The Simple Chordate *Ciona intestinalis* Has a Reduced Complement of Genes Associated with Fanconi Anemia

Edward C. Stanley[1], Paul A. Azzinaro[2], David A. Vierra[2], Niall G. Howlett[2,3] and Steven Q. Irvine[1,4]

[1]Integrative and Evolutionary Biology Graduate Specialization, [2]Cell and Molecular Biology Graduate Specialization, [3]Department of Cell and Molecular Biology, [4]Department of Biological Sciences, University of Rhode Island, Kingston, RI, USA.

ABSTRACT: Fanconi anemia (FA) is a human genetic disease characterized by congenital defects, bone marrow failure, and increased cancer risk. FA is associated with mutation in one of 24 genes. The protein products of these genes function cooperatively in the FA pathway to orchestrate the repair of DNA interstrand cross-links. Few model organisms exist for the study of FA. Seeking a model organism with a simpler version of the FA pathway, we searched the genome of the simple chordate *Ciona intestinalis* for homologs of the human FA-associated proteins. BLAST searches, sequence alignments, hydropathy comparisons, maximum likelihood phylogenetic analysis, and structural modeling were used to infer the likelihood of homology between *C. intestinalis* and human FA proteins. Our analysis indicates that *C. intestinalis* indeed has a simpler and potentially functional FA pathway. The *C. intestinalis* genome was searched for candidates for homology to 24 human FA and FA-associated proteins. Support was found for the existence of homologs for 13 of these 24 human genes in *C. intestinalis*. Members of each of the three commonly recognized FA gene functional groups were found. In group I, we identified homologs of FANCE, FANCL, FANCM, and UBE2T/FANCT. Both members of group II, FANCD2 and FANCI, have homologs in *C. intestinalis*. In group III, we found evidence for homologs of FANCJ, FANCO, FANCQ/ERCC4, FANCR/RAD51, and FANCS/BRCA1, as well as the FA-associated proteins ERCC1 and FAN1. Evidence was very weak for the existence of homologs in *C. intestinalis* for any other recognized FA genes. This work supports the notion that *C. intestinalis*, as a close relative of vertebrates, but having a much reduced complement of FA genes, offers a means of studying the function of certain FA proteins in a simpler pathway than that of vertebrate cells.

KEYWORDS: invertebrate, DNA repair, tunicate

CITATION: Stanley et al. The Simple Chordate *Ciona intestinalis* Has a Reduced Complement of Genes Associated with Fanconi Anemia. *Evolutionary Bioinformatics* 2016:12 133–148 doi: 10.4137/EBO.S37920.

TYPE: Original Research

RECEIVED: November 16, 2015. **RESUBMITTED:** February 10, 2016. **ACCEPTED FOR PUBLICATION:** February 16, 2016.

ACADEMIC EDITOR: Jike Cui, Associate Editor

PEER REVIEW: Seven peer reviewers contributed to the peer review report. Reviewers' reports totaled 1,840 words, excluding any confidential comments to the academic editor.

FUNDING: This study was supported by funding from University of Rhode Island Council for Research grant (SQI); National Institutes of Health/National Heart, Lung and Blood Institute grant R01HL101977 (NGH); Rhode Island IDeA Network of Biomedical Research Excellence (RI-INBRE) grant P20GM103430 from the National Institute of General Medical Sciences; Rhode Island Research Alliance Collaborative Grant Award; and Rhode Island Experimental Program to Stimulate Competitive Research (RI-EPSCoR) grant #1004057 from the National Science Foundation. The authors confirm that the funder had no influence over the study design, content of the article, or selection of this journal.

COMPETING INTERESTS: Authors disclose no potential conflicts of interest.

CORRESPONDENCE: sirvine@uri.edu; nhowlett@uri.edu

Introduction

DNA repair mechanisms are a major way by which organisms avoid mutations that can lead to disease, especially cancer. However, the complexity of DNA repair pathways has hindered progress in fully understanding how they work. We have examined the genome of the simple chordate animal, *Ciona intestinalis*, which is the closest invertebrate relative of vertebrates, for genes associated with the repair of DNA interstrand cross-links (ICL repair), to see if it might possess a simplified version of this DNA repair mechanism. Fanconi anemia (FA) is clinically characterized by congenital abnormalities, pediatric bone marrow failure, and increased cancer risk during early adulthood. FA is caused by mutation of one of the 19 genes linked in a complex pathway. The proteins encoded by the FA genes function together in the process of ICL repair and in the maintenance of genome stability.[1–3] ICLs are highly toxic lesions that covalently link DNA strands, thereby imposing a direct physical block to DNA replication and RNA transcription. The FA protein interaction network is extensive and includes numerous other proteins that function in ICL repair, which have not been genetically linked to FA.[4]

The FA pathway proteins have been categorized into three distinct groups[3]: group I represents the FA core complex and comprises FANCA, FANCB, FANCC, FANCE, FANCF, FANCG, FANCL, FANCM, and FANCT/UBE2T. The FA core complex catalyzes the site-specific monoubiquitination of the FANCD2 and FANCI (group II) proteins.[5–7] FANCL is a RING domain containing E3 ubiquitin ligase,[8,9] while UBE2T is an E2 ubiquitin-conjugating enzyme.[10] FANCM is a large (230 kDa) scaffold protein that possesses DNA binding and ATPase/translocase activities.[11,12] The functions of the remaining group I proteins remain poorly understood. The group II proteins FANCD2 and FANCI, when monoubiquitinated, facilitate the recruitment of several key DNA repair proteins, including FAN1, FANCP/SLX4, and CtIP, to the ICL.[13–18] The group III FA proteins comprise FANCD1/BRCA2, FANCJ/BRIP1, FANCN/PALB2,

FANCO/RAD51C, FANCP/SLX4, FANCQ/ERCC4, FANCR/RAD51, and FANCS/BRCA1 and function downstream of FANCD2 and FANCI monoubiquitination. These proteins function primarily in the homologous recombination (HR) step of ICL repair. For example, FANCD1/BRCA2, FANCN/PALB2, and FANCO/RAD51C regulate the localization and activity of FANCR/RAD51, a well established and key HR protein.[19–25] Several of the FA proteins are ubiquitous among the eukaryotes.[26] Almost every organism surveyed possesses both of the group II proteins, as well as FANCL, FANCM, and an associated ubiquitin-conjugating (E2) enzyme (Fig. 1). There is no apparent evolutionary pattern associated with the presence or absence of the group I proteins outside of the vertebrates, as some are found in insects, while others are seen in plants and red algae before seemingly reappearing in *Nematostella* and then again in the vertebrates. Echinoderms, a sister group of the chordates, possess at least four of the group I proteins.

C. intestinalis is a tunicate, the group thought to be the closest invertebrate relative of the vertebrates.[27] *C. intestinalis* has a number of characteristics that make it a promising model for human diseases. Its genome is very compact, at only 115 Mb, fully sequenced, most of which has been mapped to chromosomes. The current genebuild on Ensembl has 16,671 coding genes, as compared with 20,313 in humans.[28] Homologs of almost all human gene families are represented, but *Ciona* does not have the duplicate genes created by the genome duplications that occurred in vertebrates.[29] There are curated databases with abundant gene expression data,[30,31] as well as a proteome database.[32] While in many cases *Ciona* has lost genes reflecting adaptation to its sessile lifestyle,[33] it can still be used to model simplified pathways,[34–36] as it possesses a simplified version of the vertebrate body plan, most notably as a larva.[37]

A previous study focusing on zebrafish[38] looked into the *Ciona* FA pathway and was unable to find most of the genes. The genes that were found were concentrated in groups II and III, making it plausible that *Ciona* could at the very least be used as a model for the latter two-thirds of the pathway. A subset of the vertebrate group I proteins do appear to be present in *Ciona*, according to our study, suggesting that it may possess a minimal FA pathway.

In order to better assess the total complement of FA-associated genes in *C. intestinalis*, we have analyzed the protein structure, hydrophobicity, and phylogenetic relationships of candidates for each of the FA genes of vertebrates. These analyses indicate that *C. intestinalis* has both of the group II genes from vertebrates, as expected, but only one-third of the group I and two-thirds of the group III genes. In comparison with other animals, and even the plant *Arabidopsis*, *C. intestinalis* appears to have an extremely depauperate FA pathway. These data suggest that *C. intestinalis* may be a good model organism to study a simplified FA pathway and gain important insight into the poorly understood molecular basis of the developmental defects of FA patients.

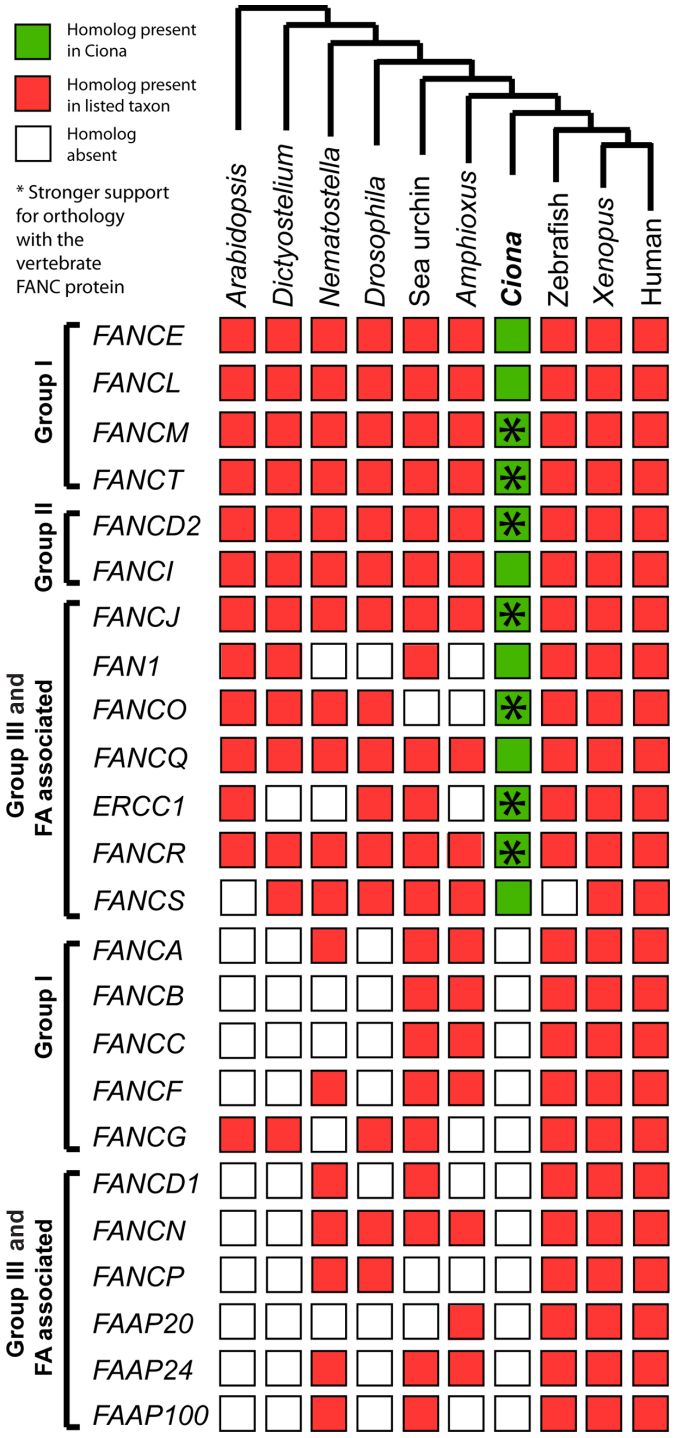

Figure 1. Presence/absence of FA gene orthologs in selected eukaryotes, as determined by this study. Filled boxes denote estimated presence of a gene in that taxon. Outside of *Ciona* and humans, presence/absence was determined only by a Delta-BLAST search of the NCBI database using the human gene as query. The dendrogram at the top of the figure denotes the relationships between organisms.

Materials and Methods

Obtaining sequences. First, a Reciprocal Best BLAST (RBB)[39] search on 24 gene products was performed, searching the human genes of the FA pathway (Table 1) against the *Ciona* proteome, taking the closest match, and then searching

the *Ciona* protein back against the human database to see if the same protein was returned as the closest result. This step was augmented with a search by the reciprocal smallest distance (RSD) method,[40] which in all but three cases returned the same protein as RBB. In these three cases the RSD candidate had a higher percentage of positive matches, so those proteins were the ones listed in Table 1.

BLAT[41] in the JGI genome portal[42] as well as OrthoDB[43] was used to look for synteny between human and *Ciona* FA genes, but none was detected for any of the candidates.

Protein information. Using ClustalX and ClustalΩ,[44] each *Ciona* FA protein sequence was aligned against the human and *Xenopus laevis* sequence. The sequences were imported into Jalview,[45] and the most closely aligned regions were isolated. Hydrophobicity plots of each sequence were created using Biopython and code built and modified from Dalke Scientific.[46] To determine whether the results were significant, the Pearson coefficients were evaluated for the *Ciona* amino acid (aa) sequence against the human and *Xenopus* sequences (again using Python), a beta distribution derived for each sequence,[47] and a comparison of the critical values to a $P < 0.002$ level of significance was made. As a standard, $P < 0.05$ level of significance with 24 tests gives about a 30% chance of a false positive (Type I error), so a more thorough bound of significance was required. The Sidak test,[48] a familywise error correction method used to reduce type I errors, suggests a P-value of $1 - (1 - 0.05)^{1/24}$, or about 0.0021, where 0.05 is the original level of significance and 24 is the number of comparison tests performed. This assumes that the genes and their products are independent – there does not appear to be any evidence that a mutation in one FA protein leads to the absence of any of the other FA proteins.

Protein structural models (Figs. 2E, F, J, and K and 4F and G) were constructed using Discovery Studio v. 3.1 (BIOVIA), based on pdb files in the RCSB Protein Data Bank, using 50 iterations with loop refinement. The protein motif diagrams were based on the information in Pfam 29.0.[49]

Phylogenetic analysis. Full protein sequences (see Supplementary Table S1 for accession numbers) were aligned using MAFFT with default settings.[50] Poorly aligned regions were excised using TrimAI v. 1.3 using the Gappyout setting on the Phylemon 2.0 web server.[51] RAxML v. 8.0.0[52] was used to construct a maximum likelihood (ML) tree with bootstrap number determined with the FC bootstrapping criterion and PROTGAMMABLOSUM62 substitution model. User supplied trees with candidate genes rearranged were statistically evaluated using the Shimodaira–Hasegawa (SH) log likelihood test in RAxML.

Results

***Ciona* has orthologs of vertebrate FA genes from each functional group.** Our analysis revealed that *Ciona* has highly conserved orthologs of genes from each of the three FA protein groups (Fig. 1). Like all the other multicellular organisms examined, *Ciona* has both members of group II: *FANCD2* and *FANCI*. However, only 4 of 9 members of group I and 5 of 8 members of group III were found, as well as only 2 of several "FA associated" proteins. In fact, *Ciona* appears to have as few or fewer members of the FA pathway of any multicellular organism examined, including plants, slime mold, and the primitive metazoan *Nematostella*.

Below we present evidence for or against orthology in *C. intestinalis* of each of the members of the FA pathway. The first analyses described are for those genes that we estimate are present in *Ciona*, organized by the functional group. We then list those that do not have orthologs in *Ciona* according to our methods. The order of the genes in the text is similar to the vertical order in Figure 1.

Group I orthologs found. *FANCE*. FANCE is part of the FA core complex with an unknown function. RBB returns an uncharacterized *C. intestinalis* protein LOC100186252 (XP_002129936). The *Ciona* candidate protein aligns well with the last 250–300 aa of vertebrate FANCE proteins ($R^2 = 0.202$), but on the whole, the correlation is only 0.08 (and the region outside the C-terminal registering at only 0.05; Fig. 2A). The *Ciona* candidate is about 400 aa in length, while vertebrate FANCE proteins are all between 550 and 600 aa. Moderate alignment is seen between the two globular domains in the *Ciona* candidate and the two C-terminal globular regions in the human protein, though no other shared secondary structure is found in the ELM analysis (data not shown). The ML best tree (Fig. 2B) groups the *Ciona* candidate LOC100186252 ("CiUP1") in a sister group to the vertebrate FANCA proteins, more closely related to the plant and fungal candidates. However, if LOC100186252 is forced to group with the vertebrate FANCE proteins (Fig. 2C), the tree is not significantly worse, while moving LOC100186252 more distant from the FANCE clade is statistically worse (Fig. 2D; $P < 0.01$), consistent with the orthology of FANCE. In addition, a crystal structure exists for human FANCE,[53] allowing us to perform structural homology modeling between the human protein and the inferred *C. intestinalis* protein (Fig. 2E and F). The 3D models indicate that the structure of LOC100186252 is potentially very similar to human FANCE. Taken as a whole, these data provide support for LOC100186252 being the homolog of FANCE in *C. intestinalis*.

FANCL. FANCL is an E3 ubiquitin ligase and a component of the FA core complex, which serves to ubiquitinate FANCD2 and FANCI.[9] RBB returns a putative *Ciona* FANCL protein with an E-value of 2×10^{-74} (Table 1). The *Ciona* candidate hydrophobicity plot shows close correspondence to the vertebrate proteins (Fig. 2G). SMART and Pfam primary sequence-based prediction analyses both detect three amino-terminal WD40 repeats and a carboxy-terminal RING domain in *Ciona* fancl (Fig. 2I), similar to that originally described for human FANCL.[9,54] Subsequent structural analyses of *Drosophila* and human FANCL have revealed

Table 1. BLAST (RBB/RSD) results. Refer to Supplemental Table S1 for additional accession numbers of sequences used in phylogenetic and other analyses.

GROUP	HUMAN GENE NAME	HUMAN ACC. NO. BLASTED	BEST *CIONA* GENE MATCH	E-VALUE HS→CI	*CIONA* ACCESSION NO.	RECIPROCAL HUMAN GENE MATCH	E-VALUE CI→HS	RECIPROCAL HUMAN ACC. NO.
I	*FANCA*	NP_000126	Trafficking protein particle complex subunit 10	1×10^{-122}	XP_009858877	Trafficking protein particle complex subunit 10	4×10^{-174}	NP_003265
	FANCB	NP_001018123	Lysine demethylase/ histidyl hydroxylase MINA	0.076	XP_002131324	Lysine demethylase/ histidyl hydroxylase MINA	6×10^{-73}	Q8IUF8
	FANCC	NP_000127	Stabilin 2	5.2	XP_002122507	Stabilin 2	0.0	CAC82105
	FANCE	NP_068741	FA Group E protein C-term. domain, LOC100186252	6×10^{-16}	XP_002129936	Fanconi Anemia Group E protein	7×10^{-65}	NP_068741.1
	FANCF	NP_073562	Peroxidasin-like	4×10^{-146}	XP_009859106	Peroxidasin	$<1.7 \times 10^{-308}$	NP_036425
	FANCG	NP_004620	Stress induced phosphoprotein 1	3×10^{-78}	XP_002128875	Stress induced phosphoprotein 1	3×10^{-77}	NP_006810
	FAAP20	NP_001139782	Polyamine-mod. fact. 1–1 (CiPMF1–1)	12	XP_002127003	Serine/Threonine Protein Kinase MRCK	2×10^{-17}	NP_003598
	FAAP24	NP_689479	DNA polymerase β	1×10^{-145}	XP_002128462	DNA polymerase β	4×10^{-145}	NP_002681
	FAAP100	NP_079437	L-fucose Kinase	0.0002	XP_002122353	L-fucose Kinase	$<1.7 \times 10^{-308}$	NP_659496
	FANCL	AAH09042	Ubiquitin ligase Fanconi Anemia group L	2×10^{-74}	XP_009861960	Ubiquitin ligase Fanconi Anemia group L	2×10^{-81}	Q9 NW38.2
	FANCM	NP_065988	FANCM protein	5×10^{-156}	XP_009862357	FANCM protein	$<1.7 \times 10^{-308}$	AAI44512.1
	UBE2T	NP_054895	UBE2–17kDa	2×10^{-71}	XP_002129339	UBE2D4	2×10^{-77}	XP_006715797
II	*FANCD2*	NP_149075	Fanconi Anemia Complement. Grp D2	$<1.7 \times 10^{-308}$	XP_002130241	Fanconi Anemia Complementation Group D2	$<1.7 \times 10^{-308}$	AAL05980.1
	FANCI	NP_060663	Fanconi Anemia Complement. Grp I	$<1.7 \times 10^{-308}$	XP_009858757	Fanconi Anemia Complementation Grp. I	0.0	ABQ63084.1
III	*FANCD1/ BRCA2*	P51587	Unchar. prot. LOC100185089	6×10^{-12}	XP_002129592	BRCA2	6×10^{-13}	AAB07223
	FANCJ	NP_114432	CiTFIIH-X	5×10^{-157}	XP_002126055[a]	CiTFIIH-X	0.0	NP_000391
	FANCN	Q86YC2	WD repeat containing protein-5 like	2×10^{-94}	XP_002127700	WD repeat containing protein-5 like	7×10^{-116}	P61964
	FANCO/ RAD51C	AAC39604	RAD51 homolog 1 isoform 1	5×10^{-118}	XP_002126934[b]	RAD51 homolog 1 isoform 1	2×10^{-93}	NP_002866.2
	FANCP	NP_115820	Kelch-like protein 10	$<1 \times 10^{-27}$	XP_002122519	Kelch-like protein 20	0.0	NP_055273.2

FANCQ/XPF	Q92889	DNA repair protein XPF	3×10^{-72}	XP_009859502	DNA repair protein XPF	2×10^{-117}	AAB50174
FANCR/ RAD51	CAG38796	RAD51 homolog 1 isoform 1	7×10^{-140}	XP_002130341	RAD51 homolog 1 isoform 1	$<1.7 \times 10^{-308}$	NP_002866.2
FANCS/ BRCA1	NP_009225	BRCA1 putative homolog	1×10^{-33}	KH2012: KH.C9.487[c]	breast cancer 1, early onset, isoform CRA_g	8×10^{-29}	EAW60929
ERCC1	P07992	ERCC1-like	9×10^{-41}	XP_009861832	ERCC1	2×10^{-70}	NP_001974
FAN1	NP_055782	Fanconi Associated Nuclease 1 C-terminal domain	9×10^{-41}	XP_004227197	Fanconi Associated Nuclease 1 C-terminal domain	1×10^{-101}	NP_055782

Notes: [a]Subsequent analysis led to the identification of another sequence, *Ciona* Fanconi anemia group J protein homolog, NCBI acc. no. XP_002120239 as a better proposed orthology of human FANCJ (see text). [b]Subsequent analysis led to the identification of another sequence, *C. intestinalis* DNA repair protein RAD51 homolog 3-like, NCBI acc. no. XP_002130341 as a better proposed orthology of human FANCO (see text). [c]A longer gene model was found in the ANISEED database, to which this accession number corresponds.

Table 2. Hydrophobicity plot correlations between identities and positive matches.

GROUP	NAME	ID%	HYDROPATHY PLOT ID% R^2	POS%	HYDROPATHY PLOT POS% R^2
I	FANCA	36%	0.052	58%	0.027
	FANCB	37%	0.045	52%	0.026
	FANCC	32%	0.121	46%	0.052
	FANCE	19%	0.082	34%	0.051
	FANCF	47%	0.039	65%	0.021
	FANCG	56%	0.034	74%	0.021
	FANCL	36%	0.373	56%	0.199
	FANCM	52%	0.369	66%	0.321
	FAAP20	18%	0.053	28%	0.027
	FAAP24	26%	0.177	45%	0.098
	FAAP100	35%	0.114	53%	0.053
	UBE2T	34%	0.311	58%	0.222
II	FANCD2	25%	0.304	44%	0.222
	FANCI	29%	0.331	50%	0.237
III	FANCD1	31%	0.127	49%	0.046
	FANCJ	37%	0.291	53%	0.236
	FANCN	18%	0.099	45%	0.061
	FANCO	41%	0.328	44%	0.345
	FANCP	32%	0.050	49%	0.030
	FANCQ/ XPF	50%	0.566	68%	0.446
	FANCR/ RAD51	82%	0.844	92%	0.847
	FANCS/ BRCA1	27%	0.117	46%	0.322
	ERCC1	47%	0.396	65%	0.446
	FAN1	41%	0.447	63%	0.387

Note: ID% refers to the Delta-BLAST result for identical amino acid matches. Pos% refers to Delta-BLAST results for positive amino acid matches, eg, aa from the same functional groups.

that FANCL encompasses three distinct domains: an amino-terminal E2-like fold, a central double RWD-like domain, and a carboxy-terminal RING domain.[55,56] Structural homology modeling of *Ciona* fancl, based on the 3.2 Å *Drosophila melanogaster* FANCL structure (PDB ID 3K1L),[55] indicates the existence of close structural similarity (Fig. 2J and K). In addition, Clustal Omega multiple sequence alignment (MSA) analyses of human, mouse, and *Ciona* FANCL indicate that K22, a predicted site of autoubiquitination, is conserved in all three species (data not shown). The ML best tree (Fig. 2H) agrees with this finding, showing that the *C. intestinalis* candidate falls in a clade with the vertebrate FANCL proteins to the exclusion of the second most similar *Ciona* and human proteins. However, moving the *C. intestinalis* candidate further from the vertebrate FANCL clade, or as a sister taxon to the vertebrate FANCL genes, does not make for a statistically worse tree (data not shown). This ambiguity indicates that the phylogenetic evidence for orthology is weak. However, based on the

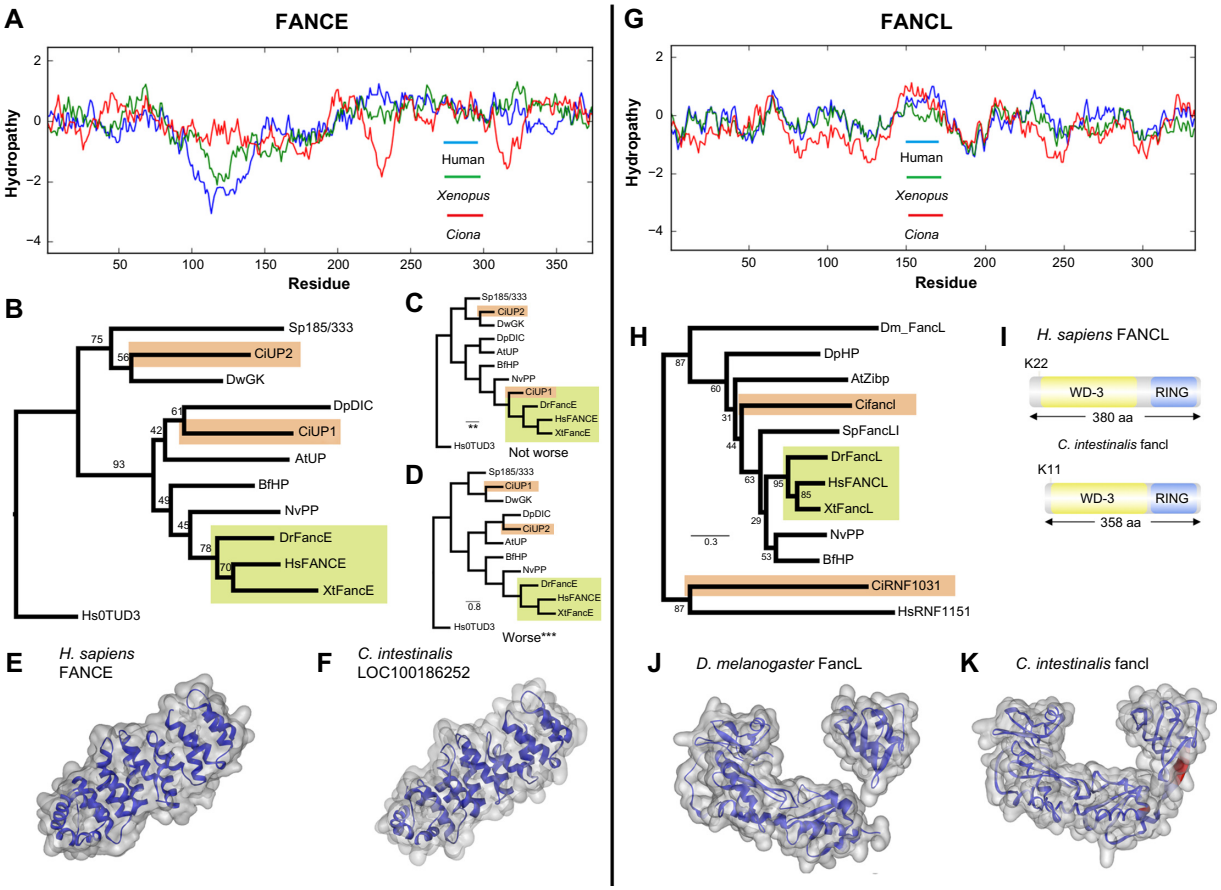

Figure 2. Analysis of FANCE (**A**–**F**) and FANCL (**G**–**K**) putative homologs in *C. intestinalis*. (**A**) Hydropathy plot of best aligning regions in human, *Xenopus*, and *Ciona* putative homologs for FANCE. (**B**) Best ML tree for alignment of FANCE and putative homologs in *C. intestinalis* and other eukaryotes. CiUP1 (LOC100186252) has 93% bootstrap support for membership in the clade with vertebrate FANCE proteins. (**C**) Forcing CiUP1 into the vertebrate FANCE clade does not result in a statistically worse tree, whereas if the locations of the two best *C. intestinalis* BLAST matches to FANCE are switched in the ML tree (**D**), the tree is worse at the $P < 0.01$ level, giving further support to LOC100186252 as the homolog of FANCE. (**E,F**) Structural modeling of human FANCE and *C. intestinalis* LOC100186252, showing extreme similarity of overall structures. (**G**) Hydropathy plot of best aligning regions in human, *Xenopus*, and *Ciona* putative homologs for FANCL. (**H**) Best ML tree for alignment of putative FANCL homologs, showing 87% bootstrap support for Cifancl clustering with vertebrate and other FANCL proteins. (**I**) Diagrammatic comparison of human and *C. intestinalis* FANCL inferred protein motifs. (**J,K**) Modeling of *D. melanogaster* and *C. intestinalis* FANCL protein structures.

structural similarities, there is reasonably strong support for the *C. intestinalis* gene to be a true ortholog of human *FANCL*.

FANCM. FANCM is also a component of the FA core complex and plays a key role in DNA replication fork remodeling and the chromatin recruitment of the group I proteins during ICL repair.[11,57–61] RBB returns a putative *Ciona* FANCM protein as the closest match. Secondary structure analysis shows that both the human and *Ciona* candidate proteins possess a DEAH-box helicase/DNA-stimulated ATPase domain (Fig. 3B). The human FANCM protein also possesses a degenerate XPF/ERCC4 endonuclease domain that the *Ciona* protein lacks.[12] The hydrophobicity plot shows high levels of correlation, especially toward the amino-terminus (Fig. 3A). In the ML tree, the *Ciona* FANCM candidate clusters with the vertebrate FANCM proteins in a clade with 94% bootstrap support (Fig. 3C). These data indicate strong support for the orthology of the *C. intestinalis* candidate.

FANCT/UBE2T. FANCT/UBE2T is one of the many E2 ubiquitin-conjugating enzymes found in the human proteome and is the specific one implicated in the monoubiquitination of FANCD2 and FANCI.[10] In humans, UBE2T interacts with FANCL to ubiquitinate FANCD2. Patient-derived mutations in the *UBE2T* gene have recently been discovered in two unrelated patients, leading to a call to denote *UBE2T* as *FANCT*.[62]

The Delta-BLAST search returns *Ciona* ube2–17kd as the closest match to human UBE2T. However, the reciprocal BLAST against human proteins returns human UBE2D4 with an E-value of 2×10^{-77} (Table 1). The RSD method returns *Ciona* ube2 J11 with an E-value of 9×10^{-76}. Apparently, these very similar E2 ubiquitin-conjugating enzymes cannot be reliably distinguished by BLAST searches (Fig. 3E). The hydropathy and phylogenetic analyses (Fig. 3D and F) do not help to resolve the exact relationship. In the hydropathy plot, it is apparent that both the *Xenopus* and *Ciona* proteins

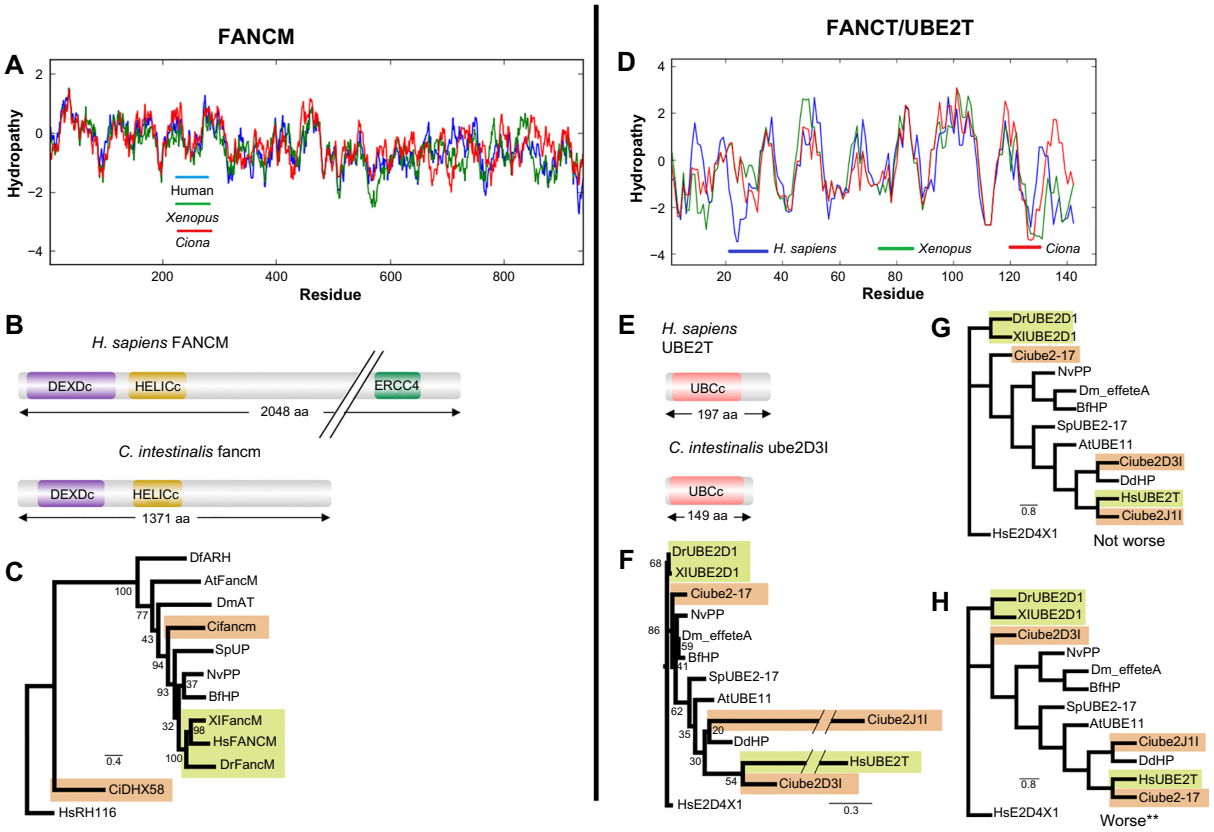

Figure 3. Analysis of FANCM (**A**–**C**) and FANCT/UBE2T (**D**–**F**) putative homologs in *C. intestinalis*. (**A**) Hydropathy plot of best aligning regions in human, *Xenopus*, and *Ciona* putative homologs for FANCM. (**B**) Diagrammatic comparison of human and *C. intestinalis* FANCM inferred protein motifs. (**C**) Best ML tree for alignment of FANCM and putative homologs in *C. intestinalis* and other eukaryotes. Cifancm has 100% bootstrap support for membership in the clade with vertebrate FANCM proteins, to the exclusion of the next most similar *C. intestinalis* protein. (**D**) Hydropathy plot of best aligning regions in human, *Xenopus*, and *Ciona* putative homologs for UBE2T. (**E**) Diagrammatic comparison of human and *C. intestinalis* UBE2T inferred protein motifs. (**F**) Best ML tree for alignment of putative UBE2T homologs, showing that the best BLAST match to UBE2T, Ciube2–17, may not be as closely related to UBE2T as Ciube2D3l (see text). (**G**) Switching the relationships of Ciube2 J1l with Ciube2D3l does not make a significantly worse tree. (**H**) However, switching the best BLAST hit, Ciube2–17, with Ciube2D3l does result in a significantly worse tree (**$P < 0.02$).

roughly follow the pattern of human UBE2T, but neither closely matches with the hydropathy of the human protein. Curiously, in the ML phylogenetic analysis, the best tree shows human UBE2T clustering with another *Ciona* ube2 protein, *Ciona* ube2D3-like, but not the *Ciona* ube2–17, or ube2J1l proteins that are the best hits in the RBB and RSD analyses (Fig. 3F). If *Ciona* ube2J1l is grouped with human UBE2T, the tree is not statistically worse (Fig. 3G), but if *Ciona* ube2D3l is swapped with *Ciona* ube2–17, the tree does become significantly worse (Fig. 3H). In short, there are multiple ube2 proteins in *Ciona* that have such high similarity to the human UBE2T that they alternately appear as putative homologs in different analytic methods. We suggest that it is likely that one of these performs the same E2 ubiquitin conjugation function as UBE2T does in the human FA pathway.

Both group II genes have orthologs in *Ciona*. *FANCD2*. FANCD2 is one of the proteins monoubiquitinated by FANCL and FANCT/UBE2T during ICL repair.[5,9,10] Both RBB and RSD returned a putative FA complementation group D2 protein in *C. intestinalis* as the closest match for this protein in humans, with the BLAST search returning 25% identity, a 44% match on positives, and an E-value of less than 1.7×10^{-308}, indicating extremely strong similarity (Table 1). The *Ciona* fancd2 protein contains 1394 aa, while the most common isoform in humans is 1451 aa long.

When the sequences are aligned and gaps removed, the smoothed hydrophobicity plots show multiple similarities (Fig. 4A). The proteins have highly similar ($R^2 \geq 0.71$) regions at around aligned *Ciona* aa 100–125, 240–280, 510–540, 660–760, 1010–1045, and 1130–1170. Both the human and *Ciona* proteins show five globular domains with moderate alignment. The phylogenetic analysis groups the *C. intestinalis* fancd2 candidate with vertebrate, fly, urchin, and amphioxus putative orthologs, although at low bootstrap support (Fig. 4B). Forcing the *C. intestinalis* candidate out of the FANCD2 clade makes the tree significantly worse at the $P < 0.02$ level (Fig. 4C).

In addition, Clustal Omega MSA analyses of human, mouse, and *Ciona* FANCD2 revealed a strong conservation of the CUE ubiquitin-binding domain,[63] the PCNA-interaction motif,[64] and the site of FANCD2 monoubiquitination K561 (Fig. 4D and E).[5] Furthermore, structural homology

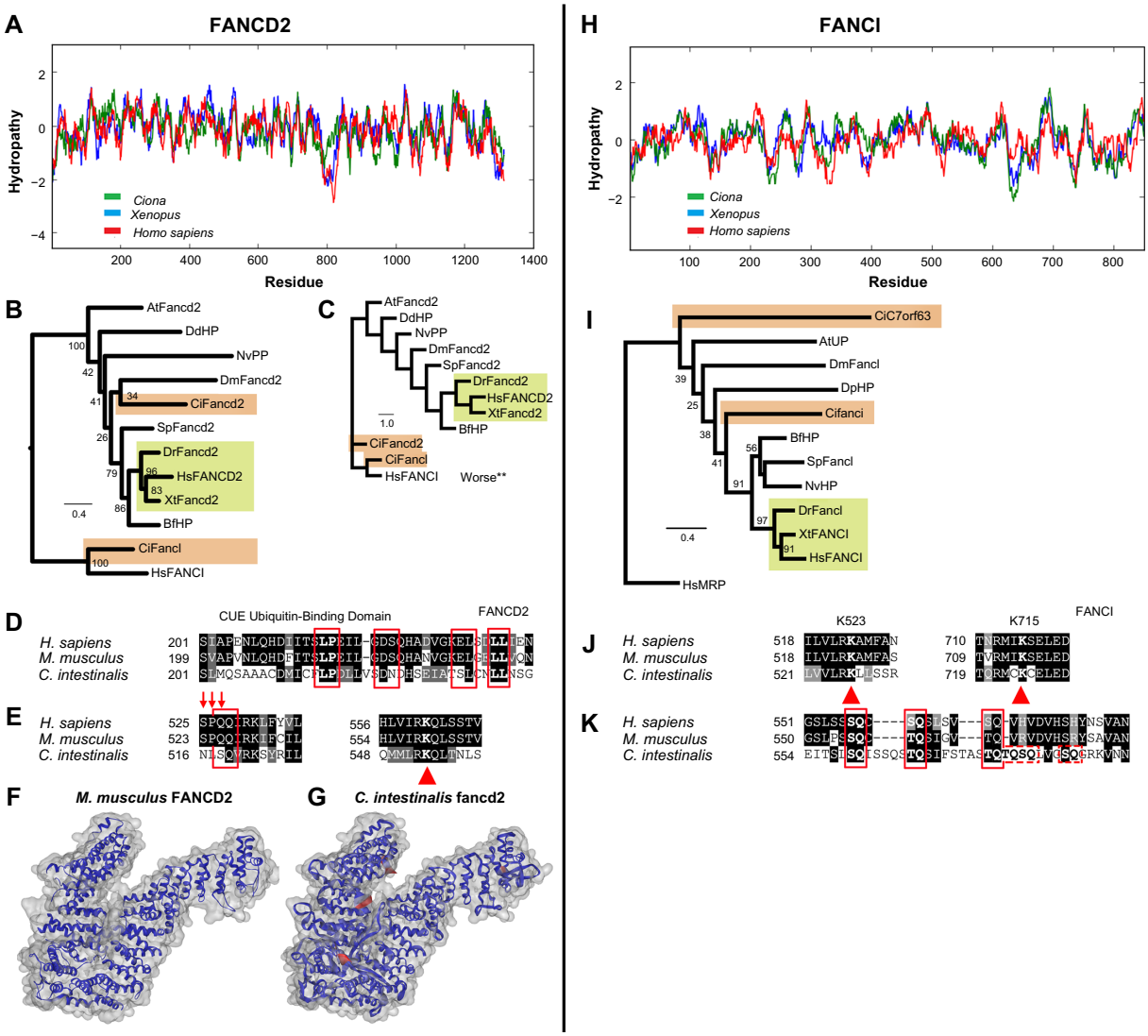

Figure 4. Analysis of FANCD2 (**A**–**G**) and FANCI (**H**–**K**) putative homologs in *C. intestinalis*. (**A**) Hydropathy plot of best aligning regions in human, *Xenopus*, and *Ciona* putative homologs for FANCD2. (**B**) Best ML tree for alignment of FANCE and putative homologs in *C. intestinalis* and other eukaryotes. Cifancd2 groups closely with FANCD2, but with low bootstrap support. (**C**) If Cifancd2 is moved out of the FANCD2 clade, the tree is statistically worse at the $P < 0.02$ level, supporting the case for Cifancd2 as a true homolog of FANCD2. (**D**) Alignment of human, mouse, and *C. intestinalis* FANCD2 protein sequences showing conservation of L215, P216, L234, and L235, critical residues of the CUE domain (red boxes).[63] (**E**) Alignment showing partial conservation of critical residues around human aa 525 (arrows, and box), as well as K561, the site of monoubiquitination[5] (red arrowhead) in *C. intestinalis*. (**F, G**) Modeling of mouse and *C. intestinalis* FANCD2 homolog protein structures, respectively. (**H**) Hydropathy plot of best aligning regions in human, *Xenopus*, and *Ciona* putative homologs for FANCI. (I) Best ML tree for alignment of putative FANCI homologs, showing weak bootstrap support for Cifancl being more closely related to FANCI than the next most similar *C. intestinalis* protein. (**J**) Alignment of human, mouse, and *C. intestinalis* FANCI protein sequences showing the conservation of K523 and K715, the site of FANCI monoubiquitination,[6,7] and, the site of FANCI SUMOylation, respectively (arrowheads). (**K**) SQ/TQ phosphosite clusters (red boxes) shown to be critical for FANCI function.[68] Dashed boxes denote possible additional functional SQ/TQ phosphosite clusters in *C. intestinalis* sequence not found in humans and mouse.

modeling of *Ciona* fancd2, based on the 3.4 Å *Mus musculus* Fancd2-Fanci heterodimer structure (PDB ID 3S4W),[65] reveals a largely favorable structural similarity (Fig. 4F and G). Taken together, we consider that these data provide good support for the presence of a *C. intestinalis fancd2* gene.

FANCI. Like FANCD2, FANCI is monoubiquitinated by FANCL and FANCT/UBE2T during ICL repair. Both RBB and RSD returned a *C. intestinalis* candidate fanci as the closest match to the human FANCI protein, with an E-value of 0. The hydrophobicity plots return an R^2 value of 0.33, but several areas, notably a 150 amino acid stretch toward the carboxy-terminal end of the protein, have much higher correlations (Fig. 4H). Both proteins show multiple globular domains with moderate alignment and no recognizable secondary motifs. Clustal Omega MSA analyses of human, mouse, and *Ciona* FANCI indicate the conservation of K523 and K715, the sites of FANCI monoubiquitination and SUMOylation, respectively (Fig. 4J).[6,7,66] In addition, *Ciona* fanci contains multiple conserved SQ/TQ ATM/ATR kinase phosphorylation motifs proximal to the putative

monoubiquitination site (Fig. 4K). In vertebrates, these sites have been demonstrated to be critical for FANCI regulation and function.[67,68] On the other hand, the ML phylogenetic analysis is inconclusive with respect to the orthology of the *C. intestinalis* candidate and FANCI. The best ML tree places the *Ciona* candidate as a sister taxon to a clade of deuterostome plus cnidarian FANCI proteins (Fig. 4I). However, forcing the *Ciona* candidate into the vertebrate FANCI clade results in a statistically worse tree, while forcing the *Ciona* candidate to group with the next most similar *Ciona* protein is not significantly different from the best ML tree (data not shown). In spite of the lack of support from the phylogenetic analysis, the sequence motif and structural data strongly suggest that *Ciona* fanci is a true FANCI ortholog.

Seven group III orthologs were found. *FANCJ/BRIP1.* In humans, FANCJ is a 5′–3′ DNA helicase that interacts directly with BRCA1.[69,70] RBB returns the ERCC2 nucleotide excision repair protein, but RSD returns human FANCJ. There is good alignment between the globular domains in human FANCJ and the *Ciona* candidate, and the hydrophobicity plot shows high correlation (Fig. 5A). The human protein is of a similar size to the *Ciona* protein, and they both possess a DEAH-box helicase domain (Fig. 5B). The ML tree groups *C. intestinalis* fancj in the vertebrate FANCJ clade at 100% bootstrap support, and moving the *C. intestinalis* candidate out of that clade makes the tree significantly worse (Fig. 5C and D). Given these data, the *C. intestinalis* fancj candidate is a clear ortholog of human FANCJ.

FAN1. Fanconi-associated nuclease 1 is a DNA repair protein known to interact with monoubiquitinated FANCD2[14] and FANCI.[71] The RBB returns a protein annotated as *Ciona* fan1, with an E-value of 4×10^{-145}. The fan1 C-terminal region shows 41% identity and 63% positive matches. The human and *Ciona* proteins align extremely well in the hydropathy plot (Fig. 5E) and both contain a 110 aa VRR nuclease domain (Fig. 5F). The ML tree clusters the *C. intestinalis* candidate with the vertebrate FAN1 proteins (Fig. 5G) and is significantly worse when the *C. intestinalis* protein is taken out of that clade (Fig. 5H; $P < 0.01$). Taken together, the evidence is strongly in favor of *Ciona* fan1 being a homolog of FAN1.

FANCQ/ERCC4. The *FANCQ* gene product, also known as ERCC4 or XPF, forms a heterodimer with ERCC1 and functions as a DNA repair structure-specific endonuclease.[72] Both search methods return a *Ciona* xpf as the most closely matching protein, with 50% identity, and 64% positive matches. The hydrophobicity plots show a high correlation, excepting one area corresponding to aa 390–430 in *Ciona* and aa 520–560 in humans (Fig. 6A). Both proteins possess an ERCC4 endonuclease domain of the same size approximately the same distance from the carboxy-terminal end of the protein (Fig. 6C). The ML analysis clusters the *C. intestinalis* xpf in the FANCQ clade (Fig. 6B), although moving the *C. intestinalis* protein out of that clade does not make the tree likelihood significantly worse (data not shown). Taken together, we conclude that *C. intestinalis* does have a FANCQ ortholog.

ERCC1. ERCC1 interacts directly with FANCQ/ERCC4. The *Ciona* candidate returned by RBB (XP_009861832) has an extremely similar hydropathy plot as the human and frog ERCC1 proteins, except at the N-terminal-most 50 residues (Fig. 6D), although the *Ciona* candidate appears to lack an intact HhH1 domain present in the human protein (Fig. 6E). The ML analysis groups the *Ciona* candidate within the vertebrate ERCC1 clade (Fig. 6F). Moving the *Ciona* protein outside that clade or grouping it with the next most similar human gene (FAAP24) makes the trees statistically worse at the $P < 0.02$ level (Fig. 6G and H). These data strongly support the orthology of the *Ciona* candidate.

FANCO/RAD51C. RAD51C is also required for the maintenance of chromosome stability by functioning in HR repair.[73] *Ciona* has five potential RAD51 family homologs if the proteins listed as lim15 and xrcc2 are included. RBB finds *Ciona* rad51 (XP_002126934) as the closest match to human FANCO. However, if the *Ciona* protein identified as rad51C in GenBank (XP_002130341) is used in the ML analysis with FANCO, *Ciona* rad51C robustly groups with FANCO to the exclusion of *Ciona* rad51 (Fig. 7C). Forcing *Ciona* rad51C out of the FANCO clade results in a statistically worse tree (Fig. 7D, $P < 0.01$). Structurally, the *Ciona* rad51C is more similar to FANCO than the higher BLAST match *Ciona* rad51 (Fig. 7A and B). Based on these analyses, we conclude that *Ciona* does have a FANCO homolog.

FANCR/RAD51. In humans, RAD51, recently gaining the name FANCR, is the major DNA strand exchange protein and is critical for the HR DNA repair process.[74,75] De novo heterozygous *RAD*51 mutations have recently been reported in two unrelated individuals with an FA-like syndrome.[76] RAD51 is known to interact with both FANCS/BRCA1 and FANCD1/BRCA2 in the cellular DNA damage response.[77] Both search methods return a *Ciona* rad51 as the most likely counterpart to the human protein. RAD51 appears to be the most highly conserved protein in the entire FA pathway. The protein possesses 82% identity between human and *Ciona* as well as a 92% level of positive matches, far outstripping any other gene product tested. The *Ciona* product is 338 aa in length, while the human product is 339 aa (Fig. 8B). Both RAD51 and the *Ciona* rad51 candidate possess a 20 amino acid helix-hairpin-helix domain starting at about amino acid 60, as well as a 187 aa AAA-ATPase domain ending 33 aa before the C-terminus. The hydrophobicity plots show extreme similarity, returning a Pearson coefficient of 0.92 (Fig. 8A). The ML analysis shows the *Ciona* rad51 candidate grouping with other deuterostome RAD51 proteins (Fig. 8C), while excluding *Ciona* rad51 from that clade results in a statistically worse tree (Fig. 8D, $P < 0.01$). It is highly likely that *Ciona* rad51 is a true ortholog of human FANCR/RAD51.

FANCS/BRCA1. The *C. intestinalis* candidate for FANCS by RBB has two BRCT (BRCA1 C-terminal domain) domains at the C-terminus, similar to BRCA1 (Fig. 8F). BRCT domains typically mediate interactions with phosphopeptides. The

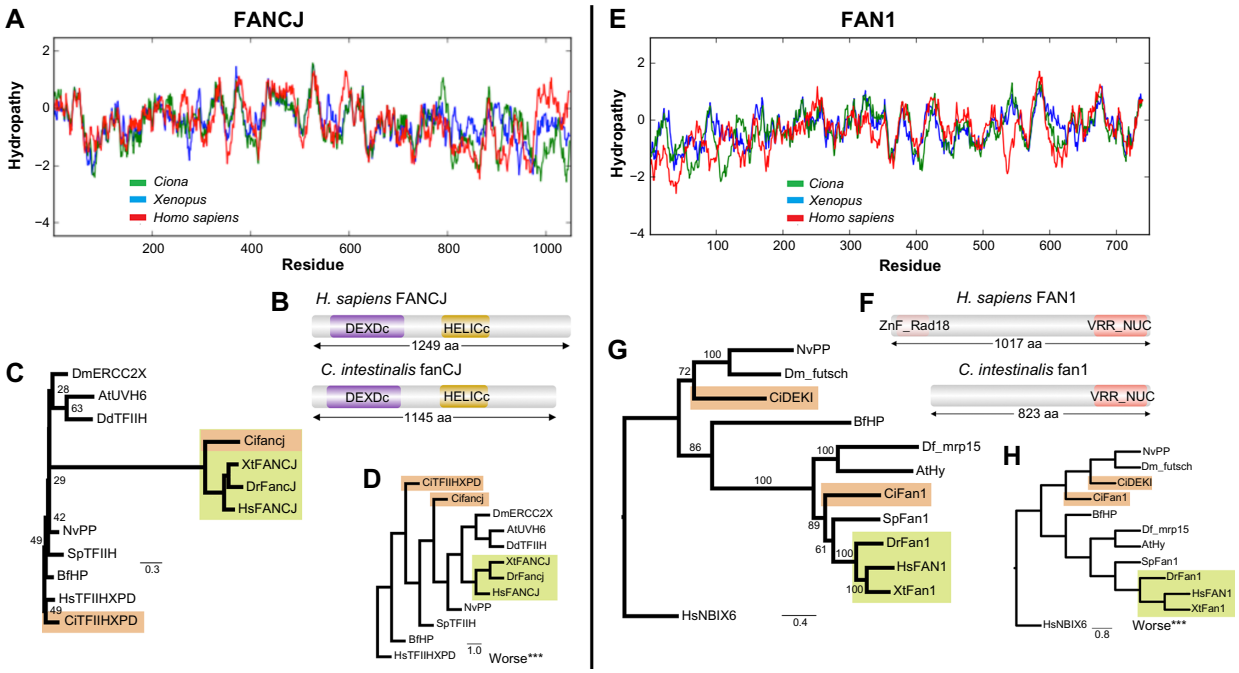

Figure 5. Analysis of FANCJ (**A**–**D**) and FAN1 (**E**–**H**) putative homologs in *C. intestinalis*. (**A**) Hydropathy plot of best aligning regions in human, *Xenopus*, and *Ciona* putative homologs for FANCJ. (**B**) Diagrammatic comparison of human and *C. intestinalis* FANCJ inferred protein motifs. (**C**) Best ML tree for alignment of FANCJ and putative homologs in *C. intestinalis* and other eukaryotes. Cifancj has 100% bootstrap support for membership in the clade with vertebrate FANCJ proteins, to the exclusion of the next most similar *C. intestinalis* protein. (**D**) If Cifancj is moved out of the FANCJ clade, the tree is statistically worse at the $P = 0.01$ level, supporting the case for Cifancj as a true homolog of FANCJ. (**E**) Hydropathy plot of best aligning regions in human, *Xenopus*, and *Ciona* putative homologs for FAN1. (**F**) Diagrammatic comparison of human and *C. intestinalis* FAN1 inferred protein motifs. (**G**) Best ML tree for alignment of putative FAN1 homologs, with Cifan1 falling in a clade with the vertebrate FAN1 homologs with 100% bootstrap support. Forcing Cifan1 out of that clade results in a statistically worse tree ($P < 0.01$).

hydropathy plot of the C-terminal 500 residues of the *C. intestinalis*, human, and frog proteins show a good degree of similarity (Fig. 8E). However, the rest of the sequence of the 1172 aa predicted *C. intestinalis* protein (from the ANISEED database as KH2012:KH.C9.487) toward the N-terminus has little resemblance to the human FANCS/BRCA1. Most likely because of this lack of alignment for a large part of the sequence, the ML analysis does not group the *C. intestinalis* protein with FANCS at a robust level (Fig. 8G). In fact, moving the *C. intestinalis* sequence either within the vertebrate BRCA1 clade or to the more distant branch of the tree makes for a statistically worse topology (data not shown). Because part of the protein is similar to its putative homolog while over half is not, we cannot say with complete confidence that "Ci-brca1" is a true homolog. However, it may still be the case that this protein in combination with one or more others is fulfilling the function served in humans by BRCA1.

FA and FA-associated proteins not found in *Ciona*. Our analyses found 11 FA or FA-associated proteins present in vertebrates but not in *Ciona*. These results were based on the four major criteria outlined for each of the predicted FA homologs, as outlined above, namely, BLAST search, structural motif similarity, hydropathy, and phylogenetic (ML) analysis. The FA proteins that we did not find homologs for in *Ciona* were as follows: FANCA, FANCB, FANCC, FANCF, FANCG, FANCD1/BRCA2, FANCN/PALB2, and FANCP/SLX4. We also failed to find the FA-associated proteins FAAP20, FAAP24, and FAAP100.

For 10 of the 11 cases, RBB and RSD failed to match a *Ciona* protein sequence with an FA-related protein (Table 1). The exception is FANCD1/BRCA2, for which a match comes up in RBB as an uncharacterized protein LOC100185089 (Table 1). However, the ML analysis results in another *C. intestinalis* protein showing a closer relationship to FANCD1/BRCA2. Rearranging the trees so that the best BLAST match is moved out of the FANCD1 clade altogether, or switching the first and second most similar *C. intestinalis* proteins in the tree, does not result in statistically worse trees, indicating that the evidence for homology of the *C. intestinalis* proteins is weak (data not shown). In addition, the hydropathy analysis shows a low correlation ($R^2 = 0.117$, Table 2). A Prosite scan indicates that LOC100185089 has two BRC repeats, which may explain why it comes up in the BLAST search. However, FANCD1 is a much larger protein (3418 aa vs. 724 aa) and has eight BRC repeats. These BRC repeats represent the major sites of interaction between RAD51 and BRCA2.[78] In addition, BRCA2/FANCD1 has an α-helical region, an oligonucleotide/oligosaccharide-binding domain, a TOWER domain, and a second oligonucleotide/oligosaccharide-binding domain.

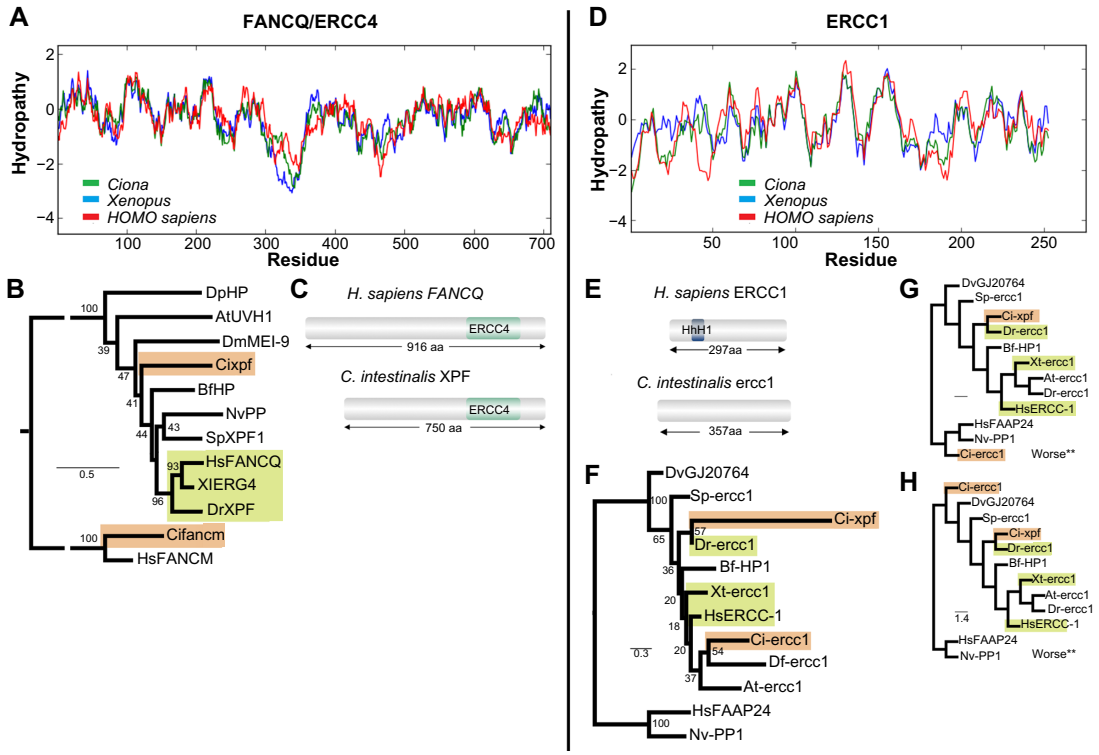

Figure 6. Analysis of FANCQ/ERCC4 (**A**–**C**) and ERCC1 (**D**–**H**) putative homologs in *C. intestinalis*. (**A**) Hydropathy plot of best aligning regions in human, *Xenopus*, and *Ciona* putative homologs for FANCQ/ERCC4. The best *C. intestinalis* BLASTmatch to FANCQ is termed XPFin GenBank (Table 1). (**B**) Best MLtree for alignment of FANCQ and putative homologs in *C. intestinalis* and other eukaryotes. CiXPFhas 100% bootstrap support for membership in the clade with vertebrate FANCQ proteins, to the exclusion of the next most similar *C. intestinalis* protein. (**C**) Diagrammatic comparison of human FANCQ and *C. intestinalis* XPFinferred protein motifs. (**D**) Hydropathy plot of best aligning regions in human, *Xenopus*, and *Ciona* putative homologs for ERCC1. (**E**) Diagrammatic comparison of human and *C. intestinalis* ERCC1 inferred protein motifs. (**F**) Best MLtree for alignment of putative ERCC1 homologs, with Ci-ercc1 shown closely related to ERCC1. However, the next closest *C. intestinalis* match, Ci-xpf, is in the same clade. However, forcing Ci-ercc1 out of the clade (**G**), or further from ERCC1 (**H**), both result in statistically worse trees ($P < 0.02$).

C. intestinalis LOC100185089 possesses two BRC repeats only. None of these other domains are present.

There is a possibility that the predicted *Ciona* protein in the NCBI database is not the full-length sequence. However, we searched a 22 kb region in the *Ciona* genome, which includes LOC100185089 and flanking regions. No significant similarity to the human sequence outside the region that aligns with LOC100185089 was found, even when the protein sequence not included in LOC100185089 was blasted against the translated *Ciona* genomic sequence. Thus, we infer that *Ciona* does not have a complete ortholog of human BRCA2.

For the other 10 of the 11 cases of unlikely homology, the hydropathy R^2 statistics are lower numbers than those for the putative homologs, ranging from 0.034 to 0.177 vs. 0.291 to 0.566, respectively (Table 2). Similarly, we did not find good evidence for homology to any *C. intestinalis* proteins by any of the other three analytical methods used (Table 1, and data not shown). Therefore, we conclude that these 11 FA and FA-associated proteins are missing from *C. intestinalis*.

Discussion

In this study, we have established that the model marine invertebrate, *C. intestinalis*, appears to contain all of the necessary functional components to reconstitute a simplified FA pathway (Fig. 9). Of the FA core complex group I proteins, we identified orthologs of FANCL, FANCT/UBE2T, and FANCM, and possibly FANCE. FANCL and FANCT/UBE2T are the E3 ubiquitin ligase and E2 ubiquitin conjugase enzymes, respectively, that monoubiquitinate FANCD2 and FANCI.[5–7,10] While FANCD2 and FANCI monoubiquitination are largely defective in FA patient cells with mutations in any of the core complex genes (*FANCA*, *B*, *C*, *E*, *F*, *G*, *L*, and *T*), several studies have established that FANCL and FANCT/UBE2T, in the presence of an E1 ubiquitin-activating enzyme and DNA, can readily promote FANCD2 and FANCI monoubiquitination *in vitro*.[8,79–81] The roles of the other FA core complex proteins in promoting FANCD2 and FANCI monoubiquitination *in vivo* remain unknown. The functions provided by these other core complex proteins may be unnecessary in *C. intestinalis*, or may be provided by other proteins. Interestingly, previous studies have established that the FANCE protein directly interacts with FANCD2, thereby bridging the core ubiquitin ligase machinery and the substrate. *C. intestinalis* fance may fulfill an analogous function. Similar to human FANCM, *C. intestinalis* fancm contains an N-terminal DEAH domain-containing Walker A and B motifs typical of an SF2 family

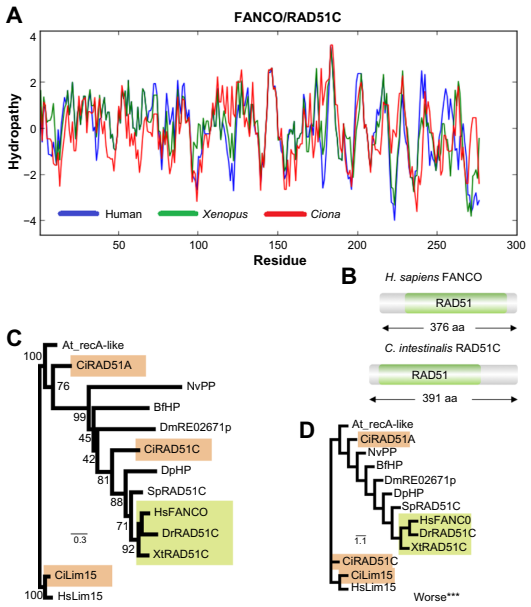

Figure 7. Analysis of FANCO/RAD51C putative orthologs in *C. intestinalis*. (**A**) Hydropathy plot of human, *Xenopus*, and *Ciona* putative FANCO proteins. (**B**) Diagrammatic comparison of *Homo sapiens* FANCO and *C. intestinalis* RAD51C protein motifs. (**C**) ML analysis has CiRAD51C grouping with the vertebrate FANCO protein sequences at moderate (86%) bootstrap support. (**D**) If CiRAD51C is excluded from the clade with the vertebrate FANCO proteins, the tree is significantly worse (***$P < 0.01$).

translocase. These proteins are capable of movement along DNA in the absence of helicase activity. FANCM translocase activity is necessary for replication fork stability and ATR-CHK1 checkpoint signaling.[82,83] The C-terminus of human FANCM contains a degenerate ERCC4 endonuclease domain, which is also the site of binding of its heterodimeric partner FAAP24; yet, this region appears absent in *C. intestinalis* fancm (Fig. 3B). Since *C. intestinalis* appears to lack a FAAP24 homolog, it is not surprising that Cifancm lacks the binding site. It has been speculated that the FANCM-FAAP24 heterodimer plays an important DNA-targeting function, and why the formation of a heterodimer might be unnecessary in *C. intestinalis* is unclear.[57,84] However, the categorization of *FANCM* as a true FA gene remains controversial.

The evidence for structural and functional conservation of the FANCD2 and FANCI proteins appears quite strong, with several protein domains and important sites of posttranslational modification being highly conserved (Fig. 4D, E, J, and K). This is consistent with the previous finding indicating considerable depth in their conservation in all eukaryotes.[26] The monoubiquitination of these proteins is a critical step in the activation of the FA pathway and in ICL repair.[5–7] In the case of FANCD2, monoubiquitination of K561 has been implicated in the recruitment of the FAN1 and FANCP/SLX4 proteins, which participate in, or facilitate, several key nucle-

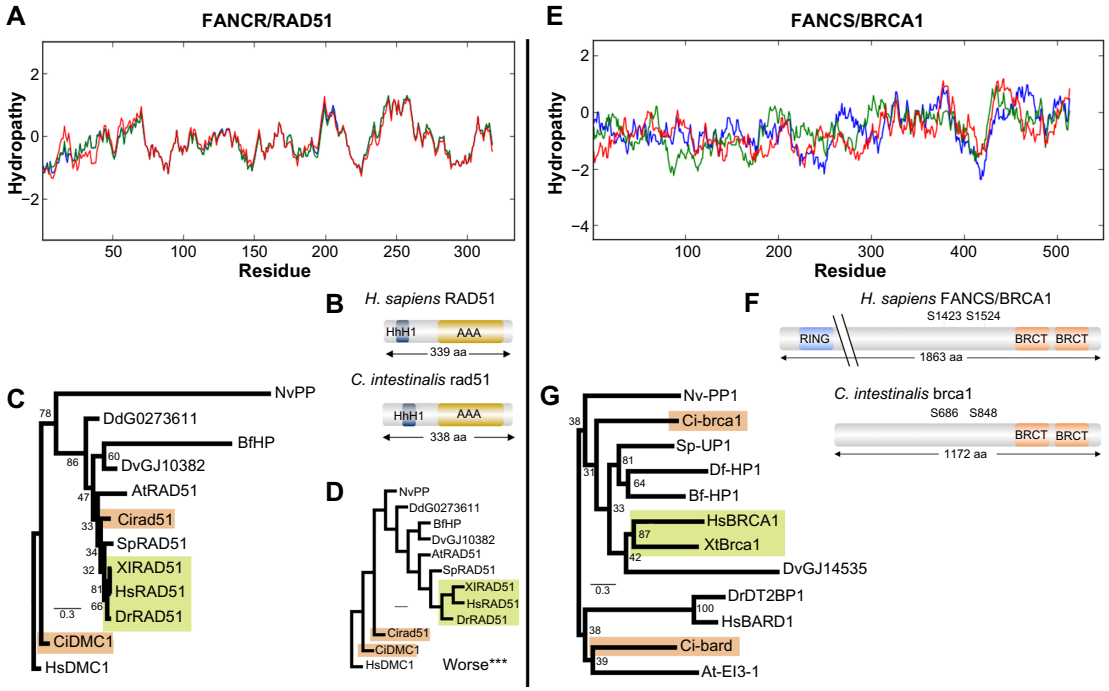

Figure 8. Analysis of FANCR/RAD51 (**A**–**D**) and FANCS/BRCA1 (**E**–**G**) putative homologs in *C. intestinalis*. (**A**) Hydropathy plot showing extremely similar hydropathy profiles in human, *Xenopus*, and *Ciona* putative homologs for RAD51. (**B**) Diagrammatic comparison of human and *C. intestinalis* RAD51 inferred protein motifs. (**C**) Best ML tree for alignment of putative RAD51 homologs, showing Ci-rad51 closely related to RAD51. (**D**) Forcing Ci-rad51 out of the clade results in a statistically worse tree ($P < 0.01$). (**E**) Hydropathy plot of best aligning regions in human, *Xenopus*, and *Ciona* putative homologs for FANCS/BRCA1, showing regions of both similar and divergent hydropathy. (**F**) Diagrammatic comparison of human FANCS/BRCA1 and putative *C. intestinalis* brca1 inferred protein motifs. The human protein has a RING domain not present in the *C. intestinalis* candidate. (**G**) Best ML tree for alignment of FANCS/BRCA1 and putative homologs in *C. intestinalis* and other eukaryotes. There is weak support for a clade with Ci-brca1 and FANCS/BRCA1.

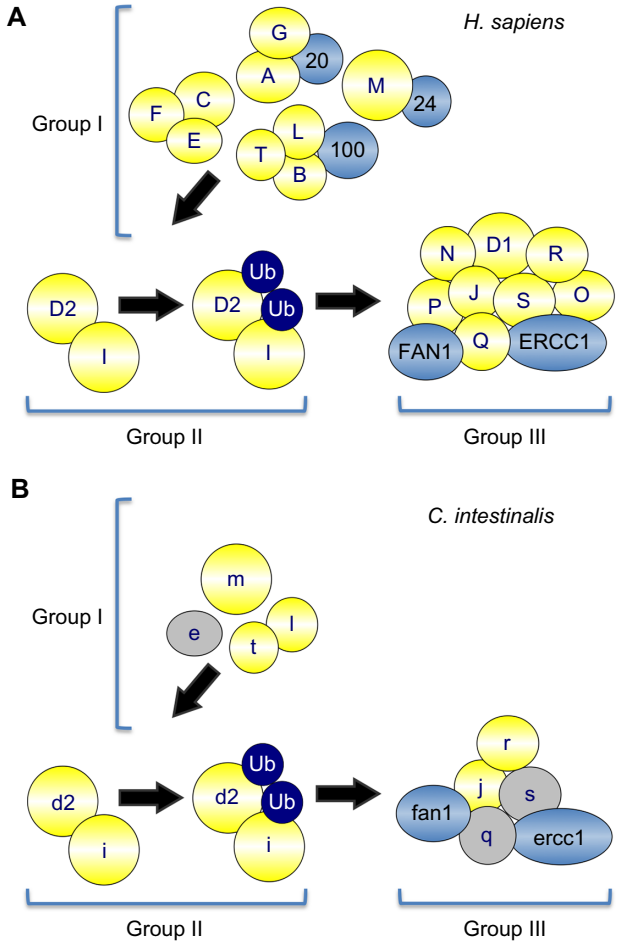

Figure 9. (**A**) A model of the FA pathway in humans. Following exposure to DNA damaging agents or during S-phase of the cell cycle, the FA core complex (group I) proteins catalyze the monoubiquitination of the FANCD2 and FANCI (group II) proteins. Following their monoubiquitination, FANCD2 and FANCI function together with the downstream FA (group III) proteins to repair damaged DNA. Modified from Cybulski and Howlett, 2011.[114] (**B**) A model of a hypothetical simplified FA pathway in *C. intestinalis* based on the reduced complement of FA gene homologs found by this study. *C. intestinalis* possesses the critical E3 ubiquitin ligase (Fancl) and E2 ubiquitin-conjugating enzyme (Fanct) to monoubiquitinate Fancd2 and Fanci, as well as a minimal set of FA group III effector proteins. Proteins shown in gray have lower support for existence in *C. intestinalis*.

olytic processing steps during ICL repair.[13–15,17] Conservation of the FANCD2 K561 and FANCI K523 monoubiquitination sites, as well as several other important sites of posttranslational modification, strongly suggests that this central step is intact in *C. intestinalis*.

Of the group III proteins, the evidence points to the existence of *C. intestinalis* orthologs of FANCJ/BRIP1, FANCQ/ERCC4, FANCR/RAD51, FANCO/RAD51C, and FANCS/BRCA1. The heterodimeric binding partner of FANCQ/ERCC4, ERCC1, is also present, as is the FANCD2-associated nuclease FAN1. Conservation of FANCR/RAD51 and FANCS/BRCA1 is not surprising, given their key roles in multiple cellular processes, including meiotic and mitotic recombination. Targeted disruption of either gene results in early embryonic lethality in mice.[85,86] However, the absence of FANCD1/BRCA2 is particularly surprising, given its strong conservation among eukaryotes.[26] FANCD1/BRCA2 plays a critical role in regulating FANCR/RAD51 nucleoprotein filament formation and DNA strand exchange.[78,87–89] It is also intriguing that *C. intestinalis* apparently lacks FANCN/PALB2. FANCN/PALB2 interacts directly with FANCD1/BRCA2 and promotes its chromatin localization.[23] Studies of the *Ustilago maydis* homolog of BRCA2 indicate that BRCA2 promotes RAD51 nucleation at junctions of single-stranded and double-stranded DNA.[90,91] However, lower eukaryotes such as *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe* lack homologs of both FANCS/BRCA1 and FANCD1/BRCA2, indicating that the functions provided by these proteins are unnecessary in certain organisms or may be provided by other proteins.

There is considerable precedent suggestive of the efficacy of studying the FA pathway in *C. intestinalis*. Study of several human diseases have benefited from the use of invertebrate model organisms. In particular, the genetically tractable invertebrates, such as *Drosophila* and *Caenorhabditis elegans*, have been used extensively.[92,93] Notably, it has recently been shown that even very simple animals, such as sponges and sea anemones, have homologs of many human disease genes.[94,95] *C. intestinalis* has only recently emerged as a model system. However, it has already been used to study certain human disease pathways, such as Huntington's Disease[96,97] and Alzheimer's disease (AD).[98] In the case of Alzheimer's, transgenic *C. intestinalis* were produced expressing the human APP gene mutant associated with familial AD. The transgenic protein resulted in the formation of amyloid-β plaques in less than 24 hours in the rapidly developing *C. intestinalis* larval brain. This result contrasts with a 2–8-month time period for plaques to form in mouse AD models. For FA, study of the pathway in invertebrate model organisms has proven valuable in several cases.[99] For example, the function of FANCJ in maintaining poly(G)/poly(C) tract stability during DNA replication was first shown in the nematode worm *C. elegans*.[100] It was later demonstrated that human FANCJ has the same helicase function.[101]

It is important to note that of all the three major constellations of FA patient phenotypes, namely, developmental defects, bone marrow failure, and increased cancer risk, the molecular bases of the developmental defects are the most poorly understood. A *C. intestinalis* model for FA could provide unique insights into these defects. Temporospatial aspects of FA gene expression and developmental consequences of disruption of FA genes using CRISPR/Cas9 or TALEN systems[102–106] could be highly informative for FA patient developmental defects. Furthermore, another unique benefit to exploring a *C. intestinalis* model for FA is the prospect of discovering the physiological function(s) of this pathway. While it is well established that FA patient cells

are hypersensitive to ICL-inducing agents, the relevance of ICLs in the physiological setting is unclear. Recent studies have established an important role for the FA proteins in mitigating endogenously arising aldehyde-mediated DNA damage.[107–109] Exploring the pathway in other model systems may lead to a broader understanding of the true function(s) of these key proteins. *C. intestinalis*, as a tunicate, is in the most closely related invertebrate group to the vertebrates.[27,110] As such, in spite of being anatomically simpler than a vertebrate, they are genetically more similar than other eukaryotes. However, it is possible that *C. intestinalis* may deploy its FA homologs differently than they function in humans. If this is the case, it may still be relevant to understanding human disease, as it will point to alternative ways of dealing with DNA lesions and may provide information on some of the other defects seen in FA patients.

In summary, our study provides compelling evidence for the existence of a simplified and potentially functional FA pathway in the model chordate *C. intestinalis*. *C. intestinalis* is an excellent model for the study of developmental processes because it is anatomically simple, its gametogenesis and development are well studied, it has a small and well-annotated genome and abundant gene expression data, and good transgenic technology exists.[29,30,111–113] Future studies will seek to determine the patterns and timing of FA gene expression in *C. intestinalis* and the developmental impacts of disruption of the pathway.

Author Contributions

Analyzed the data: ECS, PAA, DAV, NGH, SQI. Wrote the first draft of the manuscript: ECS. Contributed to the writing of the manuscript: NGH, SQI. Agreed with manuscript results and conclusions: ECS, PAA, DAV, NGH, SQI. Jointly developed the structure and arguments for the paper: ECS, NGH, SQI. Made critical revisions and approved the final version: ECS, NGH, SQI. All the authors reviewed and approved the final manuscript.

Supplementary Material

Supplemental Table S1. Accession Numbers for sequences used in analyses.

REFERENCES

1. Kim H, D'Andrea AD. Regulation of DNA cross-link repair by the Fanconi anemia/BRCA pathway. *Genes Dev*. 2012;26(13):1393–408.
2. Kottemann MC, Smogorzewska A. Fanconi anemia and the repair of Watson and Crick DNA crosslinks. *Nature*. 2013;493(7432):356–63.
3. Wang W. Emergence of a DNA-damage response network consisting of Fanconi anaemia and BRCA proteins. *Nat Rev Genet*. 2007;8(10):735–48.
4. Rego MA, Kolling FW, Howlett NG. The Fanconi anemia protein interaction network: casting a wide net. *Mutat Res*. 2009;668(1–2):27–41.
5. Garcia-Higuera I, Taniguchi T, Ganesan S, et al. Interaction of the Fanconi anemia proteins and BRCA1 in a common pathway. *Mol Cell*. 2001;7(2):249–62.
6. Sims AE, Spiteri E, Sims RJ III, et al. FANCI is a second monoubiquitinated member of the Fanconi anemia pathway. *Nat Struct Mol Biol*. 2007;14(6):564–7.
7. Smogorzewska A, Matsuoka S, Vinciguerra P, et al. Identification of the FANCI protein, a monoubiquitinated FANCD2 paralog required for DNA repair. *Cell*. 2007;129(2):289–301.
8. Alpi AF, Pace PE, Babu MM, Patel KJ. Mechanistic insight into site-restricted monoubiquitination of FANCD2 by Ube2t, FANCL, and FANCI. *Mol Cell*. 2008;32(6):767–77.
9. Meetei AR, de Winter JP, Medhurst AL, et al. A novel ubiquitin ligase is deficient in Fanconi anemia. *Nat Genet*. 2003;35(2):165–70.
10. Machida YJ, Machida Y, Chen Y, et al. UBE2T is the E2 in the Fanconi anemia pathway and undergoes negative autoregulation. *Mol Cell*. 2006;23(4):589–96.
11. Gari K, Decaillet C, Stasiak AZ, Stasiak A, Constantinou A. The Fanconi anemia protein FANCM can promote branch migration of Holliday junctions and replication forks. *Mol Cell*. 2008;29(1):141–8.
12. Meetei AR, Medhurst AL, Ling C, et al. A human ortholog of archaeal DNA repair protein Hef is defective in Fanconi anemia complementation group M. *Nat Genet*. 2005;37(9):958–63.
13. Kratz K, Schoepf B, Kaden S, et al. Deficiency of FANCD2-associated nuclease KIAA1018/FAN1 sensitizes cells to interstrand crosslinking agents. *Cell*. 2010;142(1):77–88.
14. MacKay C, Declais AC, Lundin C, et al. Identification of KIAA1018/FAN1, a DNA repair nuclease recruited to DNA damage by monoubiquitinated FANCD2. *Cell*. 2010;142(1):65–76.
15. Smogorzewska A, Desetty R, Saito TT, et al. A genetic screen identifies FAN1, a Fanconi anemia-associated nuclease necessary for DNA interstrand crosslink repair. *Mol Cell*. 2010;39(1):36–47.
16. Unno J, Itaya A, Taoka M, et al. FANCD2 binds CtIP and regulates DNA-end resection during DNA interstrand crosslink repair. *Cell Rep*. 2014;7(4):1039–47.
17. Yamamoto KN, Kobayashi S, Tsuda M, et al. Involvement of SLX4 in interstrand cross-link repair is regulated by the Fanconi anemia pathway. *Proc Natl Acad Sci U S A*. 2011;108(16):6492–6.
18. Yeo JE, Lee EH, Hendrickson EA, Sobeck A. CtIP mediates replication fork recovery in a FANCD2-regulated manner. *Hum Mol Genet*. 2014;23(14):3695–705.
19. French CA, Masson JY, Griffin CS, O'Regan P, West SC, Thacker J. Role of mammalian RAD51 L2 (RAD51C) in recombination and genetic stability. *J Biol Chem*. 2002;277(22):19322–30.
20. Godthelp BC, Wiegant WW, van Duijn-Goedhart A, et al. Mammalian Rad51C contributes to DNA cross-link resistance, sister chromatid cohesion and genomic stability. *Nucleic Acids Res*. 2002;30(10):2172–82.
21. Reid S, Schindler D, Hanenberg H, et al. Biallelic mutations in PALB2 cause Fanconi anemia subtype FA-N and predispose to childhood cancer. *Nat Genet*. 2007;39(2):162–4.
22. Vaz F, Hanenberg H, Schuster B, et al. Mutation of the RAD51C gene in a Fanconi anemia-like disorder. *Nat Genet*. 2010;42(5):406–U63.
23. Xia B, Sheng Q, Nakanishi K, et al. Control of BRCA2 cellular and clinical functions by a nuclear partner, PALB2. *Mol Cell*. 2006;22(6):719–29.
24. Zhang B, Arun G, Mao YS, et al. The lncRNA Malat1 is dispensable for mouse development but its transcription plays a cis-regulatory role in the adult. *Cell Rep*. 2012;2(1):111–23.
25. Xia B, Dorsman JC, Ameziane N, et al. Fanconi anemia is associated with a defect in the BRCA2 partner PALB2. *Nat Genet*. 2007;39(2):159–61.
26. Zhang XY, Langenick J, Traynor D, Babu MM, Kay RR, Patel KJ. Xpf and not the Fanconi anaemia proteins or Rev3 accounts for the extreme resistance to cisplatin in *Dictyostelium discoideum*. *PLoS Genet*. 2009;5(9):e1000645.
27. Delsuc F, Brinkmann H, Chourrout D, Philippe H. Tunicates and not cephalochordates are the closest living relatives of vertebrates. *Nature*. 2006;439:965–8.
28. Flicek P, Amode MR, Barrell D, et al. Ensembl 2014. *Nucleic Acids Res*. 2014;42(D1):D749–55.
29. Dehal P, Satou Y, Campbell RK, et al. The draft genome of *Ciona intestinalis*: insights into chordate and vertebrate origins. *Science*. 2002;298(5601):2157–67.
30. Tassy O, Dauga D, Daian F, et al. The ANISEED database: digital representation, formalization, and elucidation of a chordate developmental program. *Genome Res*. 2010;20(10):1459–68.
31. Satou Y, Kawashima T, Shoguchi E, Nakayama A, Satoh N. An integrated database of the ascidian, *Ciona intestinalis*: towards functional genomics. *Zoolog Sci*. 2005;22:837–43.
32. Endo T, Ueno K, Yonezawa K, et al. CIPRO 2.5: *Ciona intestinalis* protein database, a unique integrated repository of large-scale omics data, bioinformatic analyses and curated annotation, with user rating and reviewing functionality. *Nucleic Acids Res*. 2011;39:D807–14.
33. Hughes AL, Friedman R. Loss of ancestral genes in the genomic evolution of *Ciona intestinalis Evol Dev*. 2005;7(3):196–200.
34. Davidson B. *Ciona intestinalis* as a model for cardiac development. *Semin Cell Dev Biol*. 2007;18(1):16–26.
35. Philips A, Blein M, Robert AS, et al. Ascidians as a vertebrate-like model organism for physiological studies of Rho GTPase signaling. *Biol Cell*. 2003;95(5):295–302.
36. Shi WY, Levine M, Davidson B. Unraveling genomic regulatory networks in the simple chordate, *Ciona intestinalis*. *Genome Res*. 2005;15(12):1668–74.

37. Satoh N, Satou Y, Davidson B, Levine M. *Ciona intestinalis*: an emerging model for whole-genome analyses. *Trends Genet*. 2003;19(7):376–81.
38. Titus TA, Selvig DR, Qin BF, et al. The Fanconi anemia gene network is conserved from zebrafish to human. *Gene*. 2006;371(2):211–23.
39. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol*. 1990;215:403–10.
40. Satoh N, Satou Y, Davidson B, et al. Detecting putative orthologs. *Nat Genet*. 2003;23(2):2947–8.
41. Kent WJ. BLAT – the BLAST-like alignment tool. *Genome Res*. 2002;12(4): 656–64.
42. Nordberg H, Cantor M, Dusheyko S, et al. The genome portal of the Department of Energy Joint Genome Institute: 2014 updates. *Nucleic Acids Res*. 2014;42(D1):D26–31.
43. Kriventseva EV, Rahman N, Espinosa O, Zdobnov EM. OrthoDB: the hierarchical catalog of eukaryotic orthologs. *Nucleic Acids Res*. 2008;36(SI):D271–5.
44. Larkin MA, Blackshields G, Brown NP, et al. Clustal W and clustal X version 2.0. *Bioinformatics*. 2007;23(21):2947–8.
45. Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. Jalview Version 2-a multiple sequence alignment editor and analysis workbench. *Bioinformatics*. 2009;25(9):1189–91.
46. Dalke A. Hydrophobicity plots with matplotlib. http://www.dalkescientific.com/writings/NBN/plotting.html. 2011.
47. Abourizk SM, Halpin DW, Wilson JR. Fitting beta-distributions based on sample data. *J Constr Eng Manag*. 1994;120(2):288–305.
48. Sidak Z. Rectangular confidence regions for means of multivariate normal distributions. *J Am Stat Assoc*. 1967;62(318):626–33.
49. Finn RD, Coggill P, Eberhardt RY, et al. The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res*. 2016;44(D1):D279–85.
50. Katoh K, Kuma K, Toh H, Miyata T. MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Res*. 2005;33(2):511–8.
51. Sanchez R, Serra F, Tarraga J, et al. Phylemon 2.0: a suite of web-tools for molecular evolution, phylogenetics, phylogenomics and hypotheses testing. *Nucleic Acids Res*. 2011;39(2):W470–4.
52. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*. 2014;30(9):1312–3.
53. Nookala RK, Hussain S, Pellegrini L. Insights into Fanconi anaemia from the structure of human FANCE. *Nucleic Acids Res*. 2007;35(5):1638–48.
54. Gurtan AM, Stuckert P, D'Andrea AD. The WD40 repeats of FANCL are required for Fanconi anemia core complex assembly. *J Biol Chem*. 2006;281(16): 10896–905.
55. Cole AR, Lewis LPC, Walden H. The structure of the catalytic subunit FANCL of the Fanconi anemia core complex. *Nat Struct Mol Biol*. 2010;17(3): 294–U54.
56. Hodson C, Cole AR, Lewis LPC, Miles JA, Purkiss A, Walden H. Structural analysis of human FANCL, the E3 ligase in the Fanconi anemia pathway. *J Biol Chem*. 2011;286(37):32628–37.
57. Ciccia A, Ling C, Coulthard R, et al. Identification of FAAP24, a Fanconi anemia core complex protein that interacts with FANCM. *Mol Cell*. 2007;25(3):331–43.
58. Gari K, Decaillet C, Delannoy M, Wu L, Constantinou A. Remodeling of DNA replication structures by the branch point translocase FANCM. *Proc Natl Acad Sci U S A*. 2008;105(42):16107–12.
59. Huang M, Kim JM, Shiotani B, Yang K, Zou L, D'Andrea AD. The FANCM/FAAP24 complex is required for the DNA interstrand crosslink-induced checkpoint response. *Mol Cell*. 2010;39(2):259–68.
60. Kim JM, Kee Y, Gurtan A, D'Andrea AD. Cell cycle-dependent chromatin loading of the Fanconi anemia core complex by FANCM/FAAP24. *Blood*. 2008;111(10):5215–22.
61. Sobeck A, Stone S, Landais I, de Graaf B, Hoatlin ME. The Fanconi anemia protein FANCM is controlled by FANCD2 and the ATR/ATM pathways. *J Biol Chem*. 2009;284(38):25560–68.
62. Hira A, Yoshida K, Sato K, et al. Mutations in the gene encoding the E2 conjugating enzyme UBE2T cause Fanconi anemia. *Am J Hum Genet*. 2015;96(6):1001–7.
63. Rego MA, Kolling FW, Vuono EA, Mauro M, Howlett NG. Regulation of the Fanconi anemia pathway by a CUE ubiquitin-binding domain in the FANCD2 protein. *Blood*. 2012;120(10):2109–17.
64. Howlett NG, Harney JA, Rego MA, Kolling FW, Glover TW. Functional interaction between the Fanconi anemia D2 protein and proliferating cell nuclear antigen (PCNA) via a conserved putative PCNA interaction motif. *J Biol Chem*. 2009;284(42):28935–42.
65. Joo W, Xu G, Persky NS, et al. Structure of the FANCI-FANCD2 complex: insights into the Fanconi anemia DNA repair pathway. *Science*. 2011;333(6040): 312–6.
66. Gibbs-Seymour I, Oka Y, Rajendra E, et al. Ubiquitin-SUMO circuitry controls activated Fanconi anemia ID complex dosage in response to DNA damage. *Mol Cell*. 2015;57(1):150–64.
67. Chen YH, Jones MJK, Yin Y, et al. ATR-mediated phosphorylation of FANCI regulates dormant origin firing in response to replication stress. *Mol Cell*. 2015;58(2):323–38.
68. Ishiai M, Kitao H, Smogorzewska A, et al. FANCI phosphorylation functions as a molecular switch to turn on the Fanconi anemia pathway. *Nat Struct Mol Biol*. 2008;15(11):1138–46.
69. Cantor SB, Bell DW, Ganesan S, et al. BACH1, a novel helicase-like protein, interacts directly with BRCA1 and contributes to its DNA repair function. *Cell*. 2001;105(1):149–60.
70. Cantor S, Drapkin R, Zhang F, et al. The BRCA1-associated protein BACH1 is a DNA helicase targeted by clinically relevant inactivating mutations. *Proc Natl Acad Sci U S A*. 2004;101(8):2357–62.
71. Liu T, Ghosal G, Yuan J, Chen J, Huang J. FAN1 acts with FANCI-FANCD2 to promote DNA interstrand cross-link repair. *Science*. 2010;329(5992):693–6.
72. Gregg SQ, Robinson AR, Niedernhofer LJ. Physiological consequences of defects in ERCC1-XPF DNA repair endonuclease. *DNA Repair (Amst)*. 2011;10(7, SI):781–91.
73. Takata M, Sasaki MS, Tachiiri S, et al. Chromosome instability and defective recombinational repair in knockout mutants of the five Rad51 paralogs. *Mol Cell Biol*. 2001;21(8):2858–66.
74. Boulton SJ. Cellular functions of the BRCA tumour-suppressor proteins. *Biochem Soc Trans*. 2006;34(5):633–45.
75. Mazón G, Mimitou E, Symington L. SnapShot: homologous recombination in DNA double-strand break repair. *Cell*. 2010;142:646.
76. Wang AT, Kim T, Wagner JE, et al. A dominant mutation in human RAD51 reveals its function in DNA interstrand crosslink repair independent of homologous recombination. *Mol Cell*. 2015;59(3):478–90.
77. Chen JJ, Silver D, Cantor S, Livingston DM, Scully R. BRCA1, BRCA2, and Rad51 operate in a common DNA damage response pathway. *Cancer Res*. 1999;59(7, S):1752S–6S.
78. Wong AKC, Pero R, Ormonde PA, Tavtigian SV, Bartel PL. RAD51 interacts with the evolutionarily conserved BRC motifs in the human breast cancer susceptibility gene brca2. *J Biol Chem*. 1997;272(51):31941–4.
79. Longerich S, Kwon Y, Tsai MS, Hlaing AS, Kupfer GM, Sung P. Regulation of FANCD2 and FANCI monoubiquitination by their interaction and by DNA. *Nucleic Acids Res*. 2014;42(9):5657–70.
80. Longerich S, Filippo JS, Liu D, Sung P. FANCI binds branched DNA and is monoubiquitinated by UBE2T-FANCL. *J Biol Chem*. 2009;284(35):23182–6.
81. Sato K, Toda K, Ishiai M, Takata M, Kurumizaka H. DNA robustly stimulates FANCD2 monoubiquitylation in the complex with FANCI. *Nucleic Acids Res*. 2012;40(10):4553–61.
82. Blackford AN, Schwab RA, Nieminuszczy J, Deans AJ, West SC, Niedzwiedz W. The DNA translocase activity of FANCM protects stalled replication forks. *Hum Mol Genet*. 2012;21(9):2005–16.
83. Collis SJ, Ciccia A, Deans AJ, et al. FANCM and FAAP24 function in ATR-mediated checkpoint signaling independently of the Fanconi anemia core complex. *Mol Cell*. 2008;32(3):313–24.
84. Walden H, Deans AJ. The Fanconi anemia DNA repair pathway: structural and functional insights into a complex disorder. 2014;43:257–78.
85. Gowen LC, Johnson BL, Latour AM, Sulik KK, Koller BH. Brca1 deficiency results in early embryonic lethality characterized by neuroepithelial abnormalities. *Nat Genet*. 1996;12(2):191–4.
86. Tsuzuki T, Fujii Y, Sakumi K, et al. Targeted disruption of the Rad51 gene leads to lethality in embryonic mice. *Proc Natl Acad Sci U S A*. 1996;93(13):6236–40.
87. Davies AA, Masson JY, Mcllwraith MJ, et al. Role of BRCA2 in control of the RAD51 recombination and DNA repair protein. *Mol Cell*. 2001;7(2): 273–82.
88. Esashi F, Galkin VE, Yu X, Egelman EH, West SC. Stabilization of RAD51 nucleoprotein filaments by the C-terminal region of BRCA2. *Nat Struct Mol Biol*. 2007;14(6):468–74.
89. Shivji MKK, Mukund SR, Rajendra E, et al. The BRC repeats of human BRCA2 differentially regulate RAD51 binding on single- versus double-stranded DNA to stimulate strand exchange. *Proc Natl Acad Sci U S A*. 2009;106(32):13254–9.
90. Yang HJ, Jeffrey PD, Miller J, et al. BRCA2 function in DNA binding and recombination from a BRCA2-DSS1-ssDNA structure. *Science*. 2002;297(5588):1837–48.
91. Yang HJ, Li QB, Fan J, Holloman WK, Pavletich NP. The BRCA2 homologue Brh2 nucleates RAD51 filament formation at a dsDNA-ssDNA junction. *Nature*. 2005;433(7026):653–7.
92. Ségalat L. Invertebrate animal models of diseases as screening tools in drug discovery. *ACS Chem Biol*. 2007;4:231–6.
93. Sosa MAG, De Gasperi R, Elder GA. Modeling human neurodegenerative diseases in transgenic systems. *Hum Genet*. 2012;131(4):535–63.
94. Sullivan JC, Finnerty JR. A surprising abundance of human disease genes in a simple "basal" animal, the starlet sea anemone (*Nematostella vectensis*). *Genome*. 2007;50(7):689–92.
95. Perina D, Korolija M, Hadzija MP, et al. Functional and structural characterization of FAU gene/protein from marine sponge *Suberites domuncula*. *Mar Drugs*. 2015;13(7):4179–96.
96. Idris MM, Thorndyke MC, Brown ER. Evidence for dynamic and multiple roles for huntingtin in *Ciona intestinalis*. *Invertebr Neurosci*. 2013;13(2):151–65.

97. Gissi C, Pesole G, Cattaneo E, Tartari M. Huntingtin gene evolution in Chordata and its peculiar features in the ascidian *Ciona* genus. *BMC Genomics*. 2006;7:288.
98. Virata MJ, Zeller RW. Ascidians: an invertebrate chordate model to study Alzheimer's disease pathogenesis. *Dis Model Mech*. 2010;3(5–6):377–85.
99. McVey M. Strategies for DNA interstrand crosslink repair: insights from worms, flies, frogs, and slime molds. *Environ Mol Mutagen*. 2010;51(6, SI):646–58.
100. Youds JL, Barber LJ, Ward JD, et al. DOG-1 is the *Caenorhabditis elegans* BRIP1/FANCJ homologue and functions in interstrand cross-link repair. *Mol Cell Biol*. 2008;28(5):1470–9.
101. Wu Y, Shin-ya K, Brosh RM Jr. FANCJ helicase defective in Fanconia anemia and breast cancer unwinds G-quadruplex DNA to defend genomic stability. *Mol Cell Biol*. 2008;28(12):4116–28.
102. Jung M, Dunbar CE, Winkler T. Modeling human bone marrow failure syndromes using pluripotent stem cells and genome engineering. *Mol Ther*. 2015;23(12):1832–42.
103. Osborn MJ, Gabriel R, Webber BR, et al. Fanconi anemia gene editing by the CRISPR/Cas9 system. *Hum Gene Ther*. 2015;26(2):114–26.
104. Stolfi A, Gandhi S, Salek F, Christiaen L. Tissue-specific genome editing in *Ciona* embryos by CRISPR/Cas9. *Development*. 2014;141(21):4115–20.
105. Yoshida K, Treen N, Hozumi A, Sakuma T, Yamamoto T, Sasakura Y. Germ cell mutations of the ascidian *Ciona intestinalis* with TALE nucleases. *Genesis*. 2014;52(5):431–9.
106. Treen N, Yoshida K, Sakuma T, et al. Tissue-specific and ubiquitous gene knockouts by TALEN electroporation provide new approaches to investigating gene function in *Ciona*. *Development*. 2014;141(2):481–7.
107. Garaycoechea JI, Crossan GP, Langevin F, et al. Genotoxic consequences of endogenous aldehydes on mouse haematopoietic stem cell function. *Nature*. 2011;489(12):53–U67.
108. Langevin F, Crossan GP, Rosado IV, Arends MJ, Patel KJ. Fancd2 counteracts the toxic effects of naturally produced aldehydes in mice. *Nature*. 2011;475(7354):53–8.
109. Rosado IV, Langevin F, Crossan GP, Takata M, Patel KJ. Formaldehyde catabolism is essential in cells deficient for the Fanconi anemia DNA-repair pathway. *Nat Struct Mol Biol*. 2011;18(12):1432–4.
110. Dunn CW, Hejnol A, Matus DQ, et al. Broad phylogenomic sampling improves resolution of the animal tree of life. *Nature*. 2008;452:745–50.
111. Satoh N, Levine M. Surfing with the tunicates into the post-genome era. *Genes Dev*. 2005;19(20):2407–11.
112. Corbo JC, Di Gregorio A, Levine M. The ascidian as a model organism in developmental and evolutionary biology. *Cell*. 2001;106(5):535–8.
113. Stolfi A, Christiaen L. Genetic and genomic toolbox of the chordate *Ciona intestinalis*. *Genetics*. 2012;192(1):55–66.
114. Cybulski KE, Howlett NG. FANCP/SLX4 A Swiss army knife of DNA interstrand crosslink repair. *Cell Cycle*. 2011;10(11):1757–63. doi:10.4161/cc.10.11.15818.