mathematics

Supplementary

The process of WaveCluster

WaveCluster looks at RD of bins from a signal processing perspective. The high-frequency part of the signal corresponds to the region in the feature space where the RD of bins changes rapidly, i.e., the boundary of the cluster. The low-frequency part of the signal with high amplitude corresponds to the region in the feature space where bins are concentrated.

Wavelet clustering consists of the following four steps:

a. Quantize feature space, then assign samples to the units.
b. Apply wavelet transform on the feature space.
c. Find the connected components (clusters) in the transformed feature space.
d. Map bins in the feature space to clusters

To ensure user-friendliness, we introduced how to use waveCluster to detect CNV in GitHub in detail, and attached the corresponding test data in the release version.

1. Experiment: Simulated data with noise

We added random noise to each read depth of the sequencing data, and the read depth with noise fluctuated between 90% and 110% of the original read depth, as shown in Figure S1, and the 20 groups data with noise were detected by the six methods in the experiment. The detailed results are shown in Figure S2.

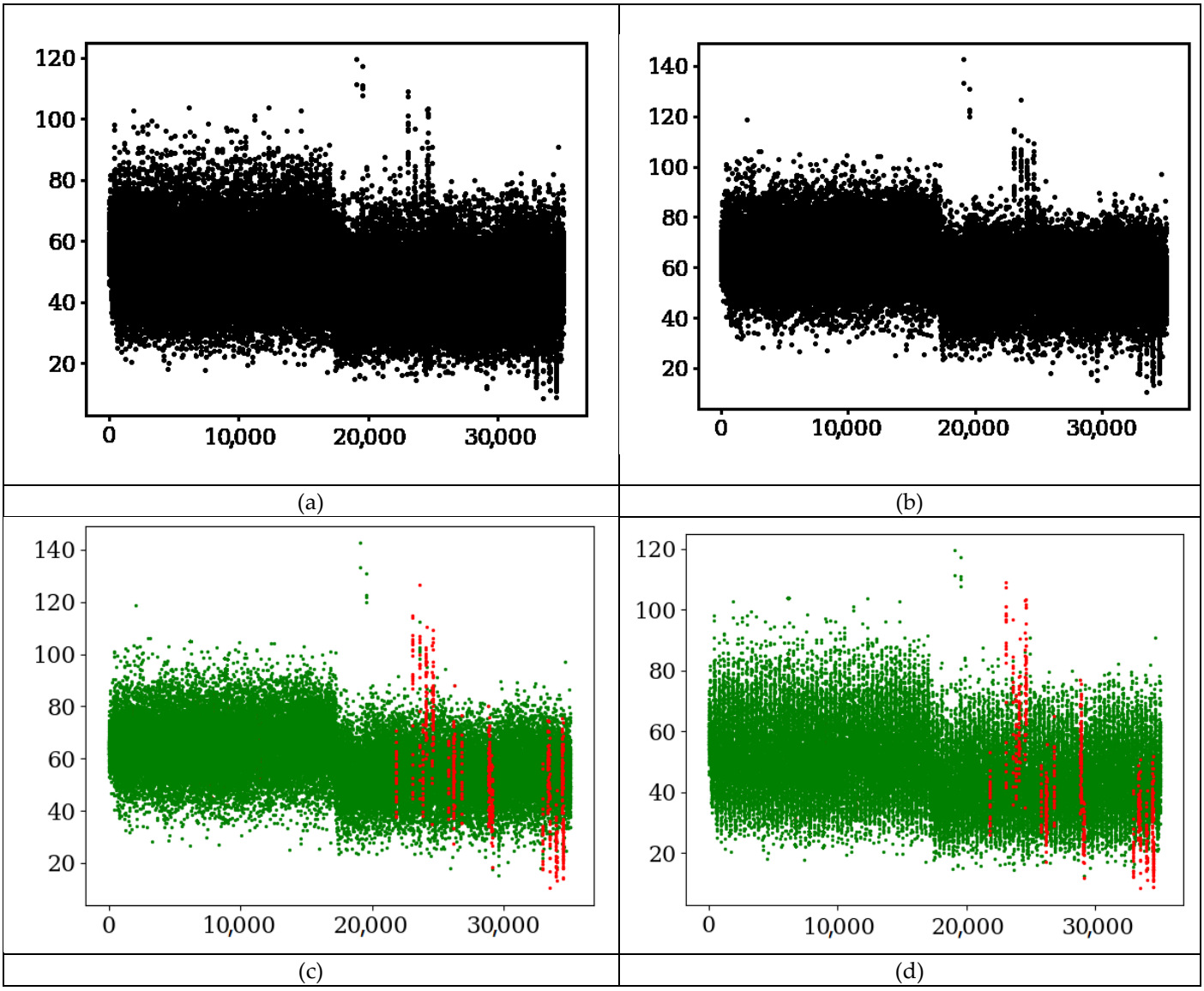
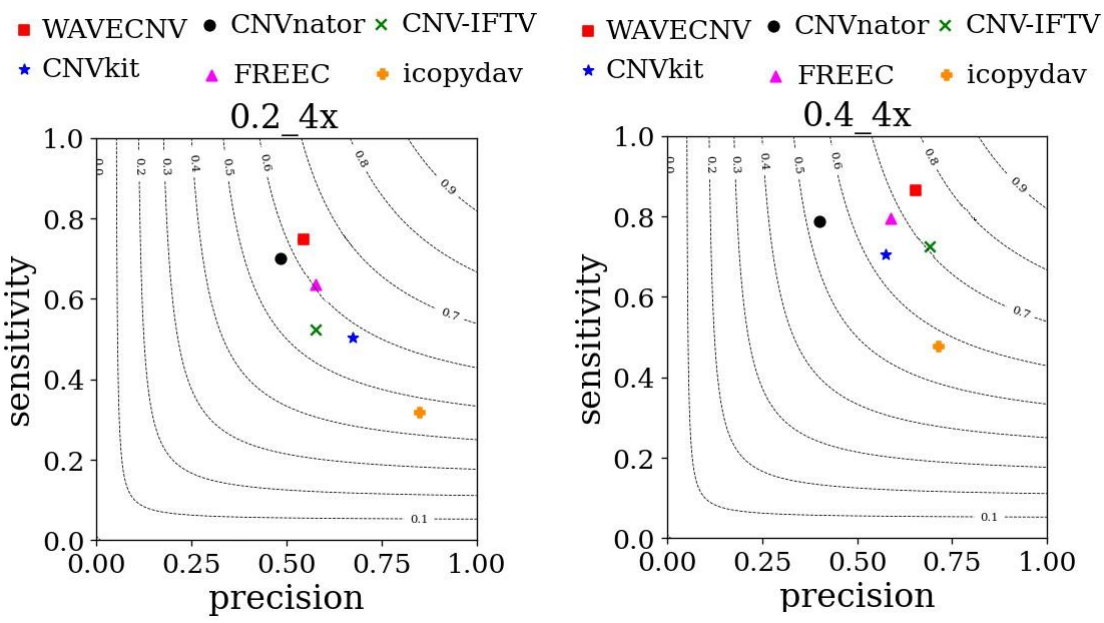


Figure S1. (a) The original sequence data. (b) The sequence data with noise. (c) The detection results of the original 313 data, green represents the normal area and red represents the CNV area (d) The detection results with noise data, green 314 represents the normal area and red represents the CNV area. The results show that the poor-quality data reduces the sensitivity and accuracy of all methods, but waveCNV still has the highest sensitivity and approximate accuracy with the existing methods among the six methods.



(a) (b)

Figure S2. The precision, sensitivity, and F1 scores of the noisy sequencing data were measured by six methods. (a) The simulated tumor concentration was 0.2 and the coverage was 4x. (b) The simulated tumor concentration was 0.4 and the coverage was 4x.