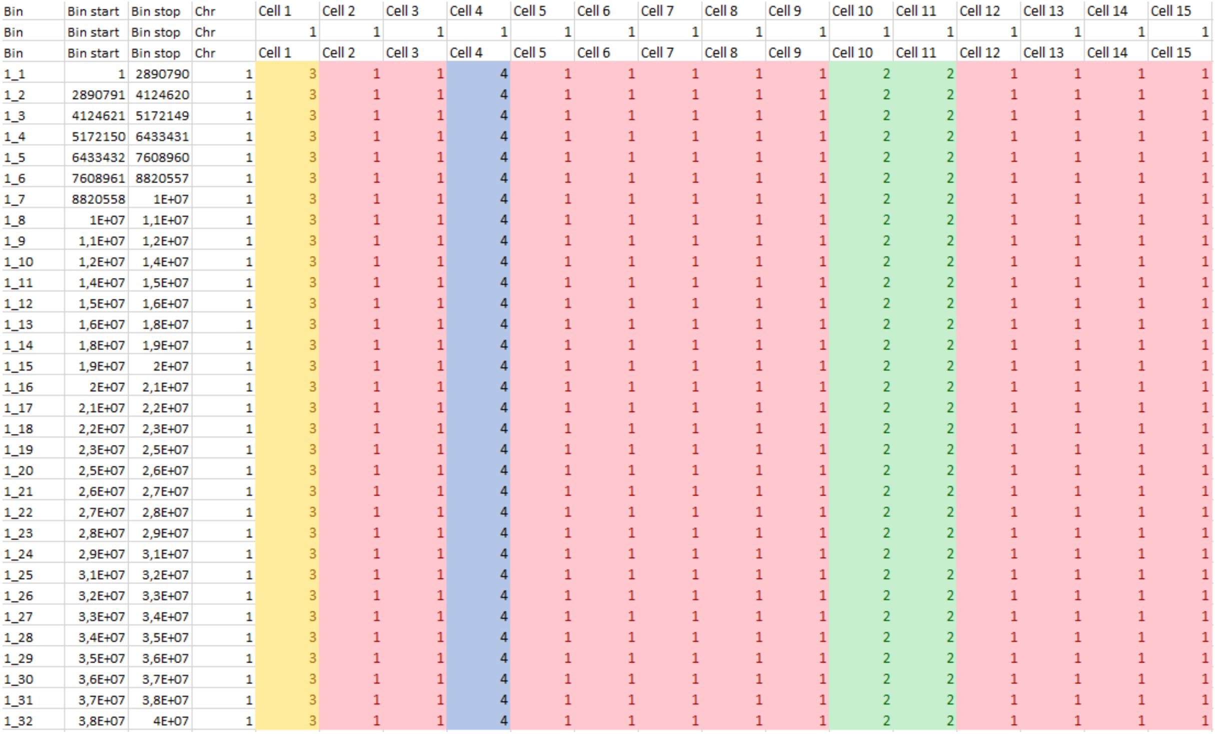
# SUPPLEMENTARY METHODS

## 1.1 SINGLE CELL WHOLE GENOME SEQUENCING DATA

Through single cell whole genome sequencing, the copy number of each 1 Mbp (mega base pair) segment of each chromosome in each individual cell is approximated. This results in a matrix with thousands of rows indicating the copy number for each chromosome segment in each analyzed cell. In **S. Figure 1** an example of a small proportion of what such a matrix might look like is visualized.



| Bin | Bin start | Bin stop | Chr | Cell 1 | Cell 2 | Cell 3 | Cell 4 | Cell 5 | Cell 6 | Cell 7 | Cell 8 | Cell 9 | Cell 10 | Cell 11 | Cell 12 | Cell 13 | Cell 14 | Cell 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Bin | Bin start | Bin stop | Chr | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| Bin | Bin start | Bin stop | Chr | Cell 1 | Cell 2 | Cell 3 | Cell 4 | Cell 5 | Cell 6 | Cell 7 | Cell 8 | Cell 9 | Cell 10 | Cell 11 | Cell 12 | Cell 13 | Cell 14 | Cell 15 |
| 1_1 | 1 | 2890790 | 1 | 3 | 1 | 1 | 4 | 1 | 1 | 1 | 1 | 1 | 2 | 2 | 1 | 1 | 1 | 1 |
| 1_2 | 2890791 | 4124620 | 1 | 3 | 1 | 1 | 4 | 1 | 1 | 1 | 1 | 1 | 2 | 2 | 1 | 1 | 1 | 1 |
| 1_3 | 4124621 | 5172149 | 1 | 3 | 1 | 1 | 4 | 1 | 1 | 1 | 1 | 1 | 2 | 2 | 1 | 1 | 1 | 1 |
| 1_4 | 5172150 | 6433431 | 1 | 3 | 1 | 1 | 4 | 1 | 1 | 1 | 1 | 1 | 2 | 2 | 1 | 1 | 1 | 1 |
| 1_5 | 6433432 | 7608960 | 1 | 3 | 1 | 1 | 4 | 1 | 1 | 1 | 1 | 1 | 2 | 2 | 1 | 1 | 1 | 1 |
| 1_6 | 7608961 | 8820557 | 1 | 3 | 1 | 1 | 4 | 1 | 1 | 1 | 1 | 1 | 2 | 2 | 1 | 1 | 1 | 1 |
| 1_7 | 8820558 | 1E+07 | 1 | 3 | 1 | 1 | 4 | 1 | 1 | 1 | 1 | 1 | 2 | 2 | 1 | 1 | 1 | 1 |
| 1_8 | 1E+07 | 1,1E+07 | 1 | 3 | 1 | 1 | 4 | 1 | 1 | 1 | 1 | 1 | 2 | 2 | 1 | 1 | 1 | 1 |
| 1_9 | 1,1E+07 | 1,2E+07 | 1 | 3 | 1 | 1 | 4 | 1 | 1 | 1 | 1 | 1 | 2 | 2 | 1 | 1 | 1 | 1 |
| 1_10 | 1,2E+07 | 1,4E+07 | 1 | 3 | 1 | 1 | 4 | 1 | 1 | 1 | 1 | 1 | 2 | 2 | 1 | 1 | 1 | 1 |
| 1_11 | 1,4E+07 | 1,5E+07 | 1 | 3 | 1 | 1 | 4 | 1 | 1 | 1 | 1 | 1 | 2 | 2 | 1 | 1 | 1 | 1 |
| 1_12 | 1,5E+07 | 1,6E+07 | 1 | 3 | 1 | 1 | 4 | 1 | 1 | 1 | 1 | 1 | 2 | 2 | 1 | 1 | 1 | 1 |
| 1_13 | 1,6E+07 | 1,8E+07 | 1 | 3 | 1 | 1 | 4 | 1 | 1 | 1 | 1 | 1 | 2 | 2 | 1 | 1 | 1 | 1 |
| 1_14 | 1,8E+07 | 1,9E+07 | 1 | 3 | 1 | 1 | 4 | 1 | 1 | 1 | 1 | 1 | 2 | 2 | 1 | 1 | 1 | 1 |
| 1_15 | 1,9E+07 | 2E+07 | 1 | 3 | 1 | 1 | 4 | 1 | 1 | 1 | 1 | 1 | 2 | 2 | 1 | 1 | 1 | 1 |
| 1_16 | 2E+07 | 2,1E+07 | 1 | 3 | 1 | 1 | 4 | 1 | 1 | 1 | 1 | 1 | 2 | 2 | 1 | 1 | 1 | 1 |
| 1_17 | 2,1E+07 | 2,2E+07 | 1 | 3 | 1 | 1 | 4 | 1 | 1 | 1 | 1 | 1 | 2 | 2 | 1 | 1 | 1 | 1 |
| 1_18 | 2,2E+07 | 2,3E+07 | 1 | 3 | 1 | 1 | 4 | 1 | 1 | 1 | 1 | 1 | 2 | 2 | 1 | 1 | 1 | 1 |
| 1_19 | 2,3E+07 | 2,5E+07 | 1 | 3 | 1 | 1 | 4 | 1 | 1 | 1 | 1 | 1 | 2 | 2 | 1 | 1 | 1 | 1 |
| 1_20 | 2,5E+07 | 2,6E+07 | 1 | 3 | 1 | 1 | 4 | 1 | 1 | 1 | 1 | 1 | 2 | 2 | 1 | 1 | 1 | 1 |
| 1_21 | 2,6E+07 | 2,7E+07 | 1 | 3 | 1 | 1 | 4 | 1 | 1 | 1 | 1 | 1 | 2 | 2 | 1 | 1 | 1 | 1 |
| 1_22 | 2,7E+07 | 2,8E+07 | 1 | 3 | 1 | 1 | 4 | 1 | 1 | 1 | 1 | 1 | 2 | 2 | 1 | 1 | 1 | 1 |
| 1_23 | 2,8E+07 | 2,9E+07 | 1 | 3 | 1 | 1 | 4 | 1 | 1 | 1 | 1 | 1 | 2 | 2 | 1 | 1 | 1 | 1 |
| 1_24 | 2,9E+07 | 3,1E+07 | 1 | 3 | 1 | 1 | 4 | 1 | 1 | 1 | 1 | 1 | 2 | 2 | 1 | 1 | 1 | 1 |
| 1_25 | 3,1E+07 | 3,2E+07 | 1 | 3 | 1 | 1 | 4 | 1 | 1 | 1 | 1 | 1 | 2 | 2 | 1 | 1 | 1 | 1 |
| 1_26 | 3,2E+07 | 3,3E+07 | 1 | 3 | 1 | 1 | 4 | 1 | 1 | 1 | 1 | 1 | 2 | 2 | 1 | 1 | 1 | 1 |
| 1_27 | 3,3E+07 | 3,4E+07 | 1 | 3 | 1 | 1 | 4 | 1 | 1 | 1 | 1 | 1 | 2 | 2 | 1 | 1 | 1 | 1 |
| 1_28 | 3,4E+07 | 3,5E+07 | 1 | 3 | 1 | 1 | 4 | 1 | 1 | 1 | 1 | 1 | 2 | 2 | 1 | 1 | 1 | 1 |
| 1_29 | 3,5E+07 | 3,6E+07 | 1 | 3 | 1 | 1 | 4 | 1 | 1 | 1 | 1 | 1 | 2 | 2 | 1 | 1 | 1 | 1 |
| 1_30 | 3,6E+07 | 3,7E+07 | 1 | 3 | 1 | 1 | 4 | 1 | 1 | 1 | 1 | 1 | 2 | 2 | 1 | 1 | 1 | 1 |
| 1_31 | 3,7E+07 | 3,8E+07 | 1 | 3 | 1 | 1 | 4 | 1 | 1 | 1 | 1 | 1 | 2 | 2 | 1 | 1 | 1 | 1 |
| 1_32 | 3,8E+07 | 4E+07 | 1 | 3 | 1 | 1 | 4 | 1 | 1 | 1 | 1 | 1 | 2 | 2 | 1 | 1 | 1 | 1 |

**S. Figure 1 A small portion of the single cell whole genotyping data.** The first column is the chromosome and bin number. The second and third column represent the chromosomal region encompassed by that bin. The fourth column is the chromosome. The first and third row are the cell names, and the second row is the number of cells encompassed by each column. If manual curation has been performed on beforehand, equal columns might have been fused and each column may thus represent multiple analyzed cells with identical genomic profiles. Each matrix element is the copy number in that bin for that cell or groups of cells.

## 1.2 PREPARING THE DATA

### 1.2.1 Loading the data

First, the data is loaded into the R environment and organized into multiple variables:

- **data:** The entire matrix where each row is a chromosomal segment, each column is a cell, and the matrix elements are the determined copy number at that position.
- **dm:** Chromosomal position described by column 1 to 4.
- **names:** Annotation for each single cell group. Row three in the dataset above.
- **nr:** Number of cells represented by each column. Row 2 in the dataset above.
- **samples:** A vector indicating to which sample each cell belong. Sometimes the cells might be from different treatment groups or timepoints. Provide a vector containing information about which column belong to which sample.
- **Ploidy:** Choose the ploidy level of the data set. The default value is 2.

- **Co:** The cutoff for an event to be considered a stem event. The default is 0.9 (90 %).

We also create an empty matrix named **d** that will contain all events encompassed by each column/cell in the dataset (**S.Figure 1, Flowchart 1**).

### 1.2.2 Fusing equal columns

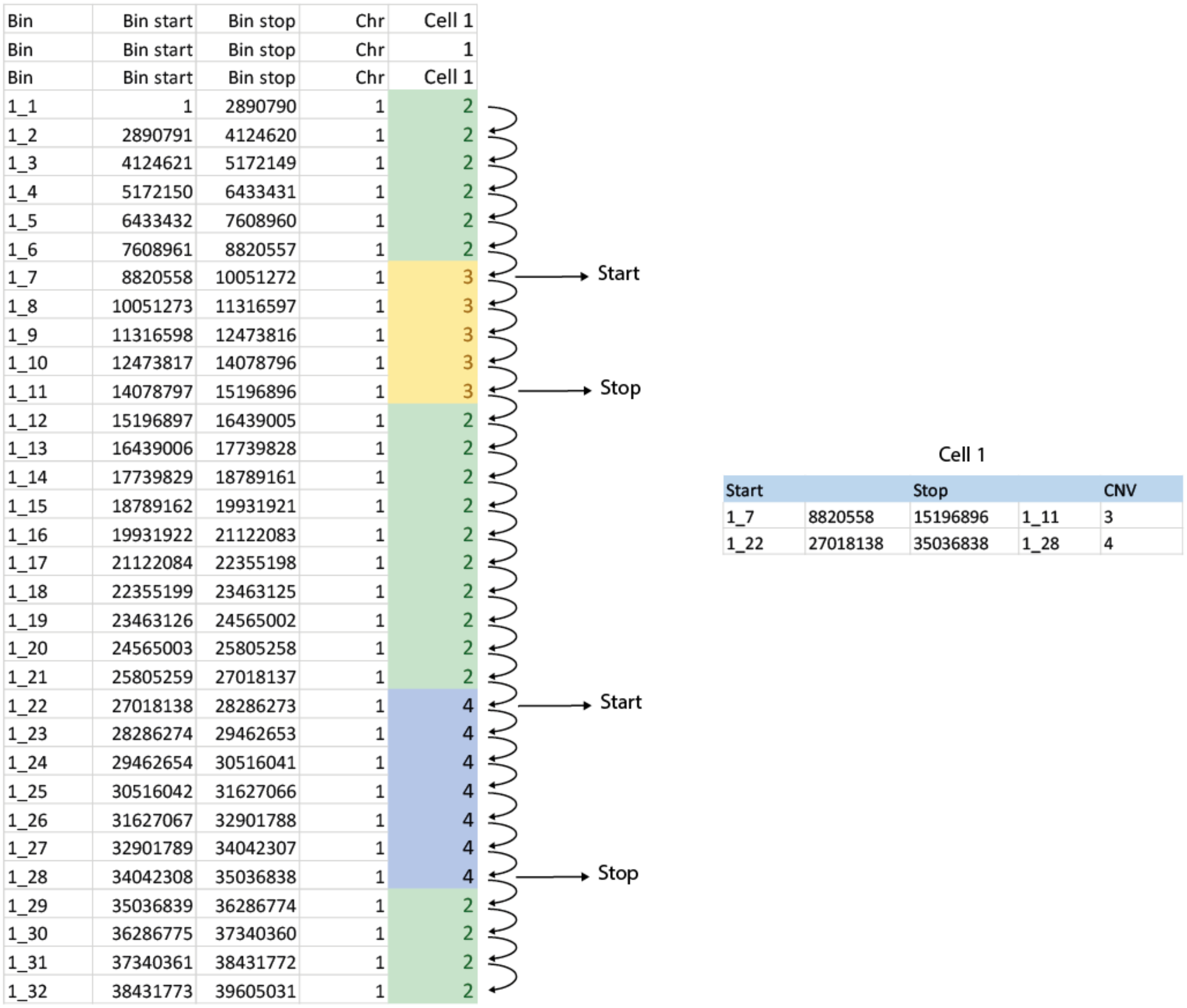
If two columns have identical genomic profiles, and belong to the same group, they are grouped together. The number of cells represented by each column in the data file is visualized in the second row (**S. Figure 1, Flowchart 1**).

## 1.3 IDENTIFYING ALL EVENTS BELONGING TO EACH INDIVIDUAL CELL

In order to deduce the evolutionary relationship between the cells in the data set, we need to construct an event matrix $E = [\hat{a}_1, \hat{a}_2 \ldots \hat{a}_k]$ illustrating the distribution of genetic alterations (rows) across the cells or groups of identical cells (columns). Each vector $\hat{a}_i$ is binary vector indicating which genetic alterations are present in cell $j$. The event matrix is subsequently used for phylogenetic reconstruction.

As a first step to deduce which genetic alterations are inherent in each cell in the data set, each column is assessed separately, without taking into consideration the other cells. A column $\boldsymbol{j}$ is chosen. The code scans through each row of column $\boldsymbol{j}$ from the first to the last (**S. Figure 2**). If the copy number in bin *i* differ from the chosen ploidy of the cell population, the cell has a genetic alteration at that position. If the copy number in bin *i* differ from the one in *i*-1, it is the starting position for a new detected alteration. The starting location of bin *i* is saved in the matrix **d** as the starting position of the new genetic alteration. Then we go on to the next row. If the consecutive bin has the same copy number as the previous one, we continue taking a step until the copy number changes again. When the copy number of bin *i* no longer equal the one of *i*-1 we are at the end of the alteration and the end position of bin *i*-1 is saved as the end position for that event. There are two special cases in which the bin's value cannot be compared to either the bin before or the one after. If we are at the first row of the column and the copy number differs from the ploidy level, this bin's starting position is the starting position for this event. If we are at the last row of the column and the copy number differs from the ploidy level, this bin's end position is the end position for this event. When the entire column has been processed in this way, it is repeated for the next column, until all have been assessed (**Flowchart 2**).

The matrix d now illustrates all cohesive segments that differ from the set ploidy level, in each column/cell. This is although not the entire story, since there might be consecutive events in locations overlapping with the previous alterations. Deducing the most probable order of these events, can be aided by combining the information from the cell under consideration together with assessment of the other cells in the dataset. Some of the most common situations that can occur in the dataset, along with explanations of how they are handled algorithmically, is described in detail in the forthcoming sections.



| Bin | Bin start | Bin stop | Chr | Cell 1 |
|---|---|---|---|---|
| Bin | Bin start | Bin stop | Chr | 1 |
| Bin | Bin start | Bin stop | Chr | Cell 1 |
| 1_1 | 1 | 2890790 | 1 | 2 |
| 1_2 | 2890791 | 4124620 | 1 | 2 |
| 1_3 | 4124621 | 5172149 | 1 | 2 |
| 1_4 | 5172150 | 6433431 | 1 | 2 |
| 1_5 | 6433432 | 7608960 | 1 | 2 |
| 1_6 | 7608961 | 8820557 | 1 | 2 |
| 1_7 | 8820558 | 10051272 | 1 | 3 |
| 1_8 | 10051273 | 11316597 | 1 | 3 |
| 1_9 | 11316598 | 12473816 | 1 | 3 |
| 1_10 | 12473817 | 14078796 | 1 | 3 |
| 1_11 | 14078797 | 15196896 | 1 | 3 |
| 1_12 | 15196897 | 16439005 | 1 | 2 |
| 1_13 | 16439006 | 17739828 | 1 | 2 |
| 1_14 | 17739829 | 18789161 | 1 | 2 |
| 1_15 | 18789162 | 19931921 | 1 | 2 |
| 1_16 | 19931922 | 21122083 | 1 | 2 |
| 1_17 | 21122084 | 22355198 | 1 | 2 |
| 1_18 | 22355199 | 23463125 | 1 | 2 |
| 1_19 | 23463126 | 24565002 | 1 | 2 |
| 1_20 | 24565003 | 25805258 | 1 | 2 |
| 1_21 | 25805259 | 27018137 | 1 | 2 |
| 1_22 | 27018138 | 28286273 | 1 | 4 |
| 1_23 | 28286274 | 29462653 | 1 | 4 |
| 1_24 | 29462654 | 30516041 | 1 | 4 |
| 1_25 | 30516042 | 31627066 | 1 | 4 |
| 1_26 | 31627067 | 32901788 | 1 | 4 |
| 1_27 | 32901789 | 34042307 | 1 | 4 |
| 1_28 | 34042308 | 35036838 | 1 | 4 |
| 1_29 | 35036839 | 36286774 | 1 | 2 |
| 1_30 | 36286775 | 37340360 | 1 | 2 |
| 1_31 | 37340361 | 38431772 | 1 | 2 |
| 1_32 | 38431773 | 39605031 | 1 | 2 |

Cell 1

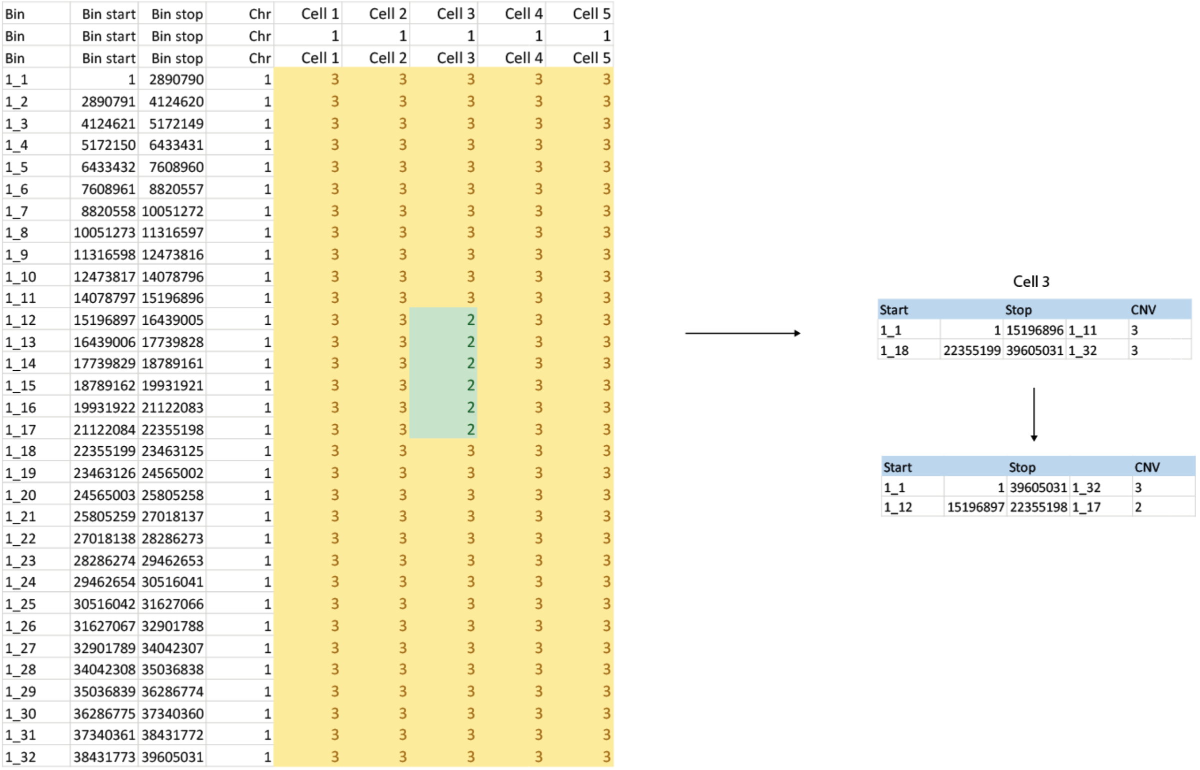
| Start | | Stop | | CNV |
|---|---|---|---|---|
| 1_7 | 8820558 | 15196896 | 1_11 | 3 |
| 1_22 | 27018138 | 35036838 | 1_28 | 4 |

**S. Figure 2** An illustration of the first assessment of the data that the algorithm does to identify the segments differing from the copy number that is chosen as the ploidy level. In this this example it is two. The start and stop position and copy number for each such section is saved in a matrix denoted d. To the right we can see how the matrix d will look like for Cell 1 in the part of the data visualized to the left.

## 1.4 Treating an event with a new event equal to the ploidy level on top of it

If a cell obtains an additional copy number alteration within a region that is already affected by a copy number change, that segment might get a copy number that equals the ploidy level. This particular part of the data set will thus, through merely using the simple algorithm above, be considered as two separate events. In **S. Figure 3** we can see a situation where a part of chromosome 1 has copy number 3 in all cells, except in cell 3, where a small segment with copy number 2 has appeared. For cell 3 the algorithm, only considering individual cells without taking into consideration the other cells, will say that this cell has two genetic alterations in this part of the data, both having a copy number of three. The part with 2:s will, in the case of a diploid cell population, not be considered an event at all. By just looking at the data, it does although seem more likely that cell 3 has had the same alteration as the other cells and then lost a small segment within it i.e., it is later in the evolutionary history of the cell population.

To solve this, we assess the matrix **d** and loop through the rows *i* belonging to one cell/column. If the event in the row *i* has the same copy number as the one in *i*-1 and they are located on the same chromosome, the start position of event *i-1* and the end position of event *i* as well as their common copy number, is stored in a vector. We also identify the size of the event in between these events. It is given by subtracting the start position of event *i* with the end position of event *i*-1. This event should make up < 50 % of the distance encompassed by event *i-1* and event *i*. If at least 3 cells have the separate events we do not form the fused version. Otherwise, we say that it is probable that the ploidal copy number alteration occurred later in the history of the cell population and the two smaller alterations are removed, the larger underlying event is created and the new ploidal event is added to the matrix **d** (**Flowchart 3**).



| Bin | Bin start | Bin stop | Chr | Cell 1 | Cell 2 | Cell 3 | Cell 4 | Cell 5 |
|---|---|---|---|---|---|---|---|---|
| Bin | Bin start | Bin stop | Chr | 1 | 1 | 1 | 1 | 1 |
| Bin | Bin start | Bin stop | Chr | Cell 1 | Cell 2 | Cell 3 | Cell 4 | Cell 5 |
| 1_1 | 1 | 2890790 | 1 | 3 | 3 | 3 | 3 | 3 |
| 1_2 | 2890791 | 4124620 | 1 | 3 | 3 | 3 | 3 | 3 |
| 1_3 | 4124621 | 5172149 | 1 | 3 | 3 | 3 | 3 | 3 |
| 1_4 | 5172150 | 6433431 | 1 | 3 | 3 | 3 | 3 | 3 |
| 1_5 | 6433432 | 7608960 | 1 | 3 | 3 | 3 | 3 | 3 |
| 1_6 | 7608961 | 8820557 | 1 | 3 | 3 | 3 | 3 | 3 |
| 1_7 | 8820558 | 10051272 | 1 | 3 | 3 | 3 | 3 | 3 |
| 1_8 | 10051273 | 11316597 | 1 | 3 | 3 | 3 | 3 | 3 |
| 1_9 | 11316598 | 12473816 | 1 | 3 | 3 | 3 | 3 | 3 |
| 1_10 | 12473817 | 14078796 | 1 | 3 | 3 | 3 | 3 | 3 |
| 1_11 | 14078797 | 15196896 | 1 | 3 | 3 | 3 | 3 | 3 |
| 1_12 | 15196897 | 16439005 | 1 | 3 | 3 | 2 | 3 | 3 |
| 1_13 | 16439006 | 17739828 | 1 | 3 | 3 | 2 | 3 | 3 |
| 1_14 | 17739829 | 18789161 | 1 | 3 | 3 | 2 | 3 | 3 |
| 1_15 | 18789162 | 19931921 | 1 | 3 | 3 | 2 | 3 | 3 |
| 1_16 | 19931922 | 21122083 | 1 | 3 | 3 | 2 | 3 | 3 |
| 1_17 | 21122084 | 22355198 | 1 | 3 | 3 | 2 | 3 | 3 |
| 1_18 | 22355199 | 23463125 | 1 | 3 | 3 | 3 | 3 | 3 |
| 1_19 | 23463126 | 24565002 | 1 | 3 | 3 | 3 | 3 | 3 |
| 1_20 | 24565003 | 25805258 | 1 | 3 | 3 | 3 | 3 | 3 |
| 1_21 | 25805259 | 27018137 | 1 | 3 | 3 | 3 | 3 | 3 |
| 1_22 | 27018138 | 28286273 | 1 | 3 | 3 | 3 | 3 | 3 |
| 1_23 | 28286274 | 29462653 | 1 | 3 | 3 | 3 | 3 | 3 |
| 1_24 | 29462654 | 30516041 | 1 | 3 | 3 | 3 | 3 | 3 |
| 1_25 | 30516042 | 31627066 | 1 | 3 | 3 | 3 | 3 | 3 |
| 1_26 | 31627067 | 32901788 | 1 | 3 | 3 | 3 | 3 | 3 |
| 1_27 | 32901789 | 34042307 | 1 | 3 | 3 | 3 | 3 | 3 |
| 1_28 | 34042308 | 35036838 | 1 | 3 | 3 | 3 | 3 | 3 |
| 1_29 | 35036839 | 36286774 | 1 | 3 | 3 | 3 | 3 | 3 |
| 1_30 | 36286775 | 37340360 | 1 | 3 | 3 | 3 | 3 | 3 |
| 1_31 | 37340361 | 38431772 | 1 | 3 | 3 | 3 | 3 | 3 |
| 1_32 | 38431773 | 39605031 | 1 | 3 | 3 | 3 | 3 | 3 |

Cell 3

| Start | | Stop | | CNV |
|---|---|---|---|---|
| 1_1 | 1 | 15196896 | 1_11 | 3 |
| 1_18 | 22355199 | 39605031 | 1_32 | 3 |

| Start | | Stop | | CNV |
|---|---|---|---|---|
| 1_1 | 1 | 39605031 | 1_32 | 3 |
| 1_12 | 15196897 | 22355198 | 1_17 | 2 |

**S. Figure 3** An example of an event having a copy number equal to the ploidy level on top of another event. In this dataset it is more likely that cell 3 have had the larger event with copy number 3 before and then lost one copy of a smaller segment within that chromosomal region, resulting in a segment with a copy number equal to the ploidy level.

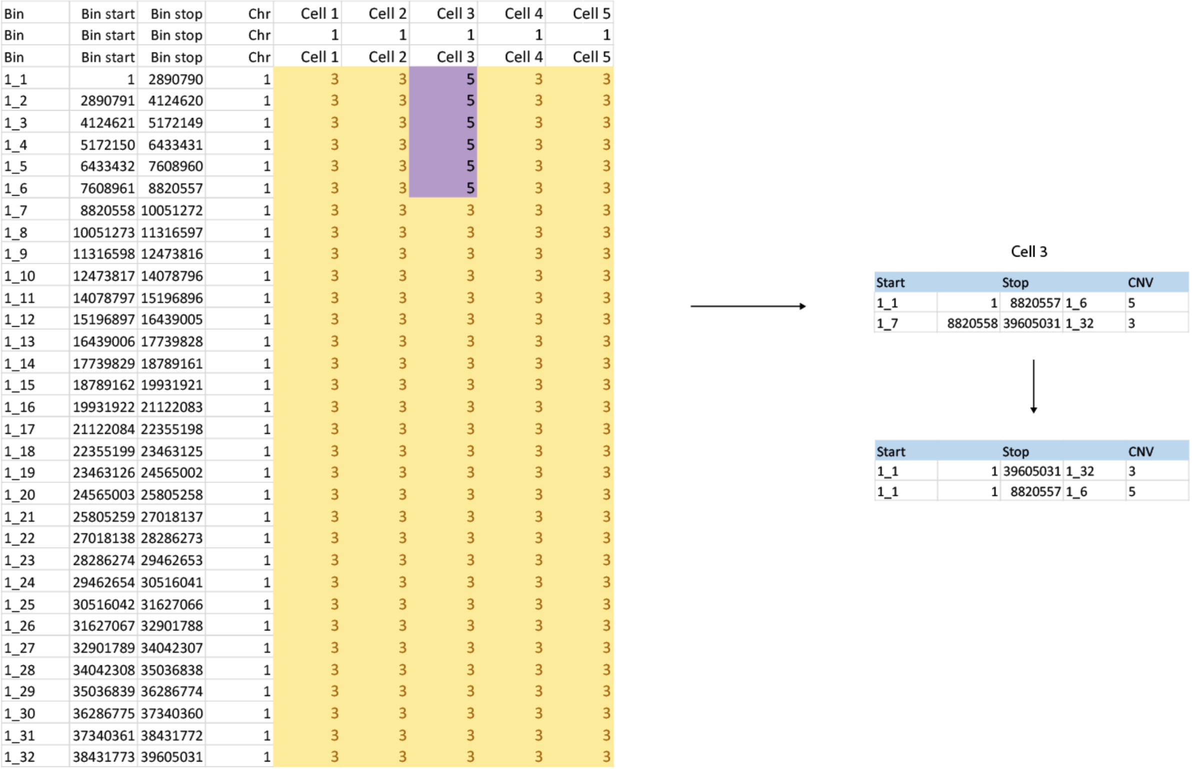
## 1.5 Combining two consecutive events to identify events on top of events

We might also have a situation similar to the one above, but where we have another copy number alteration, different from the ploidy level, on top of a previous one but with one shared breakpoint. Also, this situation will, by the simple algorithm described in section 1.3, consider these two events as separate smaller events instead of a larger event with a new on top (**S. Figure 4**). A new algorithm is needed to treat these situations (**Flowchart 4**).

The columns representing a particular cell group in the matrix d is chosen. Subsequently the algorithm loops through the rows *i* in the matrix d. The start position for event *i* and the end position of event *i*+1 is extracted. This is the interval on the chromosome spanning both of these events. Event *i* and event *i*+1 should be on the same chromosome.

First, the situation where the entire entire interval is spanned by the first event's copy number is considered. In the example in **S. Figure 4**, that would be copy number five. If this event is present in at least one other cell *and* the CNV makes up > 50 % of the entire interval or if the event is a stem event, we accept this as a probable underlying event. This event is thus added to the matrix d and the small part that is what we only see now, is removed. We also consider the other way around, in which we view it as the entire interval is encompassed by event *i*+1's copy number, in this case 3:s and do the same analysis. In this example it seems as if Cell 3 in the dataset has had the entire interval spanned by 3:s and then have gotten a new alteration on top of that, denoted by the 5:s.

This procedure is performed for all events in all cells to identify and render these types of situations. Otherwise, the program would not notice the similarity between cell 3 and the others, but merely consider it as lacking the event seen in the other cell and instead having two smaller unique alterations.



| Bin | Bin start | Bin stop | Chr | Cell 1 | Cell 2 | Cell 3 | Cell 4 | Cell 5 |
|---|---|---|---|---|---|---|---|---|
| Bin | Bin start | Bin stop | Chr | 1 | 1 | 1 | 1 | 1 |
| Bin | Bin start | Bin stop | Chr | Cell 1 | Cell 2 | Cell 3 | Cell 4 | Cell 5 |
| 1_1 | 1 | 2890790 | 1 | 3 | 3 | 5 | 3 | 3 |
| 1_2 | 2890791 | 4124620 | 1 | 3 | 3 | 5 | 3 | 3 |
| 1_3 | 4124621 | 5172149 | 1 | 3 | 3 | 5 | 3 | 3 |
| 1_4 | 5172150 | 6433431 | 1 | 3 | 3 | 5 | 3 | 3 |
| 1_5 | 6433432 | 7608960 | 1 | 3 | 3 | 5 | 3 | 3 |
| 1_6 | 7608961 | 8820557 | 1 | 3 | 3 | 5 | 3 | 3 |
| 1_7 | 8820558 | 10051272 | 1 | 3 | 3 | 3 | 3 | 3 |
| 1_8 | 10051273 | 11316597 | 1 | 3 | 3 | 3 | 3 | 3 |
| 1_9 | 11316598 | 12473816 | 1 | 3 | 3 | 3 | 3 | 3 |
| 1_10 | 12473817 | 14078796 | 1 | 3 | 3 | 3 | 3 | 3 |
| 1_11 | 14078797 | 15196896 | 1 | 3 | 3 | 3 | 3 | 3 |
| 1_12 | 15196897 | 16439005 | 1 | 3 | 3 | 3 | 3 | 3 |
| 1_13 | 16439006 | 17739828 | 1 | 3 | 3 | 3 | 3 | 3 |
| 1_14 | 17739829 | 18789161 | 1 | 3 | 3 | 3 | 3 | 3 |
| 1_15 | 18789162 | 19931921 | 1 | 3 | 3 | 3 | 3 | 3 |
| 1_16 | 19931922 | 21122083 | 1 | 3 | 3 | 3 | 3 | 3 |
| 1_17 | 21122084 | 22355198 | 1 | 3 | 3 | 3 | 3 | 3 |
| 1_18 | 22355199 | 23463125 | 1 | 3 | 3 | 3 | 3 | 3 |
| 1_19 | 23463126 | 24565002 | 1 | 3 | 3 | 3 | 3 | 3 |
| 1_20 | 24565003 | 25805258 | 1 | 3 | 3 | 3 | 3 | 3 |
| 1_21 | 25805259 | 27018137 | 1 | 3 | 3 | 3 | 3 | 3 |
| 1_22 | 27018138 | 28286273 | 1 | 3 | 3 | 3 | 3 | 3 |
| 1_23 | 28286274 | 29462653 | 1 | 3 | 3 | 3 | 3 | 3 |
| 1_24 | 29462654 | 30516041 | 1 | 3 | 3 | 3 | 3 | 3 |
| 1_25 | 30516042 | 31627066 | 1 | 3 | 3 | 3 | 3 | 3 |
| 1_26 | 31627067 | 32901788 | 1 | 3 | 3 | 3 | 3 | 3 |
| 1_27 | 32901789 | 34042307 | 1 | 3 | 3 | 3 | 3 | 3 |
| 1_28 | 34042308 | 35036838 | 1 | 3 | 3 | 3 | 3 | 3 |
| 1_29 | 35036839 | 36286774 | 1 | 3 | 3 | 3 | 3 | 3 |
| 1_30 | 36286775 | 37340360 | 1 | 3 | 3 | 3 | 3 | 3 |
| 1_31 | 37340361 | 38431772 | 1 | 3 | 3 | 3 | 3 | 3 |
| 1_32 | 38431773 | 39605031 | 1 | 3 | 3 | 3 | 3 | 3 |

Cell 3

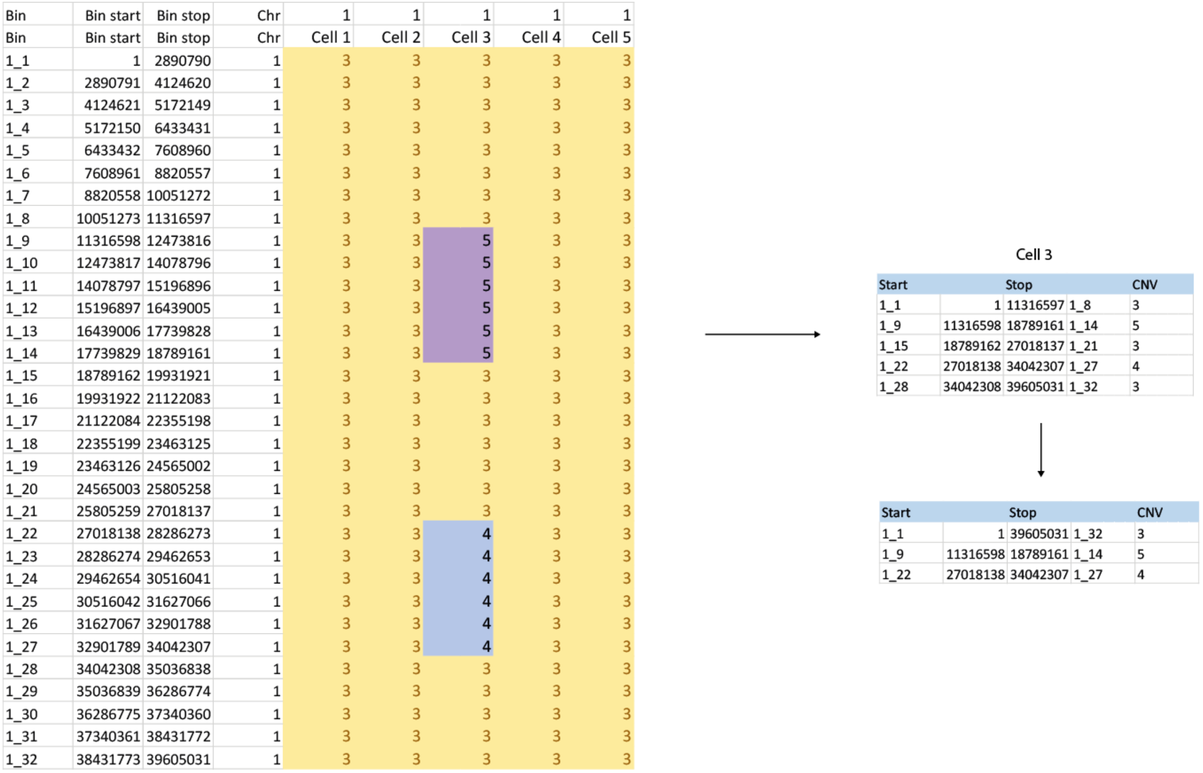
| Start | | Stop | | CNV |
|---|---|---|---|---|
| 1_1 | 1 | 8820557 | 1_6 | 5 |
| 1_7 | 8820558 | 39605031 | 1_32 | 3 |

| Start | | Stop | | CNV |
|---|---|---|---|---|
| 1_1 | 1 | 39605031 | 1_32 | 3 |
| 1_1 | 1 | 8820557 | 1_6 | 5 |

**S. Figure 4** An example of an event on top of another event that lie in the end of the underlying event. The code will initially classify Cell 3 as having two smaller events that are unique for this cell, and thus miss the obvious similarity with the other cells. By, a two step process this algorithm considers the entire interval spanned by these smaller events and either gives it the copy number of the first or the second event. It then compares that event with the other cells to draw a conclusion of whether one of these are a probable underlying genetic alteration.

## 1.6 Treating two events on top of an underlying event

We might also get a situation in which two new genetic alterations have been obtained on top of a, presumably, underlying event as in the example in **S. Figure 5**. Also, here we loop over the events identified in each cell in matrix d. If event *i*, event *i*+2 and event *i*+4 have the same copy number and are on the same chromosome, we have identified a situation such as the one visualized in **S. Figure 5**. We then extract event *i*'s start position, event *i*+4's end position and their common copy number to create a new event spanning the entire interval. If the small events are > 5 bins in size each, and each make up < 25 % of the total event size we accept them. The three smaller fragments are removed from the matrix d, and the larger underlying event is added. Here we do not compare with other cells. It is unlikely that we have 5 smaller events directly after one another rather than three ones (**S. Figure 5, Flowchart 5**).



| Bin | Bin start | Bin stop | Chr | 1 | 1 | 1 | 1 | 1 |
|---|---|---|---|---|---|---|---|---|
| Bin | Bin start | Bin stop | Chr | Cell 1 | Cell 2 | Cell 3 | Cell 4 | Cell 5 |
| 1_1 | 1 | 2890790 | 1 | 3 | 3 | 3 | 3 | 3 |
| 1_2 | 2890791 | 4124620 | 1 | 3 | 3 | 3 | 3 | 3 |
| 1_3 | 4124621 | 5172149 | 1 | 3 | 3 | 3 | 3 | 3 |
| 1_4 | 5172150 | 6433431 | 1 | 3 | 3 | 3 | 3 | 3 |
| 1_5 | 6433432 | 7608960 | 1 | 3 | 3 | 3 | 3 | 3 |
| 1_6 | 7608961 | 8820557 | 1 | 3 | 3 | 3 | 3 | 3 |
| 1_7 | 8820558 | 10051272 | 1 | 3 | 3 | 3 | 3 | 3 |
| 1_8 | 10051273 | 11316597 | 1 | 3 | 3 | 3 | 3 | 3 |
| 1_9 | 11316598 | 12473816 | 1 | 3 | 3 | 5 | 3 | 3 |
| 1_10 | 12473817 | 14078796 | 1 | 3 | 3 | 5 | 3 | 3 |
| 1_11 | 14078797 | 15196896 | 1 | 3 | 3 | 5 | 3 | 3 |
| 1_12 | 15196897 | 16439005 | 1 | 3 | 3 | 5 | 3 | 3 |
| 1_13 | 16439006 | 17739828 | 1 | 3 | 3 | 5 | 3 | 3 |
| 1_14 | 17739829 | 18789161 | 1 | 3 | 3 | 5 | 3 | 3 |
| 1_15 | 18789162 | 19931921 | 1 | 3 | 3 | 3 | 3 | 3 |
| 1_16 | 19931922 | 21122083 | 1 | 3 | 3 | 3 | 3 | 3 |
| 1_17 | 21122084 | 22355198 | 1 | 3 | 3 | 3 | 3 | 3 |
| 1_18 | 22355199 | 23463125 | 1 | 3 | 3 | 3 | 3 | 3 |
| 1_19 | 23463126 | 24565002 | 1 | 3 | 3 | 3 | 3 | 3 |
| 1_20 | 24565003 | 25805258 | 1 | 3 | 3 | 3 | 3 | 3 |
| 1_21 | 25805259 | 27018137 | 1 | 3 | 3 | 3 | 3 | 3 |
| 1_22 | 27018138 | 28286273 | 1 | 3 | 3 | 4 | 3 | 3 |
| 1_23 | 28286274 | 29462653 | 1 | 3 | 3 | 4 | 3 | 3 |
| 1_24 | 29462654 | 30516041 | 1 | 3 | 3 | 4 | 3 | 3 |
| 1_25 | 30516042 | 31627066 | 1 | 3 | 3 | 4 | 3 | 3 |
| 1_26 | 31627067 | 32901788 | 1 | 3 | 3 | 4 | 3 | 3 |
| 1_27 | 32901789 | 34042307 | 1 | 3 | 3 | 4 | 3 | 3 |
| 1_28 | 34042308 | 35036838 | 1 | 3 | 3 | 3 | 3 | 3 |
| 1_29 | 35036839 | 36286774 | 1 | 3 | 3 | 3 | 3 | 3 |
| 1_30 | 36286775 | 37340360 | 1 | 3 | 3 | 3 | 3 | 3 |
| 1_31 | 37340361 | 38431772 | 1 | 3 | 3 | 3 | 3 | 3 |
| 1_32 | 38431773 | 39605031 | 1 | 3 | 3 | 3 | 3 | 3 |

Cell 3

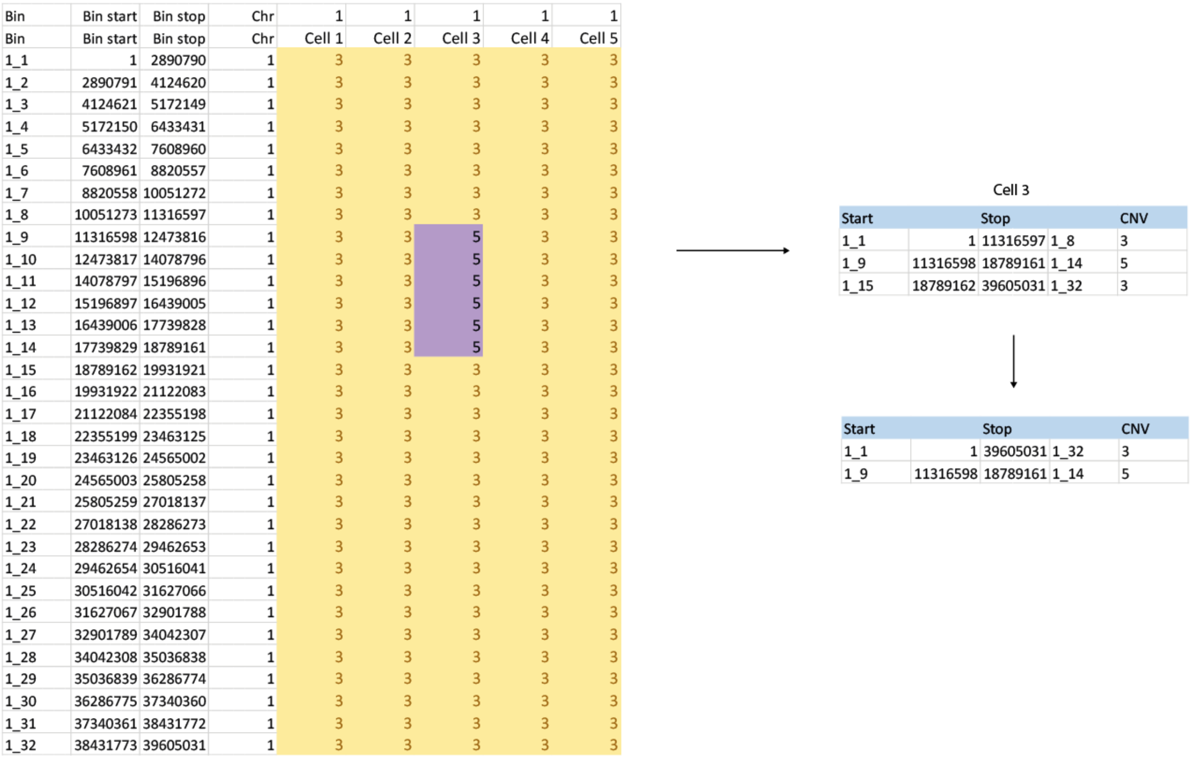
| Start | | Stop | | CNV |
|---|---|---|---|---|
| 1_1 | 1 | 11316597 | 1_8 | 3 |
| 1_9 | 11316598 | 18789161 | 1_14 | 5 |
| 1_15 | 18789162 | 27018137 | 1_21 | 3 |
| 1_22 | 27018138 | 34042307 | 1_27 | 4 |
| 1_28 | 34042308 | 39605031 | 1_32 | 3 |

| Start | | Stop | | CNV |
|---|---|---|---|---|
| 1_1 | 1 | 39605031 | 1_32 | 3 |
| 1_9 | 11316598 | 18789161 | 1_14 | 5 |
| 1_22 | 27018138 | 34042307 | 1_27 | 4 |

**S. Figure 5** An example of two events on top of another event. Initially the code will think that cell 3 have 5 small events lying directly after one another. If event *i* event *i*+2 and event *i*+4 have the same copy number, it is more likely that there has been a larger event below it and the other smaller events has occurred later in the evolution of this cell.

## 1.7 TREATING ONE EVENT IN THE MIDDLE OF ANOTHER

We might have a situation in which an event has occurred in the middle of another event. This cell will thus be considered as having three smaller events instead of two, which is unlikely. We choose one cell in matrix d and then loop over the rows. If event *i* and event *i*+2 have the same copy number, we extract the starting position of event *i* and end position of event *i*+2. Firstly, we also add their shared copy number to form a new event. If at least one other cell has this event and the copy number makes up > 50 % of the event or if this event is a stem event, we say that this cell most likely had this event before. The two smaller events are thus removed and this larger is added. In the second case we instead choose event *i*’s copy number instead and do the same comparison. If the conditions are fulfilled, we remove event *i* and add this new one (**S. Figure 6, Flowchart 6**).



| Bin | Bin start | Bin stop | Chr | 1 | 1 | 1 | 1 | 1 |
|---|---|---|---|---|---|---|---|---|
| Bin | Bin start | Bin stop | Chr | Cell 1 | Cell 2 | Cell 3 | Cell 4 | Cell 5 |
| 1_1 | 1 | 2890790 | 1 | 3 | 3 | 3 | 3 | 3 |
| 1_2 | 2890791 | 4124620 | 1 | 3 | 3 | 3 | 3 | 3 |
| 1_3 | 4124621 | 5172149 | 1 | 3 | 3 | 3 | 3 | 3 |
| 1_4 | 5172150 | 6433431 | 1 | 3 | 3 | 3 | 3 | 3 |
| 1_5 | 6433432 | 7608960 | 1 | 3 | 3 | 3 | 3 | 3 |
| 1_6 | 7608961 | 8820557 | 1 | 3 | 3 | 3 | 3 | 3 |
| 1_7 | 8820558 | 10051272 | 1 | 3 | 3 | 3 | 3 | 3 |
| 1_8 | 10051273 | 11316597 | 1 | 3 | 3 | 3 | 3 | 3 |
| 1_9 | 11316598 | 12473816 | 1 | 3 | 3 | 5 | 3 | 3 |
| 1_10 | 12473817 | 14078796 | 1 | 3 | 3 | 5 | 3 | 3 |
| 1_11 | 14078797 | 15196896 | 1 | 3 | 3 | 5 | 3 | 3 |
| 1_12 | 15196897 | 16439005 | 1 | 3 | 3 | 5 | 3 | 3 |
| 1_13 | 16439006 | 17739828 | 1 | 3 | 3 | 5 | 3 | 3 |
| 1_14 | 17739829 | 18789161 | 1 | 3 | 3 | 5 | 3 | 3 |
| 1_15 | 18789162 | 19931921 | 1 | 3 | 3 | 3 | 3 | 3 |
| 1_16 | 19931922 | 21122083 | 1 | 3 | 3 | 3 | 3 | 3 |
| 1_17 | 21122084 | 22355198 | 1 | 3 | 3 | 3 | 3 | 3 |
| 1_18 | 22355199 | 23463125 | 1 | 3 | 3 | 3 | 3 | 3 |
| 1_19 | 23463126 | 24565002 | 1 | 3 | 3 | 3 | 3 | 3 |
| 1_20 | 24565003 | 25805258 | 1 | 3 | 3 | 3 | 3 | 3 |
| 1_21 | 25805259 | 27018137 | 1 | 3 | 3 | 3 | 3 | 3 |
| 1_22 | 27018138 | 28286273 | 1 | 3 | 3 | 3 | 3 | 3 |
| 1_23 | 28286274 | 29462653 | 1 | 3 | 3 | 3 | 3 | 3 |
| 1_24 | 29462654 | 30516041 | 1 | 3 | 3 | 3 | 3 | 3 |
| 1_25 | 30516042 | 31627066 | 1 | 3 | 3 | 3 | 3 | 3 |
| 1_26 | 31627067 | 32901788 | 1 | 3 | 3 | 3 | 3 | 3 |
| 1_27 | 32901789 | 34042307 | 1 | 3 | 3 | 3 | 3 | 3 |
| 1_28 | 34042308 | 35036838 | 1 | 3 | 3 | 3 | 3 | 3 |
| 1_29 | 35036839 | 36286774 | 1 | 3 | 3 | 3 | 3 | 3 |
| 1_30 | 36286775 | 37340360 | 1 | 3 | 3 | 3 | 3 | 3 |
| 1_31 | 37340361 | 38431772 | 1 | 3 | 3 | 3 | 3 | 3 |
| 1_32 | 38431773 | 39605031 | 1 | 3 | 3 | 3 | 3 | 3 |

Cell 3

| Start | | Stop | | CNV |
|---|---|---|---|---|
| 1_1 | 1 | 11316597 | 1_8 | 3 |
| 1_9 | 11316598 | 18789161 | 1_14 | 5 |
| 1_15 | 18789162 | 39605031 | 1_32 | 3 |

| Start | | Stop | | CNV |
|---|---|---|---|---|
| 1_1 | 1 | 39605031 | 1_32 | 3 |
| 1_9 | 11316598 | 18789161 | 1_14 | 5 |

**S. Figure 6** An example of an event occurring in the middle of a previous event. In this case the code will think that Cell 3 has three separate genetic alteration directly after one another. In this situation it is more likely that Cell 3 has had the larger event with 3:s before and then gotten the event with 5:s.

## 1.8 Producing an event matrix

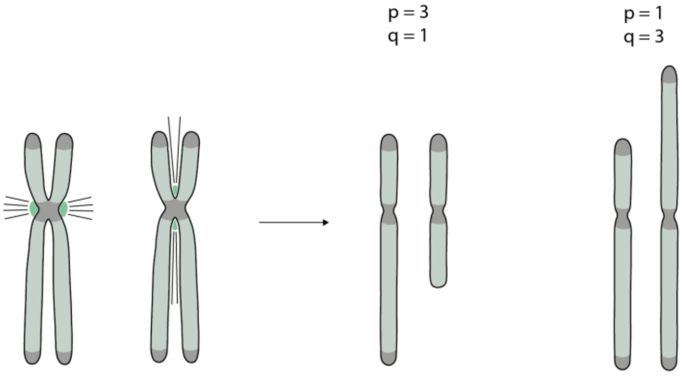
An event matrix $E = [\hat{a}_1, \hat{a}_2 \ldots \hat{a}_k]$ was created illustrating the distribution of genetic alterations (rows) across the cells or groups of identical cells (columns). Each vector $\hat{a}_i$ is binary vector indicating which genetic alterations are present in cell $j$. The event matrix is subsequently used for phylogenetic reconstruction.

## 1.9 Adding the stem events and treating lost stem events

Now we want to identify all of the genetic alterations that are stem events. First, we loop through all the rows of the event matrix. Then we compute the proportion of detected cells that has the aberration. If this proportion is larger than the event cutoff, we accept it as a stem event. All cells are thus thought to have had this event initially. Columns lacking the stem event in the event matrix at this stage are investigated further. The genomic region spanned by the stem event is extracted. Sections within this region having a copy number equal to the ploidal copy number, are added as new events in the event matrix as a subsequent event resulting in the loss of the underlying stem event (**Flowchart 7**). Note that events within this segment that differs from the ploidal copy number, is already present in the event matrix and is not needed to be added.

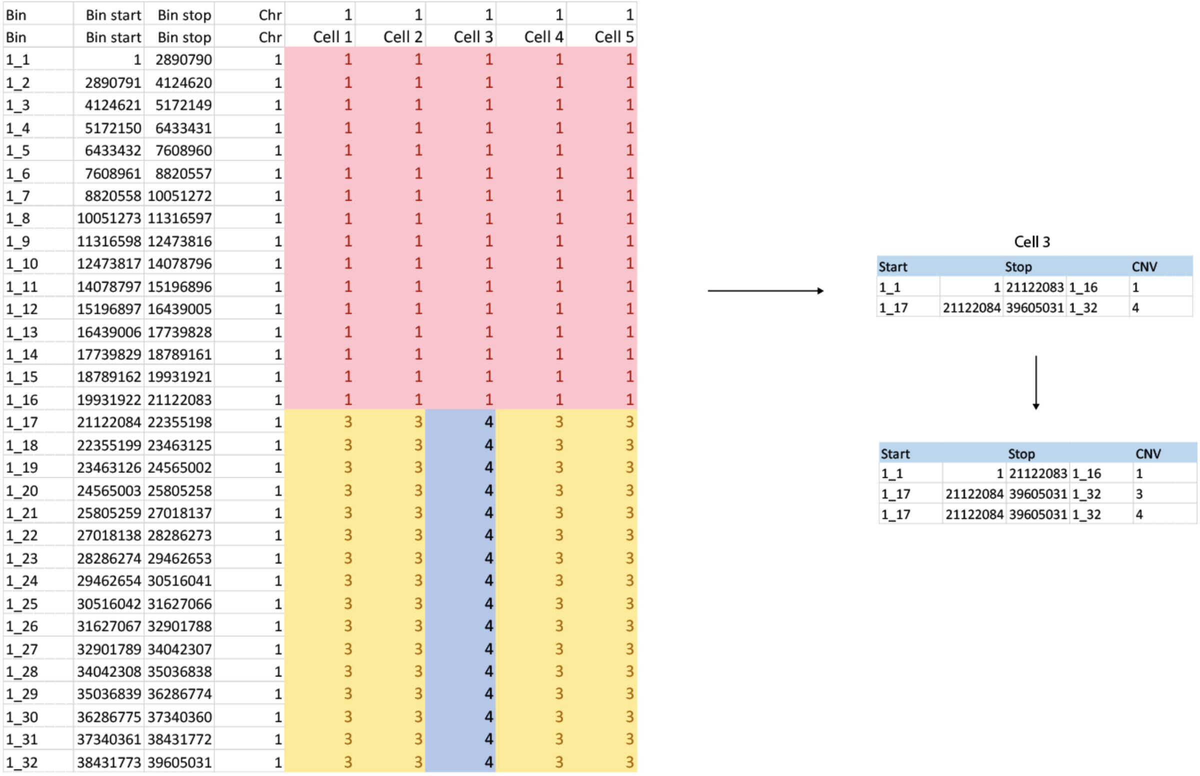
## 1.10 Isochromosome events

When the chromosomes segregate, isochromosome events may occur in which one of the chromosomes are split in the wrong direction. This results in a cell with p=3, q=1 and another cell with p=1, q=3 (**S. Figure 7**).



**S. Figure 7** An example of an isochromosome formation.

We loop through the events in the event matrix and extract the start and end position, chromosome, and copy number alteration of the alteration at the p-arm. The event should hence start at the first position of the p-arm. If the copy number at this position is 1 and it is a stem event, we create a new event from its end position +1 ranging to the end of the chromosome. The event is set to having the copy number 3. If this alteration is present in the event matrix and it is present in > 50 % of cells, we add this event to all cells (**S. Figure 8, Flowchart 8**). The same procedure is conducted for the p-arm being of copy number 3.



| Bin | Bin start | Bin stop | Chr | 1 | 1 | 1 | 1 | 1 |
|---|---|---|---|---|---|---|---|---|
| Bin | Bin start | Bin stop | Chr | Cell 1 | Cell 2 | Cell 3 | Cell 4 | Cell 5 |
| 1_1 | 1 | 2890790 | 1 | 1 | 1 | 1 | 1 | 1 |
| 1_2 | 2890791 | 4124620 | 1 | 1 | 1 | 1 | 1 | 1 |
| 1_3 | 4124621 | 5172149 | 1 | 1 | 1 | 1 | 1 | 1 |
| 1_4 | 5172150 | 6433431 | 1 | 1 | 1 | 1 | 1 | 1 |
| 1_5 | 6433432 | 7608960 | 1 | 1 | 1 | 1 | 1 | 1 |
| 1_6 | 7608961 | 8820557 | 1 | 1 | 1 | 1 | 1 | 1 |
| 1_7 | 8820558 | 10051272 | 1 | 1 | 1 | 1 | 1 | 1 |
| 1_8 | 10051273 | 11316597 | 1 | 1 | 1 | 1 | 1 | 1 |
| 1_9 | 11316598 | 12473816 | 1 | 1 | 1 | 1 | 1 | 1 |
| 1_10 | 12473817 | 14078796 | 1 | 1 | 1 | 1 | 1 | 1 |
| 1_11 | 14078797 | 15196896 | 1 | 1 | 1 | 1 | 1 | 1 |
| 1_12 | 15196897 | 16439005 | 1 | 1 | 1 | 1 | 1 | 1 |
| 1_13 | 16439006 | 17739828 | 1 | 1 | 1 | 1 | 1 | 1 |
| 1_14 | 17739829 | 18789161 | 1 | 1 | 1 | 1 | 1 | 1 |
| 1_15 | 18789162 | 19931921 | 1 | 1 | 1 | 1 | 1 | 1 |
| 1_16 | 19931922 | 21122083 | 1 | 1 | 1 | 1 | 1 | 1 |
| 1_17 | 21122084 | 22355198 | 1 | 3 | 3 | 4 | 3 | 3 |
| 1_18 | 22355199 | 23463125 | 1 | 3 | 3 | 4 | 3 | 3 |
| 1_19 | 23463126 | 24565002 | 1 | 3 | 3 | 4 | 3 | 3 |
| 1_20 | 24565003 | 25805258 | 1 | 3 | 3 | 4 | 3 | 3 |
| 1_21 | 25805259 | 27018137 | 1 | 3 | 3 | 4 | 3 | 3 |
| 1_22 | 27018138 | 28286273 | 1 | 3 | 3 | 4 | 3 | 3 |
| 1_23 | 28286274 | 29462653 | 1 | 3 | 3 | 4 | 3 | 3 |
| 1_24 | 29462654 | 30516041 | 1 | 3 | 3 | 4 | 3 | 3 |
| 1_25 | 30516042 | 31627066 | 1 | 3 | 3 | 4 | 3 | 3 |
| 1_26 | 31627067 | 32901788 | 1 | 3 | 3 | 4 | 3 | 3 |
| 1_27 | 32901789 | 34042307 | 1 | 3 | 3 | 4 | 3 | 3 |
| 1_28 | 34042308 | 35036838 | 1 | 3 | 3 | 4 | 3 | 3 |
| 1_29 | 35036839 | 36286774 | 1 | 3 | 3 | 4 | 3 | 3 |
| 1_30 | 36286775 | 37340360 | 1 | 3 | 3 | 4 | 3 | 3 |
| 1_31 | 37340361 | 38431772 | 1 | 3 | 3 | 4 | 3 | 3 |
| 1_32 | 38431773 | 39605031 | 1 | 3 | 3 | 4 | 3 | 3 |

Cell 3

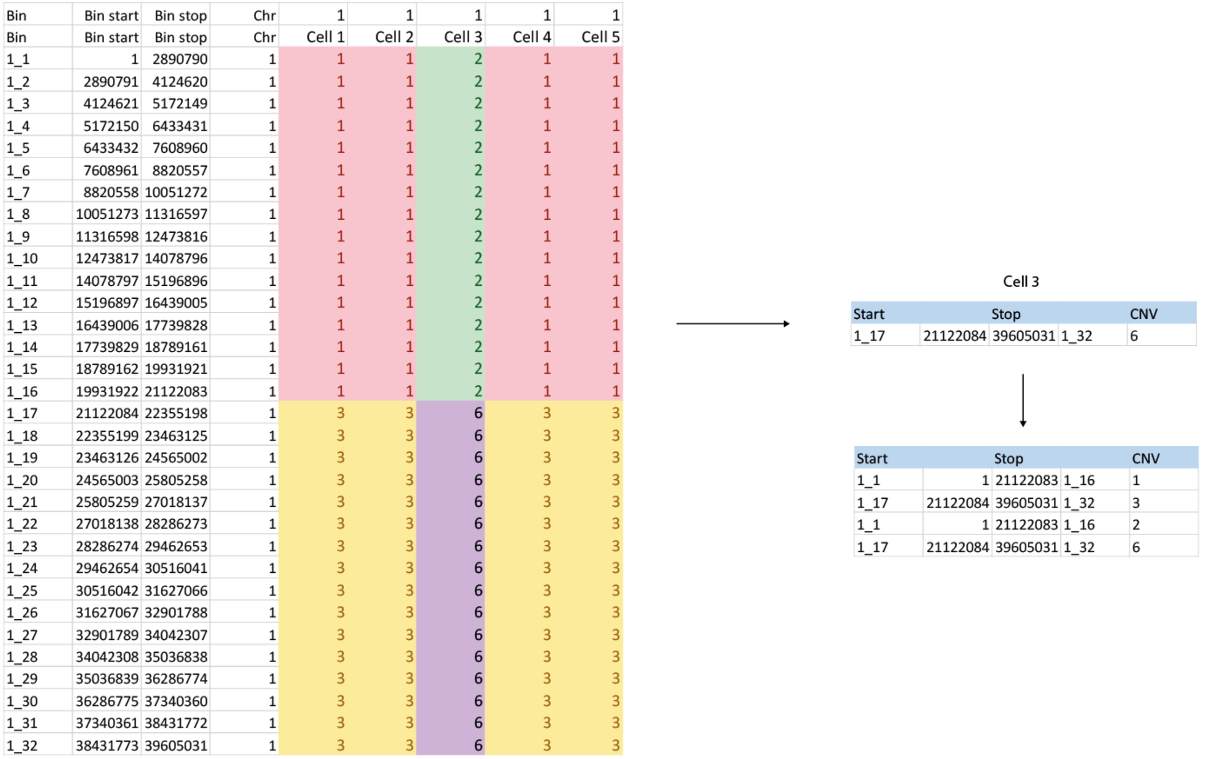
| Start | | Stop | | CNV |
|---|---|---|---|---|
| 1_1 | 1 | 21122083 | 1_16 | 1 |
| 1_17 | 21122084 | 39605031 | 1_32 | 4 |

| Start | | Stop | | CNV |
|---|---|---|---|---|
| 1_1 | 1 | 21122083 | 1_16 | 1 |
| 1_17 | 21122084 | 39605031 | 1_32 | 3 |
| 1_17 | 21122084 | 39605031 | 1_32 | 4 |

**S. Figure 8** An example of an isochromosome formation in the cell population. In this case it is likely that also cell 3 has gone through such a transition and has then gotten an extra copy of the q-arm.

## 1.11 DUPLICATIONS OF CHROMOSOMES

Other situations that can occur are duplications of entire chromosomes in which case the alterations at that chromosome will have their copy numbers doubled. To identify such situations, we loop over the events in the event matrix. If the event is present in at least 50 % of the cells we extract the start and end position as well as the copy number of that event. It should not cover the entire chromosome since we will cross over to the next chromosome in the comparison in those cases and these types of situations are taken care of in the next algorithm (section 1.12) treating consecutive events. Then we loop through the cells. If the cell has the alteration, we extract the remaining part of the chromosome. If that segment has the same copy number throughout, we save both of the events in a new matrix. When we have looped through all of the columns, we tabulate the segments. Then we loop through all of the columns again but now only choose the ones lacking the event. Then we extract the corresponding segments. We then take the copy numbers in these segments, mod the copy number in the events under consideration. If the mod is 0, we have a duplication. In this case both event 1 and 2 is added to the EM for the cell under consideration. This is repeated for all columns lacking the events. This entire comparison procedure is performed for all events in the EM (**S. Figure 9, Flowchart 9**).



| Bin | Bin start | Bin stop | Chr | 1 | 1 | 1 | 1 | 1 |
|---|---|---|---|---|---|---|---|---|
| Bin | Bin start | Bin stop | Chr | Cell 1 | Cell 2 | Cell 3 | Cell 4 | Cell 5 |
| 1_1 | 1 | 2890790 | 1 | 1 | 1 | 2 | 1 | 1 |
| 1_2 | 2890791 | 4124620 | 1 | 1 | 1 | 2 | 1 | 1 |
| 1_3 | 4124621 | 5172149 | 1 | 1 | 1 | 2 | 1 | 1 |
| 1_4 | 5172150 | 6433431 | 1 | 1 | 1 | 2 | 1 | 1 |
| 1_5 | 6433432 | 7608960 | 1 | 1 | 1 | 2 | 1 | 1 |
| 1_6 | 7608961 | 8820557 | 1 | 1 | 1 | 2 | 1 | 1 |
| 1_7 | 8820558 | 10051272 | 1 | 1 | 1 | 2 | 1 | 1 |
| 1_8 | 10051273 | 11316597 | 1 | 1 | 1 | 2 | 1 | 1 |
| 1_9 | 11316598 | 12473816 | 1 | 1 | 1 | 2 | 1 | 1 |
| 1_10 | 12473817 | 14078796 | 1 | 1 | 1 | 2 | 1 | 1 |
| 1_11 | 14078797 | 15196896 | 1 | 1 | 1 | 2 | 1 | 1 |
| 1_12 | 15196897 | 16439005 | 1 | 1 | 1 | 2 | 1 | 1 |
| 1_13 | 16439006 | 17739828 | 1 | 1 | 1 | 2 | 1 | 1 |
| 1_14 | 17739829 | 18789161 | 1 | 1 | 1 | 2 | 1 | 1 |
| 1_15 | 18789162 | 19931921 | 1 | 1 | 1 | 2 | 1 | 1 |
| 1_16 | 19931922 | 21122083 | 1 | 1 | 1 | 2 | 1 | 1 |
| 1_17 | 21122084 | 22355198 | 1 | 3 | 3 | 6 | 3 | 3 |
| 1_18 | 22355199 | 23463125 | 1 | 3 | 3 | 6 | 3 | 3 |
| 1_19 | 23463126 | 24565002 | 1 | 3 | 3 | 6 | 3 | 3 |
| 1_20 | 24565003 | 25805258 | 1 | 3 | 3 | 6 | 3 | 3 |
| 1_21 | 25805259 | 27018137 | 1 | 3 | 3 | 6 | 3 | 3 |
| 1_22 | 27018138 | 28286273 | 1 | 3 | 3 | 6 | 3 | 3 |
| 1_23 | 28286274 | 29462653 | 1 | 3 | 3 | 6 | 3 | 3 |
| 1_24 | 29462654 | 30516041 | 1 | 3 | 3 | 6 | 3 | 3 |
| 1_25 | 30516042 | 31627066 | 1 | 3 | 3 | 6 | 3 | 3 |
| 1_26 | 31627067 | 32901788 | 1 | 3 | 3 | 6 | 3 | 3 |
| 1_27 | 32901789 | 34042307 | 1 | 3 | 3 | 6 | 3 | 3 |
| 1_28 | 34042308 | 35036838 | 1 | 3 | 3 | 6 | 3 | 3 |
| 1_29 | 35036839 | 36286774 | 1 | 3 | 3 | 6 | 3 | 3 |
| 1_30 | 36286775 | 37340360 | 1 | 3 | 3 | 6 | 3 | 3 |
| 1_31 | 37340361 | 38431772 | 1 | 3 | 3 | 6 | 3 | 3 |
| 1_32 | 38431773 | 39605031 | 1 | 3 | 3 | 6 | 3 | 3 |

Cell 3

| Start | | Stop | | CNV |
|---|---|---|---|---|
| 1_17 | 21122084 | 39605031 | 1_32 | 6 |

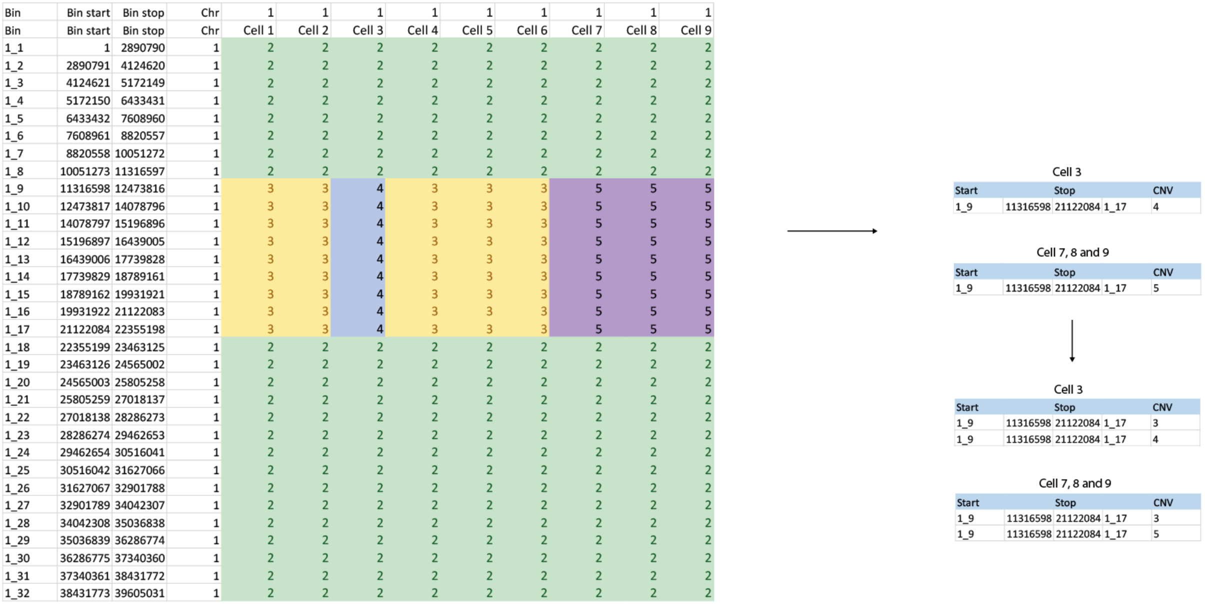
| Start | | Stop | | CNV |
|---|---|---|---|---|
| 1_1 | 1 | 21122083 | 1_16 | 1 |
| 1_17 | 21122084 | 39605031 | 1_32 | 3 |
| 1_1 | 1 | 21122083 | 1_16 | 2 |
| 1_17 | 21122084 | 39605031 | 1_32 | 6 |

**S. Figure 9** An example of a duplication of a cell in the population. Cell 3 has probably had the copy number profile just as the other cells and then duplicated.

## 1.12 Consecutive events

There might be situations in which we have gains and losses of a smaller segment. It is unlikely that another cell would get an event with exactly the same breakpoints but another copy number independently of one another. Hence, it is, in these situations, more likely that we have gotten a gain or loss of an alteration already present in a consecutive manner.

We start by looping through the events in the event matrix. We extract the start and end position of that particular event. Then we compare those breakpoints to the breakpoints of the other events in the event matrix. We extract the rows in the event matrix containing events with the same breakpoints but other copy numbers. Subsequently we tabulate the proportion of cells having each of the different versions of this segment and identify the modal number of copy numbers for the segment. If a cell has a "non-modal" number of copies of the segment, we determine that this cell has had the modal version of the segment initially, and then have gotten a new alteration in the same segment (**S. Figure 10, Flowchart 10**).



| Bin | Bin start | Bin stop | Chr | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Bin | Bin start | Bin stop | Chr | Cell 1 | Cell 2 | Cell 3 | Cell 4 | Cell 5 | Cell 6 | Cell 7 | Cell 8 | Cell 9 |
| 1_1 | 1 | 2890790 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 |
| 1_2 | 2890791 | 4124620 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 |
| 1_3 | 4124621 | 5172149 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 |
| 1_4 | 5172150 | 6433431 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 |
| 1_5 | 6433432 | 7608960 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 |
| 1_6 | 7608961 | 8820557 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 |
| 1_7 | 8820558 | 10051272 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 |
| 1_8 | 10051273 | 11316597 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 |
| 1_9 | 11316598 | 12473816 | 1 | 3 | 3 | 4 | 3 | 3 | 3 | 5 | 5 | 5 |
| 1_10 | 12473817 | 14078796 | 1 | 3 | 3 | 4 | 3 | 3 | 3 | 5 | 5 | 5 |
| 1_11 | 14078797 | 15196896 | 1 | 3 | 3 | 4 | 3 | 3 | 3 | 5 | 5 | 5 |
| 1_12 | 15196897 | 16439005 | 1 | 3 | 3 | 4 | 3 | 3 | 3 | 5 | 5 | 5 |
| 1_13 | 16439006 | 17739828 | 1 | 3 | 3 | 4 | 3 | 3 | 3 | 5 | 5 | 5 |
| 1_14 | 17739829 | 18789161 | 1 | 3 | 3 | 4 | 3 | 3 | 3 | 5 | 5 | 5 |
| 1_15 | 18789162 | 19931921 | 1 | 3 | 3 | 4 | 3 | 3 | 3 | 5 | 5 | 5 |
| 1_16 | 19931922 | 21122083 | 1 | 3 | 3 | 4 | 3 | 3 | 3 | 5 | 5 | 5 |
| 1_17 | 21122084 | 22355198 | 1 | 3 | 3 | 4 | 3 | 3 | 3 | 5 | 5 | 5 |
| 1_18 | 22355199 | 23463125 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 |
| 1_19 | 23463126 | 24565002 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 |
| 1_20 | 24565003 | 25805258 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 |
| 1_21 | 25805259 | 27018137 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 |
| 1_22 | 27018138 | 28286273 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 |
| 1_23 | 28286274 | 29462653 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 |
| 1_24 | 29462654 | 30516041 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 |
| 1_25 | 30516042 | 31627066 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 |
| 1_26 | 31627067 | 32901788 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 |
| 1_27 | 32901789 | 34042307 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 |
| 1_28 | 34042308 | 35036838 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 |
| 1_29 | 35036839 | 36286774 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 |
| 1_30 | 36286775 | 37340360 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 |
| 1_31 | 37340361 | 38431772 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 |
| 1_32 | 38431773 | 39605031 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 |

Cell 3

| Start | | Stop | | CNV |
|---|---|---|---|---|
| 1_9 | 11316598 | 21122084 | 1_17 | 4 |

Cell 7, 8 and 9

| Start | | Stop | | CNV |
|---|---|---|---|---|
| 1_9 | 11316598 | 21122084 | 1_17 | 5 |

Cell 3

| Start | | Stop | | CNV |
|---|---|---|---|---|
| 1_9 | 11316598 | 21122084 | 1_17 | 3 |
| 1_9 | 11316598 | 21122084 | 1_17 | 4 |

Cell 7, 8 and 9

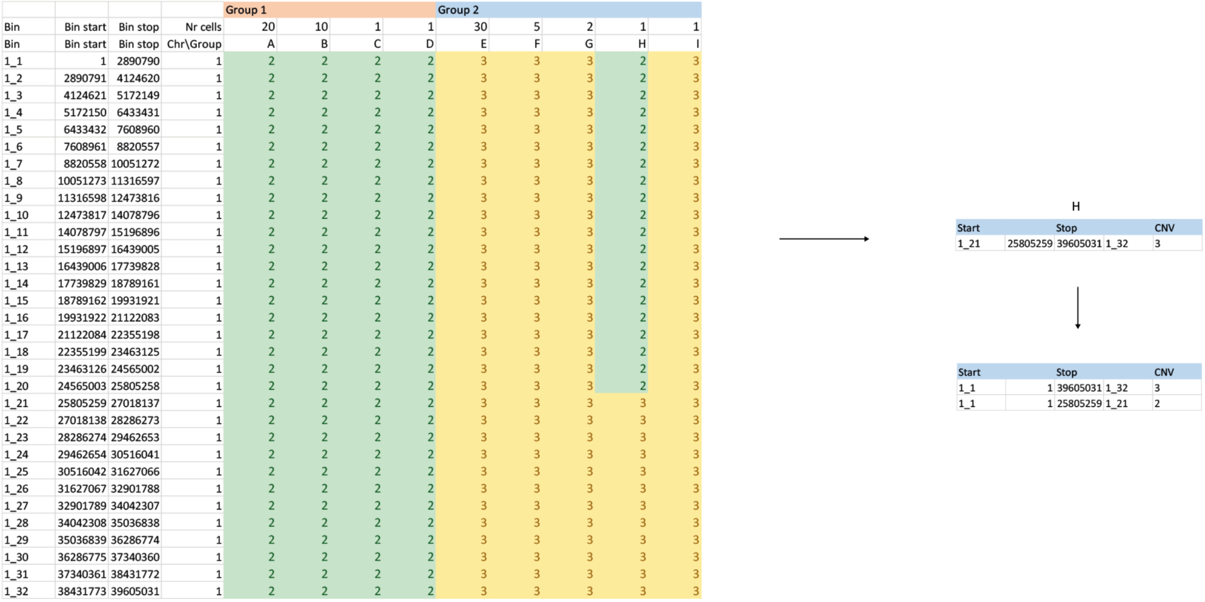
| Start | | Stop | | CNV |
|---|---|---|---|---|
| 1_9 | 11316598 | 21122084 | 1_17 | 3 |
| 1_9 | 11316598 | 21122084 | 1_17 | 5 |

**S. Figure 10** Consecutive events. The cells have most likely obtained three copies of the segment. Subsequently some cells have gained further extra copies of the segment.

## 1.13 STEM WITHIN GROUPS

The data set might contain different groups such as groups of cells extracted at a specific passaging time or location. In these cases, it might also be of value to consider similarities within these groups as well.

We loop over the events in the event matrix. Then, for each event, we calculate the proportion of cells within groups that have that genetic alteration. If this is above or equal to the predetermined cutoff for clonal events it is considered a stem event within this group. The cells not having the genetic alteration are identified. The segment encompassed by this alteration is subsequently analyzed. The cell gets the event it is missing and events with 2:s on that segment is added as new events after the stem event (**S. Figure 11, Flowchart 11**).



| | | | | Group 1 | | | | Group 2 | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Bin | Bin start | Bin stop | Nr cells | 20 | 10 | 1 | 1 | 30 | 5 | 2 | 1 | 1 |
| Bin | Bin start | Bin stop | Chr\Group | A | B | C | D | E | F | G | H | I |
| 1_1 | 1 | 2890790 | 1 | 2 | 2 | 2 | 2 | 3 | 3 | 3 | 2 | 3 |
| 1_2 | 2890791 | 4124620 | 1 | 2 | 2 | 2 | 2 | 3 | 3 | 3 | 2 | 3 |
| 1_3 | 4124621 | 5172149 | 1 | 2 | 2 | 2 | 2 | 3 | 3 | 3 | 2 | 3 |
| 1_4 | 5172150 | 6433431 | 1 | 2 | 2 | 2 | 2 | 3 | 3 | 3 | 2 | 3 |
| 1_5 | 6433432 | 7608960 | 1 | 2 | 2 | 2 | 2 | 3 | 3 | 3 | 2 | 3 |
| 1_6 | 7608961 | 8820557 | 1 | 2 | 2 | 2 | 2 | 3 | 3 | 3 | 2 | 3 |
| 1_7 | 8820558 | 10051272 | 1 | 2 | 2 | 2 | 2 | 3 | 3 | 3 | 2 | 3 |
| 1_8 | 10051273 | 11316597 | 1 | 2 | 2 | 2 | 2 | 3 | 3 | 3 | 2 | 3 |
| 1_9 | 11316598 | 12473816 | 1 | 2 | 2 | 2 | 2 | 3 | 3 | 3 | 2 | 3 |
| 1_10 | 12473817 | 14078796 | 1 | 2 | 2 | 2 | 2 | 3 | 3 | 3 | 2 | 3 |
| 1_11 | 14078797 | 15196896 | 1 | 2 | 2 | 2 | 2 | 3 | 3 | 3 | 2 | 3 |
| 1_12 | 15196897 | 16439005 | 1 | 2 | 2 | 2 | 2 | 3 | 3 | 3 | 2 | 3 |
| 1_13 | 16439006 | 17739828 | 1 | 2 | 2 | 2 | 2 | 3 | 3 | 3 | 2 | 3 |
| 1_14 | 17739829 | 18789161 | 1 | 2 | 2 | 2 | 2 | 3 | 3 | 3 | 2 | 3 |
| 1_15 | 18789162 | 19931921 | 1 | 2 | 2 | 2 | 2 | 3 | 3 | 3 | 2 | 3 |
| 1_16 | 19931922 | 21122083 | 1 | 2 | 2 | 2 | 2 | 3 | 3 | 3 | 2 | 3 |
| 1_17 | 21122084 | 22355198 | 1 | 2 | 2 | 2 | 2 | 3 | 3 | 3 | 2 | 3 |
| 1_18 | 22355199 | 23463125 | 1 | 2 | 2 | 2 | 2 | 3 | 3 | 3 | 2 | 3 |
| 1_19 | 23463126 | 24565002 | 1 | 2 | 2 | 2 | 2 | 3 | 3 | 3 | 2 | 3 |
| 1_20 | 24565003 | 25805258 | 1 | 2 | 2 | 2 | 2 | 3 | 3 | 3 | 2 | 3 |
| 1_21 | 25805259 | 27018137 | 1 | 2 | 2 | 2 | 2 | 3 | 3 | 3 | 3 | 3 |
| 1_22 | 27018138 | 28286273 | 1 | 2 | 2 | 2 | 2 | 3 | 3 | 3 | 3 | 3 |
| 1_23 | 28286274 | 29462653 | 1 | 2 | 2 | 2 | 2 | 3 | 3 | 3 | 3 | 3 |
| 1_24 | 29462654 | 30516041 | 1 | 2 | 2 | 2 | 2 | 3 | 3 | 3 | 3 | 3 |
| 1_25 | 30516042 | 31627066 | 1 | 2 | 2 | 2 | 2 | 3 | 3 | 3 | 3 | 3 |
| 1_26 | 31627067 | 32901788 | 1 | 2 | 2 | 2 | 2 | 3 | 3 | 3 | 3 | 3 |
| 1_27 | 32901789 | 34042307 | 1 | 2 | 2 | 2 | 2 | 3 | 3 | 3 | 3 | 3 |
| 1_28 | 34042308 | 35036838 | 1 | 2 | 2 | 2 | 2 | 3 | 3 | 3 | 3 | 3 |
| 1_29 | 35036839 | 36286774 | 1 | 2 | 2 | 2 | 2 | 3 | 3 | 3 | 3 | 3 |
| 1_30 | 36286775 | 37340360 | 1 | 2 | 2 | 2 | 2 | 3 | 3 | 3 | 3 | 3 |
| 1_31 | 37340361 | 38431772 | 1 | 2 | 2 | 2 | 2 | 3 | 3 | 3 | 3 | 3 |
| 1_32 | 38431773 | 39605031 | 1 | 2 | 2 | 2 | 2 | 3 | 3 | 3 | 3 | 3 |

H

| Start | | Stop | | CNV |
|---|---|---|---|---|
| 1_21 | 25805259 | 39605031 | 1_32 | 3 |

| Start | | Stop | | CNV |
|---|---|---|---|---|
| 1_1 | 1 | 39605031 | 1_32 | 3 |
| 1_1 | 1 | 25805259 | 1_21 | 2 |

**S. Figure 11** An example of a stem event within a specific group in the data set. Here it is likely that cell H have had the larger event with copy number 3 and then lost a part of it. If the subgroups had not been considered, cell H would not have the larger copy number aberration with copy number 3, but merely the smaller one.

## 1.14 PHYLOGENETIC RECONSTRUCTION

We now have the final event matrix. It can subsequently be used to reconstruct phylogenetic trees using algorithms such as the maximum likelihood or maximum parsimony methods (**Flowchart 12**).