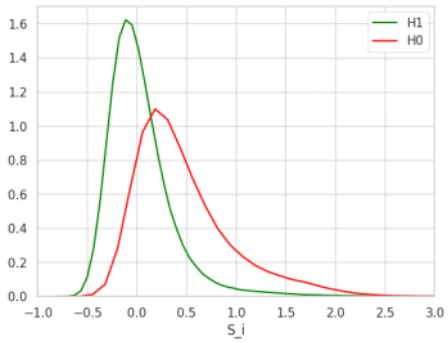
S_i empirical distribution

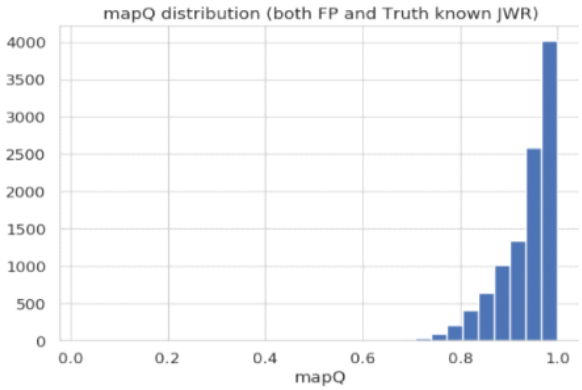
Monday, February 22, 2021 1:38 PM

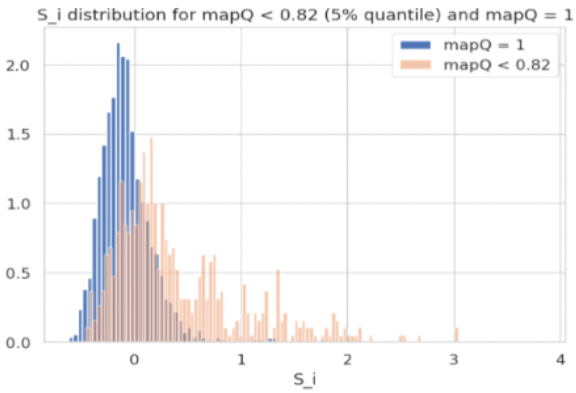
S_i distribution based on mapping quality

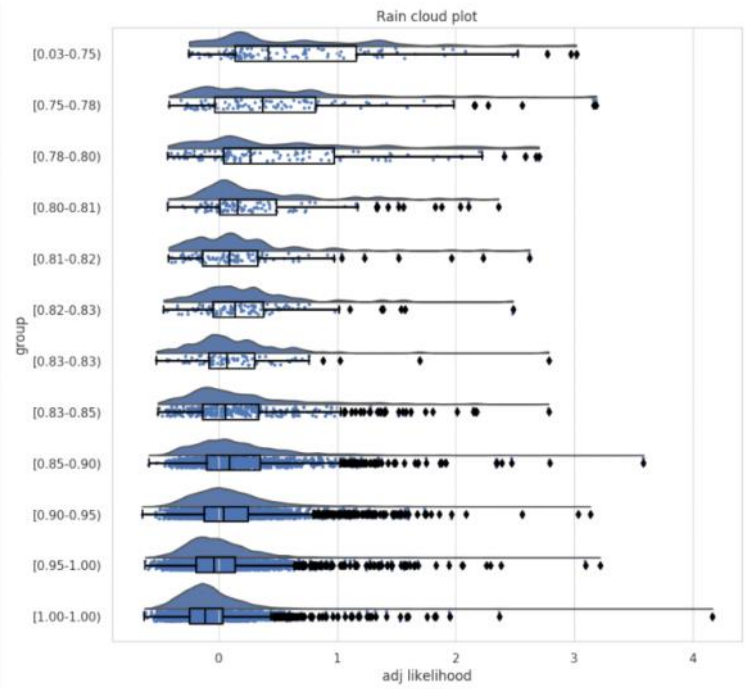
The previous empirical distribution for H1 and H0 is as follow, we think it could be more separate



So we are trying to use the mapQ by assuming that when the junction mapping quality is bad, it's more likely the mapping is wrong and the true candidate is not included.

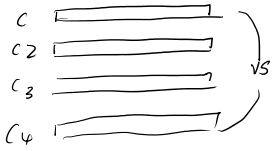


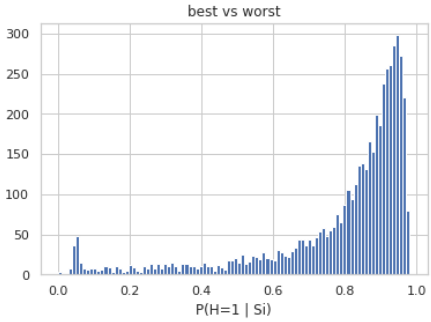


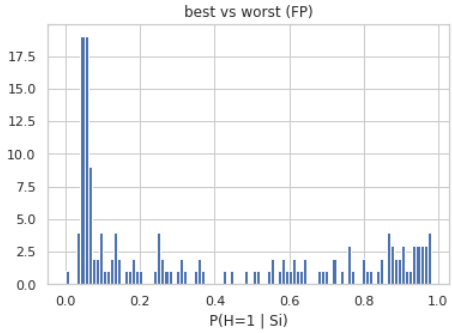


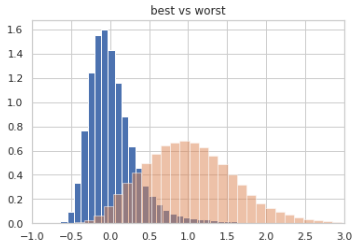
Different way of generate S_i (sequins data)

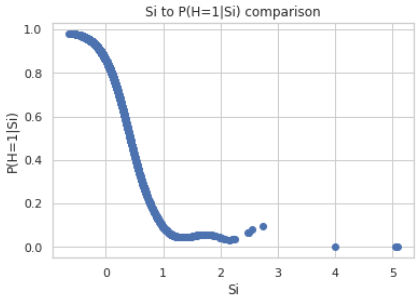
Candidate (from best to worst

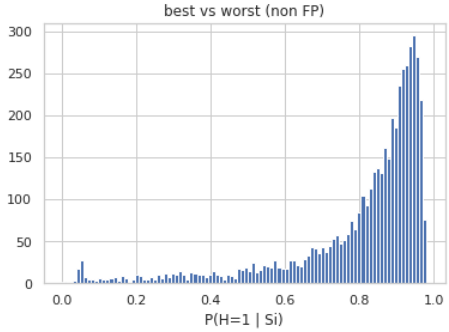










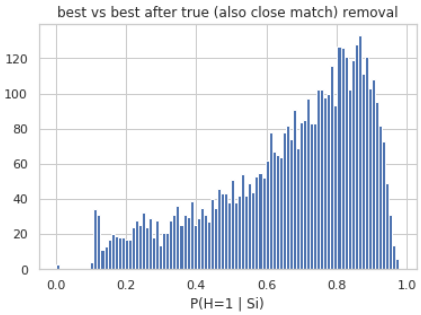


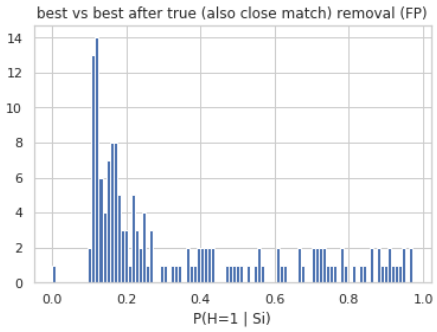
vs { C1, C2 } close matches, difference in adj_logL < 0.22

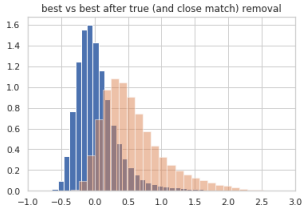
C3

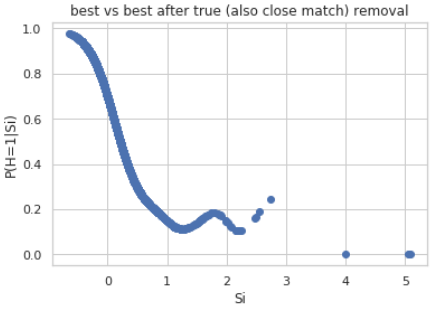
C4

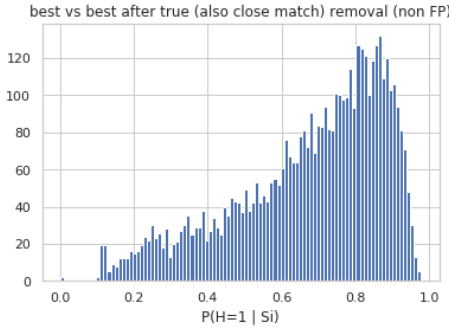
C5

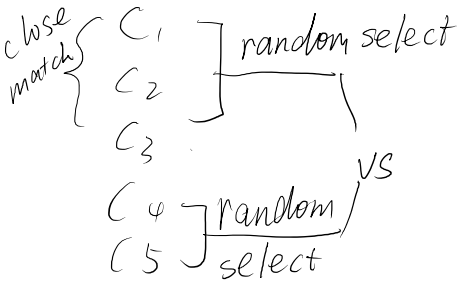


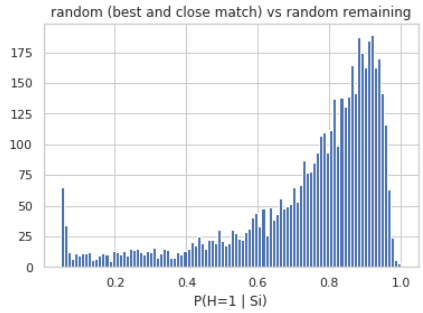


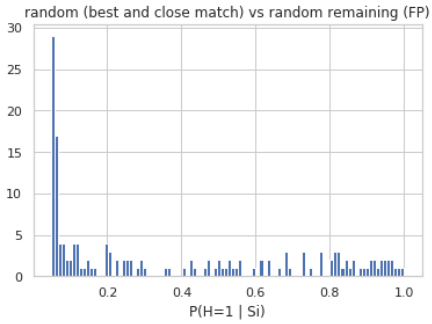


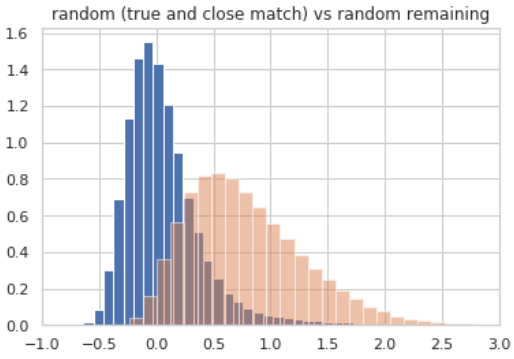


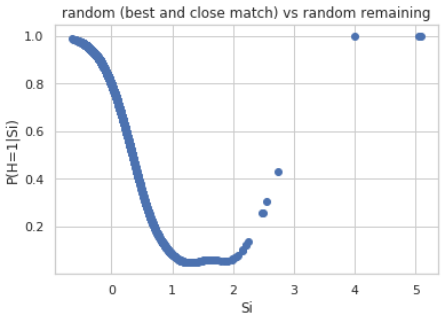


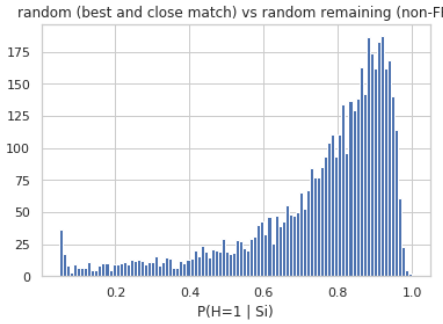

close match
C1
C2
C3
C4
C5
random select
random select
VS


random (best and close match) vs random remaining
P(H=1 | Si)


random (best and close match) vs random remaining (FP)
P(H=1 | Si)


random (true and close match) vs random remaining


random (best and close match) vs random remaining
P(H=1|Si)
Si


random (best and close match) vs random remaining (non-FP)
P(H=1 | Si)