

SUPPLEMENTARY DATA

A mathematical model for short-term vs. long-term survival in patients with
glioma

Jason B. Nikas^{1*}

¹Genomix, Inc., Minneapolis, MN 55364, USA

*Correspondence: Dr. Jason B. Nikas at: jbnikas@genomix-inc.com

Figure S1 and Tables S1-S7

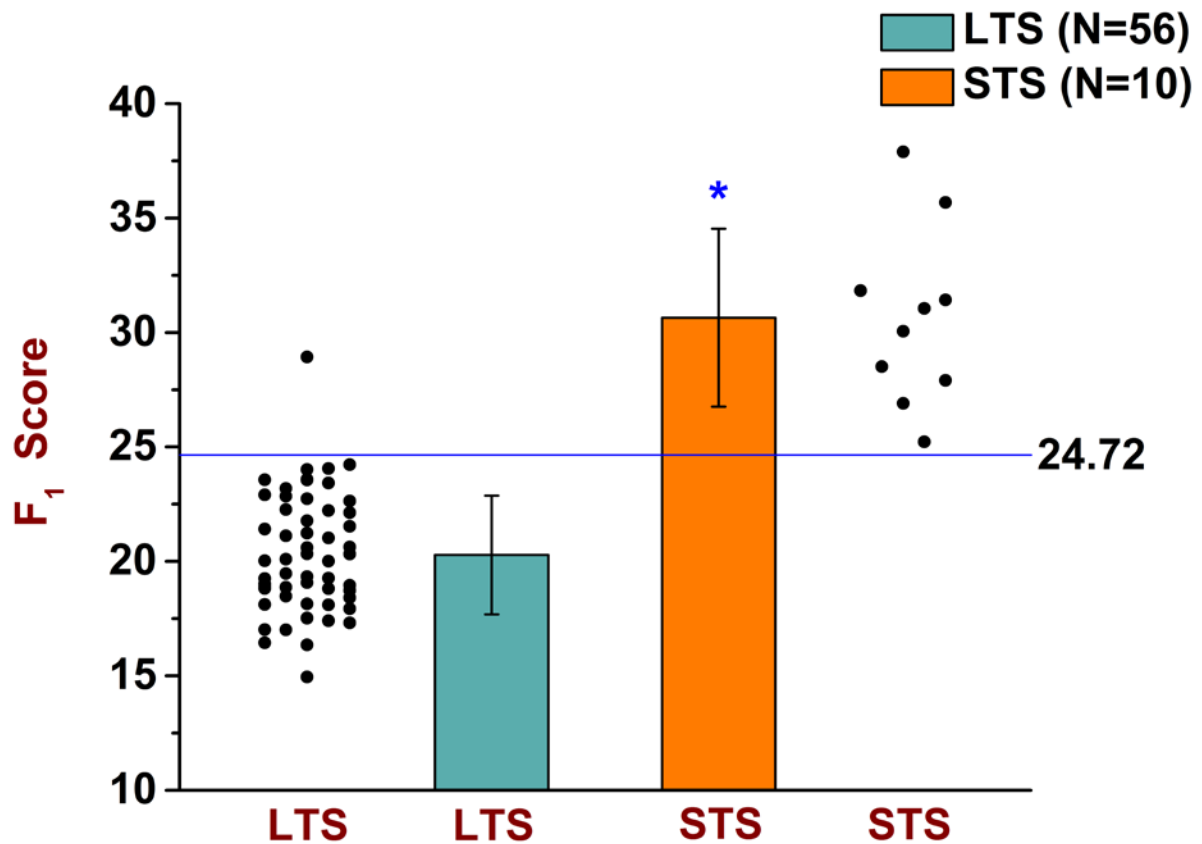


Figure S1. Results of the F_1 function in the development phase.

Sixty six subjects (56 LTS and 10 STS) were randomly selected from the total number of subjects (75 LTS and 14 STS) for the development phase. The F_1 uses 5 of the 29 most significant genes (listed in Table 1) as its input variables. Using the expression value of those 5 genes for a particular subject, the F_1 yields the F_1 score of that subject; and, based on the determined cut-off score for this phase (24.72), the F_1 classifies that subject as a long-term survivor (LTS) if the F_1 score is < 24.72 or as a short-term survivor (STS) if the F_1 score is ≥ 24.72 . As can be seen here in this dot plot & bar graph, all 10 STS subjects (orange color) had F_1 scores greater than the cut-off score of 24.72 and were therefore identified correctly by the F_1 [sensitivity = $(10/10) = 1.000$]. Regarding the LTS subjects (green color), all but one of them had F_1 scores lower than the cut-off score of 24.72 and were therefore identified correctly by the F_1 [specificity = $(55/56) = 0.982$]. The mean F_1 score of the LTS subjects was 20.281 (top of the green bar) and their standard deviation (whiskers above or below the top of the green bar) was 2.596. The mean F_1 score of the STS subjects was 30.649 (top of the orange bar) and their standard deviation (whiskers above or below the top of the orange bar) was 3.887. The significance level was set at $\alpha = 0.001$ (two-tailed), and the probability of significance for the F_1 was $P = 6.00 \times 10^{-16}$ (independent t-Test with T-value = -10.735). The F_1 is parametrically distributed with respect to both groups. The F_1 scores of all 66 subjects used in the development phase are shown in Table S3.

Table S1. Subject clinical & demographical information

Subject ID	Subject No	Gender	Age (yrs)	Tumor Type	Grade	Survival (yrs)
TCGA-CS-4938	1	F	31	Astrocytoma	G2	> 6
TCGA-CS-4942	2	F	44	Astrocytoma	G3	3.658
TCGA-CS-5390	3	F	47	Oligodendroglioma	G2	> 5
TCGA-CS-5394	4	M	40	Astrocytoma	G3	> 5
TCGA-DB-5270	5	F	38	Oligoastrocytoma	G3	> 8
TCGA-DB-5273	6	M	33	Astrocytoma	G3	> 4
TCGA-DB-5274	7	F	37	Oligoastrocytoma	G3	> 4
TCGA-DB-5275	8	M	36	Oligoastrocytoma	G3	> 4
TCGA-DB-5276	9	M	32	Oligoastrocytoma	G3	> 3
TCGA-DB-5277	10	M	34	Astrocytoma	G3	> 3
TCGA-DB-5278	11	M	17	Oligodendroglioma	G2	> 3
TCGA-DB-5279	12	M	59	Oligodendroglioma	G2	> 3
TCGA-DB-5280	13	M	43	Oligoastrocytoma	G2	> 3
TCGA-DB-5281	14	M	61	Oligoastrocytoma	G3	> 4
TCGA-DB-A4XG	15	M	34	Oligodendroglioma	G3	> 6
TCGA-DH-5141	16	M	32	Oligodendroglioma	G3	> 4
TCGA-DU-5870	17	F	34	Oligodendroglioma	G2	> 15
TCGA-DU-6392	18	F	35	Astrocytoma	G3	> 17
TCGA-DU-6393	19	M	66	Oligodendroglioma	G3	4.342
TCGA-DU-6395	20	M	31	Oligoastrocytoma	G2	4.085
TCGA-DU-6396	21	F	31	Oligoastrocytoma	G3	6.263
TCGA-DU-6397	22	M	45	Oligodendroglioma	G3	3.838
TCGA-DU-6399	23	M	54	Oligodendroglioma	G2	5.477
TCGA-DU-6401	24	F	31	Oligodendroglioma	G2	7.288
TCGA-DU-6404	25	F	24	Oligodendroglioma	G3	11.145
TCGA-DU-6407	26	F	35	Oligodendroglioma	G2	7.877
TCGA-DU-6408	27	F	23	Oligodendroglioma	G3	9.504
TCGA-DU-7007	28	M	33	Astrocytoma	G2	5.247
TCGA-DU-7008	29	F	41	Oligodendroglioma	G2	> 12
TCGA-DU-7009	30	F	32	Oligodendroglioma	G2	> 10
TCGA-DU-7011	31	M	25	Oligoastrocytoma	G2	8.767
TCGA-DU-7014	32	M	59	Oligodendroglioma	G2	9.784
TCGA-DU-7015	33	F	41	Oligodendroglioma	G2	> 6
TCGA-DU-7294	34	F	53	Oligodendroglioma	G2	> 7
TCGA-DU-7299	35	M	33	Astrocytoma	G3	3.666
TCGA-DU-7302	36	F	48	Oligodendroglioma	G3	> 4
TCGA-E1-5302	37	M	41	Astrocytoma	G3	4.178

TCGA-E1-5303	38	M	38	Astrocytoma	G3	5.619
TCGA-E1-5304	39	M	42	Astrocytoma	G3	3.427
TCGA-E1-5305	40	M	34	Astrocytoma	G3	6.663
TCGA-E1-5307	41	F	62	Astrocytoma	G3	4.827
TCGA-E1-5311	42	M	31	Oligodendroglioma	G3	11.189
TCGA-E1-5318	43	F	42	Oligodendroglioma	G2	6.515
TCGA-E1-5319	44	F	48	Oligodendroglioma	G2	7.964
TCGA-E1-5322	45	F	38	Oligoastrocytoma	G2	10.899
TCGA-HT-7480	46	M	33	Oligodendroglioma	G2	> 6
TCGA-HT-7481	47	M	39	Oligodendroglioma	G2	> 5
TCGA-HT-7482	48	F	18	Oligoastrocytoma	G2	> 6
TCGA-HT-7483	49	M	14	Oligoastrocytoma	G2	> 13
TCGA-HT-7485	50	M	42	Astrocytoma	G2	> 5
TCGA-HT-7601	51	F	30	Astrocytoma	G3	> 3
TCGA-HT-7602	52	M	21	Oligodendroglioma	G2	> 3
TCGA-HT-7603	53	M	29	Oligodendroglioma	G2	> 3
TCGA-HT-7604	54	M	50	Astrocytoma	G2	> 8
TCGA-HT-7605	55	M	38	Oligodendroglioma	G2	> 3
TCGA-HT-7606	56	F	30	Astrocytoma	G2	> 3
TCGA-HT-7608	57	M	61	Oligoastrocytoma	G2	> 4
TCGA-HT-7609	58	M	34	Oligoastrocytoma	G3	> 4
TCGA-HT-7610	59	F	25	Oligoastrocytoma	G2	> 4
TCGA-HT-7620	60	M	40	Oligodendroglioma	G3	> 4
TCGA-HT-7676	61	M	26	Oligodendroglioma	G2	> 3
TCGA-HT-7681	62	F	29	Oligoastrocytoma	G2	> 3
TCGA-HT-7684	63	M	58	Oligoastrocytoma	G3	> 3
TCGA-HT-7686	64	F	29	Astrocytoma	G3	> 3
TCGA-HT-7691	65	F	31	Astrocytoma	G2	> 3
TCGA-HT-7693	66	F	51	Oligodendroglioma	G2	> 3
TCGA-HT-7694	67	M	60	Oligodendroglioma	G3	> 3
TCGA-HT-7855	68	M	39	Astrocytoma	G3	> 3
TCGA-HT-7856	69	M	35	Oligodendroglioma	G3	> 3
TCGA-HT-7857	70	F	24	Astrocytoma	G3	> 5
TCGA-HT-7858	71	M	28	Astrocytoma	G2	> 4
TCGA-HT-7874	72	F	41	Oligodendroglioma	G3	> 3
TCGA-HT-7875	73	M	56	Oligodendroglioma	G2	> 3
TCGA-HT-8013	74	F	37	Oligoastrocytoma	G2	5.296
TCGA-HT-8113	75	F	49	Oligodendroglioma	G2	> 5
TCGA-CS-4941	76	M	67	Astrocytoma	G3	0.641
TCGA-CS-5397	77	F	54	Astrocytoma	G3	0.532
TCGA-DU-5852	78	F	61	Oligoastrocytoma	G3	0.562
TCGA-DU-6400	79	F	66	Oligodendroglioma	G2	0.101
TCGA-DU-6402	80	M	52	Astrocytoma	G3	0.584

TCGA-DU-6403	81	F	60	Oligoastrocytoma	G3	0.970
TCGA-DU-7006	82	F	60	Astrocytoma	G3	0.956
TCGA-DU-7012	83	F	74	Astrocytoma	G3	0.542
TCGA-DU-7013	84	M	59	Astrocytoma	G3	0.737
TCGA-DU-7290	85	F	45	Astrocytoma	G3	0.863
TCGA-DU-7292	86	M	69	Astrocytoma	G3	0.663
TCGA-DU-8158	87	F	57	Astrocytoma	G3	0.425
TCGA-HT-7607	88	F	61	Astrocytoma	G2	0.263
TCGA-HT-7616	89	M	75	Oligodendroglioma	G3	0.019

Clinical and demographical information about all 89 subjects used in this study. “G2” denotes grade II glioma, whereas “G3” denotes grade III glioma. Both age and survival are expressed in years.

Table S2. Control genes

Gene ID	DE (STS)	ROC AUC	FC	P	M _L	SD _L	M _S	SD _S	Notes
TBP	↓	-0.6400	-1.134	0.0991	305.345	55.917	269.260	69.251	MW
GAPDH	↓	-0.5124	-1.014	0.8892	55077.998	27507.540	54340.327	22340.622	MW
HPRT1	↑	0.6495	1.196	0.0777	337.655	341.162	403.868	278.501	MW
ALAS1	↑	0.7924	1.233	0.0005	934.375	210.949	1151.754	181.340	TT
TUBA1B	↑	0.6600	1.247	0.0585	20329.351	8211.046	25357.806	9497.810	MW
ALB	↓	-0.6333	-1.217	0.1167	10.406	7.375	8.552	9.014	MW
B2M	↑	0.7171	1.486	0.0093	20661.472	12414.934	30713.233	15691.137	MW
G6PD	↑	0.7810	1.347	0.0006	688.970	163.633	927.703	327.069	MW
POLR2A	↑	0.5029	1.019	0.9778	5400.021	1399.469	5502.813	1525.207	MW
PPIA	↑	0.5276	1.025	0.7507	3407.330	804.451	3492.089	1022.502	MW
SDHA	↓	-0.6714	-1.163	0.0457	2626.016	637.189	2258.161	537.103	TT
UBC	↑	0.6257	1.213	0.1396	32311.188	10082.128	39181.196	18045.906	MW
YWHAZ	↑	0.6010	1.112	0.2373	10450.292	3098.215	11622.114	3590.663	MW
LRP1	↑	0.6524	1.327	0.0720	20336.358	7489.631	26976.757	12183.433	MW
PGK1	↑	0.5733	1.022	0.3927	5651.227	1931.224	5776.726	1357.817	MW
HMBS	↓	-0.6314	-1.133	0.1221	413.441	140.314	365.045	110.946	MW
DIMT1L	↓	-0.7352	-1.277	0.0046	563.914	184.869	441.500	81.754	MW
LMNB1	↑	0.6886	1.707	0.0249	477.157	383.555	814.470	696.903	MW

All of those 18 control genes were analyzed for statistically significant differential expression between the two groups (14 STS and 75 LTS subjects). The arrows indicate differential expression [over-expression (↑) or under-expression (↓)] of the STS subjects as compared with the LTS subjects. The ROC AUC value, the fold change (FC) value, the P-value, the mean expression value of the LTS subjects (M_L) and their standard deviation (SD_L), the mean expression value of the STS subjects (M_S) and their standard deviation (SD_S) are listed for each gene variable.

(AW): The Aspin-Welch unequal-variance test was used for the calculation of the P-value for those variables.

(MW): The Mann-Whitney U test was used for the calculation of the P-value for those variables.

(TT): The independent t-Test for parametric variables was used for the calculation of the P-value for those variables.

As can be seen, none of the 18 control genes met the overall criterion of significance set for, and required by, this study: 1) ROC AUC \geq 0.930 (or ROC AUC \leq -0.930), 2) FC \geq 1.50 (or FC \leq -1.50), and 3) P < 2.43 x 10⁻⁶. In fact, as can be seen, none of the 18 control genes met the individual significance criterion of any of the three different and independent methods employed.

Table S3. Results of the F_1 function in the development phase

Subject ID	Subject No	F_1 Score	Notes
TCGA-CS-4938	1	20.6031	
TCGA-CS-5390	3	18.7324	
TCGA-CS-5394	4	19.0248	
TCGA-DB-5270	5	23.5807	
TCGA-DB-5273	6	20.0999	
TCGA-DB-5274	7	22.8516	
TCGA-DB-5276	9	22.2236	
TCGA-DB-5278	11	20.3211	
TCGA-DB-5280	13	16.4466	
TCGA-DB-5281	14	21.0244	
TCGA-DH-5141	16	23.1896	
TCGA-DU-6392	18	22.9112	
TCGA-DU-6393	19	24.0155	
TCGA-DU-6395	20	18.9559	
TCGA-DU-6396	21	16.3499	
TCGA-DU-6397	22	21.5289	
TCGA-DU-6399	23	19.3396	
TCGA-DU-6401	24	14.9465	
TCGA-DU-6407	26	17.0150	
TCGA-DU-6408	27	18.4863	
TCGA-DU-7007	28	23.4241	
TCGA-DU-7008	29	19.2650	
TCGA-DU-7011	31	20.0272	
TCGA-DU-7014	32	17.4065	
TCGA-DU-7015	33	18.8721	
TCGA-DU-7299	35	18.8136	
TCGA-DU-7302	36	20.3283	
TCGA-E1-5302	37	20.6246	
TCGA-E1-5303	38	21.7771	
TCGA-E1-5305	40	18.4177	
TCGA-E1-5307	41	18.1212	
TCGA-E1-5311	42	22.7442	
TCGA-E1-5319	44	21.1220	
TCGA-E1-5322	45	19.2545	
TCGA-HT-7480	46	18.1003	
TCGA-HT-7482	48	20.0125	
TCGA-HT-7483	49	18.8941	
TCGA-HT-7485	50	21.2371	

TCGA-HT-7601	51	19.4745	
TCGA-HT-7602	52	21.4121	
TCGA-HT-7604	54	17.5262	
TCGA-HT-7605	55	22.1334	
TCGA-HT-7606	56	19.0777	
TCGA-HT-7609	58	17.3158	
TCGA-HT-7610	59	18.8204	
TCGA-HT-7676	61	17.5183	
TCGA-HT-7681	62	22.2633	
TCGA-HT-7684	63	22.6417	
TCGA-HT-7691	65	24.2259	
TCGA-HT-7693	66	17.9358	
TCGA-HT-7694	67	23.5650	
TCGA-HT-7856	69	24.0601	
TCGA-HT-7857	70	28.9318	x
TCGA-HT-7858	71	17.0243	
TCGA-HT-7874	72	23.5586	
TCGA-HT-8013	74	18.1473	
TCGA-DU-5852	78	26.9029	
TCGA-DU-6402	80	37.8965	
TCGA-DU-6403	81	28.5118	
TCGA-DU-7012	83	31.4257	
TCGA-DU-7013	84	31.8338	
TCGA-DU-7290	85	35.6830	
TCGA-DU-7292	86	30.0601	
TCGA-DU-8158	87	31.0498	
TCGA-HT-7607	88	25.2198	
TCGA-HT-7616	89	27.9091	

The F_1 scores of the 66 subjects [56 LTS (# 1-75) and 10 STS (#76-89)] that were used for the development of the F_1 function are shown. As can be seen, and based on the determined cut-off score for this phase (24.72), all of the 56 LTS subjects were classified correctly except one (# 70), and all of the 10 STS subjects were classified correctly. “x” denotes a misclassification.

Table S4. Results of the F_1 function in the validation phase

Subject ID	Subject No	F_1 Score	Notes
TCGA-CS-4942	2	20.4138	
TCGA-DB-5275	8	19.5583	
TCGA-DB-5277	10	22.0500	
TCGA-DB-5279	12	24.4905	
TCGA-DB-A4XG	15	18.2326	
TCGA-DU-5870	17	20.0020	
TCGA-DU-6404	25	29.0098	x
TCGA-DU-7009	30	18.4687	
TCGA-DU-7294	34	18.7784	
TCGA-E1-5304	39	20.6672	
TCGA-E1-5318	43	21.1144	
TCGA-HT-7481	47	20.3288	
TCGA-HT-7603	53	26.1538	x
TCGA-HT-7608	57	23.0015	
TCGA-HT-7620	60	20.1659	
TCGA-HT-7686	64	21.4548	
TCGA-HT-7855	68	14.5007	
TCGA-HT-7875	73	19.1403	
TCGA-HT-8113	75	25.1102	x*
TCGA-CS-4941	76	30.8289	
TCGA-CS-5397	77	28.7981	
TCGA-DU-6400	79	22.3210	x
TCGA-DU-7006	82	33.7616	

The F_1 scores of the 23 unknown subjects [19 LTS (# 1-75) and 4 STS (#76-89)] that were used for the validation of the F_1 function are shown. As can be seen, and based on the original cut-off score that was determined in the development phase (24.72), all of the 19 LTS subjects were classified correctly except three (# 25, 53, & 75), and all of the 4 STS subjects were classified correctly except one (# 79). Using the more optimal cut-off score of 25.165, the unknown LTS subject # 75 (denoted by: x*) is no longer misclassified by the F_1 . “x” denotes a misclassification.

Table S5. Results of a 10-fold cross validation and a leave-one-out cross validation in connection with the performance of the F_1 function

a) 10-fold cross validation:

K-fold cross validation partition

N: 89

Number of Test Sets: 10

Training Size: 81 80 80 80 80 80 80 80 80 80

Test Size: 8 9 9 9 9 9 9 9 9 9

Misclassification Rate = 0.05618

Mean-squared Error = 0.05618

Confusion Matrix:

6	0	1	1
8	0	0	1
7	0	1	1
8	0	0	1
8	1	0	0
8	0	0	1
7	0	0	2
7	0	0	2
6	1	1	1
7	0	0	2

[72 2 3 12]

The Confusion Matrix to the left shows the cross validation results in each of the 10 folds (presented in the 10 rows of the matrix). The first column of the matrix shows the number of test subjects that were randomly selected from the first group (LTS) for validation purposes and were subsequently classified correctly. The second column shows the number of subjects that were randomly selected from the second group (STS) for validation purposes and were subsequently misclassified into the first group (LTS). The third column shows the number of subjects that were randomly selected from the first group (LTS) for validation purposes and were subsequently misclassified into the second group (STS). The fourth column shows the number of test subjects that were randomly selected from the second group (STS) for validation purposes and were subsequently classified correctly.

The row vector at the bottom is the summation matrix of the Confusion Matrix, i.e. its first element is the sum of all the values of the first column of the Confusion Matrix, its second element is the sum of all the values of the second column of the Confusion Matrix, etc. As can be seen from the summation matrix, following all ten folds of cross validation, of the 75 subjects of the first group (LTS), 72 were classified correctly and 3 were misclassified into the second group (STS); whereas of the 14 subjects of the second group (STS), 12 were classified correctly and 2 were misclassified into the first group (LTS).

In total, using a 10-fold cross validation, there were 5 misclassifications out of 89 subjects, and that yields a misclassification rate of 0.05618.

b) Leave-one-out cross validation:

Misclassification Rate = 0.05618

Mean-squared Error = 0.05618

Confusion Matrix:

1	0	0	0
1	0	0	0
0	0	0	1
1	0	0	0
1	0	0	0
0	0	1	0
0	0	0	1
1	0	0	0
1	0	0	0
1	0	0	0
1	0	0	0
0	0	0	1
1	0	0	0
0	1	0	0
1	0	0	0
1	0	0	0
1	0	0	0
1	0	0	0
1	0	0	0
1	0	0	0
1	0	0	0
1	0	0	0
1	0	0	0
1	0	0	0
1	0	0	0
1	0	0	0
1	0	0	0
1	0	0	0
1	0	0	0
1	0	0	0
1	0	0	0
0	0	0	1
0	1	0	0
1	0	0	0
1	0	0	0
1	0	0	0
1	0	0	0
0	0	0	1
0	0	0	1
1	0	0	0
0	0	1	0
1	0	0	0

1	0	0	0
1	0	0	0
1	0	0	0
0	0	0	1
1	0	0	0

[72 2 3 12]

The Confusion Matrix above shows the results of the leave-one-out cross validation. Since there are totally 89 subjects, there are 89 rounds of cross validation, and during each one of those, one subject is randomly left out and used for validation purposes. The first column shows the number of test subjects that were randomly left out from the first group (LTS) for validation purposes and were subsequently classified correctly. The second column shows the number of subjects that were randomly left out from the second group (STS) for validation purposes and were subsequently misclassified into the first group (LTS). The third column shows the number of subjects that were randomly left out from the first group (LTS) for validation purposes and were subsequently misclassified into the second group (STS). The fourth column shows the number of test subjects that were randomly left out from the second group (STS) for validation purposes and were subsequently classified correctly.

As can be seen from the summation matrix, at the bottom of the Confusion Matrix, following all 89 rounds of leave-one-out cross validation, of the 75 subjects of the first group (LTS), 72 were classified correctly and 3 were misclassified into the second group (STS); whereas of the 14 subjects of the second group (STS), 12 were classified correctly and 2 were misclassified into the first group (LTS).

In total, using a leave-one-out cross validation, there were 5 misclassifications out of 89 subjects, and that yields a misclassification rate of 0.05618.

Table S6. Supervised PCA results in connection with the 29 most significant genes

Subject ID	Subject No	PC1	PC2	Notes
TCGA-CS-4938	1	-2.0216	-0.2569	
TCGA-CS-4942	2	-1.5569	-0.3859	
TCGA-CS-5390	3	-1.3292	-0.3034	
TCGA-CS-5394	4	-3.2032	0.7796	
TCGA-DB-5270	5	0.6549	-1.5510	x
TCGA-DB-5273	6	-2.4310	0.3817	
TCGA-DB-5274	7	-0.0967	-1.6334	
TCGA-DB-5275	8	-2.0112	0.2033	
TCGA-DB-5276	9	-0.5895	-0.9637	
TCGA-DB-5277	10	0.3630	-1.3536	x
TCGA-DB-5278	11	-1.0077	-0.1539	
TCGA-DB-5279	12	-0.1544	-1.4953	
TCGA-DB-5280	13	-3.3818	1.2775	
TCGA-DB-5281	14	-0.9242	0.3883	
TCGA-DB-A4XG	15	-2.1429	0.7690	
TCGA-DH-5141	16	-2.3133	-0.3032	
TCGA-DU-5870	17	-1.8950	0.1495	
TCGA-DU-6392	18	3.6048	0.0206	x
TCGA-DU-6393	19	-1.1865	-1.2682	
TCGA-DU-6395	20	-2.6774	0.0132	
TCGA-DU-6396	21	-1.2391	0.8283	
TCGA-DU-6397	22	-1.2834	-0.6801	
TCGA-DU-6399	23	-2.2162	0.1046	
TCGA-DU-6401	24	-3.6034	1.0754	
TCGA-DU-6404	25	3.0144	-0.8123	x
TCGA-DU-6407	26	-2.9477	0.2656	
TCGA-DU-6408	27	-3.3722	1.1508	
TCGA-DU-7007	28	1.1953	-1.6702	x
TCGA-DU-7008	29	-1.6269	0.0520	
TCGA-DU-7009	30	-2.5524	0.8891	
TCGA-DU-7011	31	-1.2557	0.8288	
TCGA-DU-7014	32	-3.1395	1.8808	
TCGA-DU-7015	33	-2.5806	0.7692	
TCGA-DU-7294	34	-2.7889	0.8235	
TCGA-DU-7299	35	-2.4063	0.3582	
TCGA-DU-7302	36	-0.4170	-0.3568	
TCGA-E1-5302	37	-1.3578	-0.1524	
TCGA-E1-5303	38	-1.3774	-0.9160	

TCGA-E1-5304	39	-2.4572	-0.3503	
TCGA-E1-5305	40	-2.8153	0.6791	
TCGA-E1-5307	41	-2.4712	0.3040	
TCGA-E1-5311	42	-0.3180	-1.3018	
TCGA-E1-5318	43	-2.7072	0.3332	
TCGA-E1-5319	44	-1.3902	-0.6510	
TCGA-E1-5322	45	-3.1368	1.1440	
TCGA-HT-7480	46	-1.8090	0.7358	
TCGA-HT-7481	47	-1.4217	0.1430	
TCGA-HT-7482	48	-3.0045	0.8311	
TCGA-HT-7483	49	-3.1403	2.2846	
TCGA-HT-7485	50	-2.1111	-0.3562	
TCGA-HT-7601	51	-0.8142	-0.1894	
TCGA-HT-7602	52	-2.3967	-0.0180	
TCGA-HT-7603	53	-0.3573	-0.3107	
TCGA-HT-7604	54	-2.9863	1.1282	
TCGA-HT-7605	55	-0.4961	-0.6659	
TCGA-HT-7606	56	-1.7123	-0.5100	
TCGA-HT-7608	57	-2.1408	-0.5593	
TCGA-HT-7609	58	-3.1335	1.1732	
TCGA-HT-7610	59	-0.9535	0.2811	
TCGA-HT-7620	60	-1.4036	-0.4811	
TCGA-HT-7676	61	-3.4217	1.2956	
TCGA-HT-7681	62	-0.6528	-0.4944	
TCGA-HT-7684	63	-1.9025	-1.0720	
TCGA-HT-7686	64	-1.5654	0.3481	
TCGA-HT-7691	65	2.0365	1.9672	x
TCGA-HT-7693	66	-2.3374	0.5214	
TCGA-HT-7694	67	1.0140	-1.3315	x
TCGA-HT-7855	68	-3.5669	1.7929	
TCGA-HT-7856	69	1.6708	-1.5879	x
TCGA-HT-7857	70	5.7183	-0.4125	x
TCGA-HT-7858	71	-2.5942	1.1474	
TCGA-HT-7874	72	0.8565	-1.0560	x
TCGA-HT-7875	73	-2.5013	0.7811	
TCGA-HT-8013	74	-1.5152	0.2521	
TCGA-HT-8113	75	1.7669	-1.6879	x
TCGA-CS-4941	76	7.9784	-0.4064	
TCGA-CS-5397	77	7.6077	-1.0111	
TCGA-DU-5852	78	6.5219	-1.5981	
TCGA-DU-6400	79	0.0951	-1.3700	
TCGA-DU-6402	80	16.6182	7.1866	
TCGA-DU-6403	81	5.3630	-0.6441	

TCGA-DU-7006	82	10.6607	2.4647	
TCGA-DU-7012	83	10.0215	-0.2709	
TCGA-DU-7013	84	5.8626	-1.9139	
TCGA-DU-7290	85	11.8734	-2.5565	
TCGA-DU-7292	86	8.1825	-1.9295	
TCGA-DU-8158	87	10.4996	1.6978	
TCGA-HT-7607	88	1.8597	-1.5372	
TCGA-HT-7616	89	1.2813	-0.9717	

Results of the supervised PCA of all 89 subjects in connection with the 29 most significant genes. The scores of all 89 subjects according to the first two principal components (PC1 and PC2) are shown. According to the supervised PCA, if a subject has a PC1 score < 0 , then that subject is classified as LTS subject; otherwise, if a subject has a PC1 score > 0 , then that subject is classified as STS subject. As can be seen, the supervised PCA misclassified 11 subjects (# 5, 10, 18, 25, 28, 65, 67, 69, 70, 72, and 75) out of 89 subjects; that yields a misclassification rate of 0.12360. “x” denotes a misclassification.

Table S7. Overall results of the F_2 function

Subject ID	Subject No	F ₁ Score	Notes
TCGA-CS-4938	1	20.6031	
TCGA-CS-4942	2	20.4138	
TCGA-CS-5390	3	18.7324	
TCGA-CS-5394	4	19.0248	
TCGA-DB-5270	5	23.5807	
TCGA-DB-5273	6	20.0999	
TCGA-DB-5274	7	22.8516	
TCGA-DB-5275	8	19.5583	
TCGA-DB-5276	9	22.2236	
TCGA-DB-5277	10	22.0500	
TCGA-DB-5278	11	20.3211	
TCGA-DB-5279	12	24.4905	
TCGA-DB-5280	13	16.4466	
TCGA-DB-5281	14	21.0244	
TCGA-DB-A4XG	15	18.2326	
TCGA-DH-5141	16	23.1896	
TCGA-DU-5870	17	20.0020	
TCGA-DU-6392	18	22.9112	
TCGA-DU-6393	19	24.0155	
TCGA-DU-6395	20	18.9559	
TCGA-DU-6396	21	16.3499	
TCGA-DU-6397	22	21.5289	
TCGA-DU-6399	23	19.3396	
TCGA-DU-6401	24	14.9465	
TCGA-DU-6404	25	29.0098	x
TCGA-DU-6407	26	17.0150	
TCGA-DU-6408	27	18.4863	
TCGA-DU-7007	28	23.4241	
TCGA-DU-7008	29	19.2650	
TCGA-DU-7009	30	18.4687	
TCGA-DU-7011	31	20.0272	
TCGA-DU-7014	32	17.4065	
TCGA-DU-7015	33	18.8721	
TCGA-DU-7294	34	18.7784	
TCGA-DU-7299	35	18.8136	
TCGA-DU-7302	36	20.3283	
TCGA-E1-5302	37	20.6246	
TCGA-E1-5303	38	21.7771	

TCGA-E1-5304	39	20.6672	
TCGA-E1-5305	40	18.4177	
TCGA-E1-5307	41	18.1212	
TCGA-E1-5311	42	22.7442	
TCGA-E1-5318	43	21.1144	
TCGA-E1-5319	44	21.1220	
TCGA-E1-5322	45	19.2545	
TCGA-HT-7480	46	18.1003	
TCGA-HT-7481	47	20.3288	
TCGA-HT-7482	48	20.0125	
TCGA-HT-7483	49	18.8941	
TCGA-HT-7485	50	21.2371	
TCGA-HT-7601	51	19.4745	
TCGA-HT-7602	52	21.4121	
TCGA-HT-7603	53	26.1538	x
TCGA-HT-7604	54	17.5262	
TCGA-HT-7605	55	22.1334	
TCGA-HT-7606	56	19.0777	
TCGA-HT-7608	57	23.0015	
TCGA-HT-7609	58	17.3158	
TCGA-HT-7610	59	18.8204	
TCGA-HT-7620	60	20.1659	
TCGA-HT-7676	61	17.5183	
TCGA-HT-7681	62	22.2633	
TCGA-HT-7684	63	22.6417	
TCGA-HT-7686	64	21.4548	
TCGA-HT-7691	65	24.2259	
TCGA-HT-7693	66	17.9358	
TCGA-HT-7694	67	23.5650	
TCGA-HT-7855	68	14.5007	
TCGA-HT-7856	69	24.0601	
TCGA-HT-7857	70	28.9318	x
TCGA-HT-7858	71	17.0243	
TCGA-HT-7874	72	23.5586	
TCGA-HT-7875	73	19.1403	
TCGA-HT-8013	74	18.1473	
TCGA-HT-8113	75	25.1102	
TCGA-CS-4941	76	30.8289	
TCGA-CS-5397	77	28.7981	
TCGA-DU-5852	78	26.9029	
TCGA-DU-6400	79	22.3210	x
TCGA-DU-6402	80	37.8965	
TCGA-DU-6403	81	28.5118	

TCGA-DU-7006	82	33.7616	
TCGA-DU-7012	83	31.4257	
TCGA-DU-7013	84	31.8338	
TCGA-DU-7290	85	35.6830	
TCGA-DU-7292	86	30.0601	
TCGA-DU-8158	87	31.0498	
TCGA-HT-7607	88	25.2198	
TCGA-HT-7616	89	27.9091	

The F_1 scores of all 89 subjects [75 LTS (# 1-75) and 14 STS (# 76-89)] are listed. Those results of the overall performance of the F_1 function were obtained by combining the results from the development phase (listed in **Table S3**) with those from the validation phase (listed in **Table S4**). Using the final cut-off score of 25.165, if a subject has an F_1 score < 25.165 , then that subject is classified as LTS; otherwise, if a subject has an F_1 score ≥ 25.165 , then that subject is classified as STS. As can be seen, all of the 75 LTS subjects were classified correctly except three (# 25, 53, & 70), and all of the 14 STS subjects were classified correctly except one (# 79). “x” denotes a misclassification.