UM RSV Challenge Data Analysis Preliminary findings

Sept 9, 2008

- A Hero, Y, Huang, A. Puig, P. Harrington

Dept of EECS and Bioinformatics Program
University of Michigan

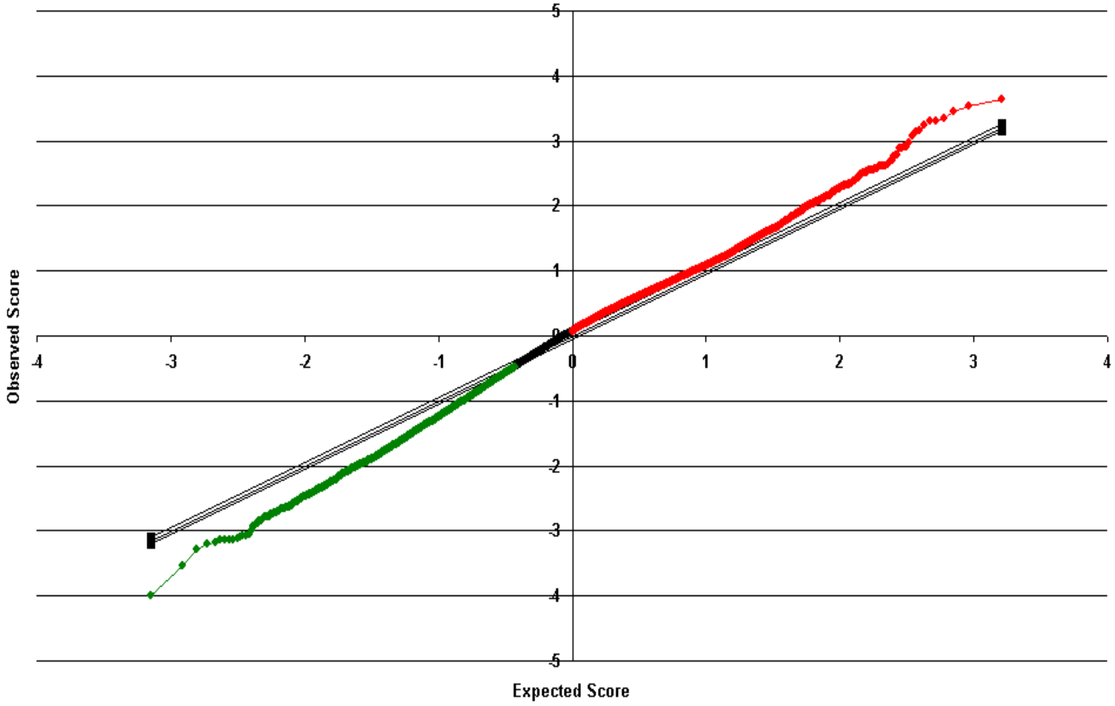
Summary

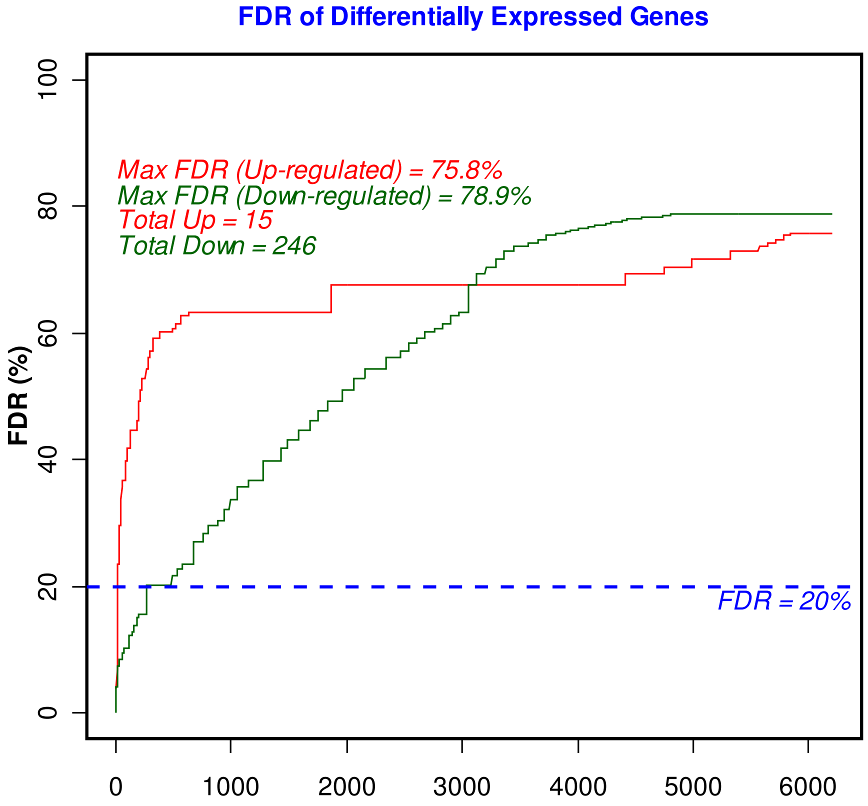
1. PAM - Univariate response model (fixed effects):
A combination of 17 mRNA probes learned from BL and PC can achieve better than 85% prediction accuracy at 0.1T, 0.2T, 0.8T

2. Random forests – Non-linear regression model:
Results not in yet

3. DMDA - Multivariate response model:
A combination of 8 mRNA probes learned from BL and PC can achieve better than 85% prediction accuracy at 0.1T, 0.2T, 0.8T and T

Differential Expression Analysis At Baseline / Pre-Challenge
Asymptomatic (n=18) vs. Symptomatic (n=14)





PAM Prediction Model of 17 Genes
Trained From Significant Genes (q-value<20%)

probesets	symbol	chr	function	relevance
54674_at	LRRN3	7q31.1	leucine rich repeat neuronal 3	
57126_at	CD177	19q13.2	CD177 molecule	
23178_at	PASK	2q37.3	PAS domain containing serine/threonine kinase	
7280_at	TUBB2A	6p25	tubulin, beta 2A	Involved in Ecoli infection
79368_at	FCRL2	1q21	Fc receptor-like 2	
3112_at	HLA-DOB	6p21.3	major histocompatibility complex, class II, DO beta	Antigen processing and presentation
10516_at	FBLN5	14q32.1	fibulin 5	
8115_at	TCL1A	14q32.1	T-cell leukemia/lymphoma 1A	Expressed only in naïve B cells and regulates early stage B-cell and T-cell differentiation
9834_at	KIAA0125	14q32.33	KIAA0125	
5023_at	P2RX1	17p13.3	purinergic receptor P2X, ligand-gated ion channel, 1	
10391_at	CORO2B	15q23	coronin, actin binding protein, 2B	
26228_at	STAP1	4q13.2	signal transducing adaptor family member 1	B cell receptor signaling
9648_at	GCC2	2q12.3	GRIP and coiled-coil domain containing 2	
4907_at	NT5E	6q14-q21	5'-nucleotidase, ecto (CD73)	
11138_at	TBC1D8	2q11.2	TBC1 domain family, member 8 (with GRAM domain)	
208_at	AKT2	19q13.1-q13.2	v-akt murine thymoma viral oncogene homolog 2	B cell receptor signaling
930_at	CD19	16p11.2	CD19 molecule	B cell receptor signaling (marker)

Predicting Infection Outcome with 17-gene Predictor --- 0.1T

0.1T Subject	**Time (hrs)**	**Class True**	**Class Predicted**	**Asx (Posterior)**	**Sx (Posterior)**
4	12	Asx	Asx	0.80	0.20
5	12	Asx	Asx	0.57	0.43
8	12	Asx	Asx	0.74	0.26
9	12	Asx	Asx	0.51	0.49
10	12	Asx	Asx	0.75	0.25
13	12	Asx	Asx	0.79	0.21
15	12	Asx	Asx	0.85	0.15
16	12	Asx	Asx	0.71	0.29
17	12	Asx	Asx	0.67	0.33
18	12	Asx	Asx	0.66	0.34
19	12	Asx	Asx	0.66	0.34
1	12	Sx	Sx	0.44	0.56
2	12	Sx	Sx	0.46	0.54
3	12	Sx	Sx	0.32	0.68
6	12	Sx	Asx	0.55	0.45
7	12	Sx	Sx	0.45	0.55
11	12	Sx	Sx	0.41	0.59
12	12	Sx	Sx	0.27	0.73
14	12	Sx	Sx	0.43	0.57
20	12	Sx	Asx	0.52	0.48

	Asx	**Sx**
0.1T	***11/11***	***7/9***
0.2T	***7/10***	***9/9***
0.8T	***10/11***	***9/9***

Predicting Infection Outcome with 17-gene Predictor --- 0.2T

0.2T Subject	Time (hrs)	Class True	Class Predicted	Asx (Posterior)	Sx (Posterior)
4	29	Asx	Asx	0.56	0.44
5	29	Asx	Sx	0.45	0.55
8	29	Asx	Asx	0.71	0.29
9	29	Asx	Asx	0.52	0.48
10	29	Asx	Asx	0.69	0.31
15	29	Asx	Asx	0.78	0.22
16	29	Asx	Asx	0.53	0.47
17	29	Asx	Sx	0.48	0.52
18	29	Asx	Sx	0.44	0.56
19	29	Asx	Asx	0.58	0.42
1	29	Sx	Sx	0.35	0.65
2	29	Sx	Sx	0.26	0.74
3	29	Sx	Sx	0.31	0.69
6	36	Sx	Sx	0.49	0.51
7	21	Sx	Sx	0.49	0.51
11	29	Sx	Sx	0.48	0.52
12	29	Sx	Sx	0.25	0.75
14	21	Sx	Sx	0.49	0.51
20	29	Sx	Sx	0.21	0.79

	Asx	Sx
0.1T	***11/11***	***7/9***
0.2T	***7/10***	***9/9***
0.8T	***10/11***	***9/9***

Note that #18 has symptom score 1 at 0.2T

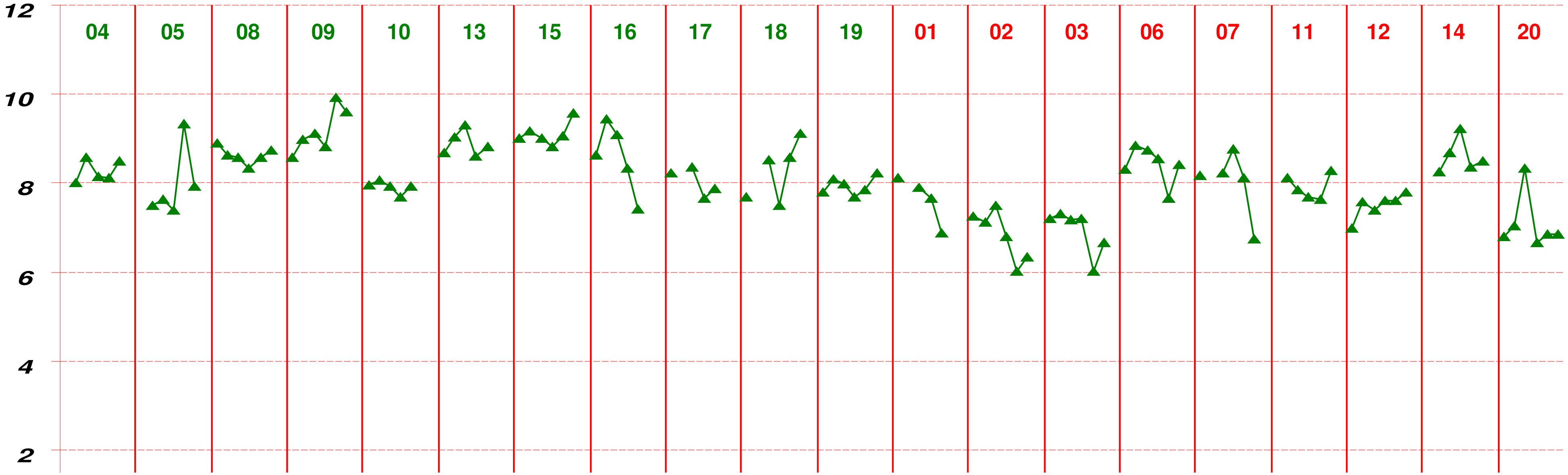
Predicting Infection Outcome with 17-gene Predictor --- 0.8T

0.8T Subject	Time (hrs)	Class True	Class Predicted	Asx (Posterior)	Sx (Posterior)
4	117	Asx	Asx	0.68	0.32
5	117	Asx	Asx	0.64	0.36
8	117	Asx	Asx	0.73	0.27
9	117	Asx	Asx	0.76	0.24
10	117	Asx	Asx	0.70	0.30
13	117	Asx	Asx	0.70	0.30
15	117	Asx	Asx	0.81	0.19
16	117	Asx	Sx	0.31	0.69
17	117	Asx	Asx	0.54	0.46
18	117	Asx	Asx	0.64	0.36
19	117	Asx	Asx	0.60	0.40
1	117	Sx	Sx	0.11	0.89
2	117	Sx	Sx	0.18	0.82
3	117	Sx	Sx	0.10	0.90
6	132	Sx	Sx	0.30	0.70
7	77	Sx	Sx	0.49	0.51
11	117	Sx	Sx	0.30	0.70
12	117	Sx	Sx	0.33	0.67
14	93	Sx	Sx	0.36	0.64
20	117	Sx	Sx	0.27	0.73

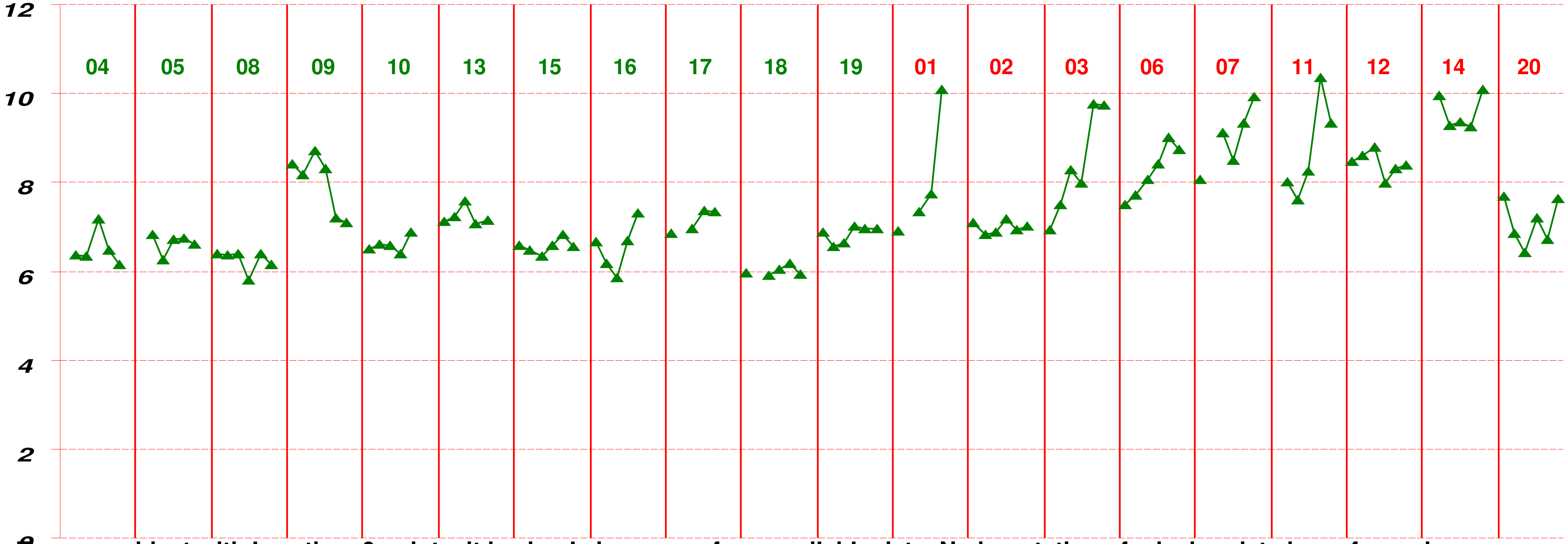
	Asx	**Sx**
0.1T	***11/11***	***7/9***
0.2T	***7/10***	***9/9***
0.8T	***10/11***	***9/9***

Temporal (BL-0T-0.1T-0.2T-0.8T-T) Expression Pattern of 17-Gene Predictor

LRRN3



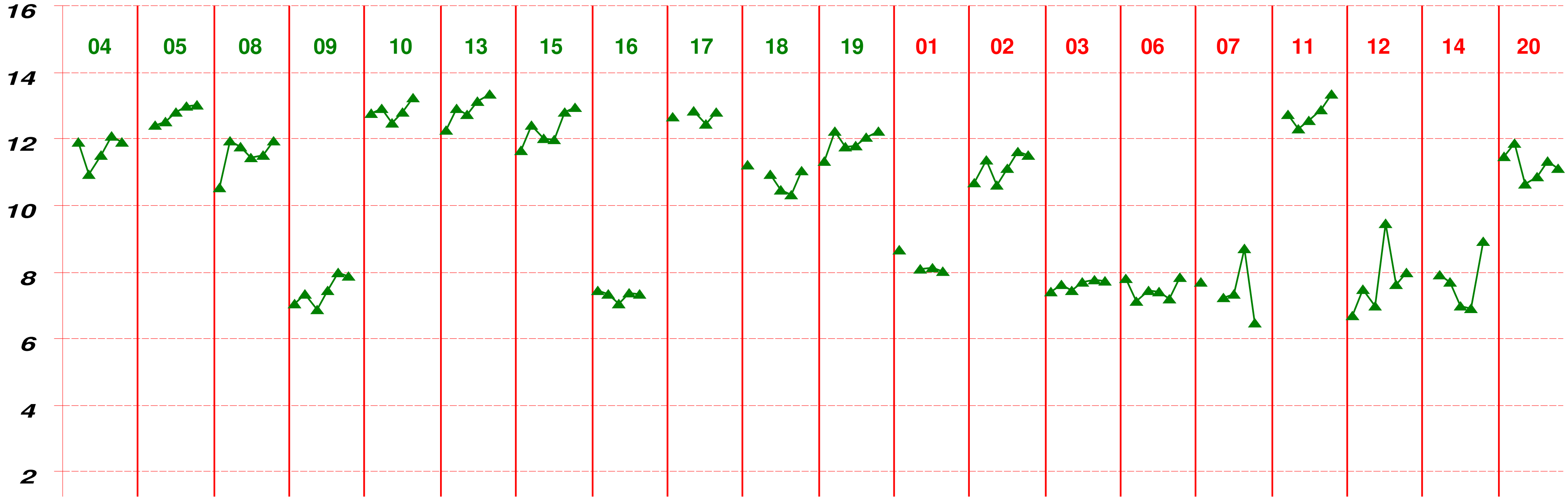
CD177



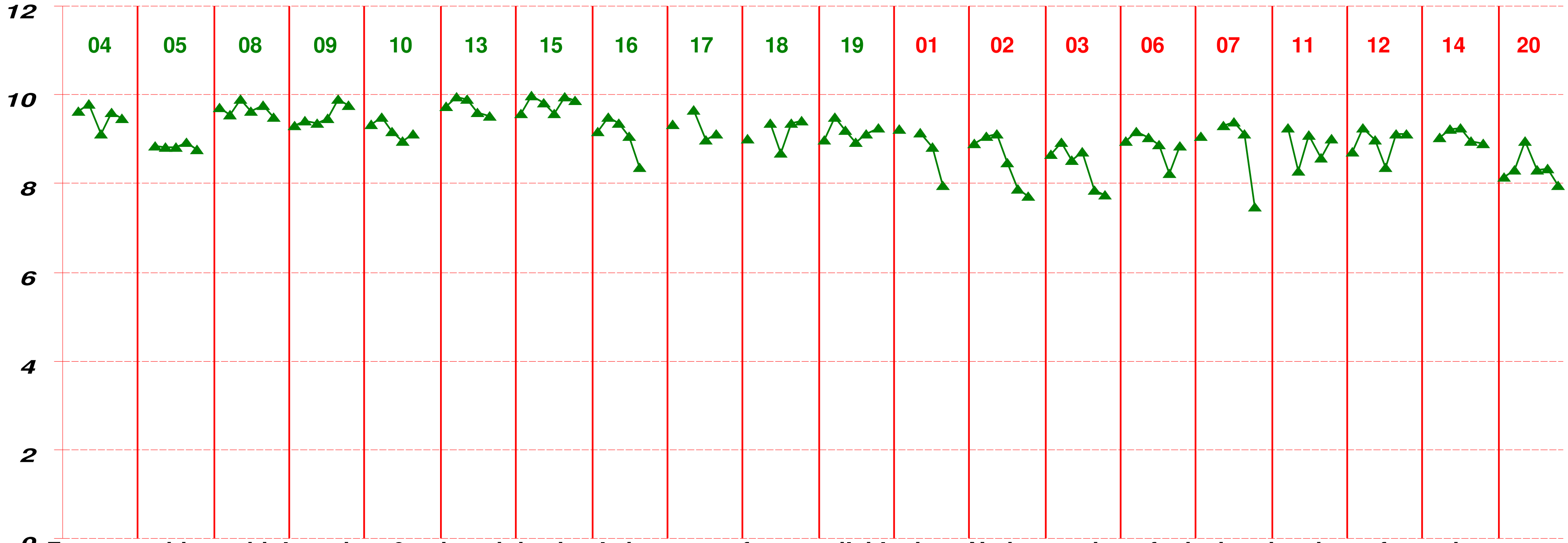
For any subject with less than 6 points, it is simply because of nonavailable data. No imputation of missing data is performed.

Temporal (BL-0T-0.1T-0.2T-0.8T-T) Expression Pattern of 17-Gene Predictor

TUBB2A



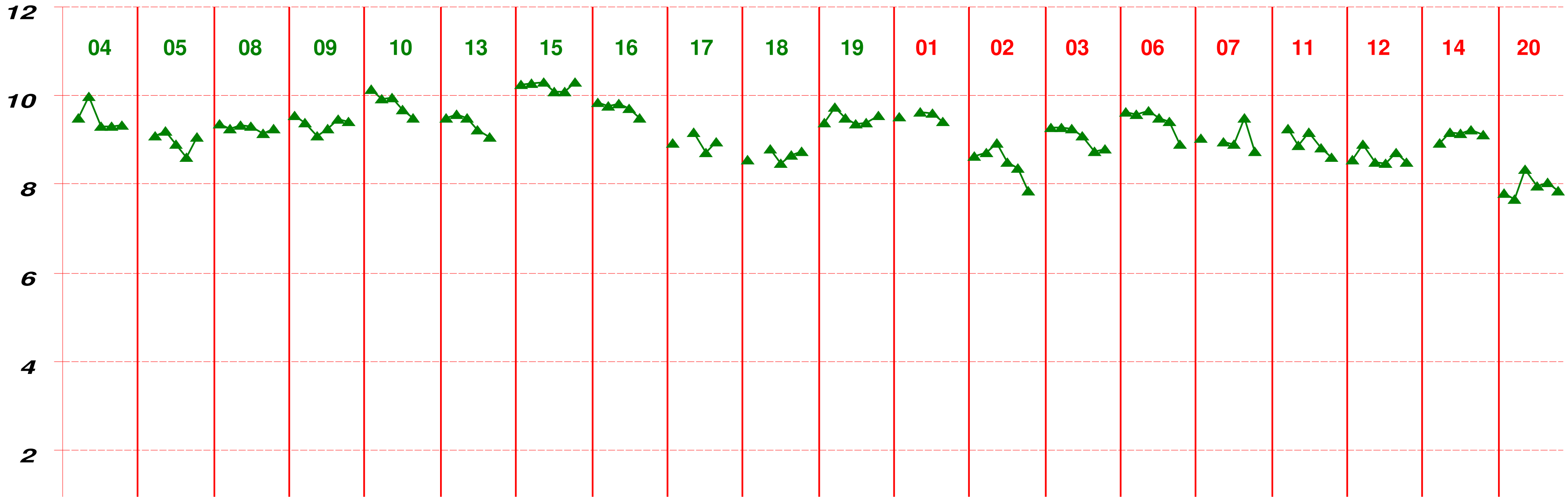
PASK



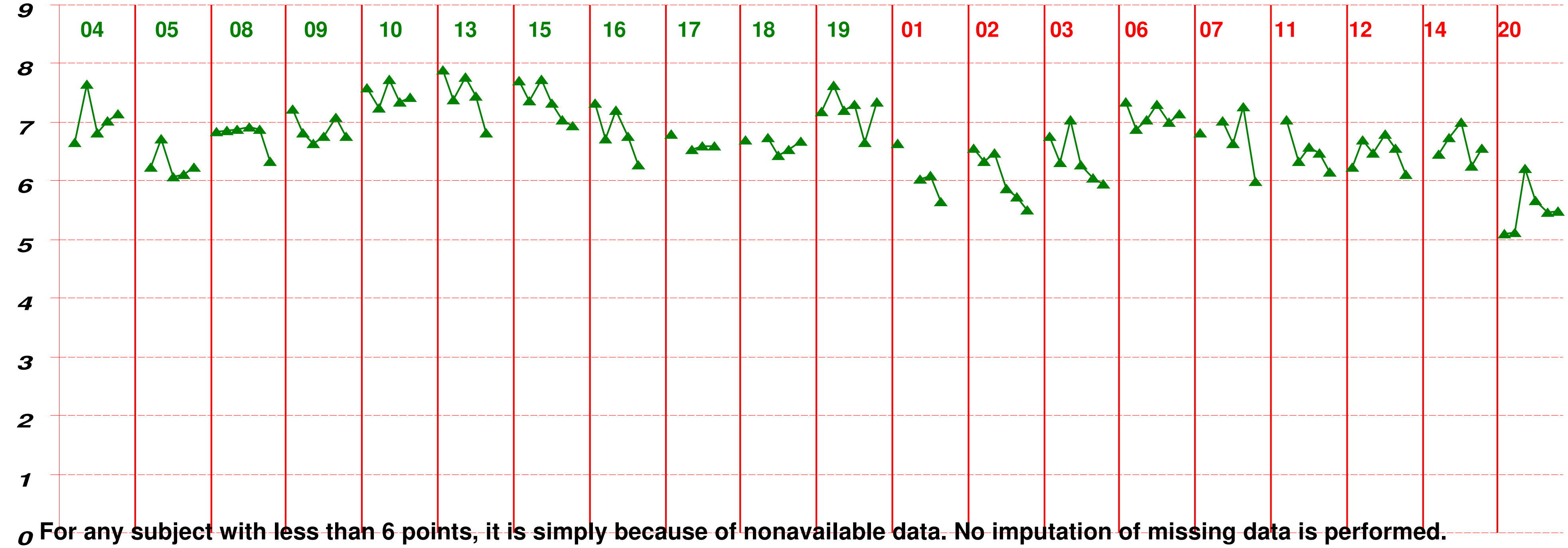
For any subject with less than 6 points, it is simply because of nonavailable data. No imputation of missing data is performed.

Temporal (BL-0T-0.1T-0.2T-0.8T-T) Expression Pattern of 17-Gene Predictor

HLA-DOB

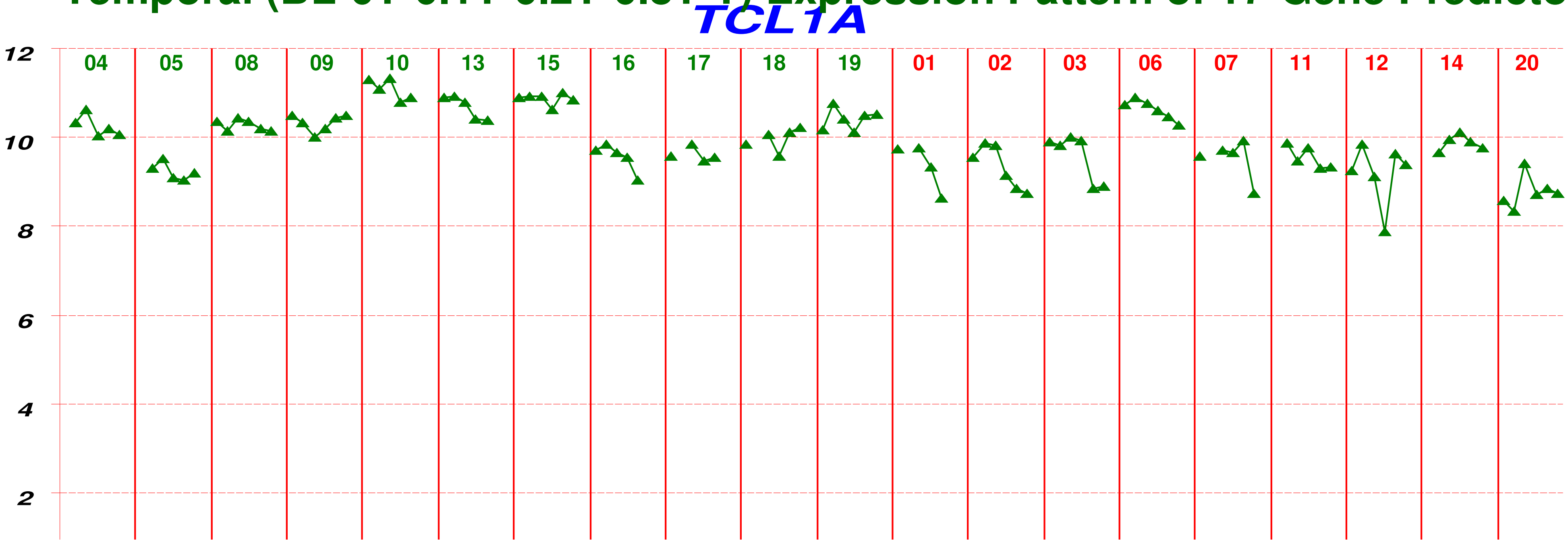


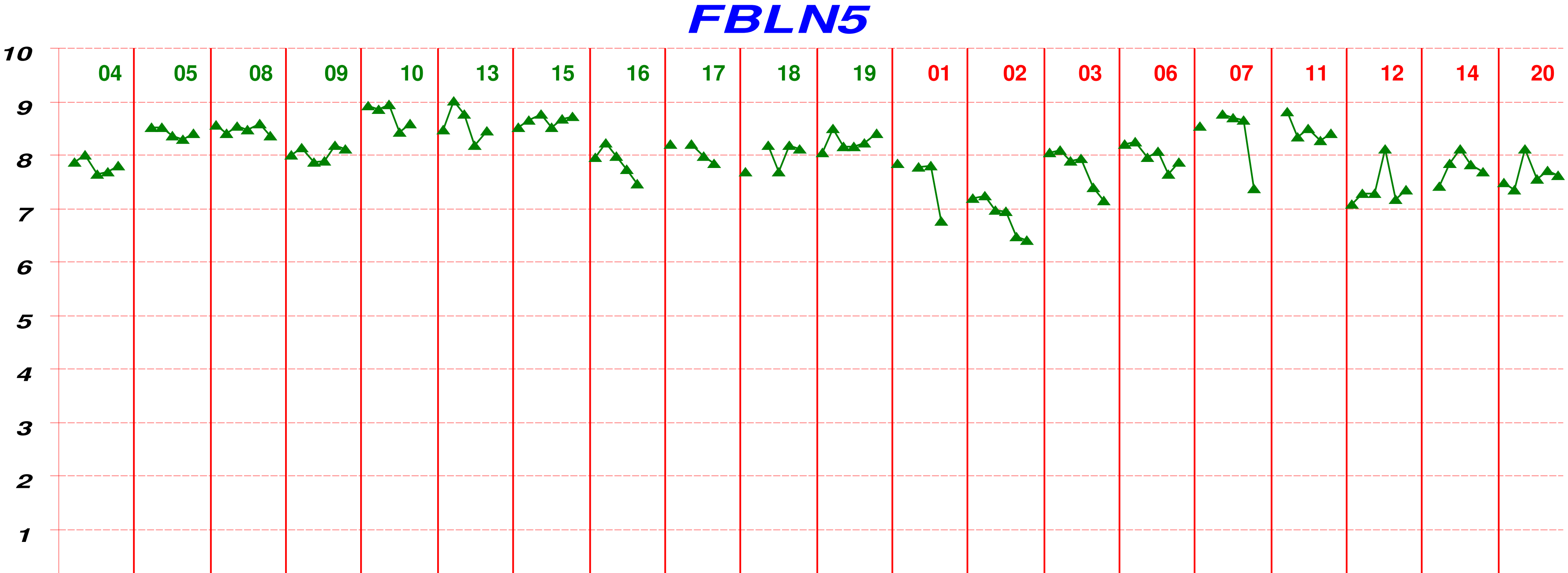
FCRL2



For any subject with less than 6 points, it is simply because of nonavailable data. No imputation of missing data is performed.

Temporal (BL-0T-0.1T-0.2T-0.8T-T) Expression Pattern of 17-Gene Predictor

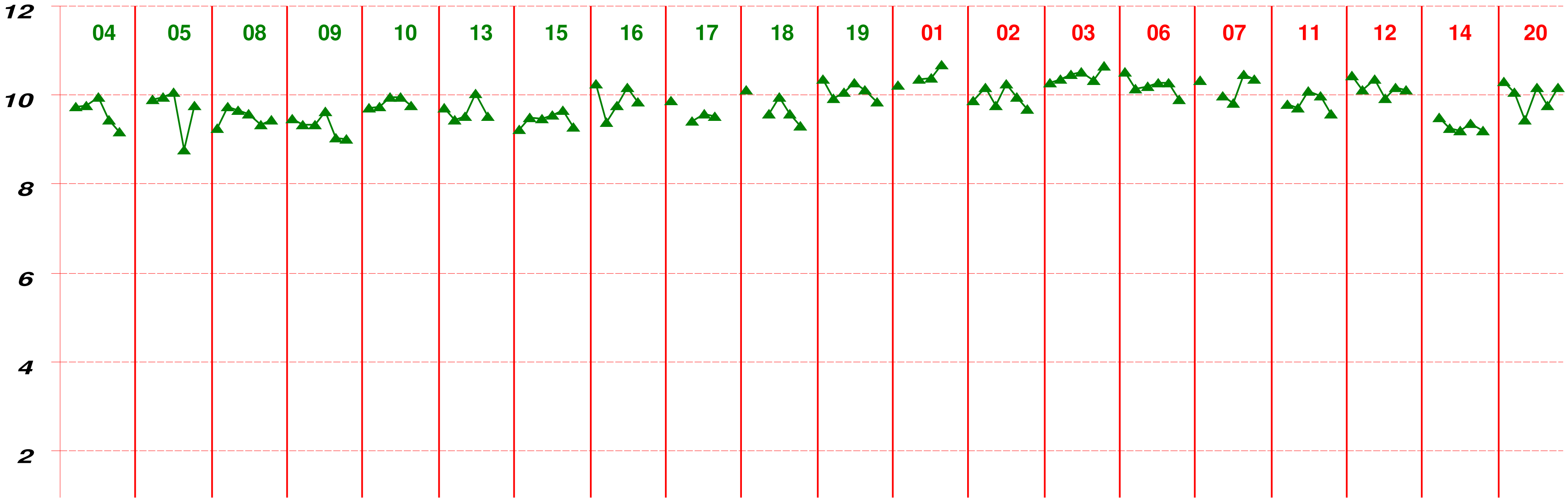




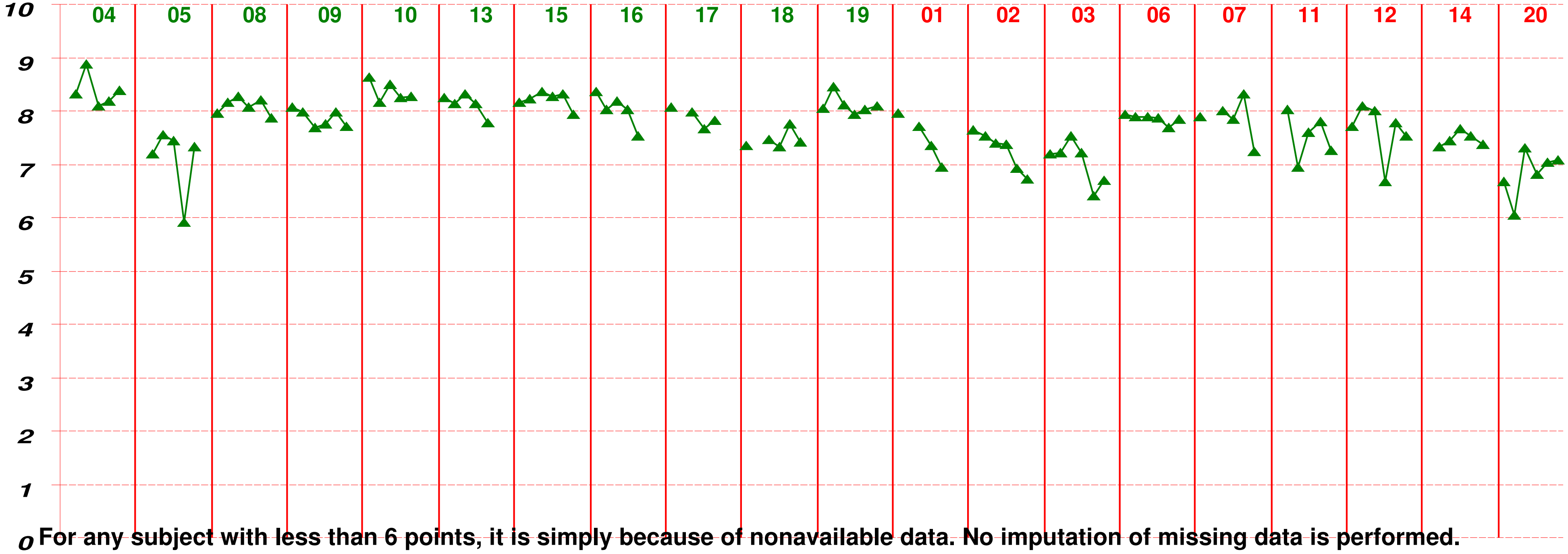
For any subject with less than 6 points, it is simply because of nonavailable data. No imputation of missing data is performed.

Temporal (BL-0T-0.1T-0.2T-0.8T-T) Expression Pattern of 17-Gene Predictor

P2RX1

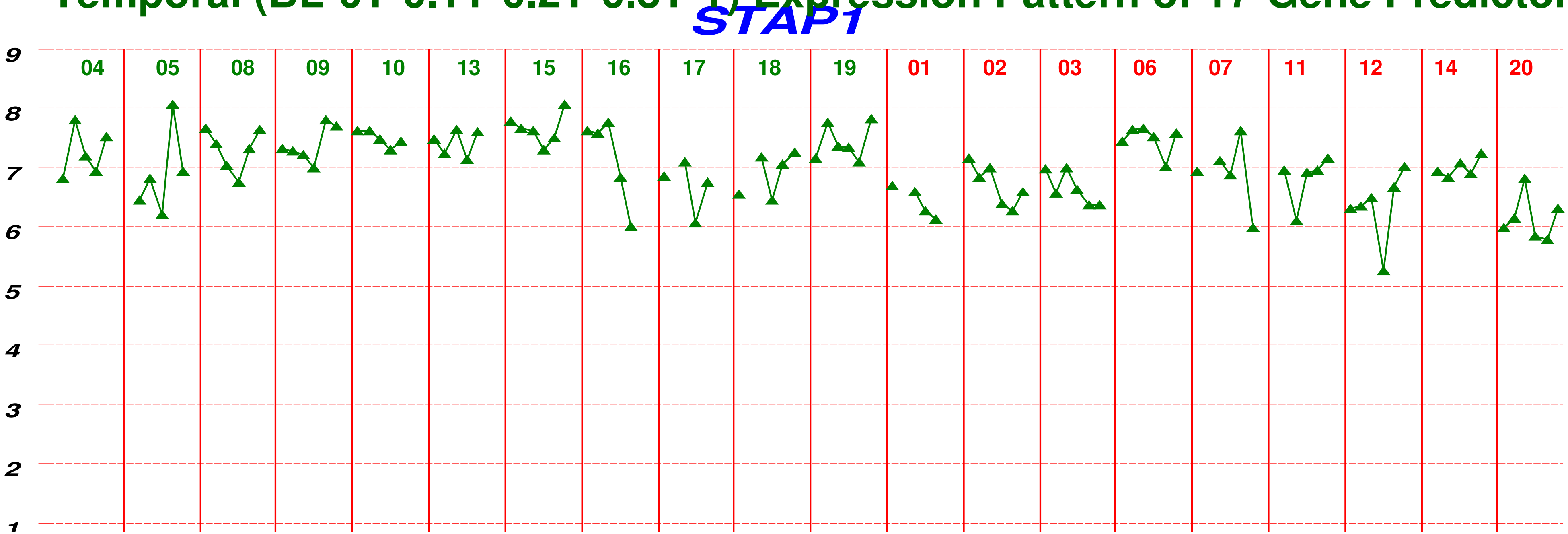


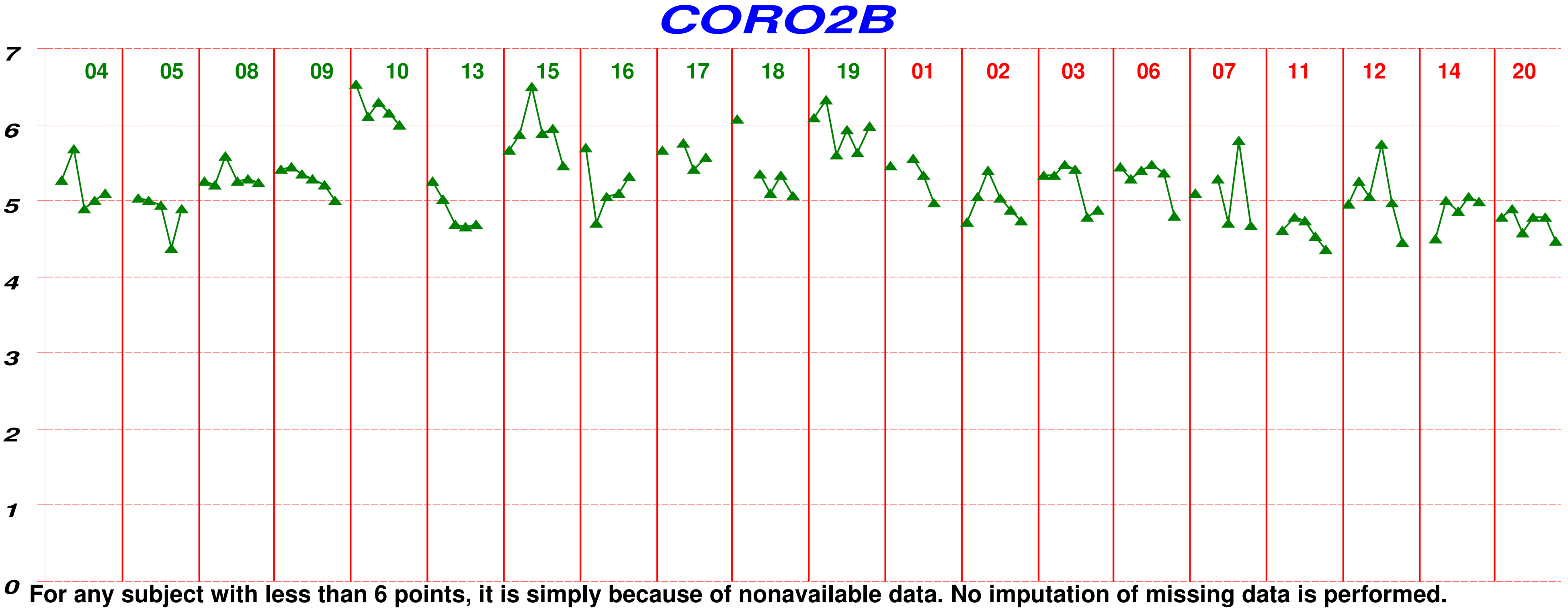
KIAA0125



For any subject with less than 6 points, it is simply because of nonavailable data. No imputation of missing data is performed.

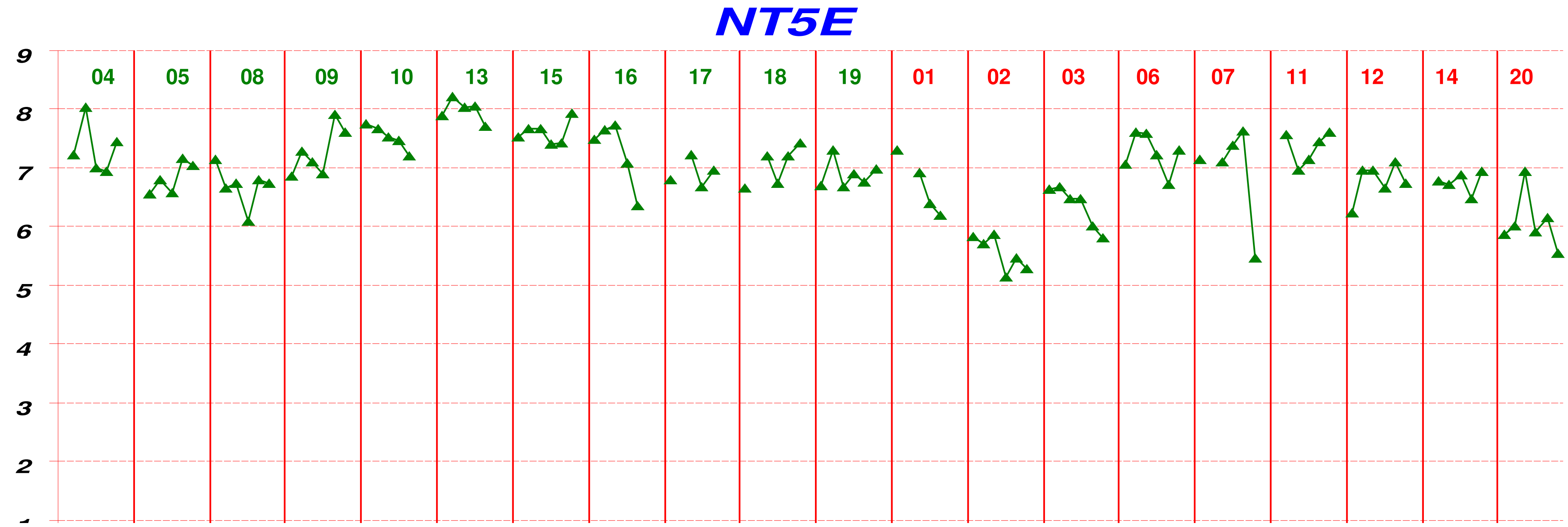
Temporal (BL-0T-0.1T-0.2T-0.8T-T) Expression Pattern of 17-Gene Predictor

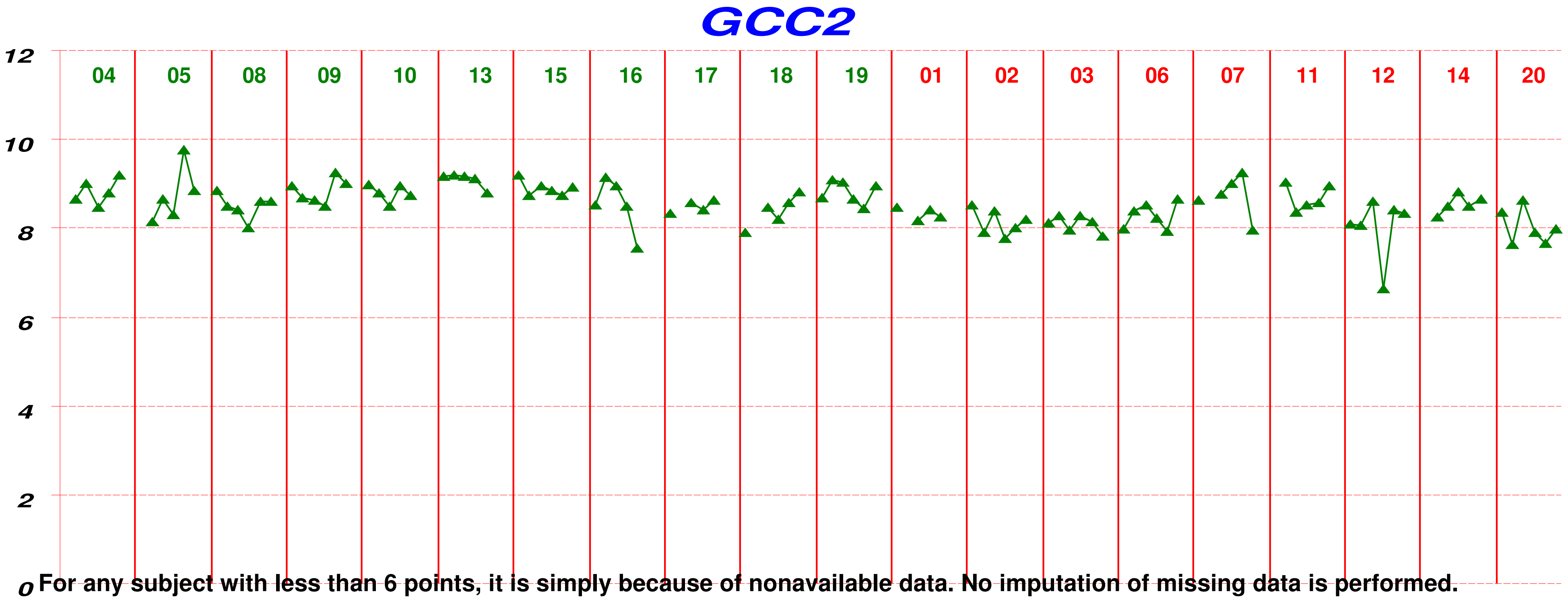




For any subject with less than 6 points, it is simply because of nonavailable data. No imputation of missing data is performed.

Temporal (BL-0T-0.1T-0.2T-0.8T-T) Expression Pattern of 17-Gene Predictor

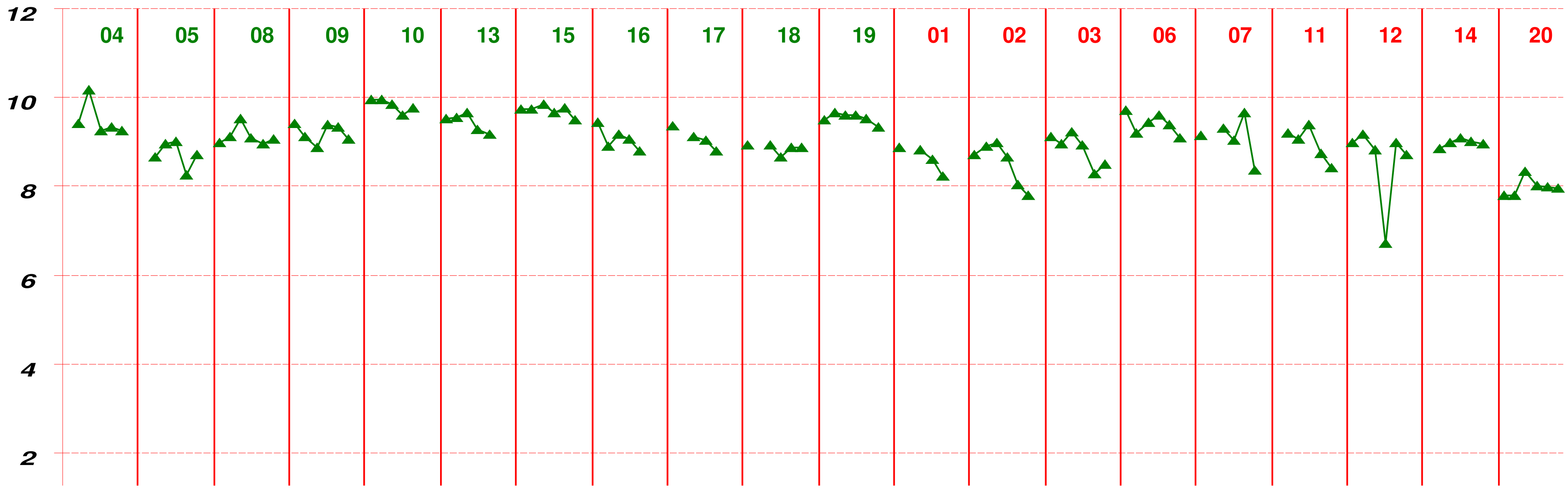




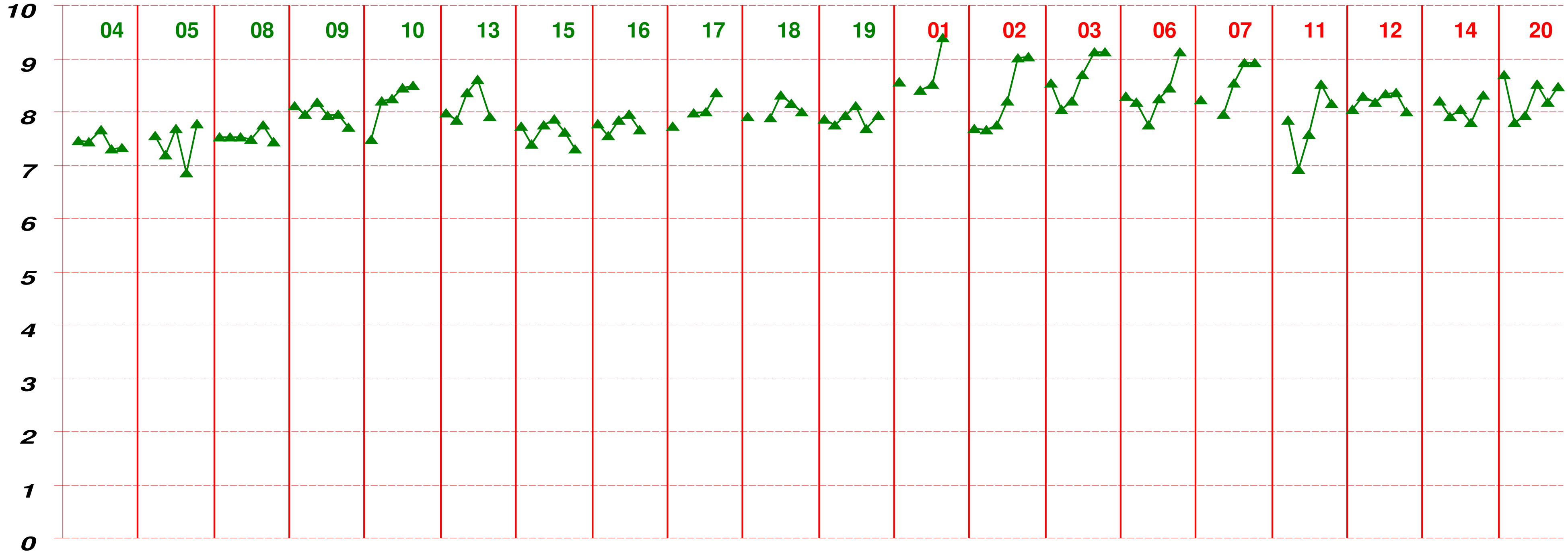
For any subject with less than 6 points, it is simply because of nonavailable data. No imputation of missing data is performed.

Temporal (BL-0T-0.1T-0.2T-0.8T-T) Expression Pattern of 17-Gene Predictor

CD19

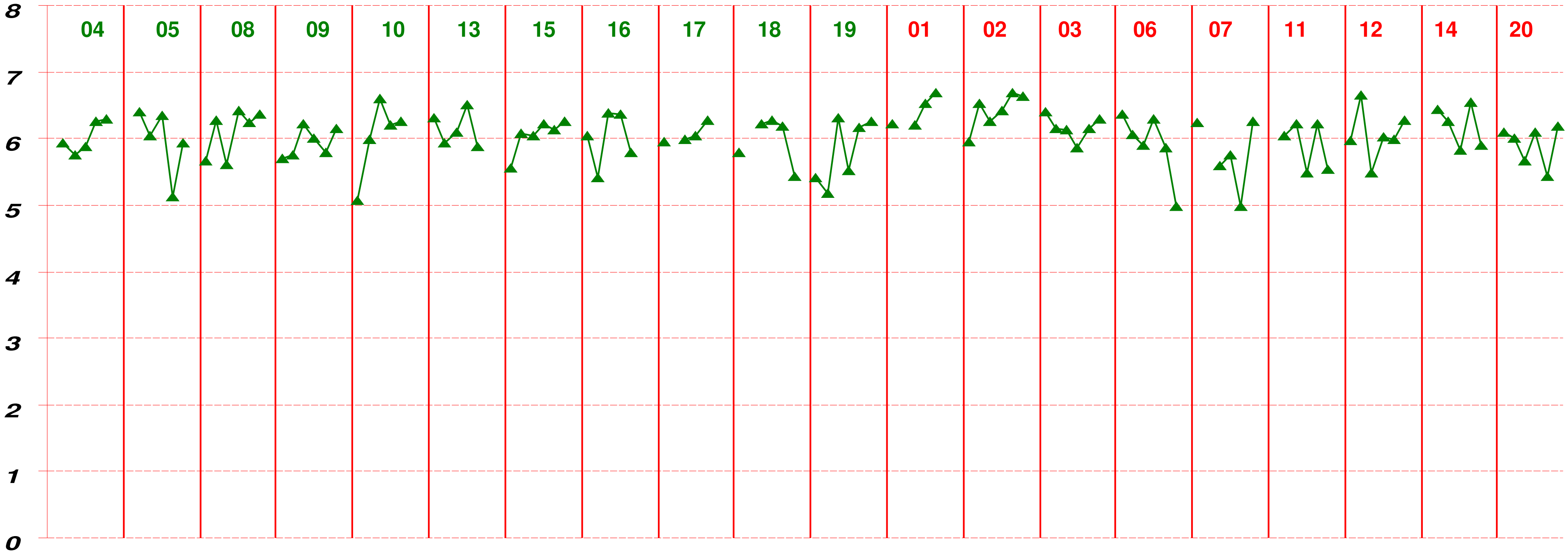


TBC1D8



Temporal (BL-0T-0.1T-0.2T-0.8T-T) Expression Pattern of 17-Gene Predictor

AKT2



For any subject with less than 6 points, it is simply because of nonavailable data. No imputation of missing data is performed.

DMDA variable selection on BL/PC

Correlation network pairs constructed on asx/sx

Retained genes that were hubs of flipping motifs

4-fold cross validation performed over 100 resamples

Obtained list of 112 genes

Used classification criterion on sx/asx in BL/PC to select best 8 (95% accuracy within BL/PC)

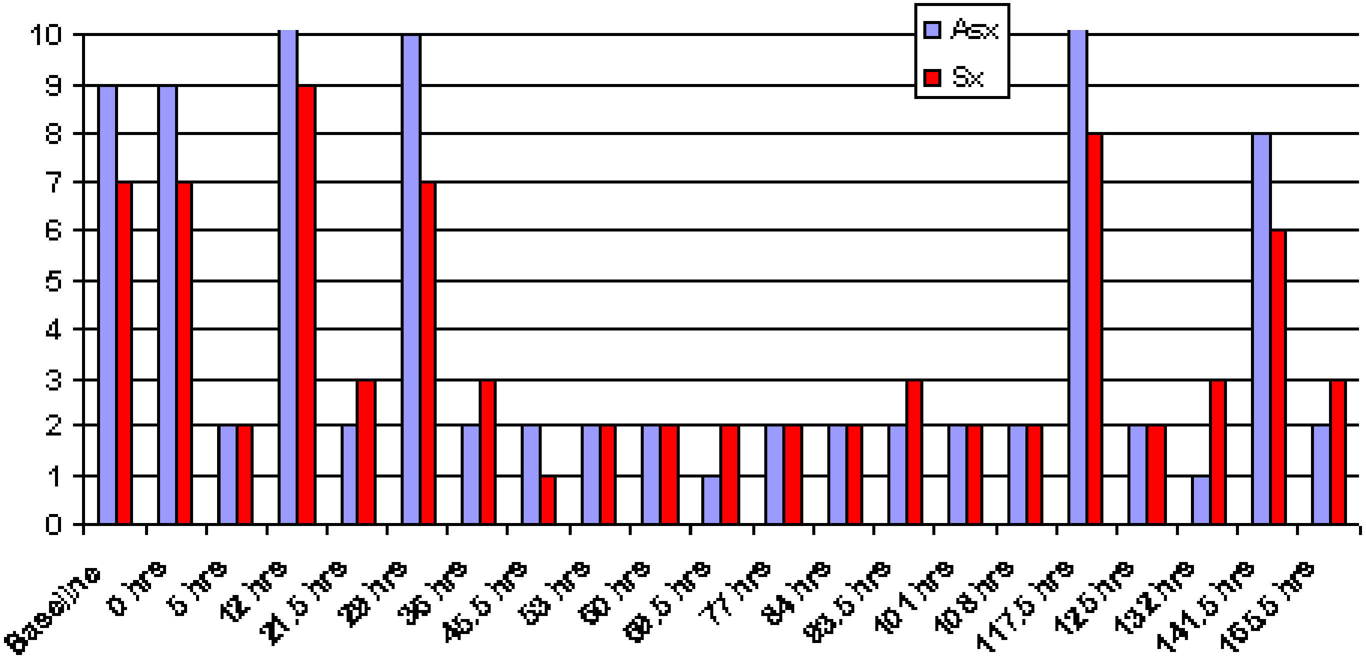
'SMARCC1'
'DDX27'
'HIVEP2'
'VPREB3'
'CD79B'
'CD19'
'ENPP2'
'VPS13D'

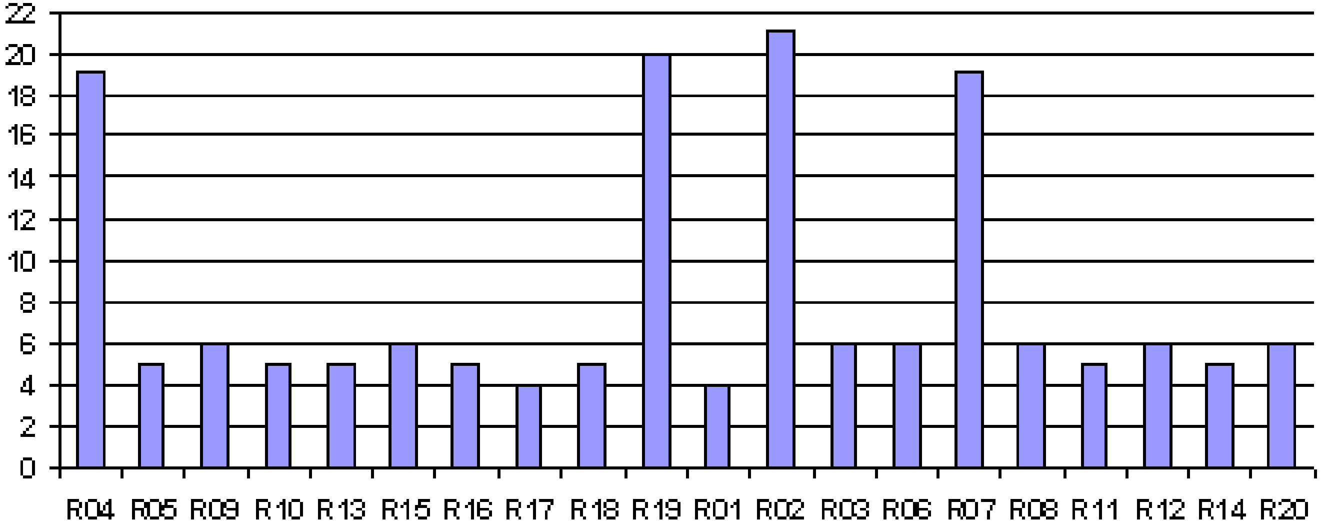
DMDA variable selection on BL/PC

These 8 genes are capable of the following prediction performance (using simple quadratic classifier)

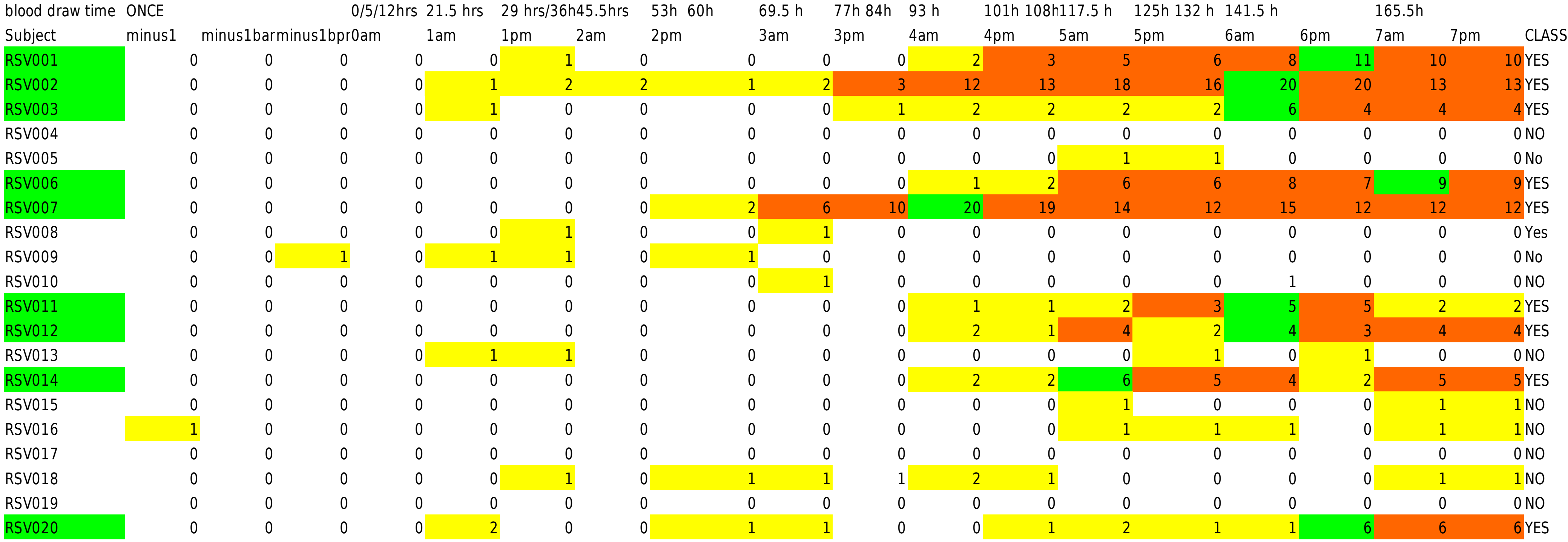
	Asx	**Sx**
0.1T	***11/11***	***5/9 (3-6-7-14S)***
0.2T	***8/10 (5-16A)***	***7/7***
0.8T	***10/11 (9A)***	***9/9***

Sample Summary Statistics





Subject symptom scores



blood draw time	ONCE			0/5/12hrs	21.5 hrs	29 hrs/36h	45.5hrs	53h 60h	69.5 h	77h 84h	93 h	101h 108h	117.5 h	125h 132 h	141.5 h		165.5h		
Subject	minus1	minus1bar	minus1bpr	0am	1am	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