as in **c**. (**e**) Transduction with mutant $RAD51C_{G773A}$ leads to a marginal decrease in G2 arrest ($41.5 \pm 2.3\%$, $n = 4$, not significant compared to mock transduction, $P = 0.24$). (**f**) Untransduced (not shown) or *pac*-transduced Rad51c-deficient hamster irs3 cells show elevated G2 phase arrest ($42.5 \pm 3.5\%$, $n = 3$) after exposure to 36 nM MMC. (**g**) Transduction of irs3 cells with human wild-type *RAD51C* rescues G2 arrest ($16.4 \pm 0.6\%$, $n = 3$, $P < 0.001$) under the same conditions as in **f**. (**h**) Transduction with mutant $RAD51C_{G773A}$ leads to a moderate decrease in G2 arrest ($30.5 \pm 2.7\%$, $n = 3$, $P < 0.005$) compared to mock transduction. (**i**) Increased survival rates indicate successful complementation of $\Delta RAD51C$-DT40 chicken cells by human wild-type *RAD51C* (dashed line) but not by the mutant $RAD51C_{G773A}$ (dotted line); the solid line is from uncorrected cells transduced with control vector. Error bars are s.d. from three experiments. (**j**) SH-2038 fibroblasts from subject IV-5 are defective in the formation of RAD51 nuclear foci after exposure to MMC ($2.8 \pm 2.2\%$ positive cells, $n = 4$). Transduction with wild-type *RAD51C* rescued their proficiency ($25.9 \pm 4.0\%$ positive cells, $n = 4$, $P < 0.001$) to a degree that was similar to unaffected control (CON) fibroblasts ($31.6 \pm 6.9\%$ positive cells). Error bars, s.d.

and the cell cycle assay. The effect of the mutation was also tested in chicken $\Delta RAD51C$-DT40 cells in which the *RAD51C* ortholog is disrupted by recombination[6]. Expression of human wild-type RAD51C complemented the sensitivity of *RAD51C*-mutant cells to MMC, whereas expression from the human *RAD51C* cDNA with the G773A mutation did not (**Fig. 2i**). The specificity of correction of the MMC-sensitive phenotype was confirmed further by the finding that expression of wild-type *RAD51C* cDNA in 18 FA cell lines from different upstream and downstream complementation groups did not rescue the characteristic G2 arrest in any of these cell lines (data not shown). These data verify that *RAD51C* is the gene responsible for the cellular phenotype in this family and that R258H is the causal mutation. We suggest a provisional assignment for the FA in this family as FA-O.

We investigated the possible structural effect of the R258H mutation by homology-modeling the structure of RAD51C from the crystal structure of the archaeal rad51 protein of *Pyrococcus furiosus*[7] (PDB 1PZN; Online Methods). The arginine residue 258 is located on helix α13 in close proximity to helix α12 and faces residues Ser304 and Glu303 of the loop-connecting strands β5 and β6 (ref. 7). In the wild-type protein, Arg258 is in close contact (hydrogen-bond distance) with the carbonyl backbone of Glu303. Substitution of arginine with histidine in the model disrupts this interaction (**Supplementary Fig. 3**) and changes the electrostatic surface of this region from slightly positive to slightly negative. Together, these structural perturbations could result in rearrangement of the neighboring secondary structure

elements, with relative displacement of the N-terminal and ATPase domains, and thus may affect higher-order structures such as the heptameric ring described for archaeal rad51 (ref. 7). The mutation does not appear to have a major effect on the stability of the protein, as the protein was readily detectable in the affected individual's fibroblasts on protein blots (data not shown).

As the loss of functional RAD51C protein in eukaryotic cells is associated with impaired formation of RAD51 foci in response to DNA damage[4–6], we looked for this phenotype in cells from subject IV-5. We found that RAD51 focus formation in response to MMC treatment was greatly reduced in the affected individual's fibroblasts and that this defect was corrected by transduction with wild-type *RAD51C* (**Fig. 2j**). This result strongly supports the genetic and functional complementation data that identified loss of RAD51C function as the primary defect in this family. The effect of the *RAD51C* mutation on RAD51 focus formation is shared with two of the downstream FA complementation groups, FA-D1 and FA-N, caused by mutations in *BRCA2* and *PALB2*, respectively, which encode two proteins that themselves have important roles in homologous recombination–based DNA repair[8]. The loss of RAD51 focus formation in response to interstrand cross-link–induced DNA damage prompted us to test fibroblasts from subject IV-5 for sensitivity to irradiation. As found in other FA complementation groups, the cells showed only modest radiosensitivity, which was, however, complemented by transduction with wild-type *RAD51C* (**Supplementary Table 2**). We also tested lymphoblastoid cells

LETTERS
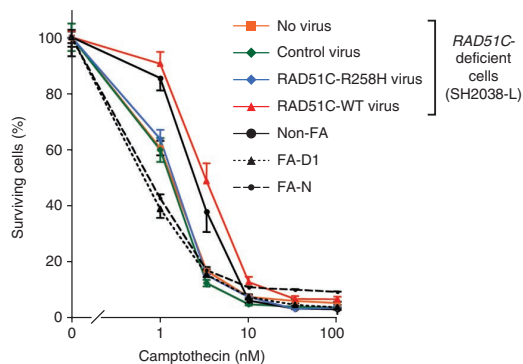
Figure 3 Camptothecin (CPT) sensitivity of lymphoblasts from subject IV-5. SH2038-L cells were tested for CPT sensitivity before transduction (no virus) or after transduction with control virus (no *RAD51C* insert), *RAD51C* with the R258H mutation, or wild-type *RAD51C*. Results for CPT-sensitive FA lymphoblastoid cell lines from FA groups FA-D1 and FA-N, and from a normal control line, are shown for comparison. The CPT sensitivity of SH2038-L is complemented by wild-type but not by mutant *RAD51C* (data shown is the mean plus or minus the one-sided s.d. from four experiments).

(SH2038-L) from the affected individual for sensitivity to the topoisomerase inhibitor camptothecin, as cells from other downstream FA groups, FA-D1 and FA-N (and FANCM-deficient cells), have recently been shown to be sensitive to this agent[9]. The *RAD51C*-deficient cells did show increased camptothecin sensitivity, which was corrected by wild-type *RAD51C* but not by cDNAs encoding the R258H mutant (**Fig. 3**). Although the camptothecin sensitivity was less marked in the affected individual's cells than in the FA-D1 and FA-N lines, these data are consistent with an FA core complex–independent role for RAD51C in addition to BRCA2, PALB2 and FANCM (ref. 9).

This study is, to our knowledge, the first report of the association of a mutation in *RAD51C* with a human disorder. The rarity of *RAD51C* mutations in humans is consistent with the fact that absence of *Rad51c* in mice causes early embryonic lethality, and there is partial embryonic lethality in *Rad51c*[ko/+] mice[10], whereas a hypomorphic *Rad51c* mouse model that expressed 5–30% of normal levels of the protein had normal growth and development but reduced fertility[11]. Thus the R258H mutation in this family may be associated with some residual function of RAD51C, as suggested by partial correction of cell cycle arrest in the hamster cell line irs3. Nonetheless, affected family members have experienced severe congenital abnormalities, some of which, such as imperforate anus and cystic kidneys with renal failure, have also been described in individuals with FA having bilateral mutations in *BRCA2* and *PALB2* (ref. 1). However, the absence of malignancies in subject IV-5 at the age of 10 years suggests that RAD51C deficiency may be associated with a less cancer-susceptible phenotype than that in the FA groups FA-D1 and FA-N.

RAD51C is known to have an important role in RAD51-mediated recombination[12,13] and recently has been shown to persist at sites of DNA damage after disassembly of RAD51 nucleoprotein filaments[14]. However, it also appears to be required for activation of the checkpoint kinase CHK2 and cell cycle arrest in response to DNA damage[14]. These characteristics suggest that *RAD51C* and the other *RAD51* paralogs may merit screening as candidate genes in families with FA-like disease and in familial cancers with undefined mutations. Indeed, in a companion paper in this issue, we demonstrate the presence of truncating and missense mutations in *RAD51C* in familial breast and ovarian cancer[15].

METHODS
Methods and any associated references are available in the online version of the paper at http://www.nature.com/naturegenetics/.

Note: Supplementary information is available on the Nature Genetics website.

1. Moldovan, G.-L. & D'Andrea, A.D. How the Fanconi anemia pathway guards the genome. *Annu. Rev. Genet.* **43**, 223–249 (2009).
2. Rosenberg, P.S. *et al.* Cancer risks in Fanconi anemia: findings from the German Fanconi Anemia Registry. *Haematologica* **93**, 511–517 (2008).
3. Abraham, J. *et al.* Eme1 is involved in DNA damage processing and maintenance of genomic stability in mammalian cells. *EMBO J.* **22**, 6137–6147 (2003).
4. Godthelp, B.C. *et al.* Mammalian Rad51C contributes to DNA cross-link resistance, sister chromatid cohesion and genomic stability. *Nucleic Acids Res.* **30**, 2172–2182 (2002).
5. French, C.A. *et al.* Role of mammalian RAD51L2 (RAD51C) in recombination and genetic stability. *J. Biol. Chem.* **277**, 19322–19330 (2002).
6. Takata, M. *et al.* Chromosome instability and defective recombinational repair in knockout mutants of the five Rad51 paralogs. *Mol. Cell. Biol.* **21**, 2858–2866 (2001).
7. Shin, D.S. *et al.* Full length archaeal Rad51 structure and mutants: mechanisms for RAD51 assembly and control by BRCA2. *EMBO J.* **22**, 4566–4576 (2003).
8. Livingston, D.M. Cancer. Complicated supercomplexes. *Science* **324**, 602–603 (2009).
9. Singh, T.R. *et al.* Impaired FANCD2 monoubiquitination and hypersensitivity to camptothecin uniquely characterize Fanconi anemia complementation group M. *Blood* **114**, 174–180 (2009).
10. Kuznetsov, S.G. *et al.* Loss of Rad51c leads to embryonic lethality and modulation of Trp53-dependent tumorigenesis in mice. *Cancer Res.* **69**, 863–872 (2009).
11. Kuznetsov, S. *et al.* RAD51C deficiency in mice results in early prophase I arrest in males and sister chromatid separation at metaphase II in females. *J. Cell Biol.* **176**, 581–592 (2007).
12. Liu, Y. *et al.* RAD51C is required for Holliday junction processing in mammalian cells. *Science* **303**, 243–246 (2004).
13. Liu, Y. *et al.* Role of RAD51C and XRCC3 in genetic recombination and DNA repair. *J. Biol. Chem.* **282**, 1973–1979 (2007).
14. Badie, S. *et al.* RAD51C facilitates checkpoint signaling by promoting CHK2 phosphorylation. *J. Cell Biol.* **185**, 587–600 (2009).
15. Meindl, A. *et al.* Germline mutations in breast and ovarian cancer pedigrees establish RAD51C as a human cancer susceptibility gene. *Nat. Genet.* advance online publication, doi:10.1038/ng.569 (18 April 2010).

ONLINE METHODS

Samples and cell lines. The family in question was referred to the Department of Clinical Genetics at Guy's Hospital. We took blood samples and skin biopsy specimens, after obtaining informed consent for all participants and ethical board approval (07/Q0702/69), to test for chromosome breakage, to obtain genomic DNA and to generate cell lines. Fibroblasts (SH2038-F) were grown from pieces of the skin explant, and a lymphoblast cell line (SH2038-L) was established by *ex vivo* EBV transformation of B lymphocytes. Since the family originated from the Lahore region of Pakistan, we obtained a panel of 47 control DNAs established by the Department of Pathology at the Army Medical College (Rawalpindi, Pakistan) to test for the presence of the *RAD51C* R258H mutation. ΔRad51c-DT40 cells were purchased from the Riken BRC Cell Bank. Rad51c-deficient irs3 cells were kindly provided from J. Thacker[5] by G. Illiakis.

Chromosome breakage test. Chromosome breakage analysis was performed according to established protocols[16]. Phytohemagglutinin-stimulated peripheral blood was cultured either with or without MMC at a final concentration of 30 ng/ml, or with or without DEB at a final concentration of 100 ng/ml, for 48–72 h in parallel with an identically treated healthy control sample. We analyzed the resulting metaphase spreads by light microscopy and scored them for chromosomal instability. We compared the affected individual's results with those of the control and established laboratory ranges. Cultured fibroblasts (SH2038-F) were tested similarly but using a 36-h exposure to MMC at a final concentration of 10 ng/ml.

Radiosensitivity assay. Confluent primary fibroblasts were trypsinized, centrifuged and resuspended at 2×10^5 per milliliter in MEM media. Aliquots were transferred into CryoTube vials and irradiated with 1, 1.5, 2, 3, 4 or 5 Gy using a 6-MV linear accelerator (Siemens) as photon source. We plated cells in triplicate into 60-mm Petri dishes at 500–6,000 cells per dish, depending on the irradiation dose applied. Earle's MEM with 15% (vol/vol) fetal bovine serum was replaced every 3–4 d. After a growth period of 2 weeks, colonies (>20 cells) were stained with 1% (wt/vol) crystal violet in 20% (vol/vol) ethanol. Clones were counted on a projection screen. Means and standard errors of the ratio of colony number relative to the number of seeded cells were calculated individually and plotted as colony survival fraction versus radiation dosage. We derived survival data from three separate experiments. Cell lines studied were RAD51C-deficient SH2038-F cells and their RAD51C-complemented isogenic counterparts; ataxia telangiectasia Aa026 (ref. 17), DNA-ligase IV-deficient GYMN (ref. 18) and RAD50-deficient F239 (ref. 19) cell lines were included as radiosensitive controls. We fitted dose-response curves to the linear quadratic model $SF = \exp(-aX - bX^2)$, where SF is the survival fraction, X the radiation dose, and a and b are fitted parameters. We did calculations using Origin 5.0 (MicroCal Software) and generated graphs using SigmaPlot 10 (Systat Software).

Camptothecin sensitivity of RAD51C-deficient cells. The lymphoblast cell line from the index individual (SH2038-L) was transduced with control virus or vectors expressing either the wild-type *RAD51C* or the mutant *RAD51C* R258H cDNAs. G418 resistant cells (3×10^5) were exposed to increasing doses (0, 1, 3.3, 10, 33.3 and 100 nM) of camptothecin (Sigma). After 5 d the cultures were harvested, stained with propidium iodide and analyzed on a FACSCalibur flow cytometer (BD Bioscience). For each data point, we collected 10^4 events using the CellQuest software. Results are shown for three or four different experiments as mean ± s.d. Survival of SH2038-L cells was compared with that of the PALB2-deficient (FA-N) LNEY cell line[20], the BRCA2-deficient (FA-D1) FA62 cell line[21] and a normal control, LCL (ref. 22).

Cell cycle analysis. We exposed native or transduced cells to 36 nM (12 ng/ml; fibroblasts, irs3 cells) or 45 nM (15 ng/ml; blood lymphocytes, lymphoblastoid cells) MMC for 48 h, harvested them and stained them with 4′-6-diamidino-2-phenylindole (DAPI) at a final concentration of 1 μg/ml in a buffer containing 154 mM NaCl, 1 mM CaCl₂, 0.5 mM MgCl₂, 0.1 M Tris, 0.2% (wt/vol) BSA and 0.1% (vol/vol) NP40 for 30 min in the dark. Alternatively, we fixed cells with 70% (vol/vol) methanol for at least 1 h at −20 °C, resuspended them in PBS and stained them with propidium iodide at a final concentration of 50 μg/ml while treating them with 25 U/ml RNase A for 30 min at 37 °C in the dark. We recorded univariate flow histograms on an analytical, triple-laser-equipped flow cytometer (LSRII, Becton Dickinson) using 355-nm Lightwave solid-state modelock laser excitation of the DAPI dye or sapphire 488-nm solid-state laser excitation of propidium iodide. We quantified the resulting cell cycle distributions, reflecting cellular DNA content, using the MPLUS AV software package (Phoenix Flow Systems).

Phenotypic correction of FA cells by retroviral transduction. The control vector S11IN, expressing an IRES-*nptII* cassette, and the vector S11RCIN, expressing additionally wild-type *RAD51C* cDNA, were constructed using methods previously described (ref. 23 and **Supplementary Fig. 2**). We generated the S11IP and the S11RCIP vectors by replacing neomycin phosphotransferase II (*nptII*) with puromycin N-acetyl-transferase (*pac*) cDNA using standard procedures. The missense mutation encoding the G773A substitution was introduced using the QuikChange Site-Directed Mutagenesis kit (Stratagene) according to the manufacturer's instructions. Generation of stable oncoretroviral cell lines and transduction of adherent and nonadherent cells were performed as described[22,24,25]. Transduced cells were selected in G418 or puromycin for 7–14 d, challenged with MMC and then assayed by flow cytometry. We analyzed transduced human fibroblasts and irs3 cells by cell cycle analysis for rescue of G2-phase arrest as described above. We analyzed ΔRad51c-DT40 cells for survival rates after 3 d, in increasing concentrations of MMC, using propidium iodide to discriminate live cells from dead cells as described[22].

Immunoblotting. We performed immunoblots with samples containing 50 μg of total protein on 7% or 3–8% (for BRCA2) NuPage Tris-acetate polyacrylamide gels (Invitrogen). Membranes were probed with mouse monoclonal anti-FANCD2 (1:800; Santa Cruz sc-20022), rabbit polyclonal anti-BRIP1 (1:1,000; Novus NB100-416) or rabbit polyclonal anti-BRCA2 (1:500; Calbiochem PC146). Rabbit polyclonal anti-PALB2 (1:1,000) was a kind gift of B. Xia. Secondary antibodies included sheep anti-mouse IgG (GE Healthcare RPN4201) or donkey anti-rabbit IgG (GE Healthcare NA934V). We used these horseradish peroxidase–linked whole antibodies at dilutions of 1:2,000 to 1:5,000 and detected them by the chemiluminescence technique using the ECL system (Amersham).

Immunofluorescence. We analyzed the capability of SH2038-F primary fibroblasts to form nuclear RAD51 foci as follows. Cells were grown on glass slides, and subconfluent cultures were exposed to 50 ng/ml of MMC for 15 h for foci induction. The cells were fixed in 4% (vol/vol) paraformaldehyde in PBS (pH 6.8) for 15 min on ice and permeabilized with 0.5% (vol/vol) Triton X-100 in PBS for 10 min on ice. After blocking with 0.5% (vol/vol) fetal bovine serum and washing in PBS, we incubated the slides with monospecific rabbit polyclonal anti-RAD51 as the primary antibody (1:800; Abcam, ab 63801). Secondary antibody was Alexa594-conjugated goat anti-rabbit IgG (1:2,000; Invitrogen/Molecular Probes, A-11037). The cells were counterstained with DAPI in Vectorshield mounting medium (Vector Laboratories). We determined the percentage of foci-positive cells (more than ten foci per nucleus) visually on a Zeiss Axio Imager.A1 fluorescence microscope. For each experiment, 200–400 nuclei were analyzed.

SNP genotyping. We undertook genome-wide linkage analysis using the GeneChip Human Mapping 10K Array Xba 142 2.0, containing 10,204 SNP markers. The median intermarker distance was 113 kb, and the mean heterozygosity of markers was 0.38. We obtained genotypes using the Affymetrix protocol for the GeneChip Mapping 10K Xba array and images using an Affymetrix Gene Chip Scanner 3000. Affymetrix GeneChip Operating Software 1.4 software was used to obtain raw allele scores. We processed scores using Affymetrix GeneChip Genotyping Analysis Software (GTYPE) to derive SNP genotypes. Genotype calls were analyzed with AutoSNPa[26], which allows visualization of genotype data across each chromosome for rapid autozygosity mapping.

DNA sequencing. We designed primers to amplify the nine exons and intron-exon boundaries of *RAD51C*. Primers and PCR conditions are shown in **Supplementary Table 3**. For DNA sequencing of PCR products, we used the BigDye v3.1 cycle sequencing kit and a 3730XL DNA sequencer (Applied Biosystems).

Modeling RAD51C structure. We modeled RAD51C structure by homology from the crystal structure of the archaeal Rad51 protein from *P. furiosus* (PDB 1PZN)[7]. We produced the sequence alignment used to build the model using PRALINE with the homology-extended alignment strategy[27]. We generated three-dimensional models using the MODELLER package[28]. The selected model was chosen on the basis of the MODELLER objective function's score. We obtained the *in silico* mutant R258H using the PyMOL mutagenesis tool (DeLano Scientific). To refine both the model and the *in silico* mutant, we performed energy minimizations with the GROMACS package[29] using the GROMOS96 force field[30].

URLs. PyMOL, http://www.pymol.org/.

16. Auerbach, A.D. *et al.* International Fanconi Anemia Registry: relation of clinical symptoms to diepoxybutane sensitivity. *Blood* **73**, 391–396 (1989).
17. Sandoval, N. *et al.* Characterization of ATM gene mutations in 66 ataxia telangiectasia families. *Hum. Mol. Genet.* **8**, 69–79 (1999).
18. Enders, A. *et al.* A severe form of human combined immunodeficiency due to mutations in DNA ligase IV. *J. Immunol.* **176**, 5060–5068 (2006).
19. Waltes, R. *et al.* Human RAD50 deficiency in a Nijmegen breakage syndrome-like disorder. *Am. J. Hum. Genet.* **84**, 605–616 (2009).
20. Reid, S. *et al.* Biallelic mutations in PALB2 cause Fanconi anemia subtype FA-N and predispose to childhood cancer. *Nat. Genet.* **39**, 162–164 (2007).
21. Antonio Casado, J. *et al.* A comprehensive strategy for the subtyping of patients with Fanconi anaemia: conclusions from the Spanish Fanconi Anemia Research Network. *J. Med. Genet.* **44**, 241–249 (2007).
22. Hanenberg, H. *et al.* Phenotypic correction of primary Fanconi anemia T cells with retroviral vectors as a diagnostic tool. *Exp. Hematol.* **30**, 410–420 (2002).
23. Kalb, R. *et al.* Hypomorphic mutations in the gene encoding a key Fanconi anemia protein, FANCD2, sustain a significant group of FA-D2 patients with severe phenotype. *Am. J. Hum. Genet.* **80**, 895–910 (2007).
24. Hanenberg, H. *et al.* Colocalization of retrovirus and target cells on specific fibronectin fragments increases genetic transduction of mammalian cells. *Nat. Med.* **2**, 876–882 (1996).
25. Hanenberg, H. *et al.* Optimization of fibronectin-assisted retroviral gene transfer into human CD34+ hematopoietic cells. *Hum. Gene Ther.* **8**, 2193–2206 (1997).
26. Carr, I.M. *et al.* Interactive visual analysis of SNP data for rapid autozygosity mapping in consanguineous families. *Hum. Mutat.* **27**, 1041–1046 (2006).
27. Simossis, V.A. *et al.* Homology-extended sequence alignment. *Nucleic Acids Res.* **33**, 816–824 (2005).
28. Martí-Renom, M.A. *et al.* Comparative protein structure modeling of genes and genomes. *Annu. Rev. Biophys. Biomol. Struct.* **29**, 291–325 (2000).
29. Berendsen, H.J.C. *et al.* GROMACS: a message-passing parallel molecular dynamics implementation. *Comput. Phys. Commun.* **91**, 43–56 (1995).
30. Daura, X. *et al.* Parametrization of aliphatic CH$_n$ united atoms of GROMOS96 force field. *J. Comput. Chem.* **19**, 535–547 (1998).

Supplementary Material

Mutation of the *RAD51C* gene in a Fanconi anemia-like disorder

Fiona Vaz, Helmut Hanenberg, Beatrice Schuster, Karen Barker, Constanze Wiek, Verena Erven, Kornelia Neveling, Daniela Endt, Ian Kesterton, Flavia Autore, Franca Fraternali, Marcel Freund, Linda Hartmann, David Grimwade, Roland G Roberts, Heiner Schaal, Shehla Mohammed, Nazneen Rahman, Detlev Schindler, Christopher G Mathew
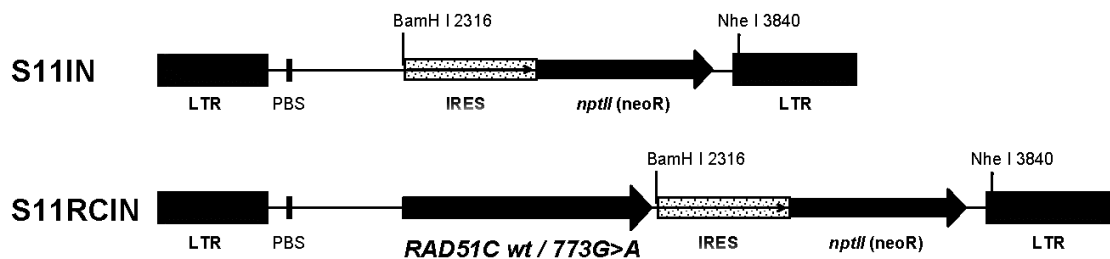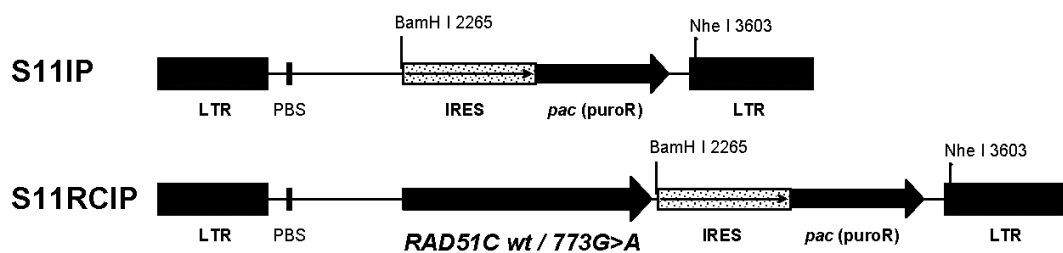
Supplementary Figure 1

Evolutionary conservation of R258 in the RAD51 protein family

Alignment of sequences of RAD51C (RAD51L2), RAD51B (RAD51L1), RAD51 and XRCC3 from human, chicken, zebrafish, sea urchin and thale cress (*Homo sapiens*, *Gallus gallus*, *Danio rerio*, *Strongylocentrotus purpuratus*, and *Arabidopsis thaliana*). Secondary structure is depicted schematically above the amino acid sequences with cylinders for α-helices and arrows for β-sheets, and is taken from Shin *et al.* 2003 (PDB 1N0W)[7] and Miller *et al.* 2004 (Nucleic Acids Res. **32**, 169-178). The sites of the Walker B motif and the R258H mutation are indicated. (The corresponding region in RAD51D and XRCC2 seems distinct in its sequence constraints and does not contain a readily identifiable R258 counterpart).
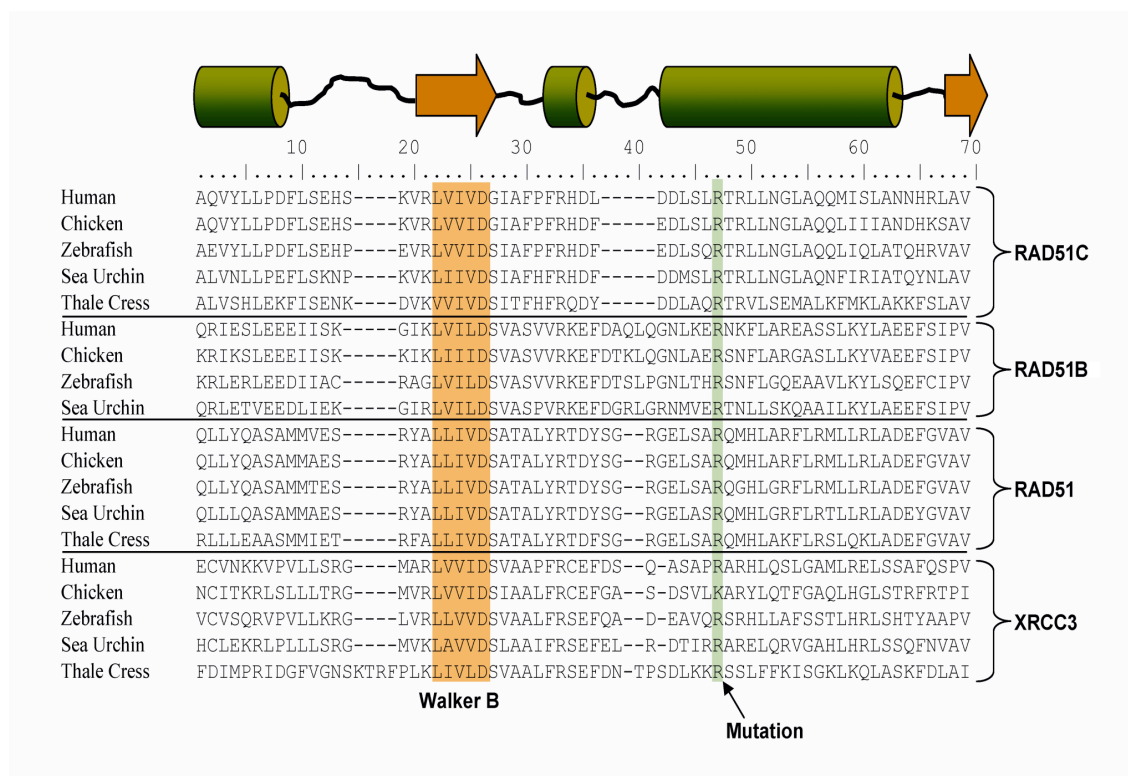
Supplementary Figure 2

Maps of the retroviral vector plasmids

Expression cassettes consisting of *RAD51C* wildtype and 773G>A mutant cDNAs linked via encephalomyocarditis virus internal ribosomal entry site (IRES) to either the neomycin phosphotransferase II (*nptII*) gene or to the puromycin N-acetyl-transferase (*pac*) gene were expressed in oncoretroviral vector off the viral 5' LTR. The same viruses without *RAD51C* cDNA were utilized as controls.

Neomycin phosphotransferase II (nptII) vectors

pac (puroR) vectors

Supplementary Figure 3

Models of the possible structural effect of the R258H mutation in RAD51C

Panels A and B show a close-up view in ribbon representation of the regions around residue 258 in wildtype arginine (A) and in mutant histidine (B) RAD51C, with loss of the interaction between H258 and the E303 carbonyl backbone. Panels C and D show the electrostatic potential surfaces of the wildtype R258 and mutant H258 RAD51C proteins respectively. The electrostatic potential surfaces have been calculated with values of the potential ranging from -6 kT (red) to the maximal positive value +6 kT (blue). The region subject to a change in the electrostatic potential from slightly positive (wildtype) to slightly negative (mutant H258) is highlighted by a square box.
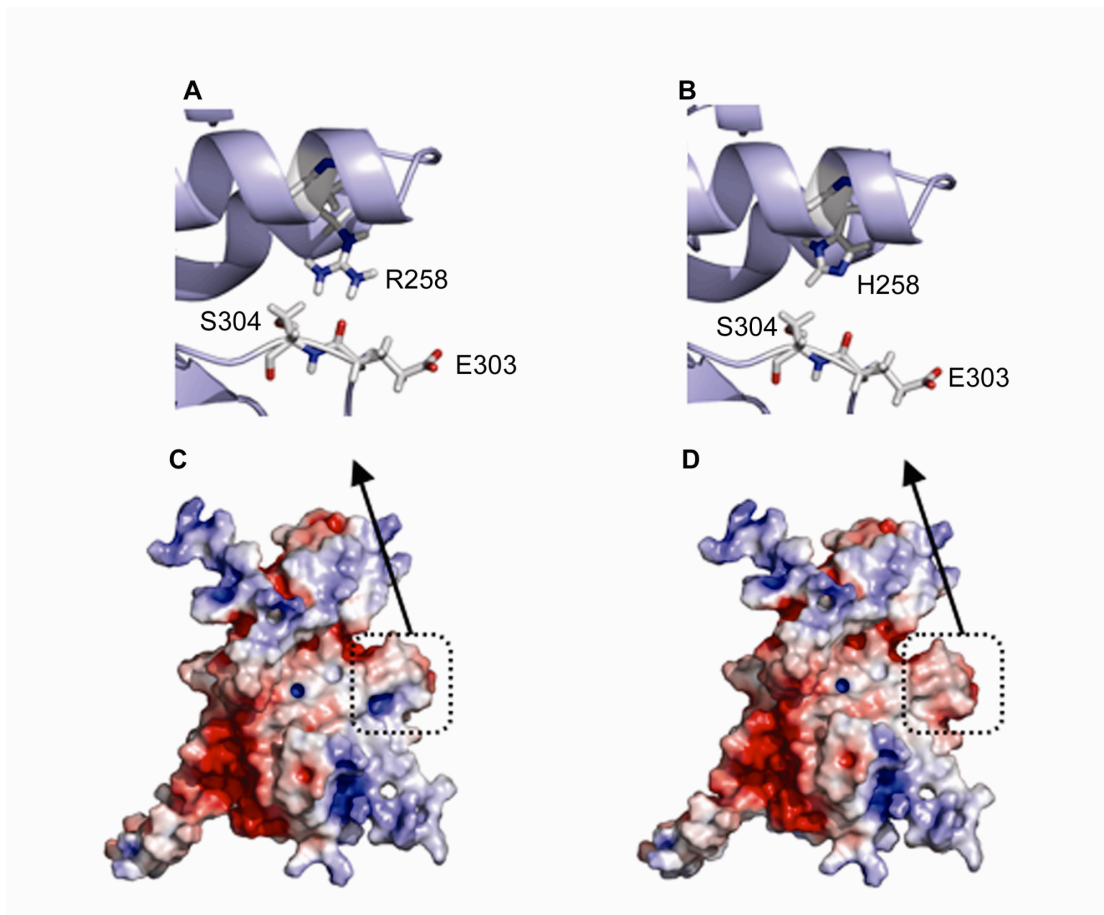
Supplementary Table 1

Large regions of shared homozygosity in the affected siblings IV-3 and IV-5 in which the unaffected sibling was heterozygous, or homozygous for the other allele

Chromosome	Region of homozygosity	Size (bp)
17	47396100 to 62034036	14637936
22	35841022 to 46610021	10768999
7	152524507 to 156156872	3632365
2	134136406 to 137702474	3566068
2	61229967 to 64700540	3470573
2	66576315 to 69741341	3165026
5	3830394 to 6912895	3082501

Supplementary Table 2

Analysis of radiosensitivity (see Online Methods) in fibroblasts from patient IV-5 (SH2038-F) before and after complementation with wildtype (wt) *RAD51C*. Data from cell lines from patients with known radiosensitivity are included for comparison. (ED_{50} = Dose for 50% clonogenic survival; Gy = Grays)

Cell line	ED_{50} (Gy)
Control human fibroblast	1.53
SH2038-F	1.22
SH2038-F + *RAD51C* -wt	1.67
Ataxia telangiectasia	0.39
DNA-ligase IV deficient fibroblasts	0.61
RAD50 deficient fibroblasts	0.77

Supplementary Table 3

Primer sequences used to amplify and sequence *RAD51C*, with annealing temperature (Tm) and amplicon size (TD – touchdown PCR, * indicates that DMSO was added to a final concentration of 10%).

Exon	Forward 5' to 3'	Reverse 5' to 3'	Tm (°C)	Size (bp)
1	AAATGGGATTTTGGGGAATC	GTAAACATGGACGTGGGAGG	TD*	471
2	AAAATTAAATGGTTGATAGAATGTTGC	TCAAGAAGGGATAATGAAGTAACAC	65	583
3	GACATTTCTGTTGCCTTGGG	GCTGTGGCATTTCTCATTTTG	65	472
4	TTTTGCTATAATTTGTCATCTTTCAG	TTGTAGGTCAAGGAAGGAAGAGA	60	413
5	TTACTGTTCCAGGCATTGGG	TGGAAACCAACCAAACGTAAC	65	430
6	GTGCATGCCACCATGTCT	TGTGTCTGGCCACTCAATAAA	68	398
7	GAATAATGATTTGCAGTATTTCCC	CAGACAAGGCAACAAAAGTGTC	65	400
8	CATACGGGTAATTTGAAGGGTG	TTTGGGGACAATGTTCTAAGC	65	384
9	CGCCTGGCCCTAGAATAAA	GGCCACATGAGATCAGCTTT	65	491

3. RESULTS

3.1 FA CANDIDATE GENES

3.1.1 A HISTONE-FOLD COMPLEX AND FANCM FORM A CONSERVED DNA-REMODELING COMPLEX TO MAINTAIN GENOME STABILITY

3.1.2 ON THE ROLE OF FAN1 IN FANCONI ANEMIA

3.2 IDENTIFICATION OF NOVEL FA GENES

3.2.1 MUTATION OF THE RAD51C GENE IN A FANCONI ANEMIA-LIKE DISORDER

3.2.2 SLX4, A COORDINATOR OF STRUCTURE-SPECIFIC ENDONUCLEASES, IS MUTATED IN A NEW FANCONI ANEMIA SUBTYPE

3.2.3 XPF MUTATIONS SEVERELY DISRUPTING DNA INTERSTRAND CROSSLINK REPAIR CAUSE FANCONI ANEMIA

3.3 GENOTYPING FANCONI ANEMIA BY NEXT GENERATION SEQUENCING

3.3.1 WHOLE EXOME SEQUENCING REVEALS NOVEL MUTATIONS IN THE RECENTLY IDENTIFIED FANCONI ANEMIA GENE SLX4/FANCP

3.3.2 GENOTYPING FANCONI ANEMIA BY WHOLE EXOME SEQUENCING: ADVANTAGES AND CHALLENGES

LETTERS

SLX4, a coordinator of structure-specific endonucleases, is mutated in a new Fanconi anemia subtype

Chantal Stoepker[1], Karolina Hain[2], Beatrice Schuster[3], Yvonne Hilhorst-Hofstee[4], Martin A Rooimans[1], Jurgen Steltenpool[1], Anneke B Oostra[1], Katharina Eirich[3], Elisabeth T Korthof[4], Aggie W M Nieuwint[1], Nicolaas G J Jaspers[5], Thomas Bettecken[6], Hans Joenje[1], Detlev Schindler[3], John Rouse[2] & Johan P de Winter[1]

DNA interstrand crosslink repair requires several classes of proteins, including structure-specific endonucleases and Fanconi anemia proteins. SLX4, which coordinates three separate endonucleases, was recently recognized as an important regulator of DNA repair. Here we report the first human individuals found to have biallelic mutations in *SLX4*. These individuals, who were previously diagnosed as having Fanconi anemia, add SLX4 as an essential component to the FA-BRCA genome maintenance pathway.

Fanconi anemia is a rare, heterogeneous chromosomal instability syndrome characterized by bone marrow failure, congenital abnormalities, hypersensitivity to DNA crosslinking agents and an increased susceptibility to cancer. Studies to unravel the genetic basis of Fanconi anemia have led to the identification of a previously unidentified genome maintenance pathway which evolved relatively late during evolution and exists—in its fully developed form—only in vertebrates. Fourteen Fanconi anemia genes have been identified[1,2], but a small percentage of individuals diagnosed with Fanconi anemia

Table 1 Clinical phenotypes of individuals with *SLX4* mutations

	EUFA1354	457-1	457-2	457-3
Age at diagnosis[a]	9	9	7 (via 457-1)	7 (via 457-1)
Growth retardation	Short stature (–2.5 s.d. at age 9; –4.5 s.d. at age 18)	Prenatal	Prenatal	Pre- and post-natal
Thumb abnormalities	Hypoplastic right thumb	No	No	No
Facial features	Almond-shaped and short palpebral fissures, bulbous nasal tip, micrognathia, microcephaly (–2.5 s.d.)	No	No	No
Skin abnormalities	Hypopigmented spot on back	No	No	Café-au-lait spots
Ear abnormalities	Bilateral hearing loss, hypoplastic malleus, narrow external ear canals	No	No	No
Kidney abnormalities	No	Horseshoe kidney	No	No
Hematology[b]	Pancytopenia, Hb 2.3 mM/l, leucocytes 3.4×10^9/l, neutrophils 0.7×10^9/l, thrombocytes 18×10^9/l. Individual transplanted at age ten with bone marrow from the mother and is currently (age 18) doing well.	Pancytopenia, Hb 6.1 mM/l, erythrocytes 3.0×10^{12}/l, leucocytes 2.7×10^9/l, neutrophils 0.42×10^9/l, thrombocytes 20×10^9/l. Individual received HSCT with BM from an unrelated donor at age 9.5 and is currently (age 11.5) doing well.	Beginning bone marrow failure, erythrocytes 3.7×10^{12}/l, leucocytes 3.7×10^9/l, neutrophils 0.93×10^9/l, thrombocytes 107×10^9/l. Current age is 8.5.	Pancytopenia, Hb 5.7 mM/l, erythrocytes 2.8×10^{12}/l, leucocytes 2.5×10^9/l, neutrophils 0.27×10^9/l, thrombocytes 20×10^9/l. Individual received HSCT with BM from unrelated donor at the current age of 8.5.
Paternal *SLX4* mutation	c.286delA, p.Thr96Leu*fs*X30	c.1093delC, p.Gln365Ser*fs*X31	c.1093delC; p.Gln365Ser*fs*X31	c.1093delC; p.Gln365Ser*fs*X31
Maternal *SLX4* mutation	c.286delA, p.Thr96Leu*fs*X30	c.1163+3dupT, p.Arg317_Phe387 del	c.1163+3dupT, p.Arg317_Phe387 del	c.1163+3dupT, p.Arg317_Phe387 del

s.d., standard deviation; HSCT, hematopoietic stem cell transplantation; BM, bone marrow.
[a]Age at diagnosis in years. [b]Normal values: Hb 7.4–9.0 mM/l, erythrocytes (male) $4.7–6.1 \times 10^{12}$/l, leucocytes $4.5–13.5 \times 10^9$/l, neutrophils $1.8–8.0 \times 10^9$/l, thrombocytes $150–450 \times 10^9$/l.

[1]Department of Clinical Genetics, Vrije Universiteit (VU) Medical Center, Amsterdam, The Netherlands. [2]Medical Research Council Protein Phosphorylation Unit, College of Life Sciences, University of Dundee, Dundee, Scotland, UK. [3]Department of Human Genetics, University of Wuerzburg, Biozentrum, Wuerzburg, Germany. [4]Department of Clinical Genetics, Leiden University Medical Center, Leiden, The Netherlands. [5]Department of Genetics, Erasmus Medical Center, Rotterdam, The Netherlands. [6]Center for Applied Genotyping Munich, Max-Planck-Institut für Psychiatrie, Munich, Germany. Correspondence should be addressed to J.P.d.W. (j.dewinter@vumc.nl), J.R. (j.rouse@dundee.ac.uk) or D.S. (schindler@biozentrum.uni-wuerzburg.de).

Figure 1 Cellular characteristics of individuals with Fanconi anemia with a defect in SLX4. (**a**) Spontaneous and mitomycin C (MMC)-induced chromosomal breakage in EBV-immortalized lymphoblasts from a healthy individual (EUFA045-L), a FANCA-deficient individual (HSC72) and individual EUFA1354. Percentages of cells with up to ten or more break events per metaphase are shown. (**b**) Growth inhibition of EUFA1354, 457/1 and 457/3 lymphoblasts upon exposure to MMC or (**c**) camptothecin (CPT). Lymphoblasts from a healthy individual (HSC93) and a FANCM-deficient individual (EUFA867-L) were analyzed as controls. Data represent means and s.e.m. from at least two independent experiments. (**d**) Immunoprecipitation and protein blot analysis showing reduced SLX4 expression in lymphoblasts from the affected siblings 457/1 and 457/3 and the absence of full-length SLX4 in lymphoblasts from EUFA1354. The mutant protein in 457/1 and 457/3 interacts with ERCC1, XPF and MUS81, whereas these proteins are not co-precipitated in EUFA1354-L. We performed immunoprecipitation with antibodies against the SLX4 N terminus (1–300) and protein blotting with an antibody against the SLX4 C terminus (1,534–1,834). SE and LE indicate short and long exposures of the blot, respectively. (**e**) Immunoprecipitation and protein blot analysis demonstrating the absence of full-length SLX4 and impaired interactions with MUS81 and ERCC1 in SV40-immortalized EUFA1354 fibroblasts. We used wildtype fibroblasts (LN9SV) as a control. (**f**) Subcellular fractionation of EUFA1354 lymphoblasts and fibroblasts showing reduced chromatin binding of ERCC1 and XPF. We analyzed the cytoplasmic fraction (CE), nuclear extract (NE) and chromatin fraction (CB) and used tubulin, p300 and histone H3 as controls for these fractions. WT, wildtype; FA-A, Fanconi anemia type A; FA-M, Fanconi anemia type M.
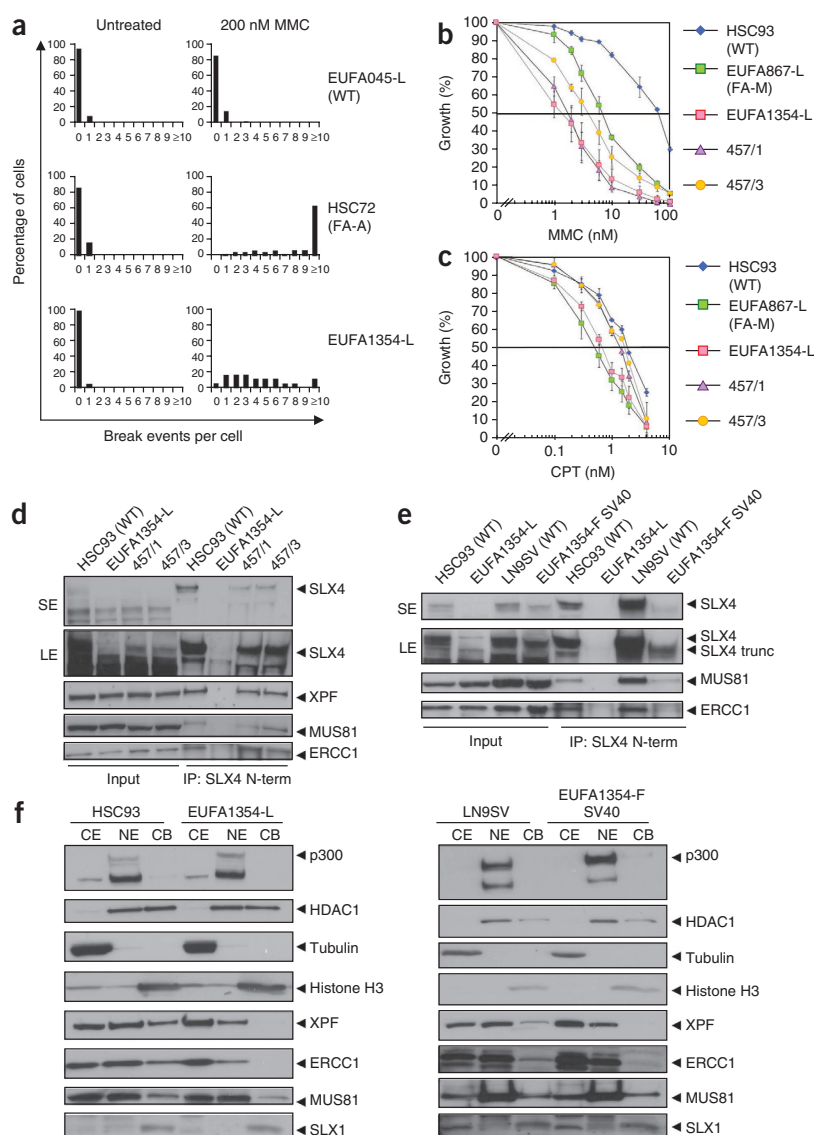
have remained unclassified, as no pathogenic mutations could be detected in the currently known Fanconi anemia genes.

One of these individuals (EUFA1354), a Dutch male with growth retardation, microcephaly, small eyes, hypopigmentation, thumb abnormalities and hearing loss, was diagnosed with pancytopenia at the age of 9 and Fanconi anemia was suspected (**Table 1**). We confirmed the Fanconi anemia diagnosis by a chromosomal breakage assay on T lymphocyte cultures, which showed increased spontaneous and excessive mitomycin C (MMC)-induced chromosomal aberrations that were well within the range established for Fanconi anemia (**Supplementary Fig. 1**). An EBV-immortalized lymphoblastoid cell line from this individual was also hypersensitive to MMC in terms of chromosomal breakage (**Fig. 1a**) and growth inhibition (**Fig. 1b** and **Supplementary Fig. 2a**). Notably, these lymphoblasts were also hypersensitive to the topoisomerase I inhibitor camptothecin (**Fig. 1c** and **Supplementary Fig. 2b**), a feature that until now was considered specific for the Fanconi anemia subgroups D1, M, N and O and which is possibly associated with defects in homologous recombination repair[2,3]. In further support of the Fanconi anemia diagnosis, we observed an elevated MMC-induced accumulation in the G2/M phase of the cell cycle, both in primary and in SV40-immortalized fibroblasts from this individual (**Supplementary Fig. 2c,d**). Somewhat surprisingly, fibroblasts were not very sensitive to the crosslinking drug when using growth inhibition or chromosomal breakage as a readout (**Supplementary Fig. 2e–g**).

Sequence analysis, MMC-induced FANCD2 monoubiquitination and normal formation of nuclear FANCD2 foci (**Supplementary Fig. 3a,b**) excluded a defect in the upstream part of the FA-BRCA pathway in this individual[4]. The induction of RAD51 and H2AX foci in this individual's fibroblasts suggested a normal DNA damage response (**Supplementary Fig. 3c–f**). We ruled out an abnormality in the downstream Fanconi anemia proteins BRCA2, PALB2 and FANCJ by cell fusion experiments and sequence analysis of the corresponding genes (data not shown), indicating that this individual represented a new Fanconi anemia subtype with a defect in a new player of the FA-BRCA pathway.

Recently, four research groups identified the human SLX4 scaffold protein, which was proposed to function in the processing of DNA

LETTERS

Figure 2 Functional correction of EUFA1354 fibroblasts and lymphoblasts by exogenous SLX4. ERCC1 foci are absent in (**a**) primary and (**b**) SV40-immortalized EUFA1354 fibroblasts. Bebu and LN9SV were used as wildtype controls. (**c**) Transient transfection of FLAG-SLX4 in SV40-immortalized EUFA1354 fibroblasts restored their ability to form nuclear ERCC1 foci. An empty vector (pcDNA3) was used as a control. (**d**) Transient transfection of GFP-SLX4 in SV40-immortalized EUFA1354 fibroblasts restored their ability to form nuclear ERCC1 foci. Transient transfection of GFP was used as a control. ERCC1 foci are more pronounced after brief pre-permeabilization with Triton-X100 before fixation, but this removed exogenous GFP. ERCC1 antibody FL297 was used for immunofluoresence, and DAPI (blue) was used as a nuclear counterstaining. (**e**) A mouse Slx4 protein that lacks the Slx1 binding site (FLAG-Slx4ΔSlx1) is stably expressed in EUFA1354 lymphoblasts and interacts with ERCC1. (**f**) FLAG-Slx4ΔSlx1 partially corrects MMC-induced growth inhibition in EUFA1354 lymphoblasts. Data represent means and s.e.m. from two independent experiments. (**g**) FLAG-Slx4ΔSlx1 partially corrects MMC-induced chromosomal breakage in EUFA1354 lymphoblasts. WT, wildtype.
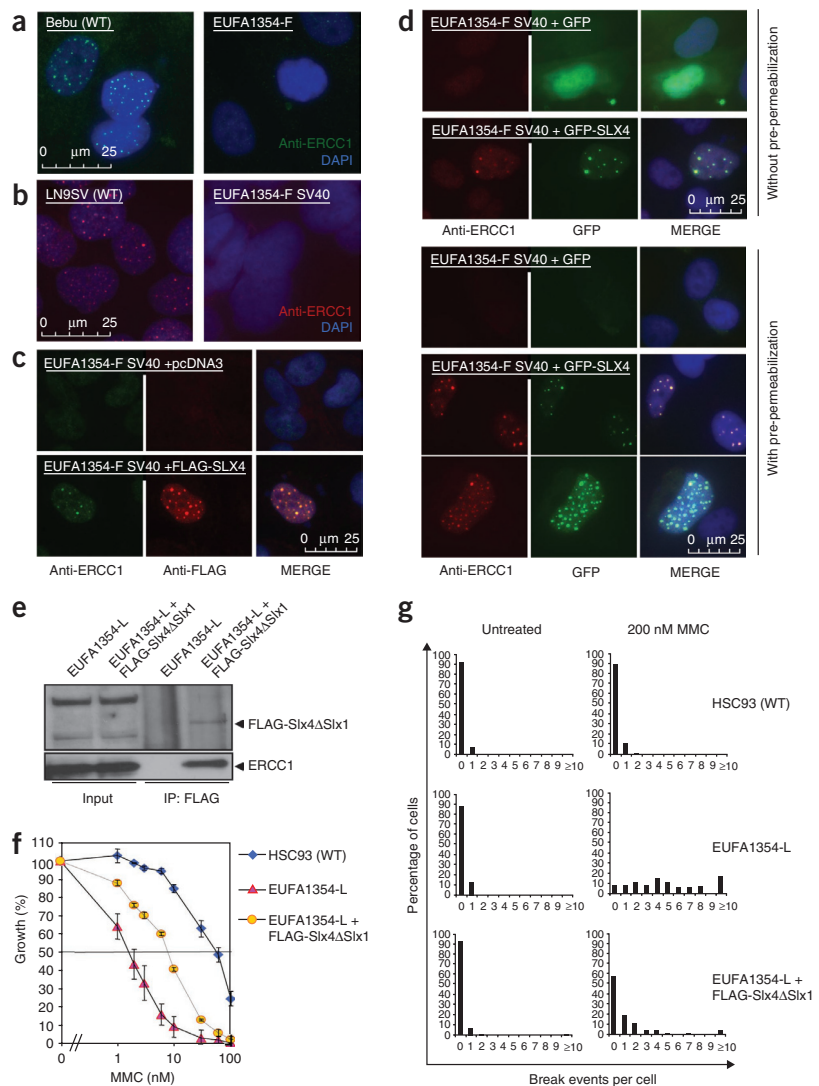
repair intermediates and crosslink repair through interaction with the structure-specific endonucleases SLX1, XPF-ERCC1 and MUS81-EME1 (refs. 5–8). SLX4-depleted cells are hypersensitive to crosslinking agents and camptothecin, similar to lymphoblasts from the affected individual EUFA1354. Therefore, we hypothesized that SLX4 might be defective in this individual. Sequence analysis of genomic DNA and complementary DNA (cDNA) from EUFA1354 indeed revealed biallelic mutations in the reading frame of *SLX4* (**Supplementary Fig. 4**). We detected a homozygous, 1-bp deletion (c.286delA) in the first exon of *SLX4* that results in a frameshift at codon 96 and a premature stop at codon 126 (p.Thr96Leu*fs*X30). The consanguineous parents and healthy sister of the affected individual were all heterozygous for this sequence variant.

We obtained additional evidence for SLX4 deficiency in individuals with Fanconi anemia by the identification of a second Fanconi anemia family with pathogenic *SLX4* mutations. Linkage analysis with a genome-wide SNP array showed a common 13.5 Mbp region around the *SLX4* locus in three unclassified German siblings with mild manifestations of Fanconi anemia and bone marrow failure (**Table 1** and **Supplementary Fig. 5**). The siblings all inherited a 1-bp deletion in *SLX4* (c.1093delC, p.Gln365Ser*fs*X31) from their father and a splice site mutation (c.1163+3dupT) from their mother (**Supplementary Fig. 6a**). The maternal mutation lead to exon 5 skipping, an in-frame deletion that disrupts the UBZ4 domain in SLX4, which may be involved in targeting SLX4 to sites of DNA damage by binding to ubiquitinated proteins (**Supplementary Fig. 6b,c**). We detected the residual mutant SLX4 protein in cell lysates, and this protein was able to interact with XPF, ERCC1 and MUS81 (**Fig. 1d**). Notably, lymphoblasts from these individuals were

sensitive to MMC (**Fig. 1b**) but not to camptothecin (**Fig. 1c**), indicating that the mutant protein is partially functional.

We also examined SLX4 protein expression in lymphoblast lysates from EUFA1354. Full-length SLX4 was undetectable by immuno-precipitation with antibodies against the N or C terminus of SLX4, but we detected very low levels of a truncated SLX4 protein with antibodies against the C terminus of SLX4 (**Fig. 1d** and **Supplementary Fig. 7a**). Consequently, the amounts of XPF-ERCC1 and MUS81 in SLX4 immunoprecipitates were reduced compared to wildtype cells (**Fig. 1d** and **Supplementary Fig. 7a**). The truncated SLX4 protein was more pronounced in SV40-immortalized fibroblasts from EUFA1354 (**Fig. 1e** and **Supplementary Fig. 7b**) and may be derived from an alternative translation initiation site present at codon 213. Reciprocal immunoprecipitations with antibodies against ERCC1, XPF or MUS81 readily co-precipitated full-length SLX4 from wildtype cells, but only some truncated SLX4 protein was co-precipitated from EUFA1354 cells (**Supplementary Fig. 7c,d**). When transiently expressed in human HEK293 cells, a truncated SLX4 protein starting from Met213 is able to interact with XPF-ERCC1, MUS81-EME1 and SLX1, similar to full-length SLX4 (**Supplementary Fig. 7e**). These data indicate that

LETTERS

the defects seen in the EUFA1354 cells are due to very low concentrations of the truncated SLX4 protein and not because truncated SLX4 is defective in interacting with a specific nuclease.

Gel filtration experiments showed that XPF-ERCC1 and SLX1, which normally exist in two subcomplexes[8], eluted only in the low molecular weight fractions from EUFA1354 fibroblast lysates, whereas the elution of MUS81 was not affected (**Supplementary Fig. 8**). Subcellular fractionation studies revealed a reduced chromatin association of XPF-ERCC1 in EUFA1354, whereas MUS81 and SLX1 were hardly affected (**Fig. 1f**). These data strongly suggest that the SLX4 defect in EUFA1354 interferes with the function of the XPF-ERCC1 endonuclease, an important player in crosslink repair.

According to a recent study[6], SLX4 accumulates in nuclear foci, where it co-localizes with XPF. We investigated the nuclear localization of ERCC1 in primary and immortalized EUFA1354 fibroblasts. We detected nuclear ERCC1 foci in wildtype cells; however, EUFA1354 fibroblasts lacked these structures (**Fig. 2a,b**). We confirmed this striking result with another ERCC1 antibody[9] (**Supplementary Fig. 9**). EUFA1354 fibroblasts, transiently transfected with SLX4 cDNA, regained their capacity to form ERCC1 foci, which co-localize with both FLAG-SLX4 and GFP-SLX4 (**Fig. 2c,d**). This confirms that the SLX4 defect in EUFA1354 affects the XPF-ERCC1 endonuclease.

To further strengthen the evidence that SLX4 deficiency causes the cellular phenotype in EUFA1354, we tried to correct the MMC hypersensitivity through SLX4 cDNA transfection in EUFA1354 lymphoblasts. Although we were unable to stably express full-length human SLX4, a truncated mouse Slx4 protein that corrects Slx4-deficient mouse embryonic fibroblasts[10] partially restored MMC resistance in terms of cell growth (**Fig. 2e,f**). Chromosomal breakage analysis on individual metaphases showed that only part of the lymphoblasts (57%) had been fully complemented (**Fig. 2g**).

In conclusion, defective SLX4 is associated with a previously unknown subtype of Fanconi anemia, Fanconi anemia-P. The presence of a truncated SLX4-FANCP protein in the individuals reported here indicates that the mutations may be hypomorphic, as has been found for individuals with an alteration in ERCC1 or XPF[11,12]. Because germline mutations in all the other downstream players of the FA-BRCA pathway predispose to breast and/or ovarian cancer, *SLX4* could also be a cancer predisposition gene.

METHODS

Methods and any associated references are available in the online version of the paper at http://www.nature.com/naturegenetics/.

1. de Winter, J.P. & Joenje, H. The genetic and molecular basis of Fanconi anemia. *Mutat. Res.* **668**, 11–19 (2009).
2. Vaz, F. *et al.* Mutation of the *RAD51C* gene in a Fanconi anemia-like disorder. *Nat. Genet.* **42**, 406–409 (2010).
3. Singh, T.R. *et al.* Impaired FANCD2 monoubiquitination and hypersensitivity to camptothecin uniquely characterize Fanconi anemia complementation group M. *Blood* **114**, 174–180 (2009).
4. Shimamura, A. *et al.* A novel diagnostic screen for defects in the Fanconi anemia pathway. *Blood* **100**, 4649–4654 (2002).
5. Fekairi, S. *et al.* Human SLX4 is a Holliday junction resolvase subunit that binds multiple DNA repair/recombination endonucleases. *Cell* **138**, 78–89 (2009).
6. Svendsen, J.M. *et al.* Mammalian BTBD12/SLX4 assembles a Holliday junction resolvase and is required for DNA repair. *Cell* **138**, 63–77 (2009).
7. Andersen, S.L. *et al. Drosophila* MUS312 and the vertebrate ortholog BTBD12 interact with DNA structures-specific endonucleases in DNA repair and recombination. *Mol. Cell* **35**, 128–135 (2009).
8. Muñoz, I.M. *et al.* Coordination of structure-specific nucleases by human SLX4/BTBD12 is required for DNA repair. *Mol. Cell* **35**, 116–127 (2009).
9. van Vuren, A.J. *et al.* Evidence for a repair enzyme complex involving ERCC1 and complementing activities of ERCC4, ERCC11 and xeroderma pigmentosum group F. *EMBO J.* **12**, 3693–3701 (1993).
10. Crossan, G.P. *et al.* Disruption of mouse Slx4, a regulator of structure-specific nucleases, phenocopies Fanconi anemia. *Nat. Genet.* advance online publication, doi:10.1038/ng.752 (16 January 2011).
11. Jaspers, N.G.J. *et al.* First reported patient with human ERCC1 deficiency has cerebro-oculo-facio-skeletal syndrome with a mild defect in nucleotide excision repair and severe developmental failure. *Am. J. Hum. Genet.* **80**, 457–466 (2007).
12. Ahmad, A. *et al.* Mislocalization of XPF-ERCC1 nuclease contributes to reduced DNA repair in XP-F patients. *PLoS Genet.* **6**, e1000871 (2010).

ONLINE METHODS

Ethics statement. The research was carried out after approval by the Institutional Review Board of the Vrije Universiteit Medical Center that adhered to local ethical standards and was initiated only after the relevant informed consents had been obtained. Information and consent of the German family used in this study was in agreement with national legal regulations and the Declaration of Helsinki.

Cell culture and transfection. HEK293 cells were cultured in Dulbecco's Modified Eagle's Medium (DMEM, GIBCO) with 10% FBS (Hyclone), 100 international units ml⁻¹ penicillin and 0.1 mg ml⁻¹ streptomycin and 1% L-glutamine (Invitrogen). Cells were transiently transfected by calcium phosphate precipitation and incubated for 24–30 h at 37 °C before lysis.

Skin fibroblasts, either primary or immortalized with SV40 large T antigen, were cultured in DMEM or Ham's F-10 with 10% FBS. Fibroblasts were transiently transfected with Amaxa Nucleofector using a human dermal fibroblasts nucleofector kit.

EBV-transformed lymphoblasts were cultured in RPMI1640 medium containing 10% FBS. Stably expressing lymphoblasts were generated by electroporation of pMEP4 constructs (Invitrogen) and selection on 100 µg/ml hygromycin.

cDNA constructs. The cDNA encoding full-length human SLX4 was generated by PCR on IMAGE clones 6527830 and 4340346 and was cloned into pcDNA5 FRT/TO-FLAG and pcDNA5 FRT/TO-GFP (Invitrogen, Flp-In T-Rex system). For GFP-SLX4 213-end, the pcDNA5 FRT/TO-GFP with full-length SLX4 was used to PCR amplify a fragment encoding SLX4 residues 213–1,834, which was shuttled into pcDNA5 FRT/TO-GFP. Constructs were verified by sequencing. A cDNA construct encoding mouse SLX4 that lacks the SLX1 binding site (amino acids 1,417–1,565 in mouse Slx4 corresponding to amino acids 1,685–1,834 in human SLX4) was a gift of K.J. Patel.

Chromosomal breakage analysis. Heparinized blood (2 ml) was diluted with 18 ml blood culture medium (Ham's F-10 + 15% FBS + phytohemagglutinin). Subsequently, 5 ml of this suspension was cultured in the presence of 0, 15 or 300 nM MMC. After 72 h, 100 µl demecolcin (10 µg/ml in Hank's Balanced Salt Solution (HBSS)) was added, and cells were incubated for 20 min at 37 °C. Cells were centrifuged, resuspended in 0.075 M KCl and incubated for 20 min at 18–22 °C. The cells were spun down, resuspended in 10 ml fixative (75% methanol and 25% acetic acid), incubated for 30 min at room temperature and centrifuged. The pellet was resuspended in 10 ml fixative and incubated for 5 min at room temperature. This step was repeated and finally the pellet was resuspended in 0.5–1.0 ml fixative. The cell suspension was dropped on a slide and allowed to dry. Slides were incubated for 15 min in 0.1 M HNO_3 and rinsed in tap water and 70% ethanol, respectively. Next, slides were stained for 5 min in a 4% Giemsa solution, rinsed in tap water, allowed to dry and coded. From each coded culture, at least 50 metaphases were examined for chromosomal damage. The presence of chromatid breaks and interchanges was expressed as break events per cell, counting chromatid interchange figures as the minimum number of break events required for their reconstruction. After scoring, the slides were decoded and the results were analyzed.

Lymphoblasts were seeded in 25 cm² tissue culture flasks at a density of 2–3 × 10⁶ per 10 ml fresh Roswell Park Memorial Institute (RPMI) medium containing 10% FBS and cultured at 37 °C for 48 h in the presence of 0 nM or 200 nM MMC. Next, 200 µl of demecolcin (10 µg/ml in HBSS) was added to each culture flask, and the cells were incubated for an additional 20 min at 37 °C. Metaphase spreads were prepared and analyzed as described above.

Fibroblasts (1 × 10⁶ cells in 15 ml Ham's F-10 + 10% FBS) were seeded in an 80 cm² tissue culture flask and cultured at 37 °C in the presence or absence of 50 nM mitomycin C. After 48 h, 300 µl of demecolcin (10 µg/ml in HBSS) was added to each culture flask, and cells were incubated for an additional 30 min at 37 °C. Cells were then trypsinized, and metaphase spreads were prepared and analyzed as described above.

Cell cycle analysis. Fibroblasts were cultured for 72 h either without or with MMC (50 nM or 100 nM) in Ham's F-10 medium supplemented with 10% FBS. Cells were harvested by trypsinization and permeabilized in buffer containing 100 mM Tris-HCL (pH 7.5), 150 mM NaCl, 0.5 mM $MgCl_2$,

1 mM $CaCl_2$, 0.2% BSA and 0.1% NP-40 followed by staining of DNA with PI/RNase staining buffer (BD Pharmingen, BD Biosciences). Cell suspensions were analyzed by flow cytometry on a BD FACSCalibur (BD Biosciences) to determine G2/M accumulation as the percentage of cells present in the G2/M phase of the cell cycle.

Lymphoblasts were stained with 4′-6-diamidino-2-phenylindole (DAPI) at a final concentration of 1 µg/ml. Flow histograms were made on an analytical, triple-laser–equipped flow cytometer (LSRII, BD Biosciences). Results were quantified with MPLUS AV software (Phoenix Flow Systems).

Growth inhibition assays. The MMC- and camptothecin-induced growth inhibition assays on lymphoblasts and fibroblasts were performed as previously described[13].

Cell fractionation. Cytoplasmic, nuclear and chromatin fractions were isolated using a Subcellular Protein Fractionation Kit (Thermo Scientific) according to the manufacturer's instructions. Controls used for the fractionation were tubulin (cytoplasmic fraction), p300 (nuclear fraction), histone H3 (chromatin fraction) and HDAC1 (nuclear and chromatin fraction).

Cell lysis, immunoprecipitation and protein blotting. Cells were lysed on ice with ice-cold lysis buffer containing 50 mM Tris-HCl (pH 7.4), 1% Triton X-100, 0.1% (v/v) 2-mercaptoethanol, 0.27 M sucrose and protease inhibitors for immunoprecipitation. For pull-downs with GFP-Trap beads (ChromoTek), cells were lysed on ice with ice-cold lysis buffer consisting of 10 mM Tris-HCl, (pH 7.5), 150 mM NaCl, 0.5 mM EDTA, 0.5% NP40 and protease inhibitors. Lysates were precleared with 50 µl protein G-Sepharose (50% slurry) for 30 min. Immunoprecipitations and pull-downs were performed for 2 h at 4 °C.

The primary antibodies used for immunoprecipitation and protein blotting were as follows: anti-XPF (Thermo MS-1381), anti-ERCC1 (Thermo MS-671 or Santa Cruz FL297), anti-MUS81 (ImmuQuest IQ285 or Abcam ab14387), anti-EME1 (ImmuQuest IQ284), anti-GFP (Roche 11814460001), anti-GAPDH (Abcam ab8245), anti-p300 (Santa Cruz SC-585), anti-HDAC1 (Santa Cruz SC7872), anti-tubulin (Abcam ab7291), anti-histone H3 (Cell Signaling Technology 9715), anti-FANCD2 (Santa Cruz sc20022) and anti-RAD50 (GeneTex 13B3). Sheep polyclonal antibodies were raised against the first 300 amino acids (SLX4-N) or the last 300 amino acids (SLX4-C) of SLX4 fused to GST. Antibodies were also raised against full-length SLX1-GST expressed in bacteria. SLX4 and SLX1 antibodies were affinity purified before use.

Immunofluorescence. Fibroblasts (100,000 cells per well) were seeded in four-well chamber slides (Nunc). After overnight culture, cells were washed with PBS and pre-permeabilized with 0.25% Triton X-100 in PBS (1 min on ice). Cells were fixed with 4% paraformaldehyde (15 min at room temperature) and permeabilized with 0.5% Triton X-100 in PBS (10–20 min at room temperature). Unspecific binding sites were blocked with PBS + 10% FBS (or PBS + 10% BSA for γH2AX staining) for 1 h at room temperature. Slides were incubated with specific antibodies in blocking buffer (2 h at room temperature). Excess antibody was removed by four washing steps with PBS + 0.2% Triton X-100. Slides were then incubated with secondary antibody labeled with Alexa 594 or Alexa 488 (Invitrogen, 1:1,000) and DAPI (1:1,000) for 1 h at room temperature. Finally, slides were washed four times with PBS + 0.2% Triton X-100 and embedded. Slides were analyzed with a fluorescent microscope (DM5000, Leica). ERCC1 foci were also analyzed 72 h after cDNA transfection.

Lymphoblasts were attached to glass slides by cytocentrifugation. The percentage of foci-positive cells (>10 foci per nucleus) was determined on a Zeiss Axio Imager A1 fluorescence microscope. For each experiment, 200–400 nuclei were analyzed.

The following antibodies were used: anti-FANCD2 (Novus NB100-182, 1:200), anti-RAD51 (gift of R. Kanaar, 1:2,500, or Abcam ab63801), anti-phospho-H2AX (Ser139) (Millipore clone JBW301, 1:400) and anti-ERCC1 (Santa Cruz FL297, 1:200).

Size exclusion chromatography. For gel filtration experiments, cell extracts (3 mg of protein) were loaded on a Superose 6 column (GE Healthcare) in buffer containing 50 mM Tris/HCl (pH 7.4), 1 mM EDTA, 0.2 M sodium chloride and 0.1% (v/v) 2-mercaptoethanol. Molecular weight markers (Bio-Rad)

were as follows: thyroglobulin (670 kDa) and bovine gamma globulin (158 kDa). Fraction was denatured and subjected to protein blot analysis.

Sequence analysis. For direct sequencing, PCR products were generated by SLX4-specific primers (**Supplementary Table 1**). PCR fragments were treated with shrimp alkaline phosphatase (30 min at 37 °C) and Exonuclease I (15 min at 80 °C). Sequencing reactions were carried out using 10 pM of primer and the Big Dye Terminator cycle sequencing kit (Applied Biosystems). Samples were analyzed in an ABI 3730 DNA analyzer (Applied Biosystems).

Genome wide SNP array. Genome wide SNP genotypes for the German family members were determined on a HumanHap300v2 Genotyping BeadChip (Illumina Inc.). After scanning the microarrays on a BeadStation 500G with the software BeadScan Ver. 3.7.5.23284, SNP genotypes were called using the software BeadStudio Ver. 3.1.3.0 with the Genotyping Module Ver. 3.2.32 (all from Illumina Inc.).

13. Joenje, H. *et al.* Classification of Fanconi anemia patients by complementation analysis: evidence for a fifth genetic subtype. *Blood* **86**, 2156–2160 (1995).

Supplementary information

SLX4, a coordinator of structure-specific endonucleases, is mutated in a new Fanconi anemia subtype

Chantal Stoepker[1], Karolina Hain[2], Beatrice Schuster[3], Yvonne Hilhorst-Hofstee[4], Martin A. Rooimans[1], Jurgen Steltenpool[1], Anneke B. Oostra[1], Katharina Eirich[3], Elisabeth T. Korthof[4], Aggie W. M. Nieuwint[1], Nicolaas G. J. Jaspers[5], Thomas Bettecken[6], Hans Joenje[1], Detlev Schindler[3#] John Rouse[2#] and Johan P. de Winter[1#]

[1]Department of Clinical Genetics, VU University Medical Center, Van der Boechorststraat 7, NL-1081 BT Amsterdam, The Netherlands, [2]MRC Protein Phosphorylation Unit, College of Life Sciences, University of Dundee, Dow street, DD1 5EH Dundee, Scotland, UK, [3]Department of Human Genetics, University of Wuerzburg, Biozentrum, Am Hubland, D-97074 Wuerzburg, Germany, [4]Department of Clinical Genetics, Leiden University Medical Center, P.O. Box 9600, NL-2300 RC Leiden, The Netherlands, [5]Department of Genetics, Erasmus Medical Center, P.O. Box 2040, NL-3000 CA Rotterdam, The Netherlands, [6]CAGT-Center for Applied Genotyping Munich, Max-Planck-Institut fuer Psychiatrie, Kraepelinstrasse 2-10, D-80804 Munich, Germany.

[#] Correspondence should be addressed to J.P.d.W. (j.dewinter@vumc.nl), J.R. (j.rouse@dundee.ac.uk) or D.S. (schindler@biozentrum.uni-wuerzburg.de)

Stoepker *et al.* Supplementary Table 1

Table S1A. SLX4 specific primers used for sequencing genomic DNA

SLX4ex1F	TGTTTAACCACAGGCCCAAT
SLX4ex1R	GCCCTTTCCAGGAAGTTTTC
SLX4ex2F	ACCAACAAGCAACCAGTCCT
SLX4ex2R	ATCCAGTGAAGTGGCAAAGG
SLX4ex3F	TTCCCGGAGTGCTGATTAGT
SLX4ex3R	ACAACAAAGCTGAGGTGCTG
SLX4ex4AF	GACCCACATTTGCTCCAATC
SLX4ex4AR	AGGGCTCTTTTTCCCTCCA
SLX4ex4BF	TTCAGGCTGTGCGGCTGCAG
SLX4ex4BR	TGCTGCGATTACAGGTGTGAACTAC
SLX4ex5F	AACTTCTGGCCTGGAATTGA
SLX4ex5R	ATACCGGGGGTTTCTTCTTG
SLX4ex6F	CCAGAAGCAGGTTTGTGTGA
SLX4ex6R	CCTTCCTGGACTTTCCATCA
SLX4ex7F	ATGTGATGGCTTCTGCAGTG
SLX4ex7R	AGAGGATTCACTCGCTGTGG
SLX4ex8F	TCTCTTACCTCCCTGGTGGA
SLX4ex8R	CTCACGGATGTCAGGATGTG
SLX4ex9F	GGGTCACTCAGAGGTTGAGG
SLX4ex9R	GCAGGAAGTGAGGGAGAGTG
SLX4ex10F	AGGCTGCAGTAAGCCATGAT
SLX4ex10R	CTGGTCATGGACTTGGGATT
SLX4ex11AF	TGTTTCTGGCAAGGAGTGTG
SLX4ex11AR	CTCCACCTTGTCCCACTGTT
SLX4ex11BF	TTTGCAGCTACTCAGCGAAA
SLX4ex11BR	TTTCTGCTTTGATGGCACAG
SLX4ex11CF	CATCACACAAGTGGGTCGTC
SLX4ex11CR	CTGATTTGGGTCTGGGAAGA
SLX4ex11DF	GCTGTTCTGTGACCGTGAGA
SLX4ex11DR	TCCAGGTTCCAGTGGTCAAT
SLX4ex11EF	GGAGACAGTGACGATGAGCA
SLX4ex11ER	CCCTGGGTCTCATGACTGAT
SLX4ex12F	GTGGCCCATTGTCTCAAGTT
SLX4ex12R	ACCAGACCCAGAGACCACAC
SLX4ex13F	ATAGGGAACGTGGAGTGTGG
SLX4ex13R	GACGGGGGTTTTTGAAGATT
SLX4ex14F	GGACCCGTAGACACCTTCCT
SLX4ex14R	CTCCAAATGCCACCCTAGAA

**Table S1B.** SLX4 specific primers used for sequencing cDNA

|	
1Afor	CAGTACTTTTTGTTCAATTGTGCAAACTC
1Arevseq	TTATCTGAGTGCCGTTTGAGGCAG
1Arev	TCACGTTCATGGCTGAGAGGTTC
1Bfor	TCATCTGTCTGCCTGTCCTG
1Brev	GCATTCGCTGTAGGACCAAT
2for	CCAAAGGATCCTCAAGAGGAGATG
2rev	CACAGAAAGCTCTGCTTGCGTTC
3for	AAGTGGAATTGTCTAGCACGCCAC
3rev	CATCTGCCCACATTGACCTCAAG
4for	GTGCCTATTGCCACTGACTCAG
4rev	GATTTCTGAGATCTGGAGCTCGAAT
5for	GCATCCTCACGCTGTCTAAAGAG
5rev	CAGCAGTCGTCAATTGGAATTGGG
6for	TTCAGCAGGCGGTTCCTGAAAC
6rev	GTGTGATGCTTTCATGATGCTTCC
7for	TAGCAGGCCGAGCTTTCTGAATTC
7forseq	GAAGCATCATGAAAGCATCACACC
7rev	AAATGCCTGTGGAGGCCTGAC

Stoepker *et al.* Supplementary Figure 1
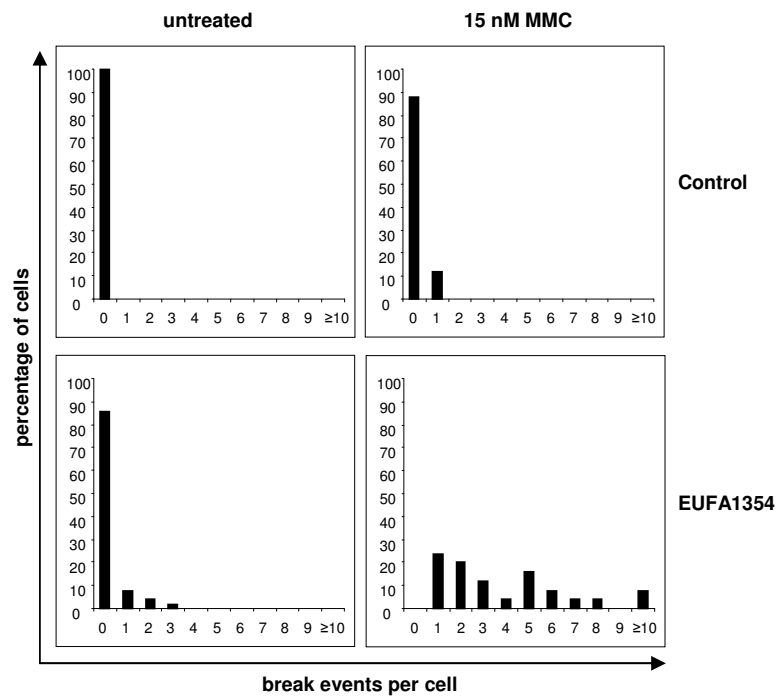
Supplementary Fig. 1. Spontaneous and MMC-induced chromosomal breakage in PHA-stimulated lymphocyte cultures from a healthy control and the affected individual (EUFA1354). Percentages of cells with up to ≥ 10 break events per cell are shown. In EUFA1354 lymphocytes treated with 15 nM MMC, all metaphases were aberrant and exhibited 1 to ≥ 10 break events per cell; at 300 nM MMC 100% of the metaphases had ≥ 10 break events per cell (results not shown).

# Stoepker *et al.* Supplementary Figure 2
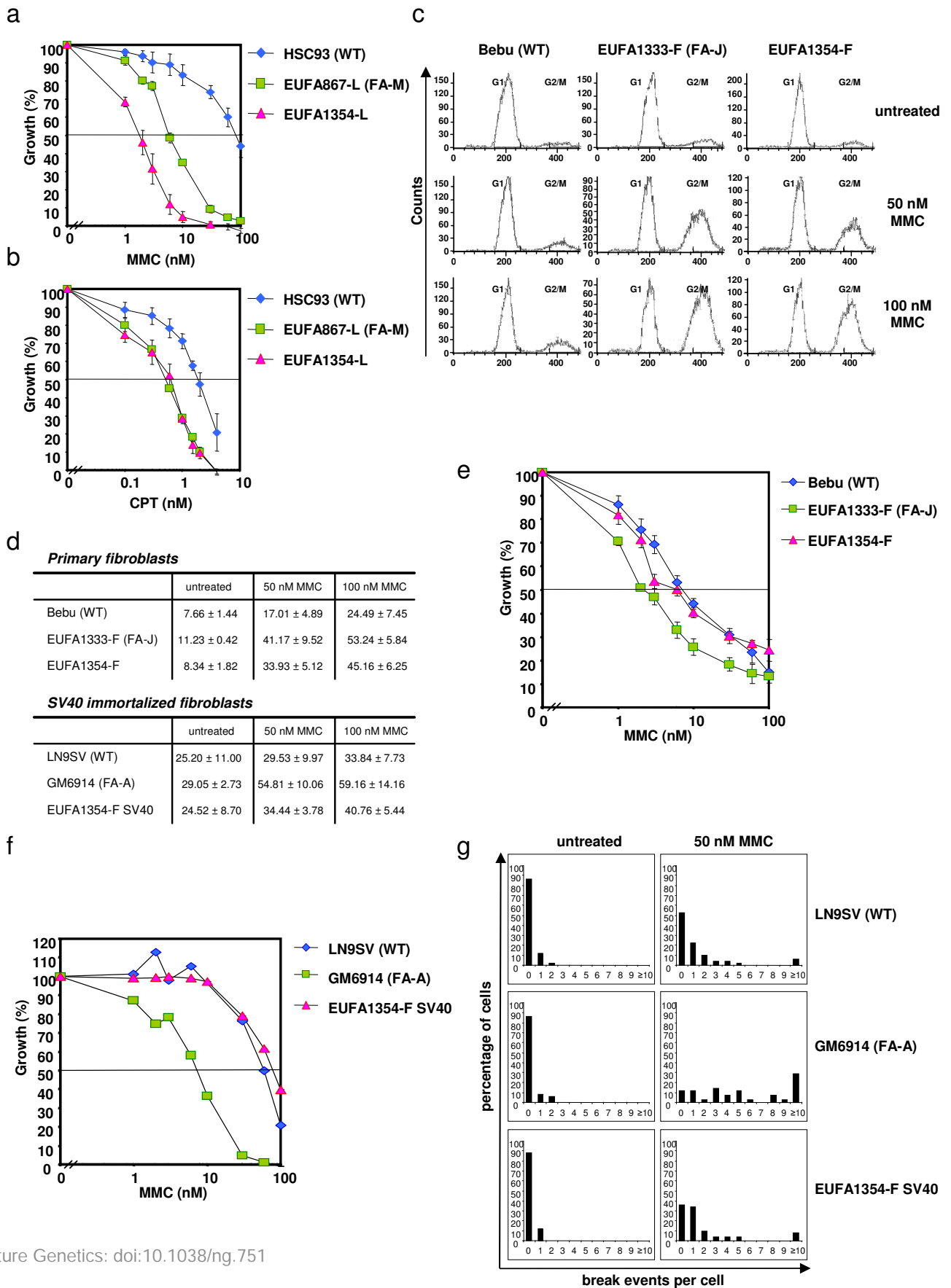
Supplementary Fig. 2. DNA cross-linker and camptothecin sensitivity in EUFA1354 lymphoblasts and fibroblasts. (**a**) Growth inhibition of EUFA1354 lymphoblasts upon exposure to mitomycin C (MMC) or (**b**) camptothecin (CPT). Lymphoblasts from a healthy individual (HSC93) and a FANCM-deficient individual (EUFA867-L) were analyzed as controls. Data represent means and standard errors of the mean from at least three independent experiments. (**c**) Cell cycle analysis of primary EUFA1354 fibroblasts reveals G2/M arrest after treatment with mitomycin C (MMC). Wild type fibroblasts (Bebu) and FANCJ-deficient fibroblasts (EUFA1333-F) were used as controls. (**d**) MMC-induced G2/M arrest in primary and SV40-immortalized EUFA1354 fibroblasts. Percentages of cells in G2/M arrest are shown. The results are the mean of two independent experiments, with standard errors of the mean. Wild type fibroblasts (Bebu and LN9SV) and FA-deficient fibroblasts (EUFA1333-F (FA-J) and GM6914 (FA-A)) were used as controls. (**e**) MMC-induced growth inhibition in primary fibroblasts. Data represent means and standard errors of the mean for three experiments. (**f**) MMC-induced growth inhibition in SV40-immortalized fibroblasts. A representative result of 3 independent experiments is shown. (**g**) Spontaneous and MMC-induced chromosomal breakage in SV40-immortalized fibroblasts. Percentages of cells with up to ≥ 10 break events per cell are shown. The number of EUFA1354 fibroblasts with zero MMC-induced break events per cell is significantly different from wild type cells (p=0.05, two-sample Chi2 test).

Stoepker *et al.* Supplementary Figure 3
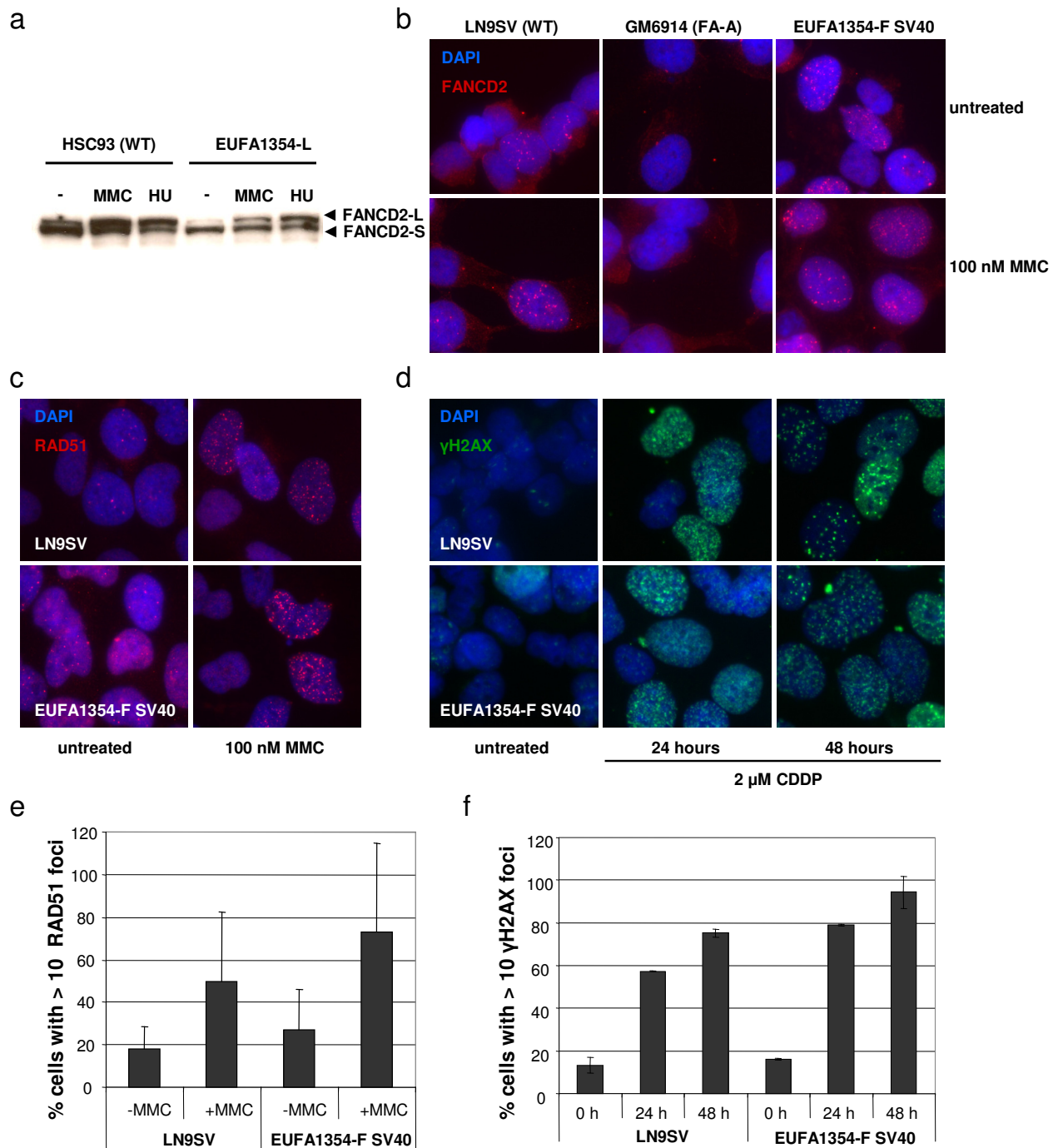
Supplementary Fig. 3. Normal FANCD2 monoubiquitination and FANCD2, RAD51 and γ-H2AX focus formation in EUFA1354. (**a**) Immunoblotting revealed induction of FANCD2 monoubiquitination after treatment with mitomycin C (MMC) or hydroxyurea (HU) in EUFA1354-L lymphoblasts. (**b**) Immunofluorescence demonstrated normal induction of FANCD2 foci (red) in SV40-immortalized EUFA1354 fibroblasts. Wild type (LN9SV) and FANCA-deficient (GM6914) fibroblasts were included as a positive and negative control, respectively. DAPI (blue) was used as a nuclear counterstaining. (**c**) MMC-induced RAD51 foci (red) in SV40-immortalized EUFA1354 fibroblasts and LN9SV wild type fibroblasts. (**d**) Cisplatin-induced γ-H2AX foci (green) in SV40-immortalized EUFA1354 fibroblasts and LN9SV wild type fibroblasts. (**e+f**) Quantification of the results in c and d. Data represent means and standard errors of the mean for two experiments.

Stoepker *et al.* Supplementary Figure 4
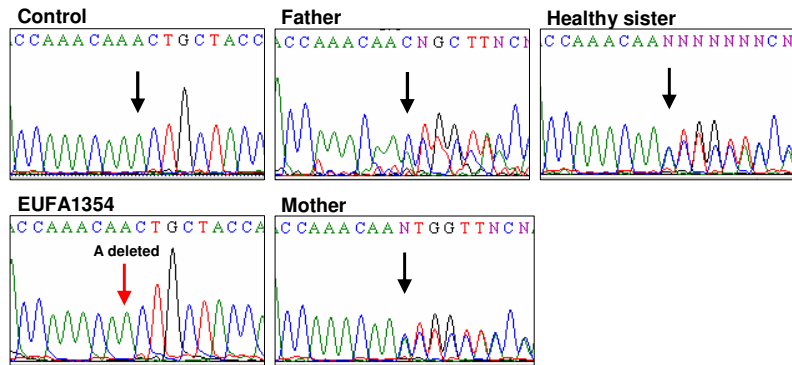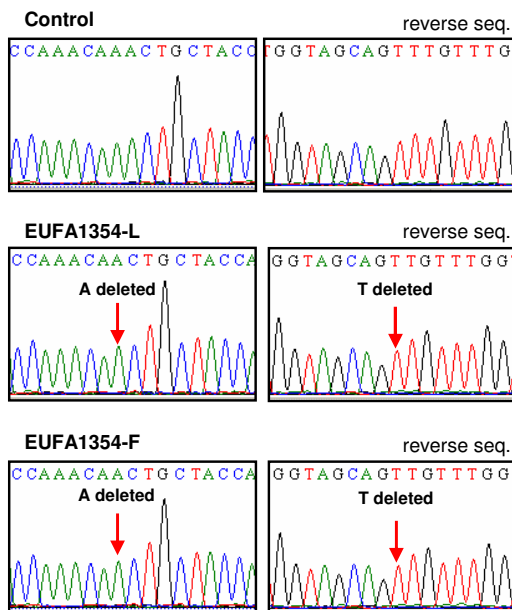
Supplementary Fig. 4. *SLX4* mutations in EUFA1354. (**a**) Sequence analysis of genomic blood DNA of EUFA1354 revealed a homozygous 1-bp deletion in exon 1 of *SLX4* (c.286delA), resulting in a frameshift at codon 96 and premature stop at codon 126 (p.T96LfsX30). The parents (second cousins) and a healthy sister are all heterozygous for the mutation. (**b**) Sequence analysis on cDNA from EUFA1354 lymphoblasts and fibroblasts detected the homozygous 1-bp deletion in *SLX4* (c.286delA).

Stoepker *et al.* Supplementary Figure 5
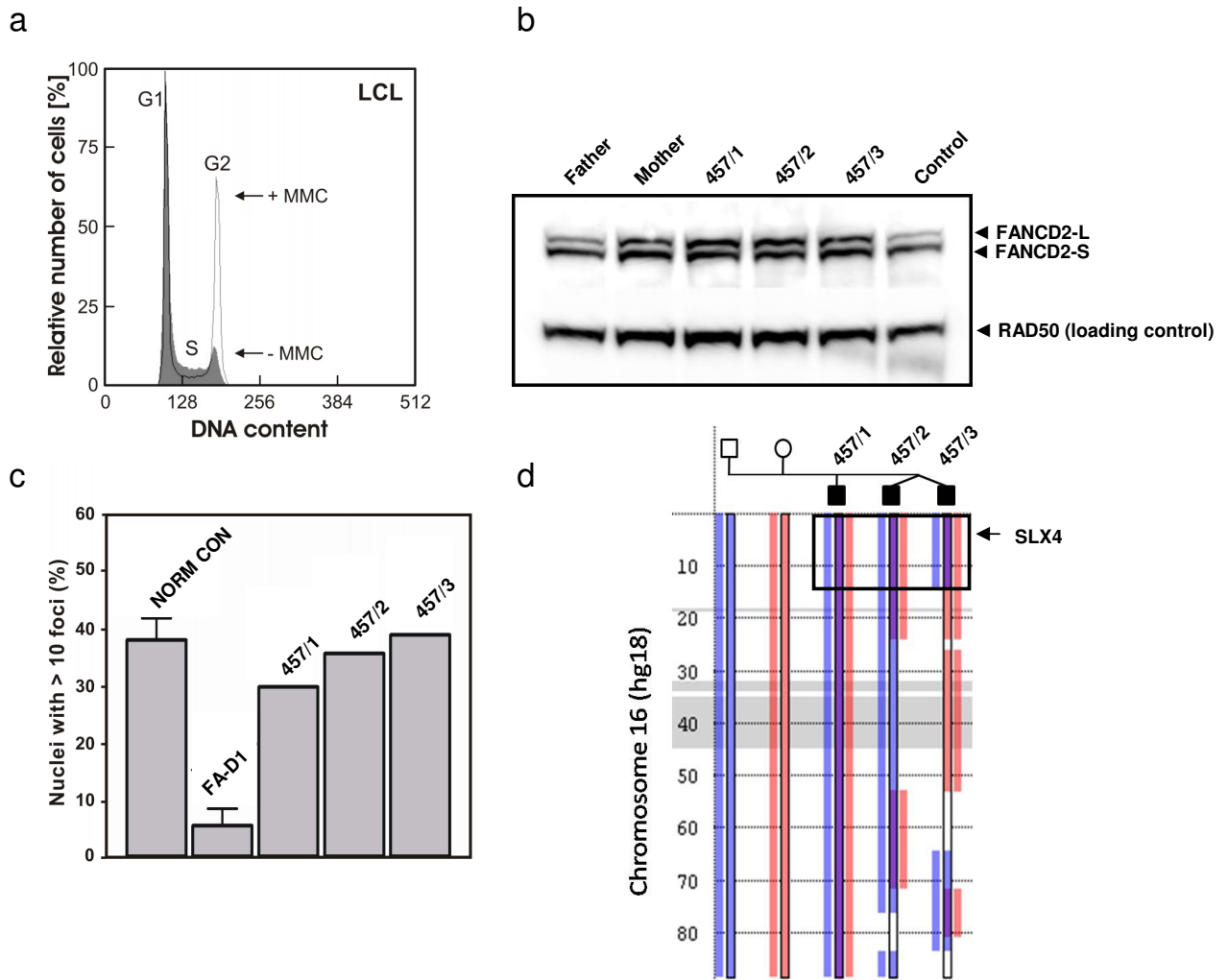
Supplementary Fig. 5. Characteristics of the German SLX4 deficient family. (**a**) G2 phase arrest in the oldest affected German sibling (457/1). Lymphoblast cultures were left untreated (grey) or exposed to 45 nM mitomycin C (MMC) for 48 h (transparent overlay). As a manifestation of FA, cells show exaggerated G2 phase blockage (arrow) in response to MMC. Quantitative analysis of the cell cycle distributions reveals 60.1% G1, 30.2% S and 9.7% G2 phase without MMC and 42.2% G1, 18.3% S and 39.5% G2 phase with MMC. (**b**) Normal FANCD2 monoubiquitination in lymphoblasts from the German FA siblings (457/1-3) suggests a downstream defect in the FA/BRCA pathway. RAD50 served as loading control. (**c**) Normal formation of RAD51 foci in lymphoblasts of the German FA siblings (457/1-3) suggests properly functioning BRCA2, PALB2 and RAD51C. This was confirmed by Western blotting (BRCA2 and PALB2) or sequencing (RAD51C). FANCJ protein levels were also normal. (**d**) Graphic presentation of linkage analysis in the German family by the program Phaser. The disease genotype of chromosome 16 in this family is defined by the first affected sibling (457/1, violet central bar) between the corresponding parental alleles (left paternal blue bar, right maternal red bar). Regions of common linkage (violet) are shown for the second affected sibling (457/2) compared to the first, and for the third affected sibling (457/3) compared to the first. A region common to all three of them extends from 0 to 13.5 K (box). SLX4 lies therein at 3.5-3.6 K (arrow). Regions where only one parental mutation allele has been inherited are shown in that colour. Gaps between the bars characterize regions where the non-mutated alleles have been transmitted. Gray is the centromeric region. This analysis also reveals that the twin siblings are dizygotic.

Stoepker *et al.* Supplementary Figure 6
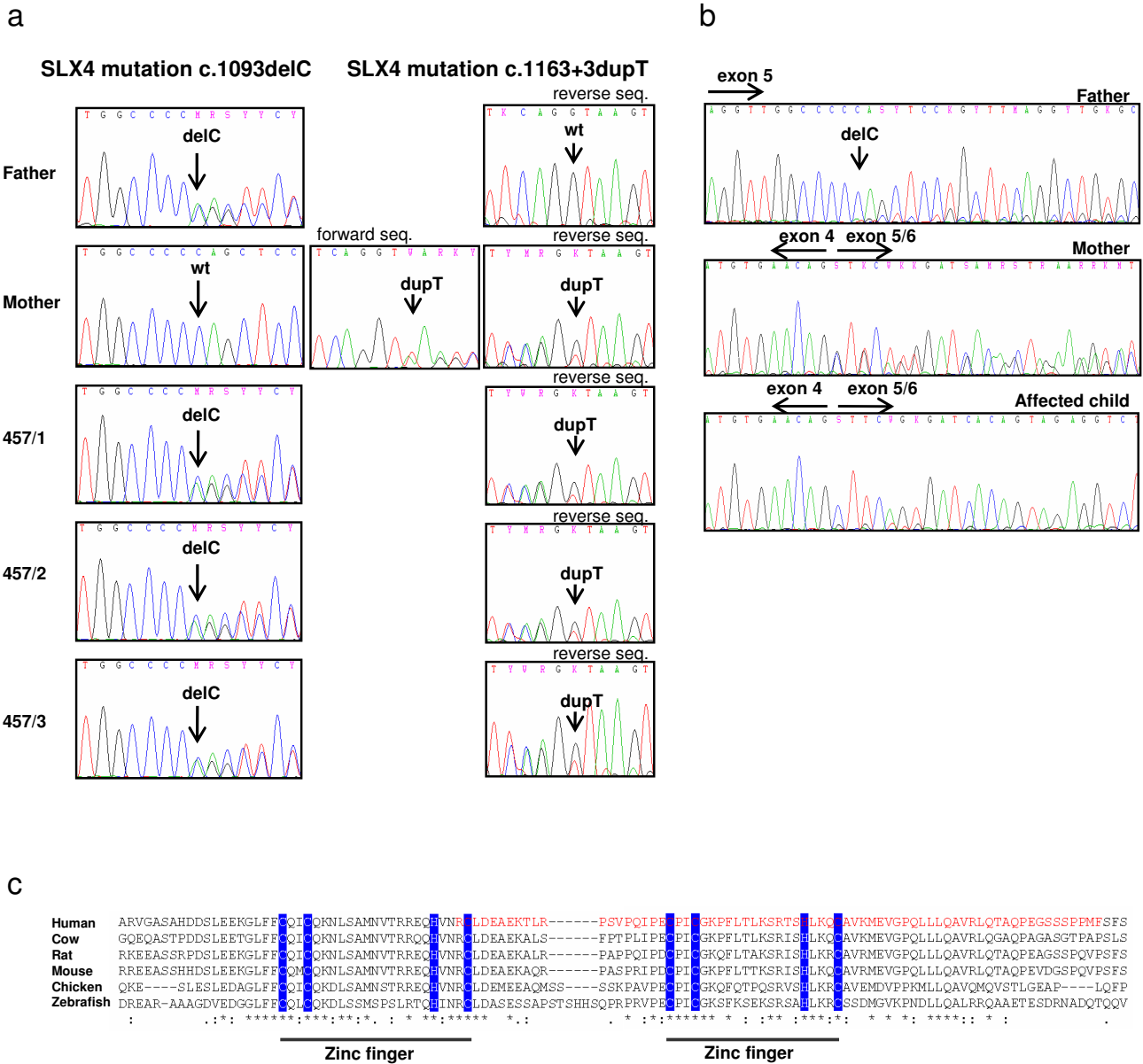


Supplementary Fig. 6. *SLX4* mutations in the unclassified German family. (**a**) Sequence analysis of genomic DNA revealed a 1-bp deletion (c.1093delC) and a splice site mutation (c.1163+3dupT) in the three affected siblings. The deletion was detected in genomic DNA from the father; the splice site mutation was present in genomic DNA from the mother. Only reverse sequences identify the duplication unequivocally in the affected siblings as forward sequences have superimposed sequence resulting from the paternal deletion. "Splicefinder" software (http://www.splicefinder.net/form.php) predicts a decrease of the splice donor score from 20.8 (wild type) to 11.9 (c.1163+3dupT). (**b**) Sequence analysis on cDNA of affected sibling 3 (457/3) shows that c.1163+3dupT causes exon 5 skipping. Underlying sequence without exon 5 skipping from the other mutant allele is minor, suggesting instability of the transcript containing c.1093delC by nonsense-mediated mRNA decay. This finding is confirmed by cDNA sequencing in the father. The transcript that lacks exon 5 sequence due to c.1163+3dupT is stable as indicated by cDNA sequencing in the mother. (**c**) The exon 5 deletion removes amino acids 317-387 (indicated in red) from the SLX4 protein, which disrupts the conserved UBZ4 domain.

Stoepker *et al.* Supplementary Figure 7
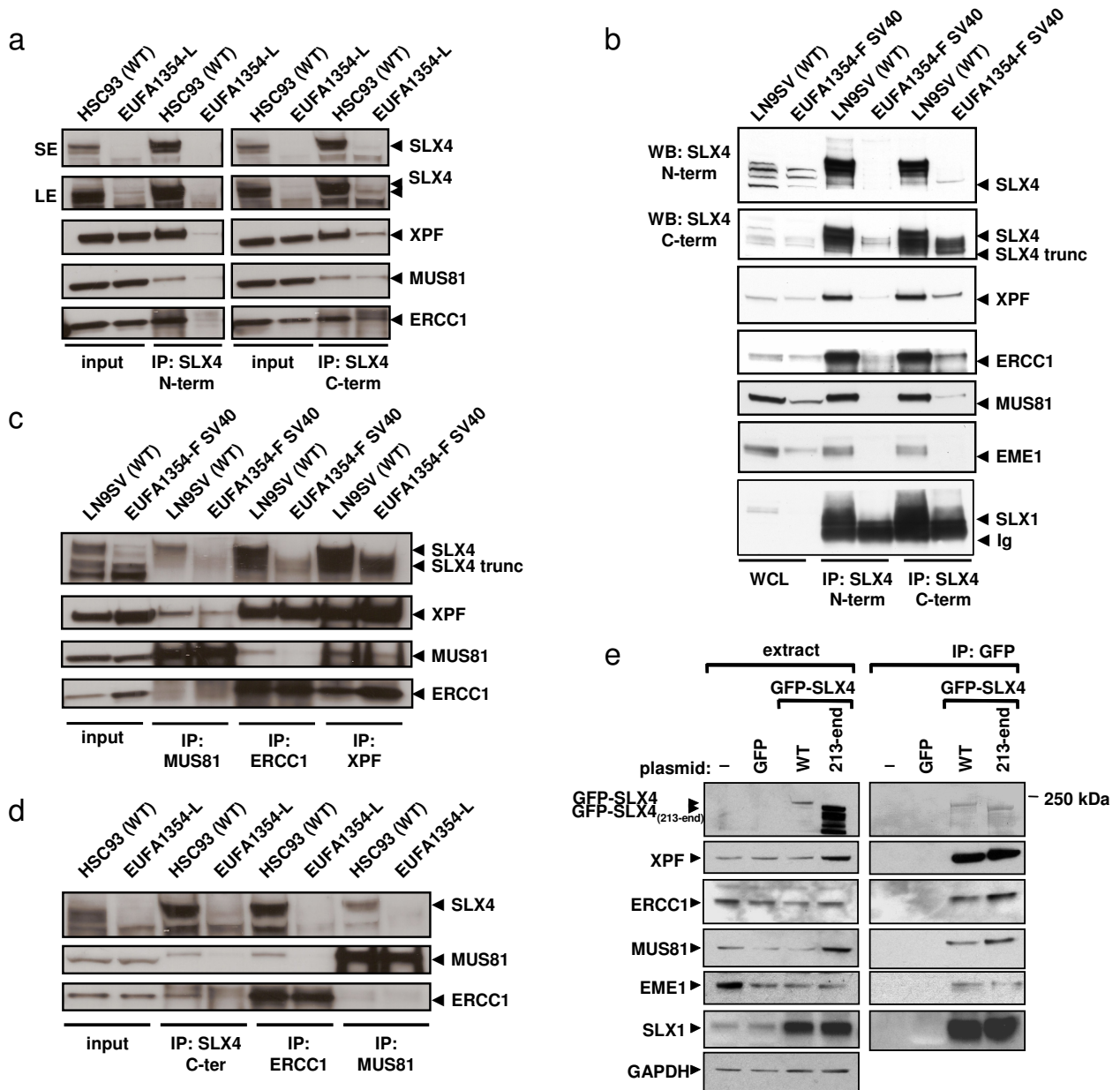
Supplementary Fig. 7. SLX4 expression and interaction with structure-specific endonucleases in EUFA1354. (**a**) Immunoprecipitation (IP) and Western blot analysis showing the absence of full-length SLX4 protein and reduced levels of ERCC1, XPF and MUS81 in SLX4 precipitates from lymphoblasts of the affected individual (EUFA1354-L). Immunoprecipitation was performed with antibodies against the SLX4 N-terminus (1-300) or C-terminus (1534-1834) and Western blotting with the antibody against the SLX4 C-terminus. SE and LE are short and long exposures of the blot. (**b**) Immunoprecipitation (IP) and Western blot analysis revealing absence of full-length SLX4 protein and reduced levels of XPF, ERCC1, MUS81, EME1 and SLX1 in SLX4 precipitates of SV40-immortalized EUFA1354-F fibroblasts. A truncated SLX4 protein is detected. Immunoprecipitation and Western blotting were performed with antibodies against the N-terminus (1-300) or C-terminus (1534-1834) of SLX4. Ig indicates the immunoglobulin light chain. (**c**) Reciprocal immunoprecipitations showing the absence of full-length SLX4 in MUS81, ERCC1 or XPF precipitates from SV40-immortalized EUFA1354 fibroblasts. The truncated SLX4 protein does coprecipitate. SV40-immortalized fibroblasts from a healthy individual (LN9SV) were analyzed as a control. (**d**) Reciprocal immunoprecipitations showing undetectable SLX4 protein in ERCC1 and MUS81 precipitates from EUFA1354-L lymphoblasts. (**e**) A truncated SLX4 protein starting at methionine 213 is able to interact with structure specific endonucleases. HEK293 cells were transiently transfected with pcDNA5.1 encoding GFP, GFP–SLX4 full–length or GFP–SLX4 (213–end) or with empty vector (–). After 48 hr, cells were lysed and extracts were subjected to immunoprecipitation with anti–GFP antibodies. Precipitates were analysed by Western blotting with the antibodies indicated. "Input" represents cell extracts.

Stoepker *et al.* Supplementary Figure 8
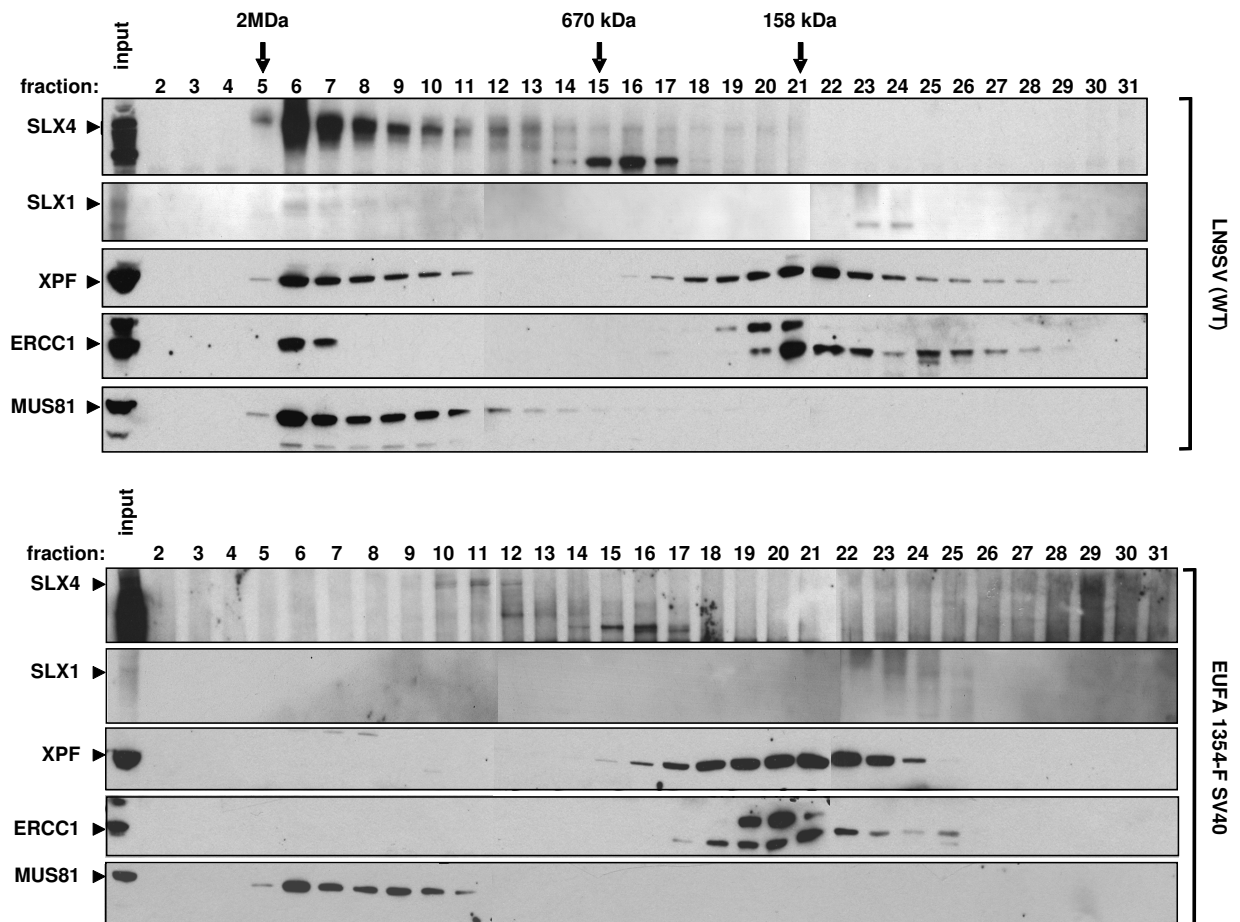
Supplementary Fig. 8. XPF/ERCC1 and SLX1 elute only in the low molecular weight fractions from EUFA1354 fibroblast lysates. Extracts of wild type fibroblasts (LN9SV) or fibroblasts from individual EUFA1354 were analyzed by size exclusion chromatography on a Superose 6 10/300 GL column in buffer containing 0.2 M NaCl. The elution positions of Dextran blue (2 MDa), thyroglobulin (670 kDa) and bovine γ–globulin (158 kDa) are shown.

Stoepker *et al.* Supplementary Figure 9
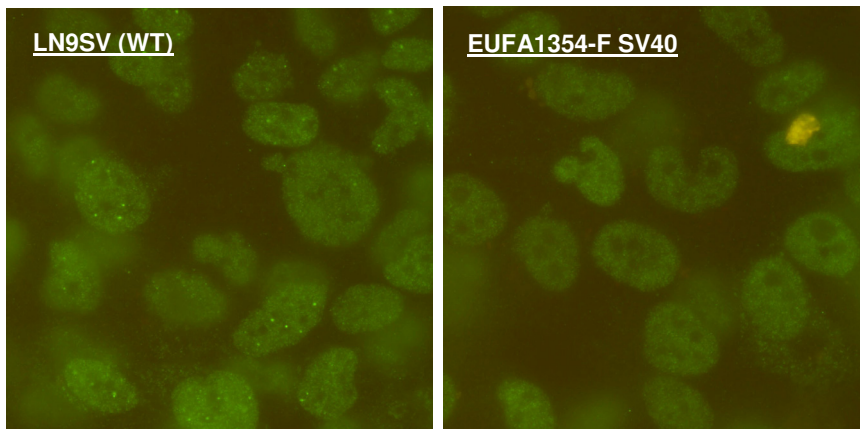
Supplementary Fig. 9. Nuclear ERCC1 foci in SV40-immortalized wild type fibroblasts, but not in SV40-immortalized EUFA1354 fibroblasts. These analyses were carried out with highly specific affinity-purified ERCC1 polyclonal antibody[7].

3. RESULTS

3.1 FA CANDIDATE GENES

3.1.1 A HISTONE-FOLD COMPLEX AND FANCM FORM A CONSERVED DNA-REMODELING COMPLEX TO MAINTAIN GENOME STABILITY

3.1.2 ON THE ROLE OF FAN1 IN FANCONI ANEMIA

3.2 IDENTIFICATION OF NOVEL FA GENES

3.2.1 MUTATION OF THE RAD51C GENE IN A FANCONI ANEMIA-LIKE DISORDER

3.2.2 SLX4, A COORDINATOR OF STRUCTURE-SPECIFIC ENDONUCLEASES, IS MUTATED IN A NEW FANCONI ANEMIA SUBTYPE

3.2.3 XPF MUTATIONS SEVERELY DISRUPTING DNA INTERSTRAND CROSSLINK REPAIR CAUSE FANCONI ANEMIA

3.3 GENOTYPING FANCONI ANEMIA BY NEXT GENERATION SEQUENCING

3.3.1 WHOLE EXOME SEQUENCING REVEALS NOVEL MUTATIONS IN THE RECENTLY IDENTIFIED FANCONI ANEMIA GENE SLX4/FANCP

3.3.2 GENOTYPING FANCONI ANEMIA BY WHOLE EXOME SEQUENCING: ADVANTAGES AND CHALLENGES

XPF mutations severely disrupting DNA interstrand cross-link repair cause Fanconi anemia

Massimo Bogliolo[1,2,#], Beatrice Schuster[3,#], Chantal Stoepker[4], Burak Derkunt[5], Yan Su[5], Anja Raams[6], Juan P. Trujillo[1], Jordi Minguillón[1], María J. Ramírez[1,2], Roser Pujol[1,2], José A. Casado[2,7], Rocío Baños[2,7], Paula Rio[2,7], Kerstin Knies[3], Sheila Zúñiga[8], Javier Benítez[2,9], Juan A. Bueren[2,7], Nicolaas G.J. Jaspers[6], Orlando D. Schärer[5], Johan P. de Winter[4], Detlev Schindler[3,*] & Jordi Surrallés[1,2,*]

[1]Genome Instability and DNA Repair Group, Department of Genetics and Microbiology, Universitat Autònoma de Barcelona, Bellaterra, Barcelona, Spain.

[2]Centre for Biomedical Network Research on Rare Diseases (CIBERER), Instituto de Salud Carlos III, Bellaterra, Barcelona, Spain.

[3]Department of Human Genetics, University of Wurzburg, Wurzburg, Germany.

[4]Department of Clinical Genetics, VU University Medical Center, Amsterdam, The Netherlands.

[5]Department of Pharmacological Sciences and Chemistry, Stony Brook University, Stony Brook, New York, USA.

[6]Department of Genetics, Erasmus University Medical Center, Rotterdam, The Netherlands

[7]Hematopoiesis and Gene Therapy Division, Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas (CIEMAT), Madrid, Spain.

[8]Department of Bioinformatics. Sistemas Genómicos SL, Valencia, Spain.

[9]Human Genetics Group, Spanish National Cancer Center, CNIO, Madrid, Spain.

#contributed equally to this work

*Correspondence should be addressed to JS (jordi.surralles@uab.es) or DS (schindler@biozentrum.uni-wuerzburg.de)

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[1]. 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[4]. 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[1,5]. 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[1]. 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 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 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)[6]. 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 [7].

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 photoproducts (6-4PP) 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).

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[7]. This feature is evident for XP-causing mutations and accentuated in cells from the patient with the XFE syndrome[8]. 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 colocalization 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[9,10]. 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[10]. 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 substrate (Fig. 4b). This is unlike wild type XPF and XPF mutants causing XP (R799W) or XFE progeria (R153P)[7], 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[12-14]. 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[15]. 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[16].

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[17-20] seems to have a minor role in ICL repair and lack of FAN1 does not cause FA in humans[21,22]. 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[23]. The nuclease hSNM1B/Apollo has also been linked to the Fanconi anemia pathway via its interaction with FANCP/SLX4[24]. 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 are 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[5]. 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[7]. 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 6-4PPs. 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[7], 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[5,7], 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[25-29].

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)* [30,31], 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.

Acknowledgements

References

1. Deans, A.J. & West, S.C. DNA interstrand crosslink repair and cancer. *Nat Rev Cancer* **11**, 467-480 (2011).

2. Raya A. *et al.* Disease-corrected haematopoietic progenitors from Fanconi anemia induced pluripotent. *Nature* **460**, 53-59 (2009).

3. Rio, P. *et al.* In vivo proliferation advantage of genetically corrected hematopoietic stem cells in a mouse model of Fanconi anemia FA-D1. *Blood* **112**,4853-4861 (2008).

4. Casado, J.A. *et al*. A comprehensive strategy for the subtyping of patients with Fanconi anaemia: conclusions from the Spanish Fanconi Anemia Research Network. *J Med Genet* **44**, 241-249 (2007).

5. Gregg, S.Q. Robinson, A.R. & Niedernhofer, L.J. Physiological consequences of defects in ERCC1-XPF DNA repair endonuclease. *DNA Repair* **10**, 781-791 (2011).

6. Castella, M. *et al*. Chromosome fragility in patients with Fanconi anaemia: diagnostic implications and clinical impact. *J Med Genet* **48**, 242-250 (2011).

7. Ahmad, A. *et al.* Mislocalization of XPF-ERCC1 nuclease contributes to reduced DNA repair in XP-F patients. *PLoS Genet* **6**, e1000871 (2010).

8. Niedernhofer L.J. *et al*. A new progeroid syndrome reveals that genotoxic stress suppresses the somatotroph axis. *Nature* **444**,1038-1043 (2006).

9. Kim Y. *et al.* Mutations of the SLX4 gene in Fanconi anemia. *Nat Genet* **43**, 142-146 (2011).

10. Stoepker, C. *et al.*, SLX4, a coordinator of structure-specific endonucleases, is mutated in a new Fanconi anemia subtype. *Nat Genet* **43**, 138-141 (2011).

11. de Laat, W.L., Sijbers, A.M., Odijk, K., Jaspers, N.G. & Hoeijmakers, J.H. Mapping of interaction domains between human repair proteins ERCC1 and XPF. *Nucleic Acids Res* **26**, 4146-4152 (1998).

12. Enzlin, J.H. & Schärer, O.D. The active site of the DNA repair endonuclease XPF-ERCC1 forms a highly conserved nuclease motif. *EMBO J* **21**, 2045-2053 (2002).

13. Staresincic, L. *et al.* Coordination of dual incision and repair synthesis in human nucleotide excision repair. *EMBO J* **28**, 1111-1120 (2009).

14. Su, Y. Multiple DNA binding domains mediate the function of the ERCC1-XPF protein in nucleotide excision repair. *J Biol Chem* **287**, 21846-21855 (2012).

15. De Silva, I.U., *et al*. Defining the roles of nucleotide excision repair and recombination in the repair of DNA interstrand cross-links in mammalian cells.. *Mol Cell Biol* **20**, 7980-7990 (2000).

16. Rothfuss, A. & M. Grompe. Repair kinetics of genomic interstrand DNA cross-links: evidence for DNA double-strand break-dependent activation of the Fanconi anemia/BRCA pathway. *Mol Cell Biol* **24**,123-134 (2004).

17. Kratz, K. *et al*. Deficiency of FANCD2-associated nuclease KIAA1018/FAN1 sensitizes cells to interstrand crosslinking agents. *Cell* **142**, 77-88 (2010).

18. Liu, T. *et al*. FAN1 acts with FANCI-FANCD2 to promote DNA interstrand cross-link repair. *Science* **329**, 693-696 (2010).

19. MacKay, C. et al. Identification of KIAA1018/FAN1, a DNA repair nuclease recruited to DNA damage by monoubiquitinated FANCD2. *Cell* **142**, 65-76 (2010).

20. Smogorzewska A, Desetty R, Saito TT, et al. A genetic screen identifies FAN1, a Fanconi anemia-associated nuclease necessary for DNA interstrand crosslink repair. *Mol. Cell* **39**, 36-47(2010).

21. Trujillo, J.P. et al. On the role of FAN1 in Fanconi anemia. *Blood* **120**, 86-89 (2012).

22. Zhou, W. et al. FAN1 mutations cause karyomegalic interstitial nephritis, linking chronic kidney failure to defective DNA damage repair. Nat Genet. 2012 Jul 8. doi: 10.1038/ng.2347. [Epub ahead of print]. (2012).

23. Wang, A.T. et al. Human SNM1A and XPF-ERCC1 collaborate to initiate DNA interstrand cross-link repair. *Genes Dev*. **25**,1859-70 (2011).

24. Salewsky, B. et al. The nuclease hSNM1B/Apollo is linked to the Fanconi snemia pathway via its interaction with FANCP/SLX4. *Hum Mol Genet*, in press.

25. Weeda, G. *et al*. Disruption of mouse ERCC1 results in a novel repair syndrome with growth failure, nuclear abnormalities and senescence. *Curr Biol.* **7**, 427-439 (1997).

26. McWhir, J., Selfridge, J., Harrison, D.J., Squires, S. & Melton, D.W. Mice with DNA repair gene (ERCC-1) deficiency have elevated levels of p53, liver nuclear abnormalities and die before weaning. *Nat Genet* **5**, 217-224 (1993).

27. Tian, M., Shinkura, R., Shinkura, N. & Alt, F.W. Growth retardation, early death, and DNA repair defects in mice deficient for the nucleotide excision repair enzyme XPF. *Mol Cell Biol* **24**, 1200-1205 (2004).

28. Jaspers, N.G. *et al.* First reported patient with human ERCC1 deficiency has cerebro-oculo-facio-skeletal syndrome with a mild defect in nucleotide excision repair and severe developmental failure. *Am J Hum Genet* **80**, 457-466 (2007).

29. Gregg, S.Q. *et al*. A mouse model of accelerated liver aging due to a defect in DNA repair. *Hepatology*. doi: 10.1002/hep.24713. [Epub ahead of print]. (2011).

30. Meindl, A. *et al*. Germline mutations in breast and ovarian cancer pedigrees establish RAD51C as a human cancer susceptibility gene. *Nat Genet* **42**, 410-414. (2010).

31. Levy-Lahad, E. Fanconi anemia and breast cancer susceptibility meet again. *Nat Genet* **42**, 368-369 (2010).

Legends to figures

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.

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.

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.

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.

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.
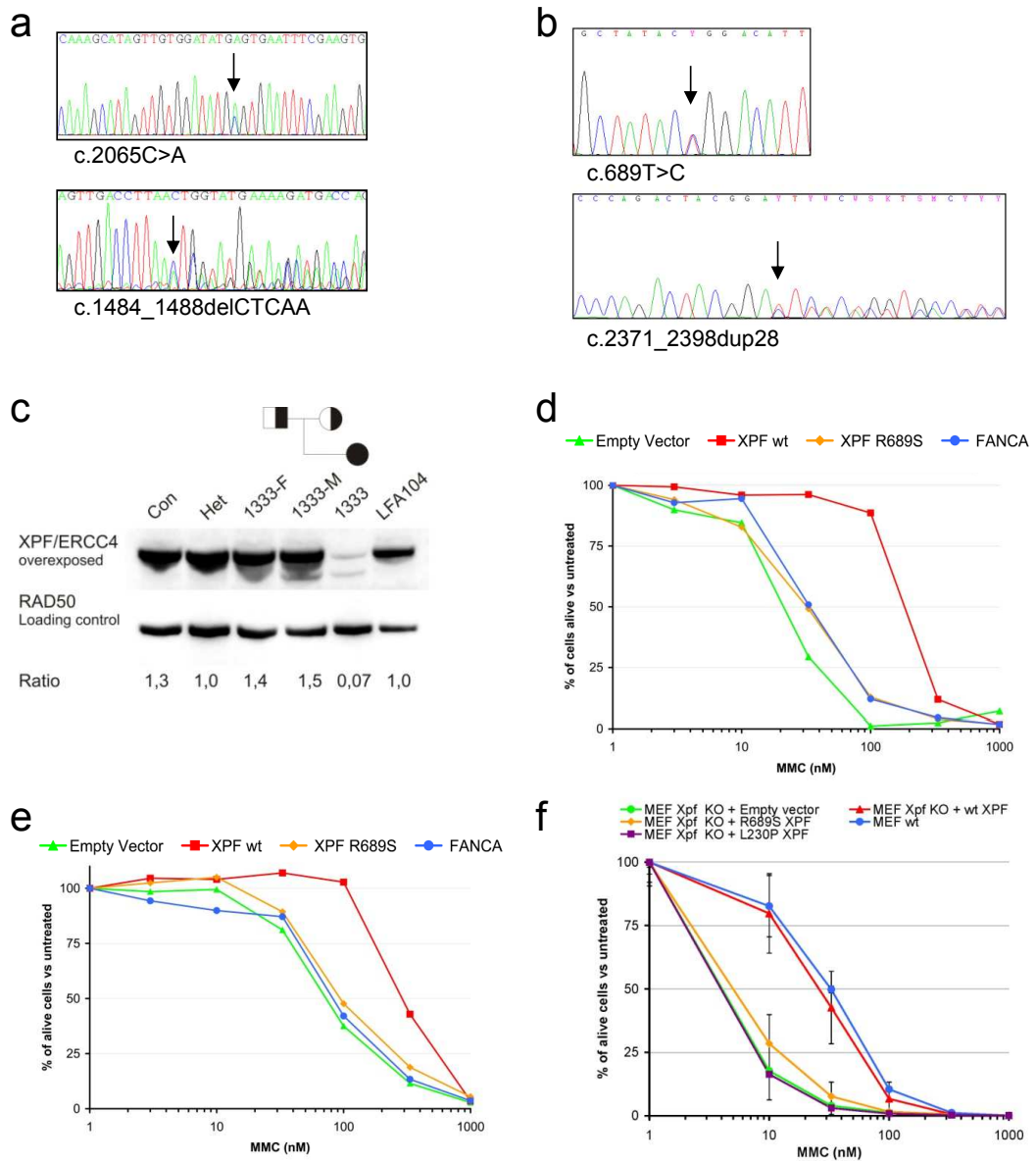
Fig.1. XPF mutations and XPF deficiency in Fanconi anemia patients
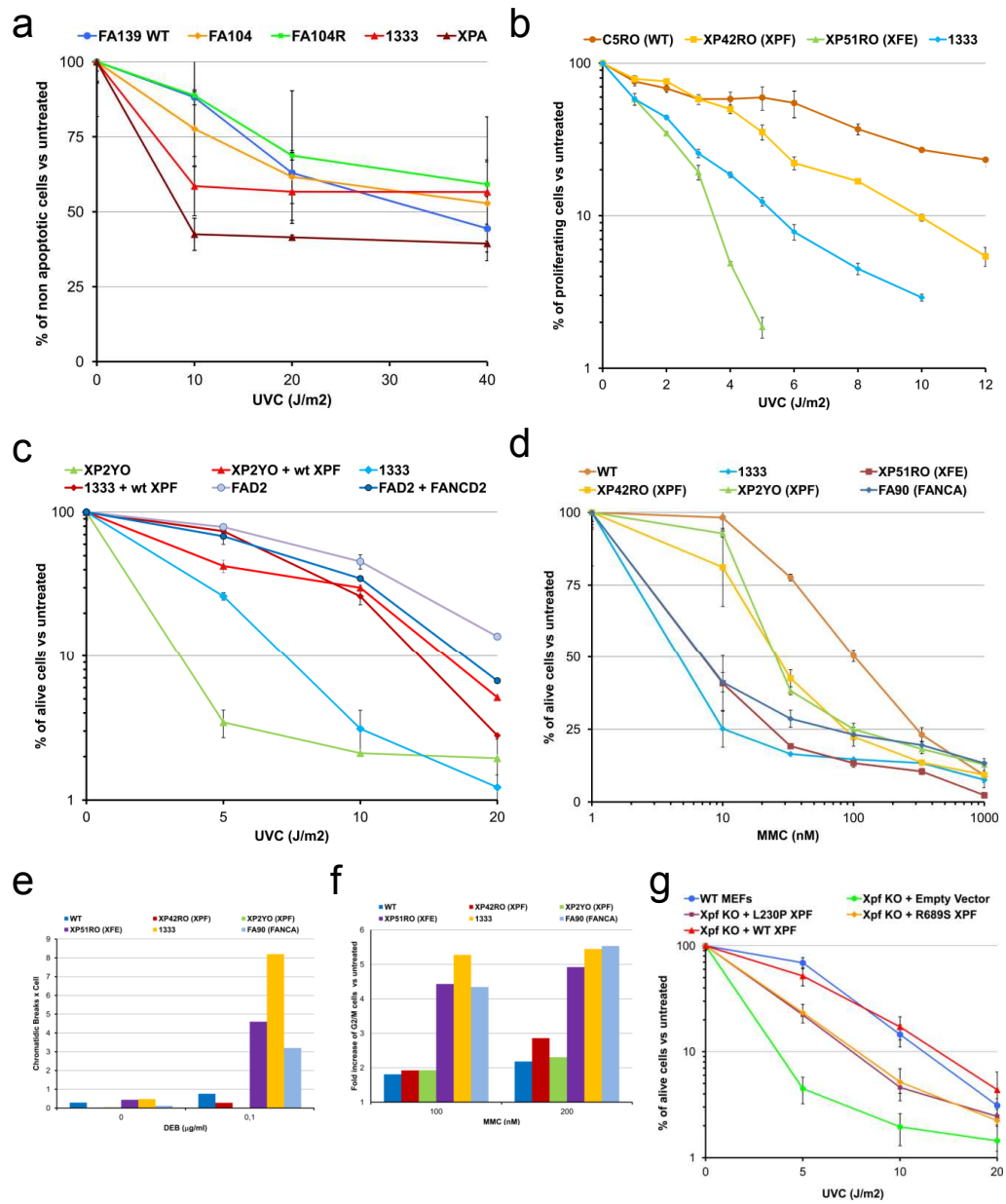
Fig.2. UV and MMC sensitivities of XPF mutants
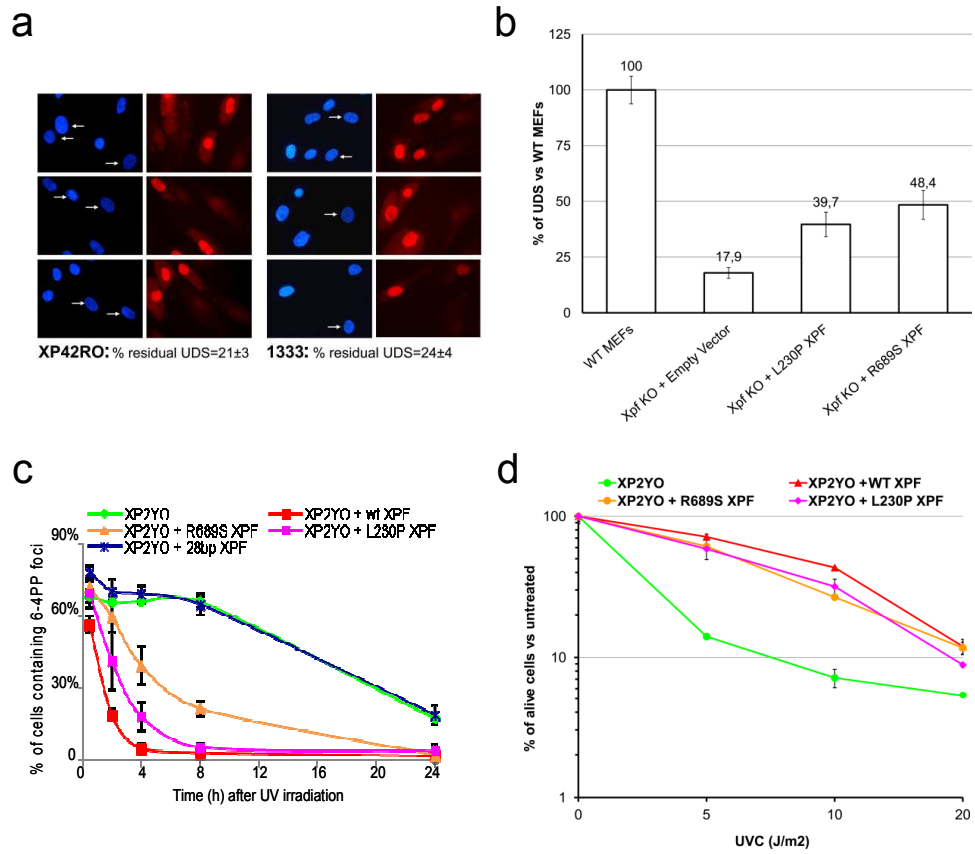
Fig.3. NER analysis of XPF mutants
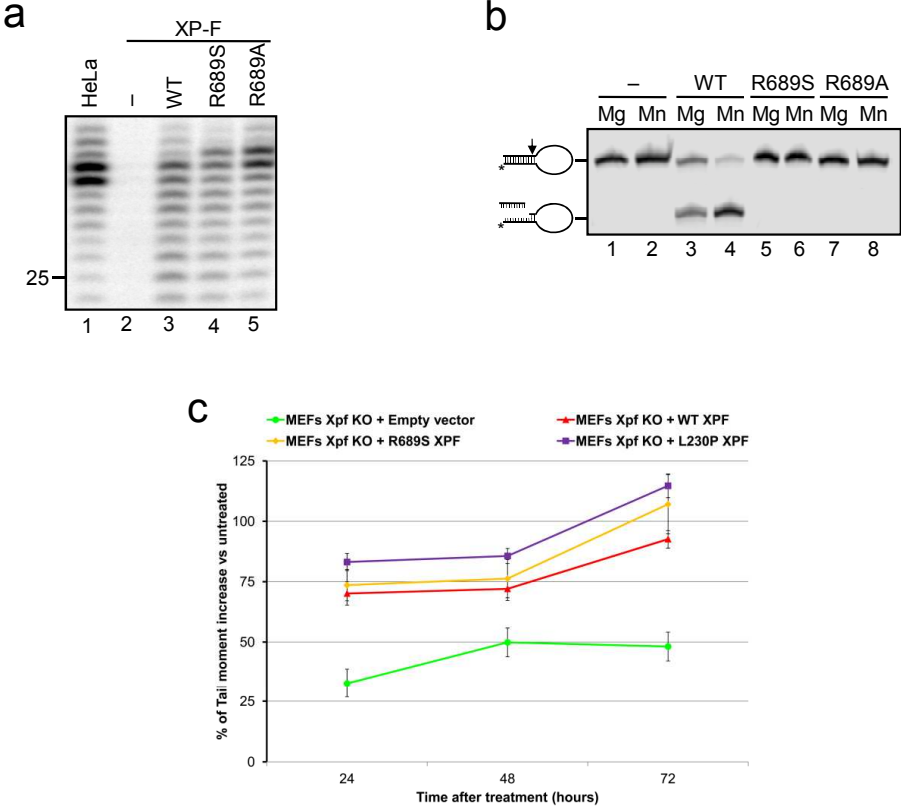
Fig.4. NER, nuclease and unhooking activities of XPF mutants
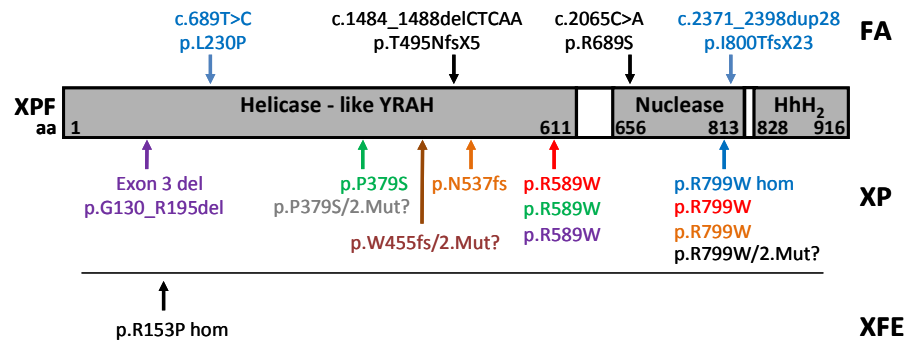
Fig.5. XPF mutations lead to three rare diseases

Supplementary information on line

XPF mutations severely disrupting DNA interstrand cross-link repair cause Fanconi anemia

Massimo Bogliolo, Beatriz 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 Surralles

Correspondence should be addressed to JS (jordi.surralles@uab.es) or DS (schindler@biozentrum.uni-wuerzburg.de)

This file includes:

Materials and Methods on line

Supplementary Text

Supplementary Figs. 1 to 4

Supplementary Tables 1 to 3

Supplementary References

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[1].

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[2,3]. 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 Eleoronics) as previously described[4]. MEFs and SV40 immortalised human fibroblasts were seeded in six wells 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% CO_2. 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[5]. 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 °C 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[6] using a mouse primary antibody to cyclobutane pyrimidine dimers (CPD) (Kamiya biomedical company) and rabbit antibody to XPF (ab76948, abcam) overnight at 4 °C. 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) photoproducts ((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%, CO$_2$. 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 µl of the cell suspension was added to 225 µl 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 Gelbond^{TM} 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: $(TMmi - TMci)/(TMci - TMc) \times 100$, 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[7] 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[9] 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 TAATATACGCGTATGGAGTCAGGGCAGCCGGCTCGACGGATTGCCATGGCGC CGCTGCTGGAGTAC and AATCTAACTAGTTCAAGCGTAATCTGGAACATCGTATGGGTAAACAACTCCG CCGTTGCATGAGG 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[6] 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[6]. 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[9,10]. For each reaction, 2 µl of repair buffer (200 mM Hepes-KOH, 25 mM MgCl$_2$, 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 °C 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 °C 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-^{32}P] dCTP for each reaction) was added. After pre-incubation at 37 ^{o}C 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 37°C 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 95°C 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)$_{22}$CCGAGCGCTGGC) 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 MgCl$_2$) by heating at 90 ^{o}C 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 MgCl$_2$ or 0.4 mM MgCl$_2$, 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 ^{o}C for 30 min and stopped by adding 10 μl 80% formamide/10mM EDTA. After heating at 95 ^{o}C 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 ^{o}C for 40 min and visualized by fluorescence by Typhoon 9400 (Amersham Biosciences).
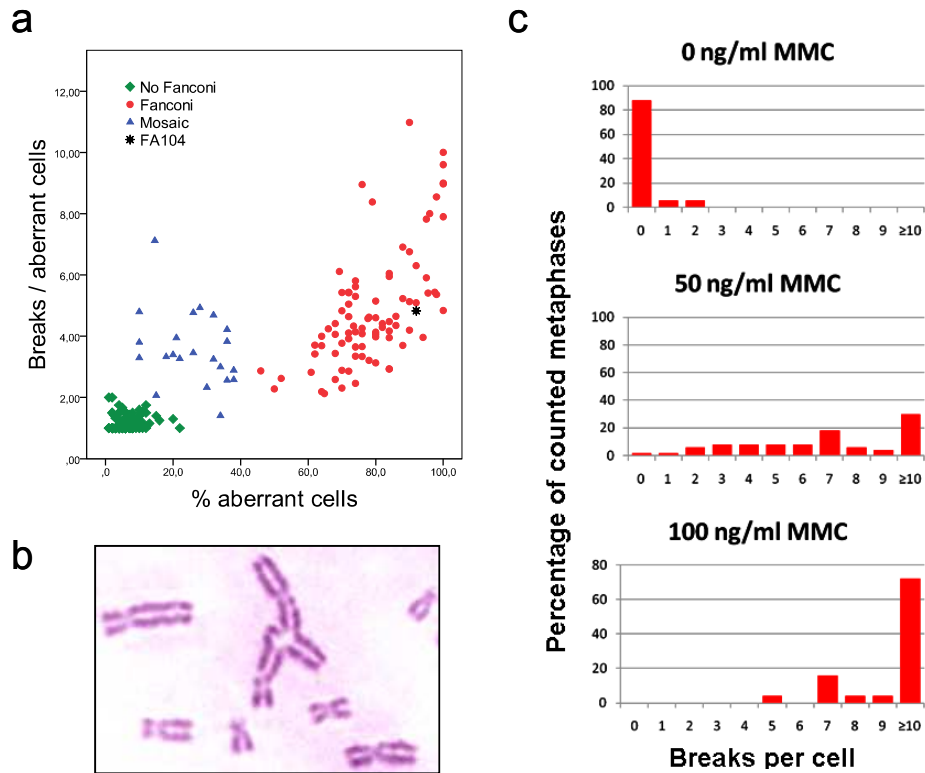
Supplementary text

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 dataset[1]. 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[11]. 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[12] 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[13] 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[14]. 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.

a

```
Homo sapiens                684   IVVDMREFRSELPSLIHRRGIDIEPVTLEVGDYILTPEMCVERKSISDLIGSL   736
Mus musculus                685   IVVDMREFRSELPSLIHRRGIDIEPVTLEVGDYILTPELCVERKSVSDLIGSL   737
Xenopus laevis              665   IIVDMREFRSELPSLIHRRGIDIEPVTLEVGDYILTPDICVERKSVSDLIGSL   717
Danio rerio                 655   IIVDMREFRSELPSLLHRRGLDIEPVTLEVGDYILTSDICVERKSVSDLIGSL   707
Drosophila melanogaster     698   VIVDMREFRSDLPCLIHKRGLEVLPLTITIGDYILTPDICVERKSISDLIGSL   750
Arabidopsis thaliana        726   VIVDMREFMSSLPNVLHQKGMKIIPVTLEVGDYILSPSICVERKSIQDLFQSF   778
Saccharomyces cerevisiae    822   VIVDTREFNASLPGLLYRYGIRVIPCMLTVGDYVITPDICLERKSISDLIGSL   874
Schizosaccharomyces pombe   668   VIVDLREFRSSLPSILHGNNFSVIPCQLLVGDYILSPKICVERKSIRDLIQSL   720
```

b

```
Homo sapiens                210   --EVVEIHVSMTPTMLAIQTAILDILNACLKELKCIINPS-LEVEDLSLENAIG   259
Mus musculus                210   --EVVEIHVSMTPAMLAIQTAILDILNACLKELKCHNPS-LEVEDLSLENALG   259
Xenopus laevis              199   --EVVELHVSMTPSMLAIQSSILDIMNACLKELKRFNPA-LEVEDLSLENAIG   248
Danio rerio                 200   --DVVELHVTLTPAMRAIQSSILDIMNACLKELKRYNPT-LEAEDLSLENSLG   249
Drosophila melanogaster     235   --QSIEMHVPISQNITSIQSHTLEIMNFLVQEIKRINRT-VDMEAVTVENCVT   284
Arabidopsis thaliana        200   --EVVDIRVSMSNYMVGIQKATIFVMDACLKEMKKTNK--VDVDDLTVESGLF   248
Saccharomyces cerevisiae    318   HNKVIEVKVSLTNSMSQIQFGLMECLKKCIAELSRKNPE-LALDWWNMENVLD   369
Schizosaccharomyces pombe   203   --NVVELNVNLSDSQKTIQSCLLTCIESTMRELRRLNSAYLDMEDWNIESALH   253
```

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**.
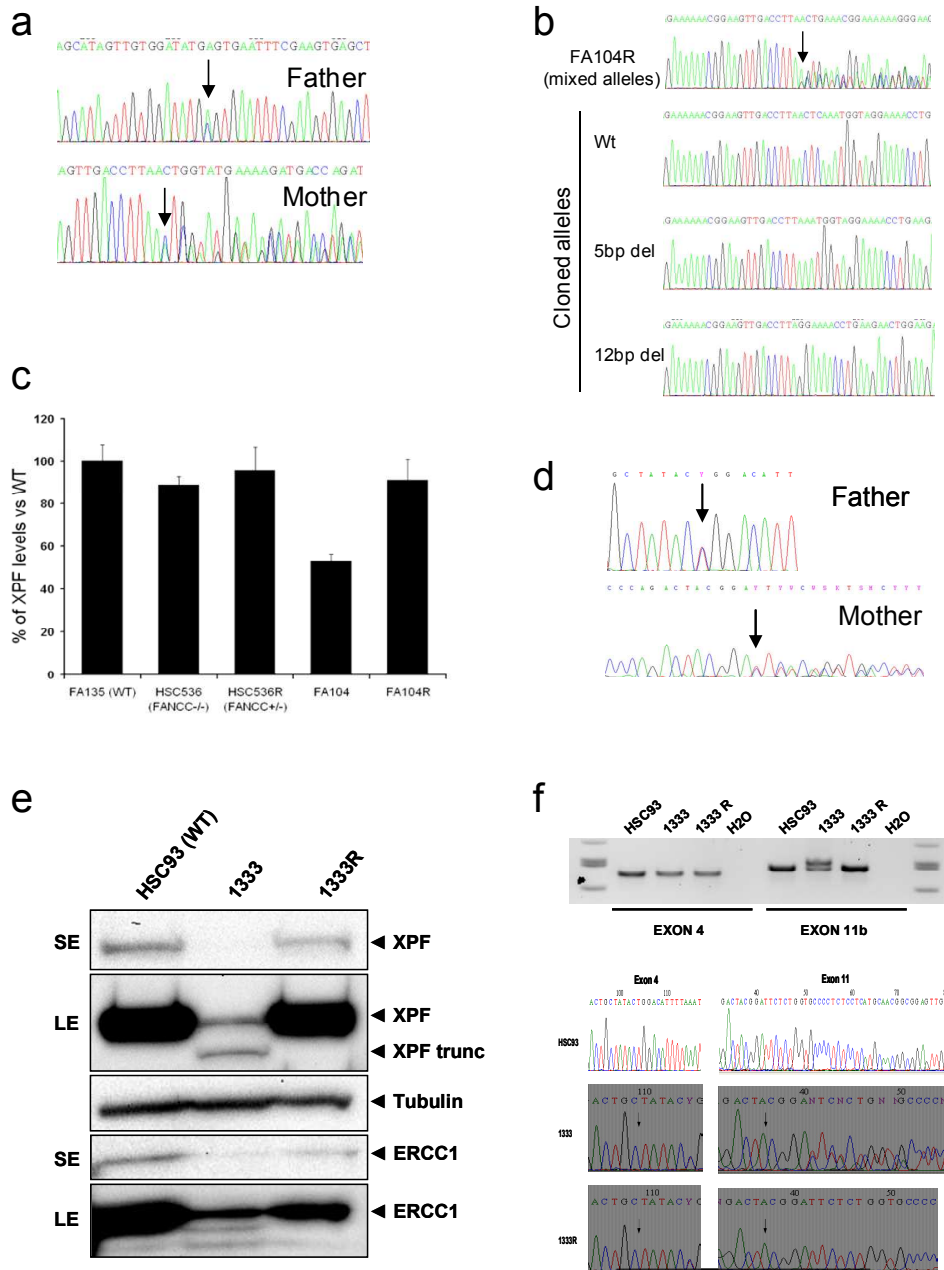
Supplementary Fig. 3. Analysis of parental DNAs and reverted cell lines. (**a**) Segregation of mutant alleles in FA104. Sequence analysis of genomic blood DNA of FA104's parents revealed that the allele with the missense mutation in exon 11

(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).
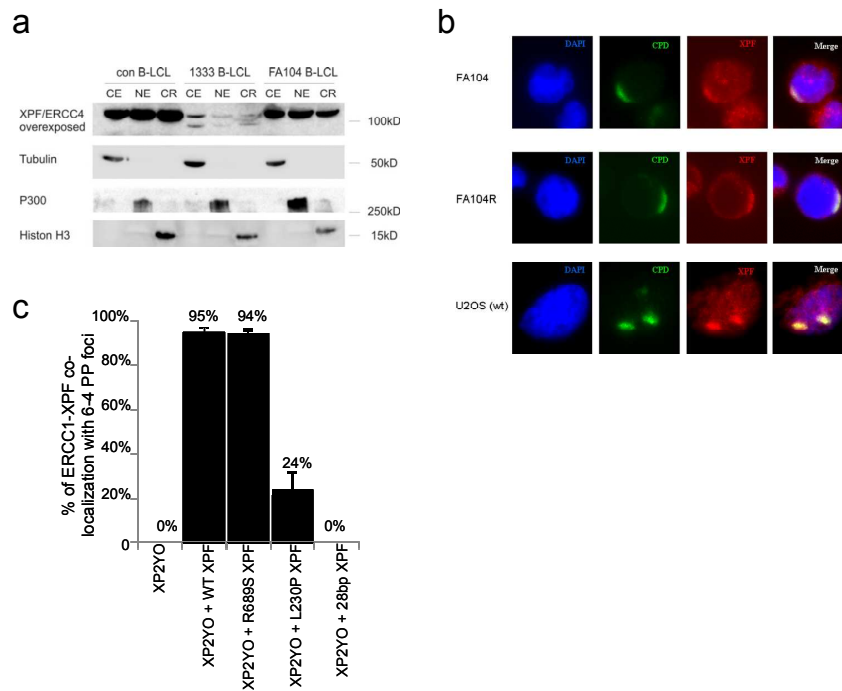
Supplementary Fig. 4. XPF cellular localization (**a**) FA104 shows an abundance and a distribution of XPF between the cytoplasmic, nuclear and chromatin compartments comparable to a normal control, whereas 1333 reveals reduced abundance and two species of that protein with sizes predicted by its mutations 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.
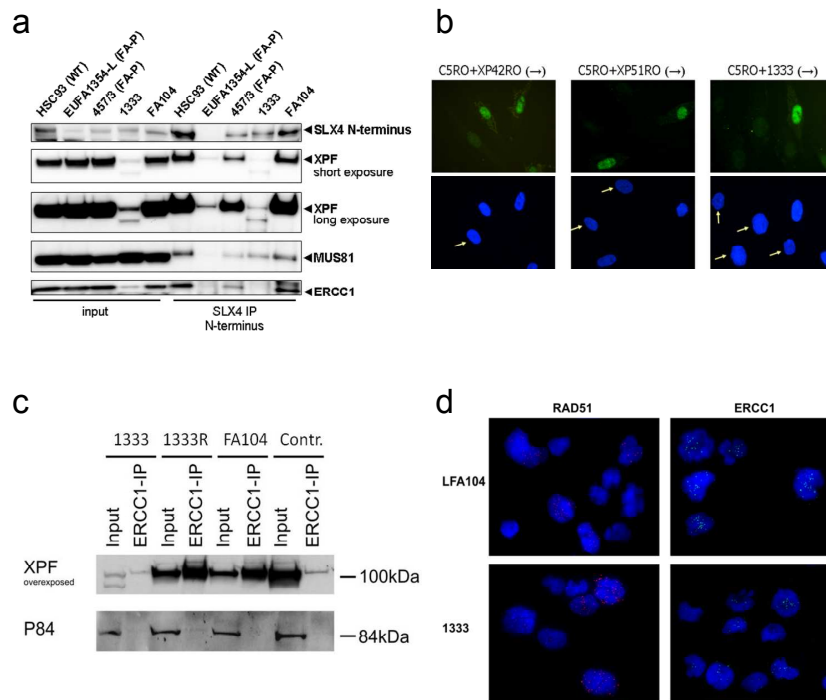
Supplementary Fig. 5. SLX4 and ERCC1 interactions in XPF-deficient FA patients (**a**) SLX4 interactions in XPF-deficient FA patients. SLX4 was immunoprecipitated with a polyclonal antibody raised against the first 300 amino acids of SLX4 (SLX4 N-terminus). Precipitated proteins were visualized by Western 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. (**d**) ERCC1 and Rad51 foci in FA104 and 1333 (foci quatification is provided within the text).

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 |

Supplementary references

1. Castella, M. *et al*. Chromosome fragility in patients with Fanconi anaemia: diagnostic implications and clinical impact. *J Med Genet* **48**, 242-250 (2011).
2. Dull, T. *et al*. A third-generation lentivirus vector with a conditional packaging system. *J Virol* **72**, 8463-8471 (1998).
3. Almarza, E. *et al*. Characteristics of lentiviral vectors harboring the proximal promoter of the vav proto-oncogene: a weak and efficient promoter for gene therapy. *Mol Ther*. **15**, 1487-1494 (2007).
4. Casado, J.A. *et al*. A comprehensive strategy for the subtyping of patients with Fanconi anaemia: conclusions from the Spanish Fanconi Anemia Research Network. *J Med Genet* **44**, 241-249 (2007).
5. Limsirichaikul, S. et al. A rapid non-radioactive technique for measurement of repair synthesis in primary human fibroblasts by incorporation of ethynyl deoxyuridine (EdU). N*ucleic Acids Res* **37**:e31 (2009).
6. Bogliolo, M. *et al*. Histone H2AX and Fanconi anemia FANCD2 function in the same pathway to maintain chromosome stability. *EMBO J* **26**, 1340-1351 (2007).
7. Enzlin, J.H. & Schärer, O.D. The active site of the DNA repair endonuclease XPF-ERCC1 forms a highly conserved nuclease motif. *EMBO J* **21**, 2045-2053 (2002).
8. Sijbers, A.M. *et al*. Xeroderma pigmentosum group F caused by a defect in a structure-specific DNA repair endonuclease. *Cell* **86**, 811-822 (1996).
9. Wood, R.D., Robins, P., & Lindahl, T. Complementation of the xeroderma pigmentosum DNA repair defect in cell-free extracts. *Cell* **53**, 97-106 (1988).
10. Moggs, J.G., Yarema, K.J., Essigmann, J.M. & Wood, R.D. Analysis of incision sites produced by human cell extracts and purified proteins during nucleotide excision repair of a 1,3-intrastrand d(GpTpG)-cisplatin adduct. *J Biol Chem* **271**, 7177-7186 (1996)
11. Milne, R.L. *et al*. ERCC4 Associated with Breast Cancer Risk: A Two-Stage Case-Control Study Using High-throughput Genotyping. *Cancer Res*. **66**, 9420 (2006).
12. Li, H. *et al*. The sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079 (2009).
13. Rios, D. *et al*. A database and API for variation, dense genotyping and resequencing data. *BMC Bioinformatics* **11**, 238 (2010).
14. Ng, S.B. *et al*. Exome sequencing identifies the cause of a mendelian disorder. *Nat Genet* **42**, 30-35 (2010).

3. RESULTS

3.1 FA CANDIDATE GENES

3.1.1 A HISTONE-FOLD COMPLEX AND FANCM FORM A CONSERVED DNA-REMODELING COMPLEX TO MAINTAIN GENOME STABILITY

3.1.2 ON THE ROLE OF FAN1 IN FANCONI ANEMIA

3.2 IDENTIFICATION OF NOVEL FA GENES

3.2.1 MUTATION OF THE RAD51C GENE IN A FANCONI ANEMIA-LIKE DISORDER

3.2.2 SLX4, A COORDINATOR OF STRUCTURE-SPECIFIC ENDONUCLEASES, IS MUTATED IN A NEW FANCONI ANEMIA SUBTYPE

3.2.3 XPF MUTATIONS SEVERELY DISRUPTING DNA INTERSTRAND CROSSLINK REPAIR CAUSE FANCONI ANEMIA

3.3 GENOTYPING FANCONI ANEMIA BY NEXT GENERATION SEQUENCING

3.3.1 WHOLE EXOME SEQUENCING REVEALS NOVEL MUTATIONS IN THE RECENTLY IDENTIFIED FANCONI ANEMIA GENE SLX4/FANCP

3.3.2 GENOTYPING FANCONI ANEMIA BY WHOLE EXOME SEQUENCING: ADVANTAGES AND CHALLENGES

Human Mutation

Whole Exome Sequencing reveals uncommon mutations in the recently identified

Fanconi anemia gene *SLX4*/*FANCP*

Beatrice Schuster[1]*, Kerstin Knies[1]*, Chantal Stoepker[2], Eunike Velleuer[3], Richard Friedl[1],

Birgit Gottwald-Mühlhauser[1], Johan P de Winter[2], Detlev Schindler[1]

[1] Department of Human Genetics, University of Wuerzburg, Wuerzburg, Germany

[2] Department of Clinical Genetics, Vrije Universteit (VU) Medical Center, Amsterdam, The Netherlands

[3] Department of Pediatric Hematology, Oncology and Clinical Immunology, University of Duesseldorf School of

  Medicine, Duesseldorf, Germany

* contributed equally

Correspondence

Beatrice Schuster, M.S.

Department of Human Genetics

University of Wuerzburg,

Biozentrum, Am Hubland,

D-97074 Wuerzburg, Germany

beatrice.schuster@biozentrum.uni-wuerzburg.de

Tel: +49-931-3184093

Fax: +49-931-3184069

1

Human Mutation

Abstract

Fanconi anemia (FA) is a rare genetic disorder characterized by congenital malformations, progressive bone marrow failure (BMF) and susceptibility to malignancies. FA is caused by biallelic or hemizygous mutations in one of 15 known FA genes, whose products are involved in the FA/BRCA DNA damage response pathway. Here we report on a patient with previously unknown mutations of the most recently identified FA gene, *SLX4*/*FANCP*. Whole Exome Sequencing (WES) revealed a nonsense mutation and an unusual splice site mutation resulting in the partial replacement of exonic with intronic bases, thereby removing a nuclear localization signal. Immunoblotting detected no residual SLX4/FANCP protein which was consistent with abrogated interactions with XPF/ERCC1 and MUS81/EME1. This cellular finding did not result in a more severe clinical phenotype than that of previously reported FA-P patients. Our study additionally exemplifies the versatility of WES for the detection of mutations in heterogenic disorders like FA.

Human Mutation

Fanconi Anemia (FA; MIM# 227650) is an autosomal or X-chromosomal recessive disorder first described in 1927 by the Swiss pediatrician Guido Fanconi (Lobitz and Velleuer, 2006). Recently, the carrier frequency in the US was estimated to be 1:181, corresponding to an incidence of FA of less than 1:100,000 (Rosenberg, et al., 2011). Higher rates have been reported for certain ethnicities or due to isolation or founder effects (Callen, et al., 2005). The clinical manifestations of FA are variable yet characteristic. Typical congenital malformations include short stature, skin hyper- or hypopigmentations, radial ray defects and malformations of ears, eyes and inner organs. Most FA patients develop progressive bone marrow failure in childhood. Furthermore, they have an increased risk of myelodysplastic syndrome (MDS) and hematological malignancies, in particular acute myelogenous leukemia (Alter and Kupfer, 1993; Rosenberg, et al., 2003; Seif, 2011). In addition, they are predisposed for solid tumors occurring in young adulthood. FA patients have an up to 700-fold increased risk for squamous cell carcinomas which arise most frequently in the mucosa of the head and neck or genital regions (Rosenberg, et al., 2008; Rosenberg, et al., 2011). The reasons for the increased susceptibility of FA patients to neoplasm are not fully understood. Most likely this is due to a DNA repair defect and genomic instability that characterize the cellular phenotype (Deans and West, 2011). FA cells show highly increased rates of chromosomal breakage especially after exposure to DNA-crosslinking agents, accumulate in the G2 phase of the cell cycle and encounter diminished survival (Auerbach, 1993; Schindler, et al., 1987; Schroeder, et al., 1964). Like the clinical phenotype, the genetic background of FA is very heterogeneous. To date 15 complementation groups (FA-A, -B, -C, -D1, -D2, -E, -F, -G, -I, -J, -L, -M, -N, -O and -P) have been delineated. The first identified FA gene, *FANCC* (MIM# 613899), was reported in 1992 (Strathdee, et al., 1992). Since 2000 nearly every year a new FA gene has been added, most recently *SLX4/BTBD12* that has been assigned the alias *FANCP* (MIM# 613278) (Kim, et al., 2011; Stoepker, et al., 2011). Its product interacts with different structure-specific endonucleases such as XPF/ERCC1 (MIM# 133520/126380),

3

Human Mutation

MUS81/EME1 (MIM# 606591/610885) and the Holliday junction resolvase SLX1 (MIM# nonexistent), by coordinating their activity in DNA repair and recombination (Andersen, et al., 2009). *FANCP* is involved in the FA/BRCA pathway and the network of DNA interstrand crosslink (ICL) repair. A key step in this pathway is FANCD2 (MIM# 613984) and FANCI (MIM# 611360) monoubiquitination through the FA core complex following DNA damage and replication fork stalling. FANCP acts downstream of these protein modifications, similar to FANCD1 (MIM# 600185), FANCJ (MIM# 605882), FANCN (MIM# 610355) and FANCO (MIM# 602774). *FANCP*-mutated cells are proficient of RAD51 foci formation, unlike FA-D1- or FA-O cells. Given these facts, a role in the coordination of DNA incision for ICL unhooking seems more likely than one in Holliday junction resolution, even though the precise function of FANCP in the FA/BRCA pathway remains elusive (Crossan, et al., 2011; Stoepker, et al., 2011). Four families with a total of six affected children have been assigned to complementation group FA-P. Their underlying *FANCP* mutations result in protein truncation and degradation. The presence of residual protein, retained function and other factors may explain the variable severity of clinical FA manifestations (Andersen, et al., 2009; Kim, et al., 2011; Stoepker, et al., 2011).

In this study, we report on an additional FA-P patient who was assigned to that complementation group due to *FANCP* mutations identified by Whole Exome Sequencing (WES). *FANCP* proved to be the only FA gene carrying compound heterozygous pathogenic sequence changes that were confirmed by Sanger sequencing. A nonsense and a splice site mutation followed Mendelian segregation in the family of the patient.

The 21 year-old girl of German descent was diagnosed with FA at the age of 5 years, showing FA-typical features including prenatal dystrophy, short stature, hypoplasia of the right thumb, microcephaly, speckles of skin hyperpigmentation at the arms and legs, minor café-au-lait and vitiligo spots, trivial mitral valve prolapse and hypothyreoidism. Apart from the platelets

4

(reduced since age 5, lowest number about 20,000/μl) her blood counts were relatively stable until she developed MDS at 19 years of age. She was successfully transplanted with hematopoietic stem cells from a 10/10 matched unrelated donor. The patient has not developed malignancies up to her current age of 21; nor does she have a strong family history of cancer.

Initially, the clinical suspicion of FA was confirmed by elevated spontaneous and mitomycin C-induced chromosome breakage rates (data not shown) and G2-phase accumulation in lymphocyte, lymphoblastoid (Fig. 1A) and fibroblast cultures, which was shown by flow cytometric cell cycle analysis as described in Vaz et al. (2010).

We isolated genomic DNA from fibroblasts using the *GeneJet™ Genomic DNA Purification Kit* (Fermentas). For isolation of RNA we employed the *Quick-RNA™ MiniPrep Kit* (Zymo Research). Translation into cDNA was performed by *SuperScript® II Reverse Transcriptase* (Invitrogen).

Since the patient was among those who remained without detected mutation or assignment to a distinct subtype, we got interested in the significance of WES for molecular diagnostics of FA. We commissioned enrichment and sequencing of the patient´s exome to a service provider. Target enrichment was achieved by means of the *SureSelect Human All Exon 50Mb Kit* (Agilent) and was followed by Next Generation Sequencing on a *SOLiD5500xl* instrument (Applied Biosystems). Afterwards we performed *in house* analysis of the WES data using *NextGENe™* v2.18 software (Softgenetics). The data revealed a total of 103,222,641 reads (Supp.Table S1). 61% of these mapped on target and resulted in an 87-fold average coverage of the exome. Altogether we detected 32,013 variants, including novel mutations as well as listed SNPs. Because of the patient's non-consanguineous descent we restricted our search to compound heterozygous changes and detected 14,715 unknown (excluding reported polymorphisms) heterozygous variants in coding sequences and adjacent intron portions. In

5

Human Mutation

particular, 15 base substitutions were detected in 21 FA and FA-associated genes (91% of exons covered by ≥ 5 reads). *FANCP* (NM_032444.2) carried two *bona fide* pathogenic variants. Even though WES could have missed pathogenic mutations in other FA genes, the compound heterozygous finding in *FANCP* makes this most unlikely. The mutated positions were covered by 6 and 15 reads respectively. We observed the nonsense mutation c.1538G>A in exon 7 resulting in a premature stop codon with the predicted effect of protein truncation, p.W513X, and the splice acceptor mutation c.1367-2A>G in intron 6 (Fig. 1B). Mutation nomenclature is based on cDNA sequence of *SLX4* transcript ENST00000294008 and nucleotide numbering reflects cDNA numbering with +1 corresponding to the A of ATG initiation codon.

Mutation validation was performed by Sanger technique on a 3130xl instrument (Applied Biosystems). The primers for gDNA sequencing included *FANCP*_exon7_for 5'-CCAGAAGCAGGTTTGTGTGA-'3 and *FANCP*_exon7_rev 5'-CCTTCCTGGACTTTCCATCA-'3. We re-sequenced the corresponding regions of *FANCP* in the patient and additionally confirmed the biallelic mutation status and Mendelian segregation of the mutations by sequencing genomic DNA from both parents. The results showed that the splice site mutation was paternally inherited, while the nonsense mutation was transmitted maternally (Supp.Fig. S1A).

We analyzed the consequences of c.1367-2A>G by Sanger sequencing of patient´s cDNA using the primers *FANCP*_c.1-65_for 5'-CAGTACTTTTTGTTCAATTGTGCAAACTC-'3 and *FANCP*_c.1570_rev 5'- CACAGAAAGCTCTGCTTGCGTTC-'3. This analysis demonstrated that a cryptic splice acceptor in exon 7 at position c.1417_1418 is used instead of the mutated in intron 6, as it was predicted by *in silico* analysis. Splice site score calculation using the web tool *http://rulai.cshl.edu/new_alt_exon_db2/HTML/score.html* revealed a score of 1.2 of the cryptic splice acceptor at positions c.1417_1418 and a score of -

6

Human Mutation

1.7 of the mutated one. Of note, this change of the splice acceptor altered the usage of the unaffected wildtype splice donor of intron 6 (score 5.4) to that of a cryptic splice donor at position c.1366+52_1366+53 (score 10). This fact is indicated by the substitution of the 5'-terminal 51 bases of exon 7 with the 5'-terminal 51 bases of intron 6. The electropherogram of cDNA sequencing demonstrates this finding by the superposition of exactly 51 bases starting at cDNA position 1367 and ending at 1417 (Fig.1C and Supp.Fig. S1B), designated as r.1367_1417delins gtttgtgtgatcagaagagtgaccctgggagaggccatcagcaggtcccgg. The length of the open reading frame does not differ as a result of this aberrant splicing pattern. Other *in silico* analyses revealed that the deduced wildtype amino acids at positions p.456_472, ENKSRKKKPPVSPPLLL, are predicted to include at positions p.460_464 one of five potential SLX4/FANCP nuclear localization signals (NLS) (*http://psort.hgc.jp/form2.html)* (Fig.2A). The mutation, denoted p.E456_L4772delinsGLCDQKSDPGRGHQQVP, results in the loss of that potential NLS.

Further experiments similar to those described by Stoepker et al. (2011) showed that no residual protein is present. Neither by immunoprecipitation (Fig. 2B), nor in a cell fractionation assay (Fig.2C) SLX4/FANCP was detected on Western Blots. We conclude that the allele carrying the stop mutation gives rise to a truncated protein that is unstable and rapidly degraded. Similarly, the allele with the splice mutation does not express a stable protein which could locate to the nucleus. Therefore, it is not surprising that interactions with the structure-specific nucleases XPF/ERCC1 and EME1/MUS81 are disrupted (Fig.2B) and that ERCC1 is not able to form nuclear foci (Fig.2D), as described for other FA-P patients (Stoepker, et al., 2011).

In summary, our study adds a seventh patient to the most recently described FA subtype, FA-P. Neither of her compound heterozygous mutations has previously been reported. They extend the mutation spectrum of the latest member of the FA gene family; *FANCP*, and have

7

Human Mutation

been added to the Fanconi Anemia Mutation Database (http://www.rockefeller.edu/fanconi/).
In contrast to the FA-P patients reported so far, cells derived from the present patient do not
seem to be able to express any SLX4/FANCP protein (Kim, et al., 2011; Stoepker, et al.,
2011). The failure of coordination of structure-specific nucleases in ICL unhooking due to the
absence of SLX4/FANCP does not result in a more severe phenotype as that of other FA-P
patients previously reported, which is not comparable to the cancer-prone phenotype of
subtypes FA-D1 or -N, but falls into the clinical spectrum of the other FA groups. These
insights were facilitated by WES that proved a valuable tool for molecular diagnostics of FA,
as of other heterogeneous diseases, by the identification of disease-causing genes so that it
may increasingly replace classical genetic approaches.

Competing interests

The authors deny competing financial or other interests.

Funding

Human Mutation

References

Alter BP, Kupfer G. 1993. Fanconi Anemia. In: Pagon RA, Bird TD, Dolan CR, Stephens K, editors. GeneReviews. Seattle (WA).

Andersen SL, Bergstralh DT, Kohl KP, LaRocque JR, Moore CB, Sekelsky J. 2009. Drosophila MUS312 and the vertebrate ortholog BTBD12 interact with DNA structure-specific endonucleases in DNA repair and recombination. Molecular cell 35:128-35.

Auerbach AD. 1993. Fanconi anemia diagnosis and the diepoxybutane (DEB) test. Experimental hematology 21:731-3.

Callen E, Casado JA, Tischkowitz MD, Bueren JA, Creus A, Marcos R, Dasi A, Estella JM, Munoz A, Ortega JJ and others. 2005. A common founder mutation in FANCA underlies the world's highest prevalence of Fanconi anemia in Gypsy families from Spain. Blood 105:1946-9.

Crossan GP, van der Weyden L, Rosado IV, Langevin F, Gaillard PH, McIntyre RE, Gallagher F, Kettunen MI, Lewis DY, Brindle K and others. 2011. Disruption of mouse Slx4, a regulator of structure-specific nucleases, phenocopies Fanconi anemia. Nature genetics 43:147-52.

Deans AJ, West SC. 2011. DNA interstrand crosslink repair and cancer. Nature reviews. Cancer 11:467-80.

Kim Y, Lach FP, Desetty R, Hanenberg H, Auerbach AD, Smogorzewska A. 2011. Mutations of the SLX4 gene in Fanconi anemia. Nature genetics 43:142-6.

Lobitz S, Velleuer E. 2006. Guido Fanconi (1892-1979): a jack of all trades. Nature reviews. Cancer 6:893-8.

Rosenberg PS, Alter BP, Ebell W. 2008. Cancer risks in Fanconi anemia: findings from the German Fanconi Anemia Registry. Haematologica 93:511-7.

Rosenberg PS, Greene MH, Alter BP. 2003. Cancer incidence in persons with Fanconi anemia. Blood 101:822-6.

Rosenberg PS, Tamary H, Alter BP. 2011. How high are carrier frequencies of rare recessive syndromes? Contemporary estimates for Fanconi Anemia in the United States and Israel. American journal of medical genetics. Part A 155A:1877-83.

Schindler D, Kubbies M, Hoehn H, Schinzel A, Rabinovitch PS. 1987. Confirmation of Fanconi's anemia and detection of a chromosomal aberration (1Q12-32 triplication) via BrdU/Hoechst flow cytometry. Am J Pediatr Hematol Oncol 9:172-7.

Schroeder TM, Anschutz F, Knopp A. 1964. [Spontaneous chromosome aberrations in familial panmyelopathy]. Humangenetik 1:194-6.

Seif AE. 2011. Pediatric leukemia predisposition syndromes: clues to understanding leukemogenesis. Cancer genetics 204:227-44.

Stoepker C, Hain K, Schuster B, Hilhorst-Hofstee Y, Rooimans MA, Steltenpool J, Oostra AB, Eirich K, Korthof ET, Nieuwint AW and others. 2011. SLX4, a coordinator of structure-specific endonucleases, is mutated in a new Fanconi anemia subtype. Nature genetics 43:138-41.

Strathdee CA, Duncan AM, Buchwald M. 1992. Evidence for at least four Fanconi anaemia genes including FACC on chromosome 9. Nature genetics 1:196-8.

Human Mutation

Figure legends

Fig. 1 G2-phase arrest and *FANCP* mutation (A) Cell cycle distribution of patient-derived lymphoblastoid cells show increased G2 arrest without (21.4% of cells in G2) and especially after exposure of cultures to 15 ng/μl MMC (55.4% of cells in G2) compared to a control cell line (0 MMC: 9.1% of cells in G2, 15 ng/μl MMC: 10.9% of cells in G2). (B) The graphic chart shows the splice site mutation c.1367-2A>G (upper panel) and the nonsense mutation c.1538G>A (lower panel) by color-space WES data according to *NextGENe™* software presentation. The cutout displays from top to bottom the number of the coding exon and the gene name, the chromosomal position, followed by the reference and the patient consensus nucleotide sequences, the corresponding amino acids and, below the horizontal line, the original WES reads. Since *FANCP (BTBD12)* is encoded on the minus-strand of gDNA, the sequences and base changes are displayed in reverse complementary manner. (C) Depicted is an excerpt from exon 6 to 7 of the *Ensembl* gDNA sequence of *SLX4/FANCP* (ENSG00000188827). Red circles indicate the wildtype and cryptic splice sites. The 51 bases of exon 7 which are interchanged with the same number of bases of intron 6 are highlighted in yellow.

Fig. 2 Mutant SLX4/FANCP expression and function (A) Ideograms of the SLX4/FANCP domain structure (modified after Svendsen et al., 2009). Wildtype (WT) protein (top) contains in addition to reported domains (UBZ, yellow; BTB, blue; SAP, purple; and SBD, green) five potential nuclear localization signals (NLS, red) predicted by the web tool *http://psort.hgc.jp/form2.html*, spanning amino acid (aa) positions 109_124 (NLS 1), 397_412 (NLS 2), 460_464 (NLS 3), 1079_1085 (NLS 4) and 1814_1830 (NLS 5). The ideograms below (middle and bottom) show the predicted protein effect of the *SLX4/FANCP* mutations in the present patient. c.1538G>A leads to protein truncation at p.W513X, while c.1367-2A>G leads to p.E456_472Ldelins17 and the loss of NLS 3. (B) Immunoprecipitation and

11

Human Mutation

protein blot analysis shows SLX4/FANCP deficiency in fibroblasts of the patient compared to WT and FA-P controls. There is no co-precipitation of XPF, MUS81 and ERCC1 with the mutant protein. SE and LE indicate short and long exposure of the blot, respectively. (C) Subcellular fractionation of patient´s fibroblasts fails to demonstrate SLX4/FANCP protein in any fraction. Chromatin loading of XPF and MUS81 were not detected. Tubulin, p300 and HDAC1 served as loading controls for the cytoplasmic fraction (CE), nuclear extract (NE) and chromatin fraction (CB). The faint band observed in NE slightly below SLX4/FANCP is unspecific. (D) Formation of nuclear ERCC1 foci is abolished in patient´s fibroblasts as in other FA-P cells (EUFA1354) in contrast to the wildtype control line (LN9SV). The ERCC1 antibody FL297 was used for immunofluoresence and TOPRO3 as a nuclear counterstain.

Figure 1
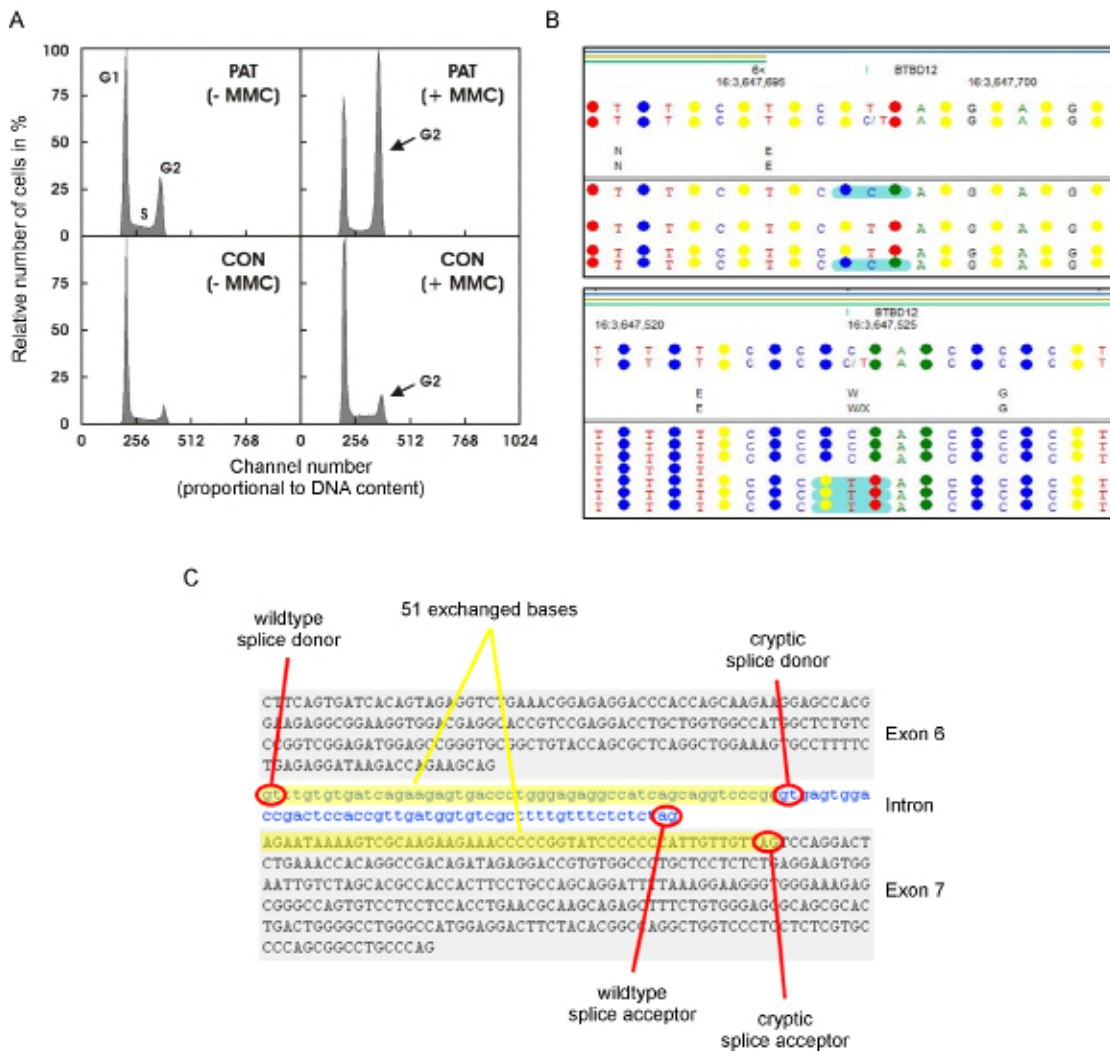
Figure 2
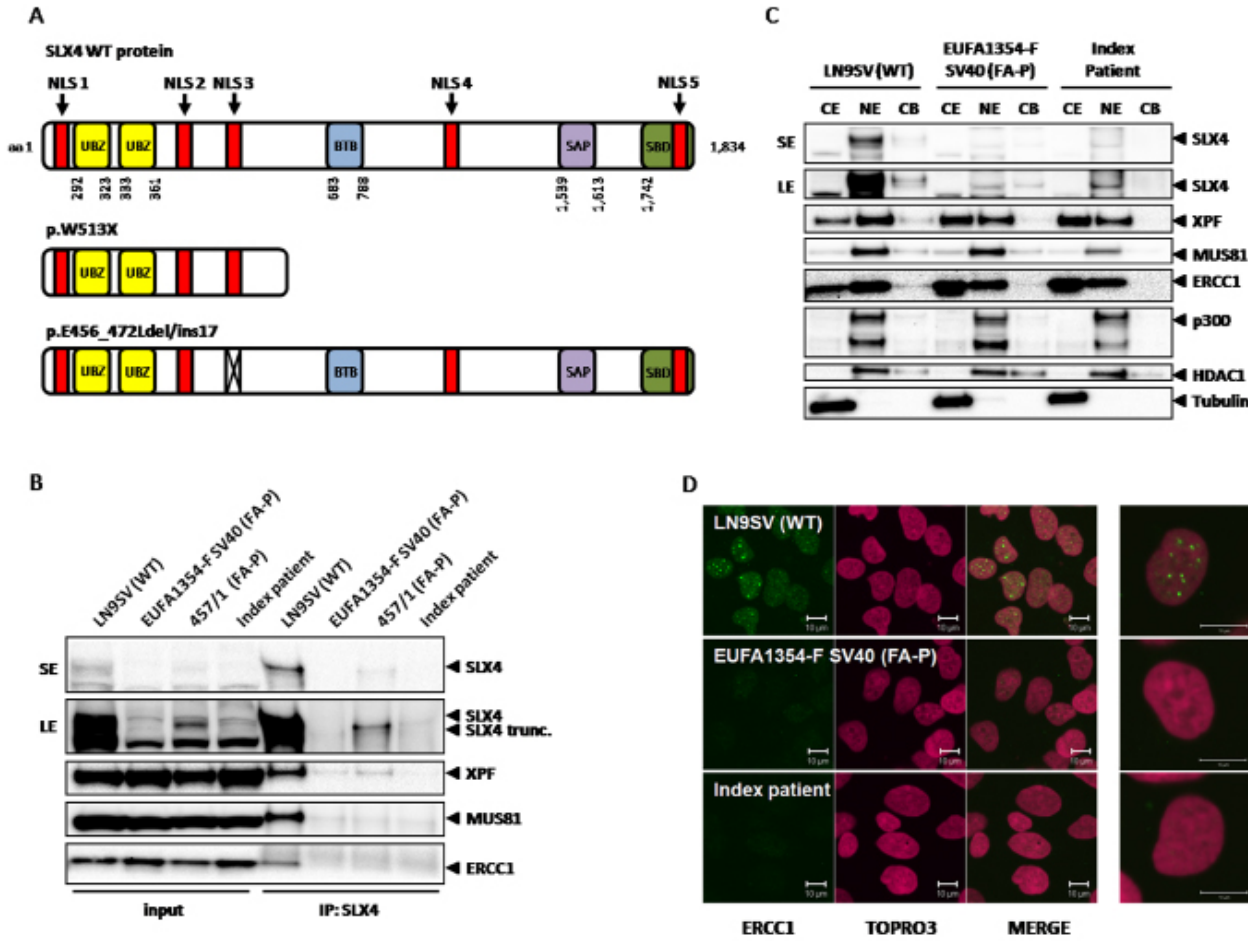
Supplementary Information

Supp. Table S1: Statistical summary of WES results including read counts and detected variants

Total read number	109,637,890
Reads passing QC	103,222,641
Reads on target	62,868,844 (61%)
Average exome coverage	87x
Total number of variants	32,013
Known SNPs/MNPs	17,194
Unknown variants in coding sequence	13,519
Unknown variants at canonicla splice sites	377
Unknown homozygous variants	91
Unknown heterozygous variants	14,715
Unknown silent variants	2,983
Unknown missense variants	9,629
Unknown nonsense variants	510
Unknown insertions/deletions	468
Genes with multiple heterozygous variants	131
Heterozygous variants in FA and FA associated genes	15

QC: quality control; SNP: single nucleotide polymorphism; MNP: multiple nucleotide polymorphism

Figure S1

A
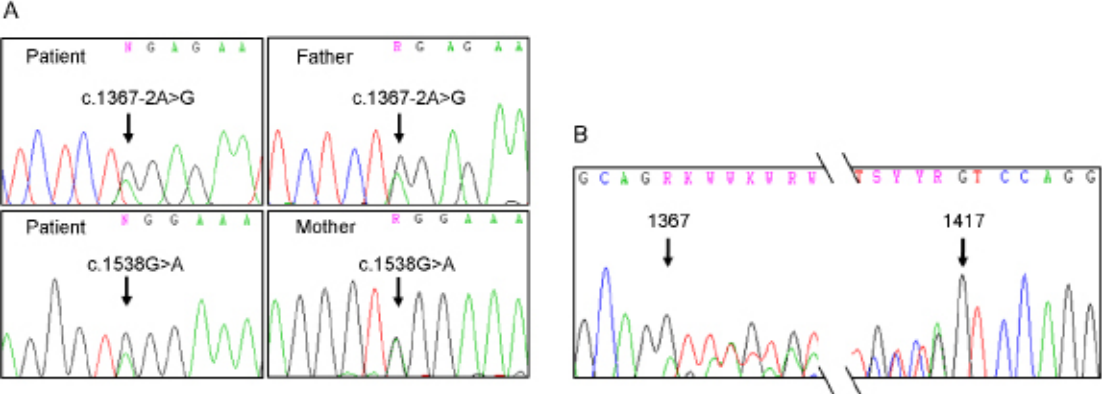
Fig. S1: Validation of Whole Exome Sequencing results by Sanger resequencing. (A) Mendelian segregation of the FANCP mutations in the patient´s family. The splice site change c.1367-2A>G was inherited from the father, while the nonsense mutation c.1538G>A is of maternal origin. (B) Consequence of c.1367-2A>G on transcript level. Sequencing the patient´s cDNA revealed the replacement of exactly 51 exonic bases starting at cDNA position 1367 and ending at 1417 by the first 51 bases of the following intron.

3. RESULTS

3.1 FA CANDIDATE GENES

3.1.1 A HISTONE-FOLD COMPLEX AND FANCM FORM A CONSERVED DNA-REMODELING COMPLEX TO MAINTAIN GENOME STABILITY

3.1.2 ON THE ROLE OF FAN1 IN FANCONI ANEMIA

3.2 IDENTIFICATION OF NOVEL FA GENES

3.2.1 MUTATION OF THE RAD51C GENE IN A FANCONI ANEMIA-LIKE DISORDER

3.2.2 SLX4, A COORDINATOR OF STRUCTURE-SPECIFIC ENDONUCLEASES, IS MUTATED IN A NEW FANCONI ANEMIA SUBTYPE

3.2.3 XPF MUTATIONS SEVERELY DISRUPTING DNA INTERSTRAND CROSSLINK REPAIR CAUSE FANCONI ANEMIA

3.3 GENOTYPING FANCONI ANEMIA BY NEXT GENERATION SEQUENCING

3.3.1 WHOLE EXOME SEQUENCING REVEALS NOVEL MUTATIONS IN THE RECENTLY IDENTIFIED FANCONI ANEMIA GENE SLX4/FANCP

3.3.2 GENOTYPING FANCONI ANEMIA BY WHOLE EXOME SEQUENCING: ADVANTAGES AND CHALLENGES

PLOS ONE

Genotyping of Fanconi Anemia Patients by Whole Exome Sequencing: Advantages and Challenges
--Manuscript Draft--

Manuscript Number:	PONE-D-12-21931
Article Type:	Research Article
Full Title:	Genotyping of Fanconi Anemia Patients by Whole Exome Sequencing: Advantages and Challenges
Short Title:	Genotyping FA patients by Whole Exome Sequencing
Corresponding Author:	Beatrice Schuster Universtiy of Wuerzburg Wuerzburg, GERMANY
Keywords:	Whole Exome Sequencing; Fanconi Anemia; Diagnostics; genotyping; Complementation group assignment
Abstract:	Fanconi anemia (FA) is a rare genomic instability syndrome. Disease-causing are biallelic mutations in any one of at least 15 genes encoding members of the FA/BRCA pathway of DNA-interstrand crosslink repair. Suggested by phenotypical manifestations, the diagnosis of FA is confirmed by the hypersensitivity of cells to DNA interstrand crosslinking agents. Customary molecular diagnostics has become increasingly cumbersome, time-consuming and expensive the more FA genes were identified. We performed Whole Exome Sequencing (WES) in four FA patients in order to investigate the potential of this method for FA genotyping. In search of an optimal WES methodology we explored different enrichment and sequencing techniques. In each case we were able to identify the pathogenic mutations so that WES provided both, complementation group assignment and mutation detection in a single approach. The mutations included homozygous and heterozygous single base pair substitutions and a two-base-pair insertion in FANCJ, -D1, or -D2. Different WES strategies had no critical influence on the individual outcome. However, database errors and in particular pseudogenes impose obstacles that may prevent correct data perception and interpretation, and thus cause pitfalls. With these difficulties in mind, our results show that WES is a valuable tool for the molecular diagnosis of FA and a sufficiently safe technique, capable of engaging increasingly in competition with classical genetic approaches.
Order of Authors:	Beatrice Schuster
	Kerstin Knies
	Najim Ameziane
	Martin Rooimans
	Thomas Bettecken
	Johan de Winter
	Detlev Schindler
Suggested Reviewers:	Jordi Surralles Universitat Autonoma de Barcelona jordi.surralles@uab.es
	Settara C. Chandrasekharappa National Human Genome Research Institute, Bethesda chandra@nhgri.nih.gov
Opposed Reviewers:	Arleen D Auerbach The Rockefeller University, New York competing interest
	Christopher G Mathew
	King´s College Lodon School of Medicin, Lodon Competing interest

Genotyping of Fanconi Anemia Patients by Whole Exome Sequencing:

Advantages and Challenges

Beatrice Schuster[1]*, Kerstin Knies[1]*, Najim Ameziane[2], Martin Rooimans[2], Thomas

Bettecken[3], Johan de Winter[2], Detlev Schindler[1]

[1] Department of Human Genetics, University of Wuerzburg, Wuerzburg, Germany

[2] Department of Clinical Genetics, Vrije Universiteit (VU) Medical Center, Amsterdam, The Netherlands

[3] Center for Applied Genotyping Munic, Max-Planck-Institut für Psychatrie, Munich, Germany

* Corresponding authors, contributed equally

Beatrice Schuster and Kerstin Knies

University of Wuerzburg

Department of Human Genetics

Biozentrum - Am Hubland

97074 Wuerzburg, Germany

Tel.: 0049-931-3184093

Fax.: 0049-931-3184069

beatrice.schuster@biozentrum.uni-wuerzburg.de

kerstin.knies@uni-wuerzburg.de

Introduction

Fanconi anemia (FA) is an autosomal or X-chromosomal recessive disorder characterized by variable yet typical developmental malformations, bone marrow failure and predisposition to leukemia and solid tumors. As much as 15 genes define corresponding complementation groups designated as FA-A, -B, -C, -D1, -D2, -E, -F, -G, -I, -J, -L, -M, -N, -O and -P. Biallelic, or in the case of FA-B hemizygous, mutations of any one of the underlying genes lead to FA, while monoallelic mutations of *FANCD1* (*BRCA2*), *FANCJ* (*BRIP1*), *FANCN* (*PALB2*) or *FANCO* (*RAD51C*) increase the risk for carriers of developing breast and ovarian cancer [1]. In contrast, FA patients commonly suffer from physical abnormalities like short stature, abnormal skin pigmentation, radial ray defects and malformations of the ears, eyes and inner organs. More than 80% of FA patients develop progressive bone marrow failure which makes pancytopenia a highly suggestive feature [2-4]. In addition, FA patients show not only greatly elevated frequencies of myelodysplastic syndrome and acute myeloid leukemia in childhood, but there is also a markedly increased prevalence of non-hematologic malignancies. They experience an up to 700-fold higher risk of squamous cell carcinomas especially of the head and neck or genital region [3,5]. Other solid tumors are less frequent among FA patients but a variety of them are still extraordinarily common compared to the general population [5]. The reason for the overall increased cancer risk may probably be due to the DNA repair defect that characterizes the cellular phenotype [6]. FA cells show elevated rates of chromosomal breakage and typical radial rearrangement figures. These features occur spontaneously but are exaggerated following exposure of cultures to DNA crosslinking agents such as diepoxybutane (DEB) or mitomycin C (MMC) [7-9]. Since this hypersensitivity is typical for FA cells, chromosomal breakage analysis is used as a diagnostic tool. Alternatively, cell cycle studies or cellular survival assays using flow cytometric methods

2

are utilized for diagnosis since FA cells are hindered to pass G2 checkpoint control, accumulate in G2 phase and show increased death rates after DNA damage induction [10,11]. On the molecular level the diagnosis is more complicated. Even though about 60% of FA patients carry mutations in *FANCA* [3], there remain 14 other FA and at least 4 associated genes (*FAAP100*, *FAAP24*, *MHF1* and *MHF2*) that may alternatively contain disease-causing defects. While there is so far no cure for FA, knowledge of the individual complementation group and the specific mutations of patients may be important for differential, prenatal or preimplantation diagnosis, prognosis or upcoming gene therapy trials. Biallelic mutations of *FANCD1*, for example, are associated with early onset of acute myelogenous leukemia and blastomas [12,13]. Because of the high number of FA and FA-associated genes and because some of these genes have more than 40 exons, Sanger sequencing is becoming increasingly tedious, time-consuming and costly.

Recently, an efficient and reliable technique using Next Generation Sequencing (NGS) emerged to improve and accelerate conventional methods of molecular diagnostics. In the present study we demonstrate the versatility of Whole Exome Sequencing (WES) in four independent projects. The four patients involved had previously been confirmed to be afflicted with FA by non-molecular procedures but were not assigned to any complementation group and thus lacking accountable mutations. Using WES we genotyped each patient by the identification of their disease-causing mutations in one of three different FA genes. Thus we consider WES an efficient tool to compete with traditional approaches for the molecular diagnosis of FA.

3

Material and methods

Study design

The study scope, patient information and consent form were approved by the ethics committee of the Medical Faculty of the University of Wuerzburg. All participants provided written informed consent.

Cell cycle analysis

For confirmation or exclusion of FA we used flow cytometric cell cycle analysis as described earlier [11,14].

DNA sample preparation

Genomic DNA was isolated from fibroblasts using the *GeneJet* Genomic DNA Purification Kit (Fermentas) following the manufacturer's instructions. For isolation of gDNA from peripheral blood we used a salting-out technique.

Whole Exome Sequencing

Enrichment and sequencing of the exomes of projects 1 to 3 were commissioned to different service providers on an exclusively commercial basis. Sample 1 was enriched by means of the *NimbleGen SeqCap EZ Human Exome Library v2.0* and sequenced *on Illumina HiSeq2000*. For sample 2 the *Agilent SureSelect Human All Exon 38 Mb Kit* (hg18) was used together with the *SOLiD4* sequencing technology by *Applied* Biosystems. In project 3 WES was performed with *SOLiD4* technology by *Applied Biosystems* after enrichment using the *Agilent SureSelect Human All Exon 50 Mb Kit* (hg19)*.* Raw data from *Illumina* sequencing were provided in *fq* format. *SOLiD* raw data were provided in *csfasta* format along with *qual* files containing corresponding quality information. For project 4 we used the *SureSelect Human All Exon Kit* (Agilent) targeting approximately 38 Mb, following the manufacturer´s instructions. The

enriched sample was sequenced on one lane of the *Illumina GAIIx* instrument using a paired-end sequencing protocol. This is available upon request.

Data analysis

Analysis of the WES data of projects 1 to 3 was performed using the alignment and analysis software *NextGENe* v2.18 by *Softgenetics.* The raw data were filtered for low quality reads before alignment. All reads passing the quality filter were aligned to the whole human genome hg18 in project 2 and hg19 in projects 1 and 3. The average exome coverage was determined using a complete list of human exons generated by the *UCSC Table Browser*. The following analytical steps were performed only with reads that matched exonic regions including exon-intron-boundaries. SNP and Insertion/Deletion (indels) analysis was done by different filtering steps depending whether consanguinity was suggested or not. In each case all variants listed in the *NCBI (national Center for Biotechnology Information) dbSNP* database were excluded first. In patient 2-1 of consanguineous descent only homozygous variants were included. In patients 1 and 3 with non-consanguineous background only genes with at least two heterozygous changes in the DNA sequence were considered to be potentially disease-causing. Mutation pathogenicity prediction was made by *MutationTaster* (*http://www.mutationtaster.org)*, *Poly-Phen2* (http://genetics.bwh.harvard.edu/pph2) and *SIFT* software (http://sift.jcvi.org).

For sample 4 we used a data analysis pipeline for the evaluation of single nucleotide variants and small indels, which was comprised of tools freely available in the web. The paired-end reads were mapped by the *Burrows-Wheeler Aligner* (*BWA*) [15] to the reference genome build according to NCBI hg19. Subsequently, SNPs and small indels were called using *Samtools* [16] and *Varscan* [17]. The resulting list of variants was annotated with *Annovar*

[18] that summons and utilizes information from external databases to assess implications and consequences of a given sequence alteration, such as an ensuing amino acid change, location in a canonical splice site, and information from dbSNP along with SNP frequency if available. Finally, a manual filtering step was executed to prioritize relevant mutations. Low frequency frameshift and truncating mutations were considered pathogenic as in the other projects. Not reported non-synonymous amino acid variants were analyzed *in silico* by *Align-GVGD* (data not shown), *MutationTaster*, *Polyphen-2* and *SIFT* to assess any potentially damaging effect.

Holding for all samples, the variant detection frequency was set at a minimum of 20% of the reads covering any aberration. A minimum coverage by 10 reads was necessary for any variant to be considered a real mutation.

Sanger sequencing

Potential mutations were validated by Sanger sequencing generally using an *Applied Biosystems 3130xl* instrument. Primer sequences and PCR conditions are available upon request.

Immunoblotting

FANCD2 expression analysis was performed with whole protein extracts isolated from patient-derived fibroblasts. Cell lines were treated with hydroxyurea or MMC before analysis. We used antibodies including mouse monoclonal anti-FANCD2 (sc20022, *Santa Cruz Biotechnology*), mouse monoclonal anti-RAD50 (GTX70228, *GeneTex*) and mouse monoclonal anti-Vinculin (sc-25336, *Santa Cruz Biotechnology*).

Autozygosity mapping

Autozygosity mapping was performed with SNP data generated with the *Illumina SNP array HumanHap300v2*. Genotypes were analyzed using AutoSNPa software [19].

Results

Confirmation of the diagnosis of FA

In each project the clinical diagnosis was confirmed by flow-cytometric cell cycle analysis. In patients 1 and 2 FA was evidenced by studies of PHA-stimulated cultures of peripheral blood lymphocytes. After 72 h incubation with 10 ng/ml MMC the ratio "sum of all G2 phases vs. growth fraction" was above 0.4 which is characteristic of FA patients (Fig. 1A) [20]. Patients 3 and 4 showed distinct cell accumulations in the G2 phase (>20%) in fibroblast cultures exposed to 12 ng/ml MMC for 48 h, likewise consistent with other FA patients (Fig. 1B).

Mutation detection by WES and validating experiments

Statistical data of each WES project are summarized in Tab. 1. The exome was covered on average between 22x and 77x. Consecutively regarded in detail were the FA and FA-associated genes among genes with homozygous, or at least two heterozygous, mutations. Assignment of the mutations to different alleles and screens for their pathogenicity revealed the following results.

Project 1

We employed 3 μg gDNA isolated from cultured fibroblasts of patient 1 to enrich the whole exome using the *NimbleGen SeqCap EZ Human Exome Library v2.0*. Sequencing was

7

performed on an Illumina HiSeq2000 instrument. It revealed two heterozygous mutations

with a score ≥ 10 (probability 1:100 for being false positive) exclusively in *FANCD2*. They

included the single-base substitution c.2314G>T, resulting in a premature stop codon at

normal amino acid sequence position p.772, and the canonical splice site change

c.3888+2T>G. Sanger sequencing confirmed the splice site alteration on gDNA level and

showed *in frame* skipping of exon 39 on cDNA level (Fig. 2A). Sanger validation of the

nonsense mutation had to be performed with a super-amplicon of exons 22 to 26 including

all pertinent introns because of the pseudogene *FANCD2-P2*, containing a copy of the active

gene region [21]. Re-sequencing of this super-amplicon did not confirm the substitution

c.2314G>T (Supplementary Fig. S1A). Thus we concluded that this variant had occurred in

the pseudogene and therefore could not be causative of FA in that patient. By decreasing

the filter settings we additionally detected the missense substitution c.2204G>A resulting in

the amino acid change p.R735Q. Even though this base exchange is also reflected in the

pseudogene sequence, its assessment by gene-specific super-amplification rendered it an

authentic *FANCD2* mutation (Fig. 2B). We confirmed the maternal segregation of p.R735Q,

whereas the splice site change was not present in maternal gDNA and may have occurred *de

novo* or was, more likely, inherited from the father, of whom no material was available.

Finally, decreased abundance of FANCD2 protein in the patient's cells confirmed our DNA

sequencing results (Fig. 2C).

Project 2

Parallel to WES we performed a genome-wide SNP study in the patient (2-1), his two

affected siblings (2-2, 2-3) and their parents. Autozygosity mapping using these data

revealed a large homozygous region on chromosome 17 (Fig. 3A). The *Agilent SureSelect

Human All Exon 38 Mb Kit* (hg18) was used to capture and enrich the whole exome from 3

µg gDNA, isolated from peripheral blood of patient 2-1. For sequencing we selected the *SOLiD4* technology by *Applied Biosystems*. Analysis of the color-spaced sequencing data revealed the homozygous single base substitution c.1878A>T in *FANCJ* (Fig. 3B) compatible with the outcome of the disease gene mapping. This point mutation results in the amino acid change p.E626D (Fig. 3B) that is predicted to be pathogenic (Supplementary Tab. S1). Sanger sequencing confirmed the homozygous mutation of the patient (Fig. 3C) and his siblings (data not shown). These results and the heterozygous detection of the mutation in both parents (Fig. 3C) were consistent with Mendelian segregation.

Project 3

3 µg gDNA of patient 3 were isolated from fibroblasts. WES was performed using *SOLiD4* technology by *Applied Biosystems* after enrichment with the *Agilent SureSelect Human All Exon 50 Mb Kit* (hg19). Initial analysis failed to show FA or FA-associated genes with biallelic mutations, so that we re-examined all unknown variants listed as SNPs. That way we identified the 2-bp insertion c.7890_7891insAA in *BRCA2/FANCD1* (Fig. 4A) with the effect p.L2631Nfs16X, and additional three reported variants in the same gene, of which only c.7795G>A was predicted to be pathogenic (Fig. 4B, Tab. S1). The SNP rs80359682 listed at this position is a deletion of three bases (c.7795_7797delGAA) which may be pathogenic, whereas our detected single-bp substitution results in the missense mutation p.E2599K likely to be deleterious. We confirmed both mutations of patient 3 by Sanger sequencing, even though the allele carrying the insertion was detectable only on low level (Fig. 4A).

Project 4

In project 4 we used again the *SureSelect Human All Exon Kit (Agilent)* targeting all human exons (hg19), in total approximately 38 Mb. The enriched sample was sequenced on one lane of the *Illumina GAIIx* instrument. In contrast to projects 1 to 3, whose data have been

analyzed by means of the alignment and analysis software *NextGENe* v2.18 by *Softgenetics*, we used for project 4 an in-house variation detection pipeline to score sequence variants (Ameziane et al. 2012, in press). We focused on rare variants within the coding and splice site regions of all known FA genes. Only one reported heterozygous base substitution was detected in *FANCD2* exon 16, c.1370T>C (p.L457P), which has previously been recognized to be pathogenic [21]. Initially we failed to detect a second mutation in *FANCD2*. After visual inspection of the mapped reads in the IGV browser, *FANCD2* exon 5 was shown to be covered by a single read. Therefore, the data appeared unreliable for mutation detection. Subsequent Sanger sequencing of that exon demonstrated the c.376A>G base substitution resulting in p.S126G that has previously been shown to be pathogenic and thus represent a missense mutation [22] (Fig. 5A). Further analysis was done by Western blotting that revealed distinct deficiency of the FANCD2 protein (Fig. 5B).

Discussion

The present study proposes the application of WES for the molecular diagnosis of FA. Major concerns with WES are ethical issues, less its performance. Potentially, WES data could be used to analyze any gene, or even all genes, for any purpose. In our projects the patients had given informed consent for FA diagnostics. We addressed and resolved the discrepancy to WES in the way that we used the whole body of data only for quality control, statistical analyses, and to apply general filter settings. For mutation screens we solely regarded FA genes and four associated genes. Defects of the latter have not yet been reported but would likely result in FA. The remainder of the WES data was secured from further access.

We performed four independent sequencing projects with disparate exome enrichment and sequencing technologies. Even if the performances differed slightly, we were able to identify the disease causing mutations of all index patients. Except for the mutations of patient 4, all of the detected pathogenic variants had not yet been listed in the *Fanconi Anemia Mutation Database* (http://www.rockefeller.edu/fanconi/) such that we consider this identification of missense, nonsense and splice site mutations as well as a 2-bp insertion in *FANCD1*, *FANCD2* and *FANCJ* a challenging task and major accomplishment. The successful outcome led us to conclude that WES generally is a reliable tool for the molecular diagnosis of FA. It also proved to be efficient in time and adequate in expense. Including sample quality control, target enrichment, sequencing and basic bioinformatics each of our projects took two to three months for completion. Alignment and mutation calling afforded another few days, followed then by validation processes. The cost of each of the four projects was highly variable and ranged from 1000$ to 4500$, but decreased in tendency over time. Sanger sequencing of all FA and the four FA-associated genes would have exceeded the projects´ current term and cost several times. Another alternative, enrichment of the FA gene regions followed by NGS would also have been more time-consuming and expensive if set-up investments were incorporated.

None of the patients who were subject of our four projects had large deletions or duplications of FA genes. If that had been the case we would probably have missed those mutations by WES alone. This drawback can, however, be avoided if WES is combined with MLPA or microarray techniques. Recently, Ameziane et al. (2012) reported the detection of large deletions after NGS of enriched FA gene regions by evaluating the Log2 ratio of the local read depth divided by a read depth reference. This method could, in principle, also be included in WES data analysis. A problem for any mutation screen remain deep intronic

11

mutations that may even escape refined variation detection pipelines, primarily due to filtering procedures [23]. None of the FA patients in our four projects had mosaicism in the hematopoietic system as shown by the diagnostic procedures upfront WES. If such situation was suggested it would be appropriate just to use fibroblast DNA for WES. Thus we think the potential problem of FA mosaics can be resolved without tedious mixing experiments to show to which extent mosaicism may be detected in blood.

Our study also raised technical issues and revealed methodical difficulties that should to be addressed. For each project we analyzed the exome coverage, in particular the coverage of the FA and FA-associated genes. Even though the average exome coverage in the *SureSelect* enrichment projects clearly exceeded the coverage in the *NimbleGen* project, we found coverage of the FA gene regions in the latter to be more consistent and complete (Supplementary Tab. S2). A similar observation was reported by Clark et al. (2011) for the whole exome in general. Most of the entirely unsequenced exons in our study had a high GC or high AT content leading to the conclusion that excess GC content is still a limiting factor for efficient hybridization and amplification during target enrichment [24]. For example, initially only one pathogenic mutation was detected in project 4, while the other mutation was missed because of insufficient coverage. In this case the GC and AT content of *FANCD2* exon 5, where the second mutation later was identified, is balanced with 44% and 56%, but probably the high AT content of the adjacent intron regions that were included in enrichment may explain the low coverage. Nevertheless, the identification of the first mutation led to close examination of that gene and subsequent identification of the second mutation. On the other hand it is a rare but recurrent experience that a single heterozygous mutation in one FA gene may accompany compound heterozygous, disease-causing mutations in another FA gene.

We observed a lower rate of sequencing errors in *SOLiD* data. The two-base-encoding technology leads to lower rates of false positive or false negative base calls and facilitates the discrimination of sequencing errors from authentic mutations [25,26]. Another obstacle to our analysis were databases of coding sequences that contained substantially different numbers of genes. Moreover, the start and end positions of single exons or the exon-intron structure of genes may vary between different databases [24]. These facts may complicate the validation process and cause confusion by wrong and incomplete mutation calling. In project 3 we experienced a problem even with SNP databases. A truly pathogenic mutation was designated as a SNP because there was a polymorphism including the mutated base pair. To avoid such pitfalls as far as possible we recommend using always the latest version of a SNP database, because mutation screening becomes easier the more polymorphisms can be excluded.

Another issue to consider during bioinformatics analysis is the choice of mutation prediction software. For the mutations addressed in this study we compared the performance of three different mutation prediction tools (*SIFT*, *PolyPhen-2* and *MutationTaster*). While *SIFT* and *Polyphen-2* often failed to ascertain the pathogenic effect of the mutations, *MutationTaster* generally was able to provide a reliable prognosis for all genes and all kind of mutations.

Finally, in this and other studies we noticed that reads containing pseudogene sequences can interfere with the sequence of the functional gene and thus cause false results. In project 1 re-sequencing showed that the c.2314G>T mutation call of *FANCD2* was erroneous and probably should have mapped to the pseudogene, *FANCD2-P2*. The missense mutation c.2204G>A likewise represented *FANCD2-P2* pseudogene sequence. In this case it proved to be a true *FANCD2* mutation at the same time. In that same exon we identified two more

base substitutions representing pseudogene sequence but the corresponding reads were misleadingly mapped to *FANCD2*. Only gene-specific re-sequencing resolved the correct sequence (data not shown). We recognized this problem not only in FA genes. For example, another project had revealed a hemizygous deletion including the *CDC27* locus, but WES unexpectedly showed heterozygous base variants. On closer inspection we found three related pseudogenes, containing the complete cDNA sequence of *CDC27* from exon 3 to 14. This led us to re-check the putative gene variants by Sanger sequencing. All of the variants turned out to be false positives resp. pseudogene sequences (data not shown). We suggest that this problem may be due to the short read length produced by *SOLiD* and *Illumina* NGS and ambiguous mapping during alignment with the genome. Pseudogenes are characterized by high sequence similarity with their corresponding functional genes and therefore ambivalent mapping in the analysis of NGS data cannot always be avoided. Regarding FA genes, special attention needs to be paid to *FANCD2* for which only *FANCD2-P1* LOC100421239 is listed in the NCBI database but not the other reported pseudogene, *FANCD2-P2* [21]. For *FANCL* and the FA-associated gene *MHF1* at least partial copies have been disclosed.

Notwithstanding the challenges with WES data analysis, we would recommend it as a valuable tool for FA genotyping. In our opinion, WES, if carefully applied, is able to compete with classical molecular approaches in diagnostics and research not only of FA genes but generally for disorders with locus heterogeneity.

Acknowledgments

References

1. Vuorela M, Pylkas K, Hartikainen JM, Sundfeldt K, Lindblom A, et al. (2011) Further evidence for the contribution of the RAD51C gene in hereditary breast and ovarian cancer susceptibility. Breast Cancer Res Treat 130: 1003-1010.
2. Alter BP, Kupfer G (1993) Fanconi Anemia. In: Pagon RA, Bird TD, Dolan CR, Stephens K, editors. GeneReviews. Seattle (WA).
3. Auerbach AD (2009) Fanconi anemia and its diagnosis. Mutat Res 668: 4-10.
4. Seif AE (2011) Pediatric leukemia predisposition syndromes: clues to understanding leukemogenesis. Cancer Genet 204: 227-244.
5. Rosenberg PS, Greene MH, Alter BP (2003) Cancer incidence in persons with Fanconi anemia. Blood 101: 822-826.
6. Deans AJ, West SC (2011) DNA interstrand crosslink repair and cancer. Nat Rev Cancer 11: 467-480.
7. Auerbach AD (1993) Fanconi anemia diagnosis and the diepoxybutane (DEB) test. Exp Hematol 21: 731-733.
8. Cervenka J, Arthur D, Yasis C (1981) Mitomycin C test for diagnostic differentiation of idiopathic aplastic anemia and Fanconi anemia. Pediatrics 67: 119-127.
9. Schroeder TM, Anschutz F, Knopp A (1964) [Spontaneous chromosome aberrations in familial panmyelopathy]. Humangenetik 1: 194-196.
10. Schindler D, Kubbies M, Hoehn H, Schinzel A, Rabinovitch PS (1987) Confirmation of Fanconi's anemia and detection of a chromosomal aberration (1Q12-32 triplication) via BrdU/Hoechst flow cytometry. Am J Pediatr Hematol Oncol 9: 172-177.
11. Seyschab H, Friedl R, Sun Y, Schindler D, Hoehn H, et al. (1995) Comparative evaluation of diepoxybutane sensitivity and cell cycle blockage in the diagnosis of Fanconi anemia. Blood 85: 2233-2237.
12. Wagner JE, Tolar J, Levran O, Scholl T, Deffenbaugh A, et al. (2004) Germline mutations in BRCA2: shared genetic susceptibility to breast cancer, early onset leukemia, and Fanconi anemia. Blood 103: 3226-3229.
13. Hirsch B, Shimamura A, Moreau L, Baldinger S, Hag-alshiekh M, et al. (2004) Association of biallelic BRCA2/FANCD1 mutations with spontaneous chromosomal instability and solid tumors of childhood. Blood 103: 2554-2559.
14. Vaz F, Hanenberg H, Schuster B, Barker K, Wiek C, et al. (2010) Mutation of the RAD51C gene in a Fanconi anemia-like disorder. Nat Genet 42: 406-409.
15. Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25: 1754-1760.
16. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, et al. (2009) The Sequence Alignment/Map format and SAMtools. Bioinformatics 25: 2078-2079.

17. Koboldt DC, Chen K, Wylie T, Larson DE, McLellan MD, et al. (2009) VarScan: variant detection in massively parallel sequencing of individual and pooled samples. Bioinformatics 25: 2283-2285.

18. Wang K, Li M, Hakonarson H (2010) ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res 38: e164.

19. Carr IM, Flintoff KJ, Taylor GR, Markham AF, Bonthron DT (2006) Interactive visual analysis of SNP data for rapid autozygosity mapping in consanguineous families. Hum Mutat 27: 1041-1046.

20. Schindler D, Friedl R, Gavvovidis I, Kalb R, Neveling K, et al. (2007) Applications of Cell Cycle testing in Fanconi Anemia. In: Schindler D, H H, editors. Fanconi Anemia - A Paradigmatic disease for the understanding of cancer and aging. Basel: Karger. pp. 110-130.

21. Kalb R, Neveling K, Hoehn H, Schneider H, Linka Y, et al. (2007) Hypomorphic mutations in the gene encoding a key Fanconi anemia protein, FANCD2, sustain a significant group of FA-D2 patients with severe phenotype. Am J Hum Genet 80: 895-910.

22. Timmers C, Taniguchi T, Hejna J, Reifsteck C, Lucas L, et al. (2001) Positional cloning of a novel Fanconi anemia gene, FANCD2. Mol Cell 7: 241-248.

23. Ameziane N, Sie D, Dentro S, Ariyurek Y, Kerkhoven L, et al. (2012) Diagnosis of fanconi anemia: mutation analysis by next-generation sequencing. Anemia 2012: 132856.

24. Clark MJ, Chen R, Lam HY, Karczewski KJ, Euskirchen G, et al. (2011) Performance comparison of exome DNA sequencing technologies. Nat Biotechnol 29: 908-914.

25. Shendure J, Ji H (2008) Next-generation DNA sequencing. Nat Biotechnol 26: 1135-1145.

26. Mardis ER (2008) The impact of next-generation sequencing technology on genetics. Trends Genet 24: 133-141.

Tables

Table 1. Summary of statistical data from four independent WES projects

	Project 1	Project 2	Project 3	Project 4
Total read number	33,661,920	121,791,357	152,961,886	27,371,419
Reads passing QC	32,251,042	82,558,019	117,526,556	-
Reads on target	29,837,615 (93%)	67,361,646 (82%)	83,597,787 (71%)	20,707,708 (76%)
Average exome coverage	22x	77x	71x	36x
Total number of variants	20,065	13,466	14,978	18,885
Known SNPs/MNPs	15,213	9,846	13,563	15,469
UV in cs	4,652	3,567	1,386	3,281
UV at ess	200	46	29	136
Homozygous UV (cs+ess)	107	201	44	30
Heterozygous UV (cs+ess)	4,745	3,419	1,371	3,387
Silent UV (cs)	1,012 het	73 hom	286 het	884
Missense UV (cs)	2,736 het	102 hom	920 het	2,289
Nonsense UV (cs)	113 het	1 hom	44 het	23
Unknown InDels (cs+ess)	766 het	24 hom	101 het	77
Multiply heterozygous mutated genes	683	-	207	411
Homozygous mutated genes	-	102	-	-

QC, quality control; UV, unknown variants; cs, coding sequence; ess, essential splice sites; SNP, single nucleotide polymorphism; MNP, multiple nucleotide polymorphism

The initial number of genes with homozygous or double heterozygous mutations was counted without filtering for low mutation scores or benign sequence changes.

Figure legends

Figure 1. (A) Graphic presentation of the results of flow cytometric cell cycle analysis. Peripheral blood lymphocytes were exposed to MMC. The ratio "sum of all G2 phases vs. growth fraction" was calculated from individual cultures and plotted against the G0/G1 phase compartment. Cells from patient 1 and the siblings 2-1, 2-2 and 2-3 show high ∑G2/GF ratios (blue squares) similar to those from other persons with FA (red diamonds), but were distinct of normal controls (green dots). (B) Flow histograms of fibroblast cultures from patients 3 and 4 show increased G2 arrest after exposure to MMC, in contrast to a control cell line (arrows).

Figure 2 (A) The heterozygous *FANCD2* splice site mutation c.3888+2T>G in patient 1. The upper panel demonstrates mutation calling in NGS data format. The lower panel shows an electropherogram of Sanger cDNA sequencing depicting heterozygous skipping of exon 39. (B) The heterozygous *FANCD2* missense mutation c.2204G>A. The upper panel demonstrates the substitution in NGS data format, while the lower panel shows the confirmation by Sanger sequencing of gDNA. (C) An immunoblot shows very faint FANCD2-S and -L bands after exposure of fibroblasts from patients 1 to MMC (lane 3). This was similar to other FA-D2 patients (example on lane 2) but contrasted markedly to normal controls (example on lane 1). RAD50 was used as loading control.

Figure 3 (A) Homozygous mutation call c.1878A>T in *FANCJ* detected in NGS data of patient 2-1. (B) Autozygosity mapping with SNP data of the family of project 2. The figure schematically presents chromosome 17 (positions in Mb) of each family member. Heterozygous SNP calls are displayed in yellow, homozygous calls in black. The three affected siblings share a homozygous region between 53.3 Mb and 68.8 Mb. (C)

18

Confirmation of homozygosity of the mutation in patient 2-1 and heterozygosity in his parents by Sanger sequencing electropherograms, consistent with Mendelian segregation.

Figure 4 (A) NGS data in the upper panel show the insertion c.7890_7891insAA in *FANCD1* detected in patient 3. The electropherogram in the lower panel demonstrates corresponding validation by Sanger sequencing. (B) The upper panel misleadingly displays the single-bp substitution c.7795G>A in the NGS data of patient 3 as a SNP, highlighted in pink. Confirmation by Sanger sequencing is shown below.

Figure 5 (A) Displayed is the *FANCD2* mutation c.1370T>C in patient 4 in NGS data as well as validated by Sanger sequencing. (B) The upper panel shows NGS data with low coverage of *FANCD2* exon 5 containing the substitution c.376A>G. The electropherogram in the lower panel depicts validation by Sanger sequencing. (C) Hydroxyurea (HU) treated (+) and untreated (-) fibroblasts of patient 4 show very low levels of both the S and L species of residual FANCD2 protein. Vinculin was used as loading control.
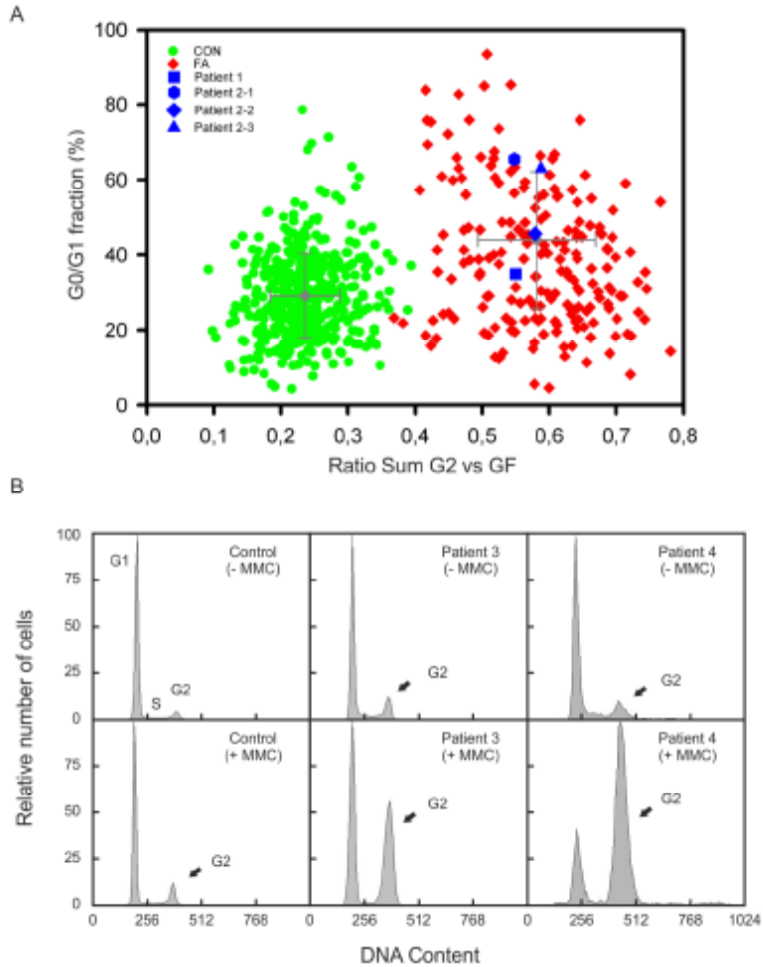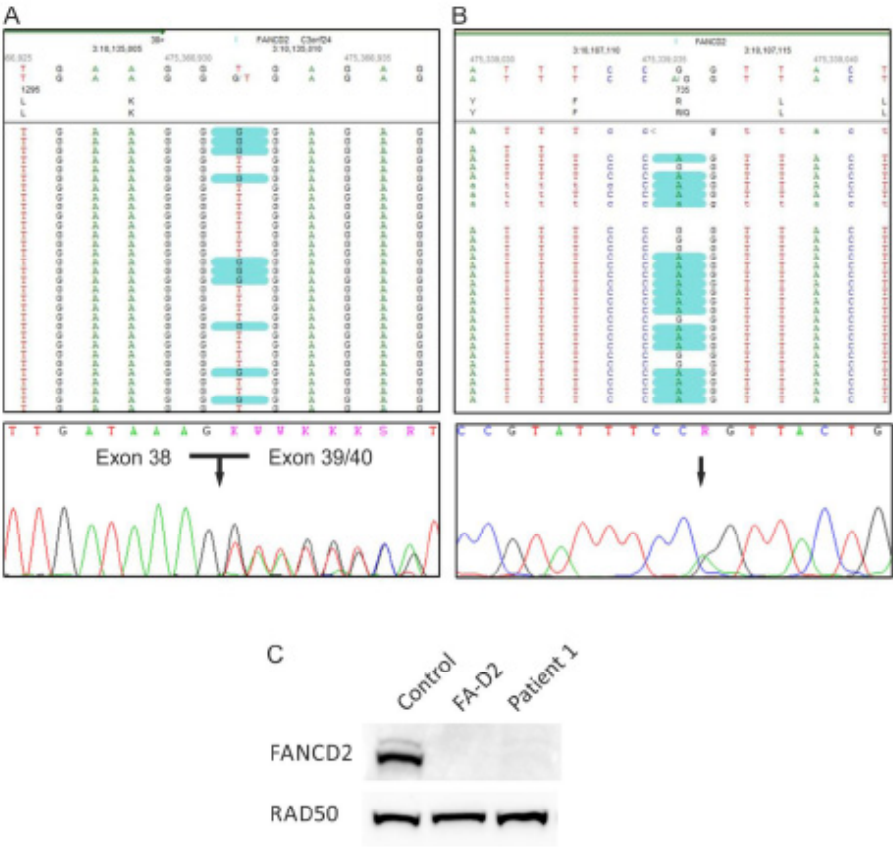
Figure 1

Figure 2

Figure 3

Figure 4

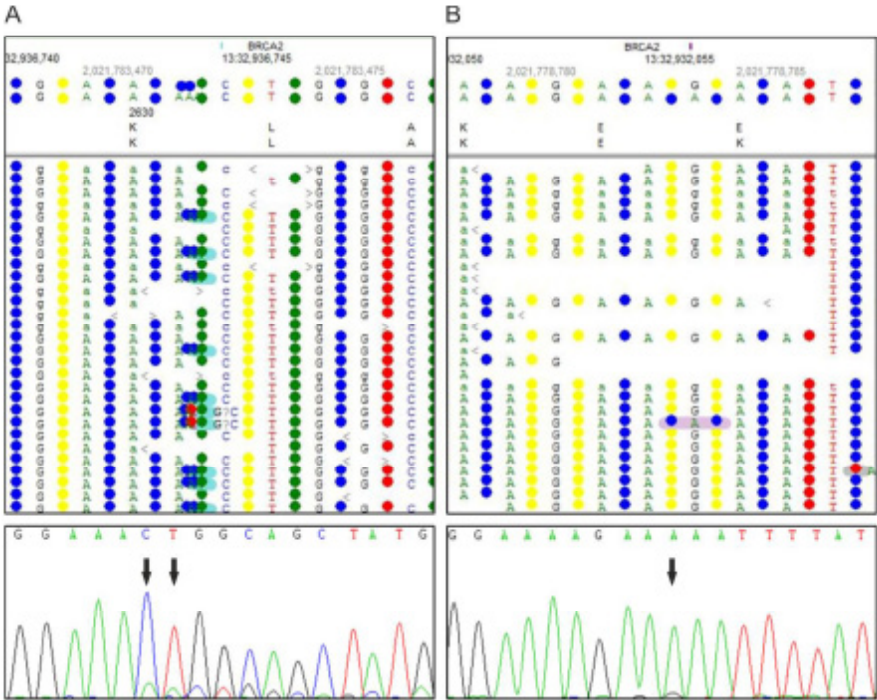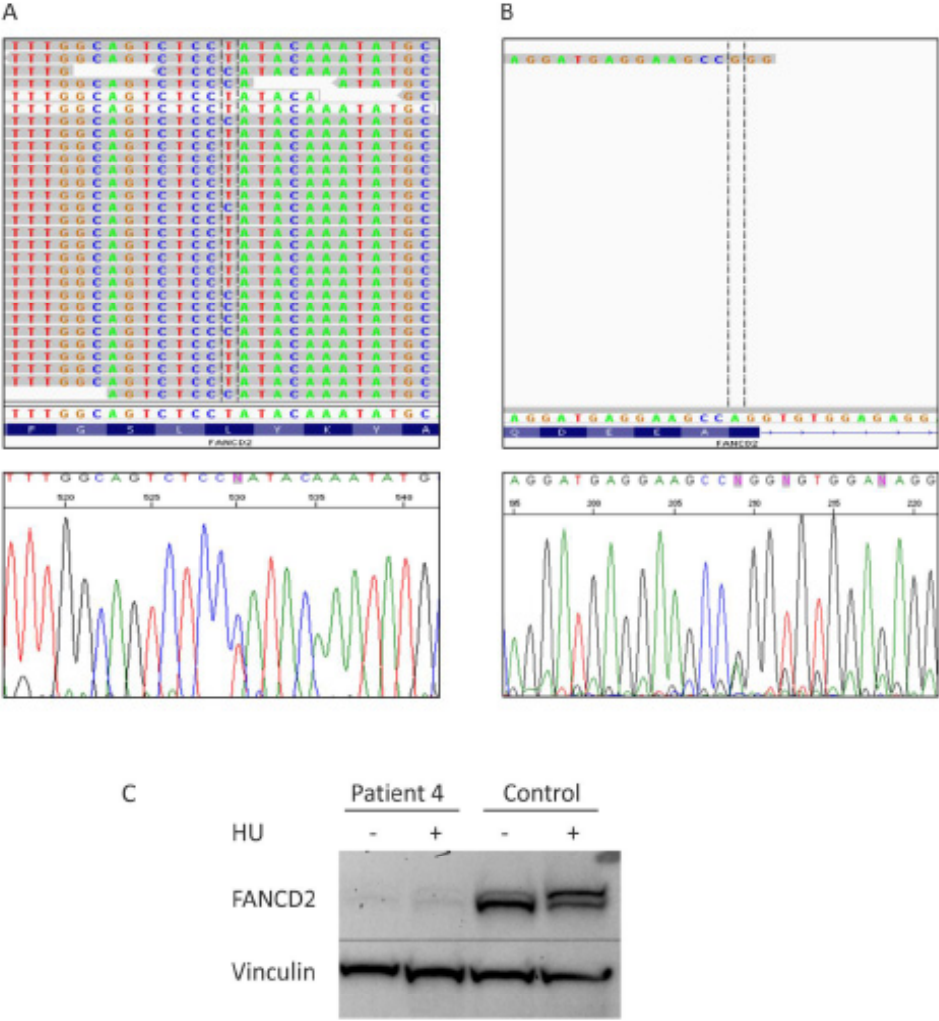Figure 5

4.   DISCUSSION

4.1   FROM KNOWN TO NOVEL GENES

This study was aimed at the identification and characterization of novel genes which might be causative for FA. For this, known and well characterized FA genes served as a basis for the identification of novel genes, which would match the FA pathway because of direct interaction or functional similarities. As already described above, co-immunoprecipitation has been used to determine proteins interacting with the FA core complex (Meetei, de Winter et al. 2003; Meetei, Sechi et al. 2003; Meetei, Levitus et al. 2004; Meetei, Medhurst et al. 2005; Ling, Ishiai et al. 2007; Yan, Delannoy et al. 2010). Two of them, MHF1 and MHF2, form together with FANCM a conserved DNA remodeling complex, which is recruited to ICLs during the S phase of the cell cycle. Since *FANCM* mutations have been found in a FA patient (Meetei, Medhurst et al. 2005), it was not unlikely to find defects in *MHF1* and *MHF2*, too. We screened unassigned FA patients with deficient FANCD2 monoubiquitination, because MHF1 and MHF2 seemed to be required for FA pathway activation and therefore acting upstream of FANCD2 monoubiquitination. However, none of the investigated patients carried mutations that would lead to altered protein expression or truncation (Yan, Delannoy et al. 2010). This result may most likely be due to the fact that the cellular phenotype of these FA patients did not exactly match that of *MHF1*- or *MHF2*-depleted cells. We distinguished patients only by being pro- or deficient for FANCD2 monoubiquitination, but we did not separate a third group with only reduced levels of ubiquitination, which would have mirrored the status in *MHF*-depleted HeLa cells more accurately. Additionally, we found *MHF*-deficient cells being sensitive for camptothecin, which is not a typical feature of cells carrying mutations in a member of the FA core complex, but has so far been found only in cells of subtypes FA-D1, -M, -N, -O and -P and is supposed to be characteristic for HR-defective cells (Andressoo, Mitchell et al. 2006; Yan, Delannoy et al. 2010; Stoepker, Hain et al. 2011). Though the MHF heterodimer seems to be required for MMC resistance and chromatin loading of the core complex, there is evidence that it has other functions independent of the FA pathway activation. Those include suppression of sister chromatid exchanges by interacting with BLM helicase and stabilization of the centromere by assembly of the kinetochore structure (Niedernhofer, Garinis et al. 2006; Yan, Delannoy et al. 2010). Accordingly it may also be possible that a $MHF^{-/-}$ phenotype would greatly differ from that of FA patients or that mutations in the corresponding genes might even be lethal. Interestingly, *FAAP24*, another important interactor of FANCM, has also never been found mutated in FA patients (unpublished data). In contrast to the MHF hetero-dimer, which is supposed to bind double-stranded DNA on stalled replication forks, FAAP24 is able to target FANCM to single-stranded DNA, an intermediate of replication

as well as HR repair (Ciccia, Ling et al. 2007; Yan, Delannoy et al. 2010). Those multiple functions of FA and FA-associated proteins make it a priori difficult, if not impossible to find a proper group of patients, to be investigated.

Something similar probably holds true for a further promising candidate, *FAN1*, which seemed to be most likely a novel FA gene. FAN1 is recruited to stalled replication forks by monoubiquitinated FANCD2, its nuclease activity is required for ICL repair and its depletion leads to increased MMC induced chromosomal breakage (Kratz, Schopf et al. 2010; Liu, Ghosal et al. 2010; MacKay, Declais et al. 2010; Smogorzewska, Desetty et al. 2010). Nevertheless we have not been able to detect an FA patient with FAN1 deficiency. In contrast, we showed that patients carrying a homozygous microdeletion including the *FAN1* locus suffer from features like mental retardation, epilepsy and muscular hypotonia, which are not typical for FA. Moreover, they did not develop hematological malignancies or BMF. Shared with FA are merely microcephaly and sensitivity to crosslinking drugs, though the latter is only mild in *FAN1* microdeletion patients and not comparable to known FA subtypes (Trujillo, Mina et al. 2012). Interestingly, Zhou et al. (2012) most recently described *FAN1* mutations as cause of late onset karyomegalic interstitial nephritis (KIN). Progressive kidney failure has not been described for *FAN1* microdeletion patients, most likely because those patients were still too young. However, it seems that the more severe the functional defect of FAN1, the more severely and earlier the phenotypical characteristics emerge. Notably, progressive kidney failure with polyploidization and renal insufficiency is a feature of aging and frequently observed in genome maintenance syndromes (Lans and Hoeijmakers 2012; Zhou, Otto et al. 2012). Certainly all those results excluded *FAN1* as FA gene, but likewise they support its importance for DNA repair and genomic stability. However, there is a connection to the FA pathway and the question on the role of FAN1 remains to be answered. FAN1 seems neither to be crucial for pathway activation nor for ICL excision, which may be performed by the heterodimeric nucleases MUS81/EME1 and XPF/ERCC1 (Kratz, Schopf et al. 2010). Initiation of HR repair, which is detectable by unimpaired RAD51 foci formation, is also not impaired by FAN1 depletion. However, the prolonged existence of these foci gives some evidence that FAN1 probably plays a role in late steps of HR repair (Kratz, Schopf et al. 2010; MacKay, Declais et al. 2010). Several studies showed that FAN1 is not only interacting with monoubiquitinated FANCD2, but also with mismatch repair (MMR) proteins like MLH1, MLH3 or PMS2 (Sijbers, de Laat et al. 1996; Kratz, Schopf et al. 2010; Smogorzewska, Desetty et al. 2010). This interplay raises the possibility that FAN1-mediated ICL resistance depends on functional MMR. Since a direct interaction between MLH1 and FANCJ was also shown to be necessary for ICL repair, it could be discussed that FAN1 might connect FANCD2 and FANCJ, whose role in the FA pathway remains elusive as well (Smogorzewska, Desetty et al. 2010).

By these and other candidate gene considerations it becomes apparent that the DNA repair

network, including the FA pathway, is pretty complicated and consists of highly complex protein interplays. Thus, the sheer interaction of a hypothetical with a known FA protein cannot guarantee a straightforward role in the FA pathway and the possibility for pathological findings leading to FA. The identification of *RAD51C* mutations in a patient with FA-like symptoms is just emphasizing that FA is not only genetically heterogeneous and molecular complex, but also has a broad phenotypical spectrum, which does not always match the expectations. *RAD51C* was considered a candidate gene by the results of homozygosity mapping and confirmed as disease causing gene by complementation analysis in a consanguineous family with one healthy and three affected children. Additionally there was an abortion, whose fetus could not be examined. Amongst the affected siblings, two died very early. One was alive and 10 years old at time of publication. He showed a typical phenotype including bilateral radial hypoplasia and undescended testes, but up to the time of publication he had not developed any hematological problems which could have further strengthened the clinical suspicion of FA. Nevertheless, his own and the affected sister´s blood lymphocytes showed increased chromosomal instability after treatment with MMC or DEB, making the diagnosis of FA most likely (Vaz, Hanenberg et al. 2010). RAD51C emerged as part of the FA/BRCA network in the course of this study. It is one of the RAD51 paralogs and required for RAD51 foci formation just like the FA proteins FANCD1 and FANCN (Takata, Sasaki et al. 2001; Godthelp, Artwert et al. 2002; Zhang, Fan et al. 2009). Additionally, RAD51C deficiency increases sensitivity towards the topoisomerase I inhibitor camptothecin, which has also been described for other FA proteins involved in HR. Even though *RAD51C* could be validated as disease causing gene, the functional effect of the R258H missense mutation remains elusive. Homology-modeling of the archaeal rad51 protein of Pyrococcus furiosus gave rise to some speculations about a structural effect disrupting the N-terminal and ATPase domains, but the mutation does not affect protein stability (Vaz, Hanenberg et al. 2010). *RAD51C* mutations being rare and hypomorphic is not surprising regarding that a complete knockout in mice leads to early embryonic lethality. Only conditional knockouts are viable and show normal growth but reduced fertility (Kuznetsov, Pellegrini et al. 2007; Kuznetsov, Haines et al. 2009). A role of RAD51C in early developmental stages may explain the severe congenital abnormalities observed in the affected siblings. Despite the cellular and molecular analogy of RAD51C with FANCD1 and FANCN, the described mutation may not be associated with early onset of cancer in FA patients. Since the only living patient has so far not developed any hematological anomalies, which is the hallmark feature of FA, we are supposed to designate *RAD51C* tentatively as FA-like gene, *FANCO*. Nevertheless, our findings led to the presumption that *RAD51C*, just like the other FA genes involved in HR, might be a breast and ovarian cancer susceptibility gene in the heterozygous state. Since then, several independent studies on European, Asian, Australian and American cohorts of breast and ovarian cancer families examined the occurrence of

RAD51C mutations and tried to estimate the relative risk for mutation carriers. Three studies on breast and ovarian cancer and one on male breast cancer patients did not reveal any deleterious sequence changes (Akbari, Tonin et al. 2010; Zheng, Zhang et al. 2010; Silvestri, Rizzolo et al. 2011; Wong, Nordfors et al. 2011). Contrary to that, eleven studies on partly large cohorts of several hundred cases indeed could confirm *RAD51C* mutations in affected families (Meindl, Hellebrand et al. 2010; Clague, Wilhoite et al. 2011; Pang, Yao et al. 2011; Pelttari, Heikkinen et al. 2011; Romero, Perez-Segura et al. 2011; Vuorela, Pylkas et al. 2011; Coulet, Fajac et al. 2012; Loveday, Turnbull et al. 2012; Osorio, Endt et al. 2012; Thompson, Boyle et al. 2012). Interestingly, pathogenic mutations have been found only in families with breast and ovarian cancer or ovarian cancer only, leading to the conclusion that there might not be an overall increased risk, but a specific risk of developing ovarian cancer. The occurrence of *RAD51C* mutations in familial breast and/or ovarian cancer seems now to be established, but their penetrance is still controversially discussed in the literature.

In analogy to the earlier discussed *FAN1*, *SLX4* became a FA candidate gene because multiple back-to-back studies highlighted its role in DNA repair and HR. The highly conserved protein was described as coordinator of the structure-specific nucleases XPF/ERCC1, MUS81/EME1 and SLX1. But especially the hypersensitivity of SLX4 depleted cells to crosslinking drugs and camptothecin prompted different research groups to investigate a possible connection to FA (Andersen, Bergstralh et al. 2009; Fekairi, Scaglione et al. 2009; Munoz, Hain et al. 2009; Svendsen, Smogorzewska et al. 2009). Our study, together with the concurrent report of Kim et al. (2011), first described FA patients carrying mutations in *SLX4*. The patients' phenotypes were diverging, yet in each case typical for FA. All of them developed hematological anomalies in early childhood and most of them had to be transplanted before the age of 10. Growth retardation, at least prenatal, seems also to be common in biallelic *SLX4* mutation carriers. Two patients appeared to be more severely affected and suffered from several FA-typical congenital malformations including radial ray defects, microcephaly and hearing loss. Both of them reached early adulthood, though one died at the age of 22 years from metastatic SCC of the tongue (Kim, Lach et al. 2011; Stoepker, Hain et al. 2011). Considering all these findings, *SLX4* doubtlessly is a novel FA gene and disease causing in the subtype FA-P. Unlike in the case of *RAD51C*, here not only the cellular manifestations are characteristic, but also the above described clinical phenotype perfectly matches with FA. Moreover we were able to detect an additional FA-P patient by NGS more than one year after the first description of *SLX4* mutations (Schuster, Knies et al. 2012, accepted for publication). This patient again showed very typical features including short stature, microcephaly and hypoplasia of the right thumb. Though she developed hematological abnormalities relatively late for FA, at the age of 19, her clinical phenotype is in complete accordance to the other patients. Altogether, until now, seven FA-P patients from five unrelated kindreds

of distinct geographical origin have been reported and their number therefore is comparable to other rare FA-subtypes. Additional confirmation for our results was provided by the publication of a *Slx4* knockout mouse, which properly phenocopies the human inherited disorder FA. Slx4 deficient mice display not only growth retardation and reduced fertility, but they are also prone to dysfunction of the hematopoietic system (Crossan, van der Weyden et al. 2011).

Since all other downstream FA genes had so far been shown to be breast and/or ovarian cancer susceptibility genes, the screening of corresponding families for *SLX4* mutations was a matter of course. Three studies on Spanish, Italian and German/Byelorussian patients did not reveal any clearly pathogenic sequence changes (Landwehr, Bogdanova et al. 2011; Catucci, Colombo et al. 2012; Fernandez-Rodriguez, Quiles et al. 2012). This could be due to a very low carrier frequency and the relative small cohorts used for analysis. But it might also be possible that *SLX4* is not linked to breast and ovarian cancer, because its function seems to differ from that of other downstream genes.

The distinct molecular role of SLX4 and its interacting nucleases in the FA pathway, remains a subject of discussion. The Holliday junction resolvase function mediated through SLX4/SLX1 interaction seems not to be important for FA. It is more likely the coordination of XPF/ERRC1 nuclease that makes SLX4 indispensable for ICL repair (Bergstralh and Sekelsky 2008; Wood 2010; Stoepker, Hain et al. 2011). While Slx4 with a defect in Slx1 interaction could restore MMC resistance in knockout mice, Slx4 deficient of Xpf/Ercc1 interaction could not (Crossan, van der Weyden et al. 2011). Furthermore, subcellular fractionation and fluorescence microscopy experiments in human FA-P cell lines showed a distinct defect in chromatin loading of XPF/ERCC1, but not of MUS81/EME1 or SLX1 (Stoepker, Hain et al. 2011; Schuster, Knies et al. 2012, accepted for publication). These findings defined the major function of SLX4 in ICL repair as XPF/ERCC1 localization, but as SLX4 is not only interacting with the already mentioned nucleases, there might be more than this. In addition to NER proteins, also the MMR heterodimer MSH2/MSH3 and telomere-binding proteins were found to be members of the SLX4-interactome, leading to the conclusion that there might be multiple DNA repair activities depending on SLX4 (Svendsen, Smogorzewska et al. 2009; Cybulski and Howlett 2011).

The identification of *SLX4* as underlying gene of the FA complementation group FA-P brought all SLX4 interacting proteins into the focus of FA candidate gene search. Consequentially, it was not really surprising that in the following year two FA patients were found to carry biallelic mutations in *XPF*, which could now probably be designated *FANCQ* (Bogliolo, Schuster et al. 2012, submitted). Interestingly, *XPF* had already been described before as disease causing gene for xeroderma pigmentosum (XP) subtype F as well as for segmental progeria (XFE) (Sijbers, de Laat et al. 1996; Niedernhofer, Garinis et al. 2006). By extensive investigations we could show that depending on the different mutations, *XPF* indeed can be causative for three distinct phenotypes.

While the hallmark features of FA are hematological abnormalities, growth retardation and the increased sensitivity to crosslinking drugs, XP is mainly characterized by skin photosensitivity, which seems to be relatively mild in XP-F patients (Gregg, Robinson et al. 2011). Cells of XP patients typically show increased sensitivity towards UV radiation, but do not accumulate in the G2 phase of the cell cycle after exposure to MMC. XFE, which has so far been described only for one patient, represents a combination of both, FA- and XP-like features. The patient described by Niedernhofer et al. (2006) suffered from severe photosensitivity of the skin, short stature and anemia. His cells were hypersensitive towards UV radiation as well as crosslinking drugs. We therefore hypothesized that FA-Q cells are deficient for ICL repair, but might have sufficient NER activity to prevent from the presence of typical XP symptoms. Functional studies showed that the FA-specific mutated XPF protein, in contrast to that responsible for XP, could localize to the nucleus and perform NER in a sufficient manner. Additional *in vitro* experiments using pathway-specific DNA substrates showed that the FA-specific point mutants were able to perform the excision step on a plasmid containing an ICL, but failed to cleave the stem-loop model resembling a stalled replication fork. Since those results could be observed for the XP causing mutants in exactly reverse manner, we concluded that the determining factor is not the catalytic activity of XPF, but considerably more its exact positioning at the DNA damage site. Even though the precise function of XPF in the FA pathway remains elusive, we excluded a role in the incision step. But it is also unlikely that it plays a role in HR, because FA-Q cells are proficient for RAD51 foci formation and did not show hypersensitivity towards campthotecin (Bogliolo, Schuster et al. submitted 2012).

Resolving the distinct roles of FA-associated nucleases and other interacting proteins is currently one of the major topics in FA research and will shed light on the highly complex enzymatic activities, which are necessary for ICL repair. Extensive studies established an important role for the MUS81/EME1 nuclease in the FA pathway by converting ICLs into one-ended DSBs and thereby making the stalled replication fork accessible for the succeeding repair machinery (Osman and Whitby 2007; Sengerova, Wang et al. 2011). But it could just as well be replaced by FANCM/FAAP24, which also belongs to the XPF/MUS81 family of structure-specific nucleases and could therefore not only be responsible for checkpoint activation and core complex loading but as well for replication fork reversal (Cybulski and Howlett 2011). The so-called ICL unhooking requires not only nucleases for incision 3' and 5' of the ICL, but also proteins with helicase activity for previous unwinding of the DNA double helix. Although there are helicases in the NER system, which could perform this task, it is more than likely that the FA protein FANCJ acts synergistically with BLM in this step of ICL repair (Wood 2010; Suhasini and Brosh 2012). Not only the crucial step of ICL unhooking, but also pathway activation and the downstream steps of translesion synthesis, recombination repair and replication fork restart are

far from being fully understood and keep several gaps which could be filled by novel FA genes. The connection of the FA/BRCA network to NER and most likely to MMR proteins via FANCJ, -D2 and -P may provide new candidates for the later steps of the FA pathway, while the most promising candidates for damage recognition and pathway activation will, like e.g. FAAP100, be associated with the core complex.

4.2   From Classical Genetic Approaches to Next Generation Sequencing

After several successful applications of WES in the search of disease causing genes for other Mendelian disorders, the previously discussed *FANCQ* was the first FA gene to be identified by this novel approach (Bogliolo, Schuster et al. 2012, submitted). In addition to the application of NGS in FA research, this work aimed to test and establish this technology also in FA diagnostics. FA is genetically very heterogeneous but yet it follows classical Mendelian inheritance and the vast majority of pathogenic mutations lies within exons or exon-intron boundaries. Therefore we decided for WES being not only the most appropriate NGS approach for FA research but also for diagnostic purposes. Reasonable outsourcing of exome capture and sequencing processes left only bioinformatic analysis as in-house challenge. Therefore, we developed an FA-optimized analysis pipeline using the commercial software NextGENe® from Softgenetics. Because of ethical issues, which will be discussed later on, the complete sequencing data set of a patient was only used for quality control and statistical analysis. For mutation screening we concentrated only on the 15 validated and the five associated FA genes *FAAP20, FAAP24, FAAP100, MHF1* and *MHF2*. After exclusion of all polymorphisms listed in the NCBI (National Center for Biotechnology Information) dbSNP database, we filtered for deleterious heterozygous or homozygous mutations depending on the patient´s family background. Indeed WES displayed a high effectivity for identification of pathogenic FA mutations. Two patients could be assigned to complementation group FA-D2, one to FA-J, another one to FA-D1 and a last one to subtype FA-P (Schuster, Knies et al. accepted for publication 2012; Schuster, Knies et al. submitted 2012). Additionally, we could validate results for one patient carrying mutations in the upstream gene *FANCG* (unpublished data). In summary, our studies showed that WES is a valuable tool for molecular FA diagnostics and is able to replace classical genetic approaches by combining complementation group assignment and mutation detection in one step. But our studies also uncovered some problems which might pose a challenge in some cases. By combining different enrichment strategies and different depths of coverage, we came to the conclusion that independent of the enrichment technique a high coverage of 100-fold in average is needed to completely cover all exons of the FA genes. As already mentioned in Majewski et al. (2011), a 30-fold coverage might be sufficient in most cases,

but 100-fold is more reliable and can diminish sequencing errors. We found that increasing the read depth can also compensate for difficult-to-enrich and difficult-to-sequence regions, which are usually but not always characterized by high GC- or AT-contents. A problem that cannot be solved so easily arises from pseudogenes. These originate either from segmental duplications or retrotranspositional events and represent highly similar, disabled, full or partial copies of a functional gene (Khurana, Lam et al. 2010). The relatively short read length used in WES in combination with pseudogene sequences can lead to misalignments, which cannot be discovered as easily as sequencing errors (Fuentes Fajardo, Adams et al. 2012). False-positives caused by pseudogenes, representing a processed and re-integrated transcript, are more likely to be recognized because those more or less exactly copy the functional exonic sequences and probably cannot be distinguished already during exome capture process. Depending on the pseudogenes' quantity and age, the misalignment will become visible in WES data by mutation call rates exceeding the expected by a multiple; good examples for this are *CDC27* as well as *MUC* and *NBPF* gene families. Young pseudogenes with few sequence changes or pseudogenes including both, exonic and intronic regions, seem to be more difficult, because misalignments are not obligatory and if occurring they are not as obvious. Amongst the FA genes, so far only *FANCD2* should be regarded with suspicion. Two related pseudogenes have been described and consist of intronic and exonic parts with a similiarity of up to 100% (Kalb, Neveling et al. 2007). General information on pseudogenes is provided e.g. on the NCBI homepage and on www.pseudgenes.org, but since probably none of the databases is complete, which is examplified on the contradictory entries on *FANCL* and *FANCN*, we recommend careful comparison. Currently the only option to be sure might be validation by specific Sanger re-sequencing or maybe using longer reads like they are generated by Roche 454® sequencing or the "Third Generation Sequencer" from Pacific Bioscience®. Other problems, which we observed, mainly concern inconsistent and incomplete database entries regarding alternative transcripts and known polymorphisms. But the more data WES, transcriptome and whole genome sequencing (WGS) projects are generating in near future, the more information will be integrated in these databases and the more reliable they will become.

For diagnostic purpose WES provides sufficient information about exonic regions and intron-exon boundaries which are most likely spots of pathogenic mutations (Majewski, Schwartzentruber et al. 2011). Nevertheless, we have to accept that this approach misses deep intronic sequence changes, which could be covered by WGS. But those alterations are rare and anyway would need further confirmation on transcript level. Even if WGS provides more information, in most cases it would be redundant, unnecessarily increasing expense and dramatically exacerbate analysis and interpretation. Transcriptome sequencing could be an alternative. This approach would decrease costs by concentrating on the coding sequence and making the enrichment step unnecessary. This technique could not only identify single and oligo nucleotide variants but also

large deletions, insertion or other rearrangements, yet it has its drawbacks. Transcriptome sequencing is tissue specific and depends on reverse transcription of RNA into cDNA, which can be inaccurate and incomplete. Low or unstable expressed transcripts might get lost and therefore cause a lot of false-negative results (Cirulli, Singh et al. 2010; Ozsolak and Milos 2011; Ku, Wu et al. 2012). The last possibility would have been NGS after enrichment of all FA genes as performed by Ameziane et al. (2012). This approach absolutely implies some advantages but in almost the same manner it again has some weakness compared to WES, which gives us the chance to search for the disease causing gene beyond the known FA genes, as required. Target enrichment can increase the coverage by low sequencing effort and it provides very detailed information, which is not only accelerating the analysis process but is also important form an ethical point of view (Ku, Cooper et al. 2012). While WES will always reveal several incidental findings, target enrichment is concentrated only on genes or regions of interest and therefore does not come into conflict with ethical issues. Up to now there is no clear consent about incidental findings in WES or WGS (Ku, Cooper et al. 2012). Some agreement might be that by detecting serious but treatable conditions, researchers have a moral obligation to forward this information to the patients. But in the case of less serious, untreatable or uncertain findings the patient´s right not to know should be weighed against the potential benefits (Wolf, Lawrenz et al. 2008; Majewski, Schwartzentruber et al. 2011). Integrating WES into routine diagnostics must give this issue serious consideration and should be accompanied by developing standard procedures. For disorders with locus heterogeneity, like FA, patients should be asked to give written informed consent not only for diagnostics, but also for research purpose. This will give us the chance to take the full advantage from WES and will facilitate to completely uncover the genetic basis of FA by identification of all involved genes. It will be interesting which candidates can be added to the FA pathway and how they will contribute to a better understanding of the DNA repair network that is able to eliminate one of the most deleterious lesions and prevents our cells from malignant degradation, premature senescence and cell death.

5. REFERENCES

Akbari, M. R., P. Tonin, et al. (2010). "RAD51C germline mutations in breast and ovarian cancer patients." Breast Cancer Res 12(4): 404.

Alter, B. P. and G. Kupfer (1993). Fanconi Anemia. Gene Reviews. R. A. Pagon, T. D. Bird, C. R. Dolan and K. Stephens. Seattle (WA).

Ameziane, N., D. Sie, et al. (2012). "Diagnosis of fanconi anemia: mutation analysis by next-generation sequencing." Anemia 2012: 132856.

Andersen, S. L., D. T. Bergstralh, et al. (2009). "Drosophila MUS312 and the vertebrate ortholog BTBD12 interact with DNA structure-specific endonucleases in DNA repair and recombination." Mol Cell 35(1): 128-135.

Andressoo, J. O., J. R. Mitchell, et al. (2006). "An Xpd mouse model for the combined xeroderma pigmentosum/Cockayne syndrome exhibiting both cancer predisposition and segmental progeria." Cancer Cell 10(2): 121-132.

Apostolou S, Whitmore S A, et al. (1996). "Positional cloning of the Fanconi anaemia group A gene. The Fanconi anaemia/breast cancer consortium." Nature Genetics 14: 324-328.

Auerbach, A. D. (2009). "Fanconi anemia and its diagnosis." Mutat Res 668(1-2): 4-10.

Bamshad, M. J., S. B. Ng, et al. (2011). "Exome sequencing as a tool for Mendelian disease gene discovery." Nat Rev Genet 12(11): 745-755.

Bergstralh, D. T. and J. Sekelsky (2008). "Interstrand crosslink repair: can XPF-ERCC1 be let off the hook?" Trends Genet 24(2): 70-76.

Bogliolo, M., B. Schuster, et al. (2012, submitted). "XPF mutations severely disrupting DNA interstrand crosslink repair cause Fanconi anemia."

Callen, E., J. A. Casado, et al. (2005). "A common founder mutation in FANCA underlies the world's highest prevalence of Fanconi anemia in Gypsy families from Spain." Blood 105(5): 1946-1949.

Catucci, I., M. Colombo, et al. (2012). "Sequencing analysis of SLX4/FANCP gene in Italian familial breast cancer cases." PLoS One 7(2): e31038.

Ciccia, A. and S. J. Elledge (2010). "The DNA damage response: making it safe to play with knives." Mol Cell 40(2): 179-204.

Ciccia, A., C. Ling, et al. (2007). "Identification of FAAP24, a Fanconi anemia core complex protein that interacts with FANCM." Mol Cell 25(3): 331-343.

Cirulli, E. T., A. Singh, et al. (2010). "Screening the human exome: a comparison of whole genome and whole transcriptome sequencing." Genome Biol 11(5): R57.

Clague, J., G. Wilhoite, et al. (2011). "RAD51C germline mutations in breast and ovarian cancer cases from high-risk families." PLoS One 6(9): e25632.

Constantinou, A. (2012). "Rescue of replication failure by Fanconi anaemia proteins." Chromosoma 121(1): 21-36.

Coulet, F., A. Fajac, et al. (2012). "Germline RAD51C mutations in ovarian cancer susceptibility." Clin Genet.

Crossan, G. P., L. van der Weyden, et al. (2011). "Disruption of mouse Slx4, a regulator of structure-specific nucleases, phenocopies Fanconi anemia." Nat Genet 43(2): 147-152.

Cybulski, K. E. and N. G. Howlett (2011). "FANCP/SLX4: a Swiss army knife of DNA interstrand crosslink repair." Cell Cycle 10(11): 1757-1763.

de Winter, J. P., F. Leveille, et al. (2000). "Isolation of a cDNA representing the Fanconi anemia complementation group E gene." Am J Hum Genet 67(5): 1306-1308.

de Winter, J. P., M. A. Rooimans, et al. (2000). "The Fanconi anaemia gene FANCF encodes a novel protein with homology to ROM." Nat Genet 24(1): 15-16.

de Winter, J. P., Q. Waisfisz, et al. (1998). "The Fanconi anaemia group G gene FANCG is identical with XRCC9." Nat Genet 20(3): 281-283.

Deans, A. J. and S. C. West (2011). "DNA interstrand crosslink repair and cancer." Nat Rev Cancer 11(7): 467-480.

Du, W., Z. Adam, et al. (2008). "Oxidative stress in Fanconi anemia hematopoiesis and disease progression." Antioxid Redox Signal 10(11): 1909-1921.

Enders, G. H. (2008). "Expanded roles for Chk1 in genome maintenance." J Biol Chem 283(26): 17749-17752.

Fanconi, G. (1927). Familiäre infantile perniziosaartige Anämie (perniziöses Blutbild und Konstitution). Jahrbuch für Kinderheilkunde und physische Erziehung. Wien. 117: 257-280.

Fekairi, S., S. Scaglione, et al. (2009). "Human SLX4 is a Holliday junction resolvase subunit that binds multiple DNA repair/recombination endonucleases." Cell 138(1): 78-89.

Fernandez-Rodriguez, J., F. Quiles, et al. (2012). "Analysis of SLX4/FANCP in non-BRCA1/2-mutated breast cancer families." BMC Cancer 12: 84.

Fuentes Fajardo, K. V., D. Adams, et al. (2012). "Detecting false-positive signals in exome sequencing." Hum Mutat 33(4): 609-613.

Garaycoechea, J. I., G. P. Crossan, et al. (2012). "Genotoxic consequences of endogenous aldehydes on mouse haematopoietic stem cell function." Nature.

Garcia-Higuera, I., T. Taniguchi, et al. (2001). "Interaction of the Fanconi anemia proteins and BRCA1 in a common pathway." Mol Cell 7(2): 249-262.

Garner, E. and A. Smogorzewska (2011). "Ubiquitylation and the Fanconi anemia pathway." FEBS Lett 585(18): 2853-2860.

Godthelp, B. C., F. Artwert, et al. (2002). "Impaired DNA damage-induced nuclear Rad51 foci formation uniquely characterizes Fanconi anemia group D1." Oncogene 21(32): 5002-5005.

Gregg, S. Q., A. R. Robinson, et al. (2011). "Physiological consequences of defects in ERCC1-XPF DNA repair endonuclease." DNA Repair (Amst) 10(7): 781-791.

Gulbis, B., A. Eleftheriou, et al. (2010). "Epidemiology of rare anaemias in Europe." Adv Exp Med Biol 686: 375-396.

Hiom, K. (2010). "FANCJ: solving problems in DNA replication." DNA Repair (Amst) 9(3): 250-256.

Hlavin, E. M., M. B. Smeaton, et al. (2010). "Initiation of DNA interstrand cross-link repair in mammalian cells." Environ Mol Mutagen 51(6): 604-624.

Howlett, N. G., T. Taniguchi, et al. (2002). "Biallelic inactivation of BRCA2 in Fanconi anemia." Science 297(5581): 606-609.

Ishiai, M., H. Kitao, et al. (2008). "FANCI phosphorylation functions as a molecular switch to turn on the Fanconi anemia pathway." Nat Struct Mol Biol 15(11): 1138-1146.

Joenje, H., F. Arwert, et al. (1981). "Oxygen-dependence of chromosomal aberrations in Fanconi's anaemia." Nature 290(5802): 142-143.

Kalb, R., K. Neveling, et al. (2007). "Hypomorphic mutations in the gene encoding a key Fanconi anemia protein, FANCD2, sustain a significant group of FA-D2 patients with severe phenotype." Am J Hum Genet 80(5): 895-910.

Khurana, E., H. Y. Lam, et al. (2010). "Segmental duplications in the human genome reveal details of pseudogene formation." Nucleic Acids Res 38(20): 6997-7007.

Kim, J. M., Y. Kee, et al. (2008). "Cell cycle-dependent chromatin loading of the Fanconi anemia core complex by FANCM/FAAP24." Blood 111(10): 5215-5222.

Kim, Y., F. P. Lach, et al. (2011). "Mutations of the SLX4 gene in Fanconi anemia." Nat Genet 43(2): 142-146.

Kitao, H. and M. Takata (2011). "Fanconi anemia: a disorder defective in the DNA damage res-

ponse." Int J Hematol 93(4): 417-424.

Kratz, K., B. Schopf, et al. (2010). "Deficiency of FANCD2-associated nuclease KIAA1018/FAN1 sensitizes cells to interstrand crosslinking agents." Cell 142(1): 77-88.

Ku, C. S., D. N. Cooper, et al. (2012). "Exome sequencing: dual role as a discovery and diagnostic tool." Ann Neurol 71(1): 5-14.

Ku, C. S., M. Wu, et al. (2012). "Exome versus transcriptome sequencing in identifying coding region variants." Expert Rev Mol Diagn 12(3): 241-251.

Kuznetsov, S., M. Pellegrini, et al. (2007). "RAD51C deficiency in mice results in early prophase I arrest in males and sister chromatid separation at metaphase II in females." J Cell Biol 176(5): 581-592.

Kuznetsov, S. G., D. C. Haines, et al. (2009). "Loss of Rad51c leads to embryonic lethality and modulation of Trp53-dependent tumorigenesis in mice." Cancer Res 69(3): 863-872.

Landwehr, R., N. V. Bogdanova, et al. (2011). "Mutation analysis of the SLX4/FANCP gene in hereditary breast cancer." Breast Cancer Res Treat 130(3): 1021-1028.

Lans, H. and J. H. Hoeijmakers (2012). "Genome stability, progressive kidney failure and aging." Nat Genet 44(8): 836-838.

Leung, J. W., Y. Wang, et al. (2012). "Fanconi anemia (FA) binding protein FAAP20 stabilizes FA complementation group A (FANCA) and participates in interstrand cross-link repair." Proc Natl Acad Sci U S A 109(12): 4491-4496.

Levitus, M., Q. Waisfisz, et al. (2005). "The DNA helicase BRIP1 is defective in Fanconi anemia complementation group J." Nat Genet 37(9): 934-935.

Levran, O., C. Attwooll, et al. (2005). "The BRCA1-interacting helicase BRIP1 is deficient in Fanconi anemia." Nat Genet 37(9): 931-933.

Ling, C., M. Ishiai, et al. (2007). "FAAP100 is essential for activation of the Fanconi anemia-associated DNA damage response pathway." EMBO J 26(8): 2104-2114.

Litman, R., M. Peng, et al. (2005). "BACH1 is critical for homologous recombination and appears to be the Fanconi anemia gene product FANCJ." Cancer Cell 8(3): 255-265.

Liu, T., G. Ghosal, et al. (2010). "FAN1 acts with FANCI-FANCD2 to promote DNA interstrand cross-link repair." Science 329(5992): 693-696.

Lo Ten Foe, J. R., M. A. Rooimans, et al. (1996). "Expression cloning of a cDNA for the major Fanconi anaemia gene, FAA." Nat Genet 14(3): 320-323.

Lobitz, S. and E. Velleuer (2006). "Guido Fanconi (1892-1979): a jack of all trades." Nat Rev

Cancer 6(11): 893-898.

Loveday, C., C. Turnbull, et al. (2012). "Germline RAD51C mutations confer susceptibility to ovarian cancer." Nat Genet 44(5): 475-476; author reply 476.

MacKay, C., A. C. Declais, et al. (2010). "Identification of KIAA1018/FAN1, a DNA repair nuclease recruited to DNA damage by monoubiquitinated FANCD2." Cell 142(1): 65-76.

Majewski, J., J. Schwartzentruber, et al. (2011). "What can exome sequencing do for you?" J Med Genet 48(9): 580-589.

McDaniel, L. D. and R. A. Schultz (2008). "XPF/ERCC4 and ERCC1: their products and biological roles." Adv Exp Med Biol 637: 65-82.

Meetei, A. R., J. P. de Winter, et al. (2003). "A novel ubiquitin ligase is deficient in Fanconi anemia." Nat Genet 35(2): 165-170.

Meetei, A. R., M. Levitus, et al. (2004). "X-linked inheritance of Fanconi anemia complementation group B." Nat Genet 36(11): 1219-1224.

Meetei, A. R., A. L. Medhurst, et al. (2005). "A human ortholog of archaeal DNA repair protein Hef is defective in Fanconi anemia complementation group M." Nat Genet 37(9): 958-963.

Meetei, A. R., S. Sechi, et al. (2003). "A multiprotein nuclear complex connects Fanconi anemia and Bloom syndrome." Mol Cell Biol 23(10): 3417-3426.

Meindl, A., H. Hellebrand, et al. (2010). "Germline mutations in breast and ovarian cancer pedigrees establish RAD51C as a human cancer susceptibility gene." Nat Genet 42(5): 410-414.

Meyer, S., H. Neitzel, et al. (2012). "Chromosomal aberrations associated with clonal evolution and leukemic transformation in fanconi anemia: clinical and biological implications." Anemia 2012: 349837.

Montenegro, G., E. Powell, et al. (2011). "Exome sequencing allows for rapid gene identification in a Charcot-Marie-Tooth family." Ann Neurol 69(3): 464-470.

Munoz, I. M., K. Hain, et al. (2009). "Coordination of structure-specific nucleases by human SLX4/BTBD12 is required for DNA repair." Mol Cell 35(1): 116-127.

Murai, J., K. Yang, et al. (2011). "The USP1/UAF1 complex promotes double-strand break repair through homologous recombination." Mol Cell Biol 31(12): 2462-2469.

Neitzel, H., J.-S. Kühl, et al. (2007). Clonal Chromosomal Aberrations in Bone Marrow Cells of Fanconi Anemia Patients: Results and Implications. Fanconi Anemia. A Paradigmatic Disease for the Understanding of Cancer and Aging. M. Schmid. Basel, Karger. 15: 79-94.

Neveling, K., R. Kalb, et al. (2007). Cancer in Fanconi Anemia and Fanconi Anemia Genes in

Cancer. Fanconi Anemia.A Paradigmatic Disease for the Understanding of Cancer and Aging. M. Schmid. Basel, Karger. 15: 59-78.

Ng, S. B., A. W. Bigham, et al. (2010). "Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome." Nat Genet 42(9): 790-793.

Ng, S. B., K. J. Buckingham, et al. (2010). "Exome sequencing identifies the cause of a mendelian disorder." Nat Genet 42(1): 30-35.

Niedernhofer, L. J., G. A. Garinis, et al. (2006). "A new progeroid syndrome reveals that genotoxic stress suppresses the somatotroph axis." Nature 444(7122): 1038-1043.

Osman, F. and M. C. Whitby (2007). "Exploring the roles of Mus81-Eme1/Mms4 at perturbed replication forks." DNA Repair (Amst) 6(7): 1004-1017.

Osorio, A., D. Endt, et al. (2012). "Predominance of pathogenic missense variants in the RAD51C gene occurring in breast and ovarian cancer families." Hum Mol Genet 21(13): 2889-2898.

Ozsolak, F. and P. M. Milos (2011). "RNA sequencing: advances, challenges and opportunities." Nat Rev Genet 12(2): 87-98.

Pang, Z., L. Yao, et al. (2011). "RAD51C germline mutations in Chinese women with familial breast cancer." Breast Cancer Res Treat 129(3): 1019-1020.

Pareek, C. S., R. Smoczynski, et al. (2011). "Sequencing technologies and genome sequencing." J Appl Genet 52(4): 413-435.

Patel, K. J. and H. Joenje (2007). "Fanconi anemia and DNA replication repair." DNA Repair (Amst) 6(7): 885-890.

Pelttari, L. M., T. Heikkinen, et al. (2011). "RAD51C is a susceptibility gene for ovarian cancer." Hum Mol Genet 20(16): 3278-3288.

Reid, S., D. Schindler, et al. (2007). "Biallelic mutations in PALB2 cause Fanconi anemia subtype FA-N and predispose to childhood cancer." Nat Genet 39(2): 162-164.

Romero, A., P. Perez-Segura, et al. (2011). "A HRM-based screening method detects RAD51C germ-line deleterious mutations in Spanish breast and ovarian cancer families." Breast Cancer Res Treat 129(3): 939-946.

Rosenberg, P. S., M. H. Greene, et al. (2003). "Cancer incidence in persons with Fanconi anemia." Blood 101(3): 822-826.

Rosenberg, P. S., H. Tamary, et al. (2011). "How high are carrier frequencies of rare recessive syndromes? Contemporary estimates for Fanconi Anemia in the United States and Israel." Am J Med Genet A 155A(8): 1877-1883.

Schindler, D., R. Friedl, et al. (2007). Applications of Cell Cycle testing in Fanconi Anemia. Fanconi Anemia - A Paradigmatic disease for the understanding of cancer and aging. Schindler D and H. H. Basel, Karger. 15: 110-130.

Schindler, D. and H. Hoehn (1988). "Fanconi anemia mutation causes cellular susceptibility to ambient oxygen." Am J Hum Genet 43(4): 429-435.

Schrader, K. A., A. Heravi-Moussavi, et al. (2011). "Using next-generation sequencing for the diagnosis of rare disorders: a family with retinitis pigmentosa and skeletal abnormalities." J Pathol 225(1): 12-18.

Schroeder, T. M., F. Anschutz, et al. (1964). "Spontaneous chromosome aberrations in familial panmyelopathy." Humangenetik 1(2): 194-196.

Schuster, B., K. Knies, et al. (2012, accepted for publication). "Whole Exome Sequencing reveals uncommon mutations in the recently identified Fanconi anemia gene SLX4/FANCP." Human Mutation

Schuster, B., K. Knies et al. (2012, in review). „Genotyping of Fanconi Anemia Patients by Whole Exome Sequencing: Advantages and Challenges." PlosOne.

Sengerova, B., A. T. Wang, et al. (2011). "Orchestrating the nucleases involved in DNA interstrand cross-link (ICL) repair." Cell Cycle 10(23): 3999-4008.

Shendure, J. A., G. J. Porreca, et al. (2008). "Overview of DNA sequencing strategies." Curr Protoc Mol Biol Chapter 7: Unit 7 1.

Sijbers, A. M., W. L. de Laat, et al. (1996). "Xeroderma pigmentosum group F caused by a defect in a structure-specific DNA repair endonuclease." Cell 86(5): 811-822.

Silvestri, V., P. Rizzolo, et al. (2011). "Mutation screening of RAD51C in male breast cancer patients." Breast Cancer Res 13(1): 404.

Smogorzewska, A., R. Desetty, et al. (2010). "A genetic screen identifies FAN1, a Fanconi anemia-associated nuclease necessary for DNA interstrand crosslink repair." Mol Cell 39(1): 36-47.

Stoepker, C., K. Hain, et al. (2011). "SLX4, a coordinator of structure-specific endonucleases, is mutated in a new Fanconi anemia subtype." Nat Genet 43(2): 138-141.

Strathdee, C. A., A. M. Duncan, et al. (1992). "Evidence for at least four Fanconi anaemia genes including FACC on chromosome 9." Nat Genet 1(3): 196-198.

Suhasini, A. N. and R. M. Brosh, Jr. (2012). "Fanconi anemia and Bloom's syndrome crosstalk through FANCJ-BLM helicase interaction." Trends Genet 28(1): 7-13.

Svendsen, J. M., A. Smogorzewska, et al. (2009). "Mammalian BTBD12/SLX4 assembles a Hol-

liday junction resolvase and is required for DNA repair." Cell 138(1): 63-77.

Takata, M., M. S. Sasaki, et al. (2001). "Chromosome instability and defective recombinational repair in knockout mutants of the five Rad51 paralogs." Mol Cell Biol 21(8): 2858-2866.

Thompson, E. R., S. E. Boyle, et al. (2012). "Analysis of RAD51C germline mutations in high-risk breast and ovarian cancer families and ovarian cancer patients." Hum Mutat 33(1): 95-99.

Timmers, C., T. Taniguchi, et al. (2001). "Positional cloning of a novel Fanconi anemia gene, FANCD2." Mol Cell 7(2): 241-248.

Tischkowitz, M. D. and S. V. Hodgson (2003). "Fanconi anaemia." J Med Genet 40(1): 1-10.

Trujillo, J. P., L. B. Mina, et al. (2012). "On the role of FAN1 in Fanconi anemia." Blood 120(1): 86-89.

Vaz, F., H. Hanenberg, et al. (2010). "Mutation of the RAD51C gene in a Fanconi anemia-like disorder." Nat Genet 42(5): 406-409.

Vuorela, M., K. Pylkas, et al. (2011). "Further evidence for the contribution of the RAD51C gene in hereditary breast and ovarian cancer susceptibility." Breast Cancer Res Treat 130(3): 1003-1010.

Wang, W. (2007). "Emergence of a DNA-damage response network consisting of Fanconi anaemia and BRCA proteins." Nat Rev Genet 8(10): 735-748.

Wolf, S. M., F. P. Lawrenz, et al. (2008). "Managing incidental findings in human subjects research: analysis and recommendations." J Law Med Ethics 36(2): 219-248, 211.

Wong, M. W., C. Nordfors, et al. (2011). "BRIP1, PALB2, and RAD51C mutation analysis reveals their relative importance as genetic susceptibility factors for breast cancer." Breast Cancer Res Treat 127(3): 853-859.

Wood, R. D. (2010). "Mammalian nucleotide excision repair proteins and interstrand crosslink repair." Environ Mol Mutagen 51(6): 520-526.

Xia, B., J. C. Dorsman, et al. (2007). "Fanconi anemia is associated with a defect in the BRCA2 partner PALB2." Nat Genet 39(2): 159-161.

Yan, Z., M. Delannoy, et al. (2010). "A histone-fold complex and FANCM form a conserved DNA-remodeling complex to maintain genome stability." Mol Cell 37(6): 865-878.

Yan, Z., R. Guo, et al. (2012). "A Ubiquitin-Binding Protein, FAAP20, Links RNF8-Mediated Ubiquitination to the Fanconi Anemia DNA Repair Network." Mol Cell 47(1): 61-75.

Zhang, F., Q. Fan, et al. (2009). "PALB2 functionally connects the breast cancer susceptibility proteins BRCA1 and BRCA2." Mol Cancer Res 7(7): 1110-1118.

Zheng, Y., J. Zhang, et al. (2010). "Screening RAD51C nucleotide alterations in patients with a family history of breast and ovarian cancer." Breast Cancer Res Treat 124(3): 857-861.

Zhou, W., E. A. Otto, et al. (2012). "FAN1 mutations cause karyomegalic interstitial nephritis, linking chronic kidney failure to defective DNA damage repair." Nat Genet 44(8): 910-915.

APPENDIX

FIGURE REFERENCES AND COPY RIGHTS

Figure 1: Guido Fanconi and his first describtion of Fanconi anemia (page 10)

Both figures have been published in: *Lobitz S. and Velleuer E. (2006). Guido Fanconi (1892-1979): a jack of all trades. Nat Rev Cancer 6(11): 893-898.*

(A) Permission for reuse in this dissertation was provided by Prof. Dr. Andreas Fanconi (Zurich, Switzerland), who holds the copy right.

(B) Permission for reuse in this dissertation was provided by Prof. Dr. Felix Sennhauser (Children´s Hospital Zurich, Switzerland), who holds the copy right.

Figure 2: Clinical features of Fanconi anemia (page 11)

(A) Kindly provided by the Shriners Hospital for Children, Philadelphia, USA. Previously published in: *Fanconi Anemia. Guidelines for Diagnosis and Managment. Third Edition, 2008. Fanconi Anemia Research Fund, Inc. Editors: Eiler M.E., Frohnmeyer D., et al. (Eugene, USA): 108*

(B and C) Reprinted from *Alter B.P., Young N.S., The bone marrow failure syndromes. In: Nathan DG, Oski FA, eds. Hematology of Infancy and Childhood, 4th ed. Philadelphia, PA: WB Saunders, Inc, 1993: 216-316*, with permission of Elsevier.

(D) Reprinted from *Konoplev S. and Bueso-Ramos C.E. (2006). Advances in the pathologic diagnosis and biology of acute myeloid leukemia. Ann Diagn Pathol10(1):39-65. Review*, with permission from Elsevier.

(E) Permission for use in this dissertation was provided by Eunike Velleuer (Medical School, University of Duesseldorf, Germany), who holds the copy right.

Figure 3: Cellular phenotype of Fanconi anemia (page 12)

(A) Reprinted from *Reid S., Schindler D., et al. (2007). Biallelic mutations in PALB2 cause Fanconi anemia subtype FA-N and predispose to childhood cancer. Nat Genet. 2007 Feb;39(2):162-4*, with per-

mission from Macmillan Publishers Ltd, Nature Pulishing Group.

(B) Reprinted from own, enclosed publication: *Vaz F., Hanenberg H., et al. (2010). Mutation of the RAD51C gene in a Fanconi anemia-like disorder. Nat Genet 42(5): 406-409.*

PAPERS AND MANUSCRIPTS

The reuse of published articles for the (co-)authors dissertation is in accordance with the journals' copy right agreements.

INDEX OF ABBREVIATIONS

GENE SYMBOLS

ATR	Ataxia telangiectasia and Rad3 related
BLM	Bloom syndrome, RecQ helicase-like
BRCA1/2	Breast cancer susceptibilty gene 1/2
BRIP1	BRCA1 interacting protein C-terminal helicase 1
CDC27	Cell division cycle 27 homolog (S. cerevisiae)
CHK1	Checkpoint kinase 1
EME1	Essential meiotic endonuclease 1 homolog (S. cerevisiae)
ERCC1	Excision repair cross-complementing rodent repair deficiency, complementation group 1
FAN1	Fanconi anemia associated nuclease 1
FANC	Fanconi anemia gene
MHF1/2	FANCM-interacting histone fold protein 1/2
MLH1/3	mutL homolog 1/3, colon cancer, nonpolyposis type 2
MSH2/3	mutS homolog 2/3, colon cancer, nonpolyposis type 1
MUC	Mucin gene, oligomeric mucus/gel-forming
MUS81	MUS81 endonuclease homolog (S. cerevisiae)
NBPF	Neuroblastoma breakpoint family
PALB2	Partner and localizer of BRCA2
PMS2	postmeiotic segregation increased 2
POLN	polymerase (DNA directed) nu
RAD51	RAD51 homolog (S. cerevisiae), RecA-like protein
RAD51C	RAD51 homolog C (S. cerevisiae), RAD51L2
REV1	REV1, polymerase (DNA directed)
SLX4	SLX4 structure-specific endonuclease subunit homolog (S. cerevisiae)
UAF1	USP1 associated factor 1
USP1	Ubiquitin specific peptidase 1
XPF	Xeroderma pigmentosum, complementation group F

GENERAL ABBREVIATIONS

A	Adenine
C	Cytosine

G	Guanine
T	Tymine
AML	Acute myeloid leukemia
BMF	Bone marrow failure
BRAFT	BLM, RPA, FA, and Topo III alpha containing
BRCA	Breast cancer
cDNA	complementary DNA
DEB	Diepoxybutane
DNA	Desoxyribonucleic acid
DSB	Double strand break
FA	Fanconi anemia
FAAP	Fanconi anemia associated protein
HR	Homologous recombination
ICL	Interstrand crosslink
ID	FANCI and FANCD2 containing
MDS	Myelodysplastic syndrome
MMC	Mitomycin C
MMR	Mismatch repair
NCBI	National Center for Biotechnology Information
NER	Nucleotide excision repair
NGS	Next Generation Sequencing
RNA	Ribonucleic acid
SCC	Squamous cell carcinoma
SNP	Single nucleotide polymorphism
TLS	Translesion synthesis
VATER/VACTERL	vertebral defects, anal atresia, cardiac defects, tracheo-esophageal fistula, renal and radial anomalies, and limb abnormalities
WES	Whole Exome Sequencing
WGS	Whole Genome Sequencing
XFE	ERCC4 associated progeroid syndrome, segmental progeria
XP	Xeroderma pigmentosum

Personal contribution to the following publication

A Histon-Fold Complex and FANCM Form a conserved DNA-Remodeling Complex to Maintain Genome Stability

Yan Z, Delannoy M, Ling C, Daee D, Osman F, Muniandy PA, Shen X, Oostra AB, Du H, Steltenpool J, Lin T, Schuster B, Décaillet C, Stasiak A, Stasiak AZ, Stone S, Hoatlin ME, Schindler D, Woodcock CL, Joenje H, Sen R, de Winter JP, Li L, Seidman MM, Whitby MC, Myung K, Constantinou A, Wang W.

Molecular Cell 2010

Beatrice Schuster was involved in this publication by screening unassigned FA patients for deficient MHF1 or MHF2 protein expression. She was also involved in proofreading of the manuscript.

D. Schinder

J.P. de Winter

A.B. Oostra

Personal contribution to the following publication

A Histon-Fold Complex and FANCM Form a conserved DNA-Remodeling Complex to Maintain Genome Stability

Yan Z, Delannoy M, Ling C, Daee D, Osman F, Muniandy PA, Shen X, Oostra AB, Du H, Steltenpool J, Lin T, Schuster B, Décaillet C, Stasiak A, Stasiak AZ, Stone S, Hoatlin ME, Schindler D, Woodcock CL, Joenje H, Sen R, de Winter JP, Li L, Seidman MM, Whitby MC, Myung K, Constantinou A, Wang W.

Molecular Cell 2010

Beatrice Schuster was involved in this publication by screening unassigned FA patients for deficient MHF1 or MHF2 protein expression. She was also involved in proofreading of the manuscript.

D. Schindler

(Weidong Wang).

Personal contribution to the following publication

On the role of FAN1 in Fanconi anemia

Juan P. Trujillo, Leonardo B. Mina, Roser Pujol, Massimo Bogliolo, Joris Andrieux, Muriel Holder, Beatrice Schuster, Detlev Schindler, and Jordi Surrallés

Blood 2012

Beatrice Schuster was involved in this publication by screening unassigned FA patients for FAN1 protein deficiency and proofreading of the manuscript.

Personal contribution to the following publication

Mutation of the RAD51C gene in a Fanconi anemia-like disorder

Vaz F, Hanenberg H, Schuster B, Barker K, Wiek C, Erven V, Neveling K, Endt D, Kesterton I, Autore F, Fraternali F, Freund M, Hartmann L, Grimwade D, Roberts RG, Schaal H, Mohammed S, Rahman N, Schindler D, Mathew CG.

Nature Genetics 2010

Beatrice Schuster was involved in this publication by the exclusion of known FA complementation groups in the index patient. She also confirmed the described mutation and analyzed its effect on protein expression and stability. Finally, she did mutation screening in other unassigned FA patients and was involved in proofreading of the manuscript.

DR DETLEV SCHINDLER

DR FLAVIA AUTORE

PROF FRANCA FRATERNALI

PROF CHRISTOPHER MATHEW

Personal contribution to the following publication

Mutation of the RAD51C gene in a Fanconi anemia-like disorder

Vaz F, Hanenberg H, Schuster B, Barker K, Wiek C, Erven V, Neveling K, Endt D, Kesterton I, Autore F, Fraternali F, Freund M, Hartmann L, Grimwade D, Roberts RG, Schaal H, Mohammed S, Rahman N, Schindler D, Mathew CG.

Nature Genetics 2010

Beatrice Schuster was involved in this publication by the exclusion of known FA complementation groups in the index patient. She also confirmed the described mutation and analyzed its effect on protein expression and stability. Finally, she did mutation screening in other unassigned FA patients and was involved in proofreading of the manuscript.

Personal contribution to the following publication

SLX4, a coordinator of structure-specific endonucleases, is mutated in a new Fanconi anemia subtype.

Stoepker C, Hain K, Schuster B, Hilhorst-Hofstee Y, Rooimans MA, Steltenpool J, Oostra AB, Eirich K, Korthof ET, Nieuwint AW, Jaspers NG, Bettecken T, Joenje H, Schindler D, Rouse J, de Winter JP.

Beatrice Schuster was involved in this publication by mutation screening in a group of unassigned FA patients. She performed linkage analysis, exclusion of known FA complementation groups, mutation detection and confirmation of Mendelian segregation in one family. She was also involved in proofreading of the manuscript.

D. Schinder

K. Eirich

Rouse Hain

J.P. de Winter

C.Stoepker

A.B Oostra

M.A.Rooimans

Personal contribution to the following publication

XPF mutations severely disrupting DNA interstrand cross-link repair cause Fanconi anemia

Massimo Bogliolo, Beatrice Schuster, Chantal Stoepker, Burak Derkunt, Yan Su, Anja Raams, Juan P. Trujillo, Jordi Minguillón, Maria J. Ramirez, Roser Pujol, José A. Casado, Rocío Banos, Paula Rio, Kerstin Knies, Sheila Zúniga, Javier Benítez, Juan A. Bueren, Nicolaas G.J. Jaspers, Orlando D. Schärer, Johan P. de Winter, Detlev Schindler and Jordi Surrallés

submitted 2012

Beatrice Schuster was involved in this publication by screening the cohort of German patients for XPF mutations. She identified and characterized the disease causing mutations in one of the described patients and therefore performed experiments on DNA and protein level. She provided data and figures to the manuscript and was involved in the proofreading process.
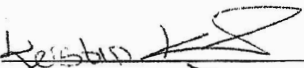
Personal contribution to the following publication

Whole Exome Sequencing reveals uncommon mutations in the recently identified Fanconi anemia gene *SLX4/FANCP*

Beatrice Schuster, Kerstin Knies, Chantal Stoepker, Eunike Velleuer, Richard Friedl, Birgit Gottwald-Mühlhauser, Johan P de Winter, Detlev Schindler

Beatrice Schuster was involved in this publication by identifying and characterizing the disease causing mutations. She designed and performed necessary experiments. Finally she wrote the manuscript and took responsibility as corresponding author.

Personal contribution to the following publication

Genotyping of Fanconi Anemia Patients by Whole Exome Sequencing: Advantages and Challenges

Kerstin Knies, Beatrice Schuster, Najim Ameziane, Martin Rooimans, Thomas Bettecken, Johan de Winter, Detlev Schindler

in review 2012 Plos One

Beatrice Schuster was involved in this publication by study design and establishing a FA-specific whole exome sequencing analysis pipeline. She performed bioinformatic analysis and performed validating experiments. Finally she wrote the manuscript and took responsibility as corresponding author.

OWN PUBLICATIONS

2010

Yan Z., Delannoy M., Ling C., Daee D., Osman F., Muniandy P.A., Shen X., Oostra A.B., Du H., Steltenpool J., Lin T., **Schuster B.**, Décaillet C., Stasiak A., Stasiak A.Z., Stone S., Hoatlin M.E., Schindler D., Woodcock C.L., Joenje H., Sen R., de Winter J.P., Li L., Seidman M.M., Withby M.C., Myung K., Constantinou A., Wang W.: **A histone-fold complex and FANCM form a conserved DNA-remodeling complex to maintain genome stability.** 2010. Molecular cell 37, 865-878.

Vaz F., Hanenberg H., **Schuster B**., Barker K., Wiek C., Erven V., Neveling K., Endt D., Kesterton I., Autore F., Fraternali F., Freund M., Hartmann L., Grimwade D., Roberts R.G., Schaal H., Mohammed S., Rahman N., Schindler D., Mathew C.G.: **Mutation of the RAD51C gene in a Fanconi anemia-like disorder**. 2010. Nature genetics 42, 406-409.

2011

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., Jaspers N.G.J., Bettecken T., Joenje H., Schindler D., Rouse J., de Winter J.P.: **SLX4, a coordinator of structure-specific endonucleases, is mutated in a new Fanconi anemia subtype**. 2011. Nature genetics 43, 138-141.

Kopic S., Eirich K., **Schuster B**., Hanenberg H., Varon-Mateeva R., Rittinger O., Schimpl G., Schindler D., and Jones N.: **Hepatoblastoma in a 4-year-old girl with Fanconi anaemia**. 2011. Acta Paediatr 100, 780-783.

2012

Trujillo J.P., Mina L.B., Pujol R., Bogliolo M., Andrieux J., Holder M., **Schuster B.**, Schindler D., and Surralles J.: **On the role of FAN1 in Fanconi anemia**. 2012. Blood 120, 86-89.

Schuster B.*, Knies K.*, Stoepker C., Velleuer E., Friedl R., Gottwald-Mühlhauser B., de Winter J.P., and Schindler D.: **Whole Exome Sequencing reveals uncommon mutations in the recently identified Fanconi anemia gene SLX4/FANCP**. 2012 (accepted for publication). Human Mutation. *contributed equally*

Schuster, B.*, Knies, K.*, Ameziane, N., Rooimans, M.A., Bettecken, T., De Winter, J.P., and Schindler, D.: **Genotyping of Fanconi Anemia Patients by Whole Exome Sequencing: Advantages and Challenges**. 2012 (in review, PlosOne) **contributed equally*

Bogliolo M.*, **Schuster B.***, Stoepker C., Derkunt B., Su Y., Raams A., Trujillo J.P., Minguillon J., Ramirez M.J., Pujol R., Casado J.A., Banos R., Rio P., Knies K., Zuniga S., Benitez J., Bueren J.A., Jaspers N.G.J., Schärer O.D., de Winter J.P., Schindler D., Surralles J.: **XPF mutations severely disrupting DNA interstrand crosslink repair cause Fanconi anemia**. 2012 (submitted) **contributed equally*

Contributions to Conferences and Symposia

First Authorships only

A novel Fanconi anemia candidate gene: EMSY

B. Schuster, L. Hughes-Davies, K. Neveling, E. Rossato, D. Schindler

Poster presentation, GFH (Gesellschaft für Humangenetik/German Society for Human Genetics) Conference, Aachen, Germany, 2009

FANCI revisted

B.Schuster, K.Neveling, H. Hoehn, A. Smogorzewska, H. Hanenberg, D. Schindler

Poster presentation, FARF (Fanconi Anemia Research Fund) Scientific Symposium, Baltimore, USA, 2009

Fanconi anemia patients of type I depend on hypomorphic mutations in FANCI

B. Schuster1, K. Neveling1, H. Hoehn1, A. Smogorzewska2, H. Hanenberg3, D. Schindler1

Poster presentation, GFH Conference, Hamburg, Germany, 2010

FANCP – the 15th Fanconi Anemia Gene

B. Schuster, C. Stoepker, K. Hain, Y. Hilhorst-Hofstee, M.A. Rooimans, J. Steltenpool, A.B. Oostra, K. Eirich, E.T. Korthof, A.W.M. Nieuwint, N.G.J. Jaspers, T. Bettecken, H. Joenje, D. Schindler, J. Rouse and J.P. de Winter

Oral presentation, GFH Conference, Regensburg, Germany, 2011

Pseudogenes: an unsolvable problem in Whole Exome Sequencing?

B. Schuster, K. Knies, D. Schindler

Poster presentation, ESHG (European Society of Human Genetics) Conference, Nürnberg, Germany, 2012

Genotyping Fanconi Anemia Patients by Whole Exome Sequencing

B. Schuster

Oral presenation, Next Generation Sequencing Symposium, Medical School MHH Hannover, Germany, 2012

Danksagung

Am Ende dieser Arbeit geht mein erster Dank an den jetzigen Institutsleiter Prof. Dr. Thomas Haaf und den ehemaligen Vorstand Prof. Dr. Holger Höhn für die Möglichkeit meine Promotion am Institut für Humangenetik der Universität Würzburg durchzuführen.

Prof. Dr. Schindler danke ich für die Vergabe des äußerst interessanten und ergiebigen Themas, die Möglichkeit in seinem Labor selbsständig zu arbeiten und selbsverständlich für die Erstellung des ersten Gutachtens.

Für die Erstellung des Zweitgutachtens bedanke ich mich vielmals bei Prof. Dr. Charlotte Förster, Leiterin des Lehrstuhls für Neurobiologie und Genetik.

Außerdem möchte ich an dieser Stelle Prof. Dr. Traute Schröder-Kurth Danke sagen. Sie sorgte nicht nur für stetigen Nachschub an Nervennahrung und aufmunternden Worten, sondern ließ mich auch an ihrem reichen Erfahrungsschatz teilhaben. Außerdem wurde ein Teil meiner Studien mit Hilfe des „Schröder-Kurth Fonds" finanziert.

Gegen Ende meiner Promotion wurde ich in das Mentoring Programm für Naturwissenschaftlerinnen aufgenommen. Die Leiterin dieses Programms, Dr. Ljubica Lozo, und mein Mentor Prof. Dr. Manfred Kubbies sind für mich seitdem eine wichtige Stütze außerhalb der Wissenschaft.

Ein dickes Dankeschön geht selbstverständlich an meine Mitstreiter und die viel zahlreicheren Mitstreiterinnen in der Arbeitsgruppe. Ohne diese gute Gemeinschaft hätte sich wohl so manche Frustrationsphase um einiges länger hingezogen. Es war schön sich zusammen zu ärgern und natürlich noch viel schöner sich zusammen zu freuen.

Meiner Familie und meinem Freund Florian gilt der größte Dank. Jahrelang haben sie mich durch die Höhen und Tiefen der Promotion begleitet, geduldig mein Jammern ertragen und sich gemeinsam mit mir über meine Erfolge gefreut.

„Danke, dass ich immer auf euch zählen konnte!"