

## SUPPLEMENTARY NOTE

### Number of FSCs per germarium and founder FCs per egg chamber: statistical methods

#### 1. Best fit for number of FSCs per germarium based on the distribution of numbers of different FSC colors observed

For statistical modeling we assumed for simplicity that all ovarioles have the same number of surviving FSC lineages at the time of analysis (9d after clone induction for the first analysis). In ovarioles 9 days after clone marking we observed experimentally that each of the three types of RFP-negative FSC clones (B, G and BG) were present at a frequency of roughly 1/9 of the total, while the three types of RFP-positive FSC clones (BR, GR, BGR) were present at a frequency of roughly 2/9. We therefore assumed in our statistical modeling that the different colors of FSC clone were present in those same proportions (B:G:BG:BR:GR:BGR = 1:1:1:2:2:2). Those proportions are exactly what would be expected if each of the nine possible genotype was present at the same frequency. This outcome was very convenient for simple modeling. It represents a balance between too few recombination events (leading to an excess of *GFP/lacZ* and *RFP/+* genotypes) and too many successive recombination events (which lead to an excess of the genotypes stable to further recombination- *GFP/GFP*, *lacZ/lacZ*, *RFP/RFP* and *+/+*).

Assuming that the nine genotypes in the FSCs are independent and identically distributed (as found experimentally), the probability of observing exactly  $k$  different FSC colors when there are  $n$  surviving FSC lineages per germarium at the time of analysis is given by the formula below.

$$\sum_{c_1=(k-3)\vee 0}^{3\wedge k} \sum_{m=c_1}^{n-(k-c_1)} \binom{n}{m} \left(\frac{1}{3}\right)^m \left(\frac{2}{3}\right)^{n-m} f(c_1, m) f(k - c_1, n - m)$$

where

$$\begin{aligned} f(0, 0) &= 1 \\ f(z, 0) &= 0 \quad \text{where } z \text{ is a non-zero integer} \\ f(1, r) &= 3(1/3)^r \\ f(2, r) &= 3((2/3)^r - 2(1/3)^r) \\ f(3, r) &= 1 - f(1, r) - f(2, r) \end{aligned}$$

The expected percentages of ovarioles with 1-6 distinct FSC colors calculated from this formula are presented below for different numbers of surviving FSC lineages (n= 4, 5, 6 etc.). The calculated distributions that most closely match observed distributions (from 50 ovarioles expressed as percentages) at 9d are shown below for (a) FSCs with matching FCs, (b) all FSCs (with or without matching FCs) and (c) the number of FC-producing FSC colors present at 5d (equal to the number of distinguishable FC patches at day 9 (Supplementary Fig. S2c)).

*(a) Best fit of predicted color distributions to experimental data for FSC clones with matching FCs at 9d.*

| k                  | 1 | 2  | 3  | 4  | 5  | 6 | $\Sigma(\text{Observed-Predicted})^2$ |
|--------------------|---|----|----|----|----|---|---------------------------------------|
| FSC+FC 9d observed | 2 | 12 | 44 | 30 | 8  | 4 |                                       |
| n=4 prediction     | 1 | 20 | 55 | 24 | 0  | 0 | 302                                   |
| n=5 prediction     | 0 | 8  | 42 | 42 | 7  | 0 | 185                                   |
| n=6 prediction     | 0 | 3  | 28 | 48 | 19 | 1 | 795                                   |

Best match is for n=5 FSCs (lowest value for sum of differences squared).

The theoretical distribution above for 5 FSCs is shown in Supplementary Fig. 2a, next to “FSC clones” data.

*(b) Best fit of predicted color distributions to experimental data for FSC clones (with or without matching FCs) at 9d.*

| k               | 1 | 2  | 3  | 4  | 5  | 6  | $\Sigma(\text{Observed-Predicted})^2$ |
|-----------------|---|----|----|----|----|----|---------------------------------------|
| FSC 9d observed | 0 | 6  | 34 | 36 | 14 | 10 |                                       |
| n=4 prediction  | 1 | 20 | 55 | 24 | 0  | 0  | 1078                                  |
| n=5 prediction  | 0 | 8  | 42 | 42 | 7  | 0  | 253                                   |
| n=6 prediction  | 0 | 3  | 28 | 48 | 19 | 1  | 295                                   |
| n=7 prediction  | 0 | 1  | 18 | 47 | 30 | 4  | 694                                   |

Best match for n= 5-6 FSCs.

The theoretical distribution above for 6 FSCs is shown in Supplementary Fig. 2b, next to “FSCs” data.

*(c) Best fit of predicted color distributions to experimental data for FSC clones with matching FCs at 5d.*

| k                  | 1 | 2 | 3  | 4  | 5  | 6  | $\Sigma(\text{Observed-Predicted})^2$ |
|--------------------|---|---|----|----|----|----|---------------------------------------|
| FSC+FC 5d observed | 0 | 6 | 18 | 20 | 32 | 24 |                                       |
| n=8 prediction     | 0 | 1 | 11 | 41 | 39 | 8  | 820                                   |
| n=9 prediction     | 0 | 0 | 7  | 34 | 45 | 14 | 622                                   |
| n=10 prediction    | 0 | 0 | 4  | 28 | 48 | 20 | 568                                   |
| n=11 prediction    | 0 | 0 | 3  | 22 | 49 | 33 | 635                                   |

Best match for n=10 FSCs.

The theoretical distribution above for 10 FSCs is shown in Supplementary Fig. 2c, next to “FC” (= FSC/FC clones at 5d) clone data.

Similar modeling for the observed distribution of distinguishable FC-producing FSC lineages at days 14 (n= 53 ovarioles), 20 (n=47 ovarioles) and 30 (n=60 ovarioles) is shown below. The average number of distinguishable lineages from day 5 to day 30 is plotted in Fig. 2j and the inferred number of surviving FSC lineages is presented in Fig. 2k.

*(d) Best fit of predicted color distributions to experimental data for FSC clones with matching FCs at 14d.*

| k                   | 1   | 2  | 3  | 4  | 5 | 6 | $\Sigma(\text{Observed-Predicted})^2$ |
|---------------------|-----|----|----|----|---|---|---------------------------------------|
| FSC+FC 14d observed | 17  | 66 | 13 | 2  | 2 | 0 |                                       |
| n=1 prediction      | 100 | 0  | 0  | 0  | 0 | 0 | 11422                                 |
| n=2 prediction      | 19  | 81 | 0  | 0  | 0 | 0 | 406                                   |
| n=3 prediction      | 4   | 44 | 52 | 0  | 0 | 0 | 2182                                  |
| n=4 prediction      | 1   | 20 | 55 | 24 | 0 | 0 | 4624                                  |

Best match for 2 FSCs.

*(e) Best fit of predicted color distributions to experimental data for FSC clones with matching FCs at 20d.*

| k                   | 1   | 2  | 3  | 4 | 5 | 6 | $\Sigma(\text{Observed-Predicted})^2$ |
|---------------------|-----|----|----|---|---|---|---------------------------------------|
| FSC+FC 20d observed | 30  | 53 | 15 | 2 | 0 | 0 |                                       |
| n=1 prediction      | 100 | 0  | 0  | 0 | 0 | 0 | 7938                                  |
| n=2 prediction      | 19  | 81 | 0  | 0 | 0 | 0 | 1079                                  |
| n=3 prediction      | 4   | 44 | 52 | 0 | 0 | 0 | 2126                                  |

Best match for 2 FSCs.

*(f) Best fit of predicted color distributions to experimental data for FSC clones with matching FCs at 30d.*

| k                   | 1   | 2  | 3  | 4 | 5 | 6 | $\Sigma(\text{Observed-Predicted})^2$ |
|---------------------|-----|----|----|---|---|---|---------------------------------------|
| FSC+FC 30d observed | 56  | 39 | 5  | 0 | 0 | 0 |                                       |
| n=1 prediction      | 100 | 0  | 0  | 0 | 0 | 0 | 3482                                  |
| n=2 prediction      | 19  | 81 | 0  | 0 | 0 | 0 | 3158                                  |
| n=3 prediction      | 4   | 44 | 52 | 0 | 0 | 0 | 4938                                  |

Best match for 1-2 FSCs (plotted as 1.5 in Fig. 2k)

## **2. Number of FSCs per germarium deduced from the percentage contribution from each FSC lineage color.**

The number of FC-producing FSCs can also be calculated as the reciprocal of the FC population contributed by a single FSC. We calculated the contribution of a single FSC lineage color to the FC population. This number underestimates the true number of FSCs because each color may include more than one FSC.

To determine the contribution of a single FSC lineage color to the entire FC epithelium of an ovariole from the 9d multicolor data we considered only G, B and BG lineages because these are present at an average frequency of 1/9 rather than 2/9 and are therefore more often initiated by single FSCs (Supplementary Fig. 1b,c). We also considered only ovarioles with at least four egg chambers and lineages that included a surviving FSC and at least one matching FC patch, even if present only in the germarium. We estimated the proportion of each egg chamber occupied by the color under examination, scoring all egg chambers in the ovariole (including those with no matching FCs). The average contribution of a single FSC color was close to 1/8 of all FCs (12.0%; n=30, SEM= 2.6%). An analogous analysis of 9-day FSC clones generated by the MARCM method showed an average contribution close to 1/7 of all FCs (15.3%; n=63, SEM=1.8%).

These data would lead to an estimate of 7-8 FSCs if each lineage color contained only a single FSC during the period of measurement (5-9d). However, even if each colored clone were initiated from a single FSC immediately after recombination marking, several FSCs would later amplify. The average amplification of surviving FSCs would be the reciprocal of the proportion of initially marked FSC lineages that survive. Hence, the number of FSCs estimated to be present during 5-9d by calculating the contributions of individual colors is equivalent to an estimate of the number of surviving FSCs during the period from 5-9d. The value of 7-8 surviving FSC lineages fits well with the time-course of surviving FSC lineages measured by counting the number of distinct colors present at different time points (Fig. 2k).

### 3. Number of founder FCs per egg chamber deduced from percentage contribution from each FC color.

The number of founder FCs per egg chamber can be calculated as the reciprocal of the average contribution of a single FC color, provided that a single color represents a single founder FC. To minimize the frequency of two or more founder FCs sharing the same color we only considered ovarioles with B, G and BG FSC clones. There were 83 patches containing B, G or BG FCs, with an average contribution of 30.4% per patch to the FC epithelium of a single egg chamber.

To derive the contribution per founding FC we must also estimate how many of the scored FC patches derived from one, two, three or more founder FCs of the same color. To do this we assumed that each founder FC is drawn independently from a group of source cells and hence that the number of founder cells of a given color will show a binomial distribution. To determine the probability,  $p$ , that a founder FC of a specific color (B, G or BG) is drawn we counted the number of egg chambers with zero FCs of the specified color (among ovarioles with an FSC clone of the specified color). For B, G and BG lineages combined we found that 102 of 185 egg chambers had no FC contribution from the interrogated genotype. Hence,  $(1-p)^k = 102/185$  where  $k$  is the number of founder FCs per egg chamber.

We then performed separate calculations of predicted binomial distributions of the number of founder FCs of the same color per egg chamber for different values of  $k$ . We present the results below only for  $k = 3, 4$  and  $5$  because these emerge as the only plausible candidates. From the  $p$  value inferred (from the experimental determination that  $(1-p)^k = 102/185$ ) we calculated the fraction of patches predicted to be founded by  $i$  ( $= 1, 2, 3, 4$  or more) founder FCs per egg chamber as  $f(i) = p^i (1-p)^{k-i} / i!(k-i)!$  and hence the inferred average number of founding FCs per patch of a given color as  $(f(1)+2f(2)+3f(3)+4f(4))/(f(1)+f(2)+f(3)+f(4))$ . Occupancy per founding FC was then calculated as occupancy per patch divided by inferred number of founding FCs per patch. The reciprocal of this value was our best estimate of the number of founding FCs per egg chamber. Even though different values of  $k$  produced small differences in the inferred number of founding FCs per patch the best estimate for the number of founding FCs per egg chamber is 4 whether we initially use values of  $k$  anywhere from 3 to 5.

*(a) Number of founder FCs per patch and hence founder FCs per egg chamber based on binomial distribution: multicolor clones*

|     | p     | Probability that an FC patch is founded by the indicated number of FCs (i) of the same color |       |       |       |   | Inferred average number of founder FCs per patch | Experimental occupancy per patch | Inferred occupancy per founder FC | Inferred # founder FCs |
|-----|-------|--|-------|-------|-------|---|--|----------------------------------|-----------------------------------|------------------------|
|     |       | 1  | 2     | 3     | 4     | 5 |  |                                  |                                   |                        |
| k=3 | 0.180 | 0.363  | 0.080 | 0.006 | 0     | 0 | 1.20   | 30.4%                            | 25.2%                             | 4.0                    |
| k=4 | 0.138 | 0.354  | 0.085 | 0.009 | 0     | 0 | 1.23   | 30.4%                            | 24.8%                             | 4.0                    |
| k=5 | 0.112 | 0.348  | 0.088 | 0.011 | 0.001 | 0 | 1.25   | 30.4%                            | 24.3%                             | 4.1                    |

A MARCM FSC clonal analysis was scored and analyzed in exactly analogous fashion. Average occupancy was 31.2% over 179 patches and 158/337 egg chambers had zero FC contribution, scoring only ovarioles that included marked FSC clones.

*(b) Number of founder FCs per patch and hence founder FCs per egg chamber based on binomial distribution: MARCM clones*

|     | p     | Probability that an FC patch is founded by the indicated number of FCs (i) of the same color |       |       |       |   | Inferred average number of founder FCs per patch | Experimental occupancy per patch | Inferred occupancy per founder FC | Inferred # founder FCs |
|-----|-------|--|-------|-------|-------|---|--|----------------------------------|-----------------------------------|------------------------|
|     |       | 1  | 2     | 3     | 4     | 5 |  |                                  |                                   |                        |
| k=3 | 0.223 | 0.404  | 0.116 | 0.011 | 0     | 0 | 1.26   | 31.2%                            | 24.8%                             | 4.0                    |
| k=4 | 0.173 | 0.391  | 0.123 | 0.017 | 0.001 | 0 | 1.30   | 31.2%                            | 23.9%                             | 4.2                    |
| k=5 | 0.141 | 0.384  | 0.126 | 0.021 | 0.002 | 0 | 1.33   | 31.2%                            | 23.4%                             | 4.3                    |

For the MARCM data the percentage occupancy per founding cell is close to 25% (for all values of k, including k=4), giving a best estimate of an average of four founding FCs per egg chamber, just as for the multicolor FC patches. It is clear, however, that there can sometimes be at least five founding FCs (Fig. 1k). Also, a variety of departures from our simplifying statistical assumptions, such as unequal numbers of FSCs of different colours and heterogeneous potential for immediate FC production (both of which we observe), would lead to under-estimating how frequently a single FC colour is founded by a single FC, and would therefore raise our estimate of the number of founding FCs to a value higher than four.