Data analysis

Friday, February 5, 2021 1:44 PM

Part. 1 Check whether or not squiggle has information to distinguishing close splice site.

- NanoSplicer accuracy (96%)

1. **Check reason of failing**
2. **S_i analysis (validate)**

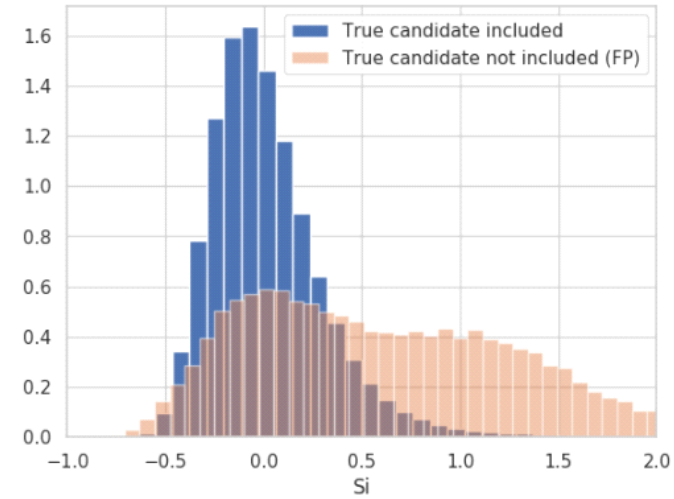
Part. 2 Sequins analysis using real data analysis pipeline (do not use transcript annotation except for the performance assessment)

- Accuracy (No S_i filtering)
  - ```
    minimap accuracy: 93.8096%
    NanoSplicer accuracy: 92.6063%
    proportion of Junction within read identified by both softeware: 91.1034%
    ```

Clarify the definition of accuracy (separate false positive and those Junction within reads within the 10nt window from the true one)

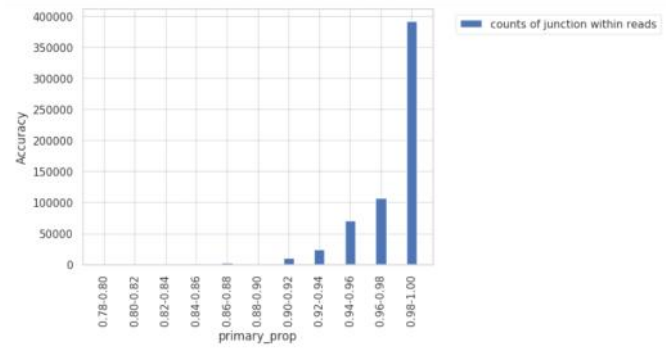
- Accuracy (With S_i filtering)
  - ```
    minimap accuracy: 94.9447%
    NanoSplicer accuracy: 94.1073%
    proportion of Junction within read identified by both softeware: 90.4494%
    ```

DO NOT filter the minimap2 result:
People would never know what to filter out before running NanoSplicer

- S_i distributionfor true positive and false positives
  - 

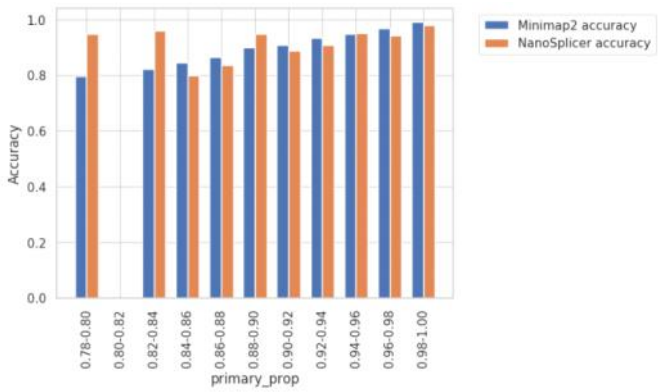
Double check the sequins annotation version. Doing analysis using v2.2?

This figure can also be used to validate Si

- Primary site proportion analysis
  - Number of junction within reads from sets with certain primary site proportion
  - 

Separate things

  - Accuracy of NanoSplicer and Minimap2


Minimap2 accuracy
NanoSplicer accuracy
Accuracy
1.0
0.8
0.6
0.4
0.2
0.0
0.78-0.80
0.80-0.82
0.82-0.84
0.84-0.86
0.86-0.88
0.88-0.90
0.90-0.92
0.92-0.94
0.94-0.96
0.96-0.98
0.98-1.00
primary_prop