Mechanistic modelling and Bayesian inference elucidates the variable dynamics of double-strand break repair

M. Woods[1] and C.P. Barnes[1,*]

[1] Department of Cell and Developmental Biology, University College London, England

*To whom correspondence should be addressed. Email: christopher.barnes@ucl.ac.uk

Abstract

DNA double-strand breaks are lesions that form during metabolism, DNA replication and exposure to mutagens. When a double-strand break occurs one of a number of repair mechanisms is recruited, all of which have differing propensities for mutational events. Despite DNA repair being of crucial importance, the relative contribution of these mechanisms and their regulatory interactions remain to be fully elucidated. Understanding these mutational processes will have a profound impact on our knowledge of genomic instability, with implications across health, disease and evolution. Here we present a new method to model the combined activation of non-homologous end joining, single strand annealing and alternative end joining, following exposure to ionizing radiation. We use Bayesian statistics to integrate eight biological data sets of double-strand break repair curves under varying genetic knockouts and confirm that our model is predictive by re-simulating and comparing to additional data. Analysis of the model suggests that there are at least three disjoint modes of repair, which we assign as fast, slow and intermediate. Our results show that when multiple data sets are combined, the rate for intermediate repair is variable amongst genetic knockouts. Further analysis suggests that the ratio between slow and intermediate repair depends on the presence or absence of DNA-PKcs and Ku70, which implies that non-homologous end joining and alternative end joining are not independent. Finally, we consider the proportion of double-strand breaks within each mechanism as a time series and predict activity as a function of repair rate. We outline how our insights can be directly tested using imaging and sequencing techniques and conclude that there is evidence of variable dynamics in alternative repair pathways. Our approach is an important step towards providing a unifying theoretical framework for the dynamics of DNA repair processes.

Introduction

Double-strand breaks (DSBs) are lesions in DNA that occur naturally by oxidative stress, DNA replication and exogenous sources [1, 2]. When left unprocessed or during erroneous repair, they cause changes to DNA structure creating mutations and potential genomic instability [3–8]. To repair DSBs, multiple mechanisms have evolved and are known to include non homologous end joining (NHEJ) [7, 9–17], homologous recombination [18] including single strand annealing (SSA) [19, 20], microhomology mediated end joining (MMEJ) [21, 22] and alternative or back-up end joining (A-EJ) [23, 24]. The choice of mechanism depends on the structure of the break point, where simple

1

breaks caused by restriction enzymes are different in structure from those caused by ionising radiation (IR) (reviewed in [25, 26]). This affects the probability of error prone repair because mutations are mechanism specific and depending on which mechanism is activated a cell might exhibit chromosome translocations [4, 5], small deletions or insertions [6, 7] or recombination leading to loss of heterozygosity [8]. For example, in mouse, error by SSA causes chromosome translocations [4] and in *Saccharomyces cerevisiae*, NHEJ of simple DSBs is associated with small deletions or insertions [7]. *In vivo* studies of DSBs have suggested that in addition to structural activation arising from the break point, cell-cycle dynamics can also play a role in repair mechanism activation (reviewed in [27]). In particular the choice of mechanism is not fixed at the time of damage and cells exhibit a pulse like repair in human U-2 OS cells [28], a behaviour supported by a molecular basis for cell cycle dependence in NHEJ, mediated by Xlf1 phosphorylation [29].

To understand how mutations are distributed in the genome, it is important to uncover the dynamic activation and interplay between different DSB repair mechanisms. This mutual activation is not fully understood, however the individual repair mechanisms and recruitment proteins of NHEJ, SSA and A-EJ have been documented. NHEJ requires little or no homology, is a mechanism of DNA end joining in both unicellular and multicellular organisms [7] and can exhibit fast repair by the binding of DNA dependent protein kinase (DNA-PK) [15]. In vertebrates, NHEJ initiates the recruitment and binding of several proteins. These have been shown to include Ku70, Ku80, DNA-PK catalytic subunit (DNA-PKcs), Artemis and Ligase IV in a cell free system [9]. Ku70 and Ku80 are subunits of the protein DNA-PK. Biochemical and genetic data suggest they bind to DNA ends and stimulate the assembly of NHEJ proteins by DNA-PKcs [10, 12]. Repair proceeds by Artemis facilitated overhang processing and end ligation via DNA Ligase IV [13, 14]. Although well studied, new regulating components of NHEJ are still being discovered, for example the protein PAXX [17]. SSA is slower than NHEJ and in yeast works at almost 100% efficiency for homologous regions of at least 400bp [6]. First described in mouse fibroblast cells [19, 20], during SSA two complementary sequences are exposed through a 5′ to 3′ exonuclease end resection and aligned. Remaining overhangs are then cut by an endonuclease and the DNA is reconstructed by DNA polymerase using the homologous sequences as a template. Some of the components that contribute to SSA have been identified in eukaryotes e.g. the complex MRN consisting of Mre11, Rad50 and Nibrin which facilitates DNA end resection [30]. Following resection, replication protein A (RPA) binds to the DNA and when phosphorylated, forms a complex with Rad52 to stimulate DNA annealing [31, 32]. Similarly to NHEJ, following gap repair, SSA is terminated with end ligation by Ligase III [33]. Data of repair kinetics for mutants defective in Rad52 show limited slow repair in comparison to wild type repair curves in gamma irradiated cells in chicken B line cells [34], suggesting that SSA may be active in the repair of DSBs caused by IR. In yeast, it has been suggested that SSA constitutes a major role in the repair of DSBs accounting for three to four times more repairs than gene conversion during M phase [35].

One interesting finding in genetic studies is that when NHEJ is compromised, DSBs are removed by alternative mechanisms, that have adopted various names in the literature, such as MMEJ in yeast [36] and back-up NHEJ (B-NHEJ) in higher eukaryotes [37] that here, we collectively refer to as A-EJ [23, 38], (reviewed in [37, 39, 40]). It is still unclear how A-EJ is regulated or interacts with other processes but there is evidence it is active in the repair of breaks with microhomology of 3-16 nu-

cleotides, reviewed by Decottignies [40]. Thought to act on break points with ends that are not complementary in the absence of NHEJ factors [38], an assortment of PARP-1, 53BP1, Lig3 and 1, Mre11, CtIP and Polθ have been proposed as regulators of A-EJ. PARP-1 is required and competes with Ku for binding to DNA ends through the PARP-1 DNA binding domain [24]. Other proteins are involved in initial binding, where activation of 53BP1 in MMEJ is dependent on Ku70 and independent of DNA-PKcs [22] and CtIP has been associated through the use of microhomology [41]. The proteins required for end joining have been identified as Lig3 and Lig1 in the absence of XRCC1 [42–44]. This pathway has never been observed in single cells and it is unclear how A-EJ is related to other mechanisms. However, targeted RNAi screening for A-EJ has uncovered shared DNA damage response factors with homologous recombination [45]. For an illustration of the three mechanisms (see Fig. 1a).

Mathematical models of DSB repair have used biphasic [46], biochemical kinetic [47–52], multiscale [53, 54], and stochastic methods [55]. In a study by Cucinotta *et al.* [48], a set of coupled nonlinear ordinary differential equations were developed. The model was based on the law of mass action and stepwise irreversible binding of repair proteins to describe NHEJ rejoining kinetics and the phosphorylation of H2AX by DNA-PKcs. Similar studies have modelled repair kinetics and protein recruitment during SSA [49, 51], NHEJ [50] and other mechanisms including NHEJ, HR, SSA and two alternative pathways under a wide range of linear energy transfer (LET) values and heavy ions [52]. These studies have been met with some controversy, for example with the argument that the biphasic model has never succeeded in providing definitive values for the repair components [56]. Recently however, the models have been further developed to model the complexity of a DSB by application to damage induced by ionising radiation of different qualities [57,58]. This can be achieved because the spectrum of DSB-DNA damage can be computed by applying Monte Carlo track structure [59], which is a method that can be used to simulate the passage of charged particles in water, for a review see [60].

Previous biochemical kinetic models have been used to reproduce the experimental data observed. This approach often uses more parameters than are required to describe sequential steps in the repair process. This can cause difficulty in identifying parameter values because multiple parameter value combinations may be able to describe the data well, an issue known as identifiability. Consequently, predictions are not unique, which can be detrimental in the design of a biological experiment. Therefore the creation of models that provide a unique interpretation of repair dynamics is a challenge.

Here we develop a statistical model that can take DSB repair curve data, such as those generated from pulse field gel electrophoresis (PFGE) or comet assays, and infer repair mechanism activation. The method relies on training a simple model against multiple data sets of DSB repair under different genetic knockouts when multiple repair mechanisms are activated. Using the most probable set of parameter values, we can then simulate the model and make predictions on the activation of different rates of repair. Unlike previous modelling approaches, we do not model individual recruitment proteins. Instead we assign parameter values to different rates of repair. This has two benefits. Firstly it provides a method to uncover different rates of repair arising from different repair mechanisms that are implicit in the data. Secondly, it reduces the number of parameters required to describe the system, leading to a more identifiable model. Our approach strikes a balance be-

3

tween a detailed mechanistic description of the biochemical components with a traditional statistical model. This enables insights into the dynamical process underling repair pathways combined with novel and testable predictions. We use this method to integrate the data from eight repair curve assays under genetic knockouts including combinations of Ku70, DNA-PKcs, Rad52 and Rad54. We first infer that there are at least three disjoint dynamical repair mechanisms that explain the combined data and that the dynamics depend on the regulating recruitment proteins. We propose that there are a number of alternative end joining dynamical processes that may or may not share a common genetic pathway. We also demonstrate that our model has predictive power on new data sets, including PARP-1 knockouts, and show that the activation of different repair processes over time depends on the speed of the underlying dynamics.

Materials and Methods

Experimental data

The experimental data used in this study are published repair curves generated from methods of pulse field gel electrophoresis, a technique that distributes the DNA according to the length of the fragment. We model the dose equivalent number of DSBs that are obtained from the fraction of DNA released into the gel [61]. Table 1 lists the experimental data that are used for inference. Cells were exposed to X-rays [24] and the number of DSBs within the population recorded over time. The eight data sets are labelled D1-D8. Data D1 is wild type and since the cell cycle phase is unrestricted, we expect all three repair processes to be present. Data D2 and D3 are DNA-PKcs knockouts in G1 and G2 phase, where we expect NHEJ to be compromised but since Ku is present we still expect the recruitment process. Data D4 is a Rad52 knockout where we expect only NHEJ and A-EJ to be present. Data D5 and D6 are Ku knockouts, where we assume the whole of the NHEJ pathway to be compromised and only SSA and A-EJ remain active. Data D7 and D8 are expected to have no repair by PARP-1 mediated A-EJ because both sets were treated with PARP-1 inhibitors. Data D7 comes from $Ku70^{-/-}$ mouse fibroblasts, where we expect to see no repair by NHEJ as well as a lack of A-EJ due to PARP-1 inhibition [24].

Modelling double-strand break repair processes

We assume DSBs caused by ionising radiation (IR) can be repaired by multiple processes with different rates corresponding to NHEJ, SSA and A-EJ (see Fig. 1b). In principle, other mechanisms could be included, such as Rad54 dependent homologous recombination (HR), but since this mechanism is is mostly active in S phase of the cell cycle, and thought to contribute little to IR induced DSBs [62], we chose not to model it explicitly. We model a DNA repair process by a stochastic reaction system, represented by

$$x + E_i \xrightarrow{K_i} y_i \tag{1}$$

$$y_i \xrightarrow{K_i'} \emptyset + E_i, \tag{2}$$
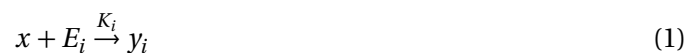
where x is the DSB, E_i is the recruitment protein for process i (Ku, MRN and PARP-1), and K_i, K_i' are the parameters for recruitment and subsequent ligation respectively. To model the limited re-

4

Data	Dose (Gy)	Phase	Mutant, Cell line	Repair
D1	20	Asynchronous	WT MEFs* [62]	NHEJ, SSA, A-EJ
D2	20	G1	DNA-PKcs$^{-/-}$ MEFs* [62]	SSA, A-EJ
D3	20	G2	DNA-PKcs$^{-/-}$ MEFs* [62]	SSA, A-EJ
D4	80	Asynchronous	Rad52$^{-/-}$ DT40* [63]	NHEJ, A-EJ
D5	80	Asynchronous	Ku70$^{-/-}$/Rad54$^{-/-}$ DT40* [34]	SSA, A-EJ
D6	54	Asynchronous	Ku70$^{-/-}$ DT40* [34]	SSA, A-EJ
D7	52	Asynchronous	Ku70$^{-/-}$ + DPQ MEFs [24]	SSA
D8	32	Asynchronous	WT + 3'-AB MEFs [24]	NHEJ, SSA

Table 1: Table of data sets used for model fitting. The data contains DSB repair kinetics for cells that are irradiated at different doses or split into different phases of the cell cycle, G1 and G2. Data was traced from current literature, or where indicated was provided by G. Iliakis (*). References to the data and cell lines are provided. We chose a combination of mouse embryonic fibroblasts (MEFs) and DT40 cells because DT40 cells remove DSBs from their genome similarly to mammalian cells [63].

sources available to the cell, we follow the approach of Cucinotta *et al* [48], and assume that the total amount of protein is conserved for each repair mechanism

$$E_i + y_i = C_i, \tag{3}$$

where C_i is the total amount of recruitment protein for each repair mechanism. This structure also captures the experimentally observed fact that mRNA levels for repair processes depend on the radiation dose [64]. The reactions result in a nonlinear coupled stochastic system which is simulated using the Gillespie algorithm [65].

The proportions of DSBs repaired by each mechanism are estimated by calculating the cumulative number of DSBs that enter each individual pathway with the integral

$$\tilde{N}_i = \int_0^\infty K_i x E_i dt. \tag{4}$$

This integral can then be used to calculate the proportion of DSBs repaired by each mechanism.

Hierarchical modelling of repair across knockout strains

To build a model that can be used to obtain unique predictions, it is advantageous to minimise the number of parameters that describe the system. To do this, we developed a hierarchical model where the individual parameter values, K, K', are lognormally distributed with a common mean, μ, across all the data sets in which they are included (see Fig. 2a). For data sets in which a repair protein is repressed downstream of the initial protein that binds, we impose an additional hyperparameter. We include this additional hyperparameter because it is not clear if a repair mechanism remains active when individual regulating proteins are repressed.

More formally, we wish to obtain $K_i^d, K_i'^d$ for $i \in \{1,2,3\}$ and eight data sets $d \in \{1,...,8\}$. These parameters are constrained by the repair dependent hyperparameters, $\gamma = \{\mu_i, \sigma^2\}$, where μ_i represents the mean of a lognormal distribution, and σ^2 the variance. The $K_i^d, K_i'^d$ are drawn from the population level distributions, where the joint density can be written

$$\pi(D, K, \gamma) = \pi(D|K)\pi(K|\gamma)\pi(\gamma), \tag{5}$$

and Bayes rule becomes

$$\pi(\gamma|D) = \frac{\pi(\gamma)\int f(D|K)\pi(K|\gamma)dK}{\pi(D)}, \tag{6}$$

which gives the posterior of the hyperparameters given the data, D. The integral indicates that we sum over (marginalise) the K values.

Parameter estimation using approximate Bayesian computation

To perform the inference we use approximate Bayesian computation sequential Monte Carlo (ABC SMC) [66–69]. This method that can be used to fit a model to multiple data sets when the likelihood is unavailable. In the Bayesian framework, we are interested in the posterior distribution $\pi_\epsilon(\theta, x|y)$, where θ is a vector of parameters and $x|y$ is the simulated data conditioned on the experimental data. To obtain samples from the posterior distribution we must condition on the data y and this is done via an indicator function $I_{\mathscr{A}_{y,\epsilon}}(x)$. We then have

$$\pi_\epsilon(\theta, x|y) = \frac{\pi(\theta)f(x|\theta)I_{\mathscr{A}_{y,\epsilon}}(x)}{\int_{\mathscr{A}_{y,\epsilon}\times\Theta}\pi(\theta)f(x|\theta)dx\,d\theta},$$

where $\mathscr{A}_{y,\epsilon} = \{x \in \mathscr{D} : \rho(x, y) \le \epsilon\}$, $\rho : \mathscr{D} \times \mathscr{D} \to R^+$ is a distance function comparing the simulated data to the observed data and π_ϵ is an approximation to the true posterior distribution. This approximation is obtained via a sequential importance sampling algorithm that repetitively samples from the parameter space until ϵ is small, such that the resulting approximate posterior, π_ϵ, is close to the true posterior.

The ABC SMC algorithm is used to calculate an approximation to the target posterior density $\pi_\epsilon(\gamma|D)$, where $D = \{D_i; i = 1, 2, ..., 8\}$. We can include the hierarchical model by simulating data, x^*, using the following scheme:

$$\begin{aligned}
\gamma &\sim U(\alpha, \beta) \\
K &\sim LN(\mu, \sigma^2) \\
D^* &\sim f(D|K), \tag{7}
\end{aligned}$$

where $f(D|K)$ is the data generating model (the solution to the reaction system represented by Equations 1 - 3). The α, β are the lower and upper limits of the uniform prior on the hyperparameters. Note that in the sequential importance sampling step, we perturb only the hyperparameters. For further details on the inference and all prior values see S1 Text.

Results

DSB repair requires fast, slow and alternative mechanisms

We fit three different models, comprising one, two and three repair processes respectively (M_1 : $\{i = 1\}$, M_2 : $\{i = 1, 2\}$ and M_3 : $\{i = 1, 2, 3\}$), and found that a three process model describes the best fit using an approximation to the Deviance Information Criterion (DIC) [70] and the Akaike information criterion (AIC) [71], based on a surrogate likelihood approach [72] (Fig. B, C in S1 Text). The final model structure is presented in Table 2, (for prior distributions see S1 Text), and a summary of the fitted parameter values is given in Table 3.

Data	Model Parameters			Hyperparameters		
	Fast	**Slow**	**A-EJ**	**Fast**	**Slow**	**A-EJ**
D1	K_{1D1}, K'_{1D1}	K_{2D1}, K'_{2D1}	K_{3D1}, K'_{3D1}	μ_1, σ^2	μ_2, σ^2	μ_3, σ^2
D2	K_{1D2}	K_{2D2}, K'_{2D2}	K_{3D2}, K'_{3D2}	μ_4, σ^2	μ_2, σ^2	μ_3, σ^2
D3	K_{1D3}	K_{2D3}, K'_{2D3}	K_{3D3}, K'_{3D3}	μ_4, σ^2	μ_2, σ^2	μ_3, σ^2
D4	K_{1D4}, K'_{1D4}	K_{2D4}	K_{3D4}, K'_{3D4}	μ_1, σ^2	μ_4, σ^2	μ_3, σ^2
D5	-	K_{2D5}, K'_{2D5}	K_{3D5}, K'_{3D5}	-	μ_2, σ^2	μ_3, σ^2
D6	-	K_{2D6}, K'_{2D6}	K_{3D6}, K'_{3D6}	-	μ_2, σ^2	μ_3, σ^2
D7	-	K_{2D7}, K'_{2D7}	-	-	μ_2, σ^2	-
D8	K_{1D8}, K'_{1D8}	K_{2D8}, K'_{2D8}	-	μ_1, σ^2	μ_2, σ^2	-

Table 2: Model parameters with their corresponding hyperparameters used in the hierarchical model. Their values are fitted using ABC SMC. For prior distributions on the hyperparameters, see Text S1.

The fit of the simulation to the data for all eight data sets is shown in Fig. 2g. The fits capture the essential aspects of the repair curves and most points are consistent with the posterior median and credible regions. When we qualitatively compare this fit to that obtained by the one and two process models, M_1 and M_2, we find a poor fit for M_2 in data sets D2-4 and generally large credible regions for M_1 (see Fig. D, E, F in S1 Text). The posterior distributions of the hyperparameters are shown in Fig. 2b. Inspection of the interquartile range of the hyperparameters confirms that a combination of fast, slow and intermediate repair is sufficient to describe the wild type and mutant data, furthermore a two sided Kolmogorov Smirnov test between the posterior distributions for the hyperparameters confirmed that the four distributions were significantly different to one another

Data	Model Parameters			Posterior median ($q_{0.25}$, $q_{0.75}$) (hrs^{-1})		
	Fast	Slow	A-EJ	Fast	Slow	A-EJ
D1	K_{1D1}	K_{2D1}	K_{3D1}	3.06 (1.9,5.36)	0.04 (0.022,0.078)	0.34 (0.19,0.62)
D2	K_{1D2}	K_{2D2}	K_{3D2}	0.05 (0.023,0.094) †	0.037 (0.019,0.067)	0.4 (0.26,0.63)
D3	K_{1D3}	K_{2D3}	K_{3D3}	0.04 (0.022,0.081) †	0.038 (0.021,0.08)	0.5 (0.31,0.84)
D4	K_{1D4}	K_{2D4}	K_{3D4}	2.24 (1.52,3.56)	0.056 (0.027,0.11) †	0.45 (0.26,0.75)
D5	-	K_{2D5}	K_{3D5}	-	0.066 (0.034,0.13)	0.22 (0.13,0.42)
D6	-	K_{2D6}	K_{3D6}	-	0.057 (0.03,0.102)	0.29 (0.16,0.53)
D7	-	K_{2D7}	-	-	0.042 (0.022,0.077)	-
D8	K_{1D8}	K_{2D8}	-	2.34 (1.75,3.16)	0.048 (0.03,0.092)	-
D1	K'_{1D1}	K'_{2D1}	K'_{3D1}	2.87 (2.03,4.49)	0.04 (0.024,0.08)	0.4 (0.22,0.65)
D2	-	K'_{2D2}	K'_{3D2}	- †	0.041 (0.021,0.08)	0.62 (0.4,0.97)
D3	-	K'_{2D3}	K'_{3D3}	- †	0.041 (0.023,0.08)	0.91 (0.58,1.33)
D4	K'_{1D4}	-	K'_{3D4}	4.35 (3.07,6.57)	- †	0.29 (0.19,0.46)
D5	-	K'_{2D5}	K'_{3D5}	-	0.025 (0.016,0.046)	0.16 (0.12,0.31)
D6	-	K'_{2D6}	K'_{3D6}	-	0.034 (0.02,0.058)	0.19 (0.15,0.28)
D7	-	K'_{2D7}	-	-	0.022 (0.015,0.032)	-
D8	K'_{1D8}	K'_{2D8}	-	3.2 (2.6,4.4)	0.041 (0.022,0.085)	-

Table 3: Median posterior parameter values with 0.5 credible regions. Values can be compared to initial binding rate constants from two models in the literature; $k = 0.5$ hrs^{-1} for NHEJ [50] and $k = 0.057$ hrs^{-1} for SSA [51]. Half times of DSBs can be calculated using $t_{1/2} = \log_e / K$. † Cases in which a protein proceeding the first complex to bind is inhibited. In this model we assume there is initial binding but no repair.

$(\mu_1, \mu_2\ D = 1, \mu_1, \mu_3\ D = 1, \mu_2, \mu_3\ D = 0.998, \mu_2, \mu_4\ D = 0.686, \mu_4, \mu_3\ D = 0.752, \mu_4, \mu_1\ D = 1$, all tests $p < 2.2e^{-16}$). For each data set (D1-D8) the posterior interquartile ranges of the parameters K_1, K_2 and K_3 were recorded (Fig. G in S1 Text). Marginal distributions for the wild type K_1, K_2, K_3 are shown in Fig. 2 c-e). Analysis of the marginal distribution shows that the parameter distributions of K_1, K_2 and K_3 deviate from the hyperparameter distributions, suggesting that although the rates are defined as fast, slow and intermediate, there is variation in activation of the mechanisms among different mutants (Fig. 2c-e). There is some overlap in parameter values K_1, K_2 and K_3 (Fig. 2f) but the interquartile ranges of the parameters K_1, K_2 and K_3 are disjoint, this is also observed in all eight data sets (Fig. G in S1 Text). For all posterior distributions of the parameters and a plot of individual DSBs and their repair in a wild type model, see Fig. H-K in S1 Text. When individual DSBs are tracked in the model, the DSBs are quickly distributed amongst the three pathways and repaired according to the predicted rate. To check that our model parameters were robust to adding additional data, we performed ABC SMC again on nine data sets. This new data set consisted of the eight data sets listed in Table 1 and an additional repair time series of xrs-6 cells deficient in Ku80 inhibited of PARP-1 by DPQ [24], in which cells were deficient in NHEJ and A-EJ. The results of the total number of DSBs repaired and our predictions on the activation of A-EJ using this data set were the same and a comparison between the eight data and nine data posterior showed a similar fit (Figs. L and M in S1 Text). In summary, we conclude that a three process model provides the best fit to the data observed (Fig. B-C in S1 Text) and that for equal prior ranges on the processes the biological data can be explained by one fast, one slow and at least one intermediate rate of repair.

The number of DSBs repaired by each mechanism depends on regulating recruitment proteins

By re-simulating from our fitted posterior distribution we were able to examine the dynamics of DSB repair across mechanisms and data sets (see Fig. 3). Data sets in which NHEJ is active exhibited a faster repair with the cumulative number of DSBs reaching to within 80% of the total within a period of 2 hours post irradiation. Next, we plotted the number of DSBs entering each repair mechanism as a time series (Fig. 3b,d). The simulated data predicts that fast repair consistently processes most of the DSBs within two hours after radiation (red curves in Fig. 3). Similarly, there were no clear differences amongst the data in the DSB processing by slow repair. Intriguingly, intermediate repair was slower in cells compromised of Ku70 (D5,D6) than those without DNA-PKcs (D2,D3) (green curves, Fig. 3b,d) To calculate the predicted number of DSBs repaired by fast, slow and alternative mechanisms, we computed the integral $\tilde{N}_i$. The results are shown in Fig. 4. Data sets for which cells were deficient in regulating components of NHEJ confirmed variation in the numbers of DSBs repaired by intermediate rates. In agreement with the results obtained from the time series plots (Fig. 3b,d) there was a difference in the ratio of slow and intermediate rates between data sets D2,D3 and D5,D6. We also observed an increase in the number of DSBs repaired by A-EJ between G1 and G2 (Figure 4 D2,D3) , agreeing with experimental results in the literature [62].

Fast and slow kinetics exhibit constrained rate parameters

To test the predictive capability of the model, we re-simulated repair curves using the posterior parameters for data sets D1-D8 and compared the simulated curves to new data sets from the literature

9

(Fig. 5). The model trajectories provide a good fit to wild type data at 40Gy (Fig. 5a, [62]). More impressively, the simulations of cells defective in PARP-1 by application of 3'-AB and Ku80 deficient xrs-6 cells exposed to DPQ defective in PARP-1 and NHEJ show good agreement with the experimental curves (Fig. 5b,c, [24]). This analysis suggests that small ranges of the rate parameters for the fast and slow repair processes, when assigned to NHEJ and SSA, can predict multiple low LET data sets under different experimental conditions. This suggests that they are constrained across experimental and biological conditions. Next we investigated whether the intermediate rates from our model could potentially fit the experimental data when activation of fast and slow rates are inhibited. We re-simulated the model again but this time set the parameters for active NHEJ and SSA to zero across the models for D1-D8. The results are shown in Fig. 5d along with the corresponding experimental data. In wild type data, we see that it is possible for all the DSBs to be removed with the predicted rates of A-EJ, although repair is slower, suggesting that the ability for A-EJ to repair DSBs is not saturated. When Rad52 is inhibited (data D4), because PARP-1 could still compete with MRN we see that it is not possible with the predicted rates for all the DSBs to be repaired in the absence of NHEJ. There was no clear difference in the repair of DNA-PKcs mutants, however when Ku70 is inhibited (data D5, D6) we see that when competition by slow repair is removed the predicted rates of A-EJ, suggest that the activation of A-EJ is not saturated and is perhaps inhibited by competition with MRN.

Alternative end joining demonstrates variable dynamics

The time taken for over half the DSBs to be repaired by A-EJ is shown in Fig. 6a. Some repair is fast, occurring within two hours, however, for cells deficient in Ku70, A-EJ adopts a slower repair with half maximum achieved at eight hours. The activation of A-EJ across the data sets is represented by the interquartile ranges of the posterior distributions for K_3 and K_3' (Fig. 6b). The rate for ligation, K_3', is low in data sets $D5$ and $D6$, suggesting that intermediate mechanisms are less active in the absence of Ku70. The rate is highest in G2 when DNA-PKcs is inhibited. These data suggest that A-EJ adopts a slow or fast repair and that the speed of repair depends on the presence or absence of DNA-PKcs and Ku70, because inhibition of Rad52 had little effect on the time until half-maximum. There are two ways in which this difference between Ku70 and DNA-PKcs mutants can be interpreted. The first is that when Ku70 is inhibited, two alternative mechanisms are activated, one that is fast and one that is slow. The other interpretation is that A-EJ is one repair mechanism that repairs at a slower rate when Ku70 is inhibited. We also modelled the complete inhibition of A-EJ by active NHEJ (A-EJ removed from data sets D1 and D4) and found that the model still captures all the observed dynamics, albeit with the rate of fast repair slightly increased in data set D4 where there is inhibition of Rad52 (see Fig. N in S1 Text).

Total activation time of repair mechanisms depends on the rate of repair

Inspection of the time series data (Fig. 3b,d) suggests that at time $t = 0.5$hrs, the majority of DSBs that are being processed are within a fast mode of repair. Fig. 6c illustrates a typical distribution of DSBs over each mechanism at different points in time. At time $t = 0$, the cells are exposed to a single dose of ionising radiation. Quickly, for example at time $t < 1$hrs, fast and possibly faster alternative repair mechanisms dominate the DSB processing. Later, after all DSBs processed by

10

the faster mechanisms have been repaired, the remaining DSBs fall within the category of breaks that require processing by slower mechanisms. This change in the activity of repair mechanisms could potentially be investigated by recording changes in the level of recruitment proteins or gene expression as time series. To quantify this change in our simulated data, we plotted the percentage of DSBs that remain in active repair mechanisms over time for the wild type data (see Fig. 6d). By inspection, we can see that at 0.5 hours after irradiation most DSBs reside in the fast mechanisms. At a time of $t = 8$hrs, the percentage approches zero for fast repair and the majority of DSBs are found within the slow repair pathways. The variation in Fig. 6d) is shown with the 25th and 75th percentiles and this is due to the variation in repair rate $K_i, K_i^{'}$ for each mechanism. Ultimately, it is the values of the parameters $K_i, K_i^{'}$ that determine the rate of repair, so to confirm if the dynamics presented in Fig. 6c are representative of the whole data set, we considered all time series for parameters K_i, a total of 9000 simulations. For each parameter at every time point, we assigned a value of 1 if the corresponding mechanism for the parameter contained over 30% of the total DSBs being processed at that time point and a value of 0 if it contained less than 30%. The results are shown in Fig. 6e, where for each parameter K_i, a red line indicates the times at which the mechanism with rate K_i is greater than 30% active. There is a clear trend showing that the percentage of total activation decreases in time with an increase in repair rate K. In other words, the model predicts the times at which different repair processes are likely to express regulating components. When repair is extremely slow the repair mechanism never reaches 30 % of the current DSBs. In summary, these results predict that if a cell experiences a sudden creation of DSBs, then gene expression for slower repair mechanisms will be maintained for longer than those required for faster repair mechanisms such as NHEJ, a result that has been implied for NHEJ and HR (Fig. 3 in [28]).

Discussion

In this study, we presented a new hierarchical model of DSB repair and applied Bayesian inference to infer the number of active repair processes and their dynamical behaviour from experimental PFGE data. Because the model assumptions are simple and exclude the full mechanistic details of the biological processes, we are able to form an identifiable model and provide unique insights on the difference in dynamics under multiple knock-out cell lines.

We have identified four major insights, each of which can be further tested experimentally. The first insight is that the data is explained by at least three independent mechanisms. Our results suggest that there are multiple dynamic regimes for the intermediate process. For example a mechanism faster than Rad52 dependent HR is required to fit the experimental data to the model in data sets D2 and D3 (knockout of DNA-PKcs). Another interesting finding is that intermediate repair is increased in G2 phase of the cell cycle. If we assume that intermediate repair corresponds to alternative end joining, then this is in agreement with experimental results in the literature, supporting the existing biological evidence of the role of A-EJ in DSB repair [62]. This agrees with genetic studies that suggest two forms of alternative end joining depending on the presence of microhomology [40].

Our second insight is that the speed of A-EJ depends on the presence of regulating components in NHEJ and SSA, and in particular we observe a slower rate when Ku70 is inhibited and a faster rate when DNA-PKcs is inhibited (data sets D5-D6 and D2-D3 respectively, Fig. 3). As reported by experimental analysis, Ku deficient cells do not produce NHEJ products due to excessive degradation

or inhibition of end joining [11], and our model suggests that a slower A-EJ is active in these cells (data sets D5-D6). In addition, inhibition of DNA-PKcs does not activate repair by PARP-1 mediated A-EJ [24] and leads to elevated levels of resection and more HR [73]. Together with our model, this suggests that an alternative mechanism that is faster than PARP-1 mediated A-EJ could be activated when DNA-PKcs is inhibited (data sets D2-D3). These results could be tested by examining DSB repair in single cells with and without inhibitors using time-lapse microscopy and existing markers such as fluorescently tagged 53BP1, a protein that co-localises with DSBs, and fluorescent tagged PARP1, a candidate protein for A-EJ [28, 74].

The third insight that is generated from our analysis is the prediction of the total number of DSBs repaired by each mechanism. We can use this to estimate the proportion of different mutations following DSB repair in wildtype and mutant cells. Some cancers are deficient in at least one repair mechanism and in these cases, alternative mechanisms of repair have been observed to compensate [75]. One example is the increase in chromosomal aberrations observed in cells compromised of NHEJ by loss of Ku80 [3]. Recently, mutations specific to alternative mechanisms have been identified, where next generation sequencing has revealed sequence specific chromosome translocations following A-EJ at dysfunctional telomeres [5]. In addition, A-EJ is error prone, giving rise to chromosome translocations, of which there are more when NHEJ is inactive, suggesting it's role as a back up mechanism in eukaryotes [44]. If we know how many DSBs are likely to be repaired by each mechanism, this information will be important in predicting the numbers and types of mutations that we expect to observe. Potentially, a better understanding of the interplay between DSB repair mechanisms could be applied to design potential synthetic lethal therapeutics in cancer [76].

The fourth insight is that the gene expression profile of the proteins within different DSB repair mechanisms should change over time, with slower repair mechanisms still remaining active many hours after the initial dose of radiation. Pulse-like behaviour has been recorded in the repair of DSBs in human cells [28] and we suggest that this prediction could be further investigated using microarrays or sequencing. In fact, the expression of repair pathway genes has recently been used to diagnose the prognosis of carcinomas [77]. Currently the genes involved in the different repair pathways - and how much they are shared - remains to be fully elucidated. Our model could be used to predict the times at which different repair pathways dominate and provide a theoretical model for the interpretation of time-course gene expression results.

The model will require further development before it can be applied to more general problems in the DNA damage response. With additional data it will be possible to extend the model and include additional terms such as explicit repressive cross-talk interactions. It may be possible that inhibition of a certain protein may not completely ablate the function of a repair pathway. While our two-step model somewhat accounts for this, the actual pathway contains many different proteins and more complex effects could arise. Additionally, the PFGE assay provides a population average of the total repair; in order to obtain information on relatively small numbers of double strand breaks, together with estimates of cellular heterogeneity and stochasticity, data obtained by methods such as live cell imaging are required. Another current limitation arises from the heterogeneity of the DNA damage spectrum, where complex breaks can account for 30-40% of the population [78]. Previous methods of Monte Carlo track structure have been able to predict the number of single strand breaks, simple and complex DSBs that are created [59]. By identifying proteins responsible for the processing of complex DSBs and analysing knockout repair curves, it will be possible to gain understanding on the repair of heterogenous breaks, and incorporate these into our model. Despite our demonstra-

tion that the model is predictive across multiple doses in the low LET regime, we modelled repair following a variety of radiation doses, and we can't rule out that complexity of DSBs further alters the dynamics of repair. Previous studies have also modelled the formation of foci by summing the DSB enzymes involved in the phosphorylation of H2AX [48,52] and using this approach, our model could be developed to take into account lower doses of radiation or the dynamics of γ-H2AX foci formation. Related complicating factors are clustered DSBs [79] and DSBs residing in heterochromatic regions, which have been shown to require Artemis for repair [80].

From our simple assumptions we have built a predictive model and generated *in silico* data that was used to produce a number of unique insights that can be tested experimentally. Mathematical modelling not only facilitates the analysis of disparate data sets but also enforces the explicit formalisation of the underlying assumptions of our hypotheses. Our framework is another step towards a theoretical understanding of the dynamics of DNA repair pathways. The DNA damage response is comprised of a large number of extremely complex interacting biological processes. As the collection of larger and more heterogeneous data sets increases, we anticipate that mathematical modelling approaches will be absolutely essential for the reverse engineering and understanding of these complex biological processes.

Funding

Acknowledgements

References

[1] Mizuuchi K, Fisher LM, O'Dea MH, Gellert M. DNA gyrase action involves the introduction of transient double-strand breaks into DNA. Proc Natl Acad Sci U S A. 1980;77(4):1847–1851.

[2] Freifelder D. Lethal Changes in Bacteriophage DNA Produced by X-Rays. Radiation Research Supplement. 1966;6:pp. 80–96.

[3] Difilippantonio MJ, Zhu J, Chen HT, Meffre E, Nussenzweig MC, Max EE, et al. DNA repair protein Ku80 suppresses chromosomal aberrations and malignant transformation. Nature. 2000 mar;404(6777):510–514.

[4] Jasin M, Richardson C. Frequent chromosomal translocations induced by DNA double-strand breaks. Nature. 2000;405(6787):697–700.

[5] Mateos-Gomez PA, Gong F, Nair N, Miller KM, Lazzerini-Denchi E, Sfeir A. Mammalian polymerase θ promotes alternative NHEJ and suppresses recombination. Nature. 2015;518(7538):254–257. Available from: http://dx.doi.org/10.1038/nature14157.

[6] Sugawara N, Haber JE. Characterization of double-strand break-induced recombination: homology requirements and single-stranded DNA formation. Molecular and Cellular Biology. 1992;12(2):563–575.

[7] Moore JK, Haber JE. Cell cycle and genetic requirements of two pathways of nonhomologous end-joining repair of double-strand breaks in Saccharomyces cerevisiae. Molecular and Cellular Biology. 1996;16(5):2164–73.

[8] Moynahan ME, Jasin M. Loss of heterozygosity induced by a chromosomal double strand break. Proceedings of the National Academy of Sciences. 1997;94(17):8988–8993.

[9] Budman J, Chu G. Processing of DNA for nonhomologous end-joining by cell-free extract. The EMBO Journal. 2005;24(4):849–860.

[10] Rathmell WK, Chu G. Involvement of the Ku autoantigen in the cellular response to DNA double-strand breaks. Proceedings of the National Academy of Sciences. 1994;91(16):7623–7627.

[11] Liang F, Romanienko PJ, Weaver DT, Jeggo PA, Jasin M. Chromosomal double-strand break repair in Ku80-deficient cells. Proceedings of the National Academy of Sciences. 1996;93(17):8929–8933.

[12] Hammarsten O, Chu G. DNA-dependent protein kinase: DNA binding and activation in the absence of Ku. Proceedings of the National Academy of Sciences. 1998;95(2):525–530.

[13] Ma Y, Pannicke U, Schwarz K, Lieber MR. Hairpin Opening and Overhang Processing by an Artemis/DNA-Dependent Protein Kinase Complex in Nonhomologous End Joining and V(D)J Recombination . Cell. 2002;108(6):781 – 794.

[14] Grawunder U, Wilm M, Wu X, Kulesza P, Wilson TE, Mann M, et al. Activity of DNA ligase IV stimulated by complex formation with XRCC4 protein in mammalian cells. Nature;388(6641):492–495.

[15] DiBiase SJ, Zeng ZC, Chen R, Hyslop T, Curran WJ, Iliakis G. DNA-dependent Protein Kinase Stimulates an Independently Active, Nonhomologous, End-Joining Apparatus. Cancer Research. 2000;60(5):1245–1253.

[16] Beucher A, Birraux J, Tchouandong L, Barton O, Shibata A, Conrad S, et al. ATM and Artemis promote homologous recombination of radiation-induced DNA double-strand breaks in G2. The EMBO Journal. 2009;28(21):3413–3427.

[17] Ochi T, Blackford AN, Coates J, Jhujh S, Mehmood S, Tamura N, et al. PAXX, a paralog of XRCC4 and XLF, interacts with Ku to promote DNA double-strand break repair. Science. 2015;347(6218):185–188. Available from: `http://www.sciencemag.org/content/347/6218/185.abstract`.

[18] Pâques F, Haber JE. Multiple Pathways of Recombination Induced by Double-Strand Breaks in Saccharomyces cerevisiae. Microbiology and Molecular Biology Reviews. 1999;63(2):349–404.

[19] Lin FL, Sperle K, Sternberg N. Model for homologous recombination during transfer of DNA into mouse L cells: role for DNA ends in the recombination process. Molecular and Cellular Biology. 1984;4(6):1020–1034.

[20] Lin FL, Sperle K, Sternberg N. Recombination in mouse L cells between DNA introduced into cells and homologous chromosomal sequences. Proc Natl Acad Sci. 1985;82(5):1391–5.

[21] Deng SK, Gibb B, de Almeida MJ, Greene EC, Symington LS. RPA antagonizes microhomology-mediated repair of DNA double-strand breaks. Nat Struct Mol Biol. 2014;21(4):405–412. Available from: `http://dx.doi.org/10.1038/nsmb.2786`.

[22] Xiong X, Du Z, Wang Y, Feng Z, Fan P, Yan C, et al. 53BP1 promotes microhomology-mediated end-joining in G1-phase cells. Nucleic Acids Research. 2015;43(3):1659–1670. Available from: `http://nar.oxfordjournals.org/content/43/3/1659.abstract`.

[23] Mansour WY, Rhein T, Dahm-Daphi J. The alternative end-joining pathway for repair of DNA double-strand breaks requires PARP1 but is not dependent upon microhomologies. Nucleic Acids Research. 2010;38(18):6065–6077. Available from: `http://nar.oxfordjournals.org/content/38/18/6065.abstract`.

[24] Wang M, Wu W, Wu W, Rosidi B, Zhang L, Wang H, et al. PARP-1 and Ku compete for repair of DNA double strand breaks by distinct NHEJ pathways. Nucleic Acids Research. 2006;34(21):6170–6182.

[25] Schipler A, Iliakis G. DNA double-strand-break complexity levels and their possible contributions to the probability for error-prone processing and repair pathway choice. Nucleic Acids Research. 2013;41(16):7589–7605.

[26] Vilenchik MM, Knudson AG. Endogenous DNA double-strand breaks: Production, fidelity of repair, and induction of cancer. Proceedings of the National Academy of Sciences. 2003;100(22):12871–12876.

[27] Branzei D, Foiani M. Regulation of DNA repair throughout the cell cycle. Nat Rev Mol Cell Biol. 2008;9(4):297–308. Available from: http://dx.doi.org/10.1038/nrm2351.

[28] Karanam K, Kafri R, Loewer A, Lahav G. Quantitative Live Cell Imaging Reveals a Gradual Shift between DNA Repair Mechanisms and a Maximal Use of HR in Mid S Phase. Molecular Cell. 2012;47(2):320 – 329.

[29] Hentges P, Waller H, Reis CC, Ferreira MG, Doherty AJ. Cdk1 Restrains NHEJ through Phosphorylation of XRCC4-like Factor Xlf1. Cell Reports. 2015;9(6):2011–2017. Available from: http://www.cell.com/cell-reports/abstract/S2211-1247(14)01010-9.

[30] Trujillo KM, Yuan SSF, Lee EYHP, Sung P. Nuclease Activities in a Complex of Human Recombination and DNA Repair Factors Rad50, Mre11, and p95. Journal of Biological Chemistry. 1998;273(34):21447–21450.

[31] Deng X, Prakash A, Dhar K, Baia GS, Kolar C, Oakley GG, et al. Human Replication Protein A Rad52 Single Stranded DNA Complex: Stoichiometry and Evidence for Strand Transfer Regulation by Phosphorylation. Biochemistry. 2009;48(28):6633–6643. PMID: 19530647.

[32] Mortensen UH, Bendixen C, Sunjevaric I, Rothstein R. DNA strand annealing is promoted by the yeast Rad52 protein. Proc Natl Acad Sci U S A. 1996;93(20):10729–34.

[33] Götlich B, Reichenberger S, Feldmann E, Pfeiffer P. Rejoining of DNA double-strand breaks in vitro by single-strand annealing. European Journal of Biochemistry. 1998;258(2):387–395.

[34] Wang H, Zeng ZC, Bui TA, Sonoda E, Takata M, Takeda S, et al. Efficient rejoining of radiation-induced DNA double-strand breaks in vertebrate cells deficient in genes of the RAD52 epistasis group. Oncogene. 2001;20(18).

[35] Fishman-Lobell J, Rudin N, Haber JE. Two alternative pathways of double-strand break repair that are kinetically separable and independently modulated. Molecular and Cellular Biology. 1992;12(3):1292–1303.

[36] Ma JL, Kim EM, Haber JE, Lee SE. Yeast Mre11 and Rad1 Proteins Define a Ku-Independent Mechanism To Repair Double-Strand Breaks Lacking Overlapping End Sequences. Molecular and Cellular Biology. 2003;23(23):8820–8828. Available from: http://mcb.asm.org/content/23/23/8820.abstract.

[37] Iliakis G. Backup pathways of NHEJ in cells of higher eukaryotes: Cell cycle dependence. Radiotherapy and Oncology. 2009;92(3):310 – 315. Special Issue on Molecular and Experimental Radiobiology Including papers from the 11th International Wolfsberg Meeting on Molecular Radiation Biology/Oncology. Available from: http://www.sciencedirect.com/science/article/pii/S0167814009003351.

[38] Wang H, Perrault AR, Takeda Y, Qin W, Wang H, Iliakis G. Biochemical evidence for Ku-independent backup pathways of NHEJ. Nucleic Acids Research. 2003;31(18):5377–5388. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC203313/.

[39] Mladenov E, Magin S, Soni A, Iliakis G. DNA Double-Strand Break Repair as determinant of cellular radiosensitivity to killing and target in radiation therapy. Frontiers in Oncology. 2013;3(113).

[40] Decottignies A. Alternative end-joining mechanisms: a historical perspective. Frontiers in Genetics. 2013;4(48). Available from: http://www.frontiersin.org/cancer_genetics/10.3389/fgene.2013.00048/abstract.

[41] Zhang Y, Jasin M. An essential role for CtIP in chromosomal translocation formation through an alternative end-joining pathway. Nat Struct Mol Biol. 2011;18(1):80–84.

[42] Audebert M, Salles B, Calsou P. Involvement of Poly(ADP-ribose) Polymerase-1 and XRCC1/DNA Ligase III in an Alternative Route for DNA Double-strand Breaks Rejoining. Journal of Biological Chemistry. 2004;279(53):55117–55126.

[43] Liang L, Deng L, Nguyen SC, Zhao X, Maulion CD, Shao C, et al. Human DNA ligases I and III, but not ligase IV, are required for microhomology-mediated end joining of DNA double-strand breaks. Nucleic Acids Research. 2008;36(10):3297–3310.

[44] Soni A, Siemann M, Grabos M, Murmann T, Pantelias GE, Iliakis G. Requirement for Parp-1 and DNA ligases 1 or 3 but not of Xrcc1 in chromosomal translocation formation by backup end joining. Nucleic Acids Research. 2014;42(10):6380–6392. Available from: http://nar.oxfordjournals.org/content/42/10/6380.abstract.

[45] Howard SM, Yanez DA, Stark JM. DNA Damage Response Factors from Diverse Pathways, Including DNA Crosslink Repair, Mediate Alternative End Joining. PLoS Genet. 2015;11(1):e1004943. Available from: http://dx.doi.org/10.1371%2Fjournal.pgen.1004943.

[46] L M, G I. Kinetics of DNA double-strand break repair throughout the cell cycle as assayed by pulsed field gel electrophoresis in CHO cells. Int J Radiat Biol. 1991;59(1).

[47] Verbruggen P, Heinemann T, Manders E, von Bornstaedt G, van Driel R, Höfer T. Robustness of DNA Repair through Collective Rate Control. PLoS Comput Biol. 2014;10(1):e1003438.

[48] Cucinotta FA, Pluth JM, Anderson JA, Harper JV, O'Neill P. Biochemical Kinetics Model of DSB Repair and Induction of gamma-H2AX Foci by Non-homologous End Joining. Radiation Research. 2008;169(2):214–222.

[49] Taleei R, Weinfeld M, Nikjoo H. Single strand annealing mathematical model for double strand break repair. journal of Molecular Engineering and Systems Biology. 2012;1(1).

[50] Taleei R, Nikjoo H. The Non-homologous End-Joining (NHEJ) Pathway for the Repair of DNA Double-Strand Breaks: I. A Mathematical Model. Radiation Research. 2013;179(5):530–539.

17

[51] Taleei R, Weinfeld M, Nikjoo H. A kinetic model of single-strand annealing for the repair of DNA double-strand breaks. Radiation Protection Dosimetry. 2011;143(2-4):191–195.

[52] Belov OV, Krasavin EA, Lyashko MS, Batmunkh M, Sweilam NH. A quantitative model of the major pathways for radiation-induced DNA double-strand break repair. Journal of Theoretical Biology. 2015;366(0):115 – 130. Available from: http://www.sciencedirect.com/science/article/pii/S0022519314005633.

[53] Friedland W, Kundrát P, Jacob P. Stochastic modelling of DSB repair after photon and ion irradiation. International Journal of Radiation Biology. 2012;88(1-2):129–136.

[54] Friedland W, Jacob P, Kundrát P. Stochastic Simulation of DNA Double-Strand Break Repair by Non-homologous End Joining Based on Track Structure Calculations. Radiation Research. 2013;173(5):677–688.

[55] Li Y, Qian H, Wang Y, Cucinotta FA. A Stochastic Model of DNA Fragments Rejoining. PLoS ONE. 2012;7(9):e44293.

[56] Bodgi L, Foray N. On the coherence between mathematical models of DSB repair and physiological reality. Mutation Research/Genetic Toxicology and Environmental Mutagenesis. 2014;761:48 – 49. Available from: http://www.sciencedirect.com/science/article/pii/S1383571814000047.

[57] Taleei R, Girard PM, Nikjoo H. DSB repair model for mammalian cells in early S and G1 phases of the cell cycle: Application to damage induced by ionizing radiation of different quality. Mutation Research/Genetic Toxicology and Environmental Mutagenesis. 2015;779:5 – 14.

[58] Taleei R, Nikjoo H. Response to the Letter of Bodgi and Foray: On the coherence between mathematical models of DSB repair and physiological reality. Mutation Research/Genetic Toxicology and Environmental Mutagenesis. 2014;761:50 – 52.

[59] Taleei R, Girard PM, Sankaranarayanan K, Nikjoo H. The Non-homologous End-Joining (NHEJ) Mathematical Model for the Repair of Double-Strand Breaks: II. Application to Damage Induced by Ultrasoft X Rays and Low-Energy Electrons. Radiation Research. 2013 2015/11/02;179(5):540–548. Available from: http://dx.doi.org/10.1667/RR3124.1.

[60] Nikjoo H, Uehara S, Emfietzoglou D, Cucinotta FA. Track-structure codes in radiation research. Radiation Measurements. 2006;41(9-10):1052 – 1074. Space Radiation Transport, Shielding, and Risk Assessment Models. Available from: http://www.sciencedirect.com/science/article/pii/S1350448706001338.

[61] Windhofer F, Wu W, Iliakis G. Low levels of DNA ligases III and IV sufficient for effective NHEJ. Journal of Cellular Physiology. 2007;213(2):475–483.

[62] Wu W, Wang M, Wu W, Singh SK, Mussfeldt T, Iliakis G. Repair of radiation induced DNA double strand breaks by backup NHEJ is enhanced in G2. DNA Repair. 2008;7(2):329–338.

[63] Iliakis G, Wang H, Perrault AR, Boecker W, Rosidi B, Windhofer F, et al. Mechanisms of DNA double strand break repair and chromosome aberration formation. Cytogenetic and Genome Research. 2004;104(1-4):14–20.

[64] Zhang LY, Chen LS, Sun R, JI SJ, Ding YY, Wu J, et al. Effects of expression level of DNA repair-related genes involved in the NHEJ pathway on radiation-induced cognitive impairment. Journal of Radiation Research. 2013;54(2):235–242. Available from: `http://jrr.oxfordjournals.org/content/54/2/235.abstract`.

[65] Gillespie DT. Exact stochastic simulation of coupled chemical reactions. J Phys Chem. 1977;81(25):2340–2361.

[66] Toni T, Welch D, Strelkowa N, Ipsen A, Stumpf MPH. Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. Journal of the Royal Society, Interface / the Royal Society. 2009;6(31):187–202.

[67] Toni T, Stumpf MPH. Simulation-based model selection for dynamical systems in systems and population biology. Bioinformatics. 2010 Jan;26(1):104–110.

[68] Liepe J, Barnes C, Cule E, Erguler K, Kirk P, Toni T, et al. ABC-SysBio–approximate Bayesian computation in Python with GPU support. Bioinformatics. 2010;26(14):1797–1799. Available from: `http://bioinformatics.oxfordjournals.org/content/26/14/1797.abstract`.

[69] Liepe J, Kirk P, Filippi S, Toni T, Barnes CP, Stumpf MPH. A framework for parameter estimation and model selection from experimental data in systems biology using approximate Bayesian computation. Nature protocols. 2014;9(2):439–456.

[70] Spiegelhalter DJ, Best NG, Carlin BP. Bayesian measures of model complexity and fit. Journal of the Royal Statistical Society, Series B. 2002;.

[71] Akaike H. A new look at the statistical model identification. Automatic Control, IEEE Transactions on. 1974 Dec;19(6):716–723. Available from: `http://dx.doi.org/10.1109/tac.1974.1100705`.

[72] Olivier F, Guillaume L. Deviance Information Criteria for Model Selection in Approximate Bayesian Computation. Statistical Applications in Genetics and Molecular Biology. 2011;10(1):1–25. Available from: `http://EconPapers.repec.org/RePEc:bpj:sagmbi:v:10:y:2011:i:1:n:33`.

[73] Shibata A, Conrad S, Birraux J, Geuting V, Barton O, Ismail A, et al. Factors determining DNA double-strand break repair pathway choice in G2 phase. The EMBO Journal. 2011;30(6):1079–1092.

[74] Haince JF, McDonald D, Rodrigue A, Dery U, Masson JY, Hendzel MJ, et al. PARP1-dependent Kinetics of Recruitment of MRE11 and NBS1 Proteins to Multiple DNA Damage Sites. Journal of Biological Chemistry. 2008;283(2):1197–1208. Available from: `http://www.jbc.org/content/283/2/1197.abstract`.

[75] Ceccaldi R, Liu JC, Amunugama R, Hajdu I, Primack B, Petalcorin MIR, et al. Homologous-recombination-deficient tumours are dependent on Polθ-mediated repair. Nature. 2015;518(7538):258–262. Available from: http://dx.doi.org/10.1038/nature14184.

[76] Cho NW, Greenberg RA. DNA repair: Familiar ends with alternative endings. Nature. 2015;518(7538):174–176. Available from: http://dx.doi.org/10.1038/nature14200.

[77] Pitroda SP, Pashtan IM, Logan HL, Budke B, Darga TE, Weichselbaum RR, et al. DNA Repair Pathway Gene Expression Score Correlates with Repair Proficiency and Tumor Sensitivity to Chemotherapy. Science Translational Medicine. 2014;6(229):229ra42–229ra42. Available from: http://stm.sciencemag.org/content/6/229/229ra42.

[78] Dobbs TA, Palmer P, Maniou Z, Lomax ME, O'Neill P. Interplay of two major repair pathways in the processing of complex double-strand DNA breaks. DNA Repair. 2008;7(8):1372 – 1383. Available from: http://www.sciencedirect.com/science/article/pii/S1568786408001845.

[79] Sutherland BM, Bennett PV, Sidorkina O, , Laval J. Clustered Damages and Total Lesions Induced in DNA by Ionizing Radiation: Oxidized Bases and Strand Breaks. Biochemistry. 2000;39(27):8026–8031.

[80] Woodbine L, Brunton H, Goodarzi AA, Shibata A, Jeggo PA. Endogenously induced DNA double strand breaks arise in heterochromatic DNA regions and require ataxia telangiectasia mutated and Artemis for their repair. Nucleic Acids Research. 2011;39(16):6986–6997. Available from: http://nar.oxfordjournals.org/content/39/16/6986.abstract.
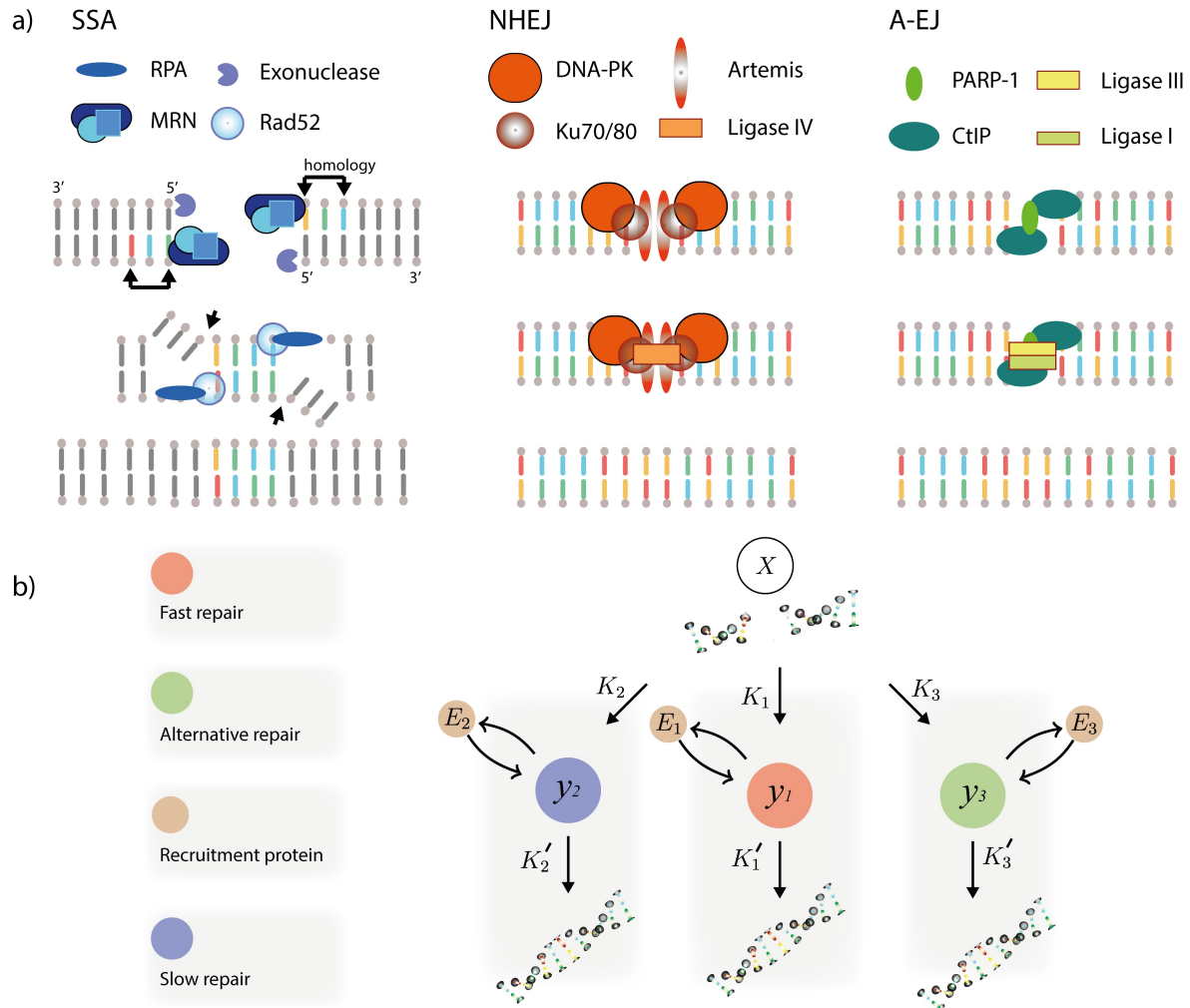
Figures

Figure 1: Modelling multiple repair mechanisms. a) Proteins and repair steps contributing to repair during SSA, NHEJ and A-EJ in mammalian cells (illustration). b) The model. Filled circles represent species and arrows represent reactions.
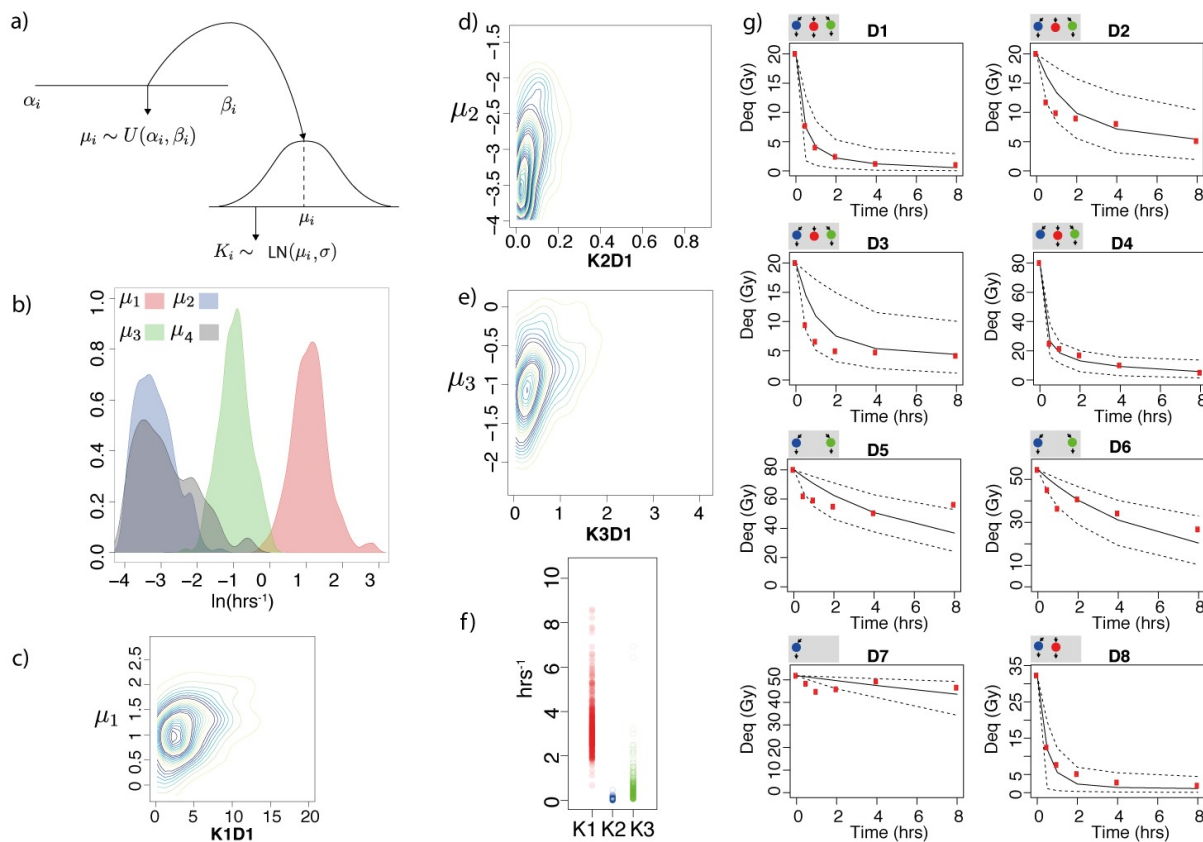
Figure 2: a) Diagram showing the parameter sampling process. Hyperparameters μ_i are drawn from a uniform distribution between α_i and β_i. Model parameters K_i are sampled from a lognormal distribution with mean μ_i and variance σ^2. b) Posterior distributions for the hyperparameters μ_{1-3} and μ_4. c-e) Posterior analysis for data set $D1$. Marginal distributions of the parameters K_{1D1}, K_{2D1}, K_{3D1} against the hyperparameters. f) Posterior distributions of the parameters K_{1D1}, K_{2D1}, K_{3D1}, showing some overlap. g) Time series plots of the experimental data and model fit. Subfigures to the side of each figure represent the active repair mechanisms. The y axis represents the dose equivalent in units of Gray (Gy) (1 DSB = 0.0286 Gy). Red, blue and green represent fast, slow and alternative repair, black solid lines are the median fit and the dashed black lines represent the 0.95 credible regions.
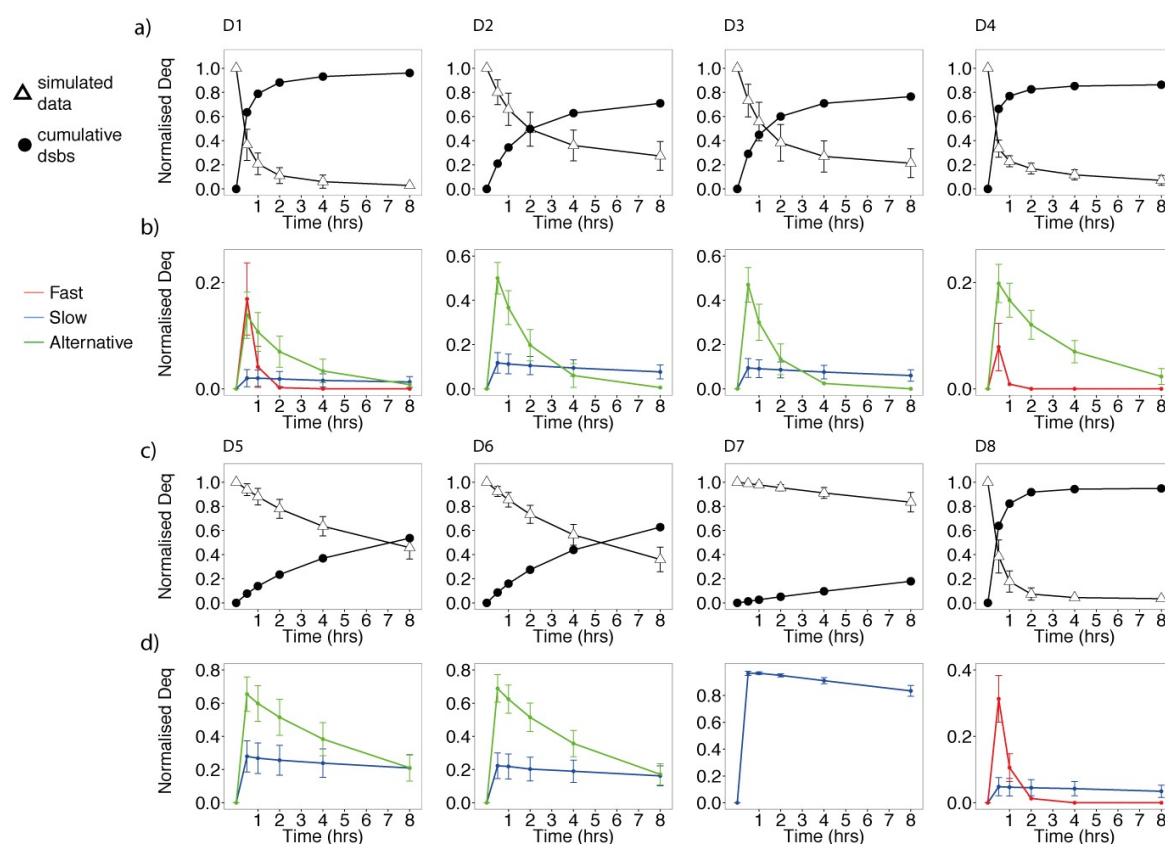
Figure 3: Normalised repair, cumulative repair and the proportion of DSBs within each mechanism as a percentage of the total DSBs for data D1-D8. Plots a) and c) show DSB repair curves simulated from the posterior together with the cumulative number of DSBs repaired. Plots b) and d) show the DSBs entering each repair mechanism. The errorbars represent the 0.5 credible regions (0.25,0.75). Red, blue and green represent fast, slow and alternative repair.
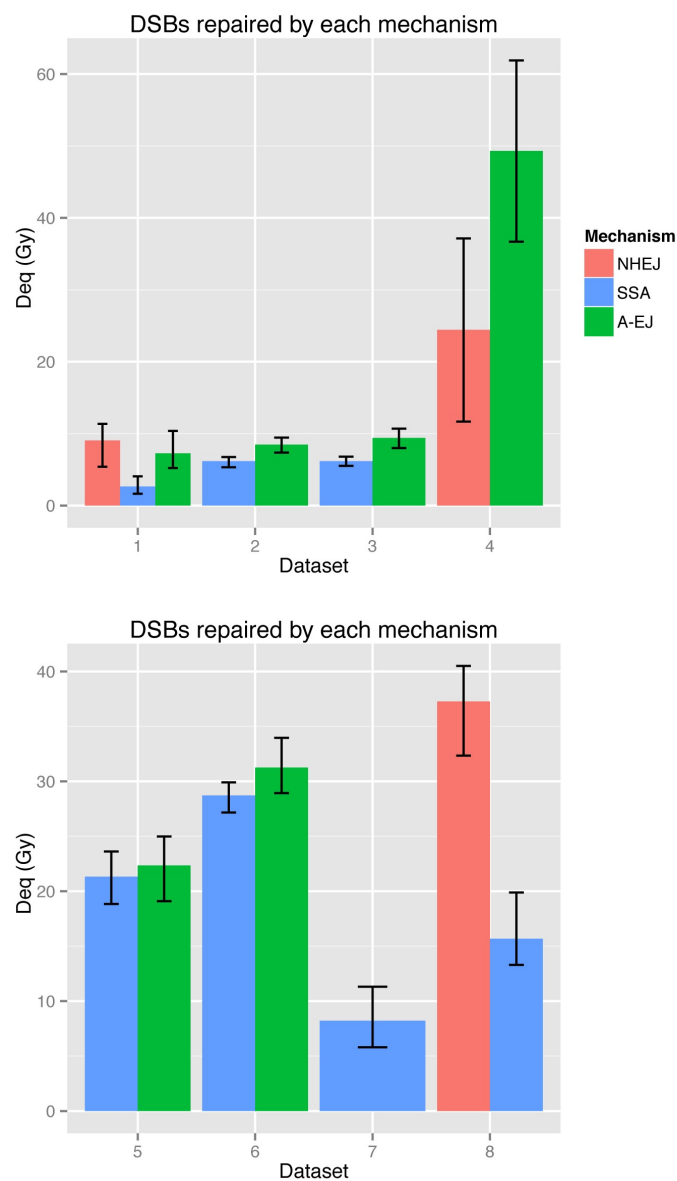
Figure 4: $\tilde{N}_i$, the total number of DSBs repaired by each mechanism for each data set (dose equivalent (Gy)). The errorbars represent the 0.5 credible regions (0.25,0.75). Red, blue and green represent fast, slow and alternative repair.
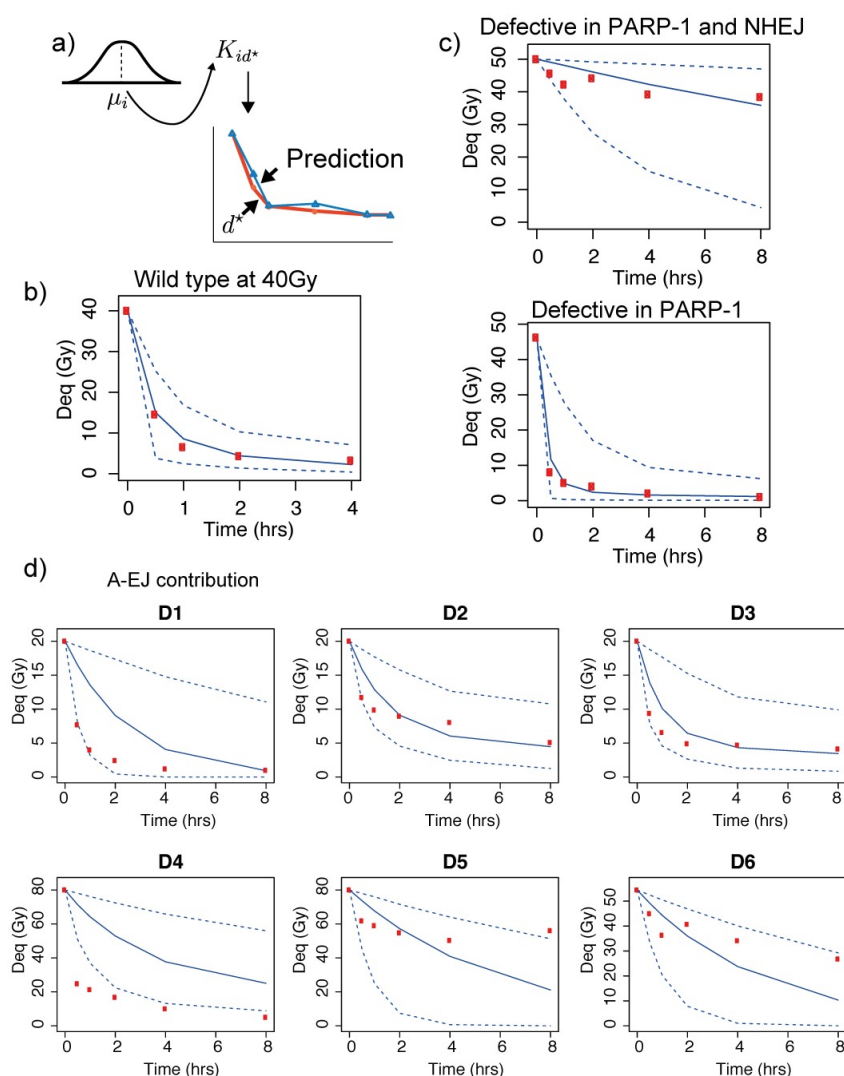
Figure 5: Model predictions. a) Parameters were drawn from the hyperparameter distribution and the code was simulated again under different initial conditions (blue curve) and compared to different data sets (red points) b). Model prediction (blue curve) of wild type repair (red points) from our initial parameter fit (Fig. 2g). c). PARP-1 inhibited repair predictions. The initial parameter fit (blue curves) provided a good fit to the experimental data. d). By setting the parameters of NHEJ and SSA to zero in our predicted posterior, we were able to simulate the expected contribution of A-EJ in our data sets where A-EJ is assumed to be active. Number of DSBs remaining in each of the data sets are shown in the grey shaded region. b-d) Number of DSBs are given in dose equivalent units (Gy). Blue solid lines are the median points, dashed blue lines are the 0.95 credible regions.
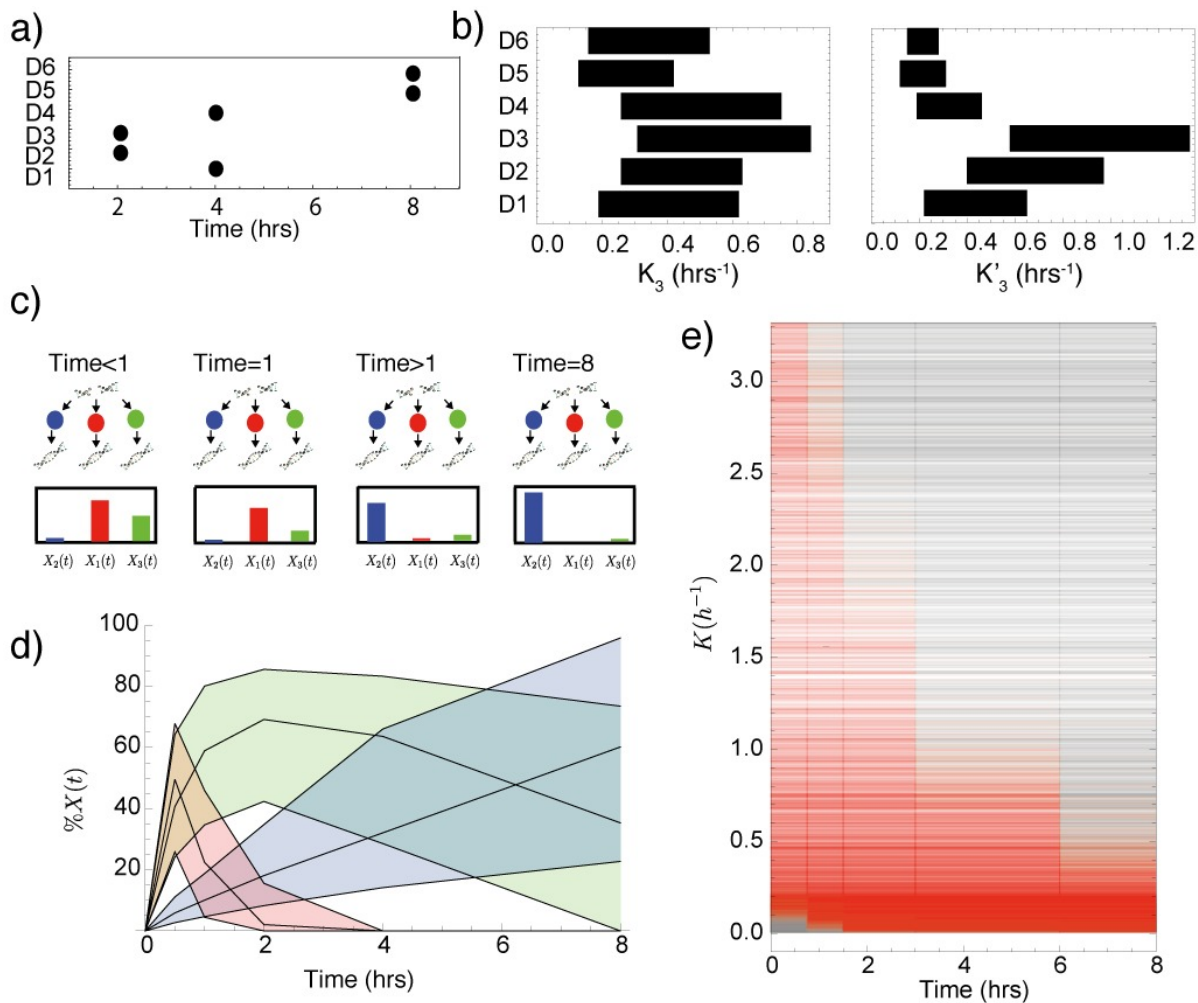
Figure 6: a). The time in which a repair curve reaches below half maximal value for each data set in which A-EJ is assumed to be active. The slowest mode of repair occurred in data sets 5 and 6, where Ku70 is inactive. b). Rectangle plot of the interquartile ranges of K_3 and K'_3 for all data sets where A-EJ is assumed to be active. c). Illustration, showing a typical distribution of the DSBs that remain to be repaired over time. For times < 1 hour a large proportion of DSBs are being repaired by fast NHEJ and faster A-EJ mechanisms, whereas at later times, the majority of DSBs reside in slower HR mechanisms. d). Time series showing the percentage of remaining DSBs in each repair pathway for the wild type data D1. e). Plot showing the time at which each repair mechanism is greater than 30% active for different parameter values. Grey indicates that the mechanism is less than 30% active and red indicates the mechanism is greater than 30% active.

26