# Details & Setup

**Time:** 20 minutes  
**Audience:** Some know some of the science / assays. Unlikely the scenario is familiar.  
**Setup**: NAb IC50 with non-responses removed.

**Issues to resolve**: 190 (time alignment), 333 (incorrect highlighting with x selection), 219 (green selection counts), 362 (undefined tooltip message accuracy)

# Backup scenario

CAVD 419 – previously replicated RV-144 in a macaque challenge trial with 40% efficacy. Wanted to see if priming with DNA or AD26 would increase the efficacy of the gp140 vaccine. ICS responses are actually HIGHER when unprimed, while ELISPOT are lower. Odd, eh?

# Big Idea

The single most important thing to remember is that you start with all of the integrated data. You don’t have to find and combine datasets to build up data for a question – all the data are already in your shopping cart. What that means is if you plot the default measure from an assay you’ll be seeing data from every available subject that was tested, from every time point in every study where it was run, and every antigen tested. That’s a lot of dots. In any presentation graphic these factors would be teased out and clearly marked. Often you’ll want to break it apart by these factors so that you can see what you’re looking for or filter things out. [NAb CLADE] For example, I can break the results out by the Clade of the assay antigen and now with a few clicks I’m comparing response by clade across 12 studies. So that’s the #1 thing: don’t forget you may be looking at more than you expected.

# Tour: Home

After you log in you’ll see groups of subjects that you’ve saved in the past over on the left. News is in the middle. Right now, the news is just instructions and reference for our beta participants.

# Tour: Learn

In the Learn section you can find detailed information about completed CAVD studies, assays, and products. We will be getting more details during beta so some pages might be a little light. It’s important to note that the Learn section describes all the completed CAVD studies but we only have subject-level data for about 17 of them. Notice on CAVD 256 right now it’s sparse but more information is coming. You can click around to see relationships. So if I click NYVAC-C it shows me the product page and I can see it was used in other studies, too, and see that those are NHP studies.

# Tour: Find Subjects

Find Subjects is where you can see relationships between subjects. It can be a good way to visually filter down to a group of interest. You can see subjects broken down by these factors. I can see how many subjects got NYVAC-C, for example: 161.

## Tour: Info Panel

If I select them, it goes over here on the right. The other bars are showing that some subjects also received these other products. Now I can choose whether I want to make it a filter. Filters are persistent and affect every view of data, so this would constrain what you can plot or export to only the subjects that received this product. However, that filter preserves everything we know about those subjects – all the assay results from all time points. Below the filters you can see some counts of important factors. These are to help understand the diversity in your filtered data. You can see how many studies used that product, for example: 3. In fact, if I switch to view Studies I can see them and how many subjects were in each.

# Tour: Plot

Next is the Plot. You can compare up to 3 variables at once across multiple studies, time points, antigens, and more. So if I plot NAb data it gives me some default choices I’ll just accept [log]. And I’m now looking at 40 subjects who received NYVAC-C. But remember the one most important idea: each subject probably has lots of dots here – multiple time points, multiple viruses. There will be counts in the info pane that help make that more clear later during beta.

So there’s lots you could do now. One useful step is to put time on the x axis. [STUDY DAYS]. Notice this option down here lets me choose an alignment. Let me show you “Last Vaccination.” This lets you see all the studies you have data for in your plot and you can hover on these icons to see what was happening in the study protocol at these times. Expand it and you can see each treatment’s schedule separately. Later on you’ll be able to click these icons to select the associated data above.

When we selected Last Vaccination, it found the last vaccination visit in each treatment and called that week 0. In reality that visit is at a different week across studies and treatments – investigators are looking for differences in the number and timing of vaccinations. The weeks after this might include the peak immune response induced and you could select and filter to it across multiple studies and arms, which we think may be very valuable and much faster and easier than in any other tool.

You can directly select specific data to get its details and filter down. In this plot, maybe I’m only interested in the top responses that are above 1000. When I make that selection you’ll notice that the data I selected turns green and some dots outside the selection turn black – those are other data from the same subjects, which can be very useful context to see breadth or consistency of response. They might be tested at other times or against other viruses. And I can see even without filtering how many subjects they come from and whether it’s 1 or both studies.

I can replace this X axis with the virus clade from and see how they compare. Notice I get box plots to help. In NYVAC-C, the “C” stands for Clade C HIV, so it makes sense that Clade C viruses had the best neutralization on NAb. I can even select only Clade C and filter to the data that match the vaccine strategy.

[DO WE NEED TO REMOVE NON-RESPONSES?]

**OPTIONAL TO EXTEND TIME**: I could see if there’s a difference between the viruses used in the assay [X VIRUS NAME], where the first one seems to have a much better response. If I simultaneously want to consider the study I can add [COLOR STUDY NAME] that as a color and see there’s no clear interaction there.

And if I want to see if there’s a relationship to T-Cell activity I can switch again to plot ELISPOT. [SHOW DIMENSIONS] One important thing to note is that I’m plotting multiple NAb antigens and multiple ELISPOT antigens so it’s going to take the MEDIAN values across those antigens for each subject at each time point in order to make a plot. You can see the tooltip right up here at the top. If you don’t want it to do that, another way to get a plot is to select only one antigen on each axis – then it can happily plot their real values.

So notice something unexpected here. I get a plot but also these gray areas. When I hover over them it explains reasons why. These are data from subjects who were tested on both assays but don’t have a matching x or y. One likely reason in this case is that these assays were sometimes tested at different times, so we don’t have a matching value.

But we do have a plot and it looks like it would be hard to conclude there’s a strong relationship.

One trick I like for dealing with these gray areas is making a selection window and then moving it to see what dots highlight in the other gray area. If the top NAb subjects also had high ELISPOT values then there could still be something there, but again, nothing convincing.

I can also get rid of the Clade C filter and see if that makes a difference. Hardly a change. If I did want to see how the assays compared from 2 different time points, I could export and make that comparison myself – this system is deliberately limited to guide me toward comparing data from the same times.

# Tour: Grid

To export, I can go to the Grid. It’s a spreadsheet view for all the data in your active filters. You can add columns to the Grid about any information we have from those subjects. Assays, time, attributes. You can also filter on all these columns. And if you click Export it gives you an Excel file for use in your own tools. The plot can be a good way to know what you want to export, but you never have to use it if you’d rather use R or JMP or Excel or whatever right away.

OPTIONAL TO EXTEND TIME: I’ve previously made an R script to produce a scatterplot matrix of each NAb virus by each ELISPOT antigen, and here’s the result, which is an example of how and why we expect users to export sometimes.

# Recap

Let’s review. By taking the tour, we actually explored a few questions. In Learn, we saw what product was used in a study. We saw which other studies used it too. In Find Subjects we filtered to only those product recipients. In Plot we looked at whether they showed a relationship between Clade C NAb response and IFNg ELISPOT response. As is often the case the answer seems to be there’s no evidence to support that idea. But you can imagine how being able to answer that question yourself quickly can help you. Maybe you move on to another idea or slightly change the question to see if something else might be happening.