1.7.2025

* Máme vždy různý počet readů, nejjednodušší by bylo mít prostě dvojice a s tím pak pracovat, ale každý read máme přečtený různěkrát
* Problém ohledně bulk seq
  + Ready mají umi, ale nevíme, jestli dva ready s různými umi jsou ze dvou sesterských buněk, které byly původně ze stejné buňky, nebo jestli to jsou ready z jedné buňky (protože RNA máme hodně kopií a můžeme osekvenovat víc RNA)
* Simulace – monte carlo, nulová hypotéza jak by to vypadalo, kdyby tam nebyl commitment, mít jednoduchý parametr, že když to nasimulujeme tisíckrát a zjistíme že takhle se chová jedna simulace ze 20 (naše p value)
  + Pro každý receptor náhodně vybrat počet vzorků ve kterých se vyskytuje a pracovat s nulovou hypotézou
  + Zkontrolovat tak, aby nám seděly součty na konci řádků, pak udělat součty na konci sloupců
  + Pro každou simulaci udělat průměr a simulaci udělat 1000 a porovnat výsledky
* Součet všech readů se dá použít jako denominator
* Stop codon značen \*
* Zchecknout jestli homozyg nebo heterozyg
* Jestli na rag knockoutu nebo ne, endogenní
* Vyfiltrovat neproduktivní sekvence
* Tabulka, kde 0 seq není, 1 seq je, vyfiltrujeme
* Cíl: průměrný commitment, 50% si buňka hodí korunou, 100% že je determinovaná
* H0: Každá buňka má 50% že nemá určený fate
* Druhá extrémní hypotéza že má 100%
* 1. věc u každého receptoru dát ohodnocení jak moc dobře ty buňky commitují
  + Horní odhad na procenta na kolik se ty buňky fakt trefí na ten dominantní fenotyp (domin fenotyp se vyhodnocuje na základě těch dat)
  + Vážený průměr – čím větší počet jedniček, tím
  + Sloupec s váhami, sloupec s dominantním fenotypem
* 2. věc Kvantifikovat kolik receptorů se opakuje jen v rámci skupiny
  + Zjistit počet čistě CD8
  + Zjistit počet čistě CD4
  + Zjistit počet průniku
  + Seznam
* Data zda buňka je CD4 či CD8, ideálně periferie a vím její alpha a beta (musí být spárované), ideálně z myši která nemá knockout, hledat buď spárované alpha a beta v bio vzorcích nebo mít jeden řetěz zafixovaný (např Vb5) natvrdo a pak se sekvenuje jen jeden řetězec
* Udělat vzorky i pro betu
  + Vyhozeno DN a DP
* Barplots
* Meeting s Jurajem ohledně machine learningu – natrénovat algoritmy na Jurajově datech a vyhodnotit jestli to sedí na mých datech
* Alignment různých sekvencí na použití machine learningu – mezery na konci sekvencí aby byly stejně dlouhé