Single Antigen Analysis

Derek Sonderegger

2019-06-25

For the human test/training data sets, we fit a model using only a single antigen on the training set (which included all the negatives but only positives from that time point) and then calculated the AUC values using the corresponding test set. The following table gives the top 10 antigens for the IgG and IgM serum AUC values.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Type | Antigen | Week1 | Week2 | Week3 | Week4 |
| IgG | LPSA | 0.7567 | 0.9734 | 0.9658 | 1 |
| IgG | BPSL2697\_GroEL | 0.6926 | 0.7354 | 0.9108 | 0.969 |
| IgG | BPSS0477\_GroEL2 | 0.6306 | 0.6921 | 0.9083 | 0.9796 |
| IgG | CPS | 0.7316 | 0.8112 | 0.8999 | 0.7511 |
| IgG | BPSS0530 | 0.6795 | 0.6925 | 0.8214 | 0.9599 |
| IgG | BPSS1652 | 0.6552 | 0.6655 | 0.7917 | 0.9641 |
| IgG | BPSL2522\_OmpA | 0.6194 | 0.6083 | 0.8357 | 0.9092 |
| IgG | BPSS0135 | 0.66 | 0.7036 | 0.809 | 0.7704 |
| IgG | BPSS0476\_GroS | 0.6095 | 0.7 | 0.797 | 0.8303 |
| IgG | LPSB | 0.6432 | 0.7023 | 0.8298 | 0.5958 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Type | Antigen | Week1 | Week2 | Week3 | Week4 |
| IgM | MSHR5855.WCL | 0.6687 | 0.5553 | 0.7594 | 0.793 |
| IgM | LPSA | 0.4977 | 0.7889 | 0.8867 | 0.5493 |
| IgM | BPSL3396\_AtpD | 0.693 | 0.6325 | 0.5471 | 0.812 |
| IgM | BPSL2522\_OmpA | 0.6681 | 0.6431 | 0.5806 | 0.7845 |
| IgM | LPSB | 0.614 | 0.5766 | 0.6229 | 0.7972 |
| IgM | BPSL2827\_DNAK | 0.5133 | 0.5404 | 0.6046 | 0.9363 |
| IgM | BPSS1652 | 0.6484 | 0.4918 | 0.6951 | 0.7539 |
| IgM | BPSL1404\_ClpX | 0.6127 | 0.6855 | 0.5293 | 0.7532 |
| IgM | BPSS0477\_GroEL2 | 0.621 | 0.5352 | 0.7181 | 0.6863 |
| IgM | BPSS1850 | 0.687 | 0.5841 | 0.6249 | 0.6641 |

For the non-human primates we do something similar, and we utilize the NHP version 2 datasets. Here the single antigen model

data('NHP\_BurkPx\_train2', package='BurkPx')  
data('NHP\_BurkPx\_test2', package='BurkPx')  
  
# Make one data frame that has both the test and training observations  
NHP\_BurkPx <- rbind(   
 NHP\_BurkPx\_test %>% mutate(Set = 'Test'),  
 NHP\_BurkPx\_train %>% mutate(Set = 'Train') )  
  
# Now actually do the work for each antigen!  
NHP\_Results <- rbind(  
 NHP\_BurkPx %>% group\_by(Type, Antigen) %>% filter( TimeGroup %in% c('Prior', 'Healthy', 'Week1') ) %>% mutate(TimeGroup = 'Week1'),  
 NHP\_BurkPx %>% group\_by(Type, Antigen) %>% filter( TimeGroup %in% c('Prior', 'Healthy', 'Week2') ) %>% mutate(TimeGroup = 'Week2'),  
 NHP\_BurkPx %>% group\_by(Type, Antigen) %>% filter( TimeGroup %in% c('Prior', 'Healthy', 'Week3') ) %>% mutate(TimeGroup = 'Week3'),  
 NHP\_BurkPx %>% group\_by(Type, Antigen) %>% filter( TimeGroup %in% c('Prior', 'Healthy', 'Week4+') ) %>% mutate(TimeGroup = 'Week4+') ) %>%  
 group\_by(Type, Antigen, TimeGroup) %>%   
 nest() %>%   
 mutate( AUC = map(data, ind\_model) ) %>%  
 select(-data) %>% unnest(AUC)

NHP\_Results %>%   
 filter( Type == 'IgG' ) %>%  
 spread(TimeGroup, AUC) %>% #group\_by(Antigen) %>%  
 mutate( best = select(., starts\_with('Week')) %>% apply(1, mean, na.rm=TRUE) ) %>%  
 arrange(desc(best)) %>% select(-best) %>%  
 slice(1:10) %>%  
 pander::pander()

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Type | Antigen | Week1 | Week2 | Week3 | Week4+ |
| IgG | LPSA | 0.9507 | 1 | NA | NA |
| IgG | CPS | 0.8867 | NA | NA | NA |
| IgG | MSHR5855.WCL | 0.6478 | 1 | NA | NA |
| IgG | BPSS1498\_HCP1.B | 0.6422 | NA | NA | 0.9511 |
| IgG | BPSS0477\_GroEL2 | 0.5234 | 1 | NA | NA |
| IgG | BPSL1743\_Arg | 0.5382 | 0.8621 | NA | 0.7356 |
| IgG | BPSL2522\_OmpA | 0.5314 | 0.8707 | NA | NA |
| IgG | BPSS1652 | 0.5246 | 0.9655 | NA | 0.6034 |
| IgG | BPSS0135 | 0.5554 | 0.8345 | NA | 0.7004 |
| IgG | BPSL2096\_AhpC | 0.5573 | NA | NA | 0.8348 |

NHP\_Results %>%   
 filter( Type == 'IgM' ) %>%  
 spread(TimeGroup, AUC) %>%  
 mutate( best = select(., starts\_with('Week')) %>% apply(1, mean, na.rm=TRUE) ) %>%  
 arrange(desc(best)) %>% select(-best) %>%  
 slice(1:10) %>%  
 pander::pander()

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Type | Antigen | Week1 | Week2 | Week3 | Week4+ |
| IgM | LPSA | 0.9889 | 1 | NA | NA |
| IgM | CPS | 0.9483 | NA | NA | NA |
| IgM | MSHR5855.WCL | 0.8381 | 0.9948 | NA | NA |
| IgM | BPSL2522\_OmpA | 0.798 | 0.9552 | NA | 0.7924 |
| IgM | LPSB | 0.7155 | 0.9293 | NA | 0.819 |
| IgM | BPSS1652 | 0.5413 | 0.9776 | NA | 0.8218 |
| IgM | BPSL1404\_ClpX | 0.7291 | 0.8655 | NA | 0.6889 |
| IgM | BPSL3222\_rpIL | 0.5616 | 0.9353 | NA | 0.7011 |
| IgM | BPSS1769\_NADH | 0.5917 | 0.8871 | NA | 0.7069 |
| IgM | BPSL1743\_Arg | 0.5468 | 0.8534 | NA | 0.7399 |

# Also make these stratified by time so that we see AUC week1, week2, etc...  
  
# They have IgA saliva data. We want to test how well the multivariate serum models perform for the saliva. Questionable, but worth a shot.  
#   
# For NHP make a version 3 - No day 1-6 animals in test/training.  
# For Humans, make a version 2 - do just positives 0-2 weeks  
#