Meeting 13.8.

* Dva modely
  + Nějaký poměr buněk determinovaných a nedeterminovaných, nějaké procento
  + Nebo můžeme hledat pravděpodobnost toho, že buňka udělá chybu, předpokládáme 100% commit
* Co budeme potřebovat
  + Kolik procent je commitovaných jak
  + Znát kolik ze sekvencí je CD4, kolik CD8
* Pokud commitment bude 50%, pak při 10 cd4 a 5 cd8 je ta pravděpodobnost 2/3 pro cd4 a 1/3 pro cd8, pravděpodobnost pro každou buňku je 1/15
* Pokud ten commitment bude 80%, pak pravděpodobnost pro každou buňku můžeme udělat pomocí vah
  + Pro cd4 0,8\*1/15\*10 = 0,533
  + Pro cd8 0,2\*1/15\*5 = 0,067
  + Pravděpodobnost pro cd4 0.533/(0.533+0.067)=0.89
* Co by se stalo, kdybychom řekli, že tam máme jen cd4 sekvence
  + Simulace
* Podívat se do paperu
  + Jak byly buňky sortované
* Nejde nám o biologický poměr mezi cd4 a cd8 sekvencemi, nám jde o to, jaký poměr je tady v našich datech
* První věc – řekneme si, že jsou všechny cd4, sestavit rovnici pro pravděpodobnost \* počet všech sekvencí
* Druhá věc – řekneme procento tomu modelu, vezmeme poměr unikátních cd4 a unikátních cd8
* Můžeme buď
  + Říct že všechny jsou cd4
  + Nastavit si hodnotu fixně a fitovat jen jeden parametr
  + Fitovat k tomu i ten poměr cd4 a cd8
* Nefitovat ten průměr, fitovat ty tři stricly cd4, strictly c8
* Udělat subset té tabulky jen na sekvence s occurrence 2