Ahoj, promiň, že píšu o víkendu, nedostala jsem se k tomu přes týden.Vygenerovala jsem tabulku, kde jsou jen aa sekvence co měli occurance ve dvou a víc vzorcích, jako sloupečky jsem dala aa sekvenci, samples (seznam všech vzorků které danou sekvenci mají), same\_mouse (true/false, true pokud jde o jen jednu myš, false jestli je to napříc víc myší). Nevím jestli dělat sloupeček same\_tissue? Nebo tam dát třeba ratio S/T? Ale to mi přijde, že by nebylo fér vzhledem k tomu, že ne všechny myši mají vzorky i z T. A pak sloupeček CD4/CD8? Nebo třeba dva sloupečky kde bude počet occurance v CD4 a occurance v CD8?Jinak jména těch vzorků jsem ještě neupravila, potřebuju se pořádně podívat po těch P5-P10 v článku, při zběžném prohledávání jsem to nenašla"Z:\48\_lab\Project Cell Fate\boehm\_paper\_tcr\250518\_repeating\_aa\_seq.xlsx"

thymus a spleen se bude analyzovat zvlast

[2:26](https://labofadaptiveimmunity.slack.com/archives/D064ZN1SNHZ/p1747567587885929)

ondrej

tzn. ze prekryv thymus/spleen nas ted tolik nazajima

:+1:1

[2:27](https://labofadaptiveimmunity.slack.com/archives/D064ZN1SNHZ/p1747567660288699)

ondrej

z te tabulky ti teda vypadly ty frekvence

[2:28](https://labofadaptiveimmunity.slack.com/archives/D064ZN1SNHZ/p1747567717000789)

ondrej

vlastne nevim, k cemu tahle tabulka je

[2:28](https://labofadaptiveimmunity.slack.com/archives/D064ZN1SNHZ/p1747567724139629)

ondrej

ta pred tim mi prisla dobra

[2:29](https://labofadaptiveimmunity.slack.com/archives/D064ZN1SNHZ/p1747567798486659)

ondrej

s tim, ze by bylo fajn rozdelit ten slopecek vzorek na vice kategorii, pak by s tou tabulkou slo pracovat. Take je dulezite, abys pochopila, o co nam jde a pak navrhla reseni, zkusila ho treba aplikovat (treba ty jednoduche kroky z toho) a pak se na tu muzeme podivat

[2:30](https://labofadaptiveimmunity.slack.com/archives/D064ZN1SNHZ/p1747567822203639)

ondrej

A pokud neni jasne, co je P5-10, tak to dame cele pryc

**Monday, May 19th**



**BělaBěla**  [12:39 PM](https://labofadaptiveimmunity.slack.com/archives/D064ZN1SNHZ/p1747647576580239)

Mně přišla ta první tabulka těžce uchopitelná, když tam je jeden sloupeček pro každý vzorek, proto jsem chtěla udělat druhou, kde bude informace o tom, v jakých všech vzorcích ta daná sekvence je, jestli se opakuje i v jiných myších, jestli je S nebo T, CD4 nebo CD8. Zároveň jsem to pochopila tak z toho, cos psal dřív, že chceš ty sekvence, co se opakují napříč různými vzorky (tedy ve dvou či více sloupcích nenulová hodnota), což mi tu první tabulku s 1473 sekvencemi zredukovalo na 943. Jak by sis to rozdělení do kategorií představoval? Že by v řádcích byly sekvence a pro každý vzorek místo jednoho sloupečku by bylo několik dalších sloupečků, kde by se opakovala informace o tom vzorku jako S nebo T atd. v každém řádku?



**ondrejondrej**  [12:43 PM](https://labofadaptiveimmunity.slack.com/archives/D064ZN1SNHZ/p1747647829165119)

Ta nova tabulka nebyla kvantitativni. Potrebujeme zjistit / kvantifikovat jako moc jednotlive TCR sekvence jsou CD4 vs CD8 biased v thymu a v periferii.

[12:43](https://labofadaptiveimmunity.slack.com/archives/D064ZN1SNHZ/p1747647838813829)

ondrej

A jak moc jsou nevyhranene.

[12:44](https://labofadaptiveimmunity.slack.com/archives/D064ZN1SNHZ/p1747647872500979)

ondrej

Sekvence, ktere se neopakuji nas sice jakoby nezajimaji, ale zas dotvareji celkovy pocet sekvenci v kazdem vzorku

[12:45](https://labofadaptiveimmunity.slack.com/archives/D064ZN1SNHZ/p1747647911718199)

ondrej

V nekterych packagich, jako je treba ruzne similarity indexy to je potreba

[12:45](https://labofadaptiveimmunity.slack.com/archives/D064ZN1SNHZ/p1747647922426429)

ondrej

predpokladam

[12:45](https://labofadaptiveimmunity.slack.com/archives/D064ZN1SNHZ/p1747647930318759)

ondrej

i kdyz celkove je to pro nas spis sum

[12:46](https://labofadaptiveimmunity.slack.com/archives/D064ZN1SNHZ/p1747648004041369)

ondrej

A myslel jsem to tak, ze u kazdeho sloupce budou nasledujici kategorie - nejake ID toho samplu, mys (cislo), organ (T/S), populace (CD4 vs CD8), 4mer (ANO/NE) (edited)

[12:47](https://labofadaptiveimmunity.slack.com/archives/D064ZN1SNHZ/p1747648056362329)

ondrej

Pak by asi bylo dobre spocitate nejake ty similarity indexy mezi jednotlivymi vzorky.

[12:48](https://labofadaptiveimmunity.slack.com/archives/D064ZN1SNHZ/p1747648106317019)

ondrej

Asi je mnoho zpusobu toho, jak se dobrat k nejakemu reseni te nasi otazky. Asi by bylo fajn, kdybys nejaky zpusob nebo zpusoby navrhla a pak se o tom muzeme pobavit. (edited)



**BělaBěla**  [12:55 PM](https://labofadaptiveimmunity.slack.com/archives/D064ZN1SNHZ/p1747648524909289)

Jen abych se ujistila, že to chápu správně, v prvním kroku teď mám tu count\_table kterou jsem posílala jako první rozšířit s tím, že mám 21 vzorků a pro každý vzorek bude 5 sloupečků s kategoriemi co jsi zmínil výše, tedy celkem 105 sloupečků a ty neopakující se sekvence nechám, takže 1473 řádků



**ondrejondrej**  [12:58 PM](https://labofadaptiveimmunity.slack.com/archives/D064ZN1SNHZ/p1747648694489509)

Tak pro kazdy vzorek muze byt porad jeden slopecek, jen nekolik prvnich radek by byly soucasti zahlavi. Pokud se s tim nepracuje blbe - mozna ano.

[12:58](https://labofadaptiveimmunity.slack.com/archives/D064ZN1SNHZ/p1747648722988919)

ondrej

Mozna je nejake lepsi reseni, ale urcite je dobre mit tabulku, kde lze jednoduse vytahnout ty vyse uvedene informace (edited)

:+1:1



**BělaBěla**  [12:59 PM](https://labofadaptiveimmunity.slack.com/archives/D064ZN1SNHZ/p1747648794009909)

Jakmile budu mít tabulku, tak vybereme strategii, jak vytáhnu z té tabulky informaci pro každou aa sekvenci ohledně toho, zda je výhradně CD4, výhradně CD8 nebo non-biased a z toho pak grafíky, chápu to správně?



**ondrejondrej**  [1:00 PM](https://labofadaptiveimmunity.slack.com/archives/D064ZN1SNHZ/p1747648807052289)

jj



**BělaBěla**  [1:02 PM](https://labofadaptiveimmunity.slack.com/archives/D064ZN1SNHZ/p1747648921405739)

No já jsem si právě říkala, jestli by třeba nemělo smysl mít dvě tabulky s tím, že jedna bude čistě numerická kde budou ty umi counts a jestli se aa seq vyskytuje v tom vzorku nebo ne (s tím že by asi dáválo víc smysl tam mít to id pro každý vzorek než ten název) a druhou tabulku jakoby na metadata, kde bude těch pět kategorií

[1:02](https://labofadaptiveimmunity.slack.com/archives/D064ZN1SNHZ/p1747648940905409)

Běla

Bylo by to přehlednější, ale zas to jsou dvě tabulky místo jedné



**ondrejondrej**  [1:02 PM](https://labofadaptiveimmunity.slack.com/archives/D064ZN1SNHZ/p1747648946341939)

To me taky ted napadlo

[1:02](https://labofadaptiveimmunity.slack.com/archives/D064ZN1SNHZ/p1747648953591549)

ondrej

ze muze byt sekundarni tabulka

[1:02](https://labofadaptiveimmunity.slack.com/archives/D064ZN1SNHZ/p1747648970722199)

ondrej

kde podle ID vzorku budou tyhle parametry

[1:03](https://labofadaptiveimmunity.slack.com/archives/D064ZN1SNHZ/p1747648980775369)

ondrej

ja takhle asi nedokazu rict, co je lepsi

[1:03](https://labofadaptiveimmunity.slack.com/archives/D064ZN1SNHZ/p1747649011081729)

ondrej

Tak muzes zkusit, co ti prijde lepsi



**BělaBěla**  [1:04 PM](https://labofadaptiveimmunity.slack.com/archives/D064ZN1SNHZ/p1747649096463179)

Taky moc nejsem schopná posoudit, co je lepší, protože doteď jsem vždycky metadata tabulku s tabulkou s výsledky spojila a pak z toho dělala grafy, ani nevím jestli by mě Rko nechalo mít pro ggplot jako source dvě tabulky

[1:05](https://labofadaptiveimmunity.slack.com/archives/D064ZN1SNHZ/p1747649134215809)

Běla

Uvidím, zkusím to otestovat



**ondrejondrej**  [1:06 PM](https://labofadaptiveimmunity.slack.com/archives/D064ZN1SNHZ/p1747649203855129)

Tu druhou tabulku bys asi pouzivala spis na to, jak si vytahnout (redukovat) veci z te hlavni tabulky.