Comprehensive mutational analysis of the BRCA1-associated DNA helicase and tumor suppressor, FANCJ/BACH1/BRIP1

Jennifer A Calvo¹, Briana Fritchman², Desiree Hernandez², Nicole S Persky², Cory M Johannessen²,³, Federica Piccioni²,⁴, Brian A. Kelch⁵, Sharon B Cantor¹

¹ Molecular Cell and Cancer Biology, University of Massachusetts Medical School, Worcester, MA 01605, USA
² The Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA
³ Current Address: Oncology, Novartis Institutes of Biomedical Research, Cambridge, MA 02142 USA
⁴ Current Address: Merck Research Laboratories, 33 Avenue Louis Pasteur, Boston, MA 02115
⁵ Biochemistry and Molecular Pharmacology, University of Massachusetts Medical School, Worcester, MA 01605, USA

Running Title: Functional analysis of FANCJ mutations

Corresponding author: Sharon B Cantor, Sharon.Cantor@umassmed.edu; University of Massachusetts Medical School, 364 Plantation Street, LRB 415, Worcester MA 01605

Potential Conflicts of Interest: The authors have declared that no competing interests exist.
ABSTRACT

FANCJ (BRIP1/BACH1) is a hereditary breast and ovarian cancer (HBOC) gene encoding a DNA helicase. Similar to HBOC genes, BRCA1 and BRCA2, FANCJ is critical for processing DNA inter-strand crosslinks (ICL) induced by chemotherapeutics, such as cisplatin. Consequently, cells deficient in FANCJ or its catalytic activity are sensitive to ICL-inducing agents. Unfortunately, the majority of FANCJ clinical mutations remain uncharacterized, limiting therapeutic opportunities to effectively use cisplatin to treat tumors with mutated FANCJ. Here we sought to perform a comprehensive screen to identify FANCJ loss-of-function (LOF) mutations. We developed a FANCJ lentivirus mutation library representing ~450 patient derived-FANCJ nonsense and missense mutations to introduce FANCJ mutants into FANCJ knockout (K/O) HeLa cells. We performed a high-throughput screen to identify FANCJ LOF mutants that, as compared to wildtype FANCJ, fail to robustly restore resistance to ICL-inducing agents, cisplatin or mitomycin C (MMC). Based on the failure to confer resistance to either cisplatin or MMC, we identified 26 missense and 25 nonsense LOF mutations. Nonsense mutations elucidated a relationship between location of truncation and ICL sensitivity, as the majority of nonsense mutations before amino acid 860 confer ICL sensitivity. Further validation of a subset of LOF mutations confirmed the ability of the screen to identify FANCJ mutations unable to confer ICL resistance. Finally, mapping the location of LOF mutations to a new homology model additional functional information.

Implications: We identify 51 FANCJ LOF mutations, providing important classification of FANCJ mutations that will afford additional therapeutic strategies for affected patients.
INTRODUCTION

FANCJ (BRCA1-associated helicase 1, (BACH1) or BRCA1 interacting protein (BRIP1)) was identified as a direct interacting partner of BRCA1. Similar to BRCA1, FANCJ was shown to function in DNA repair and to be mutated in hereditary breast cancer patients (1). Unlike BRCA1 or BRCA2 clinical mutations that informed little about how DNA repair and tumor suppression were mediated, mutations in the FANCJ helicase domain provided a direct connection between DNA metabolism and tumor suppression (2). While the link between FANCJ and hereditary breast cancer suppression has been recently questioned, deleterious germline mutations in FANCJ are significantly associated with an increased risk of ovarian cancer (OC), observed in 0.9%-2.5% of OC patients (3-8). In fact, after BRCA1 and BRCA2, FANCJ is the third most common cancer susceptibility gene in OC (5,7,8). In addition to hereditary breast and ovarian cancer (HBOC), FANCJ mutations are found in melanoma, prostate, and hereditary colon cancer, providing evidence that FANCJ mutations may be a risk factor in multiple types of cancer (9-11). Indeed, FANCJ, BRCA1, and BRCA2 are bi-allelically mutated in Fanconi anemia (FA), a bone marrow failure disease that also predisposes to cancers such as leukemia (12).

Consistent with its roles as an ATPase, DNA helicase, and translocase, FANCJ contains a highly conserved helicase homology domain with seven conserved motifs including Walker A and Walker B boxes, as well as an iron-sulfur (Fe-S) cluster that are all essential for its catalytic activity (see Figure1A). The DNA-dependent ATPase function of FANCJ catalytically unwinds a range of duplex DNA substrates as well as secondary DNA
structures such as G-quadruplexes (G4s) (13-15). These DNA unwinding activities support efficient replication and the progression of cells through S-phase, the mobilization of DNA repair protein and the activation of checkpoint responses as well as DNA repair activities during replication stress (16-20). Most notably, FANCJ catalytic activity is required for the processing of ICLs, which requires coordination with the mismatch repair (MMR) protein, MLH1, that binds lysine residues 141 and 142 within the FANCJ helicase domain. Accordingly, loss of FANCJ catalytic activity or MLH1 binding causes exceedingly elevated sensitivity to ICL-inducing agents (21). Likewise, FA-associated FANCJ clinical mutations fail to restore ICL resistance, consistent with the role of FANCJ enzyme activity in ICL repair processing (14). While other pathogenic variants that disrupt FANCJ enzyme function, expression, or splicing have been identified (17,20,22,23), the majority of FANCJ sequence changes remain unclassified, thereby limiting clinical utility.

There are also several protein interactions outside the helicase domain in the less conserved C-terminal region of FANCJ. Most notably, the direct interaction with BRCA1 is mediated by phosphorylation of FANCJ serine 990 (Figure 1A) (24). While loss of this phosphorylation and BRCA1 binding does not sensitize to ICL-inducing agents, homologous recombination (HR) is reduced and the DNA damage tolerance mechanism of translesion synthesis (TLS) is enhanced (25,26). The phosphorylation of S990 also mediates the acetylation of lysine 1249 and its subsequent interaction with CtIP, an interaction important for DNA end resection (27,28). Correspondingly, loss of this K1249 acetylation also disrupts HR and promotes TLS (27). FANCJ is also phosphorylated at threonine 1133 in response to replication stress, facilitating a direct interaction with
TOPBP1, and promotion of an ATR-dependent checkpoint in response to replication stress (29). Additional DNA repair proteins, including MRE11, RPA, and BLM bind within the FANCJ carboxy terminus (881-1249), and although the binding parameters remain to be fully characterized, these interactions have been shown to regulate FANCJ enzyme activity (30-34). While only MLH1 binding has been shown to be essential for ICL resistance, other FANCJ interactions including BRCA1 and TOPBP1 modulate its DNA repair and checkpoint activities, respectively, in a manner that could be critical for tumor suppression (reviewed in (35)).

Here, we provide a comprehensive mutational analysis of the DNA helicase, FANCJ. We generated a library of nonsense and missense mutations obtained from the cBioPortal for Cancer Genomics and evaluated their function in a high-throughput screen (HTS) to evaluate sensitivity to ICL-inducing agents (36,37). We observed a distinct LOF phenotype due to 51 distinct FANCJ mutations (25 nonsense and 26 missense), all of which are located within the FANCJ helicase domain (1-880). Mapping the missense mutations to a homology model of FANCJ provided additional insight to the mechanism of disruption. This new information about the functional consequences of clinically-relevant FANCJ mutations will provide important insights to interpret cancer risk as well as to manage prevention and treatment strategies.
MATERIALS AND METHODS

FANCJ mutations and in silico programs

FANCJ mutations included in the study were identified using cBioPortal (http://cbioportal.org); more information on mutations found in Table S1 (36,37). Negative controls included synonymous mutations at same amino acid as cBioportal mutations as well as non-pathogenic mutations seen in healthy patients (http://gnomad.broadinstitute.org/) (38). The library also contained controls spiked into the plate including; no virus, virus expressing dsRed or eGFP, virus expressing FANCJ K52R or S990A. Missense mutations were evaluated using the following in silico programs, Mutation Assessor (http://mutationassessor.org/r3/) (39), Polyphen-2 (http://genetics.bwh.harvard.edu/pph2/index.shtml) (40) and SIFT (https://sift.bii.a-star.edu.sg/) (41). Protein alignments were performed using Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/) (42).

Cell lines and Reagents

HeLa and U2OS FANCJ K/O and CRISPR control (CC) cells (26,43) were grown in DMEM supplemented with 10% fetal bovine serum (Sigma Aldrich) and 1% Penicillin/Streptomycin (Gibco). HeLa FANCJ K/O cells tested negative for mycoplasma prior to the high-throughput screen. Cisplatin (Sigma Aldrich) was prepared in 1mM saline solution, per manufacturer’s instructions and MMC (Sigma Aldrich) was prepared in water.

Immunoblotting
Cells were harvested, lysed and processed for Western blot analysis as described previously using 150mM NETN lysis buffer (20mM Tris (pH 8.0), 150mM NaCl, 1mM EDTA, 0.5% NP-40, and Halt Protease inhibitor cocktail [ThermoFisher]). Proteins were separated using SDS–PAGE and electro-transferred to nitrocellulose membranes. Membranes were blocked in 5% not fat dry milk (NFDM) phosphate-buffered saline (PBS)/Tween and incubated with primary Ab overnight at 4°C. Antibodies for Western blot analysis included anti-β actin (Sigma-Aldrich), anti-FANCJ (E67). Membranes were washed, incubated with horseradish peroxidase-linked secondary Abs (Amersham) for 1h at room temperature and detected by chemiluminescence (Amersham).

**Lentiviral Production**

The mutant clones in the FANCJ mutation library were individually generated using site-directed mutagenesis (Genscript Piscataway, NJ) using PMT-BRD025 FANCJ-WT as template, and each mutant clone was sequence verified (Genscript Piscataway, NJ). Details of standard virus production pipelines can be found at the Broad Institute’s Genetic Perturbation Platform website (https://portals.broadinstitute.org/gpp/public/). Viruses for the mutant and WT FANCJ were produced in 96 well plates using HEK293T cells transfected with packaging vector psPAX2 (100 ng), envelope plasmid CMV-VSVG (10 ng), and respective PMT-BRD025 FANCJ mutant plasmid (100 ng). Lentiviral-containing supernatants were harvested 31 hours post-transfection and stored in polypropylene plates at -80°C until use.

**High Throughput Screen Conditions**
Mutant and WT FANCJ virus was robotically arrayed into 2×384 well plates, from which virus was later transferred to cell plates. Separately, 200 HeLa FANCJ K/O cells per well were seeded into 384-well white-walled, clear-bottom plates in a volume of 40 µL. Plates were incubated at room temperature for 30 minutes to aid with even seeding within each well. Then 10µL of media containing polybrene (4µg/mL) and 6µL of the arrayed lentivirus were sequentially added using the JANUS liquid handler and the plates were centrifuged for 30 min at 2250 RPM (1,178xg) at 37° C followed by overnight incubation at 37° C (5% CO2). The following morning, using liquid handling, 50 µL of media was removed and replaced with 50µL of fresh media. Cell plates were randomly divided into 6 treatment arms in duplicate: untreated, cisplatin, and MMC. Additionally, one plate was treated with puromycin with which to calculate infection efficiencies. 48 hours post-infection we added 10µL of media or drugs was added to each well of the plates to a final concentration of either: 500nM Cisplatin, 10nM MMC, 0.15 µg/mL puro, or media to the untreated plates. 5 days post-drug addition, cell viability was quantified through addition of 10µL of CellTiterGlo (Promega) and subsequently read out on an EnVision Multilabel Reader (Perkin Elmer). Percent viabilities were determined by dividing the average luminescence value in drug by the average luminescence value in the absence of drug.

**Structural modelling**

We used the Phyre2 server to generate a homology model of the FANCJ protein (44). The model was primarily based on structures of the FANCJ homologs XPD and DinG (45,46). The resulting model was built with high confidence in the modeled regions and had good stereochemistry. To incorporate the FeS cluster into the complex, we
superposed the XPD structure (45) onto the FANCJ model and used the coordinates of FeS cluster from the superposed XPD protein. The FeS cluster is positioned with high confidence due to the substantial sequence similarity in this region. To model how ATP and DNA bind to FANCJ, we superposed the DinG structure and used the superposed coordinates of these ligands. The ATP is positioned with high confidence due to the high conservation in the region and the well understood ATPase mechanism. The DNA is placed with much less confidence, primarily because of the weaker sequence similarity in the ssDNA binding groove. Therefore, the positioning of ssDNA in the resulting model should not imply confident prediction of which residues directly bind to the DNA; rather, the modeled DNA should only be used to localize the approximate location of the DNA strand.

**Statistics**

GraphPad Prism was used to calculate correlation. For the Z' factor assay, 200 HeLa FANCJ K/O cells/well in 384 plate were spinfected with 8ul FANCJ WT or FANCJ K52R lentivirus, and 24h post infection, media was changed and 48h post infection, plates were treated with 20nM MMC. Five days post MMC treatment, survival was measured and luminescence data from 176 FANCJ WT and 176 FANCJ K52R samples were used to determine Z' using the following equation:

\[ Z' = 1 - \frac{(3 \text{ SD of sample} + 3 \text{ SD of control})}{| \text{ mean of sample} - \text{ mean of control} |} \]
RESULTS

Generation of a FANCJ mutation library from cancer-associated mutations

Next-generation sequencing has identified mutations throughout the FANCJ gene, but the functional consequences remain mostly unclassified. To fully examine the significance of FANCJ clinical mutations, we evaluated nonsense and missense mutations obtained from the cBioPortal for Cancer Genomics, (36,37) the vast majority being somatic mutations (Table S1). We generated a lentiviral expression library of 595 mutations throughout the FANCJ gene, corresponding to 40 nonsense mutations and over 460 missense mutations (Table 1, Table S2). Positive controls included FANCJ K52R, a known LOF mutation that disrupts FANCJ catalytic activity (1,48,49), and negative controls included 92 synonymous mutations and 46 non-pathogenic mutations from healthy patients identified from Genome Aggregation Database (gnomAD) (Table 1) (38). FANCJ mutations were evenly distributed throughout the FANCJ gene product (Figure S1A). Evaluation of the mutations using the in silico mutation prediction program, Mutation Assessor, which uses evolutionary conservation of amino acids to predict functional impact of substitutions, revealed approximately half of the mutations are classified as high or medium risk of functional impact whereas the other half are classified as low or neutral functional impact (39,50). Similarly, additional predictive algorithms, SIFT and Polyphen-2, confirmed that approximately half of missense mutations are predicted to be deleterious whereas the other half tolerated (Figure S2) (40,41).

Design of high-throughput screen (HTS) to identify FANCJ LOF mutations
To evaluate the functional consequence of FANCJ mutants, we first identified a FANCJ-deficient cell line amenable to high-throughput screening. We took advantage of the sensitivity of FANCJ-deficient cells to ICL-inducing agents and assessed ICL sensitivity in two CRISPR-CAS9 generated FANCJ-knockout (KO) cell lines, osteosarcoma (U2OS FANCJ K/O) and cervical adenocarcinoma (HeLa FANCJ K/O) (26,43). We observed the expected sensitivity to the ICL-inducing agents, MMC and cisplatin, in both FANCJ K/O cell lines, when compared to CRISPR controls (CC), but the HeLa FANCJ K/O cells were distinctly more sensitive than U2OS FANCJ K/O cells (Figure 1B, Figure S1B). We confirmed complementation with wildtype (WT) FANCJ conferred ICL resistance in FANCJ K/O cells. Specifically, FANCJ K/O cells were infected with virus from a lentiviral construct expressing empty vector (EV), FANCJ WT, FANCJ K52R, a known helicase-dead and LOF mutant, or FANCJ S990A, the BRCA1 interaction-defective mutant (2,24). An immunoblot revealed lentiviral infection resulted in FANCJ expression, and sensitivity assays confirmed cisplatin resistance in cells expressing either FANCJ WT or FANCJ S990A, whereas cells expressing EV or FANCJ K52R failed to confer cisplatin resistance (Figure 1C-1E). Cisplatin treatment (250nM) resulted in a 9-fold difference in survival between EV and FANCJ WT in HeLa FANCJ K/O compared to <2-fold difference in survival in U2OS FANCJ K/O cells (Figure 1D, 1E). We therefore exploited the larger window in the HeLa cells to maximize our ability to identify FANCJ clinical mutations with LOF phenotypes in the HTS.

Finally, to evaluate whether screening conditions were robust and suitable to identify LOF mutations, we determined the Z-factor (Z’), a statistical test designed to evaluate signal range as well as data variation (47). HeLa FANCJ K/O cells were infected with either
FANCJ WT or FANCJ K52R virus, treated with MMC and survival was evaluated. Although this experiment only resulted in 57% infection efficiency (IE), we obtained a $Z' = 0.492$. Since a $Z' \geq 0.5$ indicates assay conditions are ideal for the HTS, this test suggested that our screen conditions could provide the sufficient quality and robustness to identify LOF mutations (Figure S1C) (47).

Identification of FANCJ LOF mutants with HTS

The HTS was performed in 384 well plates in technical duplicates under three experimental conditions; untreated, MMC, and cisplatin (Figure 2A). The use of these two similar ICL-inducing agents provided orthogonal biological replicates to help increase confidence in LOF mutants. The doses of MMC (10nM) and cisplatin (500nM) were chosen as the lowest dose that provided the largest signal window between sensitivity of FANCJ K/O K52R and resistance in FANCJ K/O WT cells (Figure S1D). An additional condition included in the screen was puromycin treatment to assess IE; removal of clones with < 50% IE prevented the incorrect classification of low expressors as LOF mutations. After exclusion of 14 mutations for low IE, we obtained sensitivity data for > 99.9% of the mutation library, with a mean IE of > 99%, indicating suitable complementation with the mutation library (Figure 2B). The two technical replicates of each ICL-inducing agents exhibited close correlation (Figure S3A). Additionally, the sensitivity of FANCJ mutants to MMC treatment correlated closely with the sensitivity to cisplatin treatment, consistent with the fact that both are ICL-inducing agents with similar mechanisms of action (Figure 2C). The controls performed as expected, with cells expressing FANCJ WT or FANCJ S990A conferring resistance to ICL-inducing agents and exhibiting increased survival, and cells infected with FANCJ K52R, eGFP, dsRed or mock-infected cells failing to confer
resistance to MMC or cisplatin treatment (Figure 2D). Furthermore, the synonymous mutation negative controls, which result in no alteration of gene product, as well as non-pathogenic mutations from healthy individuals exhibited resistance to cisplatin and MMC treatment, indicating complementation with these presumably functional FANCJ clones conferred resistance to ICL-inducing agents in our screen (Figure S3B).

To determine whether the FANCJ mutations in our library had the ability to confer resistance to ICL-inducing agents when introduced into FANCJ K/O cells, we evaluated survival following treatment with cisplatin and MMC separately. Mutants unable to confer cellular resistance to ICL-inducing agents were classified as FANCJ LOF mutants using the criteria of ≤ 50% survival following treatment with cisplatin or MMC, as compared to untreated cells. The known LOF mutations included in the screen, as expected, were classified as the LOF mutations, including the well-characterized dominant-negative helicase dead K52R mutation, as well as mutations observed in FA patients, A349P and H396D (20,51). We also observed LOF in G690E, a mutation recently characterized as a null mutation (52).

Examination of the 40 nonsense mutations revealed 25 LOF nonsense mutations and elucidated a clear relationship between location of FANCJ truncation and cisplatin sensitivity. With few exceptions, a truncation before amino acid 860 resulted in cisplatin sensitivity, whereas after amino acid 860 resulted in cisplatin resistance (Figure 3A and Table 2). This indicates that the first 860 amino acids of FANCJ, corresponding to the helicase domain, are required for the cellular resistance to ICL-inducing agents, and the C-terminal region (881-1249) of FANCJ, including the protein interaction domains located here, is dispensable for ICL resistance, reviewed in (9,14). Although the vast majority of
nonsense mutations before amino acid 860 result in cisplatin sensitivity, this relationship
was not absolute because three mutations (E357*, E795*, and R789*) retained the ability
to confer ICL resistance (Figure 3A).

Similar to the nonsense mutations, all of the missense mutations categorized as LOF
mutations mapped to the first ~860 amino acids of FANCJ, with the vast majority in
residues evolutionarily conserved between human, mouse and chicken (Figure 3B, Table
2, Figure S4). In addition, the majority of missense LOF mutations were classified as high
functional impact using Mutation Assessor, deleterious using SIFT, and damaging using
Polyphen-2, consistent with the clustering of the mutations primarily in the evolutionary-
conserved DEAH boxes and Fe-S motifs (Figure 3C and Figure 4A). These findings are
similar to the localization of LOF mutations in another DNA damage related helicase,
BLM, which primarily localized to structural motifs within the helicase domain (53). A
discrete set of six mutations were identified as FANCJ hypomorphs using the criteria of
relative survival of >50% and <70%, relative to untreated. One hypomorph identified is
R707C, a mutation recently characterized as having diminished dimerization and helicase
processivity, as well as increased sensitivity to cisplatin (54) (Figure 3D).

Mapping LOF mutations onto a FANCJ homology model
To investigate how the LOF mutations could disrupt FANCJ structure and function, we
built a homology model of the FANCJ protein. The Phyre2 server (44) generated a
structural model using several related helicases as template, covering ~55% of the
FANCJ sequence (Figure 4B). This high-confidence model includes the ATPase, FeS
cluster and the Arch domain, but the FANCJ C-terminal region could not be accurately
modeled because it is predicted to be largely intrinsically disordered.
Mapping the LOF mutations to this homology structure identified several distinct clusters of FANCJ LOF mutations that were primarily located in regions of known function. Only one mutation (S189L) mapped to the unstructured region between residues 66 and 240, and none were found in the long unstructured C-terminal tail. We classified the LOF mutations into four major classes that disrupt either the Fe cluster, the ATPase active site, ssDNA binding, or overall structure or folding (Table S3). To test these predictions, we analyzed thirteen LOF mutations identified in the primary screen that represent these four predicted functional classes.

Several mutations are predicted to perturb key catalytic residues that are highly conserved or invariant across the broad superfamily of ATPases of which FANCJ is a member (Table S3) (55). The G49R and K52R mutations alter highly conserved residues in the Walker A motif that are used to bind ATP (56), while D393V disrupts the invariant aspartate in the Walker B motif that is necessary for ATP hydrolysis (57). The R831K mutation disrupts a key residue for both ATPase activity and transmission of conformational changes; mutation of the equivalent ‘arginine finger’ residue in the related BLM helicase causes loss of ATPase and helicase activities (58). While H396D and S614Y do not disrupt highly conserved residues, their predicted proximity to the main catalytic machinery suggests these mutations also disrupt ATPase activity. Validation experiments confirm that these predicted ATPase-defective mutants indeed disrupt FANCJ function; G49R, D393V, S614Y, G690R and L860P do not confer ICL resistance when expressed in FANCJ K/O cells (Figure 4C and Figure S5).
Our model predicts another class of LOF mutations will disrupt the FeS domain (Table S3). The C350F mutation causes the loss of a cysteine residue that directly coordinates the FeS cluster, while mutation of the adjacent residue (A349P) is expected to distort the FeS binding residues. Indeed, mutation of the equivalent residue in the related protein XPD abolishes FeS cluster formation and helicase activity (59). We predict other mutations near the FeS cluster (W335L, L340R, V341D, L347P, L358P, and F366S) would destabilize the domain, leading to loss of FANCJ activity and consistent with this prediction, sensitivity assays clearly illustrate that V341D, F366S and L347P mutations fail to confer ICL resistance (Figure 4C and Figure S5).

A third class of mutations map to a region of FANCJ that our model predicts as a binding site for single-stranded DNA (ssDNA). By superposing FANCJ on the structure of the related DinG protein bound to ssDNA (46), we approximate the positioning of the ssDNA binding region into the groove between the two ATPase domains. Interestingly, we find that the T252R, S697P, S697F, R762P, and G763C mutations map to this region. Each of these residues are conserved in the related XPD helicase (T76, S541, R601, and G602 of XPD), suggesting a shared function. Due to their location in or near the DNA binding cleft, we predict that these mutations disrupt the binding of FANCJ to ssDNA. Consistent with the importance of this functional domain, FANCJ mutations S697P and R762P fail to confer ICL resistance (Figure 4C and Figure S5).

The final class of mutations are found in buried residues that are likely important for
overall FANCJ structure and stability. Because most of residues reside in the hydrophobic cores of individual domains, we predict that these mutations will disrupt the folding or otherwise alter FANCJ structure (Table S3). Residues in this class are found in domains throughout the modeled protein, including the FeS domain mutations (V341D, L347P, and L358P), the ATPase domain 1 (F366S), the Arch domain (L415P), and ATPase domain 2 (G690E/R, P785L, and L860P), and by the disruption of these domains, interfere with their function. It is likely that the mutations in this class disrupt FANCJ function through varied molecular mechanisms, such as reduced protein half-life or altered protein structure or dynamics. We found that the FANCJ mutants chosen for validation exhibit similar expression compared to WT FANCJ (Figure 4D), indicating that the mutations do not grossly perturb levels of FANCJ. Nevertheless, the functional consequence of this class of mutations was confirmed by the failure to confer ICL resistance by the V341D, L347P, F366S, G690R, and L860P FANCJ mutants (Figure 4C and Figure S5).

DISCUSSION

To leverage genomic information obtained in the past decades for therapeutics and diagnostics, understanding the functional consequences of genetic variations will be critical (60). Mutations in FANCJ have been associated with HBOC for years, but the physiological consequences of the majority of FANCJ mutations still remains unclear. Our comprehensive screen provides an important step in elucidating the physiological consequences of FANCJ mutations, specifically in terms of sensitivity to ICL-inducing
agents. We identified 25 nonsense and 26 missense FANCJ mutations exhibiting LOF phenotypes following treatment with DNA cross-linking agents. The majority of these LOF mutations are located in evolutionarily conserved amino acids, constrained to the first 860 amino acids of FANCJ, and positioned in domains important for DNA helicase activity including ATPase domains and the Fe-S motif. The absence of LOF mutations in the C-terminal region (residues 881–1249) suggest that the FANCJ N-terminal helicase domain is essential for tumor suppression and that C-terminal interactions with BRCA1, BLM, TOPBP1 and CtIP instead modulate this activity (14).

Our development of a FANCJ homology model and subsequent mapping of the identified LOF mutations facilitated their categorization into four classes of disrupted function; mutations that disrupt the Fe cluster, the ATPase active site, ssDNA binding, or overall structure or folding. The finding that rather than being randomly arranged throughout the protein, the LOF mutations are preferentially located in known regions essential for ICL resistance provides an important validation of the screen. Furthermore, the categorization of LOF mutations into functional classes may provide important insight about the functional severity of a mutation. For example, a DNA-binding mutant would be expected to be less detrimental than a mutant capable of binding DNA but unable to translocate DNA. This type of mutation was described recently in FeS cluster mutations and indeed exhibited greater MMC sensitivity compared to FANCJ KO cells (23). Other FANCJ mutations with similar impairment (K52R and A349P) have been shown to behave as a dominant negative (2,51). Providing additional information about potential dominant negative mutations is especially relevant given that this mutation screen examines the consequence of FANCJ mutations in a FANCJ deficient background. Therefore, the
functional impact of the identified LOF variants, and any potential actionability, will likely require homozygosity, another impairment of the wildtype FANCJ allele, or a dominant negative mutation.

Our screen determined that approximately 12% of clinically relevant FANCJ mutations result in a LOF phenotype, suggesting that the vast majority of clinical mutations do not alter the ability of FANCJ to confer cellular resistance to ICL-inducing agents. These results are comparable to the low frequency (11%) of BRCA1 mutations that exhibited cisplatin sensitivity in a large-scale mutation complementation assay (61). Similarly, in a recent study, only 30% of 20 mutations in FANCJ were designated null (52). However, deleterious effects (null or hypomorph phenotypes) were found in 75% of FANCJ mutations analyzed, possibly reflecting experimental differences such as higher MMC doses (52).

The low frequency of LOF mutations in our screen has several implications. For one, sensitivity to ICL-inducing agents may not be the optimal predictor of LOF. Instead, cancer-associated FANCJ mutations may result in disruption of disparate functions of FANCJ such as G4 resolution, stabilization of microsatellites, or regulation of replication stress (14,35,43,62,63). Roles for FANCJ have been identified in the suppression of HR associated gene duplication/amplification through recruitment of CtIP to damaged sites and promotion of DNA end resection (27,28). These numerous functions provide additional potential mechanisms by which FANCJ mutations may result in pathological phenotypes without exhibiting sensitivity to ICL-inducing agents. An important next step will be to elucidate other consequences of FANCJ mutations that retain cellular resistance to ICL-inducing agents. Performing the mutation screen under conditions that interrogate
replication stress response, G4 resolution, or other FANCJ functions would likely identify additional LOF mutations.

Additional alterations to our screening parameters would likely identify FANCJ gain-of-function (GOF) mutants. As designed, the screen conditions precluded the identification of GOF mutants, which exhibited greater ICL resistance than FANCJ WT. However, this will be an important future direction since evidence suggests the existence of FANCJ GOF mutations can promote chemoresistance and additional oncogenic properties. Consistent with this idea, the first two FANCJ helicase mutations identified in hereditary breast cancer patients suggested strict regulation of FANCJ unwinding activity is important for tumor suppression; FANCJ P47A reduced ATPase/unwinding activity whereas FANCJ M299I enhanced this activity (2,64). Moreover, mutations that disrupt BRCA1 binding (S990A mutation) or acetylation (K1249A mutation), although not currently clinically observed, shift replication towards TLS at the expense of HR, resulting in hyper-resistance to ICL-inducing agents (25,27). Moreover, we recently demonstrated TLS counters replication stress from genotoxins as well as oncogenes to limit replication gaps and consequently TLS is an adaptation present in many cancer cell lines (26).

Higher levels of unregulated FANCJ could be fundamental to this adaptation as non-redundant studies from cBioPortal cancer datasets (36,37) reveal that 15% of tumors show an increase in FANCJ copy number, suggesting FANCJ expression is elevated in a range of cancers, consistent with a recent findings that increased FANCJ expression is correlated with poor patient outcomes (65,66).

In summary, our development of a comprehensive patient-derived library of FANCJ mutations will be crucial to establish the mechanisms of pathogenicity of FANCJ
mutations, since many mutations have, thus far, remained uncharacterized. Our study is especially relevant because FANCJ is often included in multigene hereditary cancer panel testing. This and other ongoing studies investigating the functional consequences of mutations in HBOC genes will allow for improved screening, prevention, and specific targeted therapeutic strategies.

ACKNOWLEDGEMENTS

We thank the members of the Cantor Laboratory for helpful discussions and critical reading of the manuscript. We thank the University of Massachusetts Medical School RNAi Core Facility for technical assistance. This research was funded by grant numbers: NIH R01 CA225018-03 (S. Cantor) and NIH R01 CA225018-02S1 (S. Cantor).
References


Figure 1. The FANCJ helicase exhibits numerous mutations in cancer, with varying predictive severity, and deficiency of FANCJ or its helicase function results in severe ICL sensitivity.

A) Schematic representation of the FANCJ protein including Fe-S cluster, DEAH boxes including Walker A and Walker B boxes and the Arch domain. The 595 FANCJ mutations are aligned with the protein schematic and further classified by type of mutation; nonsense (n=40), synonymous (n=92), and missense mutations are further separated by their Mutation Assessor determination of high (n=53), medium (176), low (n=143), and neutral (n=81) assessment. B) Cisplatin sensitivity plotted for U2OS FANCJ K/O and HeLa FANCJ K/O as compared to appropriate CRISPR controls (CC). C) Immunoblotting illustrating FANCJ expression following infection with FANCJ lentivirus. Cisplatin sensitivity plotted for HeLa FANCJ K/O cells (D) or U2OS FANCJ K/O cells (E) following infection with FANCJ virus lentivirus.

Figure 2. High-throughput screen of FANCJ mutations provides conditions to identify FANCJ LOF mutations.

A) Schematic representation of high-throughput screen (HTS). HeLa FANCJ K/O (200/well) are infected with FANCJ mutation library in 384 well plates and left untreated or treated with cisplatin (500nM), MMC (10nM), or puromycin (0.15µg/ml). Five days after treatment, survival was quantitated using Cell-Titer Glo. B) Infection efficiency of FANCJ mutation library is shown. Mutations (n=14) found under the dotted line (<50% survival).
were excluded from screen due to low infection efficiency.  C) Correlation between cisplatin survival and MMC survival.  D) Percent survival of controls to MMC and cisplatin.

**Figure 3. FANCJ LOF mutations are localized within the helicase domain**

A) Schematic of FANCJ protein mapping location of nonsense mutations (n=40) and cisplatin sensitivity. Shaded area with reduced survival following cisplatin treatment illustrates LOF mutations are localized to first 860 amino acids.  B) Localization of LOF or hypomorphic missense mutations, aligned to FANCJ protein schematic in A.  C) Percent survival of entire mutation library to MMC and cisplatin. Missense mutations are categorized by their Mutation Assessor designation of high, medium, low, and neutral, and the two FA controls (A349P and H369D).  D) Percent survival of missense mutations to MMC and cisplatin. Mutations are classified as WT, LOF, hypomorph, and the two FA controls (A349P and H369D).

**Figure 4. FANCJ HTS identifies LOF mutations, localized in highly-conserved regions**

A) The location of the 26 FANCJ LOF missense mutations identified from screen are mapped to the FANCJ protein. FA-associated mutations are marked with an asterisk and mutations chosen for further validation are red.  B) Homology model of the FANCJ protein illustrating the location of the LOF mutations; model is shown in two orientations.  C) MMC sensitivity shown for HeLa FANCJ K/O cells infected with lentivirus expressing mutants predicted to disrupt ATPase domain, the FeS cluster, or DNA binding domain. The same
WT and EV control data is shown in each plot. D) Immunoblotting illustrates FANCJ expression following infection of lentivirus expressing EV, FANCJ WT, or FANCJ mutants.
Table 1: FANCJ mutation library: Description and enumeration of controls and mutations

<table>
<thead>
<tr>
<th>Controls</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Empty (no virus)</td>
<td>124</td>
</tr>
<tr>
<td>S990A (BRCA1 binding mutant)</td>
<td>8</td>
</tr>
<tr>
<td>K52R (LOF control)</td>
<td>8</td>
</tr>
<tr>
<td>dsRed or eGFP</td>
<td>16</td>
</tr>
<tr>
<td>FANCJ WT (control)</td>
<td>16</td>
</tr>
<tr>
<td>Total controls</td>
<td>172</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Mutations</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>cBioPortal</strong></td>
<td></td>
</tr>
<tr>
<td>Nonsense mutations</td>
<td>40</td>
</tr>
<tr>
<td>Missense mutations</td>
<td>396</td>
</tr>
<tr>
<td>In-frame deletion</td>
<td>5</td>
</tr>
<tr>
<td>cBioPortal total</td>
<td>441</td>
</tr>
<tr>
<td><strong>gnomAD</strong></td>
<td></td>
</tr>
<tr>
<td>Missense mutations</td>
<td>28</td>
</tr>
<tr>
<td>Synonymous mutations</td>
<td>18</td>
</tr>
<tr>
<td>gnomAD total</td>
<td>46</td>
</tr>
<tr>
<td>Silent mutations</td>
<td>72</td>
</tr>
<tr>
<td>Additional mutations (literature)</td>
<td>36</td>
</tr>
<tr>
<td>Total mutations</td>
<td>595</td>
</tr>
</tbody>
</table>
Table 2: FANCJ LOF and hypo-morphic mutations

<table>
<thead>
<tr>
<th></th>
<th>LOF</th>
<th>Hypomorphic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Missense</td>
<td>Nonsense</td>
<td>Missense</td>
</tr>
<tr>
<td>G49R</td>
<td>Q25*</td>
<td>A89V</td>
</tr>
<tr>
<td>K52R</td>
<td>G49*</td>
<td>P351L</td>
</tr>
<tr>
<td>S189L</td>
<td>G51*</td>
<td>R707C</td>
</tr>
<tr>
<td>T252R</td>
<td>E81*</td>
<td>S960T</td>
</tr>
<tr>
<td>W335L</td>
<td>Q126*</td>
<td>E1145A</td>
</tr>
<tr>
<td>L340R</td>
<td>Y147*</td>
<td>K1146E</td>
</tr>
<tr>
<td>V341D</td>
<td>G224*</td>
<td></td>
</tr>
<tr>
<td>L347P</td>
<td>Q227*</td>
<td></td>
</tr>
<tr>
<td>A349P</td>
<td>R261*</td>
<td></td>
</tr>
<tr>
<td>C350F</td>
<td>E387*</td>
<td></td>
</tr>
<tr>
<td>L358P</td>
<td>R439*</td>
<td></td>
</tr>
<tr>
<td>F366S</td>
<td>W448*</td>
<td></td>
</tr>
<tr>
<td>D393V</td>
<td>Y461*</td>
<td></td>
</tr>
<tr>
<td>H396D</td>
<td>Q561*</td>
<td></td>
</tr>
<tr>
<td>L415P</td>
<td>R581*</td>
<td></td>
</tr>
<tr>
<td>S614Y</td>
<td>S618*</td>
<td></td>
</tr>
<tr>
<td>E626K</td>
<td>Q645*</td>
<td></td>
</tr>
<tr>
<td>G690E</td>
<td>S653*</td>
<td></td>
</tr>
<tr>
<td>G690R</td>
<td>Q685*</td>
<td></td>
</tr>
<tr>
<td>S697P</td>
<td>Q689*</td>
<td></td>
</tr>
<tr>
<td>S697F</td>
<td>E726*</td>
<td></td>
</tr>
<tr>
<td>R762P</td>
<td>S772*</td>
<td></td>
</tr>
<tr>
<td>G763C</td>
<td>S805*</td>
<td></td>
</tr>
<tr>
<td>P785L</td>
<td>Q815*</td>
<td></td>
</tr>
<tr>
<td>R831K</td>
<td>R836*</td>
<td></td>
</tr>
<tr>
<td>L860P</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 2

A

HeLa FANCJ K/O 200 cells/well

FANCJ mutation library

Media

Puro

MMC

Cisplatin

Cell Titer Glo

B

Infection Efficiency (%)

FANCJ location (amino acid)

C

r=0.9429

Percent Survival (MMCT)

Percent Survival (Cisplatin)

D

Percent Survival (MMCT)

Percent Survival (Cisplatin)

- no virus
- WT
- S990A
- K52R
- eGFP/dsRed
Figure 3

A

Walker A
NLS
276
362
406
576
Walker B
Arch
888
1063
1249
Fe-S
ATPase/Helicase domain
BRCA1 interaction

B

Missense mutations

- LOF (n=26)
- hypomorph (n=6)

C

Percent Survival (MMC)

Percent Survival (Cisplatin)

D

Percent Survival (MMC)

Percent Survival (Cisplatin)
Figure 4

A

Walker NLS 276 362 406 576 888 1063 BRCA1 interaction
A TPase/Helicase domain

B

DNA LOF mutations
FeS cluster
ATP
L415
R831
90°

C

ATPase mutants
FeS mutants
DNA binding mutants

D

FANCJ
β-actin
Molecular Cancer Research

Comprehensive mutational analysis of the BRCA1-associated DNA helicase and tumor suppressor, FANCJ/BACH1/BRIP1

Jennifer A. Calvo, Briana Fritchman, Desiree Hernandez, et al.

Mol Cancer Res  Published OnlineFirst February 22, 2021.

Updated version
Access the most recent version of this article at:
doi:10.1158/1541-7786.MCR-20-0828

Supplementary Material
Access the most recent supplemental material at:
http://mcr.aacrjournals.org/content/suppl/2021/02/20/1541-7786.MCR-20-0828.DC1

Author Manuscript
Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

E-mail alerts
Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions
To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions
To request permission to re-use all or part of this article, use this link http://mcr.aacrjournals.org/content/early/2021/02/22/1541-7786.MCR-20-0828. Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.