In silico analysis on the functional and structural impact of Rad50 mutations involved in DNA strand break repair

Juwairiah Remali [1] , **Wan Mohd Aizat** [2] , **Chyan Leong Ng** [2] , **Yi Chieh Lim** [3] , **Zeti-Azura Mohamed-Hussein** [2, 4] , **Shazrul Fazry** Corresp. [1, 5]

[1] Department of Food Sciences, Faculty of Science and Technology, Universiti Kebangsaan Malaysia, Bangi, Selangor, Malaysia

[2] Institute of Systems Biology (INBIOSIS), Universiti Kebangsaan Malaysia, Bangi, Selangor, Malaysia

[3] dKDanish Cancer Society, Research Centre Strand Boulevard, Copenhagen, Denmark

[4] Department of Applied Physics, Faculty of Science and Technology, Universiti Kebangsaan Malaysia, Bangi, Selangor, Malaysia

[5] Pusat Penyelidikan Tasik Chini, Fakulti Sains dan Teknologi, Universiti Kebangsaan Malaysia, Bangi, Selangor, Malaysia

Corresponding Author: Shazrul Fazry
Email address: shazrul@ukm.edu.my

Background. DNA double strand break repair is important to preserve the fidelity of our genetic makeup after DNA damage. Rad50 is one of the components in MRN complex important for DNA repair mechanism. Rad50 mutations can lead to microcephaly, mental retardation and growth retardation in human. However, Rad50 mutations in human and other organisms have never been gathered and heuristically compared for their deleterious effects. It is important to assess the conserved region in Rad50 and its homolog to identify vital mutations that can affect functions of the protein.

Method. In this study, Rad50 mutations were retrieved from SNPeffect 4.0 database and literature. Each of the mutations was analysed using various bioinformatic analyses such as PredictSNP, MutPred, SNPeffect 4.0, I-Mutant and MuPro to identify its impact on molecular mechanism, biological function and protein stability, respectively.

Results. We identified 103 mostly occurred mutations in the Rad50 protein domains and motifs, which only 42 mutations were classified as most deleterious. These mutations are mainly situated at the specific motifs such as Walker A, Q-loop, Walker B, D-loop and signature motif of the Rad50 protein. Some of these mutations were predicted to negatively affect several important functional sites that play important roles in DNA repair mechanism and cell cycle signalling pathway, highlighting Rad50 crucial role in this process. Interestingly, mutations located at non-conserved regions were predicted to have neutral/non-damaging effects, in contrast with previous experimental studies that showed deleterious effects. This suggests that software used in this study may have limitations in predicting mutations in non-conserved regions, implying further improvement in their algorithm is needed. In conclusion, this study reveals the priority of acid substitution associated with the genetic disorders. This finding highlights the vital roles of certain residues such as K42E, C681A/S, CC684R/S, S1202R, E1232Q and D1238N/A located in Rad50 conserved regions, which can be considered for a more targeted future studies.

In silico analysis on the functional and structural impact of Rad50 mutations involved in DNA strand break repair

Juwairiah Remali[1], Wan Mohd Aizat[2], Chyan Leong Ng[2], Yi Chieh Lim[3], Zeti-Azura Mohamed-Hussein[2,4], and Shazrul Fazry[1,5,*]

[1] Department of Food Sciences, Faculty of Science and Technology, Universiti Kebangsaan Malaysia, Bangi, Selangor, Malaysia

[2] Institute of Systems Biology (INBIOSIS), Universiti Kebangsaan Malaysia, Bangi, Selangor, Malaysia

[3] dKDanish Cancer Society, Research Centre Strand Boulevard, Copenhagen, Denmark

[4] Department of Applied Physics, Faculty of Science and Technology, Universiti Kebangsaan Malaysia, Bangi, Selangor, Malaysia

[5] Pusat Penyelidikan Tasik Chini, Fakulti Sains dan Teknologi, Universiti Kebangsaan Malaysia, Bangi, Selangor, Malaysia

Corresponding Authors:

Shazrul Fazry
Department of Food Sciences, Faculty of Science and Technology, Universiti Kebangsaan
Malaysia, Bangi, Selangor, Malaysia Email address: shazrul@ukm.edu.my

Abstract

Background. DNA double strand break repair is important to preserve the fidelity of our genetic makeup after DNA damage. Rad50 is one of the components in MRN complex important for DNA repair mechanism. Rad50 mutations can lead to microcephaly, mental retardation and growth retardation in human. However, Rad50 mutations in human and other organisms have never been gathered and heuristically compared for their deleterious effects. It is important to assess the conserved region in Rad50 and its homolog to identify vital mutations that can affect functions of the protein.
Method. In this study, Rad50 mutations were retrieved from SNPeffect 4.0 database and literature.

38   Each of the mutations was analysed using various bioinformatic analyses such as PredictSNP,
39   MutPred, SNPeffect 4.0, I-Mutant and MuPro to identify its impact on molecular mechanism,
40   biological function and protein stability, respectively.

41   **Results.** We identified 103 mostly occurred mutations in the Rad50 protein domains and motifs,
42   which only 42 mutations were classified as most deleterious. These mutations are mainly situated
43   at the specific motifs such as Walker A, Q-loop, Walker B, D-loop and signature motif of the
44   Rad50 protein. Some of these mutations were predicted to negatively affect several important
45   functional sites that play important roles in DNA repair mechanism and cell cycle signalling
46   pathway, highlighting Rad50 crucial role in this process. Interestingly, mutations
47   located at non-conserved regions were predicted to have neutral/non-damaging effects, in
48   contrast with previous experimental studies that showed deleterious effects. This suggests that
49   software used in this study may have limitations in predicting mutations in non-conserved
50   regions, implying further improvement in their algorithm is needed. In conclusion, this study
51   reveals the priority of acid substitution associated with the genetic disorders. This finding
52   highlights the vital roles of certain residues such as K42E, C681A/S, CC684R/S, S1202R,
53   E1232Q and D1238N/A located in Rad50 conserved regions, which can be considered for a
54   more targeted future studies.

55

56   **Keywords:** Rad50, Rad50 related diseases, DNA damage, Rad50 mutation, Rad50 *in silico*
57   model

58

59

Introduction

60

61

62   DNA repair process exists in all organisms including both prokaryotes and eukaryotes, and most
63   of the related proteins in this process are known to be highly conserved throughout biological
64   evolution. One of such protein complexes involved in eukaryotic DNA repair process is MRN
65   complex and it comprises of three proteins i.e. meiotic recombination 11 (Mre11), DNA repair
66   protein Rad50, and nibrin (called Nbn or Nbs1). These proteins play an important role in
67   maintaining the genomic integrity by orchestrating DNA damage checkpoint, telomere
68   maintenance, homologous recombination (HR) as well as non-homologous end joining repair
69   (NHEJ) mechanism (van den Bosch, Bree & Lowndes, 2003). MRN complex is one of the first
70   factors to be localised to DNA lesions where it has a structural role by tethering and stabilising
71   broken chromosomes (de Jager et al., 2001; van den Bosch, Bree & Lowndes, 2003).

72

73   Null mutations in MRN complex have been shown to be lethal in higher eukaryotes such as in
74   embryonic stem cells (Luo et al., 1999). In addition, mutations in the *Nbs1* gene, can cause
75   Nijmegen breakage syndrome (NBS) whereas Mre11 mutations resulted in Ataxia telangiectasia-
76   like disease syndrome (ATLD) (Carney *et al.*, 1998). So far, studies of Nbs1 and Mre11
77   deficiencies in human have been extensively investigated through cells and clinical data obtained

78 from NBS and ATLD patients (Barbi et al., 1991; Waltes et al., 2009). Unfortunately
79 investigation of the effect of Rad50 mutations on human is very limited due to only one patient
80 with fully characterized Rad50 deficiency (known as NBS like disorder (NBSLD)) has been
81 reported (Waltes et al., 2009). This NBSLD patient with microcephaly, bird-like features,
82 radiosensitivity and delayed development, revealed to inherit heterozygous mutation from her
83 parents (Barbi et al., 1991). The first mutation (c.3277C/T; p.R1093X) on exon 21 was
84 maternally inherited causing a premature termination codon, thus producing a truncated Rad50
85 protein. Whereas the second mutation on the exon 25 (c.3939A/T) was paternally inherited and it
86 has changed the stop codon of normal Rad50 to a tyrosine codon, thereby producing a larger
87 Rad50 protein (Waltes et al., 2009). Both mutation interestingly give rise to hypomorphic
88 characterization of the Rad50 expressions in this patient (Gatei et al., 2011). The cause of this
89 characteristic are still on debate to this day. Given that perturbation of Rad50 structure and
90 function could contribute to genomic instability (Assenmacher & Hopfner, 2004), it is therefore
91 important to decipher its conserved domains and genetic polymorphism.

92

93 Single nucleotide polymorphism (SNP) is one of the most common types of genetic variation in
94 human (Lee et al., 2005). Even though most of the polymorphic changes do not affect normal
95 cellular function, some variants do influence gene expression or translated protein function
96 (Risch & Merikangas, 1996; Collins, Guyer & Charkravarti, 1997). For instance, cystic fibrosis
97 (Bartoszewski et al., 2010), sickle-cell anemia (Shaikho et al., 2017), and β-thalassemia (Traeger
98 et al., 1980) are examples of diseases resulted from SNPs. Nearly half of the disease-related
99 mutations are derived from nonsynonymous SNPs (nsSNPs), a single base change that alters the
100 amino acid sequence of the encoded protein (Cargill et al., 1999; Halushka et al., 1999).
101 Although it is remarkably important to reveal the connection between SNPs and related diseases,
102 the accelerating number of known SNPs have made it very difficult to discriminate between
103 pathogenic and neutral variants through experimental validations (Tranchevent et al., 2011).
104 Therefore, bioinformatic prediction tools have become extremely critical for the initial analysis
105 of their molecular functions as well as prioritization of further experimental characterization
106 including deciphering the effects of Rad50 SNPs (Bendl et al., 2014). Furthermore, prioritization
107 of disease candidates genes from experiment and databases evidence is essential for further
108 pathological investigation (Piro & Di Cunto, 2012). Several investigations on Rad50 mutations
109 have been reported in human (Waltes et al., 2009; Gatei et al., 2011), mice (Bender et al., 2002;
110 Roset et al., 2014), yeast (Alani, Padmore & Kleckner, 1990; Chen et al., 2005), and archaea
111 (Koroleva et al., 2007) yet there are still no reports that compare these experimental results with
112 *in silico* prediction, which will be important for the protein functional annotation. Moreover, a
113 number of different SNPs for Rad50 have been deposited in SNP databases but their impact on
114 the cellular regulation have not been thoroughly investigated thus far.

115

116 Hence, the aim of this study was to identify the functional and structural effects of amino acid
117 mutations in Rad50 gathered from exhaustive literature review and SNP database (SNPeffect

118   4.0) search. Rad50 sequences in different organisms including human and selected animals
119   (chimpanzee, rats, mice, zebra fish, rabbit and fruit fly) and yeasts were compared and aligned to
120   identify their conserved residues. Mutations that contributed to the most damaging effects were
121   then analysed *in silico* using PredictSNP for the amino acid impact after the substitution,
122   MutPred for predicting molecular mechanism, SNPeffect for identification of protein or amyloid
123   aggregation as well as I-Mutant and MuPro for protein stability after the mutation. Such
124   approach was also successfully reported by several researchers studying the impact of various
125   SNPs. For example, Marín-Martín *et al.*, (2014) studied the impact of SNPs in the ABCA1
126   transporter gene by cross validating their prediction with experimentally reported data. Another
127   study by Fawzy *et al.*, (2015) also validated their *in silico* approach finding by means of
128   comparison with available literature to study gene polymorphisms in obese children and
129   adolescents. In this study, Rad50 mutations gathered from various studies are compared with
130   their *in silico* predictions. This is highly valuable in understanding Rad50 functional roles
131   especially during DNA strand break, allowing prioritization of mutations or sites to be studied in
132   future *in vivo* studies, whilst bearing in mind its possible impact on human. Ultimately, this may
133   help on the development of precision medicine for Rad50 mutations in human.
134
135
136

137   **Materials & Methods**

138

139   **Multiple sequence alignment (MSA) analysis and conserved domain analyses**

140

141   Human Rad50 protein sequence was obtained from National Center for Biotechnology
142   Information (NCBI). The sequence similarity search tool, BLASTP from the NCBI server
143   (http://blast.ncbi.nlm.nih.gov/Blast.cgi) was used to find homologs for Rad50. To investigate the
144   similarity between Rad50 protein in human and other organisms such as *Danio rerio, Mus*
145   *musculus, Rattus norvegicus, Pan troglodytes, Oryctolagus cuniculus, Drosophila melanogaster,*
146   *Saccharomyces cerevisiae and Schizosaccharomyces pombe,* a multiple sequence analysis
147   (MSA) was conducted using Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/) with
148   default settings to determine consensus and conserved regions between the multiple sequences of
149   different   organisms   (Sievers   &   Higgins,   2018).   Meanwhile,   InterPro
150   (http://www.ebi.ac.uk/interpro/) was used to identify the domains and motifs using human
151   sequence (Finn et al., 2017). InterPro results are classified into several types (families, domains,
152   motif or sites) depending on the biological entity they represent (Finn et al., 2017). Using this
153   tool, Rad50 protein sequence was classified into families and the presence of domains and
154   important sites were predicted. ClustalX software (Thompson, Gibson & Higgins, 2002) was
155   used to view and analyse the conserved regions within the domains and motifs in the selected
156   proteins.

157

Data mining of Rad50 mutation from literature and SNPs database

Rad50 mutations were identified from previous published manuscripts using PubMed database and their functional impacts were extracted for comparison. Besides that, naturally occurring single nucleotide polymorphisms (SNPs) in Rad50 were retrieved from SNPeffect 4.0 database (http://snpeffect.switchlab.org/about) (De Baets et al., 2012) (date of access: April 07th 2018). SNPeffect 4.0 database currently contains more than 60,000 human SNPs gathered from human avariance list available at UniProt website (https://www.uniprot.org/). It specifically focuses on the molecular characterization, annotation of diseases as well as polymorphism variants in human proteins (De Baets et al., 2012). All these available Rad50 protein mutations (obtained from both literature and databases) have been aligned using pairwise alignment through Clustal Omega between human sequence and other organisms' sequence, individually. From this analysis, we identified similar mutation sites in human. All the identified equivalent mutations in human were manually refined, for example removing the same residues and mutations that has been studied by several different researchers to identify the non-redundant mutations (S1 Table). Identified mutations (after converting to equivalent residues in human) were then mapped into Fig. S1.

Secondary structure prediction and analysis of 3D modeling

The Rad50 templates identified from the BLAST analysis also were used to develop secondary structure and 3D model. The PSIPRED program (http://bioinf.cs.ucl.ac.uk/psipred/) has been utilised for secondary protein structure prediction (Buchan et al., 2013). Secondary structure prediction has revealed a clear distribution of alpha helix, beta sheet and coil in *H. sapiens* (Helix: 74.69%, coil; 18.29 and beta sheet; 7.01% (S2 Figure). Databases such as UniProt (https://www.uniprot.org/) and Protein Data Bank (PDB) (https://www.rcsb.org/) were used to identify structural information regarding Rad50 protein in human. Rad50 protein sequence also has been BLAST searched against Protein Data Bank (PDB) sequence in Network Protein Sequence @nalysis (NPS@) (https://npsa-prabi.ibcp.fr/) to identify the most identical structure. The incomplete structure has been further predicted using fold recognition method using Protein Homology/analogY Recognition Engine Version 2.0 (Phyre2) (http://www.sbg.bio.ic.ac.uk/phyre2) (Kelley et al., 2015). Phyre2 is an online tool to predict and analyze protein structure, function and mutations which uses advanced remote homology detection methods to build 3D models, predict ligand binding sites and analyze the effect of amino acid variants (e.g., nonsynonymous SNPs (nsSNPs)) for a protein sequence (Kelley et al., 2015). Rad50 sequence was submitted to the webserver to interpret the secondary and tertiary structures of the model, domain composition and quality. 3D model of Rad50 was run under 'intensive' mode that generates a complete full-length model of a protein sequence by using multiple template modelling and simplified *ab initio* folding simulation (Kelley et al., 2015).

197 UCSF Chimera software was used to view and to analyse the 3D structure (Pettersen et al.,
198 2004).

199

200 **Prediction of deleterious effects of Rad50 mutations using *in silico* tools**

201

202 The Rad50 mutations were in silico predicted using PredictSNP to determine their possible
203 molecular impacts in human (https://loschmidt.chemi.muni.cz/predictsnp1/) (Bendl et al., 2014).
204 Its benchmark dataset contains over 43,000 mutations obtained from the Protein Mutant
205 Database and the UniProt database (Bendl et al., 2014). This tool incorporated six established
206 prediction tools; such as Multivariate Analysis of Protein Polymorphism (MAPP) (Stone &
207 Sidow, 2005), Predictor of human Deleterious Single Nucleotide Polymorphisms (PhD-SNP)
208 (Capriotti & Fariselli, 2017)*,* PolyPhen-1 (Ramensky, 2002)*,* PolyPhen-2 (Adzhubei, Jordan &
209 Sunyaev, 2013)*,* Sorting Intolerant from Tolerant (SIFT) (Sim et al., 2012) and Single-
210 Nucleotide Amplified Polymorphisms (SNAP) (Bromberg & Rost, 2007) to provide a more
211 accurate and robust comparison. We classified the mutations as deleterious if five to seven
212 analyses performed were identified as damaging in PredictSNP. For instance, an *in silico*
213 prediction was considered accurate when a given mutation predicted to be deleterious (as
214 performed in this study) was also found experimentally deleterious (either *in vitro* or *in vivo* with
215 phenotypes such as embryonic lethality, growth defect and/or cancer predisposition) based on
216 previous cited studies. Conversely, the prediction is inaccurate if such deleterious mutations was
217 predicted as neutral or tolerant.

218

219 **Molecular mechanism of amino acid substitutions**

220

221 To determine the molecular mechanism based on pathogenicity of amino acid substitutions in
222 Rad50, MutPred2 (Pejaver et al., 2017) (http://mutpred2.mutdb.org/index.html) analysis was
223 carried out. This program predicts the pathogenicity and molecular impacts of amino acid
224 substitutions potentially affecting the phenotype. It is trained on a set of 53,180 pathogenic and
225 206,946 unlabelled (putatively neutral) variants obtained from the Human Gene Mutation
226 Database (HGMD) (Stenson et al., 2017), SwissVar (Mottaz et al., 2010), dbSNP (Sherry, 2001)
227 and inter-species pairwise alignment (Pejaver et al., 2017). The output of MutPred contains a
228 general probability that the amino acid substitution is deleterious/disease-associated, and a list of
229 rank of specific molecular alterations potentially affecting the phenotype with its *p*-value
230 (<0.05).

231

232 **Prediction of molecular and structural effects of protein coding variants in Rad50 mutation**

233

234 Prediction of molecular and structural effects of protein coding variants in Rad50 mutations was
235 performed using SNPeffect4.0 (De Baets et al., 2012) (http://snpeffect.switchlab.org/about). The
236 analysis includes predictions of the aggregation prone regions in a protein sequence (TANGO),

237 amyloid-forming regions (WALTZ) and chaperone binding site (LIMBO). The range of
238 prediction score differences outside -50 to 50 for mutants are considered significant (De Baets et
239 al., 2012). SNPeffect also uses FoldX (Schymkowitz et al., 2005) to analyse the effect of
240 mutations on the structural stability. However, as structure quality is important for the accuracy
241 of delta G predictions for stability, model structures with less than 90% sequence identity to the
242 modelling template structure will not be modelled (De Baets et al., 2012).

243
244 **Analysis of protein stability**

245
246 The stability of Rad50 upon single amino acid residue mutations were predicted using MUpro
247 (http://mupro.proteomics.ics.uci.edu/) (Cheng, Randall & Baldi, 2006) and I-Mutant 3.0
248 (http://gpcr2.biocomp.unibo.it/cgi/predictors/I-Mutant3.0/I-Mutant3.0.cgi) (Capriotti, Fariselli
249 & Casadio, 2005) using default setting, for instance temperature was set at 25°C and pH 7. Mu-
250 Pro and I-Mutant 3.0 are valuable tools for protein stability prediction and analysis, even when
251 the protein structure is not yet known with atomic resolution. Both use support vector machines
252 (SVM)-based tool to predict protein stability changes for single amino acid mutations either from
253 both sequence or structural information which correctly predicts with over 80% accuracy using
254 cross validation methods (datasets and experimental) (Capriotti, Fariselli & Casadio, 2005;
255 Cheng, Randall & Baldi, 2006). Rad50 protein sequence was searched against the web server
256 and energy changes (ΔΔG) were recorded. Negative value for ΔΔG represents a decrease in
257 protein stability whereas positive value for ΔΔG represents an increase in stability.

258
259 # Results

260
261 **Rad50 data acquisition and MSA analysis**

262
263 Human Rad50 sequence from NCBI database contains 1312aa with the accession number of
264 AAB07119.1. Sequence homology search of the human Rad50 protein was performed against
265 NCBI nonredundant protein databases (*E*-value ≤ 1E-05) and the result was downloaded for
266 further analysis. Out of 500 sequences, six sequences were choosen for MSA analysis from
267 diverse organisms such as *D. rerio*, *M. musculus*, *R. norvegicus, P. troglodytes, O. cuniculus,*
268 and *D. melanogaster*. Two sequences, *Saccharomyces cerevisiae* and *Schizosaccharomyces*
269 *pombe* were also included due to widely being used as models in previous Rad50 studies (S1
270 Table).

271
272 **Analysis of protein domains**

273
274 Domain identification analysis showed that Rad50 contains three P-loop containing nucleoside
275 triphosphate hydrolase (P-loop NTPase) domains which belong to ATP Binding Cassette (ABC)
276 protein superfamily (De La Rosa Metzere Bierlein & and Scott W. Nelson, 2011). It is located

277 near the N- and C-terminal, at the residue number of 25-103, 130-227 and 1196-1279 (Figure
278 1a). Residue annotation showed that Rad50 has six specific motifs including Walker A and Q-
279 loop that are located at the N-terminal whereas Rad50 signature motif, Walker B, D-loop and H-
280 loop/switch region are located at C terminal (Figure 1a) (De La Rosa Metzere Bierlein & and
281 Scott W. Nelson, 2011). It also has a domain called zinc hook (635-734aa) located at C-terminal
282 region (Figure 1a) (Hopfner et al., 2002). Multiple sequence alignment (MSA) analysis between
283 human Rad50 and its homologous genes (*D. rerio*, *M. musculus*, *R. norvegicus, P. troglodytes,*
284 *O. cuniculus, D. melanogaster*, *S. cerevisiae* and *S. pombe*) also revealed that these specific
285 motifs are highly conserved (Figure 1b).

286
287

288 **Mutation datasets from the literature and database searches**

289

290 In order to identify the Rad50 mutations, literature pertaining to the topic was exhaustively
291 searched and 18 articles over the period of 1990 to 2017 were identified. All these mutations
292 from different organisms were listed in S1 Table. There are 103 mutations identified which
293 mostly occurred in the protein domains and motifs with various biological effects (S2 Table). In
294 order to obtain equivalent mutations in human, pairwise alignment was performed individually
295 between each organism *(D. rerio, M. musculus, R. norvegicus, P. troglodytes, O. cuniculus, D.*
296 *melanogaster, S. cerevisiae and S. pombe)* and the Rad50 human sequence as a reference (S1
297 Table). Then, MSA analysis was carried out between these sequences from different organisms
298 (including human) to identify consensus regions (Figure 1 and S1). Further refinement such as
299 integrating similar mutations that occurred at the same positions (for examples; S1202R, K42R,
300 S679R, P682E, V683R, R1214E, K6E, and R81I (S2 Table) from different organisms of which a
301 total of 80 different mutations or non-redundant mutation were identified. All these mutations
302 have been mapped based on equivalent residues in human (S1 Figure). From SNPeffect 4.0
303 database, another 13 SNP mutations were also identified (S3 Table). However, from the total of
304 103 mutations obtained from literature, only 42 residues of the Rad50 protein mutations were
305 known to contribute to the most damaging effects *in vitro* and *in vivo* such as embryonic lethality
306 (Bender et al., 2002; Roset et al., 2014) and growth defect (Table 1 and S2 Table) (Alani,
307 Padmore & Kleckner, 1990; Bhaskara et al., 2007; Waltes et al., 2009; He et al., 2012; Barfoot et
308 al., 2015; Hohl et al., 2015). Most of these deleterious mutations reside at the specific motifs
309 such as Walker A, Q-loop, zinc hook, Rad50 signature motif, Walker B and D-loop (Figure 1b)
310 that become our primary research focus (Figure 1b).

311

312 **3D structure modelling of Rad50**

313

314 Currently there is no complete structure of human Rad50 available. Nonetheless, a crystal
315 structure of Rad50 hook and coil-coil domain (HCC) that contains 182 residues (residue 585-
316 766) has been determined (PDB ID: 5GOX) (Park et al., 2017), which represents 13% of the

317   Rad50 structure in human. We attempted to predict a more complete human Rad50 3D structure
318   model using homology modeling. Homology modeling program Phyre2 successfully predict the
319   N-terminal of 276 residues (2-278) and C-terminal of 155 residues (1153-1306) of human Rad50
320   with 100% confidence level using the template model (PDB ID: 5DAC) from *Chaetomium*
321   *thermophilum* (Seifert et al., 2016) that share 66% sequence identity (S2 Figure). As a result, half
322   of human Rad50 protein structure was obtained. The regions with no 3D structure information
323   available are residues 279 to 584 and residues 767 to 1152 (S2 Figure), which were mainly
324   predicted to consist of alpha helices secondary structure (S2 Figure). Pairwise alignment of
325   Rad50 sequence between *C. thermophilum* (1315aa) and human (1312aa) showed about 30%
326   sequence identity (S2 Figure). The result showed that the partial structure of *C. thermophilum*
327   that has been determined (black line) are highly conserved with the human sequence (S2 Figure)
328   suggesting that the human structure should also share high structure similarity to *C.*
329   *thermophilum* at these regions. In agreement to this, results from Phyre2 prediction showed that
330   the N-terminal and C-terminal of Rad50 form a globular and coil-coil domain, similar to the
331   structure of *C. thermophilum* (Figure 2a). With the generated model, six motifs of Rad50 namely
332   Q-loop, Walker B, signature motif, D-loop, Walker A and H-loop were identified and marked in
333   the 3D structure (Figure 2a). All identified deleterious residues found in the domain were also
334   marked as shown in Figure 2b. To correlate the deleterious residues in the Rad50 HCC domain
335   with zinc hook motif that was not found in the model, the structure of 181 residues (residue 585-
336   766) that has been determined (PDB ID: 5GOX) independently was employed (Figure 2c and 2d)
337   for functional analysis.

338

339   **Analyses of Rad50 mutation deleterious effects**

340

341   All of these 42 mutations (based on mutations from other organisms mapped to human) (Figure
342   1b and 2b) were then analysed using bioinformatics analyses such as impact of amino acid
343   substitutions (PredictSNP), molecular mechanism (MutPred), structural phenotyping (protein and
344   amyloid aggregation) (SNPeffect 4.0) and protein stability (MuPro and I-Mutant 3.0) (Table 2).
345   All raw data from each analysis has been supplied as supplementary data (S3 Table for
346   PredictSNP analysis, S4 Table for MutPred analysis, S5 Table for SNPeffect analysis and finally
347   S6 Table for I-Mutant and MuPro analysis). The results showed that most of the deleterious
348   effects fall into specific motifs such as Walker A, Q-loop, Rad50 signature motif, Walker B and
349   D-loop (Table 2 and Figure 2a). Previous analysis also revealed that mutations at these motifs
350   contributed to a number of biological defects such as growth defect (Alani, Padmore & Kleckner,
351   1990; He et al., 2012; Hohl et al., 2015) embryonic lethality (Bender et al., 2002; Roset et al.,
352   2014), cancer predisposition (Bender et al., 2002; Roset et al., 2014), hematopoietic and
353   spermatogenic depletion (Bender et al., 2002; Roset et al., 2014) (Table 1). Several mutations at
354   the zinc hook region (C681A, C681S, P682E, C684R and C684S) and ATPase/coiled-coil
355   domain (K6E and K132E) also showed to be deleterious (Table 2, Figure 2c and 2d).

356

357 Furthermore, mutations located at Walker A (Figure 2a) were predicted to affect catalytic and
358 allosteric site, loss or gain of methylation, alteration of DNA binding, metal binding, ordered
359 interface and the loss of relative solvent accessibility, which all are depending on the types of
360 amino acid substitutions (Table 2). These mutations were predicted to affect ATP binding site
361 motif, N-myristolylation, casein kinase II (CK2), protein kinase A (PKA) phosphorylation site
362 and Forkhead-associated (FHA) functional sites. Mutations at the Walker A region also might
363 led to the decrement of protein stability as predicted by I-Mutant and MuPro. Mutation at the Q-
364 loop region (Q159H) (Figure 2a and 2b) also predicted to have significant deleterious effect and
365 decreased protein stability, but no effects have been identified on its molecular mechanism and
366 structural phenotyping as predicted by MutPred and SNPeffect 4.0, respectively (Table 2).
367

368 Mild deleterious effect was predicted at the mutated zinc hook domain (Table 2 and Figure 2e).
369 Subsequent analysis using MutPred also revealed that any mutation at zinc hook might affect
370 several important functional sites that involved in DNA damage repair signalling response and
371 cell   cycle   checkpoints   such   as   phosphatidylinositol   3-kinase-related   kinases   (PIKK)
372 phosphorylation site, protein kinase C (PKC) phosphorylation site and BRCA1 C-terminus
373 (BRCT) phosphopeptide ligands binding sites (Table 2). Moreover, deleterious mutation was
374 predicted at the conserved cysteine residue located at the zinc hook motif (CXXC). For example,
375 amino acid substitutions of alanine (A) and serine (S) at the cysteine residue position 681;
376 (C681(A/S)) (Figure 2d) may affect N-glycosylation, proline-directed phosphorylation and
377 mitogen-activated protein kinases (MAPK) phosphorylation site, which possibly due to the
378 affected zinc binding domain (Table 2). Another deleterious mutation, C684(R/S) was also
379 predicted to not affect its molecular mechanism but might disrupt diarginine retention/retrieving
380 signal, PKC and PIKK phosphorylation site (Table 2). Whilst P682E (Figure 2d) mutation may
381 lead to gain of helix, altered coiled coil domain, loss of N-linked glycosylation and CK2
382 phosphorylation site (Table 2).
383

384 Rad50 signature motif (Figure 2a) is a critical site which could lead to deleterious effects if
385 mutated   as   suggested   by   PredictSNP   analysis   (Table   2).   All   mutations   in   this   motif
386 (S1202A/R/M, Q1205E and K1206M/A/E) or located near this motif (G1198E, L1211W and
387 R1214A/E) (Figure 1b and 2b) were predicted to affect the protein allosteric and catalytic sites
388 (Table 2), except for R1198E. Mutations at residue S1202A/R/M (Figure 1b and 2b) might affect
389 PKA phosphorylation sites and glycosaminoglycan attachment site (Table 2). Furthermore,
390 R1214A (Figure 1b and 2b) mutation might affect ATP-binding cassette, ABC transporter-type,
391 signature   and   profile   functional   sites   (Table   2).   We   have   also   predicted   ~~identified~~ several
392 mutations in Rad50 signature motif such as Q1205E, L1211W and R1214A that contributed to
393 the total defect in the structural phenotyping such as the increment in protein and amyloid
394 aggregation and the decrement of protein stability (Table 2).
395

396   We have also predicted K6E, K132E and K105E mutations occurred at the coil-coiled domain or
397   ATPase domain to be deleterious (Figure 1b, 2b and Table 2). Specifically, the mutations at K6E
398   and K132E might lead to loss of strand or loss of helix, respectively. Additionally mutation at
399   K132E also predicted to affect casein kinase 1 (CK1) and PKC phosphorylation sites (Table 2).
400   Even though K22M and R83I (Figure 2b) were predicted to be neutral in PredictSNP analysis,
401   both of these mutations have also been predicted to increase protein aggregation tendency (Table
402   2). The mutation at R83I might contributed to the alteration of coiled coil structure domain,
403   DNA binding and ordered interface, that might affect the functional site such as protein-protein
404   interactions (PPI)-docking motif (Table 2). Another neutral mutation predicted were T191E,
405   C221E and S106E (Figure 1b and 2b), where T191E mutation might be responsible in altering
406   the coiled coil domain and may affect tumor necrosis factor receptor-associated factor (TRAF),
407   serine/threonine-protein kinase (NEK2) and PKC phosphorylation site (Table 2). On the other
408   hand, C221E and S106E (Figure 1b and 2b) were predicted to not affect any molecular
409   mechanism or protein aggregation (Table 2).

410

Discussion

412

413   Rad50 is a member of the structural maintenance of chromosomes (SMC) family of proteins that
414   participates in chromosome structural changes (Kinoshita et al., 2009). The globular ABC
415   ATPase head domain is formed by the N- and C-termini (Figure 2a) (Hohl et al., 2011). The
416   coiled-coil apex of Rad50 contains a conserved cysteine amino acid motif across the organisms,
417   which is called the zinc hook (Kinoshita et al., 2009). When DNA double strand break occurs,
418   Rad50 complex binds to the DNA early in the repair process to recognize such breaks and grips
419   them in close juxtaposition (Paull & Gellert, 1998; de Jager et al., 2001). This protein also
420   activates ATM kinase that is crucial for DNA damage signalling (Uziel et al., 2003).

421

422   Rad50 globular head domain contains conserved domains and motifs (Figure 1a and 2a) such as
423   P-loop NTPase domains and six motifs which are Walker A and B motifs, Rad50 signature
424   motif, D-loop, H-loop, and Q-loop motif (Figure 1a and 2a). P-loop NTPase domains are belong
425   to ABC protein superfamily. The ABC protein superfamily has been identified in diverse
426   organisms and is also known to be one of the most conserved protein superfamilies (Jones,
427   O'Mara & George, 2009). ABC proteins consist of six conserved motifs (Figure 1a and 2a)
428   which make up the nucleotide binding domain in Rad50 (Symington, 2002). The nucleotide
429   binding domain of ABC protein is known to play an important role in binding and hydrolyzing
430   ATP at its dimeric interface (Davidson et al., 2008). Rad50 also has a special conserved Cys-X-
431   X-Cys zinc hook motif at the center of coiled-coil domain (Figure 2c). This motif is in a hook-
432   shaped structure which dimerizes a second hook via cysteine-mediated zinc ion coordination
433   (Figure 2c) (Hopfner et al., 2002).This zinc dependent dimerization event allows the formation
434   of MRN complex which has suitable lengths and conformational arrangements to link sister
435   chromatids in HR and DNA ends in NHEJ (Hopfner et al., 2002).

436

Consistency between bioinformatics prediction and experimental evidence

438

439   PredictSNP was used in this study to provide a more accurate prediction of disease-related
440   mutations as it combines six best performing prediction tools for a consensus classifier (Bendl et
441   al., 2014). Evidently, this *in silico* analysis was consistent with the results from the previous
442   experimental studies where mutations at the Walker A, D-loop, signature motif, Q-loop and
443   Walker B have shown damaging effects (Table 1, 2 and Figure 2a).

444

445   G41D and K40E (Figure 1b and 2b) mutations at the Walker A motif (Figure 2a) and C681A and
446   C684R (Figure 1b and 2d) mutations at the cysteine residue (CXXC) in the zinc hook motif
447   (Figure 2c) conferred an identical phenotype with the Rad50 null mutation characterized by total
448   defect in the formation of viable spore in *S. cerevisiae* experiment (Table 1) (Alani, Padmore &
449   Kleckner, 1990; He et al., 2012). This analysis also identified that mutations at Q-loop (Q159H)
450   and D-loop (D1238N and D1238A) (Figure 1b, 2a and 2b) were also predicted deleterious
451   (Table 2) and were experimentally shown to interrupt all ATP-dependent activities of the
452   complex in different organisms such as *P. furiosus* and T bacteriophage respectively (Table 1)
453   (Moncalian et al., 2004a; De La Rosa Metzere Bierlein & and Scott W. Nelson, 2011).
454   Furthermore, a E1232Q (Figure 1b and 2b) mutation at the Walker B motif (Figure 2a) was also
455   predicted to be deleterious (Table 2). Similarly the mutation of Walker B at residue E798Q in
456   *Thermotoga maritima* showed low ability to respond to DNA damage (Table 1) (Rojowska et al.,
457   2014a). This suggests that this motif is important for a molecular repair process, specifically
458   during DNA binding process, which if mutated will affect the viability of an organism. Our
459   analysis using PredictSNP has identified three mutations, which were N28A (Figure 1b and 2b)
460   (De La Rosa Metzere Bierlein & and Scott W. Nelson, 2011), D1238H (Figure 1b and 2b) (De
461   La Rosa Metzere Bierlein & and Scott W. Nelson, 2011) and S1202R (Figure 1b and 2b)
462   (Kerem et al., 1989; Moncalian et al., 2004a) located at the Walker A, D-loop and Rad50
463   signature motif, respectively (Figure 2a) (Kerem et al., 1989; Moncalian et al., 2004a; De La
464   Rosa Metzere Bierlein & and Scott W. Nelson, 2011).

465

466   Mutations at the Walker A domain and Rad50 signature motif (Figure 2a) may also affect
467   important functional sites such as ATP binding site (Table 2). For example, K42R/M/E/A
468   mutation at the Walker A (Figure 1b and 2b) in *S. cerevisiae* and *D. radiodurans* has been
469   identified experimentally to cause defective in ATPase (Table 1) (Chen et al., 2005; Koroleva et
470   al., 2007) and S793R mutation in *Pyrococcus furiosus* showed the inhibition of ATP binding and
471   disrupted communication between ATP loops (Table 1) (Moncalian et al., 2004a). This mutation
472   further distorted the surface of the C-terminal domain and thus altered the interaction between
473   Rad50 monomers to prevent dimerization (Table 1) (Moncalian et al., 2004a). We have also
474   identified mutations at several motifs such as Walker A (G41D, K42M/R/E/A) and Walker B
475   (E1232Q) (Figure 1b, 2a and 2b) that might affect the binding of FHA phosphopeptide ligands

476   that plays a critical role in DNA damage repair mechanism and cell cycle (Table 2). Many FHA
477   domain–containing proteins localized to the nucleus showed to play a critical role in establishing
478   or maintaining DNA repair, cell cycle checkpoints or transcriptional regulation (Durocher et al.,
479   2000). When mutated, diseases such as Nijmegen breakage syndrome (NBS) and the hereditary
480   cancer syndrome variant Li-Fraumeni (CHK2) will be developed (Matsuura et al., 1998; Varon
481   et al., 1998; Carney et al., 1998b; Featherstone & Jackson, 1998; Bell et al., 1999) suggesting the
482   importance of these conserved residues within Rad50 for DNA repair and maintenance.
483
484   Mutations at or near the Rad50 signature motif (Figure 1b and 2a) were also known to be
485   damaging (Table 2), particularly the S1202R (Figure 1b and 2b) mutation which has been
486   studied the most due to its numerous biological defects *in vivo*. The same residue mutations of
487   the Rad50 signature motif in yeast (S12025R) and human (S1202R) also generated complexes
488   that were significantly diminished in adenylate kinase (AK) activity that was important for DNA
489   tethering (Bhaskara et al., 2007). Previously, AK deficiency was found to be associated with
490   anaemia and several cases of mental retardation and psychomotor impairment (Abrusci et al.,
491   2007), which may explain why disruption of the MRN complex also causes this phenotype on
492   patients (Waltes et al., 2009). In addition, such deleterious mutation also contributed to inviable
493   spores and significant telomere shortening in *S. cerevisiae* (Bhaskara et al., 2007). Defects in
494   telomere length in human have been known to cause the pathology of several age-related
495   diseases and premature ageing syndrome, as well as cancer and other human diseases such as
496   Hoyeraal-Hreidarsson syndrome, Coats plus syndrome, pulmonary fibrosis, dyskeratosis
497   congenita, liver fibrosis and aplastic anemia (Blasco, 2005).
498
499   Additionally, most mutations such as G1199E, S1202A/R/M, and Q1205E (Figure 1b and 2b) at
500   the Rad50 signature motif (Figure 2a) were identified to affect PKA phosphorylation site (Table
501   2) suggesting that this site is dependent upon the function of Rad50 signature motif.
502   Phosphorylation is one of the most ubiquitous and important post translational modifications of
503   proteins, and implicated in almost all kinds of cellular processes and pathways (Ptacek & Snyder,
504   2006). In neurons, enhanced PKA signalling promotes neuronal development, enhances synaptic
505   plasticity, and elevates dopamine synthesis (Dagda & Das Banerjee, 2015). However a
506   deterioration in PKA signalling has contributed to the aetiology of several neurodegenerative
507   diseases, such as Alzheimer and Parkinson (Dagda & Das Banerjee, 2015). We hypothesized that
508   the defective PKA functional sites may also lead to Nijmegen breakage syndrome associated
509   with neurological phenotype in Rad50 mutations (Waltes et al., 2009), however this potential
510   phosphorylation sites remain to be validated.
511
512   C681A and C684R mutations (Figure 1b and 2d) at the zinc hook motif (Figure 2c) were
513   identified deleterious from our analysis (Table 2) and these mutations were known to lead severe
514   defects in various DNA damage response (DDR) such as ataxia-telangiectasia mutated (ATM)
515   protein activation, homologous recombinant, irradiation sensitivity and ataxia telangiectasia and

516 Rad3 related (ATR) protein activation (He et al., 2012). These findings were consistent with our
517 bioinformatics analysis where C684S deleterious mutation at zinc hook (Figure 1b and 2d) might
518 affect a protein kinase called ataxia telangiectasia mutated (ATM) that belongs to the
519 phosphatidylinositol 3-kinase-related kinase (PIKK) family (Table 2). The ATM protein was
520 known to cause devastating ataxia-telangiectasia syndrome which is characterized by progressive
521 neurological disorder, impaired organ maturation and immunodeficiency (Shiloh & Ziv, 2013).
522 Rad50 phosphorylated ATM at S635 site (Figure 1b and 2d) of which the mutation on this site
523 showed its importance for cell cycle control signalling and DNA repair mechanism (Gatei et al.,
524 2011).

526 P682E mutation at the zinc hook motif (Figure 1b and 2d) was shown to be deleterious (Table 2),
527 where previous study has reported that the double mutation P682E and S679R at the zinc hook
528 motif have reduced zinc affinity and dimerization efficiency leading to mice lethality (Roset et
529 al., 2014). In addition, crossbreeding P682E and S679R mutant mice with wildtype mice produce
530 offsprings with hydrocephalus (accumulation of cerebrospinal fluid within the brain), defects in
531 hematopoietic stem cells and gametogenic cells. This suggests that the hook motif has strong
532 influence on the MRN complex associated with DDR signalling, tissue homeostasis and
533 tumorigenesis, as well as fertility of the organism (Roset et al., 2014). This is consistent with the
534 mutations in the yeast hook domain that has increased chromosomal fragmentation (Cahill &
535 Carney, 2007), suggesting its presence is required for the binding or tethering of chromosomal
536 ends.

538 **Limitations of *in silico* prediction**

540 Several mutations were functionally predicted to be neutral, in contrast with the previous
541 experimental findings. For example, a few mutations i.e. S635G *(H. sapiens)*, S679R, C680N,
542 P682A, V683I *(S. cerevisiae)*, V683R *(M. musculus)* and Q685S *(S. cerevisiae)* (Figure 1b and
543 2d) located at the zinc hook domain (Table 2) and mutations on K22M *(S. cerevisiae and M.
544 musculus)*, R83I *(S. cerevisiae and M. musculus)*, T191E, C221E and S106E *(T. maritima)* in the
545 ATPase domain (Figure 1b, 2b and Table 2) were experimentally validated to be deleterious
546 (some causing embryonic lethality, growth defect, cancer predisposition, as well as
547 hematopoietic and spermatogenic depletion *in vivo* (Bender et al., 2002). A few previous studies
548 have also shown discrepancies between computer prediction and experimental data. For
549 example, an extensive *in silico* analysis using PolyPhen2 and MutPred tools of the ATP-binding
550 cassette transporter ABCA1, an important target in anti-atherosclerosis treatment predicted that
551 several nsSNPs can be neutral, contradicting with previous experimental data findings (Marín-
552 Martín et al., 2014). Furthermore, another *in silico* analysis performed using PolyPhen and SIFT
553 on proteins related to several hereditary diseases such as glucose-6-phosphate dehydrogenase
554 deficiency (G6PD), the receptor 1 for tumor necrosis factor-(TNFRSF1A), and familial
555 mediterranean fever (MEFV) has concluded that some nsSNPs impact may also not be predicted

556   deleterious to correspond to previous phenotypic effect (Tchernitchko, Goossens & Wajcman,
557   2004). Moreover, *in silico* identification of PmrAB virulence targets in *Salmonella*
558   *typhimurium* also demonstrated false positive prediction when validated experimentally (Marchal
559   et al., 2004) suggesting that more work has to be done to develop a more accurate bioinformatics
560   prediction platforms in the future. In contrast, various SNP prediction software have predicted
561   that these mutations were not damaging (Table 2).

563   Such discrepancy between the computational prediction and experimental results may be due to
564   several limitations in the bioinformatic tools used in our analysis. Several web-based prediction
565   tools may supply conflicting results (Wan et al., 2008) and even with an integrated predictor,
566   PredictSNP (Bendl et al., 2014), it is also limited by the differences in algorithms, principles,
567   training datasets and information used. For example, MAPP, PANTHER and SIFT in the
568   PredictSNP used alignment scores for functional prediction whereas SNAP, PoplyPhen-1 and
569   PolyPhen-2 used neural network, support vector machine and Naïve Bayes algorithm,
570   respectively (Bendl et al., 2014). Interestingly, we identified that the software predicts most
571   accurately (in agreement with experimental results) for the motifs or sites located at the highly
572   conserved position (Figure 1b). Conversely, most residues that were predicted to be neutral are
573   located at non-conserved positions in the Rad50 protein (S3 Figure and S7 Table), suggesting
574   that these prediction software may have only been trained and preferentially biased towards
575   conserved regions (Gardner et al., 2017). This suggests that computer prediction should also
576   consider and take into account the effect of non-conserved regions outside the motifs/domains
577   too for future improvement in their algorithms. Furthermore, any subsequent prediction studies
578   should also be aware of this limitation (whether located in conserved or non-conserved regions)
579   to carefully deduce the function of their protein mutation of interest.

581   Nonetheless, we cannot rule out the possibility that these mutations derived from other
582   organisms may not be readily affected or transferred to other organisms including human. This is
583   because certain organisms may possess gene compensation to compensate or mask the effect of
584   such mutations and that the different proteins from different organisms may not have perfectly
585   superimposable function. Hence, future experiments should focus on their validation especially
586   in human cell line studies to better understand the roles of these mutated residues in Rad50
587   function.

Conclusions

592   This study compiled all mutations to date in Rad50 proteins from various organisms and predicts
593   their effects using various software tools such as PredictSNP, MutPred, SNPeffect, I-Mutant and
594   MUpro. Most predictions for SNPs occurring within conserved regions are in agreement with
595   their corresponding *in vivo* or *in vitro* experimental results. However, SNPs located at non-

596 conserved regions are less likely to be accurately predicted, and as such algorithms for these
597 software should be improved in future studies. Altogether this study has provided means to
598 prioritized mutations particularly in Rad50 protein that have biologically meaningful function for
599 DNA double-stranded maintenance.
600

Acknowledgements

602 The authors would like to acknowledge and thank INBIOSIS and Makmal Genomik 1 research
603 teams from Universiti Kebangsaan Malaysia for their help with laboratory and technical
604 analyses.

605

References

607 Abrusci, P., Chiarelli, L.R., Galizzi, A., Fermo, E., Bianchi, P., Zanella, A. & Valentini, G.
608   (2007). Erythrocyte adenylate kinase deficiency: characterization of recombinant mutant
609   forms and relationship with nonspherocytic hemolytic anemia. *Experimental Hematology,*
610   35(8): 1182–1189.
611 Adzhubei, I., Jordan, D.M. & Sunyaev, S.R. (2013). Predicting functional effect of human
612   missense mutations using PolyPhen-2. *Current Protocols in Human Genetics,* 7(Supp76)
613   :20.
614 Alani, E., Padmore, R. & Kleckner, N. (1990). Analysis of wild-type and rad50 mutants of yeast
615   suggests an intimate relationship between meiotic chromosome synapsis and recombination.
616   *Cell,* 61(3): 419–436.
617 Assenmacher, N. & Hopfner, K.P. (2004). MRE11/RAD50/NBS1: Complex activities.
618   *Chromosoma,* 113(4): 157–166.
619 De Baets, G., Van Durme, J., Reumers, J., Maurer-Stroh, S., Vanhee, P., Dopazo, J.,
620   Schymkowitz, J. & Rousseau, F. (2012). SNPeffect 4.0: On-line prediction of molecular and
621   structural effects of protein-coding variants. *Nucleic Acids Research,* 40(D1): 935-939.
622 Barbi, G., Scheres, J.M.J.C., Schindler, D., Taalman, R.D.F.M., Rodens, K., Mehnert, K.,
623   Müller, M., Seyschab, H., Muller, M. & Seyschab, H. (1991). Chromosome instability and
624   X-ray hypersensitivity in a microcephalic and growth-retarded child. *American Journal of*
625   *Medical Genetics,* 40(1): 44–50.
626 Bartoszewski, R. a., Jablonsky, M., Bartoszewska, S., Stevenson, L., Dai, Q., Kappes, J.,
627   Collawn, J.F. & Bebok, Z. (2010). A synonymous single nucleotide polymorphism in delta
628   F508 CFTR alters the secondary structure of the mRNA and the expression of the mutant
629   protein. *Journal of Biological Chemistry,* 285(37): 28741–28748.
630 Bell, D.W., Varley, J.M., Szydlo, T.E., Kang, D.H., Wahrer, D.C.R., Shannon, K.E.,
631   Lubratovich, M., Verselis, S.J., Isselbacher, K.J., Fraumeni, J.F., Birch, J.M., Li, F.P.,
632   Garber, J.E. & Haber, D.A. (1999). Heterozygous germ line hCHK2 mutations in Li-
633   Fraumeni syndrome. *Science,* 286(5449): 2528-2531.

634   Bender, C.F., Sikes, M.L., Sullivan, R., Huye, L.E., Le Beau, M.M., Roth, D.B., Mirzoeva, O.K.,
635   Oltz, E.M. & Petrini, J.H.J. (2002). Cancer predisposition and hematopoietic failure in
636   Rad50S/S mice. *Genes and Development,* 16(17): 2237–2251.
637   Bendl, J., Stourac, J., Salanda, O., Pavelka, A., Wieben, E.D., Zendulka, J., Brezovsky, J. &
638   Damborsky, J. (2014). PredictSNP: Robust and Accurate Consensus Classifier for Prediction
639   of Disease-Related Mutations. *PLoS Computational Biology*, 10(1): 1003440.
640   Bhaskara, V., Dupr, A., Lengsfeld, B., Hopkins, B.B., Chan, A., Lee, J.H., Zhang, X., Gautier, J.,
641   Zakian, V. & Paull, T.T. (2007). Rad50 Adenylate Kinase Activity Regulates DNA
642   Tethering by Mre11/Rad50 Complexes. *Molecular Cell,* 25(5): 647–661.
643   Blasco, M.A. (2005). Telomeres and human disease: Ageing, cancer and beyond. Nature
644   Reviews *Genetics,* 6(8): 611–622.
645   van den Bosch, M., Bree, R.T. & Lowndes, N.F. (2003). The MRN complex: Coordinating and
646   mediating the response to broken chromosomes. *EMBO Reports,* 4(9): 844-849.
647   Bromberg, Y. & Rost, B. (2007). SNAP: Predict effect of non-synonymous polymorphisms on
648   function. *Nucleic Acids Research,* 35(11): 3823–3835.
649   Cahill, D. & Carney, J.P. (2007). Dimerization of the Rad50 protein is independent of the
650   conserved hook domain. *Mutagenesis,* 22(4): 269–274.
651   Capriotti, E. & Fariselli, P. (2017). PhD-SNPg: A webserver and lightweight tool for scoring
652   single nucleotide variants. *Nucleic Acids Research,* 45(W1): W247–W252.
653   Capriotti, E., Fariselli, P. & Casadio, R. (2005). I-Mutant2.0: Predicting stability changes upon
654   mutation from the protein sequence or structure. *Nucleic Acids Research,* 33(SUPPL. 2):
655   306-310.
656   Cargill, M., Altshuler, D., Ireland, J., Sklar, P., Ardlie, K., Patil, N., Lane, C.R., Lim, E.P.,
657   Kalyanaraman, N., Nemesh, J., Ziaugra, L., Friedland, L., Rolfe, A., Warrington, J.,
658   Lipshutz, R., Daley, G.Q., Lander, E.S., Cargill, M., Sklar, P., Ireland, J., Ardlie, K., Patil,
659   N., Lane, C.R., Lim, E.P., Kalyanaraman, N., Nemesh, J., Ziaugra, L., Friedland, L., Rolfe,
660   A., Warrington, J., Lipshutz, R. & Daley, G.Q. (1999). Characterization of single-nucleotide
661   polymorphisms in coding regions of human genes. *Nature,* 22(July): 231–238.
662   Carney, J.P., Maser, R.S., Olivares, H., Davis, E.M., Le Beau, M., Yates, J.R., Hays, L., Morgan,
663   W.F. & Petrini, J.H. (1998a). The hMre11/hRad50 protein complex and Nijmegen breakage
664   syndrome: linkage of double-strand break repair to the cellular DNA damage response. *Cell,*
665   93(3): 477–86.
666   Carney, J.P., Maser, R.S., Olivares, H., Davis, E.M., Le Beau, M., Yates, J.R., Hays, L., Morgan,
667   W.F. & Petrini, J.H.J. (1998b). The hMre11/hRad50 protein complex and Nijmegen
668   breakage syndrome: Linkage of double-strand break repair to the cellular DNA damage
669   response. *Cell,* 93(3): 477–486.
670   Chen, L., Trujillo, K.M., Van Komen, S., Roh, D.H., Krejci, L., Lewis, L.K., Resnick, M.A.,
671   Sung, P. & Tomkinson, A.E. (2005). Effect of amino acid substitutions in the rad50 ATP
672   binding domain on DNA double strand break repair in yeast. *The Journal of Biological*
673   *Chemistry,* 280(4): 2620–2627.

674   Cheng, J., Randall, A. & Baldi, P. (2006). Prediction of protein stability changes for single-site
675   mutations using support vector machines. *Proteins,* 62(4): 1125–1132.
676   Collins, F.S., Guyer, M.S. & Charkravarti, A. (1997). Variations on a theme: cataloging human
677   DNA sequence variation. *Science*, 278(5343): 1580–1.
678   Dagda, R.K. & Das Banerjee, T. (2015). Role of protein kinase A in regulating mitochondrial
679   function and neuronal development: Implications to neurodegenerative diseases. *Reviews in*
680   *the Neurosciences,* 26(3): 359-370.
681   Davidson, A.L., Dassa, E., Orelle, C. & Chen, J. (2008). Structure, function, and evolution of
682   bacterial ATP-binding cassette systems. *Microbiology and Molecular Biology Reviews,*
683   72(2): 317–364.
684   Durocher, D., Taylor, I. a, Sarbassova, D., Haire, L.F., Westcott, S.L., Jackson, S.P., Smerdon,
685   S.J. & Yaffe, M.B. (2000). The molecular basis of FHA domain:phosphopeptide binding
686   specificity and implications for phospho-dependent signaling mechanisms. *Molecular cell,*
687   6(5): 1169-1182.
688   Featherstone, C. & Jackson, S.P. (1998). DNA repair: the Nijmegen breakage syndrome protein.
689   *Current biology,* CB 8(17): R622-5.
690   Finn, R.D., Attwood, T.K., Babbitt, P.C., Bateman, A., Bork, P., Bridge, A.J., Chang, H.Y.,
691   Dosztanyi, Z., El-Gebali, S., Fraser, M., Gough, J., Haft, D., Holliday, G.L., Huang, H.,
692   Huang, X., Letunic, I., Lopez, R., Lu, S., Marchler-Bauer, A., Mi, H., Mistry, J., Natale,
693   D.A., Necci, M., Nuka, G., Orengo, C.A., Park, Y., Pesseat, S., Piovesan, D., Potter, S.C.,
694   Rawlings, N.D., Redaschi, N., Richardson, L., Rivoire, C., Sangrador-Vegas, A., Sigrist, C.,
695   Sillitoe, I., Smithers, B., Squizzato, S., Sutton, G., Thanki, N., Thomas, P.D., Tosatto,
696   S.C.E., Wu, C.H., Xenarios, I., Yeh, L.S., Young, S.Y. & Mitchell, A.L. (2017). InterPro in
697   2017-beyond protein family and domain annotations. *Nucleic Acids Research*, 45(D1):
698   D190–D199.
699   Gardner, P.P., Paterson, J.M., Ghomi, F.A., Umu, S.U.U., McGimpsey, S. & Pawlik, A. (2017).
700   A meta-analysis of bioinformatics software benchmarks reveals that publication-bias unduly
701   influences software accuracy. *bioRxiv,* 092205.
702   Gatei, M., Jakob, B., Chen, P., Kijas, A.W., Becherel, O.J., Gueven, N., Birrell, G., Lee, J.H.,
703   Paull, T.T., Lerenthal, Y., Fazry, S., Taucher-Scholz, G., Kalb, R., Schindler, D., Waltes, R.,
704   Drk, T. & Lavin, M.F. (2011). ATM protein-dependent phosphorylation of Rad50 protein
705   Regulates DNA repair and cell cycle control. *Journal of Biological Chemistry,* 286(36):
706   31542–31556.
707   Halushka, M.K., Fan, J.B., Bentley, K., Hsie, L., Shen, N., Weder, A., Cooper, R., Lipshutz, R.
708   & Chakravarti, A. (1999). Patterns of single-nucleotide polymorphisms in candidate genes
709   for blood-pressure homeostasis. *Nature Genetics,* 22(3): 239–247.
710   He, J., Shi, L.Z., Truong, L.N., Lu, C.S., Razavian, N., Li, Y., Negrete, A., Shiloach, J., Berns,
711   M.W. & Wu, X. (2012). Rad50 zinc hook is important for the Mre11 complex to bind
712   chromosomal DNA double-stranded breaks and initiate various DNA damage responses.
713   *Journal of Biological Chemistry,* 287(38): 31747–31756.

714   Hopfner, K.-P., Craig, L., Moncalian, G., Zinkel, R. a, Usui, T., Owen, B. a L., Karcher, A.,
715   Henderson, B., Bodmer, J.-L., McMurray, C.T., Carney, J.P., Petrini, J.H.J. & Tainer, J.A.
716   (2002). The Rad50 zinc-hook is a structure joining Mre11 complexes in DNA recombination
717   and repair. *Nature,* 418(6897): 562–566.

718   de Jager, M., van Noort, J., van Gent, D.C., Dekker, C., Kanaar, R. & Wyman, C. (2001).
719   Human Rad50/Mre11 is a flexible complex that can tether DNA ends. *Molecular cell,* 8(5):
720   1129–35.

721   Jones, P., O'Mara, M. & George, A. (2009). ABC transporters: a riddle wrapped in a mystery
722   inside an enigma. *Trends in biochemical sciences*, 34(10): 520-31.

723   Kelley, L.A., Mezulis, S., Yates, C.M., Wass, M.N. & Sternberg, M.J.E. (2015). The Phyre2 web
724   portal for protein modeling, prediction and analysis. *Nature Protocols,* 10(6): 845–858.

725   Kerem, B., Rommens, J.M., Buchanan, J. a, Markiewicz, D., Cox, T.K., Chakravarti, a,
726   Buchwald, M. & Tsui, L.C. (1989). Identification of the cystic fibrosis gene: genetic
727   analysis. *Science,* 245(4922): 1073–1080.

728   Koroleva, O., Makharashvili, N., Courcelle, C.T., Courcelle, J. & Korolev, S. (2007). Structural
729   conservation of RecF and Rad50: implications for DNA recognition and RecF function. *The*
730   *EMBO journal,* 26(3): 867–77.

731   De La Rosa Metzere Bierlein & and Scott W. Nelson. (2011). An Interaction between the Walker
732   A and D-loop Motifs Is Critical to ATP Hydrolysis and Cooperativity in Bacteriophage. *The*
733   *Journal of Biological Chemistry,* 286(29): 26258–26266.

734   Lee, J.E., Choi, J.H., Lee, J.H. & Lee, M.G. (2005). Gene SNPs and mutations in clinical genetic
735   testing: Haplotype-based testing and analysis. *Mutation Research - Fundamental and*
736   *Molecular Mechanisms of Mutagenesis,* 573(1-2): 195-204.

737   Luo, G., Yao, M.S., Bender, C.F., Mills, M., Bladl, A.R., Bradley, A. & Petrini, J.H. (1999).
738   Disruption of mRad50 causes embryonic stem cell lethality, abnormal embryonic
739   development, and sensitivity to ionizing radiation. *Proceedings of the National Academy of*
740   *Sciences of the United States of America,* 96(13): 7376–81.

741   Marchal, K., De Keersmaecker, S., Monsieurs, P., van Boxel, N., Lemmens, K., Thijs, G.,
742   Vanderleyden, J. & De Moor, B. (2004). In silico identification and experimental validation
743   of PmrAB targets in Salmonella typhimurium by regulatory motif detection. *Genome*
744   *Biology*, 5(2): R9

745   Marín-Martín, F.R., Soler-Rivas, C., Martín-Hernández, R. & Rodriguez-Casado, A. (2014). A
746   Comprehensive In Silico Analysis of the Functional and Structural Impact of
747   Nonsynonymous SNPs in the ABCA1 Transporter Gene. *Cholesterol,* 2014: 1–19.

748   Matsuura, S., Tauchi, H., Nakamura, A., Kondo, N., Sakamoto, S., Endo, S., Smeets, D., Solder,
749   B., Belohradsky, B.H., Der Kaloustian, V.M., Oshimura, M., Isomura, M., Nakamura, Y. &
750   Komatsu, K. (1998). Positional cloning of the gene for Nijmegen breakage syndrome.
751   *Nature genetics, 19*(2): 179-181.

752 Moncalian, G., Lengsfeld, B., Bhaskara, V., Hopfner, K.P., Karcher, A., Alden, E., Tainer, J.A.
753   & Paull, T.T. (2004). The Rad50 Signature Motif: Essential to ATP Binding and Biological
754   Function. *Journal of Molecular Biology,* 335(4): 937–951.
755 Pejaver, V., Urresti, J., Lugo-Martinez, J., Pagel, K.A., Lin, G.N., Nam, H.-J., Mort, M., Cooper,
756   D.N., Sebat, J., Iakoucheva, L.M., Mooney, S.D. & Radivojac, P. (2017). MutPred2:
757   inferring the molecular and phenotypic impact of amino acid variants. *bioRxiv,* :134981.
758 Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C. &
759   Ferrin, T.E. (2004). UCSF Chimera--a visualization system for exploratory research and
760   analysis. *Journal Comput Chemistry,* 25(13): 1605–1612.
761 Piro, R.M. & Di Cunto, F. (2012). Computational approaches to disease-gene prediction:
762   Rationale, classification and successes. *FEBS Journal,* 279(5): 678-696.
763 Ptacek, J. & Snyder, M. (2006). Charging it up: global analysis of protein phosphorylation.
764 Ramensky, V. 2002. Human non-synonymous SNPs: server and survey. *Nucleic Acids Research,*
765   30(17): 3894–3900.
766 Risch, N. & Merikangas, K. (1996). The future of genetic studies of complex human diseases.
767   *Science,* 273(5281): 1516–1517.
768 Rojowska, A., Lammens, K., Seifert, F.U., Direnberger, C. & Feldmann, H. (2014). Structure of
769   the Rad 50 DNA double-strand break repair protein in complex with DNA. *The EMBO*
770   *journal,* 33(23): 2847–2859.
771 Roset, R., Inagaki, A., Hohl, M., Brenet, F., Lafrance-Vanasse, J., Lange, J., Scandura, J.M.,
772   Tainer, J.A., Keeney, S. & Petrini, J.H.J. (2014). The Rad50 hook domain regulates DNA
773   damage signaling and tumorigenesis. *Genes and Development,* 28(5): 451–462.
774 Schymkowitz, J., Borg, J., Stricher, F., Nys, R., Rousseau, F. & Serrano, L. (2005). The FoldX
775   web server: An online force field. *Nucleic Acids Research,* 33(SUPPL. 2).
776 Shaikho, E.M., Farrell, J.J., Alsultan, A., Qutub, H., Al-Ali, A.K., Figueiredo, M.S., Chui,
777   D.H.K., Farrer, L.A., Murphy, G.J., Mostoslavsky, G., Sebastiani, P. & Steinberg, M.H.
778   (2017). A phased SNP-based classification of sickle cell anemia HBB haplotypes. *BMC*
779   *Genomics*, 18(1): 608.
780 Shiloh, Y. & Ziv, Y. (2013). The ATM protein kinase: Regulating the cellular response to
781   genotoxic stress, and more. *Nature reviews. Molecular cell biology,* 14(4): 197-210.
782 Sievers, F. & Higgins, D.G. (2018). Clustal Omega for making accurate alignments of many
783   protein sequences. *Protein Science,* 27(1):135-145.
784 Sim, N.-L., Kumar, P., Hu, J., Henikoff, S., Schneider, G. & Ng, P.C. (2012). SIFT web server:
785   predicting effects of amino acid substitutions on proteins. *Nucleic Acids Research*, 40(W1):
786   W452–W457.
787 Stone, E.A. & Sidow, A. (2005). Physicochemical constraint violation by missense substitutions
788   mediates impairment of protein function and disease severity. *Genome Research,* 15(7):
789   978–986.
790 Symington, L.S. (2002). Role of RAD52 epistasis group genes in homologous recombination and
791   double-strand break repair. *Microbiology and molecular biology reviews,* 66(4): 630–70.

792 Tchernitchko, D., Goossens, M. & Wajcman, H. (2004). In silico prediction of the deleterious
793   effect of a mutation: Proceed with caution in clinical genetics. *Clinical Chemistry*, 50(11):
794   1974-1978.
795 Thompson, J.D., Gibson, T.J. & Higgins, D.G. (2002). Multiple Sequence Alignment Using
796   ClustalW and ClustalX. *Current Protocols in Bioinformatics*, 0(1): 2.3.1-2.3.22.
797 Traeger, J., Wood, W.G., Clegg, J.B., Weatherall, D.J. & Wasi, P. (1980). Defective synthesis of
798   HbE is due to reduced levels of βE mRNA. *Nature* 288(5790): 497–499.
799 Tranchevent, L.C., Capdevila, F.B., Nitsch, D., de Moor, B., de Causmaecker, P. & Moreau, Y.
800   (2011). A guide to web tools to prioritize candidate genes. *Briefings in Bioinformatics,*
801   12(1): 22–32.
802 Varon, R., Vissinga, C., Platzer, M., Cerosaletti, K.M., Chrzanowska, K.H., Saar, K., Beckmann,
803   G., Seemanová, E., Cooper, P.R., Nowak, N.J., Stumm, M., Weemaes, C.M., Gatti, R. a,
804   Wilson, R.K., Digweed, M., Rosenthal, a, Sperling, K., Concannon, P. & Reis, a. (1998).
805   Nibrin, a novel DNA double-strand break repair protein, is mutated in Nijmegen breakage
806   syndrome. *Cell*, 93(3): 467-476.
807 Waltes, R., Kalb, R., Gatei, M., Kijas, A.W., Stumm, M., Sobeck, A., Wieland, B., Varon, R.,
808   Lerenthal, Y., Lavin, M.F., Schindler, D. & Dörk, T. (2009). Human RAD50 Deficiency in a
809   Nijmegen Breakage Syndrome-like Disorder. *American Journal of Human Genetics,* 84(5):
810   605–616.
811 Wan, J., Kang, S., Tang, C., Yan, J., Ren, Y., Liu, J., Gao, X., Banerjee, A., Ellis, L.B.M. & Li,
812   T. (2008). Meta-prediction of phosphorylation sites with weighted voting and restricted grid
813   search parameter selection. *Nucleic Acids Research,* 36(4): 22.

Figure 1

Domain analysis and multiple sequence alignment.

Domain analysis using InterPro shows that Rad50 contains P-loop containing nucleoside triphosphate hydrolase domain belongs to the ATP Binding Cassette (ABC) protein superfamily (pink box) as well as a special domain called zinc hook, which so far does not overlap with any homologous superfamilies (blue box) (a). ABC protein consists of six conserved motifs; i.e. Walker A (WA), Q-loop (QL), signature motif (SM), Walker B (WB), D-loop (DL), and H-loop (HL) which make up the nucleotide binding domain. Zinc hook domain contains a conserved CxxC motif located at the residue number 681-684 (a). All deleterious residues identified from the literature were highlighted based on human equivalent mutation (Supplementary Table S1) and those occurring only in the conserved regions are shown in (b). Multiple sequence alignment (MSA) analysis of Rad50 sequences dataset (human, *D. rerio* (zebrafish), *M. musculus* (mouse), *R. norvegicus* (rat), *P. troglodytes* (chimpanzee), *O. cuniculus* (rabbit), *D. melanogaster* (fruit fly), *S. cerevisiae* (yeast) and *S. pombe* (yeast)) showed conserved residues in specific motifs (b). An "*" (asterisk) indicates position which has a single, fully conserved residue. A ":" (colon) indicates conservation between groups of strongly similar properties - scoring > 0.5 in the Gonnet PAM250 matrix. A "." (period) indicates conservation between groups of weakly similar properties - scoring ≤ 0.5 in the Gonnet PAM250 matrix.

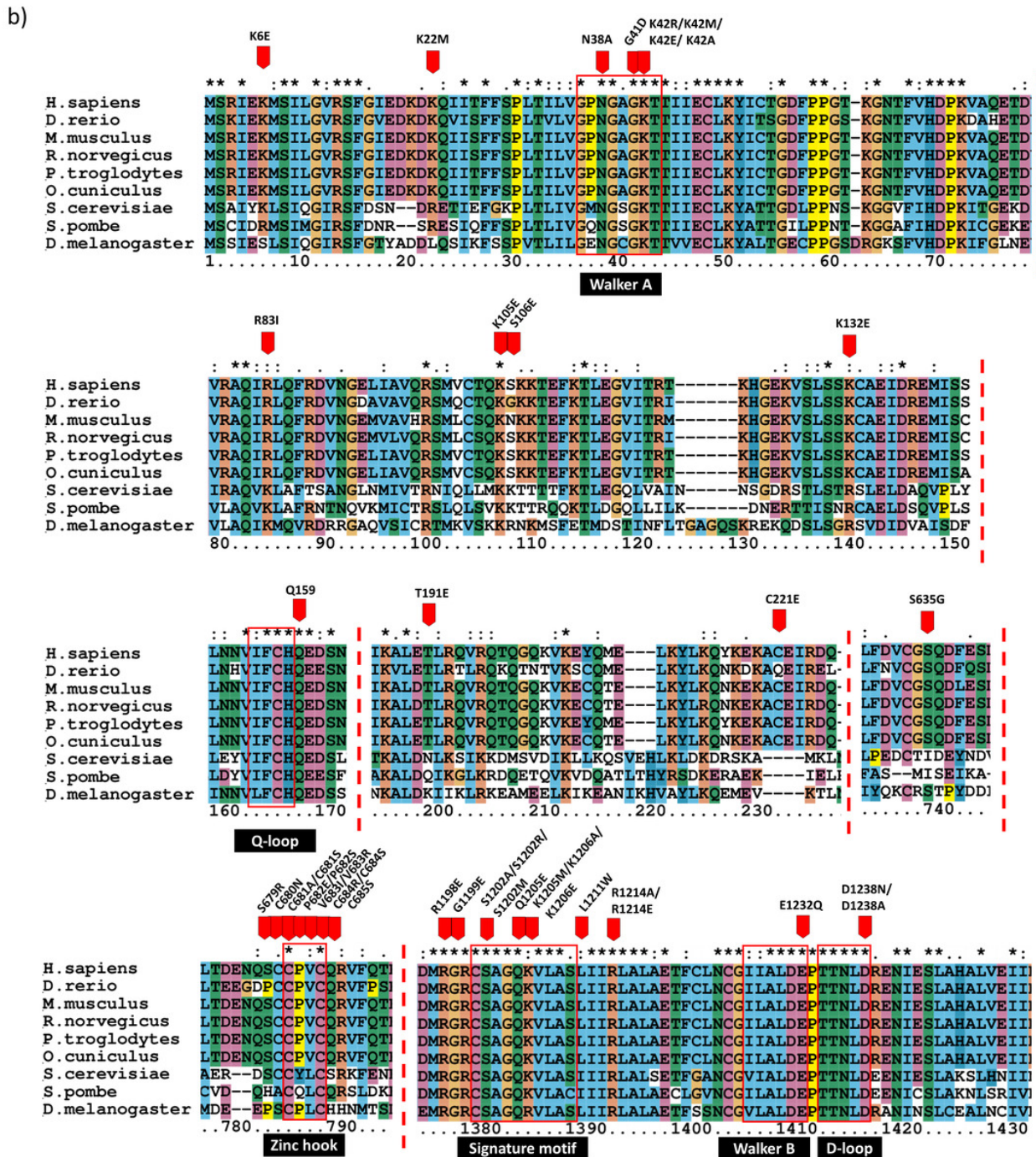
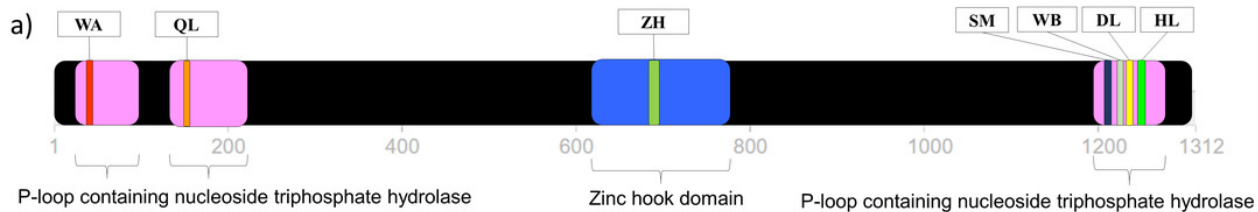



Figure 2

3D structure of Rad50.

3D structure of Rad50. A 3D structure of Rad50 human modelled using fold recognition technique Phyre2 using structure from *Chaetomium thermophilum* as a template (PDB ID: 5DAC). N-terminal of 276 residues (2-278) and C-terminal of 155 residues (1153-1306) are colored as blue and green; respectively (a). All six motifs identified are marked and represented by ball and stick representation with different colours (orange for Q-loop, blue for Walker B, purple for signature motif, yellow for D-loop, grey for Walker A and green from H-loop) (a). All mutated residues identified were marked and labelled in the 3D structure (b). Zinc hook structure of 181 residues (residue 585-766) that has been determined (PDB ID: 5GOX) and its deleterious residues also marked in the structure (c,d). All figures were generated using UCSF Chimera (Pettersen et al., 2004)

(a)

Unsolved structure D1101
C221

N-terminal (S2)

Q-loop

Walker B

Signature motif

D-loop

Walker A
H-loop

C-terminal (N1305)

Coiled-coil domain

Globular domain

(b)

Unsolved structure D1101
C221

(c)

K766
K585
K585
K766

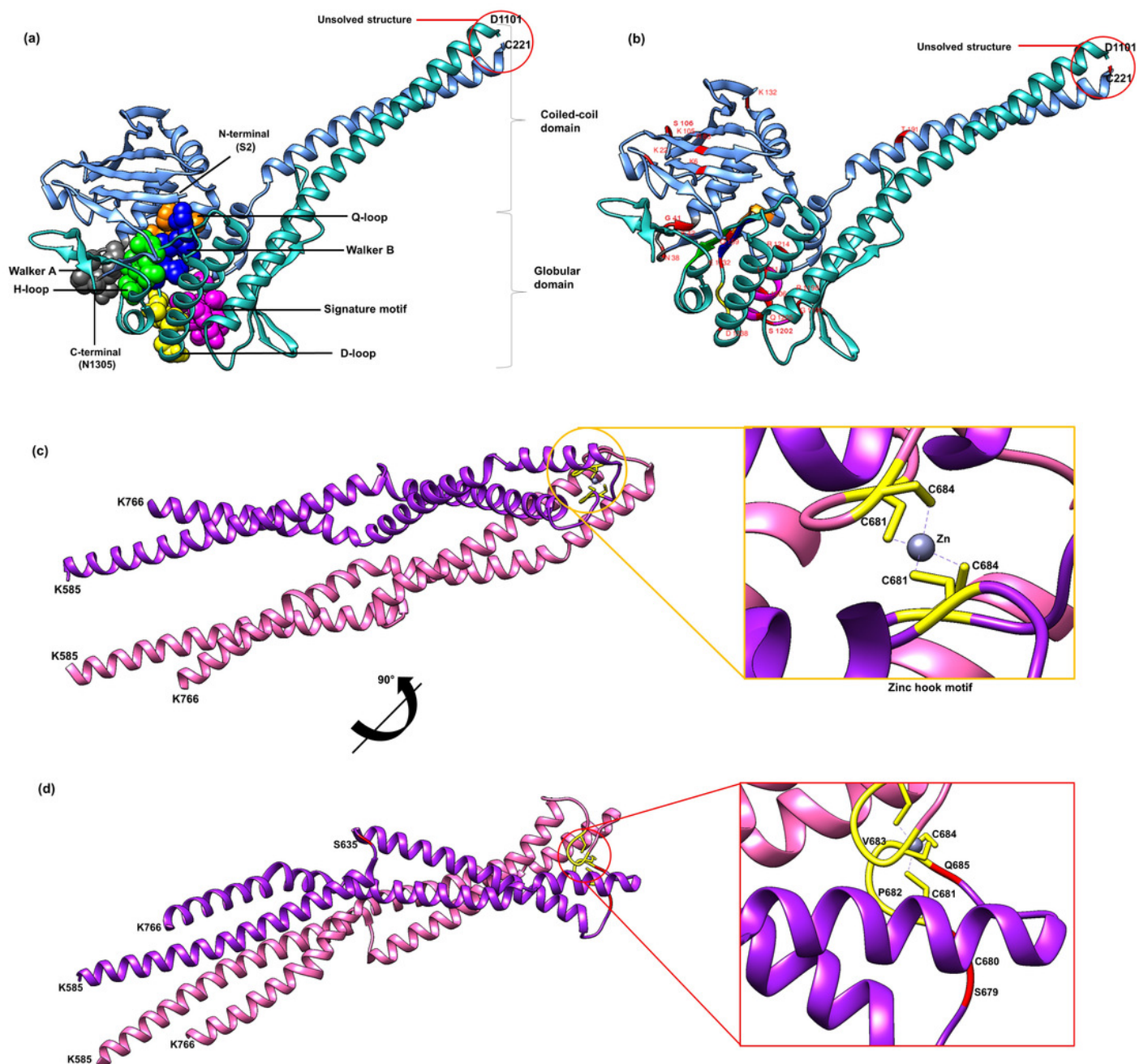
90°

C684
C681
Zn
C681 C684

Zinc hook motif

(d)

S635
K766
K585
K766
K585

V683 C684
Q685
P682 C681
C680
S679

Summary of the most damaging effects of Rad50 mutations obtained from previous *in vitro* and *in vivo* experiments*.*

Abbreviations: HR (homologous recombination), NHEJ (non-homologous end joining repair), CFTR (cystic fibrosis transmembrane conductance regulator), ATP (adenosine tri-phosphate), ATM (ataxia-telangiectasia mutated), ATR (ATM-and Rad3-Related). Refer to Table S2 for the description of all mutations.

Motif/domain	Mutations	Organism	Effects	Ref.
Walker A	K40A/R/E	*S. cerevisiae*	• HR and NHEJ defects and lower ATPase activity	(Chen *et al.*, 2005)
Walker A D-loop	N38A, D512N/A	T4 bacteriophage	• Naturally occurring mutation of CFTR protein • Reduce in ATP activity	(De La Rosa Metzere Bierlein and and Scott W. Nelson, 2011)
ATP binding domain and Walker A	G39D, K40E K81I, R20M	*S. cerevisiae*	• Total defect in formation of viable spore	(Alani, Padmore and Kleckner, 1990)
ATP binding domain	K6E, K22M, R83I	*M. musculus*	• Embryonic lethality, growth defect, cancer predisposition, hematopoietic and spermatogenic depletion	(Bender *et al.*, 2002)
Walker A	K39R, K42M	*D. radiodurans*	• Prevented ATP binding and hydrolysis	(Koroleva *et al.*, 2007)
ATPase binding domain, Walker B and Signature motif	K115E, K175E, K182E, R94E, K95E, R765E	*T. maritima*	*In vitro*: *Thermotoga maritima* • K175E, K182E, K115E Reduced DNA binding • R94E and K95E: Important for DNA binding • R765E: Diminished DNA binding • E798Q: Low affinity to DNA • S768R: Reduced DNA binding	(Rojowska *et al.*, 2014b)
	E798Q, S768R, K103E, K104E, R131E, R1202E, S1205R, E1235Q	*S. cerevisiae*	*In vivo*: *Saccharomyces cerevisiae* • S1205R and E1235Q double mutation: Unable to rescue the impaired DNA damage response • K103E, K104E and R131E: Strongly affected DNA binding and moderate reduction in telomere length • K103E and R131R (double mutation) and R1201E: Significantly reduced telomere length • S1205R: Significantly reduced telomere length	
Zinc hook	S679R, P682E V683R	*M. musculus*	• Lethality in mice. Hydrocephalus, defects in primitive hematopoietic and gametogenic cells	(Roset *et al.*, 2014)
	C684N, C685A P686A V6871 C688R Q689S	*S. cerevisiae*	• Defective to be recruited to chromosomal double strand break • Phenotype as severe as Rad50 null mutant • Defective in ATM activation, HR, sensitive to irradiation and ATR activation	(He *et al.*, 2012)
	C288S, C291S	T4 bacteriophage	• Double mutation is lethal	(Barfoot *et al.*, 2015)
	S635G	*H. sapiens*	• Chromosomal instability • Defective ATM-dependent signalling	(Gatei *et al.*, 2011)

	S685R Y688E L689R	*S. cerevisiae*	• S685R and Y688E double mutation: Sporulation efficiency and viability were severely impaired followed by L689R • Rad50-Mre11 interaction was strongly impaired, partial suppression of telomere and meiotic defects	(Hohl *et al.*, 2015)
Rad50 Signature motif	R805E L802W	*P. furiosus*	• L802W: Decrease dimerization in ATP, hydrolysis and cleavage site • R805E: Poorly grown in camptothecin; inability to repair endogenous DNA damage by HR and showed defect in resection in HO endonuclease induced	(Deshpande *et al.*, 2014)
	K1187A K1187E R1195A R1195E	*S. pombe*	• K1187A: Sensitive in higher dose of clastogens • K1187E, R1195A and R1195E: Significantly sensitive to clastogen agents and were deleterious as Rad50 null mutation	(Williams *et al.*, 2011)
	S471A/R/M, E474Q,K475M	T4 bacteriophage	• S471A/R.M, E474Q and K475M: Residues involved in the allosteric transmission between DNA and ATP binding sites	(Herdendorf and Nelson, 2011)
	S1205R S793R S1202R	*S. cerevisiae* *P. furiosus* *H. sapiens*	• S1202R: Reduced adenylate kinase • S793R: Deficient in ATP-dependent dimer formation and ATP binding • S1202R and S1205R: Low level of adenylate kinase • S1205R: Telomere shortening, not support spore viability	(Bhaskara *et al.*, 2007)
Signature motif and Q loop	S793R Q140H S1205R	*P. furiosus* *S. cerevisiae*	• S793R: Analogs to the mutation in CFTR (S549R) gene that results cystic fibrosis • S793R: Prevented ATP binding • S1205R *S. cerevisiae*: Failed to complement Rad50 deletion strain in DNA repair assay • S783R and Q140H: Halted ATP-dependent activities	(Moncalian, Lengsfeld, Bhaskara, Hopfner, Karcher, Alden, J. A. Tainer, *et al.*, 2004)
ATPase domain	R1093(stop) c.3939A/T	*H. sapiens*	• Nijmegen breakage syndrome like disorder (NBSLD)	(Waltes *et al.*, 2009)

In silico analysis of 42 deleterious mutations in Rad50.

Different tools were used to analyse all mutations as abbreviated in the table. PS
(PredictSNP), MP (MAPP), PhS (PhD-SNP), PP1 (Poly-Phen1), PP2 (Poly-Phen2), SF (SIFT), SN
(SNAP), IM (I-Mutant), MPr (MuPro), Pr (probability), AG (protein aggregation) and AM
(Amyloid aggregation). Please refer to Materials and Methods for detailed descriptions of
these tools. Note that all mutations listed above are based on the equivalent mutations in
human.

Motif	Mutation [source]	Amino acid Impact (PredictSNP) (Neutral/Deleterious)							Molecular mechanisms (MutPred2)				Structural phenotyping (SNPeffect)		Protein stability (Imutant/ MuPro)	
		PS	MP	Ph-S	PP-1	PP-2	SF	SP	Affected molecular mechanisms	Pr	P-value	Affected functional sites	AG	AM	IM	MPr
Walker A	N38A (De La Rosa Metzere Bierlein and and Scott W. Nelson, 2011)	D	D	D	D	D	D	D	Loss of catalytic site at N38 Loss of relative solvent accessibility Altered ordered interface Loss of allosteric site at N38 Altered DNA binding Altered metal binding Gain of methylation at K42	0.53 0.34 0.33 0.31 0.25 0.23 0.17	8.8e-05 3.4e-03 9.9e-03 3.0e-03 6.7e-03 0.02 8.4e-03	• ATP/GTP-binding site motif A (P-loop)	No	No	↓	↓
	G41D (Alani, Padmore and Kleckner, 1990)	D	D	D	D	D	D	D	Altered metal binding Gain of allosteric site at G41 Altered ordered interface Gain of relative solvent accessibility Gain of helix Loss of strand Altered DNA binding Loss of catalytic site at N38 Loss of methylation at K42	0.40 0.33 0.30 0.29 0.29 0.28 0.28 0.27 0.17	1.7e-04 6.4e-04 0.02 0.01 0.01 9.7e-03 4.3e-03 3.7e-03 0.01	• FHA phosphopeptide ligands • CK2 Phosphorylation site • N-myristoylation site • ATP/GTP-binding site motif A (P-loop)	No	No	↓	↓
	K42R (Chen et al., 2005) (Koroleva et al., 2007)	D	D	D	D	D	D	D	Loss of relative solvent accessibility Altered DNA binding Loss of allosteric site at T44 Loss of catalytic site at N38 Altered metal binding Loss of methylation at K42	0.30 0.28 0.28 0.27 0.25 0.20	9.6e-03 3.2e-03 7.3e-03 3.4e-03 0.01 6.9e-03	• FHA phosphopeptide ligands • PKA phosphorylation site • CK2 phosphorylation site • N-myristoylation site • ATP/GTP-binding site motif A (P-loop)	No	No	↓	↓
	K42M (Koroleva et al., 2007)	D	D	D	D	D	D	D	Altered DNA binding Loss of allosteric site at K42 Loss of relative solvent accessibility Altered ordered interface Gain of catalytic site at T43 Altered metal binding Loss of methylation at K42	0.37 0.36 0.34 0.32 0.29 0.28 0.20	7.1e-04 1.1e-03 3.1e-03 0.01 1.6e-03 5.8e-03 6.8e-03	• FHA phosphopeptide ligands • CK2 phosphorylation site • N-myristoylation site • ATP/GTP-binding site motif A (P-loop)	No	No	↑	↑
	K42E (Alani, Padmore and Kleckner, 1990) (Chen et al., 2005)	D	D	D	D	D	D	D	Altered metal binding Gain of catalytic site at T43 Altered DNA binding Loss of allosteric site at K42 Altered ordered interface Loss of relative solvent accessibility Gain of strand Loss of methylation at K42	0.44 0.33 0.33 0.30 0.29 0.29 0.27 0.20	4.5e-04 6.4e-04 1.3e-03 4.3e-03 0.02 0.01 0.03 6.8e-03	• FHA phosphopeptide ligands • CK2 phosphorylation site • Polo-like kinase phosphorylation sit • N-myristoylation site • ATP/GTP-binding site motif A (P-loop)	No	No	↓	↓
	K42A (Chen et al., 2005)	D	D	D	D	D	D	D	Loss of allosteric site at K42 Loss of relative solvent accessibility Altered DNA binding Altered ordered interface Gain of catalytic site at T43 Altered metal binding Loss of methylation at K42	0.53 0.37 0.37 0.36 0.32 0.31 0.20	5.4e-05 1.6e-03 6.0e-04 4.2e-03 8.7e-04 0.01 6.8e-03	• FHA phosphopeptide ligands • CK2 phosphorylation site • N-myristoylation site • ATP/GTP-binding site motif A (P-loop)	No	No	↓	↓
Q-loop	Q159H(Moncal ian, Lengsfeld, Bhaskara, Hopfner, Karcher, Alden, J. A. Tainer, et	D	D	D	D	D	D	D	No effect	-	-	None	No	No	↓	↓

	al., 2004)																
Zinc hook	S635G (Gatei et al., 2011)	N	N	N	D	N	N	N	No effect	-	-	None	No	No	↓	↓
	S679R (Roset et al., 2014)(Hohl et al., 2015)	N	N	N	N	N	D	N	No effect	-	-	None	No	No	↓	↓
	C680N (He et al., 2012)	N	N	N	D	N	D	D	Loss of N-linked glycosylation at N677	0.02	0.04	• N-glycosylation site	No	No	↓	↓
	C681A (He et al., 2012)	D	D	D	D	D	D	D	Gain of helix Gain of N-linked glycosylation at N677	0.30 0.02	8.0e-03 0.03	• N-glycosylation site	No	No	↓	↓
	C681S (Barfoot et al., 2015)	D	D	D	D	D	D	N	Gain of N-linked glycosylation at N677	0.02	0.04	• N-glycosylation site • Proline-directed phosphorylation • MAPK phosphorylation site	No	No	↑	↓
	P682E (Roset et al., 2014) (Hohl et al., 2015)	D	D	N	D	D	D	D	Gain of helix Altered coiled coil Loss of N-linked glycosylation at N677	0.32 0.14 0.02	3.1e-03 0.03 0.04	• CK2 phosphorylation site	No	No	↓	↓
	P682A (He et al., 2012)	N	N	N	N	D	D	N	No effect	-	-	None	No	No	↓	↓
	V683I (He et al., 2012)	N	N	N	N	N	D	N	No effect	-	-	None	No	No	↑	↑
	V683R (Roset et al., 2014)(Hohl et al., 2015)	N	N	N	N	N	D	D	Gain of helix	0.29	0.01	None	No	No	↓	↓
	C684R (He et al., 2012)	D	D	D	D	D	D	D	Gain of helix	0.30	9.5e-03	• Diarginine retention/retrieving signal	No	No	↓	↓
	C684S (Barfoot et al., 2015)	D	D	D	D	D	D	D	No effect	-	-	• PIKK phosphorylation site • PKC phosphorylation site	No	No	↓	↓
	Q685S (He et al., 2012)	N	N	N	N	N	D	N	Altered coiled coil	0.53	6.5e-03	• BRCT phosphopeptide ligands • USP7 binding motif	No	No	↓	↓
Signature motif	R1198E (Rojowska et al., 2014b)	D	D	D	D	D	D	D	Gain of catalytic site at R1200 Gain of allosteric site at R1200 Altered metal binding Altered transmembrane protein	0.25 0.21 0.14 0.10	4.8e-03 0.03 0.03 0.04	• Diarginine retention/retrieving signal	No	No	↓	↓
	G1199E (Rojowska et al., 2014b)	D	D	D	D	D	D	D	Loss of allosteric site at R1200 Loss of catalytic site at R1200 Altered transmembrane protein	0.23 0.20 0.11	0.03 0.01 0.04	• Diarginine retention/retrieving signal • PKA phosphorylation site	No	No	↓	↓
	S1202A (Herdendorf and Nelson, 2011)	D	D	D	D	D	D	D	Loss of allosteric site at R1200 Loss of catalytic site at R1200	0.23 0.20	0.03 0.01	• PKA phosphorylation site • -Glycosaminoglycan attachment site	No	No	↓	↓
	S1202R (Koroleva et al., 2007) (Moncalian, Lengsfeld, Bhaskara, Hopfner, Karcher, Alden, J. A. Tainer, et al., 2004) (Herdendorf and Nelson, 2011)	D	D	D	D	D	D	D	Gain of ADP-ribosylation at S1202 Loss of allosteric site at R1200 Loss of catalytic site at R1200	0.25 0.23 0.21	8.4e-03 0.03 1.0e-02	• PKA phosphorylation site • Glycosaminoglycan attachment site	No	No	↑	↓

	(Bhaskara *et al.*, 2007)															
	S1202M (Herdendorf and Nelson, 2011)	D	D	D	D	D	D	D	Loss of allosteric site at R1200 Gain of catalytic site at R1200	0.23 0.21	0.02 9.7e-03	• PKA phosphorylation site • Glycosaminoglycan attachment site	No	No	↑	↑
	Q1205E (Herdendorf and Nelson, 2011)	D	D	D	N	D	D	D	Gain of allosteric site at R1200 Loss of catalytic site at R1200	0.23 0.20	0.02 0.01	• PKA phosphorylation site • CK2 phosphorylation site	↑	↓	↓	↓
	K1206M (Herdendorf and Nelson, 2011)	D	D	D	D	D	D	D	Gain of catalytic site at S1202	0.09	0.04	None	↑	↓	↑	↑
	K1206A (Williams *et al.*, 2011)	D	D	D	D	D	D	D	Loss of catalytic site at K1206	0.09	0.05	None	↑	↓	↓	↑
	K1206E (Williams *et al.*, 2011)	D	D	D	D	D	D	D	Gain of catalytic site at K1206	0.11	0.03	• TRAF2 binding site • NES nuclear export signal	↑	↓	↓	↑
	L1211W (Deshpande *et al.*, 2014)	D	D	D	D	D	D	D	Loss of catalytic site at K1206	0.08	0.05	• SUMO interaction site	↑	↑	↓	↓
	R1214A (Williams *et al.*, 2011)	D	D	D	D	D	D	D	Loss of allosteric site at R1214	0.22	0.03	• ATP-binding cassette, ABC transporter-type, signature and profile	↑	↓	↓	↓
	R1214E (Deshpande *et al.*, 2014) (Williams *et al.*, 2011)	D	D	D	D	D	D	D	Loss of allosteric site at R1214	0.20	0.04	• SUMO interaction site • ATP-binding cassette, ABC transporter-type, signature and profile	↑	No	↓	↓
Walker B	E1232Q (Rojowska *et al.*, 2014b)	D	D	D	D	D	D	D	Altered metal binding Loss of catalytic site at E1232 Loss of allosteric site at P1233 Altered transmembrane protein	0.48 0.34 0.24 0.12	4.3e-03 9.8e-04 0.02 0.03	• FHA phosphopeptide ligands • SUMO interaction site	No	No	↓	↓
D-loop	D1238N (De La Rosa Metzere Bierlein and and Scott W. Nelson, 2011)	D	D	D	D	D	D	D	Altered ordered interface Altered metal binding Gain of relative solvent accessibility Gain of allosteric site at P1233 Loss of catalytic site at T1234 Altered transmembrane protein Altered coiled coil	0.3 0.31 0.27 0.25 0.17 0.12 0.08	4.3e-03 2.8e-03 0.02 0.01 0.02 0.03 0.05	▪ FHA phosphopeptide ligands • Casein kinase II phosphorylation site	No	No	↓	↓
	D1238A (De La Rosa Metzere Bierlein and and Scott W. Nelson, 2011)	D	D	D	D	D	D	D	Altered metal binding Altered ordered interface Loss of allosteric site at P1233 Loss of catalytic site at T1234 Altered transmembrane protein	0.41 0.40 0.26 0.18 0.12	3.4e-04 1.4e-03 0.01 0.02 0.02	▪ FHA phosphopeptide ligands • Casein kinase II phosphorylation site	No	No	↓	↓
ATPase domain/coiled-coil	K6E (Alani, Padmore and Kleckner, 1990) (Bender *et al.*, 2002)	D	D	D	D	D	N	D	Loss of strand Altered DNA binding Gain of N-terminal acetylation at M1	0.27 0.16 0.03	0.03 0.04 4.1e-03	None	No	No	↓	↓
	K22M (Alani, Padmore and Kleckner, 1990) (Bender	N	N	N	N	N	D	N	No effect	-	-	None	↑	No	↓	↑

et al., 2002)															
R83I (Alani, Padmore and Kleckner, 1990) (Bender *et al.*, 2002)	N	D	N	N	N	D	N	Altered ordered interface Altered DNA binding Altered coiled coil	0.29 0.22 0.10	0.03 0.02 0.04	• PP1-docking motif RVXF	↑	↑	↓	↑
K132E (Rojowska *et al.*, 2014b)	D	D	N	D	D	D	D	Loss of helix Altered transmembrane protein Gain of strand	0.28 0.27 0.27	0.02 7.3e-04 0.01	• CK1 phosphorylation site • Protein kinase C phosphorylation site	No	No	↓	↓
T191E (Rojowska *et al.*, 2014b)	N	D	N	N	N	N	N	Altered coiled coil Loss of acetylation at K187	0.28 0.28	0.01 6.2e-03	• TRAF2 binding site • NEK2 phosphorylation site • PKC phosphorylation site	No	No	↓	↓
C221E (Rojowska *et al.*, 2014b)	N	N	N	N	N	N	N	No effect	-	-	None	No	No	↓	↓
K105E (Rojowska *et al.*, 2014b)	D	N	D	D	D	D	D	No effect	-	-	None	No	No	↓	↓
S106E (Rojowska *et al.*, 2014b)	N	N	N	N	N	N	N	No effect	-	-	None	No	No	↑	↓