TITLE:

Cross-population analysis of high-grade serous ovarian cancer reveals only two robust subtypes

AUTHORS:

Gregory P. Way[a,b,c], James Rudd[a,d], Chen Wang[e], Habib Hamidi[f], Brooke L. Fridley[g], Gottfried Konecny[f], Ellen L. Goode[e], Casey S. Greene[a,b,h,1], Jennifer A. Doherty[a,d,2]

AFFILIATIONS:

[a]Quantitative Biomedical Sciences, Geisel School of Medicine at Dartmouth College, Lebanon, NH; Norris Cotton Cancer Center, Geisel School of Medicine at Dartmouth College, Lebanon, NH

[b]Department of Systems Pharmacology and Translational Therapeutics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA

[c]Genomics and Computational Biology Graduate Program, University of Pennsylvania, Philadelphia, PA 19103, USA

[d]Department of Epidemiology, Geisel School of Medicine at Dartmouth College, Lebanon, NH

[e]Department of Health Sciences Research, Mayo Clinic, Rochester, MN

[f]Department of Medicine, David Geffen School of Medicine, University of California, Los Angeles, CA

[g]Department of Biostatistics, University of Kansas Medical Center, Kansas City, KS

[h]Department of Genetics, Geisel School of Medicine at Dartmouth College, Lebanon, NH

CO-CORRESPONDING AUTHORS:

[1]10-131 SCTR 34th and Civic Center Blvd, Philadelphia, PA 19104; Phone: 215-573-2991; Fax: 215-573-9135; CSGreene@upenn.edu

[2]One Medical Center Drive, Lebanon, NH 03766; Phone: 603-653-9065; Fax: 603-653-9093; Jennifer.A.Doherty@Dartmouth.edu

CONFLICTS OF INTEREST:

The authors do not declare any conflicts of interest.

OTHER PRESENTATIONS:

Aspects of this study were presented at the 2015 AACR conference in Philadelphia.

RUNNING HEAD:

Two ovarian cancer subtypes are similar across populations

KEYWORDS:

Ovarian Cancer; Molecular Subtypes; Unsupervised Clustering

NOTES:

Words: 2,916; Figures: 4; Tables 4; Sup. Figures: 3; Sup. Tables: 6; Sup. Methods

ABSTRACT:

Background

Three to four gene expression-based subtypes of high-grade serous ovarian cancer (HGSC) have been previously reported. We sought to systematically determine the similarity of HGSC subtypes between populations.

Methods

We independently clustered (k = 3 and k = 4) five publicly-available HGSC mRNA expression datasets with >130 tumors using k-means and non-negative matrix factorization. Within each population, we summarized differential expression patterns for each cluster as moderated t statistic vectors using Significance Analysis of Microarrays. We calculated Pearson's correlations of these vectors to determine similarities and differences in expression patterns between clusters. We defined syn-clusters (SC) as sets of clusters that were strongly correlated across populations, and associated their expression patterns with biological pathways using geneset overrepresentation analyses.

Results

Across populations, for k = 3, moderated t score correlations for clusters 1, 2 and 3, respectively, ranged between 0.77-0.85, 0.80-0.90, and 0.65-0.77. For k = 4, correlations for clusters 1-4, respectively, ranged between 0.77-0.85, 0.83-0.89, 0.51-0.76, and 0.61-0.75. Within populations, comparing analogous clusters (k = 3 versus k = 4), correlations were high for clusters 1 and 2 (0.91-1.00), but were lower for cluster 3 (0.22-0.80). Results are similar using non-negative matrix factorization. SC1 corresponds to previously-reported mesenchymal-like, SC2 to proliferative-like, SC3 to immunoreactive-like, and SC4 to differentiated-like subtypes.

Conclusions

The mesenchymal-like and proliferative-like subtypes are remarkably consistent across populations and could be uniquely targeted for treatment. The other two previously described subtypes are considerably less robust, and since cross-population comparison reveals that $k = 3$ and $k = 4$ are both consistent with our results, they may not represent clear subtypes.

INTRODUCTION:

Ovarian cancer is a heterogeneous disease typically diagnosed at a late stage, with high mortality (1). The most aggressive and common histologic type is high-grade serous (HGSC) (2), characterized by extensive copy number variation, methylation events, and mutations (3). Given the genomic complexity of these tumors, mRNA expression can be thought of as a summary measure of these genomic and epigenetic alterations, to the extent that the alterations influence gene expression. Efforts to use whole genome mRNA expression analyses to stratify HGSC into clinically relevant subtypes have yielded potentially promising results, with all studies to date observing three to four subtypes with varying components of mesenchymal, proliferative, immunoreactive, and differentiated gene expression signatures (3–6), and some studies observing survival differences across subtypes (4,5). Tothill *et al*. first identified four HGSC subtypes (as well as two other non-HGSC subtypes) in an Australian population using *k*-means clustering. The authors labeled the subtypes as C1-C6, and observed that women with the C1 subtype, with a stromal-like gene signature, experienced the poorest survival compared to the other subtypes (4). Later, in The Cancer Genome Atlas (TCGA), an assemblage of tumors from various institutions throughout The United States, non-negative matrix factorization (NMF) clustering confirmed the identification of four subtypes which they labeled as mesenchymal, differentiated, proliferative, and immunoreactive, but there were no observed differences in survival (3). The TCGA group also applied NMF clustering to the Tothill data, and noted that analogous subtypes

had similar significantly differentially expressed genes (3). Konecny *et al.* also used NMF clustering in HGSC samples from the Mayo Clinic and identified four subtypes labeled as C1-C4 (5). While these subtypes are similar to those described by TCGA, the Konecny *et al.* refined classifier was better able to differentiate survival between groups in their own data, and in data from TCGA and Bonome *et al.* (6). In the Konecny *et al.* population, as similarly observed in Tothill *et al.*, the mesenchymal-like (described as stromal-like in Tothill *et al.*) and proliferative-like subtypes had poor survival, and the immunoreactive-like subtype had favorable survival (5).

While results from these studies are relatively consistent, in more recent TCGA analyses by the Broad Institute Genome Data Analysis Center (GDAC) Firehose initiative with the largest number of HGSC cases evaluated to date, three subtypes fit the data better than did four (7,8). Also, in the original analysis of the TCGA data, over 80% of the samples were assigned to more than one subtype (9), as were 42% of the Mayo samples. In both TCGA and Tothill *et al.*, ~8-15% of samples were not able to be classified. Therefore, because of this large degree of uncertainty in HGSC subtyping, further characterization of subtypes is essential in order to determine etiologic factors and to develop targeted treatments.

We characterize the underlying patterns of gene expression for three and four HGSC subtypes through a unified bioinformatics pipeline in five independent populations, and assess the robustness of subtypes across these populations. Instead of identifying subtypes in a single population and applying a classification algorithm to identify the same subtypes in other populations, we use unsupervised clustering (performed using both *k*-means clustering and NMF) separately in each population to systematically identify HGSC subtypes. We summarize the expression patterns of over 10,000 genes for each identified subtype and comprehensively characterize correlations between subtype-specific gene expression both within and between

populations. We identify a set of clusters characterized by similar differentially expressed genes that are correlated across populations, which we term "syn-clusters" (SC).

METHODS:

Data Inclusion

We applied inclusion criteria as described in the supplementary materials using data from the R package, curatedOvarianData (10; Table S1) and a separate dataset ("Mayo"; 5). We deposited the Mayo high-grade serous samples as well as other samples with mixed histologies and grades, for a total of 528 additional ovarian tumor samples, in NCBI's Gene Expression Omnibus (GEO; 11). The data can be accessed with the accession number GSE74357 (http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE74357). All tumor samples uploaded were collected with approval by an institutional review board and by the U.S. Department of Health and Human Services. After applying the unified inclusion criteria pipeline, our final analytic datasets include: TCGA (n = 499; 3,7,8); Mayo (n = 379; GSE74357; 5); Yoshihara (n = 256; GSE32062.GPL6480; (12); Tothill (n = 241; GSE9891; 4); and Bonome (n = 185; GSE26712; 7; Table 1). We restricted to the 10,930 genes measured in all 5 populations (Fig. S1). Code to replicate all analyses can be downloaded from https://github.com/greenelab/hgsc_subtypes.

Clustering

Because 3 or 4 subtypes had been reported previously, we focused on examining cluster assignment within and across populations for clusters identified using $k = 3$ or $k = 4$. As detailed in the supplemental methods, we combined the 1,500 genes with the highest variance from each

population (n = 3,698). We performed *k*-means clustering on these 3,698 genes in each population using the R package "cluster" (version 2.0.1; 13) with 20 initializations, and we characterized patterns of changes in sample assignment to clusters when $k = 3$ versus $k = 4$. We further characterized clustering solutions within populations using sample-by-sample Pearson's correlation matrices. We repeated our analyses using NMF in the R package "NMF" (version 0.20.5; 14) with 100 initializations used for each *k*.

Identification of Syn-Clusters

We performed a significance analysis of microarray (SAM) (15,16) analysis on all clusters from each population for $k = 3$ and $k = 4$ using all 10,930 genes. This resulted in a cluster-specific moderated *t* statistic for each of the input genes (17). To summarize the expression patterns of all 10,930 genes for a specific cluster in a specific population, we combined the moderated *t* statistics into a vector of length 10,930. To generate comparable labels across $k = 3$ and $k = 4$ analyses, the $k = 3$ cluster which was most strongly correlated with a $k = 4$ cluster in the TCGA data was labeled "cluster 1" and the second strongest "cluster 2" etc. Clusters in other populations that were most strongly correlated with the TCGA clusters were assigned the same label. Clusters strongly correlated across populations form a syn-cluster (SC); i.e. the clusters from each population that are strongly correlated with each other and with TCGA "cluster 1" belong to SC1. We also compared our sample assignments to subtypes reported in the Tothill, TCGA, and Konecny publications.

Identifying Biological Processes Associated with Syn-Clusters

To annotate the SCs with associated biological processes, we first identified the statistically significantly differentially expressed genes in the SAM list. We used a Bonferroni adjustment taking into account the total number of genes considered (10,930) resulting in a *p*-value cutoff of 4.6×10^{-6}. We used the intersection of these cluster-specific genesets across populations to create the final SC associated genesets. We then input these SC associated genesets into a PANTHER analysis (18) to determine SC-specific overrepresented biological pathways (Supplementary Materials).

RESULTS:

Sample Cluster Assignment

To visually inspect the consistency and distinctness of the clusters, we compared sample-by-sample correlation heatmaps (Fig. 1). For both *k* values and in each population, we observed high sample-by-sample correlations within clusters and relatively low sample-by-sample correlations across clusters (Fig. 1). The clusters in the Bonome population are depicted in gray scale because, in cross-population analyses to identify SCs, their expression patterns did not correlate with the clusters observed consistently in the four other populations (Table 2).

To better understand the changes in cluster assignment for $k = 3$ versus $k = 4$, we compared the number of samples belonging to each cluster by *k* within each population (excluding Bonome; Fig. 2). Overall, the cross-*k* pattern was consistent across populations. Cluster 1 contained essentially the same samples for both $k = 3$ and $k = 4$, as did cluster 2, but samples from cluster 3 when $k = 3$ tended to be split between clusters 3 and 4 when $k = 4$. Additionally, cluster 3 in $k = 4$ tended to have varying numbers of samples from cluster 1 in $k = 3$, and cluster 4 in $k = 4$ tended to include some samples from cluster 2 in $k = 3$ (Fig. 2).

187

188 *Correlation of Cluster-Specific Expression Patterns*

189 Within populations, we observed very high Pearson correlations of moderated *t* score
190 vectors between clusters across $k = 3$ and $k = 4$ (Table 2). We observed strong positive
191 correlations of moderated *t* score vectors between analogous clusters across TCGA, Tothill,
192 Mayo, and Yoshihara cluster assignments (Fig. 3; Table 3). However, while the clusters across k
193 $= 3$ and $k = 4$ were correlated within the Bonome data, they did not correlate strongly with
194 clusters identified in the other populations (Table 3). Because the correlations are so low
195 compared to those observed in all four other populations, the Bonome data are not included in
196 subsequent analyses. Across populations, positive correlations between clusters belonging to the
197 same SC, and negative correlations between clusters in different SCs, were stronger for clusters
198 identified when $k = 3$ than when $k = 4$ (Figure 3). We observed strong positive correlations for
199 both SC1 and SC2 across populations, and strong negative correlations between SC1 and SC2.
200 Weaker and more variable positive correlations were observed for SC3 and SC4 across
201 populations. For $k = 4$, Yoshihara cluster 3 appears to be correlated to both clusters 3 and 4 in the
202 other populations, and cluster 4 to be additionally weakly correlated to cluster 2 in the other
203 populations.

204 Within each population, clusters identified by NMF were very similar to those identified
205 using *k*-means clustering (Fig. 4). Again, both positive and negative correlations are stronger for
206 $k = 3$ than for $k = 4$. Across $k = 3$ and $k = 4$, correlations are strongest for clusters 1 and 2.
207 Sample cluster assignments for both *k*-means and NMF clusters are provided in Table S2.

208

209 *Comparison with previously-identified HGSC clusters*

Our clustering results for the Tothill, TCGA, and Mayo datasets are highly concordant with the clustering described in the original publications (3–5), as evidenced by the high degree of overlap in sample assignments to the previously-defined clusters (Table 4). Our SC1 for both *k*-means analyses was mapped to the "Mesenchymal" label from TCGA, "C1" from Tothill, and mostly to "C4" from Mayo. SC1 was the most stable in our analysis within all datasets, across $k = 3$ and $k = 4$, and across clustering algorithms. SC2 was most similar to the "Proliferative" label from TCGA, "C5" from Tothill, and "C3" from Mayo. This was the second most stable SC. SC3 for $k = 3$ was associated with both the "Immunoreactive" and "Differentiated" TCGA labels, "C2" and "C4" in Tothill, and "C1" and "C2" in Mayo. When setting *k*-means to find four clusters, SC3 was associated with "Immunoreactive", "C2", and "C1" while SC4 was associated with "Differentiated", "C4", and "C2" for TCGA, Tothill, and Mayo respectively. Pathway analysis results for all SCs are summarized in more detail in the supplementary materials and are presented in supplementary table S5.

DISCUSSION:

Previous studies have identified three to four subtypes of HGSC, but it is difficult to compare the results because each study performed analyses with different sample inclusion criteria, different gene expression platforms, and different statistical methods. In contrast, we used uniform sample inclusion criteria and applied *k*-means clustering and NMF through a standardized pipeline to five distinct publicly-available HGSC datasets including American, Australian, and Japanese women. To identify the HGSC clusters, we included only the 1,500 most variable genes in each population, as was done in TCGA analyses. However, we used the combined set of most variable genes across the five populations to perform clustering, to ensure

that important genes, which may not have met the threshold in one population but did in others, were still considered. For each cluster in each population, we summarized the differential expression of 10,930 genes, and compared these cluster-specific gene expression patterns both within and between populations to determine which genes in a specific cluster were over- or under-expressed. This process allowed us to identify syn-clusters (SC) as groups of analogous clusters observed across populations. Despite considerable diversity in the populations studied and the assay platforms used, in four of the five populations studied, we identified two very distinct SCs (SC1 and SC2), and a third SC (SC3) and potentially fourth SC (SC4) that are much less robust across populations. The results were also similar using two distinct unsupervised clustering algorithms in all populations, which further validate the presence of robust gene expression based subtypes. Compared to the clusters reported in TCGA, Tothill, and Konecny, SC1 was most similar to the mesenchymal/C1/C4 subtype and SC2 was most similar to the proliferative/C5/C3 subtype, respectively. While concordance between the original Tothill and TCGA subtypes was reported in the TCGA HGSC publication (3), our analysis included an additional 59 TCGA samples. As well, we included an additional 210 samples from Mayo that were not analyzed in the original Konecny *et al.* publication (5).

While the groupings of samples from these data-driven, agnostic analyses are quite similar to those previously reported, we did not observe any strong patterns in survival differences across the subtypes that we identified (see Supplementary Material). However, we would not necessarily expect to find differences in survival unless the biologic characteristics of the tumor subtypes translate into different responses to standard treatments. Instead, our goal is to identify robust subtypes so that they can be exhaustively characterized and targeted treatments can be developed. That SC1 and SC2 were found regardless of the number of clusters specified,

and global expression patterns were so similar in separate distinct populations, increases our confidence that each of these clusters represents a set of reproducible biological signals. As well, the strong positive correlations within and between populations indicate homogeneity of gene expression patterns across populations for SC1 and SC2. The strong negative correlations between SC1 and SC2 also indicate that they are clearly distinct from one another; this is emphasized by the inverse direction of expression for the immune system process genes. For SC3 and SC4, both positive and negative correlations are less strong, and there is some positive correlation between the two clusters, particularly in the Japanese population.

The consistency of SC1 and SC2 across *k* parameters and between diverse populations is remarkable for a number of reasons. While these studies represent the largest collections of HGSC tumors to date, given the difficulties in collecting fresh frozen tissue for large-scale gene expression studies, it is unclear how accurately any of these data sets reflect the underlying population distribution of HGSC subtypes. Results from gene expression/RNA sequencing assays in large, population-based formalin-fixed paraffin-embedded (FFPE) tumor collections will be important in further informing the definitions of HGSC subtypes. As well, given the intra-tumor heterogeneity that is likely to exist (20), our approach would be strengthened by having data on multiple areas of the tumors. Finally, since histology and grade classification have changed over time (21,22), it is unclear whether the populations we studied used comparable guidelines to determine histology and grade. We attempted to exclude all low grade serous and endometrioid samples because they often have very different gene expression patterns and more favorable survival compared to their higher grade counterparts (2). While the Bonome publication specified that they included only high-grade tumors, grade is not included in the Bonome GSE26712 data set, so we were unable to determine whether the grade distribution

differs from the other studies (7). At any rate, it is unclear why the Bonome clusters, while internally consistent across *k*, did not correspond to the clusters observed in other populations. If samples are misclassified with respect to grade or other characteristics, depending on the extent of the misclassification, lower correlations and consequently difficulty assigning SCs could result.

Our study demonstrates that two SCs of HGSC, "mesenchymal-like" and "proliferative-like", are clearly identified within and between populations. This suggests the presence of at least two robust HGSC subtypes that are either etiologically distinct, or acquire phenotypically determinant alterations through their development. These two SCs have different sets of significantly enriched pathways, which indicate distinct processes regulating and promoting tumorigenesis. The "mesenchymal-like" subtype includes aberrant regulated genes involved with extracellular matrix and cell to cell adhesion processes, while the "proliferative-like" subtype includes down-regulated immune-related genes, consistent with previous studies which have identified a negative immune signature in this subtype (5). The results also suggest that one or more additional subtypes, "immunoreactive-like" and "differentiated-like", exist but are more variable across populations or may represent, for example, steps along an immunoreactive continuum. Data on copy number alterations, mutation burden, or epigenetic effects may capture more of these clusters' variability. Because the "mesenchymal-like" and "proliferative-like" subtypes are consistently observed within and between populations, these subtypes are the best candidates for further characterization and development of subtype-specific treatment strategies. Future studies are needed to better sub-classify tumors that do not belong to either of these subtypes.

ACKNOWLEDGEMENTS:

We would like to thank Sebastian Armasu for help with statistical analyses and data processing and Emily Kate Shea for helpful discussions.

FUNDING:

This work was supported by the Institute for Quantitative Biomedical Sciences; the Norris Cotton Cancer Center Developmental Funds; the National Cancer Institute at the National Institutes of Health (R01 CA168758 to J.A.D., F31 CA186625 to J.R., R01 CA122443 to E.L.G.); the Mayo Clinic Ovarian Cancer SPORE (P50 CA136393 to E.L.G.); the Mayo Clinic Comprehensive Cancer Center-Gene Analysis Shared Resource (P30 CA15083); the Gordon and Betty Moore Foundation's Data-Driven Discovery Initiative (grant number GBMF 4552 to C.S.G.); and the American Cancer Society (grant number IRG 8200327 to C.S.G.).

FIGURE LEGENDS:

Figure 1. Sample by sample Pearson correlation matrices. Top panel: $k = 3$. Bottom panel: $k = 4$. The color bars are coded as blue, syn-cluster 1 (SC1); red, SC2; green, SC3; and purple, SC4. In the matrices, red represents high correlation, blue low correlation, and white intermediate correlation. The scales are slightly different in each population because of different correlational structures. The grey Bonome clusters indicate clusters not correlating well with any cluster from the other populations.

Figure 2. Sample membership distribution changes when setting k means to find $k = 3$ and $k = 4$. The bars represent sample cluster membership with $k = 4$ and the colors indicate the same samples' cluster assignments for when $k = 3$.

Figure 3. SAM moderated t score Pearson correlations. The color bars are coded as blue, syn-cluster 1 (SC1); red, SC2; green, SC3; and purple, SC4. (A) Correlations across datasets for *k* means $k = 3$. (B) Correlations across datasets for k means $k = 4$. The matrices are symmetrical and the upper triangle holds scatter plots for each comparison where each point represents one of the 10,930 genes measured in each population.

Figure 4. SAM moderated t score Pearson correlations of clusters formed by *k* means clustering and NMF clustering. Results are shown for both methods when setting each algorithm to find 3 and 4 clusters. The color bars are coded as blue, syn-cluster 1 (SC1); red, SC2; green, SC3; and purple, SC4.

Supplementary Figure S1. Overlapping genes assayed using either the HG-U1133 Affymetrix platform (TCGA, Tothill, Bonome) or the Agilent 4x44K platform (Mayo, Yoshihara). Differences across datasets arise from inherent array differences and/or differences in quality control preprocessing.

Supplementary Figure S2. NMF consensus matrices for datasets when (A) $k = 3$ and (B) $k = 4$. The first track represents cluster membership for k means clusters and the second track

represents silhouette widths. Note however that the NMF clusters are not mapped to the ordered k means clusters.

Supplementary Figure S3. Kaplan-Meier survival curves. For each population, the top plot is for $k = 3$ and the bottom plot is for $k = 4$.

REFERENCES:

1. Kurman RJ, Shih I-M. The Origin and Pathogenesis of Epithelial Ovarian Cancer: A Proposed Unifying Theory: Am J Surg Pathol. 2010;34:433–43.

2. Vang R, Shih I-M, Kurman RJ. Ovarian Low-grade and High-grade Serous Carcinoma: Pathogenesis, Clinicopathologic and Molecular Biologic Features, and Diagnostic Problems. Adv Anat Pathol. 2009;16:267–82.

3. The Cancer Genome Atlas. Integrated genomic analyses of ovarian carcinoma. Nature. 2011;474:609–15.

4. Tothill RW, Tinker AV, George J, Brown R, Fox SB, Lade S, et al. Novel Molecular Subtypes of Serous and Endometrioid Ovarian Cancer Linked to Clinical Outcome. Clin Cancer Res. 2008;14:5198–208.

5. Konecny GE, Wang C, Hamidi H, Winterhoff B, Kalli KR, Dering J, et al. Prognostic and Therapeutic Relevance of Molecular Subtypes in High-Grade Serous Ovarian Cancer. JNCI J Natl Cancer Inst. 2014;106.

6. Bonome T, Levine DA, Shih J, Randonovich M, Pise-Masison CA, Bogomolniy F, et al. A gene signature predicting for survival in suboptimally debulked patients with ovarian cancer. Cancer Res. 2008;68:5478–86.

7. Broad Institute TCGA Genome Data Analysis Center. Analysis Overview for Ovarian Serous Cystadenocarcinoma (Primary solid tumor cohort) - 02 April 2015. 2015. doi:10.7908/C1SQ8ZFW.

8. Broad Institute TCGA Genome Data Analysis Center. Clustering of mRNA expression: consensus NMF. 2015. doi:10.7908/C1BR8R71

9. Verhaak RGW, Tamayo P, Yang J-Y, Hubbard D, Zhang H, Creighton CJ, et al. Prognostically relevant gene signatures of high-grade serous ovarian carcinoma. J Clin Invest. 2012.

376 10. Ganzfried BF, Riester M, Haibe-Kains B, Risch T, Tyekucheva S, Jazic I, et al.
377 curatedOvarianData: clinically annotated data for the ovarian cancer transcriptome.
378 Database. 2013.

379 11. Edgar R, Domrachev M, Lash AE. Gene Expression Omnibus: NCBI gene expression and
380 hybridization array data repository. Nucleic Acids Res. 2002;30(1):207-210.

381 12. Yoshihara K, Tsunoda T, Shigemizu D, Fujiwara H, Hatae M, Fujiwara H, et al. High-Risk
382 Ovarian Cancer Based on 126-Gene Expression Signature Is Uniquely Characterized by
383 Downregulation of Antigen Presentation Pathway. Clin Cancer Res. 2012;18:1374–85.

384 13. Maechler M, Rousseeuw P, Struyf A, Hubert M, Hornik K. cluster: Cluster Analysis Basics
385 and Extensions. 2014;R package version 1.15.3.

386 14. Brunet J-P, Tamayo P, Golub TR, Mesirov JP. Metagenes and molecular pattern discovery
387 using matrix factorization. Proc Natl Acad Sci. 2004;101:4164–9.

388 15. Tusher VG, Tibshirani R, Chu G. Significance analysis of microarrays applied to the
389 ionizing radiation response. Proc Natl Acad Sci. 2001;98:5116–21.

390 16. Schwender H, Krause A, Ickstadt K. Identifying interesting genes with sigenes. RNews.
391 2006;6:45–50.

392 17. Schwender H. siggenes: Multiple testing using SAM and Efron's empirical Bayes
393 approaches. 2012;R package version 1.40.0.

394 18. Mi H, Muruganujan A, Thomas PD. PANTHER in 2013: modeling the evolution of gene
395 function, and other gene attributes, in the context of phylogenetic trees. Nucleic Acids Res.
396 2013;41:D377–86.

397 19. Rousseeuw PJ. Silhouettes: A graphical aid to the interpretation and validation of cluster
398 analysis. J Comput Appl Math. 1987;20:53-65.

399 20. Blagden SP. Harnessing Pandemonium: The Clinical Implications of Tumor Heterogeneity
400 in Ovarian Cancer. Front Oncol. 2015;5.

401 21. Silverberg SG. Histopathologic grading of ovarian carcinoma: a review and proposal. Int J
402 Gynecol Pathol Off J Int Soc Gynecol Pathol. 2000;19:7–15.

403 22. Soslow RA. Histologic Subtypes of Ovarian Carcinoma: An Overview. Int J Gynecol Pathol
404 2008

405

406

407

408

409 **Table 1**: Characteristics of the populations included in the seven analytic data sets

	TCGA	Mayo	Yoshihara *et al.*	Tothill *et al.*	Bonome *et al.*
GEO		GSE74357	GSE32062	GSE9891	GSE26712
Platform	Affy HGU1133	Agilent 4x44K	Agilent 4x44K	Affy HGU1133	Affy HGU1133
Population	United States	United States	Japan	Australia	United States
Original Sample Size	578	528	260	285	195
Analytic Sample Size[b]	499	379	256	242	185
Age [Mean (SD)]	60.0 (11.6)	62.9 (11.3)	NA	60.3 (10.3)	61.5 (11.9)
Stage					
I	10 (2%)	7 (3%)	0 (0%)	11 (5%)	0 (0%)
II	17 (4%)	11 (3%)	0 (0%)	8 (4%)	0 (0%)
III	351 (80%)	275 (73%)	202 (79%)	178 (83%)	146 (80%)
IV	63 (14%)	86 (23%)	54 (21%)	17 (8%)	36 (20%)
Grade					
2	55 (12%)	3 (1%)	130 (51%)	80 (37%)	NA
3	386 (88%)[a]	376 (99%)	126 (49%)	134 (63%)	NA
Debulking					
Optimal	325 (74%)	287 (76%)	101 (39%)	132 (62%)	89 (49%)
Suboptimal	116 (26%)	87 (23%)	155 (61%)	82 (38%)	93 (51%)

410

411 NA: Data not reported

412 [a]One sample was labeled as 'Grade 4' in TCGA

413 [b]samples without full survival data were exluded in survival analyses

414 **Table 2:** SAM moderated t score vector Pearson correlations between clusters identified using *k*

415 = 3 versus *k* = 4 within each population.

	Cluster 1	Cluster 2	Cluster 3	Cluster 4[a]
TCGA	0.99	0.98	0.69	0.53
Mayo	0.91	0.97	0.48	0.67
Yoshihara *et al.*	1.00	0.94	0.80	0.59
Tothill *et al.*	0.95	1.00	0.22	0.89
Bonome *et al.*	0.98	0.99	0.80	0.28

416 [a]Correlations for cluster 3 (*k* = 3) versus cluster 4 (*k* = 4).

417

418 **Table 3**: SAM moderated t score vector Pearson correlations between analogous clusters across

419 populations[a]

	Cluster 1	Cluster 2	Cluster 3	Cluster 4
k = 3[a]	0.77 - 0.85	0.80 - 0.90	0.65 - 0.77	NA
k = 4[a]	0.77 - 0.85	0.83 - 0.89	0.51 - 0.76	0.61 - 0.75
Bonome *k* = 3[b]	0.45 – 0.46	-0.02 - 0.12	0.22 - 0.42	NA
Bonome *k* = 4[b]	0.50 - 0.57	-0.04 - 0.04	0.13 - 0.29	0.26 - 0.43

420 [a]Correlation ranges for TCGA, Mayo, Yoshihara, and Tothill.

421 [b]Bonome is removed from gene set analyses because of low correlating clusters.

422

423

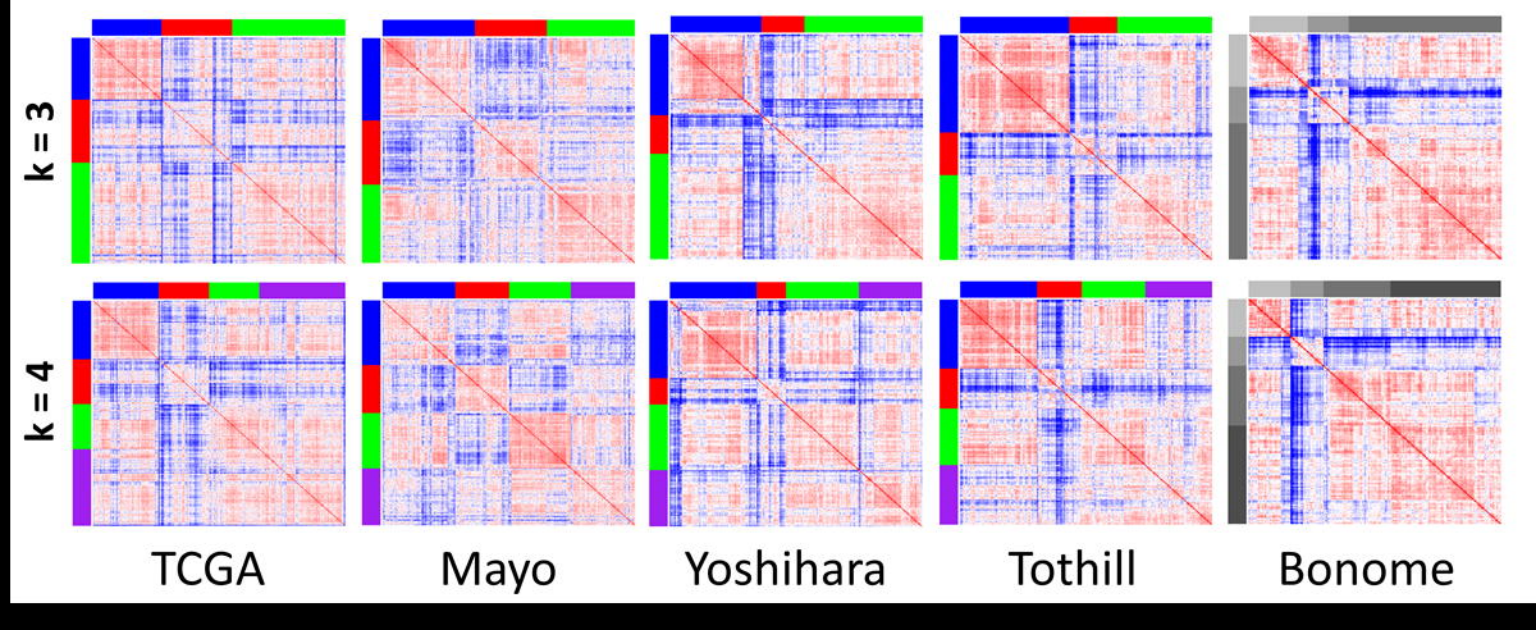
Table 4: Distributions of sample membership in the clusters identified in our study by the original cluster assignments in the TCGA, Tothill, and Konecny studies. Clusters identified in our study using *k*-means clustering with $k = 3$ and $k = 4$

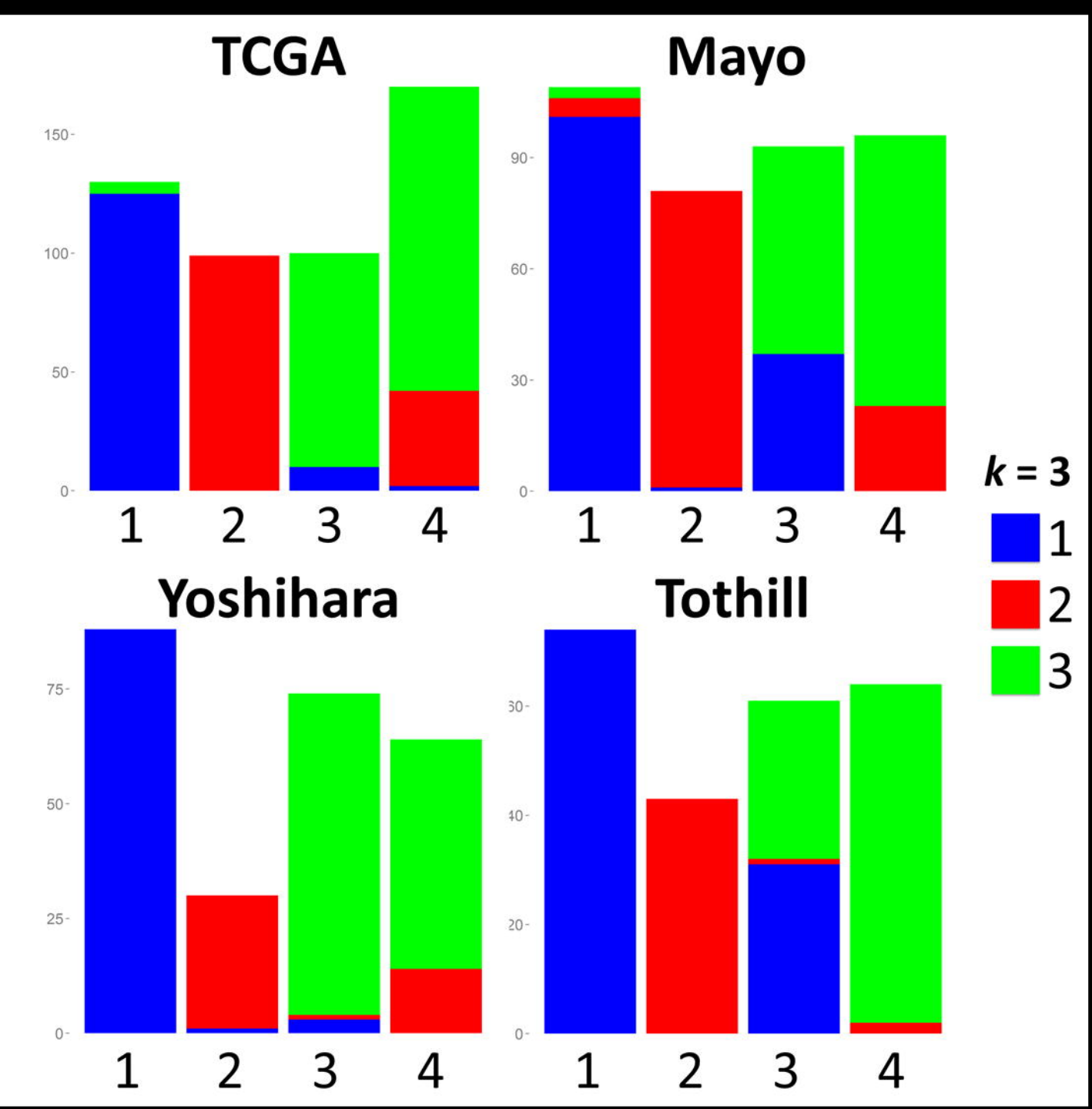
	TCGA					Tothill *et al.*							Konecny *et al.*				
	Mes	Pro	Imm	Dif	NC[a]	C1	C2	C3	C4	C5	C6	NC[a]	C1	C2	C3	C4	NA[b]
Cluster 1	98	2	20	11	6	77	22	0	0	0	0	6	16	13	2	26	82
Cluster 2	1	111	0	11	16	1	0	0	3	35	2	5	0	16	36	0	56
Cluster 3	0	21	75	106	21	0	22	6	41	0	0	22	26	31	5	0	70
	Mes	Pro	Imm	Dif	NC[a]	C1	C2	C3	C4	C5	C6	NC[a]	C1	C2	C3	C4	NA[b]
Cluster 1	97	4	12	12	5	74	0	0	0	0	0	0	7	12	3	25	62
Cluster 2	1	85	0	0	13	1	0	0	1	34	2	5	0	9	31	0	41
Cluster 3	0	5	80	3	12	3	42	0	1	1	0	14	29	6	0	1	57
Cluster 4	1	40	3	113	13	0	2	6	42	0	0	14	6	33	9	0	48

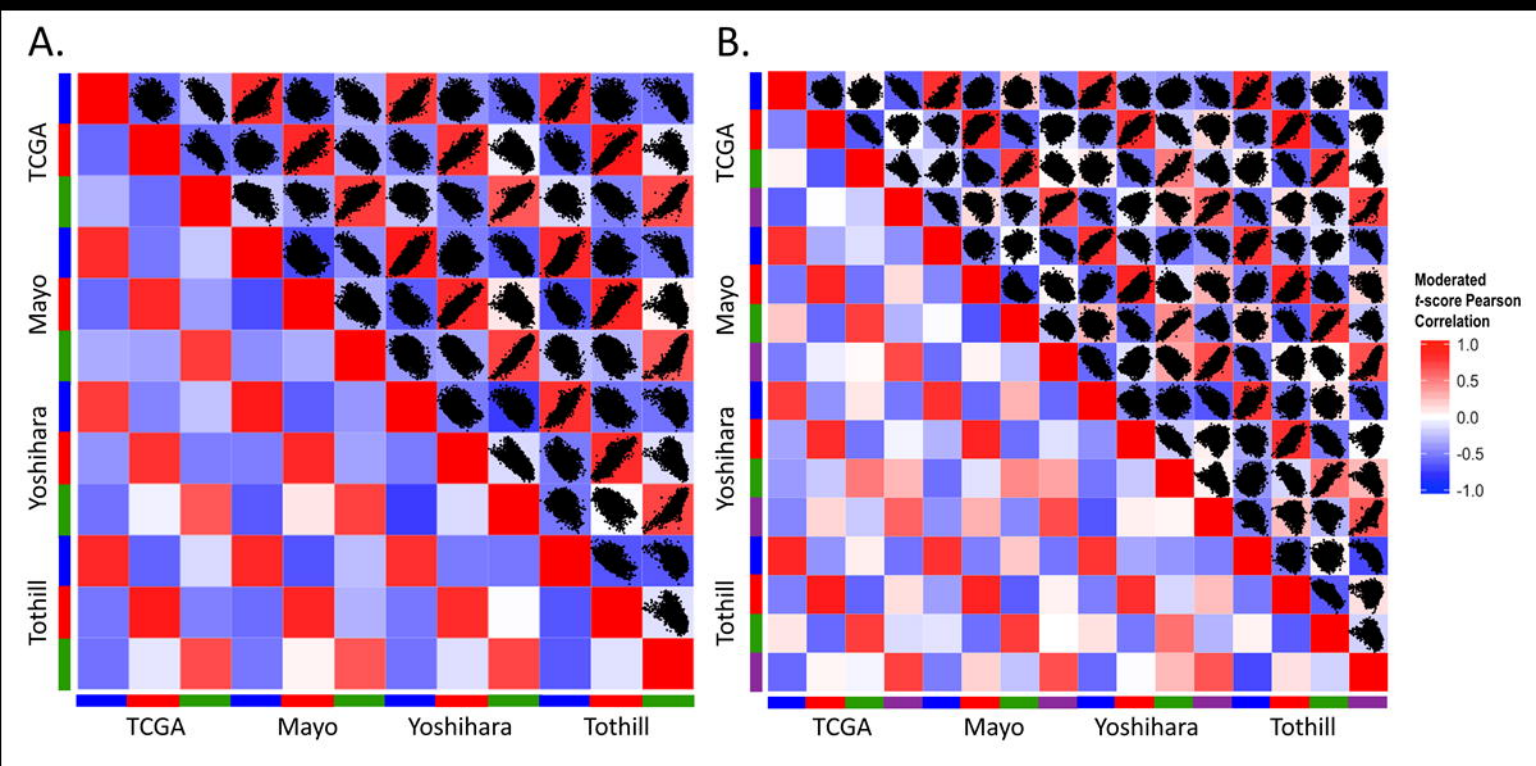
[a]NC = Samples not clustered in original publication

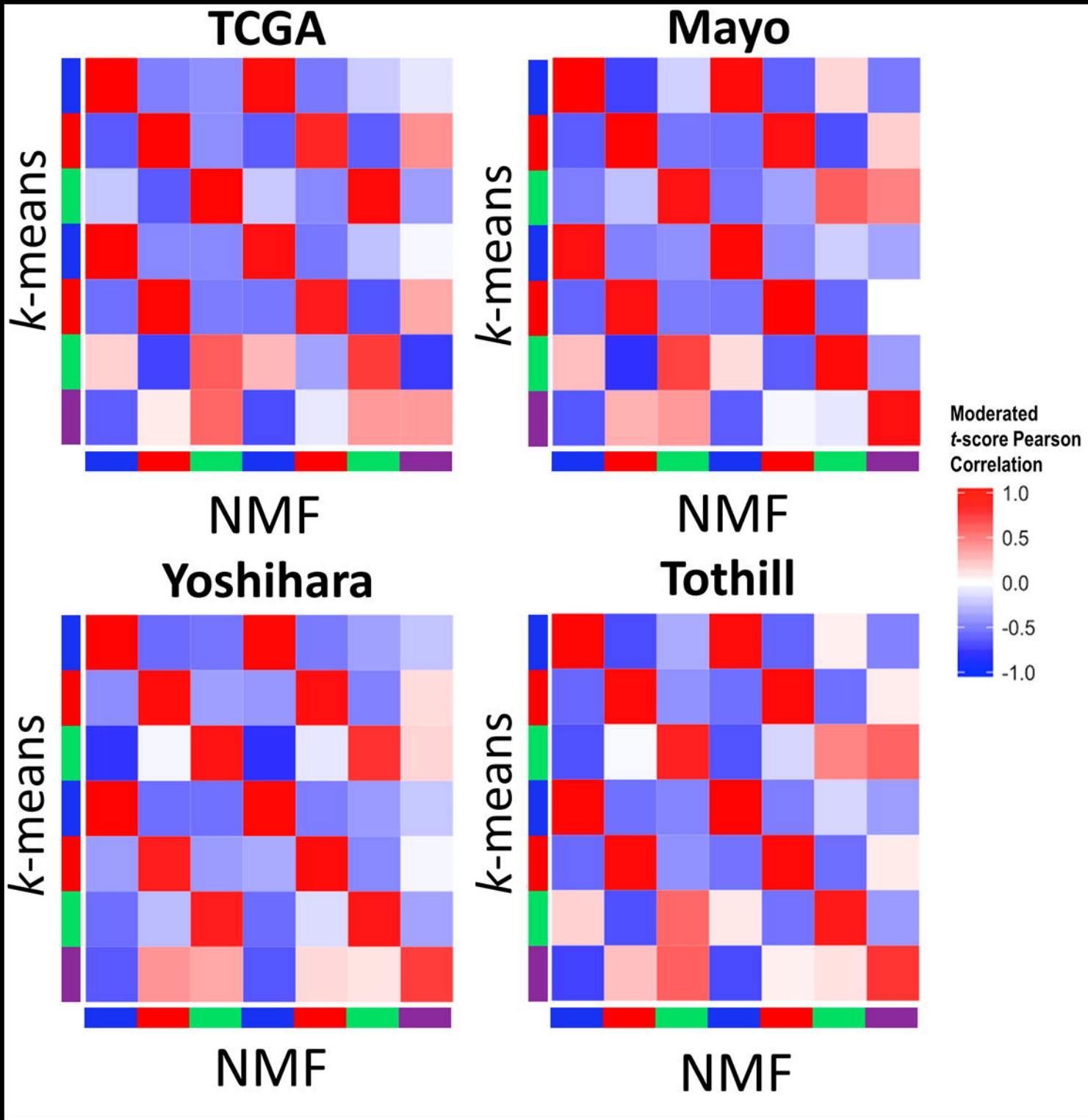
[b]NA = Samples not assessed at the time of the original publication

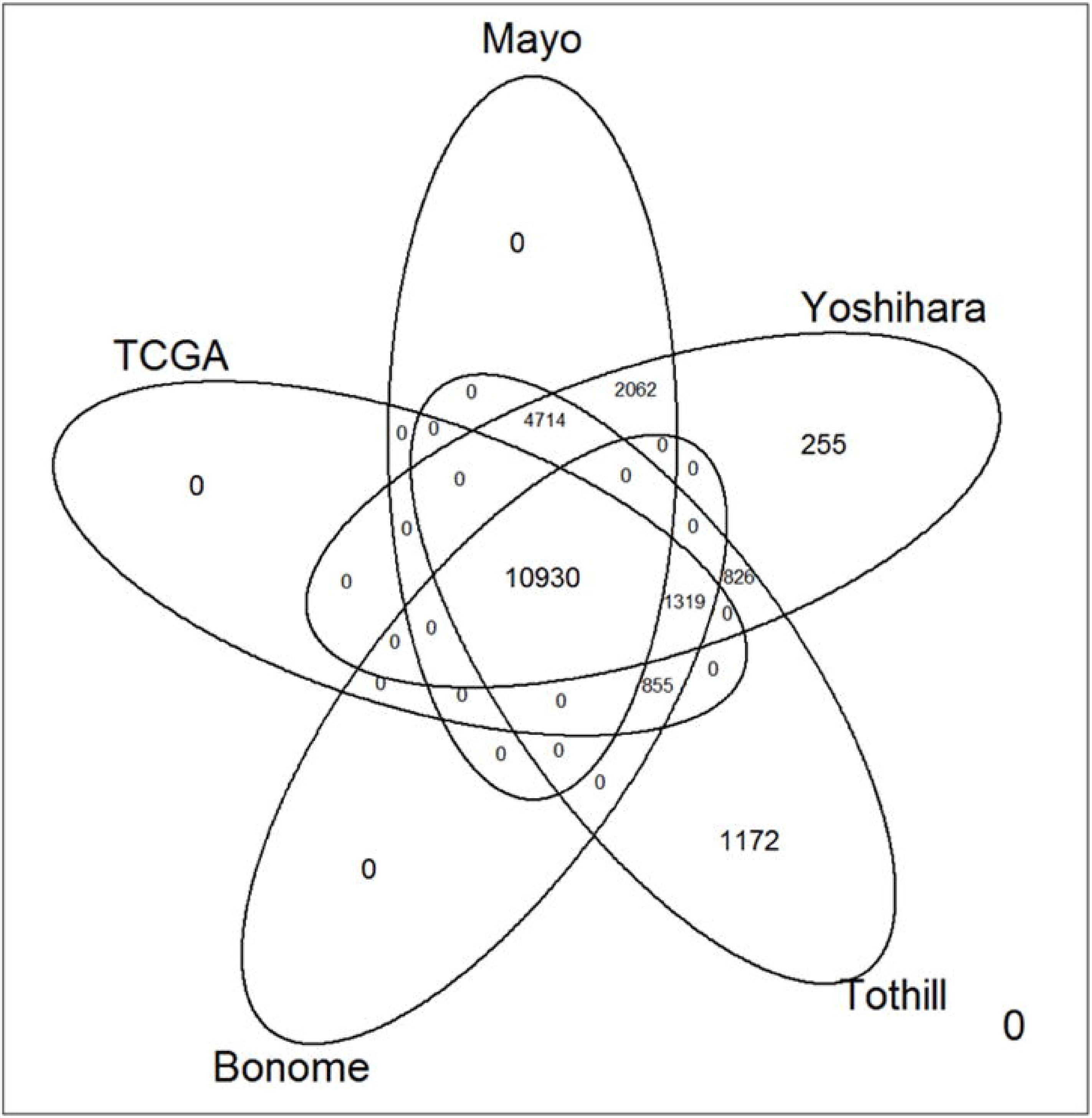
NOTE: The corresponding labels for the generally similar HGSC gene expression subtypes observed in the TCGA, Tothill, and Konecny studies are, respectively: mesenchymal/C1/C4, proliferative/C5/C3, immunoreactive/C2/C1, and differentiated/C4/C2)


k = 3
k = 4
TCGA
Mayo
Yoshihara
Tothill
Bonome

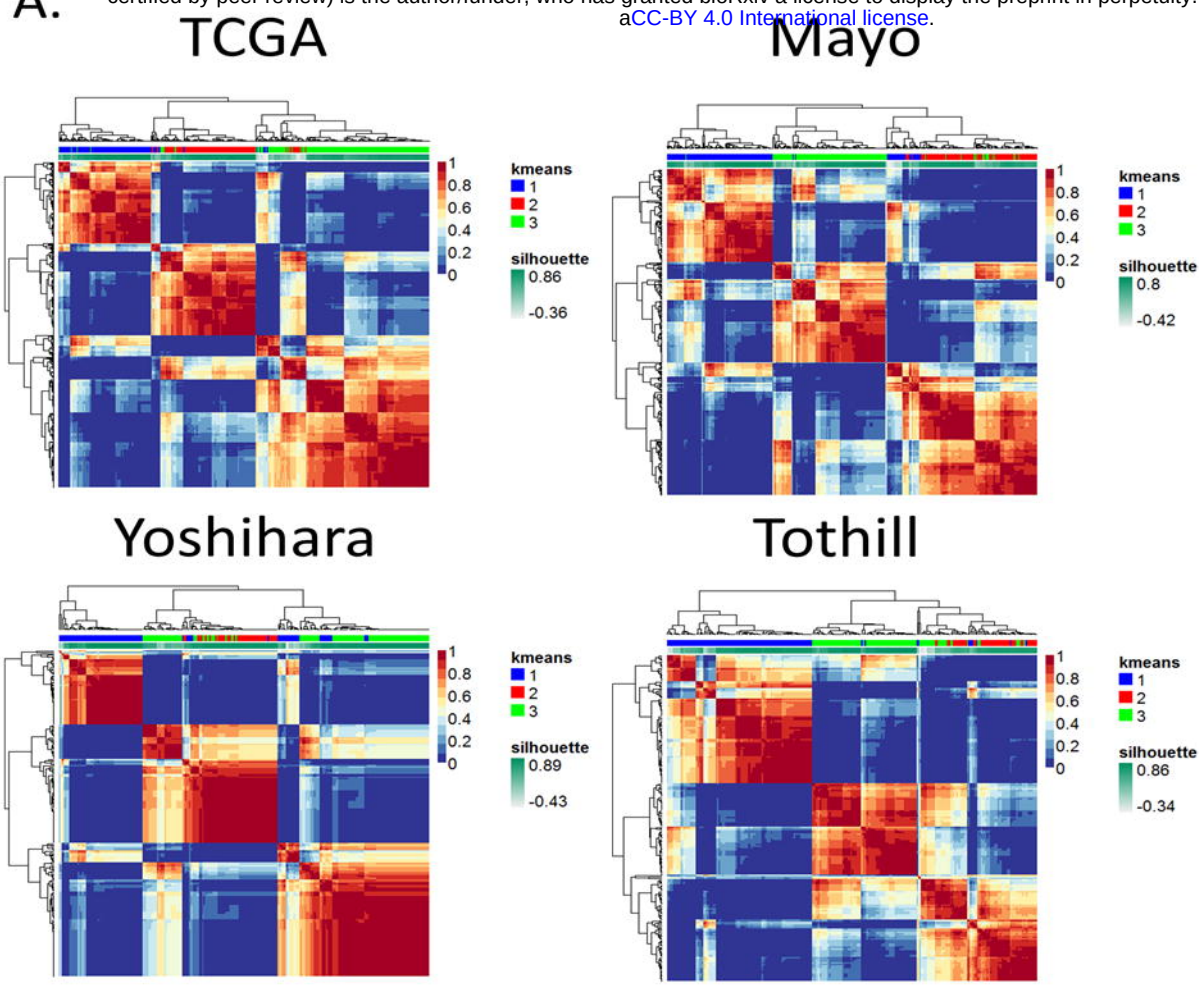

TCGA
150
100
50
0
1
2
3
4
Mayo
90
60
30
0
1
2
3
4
Yoshihara
75
50
25
0
1
2
3
4
Tothill
30
40
20
0
1
2
3
4
k = 3
1
2
3


A.
B.
TCGA
Mayo
Yoshihara
Tothill
Moderated
t-score Pearson
Correlation
1.0
0.5
0.0
-0.5
-1.0

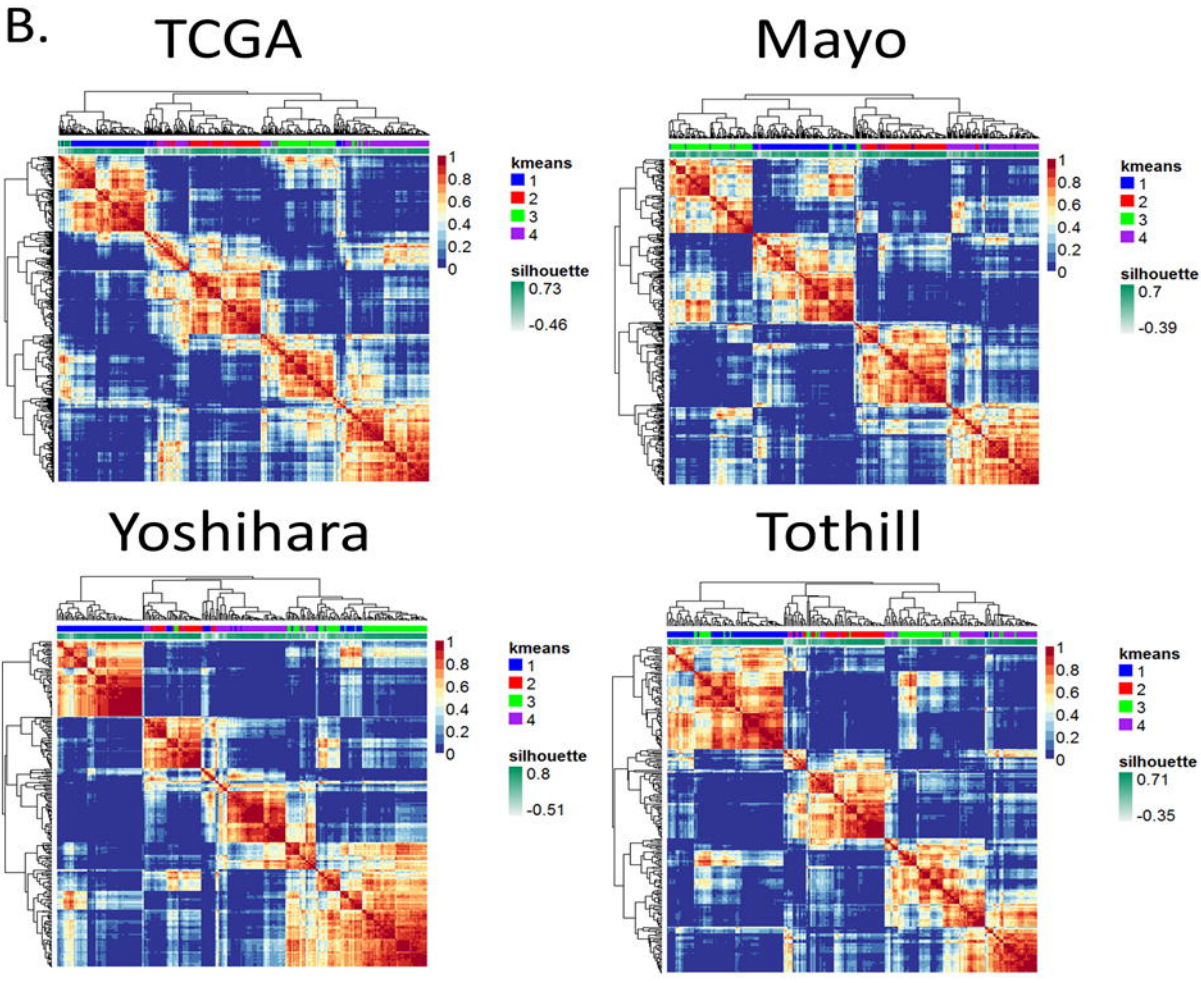

TCGA
Mayo
Yoshihara
Tothill
k-means
NMF
Moderated
t-score Pearson
Correlation
1.0
0.5
0.0
-0.5
-1.0

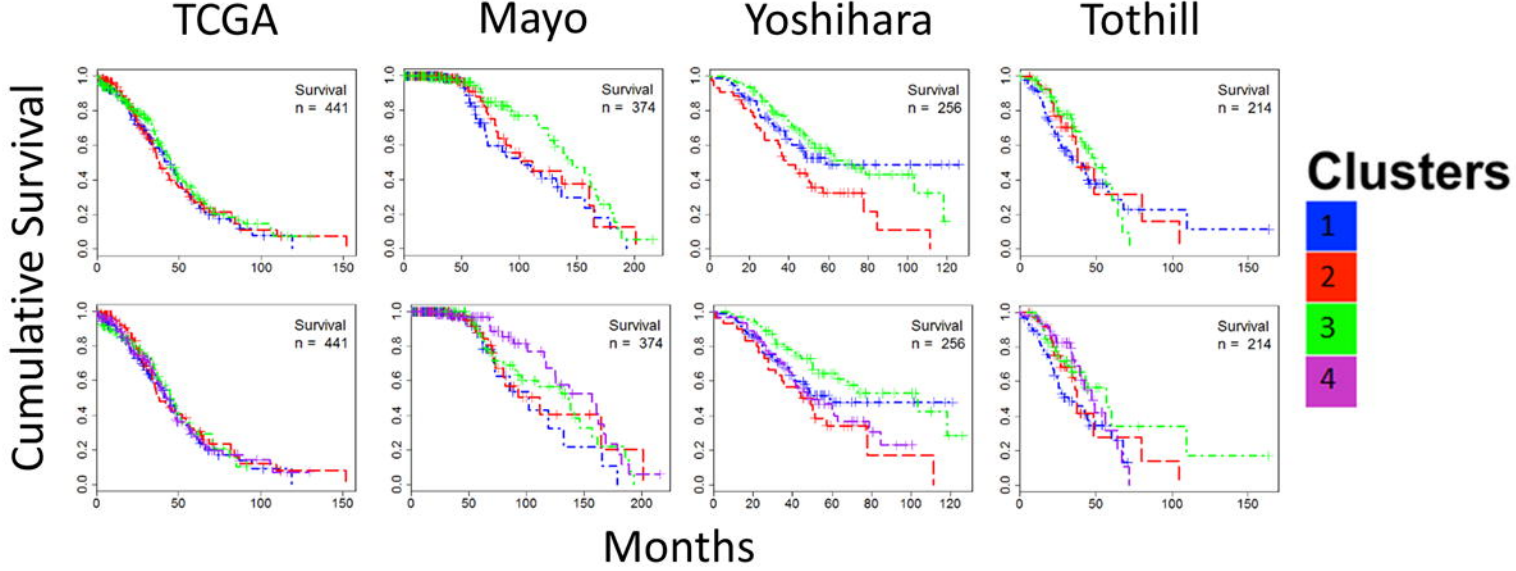

Mayo
Yoshihara
TCGA
Bonome
Tothill
0
255
0
2062
4714
10930
1319
826
855
1172
0
0

A.



B.




TCGA
Mayo
Yoshihara
Tothill
Cumulative Survival
Survival
n = 441
n = 374
n = 256
n = 214
Months
Clusters
1
2
3
4