

Here we present supplementary data in two figures. The first one shows the comparison of different peakcallers performance in the setting where we would advise to use other tools than HERON (i.e., high signal-to-noise ratio and short peaks). The second one shows a variant of Figure 3 that is using sensitivity and specificity instead of TPR and FDR measures.

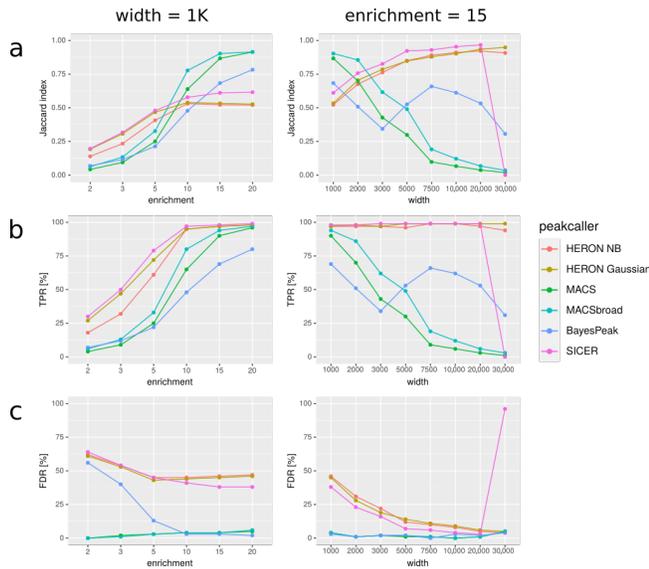


Figure 1: Comparison of peakcallers on simulated data for two different settings. In the first column, all the simulated peaks are 1K bp long, and simulated average enrichment varies. In the second column, average enrichment on peaks is constant and equal 15 and the peak length varies. In the rows three measures are shown: (a) Jaccard index:  $|\text{predicted} \cap \text{real}| / |\text{predicted} \cup \text{real}|$ , (b) TPR:  $|\text{predicted} \cap \text{real}| / |\text{real}|$ , and (c) FDR:  $|\text{predicted} \setminus \text{real}| / |\text{predicted}|$ ; where:  $|\text{predicted}|$  - summatic size of regions predicted as peaks;  $|\text{real}|$  - summatic size of real peaks, i.e., simulated ones that we wanted to discover;  $A \cap B$  - intersection;  $A \cup B$  - union;  $A \setminus B$  - difference (A and not B).

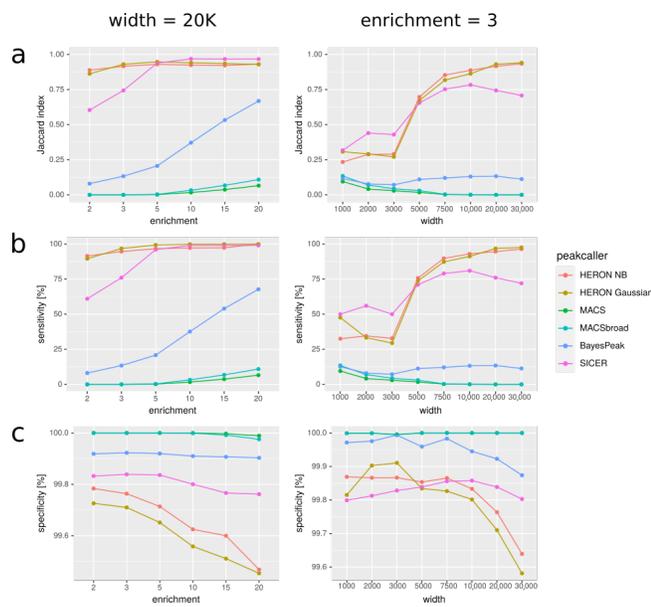


Figure 2: A version of Figure 3 presenting specificity instead of FDR. Please note the re-scaled Y axis to elucidate relatively small differences in overall high values of specificity.