## Initial Results for Flu Data

Minhua Chen 11.27.2008

## Data description

- Gene data collected for 17 subjects at 16 time points.
- Ten subjects are labeled as healthy and seven labeled as influenza infected.
- Later we will see that 2 healthy subjects (13 and 15) are suspicious, since their gene patterns are more like sick people.
- Total samples: 267 Total genes: 12023.

Time	baseline	0 hrs	5 hrs	12 hrs	21.5 hrs	29 hrs	36 hrs	45.5 hrs	53 hrs
Samples	9 + 7	10 + 7	10 + 7	10 + 7	10 + 6	9 + 7	8 + 7	10 + 7	10 + 7
Time	60 hrs	69.5 hrs	77 hrs	84 hrs	93.5 hrs	101 hrs	108 hrs		
Samples	10 + 7	10 + 7	10 + 7	10 + 7	10 + 7	10 + 7	10 + 7		

9 + 7 means 9 healthy samples and 7 influenza infected.



All factors found. Samples are sorted from early time to late time. X-axis: sample index; Y-axis factor scores. Red: healthy people; Blue: influenza infected.



Gene list:

'ATF3' 'SERPING1' 'ISG15' 'LAMP3' 'RSAD2' 'IFIT1' 'IFI44L' 'HERC5' 'IFI44' 'CCL8' 'OAS3' 'CCL2' 'OASL' 'IFIT3' 'IFIT2' 'LOC26010' 'SIGLEC1' 'IFI6' 'MX1' 'INDO' 'RTP4' 'XAF1' 'IFIH1' 'OAS1' 'OAS2' 'DDX58' 'CXCL10' 'DDX60' 'GBP1' 'SOCS1' 'LY6E' 'IFI35' 'HERC6' 'LAP3' 'IFI27' 'ZBP1' 'IFIT5' 'IRF7' 'CASP5' 'TNFAIP6' 'APOL6' 'EIF2AK2' 'IFITM3' 'CEACAM1' 'PARP12' 'DHX58' 'AIM2' 'TFEC' 'ETV7' 'TIMM10'

- The gene list agrees well with results for Rhino and RSV data.
- There are some outliers at late time. But as can be seen in the next slide, all the outliers are from subject 13 and 15, indicating that it might be more reasonable to label these two subjects as sick people. From the symptom score table (not shown here), we can also see that these two subjects have mild symptoms at late time.
- Factor 6 also shows some discriminative power. Genes in this factor: 'CDKN1C' 'MS4A4A' 'SIGLEC1' 'C1QB' 'PI3' 'C1QA' 'EPB41L3' 'RNASE2' 'C2' 'RIN2' 'CD68'...



We can see that people 13 and 15 behave more like sick people. They contributes to the outliers in this factor at late time.



45.5 hours corresponds to  $0.6T \sim 0.7T$ . This means that factor analysis can not do early time detection.



## Joint factor analysis for Rhino, RSV and Flu.

Rhino, RSV and Flu data are put together. First Rhino, then RSV, then Flu. Samples within each dataset are sorted from early time to late time. X-axis: sample index; Y-axis factor scores. Red: healthy people; Blue: influenza infected.



Gene list:

'IFI44L' 'IFIT1' 'IFI44' 'HERC5' 'OAS3' 'ISG15' 'SERPING1' 'OASL' 'LAMP3' 'RSAD2' 'IFI27' 'MX1' 'CCL2' 'CCL8' 'LOC26010' 'SIGLEC1' 'IFIT3' 'IFI6' 'IFIT2' 'CXCL10' 'OAS2' 'DDX58' 'OAS1' 'ATF3' 'XAF1' 'IFIH1' 'INDO' 'RTP4' 'DDX60' 'IFITM3' 'HERC6' 'LAP3' 'IRF7' 'LY6E' 'IFI35' 'GBP1' 'IFIT5' 'SOCS1' 'DOCK4' 'TNFAIP6' 'PARP12' 'ZBP1' 'P2RY14' 'EIF2AK2' 'DHX58' 'AIM2' 'PLSCR1' 'CASP5' 'APOL6' 'CEACAM1'



Factor 10 displayed with original label.

Factor 10 displayed with modified label. Change labels for subject 13 and 15 in Flu data from healthy to sick.





## Gene list:

'C1QA' 'SIGLEC1' 'PLXNB2' 'ALDH1A1' 'C1QB' 'EPB41L3' 'MS4A4A' 'CDKN1C' 'IFI44L' 'FER1L3' 'KLF4' 'CD36' 'SERPING1' 'TBC1D8' 'CD1D' 'MAFB' 'CSF1R' 'CD68' 'CLMN' 'CYFIP1' 'TNS3' 'CEBPA' 'SMPDL3A' 'SLC7A7' 'RIN2' 'NAGA' 'IGSF2' 'BLK' 'MTMR11' 'C3AR1' 'C2' 'TCF7L2' 'PI3' 'CARD9' 'SASH1' 'RTN1' 'IFI44' 'HMOX1' 'OAS1' 'DPYSL2' 'CAMK1' 'CX3CR1' 'PID1' 'TCN2' 'KCTD12' 'NID1' 'CPVL' 'VCAN' 'CD163' 'FBP1'

Factor 11 also shows some discriminative power at late time for the flue data, but not so clear.