Paper summary

Thursday, February 18, 2021 4:12 PM

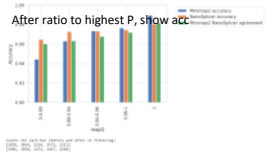
Result(Part 1) the whole part may go to supplementary

In the main text, mention what we did in this part and report the accuracy to show **Squiggle has information to distinguish close splice site**

Result(Part 2)

- **Main argument:** NanoSplicer is better than minimap2 when the basecalling/ mapping quality near the splice junctions in low **(define the mapping quality, talk with Mike, there is perhaps a better name.)**
  - It makes sense that when the basecalling/mapping quality in the JWR is relatively low, the mapped splice junction from minimap2 is less reliable, so that NanoSplicer has higher potential to get correct splice site with the help of squiggle information

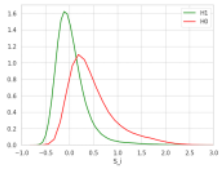
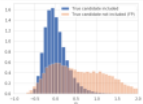
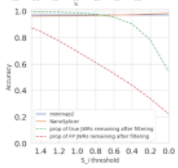
After ratio to highest P, show acc



Also show the False positive rate in the same bins and mention S_i is helpful for False positives identification

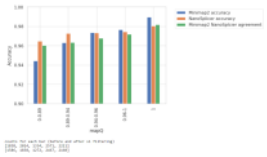
- **Example of JWR with low mapQ to show squiggle information is helpful when mapQ is bad.**
- **Examples of why the mapQ is high (bin mapQ = 1) but NanoSplicer make it wrong**
  - When the true one identified as the second likely one, show how close the likelihood is.
  - When NanoSplicer completely missed the true one (with huge difference in likelihood), check the **distinguish point.**

Supplementary:

- Empirical distribution of S_i-> appropriate threshold for S_i
  - 
- Show how is the result is sensitive to S_i threshold selection
- How S_i can distinguish FP and truth known
  - 
  - 

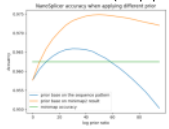
Result(Part 3)

- **Main argument: Conclusion in part2 also works on real data)**
  - (How to frame this part depends on the results of junction mapQ analysis)



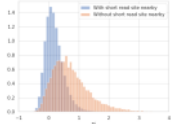
Also show the False positive rate in the same bins and mention S_i is helpful for False positives identification

  - **Adding sequence prior in real data analysis is helpful (drop minimap2 prior)**
    - The sequence prior has been implemented in minimap2 already
    - The pattern is conservative (site paper)



  - **Quantification** (depends on whether or not get good result):
    - Categorized by mapQ,
    - and use seq prior
    - and find examples

Supplementary:

- Empirical distribution of S_i-> appropriate threshold for S_i
  - Distribution of FP and Truth known JWR in real dataset
    - 
- **Example of JWR with low mapQ to show squiggle information is helpful when mapQ is bad.**
- **Examples of why the mapQ is high (bin mapQ = 1) but NanoSplicer make it wrong**
  - When the true one identified as the second likely one, show how close the likelihood is.
  - When NanoSplicer completely missed the true one (with huge difference in likelihood), check the distinguish point.
- Empirical distribution of S_i-> appropriate threshold for S_i
- How S_i can distinguish FP and truth known