

## Supplementary Models

### Model of spermatocyte cytokinesis failure frequency

The proportion of cells with 1, 2 or 4 nuclei after meiosis II can be calculated, assuming that the probability of cytokinesis failure, P is constant (Fig. S3)

$$\begin{aligned}
 1:1 & \quad 4(1-p)^3 + 4p(1-p)^2 \\
 & \quad = 4 - 8p + 4p^2 - 4p + 8p^2 - 4p^3 + 4p - 8p^2 + 4p^3 \\
 & \quad = 4(1-p)^2 \\
 2:1 & \quad 2p(1-p) + 2p(1-p)^2 + 2p^2(1-p) \\
 & \quad = 2p - 2p^2 + 2p - 4p^2 + 2p^2 \\
 & \quad = 4p(1-p) \\
 4:1 & \quad p^2
 \end{aligned}$$

The sum of these ratios is

$$\begin{aligned}
 & 4(1-p)^2 + 4p(1-p) + p^2 \\
 & = 4 - 8p + 4p^2 + 4p - 4p^2 + p^2 \\
 & = 4 - 4p + p^2 \\
 & = (2-p)^2
 \end{aligned}$$

Therefore the actual proportions of cells of types 1:1, 2:1, 4:1 are given respectively by a, b, and c in

$$\begin{aligned}
 (2-p)^2 a &= 4(1-p)^2 \quad (1) \\
 (2-p)^2 b &= 4p(1-p) \quad (2) \\
 (2-p)^2 c &= p^2 \quad (3)
 \end{aligned}$$

Where a is the proportion of cells with one nucleus, b is the proportion of cells with two nuclei and c is the proportion of cells with 4 nuclei.

Differentiating (1) with respect to p, using the notation a' for the derivative of a, we have

$$\begin{aligned}
 -2(2-p) a + (2-p)^2 a' &= -8(1-p) \\
 -8(1-p)^2 + (2-p)^3 a' &= -8(1-p)(2-p) \quad [\text{multiply by } (2-p) \text{ and substitute}] \\
 (2-p)^3 a' &= 8(1-p)^2 - 8(1-p)(2-p) \\
 &= 8(1-p)(1-p - 2+p) \\
 &= -8(1-p)
 \end{aligned}$$

Similarly from (2),

$$\begin{aligned}
 -2(2-p) b + (2-p)^2 b' &= 4(1-2p) \\
 -8p(1-p) + (2-p)^3 b' &= 4(1-2p)(2-p) \\
 (2-p)^3 b' &= 8p(1-p) + 4(1-2p)(2-p) \\
 &= 4(2p - 2p^2 + 2 - 5p + 2p^2) \\
 &= 4(2 - 3p)
 \end{aligned}$$

$$\begin{aligned}
 \text{and} \quad -2(2-p) c + (2-p)^2 c' &= 2p \\
 (2-p)^3 c' &= 2p^2 + 2p(2-p) \\
 &= 4p
 \end{aligned}$$

## Using experimental data to test the Cell Division Model

### 1. Least sum of squares

In order to test this model of spermatocyte cell division failure frequencies, it was necessary to test how well real data fits to the model. Let the observed proportions of cells containing 1, 2 or 4 nuclei be designated by  $A_i$ ,  $B_i$  and  $C_i$ . a, b and c are the theoretical values of these proportions which all correspond to a certain value of p in the model. If the theoretical values from the model fit the experimental data, then  $A_i$  is very close in value to a,  $B_i$  to b and  $C_i$  to c. To assess how good the model is, starting from experimental data, we need to find the value of p which minimises the squared differences between the expected (a, b, c) and observed ( $A_i$ ,  $B_i$ ,  $C_i$ ) frequencies for that value of p.

Use the size of the sum of squares to assess the validity of the model

The sum of squares of differences is

$$S = (A_i - a)^2 + (B_i - b)^2 + (C_i - c)^2$$

For least sum of squares this has a minimum value, and the derivative  $S' = 0$ . Hence differentiating with respect to p gives

$$\begin{aligned}
 2(A_i - a) A_i' + 2(B_i - b) B_i' + 2(C_i - c) C_i' &= 0 \\
 A_i A_i' + B_i B_i' + C_i C_i' &= a A_i' + b B_i' + c C_i'
 \end{aligned}$$

Multiplying by  $(2-p)^5$  and substituting for  $A_i, A_i'$  etc,

$$\begin{aligned} & -32(1-p)^3 + 16p(1-p)(2-3p) + 4p^3 \\ & = -8a(1-p)(2-p)^2 + 4b(2-3p)(2-p)^2 + 4cp(2-p)^2 \end{aligned}$$

Dividing by 4 and expanding terms

$$\begin{aligned} & -8 + 24p - 24p^2 + 8p^3 + 8p - 20p^2 + 12p^3 + p^3 \\ & = -2a(4 - 8p + 5p^2 - p^3) + b(8 - 20p + 14p^2 - 3p^3) + c(4p - 4p^2 + p^3) \\ & (-8 + 8a - 8b) + (32 - 16a + 20b - 4c)p + (-44 + 10a - 14b + 4c)p^2 + (21 - 2a + 3b - c)p^3 = 0 \\ & -8(1-a+b) + 4(8-4a+5b-c)p - 2(22-5a+7b-2c)p^2 + (21-2a+3b-c)p^3 = 0 \quad (4) \end{aligned}$$

The least sum of squares fit to given values  $a, b, c$  is therefore a value of  $p$ , which is a root of this cubic equation.

## 2. Existence of at least one solution for cubic equation (4) in the Cell Division Model

In the cubic function (4)

$$y = -8(1-a+b) + 4(8-4a+5b-c)p - 2(22-5a+7b-2c)p^2 + (21-2a+3b-c)p^3$$

we are expecting a root between 0 and 1 to give the value of  $p$  for the least sum of squares fit for the data  $a, b, c$  with the cell division model, where

$$a + b + c = 1, a > 0, b > 0, c > 0$$

First, we can show there is a root between 0 and 1, by showing  $y < 0$  when  $p=0$  and  $y > 0$  when  $p=1$ :

$$\begin{aligned} \text{At } p=0, \quad y &= -8(1-a+b) \\ & \leq 0 \end{aligned}$$

$$\text{since } 1-a+b \geq 1-a > 0.$$

$$\begin{aligned} \text{At } p=1, \quad y &= -8(1-a+b) + 4(8-4a+5b-c) - 2(22-5a+7b-2c) + (21-2a+3b-c) \\ & = 1 + b - c \\ & > 1 - c \\ & > 0 \end{aligned}$$

## 3. Calculating $p$

The cubic equation (4) was solved numerically, with accuracy 0.001. An EXCEL spread sheet including a script to calculate the failure frequency from the experimental frequencies of 1:1, 2:1 and 4:1 cells is available upon request (pdfoster@ntlworld.com).

Data on  $A_i, B_i$  and  $C_i$  were taken from this study and for several other mutations previously characterized to affect spermatocyte cytokinesis. Cells containing greater than four nuclei were excluded from the analysis, since the model only considers cells starting meiosis I with one nucleus. Rare cells containing three nuclei were grouped with those containing four nuclei to contribute to the value of  $C_i$ .

From the 30 genetic conditions analysed in this study, the maximum sum of squares was only 0.0938, and the average only 0.0146. The low values of the sum of squared deviations from the model (Fig. S3C) suggest that the model describes a wide range of mutant conditions well. This means that in these mutants, the probability of cytokinesis failure might be the same for cytokinesis I and II. This theory could be useful for measuring the magnitude of the effect on cytokinesis of different genetic conditions, as a probability of cytokinesis failure is a more relevant indicator of the requirement of a protein than the raw data  $A_i, B_i$  and  $C_i$ .