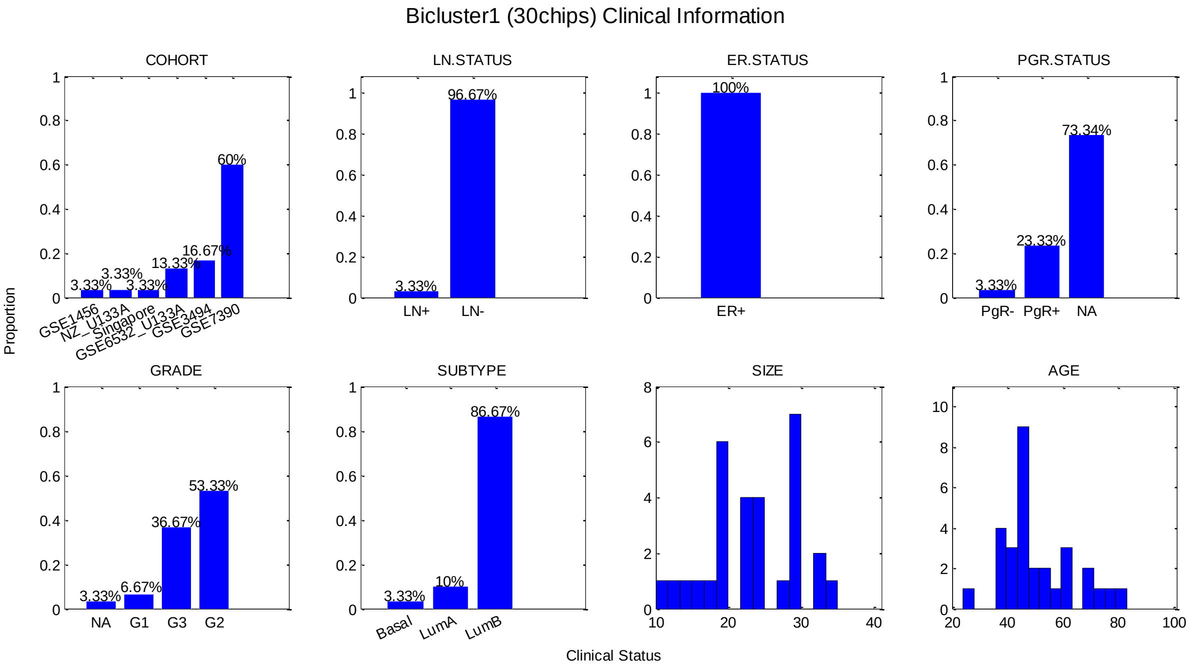
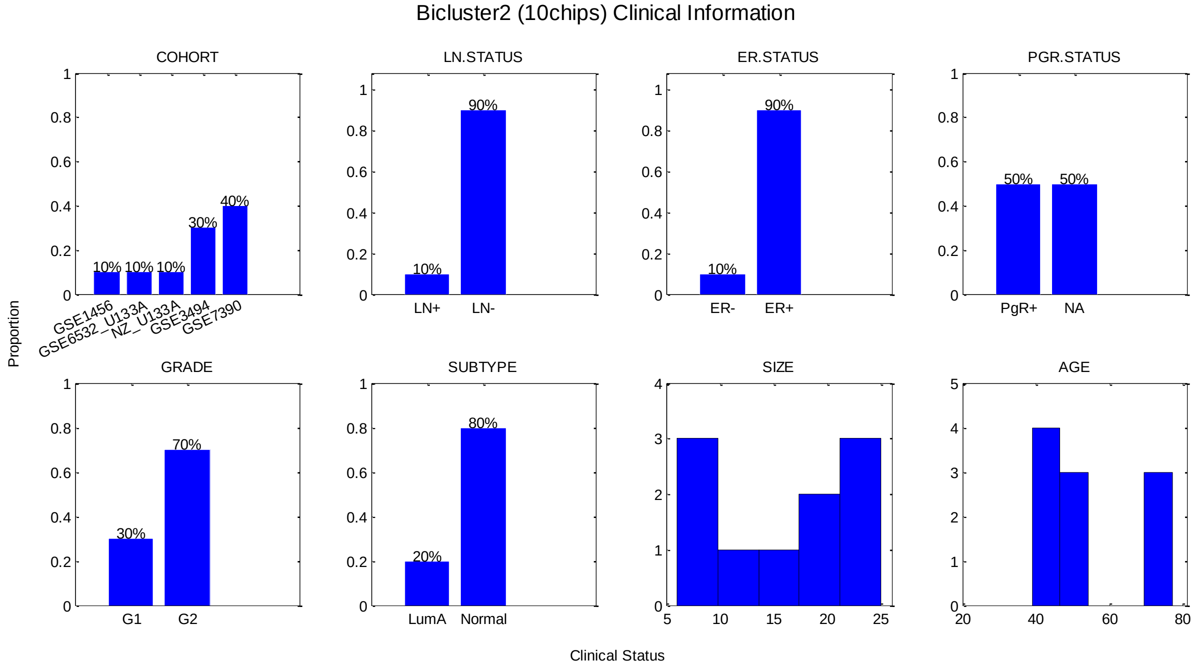
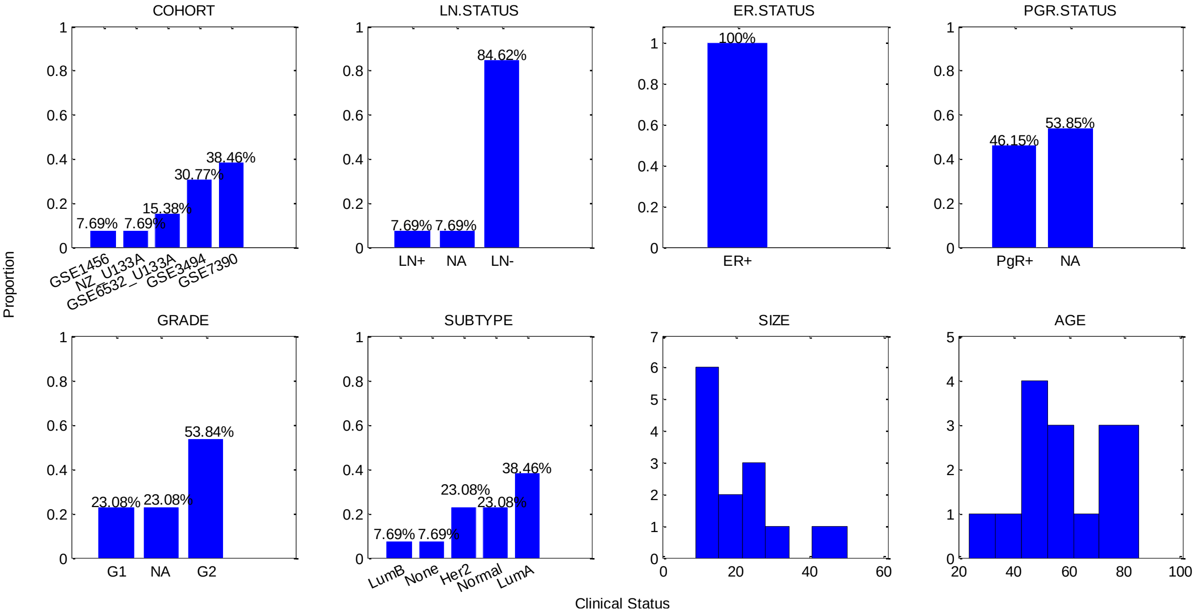
Supplementary Figure 1. Clinical information for each individual bicluster determined by cMonkey from gene expression data of breast cancer patients who had not received adjuvant treatment (437 patients). The clinical variables include cohort, lymph node status (LN.status: LN+, LN- and NA), estrogen receptor status (ER.status: ER+, ER- and NA), progesterone receptor status (PgR.status: PgR+, PgR- and NA), tumour grade (Grade: G1, G2, G3 and NA), molecular subtype (Subtype: Basal, Her2, Luminal A, Luminal B, Normal and None), tumour size (Size in mm) and patient Age (in years). Missing information is noted as NA (Not Available). The percentage noted on each bar indicates the percentage of tumour samples of the specific category in each bicluster.

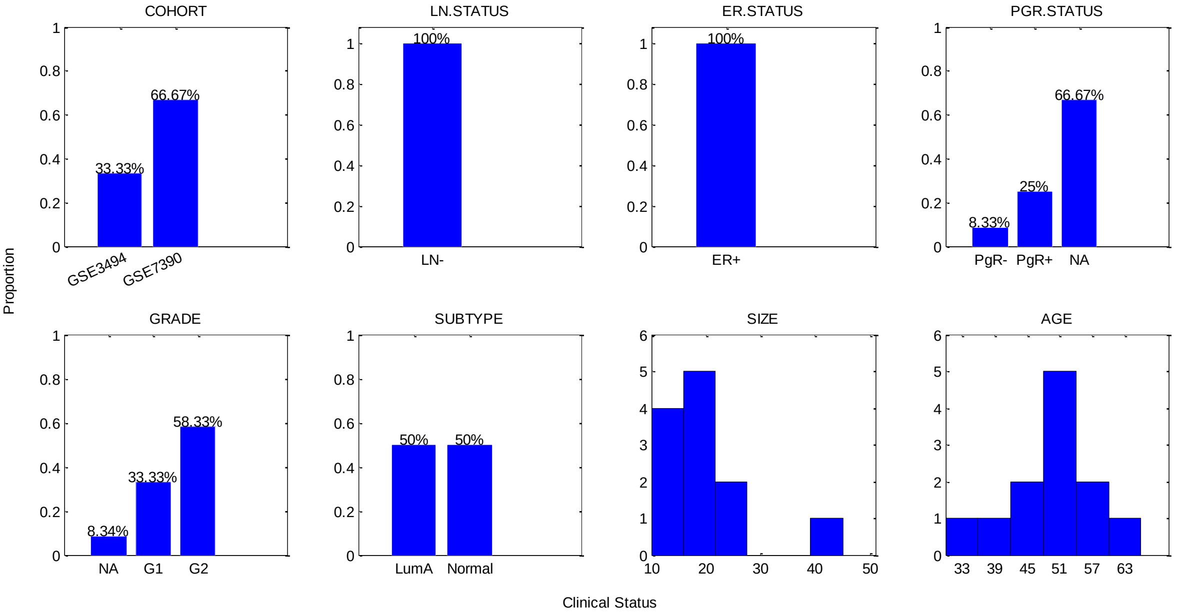




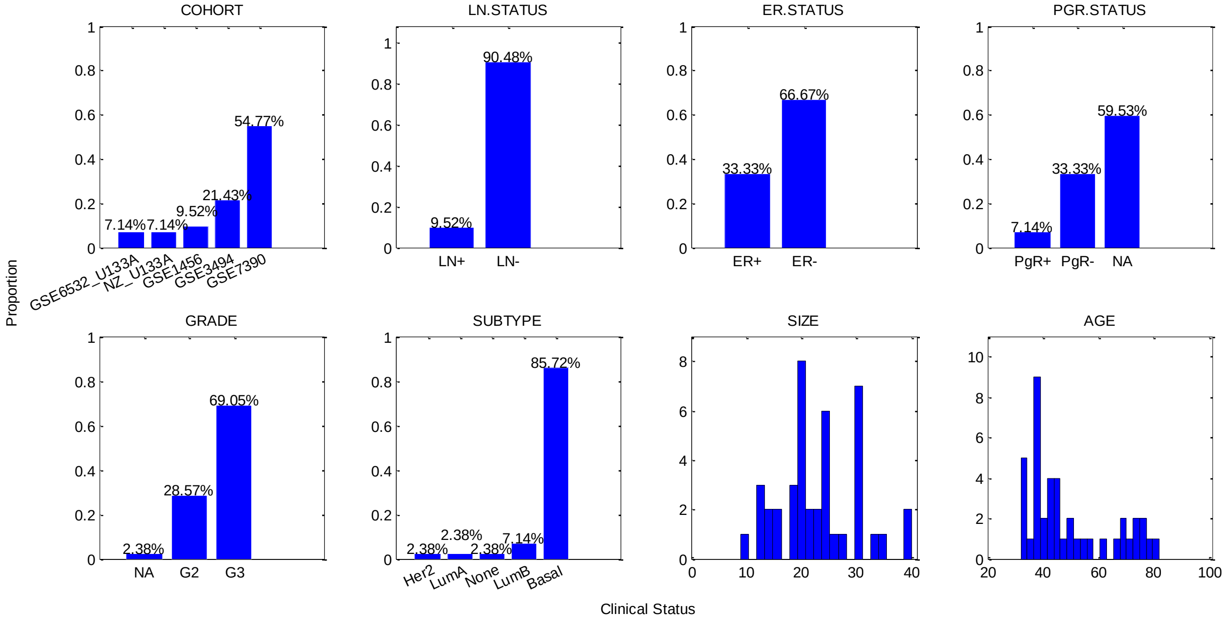
Bicluster3 (13chips) Clinical Information



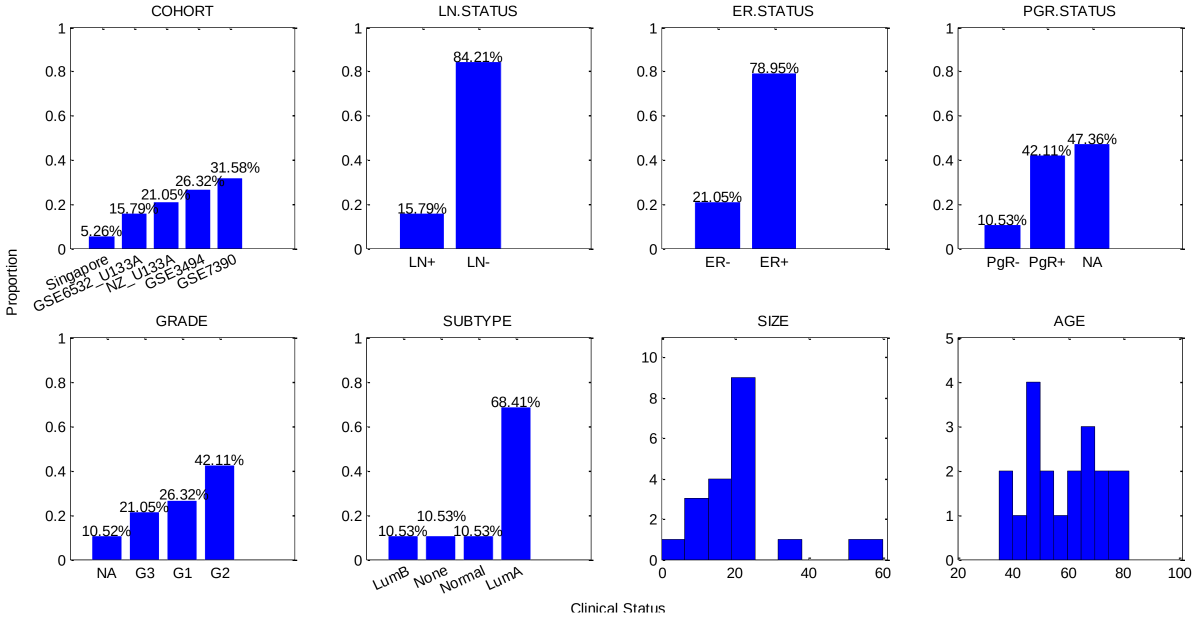
Bicluster4 (12chips) Clinical Information



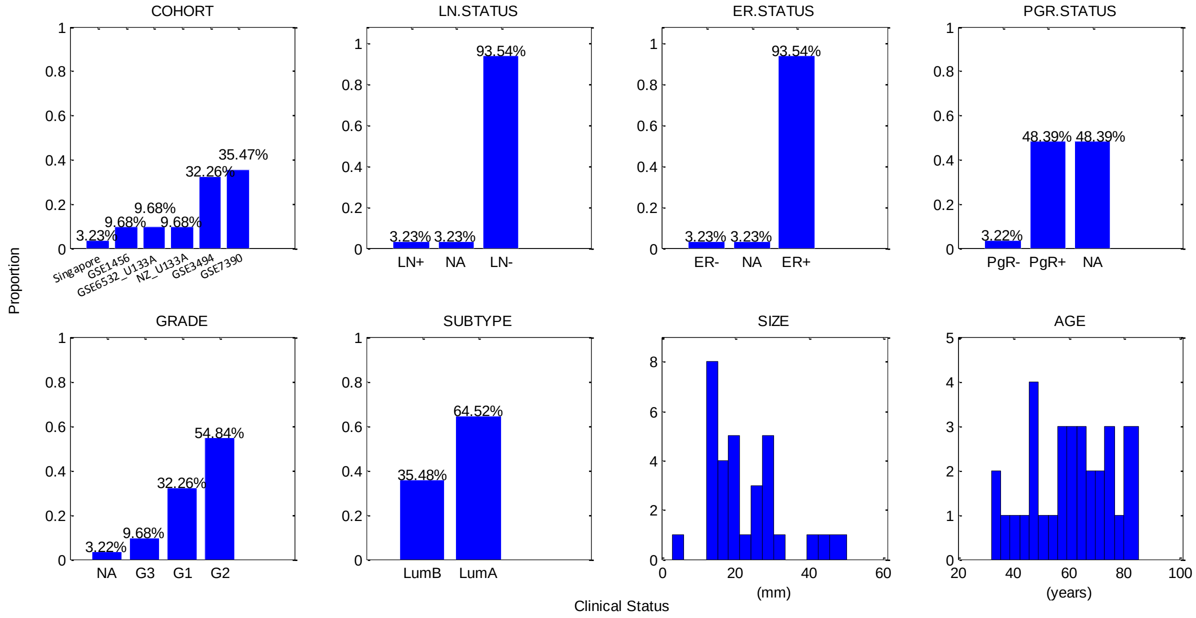
Bicluster5 (42chips) Clinical Information



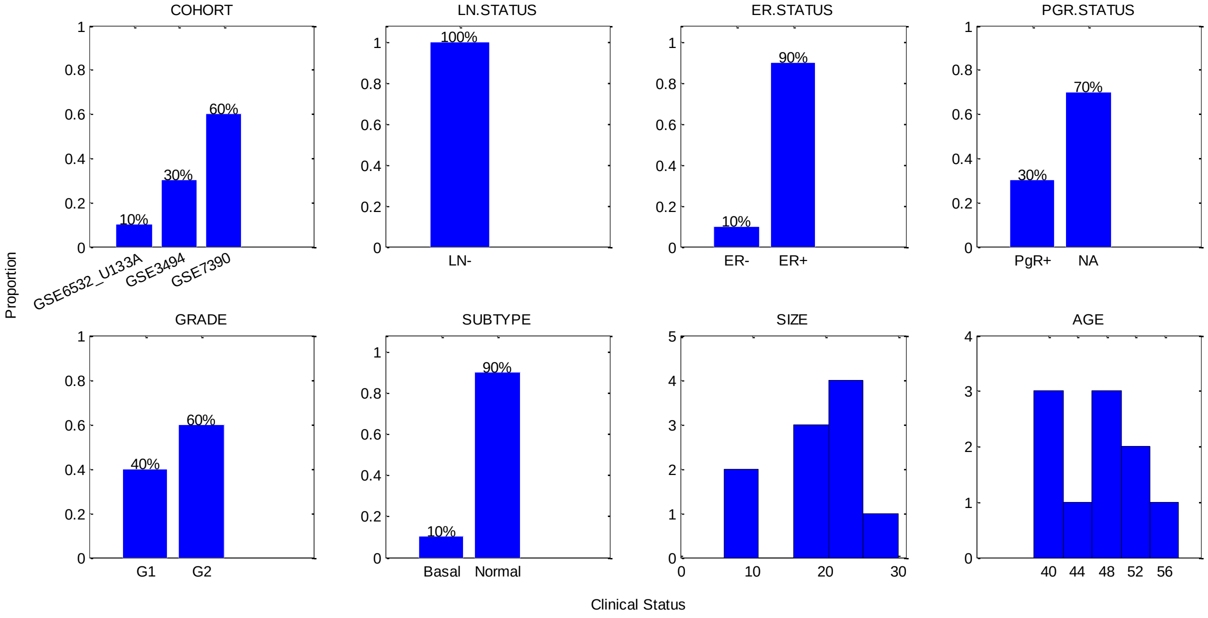
Bicluster6 (19chips) Clinical Information

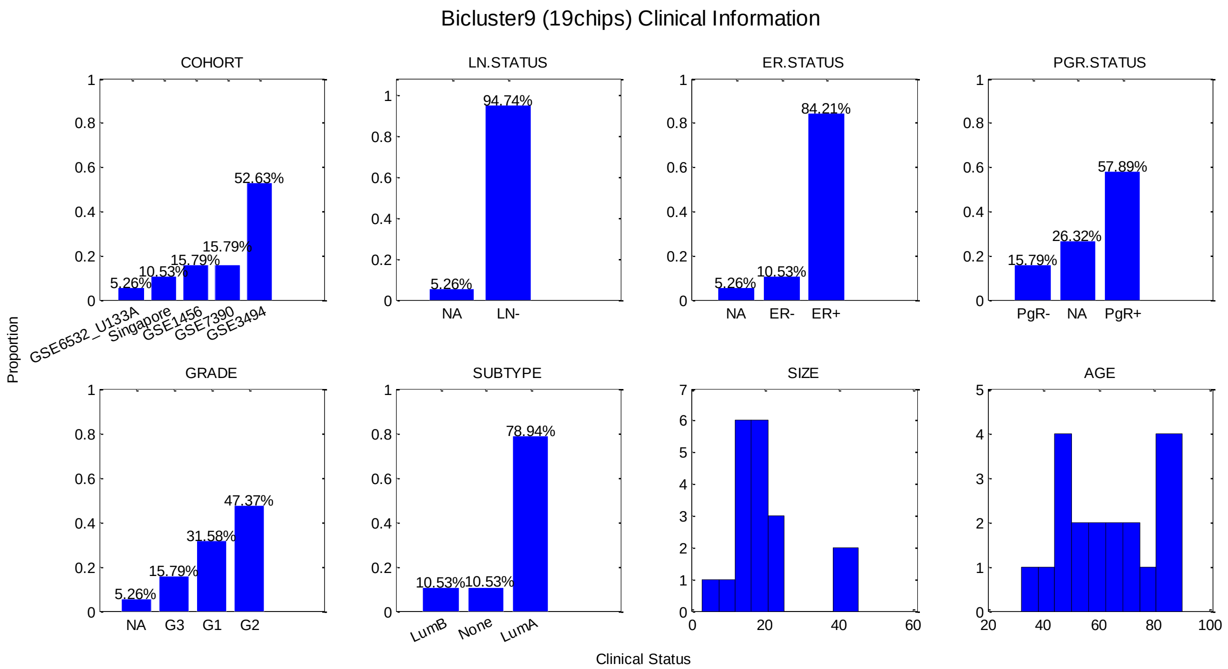


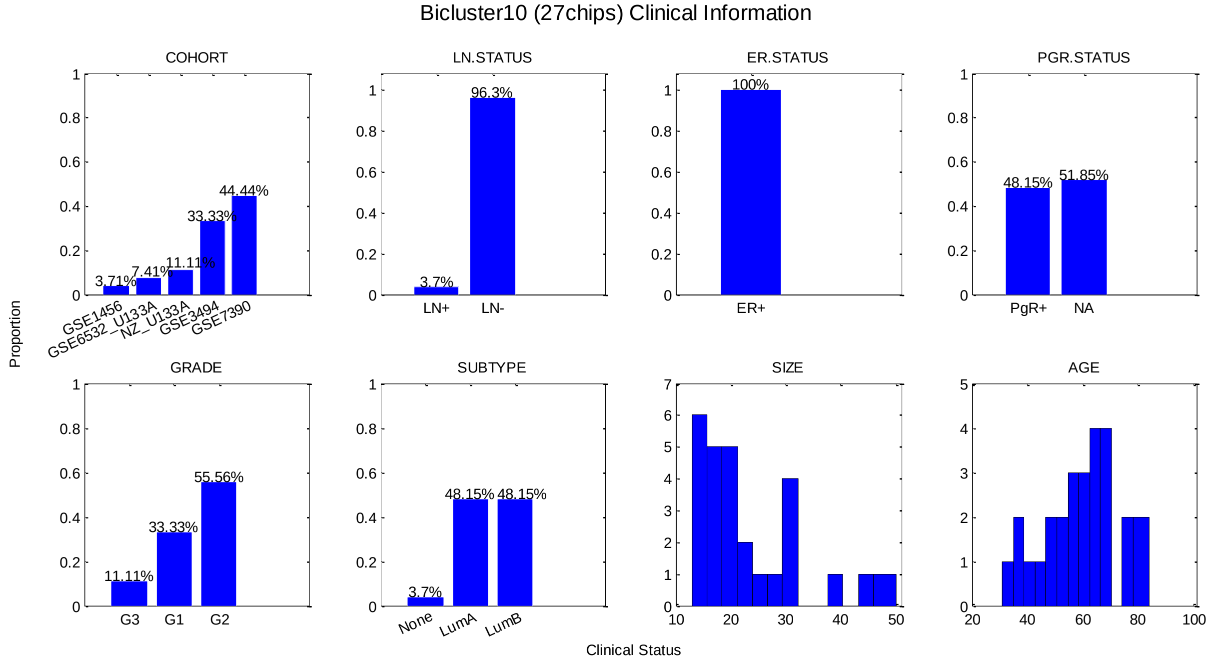
Bicluster7 (31chips) Clinical Information

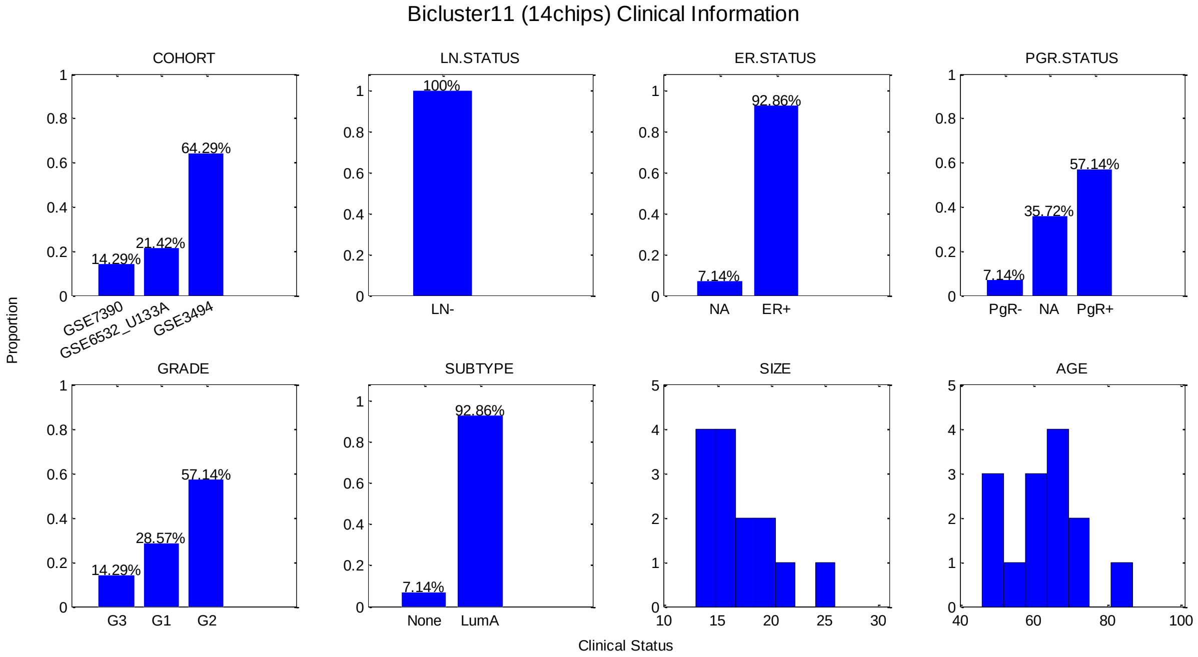


Bicluster8 (10chips) Clinical Information

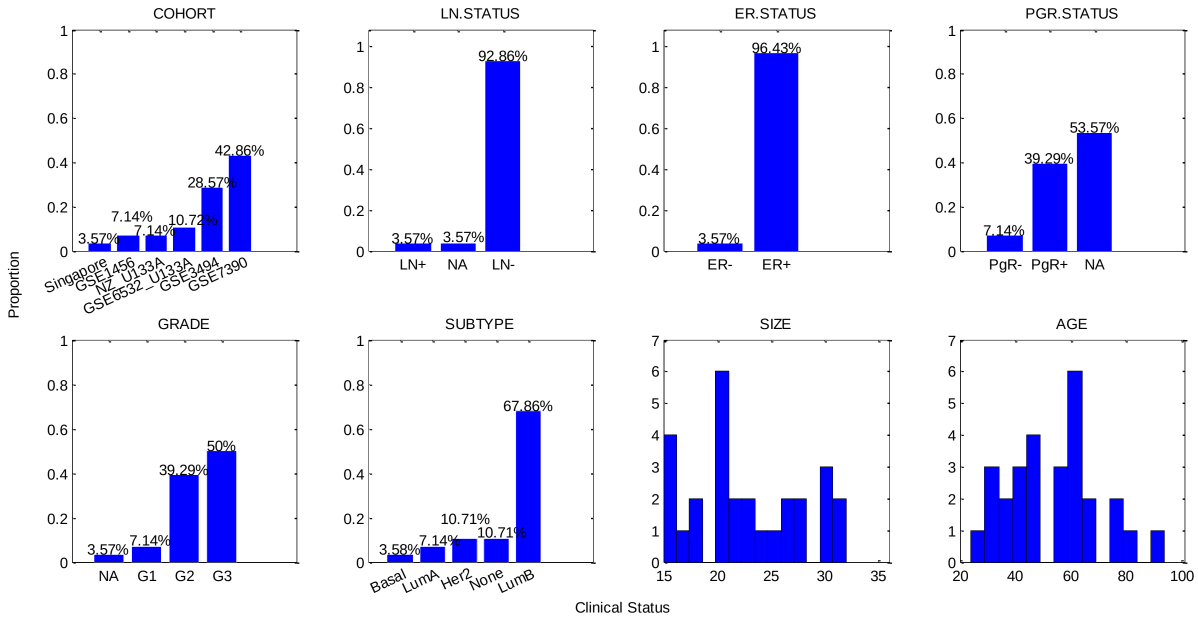



Bicluster9 (19chips) Clinical Information
COHORT
5.26%
10.53%
15.79%
15.79%
52.63%
GSE6532_U133A
Singapore
GSE1456
GSE7390
GSE3494
LN.STATUS
5.26%
94.74%
NA
LN-
ER.STATUS
5.26%
10.53%
84.21%
NA
ER-
ER+
PGR.STATUS
15.79%
26.32%
57.89%
PgR-
NA
PgR+
Proportion
GRADE
5.26%
15.79%
31.58%
47.37%
NA
G3
G1
G2
SUBTYPE
10.53%
10.53%
78.94%
LumB
None
LumA
SIZE
AGE
Clinical Status

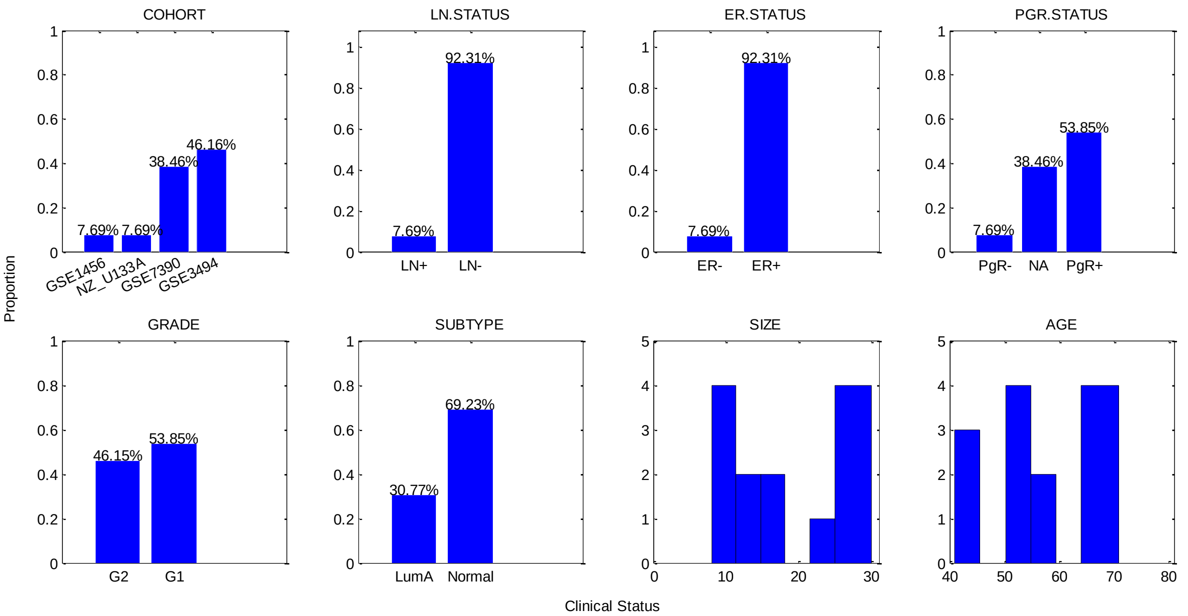

Bicluster10 (27chips) Clinical Information
COHORT
3.71%
7.41%
11.11%
33.33%
44.44%
GSE1456
GSE6532_U133A
NZ_U133A
GSE3494
GSE7390
LN.STATUS
3.7%
96.3%
LN+
LN-
ER.STATUS
100%
ER+
PGR.STATUS
48.15%
51.85%
PgR+
NA
Proportion
GRADE
11.11%
33.33%
55.56%
G3
G1
G2
SUBTYPE
3.7%
48.15%
48.15%
None
LumA
LumB
SIZE
AGE
Clinical Status


Bicluster11 (14chips) Clinical Information
COHORT
14.29%
21.42%
64.29%
GSE7390
GSE6532_U133A
GSE3494
LN.STATUS
100%
LN-
ER.STATUS
7.14%
92.86%
NA
ER+
PGR.STATUS
7.14%
35.72%
57.14%
PgR-
NA
PgR+
Proportion
GRADE
14.29%
28.57%
57.14%
G3
G1
G2
SUBTYPE
7.14%
92.86%
None
LumA
SIZE
AGE
Clinical Status

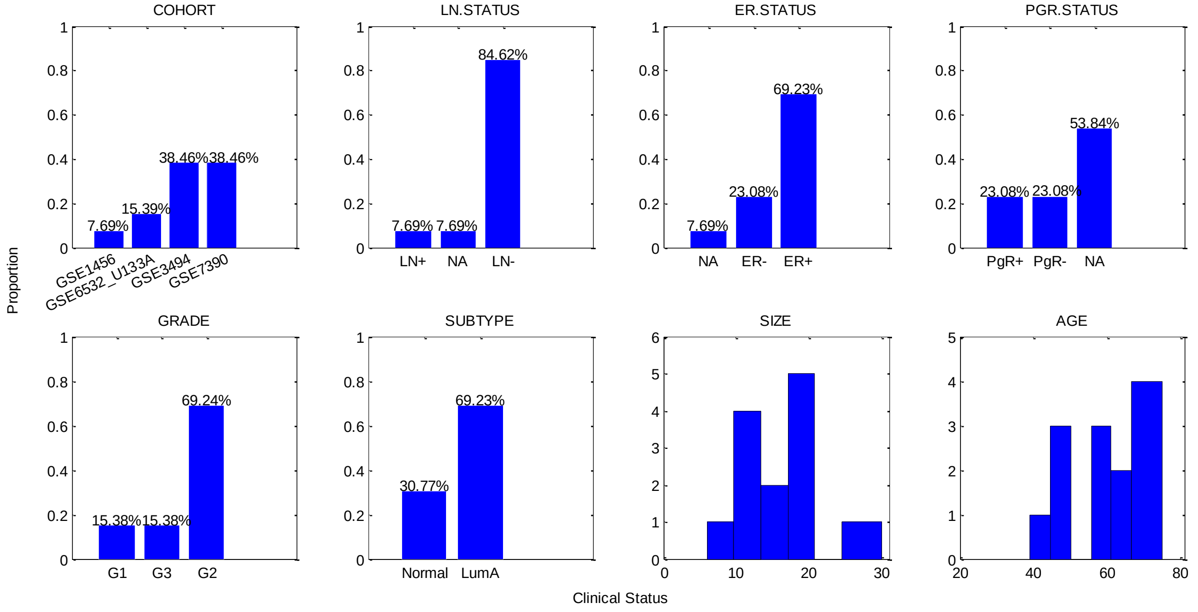
Bicluster12 (28chips) Clinical Information



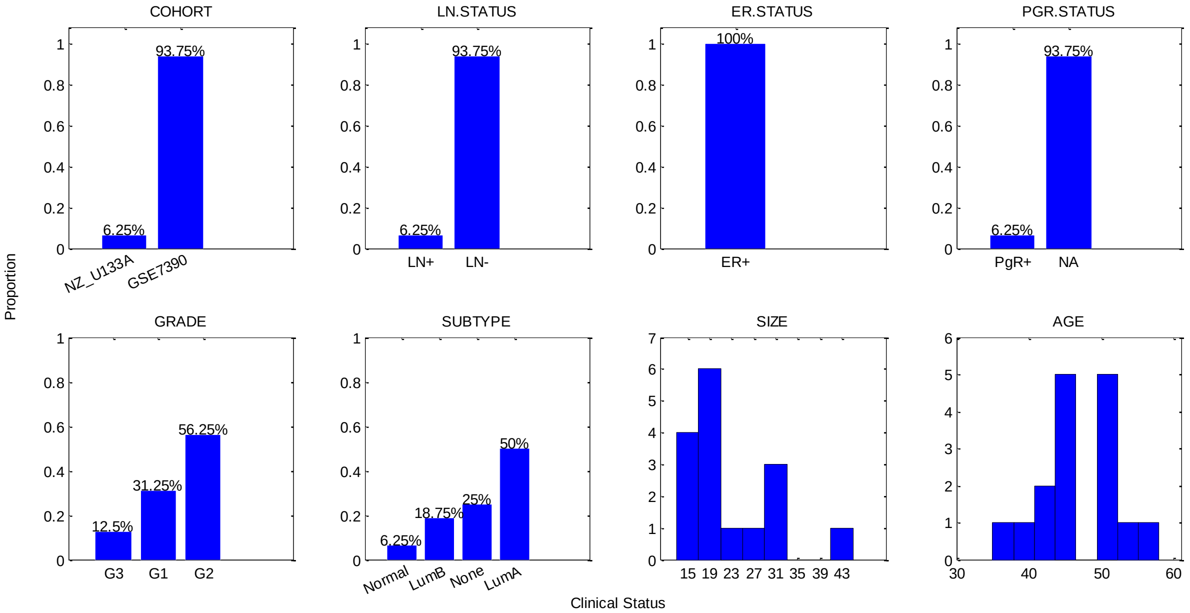
Bicluster13 (13chips) Clinical Information



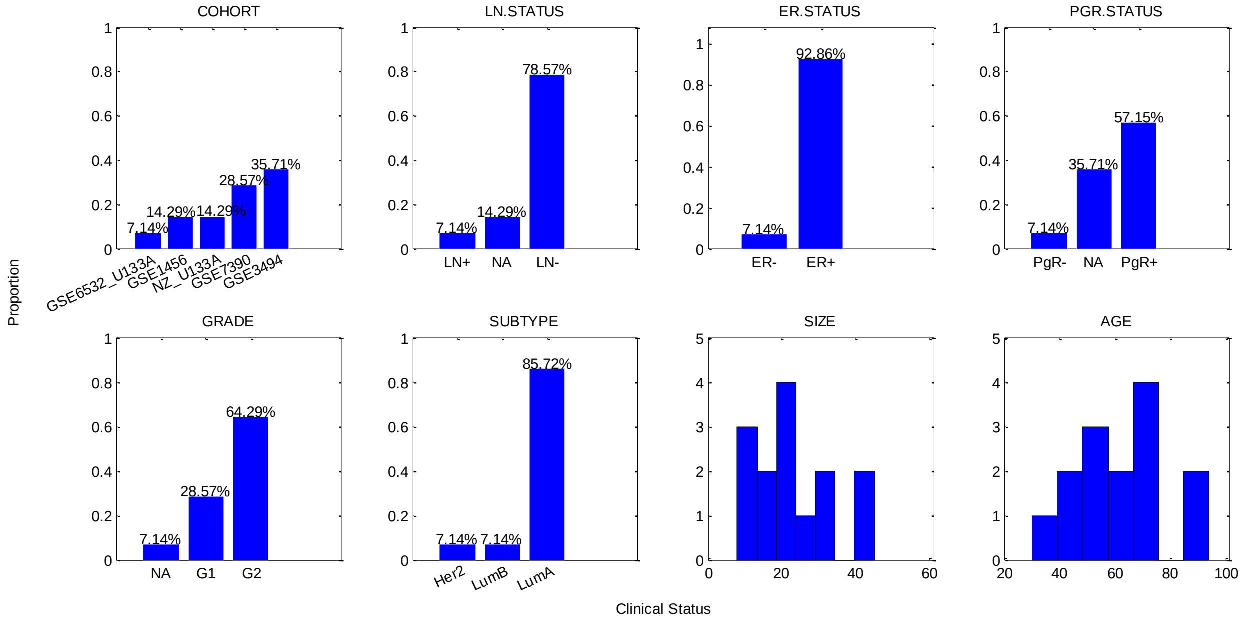
Bicluster14 (13chips) Clinical Information



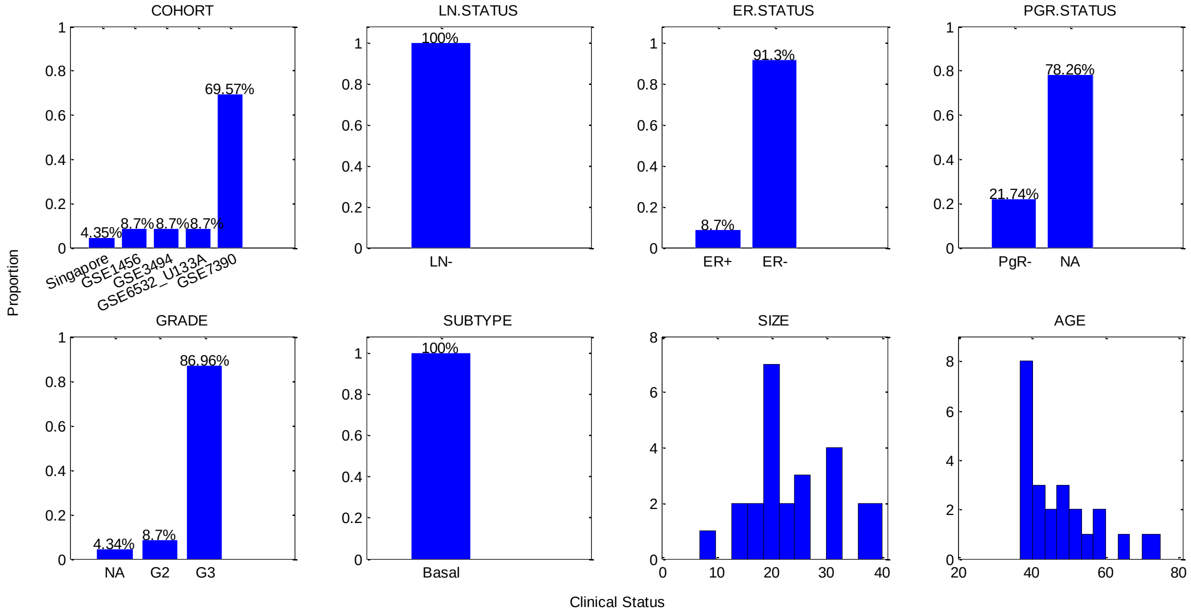
Bicluster15 (16chips) Clinical Information



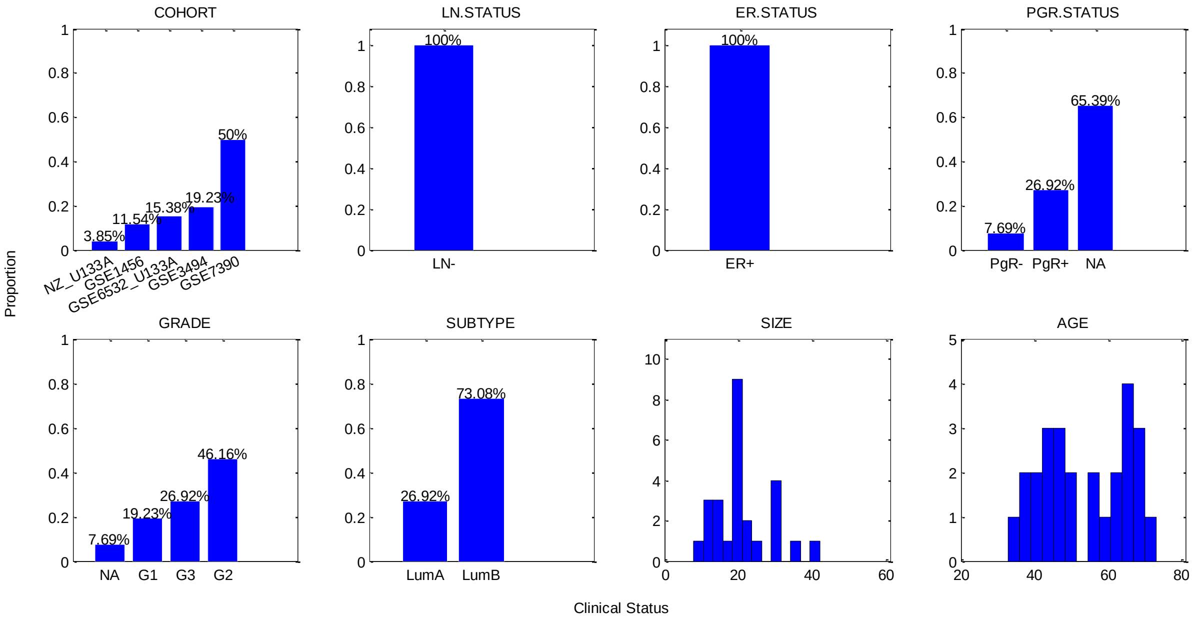
Bicluster16 (14chips) Clinical Information



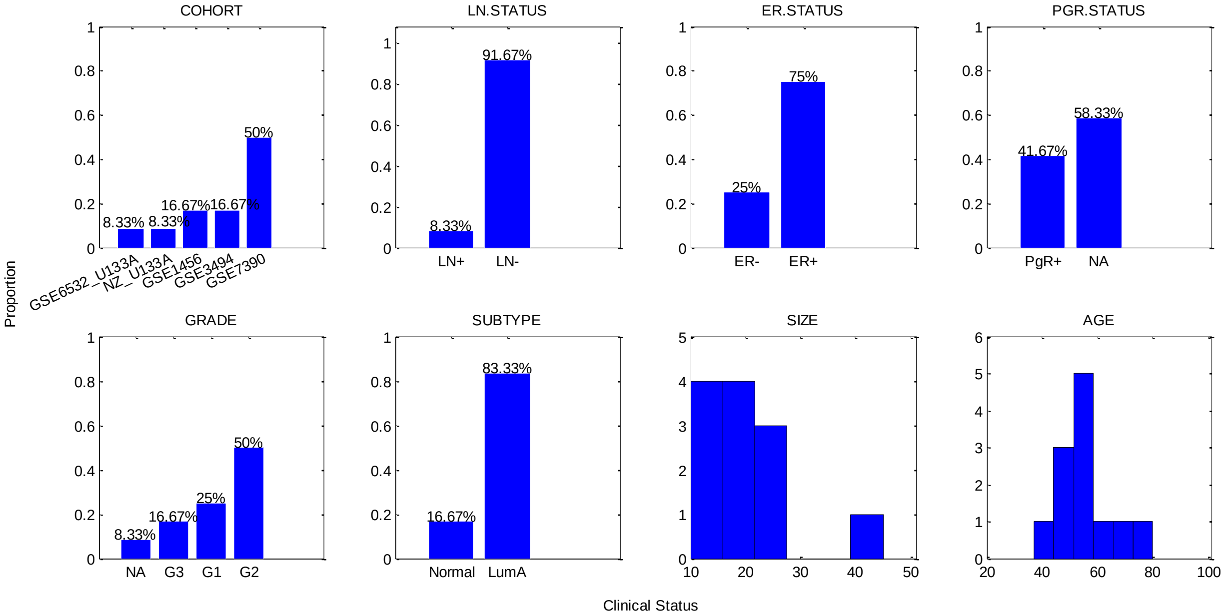
Bicluster17 (23chips) Clinical Information



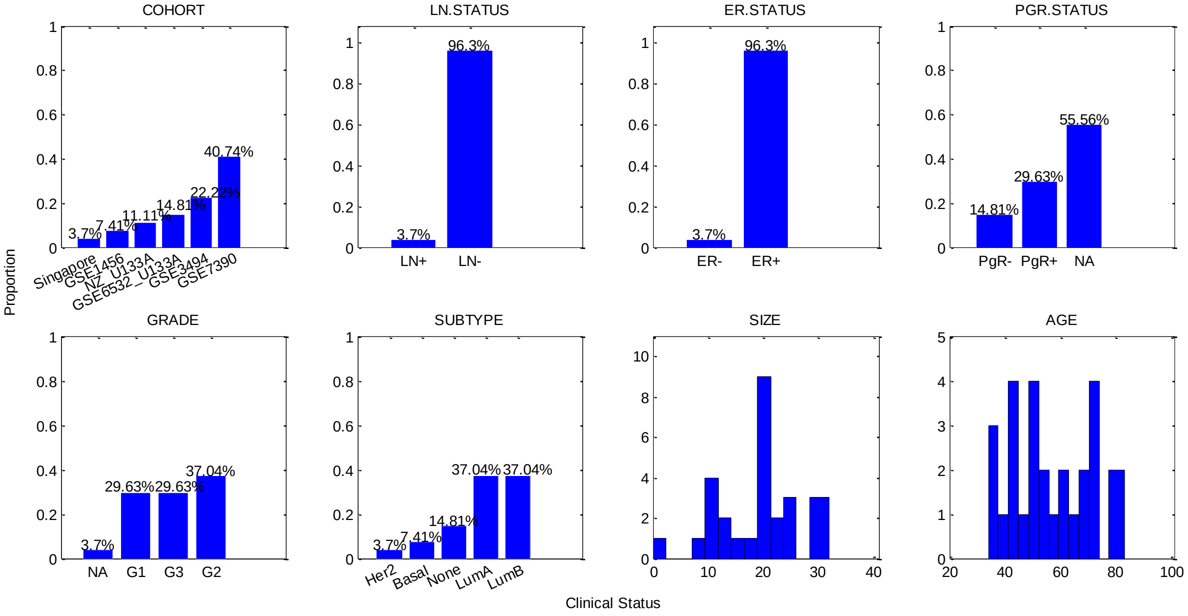
Bicluster18 (26chips) Clinical Information

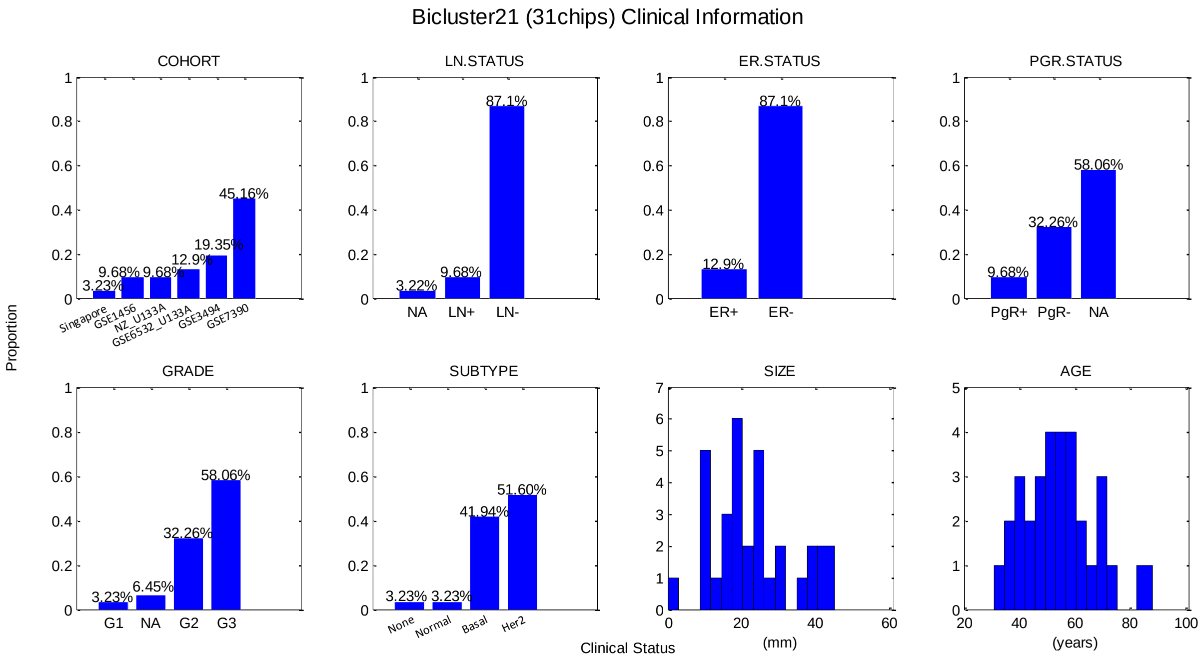


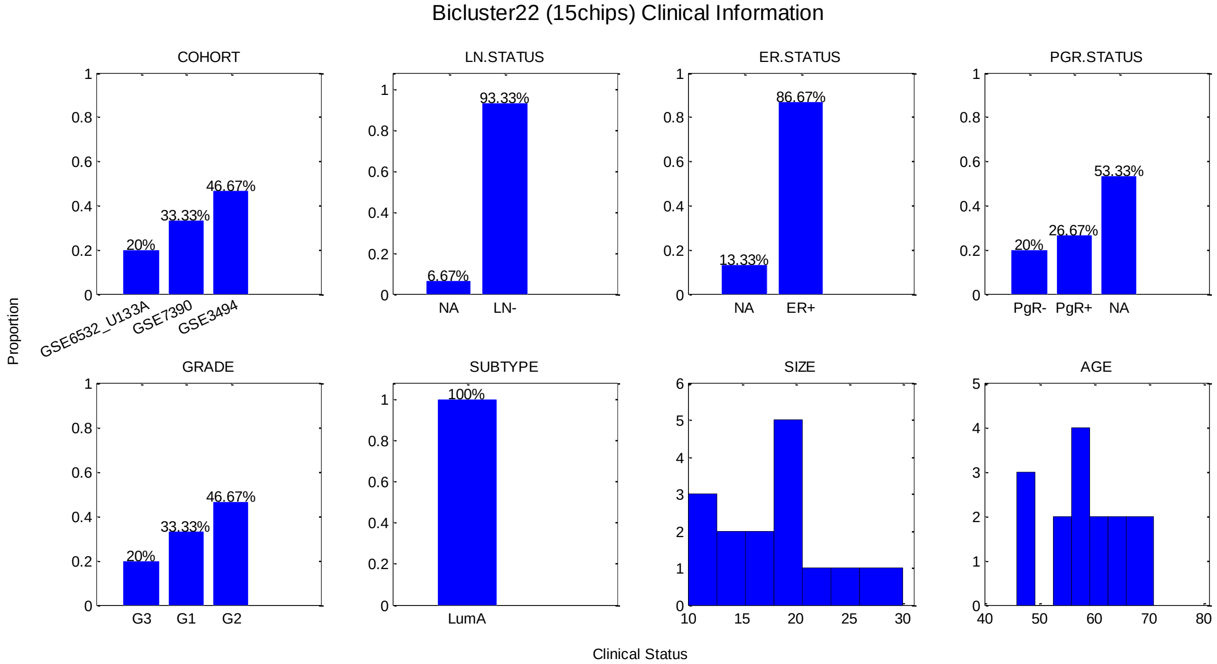
Bicluster19 (12chips) Clinical Information

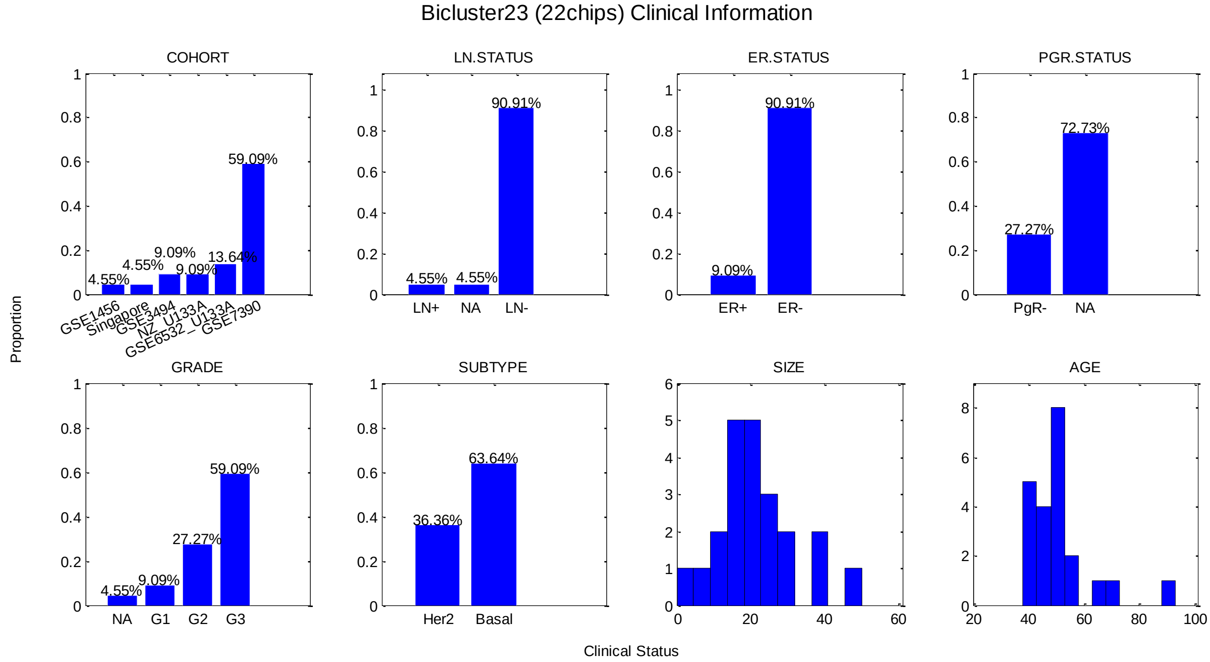


Bicluster20 (27chips) Clinical Information

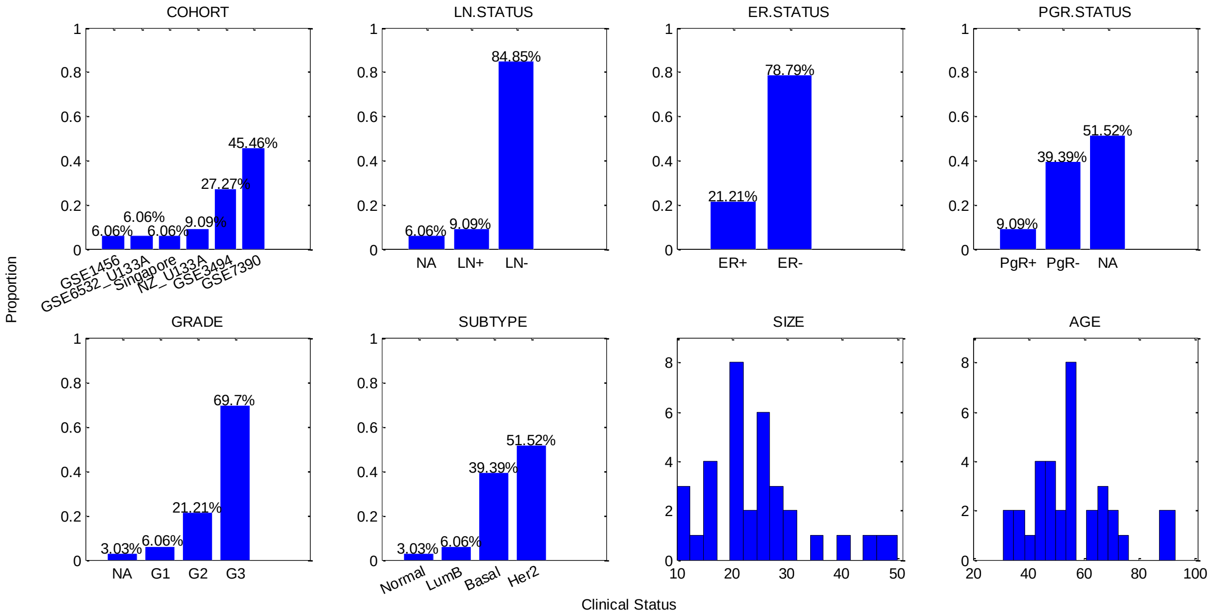



Bicluster21 (31chips) Clinical Information
Proportion
COHORT
3.23%
9.68%
9.68%
12.9%
19.35%
45.16%
Singapore
GSE1456
NZ_U133A
GSE6532_U133A
GSE3494
GSE7390
LN.STATUS
3.22%
9.68%
87.1%
NA
LN+
LN-
ER.STATUS
12.9%
87.1%
ER+
ER-
PGR.STATUS
9.68%
32.26%
58.06%
PgR+
PgR-
NA
GRADE
3.23%
6.45%
32.26%
58.06%
G1
NA
G2
G3
SUBTYPE
3.23%
3.23%
41.94%
51.60%
None
Normal
Basal
Her2
SIZE
(mm)
AGE
(years)
Clinical Status

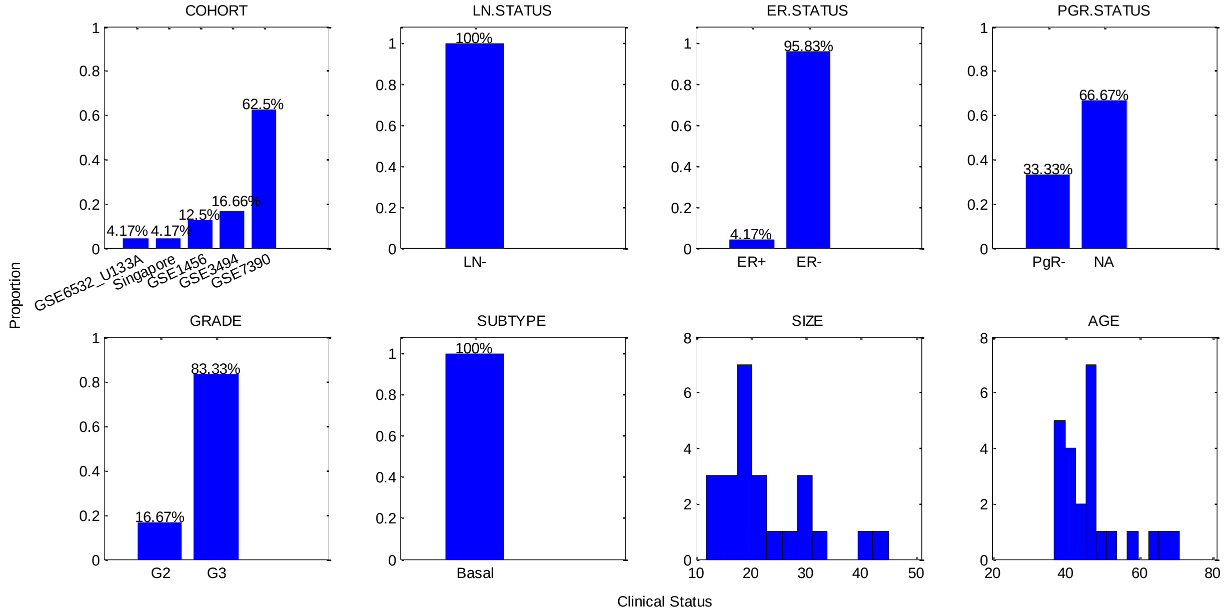

Bicluster22 (15chips) Clinical Information
Proportion
COHORT
20%
33.33%
46.67%
GSE6532_U133A
GSE7390
GSE3494
LN.STATUS
6.67%
93.33%
NA
LN-
ER.STATUS
13.33%
86.67%
NA
ER+
PGR.STATUS
20%
26.67%
53.33%
PgR-
PgR+
NA
GRADE
20%
33.33%
46.67%
G3
G1
G2
SUBTYPE
100%
LumA
SIZE
AGE
Clinical Status


Bicluster23 (22chips) Clinical Information
Proportion
COHORT
4.55%
4.55%
9.09%
9.09%
13.64%
59.09%
GSE1456
Singapore
GSE3494
NZ_U133A
GSE6532_U133A
GSE7390
LN.STATUS
4.55%
4.55%
90.91%
LN+
NA
LN-
ER.STATUS
9.09%
90.91%
ER+
ER-
PGR.STATUS
27.27%
72.73%
PgR-
NA
GRADE
4.55%
9.09%
27.27%
59.09%
NA
G1
G2
G3
SUBTYPE
36.36%
63.64%
Her2
Basal
SIZE
AGE
Clinical Status

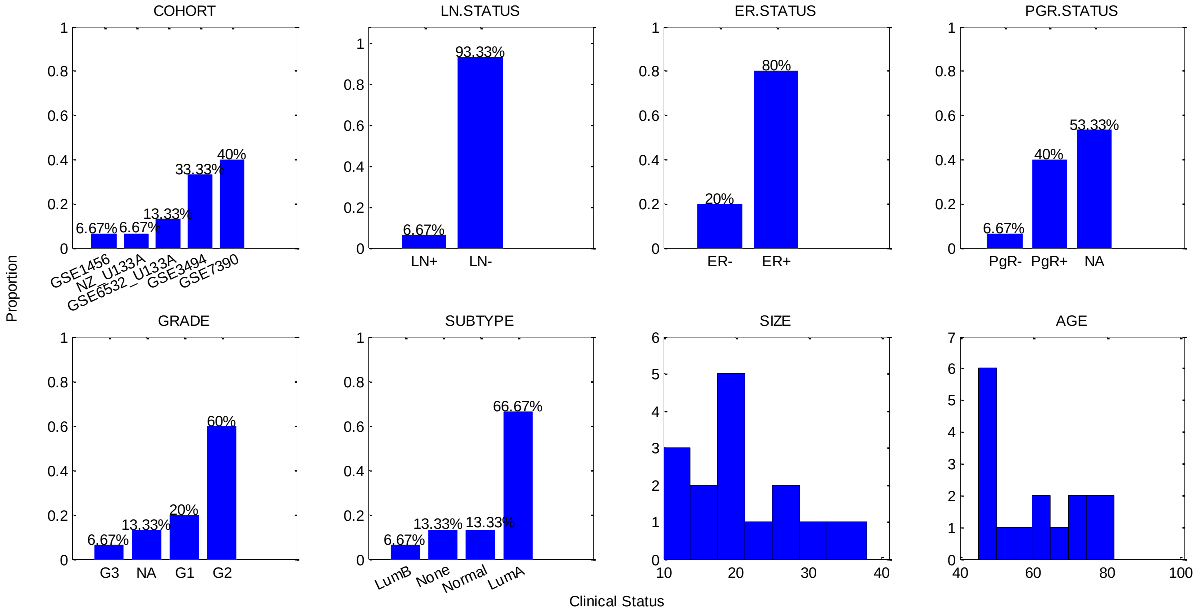
Bicluster24 (33chips) Clinical Information



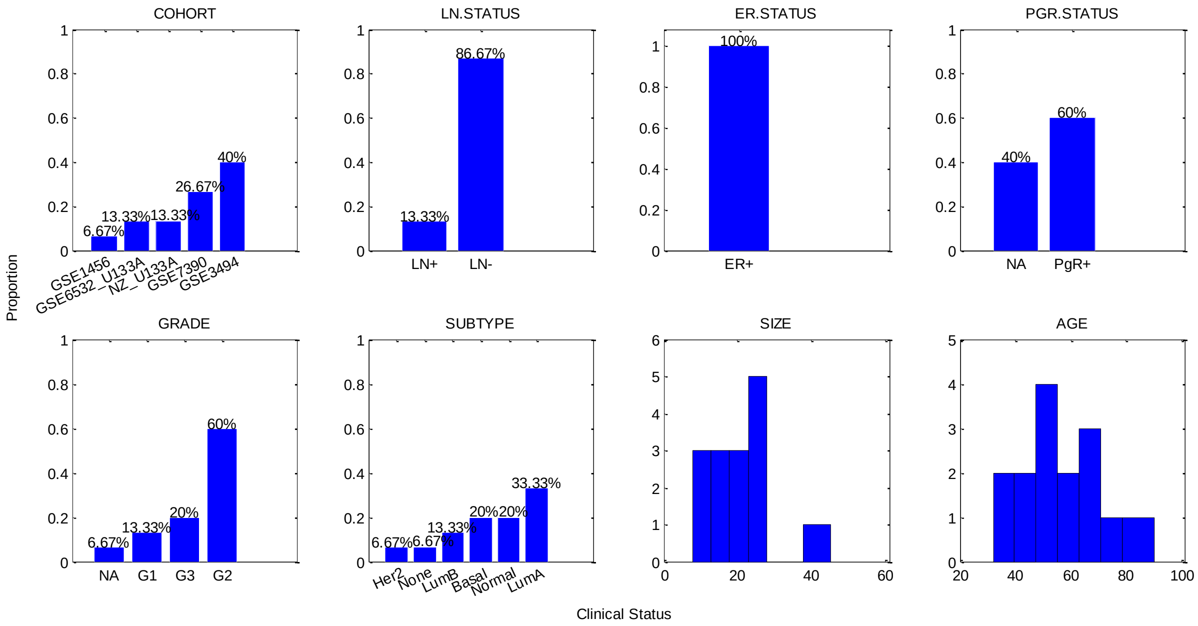
Bicluster25 (24chips) Clinical Information



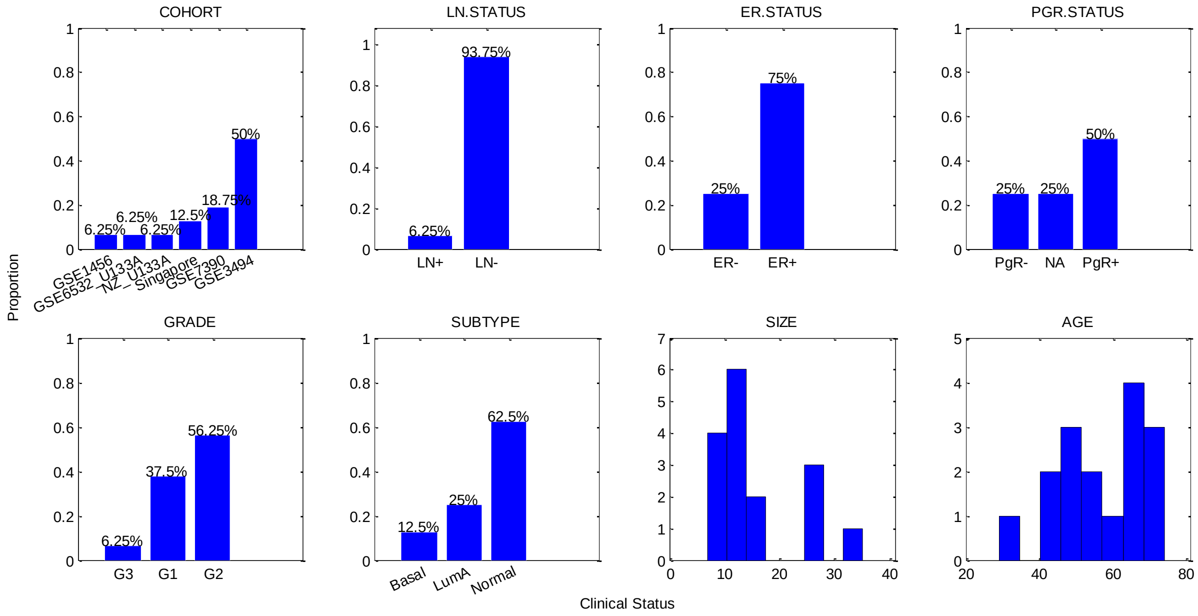
Bicluster26 (15chips) Clinical Information



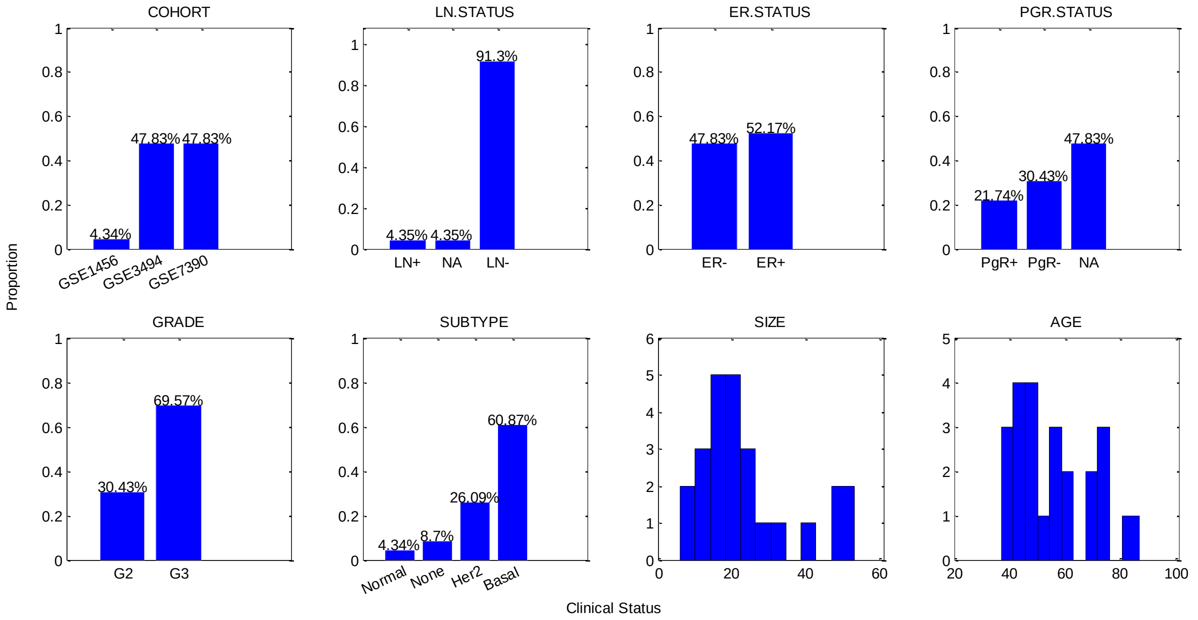
Bicluster27 (15chips) Clinical Information

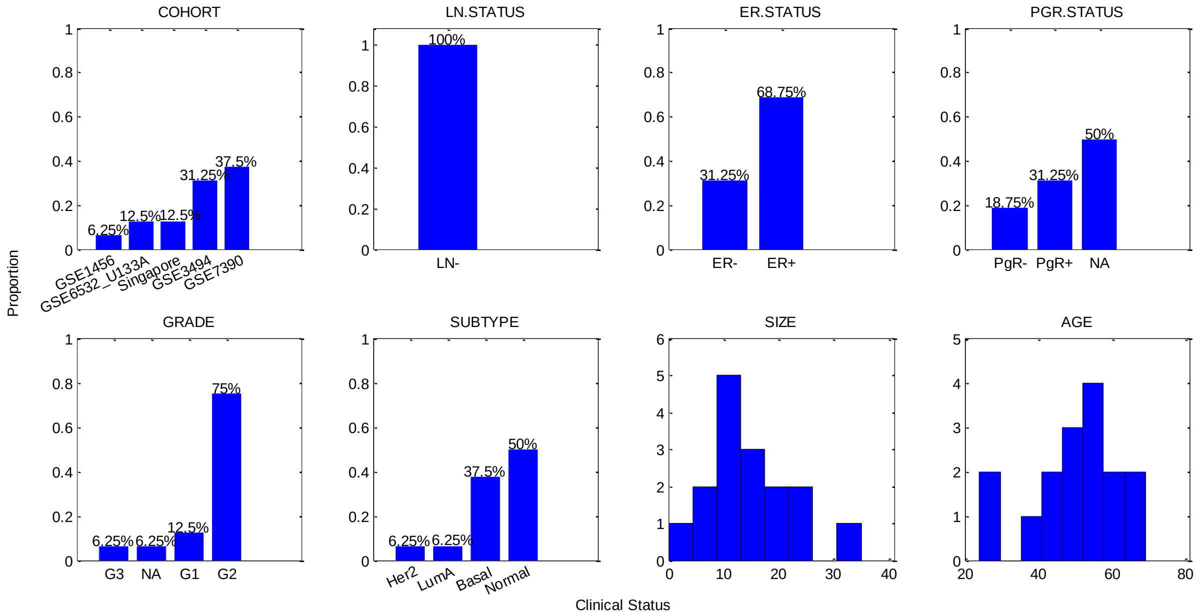


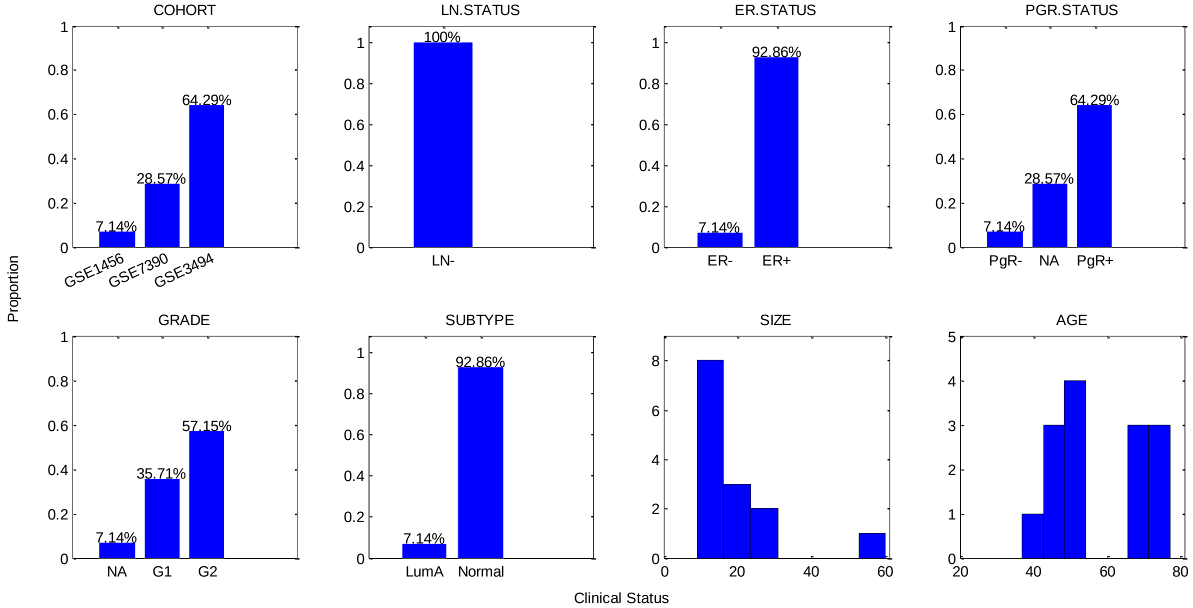
Bicluster28 (16chips) Clinical Information

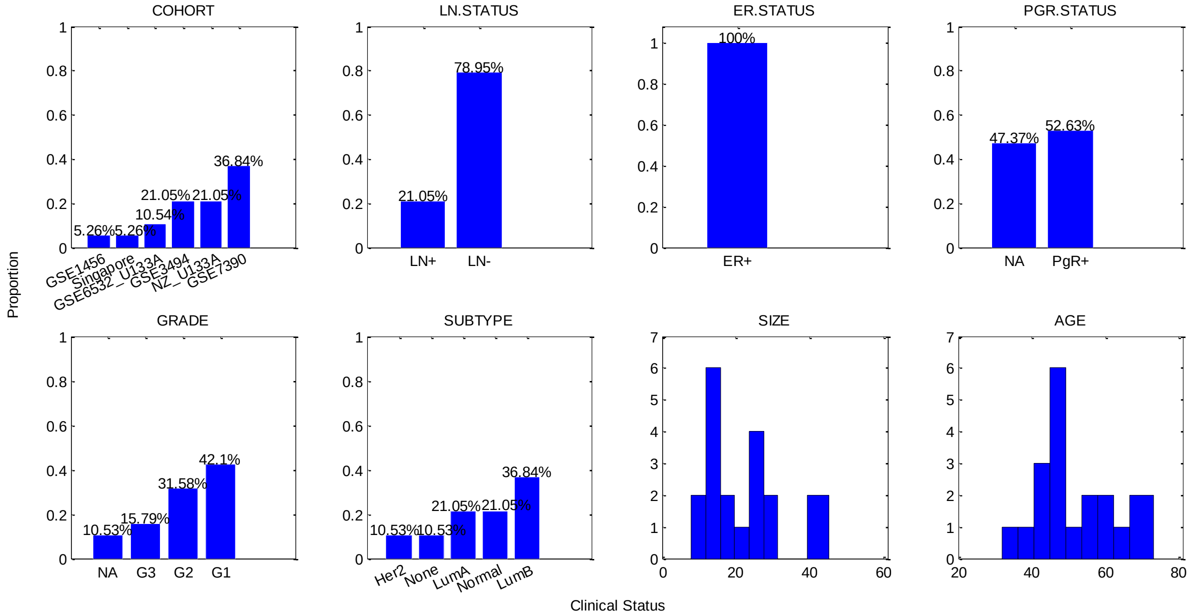


Bicluster29 (23chips) Clinical Information

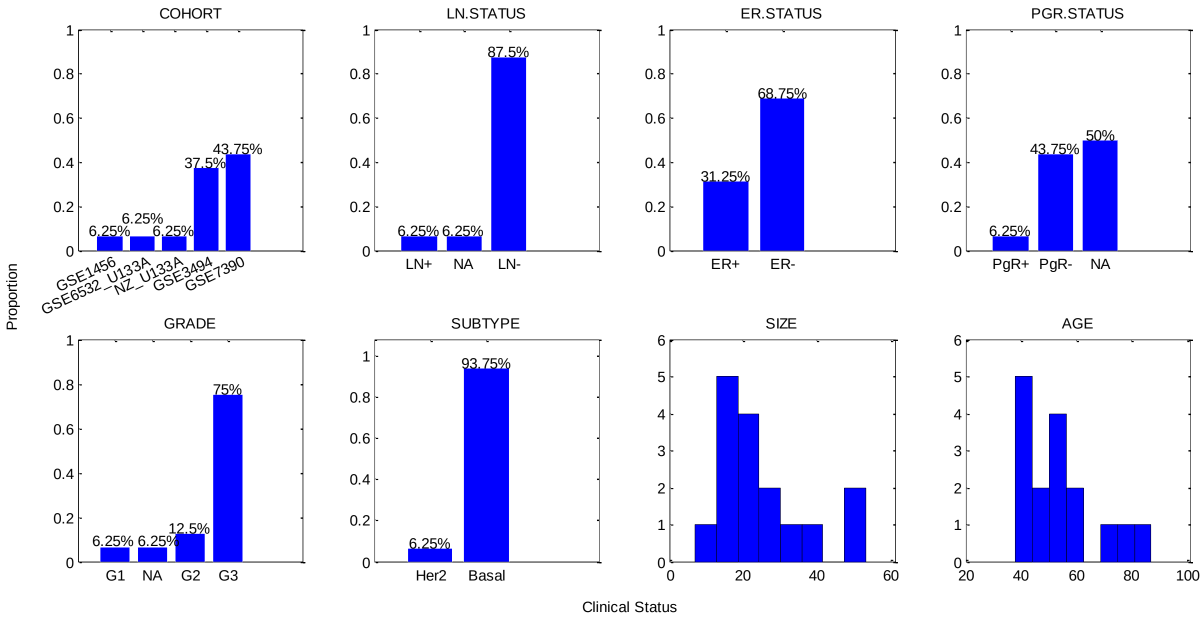



Bicluster30 (16chips) Clinical Information

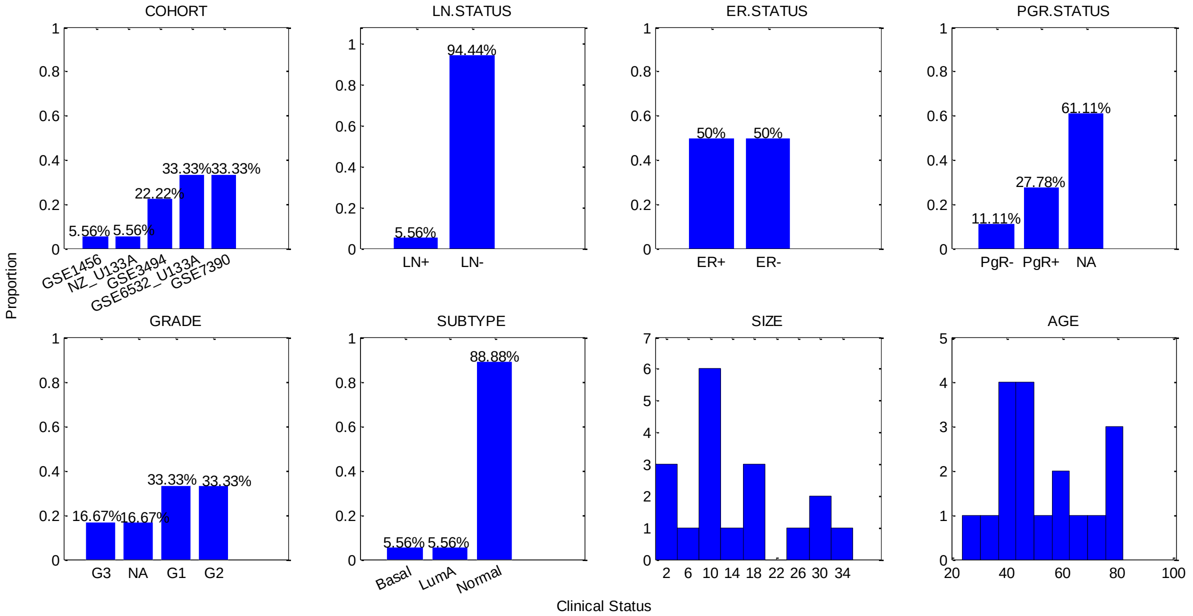

Bicluster31 (14chips) Clinical Information


Bicluster32 (19chips) Clinical Information

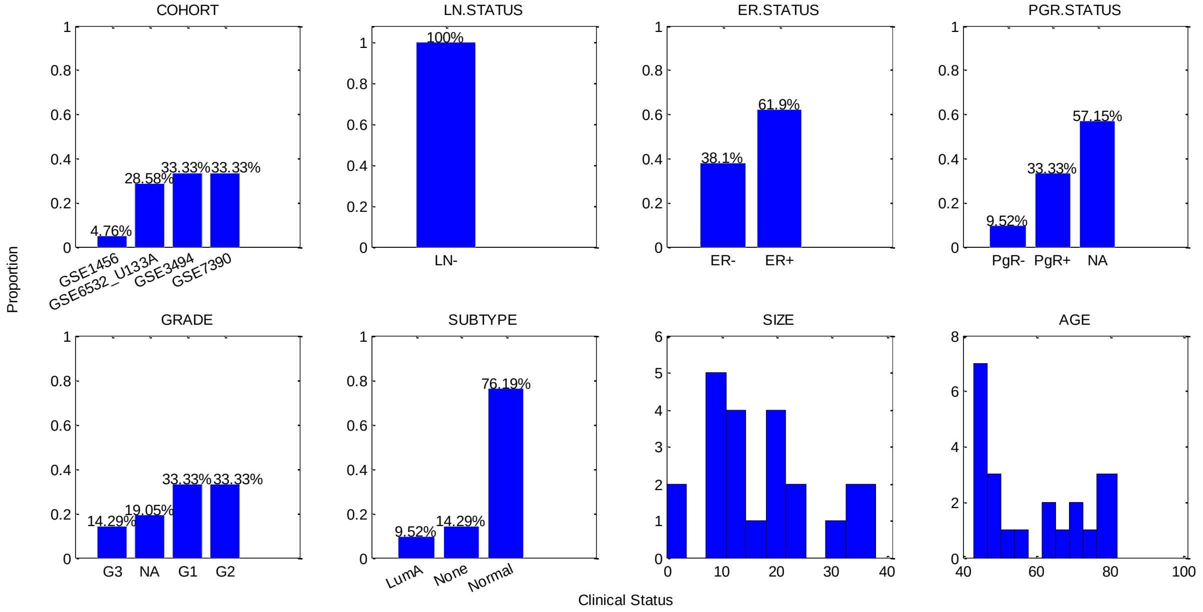
Bicluster33 (16chips) Clinical Information

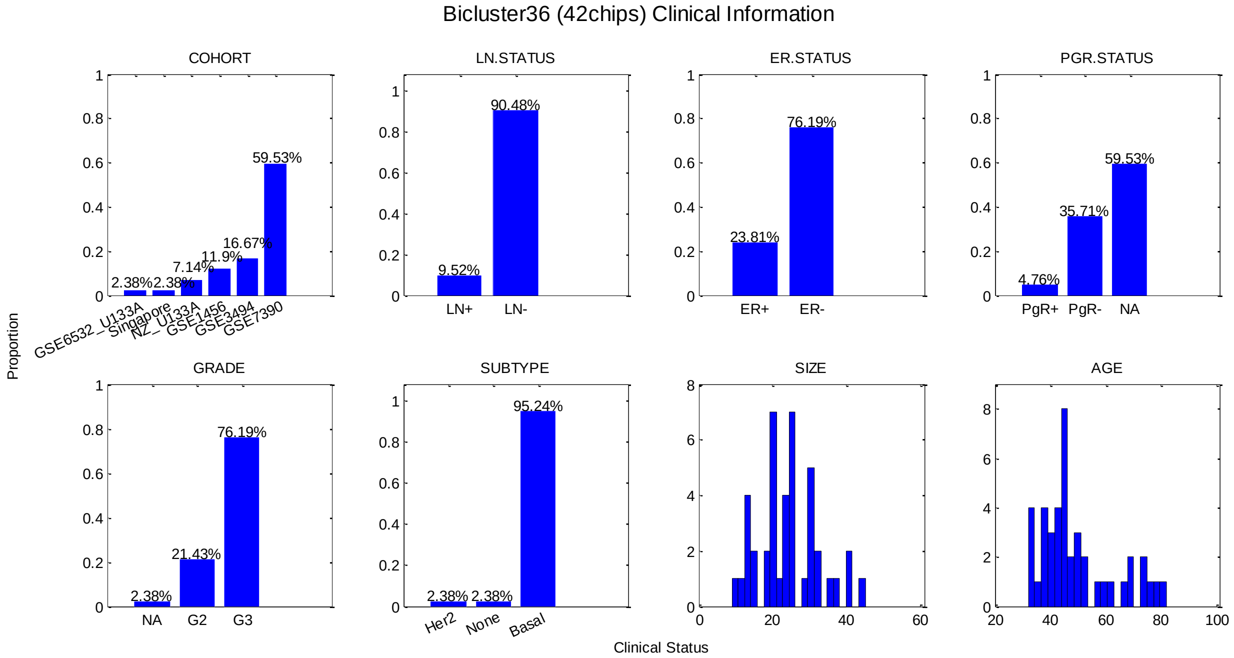


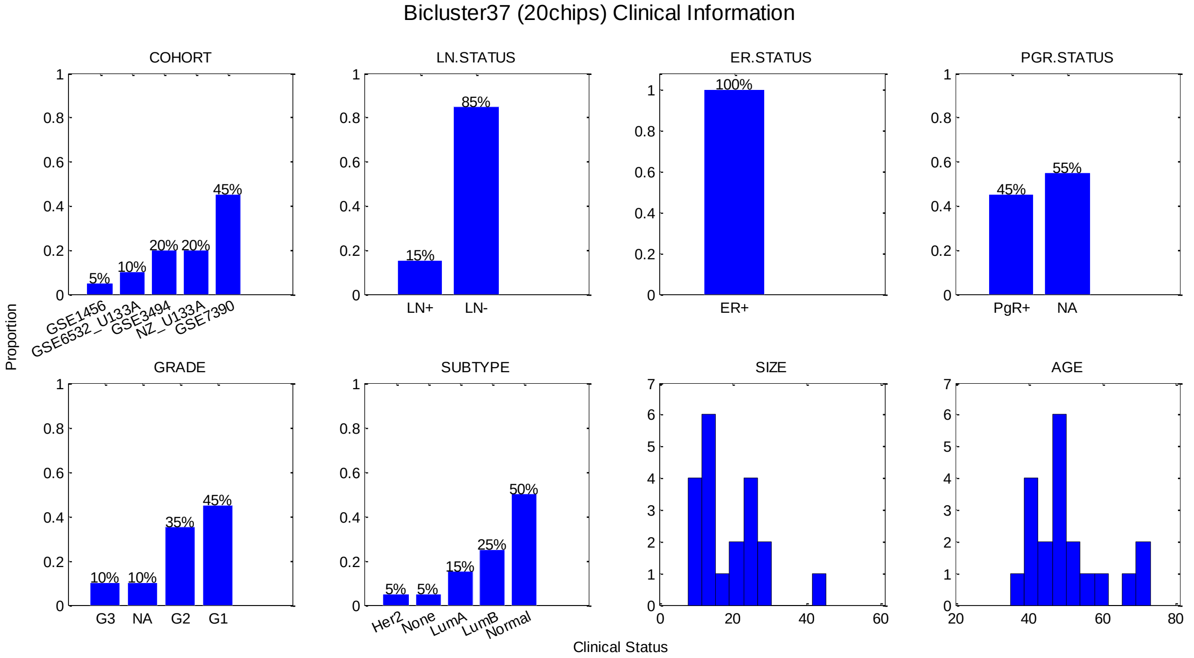
Bicluster34 (18chips) Clinical Information

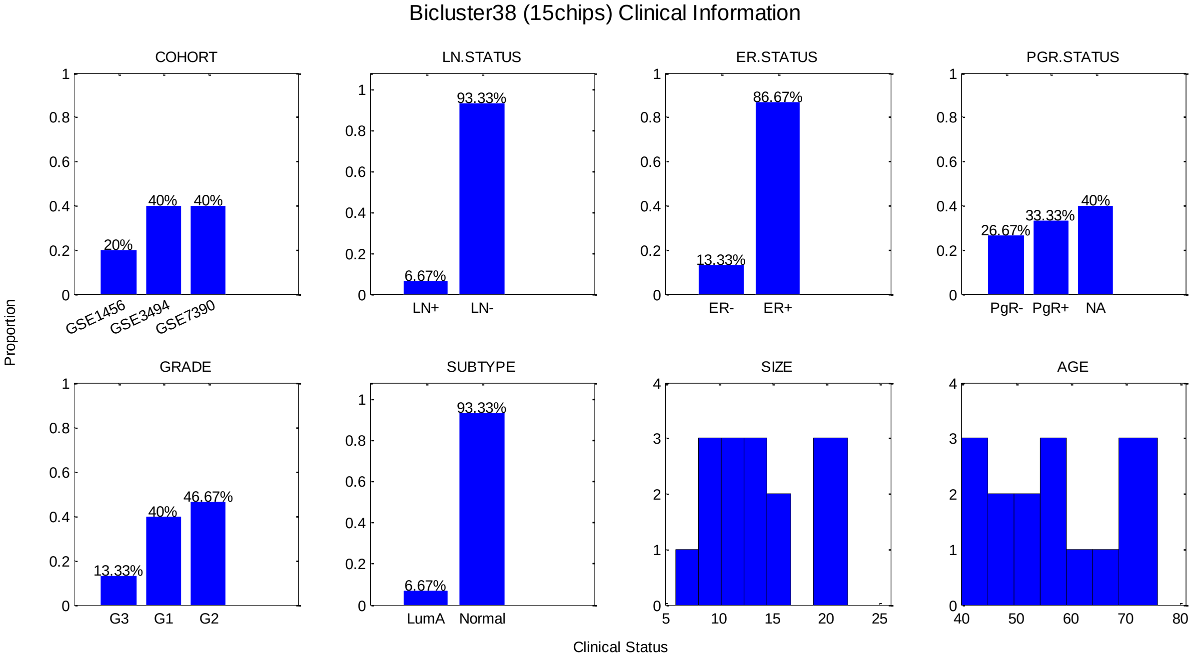


Bicluster35 (21chips) Clinical Information

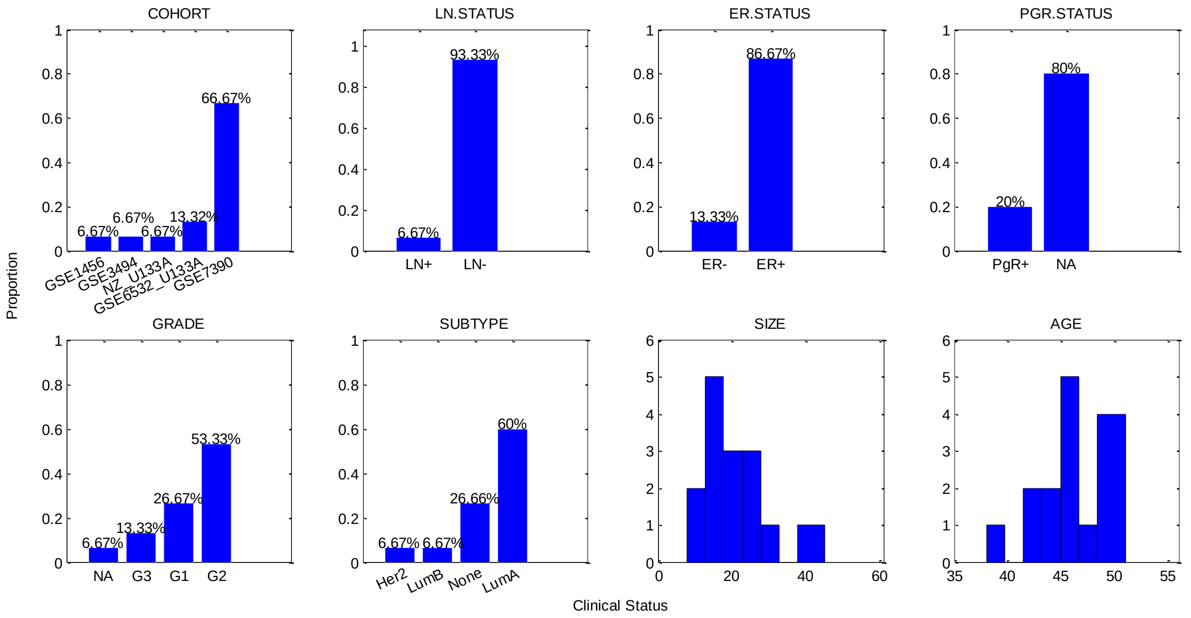



Bicluster36 (42chips) Clinical Information
COHORT
GSE6532_U133A
Singapore
NZ_U133A
GSE1456
GSE3494
GSE7390
2.38%
2.38%
7.14%
11.9%
16.67%
59.53%
LN.STATUS
LN+
LN-
9.52%
90.48%
ER.STATUS
ER+
ER-
23.81%
76.19%
PGR.STATUS
PgR+
PgR-
NA
4.76%
35.71%
59.53%
GRADE
NA
G2
G3
2.38%
21.43%
76.19%
SUBTYPE
Her2
None
Basal
2.38%
2.38%
95.24%
SIZE
AGE
Proportion
Clinical Status

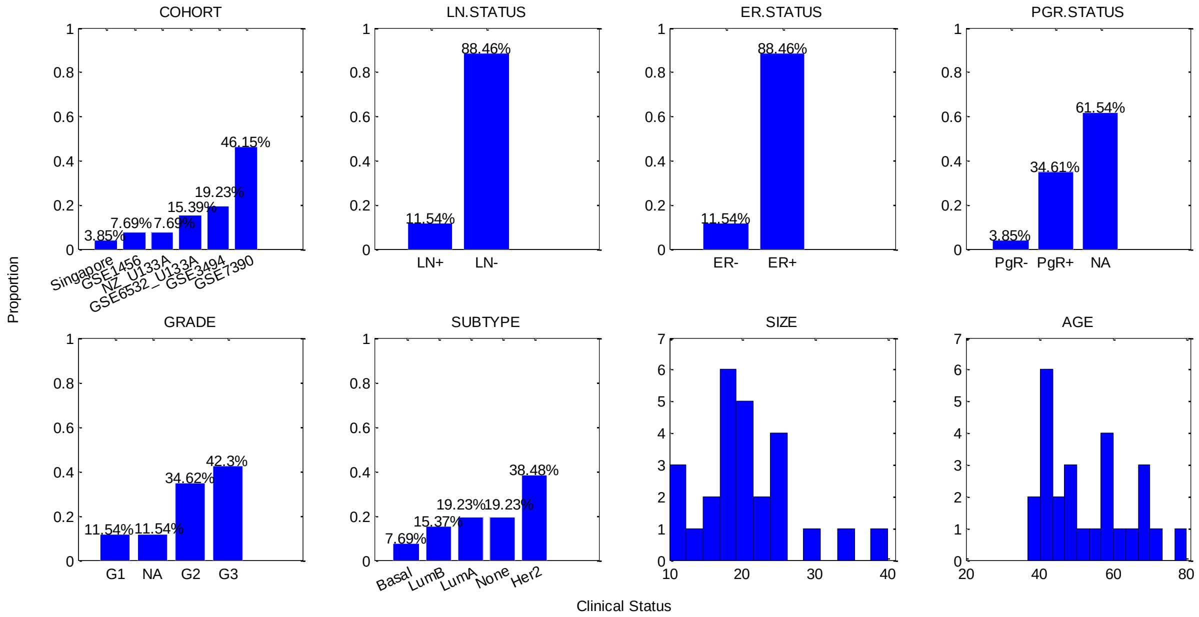

Bicluster37 (20chips) Clinical Information
COHORT
GSE1456
GSE6532_U133A
GSE3494
NZ_U133A
GSE7390
5%
10%
20%
20%
45%
LN.STATUS
LN+
LN-
15%
85%
ER.STATUS
ER+
100%
PGR.STATUS
PgR+
NA
45%
55%
GRADE
G3
NA
G2
G1
10%
10%
35%
45%
SUBTYPE
Her2
None
LumA
LumB
Normal
5%
5%
15%
25%
50%
SIZE
AGE
Proportion
Clinical Status


Bicluster38 (15chips) Clinical Information
COHORT
GSE1456
GSE3494
GSE7390
20%
40%
40%
LN.STATUS
LN+
LN-
6.67%
93.33%
ER.STATUS
ER-
ER+
13.33%
86.67%
PGR.STATUS
PgR-
PgR+
NA
26.67%
33.33%
40%
GRADE
G3
G1
G2
13.33%
40%
46.67%
SUBTYPE
LumA
Normal
6.67%
93.33%
SIZE
AGE
Proportion
Clinical Status

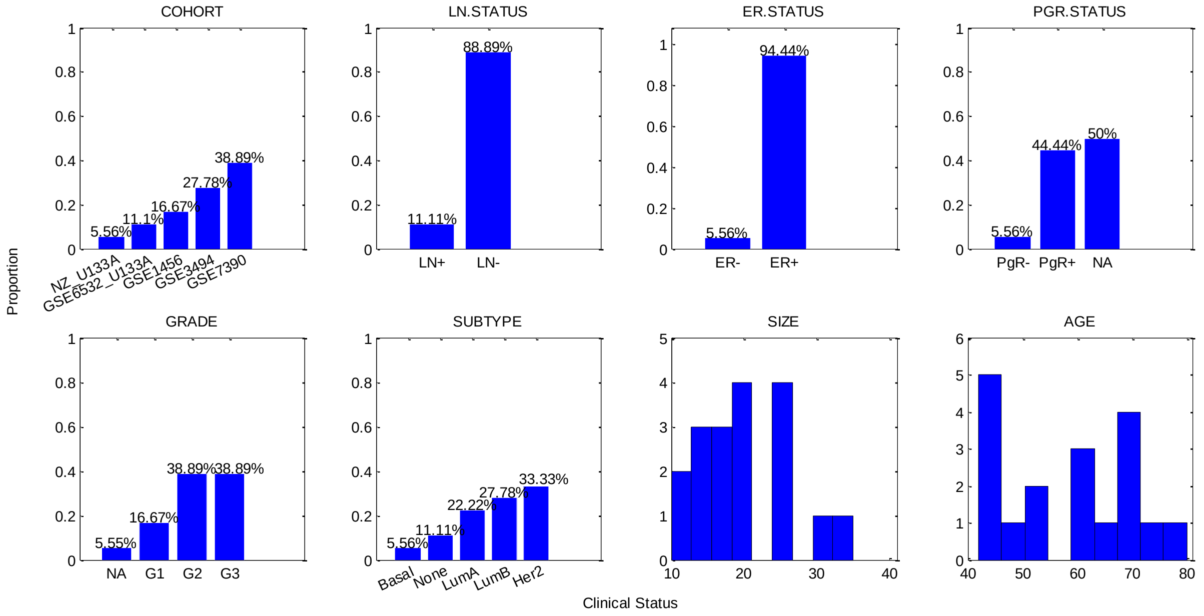
Bicluster39 (15chips) Clinical Information



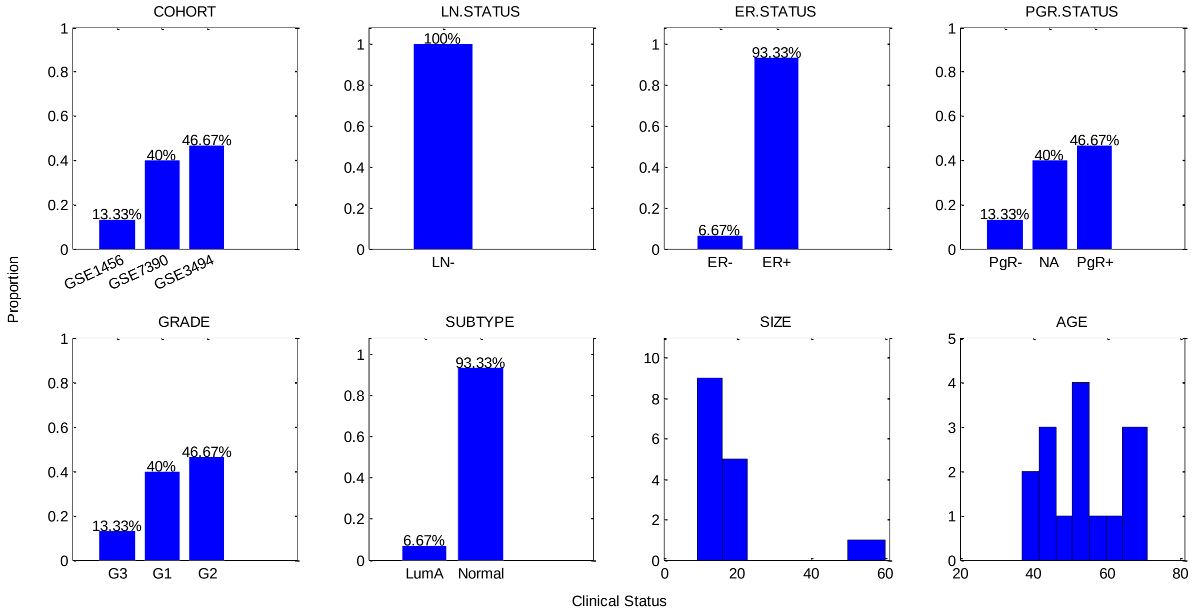
Bicluster40 (26chips) Clinical Information



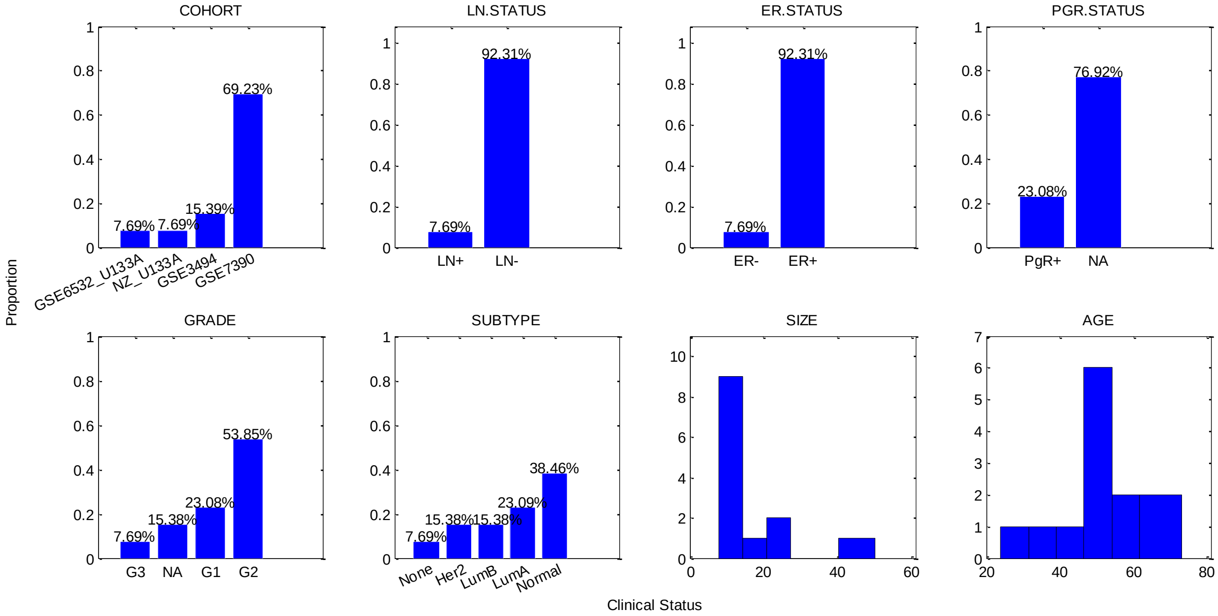
Bicluster41 (18chips) Clinical Information



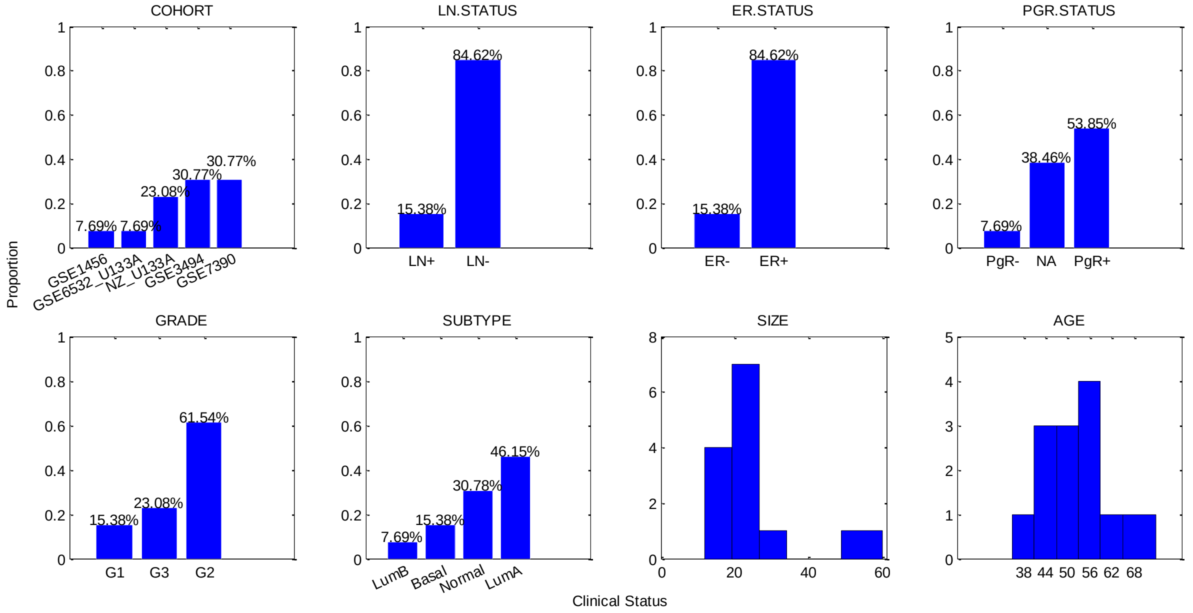
Bicluster42 (15chips) Clinical Information



Bicluster43 (13chips) Clinical Information

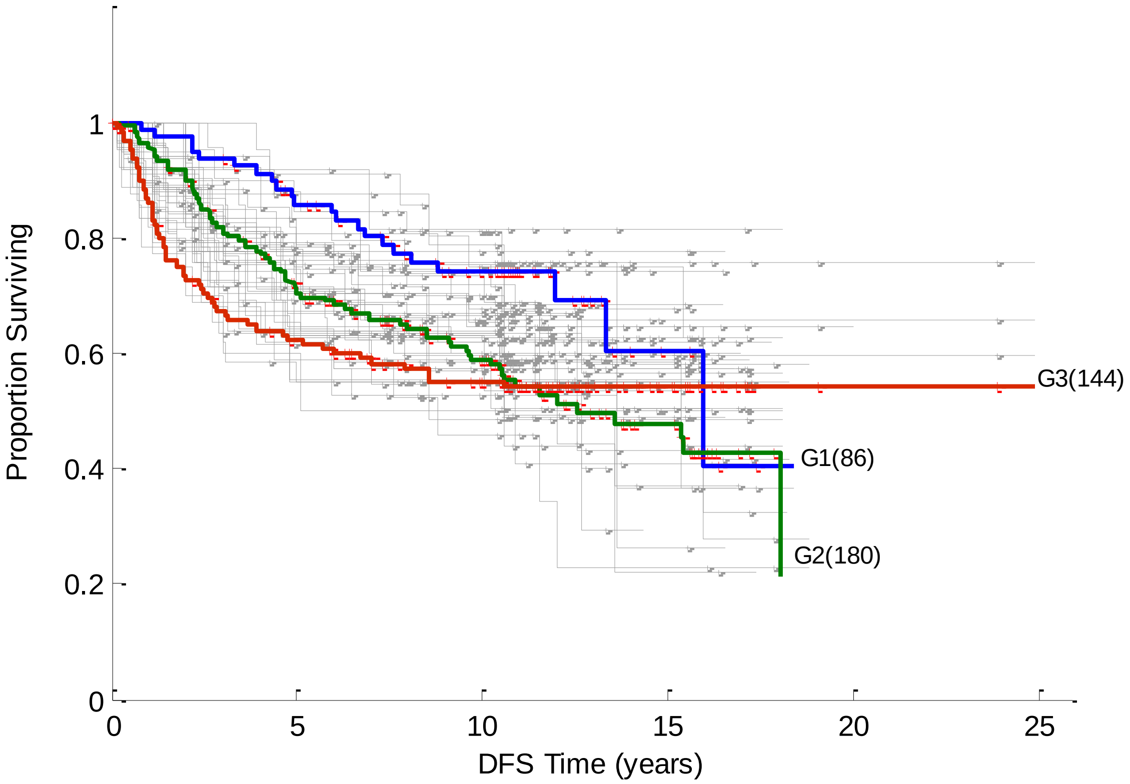


Bicluster44 (13chips) Clinical Information

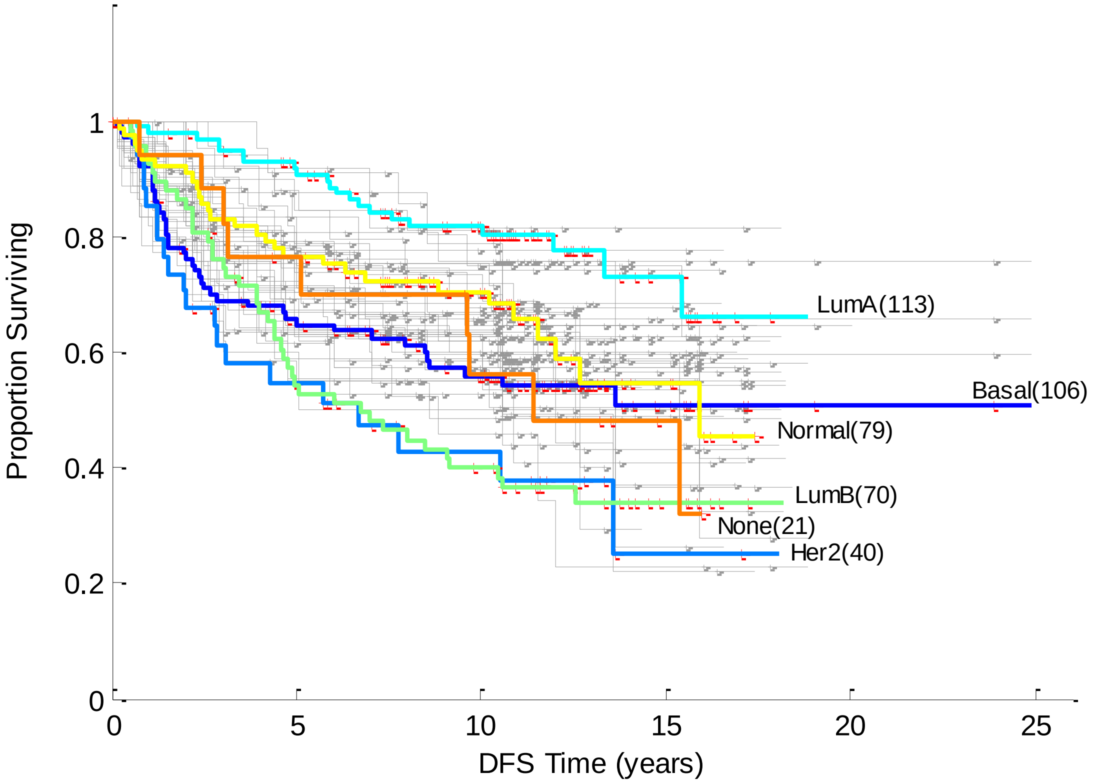


Supplementary Figure 2. Kaplan-Meier estimates of survival distribution of randomly selected patient groups. Kaplan-Meier estimate of disease free survival (DFS) for 44 groups with arbitrary patient membership (all patients were randomly allocated into 44 groups), superimposed on survival curve of (A) histological grades and (B) tumour subtypes, respectively. Distribution of bicluster-associated survival curves in Figure 4 is much broader than the distribution of patient group-associated survival curves by chance.

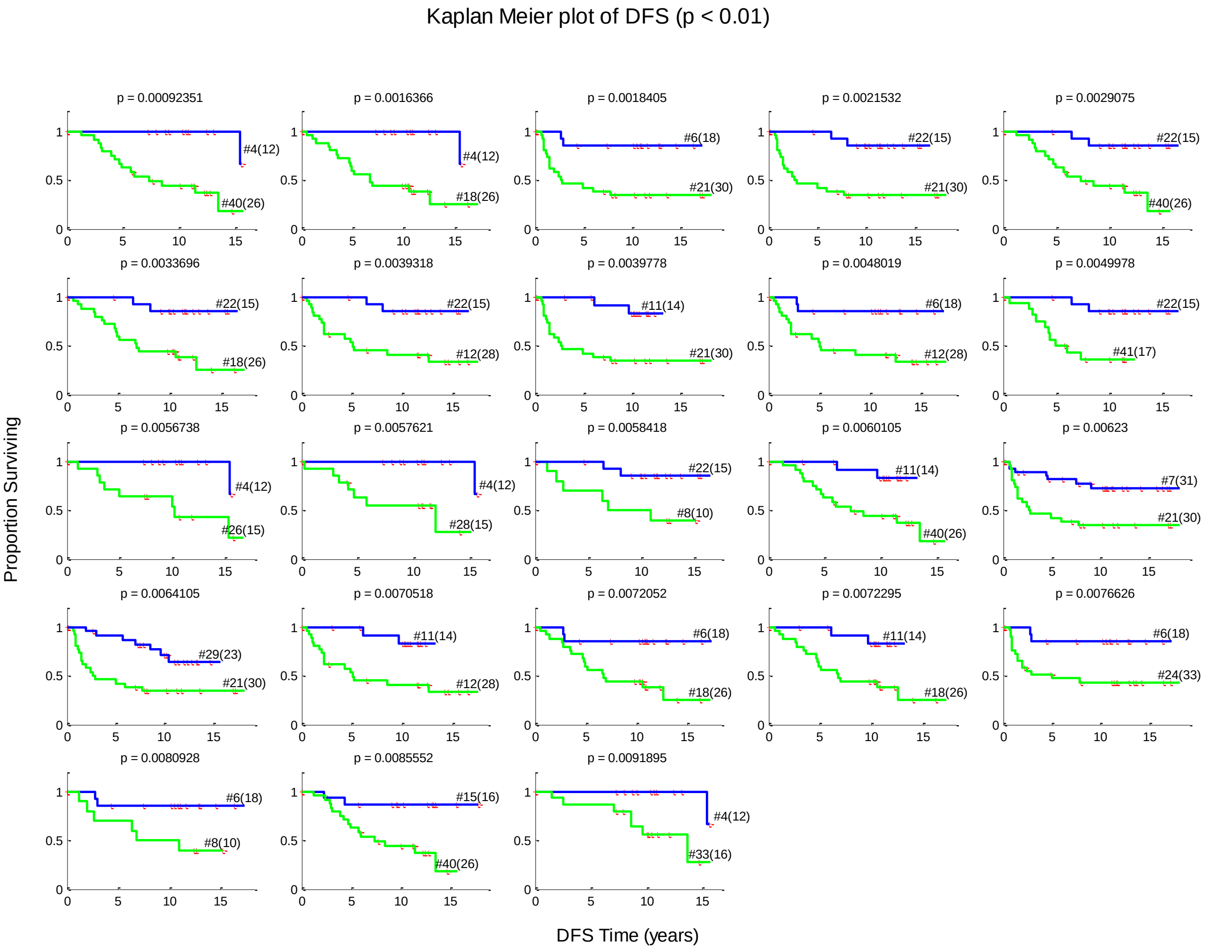
(A)



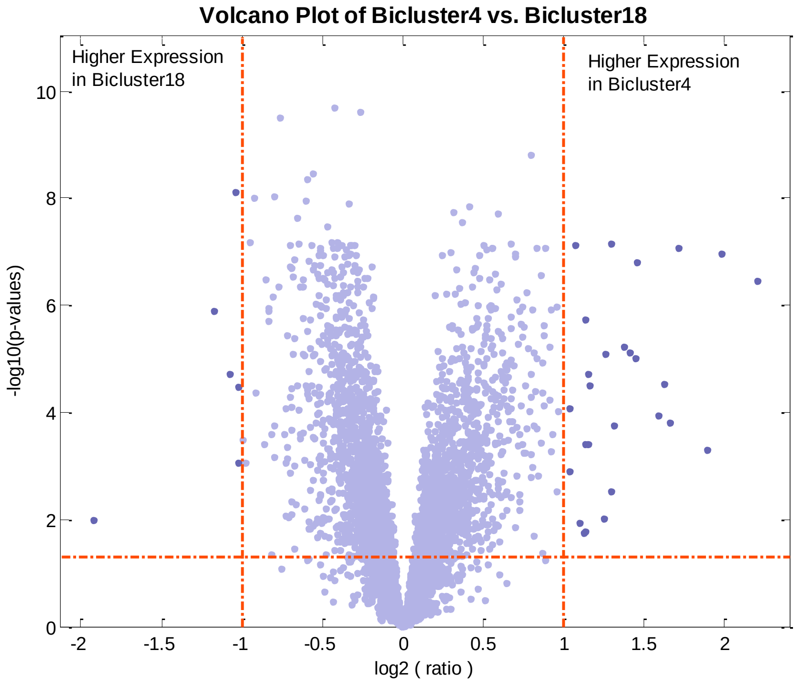
(B)

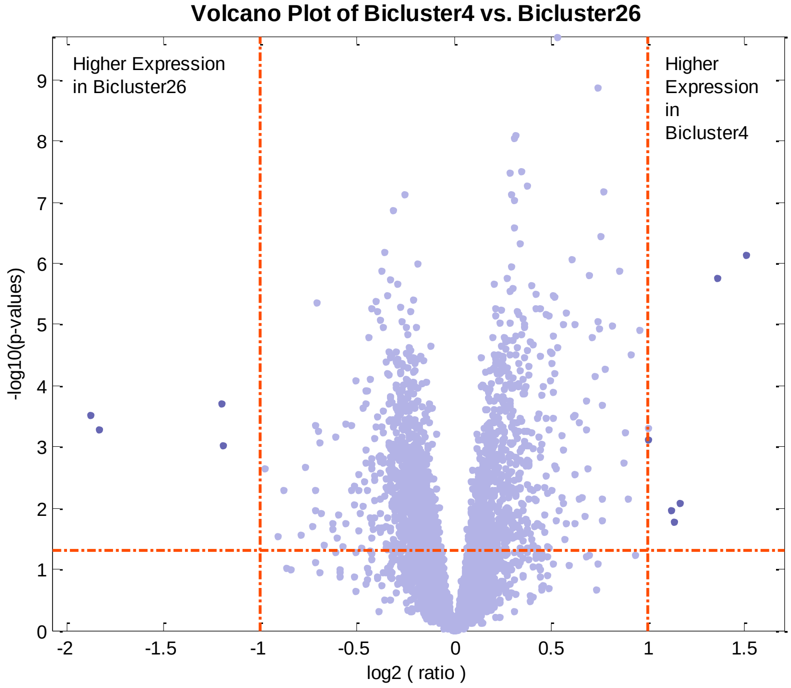


Supplementary Figure 3. Kaplan-Meier plots of pairs of biclusters showing differences in survival. Kaplan-Meier plots estimate disease free survival (DFS) distribution of pairwise biclusters comparison with significant difference in their survival distribution (logrank p value < 0.01). In each subplot, blue curve presents patients' group associated with good prognosis while curves in green represents patients' group associated with poor prognosis. Logrank p-values denotes on the top of each subplot indicates the significance of difference in survival distribution between two biclusters.

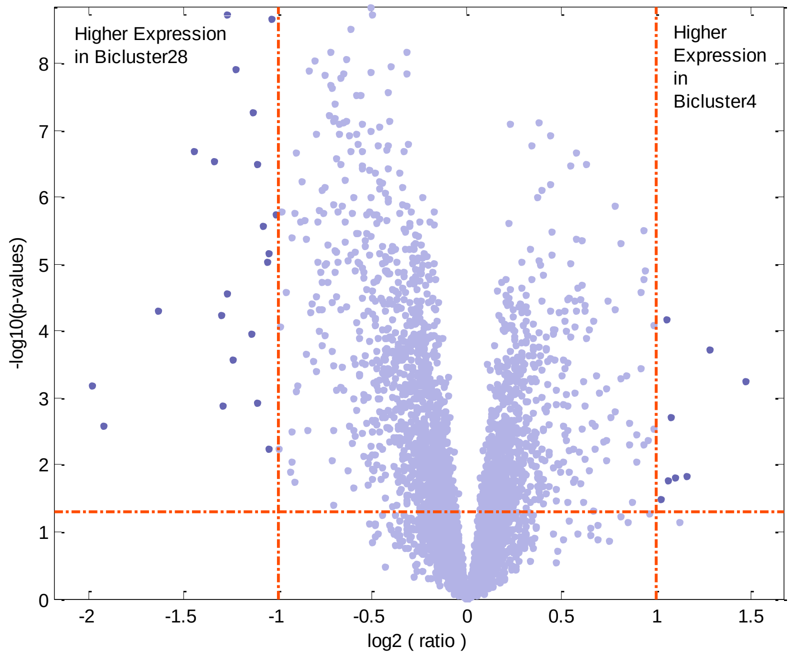


Supplementary Figure 4. Volcano plots showing genes that are differentially expressed between biclusters which have significant difference (logrank p values < 0.01) in their survival curves, associated with good and poor prognosis respectively. The statistical significance of differential gene expression was quantified by two sample t-test p-value. The fold change in expression level of each gene between two biclusters was evaluated. We identified genes whose expression level was significantly different by at least 2-fold change between biclusters associated with good and poor survival. We plot the statistical significance of differentially expressed genes against fold change in the expression levels between low and high risk biclusters to identify both statistically and biologically significant genes. By setting the cut off of significance level at 0.05 (horizontal red dashed line) and fold change at 2 (two vertical red dashed lines), the volcano plots identified genes with elevated expression levels, which represents the top-most significantly differentially expressed genes between biclusters associated with good and poor survival. The volcano plot was generated by *mavolcanoplot* in MatLab.

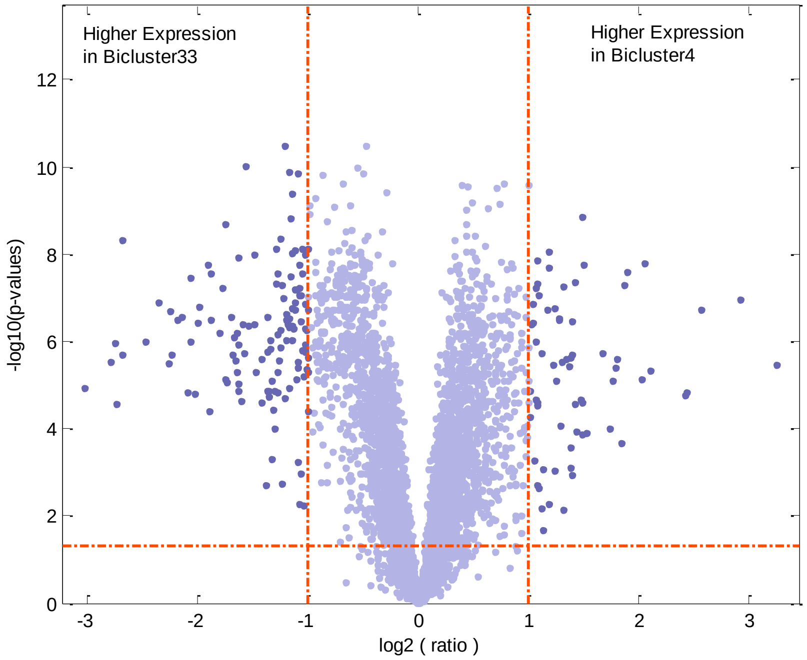




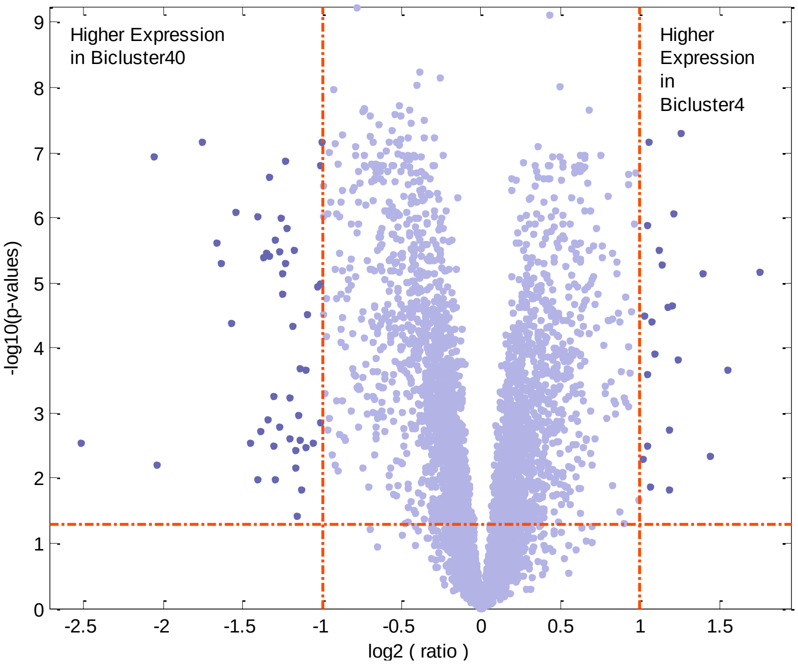
Volcano Plot of Bicluster4 vs. Bicluster28

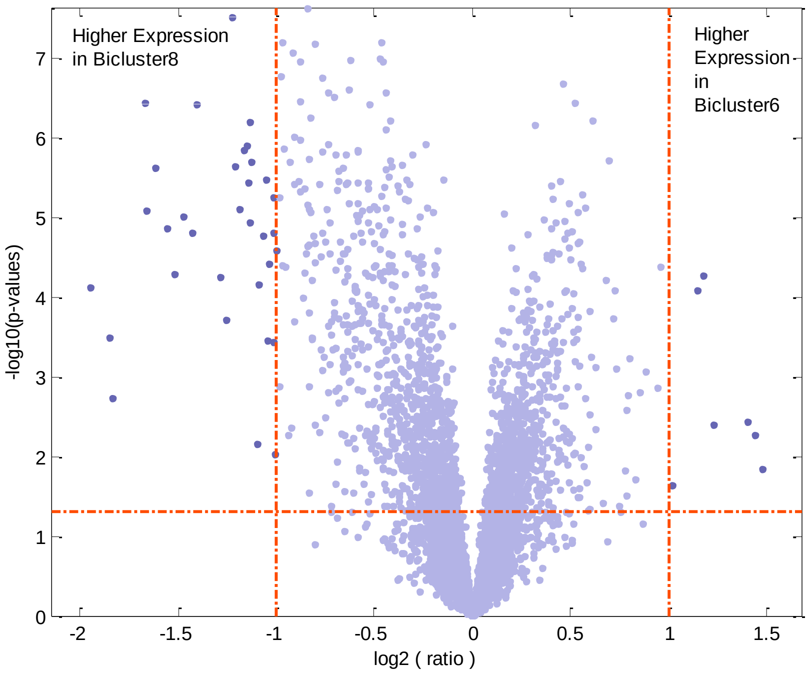


Volcano Plot of Bicluster4 vs. Bicluster33

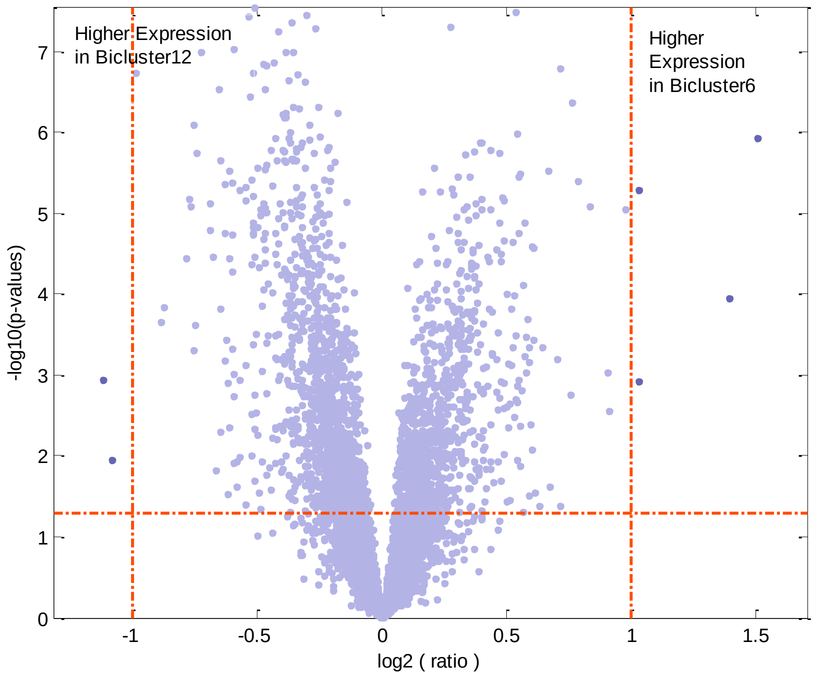


Volcano Plot of Bicluster4 vs. Bicluster40

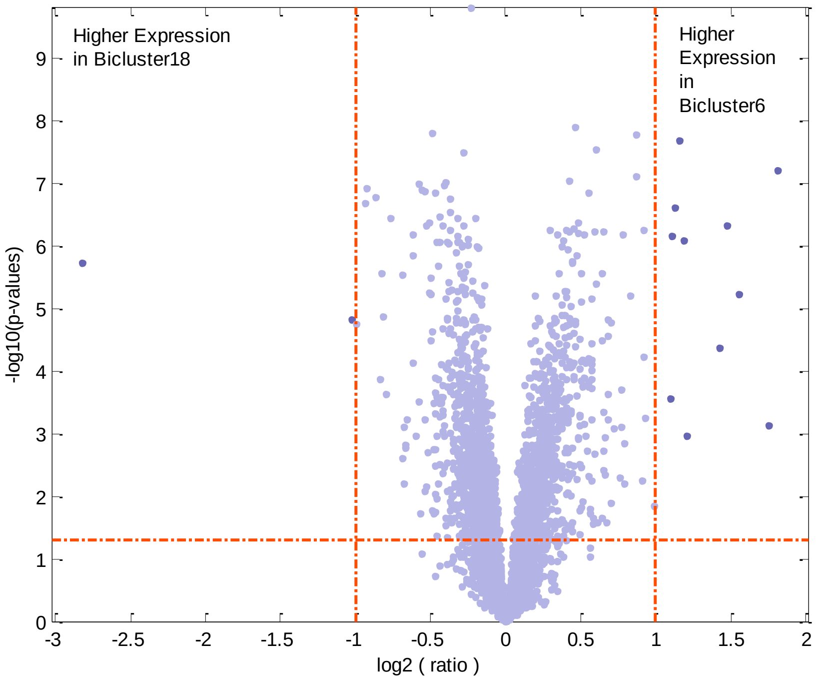




Volcano Plot of Bicluster6 vs. Bicluster8

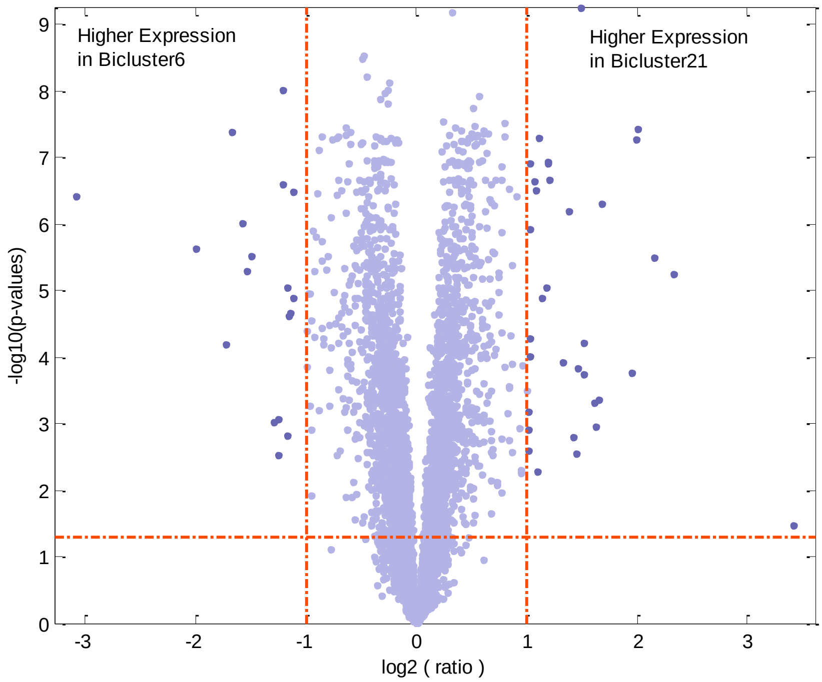


Volcano Plot of Bicluster6 vs. Bicluster12

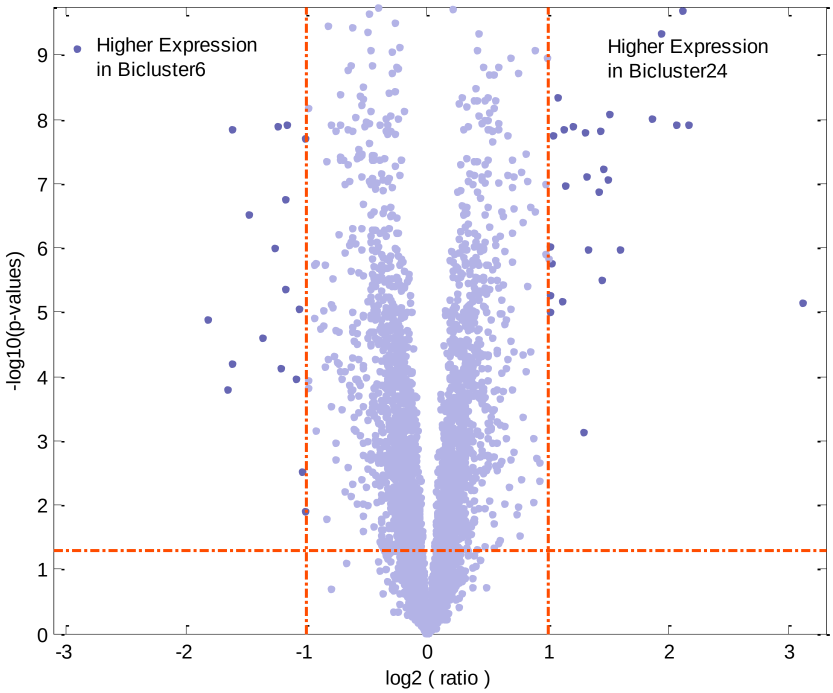


Volcano Plot of Bicluster6 vs. Bicluster18

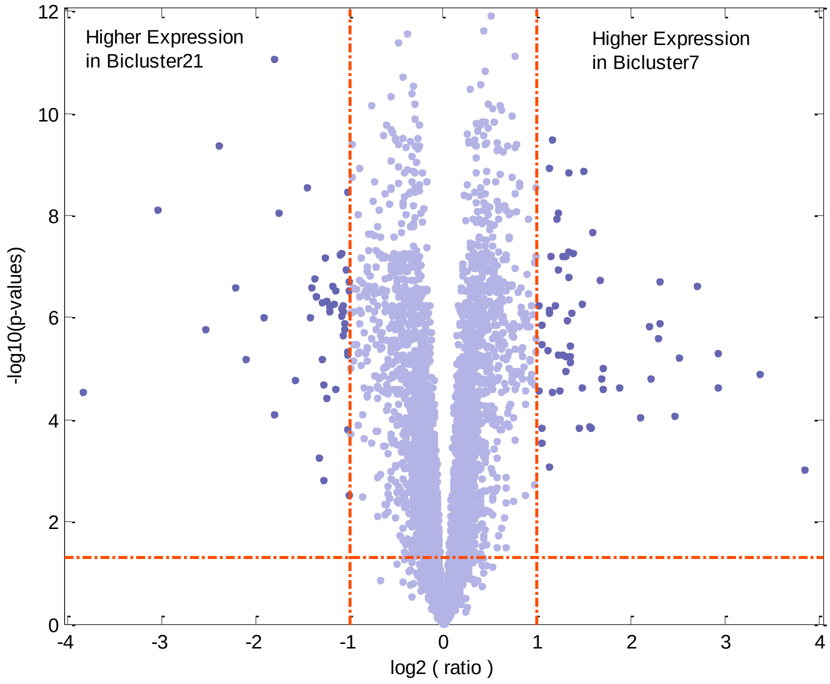
Volcano Plot of Bicluster6 vs. Bicluster21



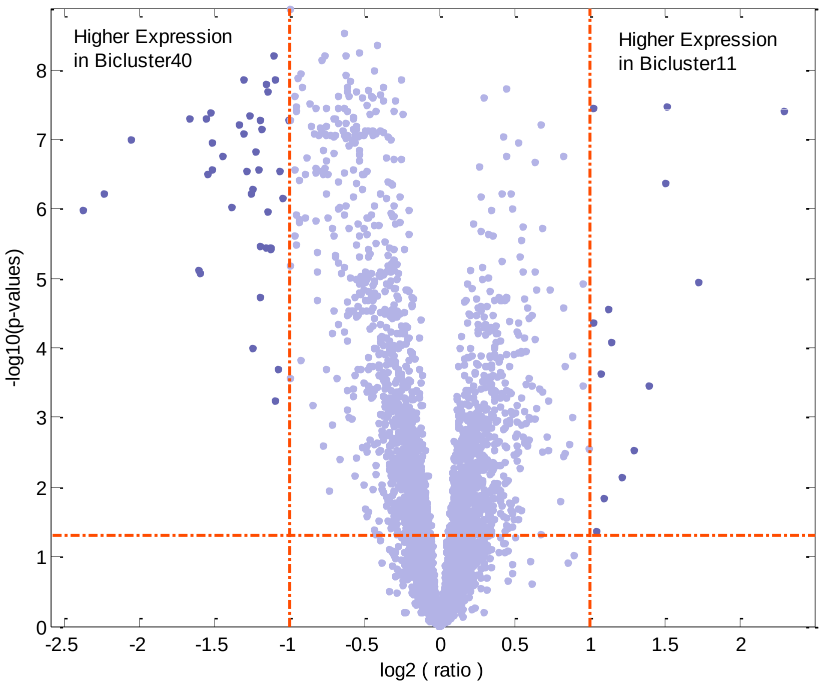
Volcano Plot of Bicluster6 vs. Bicluster24



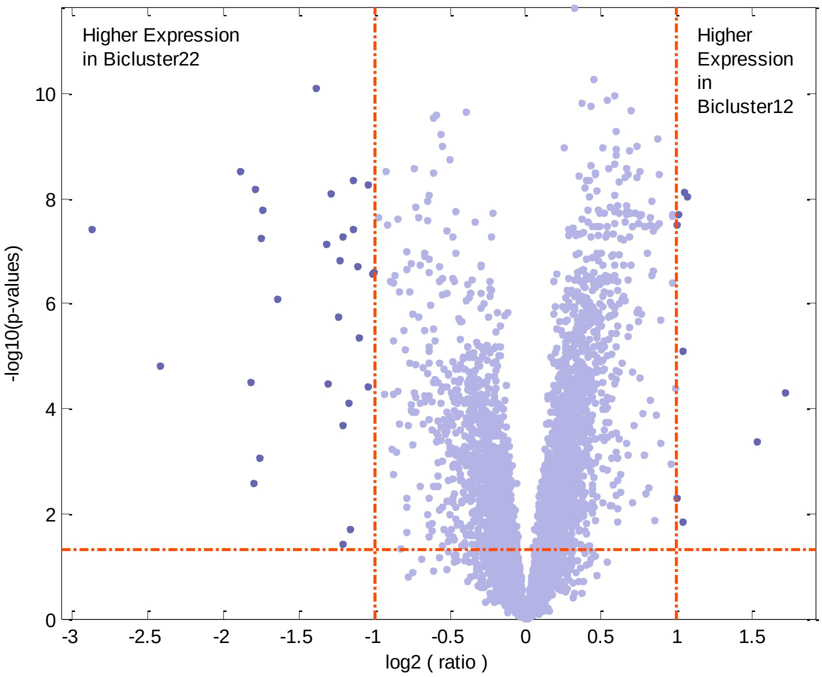
Volcano Plot of Bicluster7 vs. Bicluster21



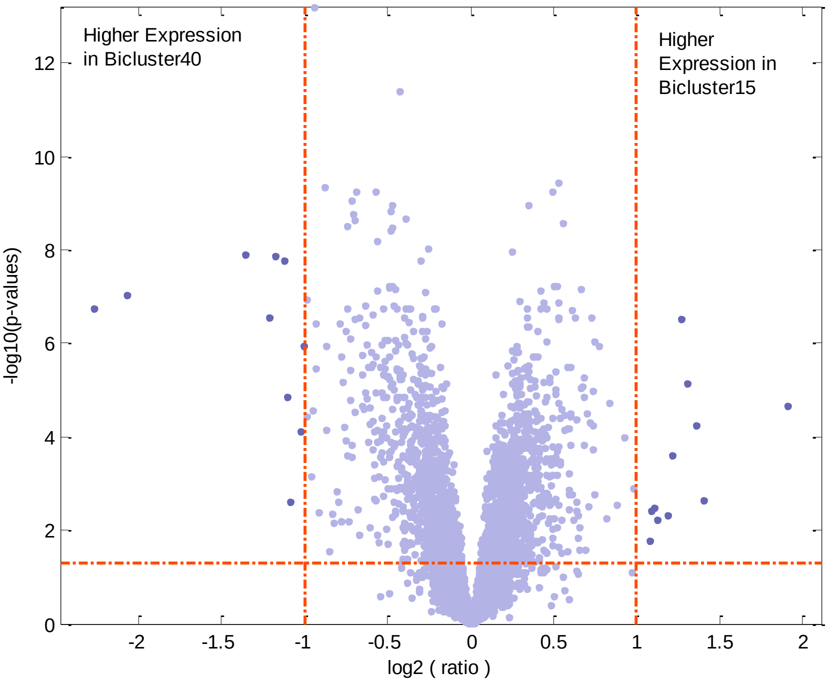
Volcano Plot of Bicluster11 vs. Bicluster40

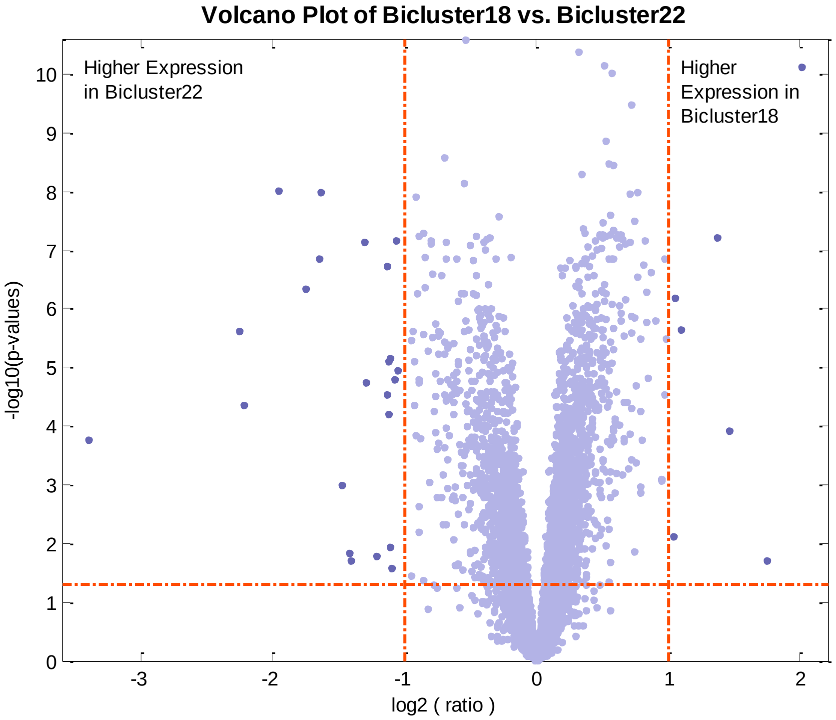


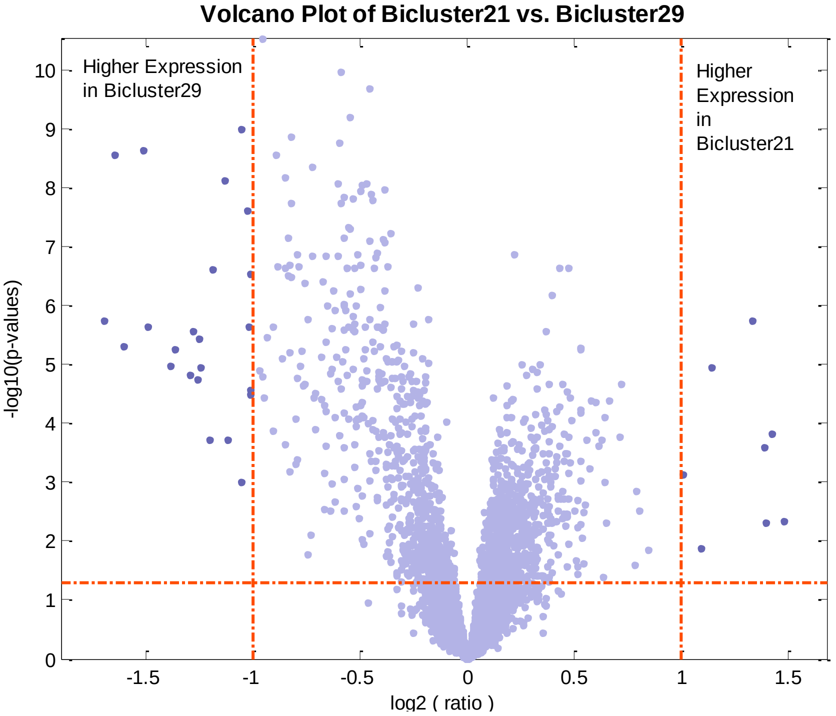
Volcano Plot of Bicluster12 vs. Bicluster22

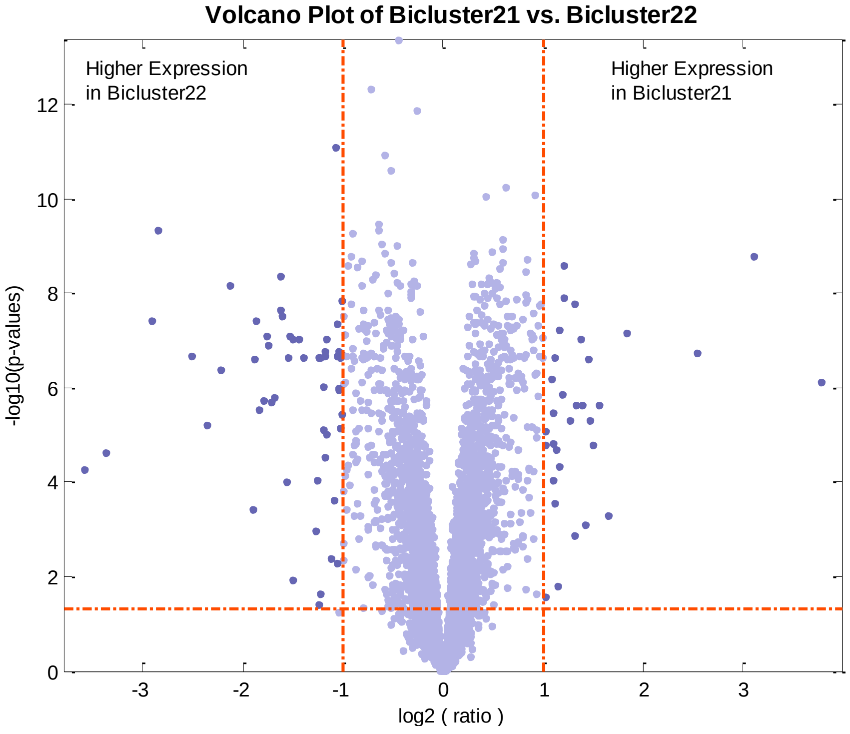


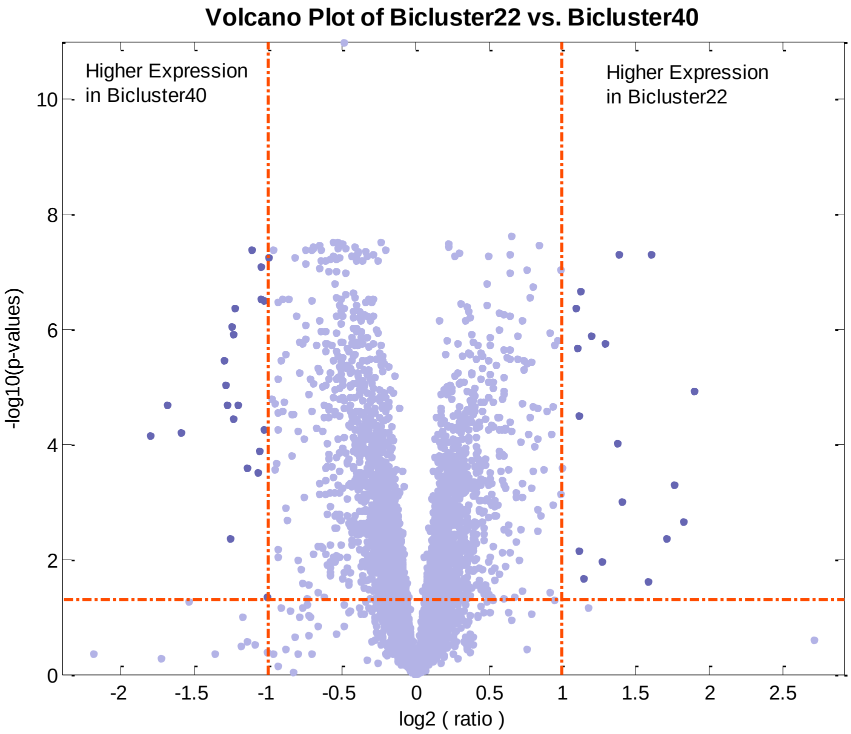
Volcano Plot of Bicluster15 vs. Bicluster40



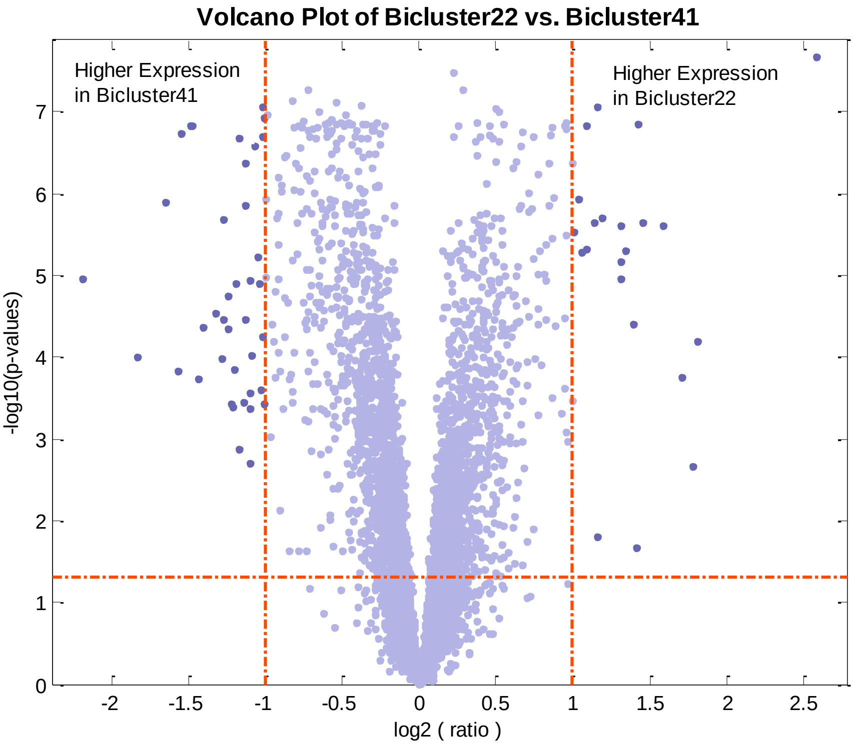

Volcano Plot of Bicluster18 vs. Bicluster22
Higher Expression in Bicluster22
Higher Expression in Bicluster18
-log10(p-values)
log2 (ratio)
0
1
2
3
4
5
6
7
8
9
10
-3
-2
-1
0
1
2


Volcano Plot of Bicluster21 vs. Bicluster29
Higher Expression in Bicluster29
Higher Expression in Bicluster21
-log10(p-values)
log2 (ratio)
0
1
2
3
4
5
6
7
8
9
10
-1.5
-1
-0.5
0
0.5
1
1.5


Volcano Plot of Bicluster21 vs. Bicluster22
Higher Expression in Bicluster22
Higher Expression in Bicluster21
-log10(p-values)
log2 (ratio)
0
2
4
6
8
10
12
-3
-2
-1
0
1
2
3



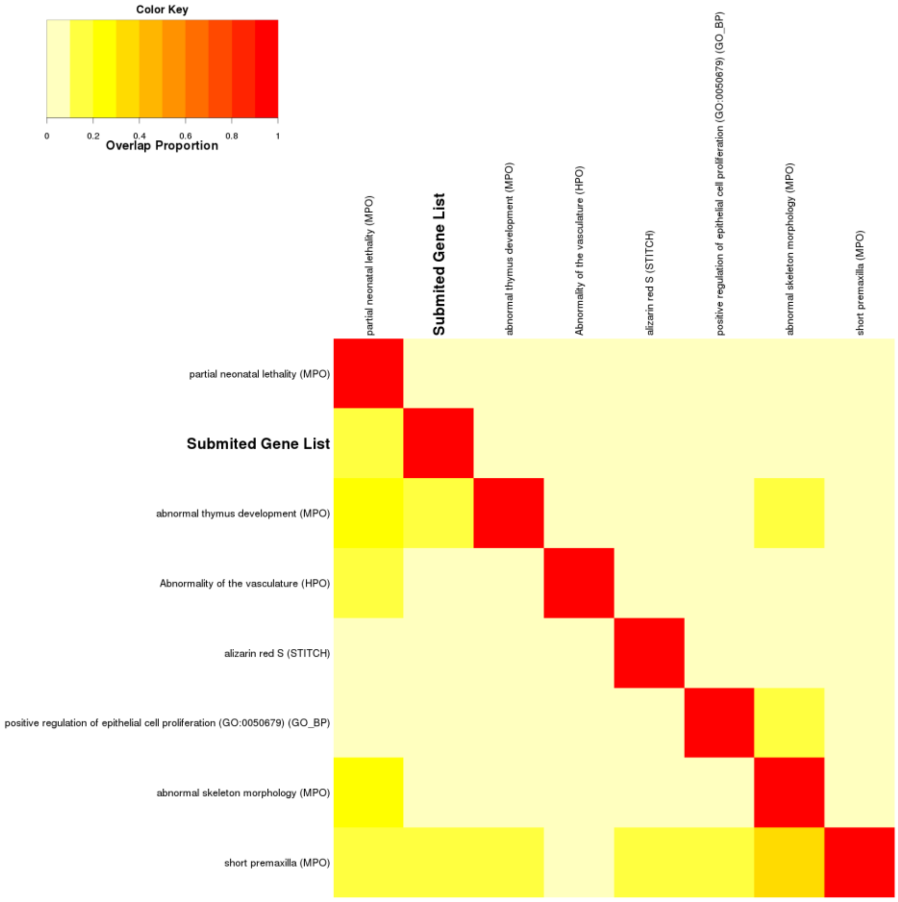
Volcano Plot of Bicluster22 vs. Bicluster40



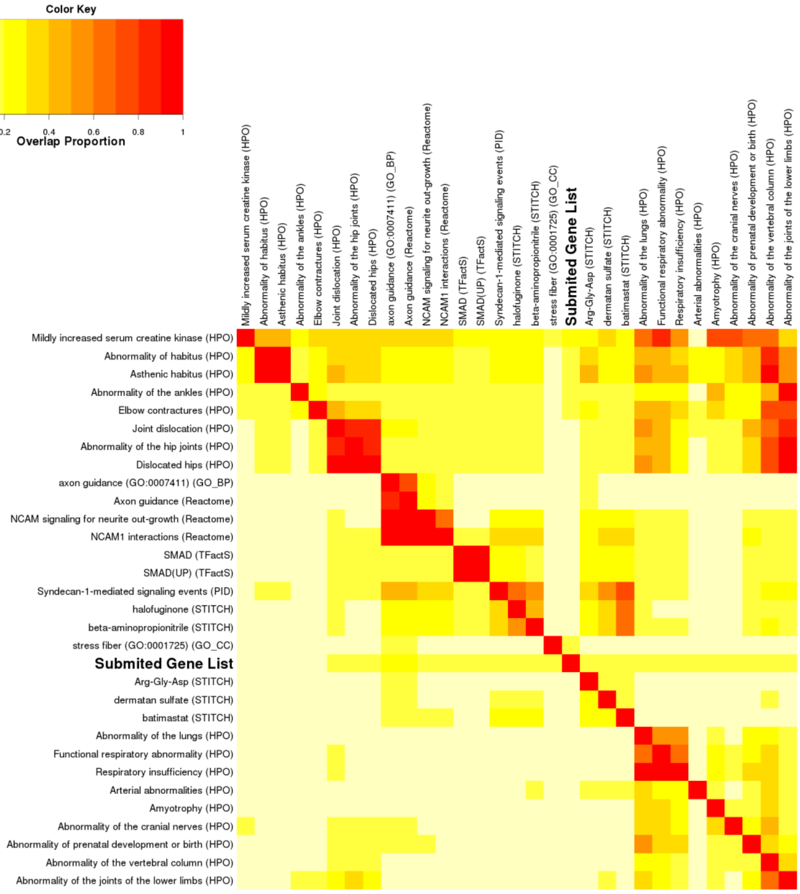
Volcano Plot of Bicluster22 vs. Bicluster41

Supplementary Figure 5. Functional enrichment heatmap for PAM gene classifiers of biclusters which have significant difference (logrank p values < 0.01) in their survival curves, associated with good and poor prognosis.

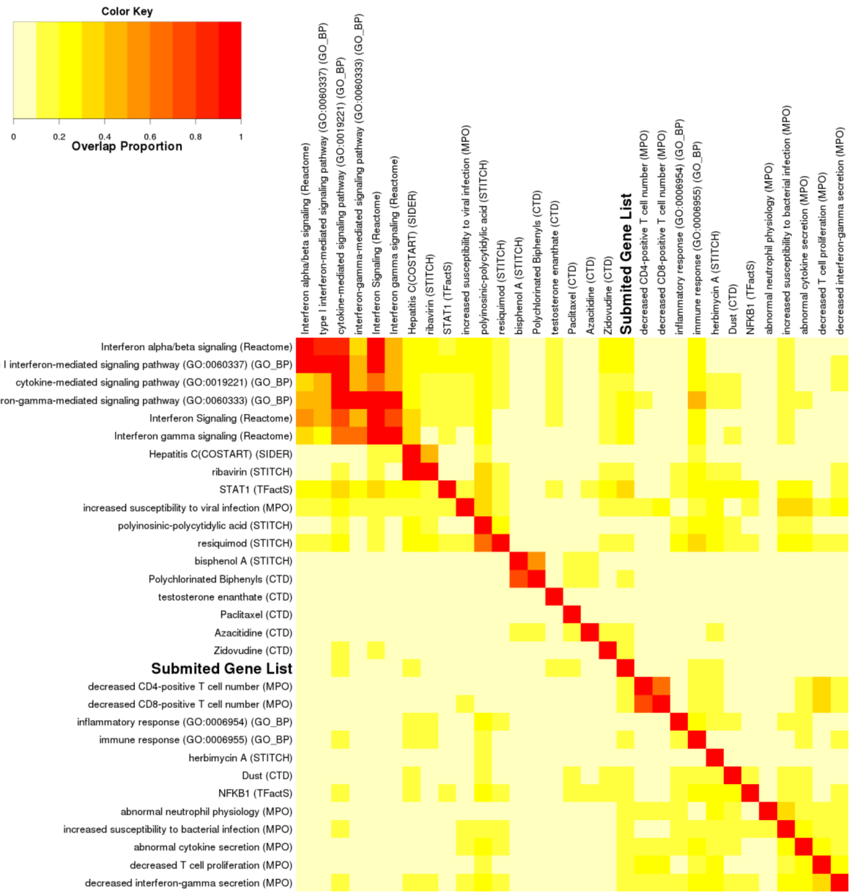
Bicluster4 vs. Bicluster18



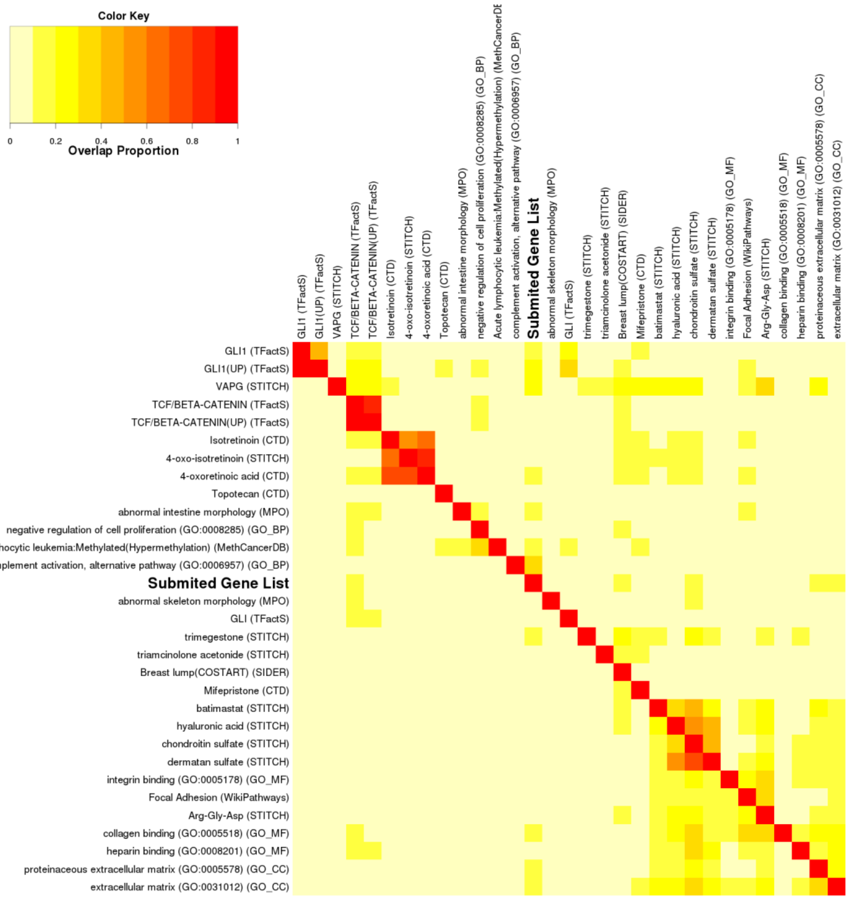
Bicluster4 vs. Bicluster26



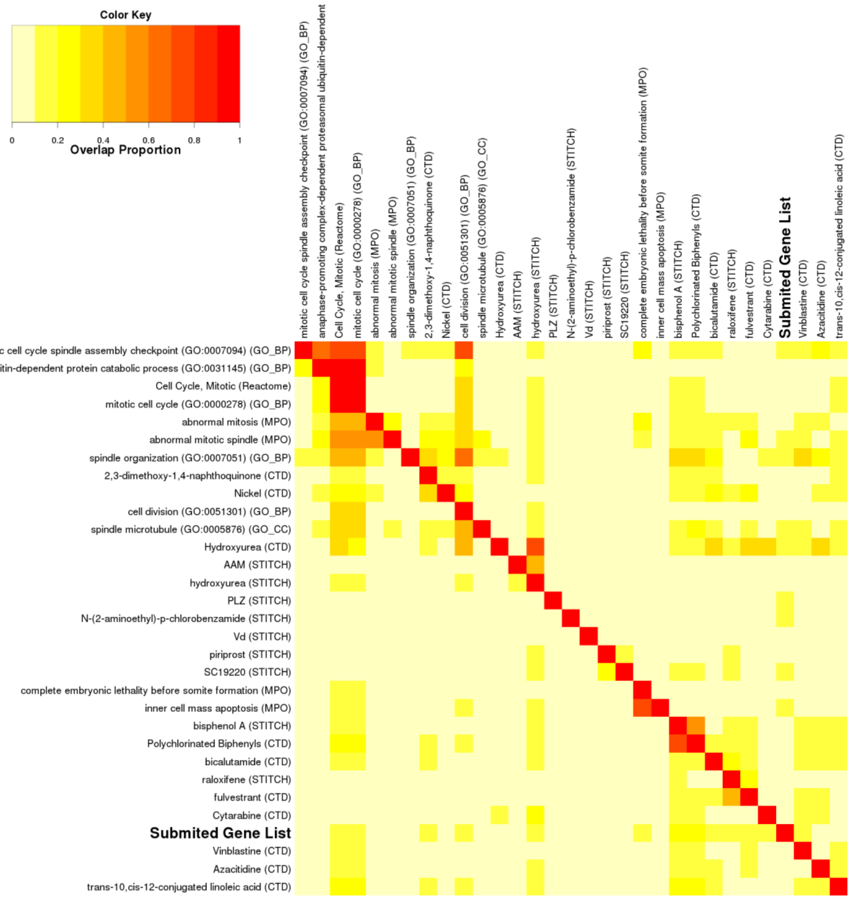
Bicluster4 vs. Bicluster33



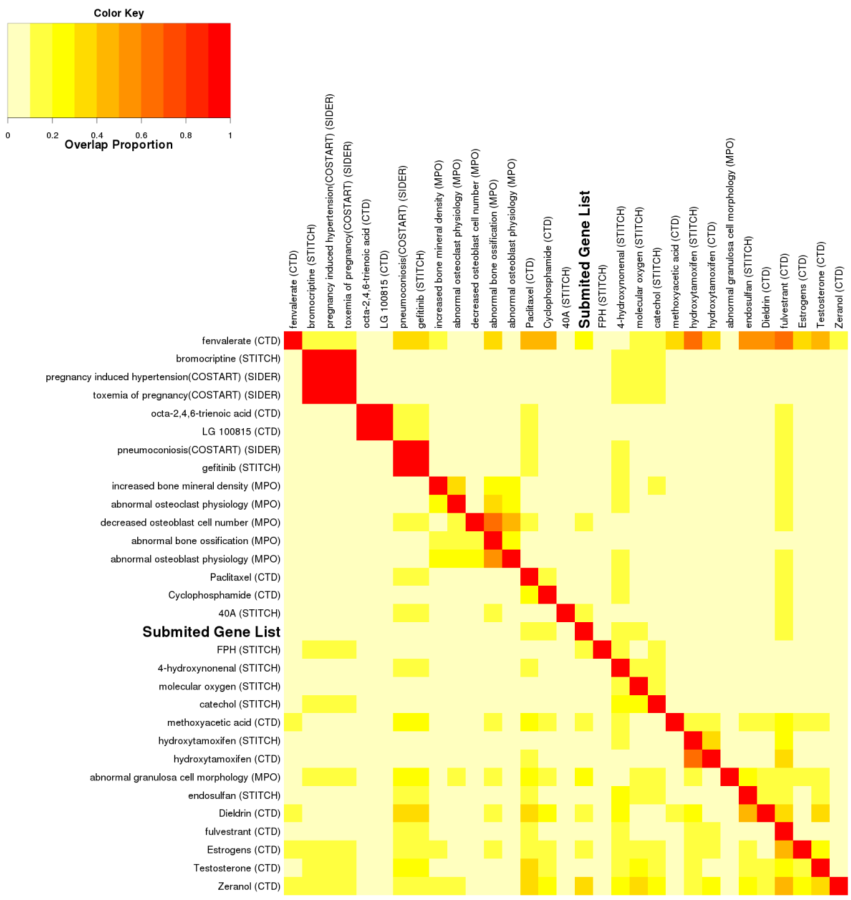
Bicluster6 vs. Bicluster8



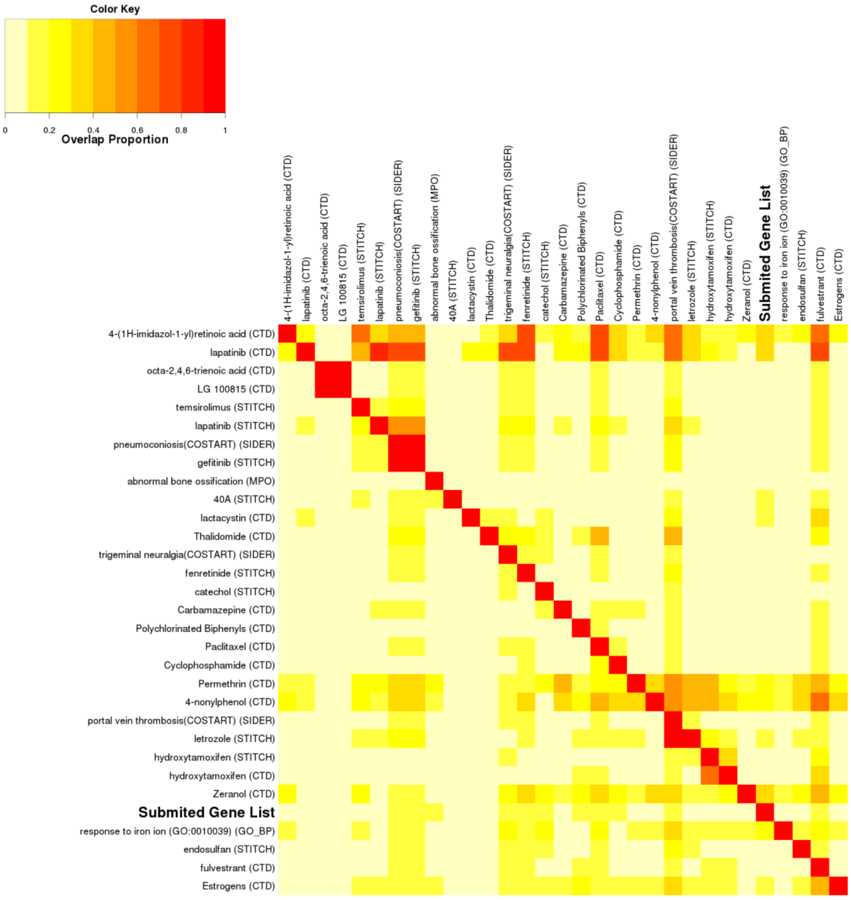
Bicluster6 vs. Bicluster12



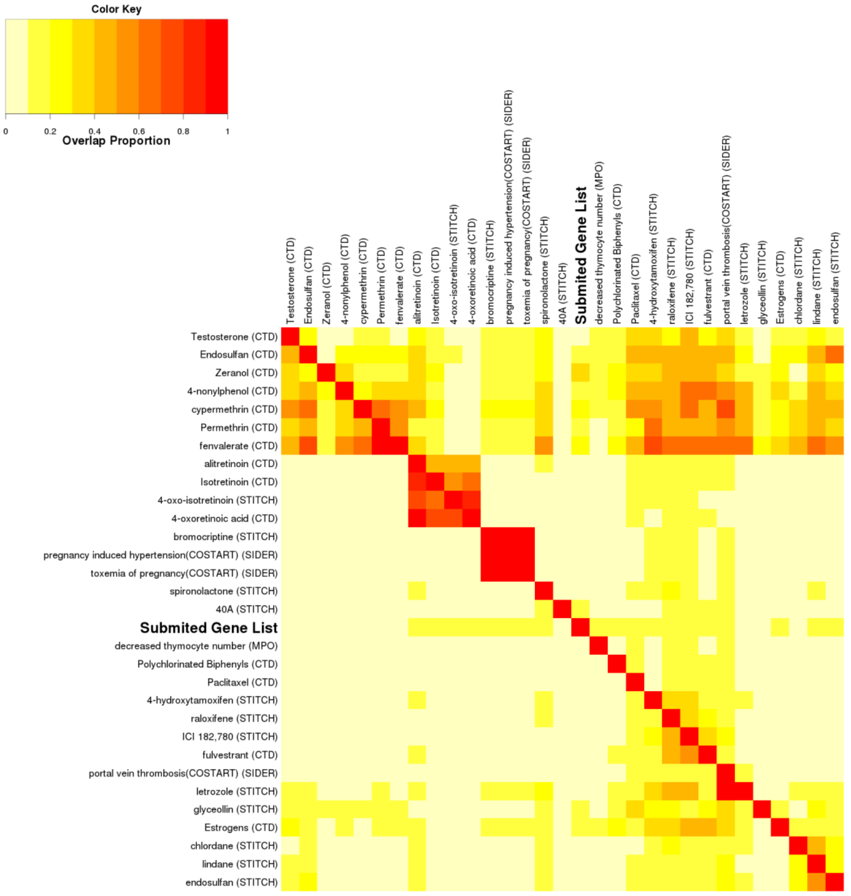
Bicluster6 vs. Bicluster21



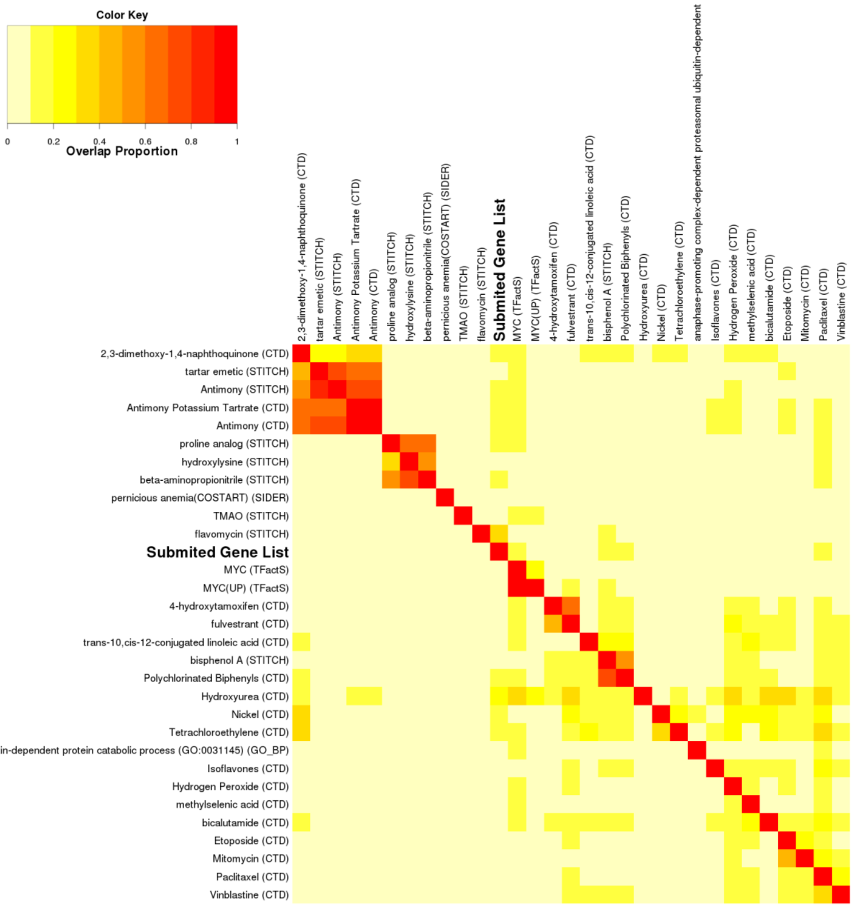
Bicluster6 vs. Bicluster24



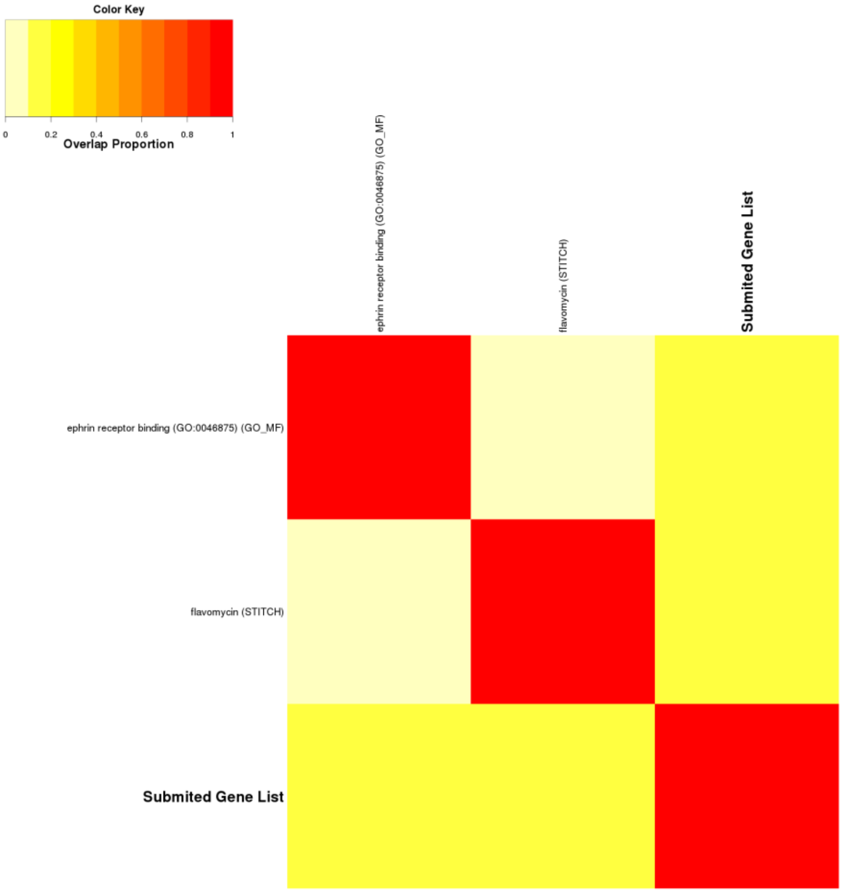
Bicluster7 vs. Bicluster21



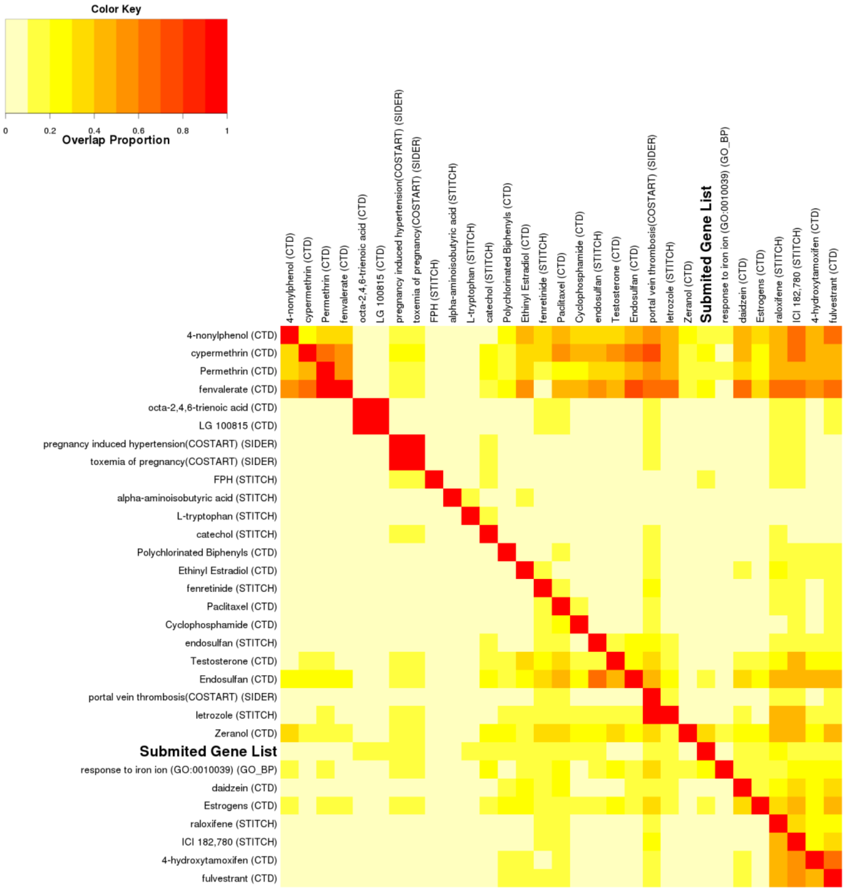
Bicluster11 vs. Bicluster12



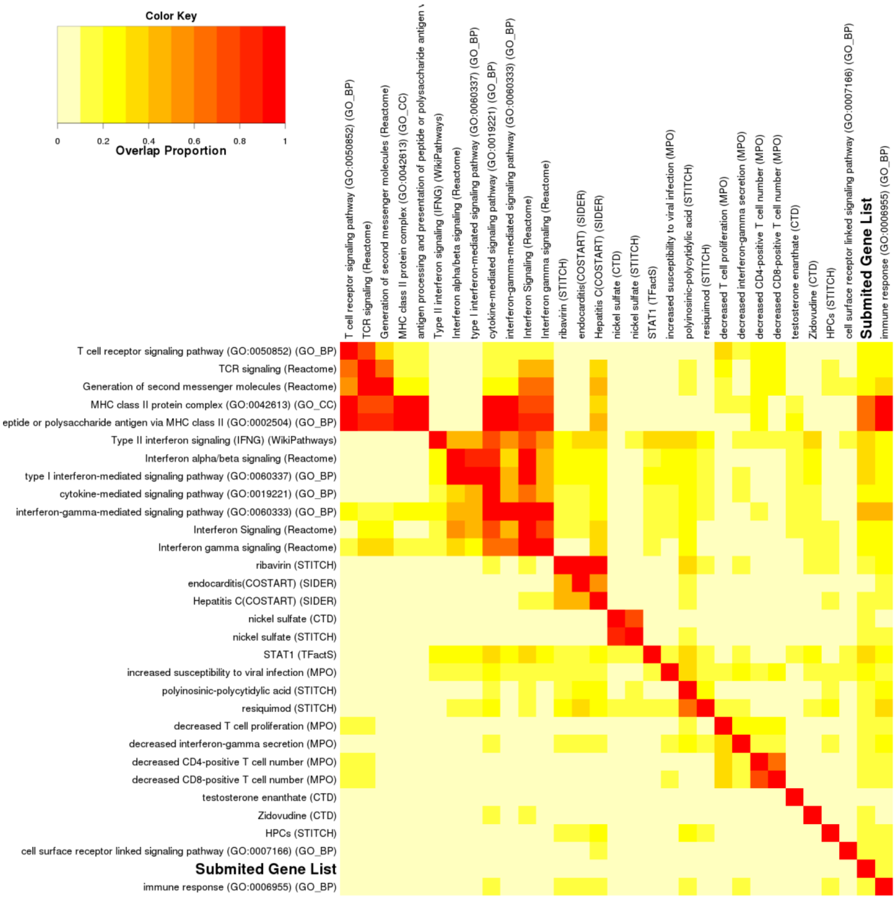
Bicluster11 vs. Bicluster18



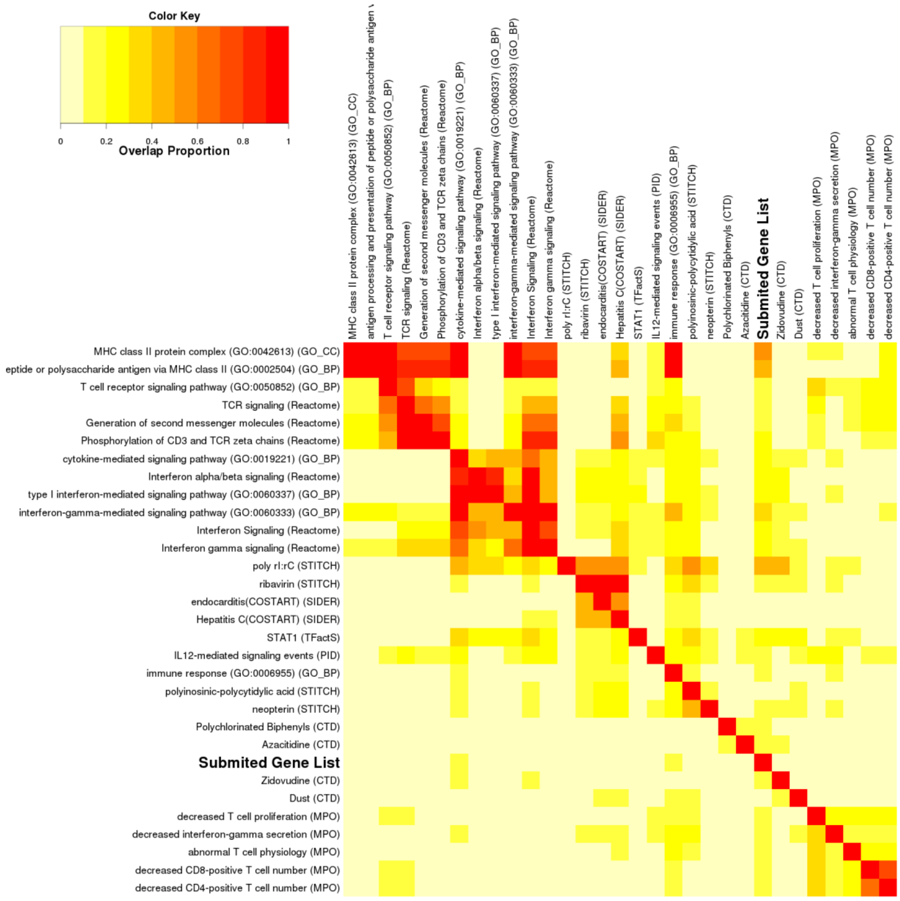
Bicluster11 vs. Bicluster21



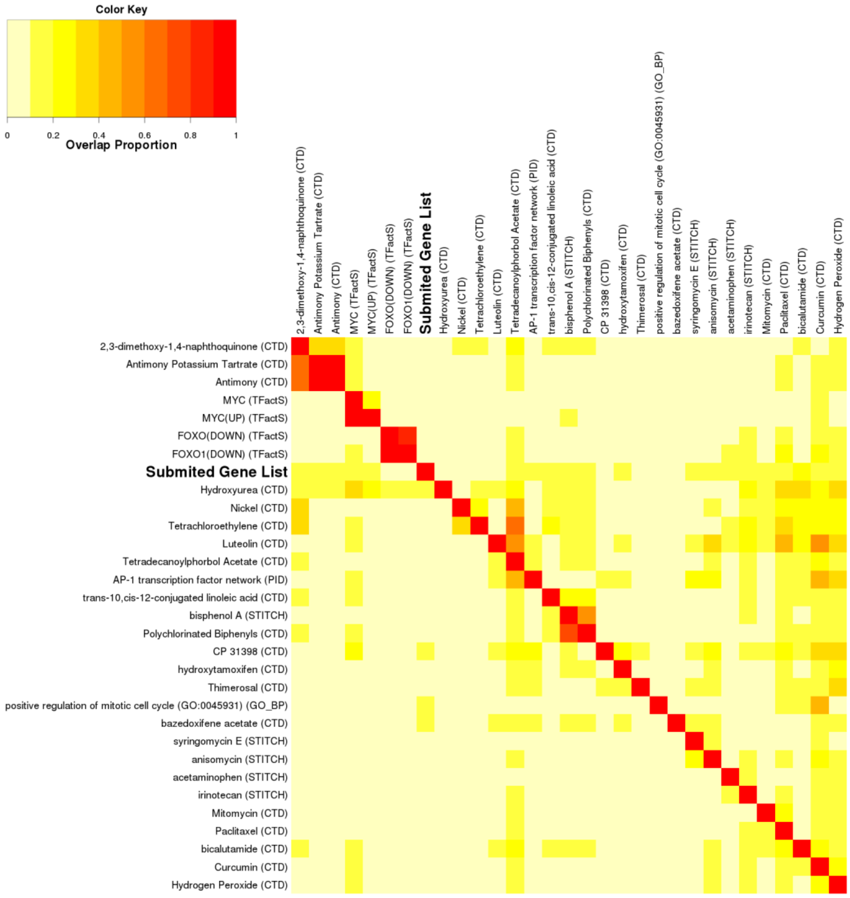
Bicluster11 vs. Bicluster40



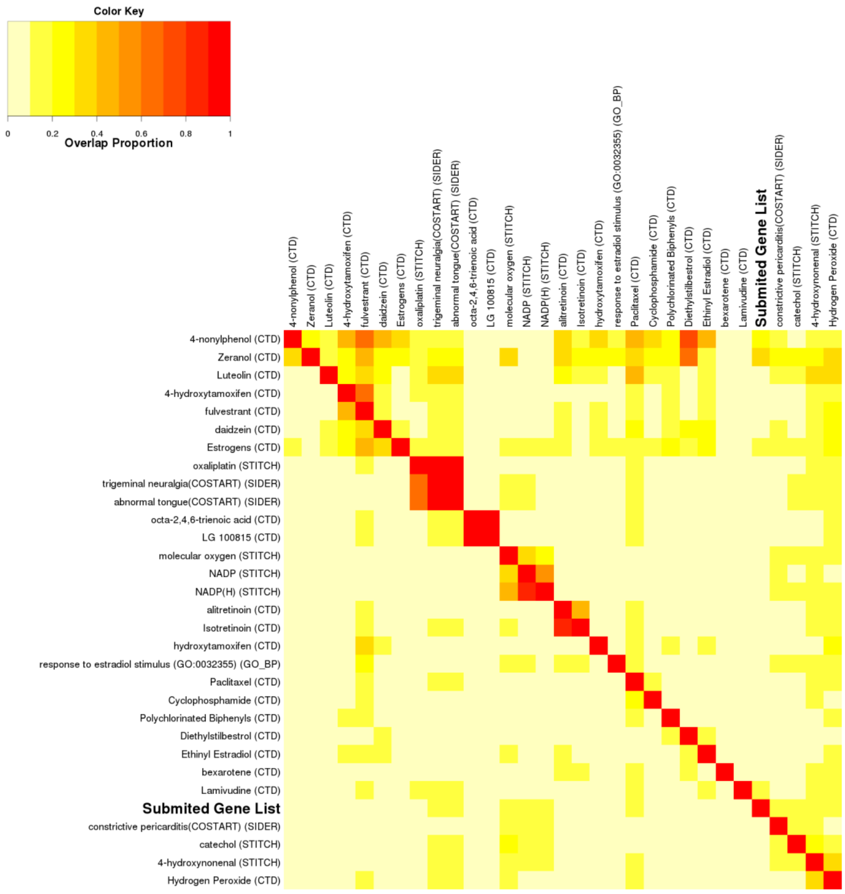
Bicluster15 vs. Bicluster40



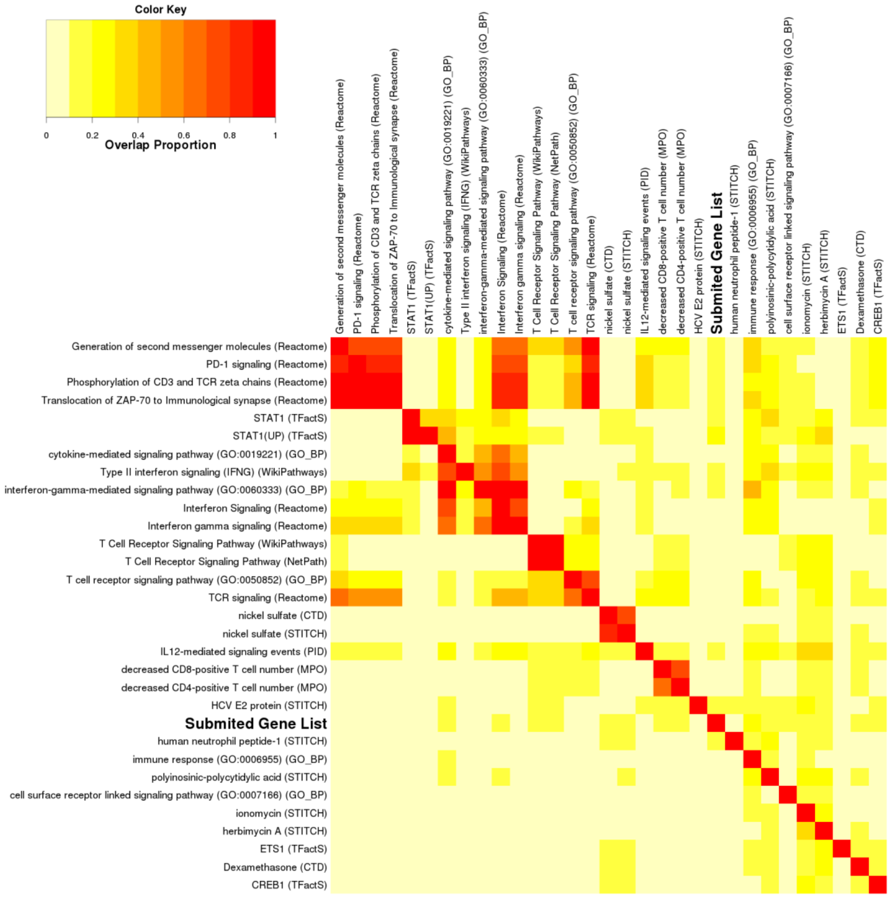
Bicluster22 vs. Bicluster8



Bicluster22 vs. Bicluster21



Bicluster22 vs. Bicluster40



Bicluster22 vs. Bicluster41