

File format

TumGrowth accepts tab delimited files. It is best to download an example to get a grasp on a *cleaned-up* structure. The formatting basics are as follows:

- First line must contain a **Grp** field that describe the treatment group and the days of sampling given in numerals. Any cell on the first line that do not correspond to a given experimental information (*Grp*, *Mid*, *Use*, *Surv*) or that it is not a number will not be considered;
- Second line contains the name of the variable associated to each time point. There will be as many measurement type as different names found under each potential time point. Columns that corresponding to a time point will be discarded if the cells on the second line are empty;
- Remaining lines: raw measurements/information for each mice.

Optional columns on the first line:

- *Mid*: to specify conveniently the Id for each animal; Default: Grp plus a number
- *Use*: if not empty, the animal will not be used for graphing and analyses but the data remains available for export; Default: all lines are considered
- *Surv*: if anything other than empty, for the animal to have survived on its last recorded time point (see point 7 below). Default: event for all

The parsing takes care of the comma/dot issue to get the numbers right but anything other than this will be considered as NA. Graphes/calculations are done using the order of the groups in the original file. Animal identities may be amended to avoid duplication: one line=one animal.

Toy dataset:

Grp	Use	Surv	Mid	0	5	10	15	15	20	d25
what	ever	is	here	V	V	V	V	W	V	V
PBS	x	x	animal2	2	4	7	8	500	14	1
PBS	x	x	animal2.5	2.5	10.3	15.6	15.6	432	15.6	like
PBS	x		animal3	3	14.3	25.6	65.6	432	85.1	
Drug	x	x	animal4	4	10	15	23	147	60	b10
Drug	x		animal2.6	2.6	5.4	9.7	0	285	eighty	0
Drug	x	x	animal6.6	6.6	0	0.7	0	120	0	613

This is interpreted as:

- Some measures named *V* has been recorded at 2/5/10/15/20 and another one named *W* at 15 (first 2 rows, mandatory);
- 3 *PBS* abd 3 *Drug*-treated mice (**Grp**, mandatory column);
- Animals are called *animal1*, *animal2* etc... (**Mid**, optional column);
- Use them all for graphing and stats (**Use**, optional column);

- Last recorded day is censored for *animal3* (**Surv**, optional column);
- Measurements in *V* at day 15 and 20 are not valid for *animal5*;
- TumGrowth is not interested in the last column as the first cell is not a number.

Parsing

For historical reasons down to pre-filled Excel sheet being in place in the Kroemerlab, the parsing proceeds in the following order:

1. Num. of digits: rounding used for import and exporting data; Toy: putting zero in the box will round-up everything to their closest integer
2. Exclude zeros starting from the end. Needed if the original spreadsheet used to record data is pre-filled with zeros; Toy: if ticked, last time point for animal5 will be 10
3. Exclude identical values starting from the end of the time course; Toy: if ticked, measurements at days 15 and 20 will be excluded for animal2. This mouse will then be considered to have survived until time point 10
4. Shifting time courses to the left: the first non zero measurement will be set to the value given in the box; Toy: putting zero in the box, days 2/5/10/15/20 become 0/3/8/13/18 for all animals and 0/5/10/15 for animal3 that has a zero at the first sampling day
5. Exclude any other zeros: should be zeros considered NA;
6. For log-transformation, zeros (if any left) are imputed by the minimum value divided by 2;
7. This applies is **Surv** is not specified in the original file. By default, animals are considered dead on their last recorded day. Option 7 overwrites this by censoring (*i.e. alive*) animals if there is a valid measurement at the **latest** day of the experiment. For more subtle or complex situations, it is best to specify **Surv** properly in the input text file;
8. Colors can be amended.

Export

Data after parsing can be exported as a tab-separated text file (*.tsv*), excluding by default the responses after transformation. The same file can re-imported for later use after amending its content.

Datasets

Test datasets:

1. **Toy**: simple dataset explaining the basic file structure and different parsing options;
2. **KO**: data published in Kolb (2008) to illustrate T/C ratios;
3. **Longitudinal test**: data published in Vetizou (Science, 2015), slightly amended for illustrative purposes;
4. **Survival test**: data published in Vetizou (Science, 2015), slightly amended for illustrative purposes;
5. **Cross-sectional test**: data published in Pietrocola (Cancer Cell, 2016), slightly amended for illustrative purposes;