Gain-of-function mutations amplify cytotoxic FAM111 protease activity in human genetic disorders

Saskia Hoffmann[1], Satyakrishna Pentakota[1], Andreas Mund[2], Peter Haahr[1], Fabian Coscia[2], Marta Gallo[1], Matthias Mann[2], Nicholas M. I. Taylor[3], Niels Mailand[1,4,#]

[1]Protein Signaling Program, [2]Proteomics Program and [3]Protein Structure and Function Program, Novo Nordisk Foundation Center for Protein Research, University of Copenhagen, DK-2200 Copenhagen, Denmark; [4]Center for Chromosome Stability, Department of Cellular and Molecular Medicine, University of Copenhagen, Blegdamsvej 3B, DK-2200 Copenhagen, Denmark

[#]Correspondence: niels.mailand@cpr.ku.dk

Abstract

Dominant missense mutations in the human serine protease FAM111A underlie perinatally lethal gracile bone dysplasia and Kenny-Caffey syndrome [1-3], yet how *FAM111A* mutations lead to disease is not known. We show that FAM111A proteolytic activity suppresses DNA replication and transcription by displacing key effectors of these processes from chromatin, triggering rapid programmed cell death by Caspase-dependent apoptosis to potently undermine cell viability. Patient-associated point mutations in FAM111A exacerbate these phenotypes by hyperactivating its intrinsic protease activity. Moreover, FAM111A forms a complex with the uncharacterized homologous serine protease FAM111B, point mutations in which cause a hereditary fibrosing poikiloderma syndrome [4], and we demonstrate that disease-associated FAM111B mutants display amplified proteolytic activity and phenocopy the cellular impact of deregulated FAM111A catalytic activity. Thus, patient-associated *FAM111A* and *FAM111B* mutations may drive multisystem disorders via a common gain-of-function mechanism that relieves inhibitory constraints on their protease activities to powerfully undermine cellular fitness.

Main text

Dominant missense mutations in human *FAM111A* underlie Kenny-Caffey syndrome and gracile bone dysplasia, which present with a broad spectrum of severe growth abnormalities with variable expressivity, including thin and brittle bones, dwarfism, facial dysmorphism and splenic hypoplasia [1-3]. FAM111A is a prospective host restriction factor that has been linked to DNA replication [5-7] but whose precise cellular function remains poorly understood. In addition to a PCNA-binding PIP box, FAM111A contains a functionally uncharacterized C-terminal serine protease domain, which

harbors a conserved catalytic triad and displays homology to stress-responsive *E. coli* Deg-type proteases (Fig. 1a; Supplementary Fig. 1). To explore the cellular role of FAM111A proteolytic activity, we generated stable cell lines conditionally expressing wild-type (WT) GFP-tagged human FAM111A or mutant alleles containing inactivating substitutions in the protease domain (D439N) or PIP box (*PIP) [7] at two- to fourfold higher levels than endogenous FAM111A (Fig. 1b; Supplementary Fig. 2). EdU incorporation analysis revealed that expression of ectopic FAM111A in U2OS cells rapidly and potently suppressed DNA replication in a manner that was fully dependent on its protease activity (Fig. 1c-e). Elevated FAM111A WT expression also reduced DNA synthesis rates in HCT116 cells (Supplementary Fig. 2). Induction of WT but not catalytically inactive FAM111A triggered a substantial decrease in the association of PCNA but not the MCM complex with chromatin in S phase cells (Fig. 1d,f,g). The resulting decline in PCNA replication foci intensity was accompanied by RPA accumulation at these sites (Supplementary Fig. 2), indicating that FAM111A-mediated dissociation of the PCNA module from active replisomes triggers replication stress. While FAM111A has been suggested to promote PCNA loading during DNA replication [7], we observed no detectable impact of CRISPR/Cas9- or siRNA-mediated depletion of endogenous FAM111A on DNA synthesis and cell proliferation (Supplementary Fig. 3), suggesting it is dispensable for DNA replication but may have a perturbation-dependent function in regulating this process. Unlike the FAM111A protease domain, mutation of the PIP box that abrogates PCNA binding [7] and localization to PCNA-positive replication foci (Supplementary Fig. 2) did not impair the ability of FAM111A to suppress DNA replication and PCNA loading (Fig. 1b,d-f). This raised the possibility that FAM111A predominantly exerts its protease-dependent impact on DNA replication via other replisome components. Supporting this idea, an unbiased FAM111A interactome analysis revealed the Replication Factor C (RFC) complex, the main cellular PCNA loader [8], but not PCNA itself as a major replication fork-associated cellular

FAM111A binding partner (Fig. 1h), in agreement with genetic data indicating a link between FAM111A and RFC subunits in restricting viral replication [9]. Co-immunoprecipitation experiments suggested that FAM111A associates with the RFC complex through interaction with the RFC1 subunit independently of the PIP box, involving a region of the extended regulatory N-terminal portion of human RFC1 that encompasses its BRCT domain [8] (Fig. 1i,j; Supplementary Fig. 4). In line with this, endogenous FAM111A and RFC1 colocalized in nucleoli in G1 and G2 phases and formed nuclear foci during S phase (Supplementary Fig. 4). Ectopic FAM111A expression diminished RFC1 chromatin association but not total cellular RFC abundance in a protease-dependent but PIP box-independent manner (Fig. 1k,l; Supplementary Fig. 4), which may account for its suppressive impact on PCNA loading. These findings suggest that FAM111A inhibits DNA replication by promoting RFC and PCNA displacement from chromatin via its intrinsic protease activity.

We noticed that induction of WT FAM111A but not the catalytically inactive D439N mutant in U2OS cells led to a drastic loss of cell viability (Fig. 2a). This was accompanied by robust formation of Caspase 3 and PARP1 cleavage products as well as extensive DNA fragmentation demarcated by accumulation of cells with sub-G1 DNA content and pan-nuclear γ-H2AX positivity, phenotypes that could be suppressed by the pan-Caspase inhibitor Z-VAD-FMK but were insensitive to p53 status (Fig. 2b-e; Supplementary Fig. 5). We therefore concluded that elevated FAM111A protease activity potently triggers programmed cell death via Caspase-dependent apoptosis. Ectopic FAM111A expression also elicited an apoptotic response in HCT116 cells (Supplementary Fig. 5). While FAM111A-mediated apoptosis had no impact on DNA replication suppression and proceeded with slower kinetics, it affected cells at all stages of interphase (Supplementary Fig. 5), suggesting that the adverse impact of FAM111A protease activity on cell

viability is not solely a consequence of disrupting DNA synthesis. Indeed, in addition to its effect on DNA replication, elevated expression of FAM111A suppressed transcription with comparable kinetics in a manner that was largely dependent on its proteolytic activity but neither the PIP box nor cell cycle status (Fig. 1c; Fig. 2f-h; Supplementary Fig. 5). Consistently, similar to its impact on RFC1, induction of catalytically active FAM111A led to a robust decrease in chromatin-bound but not total levels of RPB1, the catalytic subunit of RNA Polymerase II, and FAM111A and RPB1 formed a complex in cells (Fig. 2i-k; Supplementary Fig. 4). Thus, elevated FAM111A protease activity antagonizes essential chromatin-associated processes including replication and transcription by evicting key effectors of these processes, the collective action of which may lead to rapid cell death by Caspase-dependent apoptosis.

Dominant missense mutations in human *FAM111A*, which cluster on the periphery of the protease domain and are predicted to be surface-exposed (Fig. 3a; Supplementary Fig. 1), are causative of the skeletal disorders Kenny-Caffey syndrome and perinatally lethal gracile bone dysplasia [1-3, 10], yet how these mutations impair normal physiology and whether they act via a gain- or loss-of-function mechanism is not known. To address this, we generated a panel of cell lines conditionally expressing different patient-associated FAM111A alleles or FAM111A WT at a comparable, near-endogenous level (Fig. 3b; Supplementary Fig. 6). Strikingly, despite their low abundance, the disease mutants strongly exacerbated the impact of GFP-FAM111A on the kinetics and magnitude of DNA replication and transcription inhibition as well as apoptosis onset (Fig. 3b-f; Supplementary Fig. 6). Consistently, mutant FAM111A markedly enhanced RFC1 and RPB1 dissociation from chromatin (Fig. 3g,h; Supplementary Fig. 6). We observed similar effects in cells expressing untagged ectopic FAM111A disease mutants, ruling out a contribution of epitope tagging to these phenotypes (Supplementary Fig. 6). Moreover, complementation experiments showed that sub-

endogenous levels of disease-associated FAM111A alleles were sufficient to trigger apoptosis in cells depleted of endogenous FAM111A (Supplementary Fig. 6). Introducing an inactivating D439N substitution abrogated the impact of the patient-associated FAM111A mutants on DNA replication, transcription and apoptosis (Fig. 3b-h; Supplementary Fig. 6), raising the possibility that they aggravate FAM111A-dependent phenotypes by amplifying its catalytic activity. Supporting this notion, FAM111A cleavage fragments that required its intrinsic protease activity but were insensitive to Caspase inhibition by Z-VAD-FMK were produced by disease-associated mutants but not the WT protein (Fig. 3i,j). Importantly, using purified recombinant proteins we observed elevated auto-proteolytic cleavage of a FAM111A disease mutant *in vitro* (Fig. 3k; Supplementary Fig. 7), providing direct evidence that patient-associated FAM111A mutations amplify its protease activity. In contrast, introducing a disease mutation had little impact on the subcellular localization of FAM111A and its association with interacting proteins (Fig. 3l; Supplementary Fig. 6). While dominant mutations in FAM111A underlie Kenny-Caffey syndrome, a variant form of this disorder can also be caused by recessive mutations in the tubulin-specific chaperone TBCE [11]. Similar to elevated FAM111A protease activity, we found that loss of TBCE expression led to impaired cell survival, and FAM111A patient mutants perturbed microtubule organization in a protease-dependent manner (Supplementary Fig. 8), reminiscent of the impact of disease-associated *TBCE* mutations [11], further implicating deregulated FAM111A protease activity in Kenny-Caffey syndrome etiology. Taken together, the data suggest that patient-associated point mutations in human FAM111A exert their adverse impact on cell and organismal fitness by hyperactivating its intrinsic protease activity.

Mammalian cells encode a second, uncharacterized FAM111 family protein, FAM111B, which like FAM111A harbors a C-terminal serine protease domain that is predicted to be catalytically active

but is dispensable for normal DNA replication and cell proliferation (Fig. 4a; Supplementary Fig. 1; Supplementary Fig. 3; Supplementary Fig. 9). Heterozygous missense mutations in human *FAM111B* are causative of a Rothmund-Thomson-like syndrome that manifests primarily as poikiloderma, myopathy, pulmonary fibrosis and tendon contractures [4, 12, 13], but whose molecular basis has not been explored. As for FAM111A, patient-associated FAM111B mutations map to the protease domain boundaries and are predicted to be surface-exposed (Fig. 4a; Supplementary Fig. 1), suggesting possible commonalities in the mechanisms underlying their disease-promoting potential. Interestingly, the FAM111A interactome revealed FAM111B as a candidate FAM111A-binding protein (Fig. 1h). Indeed, an association between FAM111A and FAM111B in cells could be readily detected by reciprocal co-immunoprecipitation analysis, and *in vitro* binding experiments with purified proteins demonstrated that this interaction is direct (Fig. 4b-d). Remarkably, using cell lines expressing GFP-tagged disease-associated FAM111B alleles at endogenous levels (Fig. 4e; Supplementary Fig. 9), we found that despite FAM111A and FAM111B mutations are causative of different syndromes, FAM111B patient mutants phenocopied the ability of elevated FAM111A protease activity to potently trigger replication and transcription shutdown, disruption of microtubule network integrity and cell death via Caspase-dependent apoptosis (Fig. 4e-k; Supplementary Fig. 9), suggesting these proteases have at least partially overlapping cellular functions. In line with this, FAM111B also interacted with RFC subunits and RPB1, and expression of a FAM111B disease mutant promoted dissociation of RFC1, PCNA and RPB1 from chromatin (Supplementary Fig. 10). Unlike FAM111A, however, these phenotypes were only induced by expression of patient-associated FAM111B alleles but not the WT protein in our cell lines (Fig. 4e-k; Supplementary Fig. 9; Supplementary Fig. 10). Consistently, FAM111B disease mutants showed prominent, Caspase-independent formation of faster-migrating cleavage fragments that were not observed for WT FAM111B (Fig. 4e; Supplementary Fig. 9). Importantly, an inactivating D544N

substitution in the catalytic protease domain of FAM111B disease mutants abolished both their adverse impact on DNA synthesis, transcription, microtubule organization and apoptosis induction, as well as the appearance of FAM111B proteolytic cleavage products (Fig. 4e-k; Supplementary Fig. 9; Supplementary Fig. 10). Moreover, a disease-associated FAM111B S628N mutant displayed markedly elevated auto-proteolytic activity *in vitro* (Fig. 4l; Supplementary Fig. 7). Thus, similar to FAM111A, patient mutations in FAM111B undermine cellular fitness by eliminating inhibitory constraints on its intrinsic protease activity, imposing a block to DNA and RNA synthesis and promoting apoptotic cell death. However, despite FAM111A and FAM111B form a complex, their adverse impact on DNA-associated processes were not interdependent, as elevated FAM111A protease activity triggered these phenotypes in FAM111B-depleted cells and *vice versa* (Supplementary Fig. 11).

Our findings suggest a common gain-of-function mechanism for how dominant mutations in FAM111A and FAM111B lead to multisystem disorders, driven by illegitimate amplification of their intrinsic proteolytic activities, which are detrimental to cellular fitness. Accordingly, identifying the targets of FAM111 protease activity will be an important task for the future. Although both RFC1 and RPB1 emerge as attractive candidate FAM111 substrates based on our findings, we have so far been unable to conclusively establish such relationships. While the biological functions of FAM111 proteases are not well understood, our finding that FAM111 proteolytic activity is dispensable for normal cell growth but powerfully suppresses DNA replication, transcription and cell viability once elevated is well aligned with the putative role of FAM111A as a host restriction factor suppressing Simian Virus 40 (SV40) and poxvirus gene expression and DNA replication [5, 6, 9]. When properly controlled, these features could provide an important contribution towards a general line of defense against invading viruses and their

propagation. Given the strikingly similar cellular consequences of amplified FAM111A and FAM111B proteolytic activity, their homologous protease domains and direct interaction, it seems likely that FAM111B may also act as a host restriction factor via a FAM111A-like mechanism of action. In keeping with the notion that endogenous levels of patient-associated FAM111A and FAM111B mutants are sufficient to powerfully antagonize cellular fitness, the markedly different tissue expression profiles of *FAM111A* and *FAM111B* mRNAs (Supplementary Fig. 12) offer a plausible rationale for why mutations in these proteases have distinct clinical manifestations despite their comparable cellular impacts. It is also possible that the cellular functions and substrates of FAM111A and FAM111B do not fully overlap. Our collective results suggest a potential of specific inhibitors of FAM111A or FAM111B catalytic activity as a rational treatment strategy for patients with fibrosing poikiloderma, gracile bone dysplasia or Kenny-Caffey syndrome caused by mutations in these proteases. Moreover, they pave the way for further dissection of the clinical heterogeneity and precise molecular underpinnings of human syndromes caused by mutations in FAM111A and FAM111B, which may shed more light on the physiological functions of these proteases.

Acknowledgements

We thank Blanca Lopez Mendez, Kata Sarlós, Veronika Baráth and the Novo Nordisk Foundation Center for Protein Research imaging platform for providing reagents and technical assistance. This work was supported by Novo Nordisk Foundation (grants no. NNF14CC0001 and NNF18OC0030752), The Lundbeck Foundation (grant no. R303-2018-3212), European Research Council (ERC, grant agreement no. 616236 (DDRegulation)), Danish Cancer Society (grant no.

R231-A13972), and Danish National Research Foundation (grant no. DNRF115).

Author contributions

Conceptualization, S.H. and N.M.; Methodology, S.H., S.P., A.M., P.H., F.C., M.M., N.M.I.T. and N.M. Investigation, S.H., S.P., A.M., P.H., F.C., M.G.; Writing – Original Draft, N.M.; Supervision, M.M., N.M.I.T. and N.M.; Project Administration, N.M.; Funding Acquisition, S.H., N.M.

Competing interests statement

The authors declare no competing financial interests.

Methods

Plasmids and siRNAs

Full-length human *FAM111A* and *FAM111B* cDNAs were inserted into the destination vectors pcDNA4/TO/EGFP or pcDNA4/TO using Gateway LR Clonase (Invitrogen) to allow for doxycycline-inducible expression of the transgenes. Constructs encoding FLAG-tagged RFC1 were generated by inserting full-length human *RFC1* cDNA into pcDNA3.1+/FLAG and pFLAG-CMV2 vectors (containing C- and N-terminal FLAG tags, respectively). For expression in yeast, full length

human *FAM111A* and *FAM111B* cDNAs were cloned into the pYES-FLAG-SNAP-TopII vector using the restriction endonucleases XmaI and XhoI. RFC1 deletion fragments were amplified by PCR and inserted into pcDNA3.1+/FLAG vector. Full-length human *RFC2, RFC3, RFC4* and *RFC5* cDNAs were inserted into the destination vector pcDNA4/TO/HA-FLAG. Point mutations in FAM111A and FAM111B were introduced using the QuikChange Site-Directed Mutagenesis kit (Agilent Technologies) according to the manufacturer's protocol. The FAM111A *PIP mutant [7] was generated by introducing Y24A and F25A substitutions into full-length FAM111A. Plasmids for generation of cell lines with targeted knockout of FAM111A and/or FAM111B (*ΔFAM111A, ΔFAM111B* and *ΔFAM111A+ΔFAM111B*) using CRISPR/Cas9 were constructed as described [14], using the pX459 plasmid (Addgene #62988) for Cas9 expression and sgRNA delivery. Briefly, sgRNA sequences were ordered as complementary primers, mixed in a 1:1 ratio, and annealed. Subsequently, pX459 was digested with BbsI and the sgRNA introduced using a normal ligation reaction according to the manufacturer's instructions (New England Biolabs). The following sequences were used: *FAM111A* sgRNA #2 (forward): 5′-CACCGAAGAGCCACAACTAATACCC-3′; *FAM111A* sgRNA #2 (reverse): 5′-AAACGGGTATTAGTTGTGGCTCTTC-3′; *FAM111B* sgRNA #2 (forward): 5′-CACCGTAAACTCACAAGTTAGACGG-3′; *FAM111B* sgRNA #2 (reverse): 5′-AAACCCGTCTAACTTGTGAGTTTAC-3′.

Plasmid DNA and siRNA transfections were performed using FuGENE 6 Transfection Reagent (Promega) and Lipofectamine RNAiMAX (Invitrogen), respectively, according to the manufacturers' protocols. All siRNAs were used at a final concentration of 50 nM. The following siRNA oligonucleotides were used: Non-targeting control (CTRL): 5′-GGGAUACCUAGACGUUCUA-3′; FAM111A (#3): 5′-GGAGAAUGAUGAUUGGAAA-3′; FAM111A (#5): 5′- GGGAAGAAUAACAAGAUUA-3′; FAM111B (#1): 5′-

CCUGUUGAUCAUUGUCUAU-3′; p53 (#1): 5′-CUACUUCCUGAAAACAACG-3′; p53 (#2): 5′-GAAAUUUGCGUGUGGAGUA-3′; p53 (#3): 5′-GACUCCAGUGGUAAUCUAC-3′; RFC1: 5′-GGUAUGAGCAGUAGGCUUA-3′; TBCE (#1): 5′-CAGACUUUCUUACUGCAAU-3′; TBCE (#2): 5′-CCUUGAGUCUAACAACAUU-3′.

Cell culture

Human U2OS and HCT116 cells were obtained from ATCC. All cell lines used in this study were cultured in DMEM containing 10% FBS, and were regularly tested negative for mycoplasma infection. To generate U2OS or HCT116 cell lines inducibly expressing GFP-tagged or untagged FAM111A as well as GFP-FAM111B WT and mutant alleles, cells were co-transfected with pcDNA4/TO/GFP or pcDNA4/TO expression constructs and pcDNA6/TR (Invitrogen). Positive clones were selected by incubation in medium containing 5 µg/ml blasticidin S (Invitrogen) and 400 µg/ml zeocin (Invitrogen) for 14 days. To generate cell lines with targeted knockout of FAM111A and/or FAM111B (*ΔFAM111A, ΔFAM111B* and *ΔFAM111A+ΔFAM111B*), parental U2OS cells were transfected with pX459-sgFAM111A#2 (*ΔFAM111A)*, pX459-sgFAM111B#2 (*ΔFAM111B)* or a 1:1 mix of pX459-sgFAM111A#2 and pX459-sgFAM111B#2 (*ΔFAM111A+ΔFAM111B*) and selected briefly with puromycin during clonal selection. Clones were screened for FAM111A and FAM111B expression by immunoblotting.

Unless otherwise indicated, the following drug concentrations were used: Nocodazole (10 µM, Sigma Aldrich), Z-VAD-FMK (50 µM, ab120487, Abcam) and doxycycline (1 µg/ml, Sigma Aldrich).

Immunochemical methods

For immunoblotting and immunoprecipitation, which were performed as previously described [15], cells were lysed in EBC buffer (50 mM Tris, pH 7.5; 150 mM NaCl; 1 mM EDTA; 0.5% NP40; 1 mM DTT) supplemented with phosphatase inhibitors and protease inhibitor cocktail (Roche).

Lysates were then incubated for 10 min on ice and sonicated. For immunoprecipitations, cleared lysates were incubated with FLAG agarose (Sigma-Aldrich), GFP-Trap Agarose (Chromotek) or anti-goat RFC1 antibody (2 µg/sample) coupled to Protein G agarose beads (Thermo Fisher Scientific) for 2 h on an end-over-end rotator at 4°C, washed in EBC buffer and treated with benzonase to minimize chromatin-mediated interactions. Proteins were then resolved by SDS–PAGE and analyzed by immunoblotting. For *in vitro* binding assays, recombinant proteins were diluted in EBC buffer and incubated with FAM111A antibody (2 µg/sample) coupled to Protein A agarose beads (Thermo Fisher Scientific) at 4°C for 2 hours on a rotary wheel, followed by washes in EBC buffer. For *in vitro* auto-cleavage assays, recombinant proteins were incubated in EBC buffer.

Immunofluorescence and high-content imaging analysis

Where indicated, cells were pre-extracted in PBS containing 0.2% Triton X-100 for 2 min on ice, before fixation with 4% formaldehyde for 15 min. If cells were not pre-extracted, they were subjected to a permeabilization step with PBS containing 0.2% Triton X-100 for 5 min and incubated with primary antibodies diluted in 1% BSA-PBS for 1 h at room temperature. Following staining with secondary antibodies (Alexa Fluor; Life Technologies) and 4′,6-diamidino-2-phenylindole dihydrochloride (DAPI, 0.5 µg/ml, DNA staining) diluted in 1% BSA-PBS for 1 h at room temperature, cells were mounted onto glass slides using ProLong Gold Antifade (Invitrogen). For EdU stainings, cells were treated with EdU (10 µM) for 30 min before fixation, then stained using the Click-iT Plus EdU Alexa Fluor 647 Imaging Kit (Invitrogen) according to the manufacturer's instructions before incubation with primary antibodies. For EU stainings, cells were treated with EU (1 mM) for 60 min before fixation, then stained using the Click-iT RNA Alexa Fluor 594 imaging kit (Invitrogen) according to the manufacturer's instructions.

Images were acquired with a Leica AF6000 widefield microscope (Leica Microsystems) equipped with HC Plan-Apochromat 63x/1.4 oil immersion objective, using standard settings. Image acquisition and analysis were carried out with Leica Application Suite X software (version 3.3.3.16958; Leica Microsystems). Raw images were exported as TIFF files and if adjustments in image contrast and brightness were applied, identical settings were used on all images of a given experiment. Quantitative image-based cytometry (QIBC) was performed as described [16]. Briefly, cells were fixed, permeabilized and stained as described above. Images were acquired with a ScanR inverted high-content screening microscope (Olympus) equipped with wide-field optics, UPLSAPO dry objectives (×20, 0.75-NA), fast excitation and emission filter-wheel devices for DAPI, FITC (fluorescein isothiocyanate), Cy3 and Cy5 wavelengths, an MT20 illumination system, and a digital monochrome Hamamatsu ORCA-R2 CCD camera. Automated and unbiased image analysis was carried out with the ScanR analysis software (version 2.7.1). Data was exported and processed using Spotfire software (version 10.5.0; Tibco).

Antibodies

Antibodies used for immunoblotting included: acetylated α-Tubulin (acetyl K40, ab24610, Abcam (1:5,000 dilution)), Actin (MAB1501, Millipore (1:20,000)), Cleaved Caspase-3 (Asp175, 9661, Cell Signaling (1:1,000)), CTCF (A300-543A, Bethyl (1:500)), FAM111A (ab184572, Abcam (1:1,000)), FAM111B (HPA038637, Sigma (1:2,000)), FLAG (A00187, GenScript (1:1,000)), GFP (11814460001, Roche (1:500)) and sc-8334, Santa Cruz (1:1,000)), GTFC1 (A301-292A, Bethyl (1:1,000)), histone H2AX (2595, Cell Signaling (1:1,000)), IFI-16 (sc-8023, Santa Cruz (1:500)), MCM2 (610701, Clone 46/BM28, BD Transduction Lab (1:1,000)), P53 (sc-126, Santa Cruz (1:1,000)), PARP-1 (sc8007, Santa Cruz (1:500)), PCNA (sc-56, Santa Cruz (1:1,000)), RFC1 (ab3566, Abcam (1:1,000)), RFC3 (ab154899, Abcam (1:1,000)), RFC5 (A300-146A, Bethyl (1:1,000)), TBCE (A305-485A, Bethyl (1:1,000)), tubulin alpha (T9026, Sigma-Aldrich (1:5,000)),

Vinculin (V9131, Sigma (1:10,000)). Antibodies used for immunoprecipitation (IP) included: FAM111A (ab184572, Abcam, 2 µg/IP) and RFC1 (ab3566, Abcam, 2 µg/IP). Antibodies used for immunofluorescence included: γH2AX (05-636 (Clone JBW301), Millipore (1:500)), FAM111A (ab184572, Abcam (1:300)), MCM2 (610701, Clone 46/BM28, BD Transduction Lab (1:500)), PCNA (#2037, Triolab Immunoconcepts (1:500)), RFC1 (sc-271656, Santa Cruz (1:300) or ab3566, Abcam (1:1,000)), RPA2 (NA19L (Clone Ab-3), Roche (1:1,000)), α-Tubulin (T9026, Sigma-Aldrich (1:5,000)).

DNA fiber assays

Exponentially growing U2OS cells (1×10^6) were labeled with consecutive pulses of CldU (25 µM) and IdU (250 µM) for 25 min. Cells were then trypsinized and resuspended in PBS. Cell suspension (2 µl) was spotted onto Superfrost glass slides and lysed in buffer containing 200 mM Tris-HCl, pH 7.4; 50 mM EDTA; 0.5% SDS for 2 min. Slides were tilted at an angle to allow the DNA to run slowly down the slide. Slides were air-dried before fixing in 3:1 methanol:acetic acid. DNA fiber spreads were denatured with 2.5 M HCl for 90 min before blocking in 2% BSA-PBS with 0.1% Tween for 30 min. Slides were then incubated with rat anti-BrdU (Abcam, ab6326; 1:100 dilution) for 1 h to detect CldU. Slides were washed in PBS-Tween and PBS before antibody cross-linking in 4% formaldehyde for 10 min. Slides were then incubated with Alexa Fluor 594 goat anti-rat antibody (Thermo Fisher; 1:100) for 1 h. Following similar washes, slides were incubated with mouse anti-BrdU (BD Bioscience, #347580; 1:500) overnight at 4°C to detect IdU. Slides were washed and incubated with Alexa Fluor 488 goat anti-mouse antibody (Thermo Fisher; 1:100) for 1 h. After washing, the slides were air-dried and mounted with 50 µl Vectashield mounting medium (Vector Laboratories). Track lengths were measured using ImageJ software.

Proliferation and cell survival assays

Cell lines were seeded in 96 well plates in triplicates, and cell proliferation was determined for 4 days by incubation with 10 µg/ml of resazurin (Sigma) for 2 hr at 37ºC. Fluorescence was measured at 590 nm using a plate reader (FLUOstar® Omega, BMG Labtech). The obtained values were normalized to the values of the first day. For survival assays, cells were seeded in 6-cm dishes and treated with doxycycline to induce the expression of ectopic FAM111A or FAM111B alleles for the indicated times. After 72 hours, plates were washed once in PBS, left to dry, and stained with cell staining solution (0.5% w/v crystal violet, 25% v/v methanol). Finally, plates were washed three times in deionized water. Images were acquired using the GelCount™ (Oxford Optronix) colony counter.

Affinity purification and mass spectrometry (AP-MS)

Partial on-bead digestion was used for peptide elution from GFP-Trap Agarose (Chromotek). Briefly, 100 µl of elution buffer (2 M urea; 2 mM DTT; 20 µg/ml trypsin; 50 mM Tris, pH7.5) was added and incubated at 37°C for 30 min. Samples were alkylated with 25 mM CAA and digested overnight at room temperature before addition of 1% trifluoroacetic acid (TFA) to stop digestion. Peptides were desalted and purified with styrenedivinyl-benzene-reversed phase sulfonate (SDB-RPS) StageTips. Briefly, two layers of SDB-RPS were prepared with 100µl wash buffer (0.2% TFA in H_2O). Peptides were loaded on top and centrifuged for 5 min at 500 g, and washed with 150 µl wash buffer. Finally, peptides were eluted with 50 µl elution buffer (80% ACN and 1% ammonia) and vacuum-dried. Dried peptides were dissolved in 2% acetonitrile (ACN) and 0.1% TFA in water and stored at -20°C.

Liquid chromatography-mass spectrometry (LC-MS) analysis

Nanoflow LC-MS analysis of tryptic peptides was performed using a quadrupole Orbitrap mass spectrometer (Q Exactive HF-X, Thermo Fisher Scientific [17]) connected to an EASYnLC 1200 ultra-high-pressure system (Thermo Fisher Scientific). Approximatly 0.5 µg of peptides were

loaded on a 50 cm HPLC column (75 μm inner diameter, New Objective; in-house packed using ReproSil-Pur C18-AQ 1.9 lm silica beads; Dr Maisch GmbH, Germany). Peptides were separated using a linear gradient from 2 to 20% B in 55 min and stepped up to 40% in 40 min followed by a 5 min wash at 98% B at 350 nl/min, where solvent A was 0.1% formic acid in water and solvent B was 80% acetonitrile and 0.1% formic acid in water for a total duration of 100 min. The mass spectrometer was operated in "top-15" data-dependent mode, collecting MS spectra in the Orbitrap mass analyzer (60,000 resolution, 300-1,650 m/z range) with an automatic gain control (AGC) target of $3x10^6$ and a maximum ion injection time of 25 ms. The most intense ions from the full scan were isolated with an isolation width of 1.4 m/z. Following higher-energy collisional dissociation (HCD) with a normalized collision energy (NCE) of 27, MS/MS spectra were collected in the Orbitrap (15,000 resolution) with an AGC target of $1x10^5$ and a maximum ion injection time of 28 ms. Precursor dynamic exclusion was enabled with a duration of 30 s.

Bioinformatic analyses

Raw MS files were processed using the MaxQuant software (version 1.6.5.0) (Cox & Mann, 2008). The integrated Andromeda search engine [18] was used for peptide and protein identification at an FDR of <1% and s_0 value of 1 [19]. Missing values were imputed based on a normal distribution (width=0.15; downshift=1.8). The human UniProtKB database (January 2019) was used as forward database and the automatically generated reverse database for the decoy search. A minimum number of 7 amino acids was used for the peptide identification. Proteins that could not be discriminated by unique peptides were pooled in the same protein group [20]. Label-free protein quantification was done using the MaxLFQ algorithm [21]. Protein ratios were calculated based on median peptide ratios, and only common peptides were used for pairwise ratio calculations. The "match-between-runs" feature of MaxQuant was enabled to transfer peptide identifications across runs based on high mass accuracy and normalized retention times. All statistical and bioinformatic

analyses were performed using Perseus [22] or the R framework (https://www.r-project.org/). The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE [23] partner repository with the dataset identifier PXD017978.

Purification of recombinant FAM111 proteins

Identical conditions were used for purification of all FLAG-tagged recombinant FAM111A or FAM111B proteins. Briefly, JEL-1 yeast strain transformed with FLAG-FAM111A or FLAG-FAM111B expression plasmid was grown in SC-URA medium containing 2% glucose for 36 h at 30 °C, and subsequently in SC-URA medium containing 2% raffinose for 24 h at 30 °C. The culture was then grown in YEP medium containing 2% raffinose and grown until OD 0.8. FLAG-FAM111 protein expression was induced by adding galactose (2%) for 24 h at 20 °C. Cells were collected and resuspended in lysis buffer containing 50 mM Tris, pH 7.5; 500 mM NaCl; 10% glycerol; 1 mM EDTA; 0.1 mM DTT. Glass beads were added to the resuspended cells, which were lysed by vortexing and cleared by centrifugation at 25,000*g* at 4 °C for 30 min. The cleared lysate was incubated with FLAG M2 resin (Sigma-Aldrich) and incubated at 4 °C for 2 h. After extensive washing, FLAG-FAM111A or FLAG-FAM111B was eluted in lysis buffer supplemented with FLAG peptide (0.5 mg/mL). The elute fractions were run on a 4-12% NuPAGE Bis-Tris protein gel (Invitrogen) and stained with Instant Blue Coomassie Protein Stain (Expedeon). Fractions were concentrated on Microcon-30kDa centrifugal filters (Millipore), snap-frozen in liquid nitrogen and stored at -80 °C.

Homology modeling of FAM111A and FAM111B protease domains

Domain identification in human FAM111A and FAM111B with Conserved Domains Database (CDD) [24] (https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) identified residues 371-555 for FAM111A and residues 471-664 for FAM111B to contain conserved trypsin-like serine protease domains. Full length FAM111A and FAM111B sequences were used as input in HHpred for the

identification of experimental structures of homologous proteins (https://toolkit.tuebingen.mpg.de/tools/hhpred). Among obtained hits, the crystal structure of *E. coli* DegS protease (2R3U) at 2.6 Å resolution [25] displayed the highest degree of homology and was used as a template for generating FAM111A and FAM111B protease domain models using Modeller [26]. The quality of the models was assessed based on the Ramachandran plot using Rampage (http://mordred.bioc.cam.ac.uk/~rapper/rampage.php), which assigned 85.7% of residues to favoured regions (2.9% outliers) for FAM111A and 88.1% of residues to favoured regions (4.5% outliers) for FAM111B. Overall model quality for both FAM111A (Z-score: -4.05) and FAM111B (Z-score: -3.82) was assessed using ProSA-web Protein Structure Analysis (https://prosa.services.came.sbg.ac.at/prosa.php). PyMol (v2.0) was used for visualization of the structures.

Flow cytometry

Cells were fixed in 70% ethanol and DNA was stained with propidium iodide (0.1 mg/ml) containing RNase (20 µg/ml) for 30 min at 37 °C. Flow cytometry analysis was performed on a FACS Calibur (BD Biosciences) using CellQuest Pro software (version 6.0; Becton Dickinson). The data was analysed using FlowJo software (version 10.6).

FAM111 ***transcript profiling***

Tissue RNA-seq data (RNA HPA tissue gene data) was downloaded from The Human Protein Atlas [27] (version 19.3) and Ensembl (version 92.38).

References

1. Unger, S. *et al.* FAM111A mutations result in hypoparathyroidism and impaired skeletal development. *Am J Hum Genet* **92**, 990-995 (2013).
2. Nikkel, S.M. *et al.* Mother-to-daughter transmission of Kenny-Caffey syndrome associated with the recurrent, dominant FAM111A mutation p.Arg569His. *Clin Genet* **86**, 394-395 (2014).
3. Isojima, T. *et al.* A recurrent de novo FAM111A mutation causes Kenny-Caffey syndrome type 2. *J Bone Miner Res* **29**, 992-998 (2014).
4. Mercier, S. *et al.* Mutations in FAM111B cause hereditary fibrosing poikiloderma with tendon contracture, myopathy, and pulmonary fibrosis. *Am J Hum Genet* **93**, 1100-1107 (2013).
5. Fine, D.A. *et al.* Identification of FAM111A as an SV40 host range restriction and adenovirus helper factor. *PLoS Pathog* **8**, e1002949 (2012).
6. Tarnita, R.M., Wilkie, A.R. & DeCaprio, J.A. Contribution of DNA Replication to the FAM111A-Mediated Simian Virus 40 Host Range Phenotype. *J Virol* **93** (2019).
7. Alabert, C. *et al.* Nascent chromatin capture proteomics determines chromatin dynamics during DNA replication and identifies unknown fork components. *Nat Cell Biol* **16**, 281-293 (2014).
8. Shiomi, Y. & Nishitani, H. Control of Genome Integrity by RFC Complexes; Conductors of PCNA Loading onto and Unloading from Chromatin during DNA Replication. *Genes (Basel)* **8** (2017).
9. Panda, D., Fernandez, D.J., Lal, M., Buehler, E. & Moss, B. Triad of human cellular proteins, IRF2, FAM111A, and RFC3, restrict replication of orthopoxvirus SPI-1 host-range mutants. *Proc Natl Acad Sci U S A* **114**, 3720-3725 (2017).
10. Abraham, M.B. *et al.* Short stature and hypoparathyroidism in a child with Kenny-Caffey syndrome type 2 due to a novel mutation in FAM111A gene. *Int J Pediatr Endocrinol* **2017**, 1 (2017).
11. Parvari, R. *et al.* Mutation of TBCE causes hypoparathyroidism-retardation-dysmorphism and autosomal recessive Kenny-Caffey syndrome. *Nat Genet* **32**, 448-452 (2002).
12. Mercier, S. *et al.* Expanding the clinical spectrum of hereditary fibrosing poikiloderma with tendon contractures, myopathy and pulmonary fibrosis due to FAM111B mutations. *Orphanet J Rare Dis* **10**, 135 (2015).
13. Khumalo, N.P. *et al.* Poikiloderma, tendon contracture and pulmonary fibrosis: a new autosomal dominant syndrome? *Br J Dermatol* **155**, 1057-1061 (2006).
14. Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819-823 (2013).
15. Poulsen, M., Lukas, C., Lukas, J., Bekker-Jensen, S. & Mailand, N. Human RNF169 is a negative regulator of the ubiquitin-dependent response to DNA double-strand breaks. *J Cell Biol* **197**, 189-199 (2012).
16. Toledo, L.I. *et al.* ATR prohibits replication catastrophe by preventing global exhaustion of RPA. *Cell* **155**, 1088-1103 (2013).
17. Kelstrup, C.D. *et al.* Performance Evaluation of the Q Exactive HF-X for Shotgun Proteomics. *J Proteome Res* **17**, 727-738 (2018).

18. Cox, J. *et al.* Andromeda: a peptide search engine integrated into the MaxQuant environment. *J Proteome Res* **10**, 1794-1805 (2011).
19. Tusher, V.G., Tibshirani, R. & Chu, G. Significance analysis of microarrays applied to the ionizing radiation response. *Proc Natl Acad Sci U S A* **98**, 5116-5121 (2001).
20. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol* **26**, 1367-1372 (2008).
21. Cox, J. *et al.* Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Mol Cell Proteomics* **13**, 2513-2526 (2014).
22. Tyanova, S. *et al.* The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat Methods* **13**, 731-740 (2016).
23. Perez-Riverol, Y. *et al.* The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Res* **47**, D442-D450 (2019).
24. Lu, S. *et al.* CDD/SPARCLE: the conserved domain database in 2020. *Nucleic Acids Res* **48**, D265-D268 (2020).
25. Hasselblatt, H. *et al.* Regulation of the sigmaE stress response by DegS: how the PDZ domain keeps the protease inactive in the resting state and allows integration of different OMP-derived stress signals upon folding stress. *Genes Dev* **21**, 2659-2670 (2007).
26. Zimmermann, L. *et al.* A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at its Core. *J Mol Biol* **430**, 2237-2243 (2018).
27. Uhlen, M. *et al.* Proteomics. Tissue-based map of the human proteome. *Science* **347**, 1260419 (2015).

Figure legends

Figure 1.

FAM111A proteolytic activity displaces RFC from chromatin and inhibits DNA replication

a. Domain organization of human FAM111A. **b**. Immunoblot analysis of stable U2OS cell lines treated with Doxycycline (DOX) to induce expression of WT or mutant forms of GFP-FAM111A. **c**. DNA replication rates in U2OS/GFP-FAM111A cells treated with DOX for the indicated times, pulse-labeled with EdU and stained with DAPI were analyzed by quantifying EdU signal intensity in S phase cells using quantitative image-based cytometry (QIBC) (red bars, mean (A.U., arbitrary units); n>2000 cells per condition). **d**. Cells treated as in (c) were pre-extracted, fixed and stained with PCNA or MCM2 antibody, and subsequently analyzed by QIBC (n>2000 cells per condition). **e-g**. Quantification of data in (d) (red bars, mean). Cells in S phase were identified based on EdU positivity. **h**. Analysis of FAM111A interactors. U2OS/GFP-FAM111A WT cells were treated or not with DOX for 4 h, subjected to GFP immunoprecipitation (IP) and analyzed by mass spectrometry. Volcano plot shows enrichment of individual proteins (+DOX/-DOX ratio) plotted against the P value. Dashed lines indicate the significance thresholds (FDR<0.05; s_0=1). **i**. U2OS or U2OS/*ΔFAM111A* cells were subjected to IP with IgG (control) or RFC1 antibody followed by immunoblotting with indicated antibodies. **j**. U2OS cells transfected with empty vector (EV) or indicated RFC subunit expression plasmids were subjected to FLAG IP and immunoblotted with indicated antibodies. **k**. As in (d), except that cells were stained with RFC1 antibody (n>2000 cells per condition). **l**. Quantification of data in (k) for S phase (EdU-positive) cells (red bars, mean). Data are representative of at least three (**b-g**,**i-l**) and two (**h**) independent experiments with similar outcomes.

Figure 2.

FAM111A protease activity suppresses transcription and triggers Caspase-dependent apoptosis

a. U2OS/GFP-FAM111A cell lines were fixed at the indicated times after DOX treatment and stained with crystal violet. **b**. Immunoblot analysis of U2OS/GFP-FAM111A WT cells treated with DOX for the indicated times. **c**. Representative images of U2OS/GFP-FAM111A cell lines that were treated or not with DOX, fixed and stained with γ-H2AX antibody and DAPI. **d**. Cells in (c) were subjected to QIBC analysis of γ-H2AX signal intensity (red bars, mean (A.U., arbitrary units); *n*>2000 cells per condition). **e**. Immunoblot analysis of U2OS/GFP-FAM111A WT cells treated or not with DOX and the pan-Caspase inhibitor Z-VAD-FMK as indicated. **f-h**. Transcriptional activity in U2OS/GFP-FAM111A cells treated with DOX for the indicated times, pulse-labeled with EU and stained with DAPI were analyzed by QIBC (red bars, mean; *n*>2000 cells per condition). **i**. U2OS/GFP-FAM111A cell lines treated or not with DOX were pre-extracted, fixed and stained with RPB1 antibody, and analyzed by QIBC (*n*>2000 cells per condition). **j**. Quantification of data in (i) (red bars, mean). **k**. Representative images from the experiment in (i). Scale bars, 10 μm.

Data (**a-k**) are representative of at least three independent experiments with similar outcomes.

Figure 3.

Patient-associated mutations hyperactivate FAM111A protease activity to exacerbate its adverse impact on cellular fitness

a. Overview of heterozygous *FAM111A* mutations found in patients with gracile bone dysplasia or Kenny-Caffey syndrome. **b**. Immunoblot analysis of U2OS cell lines treated with DOX to induce expression of the indicated GFP-FAM111A alleles. U2OS/GFP-FAM111A WT(low) cells express

the transgene at a lower level than U2OS/GFP-FAM111A WT cells used in Fig. 1 and 2. **c**. U2OS/GFP-FAM111A cell lines treated with DOX, pulse-labeled with EdU and stained with DAPI were analyzed for DAPI and EdU signal intensity using QIBC. **d**. Quantification of data in (c) for S phase (EdU-positive) cells (red bars, mean (A.U., arbitrary units); n>2000 cells per condition). **e**. As in (c), except that cells were pulse-labeled with EU. **f**. Quantification of EU incorporation in cells in (e) (red bars, mean; n>2000 cells per condition). **g**. DOX-treated U2OS/FAM111A cell lines were stained with RFC1 antibody, pre-extracted and fixed, and stained with DAPI. RFC1 and DAPI signal intensities were analyzed by QIBC. **h**. As in (g), except that cells were stained with RPB1 antibody. **i**. Immunoblot analysis of U2OS/GFP-FAM111A cell lines treated with DOX in the absence or presence of Z-VAD-FMK. **j**. As in (i), but using U2OS cell lines conditionally expressing ectopic untagged FAM111A alleles. **k**. *In vitro* auto-cleavage assay. Purified recombinant FLAG-FAM111A proteins were incubated at indicated temperatures for 4 h and analyzed by immunoblotting. **l**. GFP IPs from U2OS cell lines expressing GFP-FAM111A WT or D528G mutant were analyzed by mass spectrometry. Volcano plot shows enrichment of individual proteins (WT/D528G ratio) plotted against the P value. Dashed lines indicate the significance thresholds (FDR<0.05; s_0=1).

Data are representative of at least three (**b-k**) and two (**l**) independent experiments with similar outcomes.

Figure 4.

Disease-associated FAM111B mutants have amplified proteolytic activity and phenocopy FAM111A-induced cellular phenotypes

a. Domain organization of human FAM111A and overview of disease-associated heterozygous missense mutations. **b**. U2OS/GFP-FAM111A WT cells treated or not with DOX were subjected to

GFP IP followed by immunoblotting with indicated antibodies. **c**. As in (b), but using U2OS/GFP-FAM111B WT cells. **d**. Purified recombinant FLAG-FAM111A and FLAG-FAM111B proteins were incubated separately or mixed, subjected to IP with FAM111A antibody and immunoblotted with FLAG antibody. **e**. Immunoblot analysis of U2OS/GFP-FAM111B cell lines treated with DOX in the absence or presence of Z-VAD-FMK. **f**. Cells in (e) were pulse-labeled with EdU, stained with DAPI and analyzed for DAPI and EdU signal intensity using QIBC. **g**. Quantification of data in (f) for S phase (EdU-positive) cells (red bars, mean (A.U., arbitrary units); n>2000 cells per condition). **h**. As in (f), except that cells were pulse-labeled with EU. **i**. Quantification of EU incorporation in cells in (h) (red bars, mean; n>2000 cells per condition). **j**. U2OS/GFP-FAM111B cell lines were fixed at the indicated times after DOX treatment and stained with crystal violet. **k**. Cells in (e) were fixed, immunostained with γ-H2AX antibody and subjected to QIBC analysis of γ-H2AX signal intensity (red bars, mean; n>2000 cells per condition). **l**. Immunoblot analysis of purified recombinant FLAG-FAM111B proteins and their auto-cleavage products.

Data (**b-l**) are representative of at least three independent experiments with similar outcomes.