Chapter 4

The rate of loss of heterozygosity in *Blm*-deficient ES cells

4.1 Introduction

The rate of loss of heterozygosity (LOH) in *Blm*-deficient cells has been calculated previously as 4.2×10^{-4} and 2.3×10^{-4} events/locus/cell/generation respectively in the two *Blm*-deficient ES cell lines generated (Guo *et al.*, 2004; Yusa *et al.*, 2004). These measurements are based on a single locus in each case, *Gdf9* and *Fasl* respectively. The model of LOH by crossover after mitotic recombination predicts that LOH rate should vary by position on the chromosome. As LOH occurs at all loci distal to the point of crossover, loci located closer to the telomere should have an increased chance of a crossover occurring at a proximal position and thus an increased rate of LOH. For loci very close to the centromere most mitotic recombination events, if randomly distributed, will occur distally and not affect the centromeric locus.

If the rate of LOH does vary significantly across the genome, the effective coverage of the libraries will be affected. The chance of recovering homozygous mutations in genes close to centromeres may be reduced, and genes close to telomeres increased. I decided to investigate this by determining the rate of LOH at several different chromosomal positions. Working on the assumption that mitotic chiasmata and crossovers are distributed randomly, I chose three loci along the length of chromosome 11 to investigate, including the previously measured *Gdf9* locus.

LOH rates in this context are typically measured by inserting a selectable marker at the locus to be tested. For the *Gdf9*, a *HPRT* minigene was used and LOH assessed by its loss, which produces a 6-thioguanine-resistant cell (Luo *et al.*, 2000). For *Fasl*, a mutant *neo** gene was used, and high G418 selection used to select homozygous *neo/neo* cells (Yusa *et al.*, 2004). The rates measured were similar. For the *neo** selection there was a high background of surviving *neo/+* cells that had to be corrected for by genotyping resistant cells. Negative selection may also have background, for example if the spontaneous mutation rate is high, and thus this method works on the assumption that mitotic recombination and crossover is the primary mechanism in *Blm* cells.

4.1.1 Using fluctuation analysis to measure the rate of rare events in cell culture

LOH is a rare event, so to measure it a large number of cells need to be analysed. This presents a problem as LOH can occur during the expansion of the cells to a sufficient number. This results in a large variance in the number of cells that have undergone LOH observed in the culture. It is impossible to say from a single culture whether the number of resistant cells resulted from a single early LOH event, giving rise to a cell that expanded clonally over the remaining generations, or from multiple later events. If multiple cultures are set up, each beginning from a single cell, the number of resistant cells after a set time will fluctuate between cultures, due to the the disproportionate effect of early events on the final number of resistant cells. In a seminal paper, Luria and Delbrück developed a formula to explain this fluctuation in the situation of spontaneous mutation to phage resistance in bacteria (Luria and Delbrück, 1943). LOH is analogous to a spontaneous mutation, and the same formulae and method can be used to calculate the rate.

Luria and Delbrück derived two equations that can be used to calculate the mutation rate. Both result from methods to deal with the large number of resistant cells obtained when a mutation occurs very early in the culture. The first, known as the p_0 method, simply ignores all cultures in which mutations occur and instead considers the number of cultures without mutants. The total mutations in the experiment distribute across all cultures according to a Poisson distribution, therefore the probability that no mutants occur in a culture can be calculated for a given mutation rate. Conversely, using the observed fraction of cultures that show no mutations, a mutation rate can be calculated. The p_0 method does not make efficient use of the information gathered, but is at least straightforward. Its main drawback is that a very large number of cul-

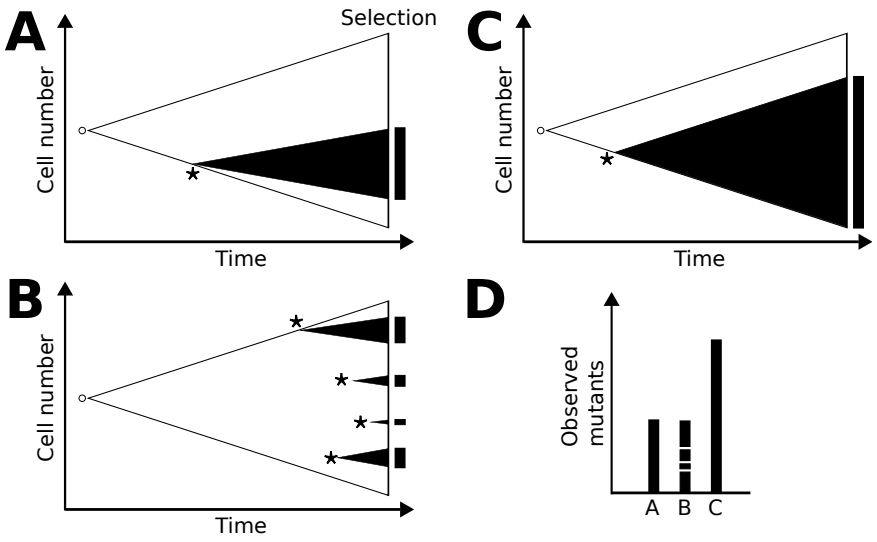

Figure 4.1: Measuring rare events in cell culture. Cultures are depicted as expansions of a single cell. Mutations (or LOH events) that arise, indicated by asterisks (*) continue to expand clonally. The number of mutant cells at the end of the culture period could result from one early event (A) or multiple later events (B). An extreme example of the effect of a very early mutation is shown in C. The figure is schematic, note that the Y axis should be a log scale if cells are growing exponentially—thus early mutations have a large effect on the final number of observed mutants (D).

tures is required to calculate a rate with any accuracy, and it relies on plating the entire culture to ensure all cells are interrogated. Both of these conditions are difficult to achieve using mammalian cell cultures.

The second method originally presented is the method of means. This uses the mean of the resistant cells to calculate a mutation rate. The problem of the long tail of the distribution is dealt with by assuming that none of the experimental cultures, which represent a small sample of the distribution of all possible mutant frequencies, are in the extreme tail. This is done in the derivation of the formula by only considering cultures that were mutation-free after a certain critical time. This allows a mutation rate to be calculated, but it is likely to be an overestimate. However, this can be used in situations where all cultures show mutations, and also in cases where only a proportion of the culture is plated.

Although the methods described in the original fluctuation analysis paper continue to be used today, several adaptations have been published for mammalian cells. As noted above, some assumptions that are acceptable for bacteria are not for mammalian cells. This applies especially to the assumption that the entire culture is plated—ES cell cultures typically have a plating efficiency of only 30–50%.

Jones *et al.* extended the principle of the p_0 method to provide an estimator of the mutation rate using the median number of resistant cells per culture. This also allows for plating of only a portion of the culture, thus plating efficiency can be incorporated. Moreover, they show that optimising the dilution such that roughly half the cultures have no mutants allows the rate to be calculated accurately with relatively few cultures (Jones *et al.*, 1994).

4.2 Results

4.2.1 Choice of loci

Two of the cell lines I used were generated as part of the TNP100 library (see Chapter 5). These are named by their well positions, D8 and F8, and both have TNP (i.e. $pu\Delta TK$-expressing) transposon integrations on chromosome 11. The co-ordinates of the insertion sites are 11:20,780,891 and 11:95,552,974 respectively (NCBI m37). I also used a cell line with $pu\Delta TK$ integrated by gene targeting at the *Gdf9* locus, which was generated by Amy Li (Li, 2010). As this locus was originally used to measure the rate of LOH in *Blm*-deficient cells (Luo *et al.*, 2000), using this cell line will allow my results to be compared directly with this rate. *Gdf9* also has the advantage of mapping almost exactly in the middle of D8 and

F8 (54 Mb from the centromere), providing a good test of whether or not LOH rate varies with distance from the centromere.

4.2.2 Calculation of mutation rate

I trypsinised cultures of these cells for at least 15 minutes and dispersed them to a single cell suspension by pipetting. I then plated 1,000 cells per 90 mm plate to obtain colonies. Each colony is a culture started from a single cell. I picked 24 colonies from each cell line after 10 days, and expanded them to a 24-well plate (via one passage on a 96-well plate). The average cell count at this stage was 751,571. Cultures with large differences from this value were discarded at this point, as the mutation rate calculation assumes that all cultures were equally expanded. One tenth of each culture was plated directly in FIAU selective medium, and the remainder diluted for counting and plating at low density (150 cells per plate) in non-selective medium to calculate the plating efficiency.

I calculated the average number of mutations per culture as follows, using the $\hat{m}_h$ median estimator derived by Jones *et al.*. For each series of cultures, I calculated the median number of FIAU resistant colonies r_m, and the mean cloning efficiency. The cloning efficiency was multiplied by the plated fraction (0.1) to obtain the effective plating p_e. The average number of mutations per culture is then given by equation (5) in Jones *et al.* (1994):

$$\hat{m}_h = \frac{r_m/p_e - ln(2)}{ln(r_m/p_e) - ln(ln(2))} \qquad (4.1)$$

The calculated mutation rates are shown in Table 4.1. The rate does appear to increase with distance from the centromere. However, the rates calculated are generally lower than those previously determined, as can be shown by comparing my rate for *Gdf9*, 2.5×10^{-5} events/cell/generation with that calculated by Luo *et al.*, 4.2×10^{-4}. Possible reasons for this are discussed below.

4.3 Discussion

4.3.1 Comparison with previously calculated rates

The rates of LOH that I calculated here are much lower than those previously determined. There are several possible reasons. First is that the different method employed here may be underestimating the number of mutations per culture. However the median estimator method gives similar results to the original formulae on other datasets, so should be applicable (Jones *et al.*, 1994). In any case, the magnitude of the difference is probably too large to be explained by features specific to one estimator. As a precaution, I did directly genotype the cell lines used in the experiment to ensure they were *Blm* mutants (not shown).

As I included the originally-measured *Gdf9* locus in my experiments, the discrepancy cannot be due to a locus-specific effect. A more likely reason is the difference in selection used between my experiments here and the previous rate calculations. I made use of the ΔTK gene for negative selection, whereas the previously reported calculations used *HPRT* or *neo** as described above. A possible mechanism by which this could affect the number of mutants recovered per culture is if the *puro*ΔTK mRNA or protein is more stable than *HPRT*, and therefore persists for longer after LOH occurs and removes the DNA. It is likely that resistance genes are expressed at high levels, as this has been artificially selected for in the choice of promoters and polyadenylation sites used in cloning vectors. If the cells are still functionally TK^+ for one or two generations after LOH, this will affect the numbers of FIAU-resistant colonies that can be obtained. Thus, using FIAU selection could result in a systematic underestimation of the mutation rate. Measuring LOH in wild-type cells using *puro*ΔTK would show whether this is the case. The off-rate of *HPRT* and *puro*ΔTK could be tested experimentally to investigate this further using, for example, Cre mediated deletion and measurement of the time taken to recover the maximum number of deleted clones. However, this is essentially the same experiment as the *Gdf9* comparison carried out here. Experiments presented in Chapter 5, also suggest that the actual rate of LOH (or at least copy number increase) is higher than that calculated here, further arguing for an effect of FIAU negative selection.

4.3.2 Implications for library coverage

The rate calculated for the most proximal locus in this analysis (D8) is about one third of the *Gdf9* rate. How relevant is this difference, over a distance of 33 Mb, with respect to library coverage? One way this can be interpreted is by considering how representative these loci are of all the genes in the genome. Plotting the positions of all genes in the genome reveals that *Gdf9* represents approximately the 30th percentile and the D8 locus approximately the 8th (Figure 4.3). Therefore, the current proto-

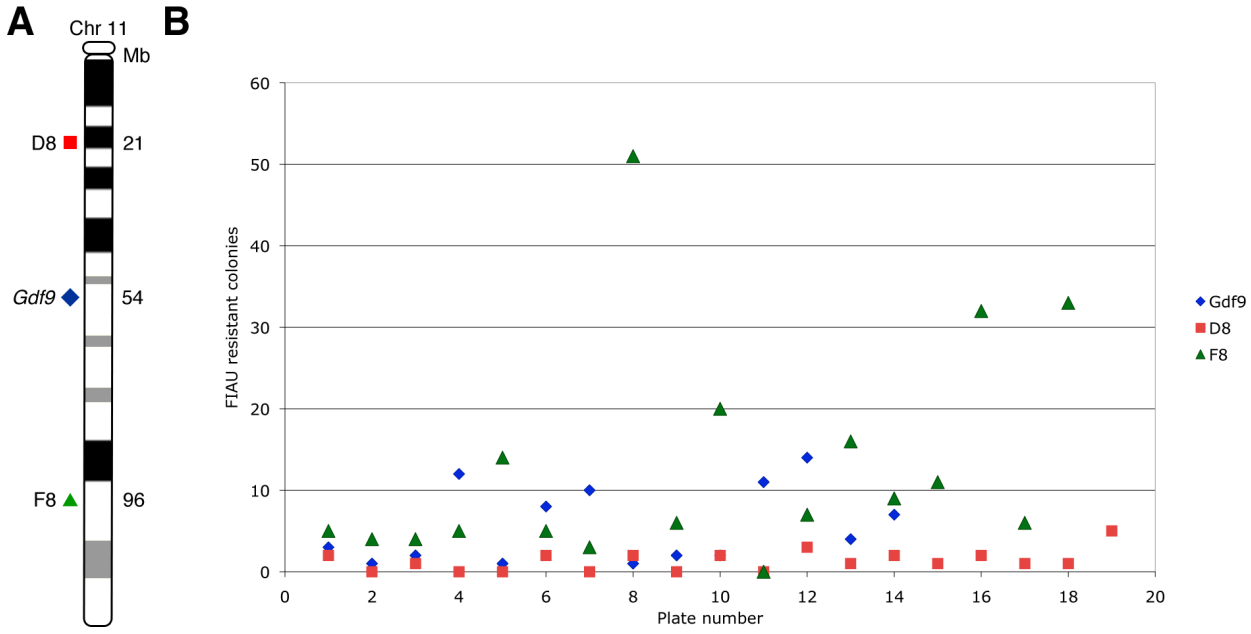

Figure 4.2: Number of LOH events observed for three loci on chromosome 11. A—Loci studied. B—Number of FIAU-resistant colonies obtained in replicate cultures for each locus. One-tenth of cultures expanded to a confluent 24-well plate was selected in each case. The ordering on the x axis is random.

Locus	Mb	Cultures	Median $FIAU^R$	Mean cloning	Mean cells/culture	$\hat{m}_h$: LOH events/culture	LOH rate
D8	21	19	1	0.35	752,914	7.53	9.9×10^{-6}
Gdf9	54	14	3.5	0.38	762,000	18.56	2.5×10^{-5}
F8	96	18	6.5	0.27	739,800	40.95	5.5×10^{-5}

Table 4.1: Calculation of LOH rate. LOH rate is events/cell/generation.

cols for library construction that are based on data for *Gdf9* should be sufficient for 70% of genes. However, the rate for the D8 locus should apply to 92% of all genes. As the rate is not drastically lower in practical terms, it should be possible to isolate LOH events at such loci with only slightly longer expansion times. These data provide a better guide for library construction, and support the hypothesis that the number of opportunities for initiation of proximal homologous recombination determines the probability of LOH at a locus.

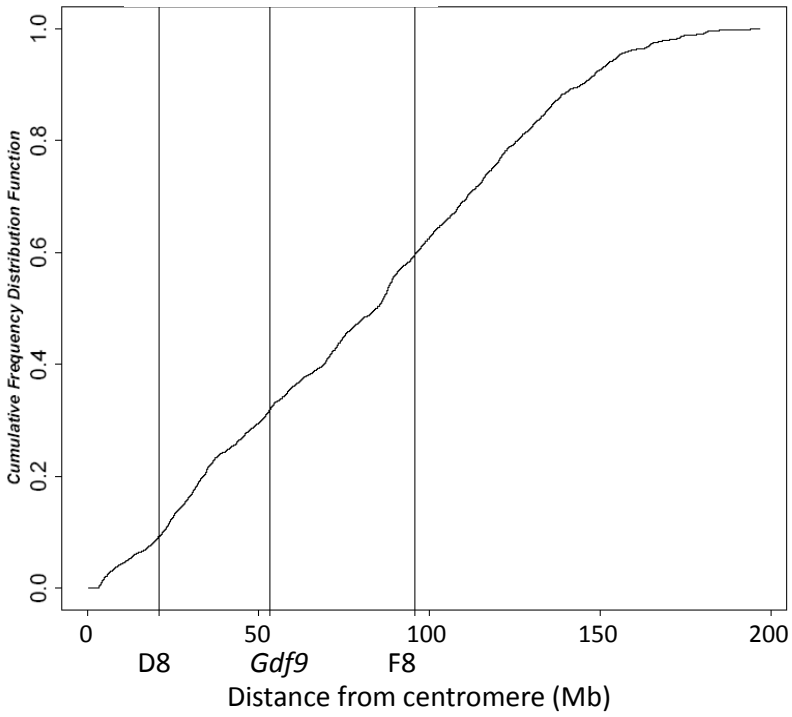

Figure 4.3: Plot of distance from centromere for all Vega curated mouse genes (Wilming *et al.*, 2008). The cumulative frequency of genes with their start (5′ end) at or before the value on the x axis is plotted. The positions of the three loci for which LOH rate was calculated are shown by vertical lines.