In April 2018, Zandrea put me in touch with Shuntai Zhou from the University of North Carolina.

He helped her previously with bioinformatics tasks.

[shuntaiz@email.unc.edu](mailto:shuntaiz@email.unc.edu)

<https://github.com/SwanstromLab/PID>

Christopher Kline is also involved in the conversation. [cjk14@pitt.edu](mailto:cjk14@pitt.edu) . He is Zandrea’s lab manager.

------------------------------------------------------------------------------------------------------------------------------------------------------

**Email from April 19th 2018 – DATA QUALITY NOT GOOD ENOUGH**

Hi Zandrea,

I finally talked to Shuntai!

So, it looks like some of the pools / samples worked, but many did not.

Here is a screenshot for one of the log files. Consensus2 is the number of reads that are good, according to the protocol from Ron Swanstrom’s lab.

Shuntai told me, however, that he only trusts the data if the Distinct to Raw is lower than 0.1 and if Resampling index is higher than 0.9. This is only the case in three of the samples.

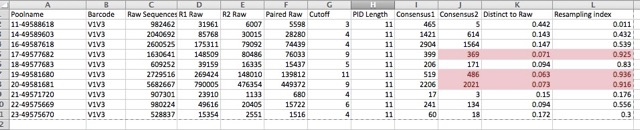
In the second log file, none of them pass that quality control.

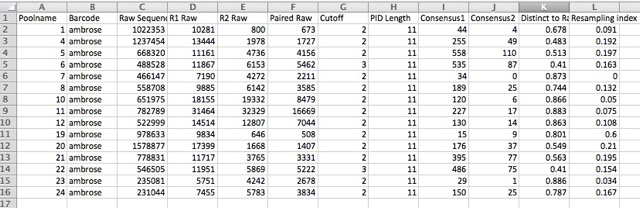
He told me that the problem likely has to do with too much adapters in the library. That’s why we see a lot of poly-A sequences in the fastq files.

Somehow your technician needs to figure out how to fix that but, Shuntai says he is happy to help.

Now, still on my list is to determine whether it would be enough to do 250 bp instead of the paired end reads. I’ll try to get you that tomorrow!

Pleuni





------------------------------------------------------------------------------------------------------------------------------------------------------

**Response from Zandrea May 18th 2018 – SWITCHING TO R01 ANIMALS WITH HIGHER VIREMIA**

Hi Pleuni,

I hope you are doing well and hopefully your semester is wrapping up and you can relax a little bit!

I just wanted to let you know that Chris and I decided to perform MiSeq troubleshooting on samples from the newer R01 study (not the R21 study that will go into a no-cost extension, as we previously have done). This involved designing and testing new primers, as the SIV in the R01 study is different from the one in the R21 study.

The reason why we decided to switch to the new study/new samples is mainly because the viremia in the animals of the R21 study was very low, which is the reason we got a new virus to infect the animals in the R01 study. The new animals have much higher plasma viremia, so we won’t use up the samples in a single run to get enough RNA to run the assay. Also, in the R21 study, Ling’s lab was isolating the viral RNA, as the samples are from Ling’s old R01, and they were not using the most optimal assay. I wanted to make sure the quality of the RNA was very good in trying to troubleshoot.

So, based on what you and Shuntai talked about, it appears that our primer concentrations may have been too high and were affecting the results, although we were using the amount in Shuntai’s protocol. But talking to another lab who does a similar assay, they found that this amount of primer was not optimal. So, Chris is directly comparing the assay with the original primer concentration and a lower primer concentration on the sample 17 plasma samples + virus challenge stock to see if the sequencing results differ. Unlike the previous rounds with the R21 samples, Chris is getting nice, prominent bands from every sample.

At the end of next week, we should have reads from both runs (high primer and low primer) to compare and hopefully to obtain actual useable sequencing data. My question to you is whether you want us to send you the raw data? Or do you want Shuntai to process the raw data for you? Whatever you prefer is fine with us.

Thanks,

Zandrea

--------------------------------------------------------------------------------------------------------------------------------------

**Chris Kline** [**cjk14@pitt.edu**](mailto:cjk14@pitt.edu) **email May 21st 2018 – ANNOUNCING NEW DATA ON THEIR WAY!**

Hi Pleuni,

I’ve already finished the first run and am working as fast as I can to get the second run on the machine tomorrow as early as possible. Mainly because the run takes about 72 hours to complete so I’d like to have it done before I leave on Friday.

Yes, the data will be on Illumina’s basespace platform like before. When both runs are complete, I can share them with you through Illumina and I will email Shuntai to see if he has the time to analyze the runs as well. If he does, I’ll make sure to send the zip files along to you. Would you prefer that I unzip them and put them into dropbox like last time?

Thanks,

Chris

--------------------------------------------------------------------------------------------------------------------------------------

**Zandrea email May 30th: DATA LOOK MUCH BETTER**

Hi Pleuni,

According to Shuntai, it looks like both of Chris’ runs worked much better than before. As a reminder, the 2 runs are on the same samples from the R01 animals that have been infected with a different virus. For the runs, Chris used Shuntai’s exact protocol (run 1) or with less primers (run 2). What we observe is:

1. All samples worked! In other words, we got paired-end reads in both runs for all samples. This is different from the previous results from the R21 animals in which many samples didn’t work. Those samples were from animals infected with a different virus with higher viral loads and the primers we used were different. In addition, Chris did the RNA extractions for himself for the R01 animals, whereas the RNA from the previous R21 animals were extracted from Ling’s lab with, in my opinion, a suboptimal protocol. Thus, in the new samples, the primers may be better optimized and the quality of the RNA may be better.
2. Using less primer increased the number of paired-end reads for nearly all samples (average reads of 1151 vs. 4985), suggesting that the original primer concentration was too high and it is best to use less.
3. The distinct to raw ratio and the resampling index numbers aren’t as great as what Shuntai had told you, but he says these are theoretical parameters and we may still have some excess primers/junk in the reactions. We are working to improve our library clean-up to get rid of this stuff, which may be more than changing primer concentration.

So, it would be great if you could take a look at the sequences and see if they look better than before!

* If yes, we can tell you the animal numbers and time points for the samples and you can even start doing analyses.
* If no, then we will need to work on optimizing the assay in our hands -- I have some ideas and I’m going to talk to Vaughn Cooper about how he prepares his libraries.

But I hope there are good reads in there! I’m encouraged by the significantly higher numbers of paired-end reads!

Thanks,

Zandrea

--------------------------------------------------------------------------------------------------------------------------------------

**The data from these two runs are on Dropbox in the folder data\_June2018\_ChrisKline\_ZandreaAmbrose**

**There is also a file with metadata named Run1SampleSheetZA.xlsx**

**Note: Kaho has started analyzing this dataset early October 2018**

--------------------------------------------------------------------------------------------------------------------------------------

**Zandrea June 4th 2018: WHAT TO DO WITH THE R21 ANIMALS?**

Hi Pleuni,

I hope you’re doing well and the semester is coming to a close for you!

Just checking on whether you’ve had a chance to look quickly at the sequences to see how they look? Chris did some more troubleshooting and is now able to get rid of more extra primers, which should help. We can’t really move forward until we know if the sequences are better or not (i.e. do we still have to optimize the assay or are we OK to continue generating sequences), so please let us know as soon as you can.

Also, we probably need to continue optimizing the primers for the R21 animals, which likely means we need to make an even shorter amplicon. I never heard from you about whether shorter sequences are OK. I think we can still use the Primer ID method to get more than 250bp, but it may be more like 350-400bp, which is hopefully OK for your analyses.

Thanks,

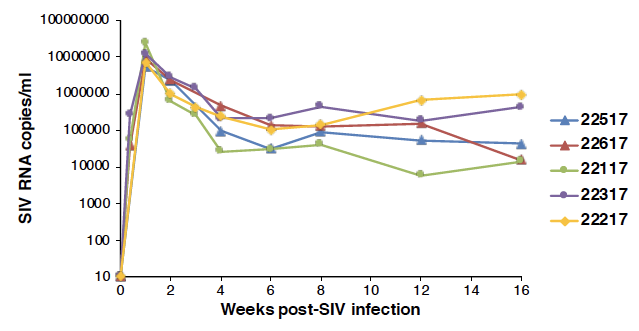
Zandrea

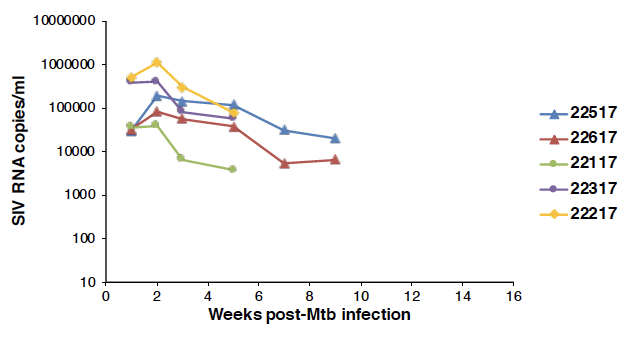
--------------------------------------------------------------------------------------------------------------------------------------

**Zandrea July 19th 2018 – MORE ABOUT THE TB INFECTIONS**

Hi Pleuni,

It was good to talk to you earlier this week! Chris just got the latest viral loads and I found out when the animals were infected with Mtb and unfortunately it was not consistent: week 20 for 22517 and 22617, week 19 for 22117 and 22317, and week 17 for 22217. But we can graph them for the first 16 weeks and then after Mtb infection:





As we discussed, it would be great if you could reanalyze the MiSeq data that Chris previously sent you (the last dataset), particularly the last time point to make sure you think the sequences are OK. We can continue to obtain some additional data for these animals! Also, I’m meeting with Ling on Tuesday, so it would be great to show her something from you now that I think we are making progress in getting sequences!

We’re also doing the last check on the new assay we optimized for the R21 animals, so we also want to go back and get sequences from those animals in the future (probably after the lab move).

Thanks!

Zandrea

**NOTE: WE NEED TO INCLUDE THE TB INFECTION DATE IN THE ANALYSIS**

--------------------------------------------------------------------------------------------------------------------

Hi Pleuni,

Our meeting is at 2pm ET.

As for env, I’m not that used to it either. Facts that I do know about it, which may or may not be useful to you:

* There are alternating so-called conserved and variable regions. The V regions correspond to loops, with V3 being the major receptor/coreceptor determinant. I believe our sequence spans the V1-V3 region (or at least a part of it), so there may be regions that behave differently than others.
* The Env protein is heavily glycosylated (~50% of the mass), which is thought to be a defense mechanism to shield against antibody recognition. My understanding is that these N-linked glycosylation sites are extremely variable and can evolve rapidly.

Not sure if that’s useful or not!

Thanks,

Zandrea

--------------------------------------------------------------------------------------------------------------------

Hi again, Zandrea!

So I did some more digging.

It turns out that there are some indels, which make the analysis a bit harder. For now, I just looked at the first 100 sites, because there are no indels there.

These first 100 sites, look conserved. Mutation frequencies are mostly low. Synonymous sites are somewhat less conserved than non-synonymous sites. See attached figures.

This is good. It means that the data look good. The next question is whether we can detect any signal in the data. Are the different time points or different animals different? How about different compartments within the animal?

My next tasks will be:

1. deal with the indels

2. determine wether there are differences between the animals.

Could you (or Chris) remind me why we have TCS\_1 and TCS\_2?

Thanks!!

Pleuni

--------------------------------------------------------------------------------------------------------------------

Cool! Yes, my understanding of Env is that there can be lots of indels. We actually see that ourselves when we compare a portion SIVmac251 (R21 virus) env with the published sequence of SIVB670 (R01 virus) env or a sequence Chris (CK) got of our SIVB670 stock:

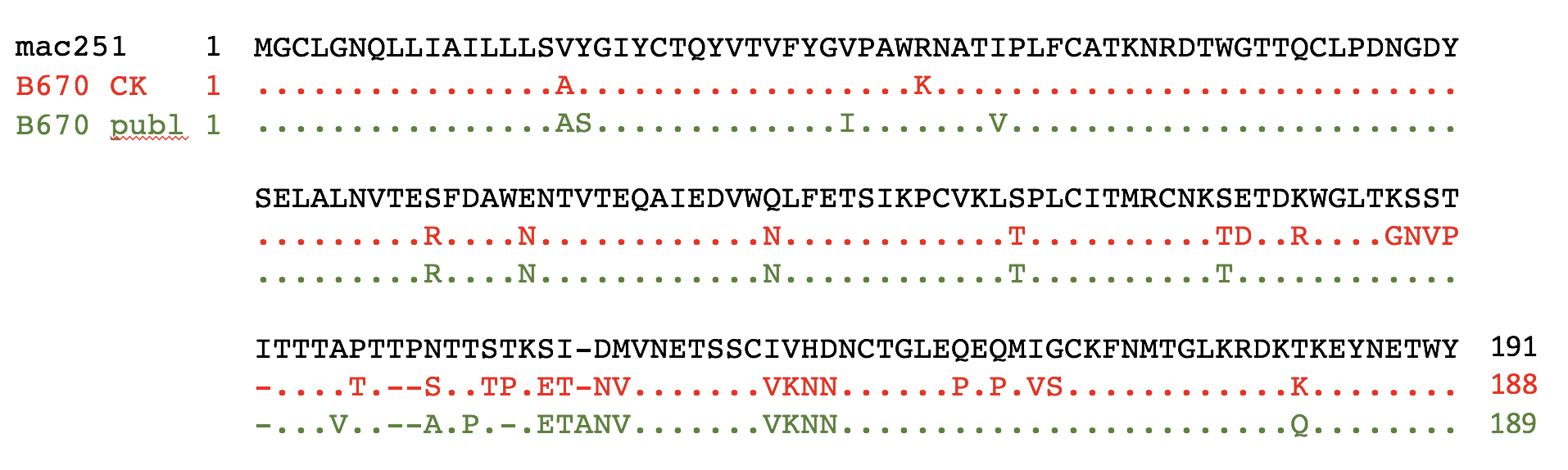
We are going to extract intracellular viral RNA from lymph node and lung cells next! It sounds like they saved serial samples from these animals. Also, they are sacrificing some of the animals already, so we’ll get more tissue samples.

We’re also finalizing our optimization of the SIVmac251 primers for MiSeq (R21 study) and we can try to get that finished up.

Do you think it would be helpful to have a monthly Skype meeting on these projects?

Thanks,

Zandrea



P.S. We’re all still baffled about why the amount of plasma virus decreases after Mtb infection. It will be VERY interesting to compare SIV sequences before and after Mtb infection. I’m wondering if some type of immune response is targeting SIV-infected cells, which would decrease diversity??

--------------------------------------------------------------------------------------------------------------------

Hi Pleuni,

Sounds good. We’re moving forward with getting more sequences and it would be great to have a standing monthly meeting with you. I’m happy to talk to Dmitri, too!

A follow up since my last email: I think the initial increase in viremia followed by decrease right after Mtb infection might be due to CD8+ T cell responses, since Ling’s lab sees an increase in both CD4+ and CD8+ T cells in the blood and lung at the same time. (They are looking at SIV-specific immune responses in both cell subsets now.) Others have shown that more CD4+ T cells (either by AIDS or CD4+ depletion by antibodies in monkeys) leads to a significant sustained increase in viremia, which wouldn’t explain our viremia data. However, CD8 depletion studies, which have mainly focused on depletion at the time of SIV challenge or in conjunction with ART, show that transient CD8 depletion leads to an immediate significant increase in plasma viremia and then a decline that eventually recovered as CD8s came back. So, we may be seeing a similar effect as new CD8 cells are being recruited to the lungs to try to fight the Mtb infection.

I could only find one study by Guido Silvestri (the editor of our PLoS Pathogens paper!) who looked at diversity after CD8 depletion:

They depleted during nonsuppressive ART (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5087330/> ) and did SGS (they call it SGA) from plasma SIV RNA and some MiSeq that was not well controlled:

To gain insight into the dynamics of virus replication and/or reactivation after CD8 depletion, we performed a longitudinal study of plasma virus by single genome analysis (SGA) of the *env* gene of SIVmac239 at three time points: (i) day 10 post SIV-infection, i.e., **peak viremia**, (ii) day 56 post SIV-infection, i.e., **immediately prior to ART initiation**, and (iii) **after CD8 depletion**. We conducted this analysis on three representative RMs (RKq11, ROw8, and RLb13) with persistent SIV suppression under ART. Circulating viruses sequenced at day 10 post-infection were relatively homogenous within each RM (99.9-100% amino acid identity). In contrast, sequences isolated day 56 post-infection showed several fixed mutations that likely represent escape from either T cell-mediated or antibody responses. **Interestingly, the viral sequences derived from plasma after CD8 depletion were more similar to those derived at peak viremia, compared to those derived immediately prior to ART (i.e., peak vs. pre ART amino acid identity = 99.5-99.7%, and peak vs. post CD8 depletion amino acid identity = 99.8-99.9). Importantly, the previously fixed escape mutations observed at the pre-ART time point represent only a small fraction of the circulating viruses after CD8 depletion.** In addition, we conducted a MiSeq deep sequence analysis of the circulating cell-associated SIV-DNA immediately prior to ART initiation and immediately before CD8 depletion. As shown in [Figure S6](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5087330/#SD1), this analysis revealed a consistent increase in inoculum-like sequences (i.e., similar to those derived at peak viremia) under ART. This latter observation is consistent with the finding that viruses similar to the initial infecting variants reemerge in plasma during the early stages of ART (data not shown), and suggests the presence of a pool of cells infected with inoculum-like virus that outlive cells infected with virus harboring escape mutations. Of note, when we conducted the same longitudinal sequence analysis of plasma virus by SGA in the SIV-infected RM that never achieved undetectable viremia (ROn13), we found a complex pattern of virus quasi-species diversity throughout the three examined time points that was consistent with continuous virus replication (data not shown). Overall, these data suggest that the plasma viruses that are present after CD8 depletion are produced from long-lived cells that were infected prior to ART initiation and before the generation of immune escape mutants.

**For us, it would be important for us to do similar analyses and also look at tissues, as we’re already planning to do. And we’re not doing a fake CD8 depletion, but are looking at a natural effect during co-infection of 2 highly pathogenic pathogens. And we’ll have deeper sequencing with the MiSeq. Also, I think the homogeneity they see early in the plasma viremia is consistent with our data.**

Have a good trip to Europe (Trump-free hopefully!) and get in touch when you’re back – I’ll be in my new office!

Sincerely,

Zandrea

PPT in Dropbox file “R01 SIV VL and CD4 update 7.24.18.pptx”

<https://www.dropbox.com/s/avtlf2uakhhslco/R01%20SIV%20VL%20and%20CD4%20update%207.24.18.pptx?dl=0>