Reproduction of Published Ovarian Subtyping Schemes

Gregory M. Chen

July 3, 2015

This document reports the results of my initial implementations of ovarian cancer subtyping classification schemes described below. Using the gene lists provided in supplementary data for each publication, I have implemented the subtyping classifiers as described in the methods or supplementary texts, replicating their methods as closely as possible.

Contents

1 Konecny et al., 2014

First, we implemented the subtyping classification scheme by Konecny et al., 2014. Their subtype classification scheme uses a nearest-centroids approach with Spearman's correlation coefficient as the distance measure. The authors provided a list of 635 selected probe sets for classifying new cases. To allow cross-platform applicability, we implemented the subtype classifier using the 575 unique Entrez gene IDs corresponding to these probe sets (with the mean value taken for multiple probe sets mapping to the same gene ID). In their supplementary materials, Konecny et al. report their predicted subtypes on a validation dataset (Bonome et al.). To assess our implementation, we compared our predicted subtypes on the Bonome dataset. The contingency matrix is given below, indicating a large degree of concordance between our implemented subtypes and the author's supplementary data. Overall, 96.15% of samples were classified identically between our implementation and the supplementary results.

	Konecny Subtypes from Supplementary			
Implemented Konecny Subtypes	c1	c2	c3	c4
c1	37	0	0	0
c2	2	64	0	0
c3	0	2	37	0
c4	3	0	0	37

Table 1: Contingency table showing strong concordance between using our implementation of the Konecny subtyping classifier and the predictions given in the supplementary materials of the Konecny manuscript. These predictions were made on the dataset of Bonome et al.

We also generated survival curves for our implemented subtypes on the Bonome set. This plot resembles Figure 3A of the main text of Konecny et al.

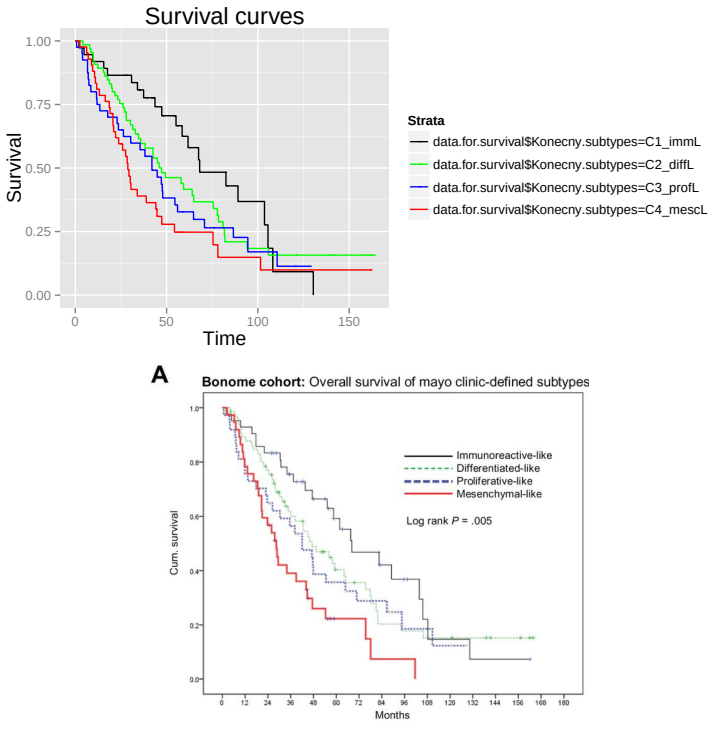


Figure 1: (Above) Survival curves the Bonome dataset using our implementation of the Konecny subtyping scheme. (Below) Corresponding Figure 3A from Konecny et al.

2 Verhaak et al., 2013 / TCGA 2011

Next, we implemented the subtype classification scheme given by Verhaak et al., 2013. The authors designed a classifier based on single-sample GSEA to classify samples into subtypes previously defined in TCGA, 2011. In their supplementary materials, the authors provide a list of four sets of gene symbols (100 total gene symbols), with each gene set associated with a subtype.

We implemented this subtype classification scheme using the provided gene sets and the ssGSEA implementation in `R` package `GSVA`. The parameters to the function `gsva` were: `method="ssgsea"`, `min.sz=10`, `tau=0.75`, which resemble default GenePattern parameters.[1]

To compare our implementation, we compared our normalized ssGSEA scores with the scores in the validation set used in the original study. In their supplementary materials, Verhaak et al. provide their normalized ssGSEA scores for a validation set consisting of the datasets of Bonome, Crijns, Denkert, Dressman, Tothill, Yoshihara, and a subset of TCGA. This validation dataset consisted of 879 patients reported in their supplementary;[2] we matched 719 patients from MetaGxOvarian data.

Due to different normalization methods, our ssGSEA scores differ in numeric range but are expected to correlate strongly with the values provided in the supplementary materials. of the 879 patients in the supplementary material, 719 were matched. We observed Pearson's correlation coefficients of 0.86, 0.92, 0.93, and 0.91 for subtypes DIF, IMR, MES, and PRO respectively.

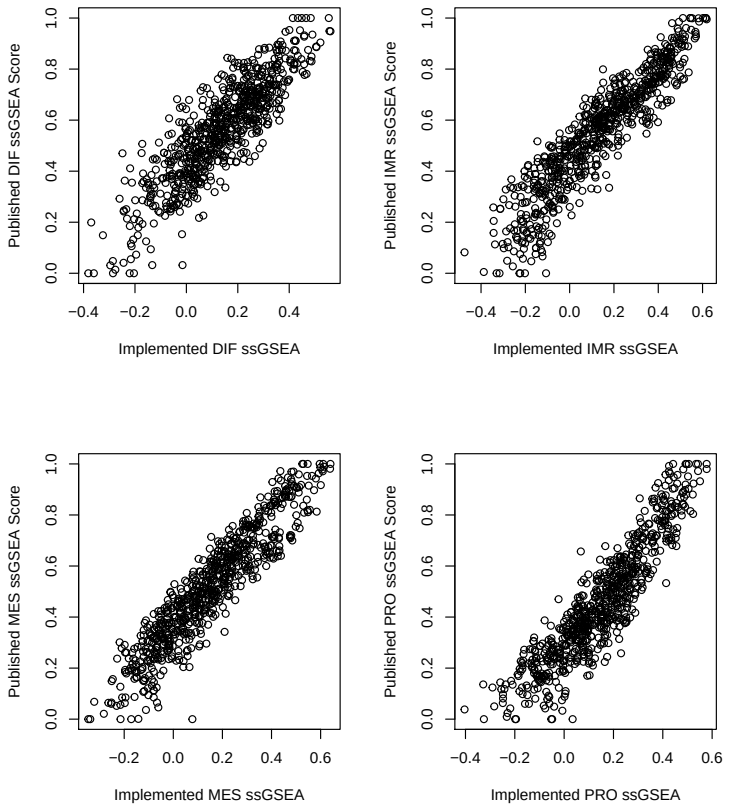


Figure 2: ssGSEA scores for the >700 patient validation set from our implementation (x axis) and the supplementary material (y axis) for each of the four subtypes.

[1] These parameters differ slightly from the function call used in Waldron et al. (2014)

[2] These numbers exclude the dataset of Crijns et al., as the patient IDs in the supplementary did not match the IDs in MetaGxOvarian.

From these normalized ssGSEA scores, a subtype classification may be performed according to one of two procedures. The first is described by Verhaak et al. and involves first classifying Immunoreactive and Mesenchymal subtypes by the higher score if corresponding ssGSEA scores exceed a given threshold.[3] Overall, this method produces a concordance of 76.36% of samples classified identically between our implementation and supplementary results.

	Verhaak Subtypes from Supplementary			
Implemented Verhaak Subtypes	DIF	IMR	MES	PRO
DIF	136	30	18	1
IMR	6	174	11	4
MES	3	13	143	0
PRO	22	20	42	96

Table 2: Contingency table showing concordance using our implementation of the Verhaak subtyping classifier and the predictions given in the supplementary of the Verhaak manuscript, using threshold values for Immunoreactive and Mesenchymal subtypes. The predictions for both implementations were made on the combined >700 sample dataset by taking the max ssGSEA subtype score.

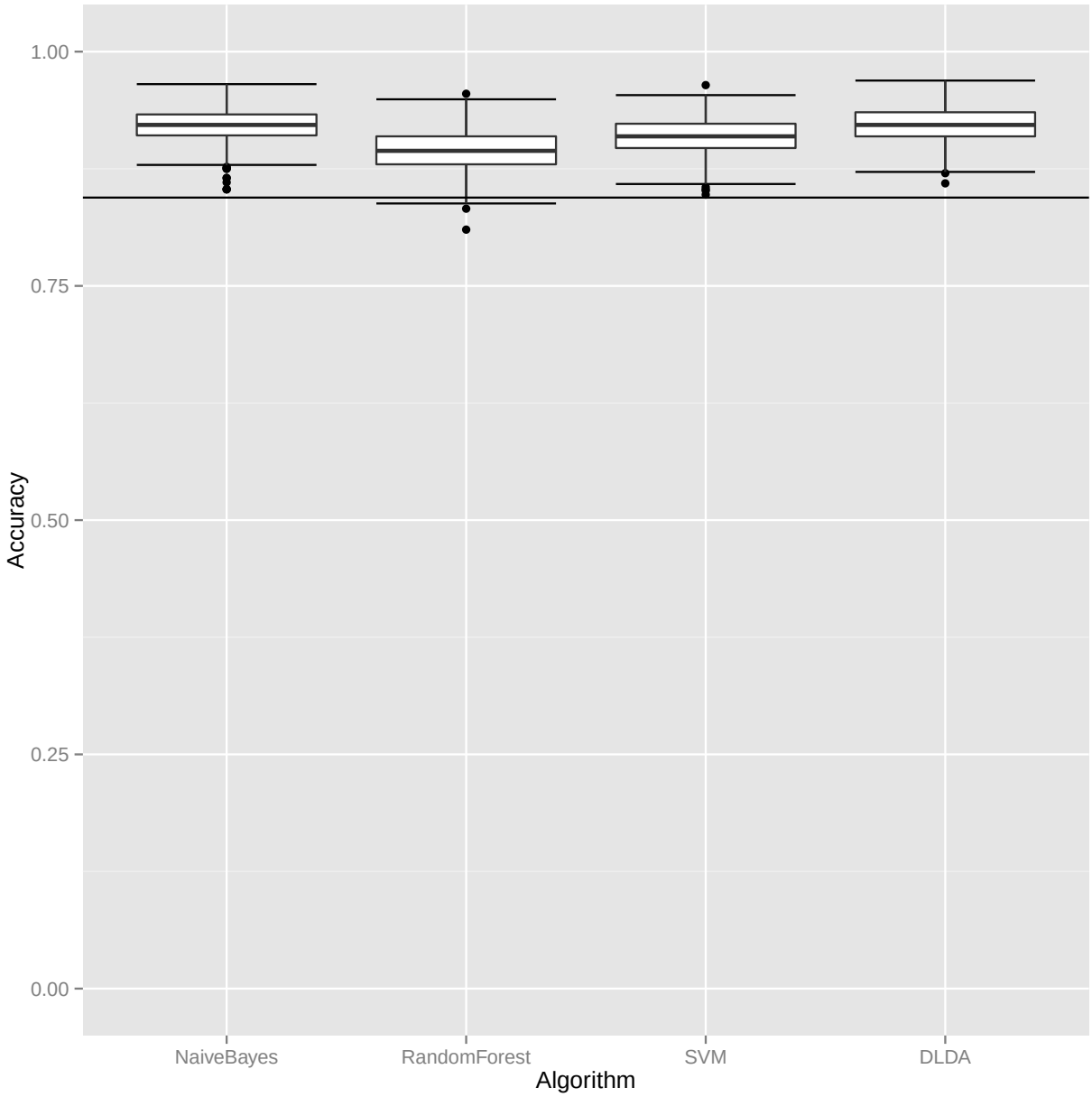
The second method is to directly assign the subtype by the max ssGSEA score. Overall, this method produces a concordance of 82.75% of samples classified identicially between our implementation and supplementary results.

	Verhaak Subtypes from Supplementary			
Implemented Verhaak Subtypes	DIF	IMR	MES	PRO
DIF	177	11	3	1
IMR	16	156	9	5
MES	17	8	126	0
PRO	22	18	14	136

Table 3: Contingency table showing concordance using our implementation of the Verhaak subtyping classifier and the predictions given in the supplementary of the Verhaak manuscript, using the max ssGSEA score for each class. The predictions for both implementations were made on the combined >700 sample dataset by taking the max ssGSEA subtype score.

In addition, we applied bootstrapping on the dataset used to define subtypes, with 1000 bootstrap replicates.

[3]The authors obtained their thresholds as the lowest ssGSEA scores for samples clustered within Immunoreactive and Mesenchymal subtypes respectively. For our implementation, we used the same TCGA train data with original cluster labels to re-compute these threshold values for each gene list used for classifying new samples.


1.00
0.75
0.50
0.25
0.00
Accuracy
NaiveBayes
RandomForest
SVM
DLDA
Algorithm

3 Helland et al., 2011 / Tothill et al., 2008

Next, we implemented the subtype classifier of Helland et al., 2011. The same group as the Tothill et al. study implemented a different classifier for their previously-described subtypes. They identified a gene list for each of their four previously-defined high-grade serous ovarian carcinoma subtypes. Using a method described in another study for breast cancer classification (Lim et al., Nat. Med. 2009), they trained a set of weights for each gene list. Classification was performed by taking a linear combination of weights and expression levels for each gene list, normalizing the scores, and classifying according to the highest-scoring subtype.

Using their published gene list and weights from the supplementary text,[4]. we implemented their subtype classifier and applied it to the TCGA dataset. The authors kindly provided a spreadsheet listing their classifier's labels on the TCGA dataset.[5] Overall, 92.49% of samples were classified identically between the authors' implementation and ours.

	Original Helland Subtypes			
Implemented Helland Subtypes	C1	C2	C4	C5
C1	122	2	3	5
C2	3	89	4	0
C4	0	8	121	2
C5	4	2	2	99

Table 4: Contingency table showing concordance of our implementation and and the predictions given by the table provided by Helland et al. Predictions were made on the TCGA dataset. Note that subtypes C3 and C6 were excluded in the original study since they are associated with non-HGS ovarian tumours.

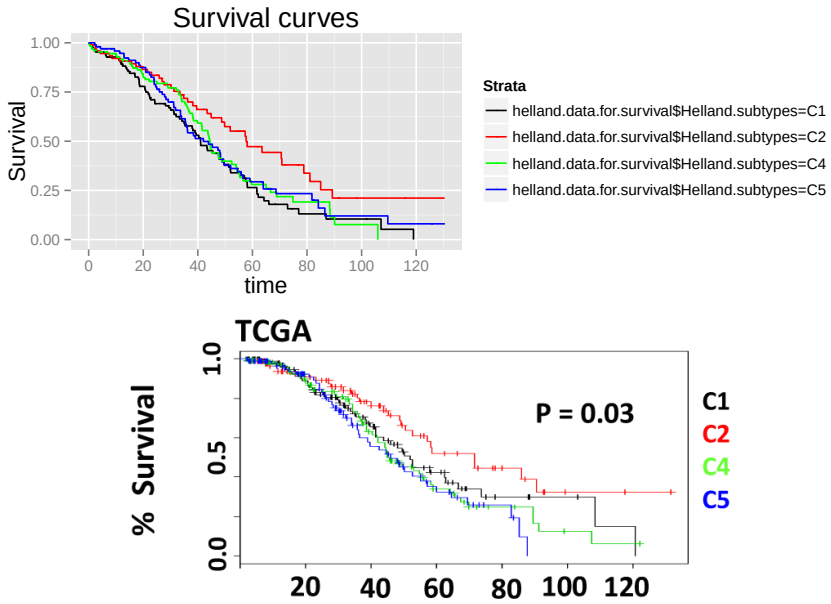


Figure 3: (Above) Survival curves the TCGA dataset using our implementation of the Helland subtyping scheme. (Below) Corresponding survival plot from Figure 1B from Helland et al.

[4]The supplementary text consists of gene symbols, Affymetrix U133 Plus 2.0 Probe IDs, and weights. In order to associate these values to a unique Entrez ID, I did the following: using Bioconductor package `annotate` and data package `hgu133plus2.db`, identify each Entrez ID based on probe ID and platform alone. Out of 1164 probe IDs in the supplementary, 1132 mapped to a unique Entrez ID, 27 mapped to multiple Entrez IDs, and 5 did not map to an Entrez ID. For the 27 genes that mapped to multiple Entrez IDs, 25 were associated to a unique Entrez ID based on the gene symbol listed in the supplementary.

[5]Of the 476 patient IDs in their spreadsheet, 466 overlapped with data from MetaGxOvarian. I am in the process of looking into this discrepancy...

4 Bentink, Haibe-Kains et al., 2012

We implemented the subtype classifier of Bentink, Haibe-Kains et al. (2012) using the function `genefu::ovcAngiogenic`, which assigned one of two subtypes based on weights of a DLDA classifier. The authors classified samples in ten independent datasets, and provide survival curves and a distribution of scores. Below are score distribution and survival plots from our implementation, using data from high-grade, late-stage serous ovarian carcinoma. Note that the distribution plots that we generated are non-parametric density estimates, whereas the distributions from the supplementary manuscript are fitted curves to normal distributions.

Validation: Tothill dataset

Our implementation of the Bentink et al. subtype classifier on the Tothill et al. dataset classified 66 patients as 'angiogenic' and 72 as 'non-angiogenic'.[6]

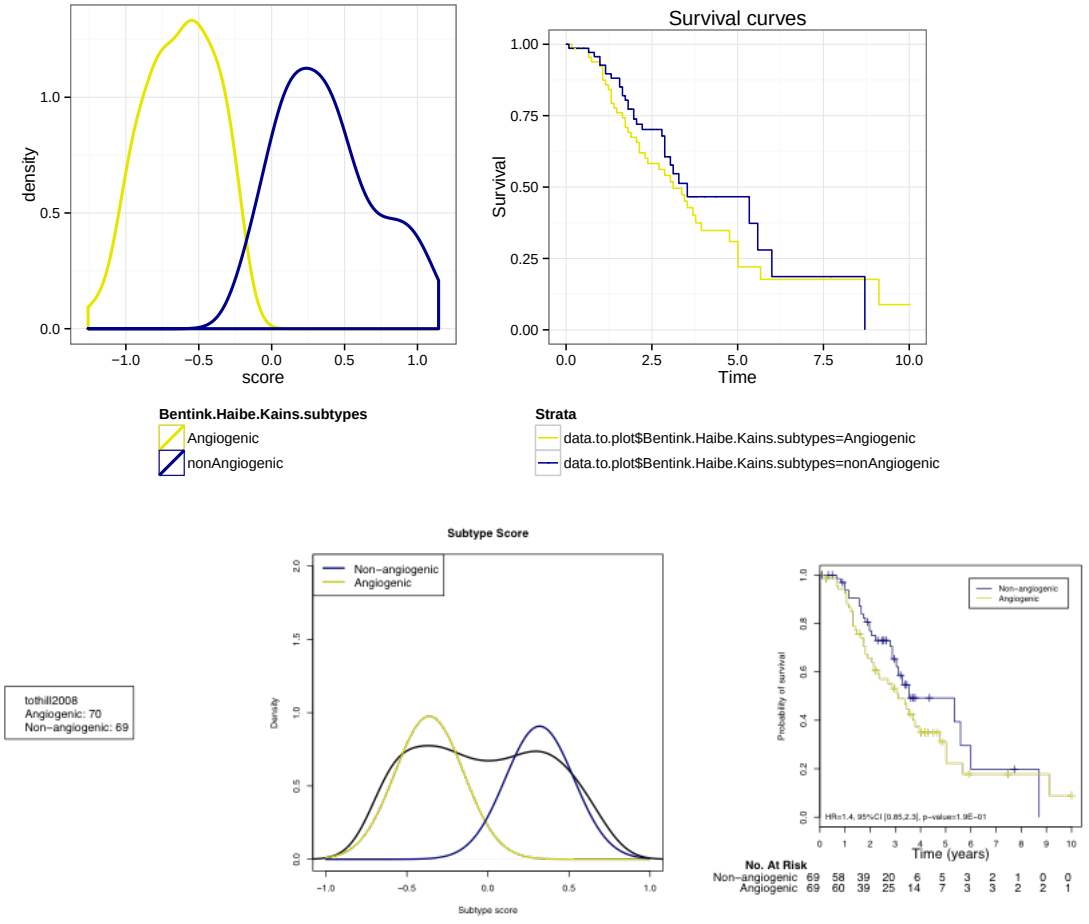


Figure 4: Tothill dataset. (Above) Score distribution and survival plot from our implementation. (Below) Score distribution fitted to a normal distribution and survival plot from the Bentink et al. supplementary.

[6]The counts and plots are similar, but not identical. Perhaps this is due to differences in gene expression normalization?

Validation: Crijns dataset

Our implementation of the Bentink et al. subtype classifier on the Crijns et al. dataset classified 36 patients as 'angiogenic' and 49 as 'non-angiogenic'.[7]

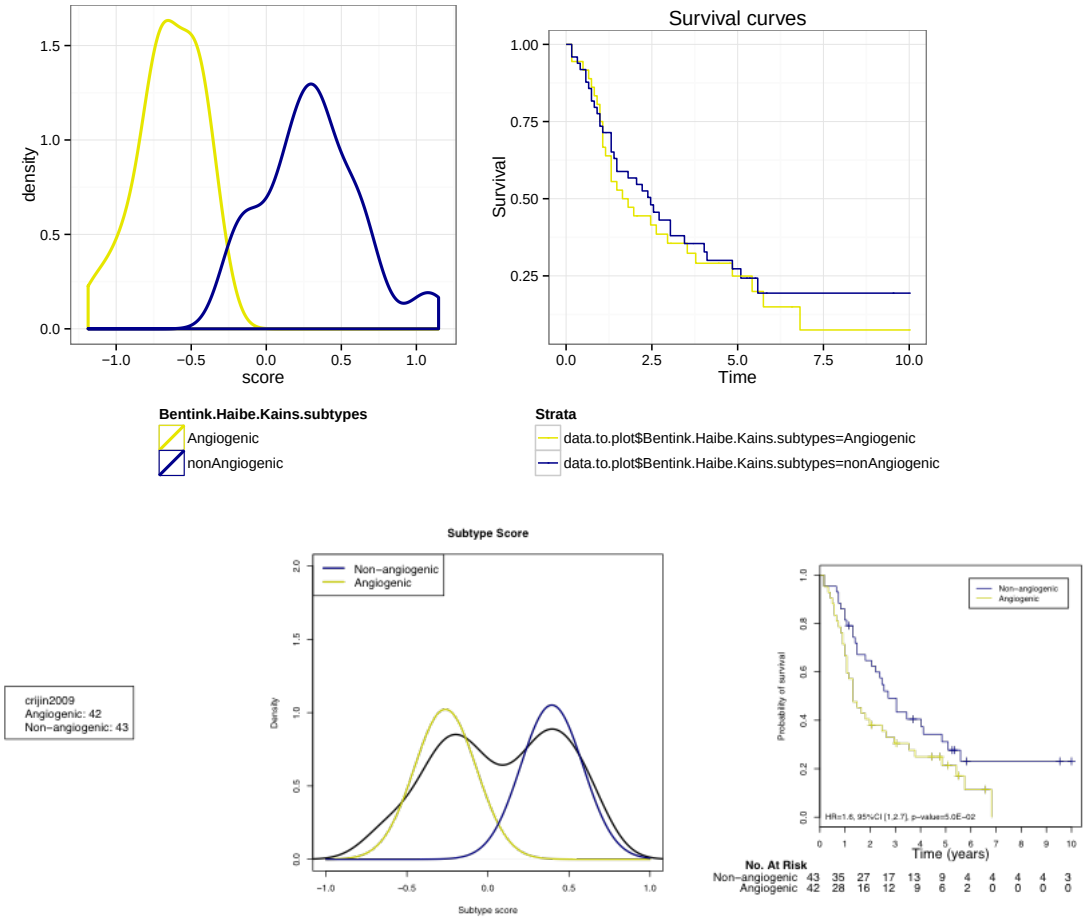


Figure 5: Crijns dataset. (Above) Score distribution and survival plot from our implementation. (Below) Score distribution fitted to a normal distribution and survival plot from the Bentink et al. supplementary.

[7]In this case, the survival curves appear to differ from the curves provided in the supplementary text.

Validation: Denkert dataset

Our implementation of the Bentink et al. subtype classifier on the Denkert et al. dataset classified 12 patients as 'angiogenic' and 29 as 'non-angiogenic'.[8]

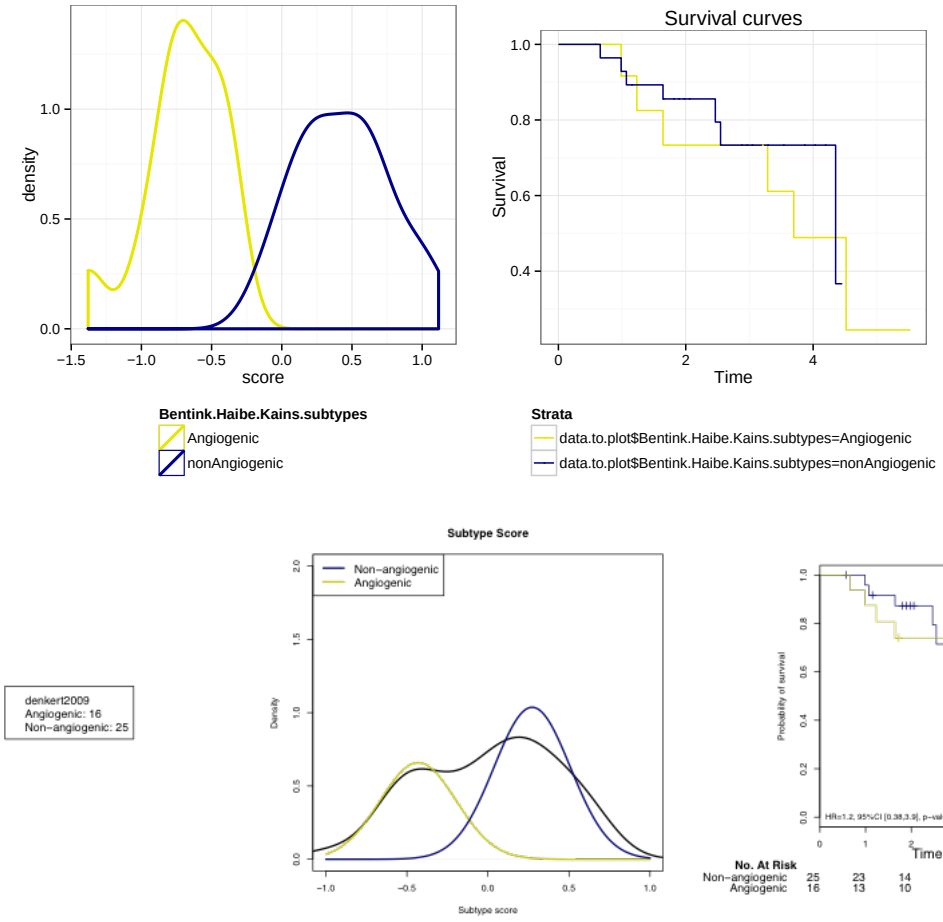


Figure 6: Denkert dataset. (Above) Score distribution and survival plot from our implementation. (Below) Score distribution fitted to a normal distribution and survival plot from the Bentink et al. supplementary.

[8]In this case, the survival curves appear to differ from the curves provided in the supplementary text.

Validation: Dressman dataset

Our implementation of the Bentink et al. subtype classifier on the Dressman et al. dataset classified 18 patients as 'angiogenic' and 37 as 'non-angiogenic'.[9]

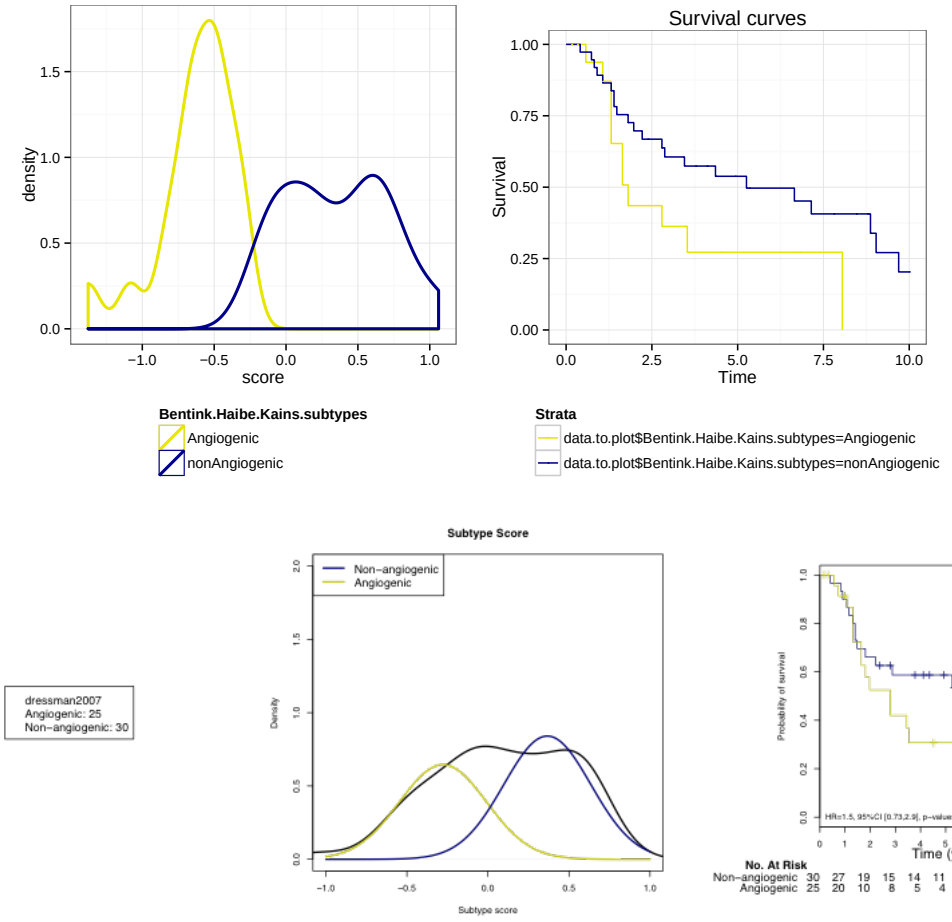


Figure 7: Dressman dataset. (Above) Score distribution and survival plot from our implementation. (Below) Score distribution fitted to a normal distribution and survival plot from the Bentink et al. supplementary.

[9]Note the different x-axis range for the survival curves, as MetaGxOvarian has access to extended survival information. Other than this, the survival plots look very similar.

Validation: Mok dataset

Our implementation of the Bentink et al. subtype classifier on the Mok et al. dataset classified 12 patients as 'angiogenic' and 41 as 'non-angiogenic'.[10]

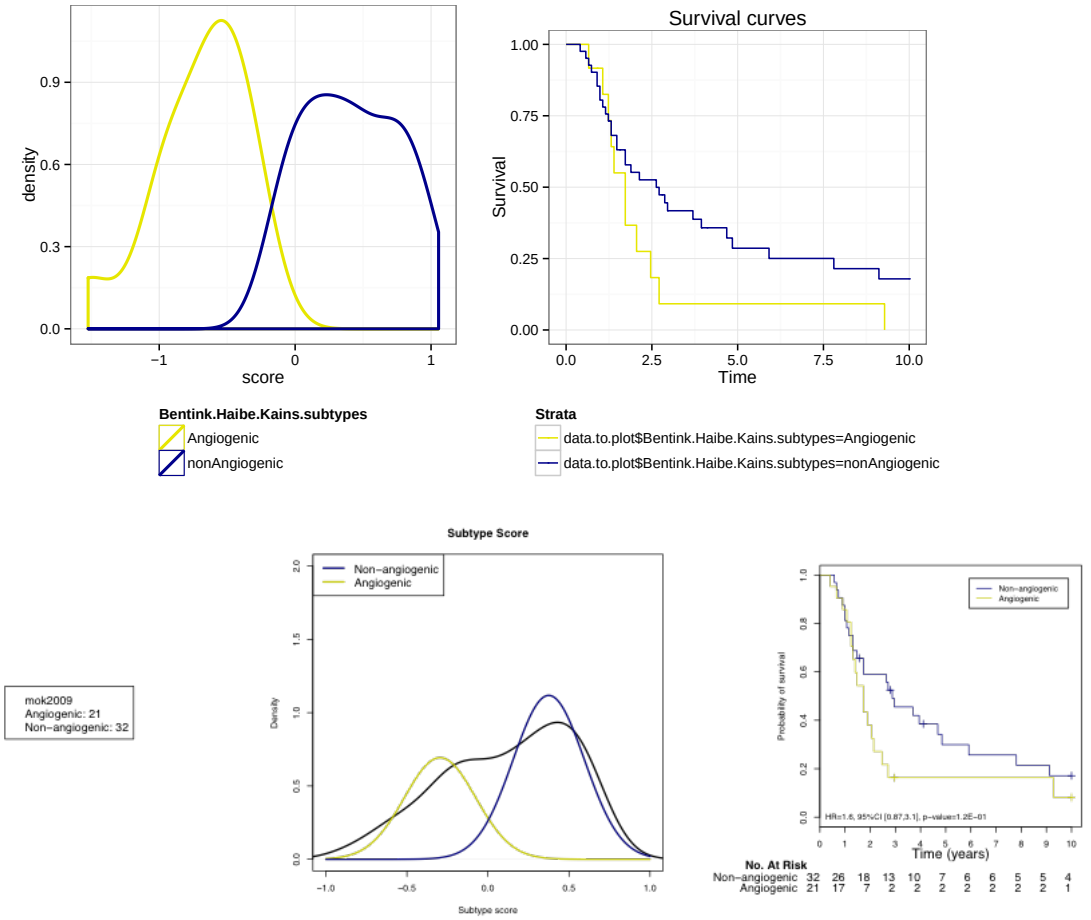


Figure 8: Mok dataset. (Above) Score distribution and survival plot from our implementation. (Below) Score distribution fitted to a normal distribution and survival plot from the Bentink et al. supplementary.

[10] In this case, the survival curves appear to differ from the curves provided in the supplementary text.

5 Appendix: Summary of Subtyping Methods

	Tothill et al., 2008.	Bentink et al., 2012.	The Cancer Genome Analysis, 2011 (Clustering), Verhaak et al., 2012 (Classification)	Konecny et al., 2014.
Probe set filtering prior to clustering	Probes with expression below 7.0, variance below 0.5 were removed; 8732 remaining probe sets.	Ratio of variance across samples / variance across replicates; selected top 1000	Top 1500 variability across patients	Top 2500 variability across patients (2040 after removing missing data)
Clustering	Consensus k-means clustering: 1000 iterations of k-means clustering. Samples which clustered together in >800 iterations were considered the 'robust' sample set (for class prediction). Using the 'robust' consensus set: use kNN and diagonal LDA to select the top n genes by LOOCV (used approx. 750). Used these genes and kNN/DLDA to classify remaining samples	Separated into training set, model selection set. Identifying splits with clear separation (ISIS) algorithm on training set, led to a large set of possible bi-partitions. To select the bi-partitions, used bootstrapping to assess the robustness of each partition; found four, selected one for biological reasons	Non-negative matrix factorization(NMF)	Non-negative matrix factorization (NMF).
Number of clusters	6 (4 serous ovarian carcinoma)	2	4	4
Probe set filtering prior to classification	Same probe sets used for clustering. The authors report the probe set identifiers common to U133A and U133Plus2.	Authors report gene names and EnsEMBL identifiers and weights; the classification algorithm may use a subset of genes that overlap with a given platform	After filtering patients with silhouette width, Significance Analysis of Microarrays (SAM) to identify gene sets of size 200 for each subtype. Used best-first search using Correlation-based Feature Subset Selection with Weka, 10-fold CV to select 100 genes.	Prediction Analysis of Microarrays (PAM) with a 1 std. dev. threshold, 10 fold CV to select 633 probe sets.
Classification	DLDA	Weighted average, using DLDA weights	The 100 genes were used to classify new samples using ssGSEA: first classified as Immunoreactive or Mesenchymal based on a cut-off, then classified into one of the four classes using ssGSEA.	Nearest centroid using Spearman's rho.
Validation	Used the Dressman dataset. Reported counts for each class, and plots of survival curves.	Report subtype score distribution and survival curves for ten independent datasets: Crijns, Denkert, Dressman, Mok, Spentzos. Zhang, TCGA, Yoshihara, Tothill, Birrer	Give class predictions for Bonome, Crijns, Denkert, Dressman, TCGA, Tothill, Yoshihara datasets	Validated on Bonome et al.