1 FLU Study PLOTS and CLASSIFICATION Results

These are the 14 genes that were used in the previous rhino and rsv study.

CXCL10 OAS2 CCL2 CCL8 IL1B IL1RN CASP7 CCR1 CCR10 TAP1 TAP2 CASP5 TNFAIP6 TNFSF10

We used this same set on the FLU data to create plots. Results of running classifier will come later.

Color coding:

- Green: Asymptomatic
- Blue : Mildly symp.
- Red : Medium symp.
- Black : Strong symptoms

The dash and line form in the plots are chosen so as to uniquely identify each subject and hence, they are not relevant.

2 Classification Results

Same classifier as before. Detailed results using different starting and end timepoints for doing the classification: timepoints | subjects classified "symp" | Bemarks

ints for doing the classification:				
timepoints	subjects classified "symp"	Remarks		
t0 - t21	none			
t0 - t29	$1,\!6$			
t0 - t36	$1,\!6,\!7,\!12$			
t0 - t45	1,5,6,7,12			
t0 - t53	1, 5, 6, 7, 8, 12, 13	13 barely makes it		
t0 - t60	1, 5, 6, 7, 8, 10, 12, 13	$10,13$ get equal my_score		
t0 - t69	1,5,6,7,8,10,12,13			
t0 - t77+	1,5,6,7,8,10,12,13,15	15 gets lowest my_score		

As more timepoints become available, the classifier can see more "overexpression signal" and hence can classify more and more subjects as being symptomatic. For example, consider the entry for t0–t36. The above table shows that using timepoints t0 to t36, Subjects 1,6,7,12 are classified "symp". These are indeed "symp" (as the table below shows). However, Subjects 5,10 are also "symp" (see below), but they are missed if we only look at t0–t36. All subjects that are not mentioned in the table above are classified as "asymp" (or, as "don't-know" depending on how you want to interpret the result of the classifier).

As per Carolyn's symptoms1.txt, subjects 1,5,6,7,8,10,12 are deemed symptomatic. If you look at the detailed symptom data, you find that Subjects 13

and 15 fall between those who are clearly "symp" and those who are clearly "asymp". Specifically, Subjects 13 and 15 had non-zero symptom measure – their total sum was 10 and 9 respectively – whereas those who were symp has total sum greater than 18 and those who were asymp had sum less than 5. Detailed sums for all subjects:

subjects	range of total sum	classification
5, 6, 7, 8	50 - 76	symp
1	39–39	symp
10,12	18 - 21	medium symp
13,15	9-10	?
9,11	5 - 7	?
2, 3, 4, 14, 16, 17	0-1	asymp

In conclusion, the over-expression of these 14 genes appears to be correlated with the (intensity) of observed symptoms. Unfortunately, they are not early indicators.

Table showing the earliest time when we detect a subject as being symp and that value as a percent of the PEAK symptom time T (as per Carolyn's symptoms1.txt file).

subject	earliest detection time	Т	percent
1	29	62	.46
5	45	98	.46
6	29	50	.58
7	36	62	.58
8	53	74	.71
10	60	110	.55
12	36	86	.41













