Normalisations for MEGA dPCR

To quantify the absolute frequency of these aberrations within the total population of alleles, copies derived from the FAM and HEX signals of the mutation assays were normalised (Norm 1) to the average number of reference assay copies amplified from two chromosomes unrelated to the edited one (Fig. 1b; Eq. 1). This initial normalisation to the reference was performed on unedited and edited cells.

$$\text{Norm 1} = \frac{\text{Assay copies per µl}}{\text{Reference copies per µl}}$$

Equation (1)

A second normalisation (Norm 2) was then performed on the relative value coming from the control unedited sample (Fig. 1b; Eq. 2).

$$\text{Norm 2} = \frac{\text{Edited Norm 1 assay}}{\text{Unedited Norm 1 assay}}$$

Equation (2)

The percentage of linked sequences was calculated at the on-target site with the Flanking assay and also with the non-targeted chromosome via the *Reference* assay. First, the probability of two unlinked sequences co-segregating into the same droplet by chance (CD) and the sum of linked and unlinked 5' and 3' sequences (λ5' and λ3', respectively) were calculated.

We calculated the chance of co-segregating sequences by multiplying the number of FAM and HEX single droplets together and dividing by the total number of negative droplets (Eq. 3).

$$C_D = \frac{S_F \times S_H}{N_D}$$

Equation (3)

The sum of the linked and unlinked 5' sequences was calculated by subtracting the natural log of the FAM negative droplets from the natural log of the accepted droplets (Eq. 4).

$$\lambda 5' = ln(\mathrm{A_D}) - ln(\mathrm{N_F})$$

Equation (4)

This was repeated for the 3' linked and unlinked sequences (Eq. 5).

$$\lambda 3' = ln(\mathrm{A_D}) - ln(\mathrm{N_H})$$

Equation (5)

Linkage was calculated by subtracting the natural log of the sum of the FAM single droplets, HEX single droplets, negative droplets, and chance co-segregated sequences from the natural log of the accepted droplets, multiplied by two, dividing this by the sum of the 5' and 3' linked and unlinked sequences, and finally multiplying by 100 (Eq. 6).

$$\%\ \mathrm{Linkage} = \frac{2\ (ln(\mathrm{A_D}) - ln(\mathrm{S_F} + \mathrm{S_H} + \mathrm{N_D} + \mathrm{C_D}))}{\lambda 5' + \lambda 3'} \times 100$$

Equation (6)

Where;

C_D = Chance co-segregation

S_F = FAM single

S_H = HEX single

N_D = Negative droplets

A_D = Accepted droplets

N_F = FAM negative droplets

N_H = HEX negative droplets

λ5' = sum of 5' linked and unlinked sequences

λ3' = sum of 3' linked and unlinked sequences

The linkage of the on-target flanking assay was normalised (Norm 3) to the linkage of the references for both the unedited and edited cells (Fig. 1b; Eq. 7).

$$\mathrm{Norm\ 3} = \frac{\%\ \mathrm{Flanking\ linkage}}{\%\ \mathrm{Reference\ linkage}}$$

Equation (7)

The reference normalised linkage of the edited cells was then normalised (Norm 4) to the reference normalised linkage of the unedited cells (Fig. 1b; Eq. 8).

$$\text{Norm 4} = \frac{\text{Edited Norm 3 linkage}}{\text{Unedited Norm 3 linkage}}$$

Equation (8)

When using a donor template where the length of the homology arms is excessively greater than the reference assay, a size-matched reference should be deployed to account for quantification bias. The number of integration copies should be normalised (Norm 5) to the number of size-matched reference copies (Fig. 1b; Eq. 9).

$$\text{Norm 5} = \frac{\text{Targeted integration copies per µl}}{\text{Size-matched reference copies per µl}}$$

Equation (9)

As no integration occurs in unedited cells, there is no requirement to normalise to this control.

Allele frequencies

To calculate the percentage of wildtype alleles, the Norm 2 Edge assay FAM (cleavage) value was multiplied by 100 (Fig. 1b; Eq. 10).

$$\%\ \text{Wildtype} = \text{Norm 2 Edge FAM} \times 100$$

Equation (10)

The Indel percentage was calculated by subtracting the Norm 2 Edge assay HEX (distal) value from the Norm 2 Edge assay FAM value and multiplying by 100 (Fig. 1b; Eq. 11).

$$\%\ \text{Indel} = (\text{Norm 2 Edge FAM} - \text{Norm 2 Edge HEX}) \times 100$$

Equation (11)

We then calculated the total non-indel aberrations by subtracting 1 from the Norm 2 Edge assay HEX value and multiplying by 100 (Fig. 1b; Eq. 12).

$$\%\ \text{Total non-indel aberrations} = (\text{Norm 2 Edge HEX} - 1) \times 100$$

Equation (12)

We calculated the percentage of linkage loss by subtracting 1 from the Norm 4 linkage value and multiplying by 100 (Fig. 1b; Eq. 13).

$$\%\ \text{Linkage loss} = (\text{Norm } 4\ - 1) \times 100$$

Equation (13)

No loss of linkage would be observed in the case of extended bidirectional end resection; therefore, the difference between the non-indel aberrations detected by the edge assay and the linkage loss quantified from the flanking assay was calculated as a large deletion (Fig. 1b; Eq. 14).

$$\%\ \text{Large deletions} = \%\ \text{Non-indel aberrations} - \%\ \text{Linkage loss}$$

Equation (14)

The percentage of linkage loss was then defined as “other aberrations,” i.e., double-strand breaks, translocations, inversions, chromothripsis, etc. (Fig. 1b; Eq. 15).

$$\%\ \text{Other aberrations} = \%\ \text{Linkage loss}$$

Equation (15)

We calculate the integration frequency for autosomes and X-linked genes in XX cells by multiplying the Norm 5 targeted integration value by 100 (Eq. 16).

$$\%\ \text{Targeted integration (Autosomes)} = \text{Norm } 5\ \times 100$$

Equation (16)

Whereas X- or Y-linked genes in XY cells, integration frequency is calculated by first multiplying the Norm 5 targeted integration value by 2 to account for half the alleles that can be integrated, then again by 100 (Fig. 1b; Eq. 17).

$$\%\ \text{Targeted integration (sex-linked XY)} = \text{Norm } 5\ \times 2\ \times 100$$

Equation (17)

The flanking assay is unable to differentiate endogenous sequences and donor templates with long homology arms and, thus, in this circumstance, it is incapable of calculating large

deletions. We calculate the resulting aberrations by subtracting the targeted integration from the total non-indel aberrations (Fig. 1b; Eq. 18).

$$\%\ \text{Large deletions}|\text{Other aberrations} = \%\ \text{Total non-indel aberrations} - \%\ \text{Targeted integration}$$

Modelling CDE equations

$$WT\ rate = dWT/dt = (k_{pr} \times DSB) - (k_{dsb} \times D(t) \times WT)$$

$$DSB\ rate = dDSB/dt = (k_{dsb} \times D(t) \times WT) - (k_{pr} + k_{in} + k_{ld} + k_{ti}) \times DSB$$

$$Precise\ repair\ rate = dPR/dt = (k_{pr} \times DSB) - (k_{dsb} \times D(t) \times PR)$$

$$Indel\ rate = dIndel/dt = (k_{in} \times DSB)$$

$$Large\ deletions\ rate = dLD/dt = (k_{ld} \times DSB)$$

$$Targeted\ integration\ rate = dTI/dt = (k_{ti} \times DSB)$$

$$D(t) = 1 - 2^{-(t/delay)}$$

Equation (18)

Where;

d[allele]/dt = difference in allele frequency/difference in time (rate)

k_{dsb}, k_{pr}, k_{in}, k_{ld}, k_{ti}, = rate coefficients of DSBs, precise repair, indels, large deletions, and targeted integration, per hour respectively.

D(t) = Cas9 nuclear trafficking delay based on the time in hours (t).