

Figure 1: a schematic graph showing the mixture model of mutational processes and signatures

Figure 1

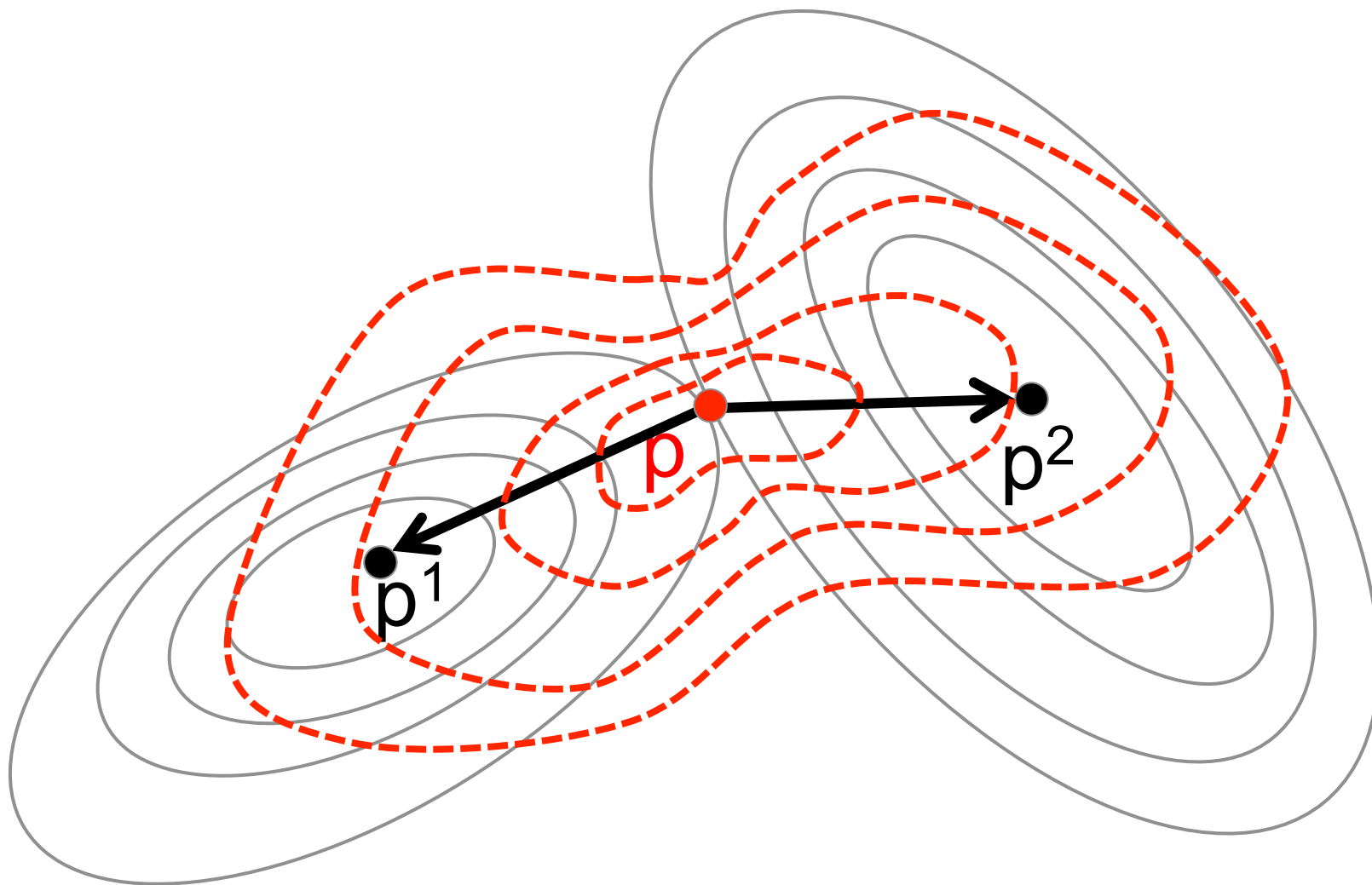


Figure 2A: contour plot of the penalty function of multinomial sampling function (optimum at p^1) and the least square of signature fitting (optimum at p^2). LIBRA tries to infer p by jointly optimizing both penalties (red contour lines, optimum at p)

Figure 2A

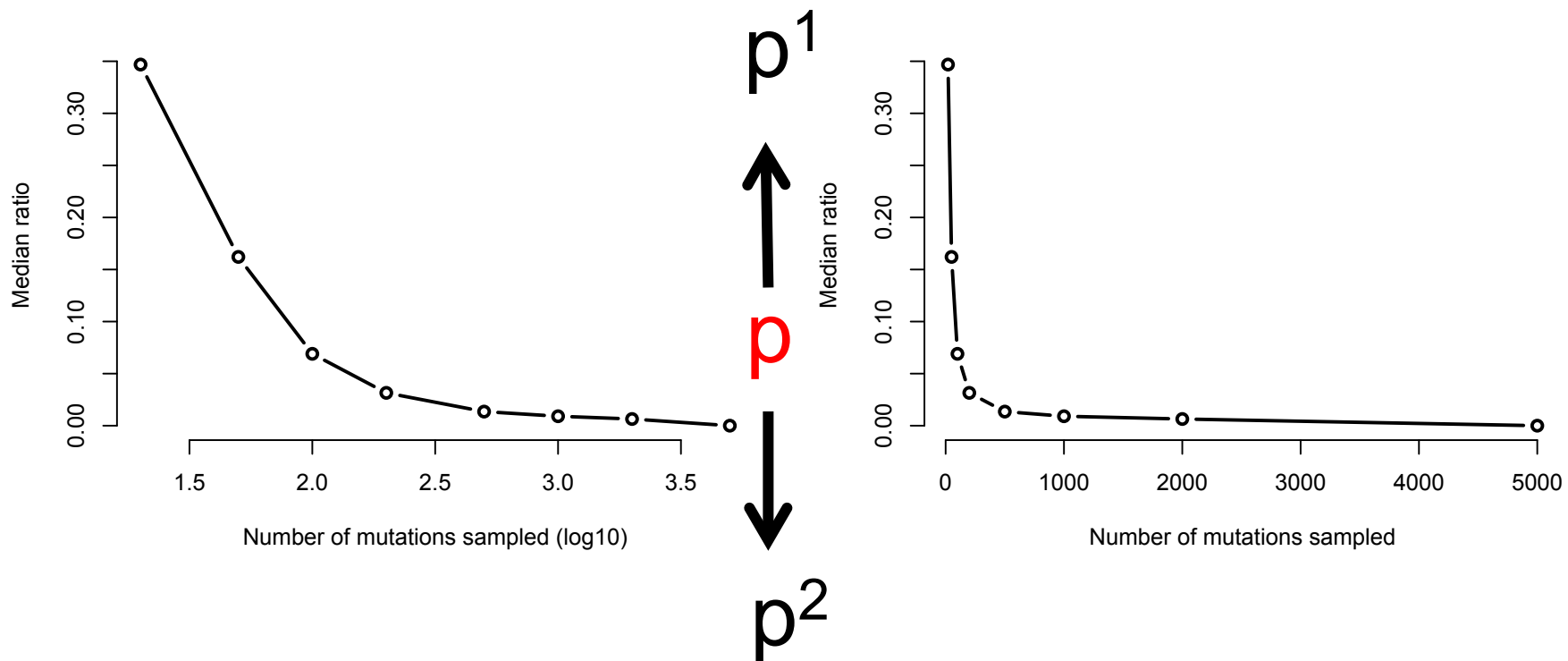


Figure 2B: As mutation number increases, the inferred p gets closer to the sampling MLE rather than the linear fitting as the variance due to sampling is smaller.

Figure 2B

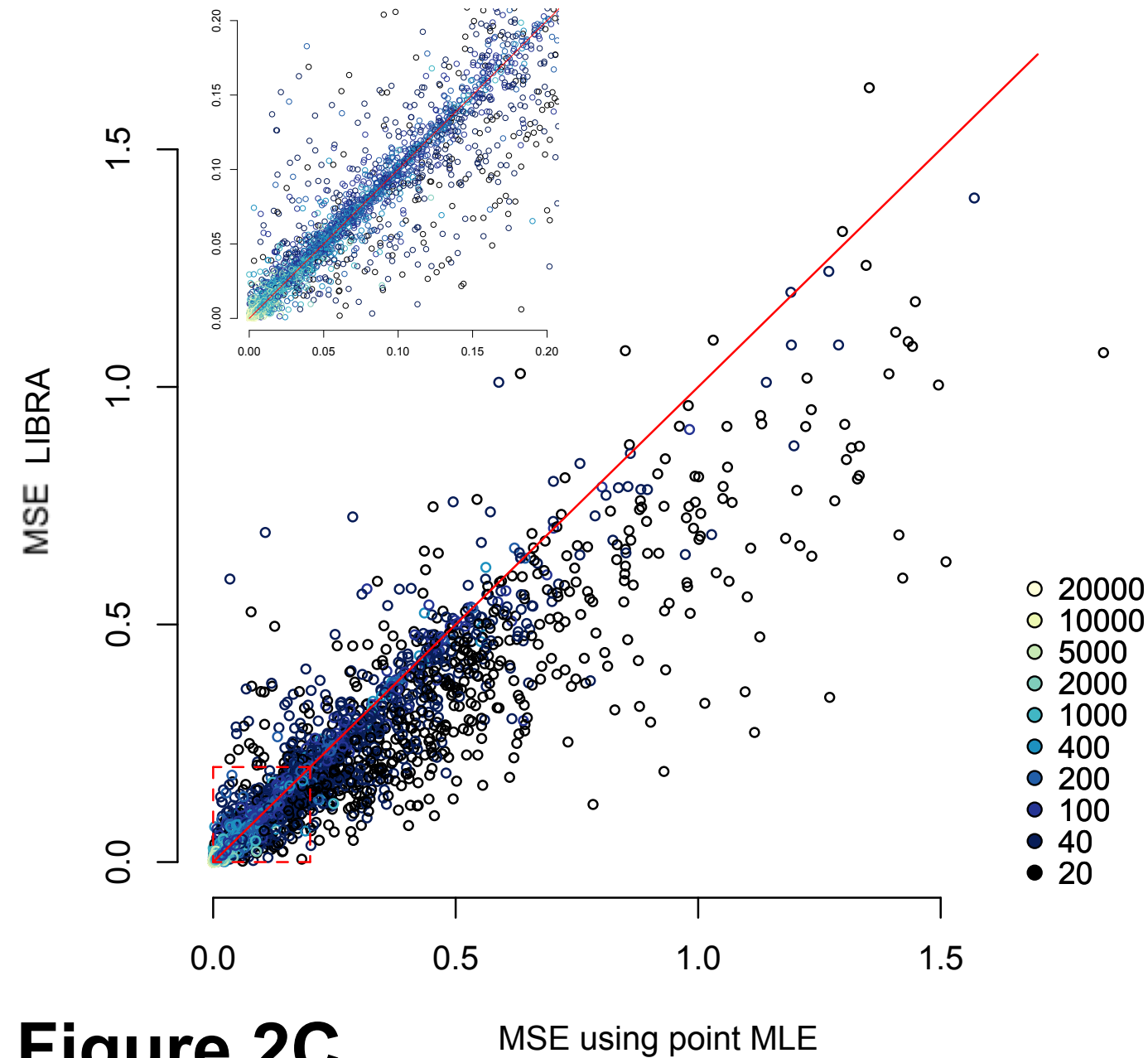


Figure 2C: MSE of LIBRA and just using the point MLE to fit the signatures. Low mutation counts profiles benefit from LIBRA the most.

- 20000
- 10000
- 5000
- 2000
- 1000
- 400
- 200
- 100
- 40
- 20

Figure 2C

MSE using point MLE

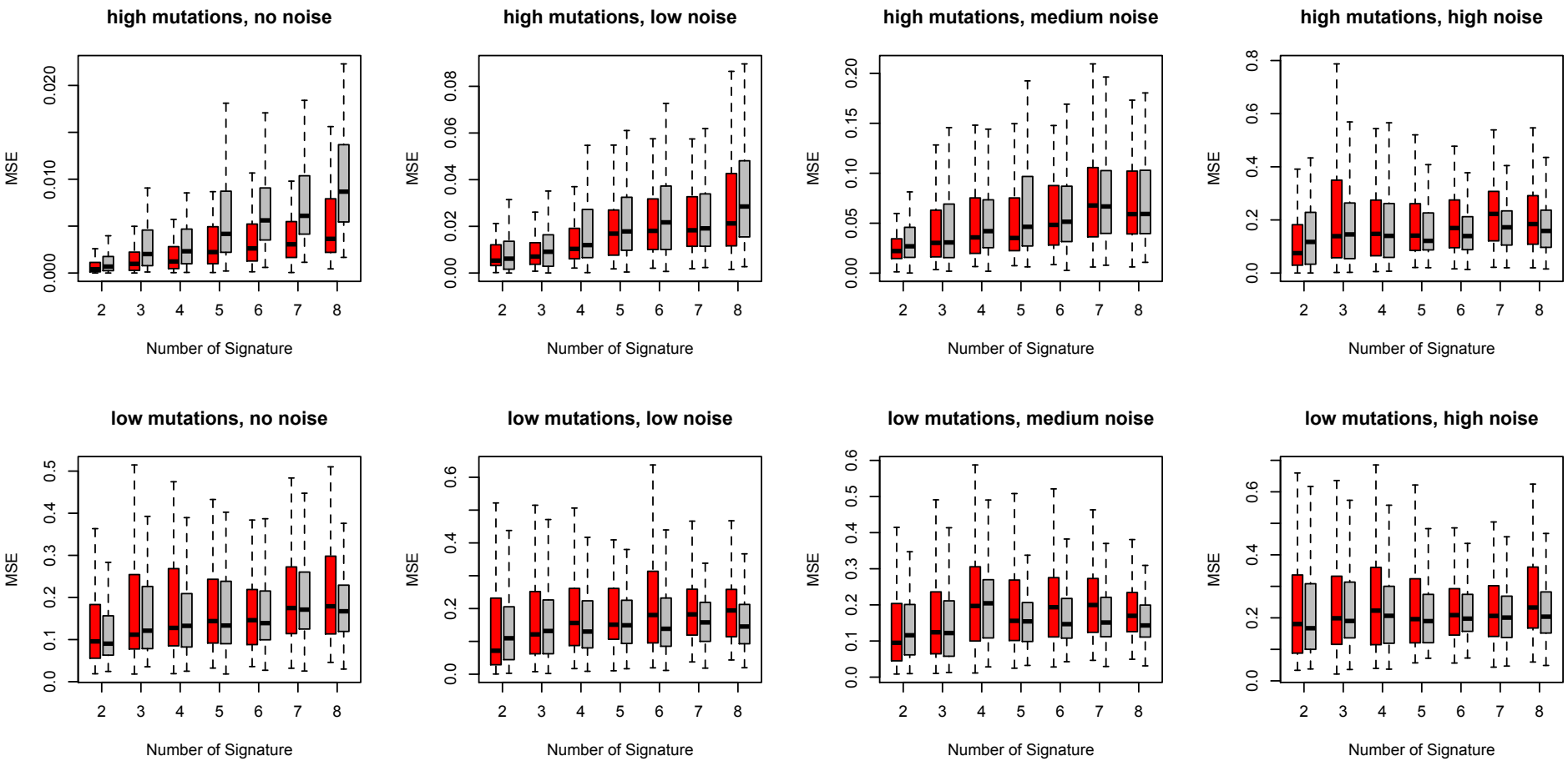


Figure 3A: Boxplots of MSE on simulated datasets.
 Red: LIBRA, grey: deconstructSigs

Figure 3A

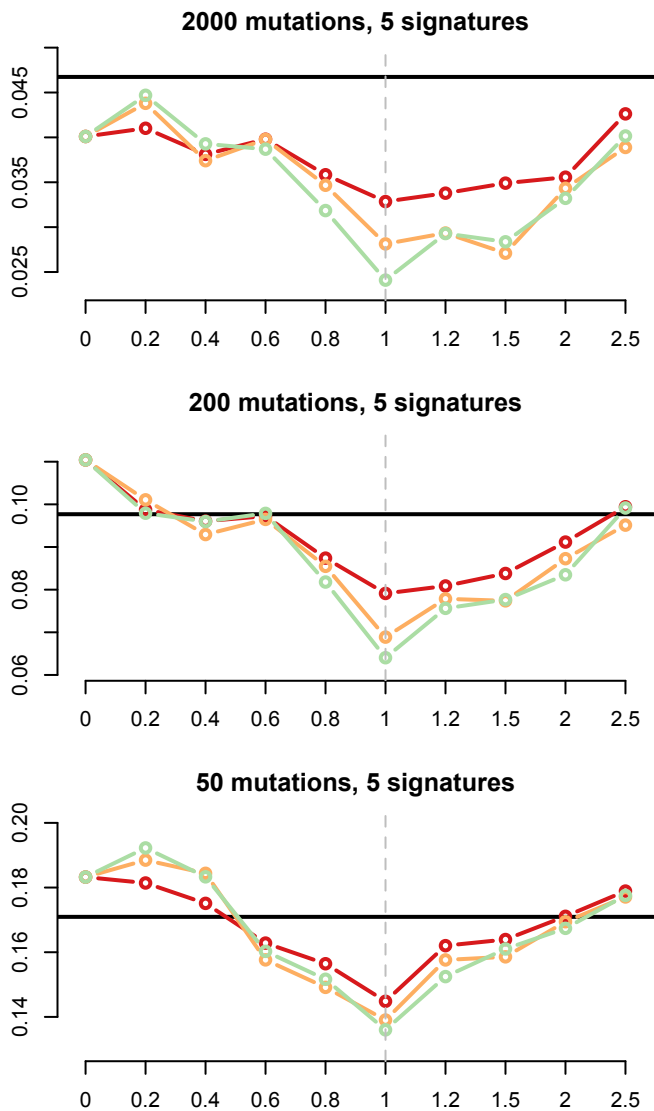


Figure 3B: MSE on simulated datasets, showing tuning the penalty weights using the prior knowledge improves performance. Penalty weights used: red, 0.5; yellow, 0.2; green, 0.1. Black line denotes deconstructSigs

Figure 3B

precision

recall

fscore

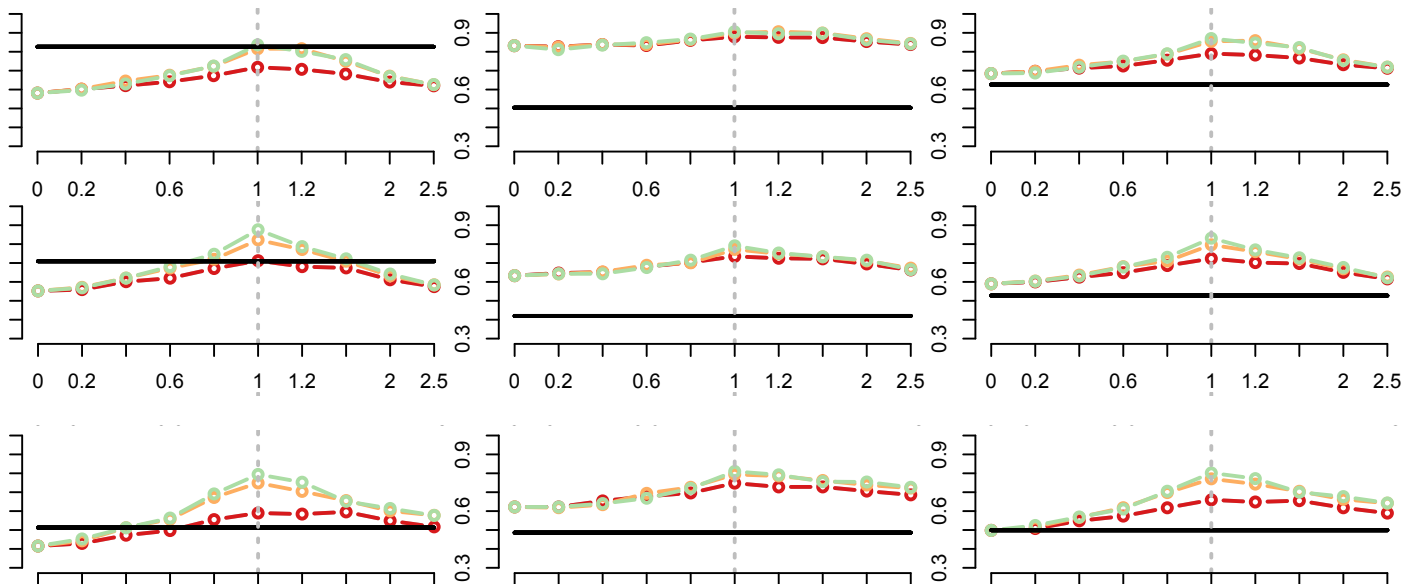
2000 mut.
5 sigs.200 mut.
5 sigs.50 mut.
5 sigs.

Figure 3C: Support recovery on simulated datasets, showing tuning the penalty weights using the prior knowledge improves performance. X-axis: fraction of true signatures given as prior (>1 indicates false signatures giving as priors). Penalty weights used: red, 0.5; yellow, 0.2; green, 0.1.

Figure 3C

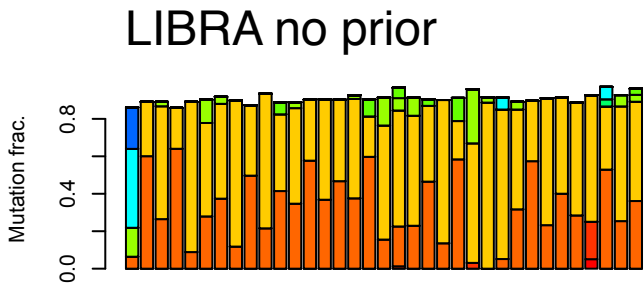
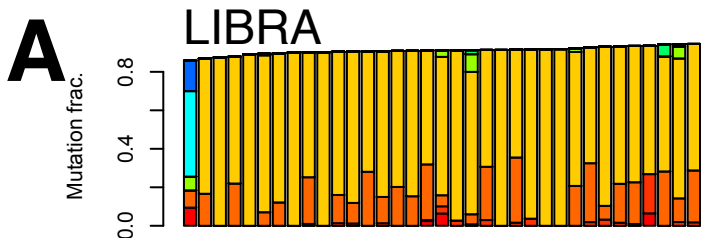
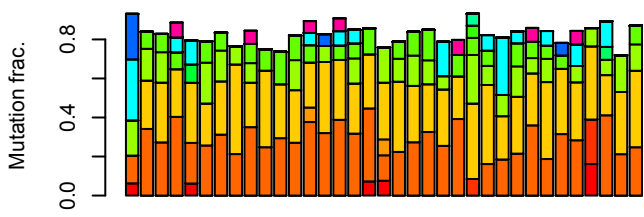


Figure 4
deconstructSigs



Only active sigs.

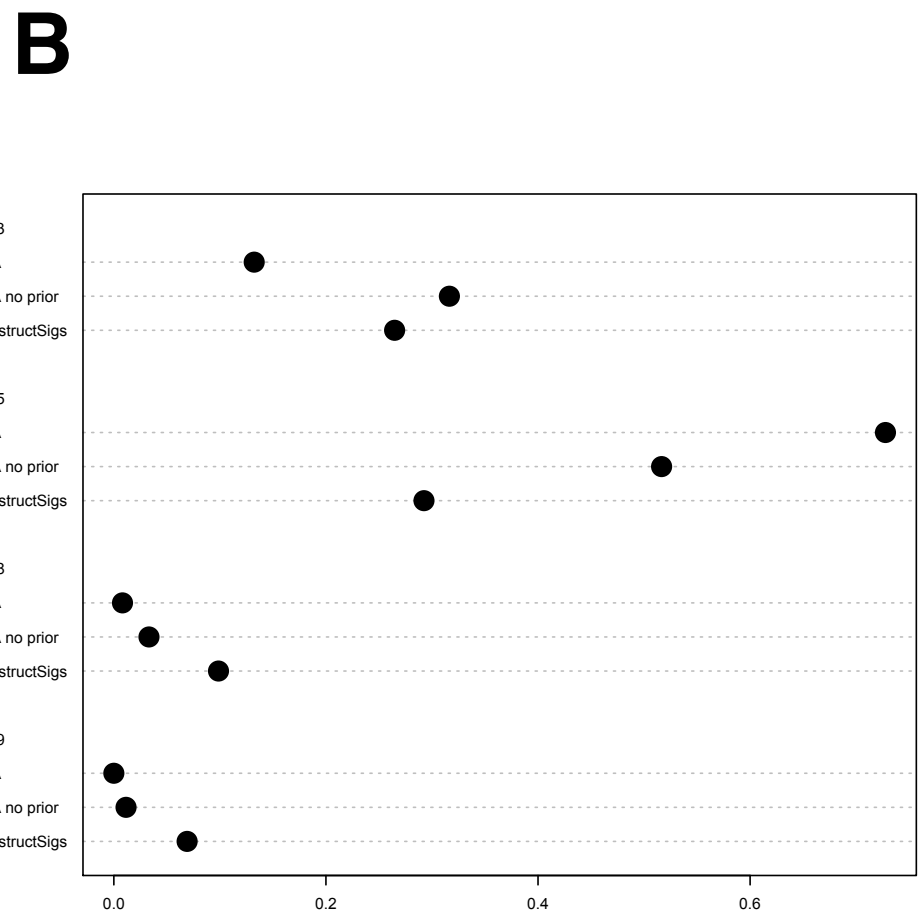
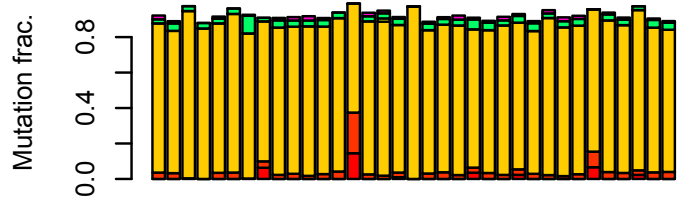
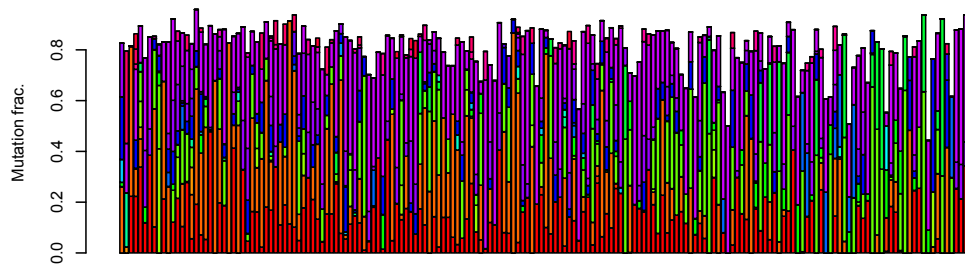


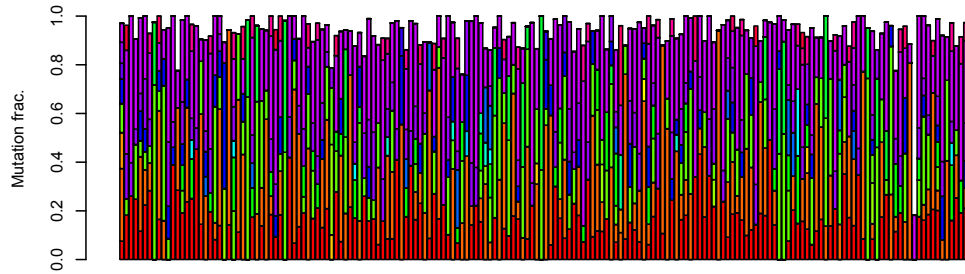
Figure 4A: Samples signature assignment for 35 WGS pRCC. Bar plots show the fractions of mutation signature assignment for each sample using LIBRA, LIBRA without prior knowledge and deconstructSigs.

Figure 4B: Dot chart showing the mean fraction of mutation signatures in each sample. Signatures contribute less than 0.05 are not shown here.

LIBRA



deconstructSigs



Total mutation

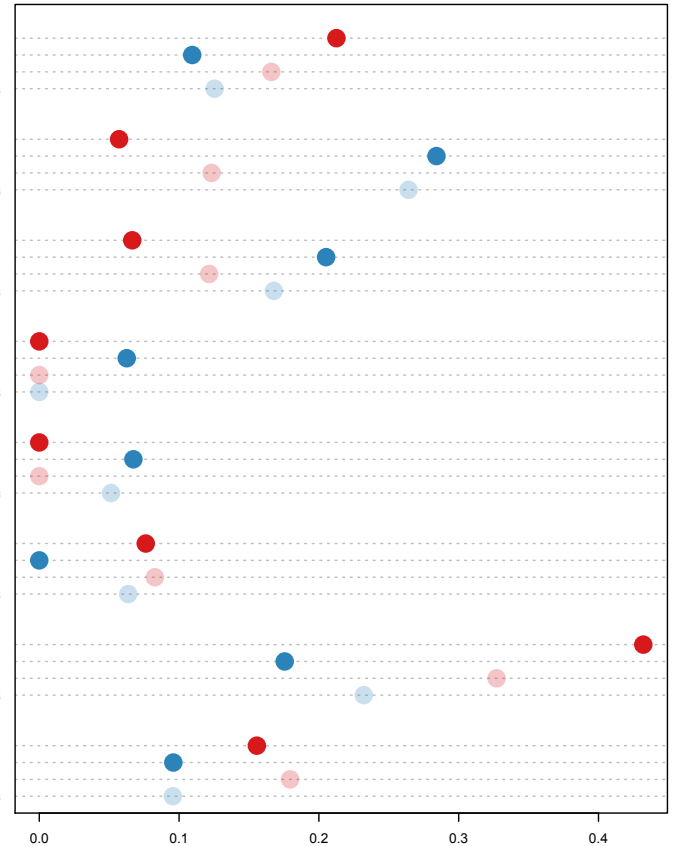
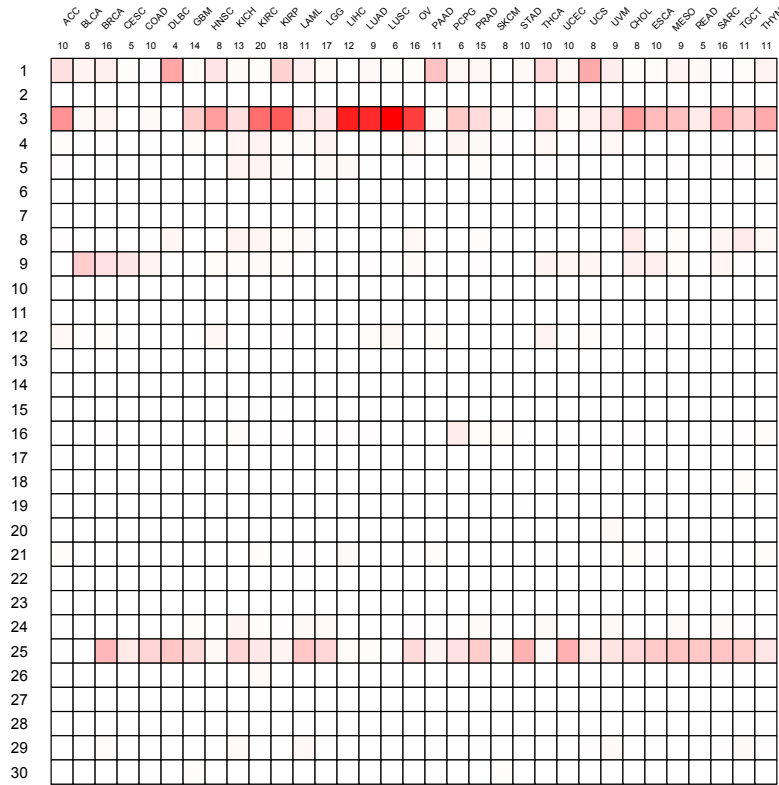


Figure 5A: Signature assignment for 182 WES ESCA samples. Bar plots show the fractions of mutation signature assignment for each sample using LIBRA and deconstructSigs.

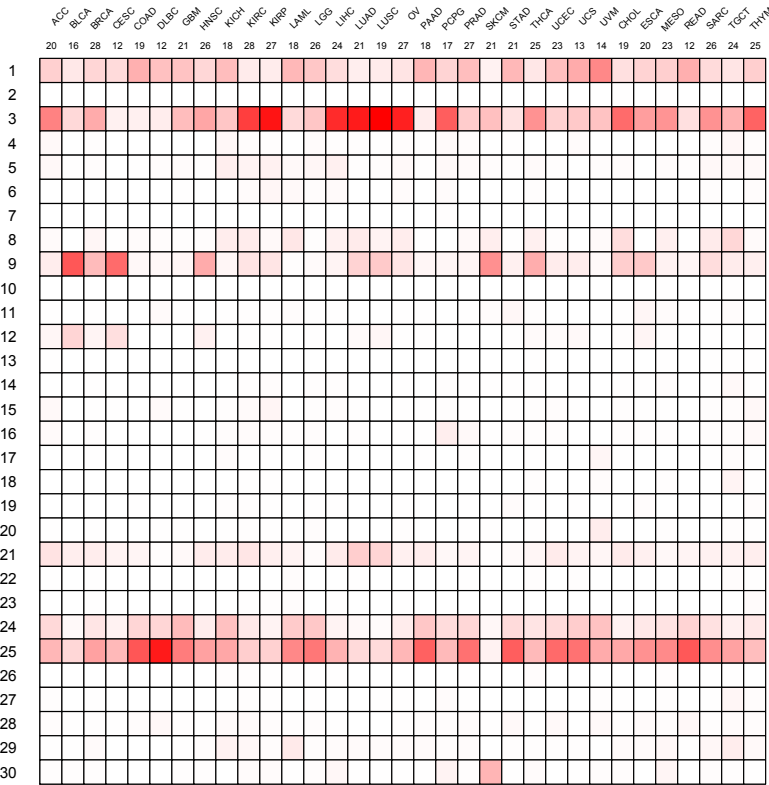
Figure 5B: Dotchart showing the mean fraction of mutation signatures in each sample, grouped by two tools and histological subtypes (adenocarcinoma/squamous). Signatures contribute less than 0.05 are not shown here.

Figure 5

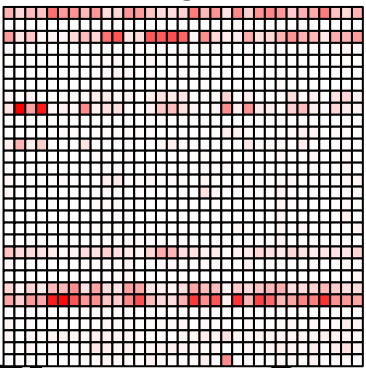
LIBRA



deconstructSigs



Simple regression



LIBRA no prior

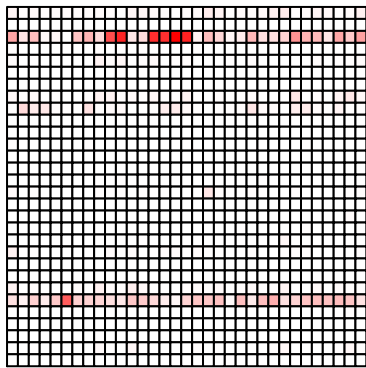
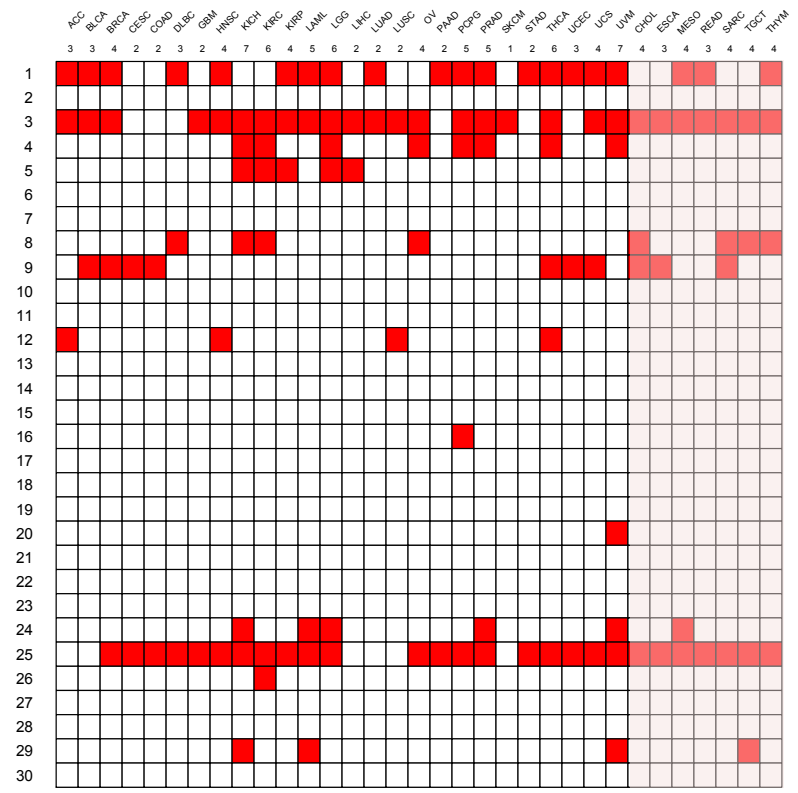


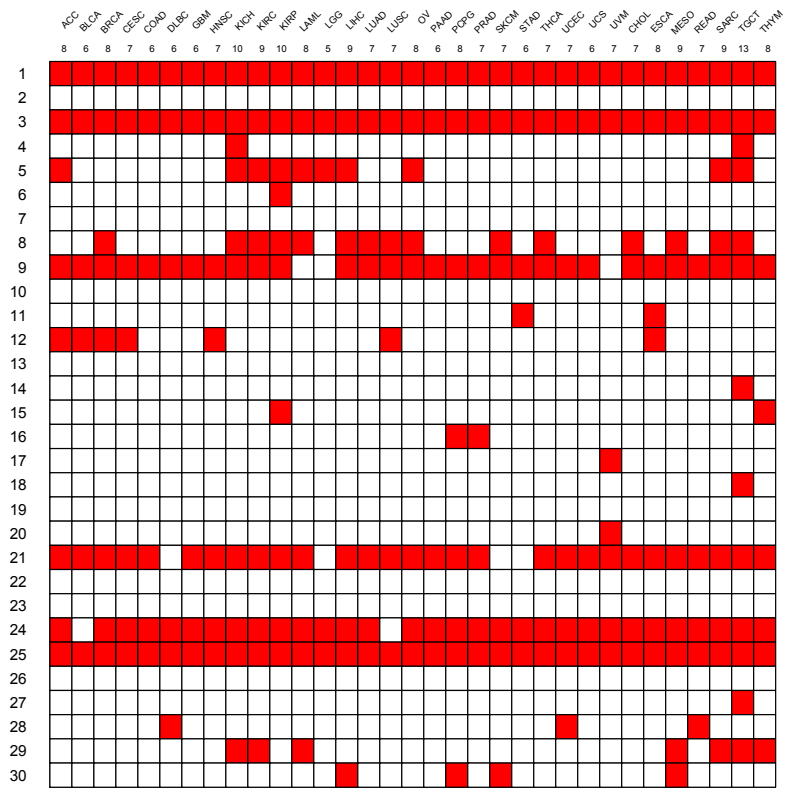
Figure 6: Mean fractions of signatures contribution of each sample in 33 cancer types. Only 26 cancer types have previously known signature distribution

Figure 6

LIBRA



deconstructSigs



Simple regression



LIBRA no prior

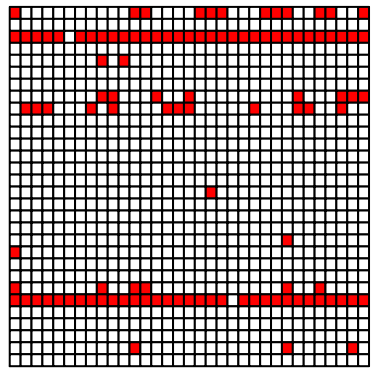


Figure 6

Figure 6: Mean fractions of signatures contribution of each sample in 33 cancer types. Only 26 cancer types have previously known signature distribution. Others are shown in shade. Signatures contribute individually less than 1% in all the samples are not shown.

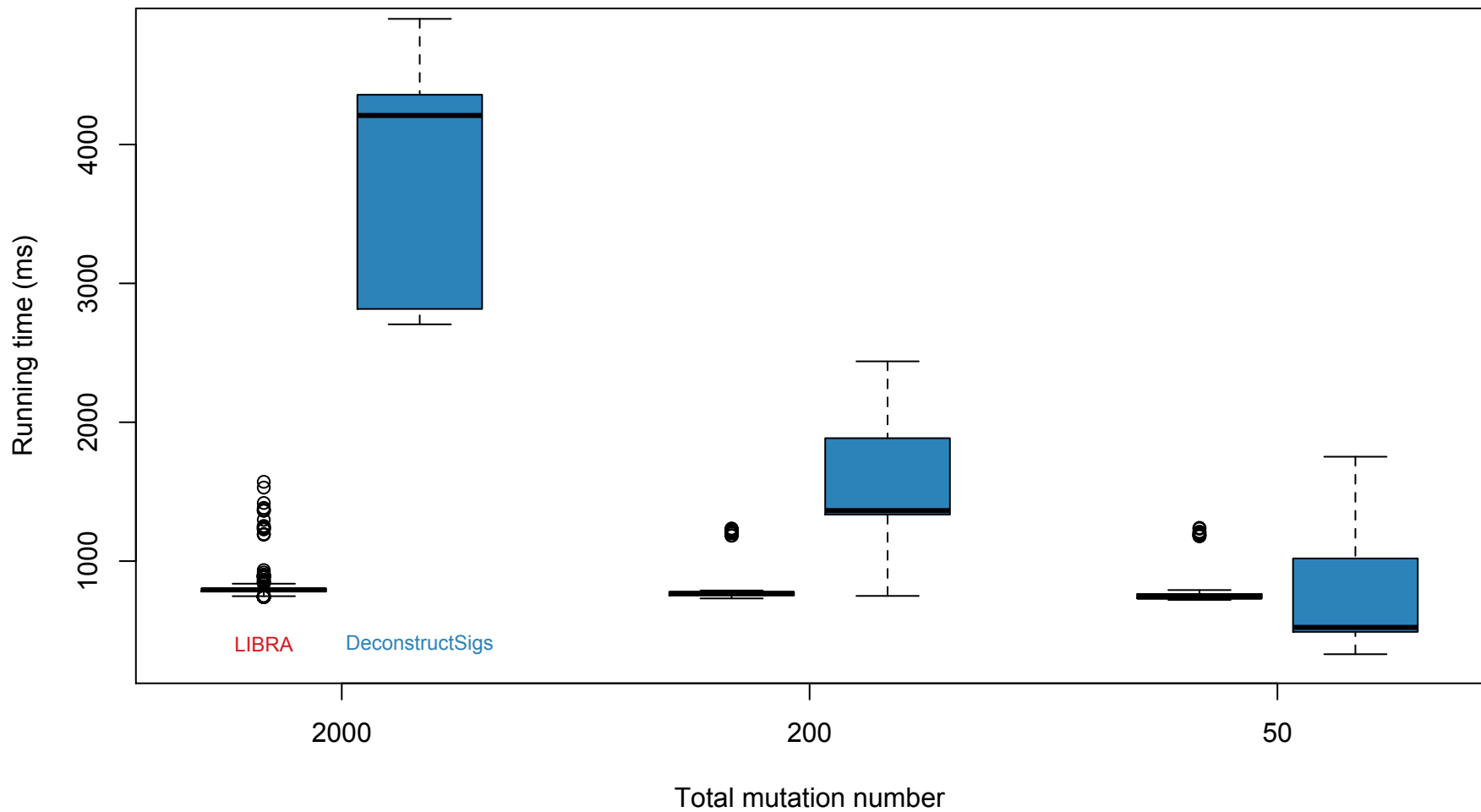


Figure 7A

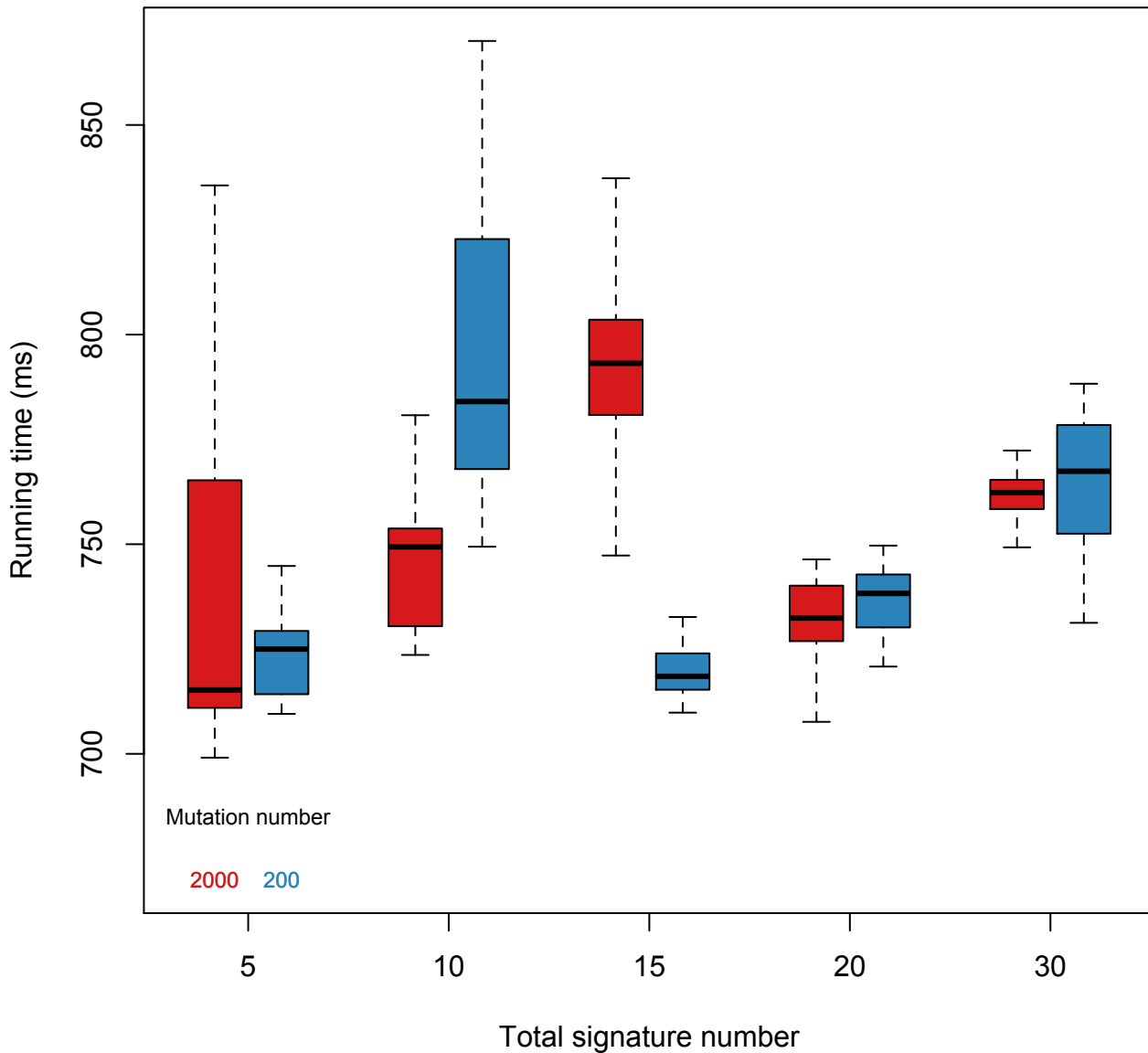


Figure 7A: Running time of LIBRA and deconstructSigs at different total mutations numbers .

Figure 7B: Running time of LIBRA at different numbers of signatures (downsampled)

Figure 7B