# Fitting ALFA-K model to passaging data

ALFA-K was fit to all lineages of data from passaging experiments by Salehi et al. A lineage was defined as two or more successive samples from the same cell line. Thus e.g. for SA532 there are 7 2-sample lineages, 6 3-sample lineages etc. Fig. A shows the relationship between samples for all 5 cell lines. Points are colored according to the $R^2$ metric score obtained after fitting ALFA-K to the longest lineage terminating at that point. We were not able to obtain good fits to SA532 or SA906 lineage B. Examining $R^2$ metric scores for all lineages obtained from the data, we found in agreement with our ABM results that longer lineages result in better fits. We next applied ALFA-K to lineages with descendants, simulated population evolution, then compared the simulated populations to the withheld samples from subsequent passages using the angle metric. We were able to obtain good predictions for the next passage, with the quality of the predictions improved when there was a larger number of samples used to train ALFA-K or when ALFA-K cross validation had a positive Rˆ2 score.
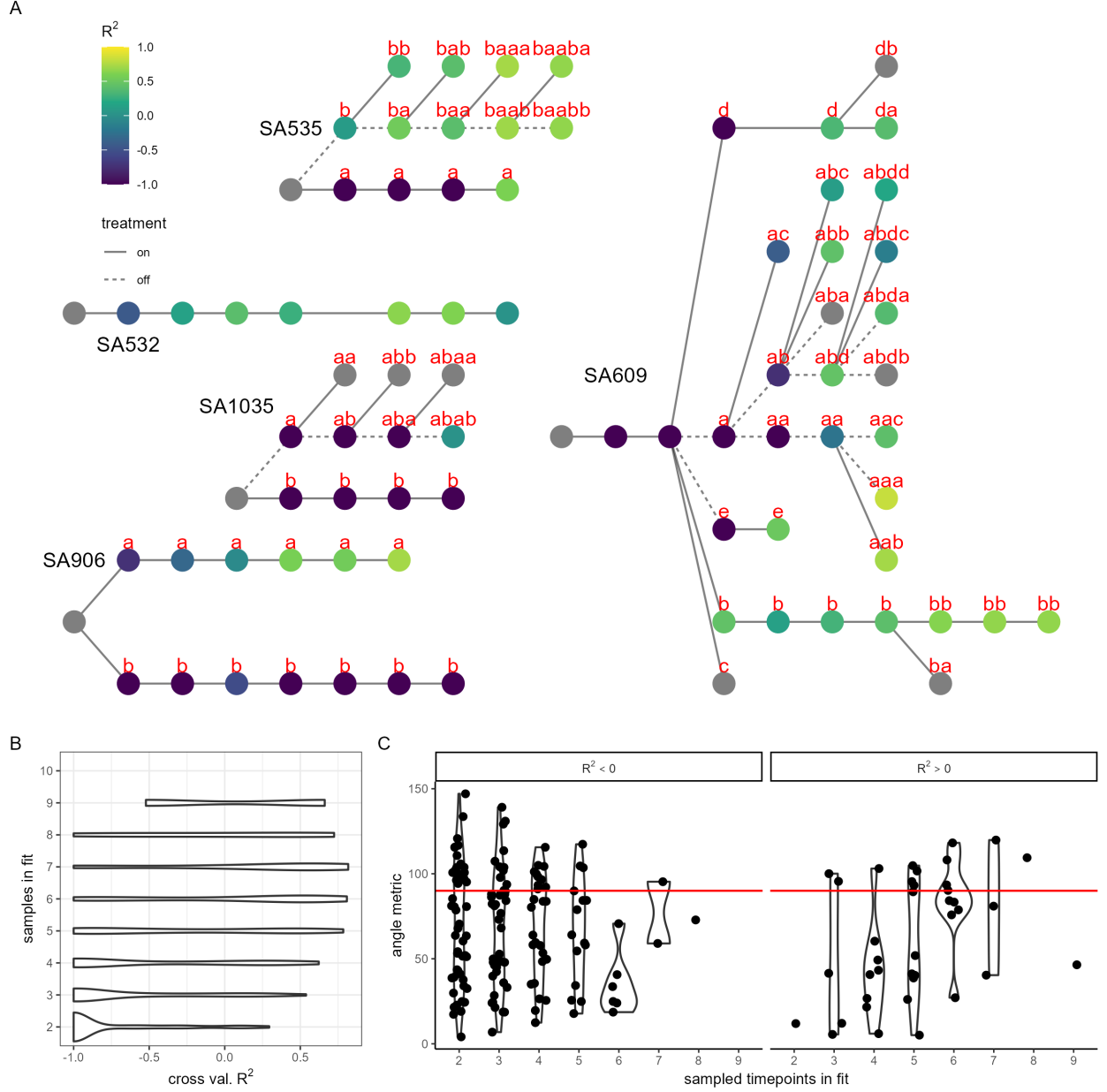
Figure 3. Fits to cell passaging data A) Maps showing the relationship cell passages that were sequenced and used as inputs to ALFA-K. Each filled circle represents a sequenced sample. Lines connect parental passages to children, with line type indicating the presence or absence of cisplatin treatment in the intervening period. Circles are colored according to the $R^2$ metric value for fitted landscapes including that sample and all its ancestors. B) $R^2$ metric scores for all sublineages in the dataset, grouped according to the number of samples used in the fitting. C) Angle metric scores for all lineages with at least one descendent, grouped according to the number of samples used in the fitting and the $R^2$ metric score. Red horizontal line indicates the expected value of the angle under the null hypothesis (i.e. that the fitted and true landscape are different).