Ahoj, mám otázku ohledně těch umi.Juraj mi doporučil zkusit začít nejdřív s daty, které mají na githubu zveřejněné a už namapované. Tam mají tabulku s jednotlivými sekvencemi (sloupec $CDR3.nucleotide.seq) a s cell type (sloupec $cell) - viz první obrázek.Nejsem si jistá ale, co se myslí tím $Umi.count. Znamená to, že jakmile je umi.count větší než 1, tak je to ta hledaná neunikátní sekvence, protože byla na více různých umi? Nebo mám hledat výskyty v tabulce pod tou sekvencí a to co se bude vyskytovat víckrát je neunikátní sekvence?

2 files



ondrejondrej  [1:17 PM](https://labofadaptiveimmunity.slack.com/archives/D064ZN1SNHZ/p1738585073902669)

Ahoj Belo. Asi se zeptej Juraje jeste, ale podle me presne UMI count znamena, kolikrat to chytli.

[1:18](https://labofadaptiveimmunity.slack.com/archives/D064ZN1SNHZ/p1738585082356909)

ondrej

Tzn. UMI count 1 = unikatni

[1:18](https://labofadaptiveimmunity.slack.com/archives/D064ZN1SNHZ/p1738585104833099)

ondrej

UMI count >1, vyskytlo se ve vzorku vicekrat

[1:19](https://labofadaptiveimmunity.slack.com/archives/D064ZN1SNHZ/p1738585190861059)

ondrej

Takze predpokladam, ze tabulka vlevo je vzdy pro kazdy vzorek. A tabulka vpravo je uz z ruznych vzorku dohromady. (edited)



BělaBěla  [1:21 PM](https://labofadaptiveimmunity.slack.com/archives/D064ZN1SNHZ/p1738585318736429)

Oba obrázky jsou ze stejné tabulky, ta první je nefiltrovaná (udělala jsem to tak, abys viděl všechny jména sloupců najednou) a druhý obrázek je ta samá tabulka filtrovaná na konkrétní náhodně vybranou sekvenci (ta která je modře vyznačená)



ondrejondrej  [1:22 PM](https://labofadaptiveimmunity.slack.com/archives/D064ZN1SNHZ/p1738585340319589)

Ok, ale tim padem jsou ze vsech vzorku dohromady.

[1:22](https://labofadaptiveimmunity.slack.com/archives/D064ZN1SNHZ/p1738585347403269)

ondrej

Obe.



BělaBěla  [1:23 PM](https://labofadaptiveimmunity.slack.com/archives/D064ZN1SNHZ/p1738585396369019)

Taky myslím, to je podle mě ten sloupec "Sample.bio.name"

[1:23](https://labofadaptiveimmunity.slack.com/archives/D064ZN1SNHZ/p1738585401350659)

Běla

image.png

[1:23](https://labofadaptiveimmunity.slack.com/archives/D064ZN1SNHZ/p1738585412273269)

Běla

Takhle to vypadá, když to nefiltruju a prostě to jen otevřu



ondrejondrej  [1:23 PM](https://labofadaptiveimmunity.slack.com/archives/D064ZN1SNHZ/p1738585415118499)

Ano, akorat v to, jak to mas vpravo

[1:23](https://labofadaptiveimmunity.slack.com/archives/D064ZN1SNHZ/p1738585431244959)

ondrej

tak tam se zda, ze dve radky maji stejny bio name

[1:23](https://labofadaptiveimmunity.slack.com/archives/D064ZN1SNHZ/p1738585435176599)

ondrej

coz je divne

[1:24](https://labofadaptiveimmunity.slack.com/archives/D064ZN1SNHZ/p1738585478931319)

ondrej

Ale trochu to vypada, ze jeden konci na I a jeden na i (tecka videt neni, ale je to kratsi - zkus zjistit, jak to je - to by mohla byt nejaka chyba)

[1:25](https://labofadaptiveimmunity.slack.com/archives/D064ZN1SNHZ/p1738585503829199)

ondrej

Ale na 95% to podle me je tak, ze to UMI je pocet opakovani v ramci jednoho vzorku

[1:25](https://labofadaptiveimmunity.slack.com/archives/D064ZN1SNHZ/p1738585534639829)

ondrej

Proto stejnou sekvenci muzes mit vicekrat, pokud se opakuje ve vice vzorcich.

:+1:1

[1:26](https://labofadaptiveimmunity.slack.com/archives/D064ZN1SNHZ/p1738585573424629)

ondrej

Zajimaji nas takove sekvence, ktere se opakuji BUD v ramci jednoho vzorku A/NEBO v ramci vice vzorku (klidne muze mit UMI 1 v obou).

:+1:1



BělaBěla  [1:26 PM](https://labofadaptiveimmunity.slack.com/archives/D064ZN1SNHZ/p1738585619513299)

image.png

[1:27](https://labofadaptiveimmunity.slack.com/archives/D064ZN1SNHZ/p1738585635581459)

Běla

Máš pravdu, bylo to i a l



ondrejondrej  [1:27 PM](https://labofadaptiveimmunity.slack.com/archives/D064ZN1SNHZ/p1738585639932079)

ok



BělaBěla  [1:33 PM](https://labofadaptiveimmunity.slack.com/archives/D064ZN1SNHZ/p1738586017835769)

No a když budeme teda pracovat u jednoho vzorku s jedním řádkem, neochudíme se o informaci ohledně proporce CD4 a CD8? Jakože tím, že je tam např umi.count bude 200, znamená to, že na 100 různých umi (teoreticky různých buněk) máme jen jednu informaci o cell type? Neochudíme se tím o tu proporci mezi CD4 a CD8? Nebo ta pravděpodost, že to umi je prostě jen stejná buňka nebo klon je tak velká, že je to zanedbatelné?



ondrejondrej  [3:18 PM](https://labofadaptiveimmunity.slack.com/archives/D064ZN1SNHZ/p1738592289409669)

Ono by bylo fajn pracovat u kazdeho vzorku s totalnim poctem UMI

[3:18](https://labofadaptiveimmunity.slack.com/archives/D064ZN1SNHZ/p1738592305735719)

ondrej

takze budeme vedet, kolik procent UMI dela ten konkretni klon klonotyp v kazdem vzorku (edited)

[3:20](https://labofadaptiveimmunity.slack.com/archives/D064ZN1SNHZ/p1738592408758709)

ondrej

Takze nebudeme pracovat s tim, ze ten klon je nebo neni, ale kolik procent toho vzorku ten klonotyp dela.

[3:20](https://labofadaptiveimmunity.slack.com/archives/D064ZN1SNHZ/p1738592456118969)

ondrej

Nejdulezitejsi jsou pro nas ty klonotypy, ktere se opakuji mezi vzorky

[3:21](https://labofadaptiveimmunity.slack.com/archives/D064ZN1SNHZ/p1738592488888099)

ondrej

A pak podle tech proporci budeme pocitat, jestli s vice shoduji mezi CD4 a CD4 vzorky nebo mezi CD4-CD8 nebo mezi CD8-CD8



BělaBěla  [3:33 PM](https://labofadaptiveimmunity.slack.com/archives/D064ZN1SNHZ/p1738593226614389)

Tedy to znamená, že potřebuju teda ke každému vzorku přidat celkový počet UMI co má a z toho vzít procento, které tvoří ta umi ke každé jeho zachycené sekvenci?



ondrejondrej  [3:35 PM](https://labofadaptiveimmunity.slack.com/archives/D064ZN1SNHZ/p1738593315125569)

jj



BělaBěla  [3:37 PM](https://labofadaptiveimmunity.slack.com/archives/D064ZN1SNHZ/p1738593440895499)

Jak bys to potom chtěl vyzualizovat? To jestli se shodují ty klonotypy napříč vzorky? Vzít to jako plot dle frekvence CD4 a tím pádem u 1 budou klony čistě na CD4, u 0 čistě na CD8 a někde mezi mix?

[3:37](https://labofadaptiveimmunity.slack.com/archives/D064ZN1SNHZ/p1738593470878619)

Běla

Jinak myslím, že se mi povedlo už vyextrahovat tu tabulku, jen se mi to háže do divného formátu



ondrejondrej  [3:37 PM](https://labofadaptiveimmunity.slack.com/archives/D064ZN1SNHZ/p1738593474718669)

\*procento UMI konkretni sekvence ze vsech UMI toho vzorku (ale to jsi asi myslela)

[3:38](https://labofadaptiveimmunity.slack.com/archives/D064ZN1SNHZ/p1738593507771319)

ondrej

Jako prvni vec bych se podival na tu vyextrahovanou tabulku (edited)

[3:39](https://labofadaptiveimmunity.slack.com/archives/D064ZN1SNHZ/p1738593552849169)

ondrej

V radku bude sekvence a ve sloupcich jednotlive vzorky a v tebulce pocet UMI dane sekvence v danem vzorku

[3:39](https://labofadaptiveimmunity.slack.com/archives/D064ZN1SNHZ/p1738593565902499)

ondrej

a protoze nam jde o specifitu, tak je dulezita AA sekvence.

[3:39](https://labofadaptiveimmunity.slack.com/archives/D064ZN1SNHZ/p1738593592040479)

ondrej

To znamena, ze pokud jsou ruzne sekvence, ktere ale tvori stejny receptor (CDR3) na AA urovni, tak bych je sloucil. (edited)

:+1:1



BělaBěla  [3:42 PM](https://labofadaptiveimmunity.slack.com/archives/D064ZN1SNHZ/p1738593738849519)

Chápu, takhle daleko ještě nejsem, já zatím vyfiltrovala jen to, co má umi count víc jak 1 nebo se vyskytuje napříč vzorky:Z:/48\_lab/Project Cell Fate/boehm\_paper\_tcr/250203\_result\_table

[3:43](https://labofadaptiveimmunity.slack.com/archives/D064ZN1SNHZ/p1738593793657509)

Běla

Akorát se to uložilo do něčeho mezi csv a excelem



ondrejondrej  [3:43 PM](https://labofadaptiveimmunity.slack.com/archives/D064ZN1SNHZ/p1738593822830949)

Jasne. Az takovou tabulku budes mit, tak bych teprve resil kvantifikaci, protoze zalezi, jak moc to bude robustni (bude se to hodne opakovat) nebo stochasticke (v zasade privatni sekvence pro konkretni vzorky s maly prekryvem jinde) (edited)

[3:44](https://labofadaptiveimmunity.slack.com/archives/D064ZN1SNHZ/p1738593883345009)

ondrej

+ je dulezite vedet, co je ktery konkretni vzorek, jestli jsou replikaty nebo jinak sortovane (+/- 4 mer), jestli treba CD8 a CD4 jsou ze stejne nebo z jine mysi



BělaBěla  [4:52 PM](https://labofadaptiveimmunity.slack.com/archives/D064ZN1SNHZ/p1738597931878209)

A mám to teda dle toho rozdělovat v té tabulce?  
Jsou tam CD4 či CD8 vždycky z každé myši. To s tetramerem si nejsem popravdě jistá, protože v článku píšou, že použili "Tetramer staining with I-Ab  
DIYKGVYQFKSV-PE-Tetramer (dilution 1:200)" na všechny buňky ještě než barvili na další markery a pak "anti-APC-tetramers" když hledali Tregy. V tabulce jsem koukala, že je jen sloupec tetramer, který je n nebo p u CD4 (asi neg a pos) a u CD8 to vůbec není, tak to asi budou ty tetramery k Tregům a ty nás nezajímají, že?



ondrejondrej  [4:53 PM](https://labofadaptiveimmunity.slack.com/archives/D064ZN1SNHZ/p1738598034762909)

V ta tabulce bude jeden sloupec jeden vzorek (nebo radek, to uz je asi jedno) s tim, ze by tam mela byt metadata o jaky typ vzorku se jedna (mys, CD4, 4 mer atp.)

[4:54](https://labofadaptiveimmunity.slack.com/archives/D064ZN1SNHZ/p1738598061350099)

ondrej

Ten 4mer asi nesouvisi s Tregy ale se specifitou na ten puvodni 4mer z LCMV

[4:54](https://labofadaptiveimmunity.slack.com/archives/D064ZN1SNHZ/p1738598070104809)

ondrej

a ten budou vazat jen CD4



BělaBěla  [4:55 PM](https://labofadaptiveimmunity.slack.com/archives/D064ZN1SNHZ/p1738598105588669)

A jo, pardon, to dává smysl



ondrejondrej  [4:55 PM](https://labofadaptiveimmunity.slack.com/archives/D064ZN1SNHZ/p1738598148112469)

takze bych to zatim pouzil vsechno. Mozna nam to drobne zkomplikuje analyzu. Bylo by fajn vedet jaky pomer byl tech 4mer+ a -. to mozna neni pro tyhle konkretni mysi, ale nejaky prumer by v tom clanku byt mohl.