Issue for consideration with the South African virus?

You are likely well aware this situation, and have a good motivating reason for your variant choice for study, but just in case you were not aware of this, I thought I should point it out: the form of 501Y.V2 that is circulating in S. Africa and spreading globally does not match the form that people seem to be testing and using for vaccine design. There is a mutation that is immunologically important in the B.1.351 variant, in the NTD supersite, a resistance mutation that is very rare globally, **R246I**.

So I'm writing just in case you weren't aware. Given this, I think (i) you might be need to do some experimental work to see if, B.1.352, the relevant epidemic form given the current public data, may be a little more sensitivity than was B.1.351, to vaccine and convalescent sera, as it might retain some sensitivity to some NTD directed NAbs. Also, as our best chance to stay ahead of the virus may well be to target multiple epitopes within a full Spike, you might need to make sure that the NTD part of the immune response to B.1.351 vaccines are as hoped for, and will ross-react with B.1.352, for example.

Please let me know if I'm thinking straight about this or not. It seemed like it was important so I'm writing to you straightaway, just in case you weren't aware and would find the information helpful for planning immunological assessment of the current 501Y.V2 vaccine candidates.

Thanks for considering this,

5.1.2e 3/14/21

Key points:

1) At least some people are testing and making immunogens based on the form of the South African virus that is not the one that is spreading, as far as I can tell, people are using B.1.351 [redacted]
2) The form of the virus that *is* actually spreading is B.1.352. Spike is found precisely in this form 596 times.
3) The R246I change is found only 13 times in B.1.351, and very rarely in nature outside of the B.1.351 virus
4) This rare form was only transiently sampled, Oct-Dec, not since.

The difference between the two forms is the following:

Number of observations
GISAID -> cov.lanl.gov 12/1/2020 - 3/14/2021

		NTD impact	RBD impact		S. Africa	Global	Num_countries	
B.1.351	D80A	ΔL242/A243/L244 **R246I**	K417N E484K N501Y	D614G A701V	13	13*	Just S. Africa	
B.1.352	D80A **D215G**	ΔL242/A243/L244	K417N E484K N501Y	D614G A701V	186	596	35 countries	

- Outside of B.1.351, R246I had been found only 23/564,368 times – its at a very low frequency in the population.

Most common patterns for Africa for this variant: Left, Spike positions written vertically over the Wuhan form of the virus, identities are periods, deletions dashes. Only the 50 most variable positions are shown. The potential problem mutation, R246I, is in red. I'm not really worried about D215G, but it _may_ also do something. The mini-alignment is followed by a tally of the counts each form is found in each continent and UK, between 12/1/2020 – 3/10/2021. L18F comes and goes, as it often does but at least both forms are common.

```
                                              1111
          1111222222223344444455666666677778991122
 1125667899034512444456801578907115778801698381523
52822790056284752234647587274010345571916968928587
LSLTQAHVDTERDYFDALALRSVTRKLSEFNAQDHQQPSATGDFSSDYVM   Global   UK  Europe  NAmer   Asia  Africa  SAmer  Oceanis  [Most common Spike that contains the pattern on the left]
........A......G.---.....N..K.Y..G.....V..........   286      25  117     7       18    109     0      10       [D80A,D215G,L242-,A243-,L244-,K417N,E484K,N501Y,D614G,A701V]
..F.....A......G.---.....N..K.Y..G.....V..........   196      12  100     3       6     73      0      2        [L18F,D80A,D215G,L242-,A243-,L244-,K417N,E484K,N501Y,D614G,A701V]
........A........---I....N..K.Y..G.....V..........   11       0   0       0       0     11      0      0        [D80A,L242-,A243-,L244-,R246I,K417N,E484K,N501Y,D614G,A701V]
........A......G.---.....N..K.Y..G................   2        0   0       0       0     2       0      0        [D80A,D215G,L242-,A243-,L244-,K417N,E484K,N501Y,D614G]
........A......G.---.....N..K....G.....V..........   2        0   0       0       0     2       0      0        [D80A,D215G,L242-,A243-,L244-,K417N,E484K,D614G,A701V]
.................---..............................   3        0   0       0       0     3       0      0        [L242-,A243-,L244-]
```

A few notes on possible importance of R246I

1. 5.1.2e is optimistic that it won't matter too much, but he is going to make and test both forms as pseudoviruses and evaluate the sensitivity to sera, and he can let us know.
2. 5.1.2e initially studied a virus that retained the R246, and it was highly resistant to sera too; their modeling suggested it was due to the deletion Δ244-246 shifting R246's location 5.1.2h
3. Δ244-246 is a clear problem in its own right (Hastie… Ollman Saphier bioRxiv) but the combination of both Δ244-246 and R246I might even be more severe and so have more limited ability to elicit x-reactive responses in a vaccine. (The convalescent sera from 510Y.V2 that you may have have been hearing about from Penny's recent studies and presentations, that were able to cross-react well with the G clade viruses, were likely from B.1.352 infections.
4. R246 is in N5; it appears clearly to be a key site for antibody sensitivity, so I think it is worth evaluating experimentally to know what you are getting. It's positive charge can enhance negatively charged antibody interactions.

1. "Residues in CDR H1 form a network of hydrogen bonds involving positively charged residues from N3, Lys147NTD, and N5, Arg246NTD, with interactions by the side chains CDR H1residuesTyr27HCand Glu31HC."

   1 quoted from: Cerutti, G. … Shapiro bioRxiv Potent SARS-CoV-2 Neutralizing Antibodies Directed Against Spike N-Terminal Domain Target a Single Supersite 5.1.2h

2. "As a stunning example of convergent binding, S2L28, S2M28 and S2X333 contact the N-terminal region (residues 14-20, NTD N-terminus), a -hairpin formed by residues 140-158 (supersite -hairpin), and a loop spanning residues 245-264 (supersite loop). These three regions collectively form an antigenic supersite on the pinnacle of the NTD on the side distal to the S2L28, S2M28 and S2X333 each mAb uses hydrophobic residues at the tip of the HCDR3 loop to contact the NTD supersite near residue R246. For S2L28, S2M28, and S2X333, this hydrophobic residue is W105, I93, and W106, respectively (**Figure 2G-I**)."
3. "The R246A substitution decreased binding of S2L28, S2M28 and S2X333 (to various extents), in agreement with the extensive interactions formed by the R246 side chain with aromatic and hydrophobic residues found in the HCDR3 of each mAb (Figure 3A). We found that another neutralizing mAb we isolated (S2X28), as well as the previously described mAb 4A8 (Chi et al., 2020) are also sensitive to the R246A substitution"
4. "all viral escape mutants resulted from a single non-synonymous nucleotide mutation. For instance, S2L28 selected for the D253E and D253G substitutions whereas S2X28 selected for the Y144C and Y1444N substitutions. Furthermore, both S2M28 and S2X28 selected for the R246G substitutions.

   2-4 quoted from McCallum, bioRxiv preprint N-terminal domain antigenic mapping reveals a site of vulnerability for SARS-CoV-2 5.1.2h