Statistical analysis of low dose rate mouse experiments with WGS technology and quantitative reproduction of mutation frequency using Whack-a-Mole model

H. Toki[1], Y. Gondo[2], A. Uchimura[3], Y. Tsunoyama[4], and M. Bando[5]

1 Research Center for Nuclear Physics, Osaka University, Osaka 567-0047, Japan
2 Department of Molecular Life Sciences, Tokai University School of Medicine, Kanagawa 259-1193, Japan
3 Department of Molecular Biosciences, Radiation Effects Research Foundation, Hiroshima, 732-0815, Japan
4 Agency for Health, Safety and Environment, Kyoto University, Kyoto 606-8501, Japan
5 Yukawa Institute for Theoretical Physics, Kyoto University, Kyoto 606-8502, Japan

Abstract

Gondo and his collaborators performed whole-genome sequencing (WGS) experiments using mice for several generations under the circumstances of radiation with low dose rates at Tokai University and Institute for Environmental Sciences (IES) in Japan. The dose rates were 0.05, 0.15, 1.0, 20mGy/day. We performed a statistical analysis of the experimental results on the single nucleotide variations (SNV) for each mouse after 4 generations. The SNVs per generation were almost unchanged below the dose rate of 1mGy/day within the 95% confidence level errors. The experimental results were reproduced by the Whack-a-Mole (WAM) model with the standard parameters obtained by the Russell mega-mouse experiments. We propose a natural dose-rate unit (NDR) to show the radiation effects at a low dose rate for a long duration to be used in society.

1 WGS experiments with low dose rates for several generations

Recently, Gondo and his collaborators performed WGS experiments using mice with low dose rates for several generations in Tokai University and IES. The dose rates were 0.05, 0.15, 1.0, 20mGy/day. The experimental details and the results will be presented orally by Gondo in the ICRP Tokyo conference [1].

We show here one of their results on mice with the family sequences for four generations under low dose-rate radiation in Fig. 1. In this paper, we present the results of the statistical analysis of these

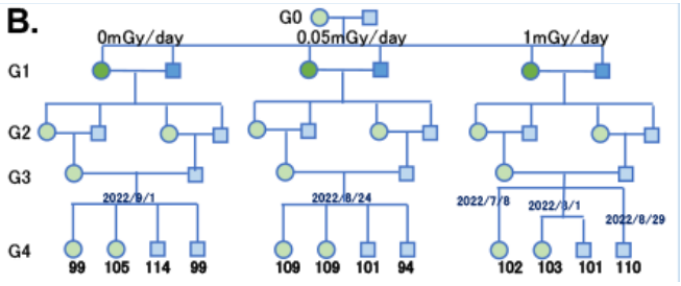

Figure 1. The family sequences of mice for four generations with the dose rates of 0.05mGy/day and 1mGy/day for about 400 days. The numbers of single nucleotide variations (SNVs) were obtained by the short read method of the WGS technology and written under each mouse in the G4 generation.

SNVs using the maximum likelihood method and the effect of the low dose-rate radiation on the numbers of SNVs per generation using the family structure under experimental conditions. We then compare the results with the Whack-a-Mole (WAM) model predictions using the standard parameter set obtained in the analysis of Russell's mega-mouse experiment [3].

2 Statistical analysis of the SNVs with long time exposure on mice

We use the maximum likelihood method with the Poisson distribution on the numbers of SNVs in various mice experiments. The numbers of de novo mutations (DNM) in various cases are listed in the paper

of Gondo and his collaborators [1]. Here, the DNMs are defined as newly detected variations in the G4 generation as compared to the DNA sequence in the G0 generation. We focus here on the SNVs identified in the WGS technology [2]. We used the likelihood function with Poisson distribution.

$$L(\lambda) = \Pi_{i=1}^{N} \frac{\lambda^{k_i}}{k_i!} e^{-\lambda} \ , \tag{1}$$

where k_i is the number of SNVs for N data points and λ is the most likelihood value in case $L(\lambda)$ is maximized. We plot the experimental data on the SNVs for the control case in Fig. 2. In the right figure, shown is the logarithm of the likelihood function LL as a function of λ. The maximum of LL provides the average value for λ and λ at $LL_{max} - 2$ provides the 95% confidence level.

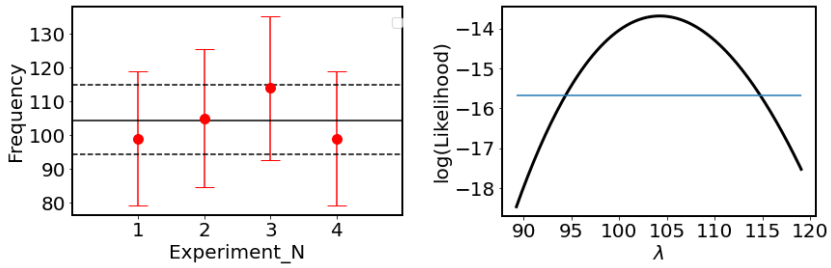

Figure 2. The experimental data points on the SNVs for the control case are shown in the left figure. The 95% error bars are shown for experimental data points. Using the most likelihood method, the average value and the 95% confidence level are shown by the solid line and the dashed lines. The log of the likelihood function $LL(\lambda)$ is given for the control case in the right figure. The 95% confidence level is provided by the value $LL_2 = LL_{max} - 2$ shown by the horizontal line.

We performed the same procedure for the exposure cases. Shown in Fig. 3 are the experimental data on the SNVs for $d = 0.05$ and 1mGy/day. The average value λ and the 95% confidence levels are provided by the solid and the dashed lines. We prefer the most likelihood method to determine the 95% confidence level since the error bars systematically decrease with the number of data points. The standard errors change statistically when the number of data points is small. The distribution of the data points approaches the Poisson distribution by increasing the number of data points. Both methods then provide the same information at a large number of data points.

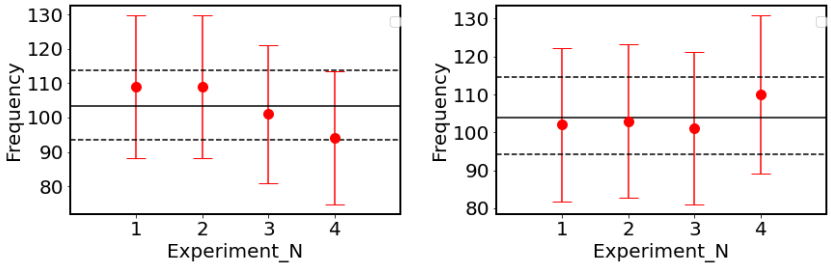

Figure 3. The experimental data on the SNVs for $d = 0.05$ and $d =$ 1mGy/day. The most likelihood value for the λ is provided by the solid line and the 95% confidence level is given by the dashed lines.

In addition to these cases with $d = 0.05$ and 1mGy/day for about 400 days with four generations, we have data on $d = 20$mGy/day for about 300 days with three generations. The family structure is shown in Fig. 4. We should note that a female mouse under $d =$ 20mGy/day radiation cannot give birth. Hence, a female mouse in the control case was used for mating with a male mouse under $d =$ 20mGy/day radiation. We used the most likelihood method for a

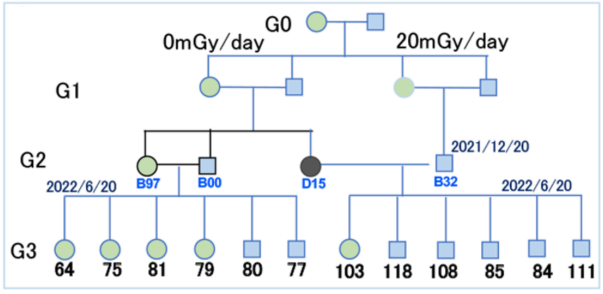

Figure 4. The family structure of a mice experiment under the dose rate of 20mGy/day. The numbers of SNVs are written under mice in the G3 generation.

statistical analysis of this experimental data. We summarize the most likelihood values λ on these SNVs in Fig. 5. In the left figure, the SNVs are shown as a function of the total dose. In this plot, it is obvious that the control value of $d = 20$mGy/day case shown at the left-lowest corner is different from those values of the lower dose-rate cases shown at the left-upper corner. Hence, we plot the SNVs per generation in the right figure of Fig. 5. This result is reasonable since the control

values taken under various conditions agree with each other. Here, the horizontal axis is the dose rate. If we were to provide the total dose per generation, we should multiply 100 days by the dose rate. We see now that the SNVs per generation are unchanged until the dose rate of 1mGy/day within the statistical error.

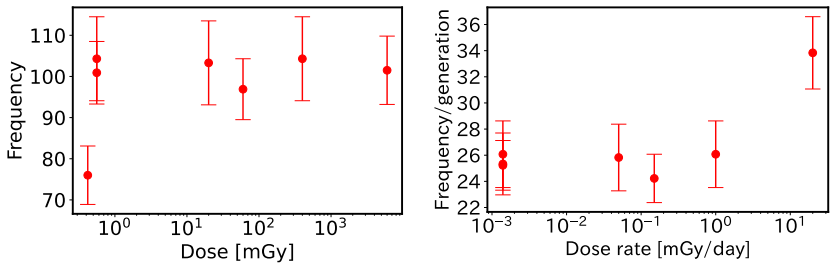

Figure 5. The SNVs for the control cases and the exposure cases are shown as a function of the total dose in the left figure. Here, we add the data points taken at 0.15mGy/day for four generations obtained at Tokai University [1]. The SNVs per generation are plotted as a function of the dose rate in the right figure. The SNVs per generation are unchanged until 1mGy/day within the 95% confidence level.

3 Theoretical analysis of SNVs per generation using Mendel's law

We have found that the SNVs per generation behave nicely for the control cases, where the values in all the control cases agree with each other within the error bars. Hence, now it is important to extract the SNVs newly produced in each generation from the average values λ obtained for the G4 mice. We assume that the SNVs for each generation are common, and the SNVs of the ancestor propagate to their descendants following Mendel's law with the resulting total SNVs measured in the G4 mice. Here, we use the sum rule of the Poisson distribution $P(i,a) = \frac{a^i}{i!}e^{-a}$.

$$\sum_i P(i,a)P(k-i,b) = P(k,a+b) \ . \tag{2}$$

This relation means a Poisson distribution with the average of a together with a Poisson distribution with the average of b provide a

Poisson distribution with the average of $a + b$ as a net result. We write the number k of SNVs for each generation following the Poisson distribution with the average value x for the control case.

Obviously, the SNVs follow the Poisson distribution $P(k, x)$ for a G4 mouse appearing newly in the fertilized egg of the G4 mouse. The male mouse in the G3 generation has SNVs of a Poisson distribution with the average x, propagating to the G4 generation by half due to Mendel's law. This rule applies to the female mouse, which contributes also $x/2$ to the G4 generation. Altogether the G3 generation contributes to SNVs in G4 by the Poisson distribution with the average $x/2 + x/2 = x$. The contribution from the G2 generation to the G4 generation can be calculated as $x/2 + x/2 = x$ using the family structure. The contribution from the G1 generation is slightly complicated due to the inter-family breading. We discuss only the average value for this case. The SNVs in the fertilized egg of the G1 male mouse propagate half of the SNVs to the G2 mouse, and the G1 female mouse also provides half of the SNVs. The net SNVs are x. In the next step, the G3 mouse gets a homo SNV and 6 hetero SNVs out of 16 possibilities. A complicated calculation for the G4 mouse coming from the G3 parents provide 1 homo SNV and 6 hetero SNVs in the G4 mouse. Since the experimental numbers for the SNVs are obtained by counting a homo SNV as one and a hetero SNV also as one SNV, we should sum up the homo and hetero contributions. Hence, the G1 contribution to the G4 mouse is $2 \times 7/16x$. Hence, we sum up all the G1, G2, G3, and G4 contributions to get the net SNVs in the G4 generation using the Poisson distribution sum rule. The average value of the net SNVs in the G4 generation is

$$x + x + x + 7/8x = 3.875x \ . \tag{3}$$

This net value has to be equated with λ extracted by the maximum likelihood method in the previous section. We get the following value for x and the 95% confidence level.

$$x = 26.9 \ \ [24.3, 29.5] \ . \tag{4}$$

As for the exposure case, the low dose rate irradiation started when the male-female couple in the G1 case were placed in the same cage to be ready for mating. Hence, the fertilized eggs of the G1 mice did not get the influence of radiation. On the other hand, the G2 mice and further generations have the influence of radiation. In this case,

we write the number of SNVs in the fertilized egg is y, which should be fixed for various dose rates. Since the family structure of the exposure case is the same as the control case, the net SNVs in the G4 generation are written as

$$y + y + y + 7/8x = 3y + 0.875x \ . \tag{5}$$

This net value has to be equated with λ for the exposure cases knowing x. The results are

$$\begin{aligned} y &= 26.6 \quad [23.1, 30.1] \quad \text{for } 0.05\text{mGy/day} \ , \\ y &= 26.8 \quad [23.3, 30.3] \quad \text{for } 1\text{mGy/day} \ . \end{aligned} \tag{6}$$

As for the 20mGy/day case, the family structure shown in Fig. 4 is different from the cases of the lower dose rate cases. In the control case, the SNVs in the fertilized egg of the male mouse is x, and that of the female mouse are also x in the G1 generation. The SNVs propagate to the G2 generation as x. Finally, to the G3 generation, it becomes $7/8$ x. We can calculate the G2 contribution to the G3 generation as x. Finally, the SNVs in the G3 generation are x. Hence the net SNVs for the control case is

$$x + x + 7/8x = 2.875x \ . \tag{7}$$

Hence, we get

$$x = 26.4 \quad [24.0, 28.9] \ . \tag{8}$$

Now we have to extract the exposure case, where the family structure is written in Fig. 4. The SNVs newly produced at the fertilized egg in the G1 mice do not get the influence of the radiation. Hence, we write the average value x. These SNVs propagate to the G2 level by half of this value x. The female mouse has independent SNVs by the average value of x. These SNVs propagate to the G3 mice. Hence, the contribution of the G1 mice to the G3 mouse is x. As for the G2 mouse in the exposure case, the male mouse has y SNVs and the female mouse has x. Hence, the G2 contribution is $(x + y)/2$. The G3 generation mouse has contributions both from the male and female mice. Assuming the contributions of male and female for the fertilized egg in the G3 generation are the same, again the SNVs of the G3 generation is $(x + y)/2$. Hence the net result is

$$2x \times 1/2 + (x + y) \times 1/2 + (x + y) \times 1/2 = 2x + y \ . \tag{9}$$

We equate this expression with the observation for the exposure 20mGy/day case. We get the following values.

$$y = 41.2 \quad [35.2, 47.3] \ . \tag{10}$$

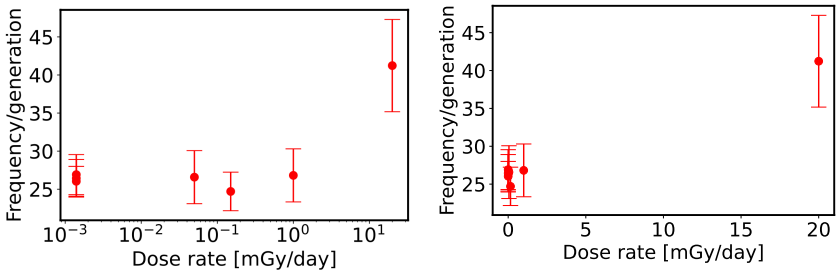

Figure 6. The calculated SNVs per generation are shown as a function of the dose rate in the log scale in the left figure. The same SNVs per generation are shown in the linear scale in the right figure.

We plot all the results in Fig. 6. The calculated SNVs per generation are shown as a function of the dose rate in the log scale in the left figure. The same SNVs per generation are shown in the linear scale in the right figure. Here, it is clear that the SNVs per generation are unchanged until 1mGy/day within the statistical errors. We can see this feature clearly in the log plot.

4 The WAM model

The WAM model was used for the analysis of the mega-mouse experiments of Russell and Kelly [3]. The WAM model is expressed by the differential equation for the mutation frequency F [4].

$$\frac{dF}{dt} = A - BF \ . \tag{11}$$

There are two terms in the WAM equation. A term increases the mutation frequency, while the B term decreases. The solution is

$$F = \frac{A}{B}(1 - e^{-Bt}) \ . \tag{12}$$

For a steady state for $t \gg 1/B$, the mutation frequency becomes

$$F = \frac{a_0 + a_1 d}{b_0 + b_1 d} \ . \tag{13}$$

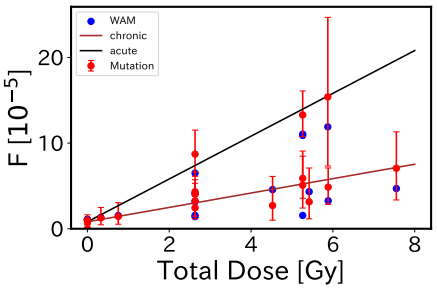

Figure 7. The mega-mouse experimental results are shown by the red points with the 95% confidence level errors. The WAM predictions are shown by the blue points. Two straight lines introduced by Russell-Kelly are also shown in this figure [3].

This mutation frequency should be used for the low dose and long exposure cases. In the above expression (13), d denotes the dose rate. When $d = 0$, $F = a_0/b_0$ provides the endogenous mutation frequency.

We are now ready to fix the parameters of the WAM model to be used for various animals. Particularly the parameters have been fixed by the mega-mouse experiments by Wada et al [5].

$$\begin{aligned} a_0 &= 3.24 \times 10^{-08}\ 1/h\ , \\ a_1 &= 2.94 \times 10^{-05}\ 1/Gy\ , \\ b_0 &= 3.00 \times 10^{-03}\ 1/h\ , \\ b_1 &= 1.36 \times 10^{-01}\ 1/Gy\ . \end{aligned} \tag{14}$$

We use these parameters as the standard parameters of the WAM model. We see then that the mutation frequency at the steady state for small d increases linearly with the dose rate.

We show the comparison of the WAM model with the mega-mouse experimental results of Russell-Kelly in Fig. 7 [3]. The experimental mutation frequencies are shown by the red points with 95% error bars, and the WAM predictions are shown by the blue points. In this figure, the two straight lines introduced by Russell-Kelly are shown, which indicate that the mutation frequencies depend on the dose rate. The WAM model provides the dose rate dependence nicely.

We applied the WAM model with the standard parameters to the new experimental results at low dose rates of Gondo and his collaborators [1]. In order to compare the SNVs per generation with the prediction of the WAM model, we have to first change the unit for the

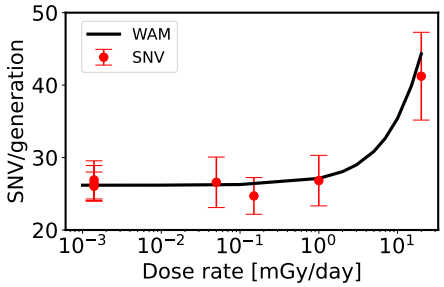

Figure 8. The SNVs per generation obtained in the low dose and long exposure mouse experiments by Gondo and his collaborators are shown by red dots with 95% confidence level error bars [1]. The WAM results are shown by the solid curve in the log plot.

mega-mouse experiment of 1 /locus/generation to 1 /bp/generation. We use for this the relation 1 locus = 1650 bp. We then multiply the number of base pairs 20×10^8 and two chromosomes for one base pair. We then get the results shown in Fig. 8 by the solid curve as compared with the experimental values of SNVs as a function of the dose rate. We find the agreement of the WAM predictions with the low dose rate experimental results is very good. The experimental feature is that the SNVs per generation stay almost constant until 1mGy/day is reproduced by the WAM model.

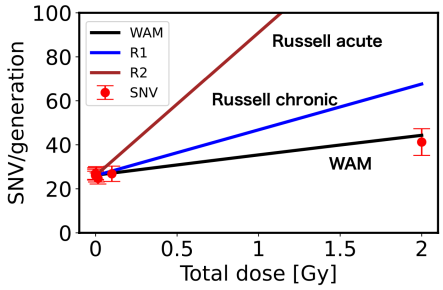

Figure 9. The SNVs per generation obtained in the low dose and long exposure mouse experiments by Gondo and his collaborators are shown by red dots with 95% confidence level error bars. The WAM results are shown by the solid curve in the linear plot. The two straight lines obtained by Russell-Keller are shown by the brown (acute) and blue (chronic) lines. The experimental value at the total dose of 2Gy is lower than these two lines.

We see a slight increase of the WAM results in the log plot in Fig. 8, and make further a linear plot in Fig. 9. The WAM results increase with the dose rate as shown by the black solid curve, which reproduces the SNV at 20mGy/day. We show here the two straight lines of Russell-Kelly by the brown and blue lines. The experimental value is much lower than the two lines. We see clearly the dose rate dependence of the SNVs per generation.

5 Conclusion

We made a statistical analysis of the SNVs for the control cases and the exposure cases. We extracted the average values and the 95% confidence levels of the SNVs using the maximum likelihood method. We found the SNVs per generation behave nicely for the control cases. We then extracted the SNVs for each generation using the family structure. The SNVs per generation were almost constant within the 95% confidence level errors until $d = 1$mGy/day. Since in one generation, the duration is about 100 days, the amount of total dose in one generation is 100mGy. Hence in terms of the total dose, the SNVs per generation were unchanged until 100mGy within the statistical errors.

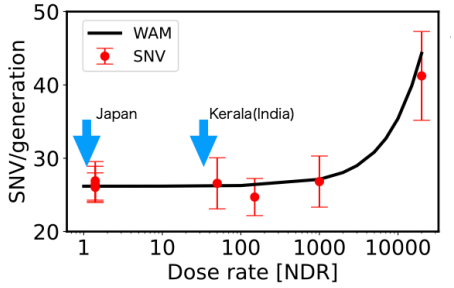

Figure 10. The SNVs per generation obtained in the low dose and long exposure mouse experiments by Gondo and his collaborators are shown by red dots with 95% confidence level error bars. The WAM results are shown by the solid curve. We propose to use the natural dose rate unit (NDR), which is used in the horizontal axis. The dose rate in Japan is 1 and that at Kerala in India is about 30.

We compared the WAM model with the experimental results taken at various dose rates. The comparison with the new data is almost perfect. It is interesting to point out that the natural dose rate in

Japan is about 0.001mGy/day and that of Kerala in India is about 0.03mGy/day. This is actually the range of dose rates taken in the Gondo experiment [1]. The WAM model should be the theory of the endogenous mutations in Kerala. It is also interesting to point out that dose rate in the cosmic environment is about 1mGy/day. The new experimental results show that the SNVs/generation (mutation rate) is almost unchanged from the endogenous SNVs.

Since the unit of radiation is difficult for society, we propose the use of the natural dose rate unit (NDR) for presentation. Namely 1NDR=0.001mGy/day=1μGy/day. We show the present results and the WAM predictions in the NDR unit in Fig. 10. This figure should be much easier to understand by the society.

References

[1] Y. Gondo et.al., 'High-precision detection of de novo mutations and application to risk assessment for the low dose-rate longtime exposure to radiation', ICRP Tokyo conference (2023).

[2] A. Uchimura et.al., 'Whole Genome Sequencing Analysis for Understanding Next Generation Effects of Radiation Exposure', ICRP Tokyo conference (2023).

[3] W.L. Russell and E.M. Kelly, 'Genetics Mutation frequencies in male mice and the estimation of genetic hazards of radiation in men', Proc. Natl. Acad. Sci. USA (1982) 542-544.

[4] M. Bando et al. 'Study of mutation from DNA to biological evolution. International Journal of Radiation Biology', Vol. 95, 014201 (2019).

[5] T. Wada et al., 'Dose and dose-rate dependence of mutation frequency under long-term exposure - a new look at DDREF from WAM model', Journal of Nuclear Science and Technology, 1824-1830 (2016).