





UM RSV Challenge Data Analysis Progress

Sept 16, 2008

A Hero, Y, Huang, A. Puig, P. Harrington Dept of EECS and Bioinformatics Program University of Michigan

Summary

1. FDR screening – discovery of sustained sx/asx differential expression TFs over post challenge time points 0.1T, 0.2T, 0.8T. Using 2 or 3 top ranked genes can achieve high prediction accuracy with naïve Bayes quadratic discriminant.

2. EDGE time course studies on 4 densely sampled subjects – three pathways (TLR/T-Cell/Cell-Adhesion) are highly expressed.

3. Random forest non-linear regression - achieves 93% current state estimation accuracy (12 missclassified chips out of 164). Prediction performance is still being investigated.

4. RSV vs HRV – 129 common genes expressed. 2 differentially

Persistency Screening

We adopt the classical **Behrens-Fisher** model for probes:

- •Gaussian distributed sample mean over asx and sx groups
- •Different size groups at each time point
- •Population variances may be different
- •Chips are conditionally independent given population mean/ variance

Known class labels

Persistently differentially expressed genes g will satisfy
min_t { |sx_t(g)-asx_t(g)| } > x
for some x>0 that can be specified by experimenter

Persistency screening

- Screen over all post-inocculation symptom times 0.1T to T
 212 genes significant at FDR of 20%
- 2. Screen over 0.1T 0.2T and 0.8T4 genes significant at FDR of 20%
- 3. Screen over 0.1T, 0.2T or at 0.1T alone No genes significant at FDR of 20%

Probes discovered at FDR of 20%

Persistent over **0.1T to T** 'TSPAN3' 'DNAL4' 'AP3S2' 'GPHN' 'CNPY2' 'CORO2B' 'ADARB1' 'TIMM44' **'TRIM38'** 'CEBPG'

Persistent over 0.1T to 0.8T 'CD177' 'TXNDC3' 'CD72' 'UPK2' These are (custom cdf) '57126_at' '51314 at' '971 at' '7370 at'

nBayes quadratic discriminant

The 0.1T-0.8T genes achieve the following prediction performance over the three benchmark times sx

0.1T	0.9091	0.8000	0.1T	0.9091	0.6622
0.2T	0.9091	0.9811	0.2T	0.9000	0.9489
0.8T	0.9127	0.9978	0.8T	0.8264	0.9967

'CD177' 'TXNDC3' 'CD72' 'CD177' 'CD72'

NB: 5-fold cross validation over 100 resampling trials



nBayes quadratic discriminant performance (CD172, TXNDC3, CD72): false positives(b) and negatives(r)







Temporal Differential Expression Analysis and Prediction

Materials

- 1. Gene expression profiles from subject #4, #19, #2, #17 (n=79)
- 2. Total of 5 missing time points are imputed by averaging two adjacent time points from same subject (#04:BL; #04:H69; #19:H132; #7:H0; #7:H45)

Methods

- 1. Differential expression analysis is performed using Extraction of Differential Gene Expression (EDGE)
- 2. Prediction model is constructed using Random Forest (RF)

Results

- 3. Total of 2,236 genes are differentially expressed temporally with significance level q-value < 1%
- 4. Prediction model selected using random forest achieved **93%** of out-of-bag accuracy on all 164 samples with mis-prediction of 12 samples
- 5. Two most important predictor genes are CD72 and CD177, in concordance with Duke group findings

Significant Genes with q-value < 1% (n=2,236) From Time Series Analysis (EDGE) On 4 Densely Sampled Individuals

-056 -0.49 -0.42 -0.42 -0.23 -0.16 -0.10 -0.16 0.19 0.19 0.19 0.23 0.23 0.23 0.49 0.42 0.42 0.49 0.45 0.45 0.45 0.55

Apoptosis / Toll-like receptor signaling / cytokine receptor interaction /L1B, FAS, TNFSF10, IRAK3, STAT1, LY96, CXCL10, TLR5, IRF7, IL151, CCR1, IL18RAP

T cell receptor signaling / E.coli infection / Fc epsilon RI signaling *ACTG1*, *AKT1*, *CD40LG*, *FYN*, *HCLS1*, *IKBKB*, *KRT18*, *NCL*, *NFATC1*, *NFATC3*, *PIK3R1*, *PLCG1*, *PLCG2*, *PPP3R1*, *PRKCA*, *PRKCB1*, *MAPK13*, *PTPRC*, *TUBA4A*, *PLA2G6*, *MAP3K14*, *PLA2G2D*, *ARPC5L*, *TUBB*

Cell adhesion molecules (CAMs) / Antigen processing and presentation / Toll-like receptor signaling / Fc epsilon RI signaling B2M, BTK, CD40, CD86, CTSB, CTSL1, FCER1G, HLA-DMA, HLA-DMB, HLA-DPA1, HLA-DRA, HLA-DRB1, HLA-G, ICAM1, IFI30, IRAK4, IRF5, ITGA4, JAM3, MAP2K4, MAP2K6, MAPK14, MPZL1, MYD88, NRAS, PDCD1LG2, PECAM1, PLA2G4A, PSME1, PSME2, PVRL2, RAF1, SDC2, SELL, SELP, TAP1, TAP2, TBK1, TICAM1, TLR4, TLR7, TNF, VCAN



Prediction Model Selection Using Random Forest With Significant Genes (q-value < 1% (n=2,236) from Time Series Analysis (EDGE)) Variable Importance Plot **Confusion matrix:** CD72 CD177 Sx Class.error Asx 0 7.0% Asx 80 6 TLR5 ĊD200 Sx 6 72 7.7% TCL1A 0 EPB41L3 KIAA0746 ···· MICAL3 C14orf94 **Misclassified individual chips** HLA-DOB GZMK 0 CD79A Class Class ····· 0····· Subject Time Sx Asx Predicted True KIAA0125 FCER2 0 R06 H00 55.8% 44.2% Asx Sx **TMEM144** •••••••• TBC1D8 0 R11 H00 76.3% 23.7% Sx Asx **OSBPL10** •••••• R06 H012 55.6% 44.4% Asx Sx C11orf75 CBX7 0 ••••• R09 H012 41.1% 58.9% Sx Asx ALDH1A10 R14 H012 54.7% 45.3% Sx Asx C1QA 0 SECTM 0 R14 H021 64.1% 35.9% Asx Sx CD22 0 P2RX5 0 R11 H029 52.9% 47.1% Sx Asx ZNF266 0 R17 H029 30.5% 69.5% Sx Asx HMOX2 0 PLEKHA1 0 R18 H029 44.0% 56.0% Sx Asx SERPING1 0 0 R05 H117 40.7% 59.3% Sx Asx LRRN3 PASK 0 R16 H117 16.8% 83.2% Sx Asx R17 H117 33.3% Sx 66.7% Asx

Densely Sampled Individuals – Comparing RSV vs. HRV Significant Genes

- 1. RSV initiates relatively stronger host response that can be detected at mRNA expression level. Expression of more genes (>12 fold) are altered significantly over time.
- 2. Significant number of differential genes appear in both two challenges, suggesting similar immune machinery is deployed by hosts in the two studies.





Sample Summary Statistics



Subject symptom scores

blood draw time	ONCE			0/5/1	2hrs 21.5 hrs	29 hrs	/36h45.5hrs	53h 60h	69.5 h	77h 84	h 93 h	101h	108h117.5 h	125h 13	32 h 141.5 h	ı	165.51	i	
Subject	minus1	m	inus1barminus	s1bpr0am	1 am	1pm	2am	2pm	3am	3pm	4am	4pm	5am	5pm	6am	6pm	7am	7pm	CLASS
RSV001		0	0	0	0	0	1	0	0	0	0	2	3	5	6	8	11	10	10 YES
RSV002		0	0	0	0	1	2	2	1	2	3	12	13	18	16	20	20	13	13 YES
RSV003		0	0	0	0	1	0	0	0	0	1	2	2	2	2	6	4	4	4 YES
RSV004		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0 NO
RSV005		0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0 No
RSV006		0	0	0	0	0	0	0	0	0	0	1	2	6	6	8	7	9	9 YES
RSV007		0	0	0	0	0	0	0	2	6	10	20	19	14	12	15	12	12	12 YES
RSV008		0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0 Yes
RSV009		0	0	1	0	1	1	0	1	0	0	0	0	0	0	0	0	0	0 No
RSV010		0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0 NO
RSV011		0	0	0	0	0	0	0	0	0	0	1	1	2	3	5	5	2	2 YES
RSV012		0	0	0	0	0	0	0	0	0	0	2	1	4	2	4	3	4	4 YES
RSV013		0	0	0	0	1	1	0	0	0	0	0	0	0	1	0	1	0	0 NO
RSV014		0	0	0	0	0	0	0	0	0	0	2	2	6	5	4	2	5	5 YES
RSV015		0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	1 NO
RSV016		1	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	1	1 NO
RSV017		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0 NO
RSV018		0	0	0	0	0	1	0	1	1	1	2	1	0	0	0	0	1	1 NO
RSV019		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0 NO
RSV020		0	0	0	0	2	0	0	1	1	0	0	1	2	1	1	6	6	6 YES