Additional data file 6 - Correction of an unbalanced chromosome

This step occurs when the common breakpoints of two nodes do not define a viable copy number profile for a chromosome, due to an imbalance between ‘up’ and ‘down’ breakpoints: \( \delta = \sum b^t_k \neq 0 \).

Let us first consider the case in which \( |\delta| = 1 \). To correct the imbalance, TuMult will either remove a breakpoint of \( \text{sign}(\delta) \) from the common breakpoint vector, or add a breakpoint of \( \text{sign}(-\delta) \) chosen from the breakpoints specific to one of the two samples. TuMult will choose the breakpoint to be modified on the basis of two criteria:

(i) First, modifying the wrong breakpoint in the common precursor may lead to an impossible scenario. A scenario is considered ‘impossible’ if a segment deviates from the set of authorized values (-2, -1, 0, 1, 2) in the common precursor, or if a segment with a homozygous deletion (-2) in the common precursor moves to another copy number in a descendant node. These scenarios are immediately discarded.

(ii) Secondly, the ‘possible’ scenario that minimizes breakpoint usage is selected. Breakpoint usage is defined as the total number of breakpoints affecting the chromosome in the scenario. Let us call \( a^{ir} \) the amplitude vector of the common precursor after correction for breakpoints imbalance. Breakpoint usage is quantified by summing the distance between this common precursor and each node, and the number of breakpoints remaining in the nodes at the \( \text{Front} \) in the next step of the algorithm:

\[
D(a^{ir},a^i) + D(a^{ir},a^j) + \sum_{f \in \text{Front}} \max(a^i)(1 - F_k).
\]

Minimizing this distance penalizes scenarios in which very rare breakpoints occur several times.

The ‘possible’ scenario minimizing breakpoint usage is therefore selected, and the profile of the common precursor is set to: \( CP(a^i,a^j) = a^{ir} \).

If \( |\delta| = n > 1 \), the same approach is followed to select the set of \( n \) breakpoints for which modification leads to a possible scenario minimizing the total number of changes.
(a) Tumor 1

Common breakpoints

Tumor 2

Poss

b1)

Common precursor

+ BC

Tumor 1

- B

Tumor 2

\[ D(a^{1'}, a^1) = 0 \]
\[ D(a^{1''}, a^1) = (1 - F_1) + (1 - F_3) = 1.79 \]
\[ \sum_i \max(a_i^1)(1 - F_i) = \sum_i \max(a_i^{1''}, a_i^1)(1 - F_i) = (1 - F_2) + (1 - F_4) + (1 - F_5) = 2.56 \]
\[ D(a^{1''}, a^1) + D(a^{1''}, a^1) + \sum_i \max(a_i^1)(1 - F_i) = 4.35 \]

(b) A B C D

\begin{array}{cccccc}
\hline
a^1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\
a^{1''} & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\
a^1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\
a^{1''} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
F_i & 0.24 & 0.13 & 0.09 & 0.11 & 0.01 & 0.03 & 0.12 & 0.07 & 0.04 & 0.27 \\
k & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \\
\end{array}

A B C D

\begin{array}{cccccc}
\hline
a^{1''} & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\
a^1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\
a^{1''} & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\
a^1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
F_i & 0.24 & 0.13 & 0.09 & 0.11 & 0.01 & 0.03 & 0.12 & 0.07 & 0.04 & 0.27 \\
k & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \\
\end{array}

(c) b2)

Common precursor

+ C

Tumor 1

Tumor 2

\[ D(a^{1''}, a^1) = (1 - F_2) + (1 - F_3) = 1.80 \]
\[ D(a^{1''}, a^1) = 0 \]
\[ \sum_i \max(a_i^1)(1 - F_i) = \sum_i \max(a_i^{1''}, a_i^1)(1 - F_i) = (1 - F_2) + (1 - F_4) + (1 - F_5) + (1 - F_6) = 3.47 \]
\[ D(a^{1''}, a^1) + D(a^{1''}, a^1) + \sum_i \max(a_i^1)(1 - F_i) = 5.27 \]

b3)

Common precursor

+ BC

Tumor 1

+ C

Tumor 2

\[ D(a^{1''}, a^1) = (1 - F_2) + (1 - F_3) = 1.83 \]
\[ D(a^{1''}, a^1) = (1 - F_2) + (1 - F_3) = 1.87 \]
\[ \sum_i \max(a_i^1)(1 - F_i) = \sum_i \max(a_i^{1''}, a_i^1)(1 - F_i) = (1 - F_2) + (1 - F_5) + (1 - F_6) = 1.60 \]
\[ D(a^{1''}, a^1) + D(a^{1''}, a^1) + \sum_i \max(a_i^1)(1 - F_i) = 5.30 \]
Figure 1 – Correction of an unbalanced chromosome

This figure shows an example of correction for an unbalanced chromosome. In this example, tumors 1 and 2 are joined in the tree and the profile of their common precursor is being reconstructed. A third tumor (tumor 3) remains in the Front at this step of tree reconstruction.

(a) Tumors 1 and 2 have only one breakpoint in common, leading to an unbalanced chromosome in the common precursor: \( \delta = \sum_k b_k^{12} = -1 \). There are three possible corrections (Poss) for this chromosome: add an ‘up’ breakpoint from among the breakpoints present in only one of the tumors (\( b_1 \) or \( b_2 \)), or remove the common breakpoint from the common precursor (\( b_3 \)).

(b) Amplitude vectors of the three tumors, and the common amplitude vector of tumors 1 and 2, \( a_{12} \). The frequency \( F_k \) of each breakpoint in the reference data set is also indicated.

(c) Scenarios corresponding to each choice. For each case, the subtree with the common precursor and its two descendant nodes is represented, together with the corrected common amplitude vector, and the details of the calculation of breakpoint usage. Adding \( b_1 \) or \( b_2 \) leads to a scenario in which a breakpoint of the common precursor is erased by a subsequent aberration targeting the neighboring segment. Removing \( b_3 \) means the two gains occurred independently in the two samples and share a common breakpoint by chance. None of these three scenarios displays an impossible feature (such as a segment deviating from authorized values), hence breakpoint usage is calculated for all three scenarios. This is achieved by adding the distances between the common precursor and each descendant node to the number of breakpoints remaining in the Front for the next step. Here, the two nodes that will remain in the Front are the common precursor and tumor 3.

Note that in scenario \( b_1 \), the common precursor and tumor 3 share a common breakpoint. This breakpoint is therefore counted only once, in \( \sum_k \max_{j \in \text{Front}} (a_k^j)(1 - F_k) \), leading to a smaller breakpoint usage for this scenario (only 5 breakpoints in total,
versus 6 for \( b_2 \) and \( b_3 \)). Thus, if a breakpoint is present in only one of the two nodes being joined, but is also present in another node in the *Front*, it will generally be included in the common precursor. Indeed, such a breakpoint is more likely to have occurred early in the tree and to have been removed by a subsequent event in one of the two nodes being joined.

In this example, breakpoint \( b_1 \) is added to the profile of the common precursor.