Mutations in DONSON disrupt replication fork stability and cause microcephalic dwarfism.

Citation for published version:

Digital Object Identifier (DOI):
10.1038/ng.3790

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Peer reviewed version

Published In:
Nature Genetics

Publisher Rights Statement:
This is the author's final peer-reviewed manuscript as accepted for publication

General rights
Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 15. Aug. 2019
Mutations in DONSON disrupt replication fork stability and cause microcephalic dwarfism.

John J Reynolds¹,³,⁶, Louise S Bicknell²,³,⁶, Paula Carrol³,⁶, Martin R Higgs¹,³,⁶, Ranad Shaheen⁴,³,⁶, Jennie E Murray², Dimitrios K Papadopulos⁵, Andrea Leitch², Olga Murina², Žygimanté Tarnauskaite², Sarah R Wesse⁶, Anastasia Zlatanou¹, Audrey Vernet¹, Alex von Kriegsheim², Rachel MA Mottram¹, Clare V Logan², Hannah Bye⁷, Yun Li⁸, Alexander Brean¹, Sateesh Maddirevula⁴, Rachel C Challis², Kassiani Skouloudaki⁵, Agaadir Almoisheer⁴, Hessa S Alsaif⁴, Ariella Amar⁷, Natalie J Prescott⁷, Michael B Bober⁹, Angela Duker⁹, Eissa Faqeih¹⁰, Mohammed Zain Seidahmed¹¹, Saeed Al Tala¹², Abdulrahman Alswaid¹³, Saleem Ahmed¹⁴,¹⁵, Jumana Yousuf Al-Aama¹⁴,¹⁵, Janine Altmüller¹⁶, Mohammed Al Balwi¹⁷, Angela F Brady¹⁸, Luciana Chessa¹⁹, Helen Cox²⁰, Rita Fischetto²¹, Raoul Heller²², Bertram D Henderson²³, Emma Hobson²⁴, Peter Nürnberg¹⁶, E Ferda Percin²⁵, Angela Peron²⁶,²⁷, Luigina Spaccini²⁶, Alan J Quigley²⁸, Seema Thakur²⁹, Carol A Wise³⁰, Grace Yoon³¹,³², Maha Alnemer³³, Pavel Tomancak³⁵, Gökhan Yigit⁸, A Malcolm R Taylor¹, Martin AM Reijns², Michael A Simpson⁷, David Cortez⁶, Fowzan S Alkuraya⁴,³⁵,Christopher G Mathew⁷,³⁴,³⁵, Andrew P Jackson²,³⁵,³ & Grant S Stewart¹,³⁵.
6Department of Biochemistry, Vanderbilt University School of Medicine, Nashville, Tennessee, USA.
7Department of Medical and Molecular Genetics, Faculty of Life Science and Medicine, King’s College London, 7th Floor Tower Wing, Guy’s Hospital, London, UK.
8Institute of Human Genetics, University Medical Center Göttingen, Göttingen, Germany.
9Nemours/Alfred I. duPont Hospital for Children, Wilmington, Delaware, USA.
10Department of Pediatric Subspecialties, Children’s Hospital, King Fahad Medical City, Riyadh, Saudi Arabia.
11Department of Pediatrics, Security Forces Hospital, Riyadh, Saudi Arabia.
13Department of Pediatrics, King Abdulaziz Medical City, Riyadh, Saudi Arabia.
14Department of Genetic Medicine, King Abdulaziz University Hospital, Jeddah, Saudi Arabia.
15Princess Al-Jawhara Al-Brahim Center of Excellence in Research of Hereditary Disorders, Jeddah, Saudi Arabia.
16Cologne Center for Genomics and Center for Molecular Medicine Cologne, University of Cologne, Cologne, Germany.
17Department of Pathology and Laboratory Medicine, King Abdulaziz Medical City, National Guard Health Affairs, Riyadh, Saudi Arabia.
18North West Thames Regional Genetics Service, London North West Healthcare NHS Trust, Harrow, UK.
19Department of Clinical and Molecular Medicine, University La Sapienza, Roma, Italy.
20West Midlands Regional Clinical Genetics Service, Birmingham Women's Hospital, West Midlands, UK.
21Pediatric Hospital Giovanni XXIII, Bari, Italy.
22Institute of Human Genetics, University Hospital of Cologne, Cologne, Germany.
23 Division of Clinical Genetics, Faculty of Health Sciences, UFS, Bloemfontein, South Africa.

24 Department of Genetics, Yorkshire Regional Genetic service, Chapel Allerton Hospital, Leeds, UK.

25 Department of Medical Genetics, Gazi University Faculty of Medicine, Besevler Ankara, Turkey.

26 Clinical Genetics Unit, Division of Maternal Fetal Medicine, Department of Obstetrics and Gynecology, V. Buzzi Children's Hospital, Universita' degli Studi di Milano, Milan, Italy.

27 Child Neuropsychiatry Unit – Epilepsy Center, San Paolo University Hospital, Department of Health Sciences, Università degli Studi di Milano, Milan, Italy.

28 Dept of Radiology, Royal Hospital for Sick Children, Edinburgh, UK.

29 Dept of Genetic and Fetal Medicine, Fortis lafemme, New Delhi, India.

30 Sarah M. and Charles E. Seay Center for Musculoskeletal Research, Texas Scottish Rite Hospital for Children, Dallas, Texas, USA.

31 Division of Clinical and Metabolic Genetics, The Hospital for Sick Children, University of Toronto, Toronto, Canada.

32 Department of Pediatrics, Division of Neurology, The Hospital for Sick Children, University of Toronto, Toronto, Canada.

33 Department of Obstetrics and Gynecology, King Faisal Specialist Hospital and Research Center, Riyadh, Saudi Arabia.

34 Sydney Brenner Institute for Molecular Bioscience, Faculty of Health Sciences, University of the Witwatersrand, South Africa.

35 These authors jointly directed this work.

36 These authors contributed equally to this work.
*Correspondence to:

GSS, g.s.stewart@bham.ac.uk; APJ, Andrew.jackson@igmm.ed.ac.uk;

CGM, christopher.mathew@kcl.ac.uk; FSA, falkuraya@kfshrc.edu.sa
Abstract

To ensure efficient genome duplication, cells have evolved numerous factors that promote unperturbed DNA replication, and protect, repair and restart damaged forks.

Here we identify DONSON as a novel fork protection factor, and report biallelic DONSON mutations in 29 individuals with microcephalic dwarfism. We demonstrate that DONSON is a replisome component that stabilises forks during genome replication. Loss of DONSON leads to severe replication-associated DNA damage arising from nucleolytic cleavage of stalled replication forks. Furthermore, ATR-dependent signalling in response to replication stress is impaired in DONSON-deficient cells, resulting in decreased checkpoint activity, and potentiating chromosomal instability. Hypomorphic mutations substantially reduce DONSON protein levels and impair fork stability in patient cells, consistent with defective DNA replication underlying the disease phenotype.

In summary, we identify mutations in DONSON as a common cause of microcephalic dwarfism, and establish DONSON as a critical replication fork protein required for mammalian DNA replication and genome stability.
Microcephalic primordial dwarfism (MPD) is the collective term for a group of human disorders characterised by intra-uterine and postnatal growth delay alongside marked microcephaly\(^1\), and includes disorders such as MOPD II, ATR/ATRIP-Seckel syndrome and Meier-Gorlin syndrome. Mutations in genes encoding either components of the DNA replication machinery (replisome) or genome stability proteins are a frequent cause of microcephalic dwarfism\(^2-14\).

During the course of normal DNA replication, a subset of replication forks may stall, causing 'replication stress'\(^15\). This stalling can be caused by endogenous or exogenous sources, such as collision of the replisome with DNA lesions or the transcriptional machinery, or replication of difficult to replicate genomic regions. To facilitate efficient genome duplication, stalled replication forks must be stabilised and protected from collapse. Multiple factors safeguard replication fork stability, many of which function within the ATR-CHK1-dependent replication stress response\(^16-18\). This pathway ensures that fork stabilisation is tightly coordinated with a global reduction in DNA synthesis, allowing stalled or damaged forks to be repaired and restarted\(^19,20\).

Exome sequencing analysis of microcephalic dwarfism patients has identified several novel factors that regulate replication and/or the replication stress response. Using this strategy, we recently identified mutations in \textit{TRAIP} in individuals with MPD\(^5\), and demonstrated that \textit{TRAIP} is required for the response to replication-blocking DNA lesions. To identify similar disease-associated genes, we carried out whole exome sequencing of genetically uncharacterised patients with microcephaly. Here, we report the identification of \textit{DONSON} as a new microcephalic dwarfism gene, and demonstrate that \textit{DONSON} is a novel replisome component that maintains genome stability by protecting stalled/damaged replication forks.
Results

**DONSON mutations identified in microcephalic dwarfism patients**

Whole exome sequencing (WES) was undertaken on 26 patients with microcephaly and reduced stature. After aligning WES reads to the reference genome, variant calling, and filtering for rare variants (MAF <0.005), analysis under a recessive model of inheritance identified rare biallelic variants in the *DONSON* (‘Downstream neighbour of SON’) gene in nine affected individuals from seven families. Sanger capillary sequencing confirmed the presence of these mutations in these patients (P1-1 to P3, P6, and P9 to P11, Table 1). Subsequent re-sequencing of an additional 230 patients with primary microcephaly, microcephaly with reduced stature, or MPD, identified five additional families with compound heterozygous mutations in *DONSON* (P4, P5, P7, P8, P12; Table 1). All variants segregated amongst family members in a manner consistent with an autosomal recessive trait, and were present at a frequency of <0.5% in the ExAC database21.

Two other concurrent molecular genetic studies provided further independent evidence to support the identification of *DONSON* as a novel human disease gene. Firstly, exome sequencing was carried out on a consanguineous Palestinian family previously reported to have a Fanconi Anaemia-like disorder22. These patients presented with microcephaly, short stature, slow growth and forearm and thumb dysplasia, although no individuals had haematological evidence of bone marrow failure. This WES analysis revealed a deleterious homozygous transition, c.1337T>C, resulting in substitution of a highly conserved residue (p.M446T) in all three affected individuals (P13-1, P13-2, P13-3; Table 1, Supplementary Fig. 1). Secondly, a study of five consanguineous families in Saudi Arabia with extreme microcephaly and short stature allowed a 1.6 Mb haplotype shared by all five families (combined multipoint LOD score Z= 8.0) to be mapped to a defined critical interval on chromosome 21 that contained *DONSON*. Whole exome and
genome sequencing identified a single rare variant at this locus in DONSON, c.786-22A>G. Capillary sequencing confirmed this intronic variant to be homozygous in all seven affected individuals from this study (P14 to P18-3; Table 1), identical to that detected in two Saudi Arabian individuals present within the first study described above (P11, P12).

Subsequently, a further five individuals from three different families with DONSON mutations were identified in additional MPD patients recruited to two of the genetic studies described above (P19 to P21-2; Table 1).

DONSON mutations give rise to severe microcephaly with short stature

Despite their identification in separate studies, all patients with DONSON mutations had similar clinical phenotypes. Marked microcephaly was present (OFC -7.5 +/- 2.4 SD), with a substantial reduction in cerebral cortical size, along with decreased gyral folding evident on neuroimaging (Fig. 1a and Supplementary Fig. 2), similar to that previously reported for other primary microcephaly and microcephalic dwarfism patients^{5,23-25}. Height was reduced (-3.2 +/- 1.4 SD), although much less so than head circumference (Fig. 1a), and to a lesser degree than observed in other microcephalic dwarfism-associated disorders (where height was typically ≤ -4 SD)^{2,3,5,8-10,24,26}. Minor skeletal abnormalities were present in several patients, including fifth finger clinodactyly, syndactyly, brachydactyly, hypoplasia of carpal/metacarpal/phalangeal bones, or radial head dislocation (Supplementary Table 1). Absent/hypoplastic patellae were present in patients P12, P20-1 and P20-2. Notably, patient P19 had bilateral hypoplasia of the radius and thumb, which, together with the limb abnormalities displayed by P13-1 and 13-2, established radial ray defects as an uncommon but recurrent phenotype. In family P21, the most extreme phenotype was observed, with substantial limb shortening/reduction in association with foetal lethality (Supplementary Fig. 3). Aside from microcephaly, neither a recognisable facial phenotype (Fig. 1b) nor recurrent malformations affecting other organ systems were evident. Intellectual disability, if present, was typically mild.
In conclusion, the number of biallelic variants identified, combined with a common clinical presentation, provided strong evidence for DONSON being a novel human disease gene, associated with autosomal recessive inheritance. We therefore investigated the consequence of these mutations on DONSON protein function.

**DONSON mutations markedly reduce protein levels**

DONSON mutations were identified in 29 individuals, and comprised a range of mutation classes (Fig. 1c). Notably, no biallelic nonsense or frameshift mutations were observed, indicating that mutations likely resulted in partial loss of DONSON function. To investigate this, we established patient-derived primary fibroblast or lymphoblastoid cell lines representing a range of mutations. Immunoblotting demonstrated marked decreases in DONSON protein levels in all cell lines tested (Fig. 2a, b), establishing that the analysed mutations affected DONSON protein expression.

Mutations in multiple families were associated with two different ancestral haplotypes (P1-1 to P7; P11, P12 and P14 to 18-3 respectively; Table 1) and were investigated in more detail. Firstly, as described above, nine individuals (P11, P12 and P14 to P18-3) were homozygous for the c.786-22A>G mutation, predicted to enhance a cryptic splice donor site within intron 4 (MaxEntScan)\textsuperscript{27}. Consistent with a common ancestral founder, five consanguineous families of Saudi Arabian origin bearing this mutation (P14 to P18-3) shared a 1.6 Mb haplotype region of chromosome 21 (Supplementary Fig. 4). qRT-PCR analysis of RNA isolated from four patients with this variant demonstrated a significant decrease in full-length transcript, and increased skipping of exon 5 (Supplementary Fig. 5). This resulted in an out-of-frame mRNA predicted to undergo NMD, explaining the substantial reduction in protein levels seen in fibroblasts homozygous for this mutation (P12; Fig 2b).
Secondly, two missense variants in DONSON, p.S28R and p.K489T, and an intronic variant (c.786-33A>G) were present in seven individuals of European ancestry and one Somali (P1-1 to P7; Table 1). These were associated with a different ancestral haplotype, comprising a shared 127.7 kb genomic region (Supplementary Table 2). No other deleterious biallelic variants were present in the four other genes within this region. Despite the close proximity of the c.786-33A>G intronic variant to the Saudi Arabian c.786-22A>G mutation, the former change did not disrupt splicing between exons 4 and 5, as assessed by either mini gene splicing assays (Supplementary Fig. 6), or RT-PCR analyses of patient cell lines (data not shown), indicating this variant is unlikely to be pathogenic. Since DONSON protein levels were severely reduced in cells from patients inheriting this haplotype allele in combination with a truncating mutation (P2, P6), this suggested that either one or both of the missense variants associated with this haplotype (p.S28R, p.K489T) compromised protein levels (Fig. 2b).

To investigate these two variants, we established isogenic HeLa cell lines expressing doxycycline-inducible, siRNA-resistant wild type (WT) or mutant (p.S28R, p.K489T) GFP-tagged DONSON. Following siRNA depletion of endogenous DONSON (Supplementary Fig. 7), and induction of exogenous GFP-DONSON, immunoblotting revealed that the p.K489T mutation, but not p.S28R, reduced protein levels in a post-transcriptional manner (Fig. 2c and Supplementary Fig. 8). This suggested that the p.K489T substitution within the second haplotype causes the decreased DONSON protein levels observed in P2 and P6 (Fig. 2b). The K489T variant is present as a rare allele in the population, observed at a frequency of 0.00099 in the ExAC database\(^1\). In the patients reported here, it is always observed \textit{in trans} with a frameshift or other protein-disrupting allele, suggesting that it is a functionally weak variant insufficient to cause disease alone, a conclusion supported by the presence of a single homozygous individual in ExAC.
Eight other point mutations were identified in patients: six missense substitutions, a two amino-acid deletion and an amino-acid insertion (Table 1 and Fig. 1c). Notably these were at highly conserved residues (Supplementary Fig. 1) and predicted to be deleterious (Alamut Visual). In agreement, exogenous expression of the p.M446T mutant resulted in significantly reduced protein levels (Fig. 2c), similar to cells derived from patients P13-1, P13-2 and P13-3. Furthermore, five of these mutations disrupted the subcellular localisation of GFP-DONSON (Supplementary Fig. 9), suggesting that these alterations also compromise DONSON protein function via protein mis-localisation.

Finally, an intronic mutation, c.1047-9A>G, was present in three individuals (P9, P21-1 and P21-2). qRT-PCR analysis of RNA isolated from the two patients homozygous for this variant (P21-1, P21-2) revealed a substantial reduction in DONSON transcript levels compared to normal controls (Supplementary Fig. 10). This variant was also observed in trans with the missense mutation F292L in patient P9. Since cells derived from this individual exhibited severely reduced levels of DONSON protein (Fig. 2b), it is likely that this intronic change also perturbs DONSON protein expression.

Taken together, the deleterious consequences of the identified mutations on splicing, transcript abundance, subcellular localisation and/or protein levels strengthened our conclusion that DONSON was a novel human disease gene. The fact that knockout of murine Donson leads to developmental lethality, together with the presence of residual DONSON protein in patient-derived cell lines (Fig. 2a, b and Supplementary Fig. 11), supports the notion that the identified mutations are hypomorphic, retaining some residual function.

DONSON stabilises replication forks during normal DNA replication

While DONSON is highly conserved in metazoa and plants, its precise function(s) remained to be defined. Humpty-dumpty (hd), the Drosophila ortholog of DONSON, has
been proposed to play a role in cell proliferation: Hd expression peaks during S-phase; *hd* mutants have an ‘egg shell’ phenotype; and clonal inactivation of *hd* impairs genome replication in larval tissues.\(^{29}\)

In light of this, we investigated whether human DONSON might play a similar role. After synchronising cells with a double thymidine block, we observed that human DONSON, like Hd, was also maximally expressed during S-phase, mirroring Cyclin A expression (Fig. 3a). Furthermore, depletion of DONSON resulted in a significant increase in BrdU-positive cells observed by FACS, consistent with a role in promoting efficient S-phase progression (Fig. 3b). Given these data, we next used DNA fibre analysis to assess whether DONSON depletion led to decreased DNA replication fork progression. Although fork progression rates did not decrease in cells lacking DONSON (Fig. 3c), compromising DONSON expression increased spontaneous replication fork stalling, with a concomitant decrease in the number of ongoing forks (Fig. 3d). Moreover, we also observed increased replication fork asymmetry in cells depleted of DONSON, indicating replication fork instability (Fig. 3e). Together, this suggests that the increase in BrdU-positive DONSON-depleted cells may reflect prolonged S-phase due to stalled replication forks, rather than a global reduction in DNA synthesis.

Since DNA replication is closely linked with genome stability\(^{15,30,31}\), we reasoned that loss of DONSON would lead to a failure to complete timely replication and increased S-phase DNA damage. To test this hypothesis, we combined immunofluorescence of γ-H2AX and 53BP1 (markers of DNA damage and DNA double strand breaks respectively) with EdU labelling to identify S-phase cells. We observed that a significant proportion of DONSON-depleted cells exhibited spontaneous γ-H2AX and 53BP1 foci (Fig. 3f), of which the majority occurred in S-phase cells (Supplementary Fig. 12a-d), consistent with the identification of DONSON as a potential genome stability regulator by high-throughput siRNA screening.\(^{32}\) Using pulsed-field gel electrophoresis, we also observed elevated
levels of DNA double strand breaks in the absence of DONSON (Supplementary Fig. 12e). Taken together, these data support a role for DONSON in maintaining replication fork stability during unperturbed DNA replication, and demonstrate that spontaneous DNA damage arises in replicating cells in the absence of DONSON.

DONSON is a component of the replisome

To shed further light on the role of DONSON in regulating replication fork stability, we carried out mass spectrometry screening to identify interaction partners of GFP-tagged DONSON. Amongst the interactors, we detected multiple replication proteins including subunits of the MCM helicase and the GINS complex (Fig. 4a and Supplementary Table 3). To confirm these findings, we carried out pull-down analyses coupled with immunoblotting to identify GFP-DONSON binding proteins. Consistent with our mass spectrometry data, we detected interactions between GFP-DONSON and the replisome components MCM2, MCM7, Treslin and PCNA (Fig. 4b), suggesting that DONSON associated with the replisome.

We next used three complementary techniques to assess whether DONSON localised to sites of DNA replication. Firstly, we carried out proximity ligation assays (PLA) of GFP-DONSON with the replication proteins PCNA and RPA. In line with DONSON being closely associated with the replication machinery, we observed robust PLA signals between GFP-DONSON and both PCNA and RPA (Fig. 4c, d). We next performed Fluorescence Cross-Correlation Spectroscopy (FCCS) in live HeLa cells stably co-expressing RFP-PCNA and GFP-DONSON, to measure the degree of co-diffusion of these molecules. Significantly increased co-diffusion of PCNA and DONSON was observed in S-phase PCNA foci, but not in nuclei of non-replicating cells (Fig. 4e-f and Supplementary Fig. 13), indicating that these proteins interacted during DNA replication. Finally, we utilised iPOND (isolation of proteins on nascent DNA) combined with mass
spectrometry to ascertain whether DONSON is present on newly replicated DNA. Crucially, this approach demonstrated that DONSON, like MCMs and RPA, was significantly enriched at replication forks compared to mature chromatin (Fig. 4g).

Collectively, these data strongly support the conclusion that DONSON is a novel replisome component that plays a role in promoting fork stability.

DONSON depletion impairs cell-cycle checkpoint activation

Since our data suggested that DONSON functions to protect replication forks during unperturbed DNA replication, we extended our findings to evaluate the role of DONSON in preventing replication fork stalling following exogenous replication stress. Exposure to the replication stress-inducing agents hydroxyurea (HU) and mitomycin C (MMC) induced significantly more fork stalling in DONSON-depleted cells than in control cells (Fig. 5a-b). DONSON depletion also resulted in a failure to suppress new origin firing upon exogenous replication stress (Fig. 5c). Since suppression of new origin firing reflects checkpoint activity, this suggests that DONSON is required for efficient activation of the intra-S phase checkpoint. To further investigate this, we measured activation of this checkpoint after inhibition of ATR (VE821; ATRi), the apical kinase which governs the replication stress response^{18}. Upon HU exposure and ATR inhibition, we observed no difference in the number of new origins fired between control or DONSON-depleted cells, indicating that DONSON and ATR may function within the same pathway to activate the intra-S phase checkpoint (Fig. 5d).

We next examined whether the ATR-dependent replication stress response was functional in the absence of DONSON. We first monitored ATR pathway activation in DONSON-depleted cells treated with HU or MMC by immunoblotting, using phospho-specific antibodies to known ATR substrates. This analysis revealed that cells lacking DONSON failed to efficiently phosphorylate a number of ATR substrates, such as CHK1.
and NBS1, in response to HU or MMC (Fig. 5e and Supplementary Fig. 14a). Moreover, ATR autophosphorylation on T1989, another marker of ATR activation\textsuperscript{36}, was reduced (Supplementary Fig. 14b). Loss of DONSON also significantly increased mitotic indices following exposure to HU or MMC as measured by FACS, demonstrating that DONSON-depleted cells fail to efficiently activate the G2/M checkpoint in response to replication stress (Fig. 5f and Supplementary Fig. 14c). We next determined whether the reduced ATR signalling observed was due to decreased levels of RPA-coated ssDNA, which is the stimulus for ATR activation. Surprisingly, DONSON-depleted cells exhibited elevated levels of RPA-coated ssDNA following HU treatment (Supplementary Fig. 15), consistent with defective activation of the ATR-dependent replication stress response.

Dysregulated DNA replication combined with impaired intra-S phase checkpoint signalling, such as in ATR-deficient cells\textsuperscript{37-41}, gives rise to extensive chromosome breakage and genome instability. Consistent with this, we observed significantly elevated levels of spontaneous micronuclei and chromatid gaps/breaks in cells lacking DONSON (Fig. 6a and Supplementary Fig. 16a), which was exacerbated by exposure to HU or MMC (Supplementary Fig. 16a-c). We also observed spontaneously-arising highly-fragmented or completely pulverised metaphase chromosomes in cells lacking DONSON, which increased upon exogenous replication stress (Fig. 6b and Supplementary Fig. 16d).

Together, these data confirm that upon exogenous replication stress, DONSON is required to stabilise stalled replication forks, efficiently activate the intra-S and G/2M cell-cycle checkpoints, and maintain genome stability.

**Cleavage of stalled replication forks in DONSON-deficient cells**

It has been proposed that the spontaneous DNA damage arising in ATR-deficient cells is due to processing of stalled/damaged forks by SLX4-associated structure-specific
nucleases, such as MUS81, SLX1 and XPF\textsuperscript{42-46}. We therefore postulated that the replication abnormalities and chromosomal aberrations of DONSON-deficient cells might arise via similar mechanisms. Indeed, the spontaneous replication fork asymmetry and H2AX phosphorylation exhibited by DONSON-depleted cells were partially reduced by co-depletion of either MUS81 or XPF (Fig. 6c, d). Moreover, co-depletion of MUS81 or XPF also reduced chromosome breakage and genomic pulverisation in these cells (Fig. 6e-g). From this, we conclude that the severe genome instability apparent in the absence of DONSON is due to nucleolytic processing of damaged replication forks by structure-specific nucleases.

Replication stress-induced genomic damage in DONSON patient cells

To link the role of DONSON in regulating replication fork stability and the phenotype of patients with DONSON mutations, we characterised replication dynamics and genomic stability of patient-derived fibroblasts. All DONSON patient-derived cell lines examined (P2, P6, P9, P10-2, P12) showed higher levels of spontaneous fork asymmetry and fork stalling than cells from unaffected individuals (Supplementary Figs. 17a, 18). Furthermore, patient-derived cells also exhibited elevated fork asymmetry and fork stalling following HU exposure, combined with defective intra-S phase checkpoint activation (Supplementary Fig. 17a, 18). Finally, levels of S-phase DNA damage and chromosome breakage were also elevated in these cell lines (Supplementary Fig. 17b-c). Together, these observations provide a potential pathological explanation for the clinical phenotype.

In addition, using isogenic cell lines inducibly expressing GFP-DONSON (Fig. 2c), we observed that expression of the haplotype-associated S28R mutant, but not the K489T variant, complemented loss of endogenous DONSON by rescuing the spontaneous fork stalling observed upon DONSON depletion (Supplementary Fig. 19). This is consistent with K489T being the pathogenic variant within the haplotype region (Fig. 2c).
Finally, we set out to demonstrate that the patient-associated cellular phenotypes were directly due to mutation of DONSON. We first established three paired isogenic cell lines via transduction of patient-derived fibroblasts with retroviral expression vectors encoding WT DONSON or an empty vector (Fig. 7a). Importantly, the spontaneous DNA damage, replication fork stalling, replication fork asymmetry and intra-S phase checkpoint defect were all corrected by expression of WT DONSON (Fig. 7b-d and Supplementary Fig. 20), confirming that these phenotypes were directly due to DONSON deficiency. Lastly, using one of these cell lines, we also observed that inhibition of ATR and mutation of DONSON are epistatic with regard to the observed replication abnormalities (Supplementary Fig. 21).
Here we identify DONSON as a novel human disease gene, and describe 29 individuals with a range of mutations in DONSON, establishing such alterations as a frequent cause of microcephalic dwarfism. Since normal embryonic development requires rapid cellular proliferation\textsuperscript{47,48} it is exquisitely sensitive to genetic perturbations that impact DNA replication\textsuperscript{1-3,6}. A failure to complete timely genome duplication will profoundly affect the number of cells generated during embryonic development. For example, hypomorphic mutations in ATR result in severe microcephaly and growth retardation, both in humans and in a murine model\textsuperscript{2,3,49,50}, due to the role that ATR plays in preventing replication stress during development\textsuperscript{49-50}. We propose that mutation of DONSON similarly reduces the number of cells generated during development via a failure to maintain replication fork stability in the presence of endogenous replication stress, thus explaining the decreased organism size observed. Furthermore, since brain development requires rapid proliferation of neural progenitor cells within a limited timeframe\textsuperscript{47}, it is particularly sensitive to disruptive genetic perturbations. This may explain why brain development is disproportionately affected in these individuals compared to growth.

DONSON has no predicted domain structure or paralogs, and previous characterisation has been limited to two previous studies: an siRNA screen proposing that DONSON regulates genome stability, and a study of its Drosophila ortholog Humpty-dumpty suggesting a role in cell proliferation\textsuperscript{29,32}. Consistent with this, we establish that DONSON is a replisome component that ensures replication fork stability, and promotes efficient activation of both intra-S and G2/M cell-cycle checkpoints upon exogenous replication stress. Loss of DONSON leads to increased spontaneous stalling of replication forks, which are subsequently cleaved into replication-associated DNA double strand breaks by structure-specific nucleases. Defective cell-cycle checkpoint activation in DONSON-deficient cells then allows transmission of these breaks into mitosis, accounting
for the elevated chromosomal damage and genome fragmentation observed (Supplementary Fig. 22). Additional studies will be important to confirm this model, and to investigate whether DONSON is a constitutive component of the replisome, or whether it is recruited to a subset of replication forks. Furthermore, establishing which replisome components DONSON directly interacts with, and the functional importance of these associations, will also inform understanding of its biological function.

The mechanism by which DONSON ensures replication fork stability and promotes checkpoint activation remains to be defined. We propose that in addition to being a replisome component, DONSON is also involved in promoting the ATR-CHK1 replication stress response, since we observed that DONSON-depleted cells exhibit defective activation of cell cycle checkpoints and reduced ATR-dependent signalling in response to exogenous replication stress. This hypothesis is supported by the observation that impaired replication alone, such as that arising from a hypomorphic mutation in MCM4 (MCM4\textsuperscript{Chaos3/Chaos3}), does not give rise to decreased CHK1 phosphorylation or increased new origin firing upon replication stress\textsuperscript{51}. However, it is unclear whether DONSON functions directly or indirectly to regulate the ATR-CHK1 pathway. Our demonstration that cells lacking DONSON do not exhibit a global reduction in replication, or decreased levels of RPA-coated ssDNA, indicates that loss of DONSON does not affect the cells ability to generate the primary stimulus for ATR pathway activation. Based on this, we propose that either DONSON directly activates ATR, in a manner similar to TOPBP1\textsuperscript{52} or ETAA1\textsuperscript{53,54}, or functions indirectly to regulate other factors necessary for efficient ATR-CHK1 signalling, such as the MRE11/RAD50/NBS1 (MRN)\textsuperscript{55} complex or TIPIN/Timeless\textsuperscript{19,20}. Since DONSON does not contain a canonical ATR activation domain, which is found in both TOPBP1 and ETAA1, we favour the latter possibility. However, how DONSON acts to promote ATR signalling is not yet clear, and future work will be critical in establishing whether this is direct or indirect.
It is also evident that the cellular phenotype of cells lacking DONSON cannot be explained solely by abnormal DNA replication or defective ATR-dependent signalling. In particular, exposure of cells lacking ATR to exogenous replication stress results in highly elevated levels of H2AX phosphorylation, a situation not observed upon DONSON loss (Fig. 5e), despite the presence of substantial amounts of DNA damage. Therefore, whilst our observations are consistent with a role for DONSON in promoting ATR-CHK1 signalling, DONSON may also impact on other pathways that promote H2AX phosphorylation at the replication fork, for example those governed by ATM or the MRN complex.

In summary, we have identified DONSON as a novel disease gene that plays a key role in regulating cellular replication and cell cycle checkpoints. Further studies examining how DONSON functions will provide fundamental insight into how cells maintain replication fork integrity, and how these pathways prevent human disease.
Acknowledgements

We would like to thank the families and clinicians for their involvement and participation.

We are grateful to R. S. Taylor (University of Manchester) and D.-J. Kleinjan (University of Edinburgh), J. Bartek and C. Bartek (Copenhagen) for their kind gift of reagents. We thank E. Freyer, J. Wills, J. Ding, A. Fluteau, C. Keith, D. Longman and the IGMM FACS, core sequencing and mass-spectrometry facilities for technical assistance and advice. The Walking With Giants Foundation and Potentials Foundation supported the Primordial Dwarfism Registry (M.B.B.). This work is supported by funding from Cancer Research UK (C17183/A13030) (G.S.S., M.R.H. and A.V.), the Medical Research Council (MR/M009882/1) (J.J.R.), Worldwide Cancer Research (13-1012) (A.Z.), Birmingham Children’s Hospital Research Foundation (BCHRF400) (R.M.A.M.), University of Birmingham (J.J.R., R.M.A.M., A.B.), Newlife Foundation for Disabled Children (P.C., L.S.B.), Medical Research Scotland (L.S.B.) and the National Institute for Health Research (NIHR) Biomedical Research Centre at Guy’s and St Thomas’ NHS Foundation Trust and King's College London (H.B., A. Amar, N.J.P, M.A.S., C.G.M.), German Federal Ministry of Education and Research (BMBF) (1GM1404; E-RARE network EuroMicro) (G.Yigit), KSCDR funding and KACST grant 09-MED941-20 (F.S.A.), Swiss National Science Foundation (P2ZHP3_158709) (O.M.), Medical Research Council, the Lister Institute for Preventative Medicine and the European Research Council (ERC, 281847) (A.P.J.).

Author Contributions


**Competing Financial Interests Statement**

The authors declare no competing financial interests.
References


Figure Legends:

**Figure 1: DONSON mutations cause severe microcephaly and short stature.**

(a) DONSON mutations result in severe prenatal-onset microcephaly, often associated with short stature. Length at birth (Lgt), current height (Hgt) and head circumference (OFC) plotted as z-scores (SD from population mean for age and sex). Black horizontal bars indicate mean values. Dashed line at -2 SD indicates cut-off for normal population distribution. Patients from the three independently identified DONSON patient cohorts are denoted by black (P1-P12 and P20), orange (P13), and blue (P14-19 and P21) circles. (b) Photographs of affected individuals with DONSON mutations demonstrating facial similarities. Written consent to publish photographs was obtained from the affected families. P, patient. (c) Schematics of the DONSON gene and protein indicating position of the identified mutations. DONSON mutations comprised a range of mutation classes (nonsense, frameshift, essential splice site, missense and intronic). The genomic structure is based on the longest ORF containing ten coding exons (white rectangles) (NM_017613.3). The positions of identified mutations affecting splicing are shown on the gene structure (top) and missense and truncating variants on the encoded protein (bottom).

**Figure 2: Mutations in DONSON affect protein levels.**

(a-b) DONSON mutations result in severely reduced levels of DONSON protein. Immunoblotting of cell extracts from lymphoblastoid (a) and fibroblast (b) cell lines derived from patients with mutations in DONSON. ATR was used as a loading control. The two blots from (a) originate from two independent gels. (c) The K489T, but not the S28R variant, associated with the DONSON haplotype affects protein levels. Cells were treated with doxycycline 48 h post siRNA transfection, and harvested for Western blot analysis 24 h later (n=2). Exogenous DONSON were detected using an anti-GFP antibody
respectively. TOPBP1 was used as a loading control. Depletion of endogenous DONSON in these cells was confirmed by immunoblotting (Supplementary Fig. 2).

**Figure 3. Loss of DONSON results in spontaneous replication fork stalling and increased genome instability.**

(a) DONSON protein levels are increased during S-phase. HeLa cells were synchronised in S-phase using a double thymidine block, released, harvested at the indicated time points, and immunoblotting was performed (n=2). Cyclin A and phospho-histone H3 Ser-10 are markers of S/G2 and M phase respectively. Vinculin represents a loading control.

(b) S-phase is prolonged upon DONSON depletion. HeLa cells transfected with the indicated siRNAs were pulsed with BrdU, fixed and analysed by FACS (n=4; error bars indicate SD). (c-e) Replication fork analysis of HeLa cells transfected with control or DONSON siRNA and pulsed with CldU and IdU. (c) Top: Schematic of DNA fibre analysis. Bottom: loss of DONSON does not decrease replication fork velocity. Replication fork speed (kb/min) was determined (n=5). (d) DONSON depletion results in spontaneous fork stalling. Percentages of ongoing replication forks, new origins and stalled replication forks in cells from (c) were quantified (n=3). (e) DONSON depletion leads to replication fork asymmetry. Top: example images; magenta arrows indicate origins of replication; white arrow denotes fork asymmetry. Bottom: plot indicates the ratio of left/right fork track lengths of bidirectional replication forks in cells from (c). Red lines denote median ratios (n=3). (f) Loss of DONSON increases spontaneous γH2AX/53BP1 foci formation. HeLa cells transfected with the indicated siRNAs were immunostained with antibodies to 53BP1 and γH2AX (left panel), and the percentage of cells with >10 53BP1 and γH2AX foci were quantified using fluorescence microscopy (right panel; n=5; >300 cells per sample per independent experiment). Scale bar; 10 µm.
**Figure 4. DONSON localizes to the replication fork.**

(a-d) DONSON interacts with multiple components of the replication machinery. (a) GFP or GFP-DONSON was precipitated by GFP-Trap, from asynchronous cells or cells accumulated in S-phase with 2 mM HU treatment for 24 h. Heatmap denotes significant interactions identified by mass spectrometry (n=3). Inset: Schematic of the mammalian replisome with selected replication factors. (b) 293FT cells were transfected with the indicated expression vectors in the presence/absence of HU. GFP or GFP-DONSON were isolated by GFP-Trap and co-precipitating proteins visualised by immunoblotting (n=2). Benzonase Nuclease was included to exclude DNA-mediated interactions. The bottom two panels are scanned images of Ponceau S-stained nitrocellulose membrane. (c-d) DONSON localises in close proximity to replication forks. (c-d) PLA was carried out on cells from (a) using the indicated antibodies in the presence/absence of HU (n=2). (c) Quantification of PLA signals. (d) Representative PLA images. (e-f) DONSON interacts with PCNA at replication foci in live cells. (e) Representative confocal images of live cells expressing GFP-DONSON and RFP-PCNA. Boxes indicate representative regions used for FCCS analysis. (f) FCCS measurements of GFP-DONSON and RFP-PCNA reveal significant cross-correlation at replication foci at similar concentrations. Average cross-correlation curves are shown from cells expressing GFP-DONSON in replication foci (red) or non-replicating (grey) cells, or GFP-expressing S-phase nuclei (purple). Inset: Mean cross-correlation amplitude values from multiple cells (error bars indicate SD; n=4, 3 and 5). Increased G(τ) values indicate higher degree of cross-correlation between GFP-DONSON and RFP-PCNA in replication foci. See also Supplementary Fig. 11. (g) iPOND was performed on 293T (n=3), HeLa (n=2) and HCT116 (n=2) cells, and EdU-coprecipitates analysed by mass spectrometry. Data represents the combination of all seven experiments. Log2 abundance denotes the ratio of proteins at nascent DNA
compared to mature chromatin. Values >0 represent proteins enriched at forks, whilst values ≤0 denote chromatin-bound factors. Scale bars; 10 µm.

**Figure 5. Depletion of DONSON compromises activation of cell cycle checkpoints.**

**(a-c)** Loss of DONSON results in replication fork instability that is exacerbated by replication stress. *(a)* HeLa cells transfected with either control or DONSON siRNA were pulsed with CldU, exposed to 2 mM HU for 2 h, and then pulsed with IdU. Alternatively, cells were exposed to 50 ng/ml MMC for 24 h, and pulsed with sequential pulses of CldU and IdU (see schematic). DNA fibres were quantified, and the percentage of *(b)* stalled forks and *(c)* new origins are displayed (in all cases n=3). *(d)* Loss of DONSON is epistatic with ATR inhibition. Replication fork analysis of HeLa cells transfected with either control or DONSON siRNA. Cells were pulsed with CldU, exposed to 2 mM HU +/- 5μM ATR inhibitor for 2 h, and then pulsed with IdU (n=3). New origins (2nd label origin) were counted as an indicator of intra-S phase checkpoint activation. *(e)* Cells lacking DONSON exhibit defective or delayed ATR activation in response to replication stress. Whole cell extracts of HeLa cells transfected with either control or DONSON siRNA were subjected to immunoblot analysis using the indicated antibodies following treatment with 1 mM HU (n=2). *(f)* The percentage of mitotic cells following exposure to 1 mM HU for 24 h (from *(e)*) was determined by flow cytometry, using antibodies to phosphorylated histone H3-Ser10 (a marker of mitosis) (n=5).

**Figure 6. Increased spontaneous chromosome breakage and fragmentation of mitotic chromosomes in DONSON-depleted cells.**

**(a,b)** Metaphases chromosomes from DONSON or control siRNA transfected HeLa cells were visualised by Giemsa staining and light microscopy. *(a)* Quantification of average numbers of chromatid gaps/breaks per metaphase (n=6; >50 metaphases per sample per
Representative images of normal chromosomes, chromosomes containing gaps/breaks, highly fragmented and pulverized chromosomes. Red arrows denote chromatid gaps/breaks; blue arrows indicate chromosomal exchanges. Scale bar; 10 µm. (c-g) Loss of the structure-specific nucleases MUS81 or XPF significantly reduces the spontaneous replication fork asymmetry and genome instability in DONSON-depleted cells. (c) Cells transfected with the indicated siRNAs were pulsed with CldU and IdU. Replication fork asymmetry was measured as in (Fig. 3e). The red lines denotes median ratios (n=3). (d) Co-depletion of MUS81 or XPF with DONSON reduces levels of spontaneous DNA damage. Extracts from cells transfected with the indicated siRNAs were subjected to SDS-PAGE and immunoblotting using the antibodies indicated. (e-f) Co-depletion of MUS81 (e) or XPF (f) reduces chromosomal aberrations in cells lacking DONSON. Quantification of the average number of chromatid gaps/breaks per metaphase in cells transfected with control, DONSON, MUS81 and/or XPF siRNA. At least 50 metaphases per experiment were counted (n=3). (g) Quantification of the average percentage of metaphases containing highly fragmented chromosomes or pulverized chromosomes in cells transfected with the indicated siRNAs. At least 50 metaphases per experiment were counted (n=3).

Figure 7: DONSON patient cells have spontaneous defects in replication fork progression that result in DNA damage

(a) Complementation of patient-derived fibroblasts with WT DONSON. Fibroblasts derived from DONSON patients P2, P6 and P9 were infected with retroviruses encoding either WT DONSON or an empty vector. DONSON expression was determined by immunoblotting. A non-specific cross-reactive protein represents a loading control. (b) Expression of WT DONSON in patient fibroblasts rescues elevated levels of spontaneous DNA damage. The percentage of cells from (a) with 53BP1/γH2AX foci was quantified by immunostaining.
(n=3). (c) DNA fibre analysis of complemented DONSON patient fibroblasts pulsed with CldU and IdU. Fork asymmetry was quantified. Plot indicates ratios of left/right fork track lengths of bidirectional replication forks. The red lines denote median ratios. (n=3). (d) The percentage of stalled forks and new origins from cells in (c) was quantified (n=3). Ongoing forks are shown in (Supplementary Fig. 19).
Table 1: Biallelic DONSON mutations identified in 29 individuals

<table>
<thead>
<tr>
<th>Patient</th>
<th>Country of Origin</th>
<th>Mutation 1</th>
<th>Mutation 2</th>
<th>Segregation</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1-1</td>
<td>Italy</td>
<td>c.1251_1256delCTCTAA, p.Asn417_Ser418del</td>
<td>haplotype</td>
<td>Het, M</td>
</tr>
<tr>
<td>P1-2</td>
<td>Italy</td>
<td>c.1251_1256delCTCTAA, p.Asn417_Ser418del</td>
<td>haplotype</td>
<td>Het, M</td>
</tr>
<tr>
<td>P2</td>
<td>UK</td>
<td>c.877C&gt;T, p.Arg293*</td>
<td>haplotype</td>
<td>Het, M, P</td>
</tr>
<tr>
<td>P3</td>
<td>UK</td>
<td>c.1254dupT, p.Lys419*</td>
<td>haplotype</td>
<td>Het, M, P</td>
</tr>
<tr>
<td>P4</td>
<td>UK</td>
<td>c.1686dupT, p.Asn563*</td>
<td>haplotype</td>
<td>Het, nps</td>
</tr>
<tr>
<td>P6</td>
<td>USA</td>
<td>c.1282C&gt;T, p.Gln428*</td>
<td>haplotype</td>
<td>Het, M, P</td>
</tr>
<tr>
<td>P7</td>
<td>USA</td>
<td>c.1282C&gt;T, p.Gln428*</td>
<td>haplotype</td>
<td>Het, nps</td>
</tr>
<tr>
<td>P8</td>
<td>Italy</td>
<td>c.1474_1475delCA, p.Gln492Glufs*18</td>
<td>c.786-7T&gt;C</td>
<td>Het, M, P</td>
</tr>
<tr>
<td>P9</td>
<td>Turkey</td>
<td>c.876C&gt;G, p.Phe292Leu</td>
<td>c.1047-9A&gt;G (SS)</td>
<td>Het, M</td>
</tr>
<tr>
<td>P10-1</td>
<td>India</td>
<td>c.1628_1630dupAAA, p.Gln543_Ile544insLys</td>
<td>c.1032C&gt;T, p.Ser344Ser (SS)</td>
<td>Het, M, P</td>
</tr>
<tr>
<td>P11</td>
<td>Saudi Arabia</td>
<td>c.786-22A&gt;G (SS)</td>
<td>c.786-22A&gt;G (SS)</td>
<td>Hom, M, P</td>
</tr>
<tr>
<td>P12</td>
<td>Saudi Arabia</td>
<td>c.786-22A&gt;G (SS)</td>
<td>c.786-22A&gt;G (SS)</td>
<td>Hom, nps</td>
</tr>
<tr>
<td>P13-1</td>
<td>Palestine</td>
<td>c.1337T&gt;C, p.Met446Thr</td>
<td>c.1337T&gt;C, p.Met446Thr</td>
<td>Hom, M, P</td>
</tr>
<tr>
<td>P13-2</td>
<td>Palestine</td>
<td>c.1337T&gt;C, p.Met446Thr</td>
<td>c.1337T&gt;C, p.Met446Thr</td>
<td>Hom, M, P</td>
</tr>
<tr>
<td>P14</td>
<td>Saudi Arabia</td>
<td>c.786-22A&gt;G (SS)</td>
<td>c.786-22A&gt;G (SS)</td>
<td>Hom, M, P</td>
</tr>
<tr>
<td>P15</td>
<td>Saudi Arabia</td>
<td>c.786-22A&gt;G (SS)</td>
<td>c.786-22A&gt;G (SS)</td>
<td>Hom, M, P</td>
</tr>
<tr>
<td>P16</td>
<td>Saudi Arabia</td>
<td>c.786-22A&gt;G (SS)</td>
<td>c.786-22A&gt;G (SS)</td>
<td>Hom, M, P</td>
</tr>
<tr>
<td>P17</td>
<td>Saudi Arabia</td>
<td>c.786-22A&gt;G (SS)</td>
<td>c.786-22A&gt;G (SS)</td>
<td>Hom, M, P</td>
</tr>
<tr>
<td>P18-1</td>
<td>Saudi Arabia</td>
<td>c.786-22A&gt;G (SS)</td>
<td>c.786-22A&gt;G (SS)</td>
<td>Hom, M, P</td>
</tr>
<tr>
<td>P18-2</td>
<td>Saudi Arabia</td>
<td>c.786-22A&gt;G (SS)</td>
<td>c.786-22A&gt;G (SS)</td>
<td>Hom, M, P</td>
</tr>
<tr>
<td>P18-3</td>
<td>Saudi Arabia</td>
<td>c.786-22A&gt;G (SS)</td>
<td>c.786-22A&gt;G (SS)</td>
<td>Hom, M, P</td>
</tr>
<tr>
<td>P19</td>
<td>Turkey</td>
<td>c.1297T&gt;C, p.Pro433Ser</td>
<td>c.1297T&gt;C, p.Pro433Ser</td>
<td>Hom, M, P</td>
</tr>
</tbody>
</table>

‘Haplotype’ indicates the presence of three co-segregating variants: c.82A>C (p.Ser28Arg); c.786-33A>G; c.1466A>C (p.Lys489Thr). Hom, homozygous in affected individual; Het, compound heterozygous in affected individual; M, mutation identified in mother; P, mutation identified in father; nps, no parental samples available; SS, Splice site mutation. Reference sequence, NM_017613.3.
Online Methods

Research subjects.

Genomic DNA from the affected children and family members was extracted from peripheral blood using standard methods or saliva samples using Oragene collection kits according to the manufacturer's instructions. Informed consent was obtained from all participating families. Ethics for the studies were approved by the Scottish Multicentre Research Ethics Committee (04:MRE00/19), by an IRB-approved research protocol (KFSHRC RAC# 2080006), and via the 'National Gene Mapping' protocol by Guy's and St. Thomas' National Health Service (NHS) Foundation Trust local research ethics committee (ref.: 08/H0802/84, "Systematic Characterization of Genes in Inherited Disorders"). In addition, ethical approval for linkage studies on the genetics of Fanconi anaemia in 1989 were obtained from the Guy's Hospital Research Ethics Committee (ref. EC89/10/27)22, with further approval for mutation analysis on existing samples in 1996 (ref. 96/3/9). Parents provided written consent for the publication of photographs of the affected individuals.

Exome sequencing and haplotype analysis.

Exome sequencing of genomic DNA and variant filtering was performed as described previously12. Cohort resequencing was performed by Sanger sequencing of PCR products representing all coding exons of DONSON (primer sequences are detailed in Supplementary Table 4), with variant calling using MutationSurveyor (SoftGenetics Inc.). Haplotype analysis was undertaken by SNP genotyping both patients using Affymetrix CytoScan 750K arrays. Genotypes were generated using Affymetrix Genotyping Console software and examined manually. The pathogenic impact of DONSON mutations was predicted using Alamut Visual Software (Interactive Biosoftware Inc).
**Cell culture and generation of cell lines.**

Lymphoblastoid cell lines (LCLs) were maintained in RPMI 1640 supplemented with 15% FBS, L-glutamine and penicillin/streptomycin antibiotics. LCLs were generated in house from peripheral blood samples by EBV transformation using standard methods. Dermal primary fibroblasts were grown from skin-punch biopsies in AmnioMax medium (Life Technologies) and then maintained in DMEM supplemented with 10% FBS, 5% L-glutamine and 5% penicillin/streptomycin antibiotics. Patient cell lines were validated using Sanger sequencing and immunoblotting. 293FT (Invitrogen) and HeLa (ATCC) cells were maintained in DMEM supplemented with 10% FBS, 5% L-glutamine and 5% penicillin/streptomycin antibiotics.

Stable cell lines were generated by Flp recombinase–mediated integration using HeLa-Flp-In T-REx host cells (gift from S. Taylor, University of Manchester) transfected with pcDNA5/FRT/TO-EGFP (vector only or EGFP-TRAIP) and pCAGGS-Flp.e (gift from D.-J. Kleinjan, University of Edinburgh). Transfected cells were selected using 5 μg/ml blasticidin and 400 μg/ml hygromycin, and the resulting colonies were then expanded for testing. Protein expression was induced with 1 μg/ml doxycycline (Sigma-Aldrich) treatment.

Primary fibroblasts derived from patients 2, 6 and 9 were immortalized with TERT retroviral supernatant with 4 μg/ml polybrene and infected with pMSCV-vector only or pMSCV-DONSON. Selection was performed using 750 ng/ml puromycin (Clontech) and 500 μg/ml neomycin (Invitrogen). Expression of the protein was verified by immunoblotting (Fig. 7a). All cell lines were routinely tested for mycoplasma.
Cell treatments.

Plasmids and siRNA oligos were transfected in Opti-MEM reduced serum medium using Lipofectamine 2000 and Oligofectamine (Life Technologies) respectively according to the manufacturer's guidelines. A custom siRNA targeting lacZ has previously been described\(^5\) and was used as a control siRNA. A DONSON siRNA SMARTpool (Dharmacon) was used for all siRNA transfections except when transfecting the HeLa Flp-In/T-Rex cells expressing an exogenous, siRNA-resistant, GFP-tagged DONSON construct. In this case a custom DONSON siRNA sequence (CCTGTGGACTGGAGTATTAdTdT) was used (Dharmacon). MUS81 siRNA SMARTpool and XPF siRNA SMARTpool (Dharmacon) were used where indicated. Transfected cells in both cases were analysed at 48-72 h post transfection. Where indicated, cells were treated with 1-2 mM hydroxyurea (Sigma-Aldrich), 50 ng/ml mitomycin C (Sigma-Aldrich) or 2 mM thymidine (Sigma-Aldrich). The ATR inhibitor (VE-821; Selleck Chemicals) was used at 5 μM. dNTP analogues EdU, CldU and IdU were purchased from Sigma Aldrich, and were used as indicated.

RT-PCR.

Total RNA was extracted from cell lines using the RNeasy kit (Qiagen) according to the manufacturer's instructions. DNA was removed by treatment with DNase I (Qiagen), and cDNA was generated using random oligomer primers and AMV RT (Roche). The RT-PCR primer pairs used are detailed in Supplementary Table 4.

DNA expression constructs

pEGFP-DONSON expression construct was created by cloning the human DONSON ORF into the pDONR221 Gateway shuttle vector (Invitrogen). WT DONSON was amplified from cDNA and recombined into the pDEST-EGFP vector to generated a GFP-tagged DONSON expression construct. The DONSON ORF was made siRNA resistant using site-
directed mutagenesis (Agilent Technologies) by altering the following nucleotides:
CCTGTGGACTGGAGTATTA was changed to CCCGTAGATTGGTATCA. Patient-
associated mutations were engineered into the pEGFP-DONSON expression plasmid
using site-directed mutagenesis according to the manufacturer’s instructions. (All primers
are detailed in Supplementary Table 4).

The retroviral expression construct expressing DONSON was created by recombination
between the pDONR221-DONSON vector and a Gateway-compatible pMSCVneo
retroviral expression construct (Clontech).

The human telomerase reverse transcriptase (hTERT) expressing retroviral construct used
to immortalise patient-derived human fibroblasts was a kind gift from Bob Weinberg
(Addgene plasmid: #1771).

Minigene splicing reporter assay
A 1.58 kb stretch of the DONSON gene encompassing the 3’ end of intron 3, exon 4,
intron 4, exon 5 and the 5’ end of intron 5 was amplified using DNA from a healthy
individual and DONSON patients (carrying mutation c.786-22A>G or c.786-33A>G) using
the DONSON-int3-SalI-F and DONSON-int5-SpeI-R primers, and cloned into the RHCGlo
vector\textsuperscript{57} using the SalI and SpeI restriction sites. Site-directed mutagenesis was used to
introduce the DONSON intron 4 splice acceptor mutation (c.786-1G>A) into the splicing
reporter construct. HeLa cells were transfected with each individual splicing mutation
reporter construct using Lipofectamine 2000 according the manufacturer’s instructions. 24
h post-transfection, cells were harvested, total cellular RNA was extracted and cDNA
generated using Superscript III reverse transcriptase first-stand synthesis system
(Invitrogen). PCR was carried out using primers (RSV_minigene_F and RSV_minigene_R)
to the 5' and 3' ends of the artificial exons present in the RHCglo vector. DONSON WT and mutant cDNA amplicons were resolved on a 2% agarose gel to visualise differences in splicing. Individual PCR products were subsequently cloned into the pGEM-T Easy Vector (Promega) and sequenced to verify the exon content of each transcript. All relevant primers are detailed in Supplementary Table 4.

iPOND

iPOND was performed as previously described\(^{35,58}\). Briefly, exponentially growing cells were incubated with 10 µM EdU for 10 min, cross-linked with 1% formaldehyde, harvested and permeabilised. For pulse-chase controls, cells were incubated in 10 µM EdU for 10 min, washed in media containing 10 µM thymidine, then incubated with media containing 10 µM thymidine for 1 h, before being cross-linked. Biotin azide was covalently attached to EdU within newly replicated DNA using a Click reaction, and EdU containing DNA was precipitated using Streptavidin agarose beads. Edu co-precipitates were then analysed by mass spectrometry. Log2 abundance values represent the ratio of proteins found in EdU-pulsed samples compared to those pulse-chased with EdU-thymidine.

Immunoblot analysis and antibodies

Whole cell extracts were obtained by sonication in UTB buffer (8 M Urea, 50 mM Tris, 150 mM β-mercaptoethanol) and analysed by SDS-PAGE following standard procedures. Protein samples were run on 6-12% acrylamide SDS-PAGE or 4-12% NuPage mini-gels (Life Technologies) and transferred onto nitrocellulose membrane. Immunoblotting was performed using antibodies to: Cyclin A (Santa Cruz, sc-751; 1:1,000), CHK1 (Santa Cruz, sc-8408; 1:1,000), CHK2 (Santa Cruz, sc-5278; 1:1000), FANCD2 (Santa Cruz, sc-20022; 1:1000), MCM2 (BD Transduction Laboratories, 610700; 1:10000), MCM7 (Santa Cruz, sc-56324; 1:1000), MUS81 (Santa Cruz, sc-53382; 1:2000); XPF (Santa Cruz, sc-136153;
1:1000); H2A (Millipore, 07-146; 1:3000), γ-H2AX (Millipore, 05-636; 1:3000), RPA2
(Millipore, NA18; 1:1000), phospho-histone H3 Ser-10-P (Millipore); pS343-NBS1 (Abcam,
47272; 1:500); NBS1 (Genetex, GTX70224; 1:10000); ATR (Bethyl Laboratories, A300-
137A; 1:1000), pS345-CHK1 (Cell Signaling Technology, 2341; 1:100), pS4/S8-RPA2
(Bethyl Laboratories, A300-245A; 1:1,000), pS966-SMC1 (Bethyl Laboratories, A300-
050A; 1:1,000), SMC1 (Bethyl Laboratories, A300-055A; 1:1,000), Treslin (Bethyl
Laboratories, A303-472A; 1:1,000); TOPBP1 (Bethyl Laboratories; A300-111A; 1:1000);
Vinculin (Sigma-Aldrich, V9264; 1:1,000); α-Tubulin (Sigma-Aldrich, T5168; 1:4000); GFP
(Roche, 11814460001; 1:500). The polyclonal anti-DONSON antibody was generated by
immunising rabbits with a GST-fusion protein encoding aa 1-125 of human DONSON.
Antibody was affinity-purified from rabbit sera (Eurogentec) and specificity established
using lysates from patient cells and DONSON siRNA-transfected cells.

Loading controls for all blots derive from reprobing the same membrane, except for
phospho-antibody blots, where paired gels were run simultaneously, and blotted in parallel
for phosphorylated and total proteins.

**Immunofluorescence and fluorescent microscopy.**
siRNA transfected HeLa cells or passage-matched TERT-immortalized fibroblasts were
seeded on coverslips 24 h before extraction/fixation. To visualise cells undergoing DNA
replication, cells incubated in medium containing 10 μM EdU for 10-30 min prior to
harvesting. To remove soluble proteins before immunofluorescence, cells were pre-
extracted for 10 min on ice with ice-cold buffer (25 mM HEPES, pH 7.4, 50 mM NaCl, 1
mM EDTA, 3 mM MgCl2, 300 mM sucrose and 0.5% Triton X-100) and then fixed with 4%
paraformaldehyde for 15 min. For analysis of cells transfected with GFP-tagged protein,
cells were fixed and permeabilised by incubation with ice-cold methanol for 20 minutes.
EdU immunolabeling was performed using Click-iT EdU Imaging Kit (Invitrogen, C10337) according to the manufacturer’s protocol. Cells were stained for 53BP1 (Novus Biologicals, NB100-304; 1:1,000) and/or γH2AX (Millipore, 05-636; 1:1000) and stained with secondary antibodies conjugated to Alexa Fluor-568 (Life Technologies) and DAPI.

For quantification of signal-integrated densities of γH2AX staining, cells were stained with an antibody specific to γH2AX (Millipore, 05-636; 1:1000), images were visualized using a Zeiss Axioplan 2 microscope with iVision software (BioVision Technologies) and captured using a 40× oil-immersion objective. For quantification of signal-integrated densities of RPA staining, cells were stained with RPA2 antibody (Millipore, NA18; 1:200), images were visualized using a Nikon Eclipse Ni microscope with NIS-Elements software (Nikon Instruments) and captured using a 100× oil-immersion objective. Nuclei were segmented on the basis of DAPI staining and then signal-integrated density of γH2AX or RPA staining quantified for each nuclear region using ImageJ software (US National Institutes of Health).

For quantification of γH2AX staining, more than 100 EdU positive cells and 50 EdU negative cells were analyzed per experiment per condition, and for quantification of RPA staining, more than 200 cells were analyzed per experiment per condition. Exposure time, binning, microscope settings and light source intensity were kept constant for all the samples in all cases.

For quantification of native BrdU foci cells were incubated in medium containing 10 μM BrdU for 24 h prior to harvesting. Six hours prior to harvesting, 2 mM HU was added to the media. To visualise ssDNA foci, cells were extracted for 10 min on ice with ice-cold buffer (25 mM HEPES, pH 7.4, 50 mM NaCl, 1 mM EDTA, 3 mM MgCl2, 300 mM sucrose and 0.5% Triton X-100) and then fixed with 4% paraformaldehyde for 15 min. After fixation,
cells were washed with PBS and blocked in 3% FCS in PBS for 30 min at room temperature. ssDNA was visualised using a BrdU antibody (Abcam, ab6326; 1:500). To denature DNA cells were incubated in 2 M HCl in PBS for 30 min prior to addition of the BrdU antibody. Images were acquired as for γH2AX quantification and foci were quantified using ImageJ-based script. Nuclei were defined on the basis of DAPI staining and native BrdU foci were detected using "Find maxima" function of ImageJ within each nuclear region. Exposure time, binning, microscope settings, light source intensity and the noise level in “Find maxima” function were kept constant for all the samples within each individual experiment. More than 100 cells were analyzed per experiment per condition.

**Metaphase spreads**

Chromosomal aberrations were scored in Giemsa-stained metaphase spreads as previously described. Briefly, demecolcine (Sigma Aldrich) was added at a final concentration of 0.2 µg/ml 4 h prior to harvesting. Cells were harvested by trypsinisation, subjected to hypotonic shock for 1 hour at 37°C in 0.3 M sodium citrate and fixed in 3:1 methanol:acetic acid solution. Cells were dropped onto acetic acid humidified slides, stained for 15 minutes in Giemsa-modified solution (Sigma; 5% v/v in H₂O) and washed in water for 5 minutes.

**DNA fibre spreading assay**

Passage-matched primary, TERT-immortalized fibroblasts or siRNA transfected HeLa cells were pulse labeled with CldU for 20 min, washed with media and damaged with 2 mM hydroxyurea for 2 h before being pulse labeled with IdU for 40 min. Alternatively, 50 ng/ml mitomycin C was added to the cells 24 h before CldU pulse labeling and left on during 20 min CldU and 20 min IdU pulse labeling. Cells were harvested by trypsinization, and cell pellets were washed in PBS. 5 × 10⁵ cells were lysed directly onto glass slides using
spreading buffer (200 mM Tris-HCl, pH 7.5, 50 mM EDTA, 0.5% SDS) and fixed in methanol/acetic acid (3:1 ratio). Following 2.5 M HCl denaturation, CldU was detected using rat anti-BrdU (clone BU1/75, ICR1; Abcam, ab6326; 1:750), and IdU was detected using mouse anti-BrdU (clone B44; BD Biosciences, 347583; 1:750). Slides were then fixed in 4% paraformaldehyde before immunostaining with secondary antibodies conjugated to Alexa Fluor-594 or Alexa Fluor-488 (Life Technologies). Labeled DNA fibers were visualized using a Nikon Eclipse Ni microscope with NIS-Elements software (Nikon Instruments). Images were captured using 40× oil-immersion objectives and were recoloured and analyzed using ImageJ software (US National Institutes of Health). For quantification of replication structures, at least 400 structures were counted per experiment. Tract lengths were measured using ImageJ (National Institutes of Health; http://rsbweb.nih.gov/ij/). To calculate fork velocity, arbitrary length values were converted into micrometers using the scale bars created by the microscope, with 1 μm equivalent to 2.59 kb. Replication fork speed (kb/min) was then determined by dividing the length of CldU and IdU tracks (in kb) from ongoing forks by the pulse time.

**FACS analysis**

For BrdU analysis, HeLa cells were pulse labeled with 10 μM BrdU for 30 min before fixation with 70% ethanol at −20 °C for 16 h. Cells were then digested with 1 mg/ml pepsin and denatured with 2 M HCl before washing with PBS and blocking in 0.5 % BSA, 0.5 % Tween-20. BrdU labeling was detected using anti-BrdU antibody (Abcam, ab6326; 1:75) and FITC-conjugated anti-rat secondary antibody. DNA content was assessed by staining with 50 μg/ml propidium iodide. Cells were sorted on a BD Biosciences FACS Aria II and data were analyzed using FlowJo software (v7.6.1, Tree Star).
For mitotic analysis and immuno-detection of phospho-histone H3 (Ser10), HeLa cells were harvested, fixed, permeabilised 24 h post exposure to HU or MMC, as previously described. Cells were analysed using an Accuri flow cytometer (BDBiosciences) in conjunction with CFlowplus software. Data represents that obtained from at least 30,000 cells.

**Immunoprecipitation and GFP-Trap**

293FT cells transfected with plasmids encoding GFP-DONSON or GFP were untreated, or exposed to 2 mM HU for 16 h and harvested. Cells were then incubated in lysis buffer (150 mM NaCl, 50 mM Tris HCl pH7.5, 2 mM MgCl₂, 1 % NP40, 90 U/ml Benzonase (Novagen) and Protease Inhibitor Cocktail EDTA free (Roche)) for 30 min with rotation at 4 °C. The resultant cell lysates were pre-cleared at 44,000 rpm at 4 °C for 30 min.

For immunoprecipitations, 3 mg of lysate was immunoprecipitated with 5 µg of antibody, immune complexes collected with Protein A-Sepharose (Sigma-Aldrich). Complexes were washed with wash buffer (150 mM NaCl, 50 mM Tris HCl pH 7.5, 0.5 % NP40, and Complete Protease Inhibitor Cocktail (Roche)) and analysed by SDS-PAGE.

For GFP-Trap, 3 mg lysates were incubated with GFP-Trap agarose beads (ChromoTek) at 4 °C for 5 h. The resulting GFP-Trap complexes were washed with wash buffer as above and analysed by SDS-PAGE. Experiments were carried out in the presence of Benzonase Nuclease to exclude the possibility of interactions being mediated by DNA.

For mass spectrometry analysis, GFP or GFP-DONSON were isolated from tetracycline induced, or uninduced, Flp-In T-REx HeLa cell extracts by incubation with GFP-trap magnetic agarose beads (Chromotek) for 2 hours on a Kingfisher Duo robotic handling
station (Thermo). Asynchronous cells and S-phase accumulated cells, using a 24 h treatment with 2 mM HU, were analysed. On-bead digest and mass spectrometry were performed as described\textsuperscript{60}. Data represents three independent experiments for each condition, analysed by back-to-back MS and quantified by Label free quantification (LFQ). Proteins were identified and quantified with the MaxQuant 1.5 software suite by searching against the Uniprot human database. M(ox) and protein N-terminal acetylation were set as variable, and carbamylation as a fixed modification, with a 1 % FDR. Contaminants and reverse data base hits were deleted. Protein significantly enriched by GFP-DONSON were selected on the basis of p-value <0.05, and >2 fold change from asynchronous to S-phase, as identified by Student t-test and ratio cut-off against the respective negative control LFQ data as determined by MaxQuant (p<0.05; 2-fold).

**Proximity ligation assay (PLA).**

PLA was carried out as described in\textsuperscript{5,56}. Briefly, cells from GFP or GFP-DONSON Flp-In T-REx HeLa cell lines were treated with 1ug/ml doxycycline and fixed/extracted after 24 h. For PCNA visualisation, cells were fixed with methanol at -20 °C for 10 min followed by a 5 min extraction in 0.3% Triton-X100 in PBS. For RPA visualisation, cells were pre-extracted in nuclear extraction buffer for 5 minutes on ice, and fixed in 3.6% paraformaldehyde for 10 minutes at room temperature. Cells were then incubated in anti-PCNA (PC10, 1:500, Santa Cruz) or anti-RPA (NA18; 1:500; Merck-Millipore) antibodies along with anti-GFP antibody (ab6556, 1:500, Abcam), and in situ proximity ligation was performed using a Duolink Detection Kit (Sigma Aldrich). Nuclear foci were imaged using a Nikon Eclipse Ni-U microscope equipped with a 100X oil lens in conjunction with a Zyla camera, and images were acquired using Elements software (Nikon). More than 200 cells were analysed per experiment per condition.
Fluorescence Cross-Correlation Spectroscopy (FCCS)

HeLa cells stably expressing GFP-DONSON and mCherry-PCNA (construct kindly provided by C. Lukas, Copenhagen; referred to as RFP-PCNA) were used for FCCS. For all details on Fluorescence Microscopy Imaging and FCS/FCCS, refer to Supplementary Note.

Statistical Analyses

Statistical differences were analyzed by: two-tailed Student T-Test (Fig. 3b, 3d, 3f, 4f, 4g, 5b-d, 5f, 6a, 6e, 6f, 7b, 7d and Supplementary Fig. 5c, 8b, 9a, 12c, 13i, 14c, 16a-c, 17a-c, 19, 20, 21); Mann-Whitney rank sum test (Fig. 3e, 6c, 7c and Supplementary Fig. 12b, 15a, 15c); and Chi-Squared Test (Fig 4c, 6g and Supplementary Fig. 16d). n refers to number of independent experiments unless indicated. Error bars represent standard error of the mean (s.e.m.) unless specified.

Data Availability

The NGS data used in the manuscript can be obtained from the European Genome-phenome Archive (EGA) under accession EGAS00001002224. NGS datasets on patients P14-P18, p21 are not available due to institutional IRB restrictions. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD005690.
Methods-only references


<table>
<thead>
<tr>
<th>Patient</th>
<th>Country of Origin</th>
<th>Mutation 1</th>
<th>Mutation 2</th>
<th>Segregation</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1-1</td>
<td>Italy</td>
<td>c.1251_1256delCTCTAA, p.Asn417_Ser418del</td>
<td>haplotype</td>
<td>Het, M</td>
</tr>
<tr>
<td>P1-2</td>
<td>Italy</td>
<td>c.1251_1256delCTCTAA, p.Asn417_Ser418del</td>
<td>haplotype</td>
<td>Het, M</td>
</tr>
<tr>
<td>P2</td>
<td>UK</td>
<td>c.877G&gt;T, p.Arg293*</td>
<td>haplotype</td>
<td>Het, M, P</td>
</tr>
<tr>
<td>P3</td>
<td>UK</td>
<td>c.1254dupT, p.Lys419*</td>
<td>haplotype</td>
<td>Het, M, P</td>
</tr>
<tr>
<td>P4</td>
<td>UK</td>
<td>c.1686dupT, p.Asn563*</td>
<td>haplotype</td>
<td>Het, nps</td>
</tr>
<tr>
<td>P6</td>
<td>USA</td>
<td>c.1282G&gt;T, p.Gln428*</td>
<td>haplotype</td>
<td>Het, M, P</td>
</tr>
<tr>
<td>P7</td>
<td>USA</td>
<td>c.1282G&gt;T, p.Gln428*</td>
<td>haplotype</td>
<td>Het, nps</td>
</tr>
<tr>
<td>P8</td>
<td>Italy</td>
<td>c.1474_1475delCA, p.Gln492Glu*+18</td>
<td>c.786-7T&gt;C</td>
<td>Het, M, P</td>
</tr>
<tr>
<td>P9</td>
<td>Turkey</td>
<td>c.876C&gt;T, p.Tyr282Cys</td>
<td>c.1047-9A&gt;G (SS)</td>
<td>Het, M</td>
</tr>
<tr>
<td>P10-1</td>
<td>India</td>
<td>c.1628_1630dupAAA, p.Gln543_Ile544insLys</td>
<td>c.1032G&gt;T, p.Ser344Ser (SS)</td>
<td>Het, M, P</td>
</tr>
<tr>
<td>P11</td>
<td>Saudi Arabia</td>
<td>c.786-22A&gt;G (SS)</td>
<td>c.786-22A&gt;G (SS)</td>
<td>Hom, M, P</td>
</tr>
<tr>
<td>P12</td>
<td>Saudi Arabia</td>
<td>c.786-22A&gt;G (SS)</td>
<td>c.786-22A&gt;G (SS)</td>
<td>Hom, nps</td>
</tr>
<tr>
<td>P13-1</td>
<td>Palestine</td>
<td>c.1337T&gt;C, p.Met446Thr</td>
<td>c.1337T&gt;C, p.Met446Thr</td>
<td>Hom, M, P</td>
</tr>
<tr>
<td>P13-2</td>
<td>Palestine</td>
<td>c.1337T&gt;C, p.Met446Thr</td>
<td>c.1337T&gt;C, p.Met446Thr</td>
<td>Hom, M, P</td>
</tr>
<tr>
<td>P13-3</td>
<td>Palestine</td>
<td>c.1337T&gt;C, p.Met446Thr</td>
<td>c.1337T&gt;C, p.Met446Thr</td>
<td>Hom, M, P</td>
</tr>
<tr>
<td>P14</td>
<td>Saudi Arabia</td>
<td>c.786-22A&gt;G (SS)</td>
<td>c.786-22A&gt;G (SS)</td>
<td>Hom, M, P</td>
</tr>
<tr>
<td>P15</td>
<td>Saudi Arabia</td>
<td>c.786-22A&gt;G (SS)</td>
<td>c.786-22A&gt;G (SS)</td>
<td>Hom, M, P</td>
</tr>
<tr>
<td>P16</td>
<td>Saudi Arabia</td>
<td>c.786-22A&gt;G (SS)</td>
<td>c.786-22A&gt;G (SS)</td>
<td>Hom, M, P</td>
</tr>
<tr>
<td>P17</td>
<td>Saudi Arabia</td>
<td>c.786-22A&gt;G (SS)</td>
<td>c.786-22A&gt;G (SS)</td>
<td>Hom, M, P</td>
</tr>
<tr>
<td>P18-1</td>
<td>Saudi Arabia</td>
<td>c.786-22A&gt;G (SS)</td>
<td>c.786-22A&gt;G (SS)</td>
<td>Hom, M, P</td>
</tr>
<tr>
<td>P18-2</td>
<td>Saudi Arabia</td>
<td>c.786-22A&gt;G (SS)</td>
<td>c.786-22A&gt;G (SS)</td>
<td>Hom, M, P</td>
</tr>
<tr>
<td>P18-3</td>
<td>Saudi Arabia</td>
<td>c.786-22A&gt;G (SS)</td>
<td>c.786-22A&gt;G (SS)</td>
<td>Hom, M, P</td>
</tr>
</tbody>
</table>

‘Haplotype’ indicates the presence of three co-segregating variants: c.82A>C (p.Ser28Arg); 786-33A>G; 1466A>C (p.Lys489Thr). Hom, homozygous in affected individual; Het, compound heterozygous in affected individual; M, mutation identified in mother; P, mutation identified in father; nps, no parental samples available; SS, Splice site mutation. Reference sequence, NM_017613.3.
Figure 1

(a) Growth Parameters

<table>
<thead>
<tr>
<th>Birth</th>
<th>Current</th>
</tr>
</thead>
<tbody>
<tr>
<td>OFC</td>
<td>Lgt</td>
</tr>
</tbody>
</table>

SD

-4  -8  -12

(b) Images of patients

P1-1  P1-2  P2  P5
P6  P12  P15  P16
P17  P18-1  P18-2  P18-3

(c) Diagram of DONSON gene variations

- c.786-33 A>G
- c.786-22A>G
- c.786-7T>C
- c.1032C>T
- c.1047-9A>G
- S28R
- C278R
- F292L
- R293*
- N417_S418del
- K419*
- Q428*
- M446T
- K489T
- Q543_L544insK
- E504K
- Q492Efs*18
- N563*
Figure 2

(a) LCLs

<table>
<thead>
<tr>
<th></th>
<th>WT3</th>
<th>Parent</th>
<th>P13-1</th>
<th>P13-2</th>
<th>P13-3</th>
<th>WT4</th>
<th>P1-1</th>
</tr>
</thead>
<tbody>
<tr>
<td>DONSON</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ATR</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(b) Fibroblasts

<table>
<thead>
<tr>
<th></th>
<th>WT1</th>
<th>P12</th>
<th>P2</th>
<th>P9</th>
<th>WT2</th>
<th>P6</th>
<th>P10-2</th>
</tr>
</thead>
<tbody>
<tr>
<td>DONSON</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ATR</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(c) DONSON siRNA

<table>
<thead>
<tr>
<th></th>
<th>-DOX</th>
<th>+DOX</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S28R</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M446T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>K489T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S28R</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M446T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>K489T</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- GFP-DONSON
- TOPBP1
Figure 3

(a) Western blot analysis showing expression levels of DONSON, Cyclin A, H3 S10-P, and Vinculin in asynchronously dividing cells (Async) and at various time points (0-24 hours) post-release. The blots indicate a change in protein expression over time.

(b) Bar graph showing the percentage of cells in different cell cycle phases (G1, S, G2/M) after treatment with Control siRNA or DONSON siRNA. The p-values are indicated for each comparison.

(c) Schematic diagram illustrating the timeline of CldU and IdU incorporation followed by harvest. The duration of each event is marked with "20 min".

(d) Distributions of fork speed in kilobases per minute (kbp/min) for Control siRNA and DONSON siRNA. The p-values for the comparison of fork speed distributions are provided.

(e) Images showing symmetrical and asymmetrical patterns of 53BP1 and γ-H2AX foci. The p-value for the comparison of foci distribution is indicated.

(f) Images comparing DAPI, 53BP1, γ-H2AX, and Y-H2AX expression levels in Control siRNA and DONSON siRNA treated cells. The percentage of cells with >10 53BP1/γH2AX foci is shown with a p-value for the comparison.

Additionally, the data indicates ongoing forks, new origins, and stalled forks with p-values for each comparison.
Figure 4

<table>
<thead>
<tr>
<th>a</th>
<th>GFP</th>
<th>GFP-DONSON</th>
</tr>
</thead>
<tbody>
<tr>
<td>Uninduced</td>
<td>Asyc</td>
<td>Uninduced</td>
</tr>
<tr>
<td>MCM2</td>
<td>USP11</td>
<td>MCM4</td>
</tr>
<tr>
<td>MCM2</td>
<td>USP11</td>
<td>MCM4</td>
</tr>
</tbody>
</table>

Fold change LFQ intensity

<table>
<thead>
<tr>
<th>b</th>
<th>2% Input</th>
<th>GFP-Trap</th>
</tr>
</thead>
<tbody>
<tr>
<td>GFP</td>
<td>DONSON</td>
<td>GFP</td>
</tr>
<tr>
<td>HU</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IB</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treslin</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MCM7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MCM2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PCNA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GFP-DONSON</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GFP</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

% PLA signal

<table>
<thead>
<tr>
<th>c</th>
<th>6+ dots/nucleus</th>
<th>1-5 dots/nucleus</th>
<th>0 dots/nucleus</th>
</tr>
</thead>
<tbody>
<tr>
<td>P &lt; 0.001</td>
<td>P &lt; 0.001</td>
<td>P &lt; 0.001</td>
<td>P &lt; 0.001</td>
</tr>
<tr>
<td>GFP</td>
<td>GFP-DONSON</td>
<td>GFP + HU</td>
<td>GFP-DONSON + HU</td>
</tr>
<tr>
<td>GFP</td>
<td>GFP-DONSON</td>
<td>GFP + HU</td>
<td>GFP-DONSON + HU</td>
</tr>
</tbody>
</table>

αGFP + αPCNA

<table>
<thead>
<tr>
<th>d</th>
<th>αGFP + αRPA</th>
</tr>
</thead>
<tbody>
<tr>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>HU</td>
<td></td>
</tr>
<tr>
<td>DAPI + PLA</td>
<td></td>
</tr>
</tbody>
</table>

αGFP + αPCNA

αGFP + αRPA

<table>
<thead>
<tr>
<th>e</th>
<th>GFP-DONSON</th>
<th>RFP-PCNA</th>
</tr>
</thead>
<tbody>
<tr>
<td>S-phase</td>
<td>Non-replicating</td>
<td></td>
</tr>
<tr>
<td>Replication foci</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-replicating</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GFP</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>f</th>
<th>G(t)</th>
</tr>
</thead>
<tbody>
<tr>
<td>P &lt; 0.001</td>
<td></td>
</tr>
<tr>
<td>P = 0.14</td>
<td></td>
</tr>
<tr>
<td>P = 9.4E-05</td>
<td></td>
</tr>
<tr>
<td>P = 8.3E-05</td>
<td></td>
</tr>
</tbody>
</table>

G(t)

<table>
<thead>
<tr>
<th>g</th>
<th>Log2 abundance (pulse/chase)</th>
</tr>
</thead>
<tbody>
<tr>
<td>DONSON</td>
<td>RPA1-3</td>
</tr>
<tr>
<td>P = 0.00023</td>
<td></td>
</tr>
<tr>
<td>P = 0.14</td>
<td></td>
</tr>
<tr>
<td>P = 8.3E-05</td>
<td></td>
</tr>
<tr>
<td>P = 9.4E-05</td>
<td></td>
</tr>
</tbody>
</table>
Figure 5

(a) Untreated

20 min 20 min

CldU IdU Harvest

HU treated

20 min 2 h 20 min

2 mM HU +/- 5 µM ATRi

CldU IdU Harvest

MMC treated

24 h 20 min 20 min

50 ng MMC CldU IdU Harvest

(b) Stalled replication forks

Control siRNA DONSON siRNA

Percentage of all CldU-labelled

Untreated + HU + MMC

P = 5.7E-05

P = 0.016

P = 0.0078

(c) New origins

Untreated + HU + MMC

P = 0.104

P = 6.7E-05

P = 0.001

(d) New origins

Untreated + HU + MMC

P = 0.1539

P = 6.7E-05

P = 0.40

(e) Time (h) after 1mM HU

Control siRNA DONSON siRNA

DONSON

FANCD2

SMC1-S966-P

SMC1

NBS1 S343-P

NBS1

CHK1 S345-P

CHK1

RPA2-S4/8-P

RPA2

γ-H2AX

H2A

(f) Percentage of mitotic cells compared to untreated control

+ HU

P = 4.6E-06
Figure 6

(a) Bar graph showing the average chromatid gaps/breaks per metaphase for Control siRNA and DONSON siRNA. The y-axis represents the number of gaps/breaks ranging from 0 to 2.5, and the x-axis shows the siRNA treatments. The error bars indicate standard error.

(b) Images illustrating Chromatid gaps/breaks for Control siRNA, DONSON siRNA, Mus81 siRNA, and DONSON/Mus81 siRNA treatments. The images depict highly fragmented and pulverised chromosomes.

(c) Graph showing the left fork length/right fork length (ratio) for different siRNA treatments. The x-axis includes Con siRNA, DONSON siRNA, MUS81 siRNA, DONSON/MUS81 siRNA, XPF siRNA, and DONSON/XPF siRNA. The y-axis represents the ratio ranging from 0 to 8.0. The error bars indicate standard error.

(d) Western blot analysis for siRNA treatments with bands for DONSON, MUS81, XPF, γ-H2AX, and H2A.

(e) Bar graph showing the average chromatid gaps/breaks per metaphase for different siRNA treatments. The x-axis includes Con siRNA, DONSON siRNA, MUS81 siRNA, DONSON/MUS81 siRNA, XPF siRNA, and DONSON/XPF siRNA. The y-axis represents the number of gaps/breaks ranging from 0 to 3. The error bars indicate standard error.

(f) Bar graph showing the average chromatid gaps/breaks per metaphase for different siRNA treatments. The x-axis includes Con siRNA, DONSON siRNA, MUS81 siRNA, DONSON/MUS81 siRNA, XPF siRNA, and DONSON/XPF siRNA. The y-axis represents the number of gaps/breaks ranging from 0 to 2. The error bars indicate standard error.

(g) Bar graph showing the percentage of metaphases with Highly Fragmented and Pulverised Chromosomes for different siRNA treatments. The x-axis includes Con, DONSON, MUS81, DONSON/MUS81, XPF, and DONSON/XPF siRNA. The y-axis represents the percentage ranging from 0 to 7. The error bars indicate standard error.