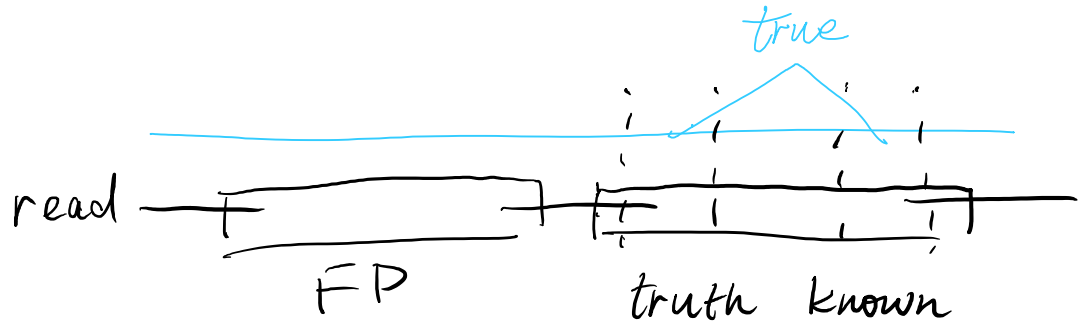


Feb 10 2021 Meeting

Wednesday, February 10, 2021 10:35 AM

1. Separate FP cases and truth known cases

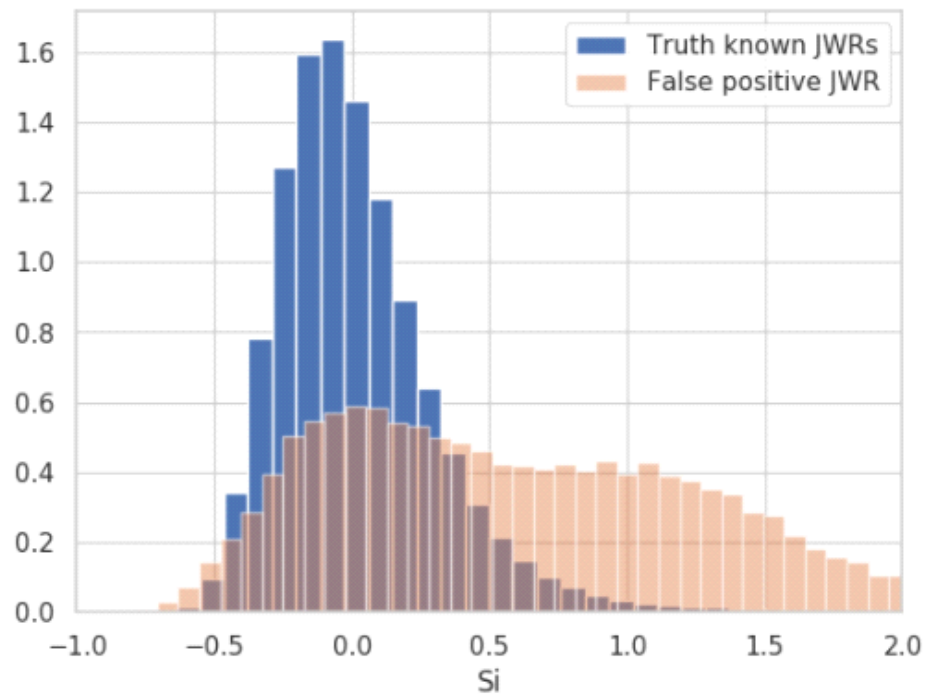


- For each reads, the true isoform is known so that the true splice site within each read can be known.
- In the analysis, junction within reads(JWR) a divided into 2 categories: **False Positive** and **Truth known**. For JWR that mapped within 10nt from both side of true splice junctions will be identified as **Truth known JWR** (613790 96.48%), **otherwise False Positive JWR** (22371 in total (3.52%))
- Accuracy for Truth known JWR:

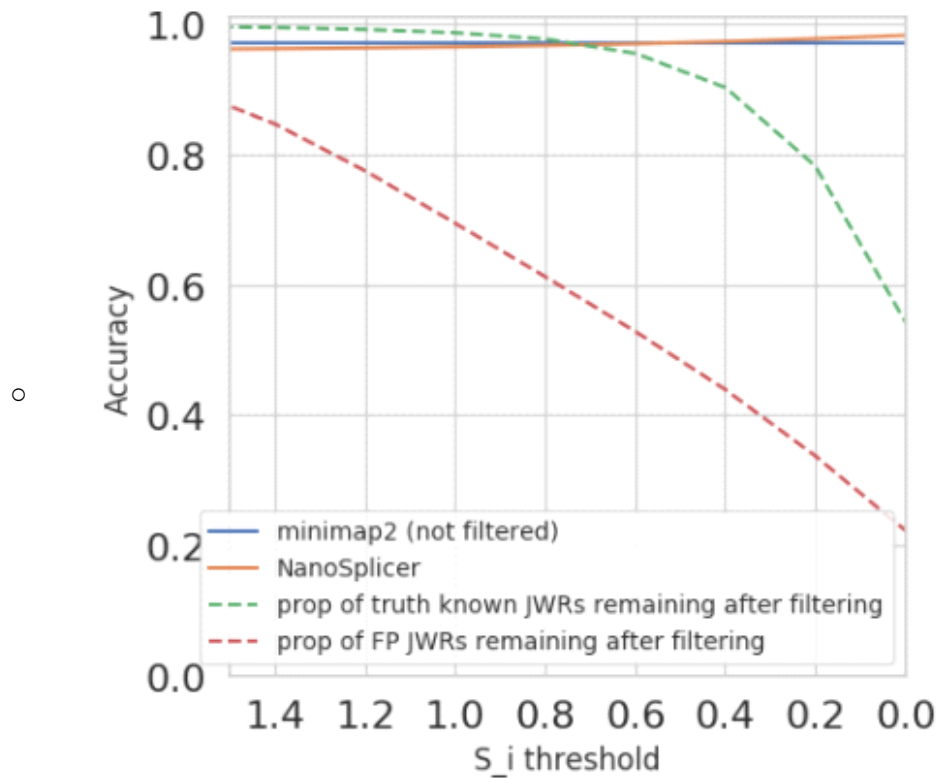
```
minimap accuracy: 97.2287%
NanoSplicer accuracy: 95.9815%
proportion of Junction within read identified by both software: 94.4238%
```

2. S_i analysis

- Si distribution for False Positive JWR and Truth known JWR



- Apply different S_i threshold
 - Minimap2 result are not filtered, so that it's a flat line
 - Accuracy is calculated on Truth known JWRs only
 - The accuracy are the same at $S_i = 0.45$, with 8.0% of truth known JWR filtered out, and 53.8% of False Positive JWR filtered out.



2.1 S_i analysis (figure out the bimodal distribution of FP JWR)

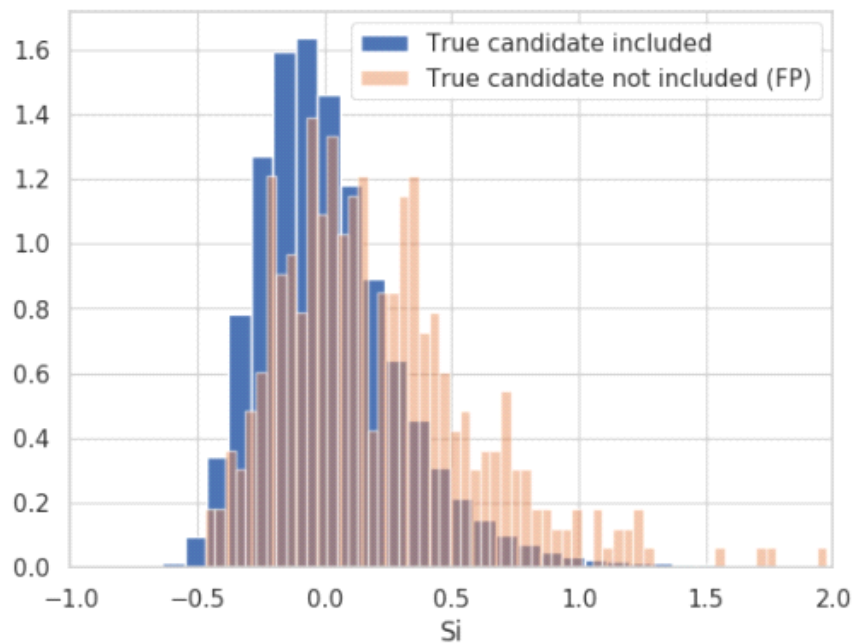
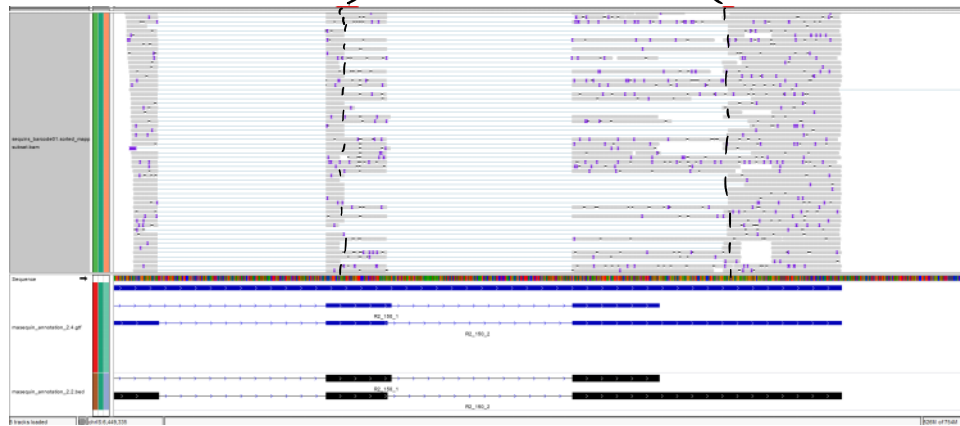
I manually checked some False positive JWR set and find some typical categories

- Junction mapping quality seem to be okay but not in annotation.



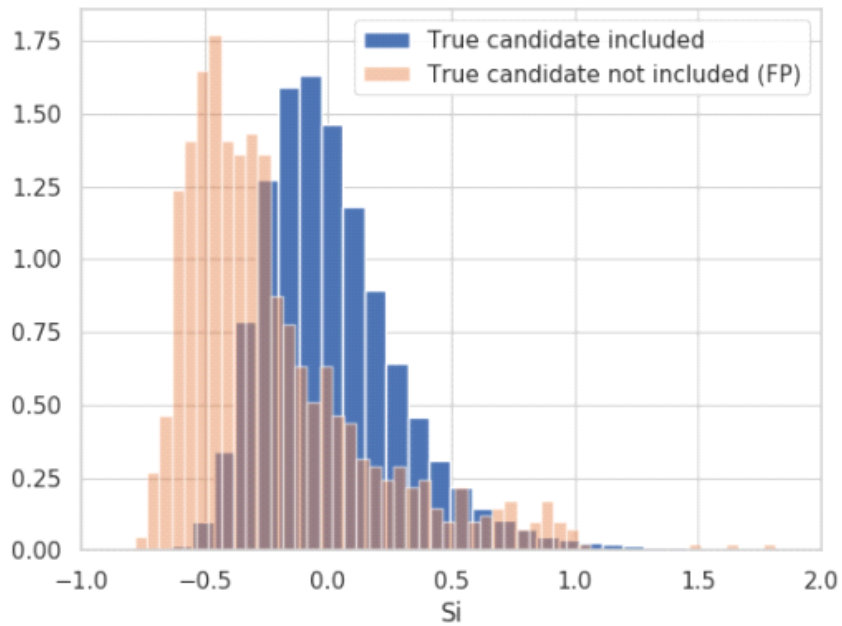
I manually checked some false positive JVK set and find some typical categories

- Junction mapping quality seem to be okay but not in annotation.



- Guppy fail to trim poly-A: Poly-A tail doesn't exist in the fastq file for most of the reads, but we can still find some reads come with poly A (planning to manually filter out them).





- False Positive introduced when minimap2 failed to identify small exons

