

Sparse Probit Regression on RSV data

Minhua Chen

09.08.2008

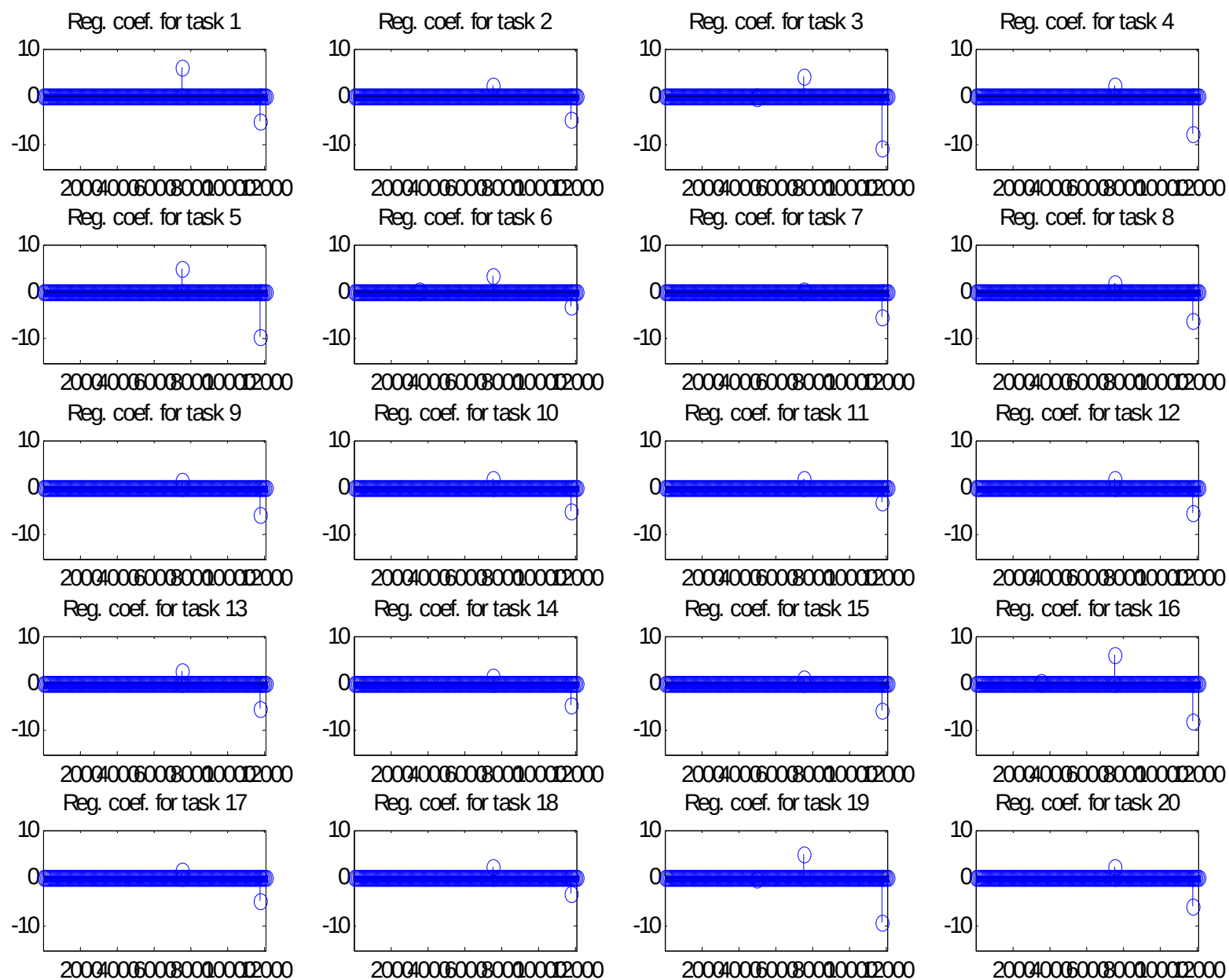
Experimental setup

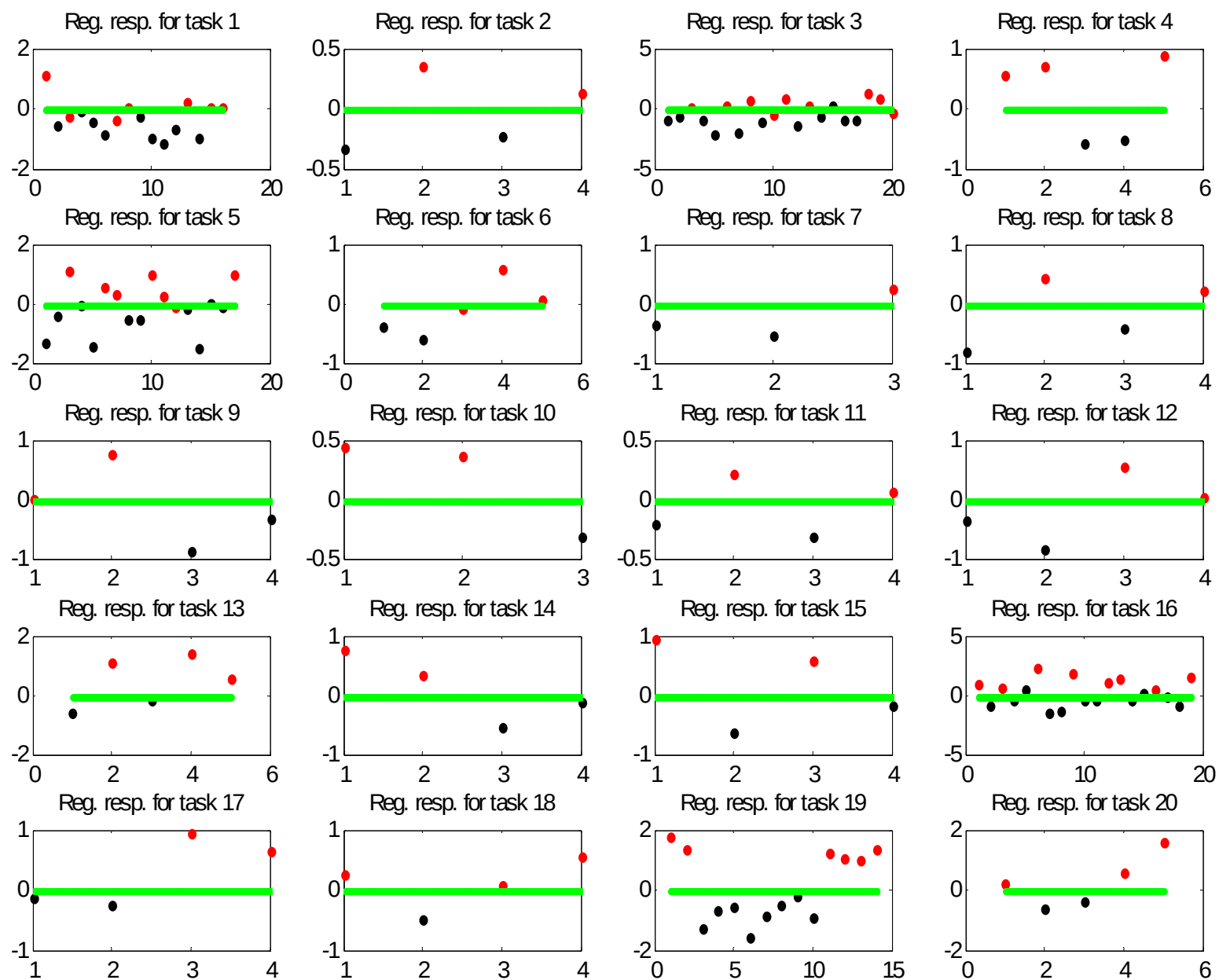
- RSV data with 12023 gene features.
- 164 samples collected at 21 time points.
- All 12023 genes and all samples except baseline are used, resulting in 20 tasks.
- Multi-task sparse probit regression with global sharing.

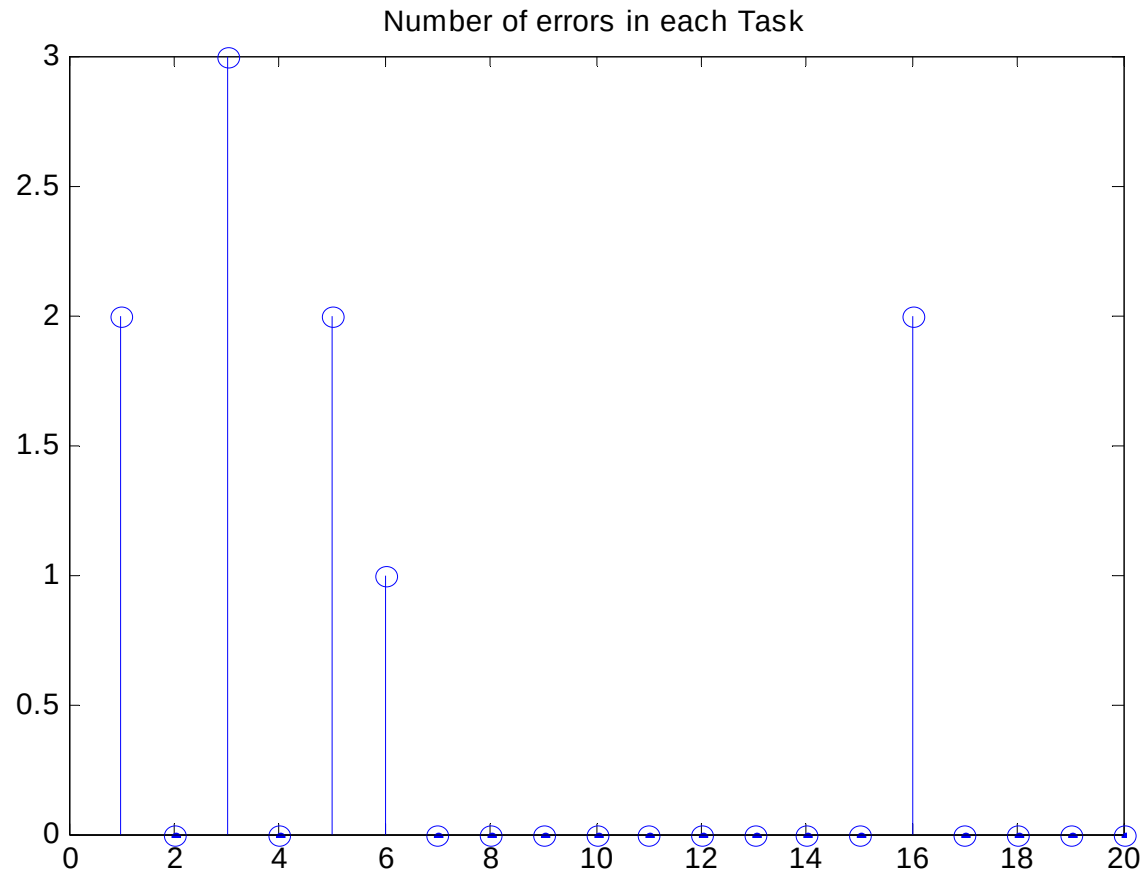
Time (hrs)	0	1	2	3	4	5	6	7	8	9	10
samples	9+7	9+7	2+2	11+9	2+3	10+7	2+3	2+1	2+2	2+2	1+2
Time (hrs)	11	12	13	14	15	16	17	18	19	20	
samples	2+2	2+2	2+3	2+2	2+2	11+8	2+2	1+3	8+6	2+3	

Red : healthy samples

Blue: RSV infected samples



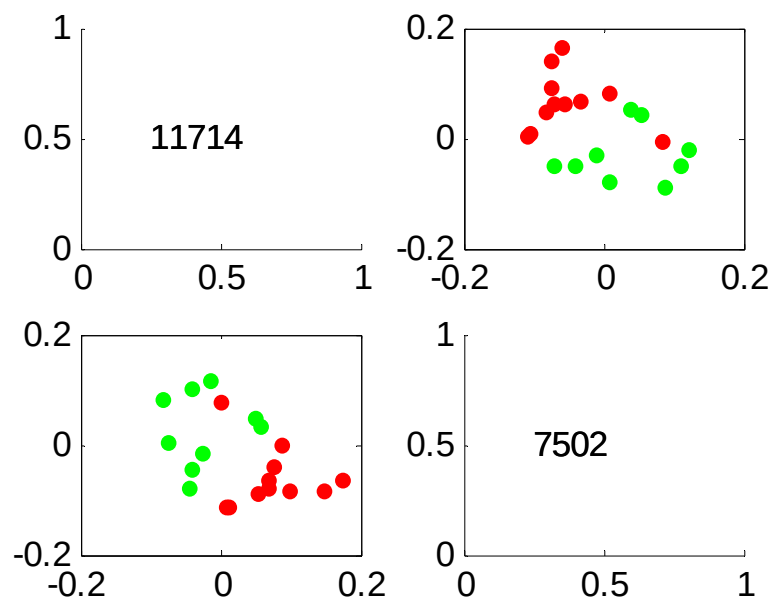




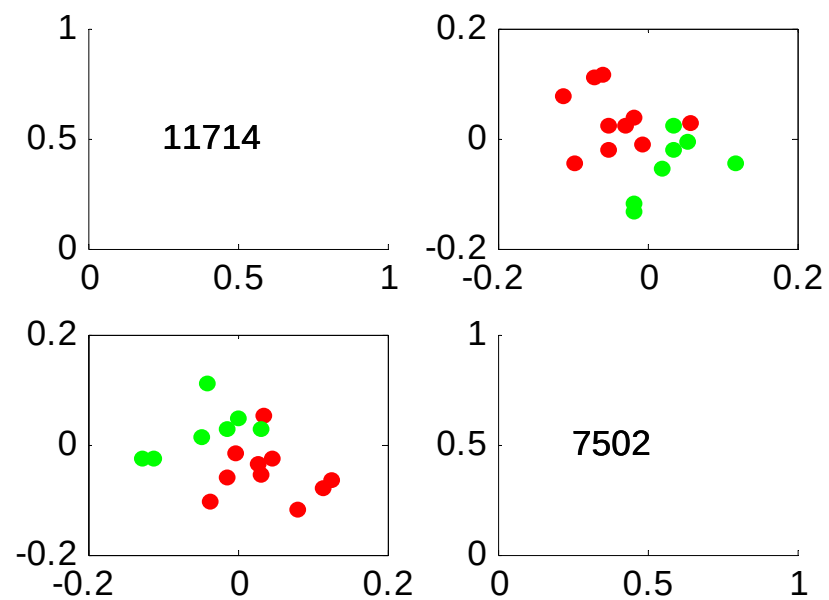
Genes: (First two are manifested in the reg. coef.)

'971_at' (11714) '57126_at' (7502)

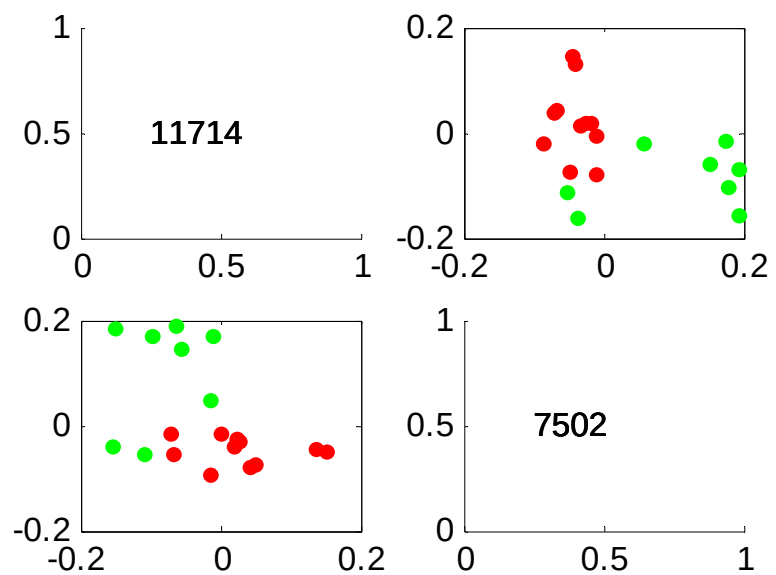
However, these two genes are **not** in the list of factor analysis.
It seems that these two genes are not biologically meaningful.



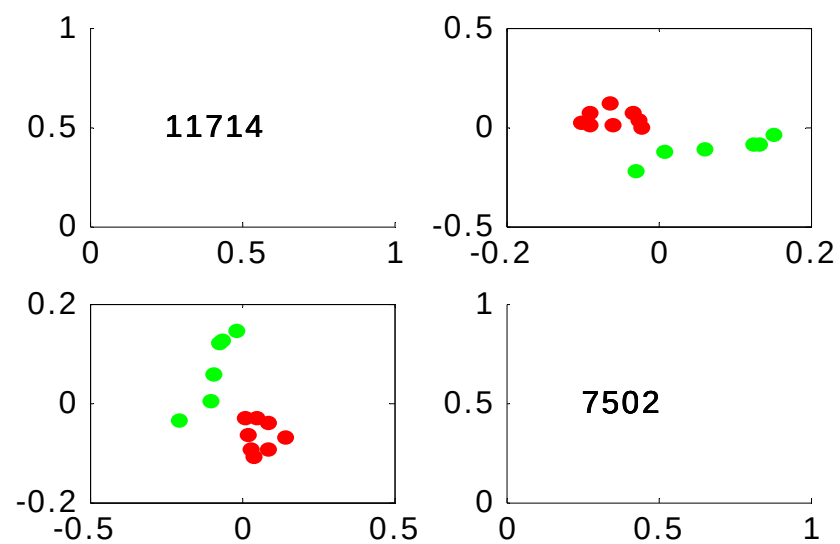
Gene scatter of task 3



Gene scatter of task 5



Gene scatter of task 16



Gene scatter of task 19

Initial result of RSV using factor analysis

Minhua Chen

09.02.2008

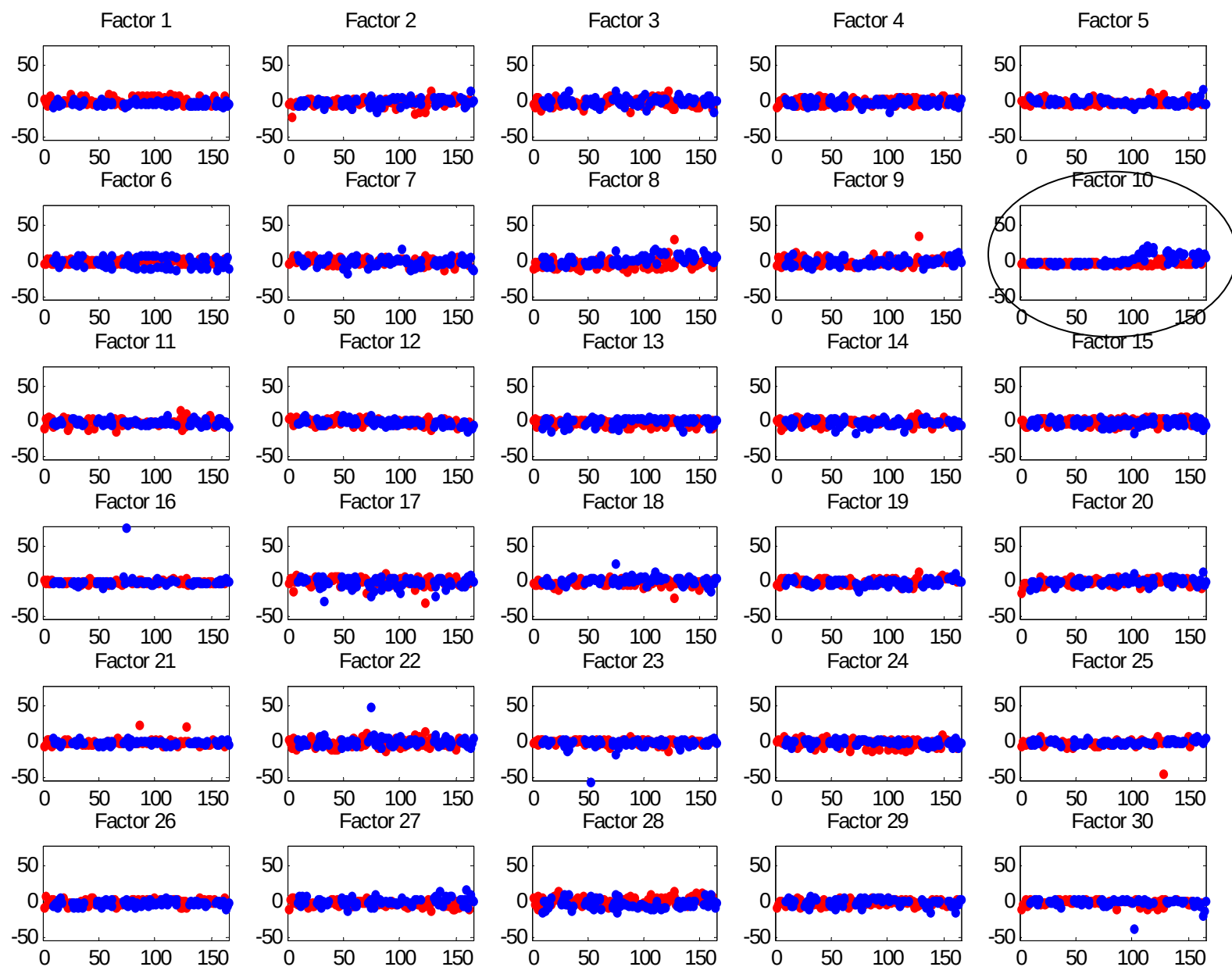
Experimental setup

- RSV data with 12023 gene features.
- 164 samples collected at 21 time points.
- All 12023 genes and 164 samples are used.
- Set factor number to be 30.

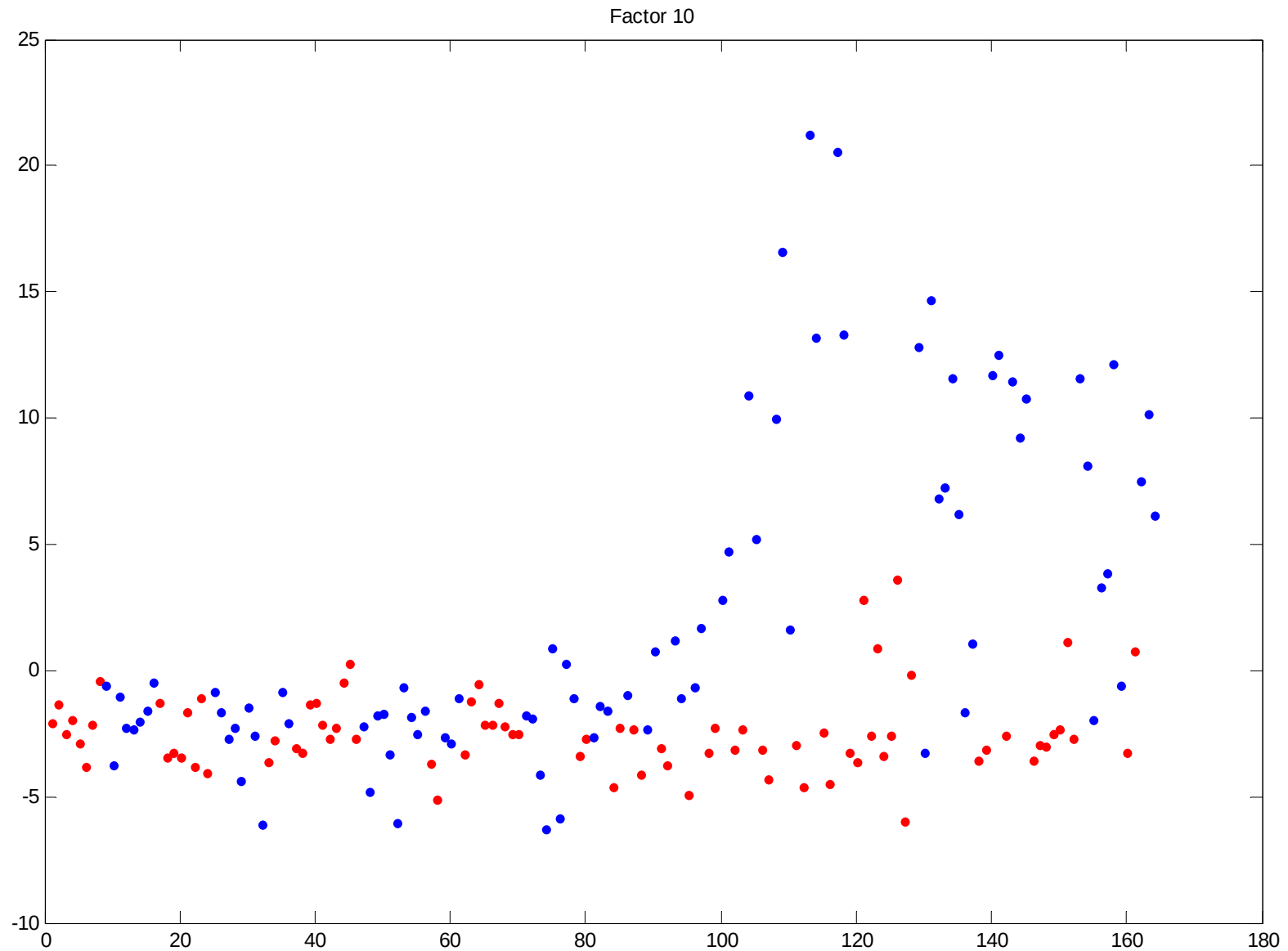
time	t1	t2	t3	t4	t5	t6	t7	t8	t9	t10	t11
samples	8+8	8+8	2+2	10+10	2+3	9+8	2+3	2+1	2+2	2+2	1+2
time	t12	t13	t14	t15	t16	t17	t18	t19	t20	t21	
samples	2+2	2+2	2+3	2+2	2+2	10+9	2+2	1+3	7+7	2+3	

Red : healthy samples

Blue: RSV infected samples



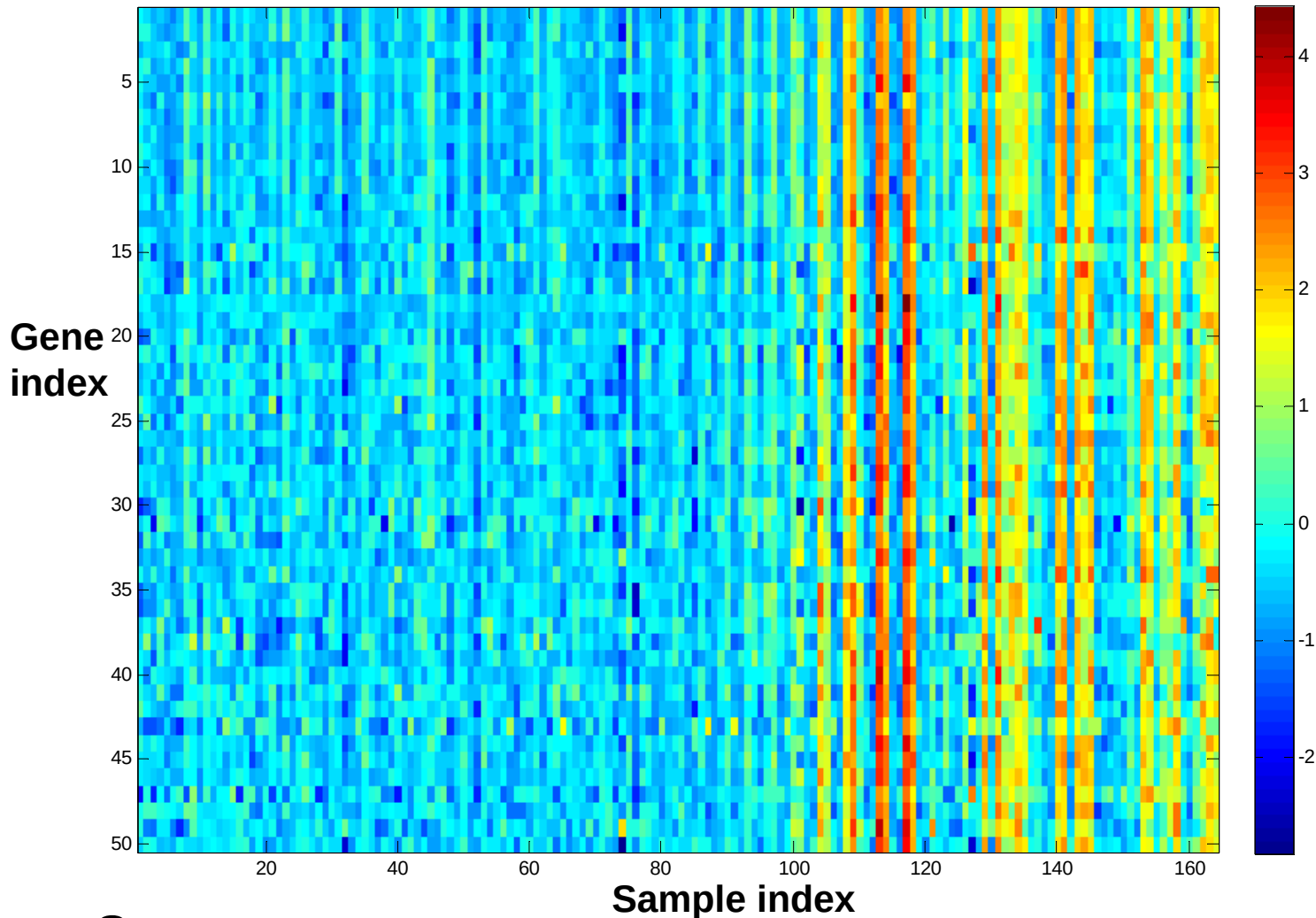
X-axis is sample index, from t1 to t21, red healthy, blue RSV infected



This is the discriminative factor found.
It starts to be discriminative around t8.
Still, there are some outliers at later times.

Gene expressions corresponding to factor 10

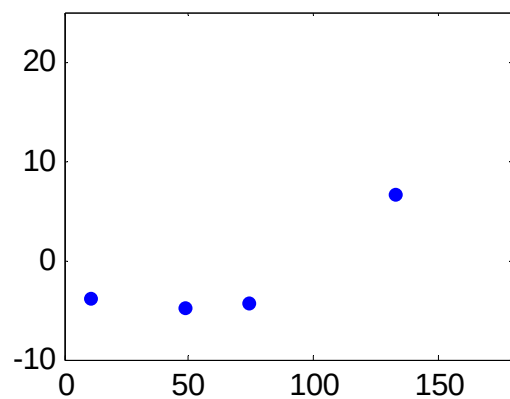
Corresponding Gene Expression Values (Normalized)



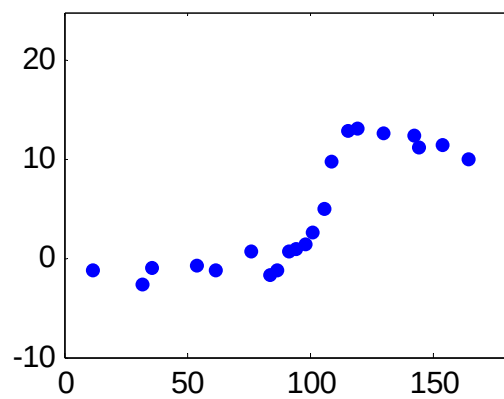
Genes:

RSAD2 IFI44L SERPING1 IFI44 LAMP3 IFIT1 HERC5 OAS3 ISG15
OASL 3437_at DDX58 GBP1 LOC26010 P2RY14 IFI6 MX1 467_at
RTP4 XAF1 IFIH1 LAP3 FLJ20035 OAS1 IFIT5 SIGLEC1 IFIT2
IFI35 OAS2 ...

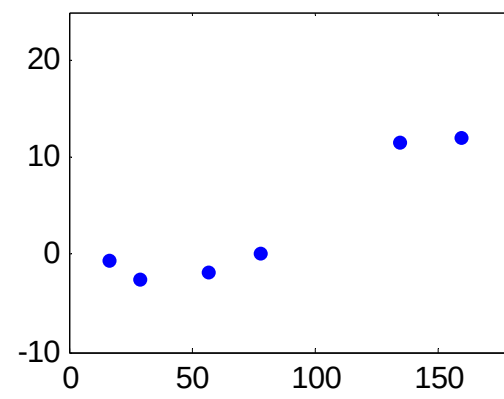
time traj. of person 1 in factor 10



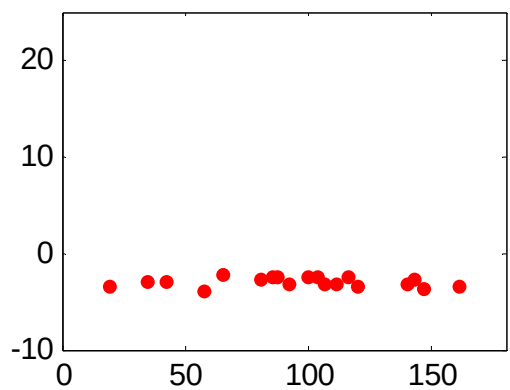
time traj. of person 2 in factor 10



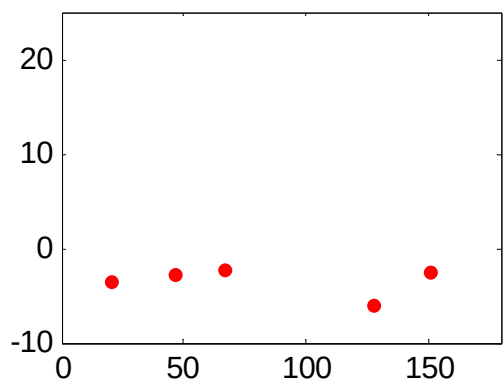
time traj. of person 3 in factor 10



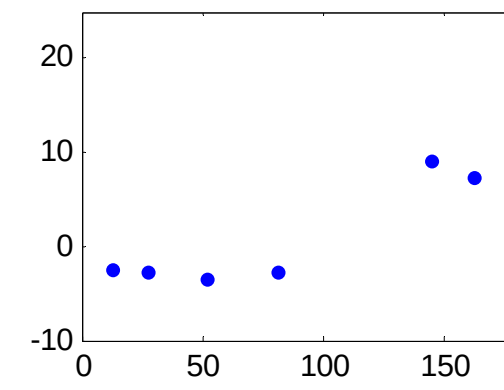
time traj. of person 4 in factor 10



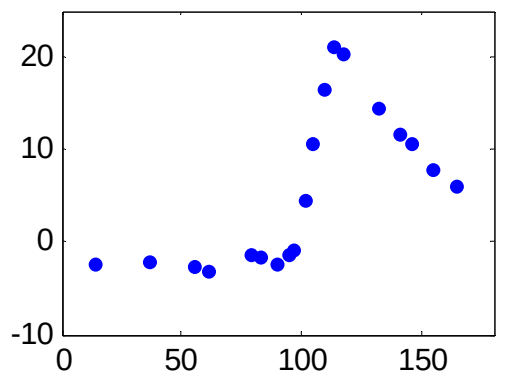
time traj. of person 5 in factor 10



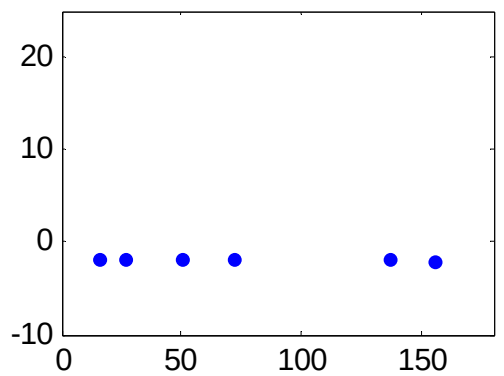
time traj. of person 6 in factor 10



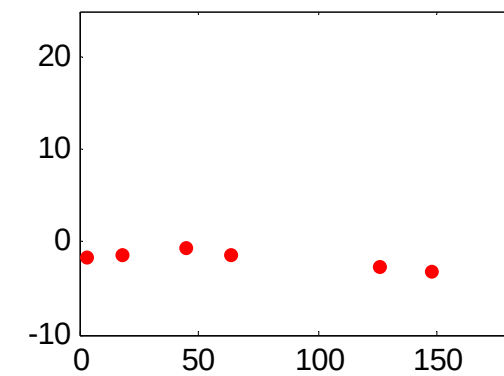
time traj. of person 7 in factor 10



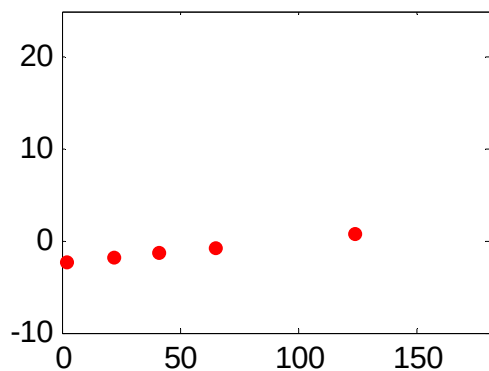
time traj. of person 8 in factor 10



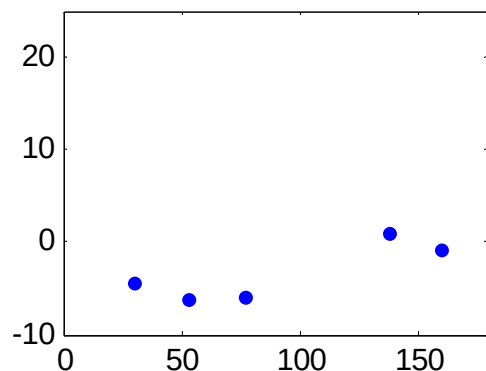
time traj. of person 9 in factor 10



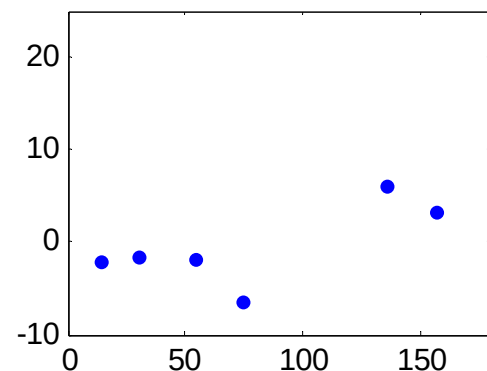
time traj. of person 10 in factor 10



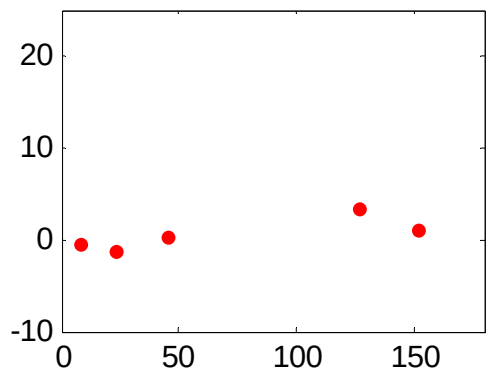
time traj. of person 11 in factor 10



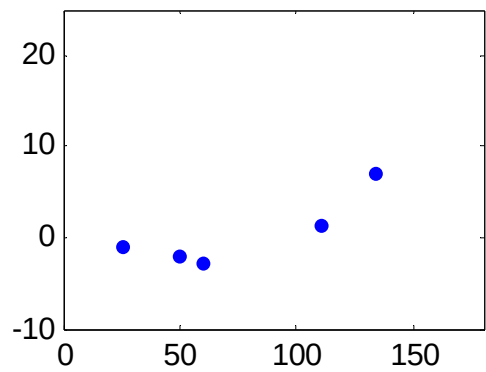
time traj. of person 12 in factor 10



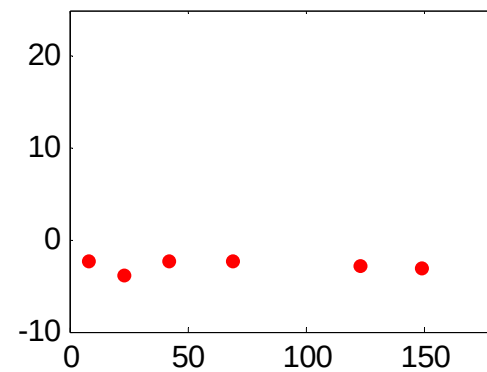
time traj. of person 13 in factor 10



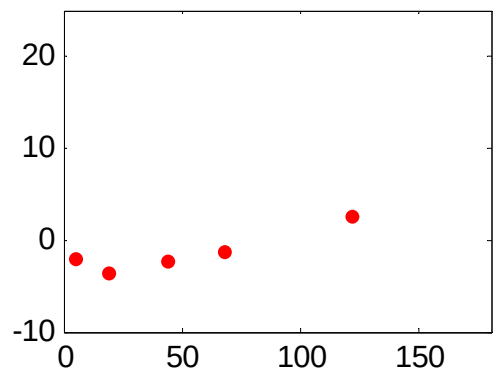
time traj. of person 14 in factor 10



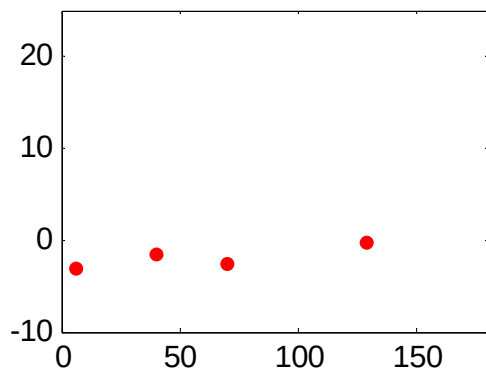
time traj. of person 15 in factor 10



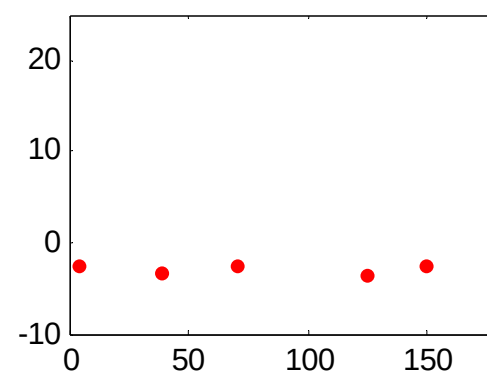
time traj. of person 16 in factor 10

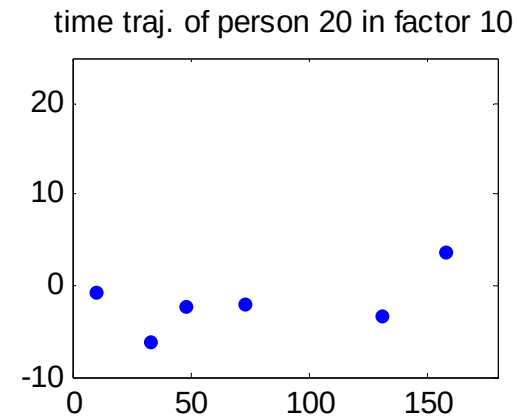
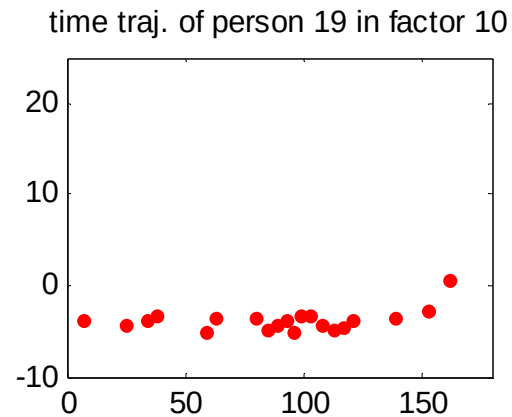


time traj. of person 17 in factor 10



time traj. of person 18 in factor 10





Summary

- One discriminative factor identified.
- This factor discriminate samples well for time points $> t_8$. Still there are some outliers at later time points.
- The genes associated with the discriminative factor are consistent with that found for Rhino and Flu data.
- Time trajectories of this factor for individual persons are plotted.