Deconvolution Algorithm Benchmarking

Today, we (Rainer, Jakob, Marian) discussed a structure for the DAB paper.

Who is our target audience?

In the deconvolution world, there are 3 possible scientist, that we may try to reach with the dab paper:
1. Guys, without any interest in bioinformatics, or method background. They just have Y, and are desperate in getting C. Any solution will be good enough.
2. Guys, that try to set up a framework for a bigger analysis. They might be interested in measuring C for a set of Y. In the end, they don't want to implement their own solution, but get the best solution on the market.
3. Guys, that implement their own algorithms, and want to compare their result fair on the same dataset.

For group (1) we do not need to put any focus onto scoring, we can just show something. For group (2) we should clearly emphasize why comparing on the current dataset is necessary, and we should discuss why and how scoring without the true C can/should be done. For group (3) more theory is necessary. Therefore, we like to target group (2).
if we satisfy group (i), group (i-1) will be satisfied as well.

Structure of the paper

(how to show that?) a) modern deconv algo's train on scRNA-seq, what is the effect of the data on the model? Can model adapted to d1, applied to d2. b) big difference between algo and model! 1. Deconvolution is a competitive field, there are many different algorithms, state of the art algorithms adapt a deconovlution model to scRNA-seq data (e.g. DTD, MUSIC, cibersortx). However, none of those models works as an one-solve-all solution. show this only on DTD, and additionally show it for e.g. MUSIC and put that into the supplement. 2. The performance can be evaluated best looking onto cor(C, $\hat{C}$) for real bulks. However, most of the time, C is not available, so we need another score 3. Using scRNA-seq data, we can create artifical mixtures, especially with various variances, and therefore rank algorithms without C for real bulks. (Rainer: this is more a side thing, will not be in the abstract.) do this on all dream data's! 4. side note, there is an R package

don't say score, but ... ??? it mimics the correlation

- show our results on "the" dream data
- using the "score" that this is just one point on the fits => extrapolating here does work

- realer score invertieren. (on paper)

plots we need: for 1) all Han models deconvolute all dream dataset. for 2) literature, no plots for 3) all Han models do Tim's score/metric on all dream datasets. are the winner the same as with (1) for 4) kind of flow chart of DAB

all: look onto the report markdown and vignette code documentation => Tim + delegieren an mich und jakob