Paula Berry

Deduper Pseudocode comments

Abbie Olsen:

Your code is very well organized and easy to read. One thing I noticed was that all of the argparse arguments were required, when in the assignment instructions several of them are listed as optional for the user.

I would have liked to see what arguments your functions would accept, instead of just what values they returned.

In the “read\_parser()” function I’m not sure you have to account for the soft clipping at the start of the CIGAR string if the strand on the minus side, because you will be counting using only the matched things anyways.

During the iterate section of code, you refer to different columns as being on different lines – will you be splitting the line into different lines and then rejoining it for writing to the deduped file again? I was unclear on exactly what was happening to the different bits of each line in the SAM file in this part.

Nicole Tran:

Your code is very easy to read! I noticed in the “Parse QNAME” function you call out the list element by counting from the beginning. (“QNAME[7]”) it might be safer to use -1 instead so you know you’re always getting the UMI at the end no matter how many extra elements the QNAME string has.

in def\_forward() you call a “ “ an empty string, but it looks like there’s a space in there. Python will not think of that as an empty string.

I had a little trouble following the sequence in the algorithm, I think you are going through all of the records, or leap-frogging over records? I at first thought leap-frogging was the way to go but apparently it is not, all the records with one chromosome need to be compared to all the other records with that chromosome.

Annie Wang:

Your markdown file is so easy to read! I liked how detailed you were with how to check the headers and what they meant and what to do with them. The actual algorithm was a little sparse, I’m not sure how you are planning on sorting through all the of the file to make sure every line that needs to be compared is compared.