



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

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

Citation for published version:

Reynolds, J, Bicknell, L, Carroll, P, Higgs, M, Shaheen, R, Murray, J, Papadopoulos, DK, Leitch, A, Murina, O, Tarnauskaite, Z, Wessel, SR, Zlatanou, A, Vernet, A, Von Kriegsheim, A, Mottram, RM, Logan, C, Bye, H, Li, Y, Brean, A, Maddirevula, S, Challis, R, Skouloudaki, K, Almoisheer, A, Alsaif, HS, Amar, A, Prescott, NJ, Bober, M, Duker, A, Faqeih, E, Seidahmed, MZ, Tala, SA, Alswaid, A, Ahmed, S, Al-Aama, JY, Altmuller, J, Al Balwi, M, Brady, A, Chessa, L, Cox, H, Fischetto, R, Heller, R, Henderson, BD, Hobson, E, Nurnberg, P, Percin, F, Peron, A, Spaccini, L, Quigley, A, Thakur, S, Wise, CA, Yoon, G, Alnemer, M, Yigit, G, Taylor, AMR, Reijns, M, Simpson, MA, Cortez, D, Alkuraya, FS, Mathew, CG, Jackson, A & Stewart, GS 2017, 'Mutations in DONSON disrupt replication fork stability and cause microcephalic dwarfism.' *Nature Genetics*. DOI: 10.1038/ng.3790

Digital Object Identifier (DOI):

[10.1038/ng.3790](https://doi.org/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.



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

3
4 John J Reynolds^{1,36}, Louise S Bicknell^{2,3,36}, Paula Carroll^{2,36}, Martin R Higgs^{1,36}, Ranad
5 Shaheen^{4,36}, Jennie E Murray², Dimitrios K Papadopoulos⁵, Andrea Leitch², Olga Murina²,
6 Žygimantė Tarnauskaitė², Sarah R Wessel⁶, Anastasia Zlatanou¹, Audrey Vernet¹, Alex
7 von Kriegsheim², Rachel MA Mottram¹, Clare V Logan², Hannah Bye⁷, Yun Li⁸, Alexander
8 Brean¹, Sateesh Maddirevula⁴, Rachel C Challis², Kassiani Skouloudaki⁵, Agaadir
9 Almoisheer⁴, Hessa S Alsaif⁴, Ariella Amar⁷, Natalie J Prescott⁷, Michael B Bober⁹, Angela
10 Duker⁹, Eissa Faqeh¹⁰, Mohammed Zain Seidahmed¹¹, Saeed Al Tala¹², Abdulrahman
11 Alswaid¹³, Saleem Ahmed^{14,15}, Jumana Yousuf Al-Aama^{14,15}, Janine Altmüller¹⁶,
12 Mohammed Al Balwi¹⁷, Angela F Brady¹⁸, Luciana Chessa¹⁹, Helen Cox²⁰, Rita
13 Fischetto²¹, Raoul Heller²², Bertram D Henderson²³, Emma Hobson²⁴, Peter Nürnberg¹⁶, E
14 Ferda Percin²⁵, Angela Peron^{26,27}, Luigina Spaccini²⁶, Alan J Quigley²⁸, Seema Thakur²⁹,
15 Carol A Wise³⁰, Grace Yoon^{31,32}, Maha Alnemer³³, Pavel Tomancak⁵, Gökhan Yigit⁸, A
16 Malcolm R Taylor¹, Martin AM Reijns², Michael A Simpson⁷, David Cortez⁶, Fowzan S
17 Alkuraya^{4,35,*}, Christopher G Mathew^{7,34,35,*}, Andrew P Jackson^{2,35,*} & Grant S Stewart^{1,35,*}.

18
19 ¹Institute of Cancer and Genomic Sciences, College of Medical and Dental Sciences,
20 University of Birmingham, Birmingham, UK.

21 ²MRC Human Genetics Unit, IGMM, University of Edinburgh, Edinburgh, UK.

22 ³Present address: Department of Pathology, Dunedin School of Medicine, University of
23 Otago, Dunedin, New Zealand.

24 ⁴Department of Genetics, King Faisal Specialist Hospital and Research Center, Riyadh,
25 Saudi Arabia.

26 ⁵Max-Planck Institute of Molecular Cell Biology and Genetics, Dresden, Germany.

27 ⁶Department of Biochemistry, Vanderbilt University School of Medicine, Nashville,
28 Tennessee, USA.

29 ⁷Department of Medical and Molecular Genetics, Faculty of Life Science and Medicine,
30 King's College London, 7th Floor Tower Wing, Guy's Hospital, London, UK.

31 ⁸Institute of Human Genetics, University Medical Center Göttingen, Göttingen, Germany.

32 ⁹Nemours/Alfred I. duPont Hospital for Children, Wilmington, Delaware, USA.

33 ¹⁰Department of Pediatric Subspecialties, Children's Hospital, King Fahad Medical City,
34 Riyadh, Saudi Arabia.

35 ¹¹Department of Pediatrics, Security Forces Hospital, Riyadh, Saudi Arabia.

36 ¹²Armed forces hospital, SR. P.D. Genetic unit. Khamis Mushayt, Saudi Arabia.

37 ¹³Department of Pediatrics, King Abdulaziz Medical City, Riyadh, Saudi Arabia.

38 ¹⁴Department of Genetic Medicine, King Abdulaziz University Hospital, Jeddah, Saudi
39 Arabia

40 ¹⁵Princess Al-Jawhara Al-Brahim Center of Excellence in Research of Hereditary
41 Disorders, Jeddah, Saudi Arabia.

42 ¹⁶Cologne Center for Genomics and Center for Molecular Medicine Cologne, University of
43 Cologne, Cologne, Germany.

44 ¹⁷Department of Pathology and Laboratory Medicine, King Abdulaziz Medical City,
45 National Guard Health Affairs, Riyadh, Saudi Arabia.

46 ¹⁸North West Thames Regional Genetics Service, London North West Healthcare NHS
47 Trust, Harrow, UK.

48 ¹⁹Department of Clinical and Molecular Medicine, University La Sapienza, Roma, Italy.

49 ²⁰West Midlands Regional Clinical Genetics Service, Birmingham Women's Hospital, West
50 Midlands, UK.

51 ²¹Pediatric Hospital Giovanni XXIII, Bari, Italy.

52 ²²Institute of Human Genetics, University Hospital of Cologne, Cologne, Germany.

53 ²³Division of Clinical Genetics, Faculty of Health Sciences, UFS, Bloemfontein, South
54 Africa.

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

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

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

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

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

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

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

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

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

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

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

75

76 ³⁵These authors jointly directed this work.

77 ³⁶These authors contributed equally to this work.

78

79 *Correspondence to:

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

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

82 **Abstract**

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

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

94 In summary, we identify mutations in *DONSON* as a common cause of
95 microcephalic dwarfism, and establish DONSON as a critical replication fork protein
96 required for mammalian DNA replication and genome stability.

97 Microcephalic primordial dwarfism (MPD) is the collective term for a group of human
98 disorders characterised by intra-uterine and postnatal growth delay alongside marked
99 microcephaly¹, and includes disorders such as MOPD II, ATR/ATRIP-Seckel syndrome
100 and Meier-Gorlin syndrome. Mutations in genes encoding either components of the DNA
101 replication machinery (replisome) or genome stability proteins are a frequent cause of
102 microcephalic dwarfism²⁻¹⁴.

103 During the course of normal DNA replication, a subset of replication forks may stall,
104 causing ‘replication stress’¹⁵. This stalling can be caused by endogenous or exogenous
105 sources, such as collision of the replisome with DNA lesions or the transcriptional
106 machinery, or replication of difficult to replicate genomic regions. To facilitate efficient
107 genome duplication, stalled replication forks must be stabilised and protected from
108 collapse. Multiple factors safeguard replication fork stability, many of which function within
109 the ATR-CHK1-dependent replication stress response¹⁶⁻¹⁸. This pathway ensures that fork
110 stabilisation is tightly coordinated with a global reduction in DNA synthesis, allowing stalled
111 or damaged forks to be repaired and restarted^{19,20}.

112 Exome sequencing analysis of microcephalic dwarfism patients has identified
113 several novel factors that regulate replication and/or the replication stress response. Using
114 this strategy, we recently identified mutations in *TRAIP* in individuals with MPD⁵, and
115 demonstrated that TRAIP is required for the response to replication-blocking DNA lesions.
116 To identify similar disease-associated genes, we carried out whole exome sequencing of
117 genetically uncharacterised patients with microcephaly. Here, we report the identification
118 of *DONSON* as a new microcephalic dwarfism gene, and demonstrate that DONSON is a
119 novel replisome component that maintains genome stability by protecting stalled/damaged
120 replication forks.

121 Results

122 ***DONSON* mutations identified in microcephalic dwarfism patients**

123 Whole exome sequencing (WES) was undertaken on 26 patients with microcephaly
124 and reduced stature. After aligning WES reads to the reference genome, variant calling,
125 and filtering for rare variants (MAF <0.005), analysis under a recessive model of
126 inheritance identified rare biallelic variants in the *DONSON* ('Downstream neighbour of
127 *SON*') gene in nine affected individuals from seven families. Sanger capillary sequencing
128 confirmed the presence of these mutations in these patients (P1-1 to P3, P6, and P9 to
129 P11, **Table 1**). Subsequent re-sequencing of an additional 230 patients with primary
130 microcephaly, microcephaly with reduced stature, or MPD, identified five additional
131 families with compound heterozygous mutations in *DONSON* (P4, P5, P7, P8, P12; **Table**
132 **1**). All variants segregated amongst family members in a manner consistent with an
133 autosomal recessive trait, and were present at a frequency of <0.5% in the ExAC
134 database²¹.

135 Two other concurrent molecular genetic studies provided further independent
136 evidence to support the identification of *DONSON* as a novel human disease gene. Firstly,
137 exome sequencing was carried out on a consanguineous Palestinian family previously
138 reported to have a Fanconi Anaemia-like disorder²². These patients presented with
139 microcephaly, short stature, slow growth and forearm and thumb dysplasia, although no
140 individuals had haematological evidence of bone marrow failure. This WES analysis
141 revealed a deleterious homozygous transition, c.1337T>C, resulting in substitution of a
142 highly conserved residue (p.M446T) in all three affected individuals (P13-1, P13-2, P13-3;
143 **Table 1, Supplementary Fig. 1**). Secondly, a study of five consanguineous families in
144 Saudi Arabia with extreme microcephaly and short stature allowed a 1.6 Mb haplotype
145 shared by all five families (combined multipoint LOD score $Z= 8.0$) to be mapped to a
146 defined critical interval on chromosome 21 that contained *DONSON*. Whole exome and

147 genome sequencing identified a single rare variant at this locus in *DONSON*, c.786-22A>G.
148 Capillary sequencing confirmed this intronic variant to be homozygous in all seven affected
149 individuals from this study (P14 to P18-3; **Table 1**), identical to that detected in two Saudi
150 Arabian individuals present within the first study described above (P11, P12).

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

154 ***DONSON* mutations give rise to severe microcephaly with short stature**

155 Despite their identification in separate studies, all patients with *DONSON* mutations
156 had similar clinical phenotypes. Marked microcephaly was present (OFC -7.5 +/- 2.4 SD),
157 with a substantial reduction in cerebral cortical size, along with decreased gyral folding
158 evident on neuroimaging (**Fig. 1a** and **Supplementary Fig. 2**), similar to that previously
159 reported for other primary microcephaly and microcephalic dwarfism patients^{5,23-25}. Height
160 was reduced (-3.2 +/- 1.4 SD), although much less so than head circumference (**Fig. 1a**),
161 and to a lesser degree than observed in other microcephalic dwarfism-associated
162 disorders (where height was typically ≤ -4 SD)^{2,3,5,8-10,24,26}. Minor skeletal abnormalities
163 were present in several patients, including fifth finger clinodactyly, syndactyly,
164 brachydactyly, hypoplasia of carpal/metacarpal/phalangeal bones, or radial head
165 dislocation (**Supplementary Table 1**). Absent/hypoplastic patellae were present in
166 patients P12, P20-1 and P20-2. Notably, patient P19 had bilateral hypoplasia of the radius
167 and thumb, which, together with the limb abnormalities displayed by P13-1 and 13-2,
168 established radial ray defects as an uncommon but recurrent phenotype. In family P21, the
169 most extreme phenotype was observed, with substantial limb shortening/reduction in
170 association with foetal lethality (**Supplementary Fig. 3**). Aside from microcephaly, neither
171 a recognisable facial phenotype (**Fig. 1b**) nor recurrent malformations affecting other
172 organ systems were evident. Intellectual disability, if present, was typically mild.

173 In conclusion, the number of biallelic variants identified, combined with a common
174 clinical presentation, provided strong evidence for *DONSON* being a novel human disease
175 gene, associated with autosomal recessive inheritance. We therefore investigated the
176 consequence of these mutations on *DONSON* protein function.

177

178 ***DONSON* mutations markedly reduce protein levels**

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

186 Mutations in multiple families were associated with two different ancestral
187 haplotypes (P1-1 to P7; P11, P12 and P14 to 18-3 respectively; **Table 1**) and were
188 investigated in more detail. Firstly, as described above, nine individuals (P11, P12 and
189 P14 to P18-3) were homozygous for the c.786-22A>G mutation, predicted to enhance a
190 cryptic splice donor site within intron 4 (MaxEntScan)²⁷. Consistent with a common
191 ancestral founder, five consanguineous families of Saudi Arabian origin bearing this
192 mutation (P14 to P18-3) shared a 1.6 Mb haplotype region of chromosome 21
193 (**Supplementary Fig. 4**). qRT-PCR analysis of RNA isolated from four patients with this
194 variant demonstrated a significant decrease in full-length transcript, and increased
195 skipping of exon 5 (**Supplementary Fig. 5**). This resulted in an out-of-frame mRNA
196 predicted to undergo NMD, explaining the substantial reduction in protein levels seen in
197 fibroblasts homozygous for this mutation (P12; **Fig 2b**).

198 Secondly, two missense variants in DONSON, p.S28R and p.K489T, and an
199 intronic variant (c.786-33A>G) were present in seven individuals of European ancestry and
200 one Somali (P1-1 to P7; **Table 1**). These were associated with a different ancestral
201 haplotype, comprising a shared 127.7 kb genomic region (**Supplementary Table 2**). No
202 other deleterious biallelic variants were present in the four other genes within this region.
203 Despite the close proximity of the c.786-33A>G intronic variant to the Saudi Arabian c.786-
204 22A>G mutation, the former change did not disrupt splicing between exons 4 and 5, as
205 assessed by either mini gene splicing assays (**Supplementary Fig. 6**), or RT-PCR
206 analyses of patient cell lines (data not shown), indicating this variant is unlikely to be
207 pathogenic. Since DONSON protein levels were severely reduced in cells from patients
208 inheriting this haplotype allele in combination with a truncating mutation (P2, P6), this
209 suggested that either one or both of the missense variants associated with this haplotype
210 (p.S28R, p.K489T) compromised protein levels (**Fig. 2b**).

211 To investigate these two variants, we established isogenic HeLa cell lines
212 expressing doxycycline-inducible, siRNA-resistant wild type (WT) or mutant (p.S28R,
213 p.K489T) GFP-tagged DONSON. Following siRNA depletion of endogenous DONSON
214 (**Supplementary Fig. 7**), and induction of exogenous GFP-DONSON, immunoblotting
215 revealed that the p.K489T mutation, but not p.S28R, reduced protein levels in a post-
216 transcriptional manner (**Fig. 2c** and **Supplementary Fig. 8**). This suggested that the
217 p.K489T substitution within the second haplotype causes the decreased DONSON protein
218 levels observed in P2 and P6 (**Fig. 2b**). The K489T variant is present as a rare allele in the
219 population, observed at a frequency of 0.00099 in the ExAC database²¹. In the patients
220 reported here, it is always observed *in trans* with a frameshift or other protein-disrupting
221 allele, suggesting that it is a functionally weak variant insufficient to cause disease alone, a
222 conclusion supported by the presence of a single homozygous individual in ExAC.

223 Eight other point mutations were identified in patients: six missense substitutions, a
224 two amino-acid deletion and an amino-acid insertion (**Table 1** and **Fig. 1c**). Notably these
225 were at highly conserved residues (**Supplementary Fig. 1**) and predicted to be
226 deleterious (Alamut Visual). In agreement, exogenous expression of the p.M446T mutant
227 resulted in significantly reduced protein levels (**Fig. 2c**), similar to cells derived from
228 patients P13-1, P13-2 and P13-3. Furthermore, five of these mutations disrupted the
229 subcellular localisation of GFP-DONSON (**Supplementary Fig. 9**), suggesting that these
230 alterations also compromise DONSON protein function via protein mis-localisation.

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

238 Taken together, the deleterious consequences of the identified mutations on
239 splicing, transcript abundance, subcellular localisation and/or protein levels strengthened
240 our conclusion that *DONSON* was a novel human disease gene. The fact that knockout of
241 murine *Donson* leads to developmental lethality²⁸, together with the presence of residual
242 DONSON protein in patient-derived cell lines (**Fig. 2a, b** and **Supplementary Fig. 11**),
243 supports the notion that the identified mutations are hypomorphic, retaining some residual
244 function.

245

246 **DONSON stabilises replication forks during normal DNA replication**

247 While *DONSON* is highly conserved in metazoa and plants, its precise function(s)
248 remained to be defined. *Humpty-dumpty* (*hd*), the *Drosophila* ortholog of *DONSON*, has

249 been proposed to play a role in cell proliferation: Hd expression peaks during S-phase; *hd*
250 mutants have an 'egg shell' phenotype; and clonal inactivation of *hd* impairs genome
251 replication in larval tissues²⁹.

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

266 Since DNA replication is closely linked with genome stability^{15,30,31}, we reasoned
267 that loss of DONSON would lead to a failure to complete timely replication and increased
268 S-phase DNA damage. To test this hypothesis, we combined immunofluorescence of γ -
269 H2AX and 53BP1 (markers of DNA damage and DNA double strand breaks respectively)
270 with EdU labelling to identify S-phase cells. We observed that a significant proportion of
271 DONSON-depleted cells exhibited spontaneous γ -H2AX and 53BP1 foci (**Fig. 3f**), of which
272 the majority occurred in S-phase cells (**Supplementary Fig. 12a-d**), consistent with the
273 identification of DONSON as a potential genome stability regulator by high-throughput
274 siRNA screening³². Using pulsed-field gel electrophoresis, we also observed elevated

275 levels of DNA double strand breaks in the absence of DONSON (**Supplementary Fig.**
276 **12e**). Taken together, these data support a role for DONSON in maintaining replication
277 fork stability during unperturbed DNA replication, and demonstrate that spontaneous DNA
278 damage arises in replicating cells in the absence of DONSON.

279

280 **DONSON is a component of the replisome**

281 To shed further light on the role of DONSON in regulating replication fork stability,
282 we carried out mass spectrometry screening to identify interaction partners of GFP-tagged
283 DONSON. Amongst the interactors, we detected multiple replication proteins including
284 subunits of the MCM helicase and the GINS complex (**Fig. 4a** and **Supplementary Table**
285 **3**). To confirm these findings, we carried out pull-down analyses coupled with
286 immunoblotting to identify GFP-DONSON binding proteins. Consistent with our mass
287 spectrometry data, we detected interactions between GFP-DONSON and the replisome
288 components MCM2, MCM7, Treslin and PCNA (**Fig. 4b**), suggesting that DONSON
289 associated with the replisome.

290 We next used three complementary techniques to assess whether DONSON
291 localised to sites of DNA replication. Firstly, we carried out proximity ligation assays (PLA)
292 of GFP-DONSON with the replication proteins PCNA and RPA. In line with DONSON
293 being closely associated with the replication machinery, we observed robust PLA signals
294 between GFP-DONSON and both PCNA and RPA (**Fig. 4c, d**). We next performed
295 Fluorescence Cross-Correlation Spectroscopy (FCCS)^{33,34} in live HeLa cells stably co-
296 expressing RFP-PCNA and GFP-DONSON, to measure the degree of co-diffusion of
297 these molecules. Significantly increased co-diffusion of PCNA and DONSON was
298 observed in S-phase PCNA foci, but not in nuclei of non-replicating cells (**Fig. 4e-f** and
299 **Supplementary Fig. 13**), indicating that these proteins interacted during DNA replication.
300 Finally, we utilised iPOND (isolation of proteins on nascent DNA)³⁵ combined with mass

301 spectrometry to ascertain whether DONSON is present on newly replicated DNA. Crucially,
302 this approach demonstrated that DONSON, like MCMs and RPA, was significantly
303 enriched at replication forks compared to mature chromatin (**Fig. 4g**).

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

306

307 **DONSON depletion impairs cell-cycle checkpoint activation**

308 Since our data suggested that DONSON functions to protect replication forks during
309 unperturbed DNA replication, we extended our findings to evaluate the role of DONSON in
310 preventing replication fork stalling following exogenous replication stress. Exposure to the
311 replication stress-inducing agents hydroxyurea (HU) and mitomycin C (MMC) induced
312 significantly more fork stalling in DONSON-depleted cells than in control cells (**Fig. 5a-b**).
313 DONSON depletion also resulted in a failure to suppress new origin firing upon exogenous
314 replication stress (**Fig. 5c**). Since suppression of new origin firing reflects checkpoint
315 activity, this suggests that DONSON is required for efficient activation of the intra-S phase
316 checkpoint. To further investigate this, we measured activation of this checkpoint after
317 inhibition of ATR (VE821; ATRi), the apical kinase which governs the replication stress
318 response¹⁸. Upon HU exposure and ATR inhibition, we observed no difference in the
319 number of new origins fired between control or DONSON-depleted cells, indicating that
320 DONSON and ATR may function within the same pathway to activate the intra-S phase
321 checkpoint (**Fig. 5d**).

322 We next examined whether the ATR-dependent replication stress response was
323 functional in the absence of DONSON. We first monitored ATR pathway activation in
324 DONSON-depleted cells treated with HU or MMC by immunoblotting, using phospho-
325 specific antibodies to known ATR substrates. This analysis revealed that cells lacking
326 DONSON failed to efficiently phosphorylate a number of ATR substrates, such as CHK1

327 and NBS1, in response to HU or MMC (**Fig. 5e** and **Supplementary Fig. 14a**). Moreover,
328 ATR autophosphorylation on T1989, another marker of ATR activation³⁶, was reduced
329 (**Supplementary Fig. 14b**). Loss of DONSON also significantly increased mitotic indices
330 following exposure to HU or MMC as measured by FACS, demonstrating that DONSON-
331 depleted cells fail to efficiently activate the G2/M checkpoint in response to replication
332 stress (**Fig. 5f** and **Supplementary Fig. 14c**). We next determined whether the reduced
333 ATR signalling observed was due to decreased levels of RPA-coated ssDNA, which is the
334 stimulus for ATR activation. Surprisingly, DONSON-depleted cells exhibited elevated
335 levels of RPA-coated ssDNA following HU treatment (**Supplementary Fig. 15**), consistent
336 with defective activation of the ATR-dependent replication stress response.

337 Dysregulated DNA replication combined with impaired intra-S phase checkpoint
338 signalling, such as in ATR-deficient cells³⁷⁻⁴¹, gives rise to extensive chromosome
339 breakage and genome instability. Consistent with this, we observed significantly elevated
340 levels of spontaneous micronuclei and chromatid gaps/breaks in cells lacking DONSON
341 (**Fig. 6a** and **Supplementary Fig. 16a**), which was exacerbated by exposure to HU or
342 MMC (**Supplementary Fig. 16a-c**). We also observed spontaneously-arising highly-
343 fragmented or completely pulverised metaphase chromosomes in cells lacking DONSON,
344 which increased upon exogenous replication stress (**Fig. 6b** and **Supplementary Fig.**
345 **16d**).

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

349

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

351 It has been proposed that the spontaneous DNA damage arising in ATR-deficient
352 cells is due to processing of stalled/damaged forks by SLX4-associated structure-specific

353 nucleases, such as MUS81, SLX1 and XPF⁴²⁻⁴⁶. We therefore postulated that the
354 replication abnormalities and chromosomal aberrations of DONSON-deficient cells might
355 arise via similar mechanisms. Indeed, the spontaneous replication fork asymmetry and
356 H2AX phosphorylation exhibited by DONSON-depleted cells were partially reduced by co-
357 depletion of either MUS81 or XPF (**Fig. 6c, d**). Moreover, co-depletion of MUS81 or XPF
358 also reduced chromosome breakage and genomic pulverisation in these cells (**Fig. 6e-g**).
359 From this, we conclude that the severe genome instability apparent in the absence of
360 DONSON is due to nucleolytic processing of damaged replication forks by structure-
361 specific nucleases.

362

363 **Replication stress-induced genomic damage in DONSON patient cells**

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

374 In addition, using isogenic cell lines inducibly expressing GFP-DONSON (**Fig. 2c**),
375 we observed that expression of the haplotype-associated S28R mutant, but not the K489T
376 variant, complemented loss of endogenous DONSON by rescuing the spontaneous fork
377 stalling observed upon DONSON depletion (**Supplementary Fig. 19**). This is consistent
378 with K489T being the pathogenic variant within the haplotype region (**Fig. 2c**).

379 Finally, we set out to demonstrate that the patient-associated cellular phenotypes
380 were directly due to mutation of *DONSON*. We first established three paired isogenic cell
381 lines via transduction of patient-derived fibroblasts with retroviral expression vectors
382 encoding WT *DONSON* or an empty vector (**Fig. 7a**). Importantly, the spontaneous DNA
383 damage, replication fork stalling, replication fork asymmetry and intra-S phase checkpoint
384 defect were all corrected by expression of WT *DONSON* (**Fig. 7b-d** and **Supplementary**
385 **Fig. 20**), confirming that these phenotypes were directly due to *DONSON* deficiency.
386 Lastly, using one of these cell lines, we also observed that inhibition of ATR and mutation
387 of *DONSON* are epistatic with regard to the observed replication abnormalities
388 (**Supplementary Fig. 21**).

389 Discussion

390 Here we identify *DONSON* as a novel human disease gene, and describe 29
391 individuals with a range of mutations in *DONSON*, establishing such alterations as a
392 frequent cause of microcephalic dwarfism. Since normal embryonic development requires
393 rapid cellular proliferation^{47,48} it is exquisitely sensitive to genetic perturbations that impact
394 DNA replication^{1-3,6}. A failure to complete timely genome duplication will profoundly affect
395 the number of cells generated during embryonic development. For example, hypomorphic
396 mutations in *ATR* result in severe microcephaly and growth retardation, both in humans
397 and in a murine model^{2,3,49,50}, due to the role that *ATR* plays in preventing replication
398 stress during development⁴⁹⁻⁵⁰. We propose that mutation of *DONSON* similarly reduces
399 the number of cells generated during development via a failure to maintain replication fork
400 stability in the presence of endogenous replication stress, thus explaining the decreased
401 organism size observed. Furthermore, since brain development requires rapid proliferation
402 of neural progenitor cells within a limited timeframe⁴⁷, it is particularly sensitive to
403 disruptive genetic perturbations. This may explain why brain development is
404 disproportionately affected in these individuals compared to growth.

405 *DONSON* has no predicted domain structure or paralogs, and previous
406 characterisation has been limited to two previous studies: an siRNA screen proposing that
407 *DONSON* regulates genome stability, and a study of its *Drosophila* ortholog *Humpty-*
408 *dumpty* suggesting a role in cell proliferation^{29,32}. Consistent with this, we establish that
409 *DONSON* is a replisome component that ensures replication fork stability, and promotes
410 efficient activation of both intra-S and G2/M cell-cycle checkpoints upon exogenous
411 replication stress. Loss of *DONSON* leads to increased spontaneous stalling of replication
412 forks, which are subsequently cleaved into replication-associated DNA double strand
413 breaks by structure-specific nucleases. Defective cell-cycle checkpoint activation in
414 *DONSON*-deficient cells then allows transmission of these breaks into mitosis, accounting

415 for the elevated chromosomal damage and genome fragmentation observed
416 (**Supplementary Fig. 22**). Additional studies will be important to confirm this model, and to
417 investigate whether DONSON is a constitutive component of the replisome, or whether it is
418 recruited to a subset of replication forks. Furthermore, establishing which replisome
419 components DONSON directly interacts with, and the functional importance of these
420 associations, will also inform understanding of its biological function.

421 The mechanism by which DONSON ensures replication fork stability and promotes
422 checkpoint activation remains to be defined. We propose that in addition to being a
423 replisome component, DONSON is also involved in promoting the ATR-CHK1 replication
424 stress response, since we observed that DONSON-depleted cells exhibit defective
425 activation of cell cycle checkpoints and reduced ATR-dependent signalling in response to
426 exogenous replication stress. This hypothesis is supported by the observation that
427 impaired replication alone, such as that arising from a hypomorphic mutation in MCM4
428 (*MCM4^{Chaos3/Chaos3}*), does not give rise to decreased CHK1 phosphorylation or increased
429 new origin firing upon replication stress⁵¹. However, it is unclear whether DONSON
430 functions directly or indirectly to regulate the ATR-CHK1 pathway. Our demonstration that
431 cells lacking DONSON do not exhibit a global reduction in replication, or decreased levels
432 of RPA-coated ssDNA, indicates that loss of DONSON does not affect the cells ability to
433 generate the primary stimulus for ATR pathway activation. Based on this, we propose that
434 either DONSON directly activates ATR, in a manner similar to TOPBP1⁵² or ETAA1^{53,54}, or
435 functions indirectly to regulate other factors necessary for efficient ATR-CHK1 signalling,
436 such as the MRE11/RAD50/NBS1 (MRN)⁵⁵ complex or TIPIN/Timeless^{19,20}. Since
437 DONSON does not contain a canonical ATR activation domain, which is found in both
438 TOPBP1 and ETAA1, we favour the latter possibility. However, how DONSON acts to
439 promote ATR signalling is not yet clear, and future work will be critical in establishing
440 whether this is direct or indirect.

441 It is also evident that the cellular phenotype of cells lacking DONSON cannot be
442 explained solely by abnormal DNA replication or defective ATR-dependent signalling. In
443 particular, exposure of cells lacking ATR to exogenous replication stress results in highly
444 elevated levels of H2AX phosphorylation, a situation not observed upon DONSON loss
445 (**Fig. 5e**), despite the presence of substantial amounts of DNA damage. Therefore, whilst
446 our observations are consistent with a role for DONSON in promoting ATR-CHK1
447 signalling, DONSON may also impact on other pathways that promote H2AX
448 phosphorylation at the replication fork, for example those governed by ATM or the MRN
449 complex.

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

454 **Acknowledgements**

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

473

474 **Author Contributions**

475 J.J.R., M.R.H., P.C., O.M., A.Z., A.L., R.M.A.M., A.B. and G.S.S. designed and performed
476 the cell biology experiments. J.E.M., L.S.B., R.S, C.V.L., F.S.A, M.A.S., C.G.M., Y.L., S.M.
477 and G. Yigit performed NGS sequencing and analysis. L.S.B., P.C., R.C.C., R.S., A.V.,
478 J.E.M., M.A.S., C.V.L., Z.T., M.A.M.R., H.B., A. Amar., S.M., A. Almoisheer, H.S.A. and
479 N.J.P. performed sequencing, genotyping, linkage analysis, analysis of splicing and other

480 molecular genetics experiments. D.C. and S.R.W. performed iPOND experiments. P.T.,
481 D.K.P. and K.S. performed FCCS analysis. A.V.K. performed mass spectrometry analysis.
482 E.F., M.Z.S., S.A.T., A. Alswaid, S.A., J.Y.A., M.A.B., A.F.B., L.C., H.C., A.D., R.F., E.H.,
483 E.F.P., A.P., L.S., S.T., G. Yoon., J.A., P.N., A.J.Q., B.D.H., M.A. and R.H. contributed
484 clinical cases and clinical data and analysis for the study. M.B.B., C.A.W., J.E.M., L.S.B.,
485 A.M.R.T., F.S.A., C.G.M. and A.P.J. recruited study cohorts, and performed review of
486 phenotypes and sample collection. J.J.R., M.R.H., L.S.B., A.P.J. and G.S.S. wrote the
487 manuscript. The study was planned and supervised by G.S.S., C.G.M., F.S.A and A.P.J.

488

489 **Competing Financial Interests Statement**

490 The authors declare no competing financial interests.

491 **References**

- 492 1. Klingseisen, A. & Jackson, A.P. Mechanisms and pathways of growth failure in
493 primordial dwarfism. *Genes Dev* **25**, 2011-24 (2011).
- 494 2. O'Driscoll, M., Ruiz-Perez, V.L., Woods, C.G., Jeggo, P.A. & Goodship, J.A. A
495 splicing mutation affecting expression of ataxia-telangiectasia and Rad3-related
496 protein (ATR) results in Seckel syndrome. *Nat Genet* **33**, 497-501 (2003).
- 497 3. Ogi, T. *et al.* Identification of the first ATRIP-deficient patient and novel mutations in
498 ATR define a clinical spectrum for ATR-ATRIP Seckel Syndrome. *PLoS Genet* **8**,
499 e1002945 (2012).
- 500 4. German, J. Bloom's syndrome. I. Genetical and clinical observations in the first
501 twenty-seven patients. *Am J Hum Genet* **21**, 196-227 (1969).
- 502 5. Harley, M.E. *et al.* TRAP promotes DNA damage response during genome
503 replication and is mutated in primordial dwarfism. *Nat Genet* **48**, 36-43 (2016).
- 504 6. Qvist, P. *et al.* CtIP Mutations Cause Seckel and Jawad Syndromes. *PLoS Genet* **7**,
505 e1002310 (2011).
- 506 7. Rosin, N. *et al.* Mutations in XRCC4 cause primary microcephaly, short stature and
507 increased genomic instability. *Hum Mol Genet* **24**, 3708-17 (2015).
- 508 8. Bicknell, L.S. *et al.* Mutations in the pre-replication complex cause Meier-Gorlin
509 syndrome. *Nat Genet* **43**, 356-9 (2011).
- 510 9. Bicknell, L.S. *et al.* Mutations in ORC1, encoding the largest subunit of the origin
511 recognition complex, cause microcephalic primordial dwarfism resembling Meier-
512 Gorlin syndrome. *Nat Genet* **43**, 350-5 (2011).
- 513 10. Guernsey, D.L. *et al.* Mutations in origin recognition complex gene ORC4 cause
514 Meier-Gorlin syndrome. *Nat Genet* **43**, 360-4 (2011).

- 515 11. Fenwick, A.L. *et al.* Mutations in CDC45, Encoding an Essential Component of the
516 Pre-initiation Complex, Cause Meier-Gorlin Syndrome and Craniosynostosis. *Am J*
517 *Hum Genet* **99**, 125-38 (2016).
- 518 12. Murray, J.E. *et al.* Extreme growth failure is a common presentation of ligase IV
519 deficiency. *Hum Mutat* **35**, 76-85 (2014).
- 520 13. Murray, J.E. *et al.* Mutations in the NHEJ component XRCC4 cause primordial
521 dwarfism. *Am J Hum Genet* **96**, 412-24 (2015).
- 522 14. Shaheen, R. *et al.* Genomic analysis of primordial dwarfism reveals novel disease
523 genes. *Genome Res* **24**, 291-9 (2014).
- 524 15. Zeman, M.K. & Cimprich, K.A. Causes and consequences of replication stress. *Nat*
525 *Cell Biol* **16**, 2-9 (2014).
- 526 16. Zou, L. & Elledge, S.J. Sensing DNA damage through ATRIP recognition of RPA-
527 ssDNA complexes. *Science* **300**, 1542-8 (2003).
- 528 17. MacDougall, C.A., Byun, T.S., Van, C., Yee, M.C. & Cimprich, K.A. The structural
529 determinants of checkpoint activation. *Genes Dev* **21**, 898-903 (2007).
- 530 18. Nam, E.A. & Cortez, D. ATR signalling: more than meeting at the fork. *Biochem J*
531 **436**, 527-36 (2011).
- 532 19. Chou, D.M. & Elledge, S.J. Tipin and Timeless form a mutually protective complex
533 required for genotoxic stress resistance and checkpoint function. *Proc Natl Acad Sci*
534 *U S A* **103**, 18143-7 (2006).
- 535 20. Kemp, M.G. *et al.* Tipin-replication protein A interaction mediates Chk1
536 phosphorylation by ATR in response to genotoxic stress. *J Biol Chem* **285**, 16562-
537 71 (2010).
- 538 21. Lek, M. *et al.* Analysis of protein-coding genetic variation in 60,706 humans. *Nature*
539 **536**, 285-91 (2016).

- 540 22. Milner, R.D., Khallouf, K.A., Gibson, R., Hajianpour, A. & Mathew, C.G. A new
541 autosomal recessive anomaly mimicking Fanconi's anaemia phenotype. *Arch Dis*
542 *Child* **68**, 101-3 (1993).
- 543 23. Germanaud, D. *et al.* Simplified gyral pattern in severe developmental
544 microcephalies? New insights from allometric modeling for spatial and spectral
545 analysis of gyrification. *Neuroimage* **102 Pt 2**, 317-31 (2014).
- 546 24. Martin, C.A. *et al.* Mutations in PLK4, encoding a master regulator of centriole
547 biogenesis, cause microcephaly, growth failure and retinopathy. *Nat Genet* **46**,
548 1283-92 (2014).
- 549 25. Trimborn, M. *et al.* Mutations in microcephalin cause aberrant regulation of
550 chromosome condensation. *Am J Hum Genet* **75**, 261-6 (2004).
- 551 26. Griffith, E. *et al.* Mutations in pericentrin cause Seckel syndrome with defective
552 ATR-dependent DNA damage signaling. *Nat Genet* **40**, 232-6 (2008).
- 553 27. Yeo, G. & Burge, C.B. Maximum entropy modeling of short sequence motifs with
554 applications to RNA splicing signals. *J Comput Biol* **11**, 377-94 (2004).
- 555 28. Mouse Genome Database (MGD) at the Mouse Genome Informatics website. Vol.
556 2016 (ed. Laboratory, T.J.) (Bar Harbor, Maine).
- 557 29. Bandura, J.L. *et al.* humpty dumpty is required for developmental DNA amplification
558 and cell proliferation in *Drosophila*. *Curr Biol* **15**, 755-9 (2005).
- 559 30. Cortez, D. Preventing replication fork collapse to maintain genome integrity. *DNA*
560 *Repair (Amst)* **32**, 149-57 (2015).
- 561 31. Errico, A. & Costanzo, V. Mechanisms of replication fork protection: a safeguard for
562 genome stability. *Crit Rev Biochem Mol Biol* **47**, 222-35 (2012).
- 563 32. Fuchs, F. *et al.* Clustering phenotype populations by genome-wide RNAi and
564 multiparametric imaging. *Mol Syst Biol* **6**, 370 (2010).

- 565 33. Papadopoulos, D.K. *et al.* Probing the kinetic landscape of Hox transcription factor-
566 DNA binding in live cells by massively parallel Fluorescence Correlation
567 Spectroscopy. *Mech Dev* **138 Pt 2**, 218-25 (2015).
- 568 34. Bacia, K. & Schwille, P. Practical guidelines for dual-color fluorescence cross-
569 correlation spectroscopy. *Nat Protoc* **2**, 2842-56 (2007).
- 570 35. Sirbu, B.M. *et al.* Analysis of protein dynamics at active, stalled, and collapsed
571 replication forks. *Genes Dev* **25**, 1320-7 (2011).
- 572 36. Liu, S. *et al.* ATR autophosphorylation as a molecular switch for checkpoint
573 activation. *Mol Cell* **43**, 192-202 (2011).
- 574 37. Durkin, S.G., Arlt, M.F., Howlett, N.G. & Glover, T.W. Depletion of CHK1, but not
575 CHK2, induces chromosomal instability and breaks at common fragile sites.
576 *Oncogene* **25**, 4381-8 (2006).
- 577 38. Ozeri-Galai, E., Schwartz, M., Rahat, A. & Kerem, B. Interplay between ATM and
578 ATR in the regulation of common fragile site stability. *Oncogene* **27**, 2109-17 (2008).
- 579 39. Toledo, L.I. *et al.* ATR prohibits replication catastrophe by preventing global
580 exhaustion of RPA. *Cell* **155**, 1088-103 (2013).
- 581 40. Brown, E.J. & Baltimore, D. ATR disruption leads to chromosomal fragmentation
582 and early embryonic lethality. *Genes Dev* **14**, 397-402 (2000).
- 583 41. Brown, E.J. & Baltimore, D. Essential and dispensable roles of ATR in cell cycle
584 arrest and genome maintenance. *Genes Dev* **17**, 615-28 (2003).
- 585 42. Forment, J.V., Blasius, M., Guerini, I. & Jackson, S.P. Structure-specific DNA
586 endonuclease Mus81/Eme1 generates DNA damage caused by Chk1 inactivation.
587 *PLoS One* **6**, e23517 (2011).
- 588 43. Couch, F.B. *et al.* ATR phosphorylates SMARCAL1 to prevent replication fork
589 collapse. *Genes Dev* **27**, 1610-23 (2013).

- 590 44. Ragland, R.L. *et al.* RNF4 and PLK1 are required for replication fork collapse in
591 ATR-deficient cells. *Genes Dev* **27**, 2259-73 (2013).
- 592 45. Hodskinson, M.R. *et al.* Mouse SLX4 is a tumor suppressor that stimulates the
593 activity of the nuclease XPF-ERCC1 in DNA crosslink repair. *Mol Cell* **54**, 472-84
594 (2014).
- 595 46. Svendsen, J.M. *et al.* Mammalian BTBD12/SLX4 assembles a Holliday junction
596 resolvase and is required for DNA repair. *Cell* **138**, 63-77 (2009).
- 597 47. Takahashi, T., Nowakowski, R.S. & Caviness, V.S., Jr. The cell cycle of the
598 pseudostratified ventricular epithelium of the embryonic murine cerebral wall. *J*
599 *Neurosci* **15**, 6046-57 (1995).
- 600 48. Snow, M.H.L. Gastrulation in the mouse: growth and regionalization of epiblast. *J.*
601 *Embryol. Exp. Morphol* **42**, 293–303 (1977).
- 602 49. Murga, M. *et al.* A mouse model of ATR-Seckel shows embryonic replicative stress
603 and accelerated aging. *Nat Genet* **41**, 891-8 (2009).
- 604 50. Despras, E., Daboussi, F., Hyrien, O., Marheineke, K. & Kannouche, P.L.
605 ATR/Chk1 pathway is essential for resumption of DNA synthesis and cell survival in
606 UV-irradiated XP variant cells. *Hum Mol Genet* **19**, 1690-701 (2010).
- 607 51. Kawabata, T. *et al.* Stalled fork rescue via dormant replication origins in
608 unchallenged S phase promotes proper chromosome segregation and tumor
609 suppression. *Mol Cell* **41**, 543-53 (2011).
- 610 52. Kumagai, A., Lee, J., Yoo, H.Y. & Dunphy, W.G. TopBP1 activates the ATR-ATRIP
611 complex. *Cell* **124**, 943-55 (2006).
- 612 53. Bass, T.E. *et al.* ETAA1 acts at stalled replication forks to maintain genome integrity.
613 *Nat Cell Biol* **18**, 1185-1195 (2016).
- 614 54. Haahr, P. *et al.* Activation of the ATR kinase by the RPA-binding protein ETAA1. *Nat*
615 *Cell Biol* **18**, 1196-1207 (2016).

- 616 55. Duursma, A.M., Driscoll, R., Elias, J.E. & Cimprich, K.A. A role for the MRN complex
617 in ATR activation via TOPBP1 recruitment. *Mol Cell* **50**, 116-22 (2013).

618 **Figure Legends:**

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

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

634

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

636 **(a-b)** *DONSON* mutations result in severely reduced levels of *DONSON* protein.
637 Immunoblotting of cell extracts from lymphoblastoid (a) and fibroblast (b) cell lines derived
638 from patients with mutations in *DONSON*. ATR was used as a loading control. The two
639 blots from (a) originate from two independent gels. **(c)** The K489T, but not the S28R
640 variant, associated with the *DONSON* haplotype affects protein levels. Cells were treated
641 with doxycycline 48 h post siRNA transfection, and harvested for Western blot analysis 24
642 h later (n=2). Exogenous *DONSON* were detected using an anti-GFP antibody

643 respectively. TOPBP1 was used as a loading control. Depletion of endogenous DONSON
644 in these cells was confirmed by immunoblotting (**Supplementary Fig. 2**).

645

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

648 **(a)** DONSON protein levels are increased during S-phase. HeLa cells were synchronised
649 in S-phase using a double thymidine block, released, harvested at the indicated time
650 points, and immunoblotting was performed (n=2). Cyclin A and phospho-histone H3 Ser-
651 10 are markers of S/G2 and M phase respectively. Vinculin represents a loading control.

652 **(b)** S-phase is prolonged upon DONSON depletion. HeLa cells transfected with the
653 indicated siRNAs were pulsed with BrdU, fixed and analysed by FACS (n=4; error bars
654 indicate SD). **(c-e)** Replication fork analysis of HeLa cells transfected with control or
655 DONSON siRNA and pulsed with CldU and IdU. **(c)** Top: Schematic of DNA fibre analysis.

656 Bottom: loss of DONSON does not decrease replication fork velocity. Replication fork
657 speed (kb/min) was determined (n=5). **(d)** DONSON depletion results in spontaneous fork
658 stalling. Percentages of ongoing replication forks, new origins and stalled replication forks
659 in cells from (c) were quantified (n=3). **(e)** DONSON depletion leads to replication fork

660 asymmetry. Top: example images; magenta arrows indicate origins of replication; white
661 arrow denotes fork asymmetry. Bottom: plot indicates the ratio of left/right fork track
662 lengths of bidirectional replication forks in cells from (c). Red lines denote median ratios
663 (n=3). **(f)** Loss of DONSON increases spontaneous γ H2AX/53BP1 foci formation. HeLa

664 cells transfected with the indicated siRNAs were immunostained with antibodies to 53BP1
665 and γ H2AX (left panel), and the percentage of cells with >10 53BP1 and γ H2AX foci were
666 quantified using fluorescence microscopy (right panel; n=5; >300 cells per sample per
667 independent experiment). Scale bar; 10 μ m.

668

669 **Figure 4. DONSON localizes to the replication fork.**

670 **(a-d)** DONSON interacts with multiple components of the replication machinery. **(a)** GFP
671 or GFP-DONSON was precipitated by GFP-Trap, from asynchronous cells or cells
672 accumulated in S-phase with 2 mM HU treatment for 24 h. Heatmap denotes significant
673 interactions identified by mass spectrometry (n=3). Inset: Schematic of the mammalian
674 replisome with selected replication factors. **(b)** 293FT cells were transfected with the
675 indicated expression vectors in the presence/absence of HU. GFP or GFP-DONSON were
676 isolated by GFP-Trap and co-precipitating proteins visualised by immunoblotting (n=2).
677 Benzonase Nuclease was included to exclude DNA-mediated interactions. The bottom two
678 panels are scanned images of Ponceau S-stained nitrocellulose membrane. **(c-d)**
679 DONSON localises in close proximity to replication forks. **(c-d)** PLA was carried out on
680 cells from (a) using the indicated antibodies in the presence/absence of HU (n=2). **(c)**
681 Quantification of PLA signals. **(d)** Representative PLA images. **(e-f)** DONSON interacts
682 with PCNA at replication foci in live cells. **(e)** Representative confocal images of live cells
683 expressing GFP-DONSON and RFP-PCNA. Boxes indicate representative regions used
684 for FCCS analysis. **(f)** FCCS measurements of GFP-DONSON and RFP-PCNA reveal
685 significant cross-correlation at replication foci at similar concentrations. Average cross-
686 correlation curves are shown from cells expressing GFP-DONSON in replication foci (red)
687 or non-replicating (grey) cells, or GFP-expressing S-phase nuclei (purple). Inset: Mean
688 cross-correlation amplitude values from multiple cells (error bars indicate SD; n=4, 3 and
689 5). Increased $G(\tau)$ values indicate higher degree of cross-correlation between GFP-
690 DONSON and RFP-PCNA in replication foci. See also **Supplementary Fig. 11.** **(g)** iPOND
691 was performed on 293T (n=3), HeLa (n=2) and HCT116 (n=2) cells, and EdU-
692 coprecipitates analysed by mass spectrometry. Data represents the combination of all
693 seven experiments. Log₂ abundance denotes the ratio of proteins at nascent DNA

694 compared to mature chromatin. Values >0 represent proteins enriched at forks, whilst
695 values ≤ 0 denote chromatin-bound factors. Scale bars; 10 μm .

696

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

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

714

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

717 **(a,b)** Metaphases chromosomes from DONSON or control siRNA transfected HeLa cells
718 were visualised by Giemsa staining and light microscopy. **(a)** Quantification of average
719 numbers of chromatid gaps/breaks per metaphase ($n=6$; >50 metaphases per sample per

720 experiment). **(b)** Representative images of normal chromosomes, chromosomes
721 containing gaps/breaks, highly fragmented and pulverized chromosomes. Red arrows
722 denote chromatid gaps/breaks; blue arrows indicate chromosomal exchanges. Scale bar;
723 10 μ m. **(c-g)** Loss of the structure-specific nucleases MUS81 or XPF significantly reduces
724 the spontaneous replication fork asymmetry and genome instability in DONSON-depleted
725 cells. **(c)** Cells transfected with the indicated siRNAs were pulsed with CldU and IdU.
726 Replication fork asymmetry was measured as in **(Fig. 3e)**. The red lines denotes median
727 ratios (n=3). **(d)** Co-depletion of MUS81 or XPF with DONSON reduces levels of
728 spontaneous DNA damage. Extracts from cells transfected with the indicated siRNAs were
729 subjected to SDS-PAGE and immunoblotting using the antibodies indicated. **(e-f)** Co-
730 depletion of MUS81 (e) or XPF (f) reduces chromosomal aberrations in cells lacking
731 DONSON. Quantification of the average number of chromatid gaps/breaks per metaphase
732 in cells transfected with control, DONSON, MUS81 and/or XPF siRNA. At least 50
733 metaphases per experiment were counted (n=3). **(g)** Quantification of the average
734 percentage of metaphases containing highly fragmented chromosomes or pulverized
735 chromosomes in cells transfected with the indicated siRNAs. At least 50 metaphases per
736 experiment were counted (n=3).

737

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

740 **(a)** Complementation of patient-derived fibroblasts with WT *DONSON*. Fibroblasts derived
741 from DONSON patients P2, P6 and P9 were infected with retroviruses encoding either WT
742 *DONSON* or an empty vector. DONSON expression was determined by immunoblotting. A
743 non-specific cross-reactive protein represents a loading control. **(b)** Expression of WT
744 *DONSON* in patient fibroblasts rescues elevated levels of spontaneous DNA damage. The
745 percentage of cells from (a) with 53BP1/ γ H2AX foci was quantified by immunostaining

746 (n=3). **(c)** DNA fibre analysis of complemented DONSON patient fibroblasts pulsed with
747 CldU and IdU. Fork asymmetry was quantified. Plot indicates ratios of left/right fork track
748 lengths of bidirectional replication forks. The red lines denote median ratios. (n=3). **(d)** The
749 percentage of stalled forks and new origins from cells in (c) was quantified (n=3). Ongoing
750 forks are shown in **(Supplementary Fig. 19)**.

Table 1: Biallelic *DONSON* mutations identified in 29 individuals

Patient	Country of Origin	Mutation 1	Mutation 2	Segregation
P1-1	Italy	c.1251_1256delCTCTAA, p.Asn417_Ser418del	<i>haplotype</i>	Het, M
P1-2	Italy	c.1251_1256delCTCTAA, p.Asn417_Ser418del	<i>haplotype</i>	Het, M
P2	UK	c.877C>T, p.Arg293*	<i>haplotype</i>	Het, M, P
P3	UK	c.1254dupT, p.Lys419*	<i>haplotype</i>	Het, M, P
P4	UK	c.1686dupT, p.Asn563*	<i>haplotype</i>	Het, nps
P5	Somalia	c.832T>C, p.Cys278Arg AND/OR c.845A>G, p.Tyr282Cys	<i>haplotype</i>	Het, M, P
P6	USA	c.1282C>T, p.Gln428*	<i>haplotype</i>	Het, M, P
P7	USA	c.1282C>T, p.Gln428*	<i>haplotype</i>	Het, nps
P8	Italy	c.1474_1475delCA, p.Gln492Glufs*18	c.786-7T>C	Het, M, P
P9	Turkey	c.876C>G, p.Phe292Leu	c.1047-9A>G (SS)	Het, M
P10-1	India	c.1628_1630dupAAA, p.Gln543_Ile544insLys	c.1032C>T, p.Ser344Ser (SS)	Het, M, P
P10-2	India	c.1628_1630dupAAA, p.Gln543_Ile544insLys	c.1032C>T, p.Ser344Ser (SS)	Het, M, P
P11	Saudi Arabia	c.786-22A>G (SS)	c.786-22A>G (SS)	Hom, M, P
P12	Saudi Arabia	c.786-22A>G (SS)	c.786-22A>G (SS)	Hom, nps
P13-1	Palestine	c.1337T>C, p.Met446Thr	c.1337T>C, p.Met446Thr	Hom, M, P
P13-2	Palestine	c.1337T>C, p.Met446Thr	c.1337T>C, p.Met446Thr	Hom, M, P
P13-3	Palestine	c.1337T>C, p.Met446Thr	c.1337T>C, p.Met446Thr	Hom, M, P
P14	Saudi Arabia	c.786-22A>G (SS)	c.786-22A>G (SS)	Hom, M, P
P15	Saudi Arabia	c.786-22A>G (SS)	c.786-22A>G (SS)	Hom, M, P
P16	Saudi Arabia	c.786-22A>G (SS)	c.786-22A>G (SS)	Hom, M, P
P17	Saudi Arabia	c.786-22A>G (SS)	c.786-22A>G (SS)	Hom, M, P
P18-1	Saudi Arabia	c.786-22A>G (SS)	c.786-22A>G (SS)	Hom, M, P
P18-2	Saudi Arabia	c.786-22A>G (SS)	c.786-22A>G (SS)	Hom, M, P
P18-3	Saudi Arabia	c.786-22A>G (SS)	c.786-22A>G (SS)	Hom, M, P
P19	Turkey	c.1297C>T, p.Pro433Ser	c.1297C>T, p.Pro433Ser	Hom, M, P
P20-1	South Africa	c.1254dupT, p.Lys419*	c.1510G>A, p.Glu504Lys	Het, M, P
P20-2	South Africa	c.1254dupT, p.Lys419*	c.1510G>A, p.Glu504Lys	Het, M, P
P21-1	Saudi Arabia	c.1047-9A>G (SS)	c.1047-9A>G (SS)	Hom, M, P
P21-2	Saudi Arabia	c.1047-9A>G (SS)	c.1047-9A>G (SS)	Hom, M, P

'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.

759 **Online Methods**

760 **Research subjects.**

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

774

775 **Exome sequencing and haplotype analysis.**

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

784

785 **Cell culture and generation of cell lines.**

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

795

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

803

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

809

810 **Cell treatments.**

811 Plasmids and siRNA oligos were transfected in Opti-MEM reduced serum medium using
812 Lipofectamine 2000 and Oligofectamine (Life Technologies) respectively according to the
813 manufacturer's guidelines. A custom siRNA targeting lacZ has previously been described⁵⁶,
814 and was used as a control siRNA. A DONSON siRNA SMARTpool (Dharmacon) was used
815 for all siRNA transfections except when transfecting the HeLa Flp-In/T-Rex cells
816 expressing an exogenous, siRNA-resistant, GFP-tagged DONSON construct. In this case
817 a custom DONSON siRNA sequence (CCTGTGGACTGGAGTATTAdTdT) was used
818 (Dharmacon). MUS81 siRNA SMARTpool and XPF siRNA SMARTpool (Dharmacon) were
819 used where indicated. Transfected cells in both cases were analysed at 48-72 h post
820 transfection. Where indicated, cells were treated with 1-2 mM hydroxyurea (Sigma-Aldrich),
821 50 ng/ml mitomycin C (Sigma-Aldrich) or 2 mM thymidine (Sigma-Aldrich). The ATR
822 inhibitor (VE-821; Selleck Chemicals) was used at 5 µM. dNTP analogues EdU, CldU and
823 IdU were purchased from Sigma Aldrich, and were used as indicated.

824

825 **RT-PCR.**

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

830

831 **DNA expression constructs**

832 pEGFP-DONSON expression construct was created by cloning the human *DONSON* ORF
833 into the pDONR221 Gateway shuttle vector (Invitrogen). WT *DONSON* was amplified
834 from cDNA and recombined into the pDEST-EGFP vector to generate a GFP-tagged
835 DONSON expression construct. The *DONSON* ORF was made siRNA resistant using site-

836 directed mutagenesis (Agilent Technologies) by altering the following nucleotides:
837 CCIGTGGACTGGAGTATTA was changed to CCCGTAGATTGGTCTATCA. Patient-
838 associated mutations were engineered into the pEGFP-DONSON expression plasmid
839 using site-directed mutagenesis according to the manufacturer's instructions. (All primers
840 are detailed in **Supplementary Table 4**).

841

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

845

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

849

850 **Minigene splicing reporter assay**

851 A 1.58 kb stretch of the *DONSON* gene encompassing the 3' end of intron 3, exon 4,
852 intron 4, exon 5 and the 5' end of intron 5 was amplified using DNA from a healthy
853 individual and *DONSON* patients (carrying mutation c.786-22A>G or c.786-33A>G) using
854 the DONSON-int3-Sall-F and DONSON-int5-SpeI-R primers, and cloned into the RHCglo
855 vector⁵⁷ using the Sall and SpeI restriction sites. Site-directed mutagenesis was used to
856 introduce the DONSON intron 4 splice acceptor mutation (c.786-1G>A) into the splicing
857 reporter construct. HeLa cells were transfected with each individual splicing mutation
858 reporter construct using Lipofectamine 2000 according the manufacturer's instructions. 24
859 h post-transfection, cells were harvested, total cellular RNA was extracted and cDNA
860 generated using Superscript III reverse transcriptase first-stand synthesis system
861 (Invitrogen). PCR was carried out using primers (RSV_minigene_F and RSV_minigene_R)

862 to the 5' and 3' ends of the artificial exons present in the RHCglo vector. *DONSON* WT
863 and mutant cDNA amplicons were resolved on a 2% agarose gel to visualise differences in
864 splicing. Individual PCR products were subsequently cloned into the pGEM-T Easy Vector
865 (Promega) and sequenced to verify the exon content of each transcript. All relevant
866 primers are detailed in **Supplementary Table 4**.

867

868 **iPOND**

869 iPOND was performed as previously described^{35,58}. Briefly, exponentially growing cells
870 were incubated with 10 μ M EdU for 10 min, cross-linked with 1% formaldehyde, harvested
871 and permeabilised. For pulse-chase controls, cells were incubated in 10 μ M EdU for 10
872 min, washed in media containing 10 μ M thymidine, then incubated with media containing
873 10 μ M thymidine for 1 h, before being cross-linked. Biotin azide was covalently attached to
874 EdU within newly replicated DNA using a Click reaction, and EdU containing DNA was
875 precipitated using Streptavidin agarose beads. Edu co-precipitates were then analysed by
876 mass spectrometry. Log₂ abundance values represent the ratio of proteins found in EdU-
877 pulsed samples compared to those pulse-chased with EdU-thymidine.

878

879 **Immunoblot analysis and antibodies**

880 Whole cell extracts were obtained by sonication in UTB buffer (8 M Urea, 50 mM Tris, 150
881 mM β -mercaptoethanol) and analysed by SDS-PAGE following standard procedures.
882 Protein samples were run on 6-12% acrylamide SDS-PAGE or 4-12% NuPage mini-gels
883 (Life Technologies) and transferred onto nitrocellulose membrane. Immunoblotting was
884 performed using antibodies to: Cyclin A (Santa Cruz, sc-751; 1:1,000), CHK1 (Santa Cruz,
885 sc-8408; 1:1,000), CHK2 (Santa Cruz, sc-5278; 1:1000), FANCD2 (Santa Cruz, sc-20022;
886 1:1000), MCM2 (BD Transduction Laboratories, 610700; 1:10000), MCM7 (Santa Cruz,
887 sc-56324; 1:1000), MUS81 (Santa Cruz, sc-53382; 1:2000); XPF (Santa Cruz, sc-136153;

888 1:1000); H2A (Millipore, 07-146; 1:3000), γ -H2AX (Millipore, 05-636; 1:3000), RPA2
889 (Millipore, NA18; 1:1000), phospho-histone H3 Ser-10-P (Millipore); pS343-NBS1 (Abcam,
890 47272; 1:500); NBS1 (Genetex, GTX70224; 1:10000); ATR (Bethyl Laboratories, A300-
891 137A; 1:1000), pS345-CHK1 (Cell Signaling Technology, 2341; 1:100), pS4/S8-RPA2
892 (Bethyl Laboratories, A300-245A; 1:1,000), pS966-SMC1 (Bethyl Laboratories, A300-
893 050A; 1:1,000), SMC1 (Bethyl Laboratories, A300-055A; 1:1,000), Treslin (Bethyl
894 Laboratories, A303-472A; 1:1,000); TOPBP1 (Bethyl Laboratories; A300-111A; 1:1000);
895 Vinculin (Sigma-Aldrich, V9264; 1:1,000); α -Tubulin (Sigma-Aldrich, T5168; 1:4000); GFP
896 (Roche, 11814460001; 1:500). The polyclonal anti-DONSON antibody was generated by
897 immunising rabbits with a GST-fusion protein encoding aa 1-125 of human DONSON.
898 Antibody was affinity-purified from rabbit sera (Eurogentec) and specificity established
899 using lysates from patient cells and DONSON siRNA-transfected cells.

900

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

904

905 **Immunofluorescence and fluorescent microscopy.**

906 siRNA transfected HeLa cells or passage-matched *TERT*-immortalized fibroblasts were
907 seeded on coverslips 24 h before extraction/fixation. To visualise cells undergoing DNA
908 replication, cells incubated in medium containing 10 μ M EdU for 10-30 min prior to
909 harvesting. To remove soluble proteins before immunofluorescence, cells were pre-
910 extracted for 10 min on ice with ice-cold buffer (25 mM HEPES, pH 7.4, 50 mM NaCl, 1
911 mM EDTA, 3 mM MgCl₂, 300 mM sucrose and 0.5% Triton X-100) and then fixed with 4%
912 paraformaldehyde for 15 min. For analysis of cells transfected with GFP-tagged protein,
913 cells were fixed and permeabilised by incubation with ice-cold methanol for 20 minutes.

914

915 EdU immunolabeling was performed using Click-iT EdU Imaging Kit (Invitrogen, C10337)
916 according to the manufacturer's protocol. Cells were stained for 53BP1 (Novus Biologicals,
917 NB100-304; 1:1,000) and/or γ H2AX (Millipore, 05-636; 1:1000) and stained with secondary
918 antibodies conjugated to Alexa Fluor-568 (Life Technologies) and DAPI.

919

920 For quantification of signal-integrated densities of γ H2AX staining, cells were stained with
921 an antibody specific to γ H2AX (Millipore, 05-636; 1:1000), images were visualized using a
922 Zeiss Axioplan 2 microscope with iVision software (BioVision Technologies) and captured
923 using a 40 \times oil-immersion objective. For quantification of signal-integrated densities of
924 RPA staining, cells were stained with RPA2 antibody (Millipore, NA18; 1:200), images
925 were visualized using a Nikon Eclipse Ni microscope with NIS-Elements software (Nikon
926 Instruments) and captured using a 100 \times oil-immersion objective. Nuclei were segmented
927 on the basis of DAPI staining and then signal-integrated density of γ H2AX or RPA staining
928 quantified for each nuclear region using ImageJ software (US National Institutes of Health).
929 For quantification of γ H2AX staining, more than 100 EdU positive cells and 50 EdU
930 negative cells were analyzed per experiment per condition, and for quantification of RPA
931 staining, more than 200 cells were analyzed per experiment per condition. Exposure time,
932 binning, microscope settings and light source intensity were kept constant for all the
933 samples in all cases.

934

935 For quantification of native BrdU foci cells were incubated in medium containing 10 μ M
936 BrdU for 24 h prior to harvesting. Six hours prior to harvesting, 2 mM HU was added to the
937 media. To visualise ssDNA foci, cells were extracted for 10 min on ice with ice-cold buffer
938 (25 mM HEPES, pH 7.4, 50 mM NaCl, 1 mM EDTA, 3 mM MgCl₂, 300 mM sucrose and
939 0.5% Triton X-100) and then fixed with 4% paraformaldehyde for 15 min. After fixation,

940 cells were washed with PBS and blocked in 3% FCS in PBS for 30 min at room
941 temperature. ssDNA was visualised using a BrdU antibody (Abcam, ab6326; 1:500). To
942 denature DNA cells were incubated in 2 M HCl in PBS for 30 min prior to addition of the
943 BrdU antibody. Images were acquired as for γ H2AX quantification and foci were quantified
944 using ImageJ-based script. Nuclei were defined on the basis of DAPI staining and native
945 BrdU foci were detected using "Find maxima" function of ImageJ within each nuclear
946 region. Exposure time, binning, microscope settings, light source intensity and the noise
947 level in "Find maxima" function were kept constant for all the samples within each
948 individual experiment. More than 100 cells were analyzed per experiment per condition.

949

950 **Metaphase spreads**

951 Chromosomal aberrations were scored in Giemsa-stained metaphase spreads as
952 previously described⁵⁶. Briefly, demecolcine (Sigma Aldrich) was added at a final
953 concentration of 0.2 μ g/ml 4 h prior to harvesting. Cells were harvested by trypsinisation,
954 subjected to hypotonic shock for 1 hour at 37°C in 0.3 M sodium citrate and fixed in 3:1
955 methanol:acetic acid solution. Cells were dropped onto acetic acid humidified slides,
956 stained for 15 minutes in Giemsa-modified solution (Sigma; 5% v/v in H₂O) and washed in
957 water for 5 minutes.

958

959 **DNA fibre spreading assay**

960 Passage-matched primary, *TERT*-immortalized fibroblasts or siRNA transfected HeLa cells
961 were pulse labeled with CldU for 20 min, washed with media and damaged with 2 mM
962 hydroxyurea for 2 h before being pulse labeled with IdU for 40 min. Alternatively, 50 ng/ml
963 mitomycin C was added to the cells 24 h before CldU pulse labeling and left on during 20
964 min CldU and 20 min IdU pulse labeling. Cells were harvested by trypsinization, and cell
965 pellets were washed in PBS. 5×10^5 cells were lysed directly onto glass slides using

966 spreading buffer (200 mM Tris-HCl, pH 7.5, 50 mM EDTA, 0.5% SDS) and fixed in
967 methanol/acetic acid (3:1 ratio). Following 2.5 M HCl denaturation, CldU was detected
968 using rat anti-BrdU (clone BU1/75, ICR1; Abcam, ab6326; 1:750), and IdU was detected
969 using mouse anti-BrdU (clone B44; BD Biosciences, 347583; 1:750). Slides were then
970 fixed in 4% paraformaldehyde before immunostaining with secondary antibodies
971 conjugated to Alexa Fluor-594 or Alexa Fluor-488 (Life Technologies). Labeled DNA fibers
972 were visualized using a Nikon Eclipse Ni microscope with NIS-Elements software (Nikon
973 Instruments). Images were captured using 40× oil-immersion objectives and were
974 recoloured and analyzed using ImageJ software (US National Institutes of Health). For
975 quantification of replication structures, at least 400 structures were counted per experiment.
976 Tract lengths were measured using ImageJ (National Institutes of Health;
977 <http://rsbweb.nih.gov/ij/>). To calculate fork velocity, arbitrary length values were converted
978 into micrometers using the scale bars created by the microscope, with 1 μm equivalent to
979 2.59 kb⁵⁹. Replication fork speed (kb/min) was then determined by dividing the length of
980 CldU and IdU tracks (in kb) from ongoing forks by the pulse time.

981

982 **FACS analysis**

983 For BrdU analysis, HeLa cells were pulse labeled with 10 μM BrdU for 30 min before
984 fixation with 70% ethanol at -20 °C for 16 h. Cells were then digested with 1 mg/ml pepsin
985 and denatured with 2 M HCl before washing with PBS and blocking in 0.5 % BSA, 0.5 %
986 Tween-20. BrdU labeling was detected using anti-BrdU antibody (Abcam, ab6326; 1:75)
987 and FITC-conjugated anti-rat secondary antibody. DNA content was assessed by staining
988 with 50 μg/ml propidium iodide. Cells were sorted on a BD Biosciences FACS Aria II and
989 data were analyzed using FlowJo software (v7.6.1, Tree Star).

990

991 For mitotic analysis and immuno-detection of phospho-histone H3 (Ser10), HeLa cells
992 were harvested, fixed, permeabilised 24 h post exposure to HU or MMC, as previously
993 described⁵⁶. Cells were analysed using an Accuri flow cytometer (BDBiosciences) in
994 conjunction with CFlowplus software. Data represents that obtained from at least 30,000
995 cells.

996

997 **Immunoprecipitation and GFP-Trap**

998 293FT cells transfected with plasmids encoding GFP-DONSON or GFP were untreated, or
999 exposed to 2 mM HU for 16 h and harvested. Cells were then incubated in lysis buffer (150
1000 mM NaCl, 50 mM Tris HCl pH7.5, 2 mM MgCl₂, 1 % NP40, 90 U/ml Benzonase (Novagen)
1001 and Protease Inhibitor Cocktail EDTA free (Roche)) for 30 min with rotation at 4 °C. The
1002 resultant cell lysates were pre-cleared at 44,000 rpm at 4 °C for 30 min.

1003

1004 For immunoprecipitations, 3 mg of lysate was immunoprecipitated with 5 µg of antibody,
1005 immune complexes collected with Protein A-Sepharose (Sigma-Aldrich). Complexes were
1006 washed with wash buffer (150 mM NaCl, 50 mM Tris HCl pH 7.5, 0.5 % NP40, and
1007 Complete Protease Inhibitor Cocktail (Roche)) and analysed by SDS-PAGE.

1008

1009 For GFP-Trap, 3 mg lysates were incubated with GFP-Trap agarose beads (ChromoTek)
1010 at 4 °C for 5 h. The resulting GFP-Trap complexes were washed with wash buffer as
1011 above and analysed by SDS-PAGE. Experiments were carried out in the presence of
1012 Benzonase Nuclease to exclude the possibility of interactions being mediated by DNA.

1013

1014 For mass spectrometry analysis, GFP or GFP-DONSON were isolated from tetracycline
1015 induced, or uninduced, Flp-In T-REx HeLa cell extracts by incubation with GFP-trap
1016 magnetic agarose beads (Chromotek) for 2 hours on a Kingfisher Duo robotic handling

1017 station (Thermo). Asynchronous cells and S-phase accumulated cells, using a 24 h
1018 treatment with 2 mM HU, were analysed. On-bead digest and mass spectrometry were
1019 performed as described⁶⁰. Data represents three independent experiments for each
1020 condition, analysed by back-to-back MS and quantified by Label free quantification (LFQ).
1021 Proteins were identified and quantified with the MaxQuant 1.5 software suite by searching
1022 against the Uniprot human database. M(ox) and protein N-terminal acetylation were set as
1023 variable, and carbamylation as a fixed modification, with a 1 % FDR. Contaminants and
1024 reverse data base hits were deleted. Protein significantly enriched by GFP-DONSON were
1025 selected on the basis of p-value <0.05, and >2 fold change from asynchronous to S-phase,
1026 as identified by Student t-test and ratio cut-off against the respective negative control LFQ
1027 data as determined by MaxQuant (p<0.05; 2-fold).

1028

1029 **Proximity ligation assay (PLA).**

1030 PLA was carried out as described in^{5,56}. Briefly, cells from GFP or GFP-DONSON Flp-In T-
1031 REx HeLa cell lines were treated with 1ug/ml doxycycline and fixed/extracted after 24 h.
1032 For PCNA visualisation, cells were fixed with methanol at -20 °C for 10 min followed by a 5
1033 min extraction in 0.3% Triton-X100 in PBS. For RPA visualisation, cells were pre-extracted
1034 in nuclear extraction buffer for 5 minutes on ice, and fixed in 3.6% paraformaldehyde for
1035 10 minutes at room temperature. Cells were then incubated in anti-PCNA (PC10, 1:500,
1036 Santa Cruz) or anti-RPA (NA18; 1:500; Merck-Millipore) antibodies along with anti-GFP
1037 antibody (ab6556, 1:500, Abcam), and *in situ* proximity ligation was performed using a
1038 Duolink Detection Kit (Sigma Aldrich). Nuclear foci were imaged using a Nikon Eclipse Ni-
1039 U microscope equipped with a 100X oil lens in conjunction with a Zyla camera, and
1040 images were acquired using Elements software (Nikon). More than 200 cells were
1041 analysed per experiment per condition.

1042

1043 **Fluorescence Cross-Correlation Spectroscopy (FCCS)**

1044 HeLa cells stably expressing GFP-DONSON and mCherry-PCNA (construct kindly
1045 provided by C. Lukas, Copenhagen; referred to as RFP-PCNA) were used for FCCS. For
1046 all details on Fluorescence Microscopy Imaging and FCS/FCCS, refer to **Supplementary**
1047 **Note.**

1048

1049 **Statistical Analyses**

1050 Statistical differences were analyzed by: two-tailed Student T-Test (**Fig. 3b, 3d, 3f, 4f, 4g,**
1051 **5b-d, 5f, 6a, 6e, 6f, 7b, 7d** and **Supplementary Fig. 5c, 8b, 9a, 12c, 13i, 14c, 16a-c, 17a-**
1052 **c, 19, 20, 21**); Mann-Whitney rank sum test (**Fig. 3e, 6c, 7c** and **Supplementary Fig. 12b,**
1053 **15a, 15c**); and Chi-Squared Test (**Fig 4c, 6g** and **Supplementary Fig. 16d**). n refers to
1054 number of independent experiments unless indicated. Error bars represent standard error
1055 of the mean (s.e.m.) unless specified.

1056

1057 **Data Availability**

1058 The NGS data used in the manuscript can be obtained from the European Genome-
1059 phenome Archive (EGA) under accession EGAS00001002224. NGS datasets on patients
1060 P14-P18, p21 are not available due to institutional IRB restrictions. The mass spectrometry
1061 proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE
1062 partner repository with the dataset identifier PXD005690.

1063 **Methods-only references**

- 1064 56. Higgs, M.R. *et al.* BOD1L Is Required to Suppress Deleterious Resection of
1065 Stressed Replication Forks. *Mol Cell* **59**, 462-77 (2015).
- 1066 57. Singh, G. & Cooper, T.A. Minigene reporter for identification and analysis of cis
1067 elements and trans factors affecting pre-mRNA splicing. *Biotechniques* **41**, 177-81
1068 (2006).
- 1069 58. Sirbu, B.M., Couch, F.B. & Cortez, D. Monitoring the spatiotemporal dynamics of
1070 proteins at replication forks and in assembled chromatin using isolation of proteins
1071 on nascent DNA. *Nat Protoc* **7**, 594-605 (2012).
- 1072 59. Conti, C. *et al.* Replication fork velocities at adjacent replication origins are
1073 coordinately modified during DNA replication in human cells. *Mol Biol Cell* **18**, 3059-
1074 67 (2007).
- 1075 60. Turriziani, B. *et al.* On-beads digestion in conjunction with data-dependent mass
1076 spectrometry: a shortcut to quantitative and dynamic interaction proteomics. *Biology*
1077 (*Basel*) **3**, 320-32 (2014).

Table 1: Biallelic DONSON mutations identified in 29 individuals

Patient	Country of Origin	Mutation 1	Mutation 2	Segregation
P1-1	Italy	c.1251_1256delCTCTAA, p.Asn417_Ser418del	<i>haplotype</i>	Het, M
P1-2	Italy	c.1251_1256delCTCTAA, p.Asn417_Ser418del	<i>haplotype</i>	Het, M
P2	UK	c.877C>T, p.Arg293*	<i>haplotype</i>	Het, M, P
P3	UK	c.1254dupT, p.Lys419*	<i>haplotype</i>	Het, M, P
P4	UK	c.1686dupT, p.Asn563*	<i>haplotype</i>	Het, nps
P5	Somalia	c.832T>C, p.Cys278Arg AND/OR c.845A>G, p.Tyr282Cys	<i>haplotype</i>	Het, M, P
P6	USA	c.1282C>T, p.Gln428*	<i>haplotype</i>	Het, M, P
P7	USA	c.1282C>T, p.Gln428*	<i>haplotype</i>	Het, nps
P8	Italy	c.1474_1475delCA, p.Gln492Glufs*18	c.786-7T>C	Het, M, P
P9	Turkey	c.876C>G, p.Phe292Leu	c.1047-9A>G (SS)	Het, M
P10-1	India	c.1628_1630dupAAA, p.Gln543_Ile544insLys	c.1032C>T, p.Ser344Ser (SS)	Het, M, P
P10-2	India	c.1628_1630dupAAA, p.Gln543_Ile544insLys	c.1032C>T, p.Ser344Ser (SS)	Het, M, P
P11	Saudi Arabia	c.786-22A>G (SS)	c.786-22A>G (SS)	Hom, M, P
P12	Saudi Arabia	c.786-22A>G (SS)	c.786-22A>G (SS)	Hom, nps
P13-1	Palestine	c.1337T>C, p.Met446Thr	c.1337T>C, p.Met446Thr	Hom, M, P
P13-2	Palestine	c.1337T>C, p.Met446Thr	c.1337T>C, p.Met446Thr	Hom, M, P
P13-3	Palestine	c.1337T>C, p.Met446Thr	c.1337T>C, p.Met446Thr	Hom, M, P
P14	Saudi Arabia	c.786-22A>G (SS)	c.786-22A>G (SS)	Hom, M, P
P15	Saudi Arabia	c.786-22A>G (SS)	c.786-22A>G (SS)	Hom, M, P
P16	Saudi Arabia	c.786-22A>G (SS)	c.786-22A>G (SS)	Hom, M, P
P17	Saudi Arabia	c.786-22A>G (SS)	c.786-22A>G (SS)	Hom, M, P
P18-1	Saudi Arabia	c.786-22A>G (SS)	c.786-22A>G (SS)	Hom, M, P
P18-2	Saudi Arabia	c.786-22A>G (SS)	c.786-22A>G (SS)	Hom, M, P
P18-3	Saudi Arabia	c.786-22A>G (SS)	c.786-22A>G (SS)	Hom, M, P
P19	Turkey	c.1297C>T, p.Pro433Ser	c.1297C>T, p.Pro433Ser	Hom, M, P
P20-1	South Africa	c.1254dupT, p.Lys419*	c.1510G>A, p.Glu504Lys	Het, M, P
P20-2	South Africa	c.1254dupT, p.Lys419*	c.1510G>A, p.Glu504Lys	Het, M, P
P21-1	Saudi Arabia	c.1047-9A>G (SS)	c.1047-9A>G (SS)	Hom, M, P
P21-2	Saudi Arabia	c.1047-9A>G (SS)	c.1047-9A>G (SS)	Hom, M, P

'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.

Figure 1

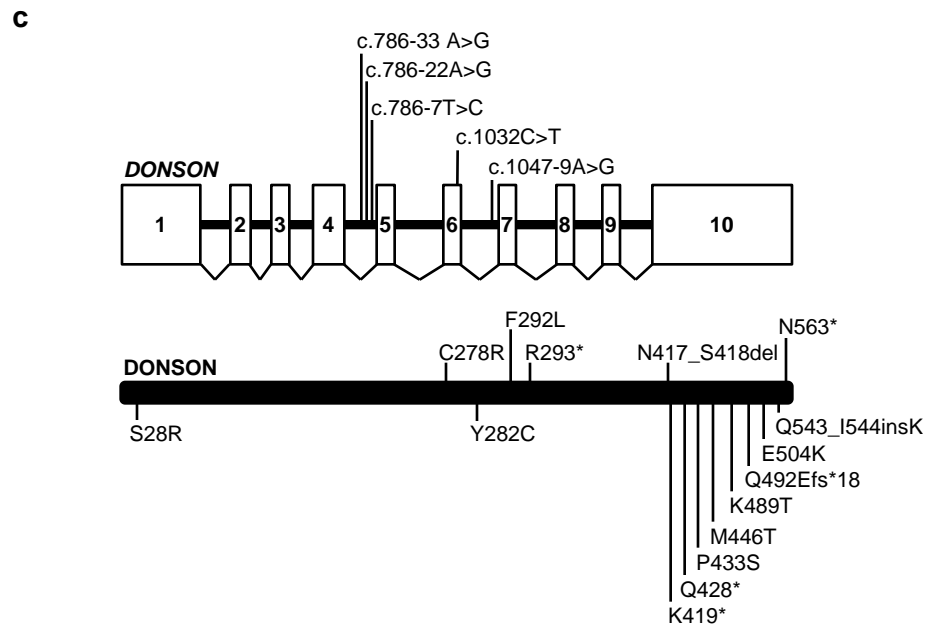
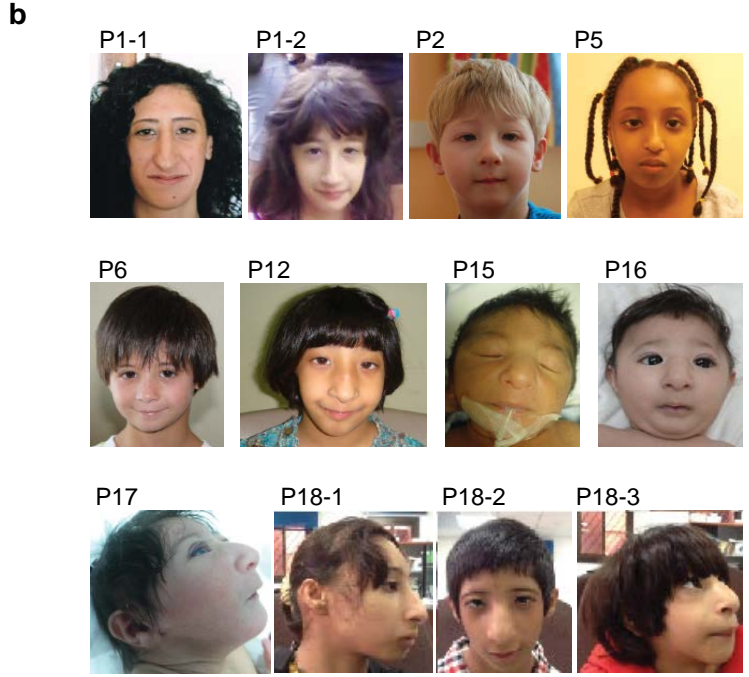
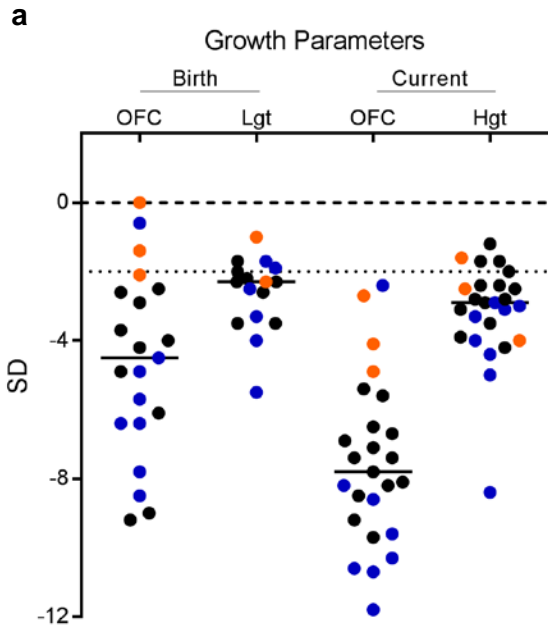


Figure 2

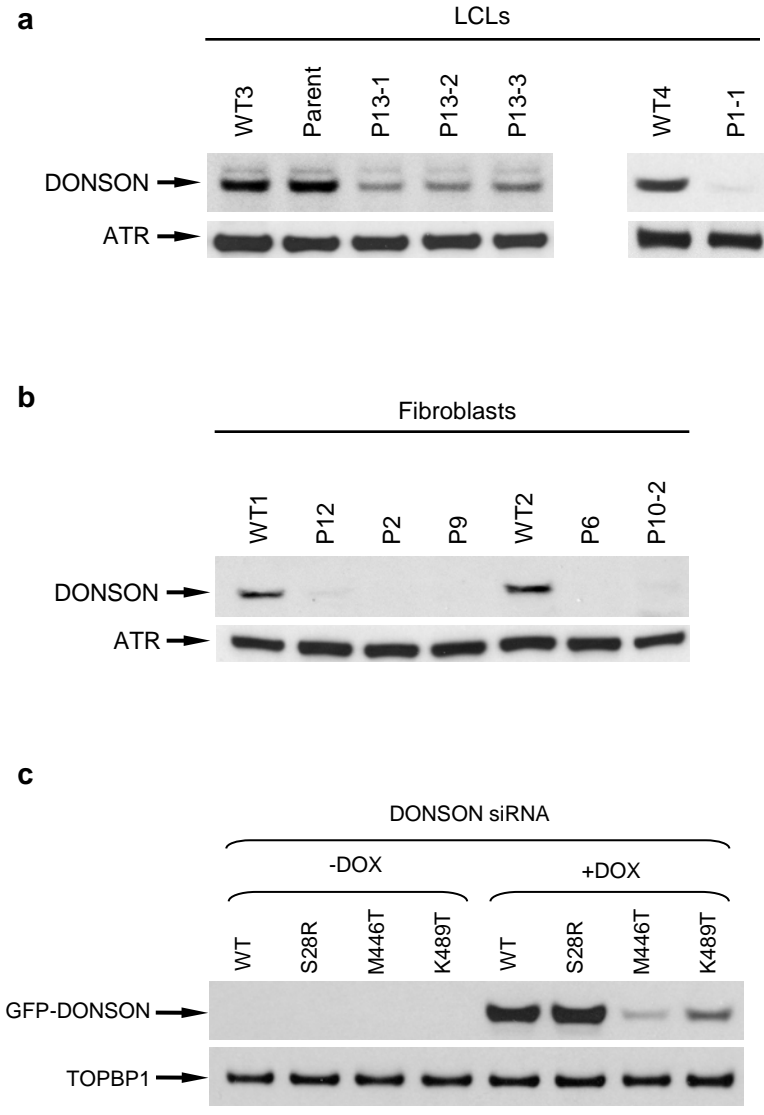


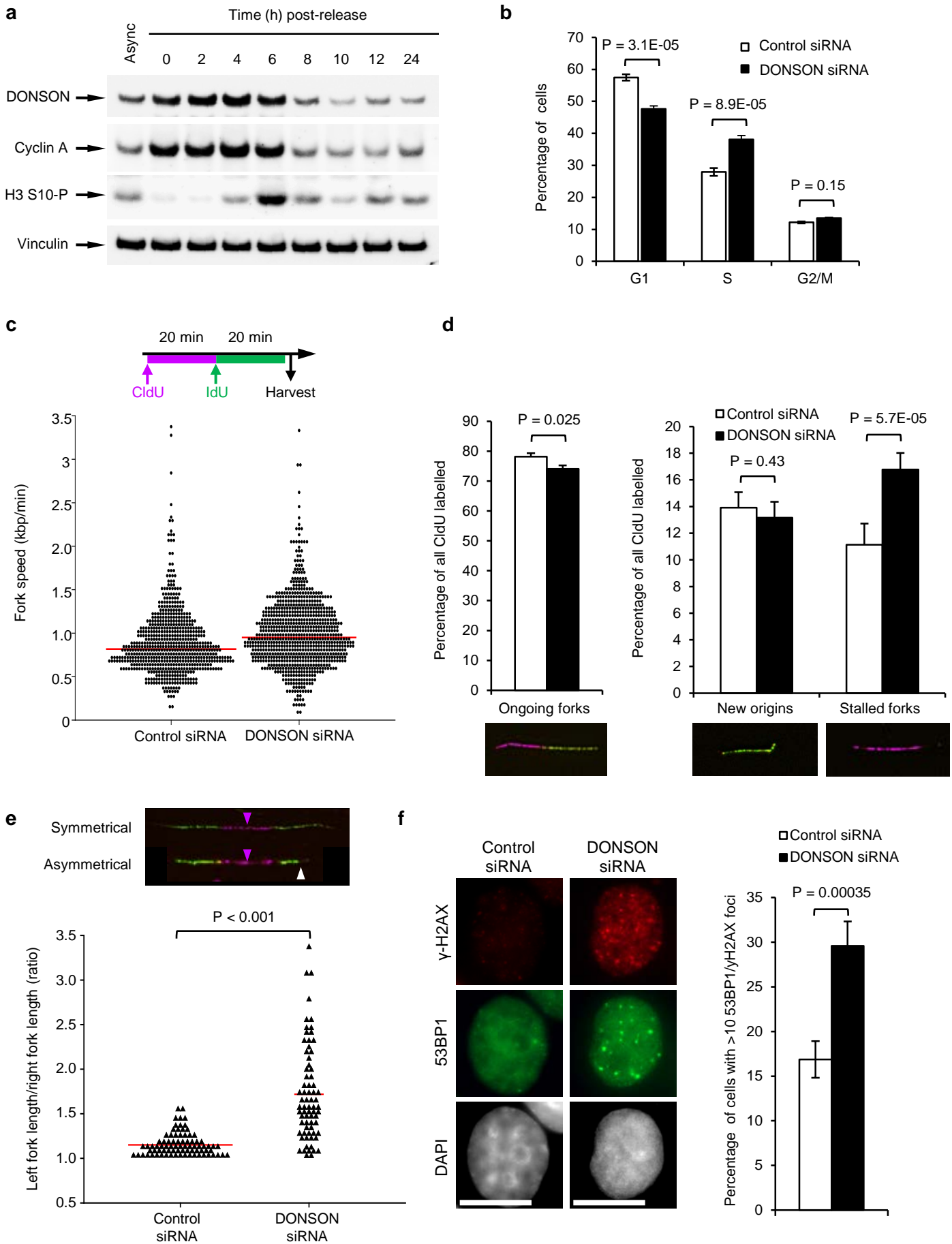
Figure 3

Figure 4

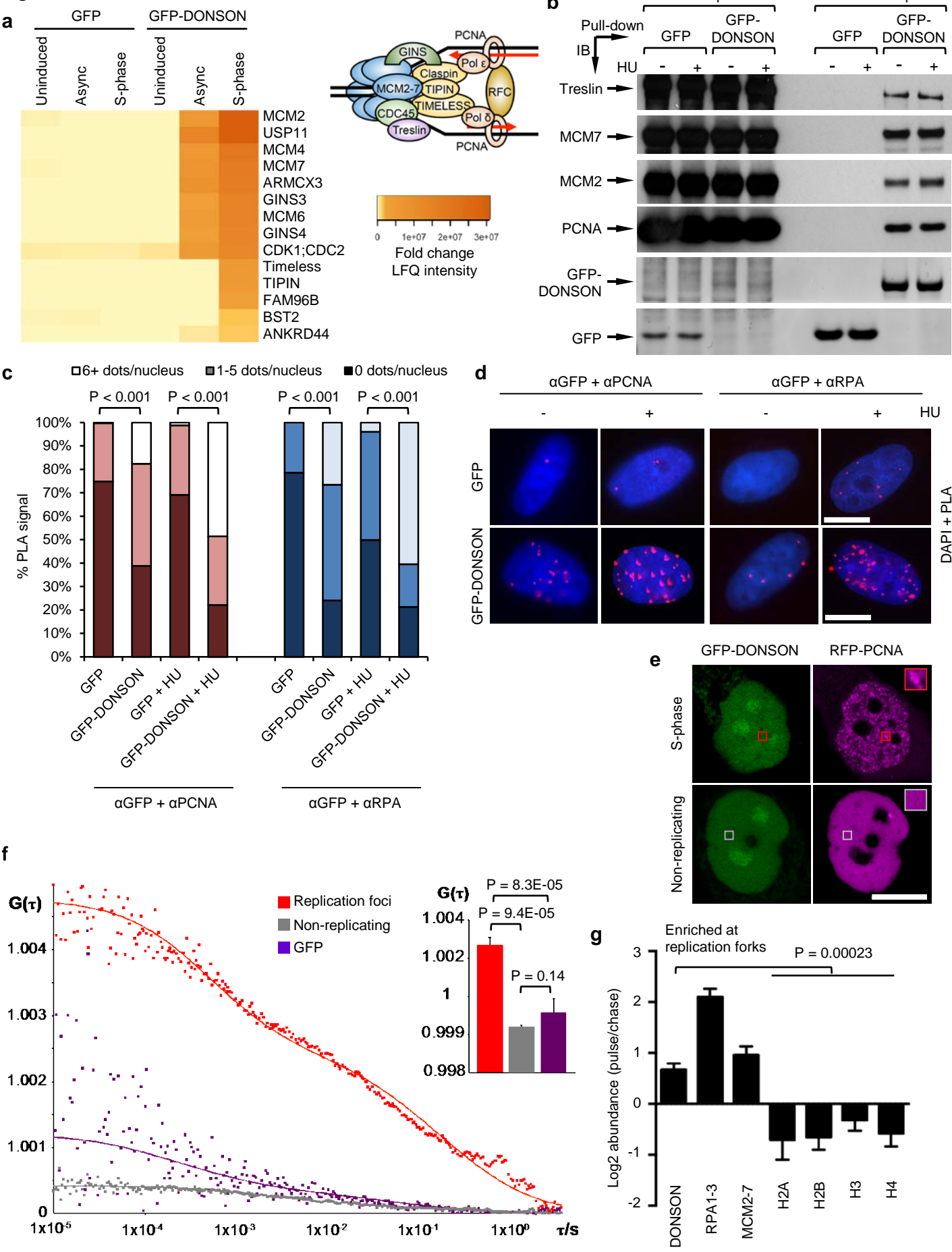


Figure 5

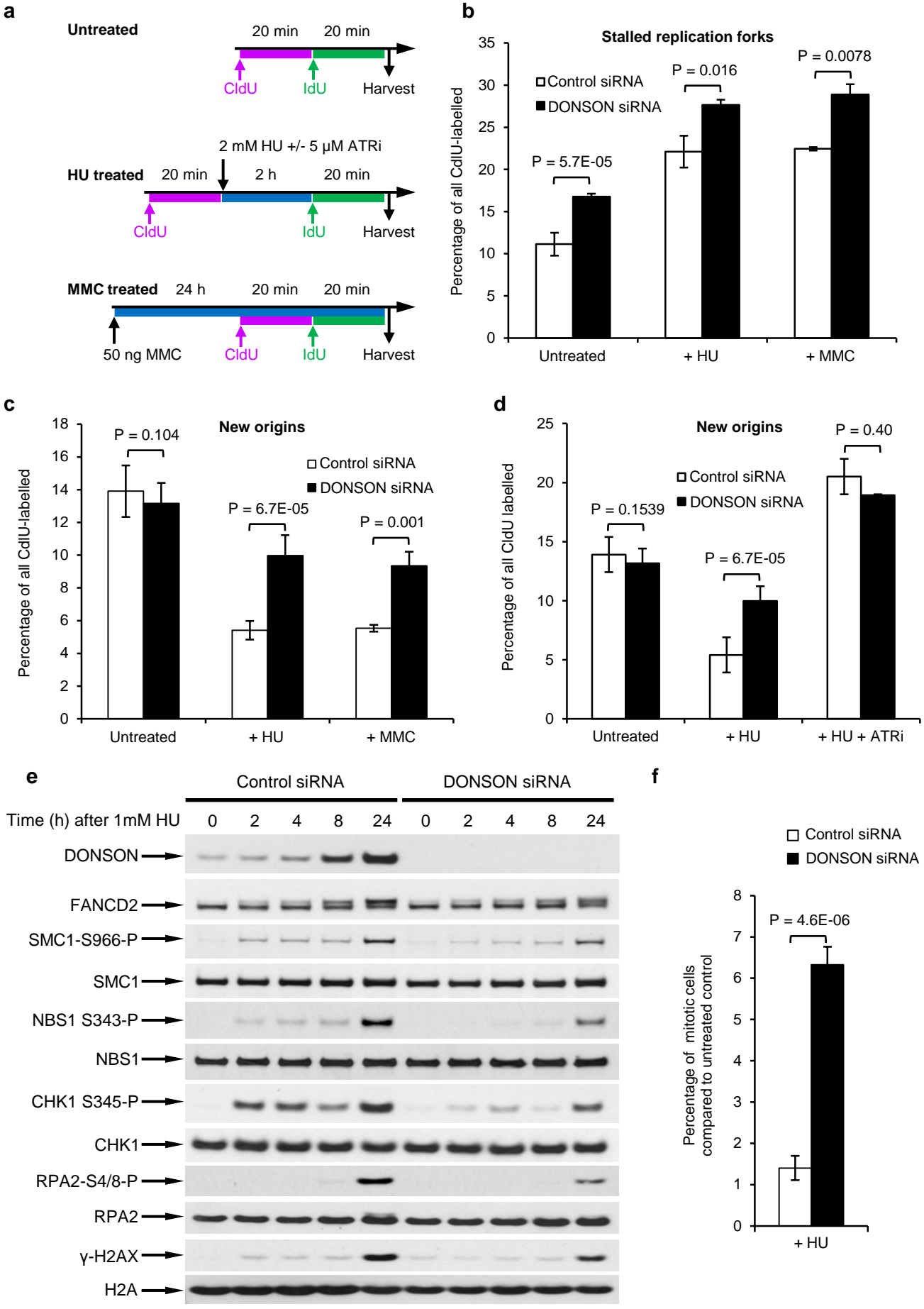


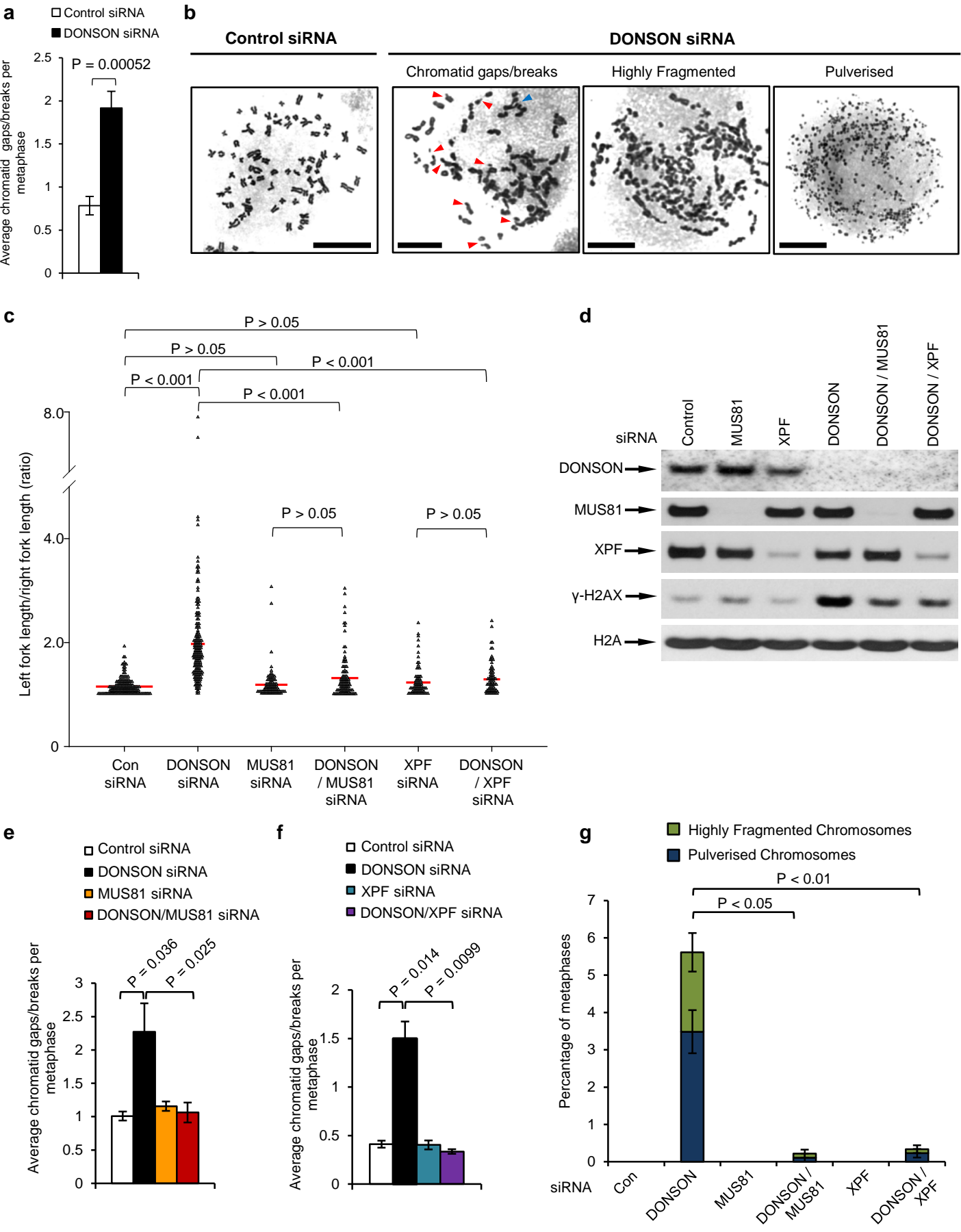
Figure 6

Figure 7

