

(19) World Intellectual Property Organization
International Bureau(43) International Publication Date
12 January 2012 (12.01.2012)(10) International Publication Number
WO 2012/004559 AI

(51) International Patent Classification:

C12Q 1/44 (2006.0 1) **C12N 9/00** (2006.0 1)
G01N 33/566 (2006.0 1) **C12Q 1/68** (2006.0 1)
G01N 33/5 74 (2006.01)

(21) International Application Number:

PCT/GB201 1/001017

(22) International Filing Date:

6 July 2011 (06.07.2011)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

101 1367.8 6 July 2010 (06.07.2010) GB

(71) Applicant (for all designated States except US): **MEDICAL RESEARCH COUNCIL** [GB/GB]; 2nd Floor
 David Phillips Building, Polaris House, North Star Avenue, Swindon, Wiltshire SN2 1FL (GB).

(72) Inventor; and

(75) Inventor/Applicant (for US only): **ROUSE, John** [IE/GB]; MRC Protein Phosphorylation Unit, University of Dundee College of Life Sciences, The Sir James Black Centre, Dow Street, Dundee DD1 5EH (GB).

(74) Agent: **PILKINGTON, Stephanie, Joan**; Potter Clarkson LLP, Park View House, 58 The Ropewalk, Nottingham NG1 5DD (GB).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM,

AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:

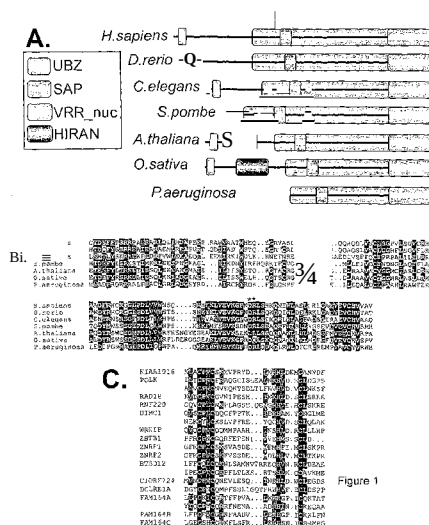
— of inventorship (Rule 4.17(iv))

Published:

— with international search report (Art. 21(3))

— before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))

(54) Title: METHODS FOR SCREENING FOR COMPOUNDS EXPECTED TO BE USEFUL IN THE MODULATING, FOR EXAMPLE INHIBITING, THE ACTIVITY OF KIAA1018/MTMR15/FAN1



(57) Abstract: A method for identifying a compound expected to be useful in modulating, for example inhibiting, the repair of a DNA inter-strand crosslink (ICL), resolution of an ICL-induced double strand break, or homologous recombination in a cell, the method comprising the steps of (1) determining whether a test compound modulates, for example inhibits, the endonuclease or 5'-exonuclease activity of a FAN1 polypeptide on a substrate polynucleotide and (2) selecting a compound which modulates, for example inhibits, the said FAN1 polypeptide endonuclease or 5'-exonuclease activity. Such a compound may be useful in enhancing the effect of platinum based chemotherapy or other ICL-inducing therapy on cancer cells.

**METHODS FOR SCREENING FOR COMPOUNDS EXPECTED TO BE USEFUL IN THE
MODULATING, FOR EXAMPLE INHIBITING, THE ACTIVITY OF
KIAA1018/MTMR15/FAN1**

The present invention relates to methods for screening for compounds expected to be useful in modulating, for example inhibiting, the activity of a polypeptide considered to be involved in repairing DNA inter-strand crosslinks.

DNA inter-strand crosslinks (ICLs) are formed when bifunctional agents covalently link the two strands in a double helix. ICLs are toxic lesions that prevent strand separation necessary for transcription and DNA replication. ICLs can be induced by drugs and also by endogenous metabolites. Cross-linking agents such as mitomycin-C (MMC) and cisplatin generate a mixture of mono-adducts and ICLs in cells but cellular toxicity correlates with the number of ICLs. Although ICLs can be repaired in G₁, the major route for ICL repair appears to occur in S-phase (Akkari et al., 2000; Rothfuss and Grompe, 2004; Taniguchi et al., 2002). Various models for the repair of ICLs have been suggested (McCabe et al., 2009; Moldovan and D'andrea, 2009), and recent studies proposed that ICL repair requires two forks to converge on the ICL (Raschle et al., 2008) (Fig. 8)

Forks that stall at ICLs recruit signaling complexes including the Fanconi Anaemia (FA) proteins and FA-associated proteins (Moldovan and D'andrea, 2009) (Fig. 8). Fanconi Anaemia is an inherited recessive condition characterized by developmental defects, skeletal abnormalities, bone marrow failure and cancer predisposition (Wang, 2007). FA falls into thirteen complementation groups and the relevant FA genes have been cloned (Patel and Joeje, 2007; Wang, 2007). Nevertheless FA patients exist where mutations in known FA genes could not be found. The central components of the FA pathway are FANCD2 and its paralogue FANCI that together form the "ID" complex (Garcia-Higuera et al., 2001; Smogorzewska et al., 2007). These two proteins are mono-ubiquitinated at Lys561 and Lys523, respectively, in S-phase and in response to ICLs (Fig. 8) (Garcia-Higuera et al., 2001; Taniguchi et al., 2002). This reaction is catalysed by the E3 ubiquitin ligase FANCL subunit of the FA core complex, that comprises FANCA, B, C, E, F, G, L, and M, and also requires the FA-associated proteins FAAP100 and FAAP24 (Ciccia et al., 2007; Collis et al., 2008; Ling et al., 2007). Furthermore, loss of FANCD2 mono-ubiquitination is observed in many FA patients (Moldovan and D'andrea, 2009).

Mono-ubiquitination of FANCD2 is necessary for ICL repair but the underlying molecular mechanisms are unclear. The mono-ubiquitinated form of the ID complex may recruit ICL repair proteins but, as yet, no ligands for ubiquitinated FANCD2 or FANCI have been reported. It was reported that mono-ubiquitination of FANCD2 is required for the

"unhooking" of the ICL in a cell-free repair system (Knipscheer et al., 2009) (Fig. 8). Unhooking involves incisions on either side of the ICL, one of which is catalysed by the structure-specific nuclease MUS81-EME1 (Fig. 8) (Hanada et al., 2007; Hanada et al., 2006). MUS81-EME1 creates a one-ended double-strand break (DSB) that can be used later to initiate homologous recombination (HR). The identity of the nuclease that catalyses the second incision to enable unhooking of the ICL is unclear. XPF-ERCC1 has been implicated but this is controversial (Bergstralh and Sekelsky, 2008; Bhagwat et al., 2009). After unhooking the resulting gap is filled in by translesion synthesis, that also appears to require FANCD2 ubiquitination (Knipscheer et al., 2009), and the unhooked lesion is removed by excision repair. The DSBs generated by unhooking are resected and one of them initiates HR to complete ICL repair (Fig. 8). Successful HR-mediated repair of the MUS81-generated DSB depends on processing of DNA repair intermediates by the SLX4 complex. SLX4 acts as a scaffold for XPF-ERCC1, MUS81-EME1 and SLX1. Cells lacking, or depleted of, SLX4 (Fekairi et al., 2009; Munoz et al., 2009; Svendsen et al., 2009) or XPF-ERCC1 (Niedernhofer et al., 2004) cannot repair the DSBs created by MUS81 after ICL induction, and exhibit defects in HR-mediated repair of DSBs. In this study we report the identification of FAN1 (FA-associated nuclease 1), also referred to herein as REND1, a novel nuclease recruited to sites of DNA damage by mono-ubiquitinated FANCD2 that is important for repair of ICLs.

The listing or discussion of a prior-published document in this specification should not necessarily be taken as an acknowledgement that the document is part of the state of the art or is common general knowledge.

A first aspect of the invention provides a method for identifying a compound expected to be useful in modulating, for example inhibiting, KIAA1018/MTMR15/FAN1 (FAN1) endonuclease activity, the method comprising the steps of (1) determining whether a test compound modulates, for example inhibits, the endonuclease activity of a FAN1 polypeptide on a branched polynucleotide substrate and (2) selecting a compound which modulates, for example inhibits, the said FAN1 endonuclease activity.

In an embodiment of the preceding aspect, the branched polynucleotide substrate may be, for example, a substrate 5' flap-containing nucleic acid or a replication fork nucleic acid.

A second aspect of the invention provides a method for identifying a compound expected to be useful in modulating, for example inhibiting, FAN1 5'-exonuclease activity, the

method comprising the steps of (1) determining whether a test compound modulates, for example inhibits, the 5'-exonuclease activity of a FAN1 polypeptide on a substrate polynucleotide and (2) selecting a compound which modulates, for example inhibits, the said FAN1 5'-exonuclease activity.

5

A further aspect of the invention provides a method for identifying a compound expected to be useful in modulating, for example inhibiting, the repair of a DNA inter-strand crosslink (ICL), resolution of an ICL-induced double strand break, or homologous recombination in a cell, the method comprising the steps of (1) determining whether a
10 test compound modulates, for example inhibits, the endonuclease or 5'-exonuclease activity of a FAN1 polypeptide on a substrate polynucleotide and (2) selecting a compound which modulates, for example inhibits, the said FAN1 polypeptide endonuclease or 5'-exonuclease activity.

15

A further aspect of the invention provides a method for identifying a compound expected to be useful in modulating, for example inhibiting, the repair of a DNA inter-strand crosslink, resolution of an ICL-induced double strand break, or homologous recombination in a cell, the method comprising the steps of (1) determining whether a
20 test compound modulates, for example inhibits, the interaction between a FAN1 polypeptide and an ubiquitinated FANCD2 or FANCI polypeptide and (2) selecting a compound which modulates, for example inhibits, the said interaction between the FAN1 polypeptide and the ubiquitinated FANCD2 or FANCI polypeptide.

25

A further aspect of the invention provides a method for identifying a compound expected to be useful in increasing cancer cells' sensitivity to ICLs or enhancing the effect of platinum based chemotherapy or other ICL-inducing therapy (for example mitomycin C, nitrogen mustards, nitrosoureas, alkylalkanesulphonate, and Cyclophosphamide, part of the FEC (fluorouracil/etoposide/cyclophosphamide) chemotherapy regimen) on cancer
30 cells, the method comprising the steps of (1) determining whether a test compound modulates, for example inhibits, the endonuclease or 5'-exonuclease activity of a FAN1 polypeptide on a substrate polynucleotide; or the interaction between a FAN1 polypeptide and an ubiquitinated FANCD2 or FANCI polypeptide, and (2) selecting a compound which modulates, for example inhibits, the endonuclease or 5'-exonuclease activity of the FAN1 polypeptide on the substrate polynucleotide; or the interaction
35 between the FAN1 polypeptide and the ubiquitinated FANCD2 or FANCI polypeptide.

The method of the immediately preceding aspect may be particularly useful in increasing cancer cells' sensitivity to ICLs or enhancing the effect of platinum based chemotherapy or other ICL-inducing therapy on cancer cells when the cancer cells are those that have intrinsic or acquired resistance to platinum compounds, as would be understood by a person skilled in the art. Examples of cancer cells that may be considered to exhibit intrinsic resistance to platinum compounds include colorectal cancer cells, prostate cancer cells, lung cancer cells and breast cancer cells. An example of cancer cells that may be considered to exhibit acquired resistance to platinum compounds includes ovarian cancer cells.

It is envisaged that a compound that may be identified by the methods of the invention may be useful in inhibiting lymphocyte, for example T-cell, activity. Such activity may be inhibited, for example, via inhibition of somatic recombination in T-cells. Thus such compound may be useful as an immunosuppressant.

A further aspect of the invention provides a method for identifying a compound expected to be useful in modulating, for example inhibiting, the repair of a DNA inter-strand crosslink, resolution of an ICL-induced double strand break, or homologous recombination in a cell, or in increasing cancer cells' sensitivity to ICLs or enhancing the effect of platinum based chemotherapy or other ICL-inducing therapy on cancer cells, the method comprising the steps of (1) determining whether a test compound modulates, for example inhibits, said repair of a DNA inter-strand crosslink, resolution of an ICL-induced double strand break, or homologous recombination in the cell, or the cell's resistance to ICLs or the effect of platinum based chemotherapy or other ICL-inducing therapy on the cell, wherein said determining is done on at least two cells, wherein said cells differ in the amount or activity of FAN1, and (2) selecting a compound which modulates, for example inhibits, the said repair of a DNA inter-strand crosslink, resolution of an ICL-induced double strand break, or homologous recombination in the cell, or the cell's resistance to ICLs or the effect of platinum based chemotherapy or other ICL-inducing therapy to a greater extent in the cell having more FAN1 than in the cell having less FAN1.

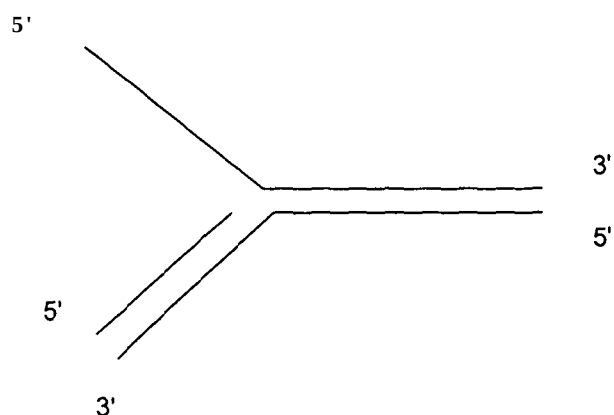
It is envisaged that the methods of the preceding aspect may be useful in screening compounds for their specificity for, for example, FAN1, and investigating their mode of action. For example, altering the expression levels of FAN1 then testing the compound activity in test cells displaying differing levels of FAN1 expression, may provide

information on potential off-target effects of test compounds, i.e. identify whether the test compound interacts with targets in the cell other than FAN1.

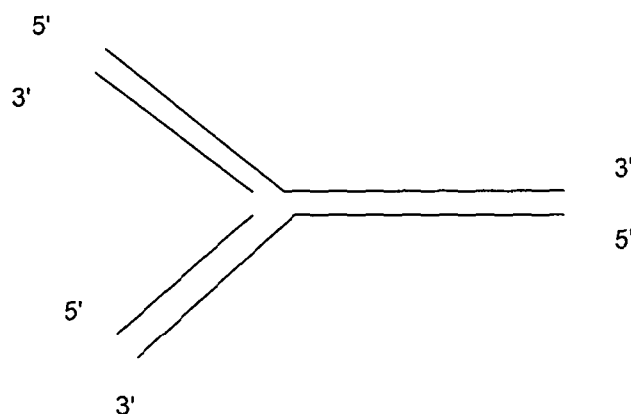
5 The endonuclease or 5'-exonuclease activity of the FAN1 polypeptide that is modulated/assessed in the screening method is endonucleolytic or 5'-exonuclease cleavage of a substrate polynucleotide as defined above. Cleavage of the substrate polynucleotide may be assessed by techniques including those discussed further below and in the Examples. For example, fluorescent or radioisotopic labelling of the polynucleotide may be used in assessing cleavage of the nucleic acid, as well known to
10 those skilled in the art. Further methods will be apparent to the skilled person on the basis of this teaching and the many known methods of assessing nucleic acid cleavage.

In an embodiment of any aspect of the invention, the substrate polynucleotide may be a branched polynucleotide. It is particularly preferred that the branched polynucleotide is a
15 5' flap-containing nucleic acid or a replication fork nucleic acid. It will be appreciated that the terms "polynucleotide" and "nucleic acid" may be used interchangeably herein.

A typical substrate 5' flap-containing nucleic acid is shown in Figure 2 A and below. The 5' flap-containing nucleic acid typically has a single-stranded portion of typically 5 or
20 more nucleotides but may be, for example, between 1 and 100 nucleotides. Preferably the single-stranded portion may be between 3 and 50 nucleotides, more preferably between 5 and 10 nucleotides. The double stranded portions of the molecule independently typically have at least 5 base pairs, for example between 5 and 100 base pairs. Preferably, the double stranded portion may be between 10 and 50 base pairs,
25 more preferably 20 base pairs.



A typical replication fork nucleic acid is shown in Figure 2A and below. Each of the three double stranded portions of the molecule independently typically has at least 5 base pairs, for example between 10 and 50 base pairs, such as 20 base pairs.



5

It is considered that the sequence of the 5' flap-containing nucleic acid or the replication fork nucleic acid is not critical. It is envisaged that the nucleic acid will be deoxyribose nucleic acid. The nucleic acids may include naturally occurring or wholly or partially synthetic DNA, e.g., cDNA and genomic DNA. The nucleic acids may be labelled, for example, with radioactive isotopes, such as ^{32}P , ^3H , and ^{14}C . Alternatively, the nucleic

acids may be labelled with non-radioactive probes, for example chemiluminescent or fluorescent probes, as would be understood in the art. The substrate nucleic acids may be labelled with a tag to enable purification, fixing to a solid support and/or other manipulation to be carried out, as would be understood in the art. For example, the nucleic acid may be labelled with biotin to allow binding to an appropriate support labelled with an avidin compound, such as streptavidin.

The substrate nucleotides are produced by annealing oligonucleotides using standard techniques known in the art.

10

The endonuclease or exonuclease activity may be increased or reduced by an alteration in the V_{max} or the K_m (or both) of the FAN1 polypeptide for a particular substrate. For example, activity may be increased by an increased V_{max} or decreased K_m . It will be appreciated that it may not be necessary to determine the value of either V_{max} or K_m in order to determine whether the FAN1 polypeptide has been activated or deactivated.

15

Activity may be measured as the amount of a substrate cleaved in a given time; a change of activity may therefore be detected as a change in the amount of substrate (for example, at a single concentration) that is cleaved in a given time. It is preferred that the activity is increased or decreased, as appropriate, by at least 2, preferably 5, 10, 15, 20, 25, 30 or 50-fold.

20

Typically the FAN1 polypeptide is wild type human FAN1 polypeptide or a fragment thereof, or a fusion either thereof. Typically the fragment comprises at least the VPR-nuc domain. Nevertheless, the VPR-nuc domain is not considered to be required for interaction with FANCD2 or FANCI. The VRR-nuc domain is located at amino acid 871-1009 of human FAN1 (the sequence of which is provided below) or residues equivalent thereto. In some embodiments, the UBZ domain may be included in the fragment, but this is not required for activity of FAN1. The UBZ domain is located at amino acids 41-69 of human FAN1 or residues equivalent thereto. For methods assaying FAN1 binding to FANCD2 or FANCI, the UBZ domain is considered to be required. In some embodiments, the fragment may also comprise the SAP domain, which is located at amino acid numbers 461-504 of human FAN1 or residues equivalent thereto.

25

30

Thus, for nuclease activity assays, it is envisaged that the fragment of FAN1 will comprise at least the VRR-nuc and generally the SAP domain. The fragment may thus correspond to at least residues 460-1017 of human FAN1, or residues equivalent

35

thereto. For FANCD2 or FANCI binding assays the fragment may comprise at least residues 41-69 of human FAN1. Full length human FAN1 may be used in all aspects of the invention, or fragments thereof containing the relevant domains according to the requirements of the assay.

5

Plasmid constructs that may be used to express FAN1 are well known to those skilled in the art. For example, FAN1 may be expressed using a bacterial expression plasmid with a NUS (N-utilization substance) and a hexahistidine tag. For example, FAN1 may be expressed using the pET43.1A bacterial expression plasmid. For expression in Sf21

10

insect cells, FAN1 may be cloned into, for example, pFASTBAC.

It is envisaged that co-expression of FAN1 with MLH1 (MutL homolog 1), a mismatch repair protein FAN1 interacts with, may improve stability of the recombinant FAN1 and is therefore also included in the methods of the invention.

15

The human FAN1 polypeptide and polynucleotide sequence is shown below and is shown at GenBank NM_014967 for nucleic acid sequence and UniProt Q9Y2M0 for amino acid sequence. The skilled person will readily be able to identify other FAN1 polypeptide sequences from databases. For example, the Homologene feature of the

20

The nucleotide sequence of human FAN1 (GenBank ref: NM_014967) (SEQ ID NO: 1) may be found below:

```

1  ggcgcgcgga gccaggtggg aggtgcgagg tggccgcgcg gggatcttgg gtgacagggc
25  61  accgagggaa ggaggacgcg agggcagcca ggcctaggg agcagggaga gtggctcggg
121  ctgagtcgcg tggccccagg aagaagaaat tgtcgagacg aataacatga ggtcatatag
181  aatcccactt ttggtgattt caagtcaaga aagtaaaagt aaaccattgc tatctttcac
241  cttaaatatc ctgtgtttta ttgctcagaa catccagttt ttctaatact catgatgtca
301  gaaggggaaac ctcctgacaa aaaaaggcct cgtagaagct tatcaatcag caagaataag
30  361  aaaaagcat ctaattctat tatttcgtgt ttaacaatg caccacctgc taaacttgcc
421  tgccccgttt gcagtaaaat ggtgcctaga tatgacttaa accggcacct tgatgaaatg
481  tgtgctaaca atgacttcgt tcaagtggat ccagggcagg ttggcttaat aaattcaa
541  gtgtctatgg tagatttaac cagtgttacc ttagaagatg taacacctaa gaagtcacca
601  ccaccaaaga caaatttaac ccctggccaa agtgattcag caaaaaggga agtaaagcag
35  661  aagatcagtc cctactttaa agtaaatgat gtggtgtgca aaaatcaaga tgagctgaga
721  aatcgtagtg tgaagtcac ttgtttggga agcctagcat ctaaattgtc cagaaaatac
781  gtaaaggcta aaaaatcaat agataaggat gaagaatttg ccggttctag tccacagagt
841  tccaaatcca cagttgttaa gagcctgatt gataactctt cagaaattga ggacgaggat
901  caaattttgg agaacagttc tcaaaaagaa aacgtgttta aatgtgattc tctaaaggaa
40  961  gagtgcattc ctgaacatat ggtaagagga agtaaaataa tggaagccga aagccaaaag

```

	1021	gctacccggg	aatgtgagaa	atcagccctc	acccttgat	tctcagataa	tgcgatcatg
	1081	ttattctcac	cagatttcac	tcttaggaat	acattaaagt	ctacttcaga	agacagtctt
	1141	gtaaagcaag	agtgtatcaa	agaagtgggt	gaaaaacgtg	aggcatgtca	ttgtgaagaa
	1201	gtaaaaatga	ctgttgcttc	agaagctaaa	atacagctgt	cagattcaga	ggcaaaatct
5	1261	catagttctg	cagatgatgc	ttctgcatgg	agtaacatcc	aagaggctcc	tctgcaggat
	1321	gacagttgct	taaacaatga	tatccctcac	agcattcctt	tggagcaggg	gtcaagctgc
	1381	aatggctctg	gtcaaacaac	cggctatcct	tactaccttc	ggagtctcct	tgtggtgctg
	1441	aaaaccgtac	ttgagaatga	agatgatatg	ttgctctttg	atgagcagga	gaagggaatt
	1501	gtaactaaat	tttatcagtt	atcagctact	ggtcagaagt	tatatgtaag	gctctttcaa
10	1561	cgtaaatata	gctggattaa	gatgaccaa	ttagagtatg	aagagattgc	cttagactta
	1621	acacctgtga	ttgaagaatt	gacgaatgca	ggctttctac	agacagaatc	tgagttgcaa
	1681	gaactctctg	aagtgttga	actcctttct	gctcctgaac	taaaatccct	agccaagacc
	1741	ttccacttgg	tgaatcccaa	tggacagaaa	cagcagctgg	tggagcctt	tctcaaattg
	1801	gccaaacagc	gttcagtctg	cacttggggc	aagaataagc	ctggaattgg	tgcatgtatt
15	1861	ttaaaaagag	ccaaagcctt	ggctggacag	tcagtacgaa	tctgtaaagg	ccccagggct
	1921	gtgttttccc	gcactttgct	actgttttcg	ttgaccgact	caatggaaga	tgaagacgcc
	1981	gcttgtggag	gtcagggaca	gctttcaaca	gtcctgttgg	tcaacctcgg	ccgaatggag
	2041	tttcctagtt	acaccatcaa	tcggaaaacc	cacatcttcc	aagacagaga	tgatcttatc
	2101	agatatgcag	cagccacgca	catgctgagt	gacatttctt	ccgcaatggc	caatgggaac
20	2161	tggaagaag	ctaaggagct	cgctcagtgt	gcaaaaaggg	attggaacag	actgaaaaac
	2221	cacccttctc	tgagatgcca	cgaagattta	ccactcttcc	tgcggtgttt	cactgttggg
	2281	tggatttata	caaggatttt	gtctcggttt	gtggaataac	tgacagagct	tcacatgtat
	2341	gaggaagccg	tcagagaact	tgaagcctt	ttgtctcaga	gaatttattg	tcctgacagc
	2401	agaggccgat	ggtgggatcg	actggccctt	aatttacacc	agcacttgaa	gcgcttgga
25	2461	ccgactatca	agtgcacac	agaggggctg	gcggatccgg	aagtcagaac	gggacaccgc
	2521	ctttcactgt	atcagcgagc	cgctcgctg	cgagagtctc	cgagctgtaa	aaagttcaag
	2581	cacctcttcc	agcagctccc	agaaatggct	gtgcaagatg	tgaacacgt	gaccatcaca
	2641	ggcaggctgt	gcccacagcg	tgggatgtgc	aagtctgtgt	ttgtgatgga	ggccggggag
	2701	gccgctgacc	ccaccacggg	cctgtgctct	gtggaggagc	tggcactggc	ccattacaga
30	2761	cgcagcgggt	ttgaccaggg	gattcatggc	gaagggtcca	ccttcagcac	cctgtatggc
	2821	ctctcctgt	gggacatcat	cttcatggat	gggattccgg	atgtcttcag	aaacgcctgt
	2881	caggcattcc	ccctggactt	gtgcacagac	agcttcttca	caagcagacg	cccagccctt
	2941	gaggccaggc	tgacgtgat	tcatgatgcc	cccaggagga	gcctgcgggc	ctgggtggca
	3001	gccacgtggc	atgagcagga	aggcagagtg	gcttcccttg	tcagctggga	tcgcttcacg
35	3061	tctcttcagc	aagctcagga	tcttgtctcc	tgcctggggg	gccctgtgct	cagtgggtgtg
	3121	tgacggcacc	tggctgtgta	ctttcgacac	tgtcgagggg	gcctccccga	cctgggtggtg
	3181	tggaactccc	agagccgtca	ctttaagctg	gtggaagtta	aaggcccca	tgatcgtctt
	3241	tcacataagc	agatgatctg	gctggctgaa	ctgcagaagc	tgggggctga	agtagaagtc
	3301	tgccatgtgg	ttgcagttgg	agctaagagc	caaagcctta	gctaaaagat	tccctacagg
40	3361	agaaaaatga	aatgaggagg	agagaaactc	cgggtgtccc	gaggtgtcgg	tgtggtgagg
	3421	gccgctggcg	ttgaagtaca	tcctgtctctg	gccagctcc	ccatagcagg	cctccagggg
	3481	gccactgctg	tggtgccgca	gcacctgtct	cagtacgtcg	acttcatcag	ccaggaggga
	3541	gagcttgtga	aaggctgtga	tggagccacc	caggctgatc	tgggcctcgg	gaaccacagc
	3601	gaagtagcac	agtttcacac	gttttatgtg	gtttccagag	acacgtggca	gaataacacc
45	3661	gtgcaggttg	gcgggtttgg	aaaaccattc	tctaaaatac	tgctccgtat	cactgttctg
	3721	gctgtcgggt	tgctgagctg	gatctggctt	tggttttaat	atcaatgaat	ttctccttgg

3781 aagtaattct tggtcactga tgattccatt ctttaaggca gacggcattc ctcttagtgt
 3841 ggagctgtag cttttctata cagaagagat tttattatgt tccggggatt cccttttttag
 3901 aaagattgaa ggatgcaatg gcaaatataa actcaatact atgaaaaatt aatggaattt
 3961 cagcctcaaa gaacattttc ctcccttcct ttgtgtcctt attctaatacc tcctcccctg
 5 4021 gaattacact tttttatgtg ttgactctac ctaggctgtt actatcagcc tgaatggggg
 4081 cgggatgaga gtacctccta tccactaatt tgcttaagga taagttctaa gacgggctag
 4141 aaaaaaact agacctggcc gattctatca agaacaatgg caaactgaac agaggcagtc
 4201 aggaggccaa atgtctgatt ctttgttctg tacctttcag tagtctgcaa attttctacc
 4261 aaaaaaaatc ccaagaattt atttgggaat tattaataaag gcaacaatg aatgttatta
 10 4321 ggacaagaat atagcagtca ggaggccatg actacatcac agccaggcgg cattccctgc
 4381 cacagtggcg gcttgaatca tcaagaaatg gataaatggg gctttagtaa atcaggcttg
 4441 caggctcaaa gctgcaatct gccactctc aggtactgag actttgtggg cctcagacac
 4501 caggaagaaa gctgggatac agtcatttga gttaaaaagg gaatgacccc tcagaaaccc
 4561 gcattagcag tgttactctt ggaagtgcct ttacttttaa cgctctctgt tctgaaaaag
 15 4621 aggtgtttgg ttacgtgtga gccaacatca cgttttgtta gctgtgattt acctttgtcc
 4681 gtttaaaaga cttcacggag ccattctgta tacaagggtg gctctttcca atgtagaagg
 4741 ggttatggaa aagggtgcga tcctttgctg taaactggag agaccagtcc caaacagagg
 4801 ggaattttta gcccttctca tcaccaatg ggatgtttt gcttatagca aattcctgca
 4861 aaataaataa ataatattt gcaaaactaa aaaaaaaaaa a
 20

The amino acid sequence of human FAN1 (Uni-prot ref: Q9Y2M0) (SEQ ID NO: 2) may be found below:

1 MMSEGKPPDK KRPRRSLSIS KNKKKASNSI ISCFNNAPPA KLACPVCSKM VPRYDLNRHL
 25 61 DEMCANNDVF QVDPGQVGLI NSNVSMVDLT SVTLEDVTPK KSPPPKTNLT PGQSDSAKRE
 121 VKQKISPYFK SNDWCKNQD ELRNRSVKVI CLGSLASKLS RKYVKAQKSI DKDEEFAGSS
 181 PQSSKSTWK SLIDNSSEIE DEDQILENSS QKENVFKCDS LKEECIPEHM VRGSKIMEAE
 241 SQKATRECEK SALTPGFSDN AIMLFSPDFT LRNTLKSTSE DSLVKQECIK EWEKREACH
 301 CEEVKMTVAS EAKIQLSDSE AKSHSSADDA SAWSNIQEAP LQDDSCLNND IPHSIPLEQG
 30 361 SSCNGPGQTT GHPYYLRSFL WLKTVLENE DDMLLFDEQE KGIVTKFYQL SATGQKLYVR
 421 LFQRKLSWIK MTKLEYEEIA LDLTPIIEEL TNAGFLQTES ELQELSEVLE LLSAPELKSL
 481 AKTFHLVNPV GQKQQLVDAF LKLAKQRSVC TWGKNKPGIG AVILKRAKAL AGQSVRICKG
 541 PRAVFSRILL LFSLTDSMED EDAACGGQGQ LSTVLIVNLG RMEFPSYTIN RKTHIFQDRD
 601 DLIRYAAATH MLSDISSAMA NGNWEAEKEL AQCAKRDWNR LKNHPSLRCH EDLPLFLRCF
 35 661 TVGWIYTRIL SRFVEILQRL HMYEEAVREL ESLLSQRIYC PDSRGRWWD R LALNLHQHLK
 721 RLEPTIKCIT EGLADPEVRT GHRLSLYQRA VRLRESPCK KFKHLFQQLP EMAVQDVKHV
 781 TITGR LCPQR GMCKSVFVME AGEAADPTTV LCSVEELALA HYRRSGFDQG IHGEGSTFST
 841 LYGLLLWDII FMDGIPDVFR NACQAFPLDL CTDSFFTSRR PALEARLQLI HDAPEESLRA
 901 WVAATWHEQE GRVASLVSWD RFTSLQQAQD LVSCLGPPVL SGVCRHLAAD FRHCRGGLPD
 40 961 LWNSQSRH FKLVEVKGN DRLSHKQMIW LAELQKLGAE VEVCHVAVG AKSQSL

The human FANCD2 polypeptide and polynucleotide sequences are shown below and are shown at GenBank ref: NM_033084.3 and Uni-Prot ref: Q9BXW9. The skilled

person will readily be able to identify other FANCD2 polypeptide sequences from databases. For example, the Homologene feature of the NCBI database may be used.

The nucleotide sequence of human FANCD2 (GenBank ref: NM_033084.3) (SEQ ID NO:

5 3) may be found below:

```

1  ggcctggcgg  gaaagtcgaa  aactacgggc  ggcgacggct  tctcggaagt  aatttaagtg
61  cacaagacat  tggTcaaaat  ggTttccaaa  agaagactgt  caaaatctga  ggataaagag
121 agcctgacag  aagatgcctc  caaaaccagg  aagcaaccac  tttccaaaaa  gacaaagaaa
181 tctcatattg  ctaatgaagt  tgaagaaaat  gacagcatct  ttgtaaagct  tcttaagata
10 241 tcaggaatta  ttcttaaaac  gggagagagt  cagaatcaac  tagctgtgga  tcaaatagct
301 ttccaaaaga  agctctttca  gaccctgagg  agacaccctt  cctatcccaa  aataatagaa
361 gaatttgTta  gtggcctgga  gtcttacatt  gaggatgaag  acagtttcag  gaactgcctt
421 ttgtcttgTg  agcgtctgca  ggatgaggaa  gccagtatgg  gtgcatctta  ttctaagagt
481 ctcatcaaac  tgcttctggg  gattgacata  ctgcagcctg  ccattatcaa  aaccttattt
15 541 gagaagttgc  cagaatattt  ttttgaaaac  aagaacagtg  atgaaatcaa  catacctcga
601 ctcatgtgca  gtcaactaaa  atggcttgac  agagtgtgTg  atggcaagga  cctcaccacc
661 aagatcatgc  agctgatcag  tattgtctca  gagaacctgc  agcatgacat  catcaccagc
721 ctacctgaga  tcctagggga  ttcccagcac  gctgatgtgTg  ggaaagaact  cagtgcacta
781 ctgatagaga  atacttcact  cactgtccca  atcctggatg  tcctttcaag  cctccgactt
20 841 gacccaaact  tcctattgaa  ggTtcgccag  ttggtgatgTg  ataagttgtc  gtctattaga
901 ttggaggatt  tacctgtgat  aataaagTtc  attcttcatt  ccgtaacagc  catggataca
961 cttgaggtaa  tttctgagct  tcgggagaag  ttggatctgc  agcattgtgt  tttgccatca
1021 cggTtacagg  cttcccaagt  aaagTtgaaa  agtaaaggac  gagcaagTtc  ctcaggaaat
1081 caagaaagca  gcggtcagag  ctgtattatt  ctctctttg  atgtaataaa  gtcagctatt
25 1141 agatatgaga  aaaccatttc  agaagcctgg  attaaggcaa  ttgaaaacac  tgcctcagta
1201 tctgaacaca  aggtgtttga  cctggtgatg  cttttcatca  tctatagcac  caatactcag
1261 acaaagaagt  acattgacag  ggtgctaaga  aataagattc  gatcaggctg  cattcaagaa
1321 cagctgctcc  agagtacatt  ctctgttcat  tacttagttc  ttaaggatat  gtgttcatcc
1381 attctgtcgc  tggctcagag  tttgcttcac  tctctagacc  agagtataat  ttcatTTggc
30 1441 agtctcctat  acaaatatgc  atttaagTtt  tttgacacgt  actgccagca  ggaagTggtt
1501 ggtgccttag  tgaccatat  ctgcagtggg  aatgaagctg  aagttgatac  tgccttagat
1561 gtccttctag  agttggtagt  gttaaaccCa  tctgctatga  tgatgaatgc  tgtctttgta
1621 aagggcattt  tagattatct  ggataacata  tcccctcagc  aaatacgaaa  actcttctat
1681 gttctcagca  cactggcatt  tagcaaacag  aatgaagcca  gcagccacat  ccaggatgac
35 1741 atgcacttgg  tgataagaaa  gcagctctct  agcaccgtat  tcaagtacaa  gctcattggg
1801 attattggTg  ctgtgaccat  ggctggcatc  atggcggcag  acagaagtga  atcacctagt
1861 ttgaccaag  agagagccaa  cctgagcgat  gagcagtgca  cacaggTgac  ctcttgtTg
1921 cagttggttc  attcctgcag  tgagcagtct  cctcaggcct  ctgcacttta  ctatgatgaa
1981 tttgccaacc  tgatccaaca  tgaaaagctg  gatccaaaag  ccctggaatg  ggttgggcat
40 2041 accatctgta  atgatttcca  ggatgccttc  gtagtggact  cctgtgtTgt  tccggaaggt
2101 gactttccat  ttctgtgTaa  agcactgtac  ggactggaag  aatacgacac  tcaggatggg
2161 attgccataa  acctcctgcc  gctgctgttt  tctcaggact  ttgcaaaaaga  tgggggtccg
2221 gtgacctcac  aggaatcagg  ccaaaaattg  gtgtctccgc  tgtgcctggc  tccgtatttc
2281 cggTtactga  gactttgtgt  ggagagacag  cataacggaa  acttgaggga  gattgatggt

```

	2341	ctactagatt	gtcctatatt	cctaactgac	ctggagcctg	gagagaagtt	ggagtccatg
	2401	tctgctaaag	agcgttcatt	catgtgttct	ctcatatttc	ttactctcaa	ctggttccga
	2461	gagattgtaa	atgccttctg	ccaggaaaca	tcacctgaga	tgaaggggaa	gggtctcact
	2521	cggtaaagc	acattgtaga	attgcaaata	atcctggaaa	agtacttggc	agtcacccca
5	2581	gactatgtcc	ctcctcttgg	aaactttgat	gtggaaactt	tagatataac	acctcatact
	2641	gttactgcta	tttcagcaaa	aatcagaaaag	aaaggaaaaa	tagaaaggaa	acaaaaaaca
	2701	gatggcagca	agacatcctc	ctctgacaca	ctttcagaag	agaaaaattc	agaatgtgac
	2761	cctacgccat	ctcatagagg	ccagctaaac	aaggagttca	caggggaagga	agaaaagaca
	2821	tcattgttac	tacataattc	ccatgctttt	ttccgagagc	tggacattga	ggcttctctt
10	2881	attctacatt	gtggacttgt	gacgaagttc	atcttagata	ctgaaatgca	cactgaagct
	2941	acagaagttg	tgcaacttgg	gccccctgag	ctgcttttct	tgctggaaga	tctctcccag
	3001	aagctggaga	gtatgctgac	acctcctatt	gccaggagag	tcccccttct	caagaacaaa
	3061	ggaagccgga	atattggatt	ctcacatctc	caacagagat	ctgccaaga	aattgttcat
	3121	tgtgtttttc	aactgctgac	cccaatgtgt	aaccacctgg	agaacattca	caactatttt
15	3181	cagtgttttag	ctgctgagaa	tcacggtgta	gttgatggac	caggagtga	agttcaggag
	3241	taccacataa	tgtcttctctg	ctatcagagg	ctgctgcaga	tttttcatgg	gctttttgct
	3301	tggagtggat	tttctcaacc	tgaaaatcag	aatttactgt	attcagccct	ccatgtcctt
	3361	agtagccgac	tgaaacaggg	agaacacagc	cagcctttgg	aggaactact	cagccagagc
	3421	gtccattact	tgcagaatth	ccatcaaagc	attcccagtt	tccagtgtgc	tctttatctc
20	3481	atcagacttt	tgatggttat	tttgagaaaa	tcaacagctt	ctgctcagaa	caaagaaaaa
	3541	attgcttccc	ttgccagaca	attcctctgt	cgggtgtggc	caagtgggga	taaagagaag
	3601	agcaacatct	ctaattgacca	gctccatgct	ctgctctgta	tctacctgga	gcacacagag
	3661	agcattctga	aggccataga	ggagattgct	gggtgtgggtg	tcccagaact	gatcaactct
	3721	cctaaagatg	catcttctctc	cacattccct	acactgacca	ggcatacttt	tgttgttttc
25	3781	ttccgtgtga	tgatggctga	actagagaag	acgggtgaaa	aaattgagcc	tggcacagca
	3841	gcagactcgc	agcagattca	tgaagagaaa	ctcctctact	ggaacatggc	tgttcgagac
	3901	ttcagtatcc	tcatcaactt	gataaaggta	tttgatagtc	atcctgttct	gcatgtatgt
	3961	ttgaagtatg	ggcgtctctt	tgtggaagca	tttctgaagc	aatgtatgcc	gctcctagac
	4021	ttcagtttta	gaaaacaccg	ggaagatgtt	ctgagcttac	tggaaacctt	ccagttggac
30	4081	acaaggctgc	ttcatcacct	gtgtgggcat	tccaagattc	accaggacac	gagactcacc
	4141	caacatgtgc	ctctgctcaa	aaagaccctg	gaacttttag	tttgagagt	caaagctatg
	4201	ctcactctca	acaattgtag	agaggctttc	tggctgggca	atctaaaaaa	ccgggacttg
	4261	cagggtgaag	agattaagtc	ccaaaattcc	caggagagca	cagcagatga	gagtgaggat
	4321	gacatgtcat	cccaggcctc	caagagcaaa	gccactgagg	tatctctaca	aaaccaccca
35	4381	gagtctggca	ctgatggttg	cattttgtta	attgttctaa	gttgggtggag	cagaactttg
	4441	cctacttatg	tttattgtca	aatgcttcta	tgccattttc	cattccctcc	ataacagctt
	4501	ctgtgcttat	ataatttttg	ggaccagaa	gaaacaacga	cacaatctta	gaatcactcc
	4561	tgagtatctc	gagttgtggc	atthgttata	gagttgacaa	ttttctgcat	tatagcctct
	4621	cattttccat	gaattcatat	ctgaaaccat	tttagaagg	agaagtcatc	gaagtatttt
40	4681	ctgagtgttg	agaagaatga	gttaaaccat	ttaaacacat	ttgaaacata	caaaaataga
	4741	aatgtgaaag	catttgggtga	aagccaaagc	acagagtcag	aagctgccac	cttagagaac
	4801	tgaaataaaa	atagaagttc	ttacgctttt	ttgtggtaca	gatgctttcg	acaatttaaa
	4861	gaaagctaaa	taaaaatgta	gacatggctg	gcgcagtggc	tcatgcttgt	aatcctagca
	4921	ctttttgagg	ccaaggtagg	aggattgctt	gagtccggga	gctcaaggca	aagctgcaca
45	4981	acataacaag	accctatctc	cacaaaaaaa	atgaaaaata	aacctgggtg	cgggtggctca
	5041	cacctgtaat	cccagcactt	tgggaggccg	atgtgggcag	atcacagggt	caggagttca

5101 agaccagcct ggccaacata gtgaaacccc atctctactg aaaatacaaa aattagctgg
 5161 gtgtggtggc acgtgcctgt tatctcagct acttgggagg ctga

5 The amino acid sequence of human FANCD2 (Uni-prot ref: Q9BXW9) (SEQ ID NO: 4) may be found below:

1 MVSKRRLSKS EDKESLTEDA SKTRKQPLSK KTKKSHIANE VEENDSIFVK LLKISGIILK
 61 TGESQNQLAV DQIAFQKKLF QTLRRHPSYP KIIIEFVSGL ESYIEDEDSF RNCLLSKERL
 121 QDEEASMGAS YSKSLIKLLL GIDILQPAII KTLFEKLPEY FFENKN_{SD}EI NIPRLIVSQL
 10 181 KWLDRWDGK DLTTKIMQLI SIAPENLQHD IITSLPEILG DSQHADVGEK LSDLLIENTS
 241 LTVPILDVLS SLRLDPN_FLL KVRQLVMDKL SSIRLEDLPV IIKFILHSVT AMDTLEVISE
 301 LREKLDLQHC VLPSRLQASQ VKLKSKGRAS SSGNQESSGQ SCII_{LL}FDVI KSAIRYEKTI
 361 SEAWIKAIEN TASVSEHKVF DLVMLFIIYS TNTQTKKYID RVL_{RN}KIRSG CIQEQLLQST
 421 FSVHYLV_LKD MCSSILSLAQ SLLHSLDQSI ISFGSLLYKY AFKFFDTYCQ QEWGALVTH
 15 481 ICSGNEAEVD TALDV_{LL}ELV VLNPSAM_{MM}N AVFVKGILDY LDNISPQQIR KLFYVLSTLA
 541 FSKQNEASSH IQDDMHLVIR KQLSSTVFKY K_LIGIIGAVT MAGIM_AADRS ESPSLTQERA
 601 NLSDEQCTQV TSL_LQLVHSC SEQSPQASAL YYDEFANLIQ HEKLDPKALE WVGHTICNDF
 661 QDAFW_DSCV VPEGDFPFV KALYGLEEYD TQDGI_AINLL PLLFSQDFAK DGGPVTSQES
 721 GQKLVSPLCL APYFRLRLC VERQHNGNLE EIDGL_{LL}DCPI FLTDLEPGEK LESMSAKERS
 20 181 FMCSLIFLTL NWFREIVNAF CQETSPEMKG KVLTRLKHIV ELQI_{IL}EKYL AVTPDYVPPL
 841 GNFDVETLDI TPHTVTAISA KIRKKGKIER KQKT_DGSKTS SSDTLSEEKN SEC_{DPT}PSHR
 901 GQLNKEFTGK EEKTS_{LL}LHN SHAFFRELDI EVFSILHCGL VTKFILDTEM HTEATEWQL
 961 GPPELLFLLE DLSQKLESL TPPIARRVPF LKNKGSRNIG FSHLQ_QRSAQ EIVHCVFQLL
 1021 TPMCNHLENI HNYFQCLAAE NHGWDGPGV KVQEYHIMSS CYQRL_LQIFH GLFAWSGFSQ
 25 1081 PENQNLLYSA LHV_LSSRLKQ GEHSQPLEEL LSQSVHYLQN FHQSIP_SFC ALYLIRLLMV
 1141 ILEKSTAS_AQ NKEKIASLAR QFLCRVWPSG DKEKSNISND QLHALLCIYL EHTESILKAI
 1201 EEIAGVG_VPE LINS_{PK}DASS STFPTL_{TR}HT FVVFFRVMMA ELEKTVKKIE PGTAADSQ_QI
 1261 HEEKLLY_WNM AVRDFSILIN LIKVFDSHPV LHVCLKYGRL FVEAFLKQCM PLLDFSFRKH
 1321 REDVLSLLET FQLD_{TR}LLHH LCGHSKIHQD TRLTQHVP_{LL} KKTLEL_LVCR VKAMLT_{LN}NC
 30 1381 REAFWLGNLK NRDLQGE_EIK SQNSQESTAD ESEDD_MSSQA SKSKATEVSL QNP_{PES}GT_DG
 1441 CILLIVLSW SRTLPTV_YYC QMLLC_{PF}FP P

The human FANCI polypeptide and polynucleotide sequences are shown below and are shown at GenBank ref: NM_001113378.1 and Uni-Prot ref: Q9NVI1. The skilled person will readily be able to identify other FANCI polypeptide sequences from databases. For example, the Homologene feature of the NCBI database may be used.

The nucleotide sequence of human FANCI (GenBank ref: NM_001113378.1) (SEQ ID NO: 5) may be found below:

1 tcttgttggtt acgggtaacg gaagtgtggc ggcgttgggt tgagcgggct ttttgaagt
 61 ttgtggcgga gttctgtgat atgagcaaca atggaccaga agattttatc tctagcagca
 121 gaaaaaacag cagacaaact gcaagaattt cttcaaacc ttagagaagg tgatttgact

	181	aatctccttc	agaatcaagc	agtgaagga	aaagttgctg	gagcactcct	gagagccatc
	241	ttcaaagggt	ccccctgctc	tgaggaagct	ggaacactta	ggagacgtaa	gatatacact
	301	tgttgatcc	agttggtgga	atcgggggat	ttgcagaaag	aaatagcgtc	tgagatcata
	361	ggattactga	tgctggaggc	tcaccatttt	ccaggaccat	tattggttga	attagccaat
5	421	gagtttatta	gtgctgtcag	agaaggcagc	ctagtgaatg	gaaaatcttt	ggagttacta
	481	cctatcattc	'tcaactgccct	ggctacgaaa	aaggaaaatc	tggcttatgg	aaaagggtga
	541	ctgagtgggg	aagaatgtaa	gaaacagttg	attaacaccc	tgtgttctgg	caggtgggat
	601	cagcaatatg	taatccaact	cacctccatg	ttcaaggatg	tccctctgac	tcgagaagag
	661	gtggaatttg	tggtggaaaa	agcattgagc	atgttctcca	agatgaatct	tcaagaaata
10	721	ccacctttgg	tctatcagct	tctggttctc	tcctccaagg	gaagcagaaa	gagtgttttg
	781	gaaggaatca	tagccttctt	cagtgcacta	gataagcagc	acaatgagga	acagagtggg
	841	gacgagctat	tggatgttgt	cactgtgcc	tcaggtgaac	ttcgtcatgt	ggaaggcacc
	901	attattctac	acattgtgtt	tgccatcaaa	ttggactatg	aactaggcag	agaactcgtg
	961	aaacacttaa	aggtaggaca	gcaaggagat	tccaataata	acttaagtcc	cttcagcatt
15	1021	gctcttcttc	tgtctgtaac	aagaatacaa	agatttcagg	accaggtgct	tgatctttta
	1081	aagacttcgg	ttgtaaagag	ctttaaggat	cttcaactcc	tccaaggctc	aaaatttctt
	1141	cagaatctag	ttcctcatag	atcttatgtt	tcaaccatga	tcttggaagt	agtgaagaat
	1201	agcgttcata	gctgggacca	tgttactcag	ggcctcgtag	aacttggttt	cattttgatg
	1261	gattcatatg	ggccaaagaa	ggttcttgat	ggaaaaacta	ttgaaaccag	cccaagtctt
20	1321	tctagaatgc	caaaccagca	tgcatgtaag	ctcggagcta	atatcctgtt	ggaaactttt
	1381	aagatccatg	agatgatcag	acaagaaatt	ttggagcagg	tcctcaacag	ggttggtacc
	1441	agagcatctt	ctcccatcag	tcatttctta	gacctgcttt	caaatatcgt	catgtatgca
	1501	cccttagttc	ttcaaagttg	ttcttctaaa	gtcacagaag	cttttgacta	tttgtccttt
	1561	ctgccccttc	agactgtaca	aaggctgctt	aaggcagtc	agccccttct	caaagtcagc
25	1621	atgtcaatga	gagactgctt	gatactgttc	cttcggaaag	ctatgtttgc	caaccagctt
	1681	gatgcccga	aatctgcagt	tgctgggttt	ttgctgctcc	tgaagaactt	taaagtttta
	1741	ggcagcctgt	catcctctca	gtgcagtcag	tctctcagtg	tcagtcaggt	tcagtgtgat
	1801	gttcacagcc	attacaattc	tgctgccaat	gaaacttttt	gccttgagat	catggatagt
	1861	ttgaggagat	gcttaagcca	gcaagctgat	gttcgactca	tgctttatga	gggtttttat
30	1921	gatgttcttc	gaaggaactc	tcagctggct	aattcagtc	tgcaaactct	gctctcacag
	1981	ttaaaacagt	tctatgagcc	aaaacctgat	ctgctgcctc	ctctgaaatt	agaagcttgt
	2041	attctgacct	aaggagataa	gatctctcta	caagaaccac	tggattatct	gctgtgttgt
	2101	attcagcatt	gtttggcctg	gtataagaat	acagtcatac	ccttacagca	gggagaggag
	2161	gaagaggagg	aggaagaggc	attctacgaa	gacctagatg	atatattgga	gtccattact
35	2221	aatagaatga	ttaagagtga	gctggaagac	tttgaactgg	ataaatcagc	agatttttct
	2281	cagagcacca	gtattggcat	aaaaaataat	atctgtgctt	ttcttgtgat	gggagtttgt
	2341	gaggttttaa	tagaatacaa	tttctccata	agtagtttca	gtaagaatag	gtttgaggac
	2401	attctgagct	tatttatgtg	ttacaaaaaa	ctctctgaca	ttcttaatga	aaaagcgggt
	2461	aaagccaaaa	ctaaaatggc	caacaagaca	agtgatagtc	ttttgtccat	gaaatttgtg
40	2521	tccagtcttc	tcaactgctc	tttcagggat	agtatccaaa	gccaccaaga	aagcctttct
	2581	gttctcaggt	ccagcaatga	gtttatgctg	tatgcagtga	atgtagctct	gcagaaagta
	2641	cagcagctaa	aggaaacagg	gcatgtgagt	ggccctgatg	gccaaaaccc	agaaaagatc
	2701	tttcagaacc	tctgtgacat	aactcgagtc	ttgctatgga	gatacacttc	aattcctact
	2761	tcagtggaa	agtcgggaaa	gaaagagaaa	ggaaagagca	tctcactgct	gtgcttgagg
45	2821	ggtttacaga	aaatattcag	tgctgtgcaa	cagttctatc	agcccaagat	tcagcagttt
	2881	ctcagagctc	tggatgtcac	agataaggaa	ggagaagaga	gagaagatgc	agatgtcagt

	2941	gtcactcaga	gaacagcatt	ccagatccgg	caatttcaga	ggtccttggt	gaatttactt
	3001	agcagtcaag	aggaagattt	taatagcaaa	gaagccctcc	tgctagtcac	ggttcttacc
	3061	agtttgtcca	agttactgga	gccctcctct	cctcagtttg	tcagatggt	atcctggaca
	3121	tcaaagattt	gcaaggaaaa	cagccgggag'	gatgccttgt	tttgcaagag	cttgatgaac
5	3181	ttgctcttca	gcctgcatgt	ttcgtataag	agtcctgtca	ttctgctgcg	tgacttgtcc
	3241	caggatatcc	acgggcatct	gggagatata	gaccaggatg	tagagggtga	gaaaacaaac
	3301	cactttgcaa	tagtgaattt	gagaacggct	gccccactg	tctgtttact	tgttctgagt
	3361	caggccgaga	aggttctaga	agaagtggac	tggtaatca	ccaagcttaa	gggacaagtg
	3421	agccaagaaa	ccttatcaga	agaggcctct	tctcaggcaa	ccctacaaa	tcagcctgtt
10	3481	gagaaagcta	tcacatgca	actgggaact	ctgcttacat	ttttccacga	gctggtgcag
	3541	acagctctgc	catcaggcag	ctgtgtggac	accttggtta	aggacttggt	caaaatgtac
	3601	accacactta	cagcccttgt	cagatattat	ctccagggtg	gtcagagctc	cggaggaatt
	3661	ccaaaaata	tgaaaagct	ggtgaagctg	tctggttctc	atctgacccc	cctgtgttat
	3721	tctttcattt	cttacgtaca	gaataagagt	aagagcctga	actatacggg	agagaaaaag
15	3781	gagaaacctg	ctgccgttgc	cacagccatg	gccagagttc	ttcgggaaac	caagccaatc
	3841	cctaacctca	tctttgccat	agaacagtat	gaaaaatttc	tcacccacct	ttctaagaag
	3901	tccaaggtga	acctgatgca	gcacatgaag	ctcagcacct	cacgagactt	caagatcaaa
	3961	ggaaacatcc	tagacatggt	tcttcgagag	gatggtgaag	atgaaaatga	agagggcact
	4021	gcatcagagc	atgggggaca	gaacaaagaa	ccagccaaga	agaaaaggaa	aaaataaatg
20	4081	aaatgcctga	gttaatgtga	actttggggc	ttctgcttca	tttttaccba	acaagcaaca
	4141	atgccccttg	tcctgtagtc	cacaccgatg	ttggcatctt	ggttctgaac	ccactgaatt
	4201	caactgcacc	ttcagttaga	aggaatcttc	ttggcaggtc	ctgctactga	aaaatggctg
	4261	gccttaggca	agcccttttg	caaaaagcac	agctgaaagc	ctgagtttgg	gagcctgcac
	4321	caccccgatg	aagctccacg	ggagcaaata	cagagcctcc	aggcagtgct	atggtccagg
25	4381	ctggcttcgt	ttttccaagg	agcctttggt	gagttcaatt	atctggtaaa	tatccagcgc
	4441	ttcacctgaa	agatagtga	aattgggttag	gatgccacct	caagaactgt	aactgagagc
	4501	tcagaagtga	gcaaaggagc	ttaatgctaa	ggtcaaaagg	agagtgaag	gttgagaaca
	4561	attgccacga	acggtaatgt	tacatgttag	gagggtctgt	tttcttttta	tataagtgtg
	4621	tcttagatat	attttaata	gaaaataagc	tttctgattt	acttgtttgg	tatttaagc
30	4681	acagtttggt	tttctgtcac	ctatagagtg	caagaatgca	ctctatagaa	taaattatct
	4741	ttaaacatt					

The amino acid sequence of human FANCI (Uni-prot ref: Q9NVI1) (SEQ ID NO: 6) may be found below:

35

	1	MDQKILSLAA	EKTADKLQEF	LQTLREGDLT	NLLQNQAVKG	KVAGALLRAI	FKGSPCSEEA
	61	GTLRRRKIYT	CCIQLVESGD	LQKEIASEII	GLLMLEAHHF	PGPLLVELAN	EFISAVREGS
	121	LVNGKSLLEL	PIILTALATK	KENLAYGKGV	LSGEECKKQL	INTLCSGRWD	QQYVIQLTSM
	181	FKDVPLTAE	VEFVVEKALS	MFSKMNLQEI	PPLVYQLLV	SSKGSRKSVL	EGIIAFFSAL
40	241	DKQHNEEQSG	DELLDWTVP	SGELRHVEGT	IILHIVFAIK	LDYELGRELV	KHLKVGQQGD
	301	SNNNLSPFSI	ALLSVTRI	RFQDQVLDLL	KTSWKSFKD	LQLLQGSKFL	QNLVPHRSYV
	361	STMILEWKN	SVHSWDHVTQ	GLVELGFILM	DSYGPKKVLD	GKTIETSPSL	SRMPNQHACK
	421	LGANILLETF	KIHEMIRQEI	LEQVLNRWT	RASSPISHFL	DLLSNIVMYA	PLVLQSCSSK
	481	VTEAFDYLSF	LPLQTVQRLL	KAVQPLLKVS	MSMRDCLILV	LRKAMFANQL	DARKSAVAGF
45	541	L.LLLKNFKVL	GSLSSSQCSQ	SLSVSQVHVD	VHSHYNSVAN	ETFCLEIMDS	LRRCLSQQAD

5
 601 VRLMLYEGFY DVLRRNSQLA NSVMQTLLSQ LKQFYEPKPD LLPPLKLEAC ILTQGDKISL
 661 QEPLDYLLCC IQHCLAWYKN TVIPLQQGEE EEEEEAFYE DLDDILESIT NRMKSELED
 721 FELDKSADFS QSTSIGIKNN ICAFLVMGVC EVLIEYNFSI SSFSKNRFED ILSLFMCYKK
 781 LSDILNEKAG KAKTKMANKT SDSLLSMKFV SLLLTALFRD SIQSHQESLS VLRSSNEFMR
 841 YAVNVALQKV QQLKETGHVS GPDGQNPEKI FQNLCDITRV LLWRYTSIPT SVEESGKKEK
 901 GKSISLLCLE GLQKIFSAVQ QFYQPKIQF LRALDVTDE GEEREDADVS VTQRTAFQIR
 961 QFQRSLLNLL SSQEEDFNSK EALLLVTVLT SLSKLLEPSS PQFVQMLSWT SKICKENSRE
 1021 DALFCKSLMN LLFSLHVSYSK SPVILLRDLS QDIHGHGDI DQDVEVEKTN HFAIVNLRTA
 1081 APTVCLLVLS QAEKVLEEVD WLITKLKGQV SQETLSEEAS SQATLPNQPV EKAIMQLGT
 10
 1141 LLTFFHELQV TALPSGSCVD TLLKDLCKMY TTLTALVRY LQVCQSSGGI PKNMEKLVKL
 1201 SGSHLTPLCY SFISYVQNK KSLNYTGEKK EKPAAVATAM ARVLRETKPI PNLIFAIEQY
 1261 EKFLIHLSKK SKVNLQHMK LSTSRDFKIK GNILDMVLRE DGEDENEEGT ASEHGGQNK
 1321 PAKKRRKK

15 The FAN1 or FANCD2 or FANCI polypeptide may comprise a tag sequence, as will be well known to those skilled in the art. For example, a tag useful in a FRET system may be used. For example a fluorescent protein tag, for example a Cherry tag may be used. A GST moiety or Green Fluorescent Protein (GFP) moiety or a FLAG moiety may alternatively be used. Typically the FAN1 or FANCD2 or FANCI is a full length FAN1 or
 20 FANCD2 or FANCI polypeptide.

The FAN1 polypeptide may be a NUS-His₆ fusion polypeptide. The FAN1 polypeptide may be recombinant.

25 The FAN1 polypeptide is typically human FAN1, but may alternatively be another mammalian FAN1, for example FAN1 of a laboratory animal or of a tissue or organ assay system considered useful in assessing a potential inhibitor of FAN1. Thus, the FAN1 may be a laboratory rodent FAN1 (for example mouse, rabbit or rat) or may be a laboratory primate FAN1, for example a monkey FAN1. An assay of the present
 30 invention may, for example, be useful in assessing the effect of a test compound on FAN1 in tissue of a laboratory animal, for example a mouse or a monkey. For example, a mouse human tumor xenograft model may be used to assess the effect of a test compound on FAN1.

35 It is particularly preferred, although not essential, that the FAN1 polypeptide has at least 30% of the enzyme activity of full-length human FAN1 on a polynucleotide substrate as defined herein. A particularly preferred substrate is the 5' flap shown in Figure 2A. It is more preferred if the FAN1 polypeptide has at least 50%, preferably at least 70% and

more preferably at least 90% of the enzyme activity of full-length human FAN1 on a polynucleotide substrate as defined herein.

By "variants" of a polypeptide we include insertions, deletions and substitutions, either conservative or non-conservative. In particular we include variants of the polypeptide where such changes do not substantially alter the endonuclease or 5'-exonuclease activity, or the interaction between FAN1 and FANCD2 or FANCI polypeptide, as appropriate. The skilled person will readily be able to design and test appropriate variants, based on, for example, comparison of sequences of examples of each polypeptide, for example from different species. The skilled person will readily be able to determine where insertions or deletions can be made; or which residues can appropriately be left unchanged; replaced by a conservative substitution; or replaced by a non-conservative substitution. The variant polypeptides can readily be tested, for example as described in the Examples. Optionally, the FAN1 polypeptide retains a functional UBZ domain, for example retains Cys44 and Cys47, as discussed in the Examples.

By "conservative substitutions" is intended combinations such as Gly, Ala; Val, Ile, Leu; Asp, Glu; Asn, Gln; Ser, Thr; Lys, Arg; and Phe, Tyr.

The three-letter or one letter amino acid code of the IUPAC-IUB Biochemical Nomenclature Commission is used herein, with the exception of the symbol Zaa, defined above. In particular, Xaa represents any amino acid. It is preferred that at least the amino acids corresponding to the consensus sequences defined herein are L-amino acids.

It is particularly preferred if the polypeptide variant has an amino acid sequence which has at least 65% identity with the amino acid sequence of the relevant human polypeptide, more preferably at least 70%, 71%, 72%, 73% or 74%, still more preferably at least 75%, yet still more preferably at least 80%, in further preference at least 85%, in still further preference at least 90% and most preferably at least 95% or 97% identity with the amino acid sequence of the relevant human polypeptide.

It is still further preferred if the FAN1 variant has an amino acid sequence which has at least 65% identity with the amino acid sequence of the catalytic domain of the human polypeptide, more preferably at least 70%, 71%, 72%, 73% or 74%, still more preferably at least 75%, yet still more preferably at least 80%, in further preference at least 83 or

85%, in still further preference at least 90% and most preferably at least 95% or 97% identity with the relevant human amino acid sequence. The location of the catalytic domain is discussed above and in the Examples.

- 5 The percent sequence identity between two polypeptides may be determined using suitable computer programs, for example the GAP program of the University of Wisconsin Genetic Computing Group and it will be appreciated that percent identity is calculated in relation to polypeptides whose sequence has been aligned optimally.
- 10 The alignment may alternatively be carried out using the Clustal W program (Thompson *et al.*, 1994). The parameters used may be as follows:
Fast pairwise alignment parameters: K-tuple(word) size; 1, window size; 5, gap penalty; 3, number of top diagonals; 5. Scoring method: x percent.
Multiple alignment parameters: gap open penalty; 10, gap extension penalty; 0.05.
- 15 Scoring matrix: BLOSUM.

The alignment may alternatively be carried out using the program T-Coffee, or EMBOSS.

- The residue corresponding (equivalent) to, for example, Cys44 of full-length human
- 20 FAN1 may be identified by alignment of the sequence of the polypeptide with that of full-length human FAN1 in such a way as to maximise the match between the sequences. The alignment may be carried out by visual inspection and/or by the use of suitable computer programs, for example the GAP program of the University of Wisconsin Genetic Computing Group, which will also allow the percent identity of the polypeptides
- 25 to be calculated. The Align program (Pearson (1994) in: *Methods in Molecular Biology, Computer Analysis of Sequence Data, Part II* (Griffin, AM and Griffin, HG eds) pp 365-389, Humana Press, Clifton). Thus, residues identified in this manner are also "corresponding residues".

- 30 It will be appreciated that in the case of truncated forms of (for example) FAN1 or in forms where simple replacements of amino acids have occurred it is facile to identify the "corresponding residue".

- It is preferred that the polypeptides used in the screen are mammalian, preferably human
- 35 (or a species useful in agriculture or as a domesticated or companion animal, for example dog, cat, horse, cow), including naturally occurring allelic variants (including splice variants). The polypeptides used in the screen may comprise a GST portion or

may be biotinylated or otherwise tagged, for example with a 6His, HA, myc or other epitope tag, as known to those skilled in the art, or as mentioned above or as described in the Examples. This may be useful in purifying and/or detecting the polypeptide(s).

- 5 The substrate polynucleotide may consist of or comprise the structure of the 5' flap polynucleotide identified in Figure 2A, which is efficiently cleaved *in vitro*. Modifications to the length and structure such that the utility of the substrate in the methods of the invention may be improved are standard in the art as would be understood by a person of skill in the art.

10

- Binding of the FAN1 polypeptide to an ubiquitinated FANCD2 or FANCI polypeptide may be assessed by any suitable technique for assessing protein:protein interaction. For example, the interaction between polypeptides may be detected using fluorescence resonance energy transfer; immunoprecipitation; subcellular fractionation or imaging; surface plasmon resonance (SPR); or isothermal titration calorimetry.
- 15

- Typically a FRET (fluorescence resonance energy transfer) technique may be used, as discussed further below. Other techniques that may be useful may make use of immunoprecipitation techniques. For example, immunoprecipitation may be with an antibody that binds specifically to FAN1; or may be with an antibody that binds specifically to an ubiquitinated FANCD2 or FANCI polypeptide, as will be apparent to the skilled person. Alternatively, immunoprecipitation may be with an antibody that binds specifically to a tag present on recombinant FAN1; or with an antibody that binds specifically to a tag present on a recombinant FANCD2 or FANCI polypeptide, as will also be apparent to the skilled person.
- 20
- 25

- As an example, it is considered that detection of FAN1 coupled with either ubiquitinated FANCD2 or FANCI co- pull down or an anti- ubiquitinated FANCD2 or FANCI antibody can be carried out using Invitrogen's Alpha-Elisa technologies, which would be useful in achieving a high throughput screening system. Multiplex assays using Luminex beads or plate based electrochemiluminescence (MSD; meso scale discovery) detection could also be used.
- 30

- Details of Alpha screen technology (Perkin Elmer) applicable to both protein:protein and phosphoprotein detection (Sure fire kits developed and sold for MAPK, JAK/STAT and AKT pathways) can be found at, for example, <http://las.perkinelmer.co.uk/Catalog/CategoryPage.htm?CategoryID=AlphaTech&M=BIO>
- 35

Details of Luminex technology applicable to phospho protein detection and protein:protein and total protein quantitation can be found at, for example, http://www.luminexcorp.com/applications/cellular_signaling.html

5

An example of the use of such technology is described in reference Khan IH, Zhao J., Ghosh, P. Ziman, M., Sweeney C, Kung HJ and Luciw PA (2010) Assay Drug Dev technology 8, 27-36.

10 In MSD technology the principles of capture onto surface of plate and antibody detection are the same as any ELISA but the mode of detection uses electrochemiluminescence via Ruthenium tagged probes, and the technology allows multiplexing in the well through an array format.

<http://www.mesoscale.com/CatalogSystemWeb/WebRoot/literature/brochures/pdf/techBr>

15

ochure.pdf

Quantitative Stable Isotope Labelling with Amino acids in Cell culture (SILAC)-based mass spectrometry may be used to identify and quantitate proteins associated with immunoprecipitates of FAN1 or ubiquitinated FANCD2 or FANCI. Other
20 immunoprecipitate methods may be used, as will be well known to those skilled in the art. For example labeled ubiquitinated FANCD2 or FANCI may be used. Some examples of such methods are described in the Examples. As noted above, other techniques for assessing protein:protein interactions in cells or cell extracts may also be used. For example, a fluorescence resonance energy transfer (FRET) based system
25 may be used if the interaction of a recombinant FAN1 and recombinant ubiquitinated FANCD2 or FANCI is being assessed, for example if both FAN1 and ubiquitinated FANCD2 or FANCI polypeptide are both tagged with a fluorescent polypeptide.

Thus, the molecular interaction between FAN1 and ubiquitinated FANCD2 or FANCI
30 proteins (and the effects of test compounds) could be investigated using a FRET-based method such as FLIM-FRET on a microscope such as a multiphoton microscope. As an example, a construct for expressing Cherry-tagged wild type ubiquitinated FANCD2 or FANCI or (as a control) an inactive mutant of FANCD2 or FANCI that does not bind FAN1 may be transfected into a cell line stably expressing wild type GFP-FAN1 or
35 control polypeptide. FRET (fluorescence resonance energy transfer) can occur when the GFP and mCherry fluorophores are brought together by virtue of the binding of FAN1 to ubiquitinated FANCD2 or FANCI which will in turn affects their fluorescence lifetime,

which can be detected. Using FLIM (fluorescence lifetime imaging microscopy) it is possible to generate a spatial distribution of the cell where sites of strong protein-protein interaction (and therefore FRET) and weak interaction or no interaction can be recognised (by colour coding: see, for example, Lieres et al. 2009 Quantitative analysis of chromatin compaction in living cells using FLIM-FRET. J Cell Biol. 2009 Nov 16;187(4):481-96.).

Commonly used FRET pairs include CFP (donor) and YFP (acceptor) as well as GFP (donor) and Cherry (acceptor). In cases where the donor and acceptor fluorophores are both excited with the same excitation light wavelength, e.g. in case of the FRET pair GFP-YFP, a special kind of FRET termed enhanced acceptor fluorescence (EAF) can be detected. Examples of further references concerning FRET techniques include Wallrabe & Periasamy (2005) Current Opinion in Biotechnology Volume 16, Issue 1, February 2005, Pages 19-27; Imaging protein molecules using FRET and FLIM microscopy; Ai et al (2008) *Nature Methods* 5, 401-403 Fluorescent protein FRET pairs for ratiometric imaging of dual biosensors; Shaner et al. (2004) Nat Biotechnol 22: 1567-1572 Improved monomeric red, orange and yellow fluorescent proteins derived from *Discosoma* sp. red fluorescent protein.

The compounds identified in the methods may themselves be useful as a drug or they may represent lead compounds for the design and synthesis of more efficacious compounds.

The compound may be a drug-like compound or lead compound for the development of a drug-like compound for each of the above methods of identifying a compound. It will be appreciated that the said methods may be useful as screening assays in the development of pharmaceutical compounds or drugs, as well known to those skilled in the art.

The term "drug-like compound" is well known to those skilled in the art, and may include the meaning of a compound that has characteristics that may make it suitable for use in medicine, for example as the active ingredient in a medicament. Thus, for example, a drug-like compound may be a molecule that may be synthesised by the techniques of organic chemistry, less preferably by techniques of molecular biology or biochemistry, and is preferably a small molecule, which may be of less than 5000 daltons. A drug-like compound may additionally exhibit features of selective interaction with a particular

protein or proteins and be bioavailable and/or able to penetrate cellular membranes, but it will be appreciated that these features are not essential.

The term "lead compound" is similarly well known to those skilled in the art, and may include the meaning that the compound, whilst not itself suitable for use as a drug (for example because it is only weakly potent against its intended target, non-selective in its action, unstable, difficult to synthesise or has poor bioavailability) may provide a starting-point for the design of other compounds that may have more desirable characteristics.

It will be understood that it will be desirable to identify compounds that may modulate the endonuclease activity or 5'exonuclease activity of a FAN1 polypeptide *in vivo*. Thus it will be understood that reagents and conditions used in the method may be chosen such that the interactions between, for example, the FAN1 polypeptide and a substrate nucleic acid, are substantially the same as between the human FAN1 and an endogenous human substrate nucleic acid. Typically a method of the invention may be performed in a human cell-based system, optionally expressing human recombinant polypeptides.

The compounds that are tested in the screening methods of the invention or in other assays in which the ability of a compound to modulate the endonuclease activity or 5'exonuclease activity of a FAN1 polypeptide, may be measured, may be (but do not have to be) compounds that have been selected and/or designed (including modified) using molecular modelling techniques, for example using computer techniques. The selected or designed compound may be synthesised (if not already synthesised) and tested for its effect on the FAN1 polypeptide, for example its effect on the endonuclease activity or 5'exonuclease activity. The compound may be tested in a screening method of the invention.

The compounds that are tested may be compounds that are already considered likely to be able to modulate the activity of an endonuclease or 5'exonuclease; or may be compounds that have not been selected on the basis of being likely to modulate the activity of an endonuclease or 5'exonuclease. Thus, the compounds tested may be compounds forming at least part of a general, unselected compound bank; or may alternatively be compounds forming at least part of a pre-selected compound bank, for example a bank of compounds pre-selected on the basis of being considered likely to modulate the activity of an endonuclease or 5'exonuclease.

It will be appreciated that screening assays which are capable of high throughput operation will be particularly preferred. Examples may include cell based assays and protein-protein binding assays.

- 5 As will be apparent to those skilled in the art, it may be desirable to assess what effect the compound has on other endonucleases or 5'exonucleases.

The screening methods of the invention may comprise the step of synthesising, purifying and/or formulating the selected compound. The compound may be formulated for
10 pharmaceutical use, for example for use in *in vivo* trials in animals or humans.

A further aspect of the invention provides a purified preparation or kit of parts comprising a FAN1 polypeptide or polynucleotide; and a substrate polynucleotide as defined above and/or a FANCD2 or FANCI polypeptide.

15

It is envisaged that suitable kits for performing *in vitro* assays may comprise a FAN1 polypeptide or polynucleotide, a fluorescently labelled DNA flap substrate, and a suitable buffer comprising divalent cations, such as Mg^{2+} , Mn^{2+} , Zn^{2+} , Co^{2+} , Cu^{2+} and/or Ca^{2+} .

- 20 Such DNA flap substrates may be assembled by annealing oligonucleotides, for example where one oligonucleotide comprises label, such as a chromophore or dye.

The preparation or kit may comprise a recombinant FAN1 polynucleotide.

- 25 The kit of parts may further comprising ubiquitin or dithiobis (succinimidyl propionate) (DSP) or a de-ubiquitinase inhibitor, for example N-ethyl maleimide (NEM).

By "purified" is meant that the preparation has been at least partially separated from other components in the presence of which it has been formed, for example other
30 components of a recombinant cell. Examples of methods of purification that may be used are described in the Examples.

The preparation may be substantially pure. By "substantially pure" we mean that the said polypeptide(s) are substantially free of other proteins. Thus, we include any
35 composition that includes at least 2, 3, 4, 5, 10, 15, 20 or 30% of the protein content by weight as the said polypeptides, preferably at least 50%, more preferably at least 70%,

still more preferably at least 90% and most preferably at least 95% of the protein content is the said polypeptides.

Thus, the invention also includes compositions comprising the said polypeptides and a
5 contaminant wherein the contaminant comprises less than 96, 95, 94, 90, 85, 80 or 70% of the composition by weight, preferably less than 50% of the composition, more preferably less than 30% of the composition, still more preferably less than 10% of the composition and most preferably less than 5% of the composition by weight.

10 The invention also includes the substantially pure said polypeptides when combined with other components *ex vivo*, said other components not being all of the components found in the cell in which said polypeptides are found.

A further aspect of the invention provides a mutant FAN1 polypeptide or polynucleotide
15 wherein Asp 981 and/or Arg982 are mutated, optionally to alanine; and/or wherein Cys44 and/or Cys47 are mutated, optionally to alanine.

A further aspect of the invention provides the use of a FAN1 polypeptide in an *in vitro* method of resolving a DNA ICL.
20

A further aspect of the invention provides an *in vitro* method of resolving a DNA ICL, the method comprising the step of using a FAN1 polypeptide.

The screening method of the invention may comprise the step of assessing whether the
25 compound modulates the repair of a DNA inter-strand crosslink, resolution of an ICL-induced double strand break, or homologous recombination in a cell, or in increasing cancer cells' sensitivity to ICLs or enhancing the effect of platinum based chemotherapy or other ICL-inducing therapy on cancer cells, in a whole cell, tissue or organism; and a compound that modulates the activity or therapeutic response is selected.

30 The method may further comprise the step of assessing whether the compound modulates the activity of a FAN1 polypeptide in the whole cell, tissue or organism, and a compound that modulates the activity is selected.

35 The method may further comprise the step of synthesising, purifying and/or formulating the selected compound.

A further aspect of the invention provides a method for preparing a compound which modulates the activity of a FAN1 polypeptide, the method comprising 1) performing a method according to the invention and 2) synthesising, purifying and/or formulating the selected compound.

5

A further aspect of the invention provides a method of characterising a patient with cancer and/or suspected Fanconi Anaemia, the method comprising the step of assessing the patient's genotype or phenotype for FAN1. The method may include the step of assessing the expression level of FAN1 polynucleotide or polypeptide in the patient. The method of characterising a patient with cancer or suspected Fanconi Anaemia may further comprise the step of assessing the patient's genotype or phenotype for SLX4 and/or SLX1. It is envisaged that such methods may aid in the diagnosis of cancer or Fanconi Anaemia. It is further envisaged that an appropriate treatment regime may be chosen based on the genotype or phenotype identified above.

15

The invention will now be described in more detail by reference to the following, non-limiting, figures and examples.

Any references referred to herein are hereby incorporated by reference.

20

Figure legends

Figure 1. The KIAA1018/MTMR15/FAN1 family of proteins. A. Schematic representation of the domain architecture of KIAA1018/MTMR15/FAN1 orthologues from different species. The relevant protein identification codes are: *Homo sapiens* Q9Y2M0; *Danio rerio* Q1LWH4; *Caenorhabditis elegans* P90740; *Schizosaccharomyces pombe* Q9Y804; *Arabidopsis thaliana* Q9SX69; *Oryza sativa* B9FRR6; *Pseudomonas aeruginosa* Q9I2N0. B. Alignment of the VRR_nuc domain of FAN1. Identical residues are shaded in black, similar residues are shaded in grey. The asterisks denote conserved residues Asp981 and Arg982 mutated in the FAN1-DR mutant. (SEQ ID NOS: 23-29) C. Alignment of the UBZ domain of FAN1. Identical residues are shaded in black, similar residues are shaded in grey. The conserved Cys and His residues that define the two dyads of the UBZ domain are shaded in red. The asterisk denotes the conserved Cys44 and Cys47 residues in the first dyad. (SEQ ID NOS: 30-48)

35

Figure 2 FAN1 has structure-specific endonuclease activity

Recombinant human FAN1 was incubated with synthetic DNA structures: splayed duplex (SD; oligos a3, b), 3' flap (3'F; oligonucleotides a3, b, d3), 5' flap (5'F; oligos a3, b, c) or

a replication fork (RF)-like structure (oligos a3, b, c, d3), each radioactively 5'-³²P-labelled on the strands indicated. WT refers to wild-type FAN1 and DR refers to the Asp981Ala-Arg982Ala FAN1 mutant. **A.** Schematic diagram of the DNA substrates used in A. Sites of DNA cleavage are indicated by arrows. (SEQ ID NOS: 49 and 50). **B.** Reaction products (10 min incubation) were subjected to denaturing PAGE. Purine-specific chemical sequencing ladders (R) were derived from oligonucleotides a3 or b. **C.** FAN1 was incubated with the 5' flap shown in A. for the time indicated (s, seconds) and reaction products were subjected to denaturing PAGE. **D.** Progress curves of cleavage of the 5' flap construct incubated with wild type (black squares) and DR (red triangles) mutant FAN1. The data have been fitted to a single (DR) or double (wild type) exponential functions (lines). From these data we have calculated observed rates of cleavage of $> 0.2 \text{ s}^{-1}$ and 0.0003 s^{-1} for wild type and DR enzymes respectively. See also Figure 9.

Figure 3 FAN1 has 5' exonuclease activity

Recombinant human FAN1 was incubated for the time indicated (s, seconds) with dsDNA (oligonucleotide a3, a3-cp), ssDNA (oligonucleotide a3) or with a 5' flap (5'F; oligonucleotides a3, b, c) radioactively 5'- or 3'-³²P-labelled on the a3 strand as shown (asterisks). WT refers to wild-type FAN1 and DR refers to the Asp981Ala-Arg982Ala FAN1 mutant. Reaction products were subjected to denaturing PAGE. **A.** Cleavage of linear DNA substrates. **B.** The cleavage products were quantitated. "Fraction DNA cut" is the ratio of the relevant cleavage product to total DNA (cleaved plus uncleaved DNA). The data are plotted as a function of time, and fitted to single or double exponential functions. **C.** Activity of WT FAN1 on radioactively 3'-³²P-labelled 5' flap. R refers to a purine-specific chemical sequencing ladder derived from the labelled strand. See also Figure 10.

Figure 4 FAN1 interacts with DNA repair proteins

A. HEK293 Flp-In cells that stably express GFP-FAN1 were lysed in the presence of dithiobis (succinimidyl propionate) (DSP) or N-ethyl maleimide (NEM). These extracts together with extracts of cells that express GFP only were subjected to immunoprecipitation with GFP-Trap beads and after extensive washing precipitates were subjected to SDS-PAGE. The gel was fixed and stained with Colloidal Blue. The gel lane containing FAN1-associated proteins was cut into slices, as indicated, and the proteins were digested with trypsin before mass spectrometric fingerprinting. **B.** HEK293 Flp-In cells that stably express GFP-FAN1 were lysed in the presence or absence of DSP and extracts were subjected to immunoprecipitation with control anti-HA (IgG) or

GFP-Trap beads. Precipitates were analysed by western blotting with the antibodies indicated. Input represents 4% of the extract used for immunoprecipitation. **C.** HEK293 cells were lysed in the presence or absence of DSP and extracts were subjected to immunoprecipitation with anti-HA (IgG) or with anti-FAN1 antibodies. Precipitates were analysed by western blotting with the antibodies indicated. Input represents 4% of the extract used for immunoprecipitation. **D.** Extracts of HEK293 cells were analysed by size exclusion chromatography on a HiLoad 26/60 Superdex 200 column in buffer containing 0.2 M NaCl, and every third fraction was denatured and analysed by western blotting with the indicated antibodies. The elution positions of Dextran blue (2 MDa), thyroglobulin (670 kDa) and bovine γ -globulin (158 kDa) are shown.

See also Fig. 15.

Figure 5 The UBZ domain of FAN1 binds mono-ubiquitinated FANCD2

A. HEK293 cells were transiently transfected with pcDNA5.1-GFP-FAN1 wild type (WT) or pcDNA5.1-GFP-FAN1-UBZ* (Cys44A/Cys47A). After 48 h, cells were lysed in the presence or absence of DSP and anti-GFP precipitates were analysed by western blotting with the antibodies indicated. "Input" represents cell extracts. **B.** U2OS cells, grown on glass coverslips, were transiently transfected with pcDNA5.1-GFP, pcDNA5.1-GFP-FAN1 wild type (WT) or pcDNA5.1-GFP-FAN1-UBZ* (Cys44A/Cys47A). Cells were treated, or not, with MMC and after 16 h GFP-FAN1 foci and FANCD2 foci were detected. **C.** Quantitation of data from B. The number of cells with >5 GFP-FAN1 foci in a sample of 500 cells were counted. **D.** U2OS cells were transfected with control siRNA, or siRNA targeting FAN1 and the number of cells with >5 FANCD2 foci were quantitated after exposure of cells to MMC. **E.** HEK293 cells stably expressing GFP-FAN1 were transiently transfected with pDEST40-lacZ ("-"), pDEST40-V5-FANCD2 wild-type (WT) or pDEST40-V5-FANCD2 K561R. After 48 h, cells were lysed in the presence of DSP and anti-GFP precipitates were analysed by western blotting with the antibodies indicated. **F.** FANCD2 -/- (PD20) cells stably transfected with empty vector (-), FANCD2 wild-type (WT) or FANCD2 K561 R were transiently transfected with GFP-FAN1. Cells were treated with MMC for 18 h and then fixed and GFP-FAN1 foci were visualised. **G.** Same as F. except that cells were treated with MMC for the times indicated and the number of cells with >5 FAN1 foci were quantitated after exposure of cells to MMC. Data in C, D, G are represented as mean $\pm$ SEM.

See also Figure 11.

Figure 6 FAN1 is required for DNA repair

A. HEK293 cells were transfected with the siRNAs indicated. Clonogenic survival assays were carried out with cisplatin or mitomycin-C (see Experimental Procedures). For each siRNA, cell viability of untreated cells is defined as 100%. B. Synchronized L1 larva stage animals of the relevant genotype were incubated with the indicated concentrations of nitrogen mustard (HN_2) or cisplatin. After 48 h, the extent of developmental progression of the worms was scored, by counting the number of worms in the adult and various larval stages (L1-L2, L3-L4). Adult worms were scored as fertile if they contained fertilized eggs and as sterile if they did not. C. HEK293 cells transfected with control siRNA or FAN1 siRNA (FAN1-1) were treated with cisplatin (1 mg/ml) for 2 h and then allowed to recover for the times indicated. The proportion of cells in each population with more than two γ -H2AX foci at each timepoint (" γ -H2AX positive") was determined. The experiment was done three times and a representative experiment is shown. D. The frequency of chromosome breaks and radial chromosomes in metaphase spreads of HEK293 cells transfected with control siRNA or FAN1-1 siRNA was measured before and after exposure to MMC (10 ng/ml; 2 h) or to IR (3 Gy; 18h) was measured as described previously (Deans and West, 2009). Data in A, C are represented as mean $\pm$ SEM.

See also Figure 12.

Figure 7 Effect of FAN1 depletion on focus formation by RAD51 and RPA

A. U2OS cells, in which an 18 bp sequence recognized by I-SceI was placed between tandem mutant copies of the gene encoding GFP, were transfected with control siRNA (luciferase) and/or siRNAs specifically targeting FAN1 (FAN-1 or FAN-2) or RAD51. After 48 h, cells were transfected with a plasmid expressing I-SceI or with an empty vector, and 24 h later, cells were tested for GFP expression by FACS analysis. The frequency of HR in cells transfected with the various siRNAs was calculated relative to cells transfected with control siRNA. B. U2OS cells transfected with control siRNA or FAN1 siRNA (FAN1-1) were treated with cisplatin (1 mg/ml) for 2 h and then allowed to recover for the times indicated. Cells were then fixed, permeabilised, washed and blocked before incubation with anti-RPA antibodies. Coverslips were incubated with secondary antibodies, mounted on glass slides and visualised. The average number of RPA foci per cell was analysed. The experiment was done three times and a representative experiment is shown. C. Same as B. except that cells were stained with anti-RAD51 antibodies. D. Same as C. except that cells were exposed to IR (3 Gy) and then allowed to recover for the times indicated. Data represented as mean $\pm$ SEM.

See also Figure 13.

Fig. 8 Model for repair of inter-strand crosslinks (ICLs), Related to Introduction

It was proposed that ICL repair is initiated by the convergence of two replication forks on the ICL. Various signaling proteins including the Fanconi Anemia (FA) complex are recruited to the vicinity of the blocked replisome. This triggers the mono-ubiquitination of FANCD2 and its paralogue FANCI at Lys561 and Lys523 respectively, that in turn directs subsequent steps of ICL repair by unknown mechanisms. Cleavage of the leading strand template of one of the forks by MUS81 in concert with cleavage of the same strand on the opposite side of the ICL by a second nuclease would unhook the ICL. This results in two one-ended DSBs and a gapped duplex with the ICL adduct on one strand that is filled in by translesion synthesis. The ICL adduct is excised and resection of one of the DSBs generated by unhooking initiates homologous recombination that completes repair.

Fig. 9 Purification and assay of FAN1, Related to Figure 2

A. Coomassie gel of the different stages of the purification of recombinant human NUS-His₅-FAN1 from bacterial cell extracts, as described in Supplemental Material. Samples from the indicated stages in NUS-His₅-FAN1 purification were denatured and subjected to SDS-PAGE and Coomassie staining. **B.** Recombinant human FAN1 was incubated with synthetic DNA structures: splayed duplex (SD), 3' flap DNA (3'F), 5' flap (5'F) or a replication fork (RF)-like structure, each radiolabeled on the strands indicated. Note that f9 was 5' end-labelled while f10 was 3' end-labelled. WT refers to wild-type FAN1 and DR refers to the Asp981Ala-Arg982Ala FAN1 mutant. Reaction products were subjected to denaturing PAGE (12% polyacrylamide, 8 M urea). **C.** Schematic diagram of the DNA substrates used in B. (SEQ ID NOS: 51 and 52).

Fig. 10 Exonuclease activity of FAN1 on recessed DNA ends, Related to Figure 3

Recombinant human FAN1 was incubated with dsDNA (oligos a3-cp and a3-l25) with a recessed end in which the shorter strand (a3-l25) on the 5' end (A) or 3' end (B). WT refers to wild-type FAN1 and DR refers to the Asp981Ala-Arg982Ala FAN1 mutant. Reaction products were subjected to denaturing PAGE (12% polyacrylamide, 8 M urea). Asterisks denotes the DNA end that was labelled.

Fig. 11 FAN1 forms subnuclear foci in response to genotoxins, Related to Figure 5

A. U2OS cells were exposed to MMC (10 ng/ml) for 18h, or left untreated. Cells were fixed, permeabilised, washed and blocked before incubation with sheep anti-FAN1 antibodies generated as part of this study, or with anti-FANCD2 antibodies. After washing, coverslips were incubated with secondary antibodies, stained with DAPI, mounted on glass slides and visualised. **B.** U2OS cells, grown on glass coverslips, were

transiently transfected with pcDNA5.1-GFP-FAN1. Cells were treated, or not, with MMC and after 18h cells were fixed, permeabilised, washed and blocked before incubation with anti-GFP or anti-FANCD2 antibodies.

Fig. 12 Effect of FAN1 depletion on cell sensitivity to genotoxins and on ubiquitination of FANCD2 and FANCI, Related to Figure 6

A. HEK293 cells were transfected with control siRNA (CON) or siRNA specifically targeting FAN1 (REND-1 or REND-2) or FANCA. After 48 h, cell extracts were subjected to western blotting with the indicated antibodies. **B.** HEK293 cells were seeded in 10 cm² dishes at 25% confluence and allowed to adhere overnight. Cells were transfected with the relevant siRNA for 48 h and cells were split and seeded in 10 cm² dishes (5000 cells/dish). Cells were allowed to adhere for 8 h and exposed to the indicated doses of IR or UV before being incubated at 37°C for 10-14 days. Alternatively, genotoxins were added at the concentrations indicated for 24 h: camptothecin, 6-thioguanine or hydroxyurea. Cells were then washed free of genotoxin and incubated in fresh medium for 10-14 days before the number of colonies of >50 cells in each dish were counted. For each siRNA, cell viability of untreated cells is defined as 100%. Data represented as mean $\pm$ SEM. **C.** HEK293 cells were transfected with the control (CON) siRNA or siRNAs targeting FAN1 or FANCA. After 48 h, cells were treated with cisplatin for the indicated times and lysed. Cell extracts were subjected to western blotting with antibodies against FANCD2 (left panels) or FANCI (right panels).

Fig. 13. Effect of FAN1 depletion on SCE, Related to Figure 7

SCEs were scored for 60 mitotic spreads for HEK293 cells transfected with the relevant siRNA (D2, FANCD2; F1, FAN1-1) in the presence the indicated dose of MMC. Each point represents the total number of SCEs in a single mitotic spread.

Fig. 14 Models for action of FAN1 during ICL repair, Related to Discussion

A. It has been proposed that initiation of ICL repair requires the convergence of two replication forks at an ICL. In this scenario, cleavage of the leading strand at one of the stalled forks (corresponding to a 3' flap-like structure) by MUS81-EME1 and cleavage of the same strand on the opposite side of the ICL (corresponding to a 5' flap-like structure) by an unknown nuclease would unhook the ICL. The unknown nuclease could be FAN1 but, as discussed in the text, the available data question this possibility. Unhooking results in two one-ended DSBs. Excision repair removes the crosslink adduct and translesion synthesis fills in the gap. The ends generated by unhooking are resected by a 5' exonuclease, generating 3' overhangs and one of these overhangs invades the

complementary duplex. It is unlikely that FAN1 is involved in DSB resection because RPA and RAD51 foci form normally in cells depleted of FAN1. After strand extension, the invading strand re-anneals to the complementary strand in the duplex it came from originally. It is possible that continued DNA synthesis on the parent strand displaces the DNA in front of it. This would generate a 5' flap that could be cleaved by FAN1. **B.** It may be that MUS81 cleaves the leading strand template on both of the forks that converge on the ICL. This would generate two one-ended DSBs, and a gapped duplex containing the ICL that could be filled in by translesion synthesis. Unhooking and excision of the ICL from the linear duplex would generate a template for strand invasion. Repair is then completed by HR. In this context it is unlikely that FAN1 would be involved in unhooking since it would have to cleave a linear duplex containing an ICL, that does not resemble the 5' flap that FAN1 prefers. **C.** It is possible that it is not always necessary for two forks to converge on an ICL for ICL repair and fork regeneration to occur. In some cases repair may be initiated by collision of a single replisome with an ICL followed by MUS81-catalysed generation of a one-ended DSB. It is possible that the 5' endonuclease activity of FAN1 makes an incision in the D-loop created by strand invasion, allowing the resulting 5' end to be ligated to the invading strand that was extended. This would regenerate an intact replication fork from which DNA replication can continue.

Fig. 15. GFP-FAN1-associated proteins identified by mass spectrometry, Related to Figure 4

Gels slices in Fig. 4A were excised and proteins were identified by mass spectrometric fingerprinting. Gel slices 9-16 came from the gel lane containing GFP-FAN1 precipitates from cells lysed in the presence of DSP (Fig. 4A). Gel slices 17-24 came from the gel lane containing GFP-FAN1 precipitates from cells lysed in the presence of NEM (Fig. 4A). Only proteins that were also present in FLAG-FAN1 precipitates (data not shown), and that were not present in GFP-only precipitates, are listed.

Example 1: KIAA1018/FAN1 (FAN1) is a flap endonuclease involved in DNA repair that is recruited to DNA damage by mono-ubiquitinated FANCD2

DNA inter-strand crosslinks (ICLs) are highly toxic because they block the progression of replisomes. The Fanconi Anemia (FA) proteins, encoded by genes that are mutated in FA, are important for repair of ICLs. The FA core complex catalyses the mono-ubiquitination of FANCD2 and this event is essential for several steps of ICL repair. However, how mono-ubiquitination of FANCD2 promotes ICL repair at the molecular

level is unknown. Here we describe a novel and highly conserved protein, KIAA1018/MTMR15/FAN1, that interacts with, and is recruited to sites of DNA damage by, the mono-ubiquitinated form of FANCD2. FAN1 exhibits endonuclease activity towards 5' flaps and has 5'-exonuclease activity and these activities are mediated by an ancient VRR-nuc domain. Depletion of FAN1 from human cells causes hypersensitivity to ICLs, defects in ICL repair and genome instability. These data at least partly explain how ubiquitination of FANCD2 promotes DNA repair.

EXPERIMENTAL PROCEDURES

General Methods

Gel filtration and analysis of the resolution of cisplatin-induced DSBs were carried out as described previously (Munoz et al., 2009). DSP (dithiobis (succinimidyl propionate; Pierce) is a homobifunctional and thiol-cleavable crosslinker that was used according to the manufacturers instructions. DSP was included in lysis buffer at 2.5 mg/ml and lysates were incubated for 30 min on ice. Excess DSP was quenched by adding 75 μ l 1 M Tris-HCl (pH 7.4) followed by an additional 30 min incubation. Crosslinks were reversed by the inclusion of dithiothreitol in SDS-PAGE sample buffer added to cell extracts or immunoprecipitates before electrophoresis. Details of immunofluorescence are given in Supplementary Information.

Antibodies, cell lysis and immunoprecipitation

The primary antibodies used in this study were the following: FAN1 (this study; sheep S420C, 4th bleed), MLH1 (BD Pharminigen, 554073), PMS2 (Santa Cruz, sc-617), PCNA (Santa Cruz, PC10), FANCI (Bethyl, A301-354), FANCD2 (Abeam, ab21 87-50), FANCD2 (Novus, NB100-182; immunofluorescence), FANCA (Cascade Biosciences, abm6202), FANCC (Cascade Biosciences, abp6305), FANCE (a kind gift from KJ Patel), FANCF, FANCG (kind gifts from Johan De Winter), FLAG (Sigma, M2), Ku80 (Cell Signaling, 2753), RAD51 (Santa Cruz, H-92), RPA70 (Cell Signaling, 2267) and anti- γ -H2AX (Bethyl, A300-081A). The FAN1 antibody, raised in sheep against full-length FAN1 fused to GST, at the Scottish Antibody Production Unit (SAPU; Carluke, Lanarkshire) was affinity purified using immobilised antigen. GFP-Trap beads were from Chromotek. Protein G Sepharose was from GE Healthcare. Cells were lysed in ice-cold buffer: [40 mM HEPES (pH 7.4), 120 mM NaCl, 1% (v/v) Triton X-100, 1 mM EDTA] with protease inhibitors (Roche). To visualize mono-ubiquitinated forms of FANCI and FANCD2, 0.5 U/ml of benzonase (Sigma) was included in the lysis buffer and lysates were incubated on ice for 30 min. All immunoprecipitations were carried out in lysis buffer for 1 h at 4°C.

Endogenous immunoprecipitations were carried out using 2 µg of FAN1 antibody coupled to 10 µl protein G sepharose per 4 mg of whole cell extract.

Purification of GFP-FAN1 from HEK293 cells

- 5 FlpIn T-Rex cells (Invitrogen) cells stably expressing GFP-FAN1 in a tetracycline-inducible manner were made according to the manufacturers instructions with FAN1 in plasmid pcDNA5-FRT-TO-GFP-FAN1. FAN1 was induced and purified according to a previously-described protocol (Munoz et al., 2009).

10 siRNA

Cells were transfected with the relevant siRNA duplex (100 nM) using the calcium phosphate precipitation method. In Fig. 7 U2OS cells were transfected in 96-well plates using siRNAs at a concentration of 20 nM and DharmaFECT 1 (Dharmacon) at a 1:1000 concentration. Cells were incubated at 37°C for 48 h. The mRNA target sequences used for siRNAs were:

- 15 FANCA (GGGUCAAGAGGGAAAAUA) (SEQ ID NO: 7), FAN1-1 (GUAAGGCUCUUUCAACGUA; exon 3) (SEQ ID NO: 8), FAN1-2 (GCAGGAAGGCAGAGUGGCU; exon 12) (SEQ ID NO: 9), MLH1 (GCAUGUGGCUCAUGUUAC) (SEQ ID NO: 10), ATR
20 (GGGAGCCUGUUGAGACAAGAU) (SEQ ID NO: 11), FANCD2 (siGenome SMARTPool from Dharmacon).

Oligonucleotides

- a3: 5'-CCTCGATCCTACCAACCAGATGACGCGCTGCTACGTGCTACCGGAA
25 GTCG (SEQ ID NO: 12)
b: 5'-CGACTTCCGGTAGCACGTAGCAGCGGCTCGCCACGAACTGCACTCTA
GGC (SEQ ID NO: 13)
c: 5'-GCCTAGAGTGCAGTTCGTGGCGAGC (SEQ ID NO: 14)
d3: 5'-CGTCATCTGGTTGGTAGGATCGAGG (SEQ ID NO: 15)
30 a3-cp: 5'-CGACTTCCGGTAGCACGTAGCAGCGCTCAACTGGTTGGT
AGGATCGAGG (SEQ ID NO: 16)

Preparation of DNA substrates

- All substrates and standards were annealed by slow cooling of one radioactively 5'-³²P-labelled oligonucleotide with the relevant unlabelled one(s). In Fig. 2 these were: splayed duplex (SD), a3, b; 3' flap (3'F) a3, b, d3; 5' flap (5'F) a3, b, c; replication fork analogue (RF), a3, b, c and d3. In Fig. 3 these were a3 and a3-cp (dsDNA), s3 (ssDNA). Synthetic

structures were then purified by electrophoresis on a native 8% polyacrylamide gel and recovered by the crush and soak method followed by ethanol precipitation.

Nuclease assays

5 Purified recombinant FAN1 (35nM) was pre-incubated for at least 10min with radiolabeled DNA substrates (5nM) at 37°C in 25 mM Tris-HCl pH 7.5, 10 mM NaCl, 15 mM KCl and 0.1 mg/ml BSA to allow binding to occur. The reaction was started by the addition of 1 mM MnCl₂ and stopped by the addition of 2 mM EDTA. The samples were then boiled at 95°C for 10 min and analysed by denaturing PAGE (15%
10 polyacrylamide and 8M urea). Gels were dried, exposed to storage Phosphor screens and analysed with the ImageGauge software (Fujifilm). For kinetics experiments, the data were plotted as the fraction of DNA in the relevant product bands as a function of time, and fitted to either one or two exponential functions. When two exponential functions were used, the rate given is the faster of the two.

15

Clonogenic survival assays

HEK293 cells were seeded in 10 cm² dishes at 25% confluence and allowed to adhere overnight. Cells were transfected with the relevant siRNA for 48 h and cells were split and seeded in 10 cm² dishes (5000 cells/dish). Cells were allowed to adhere for a
20 minimum of 8 h before cisplatin or mitomycin-C were added at the indicated concentrations for 24 h. Cells were then washed free of drugs and incubated in fresh medium for 10-14 days before the number of colonies of >50 cells in each dish were counted.

25 *C. elegans* genotoxin sensitivity assays

Worms were maintained at 20°C on NGM (Nematode Growth Media) agar plates according to standard protocols (Brenner, 1974). Alleles used were: *fcd-2* (tm1298) and C01G5.8 (tm423). The C01G5.8 mutant was generated and kindly provided by Shoehi Mitani of the National Bioresource Project for the Nematode, Japan. This strain, which
30 has a 411 bp deletion in C01G5.8 that removes exons 6-8, was outcrossed five times with N2 Bristol strain (wild-type) to eliminate secondary mutations. To access ICL sensitivity, synchronized L1 larval stage animals of the relevant genotype were incubated at 20°C for 16 h in 1 ml S-basal buffer (0.1M NaCl, 0.05M KH₂PO₄, pH 6.0, 5 mg/ml cholesterol) containing *E. coli* OP50 and the indicated concentration of nitrogen mustard
35 (HN₂) or cisplatin. After incubation worms were transferred to OP50-seeded NGM plates. After 48 h, the extent of developmental progression was scored. In each

experiment, a minimum of 60 worms was scored and the results shown are the average of three independent experiments.

GFP HR assay

- 5 Cells were transfected with siRNA in 96-well dishes and after 48 h cells, were transfected with 0.25 mg I-Sce-I vector and 0.2 mg PEI in 150 µl OptiMEM/well. GFP-positive cells were analyzed using FACS 48 h after I-Sce-I transfection as previously described (Pierce et al. 1999).

10 Expression and purification of recombinant FAN1

- pET43.1a(+) expressing NUS-His₆-FAN1 or NUS-His₆-FAN1 D981A R982A was transformed into BL21 bacterial cells. FAN1 expression was induced by the addition of IPTG (50 µM) to 0.5 L of bacteria in early exponential phase in liquid culture followed by overnight incubation at 18°C. Cells were then heat shocked at 42°C for five minutes
15 before centrifugation (4000 g, 20 mins, 4°C) and the pellet was snap frozen and thawed. Bacterial pellets were lysed in lysis buffer (25 mM HEPES pH 7.6, 10% glycerol (v/v), 0.1 mM EDTA, 150 mM KCl, 1 mM DTT, 1 mM PMSF, protease inhibitors (Roche)). Lysates were incubated on ice at 4 °C for 30 min before centrifugation (18,000 rpm, 20 mins, 4°C). The supernatant was then subjected to ultracentrifugation (37,500 rpm, 1 h, 4°C)
20 before being loaded onto a DEAE column (Amersham) that was developed with a linear gradient of KCl in buffer A (25 mM HEPES pH 7.6, 10% glycerol (v/v), 0.1 mM EDTA, 1 mM DTT, ddH₂O) using an FPLC LCC-501 (Pharmacia biotech). The flow rate was 2 ml/min and 3 ml fractions were collected. Samples from every fourth fraction were denatured and subjected to SDS-PAGE followed by Coomassie blue staining. Fractions
25 containing FAN1 were combined and diluted 1:1 in buffer A and loaded onto an SP column (Amersham) that was developed with a linear gradient of KCl in buffer A. The flow rate was 0.5 ml/min and 0.25 ml fractions were collected. Samples from every third fraction were denatured and subjected to SDS-PAGE followed by Coomassie blue staining. Fractions containing FAN1 were diluted 1:1 in buffer A and loaded onto a
30 Heparin column (Amersham) that was developed with a linear gradient of KCl in buffer A. The flow rate was 0.5 ml/min and 0.25 ml fractions were collected. Samples from every second fraction were subjected to SDS-PAGE and FAN1 containing fractions were combined. Protein concentration was estimated by subjecting FAN1 to SDS-PAGE and Coomassie blue staining in parallel with known amounts of BSA.

35

Oligonucleotides use in Figs. 9 and 10

f7: 5'-ATTGACTAGGTTACATGACTGAATGATAGT (SEQ ID NO: 17)

f8: 5'-GGAGTAAAGTACTAGGTATGTGCGACATTGA (SEQ ID NO: 18)

f9: 5'-ACTATCATTCAATCATGTAACCTAGTCAATCTGCGAGCTCGAATTCAC
TGGAGTGACCT (SEQ ID NO: 19)

5 f10: 5'-GAGGTCACTCCAGTGAATTCGAGCTCGCAGTCAATGTGCGACATAC
CTAGTACTTTACTCC (SEQ ID NO: 20)

a3-cp: 5'-CGACTTCCGGTAGCACGTAGCAGCGCGTCAACTGGTTGGTAGG
ATCGAGG (SEQ ID NO: 21)

a3-l25: 5'-CGCTGCTACGTGCTACCGGAAGTCG (SEQ ID NO: 22)

10

DNA substrates were prepared with the following oligos. Fig. 9:- splayed duplex (SD), f9, f10; 3' flap (3'F) f7, f9, f10; 5' flap (5'F) f8, f9, f10; replication fork analogue (RF), f7, f8, f9, f10. Fig. 10:- a3-l25 and a3-cp (recessed 5' ends).

15 Sister chromatid exchange assay

Cells were cultured in 5 µg/ml BrdU for 40 h followed by 1 h in 0.1 µg/ml colcemid. MMC was added 20 h prior to cell harvest where indicated. Exchanges that were obviously due to "flipping" at the centrosome were omitted from quantitation. SCEs were scoring on 60 metaphases for each genotype/treatment. All quantification was carried out blind.

20

Immunofluorescence - RAD51 and RPA

U2OS cells transfected with the relevant siRNA were treated with cisplatin (1 mg/ml) for 2 h or irradiated with 3 Gy and then allowed to recover for times indicated before fixation with 4% paraformaldehyde in PBS-Triton-X100 (0.1%). After three PBS washes and
25 blocking in 3% BSA in PBS, cells were incubated with primary antibody overnight. Another three PBS washes were followed by a 1 h incubation with secondary antibody at room temperature. DNA was stained with DAPI at 1 µg/ml and images acquired using an IN Cell Analyzer 1000 (GE Healthcare). Images were analyzed with IN Cell Analyzer 1000 Workstation software using the Multi Target Analysis module (GE Healthcare).

30

Immunofluorescence - GFP-FAN1, FANCD2 and γ-H2AX

HEK293 or U2OS cells transfected with the relevant siRNA were mock-treated, treated with 80 ng/ml MMC or 0.5 mM HU and fixed 18 h later or treated with 5 Gy IR and fixed 1 h later with 2% PFA for 10 min at room temperature. Cells were then permeabilised with
35 0.2% Triton X-100 in PBS for 10 min at room temperature. For indirect immunofluorescence analysis of FANCD2, γ-H2AX or endogenous FAN1 cells were washed several times in PBS and incubated in blocking solution (PBS containing 3% IgG

-free BSA (Jackson ImmunoResearch) and 0.2% Tween 20) for 1 h. Coverslips were then incubated with primary antibodies (1 mg/ml) in blocking solution for 1 h. After extensive washing in PBS-T (PBS containing 0.2% Tween20), coverslips were incubated with secondary antibodies (2 mg/ml) conjugated to Texas Red for 45 min.

5 Coverslips were washed thoroughly in PBS - T and mounted on glass slides. Before covering with a cover slip cells were stained with DAPI-Hydromount for 5 min. Slides were viewed using a Deltavision DV3 widefield deconvolution microscope mounted on a Nikon Diaphot inverted microscope and images were deconvolved after acquisition. For GFP tagged proteins, the protocol is the same except that no blocking or antibody
10 incubation takes place.

Results

Domain organisation of KIAA1018/MTMR15/FAN1

We noticed an uncharacterised human protein, KIAA1018/MTMR15, in the human
15 sequence databases, that has a UBZ-type ubiquitin—binding domain domain, a SAP-type DNA binding domain and a putative nuclease domain termed "VRR_nuc" domain (Fig. 1A), initially referred to as "domain of unknown function 994" (DUF994) (Iyer et al., 2006). Orthologues of KIAA1018 are found in prokaryotes and most eukaryotes with the notable exception of budding yeast (Fig. 1A).

20 We suspected that KIAA1018 is involved in DNA damage responses for a number of reasons. KIAA1018 is the only VRR-nuc domain-containing protein in eukaryotes but many bacteria and bacteriophages have genes that encode solely VRR_nuc domains. Although the functions of these genes are unknown, most of them are located in operons
25 that include known DNA repair enzymes, hence the name VRR_nuc (virus-type replication-repair nuclease) (Iyer et al., 2006). The VRR_nuc domains contain a PD-(D/E)XK motif found in the active site of many restriction nucleases (Kosinski et al., 2005) (Fig. 1B). We thus suspected that KIAA1018 might act as a repair endonuclease. A putative role for KIAA1018 in DNA repair is also implied by the presence of a UBZ4-
30 type ubiquitin-binding domain that belongs to the RAD18 family of zinc fingers, a domain commonly found in DNA damage response proteins such as DNA polymerase κ , RAD18 and WRNIP (Fig. 1C) (Hofmann, 2009). Furthermore, KIAA1018 was also found to interact with the MLH1 DNA mismatch repair protein in a genome wide screen (Cannavo et al., 2007). We therefore decided to test if KIAA1018, which we refer to hereafter as
35 FAN1 for reasons that will become clear later, has nuclease activity and if it is involved in DNA repair.

FAN1 has endonuclease activity that is specific for branch points

To test for nuclease activity, we purified recombinant FAN1, fused to an N-terminal NUS-His₆ tag, in bacteria and purified it through three steps of ion exchange chromatography (Fig. 9A). In parallel we purified a mutant version of FAN1 where the conserved Asp981 and Arg982 residues (indicated by asterixes in Fig. 1B) found in the VRR_{nuc} domain were mutated to alanine ("DR" mutant). We next tested the ability of FAN1 to cleave a range of branched DNA substrates that resemble DNA repair and replication intermediates. These included a splayed duplex, a 3' flap, a 5' flap and a nicked three-way junction that mimics a DNA replication fork (Fig. 2A). All substrates were ³²P-labelled at the 5' end of the a₃ strand or the b strand as indicated in Fig. 2A. After incubation with wild-type or mutant FAN1, reaction products were separated by gel electrophoresis under denaturing conditions.

As shown in Fig. 2B, FAN1 displayed strong endonuclease activity towards the 5' flap structure and weaker activity towards the replication fork model. Cleavage affected only one strand of these structures and occurred in the double-stranded (ds) region on the same strand as the flap, 4 nucleotides (nt) 3' to the branchpoint (Fig. 2A). Selectivity of FAN1 for these DNA structures, as opposed to specificity for DNA sequence, was confirmed by analysing the cleavage of an analogous set of branched DNA structures composed of strands with alternative sequence (Fig. 2B,C). The endonuclease activity of the FAN1 DR mutant was severely reduced compared with wild-type protein (Fig. 2B; Fig. 9B) resulting in cleavage rates approximately 1000-fold lower than for wild-type protein (Fig. 2C,D). FAN1 did not exhibit endonuclease activity towards four-way junctions (data not shown).

FAN1 has 5' exonuclease activity

Incubation of wild-type FAN1 with 5'-³²P-labelled branched substrates also produced a short 4 nt radioactive fragment (Fig. 2B,C), suggesting that FAN1 might possess an additional endo- or exonuclease activity. To further investigate this, FAN1 was incubated with linear double-stranded (ds) or single-stranded (ss) DNA in which one of the oligonucleotides was radioactively 5'- or 3'-³²P-labelled. We observed a clear 5' to 3' exonuclease activity which initiates 4 nt from the 5' end and cleaves every phosphate bond thereafter but with varying intensity (Fig. 3A). The exonuclease activity of FAN1 towards ssDNA required that the 5' end be phosphorylated (Fig. 3A). FAN1 exhibited potent 5' exonuclease activity towards DNA substrates with a recessed 5' end indicating that a blunt dsDNA end is not required for exonuclease activity (Fig. 10). The exonuclease activity of FAN1 was severely reduced by mutation of Asp981 and Arg982

in the VRR_nuc domain (Fig. 3A,B). Quantitation of these data showed that the rate of initiation of the exonuclease activity of FAN1 towards dsDNA (0.09 s^{-1} ; Fig. 3B) was approximately half that of the endonuclease towards a 5' flap ($>0.2 \text{ s}^{-1}$; Fig. 2D). The calculated rates of initiation using wild type enzyme (Fig. 3B) were 0.09 s^{-1} for ds DNA (5'- ^{32}P), 0.002 s^{-1} for ds DNA (3'- ^{32}P), 0.0005 s^{-1} for ss DNA (5'- ^{32}P) and too low to measure for ss DNA (3'- ^{32}P). The rate of cleavage of dsDNA (5'- ^{32}P) using the DR mutant was 0.0003 s^{-1} , around 300-fold lower than wild-type FAN1 (Fig. 3B).

These results raised the possibility that the endonuclease activity of FAN1 on branched substrates might be coupled with a 5'-3' exonuclease activity. We therefore examined FAN1-mediated cleavage of a 5' flap in which the a3 strand containing the flap was radioactively labelled at the 3' end (Fig. 3C). This experiment clearly revealed that the endonucleolytic incision described above (Fig. 2B,C) was followed by a 5'-3' exonuclease activity that with time generated ever-shorter products (Fig. 3C). Cleavage was observed at each phosphate bond but with varying intensity. Taken together these data show that FAN1 has a 5' to 3' exonuclease activity which initiates 4 nt from the 5' end on single- and double-stranded DNA, and 4 nt from the branchpoint on 5' flaps and nicked three-way junctions. Rate measurements suggest that FAN1 has a structural selectivity for helical branchpoints.

FAN1 interacts with FANCD2 and FANCI

In an attempt to link the nuclease activity of FAN1 towards branched DNA structures with known DNA repair pathways, we aimed to find FAN1-interacting proteins. Plasmids expressing GFP-FAN1 or GFP alone, both under the control of a tetracycline-inducible promoter, were stably integrated in HEK293 cells. Cells were lysed after induction either in the presence of the reversible protein crosslinker dithiobis (succinimidyl propionate) (DSP) or the de-ubiquitinase inhibitor N-ethyl maleimide (NEM). Extracts were subjected to immunoprecipitation with GFP-Trap beads and protein-protein crosslinking was reversed with dithiothreitol. After SDS-PAGE, strong bands at the expected molecular weights of GFP-FAN1 and GFP were observed in the respective lanes (Fig. 4A). In addition, a range of other proteins was found in GFP-FAN1 but not GFP precipitates. Mass fingerprinting revealed that most of these proteins are involved in DNA repair. Both components of the MLH1-PMS2 complex involved in mismatch repair were found in GFP-FAN1 precipitates when cells were lysed in NEM or DSP (Fig. 15) (Cannavo et al., 2007). We also found FANCD2 and FANCI, but only when DSP was included in the lysis buffer. The presence of ubiquitin in the FANCD2-containing band indicated that the ubiquitinated form of FANCD2 might co-precipitate with FAN1 (Fig.

15). The specificity of the FAN1 protein interactions was independently confirmed by analysing the immunoprecipitates of FLAG-FAN 1 under similar conditions (data not shown) and only the proteins identified in both experiments are shown in Fig. 15.

5 FAN1 interactors were confirmed by a number of experiments. Firstly, western blotting detected MLH1, FANCD2 and FANCI in GFP-FAN1 but not GFP precipitates (Fig. 4B). FANCD2 and FANCI were only found in GFP-FAN1 precipitates when DSP was present in the lysis buffer. To examine endogenous complexes, antibodies were raised in sheep against human FAN1. These antibodies recognised a protein of the expected molecular
10 mass (114 kDa) in extracts of HEK293 cells that was not detected when cells were transfected with FAN1-specific small interfering (si) RNA duplexes (Fig. 12A). These antibodies were used to immunoprecipitate FAN1 from HEK293 cell extracts. Endogenous MLH1, FANCD2 and FANCI were detected in anti-FAN1 immunoprecipitates (Fig. 4C) but not in precipitates using an antibody against an
15 unrelated epitope (HA). Again, FANCD2 and FANCI were only found in GFP-FAN1 precipitates when DSP was present in the lysis buffer. These interactions were not affected by ethidium bromide or by treatment of immunoprecipitates with DNase I or benzonase (data not shown), excluding the possibility that these interactions are DNA-dependent. Abundant DNA repair proteins such as ERCC1 or PCNA, and other FA
20 proteins such as FANCA were not detected in anti-FAN1 immunoprecipitates (Figs. 4B.C).

Size exclusion chromatography of HEK293 cell extracts showed that FAN1 elutes in two sub-complexes; one of these overlaps with MLH1 and elutes slower than the 670 kDa
25 marker, while the other sub-complex elutes faster than the 670 kDa marker and overlaps with FANCD2 and FANCI (Fig. 4D). It is interesting that FAN1 in the latter sub-complex migrates more slowly on SDS-PAGE than the form of FAN1 that co-elutes with MLH1, and this may represent a post-translationally-modified form of FAN1. Taken together these data show that FAN1 binds to MLH1, FANCD2 and FANCI. The acronym FAN1
30 stands for "FANCD2/FANCI-associated nuclease 1".

The UBZ domain of FAN1 interacts with FANCD2

FAN1 has a UBZ domain of the RAD18 type that is found in DNA repair proteins such as WRNIP, POL k and RAD18 (Fig. 1C). The POL k UBZ domain binds to mono-
35 ubiquitinated PCNA (Bienko et al., 2005), but we could not detect PCNA in FAN1 precipitates (Fig. 4C). We hypothesised that instead, the UBZ domain of FAN1 binds to the mono-ubiquitinated form of FANCD2 since we detected FANCD2 in FAN1

precipitates. To test this, the two conserved cysteine residues in the first dyad of the FAN1 UBZ domain (Cys44 and Cys47, indicated by asterixes in Fig. 1C) were both mutated to alanine residues (UBZ* mutant). Whereas wild-type GFP-FAN1 transiently expressed in cells co-immunoprecipitated with endogenous FANCD2 and FANCI, the FAN1 UBZ* mutant did not, even though this mutant retained the ability to bind MLH1 (Fig. 5A).

FANCD2 forms subnuclear "foci" at sites of DNA damage in cells after DNA damage. Endogenous FAN1 formed foci that colocalised with FANCD2 in response to MMC (Fig. 11A). GFP-FAN1 transiently transfected into HEK293 cells also formed subnuclear foci in MMC-treated cells, and these colocalised with FANCD2 (Fig. 5B). The GFP-FAN1 UBZ* mutant, however, did not form subnuclear foci in MMC-treated cells (Figs. 5B, C). Depletion of FAN1 from cells had no detectable effect on MMC-induced FANCD2 focus formation (Fig. 5D). GFP-FAN1 formed foci not just in response to MMC but also in response to HU or IR (Fig. 11B). These data show that co-localization of FAN1 at sites of DNA damage with FANCD2 requires the FAN1 UBZ domain.

FAN1 is recruited to DNA damage by mono-ubiquitinated FANCD2

We next tested the possibility that the mono-ubiquitinated form of FANCD2 interacts with FAN1. Although wild-type FANCD2 transiently transfected into cells stably expressing GFP-FAN1 was detected in GFP-FAN1 precipitates, the FANCD2 K561R mutant that cannot be ubiquitinated was not (Fig. 5E). These data indicate that FAN1 interacts with the mono-ubiquitinated form of FANCD2. To test the possibility that the mono-ubiquitination of FANCD2 might be required to recruit FAN1 to foci, we studied FANCD2 -/- (PD20) human cells stably transfected with wild-type FANCD2 or with a FANCD2 K561R mutant (Garcia-Higuera et al., 2001). We found that GFP-FAN1 did not form MMC-induced foci in FANCD2 -/- cells, but formation of foci was restored when these cells stably express wild-type FANCD2 (Fig. 5F,G). Only background levels of GFP-FAN1 foci occurred when the FANCD2 K561R mutant was expressed in these cells, at all timepoints examined (Fig. 5F,G).

FAN1 is required for cellular resistance to agents that induce ICLs

Mono-ubiquitination of FANCD2 is required for ICL repair but the underlying mechanism is unclear. We hypothesized that the association of FAN1, a nuclease with specificity for branched structures, with the mono-ubiquitinated form of FANCD2 may provide an explanation. Defective ICL repair causes hypersensitivity to agents that induce ICLs. Depletion of FAN1 with two siRNAs targeting different FAN1 exons (Fig. 12A) caused

cells to become hypersensitive to cisplatin and MMC compared with control siRNA (Fig. 6A). The hypersensitivity to ICL-inducing agents associated with depletion of FAN1 was similar to that observed when FANCA was depleted (Fig. 6A). Cells depleted of FAN1 did not show hypersensitivity to DNA damaging agents such as camptothecin, hydroxyurea, UV light, or ionizing radiation (Fig. 12B). In contrast, depletion of the ATR kinase caused hypersensitivity to all of these agents (Fig. 12B). Cells defective in mismatch repair are resistant to killing by 6-thioguanine (6-TG) (Swann et al., 1996), and consistent with this, depletion of MLH1 from HEK293 cells caused cells to become more resistant to 6-TG (Fig. 12B). However, depletion of FAN1 did not, so it is unlikely that FAN1 is involved in mismatch repair.

We also tested a *Caenorhabditis elegans* strain harbouring a deletion in the *C. elegans* orthologue of FAN1 (Figs. 1A, B), encoded by the C01G5.8 locus, for sensitivity to ICL-inducing agents. This deletion does not lead to any overt developmental defects (data not shown). L1-stage worm larvae were exposed to ICL-inducing agents and the effects on the progression to subsequent larval stages L2, L3, L4 and to adult stages were observed. As shown in Fig. 6B worms defective in *Ce-fan-1* were hypersensitive to nitrogen mustard (HN₂) and cisplatin, even more so than a deletion of the *C. elegans fcd-2* Fancd-2 orthologue. These data were confirmed by depleting *Ce-fan-1* by RNA interference (data not shown). Furthermore, the *Ce-rend-1* {tm423} deletion resulted in reduced progeny survival when L4 stage animals were exposed to nitrogen mustard (data not shown). Thus the ICL hypersensitivity associated with defects in FAN1 is evolutionarily conserved.

FAN1 is required for efficient repair of ICL-induced DNA breaks

We next sought to determine at what stage of ICL repair FAN1 acts. ICLs cause mono-ubiquitination of FANCD2 and FANCI that promotes ICL repair. Mono-ubiquitination of FANCD2 and FANCI causes reduced electrophoretic mobility (Garcia-Higuera et al., 2001). Exposure of HEK293 cells transfected with control siRNA to cisplatin or MMC resulted in damage-induced mono-ubiquitination of FANCD2 and FANCI (Fig. 12C). Consistent with previous reports, depletion of FANCA abolished FANCD2 and FANCI mono-ubiquitination. However, depletion of FAN1 has no detectable effect (Fig. 12C). Therefore FAN1 is not required for mono-ubiquitination of FANCD2 or FANCI.

Exposure of cells to ICL-inducing agents causes DSBs, judged by g-H2AX foci or pulsed field gels. These DSBs, formed as a result of replication fork cleavage by MUS81 during ICL unhooking (Hanada et al., 2007; Hanada et al., 2006) (Fig. 8), initiate the HR

step of ICL repair. We next tested whether depletion of FAN1 from cells affected the induction of, or disappearance of γ -H2AX foci induced by ICLs. HEK293 cells were transfected with control siRNA or FAN1 siRNA and were either left untreated or exposed to cisplatin for 2 h. Cells were washed free of cisplatin and incubated in fresh medium, and γ -H2AX foci were counted at various times during recovery. Around 80% of cells transfected with control siRNA had between 2 and 40 foci 24 h after transient exposure to cisplatin; cells with more than two γ -H2AX foci were scored as " γ -H2AX positive". Although the percentage of control siRNA-transfected cells that were γ -H2AX positive declined to almost basal levels by 48 h (Fig. 6C), almost no decrease in the percentage of γ -H2AX-positive cells depleted of FAN1 was observed at this time (Fig. 6C). By 96 h, over 50% of cells depleted of FAN1 were still γ -H2AX positive although γ -H2AX foci had returned to basal levels in cells treated with control siRNA (Fig. 6C). These data show that depletion of FAN1 does not affect DSB induction at replisomes blocked by ICLs but instead causes a defect in DNA repair. Consistent with the persistence of cisplatin-induced γ -H2AX foci, we observed an increase in chromosomal abnormalities in metaphase spreads of FAN1-depleted cells exposed to MMC. FAN1-depleted cells showed a substantial increase in the frequency of cells with more than one chromosome break or radial chromosome, similar to cells depleted of FANCD2 (Fig. 6D). These data are consistent with FAN1 acting in the FA pathway.

FAN1 is required for efficient HR but not for DSB resection or RAD51 loading

Human cells solely expressing FANCD2 K561R show reduced HR efficiency (Nakanishi et al., 2005). It is possible that defect in the resolution of ICL-induced DSBs in FAN1-depleted cells could reflect a defect in HR. We used a reporter system in U2OS cells to measure HR frequency. In this system, an 18 bp sequence recognized by the I-SceI meganuclease is placed between tandem mutant copies of GFP (Nakanishi et al., 2005). HR between these two copies generates a wild-type GFP open reading frame, and functional GFP expression can be detected by FACS analysis. As shown in Fig. 7A, depletion of FAN1 with two separate siRNAs reduced the efficiency of I-SceI-induced HR by 50-60%, similar to the reduction reported for depletion of FA proteins (Nakanishi et al., 2005; Smogorzewska et al., 2007). These data indicate that FAN1 promotes HR in response to DSBs. Depletion of FAN1 - or FANCD2 - from HEK293 cells did not appear to affect the frequency of MMC-induced sister chromatid exchanges (Fig. 13).

We postulated that the exonuclease activity of FAN1 could affect HR by controlling the resection of DSBs generated during ICL repair. Resection of DSBs leads to the generation of ssDNA, and the coating of ssDNA by RPA leads to RPA foci. We analysed

RPA foci after exposure of cells to a pulse of cisplatin to assess DSB resection and found that depletion of FAN1 did not prevent cisplatin-induced RPA focus formation. In fact, depletion of FAN1 caused a slight increase in the average number of RPA foci per cell and in the average number of cells with >9 RPA foci (Fig. 7B). This suggests that FAN1 is not required for resection of DSBs. RPA foci gradually disappeared during the recovery of cells treated with control siRNA or FAN1 siRNA from cisplatin (Fig. 7B).

DSB resection is followed by formation of the RAD51 nucleoprotein filament on the resected DSB so we examined formation of RAD51 foci at various times during recovery of cells from cisplatin. After 24 h recovery, cells depleted of FAN1 showed an approximately 2.5-fold increase in the number of cells with RAD51 foci, and a similar increase in the number of RAD51 foci per cell, compared with control siRNA (Fig. 7C). At subsequent times during recovery, FAN1-depleted cells continued to have around twice as many RAD51 foci as control cells and these data are consistent with a defect in HR. We also tested the effects of depleting FAN1 on RAD51 loading after IR. Four hours after exposure to IR, the number of cells with greater than nine RAD51 foci, and the average number of RAD51 foci per cell, was similar in FAN1-depleted and cells treated with control siRNA (Fig. 7D). However, whereas RAD51 foci declined to basal levels by 12-24 h post-IR in control cells, FAN1-depleted cells showed a delay in the disappearance of RAD51 foci. These data suggest that FAN1 depletion leads to failure of a late stage of HR.

Discussion

It has been known for almost a decade that mono-ubiquitination of FANCD2 is required for ICL repair (Garcia-Higuera et al., 2001). However, the molecular role of this ubiquitination event has remained elusive. Here we report that the FAN1 nuclease is recruited to sites of DNA damage by mono-ubiquitinated FANCD2 and thus might act as an effector molecule carrying out one or more nucleolytic steps required for ICL repair. The phenotypic consequences of depleting FAN1 from human cells - sensitivity to ICL-inducing agents, chromosome instability in MMC-treated cells and defects in HR - are consistent with a role in ICL repair and are similar to those seen in cells solely expressing FANCD2 K561R (Moldovan and D'andrea, 2009). These findings might at least in part explain how ubiquitination of FANCD2 promotes ICL repair.

According to our data, FAN1 is recruited to sites of ICLs by mono-ubiquitinated FANCD2. This is supported by the requirement of the FAN1 UBZ domain (Fig. 5B) and the mono-ubiquitinated form of FANCD2 (Fig. 5F) for FAN1 localization. Furthermore,

FAN1 and FANCD2 proteins co-precipitate in a manner that depends on FANCD2 K561 and on the FAN1 UBZ domain. Both ubiquitinated and non-ubiquitinated FANCD2 were detected in FAN1 immunoprecipitates even though when FANCD2 K561 is mutated, no FANCD2 is detected in FAN1 immunoprecipitates. This discrepancy may be explained by de-ubiquitination of a proportion of FANCD2 after cell lysis or by the association of modified FANCD2 with the unmodified form of the protein. Even though DNA damage stimulates FANCD2 ubiquitination, FAN1 interacts with FANCD2 even without exposure of cells to genotoxins. This is probably a reflection of basal FANCD2 mono-ubiquitination that occurs in the absence of DNA damage in S-phase cells (Taniguchi et al., 2002).

Building upon existing models for ICL repair, and what is already known about the role of FANCD2 mono-ubiquitination in this pathway, it is possible to speculate on where on the ICL repair pathway FAN1 might act. Experiments on the replication of plasmids bearing single ICLs in *Xenopus* egg extracts showed that FANCD2 mono-ubiquitination is required for ICL unhooking (Knipscheer et al., 2009) which suggests that FAN1 might act at this point. In this system it was proposed that initiation of ICL repair requires the convergence of two replication forks at an ICL (Knipscheer et al., 2009). Consequently an ICL would be located at the intersection of an X-structure shown in Fig. 14A. In this scenario cleavage of the leading strand template of one of the forks by MUS81 in concert with cleavage of the same strand on the opposite side of the ICL (which resembles a 5' flap) by a second nuclease would unhook the ICL (Fig. 14A). This nuclease could be FAN1 based on our *in vitro* data showing that this nuclease preferentially cleaves the double-stranded portion of a synthetic 5' flap structure (Fig. 2A). However, there are conceptual problems with FAN1 acting at this point of the ICL repair pathway. Firstly, this hypothesis predicts that MUS81 and FAN1 are each responsible for 50% of the one-ended DSBs in cells treated with ICL-inducing agents. Assuming that g-H2AX foci are representative of DSBs, this is not what we observe - depletion of FAN1 does not affect formation of cisplatin-induced DSBs (Fig. 6C) whereas deletion of MUS81 abolished all ICL-induced DSB (Hanada et al., 2006). Secondly, this hypothesis requires that MUS81 acts on the leading strand template of one of the two stalled replication forks (Fig. 14A) but it is difficult to see why MUS81 would not cleave the leading strand template of both forks (Fig. 14B). Cleavage of both forks by MUS81 would result in two one-ended DSBs and a linear duplex containing the ICL, and it is unlikely that unhooking of the ICL from this linear duplex would require FAN1. Thirdly, it is not yet clear if the two-fork model for ICL repair is relevant *in vivo*, and it is possible that the collision of a single replication fork with an ICL is sufficient to initiate ICL repair (Fig. 14C). In this scenario, it is difficult to see how FAN1 could be involved in unhooking since it would have to cleave linear

duplex DNA on the 5' side of the ICL, a structure that is not flap-like in nature. More experiments are required to test if the one-fork or two-fork models for ICL repair, or both, operate *in vivo* and to test if the defect in ICL unhooking in the context of the FANCD2 K561R mutant is due to a defect specifically in recruitment of FAN1.

5

After unhooking, excision repair removes the crosslink adduct and translesion synthesis fills in the gap (Fig. 14A). Although FANCD2 ubiquitination appears to be involved in translesion synthesis in the *Xenopus* cell free system, it is difficult to see how FAN1 could be involved at this stage and it is likely that other FANCD2-binding proteins are required. We found that FAN1 interacts with MLH1 that is involved in mismatch repair, consistent with a previous report (Cannavo et al., 2007). At present the significance of this interaction is not clear but it may be that the interaction of FAN1 with MLH1 promotes the correction of translesion synthesis-induced mismatches during ICL repair.

10

Regardless of how exactly MUS81 - alone or in conjunction with FAN1 or an as yet unidentified nuclease - unhooks the ICL, at least one and possibly two (one-ended) DSBs are generated (Fig. 14A). These DSBs are resected and this is a function that could be fulfilled by FAN1 based on our finding that FAN1 has 5'-3' exonuclease activity that is capable of generating 3' overhangs (Fig. 3A). However, cytological data showing that RPA loading is normal in FAN1-depleted cells (Fig. 7B) argues against this role although potential redundancy between FAN1 and other 5'-3' exonucleases would need to be investigated.

20

One of the overhangs generated by DSB resection invades the complementary duplex to initiate D-loop formation and HR (Fig. 7A,B). After extension by DNA synthesis, the invading strand re-anneals to the complementary strand in the duplex it came from originally. It is possible that continued DNA synthesis on the parent strand displaces the DNA in front of it. This would generate a 5' flap that could be cleaved by FAN1. In this scenario, FAN1 is required at a late step in HR and several observations are consistent with this hypothesis. Firstly, there is a delay in the disappearance of γ -H2AX foci induced by cisplatin or IR in cells treated with FAN1 siRNA compared with control cells. Similar results were reported recently in FA cells (Leskovac et al., 2010). At all timepoints during recovery from cisplatin and IR, there are more RAD51 foci in FAN1-depleted cells than in control cells (Fig. 7C). It is possible that disappearance of foci in FAN1-depleted cells reflects inappropriate repair perhaps by non-homologous end joining, and this may account for the increase in radial chromosomes seen in MMC-treated cells depleted of FAN1 (Fig. 6D). Secondly, depletion of FAN1 from human cells results in reduced

30

35

efficiency in I-SceI-induced HR (Fig. 7A) and similar results were reported in human cells expressing FANCD2 K561R (Nakanishi et al., 2005). Depletion of FAN1 does not affect RAD51 loading in cisplatin-treated cells (Fig. 7C) suggesting that if FAN1 functions at the HR step of ICL repair then it acts independently of RAD51 or after
5 RAD51 focus formation. Similar to FANCD2 null cells, depletion of FAN1 does not affect the frequency of MMC-induced SCEs (Fig. 13). Therefore if FAN1 does act at the HR step of ICL repair, then its role may be restricted to a sub-pathway of HR such as synthesis-dependent strand annealing that avoids crossing over.

10 The SLX4 complex of structure-specific nucleases is also required for ICL repair but it is not yet clear at what stage this complex acts (Fekairi et al., 2009; Munoz et al., 2009; Svendsen et al., 2009). SLX4-XPF-SLX1-MUS81 can cleave three-way DNA junctions, 3' flaps and 5' flaps *in vitro*, and so FAN1 specificity overlaps with the SLX4 complex in 5' flap cleavage. We could find no evidence for an interaction of FAN1 with the SLX4
15 scaffold (unpublished data). It is not yet clear why two 5' endonucleases are required during ICL repair and it will be important to test redundancy between FAN1 and the SLX4 complex.

FAN1 is the only VRR_nuc domain-containing protein in eukaryotic cells. These
20 domains are found in all kingdoms of life, but the functions of most of them are unknown (Iyer et al., 2006). Many bacteria and phages have VRR_nuc domain proteins and so it appears that the FA repair pathway which appeared relatively late during evolution was built on a more ancient VRR_nuc domain nuclease. It will be interesting to follow up on this hypothesis. Many cytotoxic anti-cancer agents act by inducing ICLs and it is
25 possible that nucleases such as FAN1 are good targets for sensitizing cancer cells to killing by ICLs. Finally, although the majority of FA patients have mutations in the known FA genes, FA patients exist where mutations in known FA genes could not be found. In this light, it is likely that FAN1 mutations will be found in some of these patients.

30 References

Akkari, Y.M.N., Bateman, R.L., Reifsteck, C.A., Olson, S.B., and Grompe, M. (2000). DNA replication is required to elicit cellular responses to psoralen-induced DNA interstrand cross-links. *Molecular and Cellular Biology* 20, 8283-8289.
Bergstralh, D.T., and Sekelsky, J. (2008). Interstrand crosslink repair: can XPF-ERCC1
35 be let off the hook? *Trends in Genetics* 24, 70-76.

Bhagwat, N., Olsen, A.L., Wang, A.T., Hanada, K., Stuckert, P., Kanaar, R., D'Andrea, A., Niedernhofer, L.J., and McHugh, P.J. (2009). XPF-ERCC1 Participates in the Fanconi Anemia Pathway of Cross-Link Repair. *Molecular and Cellular Biology* 29, 6427-6437.

Bienko, M., Green, C.M., Crosetto, N., Rudolf, F., Zapart, G., Coull, B., Kannouche, P., Wider, G., Peter, M., Lehmann, A.R., *et al.* (2005). Ubiquitin-binding domains in Y-family polymerases regulate translesion synthesis. *Science* 310, 1821-1824.

Brenner, S. (1974). Genetics of *Caenorhabditis-Elegans*. *Genetics* 77, 71-94.

Cannavo, E., Gerrits, B., Marra, G., Schlapbach, R., and Jiricny, J. (2007). Characterization of the interactome of the human MutL homologues MLH1, PMS1, and PMS2. *J Biol Chem* 282, 2976-2986.

Ciccio, 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.

Collis, S.J., Ciccio, A., Deans, A.J., Horejsi, Z., Martin, J.S., Maslen, S.L., Skehel, J.M., Elledge, S.J., West, S.C., and Boulton, S.J. (2008). FANCM and FAAP24 Function in ATR-Mediated Checkpoint Signaling Independently of the Fanconi Anemia Core Complex. *Molecular Cell* 32, 313-324.

Deans, A.J., and West, S.C. (2009). FANCM Connects the Genome Instability Disorders Bloom's Syndrome and Fanconi Anemia. *Molecular Cell* 36, 943-953.

Fekairi, S., Scaglione, S., Chahwan, C., Taylor, E.R., Tissier, A., Coulon, S., Dong, M.Q., Ruse, C., Yates, J.R., 3rd, Russell, P., *et al.* (2009). Human SLX4 is a Holliday junction resolvase subunit that binds multiple DNA repair/recombination endonucleases. *Cell* 138, 78-89.

Garcia-Higuera, I., Taniguchi, T., Ganesan, S., Meyn, M.S., Timmers, C., Hejna, J., Grompe, M., and D'Andrea, A.D. (2001). Interaction of the Fanconi anemia proteins and BRCA1 in a common pathway. *Mol Cell* 7, 249-262.

Hanada, K., Budzowska, M., Davies, S.L., van Drunen, E., Onizawa, H., Beverloo, H.B., Maas, A., Essers, J., Hickson, I.D., and Kanaar, R. (2007). The structure-specific endonuclease Mus81 contributes to replication restart by generating double-strand DNA breaks. *Nature Structural & Molecular Biology* 14, 1096-1104.

Hanada, K., Budzowska, M., Modesti, M., Maas, A., Wyman, C., Essers, J., and Kanaar, R. (2006). The structure-specific endonuclease Mus81-Eme1 promotes conversion of interstrand DNA crosslinks into double-strand breaks. *Embo Journal* 25, 4921-4932.

Hofmann, K. (2009). Ubiquitin-binding domains and their role in the DNA damage response. *DNA Repair (Amst)* 8, 544-556.

Iyer, L.M., Babu, M.M., and Aravind, L. (2006). The HIRAN domain and recruitment of chromatin remodeling and repair activities to damaged DNA. *Cell Cycle* 5, 775-782.

Knipscheer, P., Raschle, M., Smogorzewska, A., Enoiu, M., Ho, T.V., Scharer, O.D., Elledge, S.J., and Walter, J.C. (2009). The Fanconi Anemia Pathway Promotes Replication-Dependent DNA Interstrand Cross-Link Repair. *Science* 326, 1698-1701.

Kosinski, J., Feder, M., and Bujnicki, J.M. (2005). The PD-(D/E)XK superfamily revisited: identification of new members among proteins involved in DNA metabolism and functional predictions for domains of (hitherto) unknown function. *Bmc Bioinformatics* 6, -

Leskovac, D., Vujic, D., Guc-Scekic, M., Petrov, S., Joksic, I., Slijepcevic, P., and Joksic, G. (2010). Fanconi anemia is characterized by delayed repair kinetics of DNA double-strand breaks. *Tohoku J Exp Med* 221, 69-76.

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.

Mccabe, K.M., Olson, S.B., and Moses, R.E. (2009). DNA Interstrand Crosslink Repair in Mammalian Cells. *Journal of Cellular Physiology* 220, 569-573.

Moldovan, G.L., and D'andrea, A.D. (2009). How the Fanconi Anemia Pathway Guards the Genome. *Annual Review of Genetics* 43, 223-249.

Munoz, I.M., Hain, K., Declais, A.C., Gardiner, M., Toh, G.W., Sanchez-Pulido, L., Heuckmann, J.M., Toth, R., Macartney, T., Eppink, B., *et al.* (2009). Coordination of structure-specific nucleases by human SLX4/BTBD12 is required for DNA repair. *Mol Cell* 35, 116-127.

Nakanishi, K., Yang, Y.G., Pierce, A.J., Taniguchi, T., Digweed, M., D'Andrea, A.D., Wang, Z.Q., and Jasin, M. (2005). Human fanconi anemia monoubiquitination pathway promotes homologous DNA repair. *Proceedings of the National Academy of Sciences of the United States of America* 102, 1110-1115.

Niedernhofer, L.J., Odijk, H., Budzowska, M., van Drunen, E., Maas, A., Theil, A.F., de Wit, J., Jaspers, N.G., Beverloo, H.B., Hoeijmakers, J.H., *et al.* (2004). The structure-specific endonuclease Ercc1-Xpf is required to resolve DNA interstrand cross-link-induced double-strand breaks. *Mol Cell Biol* 24, 5776-5787.

Patel, K.J., and Joeje, H. (2007). Fanconi anemia and DNA replication repair. *DNA Repair* 6, 885-890.

Raschle, M., Knipscheer, P., Enoiu, M., Angelov, T., Sun, J., Griffith, J.D., Ellenberger, T.E., Scharer, O.D., and Walter, J.C. (2008). Mechanism of replication-coupled DNA interstrand crosslink repair. *Cell* 134, 969-980.

Rothfuss, A., and Grompe, M. (2004). Repair kinetics of genomic interstrand DNA cross-links: Evidence for DNA double-strand break-dependent activation of the Fanconi anemia/BRCA pathway. *Molecular and Cellular Biology* 24, 123-134.

Smogorzewska, A., Matsuoka, S., Vinciguerra, P., McDonald, E.R., Hurov, K.E., Luo, J., Ballif, B.A., Gygi, S.P., Hofmann, K., D'Andrea, A.D., *et al.* (2007). Identification of the FANCI protein, a monoubiquitinated FANCD2 paralog required for DNA repair. *Cell* 129, 289-301.

- 5 Svendsen, J.M., Smogorzewska, A., Sowa, M.E., O'Connell, B.C., Gygi, S.P., Elledge, S.J., and Harper, J.W. (2009). Mammalian BTBD12/SLX4 Assembles A Holliday Junction Resolvase and Is Required for DNA Repair. *Cell* 738, 63-77.

- Swann, P.F., Waters, T.R., Moulton, D.C., Xu, Y.Z., Zheng, Q.G., Edwards, M., and Mace, R. (1996). Role of postreplicative DNA mismatch repair in the cytotoxic action of thioguanine. *Science* 273, 1109-1111.

- 10 Taniguchi, T., Garcia-Higuera, I., Andreassen, P.R., Gregory, R.C., Grompe, M., and D'Andrea, A.D. (2002). S-phase-specific interaction of the Fanconi anemia protein, FANCD2, with BRCA1 and RAD51. *Blood* 100, 2414-2420.

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

CLAIMS

1. A method for identifying a compound expected to be useful in modulating, for example inhibiting, KIAA1018/MTMR15/FAN1(FAN1) endonuclease activity, the method
5 comprising the steps of (1) determining whether a test compound modulates, for example inhibits, the endonuclease activity of a FAN1 polypeptide on a branched polynucleotide substrate and (2) selecting a compound which modulates, for example inhibits, the said FAN1 endonuclease activity.
- 10 2. The method of claim 1, wherein the branched polynucleotide substrate is a substrate 5' flap-containing nucleic acid or replication fork nucleic acid.
3. A method for identifying a compound expected to be useful in modulating, for example inhibiting, FAN1 5'-exonuclease activity, the method comprising the steps of (1)
15 determining whether a test compound modulates, for example inhibits, the 5'-exonuclease activity of a FAN1 polypeptide on a substrate polynucleotide and (2) selecting a compound which modulates, for example inhibits, the said FAN1 5'-exonuclease activity.
- 20 4. A method for identifying a compound expected to be useful in modulating, for example inhibiting, the repair of a DNA inter-strand crosslink (ICL), resolution of an ICL-induced double strand break, or homologous recombination in a cell, the method comprising the steps of (1) determining whether a test compound modulates, for example inhibits, the endonuclease or 5'-exonuclease activity of a FAN1 polypeptide on
25 a substrate polynucleotide and (2) selecting a compound which modulates, for example inhibits, the said FAN1 polypeptide endonuclease or 5'-exonuclease activity.
5. A method for identifying a compound expected to be useful in modulating, for example inhibiting, the repair of a DNA inter-strand crosslink, resolution of an ICL-induced double strand break, or homologous recombination in a cell, the method
30 comprising the steps of (1) determining whether a test compound modulates, for example inhibits, the interaction between a FAN1 polypeptide and an ubiquitinated FANCD2 or FANCI polypeptide and (2) selecting a compound which modulates, for example inhibits, the said interaction between the FAN1 polypeptide and the
35 ubiquitinated FANCD2 or FANCI polypeptide.

6. A method for identifying a compound expected to be useful in increasing cancer cells' sensitivity to ICLs or enhancing the effect of platinum based chemotherapy or other ICL-inducing therapy (for example mitomycin C, nitrogen mustards, nitrosoureas, alkylalkanesulphonate, and Cyclophosphamide) on cancer cells, the method comprising the steps of (1) determining whether a test compound modulates, for example inhibits, the endonuclease or 5'-exonuclease activity of a FAN1 polypeptide on a substrate polynucleotide; or the interaction between a FAN1 polypeptide and an ubiquitinated FANCD2 or FANCI polypeptide, and (2) selecting a compound which modulates, for example inhibits, the endonuclease or 5'-exonuclease activity of the FAN1 polypeptide on the substrate *polynucleotide*; or the interaction between the FAN1 polypeptide and the ubiquitinated FANCD2 or FANCI polypeptide.

7. A method for identifying a compound expected to be useful in modulating, for example inhibiting, the repair of a DNA inter-strand crosslink, resolution of an ICL-induced double strand break, or homologous recombination in a cell, or in increasing cancer cells' sensitivity to ICLs or enhancing the effect of platinum based chemotherapy or other ICL-inducing therapy on cancer cells, the method comprising the steps of (1) determining whether a test compound modulates, for example inhibits, said repair of a DNA inter-strand crosslink, resolution of an ICL-induced double strand break, or homologous recombination in the cell, or the cell's resistance to ICLs or the effect of platinum based chemotherapy or other ICL-inducing therapy on the cell, wherein said determining is done on at least two cells, wherein said cells differ in the amount or activity of FAN1, and (2) selecting a compound which modulates, for example inhibits, the said repair of a DNA inter-strand crosslink, resolution of an ICL-induced double strand break, or homologous recombination in the cell, or the cell's resistance to ICLs or the effect of platinum based chemotherapy or other ICL-inducing therapy to a greater extent in the cell having more FAN1 than in the cell having less FAN1.

8. The method of any one of claims 1 to 7 wherein the FAN1 polypeptide is wild type human FAN1 polypeptide or a fragment thereof, or a fusion either thereof.

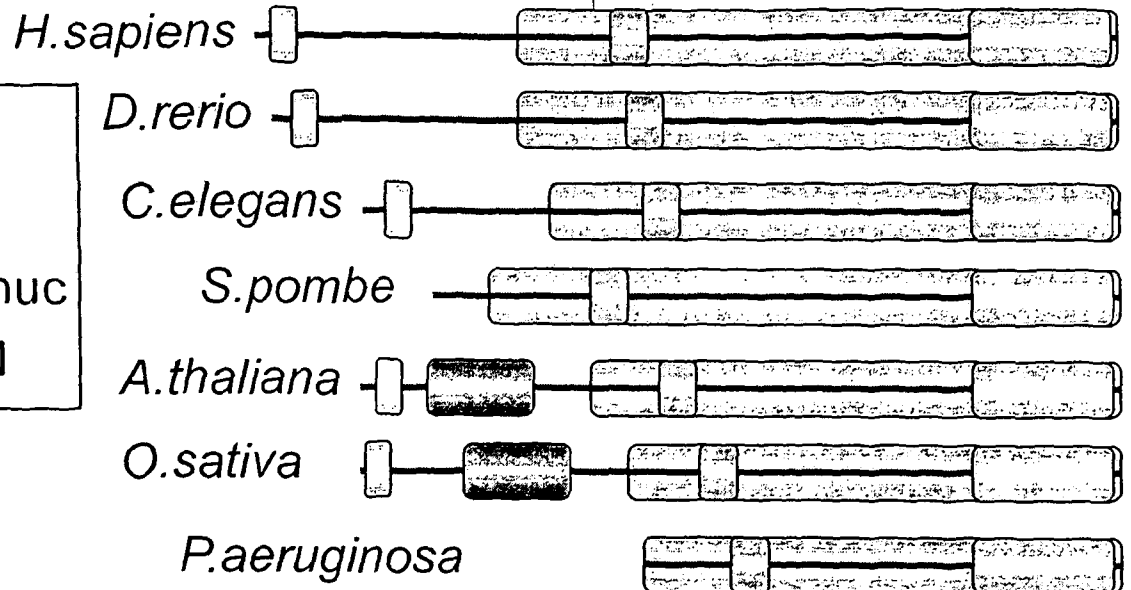
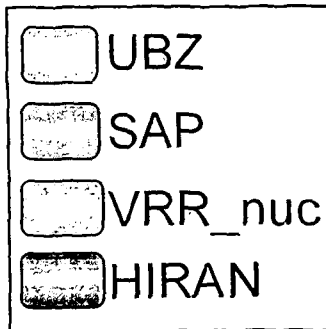
9. The method of claim 8 wherein the fragment comprises the VRR-nuc domain and, optionally, further comprises the SAP domain, and/or optionally, further comprises the UBZ domain.

10. The method of claim 8, when dependent on claims 5 or 6, wherein the fragment comprises the UBZ domain and, optionally, further comprises the SAP domain, and/or optionally, further comprises the VRR-nuc domain.
- 5 11. The method of any one of claims 8 to 9 wherein the fragment corresponds to at least residues 460-1017 of human FAN1.
12. The method of any one of the preceding claims wherein the FAN1 polypeptide is a NUS-His6 fusion polypeptide.
- 10 13. The method of any one of the preceding claims wherein the FAN1 polypeptide is recombinant.
14. The method of any one of claims 3, 4, 6 or 8 to 13, wherein the substrate
15 polynucleotide is a branched polynucleotide.
15. The method of claim 14, wherein the substrate polynucleotide consists of or comprises a 5' flap-containing polynucleotide or a replication fork polynucleotide.
- 20 16. The method of claim 2 or 15, wherein the substrate polynucleotide consists of or comprises the structure of the 5' flap polynucleotide identified in Figure 2A.
17. A purified preparation or kit of parts comprising a FAN1 polypeptide or polynucleotide; and a substrate polynucleotide as defined in claims 14 to 16 and/or a
25 FANCD2 or FANCI polypeptide.
18. The purified preparation or kit of parts of Claim 17, wherein the substrate polynucleotide is fluorescently labelled.
- 30 19. The preparation or kit of parts of claim 17 comprising a recombinant FAN1 polynucleotide.
20. The preparation or kit of parts of claims 17 to 19 further comprising ubiquitin or dithiobis (succinimidyl propionate) (DSP) or a de-ubiquitinase inhibitor, for example N-ethyl maleimide (NEM).
35

21. A mutant FAN1 polypeptide or polynucleotide wherein Asp 981 and/or Arg982 are mutated, optionally to alanine; and/or wherein Cys44 and/or Cys47 are mutated, optionally to alanine.
- 5 22. Use of a FAN1 polypeptide in an *in vitro* method of resolving a DNA ICL.
23. An *in vitro* method of resolving a DNA ICL, the method comprising the step of using a FAN1 polypeptide.
- 10 24. The method of any one of claims 1 to 16 comprising the step of assessing whether the compound modulates the repair of a DNA inter-strand crosslink, resolution of an ICL-induced double strand break, or homologous recombination in a cell, or in increasing cancer cells' sensitivity to ICLs or enhancing the effect of platinum based chemotherapy or other ICL-inducing therapy on cancer cells, in a whole cell, tissue or
15 organism; and a compound that modulates the activity or therapeutic response is selected.
25. The method of claim 24 further comprising the step of assessing whether the compound modulates the activity of a FAN1 polypeptide in the whole cell, tissue or
20 organism, and a compound that modulates the activity is selected.
26. The method of any one of claims 1 to 16, 24 or 25 further comprising the step of synthesising, purifying and/or formulating the selected compound.
- 25 27. A method for preparing a compound which modulates the activity of a FAN1 polypeptide, the method comprising:
1) performing a method according to any one of claims 1 to 16, 24 or 25; and
2) synthesising, purifying and/or formulating the selected compound.
- 30 28. A method of characterising a patient with cancer and/or suspected Fanconi Anaemia, the method comprising the step of assessing the patient's genotype or phenotype for FAN1.
29. The method of Claim 28, wherein the method further comprises the step of
35 assessing the patient's genotype or phenotype for SLX4 and/or SLX1.

30. The method of claim 5 wherein the interaction is detected using fluorescence resonance energy transfer; immunoprecipitation; subcellular fractionation or imaging; surface plasmon resonance (SPR); or isothermal titration calorimetry.

1/15

A.**B.**

H.sapiens CTDSFETSRRLPALEARLQLTHDAPEESLRAWAATWHEQ..EGRVASLYSWDRFSL.QQAQDLVSCYGGPVISGVCRHL
D.rerio HTDCFYGSRREAEARAEMREASVETLQDIIADVSTQ..EGRVCALINWERFSTP.QQAQSLVACLGGHFLSGVFLRL
C.elegans LSNITLYSKRKEKEEDREVWIDELEEQBLIEENIRKIWDLK..HNENRECSYKQSPMGAEDCVSFFOCTPRPALIILIRRL
S.pombe HTDSFYISRESTIMKRLEEIRNCKAGL...ILKDNVIREHQKRTFCVGENWSYTCE...MLEIVDCINDNGLAQIFLAL
A.thaliana ETESFYLRKETIESOLEKVANGMAEE...LLIISMETQ..RGTACRGVAVERTSLE..ELRAAVACVGGMCLASLORHL
O.sativa ETDEFYRSRKDLIESOLKKIODGIABE...LLISSWELH..QCTSCRGVNDRESLT..DRAAVVCTGCHRLASLRLHL
P.aeruginosa HSADFERORRAALFEACLGRIGEGSYRD...AIRCRVDRK..FGLQSPFVYVWELLGEE..LLEQATVCLPAHRAWFERL

H.sapiens RADERHCRGGIPDLVWNSQ...SRHFKLVEVKGEVDRLSHKQMIPLAELQKLGAIEVEVCHVAV
D.rerio AKDYRHCRGGIPDLVWSTS...SNKYKLVEVKGEVDRLSQKQOIWDELRLKLGAEVEVCHVTAT
C.elegans AENYRNSRSGEPDLTLNPE...TKRVAVVEVKGEGDRLSTKORLWLAELADSGIRAEVCHVAAQ
S.pombe TDYRNSRSGEPDLCLNPS...KKKMFSEVKSNDRLSEAKFWLSLISSEVDEVEVCHVSMH
A.thaliana AADYRSWCSGMPDLLVRFKENGYEAGEAKLVEVKSEKELSEQORAWLLMLMDSGFNVETCKVRPA
O.sativa ALDYRSWSGMPDLLVRFDERGGGEAKLVEVKGEVDRLSEQORAWLLMLMDFGFEVEVCKVSPV
P.aeruginosa LEDIPGNRAGLIPDLTQWPA...QRRYRMVEVKGEGDRLQDNOLRWLOFCREREMPVAVCYVRWH

C.

KIAA1018 KLCAPVCSKMPVRYD...INRHIDEMCANNDF
 POLK ILTCPVCFRAQGCISLEALNKHYD.ECLDGPS
 ALVCPVCFNVEQKTSDLTLFNHYD.VCLNKSF
 RAD18 KYDCPVCGVNIPESH...INKHLD.SCLSREE
 RNF220 GQVCPICNPLAGSE.QEWSRYVE.HCLSKRE
 UIMC1 QVSCPLCDQCFPPTK...IERHAM.YCNGLME
 NEKCYLCKSLVPFRE...YQCHYD.SCLQLAK
 QVQCPVCCQMMPAAH...INSHLD.RCLLLHP
 WRNIP PFRCPNCGORFETENL..YVEHMS.SCLD...
 ZBTB1 GFKCPICSKSVASDE...MEMHFI.MCLSKPR
 ZNRF1 GFKCPVCSKRVSSDE...MDLHIV.MCLTKPR
 ZNRF2 LFFCQICQNLNLSAMNVTREQHYN.RCLDEAE
 BTBD12 IPECFICGKPLTLKS..RTSHLK.QCAVKME
 IPECFICGKPLTLKS..RTSHLK.QCAVKME
 C10RF124 MVNCPVCFQNEVLESQ...INSHLD.WCLEGDS
 DCLRE1A DGYCPNCOMPFSSSLIGQTPRWVVF.ECLDSPP
 FAM164A LLPCKIKCGRTFFPVA...PKKRGF.ICQKTAT
 YIQCPYCORRFNENA...ADRHIN.FCKEQAA
 FAM164B LFPCEVCGRRFAADV...LERHGP.ICKKLFN
 FAM164C LGECSHCGRKFLSFR...LERHSN.ICSRMRG

* *

Figure 1

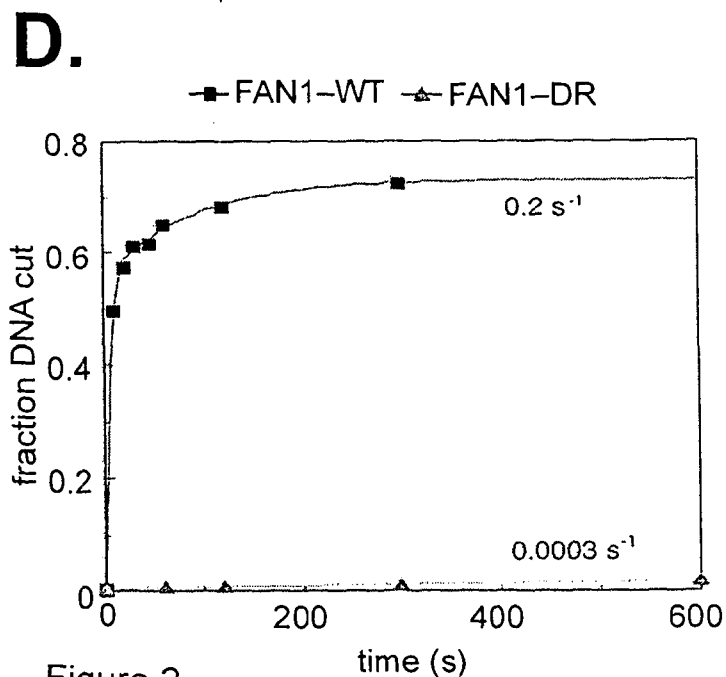
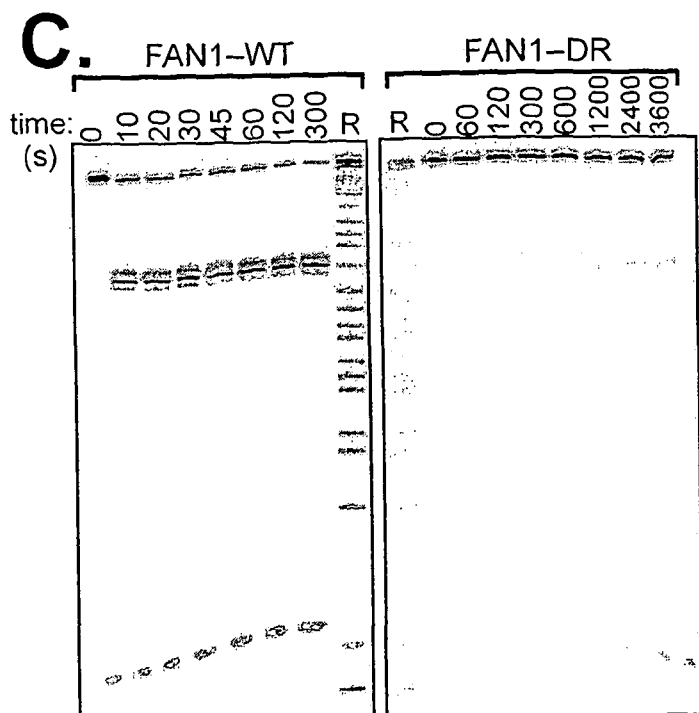
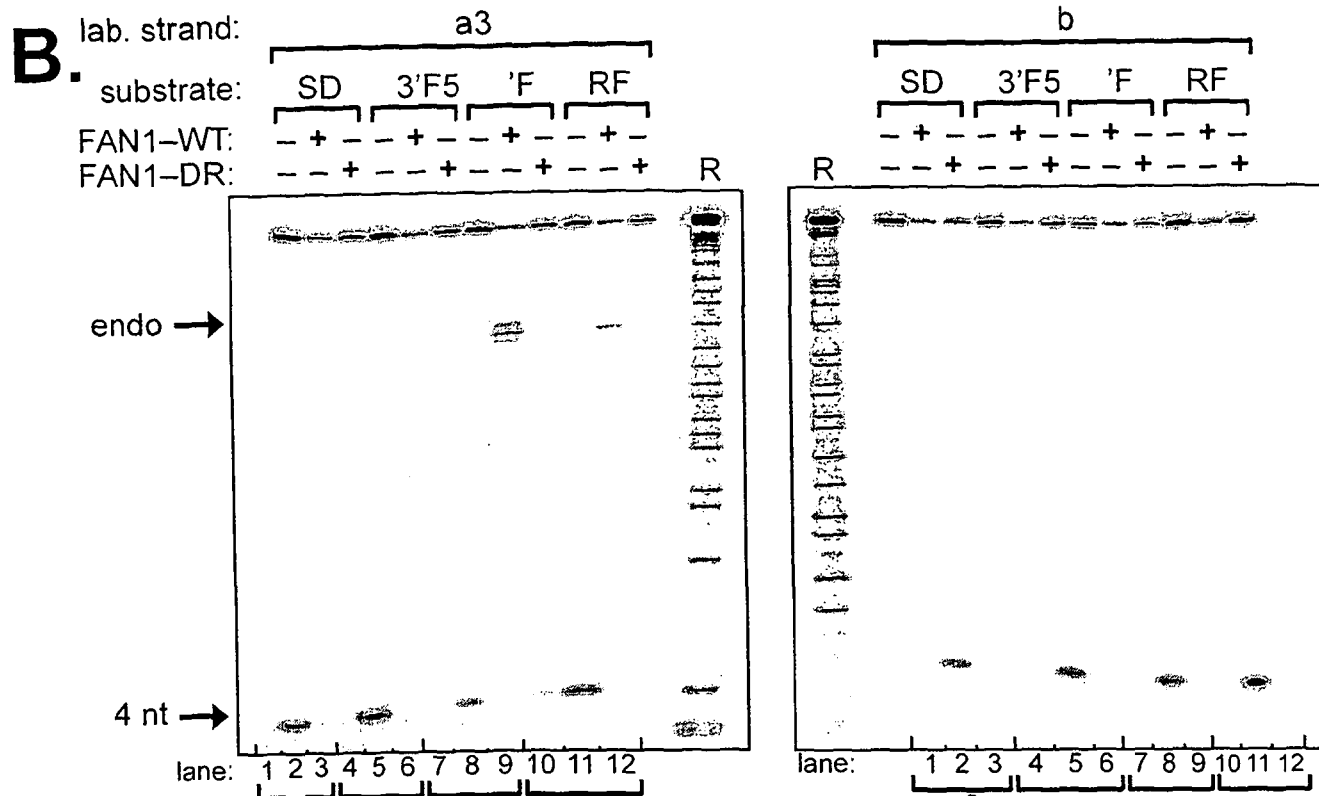
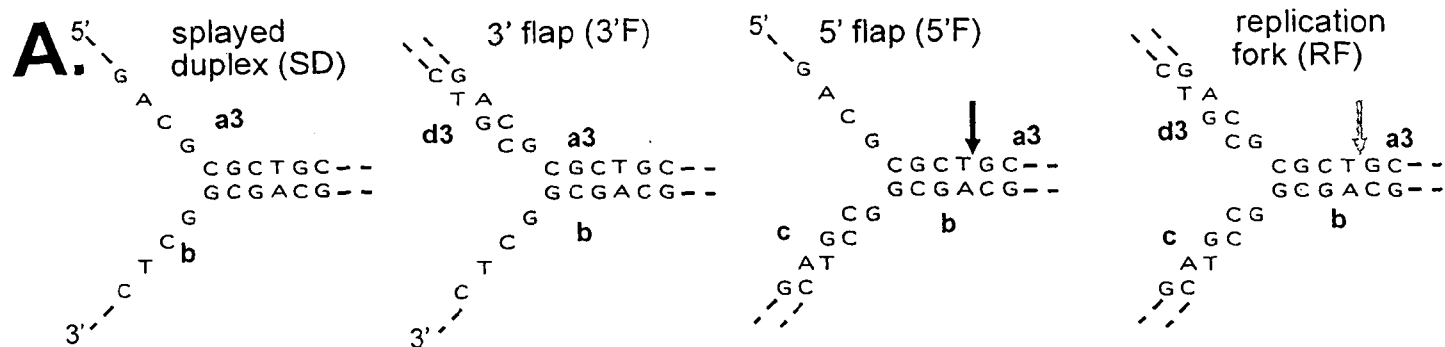
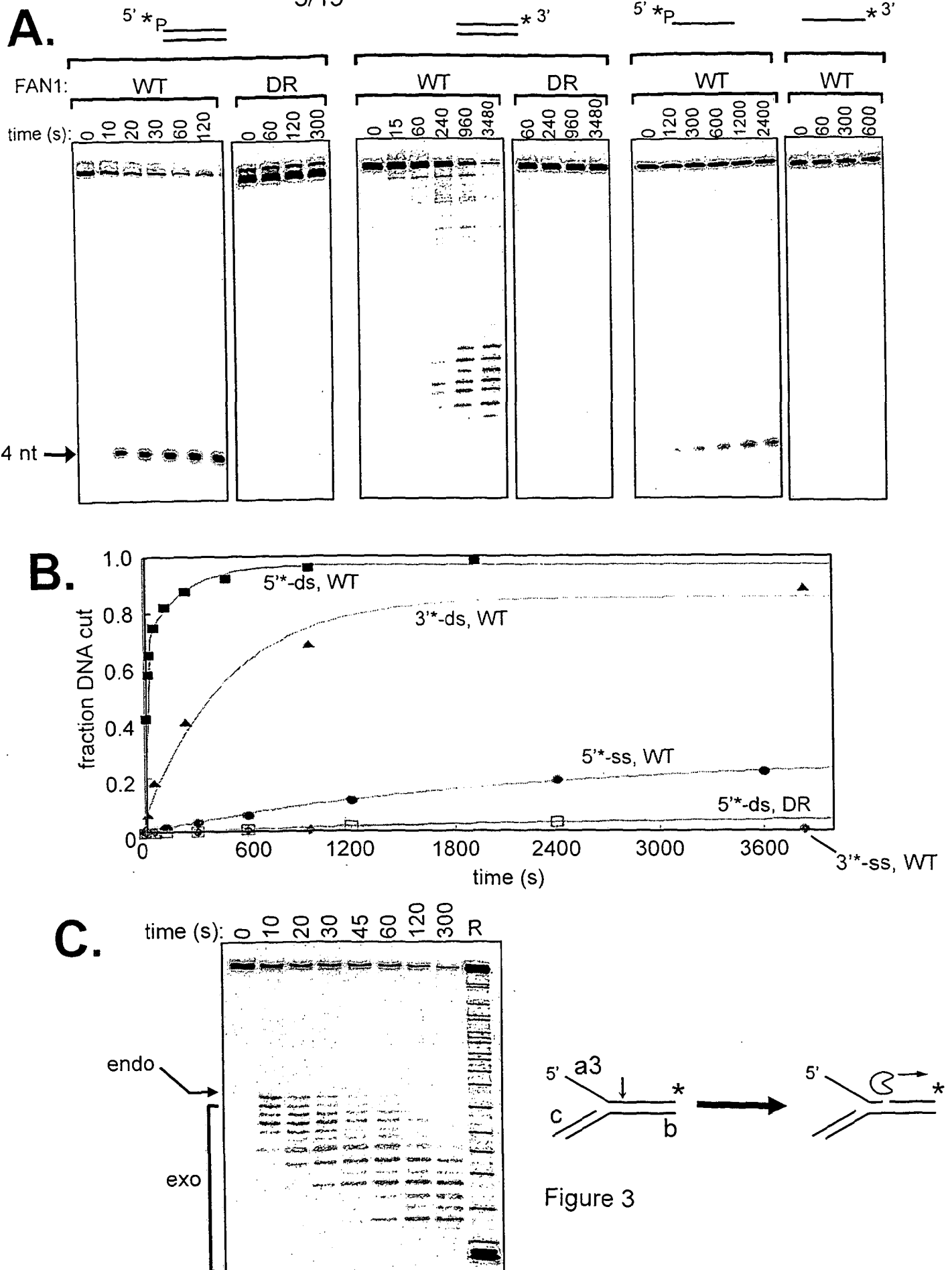


Figure 2



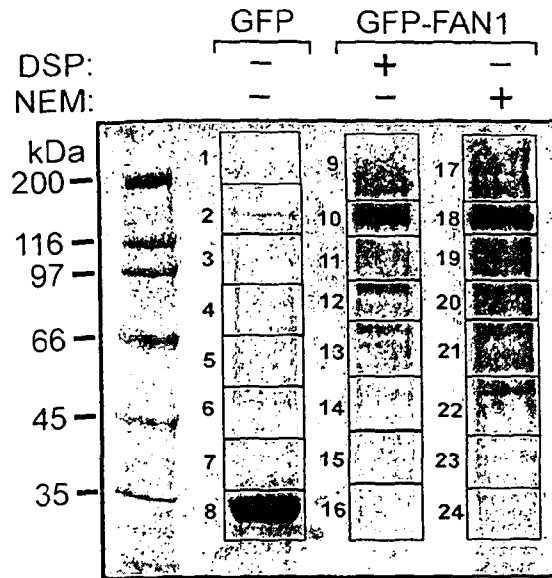
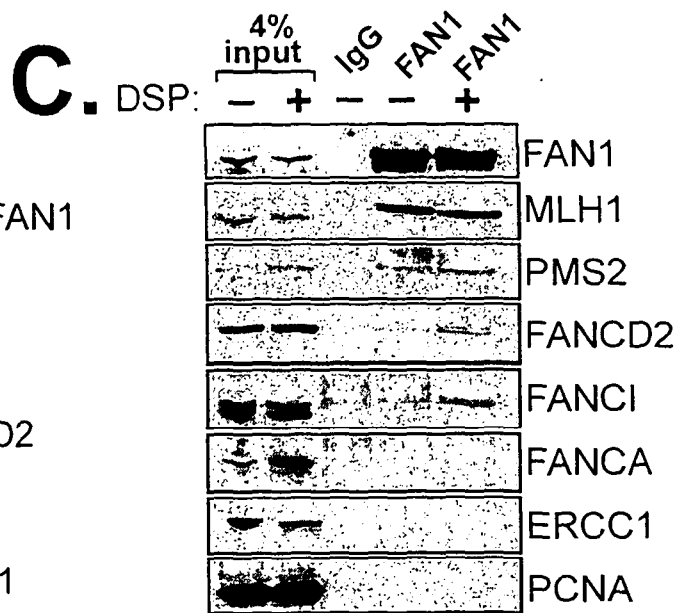
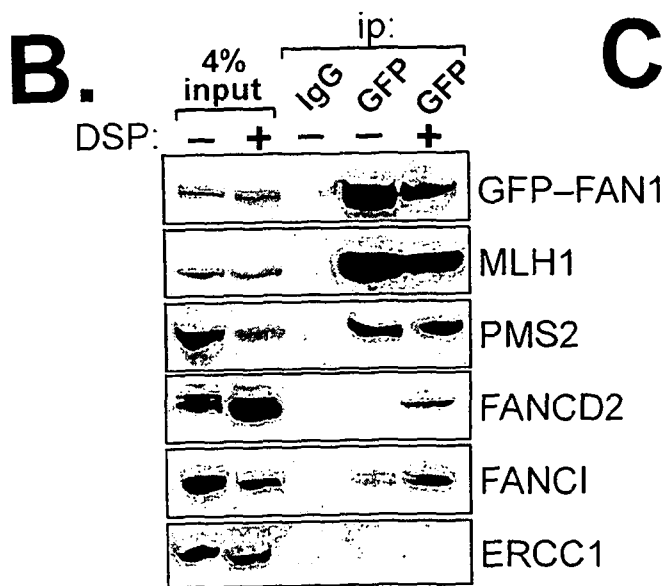


Figure 4



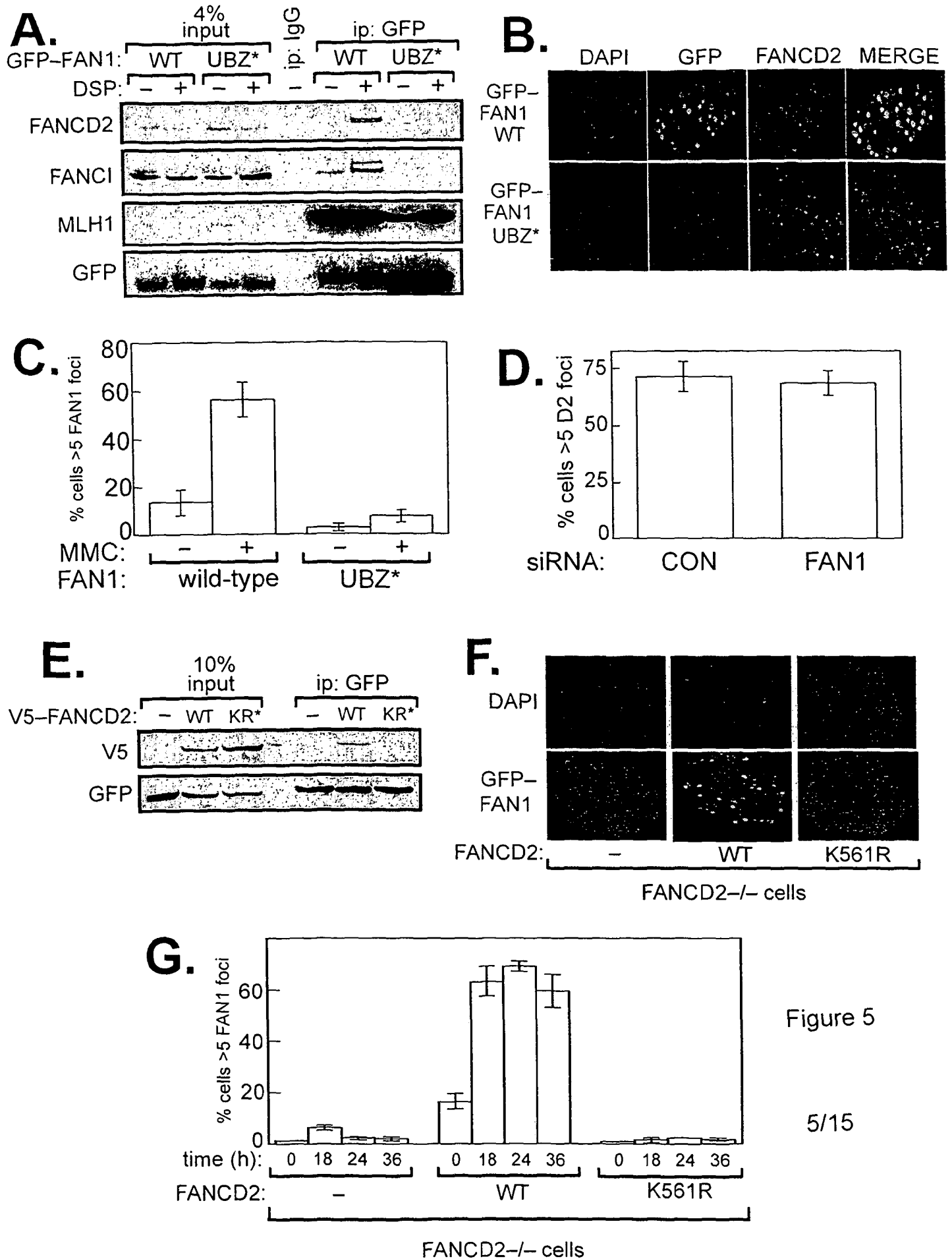


Figure 5

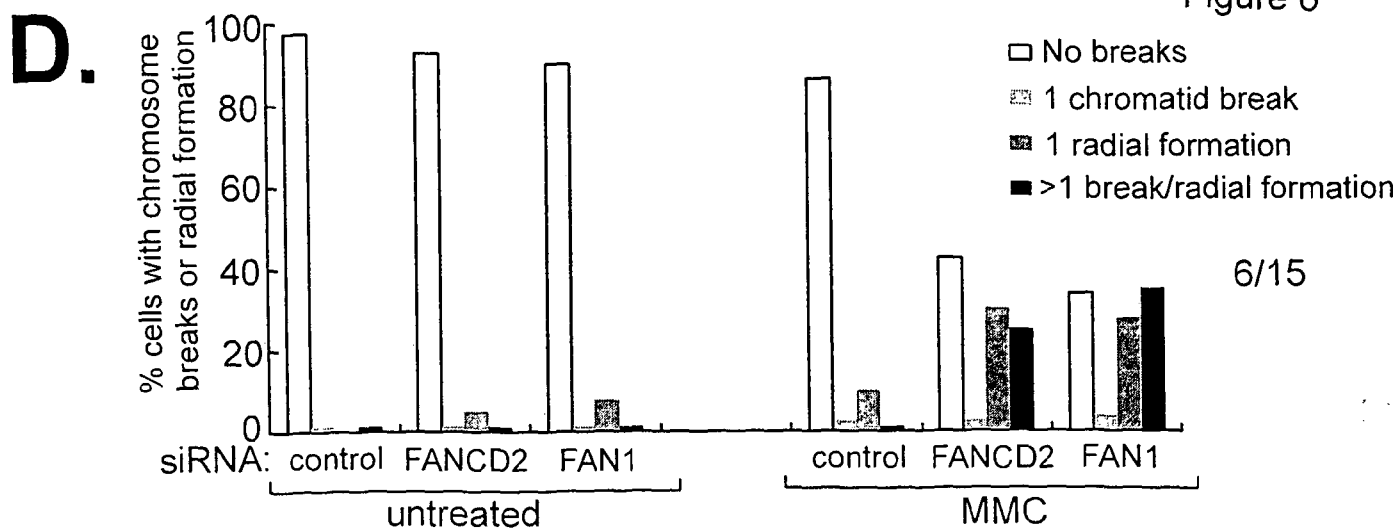
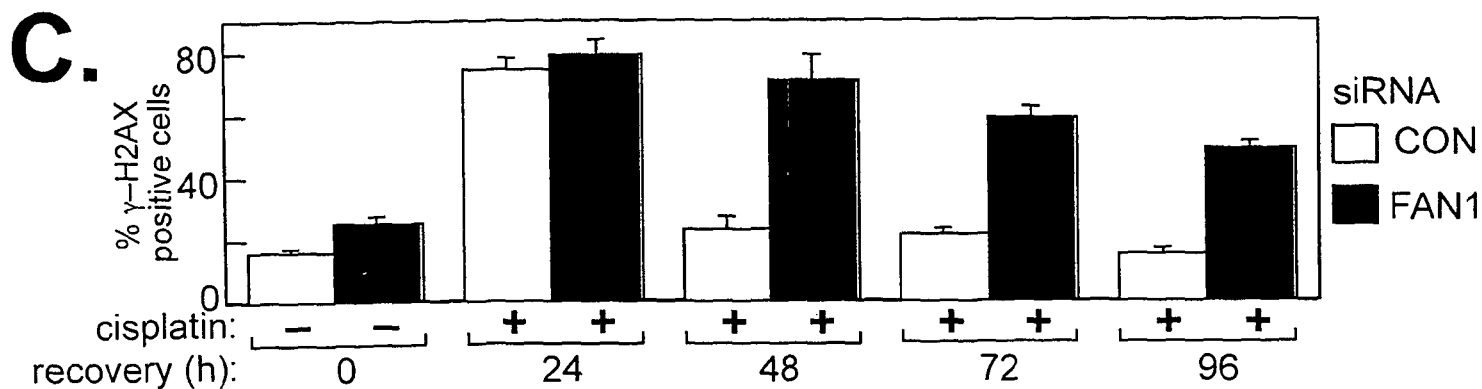
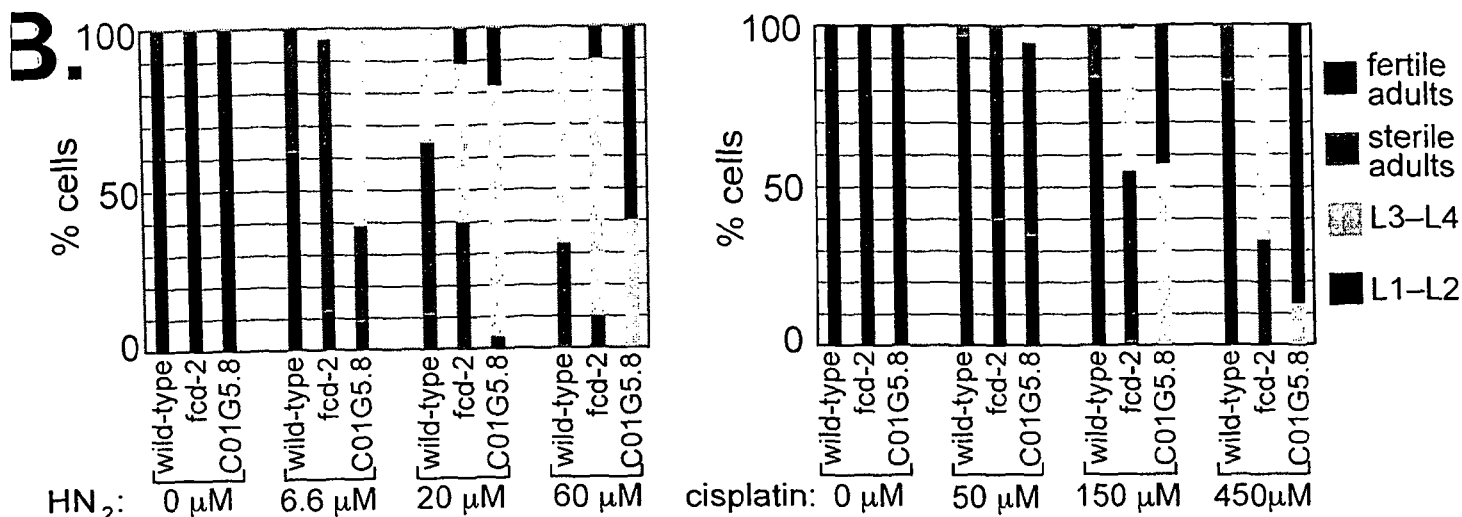
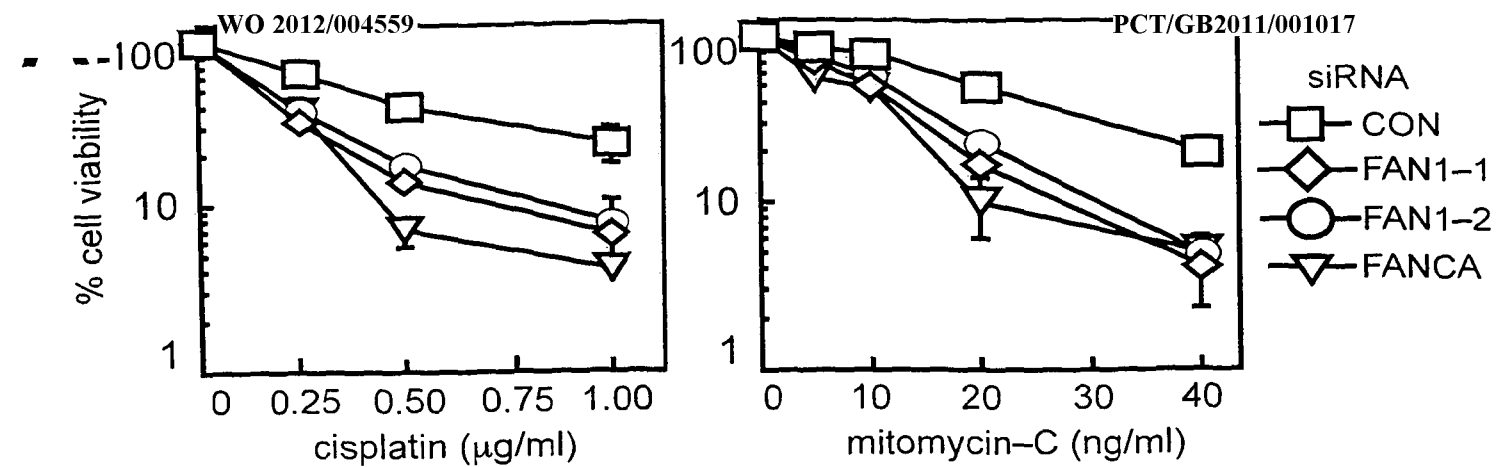
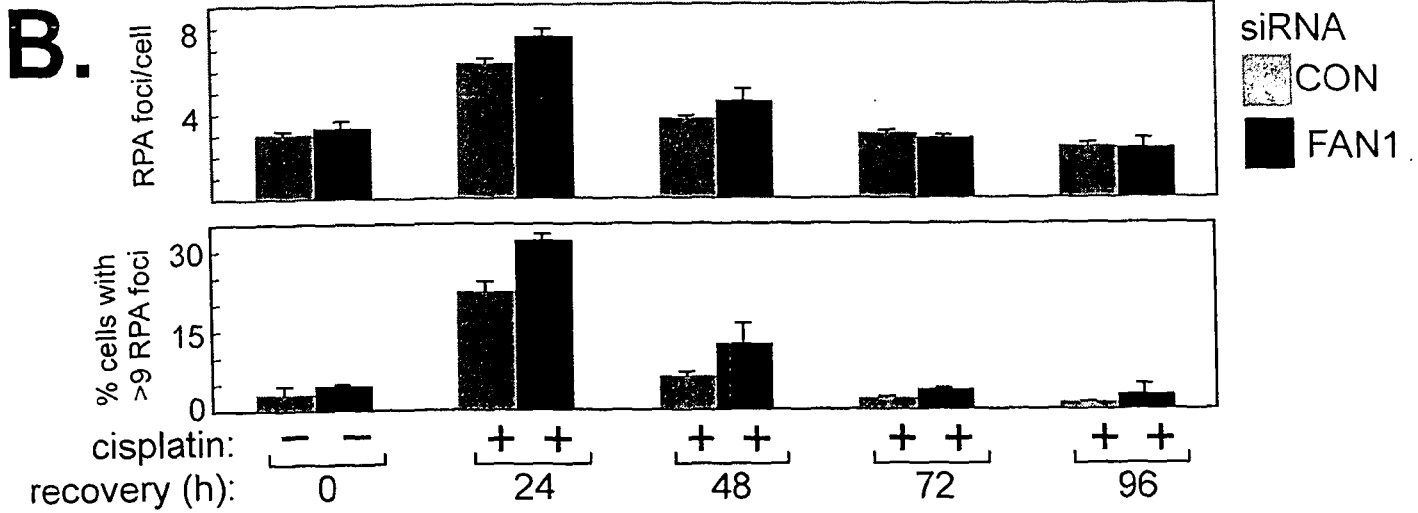
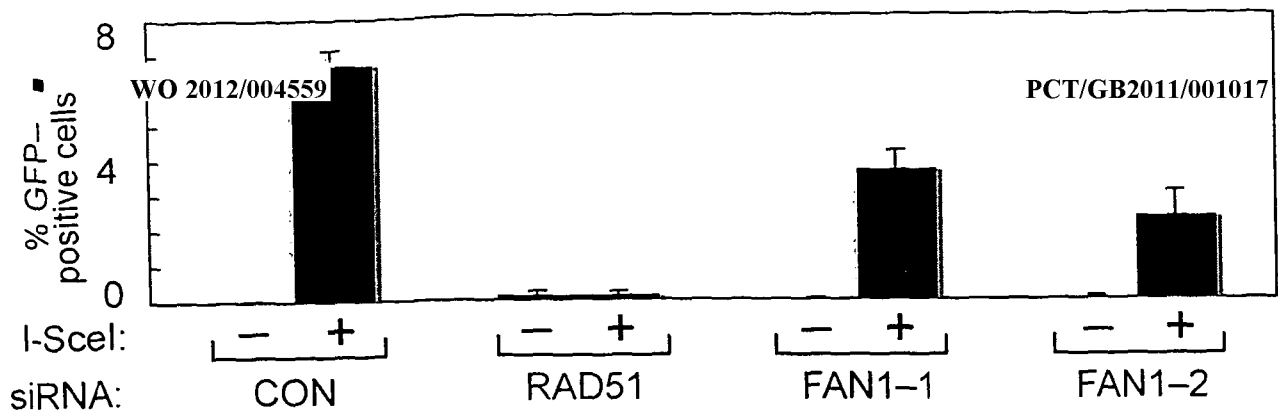


Figure 6



7/15

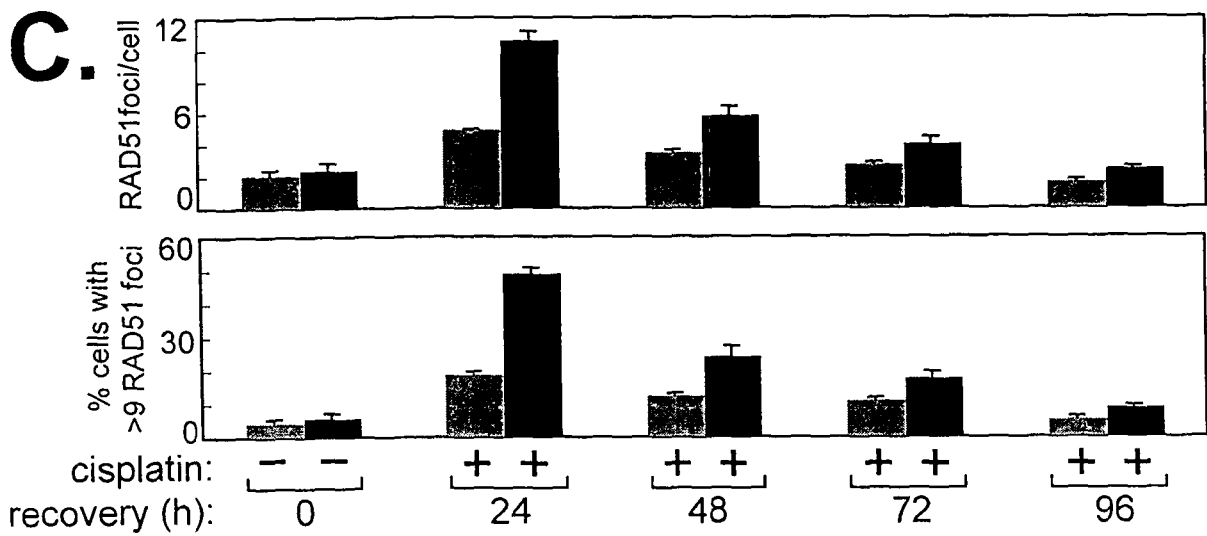
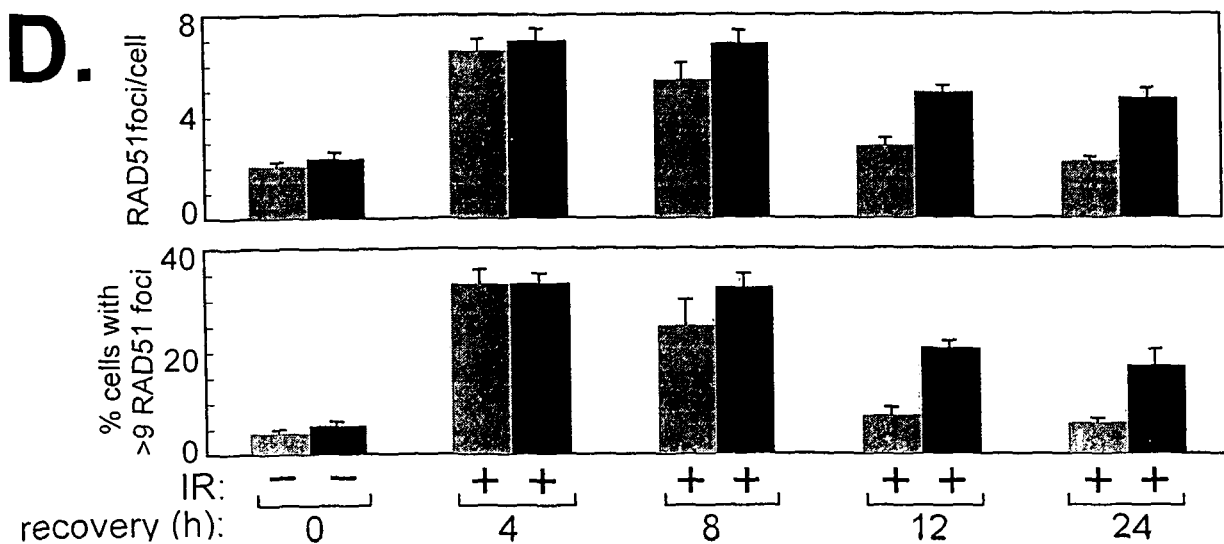
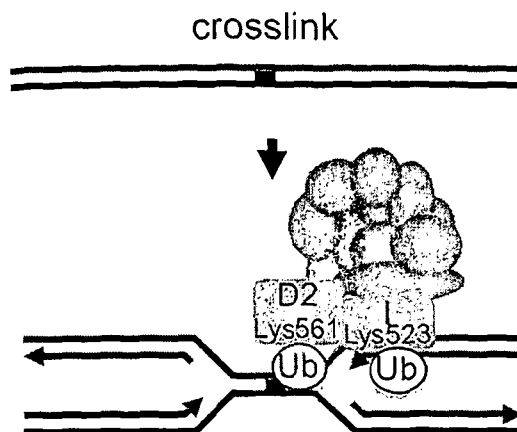


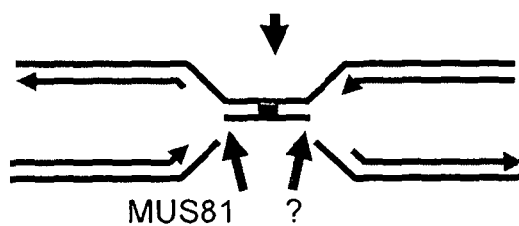
Figure 7



DNA replication
forks blocked

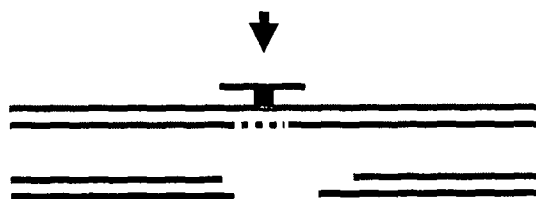


recruitment of
Fanconi anemia
complex

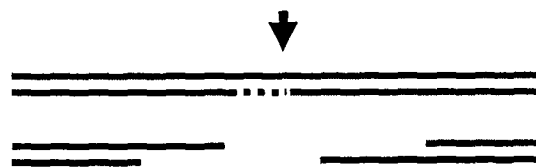


unhooking

Figure 8

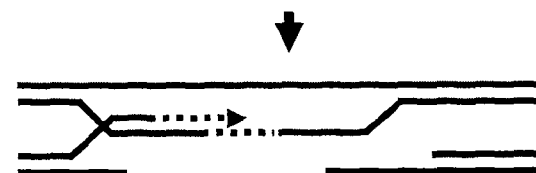


translesion
synthesis

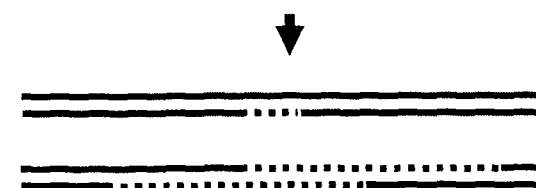


adduct excision

DSB resection



homologous
recombination



completion

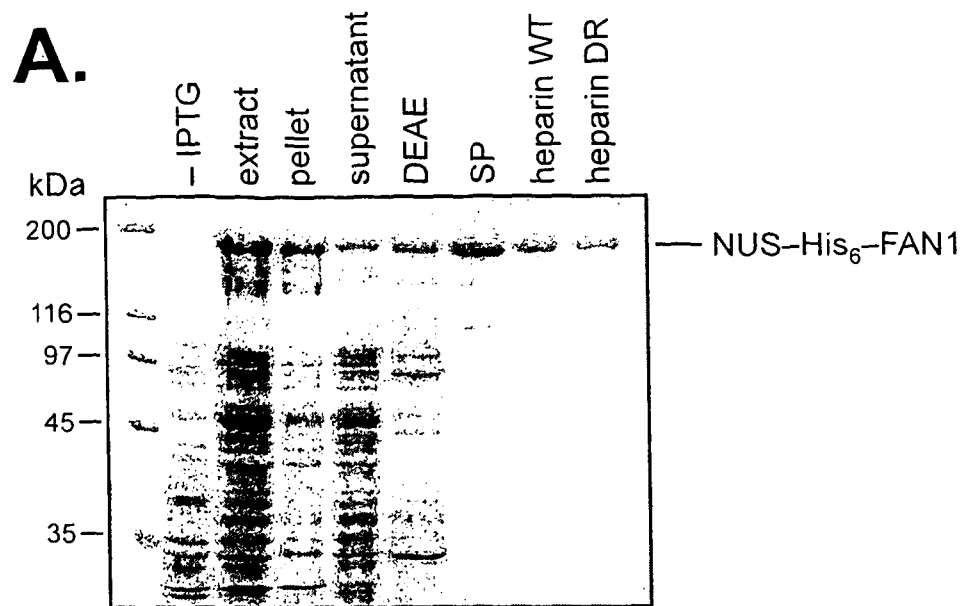
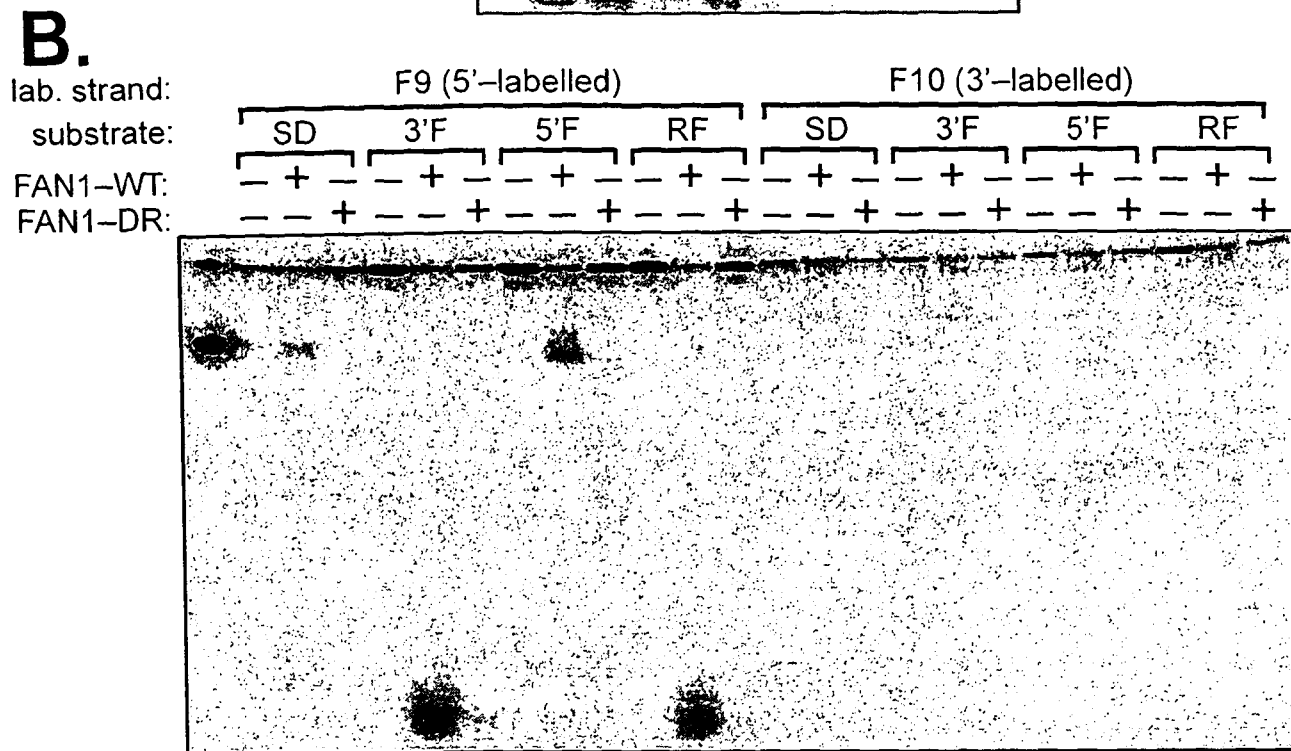
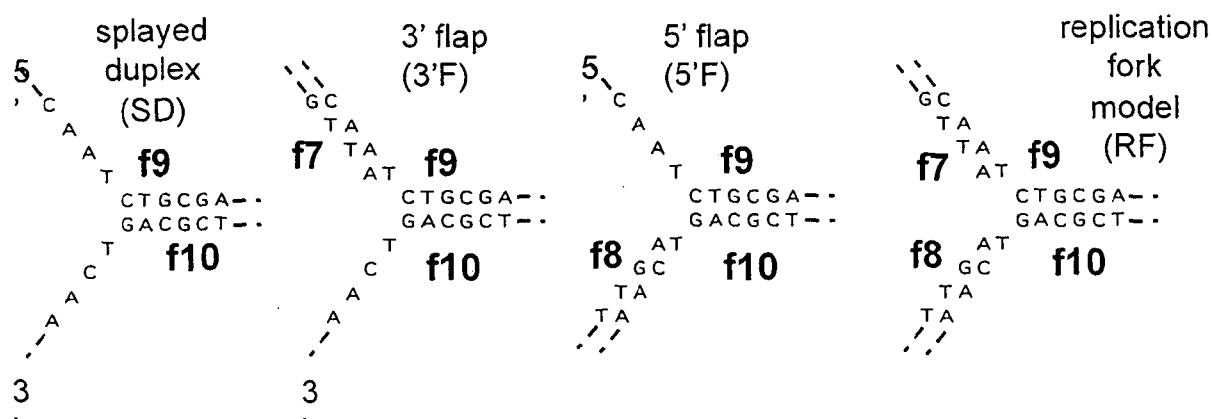


Figure 9



C.



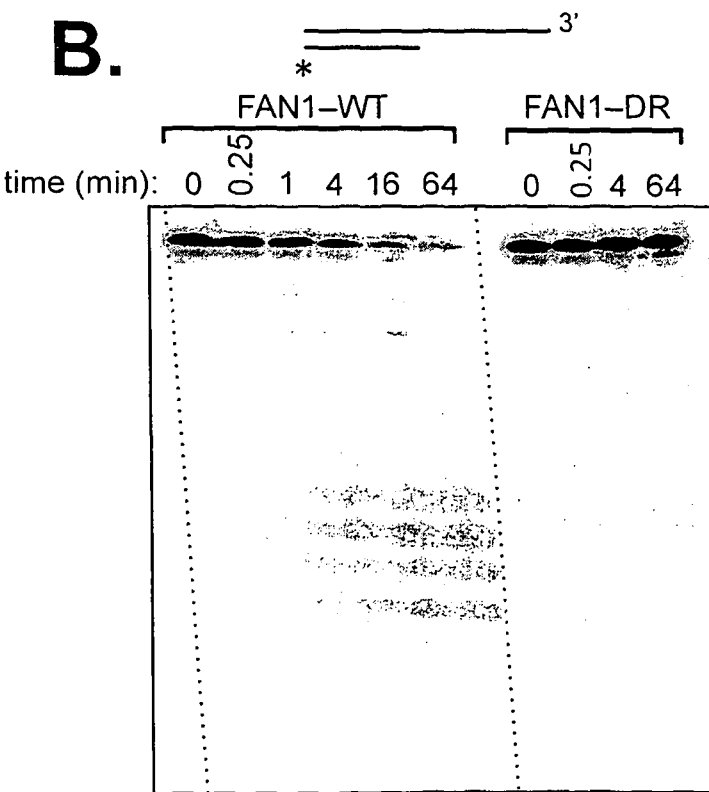
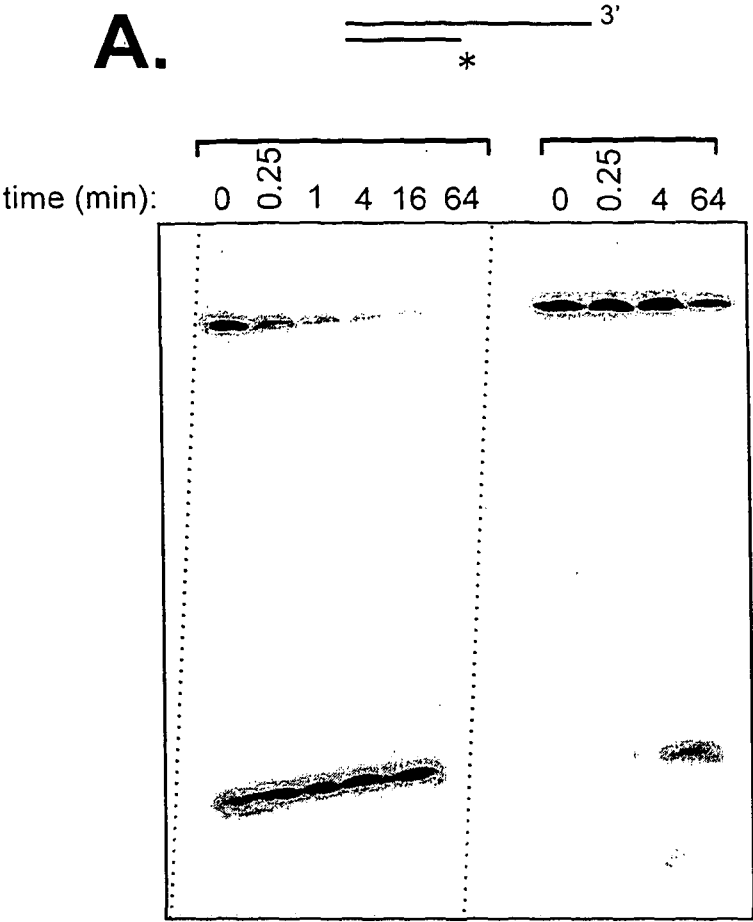


Figure 10

11/15

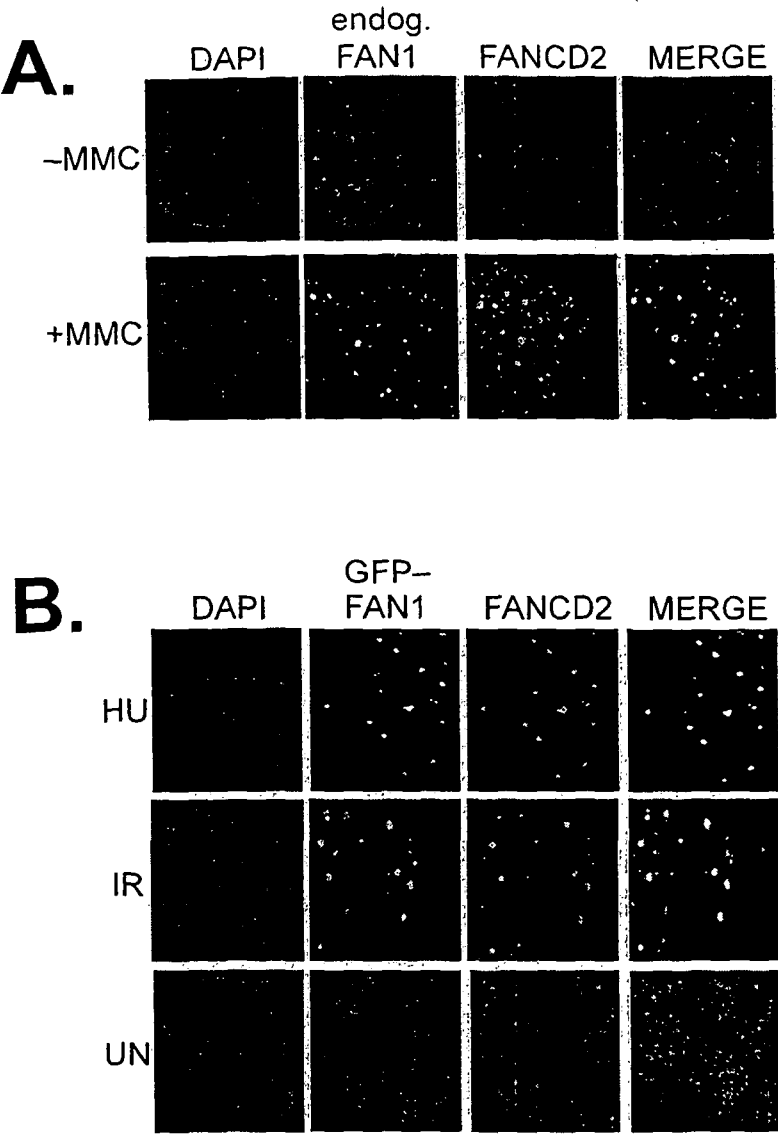


Figure 11

12/15

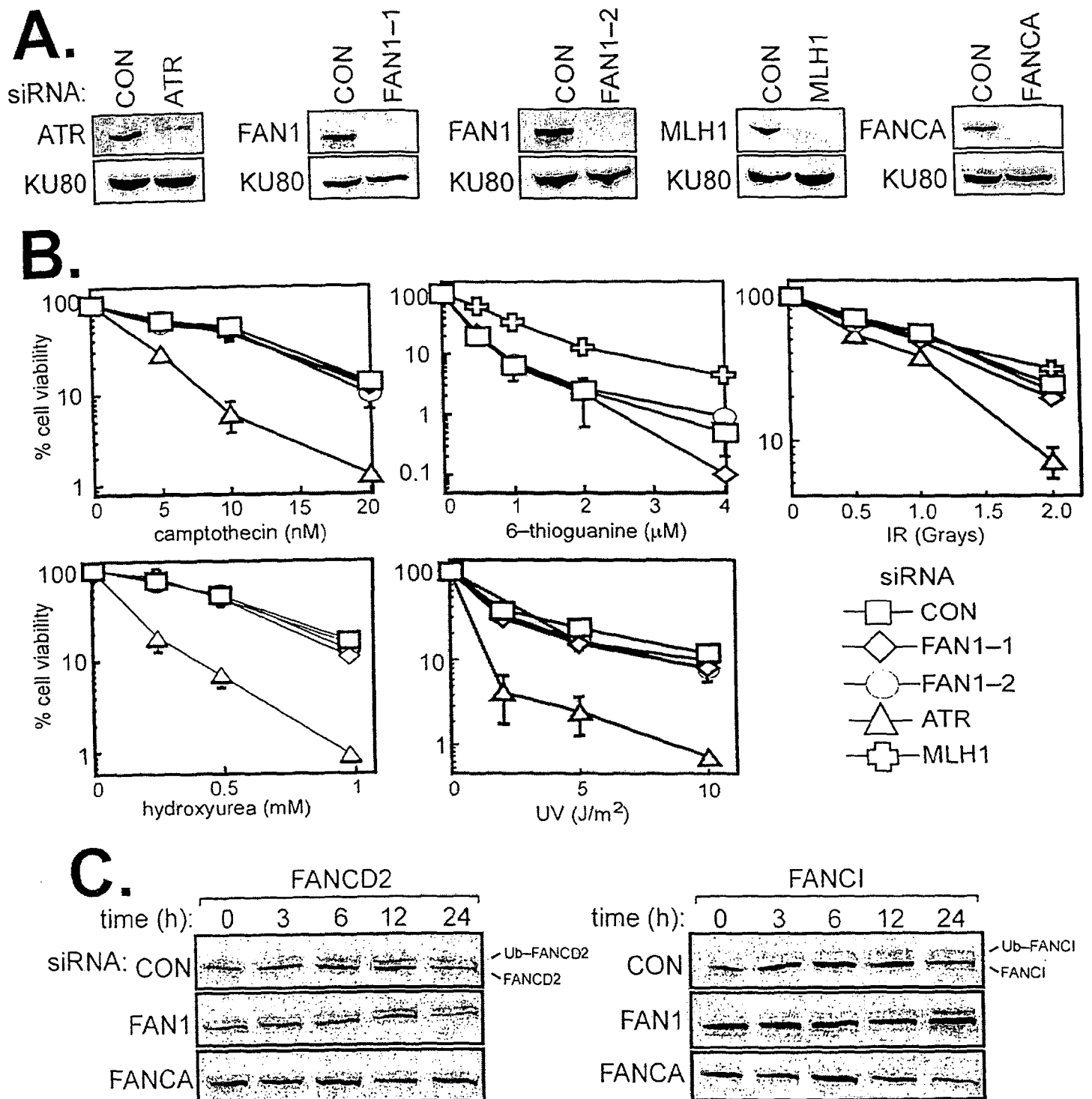


Figure 12

13/15

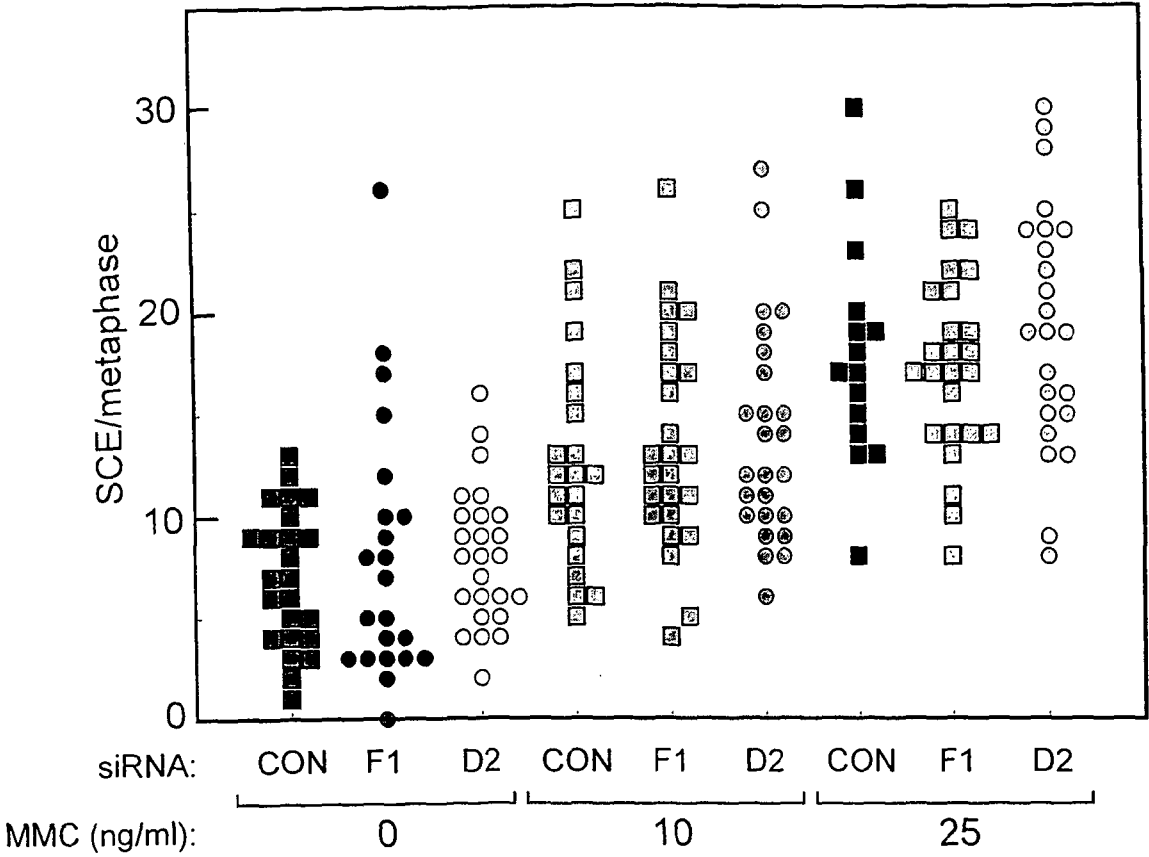


Figure 13

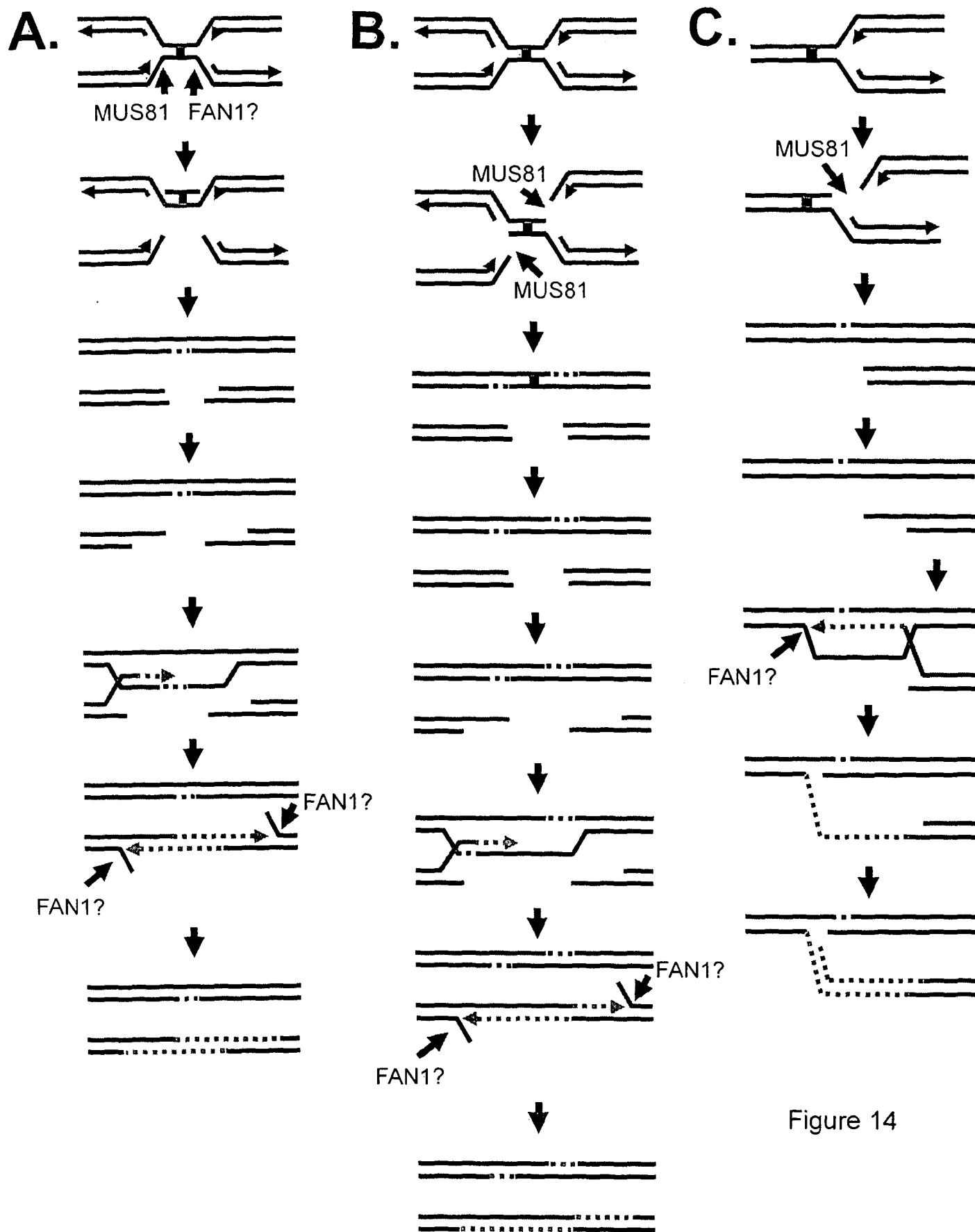


Figure 14

DSP	Band No.	Accession Number	Protein Name	Molecular Mass (Da)	No. of Peptides	Mascot Score
	9	Q9BXW9	FANCD2	168295	63	1157
	9	P62988	Ubiquitin	12245	51	565
	10	Q9Y2M0	KIAA1018	115977	692	10488
	10	Q9NV11	FANCI	150769	30	219
	11	P54278	PMS2	96935	153	2319
	11	P54277	PMS1	106619	178	2211
	11	Q14974	KPNB1	98420	86	1159
	12	P40692	MLH1	85175	265	4562
	13	O15287	FANCG	69423	2	100
	15	P40938	RFC3	41328	4	244
	15	P345249	RFC4	40170	14	129
	16	P62258	14-3-3 epsilon	29326	81	1306
	16	P63104	14-3-3 delta	27899	39	801
	16	P12004	PCNA	29092	57	774
NEM	17	P62988	Ubiquitin	12245	37	402
	18	Q9Y2M0	KIAA1018	115977	759	11416
	19	P54278	PMS2	96935	165	2470
	19	P54277	PMS1	106619	221	2158
	20	Q14974	KPNB1	98420	50	453
	23	P40692	MLH1	85175	368	6680
	23	P40938	RFC3	41328	6	244
	24	P345249	RFC4	40170	17	164
	24	P12004	PCNA	29092	65	1204
	24	P62258	14-3-3 epsilon	29326	57	1003
	24	P63104	14-3-3 delta	27899	23	430

Figure 15

INTERNATIONAL SEARCH REPORT

International application No
PCT/GB2011/001017

A. CLASSIFICATION OF SUBJECT MATTER INV. C12Q1/44 G01N33/566 G01N33/574 C12N9/00 C12Q1/68 ADD.		
According to International Patent Classification (IPC) or to both national classification and IPC		
B. FIELDS SEARCHED Minimum documentation searched (classification system followed by classification symbols) C12Q G01N C12N Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched Electronic data base consulted during the international search (name of data base and, where practical, search terms used) EPO-Internal , COMPENDEX, EMBASE, FSTA, INSPEC, WPI Data		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X,P	CRAIG MACKAY ET AL: "Identi f i c a t i o n of KIAA1018/FAN1, a DNA Repair Nuclease Recruited to DNA Damage by Monoubiquitinated FANCD2", CELL, vol . 142, no. 1, 8 July 2010 (2010-07-08) , pages 65-76, XP55009427 , ISSN: 0092-8674, DOI : 10.1016/j .cel l .2010.06.021 the whole document ----- -/- .	1-30
☒ Further documents are listed in the continuation of Box C. ☒ See patent family annex.		
* Special categories of cited documents : "A" document defining the general state of the art which is not considered to be of particular relevance "E" earlier document but published on or after the international filing date "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) "O" document referring to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art. "&" document member of the same patent family		
Date of the actual completion of the international search 2 November 2011		Date of mailing of the international search report 14/11/2011
Name and mailing address of the ISA/ European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016		Authorized officer Jacques , Patri ce

INTERNATIONAL SEARCH REPORT

International application No
PCT/GB2011/001017

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X, P	T. LIU ET AL: "FAN1 Acts with FANCD2 to Promote DNA Interstrand Cross-Link Repair", SCIENCE, vol . 329 , no. 5992 , 29 July 2010 (2010-07-29) , pages 693-696, XP55009430, ISSN: 0036-8075 , DOI : 10.1126/sci ence. 1192656 the whol e document -----	1-30
X, P	K. YOSHIKIYO ET AL: "KIAA1018/FAN1 nuclease protects cells against genomic instability induced by interstrand cross-linking agents" , PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES, vol . 107 , no. 50, 14 December 2010 (2010-12-14) , pages 21553-21557 , XP55009531 , ISSN: 0027-8424, DOI : 10.1073/pnas .1011081107 the whol e document -----	1-30
X, P	AGATA SMOGORZEWSKA ET AL: "A Genetic Screen Identifies FAN1, a Fanconi Anemia-Associated Nuclease Necessary for DNA Interstrand Crosslink Repair", MOLECULAR CELL, vol . 39 , no. 1, 8 July 2010 (2010-07-08) , pages 36-47 , XP55009530, ISSN: 1097-2765 , DOI : 10.1016/j .mol cel .2010.06.023 the whol e document -----	1-30
X, P	KATJA KRATZ ET AL: "Deficiency of FANCD2-Associated Nuclease KIAA1018/FAN1 Sensitizes Cells to Interstrand Crosslinking Agents" , CELL, vol . 142 , no. 1, 8 July 2010 (2010-07-08) , pages 77-88, XP55009534, ISSN: 0092-8674, DOI : 10.1016/j .cel l .2010.06.022 the whol e document ----- -/--	1-30

INTERNATIONAL SEARCH REPORT

International application No
PCT/GB2011/001017

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	E. CANNAVO ET AL: "Characteri zati on of the Interactome of the Human MutL Homologues MLH1, PMS1, and PMS2", JOURNAL OF BIOLOGICAL CHEMISTRY, vol . 282, no. 5, 1 February 2007 (2007-02-01) , pages 2976-2986, XP55010435 , ISSN: 0021-9258, DOI : 10. 1074/jbc.M609989200 the whol e document -----	1-30
A	P. KNI PSCHER ET AL: "The Fanconi Anemia Pathway Promotes Repl i cati on-Dependent DNA Interstrand Cross-Li nk Repai r ", SCI ENCE, vol . 326, no. 5960, 12 November 2009 (2009-11-12) , pages 1698-1701 , XP55009468, ISSN: 0036-8075 , DOI : 10. 1126/sci ence. 1182372 the whol e document -----	1-30
A	US 2009/062196 AI (D ANDREA ALAN [US] ET AL) 5 March 2009 (2009-03-05) the whol e document -----	1-30
A	US 2009/305266 AI (BAGBY GROVER C [US] ET AL) 10 December 2009 (2009-12-10) the whol e document -----	1-30

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/GB2011/001017

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2009062196	AI	05-03-2009	AU 2007325900 AI 05-06-2008
			CA 2703006 AI 05-06-2008
			EP 2092083 A2 26-08-2009
			JP 2010506939 A 04-03-2010
			WO 2008066624 A2 05-06-2008

US 2009305266	AI	10-12-2009	WO 2007130534 A2 15-11-2007
